

基于渐进消歧和Logits标准化的 偏标记学习研究

郭洪涛 宗辰辰 唐英鹏 谢明昆 黄圣君

南京航空航天大学计算机科学与技术学院
模式分析和机器智能工业和信息化部重点实验室

偏标记学习



传统监督学习

- 标记唯一性
- 标记正确性

实例

哈士奇

偏标记学习

- 多个候选标记
- 一个真实标记

实例

萨摩耶

哈士奇

雪橇犬

偏标记学习

主流思路是设计策略识别候选标记集合中的真实标记，这个过程称为**消歧**。常见的消歧方法有：

➤ 平均消歧法

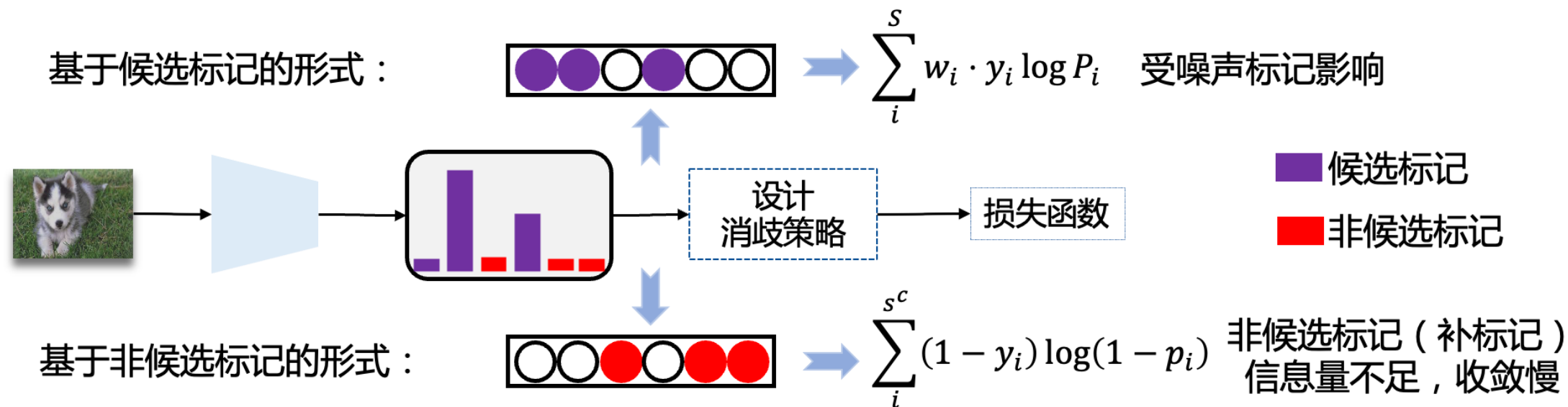
- 赋予每个候选标记**相同的权重**。例如对于一个实例，统计其 K 近邻样本的候选标记集合中出现次数最多的标记，作为该实例的真实标记 (Hullermeier et al., 2006)。
- ❖ 缺点：**容易受到候选标记集合中噪声标记的影响**，泛化效果较差。

➤ 判别消歧法

- 将真实标记看做隐变量，采用迭代的方式优化含隐变量的目标函数实现消歧。
- ❖ 缺点：**自训练迭代通常会导致错误积累** (Wang et al., 2021)。
- ❖ 缺点：**非候选标记集作为监督信号，信息量不足** (Wu et al., 2022)。



消歧的流程 & 动机



常见的消歧流程示意图

- 为了避免噪声标记的影响，在类别数目较多时也能有很好的收敛效果。
- ✓ 提出了一种高效的基于 Logits 标准化的补标记损失。
- 为了丰富迭代过程中的监督信息，避免模型训练仅依赖非候选标记集合产生偏差。
- ✓ 设计了一种新颖的双边渐进消歧策略。

