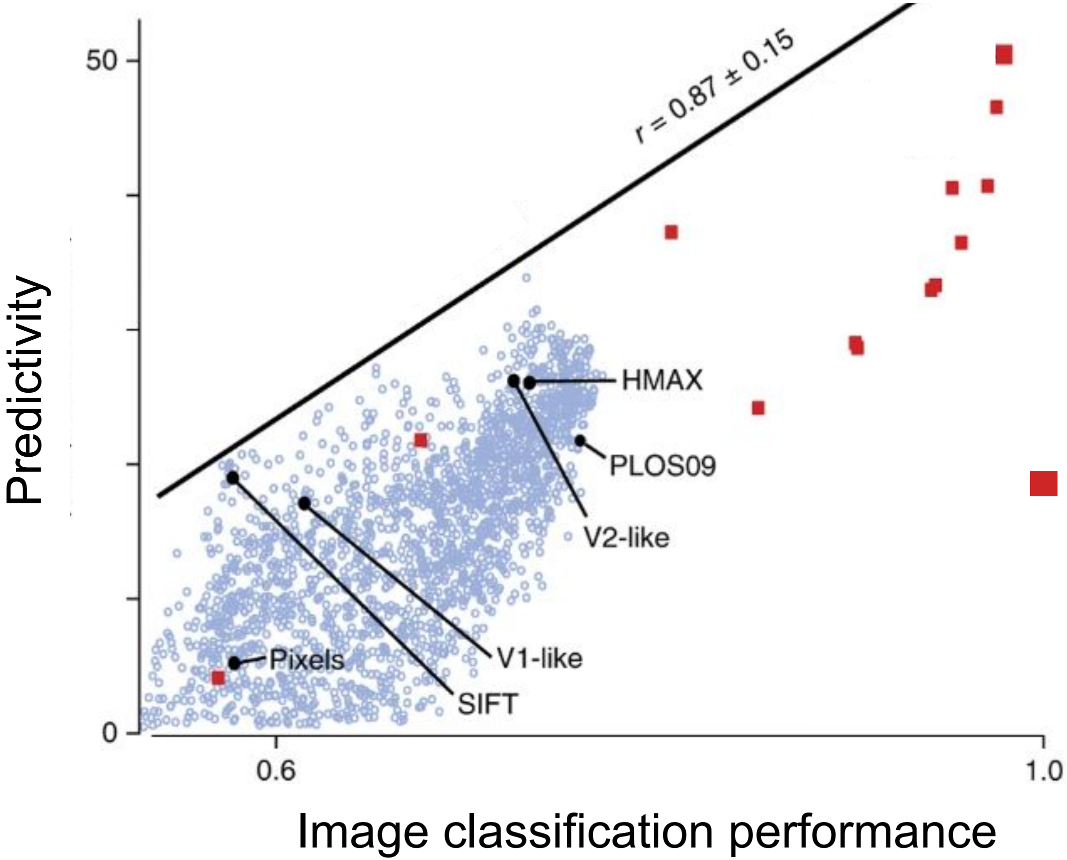
Do convolutional neural networks show brain-like unilateral processing?

Yota Kawashima
BCCN Berlin
2024/10/18

Better performance, more brain-like

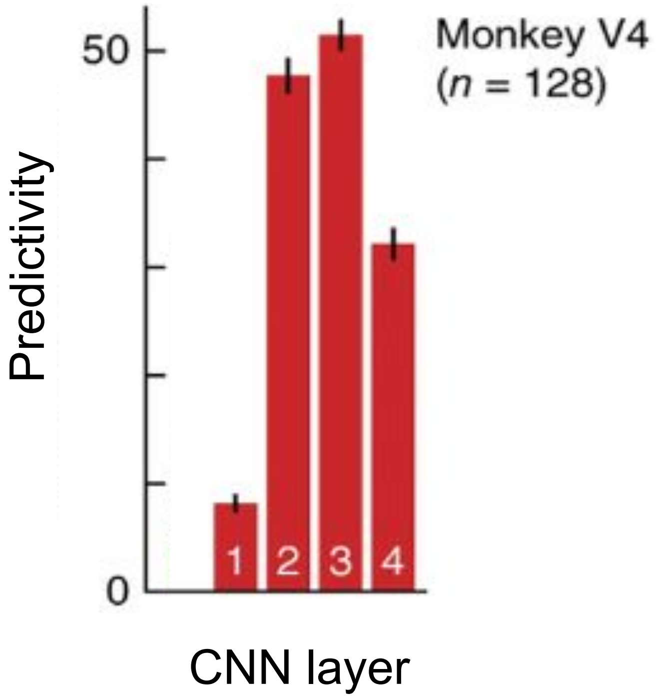


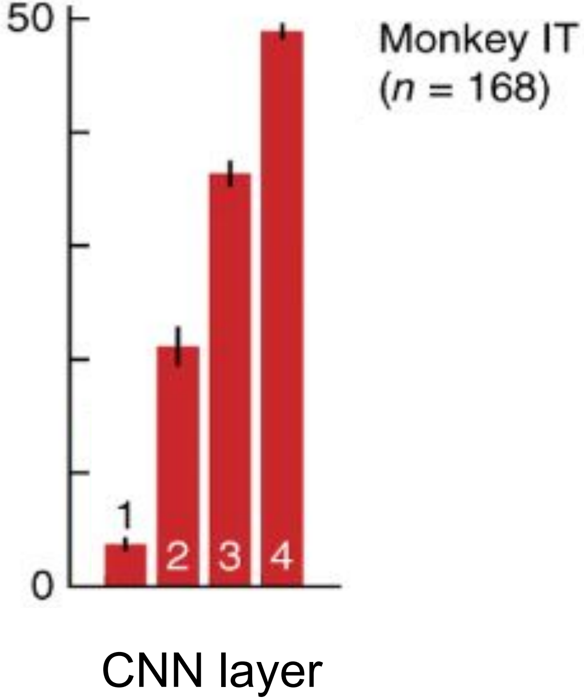
Monkey IT Electrophysiology

■ CNN at different time points during training

(Yamins et al, 2014, 2016)

Similar in what way? Hierarchical structure

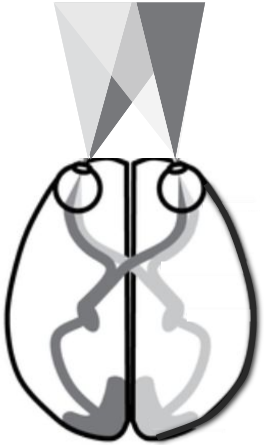




(Yamins et al, 2014, 2016)

How about unilateral procesing?

