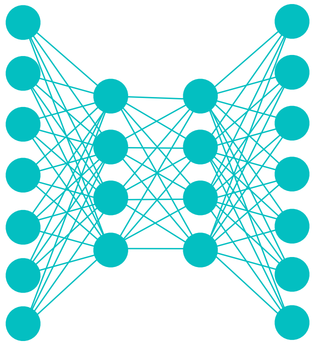
Lecture Notes for

Neural Networks and Machine Learning

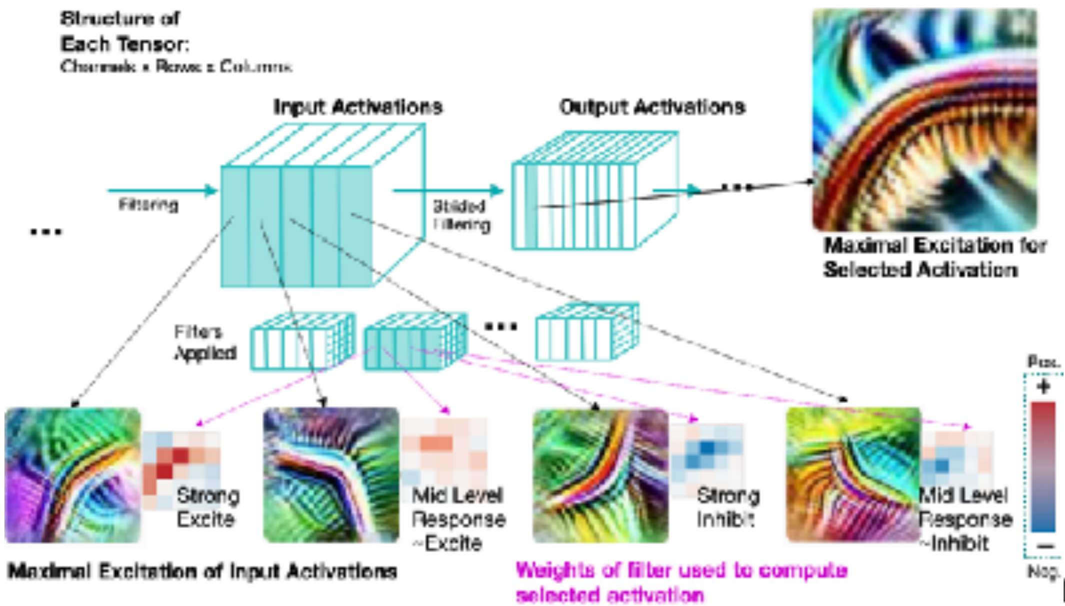
CNN Circuits Continued



Logistics and Agenda

- Logistics
  - Grading Update
  - Lab logistics!
- Agenda
  - Last Time: Circuits in CNNs
  - Continued Circuits
  - Student Paper Presentation
  - Lab Three Town Hall
  - Next Time: Fully Convolutional Networks

