DYNAMIC GAUSSIAN LOSS FUNCTION-FACE RECOGNITION

**Jenish Savaliya1,Meenakshi S.2, Rakesh K.3, Dr. S. Prabha4**

*1UG Scholar, Dept. of ECE, Hindustan Institute of Technology and Science*

*Chennai, Tamil Nadu, India*

*4Associate Professor, Dept. of ECE, Hindustan Institute of Technology and Science*

*Chennai, Tamil Nadu, India*

[*119121014@student.hindustanuniv.ac.in*](mailto:119121014@student.hindustanuniv.ac.in)*,* [*219121012@student.hindustanuniv.ac.in*](mailto:219121012@student.hindustanuniv.ac.in)*,* [*319121016@student.hindustanuniv.ac.in*](mailto:319121016@student.hindustanuniv.ac.in)*,* [*4sprabha@hindustanuniv.ac.in*](mailto:4sprabha@hindustanuniv.ac.in)

*ABSTRACT* ---**Key features of the human face are highly distinctive in nature. For accurate recognition of key features facial recognition models are supervised and trained with complex neural network architecture. To achieve an error value as negligible as possible, Suitable algorithms, data sets and weights are chosen. Recently developed and tested algorithms follow a process of fixing penalty margins to reduce errors. In real time facial recognition models these penalty margins are introduced as “Loss functions”. Choosing appropriate loss functions assist in the prediction of weights valuefor error reduction. Softmax Loss is the addition of the probability of single node output and Cross entropy error. CosFace Loss works on the basis of large margin cosine loss. It unintentionally maximised intra class variance and minimized inter class variance. Arcface Loss is highly similar to Spherephase and CosFace loss functions. It worked as an inverse principle of cosface to get the degree of differentiation between the predicted vs the actual results. Adaptive Margin Softmax loss selects intraclass samples adaptively**

**These trailblazing loss functions made the model extremely rigid in nature since the penalty margin was fixed.In this paper, elastic face model is introduced in which the inter-classes separation boundary is made flexible to give scrutinous attention to the inter class than the intra class variations. The penalty margin for the elastic face is Elastic penalty margin loss. Elastic-face shows 98% accuracy in LFW datasets and is fore-fronting than other loss functions.**

# Keywords --- Elastic penalty margin loss, Arcface Loss, Softmax loss, CosFace loss, Neural Networks, Face recognition.

# INTRODUCTION

The current upsurging world is filled with security and privacy concerns. Recognition of different types of data is one of the most viable and feasible security options. Most European countries use retinal eye

scan to check the accurate identity of the user at the ATM’s. Image Recognition techniques are used for comparing of the new image with the already existing images. Facial recognition is the cutting-edge application of image recognition. Facial recognition models are trained with data sets that contains quantitative real time images. The Real time Images are collected, pre-processed, extracted for the needed features and later separated into classes. These classes are different individuals with different identities. The collected 3D facial images are converted into simple feature vectors and is compared with the desired image for feature similarities. The entire process works on a hypothesis that real time images are highly balanced in nature, meaning manually set margin is sufficient for encapsulating each of the intra class variations. But real time images are highly unbalanced in nature which means it is highly complex to fit all the available samples with its own class. This maximizes the intra class separability than the inter class separability. To overcome this problem Softmax losses were introduced. Large scale Softmax losses reduced the overfitting of the intraclass features and encouraged in the interclass separability by adjusting the fixed margin parameter, m. Spherephase Softmax losses is an advancement of Softmax function by restricting the environment to a hypersphere where the face is mapped to their respective unique feature space. The maximal intra class distance from the centroid is expected to be smaller than the minimal inter class distance. The Softmax losses focused in the reduction of intra class variance but failed to address to sub optimal inter class variance attained. Cosface Loss addressed this from a different perspective. The distinctive facial features were projected to an angular space instead of the normal Euclidean space. Angular margin was introduced to show larger interclass variance.

1. LITERATURE REVIEW

As security systems grew and with the introduction to mobile phones facial recognition has been in the forefront of technology development. The initial methods of facial recognition utilized classifiers manually encoded data to search for similarities in facial features amongst the given samples. These algorithms are called Haar Cascades models such as Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods by T. Mantoro et al[1] and Face detection and recognition in real-time photos with haar cascade and local binary pattern histogram for automatic door locking system by Ian Haikal Amir A el at[2] utilize Haar Cascades for facial recognition. Although Haar Cascade models are faster and require less processing power compared to modern neural networks, they lack accuracy and tend to have large false positives.

The use of Principal Component Analysis for facial recognition was proposed in Face recognition and facial expression identification using PCA by S. S. Mehe el at[3] and in PCA based Facial Recognition for Attendance System by T.A Kiran el at [4]. Facial recognition is accomplished by projecting an image into a subspace covered by Eigenfaces followed by comparing its position in the face sphere of the given classes. Poor intraclass separation and large estimation are common issues that occur in the PCA facial recognition system. These limitations are negated using the LDA algorithm

The LDA Linear Discriminant Analysis algorithms are implemented using the Fisherface Method. The algorithm initially applies PCA algorithm to reduce sample image dimension to find the scatter matrices. Fischerfaces are obtained from the scatter matrics which will be used to determine similarity amongst intraclass image samples. Implementation of The LDA Algorithm for Online Validation Based on Face Recognition [6] utilizes the LDA algorithm. The LDA algorithms suffer from small sample size problem which requires large amount of intraclass data samples to be available for training.

