



Addressing the Loss-Metric Mismatch with Adaptive Loss Alignment



Chen Huang, Shuangfei Zhai, Walter Talbott, Miguel Bautista, Shih-Yu Sun, Carlos Guestrin, Josh Susskind
Apple Inc.

Loss-Metric Mismatch

Loss
function

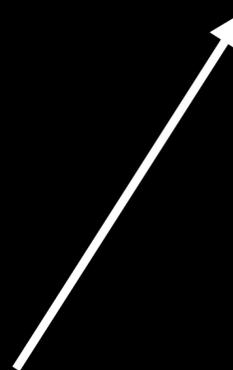
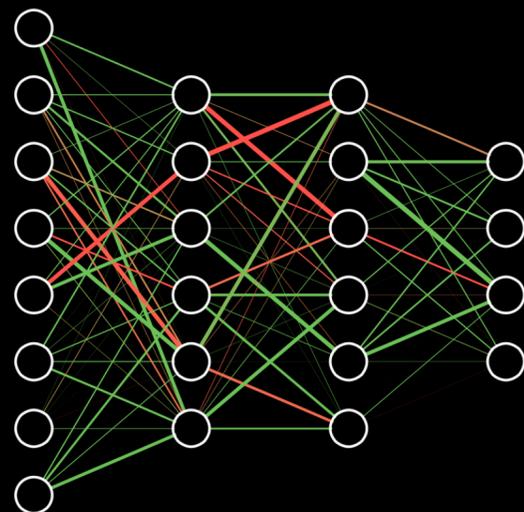
\neq

Evaluation
metric

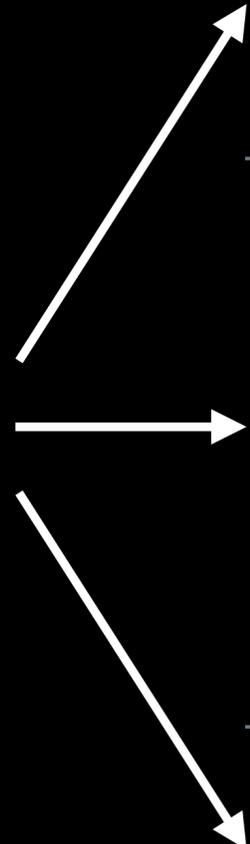
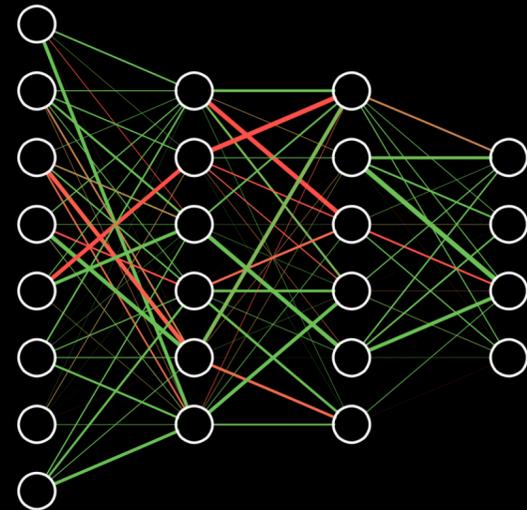
Softmax

