17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

Angular Margin Constrained Loss for Automatic Liver Fibrosis Staging

Katsuhiro Nakai
Yamaguchi University
Yamaguchi, Japan
a030vbu@yamaguchi-u.ac.jp

Xu Qiao
ShanDong University
JiNan, China
qiaoxu@sdu.edu.cn

Xian-Hua Han
Yamaguchi University
Yamaguchi, Japan
hanxhua@yamaguchi-u.ac.jp

Abstract

Automatic progression staging of liver fibrosis plays very important roles in the direct treatment and the evaluation of prognosis. In clinical site, liver biopsy is popularly used as the gold standard method of liver fibrosis staging, and has obvious drawbacks such as sampling error, heavy burden to patients and high inter-observer variability. Recently, non-invasive techniques as a diagnostic standard have attracted extensive attention. This study exploits a novel deep learning-based liver fibrosis staging framework using non-invasive MRI images. Since there exist large variance in both texture and shape of MRI liver images between patients and subtle distinctness among the progression stages of liver fibrosis, it is a challenge task for accurate progression staging of liver fibrosis.To enhance the discriminative power among the fibrosis stages with subtle difference, this study proposes to integrate angular margin penalty into the conventional softmax loss of the deep learning network, which is expected to enforce extra intra-class compactness and inter-class discrepancy simultaneously. Specifically, we explore the angular margin constrained loss in several classification neural network models such as VGG16, ResNet18, and ResNet50, and further incorporate the between-stage similarity of the training procedure to adaptively adjust the margin for boosting liver fibrosis classification performance. Experiments on the MRI image dataset provided by Shandong University, which includes three progression stages of liver fibrosis: early, middle and last stages, validate that the performance gain with the integration of the angular margin penalty are from 3% to 7% compared to the baseline models: VGG16, ResNet18, and ResNet50.

1 Introduction

Any chronic liver disease and injury can lead to the development of liver fibrosis, and without proper treatment may proceed to severe cirrhosis, which greatly increases the risk of the evolution into hepatocellular carcinoma and liver failure [1]. Thus, monitoring the progression state of liver fibrosis is of great clinical importance to make proper decision of therapeutic plan for prevention of fibrosis evolution or even reversing fibrosis [2]. In clinical site, liver biopsy [3] is popularly used as the gold standard method of liver fibrosis staging, and has many shortcomings such as sampling error, heavy pain and financial burden to the patients, and high inter-observer variability.

Recently, as an alternative tool of the liver biopsy for liver fibrosis staging, various non-invasive methods [4, 5, 6] have been investigated, such as biomarker evaluation and liver morphological analysis by ultrasound or magnetic resonance. Most existing non-invasive methods [7] for liver disease diagnosis are generally based on Ultrasonographic (US) and magnetic resonance (MR) imaging elastographic techniques, and the diagnosis is manually conducted by medical experts and radiologists, which would lead to the subjective result only and large inter-observer variability. Mojsilovic et al. [8] proposed to model texture feature of b-scan liver images, and quantitatively characterized visually-similar liver diseases while Yeh et al. [9] explored gray level concurrence and the texture feature based on non-separable wavelet transform for classifying the liver fibrosis status. With the remarkable performance boost of the deep convolutional neural network (DCNN) on various computer vision tasks such as image classification, object detection and segmentation, DCNN has been applied for Liver fibrosis diagnosis. Meng et al. [10] exploited transfer learning based on VGGNet [11] and a fully-connected network as the classifier for liver fibrosis using ultrasound images while Liu et al. [12] proposed a multi-indicator guided DNN using multiple ultrasound images, which is also based on the VGGNet architecture. Later, Hectors et al. [13] further exploited VGGNet-based transfer learning for deep analysis of gadoxetic acid–enhanced MRI to fully automated prediction of liver fibrosis. Although the the diagnosis potential of the liver fibrosis using DCNN has been verified, the prediction is far from the sufficient performance via directly using the developed network architecture for general image recognition without taking account of the characteristics in medical data.