Better CNNs seem more similar to the visual areas.



<u>Hypothesis</u>

Better CNNs explain unilateral processing better.

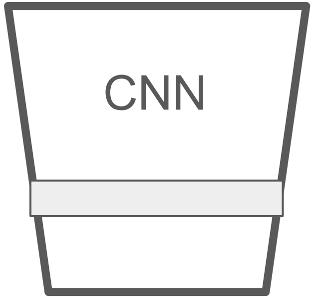
> How can we test this?

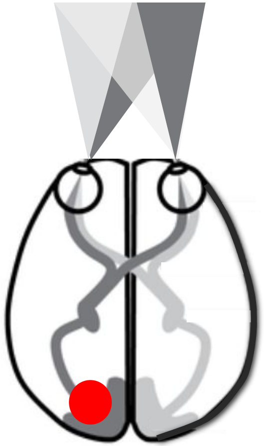
Image from wikipedia

The standard encoding model

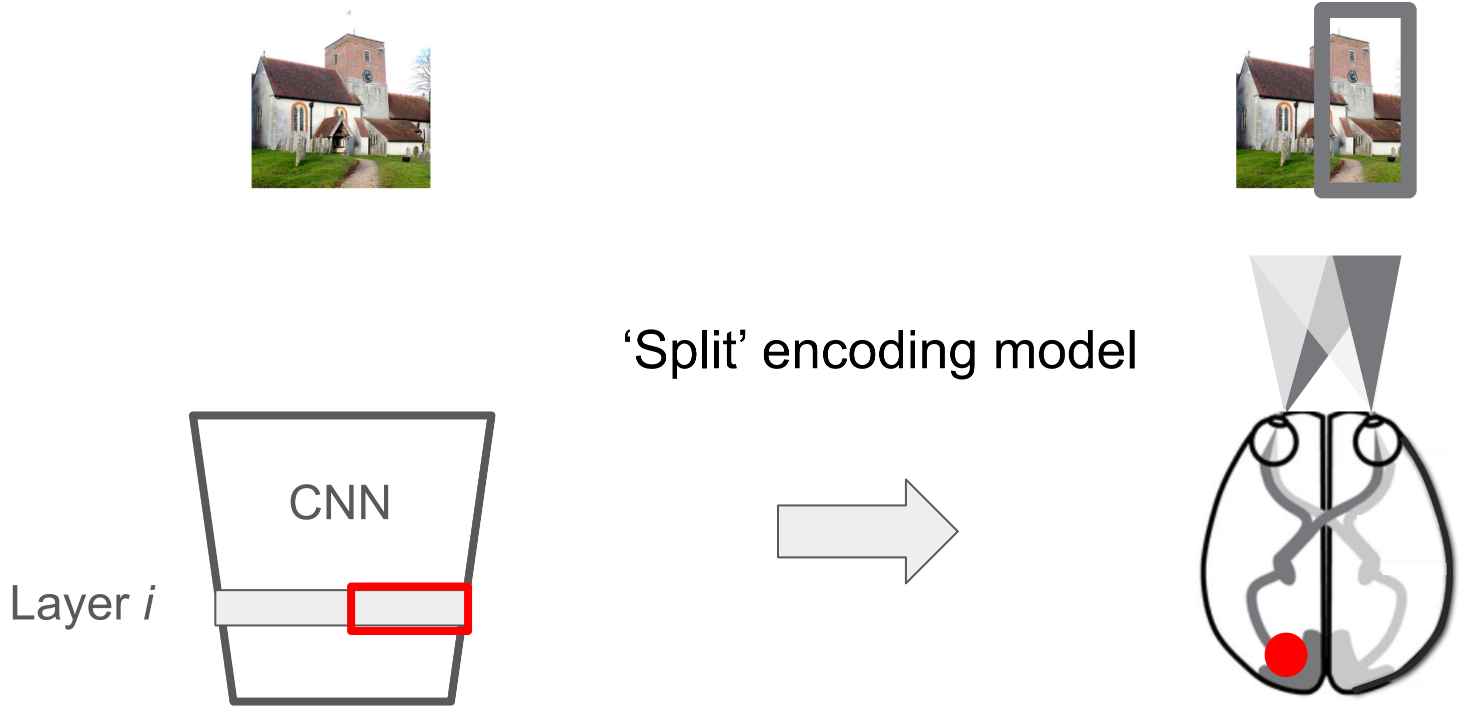
Encoding model

Layer *i*



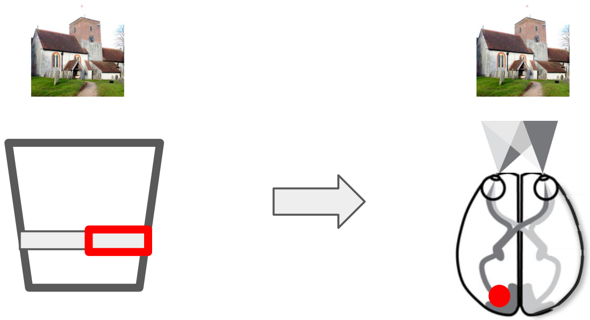


‘Spliting’ NN layers in encoding model

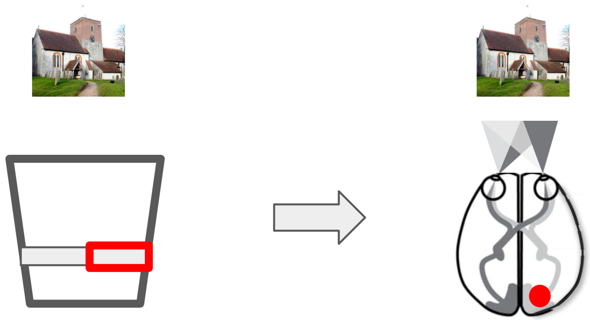


Two encoding model prediction

Contralateral prediction



Ipsilateral prediction



Contralateral prediction should be better

Contralateral prediction

Ipsilateral prediction

If CNNs are similar to the brain,

- Better contralateral prediction than ipsilateral prediction.
- Large difference in early layers.

