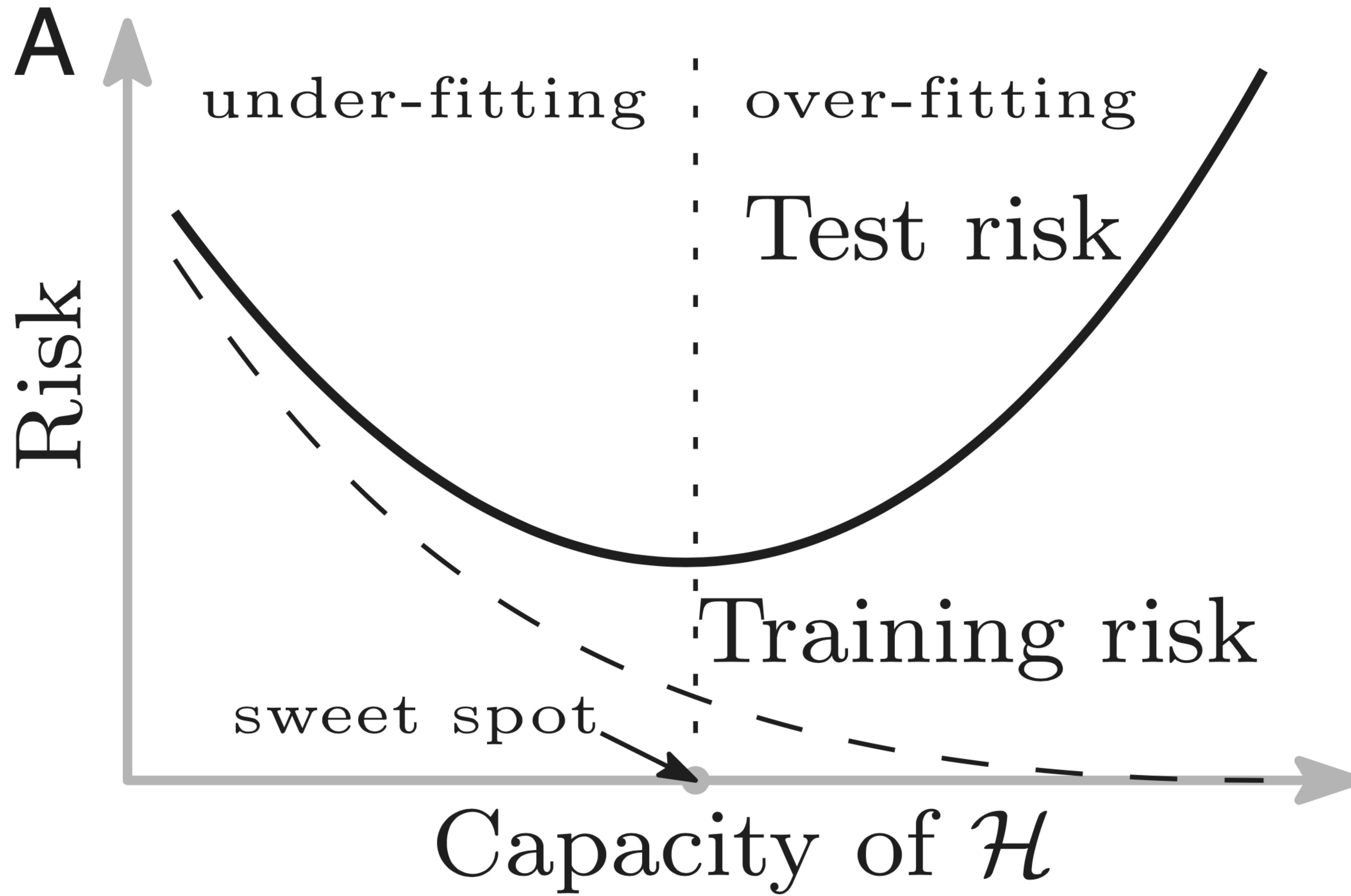
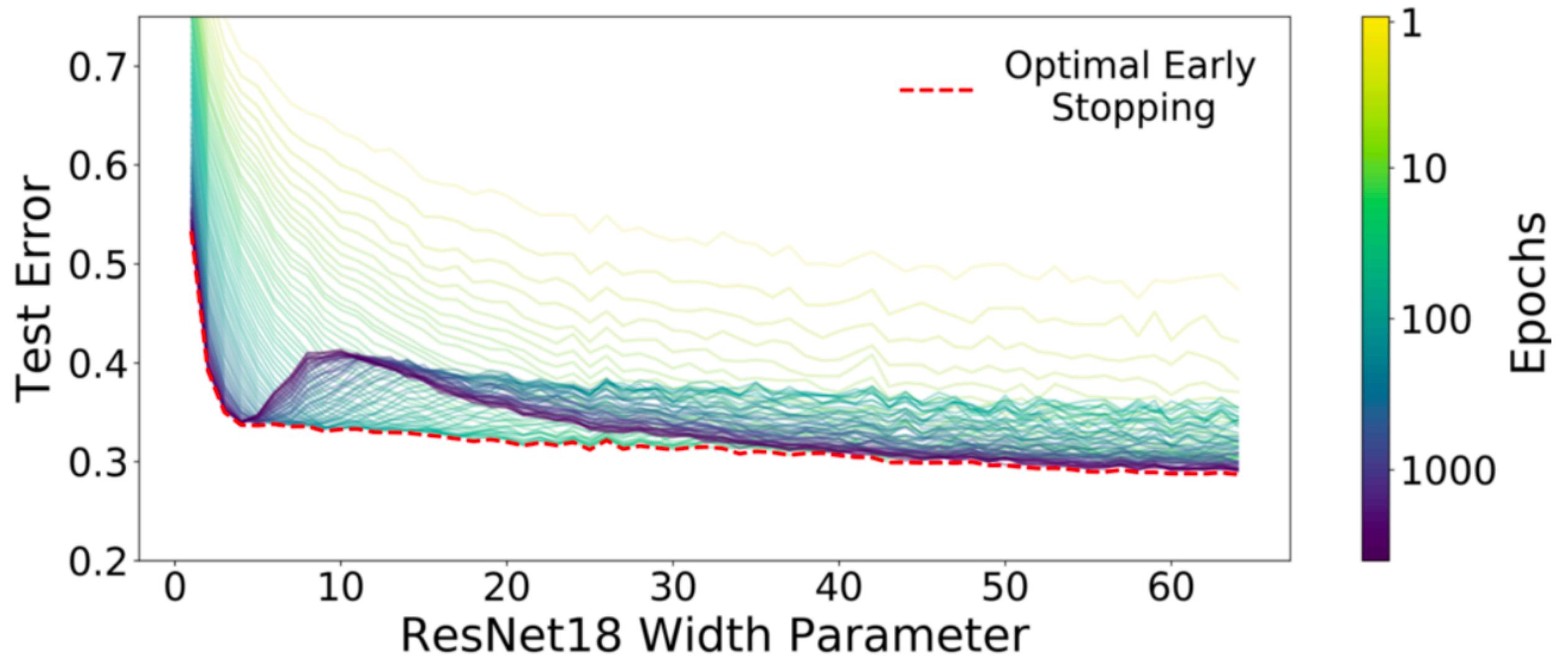
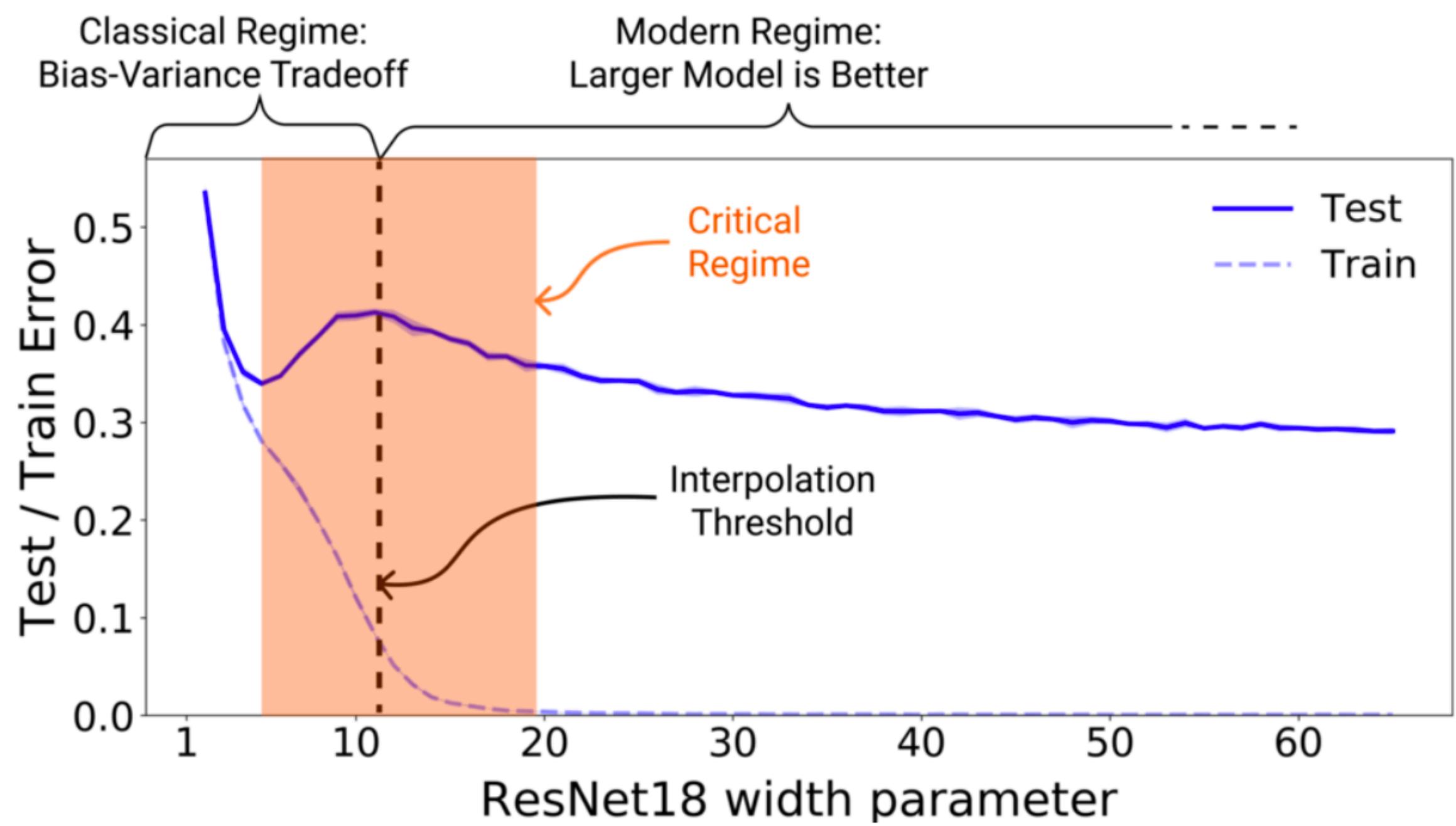


Double Descent

Смирнов Павел БПМИ172





Definition 1 (Effective Model Complexity) *The Effective Model Complexity (EMC) of a training procedure \mathcal{T} , with respect to distribution \mathcal{D} and parameter $\epsilon > 0$, is defined as:*

$$\text{EMC}_{\mathcal{D}, \epsilon}(\mathcal{T}) := \max \{n \mid \mathbb{E}_{S \sim \mathcal{D}^n} [\text{Error}_S(\mathcal{T}(S))] \leq \epsilon\}$$

where $\text{Error}_S(M)$ is the mean error of model M on train samples S .

