

# A Discriminative and Robust Feature Learning Approach for EEG-Based Motor Imagery Decoding (Student Abstract)

Xiuyu Huang<sup>1</sup>, Nan Zhou<sup>2,1</sup>, Kup-Sze Choi<sup>1</sup>

<sup>1</sup>Centre for Smart Health, The Hong Kong Polytechnic University

<sup>2</sup> College of Control Engineering, Chengdu University of Information Technology  
xiuyu.huang@connect.polyu.hk, {nan2020.zhou, thomasks.choi}@polyu.edu.hk

## Abstract

Convolutional neural networks (CNNs) have been commonly applied in the area of the Electroencephalography (EEG)-based Motor Imagery (MI) classification, significantly pushing the boundary of the state-of-the-art. In order to simultaneously decode the discriminative features and eliminate the negative effects of non-Gaussian noise and outliers in the motor imagery data, in this abstract, we propose a novel robust supervision signal, called Correntropy based Center Loss (CCL), for CNN training, which utilizes the correntropy induced distance as the objective measure. It is encouraging to see that the CNN model trained by the combination of softmax loss and CCL loss outperforms the state-of-the-art models on two public datasets.

## Introduction

Center loss (Wen et al. 2016) was initially proposed to enhance the discriminative power of the deeply learned features in the computer vision classification field. It usually integrates with softmax loss and guides the CNN model training. It concurrently learns a feature center for each class and minimizes the distance between each deep feature point and its corresponding class center. Attributed to its efficiency and easy implementation, center loss was also widely applied to the EEG-based MI classification task. However, it is sensitive to the non-Gaussian noise and outliers, as it is based on quadratic  $L2$  norm distance. Few outliers far from the class centers may dominate the objective function and degenerate the model performance.

Concerning this issue, we propose a novel loss function, called Correntropy based Center Loss (CCL), based on correntropy induced distance (CID). It can simultaneously learn the class centers and penalizes the CID between features and corresponding class centers to minimize the intra-class variation. Unlike the original center loss that heavily penalizes the feature points far away from the centers and can be easily dominated by the outliers, the CCL assigns minor significance to these further feature points. Such a mechanism can eliminate the negative effect of the outliers during the model training, enabling the CNN model to learn a more robust feature pattern from the EEG-based motor imagery (MI) data, which usually contain much noise.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

## Methodology

In this section, we first introduce the proposed approach. Then, we show the implementation of a toy example used to intuitively demonstrate different characteristics between softmax loss, center loss, and CCL. Lastly, we describe the experimental evaluation of our method on the real-world MI data.

### Correntropy Based Center Loss

The CCL is the integration of the maximum correntropy criterion and the original center loss. It is formulated as follows.

$$\mathcal{L}_{CCL} = \frac{1}{2} \sum_{i=1}^m [1 - \exp(-\frac{\|x_i - c_{y_i}\|_2^2}{2\sigma^2})] \quad (1)$$

where  $x_i \in \mathbb{R}^d$  represents the  $i$ th sample's deeply learned features with a label of  $y_i$ .  $d$  denotes the feature dimension.  $c_{y_i}$  is the  $y_i$ th class center of deep features, and  $m$  is the sample size.  $\sigma$  is the size of the Gaussian kernel. The merit of this loss is that we can dynamically adjust kernel size  $\sigma$  to "control" the significance of the feature points that are far from the centers. We implement a two-step approach to update the deep feature  $x_i$  and class center  $c_{y_i}$  in succession within one epoch. We first use the gradient descent method in each epoch to update  $x_i$  using the following equation.

$$\frac{\partial \mathcal{L}_{CCL}}{\partial x_i} = \frac{1}{2\sigma^2} \exp(-\frac{\|x_i - c_{y_i}\|_2^2}{2\sigma^2})(x_i - c_{y_i}) \quad (2)$$

Then, we perform the half quadratic (HQ) (He et al. 2011) strategy to update class centers. Finally, the objective function for the CNN training is the joint supervision of softmax loss ( $\mathcal{L}_S$ ) and CCL loss.

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_{CCL} \quad (3)$$

where the  $\lambda$  is the trade-off scalar for balancing these two losses.

