Large Age Gap Face Verification by Learning GAN Synthesized Prototype Representations

Swastik Jena[1][a], Bunil Kumar Balabantaray[1][b] and Rajashree Nayak[2][c]

[1]*Dept. of Computer Science and Engineering, National Institute of Technology Meghalaya, India*

[2]*School of Applied Sciences, Birla Global University, Bhubaneshwar, India*

Keywords: SimSwap GAN, ArcFace, PFA, LAG Dataset, Siamese Neural Network, Deep Metric Learning.

Abstract: A phenomenal growth in the field of face recognition has been witnessed over the last few years. Existing deep learning-based face recognition methodologies employ auxiliary age classifiers and intermediate age synthesizers to address the discrepancies in facial appearance due to aging. However, even after training on large amount of annotated data samples and by utilizing prior information the existing methodologies still underperform in recognizing the large intra-class age variance posed by images of same identity. LAG is a challenging face verification benchmark dataset having very few samples per identity with large age variance and no age annotations. This paper aims to perform face verification on the LAG dataset by learning the large intra-class variance posed by aging. The proposed work integrates a new training regime for the face verification task. SimSwap GAN is used for generating hybrid faces from young and adult images present in the LAG dataset. A Prototype Feature Activation (PFA) network is used to extract the feature embeddings of the hybrid faces and a modified Siamese Neural Network is trained to learn the face embeddings combined with attention-enhanced feature fusion. Extensive experiments highlight the outperforming performance of the proposed approach compared with existing baseline face verification methods on the LAG dataset.

1 INTRODUCTION

Age-related face verification has been an open challenge for the research community. Extensive work has been done for developing real world face recognition and verification systems in which age information plays a vital role (Jain and Li, 2011), (El Khiyari and Wechsler, 2016). Face is an integral attribute for bio-metric verification and identification systems, as it provides the necessary information about individual's identity (Grm et al., 2018). However, this attribute is the most vulnerable to the aging process, which significantly affects the performance of face verification algorithms (Lanitis, 2009). Face verification of identities with significant age difference is a challenging and resource intensive task. Issue of aging is highly challenging due to various factors such as:

- Aging is non-deterministic in nature, it cannot be controlled also it is not possible to eliminate aging variation during the process of image capture.
- Aging affects people differently, this may be due to gender, ethnicity, lifestyle, environment.

Table 1 depicts the brief details of the publicly available face verification/recognition datasets. The task becomes more challenging due to the fact that the child images are excessively scarce in the most commonly used publicly available datasets like CASIAWebFace (Yi et al., 2014), MSCeleb1M (Guo et al., 2016). The existing datasets as shown in Table 1 can be divided into two main groups: datasets sampled in controlled environments where pose, illumination, facial obstruction etc. are manually supervised and datasets sampled in uncontrolled environments where samples are captured in an unconstrained (no manual interference with the pose or occlusion or lighting) manner referred to as in the wild images. Early datasets belonged to the first category, a few of them are : CMU PIE (Sim et al., 2002), FERET. As for the case of uncontrolled environment the most widely used datasets are : LFW dataset (Huang et al., 2008), CASIAWebFace (Yi et al., 2014) and LAG dataset (Bianco, 2017). The LFW dataset (Huang et al., 2008) pruned news programs and provided a total of 13,233 images for 5749 unique identities. CASI-

[a] https://orcid.org/0009-0004-0241-4378
[b] https://orcid.org/0000-0002-2769-7122
[c] https://orcid.org/0000-0002-1303-1979

Jena, S., Balabantaray, B. and Nayak, R.
Large Age Gap Face Verification by Learning GAN Synthesized Prototype Representations.
DOI: 10.5220/0012398800003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 70-78
ISBN: 978-989-758-684-2; ISSN: 2184-4313

AWebFace dataset (Yi et al., 2014) is one of the largest publicly available dataset that is commonly used for pre-training the feature extractors of various models, it contains a total of 986,912 images of 10,575 people. The aforementioned datasets are quite restrictive in nature as they are tailor made specifically for face recognition and verification tasks and most importantly they contain almost no age variation in their data samples.