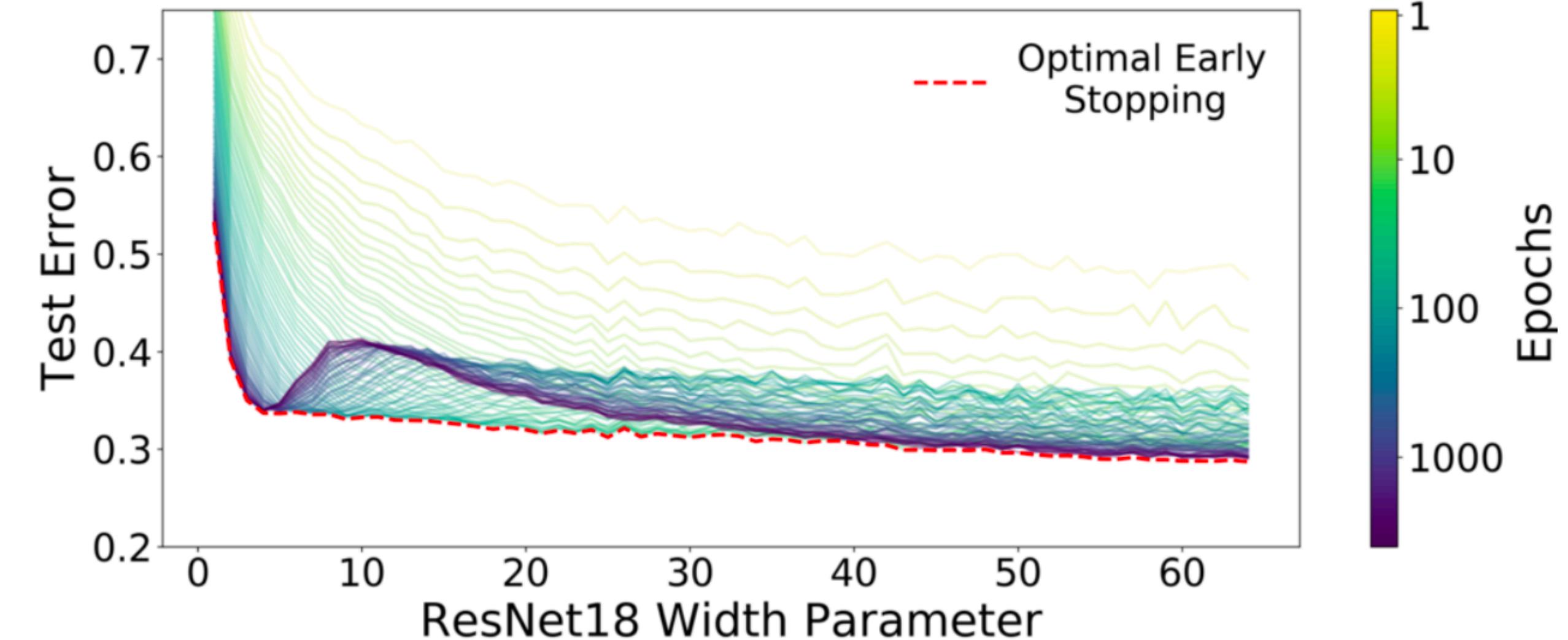
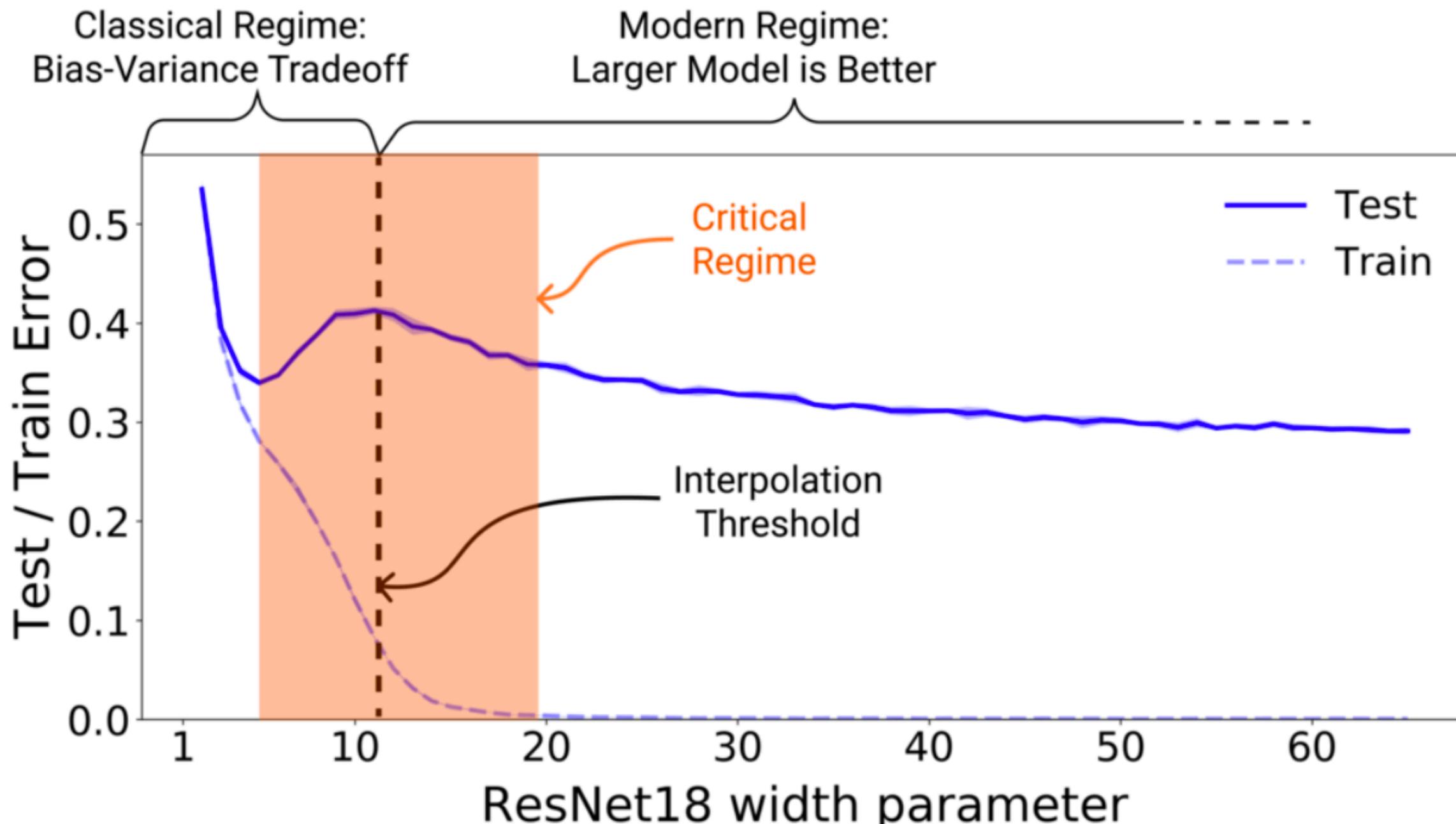
Our main hypothesis can be informally stated as follows:

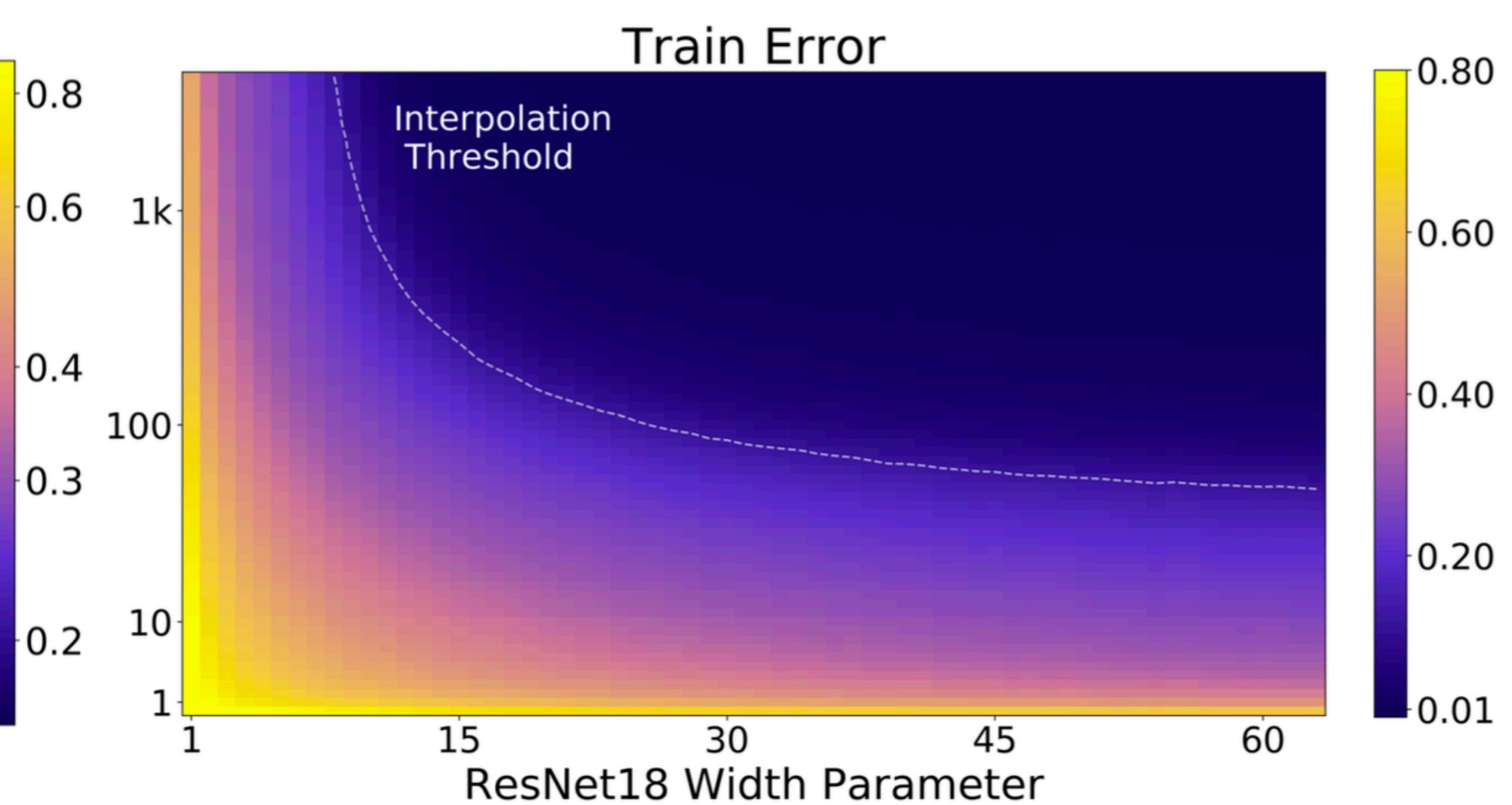
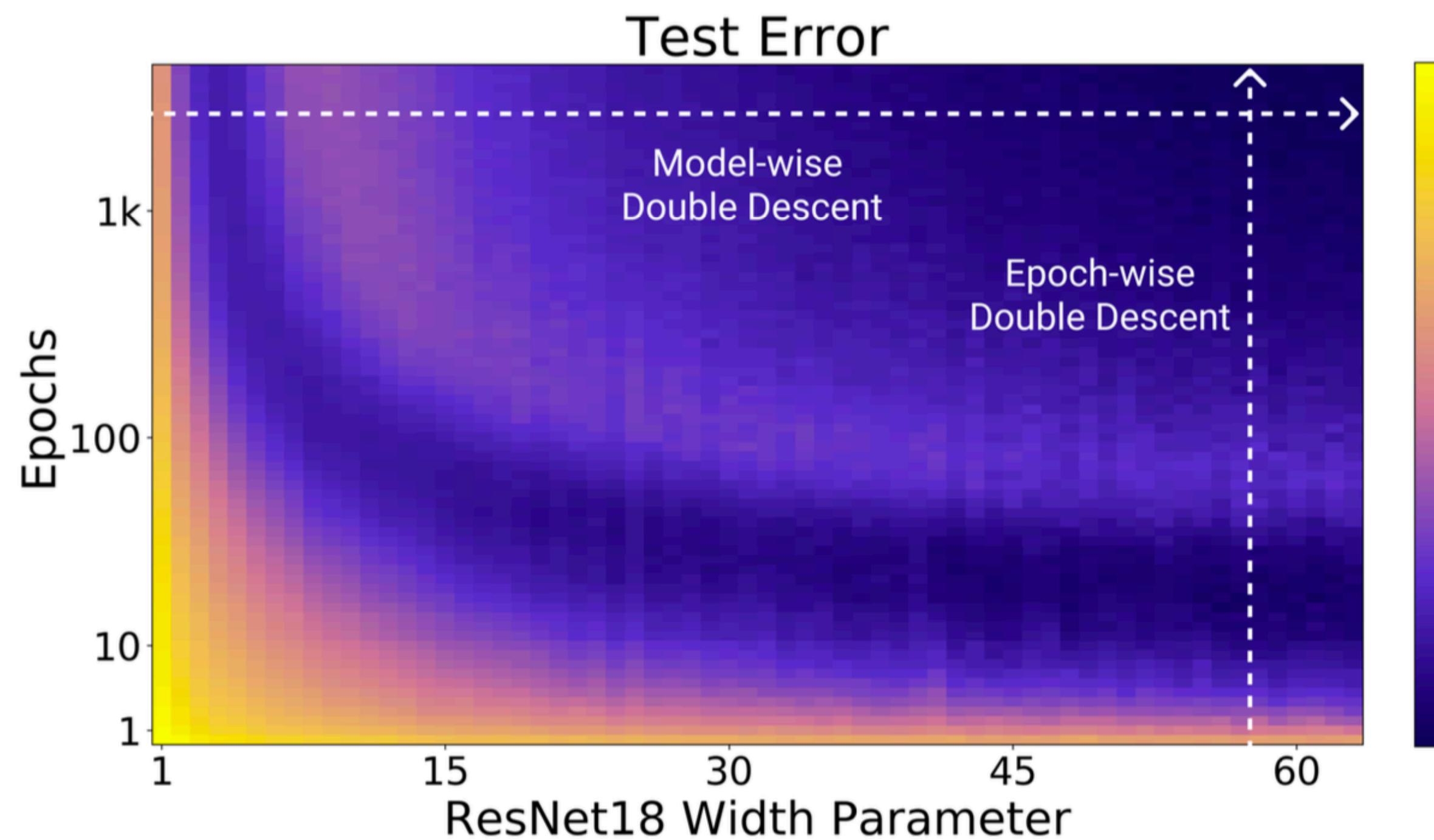
Hypothesis 1 (Generalized Double Descent hypothesis, informal) *For any natural data distribution \mathcal{D} , neural-network-based training procedure \mathcal{T} , and small $\epsilon > 0$, if we consider the task of predicting labels based on n samples from \mathcal{D} then:*

