

# 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

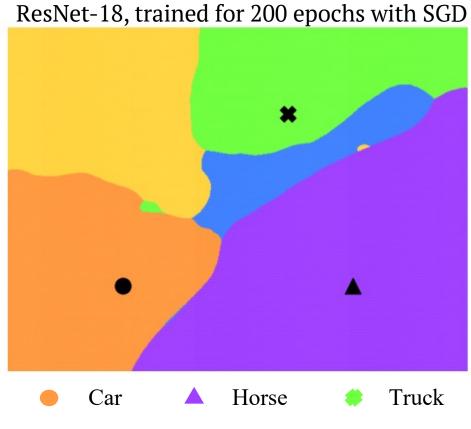
- Oneural nets learn the same model twice?
- Observation Do different neural architectures have measurable differences in inductive bias?
- How are decision regions changing in double descent phenomenon in neural networks?

Drawing decision regions

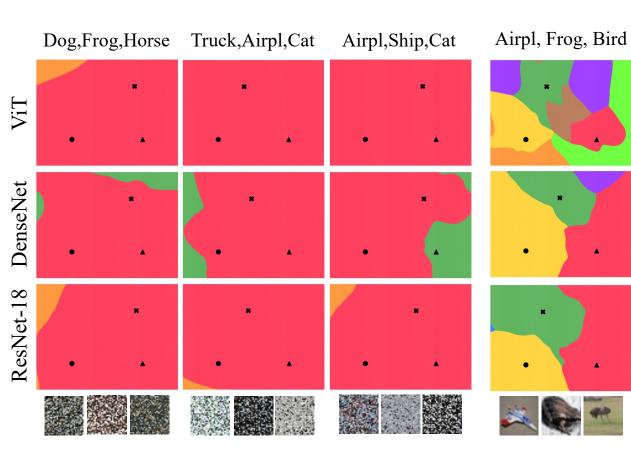
 $(x_1, x_2, x_3) \sim \mathcal{D}^3~$  Randomly sampled triplet from input space