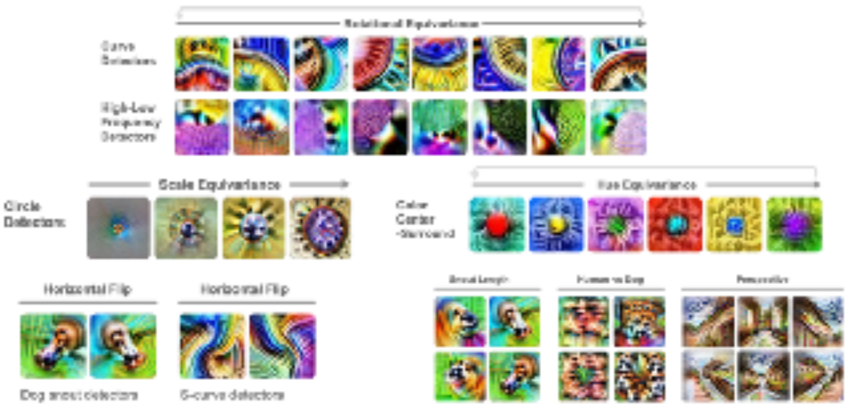
Last Time



Neuron 4b:409

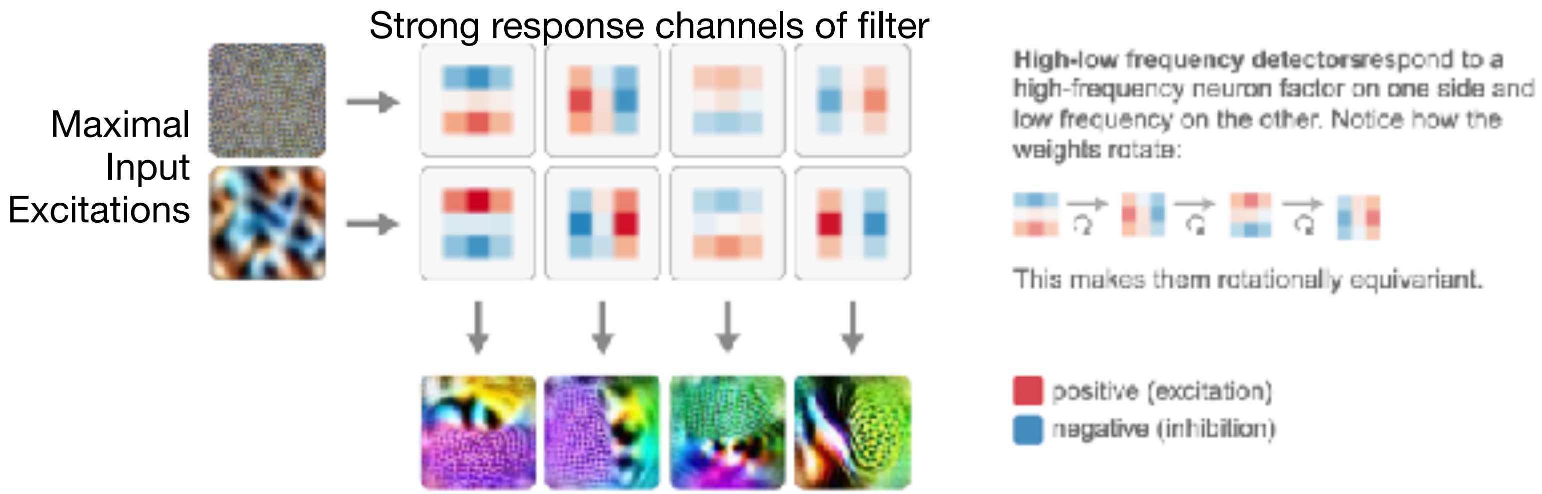
Dataset examples for neuron 4b:409

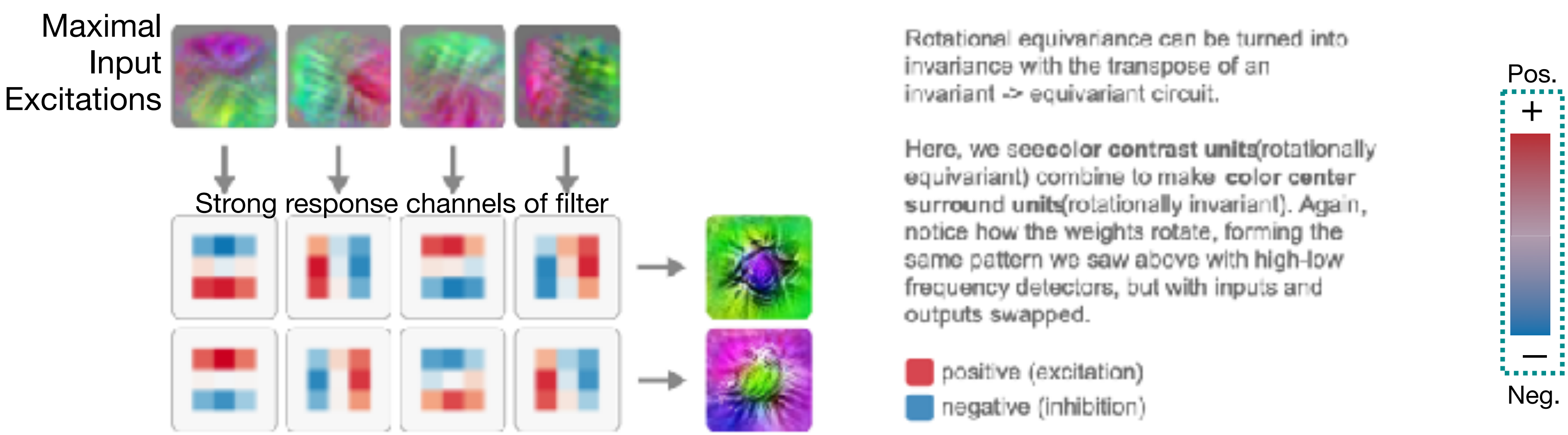
Many features that are part of a circuit are clearly designed for rotation, hue, and other invariance