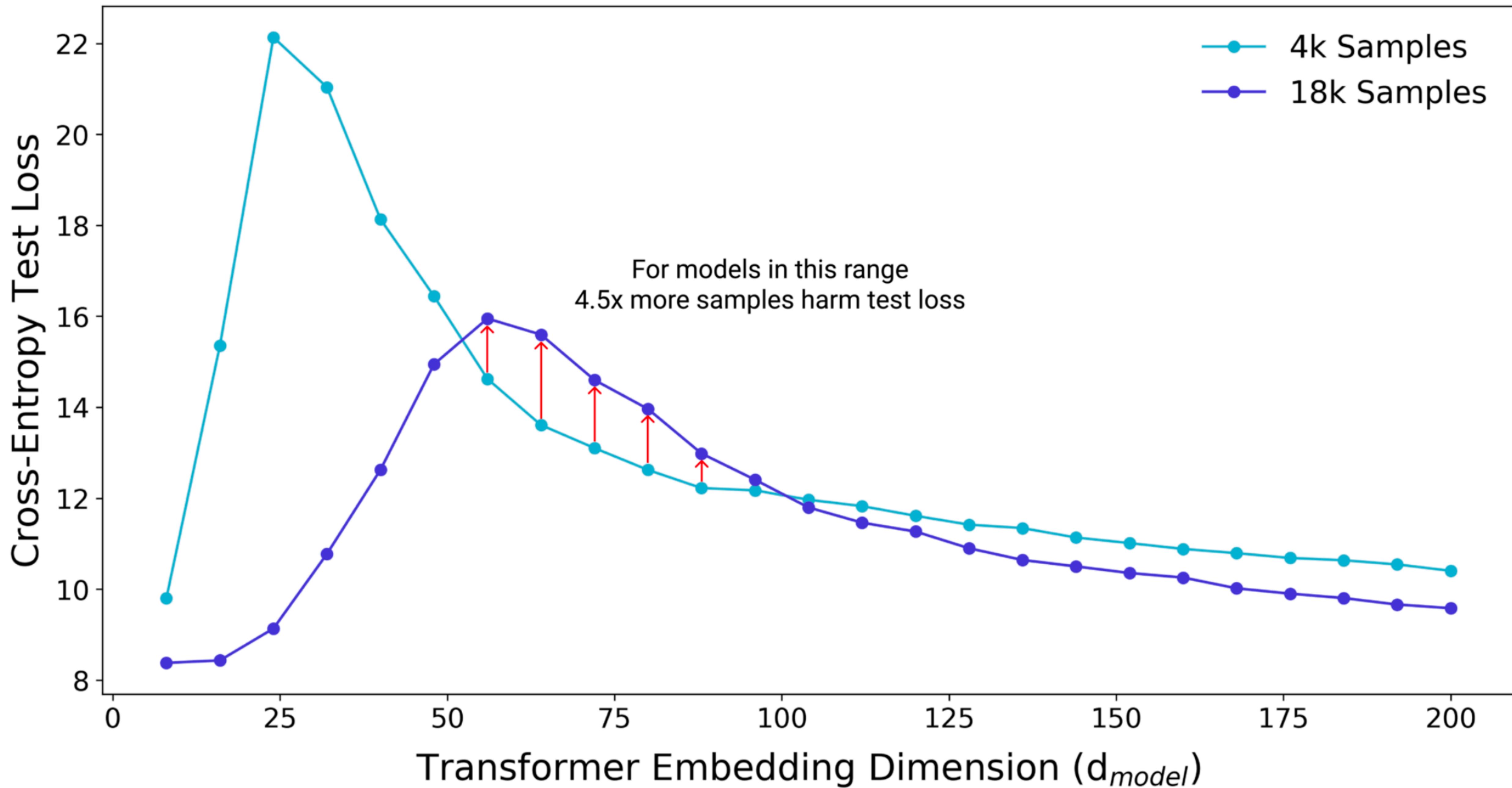
Under-parameterized regime. *If $\text{EMC}_{\mathcal{D}, \epsilon}(\mathcal{T})$ is sufficiently smaller than n , any perturbation of \mathcal{T} that increases its effective complexity will decrease the test error.*

