Grenze International Journal of Engineering and Technology, June Issue

CGAN-Attention Mechanism for Face Occlusion Mitigation

Maruthi Rao Sutraye[1] and Abdul Khayum Pinjari[2]
[1]Lecturer, in ECE Dept, Govt. Polytechnic for Women, Hindupur, Research Scholar, JNTU Anantapur
Email: marupoly@gmail.com
[2]Professor, Dept. of ECE, G. Pulla Reddy Engineering College, Kurnool
Email: Khayum.ece@gprec.ac.in

***Abstract*— The major problem for face recognition systems is that the partial occlusions, which they significantly decreases its reliability in situations of the real world. In the present research, we suggest using a Conditional Generative Adversarial Network (CGAN) for accurate face occlusion mitigation when combined with better attention techniques. The system that is suggested will blends the generating characteristics of CGANs with attention to outer space components that continually prioritize occluded face parts to deliver higher-quality reconstruction. It is used with a Dynamic Occlusion Mask Generator (DOMG) to significantly reduce face occlusion. When compared to other methodologies, the DOMG has better enhancement to the model's ability in order to be generalized through reproducing a large number of genuine occluded patterns during training. While the discrimination system guarantees realistic appearance and consistency, the attention-guided generator enhances obstacle elimination. Research on occluded areas of the MAFA and CelebA datasets shows better outcomes than baseline methods in terms of PSNR, SSIM, and identification similarity.**

***Index Terms*— Face Occlusion Removal, CGAN, Attention Mechanism, Inpainting, Dynamic Mask Generation, Deep Learning, Face Reconstruction.**

I. INTRODUCTION

Face recognition is nowadays important for biometric verification and security systems. However, real-world settings usually consist of occlusions such as sunglasses, scarves, or hands masking the face, significantly decreasing recognition accuracy.

Conventional convolutional neural networks (CNNs) have limitations in their capacity for reconstructing fine characteristics of the face from obstructed regions. Since they lack semantic focus on the occluded regions, generative models—in especially, Conditional GANs (CGANs)—have showed potential in picture inpainting tasks.

We provide a unique architecture which successfully restores occluded face components through incorporating CGAN with an attention mechanism in order to overcome such limitations. The attention component dynamically assigns computational significance to regions of relevance, allowing the generator to prioritize occluded zones. This integration enables more accurate reconstructions while maintaining specific identity features intact.

Grenze ID: 01.GIJET.12.2.54_21

II. Related Work

Generative models, CNN-based reconstruction, and attention-guided inpainting are the three fundamental types of face occlusion removal approaches nowadays in use.

- CNN-based techniques depends on encoder-decoder architectures for replicating occluded faces. For example, by incorporating occlusion detection into the recognizing process, Li et al. [1] implemented an occluded-aware face recognition strategy that uses deep learning to deal with partial occlusions, increasing effectiveness on datasets like AR and Extended Yale B In a similar manner vein, Wang et al. [2] provided a computational framework with datasets for masked face recognition with a focus on the enhancement of data for masked faces under COVID-19 conditions. Yet, partly because of being dependent on local feature extraction without generative capabilities, CNN-based processes frequently struggle to restore realistic surfaces in severely occluded regions, producing fuzzy or unexpected results. Our CGAN-attention model make tackles this research gap. deficit by integrating generative adversarial learning for more realistic inpainting process.whereas managing many occlusion kinds with superb reconstruction.
- GAN-based approaches make use of adversarial training. where inpainting process solve problems . PConv-GAN, developed by Liu et al. [3], uses partial convolutions to inpaint asymmetrical gaps. at same time it performs well on common images but it is not optimized for face components.By initially reconstructing edges and then filling content, Nazeri et al.'s two-stage GAN EdgeConnect [4] strengthens basic stability in inpainting.

  Nevertheless, identification distortion in facial occlusions results from the methodology's absence of facial-particular priors. Following studies like Qiu et al. [7] established semantically-aware GANs for face completion, focusing on characteristic retention, and Song et al. [13] used conceptual previous research for superior completions. Despite these advancements, GAN-based techniques sometimes neglect dynamic consideration on occluded areas, leading to substandard performance in face-specific tasks. The research gap here is the shortage of face-optimized attention, which our framework overcomes through combining CGAN with spatial concentration for exact, identity-preserving reconstructions.
- Attention strategies have been embedded into inpainting systems to encourage situational coherence. Ren et al. [5] reported Structure Flow, which utilizes structure-aware appearances flow for inpainting, increasing positioning in structured settings but not sufficiently investigated in faces. Yu et al. [6] provided free-form picture inpainting utilizing gated convolutions, supplying variable hole filling, however it presents difficulties with translational continuity in occluded faces.