For the task of age estimation and cross-age face recognition, the most popular datasets are: FGNet (Cootes and Lanitis, 2008) and MORPH (Ricanek and Tesafaye, 2006). FGNET dataset (Cootes and Lanitis, 2008) was proposed quite early and it is a very compact dataset consisting of only 1002 images corresponding to 82 people. MORPH (Ricanek and Tesafaye, 2006) is mainly a facial age estimation dataset providing 55,314 image samples for 13,168 unique face identities incorporating an age range of 16 to 77 years with an age gap of upto 5 years. These datasets have considerable amount of samples per class and the intra-class age difference is not very high. Most importantly these datasets provide age information regarding image samples in the form of labels and annotations. These age annotations restrict the generalizing power and robustness of models as image samples captured for practical and real-time applications do not contain these additional age annotations.

The problem becomes quite compelling with the consideration of large age gaps. To the best of our knowledge, the LAG dataset (Bianco, 2017) is the one publicly available dataset which contains only 3828 images for 1010 classes with large age gaps between data samples and no age annotations. Face verification for images in this dataset is much like the real world scenario and a difficult task. We have chosen LAG dataset (Bianco, 2017) to provide a challenging setting for our model, and analyzed how it learns to capture the large age variance with only a few samples of varying quality and no other auxiliary age information. In the LAG dataset (Bianco, 2017) large age gap is captured in the following manner:

- Extreme age difference e.g., Young (Y) vs Old (O)
- Significant variation in facial attributes due to aging e.g., Baby vs Teenager/Adult

The capability to perceive people across large age gaps is highly instrumental for many applications like:

- Identification of long-lost and found persons
- Avoid updating large facial databases with more recent images
- Recognition and identification of wanted criminals by comparing suspect's face to mugshots that could have been taken years before by law enforcement agencies
- Diagnosis of disease by discovering the premature aging of a person
- Providing utility to photo sharing websites

This wide range of applications of face recognition based on large age gap motivated us to propose large age gap based face verification model by learning GAN synthesized prototype representations. The suggested work enables to provide excellent face verification accuracy by learning the large intra-class variance posed by aging in a challenging setting. The LAG dataset (Bianco, 2017) without any auxiliary age information or annotation has been used to validate our proposed work. The contributions of this work are encapsulated as follows:

- Proposition of a new training paradigm with the help of SimSwap GAN (Chen et al., 2020) to train on scarce image samples with no auxiliary age annotations or any pseudo labels.
- Proposition of Prototype Feature Activation enhanced Siamese neural network architecture for attention based feature fusion.
- In-depth analysis of face verification task using Metric Learning based loss functions on the LAG dataset (Bianco, 2017).

The rest of the paper is organized as follows: in Section 2, we review some of the existing methods and recent advances in the field of face verification and recognition with emphasis on face verification in the context of large age gaps. In Section 3, we elucidate our proposed methodology using face swap GANs to produce class prototypes to train on the scarce dataset utilizing our proposed PFA enhanced Siamese neural network architecture. Section 4 reports the experimental results of our proposed model and its comparison with some of the existing methods. Section 5 concludes the paper by elucidating the scope for future work.

2 RELATED WORK

Age-related face verification methodologies fall under two broad categories. The first one utilizes discriminative models trained on huge datasets using deep convolutional neural networks (DCNN) (El Khiyari and Wechsler, 2016), (Wang et al., 2018b), (El Khiyari et al., 2017), (Sajid et al., 2018). The second line of work focuses on generative methods

Table 1: State of the art Face recognition/verification Data bases.

Datasets	Number of Samples	Number of Persons	Samples per Person	Prior Information	Age Variance	Environment Setting
LFW (Huang et al., 2008)	13,233	5749	2.3	No	Moderate	Uncontrolled
FGNet (Cootes and Lanitis, 2008)	1002	82	12.2	Age Label	0-45	Controlled
MORPH (Ricanek and Tesafaye, 2006)	55,134	13,618	4.1	Age Label	0-5	Controlled
CACD (Chen et al., 2015)	163,446	2000	80	Age Range	16-62	Uncontrolled
CASIA (Yi et al., 2014)	986,912	10,575	93.3	No	Moderate	Uncontrolled
MSCeleb1M (Guo et al., 2016)	1,000,000	100,000	10	No	Moderate	Uncontrolled
LAG (Bianco, 2017)	3828	1010	3.1	No	Large	Uncontrolled

