CENG 506 DEEP LEARNING

The Investigation on Cosine Derived Loss Functions for Deep Learning Domains

The Curiosity

Burak Topçu 283078027

Güliz Akkaya 283078008

Overview

- Introduction
- Baseline Paper
- Approach
- Experiments
- Conclusion
- References

Introduction

- Deep Convolutional Neural Networks is one of the deep learning approaches that are used for various tasks such as face recognition.
- Appropriate design of DCNNs is significant to provide efficiency in terms of accuracy performance and computational workload.
- Many studies [1][2][3] are performed to enhance DCNN architectures that focus on the design of loss functions.

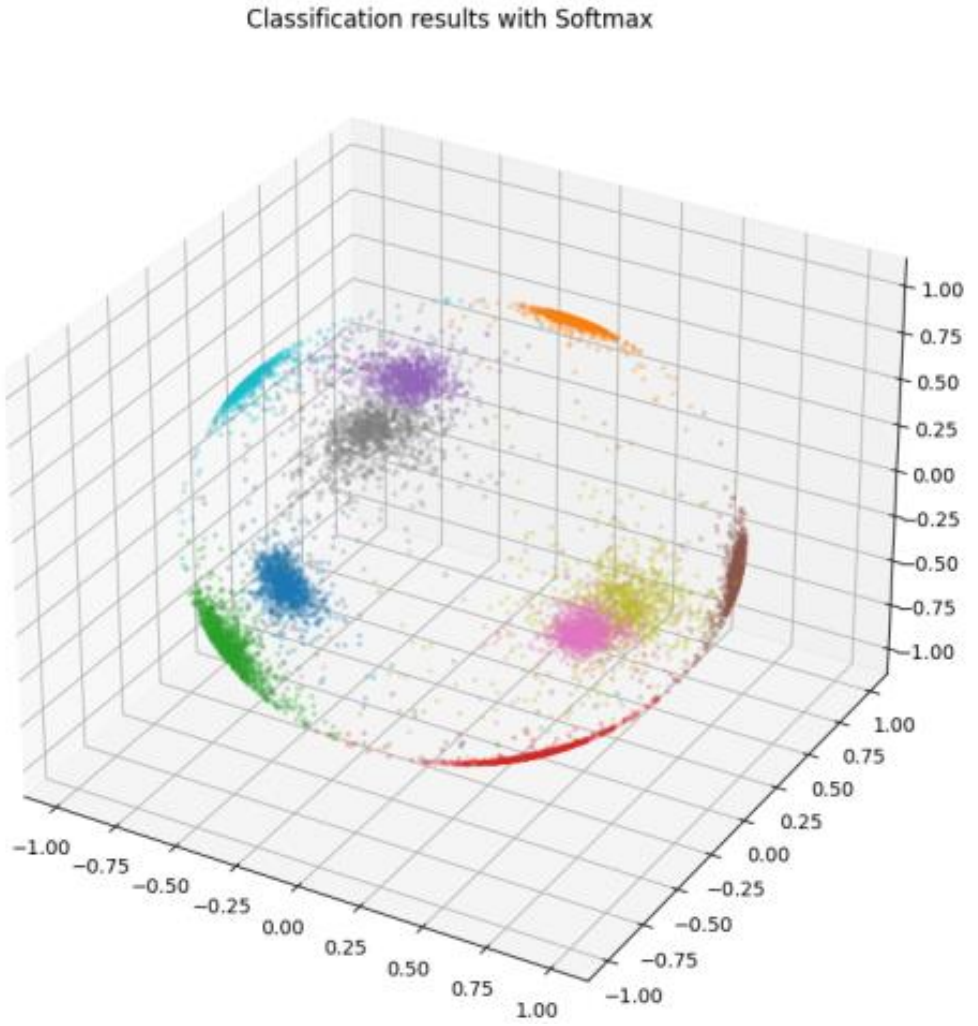
[1] ArcFace, 2019

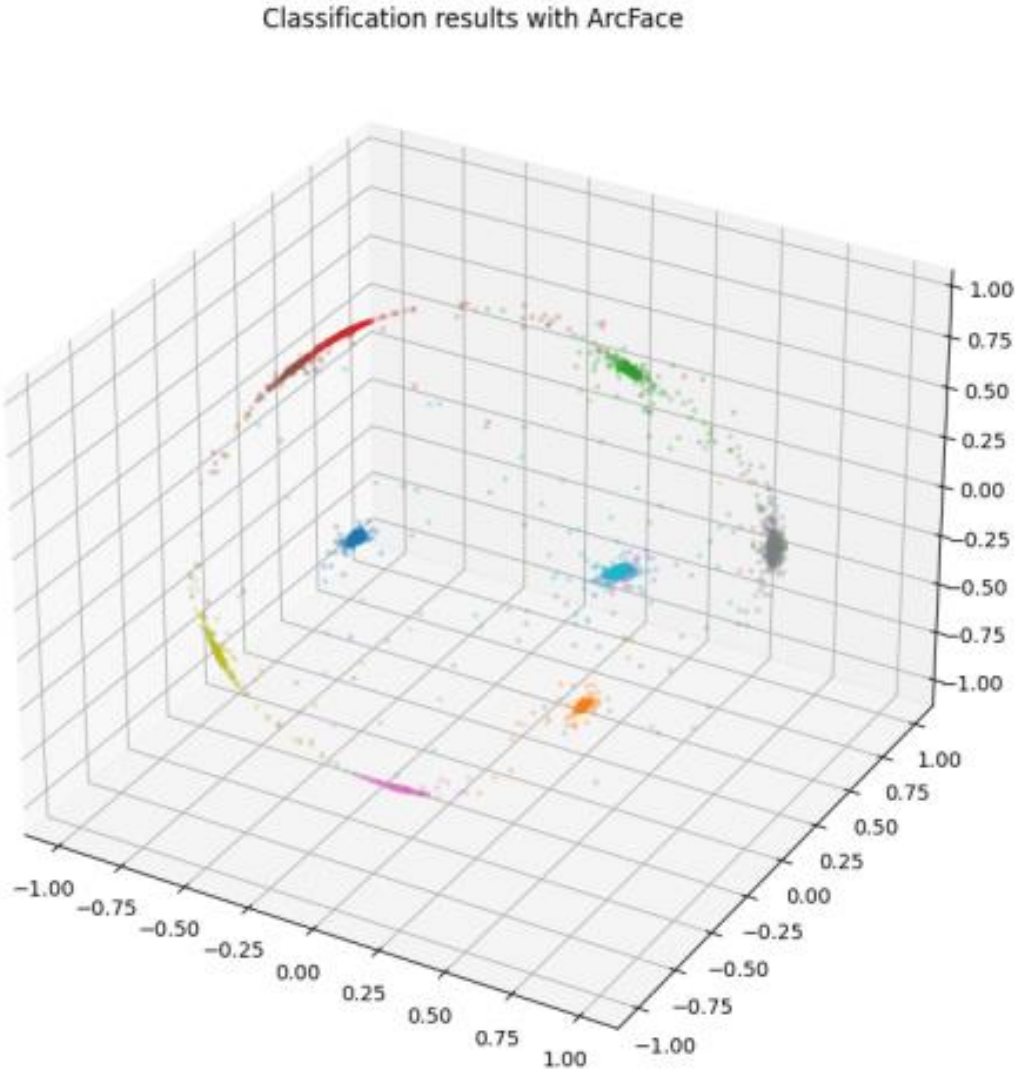
[2] CosFace, 2018

[3] SphereFace, 2017

Baseline Paper

- ArcFace: Additive Angular Margin Loss for Deep Face Recognition [1]
- Additive angular margin to enhance traditional softmax loss function
  - ✓ Higher intra-class compactness
  - ✓ Higher inter-class discrepancy
  - ✓Stabilized training process for face recognition


Classification results with Softmax
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


Classification results with ArcFace
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Approach

