E9: 309 ADL 14-12-2020

http://leap.ee.iisc.ac.in/sriram/teaching/ADL2020/

Housekeeping

Midterm project II - Abstract submission deadline 15/12/2020

Presentation deadline - Dec. 29th, 30th (time will be announced)

Recap from previous lecture

- ✴ Analyzing trained neural networks
    - ✓ Hierachical representations

Maximizing activations

Visualizing Higher-Layer Features of a Deep Network

Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent
Dept. IRO, Université de Montréal
P.O. Box 6128, Downtown Branch, Montreal, H3C 3J7, QC, Canada
first.last@umontreal.ca

Technical Report 1341
Département d'Informatique et Recherche Opérationnelle

Learning the input pattern of a trained network

- Choose a trained neural network

- Find input patterns that maximize the activations from that neuron

- Solved using gradient ascent

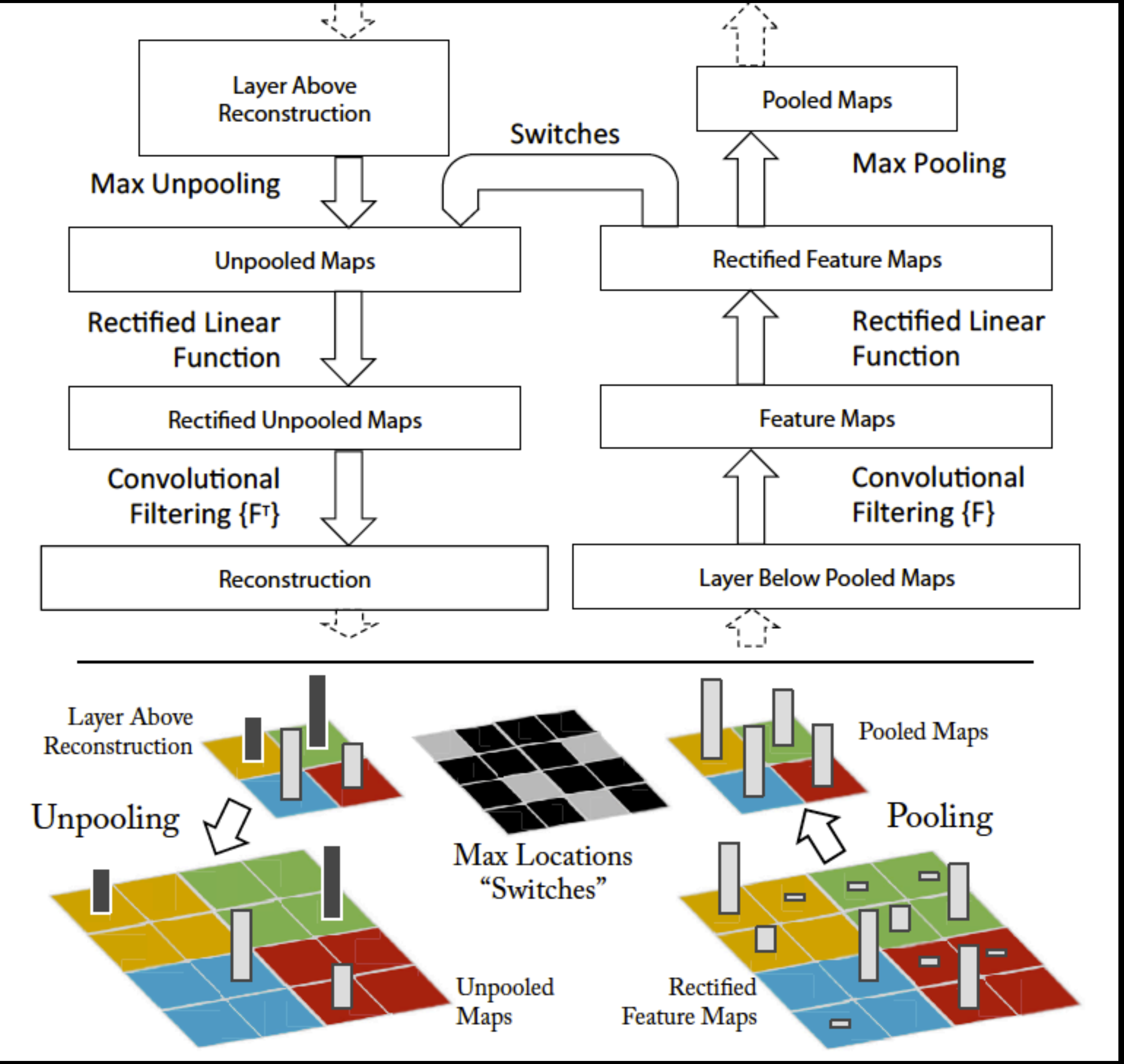
Hierachical representations in deep networks

Visualizing and Understanding Convolutional Networks

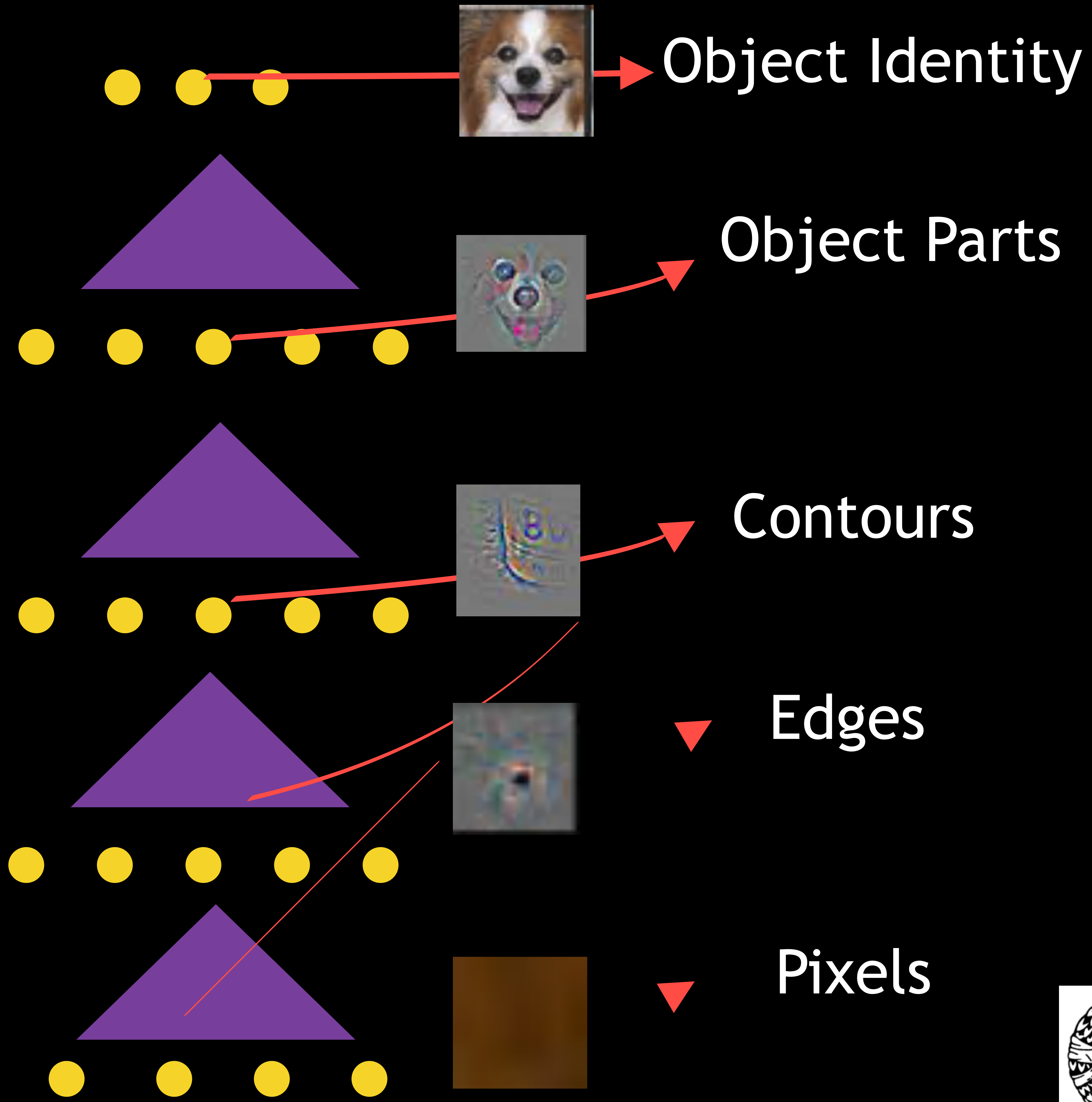
Matthew D. Zeiler and Rob Fergus

Dept. of Computer Science,
New York University, USA
{zeiler,fergus}@cs.nyu.edu

Hierachical representations in deep networks



Hierachical representations in deep networks



[Zeiler, 2014]

UNDERSTANDING HOW DEEP BELIEF NETWORKS PERFORM ACOUSTIC MODELLING

Garcia-Romero, Daniel, et al. "Speaker diarization using deep neural network embeddings." *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.

