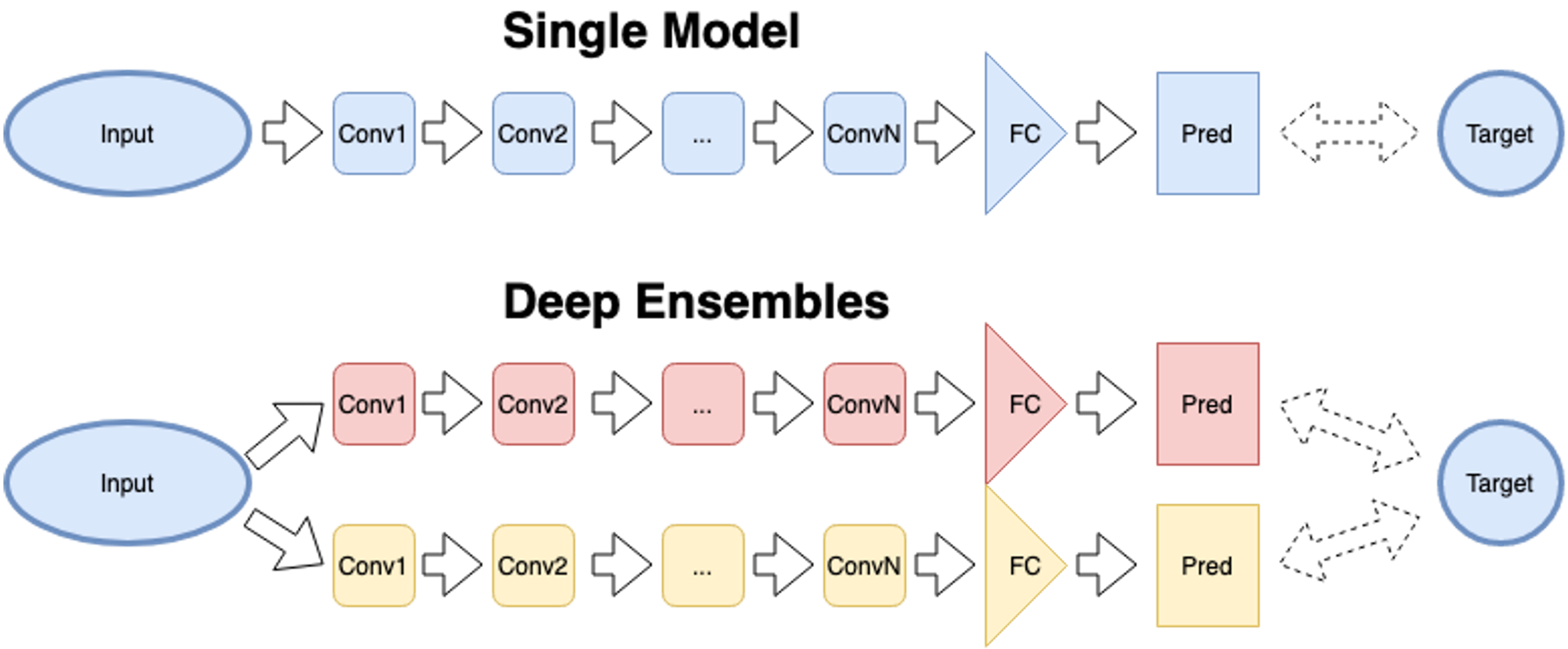
MixMo:

