Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 246 (2024) 1752–1761

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

Towards Efficient AI Solutions for Facial Recognition in the Wild

Asmail Muftah[a], Osama Almurshed[b,c], Mohamed Bennasar[d]*, Blaine Price[d], Sarah Laurence[e], Graham Pike[e]

[a]School of Science, Azzaytuna University, Tarhounah, Libya
[b]School of Computer Science and Information, Cardiff University, UK
[c]Department of Computer Science, Prince Sattam Bin Abdulaziz University, Saudi Arabia
[d]School of Computing and Communications, The Open University, UK
[e]School of Psychology and Counselling, The Open University, UK

Abstract

In addressing the challenges of facial recognition in in the wild, our study initially investigated computationally efficient approaches for facial recognition in uncontrolled environments rather than conventional, computationally intensive techniques such as generative adversarial networks (GANs) and 3D reconstruction. We leveraged the capabilities of an off-the-shelf deep learning model, namely VGGNet, for efficiency and practical deployment in real-world scenarios. Our methodology included a dual phase training approach, starting with comprehensive training of the entire model, followed by selective fine-tuning of specific layers. This process was conducted using the CelebA dataset, known for its diversity and relevance to facial recognition research. The study demonstrates that this approach not only maintains robust generalisation across diverse conditions but also significantly reduces computational demands. Despite a slight trade-off in accuracy compared to more traditional methods, the benefits of the increased efficiency and the potential for real-time application deployment, such as in surveillance systems requiring quick processing, present a compelling case for further investigation and development within the field of facial recognition technology.

Keywords: Face Recognition in the Wild; Deep Learning; Transfer Learning;

* Corresponding author. Tel.: +44 (0) 1908655019.
E-mail address: Mohamed.Bennasar@open.ac.uk

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems
10.1016/j.procs.2024.09.672

1. Introduction

The development of facial recognition systems capable of operating effectively "in-the-wild" scenarios remains a challenging endeavour in the field of computer vision [17]. There are a set of technologies in computer vision that are leveraged to solve some of the challenges in facial recognition in the wild. Among these, Generative Adversarial Networks (GANs), 3D reconstruction, and data augmentation stand out as key techniques. GANs have proven to be incredibly effective at crafting realistic facial images from minimal data, marking a significant step forward in creating viable recognition data. 3D reconstruction dives into the complex geometry of the face, enabling accurate modelling under a variety of conditions. Additionally, data augmentation is instrumental in broadening training datasets, substantially boosting the robustness of recognition models. Central to overcoming the hurdles of diverse facial orientations and expressions are technologies focused on face fractalisation and the adjustment of face rotation and pose. These methods collectively represent the forefront of efforts to enhance the precision and dependability of facial recognition technology.

In the domain of facial recognition technology, especially for "in the wild" scenarios, a variety of approaches have been proposed in literature that aimed at improving recognition accuracy amidst challenges posed by variability in facial orientation and expression. The technique of face frontalisation and pose correction leverages generative adversarial networks to convert profile images into neutral-pose face images, significantly enhancing face recognition performance under extensive pose variations. In addressing face frontalisation and pose correction, FF-GAN, detailed in [1], utilizes a GAN combined with a deep 3D Morphable Model (3DMM) to generate neutral-pose face images from profiles in unconstrained settings with large pose variations. Similarly, the Feature-Improving Generative Adversarial Network (FI-GAN) as described in [2], aligns profile face features with frontal views to enhance face recognition under large pose variations.

Adjusting face rotation and viewpoint enhances facial image alignment, using advanced techniques for better training and synthesis. The previously published paper on Disentangled Representation-learning Wasserstein GAN (DR-WGAN) [3] expands the DR-GAN model for face recognition and synthesis. This model focuses on overcoming challenges in disentangled facial representation learning and utilizes a nonlinear 3D Morphable Model (3DMM) to augment data, aimed at improving training effectiveness. Strategies for face synthesis and recognition tackle occlusions and pose variations, focusing on enhancing identity and texture information for more reliable recognition. The introduction of BoostGAN [4] targets the de-occlusion, frontalisation, and recognition of occluded profile faces. This method is designed to enhance identity and texture information capture through the use of a deep GAN for coarse synthesis coupled with a shallow net for fine detail, addressing occlusions and pose variations. Developing high-resolution face frontalisation involves sophisticated texture mapping to produce realistic, identity-preserving frontal views from profiles. The High Fidelity Pose Invariant Model (HF-PIM) [5] approaches face frontalisation by synthesizing a frontal view from a profile image. It employs a texture warping procedure and a dense correspondence field for realistic texture mapping, aiming to provide identity-preserving results in high resolution and improve the process of frontalisation. Data augmentation in facial images is essential for improving training datasets, employing techniques to create realistic, rotated faces from single views, thus boosting recognition system robustness. Methods like CAPG-GAN [6] and the "Rotate-and-Render" [7] focus on generating realistic rotated faces from single-view images. These approaches employ various techniques, including landmark heatmaps and 3D modelling, to achieve texture refinement and identity preservation, supporting data augmentation in face recognition systems.