Review of Equivariant Circuits

- Possible to reveal patterns of circuits via sets of weights

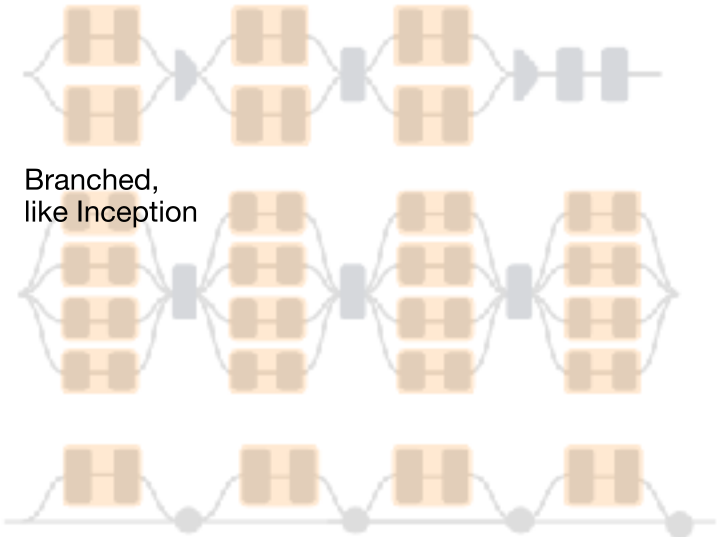




Olah, et al., "Naturally Occurring Equivariance in NN", 2021.