to produce age synthesized data for training (Ramanathan and Chellappa, 2006), (Ramanathan and Chellappa, 2008), (Huang et al., 2021) or prototype samples to aid in learning class specific general features (Kemelmacher-Shlizerman et al., 2014), (Rowland and Perrett, 1995) . With recent advances in metric-learning based loss function (Wang et al., 2017b), (Wen et al., 2016) (Deng et al., 2019), (Hermans et al., 2017), (Liu et al., 2017), (Zhang et al., 2019), (Liu et al., 2019), discriminative models have provided impressive results but their over reliance on huge annotated datasets poses a major roadblock. Especially, in the case of datasets containing samples with large intra-class variance it's very challenging to train these family of models. On the contrary, the generative methods are very volatile in nature as the training process is guided by generative adversarial networks (GANs). Moreover, these methods are highly susceptible to the quality and nature of the training samples. Below subsections briefly describes some of the significant face recognition methods of both of these categories.

2.1 Discriminative Methods

In recent years, we have seen the unprecedented success of deep convolutional neural network based face recognition techniques. Some of the remarkable methodologies for super-human level performance on face identification and verification tasks are DeepFace (Parkhi et al., 2015), DeepID (Sun et al., 2014), FaceNet (Schroff et al., 2015). Availability of large-scale training data and evolution of impressive training losses for the deep network architectures help to achieve the superior recognition performance. Earlier proposed works were reliant on metric-learning based loss, like Contrastive Loss, Triplet Loss (Hermans et al., 2017), N-Pair Loss (Sohn, 2016), Angular Loss (Wang et al., 2017b), etc. The major bottleneck here is that these methods suffer from combinatorial explosion when we construct image pairs for training. Moreover, these embedding based methodologies are proved to be inefficient while training on large-scale datasets with large number of classes. Therefore, the focus of research in the context of deep learning based face recognition has shifted towards formulating more effective and efficient classification-based loss. Wen et al. (Wen et al., 2016) developed Center Loss which enhanced the compactness of intra-class features by learning the class centers for each identity. The L2-softmax (Ranjan et al., 2017) and NormFace (Wang et al., 2017a) based methods constrained both the features and weights during the training process via L2 normalization. Moving forward, the family of angular margin based losses, such as SphereFace (Liu et al., 2017), CosFace (Wang et al., 2018a), ArcFace (Deng et al., 2019) surfaced, progressively improving the performance to new heights. More recently, AdaptiveFace (Liu et al., 2019), AdaCos (Zhang et al., 2019) have introduced adaptive margin strategy for automatic fine-tuning of hyperparameters and enforce more effective supervision during training.

However, these methods provide improved face recognition performance when trained on large scale datasets (Guo et al., 2016), (Huang et al., 2008), (Yi et al., 2014), in a controlled environment where image samples do not exhibit any large age gaps and usually the auxiliary age information is provided in some form. These methodologies provide considerably low performance in recognizing images of same identity with large age gaps such as child-adult pairs. Moreover, in real time providing auxiliary age information requires manual annotation followed by training a separate age classifier model. This limits their practical applications and integration in real-time systems.

2.2 Generative Methods

The existing generative methods can be roughly divided into prototype and deep generative model-based approaches. Prototype-based methodologies (Rowland and Perrett, 1995), (Kemelmacher-Shlizerman et al., 2014) accomplish face aging/rejuvenation by utilizing the mean of faces in each age group, but this conversely affects the preservation of facial identity.

Deep generative model-based methods (Nhan Duong et al., 2017), (Wang et al., 2016) are used for generation of age synthesized samples. Recurrent Face Aging (RFA) (Wang et al., 2016) propose a recurrent neural network for modelling the transition of intermediate states corresponding to age progression/regression. However, the focus of these methods is more skewed towards improving the visual aesthetics of generated faces rather than enhancing the performance on recognition or verification tasks. The generation process produces artifacts pertaining to group-level face transformation leading to unintended discrepancy in the face identity. MTLFace (Huang et al., 2021) tackles this issue by enhancing the visual quality with identity-related information for face recognition. It introduces an identity conditional module (ICM) aimed at attaining an identity-level face age synthesis compared to the earlier proposed group-level face age synthesis. The major limitation of these methods is that the performance is heavily dependent on the cross-age data samples which are not publicly available and are quite resource-intensive to build. Secondly, training an additional GAN model on a custom dataset increases the training time and the computational cost. Moreover, it restricts their generalizing power. Thus making these methods unsuitable for practical real-time applications.

