Angular Visual Hardness

Presenter: Beidi Chen

Rice University

Collaborators: Weiyang Liu, Animesh Garg, Zhiding Yu, Anshumali Shrivastava, Anima Anandkumar

Gap between human visual system and CNNs

Easy for Human and **Hard** for CNNs

Class Name

Golf Ball
0.001
1.0

Softmax Score

Radio
0.001
1.0

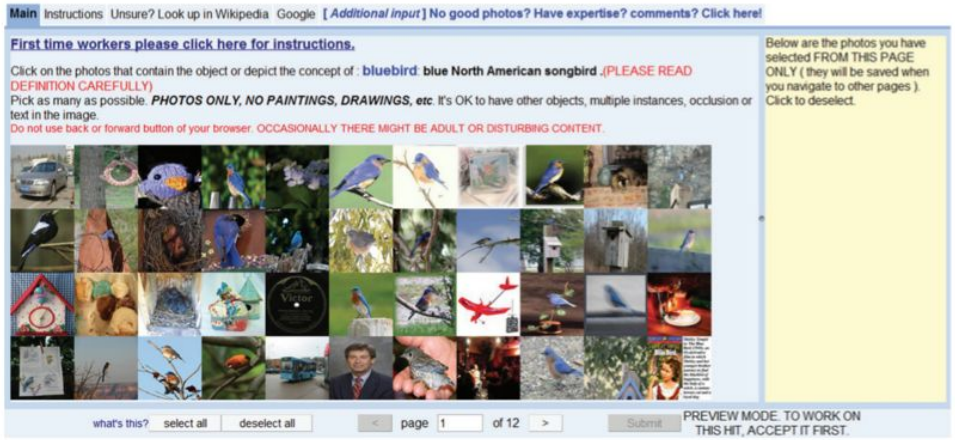
Human Selection Frequency

Hard for Human and **Easy** for CNNs

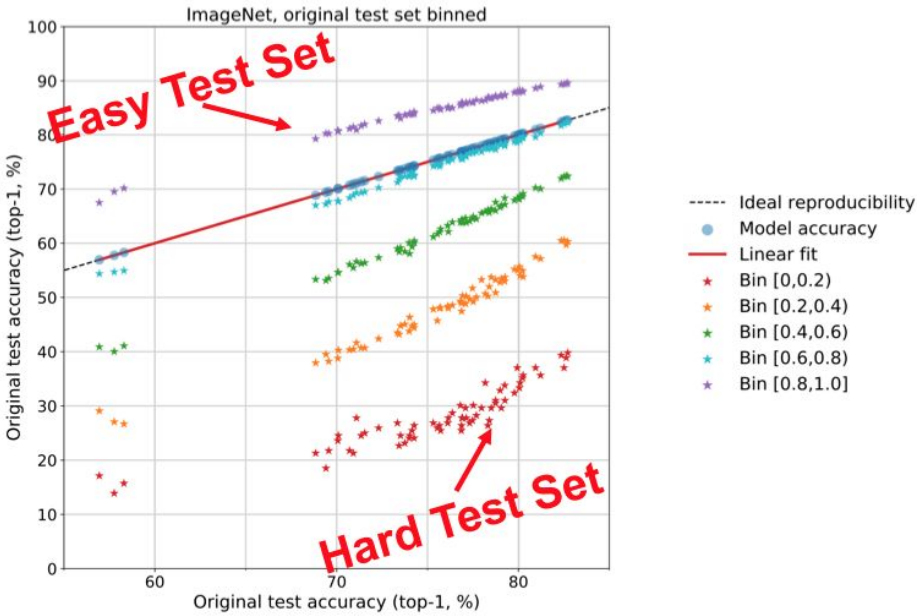
Nail
0.93
0.2

Oil Filter
0.998
0.2

Inspiration: Do ImageNet Classifiers Generalize to ImageNet?



