Semantic Cognition in Dense Convolutional Networks

Apurva Bhargava (ab8687@nyu.edu)

DSGA-1016 Computational Cognitive Modeling

New York University

Abstract

Convolutional neural network-based architectures are now the state-of-the-art solutions for human cognition tasks, often performing at the same level as or even better than humans. These architectures were inspired by the human biological neural systems. In this paper, I explore the similarity between the two by studying the pattern of learning (differentiating) and forgetting (dementia) with object recognition on CIFAR-100 dataset as the task and DenseNet-BC as the model. I also explore category typicality and effect of distortion using class ranking correlations.

Keywords: semantic cognition, differentiation, learning, semantic dementia, conv-nets, category typicality

Introduction

Cognitive functions are brain-based skills that are required to carry out any task, simple or complex, and are related with the mechanisms of learning, memorization, problem-solving, paying attention, recognition, etc. (Zhang, 2019). Semantic cognition refers to the ability to use, manipulate and generalize knowledge that is acquired over the lifespan to support innumerable verbal and non-verbal behaviours. It relies on two principal interacting neural systems: semantic representation and control processes (Ralph, Jefferies, Patterson, & Rogers, 2017). Cognitive processes in humans can be explained by creating such a representation and control mechanism using computational models. One of the most popular approaches in the artificial intelligence and machine learning paradigms, neural networks, are inspired by the way biological neural systems, in particular, the brain, process data. The artificial neural networks are, thus, a suitable candidate for modeling cognitive tasks. Such explorations allow me to explain the neurological processes in terms of some computational functions and validate them, as well as provide insight into the complexities of the human behaviour. In this report, I model object recognition, a task enabled by semantic cognition, using convolutional neural networks. I will explore the process of coarse-to-fine learning and progressing dementia, and also look at the results of small-scale human-similarity tests conducted on category typicality and response to noisy stimuli.

Dataset and Model Selection

In order to have enough classes for the coarse-to-fine differentiation study as well as to keep the dataset size manageable for computation, I chose the CIFAR-100 dataset (Krizhevsky, 2009). This dataset has 100 classes containing 600 images each. There are 500 training images and 100 testing images per class. The 100 classes in the CIFAR-100 are grouped into 20 superclasses.

I used Pytorch implementation of Dense Convolutional Network or DenseNet-BC with layers (L) = 100 and growth rate (k) =12 from (Yang, n.d.) for the classification task. My selection was driven by the compactness of the model (only 0.8 million parameters) and its error rate of only 22.88% on CIFAR-100. The models with lower error rate (ResNeXt-29 8x64, DenseNet-BC L=190 k=40 and WRN-28-10) had over 34 million parameters and the other models with reasonable size (ResNet-110, PreResNet-110, AlexNet) had higher error rates. The benchmarks I consulted are also from (Yang, n.d.). I achieved a validation accuracy of 77.1% on CIFAR-100 after training my model for 154 epochs (Figure 1).

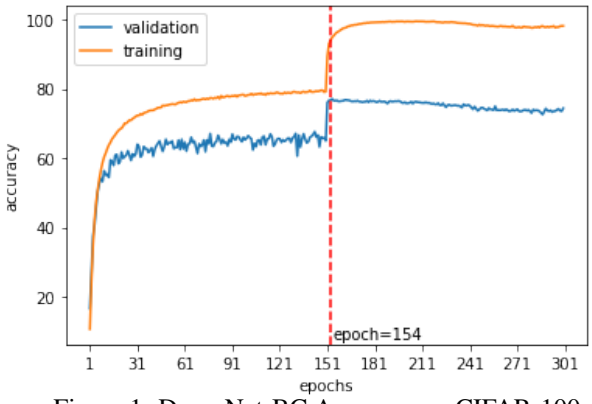

Figure 1: DenseNet-BC Accuracy on CIFAR-100

DenseNet-BC architecture: DenseNet-BC is a network architecture where each layer is directly connected to every other layer in the same dense block in a feed-forward fashion (see Figure 2). Unlike traditional convolutional networks where L layers have L connections (one between each layer and its subsequent layer), DenseNet has L(L+1)/2 direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers (Huang, Liu, & Weinberger, 2017).