This study exploits a novel discrimination-enhanced deep learning network for liver fibrosis staging using non-invasive MRI images. Due to the large variance in both texture and shape of MRI liver images between patients and subtle distinctness among the progression stages of liver fibrosis, it is a challenge task for accurate progression staging of liver cirrhosis via simply minimizing the traditional softmax loss of the DCNN-based classifiers, which is verified being insuffi-

2021 17th International Conference on Machine Vision and Applications (MVA) | 978-4-901122-20-7/21/$31.00 ©2021 10.23919/MVA51890.2021.9511356

cient to capture the discriminating power for classification [14, 15]. To enhance the capability of better discriminating capturing, several variants have been proposed to adopt multi-loss learning [16, 17, 18, 19, 20] via integrating auxiliary losses into the traditional softmax loss. Therein, angular margin has been widely explored due to its intrinsic consistency of angle's cosine measure with softmax, and illustrated remarkable performance gain and explainable interpretation for face recognition tasks. This study leverages the angular margin to enhance the discriminatingly learning power of the DCNN model for liver fibrosis staging, which is expected to enforce extra intra-class compactness and inter-class discrepancy simultaneously. Moreover, we further incorporate the between-stage similarity degree in the ongoing training procedure as a dynamical fine-tuned margin penalty to guide the network learning into the proper direction. Specifically, we implement the angular margin constrained loss in several classification neural network models such as VGG16, ResNet18, and ResNet50, and conduct live fibrosis staging on an MRI image dataset provided by Shandong University, which includes three progression stages of liver fibrosis: early, middle and last stages. Extensive experiments illustrate that the performance gain with the integration of the angular margin penalty are from 3% to 7% compared to the baseline models: VGG16, ResNet18, and ResNet50.

2 Proposed Approach

This study investigates various deep CNN architectures for liver fibrosis staging using MRI images. To encourage discriminating feature learning, we exploit an angular margin constrained loss (AMCL) instead of the conventional Softmax loss, and further integrate the between-class similarity of the class-corresponding normalized vectors to fine-turn the angular margin for regularizing network training procedure toward the proper direction. The overview of the proposed liver fibrosis staging network with AMCL is illustrated in Fig. 1, where the cropped rough liver region from the raw MRI volume instead of finer segmented liver is used as the network input. Given the $i-th$ cropped liver region, belonging to y_i-th class, we can use any DCNN backbone architecture such as VGG and Resnet to extract d-dimensional deep feature $\boldsymbol{x}_i \in \mathbb{R}^d$, and the most widely used classification loss function is softmax loss, which is expressed as:

$$L_{SM} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{f_{y_i}(\boldsymbol{x_i})}}{\sum_{c=1}^{C} e^{f_c(\boldsymbol{x}_i)}} \\ f_{y_i}(\boldsymbol{x}_i) = \boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i + b_{y_i}, f_c(\boldsymbol{x}_i) = \boldsymbol{W}_c^{\mathrm{T}}\boldsymbol{x}_i + b_c \tag{1}$$

where $\boldsymbol{W}_c \in \mathbb{R}^d$ denotes the $c-th$ column of the class weight $\boldsymbol{W} \in \mathbb{R}^{d\times C}$ and b_c is the bias term for computing the $c-th$ class score $f_c(\boldsymbol{x}_i)$ of the deep feature $\boldsymbol{x}_i$. N and C are the batch size and the class number, respectively. Then deep classification problem attempts to optimize the network parameters to minimize the softmax loss: L_{SM}. However, the softmax loss function is insufficient to explicitly enforce the deep feature embedding with higher intra-class similarity and large inter-class diversity, which unavoidably results in performance drop especially for liver fibrosis staging task under large intra-class texture and shape variations among the same stages' MRI images of different individuals and subtle distinctiveness among different stages' images of the same patient. Therefore, this study exploits an angular margin constrained loss (AMCL) function to enhance the discriminating feature learning power, and further integrates dynamic regularization on AMCL using the between class similarity of the class weights in the ongoing training procedure.