As a solution, we have proposed a new generative methodology for the accurate face verification of same identity with large age gap in a challenging environment (small number of training samples and no prior information such as age, pseudo labels etc). Instead of generating cross-age images from scratch by using GAN, we have used a pre-trained SimSwap GAN (Chen et al., 2020) to perform face-swapping for generating hybrid faces from young and adult images present in the LAG dataset (Bianco, 2017). These hybrid faces preserve the facial anatomy of the class samples and produce relatively less artifacts as compared to images generated via conventional image-to-image translation. Moreover, the image generation process effectively preserve identity as well as discriminative features of each class while maintaining relatively less computational overhead. Our aim is to synthesize hybrid faces that function as class prototypes capturing the necessary intra-class variances present in the LAG dataset (Bianco, 2017) samples to aid in the training of our face verification model.

3 PROPOSED METHODOLOGY

We have proposed an enhanced deep learning based model for the verification of face identities with large intra-class variance posed by aging in a challenging setting. Proposed model trained with only a few data samples per class without the use of any auxiliary age information or annotation and provides an excellent prediction accuracy as compared to contemporary methods. Fig. 1 illustrates an overview of our proposed methodology. Contributions of our suggested method are:

- Similarity sampling for each class present in the LAG dataset (Bianco, 2017) using different deep metric learning based loss functions to find suitable images for hybrid face generation. Based on the similarity scores two images i.e., with highest and lowest similarity scores are chosen and fed into the SimSwap GAN (Chen et al., 2020) model.
- Face swapping and generation of hybrid prototype faces by using SimSwap GAN (Chen et al., 2020).
- Training a Prototype Feature Activation (PFA) network using the generated hybrid faces to learn the prototype feature space (hybrid face feature embeddings) for each class.
- Training a Siamese neural network for extracting the face embeddings of LAG dataset (Bianco, 2017) images.
- Proposition of attention enhanced feature fusion by integrating the hybrid face feature embeddings and face embeddings of LAG dataset (Bianco, 2017) images to perform the face verification task.

We experimented with various deep metric learning based loss functions to assess the discriminative power of generated embeddings. Our proposed training paradigm offers considerable performance gain over the earlier methods.

3.1 Image Similarity Sampling

3.1.1 Deep Metric Learning Loss Selection

We performed extensive experimental analysis by using various metric learning based loss functions in order to select the ideal similarity metric for our image sampling process. For all experiments we have employed a standard Siamese neural network with InceptionResNetv1 feature extractor (pre-trained on CASIAWebFace dataset (Yi et al., 2014)) with 512 dimension embedding space and checked with different loss functions. Table 2 shows the performance accuracy of the model by using various loss functions when tested

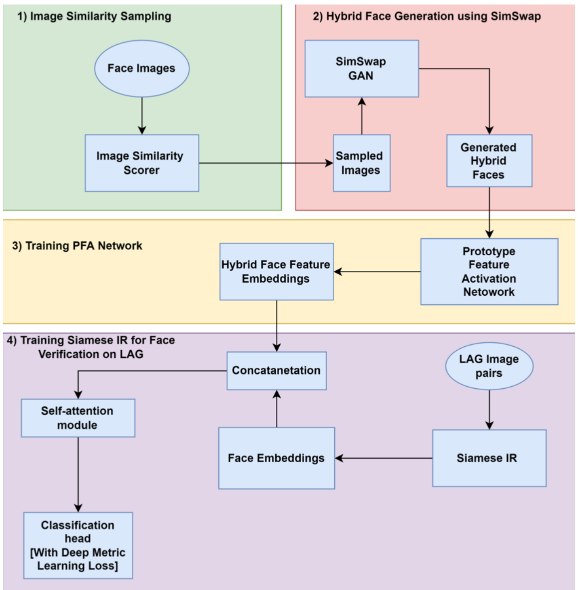

Figure 1: Block Diagram of Proposed Methodology.

on LAG dataset (Bianco, 2017). We choose ArcFace (Deng et al., 2019) as our similarity scoring metric for the next step as it gives the best performance compared to the other metrics.