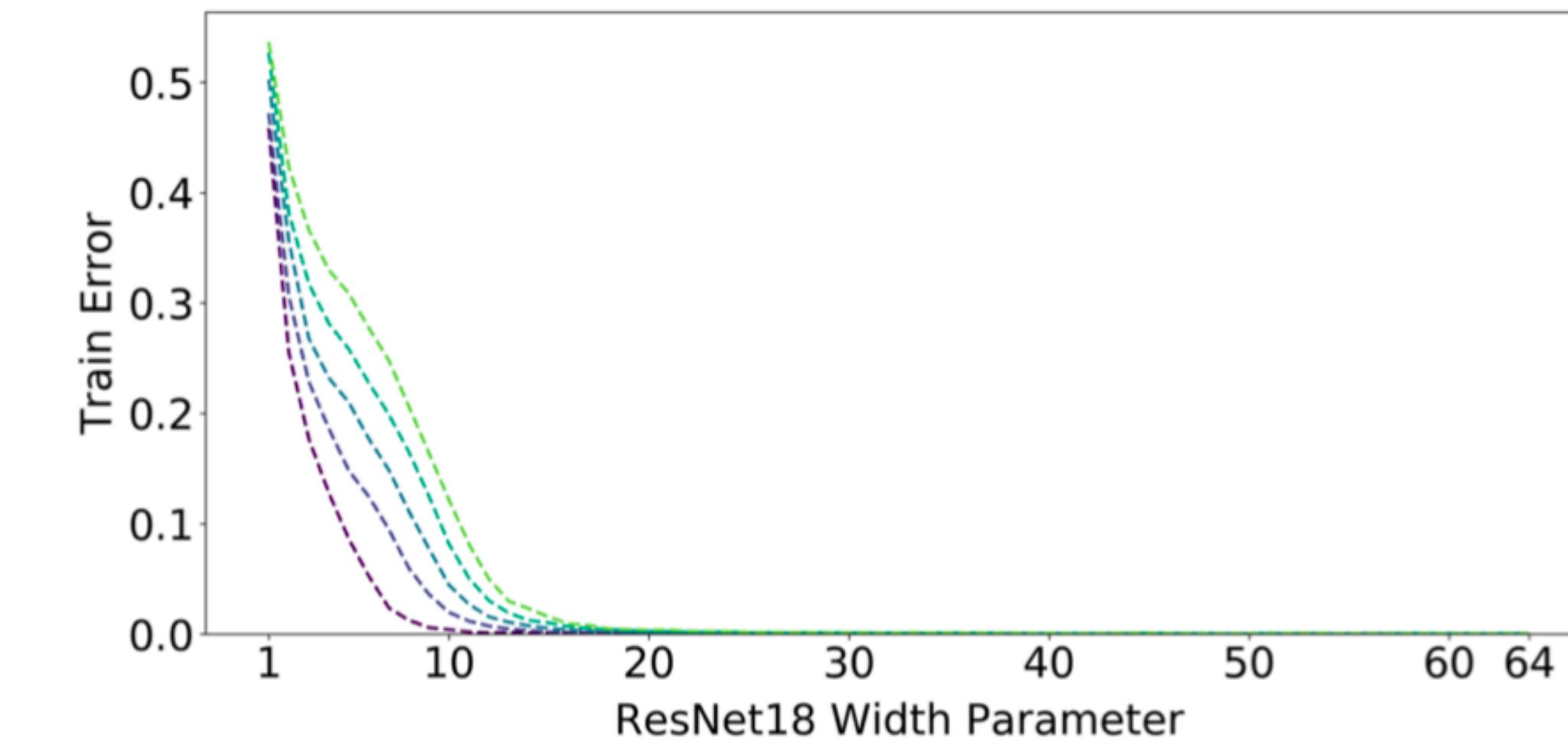
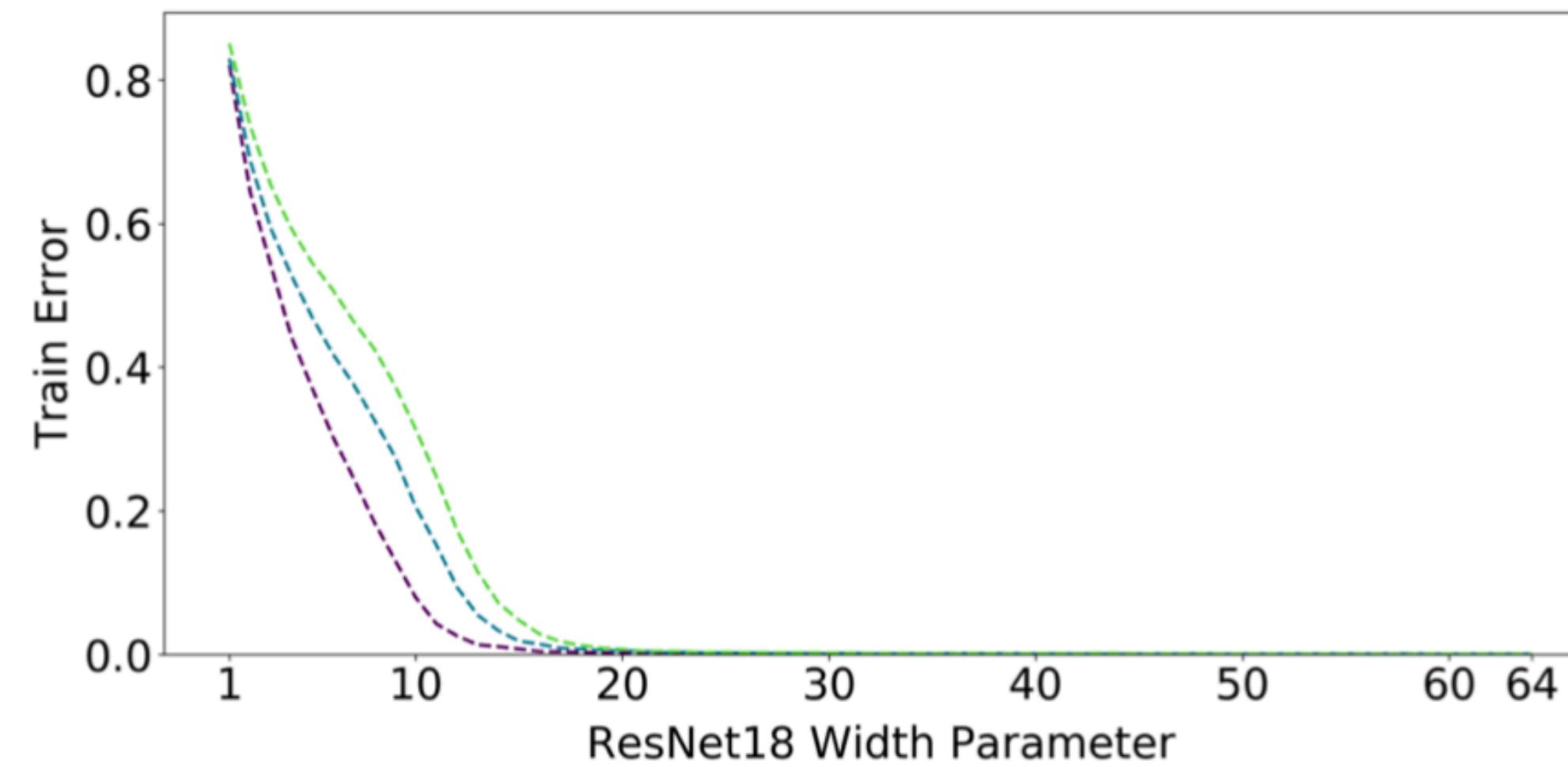
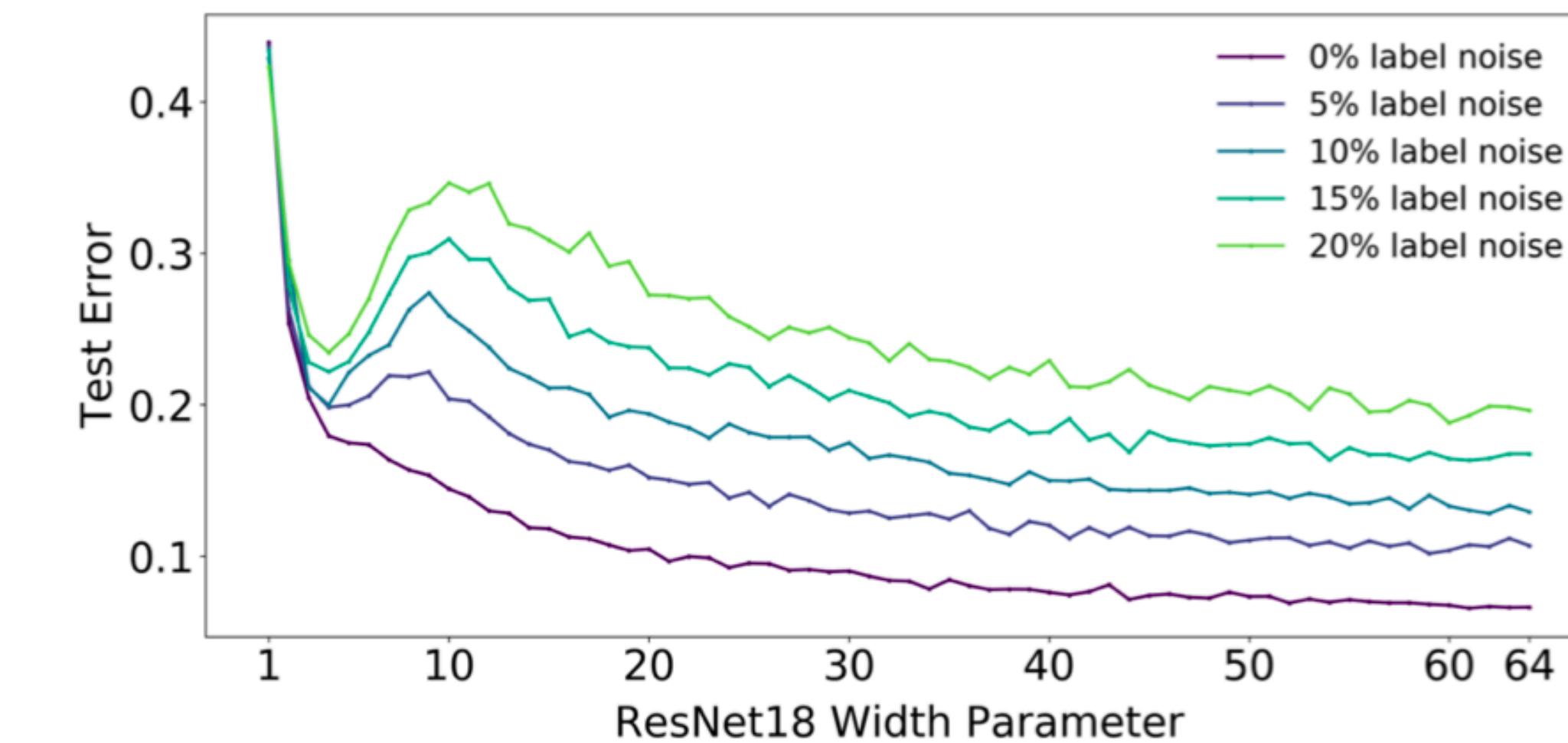
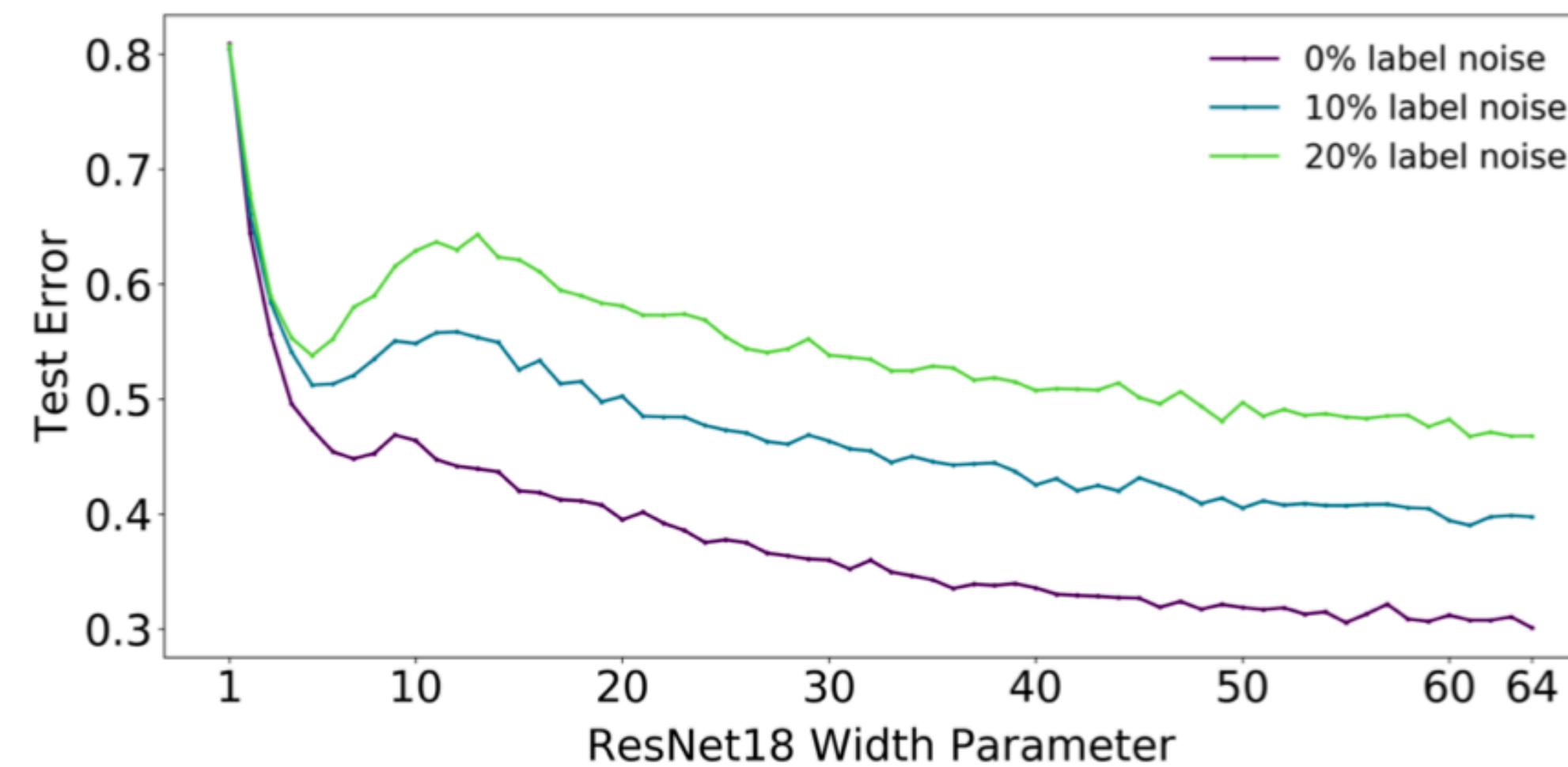
Over-parameterized regime. *If $\text{EMC}_{\mathcal{D}, \epsilon}(\mathcal{T})$ is sufficiently larger than n , any perturbation of \mathcal{T} that increases its effective complexity will decrease the test error.*