Branch Specialization

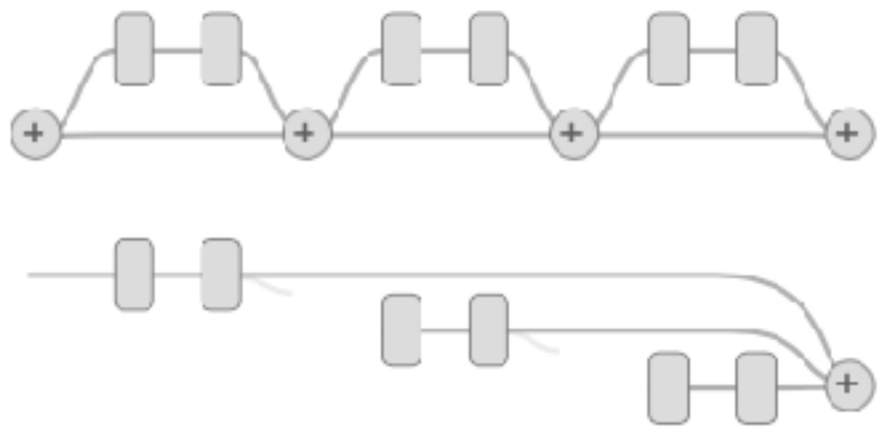
Feedforward



Branched,
like Inception

Residual,
like ResNet

Two ways of looking at residual networks



Primitives

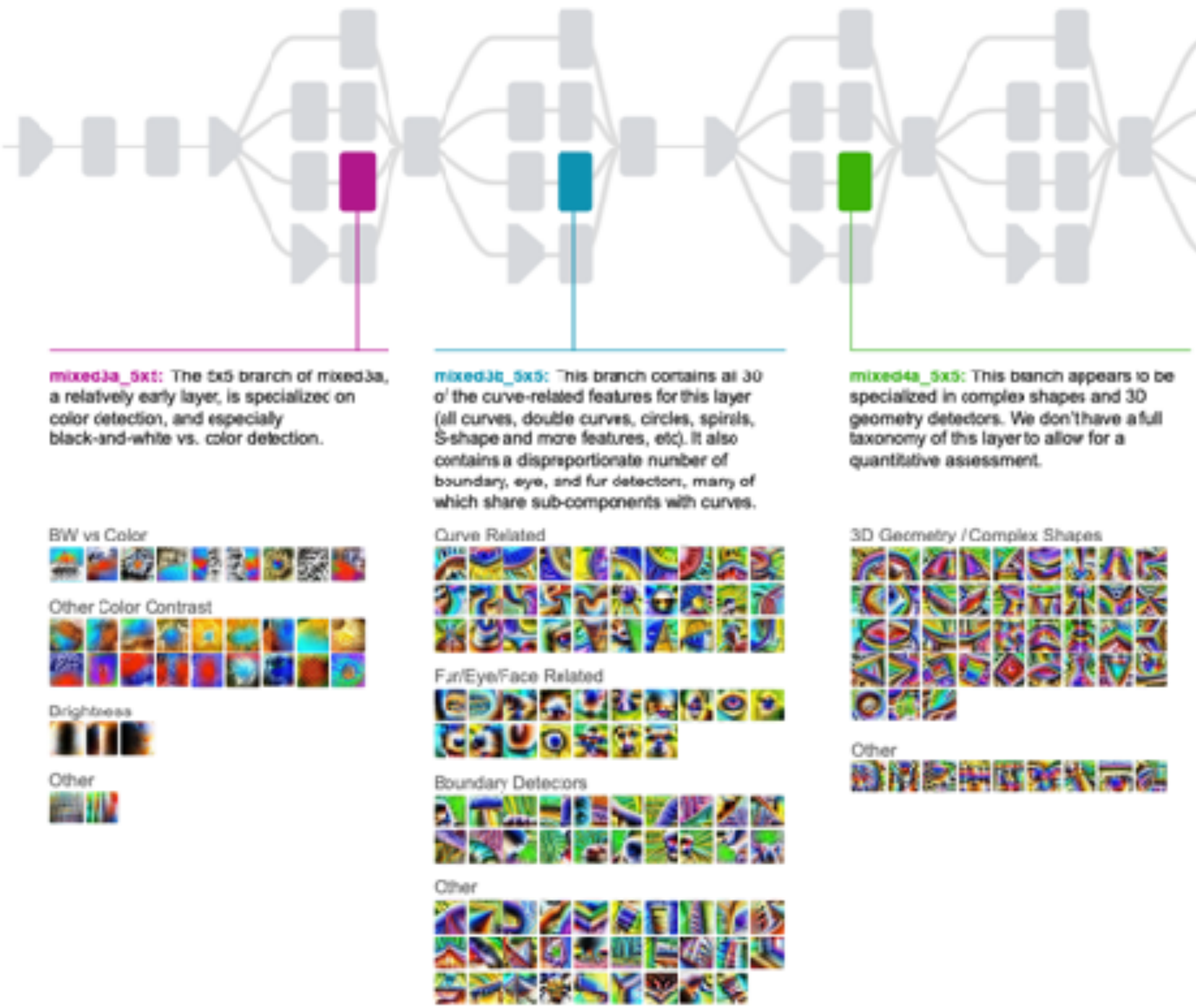
More Complex
Downstream Circuits

- Specialized branches are consistent across many architectures, support the idea of an interconnected graph of operations

https://distill.pub/2020/circuits/branch-specialization/

Voss, et al., “Branch Specialization", Distill, 2021.

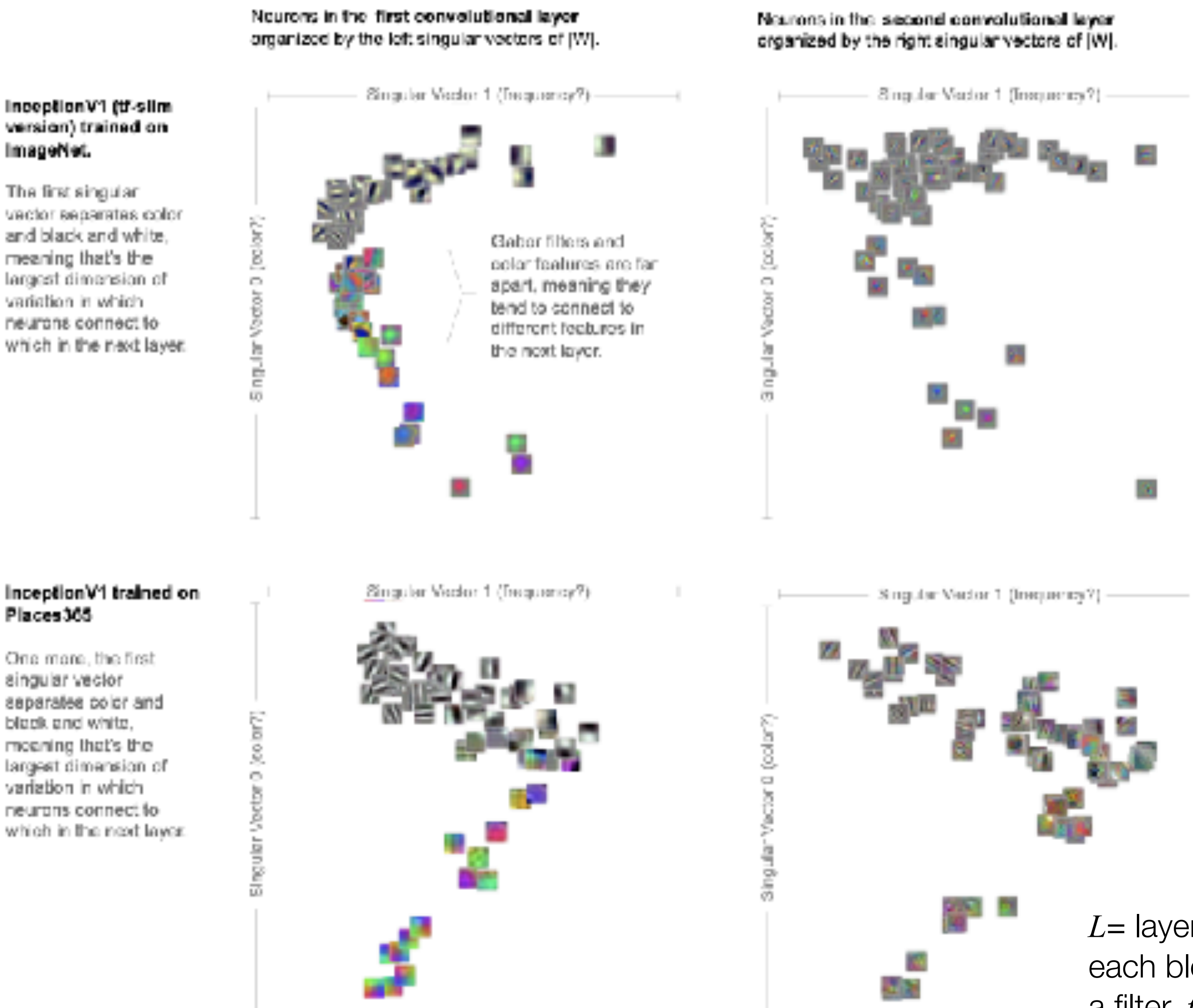
Branch Specialization



Motifs appear in branches. Similar clusters of operations can be found across different architectures

Voss, et al., "Branch Specialization", Distill, 2021.

