Can Neural Nets Learn the Same Model Twice? Investigating Reproducibility and Double Descent from the Decision Boundary Perspective

Gowthami S., Liam F., Arpit B., Ping Y., Yehuda D., Richard B., Micah G. & Tom G.

MOTIVATION

🤔 Do neural nets learn the same model twice?

🤔 Do different neural architectures have measurable differences in inductive bias?

🤔 How are decision regions changing in double descent phenomenon in neural networks?

Drawing decision regions

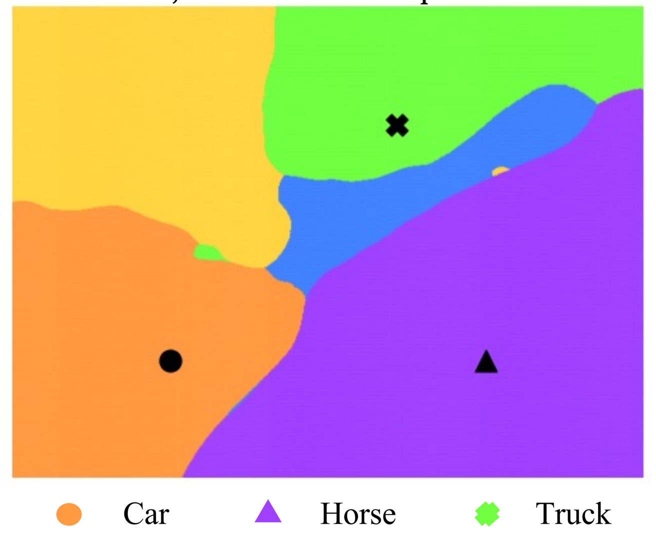
$(x_1, x_2, x_3) \sim \mathcal{D}^3$ Randomly sampled triplet from input space

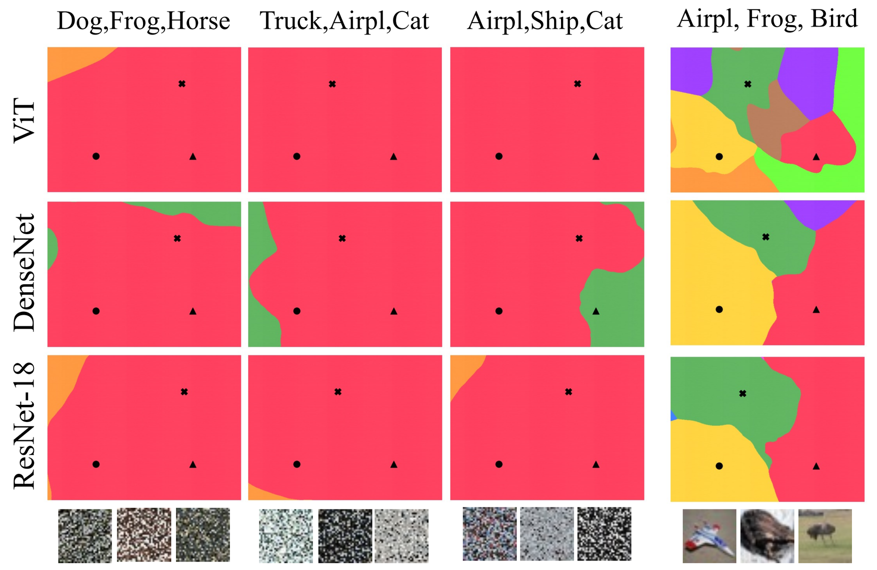
$\vec{v_1} = x_2 - x_1, \vec{v_2} = x_3 - x_1$

$-0.1 \leq \alpha, \beta \leq 1.1$

$$\alpha \cdot \max(\vec{v_1} \cdot \vec{v_1}, |\text{proj}_{\vec{v_1}} \vec{v_2} \cdot \vec{v_1}|)\vec{v_1} + \beta(\vec{v_2} - \text{proj}_{\vec{v_1}} \vec{v_2})$$