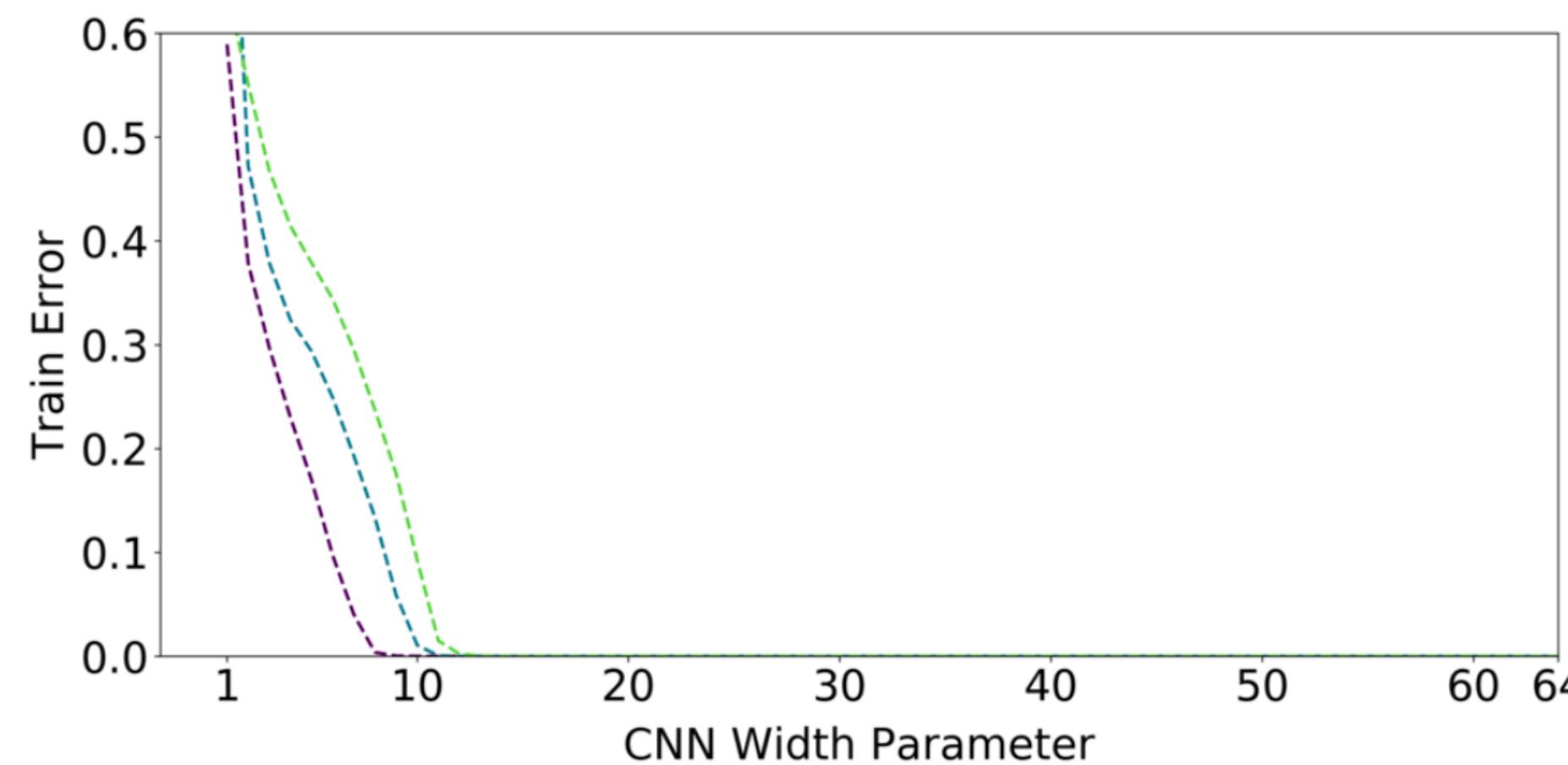
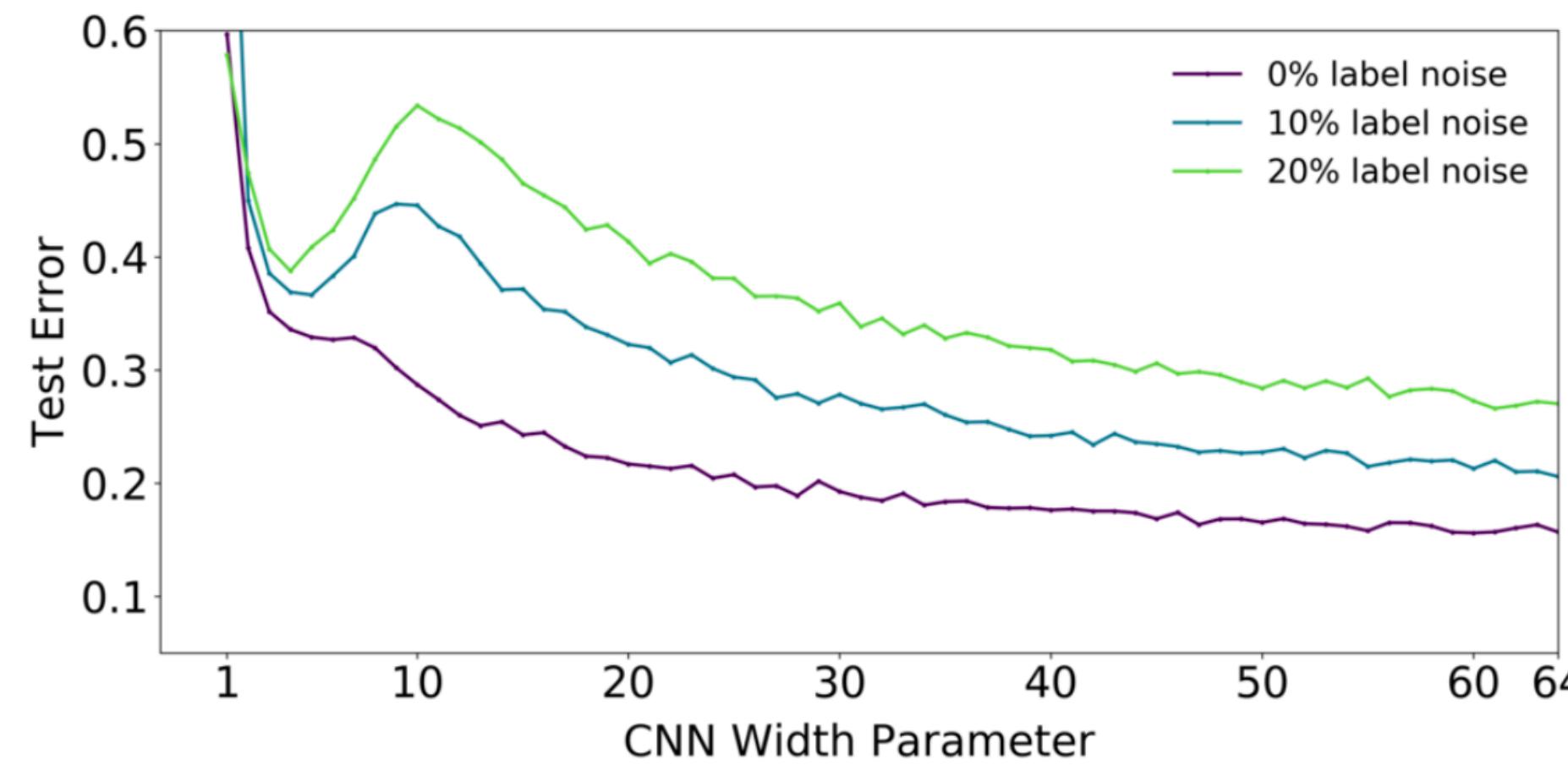
Critically parameterized regime. *If $\text{EMC}_{\mathcal{D}, \epsilon}(\mathcal{T}) \approx n$, then a perturbation of \mathcal{T} that increases its effective complexity might decrease or increase the test error.*



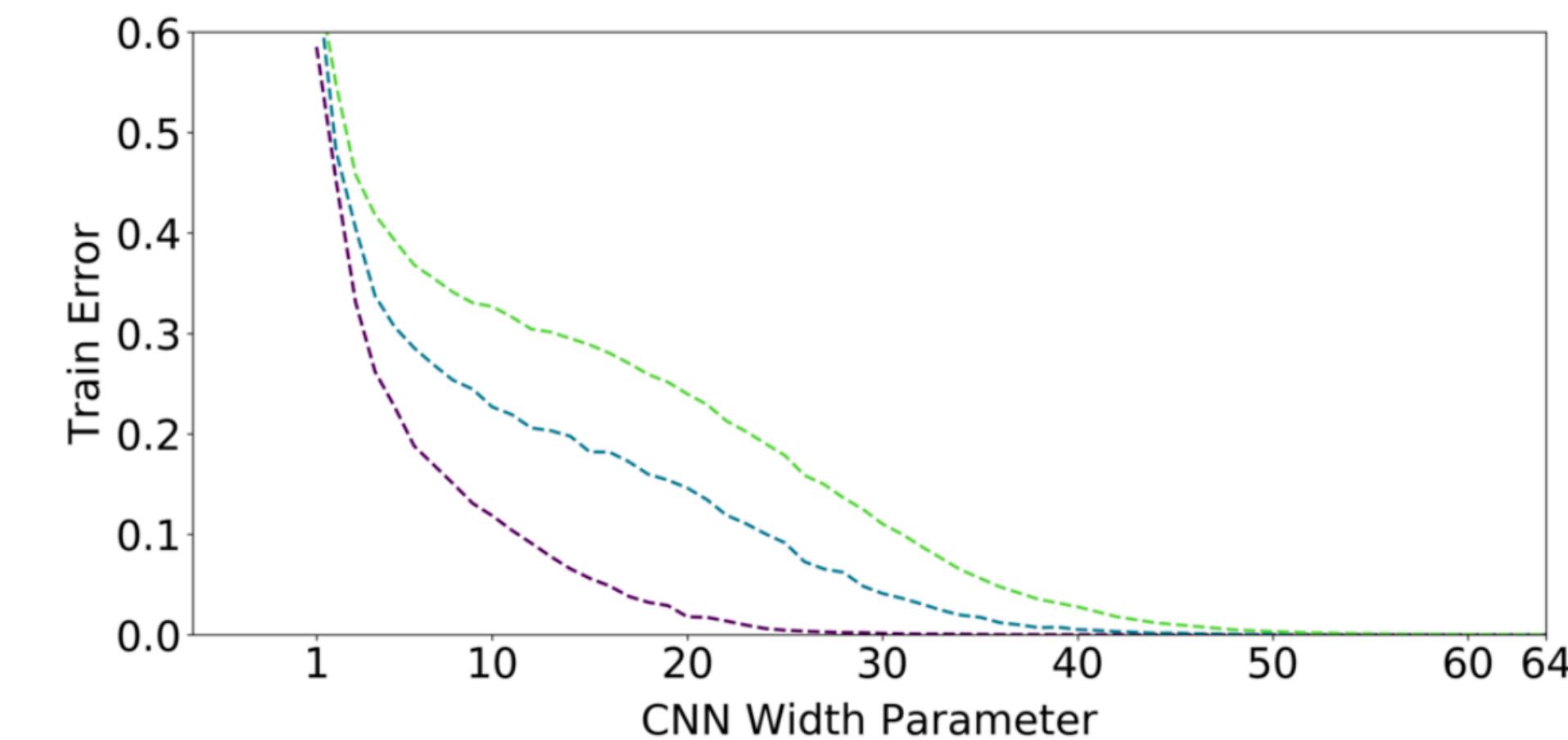
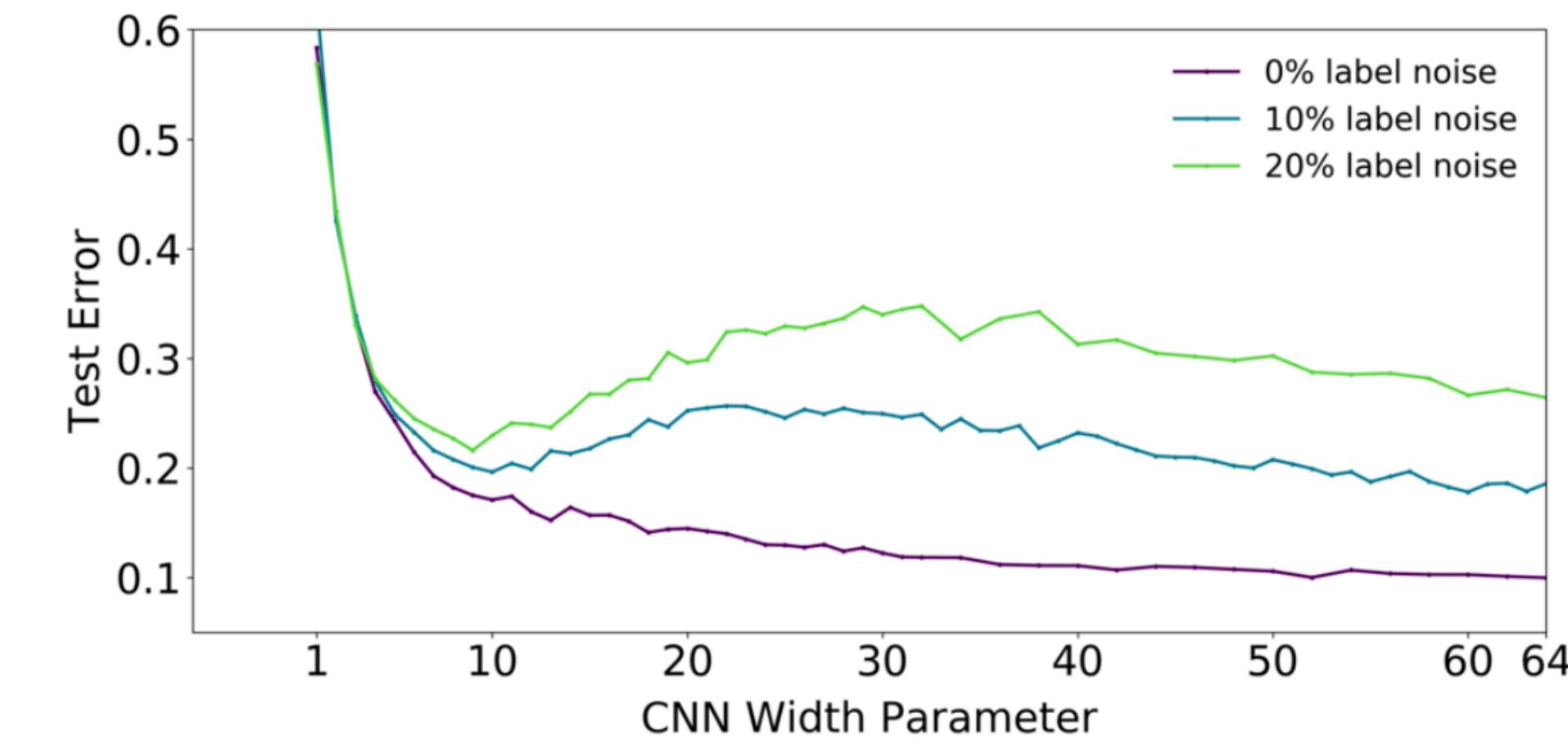