Mixing Multiple Inputs for Multiple Outputs via Deep Subnetworks

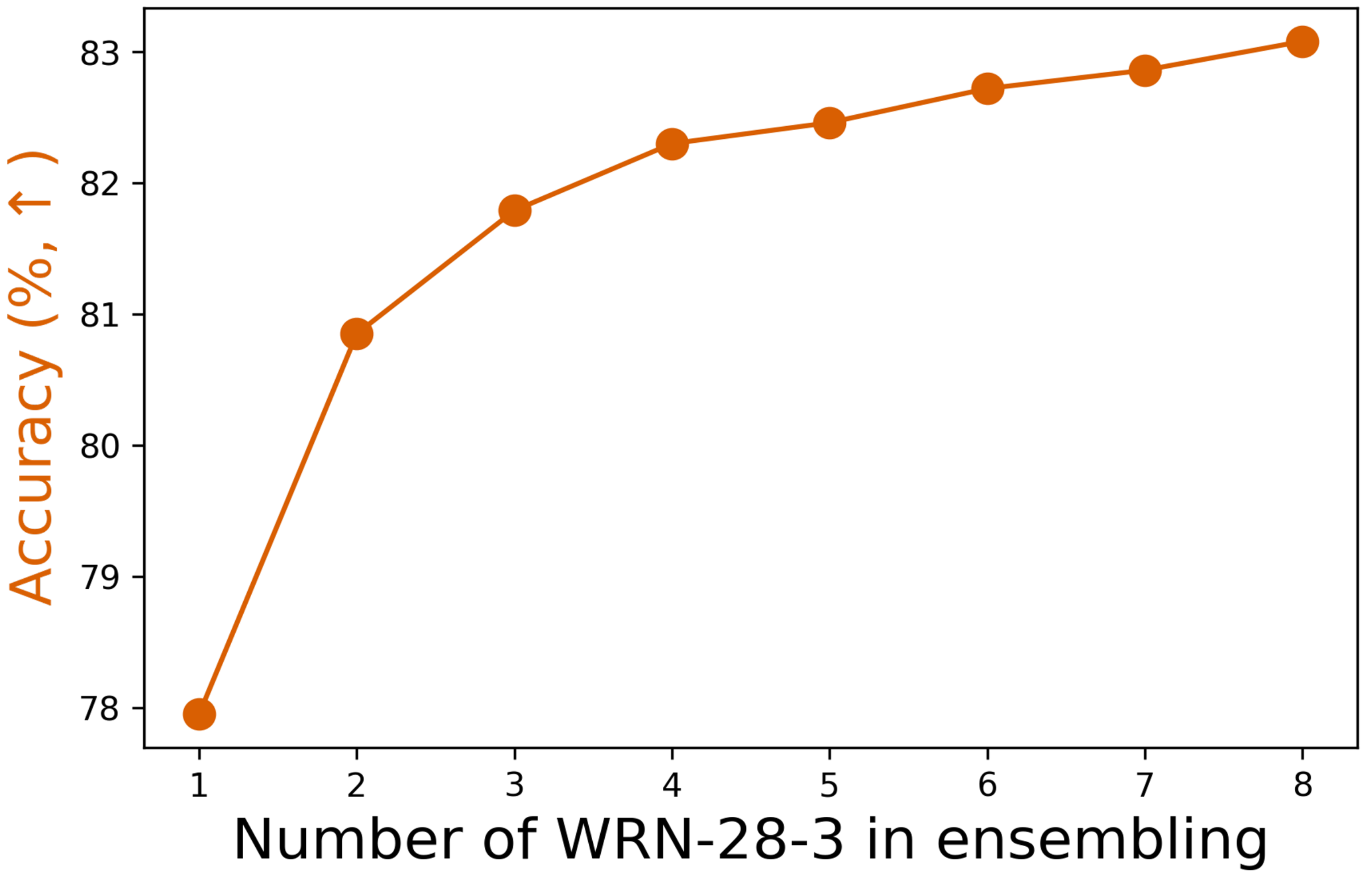
Alexandre Ramé, Rémy Sun
and Matthieu Cord

Deep ensembles



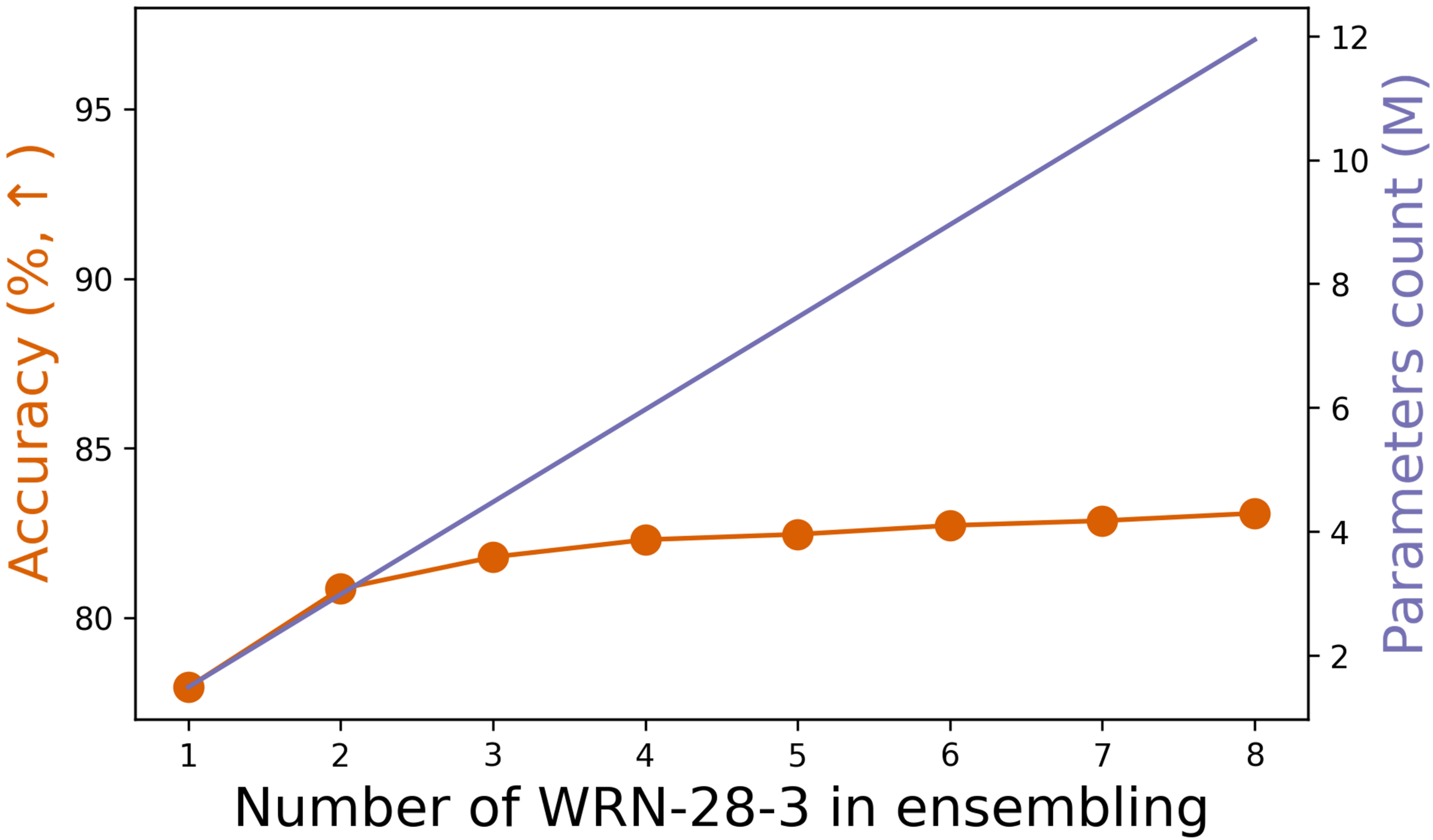
[1] Simple and scalable predictive uncertainty estimation using deep ensembles. Lakshminarayanan *et al.*, in *NeurIPS* 2017.

