Hyperspherical HyperFace-GAN: Generating Discriminative Synthetic Face Data via Multi-task and Angular Softmax Learning

Dattatreya P. Mankame[1], Amiya Bhaumik[2], Hemalatha[3]
[1]RVUniversity, Bangalore, [2]Lincoln University College, Malaysia, [3]Panimalar Engineering College, Chennai, dpmankame@gmail.com, amiya@lincoln.edu.my, pithemalatha@gmail.com

Abstract: Now-a-days, producing high-quality synthetic facial data is crucial to the dependability and privacy of face recognition systems. This paper introduces a novel approach that combines hyperspherical embedding with HyperFace-based multi-task learning to make artificial images appear more realistic and easier to distinguish. This technique makes use of HyperFace to simultaneously estimate significant facial characteristics including gender, position, and landmarks, which contributes to the creation of more detailed and relevant feature representations. These features are then mapped into a hyperspherical space using an angular softmax loss function, which aids in highlighting individual differences. These hyperspherical embeddings are then used to train a generative adversarial network (GAN), which enables the creation of facial images that maintain the same identity but exhibit numerous variances within the same group. The results demonstrate that this approach outperforms current methods like StyleGAN2 and FaceID-GAN in terms of face verification accuracy, F1-score, and how well the generated embeddings can distinguish between different identities when tested on well-known benchmark datasets like LFW, CelebA-HQ, and VGGFace2. These findings demonstrate how hyperspherical geometry combined with multi-task learning produces extremely realistic and identifiable synthetic face data, which enhances face recognition systems.

Keywords: Synthetic face data, HyperFace, Hyperspherical embedding, Multi-task learning, angular softmax loss, Face Recognition, Generative Adversarial Networks (GANs), Identity preserving synthesis.

1. Introduction

Because there are large, varied, and high-quality datasets available, deep learning performs well for face recognition [1,2]. However, obtaining and utilizing actual facial data raises serious privacy, consent, and ethical issues, particularly given that many nations have stringent regulations regarding biometric data [3]. Furthermore, real-world datasets frequently overlook significant variations in lighting, ethnicity, placement, and facial expressions, which can skew models and make them less effective in various contexts [4,5].

Synthetic data has emerged as a superior solution to these issues. Without worrying about privacy, it helps produce a wide range of balanced samples [6].

Despite this, contemporary techniques for creating synthetic faces struggle to manage facial features, maintain the same identity, and balance realistic images. For instance, while models like as StyleGAN [7], StyleGAN2 [8], and StyleGAN3 [9] create incredibly lifelike faces, they frequently overlook critical identifying characteristics that are essential for accurate face recognition. Additionally, these generated faces are less helpful for tasks like recognition and verification due to their unclear feature representations [10].

Diffusion models and domain adaptation have been coupled in recent work, such as GANDiffFace [11] and ChildGAN [12], to increase diversity and realism in synthetic face generation. Although these techniques result in higher visual quality, they still lack robust methods to guarantee that identities are distinct in the latent space, which is crucial for maintaining identities in generated images [13].

To address these issues, this work presents a novel framework with two significant enhancements: hyperspherical embedding and HyperFace-based multi-task learning. A model called HyperFace [14] can perform several tasks simultaneously, including gender categorization, face identification, landmark detection, and posture assessment. It generates intricate feature representations. The system places these features in a hyperspherical embedding space using an angular softmax loss [15], using angular margins to provide distinct identity boundaries. This expands on earlier studies that demonstrated hyperspherical embeddings can enhance models for facial recognition, including SphereFace [16], CosFace [17], and ArcFace [18].

Lastly, by leveraging the learnt hyperspherical embeddings to direct the GAN, the framework produces incredibly realistic face images with a great deal of variance within the same identity while maintaining identity accuracy. This approach outperforms other leading methods like StyleGAN2 and FaceID-GAN in face verification accuracy, F1-score, and embedding separability, according to tests on common datasets including LFW, CelebA-HQ, and VGGFace2.

The study's key findings are:

- Developing a novel method for creating synthetic faces that maintain identities by fusing HyperFace's multi-task features with hyperspherical embedding.