fMRI dataset and CNN models

The Algonauts project fMRI dataset

- 8 participants
- 872 natural scence images

The Efficient network (Tan & Le, 2019)

- Relatively few parameters
- Good performance

Models to be compared

- Trained (ImageNet-1K)
- Not-trained (random weights)

Stage i	Operator $\hat{\mathcal{F}}_i$
1	Conv3x3
2	MBConv1, k3x3
→ 3	MBConv6, k3x3
4	MBConv6, k5x5
5	MBConv6, k3x3
→ 6	MBConv6, k5x5
7	MBConv6, k5x5
8	MBConv6, k3x3
→ 9	Conv1x1 & Pooling & FC

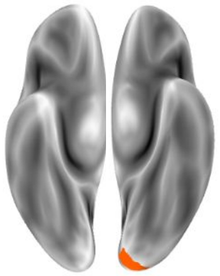
The Algonauts project, Pytorch pretrained models

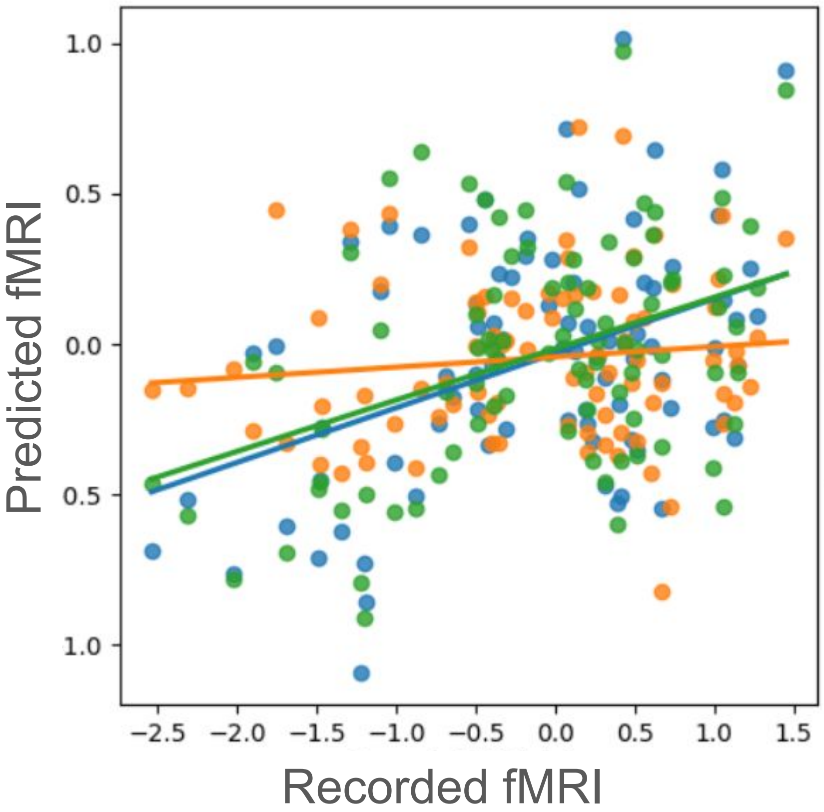
Result: single vertex prediction (*N* = 1)

Trained model

The early layer

Early visual area (left)

