Contribution ID: **27** Type: **not specified**

Enhancing Trust in AI Systems through Adaptive Image Quality Compensation

Perception is one of the main applications in which neural networks are far superior to conventional algorithms. One example is AI systems for automated driving, which can detect pedestrians based on image data and avoid them accordingly.
One problem with these systems is that their output depends heavily on the quality of the input images. For example, if an image is of inferior quality because it is heavily contaminated with noise or is too dark, accurate predictions are hardly feasible. In addition, different types of errors can occur that are of different relevance to the trustworthiness of the underlying system. For example, it may be more critical not to recognize an existing person than to recognize a person where there is none. We want to show that we can still avoid the most critical errors in situations of poor image quality. To do this, we want to compare two different approaches. In the first approach, we lower the networks' confidence threshold based on the estimated image quality to make them perceive things more cautiously in uncertain situations. In the second approach, we learn more cautious behavior directly during training by modifying the loss function by penalizing different types of errors depending on the image quality of the training data. We also aim to demonstrate that, in practice, combining these two approaches can yield the best results.
In summary, we will present a design strategy for AI-based systems that can deal with poor-quality input data without resorting to fallback solutions. In our example, this can be achieved by making a system react with varying degrees of caution. Such measures strengthen trust in AI-based systems and increase safety under unfavorable conditions.

Primary author: KEES, Yannick

Co-authors: Dr HOEMANN, Elena; Prof. KÖSTER, Frank; Dr HALLERBACH, Sven

Session Classification: Poster Session & Fingerfood