The Emergence of Semantics in Neural Network Representations of Visual Information: Supplementary Material

Dhanush Dharmaretnam and Alona Fyshe
University of Victoria
Department of Computer Science,
3800 Finnerty Rd, Victoria, Canada
[dhanushd|afyshe]@uvic.ca

Here we provide the architecture diagrams for each of the 3 CNNs we explored in the main text:

- Figure 1 shows **VGG 16** (Simonyan and Zisserman, 2014),
- Figure 2 shows **Inception V3** (Szegedy et al., 2015), and
- Figure 3 shows **ResNet 50** (He et al., 2015).

Each of these diagrams is annotated with colors corresponding to the 2 vs. 2 accuracy calculated with the hidden representation from that element of the architecture. Due the the large size of these CNNs, the figures are best viewed electronically at very high magnification.

Figure 4 shows the correlation between matched and mismatched vectors (as described in Section 2 of the main text). The correlation between mismatched vectors remains fairly stable through the layers of the network, whereas the correlation of matched vectors grows as the representation spaces converge to a similar space.

References

Kaiming He, Xiangyu Zhang, Shaoquing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1512.03385* pages 1–17.

Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint* pages 1–14.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the Inception Architecture for Computer Vision. *arXiv preprint* .

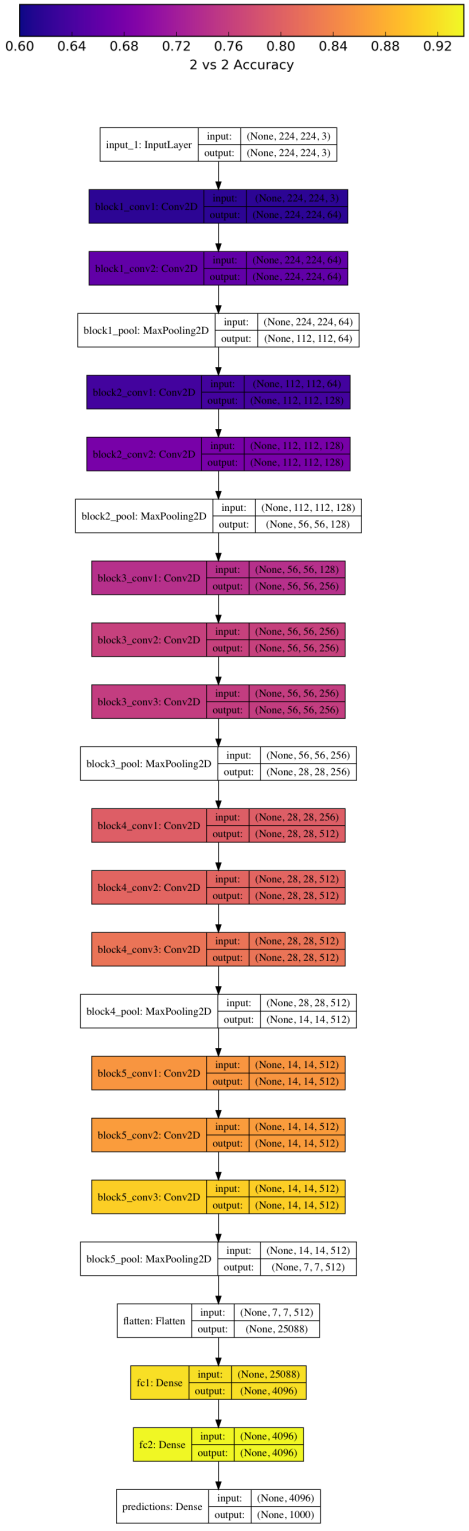

Figure 1: VGG 16 architecture. The type of node is annotated on the left of each block. On the right is the size of the input to that node, and the size of the resulting output. The nodes for which we computed 2 vs. 2 accuracy against SkipGram are colored according to the accuracy scale.

