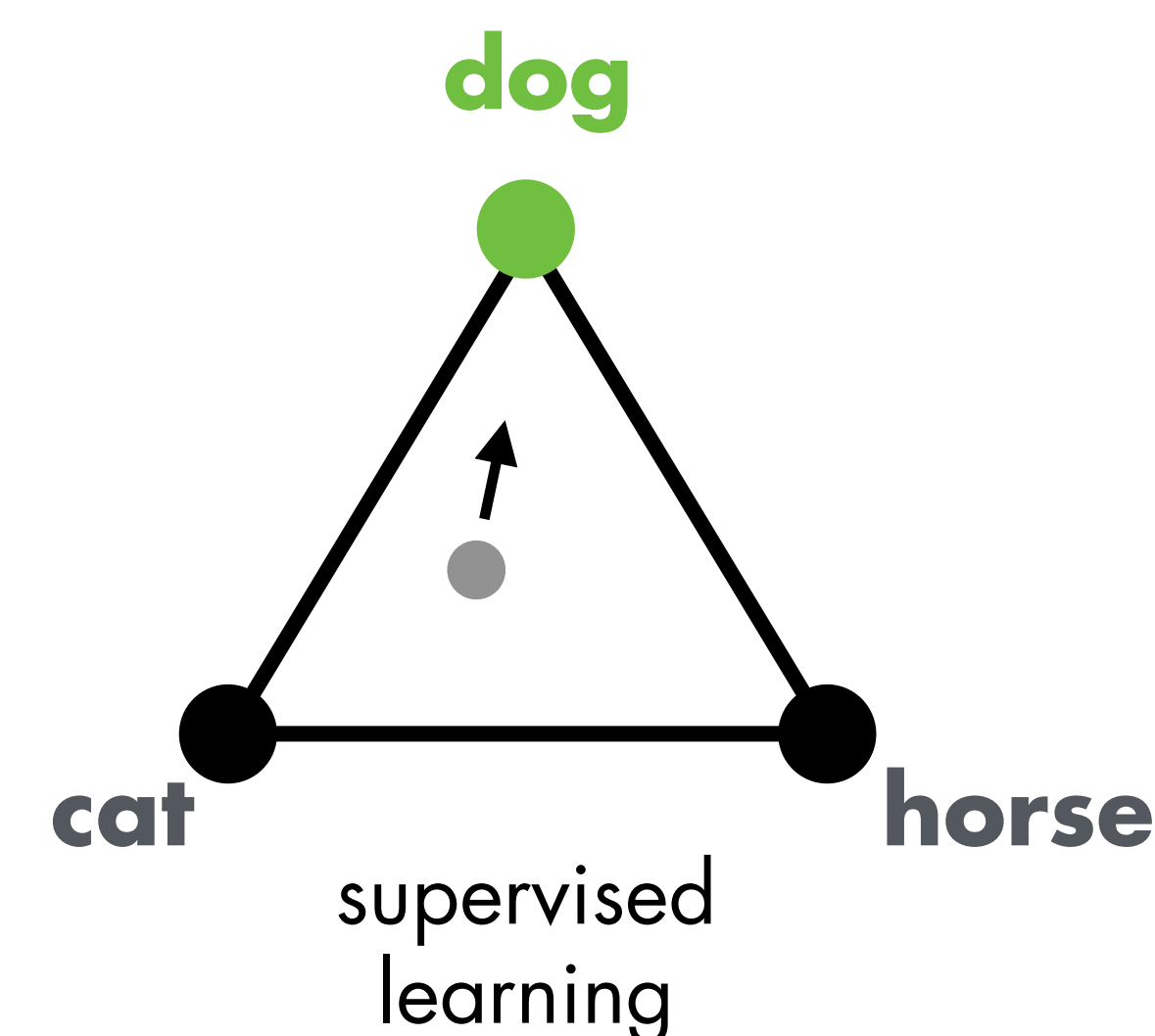


MEAN TEACHERS ARE BETTER ROLE MODELS

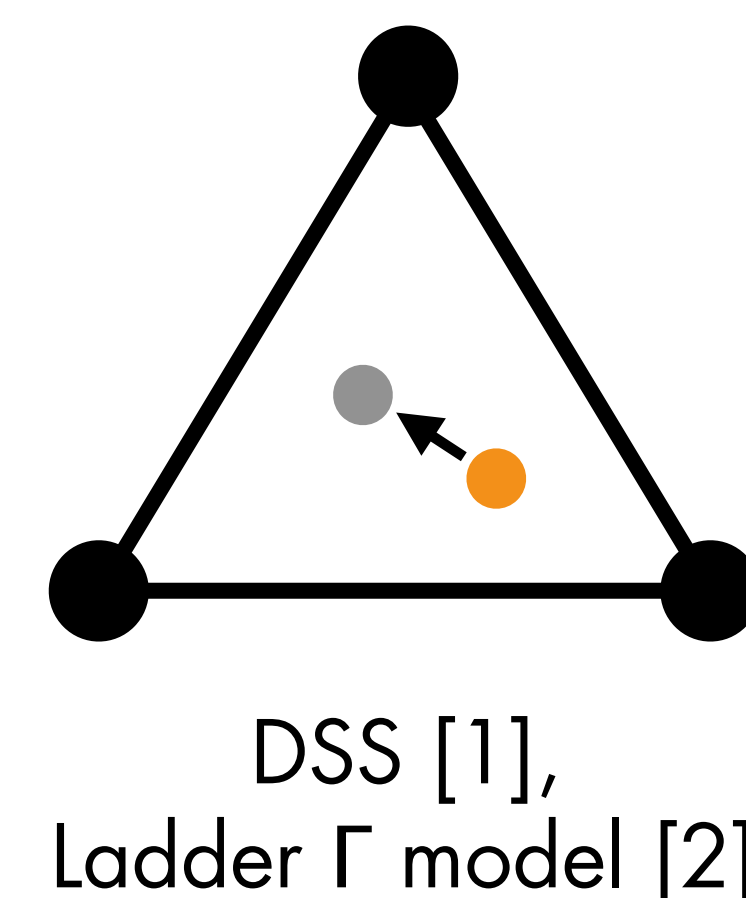
Weight-averaged consistency targets improve semi-supervised deep learning results

