[1]Anjali Yeole
[2]Prathmesh Pawar
[3]Satyam Dubey
[4]Yash Sarang
[5]Arunim Chakraborty
[6]Tanvi Kate
[7]Madhusudhana Naidu
[8]Kshitij Shidore

Image Inpainting for Missing Facial Data Recovery in Security Settings

Abstract: - The *Pervasive* adoption of face masks during the pandemic of COVID-19 has introduced significant complexities for facial recognition systems, particularly in the realms of security and authentication. This paper tackles the intricate problem of incomplete facial data resulting from mask occlusion, a *pervasive* issue across various image processing and recognition domains. Our approach centers on a sophisticated high-resolution face image completion technique utilizing Generative Adversarial Networks (GANs). The Generator network, meticulously crafted in a U-Net architecture fashion, integrates crucial skip connections to retain vital spatial information from the U-Net's encoder path. We optimize our methodology through a holistic total Generator Loss function, augmented with Mean Absolute Error (MAE) loss and a regularization term (λ) to prevent overfitting. Training our model on the esteemed CelebA-HQ dataset, we generate synthetic masked faces by realistically simulating mask placements on original images. Our method showcases exceptional performance, achieving a Peak Signal-to-Noise Ratio (PSNR) of 22.25 and a Structural Similarity Index Measure (SSIM) of 0.874. These results surpass conventional GANs, non-learning-based patch-matching methods, and even certain diffusion-based techniques by a significant margin of 1.16%. This capability to reconstruct human faces despite masks plays a pivotal role in enhancing security protocols during the ongoing pandemic.

Keywords: Generative adversarial network; Face Image Inpainting; U-Net.

I. Introduction

The extensive adoption of face coverings as a preventive measure against COVID- 19 transmission presents a notable dilemma for law enforcement agencies. Facial recognition systems, integral for perpetrator identification, face diminished accuracy due to obscured facial features. This issue is especially worrisome as individuals resort to masks to obscure their identities while engaging in criminal activities. Law enforcement requires swift access to technologies capable of identifying individuals involved in unlawful acts or breaches of security, even when their faces are partially concealed by masks.

Image inpainting is a crucial aspect of computer vision, particularly in scenarios where incomplete or damaged images need restoration. One of the notable applications of image inpainting is in face reconstruction from masked images. This task involves filling in missing or obscured parts of a face image, which is essential for various applications such as privacy protection, forensic analysis, and digital image restoration. Conventional inpainting methods often rely on patch-based techniques using low-level features like RGB values or SIFT descriptors [12, 13]. However, these methods may struggle with complex structures and overlapping foreground objects in face images. Deep learning-based approaches, including GANs [5], have shown promise in accurately reconstructing facial features from incomplete data. By learning from large-scale datasets, GANs can generate realistic and coherent facial reconstructions, contributing significantly to the field of face image inpainting.

[1] Deputy HOD, AI&DS Dept., VESIT, anjali.yeole@ves.ac.in
[2] AI&DS Dept., VESIT, prathmeshpawar2001@gmail.com
[3] AI&DS Dept., VESIT, dsatyam5302@gmail.com
[4] AI&DS Dept., VESIT, *yash.sarang9@gmail.com*
[5] AI&DS Dept., VESIT, *arunimchakraborty@gmail.com*
[6] AI&DS Dept., VESIT, *tanvikate02@gmail.com*
[7] AI&DS Dept., VESIT, *mnaidu1025@gmail.com*
[8] AI&DS Dept., VESIT, *kshitijshidore19@gmail.com*
Corresponding Author: Prathmesh Pawar[2]: prathmeshpawar2001@gmail.com
Copyright © JES 2024 on-line: journal.esrgroups.org

Fig. 1: Central Intelligence Agency's Watchlist for Masked Offenders

While it is commonly understood that obscured or partially covered faces can reduce identification accuracy, there is a noticeable gap in research regarding the effects of facial coverings on perceptual face matching (PFM) tasks [2]. Previous studies exploring how occlusion impacts identification accuracy have primarily centered on tasks related to recognition memory, which differ significantly from PFM tasks where learned and test faces are presented sequentially [3]. These studies have emphasized the varying importance of facial features in recognition, with the peri- orbital region, especially the eyes, playing a more significant role in face recognition accuracy compared to other facial features like the nose, mouth, and chin [1, 3].

