# Entwicklung und Optimierung einer bewegungssensitiven adaptiven ROI-Extraktion für die remote-Photoplethysmographie

**Development and Optimization of a motion sensitive adaptive ROI extraction for remote photoplethysmography** Masterarbeit von Philipp Witulla Tag der Einreichung: 05. Februar 2024

Gutachten: Prof. Dr.-Ing. Christoph Hoog Antink, KIS\*MED
 Gutachten: Maurice Rohr, M.Sc., KIS\*MED
 Darmstadt



TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Elektrotechnik und Informationstechnik

KIS\*MED - Künstlich Intelligente Systeme der Medizin

Studiengang Elektro- und Informationstechnik

#### Erklärung zur Abschlussarbeit gemäß § 22 Abs. 7 APB TU Darmstadt

Hiermit erkläre ich, Philipp Witulla, dass ich die vorliegende Arbeit gemäß § 22 Abs. 7 APB der TU Darmstadt selbstständig, ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt habe. Ich habe mit Ausnahme der zitierten Literatur und anderer in der Arbeit genannter Quellen keine fremden Hilfsmittel benutzt. Die von mir bei der Anfertigung dieser wissenschaftlichen Arbeit wörtlich oder inhaltlich benutzte Literatur und alle anderen Quellen habe ich im Text deutlich gekennzeichnet und gesondert aufgeführt. Dies gilt auch für Quellen oder Hilfsmittel aus dem Internet.

Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Falle eines Plagiats (§ 38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei einer Thesis des Fachbereichs Architektur entspricht die eingereichte elektronische Fassung dem vorgestellten Modell und den vorgelegten Plänen.

Darmstadt, 05. Februar 2024

P.Witulla

# Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Arbeit an meiner Master-Thesis unterstützt haben.

Ich möchte meinem Betreuer Maurice Rohr danken, für das außerordentlich aufregende Thema dieser Arbeit, das mein Interesse an der Forschung zur künstlichen Intelligenz in der Medizintechnik aufs Neue entfacht hat, sowie für seine Unterstützung und den konstruktiven Anregungen, die zum Erfolg dieser Arbeit maßgeblich beigetragen haben.

Weiterhin möchte ich meinen Eltern danken, die mein Studium erst ermöglicht und mich dabei stets unterstützt haben.

Ebenso möchte ich allen Versuchsteilnehmern danken, die der Forschung ihre Zeit gewidmet haben und so ermöglichten, das Ziel dieser Arbeit erreichen zu können.

Ich möchte diese Gelegenheit nutzen, der Python-Community und den Entwicklern der Open-Source-Pakete meinen Dank auszusprechen, die es ermöglicht haben, eine leistungsfähige Umgebung zur wissenschaftlichen Datenanalyse bereitzustellen und das Forschungsgebiet der künstlichen Intelligenz und der remote-Photoplethysmographie für jeden Interessierten zugänglich gemacht haben.

# Abstract

In this work, an adaptive Region of Interest (ROI) extraction for the segmentation of facial regions from facial videos has been developed and implemented into a system for extracting blood volume pulse signals using established algorithms of remote photoplethysmography (rPPG).

At first, a data pre-processing pipeline has been developed utilizing the MediaPipe library, which allows the segmentation of the facial region based on configurable parameters. The developed data pre-processing steps were integrated into the data processing pipeline of the existing rPPG-Toolbox. Subsequently, this data processing pipeline was then deployed to conduct studies using six publicly available rPPG datasets (UBFC-rPPG, COHFACE, VIPL-HR-V1, PURE, MMPD, RLAP-rPPG). This involved the extension of the rPPG-Toolbox with additional data loading programs for the RLAP, COHFACE and VIPL-HR-V1 datasets.

Furthermore, a separate KISMED dataset was created as part of this work. The objective of this dataset is to capture challenging cases with known interference factors for signal reconstruction using rPPG algorithms. When designing the data set, special care was taken to simulate real world situations and to include the environmental influences that occur. The data set will be made available to the institute for Artificial Intelligent Systems in Medicine at the Technical University of Darmstadt for future internal research work.

Investigations were performed on the rPPG datasets using six unsupervised rPPG algorithms (ICA, POS, CHROM, GREEN, LGI, PBV) to identify significant facial regions that contain strong pulsatile blood volume changes and are therefore particularly suitable for extracting a blood volume pulse signal. The results indicate that the restriction of the facial area to the regions of the forehead and cheeks yield the best prediction performance for the unsupervised rPPG algorithms.

The optimal forehead and cheek ROI provided the foundation for the input data of an existing neural network architecture (*TS-CAN* (Temporal Shift-Convolutional Attention Network)), which was trained using the PURE and UBFC-rPPG dataset. In order to evaluate the effect of the face segmentation, the self-implemented face extraction was compared with the pre-implemented data processing within the rPPG toolbox. The evaluation using the motion-rich scenarios in the unknown datasets results in an average reduction of the MAE by 3, 49 for the *TS-CAN* model trained with the PURE dataset when segmenting the facial area into the forehead and cheek ROI. The RMSE decreases by 3, 50, the Pearson correlation coefficient increases by 0, 1397, the SNR by 2,36 dB and the Accuracy by 10, 71 %.

Comparable improvements can be achieved for the model trained with the UBFC-rPPG dataset by applying ROI segmentation. The MAE decreases on average by 2,06 and the RMSE by 3,71. The Pearson correlation coefficient increases by 0,1219, the SNR by 2,99 dB and the Accuracy by 10,05 %.

Finally, the attention masks of the trained models were analyzed. The visualization demonstrates how the *TS-CAN* model autonomously learns to focus on meaningful facial areas on the forehead, cheeks and nose for signal extraction. In summary, face segmentation has made a significant contribution to improving the prediction performance of rPPG algorithms, as evidenced by the average improvements in the evaluation metrics.

# Zusammenfassung

In dieser Arbeit ist eine adaptive Extraktion von *Regions of Interest* (ROI) zur Segmentierung von Gesichtsbereichen aus Gesichtsvideos entwickelt und in ein System zur Extraktion von Blutvolumenpulssignalen durch bewährte Algorithmen der remote-Photoplethysmographie (rPPG) implementiert worden.

Zunächst wurde unter Verwendung der MediaPipe-Bibliothek eine Datenvorverarbeitungspipeline entwickelt, die es ermöglicht, den Gesichtsbereich anhand von definierbaren Parametern zu segmentieren. Die entwickelten Datenvorverarbeitungsschritte wurden in die Datenverarbeitungspipeline der bestehenden rPPG-Toolbox integriert. Die Datenverarbeitungspipeline wurde anschließend dazu eingesetzt, um mithilfe von sechs für die Forschung zugänglichen rPPG-Datensätzen (UBFC-rPPG, COHFACE, VIPL-HR-V1, PU-RE, MMPD, RLAP-rPPG) Untersuchungen durchzuführen. Hierzu wurde die rPPG-Toolbox um zusätzliche Datenladeprogramme für den RLAP-, COHFACE- und VIPL-HR-V1-Datensatz erweitert.

In Ergänzung dazu wurde im Rahmen dieser Arbeit ein eigener KISMED-Datensatz erstellt. Dieser hat das Ziel, Härtefälle mit bekannten Störfaktoren für die Signalrekonstruktion durch rPPG-Algorithmen aufzuzeichnen. Bei der Konzipierung des Datensatzes wurde darauf geachtet, real vorkommende Alltagssituationen nachzubilden und die dabei vorkommenden Umwelteinflüsse miteinzubeziehen. Der Datensatz wird dem Institut für künstlich intelligente Systeme in der Medizin der Technischen Universität Darmstadt für zukünftige interne Forschungsarbeiten zur Verfügung gestellt.

Mithilfe der rPPG-Datensätze wurden Untersuchungen anhand von sechs unüberwachten rPPG-Algorithmen (ICA, POS, CHROM, GREEN, LGI, PBV) durchgeführt, um die aussagekräftigsten Gesichtsregionen zu ermitteln, welche ausgeprägte pulsatile Blutvolumenänderungen enthalten und sich damit besonders für die Extraktion eines Blutvolumenpulssignals eignen. Die Resultate haben gezeigt, dass die Einschränkung des Gesichtsbereichs in die Regionen der Stirn und Wangen zur besten Vorhersageleistung der unüberwachten rPPG-Algorithmen führt.

Die optimale Stirn- und Wangen-ROI diente anschließend als Grundlage für die Eingangsdaten einer bestehenden neuronalen Netzwerkarchitektur [*TS-CAN* (Temporal Shift-Convolutional Attention Network)], welche mithilfe des PURE- und UBFC-rPPG-Datensatzes trainiert wurde. Um die Auswirkung der vorgenommenen Gesichtssegmentierung beurteilen zu können, wurde die selbst implementierte Gesichtsextraktion mit der vorimplementierten Datenverarbeitung innerhalb der rPPG-Toolbox verglichen. Bei der Evaluation anhand der bewegungsreichen Szenarien in den unbekannten Datensätzen ergibt sich beim *TS-CAN*-Modell, das mit dem PURE-Datensatz trainiert wurde, eine durchschnittliche Verringerung des MAE um 3, 49, wenn die Segmentierung des Gesichtsbereichs in die Stirn- und Wangen-ROI vorgenommen wird. Der RMSE verringert sich um 3, 50, der Pearson-Korrelationskoeffizient steigt um 0, 1397, das SNR um 2,36 dB und die Accuracy um 10, 71 %. Beim Modell, das mit dem UBFC-rPPG-Datensatz trainiert wurde, können vergleichbare Verbesserungen durch den Einsatz der ROI-Segmentierung erzielt werden. Der MAE sinkt durchschnittlich um 2,06 und der RMSE um 3,71. Der Pearson-Korrelationskoeffizient steigt um 0,1219, das SNR um 2,99 dB und die Accuracy um 10,05 %.

Zuletzt wurden die Aufmerksamkeitsmasken der trainierten Modelle betrachtet. Die Visualisierung zeigt eindrücklich, wie das *TS-CAN*-Modell selbstständig lernt, den Fokus zur Signalextraktion auf aussagekräftige Gesichtsbereiche an Stirn, Wangen und Nase zu legen. Zusammenfassend hat die Gesichtssegmentierung einen signifikanten Beitrag zur Verbesserung der Vorhersageleistung von rPPG-Algorithmen geleistet, wie durch die durchschnittlichen Verbesserungen der Metriken deutlich geworden ist.

# Abbildungsverzeichnis

1.	Ablaufdiagramm zum Erreichen des Forschungsziels	4
2. 3. 4.	Hautfarbenskala nach Fitzpatrick (Abbildung angelehnt an [17]) Absorptionskoeffizient der Epidermis bei variierenden Volumenanteilen von Melanin Molarer Extinktionskoeffizient von Hämoglobin in Abhängigkeit der Wellenlänge	8 9 9
5. 6. 7. 8.	Funktionsweisen der Transmissions-PPG und Reflexions-PPG	11 11 12 16
9. 10. 11.	Beispielbilder der Beleuchtungssituationen im COHFACE-Datensatz	21 27 30
<ol> <li>12.</li> <li>13.</li> <li>14.</li> <li>15.</li> <li>16.</li> <li>17.</li> <li>18.</li> <li>19.</li> <li>20.</li> <li>21.</li> <li>22.</li> </ol>	Tiefpassfilterung der x-Koordinaten des Gesichtspunkts an der Nasenspitze Darstellung der Gesichtsparkettierung mit farblich hervorgehobenen Reflexionswinkeln Anwendung einer histogrammbasierten Hautsegmentierungsmaske	34 36 37 38 41 42 42 43 44 45
23. 24. 25. 26. 27.	Bei der Untersuchung des Winkelschwellenwerts extrahierte Dreiecke Bei der Untersuchung zur ROI-Begrenzung extrahierte Gesichtsbereiche Vergleich zwischen verschiedenen Ansätzen zur Segmentierung des Gesichtsbereichs Vergleich der gemittelten Reflexionswinkel über alle Videos in den (Teil-)Datensätzen Vergleich der gemittelten ROI-Pixelflächen über alle Videos in den (Teil-)Datensätzen	50 54 56 61 63
28. 29. 30. 31. 32.	Trainings- und Validierungsverlust über eine Trainingslaufzeit von mehreren Epochen.Anpassung der Lernrate während des Trainings.Struktur der Eingangsdaten für das TS-CAN-Modell.Trainings- und Validierungsverlust während des Trainings mit dem PURE-Datensatz.Visualisierung der Aufmerksamkeitsmasken im TS-CAN-Modell (entnommen und angepasst	70 71 72 73
	aus [35])	74

33.	Bland-Altman-Diagramme der Differenz zwischen den vorhergesagten Herzraten und den tatsächlichen Herzraten über dem Mittelwert der beiden Messgrößen	77
34.	Bland-Altman-Diagramme bei der Evaluation der beiden mit dem PURE-Datensatz trainierten	
	TS-CAN-Modelle am unbekannten UBFC-rPPG-Datensatz.	107
35.	Bland-Altman-Diagramme bei der Evaluation der beiden mit dem UBFC-rPPG-Datensatz	
	trainierten TS-CAN-Modelle am unbekannten PURE-Datensatz.	108
36.	Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbe-	
	kannten COHFACE-Datensatz.	109
37.	Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbe-	
	kannten VIPL-HR-V1-Datensatz.	110
38.	Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbe-	
	kannten MMPD-Datensatz	111
39.	Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbe-	
	kannten RLAP-rPPG-Datensatz.	112
40.	Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbe-	
	kannten KISMED-Datensatz.	113

# Tabellenverzeichnis

1.	Leistungsvergleich der vorgestellten unüberwachten und überwachten Modelle.	18
2. 3. 4. 5.	Verteilung der Merkmale im MMPD-Datensatz [52][53]	22 24 29 31
<ol> <li>6.</li> <li>7.</li> <li>8.</li> <li>9.</li> <li>10.</li> <li>11.</li> <li>12.</li> </ol>	Vergleich der Begrenzung des Gesichtsbereichs anhand unterschiedlicher Winkelschwellen- werte für den UBFC-rPPG-Datensatz	51 52 53 55 57 62 64
13. 14. 15.	Evaluation des Modells nach dem Training mit dem PURE-Datensatz über 30 Epochen Evaluation des Modells nach dem Training mit dem UBFC-rPPG-Datensatz über 30 Epochen. Vergleich der Evaluationsmetriken in Abhängigkeit des Hauttyps nach Fitzpatrick für das Rotationsszenario im MMPD-Datensatz	75 76 78
16. 17. 18.	Datenblatt des <i>Contec Medical CMS-50E</i> -Fingerclippulsoximeters [70] Datenblatt des <i>Pulox PO-250</i> -Fingerclippulsoximeters [71] Vergleich der Begrenzung des Gesichtsbereichs anhand unterschiedlicher Winkelschwellen-	92 93
<ol> <li>19.</li> <li>20.</li> <li>21.</li> <li>22.</li> <li>23.</li> <li>24.</li> <li>25.</li> <li>26.</li> <li>27.</li> <li>28.</li> </ol>	werte für den UBFC-rPPG-Datensatz	94 95 96 97 98 98 99 99 99 99
29.	Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für den MMPD-Datensatz.	100

30.	Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Bewegungsszenario im RLAP-rPPG-Datensatz.	100
31.	Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Rotationsszenario im	101
วา	NISMED-Datellisatz	101
32. 33.	Vergleich zur Segmentierung des Gesichtsbereichs bei einer natürlichen Sonnenlichtbestrah-	101
	lung im COHFACE-Datensatz.	101
34.	Vergleich zur Segmentierung des Gesichtsbereichs bei einer künstlichen Beleuchtung im	
	COHFACE-Datensatz.	102
35.	Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im VIPL-HR-V1-	
	Datensatz	102
36.	Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im PURE-Datensatz	z.102
37.	Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im MMPD-	
	Datensatz	103
38.	Vergleich zur Segmentierung des Gesichtsbereichs beim Videospielszenario im RLAP-rPPG-	
	Datensatz	103
39.	Vergleich zur Segmentierung des Gesichtsbereichs beim Rotationsszenario im KISMED-	
	Datensatz	103
40.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs für den	
	UBFC-rPPG-Datensatz.	104
41.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs bei der	
	künstlichen Beleuchtung im COHFACE-Datensatz.	104
42.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs bei der	
	natürlichen Sonnenlichtbestrahlung im COHFACE-Datensatz	105
43.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs beim Bewe-	
	gungsszenario im VIPL-HR-V1-Datensatz.	105
44.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs beim Bewe-	
	gungsszenario im PURE-Datensatz.	105
45.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs beim Szenario	
	nach sportlicher Aktivität im MMPD-Datensatz.	106
46.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs beim Video-	
	spielszenario im RLAP-Datensatz.	106
47.	Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs beim Rotati-	
	onsszenario im KISMED-Datensatz.	106

# Inhaltsverzeichnis

1.	Motivation	1
	1.1. Formulierung der Forschungsfrage	2
	1.2. Gliederung der Arbeit	3
2.	Physiologische und anatomische Grundlagen	5
	2.1. Herz-Kreislauf-System	5
	2.1.1. Herzrate	6
	2.2. Aufbau der Haut	7
	2.2.1. Einfluss der Hautfarbe	7
3.	Stand der Technik	10
	3.1. Hintergrund der kameragestützten Photoplethysmographie	10
	3.1.1. Photoplethysmographie	10
	3.1.2. Remote-Photoplethysmographie	13
	3.2. Signalverarbeitungsalgorithmen	14
	3.3. rPPG-Verfahren auf Basis neuronaler Netze	15
	3.4. Auswahl aussagekräftiger Gesichtsbereiche	18
4.	Übersicht über bestehende rPPG-Datensätze	20
	4.1. UBFC-rPPG	20
	4.2. COHFACE	20
	4.3. VIPL-HR-V1	21
	4.4. PURE	21
	4.5. MMPD	22
	4.6. RLAP	22
	4.7. Entwicklung eines eigenen KISMED-Datensatzes	25
	4.7.1. Messsetup und verwendete Hardware	25
	4.7.2. Durchführung der Datenerhebung	26
	4.7.3. Versuchsszenarien für Härtefälle und Negativbeispiele	26
	4.7.4. Informationen zum aufgezeichneten KISMED-Datensatz	28
	4.7.5. Auswertung der Signalqualität	28
	4.7.6. Diskussion	29
5.	Datenverarbeitungsschritte	32
	5.1. Erweiterung der bestehenden rPPG-Toolbox zur Datenverarbeitung	32
	5.2. Tiefpasssfilterung der extrahierten Landmarks durch MediaPipe	33
	5.3. Berechnung der Reflexionswinkel zwischen Hautoberfläche und Kamera	34
	5.4. Histogrambasierte Hautsegmentierung/Gesichtshaarentfernung	35
	5.5. Segmentierung der Region of Interest	38

Ab	okürzungsverzeichnis	114
Α.	Anhang	91
9.	Fazit und Ausblick         9.1. Ausblick	<b>81</b> 83
8.	Diskussion	79
7.	Entwicklung der rPPG auf Basis neuronaler Netze         7.1. Einführung in den Entwicklungsprozess neuronaler Netze         7.1.1. Einführung in die Architektur neuronaler Netze         7.1.2. Trainingsprozess         7.1.3. Optimierungsprozess         7.2. Struktur der Eingangsdaten         7.3. Training des TS-CAN-Modells         7.4. Evaluation anhand unbekannter Testdatensätze	67 67 68 69 71 72 75
6.	5.7. Nachbearbeitung der extrahierten BVP-Signale         5.8. Evaluationsverfahren         5.8. Evaluationsverfahren         6.1. Untersuchung des optimalen Winkelschwellenwertes         6.2. Untersuchung zur Hautsegmentierung         6.3. Untersuchung zur Begrenzung der ROI         6.4. Gewichtete Mittelung mehrerer extrahierten Herzraten         6.5. Untersuchung der optimalen Zeitfenstergröße         6.6. Diskussion und Fazit	43 45 48 48 51 54 57 62 66
	5.6. Interpolation der Gesichtswinkel	40

### 1. Motivation

Die Messung von Vitaldaten des menschlichen Körpers ist ein essenzieller Bestandteil in der Diagnostik von Gesundheitsbeschwerden. Insbesondere Vitaldaten wie die Herzrate, Sauerstoffsättigung und Herzratenvariabilität bieten wichtige Einblicke in den Zustand des menschlichen Körpers. Eine Hürde bei der Messung der Vitaldaten im Alltag stellt die Messsensorik dar, welche je nach Messregion die Anschaffung stationärer Geräte voraussetzt und den Patienten dadurch an den Ort der Messeinrichtung bindet. Aus diesem Grund wird intensiv an Konzepten geforscht, welche sich für den Einsatz zur Vermessung menschlicher Vitaldaten aus der Ferne eignen. Als nichtinvasive und mittlerweile mobil einsetzbare Methode eignet sich die Photoplethysmographie (PPG) dazu, Vitaldaten wie den Herzschlag oder die Blutsauerstoffsättigung zu ermitteln.

Bei der remote-Photoplethysmographie (rPPG) hingegen können die Vitaldaten über eine Kamera erfasst werden, die das Gesicht oder andere Körperregionen aus der Ferne betrachtet. Die Entwicklung von nichtinvasiven Methoden wie die rPPG ermöglicht es, diese Daten ohne den Einsatz stationärer Geräte zu erfassen, was besonders in der häuslichen Umgebung von Vorteil ist. Mit den wachsenden Fortschritten der Digitalisierung im Gesundheitswesen gewinnt die Telemedizin zunehmend an Bedeutung. Seit Beginn der COVID-19-Pandemie stieg der Anteil der gesetzlich krankenversicherten Patienten in Deutschland, die telemedizinische Leistungen erhalten haben, von 0, 2% im Jahr 2017 auf 1,9% im Jahr 2021 [1]. Dies entspricht einem Anstieg um 823,6%, von 163.825 telemedizinischen Konsultationen im Jahr 2017 auf 1.349.208 Konsultationen im Jahr 2021. Der Einsatz der rPPG kann hier zukünftig einen wertwollen Beitrag zur Unterstützung der Ferndiagnostik leisten.

Zudem finden Rückschlüsse aus den Vitaldaten Anwendung im Bereich der psychologischen Erforschung von Stress, Müdigkeit und des allgemeinen Emotionsbefindens. Die Erkenntnisse können zukünftig in Bereichen wie der Arbeitsplatzsicherheit oder im Verkehrswesen hilfreich dabei sein, Maschinen- und Anlagenbedienern oder Autofahrern eine Pause nahezulegen [2][3]. Ein weiteres Einsatzszenario, das sich seit dem Aufkommen von realistisch wirkenden, allerdings künstlich erstellten Deep-Fake-Videos abzeichnet, ist es, diese auf ihre Echtheit und Authentizität hin zu überprüfen [4][5].

Obwohl derzeit noch kein kommerziell verfügbares Gerät auf Grundlage der rPPG eine Zulassung im medizinischen Anwendungsbereich besitzt [6], findet sie bereits Anwendung in Smart Devices, welche die Gesundheitsüberwachung für Endverbraucher zugänglich machen. Mit dem *Anura® MagicMirror* wurde ein Gerät vorgestellt, der es dem Nutzer ermöglichen soll, anhand von rPPG-Messungen des gefilmten Gesichts Vitalparameter wie die Herzrate, Herzratenvariabilität oder die Atemfrequenz zu messen [7]. Ein Nachteil für den Anwender besteht darin, dass für akkurate Messungen eine optimale Ausleuchtung des Gesichts notwendig ist und der Anwender während des Messzeitraums in Ruhe verharren muss.

Die vorgestellten Einsatzszenarien stellen nur einen Auszug der potenziellen Anwendungen der rPPG im Alltag dar. Damit zeichnet sich bereits ein hohes Potenzial für die Anwendung der rPPG in medizinischen und nichtmedizinischen Bereichen ab. Insbesondere für die Zulassung im medizinischen Anwendungsbereich sind allerdings noch mehrere Herausforderungen zu bewältigen. Zu den Schwierigkeiten gehört die Gewährleistung der Zuverlässigkeit von rPPG-Messungen, insbesondere unter variablen Umweltbedingungen. Verschiedene Faktoren können die Messergebnisse beeinflussen, wie die Lichtverhältnisse, Hautfarbe und Bewegungen [8]. Daraus entwickelt sich die Notwendigkeit, standardisierte Validierungsverfahren zu entwickeln, die den Anforderungen medizinischer Zulassungsbehörden gerecht werden.

Diese Arbeit fokussiert sich auf die Auslegung einer robusteren Herzratenvorhersage in alltäglichen Situationen und insbesondere bei bewegten Personen. Die konkrete Forschungsfrage wird nachfolgend vorgestellt.

#### 1.1. Formulierung der Forschungsfrage

Im Folgenden werden die zentralen Problemstellungen vorgestellt, die im Rahmen dieser Arbeit untersucht werden:

- 1. Welcher Gesichtsbereich enthält die aussagekräftigsten pulsinduzierten Informationen, die eine Extraktion der Herzrate begünstigen?
- 2. Ist eine Anpassung der Region of Interest (ROI) bei Bewegungen notwendig beziehungsweise vorteilhaft für die Signalextraktion durch rPPG-Algorithmen?
- 3. Trägt die Integration der ROI-Segmentierung in ein neuronales Netz dazu bei, zu jedem Zeitpunkt den aussagekräftigsten Gesichtsbereich zu selektieren und damit die Genauigkeit der vorhergesagten Herzrate zu erhöhen?

Das Gesamtziel dieser Arbeit ist es, ein bestehendes neuronales rPPG-Modell auf Basis von Convolutional Attention Networks (CAN) um eine adaptive ROI-Selektion zu erweitern. Durch den Einsatz einer Attention Mask (Aufmerksamkeitsmaske) bietet das CAN die Möglichkeit, den Fokus eines weiteren bildverarbeitenden neuronalen Netzes auf aussagekräftige Gesichtsbereiche zu lenken. Die Idee der Aufmerksamkeitsmaske ist es, den Hautbereichen mit stärkeren Signalen eine höhere Gewichtung zuzuweisen.

Durch eine laufende Aktualisierung der untersuchten Gesichtsbereiche soll dies insbesondere bei Bewegungen und Rotationen des Kopfes eine robustere Signalextraktion ermöglichen. Hierfür sollen zunächst anhand von Literaturrecherchen und Experimenten geeignete ROIs im Gesicht des gefilmten Probanden definiert werden. Anschließend sollen bei Bewegungen der Probanden laufend Aktualisierungen der ROIs stattfinden. Zudem soll untersucht werden, ob die Signalqualität des extrahierten PPG-Signals durch den Fokus auf Gesichtsbereiche mit geringem Winkel zwischen Haut und Kamera zunimmt. Diese Erkenntnisse sollen in die Entwicklung der bewegungsabhängigen adaptiven ROI-Selektion einfließen, welche zuletzt mit einem bestehenden End-to-End-Netzwerk verglichen wird (Abbildung 1).

#### 1.2. Gliederung der Arbeit

Einen Schwerpunkt dieser Arbeit stellt die Entwicklung einer geeigneten Datenverarbeitung für Videodaten dar, die es ermöglicht, den Gesichtsbereich anhand von definierbaren Parametern in ROIs zu segmentieren. Um dies zu erreichen, werden in Kapitel 2 zunächst relevante physiologische und anatomische Grundlagen für das Verständnis der Arbeit erläutert. Darauf aufbauend wird in Kapitel 3 die Funktionsweise und bestehende Herausforderungen der rPPG verdeutlicht. Weiterhin werden bestehende rPPG-Signalverarbeitungsmethoden und Forschungsergebnisse zur Auswahl aussagekräftiger Gesichtsbereiche vorgestellt.

Auf Grundlage der Signalverarbeitungsalgorithmen werden Untersuchungen durchgeführt, die zur Klärung der Forschungsfrage führen. Hierfür ist es zusätzlich notwendig, Datensätze heranzuziehen, die sowohl Videos von Probanden, als auch deren Blutvolumenpulssignale als Vergleichswert für die Evaluation der rPPG enthalten. Kapitel 4 bietet eine Übersicht über sechs für die Forschung zugänglichen rPPG-Datensätze. Bei der Begutachtung der Datensätze hat sich die Notwendigkeit herausgestellt, einen eigenen Datensatz aufzuzeichnen, der insbesondere Härtefallszenarien für die rPPG-Signalextraktion beinhaltet. Die Anforderungen an den Datensatz, die Gestaltung der Probandentests und die Methodik zur Durchführung werden in Abschnitt 4.7 beschrieben.

Die vorgestellten Datensätze werden im weiteren Verlauf der Arbeit dazu eingesetzt, anhand von bewährten rPPG-Algorithmen und geeigneten Evaluationsmetriken Untersuchungen durchzuführen. Mithilfe der Untersuchungen wird aus dem gesamten Gesichtsbereich die Region of Interest ermittelt, die die aussagekräftigste Extraktion eines Blutvolumenpulssignals ermöglicht. Hierfür sind zunächst einige Schritte zur Vorverarbeitung der Datensätze notwendig, deren Implementierungen in Kapitel 5 beschrieben werden.

Einen weiteren Schwerpunkt dieser Arbeit stellen die Untersuchungen dar, die zur Ermittlung der optimalen ROI in realen Einsatzszenarien führen. Anhand der Bewertungsmetriken sollen die Auswirkungen der untersuchten Einschränkungen im Gesichtsbereich für die rPPG-Signalextraktion evaluiert werden. Diese werden in Kapitel 6 beschrieben und ausgewertet. Die Erkenntnisse aus den Untersuchungen führen schrittweise zur Definition des aussagekräftigsten Gesichtsbereichs für die Signalextraktion. Der optimale Gesichtsbereich wird dabei insbesondere mit dem Fokus auf bewegungsreiche Einsatzszenarien ausgelegt.

In Kapitel 7 definiert dieser Gesichtsbereich die Ausgangsbasis für die Aufmerksamkeitsmaske des neuronalen Netzwerks, welches dadurch den Fokus der Signalextraktion auf den ermittelten Gesichtsbereich lenken soll. Dabei werden die in dieser Arbeit implementierten Datenvorverarbeitungsschritte der vorimplementierten Gesichtsextraktion aus der rPPG-Toolbox gegenübergestellt und verglichen. Zuletzt werden die entstehenden Aufmerksamkeitsmasken betrachtet, um die interne Signalverarbeitung der neuronalen Netze zu durchleuchten.



Abbildung 1.: Ablaufdiagramm zum Erreichen des Forschungsziels: Die eckigen Blöcke stellen größere Schwerpunkte dar. Die abgerundeten Blöcke stellen notwendige Schritte oder Untersuchungen innerhalb des Schwerpunktes dar.

# 2. Physiologische und anatomische Grundlagen

In diesem Kapitel werden die physiologischen und anatomischen Grundlagen erläutert, die für das Verständnis der Arbeit notwendig sind. Zunächst werden die physiologischen Grundlagen des Herz-Kreislauf-Systems und die optischen Eigenschaften der Haut als Messregion dargelegt.

#### 2.1. Herz-Kreislauf-System

Das menschliche Herz ist ein lebenswichtiges Organ, das als Antrieb des Kreislaufsystems fungiert. Seine primäre Funktion besteht darin, Blut durch den Körper zu pumpen, um Gewebe und Organe mit Sauerstoff und Nährstoffen zu versorgen. Die Blutgefäße dienen als Transportwege, um das sauerstoffarme Blut in der Lunge mit Sauerstoff anzureichern und anschließend Organe und Gewebe mit sauerstoffreichem Blut zu versorgen. In den feinsten Blutgefäßen der Gewebe, den Kapillaren, erfolgt der Sauerstoff- und Nährstoffaustausch mit dem umliegenden Gewebe. Von den Kapillaren wird das sauerstoffarme Blut über die Venen wieder zum Herzen befördert. Im nächsten Kreislaufzyklus wird das Blut wieder in der Lunge mit Sauerstoff angereichert, bevor es zum Herzen zurückgeleitet und dann in den restlichen Körper transportiert wird [9].

Der Puls beschreibt die Druck- und Volumenschwankungen der Blutgefäße, welche durch den Herzschlag hervorgerufen werden. Die Pulswelle ist eine wellenförmige Bewegung des Blutes in den Arterien. Sie beginnt mit einem Anstieg des Drucks, wenn das Herz kontrahiert und Blut in die Aorta pumpt (systolischer Druck), und einem Abfall des Drucks, wenn das Herz sich entspannt (diastolischer Druck). Angetrieben durch den systolischen Druck breitet sich die Pulswelle von der Aorta aus und bewegt sich durch das Arteriensystem. Dabei gelangt sie in immer kleinere Arterien und schließlich in die feinen Kapillaren, wo der Gasaustausch im Gewebe stattfindet.

Pulswellen besitzen Eigenschaften, die analog zu denen akustischer Wellen sind. So kann eine Pulswelle durch ihre Frequenzanteile und Amplitude beschrieben werden [10]. Die Pulswelle setzt sich aus überlagerten Schwingungen unterschiedlicher Frequenzen zusammen. Bei Betrachtung des Frequenzspektrums einer Pulswelle stellt die Grundschwingung die Herzrate dar. Bei Reflexionen an Gefäßwänden kann es zu Überlagerungen von fortschreitenden und reflektierten Wellen kommen, sodass sich die Pulswellen innerhalb eines Gefäßsystems gegenseitig verstärken oder dämpfen können. Die Einflüsse der Gefäßwände auf die Pulswelle sind im Frequenzspektrum als Vielfache der Grundschwingung zu erkennen.

Bei der Ausbreitung von Pulswellen im menschlichen Gefäßsystem kann zwischen drei verschiedenen Arten unterschieden werden: einem Druckpuls durch Blutdruckänderungen, einem Strompuls durch Änderungen der Blutströmung und einem Querschnittspuls (Volumenpuls) durch Querschnitts- beziehungsweise Volumenänderungen der Blutgefäße [11]. Das im Rahmen dieser Arbeit verwendete Messprinzip der Photoplethysmographie beruht auf der Erfassung und Analyse des Blutvolumenpulses (BVP).

#### 2.1.1. Herzrate

Die Herzrate gibt die Anzahl der Herzschläge pro Minute an. Sie wird in Schlägen pro Minute [*engl.:* beats per minute (BPM)] gemessen. Sie stellt eine grundlegende Messgröße für die Aktivität des Herzens dar und ist von entscheidender Bedeutung für die Regulation des Blutflusses im gesamten Kreislaufsystem des menschlichen Körpers. Die Herzrate unterliegt verschiedenen Einflüssen, einschließlich des autonomen Nervensystems, hormoneller Regulation und äußerer Faktoren wie dem Alter, körperlicher Aktivität, Stress und der Temperatur [11].

Während Ruhe und Entspannung ist die Herzrate typischerweise niedriger, während sie bei körperlicher Anstrengung und in Stresssituationen ansteigt, um die gesteigerte Versorgung von Sauerstoff und Nährstoffen zu gewährleisten. Bei einem gesunden erwachsenen Menschen misst die Herzrate in Ruhe 60 bis 100 BPM. Eine verringerte Herzrate wird als Bradykardie, eine erhöhte Herzrate als Tachykardie bezeichnet. Von einer Bradykardie spricht man, wenn die Ruhefrequenz des Herzen unterhalb von 60 BPM liegt. Bei einer Ruhefrequenz über 100 BPM liegt eine Tachykardie vor. Unter lebensbedrohlichen Umständen, wie dem Kammerflimmern oder Kammerflattern, kann die Herzrate Extremwerte von über 250 BPM erreichen [11][12].

Bei größtmöglicher körperlicher Anstrengung kann die Herzrate eines gesunden Menschen vorübergehend Maximalwerte von 160 bis 220 BPM erreichen. Die maximale Herzrate hängt von mehreren Faktoren, wie dem Alter, Geschlecht und körperlicher Verfassung ab. Sie wird näherungsweise bestimmt über die Formel:

$$HR_{max} = 220 - Lebensalter \tag{2.1}$$

Nach Thews et al. [11] lässt sich die momentane Herzrate  $HR_{momentan}$  aus dem Zeitintervall  $\Delta t_{RR}$  (in Sekunden) zwischen zwei aufeinanderfolgenden Herzschlägen (R-R-Intervall) in der Anzahl der Schläge pro Minute berechnen:

$$HR_{momentan} = \frac{60 \frac{s}{\min}}{\Delta t_{RR}}$$
(2.2)

Die momentane Herzrate ist zu unterscheiden von der üblicherweise angegebenen mittleren Herzrate. In der Regel wird die mittlere Herzrate HR aus der Anzahl der Herzschläge N über eine Zeitspanne  $\Delta t_{Mittelung}$  angegeben:

$$HR = \frac{N}{\Delta t_{Mittelung}} \tag{2.3}$$

Pulsmessungen geben üblicherweise die über einen Zeitraum von einer Minute gemittelte Herzrate an.

#### 2.2. Aufbau der Haut

Die Haut ist mit einer Fläche von bis zu zwei Quadratmetern das größte Organ des Menschen und erfüllt zahlreiche wichtige Funktionen. Sie besteht aus drei Schichten: der Epidermis (Oberhaut), der Dermis (Lederhaut) und der Hypodermis (Unterhaut). Diese Schichten bilden eine Schutzbarriere für den Körper und sind für verschiedene physiologische Prozesse verantwortlich. Die Haut wird von einem Netzwerk aus Arterien und Venen in der Hypodermis durchblutet (siehe Abbildung 6). Diese Durchblutung ist entscheidend für die Versorgung der Hautzellen mit Sauerstoff und Nährstoffen. Darüber hinaus spielt die Haut eine wichtige Rolle bei der Regulierung der Körpertemperatur, indem sie durch Schwitzen und Durchblutungsanpassungen die Wärmeabgabe reguliert [9].

In Bezug auf die nichtinvasive Durchblutungsmessung ist die Haut eine geeignete Region, da sie eine hohe Dichte an Blutgefäßen an der Körperoberfläche aufweist. Dies ermöglicht die Ableitung von Vitalparametern wie die Herzrate und Sauerstoffsättigung, was in der medizinischen Diagnostik und Forschung von Nutzen ist. Insbesondere das Gesicht weist eine hohe Hautdurchblutung auf und hat sich in bestehenden rPPG-Forschungen als geeignete Region für nichtinvasive Messungen des Blutvolumenpulses bewährt [13][14].

Die Gesichtshaut ist im Allgemeinen dünner im Vergleich zur Haut an anderen Körperstellen. Die Dicke der Gesichtshaut beträgt in der Regel etwa 0,7 bis 1,5 mm. Im Gesicht ermöglicht eine dünnere Haut die feinere Kontrolle über die Mimik. Die Haut ist besonders dick in Regionen, wo sie mechanischen Belastungen und Reibung ausgesetzt ist, wie den Handflächen und Fußsohlen. Dies gewährleistet eine höhere Widerstandsfähigkeit und Schutzwirkung gegenüber Verletzungen. Darüber hinaus gibt es Unterschiede in der Hautdicke zwischen den Geschlechtern, wobei Männer tendenziell eine um 10% bis 20% dickere Gesichtshaut haben als Frauen [15][16].

Das Signal-zu-Rausch-Verhältnis (SNR) bei der remote-Photoplethysmographie (rPPG) wird von mehreren Faktoren beeinflusst, darunter die Dicke der Hautschichten. Je dünner die Hautschichten sind, desto höher ist in der Regel das SNR, da Licht leichter in tiefer gelegene Hautschichten eindringen kann, um das Blutvolumen in den Gefäßen zu erfassen. Da diese Schichten typischerweise im Gesicht besonders dünn sind, weist das rPPG-Signal hier potenziell ein besseres SNR auf als an anderen Körperstellen.

Allerdings hängt das SNR nicht nur von der Hautdicke ab, sondern auch von der Hautpigmentierung. Hautpigmente wie Melanin absorbieren Licht und können die Intensität des innerhalb der Haut reflektierten Lichts verringern. Dadurch wird die Erfassung eines BVP-Signals durch den rPPG-Algorithmus beeinträchtigt [8]. Im folgenden Abschnitt wird eine Einteilung der Bevölkerung in Abhängigkeit des Melaningehalts in der Epidermis vorgestellt, sowie die Auswirkung, die dieser auf die rPPG-Signalextraktion hat.

#### 2.2.1. Einfluss der Hautfarbe

Die Hautfarbenskala nach Fitzpatrick unterteilt die Hautfarben in sechs Kategorien, welche auf ihrer jeweiligen Reaktion auf UV-Licht beruhen (Abbildung 2). Typ I entspricht einer hellen weißen Hautfarbe, Typ II einer hellen Hautfarbe, Typ III einer dunkleren weißen Hautfarbe, Typ IV einer hellbraunen Hautfarbe, Typ V einer braunen Hautfarbe und Typ VI einer dunkelbraunen oder schwarzen Haut. Unter den sechs Hauttypen nimmt die Melanin-Pigmentierung mit jedem Hauttyp zu, was negativ mit der Wahrscheinlichkeit

korreliert, bei UV-Bestrahlung Strahlungsschäden wie Sonnenbrand zu erleiden. Hauttyp I neigt mit seinem vergleichsweise geringen Melanin-Gehalt am ehesten dazu, Schäden durch UV-Strahlung zu erleiden, wohingegen Hauttyp VI am wenigsten zu Hautschäden durch UV-Strahlung neigt [17][18].



Abbildung 2.: Hautfarbenskala nach Fitzpatrick (Abbildung angelehnt an [17]). Der Gehalt an epidermalen Melanin in der obersten Hautschicht (Epidermis) nimmt mit jedem Hautfarbentyp zu. Damit zusammenhängend nimmt das Risiko an strahlungsbedingten Hauterkrankungen zu erleiden mit steigendem Melaningehalt ab.

Gemäß früherer Studien schwankt der Melanin-Volumenanteil in der Epidermis im Durchschnitt zwischen 1% und 5% innerhalb der Bevölkerung mit Fitzpatrick-Typ I und II, zwischen 10% und 20% innerhalb der Bevölkerung mit Fitzpatrick-Typ III und IV und zwischen 20% und 40% innerhalb der Bevölkerung mit Fitzpatrick-Typ V und VI [19][20]. Melanin absorbiert Licht in Abhängigkeit seiner Wellenlänge unterschiedlich stark (Abbildung 3). Die Lichtabsorption im nahem Infrarotbereich (NIR) mit einer Wellenlänge von über 780 nm ist niedriger als im Bereich des sichtbaren Lichts zwischen 380 nm und 780 nm.

Licht im NIR-Bereich dringt daher tiefer in pigmentierte Haut ein. Zahlreiche PPG-Geräte verwenden jedoch grünes Licht mit einer Wellenlänge von etwa 550 nm. Dies ist auf die vergleichsweise hohe Lichtabsorption durch Hämoglobin in diesem Wellenlängenbereich zurückzuführen (Abbildung 4). Der reflektierte Lichtanteil nimmt dabei mit steigendem Blutvolumen absorptionsbedingt ab. Durch diese Absorption können Veränderungen im Blutvolumen erfasst und als pulsierendes Signal, dem Blutvolumenpuls (BVP), gemessen werden. Die unterschiedlichen spektralen Eigenschaften von sauerstoffgesättigtem und sauerstoffarmem Hämoglobin werden zur Bestimmung der Blutsauerstoffsättigung durch die Pulsoximetrie ausgenutzt. Eine detaillierte Beschreibung dieses Funktionsprinzips folgt in Kapitel 3.1.1.

Im nächsten Kapitel werden die Funktionsweise, die messtechnischen Grundlagen und die praktischen Anwendungen der PPG-Messung näher erläutert. Weiterhin werden die spezifischen Herausforderungen und Algorithmen zur Signalextraktion durch die rPPG vorgestellt.



Abbildung 3.: Absorptionskoeffizient der Epidermis bei variierenden Volumenanteilen von Melanin. Die Volumenanteile repräsentieren die sechs verschiedenen Hauttypen nach Fitzpatrick (entnommen und übersetzt aus [21]).



Abbildung 4.: Molarer Extinktionskoeffizient von sauerstoffgesättigtem Hämoglobin (HbO<sub>2</sub>) und sauerstoffarmem Hämoglobin (Hb) in Abhängigkeit der Wellenlänge (in Anlehnung an Zhang et al. [22]). Sauerstoffarmes Hämoglobin weist ein Absorptionsmaximum bei 555 nm auf. Sauerstoffgesättigtes Hämoglobin weist zwei lokale Absorptionsmaxima bei Wellenlängen von 541 nm und 577 nm auf [11].