HISTORY TOUR: TRAINING TO BE CONSISTENT WITH SELF-GENERATED LABELS

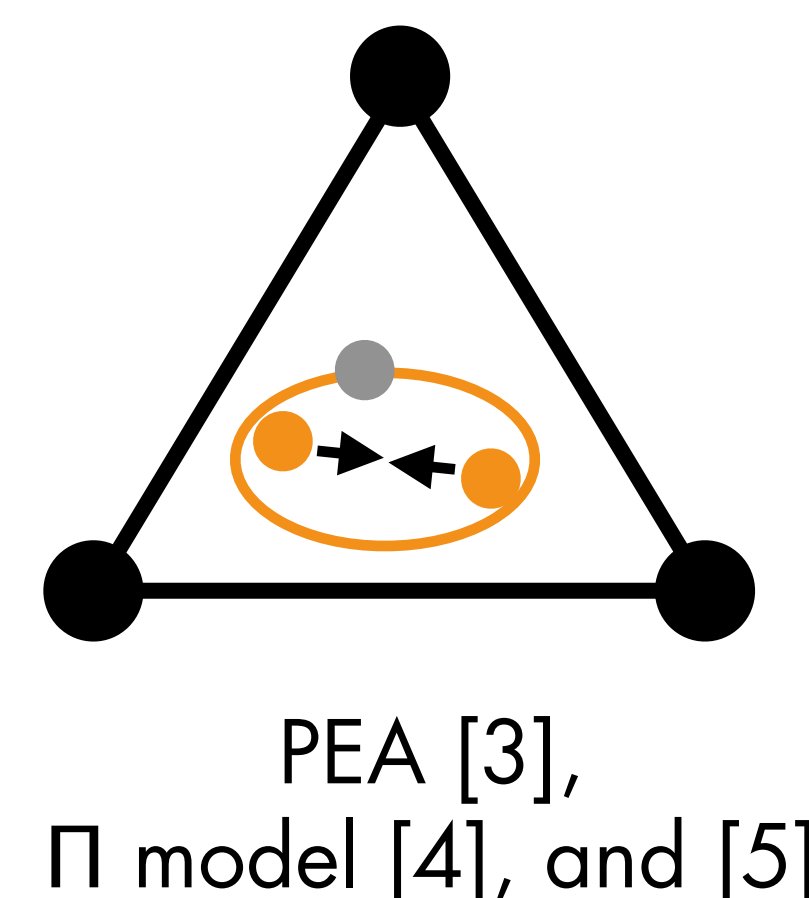
Let's train a classifier to recognise images of cats, dogs, and horses. A **known label** pulls the **prediction** to its direction.