ResNet-18, trained for 200 epochs with SGD



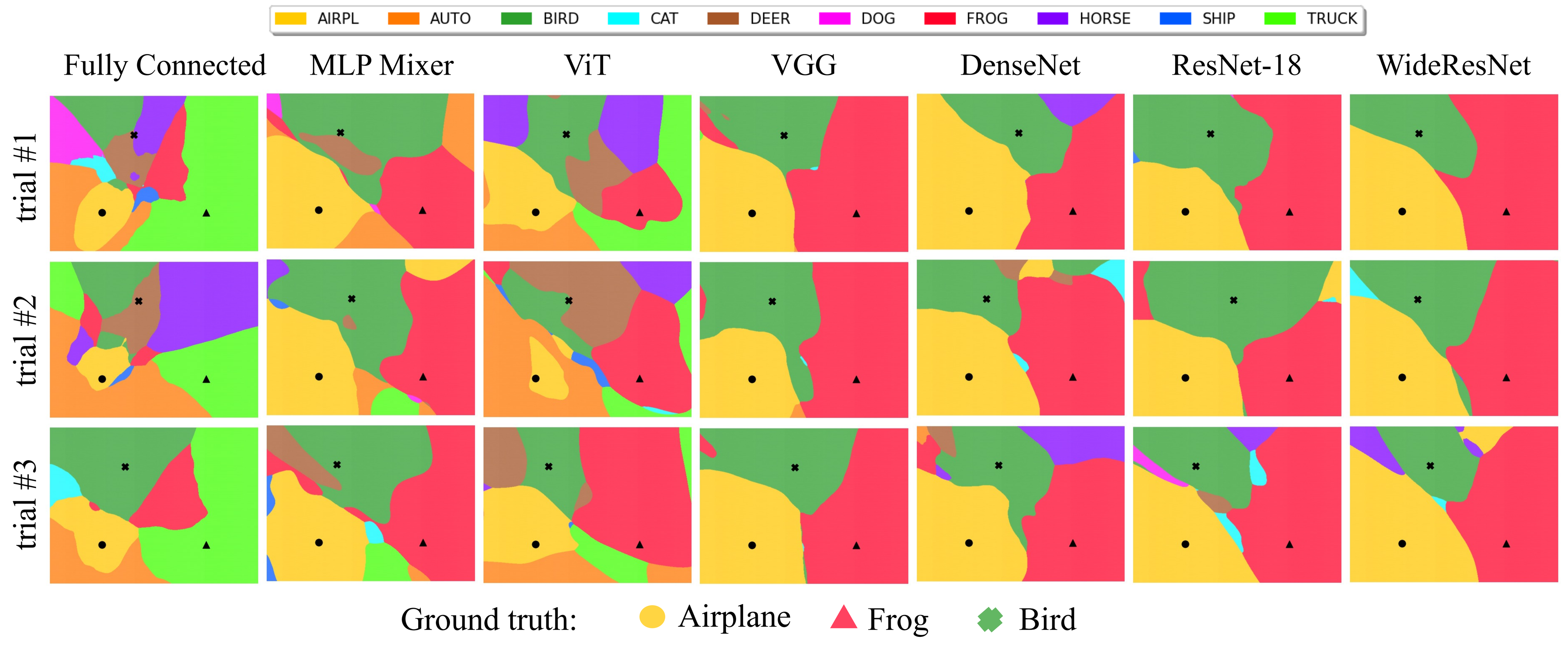


- The training process, which structures decision boundaries near the data manifold fails to produce strong structural effects far from the manifold.
- The uniform off-manifold behavior is an in- evitable consequence of the concentration of measures phenomenon