In DeepFace: Closing the Gap to Human-Level Performance in Face Verification el at[3] the authors utilize a neural network for facial recognition. The model first extracts key facial points from the given sample, then crops the face based on the key points to create a frontal face model using the key points and fiducial point map. The obtained frontal face is passed into a neural network whose output is used for classification. Although this model utilizes trained neural network it involves large quantity of image pre-processing before being transferred to a neural network.

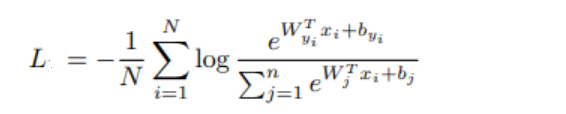
After DeepFace, Google launched its own facial recognition model FaceNet: A unified embedding for face recognition and clustering by F. Schroff el at [7]. This model uses end to end process which reduced the image pre-processing step. FaceNet utilized triplet loss where the three data samples are used two from the same intraclass A and one from a intraclass B. This enabled the neural network to maximize the difference between intraclasses. The limitations of FaceNet is that the model contains numerous hidden layers and as the dataset requires three images to be fed to the neural network it becomes difficult to handle.

To overcome the issue of the model requiring three images for training, Siamese Networks are implemented for face recognition these Siamese Networks are two identical neural networks whose final hidden layers are fused together to find their respective Euclidean distance between the outputs of the two identical networks to provide a binary output. Dataset samples contain two images of similar class and asimilar class for training. Small sample face recognition algorithm based on novel Siamese network el at Zhang [8] utilizes a Siamese network for face recognition. Although the Siamese network does not use three sample images such as FaceNet it still requires two images which becomes cumbersome on large datasets.

To overcome the shortcomings of a vanilla Softmax function the authors of A Discriminative Feature Learning Approach for Deep Face Recognition by Wen, Y el at [9] propose a Center loss which explicitly penalizes intra-class variance by adding the sum of squared distance of all the points in the batch from their class centroid. As the proposed model does not require any special dataset pre-processing it is easier to train and run multiple test iterations, but the model performs poorly when the number of classes is large.

1. METHODOLOGY

The most used method of loss function used for facial recognition and classification is called Softmax loss function and the equation is given below.

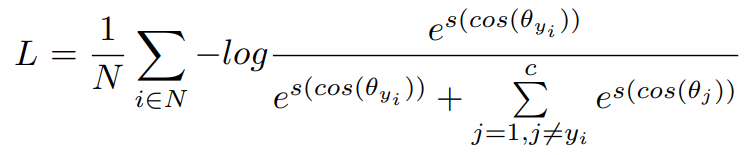


Eq 3.1 Softmax Loss Function

Where 𝑥i ∈ ℝd specifies the feature of the ɨth sample from the 𝒚ith  class. 𝑊j ∈ ℝd specifies the jth column of the weight 𝑊 ∈ ℝd×n and bj ∈ ℝn specifies the bias term. As the Softmax function is one of the oldest loss functions used, older facial recognition models utilise the Softmax loss function. However, the Softmax loss function is a general loss function and it lacks in effectively utilising the feature embeddings to decrease intraclass variance and increase inter-class variance. This results in drastic variations in performance where datasets contain large age differences amongst the same classes and in datasets with large classes over a million pairs

The current models which are widely used are Cos face and Arc face models.

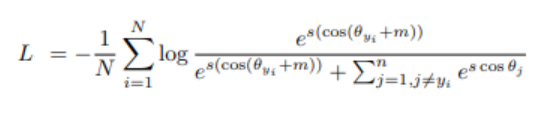
CosFace subjects the Softmax function to Cosine of the angle between the weights of the final hidden layer and the feature vector. The cosine face loss function can be formulated as follows



Eq 3.2 CosFace loss function

Where the 𝑥i𝑊*𝗒iT* = *||*𝑥*i||* *||*𝑊𝗒i||*cos(θ𝒚i )*, and *(θ𝒚i )* is the angle between the weights of final hidden layer 𝑊𝗒 and the feature vector 𝑥*i*. Although the addition of cosine to a Softmax function increases the accuracy of the model it still lacks in optimising the feature embeddings

ArcFace consists of an additive margin penalty which is added to an existing Softmax function and the given penalty is that of the geodesic distance in the non-normalised hypersphere.