Despite the good performance of GANs and 3D reconstruction, they entail complexities that often hinder their deployment in real-world applications due to high computational demands. Recognising these limitations, this paper sets out to explore alternative strategies that can efficiently handle incomplete visual information in facial images without relying on resource-intensive generative models.

In this context, our research marks the beginning of an extensive investigation aimed at determining the most suitable approach for dataset selection and model development that promises both computational efficiency and effectiveness in uncontrolled conditions. This paper primarily focuses on laying the groundwork for future studies, outlining the potential benefits of employing specific AI solutions over traditional generative models for facial image processing tasks [1] [8] [9].

1.1. Challenges and Constraint

Based on current approaches, which rely on generative solutions, there could be limitations in handling the intrinsic variability and unpredictability of real-world environments. Computational efficiency, scalability, and the need for real-time processing stand out as primary challenges. Additionally, the complexity of extreme environmental conditions, including varied poses and lighting, further strains recognition accuracy. The dependency on extensive and diverse training datasets restricts model adaptability and robustness, while achieving high-quality, photorealistic facial rendering amidst occlusions or extreme angles remains a daunting task. These factors collectively highlight the critical need for advancing research towards more versatile, efficient, and accurate facial recognition technologies capable of navigating the challenges presented by "in the wild" scenarios. Consequently, there is a pressing need for lightweight solutions that can perform analysis with fewer image generations, akin to advancements seen in GAN solutions, offering a pathway to more efficient and effective facial recognition capabilities.

The quest for computational efficiency and scalability is a notable challenge within the Pose Invariant Model (PIM) landscape, where significant optimizations are needed to enhance suitability for real-time applications and deployment on various platforms, especially those with constrained resources [9]. Models often struggle to handle the extreme conditions of diverse poses, occlusions, and lighting variations that prevail in unregulated environments. This issue substantially impedes the ability to accurately reconstruct facial features, marking a critical area for improvement [1] [8]. A pronounced need for generalisation is evident as models excel in controlled settings yet fall short in adapting to the vast spectrum of real-world conditions, including variations in poses, expressions, and lighting. This gap underscores the necessity for enhanced adaptability [1] [8]. Deep learning models' effectiveness is closely tied to the breadth and diversity of their training data. The challenge of securing comprehensive, varied datasets poses a significant barrier, especially in scenarios where acquiring such data is challenging [9]. The computational intensity of models, notably FF-GAN and HF-PIM, places a heavy burden on resources, limiting their deployment and scalability in resource-constrained environments. This complexity demands a streamlined approach to ensure broader applicability [1] [8]. Achieving high-quality, photorealistic facial frontalisation from profile views, particularly under extreme conditions of poses or occlusions, remains a daunting challenge. This limitation directly impacts the visual quality and effectiveness of recognition tasks, necessitating innovative solutions to overcome these hurdles [1] [8]. These limitations, underline the areas where future research is directed. The focus is on enhancing model efficiency, improving generalisation to diverse conditions, reducing dependency on extensive training datasets, and addressing the challenges posed by extreme environmental conditions and occlusions.

2. Tools, Datasets and Models

2.1. Computational Infrastructure and Data Management

Our computational research leverages a high-performance workstation with an 11th Gen Intel® Core™ i7-11700 8-core processor and 32GiB of DDR4 RAM, optimized for swift data processing and effective multitasking. An NVIDIA GeForce RTX 2080 Ti GPU enhances our capacity for graphics-intensive tasks and parallel computations. The system operates on Ubuntu 20, providing a stable and flexible environment for various scientific applications. For our machine learning experiments, we utilise Keras with TensorFlow as the backend, facilitating model development and training. Data for these experiments are sourced from the Kaggle repository, ensuring access to a comprehensive and diverse dataset collection. This setup ensures robust performance across our computational research activities. Table 1 concisely outlines the specifications of our computational setup.