In the field of face image inpainting, researchers have explored various strategies to achieve high-quality reconstructions. Traditional approaches focused on patch matching and low-level feature synthesis, often resulting in perceptual defects and unrealistic outcomes, particularly with complex facial structures. Recent progress in deep learning-based inpainting methods [6, 7, 14, 15] has transformed face image inpainting into an end-to-end mapping task. Among these techniques, GAN-based methods excel in producing visually realistic and structurally coherent reconstructions. By utilizing adversarial training, GAN-based inpainting networks can generate novel facial features while maintaining attribute consistency and topological structure. However, challenges persist in accurately predicting structures within voids, especially in situations where foreground and background elements overlap. Overcoming these obstacles is essential for advancing the effectiveness and applicability of GANs-based approaches in face image inpainting.

Our main contribution to this face reconstruction research lies in the following key points:

- Proposing a novel approach that automatically removes masks from facial images and replaces the affected areas with detailed reconstructions, while preserving the original facial structure.
- Improving structural and visual consistency in the reconstructed faces was achieved by introducing a penalty for High Mean Absolute Error loss during generator training. This penalty helped the model focus on learning the unmasked regions of the face accurately, utilizing the U-Net architecture's Skip connections to facilitate a coarse-to-fine image completion process specifically for the deep missing regions.
- Addressing the challenge of limited data availability for masked images by curating a synthetic paired dataset, combining resources from the widely used CelebA-HQ [6] dataset and the Mask-The-Face [7] Script.

This paper employs a structured methodology, beginning with an introduction that outlines the scope of our research. Section II elaborates on the proposed system for reconstructing masked and damaged face images. In Section III, we delve into the methodology, detailing the techniques and algorithms utilized. Following this, Section IV presents the results of our experiments, while Section V concludes with significant findings and provides suggestions for future research directions.

II. Literature Review

Traditional image inpainting techniques, such as patch-based methods [18] and diffusion-based approaches [19], traditionally functioned by identifying and integrating similar image patches into the missing regions. Patch-based methods search for pixels in unaffected areas to complete the missing regions, relying on the availability of relevant content in these areas. Conversely, diffusion-based methods propagate neighboring content to fill in the gaps but may struggle with reconstructing meaningful structures, especially in larger missing areas. For example, Patch Match [20] is a rapid nearest-neighbor field algorithm widely utilized in various image editing applications, including inpainting. Although patch-based methods have their utility, they are limited in generating semantic or innovative content due to their reliance on basic features for patch matching.

In comparison, conventional techniques, notably convolutional neural networks (CNNs), have garnered substantial attention in various image processing applications, including image inpainting. The Context Encoder [21] stands out as an early CNN-based method for image inpainting, showing promising results. However, its effectiveness is primarily observed in fixed-size and low-resolution tasks, often leaving noticeable traces during repairs [22]. Recent advancements by researchers like Iizuka et al. [23] have improved the ability to repair irregularly shaped images and areas with significant defects. Nevertheless, these methods may require additional post-processing steps to achieve the desired repair effects. Wang et al. [24] have proposed further enhancements through a generative multi-column CNN architecture and a confidence-driven image inpainting algorithm, demonstrating impressive visual outcomes. However, challenges persist, particularly in dealing with complex datasets containing diverse objects or scenes.

Pathak et al. [25] utilized a context encoder within a GAN framework for image restoration tasks. Their approach, dividing the generator into an encoder and decoder, emphasizes feature compression and extraction from incomplete images for restoration. Although effective, this method's utilization of generation antagonism losses primarily addresses local information rather than holistic semantic coherence within incomplete regions. Subsequent studies by Iizuka et al. [23] introduced a global-local double discriminator to augment the context encoder's performance, ensuring both accurate restoration in incomplete areas and coherence in the final results. However, challenges persist in accurately predicting repairs for large areas in facial images.

