ID 503

CAMRI Loss: Improving Recall of a Specific Class without Sacrificing Accuracy

Daiki Nishiyama†, Kazuto Fukuchi†, Youhei Akimoto†, Jun Sakuma†

† University of Tsukuba, RIKEN AIP

WCCI IJCNN 2022

Background

Focus on a real-world multi-class classification problem

- There are important classes that must not be misclassified (e.g., **Fig 1**)
- Improving recall of an important class is easy while sarcrificing accuracy (e.g., threshold tuning[1])
- Need to improve recall of an important class without sarcrificing accuracy

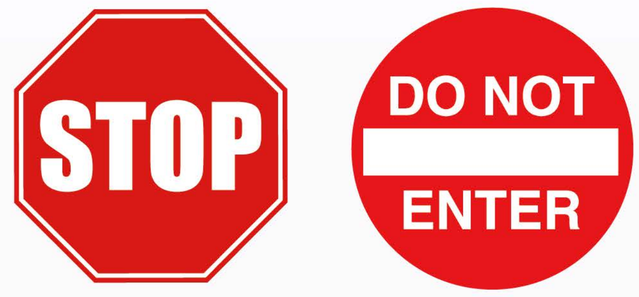
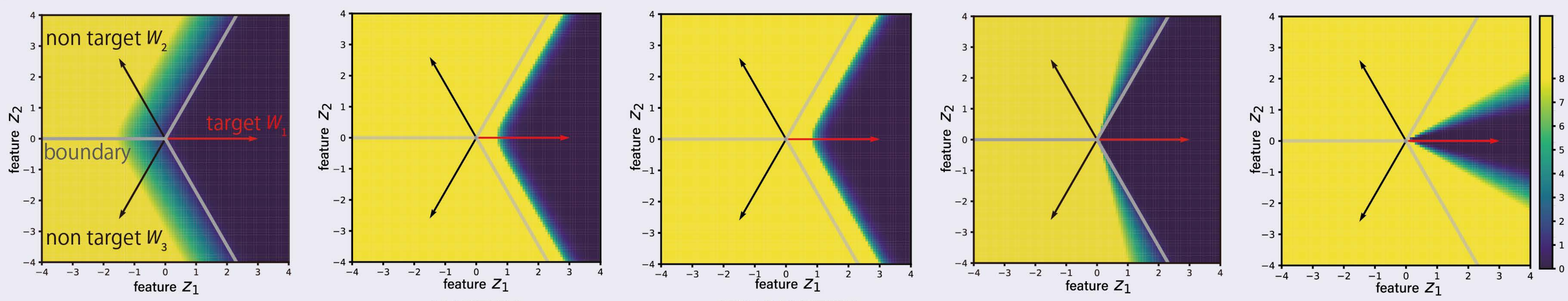


Fig 1. Examples of important classes in sign classification in automatic driving

Analysis of class-sensitive separation



WCE[2] CRWWCE[3]

(a) cross-entropy loss **(b)** approaches to penalize cross-entropy loss **(c)** L2-constrained softmax loss **(d-1)** ArcFace[4] **(d-2)** CAMRI loss

Fig 2. The contour plots of the loss values formed by each loss function in the space of 2D feature z. The last FC layer's weight vectors W_1, W_2, and W_3 correspond to three classes, and target class is 1 corresponding to W_1.

- Cannot improve separation when (b) & (c)
- Can improve separation when
  - **(d-1)** additive angular margin penalty[4]
- Need better separation of the important class than others
- ✓ Propose ***Class-sensitive additive Angular MaRgIn*** (CAMRI)
- ✓ CAMRI loss can improve class-sensitive separation
  - **(d-2)** Add angular penalty only for the important class

N: sample num, K: class num, y: teacher label, h: probability vector, s: temperature parameter, m: margin parameter

(a) $$\mathcal{L}_{\mathrm{ce}} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\{y_{n,k}\log(h_{n,k})\}$$

(c) $$\mathcal{L}_{L2} = -\frac{1}{N}\sum_{n=1}^{N}\log\left(\frac{\exp(s\cos(\theta_{t_n}))}{\sum_{k=1}^{K}\exp(s\cos(\theta_k))}\right) \text{,where } \theta_{t_n} = \arccos\left(\frac{\boldsymbol{W}_{t_n}^{T}\boldsymbol{z}_n}{\|\boldsymbol{W}_{t_n}\|\|\boldsymbol{z}_n\|}\right)$$