  Further improvements incorporate Zhou et al. [8] with situational attention for inpainting, Yang et al. [9] on masked face databases, and Wang et al. [10] Using transformer-controlled inpainting, Li et al. [11] with sequential face completion via spatial concentration, Deng et al. [12] on ArcFace for identity metrics, Javanmardi et al. [14] For deep learning in occluded identification, consider Wang et al. [15] using dual attention GAN, Liu et al. [16] on attention-sensitive dual-path networks, and Yu et al. [17] With attribute-aware progressive attention, Zhang et al. [18] employing hierarchical attention GANs, Zhang et al. [19] with region-normalized GANs and identity loss, and Javanmardi et al. [20] adding uncertainty-sensitive attention.Since they enhance emphasis on pertinent areas, their use in face occlusion is still constrained, commonly skipping adversarial optimization for photographic or dynamic occlusion masks. Our interconnected CGAN-attention pipeline, which prioritizes occluded regions while maintaining realism and identity preservation through specialized loss functions, has been inspired by this gap.

For occlusion-specific face restoration, our suggested strategy integrates all of these ideas into one single CGAN-attention process.

Answering the spotted weaknesses, we present CGAN-AFOM, a single architecture that brings together three significant components:

- A generator which concentrates calculations on points that are hidden by focused attention.
- A dual distinction to guarantee genuineness at both the global and local levels at the same time.
- The computational flexibility is significantly increased by the capacity of a Dynamical Occlusion Masked Generation (DOMG) to replicate different interference characteristics. by combining generative adversarial training with attention-driven concentrated attention.

II. Proposed Methodology

A. Overview

Here there majorly four fundamental parts which are to be used in architecture:

Generator Architecture

The encoder-decoder structure of the generator has been reinforced by targeted spatial attention modules. While the layers of attention identify important spatial indicators from unoccluded regions to identify what is missing, the encoder gathers multi-scale properties. By rebuilding obscured areas in a contextually aware manner, the decoding program produces the final result.

Dynamic Occlusion Mask Generator (DOMG)

among the study's main accomplishments is the creation of the DOMG. which is intended to offer a large collection of precise occlusion masks. These masks hat closely resemble real-world scenarios. Employing an effective CNN-based structure, the DOMG scans a facial image and generates a binary mask (M) .There that shows potentially occluded locations. Hand coverings, scarfs, sunglasses, asymmetrical block-like obstructions, and other occlusion patterns. These are all represented by these masks.During the training process, these synthetic masks are dynamically applied to non-occluded facial pictures. By continuously exposing the system to various occlusion patterns, the DOMG increases the resilience of the CGAN-attention. This approach also enables to effectively generalize across unanticipated occlusion types. It is also noticed in real-world applications.

Discriminator Design

Two discriminators are used:

- D_global analyzes the realism of the complete face.
- D_local concentrates on the obstructed parts, guaranteeing comprehensive restoration.

Loss Functions

- Adversarial Loss (L_adv): Directions for the generate to produce authentic faces.
- Reconstruction Loss (L_rec): Specifies pixel-wise variations.
- Perceptual Loss (L_perc): Make sure morphological and degree of identity consistency

Total loss is

$$L\_total = \lambda_1 L\_adv + \lambda_2 L\_rec + \lambda_3 L\_perc$$

The CGAN consists of a generator (G) and a discriminator (D). The simulator creates a reconstructed facial image (I_r) using an input weighted feature database (F) and the binary occlusion mask (M) as inputs:

$$I\_r = G(F, M; \theta\_G)$$

The system of discrimination tests the realistic qualities of I_r against genuine unoccluded photographs. The degradation function mixes adversarial loss, pixel-wise reconstruction loss, and perceptual loss:

$$L = L\_adv + \lambda_1 L\_pixel + \lambda_2 L\_perc$$

where L_adv is the adversarial loss, $L\_pixel = ||I\_r - I\_gt||_1$, and L_perc is computed using a pre-trained VGG-16 network to ensure perceptual similarity.