Table 1. Image Preprocessing Parameters for Data Augmentation

Component	Applied Value
Processor	11th Gen Intel® Core™ i7-11700 (8 cores)
Memory	32GiB DDR4 RAM
GPU	NVIDIA GeForce RTX 2080 Ti
Operating System	Ubuntu 20
Machine Learning Framework	Keras with TensorFlow
Data Source	Kaggle repository

2.2. CelebA Dataset for Facial Recognition in the Wild

In our approach, we utilised the CelebFaces Attributes (CelebA) Dataset [10], which consists of over 200,000 celebrity images. Each image is annotated with 40 facial attributes, covering a broad spectrum of facial characteristics, from expressions to accessory-related features. This extensive collection supports a variety of supervised learning projects aimed at attribute prediction. Access to CelebA is regulated to ensure ethical use, particularly in relation to the privacy and rights of the individuals represented in the dataset. Furthermore, CelebA includes 5 landmark locations per image, marking critical facial points such as the eyes, nose, and mouth corners. This level of detail is crucial for tasks requiring precise facial feature localisation, including facial alignment and manipulation. CelebA has significantly contributed to the development and improvement of algorithms for facial attribute recognition, face detection, and facial landmark detection. Additionally, it provides a foundation for more complex applications, such as face editing and synthesis. This dataset serves as a robust resource for training and evaluating machine learning models in the field of facial analysis. This dataset, rich in face identities and attributes, is key for advancing face attribute prediction in 'in the wild' scenarios. Its diversity helps in developing models that accurately recognise attributes amid real-world complexities, demonstrating its essential role in refining facial recognition technologies.

2.3. Enhancing Image Recognition with VGGNet Architecture

We chose VGGNet [11] as our training framework due to its straightforward yet effective approach to tasks such as feature extraction, classification, and localisation. This architecture distinguishes itself with a series of 3x3 convolutional layers that deepen progressively, with max pooling handling the reduction in volume. Two dense layers, each hosting 4096 nodes, follow this setup and culminate in a softmax classifier. The VGGNet investigation highlighted the importance of depth in the network's structure for superior performance. Experiments with network depths ranging from 11 to 19 layers established that a deeper configuration could significantly improve accuracy, provided it effectively addresses the computational challenges. Among the suite of models developed, VGG16 and VGG19 are particularly noteworthy, comprising 16 and 19 layers respectively. Although VGGNet requires more computational resources than other CNN models, such as Inception, its widespread adoption within the research community and its role as a comparative benchmark for new models highlight its efficacy and elegance of design.

2.4. Harnessing the VGG Face Dataset for Advanced Facial Analysis

We adopted the VGG16 model, a leading deep learning architecture pre-trained on the extensive VGG Face Dataset [12]. We rigorously trained this particular model, renowned for its depth with 16 distinct layers, using over 2.6 million images from 2,622 different identities found within the VGG Face Dataset. Such a detailed training process has equipped the model with a profound ability to recognise a wide range of facial attributes, including minor expressions, age progression, and accessory-induced variations. Mastery over these facial nuances is essential for executing advanced facial recognition tasks with high precision. By starting with the VGG-16 model, pre-trained on the VGG Face Dataset, we laid the solid groundwork for a supervised learning project focused on facial recognition. The aim is not only to refine existing security and personal device authentication mechanisms but also to enhance the model's ability to distinguish between individual faces accurately. To tailor the model more closely to real-world applications, we plan to further fine-tune it with a new dataset, which is CelebA [10]. We designed this subsequent step to adapt the model for enhanced performance in "in-the-wild" scenarios, where the complexity and unpredictability of environments pose additional challenges to facial recognition technologies. The capability of the VGG-16 model to process and recognise facial features effectively highlights its potential to improve real-time face tracking and create more personalised interactions in digital environments. Through the initial training with the VGG Face Dataset and the subsequent fine-tuning with new, diverse data, we aim to enhance the model's facial recognition and make it more adaptable and reliable in everyday scenarios.