Investigating Connection Clusters via SVD



Voss, et al., "Branch Specialization", Distill, 2021.

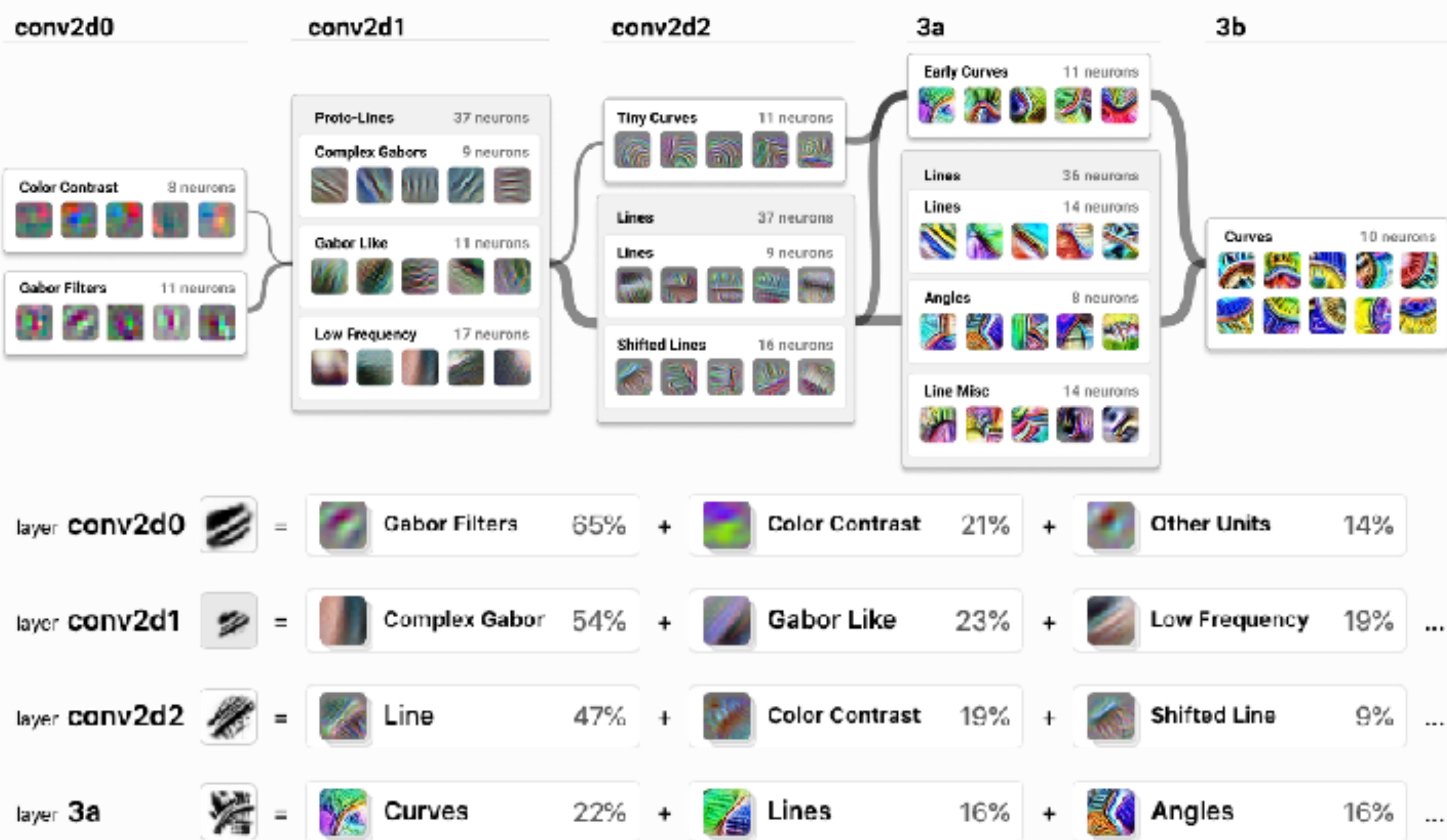
- Singular Value Decomposition (SVD) decomposes a matrix into three elements
  - $W = U\Sigma V^T$
  - U is eig-vec of WW^T
    V is eig-vec of W^TW
  - U and V are orthogonal such that
    $UU^T=I \quad VV^T=I$
  - Σ is a diagonal matrix of the singular values
  - These values characterize the variability in a matrix