Classification error
AUCPR

Classification
Cat



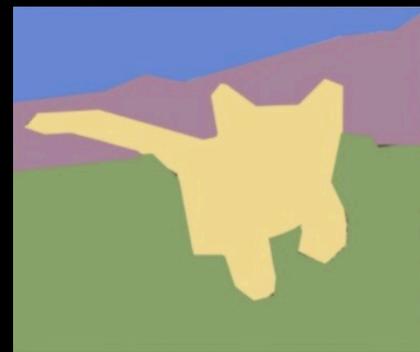
Loss-Metric Mismatch



Classification
Cat



Detection



Segmentation

Loss
function

≠

Evaluation
metric

Softmax

Classification error
AUCPR

L1 loss

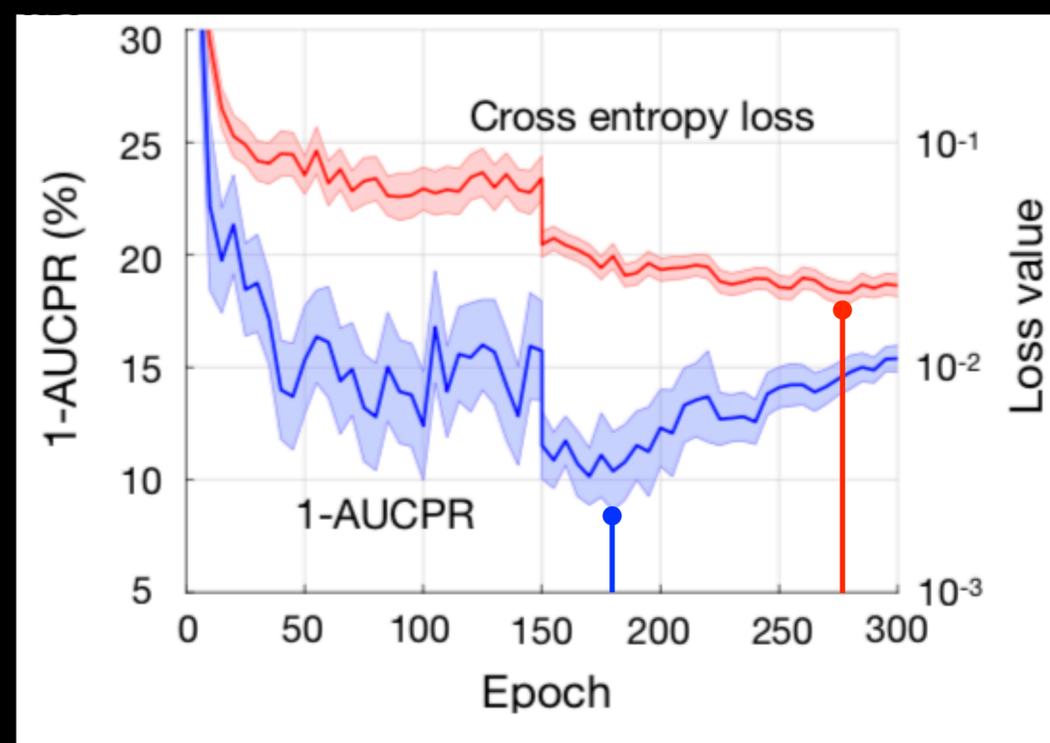
Average
Precision

Cross
Entropy

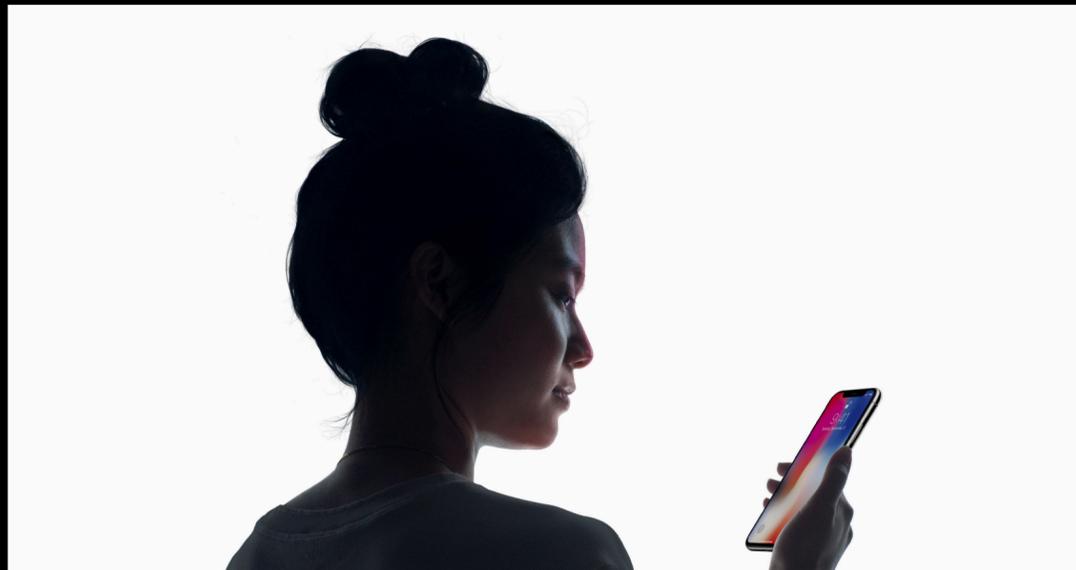
IoU

Loss-Metric Mismatch

- Loss functions are often designed to be differentiable (convex and smooth)
- Many evaluation metrics are nonlinear, non-continuous, non-decomposable



Adaptive Loss Alignment (ALA) at Apple



Loss function

vs.