2.1 Angular margin constrained loss

As formulated in Eq. 1, the $c-th$ class score f_c of the $i-th$ deep feature can be computed as: $\boldsymbol{W}_c^{\mathrm{T}}\boldsymbol{x}_i = \|W_c\|\|\boldsymbol{x}_i\|\cos\theta_{c,i}$ via fixing the bias $b_c = 0$ for simplicity, where $\theta_{c,i}$ is the angle between the $c-th$ class weight $\boldsymbol{W}_c$ and the feature $\boldsymbol{x}_i$. Following [19, 20, 21], we fix the individual class weight $\|\boldsymbol{W}_c\| = 1$ and the embedding deep feature $\|\boldsymbol{x}_i\| = s$ by l_2 normalization. Thus, the l_2 normalization makes the learned embedding features distributing on a hypersphere with a radius of s, and the softmax loss in Eq. 1 can be reformulated according to the cosine value of the angle $\theta_{c,i}$ (ASM) as

$$L_{ASM} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i,i})}}{\sum_{c=1}^{C} e^{s\cos(\theta_{c,i})}} \tag{2}$$

From the Eq. 2, it can be seen that the probability of the $i-th$ sample to the ground-truth y_i class should be higher than other classes given $\theta_{y_i,i} < \theta_{c,i}$ (or $\cos\theta_{y_i,i} > \cos\theta_{c,i}$). However, this lax loss function cannot enforce the embedding feature to maximally close to the ground-truth class weight and far from the other class weights, and thus may result in some embedding features in the margin regions and insufficient discriminating capability. To enhance the intra-class compactness and inter-class discrepancy, we add an angular margin penalty inside the cosine function to impose that the angle between the embedding feature and the ground-truth class weight should be more than margin value smaller than other classes. The AMCL is then expressed as

$$L_{AMCL} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i,i}+m)}}{e^{s\cos(\theta_{y_i,i}+m)} + \sum_{c=1,c\neq y_i}^{C} e^{s\cos(\theta_{c,i})}} \tag{3}$$

where m is the fixed angular margin and needs to be empirically decided while s is scalar value for adjusting

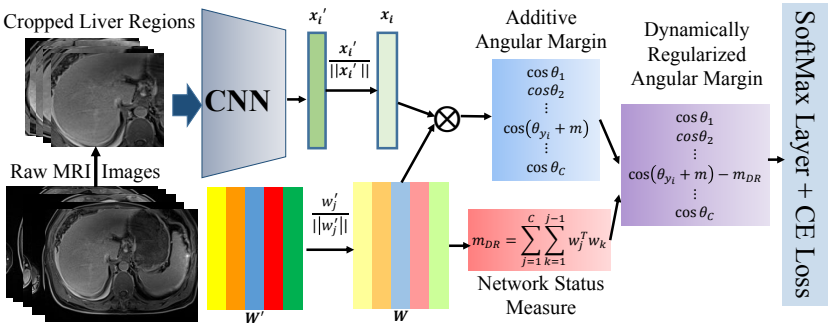

Figure 1. **The overview of the proposed AMCL framework.** With a CNN backbone network, we can extract the corresponding deep feature $\boldsymbol{x}$ following l_2 normalization. The cosine of angle between the deep feature and the class vector is utilized to formulate the softmax loss. To encourage discriminating feature learning, the angular margin constrained loss (AMCL) and its dynamic regularization version (DR-AMCL) is exploited in the right part.

the exponential function to give plausible probability distribution. In the next subsection, we will introduce the probability function evolution with the angular under various angular margins m and the scalar s, and then attempt to obtain the valid values of m and s for our liver fibrosis staging task. As we know that the network parameters are updated according to the computed loss of the current batch samples, and thus may result in unstable network learning with the sample change in each training step. While the learned class weights $\boldsymbol{W}_c$ in the ongoing training procedure can provide an indicator of the network status, and the between class similarities of the weight vector of different classes possibly measure the discriminating capability of the current network. Thus, we adaptively exploit the between-class similarities as a dynamical margin to regularize the network training toward the proper direction, dubbed as dynamically regularized AMCL (DR-AMCL). The dynamical margin (DM) using the class vector is calculated as

$$m_{DM} = \sum_{j=1}^{C}\sum_{k=1}^{j-1} \boldsymbol{W}_j^{\mathrm{T}} \boldsymbol{W}_k \tag{4}$$

The schematic illustration of the proposed DR-AMCL is given in the right part of Fig. 1, and can then be formulated as:

$$L_{DR} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\cos(\theta_{y_i,i}+m)} - m_{DM}}{e^{s\cos(\theta_{y_i,i}+m)-m_{DM}} + \sum_{c=1,c\neq y_i}^{C} e^{s\cos(\theta_{c,i})}} \tag{5}$$

