

# A study of the impact of CNN architecture variation on predicting brain activity using feature-weighted receptive fields

AUTHOR: VLAD MURGOCI

Supervisor: Xucong Zhang  
Thesis Committee: Xucong Zhang, Nergis Tömen

## 01. Introduction and Objective

Functional Magnetic Resonance Imaging (fMRI) allows for an analysis of brain activity by measuring the levels of oxygen in the blood. A brain encoding model predicts brain activity based on an input (visual stimulus). By leveraging the similarities of Convolutional Neural Networks and the visual cortex, we can create a brain encoding model using a pre-trained CNN and a feature-weighted receptive field model (fwRF) [1].

### The research questions:

How does variation of the CNN architecture affect the performance of the fwRF encoding model?  
Do CNNs with better accuracy on the image classification task produce better predictions for brain activity?

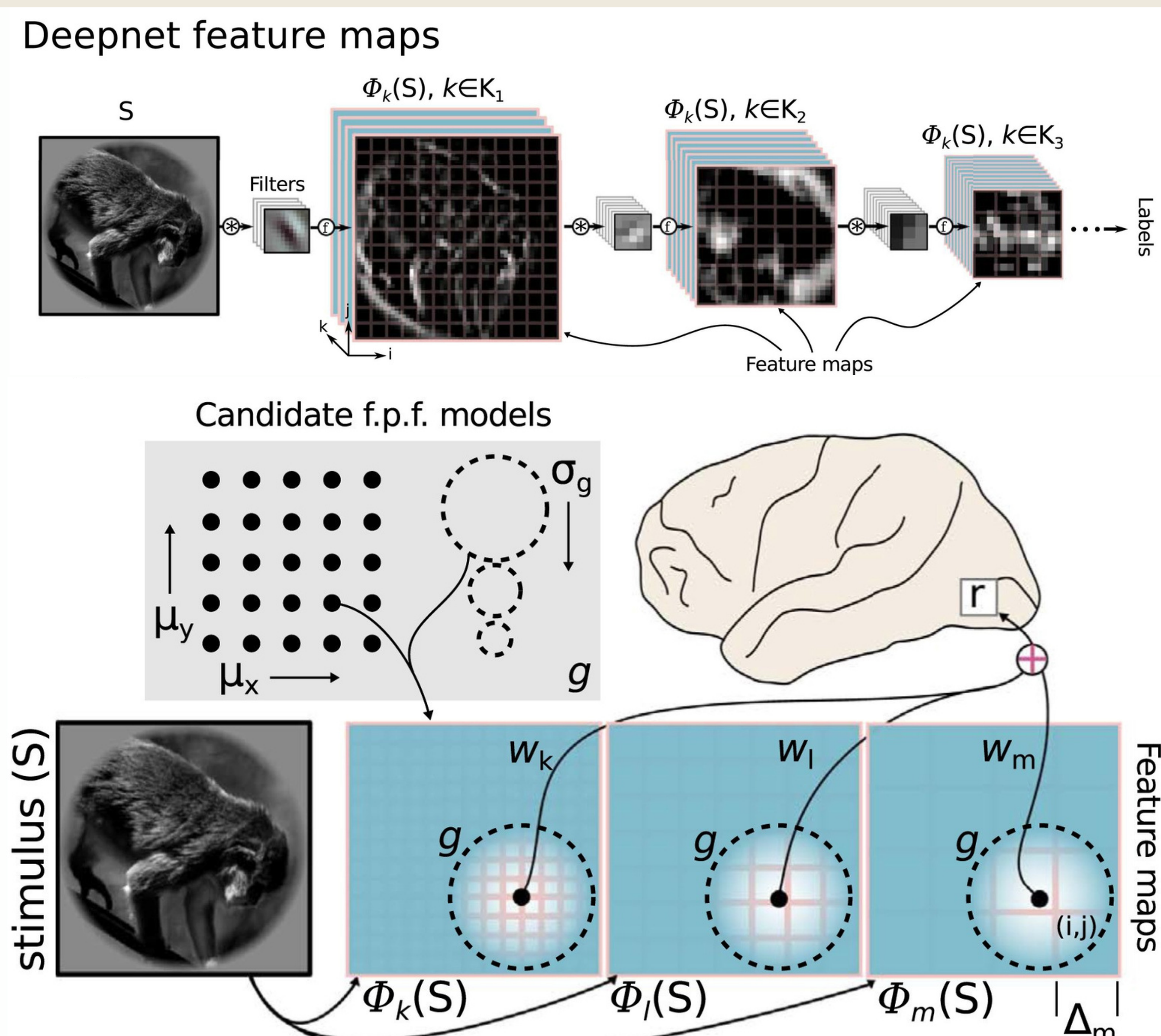


Figure 1: The training process of the fwRF model, adapted from [1]

## 02. Methodology

- Data was collected from the Natural Scenes Dataset [2].
- CNN models are pre-trained on the ImageNet dataset.
- The fwRF encoding model is trained for each model and subject, based on feature maps extracted from the CNN.
- The selected models were:

Model Name	Top-1 Accuracy
GoogleNet	69.778%
VGG13	69.928%
ResNet-18	69.758%
AlexNet	56.522%
EfficientNetV2_S	84.228%

