Anchor Loss

Modulating Loss Scale based on Prediction Difficulty

Serim Ryou Seong-Gyun Jeong Pietro Perona

ICCV 2019 Seoul, Korea

Caltech

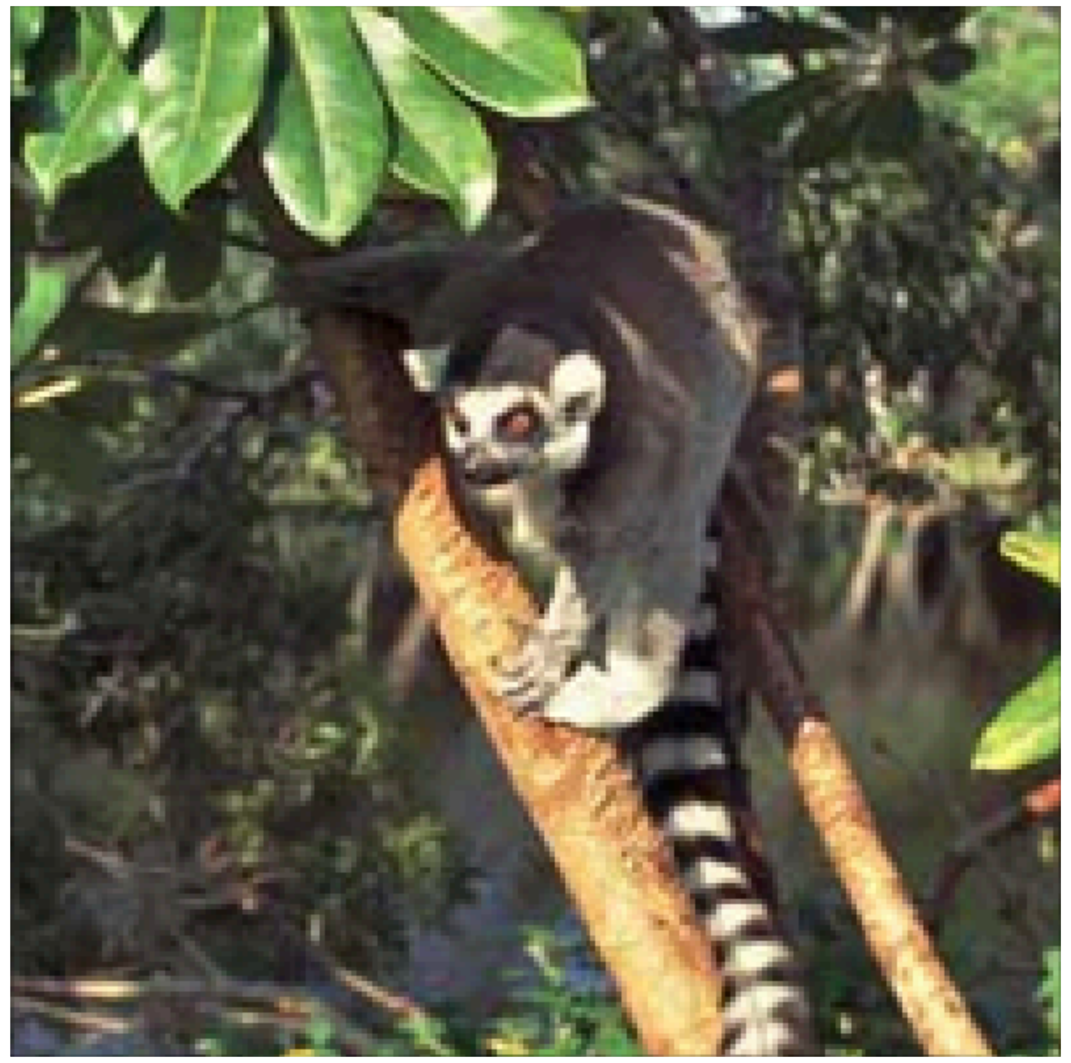
CODE42

What’s in this image?

Too easy :)

I could...

Whaaat??



container ship

life boat

pickup

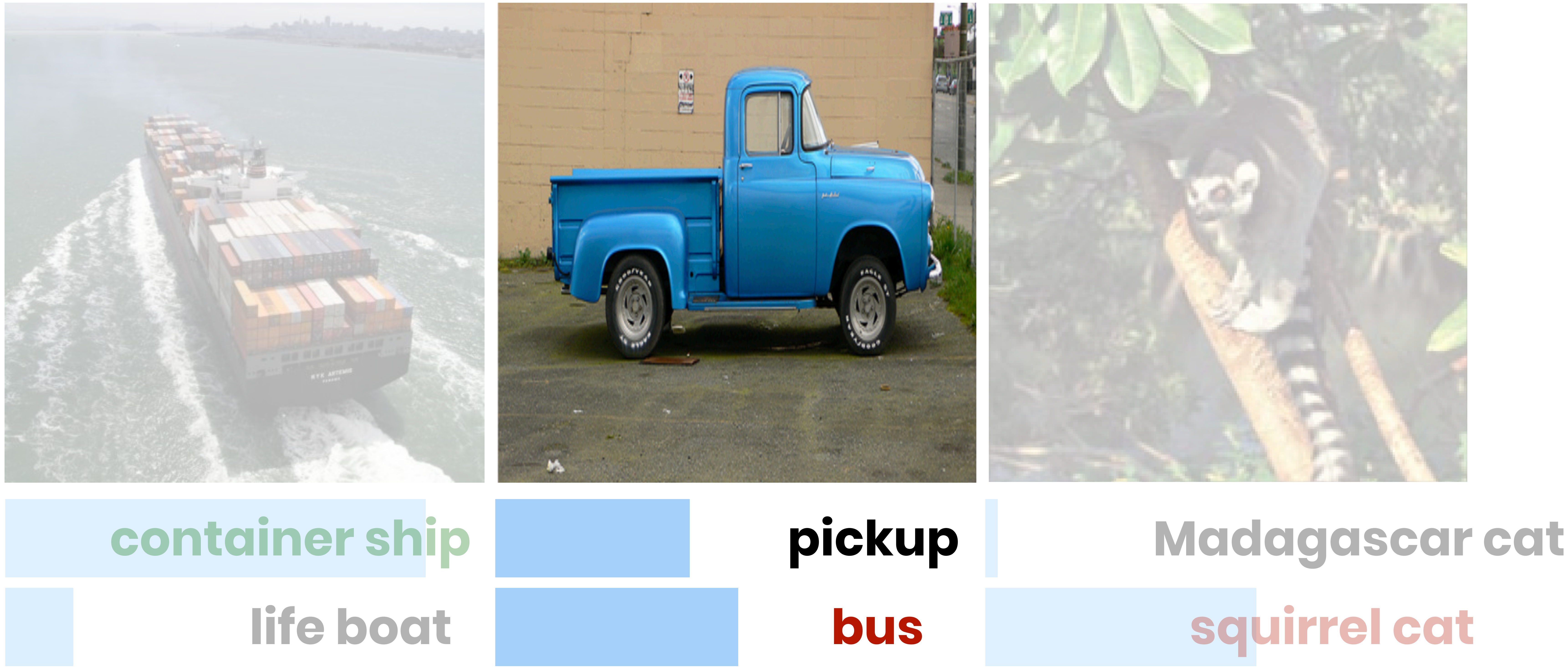
bus

Madagascar cat

squirrel cat

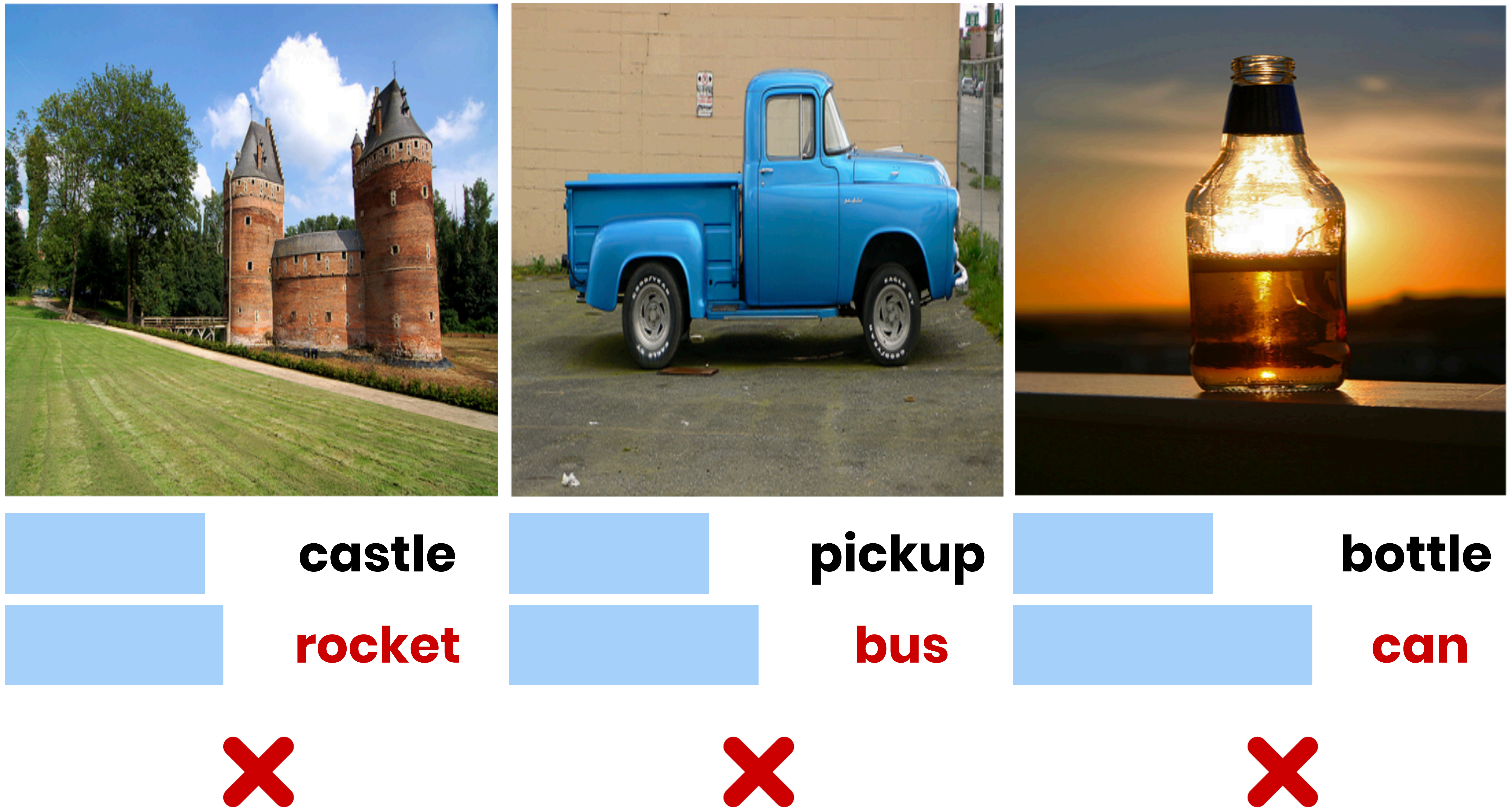
Motivation

Network often produces **bi-modal prediction** with the presence of visually confusing cases



Motivation

Network often produces **bi-modal prediction** with the presence of visually confusing cases



Goal

Give the network good learning signal by leveraging ***prediction difficulty***

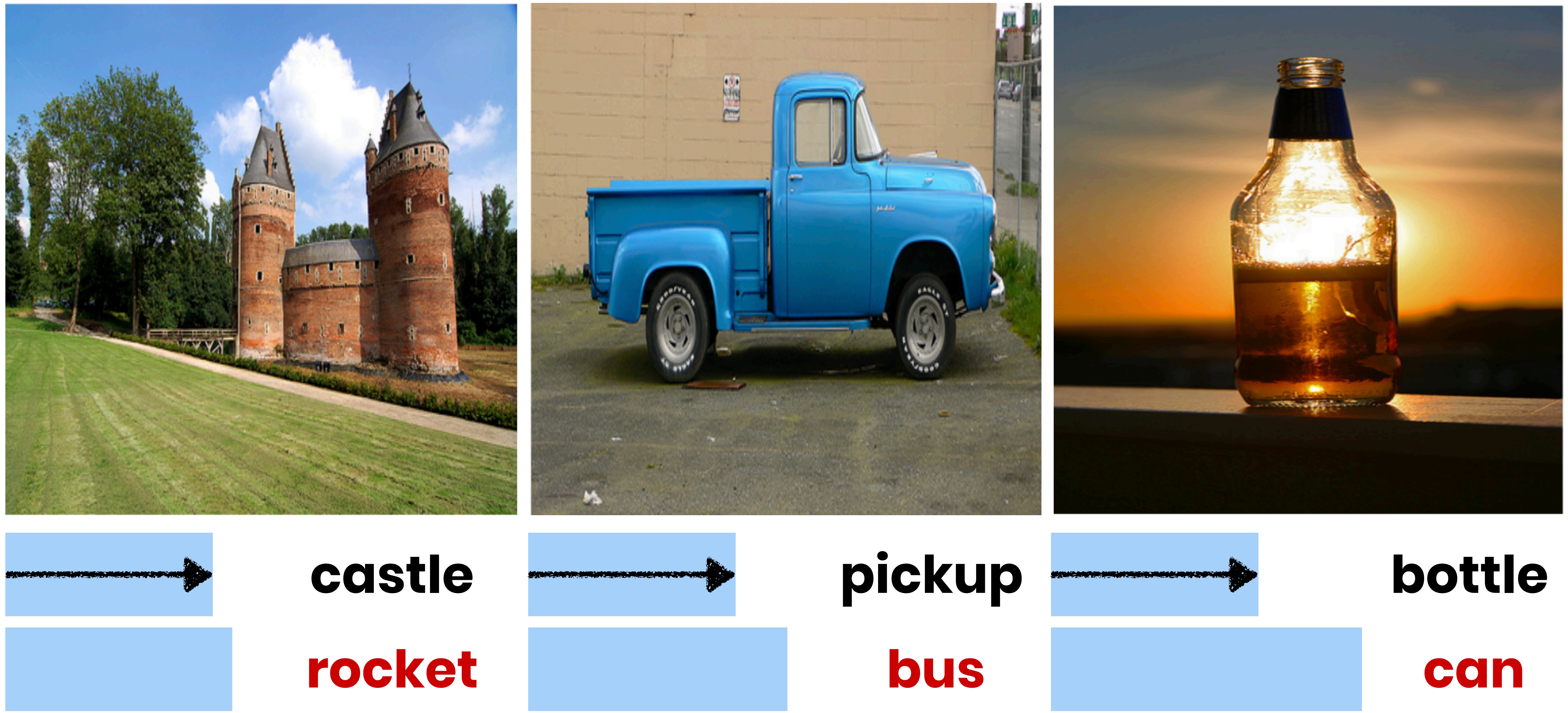
castle
rocket

pickup
bus

bottle
can

Goal

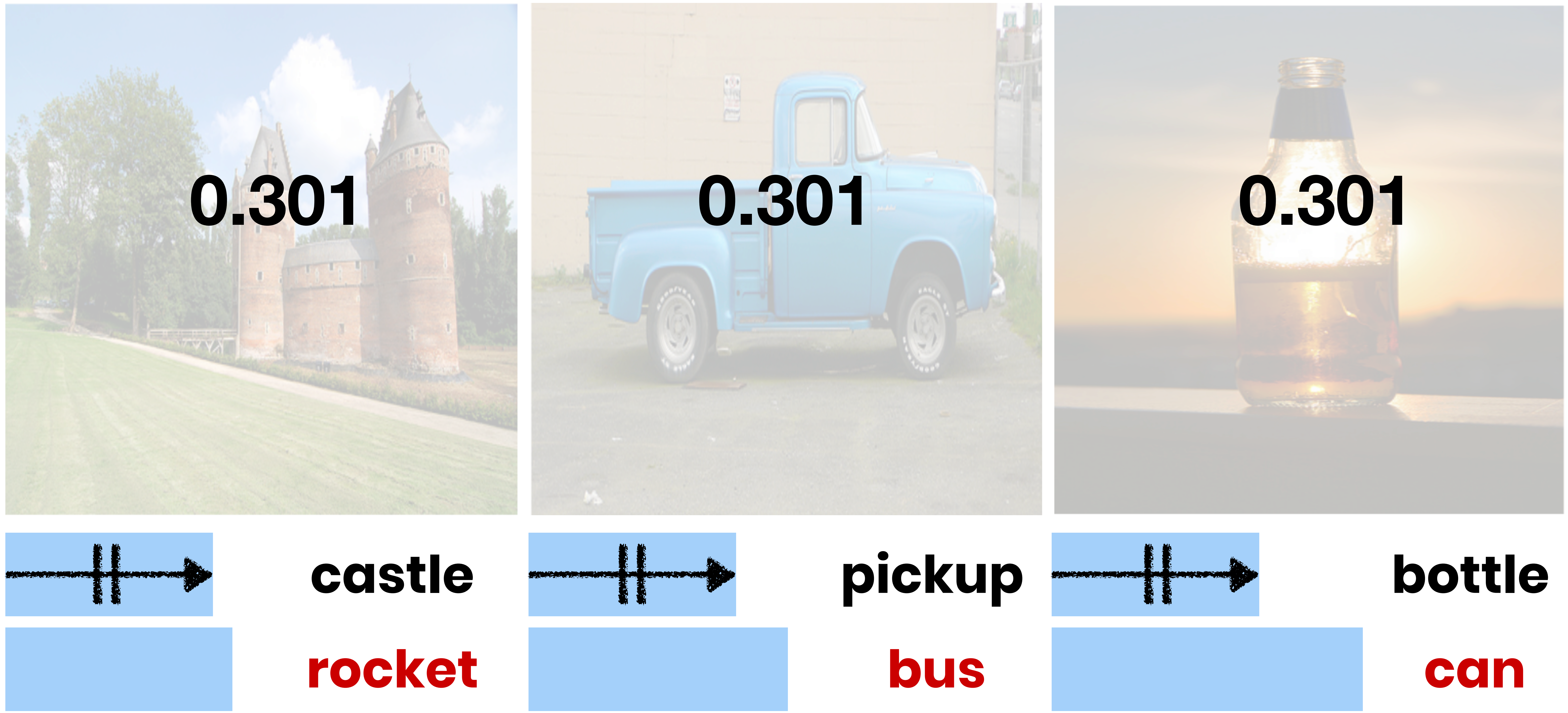
Apply different loss values based on the divergence of output predictions