Yang et al. [26] introduced networks for content and texture generation aimed at addressing high-resolution image restoration tasks. Their approach, which utilizes separate networks for semantic inference and high-frequency detail generation, successfully addressed challenges encountered in prior methods, especially in high- resolution scenarios. Additionally, they extended the U-Net network by incorporating a shift-connection layer to facilitate pixel information transfer between known and missing regions. However, despite these advancements, limitations persist in restoring facial images with blurred edges due to structural shortcomings in the algorithm.

While GANs are widely employed in image inpainting tasks, challenges persist in addressing complex texture structures, fuzzy appearances, and semantic inconsistencies, particularly in facial image contexts [26]. To overcome these challenges, this research introduces a method for facial image inpainting using a multistage GAN with a global attention mechanism (GAM) referred to as CLGN. Standardizing feature layer outputs and leveraging the guiding function of the attention mechanism, enhances training speed and stability. Incorporating U-Net skip connections reduces information loss during down-sampling, thereby improving texture coherence. Additionally, a comprehensive loss function, comprising weighted reconstruction, perceptual, style, and total variation losses, optimizes the network training for superior image quality [27].

Despite these advancements and insights from traditional and deep learning-based methods, existing works have not fully addressed the challenges inherent in recovering missing facial data within security settings. While traditional patch-based and diffusion-based methods offer valuable insights, deep learning approaches, although promising, still face limitations in capturing global structures and achieving ideal repair effects, particularly in scenarios with large defect areas. Therefore, there is a clear need for further research and development, especially in GANs- based approaches like the Unet-based Encoder-Decoder with a penalty for L1 loss, to specifically address these challenges and improve missing facial data recovery within security settings.

III. Methodology

A. Synthetic Dataset

To facilitate our research, we assembled a unique dataset by overlaying face masks onto images sourced from the CelebA-HQ dataset [6], which offers 30,000 high- resolution images at 1024×1024 pixels. Due to the lack of publicly available paired images showing masked and unmasked faces for unsupervised training, we utilized the MaskTheFace[7] computer vision script to generate our masked face dataset. This dataset comprises masks of varying types, including surgical masks, N95 masks, KN95 masks, cloth masks, and gas masks. Figure 2 illustrates the diversity of these mask types, which played a crucial role in assessing the effectiveness of advanced image inpainting techniques for facial reconstruction. The flexibility of the tool in accommodating factors like facial tilt, angle, and positioning was enhanced by offering a range of mask options, making the dataset highly relevant for our research.

Fig. 2: Various Masks in Our Training Dataset.

B. Network Architecture

Our image inpainting framework utilizes a generator architecture similar to U-Net, as depicted in Figure 3, comprising approximately 46.9 million trainable parameters and 10,880 non-trainable parameters. This network incorporates both downsampling and upsampling paths connected by skip connections to gradually extract high-level features and reconstruct missing image regions. The downsampling path includes eight downsample blocks with strides of 4 and increasing filter counts (ranging from 64 to 512), effectively capturing semantic information from the input image. Notably, skip connections are established at each downsampling stage to preserve these intermediate outputs.

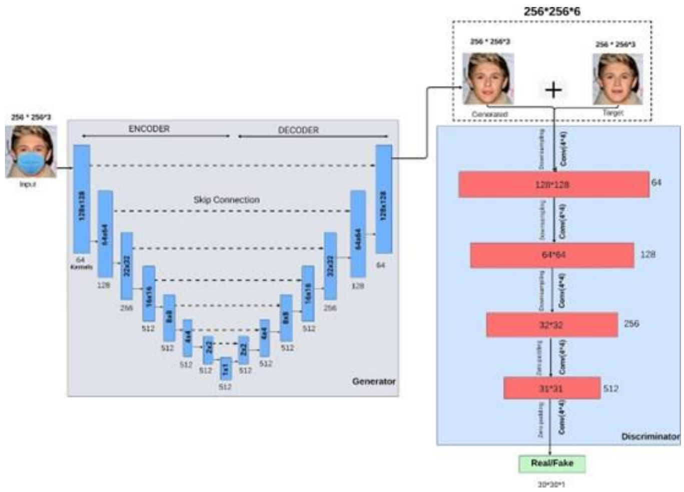

Fig. 3: An illustration showcasing a U-Net-based GAN's architecture with skip connections for optimized information flow and model performance.