ConvNets and the Human Vision

The visual cortex is the primary cortical region of the brain that receives, integrates, and processes visual information relayed from the retinas. It is divided into five different areas (V1 to V5) based on function and structure. The complexity of the visual data processing increases for each region as the visual information gets passed along. Simple cells, which are found mostly in V1, respond to specific types of visual cues such as the orientation of edges and lines. Complex

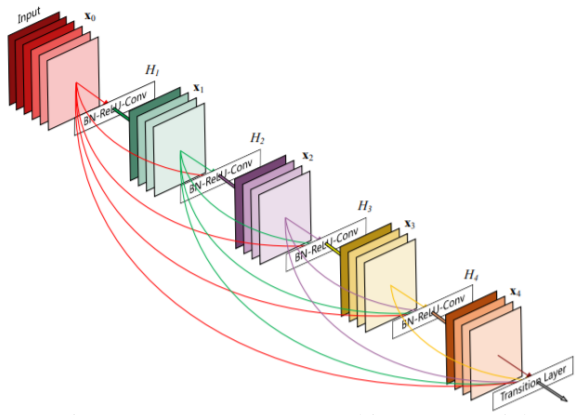

Figure 2: DenseNet-BC Architecture, L=5, k=4

(Huang et al., 2017)

cells, which occur in the next areas, respond to the summation of several receptive fields that become integrated from many simple cells, these may be more complex outlines of objects, textures, etc.(Huff, Mahabadi, & Tadi, 2021) Convolutional neural networks work similarly as they also use restricted receptive fields, and a hierarchy of layers which progressively extract more and more abstracted features (Kheradpisheh, Ghodrati, Ganjtabesh, & Masquelier, 2016).

I verified this by taking an image of a 'snail' from the CIFAR-100 testset and generating feature maps from the learned weights of successive convolution layers. As seen in Figures 3, 4 and 5, the initial convolutional layers learn information about the edges and contours and further layers learn more detailed representations.

Why use convolutional neural networks for studying dynamics of differentiation and dementia? As discussed above, convolutional neural networks learn representations that are apt for describing the information obtained by the visual cortex and passed to neurons for further processing, which could be likened to the dense layers in the DenseNet-BC. Thus, this is a suitable model for studying semantic cognition in vision related tasks.