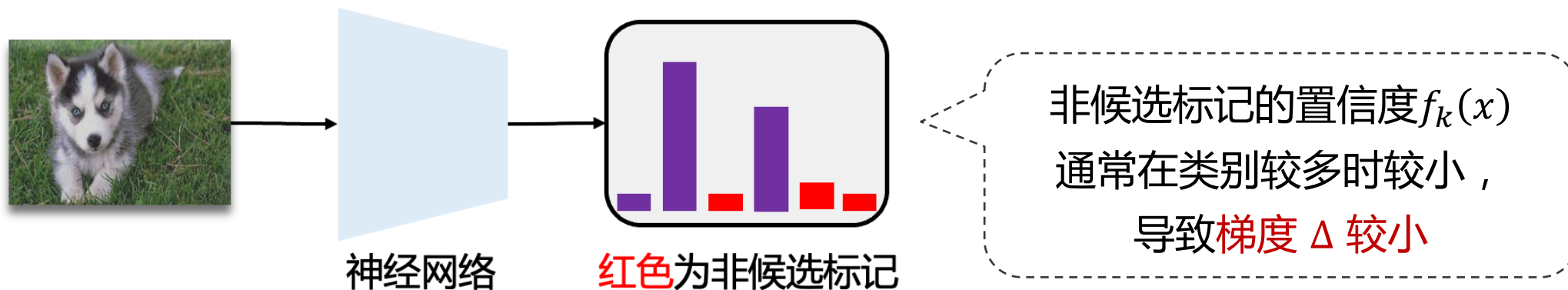
优化补标记损失

- 我们在非候选标记集 \mathcal{S}_i^c 上计算补标记损失，作为模型的监督信息，其中 $f_k(\mathbf{x}_i; \theta)$ 是模型输出：

$$\mathcal{L}_{\text{sup}}(\mathbf{x}_i, \mathcal{S}_i^c) = -\sum_{k \in \mathcal{S}_i^c} \log(1 - f_k(\mathbf{x}_i; \theta)), \quad (1)$$

- 然而式(1)的梯度为：

$$\Delta = \frac{\partial \mathcal{L}_{\text{sup}}(\mathbf{x}, \mathcal{S}^t)}{\partial f_k(\mathbf{x})} = -\frac{1}{1 - f_k(\mathbf{x})} \nabla f_k(\mathbf{x}), \quad (2)$$