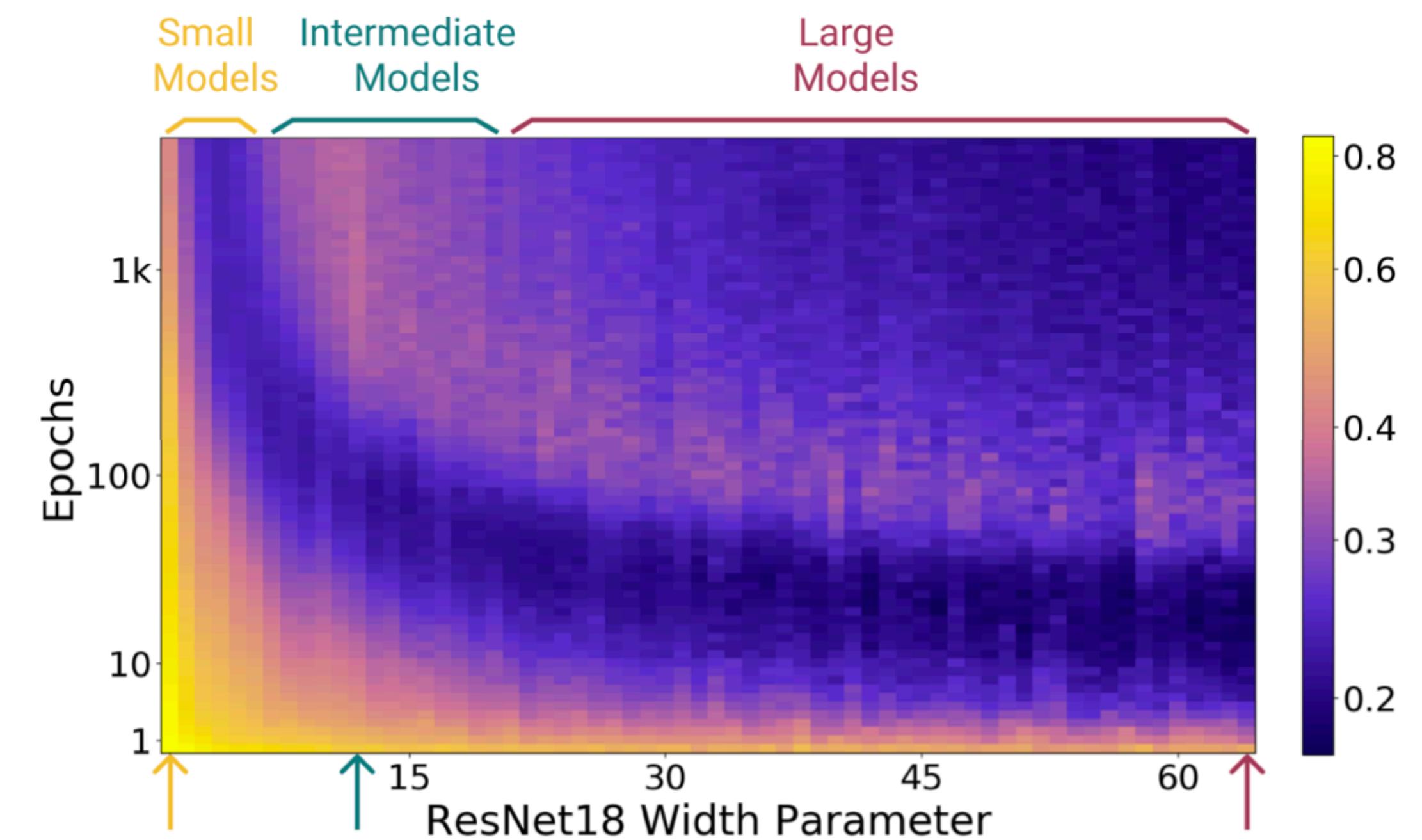
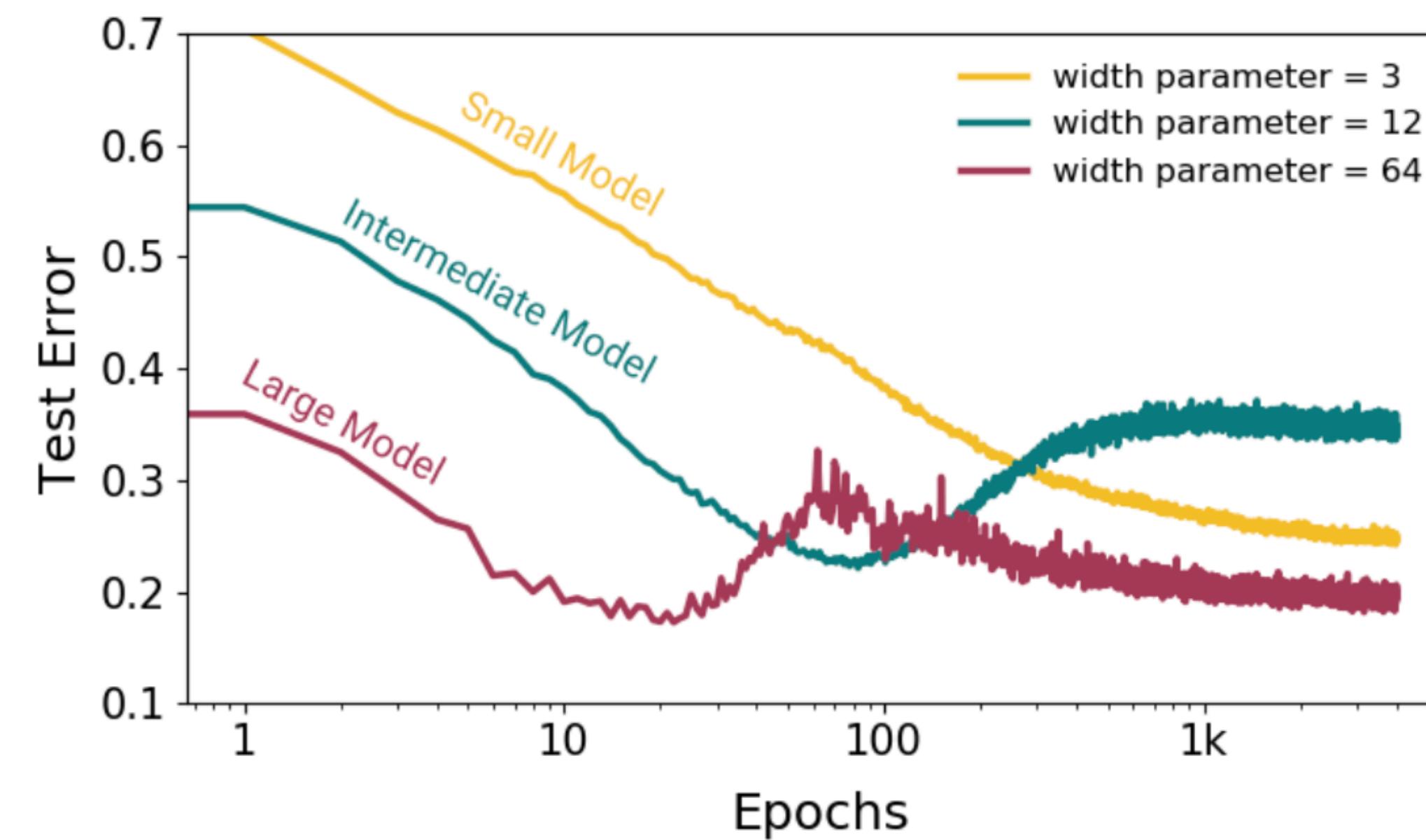


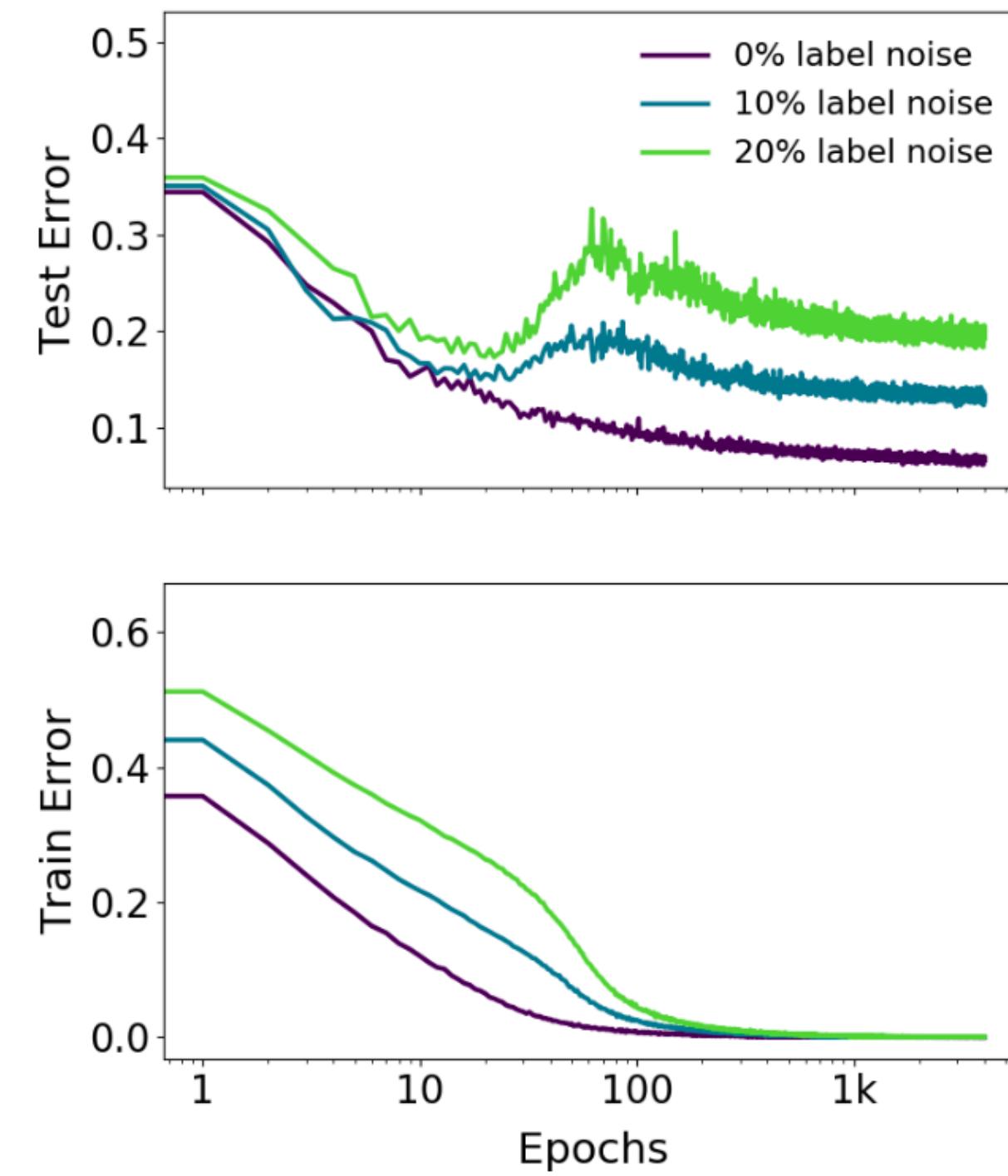


(a) Without data augmentation.