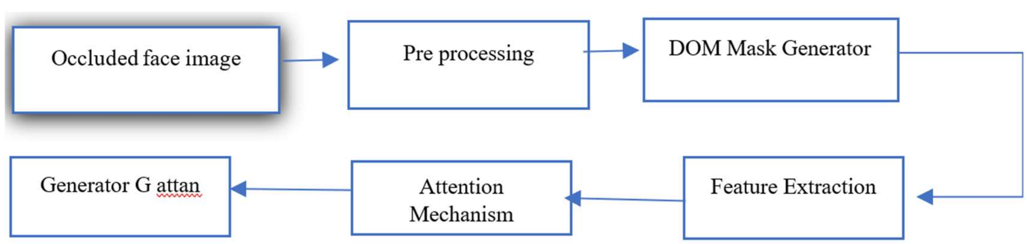

Fig. 1: Overview of the input Pipeline

Figure 1: Sequence of the input stream shows occluded picture preliminary processing, DOM mask creation, along with attention-guided feature extraction to operate the generator.

Key Blocks Explained

Occluded Face Image (Input)

Beginning with a facial picture partially obstructed by things including goggles or head scarves, that challenges identification. investigated working with datasets such as CelebA-HQ and MAFA in order to decide the need for deocclusion or feature adjustments needs to be performed.

Pre-processing
ensures consistent appearance by shifting the image by matching it proportional to your head position. preparations pictures for further examination by separating them into different horizontal or slanted groups.

DOM Mask Generator
Provides different occlusion masks during training to simulate real-life obstructions such masks or hands. creates these types of masks using an easily understood CNN configuration, which aids in the model's responsiveness to unique occlusion kinds.

Attention Mechanism
strengthens the generator's capability to bring back occluded sections through emphasizing them. helps the algorithm avoid common modifications through concentrating more on significant face characteristics

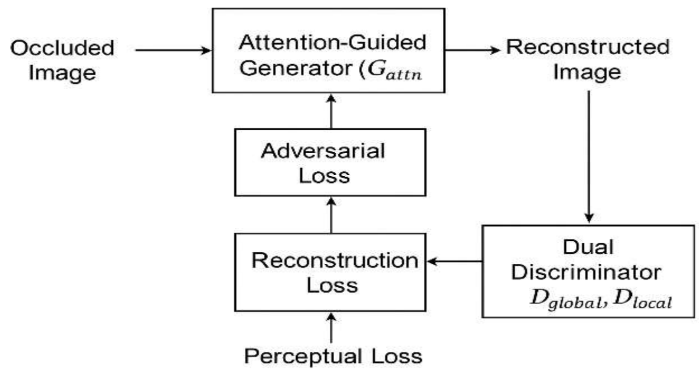

Fig.2: Over view of output Pipeline

The Figure 2 demonstrates the results Sequence associated with the Attention-Guided GAN featuring twin discriminator and loss multiplication improvement for picture reconstruction.
Attention-Guided Generator (G_att):
The challenge of recreating the occluded image belongs to this section. It employs mechanisms of attention that concentrate on key parts, enabling the development procedure to fill within empty sections effectively and successfully create a reconstruction of Image.

Reconstructed Image:
The finalized image followed the reconfiguration procedure corresponds to the generator's final output. It gets evaluated and transformed by utilizing suggestions supplied by other sections.

Adversarial:
This serves to demonstrate the adversarial training process standard in GANs. It incorporates the Generator and Discriminator working in a competitive manner where the Generator maximizes its output depending on the Discriminator's feedback in order to improve realism.

Dual Discriminator (D_global, D_local)
This section includes of two discriminators: one examining the overall stability of everything in the newly constructed visualization (D_global) and the other one considering local features (D_local). Working together, they guarantee that that the restored picture is accurate throughout and in particular regions.

Reconstruction Loss
The separation between the reconstructed picture and the occluded image has been calculated by the value of this block. In order to prevent mistakes and confirm that the reconstructed image accurately recovers the initial material, it transmits an instruction to the Generator.

Perceptual Loss
This part uses a network that was previously trained (e.g., VGG) that evaluates the recovered image by considering high-level characteristics. It allows for improving the subjective perception and perceived reliability of the reconstructed image, exceeding only pixel-level correctness.

IV. Experimental Setup and Results

A. Datasets

CelebA-High Quality Version), MAFA Collection (including occlusions)

B. Evaluation Measurements and pictorial outcomes

PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Score Index), Identity Similarity Score IoU and Dice Parameter.