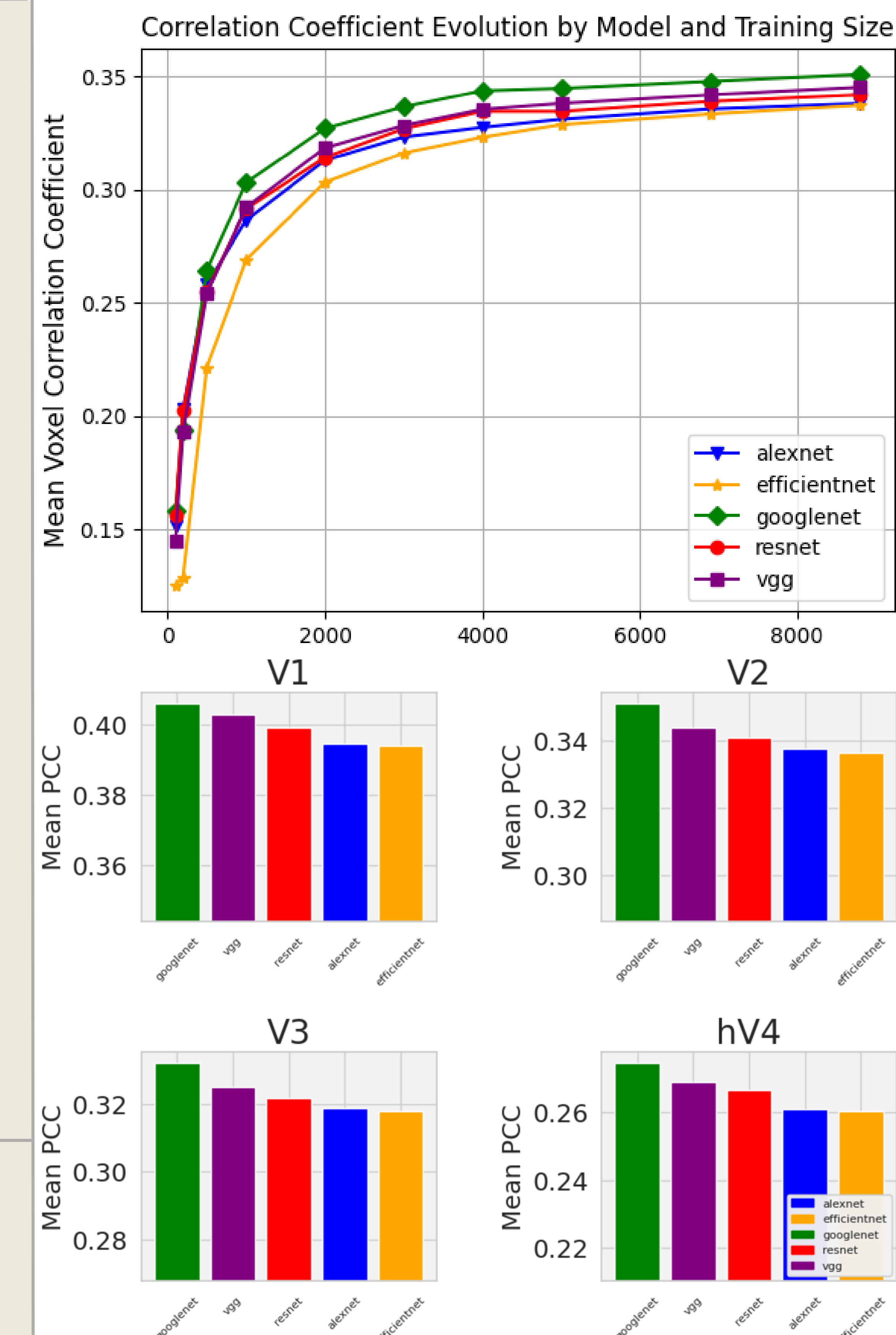
## 03. Results

- GoogleNet outperforms the other models in every experimental setting.
- EfficientNet presents the worst performance.
- All models experience a **drop** in accuracy for **latter regions of interest (ROI)** in the visual cortex.

Model Name	Mean Accuracy (PCC)
GoogleNet	0.3
VGG13	0.293
ResNet-18	0.291
AlexNet	0.288
EfficientNetV2_S	0.279

## 04. Conclusion and Future Work

- Results indicate that the choice of architecture has an impact on the quality of predictions. GoogleNet may outperform due to its Inception architecture.
- CNN accuracy on image classification seems to not be a decisive factor on the quality of the encoding model.
- Brain-optimized CNNs [3] could capitalize on the inconsequential classification accuracy.
- Voxel-to-voxel [4] DNNs are a solution for the decreased accuracy in latter ROIs.



## References

- [1] G. St-Yves and T. Naselaris, "The feature-weighted receptive field: an interpretable encoding model for complex feature spaces," *NeuroImage*, vol. 180, pp. 188–202, 2018. New advances in encoding and decoding of brain signals.
- [2] E. J. Allen, G. St-Yves, Y. Wu, J. L. Breedlove, J. S. Prince, L. T. Dowdle, M. Nau, B. Caron, F. Pestilli, I. Charest, J. B. Hutchinson, T. Naselaris, and K. Kay, "A massive 7t fMRI dataset to bridge cognitive neuroscience and artificial intelligence," *Nature Neuroscience*, vol. 25, pp. 116–126, Jan 2022.
- [3] G. St-Yves, E. J. Allen, Y. Wu, K. Kay, and T. Naselaris, "Brain-optimized neural networks learn non-hierarchical models of representation in human visual cortex," *bioRxiv*, 2022.
- [4] W. Xiao, J. Li, C. Zhang, L. Wang, P. Chen, Z. Yu, L. Tong, and B. Yan, "High-level visual encoding model framework with hierarchical ventral stream-optimized neural networks," *Brain Sciences*, vol. 12, no. 8, 2022. All Open Access, Gold Open Access, Green Open Access.