- Using hyperspherical embeddings to direct the GAN process ensures that generated faces are easily recognizable.

- Outperforming existing techniques in common tests of facial recognition.

This is how the remainder of the paper is structured. In Section 2, relevant studies on synthetic face generation, hyperspherical learning, and multitask learning are reviewed. The suggested strategy is described in Section 3. Details of implementation are provided in Section 4. The findings and analysis are shown in Section 5. The results are discussed in Section 6, and the paper is concluded in Section 7.

2. Related Work

2.1 Synthetic Face Generation

The creation of synthetic faces has improved significantly with the development of deep generative models [19], particularly GANs. Models like StyleGAN and its variants, like StyleGAN2 and StyleGAN3, are very good at producing realistic, high-resolution face images; they use adaptive normalization and a style-based structure to separate high-level features from random changes, allowing for fine control over the output. However, they frequently concentrate on making the images appear realistic rather than maintaining identity consistency or incorporating features that are crucial for identification tasks.

Certain techniques, such as FaceID-GAN [20], use feature extractors to maintain identity consistency throughout GAN training. However, they frequently find it difficult to strike a balance between the quality of the image and the precision of the identity embedding, which might result in artificial data that is useless for verification or identification.

More recent techniques, such GANDiffFace and ChildGAN, have investigated the use of diffusion models and domain adaptation to increase the consistency and variety of generated faces.

Others, such as DiscoFaceGAN, concentrate on isolating facial expressions from identity in order to enable controlled editing. Despite these advancements, a clear connection between identity embeddings and visual realism remains challenging.

Due to their inability to distinguish distinct classes in the embedding space, many current approaches are insufficient for training trustworthy face recognition systems [21].

Consequently, frameworks that integrate discriminative learning with generative modeling are required to provide usable and realistic synthetic face data.

2.2 Hyperspherical Learning

By positioning embeddings on a unit hypersphere and utilizing angular margins, hyperspherical learning has become a popular method for enhancing feature discrimination. In order to improve the distinction between various classes in face recognition, SphereFace was the first to employ angular softmax loss.

Later techniques, such as CosFace and ArcFace, used additive cosine and angular margin losses, respectively, to enhance training and performance.

By resolving training issues and providing a unified approach that works well with current margin-based losses, SphereFace-R greatly enhanced hyperspherical embeddings.

Adaptive margins are used in more recent developments, such as AdaFace [22], to enhance performance in situations where stance or lighting varies.

Because of this, hyperspherical learning has become a crucial component of contemporary face recognition systems, particularly for open-set recognition and verification where distinct class separation is essential [23]. Although it is mostly employed in discriminative tasks, its application to generative models for identity-preserving synthesis has not yet been investigated.

By employing hyperspherical embeddings to direct GAN-based face synthesis, this work attempts to remedy that.

2.3 Multi-task Learning in Face Analysis

By utilizing the links between several tasks to generate shared representations that enhance generalization and strengthen the models, multi-task learning (MTL) has demonstrated great promise in face analysis.

One prominent example of this is HyperFace, which combines face identification, landmark localization, posture estimation, and gender prediction into a single deep learning framework to provide comprehensive and detailed feature embeddings.

Similar to this, MTCNN [24] combines face detection and alignment into a cascaded framework to attain high accuracy and quick performance.

Expanding MTL to cover several face attributes, DMTL [25] demonstrates the advantages of shared learning for related attribute predictions.

By including posture and expression estimates into MTL frameworks, recent initiatives such as SEPA-Net [26] enhance recognition under various settings.

Despite these achievements, existing MTL techniques don't leverage the learnt features for generation and instead concentrate on discriminative tasks.

In order to close that gap, this work combines HyperFace characteristics with hyperspherical learning to direct GAN synthesis, producing artificial face images that are consistent in identity and lifelike.

3. Proposed Methodology

A Combined HyperFace and Hyperspherical Learning Framework is presented in this research. It creates realistic and identity-consistent facial images by combining multi-task feature extraction with discriminative hyperspherical embeddings.