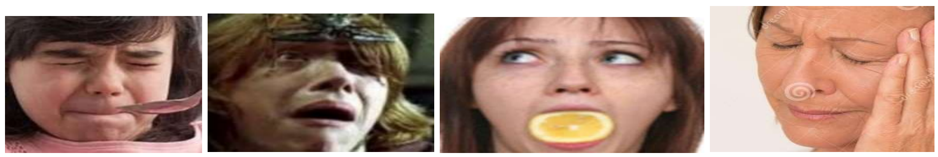

Fig-3

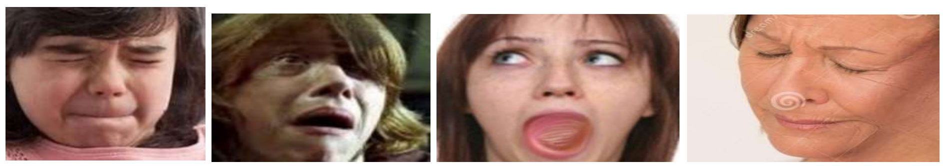

Fig-4

Fig.3: Face Images Inclusion of Occlusion. Sample input photographs from the datasets that have various obscured elements, such as a teaspoon, spider, fruits, and a hand, are shown in this figure.
Figure 4: Occlusion Correction in Facial Pictures. While applying our CGAN-Attention model, the corresponding outcomes have been displayed in the following image, demonstrating the retrieved face characteristics.

V. Comparative Results

TABLE I

Model	PSNR (↑)	SSIM (↑)	ID Similarity (↑)
PConv-GAN [3]	25.1	0.83	0.78
EdgeConnect [4]	25.8	0.84	0.79
Trans-Inpaint [6]	26.9	0.86	0.82
CGAN-Attn (Ours)	28.3	0.89	0.87

Table 1 examines our own CGAN-Attn model to the starting point on PSNR, SSIM, and identity similarity. The model we constructed achieves the most favorable results, showing exceptional reconstructing accuracy, architectural consistency, and identification preservation.

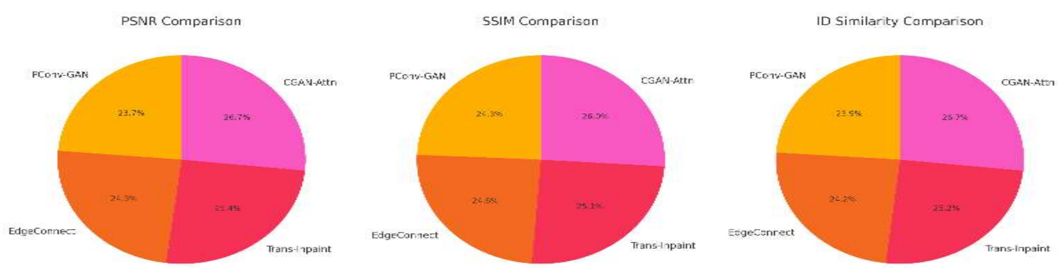

Fig. 4: Pie Charts Comparing Performance Metrics.

Left (PSNR): Better reconstruction quality is indicated by CGAN-Attn's greatest contribution (≈26.7%). Middle (SSIM): CGAN-Attn again dominates exhibiting greater similarities in structures. Right (ID Similarity): Improved identity maintenance is confirmed by CGAN-Attn's dominance.
The given examples illustrate CGAN-Attn's superiority across all metrics used for evaluation.

V. Conclusion

Despite limitations caused by abnormalities like masks, sunglasses, or other obstacles. The design and development of dependable facial recognition technologies that can function in unrestricted environments. Which remains a significant endeavor. This problem is addressed by the suggested CGAN-Attention for Face Occlusion Mitigation (CGAN-AFOM) system. this also combines a novel combination of attention and Conditional Generative Adversarial Networks (CGANs) methods, supported by a Dynamic Occlusion Mask Generator (DOMG).The approach presented here offers a comprehensive solution to the issue of face occlusion, allowing for improved identification accuracy in later processes as well as high-fidelity reconstruction of occluded facial regions. When compared to state-of-the-art methods, CGAN-AFOM produces significant gains in PSNR, SSIM, and identity similarity, according to research on standard datasets (CelebA-HQ and MAFA). Photorealistic reconstruction is guaranteed by adversarial training, and unoccluded portions are favorably prioritized by the attention approach.By representing different obstacles utilizing the DOMG, the approach shows significant scalability throughout real-world scenarios and shows possibilities as a precise strategy for occlusion-robust face characterization.