Ensembling improves accuracy ...

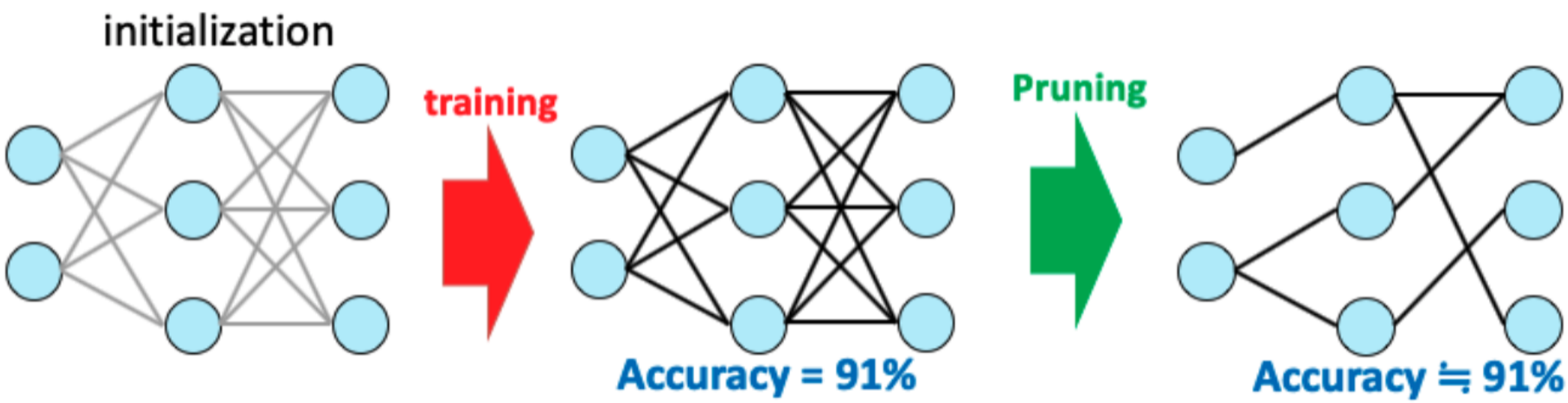


[1] On Power Laws in Deep Ensembles. Lobacheva *et al.*, in *NeurIPS* 2020.