Evaluation metrics:

False Acceptance Rate (FAR)

False Rejection Rate (FRR)

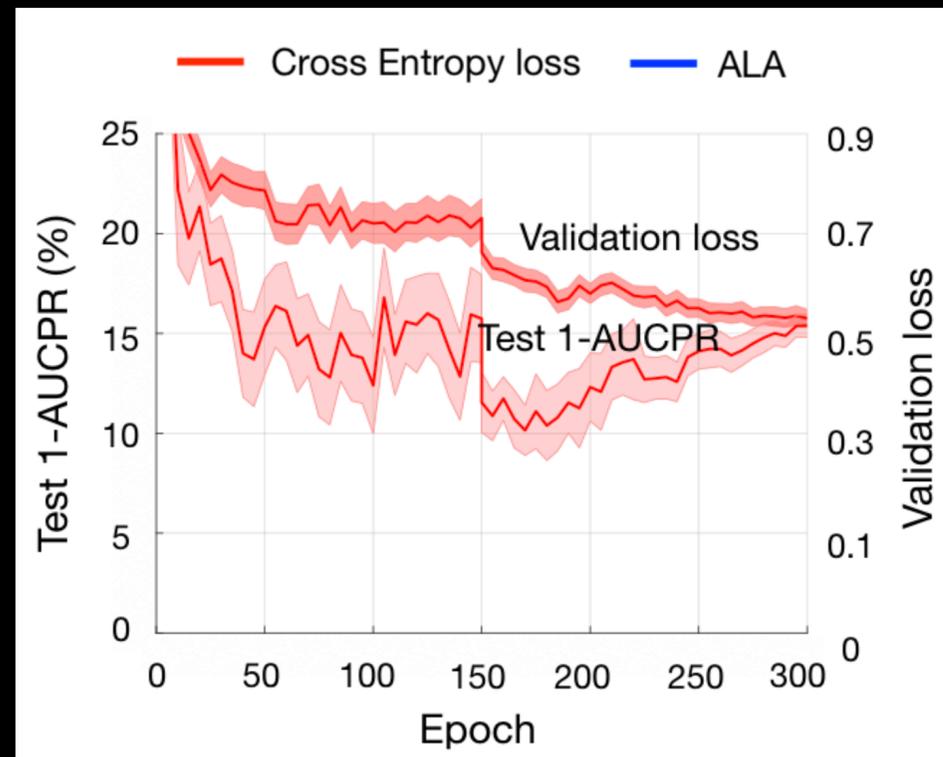
...

Adaptive Loss Alignment (ALA)

- Reinforcement learning of dynamic loss function
- Align loss with the target evaluation metric

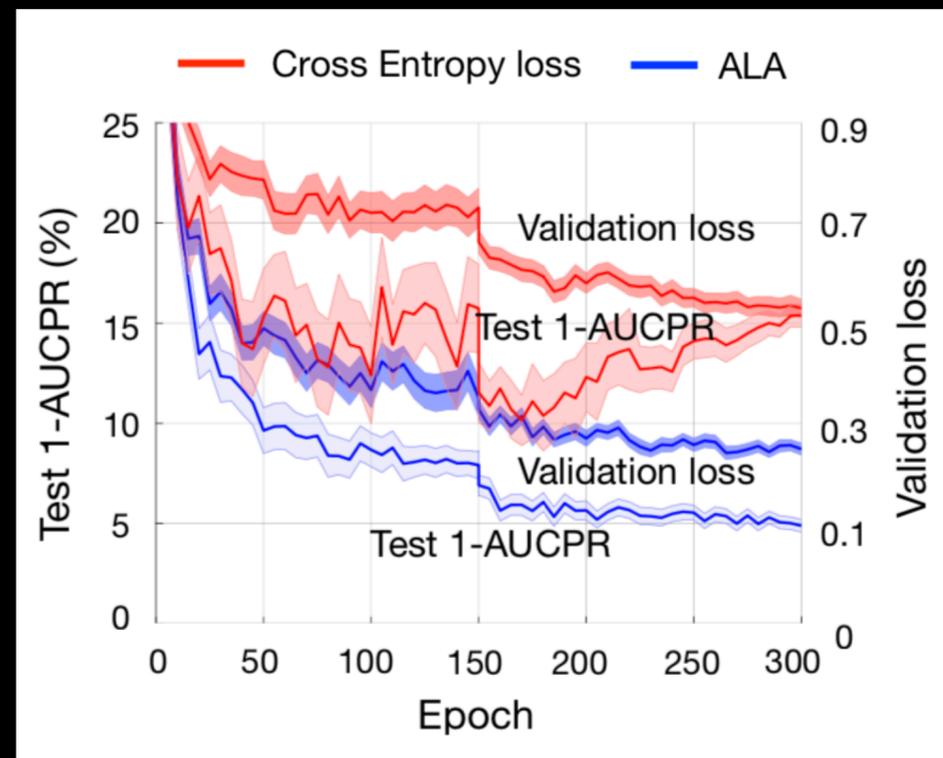
Adaptive Loss Alignment (ALA)