Abdel-rahman Mohamed, Geoffrey Hinton, and Gerald Penn

2012

Department of Computer Science, University of Toronto

t-SNE embeddings for visualization

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

Visualizing the Hidden Activity of Artificial Neural Networks

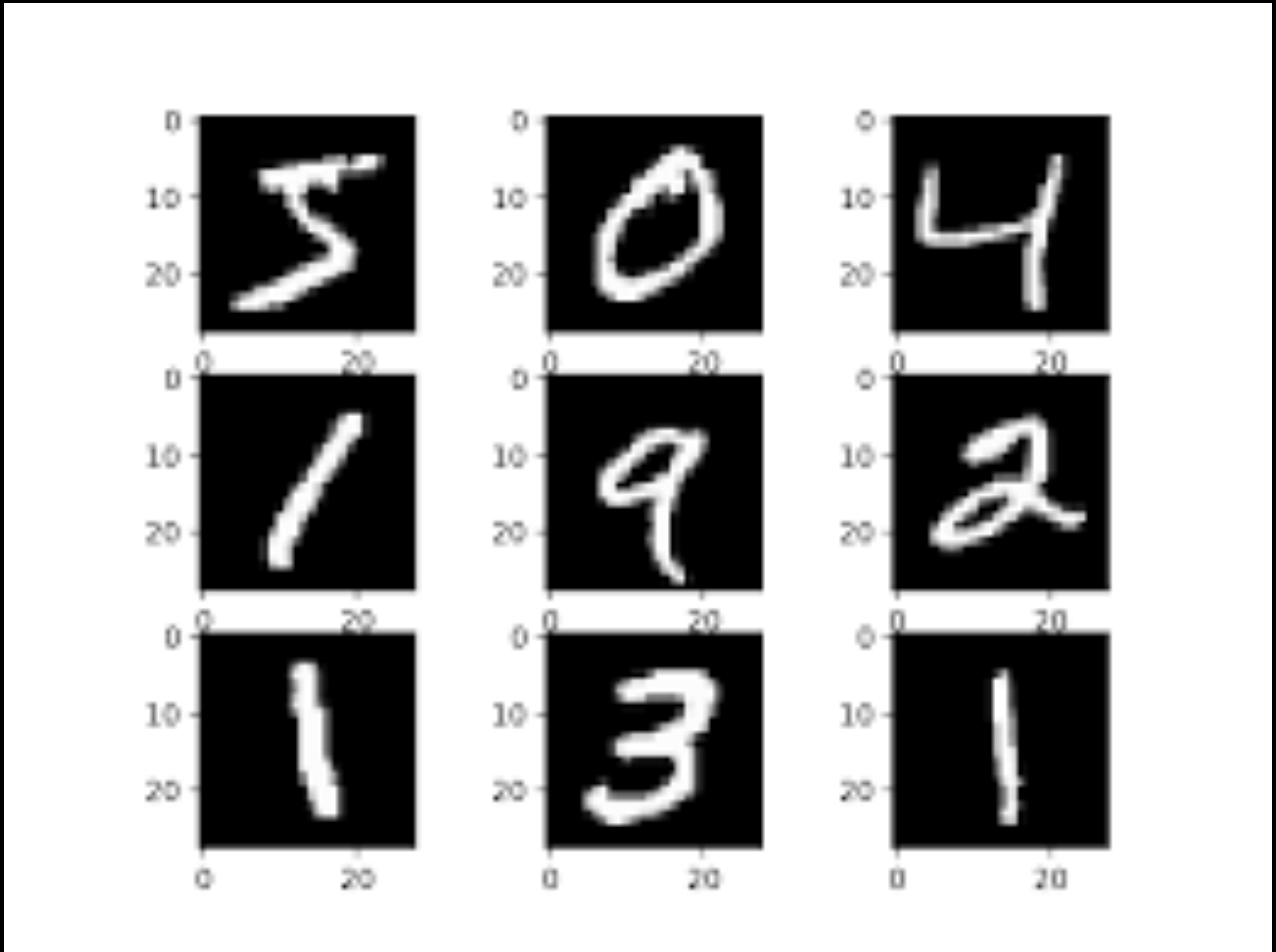
Paulo E. Rauber, Samuel G. Fadel, Alexandre X. Falcão, and Alexandru C. Telea

t-SNE embeddings for visualization

SVHN dataset

t-SNE embeddings for visualization

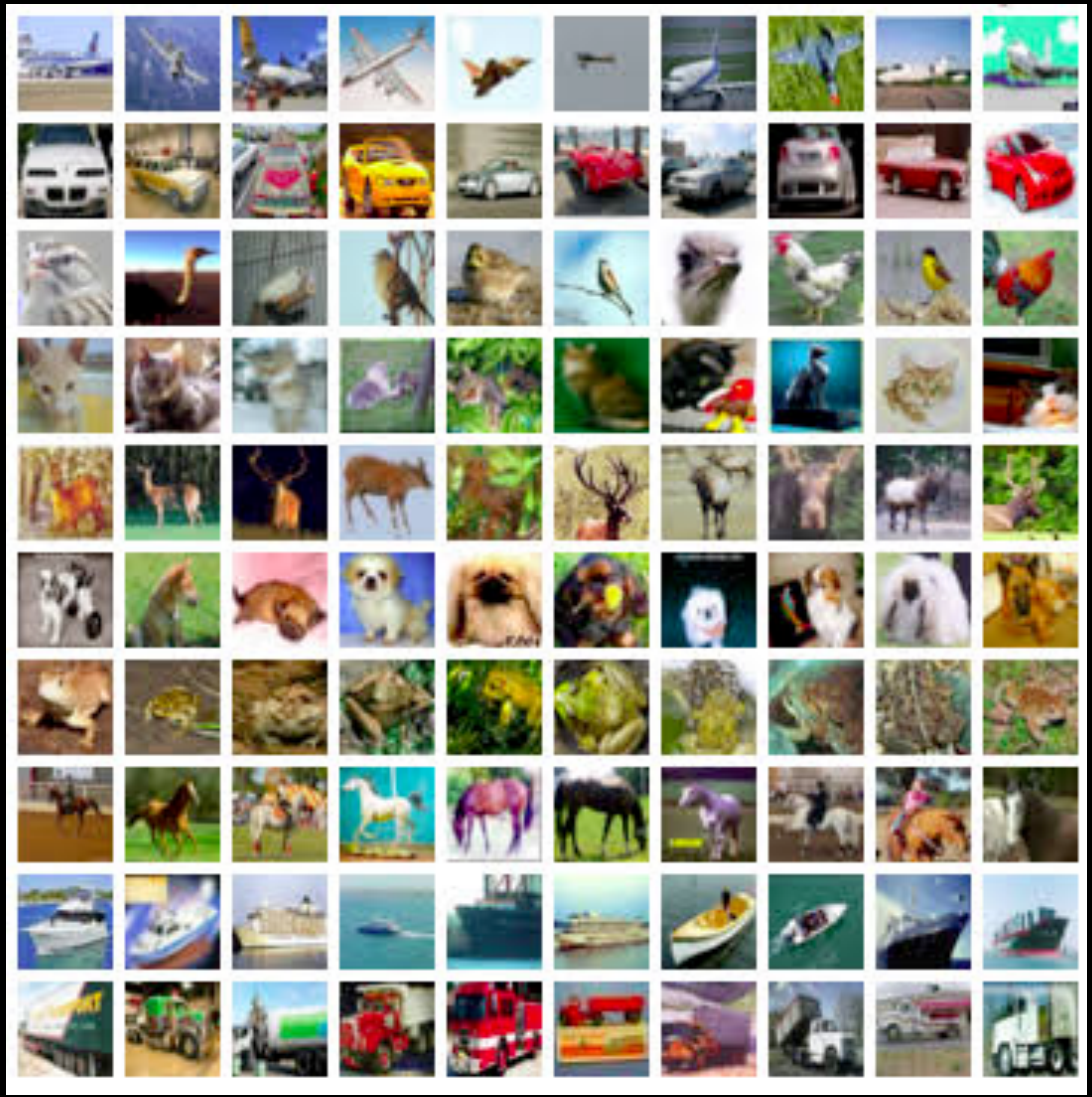
MNIST dataset



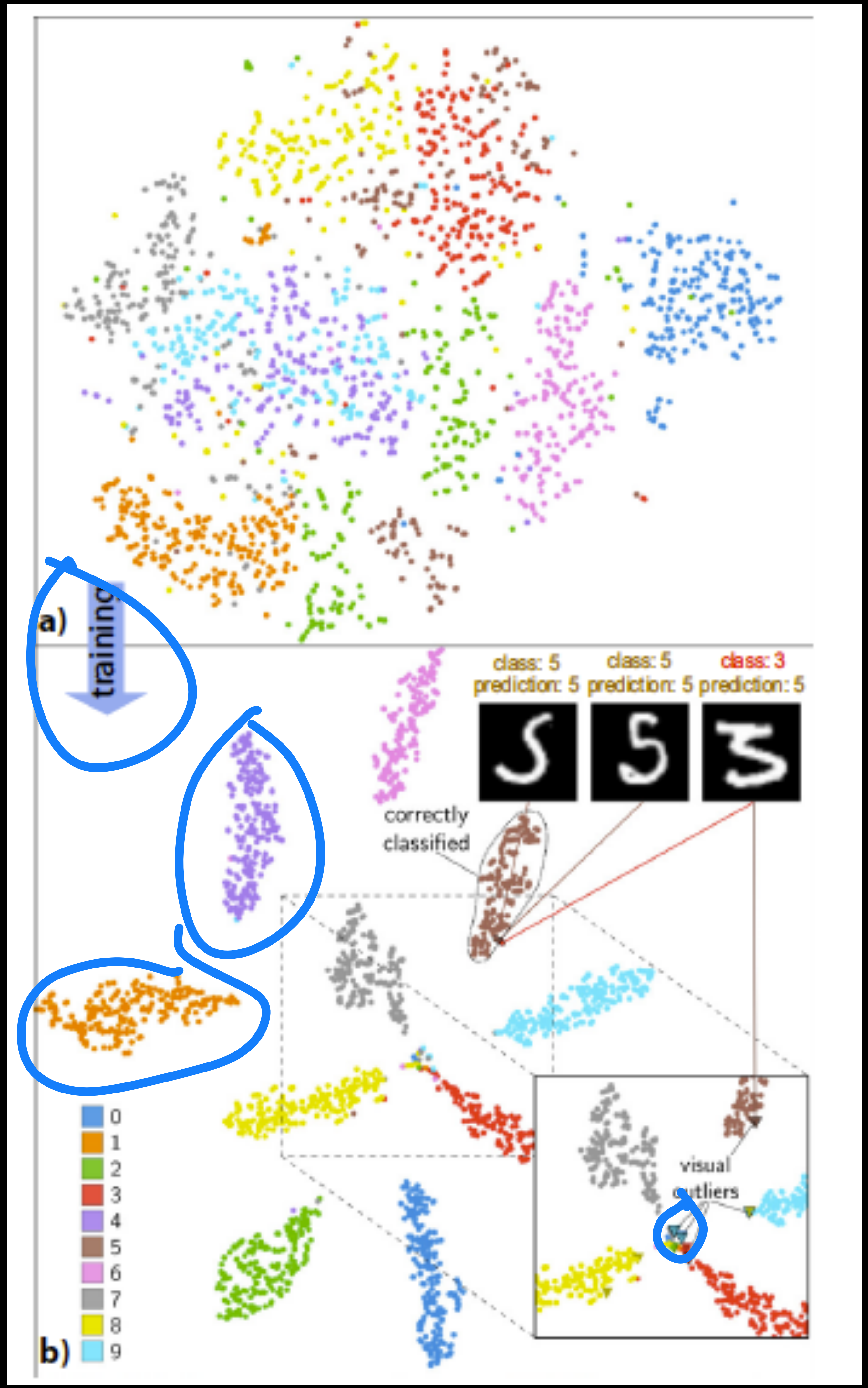
t-SNE embeddings for visualization

CIFA10 dataset

CIFAR 10



Understanding Deep Networks



tSNE
projection
of last layer
of the neural network.