- $\vec{v_1} = x_2 x_1, \vec{v_2} = x_3 x_1$
- $-0.1 \le \alpha, \beta \le 1.1$

SCAN ME

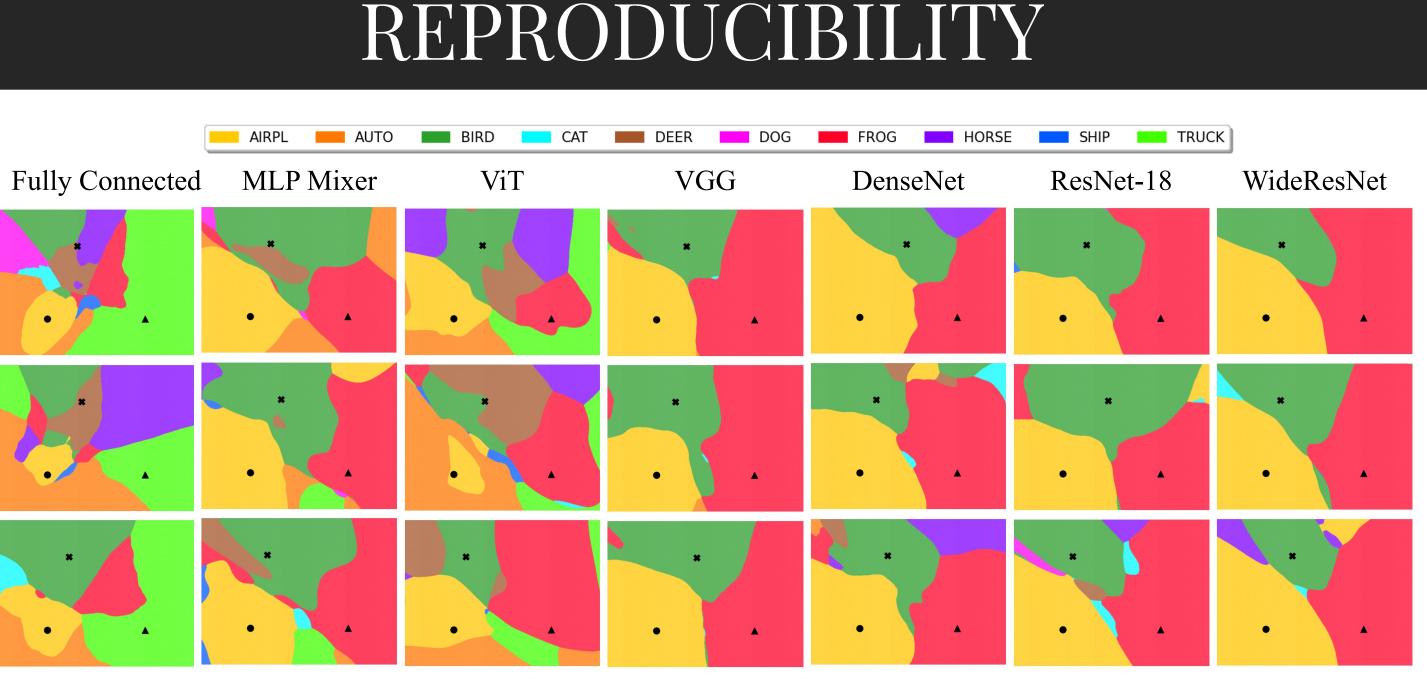


 $\alpha \cdot \max(\vec{v_1} \cdot \vec{v_1}, |\operatorname{proj}_{\vec{v_1}} \vec{v_2} \cdot \vec{v_1}|) \vec{v_1} + \beta(\vec{v_2} - \operatorname{proj}_{\vec{v_1}} \vec{v_2})$ 



- > 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

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



Frog

FullyCon

$S_i$ De	ecisi	on re	egior	n spa	nnec	l by '	$T_{i}$			
$f_{\theta_1}, f_{\theta_2}$	Sai	me a	rchit	ectu	re, tı	aine	d dif	ferei	ntly	
0 1 / 0 1 2					,				2	
WideRN30	0.87	0.85	0.85	0.82	0.81	0.78	0.63	0.61	0.45	-0.85
WideRN20	0.85	0.86	0.85	0.82	0.81	0.78	0.63	0.6	0.44	-0.80
WideRN10	0.85	0.85	0.86	0.81	0.81	0.78	0.63	0.6	0.44	-0.75 0.75
ResNet18	0.82	0.82	0.81	0.83	0.81	0.78	0.63	0.61	0.45	- 0.75 200 - 0.70 - 0.0 - 0.0 - 0.0 - 0.00 - 0.60 0.0 0 - 0.
DenseNet	0.81	0.81	0.81	0.81	0.82	0.77	0.64	0.61	0.44	-0.65 mil
VGG	0.78	0.78	0.78	0.78	0.77	0.79	0.63	0.6	0.44	S 0.60 0
ViT	0.63	0.63	0.63	0.63	0.64	0.63	0.75	0.64	0.47	- 0.55 <sup></sup>
MLPMixer	0.61	0.6	0.6	0.61	0.61	0.6	0.64	0.67	0.46	-0.50
	~ <b>. . .</b>								0.00	