$$SVD(|\mathbf{W}_f^{(L)}|)$$

$$SVD(|\mathbf{W}_f^{(L+1)}|)$$

L= layer
each block is a filter, f, in layer

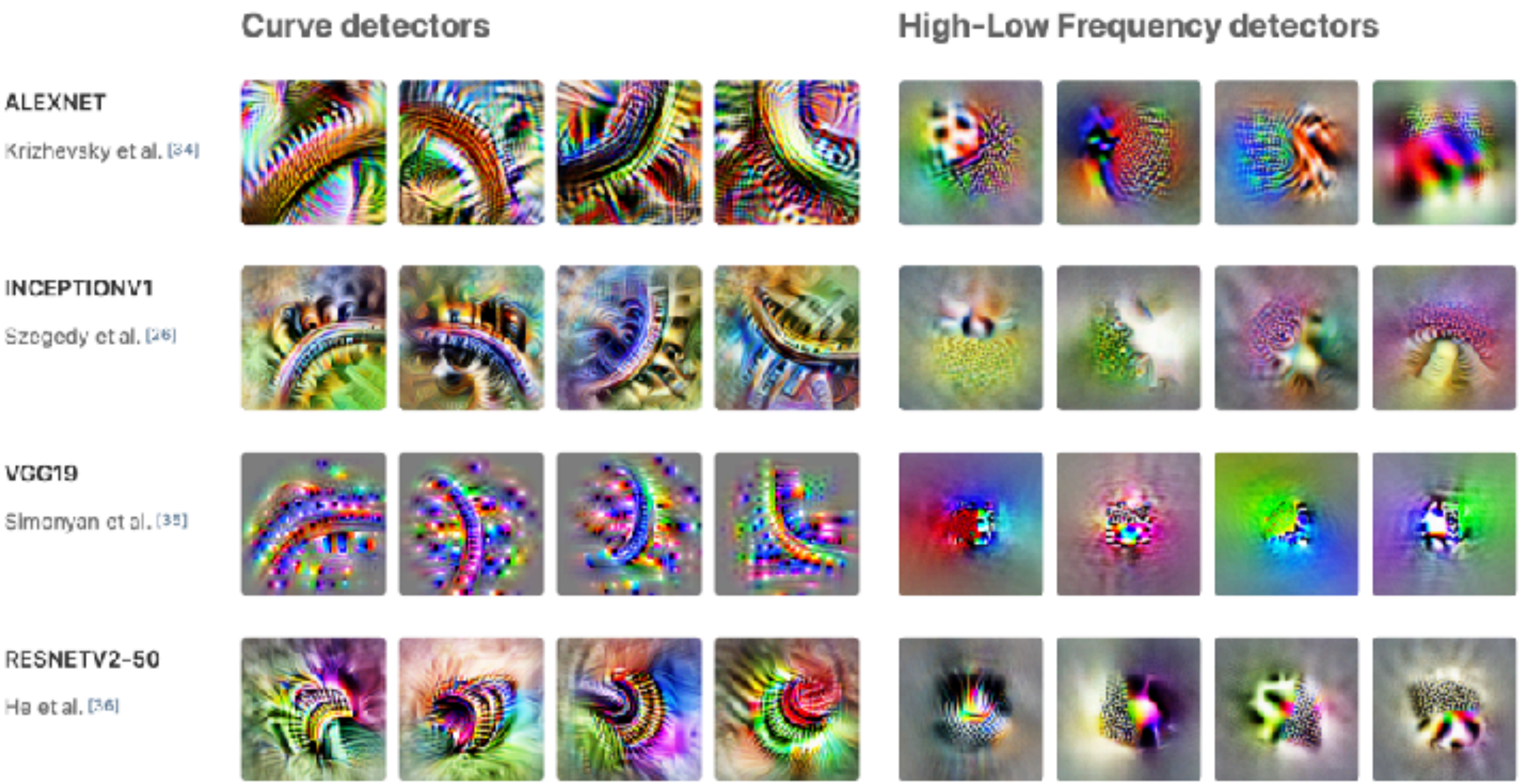
Neural Nets: Directed Graph of Circuits



Voss, et al., "Branch Specialization", Distill, 2021.