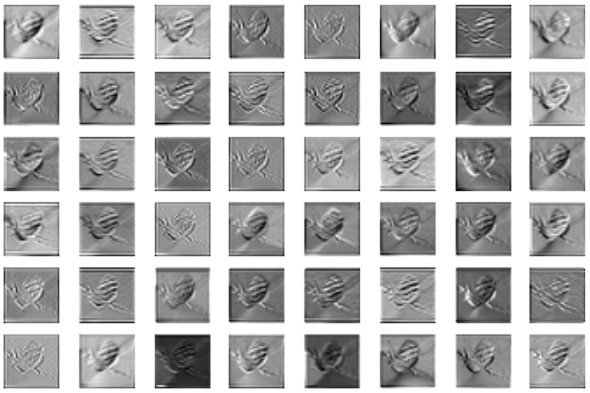

Figure 3: Feature maps for conv-layer 6

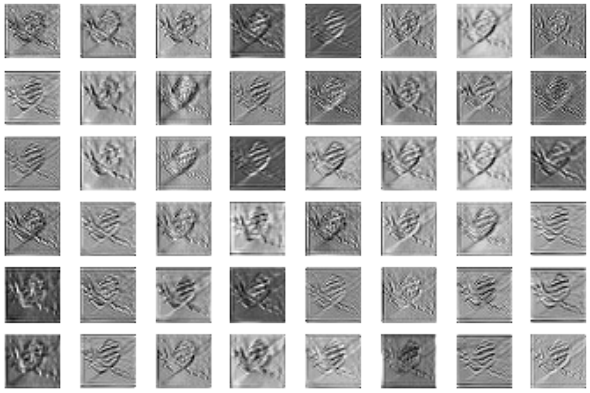

Figure 4: Feature maps for conv-layer 14

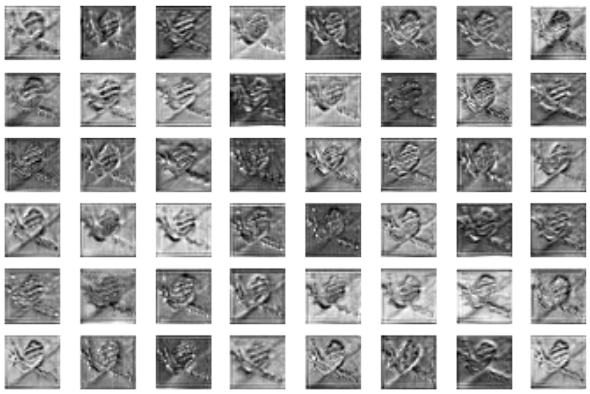

Figure 5: Feature maps for conv-layer 26

Dynamics of Differentiation in DenseNet-BC

Analysis

In order to study the dynamics of differentiation in the object recognition problem, I create some hierarchies within the 100 classes based on visual and semantic similarity. Here, the actual classes will be the finest or most-specific level of categorization, while the groupings of these classes or the groupings of these groupings of classes will be the coarser or more general level of categorization. At five stages of learning (or 5 different epochs), for all examples of a given class, I find the number of predictions that are accurate (predicting the correct) and the number of predictions that correspond to the coarser groups. I count out the completely irrelevant misclassification, attributing it to model limitation, and then compute the proportions of the predictions corresponding to the different specificity levels of the hierarchy and plot them. The hierarchies considered are given in Figure 6 as trees.

Results

In Figures 7, 8, 9 and 10, the green coloured-bars correspond to the highest level of specificity or the leaves in the hierarchy tree diagrams. The light and dark blue bars show the coarser levels or most general classes. Initially, the model makes more general predictions (either the ground truth label or a similar class). As the training epochs increase, i.e., the level of learning increases, the model makes more and more

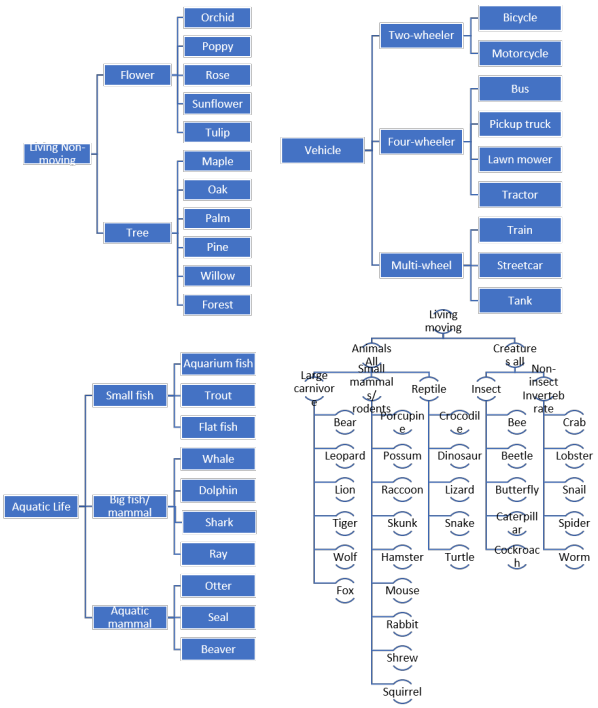

Figure 6: Hierarchy tree diagrams

specific predictions, increasing height of green bars (highest specificity level) and decreasing the heights of the blue bars in the plots. Thus, the model differentiates in a coarse to fine fashion.

I also visualized the learning in an alternative way, by taking the final sequential layer's weights associated with every class (dimensions: 324×100) and clustering them to obtain dendrograms depicting the learning of characteristics specific to sub-classes during 3 stages of the training (Figure 9, 10 and 11).