Human Labelling Interface



Recht et al. “Do ImageNet Classifiers Generalize to ImageNet?” ICML 2019

Loss function of CNNs in visual recognition

- Softmax cross-entropy loss - one of the most popular loss functions in CNN

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)$$

where,

$$L_i = -\log\left(\frac{e^{\|\boldsymbol{W}_{y_i}\|\|\boldsymbol{x}_i\|\cos(\theta_{y_i})}}{\sum_j e^{\|\boldsymbol{W}_j\|\|\boldsymbol{x}_i\|\cos(\theta_j)}}\right)$$

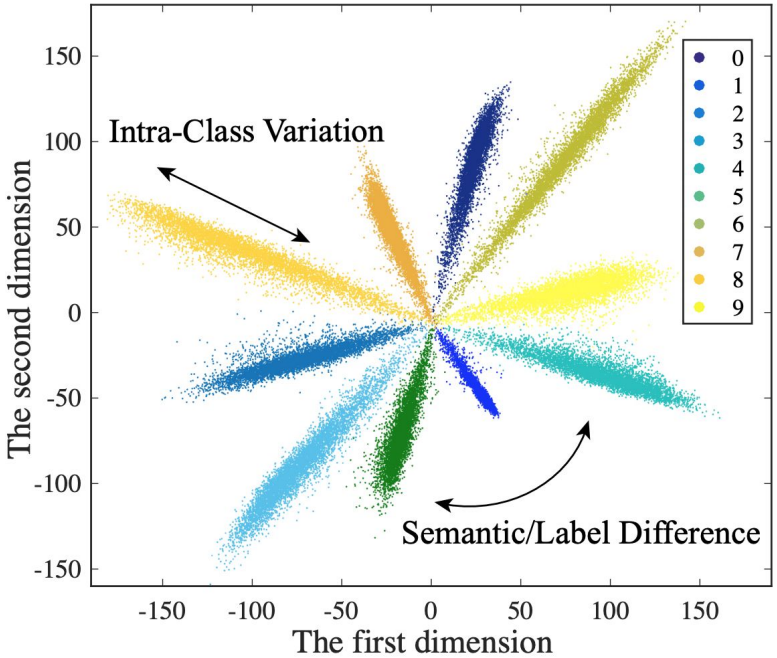
The magnitude information

The angle between feature and classifier

Model Confidence

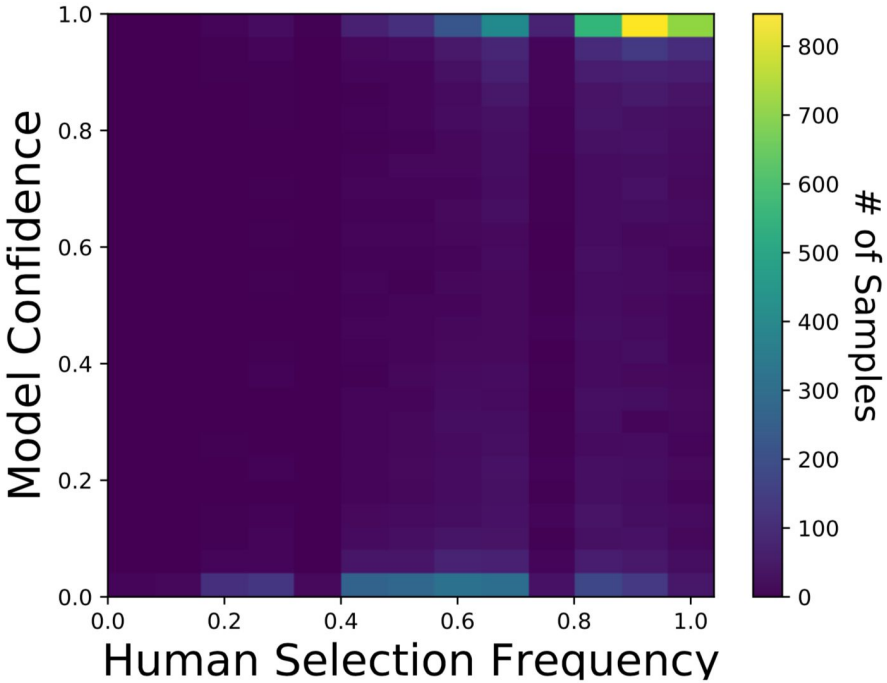
2D feature embedding on MNIST

- Deeply learned features are naturally decoupled with angle and norm.
- The angles reflect the semantic difference.