3. Experimental Work

3.1. VGG-16 for Facial Recognition

We employed three distinct approaches to investigate the efficacy of the VGG-16 model in facial recognition tasks. Our first method involved training the VGG-16 model from scratch using the CelebA dataset (VGG16), aiming to establish a foundational understanding of its capabilities. Secondly, we implemented a transfer learning technique, starting with a pre-trained VGG-16 model on the extensive ImageNet dataset, and fine-tuning it using the CelebA dataset (VGGImageNet). This strategy sought to adapt the pre-acquired general visual recognition abilities of the model to our specific task. Lastly, our third approach utilized a VGG-face model, already pre-trained for facial recognition tasks using the VGG-face dataset, and further refined it through transfer learning with the CelebA dataset (VGGFace). We were able to compare the effects of direct training with the more complex benefits of transfer learning from both a general and a task-specific model that had already been trained.

We refined three models using the CelebA dataset, beginning with transfer learning and data augmentation to enhance their initial capabilities. After selectively freezing and fine-tuning specific layers, we evaluated their performance to identify the most effective model. We then optimized this through reducing unnecessary parameters, resulting in a streamlined and efficient model with improved accuracy.

3.2. Evaluating VGG-16's Performance in Facial Recognition

For the data pre-processing and augmentation, we implemented horizontal flipping to generate mirrored images, thereby expanding the dataset's variance. This step is crucial for face recognition in uncontrolled environments where subjects may appear in varying orientations. Additionally, we applied shifting transformations, with a width and height shift range of 0.2, as well as a zoom range of 0.2, to emulate common changes in perspective and distance in real-world scenarios. These transformations prepare the model to handle the diverse positions and scales of faces found in natural settings. Moreover, we normalized the pixel values with a scaling factor of 1/255, which is instrumental in aiding the model's faster convergence by maintaining numerical stability during computations. Effectively improving the face recognition model's robustness and generalisability, this thorough pre-processing pipeline lets it work reliably in a wide range of real-world situations. Table 2 displays the list of data preparation steps applied to the image data for model generalisation.

Table 2. Image Pre-processing Parameters for Data Augmentation

Pre-processing Step	Applied Value
Rotation	90 degrees
Width Shift	20%
Height Shift	20%
Zoom	20%
Pixel Value Rescaling	1/255

3.3. A Structured Approach to Model Training

The training setup for the models involved a structured approach, starting with the application of transfer learning techniques using the CelebA dataset, which was conducted over 50 epochs. During the experiment, the dataset was divided such that 80% was allocated for training purposes, while the remaining 20% was used for testing. Subsequently, we froze all layers except the last 10 to refine the model's adaptability and focus the learning process. This step was critical in maintaining the learned features while allowing the model's output layers to be adjusted to better suit the specific task at hand. After freezing the layers, we initiated a fine-tuning phase that spanned 50 epochs. This phase is aimed at optimizing the performance of the unfrozen layers, thereby enhancing the model's accuracy and its ability to generalise over unseen data. We evaluated the models after the fine-tuning process to determine which one performed the best. This leading model then underwent a process of parameter reduction. We meticulously removed some of the parameters to further improve the model's accuracy, ensuring it not only performed well but also operated with greater efficiency. This comprehensive approach to training and refining the models was instrumental in achieving optimal results. Figure 1 shows a step-by-step approach to generate the best model.