# 3. Stand der Technik

In diesem Kapitel werden die messtechnischen Grundlagen zur Erfassung der Herzaktivität durch die PPG erläutert und der Stand der Technik zur Erfassung der kamerabasierten rPPG vorgestellt. Anschließend werden für diese Arbeit relevante Forschungsarbeiten beschrieben und bestehende Probleme bei der rPPG-Messung verdeutlicht. Zuletzt werden einige öffentlich zugängliche Datensätze vorgestellt, die im Rahmen dieser Arbeit für die rPPG-Analyse verwendet werden.

#### 3.1. Hintergrund der kameragestützten Photoplethysmographie

Die kameragestützte Messung von Vitaldaten hat sich zu einem vielversprechenden Bereich innerhalb der computergestützten Diagnostik und der Computer Vision entwickelt. Im Bereich der Telemedizin ist die Messung von Vitaldaten aus der Ferne ein wichtiges Instrument für die Untersuchung des Patienten und die Feststellung der Diagnose. Im Gegensatz zu invasiven Verfahren, die eine Punktion der Haut oder des Körpers erfordern, erlauben nichtinvasive kameragestützte Messverfahren die Erfassung von Vitaldaten ohne Eingriffe in den Körper des Patienten. Dies reduziert das Risiko von Infektionen, Unannehmlichkeiten und Komplikationen, die mit invasiven Verfahren verbunden sein können.

Außerdem sind Kameras heutzutage eine allgegenwärtige Form von Sensoren, die in fast jedem digitalen Kommunikationsgerät (Mobiltelefon, PC, etc.) vorhanden sind. Dies erlaubt es, die Vitalparameter auch außerhalb spezialisierter medizinischer Einrichtungen zu überwachen, aufzuzeichnen und gegebenenfalls zur Diagnostik heranzuziehen. Im Folgenden werden nichtinvasive Verfahren vorgestellt, die es ermöglichen, Vitalparameter wie die Herzrate eines Patienten in Echtzeit zu überwachen.

#### 3.1.1. Photoplethysmographie

Die Photoplethysmographie ist ein optisches Messverfahren, das es erlaubt, die Herzaktivität anhand von Blutvolumenschwankungen innerhalb der hypodermalen Blutgefäße messtechnisch zu erfassen. Bei der herkömmlichen PPG liegen Lichtquelle und Photosensor auf der Haut an, um pulsbedingte Veränderungen der Lichtabsorption durch die Blutgefäße zu erfassen. Hierbei wird anhand der Positionierung von Lichtquelle und Photosensor zwischen Transmissions- und Reflexions-PPG unterschieden (Abbildung 5).

Bei der Transmissions-PPG befinden sich Lichtquelle und Photosensor an gegenüberliegenden Stellen, wodurch der Photosensor das durch das Gewebe transmittierte Licht empfängt. Bei der Reflexions-PPG befinden sich Lichtquelle und Photosensor auf derselben Seite des Gewebes, wobei der Sensor die reflektierte Lichtintensität misst. Die kamerabasierte rPPG ähnelt in ihrer Funktionsweise der Reflexions-PPG, mit dem



Abbildung 5.: PPG Funktionsweisen: Transmissions-PPG (links) und Reflexions-PPG (rechts) (Abbildung angelehnt an [23]). Jedes PPG-Gerät besteht in seiner Minimalausstattung aus einer Lichtquelle und einem Photosensor.

Unterschied, dass Lichtquelle und Photosensor nicht mehr auf der Haut aufliegen.

In vielen praktischen Anwendungen wie tragbaren PPG-Geräten kommt die Reflexions-PPG zum Einsatz. Dies ist zum Teil auf die kompakte Bauweise zurückzuführen, da Lichtquelle und Photosensor nebeneinander verbaut an derselben Hautregion anliegen. Hierbei wird der Umstand ausgenutzt, dass die Haut bis zu den Blutgefäßen durchlässig für Licht ist. Nach dem Hautreflexionsmodell nach Wang et al. [24] wird das Licht zum einen Teil an der Hautoberfläche spiegelnd reflektiert und dringt zum anderen Teil in die Haut ein, wo das Licht an den pulsierenden Blutgefäßen diffus reflektiert wird (Abbildung 6).



Abbildung 6.: Hautreflexionsmodell, bei dem Umgebungslicht zum einen Teil an der Hautoberfläche spiegelnd reflektiert wird und zum anderen Teil in die Haut eindringt und diffus reflektiert wird. Nur die diffuse Reflexion enthält Informationen über den Pulsschlag [24].

Wie bereits in Kapitel 2.2.1 beschrieben entsteht der ausschlaggebende Effekt für die Absorption des diffus reflektierten Lichts durch die Lichtabsorption durch den Blutfarbstoff Hämoglobin. Mit steigendem Blutvolumen in den hypodermalen Blutgefäßen nimmt der Anteil des diffus reflektierten Lichts ab. Durch die pulsbedingten Schwankungen des Blutdrucks verändert sich das Volumen der Blutgefäße, wodurch sich die Intensität des diffus reflektierten Lichtstrahls analog zum Puls ändert. Durch einen Photosensor können die Blutvolumenschwankungen als Veränderungen der Lichtintensität gemessen werden (Abbildung 7). Aus dem gemessenen PPG-Signalverlauf lassen sich anschließend Vitaldaten wie die Herzrate oder die Herzratenvariabilität ableiten [24].

Auf dem gleichen Funktionsprinzip basiert die Pulsoximetrie. Hierbei sind die Lichtquellen und der optische Sensor meist in einen Fingerclip verbaut, um Störeinflüsse durch Umgebungslicht zu reduzieren. Zur Lichterzeugung werden alternierend zwei monochromatische LEDs eingesetzt, eine rote LED mit einer Wellenlänge von 660 nm und eine Infrarot-LED mit Wellenlängen zwischen 905 nm bis 920 nm. Durch die Analyse der Lichtabsorption bei den beiden Wellenlängen kann das Gerät die Sauerstoffsättigung im Blut berechnen. Das sauerstoffreiche Hämoglobin absorbiert Licht bei diesen Wellenlängen im Vergleich zu sauerstoffarmem Hämoglobin unterschiedlich stark (siehe Abbildung 4).



Abbildung 7.: Variation in der Lichtabsorption durch schwankendes Blutvolumen in den Kapillaren [25]. Die gemessene PPG-Wellenform umfasst die zeitveränderlichen pulsatilen Anteile, die durch den Herzschlag verursacht werden und annähernd zeitkonstante Anteile, die durch die Gewebestruktur der Haut verursacht werden. Die Grundfrequenz des pulsatilen Anteils repräsentiert die Herzrate.

Durch den Photosensor kann dann die Schwankung der transmittierten oder reflektierten Lichtintensitäten detektiert und in ein elektrisches Signal umgewandelt werden. Aus dem Verhältnis der gemessenen Lichtintensitäten kann die Sauerstoffsättigung ( $SpO_2$ ) im Blut in Echtzeit bestimmt werden. Die Pulsoximetrie ermittelt also den Prozentsatz des Hämoglobins im Blut, der mit Sauerstoff beladen ist. Zusätzlich erhält man bei der Pulsoximetrie auch das Photoplethysmogramm, womit gleichzeitig die momentane Herzrate bestimmt werden kann.

PPG und Pulsoximetrie bieten damit eine nichtinvasive Methode zur Überwachung der Herzaktivität. Die auf Photosensoren basierenden Messverfahren unterliegen Einschränkungen in ihrer Genauigkeit, die auf Störeinflüsse aus der Umgebung und Einflüsse durch individuelle physiologische Unterschiede der Patienten beruhen [8]. Zu den Störeinflüssen und Fehlerquellen bei der PPG-Messung gehören:

- **Bewegungen:** Bewegungen des Patienten können die PPG-Messungen stark beeinflussen. Bereits minimale Bewegungen können Artefakte erzeugen und zu gestörten Signalen führen, was zu Ungenauigkeiten in den ermittelten Vitalparametern führen kann.
- **Fremdlicht:** Starke Lichteinwirkung von externen Lichtquellen oder direktes Sonnenlicht können die Differenzierung des diffus reflektierten Lichts vom spiegelnd reflektierten Licht an der Hautoberfläche beeinträchtigen.
- **Hautpigmentierung:** Unterschiedliche Hautpigmentierungen der Patienten führen zu variierenden Eigenschaften in der Lichtabsorption durch die Haut (siehe Abschnitt 2.2.1). Mit einem höheren Anteil an epidermalem Melanin werden die in die Haut eindringenden Lichtanteile zunehmend absorbiert, wodurch der informationstragende Signalanteil und damit das SNR reduziert wird.
- **Durchblutungsstörungen:** Die Intensität des reflektierten Lichts hängt von der Blutmenge in den hautnahen Blutgefäßen ab. Zustände wie niedriger Blutdruck, periphere vaskuläre Krankheiten oder eine schlechte Durchblutung können die Durchblutung der Hypodermis verringern. Dies kann dazu führen, dass die pulsbedingte Volumenänderung der hypodermalen Blutgefäße verringert wird oder der informationstragende Signalanteil mit anderen Störeinflüssen überlagert wird.
- Individuelle anatomische Unterschiede: Die anatomische Struktur der Hautgewebe unterscheidet sich zwischen den Patienten. Zum Beispiel variieren die Dicke der Hautschichten oder die Flexibilität der Blutgefäße in Anhängigkeit von Geschlecht, Alter und körperlicher Verfassung. Damit zusammenhängend variiert auch die individuelle Ausprägung der Blutvolumenpulse in der Gesichtshaut.

Für einen umfassenden Überblick über die Störfaktoren und Fehlerquellen bei der PPG-Messung und deren spezifische Einflüsse auf die Signalqualität sei auf die Arbeit von Fine et al. verwiesen [8]. Viele der identifizierten Störfaktoren und Fehlerquellen bei der PPG-Messung sind auch auf die rPPG-Messung übertragbar. Trotz des potenziellen Vorteils der rPPG durch die kontaktlose Erfassung von Vitalparametern können hierbei also ähnliche Herausforderungen hinsichtlich der Signalqualität auftreten. Im Folgenden wird die Funktionsweise der kamerabasierten rPPG-Messung erläutert, sowie technische Vorkehrungen, um die Auswirkungen der Störeinflüsse auf die Signalqualität gering zu halten.

#### 3.1.2. Remote-Photoplethysmographie

Im Gegensatz zur PPG liegen Lichtquelle und Photosensor bei der remote-Photoplethysmographie (rPPG) nicht mehr auf der Haut an. Bei der rPPG-Messung werden Schwankungen des Blutvolumenpulses (BVP) in der Haut über eine Kamera erfasst, die das Gesicht oder andere Körperregionen aus der Ferne betrachtet [26]. Als Lichtquelle wird das Umgebungslicht genutzt, als Photosensor eine handelsübliche Kamera.

Die spektrale Zusammensetzung des Umgebungslichts wird nach dem Hautreflexionsmodell als konstant und der Ort der Lichtquelle als statisch angenommen [24]. Die von der Kamera aufgenommene Intensität hängt von der Entfernung der Lichtquelle zum Hautgewebe und zum Kamerasensor ab. Die von der Kamera aufgenommene Hautfarbe ändert sich mit der Zeit aufgrund von bewegungsbedingten Lichtintensitätsschwankungen und pulsbedingten Blutvolumenschwankungen. Insbesondere die durch Beleuchtungsänderung oder Bewegung induzierten Störeinflüsse sind problematisch für die Qualität des gemessenen BVP-Signals. Durch Beleuchtungsänderungen oder Bewegungen erhalten die statischen Anteile der Oberflächenreflexion einen dynamischen und ebenfalls zeitveränderlichen Charakter. Dies wird besonders problematisch, wenn die Störkomponenten Frequenzanteile im Bereich der Herzrate oder deren Harmonische aufweisen [8]. Die Herausforderung in der Extraktion eines BVP-Signals aus Kameraaufzeichnungen besteht darin, geeignete

Signalverarbeitungstechniken anzuwenden, um diese Störungen zu reduzieren und nur den pulsbedingten Anteil aus den Schwankungen der Hautfarbe zu isolieren.

#### 3.2. Signalverarbeitungsalgorithmen

Die Anfänge der rPPG-Forschung beschäftigten sich mit der Entwicklung geeigneter optoelektronischer Messmethoden und Signalverarbeitungsalgorithmen zur Erfassung rhythmischer Schwankungen der Hautdurchblutung [27][28]. Mit dem Aufkommen und der allgegenwärtigen Verbreitung von Digitalkameras für Endverbraucher wurde der Fokus der rPPG-Forschung zunehmend auf den Einsatz von Kameras als Photosensoren gelegt. Gegenstand der weiteren rPPG-Forschung war nun die Entwicklung von Signalverarbeitungsalgorithmen, um die BVP-Schwankungen aus Videos extrahieren zu können.

In Abschnitt 2.2.1 wurde bereits beschrieben, dass sich aufgrund des ausgeprägten Messeffekts im grünen Wellenlängenbereich besonders RGB-Kameras als Sensoren für rPPG eignen. Verkruysse et al. [29] stellten 2008 einen Algorithmus vor, durch den das BVP-Signal anhand des zeitlichen Verlaufs des grünen Farbkanals aus den Bildsequenzen des Videos bestimmt werden konnte (GREEN-Algorithmus). Die Einzel-Kanal-Methode erzielt im Durchschnitt die schlechteste Leistung, was darauf zurückzuführen ist, dass keine Maßnahmen unternommen werden, um Störungen durch die Kombination von Signalen aus verschiedenen Farbsensoren zu beseitigen. Dies deutet darauf hin, dass es vorteilhaft ist, von den Informationen aller Farbkanäle zu profitieren, insbesondere wenn die Pulsdynamik in den verschiedenen Kanälen ungleichmäßig verteilt ist. Selbst wenn ein Kanal keine pulsatilen Informationen enthält, kann er dennoch als Rauschsensor verwendet werden, um eine Methode zu entwickeln, die von diesem Rauschen unabhängig ist [24]. Technisch ausgefeiltere Algorithmen verfolgen diesen Ansatz, wie die Unabhängigkeitsanalyse [*engl.:* Independent Component Analysis (ICA)], welche später zur Forschung an chrominanzbasierten Verfahren führten, wie CHROM oder die Plane-Orthogonal-to-Skin (POS) Methode.

Die ICA [30] ist ein statistisches Verfahren aus der Blind-Source-Separation zur Berechnung unabhängiger Komponenten in einer Mischung statistisch unabhängiger Variablen. Im Fall der rPPG wird die ICA mit dem Ziel eingesetzt, durch geschickte Trennung der in den Farbkanälen enthaltenen Mischsignale die Pulsinformationen zu isolieren um daraus Vitalparameter zu extrahieren. Dabei wird angenommen, dass die Pulsinformationen nicht nur im grünen Farbkanal, sondern in den drei RGB-Farbkanälen gemischt vorliegen.

In modellbasierten Verfahren wurde Vorwissen über die Verteilung der Signalkomponenten in den RGB-Farbkanälen genutzt, um daraus rPPG-Algorithmen wie POS, CHROM oder PBV zu entwickeln. Die Algorithmen basieren auf einem Interaktionsmodell zwischen Licht und Gewebe zur Bestimmung eines Projektionsvektors. Die Grundidee dieser Algorithmen liegt in der Projektion der RGB-Signale in eine neue Ebene im Farbraum.

Der POS-Algorithmus [24] beruht darauf, die Dimensionalität des Entmischungsproblems auf die Ebene orthogonal zum Hautton im zeitlich normalisierten RGB-Signal zu reduzieren, um das rPPG-Signal zu extrahieren. Dieser Ansatz versucht, die spiegelnde Reflexion an der Hautoberfläche zu minimieren und so die Erfassung der subtilen pulsinduzierten Farbänderungen zu verbessern. Die chrominanzbasierte Methode (CHROM) [31] erzeugt ein rPPG-Signal, indem sie das von der Lichtreflexion verursachte Rauschen anhand eines Verhältnisses der normalisierten Farbkanäle entfernt. Der PBV-Algorithmus [32] (benannt nach dem berechneten Blutvolumenpulsvektor  $\vec{P}_{bv}$ ) nutzt den charakteristischen Signalverlauf von Blutvolumenschwankungen in verschiedenen Wellenlängenbereichen, um die pulsinduzierten Farbänderungen im RGB-Signal von Bewegungsrauschen zu unterscheiden.

Der LGI-Algorithmus (Local Group Invariance) [33] beruht auf der Idee, dass sich bestimmte lokale Pixelgruppen im Bild weniger stark ändern als andere, unabhängig von den Herzschlägen. Das Verfahren zielt darauf ab, stabile lokale Pixelgruppen zu identifizieren, die weniger anfällig für Bewegungsartefakte und andere Störungen sind, und sie von den sich ändernden Pixeln zu trennen. Diese Gruppen können aufgrund ihrer räumlichen Nähe oder ähnlichen Intensität zusammengefasst werden. Die stabilen Pixelgruppen werden als potenziell zuverlässigere Regionen für die BVP-Extraktion betrachtet.

#### 3.3. rPPG-Verfahren auf Basis neuronaler Netze

Die genannten Forschungsschwerpunkte wurden mit dem Aufkommen von Deep Learning zunehmend durch neuronale End-to-End-Architekturen ergänzt. Mit *DeepPhys* wurde 2018 eines der ersten neuronalen Netzwerke in der rPPG-Forschung veröffentlicht [34]. Die Architektur des *DeepPhys*-Modells besteht aus zwei Zweigen mit zweidimensionalen Convolutional Neural Networks (CNNs). Die Eingangsbilder werden von beiden Zweigen parallel verarbeitet, wobei der Erscheinungsbildzweig den Bewegungszweig über einen Gated-Attention-Mechanismus steuert. Im Bewegungszweig wird die Differenz aufeinanderfolgender Bilder gebildet, um Bewegungsartefakte zu minimieren und den Lernprozess auf Pixelfarbunterschiede zwischen einzelnen Bildern zu fokussieren [35]. Durch den Einsatz einer Attention Mask (Aufmerksamkeitsmaske) kann ein zweites Convolutional Attention Network (CAN) im Erscheinungsbildzweig den Fokus des CNNs im Bewegungszweig auf aussagekräftige Gesichtsbereiche lenken (vergleiche Abbildung 8). Die Idee der Attention Mask ist es, den Hautbereichen mit stärkeren Signalen eine höhere Gewichtung zuzuweisen.

Die Architektur des DeepPhys-Modells basiert auf zweidimensionalen CNNs, die insbesondere für die Mustererkennung in der Verarbeitung räumlicher Daten ausgelegt sind. Yu et al. [36] stellten 2019 mit dem *PhysNet*-Modell das erste räumlich-zeitliche End-to-End-Netzwerk zur Messung von rPPG-Signalen vor. Dabei wurden zwei Modellarchitekturen entwickelt und miteinander verglichen. Die erste Modellarchitektur basiert auf einem dreidimensionalen CNN (3D-CNN). 3D-CNNs sind in der Videoauswertung durch neuronale Netze weit verbreitet, da sie den räumlichen und zeitlichen Kontext der Videodaten gleichzeitig erfassen können.

Die zweite Modellarchitektur besteht aus der Kombination eines zweidimensionalen CNNs und eines rekurrenten neuronalen Netzes [*engl.:* Recurrent Neural Network (RNN)]. RNNs zeichnen sich dadurch aus, zeitliche Informationen aus früheren Eingaben in die Verarbeitung der aktuellen Ein- und Ausgabe miteinzubeziehen. Beim *PhysNet*-Modell dient die Ausgabe der CNN-Komponente als Eingabe für ein Long Short-Term Memory (LSTM) Netzwerk. LSTMs sind eine Unterform von RNNs und sind in ihrer Architektur darauf spezialisiert, Langzeitabhängigkeiten in sequentiellen Daten zu modellieren. Durch die Kombination aus 2D-CNN und RNN-Komponenten wird also ein Netzwerk strukturiert, das sowohl räumlich als auch zeitlich kodierte Informationen verarbeiten kann. Im Vergleich beider Modellarchitekturen anhand des mittleren quadratischen Fehlers [*engl.:* Root Mean Square Error (RMSE)] zwischen der vorhergesagten und tatsächlichen Herzrate schnitt das 3D-CNN mit einem RMSE von 2, 143 besser ab als die Kombination aus

#### 2D-CNN und LSTM mit einem RMSE von 3,139 [36].

In darauffolgenden Forschungen entstand im Jahr 2020 das *TS-CAN* (Temporal Shift-Convolutional Attention Network) [35], bei dem die Architektur des *DeepPhys*-Modells um ein Zeitverschiebungsmodul [*engl.:* Temporal Shift Module (TSM)] erweitert wurde. Das TSM ist eine Komponente innerhalb von CNNs, die speziell entwickelt wurde, um zeitliche Variationen von Pixelfarbwerten in Videobildern zu modellieren und so zeitliche Abhängigkeiten zwischen aufeinanderfolgenden Bildern zu erfassen. Das TSM im Bewegungszweig soll die Effekte der dreidimensionalen Faltungsoperation nachbilden. Der Vorteil liegt darin, dass das Modell die Leistung von dreidimensionalen CNNs erreichen kann, aber die Komplexität von zweidimensionalen CNNs beibehält [37]. Als Eingangsbild des Erscheinungsbildzweigs wird ein einziges gemitteltes Bild aus einer Fenstergröße von *N* Bildern verwendet. In der Veröffentlichung von Liu et al. wurde die Fenstergröße zu N = 10 Bildern gewählt. Dadurch reduziert das Modell die Rechenzeit, da die Aufmerksamkeitsmaske nur einmal pro Fenster berechnet wird und verringert durch die Mittelung über mehrere Bilder den Quantisierungsfehler der Kamera.



Abbildung 8.: Modellarchitektur des Temporal Shift-Convolutional Attention Network (TS-CAN) (entnommen und übersetzt aus [35]). Aufeinanderfolgende Eingangsbilder dienen als Eingangsgrößen für zwei Zweige mit unterschiedlichen Berechnungsschwerpunkten, dem Bewegungszweig und dem Erscheinungsbildzweig. Im Bewegungszweig wird die Bewegung anhand der normalisierten Differenz zweier aufeinanderfolgender Bilder modelliert. Der Erscheinungsbildzweig dient zur Extraktion aussagekräftiger Gesichtsbereiche. Das ermittelte Signal ist die erste zeitliche Ableitung der PPG-Wellenform.

Im Jahr 2023 veröffentlichten Liu et al. [38] mit dem *EfficientPhys*-Modell ein 2D-CNN mit Zeitverschiebungsmodulen, das mit seiner geringen Netzwerkkomplexität für Echtzeitberechnungen auf leistungsarmen Geräten ausgelegt worden ist. Ziel des Modells ist es, Videoaufzeichnungen ohne zusätzliche Datenvorverarbeitungsschritte zu verarbeiten und dadurch die benötigte Rechenzeit zur Bestimmung des BVP-Signals zu minimieren. Die Modellarchitektur verfügt über ein Normalisierungsmodul, das Bilddifferenzen aufeinanderfolgender Eingangsbilder über eine erlernbare Normalisierung berechnet, sowie über ein Aufmerksamkeitsmodul zur Selbstbeobachtung, das dem Netzwerk dabei hilft, sich auf die mit dem PPG-Signal zusammenhängenden Hautpixel zu konzentrieren [39]. Der Unterschied zum Aufmerksamkeitszweig im *TS-CAN*-Modell liegt darin, dass die Berechnungen beim Aufmerksamkeitsmodul im *EfficientPhys*-Modell innerhalb desselben Datenverarbeitungszweigs erfolgen und so die Netzwerkkomplexität gering gehalten wird. Bei der Modellevaluation anhand der PURE und UBFC-rPPG Datensätze schneidet das *TS-CAN*-Modell hinsichtlich seines mittleren absoluten Fehlers [*engl.*: Mean Absolute Error (MAE)] besser ab als das *Efficient*-*Phys*-Modell (siehe Tabelle 1) [39]. Die Überlegenheit des *EfficientPhys*-Modells zeigt sich in der geringeren Rechenzeit von 40 ms pro Videobild, im Vergleich zu 63 ms beim *TS-CAN*.

Unter den vorgestellten neuronalen Netzen zeigt *TS-CAN* konsistent eine gute Leistung bei der Vorhersage der Herzrate aus Videodaten. Darüber hinaus bietet es die Flexibilität, die Aufmerksamkeitsmaske anzupassen, indem selbst implementierte Datenvorverarbeitungsschritte am Eingang des Erscheinungsbildzweigs integriert werden können. Diese Anpassungsfähigkeit ist im Rahmen dieser Arbeit entscheidend, da sie es ermöglicht, spezifische Änderungen am Erscheinungsbild vorzunehmen und den Fokus der Aufmerksamkeitsmaske von vornherein auf definierte Gesichtsmerkmale zu lenken. Weiterhin ist das Modell vergleichsweise schnell in der Ausführung, was für den Einsatz in leistungsschwachen Geräten oder bei der Verarbeitung großer Mengen an Videodaten von Relevanz ist. Diese Kombination aus hoher Vorhersagegenauigkeit, Anpassungsmöglichkeit der Aufmerksamkeitsmaske und schneller Verarbeitungsgeschwindigkeit macht das *TS-CAN*-Modell zu einer optimalen Wahl für die Anforderungen dieser Arbeit.

Tabelle 1.: Leistungsvergleich der vorgestellten Modelle (entnommen und übersetzt aus [39]). Innerhalb der rPPG-<br/>Toolbox (mehr dazu in Kapitel 5) wurden die unüberwachten Modelle unter Verwendung der UBFC-rPPG-,<br/>PURE- und MMPD-Datensätze evaluiert. Die überwachten neuronalen Modelle wurden mithilfe des UBFC-<br/>rPPG- oder PURE-Datensatzes trainiert und an den übrigen beiden Datensätzen getestet.

-RPPG MAPE	MN MAE	MADE
MAPE	MAE	MADE
18 78		WIAT L
10./0	21.68	24.39
14.34	18.60	20.88
3.78	13.66	15.99
14.70	17.08	18.98
15.17	17.95	20.18
3.86	12.36	14.43
N/A 1.50	14.00 13.93	15.47 15.14
N/A 1.68	10.23 13.21	12.46 14.73
N/A 1.42	17.49 16.92	19.26 18.54
N/A 2.10	13.78 14.03	15.15 15.31
	3.78 14.70 15.17 3.86 N/A 1.50 N/A 1.68 N/A 1.42 N/A 2.10	3.78         13.66           14.70         17.08           15.17         17.95           3.86         12.36           N/A         14.00           1.50         13.93           N/A         10.23           1.68         13.21           N/A         17.49           1.42         16.92           N/A         13.78           2.10         14.03

MAE = Mittlerer Absoluter Fehler (Herzschläge/Min)

MAPE = Mittlerer Prozentualer Fehler (%)

#### 3.4. Auswahl aussagekräftiger Gesichtsbereiche

Weitere Forschungsarbeiten untersuchen die Gesichtsbereiche, in denen sich die pulsinduzierten Pixelfarbwertschwankungen am deutlichsten vom intrinsischen Rauschen des Kamerasensors abheben. Eine praktikable Metrik hierfür ist das SNR, welches das Verhältnis zwischen Nutz- und Rauschanteil im Signal misst. Der Bereich von Interesse [*engl.*: Region of Interest (ROI)] schränkt den gesamten sichtbaren Gesichtsbereich auf eine kleinere Region ein, die anschließend zur rPPG-Messung der Vitalparameter verwendet wird. Zu den am häufigsten verwendeten Messregionen gehören die gut durchbluteten Bereiche der Stirn und Wangen [13][40][41] oder das komplette Gesicht [30][42]. Meinzer et al. [43] haben 2013 in einer Untersuchung das SNR jedes Gesichtspixel berechnet und herausgefunden, dass die Signalqualität ungleichmäßig im Gesicht verteilt ist. Durch Begrenzung der ROI auf die Wangen oder Stirn konnte eine bessere Signalqualität erzielt werden als durch Signalextraktion aus dem gesamten Gesichtsbereich.

Kwon et al. [40] haben 2015 untersucht, wie sich die Wahl des Gesichtsbereichs auf das SNR auswirkt. Dafür wurde jedes Videobild in ein kleineres Raster aus 20x20 Pixeln unterteilt und das SNR und Pearson's Korrelationskoeffizient für jede Rasterzelle berechnet. Das Gesicht wurde anschließend in sieben ROIs unterteilt, darunter die Stirn, linke und rechte Wange, Nase, Mund, Kinn und der Bereich zwischen den Augen. Die Zuordnung der Rasterzellen zu den ROIs zeigte, dass Kinn und Mund das niedrigste SNR aufweisen, während die Nase und der Bereich zwischen den Augen das höchste SNR erzeugten. Bezogen auf die Fläche der ROIs kamen die Autoren der Studie zum Schluss, dass die Stirn- und Wangen-ROIs das größte Potenzial zur rPPG-Signalextraktion bieten, da diese einen hohen Anteil an Rasterzellen mit hohem SNR im Vergleich zur Größe der ROI aufweisen. Darüber hinaus ist die absolute Größe dieser ROIs größer als die anderer Gesichtsregionen. Bei vorhandenen Störeinflüssen wie Bewegungen, Lichtreflexionen oder Gesichtsbedeckungen ist die Wahrscheinlichkeit bei einer großen ROI höher, Rasterzellen mit hohem SNR aufzuzeichnen. Dies erhöht die Robustheit der Signalextraktion bei Störeinflüssen.

Fouad et al. [41] verwendeten 2019 einen Ansatz zur Hautsegmentierung, um solche Störeinflüsse zu reduzieren und den informationstragenden Hautbereich aus dem Videobild zu isolieren. Zusätzlich zur Hautsegmentierung wurde der Gesichtsbereich auf die Stirn- und Wangen ROIs unterteilt. Die Resultate haben gezeigt, dass die Extraktion des rPPG-Signals aus der ROI mit Hautsegmentierung eine höhere Vorhersagegenauigkeit der Herzrate bietet als die Extraktion aus der ROI ohne Hautsegmentierung oder aus dem gesamten Gesicht.

Kim et al. [13] führten 2021 eine Bewertung von ROI-Kandidaten unter 31 Gesichtsregionen anhand des MAE und RMSE des extrahierten rPPG-Signals zum Referenzblutvolumenpulssignal durch. Die ROIs wurden auf Grundlage einer Studie von Chopra et al. [14] definiert, in der die mittlere Dicke der Gesichtshaut für jede dieser ROIs gemessen wurde. Die Untersuchung hat ergeben, dass die ROIs der Stirn- und Wangen die besten Metriken erzielen, während die ROIs um die Nase herum die höchsten Abweichungen von der tatsächlichen Herzrate erzeugen. Die am besten abschneidenden ROIs weisen Hautdicken zwischen 1086  $\mu$ m und 1386  $\mu$ m auf, wohingegen die am schlechtensten abschneidenden ROIs Hautdicken zwischen 1496  $\mu$ m und 2016  $\mu$ m aufweisen. Durch die Untersuchungen wurde der Schluss nahegelegt, dass dünnere und größere Hautregionen präzisere Vorhersage des BVP durch den rPPG-Algorithmus erlauben [13]. Dies deckt sich mit der Eigenschaft der Haut, dass mit zunehmender Hautdicke die optische Weglänge und damit die Absorption des diffus reflektierten Lichts steigt (siehe Abschnitt 2.2).

Darüber hinaus wurden Verfahren untersucht, die ROIs bei vorhandenen Kopfbewegungen nicht mehr statisch zu definieren, sondern diese über den Zeitverlauf hinweg anzupassen. Wong et al. [44] haben 2022 die erste Studie zu den Auswirkungen der Oberflächenorientierung durchgeführt und hierzu eine sogenannte Winkelkarte entwickelt, die für jedes Gesichtspixel den Reflexionswinkel zwischen seinem Oberflächennormalenvektor und der Kameraachse misst. Es konnte gezeigt werden, dass eine Verringerung des Reflexionswinkels die Korrelation zwischen dem extrahierten rPPG-Signal und dem tatsächlichen BVP-Signal signifikant erhöht [44]. Die Abhängigkeit der extrahierten Signalqualität von der Oberflächenorientierung erklärt, warum die ROIs um die Wangen- und Stirnregionen von frontal zugewandten Personen meist stärkere rPPG-Signale enthielten. Bei Kopfrotationen bietet die Winkelkarte die Möglichkeit zur dynamischen Auswahl von ROIs, indem nur noch Gesichtsbereiche unterhalb eines definierten Winkelschwellenwerts in der aktiven ROI berücksichtigt werden. Da das SNR in den verschiedenen Gesichtsbereichen variiert, ist die Anpassung der ROIs eine hilfreiche Technik, um die Auswirkungen von Bewegungsartefakten und Beleuchtungsvariationen zu minimieren und gleichzeitig die Bereiche auszuwählen, die die meisten physiologischen Informationen enthalten.

Die vorgestellten Datenvorverarbeitungsverfahren bieten verschiedene Ansätze, um die Vorhersagegenauigkeit der rPPG-Messung zu erhöhen. Allerdings bestehen weiterhin Herausforderungen, wie zum Beispiel Artefakte durch Bewegungen und Videokompression oder Einflüsse unterschiedlichen Umgebungslichts, die die Genauigkeit beeinträchtigen können [45]. Darüber hinaus ist die Vorhersagegenauigkeit stark vom epidermalen Melaningehalt abhängig, da mit zunehmendem Melaningehalt die Absorption des innerhalb der Haut diffus reflektierten Lichstrahls steigt. In der Folge sinkt die Lichtintensität des informationstragenden Signals, wodurch folglich auch das SNR sinkt [46][8][47]. Um die Qualität der entwickelten Signalverarbeitungsalgorithmen unabhängig von den untersuchten Populationen und Aufnahmesituationen bewerten zu können, sind geeignete Datensätze notwendig, die diese Variationen umfassen. Nachfolgend werden Datensätze vorgestellt, die in der Forschung an rPPG-Verfahren häufig verwendet werden, um die Qualität der entwickelten Algorithmen miteinander vergleichen zu können.

# 4. Übersicht über bestehende rPPG-Datensätze

In diesem Kapitel werden Datensätze vorgestellt, die in der Forschung an rPPG-Verfahren häufig verwendet werden, um die Qualität der entwickelten Verfahren miteinander vergleichen zu können. Die Datensätze beruhen auf Videoaufnahmen mit variierenden Auflösungen, Videokompressionen und Aufnahmeumgebungen. Neben den Videos verschiedener Probanden sind in den Datensätzen auch deren BVP-Signale enthalten, welche parallel zur Videoaufzeichnung durch einen medizinisch zugelassenen PPG-Fingerclipsensor aufgezeichnet werden.

#### 4.1. UBFC-rPPG

Der UBFC-rPPG-Datensatz umfasst Videoaufnahmen von 42 Probanden und wurde mit einer handelsüblichen Webcam (*Logitech C920 HD Pro*) mit 30 Bildern pro Sekunde und einer Auflösung von 640x480 im unkomprimierten 8-Bit-RGB-Format erstellt. Ein Pulsoximeter (*pulox CMS50E*) wurde verwendet, um parallel zur Aufzeichung die Referenz-PPG-Daten zu erfassen, die sowohl die PPG-Wellenform als auch die PPG-Herzraten umfassen. Während der Aufzeichnung sitzt die Testperson etwa 1 m von der Kamera entfernt. Alle Aufzeichnungen wurden in Innenräumen mit unterschiedlichen Sonneneinstrahlungen und Beleuchtungsstärken durchgeführt [42]. Der Datensatz gilt als verlässlich und wird in der rPPG-Forschung häufig als Referenz eingesetzt, ist aber aufgrund der begrenzten Vielfalt an enthaltenen Hauttypen und Bewegungen nur bedingt aussagekräftig für die Qualität der Signalverarbeitungsalgorithmen.

#### 4.2. COHFACE

Der COHFACE-Datensatz umfasst 160 einminütige Videosequenzen von 40 Probanden (12 weibliche und 28 männliche), die mit dem BVP und der Atemfrequenz der aufgenommenen Personen synchronisiert sind. Die Videosequenzen wurden mit einer Auflösung von 640x480 Pixeln und einer Bildrate von 20 Hz aufgenommen [48]. Jeder Proband sitzt bei der Aufzeichnung in einem Meter Abstand vor der Kamera und wurde in vier Durchläufen mit einer Länge von je einer Minute gefilmt. Die Lichtverhältnisse werden nach zwei Durchläufen geändert. Für die ersten beiden Durchläufe wird der Proband nur von einer künstlichen Lichtquelle beleuchtet. Hierzu werden die Jalousien geöffnet und der Proband wird von einer seitlichen Sonnenlichteinstrahlung durch Fenster ohne zusätzlichen Lichtquellen im Raum beleuchtet (Abbildung 9b).

#### 4.3. VIPL-HR-V1

Der VIPL-HR-V1-Datensatz enthält Videoaufnahmen von 107 Probanden bei unterschiedlichen Lichtverhältnissen und verschiedenen Kopfbewegungen. Die Videoaufnahmen wurden gleichzeitig von vier verschiedenen Kameras mit variierenden Auflösungen, Bildraten und kameraspezifischem Bildrauschen aufgezeichnet, darunter zwei Webcams, eine Smartphone-Kamera und eine Nahinfrarot-Kamera. Parallel wurden die physiologischen Messwerte des Probanden wie die Herzrate, die Sauerstoffsättigung (SpO<sub>2</sub>) und der BVP durch ein Fingerpulsoximeter (*CONTEC CMS60C*) aufgezeichnet [49][50].



(a) Künstliche Lichtquelle



(b) Natürliche Sonnenlichtbestrahlung

Abbildung 9.: Beispielbilder der Beleuchtungssituationen im COHFACE-Datensatz

Die Probanden befinden sich in Abständen von 1 m und 1,5 m zu den Kameras. Neben den Kameras ist eine Lampe zur Variation der Lichtverhältnisse installiert. Die Versuchspersonen sitzen in einer natürlichen Haltung vor den Kameras und wurden während der Videoaufnahme zu alltäglichen Aktivitäten wie Sprechen und Umhersehen gebeten. Es wurden zusätzliche Situationen aufgenommen, in denen die testende Person ruhig sitzt. Dabei wurde die Umgebungsbeleuchtung in zwei verschiedenen Versuchsszenarien erhöht beziehungsweise verringert. In einem weiteren Szenario wurde der Abstand des Probanden zur Kamera vergrößert. Darüber hinaus wurden die Probanden aufgezeichnet, nachdem sie vorher eine zweiminütige sportliche Betätigung (Seilspringen) durchgeführt haben.

#### 4.4. PURE

Der PURE-Datensatz enthält Videoaufnahmen von zehn Personen, die vor einer Kamera verschiedene, kontrollierte Kopfbewegungen ausführen. Während dieser Bewegungsabläufe wurden sowohl Videoaufnahmen des Kopfes als auch Referenzpulsmessungen aufgezeichnet. Die zehn Personen (acht männlich, zwei weiblich) wurden in sechs verschiedenen Szenarien aufgenommen, was eine Gesamtzahl von 60 Aufnahmen von je einer Minute Länge ergibt. Zu den Szenarien gehört das statische Sitzen vor der Kamera, Sprechen, sowie langsame und schnelle Translations- und Rotationsbewegungen des Kopfes.

Die Videos wurden mit einer Bildfrequenz von 30 Hz und einer Auflösung von 640x480 Pixeln aufgezeichnet. Parallel dazu wurden Referenzdaten mit einem Fingerpulsoximeter (*Contec Medical CMS50E*) erfasst, das Messwerte zur Herzrate und Sauerstoffsättigung mit einer Abtastrate von 60 Hz liefert. Die Videos und zugehörigen Messwerte des Pulsoximeters wurden anschließend anhand der UNIX-Zeitstempel miteinander synchronisiert. Die Versuchspersonen wurden in einem durchschnittlichen Abstand von 1,1 m vor der Kamera platziert. Die Ausleuchtung erfolgt durch Tageslicht, das durch ein großes Fenster frontal auf das Gesicht fällt, wobei sich die Lichtverhältnisse im Laufe der Aufnahmen durch Wolken geringfügig ändern [51]. Die Schwäche des Datensatzes liegt darin, dass die Bewegungen unter künstlich hergestellten Bedingungen innerhalb eines kleinen Rahmens stattfinden, wodurch keine realen Bewegungen im Alltag in den Aufnahmen enthalten sind.

#### 4.5. MMPD

Der MMPD-Datensatz umfasst 33 Personen, die in jeweils 20 Versuchsdurchläufen mit einer Smartphone-Kamera (*Samsung Galaxy S22 Ultra*) aufgezeichnet wurden. Die Videos wurden mit einer Auflösung von 1280x720 Pixel bei einer Bildwiederholrate von 30 Hz aufgezeichnet und in eine Auflösung von 320x240 Pixel komprimiert. Parallel dazu wurden Referenzdaten mit einem Fingerpulsoximeter (*HKG-07C+*) erfasst, das eine Abtastrate von 200 Hz liefert. Die PPG-Signale wurden anschließend auf 30 Hz heruntergetastet, um sie an die Bildfrequenz der Videos anzupassen.

Ziel des Datensatzes war es, Videos mit einer umfassenden Bandbreite an unterschiedlichen Hauttönen, Bewegungsmustern und Lichtverhältnissen aufzuzeichnen [52]. Die Auswahl der Probanden umfasst nach der Fitzpatrick-Skala die Hauttypen III - VI (siehe Kapitel 2.2.1). Zudem ist der Datensatz hinsichtlich der Geschlechterverteilung ausgeglichen und enthält Kennzeichnungen, ob die Versuchsperson eine Brille oder Makeup trägt, oder das Gesicht teilweise von Haaren bedeckt wird (Tabelle 2). Unter vier verschiedenen Beleuchtungssituationen (drei künstliche und eine natürliche Lichtquelle) wurde jeder Proband in fünf verschiedenen Versuchsszenarien aufgezeichnet (Stationär, Kopfrotation, Sprechend, Laufend, Stationär nach Sportaktivität).

Verteilung	Ing Fitzpatrick Hauttyp Geschlecht		Fitzpatrick Hauttyp		Brillenträger		Bedeckung durch Haare		Makeup			
	III	IV	V	VI	Männlich	Weiblich	Ja	Nein	Ja	Nein	Ja	Nein
Anzahl	16	5	6	6	16	17	10	23	8	23	4	29

Tabelle 2.: Verteilung der Merkmale im MMPD-Datensatz [52][53].

# 4.6. RLAP

Der "Remote Learning Affect and Physiologic (RLAP)"-Datensatz besteht aus Videos von 58 Probanden (16 männliche und 42 weibliche), die in 13 Versuchsszenarien mit einer Logitech C930c Webcam aufgezeichnet werden. Parallel dazu wird der zugehörige Blutvolumenpuls durch ein Fingerpulsoximeter (*CONTEC CMS50E*) aufgezeichnet. Für die Datenaufzeichnung wurde das von Wang et al. [54] eigens entwickelte Tool *PhysRecorder* eingesetzt.