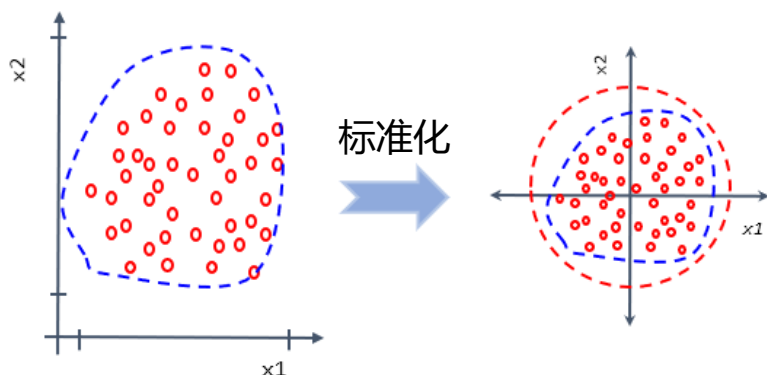
- 优化的点：让模型输出更加均衡的置信度分布，提升模型在非候选标记集合上的梯度

Logits 标准化

➤ Logits标准化目标：使模型输出相对均衡的置信度分布，然后计算补标记损失。实现方式

1. 计算原始 Logits 的均值和方差，进行标准化；
2. 将标准值输入线性网络 F_{linear} ，提升模型鲁棒性，得到最终标准化后的 Logits

步骤一：

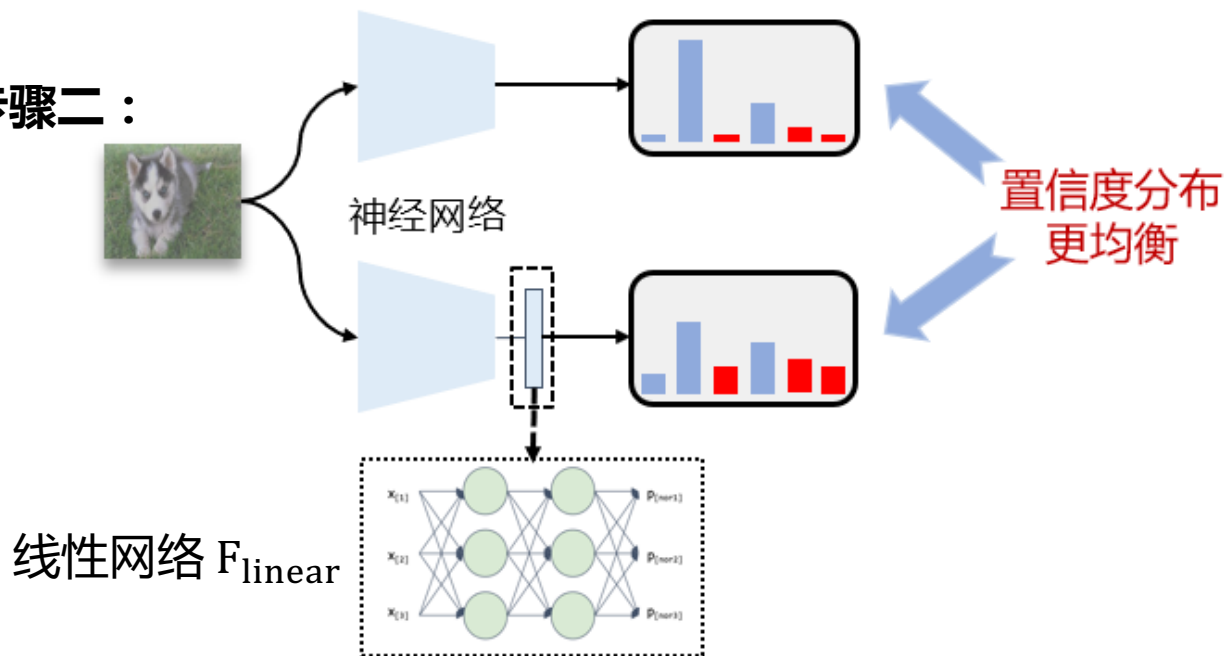


伪代码：

输入: 一个样本的原始输出Logits: $l \in \mathbb{R}^K$,
K 为类别数目, 学习参数 γ, β .
输出: $l_n = \text{Normalization}(l)$.

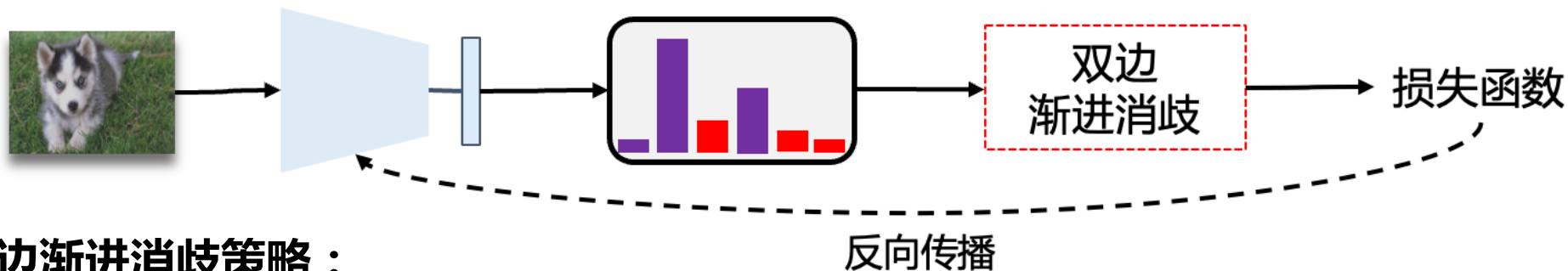
- ① 计算 l 的均值 $\mu \leftarrow \frac{1}{K} \sum_{i=1}^K l_i$;
- ② 计算 l 的方差 $\sigma^2 \leftarrow \frac{1}{K} \sum_{i=1}^K (l_i - \mu)^2$;
- ③ 对 l 进行标准化 $\hat{l} \leftarrow \frac{l - \mu}{\sqrt{\sigma^2 + \epsilon}}$;
- ④ 经过线性网络 $\text{output} = f(\hat{l})$