Fig. 3. Projection of the last MLP hidden layer activations, MNIST test subset. a) Before training (NH: 83.78%). b) After training (NH: 98.36%, AC: 99.15%). Inset shows classification of visual outliers.

Understanding Deep Networks

Fig. 4. Projection of the last MLP hidden layer activations before training, *SVHN* test subset (NH: 20.94%). Poor class separation is visible.

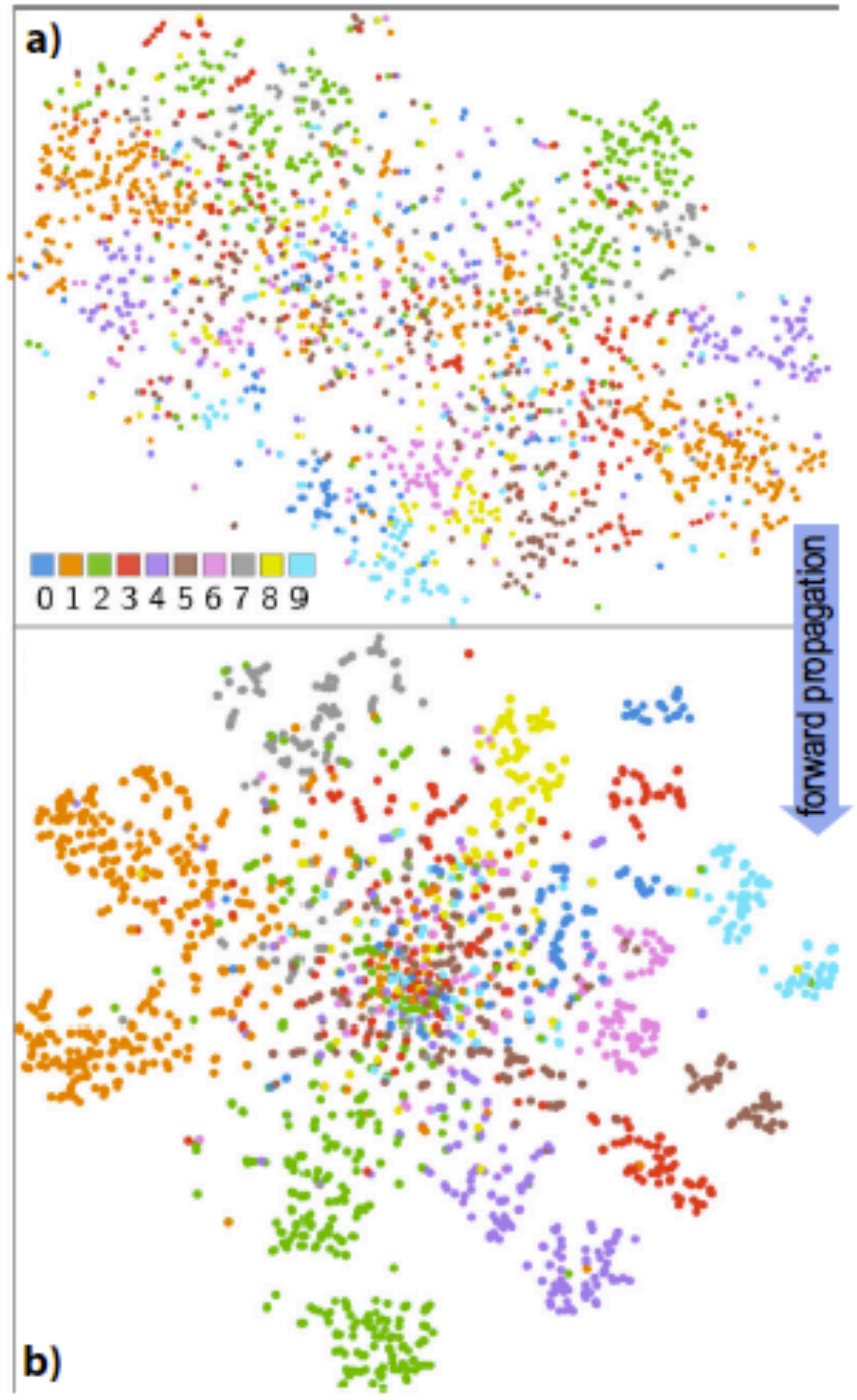


Fig. 5. Projection of the MLP hidden layer activations after training, SVHN test subset. a) First hidden layer (NH: 52.78%). b) Last hidden layer (NH: 67%).

Understanding deep networks

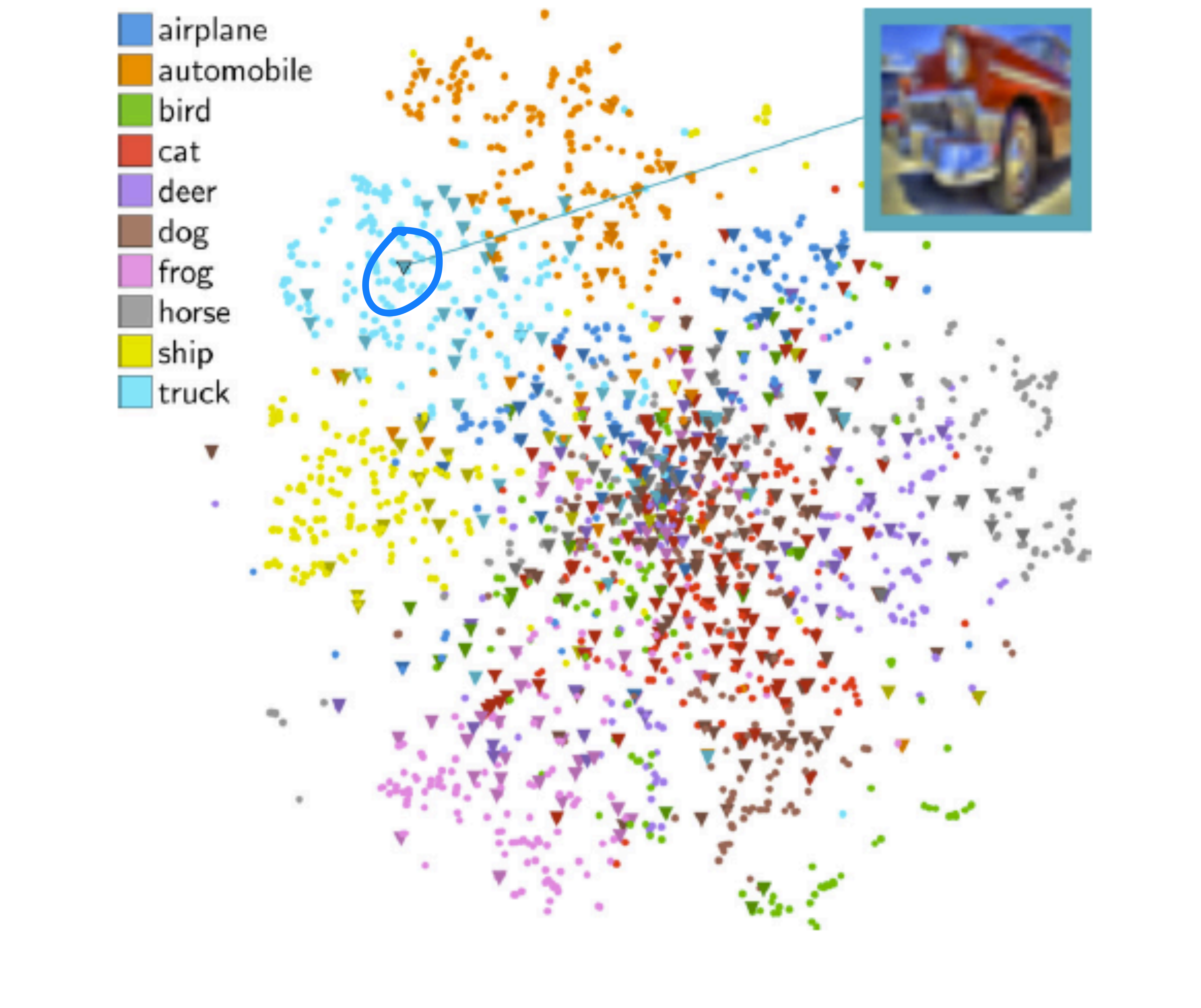


Fig. 9. Projection of last CNN hidden layer activations after training, *CIFAR*-10 test subset (NH: 53.43%, AC: 78.7%).

