Exercise 1 - Understanding Neuroscience Papers

Julian Büchel

September 2020

1 Summary

Previous work has shown that the "neural" activation in artificial neural networks (ANNs) shows remarkable similarity to neural activation recorded in the ventral visual stream, which was seen as evidence that these models capture the functionality of the visual cortex. Bashivan and colleagues [1] build on this work to show that these models not only resemble the brain but but can be used for neural population control.

The authors however raise two fundamental issues that they try to tackle in this paper:

1. The authors claim that the fact that this model shows similar activations is not direct evidence of the same *functionality*. The trained model could merely be a copy of the recorded activations and would not lead to a better understanding of the ventral system.

2. The authors raise the issue that the data that the model was trained and then later validated on stems from the same distribution and that the ANN predictions might fail if stimuli from another distribution were used.

To show that the model actually *did* implement some functionality of the ventral visual stream, the authors generated luminous images using the model ANN that would trigger particular events in V4. To validate the synthesized images, the authors recorded activation of a total of 107 neural sites in three rhesus macaques using implanted microelectrode arrays.
In the first experiment, where the authors aimed at simply stretching the activity of one particular neural site, out of the x neural sites recorded in each

monkey, the synthesis algorithm found at least one image for 79% of the neural sites that would drive that particular site at least 10% above its maximal firing rate when presented with naturalistic images.
In the second experiment, the authors aimed at activating only one neural site while suppressing the others. They found that their synthesized images increased the soft-max score, which is the score quantifying the confidence of the one-hot encoding, by more than 50% when compared to the maximal score produced by any naturalistic image.
To address the second issue that the authors raised, namely the question whether the model would perform equally well on novel images, the authors argued that the synthesized images used in the previously described experiments already differed so much from the images used in training the ANN model and the mapping between ANN activations and V4 activity.
In summary, the authors have shown that images, synthesized from a trained ANN model, can be used to control individual neural sites with statistical significance. They have also shown that their ANN indeed did improve the "controllability" of the neural activity when compared to simply finding the best naturalistic image. These results are an important contribution to the field.

References

[1] Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. *Science*, 2018.