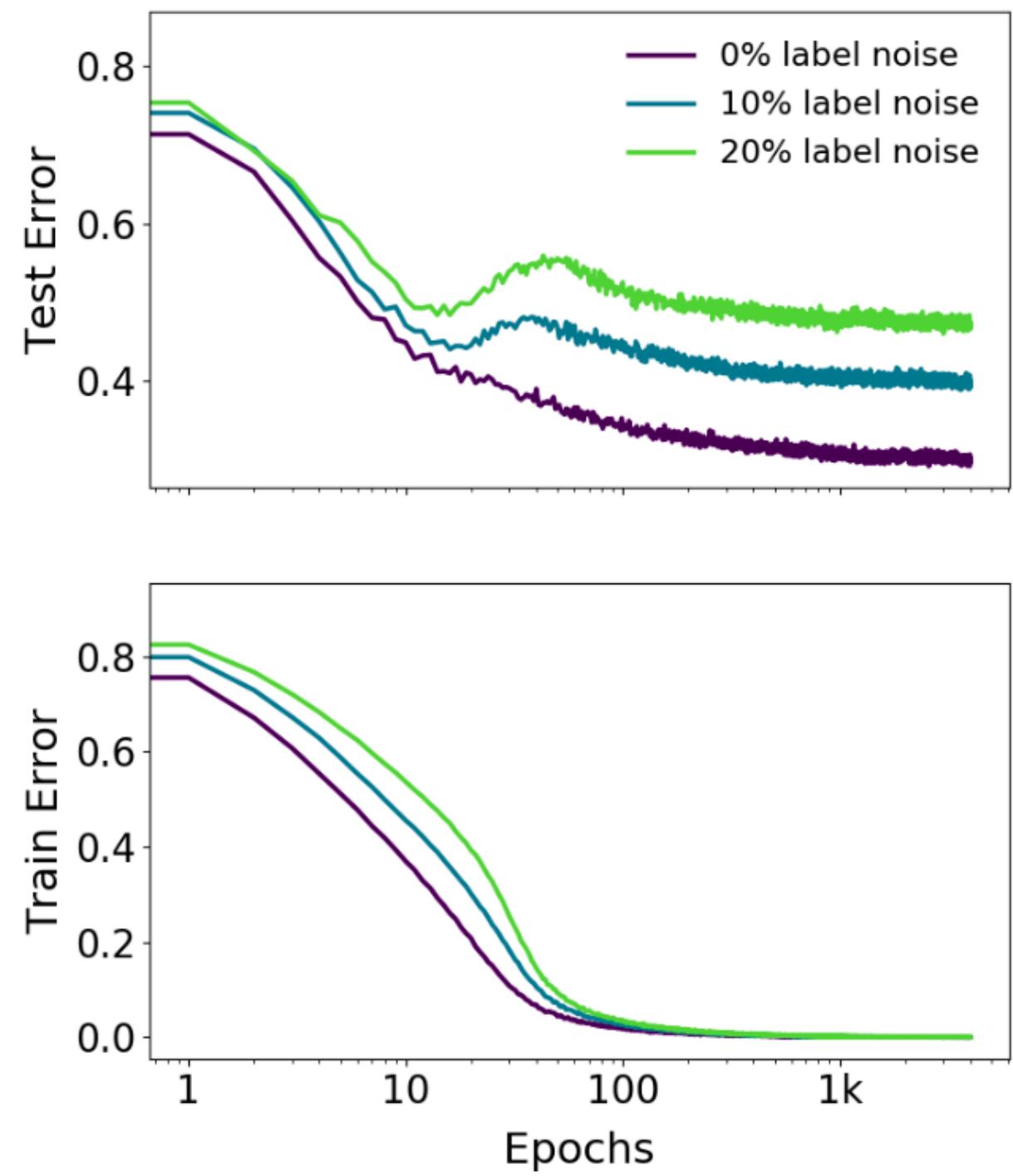


(b) With data augmentation.

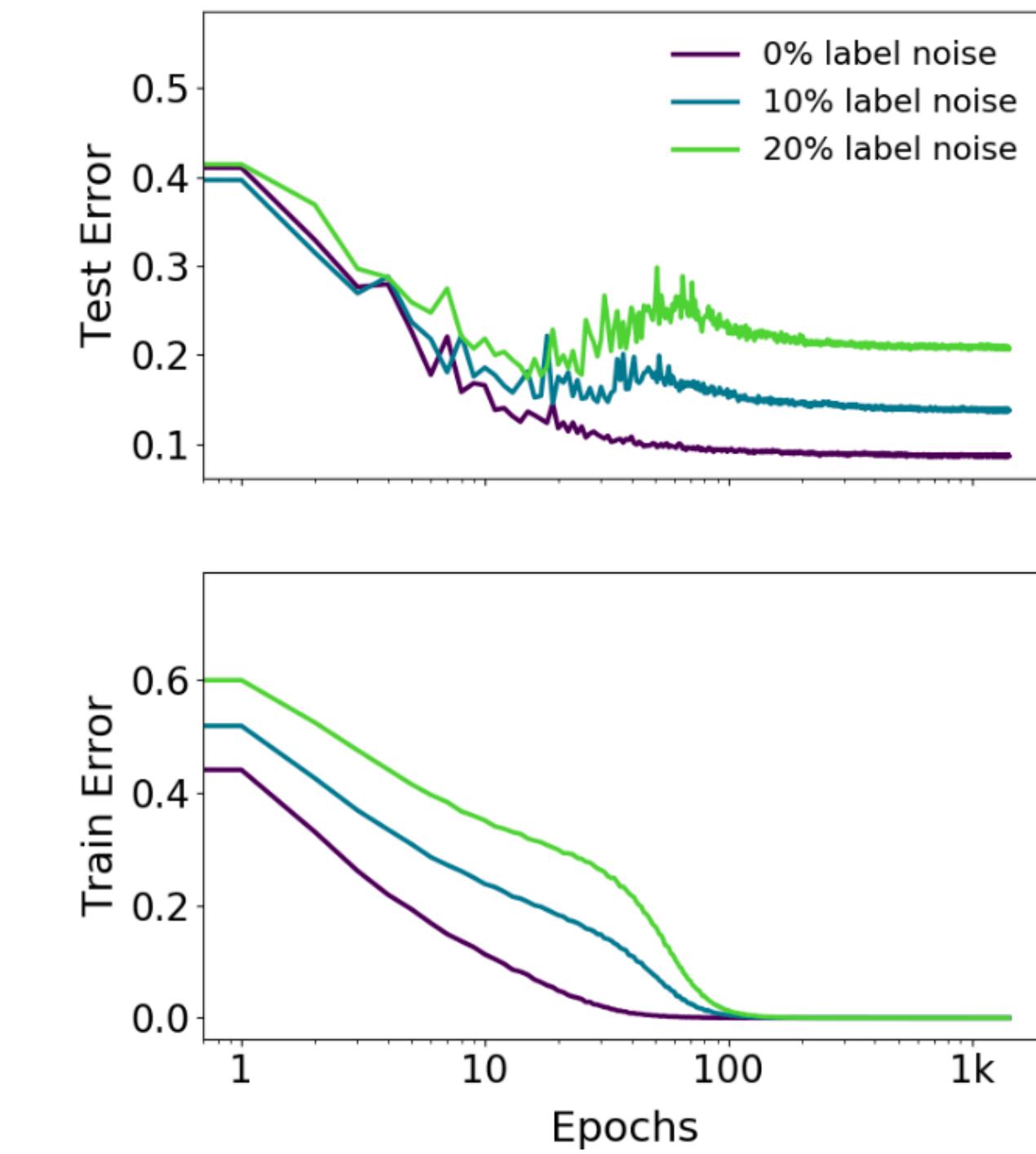




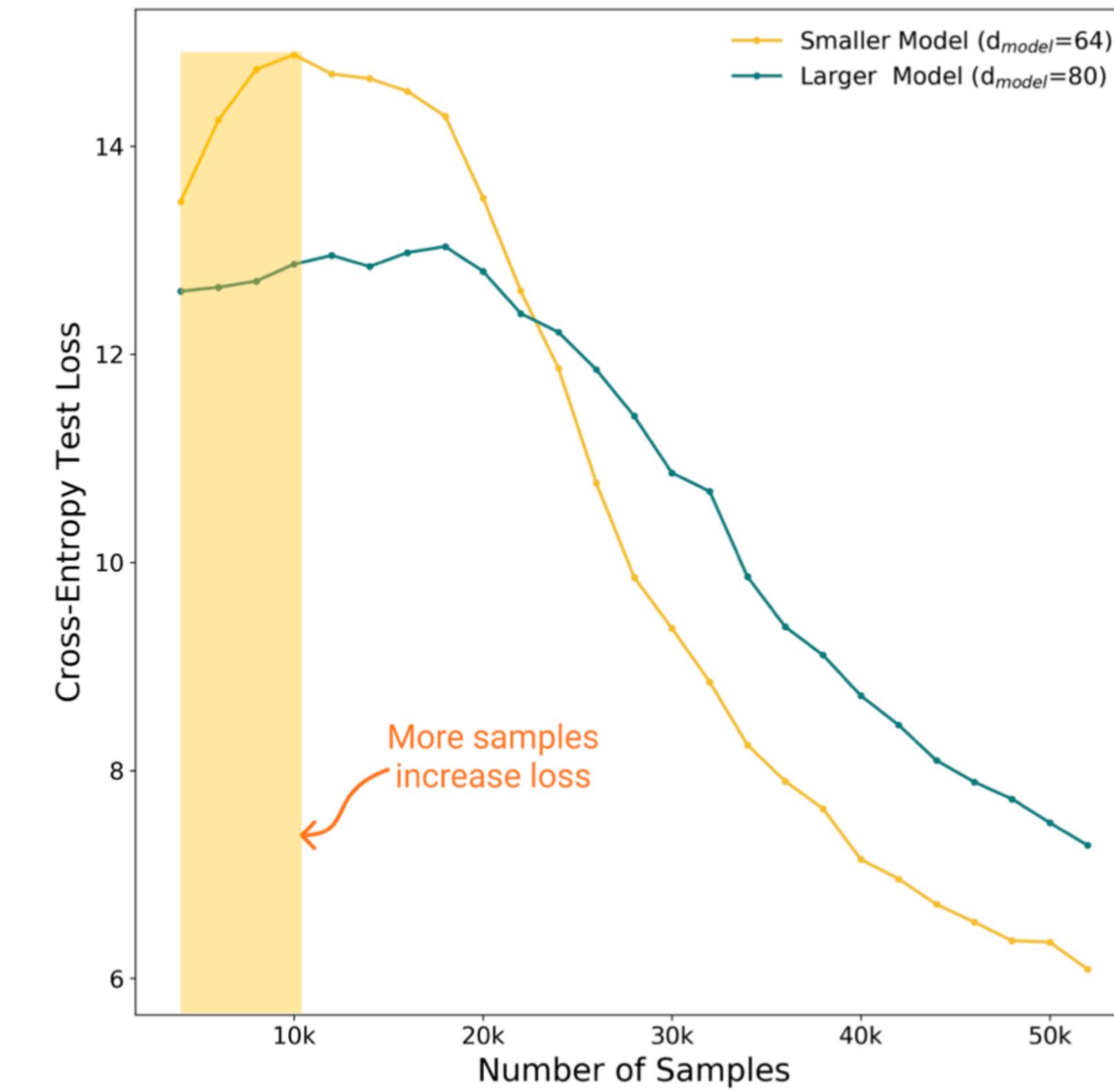
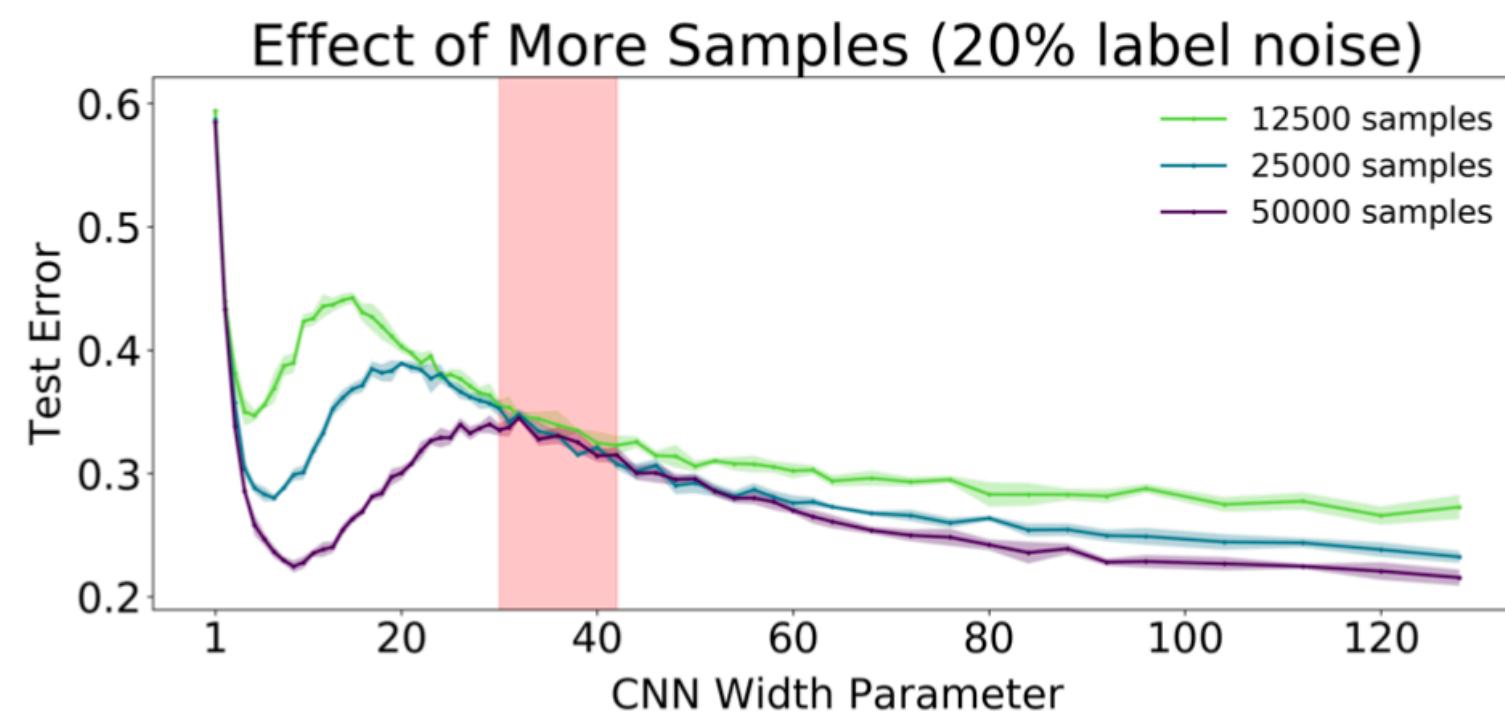
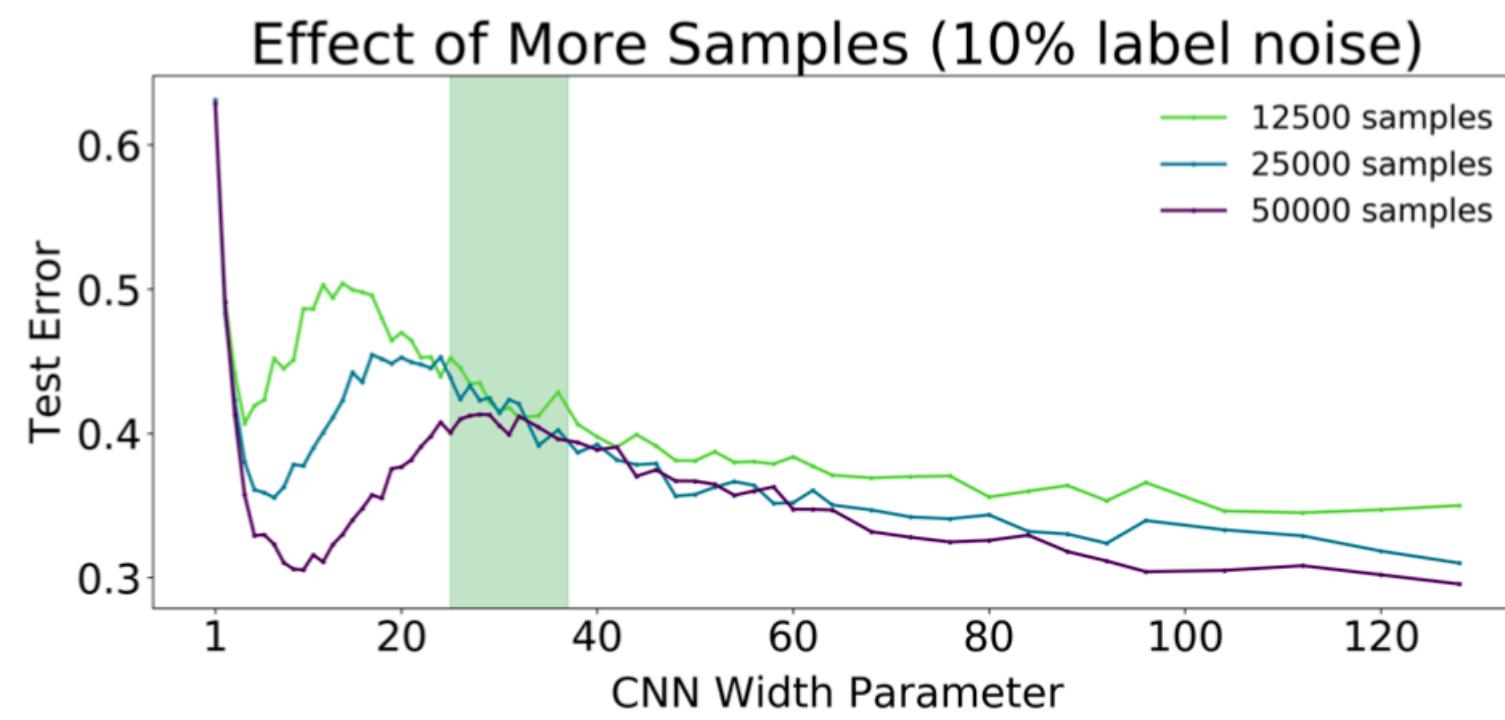
(a) ResNet18 on CIFAR10.