- Reinforcement learning of dynamic loss function
- Align loss with the target evaluation metric



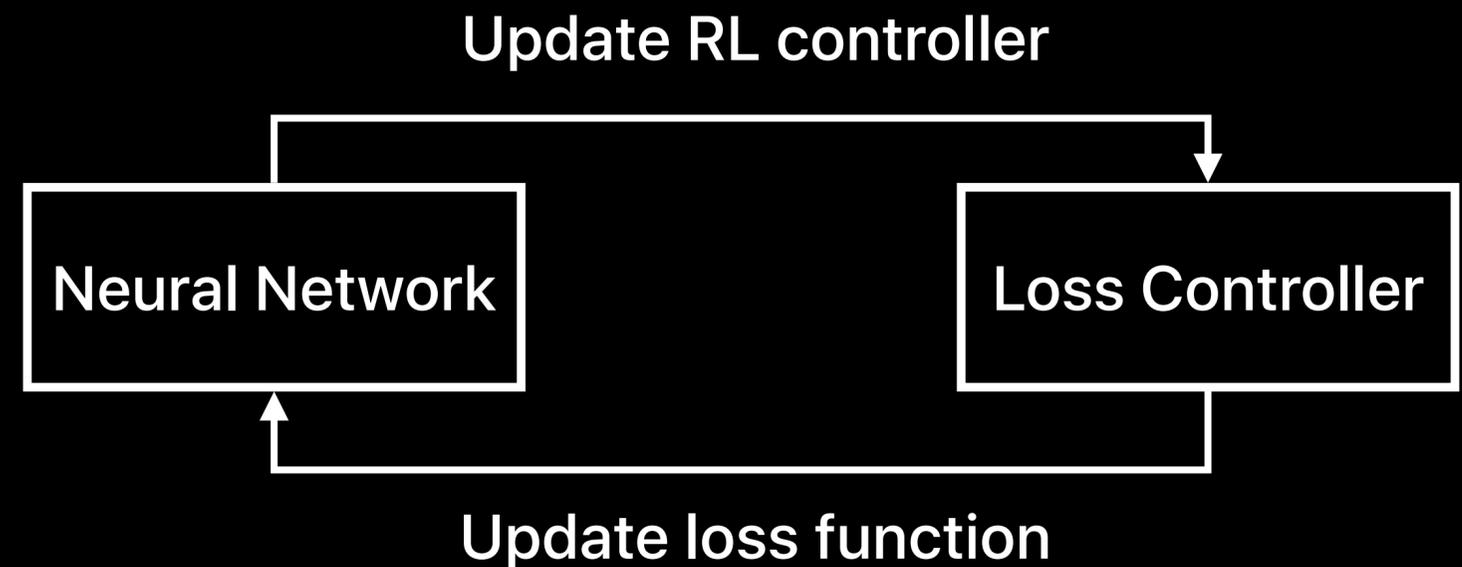
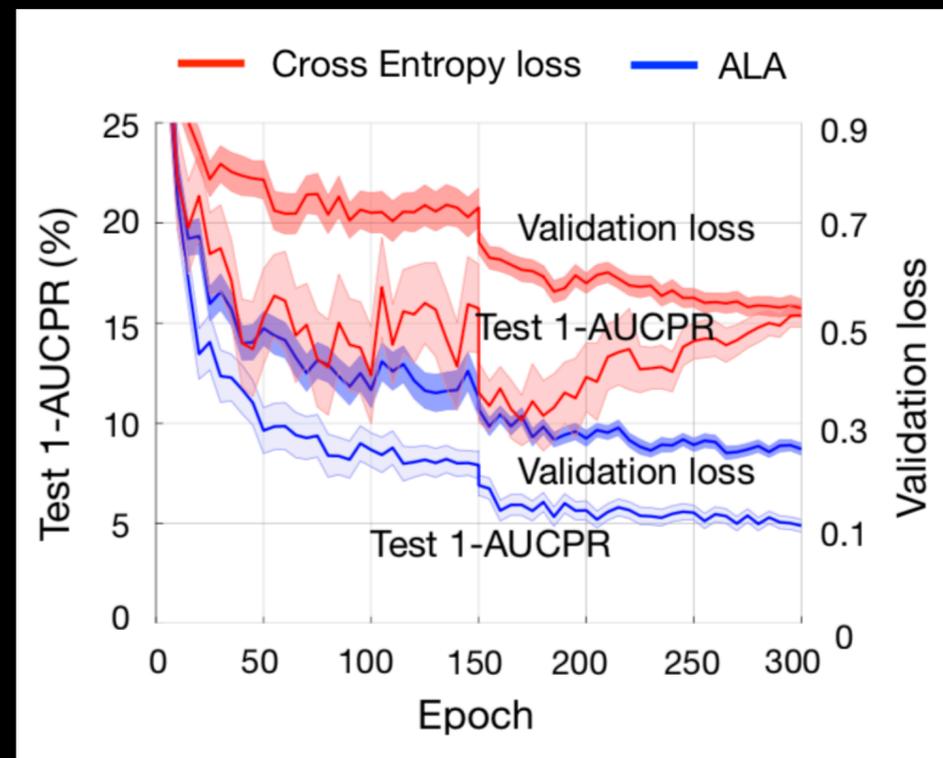
Adaptive Loss Alignment (ALA)

