

Exploring the effects of linguistic labels on learned visual representations using convolutional neural networks

Seoyoung Ahn; Gregory Zelinsky; Gary Lupyan

+ Author Affiliations

Journal of Vision October 2020, Vol.20, 612. doi:<https://doi.org/10.1167/jov.20.11.612>

Abstract

Deep convolutional neural networks (CNNs) trained on large numbers of images are now capable of human-level visual recognition in some domains (Rawat and Wang, 2017). Analyses of the visual representations learned by CNNs show some resemblance to human visual representations (Kriegeskorte and Douglas, 2018), suggesting that these networks may offer a good model of human object recognition. Here, we use CNNs as a model of visual recognition for the purpose of exploring the effects of different types of labels on learning of visual categories. To what extent are labels necessary to distinguish between different kinds of tools, animals, and foods? One idea is that certain visual categories are so distinct that no guidance from labels is necessary. Alternatively, labels may help or even be necessary to discover certain types of categories. This question is difficult to answer in human learners because we cannot control their prior visual and semantic experience. We trained multiple CNNs on the same set of images while manipulating the labels they receive: none, basic-level labels (dog, hummingbird, hammer, van), superordinate labels (mammal, bird, tool, vehicle), or various combinations. We then correlated the model's choices on a triad task (given three images, select the one that is most different) with people's choices. The performance of unsupervised models was strongly dependent on low-level visual differences, highlighting the importance of labels to the training process. More surprisingly, the best performance was achieved by models trained using the coarser-grained superordinate labels (vehicle, tool, etc.) rather than basic-level labels, even when predicting triads where all three objects came from the same

This site uses cookies. By continuing to use our website, you are agreeing to [our privacy policy](#).

Accept

or from different superordinate categories (e.g., banana, a bee, and a screwdriver). The benefits of training with superordinate-level labels will be further discussed in the context of representational efficiency and generalizability.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

