Adversarial Attacks and Defense Mechanisms in Facial Recognition Systems: A Comprehensive Analysis

1^{st} Ricky(GuangLiang) Yang
Auckland University of Technology

***Abstract*—Facial recognition technology is widely used in security, authentication, and various social services. Despite its widespread adoption, this technology is vulnerable to adversarial attacks involving subtle input image processing to trick the recognition model, leading to serious security vulnerabilities and misrecognition. Addressing these vulnerabilities is critical as they threaten individual privacy and organizational security. This paper comprehensively reviews adversarial attack methods, including gradient-based attacks, optimization-based attacks, and generative model-based attacks. This paper also investigates defense strategies such as adversarial training, defense distillation, gradient masking, input preprocessing, ensemble methods, and adaptive defense mechanisms. Ensuring the robustness and reliability of face recognition systems is essential to mitigate these evolving threats.**

***Index Terms*—Facial Recognition, Adversarial Attacks, Security, Identity Authentication, Defensive Strategies, Hybrid Defensive Models, Resilience, Adaptive Strategies**

I. Introduction

Facial recognition technology has become an integral component of modern society, with applications in security monitoring, identity authentication, and various social services. Its capability to swiftly and accurately identify individuals has rendered it invaluable for enhancing security and convenience across numerous sectors [1]. However, the increasing reliance on these systems has revealed significant vulnerabilities, mainly through adversarial attacks. These attacks involve the subtle manipulation of input images to deceive recognition models into making incorrect predictions [2]. Such manipulations can result in severe consequences, including security breaches, unauthorized access, and privacy violations.

Understanding and mitigating adversarial attacks on facial recognition systems is crucial. As these systems increase, their potential exploitation by malicious actors grows, posing severe threats to individual privacy and organizational security. Adversarial attacks can disrupt recognition processes, leading to misidentifications that could cause wrongful accusations, identity theft, and unauthorized access to secure areas. These risks underscore the critical need for robust defense mechanisms to ensure the integrity and reliability of facial recognition technology.

This paper aims to provide a comprehensive overview of the current landscape of adversarial attacks and defenses in facial recognition systems. Through an extensive literature review, we analyze various adversarial attack methods, including gradient-based attacks like the Fast Gradient Sign Method (FGSM) [3] and Basic Iterative Method (BIM) [4], optimization-based attacks such as Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) and Carlini & Wagner (C&W), and generative model-based attacks using Generative Adversarial Networks (GANs) [5]. To understand their effectiveness, we evaluate each method's advantages, disadvantages, and impact on recognition models.

The paper examines a range of defense strategies in response to these threats. These include adversarial training, defensive distillation, gradient masking, input preprocessing, ensemble methods, external network supplementation, and adaptive defense mechanisms. By assessing these strategies' practical applications and effectiveness, we provide insights into how facial recognition systems can be better protected against adversarial attacks [6], [7], [8], [9].

In conclusion, ensuring the security and reliability of facial recognition systems amidst evolving adversarial threats is paramount. This research highlights the necessity for continuous innovation in defensive strategies to safeguard personal privacy and organizational security, emphasizing the need for robust and resilient mechanisms to counteract sophisticated adversarial attacks.

II. Literature review

A. Methodology

This literature review identifies recent research on adversarial attacks and defenses in facial recognition systems, focusing on current attack methods and defensive strategies.

1) **Database Search:** We searched databases, including AUT Library, Google Scholar, IEEE Xplore, and ACM Digital Library, using keywords such as "Facial Recognition," "Adversarial Attacks," and "Defensive Strategies."
2) **Selection Criteria:** From over 50 papers, around 20 were selected based on:
    - **Publication Date:** From 2020 onward.
    - **Relevance:** Targeting adversarial attacks or defensive strategies.
    - **Quality and Innovation:** Novel techniques or comprehensive analyses.
    - **Methodological Rigor:** Thorough evaluation and testing.

3) **Review Process:** Each paper was subjected to a rigorous analysis, ensuring clarity, methodological soundness, and practical applications. The findings were then categorized to identify common themes, research gaps, and potential study areas, providing a comprehensive and reliable overview of the current state of research on adversarial attacks and defenses in facial recognition systems.