… but ensembling is costly



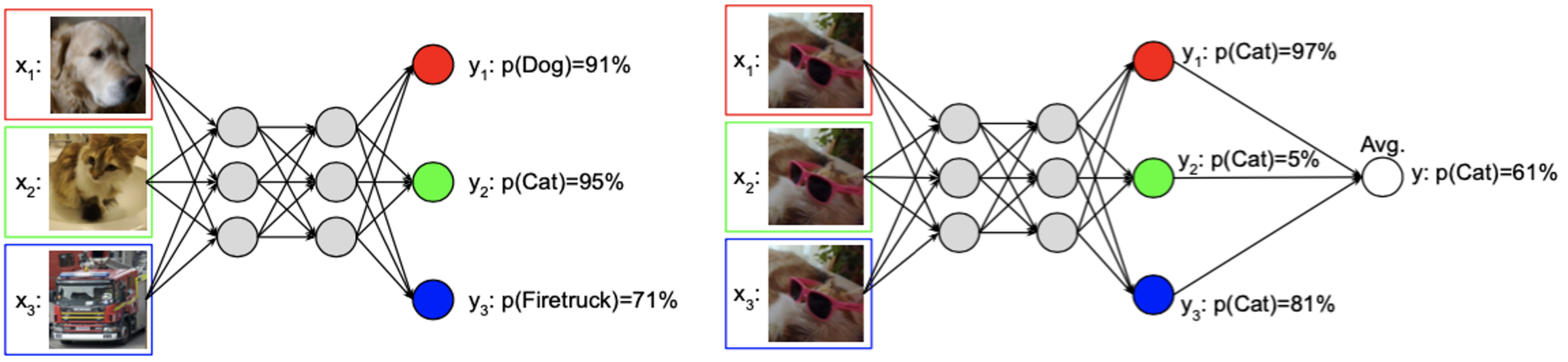
Idea: leverage network sparsity



[1] Pruning filters for efficient convnets. Li *et al.,* ICLR 2017
[2] The lottery ticket hypothesis: Finding sparse, trainable neural networks. Jonathan Frankle and Michael Carbin, ICLR 2019