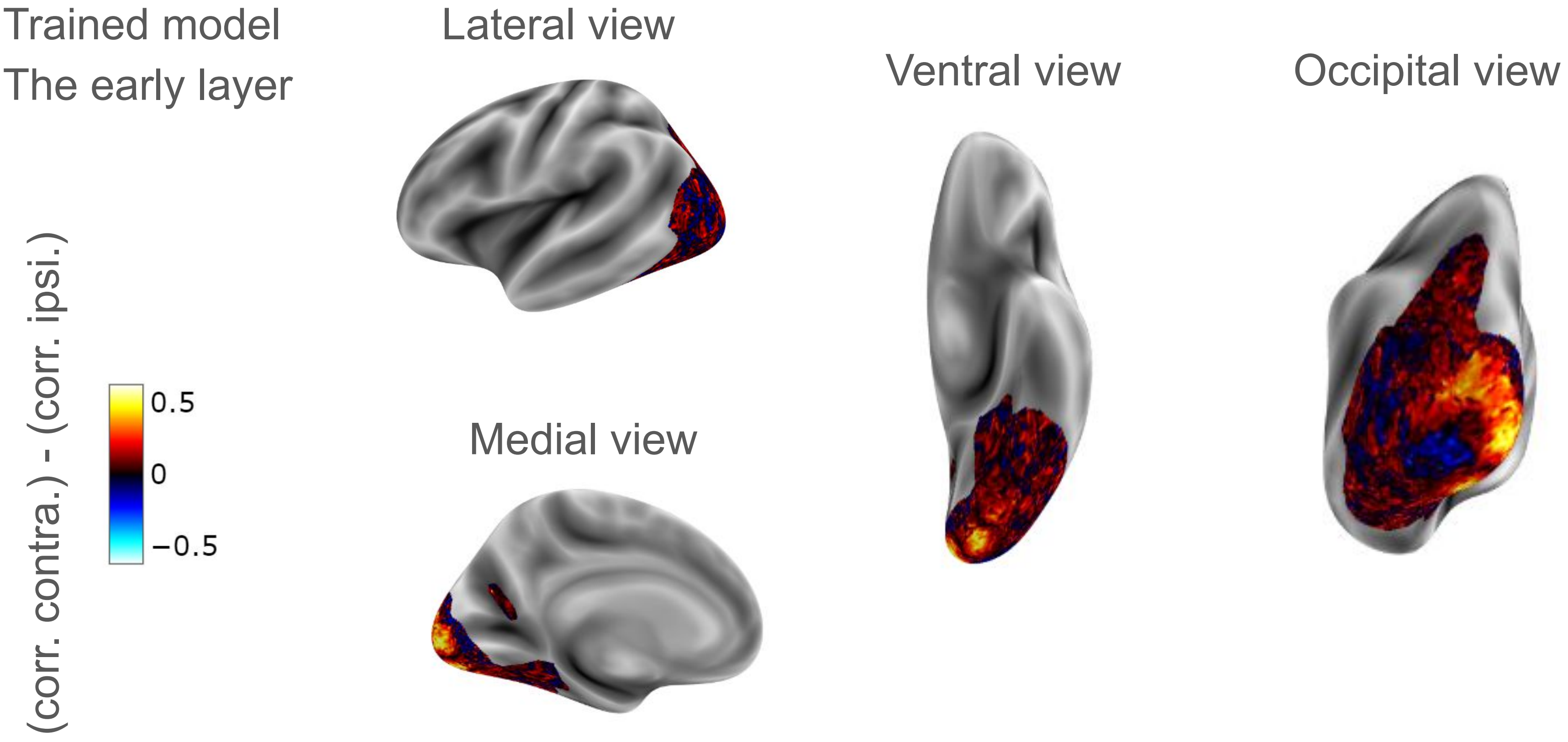




	corr.
Full:	0.55
Ipsilateral:	0.19
Contralateral:	0.67

Contralateral prediction seems better (N = 1)

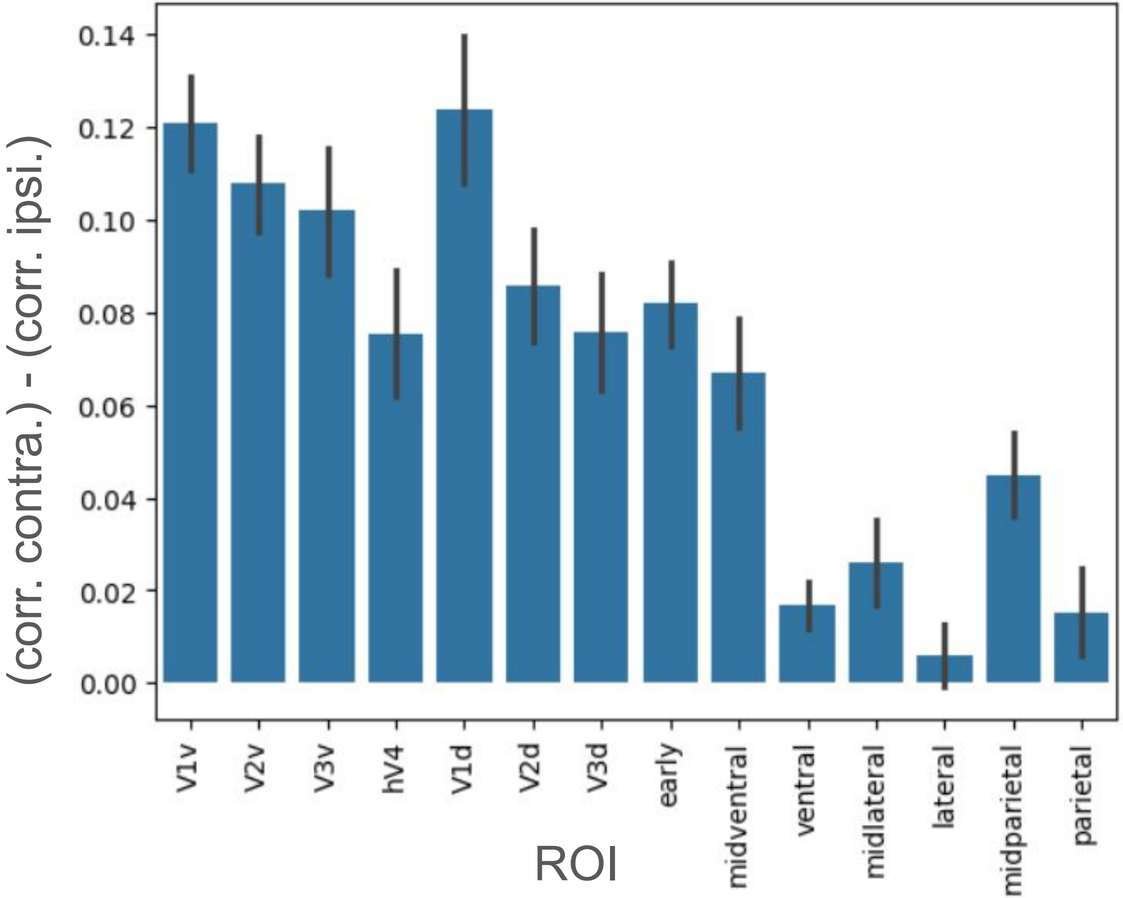
Trained model
The early layer



Contralateral prediction is better (*N* = 8)