During the upsampling phase, seven upsample blocks with transposed convolutions and strides of 2 progressively enlarge the feature maps and spatial resolution. Notably, these upsampled features are then combined with their corresponding skip connections, allowing the generator to integrate both high-level semantic information and detailed information from earlier stages. This fusion process empowers the network to produce realistic and contextually aware inpaintings within the masked regions. Skip connections play a crucial role in this process, accelerating training, addressing the vanishing gradient issue [7], and facilitating direct information transmission from the downsampling to the upsampling path. This ensures the preservation of precise details in non-masked areas while facilitating the generation of authentic content within the mask.

The discriminator network utilizes a CNN architecture with approximately 2.7 million trainable parameters to discern real image pairs from fake ones. It takes both the masked image and its corresponding ground truth image as inputs, concatenating them to assess the relationship between the masked region and its surrounding context. Three consecutive downsample blocks with LeakyReLU activations progressively extract features, while zero-padding and batch normalization layers pre- serve spatial information and enhance training stability. A final convolutional layer with 512 filters learns to distinguish between real and fake pairs, followed by a layer with one filter that outputs the probability of the input pair's authenticity. This feedback loop guides the generator during GANs training.

C. Loss Function

In our study, the generator loss function combines Binary Cross-Entropy (BCE) and Mean Absolute Error (MAE or L_1) losses, essential for training the generator effectively. The BCE loss assesses the adversarial or GAN loss, ensuring that the generator can convincingly deceive the discriminator by generating images that are classified as real. Additionally, the MAE loss calculates the pixel-wise difference between the generated output (G) and the target image (T), preserving structural details and features in the reconstructed images. The total generator loss is formulated as follows:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} \left(T_{R_i} \cdot log\, log\left(G_{G_i}\right) + \left(1 - T_{R_i}\right) \cdot log\, log\left(1 - G_{G_i}\right)\right) \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_{R_i} - y_{G_i}\right| \quad (2)$$

$$Total\ Generator\ Loss = BCE + (\lambda \cdot MAE) \quad (3)$$

Here, λ determines the trade-off between the adversarial loss and the reconstruction fidelity, allowing for fine-tuning the importance of each component in the training process.

$$L_{disc} = -\frac{1}{N}\sum_{i=1}^{N} \left(log\, log\left(D_{R_i}\right) + log\, log\left(1 - D_{G_i}\right)\right) \quad (4)$$

The discriminator incorporates Ldiscivotal loss components: the real loss (DR) and the generated loss (DG), crucial for effective discriminator training. The real loss (DR) is determined using the BCE loss, assessing the model's accuracy in classifying real images. Conversely, the generated loss (DG) is also computed using BCE loss, but for generated images, evaluating the discriminator's ability to distinguish between real and generated samples. The overall discriminator loss (Ldisc) is the sum of these losses, guiding the discriminator's training to improve its discrimination accuracy, a critical aspect in GANs' adversarial training dynamics.

IV. Results And Discussions

The GANs-based approach proposed for unmasking masked faces was executed on a Kaggle GPU-P100 platform. Our experimental setup involved partitioning a synthetic dataset comprising 30,000 images, each depicting five types of face masks, for training and testing the SecureGAN model. To train the generator module in producing realistic images of unmasked faces, we utilized 80% of the dataset, amounting to 24,000 images with and without masks, converted to 256×256 pixels to strike a balance between image detail and computation. The discriminator module, responsible for distinguishing between real and generated images, was trained using this set alongside 24,000 real images depicting unmasked faces.

To assess the effectiveness of our method, we set aside 20% of the dataset, specifically 6,000 images of masked faces, for testing purposes, excluding them from the training phase. The GAN model underwent training for 70,000 iterations using a batch size of 16 and the Adam optimizer with a learning rate set at 2e-4 and a beta value of 0.5. This configuration allowed us to evaluate the performance and accuracy of our approach in revealing faces and distinguishing between authentic and generated facial images.