3.1.2 Image Similarity Scoring

Our aim is to generate hybrid faces that capture the most wide array of variations present in a class, in turn providing us with highly discriminative, feature rich samples to train our verification model. Fig. 2 illustrates the basic overview of this similarity sampling method. We implement an ArcFace (Deng et al., 2019) based image similarity scorer pre-trained on CASIAWebFace dataset (Yi et al., 2014) with InceptionResNetv1 back bone and 512 dimension embedding space to extract two images for each class. The first image is the one exhibiting highest similarity score with all other images present in a class and is denoted as O. The second image is the one exhibiting lowest similarity score with the rest of the images present in the class and is denoted by Y. We effectively extract the easiest and the hardest sample to learn for a verification model. These two images i.e., [O, Y] act as our input for the SimSwap GAN (Chen et al., 2020) to generate hybrid face samples as explained ahead. The ArcFace Loss (Deng et al., 2019) that we use throughout our experiments is given below :

$$L = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j\neq y_i} e^{s\cos\theta_j}} \quad (1)$$

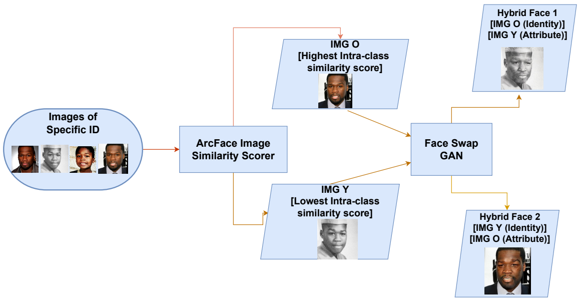

Figure 2: Similarity Sampling Method.

3.2 Hybrid Prototype Face Generation with SimSwap GAN

Instead of generating hybrid faces from scratch or synthesizing age progression/regression images we adopt SimSwap GAN (Chen et al., 2020) to perform face swap on the two face images that we obtained through the similarity sampling process. We choose SimSwap (Chen et al., 2020) as it provides unique benefits over other methods (Arjovsky et al., 2017), (Nirkin et al., 2019) in the form of identity extraction and facial attribute preservation which closely align with our objective. SimSwap (Chen et al., 2020) accepts a source image whose identity is extracted and a target image whose facial attributes are preserved. The resultant image that is generated belongs to the target image's domain and facial attributes (expression, posture, lighting) but exhibits the facial identity of the source image. We generate two hybrid faces from the image pairs obtained for each class. The first image O (having highest similarity score) acts as source image and the second image Y (having lowest similarity score) acts as target image to generate Hybrid Face 1. Now we interchange the source image and target image to generate Hybrid Face 2. Using the aforementioned technique we generate two unique hybrid faces integrating attributes of both image O and image Y for each class identity. We experiment with three pre-training schemes to analyse the effect of incorporating the generated hybrid faces in the pre-training data. They are denoted as follows: (i) Hybrid OY when trained using Hybrid Face 1 images; (ii) Hybrid YO when trained using Hybrid Face 2 images; (iii) Hybrid (OY + YO) when trained using both Hybrid Face 1 and Hybrid Face 2 images. Some of the hybrid faces generated using SimSwap GAN (Chen et al., 2020) are displayed in Fig. 3

Table 2: Comparison of deep metric learning based loss functions for face verification task on the LAG dataset (Bianco, 2017).

Feature Extractor	Loss Function	Accuracy
InceptionResNetv1	Triplet Loss (Hermans et al., 2017)	0.7494
InceptionResNetv1	ProxyAnchor (Kim et al., 2020)	0.7233
InceptionResNetv1	ProxyNCA (Movshovitz-Attias et al., 2017)	0.7543
InceptionResNetv1	Contrastive Loss	0.7730
InceptionResNetv1	CosFace (Wang et al., 2018a)	0.8012
InceptionResNetv1	**ArcFace (Deng et al., 2019)**	**0.8286**

Figure 3: Hybrid Face Samples generated using SimSwap GAN.

3.3 Prototype Feature Activation Network

The PFA network is an image embedding generator operating in the feature space of the hybrid faces that are generated using SimSwap GAN (Chen et al., 2020). It is an InceptionResnetv1 (pre-trained on CASIAWebFace dataset (Yi et al., 2014)) feature extractor with ArcFace loss (Deng et al., 2019) trained for face recognition task on the generated hybrid faces for each class. The network uses 512 dimension embedding space and classification head of dimension 1010 (number of classes in LAG dataset (Bianco, 2017)). The main purpose of this network is to learn the prototype activations present in the feature vectors of the hybrid faces to aid in feature fusion when integrated with the InceptionResnetv1 based Siamese neural network (Siamese IR) that we use for face verification task on the LAG dataset (Bianco, 2017).