Main idea: multiple subnetworks inside one base network

Multi-input Multi-output strategy

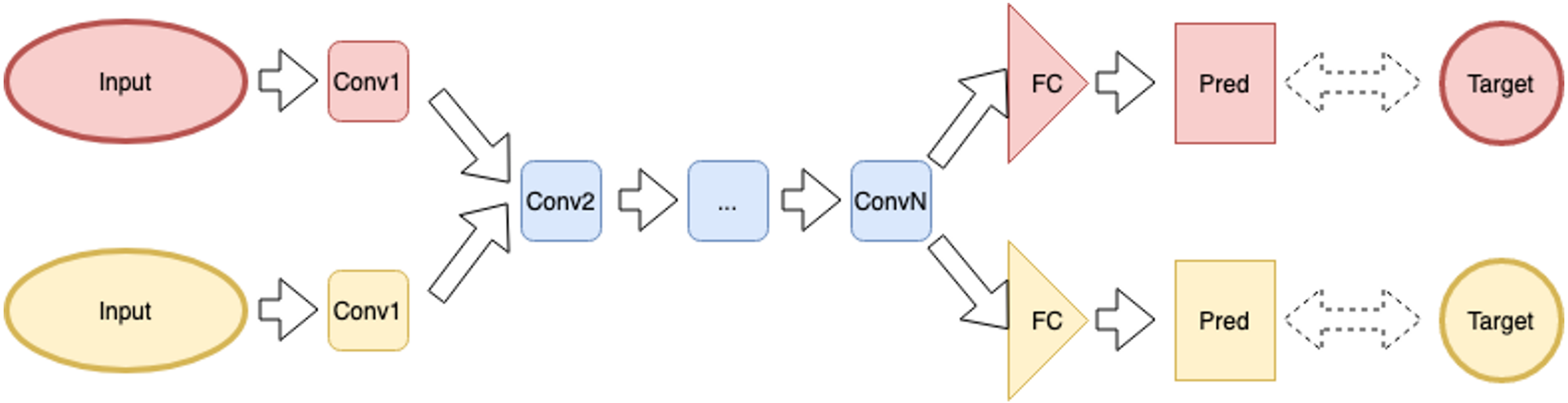


(a) Training

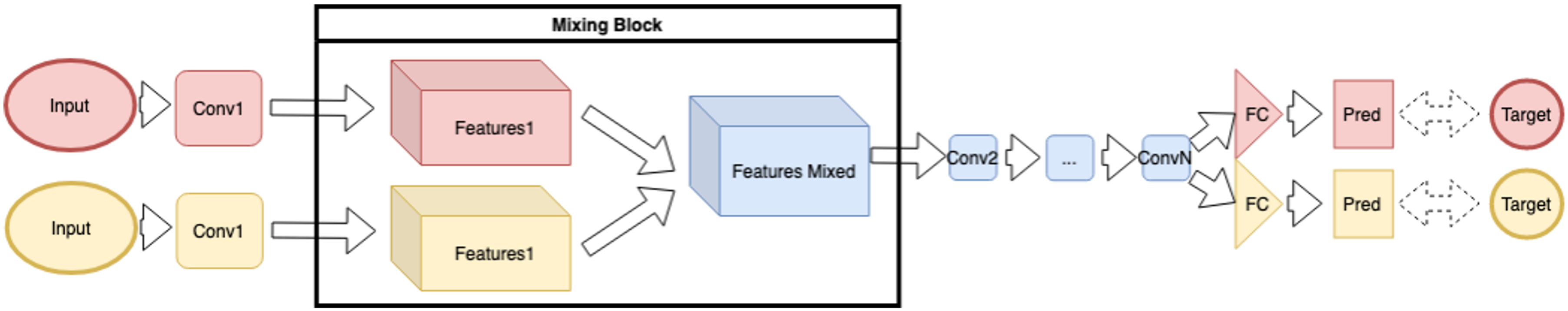
(b) Testing

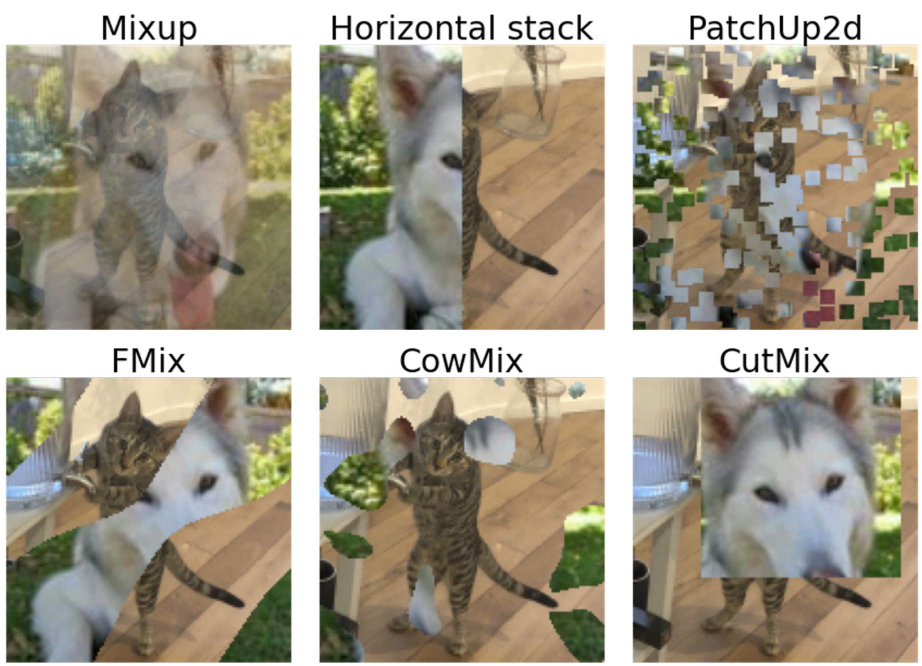
[1] Training independent subnetworks for robust prediction. Havasi *et al.*, ICLR 2021

Architecture: all layers shared except the encoders and the classifiers



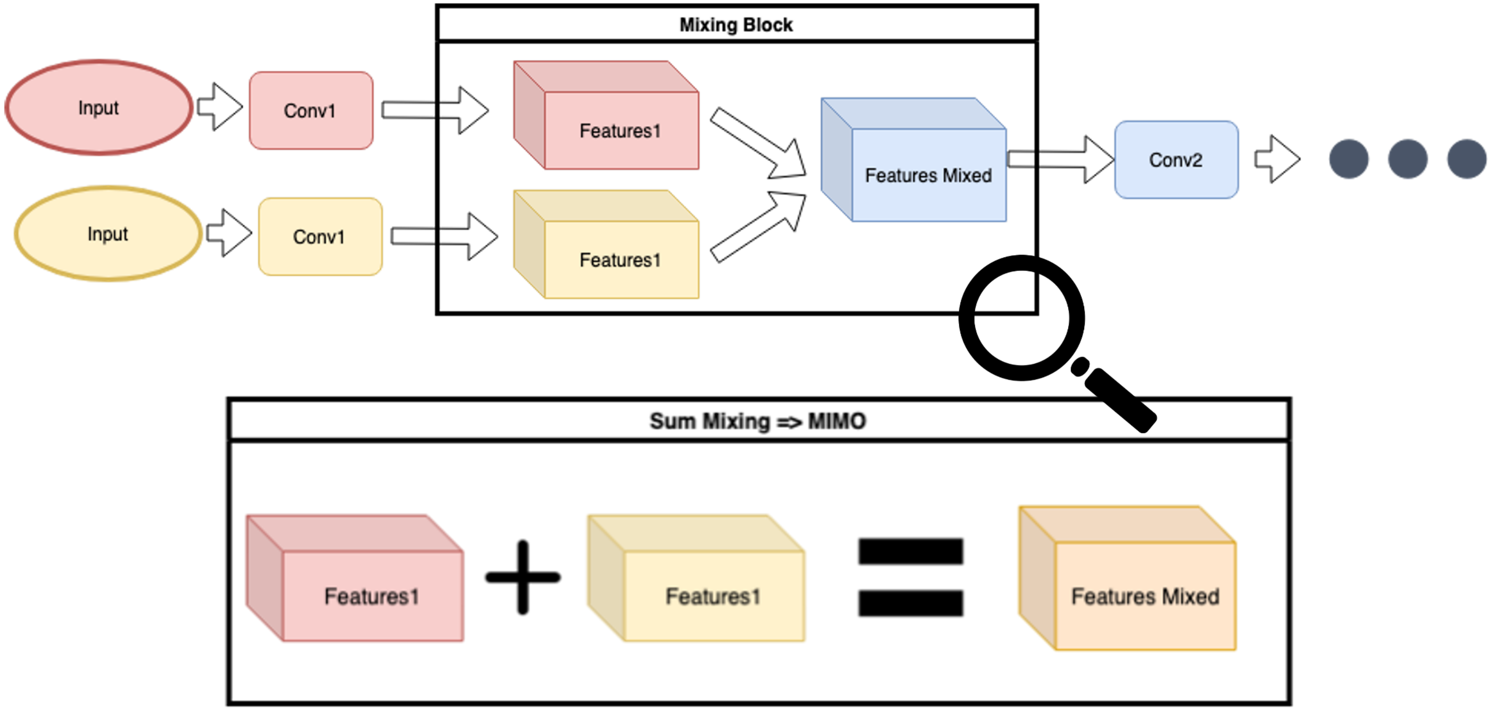
Mixing block combining the inputs' features