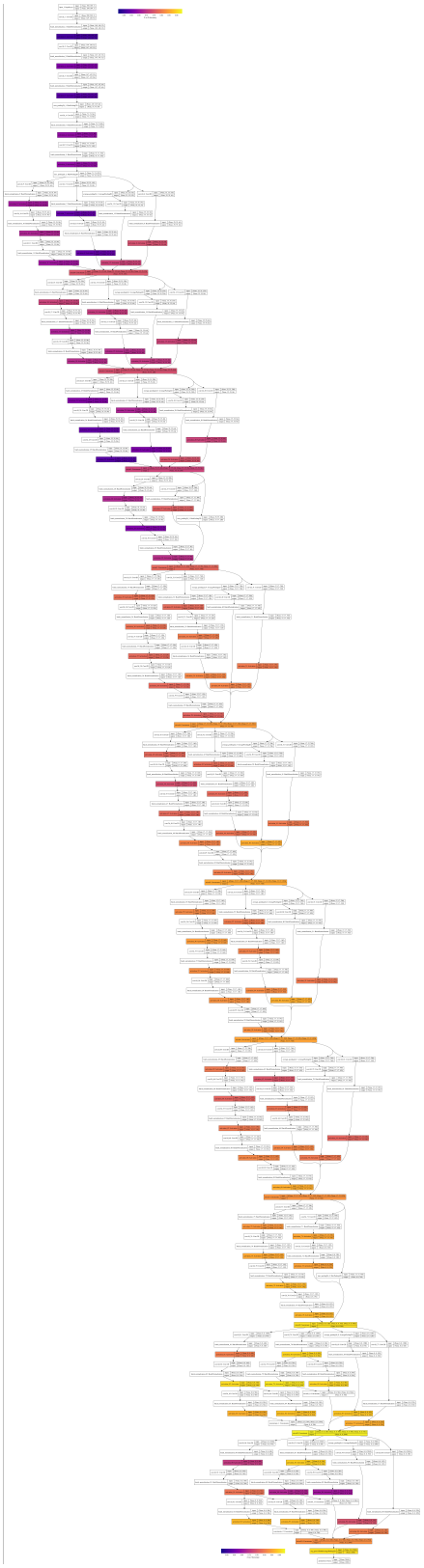

Figure 2: Inception V3 architecture. The type of node is annotated on the left of each block. On the right is the size of the input to that node, and the size of the resulting output. The nodes for which we computed 2 vs. 2 accuracy against SkipGram are colored according to the accuracy scale. In the main paper, only the mixed layers are plotted, but the accuracy of activation layers is shown here also.

Figure 3: ResNet 50 architecture. The type of node is annotated on the left of each block. On the right is the size of the input to that node, and the size of the resulting output. The nodes for which we computed 2 vs. 2 accuracy against SkipGram are colored according to the accuracy scale.

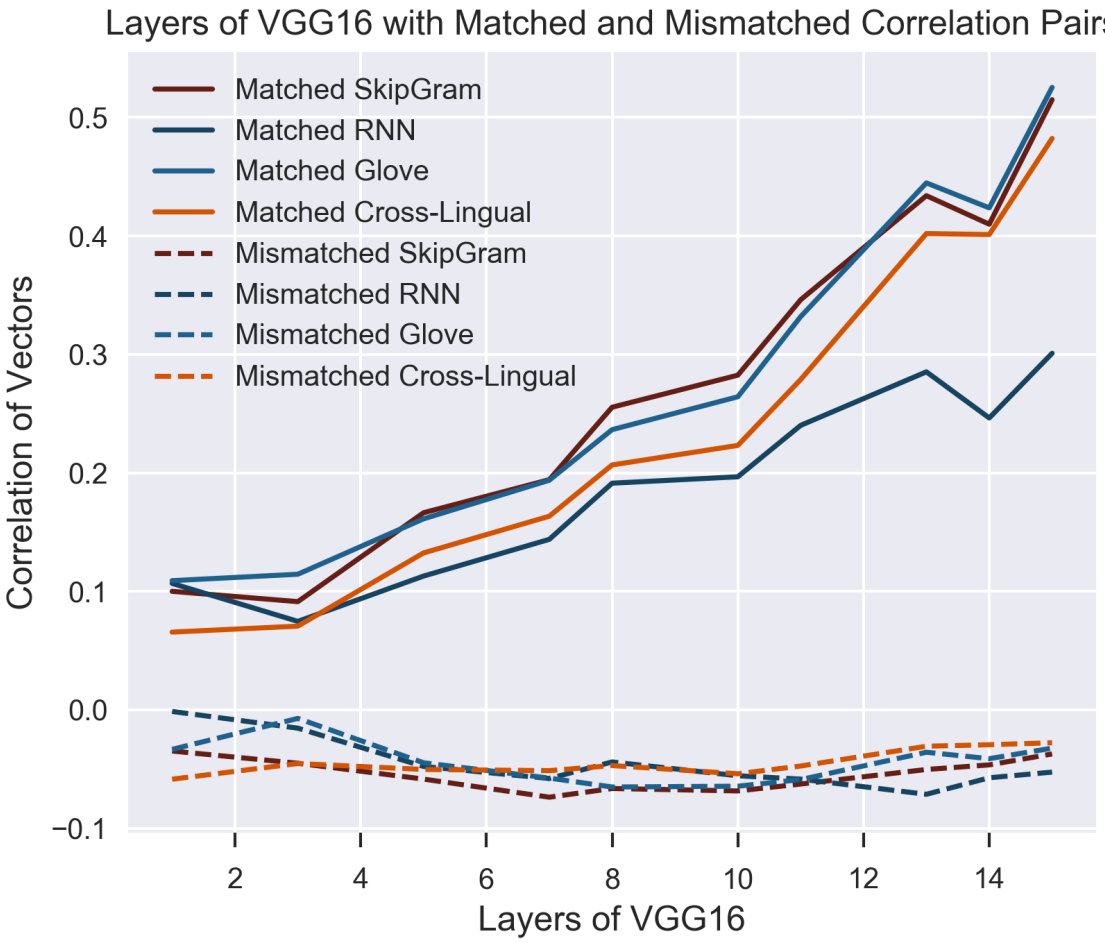

Figure 4: Comparison of the matched and mismatched vector correlations as a function of the depth in VGG 16. Note that the correlation of mismatched vectors remains largely unchanged, whereas the correlation of matched vectors grows with the depth of the network.