2.2 Effect of the hyper-parameter: s and m

The goal of the classification network training is to get the plausible high probability $P_{y_i,i}$ with enough small angle $\theta_{y_i,i}$ regardless to the hyper-parameter values. It is obvious that $P_{y_i,i}$ has the value range [0, 1] while $\theta_{y_i,i}$ corresponds to the range $[\frac{\pi}{2}, 0]$. As verified in [22], The angles $\theta_{c,i}$ for non-corresponding classes $c \neq y_i$ during the training procedure almost remain around $\frac{\pi}{2}$, and thus $\sum_{c=1,c\neq y_i}^{C} e^{s\cos(\theta_{c,i})}$ in Eq. 5 can be simplified to $\sum_{c=1,c\neq y_i}^{C} e^{s\cos(\frac{\pi}{2})} = C-1$ as a constant. With this empirical insight, we validate the evolution of the corresponding class probability $P_{y_i,i}$ $w.r.t$ $\theta_{y_i,i}$ under various hyper-parameter values s and m. For our liver fibrosis staging task ($C = 3$), the curves of $P_{y_i,i}$ $w.r.t$ $\theta_{y_i,i}$ under different s and m are plotted in Fig. 2, which manifests the influence on the $P_{y_i,i}$ especially by the scale parameter s. For network training, we prefer that the training samples even with high confidence on the corresponding class still can penalize the classification results for network updating, and thus we can see from the curve that s with values from 20 to 30 while m with values from 0.4 to 0.6 have this satisfactory properties. In our experiments, we fix s as 30 and change m from 0.4 to 0.6 with 0.05 interval for verifying the liver staging performances.

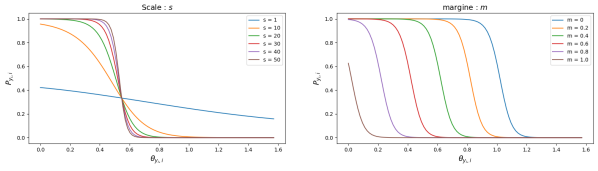

Figure 2. The evolution of $P_{y_i,i}$ $w.s.t$ $\theta_{y_i,i}$ under various scales s and margins m for our liver fibrosis staging task ($C = 3$). **Left**: The probability $P_{y_i,i}$ with various scales s and the fixed margin $m = 0.5$. **Right**: The probability $P_{y_i,i}$ with various margins m and the fixed scale $s = 30$. .

3 Experimental results

3.1 Experimental setup

Dataset: We validate the liver staging performance of our proposed (DR-) AMCL framework using the

Table 1. The compared overall accuracy and mis-staging rates using different backbone architectures w/o the AMCL or DR-AMCL constraints.

Methods	Acc	MR: A→C	MR: C→A
VGG16	0.6461	**0.0092**	**0.1711**
VGG16+AMCL	**0.6813**	0.0474	0.2201
VGG16+DR-AMCL	0.6812	0.0189	0.2197
ResNet18	0.6658	0.0207	0.0617
ResNet18+AMCL	0.7175	0.0194	0.0199
ResNet18+DR-AMCL	**0.7196**	**0.0097**	**0.0179**
ResNet50	0.6514	0.0496	0.0369
ResNet50+AMCL	0.7277	0.0340	**0.0087**
ResNet50+DR-AMCL	**0.7302**	**0.0094**	0.0512

MRI image dataset provided by Shandong University, China. This dataset consists of 117 patients MRI volumes with three liver fibrosis stages: early (A), middle (B) and last (C), and each is captured in a short period of time from an individual patient. From the 117 patients MRI volumes, we roughly cropped 2805 liver regions including 934 A-state images, 1269 B-stage images, and 602 C-stage images, without fine segmentation as shown in the left part of Fig. 1, where 80% of them were used for training, and the remaining images from different volumes with the training set were for testing.