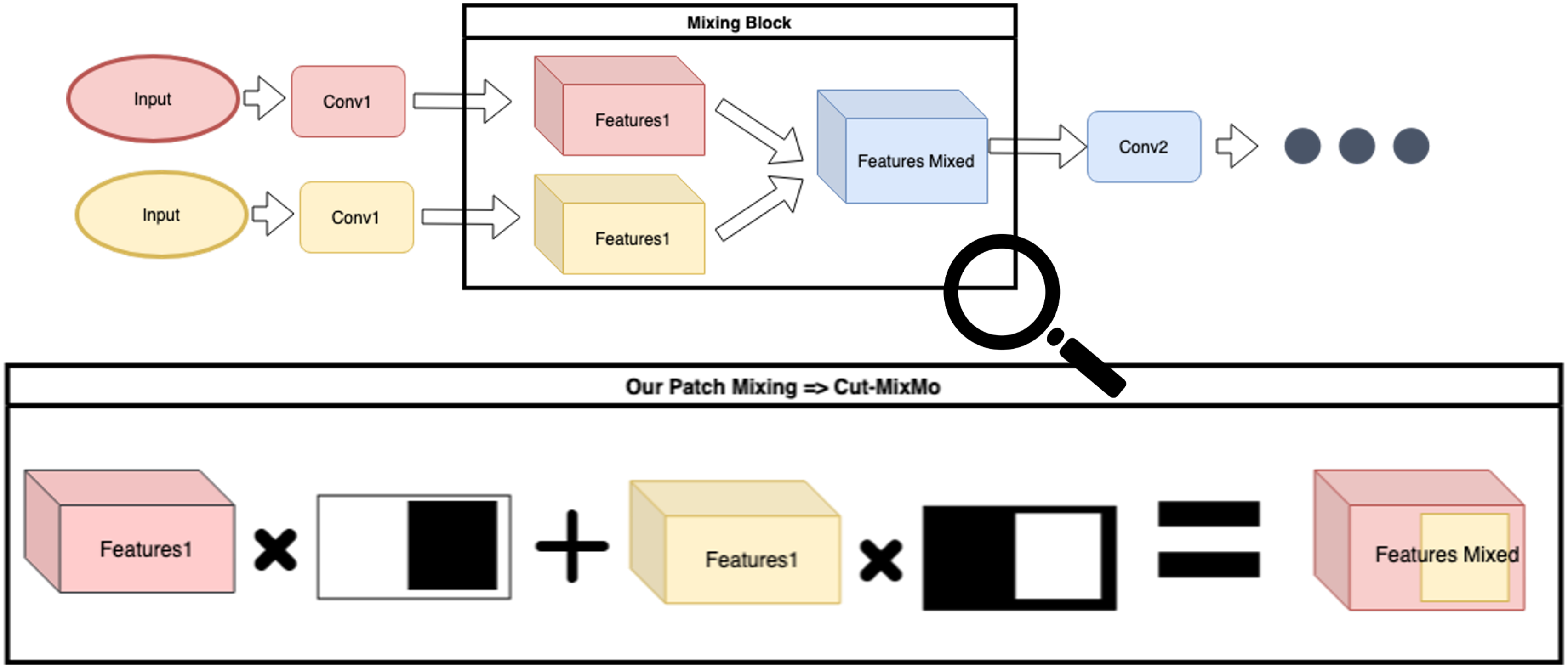
Mixup as mixing block

Mixup

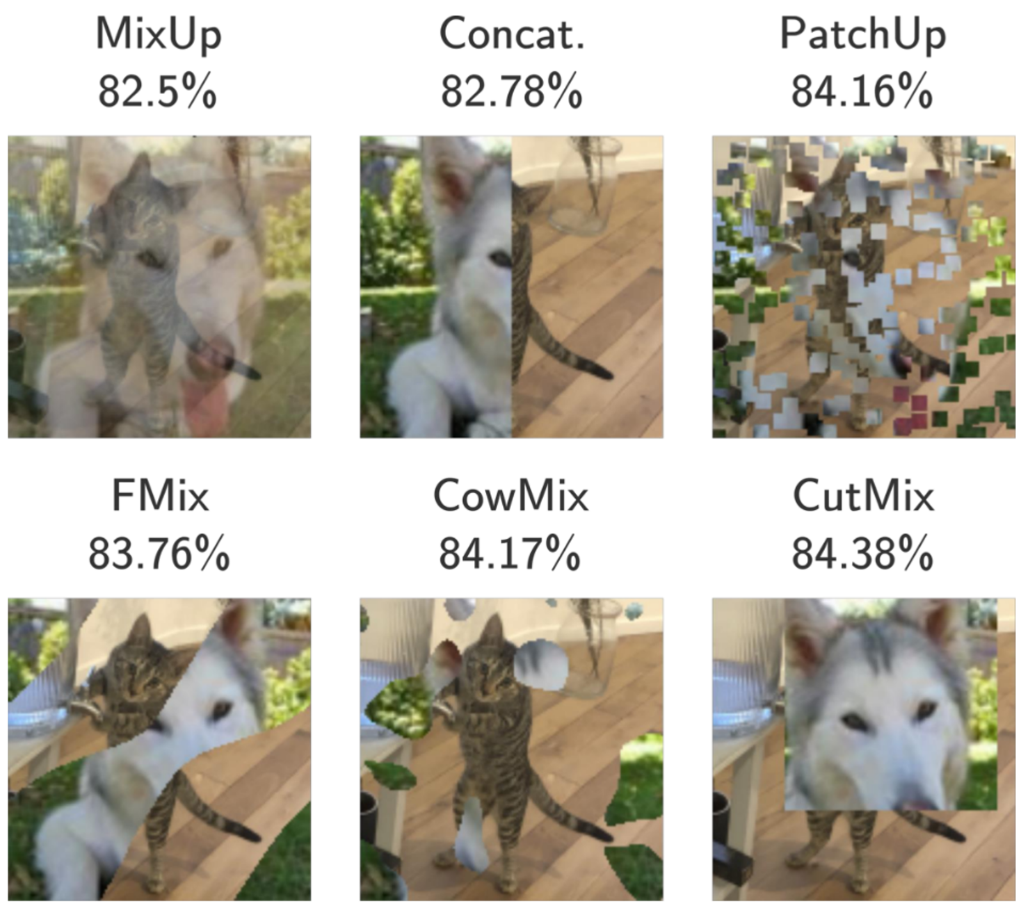


[1] Training independent subnetworks for robust prediction. Havasi *et al.*, ICLR 2021

