Correlating Brains and DNNs for Images and Colors

Ayushi Arora, Deeksha Vijay, Diksha Banka, Dravya Marwaha, Gaurvika Kapoor, Praveen Prabhat, Twinkle Arora[1]

[1]IIT Kanpur

July 28, 2021

Objective and Overview

Convolutional neural networks (CNNs) have made great advances in the computer vision fields. In this analysis, we study representation of colors in CNNs and in brains using the Python toolbox DNNBrain. We started with a literature survey, studying papers on how colors are perceived by humans and macaque brains. Then we moved to the DNNBrain toolbox, using its functionality to study how CNNs encode images as well as colors. We carry out a number of experiments (provided by DNNBrain) to do this. Finally, we carry out a classification task on color classes. Here we analyse two methods:

1. Transfer learning
2. Learning from scratch

Our hypothesis is that transfer learning will give better classification accuracy since features are extracted from CNNs which already contain some information of the input picture, as seen from the experiments carried out.

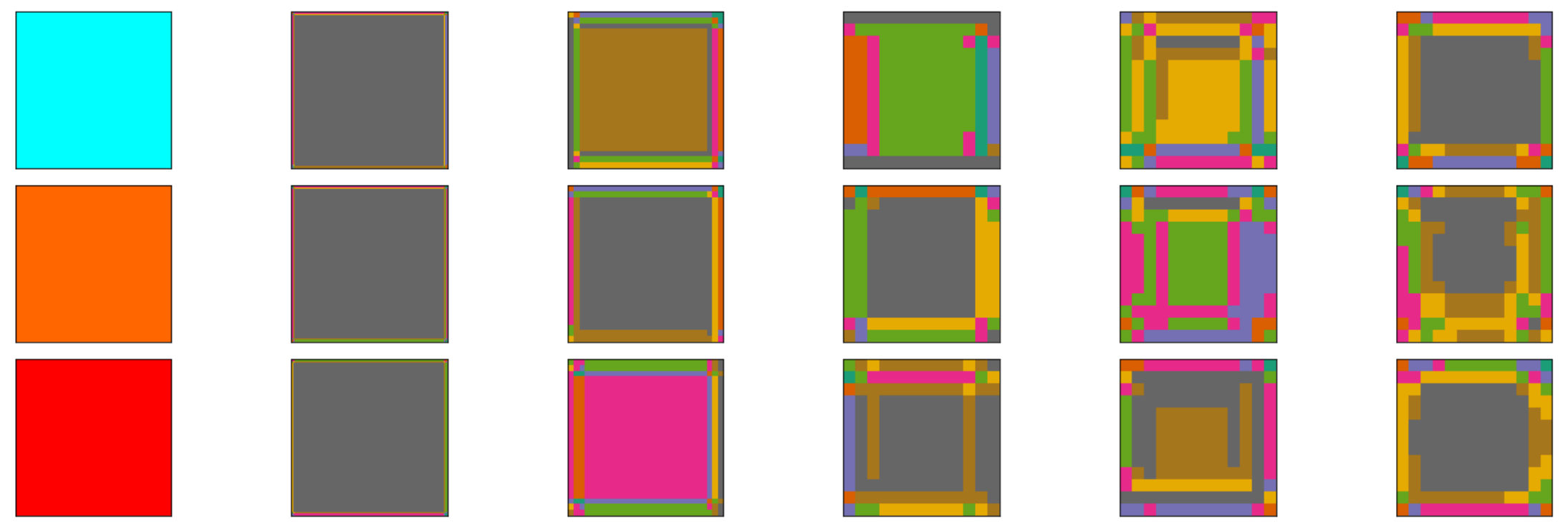
Dataset and Toolbox

DNNBrain[1] is a modular python toolbox to integrate DNN software packages and brain mapping tools. It integrates representations of DNNs and the neural representations of the brain. It assembles stimuli, artificial activity data, and biological neural activity data together, helping in comparing representations of DNNs and brains.
AlexNet [2] model was primarily used as our DNN model which consists of eight layers and learnable parameters. The architecture is composed of five convolutional layers and three fully connected layers that receives inputs from all units in previous layer followed by a 1000 way softmax classifier.
BOLD5000[3] is a large-scale publicly available dataset of practical study of human FMRI with 5000 real-world images as stimuli, which also accounts for image diversity and overlapping with standard computer vision datasets by incorporating images from the Scene Understanding (SUN), Common Objects in Context (COCO), and ImageNet datasets.