Understanding Deep Networks

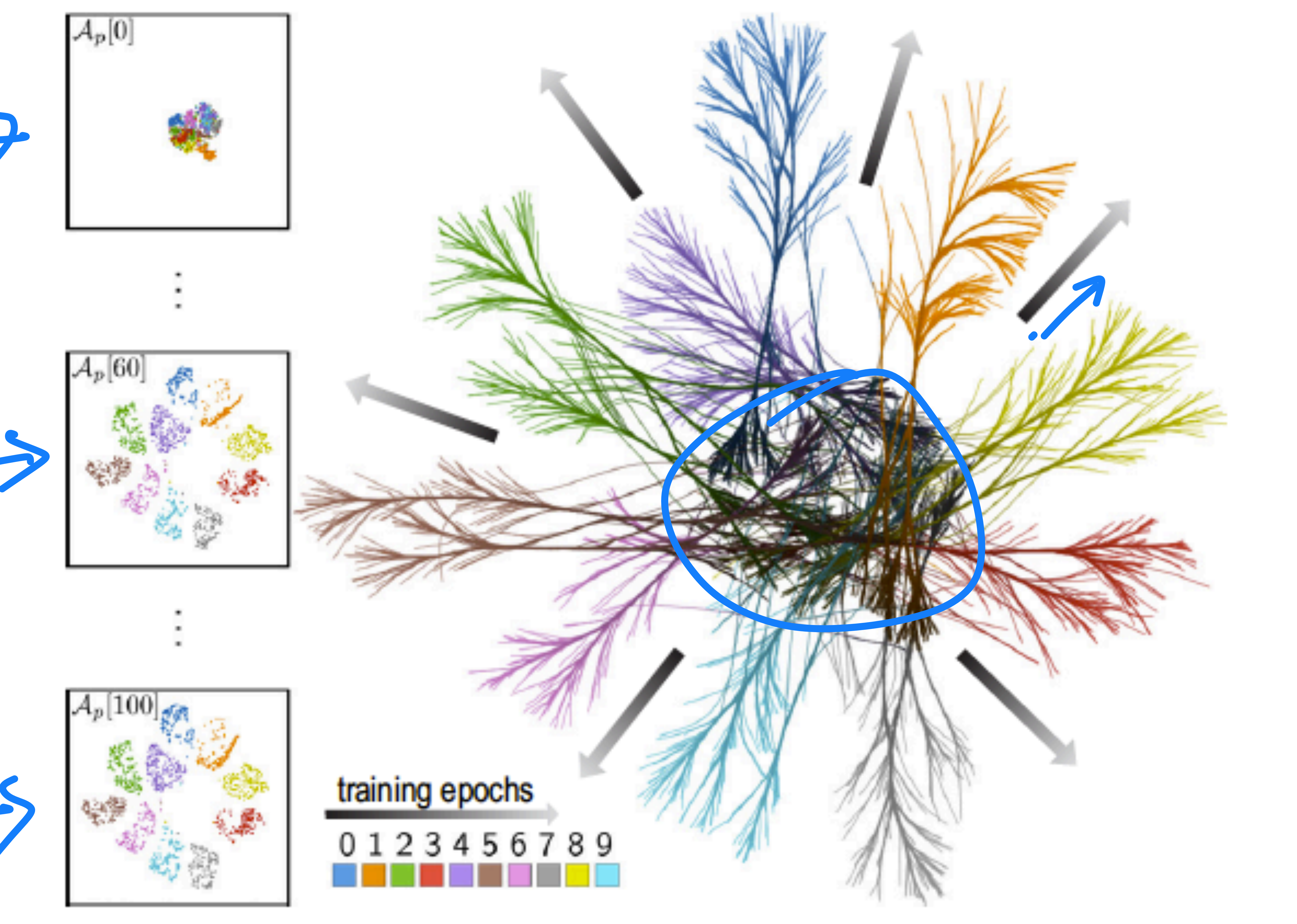


Fig. 11 Inter-epoch evolution, last CNN hidden layer, epochs 0-100, in steps of 20, *MNIST* test subset. Brighter trail parts show later epochs.

Understanding Deep Networks

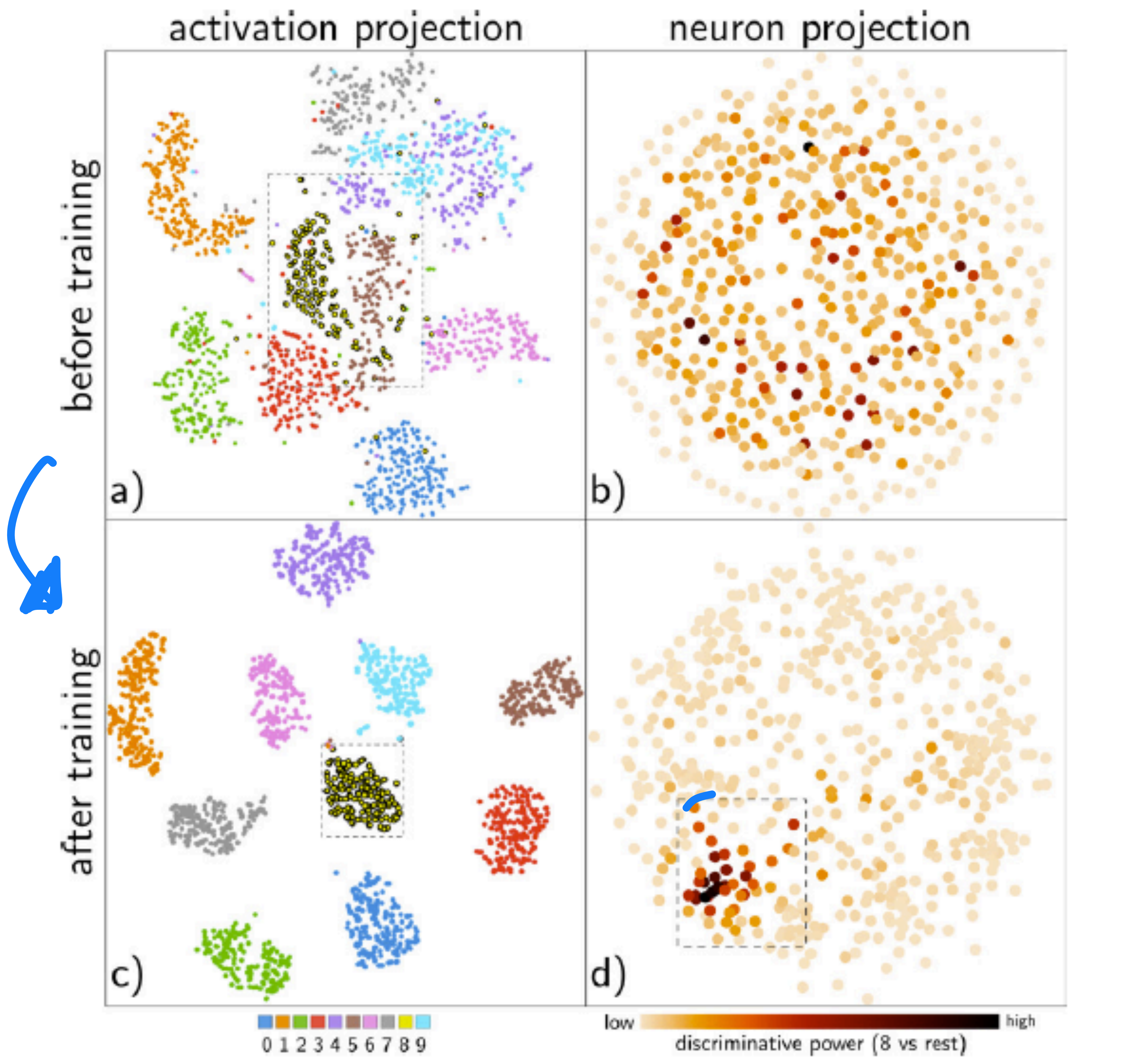
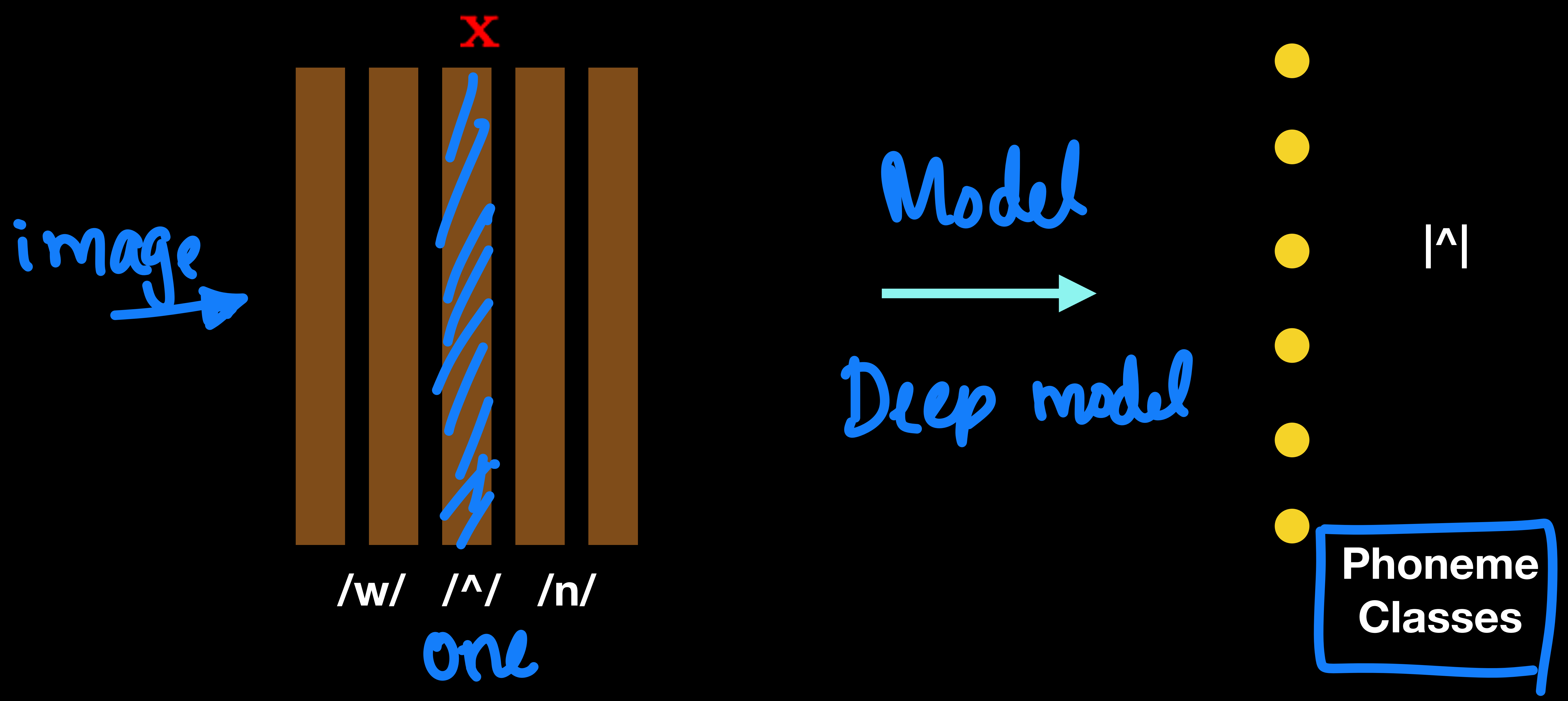