### A Toy Example

As the proposed loss is motivated by the center loss, we also conduct a toy example on the MINST dataset, the same one used in the original paper of center loss (Wen et al. 2016), for a fair comparison. We adopt all the settings of the toy

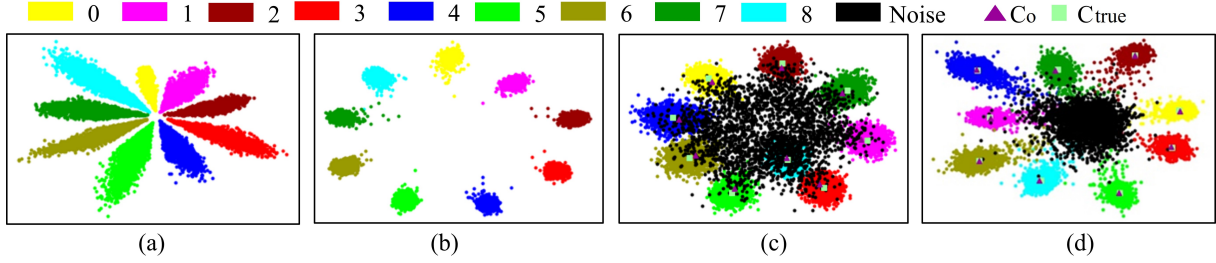


Figure 1: The distribution of deeply learned features under different schemes: (a) softmax loss and clean dataset; (b) joint loss function 1 (softmax loss and centre loss) and clean dataset; (c) joint loss function 1 and noisy dataset; (d) joint loss function 2 ( $\mathcal{L}$ ) and noisy dataset,  $\lambda = 1$ ;  $C_o$ s are updated during the training.  $C_{true}$ s are computed by the mean of feature points respectively in each class excluded from noise.

example described in that article, except that we assign the data points originally belonging to the class digit '9' to each of other labels (300 for each one, totally 2700 noise data points), which can be regarded as noise.

### Experiments on Real MI Data

We also benchmark the performance of the proposed method on two public MI datasets, called BCI Competition IV-2a and IV-2b. The CNN architecture is based on the modified EEGNet. We compare our method to four baselines, called FBCSP, SMM, EEGNet, and ConNet.

### Result and Discussion

The distributions of the deep features (toy example) trained via different schemes are shown in Figure 1. We have four key observations from Figure 1. (i) The joint supervision of the softmax loss ( $\mathcal{L}_S$ ), and center loss or CCL can ensure both a significant inter-class distance and a compact intra-class variation. (ii) The  $C_o$ s deviate from corresponding actual centers  $C_{true}$ s during the optimization when using the center loss (Figure 1 (c)). (iii) The feature distributions are significantly affected by the noise when using center loss, where the features of noise and clean data are pulled together. The latent distance between different classes may occasionally decrease (see Figure 1 (b) and (c)). (iv) As shown in Figure 1 (d), when using the CCL, the noise has barely influence on the feature distribution of the clean data, and all  $C_o$ s nearly overlap with  $C_{true}$ s.

The averaged accuracies of the different approaches for experiments using the MI data are shown in Table 1. It is observed that our approach outperforms all other baselines. The classification for each subject of both datasets can be seen in the supplemental file. This result indicates that the proposed method can effectively handle the real-world EEG data that are often mixed with signal and noise in one trial.

We also evaluate the efficiency of the proposed method in decreasing the negative effect of pure noise. We use the data of the subject 'A07' (random selection) in the IV-2a dataset for visualization. The most common noise of the EEG, electric flow signal (50 Hz), is chosen as the external noise. Twenty pure noise trials with each ten labeled 'left hand' or 'right hand' are added to the data. As shown in Figure 2,

	FBCSP	SMM	EEGNet	ConNet	Ours
IV-2a	67.75	70.72	74.50	72.53	<b>79.07</b>
IV-2b	80.01	72.68	75.88	79.37	<b>85.29</b>

Table 1: Averaged accuracy (%) obtained with IV-2a and 2b

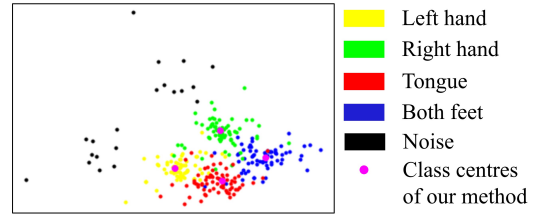


Figure 2: Visualization of feature distribution of subject A07

the pure noise has little impact on the model training when using our method.

### Conclusion

We proposed a novel loss function, referred to as CCL. By combining softmax loss with CCL to jointly supervise the learning of CNNs, the discriminative power and robustness of the deeply learned features can be significantly increased for the MI classification. Experiments on two famous MI datasets demonstrate the effectiveness of the CCL.

### Acknowledgments

This work was supported in part by GRF grant PolyU 152006/19E and NSFCs (Nos. 61802036 and 11901063).

### References

- He, R.; Hu, B.-G.; Zheng, W.-S.; and Kong, X.-W. 2011. Robust principal component analysis based on maximum correntropy criterion. *IEEE Transactions on Image Processing*, 20(6): 1485–1494.
- Wen, Y.; Zhang, K.; Li, Z.; and Qiao, Y. 2016. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, 499–515. Springer.