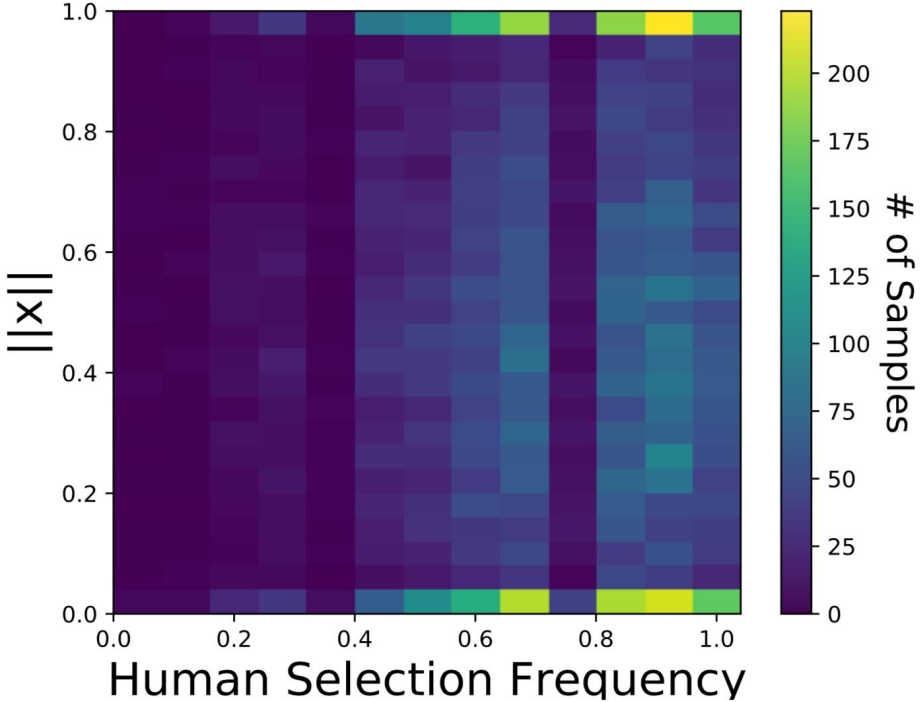
Model confidence is not aligned with human frequency

- Human Selection Frequency vs. Model Confidence in ResNet50



Magnitude does not correlate with human frequency

- Human selection frequency vs. feature L2 norm



Bridging the gap between human visual hardness and model predictions -- Angular Visual Hardness

- Definition of angular visual hardness (AVH):

  Given a sample x with label y:

$$\boldsymbol{AVH}(x) = \frac{\mathcal{A}(x, w_y)}{\sum_{i=1}^{C} \mathcal{A}(x, w_i)}$$