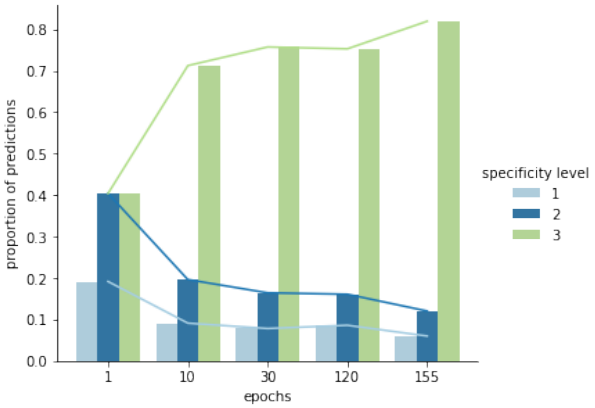

Figure 7: General to specific learning of aquatic animals

Semantic Dementia in DenseNet-BC

Analysis

I modeled dementia in DenseNet-BC through two separate ways– (1) addition of Gaussian noise to a percentage of ran-

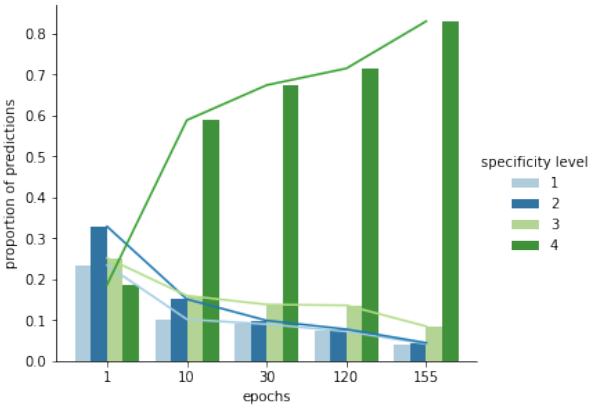

Figure 8: General to specific learning of animals

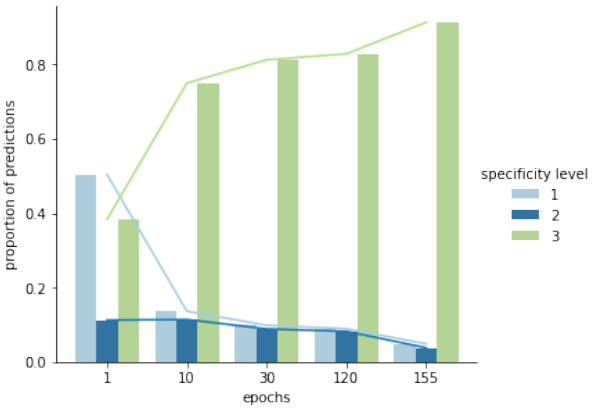

Figure 9: General to specific learning of vehicles

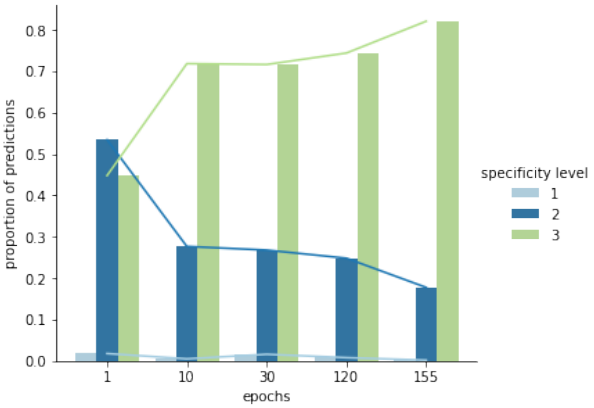

Figure 10: General to specific learning of plant-life

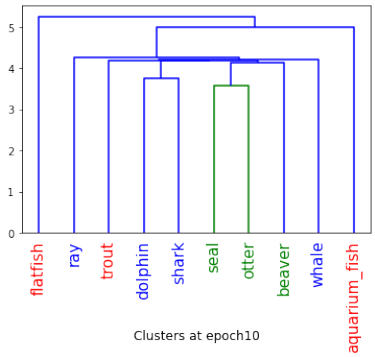

Figure 11: Dendrogram: Clustering using sequential layer

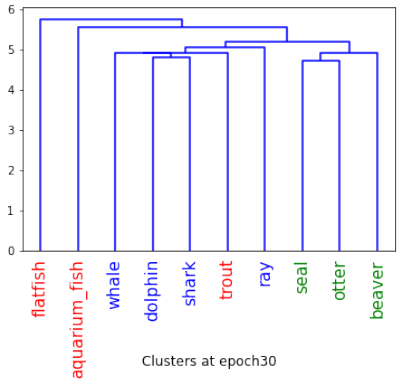

Figure 12: Dendrogram: Clustering using sequential layer

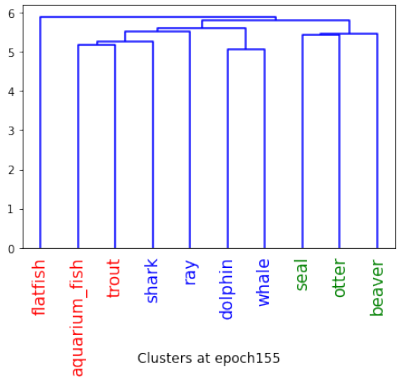

Figure 13: Dendrogram: Clustering using sequential layer

dom weights of the convolutional layer, and (2) masking of a percentage of weights of the convolutional layer randomly by setting their values to zero. In both cases, the weights to be perturbed or masked were chosen randomly through a Bernoulli distribution