步骤二：



设计消歧策略

- 在标准化基础上如何设计消歧策略？



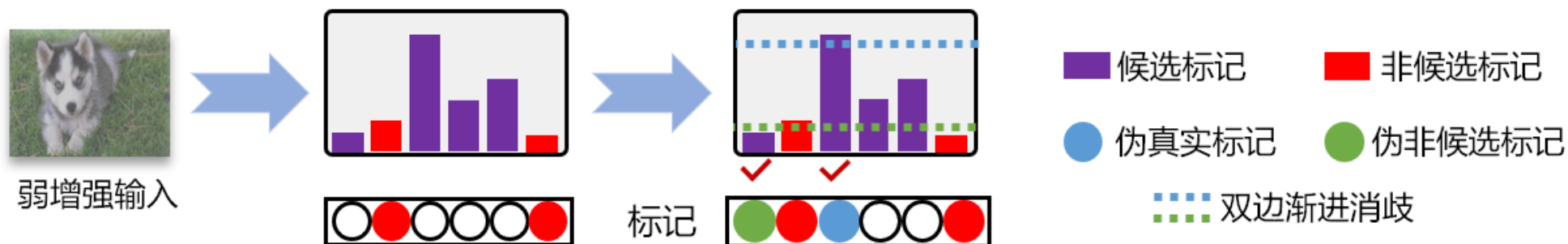
- 双边渐进消歧策略：

1. 动态更新非候选标记集合，设置较小的阈值：

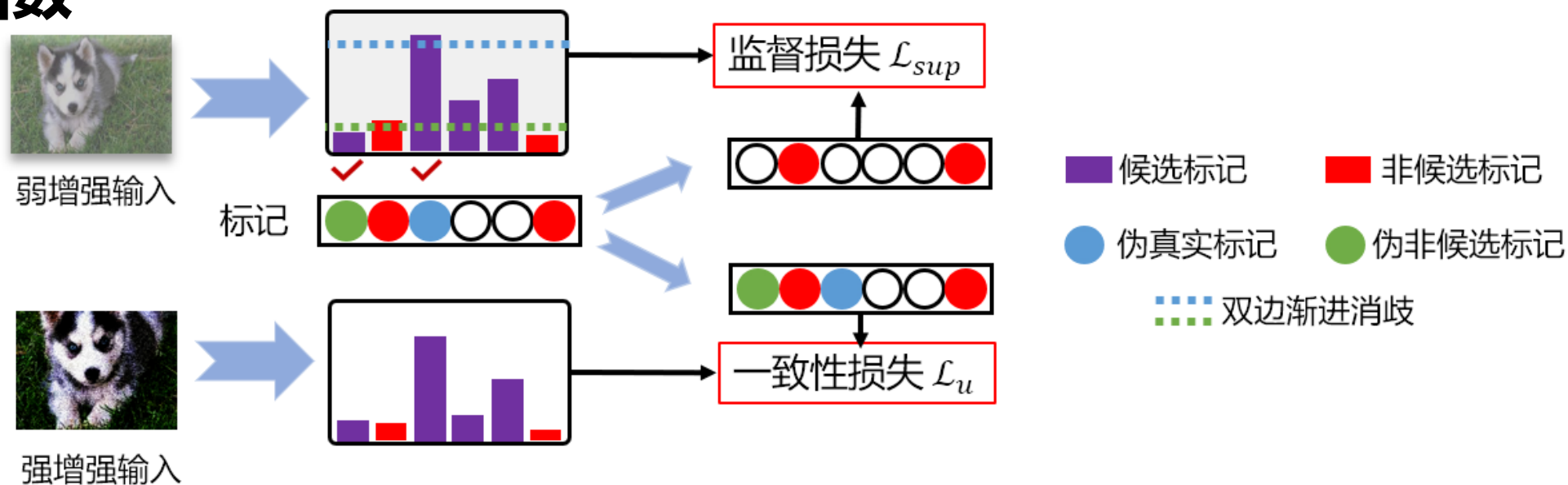
$$\tilde{\mathcal{S}}^c = \tilde{\mathcal{S}}^c \cup \{k \mid f_k(\mathbf{x}) < \tau_{\min} \mid k \in \mathcal{S}\}, \quad (3)$$

2. 动态更新真实标记，设置较大的阈值：

$$\hat{\mathbf{y}} = \operatorname{argmax} \left(f((\alpha(\mathbf{x}); \theta)) \mid f(\cdot) > \tau_{\max} \right) \quad (4)$$