Goal

Apply different loss values based on the divergence of output predictions

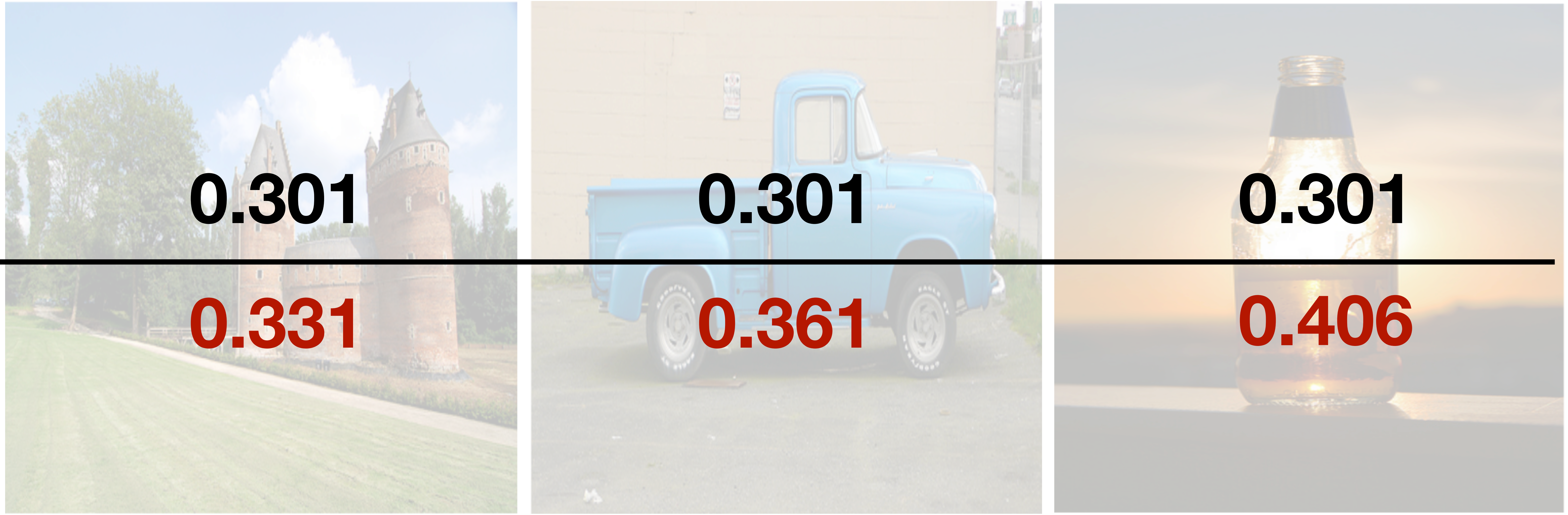
CE

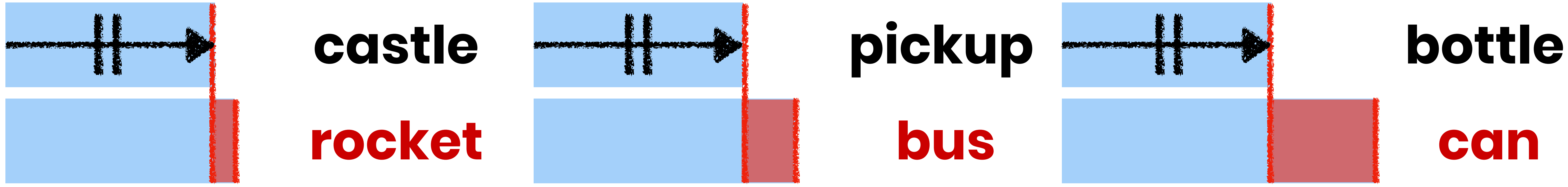


Goal

Apply different loss values based on the divergence of output predictions

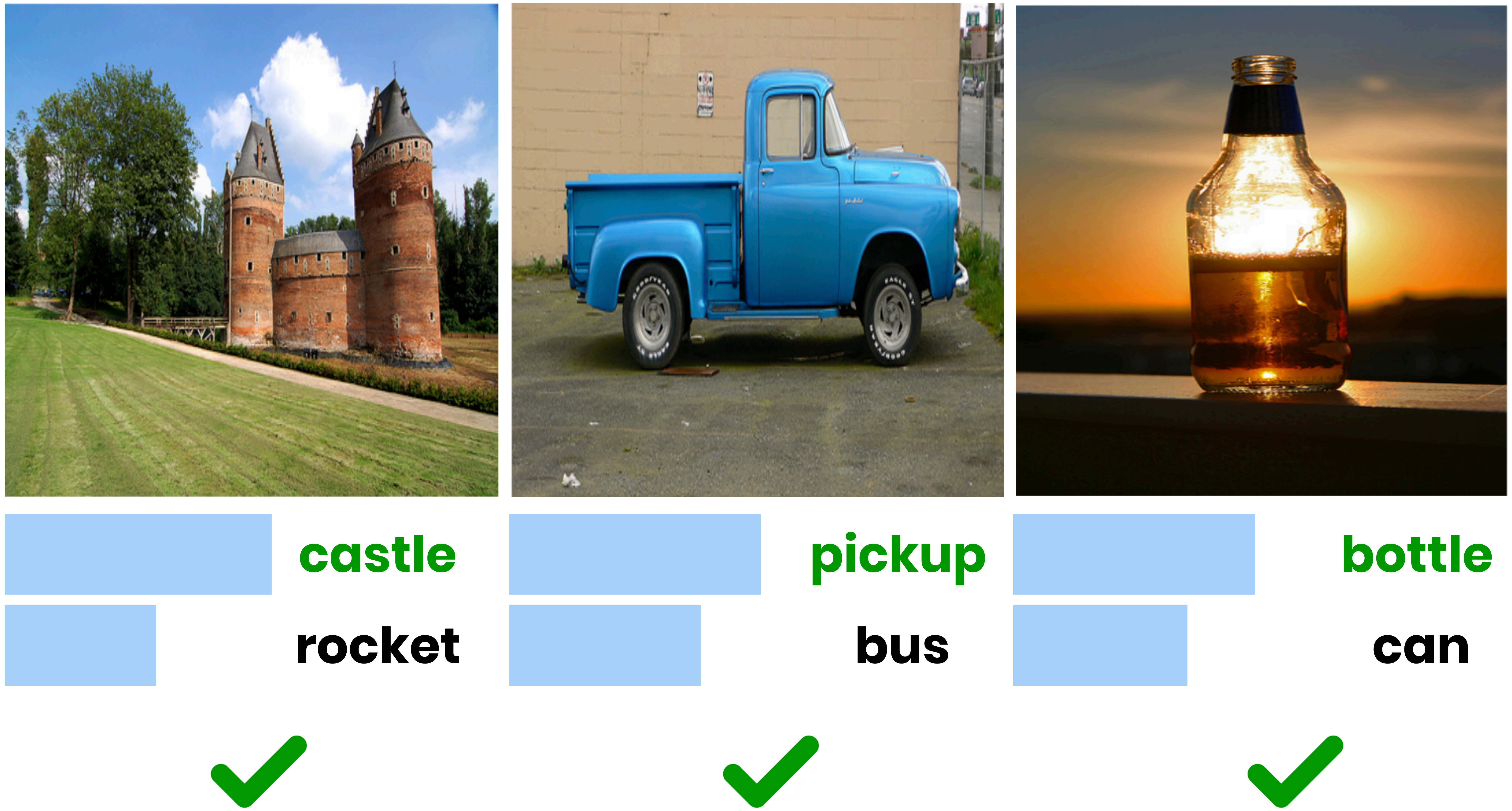
CE	0.301	0.301	0.301
Our loss	0.331	0.361	0.406





Goal

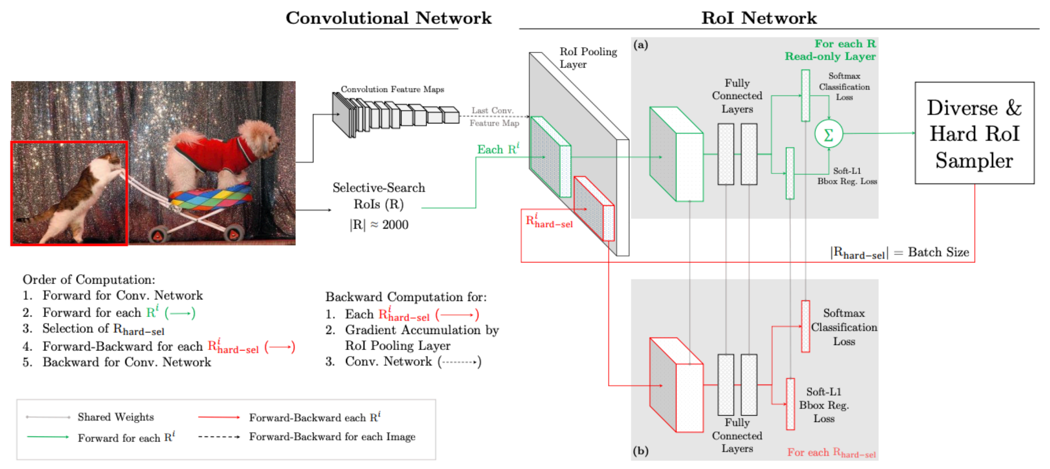
Apply different loss values based on the divergence of output predictions



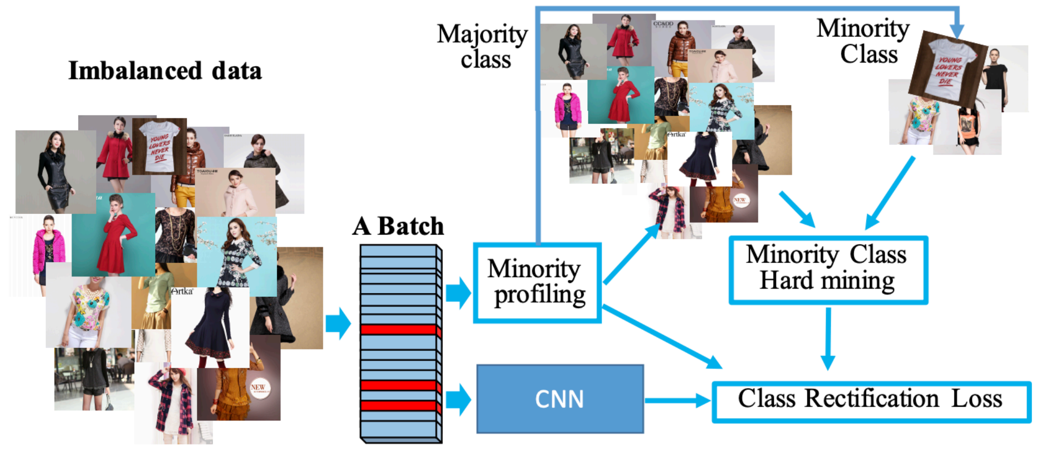
Related work

1. Hard example mining

- Online hard example mining (OHEM) [Shrivastava *et al.*]



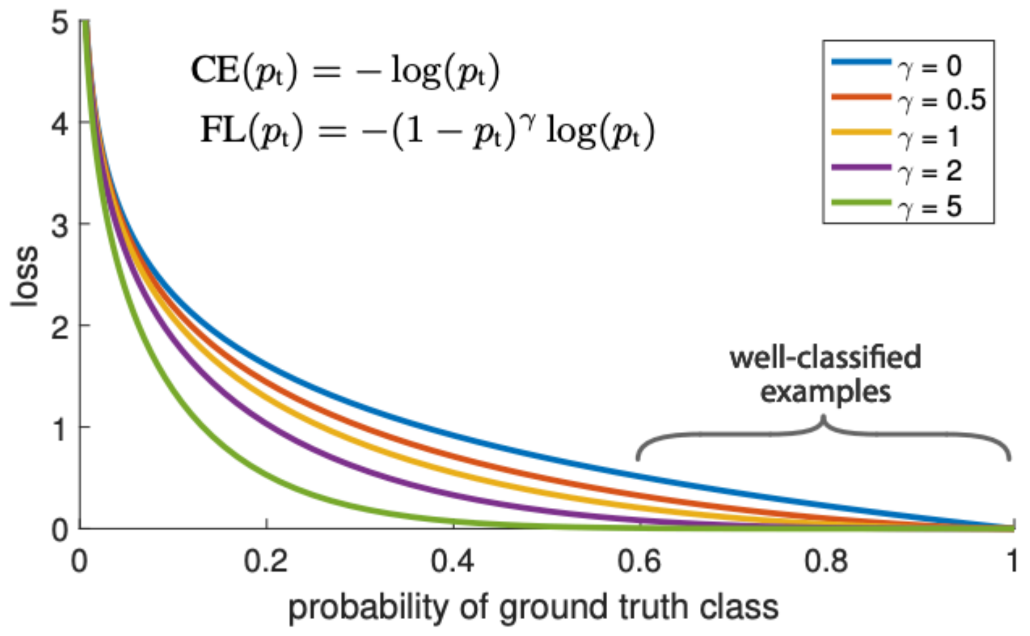
- Hard sample mining of minority attributes [Dong *et al.*]



- Online hard keypoint mining [Chen *et al.*]

2. Reweighing scheme

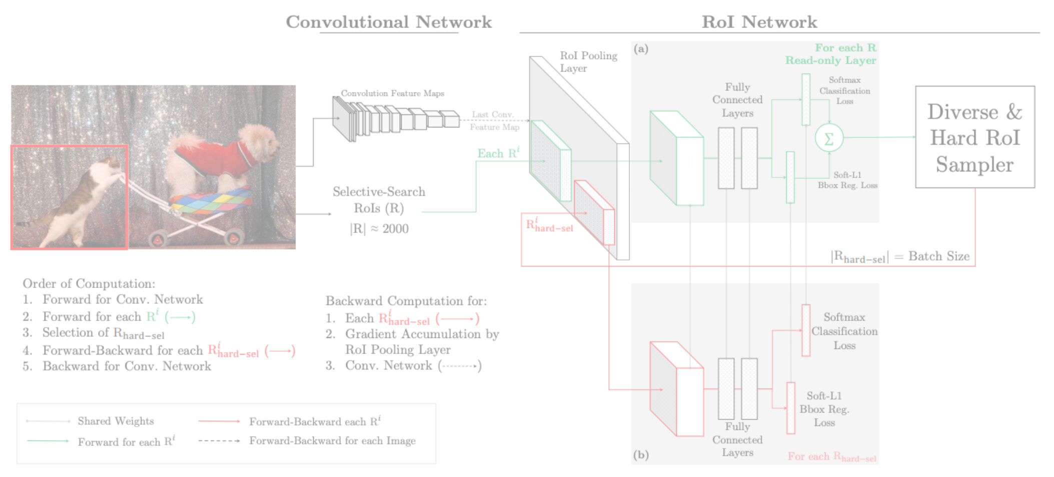
- Focal loss [Lin *et al.*]



Related work

1. Hard example mining

- Online hard example mining (OHEM) [Shrivastava *et al.*]