SCAN ME

Code and more materials available at https://somepago.github.io/dbviz

REPRODUCIBILITY



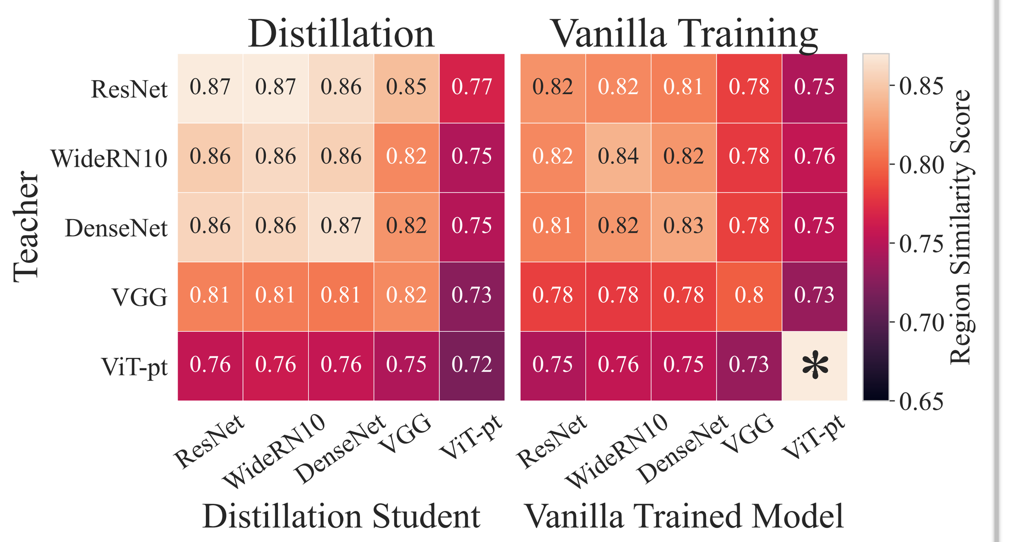
Region Similarity Score

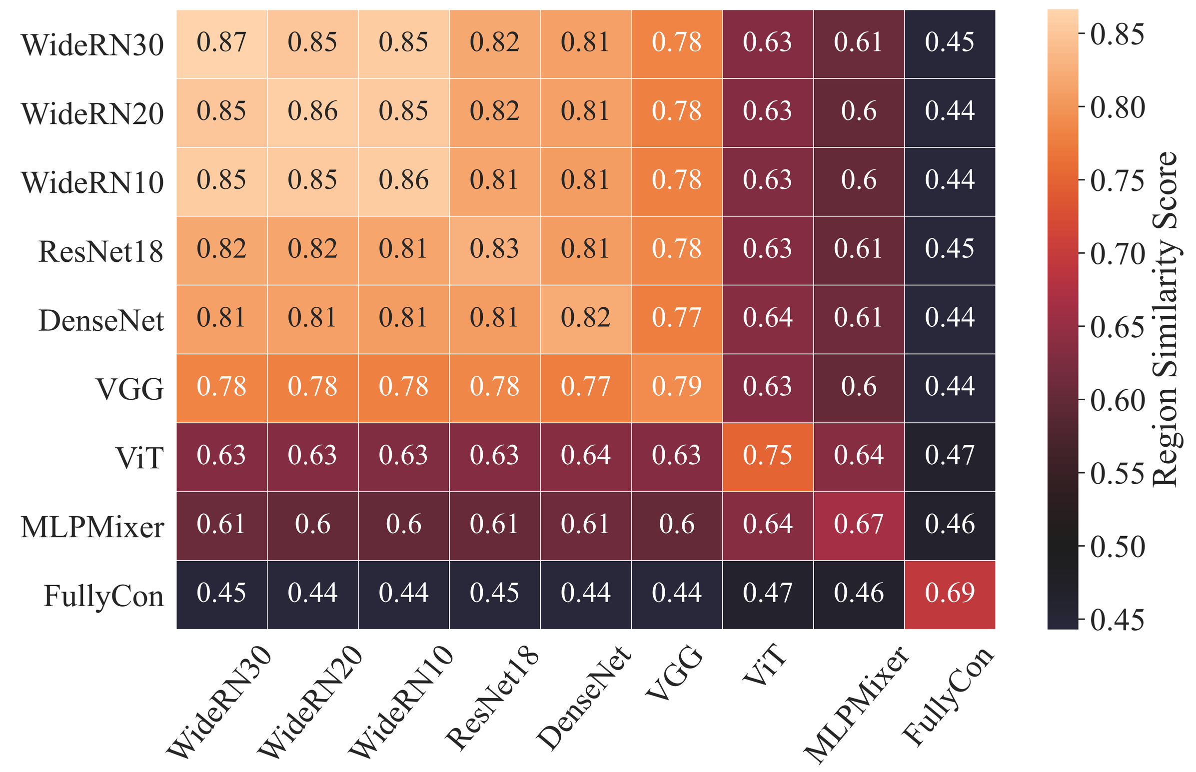
$$R(\theta_1, \theta_2) = \mathbb{E}_{T_i \sim \mathcal{D}}\left[(|f_{\theta_1}(S_i) \cap f_{\theta_2}(S_i)|)/|S_i|\right]$$