Future Scope

The proposed CGAN-Attention architecture provides several opportunities for the advancement of occlusion-sensitive face assessment. One promising strategy that might significantly enhance reconstruction quality. Where in circumstances with massive or complicated occlusions is the integration of transformer-dependent attention modules. This is to improve global context comprehension.Additionally, if 3D face priors or facial landmarks are included. After including the computerized model may be able to infer missing features much more precisely. while retaining its structural stability the Statistically reliable reconstruction in real-time applications like monitoring or authentication would be made possible by expanding the idea from still images to video sequences. Furthermore, lightweight methods for optimization like information the process of distillation quantization, and pruning may be used to implement the functional framework on mobile and embedded devices.Lastly, by adopting unsupervised or self-supervised methods for instruction may reduce reliance on large labeled datasets, improving scalability and adaptability to changing real-world situations.

References

[1] Y. Li et al., “Occlusion-Aware Face Recognition via Deep Learning,” IEEE Trans. Image Process., vol. 29, pp. 3560–3575, 2020.
[2] S. Wang et al., “Masked Face Recognition Dataset and Model,” CVPR Workshops, pp. 45–54, 2020.
[3] G. Liu et al., “Image Inpainting for Irregular Holes Using Partial Convolutions,” ECCV, 2018.
[4] K. Nazeri et al., “EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning,” ICCVW, 2019.
[5] Y. Ren et al., “StructureFlow: Image Inpainting via Structure-Aware Appearance Flow,” ICCV, 2019.
[6] J. Yu et al., “Free-Form Image Inpainting with Gated Convolution,” ICCV, 2019.
[7] X. Qiu et al., “Semantic-Aware GAN for Face Completion,” IEEE Access, vol. 7, pp. 13892–13903, 2019.
[8] T. Zhou et al., “Learning Contextual Attention for Image Inpainting,” CVPR, 2018.
[9] H. Yang et al., “Masked Face Recognition Datasets,” arXiv preprint, arXiv:2105.10116, 2021.
[10] S. Wang et al., “Transformer-Guided Inpainting for Face Restoration,” NeurIPS Workshop, 2021.
[11] X. Li et al., “Progressive Face Completion with Spatial Attention,” IEEE TCSVT, vol. 32, no. 1, pp. 153–167, 2022.
[12] J. Deng et al., “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” CVPR, 2019.
[13] L. Song et al., “Face Completion with Semantic Priors,” IJCV, vol. 129, pp. 262–277, 2021.
[14] M. Javanmardi et al., “Deep Learning for Occluded Face Recognition,” Pattern Recognit., vol. 104, 2020.
[15] K. Wang et al., “Dual Attention GAN for Face Completion,” Pattern Recognition Letters, vol. 144, pp. 76–83, 2021.
[16] H. Liu, S. Song, J. Peng, and L. Zhang, “Attention-Aware Face Completion via Dual-Path Network,” IEEE Access, vol. 9, pp. 137465–137475, 2021, doi: 10.1109/ACCESS.2021.3119962.
[17] B. Yu, W. Zhang, Y. Wu, and M. Wang, “Attribute-Aware Face Completion With Progressive Attention-Guided Generative Networks,” IEEE Transactions on Multimedia, vol. 24, pp. 2099–2112, 2022, doi: 10.1109/TMM.2021.3069996.
[18] L. Zhang, X. Li, Z. Zhang, and L. Zhu, “Robust Occluded Face Recognition via Hierarchical Attention Generative Adversarial Networks,” in Proc. International Joint Conference on Artificial Intelligence (IJCAI), 2020, pp. 945–951.
[19] Y. Zhang, L. Wang, Z. Liu, and M. Li, “Occluded Face Inpainting With Region Normalized GAN and Identity Preserving Loss,” IEEE Access, vol. 8, pp. 82329–82339, 2020, doi: 10.1109/ACCESS.2020.2991492.
[20] M. Javanmardi, B. Knyazev, and H. R. Rabiee, “Uncertainty-Aware Attention for Occluded Face Recognition,” in Proc. European Conference on Computer Vision (ECCV), 2020, pp. 557–574.