Abstract

Character recognition has been one of the most contemporary and challenging areas of research for decades. The past research on developing character recognition models using image processing and computer vision-based techniques shows that 'character recognition' is assumed to be a pure machine vision optimization problem. However, humans are still better at recognizing many characters especially distorted, ornamental or calligraphic characters as compared to the highly sophisticated recognition models. Although understanding the mechanism of character recognition by humans may give us some cues leading to better recognition abilities, the appropriate methodological approach of using these cues has not been much explored for developing character recognition models. Therefore, in this thesis, we propose to perform some cognitive experiments with humans to obtain some cues which will be used to develop intelligent character recognition models.

We start the investigation by understanding the reading behavior followed by distorted and ornamental character recognition through the analysis of the eye movement data. The experimental results show that humans direct their gaze to specific regions contributing to the character identity and recognize the characters using those key character regions. To highlight the character regions influencing the model's decision, we have proposed a 'modified Grad-CAM visualization technique', which is one of the contributions in the domain of explainable AI. Qualitative comparison between visualization maps and eye-fixation maps reveals that the deep learning model considered similar regions in character, which humans have fixated on, in the case of correctly classified characters. On the other hand, when the focused regions are different for humans and deep nets, the characters are typically misclassified by the latter. Both visualization maps and fixation maps are used to develop a visual explanation guided attention model (VEGAM). Furthermore, this thesis presents distinct attention models such as the deeply supervised and collaborative attention models that signify the importance of foveal and parafoveal vision in recognition. The experimental design, data analysis, and results presented in this thesis demonstrate the effectiveness of the proposed approaches.

Owing to a sophisticated vision system, vast visual experience, rich cognitive and perceptual processes, humans have a distinct edge over machines in the recognition task. The behavioral experiments using Bubble and proposed RISE (Random image structure evolution) stimuli emphasize the presence of some higher-level cognitive processes aiding successful recognition. In this era of AI, researchers are striving hard to build systems that can match human intelligence. One way to bridge the gap between machine and human intelligence is to design the artificial systems using cues derived from their natural counterpart, such as through cognitive experiments on human participants. This thesis, henceforth, motivates the need for substantial efforts in exploring various aspects of the cognitive processes to develop the next generation collaborative models for mimicking the human intelligence in AI-driven Engineering systems.