Bei der Aufzeichnung des RLAP-Datensatzes lag ein besonderer Fokus auf der Synchronisation zwischen dem Video und dem PPG-Signal. Analog zum Vorgehen im PURE-Datensatz wurden die Videos und zugehörigen

Messwerte des Pulsoximeters anhand der UNIX-Zeitstempel synchronisiert. Zudem wurde Wert darauf gelegt, einen Teil der aufgezeichneten Versuchsszenarien für die rPPG Messung zu optimieren. Deshalb wurden vier rPPG-spezifische Versuchsszenarien gesondert mit einer Auflösung von 640x480 Pixel bei 30 FPS im YUY2-Kameracodec aufgezeichnet, da die einzelnen Bilder in diesem Datenformat unkomprimiert gespeichert werden. Die übrigen neun Versuchsszenarien wurden mit einer Auflösung von 1920x1080 Pixel bei 30 FPS im MJPG Datenformat aufgezeichnet. Der rPPG-Teildatensatz umfasst unterschiedliche Helligkeitsstufen, geistige Anstrengungen (in Ruhe, beim Lesen und beim Spielen eines Videospiels) und Bewegungen der Gesichtsmuskel.
Datensatz	Probanden	Kopf- bewegungen	Beleuchtung	Videolänge in s	Auflösung & Bildwiederholrate	Gesamtzahl der Videos
LIBEC-rPPG [42]	42	B	T	~120	640x480	42
	12	it it	Ц	120	@ 30 FPS	14
	40	D	I /D	10	640x480	160
CONFACE [40]	40	ĸ	L/ D	10	@ 20 FPS	100
	107	D /D /C		20	1280x720	0070
VIPL-IIK-VI[49]	107	K/ D/ 3	L/ D/ П	30	@ ~24 FPS	23/0
	10	D /D /C	т	60	640x480	60
PURE [31]	10	N/ D/ 3	L	00	@ 30 FPS	00
	22	D /D /C		60	320x240	660
	33	K/ D/ 3	L/ D/ П	00	@ 30 FPS	000
	EQ	D /D /C		100	640x480	
KLAP-IPPG [54]	50	K/ D/ 3	L/ D/ П	120	@ 30 FPS	232
RLAP-Emotion	EQ	D /D /C		20.490	1920x1080	E00
& Engagement <sup>[34]</sup>	58	K/ B/ S	L/ D/ H	20-480	@ 30 FPS	522

Tabelle 3.: Übersicht über die öffentlich zugänglichen rPPG-Datensätze

Kopfbewegungen: R - Ruhig, B - Bewegend, S - Sprechend

Beleuchtung: L - Laborumgebung, D - Dunkle Umgebung, H - Helle Umgebung

Die bestehenden rPPG-Datensätze weisen einige Kriterien auf, die in der rPPG-Forschung von Relevanz sind. Die Datensätze variieren in Bezug auf verschiedene Faktoren wie Aufnahmesituationen, Probandenvielfalt und Geräten zur Datenakquirierung. Die Datensätze enthalten Videos unter verschiedenen Lichtbedingungen, Hintergründen oder Aktivitäten, welche es ermöglichen, die Robustheit der rPPG-Extraktion in unterschiedlichen Umgebungen und Alltagssituationen zu testen. Bis auf einzelne Versuchsszenarien im VIPL-HR-V1-, MMPD- und RLAP-Datensatz haben die Datensätze die Gemeinsamkeit, dass die Lichtbedingungen und Hintergründe während der Videoaufnahme statisch bleiben und sich nicht verändern.

Die Datensätze umfassen in ihrer Gesamtheit verschiedene Altersgruppen, ethnische Gruppen und Hauttypen, um die Generalisierungsfähigkeit der rPPG-Algorithmen über unterschiedliche Populationen hinweg testen zu können. Beim MMPD-Datensatz liegen umfassende Annotationen vor, die es ermöglichen, die Probanden anhand ihrer Erscheinungsmerkmale, wie Hautfarbe, Geschlecht, Brillenträger oder aufgetragenem Make-up zu unterscheiden.

Bei den meisten Datensätzen, bis auf den MMPD- und VIPL-HR-V1-Datensatz, wurden vergleichbare Kameras oder Pulsoximeter als Aufnahmegeräte eingesetzt, was eine bessere Vergleichbarkeit ermöglicht, allerdings die Identifikation und Kompensation gerätespezifischer Fehler erschwert. Das Videokompressionsformat ist entscheidend für eine präzise Extraktion der Herzrate [54][55]. Die PURE- und RLAP-Datensätze wurden als unkomprimierte Bilder abgespeichert. Beim RLAP-Datensatz wurden parallel noch unkomprimierte Videos aufgezeichnet. Die Videoaufzeichnungen im UBFC-rPPG-Datensatz liegen ebenso im unkomprimierten Videoformat vor.

Wang et al. [54] haben die Synchronisation zwischen den Videoaufzeichnungen und den Messungen des Fingerclippulsoximeters untersucht und festgestellt, dass die Videoaufzeichnungen und BVP-Signale beim UBFC-rPPG- und VIPL-HR-V1-Datensatz eine Zeitverschiebung von über 0,5 Sekunden aufweisen. Der MMPD-Datensatz weist eine Zeitverschiebung von weniger als 0,2 Sekunden auf, wohingegen beim PURE-

und RLAP-Datensatz aufgrund der Synchronisation anhand der UNIX-Zeitstempel keine Zeitverschiebung vorliegt.

Die Gemeinsamkeit der Datensätze liegt darin, dass die Versuchsszenarien bis auf wenige Ausnahmen im VIPL-HR-V1-, MMPD und RLAP-Datensatz unter kontrollierten Laborbedingungen aufgezeichnet wurden. Dadurch wird die Extraktion des BVP-Signals durch rPPG-Algorithmen erleichtert, allerdings geschieht dies zum Nachteil der Verallgemeinerungsfähigkeit der Algorithmen in realen Anwendungsfällen außerhalb kontrollierter Laborumgebungen. Zudem weisen die Datensätze in der Gestaltung ihrer Versuchsszenarien Überschneidungen auf, wodurch Variationen in den Geräten zur Datenakquirierung und in der Probandenvielfalt entstehen, aber wenige Härtefälle und Negativbeispiele mit enthaltenen Störfaktoren provoziert werden. Aus diesem Grund wird im Rahmen dieser Arbeit ein Datensatz erstellt mit dem Ziel, solche Härtefälle mit bekannten Störfaktoren aufzuzeichnen.

## 4.7. Entwicklung eines eigenen KISMED-Datensatzes

Das Ziel dieser Datenerhebung besteht darin, einen umfassenden Datensatz für die Erforschung der rPPG zu erstellen, der verschiedene Alltagssituationen und herausfordernde Szenarien umfasst. Der Fokus liegt auf der Erfassung von Härtefällen und Negativbeispielen für die Extraktion von rPPG-Signalen. Hierzu werden verschiedene Faktoren variiert, die die Erfassung von Herzschlagsignalen aus Gesichtsvideos erschweren können.

#### 4.7.1. Messsetup und verwendete Hardware

- 1. **Kamera:** Als Kamera wird eine *Logitech C930e* USB-Webcam eingesetzt. Diese Kamera gewährt eine maximale Auflösung von 1920x1080 Pixeln bei 30 FPS, wobei zum Zweck einer niedrigen Datenkompression eine Auflösung von 640x480 Pixeln unter Verwendung des YUY2-Datencodierungsformats verwendet wird. Parallel zur Aufnahme eines Videos wird zudem jedes Videobild gesondert abgespeichert. Zusätzlich wird als zweite Kamera eine spiegellose Vollformatkamera (*SIGMA fp*) eingesetzt. Die Kamera gewährt eine Auflösung von 1920x1080 Pixeln bei 29,97 FPS und eine verlustfreie Speicherung der Video-Rohformatdaten mit einer Farbtiefe von 12-Bit durch den *CinemaDNG*-Standard.
- 2. **Positionierung der Kameras:** Die Kameras bleiben stationär und werden auf Augenhöhe des Probanden positioniert. Die Entfernung der Kameras zum Probanden soll standardmäßig 1 m betragen.
- 3. Zeitbedarf pro Proband: Sofern es in der nachfolgenden Beschreibung der Versuchsszenarien nicht anders angegeben ist, dauert jedes aufgezeichnete Video 60 Sekunden. Bei Szenario V04 und V05 beträgt die Aufnahmedauer 120 Sekunden und bei Szenario V06 270 Sekunden. Die gesamte Datenerhebung für einen Probanden dauert etwa eine Stunde, einschließlich der Einweisungen durch den Versuchsleiter.
- 4. **BVP-Aufzeichnung:** Parallel zur Videoaufnahme wird der Blutvolumenpuls des Probanden durch ein Fingerclippulsoximeter [*Pulox PO-250* (Datenblatt in Anhang A.1)] aufgezeichnet, welches am linken oder rechten Zeigefinger angebracht wird.

- 5. **Synchronisierung der Daten:** Die aufgezeichneten Daten vom Pulsoximeter und der *Logitech*-Kamera werden anhand der UNIX-Zeitstempel synchronisiert. Zusätzlich werden die Videoaufzeichnungen beider Kameras zu Beginn und am Ende jedes Versuchsszenarios durch ein Lichtsignal synchronisiert. Das Lichtsignal wird durch einen *Arduino Uno* emittiert, der über eine serielle Schnittstelle mit dem Aufzeichnungs-Tool gekoppelt ist.
- 6. Lichtquellen: Als Lichtquellen werden die Deckenbeleuchtung und das Tageslicht durch Fenster eingesetzt. Für einzelne Szenarien wird ein flimmerfreier Halogen-Baustrahler als zweite Lichtquelle seitlich des Probanden positioniert. Der Halogen-Baustrahler weist ein breites Lichtspektrum auf, das vergleichbar zu natürlichem Sonnenlicht ist.

#### 4.7.2. Durchführung der Datenerhebung

- 1. Der Proband wird über den Zweck der Datenerhebung und den Datenschutz aufgeklärt.
- 2. Der Proband wird über die verschiedenen Szenarien und Aufgaben informiert.
- 3. Die Kameras werden dem Versuchsszenario entsprechend im gleichen Abstand zum Probanden positioniert und eingestellt. Die Synchronisierung der Kameravideos erfolgt anhand der UNIX-Zeitstempel und des zusätzlichen Einsatzes einer Synchronklappe zu Beginn jeder Videoaufzeichnung.
- 4. Der Proband nimmt verschiedene Positionen ein und führt die definierten Szenarien durch, während die Kameras die Gesichtsvideos aufzeichnen. Parallel wird durch eine angepasste Version des PhysRecorder-Tools [54] ein synchronisiertes Blutvolumenpulssignal vom Fingerclippulsoximeter aufgezeichnet.
- 5. Die Videos werden entsprechend den Szenarien benannt und für die spätere Analyse verlustfrei gespeichert.
- 6. Der Versuchsleiter steuert und überwacht den Ablauf und steht für eventuelle Rückfragen oder Anpassungen bereit.

#### 4.7.3. Versuchsszenarien für Härtefälle und Negativbeispiele

- 1. Variation der Lichtverhältnisse: Der Proband soll stillsitzend in einem Abstand zur Kamera positioniert werden. Die Deckenbeleuchtung ist während der Videoaufzeichnung durchgängig ein- oder ausgeschaltet, um verschiedene Lichtverhältnisse zu simulieren. In einem dritten Durchgang wird die Deckenbeleuchtung in definierten Zeitpunkten an- und ausgeschaltet.
- 2. **Ungleichmäßige Gesichtsbeleuchtung:** Eine zweite Lichtquelle seitlich des Probanden soll einen Schattenwurf über eine Gesichtshälfte erzeugen, um asymmetrische Lichtverhältnisse zu schaffen. In einem weiteren Szenario soll dabei die Deckenbeleuchtung ausgeschaltet und nur noch die seitliche Lichtquelle verwendet werden. Die seitliche Beleuchtung erfolgt durch einen Baustrahler, der alternierend alle 20 Sekunden ein- und ausgeschaltet wird. Die Gesamtdauer der Videos weisen in diesem Szenario als einzige eine Dauer von 120 Sekunden auf.
- 3. **Variation im Abstand zur Kamera:** Der Proband ändert den Abstand zur Kamera während der Aufnahme, um unterschiedliche Größen der Gesichtsmerkmale zu erzeugen. Ausgehend von einem Abstand von 1 m soll dieser alle 30 Sekunden um 0,5 m erhöht werden, bis ein Abstand von 3 m erreicht

wird. Anschließend soll der Abstand nach derselben Vorgehensweise wieder auf 1 m reduziert werden. Die resultierenden Videos weisen in diesem Szenario als einzige eine Dauer von 270 Sekunden auf.

- 4. **Aufzeichnung während körperlicher Aktivität:** Der Proband führt Kniebeugen mit Blick in die Kamera durch, um Bewegungsartefakte und Veränderungen in der Gesichtspositionierung zu provozieren.
- 5. Variation der Gesichtsbedeckung: Bei der Auswahl der Probanden wird darauf geachtet, dass unter allen Aufnahmen verschiedene Arten von Gesichtsbedeckungen wie Make-up, Schals oder Brillen mit einbezogen werden.
- 6. Variation der Gesichtsorientierung zur Kamera: Der Proband soll im Laufe der Versuchsszenarien verschiedene Positionen und Gesichtsorientierungen zur Kamera einnehmen, um Translations- und Rotationsbewegungen im Datensatz aufzunehmen. Dafür erhält der Proband Anweisungen über eine grafische Benutzeroberfläche [*engl*. Graphical User Interface (GUI)], die ihn zum Ansehen von Markierungen im Raum auffordern (Abbildung 10). Zur einfacheren Durchführung der Translationsbewegungen werden die Probanden gebeten, die Bewegungen im Stehen durchzuführen. Um das Gesicht weiterhin frontal aufzuzeichnen, wird die Höhe des Kamerastativs entsprechend angepasst.



Abbildung 10.: Versuchsaufbau bei der Aufnahme des KISMED-Datensatzes (angelehnt an den Versuchsaufbau bei der Aufzeichnung des PURE-Datensatzes nach Stricker et al. [51]). Über die GUI erhält der Proband Anweisungen, eine zufällig ausgewählte Markierung zu betrachten. Bei einem Abstand von 1 m ergeben die einpfeiligen Markierungen einen Rotationswinkel von 20° und die doppelpfeiligen Markierungen einen Rotationswinkel von 40°.

Für die einzelnen Versuchsszenarien werden zum Teil mehrere Durchgänge unter Variation der Beleuchtungssituation oder des Bewegungsmusters aufgezeichnet. Daraus ergeben sich folgende zwölf Szenarien:

- V01: Stationär mit eingeschalteter Deckenbeleuchtung
- **V02:** Stationär mit ausgeschalteter Deckenbeleuchtung
- V03: Stationär mit alternierender Deckenbeleuchtung
- **V04:** Stationär mit Variation der seitlichen Beleuchtung mit Deckenbeleuchtung

V05: Stationär mit Variation der seitlichen Beleuchtung ohne Deckenbeleuchtung

- V06: Variation der Entfernung zur Kamera
- V07: Kniebeugen mit Blick in die Kamera
- V08: Variation der Gesichtsbedeckung
- V09: Natürliches Verhalten (sprechend, lesend, auf Handy schauend, Kaugummi kauend)
- V10: Translationsbewegung
- V11: Rotationsbewegung
- V12: Kombination aus Rotation und Translation

#### 4.7.4. Informationen zum aufgezeichneten KISMED-Datensatz

Der aufgezeichnete *KISMED*-Datensatz enthält Videoaufzeichnungen von zehn Probanden (sieben männlich und drei weiblich mit einem mittleren Alter von 26, 2 Jahren). Unter den Probanden tragen 6 Personen eine Brille und zwei Personen Make-Up. Unter den Probanden ist der Fitzpatrick Hauttyp III insgesamt acht Mal vertreten, sowie je einmal der Hauttyp II und Hauttyp V.

Die Dateigröße eines einminütigen unkomprimierten Videos von der *Logitech*-Kamera beträgt etwa 2,06 Gigabyte, während die parallel abgespeicherten unkomprimierten Bilder eine Gesamtgröße von etwa 540 Megabyte aufweisen. Die unkomprimierten Videobilder der *Sigma fp*-Kamera weisen bei einem einminütigen Video eine Dateigröße von etwa 6,10 Gigabyte auf. Aus Gründen der Datenspeicherung und des beschleunigten Dateitransfers wurden die Einzelbilder der *Sigma fp*-Kamera komprimiert als .zip-Archiv abgespeichert. Bei einer Datenmenge von zehn Probanden, die in zwölf Versuchszenarien aufgezeichnet wurden, weist der entstandene Datensatz eine Gesamtgröße von 1,22 Terabyte auf. Der *KISMED*-Datensatz wird dem Institut für künstlich intelligente Systeme der Medizin (*KISMED*) der Technischen Universität Darmstadt für zukünftige institutsinterne Forschungsarbeiten zur Verfügung gestellt. Eine Veröffentlichung des Datensatzes oder die Abbildung der Probanden innerhalb von Publikationen ist durch die Datenschutzerklärung ausgeschlossen worden.

#### 4.7.5. Auswertung der Signalqualität

Die Auswertungen der Videoaufzeichnungen sind für jedes Szenario separat in den Tabellen 4 und 5 aufgeführt. Es ist zu erkennen, dass die Vorhersageleistung durch die unüberwachten rPPG-Algorithmen bei Translationsbewegungen des Probanden keine zuverlässigen Ergebnisse liefert (Szenarien V06, V07, V10 und V12). Dies ist vor allem auf Bewegungsartefakte innerhalb der Referenz-BVP-Signale zurückzuführen, welche die Berechnung der momentanen Referenzherzrate beeinflussen (siehe Abbildung 11). Bei den

Szenarien mit Variationen in der Umgebungsbeleuchtung V03, V04 und V05 ist ebenso eine deutliche Verschlechterung der Bewertungsmetriken zu verzeichnen. Eine denkbare Erklärung für diesen Umstand liegt in der automatischen Helligkeitsanpassung der eingesetzten *Logitech*-Kamera, welche einen zusätzlichen Störeinfluss auf die Extraktion der BVP-Signale durch die rPPG-Algorithmen ausmacht.

Datensatz: KISMED																				
Szenario	Stati	onär m	it Decker (V01)	beleuc	ntung		Statio	onär oh	ne Decke (V02)	nbeleuc	htung		Stationär mit alternierender Deckenbeleuchtung (V03)							
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc			
ICA	2,20	4,77	0,9275	2,55	75,00		11,33	22,66	0,4037	-4,49	50,00		19,25	25,04	0,1197	-13,63	20,00			
POS	1,05	1,93	0,9898	6,97	80,00		4,20	11,56	0,8375	1,47	83,33		14,24	18,11	-0,4154	-12,80	20,00			
CHROM	0,79	1,52	0,9948	5,98	85,00		5,57	16,16	0,6595	-0,28	88,89		11,07	14,87	0,2509	-11,09	25,00			
GREEN	2,11	4,72	0,9318	3,67	70,00		9,18	20,99	0,4938	-2,26	66,67		20,74	24,10	0,4389	-14,69	5,00			
LGI	1,05	1,93	0,9898	5,83	80,00		5,86	15,54	0,6861	-0,25	83,33		19,86	24,34	-0,3071	-15,87	15,00			
PBV	5,19	10,04	0,7581	-3,00	60,00		6,15	15,98	0,6418	-4,26	83,33		23,38	26,73	-0,4726	-16,95	5,00			
Szenario	Stat une	ionär n d Decke	nit Seiten enbeleuch	beleuch tung (\	tung /04)		Stat ohn	ionär n e Deck	nit Seiten enbeleucl	beleuch htung (\	tung V05)		Variation der Entfernung zur Kamera (V06)							
ICA	12,44	16,93	0,072	-11,80	31,58	2	21,53	24,94	0,1676	-16,76	5,56		15,21	20,55	0,1386	-11,79	25,56			
POS	5,37	8,87	0,7769	-6,33	55,26		9,96	13,26	0,3737	-12,27	36,11		5,27	9,11	0,7269	-4,51	52,22			
CHROM	5,13	9,32	0,7542	-7,67	60,53		8,01	11,29	0,4677	-11,51	36,11		6,43	10,94	0,6638	-5,84	48,89			
GREEN	13,18	17,27	0,2111	-12,77	26,32	1	22,71	25,84	0,1456	-15,44	5,56		19,30	24,97	0,0034	-15,13	15,56			
LGI	5,87	10,79	0,6749	-10,50	57,89		18,02	22,48	-0,1181	-15,22	22,22		10,72	17,25	0,361	-8,14	43,33			
PBV	10,45	15,1	0,5098	-10,94	34,21		13,18	16,94	-0,1049	-13,16	19,44		11,91	17,82	0,2666	-9,50	36,67			

 Tabelle 4.: Vergleich der ersten sechs Szenarien V01 bis V06 im aufgezeichneten KISMED-Datensatz mit der vorimplementierten Gesichtsextraktion durch die rPPG-Toolbox [39].

#### 4.7.6. Diskussion

Der KISMED-Datensatz ist mit dem Ziel aufgezeichnet worden, Härtefallszenarien und Negativbeispiele mit Störfaktoren für die rPPG-Analyse zu enthalten. Bei der Auswertung der vorhergesagten rPPG-Signale durch den POS-Algorithmus [24] zeigt sich für Szenarien mit geringer körperlicher Bewegung des Probanden eine qualitativ gute Signalextraktion. Bei den bewegungsreichen Szenarien V06, V07, V10 und V12, welche translatorische Bewegungen enthalten, treten im Referenz-BVP-Signal des Fingerclippulsoximeters Bewegungsartefakte auf. Diese Bewegungsartefakte beeinträchtigen die Berechnung der Referenzherzrate, was im späteren Vergleich der Vorhersagegüte zu deutlichen Abweichungen zwischen den rPPG-Algorithmen und der Referenz führt (vergleiche Tabelle 5). Für zukünftige Forschungen mit dem Datensatz ist eine zusätzliche Datenvorverarbeitung mit dem Ziel der Artefaktminimierung in den artefaktbelasteten Signalen denkbar. Die Auswahl geeigneter Methoden sollte jedoch sorgfältig erfolgen, um die Erhaltung relevanter physiologischer Informationen sicherzustellen und Informationsverluste zu minimieren.

Die in diesem Kapitel vorgestellten Datensätze werden im weiteren Verlauf der Arbeit eingesetzt, um anhand der unüberwachten rPPG-Algorithmen und geeigneten Metriken Untersuchungen durchzuführen. Mithilfe



Abbildung 11.: Ausschnitte der Referenz- und rPPG-BVP-Signale für jedes Szenario im KISMED-Datensatz. Die rPPG-Signale wurden mit dem POS-Algorithmus aus der Region der Stirn und Wangen extrahiert. Aus Darstellungsgründen sind sämtliche Signale auf einen Wertebereich von [-1,1] normiert. Die Signalausschläge in den Szenarien V03, V04 und V05 entstehen durch die Variation der Umgebungsbeleuchtung. Die Szenarien mit translatorischen Bewegungen V06, V07, V10 und V12 weisen unter Umständen Bewegungsartefakte in den Referenzsignalen auf.

Tabelle 5.: Vergleich der letzten sechs Szenarien V07 bis V12 im aufgezeichneten KISMED-Datensatz mit der vorimplementierten Gesichtsextraktion durch die rPPG-Toolbox [39]. Aufgrund einer unzureichenden Gesichtsverfolgung wurde für die Bewegungsszenarien V07, V10 und V12 eine selbst implementierte Extraktion des Gesichtsbereichs nach dem Vorgehen in Kapitel 6.1 mit einem Winkelschwellenwert von 90° eingesetzt. Die vergleichsweise schlechten Metriken in den Szenarien mit Translationsbewegungen der Probanden sind auf Bewegungsartefakte in den Referenz-BVP-Signalen zurückzuführen.

Datensatz: KISMED																			
Szenario	К	niebeu	genszena	rio (V07	7)		Variati	ion der (	Gesichtsbe	deckung	(V08)		Natürliches Verhalten (V09)						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc		
ICA	37,53	45,75	0,0349	-18,55	20,00		5,71	12,68	0,4106	-1,76	70,00		20,30	25,57	-0,2288	-12,98	10,00		
POS	32,34	40,85	-0,0348	-16,70	20,00		2,20	4,03	0,9404	3,27	75,00		6,42	10,66	0,3976	-4,21	45,00		
CHROM	35,95	43,53	-0,0651	-19,87	20,00		1,93	3,34	0,9688	2,93	80,00		8,26	11,45	0,3249	-8,51	30,00		
GREEN	47,64	54,23	-0,6000	-28,73	10,00		5,71	12,78	0,4056	-0,62	75,00		19,95	25,52	-0,1570	-14,81	5,00		
LGI	43,68	52,32	-0,1883	-24,53	20,00		2,72	5,74	0,8757	1,57	75,00		13,45	19,44	0,0170	-11,09	25,00		
PBV	42,89	51,79	-0,3252	-25,10	20,00		5,36	13,35	0,4346	-2,89	65,00		14,33	20,31	0,0128	-11,11	15,00		
Szenario	Trans	ationst	pewegung	g (V10)			Rotati	onsbew	vegung (\	/11)			Rotation- & Translation (V12)						
ICA	28,81	31,93	-0,2905	-13,90	5,00		7,47	12,91	0,4161	-3,46	55,00		20,74	25,24	0,0331	-13,65	10,00		
POS	6,05	9,57	0,5952	-3,70	50,00		0,88	1,57	0,9901	4,35	90,00		8,70	13,21	0,4758	-6,13	45,00		
CHROM	8,98	14,51	0,4841	-6,56	35,00		1,32	2,75	0,9669	3,15	85,00		7,56	12,45	0,6554	-7,87	60,00		
GREEN	24,61	29,69	-0,2174	-16,88	10,00		15,38	19,56	0,1416	-10,42	25,00		25,22	28,67	0,0063	-16,12	0,00		
LGI	17,77	26,47	-0,0297	-7,61	35,00		3,60	9,29	0,682	1,97	85,00		15,38	22,84	0,3588	-10,47	45,00		
PBV	19,14	25,21	0,5334	-10,62	27,78		6,86	13,85	0,4718	-3,28	65,00		19,6	26,19	0,1610	-12,67	25,00		

der Untersuchungen soll aus dem gesamten Gesichtsbereich die Region of Interest ermittelt werden, die die aussagekräftigste Extraktion eines BVP-Signals ermöglicht. Hierfür sind zunächst einige Schritte zur Vorverarbeitung der Datensätze notwendig, welche im nachfolgenden Kapitel vorgestellt werden.

# 5. Datenverarbeitungsschritte

In diesem Kapitel werden die im Rahmen dieser Arbeit implementierten Schritte zur Vorverarbeitung der Videoaufzeichnungen und der Referenz-BVP-Signale beschrieben. Die vorgestellten Schritte zur Datenvorverarbeitung sollen im späteren Verlauf der Arbeit in Untersuchungen dazu eingesetzt werden, eine ROI im Gesichtsbereich zu ermitteln, welche aussagekräftige pulsatile Informationen zur Extraktion eines BVP-Signals enthält. Weiterhin werden die Bewertungsmetriken vorgestellt, mit denen die Auswirkung der Datenvorverarbeitung für die rPPG-Signalextraktion evaluiert werden soll. Nachfolgend werden die notwendigen Grundlagen der Bild- und Signalverarbeitung vorgestellt, um die darauf aufbauenden Analysen und Auswertungen in dieser Arbeit zu verstehen.

Die Schritte zur Datenvorverarbeitung werden in Python 3.8 programmiert. Die Berechnungen laufen auf einem Intel Core i7-4702MQ Prozessor mit 4 Kernen à 2,2 GHz und Hyperthreading. Der Arbeitsspeicher beträgt 16 GB und die Datensätze sind auf einer externen Festplatte mit einer Datenübertragungsrate von etwa 100 MB/s gespeichert. Aufgrund fehlender NVIDIA CUDA Unterstützung wird keine GPU-Hardwarebeschleunigung angewandt.

#### 5.1. Erweiterung der bestehenden rPPG-Toolbox zur Datenverarbeitung

Im Rahmen dieser Arbeit wird die öffentlich verfügbare rPPG-Toolbox [39] eingesetzt und um selbst entwickelte Funktionen erweitert. Die rPPG-Toolbox ist eine End-to-End-Toolbox für kamerabasierte physiologische Messungen und bietet eine leistungsstarke Plattform zur nichtinvasiven Erfassung von Herzfrequenzsignalen aus Videodaten und Bildern. Der Funktionsumfang umfasst: 1) Unterstützung für sechs öffentlich verfügbare Datensätze, 2) Vorverarbeitungscode zur Formatierung der Datensätze für das Training neuronaler Modelle, 3) Implementierungen von fünf neuronalen Modellarchitekturen und sechs unüberwachten Lernmethoden, 4) Evaluierungs- und Inferenzpipelines für unüberwachte Algorithmen und überwachte Lernmethoden zur Reproduzierbarkeit der Datenverarbeitung [39].

Die rPPG-Toolbox bietet ein System zur Einstellung von Konfigurationsdateien, das es dem Benutzer ermöglicht, einen Großteil der verwendeten Parameter im Vorverarbeitungsprozess, Training, Nachverarbeitungsprozess und in der Auswertung zu ändern. Für jeden Datensatz und jedes implementierte neuronale Modell wird eine YAML-Datei bereitgestellt, die Einstellungsmöglichkeiten für die Datenvor- und nachbearbeitung, sowie Modellhyperparameter für das Training, die Validierung und anschließende Tests bietet.

Um die Funktionalität und die Anwendungsbreite der rPPG-Toolbox zu erweitern, wurden im Rahmen dieser Arbeit einige Erweiterungen vorgenommen. Diese Erweiterungen beinhalten die Integration zusätzlicher Dataloader (Ladeprogramme) für den RLAP-, COHFACE- und VIPL-HR-V1-Datensatz und die Implementierung eines Logging-Systems, um die Benutzerfreundlichkeit zu verbessern und nachfolgende Datenverarbeitungsfunktionen zu ermöglichen. Die Integration der Dataloader in die rPPG-Toolbox ermöglicht es, Datensätze effizient und systematisch zu verwalten und zu verarbeiten. Mit dem Dataloader kann problemlos auf eine Vielzahl von Datenquellen zugegriffen werden, darunter die benötigten Videodateien und Aufzeichnungen der Referenz-PPG-Signale. Dies erleichtert die Dateneingabe und -verarbeitung erheblich und bietet außerdem die Flexibilität, eigene Datensätze oder kleinere Teildatensätze zu nutzen. Darüber hinaus ermöglicht der Dataloader die einfache Verwaltung von Datensätzen, was bei umfangreichen Verarbeitungen mehrerer Datensätze von großem Vorteil ist.

Die Implementierung des Logging-Systems in der rPPG-Toolbox bietet Nutzern die Möglichkeit, die Verarbeitungsschritte und Ergebnisse ihrer rPPG-Analysen zu verfolgen und zu dokumentieren. Dies ist von Bedeutung, um die Transparenz und Reproduzierbarkeit von rPPG-Analysen sicherzustellen. Das Logging-System zeichnet alle Schritte der Datenverarbeitung auf, einschließlich der ausgewählten Konfigurationsparametern, der angewandten rPPG-Methoden und der Ergebnisse. Dies ermöglicht es den Nutzern, ihre Analysen zu überprüfen und die Qualität der rPPG-Methoden sicherzustellen.

## 5.2. Tiefpasssfilterung der extrahierten Landmarks durch MediaPipe

Die von Google entwickelte MediaPipe-Bibliothek stellt ein leistungsstarkes Framework zur Verfügung, um verschiedene Aufgaben im Bereich der Computer Vision zu bewältigen. Eines der Anwendungsfelder von MediaPipe ist die Gesichtserkennung, die es ermöglicht, die dreidimensionalen kartesischen Bildkoordinaten von insgesamt 478 biometrischen Gesichtspunkten, auch als Landmarks bezeichnet, zu berechnen [56]. Diese Gesichtspunkte umfassen wichtige Merkmale wie die Augen, die Nase, den Mund und andere charakteristische Punkte im Gesicht. Diese Landmarks sind entscheidend für eine präzise Gesichtsanalyse und können für verschiedene Anwendungen wie Emotionsanalyse, Gesichtserkennung oder Augmented Reality Gesichtsanimationen genutzt werden.

Allerdings neigen die von Media<br/>Pipe ermittelten Landmarks dazu, zu zittern beziehungsweise zu rauschen.<br/>Dieses Zittern kann auf verschiedene Faktoren zurückzuführen sein, wie verrauschte Bildaufnahmen, kame-<br/>rainternes Rauschen, unerwünschte Bewegungen des Gesichts oder Ungenauigkeiten im Tracking-Prozess.<br/>Um das Rauschen der Landmarkkoordinaten zu reduzieren, wird zu Beginn der Datenvorverarbeitung<br/>eine Tiefpassfilterung angewandt. Die Bildkoordinaten jedes Gesichtspunkts werden während der Vorver-<br/>arbeitung jedes Videobilds in einem NumPy-Array gespeichert. Anschließend wird der Zeitverlauf jeder Koordinate durch einen Tiefpass 40. Ordnung mit einer Grenzfrequenz von  $f_g = 3,5$  Hz gefiltert, um die hochfrequenten Rauschanteile der Bildkoordinaten zu reduzieren (Abbildung 12). Dies bedeutet, dass die Filterung hochfrequente Rauschanteile der Bildkoordinaten reduziert und somit für eine räumlich geglättete und stabilere Darstellung der Landmarks sorgt.

Die Wahl der Grenzfrequenz ist hierbei kritisch. Sie bestimmt, welche Frequenzen im Signal erhalten bleiben und welche unterdrückt werden. Eine zu niedrige Grenzfrequenz kann dazu führen, dass auch nützliche Informationen schneller Bewegungen verloren gehen, während eine zu hohe Grenzfrequenz das Rauschen nicht ausreichend reduziert. Daher ist die Wahl der richtigen Grenzfrequenz entscheidend, um einen praktikablen Kompromiss zwischen Rauschreduzierung und Signalqualität zu finden. Die Wahl der Grenzfrequenz zu 3,5 Hz basiert auf einer Untersuchung von Chen et al. [57], in der die Winkelgeschwindigkeit bei Kopfbewegungen für die Analyse körperlicher und geistiger Aktivitäten quantifiziert wurde. In der Untersuchung



Abbildung 12.: Tiefpassfilterung der x-Koordinaten des Gesichtspunkts an der Nasenspitze bei ruhig gehaltenem Kopf (a) und bei Kopfrotation (b).

von 24 Probanden wurde festgestellt, dass die maximale Rotationsfrequenz des Kopfes bei 3,5 Hz liegt. Dieser Befund steht in Einklang mit Ergebnissen von Grossman et al. [58], die vertikale und horizontale Kopfbewegungen von 20 Probanden unter Berücksichtigung der höchstmöglichen Rotationsgeschwindigkeit untersuchten. Die Auswertung ergab, dass der Median der Grundfrequenzen aller 20 Kopfrotationen bei horizontalen Bewegungen 2,6 Hz beträgt, während er bei vertikalen Bewegungen 2,4 Hz beträgt.

## 5.3. Berechnung der Reflexionswinkel zwischen Hautoberfläche und Kamera

Bei der Gesichtsanalyse generiert MediaPipe insgesamt 478 Koordinaten, die die Positionen biometrischer Gesichtspunkte im dreidimensionalen kartesischen Koordinatensystem beschreiben. Um die Darstellung des frontalen Gesichtsbereichs zu optimieren und eine kohärente Repräsentation zu schaffen, verbindet MediaPipe diese benachbarten Gesichtspunkte miteinander, um Dreiecke zu bilden. Dies führt zu einer sogenannten Parkettierung des frontalen Gesichtsbereichs, bei der die Gesichtspunkte in einem regulären Dreiecksgitter angeordnet sind.

Die Definition der Eckpunkte aller Dreiecke, die diese Gesichtsparkettierung bilden, ist ein essenzieller Schritt in der Gesichtsanalyse. Insgesamt ergibt sich aus dieser Vorgehensweise eine Anzahl von 852 Dreiecken. Diese Dreiecke bilden die Grundlage für die nachfolgende Segmentierung des Gesichts und ermöglichen eine präzise Verfolgung und Analyse der Gesichtsmerkmale.

In Abschnitt 3.1.2 wurden Forschungsergebnisse vorgestellt, die nahelegen, dass die Begrenzung des Gesichtsbereichs auf einen Teilbereich mit geringen Reflexionswinkeln zwischen der Oberflächennormale und der Kameraachse ein höheres SNR für das extrahierte rPPG-Signal liefern können. Angelehnt an die Methode von Wong et al. [44] wird für jedes Dreieck wird der Oberflächennormalenvektor mithilfe der generierten Koordinaten berechnet. Dafür werden zwei Vektoren von einem Eckpunkt  $p_j$  zu den übrigen

beiden Eckpunkten  $p_i$  und  $p_k$  des Dreiecks berechnet:

$$\vec{v}_{ji} = p_j - p_i$$
  

$$\vec{v}_{jk} = p_j - p_k$$
(5.1)

wobei  $\vec{v}_{ji}$  den Vektor von Punkt j zu einem benachbarten Punkt i bezeichnet und  $\vec{v}_{jk}$  den Vektor von Punkt j zu einem anderen benachbarten Punkt k.

Das Kreuzprodukt der entstehenden Vektoren  $\vec{v}_{ji}$  und  $\vec{v}_{jk}$  liefert den Oberflächennormalenvektor  $\vec{n}_j$ :

$$\vec{n}_j = \vec{v}_{ji} \times \vec{v}_{jk} \tag{5.2}$$

Schließlich wird der Reflexionswinkel  $\theta_j$  des Gesichtspunktes j zum Einheitsvektor der Kameraachse  $\vec{z}$  wie folgt berechnet:

$$\theta_j = \arccos(\frac{\vec{n}_j \cdot \vec{z}}{|\vec{n}_j||\vec{z}|}) \tag{5.3}$$

wobei der Einheitsvektor der Kameraachse  $\vec{z}$  orthogonal auf der Bildebene liegt und in Richtung der Kamera zeigt. Dieser wird daher angenommen als:

$$\vec{z} = \begin{pmatrix} 0\\0\\-1 \end{pmatrix}$$
(5.4)

Die Berechnung der Reflexionswinkel jedes Dreiecks ergibt die in Abbildung 13a dargestellte Gesichtsparkettierung, wobei die Reflexionswinkel basierend auf ihrem Betrag farblich hervorgehoben sind.

### 5.4. Histogrambasierte Hautsegmentierung/Gesichtshaarentfernung

Die Hautsegmentierung hat den Zweck, Bildbereiche im Gesichtsbereich zu identifizieren und zu entfernen, die nicht zur Haut gehören und damit keine pulsatilen Informationen enthalten. Dies schließt Elemente wie Haare, Bärte, Brillen, Kopfbedeckungen und weitere Bedeckungen der Haut ein. Die Vorgehensweise zur Segmentierung der Haut orientiert sich am Vorgehen von Stricker et al. [51], bei dem die Farbverteilung der relevanten Hautregionen Ausgangspunkt für die Segmentierung ist. Dort wird das Bild zunächst in den normalisierten RGB-Farbraum konvertiert. Anschließend wird für jeden Farbkanal ein Histogramm aller Bildpixel erstellt, das 64 Farbwertunterteilungen (Intervallklassen) enthält. Der Farbbereich der Haut wird als Höhepunkt im Farbwerthistogramm angenommen.

Im nächsten Schritt wird der zulässige Hautbereich durch die Bildpixel erweitert, die entweder in den Höhepunkt oder seine benachbarten Intervallklassen fallen. Anhand dieses stark reduzierten Hautbereichs wird eine provisorische Hautsegmentierungsmaske erzeugt. Nach dem Vorgehen von Stricker et al. soll mithilfe des GrabCut-Algorithmus [59] der Bildvordergrund (die Gesichtshaut) vom Bildhintergrund separiert werden. Der Grabcut-Algorithmus muss mit einer Region im Bild initialisiert werden, die als Vordergrund





Abbildung 13.: (a) Die Kombination der 478 biometrischen Gesichtspunkte aus MediaPipe's Gesichtserkennung erzeugt eine Gesichtsparkettierung aus 852 Dreiecken. (b) Der Reflexionswinkel  $\theta_i$  des Gesichtspunktes *j* folgt aus dem Winkel zwischen dem Oberflächennormalenvektor  $\vec{n}_j$  zum Einheitsvektor in Richtung zur Kamera  $\vec{z}$ . Die Reflexionswinkel sind farblich hervorgehoben. Blau repräsentiert Winkel unterhalb von 30°, grün repräsentiert Winkel zwischen 30° und 60° und rot repräsentiert Winkel über 60°.

gekennzeichnet ist, und mit einer Region, die als Hintergrund gekennzeichnet ist. In einem iterativen Prozess werden die verbleibenden Bildpixel anhand des Farbwerthistogramms beider Kategorien, dem Vordergrund oder Hintergrund zugeordnet. Dabei wird angenommen, dass die vorab definierte Hautsegmentierungsmaske dazu beiträgt, dass auch die Gesichtshaut außerhalb der definierten Grenzen im Histogramm als Vordergrund identifiziert wird, während die übrigen Bildbereiche als Hintergrund klassifiziert und aus dem resultierenden Bild entfernt werden.

Die im Rahmen dieser Arbeit implementierte Vorgehensweise orientiert sich an diesem Vorgehen. Anhand der durch MediaPipe erzeugten Gesichtsparkettierung wird zunächst der Gesichtsbereich von übrigen Bildbereichen isoliert (Abbildung 14b). Anschließend werden die Farbwerthistogramme für jeden Farbkanal im RGB-Farbraum anhand des vom Hintergrund isolierten Gesicht berechnet. Der GrabCut-Algorithmus konnte am isolierten Gesichtsbereich keine erfolgreiche Unterscheidung zwischen Gesichtshaut und hautbedeckenden Elementen vornehmen, weshalb stattdessen im weiteren Vorgehen histogrammbasierte Schwellenwerte für die zulässigen Pixelfarbwerte definiert werden. Dies ermöglicht eine alternative Vorgehensweise aufgrund der Unzulänglichkeit, die der GrabCut-Algorithmus bei der Unterscheidung zwischen Gesichtshaut und hautbedeckenden Elementen hatte.

Für jeden Farbkanal im RGB-Farbraum wird eine individuelle Fallunterscheidung anhand des Höhepunktes im Histogramm getroffen: Wenn der Höhepunkt oberhalb des 25%-Quartils liegt, wird das 25%-Quartil als unterer Schwellenwert festgelegt (Abbildung 15a). Wenn der Höhepunkt unterhalb des 25%-Quartils liegt, wird der Schwellenwert auf die untere Grenze des Histogramintervalls festgelegt, das zwei Intervalle unterhalb des Peaks liegt (Abbildung 15b). Die ermittelten Farbwertschwellenwerte werden auf dem isolierten Gesichtsbereich angewandt, was zur histogrammbasierten Hautsegmentierungsmaske führt (Abbildung 14c).



(a) Ausgangsbild



(c) Histogrammbasierte Hautmaske

(d) Segmentierte Gesichtshaut

Abbildung 14.: Anwendung einer histogrammbasierten Hautsegmentierungsmaske. Ausgehend vom Ausgangsbild (a) wird der von MediaPipe erkannte Gesichtsbereich isoliert (b). Anschließend wird die histogrammbasierte Hautsegmentierungsmaske berechnet (c), die nach Anwendung zur segmentierten Gesichtshaut führt (d).

An der Hautsegmentierungsmaske werden eine Reihe von Bildverarbeitungstechniken zur Reduktion von Bildrauschen durchgeführt. Das Bildrauschen entsteht durch einzelne Hautpixel, die durch den histogrammbasierten Schwellenwert als ungültig deklariert wurden. Das Bildrauschen wird zunächst durch einen Median-Unschärfefilter reduziert. Anschließend wird eine morphologische Öffnungsoperation, bestehend aus einer Erosion und gefolgt von einer Dilation, an der Hautsegmentierungsmaske angewandt. Dieser Prozess zielt darauf ab, einzelne Hautpixel, die zuvor außerhalb des akzeptierten Farbwertbereichs lagen, durch die benachbarten gültigen Hautpixel zu glätten, um sie in den akzeptierten Hautbereich zu integrieren. Die boolsche UND-Verknüpfung des isolierten Gesichtsbereichs und der geglätteten Hautsegmentierungsmaske liefert die segmentierte Gesichtshaut (Abbildung 14d).



Abbildung 15.: Histogramm der drei Farbkanäle im segmentierten Gesichtsbereich. Das 25%-Quartil wird durch die schwarz gestrichelte Linie dargestellt. Die rot gestrichelte Linie repräsentiert die Schwellenwerte im grünen und blauen Farbkanal. Pixel werden aus dem Gesicht entfernt, wenn sie in mindestens einem Bei der Hautse Entherhalb des Schwellenwerts liegen Hautbereichen kritisch. Der Algorithmus zur Hautsegmentierung wird mit dem Ziel ausgelegt, dunkle Gesichtsbereiche, die Gesichtsbehaarung enthalten

können, zu entfernen, während unbedeckte Hautbereiche enthalten bleiben sollen. Da die Bildhelligkeit in Anhängigkeit der vorherrschenden Beleuchtung zwischen den Datensätzen und den einzelnen Szenarien stark variiert, wird keine obere Grenze für gültige Farbwerte im Histogramm festgelegt, um stark belichtete Hautbereiche nicht zu entfernen. Der Nachteil dieser Vorgehensweise liegt darin, dass die Segmentierung nur für dunkle Bildbereiche ausgelegt ist und weiterhin Gesichtsbedeckungen durch farbige Kleidungsstücke oder Brillen im segmentierten Hautbereich enthalten sein können.