with some probability p. This probability p was used as the setting to increase or decrease noise level. To portray the progression of dementia, the value of p is increased iteratively, leading to increasing drop in the model's performance. Similar to the previous (differentiation) analysis, at each stage of progression, the proportions of the predictions for specific classes and the more generalized classes were recorded.

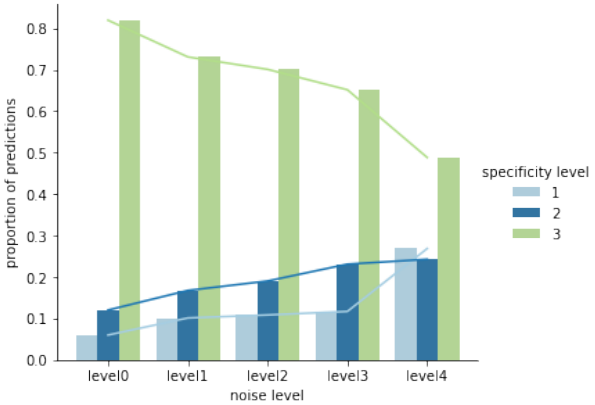

Figure 14: Losing specificity for aquatic life

Results

Figures 14, 15, 16 and 17 are for added Gaussian noise method and Figures 19 and 20 depict the plots for masked weights method. In all cases, it can be observed that the number of correct class predictions decrease as more weights are perturbed or masked. The predictions start becoming more general, as shown by the increasing height of the blue bars that correspond to coarser classes (model predicts similar classes but not the exact ground truth). For additive noise and masking, p value sets used were 0, 0.15, 0.25, 0.35, 0.45 and 0, 0.025, 0.05, 0.075, 0.1 respectively.