损失函数



对非候选标记集合 \mathcal{S}_i^c 计算补标记损失：

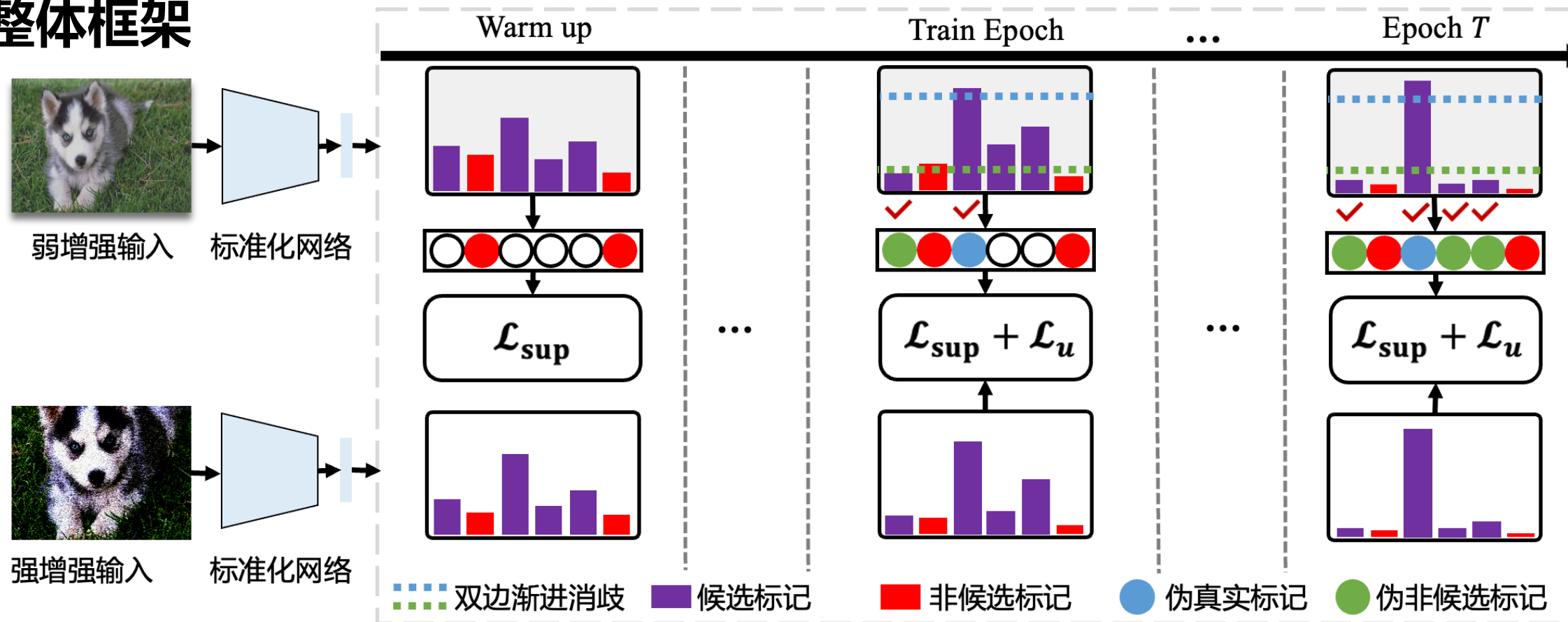
$$\mathcal{L}_{sup}(\mathbf{x}, \mathcal{S}_i^c) = - \sum_{k \in \mathcal{S}_i^c} \log(1 - f_k(\mathbf{x}; \theta)), \quad (5)$$

对伪标记集合 $L = \tilde{\mathcal{S}}^c \cup \hat{y}$ ，计算一致性损失，其中 $H(\cdot)$ 表示交叉熵， $\mathcal{A}(\cdot)$ 表示强增强输出：

$$\mathcal{L}_u(\mathbf{x}, L) = \frac{1}{B} \sum_{b=1}^B H(l, f(y | \mathcal{A}(\mathbf{x}_b); \theta)), l \in L \quad (6)$$

损失函数为： $\mathcal{L} = \mathcal{L}_{sup} + \lambda \mathcal{L}_u$

整体框架



主要贡献

- 提出了**双边渐进消歧**策略，通过双边伪标记机制丰富监督信息。
- 提出了基于**Logits标准化的补标记损失**，提升模型收敛速度和对噪声标记的鲁棒性。
- 实验表明，我们的方法在多个数据集的不同设定下，**性能优于SOTA方法**。

实验对比

➤ 实验设置

- 在MNIST和CIFAR上构造不同偏标记率的均匀候选标记集合。实验结果取多次实验准确率的均值和方差。
- 实验使用 WideResNet 和 LeNet 网络。

