CC-LOSS: CHANNEL CORRELATION LOSS FOR IMAGE CLASSIFICATION

Zeyu Song, Dongliang Chang, Zhanyu Ma, Xiaoxu Li, and Zheng-Hua Tan

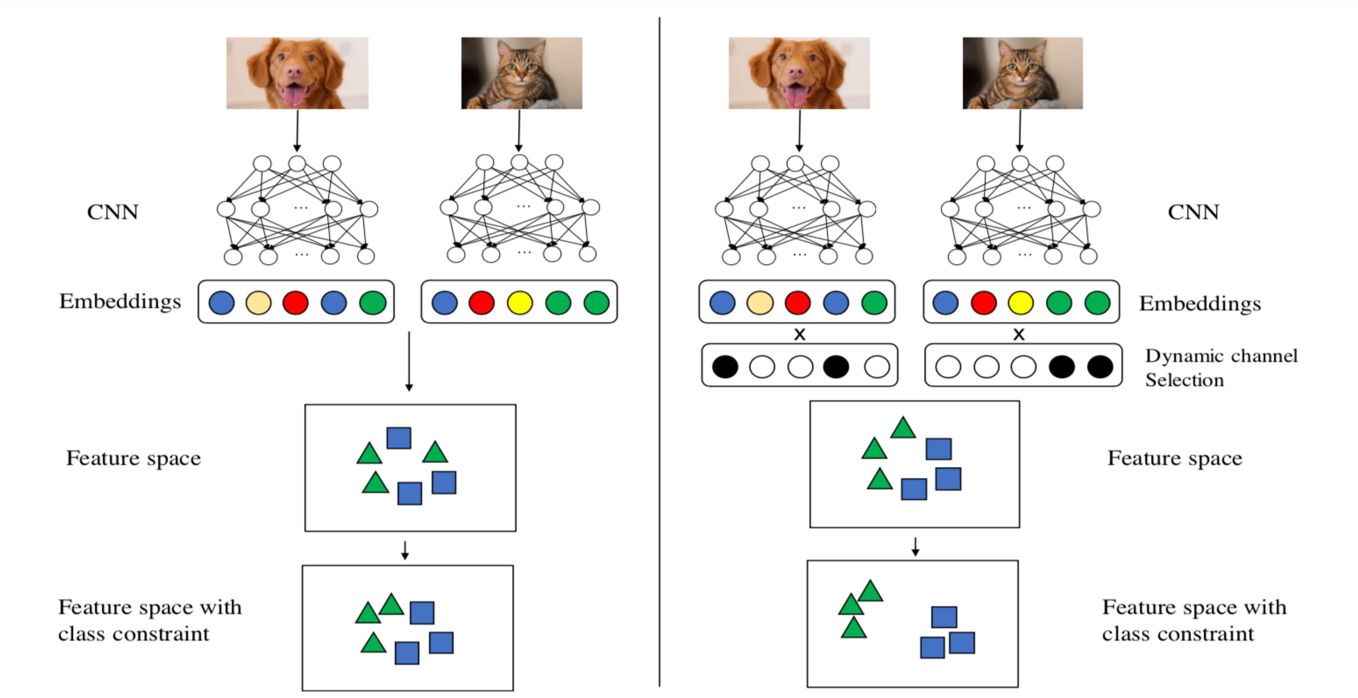
Bejing University of Posts and Telecommunications, Lanzhou University of Technology, Aalborg University