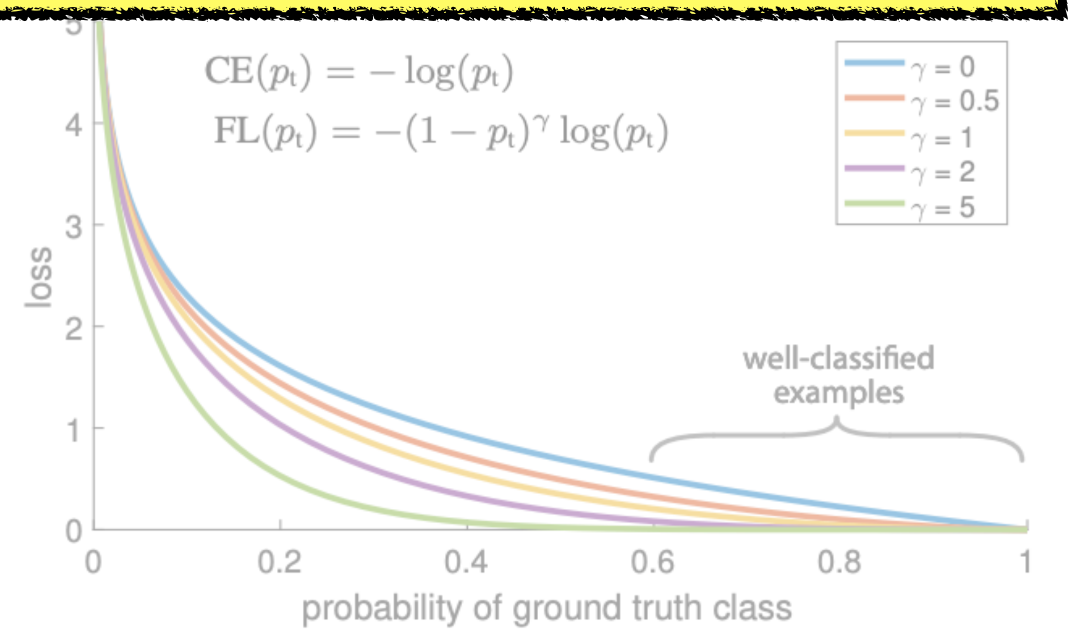


However, needs data statistics* or sampling heuristics

**e.g.,* class distribution, foreground vs background

2. Reweighing scheme

- Focal loss [Lin *et al.*]