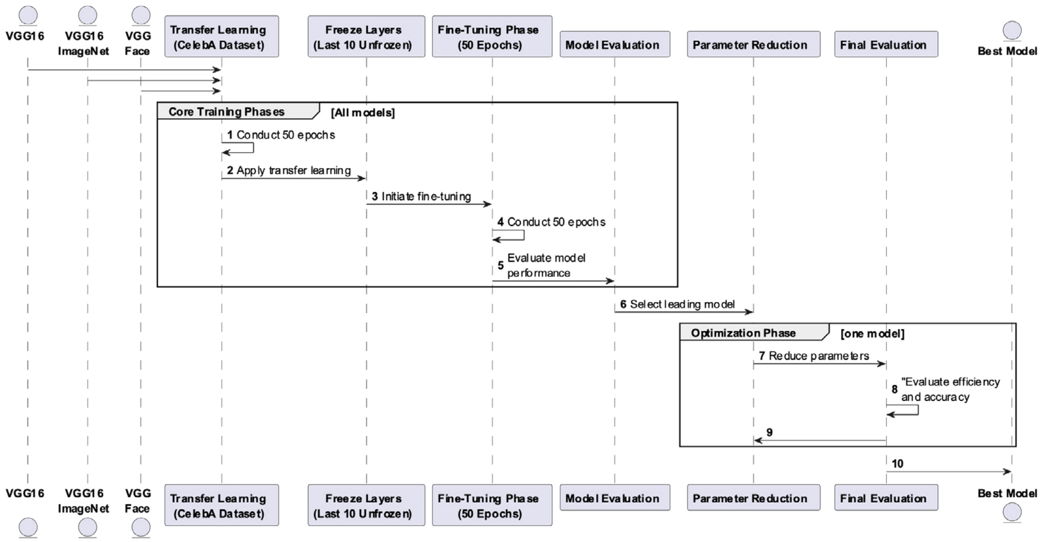

Fig. 1. Model Training and Optimization Process.

4. Results

The results of this study are shown in Table 3. The accuracy metrics derived from the different training approaches of the VGG-16 model—VGG16, VGGImageNet, and VGGFace—illustrate varying levels of performance enhancement attributable to each methodology. The model trained from scratch, VGG16, achieved an accuracy of 74.53%, serving as a baseline. The VGGImageNet model, which was pre-trained on the ImageNet dataset and fine-tuned on CelebA, showed a significant improvement in accuracy, achieving 80.08%. This indicates the effectiveness of applying transfer learning from a generalist model to a specific task. The highest accuracy was observed in the VGGFace model, at 84.65%, which was specifically pre-trained for facial recognition tasks and further fine-tuned on CelebA. This underscores the advantage of using a specialised pre-trained model where domain-specific features are more aligned with the task, resulting in superior performance. These results collectively demonstrate that transfer learning, particularly from a task-specific pre-trained model, can substantially optimize performance in facial recognition applications.

Table 3. Accuracy of VGG-16 Models on CelebA Dataset

Model	Accuracy (%)
VGG16	74.53
VGGImageNet	80.08
VGGFace	84.65

The experiment was conducted on distinctive models under a two-phase training group. Initially, all layers of the network were trained for half of the epochs. For the remaining epochs, we shifted strategies, freezing all but the last 10 layers of the model to refine the higher-level feature representations. Additionally, we implemented early stopping to prevent overtraining and to conclude training when the validation score ceased to improve. Figure 2 shows the experiment's output results.

Throughout the training session, the loss metrics consistently decreased, reflecting the model's improving proficiency in making accurate predictions. The initial phase of training all layers resulted in a significant drop in loss, while the subsequent phase of focused training on the last 10 layers allowed for fine-tuning that continued this trend.

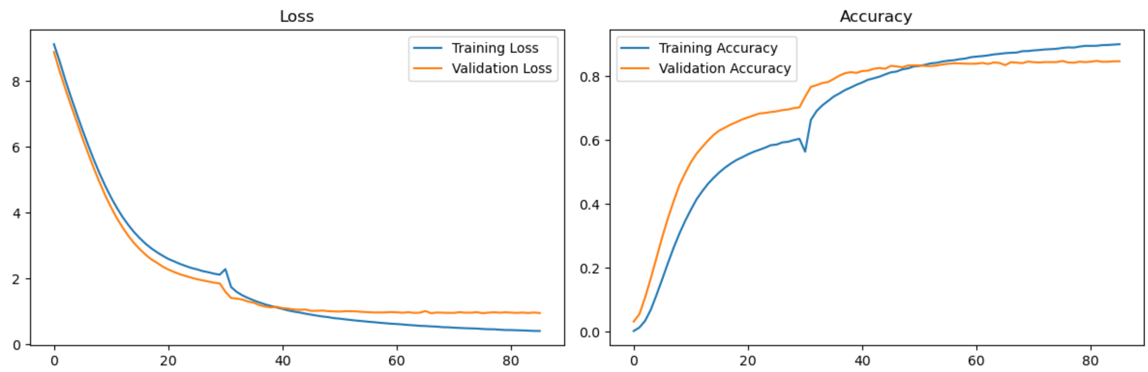

Fig. 2. Adaptive Learning Curves: Accuracy and Loss Values

Training loss exhibited a substantial decrease during the unfrozen phase and continued to decline at a steadier rate once the selective training was implemented. *Validation loss* closely mirrored the training loss, decreasing substantially during the full-network training and plateauing during the focused training phase, underscoring the model's generalisation capabilities.