Detail implementation: We adopted VGG16, ResNet18, and ResNet50 as the backbone network architectures, and Adam was used as the optimization method. The network was trained with 100 epochs and the used learning rate: 0.00005, To avoid variation of staging accuracy due to the division of the training and testing subsets, we conducted 6 runs of experiments with randomly selected training and testing subsets, and calculated the average values of 6 runs as the final evaluation for comparison. Two evaluation metrics including the overall accuracy of all stages and the F-score for each stage are used in performance comparison. Moreover, since the mis-staging from A to C or C to A is eager to be avoid in the clinic site, we further compute the mis-staging rate (MR: A→C, MR: C→A) for verifying our proposed liver fibrosis staging method.

3.2 Experimental result

Table 1 manifests the compared overall accuracy of the liver fibrosis staging, and the mis-staging rates from A to C or C to A using different backbone architectures w/o AMCL or DR-AMCL. It can observe from Table 1 that the integration of the AMCL into all network architectures can greatly boost the staging performance of liver fibrosis, and further performance gains on the ResNet 18 and ResNet50 has been obtained with the dynamical regularization on AMCL (DR-AMCL). Moreover, the mis-staging rates (A→C and C→A) using ResNets are also decreased with AMCL and DR-AMCL methods. Table 2 gives the compared F-scores of all stages (A, B and C), and validates that the integration of the AMCL and DR-AMCL into the ResNet architectures can greatly improve the F-scores of the severe liver fibrosis samples, which are especially important to get the precise staging scores for these severe patients in clinic site.

Table 2. The compared F-score of each liver fibrosis stage using different backbone architectures w/o the AMCL or DR-AMCL constraints.

Methods	F: A	F: B	F: C
VGG16	0.8060	0.7081	0.1510
VGG16+AMCL	0.7949	**0.7106**	**0.3138**
VGG16+DR-AMCL	**0.8138**	0.6911	0.2602
ResNet18	0.8598	0.6517	0.3439
ResNet18+AMCL	**0.8645**	0.6812	0.4837
ResNet18+DR-AMCL	0.8477	**0.7087**	**0.5118**
ResNet50	**0.8644**	0.6238	0.3578
ResNet50+AMCL	0.8641	0.7133	0.4574
ResNet50+DR-AMCL	0.8336	**0.7160**	**0.5506**

Finally, the compared accuracies using the ResNet50 backbone architecture with different angular margins m are shown in Table 3. It can be seen that the AMCL provides better performance with small margins while the proposed DR-AMCL achieves better performance with large margins. We are going to exploit the optimal margins in different deep backbone architectures for the liver fibrosis staging task.

Table 3. The compared accuracy on the ResNet50 architecture with different angular margins m

m	0.40	0.45	0.50	0.55	0.60
AMCL	**0.7206**	**0.7277**	0.7142	0.6917	0.7048
DR-AMCL	0.7180	0.7180	**0.7250**	**0.6991**	**0.7302**

4 Conclusion

This study proposed a novel deep learning-based liver fibrosis staging framework using non-invasive MRI images. To enforce high discriminating capability, we exploited a angular margin constrained loss (AMCL) to train more generalized deep models. Moreover, to dynamically adjust the network training procedure toward proper direction, we investigated the between-class similarity, which inherently evaluates the network status, to regularize the predefined angular margin, and then formulated a dynamically regularized AMCL (DR-AMCL) for liver fibrosis staging. Experiments on an MRI dataset validated our proposed deep methods with the AMCL and DR-AMCL outperformed the baseline models: VGG16, ResNet18, and ResNet50.

Acknowledgment: This research was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20K11867, and JSPS KAKENHI Grant Number JP12345678.

References

[1] Scott L FriedmanL: "Liver fibrosis – from bench to bedside," *J Hepatol*, 38(Suppl 1): S38–S53, 2003.

[2] Detlef Schuppan, Muhammad Ashfaq-Khan, Ai Ting Yang, Yong Ook Kim: "Liver fibrosis: direct antifibrotic agents and targeted therapies," *Matrix Biol.*, 68-69: 435–451, 2018.

[3] A. Dohan, Y. Guerrache, M. Boudiaf, J.-P. Gavini, R. Kaci, and P. Soyer: "Transjugular liver biopsy: indications, technique and results," *Diagnostic and interventional imaging*, Vol. 95, No. 1, pp: 11-15, 2014.