T_i Randomly chosen triplet

S_i Decision region spanned by T_i

$f_{\theta_1}, f_{\theta_2}$ Same architecture, trained differently





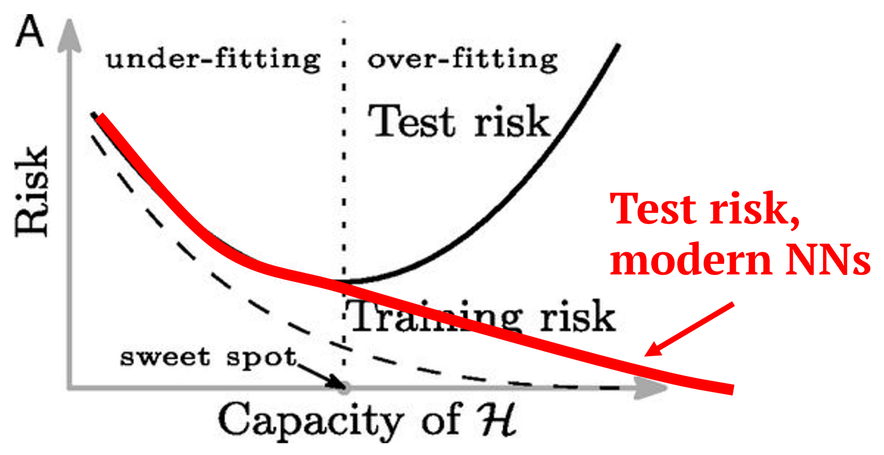
Region Similarity Scores

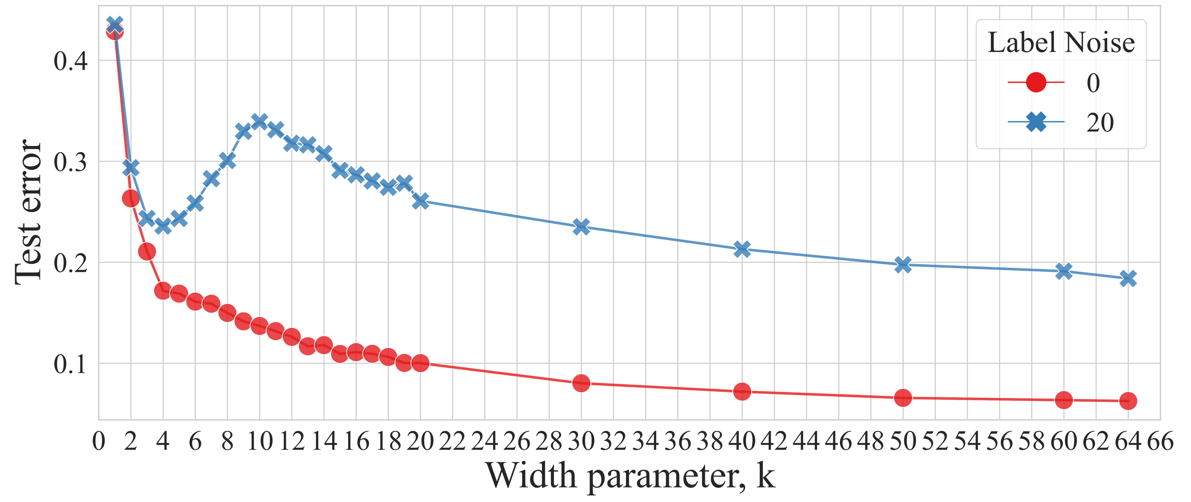
	Adam	SGD	SGD + SAM
ResNet-18	79.81	83.74	**87.22**
VGG	81.19	80.92	**84.21**
MLPMixer	67.80	66.51	**68.06**
VIT	69.55	75.13	**75.19**