The accuracy metrics demonstrated considerable gains, indicative of the model's progressive learning trajectory. The full-network training phase produced a swift rise in accuracy, while the focused training phase of the last 10 layers saw a continued but more gradual increase. *Training accuracy* initially saw rapid improvements, which then progressed at a moderate pace following the shift to selective layer training. *Validation accuracy* followed a similar pattern, with a notable increase during the early epochs, levelling off as the model underwent the fine-tuning process.

The dual-phase training approach, combined with early stopping, resulted in a model that enhanced its predictive abilities while maintaining robust generalisation performance. The clear reduction in both training and validation losses—along with the rising accuracy rates during the initial unfrozen training phase and the subsequent fine-tuning—demonstrates a successful training strategy. In the second phase, the application of early stopping ensured that training ceased at an optimal point, avoiding unnecessary computations and potential overfitting. This approach allowed for an efficient use of computational resources while achieving high performance.

Our findings are significant within the context of existing literature, particularly when considering that the majority of studies rely on generative models. In contrast, our approach utilises simpler, deep learning models or off-the-shelf solutions without incorporating generative techniques such as GANs and 3D/2D construction. This distinction underlines the effectiveness of our methodology compared to the norm in this field of research.

Our approach yields accuracy that is 5–10% lower than alternative methods documented in the literature. We intentionally designed this performance discrepancy to reduce the time complexity and computational burden associated with the generative techniques reported in the literature.

Looking ahead, we aim to adopt additional AI methodologies, such as symbolic AI, which involve defining a specific set of rules. This will allow us to use our current approach as an approximation function, thereby improving the accuracy and effectiveness of our predictive outcomes.

Our goal is to complement our existing solution with alternative methods that are less computationally intensive than the generative solutions found in the current literature. This strategy not only serves as a viable alternative to the generative approach but also enables the deployment of our solutions in real-time applications, enhancing both efficiency and practicality. For instance, our streamlined model could drastically cut down the time required for processing and analysing video feeds in building surveillance systems. This would facilitate the immediate identification and alerting of security incidents stemming from unauthorised access or unusual activities, thereby guaranteeing prompt action. By reducing the latency in processing and decision-making, our model facilitates a more secure and responsive surveillance operation, making buildings safer and more efficiently monitored.

5. Conclusion

This study presents a robust evaluation of alternative facial recognition technologies that prioritize computational efficiency and practical applicability in uncontrolled environments. By utilizing the VGG-16 model in a dual-phase training approach, we have demonstrated that less computationally intensive methods can still achieve significant effectiveness, providing a feasible solution for real-world applications such as surveillance systems. The results obtained from training the VGG-16 model using the CelebA dataset illustrate that, despite a slight reduction in accuracy, the benefits of reduced computational demand and enhanced real-time performance present a compelling alternative to more traditional, resource-intensive generative methods like GANs and 3D reconstruction.

Our findings reinforce the notion that efficiency and adaptability in facial recognition technology do not necessarily require the highest accuracy obtainable, but rather an optimal balance between performance and practical deployment capabilities. The methodologies employed show promising potential for extending facial recognition applications beyond traditional setups, facilitating wider adoption in scenarios where rapid processing and low computational overhead are crucial. Future research will focus on refining these techniques and exploring further integration with emerging AI methodologies to create more sophisticated, yet efficient, facial recognition systems. This study paves the way for continued innovation in the field, aiming to achieve more adaptable, efficient, and reliable facial recognition solutions suitable for dynamic and unpredictable real-world environments. In our future work, we plan to conduct further experiments with diverse datasets to enhance and generalize our solution.

Future research will also focus on refining these techniques and integrating emerging AI methodologies to create more sophisticated and efficient facial recognition systems. We plan to incorporate attention mechanisms to help the model focus on relevant facial features, improving its ability to handle variations in orientation, expression, and lighting. Additionally, we will explore ensemble learning techniques that combine multiple VGG models to enhance accuracy without sacrificing computational efficiency. This study paves the way for continued innovation in the field, aiming to achieve more adaptable, efficient, and reliable facial recognition solutions suitable for dynamic and unpredictable real-world environments.