(d-1) $$\mathcal{L}_{\mathrm{ArcFace}} = -\frac{1}{N}\sum_{n=1}^{N}\log\frac{\exp(s\cos(\theta_{t_n}+m))}{\exp(s\cos(\theta_{t_n}+m))+\sum_{k\neq t_n}\exp(s\cos\theta_k)}$$

m_{t_n}: class-sensitive margin param

(d-2) $$\mathcal{L}_{\mathrm{CAMRI}} = -\frac{1}{N}\sum_{n=1}^{N}\log\frac{\exp(s\cos(\theta_{t_n}+m_{t_n}))}{\exp(s\cos(\theta_{t_n}+m_{t_n}))+\sum_{k\neq t_n}\exp(s\cos\theta_k)}$$

Experiments

Table 1. Mean of recall (upper row) and accuracy (lower row) among maintaining accuracy over ten trials.

	CIFAR-10			GTSRB			AwA2		
	worst1	worst2	median	worst1	worst2	median	worst1	worst2	median
CAMRI (proposal)	0.7919 (0.8821)	**0.8505** (0.8821)	**0.9171** (0.8798)	**0.8353** (0.9801)	**0.9600** (0.9842)	**0.9952** (0.9812)	**0.1296** (0.6671)	**0.2385** (0.6639)	**0.6457** (0.6558)
ArcFace	0.7656 (0.8805)	0.8427 (0.8797)	0.9109 (0.8805)	0.8080 (0.9806)	0.9416 (0.9823)	0.9938 (0.9819)	0.1018 (0.6660)	0.1877 (0.6660)	0.6123 (0.6767)
CRWWCE	- -	- -	- -	0.7966 (0.9809)	0.9233 (0.9808)	- -	- -	- -	- -
WCE	**0.8023** (0.8797)	- -	- -	0.8293 (0.9807)	0.945 (0.9814)	- -	- -	- -	- -
Cross-Entropy (baseline)	0.7383 (0.8796)	0.8342 (0.8796)	0.8886 (0.8796)	0.7600 (0.9800)	0.8700 (0.9800)	0.9896 (0.9800)	0.1185 (0.6514)	0.1456 (0.6514)	0.5599 (0.6514)

- CAMRI loss improves recall the most while maintaining accuracy in 8 out of 9 ways (**Table 1**)

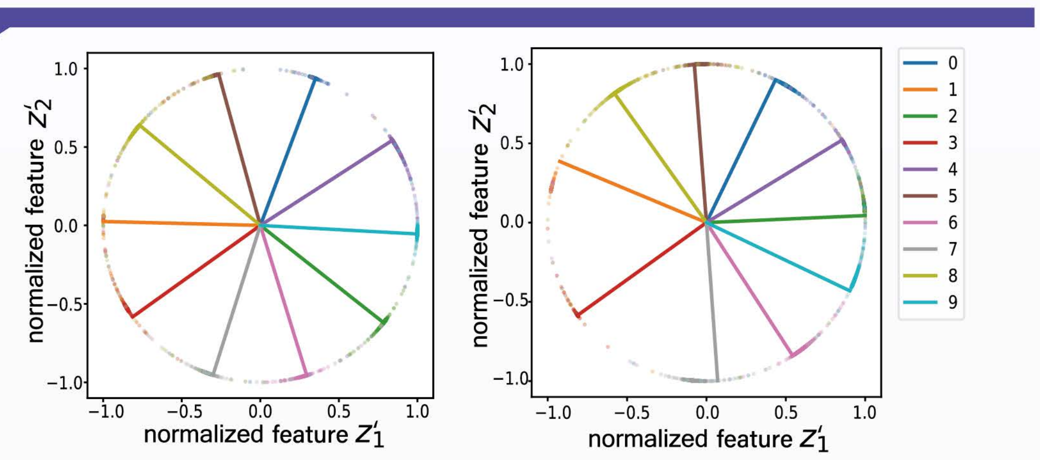


Fig 3. Normlized feature vectors of MNIST and weight vectors in the 2D feature space. "3" class is important class setting.

- The sparbility of the important class increase by training the model with CAMRI loss (**Fig 3**)

[1] D. D. Margineantu, T. G. Dietterich et al., "Bootstrap methods for the cost-sensitive evaluation of classifiers," 2000.
[2] S. Panchapagesan et al., "Multi-task learning and weighted cross-entropy for dnn-based keyword spotting." in Interspeech, 2016.
[3] Y. Ho et al., "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," IEEE Access, 2019.
[4] J. Deng et al., "Arcface: Additive angular margin loss for deep face recognition," CVPR, 2019.