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Introduction

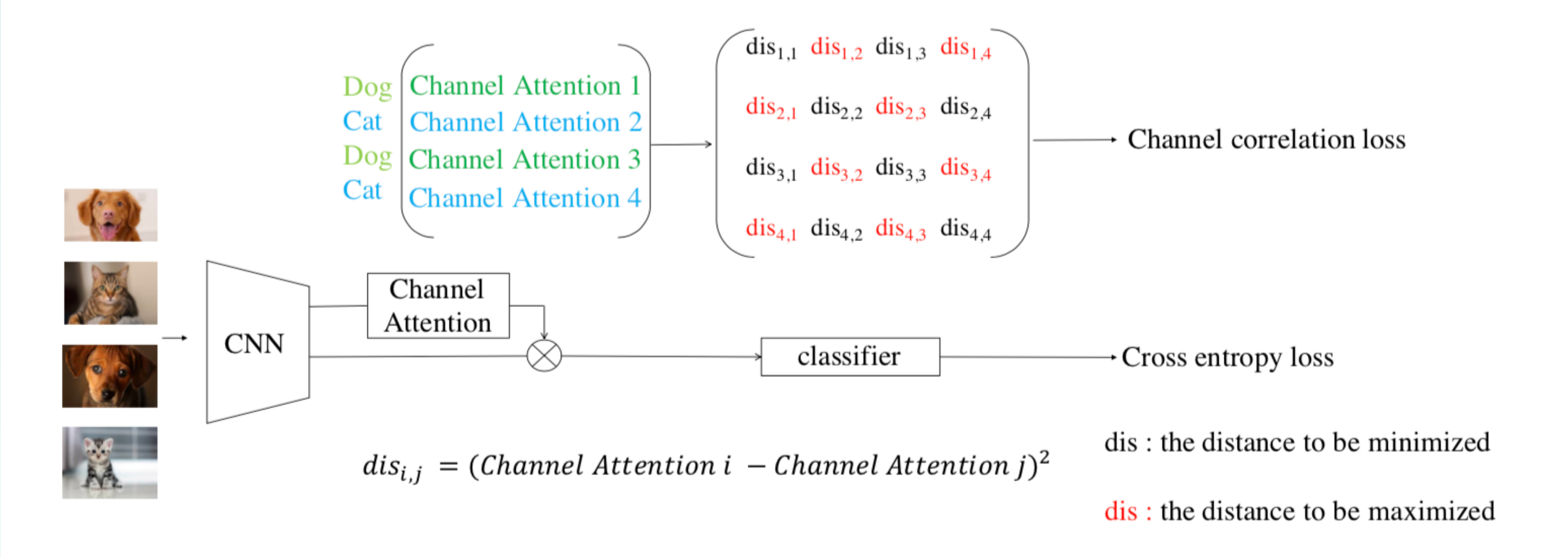
The loss function is a key component in deep learning mod- els. A commonly used loss function for classification is the cross entropy loss, which is a simple yet effective application of information theory for classification problems. Based on this loss, many other loss functions have been proposed, *e.g.*, by adding intra-class and inter-class constraints to enhance the discriminative ability of the learned features. However, these loss functions fail to consider the connections between the feature distribution and the model structure. Aiming at ad- dressing this problem, we propose a channel correlation loss (CC-Loss) that is able to constrain the specific relations be- tween classes and channels as well as maintain the intra-class and the inter-class separability. CC-Loss uses a channel attention module to generate channel attention of features for each sample in the training stage. Next, an Euclidean distance matrix is calculated to make the channel attention vectors as- sociated with the same class become identical and to increase the difference between different classes. Finally, we obtain a feature embedding with good intra-class compactness and inter-class separability. Experimental results show that two different backbone models trained with the proposed CC-Loss outperform the state-of-the-art loss functions on three image classification datasets.

Methods



Comparison between the CC-Loss and other related loss functions. The left column shows loss functions that only consider intra- and inter-class constraints. The right column shows the CC-Loss that enhances both the class relation constraints and feature discrimination.

Methods



The pipeline of the proposed CC-Loss. Channel attention vectors extracted from each sample are labeled with ground truth classes. The labeled channel attention vectors will then be used to calculate a distance matrix, representing the similarity between channel attentions of different samples. The similarity of samples belonging to the same class should be minimized while the similarity of samples belonging to the different class should be maximized. The distances between samples from the same class is marked in black and those from different classes are marked in red.

Results

Loss Function	Backbone Model	MNIST	CIFAR-100	Cars-196
CE Loss	VGG16/ResNet18	97.43/97.52	74.49/77.38	88.02/85.77
Focal Loss	VGG16/ResNet18	97.64/97.68	74.46/77.63	88.21/85.98
A-softmax Loss	VGG16/ResNet18	98.10/98.32	74.55/77.78	90.02/87.22
MC Loss	VGG16/ResNet18	98.20/98.45	72.51/70.18	**92.80/-**
ICAL	VGG16/ResNet18	97.83/98.21	74.79/77.71	89.32/86.67
FICAL	VGG16/ResNet18	98.22/98.40	74.98/78.18	89.70/87.38
CC-Loss	VGG16 + CAM/ResNet18 + CAM	**98.32 ± 0.08/98.52 ± 0.09**	**75.49 ± 0.15/78.23 ± 0.07**	91.46 ± 0.09/**88.41 ± 0.06**

Classification accuracy(%) with different loss functions. The backbone models were VGG16 and ResNet18 and evaluated on three datasets. All the CC-loss results are the average of five rounds evaluations.
We further analyze the results of the MC- Loss. The MC-Loss has better performance on the Cars-196 dataset, which is a FGVC dataset. But it has worse performance on image classification datasets, *i.e.*, MNIST and CIFAR-100.