3.4 Siamese Network with Attention Enriched Feature Fusion

Our architecture as illustrated in Fig. 4, is the main model that we use to perform the face verification task on the LAG dataset (Bianco, 2017) by taking an image pair as input and predicting whether they belong to the same person or not. It is a Siamese neural network with InceptionResnet backbone pre-trained on CASIAWebFace dataset (Yi et al., 2014) and has a 512 dimension embedding space. Our proposed architecture integrates a PFA network to perform feature fusion via attention module on the embeddings generated by both the networks to produce rich feature representations. We experiment with different training schemes i.e., Hybrid OY, Hybrid YO and Hybrid (OY + YO) to study the performance impact caused by the choice of hybrid faces used for training.

4 RESULTS AND ANALYSIS

In this section we describe the performance of our proposed methodology, Table 3 shows the different performance measure parameters of our proposed *Hybrid OY, Hybrid YO and Hybrid (OY + YO)* models compared to the state-of-the art method in (Bianco, 2017). To have a fair comparison we have implemented a backbone feature extractor i.e., InceptionResnetv1 pre-trained on the CASIAWebFace dataset (Yi et al., 2014) for all above cases. We maintained embedding dimension of 512 for all metric learning based loss functions. The images are resized to (160×160) pixels. We generated 5051 matching image pairs belonging to the same person and an equal amount of non-matching pairs are also generated. We have generated hybrid face pairs by using SimSwap

Table 3: Comparison of our proposed methodology with baseline validated on LAG dataset (Bianco, 2017).

Method	Feature Extractor	Loss Function	Pre-training dataset	Accuracy
Siamese DCNN + Feature Injection (Bianco, 2017)	DCNN	Contrastive Loss	CASIAWebFace (Yi et al., 2014)	0.8495
Siamese IR + PFA network	InceptionResnetv1	ArcFace (Deng et al., 2019)	CASIAWebFace (Yi et al., 2014) + Hybrid YO via SimSwap (Chen et al., 2020)	0.8693
Siamese IR + PFA network	InceptionResnetv1	ArcFace (Deng et al., 2019)	CASIAWebFace (Yi et al., 2014) + Hybrid OY via SimSwap (Chen et al., 2020)	0.8862
Siamese IR + PFA network	InceptionResnetv1	ArcFace (Deng et al., 2019)	**CASIAWebFace (Yi et al., 2014) + Hybrid (OY+YO) via SimSwap (Chen et al., 2020)**	**0.8970**

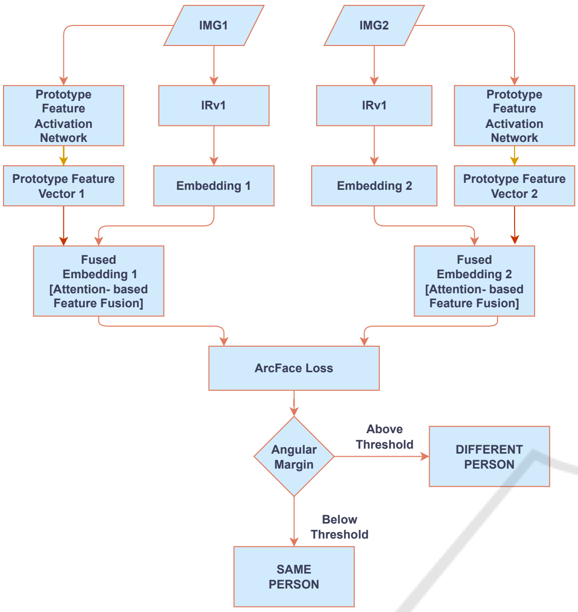

Figure 4: Overall Steps for the Face Verification Task.