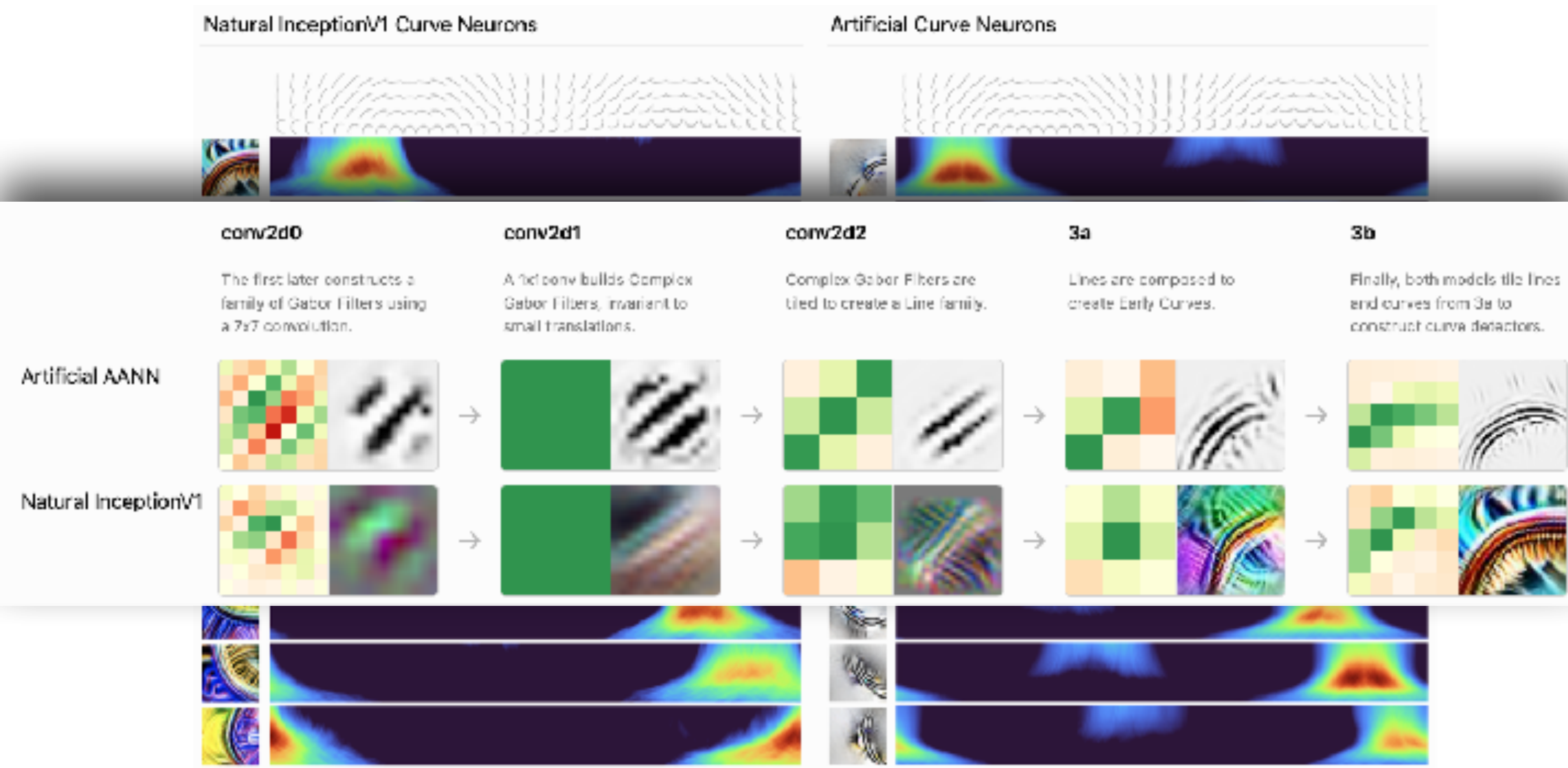
Universality of Circuits

- Analogous features and circuits form across models and tasks



Reverse Engineering a Circuit

- With assumption of what feature is, a circuit can be implemented by hand that nearly identically follows the assumed functionality



Closing Thoughts from OpenAI Researchers

Closing Thoughts

We take it for granted that the microscope is an important scientific instrument. It's practically a symbol of science. But this wasn't always the case, and microscopes didn't initially take off as a scientific tool. In fact, they seem to have languished for around fifty years. The turning point was when Robert Hooke published Micrographia [1], a collection of drawings of things he'd seen using a microscope, including the first picture of a cell.

Our impression is that there is some anxiety in the interpretability community that we aren't taken very seriously. That this research is too qualitative. That it isn't scientific. But the lesson of the microscope and cellular biology is that perhaps this is expected. The discovery of cells was a qualitative research result. That didn't stop it from changing the world.