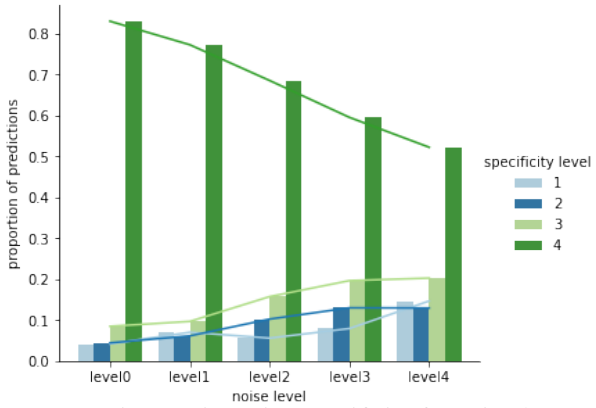

Figure 15: Losing specificity for animals

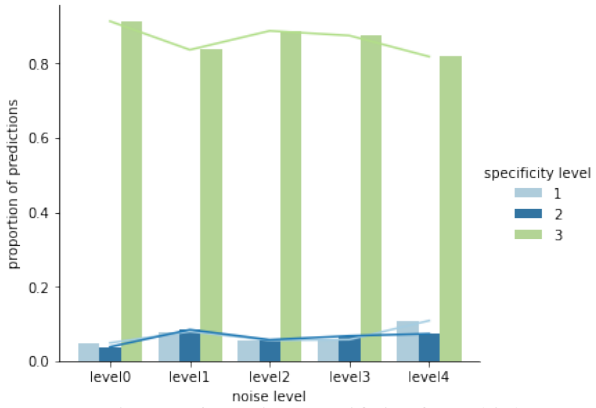

Figure 16: Losing specificity for vehicles

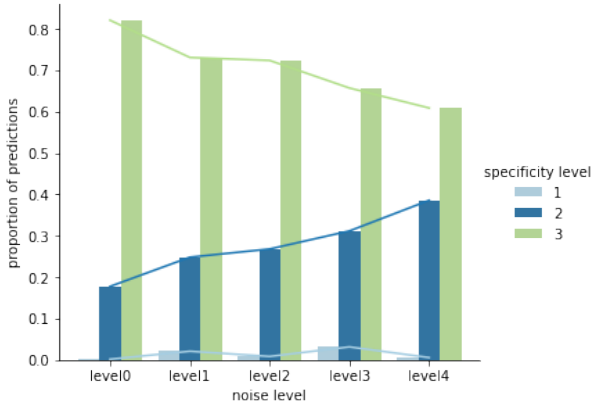

Figure 17: Losing specificity for plants

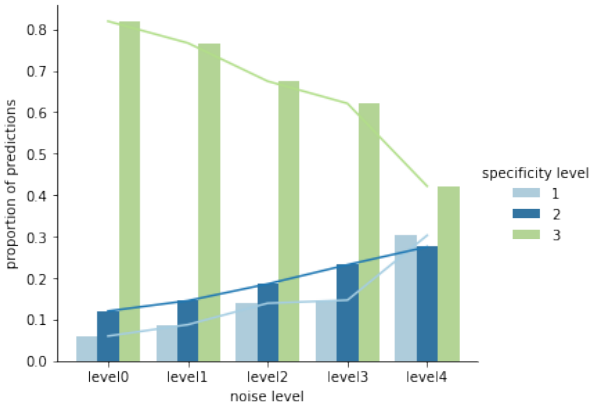

Figure 18: Masking: Losing specificity for aquatic life

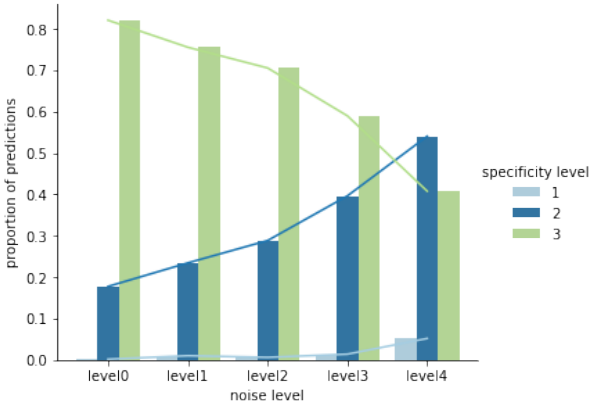

Figure 19: Masking: Losing specificity for plants

Category Typicality

Analysis

In object classification, the typicality effect describes a phenomenon whereby typical items are more easily judged as members of a category than atypical items (Wang et al., 2016). This effect has been studied in the context of category verification. In the context of my work, I can use this to analyse whether the dataset is aligned with human perspective and if it over-represents or under-represents typical examples in certain categories, thereby causing the model to learn those biases during training. I selected 5 images per class for 8 classes and asked 12 people to rank them by how representative or typical the images were for the given class. I computed the model's rank for those images by using the probability of prediction for that class. Finally, I computed the correlation between averaged human ranks and model ranks, inspired by (Lake, Zaremba, Fergus, & Gureckis, 2015). The chosen 8 classes and 5 images are illustrated in Figure 20.