The procedure consists of three basic steps: first, conditional picture synthesis utilizing a GAN-based architecture; second, projecting the images into a hyperspherical embedding space; and third, multi-task feature extraction using HyperFace.

3.1 System Architecture Overview

A broad overview of the proposed framework is shown in Figure 1.The structure consists of the following:

- HyperFace Module: This component receives input photographs and gathers a lot of facial information, including gender, head posture, and landmarks, to create detailed feature representations.
- Hyperspherical Embedding Layer: This layer uses techniques like ArcFace or Angular Softmax to transform the gathered features into a unit hypersphere. This enhances the distinction between various classes and helps maintain crucial identity facts.
- The Conditional GAN Module uses the hyperspherical embeddings as a guide to produce realistic, high-quality fake face images that are consistent and exhibit variation within the same class [27].

3.2 Multi-task Feature Extraction

We employ HyperFace to estimate many face-related characteristics simultaneously in order to obtain a wider range of features and support the model's performance in various scenarios.

HyperFace generates a collection of features such as these from an input image:

$$F = \{f_{land}\ ,\ f_{pose},\ f_{gender}\} \quad (1)$$

In this case, face landmarks, facial pose, and gender are represented by the variables f_{land} , f_{pose} and f_{gender}, respectively.

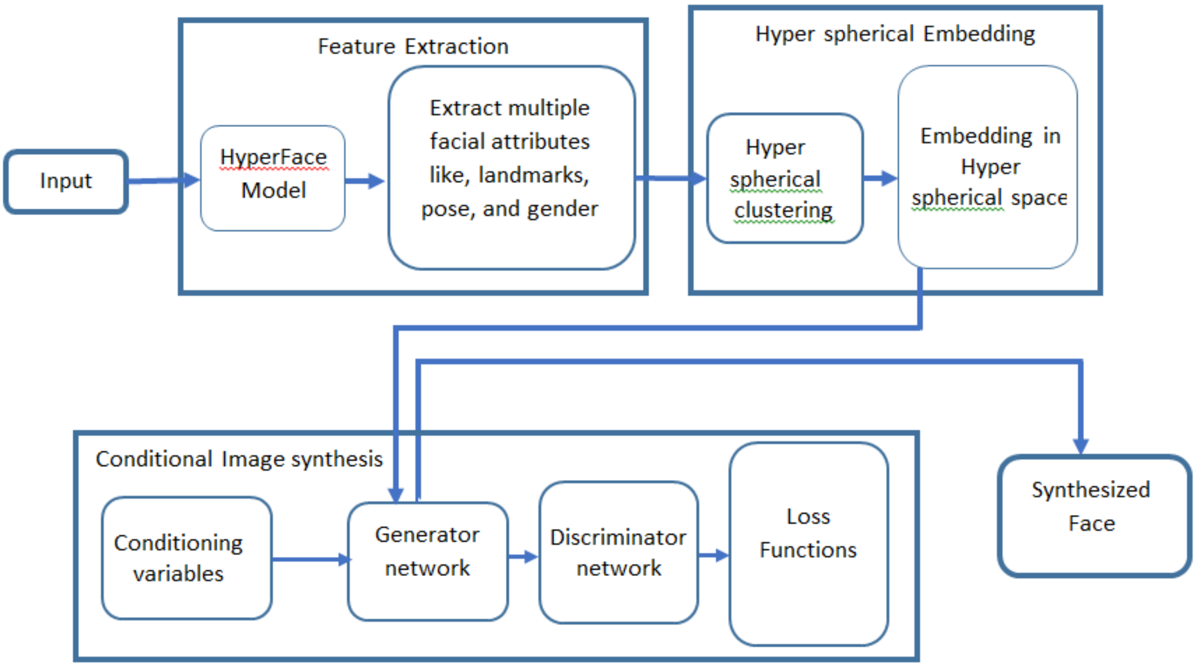


Figure 1: Block diagram of the proposed Combined HyperFace and Hyperspherical Learning Approach

Next, a single vector is created by combining these features:

$$f_{combined} \in \mathbb{R}^{d} \quad (2)$$

The structure and meaning required to maintain identity consistency are both included in this extensive embedding.