Trained model
The early layer

Mean & std across
participants & hemispheres

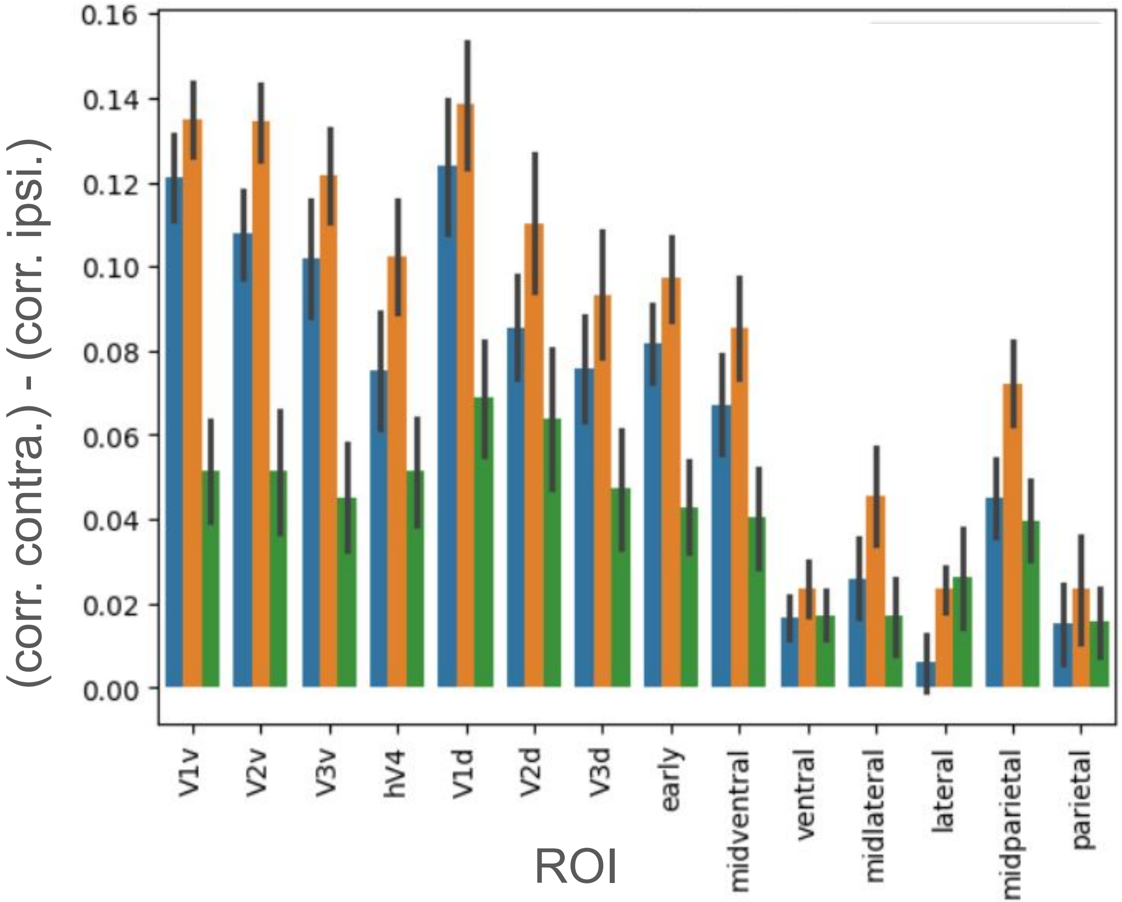


Early layers show larger difference

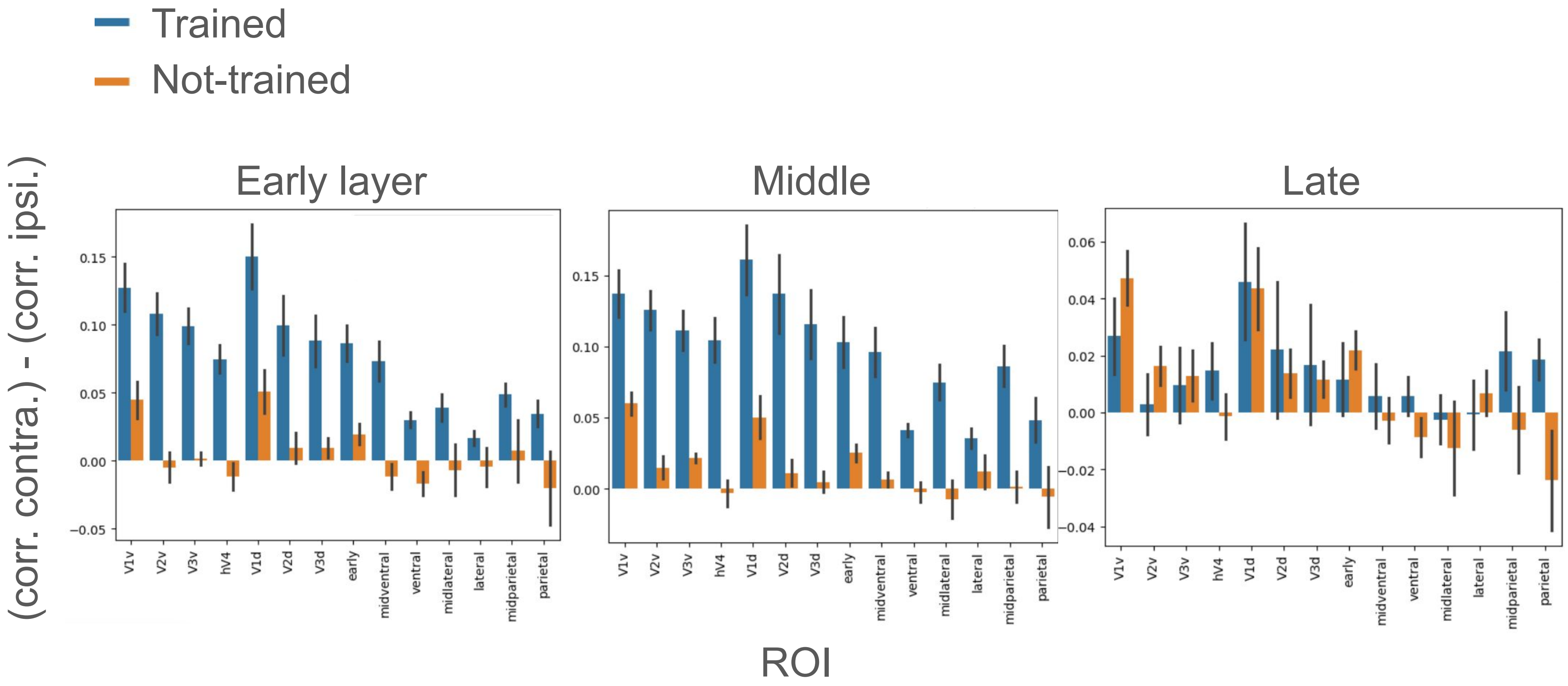
Trained model

Mean & std across participants & hemispheres

Early layer
Middle
Late



Trained model shows larger difference



Summary & Discussion

- Better CNNs seem to have more brain-like unilateral processing.
    - ‘Split’ encoding model for the trained effNet
    - Better contralateral prediction than ipsilateral prediction.
    - Larger difference in early layers.
    - But not for the not-trained effNet

- Half-trained (e.g. different time points during training)
- Architecture is important! (Yamins et al 2014)
    - ‘Split’ in architecture
    - Bilateral visual field maps in a patient with one hemisphere (Muckli et al 2009)

