The Role of Image Quality in Shaping Neural Network Representations and Performance

Jason Lee (jason.j.lee@vanderbilt.edu)
Computer Science Department, Vanderbilt University

Prince Owusu Nkrumah (prince.owusu-nkrumah@vanderbilt.edu)
Computer Science Department, Vanderbilt University

Stephen Chong Zhao (chong.zhao.1@vanderbilt.edu)
Data Science Institute, Vanderbilt University

William J. Quackenbush (william.quackenbush@vanderbilt.edu)
Vanderbilt Brain Institute, Vanderbilt University

Adaline Leong (jia.yin.leong@vanderbilt.edu)
Computer Science Department, Vanderbilt University

Trisha Mazumdar (trisha.mazumdar@vanderbilt.edu)
Computer Science Department, Vanderbilt University

Mark Wallace (mark.wallace@vanderbilt.edu)
Psychology Department, Vanderbilt University

David A. Tovar (david.tovar@vanderbilt.edu)
Psychology Department, Vanderbilt University

Abstract

Neural networks, recognized as robust models of the brain, depend on various factors including architectures, training data, algorithms, and objective functions. This study explores the influence of image quality in training data on the representation and performance of neural networks, and consequently, on their capability to model brain functions. The "visual diet"—the quality and variety of images—present in training sets such as ImageNet and EcoSet, can significantly affect how these models learn and perform across different classes. By examining how variations in image quality impact the networks' internal representations and overall performance, we aim to better understand how training data affects the correspondence between neural network models and the brain's processing mechanisms. Our main finding is that high-quality training data varies on the performance metric. A diverse range of image quality in the training set produces the most expansive representational spaces. However, the highest performance in terms of top 1 and top 5 is biased slightly more towards images with high image quality.

Keywords: Image Quality; Representations; Brain Models; Neural Networks

Introduction

The composition of training datasets for neural networks is a key area of interest in the field of artificial intelligence and cognitive neuroscience. One component of a visual training set is the overall quality of the images used to train the network. Image quality metrics are generally differentiated by whether the metric utilizes a reference. Reference metrics, such as peak Signal-to-Noise Ratio (pSNR) and Structural Similarity Index Measure (SSIM), compare an image against some representation of an "original" to determine the distortion of the input (Sara, Akter, & Uddin, 2019). By contrast, no-reference metrics, including Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Natural Image Quality Evaluator (NIQE), do not use a reference image and instead use the features of the input image itself to determine its quality (Mittal, Moorthy, & Bovik, 2012; Mittal, Soundararajan, & Bovik, 2013). Given that the training sets we are evaluating do not have a reference, we focused our image quality metrics to no-reference metrics.

Our interest in neural networks is due to their ability to capture various aspects of human vision (Cadieu et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Yamins & DiCarlo, 2016). Traditionally, many of these networks have been trained using ImageNet. However, studies suggest that manipulating the training sets can significantly enhance their correspondence with brain activity. For instance, networks trained on more ecologically valid categories (Mehrer, Spoerer, Jones, Kriegeskorte, & Kietzmann, 2021), those that mirror developmental trajectories (Avbersek, Zeman, & Op de Beeck, 2021), or use images from cameras mounted on infants, have been found to enhance brain correspondence, network performance, and generalization capabilities (Bambach, Crandall, Smith, & Yu, 2018; Vogelsang et al., 2018).

For performance we are measuring the neural network's ability to classify a given input correctly. Researchers often use "top-five" and "top-one" performance indicators, which measure how often the correct label appears in the best five and one predictions of the model, respectively, to rank network performance (Krizhevsky, Sutskever, & Hinton, 2012). However, these metrics can only show the final result of a network's calculations, and provide no insight into the actual relationships the network has learned through training. To see these representations, we extract the last fully connected layer of the model as an embedding and compare it to embeddings taken while processing other data inputs.