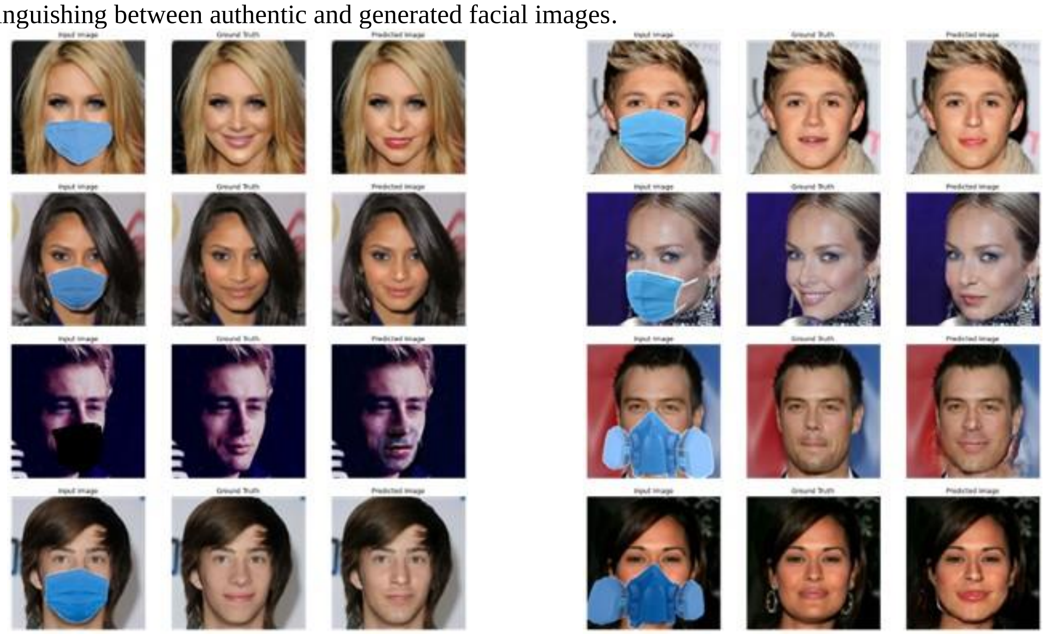

Fig. 4: Results of the Proposed Model

A. Evaluation Metrics

Our evaluation of image inpainting methods centered on assessing the quality of reconstructed face images and their alignment with ground truth. We conducted comparisons between reconstructed faces from the MaskedFace-CelebA-HQ dataset, Ground Truth images, masked images, and original images. For quantitative assessment, we utilized non-linear full reference image quality metrics such as PSNR and SSIM. PSNR gauges human perception of reconstruction quality, where higher values signify superior inpainting quality. SSIM evaluates image similarity by analyzing structural and perceptual features, with higher values indicating greater likeness between images. This comprehensive methodology enriches our comprehension of perceptual fidelity in image reconstruction.

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (5)$$

Here, X and Y denote the Generated and Target images, respectively. μ_X and μ_Y represent their mean intensities, σ_X^2 and σ_Y^2 stand for their variances, σ_{XY} signifies the covariance between them, and C_1 and C_2 are predefined constants utilized in the computation of the Structural Similarity Index (SSIM).

$$PSNR(X,Y) = 10 \cdot \left(\frac{MAX^2}{MSE}\right) \quad (6)$$

where MAX denotes the maximum pixel value, and MSE represents the Mean Square Error between images *X* and *Y*.

Table 1: Analyzing Method Performance: SSIM & PSNR Comparison

Papers	**MaskedFace (CelebA & CelebA-HQ)**	
	SSIM	***PSNR***
[17]	-	19.4dB
[10]	0.789	18.02dB
[12]	0.833	25.91dB
[9]	0.864	26.19dB
[Proposed]	0.874	22.25dB

Our model's performance is illustrated in Figure 4, showcasing its effectiveness across a diverse range of real-world test images. These samples encompass various backgrounds, from simple to complex, and diverse mask attributes including size, shape, color, and texture. Each test image includes a mask covering significant facial features, challenging our model to accurately remove the masked region and produce realistic outputs while ensuring structural coherence.