(Chen et al., 2020). Fig. 3 depicts some sample generated hybrid faces. ROC curves of compared methods are depicted in Fig. 5. By analyzing Table 3 and Fig. 5 it is quite evident that, our proposed *Hybrid OY, Hybrid YO* and *Hybrid (OY + YO)* models outperform the baseline method (Bianco, 2017) by a significant margin. Proposed *Hybrid OY, Hybrid YO* and *Hybrid (OY + YO)* models provide average gain in classification accuracy by an amount of 2.27 %, 4.14% and 5.3% over the original method respectively. Although we combine method (Bianco, 2017) with the deep metric loss functions, it struggled to capture the large intra-class variance from scarce data samples. By augmenting the training paradigm and by not making use of any kind of pseudo labels or auxiliary age classifiers our proposed methodology is able to outperform all other methods as shown by the ROC curves. The computational overhead incurred during the hybrid face generation process is relatively light compared to generative methods that use recurrent networks and generate images from scratch. Our design philosophy of re-framing the generation task as an image-to-image translation task allows us to capture diverse specific class features and generate high quality samples with minimal artifacts. Through the image similarity sampling process we carefully segregate images with highly discriminative features with respect to a class as well as images with features that are very hard to learn. Our experiments reveal that more than 90 % images with the lowest similarity score per class are child images. This highlights the major issue faced by current face verification/recognition systems. The child images do not provide enough discriminative features to the model to perform proper verification. Thus using a face swap GAN to produce hybrid images using the most discriminative feature rich image and the least discriminative feature rich image for a class we capture the entire spectrum of feature variance present in a class. Our experiments by varying the target and source image between young and old images results in considerable impact on model performance. Secondly by implementing an angular margin based metric learning loss function like ArcFace (Deng et al., 2019) we provide additional motivation for the model to learn inter-class features and generate feature rich discriminative embeddings for the face verification task.

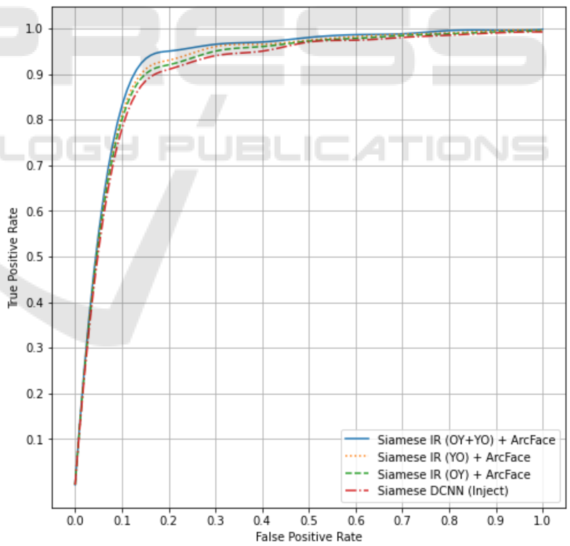

Figure 5: ROC Curves for Proposed and Compared Methods.

5 CONCLUSION

In this paper we have suggested a highly efficient face verification methodology which enabled to provide excellent accuracy for verifying faces of same identity with high age variance. As compared to existing similar methodologies, our proposed deep learning methodology is trained in a highly challenging environment i.e., with very few number of samples per individual class and use absolutely no prior information such as age label or age range during the training process. Instead of using conventional GAN generated face images or age synthesized data samples to boost up the training samples, our method fed the images in LAG dataset (Bianco, 2017) into SimSwap GAN (Chen et al., 2020) to generate two hybrid images per each class. The generated hybrid images preserve the facial anatomy and attributes of the class samples and produce relatively less artifacts. The expressive face embeddings (from both Hybrid faces and LAG dataset (Bianco, 2017) faces) coupled with attention enhanced feature fusion provides nice guidance for the verification task and results in 5.3% average improvement in the verification accuracy as compared to the state-of-the-art method.

REFERENCES

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Bianco, S. (2017). Large age-gap face verification by feature injection in deep networks. *Pattern Recognition Letters*, 90:36–42.

Chen, B.-C., Chen, C.-S., and Hsu, W. H. (2015). Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6):804–815.

Chen, R., Chen, X., Ni, B., and Ge, Y. (2020). Simswap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2003–2011.

Cootes, T. and Lanitis, A. (2008). The fg-net aging database. *ed*.

Deng, J., Guo, J., Xue, N., and Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699.

El Khiyari, H. and Wechsler, H. (2016). Face recognition across time lapse using convolutional neural networks. *Journal of Information Security*, 7(3):141–151.