Code

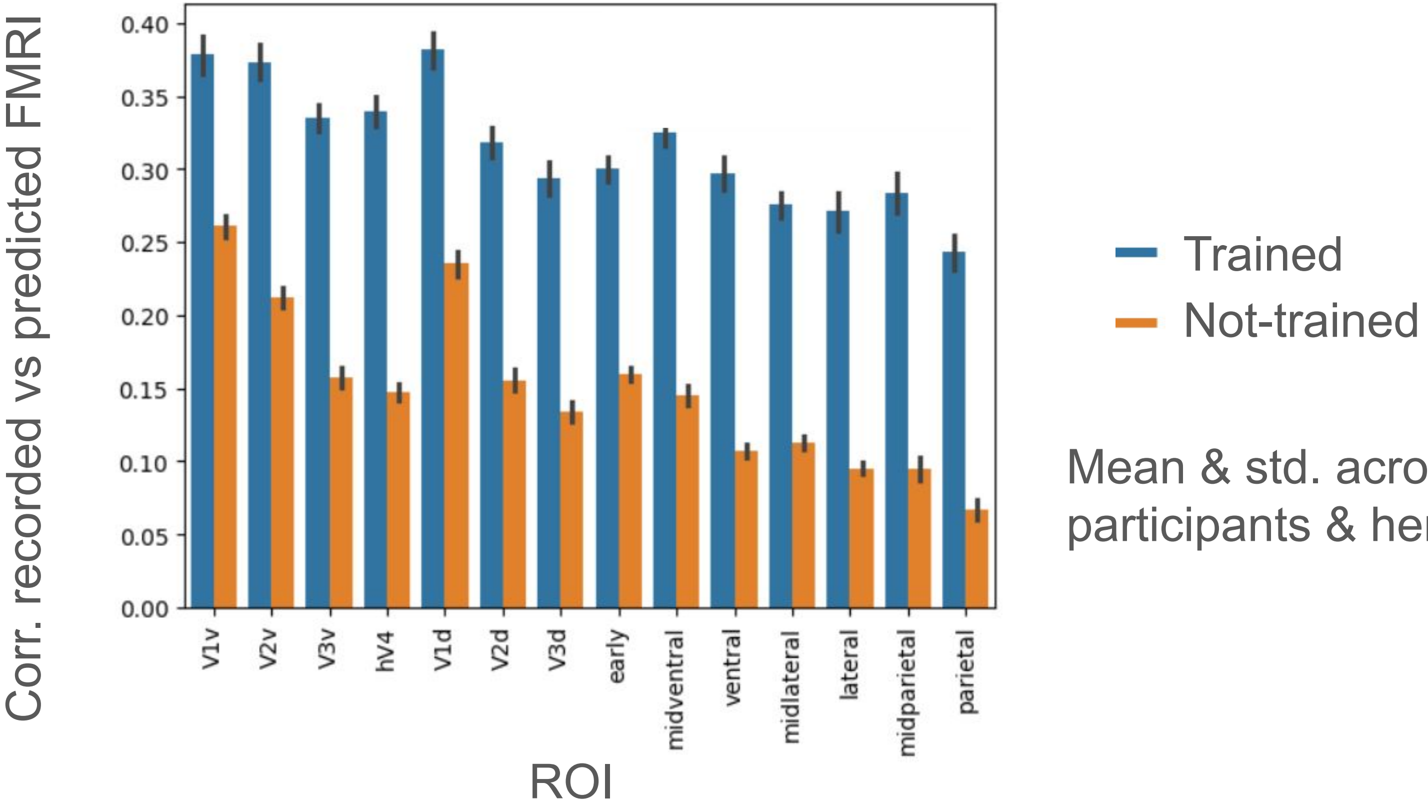
- https://drive.google.com/drive/folders/1zDgVxCHiWDRAMLSx30tVdc9bftGcKZXq?usp=sharing

Reference

- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111
- Yamins, D., DiCarlo, J. Using goal-driven deep learning models to understand sensory cortex. Nat Neurosci 19, 356–365 (2016). https://doi.org/10.1038/nn.4244
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional Neural Networks. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/1905.11946
- Muckli, L., Naumer, M. J., & Singer, W. (2009). Bilateral visual field maps in a patient with only one hemisphere. Proceedings of the National Academy of Sciences of the United States of America, 106(31), 13034–13039. https://doi.org/10.1073/pnas.0809688106
- The Algonauts project http://algonauts.csail.mit.edu/challenge.html
- Pytorch pretrained models: https://pytorch.org/vision/stable/models.html#table-of-all-available-classification-weights