Bei der experimentellen Untersuchung der implementierten Vorgehensweise konnte festgestellt werden, dass die schwellenwertabhängige Einschränkung des roten Farbkanals zur großflächigen Entfernung von Hautpixeln führt. Dies ist darauf zurückzuführen, dass die Hautfarbe insbesondere im roten Farbkanal über ein breites Spektrum an Pixelfarbwerten verteilt ist [46]. Da das Ziel der implementierten Hautsegmentierung die Erhaltung gültiger Hautbereiche ist, wird die Einschränkung der gültigen Farbwerte nur auf dem grünen und blauen Farbkanal vorgenommen.



(a) Ausgangsbild

(b) Isolierter

Gesichtsbereich

(c) Histogrammbasierte Haut-(d) Segmentierte Gesichtshaut maske

Abbildung 16.: Anwendung der Hautsegmentierungsmaske in dunklen Beleuchtungsumgebungen. Im oberen Fall ist der Farbwertschwellenwert gleich dem 25%-Ouartil, im unteren Fall zwei Intervallklassen unterhalb der prävalenten Farbwertunterteilung. Aufgrund der niedrigen Pixelfarbwerte im zweiten Fall wird effektiv keine Hautsegmentierung durchgeführt, wodurch die dunklen Hautbereiche im resultierenden Gesichtsbereich erhalten bleiben.

## 5.5. Segmentierung der Region of Interest

Im Rahmen der Datenvorverarbeitung wird die Segmentierung des gesamten Gesichtsbereichs in aussagekräftigere ROIs implementiert. Die entwickelte Segmentierung kann entweder anhand eines zuvor

definierten ROI-Bereichs, eines Schwellenwerts für den Reflexionswinkel oder der Kombination aus beidem erfolgen. Bei der Gesichtssegmentierung anhand des Winkelschwellenwerts werden sämtliche Dreiecke der Gesichtsparkettierung in den segmentierten Gesichtsbereich aufgenommen, deren Reflexionswinkel unterhalb des Winkelschwellenwerts liegt.

In Abschnitt 3.1.2 wurden bereits Erkenntnisse zur ROI-Segmentierung aus vorangegangenen Forschungsarbeiten vorgestellt. Die Definition aussagekräftiger ROIs erfolgt in Anlehnung an die Untersuchung von Kim et al. [13], in der die rPPG-Signalextraktion aus den Stirn- und Wangen-ROIs die höchste Übereinstimmung mit dem tatsächlichen BVP lieferte. Die in dieser Arbeit definierten ROIs setzen sich zusammen aus den Dreiecken in MediaPipe's Gesichtsparkettierung, die aus den folgenden Landmarks zusammengesetzt sind:

- Stirn: 8, 9, 10, 55, 66, 67, 69, 107, 108, 109, 151, 285, 296, 297, 299, 336, 337, 338
- Linke Wange: 36, 50, 100, 101, 117, 118, 119, 120, 123, 142, 187, 203, 205, 206, 207, 216
- Rechte Wange: 266, 280, 329, 330, 346, 347, 348, 349, 352, 371, 411, 423, 425, 426, 427, 436

Zur Verbesserung der Benutzerfreundlichkeit wurde die rPPG-Toolbox um Konfigurationsparameter erweitert, die es ermöglichen die gewünschte ROI-Segmentierung während der Verarbeitung des Datensatzes zu konfigurieren. Die Einstellungen legen fest, welche Gesichtsbereiche in die rPPG-Analyse miteinbezogen werden und inwieweit die Segmentierung basierend auf Winkelschwellenwerten oder bestimmten vordefinierten Gesichtsbereichen erfolgt (Abbildung 17). Die eingeführten Parameter sind die Folgenden:

- **DO\_SEGMENTATION:** Boolesche Bedingung, die festlegt, ob die ROI-Segmentierung durchgeführt wird oder die vorimplementierte Gesichtsextraktion aus der rPPG-Toolbox vorgenommen wird.
- **THRESHOLD:** Ein Schwellenwert im Bereich von 0° bis 90° für den zulässigen maximalen Reflexionswinkel jedes Dreiecks in der Gesichtsparkettierung. Dreiecke mit einem Reflexionswinkel über diesem Schwellenwert werden aus der ROI-Segmentierungsmaske entfernt.
- **CONSTRAIN\_ROI:** Boolesche Bedingung, die festlegt, ob neben dem Winkelschwellenwert auch eine Einschränkung des extrahierten Gesichtsbereichs auf die vom *ROI\_MODE* definierte ROI vorgenommen wird.
- **ROI\_MODE:** Definition, welcher Gesichtsbereich für die rPPG-Extraktion segmentiert werden soll, sofern *CONSTRAIN\_ROI* als *wahr* gesetzt ist. Zur Auswahl stehen Stirn, linke Wange, rechte Wange, sowie die Kombination aus Stirn- und Wangen-ROIs. Im Programmcode werden die ROIs angegeben als: *"forehead", "left\_cheek", "right\_cheek" und "optimal\_roi"*.
- USE\_CONVEX\_HULL: Boolesche Bedingung, die festlegt, ob eine konvexe Hülle auf den segmentierten Gesichtsbereich angewandt wird. Dies ist insbesondere von Vorteil, wenn *CONSTRAIN\_ROI* aktiviert ist und der Reflexionswinkel einzelner Dreiecke nahe am Winkelschwellenwert liegt. In solchen Fällen können die Dreiecke dazu neigen, zwischen aufeinanderfolgenden Bildern beim Akzeptanzcheck zu fluktuieren. Die segmentierte ROI wird durch den Einsatz der konvexen Umhüllung geglättet, wodurch zuvor nicht akzeptierte Dreiecke nahe des Winkelschwellenwerts dennoch in der ROI erhalten bleiben.
- USE\_OUTSIDE\_ROI: Boolesche Bedingung, mit der eine flächenmäßig gleichgroße ROI außerhalb der durch den *ROI\_MODE* definierten ROI erzeugt wird. Dabei wird zusätzlich die Interpolation der Reflexionswinkel für jedes Gesichtspixel aktiviert.

Bei ruhig gehaltenem Gesicht können einzelne Dreiecke, deren Reflexionswinkel den Winkelschwellenwert leicht übersteigt oder dem Winkelschwellenwert nahe liegen, in ihrer Berücksichtigung in der segmentierten ROI fluktuieren. Dies kann bei aufeinanderfolgenden Bilden zu Störeinflüssen bei der rPPG-Extraktion führen, da die informationstragenden Bildbereiche zwischen aufeinanderfolgenden Bildern nicht konstant erhalten bleiben. Um dieses Zittern in der Berücksichtigung der Dreiecke in der ROI zu reduzieren, wurde mit dem *Akzeptanzcheck* eine Überprüfung des aktuellen und der vergangenen Reflexionswinkel jedes Dreiecks implementiert. Das grundlegende Konzept des Akzeptanzchecks besteht darin, zu evaluieren, ob ein Dreieck unterhalb des vordefinierten Winkelschwellenwerts liegt oder dies in der Vergangenheit, innerhalb der letzten fünf verarbeiteten Bildern, der Fall war. Dieser Prozess dient der Berücksichtigung eines leicht überschwelligen Dreiecks in die segmentierte ROI.

Der Akzeptanzcheck erfolgt durch die Erfassung des aktuellen Reflexionswinkels und Aktualisierung der Winkelhistorie jedes Dreiecks über die letzten fünf Videobilder. Der aktuelle Reflexionswinkel jedes Dreiecks wird mit dem vorgegebenen Schwellenwert verglichen, um zu ermitteln, ob das Dreieck akzeptiert werden soll oder nicht. Wenn das Dreieck wegen eines zu hohen Reflexionswinkels nicht akzeptiert wird, wird anschließend berücksichtigt, ob das Dreieck in der Mehrheit der vergangenen fünf Bilder bereits akzeptiert wurde.

# 5.6. Interpolation der Gesichtswinkel

Im Anschluss an die Berechnung der Reflexionswinkel für jedes Dreieck der Gesichtsparkettierung kann bei Bedarf einer höheren Winkelauflösung eine Interpolation der Reflexionswinkel für jedes Gesichtspixel vorgenommen werden. Hierzu wird dem Mittelpunkt jedes Dreiecks der zugehörige Reflexionswinkel zugeschrieben. Das entstehende Punktraster wird anschließend mithilfe der *griddata*-Funktion aus der *SciPy*-Bibliothek für jedes Gesichtspixel linear interpoliert [60]. Die interpolierten Gesichtswinkel kommen insbesondere zum Einsatz bei der Erzeugung einer ROI, die außerhalb der Stirn- und Wangen Regionen liegt (Abbildung 18)).

Die Interpolation der Reflexionswinkel für jedes Gesichtspixel ist der zeitintensivste Arbeitsschritt in der Datenvorverarbeitung. Die Laufzeit der Datenvorverarbeitung wurde durch den Einsatz der Python Bibliothek *numba* beschleunigt. Numba ist ein Open-Source-JIT-Compiler, der eine bestimmte Auswahl an Python- und NumPy-Funktionen in schnelleren Maschinencode übersetzt [61]. Ein Teil der durch numba beschleunigten Funktionen wird während der Verarbeitung jedes Videobilds mehrere Hundert mal ausgeführt.

Mithilfe der numba-Bibliothek konnte die Laufzeit einer einzelnen Bildverarbeitung durchschnittlich um ein Drittel reduziert werden. So konnte die Laufzeit zur Verarbeitung der zehn Videos aus dem Bewegungsszenario des PURE-Datensatzes von 2131 Sekunden auf 1419 Sekunden reduziert werden, was eine Beschleunigung der Gesamtrechenzeit um 50% darstellt, beziehungsweise eine Reduktion der Laufzeit um 33% (Abbildung 19).

Die blau markierten Datenvorverarbeitungsschritte ohne aktiver Interpolation sind in Abbildung 20 dargestellt. Bei aktiver Interpolation (in orange) wird im Anschluss an die Winkelberechnung im dritten



Abbildung 17.: Schaubild über die Gesichtssegmentierung in Abhängigkeit der Konfigurationsparameter. Bei den dargestellten Abbildungen ist der Parameter USE\_CONVEX\_HULL auf True gesetzt.



(a) Gesichtsframe mit eingezeichneten Dreiecken der Stirn- und Wangen-ROIs



(b) Interpolierte Reflexionswinkel für jedes Gesichtspixel



- (c) Darstellung einer angewandten ROI-Maske und zusätzlich eingezeichneten Dreiecken der Stirn- und Wangen-ROIs zum Größenvergleich.
- Abbildung 18.: Anwendung einer ROI-Maske anhand der niedrigsten interpolierten Gesichtswinkel, die außerhalb der Stirn- und Wangen-ROIs liegen. Zur Erzeugung der Maske wurden die ersten N Gesichtspixel außerhalb der Stirn- und Wangen-ROIs mit den niedrigsten Winkeln verwendet, wobei N der Anzahl an Pixeln in der Stirn- und Wangen-ROI entspricht.



Abbildung 19.: Vergleich der benötigten Laufzeit zur Segmentierung der ROI eines Frames. Die Code-Beschleunigung durch numba reduziert die Rechenzeit im Szenario ohne Winkelinterpolation von 225,8 ms auf 128,5 ms. Dies stellt eine Beschleunigung der Gesamtrechenzeit um 75% dar und eine Reduktion der Laufzeit um 43%.

Schritt zusätzlich die Interpolation des Reflexionswinkels für jedes Gesichtspixel durchgeführt. Die übrigen Datenvorverarbeitungsschritte bis zur ROI-Segmentierung werden bei beiden Verfahren gleichermaßen durchgeführt. Bei aktiver Interpolation wird im Anschluss an die ROI-Segmentierung eine flächenmäßig gleich große Pixelfläche aus den Gesichtspixeln mit den niedrigsten interpolierten Reflexionswinkeln als resultierende ROI erstellt (siehe Abbildung 18c). Die segmentierte ROI wird letztlich mithilfe einer Bounding Box im Bild zentriert. Die entstehenden Videos mit dem segmentierten Gesichtsbereich werden für die anschließende Signalextraktion durch die sechs unüberwachten rPPG-Algorithmen auf der SSD gespeichert.



Abbildung 20.: Schaubild der durchgeführten Schritte zur Vorverarbeitung der Videos, um anhand der rPPG-Methoden ein Blutvolumenpulssignal aus den Videos zu extrahieren.

## 5.7. Nachbearbeitung der extrahierten BVP-Signale

In ihrer Forschungsarbeit zur rPPG-Toolbox [39] beschreiben Lui et al. mehrere implementierte Schritte zur Nachbearbeitung und Auswertung der Daten, die sich in früheren Arbeiten zur rPPG-Forschung bewährt haben und in der Regel zur Verbesserung der Modellvorhersagen führen. Unter anderem wird das PPG-Signal einem *detrending* (Langzeittrendbereinigung) unterzogen, um langfristige Trends und Drifts in den rPPG-Signalen zu entfernen und die Herzfrequenzkomponente hervorzuheben. Dies ermöglicht eine bessere Identifizierung und Analyse der Herzfrequenzvariationen.

Bei der Signalverarbeitung von Blutvolumenpulssignalen können verschiedene Artefakte und Störungen auftreten, die die Herzfrequenzmessung beeinflussen können. Diese Artefakte können sowohl niedrige als auch hohe Frequenzen umfassen. Zur Filterung der vorhergesagten PPG-Wellenform ein Butterworth-Filter 2. Ordnung verwendet, um unerwünschte Frequenzen und Rauschen zu unterdrücken, während die Herzfrequenzkomponente erhalten bleibt. Die voreingestellten Grenzfrequenzen betragen 0,75 Hz und 2,5 Hz, was die ermittelbaren Herzraten auf einen Bereich von 45 bis 150 BPM einschränkt. Die Wahl der Filterparameter kann einen erheblichen Einfluss auf die nachfolgenden Ergebnisse haben, wie beispielsweise Abweichungen der Herzrate.

Die Herzrate eines Menschen kann erheblich variieren, abhängig von verschiedenen Faktoren wie Alter, Fitnessniveau, Stress und Aktivitätsniveau. Das Spektrum der bei Menschen auftretenden Herzraten erstreckt sich von unter 30 BPM (z.B. während des Schlafs) bis über 250 BPM (z.B. während intensiver körperlicher Aktivität). Die Auswahl der Grenzfrequenzen des Bandpassfilters zu 0,5 Hz und 4,2 Hz ermöglicht die Erfassung eines breiteren Bereichs physiologisch relevanter Herzraten zwischen 30 und 250 BPM [11][12].

Um die Auswirkungen dieser Verarbeitungsschritte zu veranschaulichen und die Qualität der rPPG-Signale zu überprüfen, werden die übereinander gelagerten Signalverläufe dargestellt (Abbildung 21). Diese

Darstellung umfasst das unbearbeitete Ausgangssignal, das trendbereinigte Signal, das Bandpass-gefilterte Signal und die Referenzherzrate des Fingerclippulsoximeters.



Abbildung 21.: Schaubild der durchgeführten Schritte zur Nachbearbeitung eines extrahierten BVP-Signals durch den POS-Algorithus. Der Butterworth-Filter reduziert niederfrequente Signal- und Rauschanteile mit Frequenzen unter 0,5 Hz und hochfrequente Signal- und Rauschanteile mit Frequenzen über 4,2 Hz. Die Auswirkung des Detrending wird bei einem längeren Zeitverlauf deutlicher ersichtlich.

Auf das gefilterte Signal wird anschließend wahlweise eine Fast-Fourier-Transformation (FFT) oder ein Peak-Detection-Algorithmus angewandt, um die resultierende Herzrate zu berechnen (Abbildung 22). Nachdem die ermittelten Herzraten und die tatsächlichen Herzraten für jedes Video im Datensatz vorliegen, werden die eingesetzten rPPG-Algorithmen auf ihre Vorhersagegenauigkeit hin evaluiert. Um die Genauigkeit der Algorithmen zu bewerten, werden die ermittelten Herzraten mit den tatsächlichen Herzraten auf Übereinstimmung untersucht. Im nachfolgenden Abschnitt werden die Metriken vorgestellt, die es erlauben die Qualität der Herzratenschätzungen durch die rPPG-Algorithmen zu bewerten.

In der rPPG-Toolbox sind zusätzliche optionale Nachbearbeitungsschritte verfügbar. Die rPPG-Toolbox ermöglicht die Anwendung kleinerer Videofenster aus den aufgezeichneten Videodaten. Die Toolbox speichert die verarbeiteten Videos standardmäßig mit einer Auflösung von 72x72 Pixeln im .npy-Format auf der SSD. Die Wahl der Auflösung von 72x72 Pixeln kann je nach den Anforderungen des eingesetzten Algorithmus über die Konfigurationsdatei des Datensatzes angepasst werden.



Abbildung 22.: FFT des Referenz- und des extrahierten BVP-Signals durch den POS-Algorithus. Die Herzrate entspricht normalerweise dem Hauptpeak im Frequenzspektrum des BVP-Signals. Die rot gestrichelte Linie markiert das Maximum des Frequenzspektrums, welches die Herzrate repräsentiert.

#### 5.8. Evaluationsverfahren

Die Auswahl der Evaluationsmetriken für die Beurteilung der Modellqualität basiert auf etablierten Standards zur Bewertung der Genauigkeit der prognostizierten Herzrate im Vergleich zur Referenzherzrate. Diese Metriken wurden spezifisch wegen ihrer Fähigkeit gewählt, verschiedene Aspekte der Vorhersagegenauigkeit abzubilden und eine umfassende Modellbewertung zu ermöglichen, wobei der Fokus auf die Genauigkeit der Herzraten, deren Korrelation und die Anzahl der Abweichungen gelegt wird.

Die rPPG-Toolbox verfügt bereits über fünf vorimplementierte Metriken zur Bewertung der berechneten Herzrate im Vergleich zur Referenzherzrate. Diese gehören zu den gängigen Metriken innerhalb der rPPG-Forschung, um die Qualität eines Vorhersagemodells zu evaluieren.

Die fünf implementierten Metriken sind [39]:

• Mittlerer absoluter Fehler (MAE): Der MAE ist eine Metrik, die die durchschnittliche absolute Differenz zwischen den Vorhersagen des Modells  $R_p$  und den tatsächlichen Werten  $R_{gt}$  über eine Stichprobengröße von n Vorhersagen misst:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |R_{gt,i} - R_{p,i}|$$
(5.5)

Ein niedriger MAE zeigt an, dass das Modell im Durchschnitt geringe Abweichungen bei den Vorhersagen aufweist.

• Mittlerer Quadratischer Fehler (RMSE): Der RMSE ist ähnlich dem MAE, misst jedoch die Quadratwurzel des durchschnittlichen quadratischen Fehlers zwischen Vorhersagen und tatsächlichen Werten:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (R_{gt,i} - R_{p,i})^2}$$
(5.6)

Der RMSE neigt dazu, größere Fehler stärker zu bestrafen als der MAE.

• Mittlerer Absoluter Prozentualer Fehler (MAPE): Der MAPE ist eine relative Fehlermetrik, die den durchschnittlichen prozentualen Unterschied zwischen Vorhersagen und tatsächlichen Werten misst:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{|R_{gt,i} - R_{p,i}|}{R_{gt,i}} \right) \times 100\%$$
(5.7)

Der MAPE ist nützlich, um die relative Genauigkeit des Modells zu bewerten, insbesondere wenn die Skalen der Werte variieren.

 Signal-zu-Rausch-Verhältnis (SNR): Das SNR ist ein Maß dafür, wie stark das Signal (die relevanten Informationen) im Vergleich zum Rauschen (unerwünschte Informationen) ist. In der rPPG-Toolbox ist die Berechnung des SNR für ein vorhergesagtes Signal implementiert nach der Methode von De Haan et al. [31]. Hiernach berechnet sich das SNR aus dem Verhältnis des Flächeninhalts des Leistungsspektrums S um die ersten beiden Harmonischen der Grundfrequenz und des Flächeninhaltes des restlichen Leistungsspektrums:

$$SNR = 10 \cdot \log_{10} \left( \frac{\sum_{30}^{250} (U_t(f)S(f))^2}{\sum_{30}^{250} ((1 - U_t(f))S(f))^2} \right)$$
(5.8)

 $U_t(f)$  beträgt 1 im Bereich der ersten und zweiten Harmonischen des tatsächlichen rPPG-Signals, während es im übrigen Leistungsspektrum 0 beträgt. Zum Zweck des statistischen Vergleichs wird aus den SNR-Werten jedes Videos der Mittelwert gebildet, um das SNR des gesamten Datensatzes zu erhalten:

$$SNR = \frac{1}{n} \sum_{i=1}^{n} SNR_i$$
(5.9)

• Pearson-Korrelations-Koeffizient ( $\rho$ ): Der Pearson- Korrelationskoeffizient misst die Stärke und Richtung der linearen Beziehung zwischen Vorhersagen und tatsächlichen Werten. Der Pearson-Korrelationskoeffizient ergibt sich aus der vorhergesagten Herzrate  $R_p$ , der tatsächlichen Herzrate  $R_{gt}$ , nVorhersagen und  $\overline{R}$  als Mittelwert von R für n Stichproben:

$$\rho = \frac{\sum_{i=1}^{n} \left( R_{gt,i} - \overline{R}_{gt} \right) \left( R_{p,i} - \overline{R}_{p} \right)}{\sqrt{\left( \sum_{i=1}^{n} R_{gt,i} - \overline{R}_{gt} \right)^{2} \left( \sum_{i=1}^{n} R_{p,i} - \overline{R}_{p} \right)^{2}}}$$
(5.10)

 $\rho=1$  be deutet eine perfekte positive lineare Beziehung, während  $\rho=-1$  eine perfekte negative lineare Beziehung anz eigt. Zusätzlich wird der Standardfehler für ein besseres Verständnis der Genauigkeit der oben genannten Metriken berechnet. Der Standardfehler ist ein statistisches Maß für die Genauigkeit einer Schätzung. Der Standardfehler gibt die Standardabweichung sämtlicher Stichproben an, die aus der Grundgesamtheit gezogen werden. Dies ist insbesondere bei der Analyse von rPPG-Datensätzen relevant, wo die Stichprobengröße von Datensatz zu Datensatz erheblich variieren kann. Je kleiner der Standardfehler ist, desto genauer ist die Schätzung. Ein größerer Standardfehler weist auf größere Unsicherheit in der Schätzung hin.

Für alle Metriken mit Ausnahme der Pearson-Korrelation ( $\rho$ ) berechnet sich der Standardfehler SE zu:

$$SE = \frac{\sigma}{\sqrt{n}} \tag{5.11}$$

Mit der Standarabweichung  $\sigma$  und der Stichprobengröße n. Bei der Pearson-Korrelation berechnet sich der Standardfehler aus dem Pearson-Korrelationskoeffizienten  $\rho$  und der Stichprobengröße n zu:

$$SE_{\rho} = \sqrt{\frac{1-\rho^2}{n-2}}$$
 (5.12)

Zusätzlich zu den bereits implementierten Metriken wurde rPPG-Toolbox im Rahmen dieser Arbeit um eine weitere Genauigkeitsmetrik (Accuracy) erweitert. Die Accuracy-Metrik ist eine weit verbreitete Bewertungsmetrik im Bereich des Machine-Learning und der Klassifikationsstatistik. Sie wird verwendet, um die Leistung eines Klassifikationsmodells zu bewerten. Die Accuracy-Metrik gibt an, wie viele der insgesamt vorhergesagten Herzraten korrekt klassifiziert wurden, im Verhältnis zur Gesamtanzahl der untersuchten Videoausschnitte.

Die Messgenauigkeit der verwendeten Fingerclippulsoximeter zur Aufzeichnung der vefügbaren Datensätze beträgt 2 BPM beim *Contec CMS50E*-Gerät und 1 BPM beim *Contec CMS60C*-Gerät. Zum *HKG-07C*-Gerät, das bei der Aufzeichnung vom MMPD-Datensatz zum Einsatz kam, liegt kein Datenblatt vor. Die hier eingeführte Definition der Accuracy orientiert sich an der maximalen Messungenauigkeit, die in den aufgezeichneten Referenzsignalen vorliegen kann. Eine vorhergesagte Herzrate wird als korrekt gewertet, wenn die Vorhersage weniger als 2 BPM von der tatsächlichen Herzrate abweicht. Die Accuracy berechnet sich aus dem Verhältnis der korrekten Herzraten  $n_{korrekt}$  zur Anzahl aller vorhergesagten Herzraten n zu:

$$ACC = \frac{n_{korrekt}}{n} \cdot 100\%$$
(5.13)

Die vorgestellten Schritte zur Datenvorverarbeitung werden zur Klärung der Forschungsfrage eingesetzt, um die optimale ROI im Gesichtsbereich zu ermitteln, welche im weiteren Verlauf der Arbeit als Eingang des Erscheinungsbildzweigs im *TS-CAN*-Modell zum Einsatz kommen soll (siehe Abschnitt 3.1.2). Anhand der Bewertungsmetriken sollen die Auswirkungen der untersuchten Einschränkungen im Gesichtsbereich für die rPPG-Signalextraktion evaluiert werden. Im nachfolgenden Kapitel werden die Untersuchungen vorgestellt und ausgewertet, die zur Ermittlung der optimalen ROI in realen Einsatzszenarien führen.

# 6. Entwicklung der adaptiven ROI

In diesem Kapitel werden Untersuchungen vorgestellt, die mit dem Ziel durchgeführt werden, die optimale ROI-Definition bei realen Anwendungsszenarien zu ermitteln. Grundlage hierfür sind die in Kapitel 4 vorgestellten Datensätze zur rPPG-Forschung. Anhand der Schritte zur Datenvorverarbeitung, wie sie in Kapitel 5 beschrieben sind, wird evaluiert, wie sich die Einschränkung des Gesichtsbereichs in kleinere Teilbereiche auf die Qualität der rPPG-Signalextraktion auswirkt. Während den Untersuchungen werden sechs unüberwachte rPPG-Algorithmen (ICA, POS, CHROM, GREEN, LGI und PBV) verwendet, um die rPPG-Signale aus den Gesichtsvideos mit Hilfe der rPPG-Toolbox [39] zu extrahieren (siehe Abschnitt 3.1.2). Die Qualität der Extraktion wird mithilfe der Evaluationsmetriken aus Abschnitt 5.8 bewertet.

Die Erkenntnisse aus den Untersuchungen sollen schrittweise zur Definition des aussagekräftigsten Gesichtsbereichs für die Signalextraktion führen. Der optimale Gesichtsbereich soll insbesondere mit dem Fokus auf bewegungsreiche Einsatzszenarien eingesetzt werden, indem dieser im weiteren Verlauf der Arbeit die Aufmerksamkeitsmaske des *TS-CAN*-Modells definiert (siehe Abschnitt 3.1.2) und dadurch den Fokus des Bewegungszweigs auf den ermittelten Gesichtsbereich lenken soll.

Unter Berücksichtigung von Forschungserkenntnissen aus vorangegangenen Arbeiten werden zur Klärung der Forschungsfrage mehrere Untersuchungen durchgeführt, die schrittweise zu einer Annäherung an die optimale Gesichtsregion führen sollen. Die erste Untersuchung baut auf den Erkenntnissen von Wong et al. [44] auf und beschäftigt sich mit der Auswirkung des Reflexionswinkels zwischen der Gesichtsoberfläche und der optischen Achse der Kamera auf die Qualität der rPPG-Signalextraktion (siehe Abschnitt 3.1.2) Ein weiteres Untersuchungsgebiet wird die Hautsegmentierung sein, die mit der Motivation durchgeführt wird, Gesichtsbereiche, die keine pulsatilen Informationen enthalten, für die Signalextraktion zu entfernen und gleichzeitig den Einfluss der epidermalen Melaninkonzentration auf die Signalextraktion zu untersuchen. Weiterhin wird die Begrenzung der Gesichtsregion auf kleinere ROIs im Bereich der Stirn und der Wangen untersucht. Zusätzlich wird geprüft, ob die Kombination dieser ROI-Begrenzung mit der Begrenzung des maximal zulässigen Reflexionswinkels zu einer verbesserten Vorhersage der Herzrate führt. Darauf aufbauend wird untersucht, ob die Kombination mehrerer extrahierter Herzraten aus den separat analysierten ROIs der Stirn und Wangen zu einer robusteren Vorhersage der Herzrate führt. Zuletzt wird die Auswirkung unterschiedlich langer BVP-Signalverläufe auf die Vorhersageleistung der rPPG-Algorithmen analysiert.

#### 6.1. Untersuchung des optimalen Winkelschwellenwertes

Aufbauend auf den Erkenntnissen von Wong et al. (siehe Abschnitt 3.1.2) wird in dieser Untersuchung die Auswirkung verschiedener Winkelschwellenwerte auf die Leistung der rPPG-Methoden bewertet. Die

Ergebnisse von Wong et al. [44] legen nahe, dass eine Anpassung der ROI bei variierenden Gesichtsorientierungen die Möglichkeit bietet, die Auswirkungen von Bewegungsartefakten und Beleuchtungsvariationen zu minimieren und gleichzeitig die Bereiche auszuwählen, die die meisten physiologischen Informationen enthalten.

Ziel dieser Untersuchung ist es, die dynamische Anpassung der ROI während der Analyse von Gesichtsvideos zu evaluieren und die Abhängigkeit der Signalqualität von der Oberflächenorientierung zu prüfen. Die Winkelschwellenwerte definieren, welche Dreiecke aus der Gesichtsparkettierung in den Gesichtsbereich einfließen, aus dem die rPPG-Algorithmen ein BVP-Signal extrahieren. Liegt der Reflexionswinkel eines Dreiecks unterhalb des Winkelschwellenwerts, so wird das Dreieck in jedem Fall in die ROI mit aufgenommen. Ist der Reflexionswinkel leicht überschwellig, wird mithilfe des Akzeptanzchecks (siehe Abschnitt 5.5) überprüft, ob das Dreieck aufgrund seiner vergangenen fünf Reflexionswinkel in die ROI aufgenommen wird. In allen anderen Fällen wird das Dreieck aus der Gesichtsberginswinkel einfernt und der Gesichtsbereich erscheint geschwärzt.

Speziell soll ermittelt werden, ob die Senkung des Schwellenwerts von 90° auf 15°, 30°, 45° oder 60° eine Verbesserung der Vorhersagegenauigkeit ergibt oder ob diese Änderungen keine signifikanten Auswirkungen auf die Ergebnisse haben (Abbildung 23). Aus Gründen der Übersichtlichkeit werden in diesem Abschnitt die Ergebnisse aus der Verarbeitung des UBFC-rPPG-Datensatzes vorgestellt. Die Evaluation erfolgt anhand der Bewertungsmetriken, die durch die Anwendung der unüberwachten rPPG-Methoden entstehen. Die vorhergesagten und tatsächlichen BVP-Signale werden im Rahmen der Evaluation in sich nicht überlappende Zeitfenster von 30 Sekunden Länge unterteilt, um eine höhere zeitliche Auflösung für die Herzratenvorhersage zu erhalten. Umfassende Ergebnisse zu allen Datensätzen befinden sich in Anhang A.2.

Die adaptive Anpassung der ROI soll speziell für Alltagsszenarien konzipiert werden, in denen Bewegungen auftreten. Das Bewegungsszenario stellt eine Herausforderung für rPPG-Algorithmen dar, da Bewegungsartefakte, die Ortsveränderung der ROI oder Variationen in der Gesichtsbeleuchtung das extrahierte Signal stören können. Durch die Analyse der Leistungsfähigkeit von rPPG-Methoden unter diesen Bedingungen soll untersucht werden, wie robust diese Algorithmen gegenüber Bewegungen sind. Daher werden im Rahmen dieser Untersuchung Bewegungsszenarien betrachtet, sofern diese in den Datensätzen enthalten sind. Beim PURE-, VIPL-HR-V1- und MMPD-Datensatz wird ausschließlich das Bewegungsszenario verwendet, während beim RLAP-rPPG-Datensatz ausschließlich das Videospielszenario berücksichtigt wird, bei dem vergleichbare körperliche Aktivitäten in Kombination mit geistiger Anstrengung vorliegen. Der KISMED-Datensatz weist in den Bewegungsszenarien, die translatorische Bewegungen des Probanden enthalten, Bewegungsartefakte in den Referenz-BVP-Signalen auf, welche den Vergleich zwischen den vorhergesagten und den Referenz-BVP-Signalen ohne zusätzlicher Datenaufbereitung des artefaktbelasteten Referenzsignals beeinträchtigen. Aus diesem Grund wird für diese Untersuchung nur das Rotationsszenario (V11) des KISMED-Datensatzes untersucht. Dieses Szenario erfüllt die Anforderung, Variationen in der Gesichtsorientierung zu enthalten und dabei zuverlässige Referenz-BVP-Signale aufzuweisen.

Die Untersuchung ergibt, dass die Senkung des Winkelschwellenwerts von 90° auf geringere Winkelschwellenwerte keine verbesserte Vorhersagegenauigkeit der Herzrate liefert, sondern zu einer Verschlechterung führt (Tabelle 6). Die höchste Vorhersagegenauigkeit bei der rPPG-Extraktion liefert ein Winkelschwellenwert von 90°. Dieses Ergebnis trifft für sämtliche untersuchte rPPG-Datensätze und die meisten unüberwachten



(a) Winkelschwellenwert =  $15^{\circ}$ 



(b) Winkelschwellenwert =  $30^{\circ}$ 



(c) Winkelschwellenwert =  $45^{\circ}$ 



(d) Winkelschwellenwert =  $60^{\circ}$ 

Abbildung 23.: Während dieser Untersuchung extrahierte Dreiecke der Gesichtsparkettierung, deren Reflexionswinkel unterhalb des Schwellenwerts liegen.

rPPG-Algorithmen zu. Lediglich beim UBFC-rPPG-Datensatz führt ein niedrigerer Winkelschwellenwert von 60° bei Anwendung des CHROM-Algorithmus zu einer Verbesserung der Evaluationsmetriken.

Die Diskrepanz zu den Ergebnissen von Wong et al. lässt sich dadurch erklären, dass in ihrer Studie nur der UBFC-rPPG-Datensatz verwendet und der CHROM-Algorithmus eingesetzt wurde, um das rPPG-Signal aus dem Gesichtsbereich zu extrahieren. Dies deckt sich damit, dass in der hier durchgeführten Untersuchung des UBFC-rPPG-Datensatzes der CHROM-Algorithmus der einzige ist, bei dem niedrigere Winkelschwellenwerte zu einer Verbesserung der Evaluationsmetriken führt. Beim RLAP- und MMPD-Datensatz zeigt sich ebenfalls eine Verbesserung der Vorhersageleistung des CHROM-Algorithmus bei niedrigeren Winkelschwellenwerten, während die anderen unüberwachten rPPG-Algorithmen unabhängig vom Datensatz ihr Optimum bei einem Schwellenwert von 90° aufweisen.

Dies deutet darauf hin, dass ein niedrigerer Schwellenwert grundsätzlich nicht vorteilhaft ist. Stattdessen legen die Ergebnisse nahe, dass eine größere Hautfläche einen signifikanten Vorteil zur Extraktion des BVP-Signals liefert. Dies ist darauf zurückzuführen, dass die angewandten rPPG-Algorithmen, abgesehen vom GREEN-Algorithmus, eine Mittelung der Pixelfarbwerte über den Bildbereich vornehmen. Die Nutzung eines größeren Gesichtsbereichs ermöglicht es den rPPG-Algorithmen, eine größere Anzahl von Hautpixeln, die pulsatile Informationen beinhalten, für die Signalextraktion zu verwenden. Gleichzeitig können störende Faktoren wie Bildrauschen, das als lokale Störung der Bildintensität erscheint, durch die Mittelung über einen größeren Bereich kompensiert werden. Dadurch wird das Verhältnis von pulsatilen Nutzsignal zum Störsignal verbessert, was letztlich zu einer höheren Signalqualität des BVP-Signals führen kann.

Dieser Umstand soll im nachfolgenden Abschnitt näher untersucht werden. Nachdem die Auswirkung der Oberflächenorientierung auf die Vorhersagegenauigkeit der rPPG-Algorithmen analysiert wurde, konzentriert sich die folgende Untersuchung auf die Segmentierung der Haut von bedeckten Hautbereichen, die keine pulsierenden Informationen enthalten. Ein weiteres Augenmerk gilt hierbei der Erforschung des Einflusses der Hautpigmentierung. Wie bereits in Kapitel 2.2.1 beschrieben wird, führt eine erhöhte Konzentration von epidermalem Melanin zu einer stärkeren Absorption des innerhalb der Blutgefäße diffus reflektierten Lichts, was zu einer verminderten Amplitude des Nutzsignals führt. Diese nachfolgende Analyse soll Aufschluss darüber geben, wie sich die Entfernung von Hautbereichen ohne pulsierende Informationen auf die Vorhersagegenauigkeit der rPPG-Algorithmen auswirkt und inwiefern die Hautpigmentierung dabei eine Rolle spielt.

Tabelle 6.: Vergleich der Begrenzung des Gesichtsbereichs anhand unterschiedlicher Winkelschwellenwerte für den<br/>UBFC-rPPG-Datensatz. Die Auswirkung der Winkelschwellenwerte zwischen 15° und 90° werden mit der<br/>vorimplementierten Datenvorverarbeitung der rPPG-Toolbox verglichen. Im Allgemeinen Trend verbessern<br/>sich die Evaluationsmetriken mit zunehmendem Winkelschwellenwert.

	Datensatz: UBFC-rPPG																	
Winkel- schwellwert	1	rPPG-T	oolbox B	enchma	rk			15°				30°						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	M	IAE	RMSE	Pearson	SNR	Acc		
ICA	13,92	23,78	0,56	-8,88	52,38	6,03	14,37	0,70	-2,47	59,52	3	,93	11,20	0,86	0,33	76,19		
POS	3,86	9,01	0,90	-2,59	64,29	3,11	6,96	0,92	0,23	66,67	1	,63	4,04	0,97	2,48	78,57		
CHROM	3,76	11,06	0,84	-3,51	78,57	4,11	11,15	0,81	-1,84	59,52	1	,64	3,41	0,98	-0,11	71,43		
GREEN	19,12	30,27	0,41	-11,16	40,48	29,35	34,69	0,12	-18,41	4,76	26	5,42	31,97	0,20	-17,34	11,90		
LGI	15,68	28,57	0,36	-8,13	52,38	1,54	3,89	0,98	-1,97	73,81	1	,51	3,77	0,98	0,66	73,81		
PBV	15,77	26,37	0,48	-9,32	45,24	11,38	19,82	0,46	-6,67	45,24	7	,60	16,26	0,67	-5,79	59,52		
Winkel- schwellwert			45°					60°						90°				
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	M	IAE	RMSE	Pearson	SNR	Acc		
ICA	3,58	10,49	0,87	0,21	76,19	3,39	9,85	0,88	0,06	73,81	2	,09	6,24	0,94	0,90	78,57		
POS	1,73	4,01	0,97	2,91	73,81	1,55	3,79	0,98	2,97	76,19	1	,48	3,74	0,98	2,97	78,57		
CHROM	1,91	4,35	0,97	0,19	71,43	1,75	4,28	0,97	-0,004	78,57	2	,10	5,18	0,96	0,19	76,19		
GREEN	24,83	30,49	0,21	-16,15	16,67	23,89	30,11	0,07	-15,53	16,67	23	3,48	32,02	-0,04	-15,01	21,43		
LGI	1,46	3,74	0,98	0,69	76,19	1,44	3,76	0,98	0,68	76,19	1	,40	3,89	0,98	0,31	78,57		
PBV	6,95	15,31	0,73	-3,09	64,29	5,69	13,28	0,75	-1,14	64,29	5	,02	12,39	0,75	-1,93	66,67		

# 6.2. Untersuchung zur Hautsegmentierung

Ziel dieser Untersuchung ist es, die Auswirkung des Melaningehalts in der Epidermis auf die Signalextraktion durch die unüberwachten rPPG-Algorithmen zu untersuchen. Hierfür wird das Bewegungsszenario im MMPD-Datensatz eingesetzt. Beim MMPD-Datensatz liegen umfassende Annotationen vor, die es ermöglichen, die Probanden anhand ihrer Erscheinungsmerkmale, wie Hautfarbe, Geschlecht, Brillenträger oder aufgetragenem Make-up zu unterscheiden. Die Datenvorverarbeitung beschränkt sich im Rahmen dieser Untersuchung auf die Segmentierung des gesamten Gesichtsbereichs. Hierzu wird der Schwellenwert für den maximalen zulässigen Reflexionswinkel eines Dreiecks in der Gesichtsparkettierung zu 90° festgelegt, wodurch der gesamte Gesichtsbereich vom Hintergrund segmentiert wird.

Zugleich soll die Auswirkung der implementierten histogrammbasierten Hautsegmentierung auf die Vorhersagegenauigkeit evaluiert werden. Hierzu wird jeder Hauttyp nach Fitzpatrick, der im Bewegungsszenario des MMPD-Datensatzes vorhanden ist, in zwei Durchgängen evaluiert. Im ersten Durchgang wird keine histogrammbasierte Hautsegmentierung vorgenommen (siehe Tabelle 7), im zweiten Durchgang (siehe Tabelle 8) ist die Hautsegmentierung aktiv und entfernt nach dem in Abschnitt 5.4 beschriebenen Vorgehen diejenigen Hautpixel, deren grüne und blaue Farbkanalwerte unterhalb des jeweiligen Farbwertschwellenwerts liegen. Eine weitere Segmentierung des Gesichtsbereichs in spezifische ROIs wird nicht vorgenommen, um die Auswirkung der Hautsegmentierung isoliert von zusätzlichen Einschränkungen des signaltragenden Hautbereichs zu untersuchen.

				Dat	tensatz	z:	MMPD									
Fitzpatrick Hauttyp		III (	68 Szen	arien)				IV (1	6 Szena	rien)						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc					
ICA	16,80	24,81	0,0522	-10,61	35,29		30,71	36,45	-0,1212	-14,86	9,38					
POS	8,28	15,84	0,4741	-4,57	59,56	1	20,82	28,79	0,0506	-11,55	25,00					
CHROM	14,32	22,60	0,1130	-7,91	42,65		27,96	34,82	-0,1229	-14,89	15,62					
GREEN	18,66	23,53	0,0772	-13,55	14,71		26,64	34,43	-0,2316	-15,37	6,25					
LGI	11,36	19,64	0,3477	-6,14	52,94		29,50	36,85	0,0305	-13,86	18,75					
PBV	20,20	27,35	0,0707	-12,05	25,00	ĺ	33,12	38,06	0,1998	-17,40	6,25					
Fitzpatrick Hauttyp		V (	24 Szena	arien)				VI (2	4 Szena	narien)						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc					
ICA	28,42	35,95	0,0979	-16,84	12,50		30,94	37,46	-0,0998	-18,03	10,42					
POS	26,33	32,30	0,1100	-15,20	12,50		25,60	33,01	0,1168	-15,73	14,58					
CHROM	29,48	35,68	0,0644	-17,70	10,42		29,92	34,91	0,1694	-17,43	4,17					
GREEN	28,86	34,35	-0,1148	-17,03	8,33		29,92	36,94	-0,1043	-16,83	6,25					
LGI	29,55	36,28	0,0607	-16,55	14,58		31,13	38,18	-0,0451	-16,70	10,42					
PBV	32,19	38,37	0,1495	-18,09	10,42	ĺ	29,74	36,95	0,0234	-18,85	8,33					

Tabelle 7.: Evaluation der sechs unüberwachten rPPG-Algorithmen für die Hauttypen nach Fitzpatrick von III bisIV anhand des Bewegungsszenarios im MMPD-Datensatz ohne histogrammbasierter Segmentierung der<br/>Haut.

Die vorhergesagten und tatsächlichen BVP-Signale werden im Rahmen der Evaluation in sich nicht überlappende Zeitfenster von 30 Sekunden Länge unterteilt. Die Ergebnisse der Evaluation zeigen, dass die rPPG-Algorithmen mit zunehmendem Melaningehalt schlechtere Resultate bei der Extraktion des rPPG-Signals liefern. Dieser Umstand ist auf die erhöhte Lichtabsorption durch den höheren Melaningehalt zurückzuführen, was wiederum zu einem geringeren Anteil an diffus reflektiertem Licht im Vergleich zum spiegelnd reflektierten Licht führt. Dies zeigt sich besonders deutlich in der Abnahme des SNR bei den melaninreicheren Hauttypen IV, V und VI.