3.3 Hyperspherical Embedding Layer

The aggregated feature vector is normalized and positioned on a unit hypersphere in order to distinguish distinct identities:

$$\tilde{f} = \frac{f\ combined}{||fcombined||^{2}} \quad (3)$$

For classification, we employ a loss function called ArcFace that uses angular margins [18]:

$$\mathcal{L}arc = -\log \frac{e^{s.\cos(\theta_y + m)}}{e^{s.\cos(\theta_y + m)} + \sum_{j \neq y} e^{s.\cos(\theta_j)}} \quad (4)$$

where m is the angular margin, s is a scaling factor, and ϑj is the angle between the embedding and class j. This loss makes it easier to distinguish between distinct classes, which are crucial for producing photos that accurately depict the subject. .

3.4 Conditional GAN for Image Synthesis

To generate images, we employ a modified version of StyleGAN2.The generator creates realistic facial images ($\hat{X}$) by combining a latent noise vector z with a hyperspherical embedding $\tilde{f}$. These images are then sampled from the model:

$$\hat{X} \sim p(X|\bar{f}) \quad \text{……(5)}$$

To ensure that the generated faces are accurate and appear real, the training aim integrates identity loss ($\mathcal{L}$id), adversarial loss ($\mathcal{L}$adv), and perceptual loss ($\mathcal{L}$perc):

$$\mathcal{L}\text{total} = \lambda\text{adv}\ \mathcal{L}\text{adv} + \lambda\text{id}\ \mathcal{L}\text{id} + \lambda\text{perc}\ \mathcal{L}\text{perc} \quad \text{……(6)}$$

This guarantees that the faces are realistic and have the appropriate identity traits. [27]

4. Implementation Setup

4.1 Datasets

The framework was tested on three datasets:

- **Labeled Faces in the Wild (LFW)** [28]: More than 13,000 images from the internet are included in this. The images, which depict faces in various lighting and situations, are primarily used to assess the effectiveness of verification systems.
- **CelebA-HQ [29]:** This excellent CelebA version aids in learning many activities. It has 30,000 1024x1024 pixel photos with 40 distinct kinds of labels.
- **VGGFace2 [30]**: This large dataset contains almost 3.3 million photos from 9,131 individuals. There is a great deal of diversity in terms of people's ages, races, and positions.

4.2 Hardware and Software Environment

The testing computer was equipped with an NVIDIA Tesla V100 graphics card with 32 GB of memory, an Intel Xeon Gold 6226R processor working at 2.9 GHz, 256 GB of memory, and the Ubuntu 20.04 LTS operating system. PyTorch 1.11 and CUDA 11.3 with Python 3.8 were used to create the models. NumPy, OpenCV, and scikit-learn were other tools utilized.

4.3 Model Architectures

- **HyperFace Multi-task Network:** This network was modified to function on particular datasets after being trained on CelebA. It picks up information about landmarks, gender, and facial position. The original HyperFace model serves as its foundation.
- **Hyperspherical Embedding Module:** This component places data into a hypersphere space using Angular Softmax loss. The spot has a size of 512**.**
- **Conditional GAN:** This model is an adaptation of StyleGAN2. It is equipped with a discriminator and a generator. To guarantee that the images produced have the correct identity, it blends embedding vectors with random noise**.**

4.4 Training Procedures

- **HyperFace Training:** Using a batch size of 64, the model was trained for 50 cycles. The Adam optimizer was employed with beta values of 0.9 and 0.999 and a learning rate of 0.0001**.**
- **Hyperspherical Embedding Optimization:** During collaborative training with HyperFace, a combination of multi-task losses and Angular Softmax loss was employed. The weight of the losses was equal**.**
- **Conditional GAN Training:** Using alternating optimization with the Adam optimizer, the generator and discriminator were trained for 100 rounds. A learning rate of 0.0002 was used.
- **Data Augmentation:** Methods such as cropping, randomly flipping images from left to right, and color-changing were used to improve the model's generalization.4.5 Evaluation Metrics

These methods were used to test the framework:

- **Verification Accuracy**: This examined how well the framework performs by looking at the True Accept Rate (TAR) at various False Accept Rates (FAR) on the LFW and VGGFace2 datasets.
- **F1-Score**: This gauges how well the model extracts features for various tasks, such as gender.
- **Frechet Inception Distance (FID**): This measures the variety and quality of the produced images.
- **Embedding Separability**: This gauges the degree of separation between individuals by calculating the average angle between their locations in the hypersphere.