Quantitatively, our proposed net achieves an average PSNR of 22.25 and an average SSIM of 0.874 across 70,000 iterations (refer to Table 1 for detailed results), utilizing the available credits in our Kaggle free account. A comparative analysis against other training models reveals that our modified model significantly enhances image quality. For instance, Iizuka et al. [23] successfully generate plausible new content in scene images but struggle with producing realistic results for facial images with substantial missing regions, characterized by significant structural and appearance variations.

V. Conclusion

Our research represents a significant advancement in image inpainting, specifically designed to tackle the challenges brought about by the widespread adoption of face masks during the COVID-19 pandemic. Through the integration of fully convolutional and generative adversarial network principles, our model prioritizes both comprehensive semantic comprehension and the harmonization of global and local attributes in image reconstruction. A pivotal contribution lies in our innovative methodology, seamlessly removing mask elements from facial images while preserving intricate facial structural details. We emphasize structural and visual consistency through penalty mechanisms during training, enhancing the efficacy of skip connections within the U-Net architecture embedded in GANs. Noteworthy is our use of Mean Absolute Error to prevent excessive penalties in the loss function, ensuring comprehensive learning of facial attributes within manageable constraints. This strategy yields perceptually high-quality results that outperform existing methods, laying the groundwork for robust and adaptable image inpainting solutions relevant across various domains and dynamic challenges.

While our techniques demonstrate satisfactory inpainting outcomes using incomplete images, there exist inherent differences in color and texture nuances when compared to real values. The assistance of structural information helps in enhancing overall structure to a certain degree; however, accurately reproducing precise details for high-level semantic restorations such as human eyes and mouths remains a formidable task. Utilizing extensive datasets improves network adaptation during training but can result in prolonged training durations. Hence, future endeavors should concentrate on enhancing facial image inpainting with a focus on higher-level semantics to elevate result authenticity. Designing networks for large datasets should prioritize efficiency while minimizing training time. Moreover, we aspire to achieve a balanced dataset representation, including equal proportions of male and female subjects, to alleviate potential data biases, particularly evident in the predominance of female celebrities in the CelebA-HQ dataset.

VI. Acknowledgment

We express our sincere appreciation to Dr. Anjali Yeole for her invaluable mentor- ship and unwavering support during this project. Her expertise and guidance significantly influenced the direction of our research. We are also grateful to the faculty members in the Artificial Intelligence and Data Science department at VESIT, our peers for their valuable insights, and the institution for providing resources essential to our work.

References

[1] D. J. Carragher and P. J. B. Hancock, "Surgical Face Masks Impair Human Face Matching Performance for Familiar and Unfamiliar Faces," *Cognitive Research: Principles and Implications*, vol. 5, no. 1, Nov. 2020, doi: https://doi.org/10.1186/s41235-020-00258-x.

[2] AP News, "Coronavirus masks a boon for crooks who hide their faces," May 16, 2020. [Online]. Available: https://apnews.com/article/ f97b4914b4159dec0c98359fac123d52. [Accessed: Apr. 04, 2024].

[3] M. F. D. Martello and L. T. Maloney, "Where are kin recognition signals in the human face?", *Journal of Vision*, vol. 6, no. 12, pp. 2–2, Nov. 2006, doi: https://doi.org/10.1167/6.12.2.

[4] M. Ramgopal, M. Sai Roopesh, M. Veeranna Chowdary, M. Madhav, and K. Shanmuga, "Masked Facial Recognition in Security Systems Using Transfer Learning," *SN Computer Science*, vol. 4, no. 1, Oct. 2022, doi: https://doi.org/10.1007/s42979-022-01400-w.

[5] G. Modak, S. S. Das, M. A. Islam Miraj and M. K. Morol, "A Deep Learning Framework to Reconstruct Face under Mask," 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2022, pp. 200-205, doi: https://doi.org/10.1109/CDMA54072.2022.00038.

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," *arXiv.org*, 2017. https://arxiv.org/abs/1710.10196.