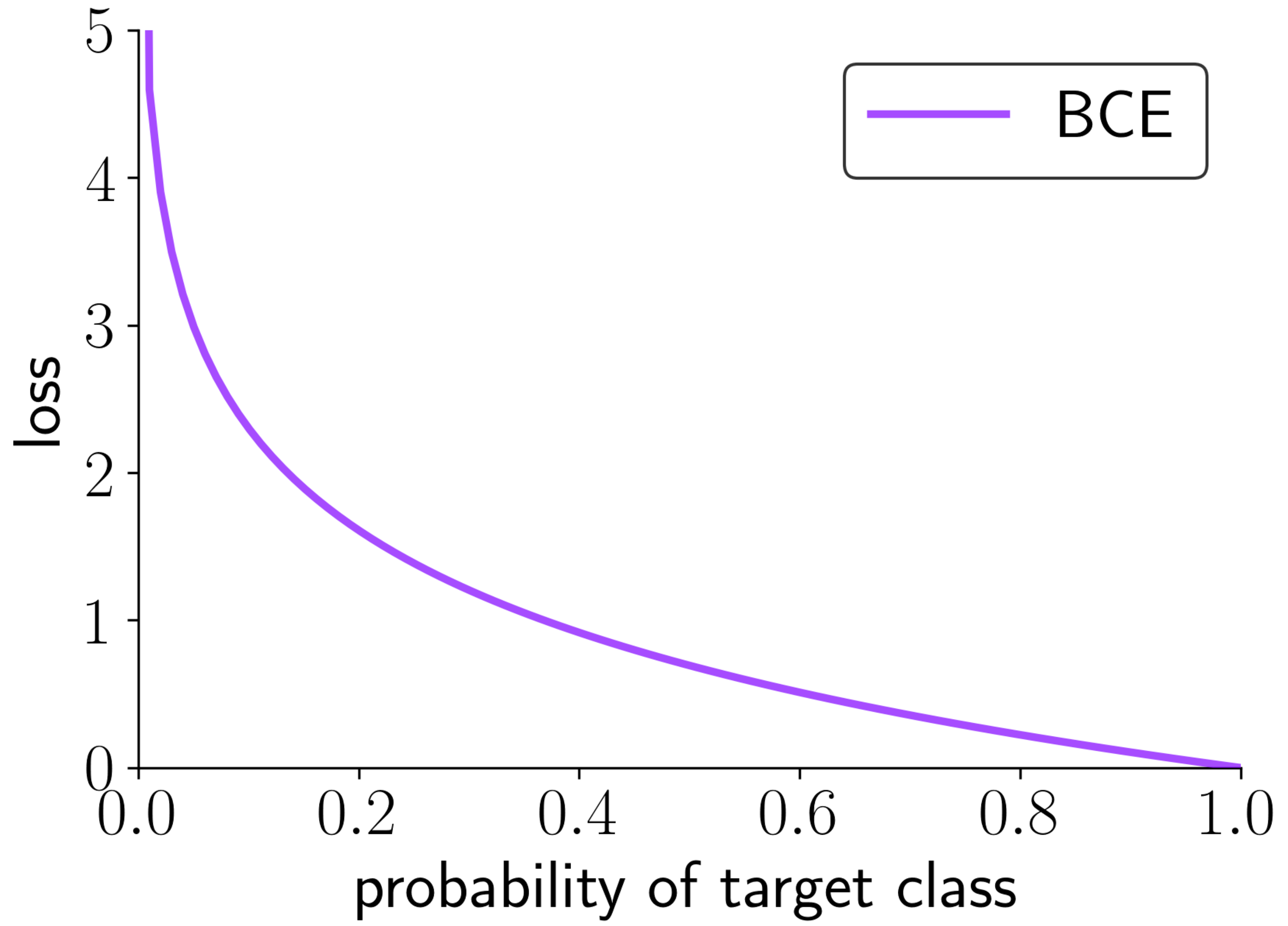
Our approach

Network decides the **prediction difficulty** by itself

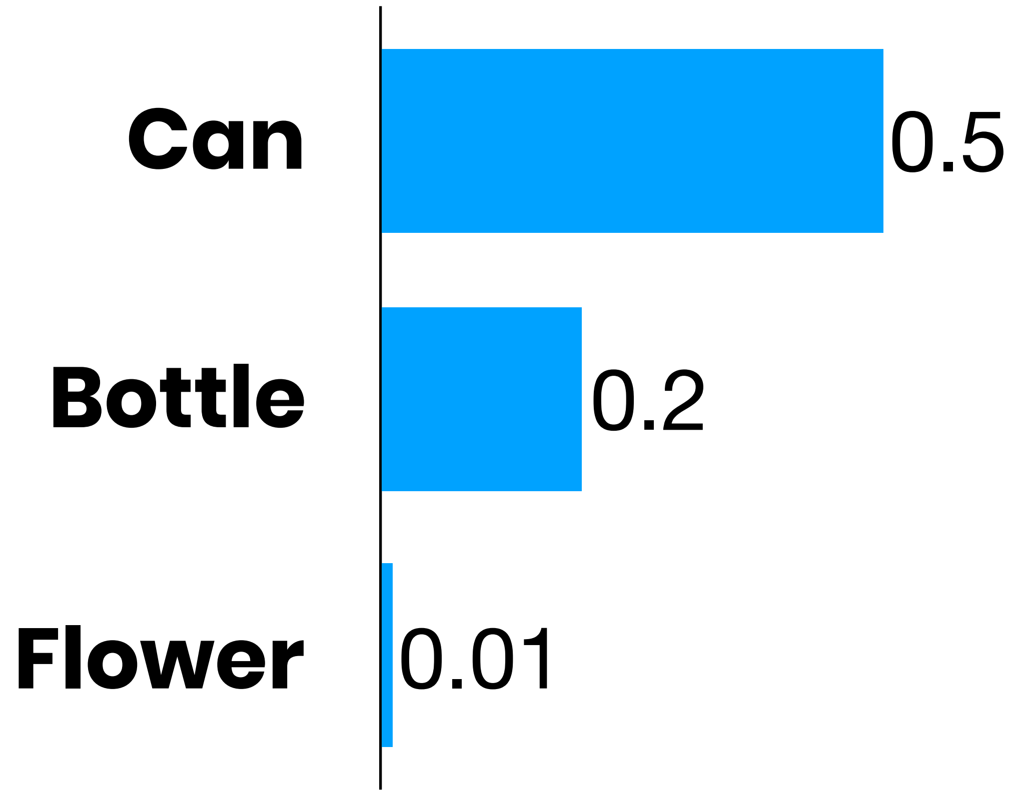
Input image

Bottle

Loss function



Output distribution



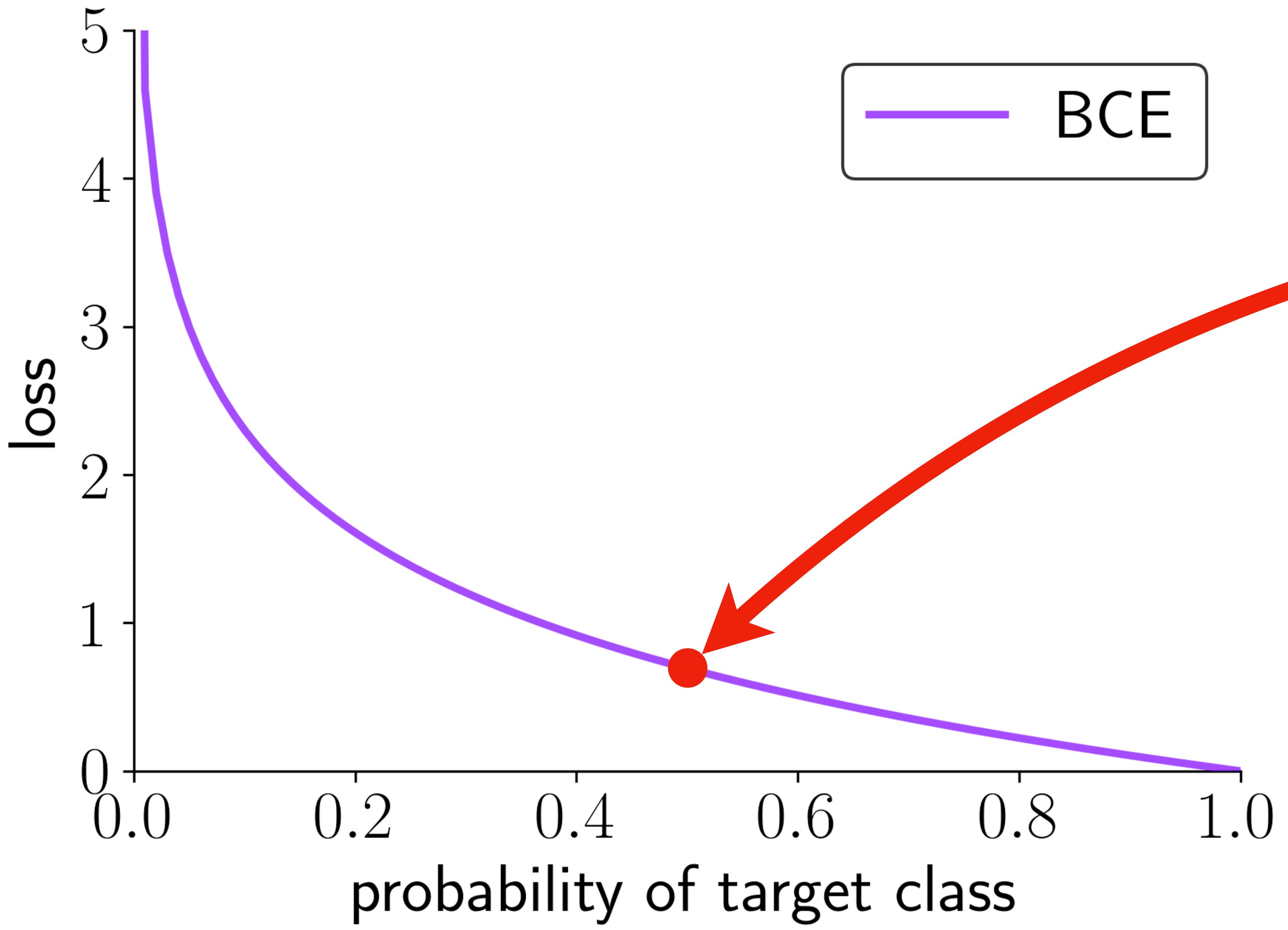
Our approach

Anchor probability is set to the highest prediction score among bg classes

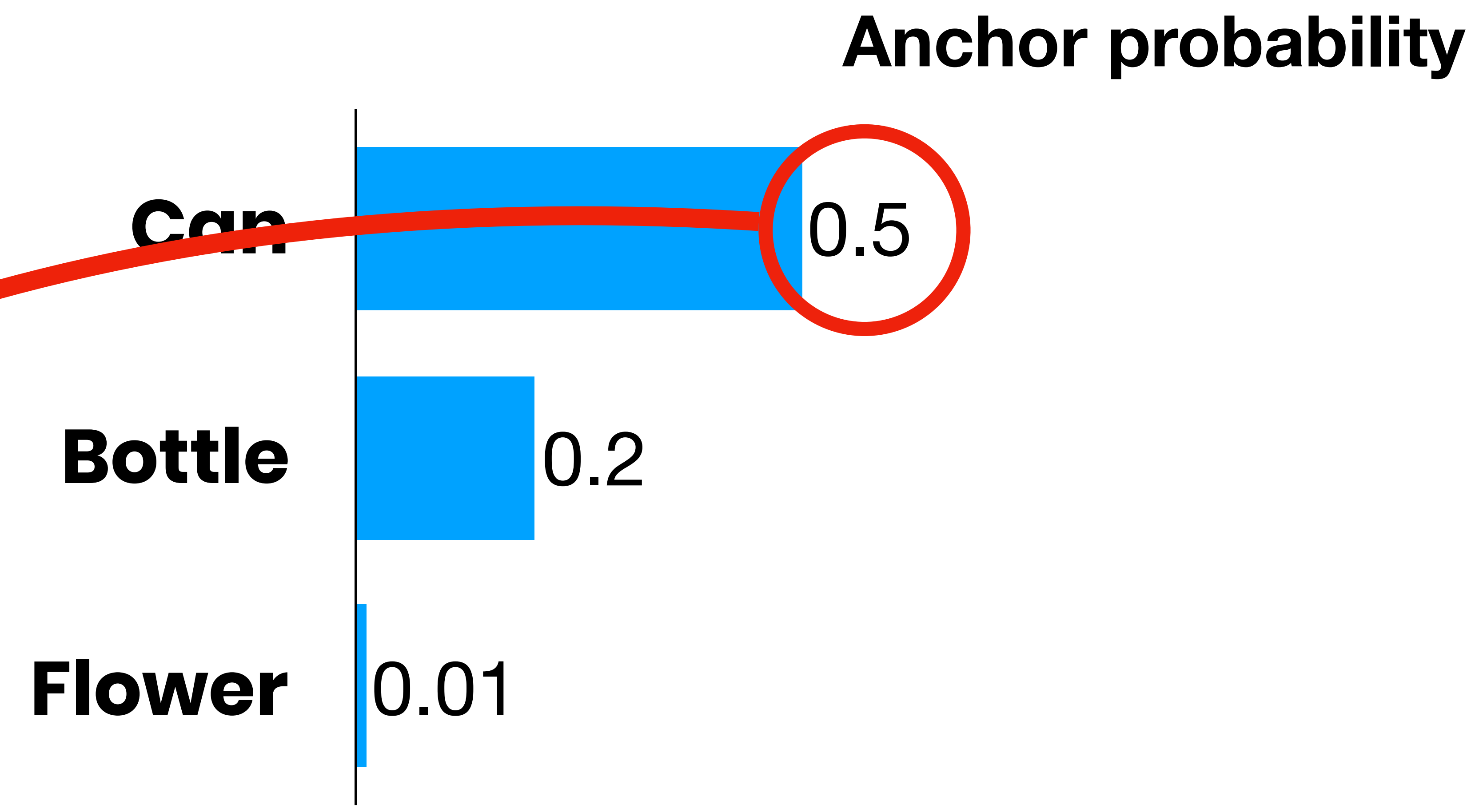
Input image

Bottle

Loss function



Output distribution



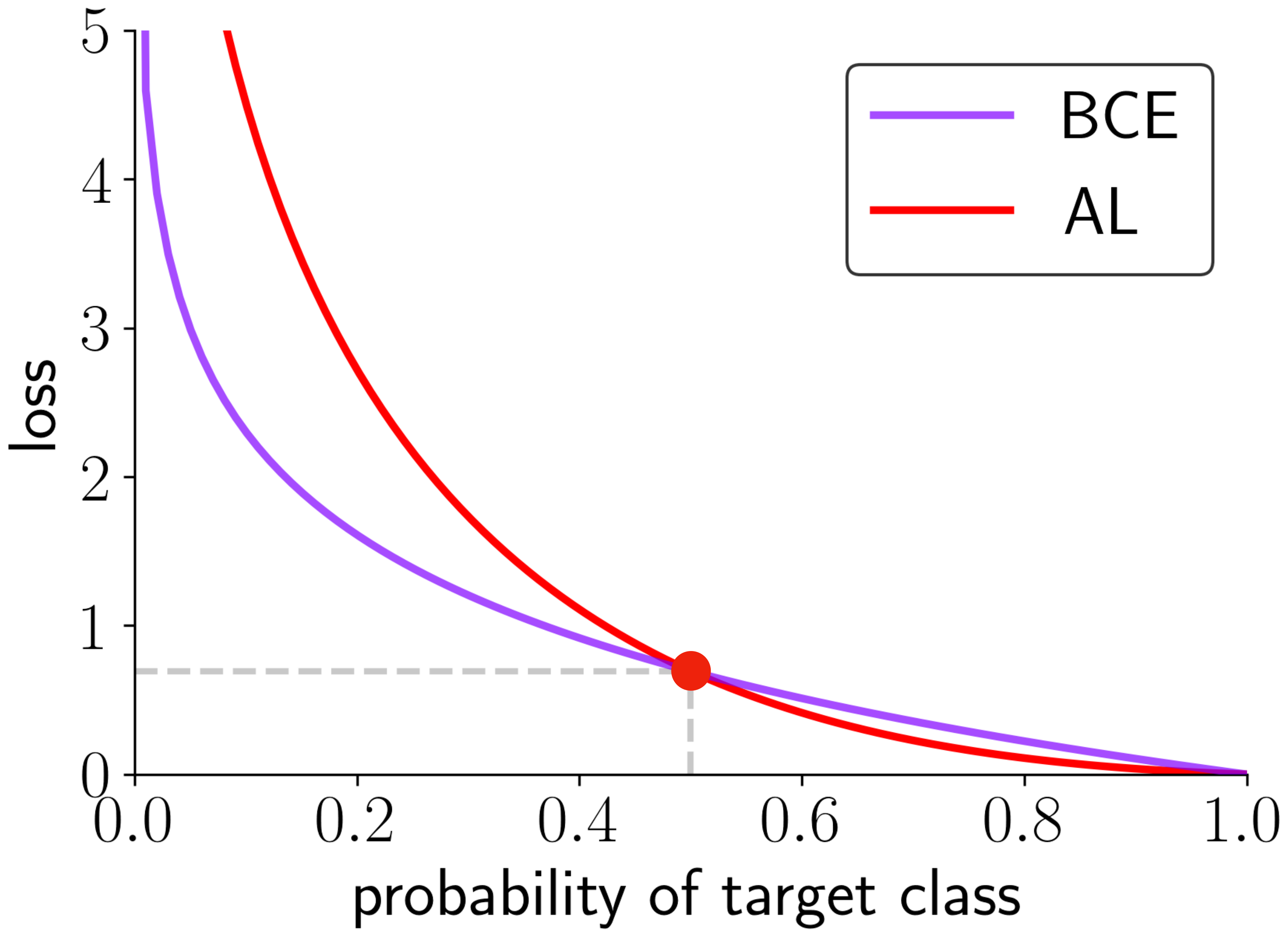
Our approach

Anchor loss (AL) modulates the loss function on ground truth probability

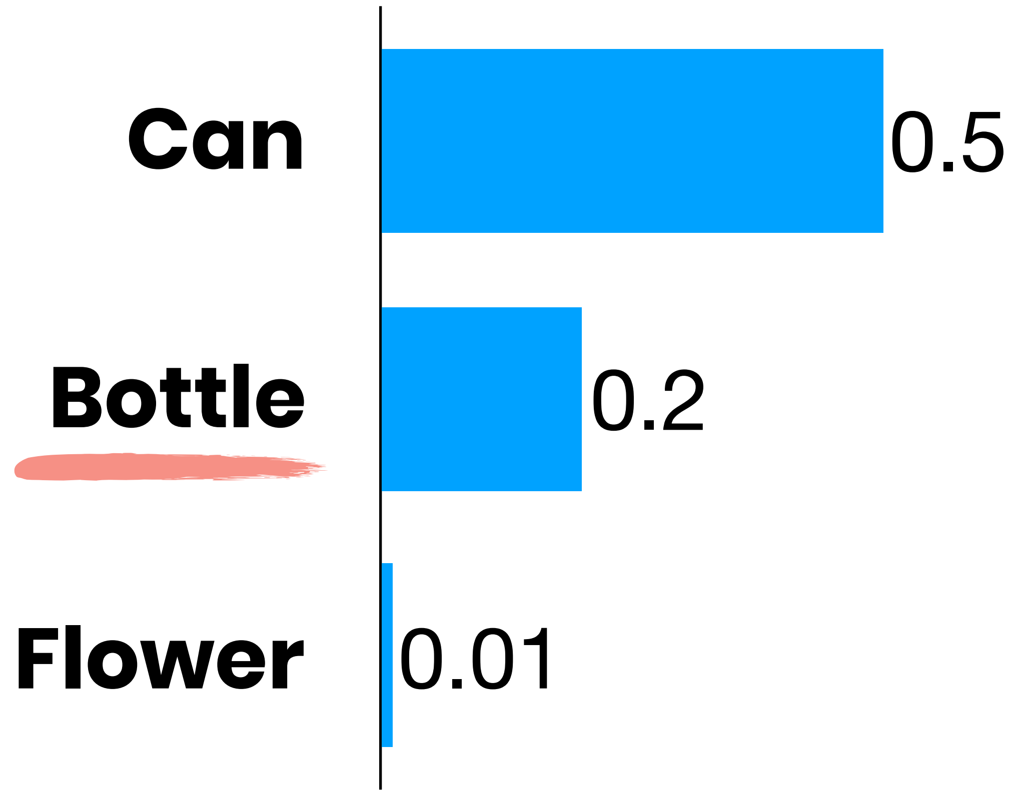
Input image

Bottle

Loss function



Output distribution



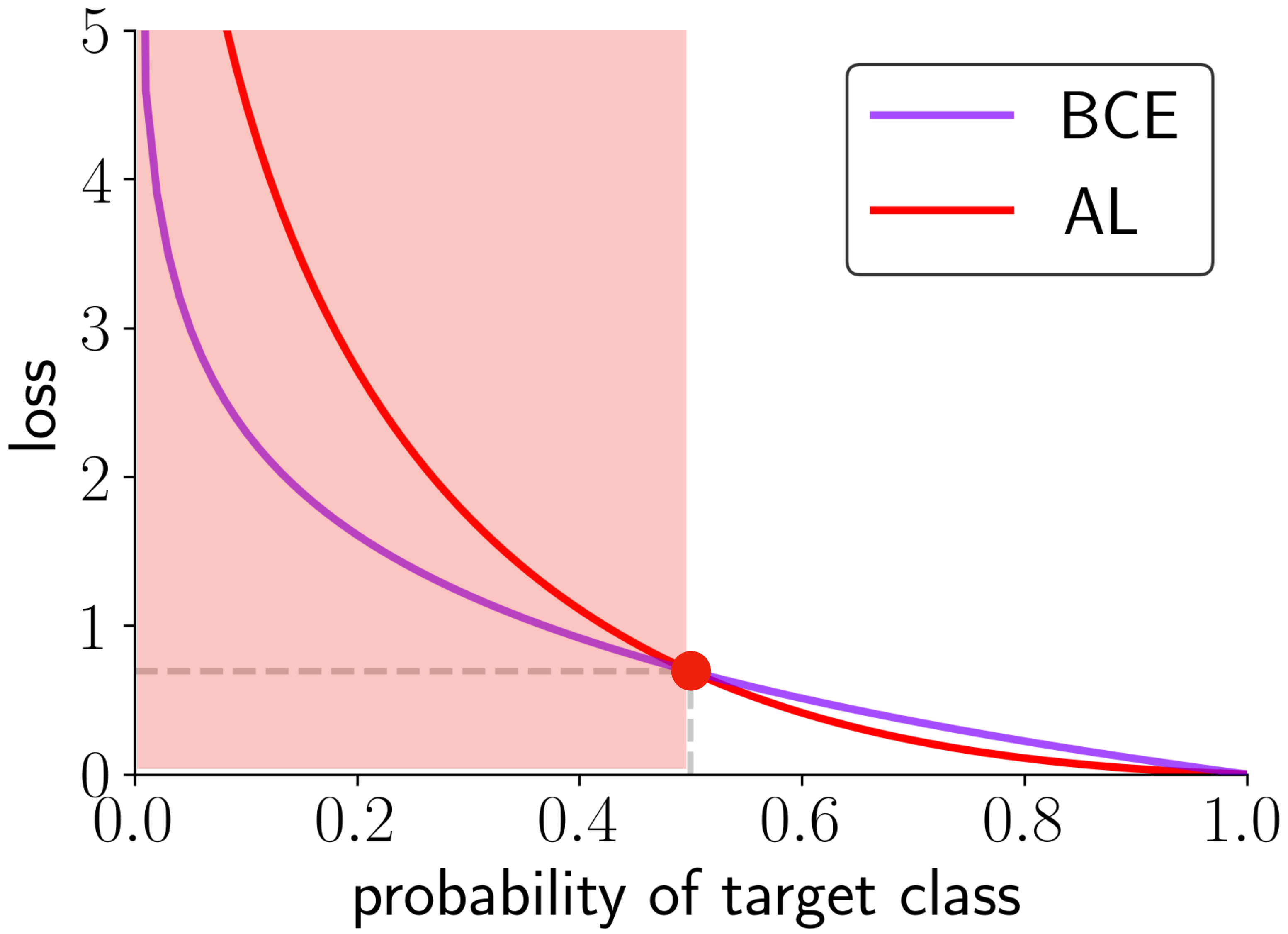
Our approach

Anchor loss (AL) penalizes more than the cross entropy when..

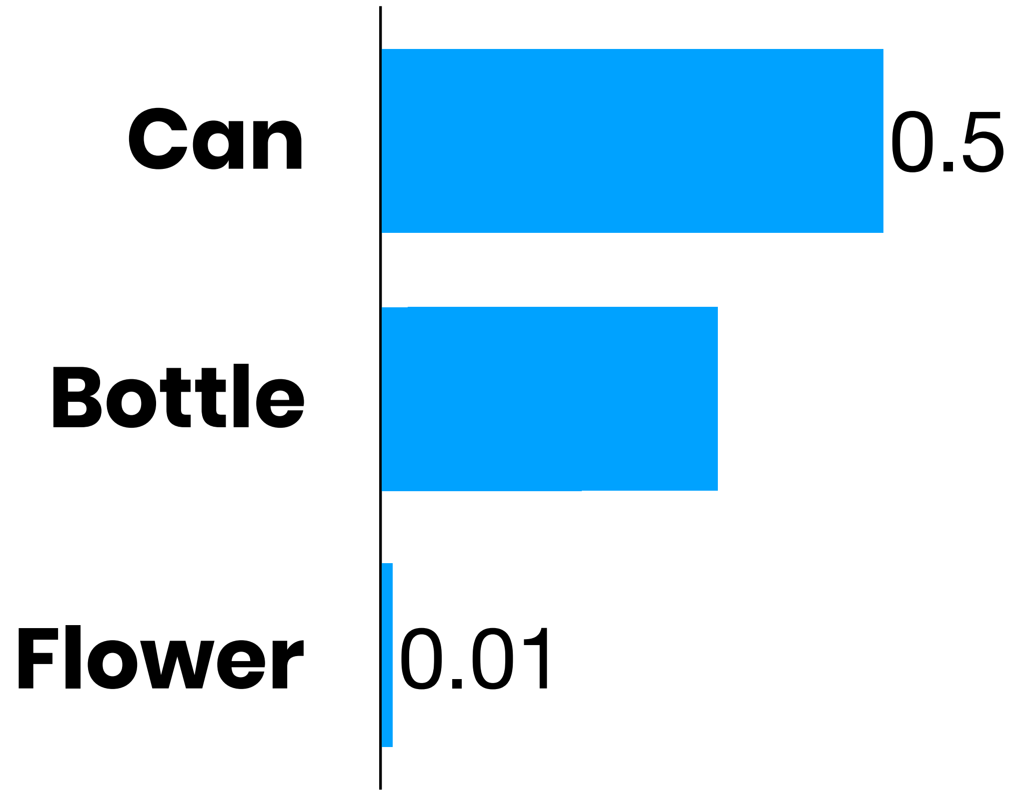
Input image

Bottle

Loss function



Output distribution



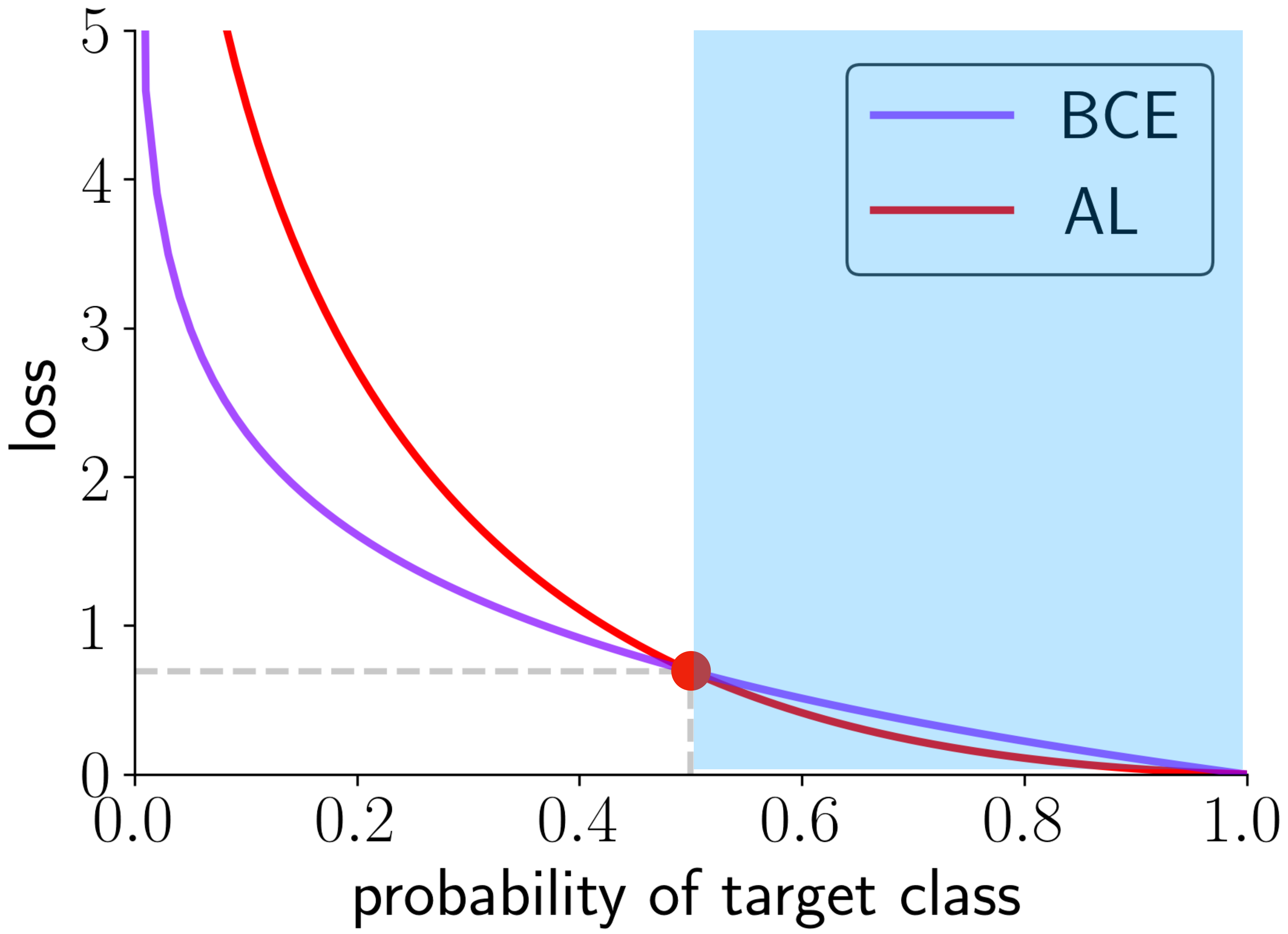
Our approach

Anchor loss (AL) penalizes less than the cross entropy when..