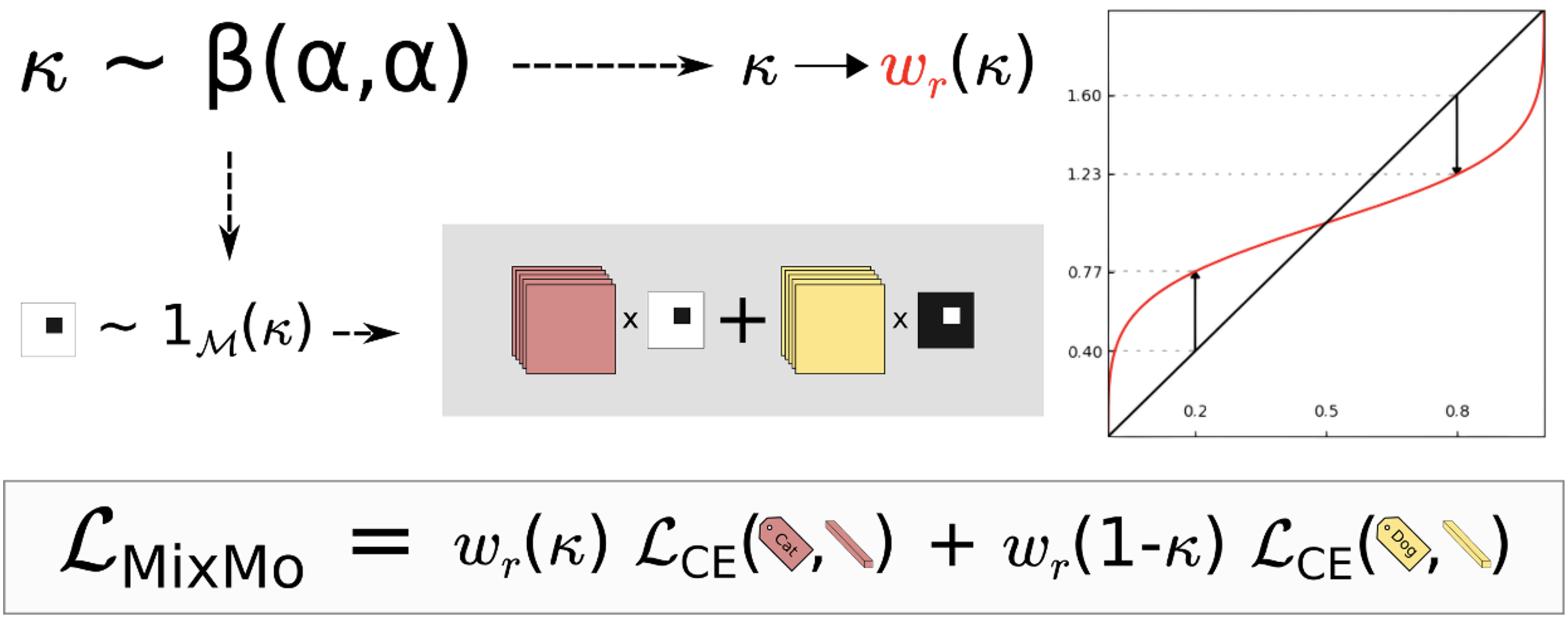
CutMix as mixing block



Binary mixing as mixing block improves individual accuracies & ensemble diversity