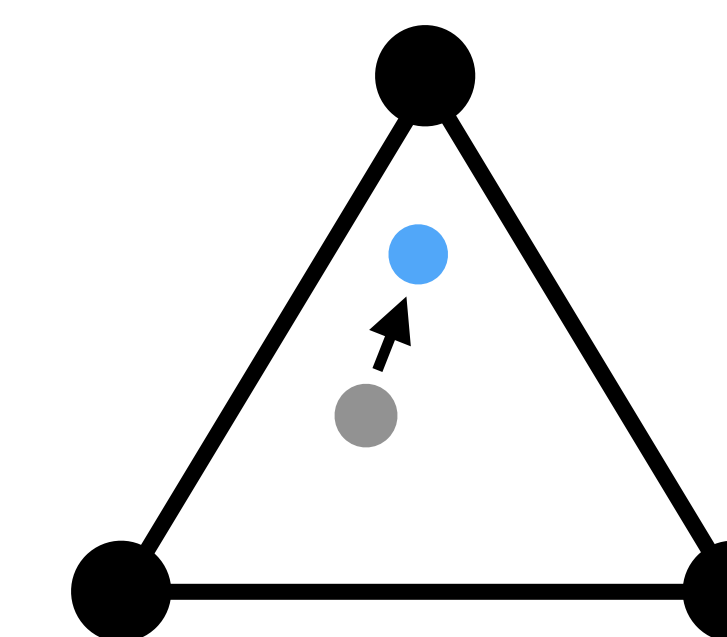
To train on an unlabelled example, we can add noise to the input and let the **clean prediction** pull the **noisy prediction**.



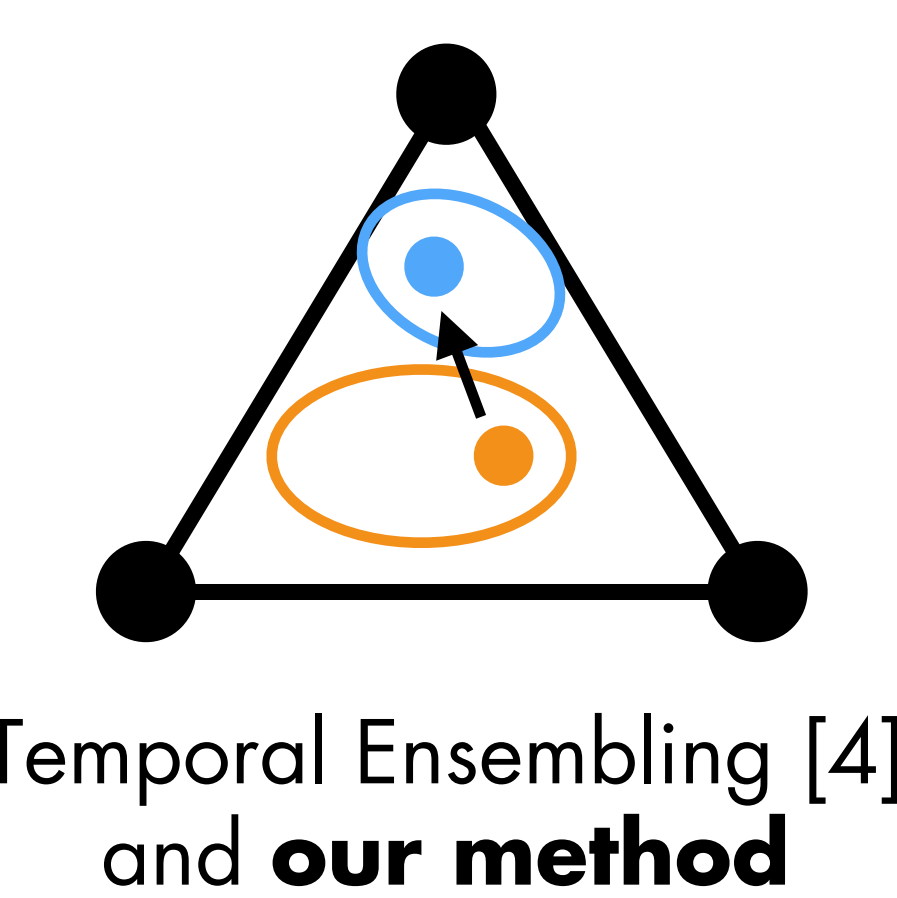
But the **clean prediction** may itself be an outlier, so it is better to let **two noisy predictions** pull each other.