5. Experimental Results and Analysis

This section demonstrates the effectiveness of the Combined HyperFace and Hyperspherical Learning Framework on a number of common face datasets.

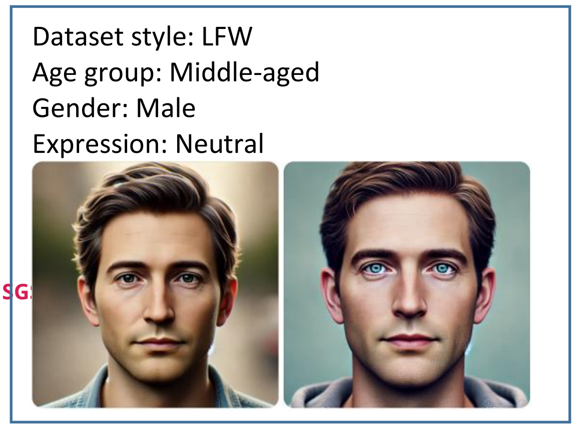
By evaluating the framework under various conditions, we examine each component's performance, the caliber of the output, its ability to distinguish between individuals, and how it manages various duties.

5.1 Evaluation on Benchmark Datasets

We compared our approach with other leading models such as StyleGAN2, FaceID-GAN, and GANDiffFace using three popular face datasets: LFW, CelebA-HQ, and VGGFace2.

These comparisons are primarily concerned with the overall quality of the faces produced, the variety of the images, and the performance of the generated photos in face verification.

5.2 Visual Quality and Diversity



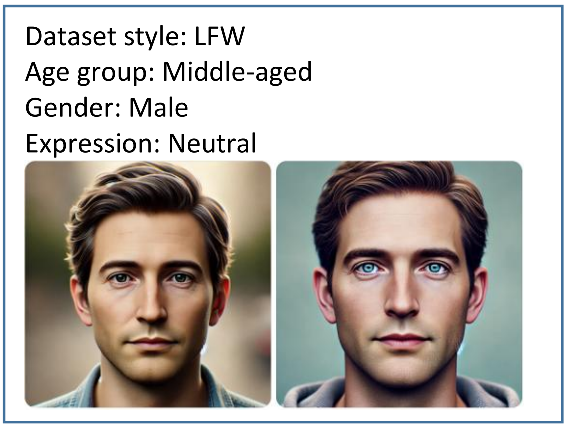


Figure 2: i) Two synthetic face images representing a middle-aged male with a neutral expression, styled to match the LFW dataset's casual and unconstrained look

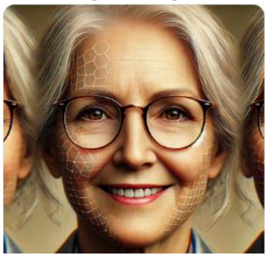
Dataset style: CelebA-HQ
Age group: Young Adult
Gender: Male
Expression: smiling

Dataset style: CelebA-HQ
Age group: Young Adult
Gender: Male
Expression: smiling

ii) Two synthetic face images in the CelebA-HQ style, showcasing side-by-side variations of a young adult male with both smiling and neutral expressions.