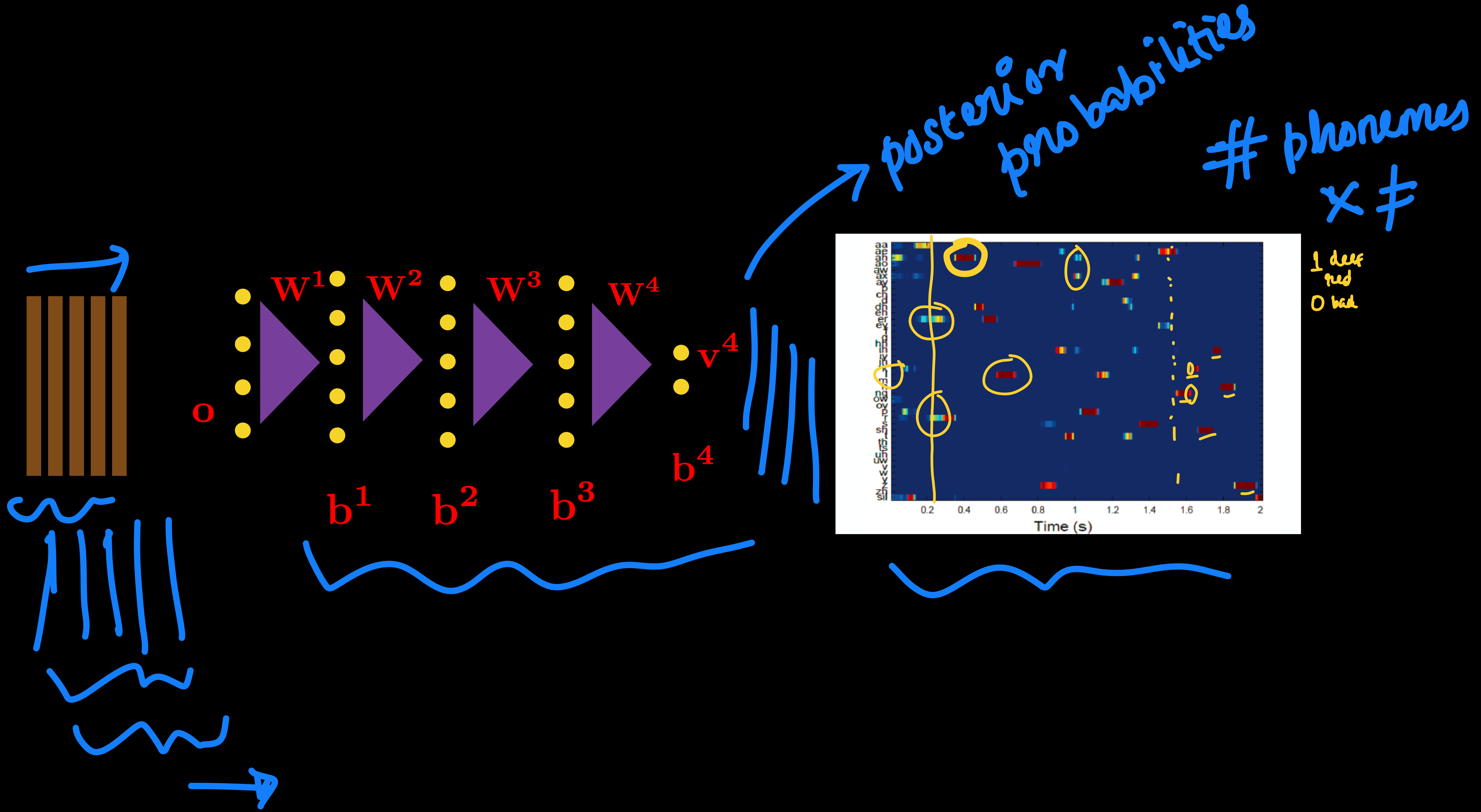
Fig. 12. Activation and neuron projections of last CNN hidden layer activations before and after training, *MNIST* test subset. Neuron projection colors show the neurons' power to discriminate class 8 *vs* rest.

Speech Recognition (Acoustic modeling)



- Classical machine learning - train a classifier on speech training data that maps to the target phoneme class.

Speech recognition



Understanding DNNs for speech

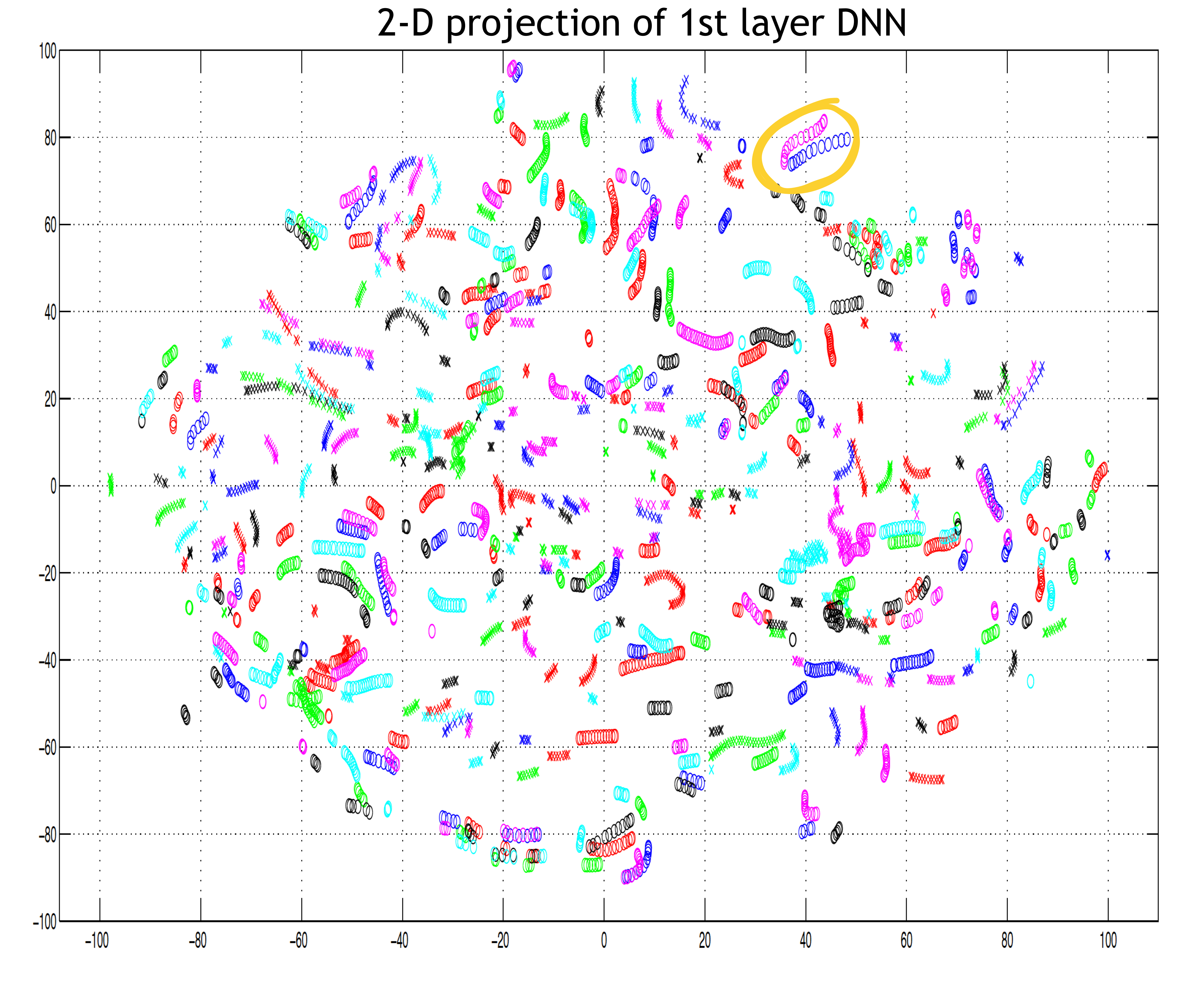
Mohamed, Abdel-rahman, Geoffrey Hinton, and Gerald Penn. "Understanding how deep belief networks perform acoustic modelling." 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2012.

Understanding DNNs for Speech

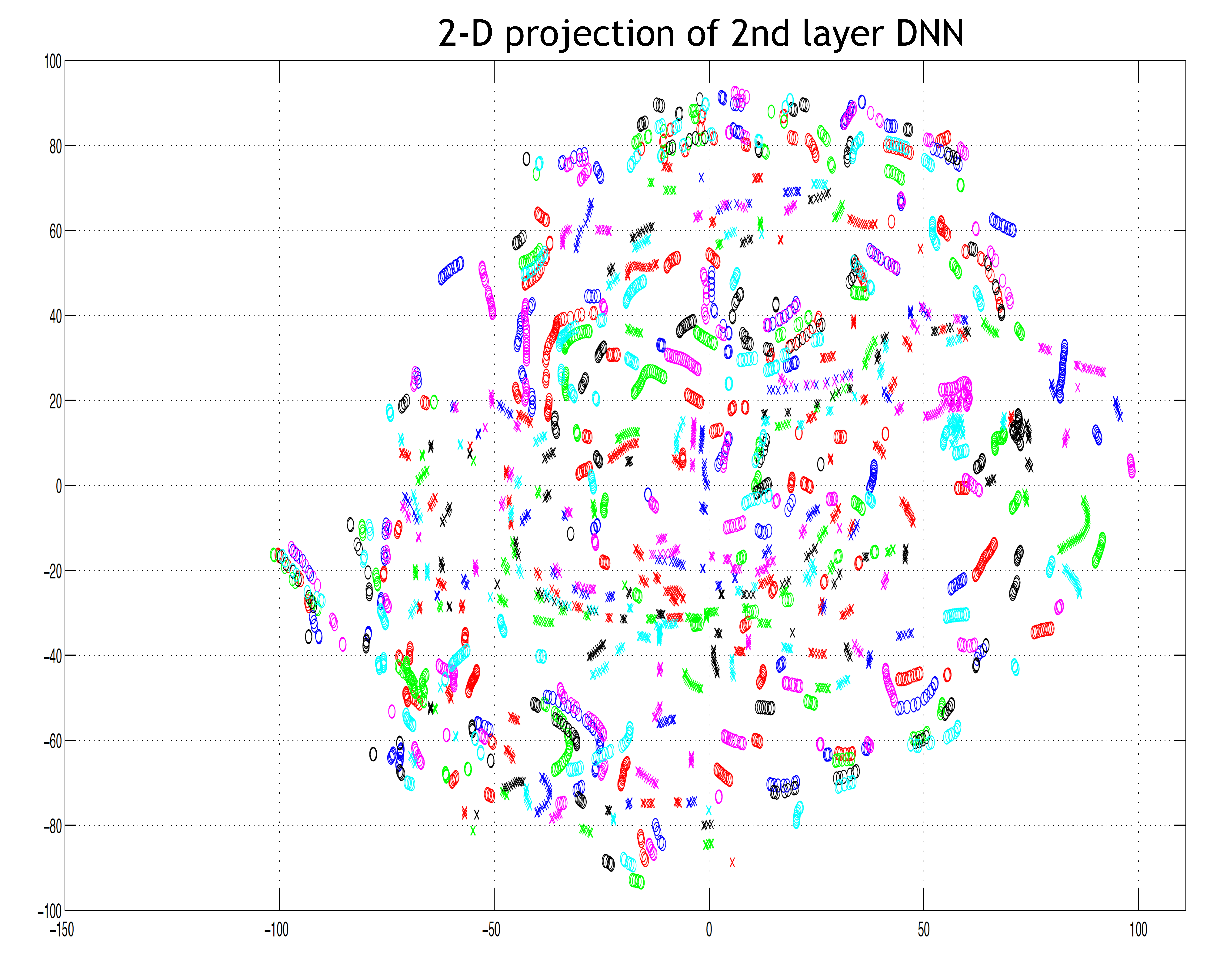
[Abdel Rahman, 2012]