**Region Similarity Score** 

Randomly chosen triplet

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

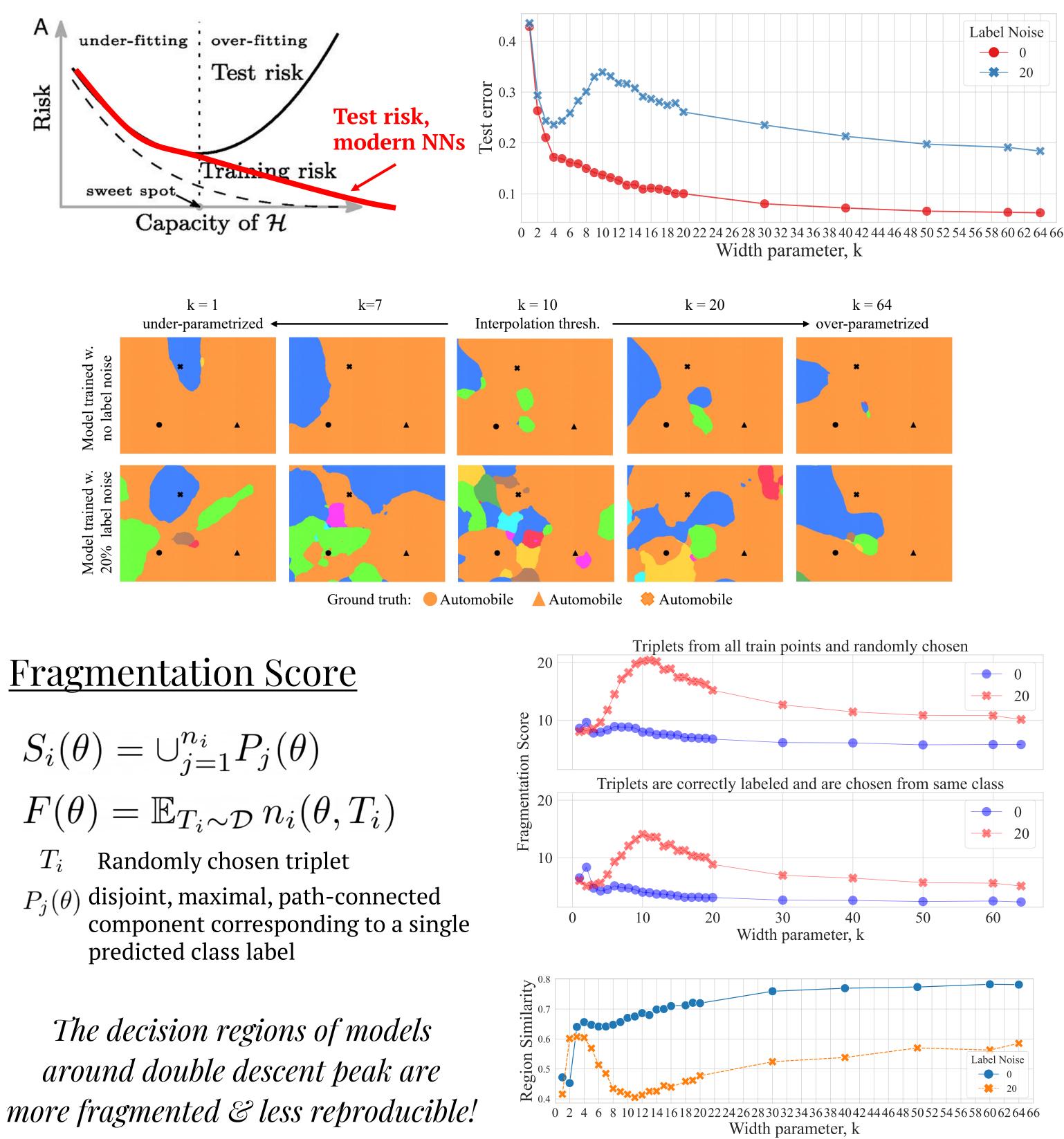
Airplane

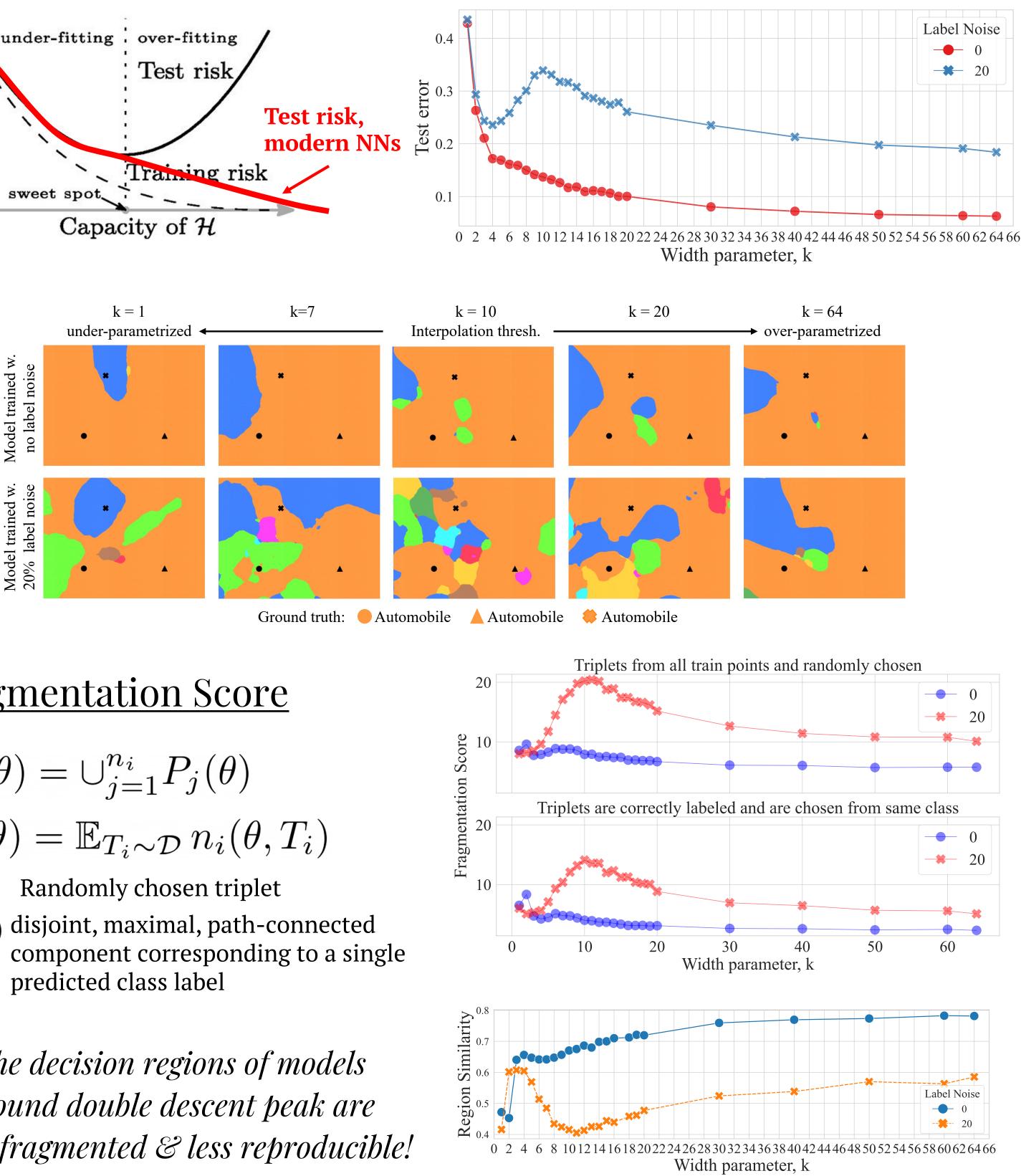
Hider Hider Hider Dester Per vice vit

Bird

Distillation Vanilla Training ResNet 0.87 0.87 0.86 0.85 0.77 0.81 0.78 0.75 -0.85 0.84 0.82 0.78 0.76 -0.80 WideRN10 0.86 0.86 0.86 DenseNet 0.86 0.86 0.87 0. 0.82 0.83 0. VGG 0.73 0.78 0.78 0.78 0.8 0.73 ViT-pt 0.76 0.76 0.76 0.75 0.72 0.75 0.76 0.75 0.73 ResNet deRNIO penservet GG ViT-pt ResNet deRNIO penservet GG ViT-pt Distillation Student Vanilla Trained Model

Re	gion Sim	ilarity Sc	cores						
	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						





### Fragmentation Score

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



### DOUBLE DESCENT