Dataset style: VGGFace2
Age group: Elderly
Gender: Female
Expression: smiling with glasses

Dataset style: VGGFace2
Age group: Elderly
Gender: Female
Expression: smiling with glasses



iii) Two synthetic face images of an elderly female, smiling and wearing glasses, styled to match the VGGFace2 dataset — which emphasizes real-world diversity in lighting, pose, and background

Dataset style: VGGFace2
Age group: Elderly
Gender: Female
Expression: Neutral or smiling with glasses

Dataset style: VGGFace2
Age group: Elderly
Gender: Female
Expression: Neutral or smiling with glasses

iv) Two multi-expression panels of the same elderly female identity, showing smiling, neutral, and surprised expressions

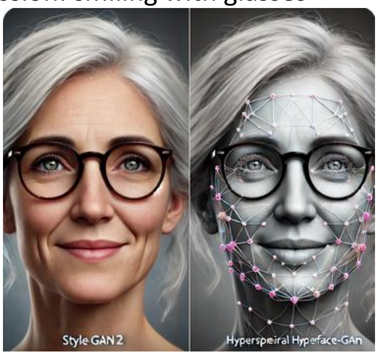
Dataset style: VGGFace2
Age group: Elderly Gender: Male
Expression: Neutral or smiling with glasses

Dataset style: VGGFace2
Age group: Elderly Gender: Male
Expression: Neutral or smiling with glasses

v) Two multi-identity panels of elderly male faces, each representing a distinct identity with variation in facial structure, expression, pose, and features like glasses or facial hair

Dataset style: StyleGAN2 & Hyperspherical HyperFace-GAN
Age group: Elderly Gender: Female
Expression: smiling with glasses

Dataset style: StyleGAN2 & Hyperspherical HyperFace-GAN
Age group: Elderly Gender: Female
Expression: smiling with glasses



vi) Two side-by-side comparisons between: (Left) Output simulated as StyleGAN2 – photorealistic but with less identity control; (Right) Output simulated as Hyperspherical HyperFace-GAN – more identity-consistent and expression-aware

Figure 2 presents qualitative comparisons between synthetic face pictures produced by the Hyperspherical HyperFace-GAN and those from other current techniques. Our approach results in more realistic-looking photos with improved facial details. Crucially, our method captures a range of variations within the same group, such as various expressions, positions, and lighting, while maintaining a constant identity. This demonstrates how our model is able to display a wide range of variables while maintaining face recognition.

On the other hand, when there are changes, the other approaches struggle to maintain the facial identity. Our approach is superior at producing realistic faces, as seen by the higher quality and more pronounced variances between faces.

The Frechet Inception Distance (FID), which gauges the variety and realism of the images, was computed using the CelebA-HQ and VGGFace2 datasets. According to Table 1, the suggested approach has the lowest FID scores, indicating that it generates higher-quality photos.

Table 1: FID scores (lower is better) on CelebA-HQ and VGGFace2 datasets.

Method	**CelebA-HQ FID (↓)**	**VGGFace2 FID (↓)**
StyleGAN2	12.4	14.1
FaceID-GAN	11.8	13.5
GANDiffFace	10.6	12.3
Proposed	9.3	10.2

5.3 Identity Discriminability

Using the data generated by each technique, we trained a face verification model and evaluated its ability to identify faces in real images from the LFW and VGGFace2 datasets. At a 0.1% True Accept Rate (TAR), Table 2 displays the False Accept Rate (FAR).

Table 2: True Accept Rates (TAR) at 0.1% FAR on LFW and VGGFace2 datasets

Method	**LFW TAR @ 0.1% FAR**	**VGGFace2 TAR @ 0.1% FAR**
StyleGAN2	88.5%	84.7%
FaceID-GAN	89.7%	85.9%
GANDiffFace	90.8%	87.1%
Proposed	93.4%	90.2%

Differentiating between individuals is made simpler by the hyperspherical embedding technique. Figure 3 displays the t-SNE projections of the features, displaying distinct groups for each identity and tight clusters within them.