- Reinforcement learning of dynamic loss function
- Align loss with the target evaluation metric

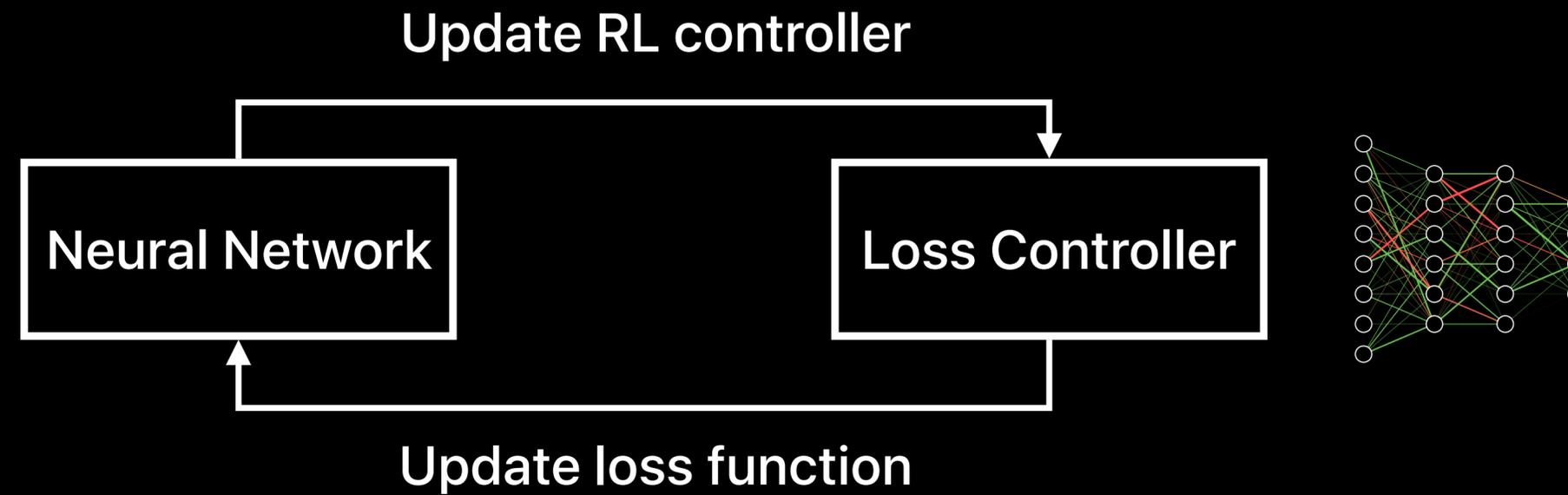


Adaptive Loss Alignment (ALA)