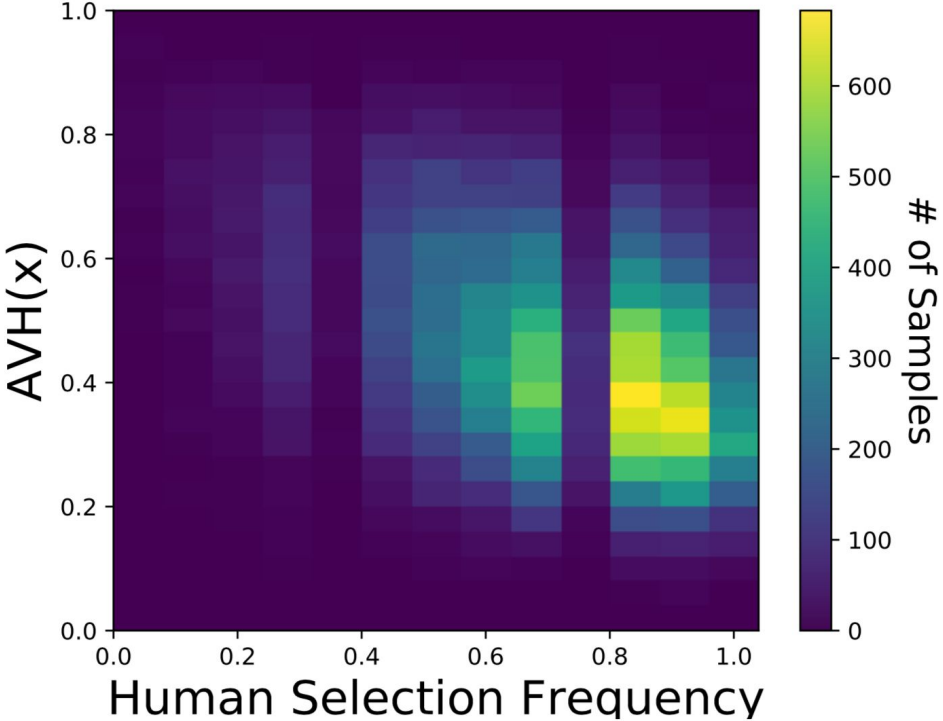
where,

$$\mathcal{A}(\boldsymbol{u}, \boldsymbol{v}) = \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|}\right)$$

w_i is the classifier for the i-th class.

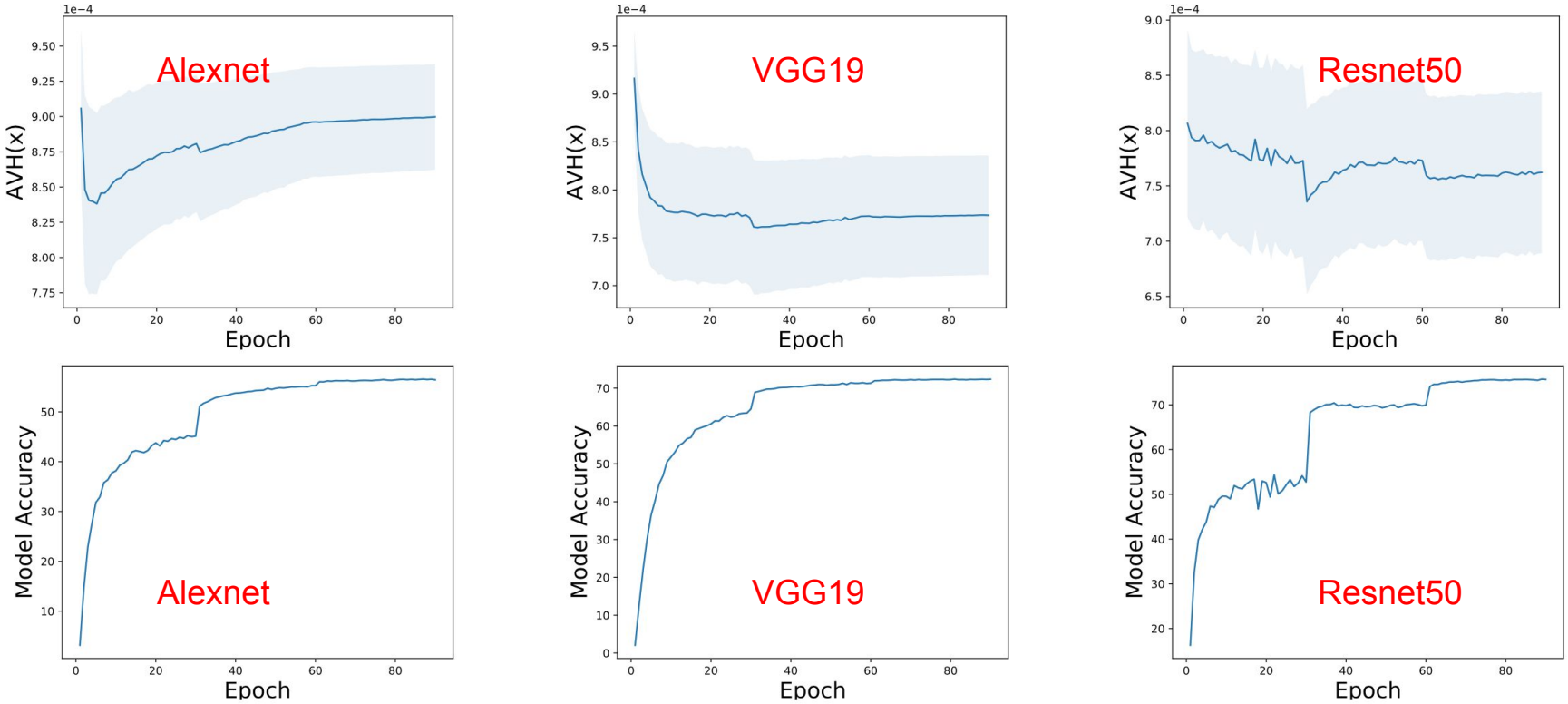
AVH is well aligned with human frequency