Die Hautsegmentierung wird mit dem Ziel eingesetzt, die Gesichtshaut von den gesichtsbedeckenden Elementen zu trennen und dadurch die Signalextraktion auf Hautbereiche zu fokussieren, die möglicherweise pulsatile Informationen enthalten. Die Evaluationsmetriken wie der MAE, RMSE und insbesondere das SNR werden durch die aktive Hautsegmentierung geringfügig verbessert, wohingegen die Vorhersagegenauigkeit durch den Einsatz der Hautsegmentierung abnimmt. Insbesondere die Zunahme des SNR lässt darauf schließen, dass durch die Hautsegmentierung erfolgreich Hautbereiche entfernt werden, die keine pulsatilen

**Tabelle 8.:** Evaluation der sechs unüberwachten rPPG-Algorithmen für die Hauttypen nach Fitzpatrick von III bis IV<br/>anhand des Bewegungsszenarios im MMPD-Datensatz mit aktiver histogrammbasierter Segmentierung<br/>der Haut für jedes Videobild.

	Datensatz: MMPD																
Fitzpatrick Hauttyp		III (	(68 Szen	arien)			IV (16 Szenarien)										
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc						
ICA	15,83	23,55	0,1559	-9,96	34,56		30,82	38,31	-0,094	-15,30	9,38						
POS	8,27	15,80	0,4923	-4,38	55,88		21,92	29,28	-0,2072	-11,34	21,88						
CHROM	13,44	22,32	0,0907	-7,40	46,32		28,89	34,53	-0,0622	-14,06	12,50						
GREEN	18,55	23,14	0,1167	-13,54	13,24		26,04	30,92	-0,1025	-15,83	9,38						
LGI	11,35	19,90	0,3101	-5,78	52,94		28,62	34,04	0,2176	-13,43	15,62						
PBV	18,66	25,63	0,0646	-11,40	27,21		32,08	37,66	-0,0089	-16,72	6,25						
Fitzpatrick Hauttyp		V (	24 Szena	arien)			VI (24 Szenarien)										
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc						
ICA	27,54	33,67	0,1647	-17,32	8,33		27,21	34,20	0,1877	-17,33	6,25						
POS	23,03	29,41	0,2807	-15,78	14,58		24,65	31,31	0,3149	-15,41	12,50						
CHROM	27,21	34,06	0,1387	-16,82	12,50		28,49	34,57	0,1642	-16,75	6,25						
GREEN	27,87	34,19	-0,0735	-16,36	8,33		29,88	35,79	-0,3807	-17,01	2,08						
LGI	27,28	33,94	0,2939	-16,51	16,67		26,29	31,82	0,3459	-16,94	12,50						
PBV	26,70	34,02	0,1371	-16,98	14,58		30,94	36,29	0,007	-17,29	4,17						

Informationen und daher kein Nutzsignal für die Extraktion des BVP-Signals enthalten. Jedoch erreicht die Hautsegmentierung im Vergleich zum Vorgehen ohne aktive Hautsegmentierung nicht den erwarteten Nutzen bezüglich einer gesteigerten Vorhersagegenauigkeit, wie sie durch die Accuracy-Metrik definiert ist.

In vorangegangenen Arbeiten stellten Kwon et al. [40] und Fouad et al. [41] fest, dass eine Begrenzung des Gesichtsbereichs in Regionen mit vergleichsweise geringer Hautdicke und hoher Durchblutung eine signifikante Auswirkung auf die Signalqualität haben, da in diesen Hautregionen das Verhältnis zwischen pulsatilem Nutzsignal und rauschbehafteten Störsignal besonders hoch ausfällt (siehe Kapitel 3.1.2). Hierbei hat sich in den vorangegangenen Forschungen wiederholt der Bereich an der Stirn und an der linken und rechten Wange als praktikable Messregion erwiesen. Im nachfolgenden Abschnitt wird untersucht, inwiefern die Begrenzung des Gesichtsbereichs in zuvor definierte ROIs, die auf die Stirn- und Wangenbereiche beschränkt sind, auf die Vorhersagegenauigkeit der sechs unüberwachten rPPG-Algorithmen bei bewegten Probanden auswirkt. Speziell soll untersucht werden, ob die Segmentierung des Gesichtsbereichs in die genannten ROIs im Vergleich zur Verwendung des gesamten Gesichtsbereichs eine Steigerung des SNR hervorruft.

# 6.3. Untersuchung zur Begrenzung der ROI

Das Ziel dieser Untersuchung besteht darin, die Auswirkungen der Segmentierung des Gesichtsbereichs in kleinere ROIs auf die Leistung der rPPG-Algorithmen zu bewerten. Es soll insbesondere untersucht werden, ob die Begrenzung der ROIs auf die Stirn- und Wangenregionen bei Bewegungsszenarien eine Steigerung des SNR hervorruft und einen Vorteil für die Vorhersagegenauigkeit der Herzrate liefert. Zusätzlich wird geprüft, ob die Anpassung der ROI-Begrenzung in bewegungsintensiven Szenarien – beispielsweise durch die Einführung eines geeigneten Winkelschwellenwerts – die Vorhersagegenauigkeit und Robustheit der rPPG-Algorithmen gegenüber Bewegungen und geänderten Lichtverhältnissen verbessert.

Es werden verschiedene Ansätze zur ROI-Begrenzung getestet, darunter die Segmentierung der Gesichtsregion in ROIs, die auf die Stirn- und Wangenbereiche beschränkt sind (Abbildung 24c) oder Kombination der ROI-Begrenzung auf diese Gesichtsregionen mit der zusätzlichen Einschränkung des maximal zulässigen Reflexionswinkels auf 45° (Abbildung 24d). Die Definition der Stirn- und Wangen-ROIs erfolgt gemäß dem Verfahren, dass in Abschnitt 5.5 beschrieben wird. Diese beiden Ansätze zur ROI-Segmentierung werden mit der vorimplementierten Datenvorverarbeitung in der rPPG-Toolbox (Abbildung 24a) und mit der selbst implementierten Extraktion des gesamten Gesichtsbereichs verglichen (Abbildung 24b).



(a) Voreingestellte Gesichtsextraktion in der rPPG-Toolbox



(b) Gesichtsbereich mit Reflexionswinkeln unterhalb 90°



(c) Stirn- & Wangen- ROI mit Reflexionswinkeln unter 90°



(d) Stirn- & Wangen- ROI mit Reflexionswinkeln unter 45°

Abbildung 24.: Während dieser Untersuchung extrahierte Gesichtsbereiche. Abbildung 24a stellt einen anderen Probanden als die übrigen Abbildungen dar.

Die Untersuchung hat das Ziel, den potenziellen Vorteil der ROI-Segmentierung bei für die Signalextraktion herausfordernden Aktivitäten zu untersuchen. Sofern vorhanden, wird daher nur das Bewegungsszenario in den Datensätzen untersucht. Die Datensätze, für die gesondert ein Bewegungsszenario vorliegt, sind der PURE-, VIPL-HR-V1- und MMPD-Datensatz. Beim RLAP-rPPG-Datensatz wird aufgrund vergleichbarer körperlicher Aktivität das Videospielszenario verwendet. Beim KISMED-Datensatz wird das Rotationsszenario (V11) untersucht. Der UBFC-rPPG- und COHFACE-Datensatz wird vollständig untersucht. Wie in den zuvor durchgeführten Untersuchungen erfolgt die Extraktion der rPPG-Signale durch die sechs unüberwachten rPPG-Algorithmen. Anschließend werden die extrahierten Signale mithilfe der Evaluationsmetriken bewertet, um die Vorhersagegenauigkeit der rPPG-Algorithmen zu vergleichen. Die vorhergesagten und tatsächlichen BVP-Signale werden im Rahmen der Evaluation in sich nicht überlappende Zeitfenster von 30 Sekunden Länge unterteilt, um eine höhere zeitliche Auflösung für die Herzratenvorhersage zu erhalten. Die Ergebnisse des UBFC-rPPG-Datensatzes sind in Tabelle 9 aufgetragen. Die Resultate der übrigen Datensätze befinden sich in Anhang A.3.

Tabelle 9.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für den UBFC-rPPG-Datensatz. Beim ICA-,<br/>POS-, und PBV-Algorithmus zeigen die Ergebnisse, dass die Extraktion des rPPG-Signals aus der Stirn- und<br/>Wangen-ROI die besten Metriken erzielt. Zudem konnte die ROI-Segmentierung das Signal-zu-Rausch-<br/>Verhältnis (SNR) verbessern. Beim CHROM-, LGI- und PBV-Algorithmus erweist sich die ROI-Segmentierung<br/>und die Begrenzung des Reflexionswinkels auf maximal 45° als vorteilhaft für die Signalextraktion.

	Datensatz: UBFC-rPPG																					
Segmen- tierung	rPPG-Toolbox Benchmark						° ohne	ROI Segr	nentierı	ıng	9	0° mit I	ROI Segm	entieru	ng	4	45° mit ROI Segmentierung					
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc		
ICA	13,92	23,78	0,5596	-8,88	52,38	2,09	6,24	0,9428	0,90	78,57	1,08	2,40	0,9909	1,96	80,42	3,72	10,13	0,8598	0,82	73,81		
POS	3,86	9,01	0,8954	-2,59	64,29	1,48	3,74	0,9782	2,97	78,57	1,42	3,59	0,9753	3,61	78,57	1,50	3,93	0,9760	3,57	76,19		
CHROM	3,76	11,06	0,8413	-3,51	78,57	2,10	5,18	0,9594	0,19	76,19	1,89	4,43	0,9679	0,90	73,81	1,77	4,14	0,9740	0,78	71,43		
GREEN	19,12	30,27	0,4093	-11,16	40,48	23,48	32,02	-0,0404	-15,01	21,43	21,91	28,27	0,0668	-14,95	14,29	22,73	27,74	0,0508	-15,75	11,90		
LGI	15,68	28,57	0,3608	-8,13	52,38	1,40	3,89	0,9761	0,31	78,57	1,40	3,89	0,9761	1,70	78,57	1,38	3,84	0,9769	1,73	78,57		
PBV	15,77	26,37	0,4791	-9,32	45,24	5,02	12,39	0,7527	-1,93	66,67	4,46	11,92	0,7824	-1,16	71,43	3,22	9,29	0,8749	-0,48	73,81		

Die Ergebnisse zeigen eine konsistente Tendenz über die verschiedenen Datensätze hinweg. Allerdings unterscheidet sich die Tendenz zwischen den rPPG-Algorithmen. Bei Anwendung des GREEN-Algorithmus ergibt die ROI-Begrenzung keinen Vorteil für die Signalextraktion, was durch die Verschlechterung der Evaluationsmetriken und insbesondere durch die Verringerung der Vorhersagegenauigkeit deutlich wird. Während beim CHROM-Algorithmus ebenso eine Verschlechterung der Vorhersagegenauigkeit durch den Einsatz der ROI-Segmentierung beim UBFC-rPPG- und COHFACE-Datensatz zu verzeichnen ist, weist diese bei den restlichen fünf Datensätzen eine Verbesserung auf.

Gemessen an der Accuracy-Metrik hat sich für den ICA-, POS- und LGI-Algorithmus einheitlich eine Steigerung durch die ROI-Segmentierung auf den Stirn- und Wangenbereich bei einem Winkelschwellenwert von 90° ergeben. Eine Ausnahme bildet hierbei die Signalextraktion durch den POS-Algorithmus beim COHFACE-Datensatz, welche durch durch die ROI-Segmentierungen nicht profitiert. Hinsichtlich der übrigen Evaluationsmetriken konnte durch die ROI-Segmentierung ebenso eine Verbesserung erzielt werden. Für den ICA-, POS- und LGI-Algorithmus wird über sämtliche Datensätze hinweg durch die ROI-Segmentierung der MAE und RMSE im Vergleich zur Signalextraktion aus dem gesamten Gesichtsbereich verringert. Im allgemeinen Trend wird das SNR durch die Fokussierung der Signalextraktion auf den Stirn- und Wangenbereich gesteigert. Der Vergleich zwischen der ROI-Segmentierung mit einem Winkelschwellenwert von 90° und 45° legt nahe, dass bei Bewegungen ein größerer Gesichtsbereich vorteilhaft ist und die robuste Signalextraktion nicht begünstigt wird durch die Einschränkung des Reflexionswinkels.

Diese Ergebnisse legen nahe, dass die Begrenzung des gesamten sichtbaren Gesichtsbereichs auf anatomisch günstige Bereiche zur präziseren Extraktion der BVP-Signale führen kann. Wie bereits in Abschnitt 3.1.2 beschrieben, weist die Haut im Bereich der Stirn und der Wangen eine vergleichbar geringe Dicke auf, was mit einem ausgeprägten Gehalt an pulsatilen Informationen einhergeht. Zudem liegen diese Teile des Gesichtsbereichs in etwa auf einer räumlichen Ebene, die näherungsweise orthogonal zur optischen Achse der Kamera liegt. Dies führt dazu, dass eine erhöhte Anzahl an von den Blutgefäßen reflektierten Photonen zeitgleich auf den Bildsensor trifft. Infolgedessen zeigt dieser Gesichtsbereich eine ausgeprägte pulsinduzierte Helligkeitsänderung, welche aus den Blutvolumenschwankungen resultiert. Diese Erkenntnis deckt sich mit den vorangegangenen Forschungsarbeiten, die vergleichbare Resultate bei der Begrenzung des gesamten Gesichtsbereicht auf die Regionen der Stirn- und Wangen erzielten (siehe Kapitel 3.1.2). Die

zusätzliche Einschränkung des zulässigen maximalen Reflexionswinkels auf 45° hat in diesem Fall keinen weiteren Vorteil für die Vorhersagegenauigkeit oder zur Steigerung des SNR erbracht.

In einer weitergehenden Untersuchung wird der informationsreiche Gesichtsbereich der Stirn- und Wangen (Abbildung 25a) verglichen mit einem gleich großen Gesichtsbereich, der außerhalb der Stirn- und Wangen liegt (Abbildung 25b), sowie mit der Kombination aus beidem (Abbildung 25c). Das Ziel dieser Untersuchung ist der Vergleich des Informationsgehalts an pulsinduzierten Farbwertschwankungen zwischen der definierten Stirn- und Wangen-ROI und den übrigen Gesichtsbereichen.



(a) Stirn- & Wangen- ROI mit Reflexionswinkeln unter 90°



(b) Gesichtspixel mit den niedrigsten interpolierten Reflexionswinkeln, außerhalb der Stirn- und Wangen-ROI



(c) Kombination der ersten beiden ROIs

Abbildung 25.: Vergleich zwischen der Stirn- und Wangen-ROI mit einer gleich großen ROI, bestehend aus den niedrigsten interpolierten Reflexionswinkeln. Zusätzlich wird die Kombination aus beiden ROIs untersucht.

Hierfür wird nach der Berechnung der Reflexionswinkel jedes Dreiecks in der Gesichtsparkettierung eine Interpolation der Reflexionswinkel für jedes Gesichtspixel vorgenommen. Um die Vergleichbarkeit der Signalextraktion zu gewährleisten, wird der Gesichtsbereich außerhalb der Stirn- und Wangen-ROI mit der gleichen Anzahl an Bildpixeln wie die Stirn- und Wangen-ROI erzeugt (vergleiche Abbildung 18c). Die Kombination aus diesen beiden ROIs erfolgt durch boolesche ODER-Verknüpfung der Segmentierungsmasken und ihrer Anwendung auf den gesamten Gesichtsbereich.

Die Resultate dieser Untersuchung machen deutlich, dass die Signalextraktion aus der Stirn- und Wangen-ROI die besten Evaluationsmetriken liefert (Tabelle 10). Dies lässt sich unter anderem mit der stärkeren Hautdurchblutung dieser Regionen und der im Vergleich zu anderen Regionen geringeren Zahl an bedeckten Hautbereichen begründen. Die interpolierte ROI schneidet in diesem Vergleich am schlechtesten ab. Dies ist zum Teil darauf zurückzuführen, dass die extrahierte ROI ungünstige Gesichtsbereiche für die rPPG-Extraktion enthält, wie die Nase, Augen, Augenbrauen oder von Haaren und Bärten bedeckten Hautbereichen.

Die kombinierte ROI zeigt im Allgemeinen vergleichbare, wenn auch etwas schlechtere Ergebnisse wie die Stirn- und Wangen-ROI. Lediglich beim Szenario mit künstlicher Laborbeleuchtung im COHFACE-Datensatz zeigt die Kombination der Stirn- und Wangen-ROI mit der ROI, welche aus den niedrigsten interpolierten Reflexionswinkeln besteht, eine Verringerung des MAE und RMSE. Gleichzeitig sinkt allerdings auch das SNR durch die Kombination der ROIs, was sich auch in einer geringeren Vorhersagegenauigkeit, gemessen

Datensatz: UBFC-rPPG																		
ROI	Stirn-	und Wa	angen-RO	I		Interp	olierte	ROI				Kombinierte ROI						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc		
ICA	1,08	2,40	0,9909	1,96	80,42	5,43	13,51	0,7630	-0,04	71,60		3,04	11,00	0,8063	0,85	78,72		
POS	1,42	3,59	0,9753	3,61	78,57	2,08	6,21	0,9351	2,60	74,25		1,69	5,72	0,9433	2,65	77,95		
CHROM	1,89	4,43	0,9679	0,90	73,81	2,69	7,13	0,9173	0,68	71,31		1,74	5,73	0,9445	0,61	72,72		
GREEN	21,91	28,27	0,0668	-14,95	14,29	23,20	30,68	0,0882	-15,20	18,52		21,77	30,31	0,0851	-13,80	25,93		
LGI	1,40	3,89	0,9761	1,70	78,57	2,99	8,91	0,8754	0,68	76,48		1,67	5,99	0,9375	1,27	78,57		
PBV	4,46	11,92	0,7824	-1,16	71,43	8,66	18,94	0,5644	-3,05	59,26		7,03	17,12	0,6124	-0,64	69,14		

 Tabelle 10.:
 Vergleich zwischen den verschiedenen Ansätzen, den Gesichtsbereich zu segmentieren.

an der Accuracy-Metrik, ausdrückt. Beim POS- und LGI-Algorithmus nimmt die Vorhersagegenauigkeit durch die Kombination beider ROIs um über fünf Prozent zu, während sie bei den übrigen Signalverarbeitungsalgorithmen um über fünf Prozentpunkte abnimmt.

Beim VIPL-HR-V1-Datensatz kann durch die Kombination der Stirn- und Wangen-ROI mit der ROI aus den niedrigsten interpolierten Reflexionswinkeln beim POS-, CHROM- und LGI-Algorithmus eine geringfügige Reduzierung des MAE und RMSE festgestellt werden. Allerdings ergeben sich keine Verbesserungen bezüglich des SNR oder der Vorhersagegenauigkeit. Die tabellarisch aufgeführten Resultate für sämtliche untersuchte Datensätze befinden sich in Anhang A.4.

Diese Untersuchungen haben untermauert, dass sich die optimale Gesichtsregion zur Extraktion des BVP-Signals durch rPPG-Algorithmen im Bereich der Stirn und Wangen befindet. Zudem hat sich herausgestellt, dass die Nutzung eines größeren Gesichtsbereichs grundsätzlich nicht zu einer verbesserten Signalextraktion und einer höheren Genauigkeit in der Vorhersage der gemittelten Herzrate führt. Insbesondere hat sich gezeigt, dass die ROI, welche aus den niedrigsten interpolierten Reflexionswinkeln besteht, eine signifikante Verschlechterung der Evaluationsmetriken für nahezu jeden betrachteten Datensatz und rPPG-Algorithmus aufweist. Des Weiteren ist anzumerken, dass die Berechnung der interpolierten Reflexionswinkel mit einer zusätzlichen Rechenzeit von 100 ms pro Videobild einhergeht (siehe Abbildung 19), welche durch den mangelnden Vorteil bei der Signalextraktion nicht gerechtfertigt ist. Aus diesen Erkenntnissen wird für das weitere Vorgehen in dieser Arbeit der Fokus ausschließlich auf die Verwendung der Stirn- und Wangen-ROI gelegt.

## 6.4. Gewichtete Mittelung mehrerer extrahierten Herzraten

Die vorangegangenen Untersuchungen haben gezeigt, dass die Segmentierung des Gesichtsbereichs auf die Stirn- und Wangenregion im Allgemeinen die höchste Vorhersagegenauigkeit liefern. Diese Erkenntnis wird im Folgenden aufgegriffen. In dieser Studie wird die Auswirkung der Kombination von Gesichtsregionen auf die Vorhersagegenauigkeit der Herzrate untersucht. Hierfür werden vier verschiedene Methoden zur Kombination der Herzraten evaluiert, die aus den drei einzelnen Sub-ROIs aus der Stirn- und Wangen-ROI berechnet werden. Spezifisch wird untersucht, ob die separate Analyse der Sub-ROIs oder die Kombination aus den vorhergesagten Herzraten jeder Sub-ROI die Vorhersagegenauigkeit verbessern und insbesondere bei Bewegungen robuster gestalten können.

In der Datenvorverarbeitung werden die Regionen des Gesichts segmentiert, wobei zuerst die Kombination aus der Stirn-ROI und den ROIs der linken und rechten Wange extrahiert wird. Anschließend werden diese drei ROIs separat analysiert. Für jede ROI werden die Pixelgröße und die gemittelten Reflexionswinkel für jedes Videobild gespeichert. Anschließend wird die Herzrate für jede ROI separat geschätzt. Die Videos werden für diese Untersuchung in zehnsekündige Fenster unterteilt. Die Unterteilung der Videos in kürzere Zeitfenster dient dazu, vorübergehende Variationen in der Orientierung der ROIs präziser zu erfassen. Insbesondere ermöglichen kürzere Zeitfenster eine höhere zeitliche Auflösung und damit eine bessere Berücksichtigung der zeitlichen Dynamik und schnellen Veränderungen in der Orientierung der ROIs. Zur Berechnung der mittleren Herzrate jedes Videofensters werden die sechs unüberwachten rPPG-Algorithmen aus Kapitel 3.1 verwendet.

Im Rahmen dieser Untersuchung werden verschiedene Methoden zur Kombination der Herzraten aus den Stirn- und Wangen-ROIs erforscht. Jede dieser Methoden bietet eine andere Herangehensweise zur Verbesserung der Vorhersagequalität. Im Folgenden werden diese Methoden detailliert erläutert.

#### **Gewichtete Mittelung**

Die erste Methode zur Herzratenkombination ist die gewichtete Mittelung. Dieser Ansatz zielt darauf ab, die Herzraten aus den einzelnen ROIs anhand einer gewichteten Summe zu kombinieren. Hierbei erfolgt die Kombination der Herzraten unter Berücksichtigung der gemittelten Größe einer ROI und ihres gemittelten Reflexionswinkels. Die Mittelung des Reflexionswinkels erfolgt zunächst räumlich, indem der Reflexionswinkel über sämtliche Dreiecke in der ROI gemittelt wird, und zeitlich, indem der räumlich gemittelte Reflexionswinkel über sämtliche Videobilder im zehnsekündigen Zeitfenster gemittelt wird. Die Gewichtung erfolgt anhand dieser beiden Einflussgrößen auf die Signalextraktion aus der ROI, welche die beiden Gewichtungsfaktoren bestimmen: einen Größengewichtungsfaktor  $w_A$  und einen Winkelgewichtungsfaktor  $w_{\theta}$ . Die Größengewichtung erfolgt proportional zur Größe der ROI, die Winkelgewichtung erfolgt antiproportional zum Reflexionswinkel.

Zur Berechnung der Winkelgewichtungsfaktoren für die Menge von n = 3 Sub-ROIs werden zunächst für jede Sub-ROI k die räumlich gemittelten Reflexionswinkel über das zehnsekündige Zeitfenster, bestehend aus N Einzelbildern, gemittelt. Dies ergibt den räumlich und zeitlich gemittelten Reflexionswinkel  $\overline{\theta}_k$  für die k-te Sub-ROI:

$$\overline{\theta}_k = \frac{\sum_{i=1}^N \theta_{k,i}}{N} \tag{6.1}$$

Dabei steht  $\theta_{k,i}$  für den räumlich gemittelten Reflexionswinkel der *k*-ten Sub-ROI im *i*-ten Einzelbild. Aus dem räumlich und zeitlich gemittlten Reflexionswinkel  $\overline{\theta}_k$  wird der Winkelgewichtungsfaktor  $w_{\theta_k}$  durch den Kehrwert gebildet:

$$w_{\theta_k} = \frac{1}{\overline{\theta}_k} \tag{6.2}$$

Diese Winkelgewichtungsfaktoren werden normalisiert, sodass die Summe der Gewichte eins ergibt:

$$w_{norm,\theta_k} = \frac{w_k}{\sum_{k=1}^n w_k} \tag{6.3}$$

Die Größengewichtungsfaktoren werden analog berechnet, wobei keine Kehrwertbildung mehr durchgeführt wird. Zunächst werden für jede Sub-ROI k die Pixelflächen  $A_{k,i}$  in jedem Videobild i über das zehnsekündige Zeitfenster aus N Einzelbildern gemittelt. Dies ergibt die zeitlich gemittelte Pixelgröße  $A_k$  der k-ten Sub-ROI:

$$A_{k} = \frac{\sum_{i=1}^{N} A_{k,i}}{N}$$
(6.4)

Hierbei steht  $A_{k,i}$  für die Pixelfläche des *i*-ten Einzelbildes in der *k*-ten Sub-ROI. Die zeitlich gemittelten Pixelgrößen werden durch die Gesamtpixelfläche der drei Sub-ROIs normalisiert. Dabei erhält die größte Fläche das höchste Gewicht:

$$w_{norm,A_k} = \frac{A_k}{\sum_{k=1}^n A_k} \tag{6.5}$$

Die kombinierten Gewichtungsfaktoren  $w_k$  werden berechnet, indem die zuvor berechneten Winkel- und Größengewichte für jede Sub-ROI gemittelt werden:

$$w_k = \frac{w_{norm,A_k} + w_{norm,\theta_k}}{2} \tag{6.6}$$

Der gewichtete Mittelwert  $\overline{HR}$  der vorhergesagten Herzraten  $HR_k$  für n = 3 Sub-ROIs wird schließlich mithilfe des kombinierten Gewichtungsfaktors jeder Sub-ROI berechnet zu:

$$\overline{HR} = \sum_{k=1}^{n} w_k \cdot HR_k \tag{6.7}$$

Durch die Berücksichtigung von Größe und Reflexionswinkel wird beabsichtigt, die Genauigkeit der kombinierten Herzraten bei Bewegungseinflüssen auf die ROIs zu maximieren, indem innerhalb jedes Zeitfensters derjenigen ROI mit dem geringsten gemittelten Reflexionswinkel die höchste Winkelgewichtung zugeteilt wird. Zugleich soll die Variation in der Pixelgröße jeder ROI, welche insbesondere bei Kopfrotationen auftreten, in die Größengewichtung miteinbezogen werden. Einer flächenmäßig großen ROI mit niedrigem Reflexionswinkel wird demnach eine höhere Gewichtung zugeteilt, als einer kleinen ROI mit hohem Reflexionswinkel.
#### Winkelgewichtung

Die zweite Methode zur Herzratenkombination orientiert sich am Vorgehen der gewichteten Mittelung, wobei allerdings die Größen der ROIs nicht mehr von Relevanz sind und daher nicht berücksichtigt werden. Bei der Methode der *Winkelgewichtung* ist nur noch der gemittelte Reflexionswinkel ausschlaggebend für den Gewichtungsfaktor. Jede der ROIs besteht aus einer Vielzahl von Dreiecken aus der Gesichtsparkettierung, von denen jedes einen eigenen Reflexionswinkel zwischen seiner Oberflächennormale und der Kameraachse aufweist.

Der gewichtete Reflexionswinkel resultiert aus der räumlichen Mittelung aller Reflexionswinkel innerhalb einer ROI. Die Mittelung wird zunächst für jedes Videobild durchgeführt und anschließend zeitlich über das Zeitfenster von 10 s gemittelt (vergleiche Formel 6.1). Daraus folgt ein einziger räumlich und zeitlich gemittelter Reflexionswinkel  $\overline{\theta}_k$  für jede ROI k und jedes Videozeitfenster, der zur Berechnung der Winkelgewichtungsfaktoren  $w_{norm,\theta_k}$  dient (analog zu den Formeln 6.2 und 6.3) und mit dem letztlich die Herzratenkombination eines Zeitfensters  $\overline{HR}$  berechnet wird zu:

$$\overline{HR} = \sum_{k=1}^{n} w_{norm,\theta_k} \cdot HR_k$$
(6.8)

#### ROI mit dem niedrigsten gemittelten Reflexionswinkel

Bei der dritten Methode wird für jedes Zeitfenster nur die Herzrate einer einzigen ROI mit dem niedrigsten gemittelten Reflexionswinkel als resultierende Herzrate verwendet. Diese Methode baut auf der Forschung von Wong et al. [44] und der Annahme auf, dass ein geringerer Reflexionswinkel zu einer besseren Erfassung des Blutvolumenpulses und somit zu einer präziseren Herzratenbestimmung führen könnte.

### Mehrheitsentscheid

Die vierte Methode zur Herzratenkombination basiert auf einem Mehrheitsentscheid aus den drei Stirn- und Wangen-ROIs. Hierbei wird die resultierende Herzrate durch die Herzrate bestimmt, die von der Mehrheit der ROIs vorhergesagt wird. Falls zwei oder mehr ROIs übereinstimmen, ergibt diese Übereinstimmung die resultierende Herzrate. Sollten alle drei ROIs unterschiedliche Vorhersagen liefern, wird nach meiner Implementierung die Herzrate der Stirn verwendet. Diese Entscheidung stützt sich auf Beobachtungen aus zuvor untersuchten Szenarien. Diese zeigen, dass die Stirn-ROI konsistent eine höhere Vorhersagegenauigkeit bietet als die einzelnen Wangen-ROIs. Dies kann auf den geringeren Reflexionswinkel (siehe Abbildung 26) und die höhere Pixelfläche zurückzuführen sein (siehe Abbildung 27).

Die Vorhersagegenauigkeit der Herzraten wird für jede Kombinationsmethode ausgewertet und mit der separaten Analyse jeder ROI verglichen. Die Analyse wird exemplarisch anhand des UBFC-rPPG-Datensatzes vorgestellt (Tabelle 11). Die tabellarisch aufgeführten Resultate für sämtliche untersuchte Datensätze befinden sich in Anhang A.5.



Abbildung 26.: Vergleich der gemittelten Reflexionswinkel über alle Videos in den (Teil-)Datensätzen.

Die Ergebnisse zeigen, dass die Kombination der Herzraten aus den separaten Analysen der verschiedenen Gesichtsregionen bei keiner Kombinationsmethode zu einer Verbesserung der Vorhersagegenauigkeit führt und diese Vorgehensweise nicht den erwarteten Nutzen bietet. Die Kombination der Herzraten durch die gewichtete Mittelung kann zum Teil dazu beitragen, den MAE und RMSE zu senken, wie es den Ergebnissen vom COHFACE- und VIPL-HR-V1-Datensatz zu entnehmen ist. Diese Verbesserung ist dadurch zu erklären, dass große Abweichungen in den drei Vorhersagen durch die Mittelung gegenseitig kompensiert werden. Gleichzeitig werden aber auch korrekte Vorhersagen aus einer ROI durch die Mittelung mit ungenauen Vorhersagen verschlechtert. Dies zeigt sich darin, dass die Vorhersagenauigkeit bei jedem Datensatz durch die Mittelung abnimmt.

Die durchgeführte Untersuchung zur Kombination von Herzraten aus verschiedenen Sub-ROIs bietet einen Ansatz, um die Genauigkeit der Herzratenschätzung zu verbessern. Allerdings offenbart die Auswertung des vorgestellten Vorgehens verschiedene Aspekte, die eine eingehendere Reflektion erfordern. Die Methodik zur Segmentierung der ROIs basiert auf den vorangegangenen Untersuchungen, die im Rahmen dieser Arbeit auf Grundlage von früheren Forschungserkenntnissen durchgeführt wurden. Der Fokus auf die Kombination der Herzraten anhand der Größe und des Reflexionswinkels demonstriert einen adaptiven Ansatz, um Orientierungsvariationen der ROIs bei der Ermittlung der Herzrate zu berücksichtigen. Das Ergebnis dieser Untersuchung deutet darauf hin, dass diese Gewichtungsfaktoren allein nicht ausreichen, um die Genauigkeit der Herzraten vorhersage signifikant zu verbessern. Es gibt weitere Faktoren wie die Hautbeschaffenheit, Bewegungsartefakte oder Beleuchtungsvariationen, wie Reflexionen oder Helligkeitsunterschiede zwischen den ROIs, welche die Signalqualität beeinflussen können und daher für eine weiterentwickelte Herzratenkombination berücksichtigt werden sollten.

Eine denkbare alternative Vorgehensweise, um die Signale aus den ROIs miteinander zu kombinieren, stellt die Kombination der BVP-Signale dar, bevor eine Herzrate extrahiert wird. Vergleichbar zur Sensordatenfusi-

Tabelle 11.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Im Anschluss an die<br/>Datenvorverarbeitung wird jedes Video im UBFC-rPPG-Datensatz in ein kürzeres Zeitfenster von 10s<br/>unterteilt. Die Evaluationsmetriken beziehen sich auf sämtliche Videofenster, die aus dem Datensatz<br/>entstehen. Die separate Auswertung der Sub-ROIs oder deren Kombination hat keine Verbesserung in<br/>der Vorhersageleistung erzielt. Die besten Evaluationsmetriken liefert die Begrenzung der ROI auf die<br/>Stirn- und Wangenregion.

	Datensatz	z: UBFC-rl	PPG												
	Methode	Methode: ICA													
ROI	Stirn &	Stirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-							
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid							
MAE ↓	1,82	7,53	6,97	8,09	6,40	6,41	6,59	4,50							
$RMSE \downarrow$	6,03	18,01	16,50	16,57	12,46	13,76	16,56	11,62							
Pearson ↑	0,9371	0,5403	0,6329	0,7177	0,7692	0,7103	0,6011	0,7868							
SNR (dB) ↑	5,59	0,39	-0,40	-0,92	2,38	2,20	1,38	2,24							
Acc (%) ↑	86,42	69,14	67,90	64,73	50,77	51,54	64,62	65,38							

on könnten die extrahierten BVP-Signale miteinander fusioniert werden, bevor die FFT auf das resultierende Signal angewandt und die Herzrate bestimmt wird. Diese Vorgehensweise konnte durch Einschränkungen in der Architektur der rPPG-Toolbox nicht im Rahmen dieser Arbeit implementiert und untersucht werden, da dort jede ROI einzeln und nacheinander verarbeitet wird, und deshalb die BVP-Signale aus den einzelnen Durchläufen nicht zum gleichen Zeitpunkt zur Datenverarbeitung vorliegen. Die Fusion der Signale mehrerer ROIs bleibt somit Gegenstand zukünftiger Forschungen.

Darüber hinaus ist die Wahl der optimalen Zeitfenstergröße bei der implementierten Methode kritisch. Die Herzrate ist ein zeitlich veränderlicher physiologischer Parameter, der sich über einen bestimmten Zeitraum stabilisieren muss, um eine zuverlässige Schätzung zu gewährleisten. Durch die Verwendung zu kurzer Zeitfenster könnte die erforderliche Dauer des BVP-Signals für eine präzise Herzratenschätzung möglicherweise nicht ausreichend sein. Kürzere Zeitfenster können dazu beitragen, schnelle Veränderungen in den Orientierungen der ROIs genauer zu erfassen, da die Mittelung ihrer zeitveränderlichen Reflexionswinkel und Pixelflächen über kürzere Zeiträume eine geringere Abweichung des Mittelwerts zum tatsächlichen Wert während jedes Zeitpunktes ermöglicht. Insbesondere bei Bewegungen oder schnellen Gesten sind kurze Zeitfenster hilfreich, um die zeitliche Auflösung zu erhöhen und so präzisere Daten über die Bewegungsdynamik zu liefern. Somit stellt die Wahl des Zeitfensters einen Kompromiss zwischen der Erfassung von Orientierungsvariationen und der Genauigkeit der Herzratenschätzung dar.

Es bedarf einer sorgfältigen Abwägung, um die optimale Zeitspanne zu bestimmen, die einerseits die Bewegungsdynamik der ROIs erfasst und andererseits einen hinreichend langen Signalverlauf für eine zuverlässige Herzratenschätzung liefert. Im folgenden Abschnitt wird untersucht, inwiefern sich die Wahl der Zeitfensterlänge für die Unterteilung der Videos auf die Vorhersagegenauigkeit der Herzrate auswirkt.

## 6.5. Untersuchung der optimalen Zeitfenstergröße

Diese Untersuchung zielt darauf ab, die Auswirkung der Videodauer in Form von verschiedenen Zeitfensterlängen auf die Vorhersagegenauigkeit der sechs unüberwachten rPPG-Algorithmen zu analysieren. Die



Abbildung 27.: Vergleich der gemittelten ROI-Pixelflächen über alle Videos in den (Teil-)Datensätzen.

Wahl der Zeitfensterlänge bei der BVP-Signalverarbeitung spielt eine entscheidende Rolle in Bezug auf die Genauigkeit der Herzfrequenzvorhersage. Die Motivation hinter dieser Untersuchung liegt darin, das optimale Zeitfenster für die Herzratenschätzung zu ermitteln, welches sowohl eine angemessene zeitliche Auflösung für präzise Vorhersagen gewährleistet als auch ausreichend Datenpunkte für eine zuverlässige Signalverarbeitung bereitstellt.

Vor dem Hintergrund des bevorstehenden Einsatzszenarios in der remote-Photoplethysmographie werden zur Untersuchung der optimalen Zeitfensterlänge annähernd gleichbleibende Aufnahmebedingungen während des untersuchten Zeitfensters angestrebt. Dadurch soll die Abweichung der über das Zeitfenster gemittelten vorhergesagten Herzrate vom tatsächlichen Mittelwert der Herzrate klein gehalten werden. Weiterhin soll die Erkennung von Belastungsveränderungen im Verlauf der Videoaufzeichnung durch kurze Zeitfenster motiviert werden. Im Optimalfall ist die tatsächliche Herzrate während des Zeitfensters konstant, sodass durch die Mittelung keine Informationen über die Herzratenvariabilität verloren geht.

Die Algorithmen zur Vorhersage der Herzrate erfolgen in dieser Untersuchung für kleinere Zeitfenster mit einer vordefinierten Dauer von 10, 15, 30 oder 60 Sekunden. Das Ziel der kürzeren Zeitfenster ist es, die Mittelung der momentanen Herzrate über einen möglichst kurzen Zeitraum vorzunehmen, sodass die zeitliche Auflösung der vorhergesagten Herzrate steigen und zugleich der Quantisierungsfehler durch die höhere zeitliche Auflösung reduziert werden soll. Für die Untersuchung werden die Bewegungsszenarien beziehungsweise die Szenarien nach körperlicher Betätigung der Probanden aus dem RLAP-rPPG-, MMPD- und VIPL-HR-V1-Datensatz den jeweiligen stationären Szenarien mit ruhenden Probanden gegenübergestellt.

Die Extraktion des BVP-Signals erfolgt durch die sechs unüberwachten rPPG-Algorithmen ICA, POS, CHROM, GREEN, LGI und PBV, welche das BVP-Signal des gesamten Videos extrahieren. Das extrahierte BVP-Signal und das Referenz-BVP-Signal werden einer Datennachbearbeitung gemäß der Schilderung in Abschnitt 5.7 unterzogen, die aus der Langzeittrendbereinigung und einer Butterworth-Filterung besteht. Anschließend

wird das nachbearbeitete BVP-Signal und das Referenz-BVP-Signal basierend auf dem gewählten Zeitfenster in kürzere Signalsequenzen unterteilt. Hieraus wird die momentane Herzrate der Signalsequenz bestimmt, indem die Peak-Frequenz im FFT-Frequenzspektrum ermittelt wird. Die Anhängigkeit der Accuracy-Metrik von der Zeitfensterlänge ist in Tabelle 12 dargestellt.

				Date	nsatz				
		RL	AP-rPPG	M	MPD	VIPL-HR-V1			
rPPG-	Zeitfenster	Ruheszenario	beim Videospielen	Buheszenario	nach Bewegung	Ruheszenario	nach Bewegung		
Methode	(s)	Tuneszenario	benn videospielen	Tuneszenario	nach bewegung	Runeszenario	naen bewegung		
	10	55,56	64,51	34,60	21,34	31,84	22,70		
	15	55,48	71,12	38,45	19,70	31,36	17,11		
ICA	30	66,67	84,59	46,59	27,27	56,14	31,58		
	60	66,54	73,28	45,45	26,52	57,01	29,25		
	-	52,63	65,52	45,45	26,52	57,01	29,25		
	10	62,43	75,00	44,95	37,25	37,15	30,17		
	15	61,84	79,74	52,46	37,31	34,75	24,12		
POS	30	75,00	91,22	63,26	40,91	52,63	35,96		
	60	73,68	85,34	62,42	40,91	58,88	31,13		
	-	64,91	79,31	62,12	40,91	58,88	31,13		
	10	64,91	72,13	35,48	28,79	28,33	20,23		
	15	64,69	79,96	40,91	27,27	31,20	16,67		
CHROM	30	78,07	90,09	52,65	32,58	47,37	28,07		
	60	71,93	83,62	56,06	28,79	51,40	22,64		
	-	70,18	75,86	56,06	28,79	51,40	22,64		
	10	19,30	18,25	11,24	7,45	15,08	11,49		
	15	20,83	21,55	13,45	7,95	17,80	10,09		
GREEN	30	32,46	34,05	16,29	9,47	29,82	18,52		
	60	34,21	32,76	18,94	15,15	33,64	16,98		
	-	24,56	41,38	18,94	15,15	33,64	16,98		
	10	65,50	71,84	43,31	34,85	30,17	22,99		
	15	64,69	79,31	50,38	33,90	33,05	20,17		
LGI	30	78,07	86,40	61,74	38,64	50,88	28,95		
	60	72,81	81,90	59,85	37,88	52,34	28,30		
	-	68,42	74,14	59,85	37,88	53,27	28,30		
	10	52,05	56,18	25,88	20,20	28,21	16,09		
	15	49,12	64,87	30,11	18,75	26,69	16,67		
PBV	30	60,96	73,41	38,26	23,86	43,86	21,93		
	60	58,77	68,97	34,85	22,73	49,53	18,87		
	-	50,88	67,24	34,85	22,73	48,60	18,87		

 Tabelle 12.:
 Auswirkung unterschiedlich langer Zeitfenster auf die Accuracy-Metrik.

Die Resultate zeigen, dass die Anzahl der Abweichungen unterhalb von 2 BPM zwischen der vorhergesagten Herzrate und der tatsächlichen Herzrate bei einer Zeitfensterlänge von 30 Sekunden am geringsten ausfallen, was sich in der höchsten Accuracy-Metrik widerspiegelt. Lediglich beim GREEN-Algorithmus weist eine längere Signaldauer einen Vorteil für die Herzratenprognose auf, da in diesem Fall weniger fehlerhafte Vorhersagen auftreten. Beim Videospielszenario des RLAP-rPPG-Datensatzes konnte durchweg eine höhere Genauigkeit erzielt werden, als beim Ruheszenario. Eine denkbare Erklärung für diesen Umstand liegt in einer potenziell höheren Durchblutung der Gesichtshaut während des Videospielszenarios, bei dem die Probanden eine geistig anspruchsvolle Aktivität ausüben. Im Vergleich der gemittelten Herzraten in beiden Versuchsszenarien konnten grundsätzlich höhere Herzraten beim Videospielszenario festgestellt werden. Die verbesserte Vorhersagegenauigkeit lässt auch vermuten, dass die Probanden beim Videospielszenario womöglich eine ruhigere Körper- beziehungsweise Kopfhaltung einnehmen, als beim Ruheszenario. Dies konnte durch stichprobenartige Begutachtung einzelner Videos aus dem Videospielszenario ausgeschlossen werden, welche mehr rotatorische Kopfbewegungen aufweisen als die Videos des gleichen Probanden beim Ruheszenario. Mit Ausnahme des GREEN-Algorithmus wird bei beiden gegenübergestellten Szenarien des RLAP-rPPG-Datensatzes die höchste Vorhersagegenauigkeit durch Anwendung von 30-sekündigen Zeitfenstern erzielt.