https://distill.pub/2020/circuits/zoom-in/

Student Paper Presentation

Group Normalization

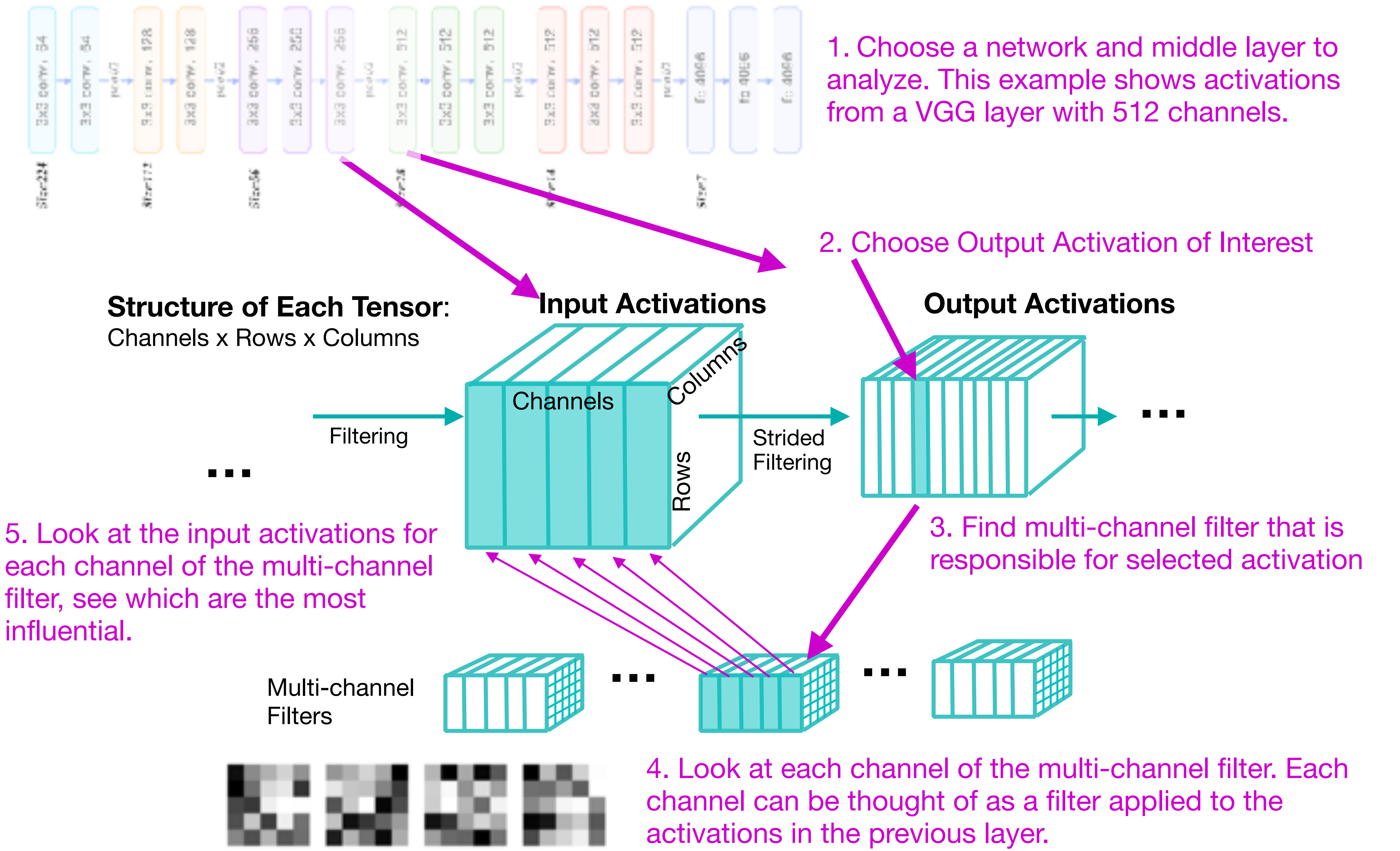
Yuxin Wu Kaiming He

Facebook AI Research (FAIR)

Abstract Batch Normalization (BN) is a milestone technique in the development of deep learning, enabling various networks to train. However, normalizing along the batch dimension introduces problems — BN's error increases rapidly when the batch size becomes smaller, caused by inaccurate batch statistics estimation. This limits BN's usage for training larger models and transferring features to computer vision tasks including detection, segmentation, and video, which require small batches constrained by memory consumption. In this paper, we present Group Normalization (GN) as a simple alternative to BN. GN divides the channels into groups and computes within each group the mean and variance for normalization. GN's computation is independent of batch sizes, and its accuracy is stable in a wide range of batch sizes. On ResNet-50 trained in ImageNet, GN has 10.6% lower error than its BN counterpart when using a batch size of 2; when using typical batch sizes, GN is comparably good with BN and outperforms other normalization variants. Moreover, GN can be naturally transferred from pre-training to fine-tuning. GN can outperform its BN-based counterparts for object detection and segmentation in COCO, and for video classification in Kinetics, showing that GN can effectively replace the powerful BN in a variety of tasks. GN can be easily implemented by a few lines of code.

Lab Three Town Hall

Figure for Circuits Lab



Lecture Notes for

Neural Networks and Machine Learning

CNN Circuits

Next Time:
Fully Convolutional Learning

Reading: Chollet 5.4