- Human selection frequency vs. AVH

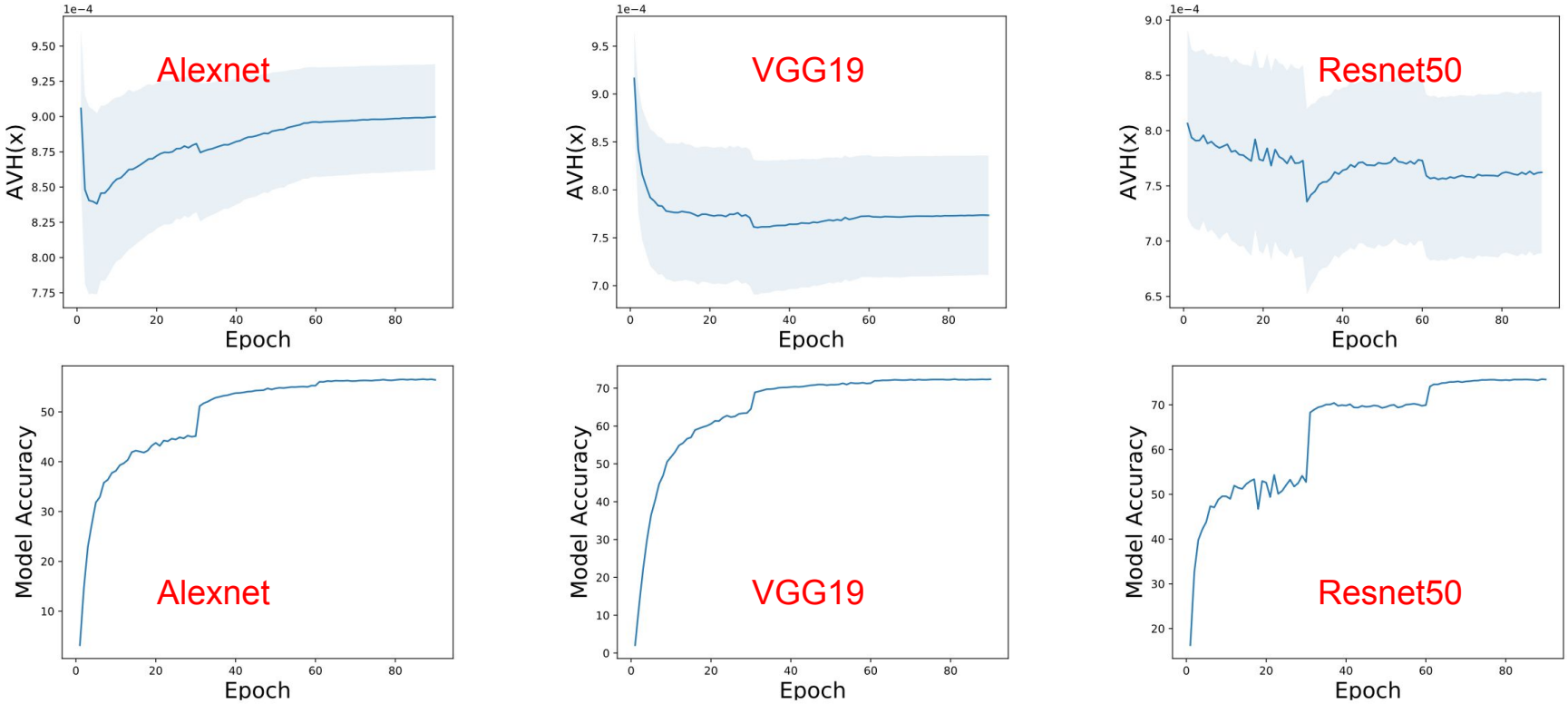


What role does AVH play during the training process?