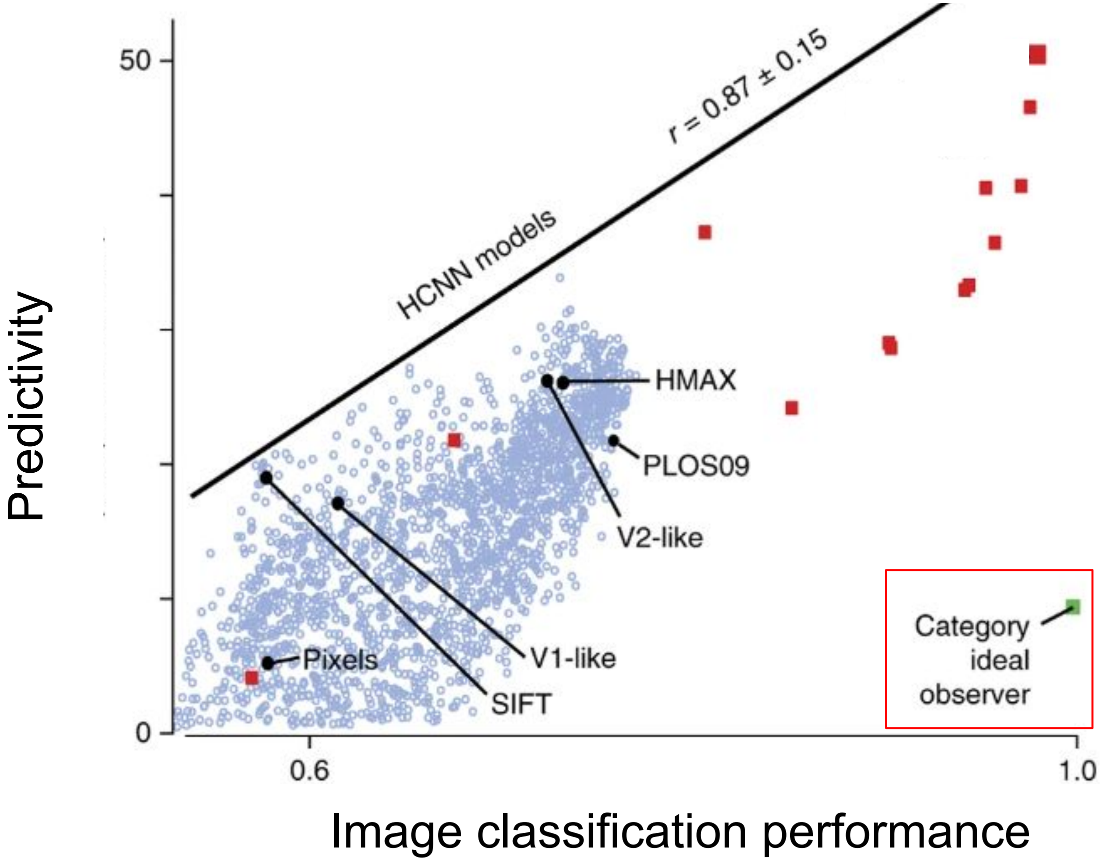
Note

Performance: the standard encoding model



Mean & std. across participants & hemispheres

Category ideal observer (semantic model)

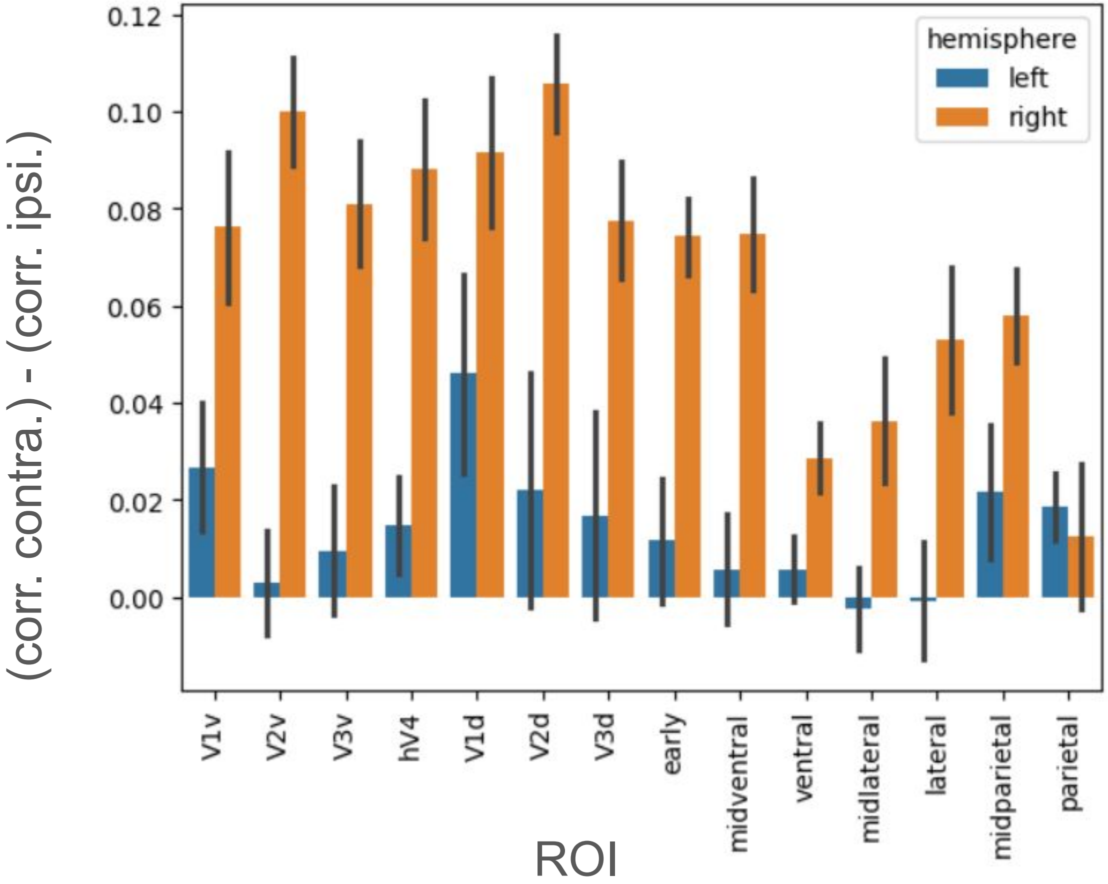


Comparing hemispheres in the last layer

Trained model
The last layer

Mean & std across participants

There are no difference between hemisphere in the other two early layers. I don't know why we see this in the late layer…