El Khiyari, H., Wechsler, H., et al. (2017). Age invariant face recognition using convolutional neural networks and set distances. *Journal of Information Security*, 8(03):174.

Grm, K., Štruc, V., Artiges, A., Caron, M., and Ekenel, H. K. (2018). Strengths and weaknesses of deep learning models for face recognition against image degradations. *Iet Biometrics*, 7(1):81–89.

Guo, Y., Zhang, L., Hu, Y., He, X., and Gao, J. (2016). Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pages 87–102. Springer.

Hermans, A., Beyer, L., and Leibe, B. (2017). In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*.

Huang, G. B., Mattar, M., Berg, T., and Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Huang, Z., Zhang, J., and Shan, H. (2021). When age-invariant face recognition meets face age synthesis: A multi-task learning framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7282–7291.

Jain, A. K. and Li, S. Z. (2011). *Handbook of face recognition*, volume 1. Springer.

Kemelmacher-Shlizerman, I., Suwajanakorn, S., and Seitz, S. M. (2014). Illumination-aware age progression. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3334–3341.

Kim, S., Kim, D., Cho, M., and Kwak, S. (2020). Proxy anchor loss for deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3238–3247.

Lanitis, A. (2009). Facial biometric templates and aging: Problems and challenges for artificial intelligence. In *AIAI Workshops*, pages 142–149.

Liu, H., Zhu, X., Lei, Z., and Li, S. Z. (2019). Adaptiveface: Adaptive margin and sampling for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11947–11956.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., and Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220.

Movshovitz-Attias, Y., Toshev, A., Leung, T. K., Ioffe, S., and Singh, S. (2017). No fuss distance metric learning using proxies. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 360–368.

Nhan Duong, C., Gia Quach, K., Luu, K., Le, N., and Savvides, M. (2017). Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 3735–3743.

Nirkin, Y., Keller, Y., and Hassner, T. (2019). Fsgan: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7184–7193.

Parkhi, O., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association.

Ramanathan, N. and Chellappa, R. (2006). Modeling age progression in young faces. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 387–394. IEEE.

Ramanathan, N. and Chellappa, R. (2008). Modeling shape and textural variations in aging faces. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8. IEEE.

Ranjan, R., Castillo, C. D., and Chellappa, R. (2017). L2-constrained softmax loss for discriminative face verification. *arXiv preprint arXiv:1703.09507*.

Ricanek, K. and Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. In *7th international conference on automatic face and gesture recognition (FGR06)*, pages 341–345. IEEE.

Rowland, D. A. and Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE computer graphics and applications*, 15(5):70–76.

Sajid, M., Shafique, T., Manzoor, S., Iqbal, F., Talal, H., Samad Qureshi, U., and Riaz, I. (2018). Demographic-assisted age-invariant face recognition and retrieval. *Symmetry*, 10(5):148.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823.

Sim, T., Baker, S., and Bsat, M. (2002). The cmu pose, illumination, and expression (pie) database. In *Proceedings of fifth IEEE international conference on automatic face gesture recognition*, pages 53–58. IEEE.

Sohn, K. (2016). Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems*, 29.

Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1891–1898.

Wang, F., Xiang, X., Cheng, J., and Yuille, A. L. (2017a). Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049.

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., and Liu, W. (2018a). Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274.

Wang, J., Zhou, F., Wen, S., Liu, X., and Lin, Y. (2017b). Deep metric learning with angular loss. In *Proceedings of the IEEE international conference on computer vision*, pages 2593–2601.

Wang, W., Cui, Z., Yan, Y., Feng, J., Yan, S., Shu, X., and Sebe, N. (2016). Recurrent face aging. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2378–2386.

Wang, Y., Gong, D., Zhou, Z., Ji, X., Wang, H., Li, Z., Liu, W., and Zhang, T. (2018b). Orthogonal deep features decomposition for age-invariant face recognition. In *Proceedings of the European conference on computer vision (ECCV)*, pages 738–753.

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515. Springer.

Yi, D., Lei, Z., Liao, S., and Li, S. Z. (2014). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*.

Zhang, X., Zhao, R., Qiao, Y., Wang, X., and Li, H. (2019). Adacos: Adaptively scaling cosine logits for effectively learning deep face representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10823–10832.