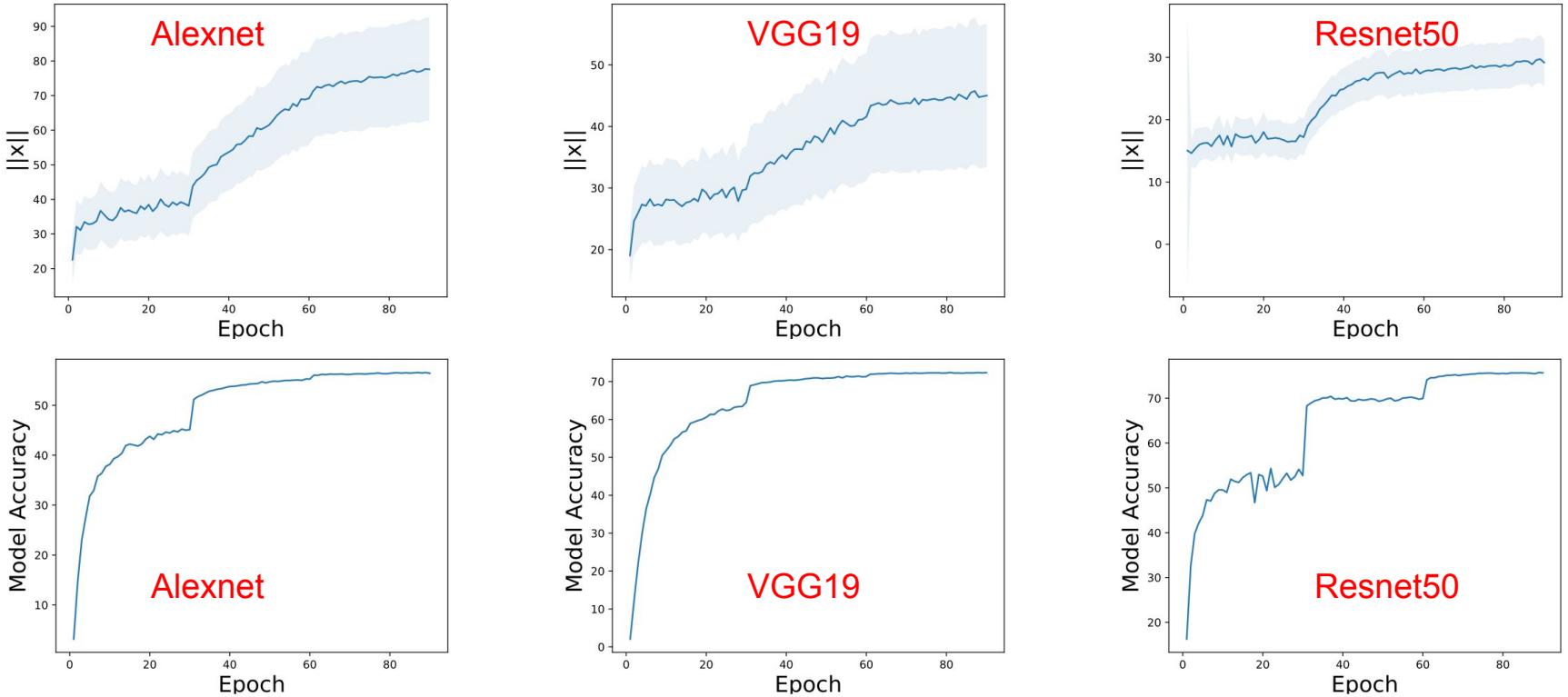
AVH hits a plateau very early even when the accuracy or loss is still improving.