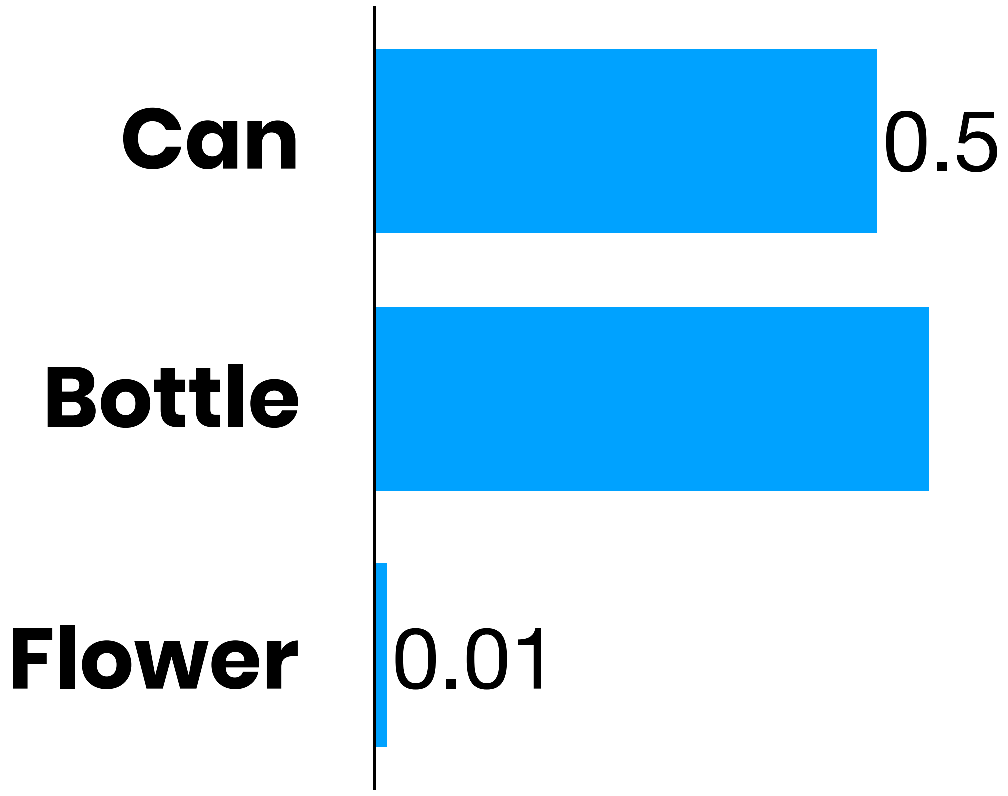
Input image

Bottle

Loss function



Output distribution



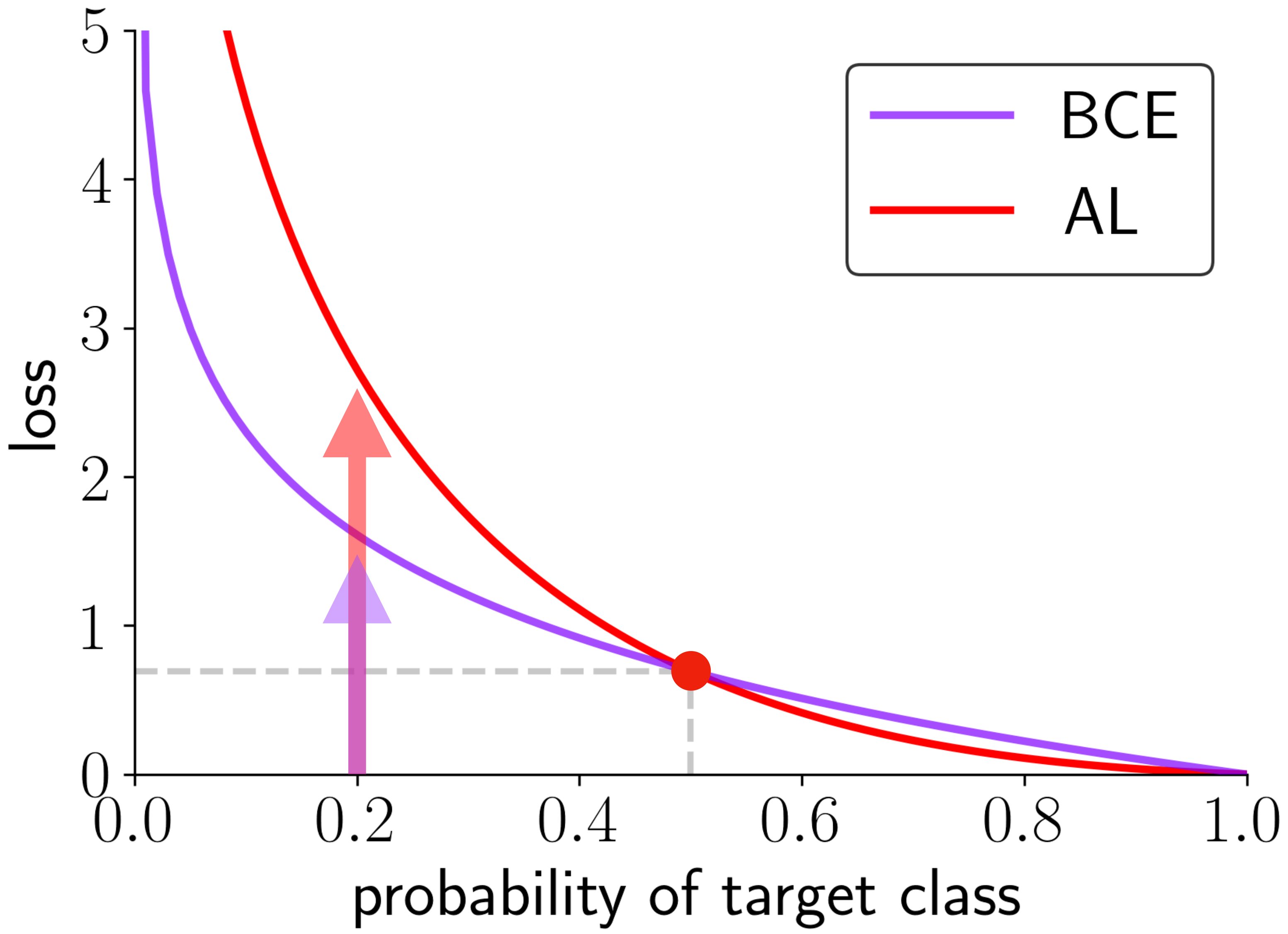
Our approach

Anchor loss (AL) gives higher penalty on ground truth probability

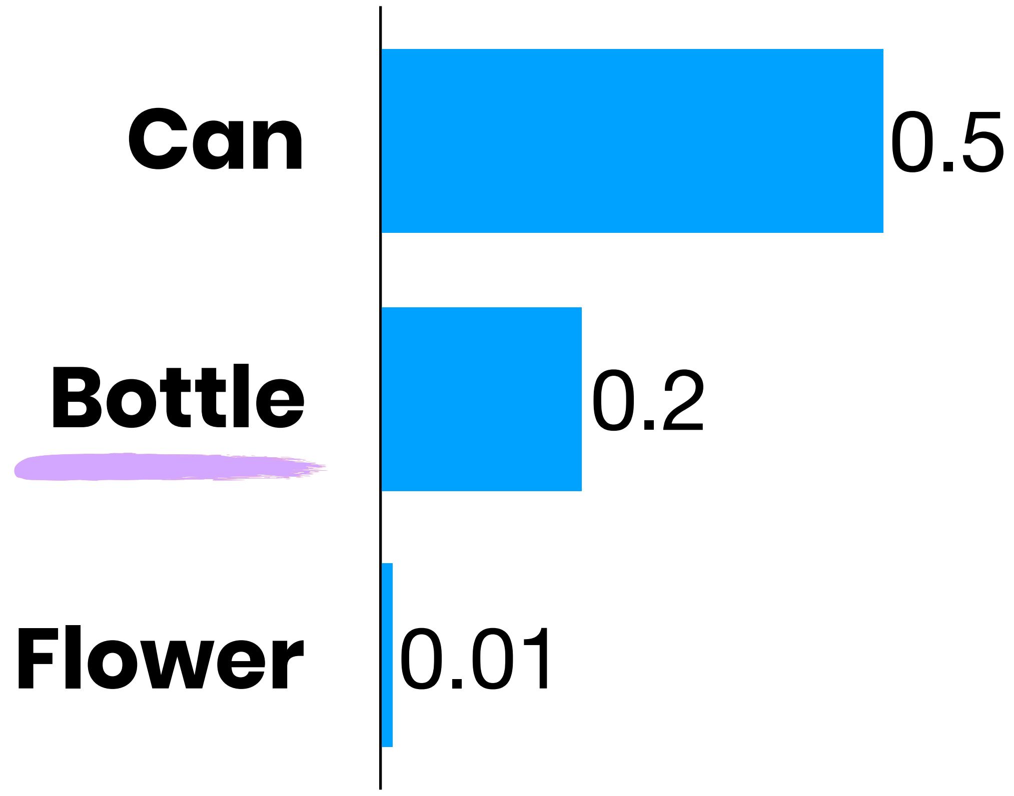
Input image

Bottle

Loss function



Output distribution



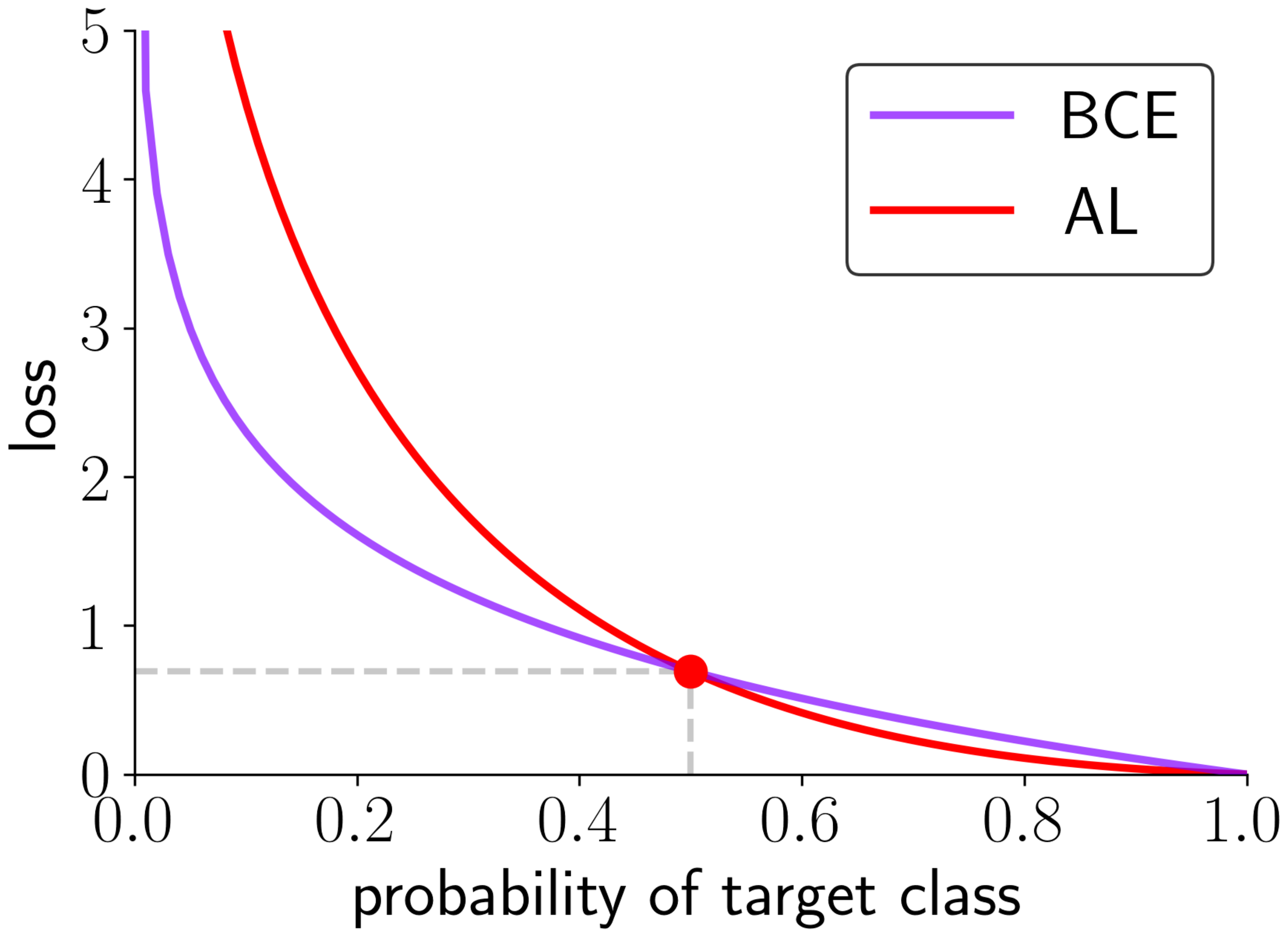
Our approach

Finally, network produces correct predictions

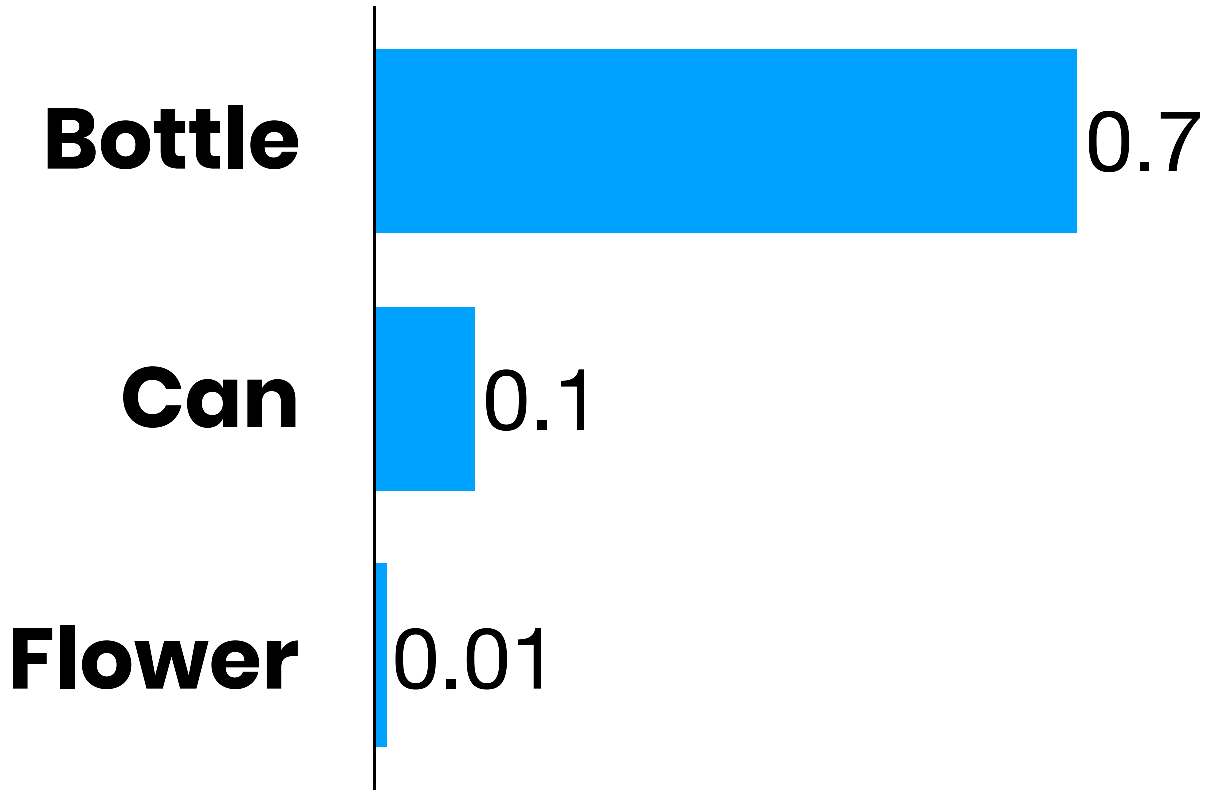
Input image

Bottle

Loss function



Output distribution



Anchor loss

Modulates cross entropy loss function using prediction difficulty

$$\ell_{AL}(q_t; \gamma, q_*) = -\underbrace{(1 - (q_t - q_*))^{\gamma}}_{\text{Modulator}} \underbrace{\log(q_t)}_{\text{Cross entropy}}$$

Anchor loss

Anchor probability leads the network to regulate the prediction difficulties

Prediction difficulty

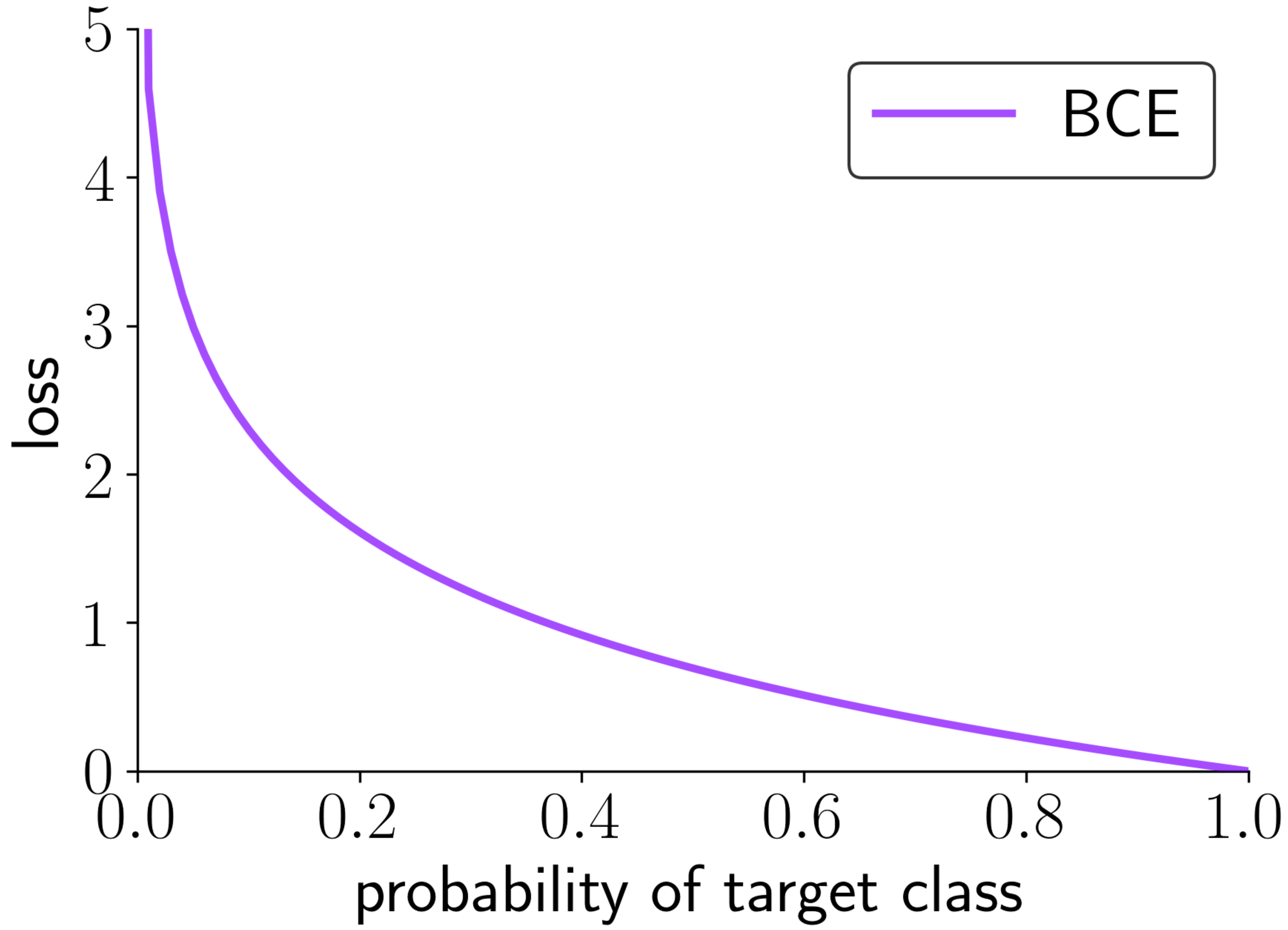
$$\ell_{AL}(q_t;\gamma,q_*) = \underbrace{-(1-\overbrace{(q_t - q_*)}^{\text{Prediction difficulty}})^{\gamma}}_{\text{Modulator}}\underbrace{\log(q_t)}_{\text{Cross entropy}}$$