B. Background

Facial recognition systems are susceptible to adversarial attacks involving imperceptible perturbations that mislead machine learning models. First identified by Szegedy et al. in 2014 [10], these attacks have evolved to include the Fast Gradient Sign Method (FGSM) [3], Basic Iterative Method (BIM) [4], and Generative Adversarial Networks (GANs) [5]. These attacks pose significant financial, surveillance, and autonomous vehicle risks, enabling identity impersonation, security breaches, and system disruptions [11], [12], [2]. Defensive measures like adversarial training, input transformation, and detection mechanisms offer defense but face challenges like increased computational complexity and reduced accuracy on clean data.

C. Definition

Adversarial attacks use small perturbations to deceive machine learning models, particularly in facial recognition, by causing misjudgment of correct inputs. Defense mechanisms aim to maintain the model's accuracy and robustness against these attacks. [12].

D. Attack Research

Adversarial attacks on facial recognition (FR) systems exploit vulnerabilities to mislead models into incorrect predictions. Vakhshiteh, Nickabadi, and Ramachandra provide a detailed analysis, categorizing attacks into physical and digital types. Physical attacks involve artifacts like photos and masks, while digital attacks use adversarial examples to deceive FR systems [13]. Essential methods include L-BFGS, FGSM, and DeepFool, each evaluated for effectiveness in evading detection. Defensive strategies such as adversarial training and robust model development are discussed, emphasizing the need for ongoing research to enhance FR systems' resilience against evolving adversarial techniques.

Akhtar et al. classify adversarial attacks into gradient-based, black-box, and unrestricted types. They highlight advancements in attack methodologies, including transfer-based attacks that improve perturbation transferability across models [14]. The study underscores the importance of robust training techniques and hybrid models to counter adversarial threats effectively. The review addresses model inversion and backdoor attacks, stressing the need for adaptive strategies to circumvent specific defenses.

Nishchal Jagadeesha's research focuses on privacy preservation in FR systems using adversarial techniques. The study demonstrates how imperceptible adversarial perturbations can fool FR systems, maintaining image integrity while protecting privacy [15]. FGSM is used for white-box attacks and generic perturbations for black-box attacks, achieving high misclassification rates without prior model knowledge. This dual focus on privacy and robustness highlights the need for improved models to handle feature-based FR systems.

Lu Yang, Qing Song, and Yingqi Wu introduce the Attentional Adversarial Attack Generative Network (A3GN), which incorporates a conditional variational autoencoder and attention modules to enhance the realism and effectiveness of adversarial examples [16], [17], [18], [19]. A3GN employs a three-player GAN setup, including a face recognition network for precision. This approach overcomes the limitations of traditional methods by achieving higher visual quality and adaptability in both white-box and black-box scenarios. Using a substitute network for feature estimation in black-box attacks marks a significant advancement, improving the generalizability and impact of adversarial attacks across different FR systems.

Divya Saxena and Jiannong Cao provide a systematic overview of GAN advancements, addressing challenges like mode collapse, non-convergence, and instability [20]. They propose solutions such as re-engineered architectures, new loss functions, and alternative optimization algorithms. Their taxonomy categorizes solutions into conditional generation, multi-generator setups, and memory networks. Improving GANs' generalization and stability is crucial for developing effective adversarial attacks and defenses.

E. Main Attack Methods Classification

Adversarial attacks on machine learning in computer vision are extensively studied [21], classified by implementation strategies, physical realizations, attack characteristics, adaptability, and knowledge requirements(Table I).

F. Defensive Research

In a comprehensive study by Vakhshiteh, Nickabadi, and Ramachandra, defense strategies against adversarial attacks on facial recognition (FR) systems are classified into three primary categories: altering the training during learning, changing the network, and supplementing the primary model with external networks [13]. The first category involves injecting adversarial examples into the training data to enhance robustness. The second focuses on modifying the network architecture, such as adjusting layers or activation functions. The third involves adding external models that assist the primary network in classifying unseen samples. This study highlights the effectiveness of these mechanisms, noting that incorporating adversarial examples improves robustness, network modifications mitigate adversarial noise, and external networks, like blockchain-based security, resist tampering.