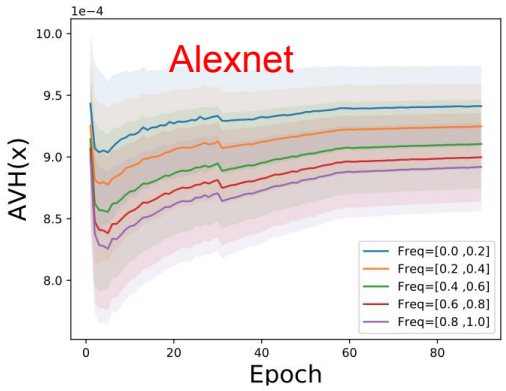
AVH is an indicator of model's generalization ability

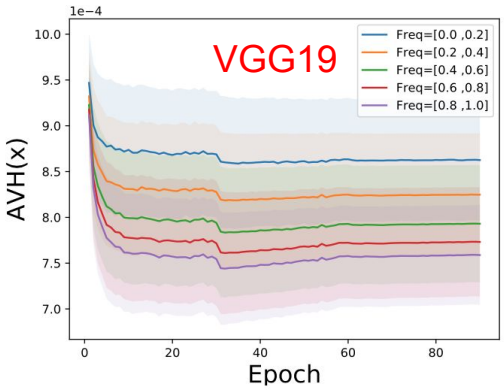


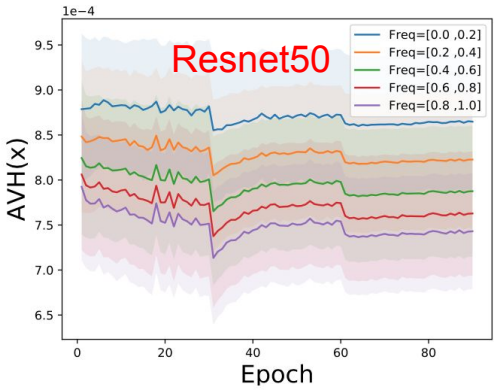
The norm of feature embeddings keeps increasing during training.



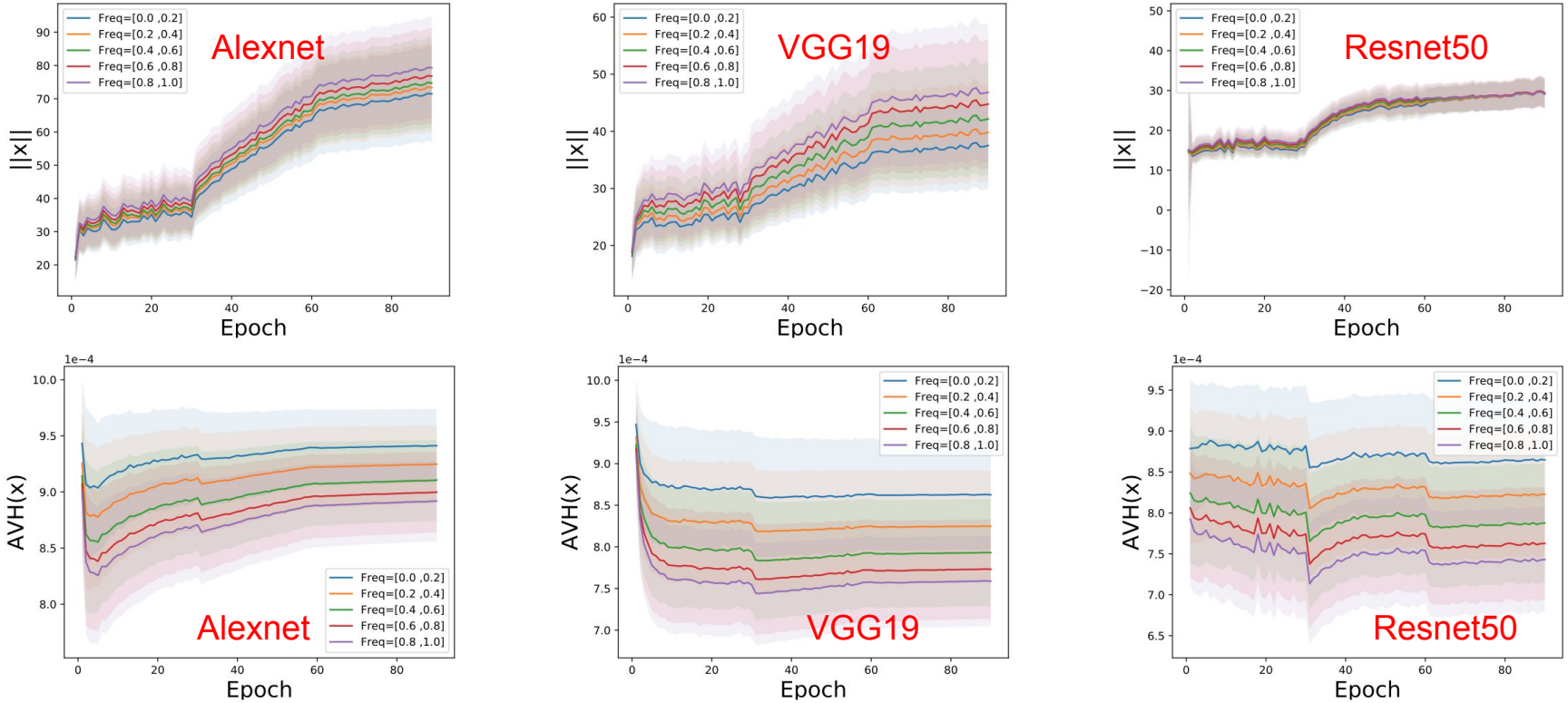
AVH's correlation with human selection frequency consistently holds across models throughout the training process.