[7] A. Anwar and A. Raychowdhury, "Masked Face Recognition for Secure Authentication," *arXiv:2008.11104 [cs, eess]*, Aug. 2020. Available: https://arxiv.org/abs/2008.11104.

[8] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: https://doi.org/10.1109/CVPR.2016.90.

[9] N. Ud Din, K. Javed, S. Bae and J. Yi,”A Novel GAN-Based Network for Unmasking of Masked Face,” in *IEEE Access*, vol. 8, pp. 44276-44287, 2020, doi: https://doi.org/10.1109/ACCESS.2020.2977386.

[10] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patch- Match,” *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 1–11, Jul. 2009, doi: https://doi.org/10.1145/1531326.1531330.

[11] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, ”Image quality assessment: from error visibility to structural similarity,” in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004, doi: https://doi.org/10.1109/TIP.2003.819861.

[12] M. K. J. Khan, N. Ud Din, S. Bae, and J. Yi, “Interactive Removal of Microphone Object in Facial Images,” *Electronics*, vol. 8, no. 10, p. 1115, Oct. 2019, doi: https://doi.org/10.3390/electronics8101115.

[13] A. Efros and T. Leung, ”Texture Synthesis by Non-parametric Sampling,” in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, 2000.

[14] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, “Simultaneous structure and texture image inpainting,” *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 882–889, Aug. 2003, doi: https://doi.org/10.1109/ TIP.2003.815261.

[15] W. Xiong et al., ”Foreground-Aware Image Inpainting,” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 5833-5841, doi: https://doi.org/10.1109/ CVPR.2019.00599.

[16] C. Agarwal and C. Bhatnagar, “Unmasking the potential: evaluating image inpainting techniques for masked face reconstruction,” *Multimedia Tools and Applications*, May 2023, doi: https://doi.org/10.1007/ s11042-023-15807-x.

[17] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson and M. N. Do, ”Semantic Image Inpainting with Deep Generative Models,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6882-6890, doi: https://doi.org/10.1109/CVPR.2017.728.

[18] D. Ding, S. Ram, and J. J. Rodr´ıguez, ”Image Inpainting Using Nonlocal Texture Matching and Nonlinear Filtering,” in *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1705-1719, April 2019, doi: https://doi.org/10.1109/TIP.2018.2880681.

[19] H.-Y. Lu, Q. Liu, M. Zhang, N. C. Beaulieu, and X. Deng, “Gradient-based low rank method and its application in image inpainting,” *vol. 77, no. 5*, pp. 5969–5993, Mar. 2018, doi: https://doi.org/10.1007/ s11042-017-4509-0.

[20] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, “The Generalized PatchMatch Correspondence Algorithm,” *Lecture Notes in Computer Science*, pp. 29–43, Sep. 2010, doi: https://doi.org/10.1007/ 978-3-642-15558-1_3.

[21] A. Criminisi, P. Perez, and K. Toyama, ”Region filling and object removal by exemplar-based image inpainting,” in *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200-1212, Sept. 2004, doi: https://doi.org/10.1109/TIP.2004.833105.

[22] Z. Yanhong, F. Jianlong, C. Hongyang, and G. Baining, “Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting,” Jan. 2019, doi: https://doi.org/10.48550/arxiv.1904.07475.

[23] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and locally consistent image completion,” *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, Jul. 2017, doi: https://doi.org/10.1145/3072959.3073659.

[24] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, “Image Inpainting via Generative Multi-column Convolutional Neural Networks,” *arXiv.org*, Oct. 20, 2018. Available: https://arxiv.org/abs/1810.08771.

[25] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, ”Context Encoders: Feature Learning by Inpainting,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2536-2544, doi: https://doi.org/10.1109/CVPR.2016. 278.

[26] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, ”High- Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 4076-4084, doi: https://doi.org/ 10.1109/CVPR.2017.434.

[27] X. Gao and Y. Fang, “A note on the generalized degrees of freedom under the L1 loss function,” *Journal of Statistical Planning and Inference*, vol. 141, no. 2, pp. 677–686, Feb. 2011, doi: https://doi.org/10.1016/j. jspi.2010.07.006.