Samer Y. Khamaiseh et al. reviews defense strategies against adversarial attacks on deep neural networks (DNNs), including adversarial training, defensive distillation, gradient masking, input preprocessing, and ensemble methods [22]. Adversarial training improves robustness but is resource-intensive. Defensive distillation smooths decision boundaries but can be

TABLE I
CLASSIFICATION OF ADVERSARIAL ATTACKS IN FACIAL RECOGNITION SYSTEMS

Classification Type	**Category**	**Examples**
By Implementation Method	Gradient-based Attacks	FGSM, BIM, DeepFool, C&W, JSMA
	Optimization-based Attacks	L-BFGS, C&W, DeepFool
	Generative Model-based Attacks	GANs, AdvFaces
By Physical Realization	Digital Attacks	Gradient-based Attacks (e.g., FGSM, BIM) Optimization-based Attacks (e.g., L-BFGS, C&W)
	Physical Attacks	Eyeglass Accessory Printing Visible Light-based Attack Infrared Dot Projector Attack
By Attack Characteristics	Targeted Attacks	FGSM (in targeted mode) BIM (in targeted mode) C&W
	Non-targeted Attacks	FGSM (in non-targeted mode) BIM (in non-targeted mode) DeepFool
By Adversarial Sample Adaptability	Image-specific Perturbations	FGSM, BIM, C&W
	Universal Perturbations	Universal Adversarial Perturbations
By Knowledge Requirement	White-box Attacks	FGSM, BIM, C&W
	Black-box Attacks	Evolutionary Attack, One Pixel Attack

bypassed by advanced attacks. Gradient masking obscures gradient information, yet it needs to be foolproof. Input preprocessing, such as image transformations, mitigates adversarial effects, and ensemble methods aggregate multiple model predictions, enhancing overall robustness but requiring more computational power. Both studies emphasize that combining various strategies is essential for comprehensive protection and call for future research to develop adaptive, resilient defenses that integrate seamlessly into real-world applications, ensuring long-term security and reliability.

G. Defense Strategies Classification

Defending facial recognition systems from adversarial attacks involves various techniques, each with strengths and weaknesses.

H. Risks

Adversarial attacks on facial recognition (FR) systems pose significant risks to individuals and organizations by exploiting vulnerabilities in deep neural networks. These attacks can manipulate FR systems, leading to unauthorized access, identity theft, and privacy breaches. Critical studies highlight these risks and their adverse consequences.

Samer Y. Khamaiseh et al. emphasize that adversarial attacks compromise the security and privacy of individuals and organizations [22]. For individuals, these attacks result in unauthorized access to personal information, wrongful accusations, and privacy violations. Misidentification by FR systems due to adversarial examples can deny services or grant unauthorized access to secure areas.

Organizations face substantial security risks from adversarial attacks, which can breach sensitive information and critical infrastructure. These attacks enable unauthorized access to restricted areas, allowing adversaries to evade detection or impersonate authorized personnel. The study notes that evolving techniques, such as FGSM, DeepFool, and GAN-based methods, challenge organizations to maintain robust defenses.

Akhtar et al. detail how small perturbations degrade FR system performance, causing security and operational disruptions [14]. Organizations relying on FR for security and access control are vulnerable to breaches, data theft, and operational downtimes. The economic impact includes mitigation costs and potential loss of customer trust and reputation.

Nishchal Jagadeesha highlights the dual-use nature of adversarial attacks [15]. While these techniques can protect individual privacy by confusing FR systems, they pose significant risks if misused. Malicious actors can exploit adversarial attacks to bypass security measures, leading to unauthorized surveillance and privacy violations.

In conclusion, adversarial attacks on FR systems present substantial risks to individuals and organizations. These risks necessitate developing robust, adaptive defense mechanisms to protect personal privacy and organizational security.

III. ANALYSIS AND DISCUSSION

A. Research Findings Significance

Researching adversarial attacks and defenses in facial recognition (FR) systems is crucial in the digital age, where FR technology is extensively used in security, authentication, and surveillance. These findings uncover FR systems' vulnerabilities and emphasize the need for robust defense mechanisms to protect personal privacy and organizational security.

Vakhshiteh et al. categorize adversarial attacks into physical and digital types, illustrating diverse methods of exploiting FR systems [13]. Akhtar et al. elaborate on advanced techniques like gradient-based and transfer-based attacks [2], highlighting the evolving nature of adversarial threats. These studies underscore the critical need for ongoing research and innovation in defensive strategies to counter sophisticated adversarial techniques.