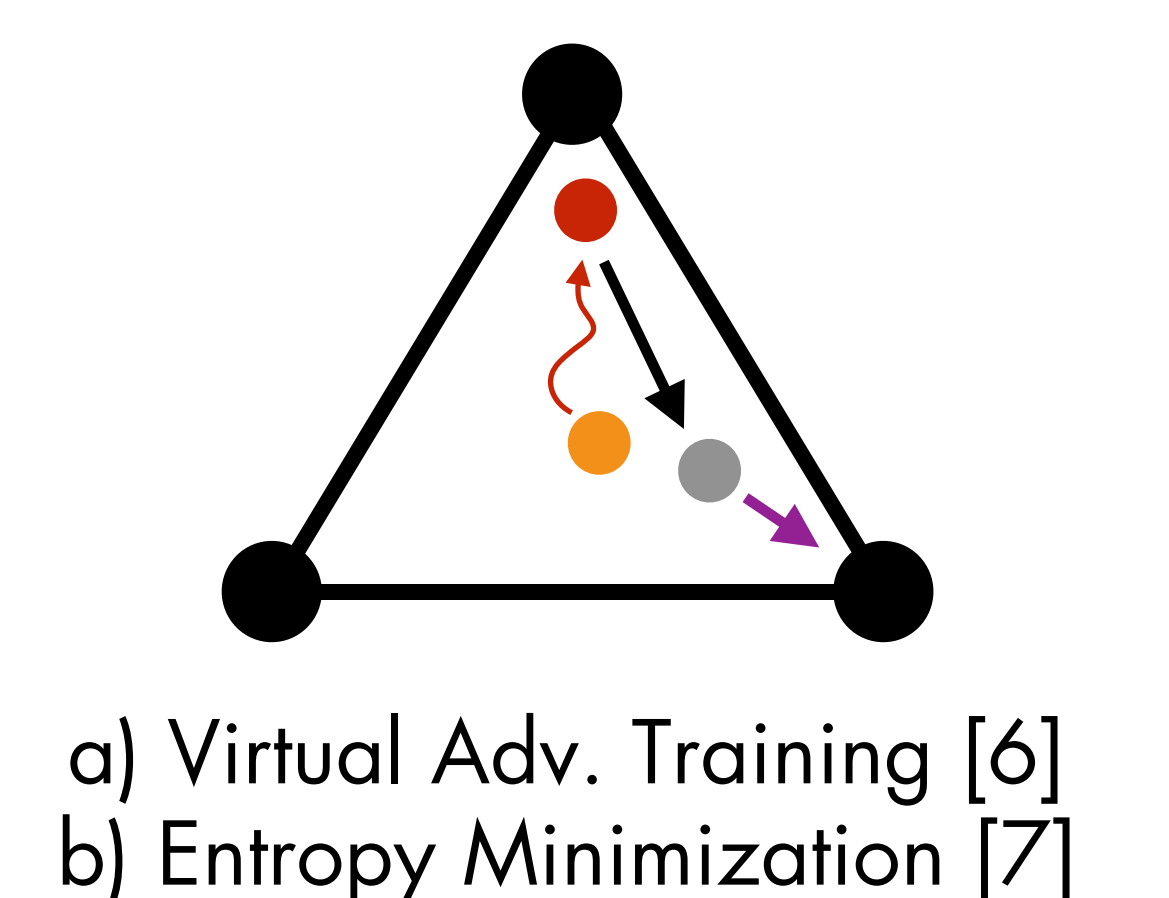
Another way is to improve the **prediction** by (pseudo-)ensembling many models to form a **teacher prediction**.