- From starting the $W_j^T * x_i^n = \|\mathbf{W_j}\| * \|\mathbf{x_i}\| * \cos(\theta_j)$ and $-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}$ the below functions are derived:

ArcFace Loss [1]

$$-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}$$

Combined (ArcFace[1] + Cosine[2] + SphereFace [3]) Loss

$$-\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(m_1\theta_{y_i}+m_2)-m_3)}}{e^{s(\cos(m_1\theta_{y_i}+m_2)-m_3)} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}$$

Experiments

- Architectural models in the experiments are built as:
  - model = cnn_learner(dls, resnet18, metrics=accuracy, loss_func = CrossEntropyLossFlat())
- And loss_funct are ArcFace, combined (ArcFace + Cosine + SphereFace) and cross entropy losses.
- In the first submission phase, different hyper parameters are trained and the optimals are found as m1 = 1, m2 = 0.4 and m3 = 0.2.
- MNIST (10 class, 60000 samples, 28*28 hand-written digits), CIFAR10(10 class, 60000 samples, 32*32 real world images) and Caltech-UCDS-Birds 2011 (200 class, 11 788 samples) datasets are used for the experiments.

Experiments

First Submission Results

Accuracy Results			
ArcFace Method		Combined Method ArcFace+SphereFace+ CosFace m2+(m1=1)+m3	
m = 0.4	98.84	**m2 = 0.4 m3 = 0.2**	**98.97**
m = 0.41	98.75	m2 = 0.41 m3 = 0.2	95.93
m = 0.42	98.66	m2 = 0.42 m3 = 0.2	98.80
m = 0.43	98.77	m2 = 0.43 m3 = 0.2	98.89
m = 0.44	98.77	m2 = 0.44 m3 = 0.2	98.82
m = 0.45	98.70	m2 = 0.4 m3 = 0.25	98.78
m = 0.46	98.85	m2 = 0.41 m3 = 0.25	98.69
m = 0.47	98.71	m2 = 0.42 m3 = 0.25	98.83
m = 0.48	98.79	m2 = 0.43 m3 = 0.25	98.80
m = 0.49	98.66	m2 = 0.44 m3 = 0.25	98.88
m = 0.50	**98.93**	m2 = 0.4 m3 = 0.3	98.84
m = 0.51	98.84	m2 = 0.41 m3 = 0.3	97.73
m = 0.52	98.65	m2 = 0.42 m3 = 0.3	98.77
m = 0.53	98.70	m2 = 0.43 m3 = 0.3	98.69
m = 0.54	98.87	m2 = 0.44 m3 = 0.3	98.81
m = 0.55	98.76	m2 = 0.4 m3 = 0.35	98.73
m = 0.56	97.58	m2 = 0.41 m3 = 0.35	98.71
m = 0.57	98.86	m2 = 0.42 m3 = 0.35	98.71
m = 0.58	98.83	**m2 = 0.43 m3 = 0.35**	**98.95**
m = 0.59	96.01	m2 = 0.44 m3 = 0.35	96.83

Second Submission Results

Mean Accuracy Results after 6 trainings

Mean Accuracy Results			
Datasets	ArcFace Loss	Combined Loss	CE Loss
MNIST	99.21	99.24	**99.29**
CIFAR-10	73.46	73.59	**78.16**
CUB-200-2011	66.36	66.51	**75.24**

For 6 trainings, standard error values

Standard Error Values			
Datasets	ArcFace Loss	Combined Loss	CE Loss
MNIST	0.0194	**0.0126**	0.0130
CIFAR-10	0.2343	**0.1050**	0.0778
CUB-200-2011	0.4573	**0.2678**	0.3609

Conclusion

- We investigated the effects of ArcFace loss, combined loss and cross entropy & softmax loss functions by performing different experiments on 3 different datasets
- With respect to our findings, these cosine based loss functions can not be generalized to be used in other deep learning domains since the results for cross entropy and softmax loss converges in earlier epochs and cosine based loss functions do not result in as high accuracy as in cross entropy and softmax