B. Challenges

The primary challenge identified in the research is the evolving sophistication of adversarial attacks. Traditional defense

TABLE II
CLASSIFICATION OF DEFENSE STRATEGIES AGAINST ADVERSARIAL ATTACKS ON FACIAL RECOGNITION SYSTEMS

Defense Strategy	Description	Strengths	Weaknesses
Adversarial Training	Using adversarial examples in training to enhance robustness.	Improves resistance to adversarial perturbations.	Computationally intensive, may reduce clean data accuracy.
Defensive Distillation	Training a distilled model at higher temperatures to smooth decision boundaries.	Increases resilience to small perturbations.	Can be bypassed by advanced attacks.
Gradient Masking	Obscuring gradient information to hinder adversarial example generation.	Simple to implement, hinders some attacks.	Superficial defense, can be circumvented by advanced attacks.
Input Preprocessing	Preprocessing data to remove adversarial perturbations.	Effective against many attacks.	May degrade data quality, reduce accuracy.
Ensemble Methods	Using multiple models to make decisions.	Reduces likelihood of simultaneous deception.	Requires more resources, complex to manage.
External Network Supplementation	Adding external models or systems.	Adds security layer, helps mitigate tampering.	Introduces complexity, needs integration.
Adaptive Defense Mechanisms	Dynamically adjusting defenses based on detected attacks.	Flexible and robust against various methods.	Complex, resource-intensive.

mechanisms, such as adversarial training and input preprocessing, often become ineffective against advanced attacks. This dynamic nature necessitates constant updates to defense strategies, complicating organizational efforts to maintain robust security measures.

Another significant challenge is balancing robustness with computational efficiency. Defensive strategies like adversarial training enhance robustness but increase computational complexity, which is problematic for real-time applications like surveillance and autonomous vehicles, where high-speed processing is crucial.

Additionally, the issue of inaccuracy in FR systems, highlighted by Moraes et al., poses further challenges. Misidentifications can lead to severe consequences, such as wrongful accusations and security breaches, complicating effective defense implementation [23].

C. Strategies to Address Challenges

Several strategies have been proposed to address these challenges. One practical approach is integrating hybrid defensive models, combining adversarial training, defensive distillation, and input preprocessing to create a comprehensive defense. This approach leverages the strengths and mitigates the weaknesses of each technique.

Another promising strategy is the development of adaptive defense mechanisms, which dynamically adjust based on the detected adversarial attack's type and sophistication. These systems, incorporating machine learning algorithms, offer a flexible and robust solution by identifying and responding to new threats in real-time.

The use of ensemble methods, discussed by Khamaiseh et al., is also viable. By aggregating predictions from multiple models, ensemble methods reduce the likelihood of simultaneous deception, enhancing the FR system's overall robustness [22].

D. Manage Risks

For Individuals

Individuals can take several steps to manage the risks posed by adversarial attacks on facial recognition (FR) systems:

- **Awareness and Education:** Individuals should be informed about the risks of using FR systems. Understanding how adversarial attacks work and the possible consequences can help individuals take proactive steps to protect their privacy.
- **Privacy Tools:** Utilize tools and techniques to protect personal data from being misused by FR systems. This includes using face masks, sunglasses, or makeup that can confuse FR systems, known as adversarial fashion.
- **Control Over Data:** Be mindful of where and how personal biometric data is shared. Opt out of facial recognition databases whenever possible and leverage privacy settings offered by platforms using FR technology.
- **Legal Recourse:** Know the legal rights and protections available. In some jurisdictions, individuals can take legal action against misuse or unauthorized collection of their biometric data.

For Organizations

Organizations need to adopt a comprehensive approach to manage the risks associated with adversarial attacks on FR systems:

- **Robust Defensive Strategies:**
  - **Adversarial Training:** Incorporate adversarial examples into the training data to enhance the robustness of FR systems. This involves training the model to recognize and correctly classify images even when adversarial perturbations exist.
  - **Defensive Distillation:** Implement defensive distillation techniques to make the model less sensitive to adversarial perturbations. This involves training the model at higher temperatures to smooth the decision boundaries.
  - **Input Preprocessing:** Apply input preprocessing techniques such as image transformations, feature squeezing, and noise reduction to filter out adversarial perturbations before they reach the FR system.
  - **Ensemble Methods:** Use multiple models to make decisions rather than relying on a single FR system. This can reduce the likelihood of all models being

fooled by the same adversarial example.

- **Regular Security Audits:** Conduct frequent security assessments and audits of FR systems to identify and mitigate vulnerabilities. This includes penetration testing to simulate adversarial attacks and evaluate the system's robustness.
- **Algorithmic Transparency:** Implement transparent algorithms and maintain logs to track FR systems' decisions. This can help identify suspicious activities and understand the impact of adversarial attacks.
- **Access Control and Monitoring:** Strengthen access control mechanisms and monitor FR systems' usage to detect and promptly respond to unauthorized attempts. This includes employing multi-factor authentication and continuous monitoring for anomalies.
- **Policy and Governance:** Establish clear policies and governance frameworks to regulate the use of FR systems. This includes setting data collection, storage, and processing guidelines to ensure compliance with privacy laws and ethical standards.
- **Collaboration and Standards:** Participate in industry collaborations and adopt standards for FR technologies. Engaging with industry groups and contributing to the development of best practices can enhance collective defense against adversarial attacks.
- **Legal and Regulatory Compliance:** Ensure compliance with relevant laws and regulations governing the use of biometric data. This includes adhering to data protection regulations such as GDPR and CCPA and being prepared for audits and inspections by regulatory bodies.