Anchor probability

Modulator

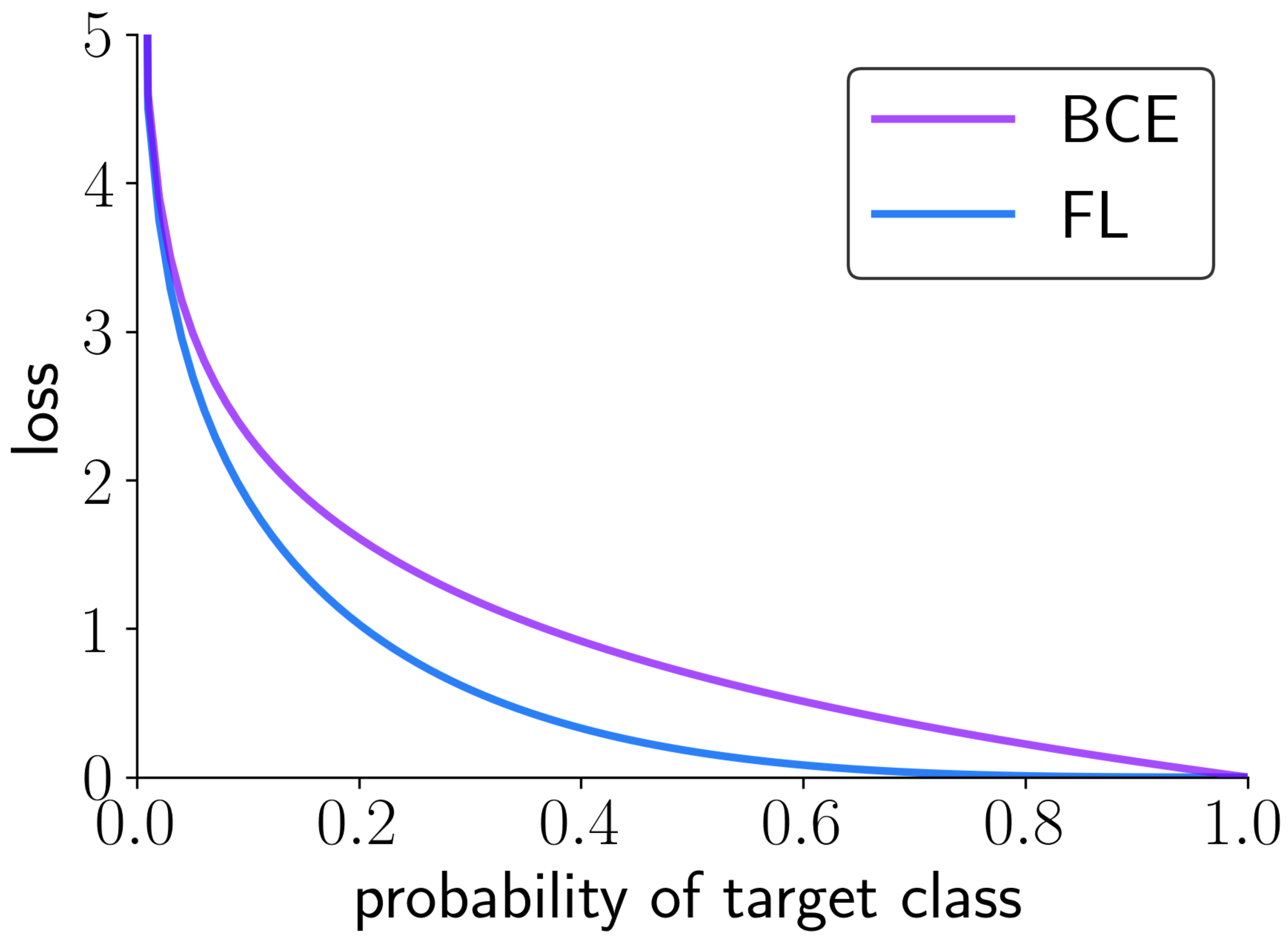
Cross entropy

Comparison to other loss functions



$$\ell_{CE}(q_t) = -\log(q_t)$$

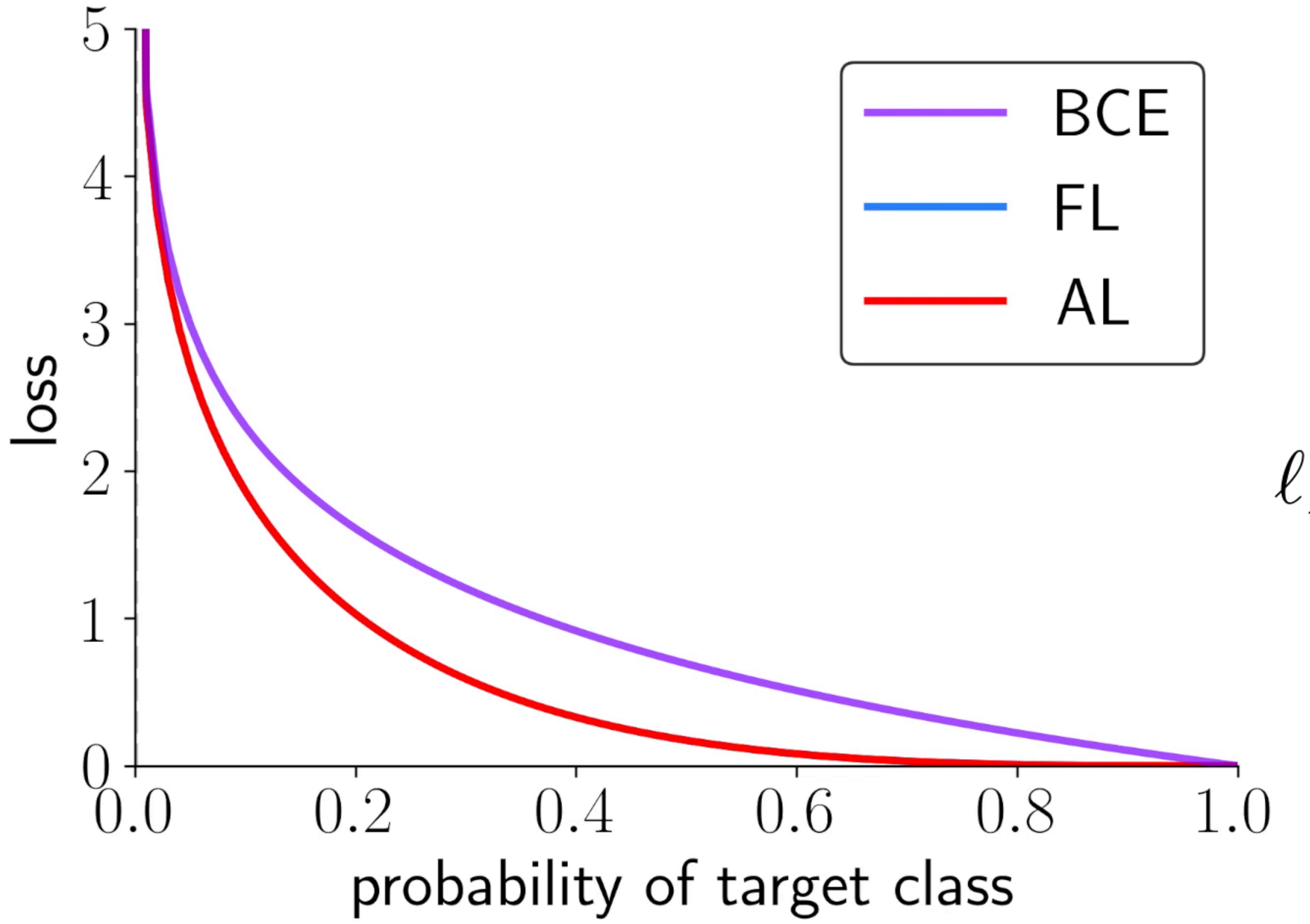
Comparison to other loss functions



$$\ell_{CE}(q_t) = -\log(q_t)$$

$$\ell_{FL}(q_t;\gamma) = -(1-q_t)^{\gamma}\log(q_t)$$

Comparison to other loss functions



$$\ell_{CE}(q_t) = -\log(q_t)$$

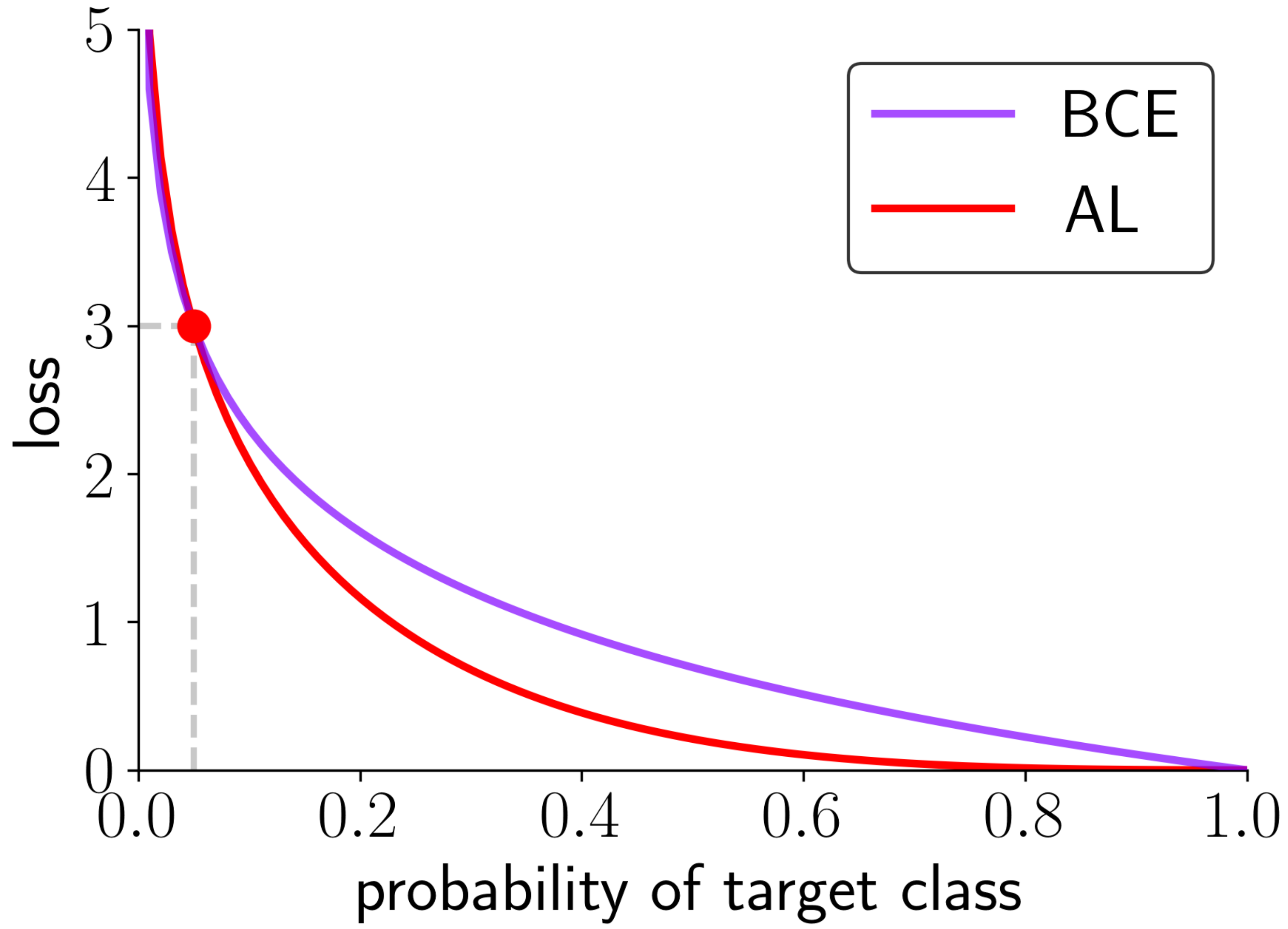
$$\ell_{FL}(q_t;\gamma) = -(1-q_t)^{\gamma}\log(q_t)$$