Acknowledgements

This research is funded by the Engineering and Physical Sciences Research Council (EPSRC) EP/R013144/1 (SAUSE).

References

[1] Yin, Xi, Yu, Xiang, Sohn, Kihyuk, Liu, Xiaoming, and Chandraker, Manmohan. (2017) "Towards large-pose face frontalization in the wild," in Proceedings of the IEEE International Conference on Computer Vision, pp. 3990–3999.

[2] Rong, Changle, Zhang, Xingming, and Lin, Yubei. (2020) "Feature-improving generative adversarial network for face frontalization," IEEE Access, vol. 8, pp. 68842–68851.

[3] Hsu, Gee-Sern Jison, Tang, Chia-Hao, and Yap, Moi Hoon. (2019) "Face synthesis and recognition using disentangled representation-learning wasserstein gan," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2371–2379.

[4] Duan, Qingyan, and Zhang, Lei. (2019) "BoostGAN for occlusive profile face frontalization and recognition," arXiv preprint arXiv:1902.09782.

[5] Cao, Jie, Hu, Yibo, Zhang, Hongwen, He, Ran, and Sun, Zhenan. (2018) "Learning a high fidelity pose invariant model for high-resolution face frontalization," Advances in Neural Information Processing Systems, vol. 31.

[6] Hu, Yibo, Wu, Xiang, Yu, Bing, He, Ran, and Sun, Zhenan. (2018) "Pose-guided photorealistic face rotation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8398–8406

[7] Zhou, Hang, Liu, Jihao, Liu, Ziwei, Liu, Yu, and Wang, Xiaogang. (2020) "Rotate-and-render: Unsupervised photorealistic face rotation from single-view images," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5911–5920.

[8] Cao, Jie, Hu, Yibo, Zhang, Hongwen, He, Ran, and Sun, Zhenan. (2020) "Towards high fidelity face frontalization in the wild," International Journal of Computer Vision, vol. 128, pp. 1485–1504, Springer.

[9] Zhao, Jian, Cheng, Yu, Xu, Yan, Xiong, Lin, Li, Jianshu, Zhao, Fang, Jayashree, Karlekar, Pranata, Sugiri, Shen, Shengmei, Xing, Junliang, et al. (2018) "Towards pose invariant face recognition in the wild," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2207–2216.

[10] Liu, Ziwei, Luo, Ping, Wang, Xiaogang, and Tang, Xiaoou. (2015) "Deep learning face attributes in the wild," in Proceedings of the IEEE International Conference on Computer Vision, pp. 3730–3738.

[11] Simonyan, Karen, and Zisserman, Andrew. (2014) "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556.

[12] Parkhi, Omkar, Vedaldi, Andrea, and Zisserman, Andrew. (2015) "Deep face recognition," in BMVC 2015-Proceedings of the British Machine Vision Conference 2015.

[13] Li, Peipei, Wu, Xiang, Hu, Yibo, He, Ran, and Sun, Zhenan. (2019) "M2FPA: A multi-yaw multi-pitch high-quality dataset and benchmark for facial pose analysis," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10043–10051.

[14] Li, Yi, Huang, Huaibo, He, Ran, Tan, Tieniu, et al. (2020) "Topology-Variant Synthesis," in Heterogeneous

Facial Analysis and Synthesis, pp. 53–94, Springer.

[15] Luo, Mandi, Cao, Jie, Ma, Xin, Zhang, Xiaoyu, and He, Ran. (2021) "FA-GAN: Face augmentation GAN for deformation-invariant face recognition," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2341–2355, IEEE.

[16] Yin, Yu, Jiang, Songyao, Robinson, Joseph P, and Fu, Yun. (2020) "Dual-attention GAN for large-pose face frontalization," in 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), pp. 249–256.

[17] Ibrahim, Mostafa EA, Abbas, Qaisar, Daadaa, Yassine, and Ahmed, Alaa ES. (2023) "A Novel PPG-Based Biometric Authentication System Using a Hybrid CVT-ConvMixer Architecture with Dense and Self-Attention Layers," Sensors, vol. 24, no. 1, p. 15, MDPI.