➤ 实验结果

- 我们的方法在多个数据集的多种偏标记率下基本都取得最优效果。
- 当标记空间较大和偏标记率较高时，我们算法的优势尤为明显。

Table 1 Experimental results on multiple datasets

表 1 在多个数据集上的实验结果

数据集	方法	$q = 0.1$	$q = 0.3$	$q = 0.5$	$q = 0.7$
Fashion-MNIST	PiCO	93.36 \pm 0.09%	93.41 \pm 0.08%	92.88 \pm 0.03%	91.73 \pm 0.07%
	PRODEN	89.15 \pm 0.58%	89.10 \pm 0.26%	88.22 \pm 0.35%	85.87 \pm 0.28%
	LWS	91.44 \pm 0.13%	91.85 \pm 0.14%	90.59 \pm 0.18%	89.46 \pm 0.16%
	RC	92.64 \pm 0.14%	92.08 \pm 0.03%	92.01 \pm 0.04%	90.83 \pm 0.35%
	ReCR	93.79 \pm 0.05%	93.72 \pm 0.20%	93.38 \pm 0.08%	92.19 \pm 0.03%
	Our	93.76 \pm 0.06%	93.73 \pm 0.05%	93.42 \pm 0.10%	92.34 \pm 0.09%
Kuzushiji-MNIST	PiCO	97.68 \pm 0.06%	97.34 \pm 0.07%	97.15 \pm 0.03%	91.90 \pm 0.04%
	PRODEN	94.61 \pm 0.39%	93.08 \pm 0.46%	90.15 \pm 0.51%	81.10 \pm 0.78%
	LWS	96.22 \pm 0.10%	96.15 \pm 0.24%	95.43 \pm 0.02%	93.63 \pm 0.04%
	RC	96.84 \pm 0.09%	96.31 \pm 0.15%	96.17 \pm 0.06%	95.84 \pm 0.12%
	ReCR	98.27 \pm 0.07%	98.08 \pm 0.03%	97.44 \pm 0.04%	95.93 \pm 0.11%
	Our	98.28 \pm 0.08%	98.15 \pm 0.06%	97.64 \pm 0.05%	96.55 \pm 0.12%
CIFAR-10	PiCO	95.78 \pm 0.05%	95.25 \pm 0.06%	94.73 \pm 0.11%	92.73 \pm 0.08%
	PRODEN	91.94 \pm 0.32%	91.10 \pm 0.50%	89.82 \pm 0.47%	86.48 \pm 0.47%
	LWS	86.47 \pm 0.20%	84.31 \pm 0.14%	54.73 \pm 0.19%	38.49 \pm 0.24%
	RC	88.96 \pm 0.06%	87.49 \pm 0.17%	83.48 \pm 0.19%	75.01 \pm 0.19%
	ReCR	97.45 \pm 0.04%	97.28 \pm 0.02%	97.05 \pm 0.05%	95.77 \pm 0.08%
	Our	97.64 \pm 0.07%	97.53 \pm 0.03%	97.19 \pm 0.04%	96.74 \pm 0.10%
数据集	方法	$q = 0.01$	$q = 0.05$	$q = 0.1$	$q = 0.2$
CIFAR-100	PiCO	74.39 \pm 0.15%	73.97 \pm 0.09%	51.94 \pm 0.11%	20.29 \pm 0.04%
	PRODEN	72.55 \pm 0.77%	71.55 \pm 0.94%	70.84 \pm 0.87%	58.86 \pm 0.85%
	LWS	58.54 \pm 0.12%	55.19 \pm 0.23%	40.12 \pm 0.34%	23.90 \pm 0.18%
	RC	64.95 \pm 0.23%	62.48 \pm 0.14%	57.48 \pm 0.04%	44.13 \pm 0.23%
	ReCR	83.12 \pm 0.20%	82.77 \pm 0.10%	82.24 \pm 0.07%	80.97 \pm 0.29%
	Our	84.05 \pm 0.12%	83.23 \pm 0.20%	82.79 \pm 0.10%	82.03 \pm 0.20%

消融实验

➤ 不同网络结构的消融实验

- 我们使用 ResNet 和 WideResNet 对不同网络结构进行消融实验。
- 实验表明我们的方法在多种网络结构也能取得明显的性能领先。

Table 2 Ablation experiments on ResNet18
表 2 CIFAR-100 数据集上使用 ResNet-18 的效果对比

数据集	方法	$q = 0.01$	$q = 0.05$	$q = 0.1$
CIFAR-100	PiCO	73.09 \pm 0.34%	72.74 \pm 0.30%	69.91 \pm 0.24%
	ReCR	79.54 \pm 0.12%	78.96 \pm 0.06%	77.72 \pm 0.08%
	Our	79.64 \pm 0.10%	79.49 \pm 0.07%	78.69 \pm 0.10%