Table 1: Category Typicality Rank Correlations

Class name	Rank Corr.	Class name	Rank corr.
Lamp	0.9892	Mountain	0.6502
Bowl	0.9546	Mushroom	0.8976
Wolf	0.7844	Skyscraper	0.2124
Shark	0.6981	Bus	-0.2759

Results

The results are summarized in Table 1. While the numbers of data samples, chosen classes and participants are not large enough to reach definitive conclusions, but certain inferences can be drawn while maintaining the specificity of the examples as an important consideration. It was observed that in case of inanimate common household-objects, such as lamp and bowl, the rank correlations are higher than 0.95. These images are clear with mostly neutral backgrounds. The machine-learning model and human rankings also see a good match for wolf and shark wherein it is possible to easily tag one image more typical than the other. The correlation is also good for mountain and mushroom, which are images within a natural setting. The rank correlations drop for man-made objects such as skyscrapers (0.21) and buses, where the correlation is negative. This could possibly be attributed to the variety of images in those classes in the dataset, and the disagreement between human and the model on what feature is important. Humans also have subjective opinions on what is the most appropriate exemplar.

These results can be attributed to either the representatives of a class within a dataset, the size of the dataset, the features or representation that the model deems important or even the lack of objectivity in human participants.

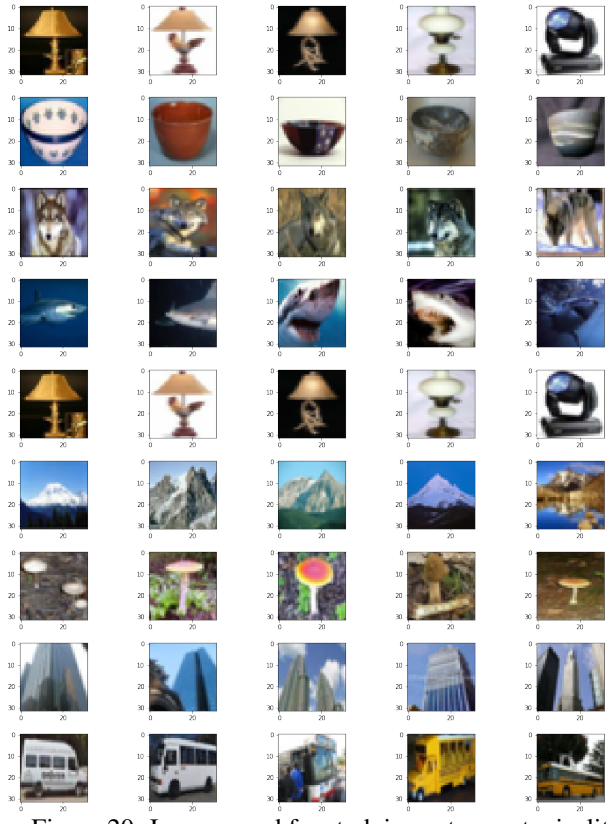
Figure 20: Images used for studying category typicality

Table 2: Distortion Study Rank Correlations

Noise:-	level 1	level2	level3
Random noise	0.8784	0.5623	0.2322
Blur	0.8312	0.8021	0.3200
Contrast	0.3544	0.4084	0.4230

Response to Stimulus Distortion

Analysis

Studying the model versus human response to noisy signal allows me to do a comparative study of the rigid learned representations of a ML model versus the flexible human perception guided by experience, based on the types and levels of distortion. This study is partly inspired by (Geirhos et al., 2017). I introduce different types of noises or distortions, namely, random noise, blur and contrast at three different levels of intensities into images. I selected 4 images per noise level per distortion type, i.e., a total of 36 images. For every image, top 5 prediction classes were generated using the clean (original) version. Then, the model was used for making a prediction on the noisy image, the probabilities for the 5 original classes were used for model's ranks. Four human participants were asked to rank the 5 classes in order of suitability to the image. For every distortion and noise level, the correlations of rankings of humans' average and model were computed and then averaged over the 4 images. The distortion type and levels are illustrated using example images in Figure 21.