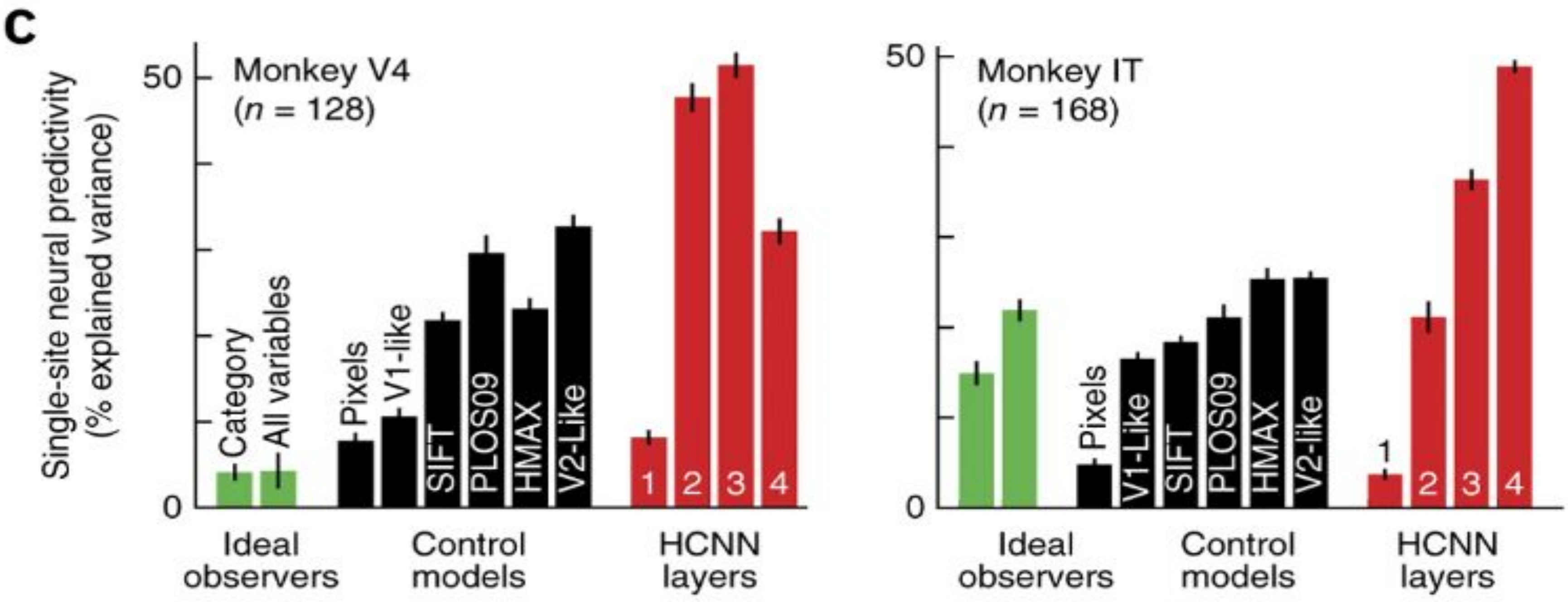


Efficient network architecture

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

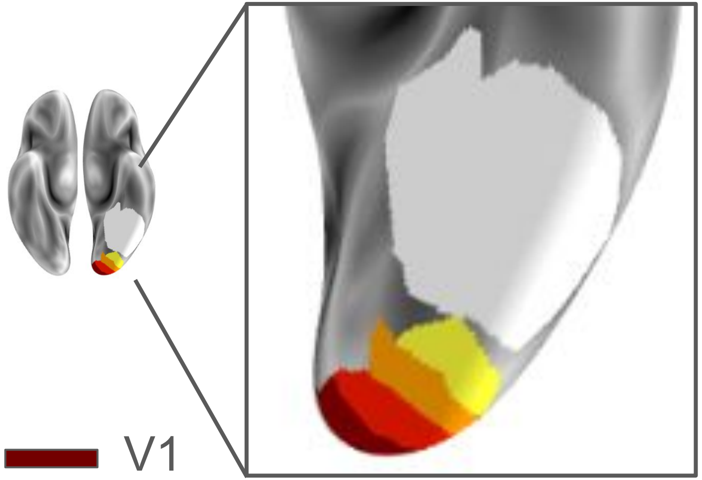
Weight	Acc@1	Acc@5	Params
AlexNet_Weights.IMAGENET1K_V1	56.522	79.066	61.1M
ConvNeXt_Base_Weights.IMAGENET1K_V1	84.062	96.87	88.6M
ConvNeXt_Large_Weights.IMAGENET1K_V1	84.414	96.976	197.8M
ConvNeXt_Small_Weights.IMAGENET1K_V1	83.616	96.65	50.2M
ConvNeXt_Tiny_Weights.IMAGENET1K_V1	82.52	96.146	28.6M
DenseNet121_Weights.IMAGENET1K_V1	74.434	91.972	8.0M
DenseNet161_Weights.IMAGENET1K_V1	77.138	93.56	28.7M
DenseNet169_Weights.IMAGENET1K_V1	75.6	92.806	14.1M
DenseNet201_Weights.IMAGENET1K_V1	76.896	93.37	20.0M
EfficientNet_B0_Weights.IMAGENET1K_V1	77.692	93.532	5.3M
EfficientNet_B1_Weights.IMAGENET1K_V1	78.642	94.186	7.8M
EfficientNet_B1_Weights.IMAGENET1K_V2	79.838	94.934	7.8M
EfficientNet_B2_Weights.IMAGENET1K_V1	80.608	95.31	9.1M
EfficientNet_B3_Weights.IMAGENET1K_V1	82.008	96.054	12.2M
EfficientNet_B4_Weights.IMAGENET1K_V1	83.384	96.594	19.3M
EfficientNet_B5_Weights.IMAGENET1K_V1	83.444	96.628	30.4M
EfficientNet_B6_Weights.IMAGENET1K_V1	84.008	96.916	43.0M
EfficientNet_B7_Weights.IMAGENET1K_V1	84.122	96.908	66.3M
EfficientNet_V2_L_Weights.IMAGENET1K_V1	85.808	97.788	118.5M
EfficientNet_V2_M_Weights.IMAGENET1K_V1	85.112	97.156	54.1M
EfficientNet_V2_S_Weights.IMAGENET1K_V1	84.228	96.878	21.5M
GoogLeNet_Weights.IMAGENET1K_V1	69.778	89.53	6.6M
Inception_V3_Weights.IMAGENET1K_V1	77.294	93.45	27.2M
MNASNet0_5_Weights.IMAGENET1K_V1	67.734	87.49	2.2M
MNASNet0_75_Weights.IMAGENET1K_V1	71.18	90.496	3.2M
MNASNet1_0_Weights.IMAGENET1K_V1	73.456	91.51	4.4M
MNASNet1_3_Weights.IMAGENET1K_V1	76.506	93.522	6.3M
MaxVit_T_Weights.IMAGENET1K_V1	83.7	96.722	30.9M
MobileNet_V2_Weights.IMAGENET1K_V1	71.878	90.286	3.5M
MobileNet_V2_Weights.IMAGENET1K_V2	72.154	90.822	3.5M
MobileNet_V3_Large_Weights.IMAGENET1K_V1	74.042	91.34	5.5M
MobileNet_V3_Large_Weights.IMAGENET1K_V2	75.274	92.566	5.5M
MobileNet_V3_Small_Weights.IMAGENET1K_V1	67.668	87.402	2.5M

From Yamins paper



Single vertex prediction

Left hemisphere (axial)



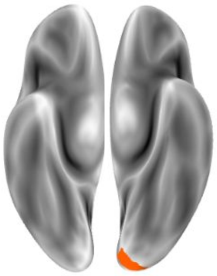
V1

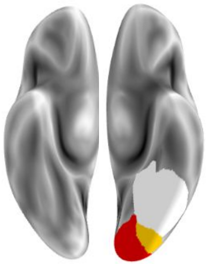
V2

V3

V4

Ventral (higher)





Contralateral prediction seems better (*N* = 1)

Trained model

The middle layer