t-SNE

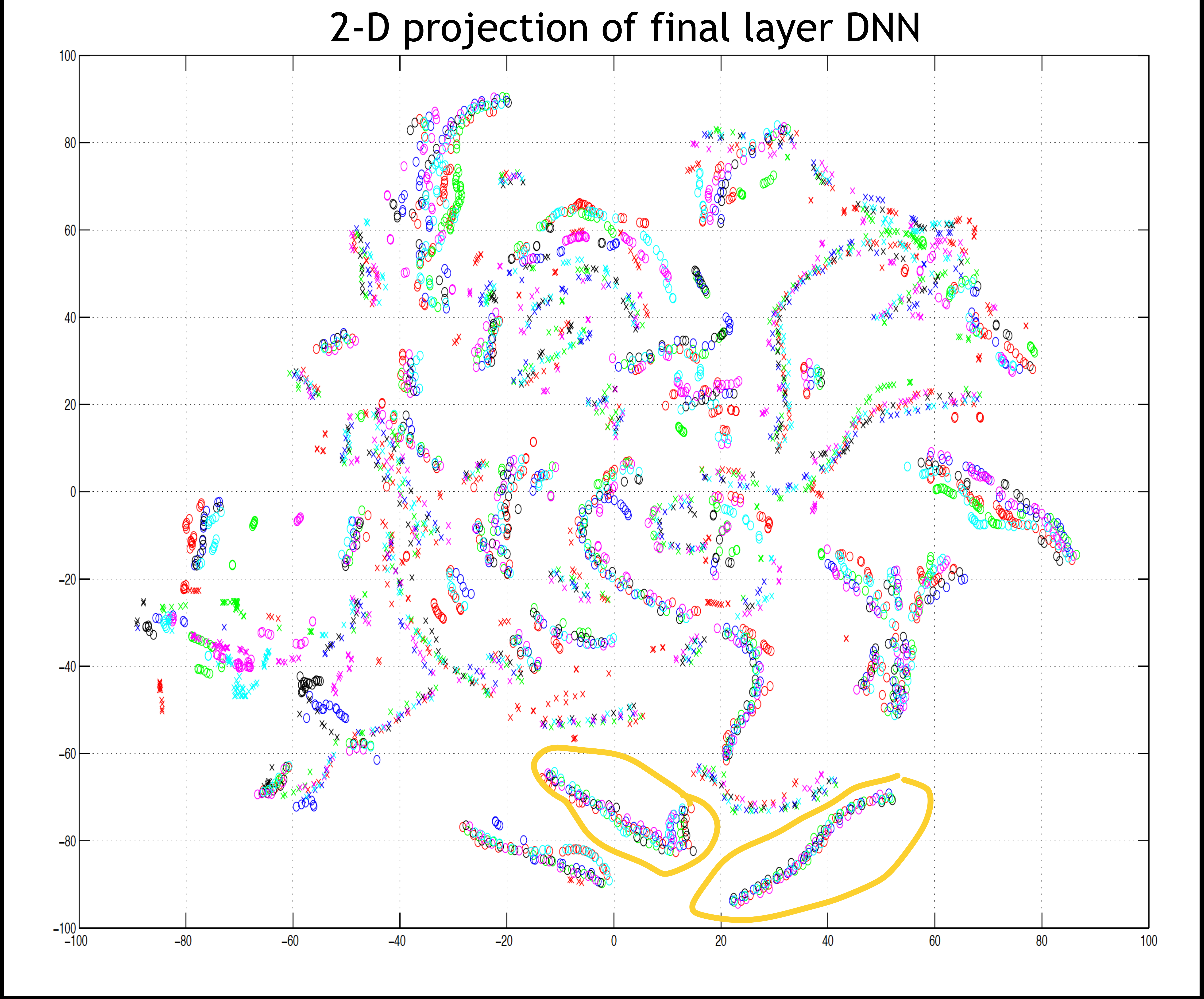
8-layer DNN



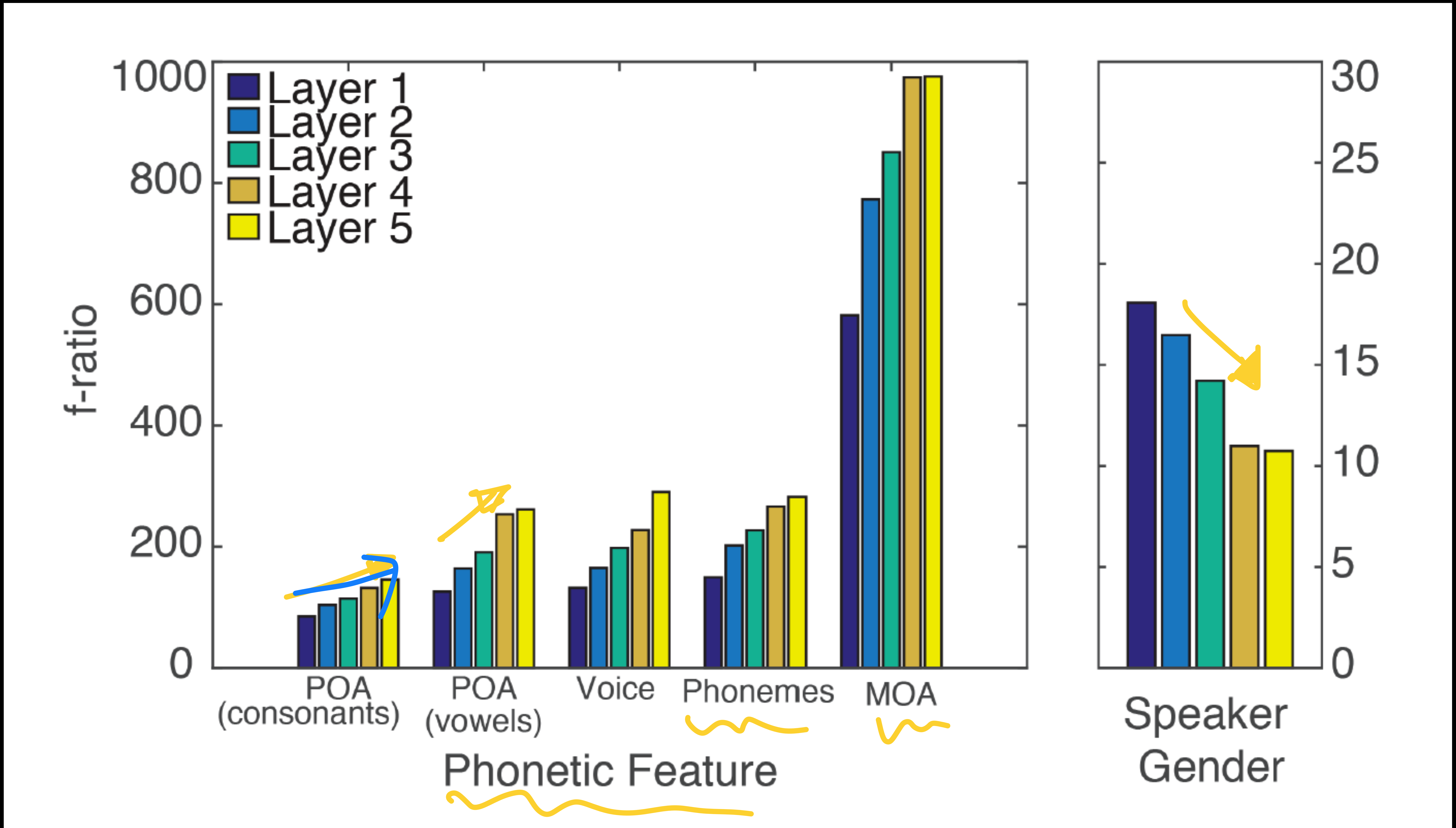
Understanding DNNs for Speech



Understanding DNNs for Speech



Understanding DNNs for Speech



↑ $$\text{f-ratio} = \frac{\|m_1 - m_2\|^2}{Tr(\sigma_1^2 + \sigma_2^2)}$$

↑ separable

m_1, m_2

σ_1^2, σ_2^2

C_1 C_2

Sentence Talker

S_1, T_1 200

→ S_2, T_1

→ S_1, T_2

→ S_2, T_2

Mod

L_1 200

L_2 200

L_3 200

L_4 200

L_5 200

200

P1 P2 P28

F B F

Summary thus far

★ Deep neural networks perform hierarchical data abstractions

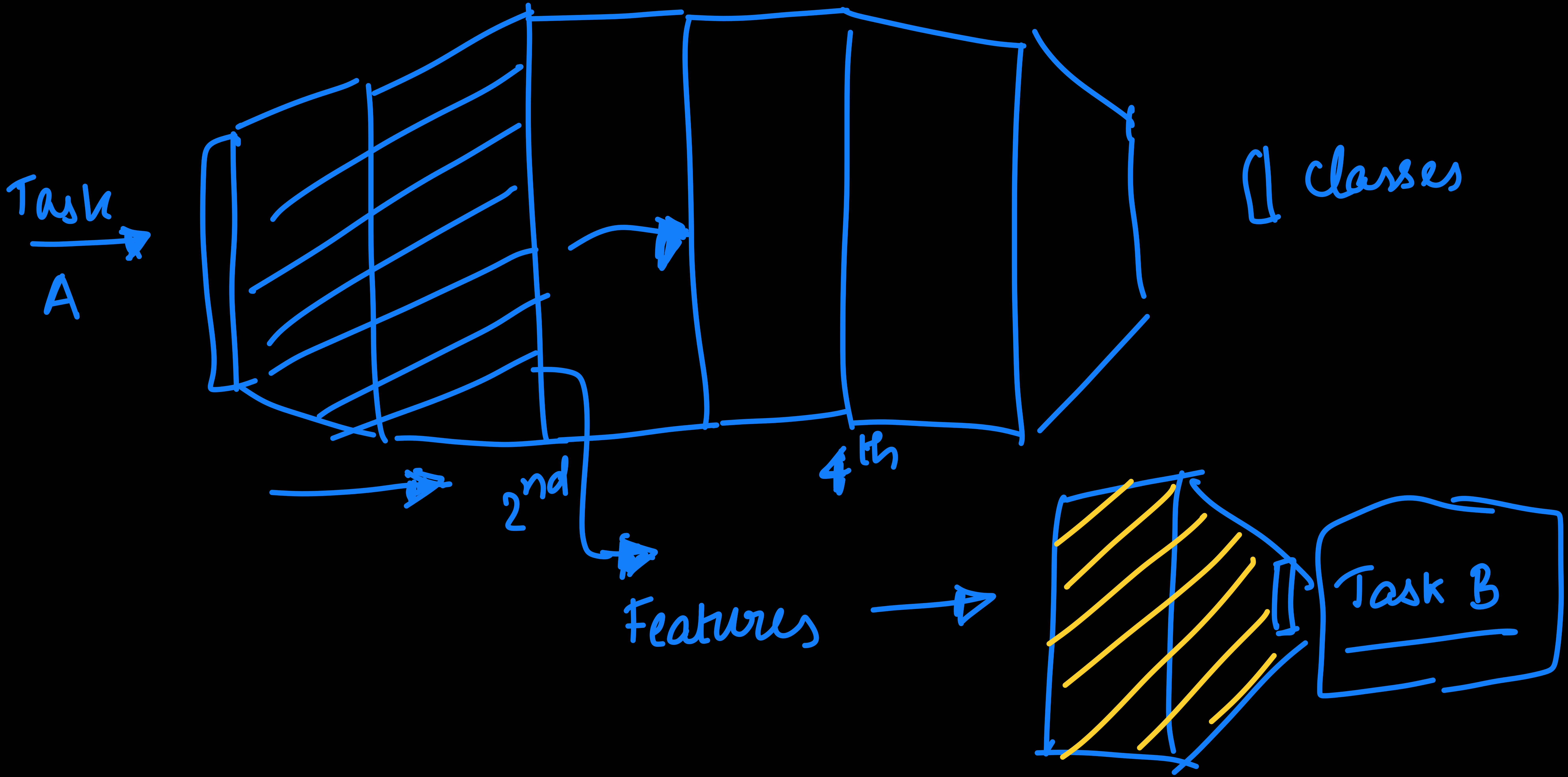
✓ Early layers form representations that are less oriented towards the task.

✓ Later layers form representations that more oriented to the task.

★ Connections with biological processing of audio/images.

Questions about representations

- Can we quantify the degree to which a particular layer is general or specific?
- Does the transition occur suddenly at a single layer, or is it spread out over several layers?
- Where does this transition take place: near the first, middle, or last layer of the network?


Task A
classes
2nd
4th
Features
Task B

Questions about representations

How transferable are features in deep neural networks?

2013

Jason Yosinski,[1] **Jeff Clune,**[2] **Yoshua Bengio,**[3] **and Hod Lipson**[4]