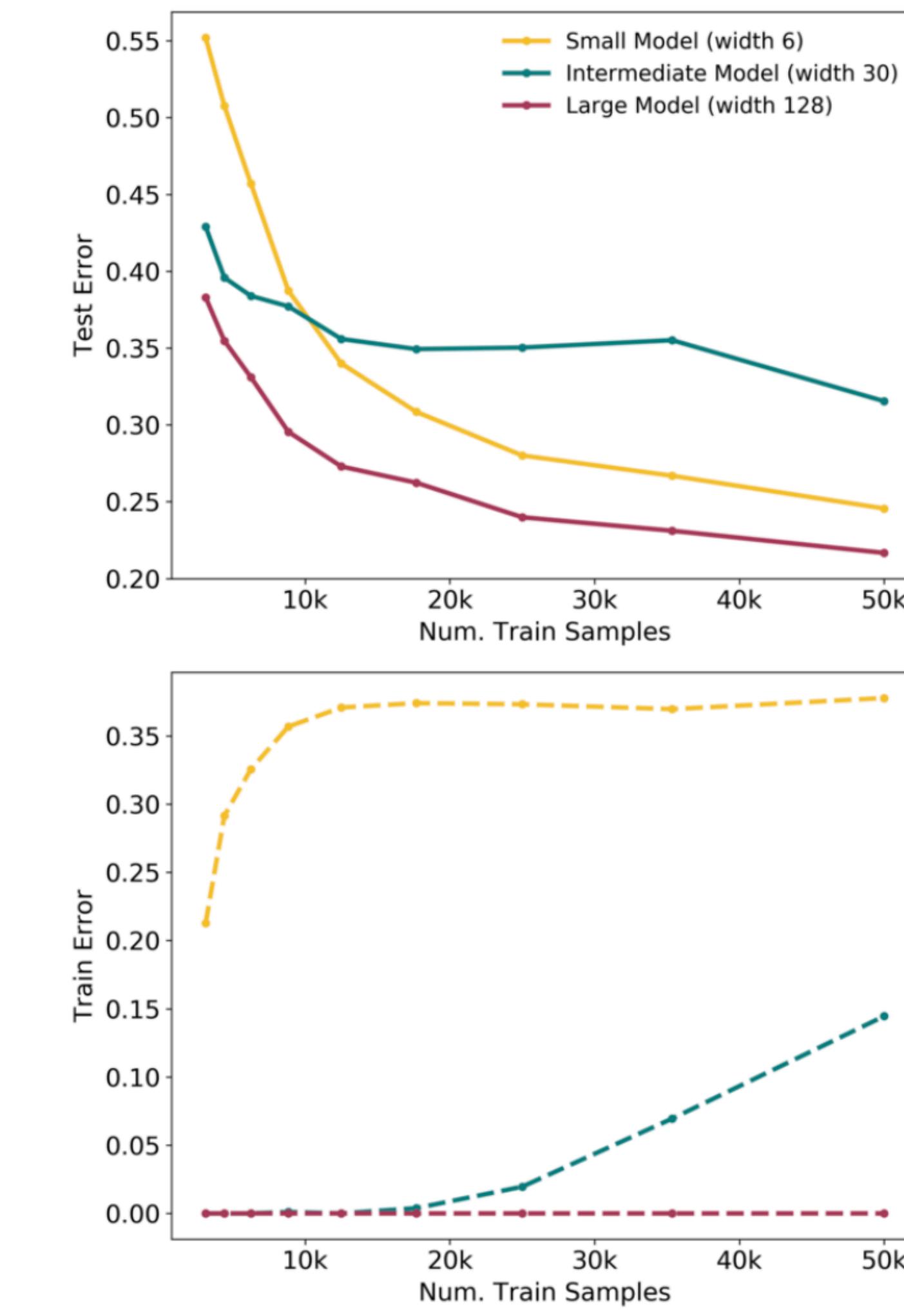
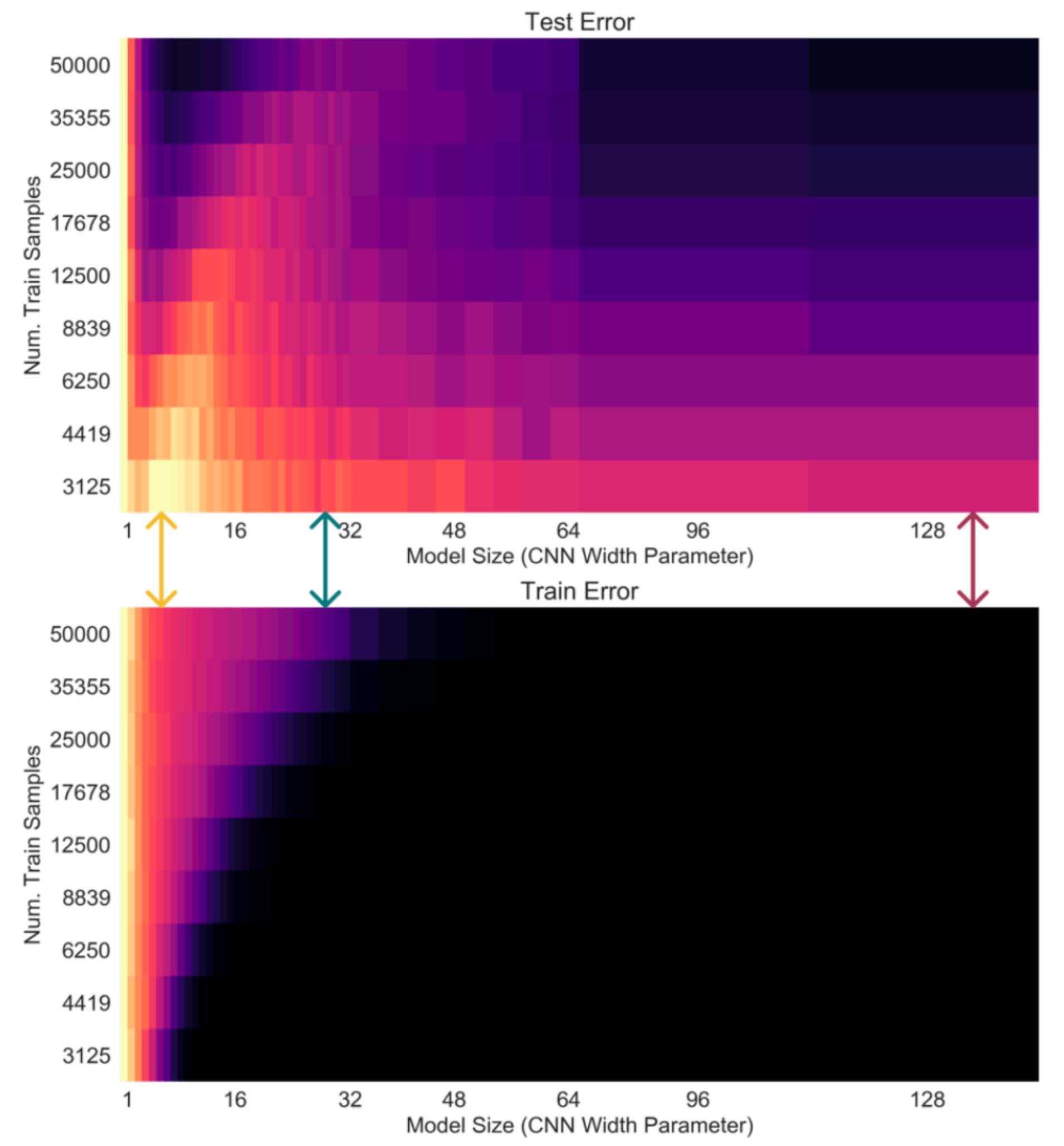


(b) ResNet18 on CIFAR100.



(c) 5-layer CNN on CIFAR 10.





Источники

<https://pdfs.semanticscholar.org/f86f/1748d1b6d22870f4347fd5d65314ba800583.pdf>

<https://arxiv.org/abs/1912.02292>