[4] Laurent Sandrin, Bertrand Fourquet, Jean-Michel Hasquenoph et al.: "Transient elastography: a new noninvasive method for assessment of hepatic fibrosis," *Ultrasound Med. Biol.*, Vol. 29, No. 12, pp: 1705-1713, 2003.

[5] Roberta Catania, Alessandro Furlan, Andrew D Smith et al.: "Diagnostic value of MRI-derived liver surface nodularity score for the non-invasive quantification of hepatic fibrosis in non-alcoholic fatty liver disease," *Eur Radiol.*, Vol. 31, No. 1, pp. 356-263, 2020.

[6] Jinning Li J, Huanhuan Liu, Caiyuan Zhang et al.: "Native T1 mapping compared to ultrasound elastography for staging and monitoring liver fibrosis: an animal study of repeatability, reproducibility, and accuracy," *Eur Radiol.*, Vol. 30, No. 1, pp. 337-345, 2020.

[7] Paul Kennedy, Mathilde Wagner, Laurent Castera et al.: "Quantitative Elastography Methods in Liver Disease: Current Evidence and Future Directions," *Radiology*, Vol. 286, No. 3, 2018.

[8] Aleksandra Mojsilovic, Miodrag Popovic, Srdjan Markovic et al.: "Characterization of visually similar diffuse diseases from b-scan liver images using nonseparable wavelet transform," *IEEE Transactions on Medical Imaging*, Vol. 17, No. 4, pp. 541-549, 1998.

[9] Wen-Chun Yeh, Sheng-Wen Huang, and Pai-Chi Li et al.: "Liver
fibrosis grade classification with b-mode ultrasound," *Ultrasound in medicine biology*, Vol. 29, No. 9, pp. 1229-1235, 2003.

[10] Dan Meng, Libo Zhang, Guitao Cao, Wenming Cao, Guixu Zhang, and Bing Hu: "Liver fibrosis classification based on transfer learning and fcnet for ultrasound images," *Ieee Access*, Vol. 5, pp. 5804-5810, 2017.

[11] Karen Simonyan and Andrew Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition," *International Conference on Learning Representations*, 2015.

[12] Jiali Liu, Wenxuan Wang, Tianyao Guan et al.: "Ultrasound Liver Fibrosis Diagnosis using Multi-indicator guided Deep Neural Networks," *International Workshop on Machine Learning in Medical Imaging*, pp. 230-237, 2019.

[13] Stefanie J. Hectors, Paul Kennedy, Kuang-Han Huang et al.: "Fully automated prediction of liver fibrosis using deep learning analysis of gadoxetic acid–enhanced MRI," *European Radiology*, pp. 230-237, 2020.

[14] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song: "SphereFace: Deep Hypersphere Embedding for Face Recognition," *In Conference on Computer Vision and Pattern Recognition (CVPR),* 2017.

[15] W. Liu, Y. Wen, Z. Yu, and M. Yang: "Large-Margin SoftmaxLoss for Convolutional Neural Networks," *In International Conference on Machine Learning (ICML)*, 2016.

[16] Y. Wen, K. Zhang, Z. Li, and Y. Qiao: "A discriminative fea-ture learning approach for deep face recognition," *In European Conference on Computer Vision (ECCV)*, pp. 499–515, 2016.

[17] E. Hoffer and N. Ailon: "Deep metric learning using triplet network," *In International Workshop on Similarity-Based Pattern Recognition*, 2015.

[18] S. Chopra, R. Hadsell, and Y. LeCun: "Learning a similarity metric discriminatively, with application to face verification," *In Conference on Computer Vision and Pattern Recognition(CVPR)*, 2005.

[19] Hao Wang, Yitong Wang, Zheng Zhou et al.: "CosFace: Large Margin Cosine Loss for Deep Face Recognition," *In Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018.

[20] Jiankang Deng, Jia Guo and Niannan Xue: "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *In Conference on Computer Vision and Pattern Recognition(CVPR)*, 2019.

[21] F. Wang, X. Xiang, J. Cheng, and A. L. Yuille: "Normface:l2hypersphere embedding for face verification," *arXiv: 1704.06369*, 2017

[22] Xiao Zhang, Rui Zhao, Yu Qiao, Xiaogang Wang, Hongsheng Li: "AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations," *In Conference on Computer Vision and Pattern Recognition(CVPR)*, 2019.