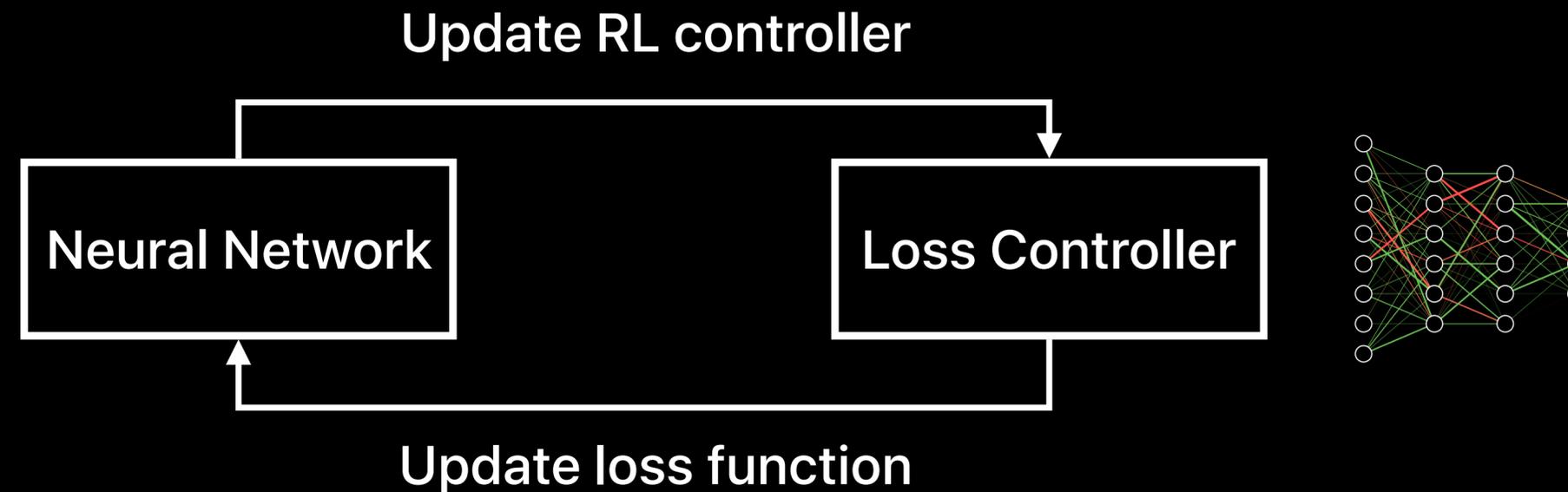
- Reinforcement learning of dynamic loss function
- Align loss with the target evaluation metric



Adaptive Loss Alignment (ALA)



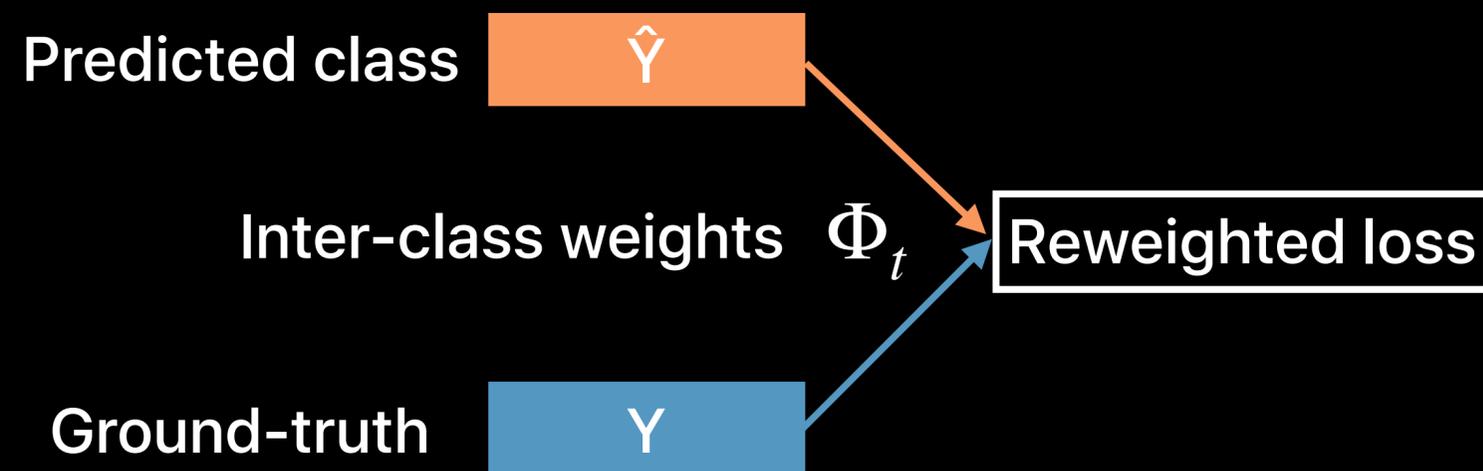
Adaptive Loss Alignment (ALA)



