What's in a scene? - Investigating generated scene information at different visual processing stages.

Aylin Kallmayer[1] & Melissa L.-H. Vo[1]

[1]*Scene Grammar Lab, Department of Psychology, Goethe University Frankfurt, 60323 Frankfurt am Main, Germany*

In recent years, deep-learning research has produced a range of neural networks that are able to classify and generate images from a variety of stimulus domains. At the same time, advances in cross-domain analysis methods (e.g., representational similarity analysis, RSA) have allowed us to make inferences about the human visual system based on the representational spaces of these networks. State-of-the-art generative networks provide a useful testbed for probing and manipulating representations at different stages of scene generation which could help us better understand "what makes a scene?". To investigate how well a generative network (PROGGAN) captures real-world scene gist and whether the information content of generated scenes is rich enough to trick human observers we conducted an online experiment. Participants had to detect real from generated images of scenes for varying presentation-times (50 vs. 500ms) and rate the degree of realism for each generated scene. We found that for short presentation times participants performed only slightly above chance, while for longer presentation times sensitivity (d') increased significantly and response bias became more conservative. Interestingly, realism ratings correlated with the false alarm rate of generated images in both conditions. Our results imply that the information contained in generated scenes triggers similar processes in the visual system during a first glimpse, but that generated scene information is perceived as less realistic with time. Based on these explicit and implicit measures of realism, we can use network dissection on the generator at different representational stages to better understand the content and structure of real-world scene representations in the human visual system.

Character count: 1780 (max 1800)