Beim Vergleich der beiden untersuchten Szenarien innerhalb der VIPL-HR-V1- und MMPD-Datensätze lässt sich feststellen, dass die Vorhersagegenauigkeit der rPPG-Algorithmen bei ruhenden Probanden wesentlich höher ausfällt, als in den Szenarien nach körperlicher Betätigung, bei denen die Herzrate während der Videoaufzeichnung zum Ruhepuls hin abfällt. Mithilfe des POS-Algorithmus und einer Zeitfensterlänge von 30 Sekunden konnte die Herzrate beim Ruheszenario des MMPD-Datensatzes in 63, 26 % der Fälle innerhalb der zulässigen Abweichung von 2 BPM zur Referenzherzrate vorhergesagt werden. Beim Bewegungsszenario erzielt der POS-Algorithmus als bester unter den sechs verglichenen Algorithmen eine Genauigkeit von 40, 91 %. Beim Ruheszenario des VIPL-HR-V1-Datensatz konnte bei einer Zeitfensterlänge von 30 Sekunden die höchste Vorhersagegenauigkeit von 56, 14 % durch den ICA-Algorithmus erzielt werden, im Vergleich zur höchsten Genauigkeit von 34, 95 %, die durch den POS-Algorithmus im Szenario nach körperlicher Betätigung erreicht wird. Die Resultate des VIPL-HR-V1-Datensatzes zeigen auf, dass insbesondere beim Bewegungsszenario die Unterteilung der BVP-Signale in kürzere Zeitfenster von Vorteil für die Herzratenprognose ist, während beim Ruheszenario eine Mittelung über einen Zeitraum von 60 Sekunden weniger Abweichungen zwischen den prognostizierten und tatsächlichen Herzraten verursacht. Dies geht allerdings einher mit einer geringeren zeitlichen Auflösung für die ermittelten Herzraten.

Basierend auf den erlangten Erkenntnissen lässt sich schlussfolgern, dass ein Zeitfenster von 30 Sekunden für die rPPG-Analyse von BVP-Signalen den besten Kompromiss zwischen zeitlicher Auflösung und Vorhersagegenauigkeit bietet. Hinsichtlich einer noch höheren zeitlichen Auflösung lässt sich die Unterteilung in kürzere Zeitfenster verbessern, indem das BVP-Signal des gesamten Videos nicht mehr in aufeinanderfolgende, sich nicht überschneidende Zeitfenster unterteilt wird, sondern in sich überlappende Zeitfenster von gleicher Länge. Die Moving-Window-Methode ist eine Technik, die in der Analyse von Zeitreihendaten verwendet wird, um Trends zu identifizieren oder Vorhersagen zu erstellen. Das Verfahren verwendet ein gleitendes Fenster (engl.: *moving window*), das sich über die Zeitreihe bewegt und jeweils eine bestimmte Anzahl von Datenpunkten umfasst. Dabei wird die nachfolgende Herzratenschätzung auf Basis dieser Signalsequenz durchgeführt. Dieser Prozess wird iterativ wiederholt, indem das Fenster weiterbewegt wird, bis das Ende der Zeitreihe erreicht ist. Mit dieser Vorkehrung bietet die Datenverarbeitungspipeline das Potenzial für eine Echtzeit-Anwendung, die zur kontinuierlichen Überwachung der Herzrate verwendet werden kann. Die Implementierung und Anwendung der Moving-Window-Methode bleibt Gegenstand zukünftiger Forschungsarbeiten, in denen der Einfluss der Fenstergröße oder der Anteil von sich überlappenden Zeitfenstern eingehender untersucht und optimiert werden kann.

### 6.6. Diskussion und Fazit

In diesem Kapitel wurden Untersuchungen zur Ermittlung des optimalen Gesichtsbereichs für die rPPG beschrieben, welche es durch die Definition unterschiedlicher Gesichtsbereiche ermöglichen, den Einfluss von verschiedenen Gesichtsorientierungen, sowie Licht- und Bewegungsartefakten auf die Signalqualität zu analysieren. Mit Aufnahmen aus sechs für die Forschung bereitgestellten Datensätzen dienen die Untersuchungsergebnisse als Grundlage des weiteren Systementwurfs. In diesem Zusammenhang wurde das Gesichtserkennungsframework in der *MediaPipe*-Bibliothek [56] eingesetzt, um biometrische Gesichtspunkte zu erhalten darauf aufbauend die Untersuchungen innerhalb einer modifizierten Version der rPPG-Toolbox [39] durchzuführen. Auf dieser Grundlage wurden mehrere Verfahren zur Auswahl des Gesichtsbereichs vorgestellt, darunter die feste Definition des maximalen Reflexionswinkels der Hautoberfläche, einer histogrammbasierten Hautsegmentierung, die feste Definition der ROI im Bereich der Stirn- und Wangen, sowie die Kombination mehrerer Herzraten durch separate Analyse der einzelnen Sub-ROIs. Die Vorteile der ROI-Definition konnten durch Auswertung des SNR und der Accuracy-Metrik für verschiedene Szenarien innerhalb der Datensätze demonstriert werden. Speziell für den Einsatz des rPPG-Verfahrens in bewegten Szenarien mit wechselnden Gesichtsorientierungen ist die Kombination aus präziser Gesichtserkennung und adaptiver Verfolgung der ROIs eine wichtige Voraussetzung für die erfolgreiche Extraktion des BVP-Signals.

Mit Blick auf die unterschiedlichen Messregionen im Gesicht zeigt die Auswertung der Resultate, dass die Stirn- und Wangenbereiche insbesondere für den Einsatz des rPPG-Systems in bewegten Szenarien eine Extraktion des BVP-Signals mit den geringsten Abweichungen zur tatsächlichen Herzrate ermöglichen. Die adaptive ROI besteht demnach aus der vordefinierten Stirn- und Wangen-ROI, ohne weiterer Limitierung des Winkelschwellenwerts, dem Einsatz der interpolierten ROI oder der Kombination mehrerer Herzraten. Die Adaption zeichnet sich letztlich dadurch aus, dass sich die ROI kontinuierlich für jedes Videobild aktualisiert und dadurch in der Lage ist, Änderungen in der Gesichtsorientierung folgen zu können. Damit eignet sich dieser Ansatz zur Gesichtssegmentierung zur Extraktion von BVP-Signalen aus sich bewegenden Personen.

Die Ergebnisse der durchgeführten Untersuchungen liefern letztlich die Definition des Gesichtsbereichs, der im Rahmen dieser Arbeit erforscht wurde, um damit die Aufmerksamkeitsmaske eines neuronalen Netzes bei Bewegungen zu aktualisieren. Die erlangten Erkenntnisse werden in den nächsten Schritten verwendet, um die adaptive ROI für die Aktualisierung der Aufmerksamkeitsmaske des *TS-CAN*-Modells einzusetzen.

# 7. Entwicklung der rPPG auf Basis neuronaler Netze

In diesem Kapitel wird beschrieben, wie die Untersuchungserkenntnisse aus dem vorangegangenen Kapitel verwendet werden, um die ROI zu definieren, die zur Aktualisierung der Aufmerksamkeitsmaske des *TS-CAN*-Modells eingesetzt werden soll. Aufbauend auf der Modellstruktur des *TS-CAN*-Modells aus Abbildung 8 wird die in Kapitel 5 beschriebene Datenvorverarbeitung zur Segmentierung der Stirn- und Wangen-ROI durchgeführt, um die Gesichtsvideos für die Eingabe in den Aufmerksamkeitszweig des *TS-CAN*-Modells vorzubereiten. Dabei werden die in dieser Arbeit implementierten Datenvorverarbeitungsschritte der vorimplementierten Gesichtsextraktion aus der rPPG-Toolbox gegenübergestellt und verglichen. Hierfür wird zunächst die zur Verarbeitung durch das neuronale Netz notwendige Datenverarbeitung und daraus resultierende Struktur der Eingangsgrößen für die verwendete Netzwerkarchitektur erklärt. Anschließend wird der Trainingsprozess geschildert und die Evaluation der trainierten Modelle anhand der bewegungsreichen Szenarien innerhalb der Datensätze aus Kapitel 4 durchgeführt.

### 7.1. Einführung in den Entwicklungsprozess neuronaler Netze

Die sechs zuvor eingesetzten rPPG-Algorithmen gehören zur Klasse der unüberwachten Algorithmen. Unüberwachte Algorithmen zeichnen sich dadurch aus, dass die Eingangsdaten für den Algorithmus nur aus den zu verarbeitenden Daten bestehen, wie im vorliegenden Fall der rPPG-Extraktion die Sequenzen aus Videobildern. Dem Algorithmus ist während der Datenverarbeitung das zugehörige BVP-Signal unbekannt und die Extraktion der Ausgangsdaten erfolgt selbstständig durch den implementierten Algorithmus. Ein wesentlicher Unterschied hierzu stellt die Klasse der überwachten Algorithmen dar. Hierbei werden dem Algorithmus neben den Sequenzen aus Videobildern noch die zugehörigen BVP-Signale als Referenz für die Vorhersage bereitgestellt. Der Algorithmus wird dann dazu trainiert, selbstständig Muster innerhalb der Eingangsdaten zu identifizieren und daraus eine Abbildung der Eingangsdaten auf die erwarteten Referenzausgangsdaten zu erstellen. Neuronale Netze gehören zu den überwachten maschinellen Lernalgorithmen, zu denen auch das im Rahmen dieser Arbeit eingesetzten *TS-CAN*-Modell gehört. Der nachfolgende Abschnitt widmet sich der Architektur des *TS-CAN*-Modells.

### 7.1.1. Einführung in die Architektur neuronaler Netze

Die Architektur des *TS-CAN*-Modells besteht im Wesentlichen aus zwei Zweigen mit zweidimensionalen Convolutional Neural Networks (CNN). CNNs sind eine Klasse von neuronalen Netzen, die sich in der Verarbeitung von zweidimensionalen Daten wie Bildern bewährt haben. CNNs haben sich insbesondere bei der Erkennung von Mustern in visuellen Daten als effektiv erwiesen und werden beispielsweise in der Bildklassifizierung oder zur Erkennung von Objekten oder Gesichtern eingesetzt. Ein grundlegender Baustein eines CNNs ist das Convolutional-Layer (Faltungsschicht). Diese Schicht führt Faltungsoperationen auf den Eingabedaten mithilfe sogenannter Filterkernen durch, wodurch lokale Muster und Merkmale innerhalb der Daten erkannt werden. Durch die Verwendung von individuell auf die Problemstellung antrainierten Filterkernen kann das Netzwerk dazu verwendet werden, relevante Merkmale zu extrahieren, die spezifisch für die Charakteristika von Blutvolumenpuls-Signalen sind. Diese charakteristischen Merkmale können im Fall der rPPG-Analyse subtile Veränderungen in der Intensität der Pixel im zeitlichen Verlauf sein, welche auf den pulsbedingte Farbwertvariationen hinweisen. Durch das Training des Netzwerks mit den zweidimensionalen Eingabedaten und den zugehörigen gelabelten Ausgabedaten werden die Gewichtungen der Filter so angepasst, dass das Netzwerk in der Lage ist, diese spezifischen Muster innerhalb der Eingangsdaten zu identifizieren, welche letztlich zur Vorhersage der Ausgangsdaten führen. Die Gewichte der Filterkerne sind demnach ein zentrale Trainingsvariable.

Ein weitere Komponente von CNNs ist das Pooling-Layer, das darauf abzielt, die räumliche Dimension der verarbeiteten Daten zu reduzieren und die Anzahl an zu trainierenden Parameter im Netzwerk zu verringern. Dadurch soll die Effizienz des Netzwerks hinsichtlich der benötigten Rechenleistung erhöht und zugleich die Anfälligkeit gegenüber Überanpassungen an die Trainingsdaten verringert werden. CNNs bestehen typischerweise aus mehreren Aneinanderreihungen von Convolutional- und Pooling-Layer. Im Fall des TS-CAN-Modells besteht jeder Zweig aus vier Convolutional-Layer, mit jeweils einem Average-Pooling- und Dropout-Layer nach dem zweiten und vierten Convolutional Layer [35]. Das Dropout-Layer ist eine weitere Regularisierungstechnik zur Vermeidung von Überanpassungen an die Trainingsdaten. Dies wird erreicht, indem während jedes Trainingsdurchlaufs (Epoche) eine zuvor fest definierte Anzahl an zufällig ausgewählten Neuronen in der vorherigen Schicht deaktiviert wird und deren Ausgang für die nachfolgenden Berechnungen gleich Null gesetzt werden. Dies bewirkt, dass die Merkmalsextraktion der deaktivierten Neuronen von den übrig gebliebenen Neuronen mitgelernt werden muss. Die Definition des Prozentsatzes an auszuschaltenden Neuronen wird innerhalb der Netzwerkarchitektur festgelegt und beträgt beim TS-CAN-Modell 25% beim ersten Dropout-Layer und 50% beim zweiten Dropout-Layer. Durch das zufällige Ausschalten von Neuronen soll verhindert werden, dass die Merkmalsextraktion des Modells von einzelnen Neuronen abhängig wird und möglicherweise Muster extrahiert werden, welche nur in den Trainingsdaten vorhanden sind. Durch den Einsatz von Dropout-Layern kann das Modell also robuster gegenüber unbekannten Variationen in den Eingangsdaten gestaltet und ein breiteres Spektrum an spezifischen Mustern extrahiert werden. Während der an das Training anschließenden Testphase wird das Dropout-Layer deaktiviert und alle Neuronen aktiviert. Dies soll eine konsistente Vorhersage gewährleisten, da das Modell während des Testens nicht mehr von zufälligen Variationen der aktiven Modellstruktur beeinflusst wird.

### 7.1.2. Trainingsprozess

Zu Beginn des Trainings eines neuronalen Netzes ist es bewährte Praxis, den vorliegenden Datensatz in Teildatensätze aufzuteilen. Dabei wird aus einem vollständigen Datensatz ein Trainings- und ein Validierungsdatensatz erstellt. Der Trainingsdatensatz wird dazu verwendet, die neuronalen Gewichte innerhalb der Modellstruktur für die erfolgreiche Vorhersage der erwarteten Ausgangsdaten anzupassen. Eine in der Praxis gängige und vielfach bewährte Aufteilung des Datensatzes in die beiden Teildatensätze erfolgt in einem Verhältnis von 80:20. Konkret werden aus dem vollständigen Datensatz 80 % der Daten in den Trainingsdatensatz übernommen und die restlichen 20 % in den Validierungsdatensatz. Die Aufteilung des Datensatzes in den Trainings- und Validierungsdatensatz erfolgt dabei so, dass ein Datenleck (*engl.* data leakage) zwischen beiden Teildatensätzen ausgeschlossen wird. Dadurch soll sichergestellt werden, dass keine Gesichtsvideos desselben Probanden in beiden Teildatensätzen enthalten sind. So soll verhindert werden, dass das Modell bereits während des Trainings Zugang zu Daten erhält, welche für die anschließende Bewertung der Generalisierungsfähigkeit des Modells verwendet werden. Der Validierungsdatensatz ermöglicht es, die Leistung des trainierten Modells auf einem unbekannten Datensatz bereits während des Trainings zu bewerten. So kann evaluiert werden, ob das Modell zur Überanpassung an die Trainingsdaten neigt und sich zu sehr auf die in den Trainingsdaten vorliegenden Muster fokussiert, welche möglicherweise nicht repräsentativ für die allgemeine Grundgesamtheit der zu erwartenden Eingangsdaten sind. Im vorliegenden Fall der rPPG-Analyse bestehen die Eingangsdaten aus Bildsequenzen von Gesichtern und den Referenz-BVP-Signalen und die Ausgabedaten repräsentieren das aus den Bildsequenzen extrahierte BVP-Signal. Eine Anforderung an die Trainingsdaten ist es, möglichst viel Variationen in den Daten abzubilden, welche auch in der Realität beziehungsweise dem Einsatzszenario des trainierten Modells vorkommen können.

### 7.1.3. Optimierungsprozess

Die Anpassung der neuronalen Gewichte an die spezifischen Anforderungen stellt mathematisch betrachtet ein Optimierungsproblem dar. Zwei Metriken, um die Leistung des Modells während des Optimierungsprozesses beim Training zu bewerten, sind der Trainingsverlust (*engl.*: training loss) und der Validierungsverlust (*engl.*: validation loss). Die Verlustfunktion misst für jeden Datenpunkt *i* aus der Gesamtheit aller Datenpunkte innerhalb der BVP-Signale *N* den mittleren quadratischen Fehler [*engl.*: Mean Squared Error (MSE)] zwischen dem vorhergesagten BVP-Signal  $\hat{Y}$  und dem Referenz-BVP-Signal *Y*:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2$$
(7.1)

Während des Trainingsprozesses wird das Modell darauf optimiert, den MSE zu minimieren. Das bedeutet, dass die neuronalen Gewichte dahingehend angepasst werden, dass die Vorhersagen des Modells den tatsächlichen Werten nahekommen. Der MSE ist demnach ein Indikator, um eine Überanpassung oder Unteranpassung an die Trainingsdaten zu überwachen. Ein zu niedriger MSE auf dem Validierungsdatensatz kann auf eine Überanpassung hinweisen, während ein hoher MSE auf dem Validierungsdatensatz auf eine Unteranpassung hinweisen kann. Im Fall einer Unteranpassung ist die Modellstruktur möglicherweise nicht dazu geeignet, die relevanten Merkmale aus den Eingangsdaten zu extrahieren. Im Fall einer Überanpassung lernt das Modell die im Trainingsdatensatz enthaltenen spezifischen Muster, ohne dabei noch auf unbekannte Daten generalisierungsfähig zu sein und über den Umfang der gelernten Trainingsdaten hinaus noch zuverlässige Vorhersagen treffen zu können. Grund für Überanpassungen können zu wenige Trainingsdaten sein, oder ein zu langes Training über zu viele Epochen. Die Epochen geben im Kontext maschineller Lernprozesse die Anzahl der vollständigen Durchläufe durch den Trainingsdatensatz während des Trainings an. Damit ein Modell effektiv und robust lernt, ist es entscheidend, dass es beim Training mehrere Epochen durchläuft. Eine einzige Epoche könnte nicht ausreichen, um alle relevanten Muster in den Daten zu erfassen und das Modell ausreichend für seinen Einsatzzweck zu trainieren. Mehrere Epochen ermöglichen es dem Modell, verschiedene Aspekte der in den Daten liegenden Muster zu identifizieren. Daher ist es in der Entwicklungspraxis bedeutend, die Vorhersageleistung während des Trainingsprozesses zu kontrollieren und das Training gegebenfalls vorzeitig zu beenden, wenn die Leistung auf dem separaten Validierungsdatensatz nicht weiter verbessert wird. Dieses Vorgehen wird als Early-Stopping bezeichnet

(Abbildung 28).



Abbildung 28.: Trainings- und Validierungsverlust über eine Trainingslaufzeit von mehreren Epochen (entnommen und übersetzt aus [62]). Das Training kann vorzeitig beendet werden, sobald der Validierungsverlust nicht weiter abnimmt, obwohl der Trainingsverlust sinkt. Alternativ kann man das Modell nach jeder Epoche speichern und das Modell mit dem niedrigsten Validierungsverlust für das vorgesehene Szenario einsetzen.

Nachdem das Modell mithilfe der Trainings- und Validierungsdaten getestet und evaluiert wurde, erfolgt ein weitere Evaluation seiner Vorhersageleistung auf gänzlich unbekannten Testdaten. Hierfür dient ein Testdatensatz zur Simulation von echten Anwendungsszenarien. Mithilfe des Testdatensatzes lässt sich überprüfen, ob das Modell nur während des Trainings eine zufriedenstellende Vorhersageleistung erreicht oder ob dieses auch noch bei neuen und unbekannten Daten generalisierungsfähig ist.

Die Konvergenz des Modells auf ein globales Minimum seiner Verlustfunktion wird durch den Einsatz vom *AdamW*-Optimierer [63] und einer Anpassung der Lernrate durch die *One-Cycle-Strategie* [64] angestrebt. Bei der *One-Cycle-Strategie* wird die Lernrate von einer anfänglichen Lernrate auf eine maximale Lernrate von 0.009 gesteigert und im späteren Trainingsverlauf auf eine minimale Lernrate reduziert [65] (siehe Abbildung 29). Bei der *One-Cycle-Strategie* wird die Lernrate nach jedem Batch, also nach jeder Durchlaufsiteration, geändert. Die Batch-Size ist ein Hyperparameter für den Trainingsprozess, der die Anzahl der zu bearbeitenden Trainingsbeispiele angibt, bevor die internen Parameter des Modells aktualisiert werden. Für eine detaillierte Beschreibung der Trainingsstrategien sei auf die entsprechende Fachliteratur und die Veröffentlichung zum *TS-CAN*-Modell [35], sowie zur rPPG-Toolbox [39] verwiesen.

Um die Vergleichbarkeit mit dem bestehenden *TS-CAN*-Modell zu gewährleisten, wird die Netzwerkarchitektur im Rahmen dieser Arbeit nicht weiter auf Optimierungspotential untersucht, sondern als Vergleichsmaßstab für die selbst implementierte Datenvorverarbeitung festgelegt. Im Folgenden wird die Struktur der Eingangsdaten beschrieben, anhand derer das Modell die Merkmalsextraktion trainiert, um aus Bildsequenzen des Gesichts ein BVP-Signal rekonstruieren zu können.



Abbildung 29.: Anpassung der Lernrate während des Trainings.

### 7.2. Struktur der Eingangsdaten

Die Eingänge des Modells sind gesichtssegmentierte, zeitkontinuierliche Bildsequenzen, sowie die zugehörigen Referenz-BVP-Signale. Aus den eingehenden Bildsequenzen werden die Signalverläufe des Blutvolumenpulses anhand von markanten Farbpixelunterschieden innerhalb der RGB-Kanäle aufeinanderfolgender Bilder extrahiert. Die Gesichtsvideos und Referenz-BVP-Signale werden im Rahmen der Datenvorverarbeitung in kürzere Sequenzen mit einer Länge von 180 Bildern beziehungsweise Datenpunkten unterteilt. Zudem werden die zuvor segmentierten Bildausschnitte, die das Gesicht beziehungsweise die ROI enthalten, auf eine Auflösung von 72x72 Pixel heruntergesampelt. Zusammen mit den drei Farbkanälen für die roten, grünen und blauen Pixelfarbwerte ergeben die Bildsequenzen einen Tensor mit der Dimension (180, 72, 72, 3).

Die Struktur der Eingangsdaten für das neuronale Netz ist zweigeteilt und beinhaltet sowohl die Eingabedaten für den Bewegungszweig als auch für den Aufmerksamkeitszweig, wobei für beide Zweige unterschiedliche Vorverarbeitungsschritte durchgeführt werden. Um die zeitlichen Veränderungen im Bewegungsbereich zu erfassen, wird die Differenz zwischen aufeinanderfolgenden Videobildern berechnet. Diese Differenzen werden anschließend normalisiert, indem sie durch die Standardabweichung aller Pixelfarbwerte innerhalb des Eingangstensors geteilt werden.

Für den Aufmerksamkeitszweig werden die Videosegmente z-score standardisiert. Dies bedeutet, dass der Mittelwert aller Pixelfarbwerte im gesamten Tensor von jedem Pixelfarbwert subtrahiert wird. Dadurch wird sichergestellt, dass der Mittelwert der Pixelwerte im gesamten Tensor auf einen Erwartungswert von  $\mu = 0$  zentriert ist. Anschließend wird jeder Pixelwert durch die Standardabweichung aller Pixelfarbwerte im gesamten Tensor geteilt. Hierdurch wird der Wertebereich der Pixelfarbwerte vereinheitlicht und weist eine Standardabweichung von  $\sigma = 1$  auf. Durch dieses Vorgehen werden die Pixelfarbwerte auf eine Standardnormalverteilung normiert. Das Ergebnis ist eine Datenrepräsentierung, bei der die Intensitätswerte in den Bildern vergleichbar sind und einen einheitlichen Mittelwert sowie Standardabweichung aufweisen. Die Dimensionen des resultierenden Tensors für den Aufmerksamkeitszweig bleiben dabei gleich (180, 72, 72, 3). Schließlich werden die beiden Tensoren aneinandergefügt, sodass sie einen Tensor im Format (180, 72, 72, 6) ergeben. Anders als in der ursprünglichen Datenvorverarbeitung des *TS-CAN*-Modells, wird keine Mittelung

der Eingangsdaten für den Aufmerksamkeitszweig vorgenommen (siehe Abschnitt 3.3). Dies soll gewährleisten, dass die Aufmerksamkeitsmaske stetig aktualisiert wird und bei bewegungsreichen Szenarien einer Änderung der Gesichtsorientierung folgt. Zu Beginn der Eingabe in das Modell wird der resultierende Eingangstensor wieder in separate Tensoren getrennt (Abbildung 30), welche im zugehörigen Zweig des *TS-CAN*-Modells für die Merkmalsextraktion verarbeitet werden.



(a) Vorimplementierte Datenvorverarbeitung in der rPPG-Toolbox [39].



(b) Selbst implementierte Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.

Abbildung 30.: Struktur der Eingangsdaten für das *TS-CAN*-Modell. Die linke Hälfte stellt die normalisierten Differenzbilder zweier aufeinanderfolgender Frames dar, welche den Eingang des Bewegungszweigs bilden. Die rechte Hälfte repräsentiert das Z-Score-standardisierte Videobild als Eingabe für den Aufmerksamkeitszweig.

Die Eingangstensoren werden von beiden Zweigen parallel verarbeitet, wobei der Erscheinungsbildzweig den Bewegungszweig über einen Gated-Attention-Mechanismus steuert. Die Merkmalsextraktion im Bewegungszweig erfolgt anhand der Differenz aufeinanderfolgender Bilder, um Bewegungsartefakte zu minimieren und den Lernprozess auf Pixelfarbunterschiede zwischen einzelnen Bildern zu fokussieren [35]. Durch den Einsatz der Attention Mask (Aufmerksamkeitsmaske) lenkt ein zweites Convolutional Attention Network (CAN) im Erscheinungsbildzweig den Fokus des CNNs im Bewegungszweig auf aussagekräftige Gesichtsbereiche (vergleiche Abbildung 8).

## 7.3. Training des TS-CAN-Modells

Das *TS-CAN*-Modell wird über 30 Epochen mit den in Abschnitt 7.1.3 beschriebenen Optimierungsstrategien trainiert. Es werden zwei Modelle mit jeweils anderen Trainings- und Validierungsdatensätzen erstellt. Ein Modelltraining erfolgt mit dem PURE-Datensatz (siehe Abbildung 31) und ein weiteres mit dem UBFC-rPPG-Datensatz. Die Wahl dieser beider Datensätze beruht darauf, dass beide Datensätze in einem verlustfreien Videoformat mit einer Farbtiefe von 8-Bit vorliegen [42][51]. Eine Forschungsarbeit von McDuff et al. hat die Auswirkung von Videokompressionsverfahren auf die rPPG-Signalextraktion untersucht und dabei festgestellt, dass bereits eine niedrige Videokompression das SNR des extrahierten BVP-Signals erheblich verschlechtert [55]. Darüber hinaus weist der PURE-Datensatz eine hohe Synchronisation zwischen den Videoaufzeichnungen und den Referenz-BVP-Signalen auf [54][66].

Es werden für den PURE- und UBFC-rPPG-Datensatz je zwei Modelle trainiert. Das erste Modell dient als Vergleichsmaßstab für das zweite Modell. Dabei wird das erste Referenzmodell mit der in der rPPG-Toolbox

vorimplementierten Datenvorverarbeitung trainiert (siehe Abbildung 30a). Mit dem zweiten Modell soll der Einfluss der in Kapitel 6 ermittelten ROI auf die Vorhersagegenauigkeit mit dem Referenzmodell verglichen werden. Die Eingangsdaten des zweiten Modells werden mit der in Abschnitt ROI-Extraktion nach Abschnitt 7.2 vorverarbeitet (siehe Abbildung 30b).

Zur Beschleunigung der beim Training anfallenden Berechnungen erfolgt das Training der Modelle auf einem zweiten PC mit einem Intel Core i7-6700K Prozessor mit 4 Kernen à 4,0 GHz und Hyperthreading. Der Arbeitsspeicher beträgt 16 GB und die Datensätze sind auf einer externen Festplatte mit einer Datenübertragungsrate von etwa 100 MB/s gespeichert. Die Berechnungen der Tensoroperationen werden auf einer externen Grafikkarte (NVIDIA GeForce GTX 980 Ti) mit NVIDIA CUDA Unterstützung durch die CUDNN-Bibliothek in der Version 11.8 durchgeführt.



Abbildung 31.: Trainings- und Validierungsverlust während des Trainings mit dem PURE-Datensatz über eine Trainingslaufzeit von 30 Epochen.

Dem Verlauf des Trainings- und Validierungsverlusts ist zu entnehmen, dass das Modell mit zunehmender Trainingsdauer erfolgreich Merkmale aus dem Trainingsdatensatz extrahiert, welche für die Rekonstruktion des BVP-Signals relevant sind (Abbildung 32). Der Grund, warum der Validierungsverlust stets niedriger als der Trainingsverlust ausfällt, ist der Einsatz der Dropout-Layer, wodurch während des Trainings nicht das gesamte neuronale Netz aktiv an der Signalextraktion beteiligt ist, sondern erst bei der Validierung und dem nachfolgenden Testen auf unbekannten Daten vollständig eingesetzt wird.





(b) Selbst implementierte Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.

Abbildung 32.: Visualisierung der Aufmerksamkeitsmasken im *TS-CAN*-Modell (entnommen und angepasst aus [35]). Das Convolutional Attention Network im Aufmerksamkeitszweig lernt selbstständig, aussagekräftige Bildbereiche für die BVP-Signalextraktion zu identifizieren. Die Aufmerksamkeitsmasken werden jeweils mit den Ausgängen der zweiten beziehungsweise vierten Faltungsschicht im Bewegungszweig multipliziert, wodurch die Gated-Attention-Mask erzeugt werden. Diese dienen als Eingänge für die nachfolgenden Schichten im Bewegungszweig.

# 7.4. Evaluation anhand unbekannter Testdatensätze

Das Ziel dieser Arbeit ist es, die Vorhersage der momentanen Herzrate in bewegungsreichen Szenarien robuster zu gestalten. Die Untersuchung hat das Ziel, den potenziellen Vorteil der ROI-Segmentierung bei für die Signalextraktion herausfordernden Aktivitäten zu untersuchen. Aus diesem Grund werden für die nachfolgende Analyse der Vorhersagegenauigkeit die Bewegungsszenarien aus den in Kapitel 4 vorgestellten Datensätzen eingesetzt. Die Evaluation der Modelle erfolgt anhand der beim Training vorenthaltenen Datensätze. Sofern vorhanden, wird nur das Bewegungsszenario in den Datensätzen untersucht. Die Datensätze, für die gesondert ein Bewegungsszenario vorliegt, sind der PURE-, VIPL-HR-V1- und MMPD-Datensatz. Beim RLAP-rPPG-Datensatz wird aufgrund vergleichbarer körperlicher Aktivität das Videospielszenario verwendet. Beim KISMED-Datensatz wird das Rotationsszenario (V11) untersucht. Der UBFC-rPPG- und COHFACE-Datensatz wird vollständig untersucht. Die Datenvorverarbeitung der Testdatensätze erfolgt analog zur Datenvorverarbeitung der Trainings- und Validierungsdatensätze.

Die vorhergesagten und tatsächlichen BVP-Signale werden im Rahmen der Evaluation in sich nicht überlappende Zeitfenster von 30 Sekunden Länge unterteilt, um eine höhere zeitliche Auflösung für die Herzratenvorhersage zu erhalten. Die Ergebnisse der beiden mit dem PURE-Datensatz trainierten Modelle sind in Tabelle 13 aufgetragen. Die Resultate der beiden mit dem UBFC-rPPG-Datensatz trainierten Modelle befinden sich in Tabelle 14.

	Train	ingsdat	ensatz: I	PURE											
Segmen- tierung	r	PPG-To	olbox Ben	chmark	C	Stirn- & Wangen-ROI									
Metrik Test- datensatz	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc				
PURE	-	-	-	-	-		-	-	-	-	-				
UBFC-rPPG	1,55	5,23	0,9591	1,81	86,00		1,02	3,19	0,9835	1,20	86,00				
MMPD Rotations - szenario	12,13	17,94	-0,0060	-8,39	35,86		9,78	15,36	0,2165	-8,07	45,26				
COHFACE	39,74	45,86	-0,1864	-22,81	10,03		29,44	39,11	-0,2888	-17,54	26,02				
VIPL-HR-V1 Bewegungs- szenario	12,51	15,51	0,1189	-11,20	16,52		10,62	14,44	0,2943	-9,53	23,14				
RLAP Videospiel - szenario	4,35	8,63	0,6911	-2,48	63,36		2,63	6,80	0,7938	0,53	80,60				
KISMED Rotations - szenario	5,98	10,43	0,5288	-2,57	60,00		1,85	3,71	0,9442	1,94	75,00				

 

 Tabelle 13.: Evaluation des Modells nach dem Training mit dem PURE-Datensatz über 30 Epochen anhand unbekannter Testdatensätze.

Gemittelt über die sechs evaluierten Datensätze ergibt sich durch die Segmentierung des Gesichtsbereichs in die Stirn- und Wangen-ROI eine Verringerung des MAE um 3, 49 beim Modell, das mit dem PURE-Datensatz

	Train	ingsdat	ensatz:	UBFC-1	rPPG					
Segmen- tierung	r	PPG-To	olbox Ben	chmark	C C		Stirn-	& Wange	n-ROI	
Metrik Test- datensatz	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc
PURE	5,10	15,95	0,7240	3,80	79,47	1,21	6,15	0,9626	9,33	90,07
UBFC-rPPG	-	-	-	-	-	-	-	-	-	-
MMPD Rotations - szenario	13,57	18,24	0,0507	-10,05	24,24	10,36	15,68	0,2123	-7,60	41,05
COHFACE	15,81	28,29	0,1238	-10,12	55,17	29,06	38,85	-0,2420	-17,10	26,96
VIPL-HR-V1 Bewegungs- szenario	9,72	13,26	0,3427	-10,28	28,10	10,10	13,82	0,3062	-8,87	25,62
RLAP Videospiel - szenario	2,45	6,05	0,8367	2,07	77,59	1,05	3,20	0,9511	4,23	87,93
KISMED Rotations - szenario	3,34	6,11	0,8494	-1,05	75,00	1,14	2,19	0,9806	2,34	90,00

**Tabelle 14.:** Evaluation des Modells nach dem Training mit dem UBFC-rPPG-Datensatz über 30 Epochen anhand<br/>unbekannter Testdatensätze.

trainiert wurde. Der RMSE verringert sich um 3, 50, der Pearson-Korrelationskoeffizient steigt um 0, 1397, das SNR um 2,36 dB und die Accuracy um 10, 71 %.

Beim Modell, das mit dem UBFC-rPPG-Datensatz trainiert wurde, können vergleichbare Verbesserungen durch den Einsatz der ROI-Segmentierung erzielt werden, wenn die schlechte Performance beim COHFACE-Datensatz bei der Berechnung des Durchschnittswertes nicht berücksichtigt wird. Der MAE sinkt durchschnittlich um 2,06 und der RMSE um 3,71. Der Pearson-Korrelationskoeffizient steigt um 0,1219, das SNR um 2,99 dB und die Accuracy um 10,05 %.

Die Abweichungen zwischen den vorhergesagten und den tatsächlichen Herzraten sind in Abbildung 33 für den PURE- und UBFC-rPPG-Datensatz in Form von Bland-Altman-Diagrammen dargestellt. Hierbei werden die Modelle eingesetzt, bei denen für den Erscheinungsbildzweig die Gesichtssegmentierung in die Stirn- und Wangenbereiche erfolgt. Den Grafiken ist zu entnehmen, dass der mittlere Fehler zwischen den vorhergesagten und den tatsächlichen Herzraten innerhalb des Toleranzbereichs von  $\pm 2$  BPM liegen. Weitere Diagramme zu den übrigen Datensätzen befinden sich in Anhang A.6.

Nachfolgend werden potenzielle Ursachen für die vergleichsweise schlechte Vorhersageleistung der Modelle beim MMPD-, COHFACE- und VIPL-HR-V1-Datensatz aufgeführt:

• **MMPD [52]:** Der MMPD-Datensatz enthält Aufnahmen von 33 Probanden, wovon 16 den Hauttyp III nach Fitzpatrick aufweisen. Die übrigen 17 Probanden weisen die Hauttypen IV bis VI auf. Diese



(a) Training mit dem PURE-Datensatz Evaluation am unbekannten UBFC-rPPG-Datensatz

(b) Training mit dem UBFC-rPPG-Datensatz Evaluation am unbekannten PURE-Datensatz

Abbildung 33.: Bland-Altman-Diagramme der Differenz zwischen den vorhergesagten Herzraten und den tatsächlichen Herzraten über dem Mittelwert der beiden Messwerte für (a) den UBFC-rPPG-Datensatz, nachdem das Modell mit dem PURE-Datensatz trainiert wurde und (b) den PURE-Datensatz, nachdem das Modell mit dem UBFC-rPPG-Datensatz trainiert wurde. Die Verteilung der Datenpunkte wird durch einen Gaußschen Kerndichteschätzer farblich hervorgehoben, wobei die höchste Datenpunktdichte gelb hervorgehoben ist.

Hauttypen sind in den Trainingsdatensätzen unterrepräsentiert. Das trainierte Modell weist deshalb Schwierigkeiten darin auf, für diese Hauttypen generalisierungsfähig zu sein (siehe Tabelle 15). Darüber hinaus wurde der MMPD-Datensatz mit einer *Samsung Galaxy S22 Ultra*-Smartphonekamera mit einer Auflösung von 1280x720 Pixel aufgezeichnet und aus Gründen der Datenspeicherung und -verbreitung auf eine Auflösung von 320x240 Pixel reduziert. Aus der Fachliteratur zum MMPD-Datensatz wird nicht ersichtlich, ob ein Videokompressionsverfahren eingesetzt wurde, allerdings lassen die Ergebnisse vermuten, dass eine potenzielle Kompression der Videos eine Ursache für die unzureichende Vorhersageleistung der überwachten und unüberwachten rPPG-Algorithmen ist.

- **COHFACE [48]:** Der COHFACE-Datensatz ist im MPEG-4-Format gespeichert. McDuff et al. haben die Auswirkung verschiedener Constant Rate Factors bei der Videokompression auf die BVP-Rekonstruktion durch rPPG-Algorithmen untersucht und festgestellt, dass das SNR mit steigendem Kompressionsfaktor zunehmend abnimmt [55]. Aus der Fachliteratur zum COHFACE-Datensatz wird nicht ersichtlich, welcher Videokompressionsfaktor eingesetzt wurde, allerdings lassen die Ergebnisse vermuten, dass die Kompression der Videos eine Ursache für die unzureichende Vorhersageleistung der überwachten und unüberwachten rPPG-Algorithmen ist.
- VIPL-HR-V1 [49]: Die untersuchten Videos des VIPL-HR-V1-Datensatzes wurden mit der Frontkamera eines Huawei P9-Smartphones aufgezeichnet. In der Veröffentlichung zum Datensatz beschreiben die Autoren, dass aus Gründen der Datenspeicherung und -verbreitung die Bildbereiche auf einen kleineren Bereich zugeschnitten wurden, sodass vornehmlich das Gesicht in den veröffentlichten Videos enthalten geblieben ist. Außerdem wurde ein Videokompressionsverfahren beruhend auf

dem MJPG-Codec eingesetzt und eine Reduktion der Auflösung auf zwei Drittel der ursprünglichen Auflösung von 1920x1080 Pixeln durchgeführt. Die Nachbearbeitung der Datensatzes bietet Grund zur Annahme, dass dabei relevante pulsatile Farbwertinformationen verloren gegangen sind, was die unzureichende Vorhersageleistung der rPPG-Algorithmen erklären könnte.

Tabelle 15.: Vergleich der Evaluationsmetriken in Abhängigkeit des Hauttyps nach Fitzpatrick für das Rotationsszena-<br/>rio im MMPD-Datensatz. Das verwendete Modell ist mit dem PURE-Datensatz trainiert worden, welcher<br/>nach subjektiver Beurteilung vornehmlich Hauttyp III enthält.

		Tes	stdate	nsatz:	MMP	D	Rota	tions	szena	rio				
Segmen- tierung	r	PPG-To	olbox B	enchma	rk		Stirn- & Wangen-ROI							
Metrik Hauttyp	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc			
III	5,96	10,47	0,4921	-7,01	55,88		5,10	9,44	0,5619	-6,21	62,75			
IV	17,29	24,66	-0,3170	-12,86	16,67		15,16	21,61	0,1416	-11,92	41,67			
v	21,48	26,69	-0,2691	-12,10	16,67		16,26	20,83	-0,0372	-12,22	17,86			
VI	24,46	29,16	-0,1365	-12,61	5,56		19,97	24,74	-0,0548	-14,03	11,11			

# 8. Diskussion

Im diesem Kapitel werden die Ergebnisse dieser Arbeit diskutiert. Dabei liegt der Fokus auf der Trainierbarkeit datenbasierter Algorithmen zur Rekonstruktion von Blutvolumenpulssignalen aus realistischen Versuchsszenarien und den Erkenntnissen, die aus dem Vergleich verschiedener Gesichtsbereiche, der Qualität der Datensätze und unterschiedlicher rPPG-Algorithmen gewonnen werden konnten.

### Vergleich der Gesichtsbereiche

Im Vergleich des Informationsgehalts der unterschiedlichen Gesichtsbereiche haben sich die Erkenntnisse aus vorangegangenen Arbeiten [13][40][41][43] bestätigt, dass der Gesichtsbereich der Stirn und Wangen ausgeprägtere pulsatile Informationen enthält, was für die Signalextraktion durch rPPG-Algorithmen von signifikantem Vorteil ist. Anders als zunächst durch die Forschung von Wong et al. [44] erwartet, hat die Reduktion des maximal zulässigen Reflexionswinkel für den Gesichtsbereich nicht den erwarteten Nutzen für die Verbesserung der Vorhersagegenauigkeit bei bewegungsreichen Einsatzszenarien gebracht.

#### Vergleich der rPPG-Algorithmen

Es konnte gezeigt werden, dass die unüberwachten rPPG-Algorithmen durch gezielte Einschränkung des analysierten Gesichtsbereichs eine bessere Signalrekonstruktion erzielen konnten, als bei der Verwendung des gesamten Gesichtsbereichs. Im Vergleich zum neuronalen End-to-End-Netzwerk konnte die Signalextraktion und Herzratenvorhersage durch die Integration der ROI-Segmentierung in die Datenvorverarbeitung verbessert werden. Im Vergleich der neuronalen Methoden mit den unüberwachten Algortihmen hat sich gezeigt, dass die neuronalen Methoden nach dem Training mit nur einem Datensatz über 30 Epochen noch unterlegen sind. Ziel dieser Arbeit war es, vorrangig die Vorhersagegenauigkeit der Algorithmen und die Auswirkung der ROI-Segmentierung bei bewegungsreichen Szenarien zu untersuchen. Allerdings liegt die Vermutung nahe, dass hier noch weiteres Verbesserungspotenzial besteht, sofern das Training über einen längeren Zeitraum und mit einer umfassenderen Datenbasis fortgesetzt wird.