- **Action:** Update loss parameters at small steps
- **State:** Learning statistics on validation set + current iteration number
- **Reward:** Improvement of validation metric

Adaptive Loss Alignment (ALA)

Classification loss parameterization



Adaptive Loss Alignment (ALA)

Classification loss parameterization



Adaptive Loss Alignment (ALA)

Classification loss parameterization



Automobile



20 epoch

40 epoch

60 epoch

Truck



Ship



Dog



Cat



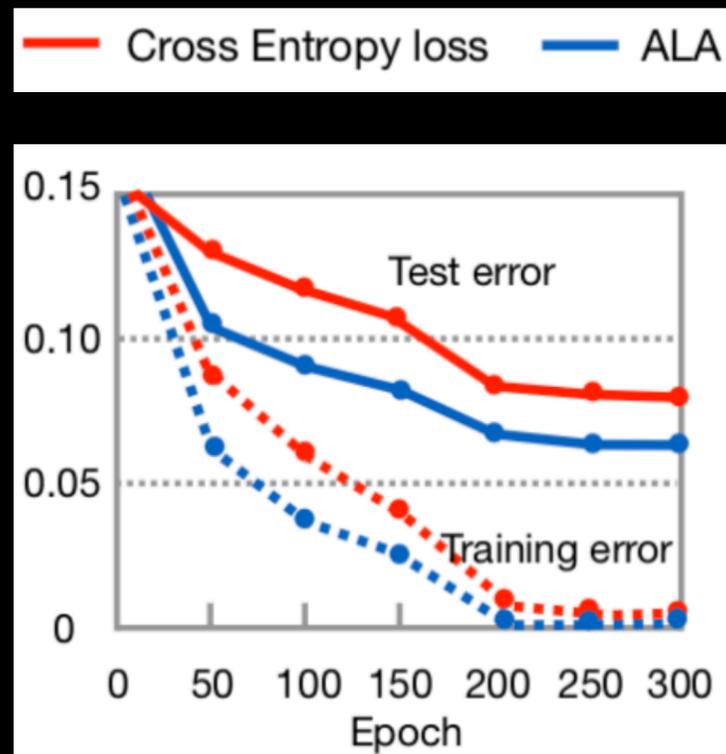
Hierarchical
classification
curriculum



Adaptive Loss Alignment (ALA) Results

ALA improves optimization + generalization

SOTA performance



Classification example on CIFAR-10

CIFAR10 Classification*		Metric learning (Products retrieval)		
Method	Error	Method	Recall@1	Recall@10
CrossEntropy	7.51%	Triplet (2015)	66.7%	82.4%
L2T (2018) [1]	7.10%	Margin (2017)	72.7%	86.2%
L2T-DLF (2018) [2]	6.95%	ABE-8 (2018)	76.3%	88.4%
ALA (1-Net)	6.85%	Triplet+ALA	75.7%	89.4%
ALA (10-nets)	6.79%	Margin+ALA	78.9%	90.7%

* Using ResNet-32 network

[1] Fan et al. "Learning to Teach", ICLR'18

[2] Wu et al. "Learning to Teach with Dynamic Loss Functions", NeurIPS'18



Poster discussion

6:30-9:00pm @ Pacific Ballroom #23