$$\ell_{AL}(q_t;\gamma,q_*) = -(1-(q_t-q_*))^{\gamma}\log(q_t)$$

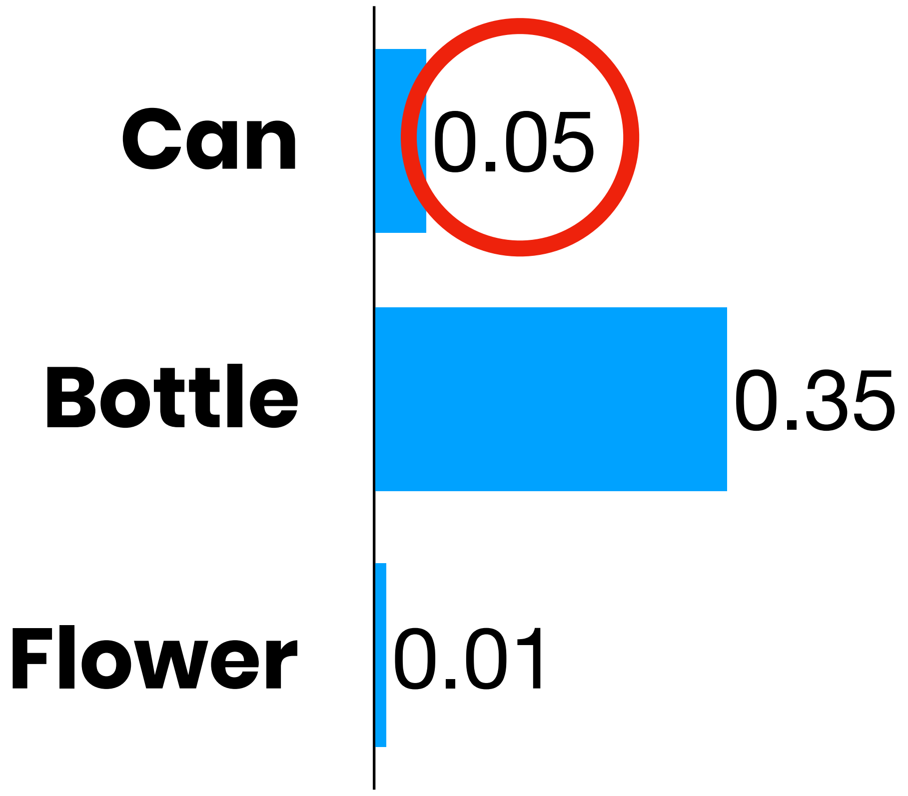
Easy case

Input image
Bottle

Loss function



Output distribution

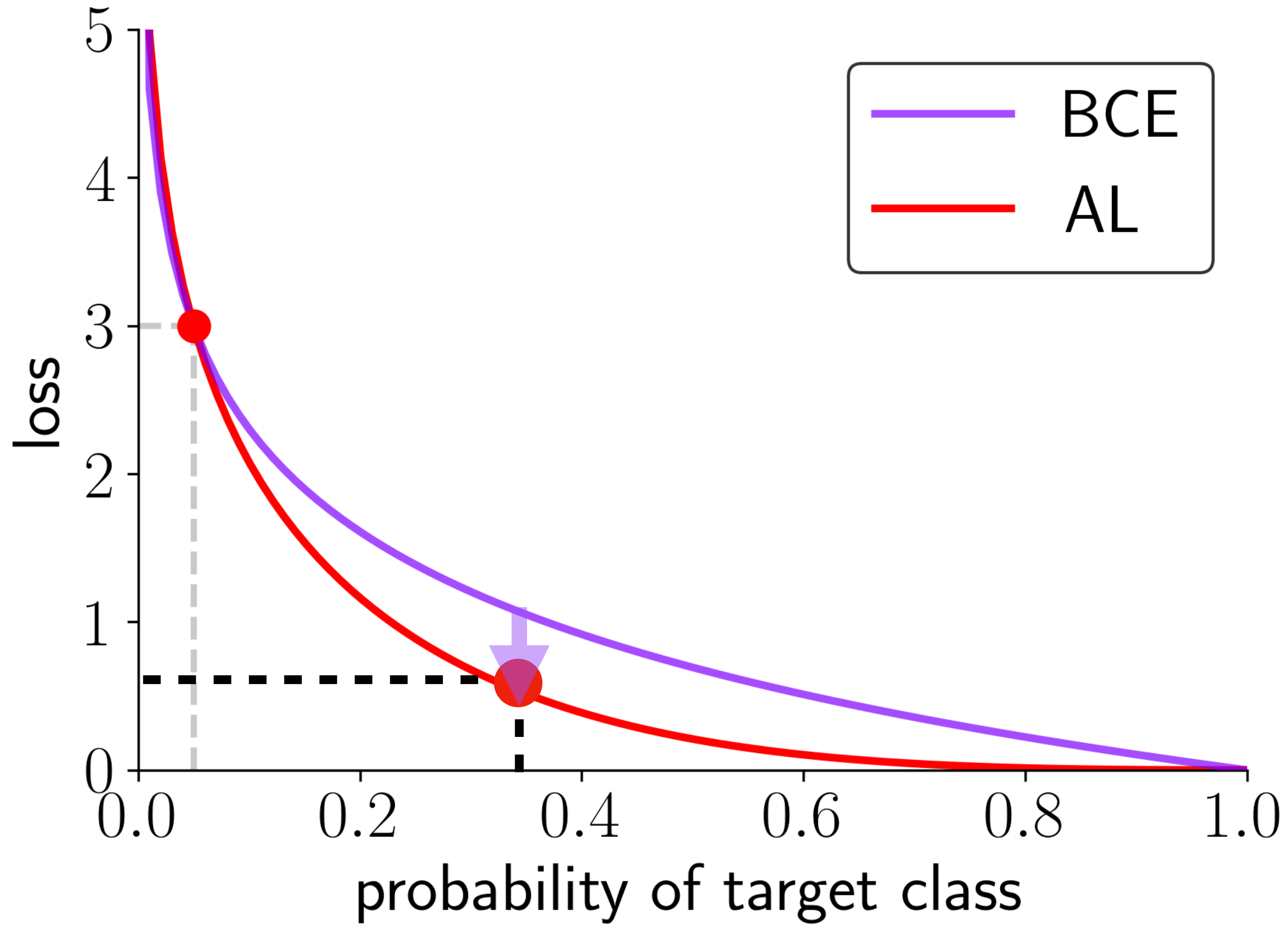


Easy case

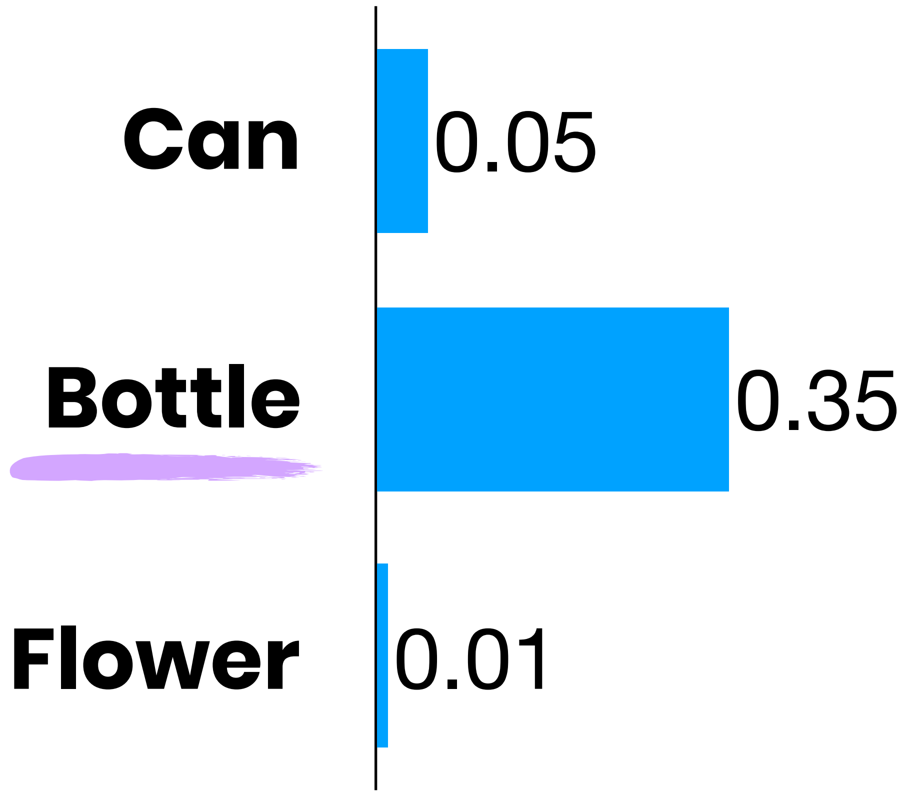
Input image

Bottle

Loss function



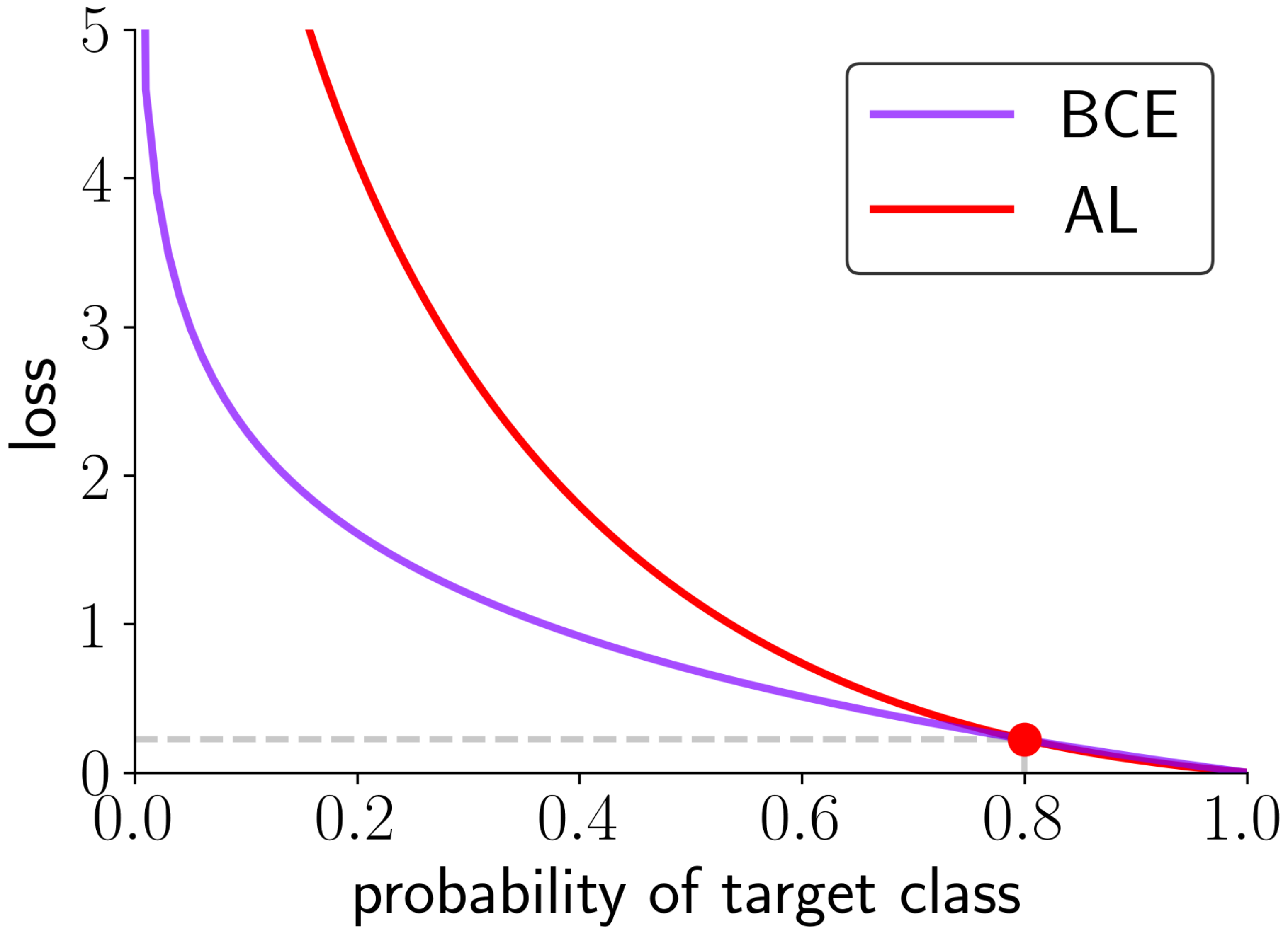
Output distribution



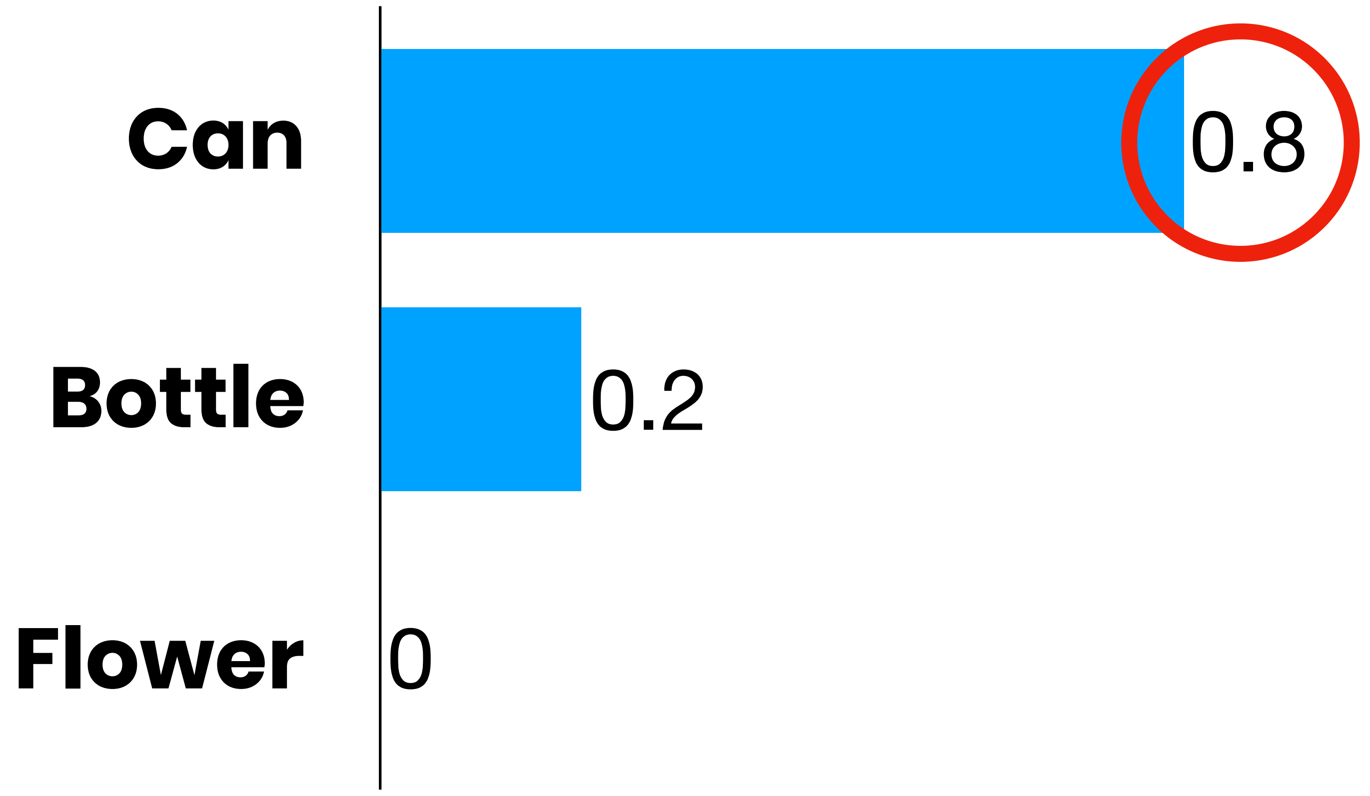
Hard cases

Input image
Bottle

Loss function



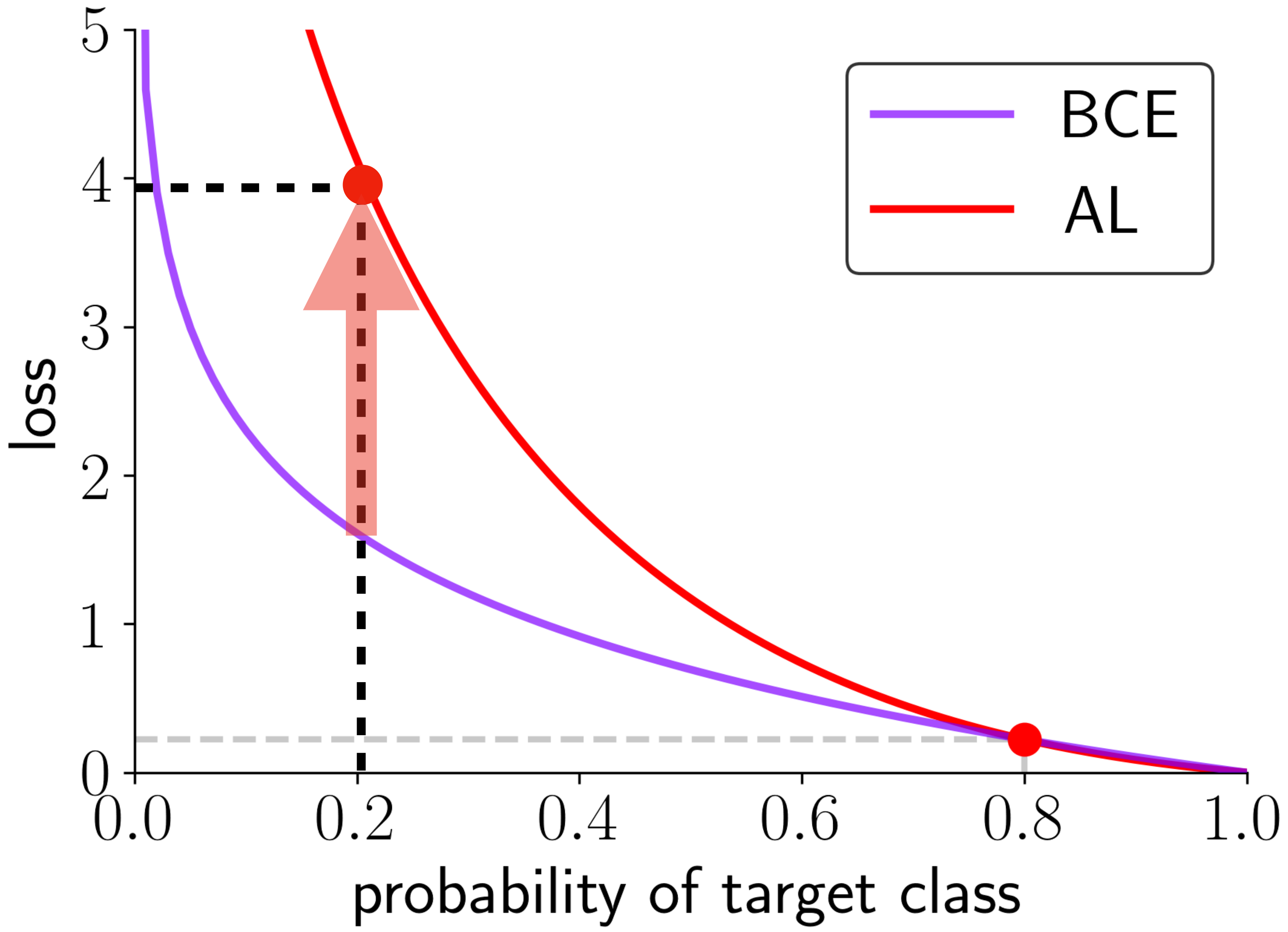
Output distribution



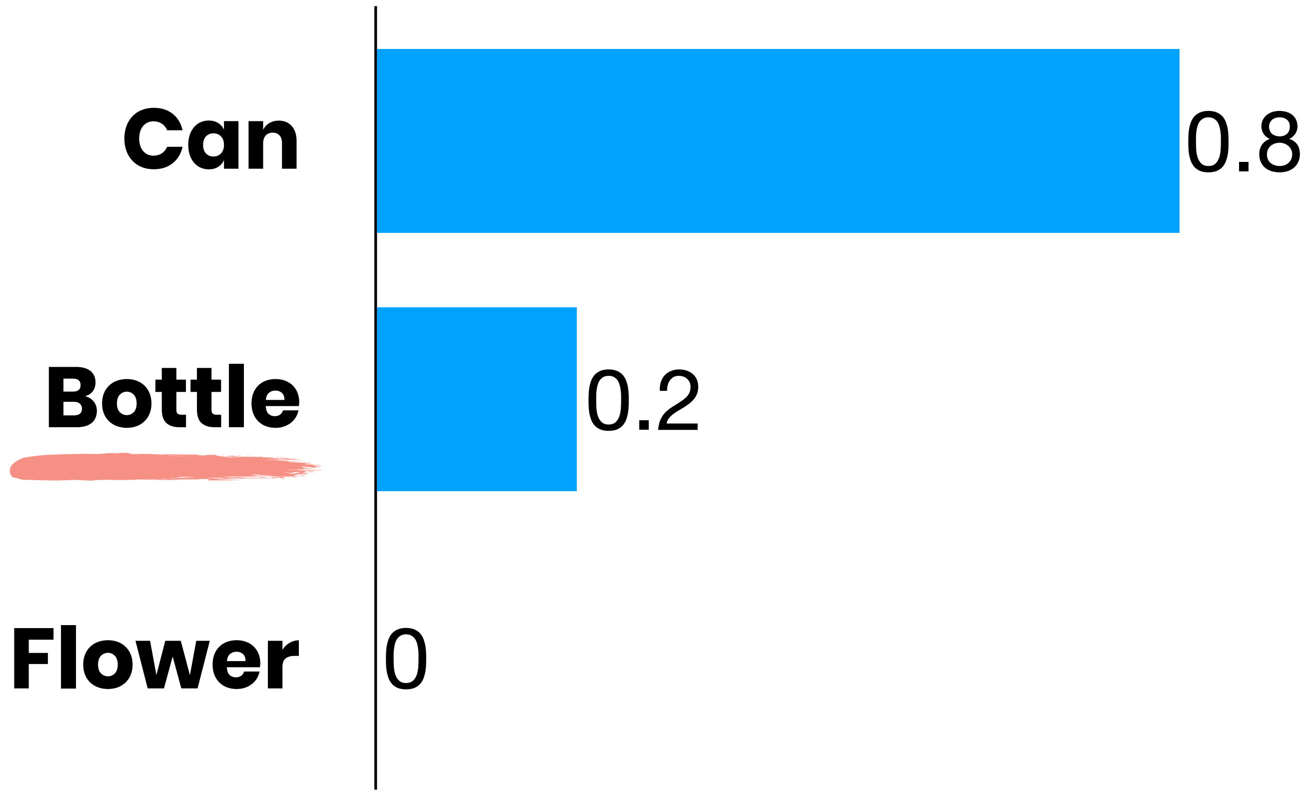
Hard cases

Input image
Bottle

Loss function



Output distribution



Anchor Loss on Pose Estimation

Anchor probability on human pose estimation

TARGET: Right Shoulder

- Spatial dependency between adjacent pixel locations

- Apply anchor loss on the background pixels

Anchor probability on human pose estimation

TARGET: Right Shoulder

Anchor Probability:

$$q_* = \max_{i \forall p_i > 0.5} q_i$$

Anchor probability on human pose estimation

TARGET: Right Shoulder

Anchor Probability:

$$q_* = \max_{i \forall p_i > 0.5} q_i$$

Anchor loss on background

$$M(p) = \begin{cases} 1 & \text{if } p = 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$l_{pose}(p, q; \gamma) = (M(p) * (1 + q - q_*)^\gamma + (1 - M(p))) * l_{BCE}(p, q)$$

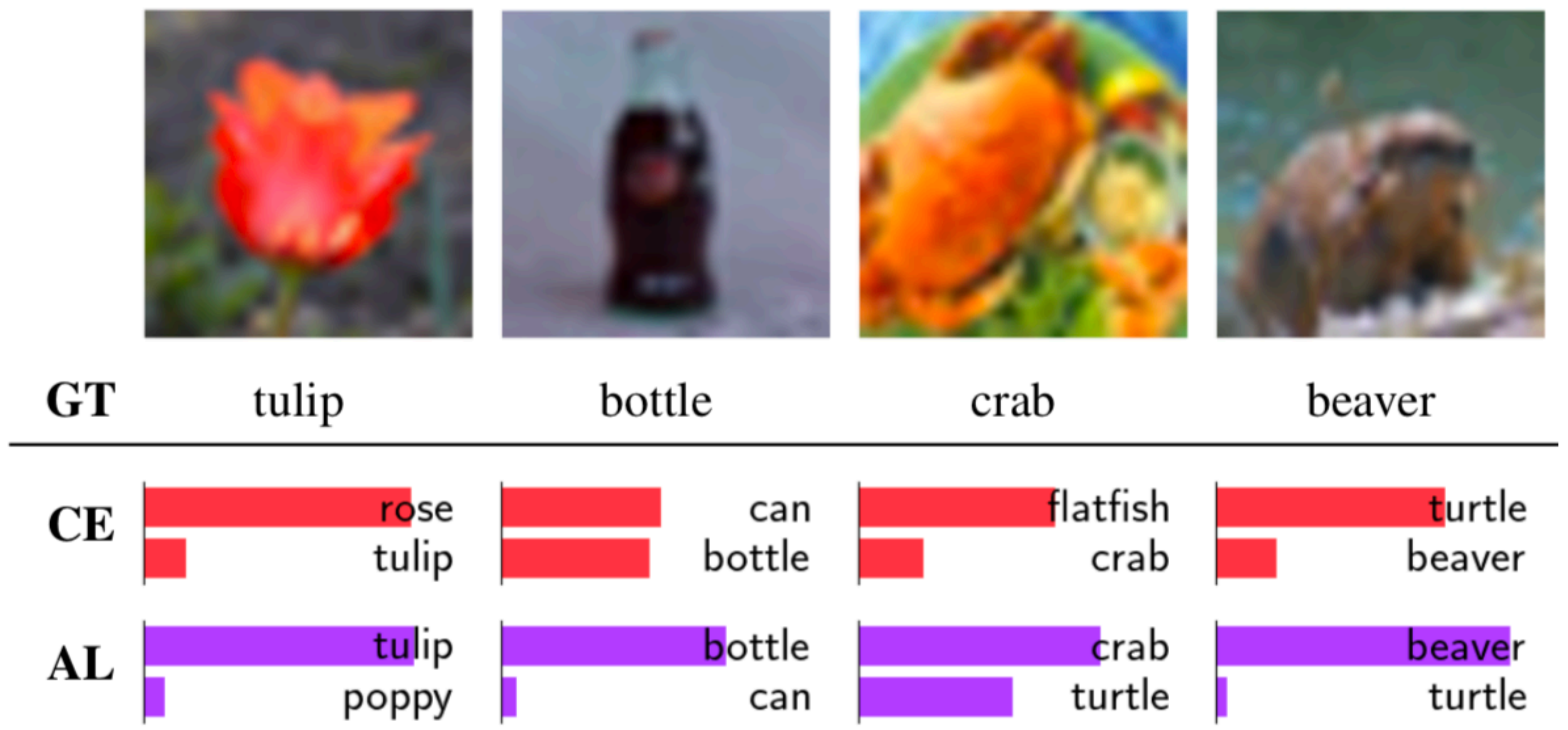
Results

Results: Image classification

- CIFAR

		CIFAR-10	CIFAR-100	
Loss Fn.	Parameter	Top-1	Top-1	Top-5
CE		93.91 ± 0.12	72.98 ± 0.35	92.55 ± 0.30
BCE		93.69 ± 0.08	73.88 ± 0.22	92.03 ± 0.42
OHEM	$\rho = 0.9, 0.9$	93.90 ± 0.10	73.03 ± 0.29	92.61 ± 0.21
FL	$\gamma = 2.0, 0.5$	94.05 ± 0.23	74.01 ± 0.04	92.47 ± 0.40
Ours				
AL	$\gamma = 0.5, 2.0$	$\mathbf{94.17 \pm 0.13}$	$\mathbf{74.38 \pm 0.45}$	92.45 ± 0.05

**With ResNet-110 model*



Results: Human pose estimation

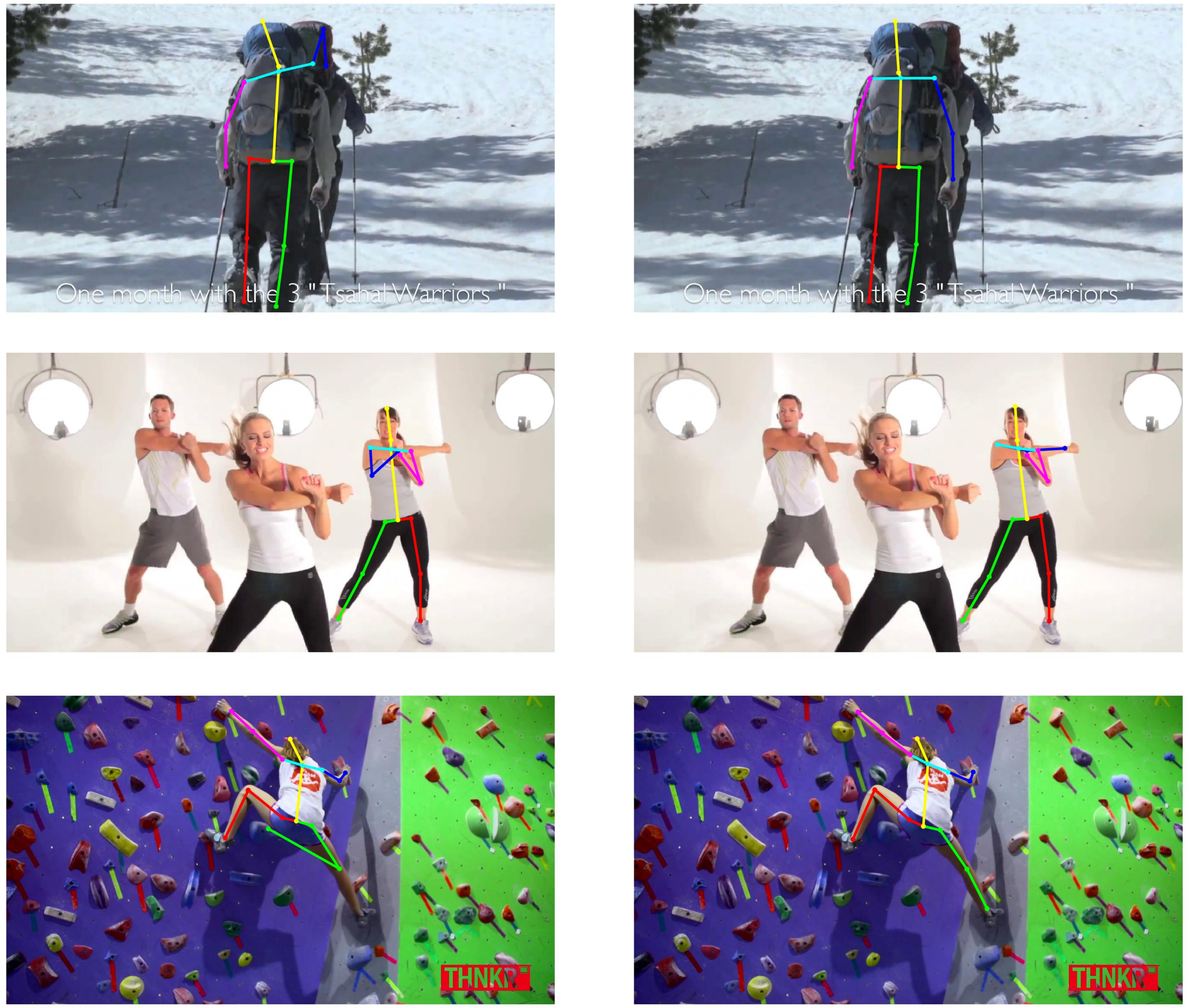
- MPII

Hourglass model variants	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Hourglass + MSE (Baseline)	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9
Hourglass + AL (Ours)	**98.6**	96.6	92.3	87.8	90.8	88.8	86.0	91.9
Chu *et al.*	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
Chen *et al.*	98.1	96.5	92.5	88.5	90.2	**89.6**	86.0	91.9
Yang *et al.*	98.5	96.7	92.5	**88.7**	**91.1**	88.6	86.0	92.0
Ke *et al.*	98.5	**96.8**	**92.7**	88.4	90.6	89.3	**86.3**	**92.1**

- LSP

Hourglass model variants	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Chu *et al.*	98.1	93.7	89.3	86.9	93.4	94.0	92.5	92.6
Yang *et al.*	98.3	94.5	92.2	88.9	**94.4**	**95.0**	93.7	93.9
Hourglass + AL (Ours)	**98.6**	**94.8**	**92.5**	**89.3**	93.9	94.8	94.0	**94.0**

MSE AL



Results: Human pose estimation

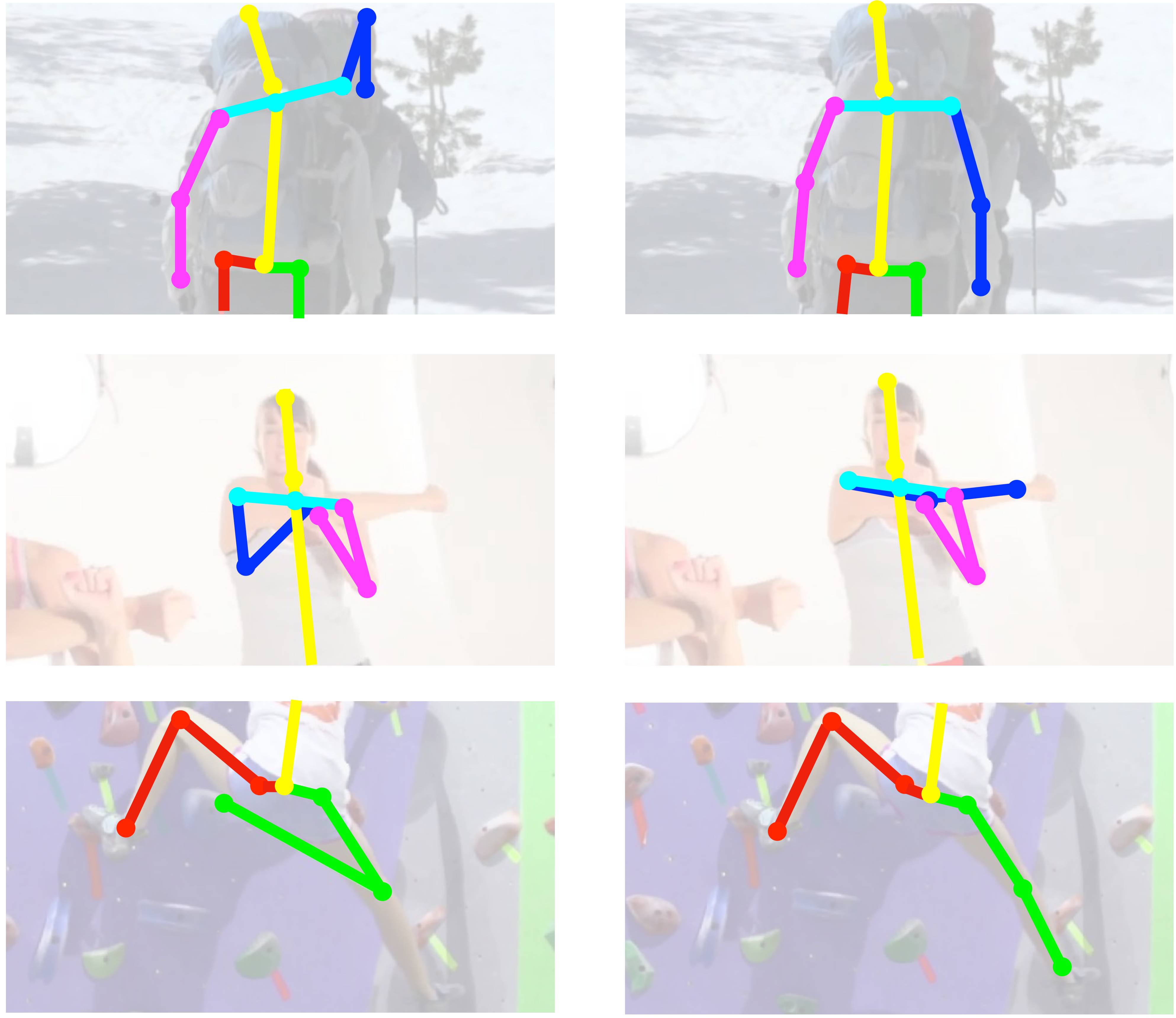
- MPII

Hourglass model variants	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Hourglass + MSE (Baseline)	98.2	96.3	91.2	87.1	90.1	87.4	83.6	90.9
Hourglass + AL (Ours)	**98.6**	96.6	92.3	87.8	90.8	88.8	86.0	91.9
Chu *et al*.	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
Chen *et al*.	98.1	96.5	92.5	88.5	90.2	**89.6**	86.0	91.9
Yang *et al*.	98.5	96.7	92.5	**88.7**	**91.1**	88.6	86.0	92.0
Ke *et al*.	98.5	**96.8**	**92.7**	88.4	90.6	89.3	**86.3**	**92.1**

- LSP

Hourglass model variants	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Total
Chu *et al*.	98.1	93.7	89.3	86.9	93.4	94.0	92.5	92.6
Yang *et al*.	98.3	94.5	92.2	88.9	**94.4**	**95.0**	93.7	93.9
Hourglass + AL (Ours)	**98.6**	**94.8**	**92.5**	**89.3**	93.9	94.8	94.0	**94.0**

MSE AL



Results

Input Image | **Areas where AL > BCE** | **Heatmap**