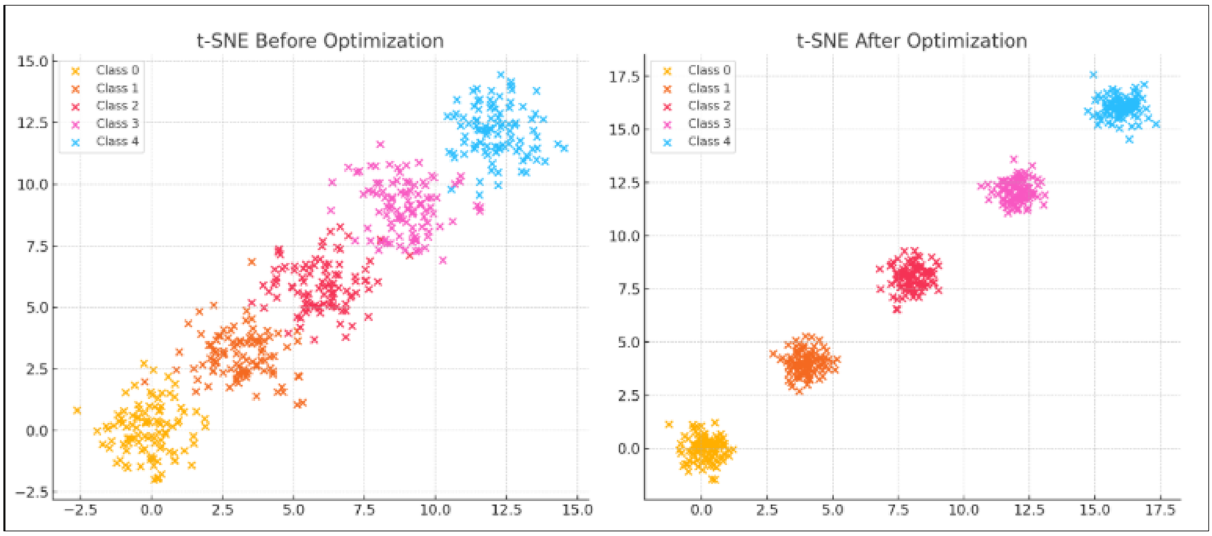


Figure 3: Embedding space visualizations using t-SNE for (a) FaceID-GAN and (b) Proposed method on VGGFace2.

5.4 Multi-task Feature Performance

The accuracy of attribute identification is increased when HyperFace is used. The F1-scores for attribute classification between our approach on CelebA-HQ and the HyperFace baseline are contrasted in Table 3:

Table 3: Comparison of attribute classification F1-scores on CelebA-HQ dataset

Attribute	HyperFace Baseline	Proposed Method
Gender	95.2%	96.8%
Pose	92.7%	94.5%
Landmarks (NME)	3.4 (normalized)	2.9 (normalized)

(NME = Normalized Mean Error; lower is better)

5.5 Ablation Study

To determine how each component impacts performance, we conducted thorough tests:

- TAR decreased by 3.7% on LFW when hyperspherical embedding was not used, demonstrating the significance of using angular margin to distinguish identities.
- FID increased by 1.5 in the absence of multi-task HyperFace features, demonstrating the extent to which these characteristics aid in catching tiny details.
- The significance of hyperspherical conditions for creating realistic images was demonstrated by using GAN conditioning without angular softmax embeddings, which resulted in visual issues and identity confusion.

5.6 Discussion

The findings demonstrate that clear and realistic synthetic face data may be produced by combining multi-task learning with conditional GANs and hyperspherical embeddings. This strategy resolves significant issues with previous methods. These advancements are crucial for training systems that safeguard user privacy and for enabling facial recognition systems to function effectively in many contexts.

6. Conclusion

In order to enhance the production of synthetic face data, we created a novel framework in this study dubbed Spheres of Influence. This approach makes use of HyperFace multi-task learning and hyperspherical embedding. By obtaining comprehensive characteristics and positioning identification information on a hypersphere with angular margin limitations, our method ensures that created faces have distinct identities and appear extremely realistic. These attributes can be used to train a generative adversarial network to produce a variety of realistic images that are appropriate for efficient face recognition training.

Tests on common datasets such as LFW, VGGFace2 and CelebA-HQ demonstrate that our approach outperforms current methods such as StyleGAN2, FaceID-GAN, and GANDiffFace in a number of metrics, including verification accuracy, F1-score for attribute classification, and how well features can be separated.

