

Improving the reliability of CNNs for digital artwork analysis

Eric Postma

*Cognitive Science & AI department
Tilburg University, the Netherlands*

The use of convolutional neural networks (CNNs) for the digital analysis of artworks is now commonplace. Although CNNs achieve superior performance in a wide variety of visual tasks, their internal models are fallible which can give rise to erroneous predictions. CNNs misclassify adversarial instances (i.e., dataset instances that are intentionally distorted by small perturbations), which indicates that their successful performances on visual tasks are based on visual cues that are inferior to those of humans. Clearly, this limitation of CNNs hampers their applicability on the domain of digital artwork analysis. The presentation reports on our attempts to encourage CNNs to employ human-like visual cues. We show that more human-like cues give rise to improved predictions and strengthen the ability to deal with adversarial instances. With the improvement, CNNs become more suitable to support in the analysis of artworks.