Methods

Dataset and Metric Choice

The goal of this study was to thoroughly explore the effects of differences in image quality on the performance of image classifier neural networks. Thus, we used the EcoSet and ImageNet datasets, both often used to train image classifier models, as well as the four commonly used neural network

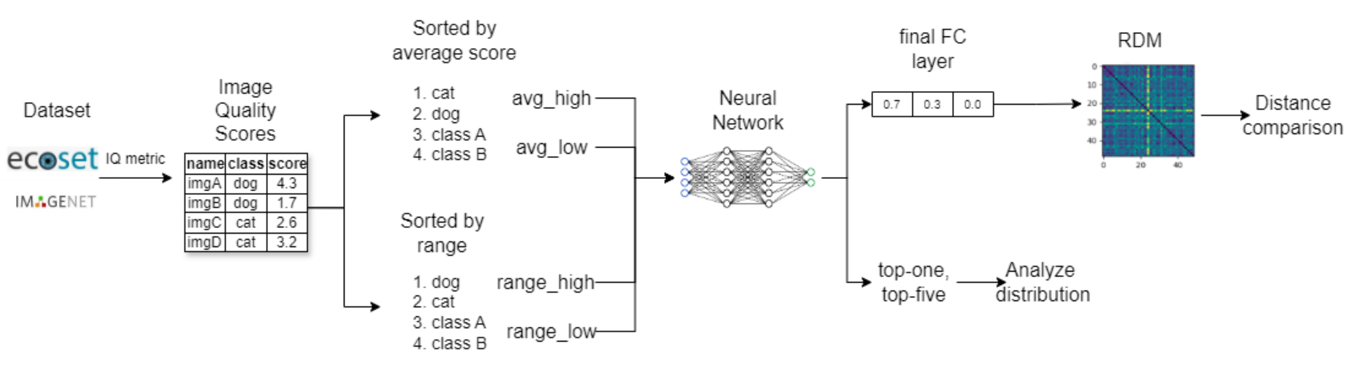

Figure 1: Experiment Schematic and Analysis

architectures, being AlexNet, Inception-v3, ResNet-50, and VGG16 (Mehrer et al., 2021). We also took differences between image quality (IQ) metrics into account, and chose to evaluate training set images using BRISQUE, CLIP-IQA, CNN-IQA, DBCNN, MUSIQ, and NIQE, due to their demonstrated correlation with brain data.

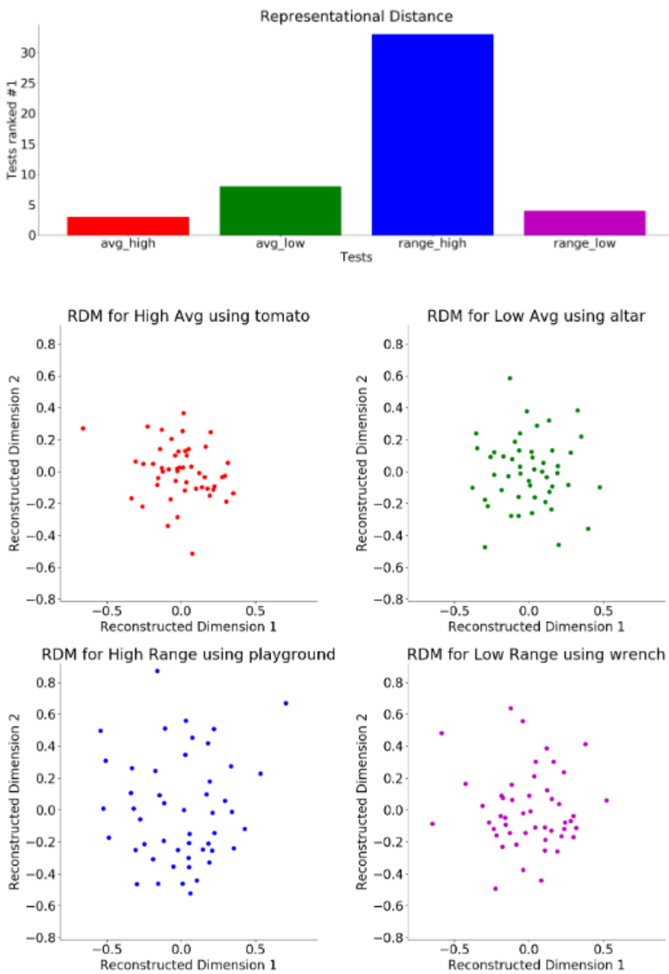

Figure 2: Outcomes of Fully Connected layer comparison. Bar graph (top) represents representational distances with MDS of sample categories (bottom)

Quality Ranges

After applying all aforementioned metrics to all training images in EcoSet and ImageNet, we sorted the categories within each dataset by the average IQ score achieved as well as the IQ score range present within the category's training set. This allowed us to investigate four potential candidates for an ideal training dataset: the top and bottom 5% of categories when sorted by average IQ score and by IQ range, labeled "average high", "average low", "range high", and "range low".