Experiments Carried out

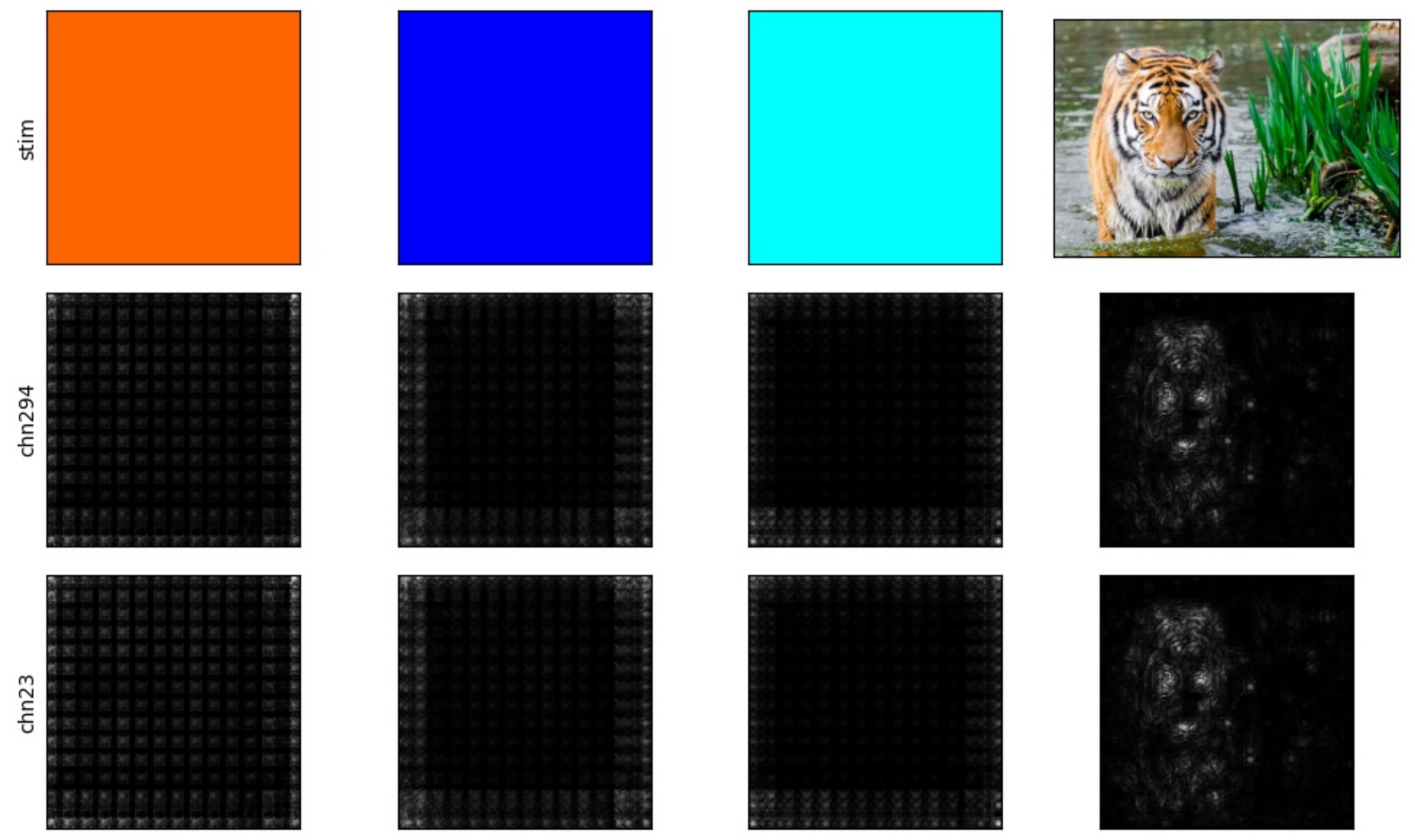
Many experiments were carried out to study the encoding of images in CNNs and correlations between CNNs and brains. Here we report only the ones we were able to extend to color inputs:

