Use of Local Image Information in Depth Edge Classification by Humans and Neural Networks

Humans can use local cues to distinguish image edges caused by a depth change from other types of edges. [illegible] automatically label edges in images of natural scenes as depth or non-depth. We use this ground truth to identify the cues used by human observers and convolutional neural networks (CNNs) for edge classification. Eight observers viewed square image patches, each centered on an image edge, ranging in width from 0.6 to 2.4 degrees (8 to 32 pixels). Human judgments (depth/non-depth) were compared to responses of a CNN trained on the same task. Human performance improved with patch size (65%-74% correct) but remained well below CNN accuracy (82-86% correct). Agreement between humans and the CNN was above chance but lower than human-human agreement. Decision Variable Correlation (Sebastian & Geisler, in press) was used to evaluate the relationships between depth responses and local edge cues. Humans seem to rely primarily on contrast cues, specifically luminance contrast and red-green contrast across the edge. The CNN also relies on luminance contrast, but unlike humans it seems to make use of mean luminance and red-green intensity as well. These local luminance and color features provide valid cues for depth edge discrimination in natural scenes.

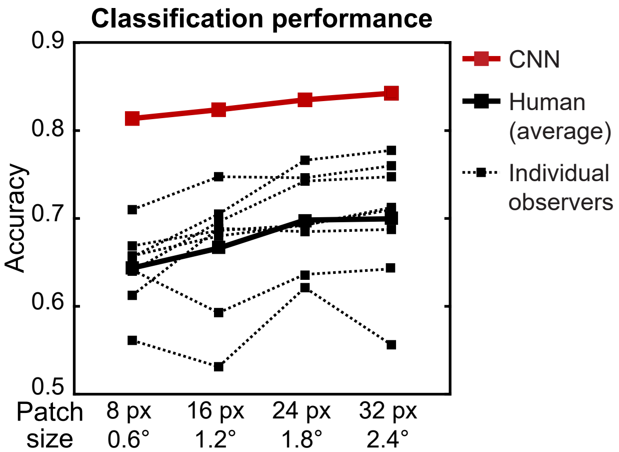

Figure 1. Performance of human observers and the CNN in the depth vs. non-depth edge classification task.

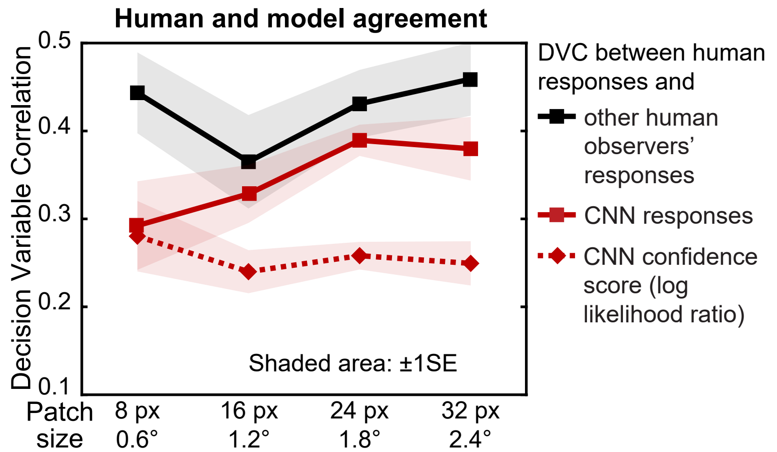
Human and model agreement
0.5
0.4
0.3
0.2
0.1
Decision Variable Correlation
DVC between human responses and
other human observers' responses
CNN responses
CNN confidence score (log likelihood ratio)
Shaded area: ±1SE
Patch size
8 px 0.6°
16 px 1.2°
24 px 1.8°
32 px 2.4°

Figure 2. Decision variable correlation (DVC) between individual observers' responses and responses of other humans (black line) or the CNN (red lines).

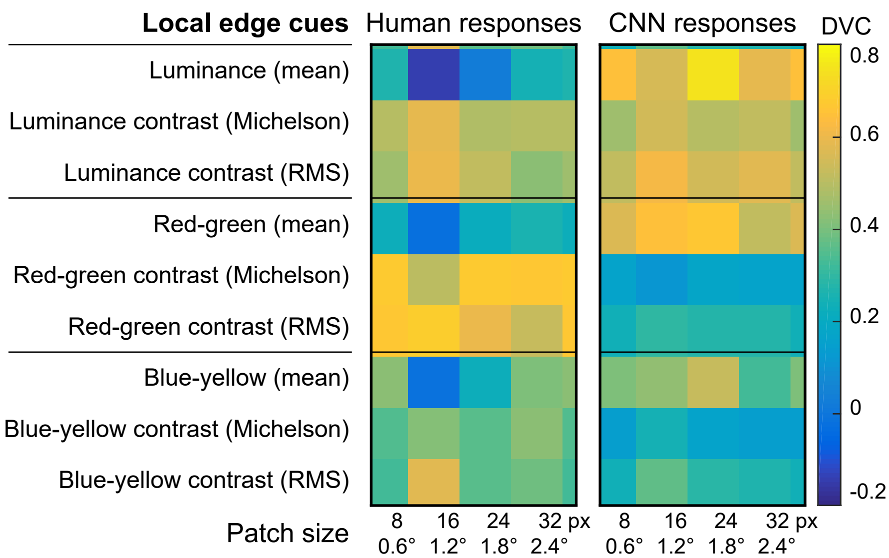
Local edge cues
Human responses
CNN responses
DVC
Luminance (mean)
Luminance contrast (Michelson)
Luminance contrast (RMS)
Red-green (mean)
Red-green contrast (Michelson)
Red-green contrast (RMS)
Blue-yellow (mean)
Blue-yellow contrast (Michelson)
Blue-yellow contrast (RMS)
0.8
0.6
0.4
0.2
0
-0.2
Patch size
8 0.6°
16 1.2°
24 1.8°
32 px 2.4°

Figure 3. Decision variable correlation (DVC) between the log likelihood of local edge cues and human "depth" responses (left panel) or CNN "depth" responses (right panel). Human responses are most strongly associated with luminance contrast and red-green contrast. CNN responses are most strongly associated with luminance contrast, mean luminance, and mean red-green values at the edge. In general, human observers seem to rely more on contrast cues than mean luminance/color when deciding whether an edge is due to a depth discontinuity, while the CNN seems to rely more on the mean luminance/color.

Adams, W.J., Elder, J.H., Graf, E.W., Leyland, J., Lugtigheid, A.J., & Muryy, A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports*, 6, 35805.
Sebastian, S. & Geisler, W. S. (in press). Decision-Variable Correlation. *Journal of Vision*.
Vilankar, K.P., Golden, J.R., Chandler, D.M., & Field, D.J. (2014). Local edge statistics provide information regarding occlusion and nonocclusion edges in natural scenes. *Journal of Vision*, 14, 13.