We inspected each model's last fully connected layer, which serves as an embedding prior to classification, and can provide insights into the distinctions the model makes, as well as the top5 and top1 accuracy scores.

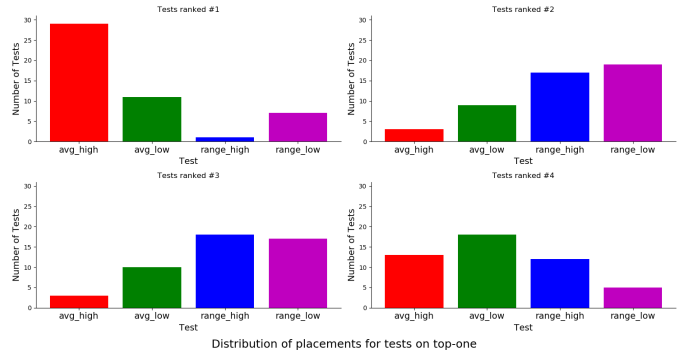

Figure 3: Outcomes of Top-one performance. Each plot represents the distribution of tests when compared against each other. Rank 1 corresponds to the test with the best top-one accuracy score for the given combination of dataset, network architecture, and metric, and similarly for rank 2, 3, and 4.

Results

Interestingly, differing results were seen in comparing the model's embeddings of test images and comparing model performance itself. "Range_high" performed overwhelmingly well for embedding distance covered; when comparing the total distance covered in the embedding for each of the four candidate categories, "Range_high" had the highest distance measure in 33 of 48 total dataset-architecture-metric combinations.

For top-one and top-five performance indicators, we measured the accuracy of the model across all image categories within the test case before comparing the result to other test cases used on the same dataset, network architecture, and IQ metric. We found that results differed not only with those from the FC layer analysis but also between datasets. Of the models used on EcoSet, categories with high average image quality scored highest on both top-one (16 of 24 architecture/metric pairs) and top-five indicators (16 of 24), with range_high categories usually achieving second (17 and 19 of 24, respectively). On ImageNet, however, avg_high had the best top-one performance, with range_low scoring second 16 times in top-one and 19 times in top-five. Range_high's ranked consistently third and fourth place, with a 13-11 ratio for top-one and an even 12-12 split for top-five.

Discussion and Future Directions

Our results show that image quality can affect both the performance and internal representation of a neural network, albeit in different ways. The next steps are to understand how these difference affect neural network correspondence with the brain as well as probe more causal relationships. We specifically aim to manipulate the image quality training data – such as increasing the quality of images in categories marked low-quality or decreasing the quality of images in categories scoring high or manipulating the ranges. We will then retest how these changes affect internal representations and performance, as well as correspondence with brain representations.

Acknowledgments

This work was supported by Vanderbilt's IDD-Reads Grant

References

Avbersek, L. K., Zeman, A., & Op de Beeck, H. (2021). Training for object recognition with increasing spatial frequency: A comparison of deep learning with human vision. *bioRxiv*.

Bambach, S., Crandall, D. J., Smith, L. B., & Yu, C. (2018). Toddler-inspired visual object learning. In *Advances in neural information processing systems* (Vol. 2018-Decem, pp. 1201–1210).

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., ... DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Biology*, *10*(12).

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology*, *10*(11).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. doi: 10.1145/3065386

Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N., & Kietzmann, T. C. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences*, *118*(8), e2011417118. doi: 10.1073/pnas.2011417118

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, *21*(12), 4695–4708. doi: 10.1109/tip.2012.2214050

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212. doi: 10.1109/lsp.2012.2227726

Sara, U., Akter, M., & Uddin, M. S. (2019). Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, *7*(3), 8–18. doi: 10.4236/jcc.2019.73002

Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), 11333–11338.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.