

# Improved Metric Learning Method Implemented in Classification Task

Xiaochen Zhou, Jiahao Li, Xiwen Li

zhouxiaochen@wustl.edu, xx@wustl.edu, xx@wustl.edu

## Abstract

*As a powerful method for recognition and re-identification task, metric learning greatly improve the accuracy and performance of tasks which require distance optimization between different features. Taking this property into consideration, we come out with the method to optimize the classification method with triplet loss method and improve this method to better suit the requirement of classification task. We will implement the method with convolutional neural network and make the comparison of the performance of classification between original CNN, CNN with triplet loss and CNN with our improved triplet loss.*

## 1 Introduction

Metric learning, highly related to similarity learning, is an area of supervised machine learning in artificial intelligence. As a powerful method to learn from examples a similarity function that measures how similar or related two objects are, metric learning is widely use in tasks like face recognition and re-identification tasks, where the difference between special sections are much more important than the distance of the global features. Compared with classification tasks, datasets for these tasks share high similarity and the global feature supervised by softmax-entropy loss is not strong enough to clarify the difference between two faces. Metric learning method, like triplet loss which is used in this paper, is invented for further optimization. By enlarging the distance between classes, called intra-class distance and lowering the distance in the same classes, called inter-class distance, metric learning make it possible for the machine learning model to tell the difference between two similar objects.

Considering the property of metric learning, some supervising signal like triplet loss can be exerted in model training as the auxiliary loss. This method will improve the performance of classification for neural network, especially for some hardcore classification task, like sofa and chair, sofa and bed etc., which shares relatively higher similarity. Also, the auxiliary loss can cluster the classes much tighter, to some extent avoiding overfitting for the

small dataset.

In this paper, we optimize triplet loss as a new supervising signal, semi-hardcore triplet loss, to help improve the performance of classification task with the assistance of convolutional neural network (CNN). Several researchers paid attention to the improvement of softmax-entropy loss function and intra-class distance optimization. Lee et.al [4] takes the advantages of dropout method to lower the inter-class distance. Deliége et.al [1] developed a new hit-and-miss layer for the better performance of softmax-entropy loss. Useful as these method, new layers implemented in neural network increases the computation complexity and makes the training process longer. Using metric learning method as auxiliary loss function, however, effectively avoid the increasing computation and improve the classification performance at the same time, making the neural network training more efficient. Then, our method will improve the capability for the loss to tighten the inter-class distance and enlarge the intra-class distance.

An overview of the following sections are as follow: In section 2 we will introduce the related work on metric learning. In section 3, algorithms for our method will be given and analyzed. Then section 4 will show the performance of classification among bared softmax-entropy loss, loss with triplet loss and loss with improved triplet loss. Last section will be the conclusion.

## 2 Background & Related Work

Feat by Xiwen Li P.S. It should work to search information on wiki about metric learning

## 3 Algorithm

Triplet loss is first invented in Scgriff et.al [5], which is exerted in face recognition task. The embedding system of triplet loss is functioning as follows. An image is set as the input of the neural network and the feature is extracted in a  $d$ -dimensional Euclidean space. To improve the performance of face recognition, another two images

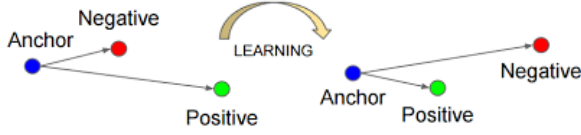


Figure 1: Visualization of triplet loss function

are sent into network as a triplet pair, where the original input is called image  $x_i^a$  (anchor), the image sharing the same class with anchor image  $x_i^p$  (positive) and the image with different class with anchor image  $x_i^n$  (negative). To enlarge the inter-class distance and lower the intra-class distance, the equation can be written as follows:

$$\|x_i^a - x_i^p\|_2^2 + \alpha < \|x_i^a - x_i^n\|_2^2, (x_i^a, x_i^p, x_i^n) \in \Gamma$$

where  $\alpha$  is the margin distance between positive pairs and negative pairs and  $\Gamma$  is the set of all possible triplet pairs  $N$  in training period. Figure 3 visualize the process of optimization for triplet loss. Then the loss function can be expressed as follows:

$$Loss = \sum_i^N [\|x_i^a - x_i^p\|_2^2 - \|x_i^a - x_i^n\|_2^2 - \alpha]_+$$

Then to improve the performance of classification, we need to do some changes on data selection. Owing to the differences for recognition tasks are relatively small, using metric learning can making the training period unconverging. However, considering that the difference between class in classification task is usually larger than recognition tasks, we do not need to worry too much about the negative influence from dataset. On contrast, we need to modify the loss function, making to stronger for classification tasks.