Eq 3.3 ArcFace Loss Function

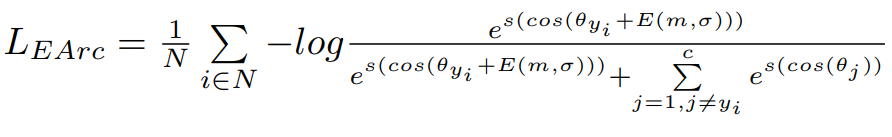
An additive angular margin penalty *m* is added to feature vector 𝑥*i* and the weights of the final hidden layer 𝑊𝒚

The proposed model, unlike previous prominent models such as ArcFace and CosFace does not utilise a fixed penalty instead the penalty is provided by randomly drawing values from a Gaussian distributed function. This approach increases the differences amongst interclass and due to the fortuity of the penalty, it reduces overfitting by acting similar to a dropout layer in a feedforward neural network. Due to the versatility of the proposed model the randomised penalty can be incorporated into any angular margin-based Softmax losses. The penalty is based on values randomly drawn from a Gaussian distribution function which is given below



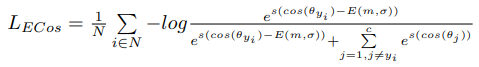
Eq 3.4 Gaussian Distribution Equation

Where µ is the function's mean distribution and σ is the standard deviation. As stated earlier the proposed random marginalised penalty can be incorporated into both ArcFace as ElasticFace-Arc and in the case of CosFace as ElasticFace-Cos. The loss equation of ElasticFace-Arc is given below



Eq 3.5 ElasticFace-Arc

And the equation of ElasticFace-Cos is given below



Eq3.6 ElasticFace-Cos

Where E(m, σ) is the Gaussian distribution function which would provide a random penalty based on mean m and standard deviation σ.

During training, the model perceives the intraclass variations and assigns a margin value based on their relative distance from their respective class centre. This allows outlines from the same class towards the class centre.

1. RESULT

The pretrained model is built using the PyTorch framework. Due to the efficiency of PyTorch in the field of research deep learning it is chosen over TensorFlow which is greatly used in edge devices and for real-time deployment purposes. The testing was done using Google Colab with a 12GB NVIDIA Tesla K80 GPU. The model was tested over benchmark datasets such as LFW[10], CFP-FP and AgeDB[11] and the model managed to perform better than the state of the art models in two datasets. The AgeDB dataset contains datasets with large age differences between the data samples in which the proposed model managed to get an accuracy score of 98.35% while the state of the art models reached 98.32% accuracy which was scored by CurricularFace[12]. In the case of LFW, which is one of the oldest datasets for facial recognition and where the proposed models scored 99.80% slightly behind the current state-of-the-art model VarGFaceNet[13] at 99.85%. Given below is the testing log of the proposed ElasticFace model in LFW dataset.

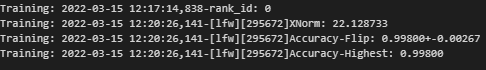


Fig 4.1 Testing log on LFW dataset

1. CONCLUSION

Elastic Margin loss for facial recognition has been interpreted and tested. The model is efficient in detecting multiple variations amongst intraclass variations, this allows for robust usage in real-time technology where different poses and face orientations can be detected and properly classified. The increase in accuracy and reduction of overfitting along with the ability to incorporate the loss function onto existing ArcFace or CosFace models allows for ease of switching from the previous versions to the proposed model. As only the loss function is changed without much changes to the neural network, engineers can update their existing models on their edge devices to the proposed Elastic Margin loss without considerable hardware changes.

1. REFERENCE
2. T. Mantoro, M. A. Ayu and Suhendi, "Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods," 2018 6th International Conference on Multimedia Computing and Systems (ICMCS), 2018, pp. 1-5.
3. Ian Haikal Amir A., Raihan Nugroho P., Wiwit Ria Rahayu, Dea Faiza Febrianty, Nailil Farihah, Winda Nur Azizah, Iwan Cony Setiadi, Yonatan Davidson Gultom, "Face detection and recognition in real-time photos with haar cascade and local binary pattern histogram for automatic door locking system," Proc. SPIE 11789, Fourth International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2020), 1178908 (12 March 2021).
4. T. A. Kiran, N. D. K. Reddy, A. I. Ninan, P. Krishnan, D. J. Aravindhar and A. Geetha, "PCA based Facial Recognition for Attendance System," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 248-252.
5. S. S. Meher and P. Maben, "Face recognition and facial expression identification using PCA," 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 1093-1098.
6. Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708.
7. Zainuddin, Zahir & Laswi, A. (2017). Implementation of The LDA Algorithm for Online Validation Based on Face Recognition. Journal of Physics: Conference Series.
8. F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823.
9. Zhang, Jianming & Jin, Xiaokang & Liu, Y. & Sangaiah, A.K. & Wang, J.. (2018). Small sample face recognition algorithm based on novel Siamese network. Journal of Information Processing Systems.
10. Wen, Y., Zhang, K., Li, Z., Qiao, Y. (2016). A Discriminative Feature Learning Approach for Deep Face Recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9911. Springer, Cham.
11. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments Gary B. Huang,Manu Ramesh, Tamara Berg, and Erik Learned-Miller.
12. S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia and S. Zafeiriou, "AgeDB: The First Manually Collected, In-the-Wild Age Database," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1997-2005.
13. CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, Feiyue Huang
14. VarGFaceNet: An Efficient Variable Group Convolutional Neural Network for Lightweight Face Recognition Mengjia Yan, Mengao Zhao, Zining Xu, Qian Zhang, Guoli Wang, Zhizhong Su