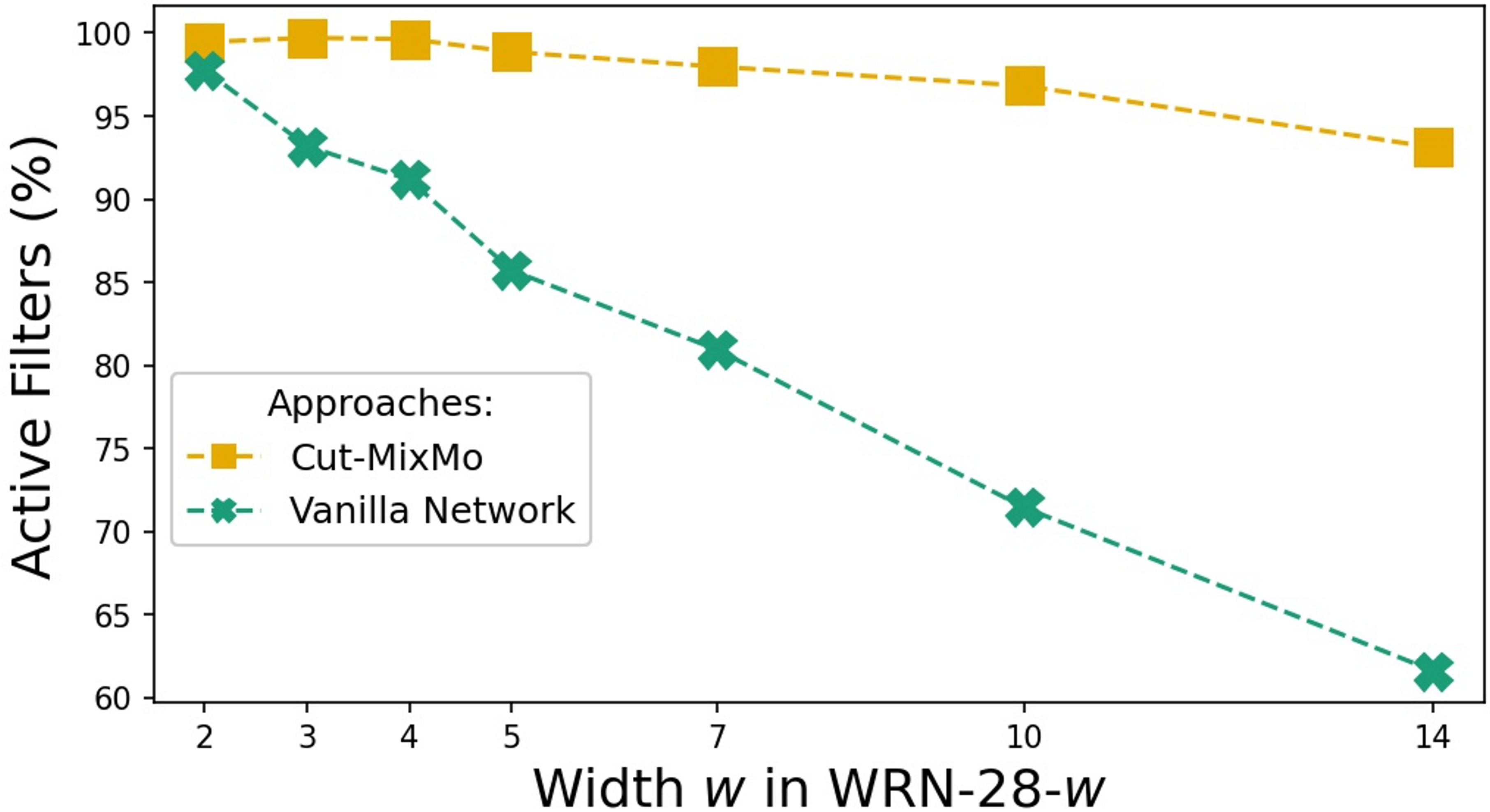
Balancing the training losses



State of the art on CIFAR and TinyImageNet

Approach	#Params	WRN-28-10		ResNet-18-3
		CIFAR100	CIFAR10	TinyImageNet
Vanilla	1.0	81.63	96.34	65.78
CutMix	1.0	84.05	97.23	68.95
Deep Ens.	2.0	83.17	96.67	68.38
MIMO	1.002	83.06	96.74	68.48
Cut-MixMo	1.002	**85.40**	**97.51**	**70.24**

Better leverages over-parameterization



Contributions

❖ Theoretically

Unifying framework for multi-input multi-output ensembling

Connection with data augmentation

❖ Empirically

State of the art at same inference cost as a vanilla network

More in paper: ImageNet, robustness, memory split advantage ….

https://github.com/alexrame/mixmo-pytorch

Merci !