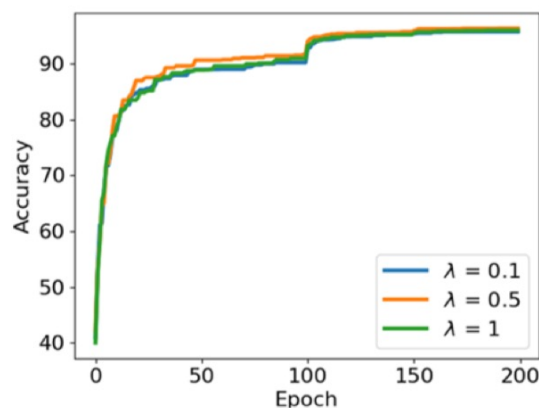
Table 3 Ablation study on CIFAR-10
表 3 对 Logits 标准化的消融实验

数据集	方法	$q = 0.1$	$q = 0.3$	$q = 0.5$	$q = 0.7$
CIFAR-10	Our w/o. Nor.	97.67	97.37	97.18	96.31
	Our	97.74(+0.07)	97.53(+0.16)	97.19(+0.01)	96.74(+0.43)
数据集	方法	$q = 0.01$	$q = 0.05$	$q = 0.1$	$q = 0.2$
CIFAR-100	Our w/o. Nor.	82.90	82.19	80.93	78.89
	Our	84.05(+1.15)	83.23(+1.04)	82.79(+1.86)	82.03(+3.14)

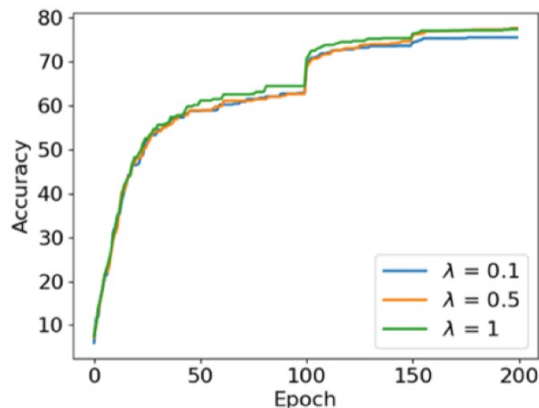
➤ Logits标准化的消融实验

- 我们在多个数据集多种偏标记率下进行了 Logits 层的消融实验。
- 消融实验表明了 Logits 标准化能有效提升模型性能。

消融实验



不同 λ 在CIFAR-10 分类效果



不同 λ CIFAR-100 分类效果

Figure 3 Ablation study on λ

图 3 不同 λ 的分类效果

➤ 超参数敏感分析

使用 ResNet18 网络对监督损失和一致性损失的平衡损失因子 $\lambda=\{0.1, 0.5, 1\}$ 进行了消融实验。实验表明提出的方法有**较强的稳定性**。

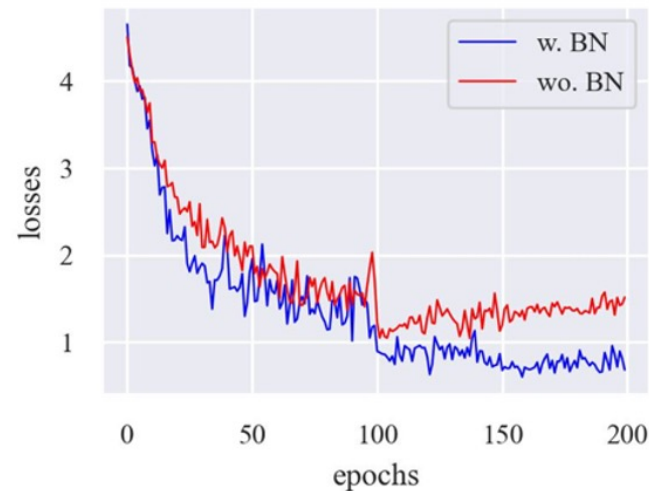


Figure 4 Ablation study on BN

图 4 Logits 标准化对收敛速度的影响

➤ Logits标准化层收敛速度实验

通过设定不添加标准化层的对照组实验，结果表明 Logits 标准化层能**加快模型收敛速度**。

敬请各位专家
批评指正！