E. Impacts

These research findings have significant implications for both individuals and organizations. For individuals, enhanced defensive strategies protect personal privacy and reduce the risk of identity theft and unauthorized surveillance. Organizations benefit from improved security measures that safeguard sensitive information and critical infrastructure.

The economic impact is also considerable. Implementing robust defensive strategies can lower the costs associated with security breaches, including data loss, reputational damage, and legal penalties. Additionally, companies prioritizing security can build greater customer trust, increasing loyalty and business growth.

F. Comparisons with Other Fields

Several parallels emerge when comparing challenges and strategies in facial recognition (FR) systems to fields such as cybersecurity and autonomous systems. Cybersecurity's constantly evolving threats necessitate adaptive, multi-layered defense strategies similar to those required for FR systems. Ensemble methods and hybrid models, which are standard in cybersecurity, enhance robustness in FR systems.

Real-time processing and accuracy are crucial in autonomous systems akin to FR systems. Both fields face the trade-off between robustness and computational efficiency, underscoring the need for innovative solutions that effectively balance these requirements.

IV. Conclusion

Studying adversarial attacks and defenses in facial recognition (FR) systems is crucial in cybersecurity. This paper highlights critical points regarding these attacks' nature, challenges, and current mitigation strategies.

The literature review reveals that adversarial attacks significantly threaten the reliability and security of FR systems. Various methods, such as gradient-based, optimization-based, and generative model-based attacks, exploit these systems' vulnerabilities. Both physical and digital attacks, including adversarial examples and presentation attacks, demonstrate the diverse approaches used to deceive FR systems. Studies by Vakhshiteh et al. [13] and Akhtar et al. [2] provide comprehensive analyses, emphasizing these threats' sophistication and evolving nature.

Several defensive strategies have been proposed to counter these attacks, including adversarial training, defensive distillation, input preprocessing, and ensemble methods. Each strategy has its strengths and limitations. For instance, adversarial training enhances robustness but is computationally intensive and may reduce accuracy on clean data. Defensive distillation smooths decision boundaries but can be bypassed by advanced attacks. Input preprocessing can filter out adversarial perturbations but might degrade data quality. Ensemble methods reduce the likelihood of simultaneous deception but require more resources and are complex to manage. The review underscores the need for a hybrid approach, combining multiple strategies to secure FR systems against diverse threats effectively.

Despite advancements, significant limitations persist in current research. One major issue is the trade-off between robustness and computational efficiency. Many effective defensive strategies are resource-intensive, making them impractical for real-time applications requiring high-speed processing. Additionally, the dynamic nature of adversarial attacks necessitates continuous updates to defense mechanisms, posing challenges to maintaining robust security measures.

Further research is crucial to address these limitations and enhance FR systems' security. One promising area is the development of adaptive defense mechanisms that dynamically adjust based on the type and sophistication of detected attacks. This involves incorporating machine learning algorithms to identify and respond to new threats in real-time, offering a flexible and robust solution. Another area of interest is integrating privacy-preserving techniques with adversarial defenses. As highlighted by Jagadeesha, combining methods to protect user privacy while ensuring robust security can address technical and ethical concerns [15]. Additionally, improving the generalization capabilities and stability of generative adversarial networks (GANs) can enhance the effectiveness of attacks and defenses, as discussed by Saxena and Cao [20].

In conclusion, while significant progress has been made in understanding and mitigating adversarial attacks on FR systems, ongoing research is essential to keep pace with the evolving threat landscape. By addressing current limitations and exploring new defense strategies, researchers can contribute to developing more secure and resilient FR systems, ultimately enhancing personal privacy and organizational security.

References

[1] R. Hwang, J. Lin, and H. Lin, "Adversarial patch attacks on deep-learning-based face recognition systems using generative adversarial networks," *Sensors*, vol. 23, no. 2, p. 853, 2023.