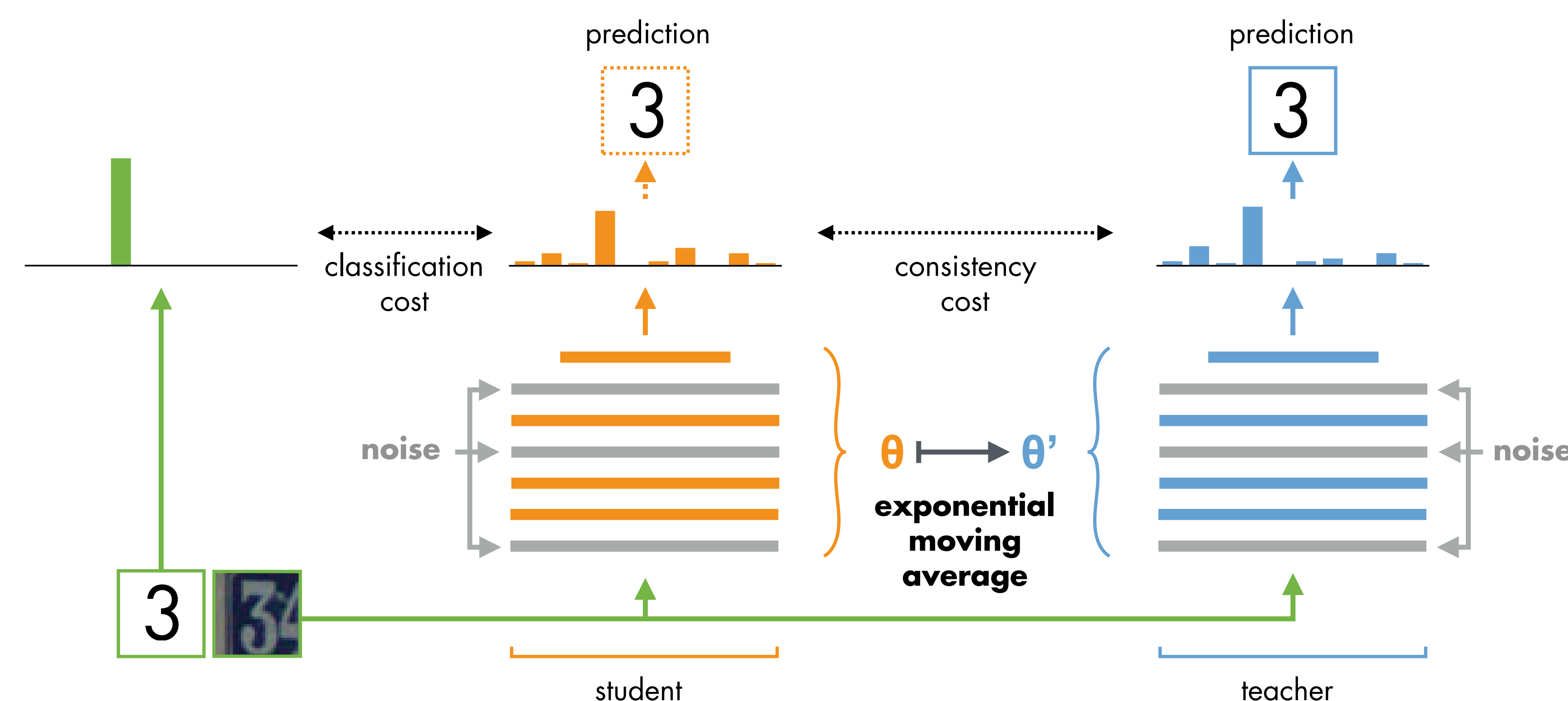
Combining these ideas works even better: a noisy pseudo-ensembled **teacher prediction** pulls a noisy **student prediction**.



Other approaches:
a) make **noisy sample adversarial**, then pull.
b) pull towards the **closest class**.



The model



Whereas Temporal Ensembling [4] averages student predictions over time to form teacher predictions, we average **student weights** over time to form a **mean teacher**. This improves test accuracy, enables training with fewer labels and works on large datasets.

Results

ImageNet using 10% of the labels	
Variational Auto-Encoder [8]	top-5 validation error 35.42 \pm 0.90
Mean Teacher ResNet-18	19.76 \pm 0.05
Mean Teacher ResNet-152	9.11 \pm 0.12
All labels, state of the art [9]	3.79

CIFAR-10 using 4000 labels	
Supervised only	test error 20.66 \pm 0.57
Π model [4]	12.36 \pm 0.31
Temporal ensembling [4]	12.16 \pm 0.31
Virtual Adversarial Training + EntMin [5]	10.55
CT-GAN [10]	9.98 \pm 0.21
Mean Teacher CNN	12.31 \pm 0.28
Mean Teacher ResNet-26	6.28 \pm 0.15
All labels, state of the art [11]	2.86