Results

The results are summarized in Table 2. It was observed that for random noise and blur, an increase in distortion level resulted in lower correlation with the response of the model. This could mean that the model and humans had dissimilar perception under the application of distortion. Humans could perhaps still identify the object when the model failed. This does not hold true for contrast. This may be attributed to the corruption of image beyond either human perception, leading to chance matching of rankings.

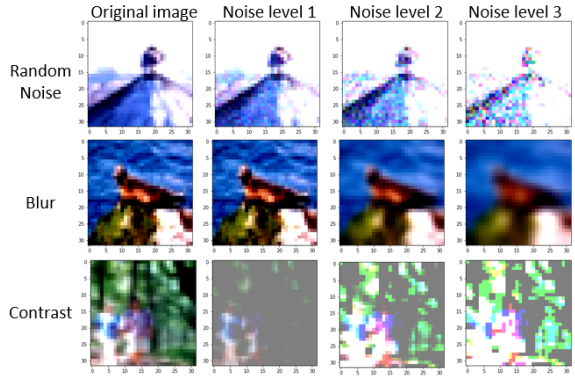

Figure 21: Distortion Types and Levels

Conclusions

In this project, I compared deep representations of objects learned by Dense Convolutional Network for object classification to human neural representations. I identified that the model starts from general (coarse) information and learns increasingly specific characteristics to identify the object as the training progresses. On increased addition of noise, the model forgets the most specific differences first, leading to prediction of similar classes but not the exact ground truth class. These are both similar to the processes of differentiating and dementia in humans. It should be noted that instead of just the first prediction, all class probabilities can be used for a more meticulous analysis. Finally, I learned that the model and human agree on some generic objects being the typical example of their class, but disagree where more subjective factors come into play.

References

Geirhos, R., Janssen, D. H. J., Schütt, H. H., Rauber, J., Bethge, M., & Wichmann, F. (2017). Comparing deep neural networks against humans: object recognition when the signal gets weaker. *ArXiv*, *abs/1706.06969*.

Huang, G., Liu, Z., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269.

Huff, Mahabadi, & Tadi. (2021). Neuroanatomy, visual cortex. *StatPearls [https://www.ncbi.nlm.nih.gov/books/NBK482504/]. Treasure Island (FL): StatPearls*.

Kheradpisheh, S. R., Ghodrati, M., Ganjtabesh, M., & Masquelier, T. (2016). Deep networks can resemble human feed-forward vision in invariant object recognition. *Scientific Reports*, *6*.

Krizhevsky, A. (2009). Learning multiple layers of features from tiny images..

Lake, B., Zaremba, W., Fergus, R., & Gureckis, T. (2015). Deep neural networks predict category typicality ratings for images. In R. Dale et al. (Eds.), *Proceedings of the 37th annual conference of the cognitive science society.* Cognitive Science Society.

Ralph, M. L., Jefferies, E., Patterson, K., & Rogers, T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42-55.

Wang, X., Tao, Y., Tempel, T., Xu, Y., Li, S., Tian, Y., & Li, H. (2016). Categorization method affects the typicality effect: Erp evidence from a category-inference task. *Frontiers in Psychology*, *7*.

Yang, W. (n.d.). *Classification on cifar-10/100 and imagenet with pytorch.* https://pythonawesome.com/classification-on-cifar-10-100-and-imagenet-with-pytorch/. (Accessed: 2021-04-10)

Zhang, J. (2019). Cognitive functions of the brain: Perception, attention and memory. *ArXiv*, *abs/1907.02863*.