#### Vergleich der Güte des Datensatzes

Es hat sich herausgestellt, dass die Leistung der eingesetzen Algorithmen stark vom verwendeten Datensatz abhängen. Insbesondere beim COHFACE-, VIPL-HR-V1- und MMPD-Datensatz haben sowohl bei die unüberwachten Algorithmen, als auch das überwachte neuronale Netz vergleichsweise schlechte Vorhersageergebnisse geliefert. Ein Grund hierfür können die in Abschnitt 7.4 genannten Auswirkungen von Videokompressionsverfahren sein. Für die PURE-, UBFC-rPPG-, RLAP-rPPG, und KISMED-Datensätze, die im unkomprimierten Bildformat vorliegen, konnten für beide Klassen von rPPG-Algorithmen zufriedenstellende Vorhersageleistungen erzielt werden, insbesondere bei Bewegungsszenarien. Die Fokussierung des Gesichtsbereichs auf die Stirn- und Wangen-ROI konnte hierbei sehr gute Ergebnisse erzielen. Lediglich in den Datensätzen, die im komprimierten Format vorliegen, scheinen nicht mehr die Bewegungsartefakte, sondern die Videokompression der größte Störfaktor zu sein. Die Auflösung der Datensätze spielt nur eine untergeordnete Rolle für die Signalextraktion [55][67], wohingegen das Videokompressionsformat entscheidend für eine präzise Extraktion des BVP-Signals [54][55] ist. Für zukünftige Arbeiten sollte demnach besonderes Augenmerk darauf gelegt werden, dass die Gesichtsaufzeichnungen in einem verlustfreien Video- beziehungsweise Bildformat vorliegen.

Im Rahmen dieser Arbeit wurde ein umfangreicher KISMED-Datensatz erstellt, der diverse bewegungsreiche Szenarien für die remote-Photoplethysmographie umfasst. Bei der Analyse dieser Szenarien konnte festgestellt werden, dass insbesondere solche mit translatorischen Bewegungen der Probanden Bewegungsartefakte in den Referenz-BVP-Signalen aufweisen, die die Berechnung der Referenzherzrate beeinträchtigen und damit den Vergleich mit den rekonstruierten BVP-Signalen erschweren können.

Ein Ansatz für zukünftige Forschungsschritte besteht darin, sich auf die Minimierung der Bewegungsartefakte in den Referenz-BVP-Signalen zu fokussieren.

# 9. Fazit und Ausblick

Im Rahmen dieser Forschungsarbeit ist eine adaptive ROI-Extraktion zur Segmentierung von Gesichtsbereichen aus Gesichtsvideos entwickelt und in ein System zur Extraktion von Blutvolumenpulssignalen durch rPPG-Algorithmen implementiert worden. Ausgehend von Literaturrecherchen wurden Experimente durchgeführt, um geeignete ROIs im Gesicht des gefilmten Probanden zu ermitteln. Zuletzt wurde ein bestehendes neuronales rPPG-Modell auf Basis von Convolutional Attention Networks (CAN) um die ermittelte optimale ROI erweitert. Dabei wurde evaluiert, ob die Signalqualität des extrahierten BVP-Signals durch den Fokus auf die aussagekräftigen Gesichtsbereiche zunimmt und die Genauigkeit der vorhergesagten Herzrate steigt.

Zunächst wurde unter Verwendung der MediaPipe-Bibliothek eine Datenvorverarbeitungspipeline entwickelt, die es ermöglicht, den Gesichtsbereich anhand von definierbaren Parametern zu segmentieren (Kapitel 5). Zu den Parametern gehören die maximalen zulässigen Reflexionswinkel, die zwischen der Gesichtsoberfläche und der Kamera auftreten dürfen, sowie die Auswahl von vordefinierten ROIs um den gesamten Gesichtsbereich in einen kleineren einzuschränken. Zudem kann bei Bedarf eine histogrammbasierte Hautsegmentierung durchgeführt werden, sowie die Interpolation der Reflexionswinkel, um eine höhere Auflösung aller Relexionswinkel im Gesichtsbereich zu erlangen. Die entwickelten Datenvorverarbeitungsschritte wurden schließlich in die Datenverarbeitungspipeline der rPPG-Toolbox [39] integriert.

Die Datenverarbeitungspipeline wurde anschließend dazu eingesetzt, um mithilfe von sechs für die Forschung zugänglichen rPPG-Datensätzen (UBFC-rPPG [42], COHFACE [48], VIPL-HR-V1[49], PURE [51], MMPD [52], RLAP-rPPG [54]) Untersuchungen durchzuführen (Kapitel 4). Hierzu wurde die rPPG-Toolbox um zusätzliche Datenladeprogramme für den RLAP-, COHFACE- und VIPL-HR-V1-Datensatz erweitert. Die bestehenden Datensätze weisen einige Kriterien auf, die in der rPPG-Forschung von Relevanz sind. Die Datensätze variieren in Bezug auf verschiedene Faktoren wie Aufnahmesituationen, Probandenvielfalt und Geräten zur Datenakquirierung und enthalten Aufnahmeszenarien unter verschiedenen Lichtbedingungen, Hintergründen oder Aktivitäten, welche es ermöglichen, die Robustheit der rPPG-Extraktion in unterschiedlichen Umgebungen und Alltagssituationen zu testen. Die Gemeinsamkeit der Datensätze liegt darin, dass die Versuchsszenarien größtenteils unter kontrollierten Laborbedingungen aufgezeichnet wurden. Dadurch wird die Extraktion des BVP-Signals durch rPPG-Algorithmen erleichtert, allerdings geschieht dies zum Nachteil der Verallgemeinerungsfähigkeit der Algorithmen in realen Anwendungsfällen außerhalb kontrollierter Laborumgebungen.

Ein weiterer Nachteil der bestehenden Datensätze ist es, dass wenige Härtefälle und Negativbeispiele mit enthaltenen Störfaktoren provoziert werden. Aus diesem Grund wurde im Rahmen dieser Arbeit der KISMED-Datensatz erstellt (Kapitel 4.7). Dieser hat das Ziel, solche Härtefälle mit bekannten Störfaktoren aufzuzeichnen. Bei der Konzipierung des Datensatzes wurde darauf geachtet, real vorkommende Alltagssituationen nachzubilden und die dabei vorkommenden Umwelteinflüsse miteinzubeziehen. Die

Umwelteinflüsse setzen sich zusammen aus Variationen in der Umgebungsbeleuchtung, Orientierungen des Probanden zur Kamera, wie Variationen in der Distanz, Gesichtsrotationen oder Translationen in der Bildebene, sowie variierenden Gesichtsbedeckungen. Der Datensatz wird dem Institut für künstlich intelligente Systeme in der Medizin der Technischen Universität Darmstadt für zukünftige interne Forschungsarbeiten zur Verfügung gestellt.

Mithilfe der rPPG-Datensätze wurden Untersuchungen anhand von sechs unüberwachten rPPG-Algorithmen (ICA [30], POS [24], CHROM [31], GREEN [29], LGI [33], PBV [32]) durchgeführt, um die aussagekräftigsten Gesichtsregionen zu ermitteln, welche ausgeprägte pulsatile Blutvolumenänderungen enthalten und sich damit besonders für die Extraktion eines Blutvolumenpulssignals eignen (Kapitel 6). Basierend auf Literaturrecherchen wurden hierzu Einschränkungen des Gesichtsbereichs in Abhängigkeit ihres Reflexionswinkels zwischen der Kamera und dem Normalenvektor der Gesichtsoberfläche durchgeführt. Die Untersuchung ergab, dass eine Verringerung des maximal gültigen Winkelschwellenwerts abseits des CHROM-Algorithmus keinen nennenswerten Vorteil für die Vorhersage der gemittelten Herzrate brachte. Weiterhin wurde die Vorhersagegenauigkeit der rPPG-Algorithmen unter Betrachtung verschiedener Melaninkonzentrationen in der Epidermis untersucht. Es konnten Ergebnisse aus vorangegangenen Studien [8][46][47] bestätigt werden, indem gezeigt wurde, dass die Algorithmen mit steigender Melaninkonzentration zunehmend schlechtere Resultate liefern. Des Weiteren wurde die Einschränkung des Gesichtsbereichs in die Regionen der Stirn und Wangen untersucht. Die Resultate haben gezeigt, dass diese Vorkehrung zur besten Vorhersageleistung der unüberwachten rPPG-Algorithmen führt.

Die optimale Stirn- und Wangen-ROI diente anschließend als Grundlage für die Eingangsdaten des bestehenden *TS-CAN*-Modells [35] (Kapitel 7). Um die Auswirkung der vorgenommenen Gesichtssegmentierung beurteilen zu können, wurde die selbst implementierte Gesichtsextraktion mit der vorimplementierten Datenverarbeitung innerhalb der rPPG-Toolbox verglichen. Es wurden vier Modelle trainiert, jeweils zwei mit dem PURE- und mit dem UBFC-rPPG-Datensatz.

Bei der Evaluation anhand der bewegungsreichen Szenarien in den vorenthaltenen Datensätzen ergibt sich beim *TS-CAN*-Modell, das mit dem PURE-Datensatz trainiert wurde, eine durchschnittliche Verringerung des MAE um 3, 49, wenn die Segmentierung des Gesichtsbereichs in die Stirn- und Wangen-ROI vorgenommen wird. Der RMSE verringert sich um 3, 50, der Pearson-Korrelationskoeffizient steigt um 0, 1397, das SNR um 2,36 dB und die Accuracy um 10, 71 %.

Beim Modell, das mit dem UBFC-rPPG-Datensatz trainiert wurde, können vergleichbare Verbesserungen durch den Einsatz der ROI-Segmentierung erzielt werden. Der MAE sinkt durchschnittlich um 2,06 und der RMSE um 3,71. Der Pearson-Korrelationskoeffizient steigt um 0,1219, das SNR um 2,99 dB und die Accuracy um 10,05 %.

Weiterhin wurden die Aufmerksamkeitsmasken der trainierten Modelle betrachtet. Die Visualisierung zeigt eindrücklich, wie das *TS-CAN*-Modell selbstständig lernt, den Fokus zur Signalextraktion auf aussagekräftige Gesichtsbereiche an Stirn, Wangen und Nase zu legen. Zusammenfassend hat die Gesichtssegmentierung einen signifikanten Beitrag zur Verbesserung der Vorhersageleistung der unüberwachten und überwachten neuronalen rPPG-Algorithmen geleistet, wie durch die durchschnittlichen Verbesserungen der Metriken deutlich geworden ist.

### 9.1. Ausblick

Bei der Evaluation längerer Videoaufnahmen werden die Daten im Rahmen dieser Arbeit in dreißigsekündige Zeitfenster Unterteilt. Dadurch wird ein Kompromiss zwischen einer höheren Vorhersagegenauigkeit zu Ungunsten einer niedrigeren zeitlichen Auflösung der vorhergesagten Herzraten eingegangen. In Kapitel 6.5 wird vorgeschlagen, wie sich die zeitliche Auflösung durch Anwendung der Moving-Window-Methode erhöhen ließe. Die Implementierung und Anwendung der Moving-Window-Methode bleibt Gegenstand zukünftiger Forschungsarbeiten, in denen der Einfluss der Fenstergröße oder der Anteil von sich überlappenden Zeitfenstern eingehender untersucht und optimiert werden kann. Aus kürzeren Zeitfenstern ergibt sich der Vorteil, dass kurzfristige Störeinflüsse, wie etwa durch Umgebungslichtänderungen, kompensiert werden, indem die Signale davor und danach stabil anliegen und man nur noch für den Zeitraum der Helligkeitsanpassung der Kamera einen Störeinfluss auf die rPPG-Signalextraktion hat.

Die Evaluation der rPPG-Algorithmen an den verschiedenen Hauttypen nach Fitzpatrick im MMPD-Datensatz hat gezeigt, dass diese hauptsächlich für den Hauttyp III geeignet sind. Mit steigender Melaninkonzentration in der Epidermis steigt die Absorption des innerhalb der Haut diffus reflektierten Lichts, was das SNR zunehmend verschlechtert (siehe Kapitel 2.2.1). Die Abhängigkeit der Lichtabsorption in der Epidermis vom epidermalen Melaningehalt wird in Abbildung 3 veranschaulicht. Es zeigt sich, dass die Absorption bei Wellenlängen im nahen Infrarotbereich (NIR) reduziert ist. Allerdings liegt in diesem Bereich eine niedrigere Lichtabsorption durch Hämoglobin vor (Abbildung 4). Diese Lichtabsorption durch Hämoglobin ist entscheidend für die algorithmische Erkennung pulsinduzierter Blutvolumenschwankungen. Eine mögliche Strategie zur Erhöhung der Robustheit der Herzratenschätzung für dunklere Hauttypen besteht in der kombinierten Auswertung von RGB- und NIR-Videos. Dieser Ansatz zur Herzratenschätzung durch rPPG-Verfahren konnte bereits in vorangegangenen Forschungsarbeiten erfolgreich demonstriert werden [68][69]. Ein Ansatz für zukünftige Forschungsarbeiten wäre die Untersuchung, ob die Fusion von RGBund NIR-Videos eine Verbesserung der Herzratenschätzung bei dunklen Hauttypen bewirkt.

Ein weiterer vielversprechender Schwerpunkt für zukünftige Forschungen liegt in der Vertiefung zur Visualisierung der Aufmerksamkeitsmaske oder grundsätzlich im Nachvollzug der Informationsverarbeitung innerhalb des neuronalen Netzes. In Hinblick auf die Merkmalsextraktion des angewandten neuronalen Netzes könnten Szenarien näher untersucht werden, in denen Gesichtsbereiche wie die Stirn, Wangen und die Nase bedeckt sind. So eröffnet sich die Möglichkeit, um herauszufinden, welcher verbleibende Bereich die besten Voraussetzungen für eine präzise Blutvolumenpulsrekonstruktion bietet.

Abschließend kann gesagt werden, dass die Rekonstruktion des Blutvolumenpulses aus Videos durch den Einsatz von rPPG-Algorithmen selbst unter herausfordernden Situationen sinnvoll eingesetzt werden kann, sofern entsprechende Vorkehrungen getroffen werden. Hierzu zählen die Segmentierung des Gesichtsbereichs in Regionen mit hoher Durchblutung und ausgeprägten pulsatilen Farbwertänderungen, sowie die Notwendigkeit von verlustfrei gespeicherten Videoaufnahmen. Angesichts der weit verbreiteten Verfügbarkeit kostengünstiger Kameras und des vergleichsweise geringen Rechenaufwands zur Bestimmung der Herzrate ist ein Einsatz der remote-Photoplethysmographie in der medizinischen Diagnostik oder telemedizinischen Anwendungen in Zukunft denkbar. Hierzu besteht allerdings noch weiterer Forschungsbedarf, etwa um die Anwendbarkeit unabhängig von der vorherrschenden Hautfarbe zu gewährleisten. Die kontinuierliche Forschung auf diesem Gebiet kann dazu beitragen, die Robustheit, Genauigkeit und Anwendbarkeit von rPPG-Algorithmen zu erhöhen und somit ihre potenzielle Integration in der medizinischen Diagnostik und Telemedizin voranzutreiben.

# Literatur

- Joachim Heuer, Annika Osterwald, Manas K Akmatov u. a. "Telemedizin als alternativer Zugang zu vertragsärztlicher ambulanter Versorgung Trends im Zeitraum 2017 bis 2021". In: *Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi). Versorgungsatlas-Bericht Nr. 23/06* (2023). DOI: 10.20364/VA-23.06.
- [2] Masaki Hasegawa, Kotaro Hayashi und Jun Miura. "Fatigue Estimation using Facial Expression features and Remote-PPG Signal". In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019 (2019). DOI: 10.1109/RO-MAN46459.2019.8956411.
- [3] Lingjian Kong, Kai Xie, Kaixuan Niu u. a. "Remote Photoplethysmography and Motion Tracking Convolutional Neural Network with Bidirectional Long Short-Term Memory: Non-Invasive Fatigue Detection Method Based on Multi-Modal Fusion." In: *Sensors (Basel, Switzerland)* 24.2 (2024), S. 455. ISSN: 1424-8220. DOI: 10.3390/S24020455. URL: http://www.ncbi.nlm.nih.gov/ pubmed/38257546.
- [4] Umur Aybars Ciftci, Ilke Demir und Lijun Yin. "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence X.X* (2020), S. 1–1. ISSN: 0162-8828. DOI: 10.1109/tpami.2020.3009287. arXiv: 1901.02212.
- [5] Jiahui Wu, Yu Zhu, Xiaoben Jiang u.a. "Local attention and long-distance interaction of rPPG for deepfake detection". In: *Visual Computer* 40.2 (2023), S. 1083–1094. ISSN: 01782789. DOI: 10.1007/s00371-023-02833-x. URL: https://doi.org/10.1007/s00371-023-02833-x.
- [6] Pireh Pirzada, Adriana Wilde, Gayle H Doherty u. a. "Remote Photoplethysmography (rPPG): A State-of-the-Art Review". In: *medRxiv* (2023). DOI: 10.1101/2023.10.12.23296882. eprint: https://www.medrxiv.org/content/early/2023/10/12/2023.10.12.23296882. full.pdf. URL: https://www.medrxiv.org/content/early/2023/10/12/2023.10. 12.23296882.
- [7] NuraLogix. *Anura*® *MagicMirror NuraLogix*. Jan. 2024. URL: https://xenodochial-johnson 9411.on.getshifter.io/magic-mirror/.
- [8] Jesse Fine, Kimberly L. Branan, Andres J. Rodriguez u. a. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Bd. 11. 4. 2021. ISBN: 1979845794. DOI: 10.3390 / bios11040126.
- [9] Stefan Silbernagl. *Taschenatlas Physiologie*. DeL. Thieme, 2012. ISBN: 9783131525383. URL: https://books.google.de/books?id=nvff2\_nmqn4C.
- [10] Siegfried Wassertheurer. "Pulswelle und Blutdruck: Kurz und bündig". In: *Journal für Hypertonie* 14 (Jan. 2010).
- [11] Peter Vaupel Gerhard Thews. *Vegetative Physiologie*. 5. Aufl. Springer-Verlag, 2005. ISBN: 3540240705. DOI: 10.1007/b137403.

- [12] Robert F. Schmidt, Florian Lang und Gerhrd Thews. *Physiologie des Menschen mit Pathophysiologie* (29. Auflage). 2005, S. 994. ISBN: 3540218823.
- [13] Daeyeol Kim, Kwangkee Lee und Chae Bong Sohn. "Assessment of ROI selection for facial video-based rPPG". In: *Sensors* 21.23 (2021). ISSN: 14248220. DOI: 10.3390/s21237923.
- [14] Karan Chopra, Daniel Calva, Michael Sosin u. a. "A comprehensive examination of topographic thickness of skin in the human face". In: *Aesthetic Surgery Journal* 35.8 (2015), S. 1007–1013. ISSN: 1527330X. DOI: 10.1093/asj/sjv079.
- [15] Steven H. Bailey, Georgette Oni, Spencer A. Brown u. a. *The use of non-invasive instruments in characterizing human facial and abdominal skin.* 2012. DOI: 10.1002/lsm.21147.
- [16] S. Rahrovan, F. Fanian, P. Mehryan u. a. "Male versus female skin: What dermatologists and cosmeticians should know". In: *International Journal of Women's Dermatology* 4.3 (2018), S. 122–130. ISSN: 23526475. DOI: 10.1016/j.ijwd.2018.03.002. URL: https://doi.org/10.1016/ j.ijwd.2018.03.002.
- [17] John D'Orazio, Stuart Jarrett, Alexandra Amaro-Ortiz u. a. "UV radiation and the skin". In: International Journal of Molecular Sciences 14.6 (2013), S. 12222–12248. ISSN: 14220067. DOI: 10.3390/ ijms140612222.
- [18] Jonathan P. Massie, Daniel Y. Cho, Cameron J. Kneib u. a. "Patient Representation in Medical Literature: Are We Appropriately Depicting Diversity?" In: *Plastic and Reconstructive Surgery - Global Open* 7.12 (2019), E2563. ISSN: 21697574. DOI: 10.1097/GOX.00000000002563.
- [19] Marks R LU Hongguang Pearse A. "Racial Differences of Melanin Content and Morphology in Surface Corneocytes–A Comparison Between Black, White and Oriental Skins". In: Chinese Journal of Dermatology 33.4, 254 (2000), S. 254. URL: http://www.pifukezazhi.com/EN/abstract/ article\_5017.shtml.
- [20] Hangxing Hu, Jin Li und Xiang Chen. "The effect of Skin Melanin Concentration on Wrist Reflectance Photoplethysmography based on Monte Carlo Simulation". In: *Proceedings - 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2022* (2022), S. 1–6. DOI: 10.1109/CISP-BMEI56279.2022.9979897.
- [21] Akhil Kallepalli und David B. James. "Quantification and influence of skin chromophores for remote detection of anemic conditions". In: February (2020), S. 46. ISSN: 16057422. DOI: 10.1117/12. 2545784.
- [22] Wei-Ting Zhang, Tzu-Hao Harry Chao, Yue Yang u. a. "Spectral Fiber-Photometry Derives Hemoglobin-Absorption Changes for Accurate Measurement of Fluorescent Sensor Activity". In: SSRN Electronic Journal (2021). DOI: 10.2139/ssrn.3943615.
- [23] N. De Pinho Ferreira, C. Gehin und B. Massot. "A Review of Methods for Non-Invasive Heart Rate Measurement on Wrist". In: Irbm 42.1 (2021), S. 4–18. ISSN: 18760988. DOI: 10.1016/j.irbm. 2020.04.001.
- [24] Wenjin Wang, Albertus C. Den Brinker, Sander Stuijk u. a. "Algorithmic Principles of Remote PPG". In: *IEEE Transactions on Biomedical Engineering* 64.7 (2017), S. 1479–1491. ISSN: 15582531. DOI: 10.1109/TBME.2016.2609282.
- [25] Toshiyo Tamura, Yuka Maeda, Masaki Sekine u. a. "Wearable photoplethysmographic sensors—past and present". In: *Electronics* 3.2 (2014), S. 282–302. ISSN: 20799292. DOI: 10.3390/electronic s3020282.

- [26] Chun Hong Cheng, Kwan Long Wong, Jing Wei Chin u. a. "Deep learning methods for remote heart rate measurement: A review and future research agenda". In: *Sensors* 21.18 (2021), S. 1–32. ISSN: 14248220. DOI: 10.3390/s21186296.
- [27] V. Blazek, T. Mühl, H. J. Schmitt u. a. "Optoelektronische Erfassung und rechnerunterstützte Analyse der Mikrozirkulations-Rhythmik". In: *Biomedizinische Technik/Biomedical Engineering* 30.s1 (1985), S. 121–122. ISSN: 0013-5585. DOI: 10.1515/bmte.1985.30.s1.121. URL: https://www.degruyter.com/document/doi/10.1515/bmte.1985.30.s1.121/html.
- [28] Vladimir Blazek und Steffen Leonhardt. Studies in Skin Perfusion Dynamics. 2021. ISBN: 9789811554476.
- [29] Wim Verkruysse, Lars O Svaasand und J Stuart Nelson. "Remote plethysmographic imaging using ambient light". In: *Optics Express* 16.26 (2008), S. 21434. ISSN: 1094-4087. DOI: 10.1364/oe.16.021434.
- [30] Ming Zher Poh, Daniel J. McDuff und Rosalind W. Picard. "Advancements in noncontact, multiparameter physiological measurements using a webcam". In: *IEEE Transactions on Biomedical Engineering* 58.1 (2011), S. 7–11. ISSN: 00189294. DOI: 10.1109/TBME.2010.2086456.
- [31] Gerard De Haan und Vincent Jeanne. "Robust pulse rate from chrominance-based rPPG". In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), S. 2878–2886. ISSN: 00189294. DOI: 10. 1109/TBME.2013.2266196.
- [32] G. De Haan und A. Van Leest. "Improved motion robustness of remote-PPG by using the blood volume pulse signature". In: *Physiological Measurement* 35.9 (2014), S. 1913–1926. ISSN: 13616579. DOI: 10.1088/0967-3334/35/9/1913.
- [33] Christian S. Pilz, Sebastian Zaunseder, Jarek Krajewski u. a. "Local group invariance for heart rate estimation from face videos in the wild". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2018-June (2018), S. 1335–1343. ISSN: 21607516. DOI: 10.1109/CVPRW.2018.00172.
- [34] Weixuan Chen und Daniel McDuff. "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks". In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11206 LNCS (2018), S. 356–373. ISSN: 16113349. DOI: 10.1007/978-3-030-01216-8\_22. arXiv: 1805.07888.
- [35] Xin Liu, Josh Fromm, Shwetak Patel u. a. "Multi-task temporal shift attention networks for ondevice contactless vitals measurement". In: Advances in Neural Information Processing Systems 2020-Decem.NeurIPS (2020), S. 1–12. ISSN: 10495258. arXiv: 2006.03790.
- [36] Zitong Yu, Xiaobai Li und Guoying Zhao. "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks". In: *30th British Machine Vision Conference 2019, BMVC 2019* (2020). arXiv: 1905.02419.
- [37] Si Qi Liu und Pong C. Yuen. "A General Remote Photoplethysmography Estimator with Spatiotemporal Convolutional Network". In: *Proceedings 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020* (2020), S. 481–488. DOI: 10.1109/FG47880.2020.00109.
- [38] Xin Liu, Brian Hill, Ziheng Jiang u. a. "EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Cardiac Measurement". In: Proceedings - 2023 IEEE Winter Conference on Applications of Computer Vision, WACV 2023 (2023), S. 4997–5006. DOI: 10.1109/WACV56688.2023.00498. arXiv: 2110.04447.
- [39] Xin Liu, Xiaoyu Zhang, Girish Narayanswamy u. a. "rPPG-Toolbox: Deep Remote PPG Toolbox". In: (2022), S. 1–19. arXiv: 2210.00716. URL: http://arxiv.org/abs/2210.00716.

- [40] Sungjun Kwon, Jeehoon Kim, Dongseok Lee u. a. "ROI analysis for remote photoplethysmography on facial video". In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2015-Novem (2015), S. 4938–4941. ISSN: 1557170X. DOI: 10.1109/EMBC.2015.7319499.
- [41] R. M. Fouad, Osama A. Omer und Moustafa H. Aly. "Optimizing Remote Photoplethysmography Using Adaptive Skin Segmentation for Real-Time Heart Rate Monitoring". In: *IEEE Access* 7 (2019), S. 76513–76528. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2922304.
- [42] Serge Bobbia, Richard Macwan, Yannick Benezeth u. a. "Unsupervised skin tissue segmentation for remote photoplethysmography". In: *Pattern Recognition Letters* 124 (2019), S. 82–90. ISSN: 01678655. DOI: 10.1016/j.patrec.2017.10.017.
- [43] Hans Peter Meinzer, Thomas Martin Deserno, Heinz Handels u. a. "Bildverarbeitung für die medizin 2013: Algorithmen Systeme Anwendungen: Proceedings des Workshops vom 3. bis 5. März 2013 in Heidelberg". In: *Informatik aktuell* (2013), S. 99–103. ISSN: 1431472X. DOI: 10.1007/978-3-642-36480-8.
- [44] Kwan Long Wong, Jing Wei Chin, Tsz Tai Chan u. a. "Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 2022-June.1 (2022), S. 2164–2170. ISSN: 21607516. DOI: 10.1109/CVPRW 56347.2022.00235.
- [45] Daniel McDuff. "Camera Measurement of Physiological Vital Signs". In: *ACM Computing Surveys* 55.9 (2023), S. 1–40. ISSN: 15577341. DOI: 10.1145/3558518. arXiv: 2111.11547.
- [46] Hannes Ernst, Matthieu Scherpf, Hagen Malberg u. a. "Optimal color channel combination across skin tones for remote heart rate measurement in camera-based photoplethysmography". In: *Biomedical Signal Processing and Control* 68 (2021). ISSN: 17468108. DOI: 10.1016/j.bspc.2021.102644.
- [47] Pradyumna Chari, Krish Kabra, Doruk Karinca u.a. "Diverse R-PPG: Camera-Based Heart Rate Estimation for Diverse Subject Skin-Tones and Scenes". In: (2020). arXiv: 2010.12769. URL: http://arxiv.org/abs/2010.12769.
- [48] Guillaume Heusch, André Anjos und Sébastien Marcel. "A Reproducible Study on Remote Heart Rate Measurement". In: (2017). arXiv: 1709.00962. URL: http://arxiv.org/abs/1709.00962.
- [49] Xuesong Niu, Hu Han, Shiguang Shan u. a. "VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-Constrained Face Video". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11365 LNCS (2019), S. 562–576. ISSN: 16113349. DOI: 10.1007/978-3-030-20873-8\_36. arXiv: 1810.04927.
- [50] Xuesong Niu, Shiguang Shan, Hu Han u. a. "RhythmNet: End-to-End Heart Rate Estimation from Face via Spatial-Temporal Representation". In: *IEEE Transactions on Image Processing* 29 (2020), S. 2409–2423. ISSN: 19410042. DOI: 10.1109/TIP.2019.2947204. arXiv: 1910.11515.
- [51] Ronny Stricker, Steffen Muller und Horst Michael Gross. "Non-contact video-based pulse rate measurement on a mobile service robot". In: IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions (2014), S. 1056–1062. DOI: 10.1109/ROMAN.2014.6926392.
- [52] Jiankai Tang, Kequan Chen, Yuntao Wang u. a. "MMPD: Multi-Domain Mobile Video Physiology Dataset". In: (2023). arXiv: 2302.03840. urL: http://arxiv.org/abs/2302.03840.

- [53] Jiankai Tang, Kequan Chen, Yuntao Wang u. a. GitHub McJackTang MMPD[EMBC 2023 Oral]. März 2023. URL: https://github.com/McJackTang/MMPD\_rPPG\_dataset (besucht am 24.01.2024).
- [54] Kegang Wang, Yantao Wei, Mingwen Tong u. a. "PhysBench: A Benchmark Framework for Remote Physiological Sensing with New Dataset and Baseline". In: (2023). arXiv: 2305.04161. URL: http: //arxiv.org/abs/2305.04161.
- [55] Daniel J. McDuff, Ethan B. Blackford und Justin R. Estepp. "The Impact of Video Compression on Remote Cardiac Pulse Measurement Using Imaging Photoplethysmography". In: Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 - 1st International Workshop on Adaptive Shot Learning for Gesture Understanding and Production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heterogeneous Face Recognition, HFR 2017, Joint Challenge on Dominant and Complementary Emotion Recognition Using Micro Emotion Features and Head-Pose Estimation, DCER and HPE 2017 and 3rd Facial Expression Recognition and Analysis Challenge, FERA 2017 (2017), S. 63–70. DOI: 10.1109/FG.2017.17.
- [56] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash u. a. "MediaPipe: A Framework for Building Perception Pipelines". In: (2019). arXiv: 1906.08172. uRL: http://arxiv.org/abs/1906.08172.
- [57] Siyuan Chen und Julien Epps. "Atomic Head Movement Analysis for Wearable Four-Dimensional Task Load Recognition". In: *IEEE Journal of Biomedical and Health Informatics* 23.6 (2019), S. 2464–2474. ISSN: 21682208. DOI: 10.1109/JBHI.2019.2893945.
- [58] G. E. Grossman, R. J. Leigh, L. A. Abel u. a. "Frequency and velocity of rotational head perturbations during locomotion". In: *Experimental Brain Research* 70.3 (1988), S. 470–476. ISSN: 00144819. DOI: 10.1007/BF00247595.
- [59] Carsten Rother, Vladimir Kolmogorov und Andrew Blake. "GrabCut Interactive foreground extraction using iterated graph cuts". In: ACM SIGGRAPH 2004 Papers, SIGGRAPH 2004 1.212 (2004), S. 309–314. DOI: 10.1145/1186562.1015720.
- [60] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant u. a. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), S. 261–272. DOI: 10.1038/s41592-019-0686-2.
- [61] Siu Kwan Lam, Antoine Pitrou und Stanley Seibert. "Numba: A LLVM-based Python JIT Compiler". In: Proceedings of LLVM-HPC 2015: 2nd Workshop on the LLVM Compiler Infrastructure in HPC - Held in conjunction with SC 2015: The International Conference for High Performance Computing, Networking, Storage and Analysis 2015-Janua (2015). DOI: 10.1145/2833157.2833162.
- [62] Ryanholbrook. Overfitting and underfitting. 20. Apr. 2023. URL: https://www.kaggle.com/ code/ryanholbrook/overfitting-and-underfitting (besucht am 18.01.2024).
- [63] Ilya Loshchilov und Frank Hutter. "Decoupled weight decay regularization". In: *7th International Conference on Learning Representations, ICLR 2019* (2019). arXiv: 1711.05101.
- [64] Leslie N. Smith und Nicholay Topin. "Super-convergence: very fast training of neural networks using large learning rates". In: (2019), S. 36. ISSN: 1996756X. DOI: 10.1117/12.2520589. arXiv: 1708.07120.
- [65] OneCycleLR PyTorch 2.1 Documentation. 2023. URL: https://pytorch.org/docs/stable/ generated/torch.optim.lr\_scheduler.OneCycleLR.html (besucht am 18.01.2024).
- [66] KegangWangCCNU. GitHub KegangWangCCNU/PhysBench: Simple, fast, and fair evaluation of remote physiological sensing models. uRL: https://github.com/KegangWangCCNU/PhysBench (besucht am 23.01.2024).

- [67] Ethan B. Blackford und Justin R. Estepp. "Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography". In: *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging* 9417 (2015), S. 94172D. ISSN: 16057422. DOI: 10.1117/12.2083940.
- [68] Shiika Kado, Yusuke Monno, Kenta Moriwaki u. a. "Remote Heart Rate Measurement from RGB-NIR Video Based on Spatial and Spectral Face Patch Selection". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2018-July (2018), S. 5676– 5680. ISSN: 1557170X. DOI: 10.1109/EMBC.2018.8513464.
- [69] Juan Cheng, Ping Wang, Rencheng Song u. a. "Remote Heart Rate Measurement from Near-Infrared Videos Based on Joint Blind Source Separation with Delay-Coordinate Transformation". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021). ISSN: 15579662. DOI: 10.1109/TIM. 2020.3041083.
- [70] Contec CMS-50E Bedienungsanleitung. URL: %7Bhttps://www.libble.de/contec-cms-50e/p/652636%7D (besucht am 01.02.2024).
- [71] pulsoximeter PULOX po-250 Bedienungsanleitung PO-250 Bedienungsanleitung.pdf. URL: %7Bhttps: //www.pulox.de/downloads/PO-250\_Bedienungsanleitung.pdf%7D (besucht am 01.02.2024).

# A. Anhang

# A.1. Datenblätter

Tabelle 16.: Datenblatt des zur Aufzeichnung vieler Datensätze verwendeten Fingerclippulsoximeters Contec Medical<br/>CMS-50E [70]. Das Pulsoximeter verfügt über einen Messbereich von 25 bis 250 BPM bei einer Auflösung<br/>von 1 BPM und weist eine Messgenauigkeit von 2 BPM bzw. 2% auf.

#### 10. Function Specification

Information	Display Mode
The Pulse Oxygen Saturation (%SpO <sub>2</sub> )	2-digit digital OLED display
Pulse Rate (PR)	3-digit digital OLED display
Pulse Intensity (bar-graph)	bar-graph OLED display
SpO2 Parameter Specification	
Measuring range	0%~100%, (the resolution is 1%).
Accuracy	70%~100%: ±2% ,Below 70% unspecified.
Pulse Parameter Specification	
Measuring range	25bpm~250bpm, (the resolution is 1bpm)
Accuracy	$\pm 2$ bpm or $\pm 2\%$ (select larger)
Safety Type	Interior Battery, B F Type
Pulse Intensity	
Range	Continuous bar-graph display, the higher display indicate the stronger pulse.

Tabelle 17.: Datenblatt des zur Aufzeichnung des eigenen Datensatzes verwendeten Fingerclippulsoximeters Pulox<br/>PO-250 [71]. Das Pulsoximeter verfügt über einen Messbereich von 30 bis 250 BPM bei einer Auflösung<br/>von 1 BPM und weist eine Messgenauigkeit von 2 BPM bzw. 2 % auf.

10. Function Specification	
Information	Display Mode
Pulse Oxygen Saturation (SpO2)	2-digit digital OLED display
Pulse Rate ( PR )	3-digit digital OLED display
Pulse Intensity (bar-graph)	Bar-graph OLED display
SpO2 Parameter Specification	
Measuring range	0%~100%, (the resolution is 1%).
Accuracy	70%~100%:±2%, Below 70% unspecified.
Average value	Calculate the Average value in every 4 measure value. The deviation between average value and true value does not exceed 1%.
Pulse Parameter Specification	
Measuring range	30bpm~250bpm, (the resolution is 1bpm)
Accuracy	±2bpm or±2% (select larger)
Average pulse rate	Moving calculate the Average pulse rate every 4 cardio-beat,s cycle. The deviation between average value and true value does not exceed 1%
Safety Type	Interior Battery, BF Type
Pulse Intensity	
Range	Continuous bar-graph display, the higher display indicates the stronger pulse.
Battery Requirement	
1.5V (AAA size) alkaline batteries ×2	<u>.</u>
Battery working life	
Two1.5 V (AAA size) 600 mAh alkali for 24 hours	ne batteries can work continually
Dimensions and Weight	
Dimensions	58.5(L) × 31(W) × 32 (H) mm
Weight	About 50g (with batteries)

10. Function Specification

# A.2. Resultate aus der Untersuchung zum optimalen Winkelschwellenwert (Ergänzung zu Kapitel 6.1)

						Dat	ensat	z: UBF	-rPI	۶G							
Winkel- schwellwert	1	rPPG-T	oolbox B	enchma	rk			15°		30°							
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MA	E F	RMSE	Pearson	SNR	Acc	
ICA	13,92	23,78	0,56	-8,88	52,38	6,03	14,37	0,70	-2,47	59,52	3,9	3 1	11,20	0,86	0,33	76,19	
POS	3,86	9,01	0,90	-2,59	64,29	3,11	6,96	0,92	0,23	66,67	1,6	3	4,04	0,97	2,48	78,57	
CHROM	3,76	11,06	0,84	-3,51	78,57	4,11	11,15	0,81	-1,84	59,52	1,6	4	3,41	0,98	-0,11	71,43	
GREEN	19,12	30,27	0,41	-11,16	40,48	29,35	34,69	0,12	-18,41	4,76	26,	12 3	31,97	0,20	-17,34	11,90	
LGI	15,68	28,57	0,36	-8,13	52,38	1,54	3,89	0,98	-1,97	73,81	1,5	1	3,77	0,98	0,66	73,81	
PBV	15,77	26,37	0,48	-9,32	45,24	11,38	19,82	0,46	-6,67	45,24	7,6	0 1	16,26	0,67	-5,79	59,52	
Winkel- schwellwert			45°					60°				90°					
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MA	E F	RMSE	Pearson	SNR	Acc	
ICA	3,58	10,49	0,87	0,21	76,19	3,39	9,85	0,88	0,06	73,81	2,0	9	6,24	0,94	0,90	78,57	
POS	1,73	4,01	0,97	2,91	73,81	1,55	3,79	0,98	2,97	76,19	1,4	8	3,74	0,98	2,97	78,57	
CHROM	1,91	4,35	0,97	0,19	71,43	1,75	4,28	0,97	-0,004	78,57	2,1	0	5,18	0,96	0,19	76,19	
GREEN	24,83	30,49	0,21	-16,15	16,67	23,89	30,11	0,07	-15,53	16,67	23,	18 3	32,02	-0,04	-15,01	21,43	
LGI	1,46	3,74	0,98	0,69	76,19	1,44	3,76	0,98	0,68	76,19	1,4	0	3,89	0,98	0,31	78,57	
PBV	6,95	15,31	0,73	-3,09	64,29	5,69	13,28	0,75	-1,14	64,29	5,0	2 1	12,39	0,75	-1,93	66,67	

 Tabelle 18.: Vergleich der Begrenzung des Gesichtsbereichs anhand unterschiedlicher Winkelschwellenwerte für den UBFC-rPPG-Datensatz.

						Dat	ensat	z: COI	IFAC	Έ								
Winkel- schwellwert	1	rPPG-T	oolbox B	enchma	rk			15°				30°						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc		
ICA	10,35	15,00	0,22	-5,89	31,97	12,34	16,86	0,14	-7,60	23,36		11,26	17,26	0,19	-5,91	32,35		
POS	13,18	17,19	0,11	-8,16	18,65	11,63	16,89	0,19	-7,04	30,33		8,99	15,49	0,33	-4,43	46,36		
CHROM	11,57	15,31	0,13	-7,82	19,67	14,66	19,25	0,04	-8,74	14,75		12,51	17,12	0,10	-7,36	20,45		
GREEN	6,48	11,64	0,56	-2,67	52,05	15,75	19,91	-0,01	-9,84	11,48		14,63	19,47	-0,01	-8,36	19,59		
LGI	12,42	16,16	0,07	-8,5	17,42	9,81	14,70	0,32	-6,62	35,45		8,21	14,07	0,40	-3,82	46,95		
PBV	13,13	17,02	0,07	-8,43	16,8	11,96	17,12	0,14	-7,92	25,82		12,50	17,51	0,10	-7,11	27,91		
Winkel- schwellwert			45°			60°						90°						
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc		
ICA	11,07	16,69	0,23	-5,96	33,83	11,67	17,59	0,20	-5,74	32,72		10,76	16,40	0,28	-5,38	35,12		
POS	8,14	14,45	0,38	-3,62	49,54	7,94	14,30	0,42	-3,57	49,91		7,72	13,69	0,40	-3,28	51,39		
CHROM	11,10	15,78	0,13	-6,96	26,83	11,10	16,42	0,17	-6,48	28,71		10,49	15,72	0,20	-6,42	30,77		
GREEN	14,28	20,11	0,09	-7,93	23,48	14,06	19,76	0,01	-7,53	24,21		12,65	18,43	0,18	-6,96	30,50		
LGI	7,51	13,84	0,44	-3,52	51,02	7,39	13,32	0,45	-3,24	50,65		7,24	13,38	0,45	-3,18	52,31		
PBV	12,15	17,94	0,10	-6,52	31,05	12,30	18,36	0,13	-6,63	29,76		11,33	17,20	0,16	-5,81	33,83		

 Tabelle 19.:
 Winkelschwellenwertvergleich für den COHFACE-Datensatz.

# A.3. Resultate aus der Untersuchung zur ROI-Segmentierung (Ergänzung zu Kapitel 6.3)
				Da	tensat	z: VII	L-HR-	-V1 (B	eweg	ungss	SZ	enar	io)			
Winkel- schwellwert	1	PPG-To	oolbox B	enchma	rk			15°						<b>30</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	14,52	17,92	0,1109	-14,18	8,65	14,4	8 17,57	-0,0619	-11,98	11,40		11,41	14,96	0,1709	-11,22	24,56
POS	12,39	16,63	0,0753	-13,28	13,26	11,2	14,39	0,1719	-10,78	19,30		11,69	15,34	0,0926	-10,10	24,56
CHROM	12,75	16,91	0,0928	-12,81	9,88	11,5	14,79	0,0968	-10,59	17,70		11,06	14,87	0,0912	-9,96	21,24
GREEN	16,32	19,86	-0,0107	-15,16	9,80	14,7	18,47	-0,1242	-12,32	7,02		14,90	18,29	-0,0855	-12,63	8,77
LGI	13,28	17,02	0,09	-13,51	12,39	13,9	17,52	-0,0522	-11,72	12,28		12,43	15,99	0,0745	-11,06	21,05
PBV	13,97	17,25	0,14	-14,25	11,53	12,8	15,78	0,0005	-12,25	15,79		12,54	16,11	0,0375	-10,92	17,54
Winkel- schwellwert			45°					60°						90°		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	12,8	16,36	-0,0661	-11,53	21,93	11,8	15,16	0,1556	-11,33	21,05		11,50	15,29	0,1943	-10,34	24,56
POS	9,88	13,67	0,1637	-9,48	30,70	9,59	13,29	0,2120	-9,45	31,58		9,44	13,06	0,2243	-9,47	32,46
CHROM	10,05	13,61	0,2129	-9,92	27,43	10,2	13,53	0,1680	-9,70	23,01		10,38	14,00	0,0638	-9,42	24,78
GREEN	13,77	17,02	-0,0468	-12,34	10,53	13,8	2 17,15	0,0435	-12,41	9,65		14,54	18,44	-0,0257	-12,16	15,79
LGI	12,66	16,90	0,0656	-10,80	24,56	12,7	8 16,73	0,0877	-10,90	18,42		13,55	18,06	-0,0327	-10,86	19,30
PBV	13,62	17,42	-0,0401	-11,51	17,54	15,2	18,89	0,0474	-11,84	16,67		14,97	18,84	0,0436	-11,72	19,30

 Tabelle 20.:
 Winkelschwellenwertvergleich f
 f
 ir das Bewegungsszenario im VIPL-HR-V1-Datensatz.