[2] N. Akhtar and A. Mian, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[3] A. Musa, K. Vishi, and B. Rexha, "Attack analysis of face recognition authentication systems using fast gradient sign method," *Applied artificial intelligence*, vol. 35, no. 15, pp. 1346–1360, 2021.

[4] A. Velykorusova, E. K. Zavadskas, L. Tupenaite, L. Kanapeckiene, D. Migilinskas, V. Kutut, I. Ubarte, Z. Abaravicius, and A. Kaklauskas, "Intelligent multi-criteria decision support for renovation solutions for a building based on emotion recognition by applying the copras method and bim integration," *Applied Sciences*, vol. 13, no. 9, p. 5453, 2023.

[5] L. Yang, Q. Song, and Y. Wu, "Attacks on state-of-the-art face recognition using attentional adversarial attack generative network," *Multimedia tools and applications*, vol. 80, pp. 855–875, 2021.

[6] A. Zolfi, S. Avidan, Y. Elovici, and A. Shabtai, "Adversarial mask: Real-world adversarial attack against face recognition models," *arXiv preprint arXiv:2111.10759*, vol. 2, no. 3, 2021.

[7] H. Zolfi, J.-Y. Lin, S.-Y. Hsieh, H.-Y. Lin, and C.-L. Lin, "Adversarial patch attacks on deep-learning-based face recognition systems using generative adversarial networks," *Sensors*, vol. 23, no. 2, p. 853, 2023.

[8] R. Saxena, A. S. Adate, and D. Sasikumar, "A comparative study on adversarial noise generation for single image classification," *International Journal of Intelligent Information Technologies*, vol. 16, pp. 75–87, 2020.

[9] E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: Convnet," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[10] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

[11] I. Fursov, M. Morozov, N. Kaploukhaya, E. Kovtun, R. Rivera-Castro, G. Gusev *et al.*, "Adversarial attacks on deep models for financial transaction records," *arXiv preprint arXiv:2106.08361*, 2021.

[12] A. Vassilev, A. Oprea, A. Fordyce, and H. Anderson, "Adversarial machine learning: A taxonomy and terminology of attacks and mitigations," NIST, Tech. Rep., 2024.

[13] F. Vakhshiteh, A. Nickabadi, and R. Ramachandra, "Adversarial attacks against face recognition: A comprehensive study," *IEEE Access*, vol. 9, pp. 92 735–92 756, 2021.

[14] N. Akhtar, A. Mian, N. Kardan, and M. Shah, "Advances in adversarial attacks and defenses in computer vision: A survey," *IEEE Access*, vol. 9, pp. 155 161–155 196, 2021.

[15] N. Jagadeesha, "Facial privacy preservation using fgsm and universal perturbation attacks," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1. IEEE, 2022.

[16] L. Yang, Q. Song, and Y. Wu, "Attacks on state-of-the-art face recognition using attentional adversarial attack generative network," *Multimedia tools and applications*, vol. 80, pp. 855–875, 2021.

[17] X. Wang, F. Ye, Y. Li, and Y. Dai, "Biphasic face photo-sketch synthesis via face semantic-aware cyclegan," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. IEEE, 2023, pp. 1172–1179.

[18] P. Srinivasan, A. M. K, M. Saraogi, J. Nataraju, G. Mishra, S. K. S, and S. A N, "Image inpainting for facial recognition using generative networks," in *2024 3rd International Conference for Innovation in Technology (INOCON)*, 2024, pp. 1–8.

[19] A. Sevastopolsky, Y. Malkov, N. Durasov, L. Verdoliva, and M. Niesner, "How to boost face recognition with stylegan?" in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023, pp. 20 867–20 877.

[20] D. Saxena and J. Cao, "Generative adversarial networks (gans): Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, may 2021. [Online]. Available: https://doi.org/10.1145/3446374

[21] Y. M. Khedr, Y. Xiong, and K. He, "Semantic adversarial attacks on face recognition through significant attributes," *International Journal of Computational Intelligence Systems*, vol. 16, no. 196, 2023.

[22] S. Y. Khamaiseh, D. Bagagem, A. Al-Alaj, M. Mancino, and H. W. Alomari, "Adversarial deep learning: A survey on adversarial attacks and defense mechanisms on image classification," *IEEE Access*, vol. 10, pp. 102 266–102 291, 2022.

[23] T. G. Moraes, E. C. Almeida, and J. R. L. de Pereira, "Smile, you are being identified! risks and measures for the use of facial recognition in (semi-) public spaces," *AI and Ethics*, vol. 1, no. 2, pp. 159–172, 2021.