The norm's correlation with human selection frequency is not consistent.

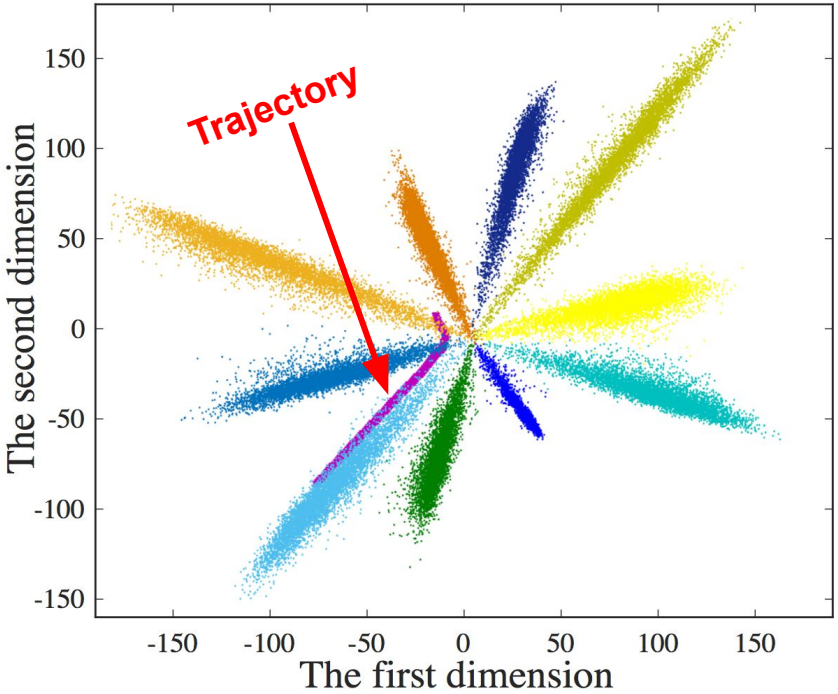


Conjecture on training dynamic of CNNs

- Softmax cross-entropy loss will first optimize the angles among different classes while the norm will fluctuate and increase very slowly.
- The angles become more stable and change very slowly while the norm increases rapidly.
- Easy examples: the angles get decreased enough for correct classification, the softmax cross-entropy loss can be well minimized by increasing the norm.
- Hard examples: the plateau is cause by unable to decrease the angle to correctly classify examples or increase the norms otherwise hurting loss

Special Case: Adversarial Example

- Trajectory of an adversarial example switching from one class to another



Future Work

- Application to sampling in deep metric learning, knowledge transfer and curriculum learning.

- How to design better loss function rather than the current softmax cross-entropy loss, which can directly optimize AVH.

Thank You

Welcome to stop by our poster!