Ablation Study

Loss Function	Backbone Model	MNIST	CIFAR-100	Cars-196
CE Loss	VGG16/ResNet18	97.43/97.52	74.49/77.38	88.02/85.77
CE Loss	VGG16 + CAM/ResNet18 + CAM	97.66 ± 0.02/97.67 ± 0.01	74.54 ± 0.02/77.57 ± 0.01	88.15 ± 0.03/86.04 ± 0.02
CC Loss	VGG16 + CAM/ResNet18 + CAM	98.32 ± 0.08/98.52 ± 0.09	75.49 ± 0.15/78.23 ± 0.07	91.46 ± 0.09/88.41 ± 0.06

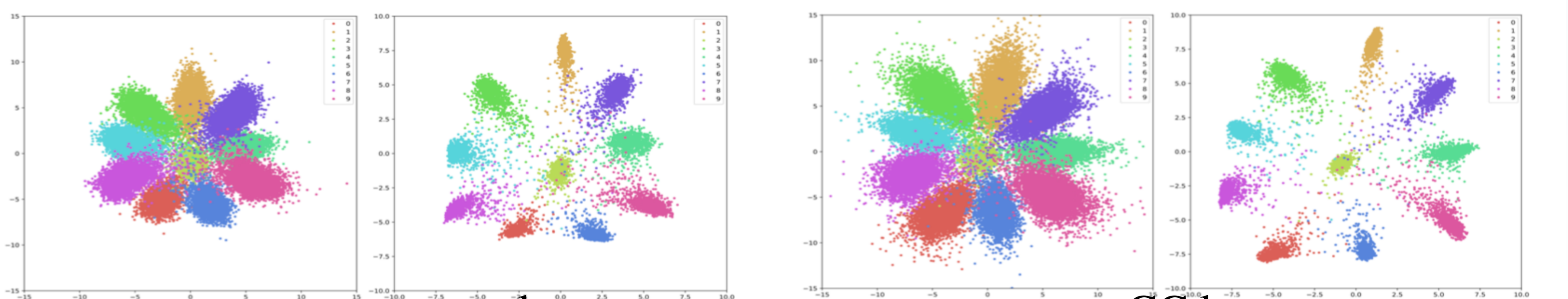
Effect on Channel Attention Module. In the case of w/o CC-Loss, dynamic channel selection gives slight improvement against the original backbone. When CC-Loss is used, it provides extra class relation constraint to boost the performance. Also, it is worth to note that CC-Loss can not be implemented without the CAM module.

λ	MNIST	CIFAR-100	Cars-196
0	97.66 ± 0.02	74.54 ± 0.02	88.15 ± 0.03
0.5	98.04 ± 0.05	75.12 ± 0.03	89.53 ± 0.07
1.0	98.32 ± 0.08	75.49 ± 0.15	91.46 ± 0.09
1.5	97.95 ± 0.06	75.04 ± 0.09	89.42 ± 0.05
2.0	97.53 ± 0.03	74.32 ± 0.12	87.84 ± 0.02

Batch Size	CIFAR-100	Cars-196
8	74.9 ± 0.12	91.3 ± 0.12
16	75.3 ± 0.07	91.5 ± 0.09
32	75.5 ± 0.15	91.3 ± 0.16
64	75.1 ± 0.14	90.6 ± 0.18
128	74.5 ± 0.14	89.9 ± 0.15

Effect on λ and batch size. We first evaluated the performance of different λ, representing the influence of CC-Loss. The best result is obtained when λ is 1. Further increasing λ will cause performance drop because we need class-label relationship pro- vided by the CE-Loss. Since we optimize the channel attention distances within mini-batch, the batch size is an important hyper-parameter to be tuned. We analyzed the affect of batch size on CIFAR- 100 and Cars-196 datasets with VGG16 as the backbone. A smaller batch size leads to a higher classification accuracy, since the class relationship in the distance matrix is simpler and easier to optimize. However, when the batch size tends to be very small, the class relationship will disappear and thus the accuracy is decreased.

Visualization



cross entropy loss　　CC loss

From the figure above, it can be observed that the features from the CC- Loss are more compact within the same class while more separable among the neighbor classes. This demonstrates the effectiveness of intra-class and inter-class component of CC-Loss.