[1] Dept. Computer Science, Cornell University
[2] Dept. Computer Science, University of Wyoming
[3] Dept. Computer Science & Operations Research, University of Montreal
[4] Dept. Mechanical & Aerospace Engineering, Cornell University

Questions about representations

A 500

B 500

beer

pizza

Imagenet Dataset

1000 images
1000 classes

Questions about representations

B3B → B3B⁺
learnt

- A selffer network B3B: the first 3 layers are copied from baseB and frozen. The five higher layers (4–8) are initialized randomly and trained on dataset B. This network is a control for the next transfer network.

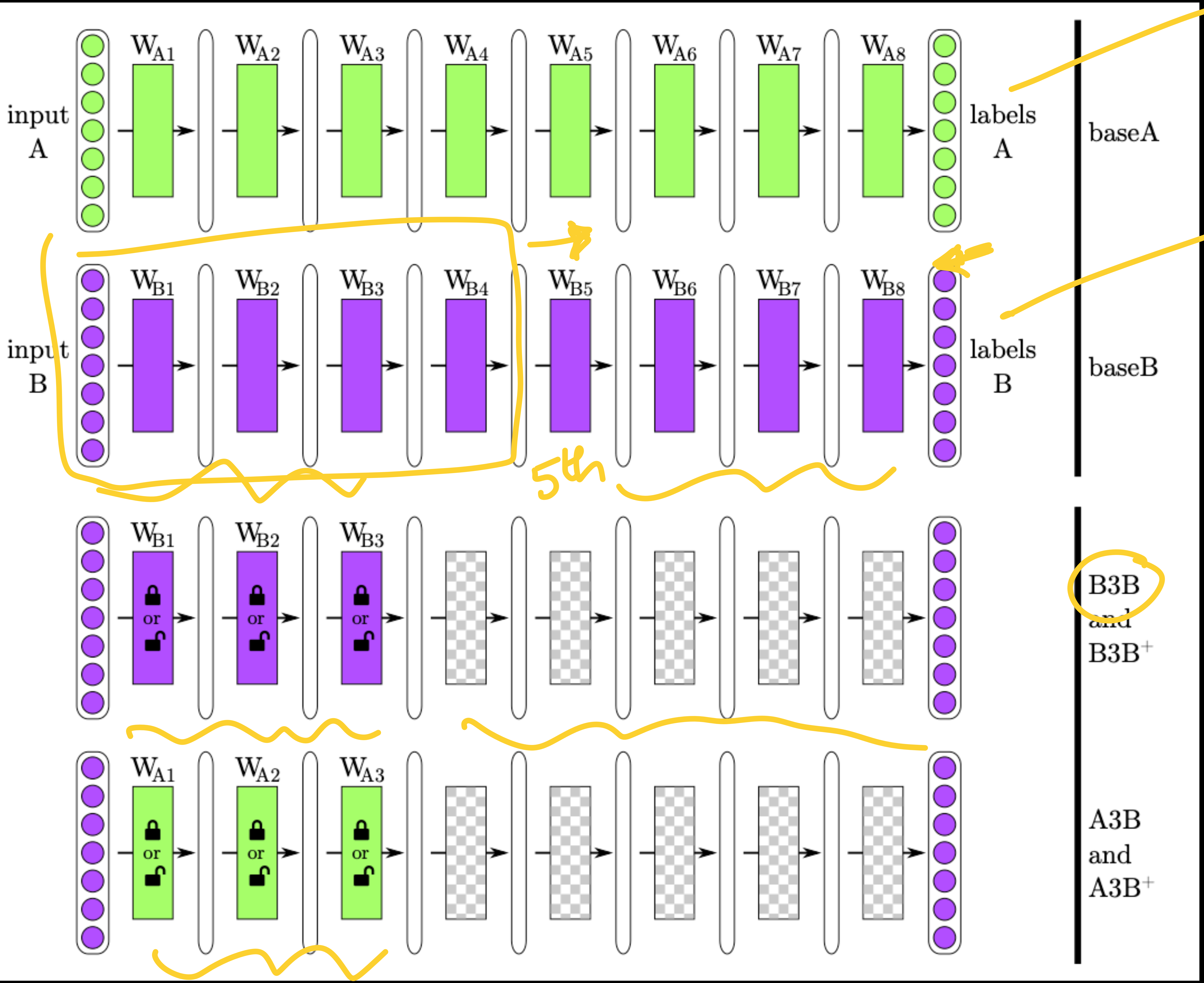
- A transfer network A3B: the first 3 layers are copied from baseA and frozen. The five higher layers (4–8) are initialized randomly and trained toward dataset B. Intuitively, here we copy the first 3 layers from a network trained on dataset A and then learn higher layer features on top of them to classify a new target dataset B.

Questions about representations

- Can we quantify the degree to which a particular layer is general or specific?
- Does the transition occur suddenly at a single layer, or is it spread out over several layers?
- Where does this transition take place: near the first, middle, or last layer of the network?