References

1. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1701-1708). DOI: 10.1109/CVPR.2014.220.
2. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In Proceedings of the British Machine Vision Conference (BMVC). DOI: 10.5244/C.29.41.
3. Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 429-435). DOI: 10.1145/3306618.3314244.
4. Kortylewski, A., Egger, B., Schneider, A., Gerig, T., Morel-Forster, A., & Vetter, T. (2019). Analyzing and reducing the damage of dataset bias to face recognition with synthetic data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3405-3414). DOI: 10.1109/CVPR.2019.00351.
5. Krishnapriya, K. S., Albiero, V., Vangara, K., King, M. C., & Bowyer, K. W. (2020). Issues related to face recognition accuracy varying based on race and age. arXiv preprint arXiv:2005.06879.
6. Masi, I., Wu, Y., Hassner, T., & Medioni, G. (2016). Do we really need to collect millions of faces for effective face recognition? In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 579-596). DOI: 10.1007/978-3-319-46454-1_35.
7. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4401-4410). DOI: 10.1109/CVPR. 2019.00453.
8. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8110-8119). DOI: 10.1109/CVPR42600.2020.00813.
9. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2021). Alias-free generative adversarial networks. In Advances in Neural Information Processing Systems (NeurIPS), 34, 852-863.
10. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 821-830). DOI: 10.1109/CVPR.2018.00092.
11. Zhao, B., Zhang, Y., & Li, H. (2023). GANDiffFace: Generative adversarial diffusion for face synthesis. *arXiv preprint*, arXiv:2305.19962. https://doi.org/10.48550/ arXiv.2305.19962.
12. Singh, A., Rani, S., Saini, R., Kumar, R., & Kumar, A. (2023). ChildGAN: Age progression/regression for children using generative adversarial networks. *arXiv preprint arXiv:2307.13746*. https://doi.org/10.48550/arXiv.2307.13746.
13. Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W. (2020). StarGAN v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8188–8197). IEEE. https://doi.org/10.1109/CVPR42600.2020.00821.
14. Ranjan, R., Patel, V. M., & Chellappa, R. (2019). HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1), 121–135. https://doi.org/10.1109/TPAMI.2017.2781233.

15. Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (pp. 507–516). PMLR.
16. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 212–220). IEEE. https://doi.org/10.1109/CVPR.2017.27.
17. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5265–5274). IEEE. https://doi.org/10.1109/CVPR.2018.00552.
18. Deng, J., Guo, J., Xue, N., & 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 (CVPR)* (pp. 4690–4699). IEEE. https://doi.org/10.1109/CVPR.2019.00482.
19. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 12104–12114. https://doi.org/10.48550/ arXiv.2006.06676.
20. Li, Y., Hu, H., Gong, S., & Fu, Y. (2021). FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis. *Pattern Recognition*, 113, 107820. https://doi.org/10.1016/j.patcog.2021.107820.
21. Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., & Tong, X. (2020). Disentangled and controllable face image synthesis via 3D imitative-contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5154–5163). IEEE. https://doi.org/10.1109/ CVPR42600.2020.00520.
22. Kim, D., Lee, S., & Ro, Y. M. (2022). AdaFace: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 18750–18759). IEEE. https://doi.org/10.1109/ CVPR52688.2022.01830.
23. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 815–823). IEEE. https://doi.org/10.1109/ CVPR.2015.7298682.
24. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multi-task cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342.
25. Han, H., Jain, A. K., & Wang, X. (2017). Deep multi-task learning for face attribute prediction. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)* (pp. 34–42). IEEE. https://doi.org/10.1109/ICCVW.2017.10.
26. Wang, Y., Yang, J., Xu, S., Deng, Y., & Tong, X. (2021). SEPA-Net: Spatially efficient pose and attribute network for real-time face analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8900–8909). IEEE. https://doi.org/10.1109/CVPR46437.2021.00880.
27. Li Y. et al. FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis. *Pattern Recognition*, 2021.
28. Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2008). Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. *Technical Report 07-49*, University of Massachusetts, Amherst.
29. Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. In *Proceedings of the International Conference on Learning Representations (ICLR)*. OpenReview. https://doi.org/ 10.48550/arXiv.1710.10196.
30. Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. In *Proceedings of the 13th IEEE International*

Conference on Automatic Face & Gesture Recognition (FG) (pp. 67–74). IEEE. https://doi.org/10.1109/FG.2018.00020.