For training period, when anchor image is set, we search the dataset to find an image with the largest distance in the same class and to find an image with the lowest distance in different classes. The equation can be written as follows:

$$Loss = \min_{x_i^p} \max_{x_i^n} \sum_i^N [\|x_i^a - x_i^p\|_2^2 - \|x_i^a - x_i^n\|_2^2 - \alpha]_+$$

This method make the loss function more hardcore than the function for recognition. However, searching for the largest distance for positive image and lowest distance for negative image is time-consuming, especially when the dataset is large. To optimize the data selection, instead of search for the whole dataset, a semi-hardcore method is implemented in this method. Considering our method is

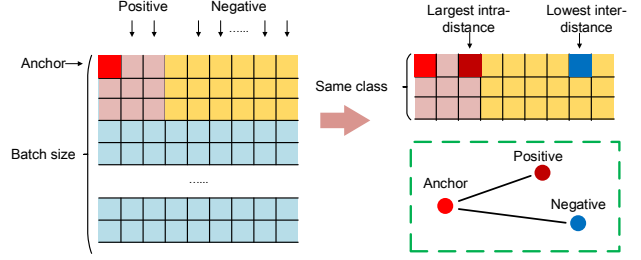


Figure 2: Visualization of semi-hardcore data selection for triplet loss function

use in neural network, each time we send data into the network, we cluster the data via classes, which means, images in the same class will be gathered together. In this case, we can calculate the distance for each two images in one batch, then rank the distance to achieve the best selection in one batch. Figure 3 clarifies the process of optimization.

## 4 Experimental Results

### 4.1 Dataset

In the experiment, we implement the neural network in Caltech 101 dataset [2]. The number of classes is 101, about 40 to 800 images per category. Most categories have about 50 images. Collection was finished in September 2003 by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato. The size of each image is roughly 300 x 200 pixels. For training period, 80% of images in one class are selected as training set and 20% of images are picked as testing set. For the baseline, we select Alexnet invented by Alex et.al [3] for classification task. We extract features from the last dropout layer and send it to both fully-connection layer for softmax-entropy loss, triplet loss and semi-hardcore triplet loss as supervising signal.

### 4.2 Data organization & Experiment Environment

Feat Xiwen Li for data organization and Experiment Environment.

(P.S. Here the experiment environment should be noticed by Jiahao)

(P.P.S Check out the data partition in section Dataset. I forget the percentage of partition for training and testing. Thanks!)

### 4.3 Comparison

Feat Jiahao Li for experiment comparison

(P.S. A table for comparison is good for visualization. The result of improvement can be find in section Algorithm. It's ok to copy these part for comparison. Thanks!)

## 5 Conclusion

We exert a classic metric learning method, triplet loss, into classification task and do some optimizations specially for classification. Owing to the property of classification and neural network training, semi-hardcore model for data selection is used for this task.

## Acknowledgments

Describe what sources you had help from, and so on. But try to put as much as possible as direct references in the body of the paper (like to say, “adapting the code provided by the authors of [?]).

## References

- [1] A. Delière, A. Cioppa, and M. Van Droogenbroeck. Hitnet: a neural network with capsules embedded in a hit-or-miss layer, extended with hybrid data augmentation and ghost capsules. *arXiv preprint arXiv:1806.06519*, 2018.
- [2] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems* 25, pages 1097–1105. Curran Associates, Inc., 2012.
- [4] H. B. Lee, J. Lee, S. Kim, E. Yang, and S. J. Hwang. Drop-max: Adaptive variational softmax. In *Advances in Neural Information Processing Systems*, pages 925–935, 2018.
- [5] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.