1. **Scan DNN:** To understand the perception of colors by the DNN, we passed plain color images in the DNN and observed the representations in different convolutional layers. From the result, it was evident that the initial layers perceive the colors more or less the same but as we go deeper, it starts differentiating the colors as shown in the figure below.



2. **Saliency Image:** It highlights pixels of the image that increase the unit's activation most when its value changes.The purpose of the saliency map is to find the regions which are prominent or noticeable at every location in the visual field and to guide the selection of attended locations, based on the spatial distribution of saliency. Saliency image for a color input is shown below.

3. **Minimal Image:** It is a way to simplify a stimulus into a minimal part which could cause almost equivalent activation as the original stimulus for a DNN unit. Some minimal images for colors are shown below.



Classification Task

A classification task is carried out for a dataset consisting of colors. Our hypothesis is that transfer learning will give better accuracies than the model which learns from scratch. We test this hypothesis.
Generating the dataset: To check the performance of both the models, we generated a dataset by creating images using different RGB values and creating numpy arrays using these RGB values. We generated 11 Basic Color Categories for Classification and labelled around 5000 RGB colors.
Learning from scratch: We defined a DNN which consisted one convolutional layer and two fully connected layers along with ReLU activation and 2D Max-Pooling. This model is designed to receive a tensor with shape as (nsample, 3, 224, 224), and do the classification.
Transfer learning: We used two major transfer learning scenarios:

1. Fixed feature extractor: Here, we replace the last fully-connected layer, replacing it with a linear layer, training just this layer for classification.
2. Fine tuning: The model remains same as above, but the losses are propagated backwards, fine-tuning the layers of the DNN itself.

Accuracy Data from the DNNs:

Model	Train split	Train Accuracy	Test split	Test Accuracy
Learning from Scratch	0.85	0.85	0.15	0.83
Transfer Learning (Feature Extractor)	0.85	0.88	0.15	0.87
Transfer Learning (Fine-Tuning)	0.85	0.86	0.15	0.85

Results and discussion

Our initial experiments showed how CNNs don't work quite the same compared to when there are salient parts present in the image. This is because CNNs are based on the shapes and edges present in the image. But, this still does not prove whether CNNs do not contain any usable information for color images.
Comparing the accuracy scores for the classification task, it is clear that the transfer learning approaches perform better. This is because CNN layers do capture some information which is usable in distinguishing one color from another.

GitHub Repository

`https://github.com/shivigup/Color-DL-Brain`

References

[1] Xiayu Chen, Ming Zhou, Zhengxin Gong, Wei Xu, Xingyu Liu, Taicheng Huang, Zonglei Zhen, and Jia Liu. Dnnbrain: A unifying toolbox for mapping deep neural networks and brains. *Frontiers in Computational Neuroscience*, 14:105, 2020.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[3] Nadine Chang, John A. Pyles, Austin Marcus, Abhinav Gupta, Michael J. Tarr, and Elissa M. Aminoff. Bold5000, a public fmri dataset while viewing 5000 visual images. *Scientific Data*, 6(1), May 2019.

Contributions

- **Members:** Ayushi Arora, Deeksha Vijay, Diksha Banka, Dravya Marwaha, Gaurvika Kapoor, Praveen Prabhat, Twinkle Arora
- **Mentor:** Shivi Gupta