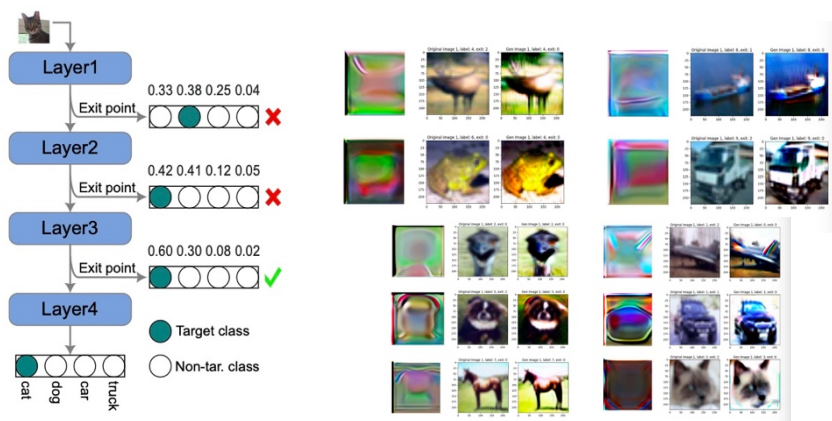


# Explainable AI for Efficient Inference on Hardware

Deep neural networks (DNNs) have achieved great breakthroughs in the past decades. However, DNNs require massive multiply-accumulate (MAC) operations and their execution on digital hardware, e.g., GPUs, causes enormous energy consumption. This poses critical risks to performance and energy sustainability in Artificial Intelligence (AI) computing systems. State-of-the-art solutions are still focusing on reducing the number of MAC operations in DNNs without exploring their explainability for efficient inference on hardware. As DNNs are deployed in a wide range of computing scenarios from edge devices to data centers, a new perspective to study DNN inference on hardware becomes necessary to meet various requirements of performance and energy efficiency. To address this challenge, this master thesis explores explainable AI approaches to enable image data to exit earlier in the inference of DNNs. With earlier-exit, efficient DNN acceleration on hardware can be achieved. This interdisciplinary research aims to develop new solutions for green AI to benefit environment.

In this thesis, first, early-exit strategy in CNNs as shown in the left figure as follows will be realized to enhance energy efficiency. Second, images are modified to enable earlier-exit in CNNs with explainable AI approaches as shown in the right figure as follows. Third, the modification techniques will be summarized to rules, so that such rules can be applied to any image data to enable earlier-exit in CNNs.



Left figure: The concept of early-exit in CNNs. Right figure: Modification of image data to enable earlier-exit in CNNs

References: <https://arxiv.org/pdf/2309.13443>

If you are interested in this topic for master thesis, please contact:

**Prof. Dr.-Ing. Li Zhang (grace.zhang@tu-darmstadt.de) with your CV and transcripts.**