Questions about representations

DCN



5th

500

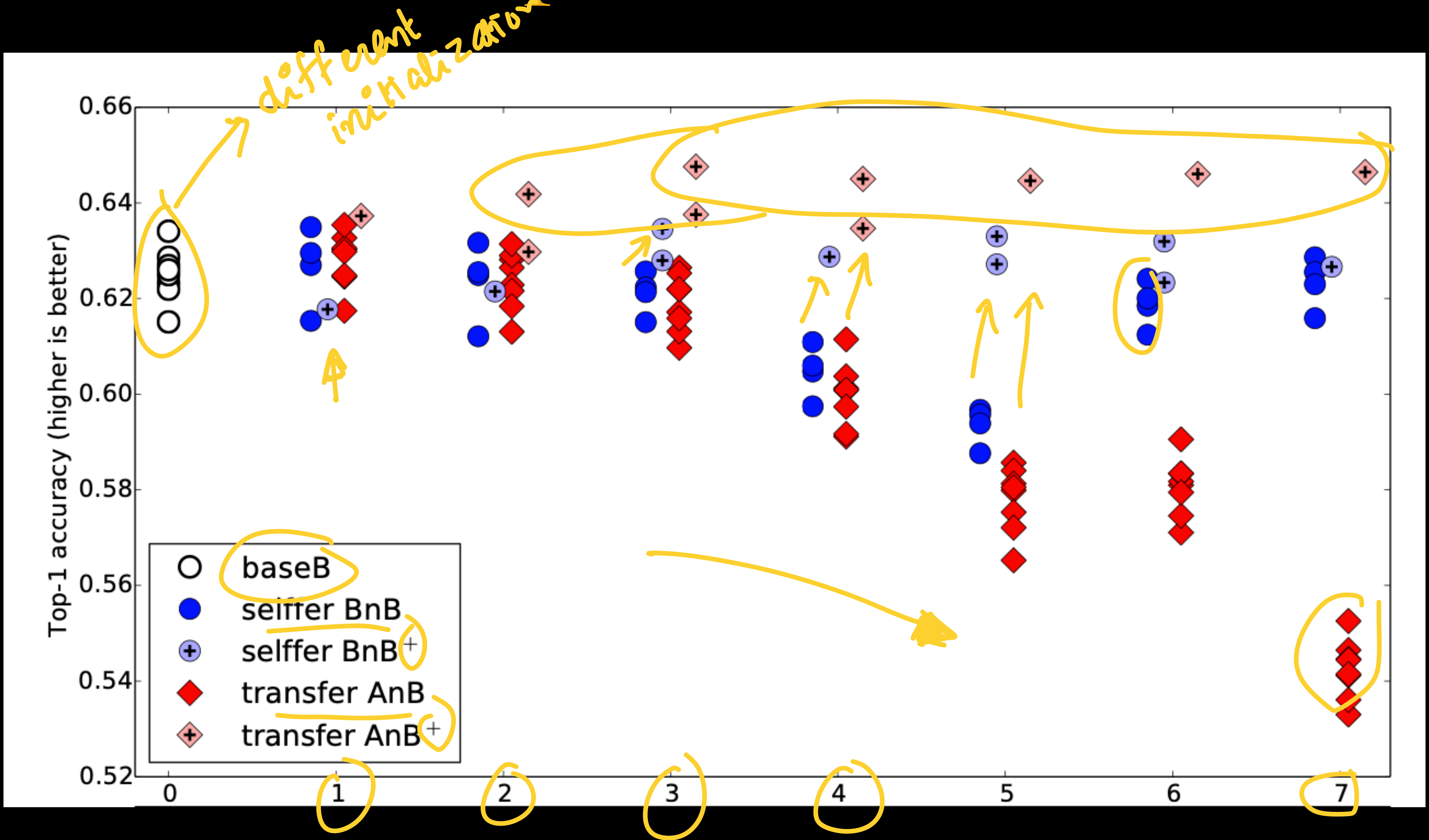
500

B3B

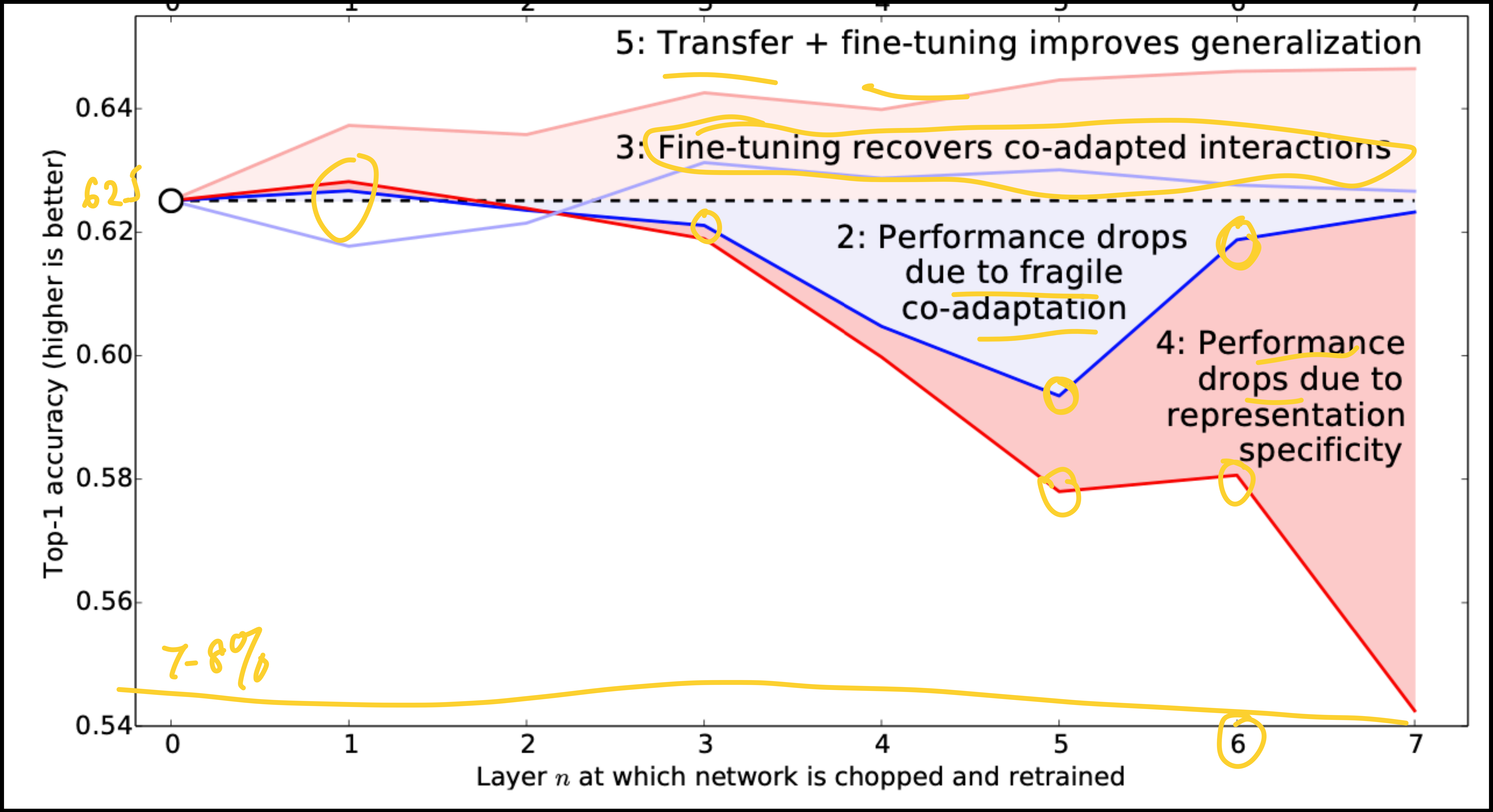
B3B → B3B$^+$

A3B → A3B$^+$

Questions about representations

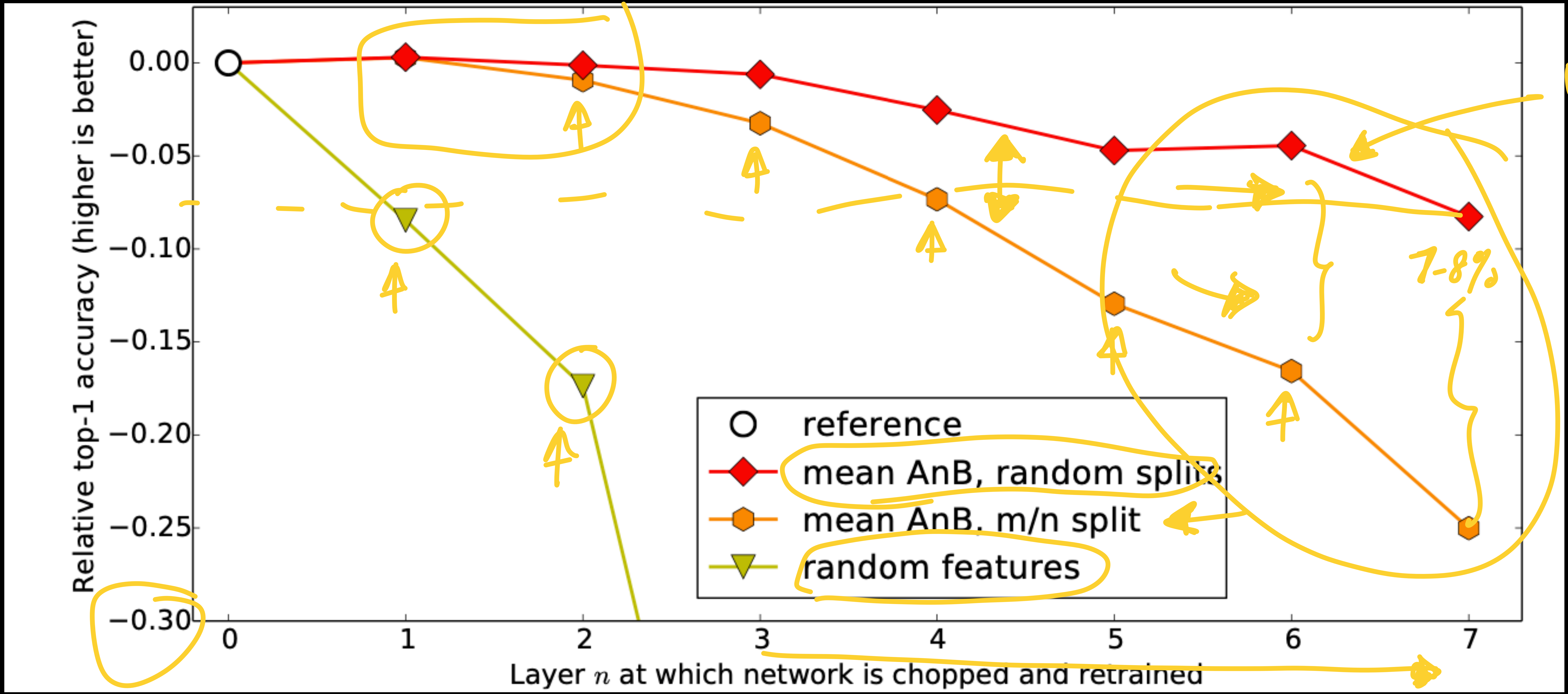


Questions about representations



Questions about representations

living - 550
non-living - 450



Red figure

Transfer learned