Tabelle 21.	: Winkelschwe	ellenwertvergleic	h für das	Bewegungssz	enario in	PURE-Datensatz.

					Dater	15	satz:	PURI	E (Bew	vegun	gssze	n	ario)				
Winkel- schwellwert	1	PPG-To	oolbox B	enchma	rk				15°						<b>30°</b>		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	7,03	21,43	0,4768	7,76	90,00		8,00	22,43	0,3946	0,32	80,00		0,53	1,47	0,9985	11,19	90,00
POS	0,44	1,04	0,9992	10,51	95,00		0,62	1,71	0,9977	4,42	90,00		0,18	0,56	0,9997	10,94	100,00
CHROM	0,44	1,30	0,9989	9,17	95,00		0,53	1,36	0,9987	3,01	95,00		0,53	1,36	0,9987	9,14	95,00
GREEN	11,6	26,58	-0,1592	-3,50	60,00		17,14	25,31	-0,0292	-11,56	5,00		17,93	24,19	0,0602	-11,25	10,00
LGI	4,66	17,76	0,6732	5,10	85,00		0,53	1,47	0,9985	4,06	90,00		0,53	1,47	0,9985	12,83	90,00
PBV	1,32	2,75	0,9931	5,65	80,00		1,85	3,95	0,9865	-0,96	80,00		8,17	21,88	0,3938	3,19	80,00
Winkel- schwellwert			45°						60°						<b>90</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	0,44	1,30	0,9989	13,83	95,00		0,26	0,68	0,9996	14,43	100,00		0,22	0,56	0,9997	14,55	100,00
POS	0,44	1,30	0,9989	13,14	95,00		0,35	1,24	0,9991	13,76	95,00		0,09	0,39	0,9999	14,07	100,00
CHROM	0,44	1,30	0,9989	10,52	95,00		0,53	1,36	0,9987	10,88	95,00		0,53	1,36	0,9987	10,88	95,00
GREEN	15,64	26,45	-0,2906	-9,37	20,00		10,81	26,85	-0,1517	-6,01	65,00		12,74	27,53	-0,2343	-4,87	60,00
LGI	0,35	1,24	0,9991	13,18	95,00		0,18	0,56	0,9997	13,27	100,00		0,09	0,39	0,9999	14,53	100,00
PBV	2,72	8,51	0,9337	7,06	85,00		5,19	14,26	0,7902	7,64	85,00		6,86	19,35	0,6874	7,55	85,00

						D	atens	satz: N	AMPE	)					
Winkel- schwellwert	1	PPG-To	oolbox B	enchma	rk			15°					<b>30</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	28,88	35,92	-0,0749	-16,14	14,02	22,76	29,76	0,0634	-15,05	17,8	22,94	30,49	0,0161	-14,10	20,45
POS	20,21	27,28	0,0647	-13,98	19,32	18,09	25,65	0,091	-12,43	26,89	16,29	24,86	0,1467	-10,19	35,23
CHROM	24,54	31,47	-0,0847	-15,37	15,91	20,78	28,02	0,0528	-13,47	21,97	19,56	27,25	0,1321	-11,77	25,76
GREEN	29,52	37,11	-0,0936	-15,53	15,91	24,54	30,47	0,0678	-16,36	8,33	25,18	30,94	0,0062	-16,15	7,20
LGI	23,83	30,85	0,0928	-15,31	17,42	22,07	29,60	0,0114	-14,20	20,45	19,85	28,81	0,0798	-11,54	33,33
PBV	27,46	33,57	-0,0073	-16,65	10,61	24,82	31,30	-0,0649	-15,49	12,50	23,65	30,94	0,0252	-14,17	18,94
			•												
Winkel- schwellwert			45°					60°					90°		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	23,17	31,12	-0,0158	-14,22	20,45	22,51	29,98	0,0994	-13,79	22,35	22,04	29,34	0,1089	-13,61	23,48
POS	16,23	24,85	0,1590	-9,38	38,64	16,06	24,83	0,1840	-9,36	38,26	15,76	24,68	0,1811	-9,22	39,77
CHROM	20,91	28,98	0,0437	-11,88	26,52	20,81	28,56	0,1040	-12,01	25,00	21,57	29,38	0,0361	-12,27	26,14
GREEN	25,22	30,66	0,0519	-15,96	6,44	24,36	30,32	0,0285	-15,47	9,09	23,53	29,86	-0,054	-15,00	10,98
LGI	19,48	28,61	0,0919	-10,77	33,71	18,87	27,98	0,1574	-10,71	35,61	20,46	29,45	0,0800	-10,89	34,09
PBV	24,97	32,30	0,0212	-14,71	18,94	24,64	31,78	0,0771	-14,68	18,18	25,68	32,81	0,0723	-15,03	17,05

 Tabelle 22.:
 Winkelschwellenwertvergleich f
 f
 in das Kniebeugen-Szenario im MMPD-Datensatz.

## A.4. Resultate aus der Untersuchung zur Segmentierung des Gesichtsbereichs (Ergänzung zu Kapitel 6.3)

				Da	atensat	z: RL	AP-rP	PG (V	ideos	pielsz	enari	o)			
Winkel- schwellwert	1	rPPG-To	oolbox B	enchma	ırk			15°					<b>30</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	6,18	11,50	0,4977	-4,39	58,19	5,09	9,91	0,5657	-2,86	63,79	2,67	6,54	0,7971	1,24	79,31
POS	2,17	6,2	0,8188	0,79	80,17	1,72	5,18	0,8698	0,91	84,05	0,95	2,69	0,9652	4,69	89,22
CHROM	1,83	4,76	0,8986	0,63	80,17	1,59	3,75	0,9314	-0,22	81,47	1,09	2,97	0,9582	2,74	87,07
GREEN	6,77	11,90	0,5016	-5,66	54,31	11,03	14,59	0,1424	-11,31	20,26	10,25	13,99	0,2383	-10,45	27,16
LGI	3,33	7,35	0,7446	-1,54	72,84	2,68	7,00	0,7594	-0,5	78,88	1,64	4,49	0,9006	3,04	82,76
PBV	5,45	9,99	0,5889	-4,30	59,48	4,33	8,89	0,6254	-3,93	68,1	3,77	8,27	0,6872	-2,00	69,4
Winkel- schwellwert			45°					60°					90°		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	2,27	6,55	0,7943	2,62	82,33	1,99	5,87	0,8319	3,29	84,05	1,66	4,97	0,8778	3,21	84,48
POS	0,95	2,69	0,9659	5,66	88,79	0,90	2,31	0,9754	5,96	88,36	0,84	2,19	0,9779	5,97	88,79
CHROM	0,86	2,23	0,9808	3,89	87,5	0,87	2,42	0,9720	4,13	88,36	0,95	2,53	0,9696	4,21	87,93
GREEN	9,68	13,17	0,2763	-9,77	27,59	9,11	13,20	0,2992	-9,49	33,19	9,27	13,75	0,2416	-9,26	33,19
LGI	1,36	3,65	0,9354	3,79	84,48	1,41	3,73	0,9329	3,97	83,62	1,18	3,04	0,9550	3,89	85,34
PBV	3,11	6,93	0,7568	-1,17	73,28	3,32	7,01	0,7508	-1,17	70,69	3,17	7,26	0,7467	-0,58	73,71

Tabelle 23.	: Winkelschwell	enwertvergleich für	das Bewegungsszenari	o im RLAP-rPPG-Datensatz.
-------------	-----------------	---------------------	----------------------	---------------------------

 Tabelle 24.:
 Winkelschwellenwertvergleich f
 ür das Rotationsszenario im KISMED-Datensatz.

					Daten	S	atz:	KISM	ED (R	otatio	nssze	n	ario)				
Winkel- schwellwert	1	PPG-To	oolbox B	enchma	rk				15°						<b>30</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	7,47	12,91	0,4161	-3,46	55,00		6,77	11,11	0,6353	-4,87	60,00		4,57	9,37	0,7096	-0,61	70,00
POS	0,88	1,57	0,9901	4,35	90,00		2,99	4,65	0,9045	-2,05	65,00		1,58	2,83	0,9649	2,09	85,00
CHROM	1,32	2,75	0,9669	3,15	85,00		2,29	4,45	0,9130	-2,64	80,00		1,14	2,33	0,9773	6,95	90,00
GREEN	15,38	19,56	0,1416	-10,42	25,00		11,16	14,39	0,2343	-11,54	10,00		13,18	17,01	0,1974	-11,20	10,00
LGI	3,60	9,29	0,682	1,97	85,00		3,87	7,10	0,7940	-4,31	60,00		1,49	2,97	0,9622	3,84	80,00
PBV	6,86	13,85	0,4718	-3,28	65,00		6,15	9,72	0,6738	-3,75	45,00		3,34	6,09	0,8518	-1,95	65,00
								•									
Winkel- schwellwert			45°						60°						<b>90</b> °		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	3,34	6,53	0,8209	3,57	70,00		2,64	5,36	0,8773	3,16	75,00		2,55	4,50	0,9162	3,63	75,00
POS	1,58	3,04	0,9617	3,25	90,00		0,97	1,97	0,9846	4,55	90,00		0,79	1,30	0,9927	6,48	90,00
CHROM	0,79	1,62	0,9906	5,01	90,00		0,97	2,26	0,9821	4,49	90,00		1,23	1,89	0,9785	2,96	90,00
GREEN	11,07	14,26	0,4203	-9,70	10,00		10,63	13,81	0,3920	-10,79	15,00		9,78	12,05	0,4106	-9,48	20,00
LGI	1,23	2,55	0,9719	4,89	85,00		0,88	1,47	0,9935	3,59	80,00		0,79	1,62	0,9908	4,39	85,00
PBV	3,25	5,26	0,8920	1,03	70,00		4,22	6,08	0,8608	1,51	75,00		2,11	4,38	0,9162	1,84	80,00

									Date	ensatz:	UBFC-	rPPG								
Segmen- tierung	r	PPG-To	olbox Ber	nchmarl	ĸ	90	° ohne	ROI Segr	nentieru	ing	9	0° mit F	OI Segm	entieru	ng	45	5° mit R	OI Segm	entierur	1g
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	13,92	23,78	0,5596	-8,88	52,38	2,09	6,24	0,9428	0,90	78,57	1,08	2,40	0,9909	1,96	80,42	3,72	10,13	0,8598	0,82	73,81
POS	3,86	9,01	0,8954	-2,59	64,29	1,48	3,74	0,9782	2,97	78,57	1,42	3,59	0,9753	3,61	78,57	1,50	3,93	0,9760	3,57	76,19
CHROM	3,76	11,06	0,8413	-3,51	78,57	2,10	5,18	0,9594	0,19	76,19	1,89	4,43	0,9679	0,90	73,81	1,77	4,14	0,9740	0,78	71,43
GREEN	19,12	30,27	0,4093	-11,16	40,48	23,48	32,02	-0,0404	-15,01	21,43	21,91	28,27	0,0668	-14,95	14,29	22,73	27,74	0,0508	-15,75	11,90
LGI	15,68	28,57	0,3608	-8,13	52,38	1,40	3,89	0,9761	0,31	78,57	1,40	3,89	0,9761	1,70	78,57	1,38	3,84	0,9769	1,73	78,57
PBV	15,77	26,37	0,4791	-9,32	45,24	5,02	12,39	0,7527	-1,93	66,67	4,46	11,92	0,7824	-1,16	71,43	3,22	9,29	0,8749	-0,48	73,81

Tabelle 25.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für den UBFC-rPPG-Datensatz.

 Tabelle 26.:
 Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für den COHFACE-Datensatz.

						-			Date	ensatz	COH	FACE								
Segmen- tierung	1	rPPG-To	olbox Be	nchmar	k	90	° ohne I	ROI Segm	entieru	ıng	90	)° mit R	OI Segme	entieru	ng	4:	5° mit R	OI Segm	entierun	ıg
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	10,35	15,00	0,2245	-5,89	31,97	11,05	16,78	0,2466	-5,38	34,01	10,76	16,40	0,2818	-6,21	35,12	10,59	16,55	0,2778	-6,71	34,22
POS	13,18	17,19	0,1113	-8,16	18,65	7,72	13,69	0,4026	-3,28	51,39	7,67	13,26	0,3678	-4,20	47,87	9,31	14,40	0,3678	-4,90	42,62
CHROM	11,57	15,31	0,1305	-7,82	19,67	10,49	15,72	0,1974	-6,42	30,77	11,23	18,09	0,1194	-6,92	27,58	12,02	16,37	0,1499	-7,57	21,11
GREEN	6,48	11,64	0,5610	-2,67	52,05	12,65	18,43	0,1782	-6,96	30,50	14,54	19,39	0,0965	-7,80	22,55	15,89	20,98	0,0446	-9,31	16,39
LGI	12,42	16,16	0,0697	-8,50	17,42	7,24	13,38	0,4533	-3,18	52,31	7,18	13,10	0,4751	-2,68	53,54	7,38	13,86	0,4390	-4,71	49,13
PBV	13,13	17,02	0,0716	-8,43	16,8	11,33	17,20	0,1590	-5,81	33,83	11,97	17,98	0,1610	-6,32	30,68	12,41	17,73	0,1078	-7,90	25,20

 

 Tabelle 27.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Bewegungsszenario im VIPL-HR-V1-Datensatz.

								Datensa	tz: VIP	L-HR-V	1 Bew	egungs	sszenari	D						
Segmen- tierung	r	PPG-To	olbox Be	nchmarl	ĸ	90	° ohne	ROI Segn	nentieru	ng	9	0° mit R	OI Segm	entierui	ıg	45	5° mit R	OI Segm	entierun	ıg
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	14,52	17,92	0,11	-14,18	8,65	11,50	15,29	0,1943	-10,34	24,56	10,63	14,77	0,2486	-10,05	25,54	12,92	16,95	0,1571	-10,83	22,83
POS	12,39	16,63	0,08	-13,28	13,26	9,44	13,06	0,2243	-9,47	32,46	9,02	12,11	0,2394	-9,45	34,95	11,90	16,77	0,1121	-10,83	24,91
CHROM	12,75	16,91	0,09	-12,81	9,88	10,38	14,00	0,0638	-9,42	24,78	10,31	13,90	0,1654	-9,24	24,78	13,20	18,55	0,1232	-11,55	18,60
GREEN	16,32	19,86	-0,01	-15,16	9,80	14,54	18,44	-0,0257	-12,16	15,79	13,62	17,24	-0,0532	-11,26	17,54	14,27	18,47	0,1759	-12,74	15,38
LGI	13,28	17,02	0,09	-13,51	12,39	13,55	18,06	-0,0327	-10,86	19,30	12,20	15,85	0,2102	-10,14	21,63	12,19	16,53	0,3158	-11,28	17,54
PBV	13,97	17,25	0,14	-14,25	11,53	14,97	18,84	0,0436	-11,72	19,30	14,69	18,53	-0,2188	-11,08	18,26	13,20	17,82	0,2015	-11,83	14,91

								Dat	ensatz	: PURE	Beweg	gungssz	zenario							
Segmen- tierung	r	PPG-To	olbox Bei	nchmarl	ĸ	90	° ohne	ROI Segn	nentier	ung	9	0° mit I	ROI Segm	entieru	ng	4	5° mit F	tOI Segm	entieru	ng
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	7,03	21,43	0,4768	7,76	90,00	0,22	0,56	0,9997	14,55	100,00	0,18	0,39	0,9989	13,23	100,00	0,53	1,36	0,9987	12,49	95,00
POS	0,44	1,04	0,9992	10,51	95,00	0,09	0,39	0,9999	14,07	100,00	0,09	0,39	0,9999	14,32	100,00	0,09	0,39	0,9999	13,12	100,00
CHROM	0,44	1,30	0,9989	9,17	95,00	0,53	1,36	0,9987	10,88	95,00	0,44	1,27	0,9989	10,12	95,00	0,44	1,30	0,9989	10,17	95,00
GREEN	11,6	26,58	-0,1592	-3,50	60,00	12,74	27,53	-0,2343	-4,87	60,00	13,36	26,55	-0,1285	-8,99	55,00	17,49	25,27	0,0734	-10,39	20,00
LGI	4,66	17,76	0,6732	5,10	85,00	0,09	0,39	0,9999	14,53	100,00	0,09	0,39	0,9999	15,27	100,00	0,09	0,39	0,9999	12,95	100,00
PBV	1,32	2,75	0,9931	5,65	80,00	6,86	19,35	0,6874	7,55	85,00	8,00	22,30	0,4259	4,56	80,00	7,56	18,73	0,6893	2,10	75,00

 

 Tabelle 28.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Bewegungsszenario im PURE-Datensatz.

Tabelle 29.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für den MMPD-Datensatz.

									E	atensa	tz: MN	IPD								
Segmen- tierung	1	PPG-To	olbox Be	nchmar	k	90	)° ohne	ROI Segr	nentieru	ing	9	0° mit R	OI Segm	entierur	ıg	45	5° mit R	OI Segme	ntierun	g
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	28,88	35,92	-0,0749	-16,14	14,02	23,17	31,12	-0,0158	-13,61	23,48	21,11	29,09	0,1039	-13,09	27,27	22,37	30,29	-0,0077	-13,31	25,76
POS	20,21	27,28	0,0647	-13,98	19,32	16,23	24,85	0,1811	-9,38	38,64	14,10	22,17	0,3309	-8,85	40,91	14,36	22,59	0,3134	-8,86	39,02
CHROM	24,54	31,47	-0,0847	-15,37	15,91	21,57	29,38	0,0361	-12,27	26,52	18,24	26,67	0,2298	-11,43	32,58	18,74	27,12	0,1961	-11,40	32,20
GREEN	29,52	37,11	-0,0936	-15,53	15,91	23,53	29,86	-0,054	-15,00	10,98	23,17	28,67	-0,0438	-15,36	9,47	23,20	29,10	-0,0602	-15,60	8,33
LGI	23,83	30,85	0,0928	-15,31	17,42	20,46	29,45	0,0800	-10,89	34,09	17,67	26,73	0,2691	-10,22	38,64	17,80	26,75	0,2421	-10,25	36,74
PBV	27,46	33,57	-0,0073	-16,65	10,61	25,68	32,81	0,0723	-15,03	17,05	22,74	30,31	0,1744	-13,77	23,86	23,44	30,99	0,0223	-13,96	21,21

 

 Tabelle 30.:
 Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Bewegungsszenario im RLAP-rPPG-Datensatz.

								Dat	ensatz:	RLAP	Videos	pielsze	nario							
Segmen- tierung	1	rPPG-To	olbox Be	nchmar	k	90	)° ohne	ROI Segn	nentieru	ing	9	0° mit R	OI Segm	entierur	ıg	4	5° mit R	OI Segm	entierur	ıg
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	6,18	11,50	0,4977	-4,39	58,19	1,66	4,97	0,8778	3,21	84,48	1,27	4,29	0,8714	3,44	84,59	1,86	5,47	0,8533	3,21	82,02
POS	2,17	6,20	0,8188	0,79	80,17	0,84	2,19	0,9779	5,97	88,79	0,78	1,98	0,9831	5,07	91,22	0,99	2,78	0,9632	5,01	88,16
CHROM	1,83	4,76	0,8986	0,63	80,17	0,95	2,53	0,9696	4,21	87,93	0,86	2,29	0,9753	3,55	90,09	0,96	2,86	0,9617	3,42	89,47
GREEN	6,77	11,90	0,5016	-5,66	54,31	9,27	13,75	0,2416	-9,26	33,19	8,80	12,97	0,3198	-8,80	34,05	10,25	14,39	0,2398	-9,76	25,44
LGI	3,33	7,35	0,7446	-1,54	72,84	1,18	3,04	0,9550	3,89	85,34	1,24	3,10	0,9534	3,82	86,40	1,08	2,91	0,9594	3,75	84,91
PBV	5,45	9,99	0,5889	-4,30	59,48	3,17	7,26	0,7467	-0,58	73,71	2,90	6,28	0,8197	-0,64	73,41	3,49	7,58	0,7464	-0,94	72,37

								Date	ensatz	: KISME	D Rot	ationss	zenario							
Segmen- tierung	r	PPG-To	olbox Bei	nchmarl	ĸ	90	° ohne	ROI Segn	nentier	ung	9	0° mit l	ROI Segm	entieru	ng	4	5° mit F	ROI Segm	entierui	ıg
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	7,47	12,91	0,4161	-3,46	55,00	2,55	4,50	0,9162	3,63	75,00	2,20	3,91	0,9324	3,86	80,00	5,62	11,84	0,5139	2,13	70,00
POS	0,88	1,57	0,9901	4,35	90,00	0,79	1,30	0,9927	6,48	90,00	0,53	1,11	0,9948	6,82	95,00	1,41	2,78	0,9684	4,25	85,00
CHROM	1,32	2,75	0,9669	3,15	85,00	1,23	1,89	0,9785	2,96	90,00	1,05	2,29	0,9778	5,42	90,00	0,79	1,89	0,9846	4,16	90,00
GREEN	15,38	19,56	0,1416	-10,42	25,00	9,78	12,05	0,4106	-9,48	20,00	8,53	11,77	0,4845	-9,59	25,00	13,45	17,06	0,2673	-10,72	20,00
LGI	3,60	9,29	0,682	1,97	85,00	0,79	1,62	0,9908	4,39	85,00	0,70	1,24	0,9908	4,68	90,00	1,49	2,97	0,9622	3,27	85,00
PBV	6,86	13,85	0,4718	-3,28	65,00	2,11	4,38	0,9162	1,84	80,00	1,58	3,77	0,9395	1,71	90,00	3,54	6,09	0,7288	-4,04	80,00

 

 Tabelle 31.: Untersuchung zur ROI-Segmentierung des Gesichtsbereichs für das Rotationsszenario im KISMED-Datensatz.

 Tabelle 32.:
 Vergleich zur Segmentierung des Gesichtsbereichs für den UBFC-rPPG-Datensatz.

						Datens	atz: U	BFC-rPF	G						
ROI	Stirn-	und Wa	angen-RC	I		Interp	olierte	ROI			Komb	inierte	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	1,08	2,40	0,9909	1,96	80,42	5,43	13,51	0,7630	-0,04	71,60	3,04	11,00	0,8063	0,85	78,72
POS	1,42	3,59	0,9753	3,61	78,57	2,08	6,21	0,9351	2,60	74,25	1,69	5,72	0,9433	2,65	77,95
CHROM	1,89	4,43	0,9679	0,90	73,81	2,69	7,13	0,9173	0,68	71,31	1,74	5,73	0,9445	0,61	72,72
GREEN	21,91	28,27	0,0668	-14,95	14,29	23,20	30,68	0,0882	-15,20	18,52	21,77	30,31	0,0851	-13,80	25,93
LGI	1,40	3,89	0,9761	1,70	78,57	2,99	8,91	0,8754	0,68	76,48	1,67	5,99	0,9375	1,27	78,57
PBV	4,46	11,92	0,7824	-1,16	71,43	8,66	18,94	0,5644	-3,05	59,26	7,03	17,12	0,6124	-0,64	69,14

 Tabelle 33.: Vergleich zur Segmentierung des Gesichtsbereichs bei einer natürlichen Sonnenlichtbestrahlung im COHFACE-Datensatz.

			1	Datens	satz: C	OHE	ACE (	Son	nenlich	tbestı	ahlung	g)					
ROI	Stirn-	und Wa	angen-RO	I		In	erpoli	erte	ROI				Komb	inierte	ROI		
Metrik	MAE	RMSE	Pearson	SNR	Acc	M	E RN	ASE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
Methode	Ļ	Ļ		(ub) ↑	(%) ↑												
ICA	10,47	13,98	0,1606	-9,26	18,75	11	26 14	l,91	0,1564	-9,35	20,00		10,62	14,52	0,1952	-8,77	21,88
POS	9,57	13,92	0,3006	-8,35	27,50	10	50 14	1,27	0,2858	-8,87	19,38		9,83	14,29	0,2615	-7,75	30,00
CHROM	10,13	13,50	0,2460	-8,92	23,75	10	53 14	l,14	0,2103	-9,05	16,88		9,90	13,97	0,1802	-8,31	26,88
GREEN	12,34	17,24	0,1586	-9,27	26,25	12	32 17	7,30	0,0817	-9,48	18,75		12,50	17,17	0,1226	-9,30	23,12
LGI	9,49	14,03	0,2959	-8,32	33,12	11	19 15	5,60	0,1082	-8,82	21,88		9,49	14,32	0,3082	-7,84	32,50
PBV	11,48	15,38	0,0965	-9,49	16,88	10	95 14	1,63	0,1437	-9,57	18,75		9,75	13,93	0,2362	-8,61	27,50

			]	Daten	satz: C	CC	OHFAC	E (kür	stliche	Beleu	chtun	g)					
ROI	Stirn-	und Wa	angen-RO	I			Interp	olierte	ROI				Komb	inierte	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc		MAE	RMSE	Pearson	SNR	Acc
ICA	9,30	14,76	0,3129	-4,62	41,47		11,35	15,97	0,1267	-8,75	24,54		7,74	11,51	0,5028	-6,56	37,42
POS	8,04	15,47	0,3855	-3,24	57,14		7,38	11,99	0,4701	-7,22	46,01		4,58	9,03	0,7009	-3,86	65,64
CHROM	10,97	16,24	0,0752	-6,90	35,41		12,88	16,37	-0,0238	-9,27	14,11		9,84	13,56	0,2905	-7,53	29,45
GREEN	11,32	18,46	0,2916	-6,59	43,32		11,93	16,35	0,2293	-9,10	25,15		9,29	14,06	0,3866	-8,27	37,42
LGI	6,31	13,76	0,5025	-2,50	64,52		7,67	13,26	0,3659	-6,33	47,24		4,39	10,01	0,6236	-3,35	71,17
PBV	10,30	17,28	0,2459	-4,80	40,55		11,25	15,85	0,1265	-8,40	26,38		9,41	14,20	0,2134	-6,67	35,58

 

 Tabelle 34.: Vergleich zur Segmentierung des Gesichtsbereichs bei einer künstlichen Beleuchtung im COHFACE-Datensatz.

 Tabelle 35.:
 Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im VIPL-HR-V1-Datensatz.

				Date	ensatz	VIPL-	HR-V1	Bewegu	ingssz	enario					
ROI	Stirn-	und Wa	angen-RO	I		Inter	polierte	ROI			Komb	inierte	ROI	_	
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	10,63	14,77	0,2486	-10,05	25,54	12,34	15,41	0,1455	-11,18	14,91	10,51	14,89	0,0914	-11,04	20,91
POS	9,02	12,11	0,2394	-9,45	34,95	11,09	14,58	0,1621	-10,66	20,18	8,54	11,30	0,1516	-9,59	28,07
CHROM	10,31	13,90	0,1654	-9,24	24,78	10,16	13,54	0,0681	-10,16	23,01	9,52	12,79	0,2924	-9,49	24,78
GREEN	13,62	17,24	-0,0532	-11,26	17,54	12,01	15,25	0,1812	-11,76	17,54	13,75	17,29	0,1310	-11,63	15,79
LGI	12,20	15,85	0,2102	-10,14	21,63	13,86	17,10	-0,1092	-11,25	9,65	12,26	16,71	0,1040	-10,40	21,93
PBV	14,69	18,53	-0,2188	-11,08	18,26	15,02	18,12	-0,0527	-11,60	9,65	14,28	17,59	0,0673	-11,60	9,65

Tabelle 36.: Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im PURE-Datensatz.

				Ι	Datensa	tz: PU	RE Bew	vegungs	szena	rio					
ROI	Stirn-	und Wa	angen-RO	I		Inte	rpolierte	ROI			Komb	inierte	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑	MAI	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	0,18	0,39	0,9989	13,23	100,00	1,32	4,39	0,9828	4,09	95,0	0,18	0,56	0,9997	12,29	100,0
POS	0,09	0,39	0,9999	14,32	100,00	0,09	0,39	0,9999	6,35	100,0	0,35	1,24	0,9991	12,45	95,0
CHROM	0,44	1,27	0,9989	10,12	95,00	0,53	1,24	0,9987	4,70	90,0	0,44	1,30	0,9989	9,22	95,0
GREEN	13,36	26,55	-0,1285	-8,99	55,00	15,4	7 25,41	0,0855	-7,97	35,0	11,87	26,73	-0,1409	-6,54	60,0
LGI	0,09	0,39	0,9999	15,27	100,00	0,62	1,42	0,9985	5,14	85,0	0,26	0,88	0,9993	12,06	95,0
PBV	8,00	22,30	0,4259	4,56	80,00	5,19	14,52	0,7795	1,53	80,0	5,89	17,29	0,7213	8,32	85,0

				D	atensa	at	z: MM	PD Be	wegung	sszena	ario					
ROI	Stirn-	und Wa	ngen-RO	I			Interp	olierte	ROI			Komb	inierte	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	21,11	29,09	0,1039	-13,09	27,27		21,46	28,65	0,1309	-13,93	23,48	21,90	29,20	0,1257	-13,49	23,48
POS	14,10	22,17	0,3309	-8,85	40,91		16,09	24,03	0,2389	-10,24	37,88	15,53	24,35	0,2204	-8,99	40,67
CHROM	18,24	26,67	0,2298	-11,43	32,58		20,10	27,99	0,1442	-12,33	28,41	20,19	28,28	0,1435	-11,64	31,06
GREEN	23,17	28,67	-0,0438	-15,36	9,47		23,00	29,07	-0,0246	-15,38	9,47	22,74	28,83	-0,0399	-15,43	8,71
LGI	17,67	26,73	0,2691	-10,22	38,64		20,21	28,74	0,1019	-11,66	32,58	18,72	27,78	0,2059	-10,44	37,88
PBV	22,74	30,31	0,1744	-13,77	23,86	1	23,55	30,88	0,1045	-14,06	22,35	22,47	30,01	0,1979	-13,29	25,38

 Tabelle 37.:
 Vergleich zur Segmentierung des Gesichtsbereichs beim Bewegungsszenario im MMPD-Datensatz.

Tabelle 38.: Vergleich zur Segmentierung des Gesichtsbereichs beim Videospielszenario im RLAP-rPPG-Datensatz.

				Ι	Datens	at	tz: RL	AP Vid	eospiels	szena	rio					
ROI	Stirn-	und Wa	angen-RC	DI			Interp	olierte	ROI			Komb	inierte	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	1,27	4,29	0,8714	3,44	84,59		4,16	9,17	0,6304	-2,34	70,26	1,95	6,21	0,8123	2,80	84,91
POS	0,78	1,98	0,9831	5,07	91,22		1,41	3,96	0,9218	1,79	84,91	0,95	2,69	0,9652	5,32	88,36
CHROM	0,86	2,29	0,9753	3,55	90,09		1,63	5,10	0,8714	0,62	82,76	1,13	3,66	0,9353	3,65	87,07
GREEN	8,80	12,97	0,3198	-8,80	34,05		10,26	13,99	0,2727	-9,72	26,72	8,99	12,76	0,3460	-9,12	33,62
LGI	1,24	3,10	0,9534	3,82	86,40		1,80	4,74	0,8888	0,48	80,60	1,31	3,59	0,9373	3,77	84,91
PBV	2,90	6,28	0,8197	-0,64	73,41		4,33	8,85	0,6549	-3,51	65,95	3,89	8,93	0,6336	-1,32	70,69

Tabelle 39.: Vergleich zur Segmentierung des Gesichtsbereichs beim Rotationsszenario im KISMED-Datensatz.

				]	Datens	a	tz: KI	SMED	Rotatio	nsszen	ario					
ROI	Stirn-	und W	angen-RC	DI			Interp	olierte	ROI			Kombi	inierte 1	ROI		
Metrik Methode	MAE ↓	RMSE ↓	Pearson ↑	SNR (dB) ↑	Acc (%) ↑		MAE	RMSE	Pearson	SNR	Acc	MAE	RMSE	Pearson	SNR	Acc
ICA	2,20	3,91	0,9324	3,86	80,00		6,77	10,20	0,6029	-6,55	30,00	4,66	7,99	0,7582	-0,73	55,00
POS	0,53	1,11	0,9948	6,82	95,00		2,29	4,16	0,9226	1,21	75,00	1,32	3,07	0,9613	5,39	85,00
CHROM	1,05	2,29	0,9778	5,42	90,00		1,93	3,65	0,9424	-1,35	75,00	1,05	2,29	0,9778	3,50	90,00
GREEN	8,53	11,77	0,4845	-9,59	25,00		12,83	15,61	-0,0512	-11,75	5,00	12,22	15,81	0,0691	-10,83	15,00
LGI	0,70	1,24	0,9908	4,68	90,00		2,29	3,47	0,948	-2,06	60,00	1,32	2,12	0,9808	3,26	75,00
PBV	1,58	3,77	0,9395	1,71	90,00		3,16	5,06	0,8993	-4,6	60,00	3,43	7,45	0,8062	-0,75	65,00

## A.5. Resultate aus der Kombination mehrerer extrahierter Herzraten (Ergänzung zu Kapitel 6.4)

Tabelle 40.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Im Anschluss an die<br/>Datenvorverarbeitung wird jedes Video im UBFC-rPPG-Datensatz in ein kürzeres Zeitfenster von 10 s<br/>unterteilt. Die Evaluationsmetriken beziehen sich auf sämtliche Videofenster, die aus dem Datensatz<br/>entstehen. Die separate Auswertung der Sub-ROIs oder deren Kombination hat keine Verbesserung in<br/>der Vorhersageleistung erzielt. Die besten Evaluationsmetriken liefert die Begrenzung der ROI auf die<br/>Stirn- und Wangenregion.

		0 0						
	Datensat	z: UBFC-r	PPG					
	Methode	: ICA						
ROI	Stirn &	Ctirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
$MAE\downarrow$	1,82	7,53	6,97	8,09	6,40	6,41	6,59	4,50
$RMSE\downarrow$	6,03	18,01	16,50	16,57	12,46	13,76	16,56	11,62
Pearson ↑	0,9371	0,5403	0,6329	0,7177	0,7692	0,7103	0,6011	0,7868
SNR (dB) ↑	5,59	0,39	-0,40	-0,92	2,38	2,20	1,38	2,24
Acc (%) ↑	86,42	69,14	67,90	64,73	50,77	51,54	64,62	65,38

Tabelle 41.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Aus dem COHFACE-<br/>Datensatz wurden die beiden Szenarien mit künstlicher Beleuchtung verwendet. Die Videos wurden in<br/>Zeitfenstern von 10 s analysiert.

	Datensatz	z: COHFA	CE (künst	liche Bele	uchtung)			
	Methode	: ICA						
ROI	Stirn &	Stirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
$MAE \downarrow$	9,30	9,60	12,53	10,90	9,00	8,82	8,36	8,13
RMSE ↓	14,76	16,06	17,33	17,80	11,93	12,32	13,50	13,12
Pearson ↑	0,3129	0,2917	0,0721	0,0899	0,3272	0,3277	0,3496	0,3802
SNR (dB) ↑	-4,62	-4,27	-6,49	-7,26	-6,69	-6,61	-6,41	-6,56
Acc (%) ↑	41,47	46,08	26,27	39,63	11,23	20,35	24,56	23,51

Tabelle 42.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Aus dem COHFACE-<br/>Datensatz wurden die beiden Szenarien mit seitlicher Sonnenlichtbestrahlung verwendet. Die Videos<br/>wurden in Zeitfenstern von 10 s analysiert. Es wird nur die rechte Gesichtshälfte von der Sonne beleuchtet.<br/>Die linke Gesichtshälfte liegt im Schatten (siehe Abbildung 9b). Die rPPG-Extraktion aus der ROI der<br/>rechten Wange weist die höchste Vorhersagegenauigkeit auf.

	Datensatz	Datensatz: COHFACE (natürliche Sonnenlichtbestrahlung)						
	Methode	Methode: ICA						
ROI	Stirn &	Stirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
MAE ↓	10,47	11,74	12,39	12,03	10,64	11,02	12,10	11,79
$RMSE\downarrow$	13,98	15,48	15,63	16,64	13,42	14,04	15,95	15,50
Pearson ↑	0,1606	0,1218	0,0199	0,0503	0,1108	0,0921	0,0913	0,1225
SNR (dB) ↑	-9,26	-9,79	-10,24	-9,32	-9,11	-9,10	-9,97	-9,92
Acc (%) ↑	18,75	17,5	6,25	21,88	6,72	9,70	7,46	7,46

**Tabelle 43.:** Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Als Datensatz wurde das Bewegungsszenario im VIPL-HR-V1-Datensatz bei einer Zeitfenstergröße von 10s verwendet. Die besten Evaluationsmetriken liefert die Begrenzung der ROI auf die Stirn- und Wangenregion.

besten Evaluationsmetriken herert die begrenzung der her dur die blahr und Hangemegroni								
	Datensatz	Datensatz: VIPL-HR-V1 Bewegungsszenario						
	Methode	: ICA						
ROI	Stirn &	Ctirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
MAE ↓	11,99	12,80	14,04	12,64	10,62	11,59	13,00	12,80
$RMSE \downarrow$	15,58	16,52	17,92	16,73	13,27	14,50	16,14	15,86
Pearson ↑	0,2779	0,2194	0,1384	0,2226	0,1237	0,0983	0,0755	0,1350
SNR (dB) ↑	-11,96	-12,71	-13,10	-12,06	-11,75	-12,51	-13,92	-13,84
Acc (%) ↑	17,31	15,63	14,42	16,58	6,07	6,07	9,54	8,67

Tabelle 44.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Als Datensatz wurde<br/>das Bewegungsszenario im PURE-Datensatz bei einer Zeitfenstergröße von 10s verwendet. Die besten<br/>Evaluationsmetriken liefert die Begrenzung der ROI auf die Stirn- und Wangenregion.

			-	-				
	Datensat	Datensatz: PURE Bewegungsszenario						
	Methode	Methode: ICA						
ROI	Stirn &	Ctirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
$MAE \downarrow$	0,60	0,95	8,09	7,23	4,65	4,04	3,36	1,81
$RMSE\downarrow$	1,88	2,34	22,48	18,43	11,33	10,14	11,71	7,91
Pearson ↑	0,9967	0,9946	0,32	0,6314	0,9405	0,9301	0,8558	0,9372
SNR (dB) ↑	8,85	7,31	1,71	-1,49	6,10	6,62	6,11	5,53
Acc (%) ↑	87,14	80,00	71,43	58,57	61,43	65,71	74,29	80,00

				-				
	Datensatz	Datensatz: MMPD Bewegungsszenario						
	Methode	Methode: ICA						
ROI	Stirn &	Ctimp	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
$MAE \downarrow$	23,17	22,94	23,92	22,78	22,22	22,62	22,97	22,65
$RMSE \downarrow$	31,12	30,29	31,65	30,55	28,69	29,47	30,17	30,17
Pearson ↑	-0,0158	0,0029	-0,0182	0,0688	0,0051	0,0054	-0,0149	0,0354
SNR (dB) ↑	-13,64	-13,90	-14,49	-13,75	-12,86	-13,51	-13,84	-13,42
Acc (%) ↑	23,48	18,94	20,08	21,59	6,14	8,33	8,55	11,40

Tabelle 45.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Als Datensatz wurden<br/>die Videoaufzeichnungen des ruhenden Probanden nach sportlicher Aktivität im MMPD-Datensatz<br/>verwendet. Die Größe des Zeitfensters beträgt 10 s.

Tabelle 46.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Als Datensatz wurde<br/>das Videospielszenario im RLAP-Datensatz bei einer Zeitfenstergröße von 10s verwendet.

	Datensatz	Datensatz: RLAP Videospielszenario						
	Methode	Methode: ICA						
ROI	Stirn &	Ctirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid
$MAE \downarrow$	3,04	4,29	5,17	5,08	4,49	4,42	4,57	4,17
$RMSE \downarrow$	8,81	10,24	10,94	10,08	7,49	7,65	9,42	8,93
Pearson ↑	0,8725	0,8310	0,8037	0,8333	0,6984	0,6902	0,6023	0,6321
SNR (dB) ↑	1,17	-1,77	-3,44	-3,70	-1,69	-1,61	-1,97	-1,95
Acc (%) ↑	81,18	69,69	64,81	63,07	35,78	40,09	52,59	55,17

Tabelle 47.: Auswirkung der Kombination von vorhergesagten Herzraten aus mehreren ROIs. Als Datensatz wurde<br/>das Rotationsszenario im KISMED-Datensatz bei einer Zeitfenstergröße von 10s verwendet.

	Datensatz: KISMED Rotationsszenario								
	Methode	Methode: ICA							
ROI	Stirn &	Stirn	Linke	Rechte	Gewichtete	Winkel-	Niedrigster	Mehrheits-	
Metrik	Wangen	Sum	Wange	Wange	Mittelung	gewichtung	Winkel	entscheid	
$MAE \downarrow$	4,86	6,21	10,31	12,66	7,25	7,27	5,68	6,50	
$RMSE \downarrow$	10,28	11,01	15,46	16,68	10,20	10,36	9,93	11,35	
Pearson ↑	0,6521	0,6652	0,2643	0,1988	0,6112	0,6090	0,6740	0,6392	
SNR (dB) ↑	-1,20	-3,69	-7,55	-10,45	-2,53	-2,54	-3,36	-3,60	
Acc (%) ↑	68,33	51,67	33,33	23,33	21,67	25,00	51,67	50,00	

## A.6. Bland-Altman-Diagramme der trainierten TS-CAN-Modelle (Ergänzung zu Kapitel 7.4)



Abbildung 34.: Bland-Altman-Diagramme bei der Evaluation der beiden mit dem PURE-Datensatz trainierten TS-CAN-Modelle am unbekannten UBFC-rPPG-Datensatz für (a) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 35.: Bland-Altman-Diagramme bei der Evaluation der beiden mit dem UBFC-rPPG-Datensatz trainierten TS-CAN-Modelle am unbekannten PURE-Datensatz für (a) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 36.: Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am unbekannten COHFACE-Datensatz für (a&c) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b&d) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 37.: Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am Bewegungsszenario des unbekannten VIPL-HR-V1-Datensatzes für (a&c) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b&d) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 38.: Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am Rotationsszenario des unbekannten MMPD-Datensatzes für (a&c) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b&d) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 39.: Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am Videospielszenario des unbekannten RLAP-rPPG-Datensatzes für (a&c) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b&d) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.



Abbildung 40.: Bland-Altman-Diagramme bei der Evaluation der vier trainierten TS-CAN-Modelle am Rotationsszenario des unbekannten KISMED-Datensatzes für (a&c) die vorimplementierte Datenvorverarbeitung der rPPG-Toolbox und (b&d) der selbst implementierten Gesichtsextraktion für den Bewegungszweig und ROI-Segmentierung für den Aufmerksamkeitszweig.

## Abkürzungsverzeichnis

BPM	Beats-per-minute (Herzschläge pro Minute)
BVP	Blutvolumenpuls
CAN	Convolutional Attention Network
CHROM	Chrominanzbasiertes rPPG-Verfahren
FFT	Fast-Fourier-Transformation
FPS	Frames per second (Videobilder pro Sekunde)
GUI	Graphical User Interface (Grafische Benutzeroberfläche)
HR	Herzrate
ICA	Independent Component Analysis (unabhängige Komponentenanalyse)
KISMED	Künstlich intelligente Systeme der Medizin
LGI	Local Group Invariance (Lokale Gruppeninvarianz)
LSTM	Long Short-Term Memory (Langes Kurzzeitgedächtnis)
MAE	Mean Absolute Error (Mittlerer absoluter Fehler)
MAPE	Mean Absolute Percentage Error (Mittlerer absoluter prozentualer Fehler)
MSE	Mean Squared Error (Mittlerer quadratischer Fehler)
NIR	Naher Infrarotbereich
POS	Plane Orthogonal to Skin
PPG	Photoplethysmographie
RMSE	Root Mean Square Error (Mittlerer Quadratischer Fehler)
RNN	Recurrent Neural Network (Rekurrentes neuronales Netz)
ROI	Region of Interest
rPPG	remote-Photoplethysmographie
SE	Standard Error (Standardfehler)
SNR	Signal-to-Noise Ratio (Signal-zu-Rausch-Verhältnis)
SpO2	Sauerstoffsättigung im Blut

TS-CAN	Temporal Shift Convolutional Attention Network
TSM	Temporal Shift Module (Zeitverschiebungsmodul)