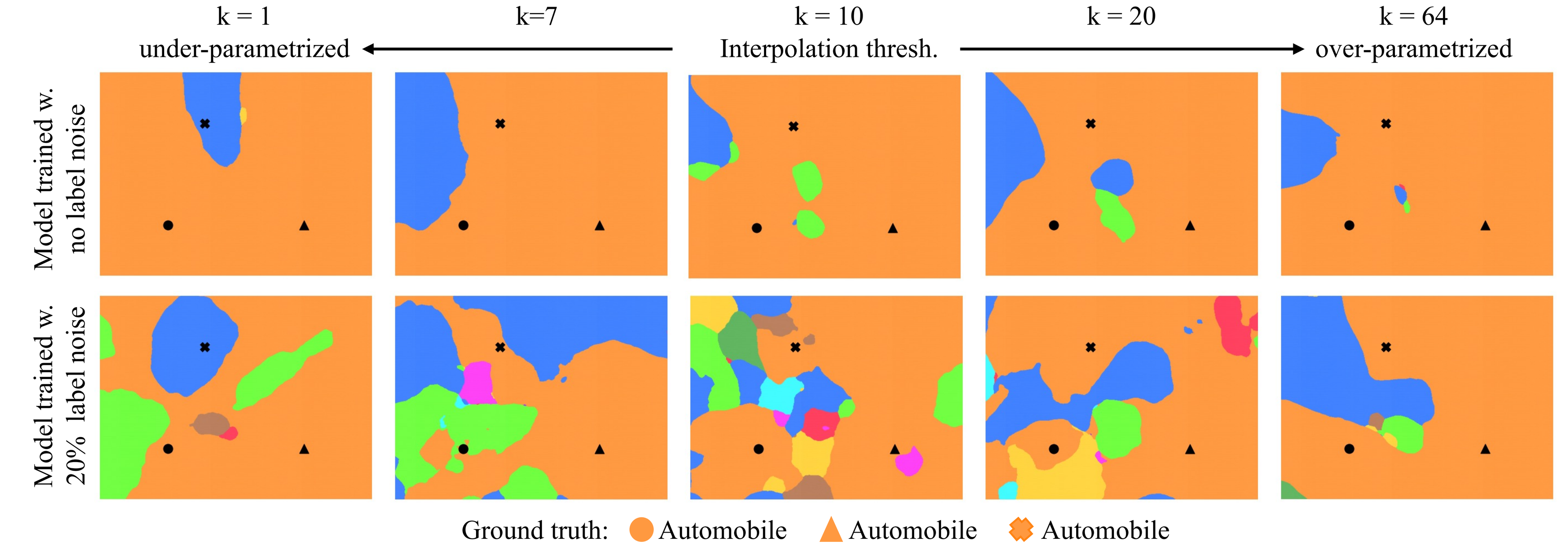
Test Accuracy

	Adam	SGD	SGD + SAM
ResNet-18	93.04	95.30	**95.68**
VGG	92.87	93.13	**93.90**
MLPMixer	**82.22**	82.04	82.18
VIT	70.89	**75.49**	74.72

DOUBLE DESCENT







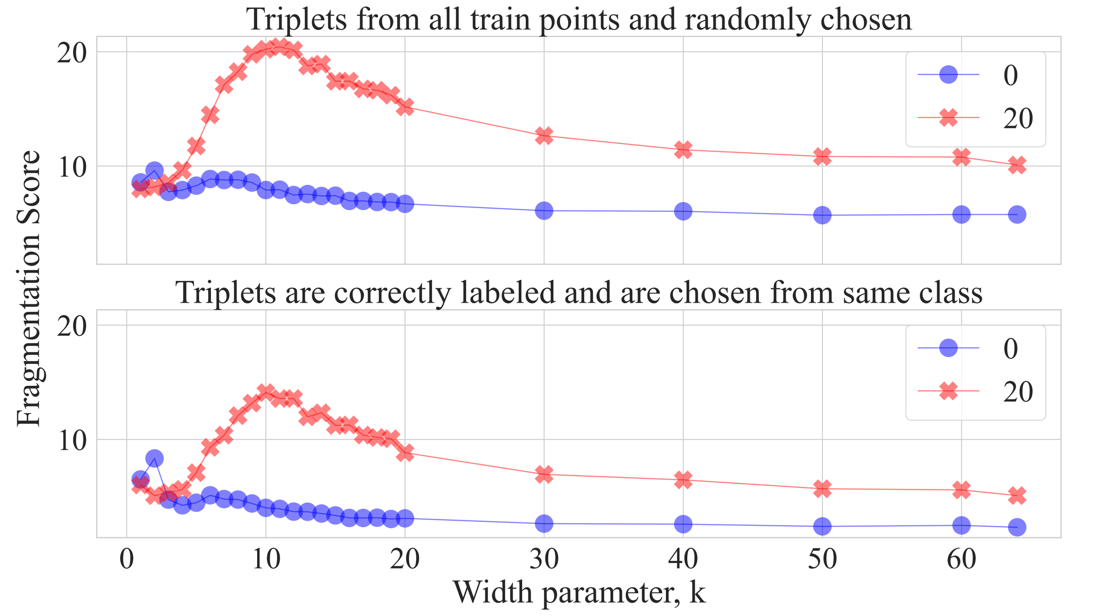
Fragmentation Score

$$S_i(\theta) = \cup_{j=1}^{n_i} P_j(\theta)$$

$$F(\theta) = \mathbb{E}_{T_i \sim \mathcal{D}}\, n_i(\theta, T_i)$$

T_i Randomly chosen triplet

$P_j(\theta)$ disjoint, maximal, path-connected component corresponding to a single predicted class label



The decision regions of models around double descent peak are more fragmented & less reproducible!