Securing AI-Generated Media: Rethinking Deepfake Vulnerabilities in Side-Face Perspectives

Bibek Das[1], Chandranath Adak[1], Anurag Deo[1], Aditya Bangar[2], Rahul Verma[1], Zahid Akhtar[3], Soumi Chattopadhyay[4], Soumya Dutta[2], Ashutosh Parihar[4], Anand Suralkar[4], Deepak Nagar[1], Vishal Kumar[2]
[1]Dept. of CSE, Indian Institute of Technology Patna, India–801106
[2]Dept. of CSE, Indian Institute of Technology Kanpur, India–208016
[3]State University of New York Polytechnic Institute, USA–13502
[4]Dept. of CSE, Indian Institute of Technology Indore, India–453552
{bibek_2422cs27,chandranath}@iitp.ac.in

***Abstract*—Deepfake technology has advanced significantly, producing highly sophisticated fake images that challenge detection mechanisms. However, existing deepfake generators struggle to maintain realism in side-face perspectives, particularly under diverse indoor and outdoor lighting conditions. This limitation is further pronounced for individuals of Indian ethnicity, where variations in skin tone, hairstyles, facial hair, and image capture distance from the camera introduce additional challenges. In this paper, we critically examine the performance of state-of-the-art deepfake generators in these scenarios, highlighting key vulnerabilities in side-face synthesis. We also assess the effectiveness of current detection frameworks in identifying these inconsistencies. Furthermore, we discuss the broader implications of generative models in security-sensitive applications and propose future research directions to enhance the robustness of deepfake synthesis and detection. Our recommendations include improving dataset diversity, developing adaptive generative models, and leveraging multimodal approaches to strengthen detection mechanisms, ensuring more secure and reliable AI-driven media applications.**

***Index Terms*—Deepfake, Generative models, Discriminators, Side-face**

I. Introduction

Deepfake (DF) technology, powered by generative deep neural networks, has revolutionized digital media by enabling the creation of highly realistic synthetic images and videos. These advancements have unlocked innovative applications in entertainment, education, and accessibility, offering new ways to engage audiences and enhance experiences [1]. However, this technology poses significant risks, including misinformation, identity fraud, and security threats. As DF generation techniques continue to evolve, the ability to distinguish between real and synthetic content has become increasingly challenging, prompting extensive research into DF detection methods. Despite these efforts, a critical gap remains in the analysis of DF vulnerabilities in side-face perspectives, particularly under diverse real-world conditions such as varying lighting, camera angles, and ethnic diversity [2].

Most existing DF research and datasets focus predominantly on frontal-face perspectives, where facial features are fully visible, and detection models can leverage symmetrical patterns and high-resolution details. In contrast, side-face DFs present unique challenges due to partial facial visibility, occlusions, and variations in illumination [3]. These factors make side-face synthesis an underexplored yet highly relevant problem, especially in security-sensitive domains [4]. Our analysis reveals that current DF generators struggle to produce realistic side-face images, often resulting in artifacts such as boundary distortions, texture inconsistencies, and lighting mismatches. These limitations are pronounced for individuals with diverse skin tones, hairstyles, and facial structures, and are even more evident when generating DFs for Indian ethnicities, where variations in complexion, hair texture, and facial attributes introduce additional synthesis challenges [5].

In this paper, we study state-of-the-art DF generators in side-face synthesis and evaluate their vulnerabilities when dealing with realistic variations in ethnicity, pose, and environmental conditions. We also assess the efficacy of current DF detection models in identifying these inconsistencies, providing insights into their applicability in real-world settings. Beyond identifying weaknesses, we discuss the broader implications of generative models in security-sensitive applications and propose future research directions to enhance both DF synthesis and detection. Our recommendations include improving dataset diversity, developing adaptive generative models, and leveraging multimodal approaches to strengthen detection mechanisms. By addressing these critical gaps, we aim to foster more secure and reliable AI-driven media applications while advancing the understanding of DF vulnerabilities in non-frontal perspectives. Improving side-face DF generation and detection could impact fields like cybersecurity, law enforcement, and digital forensics.

Our primary **contributions** to this paper are as follows:
(i) Deepfake vulnerability analysis: We identify and categorize the key limitations of existing DF generators in synthesizing side-face DFs, particularly under diverse lighting conditions and for Indian ethnicity variations.
(ii) Detection framework assessment: We examine the effectiveness of major representatives of state-of-the-art DF detection mechanisms in identifying manipulated side-face images and analyze their robustness against generator-specific inconsistencies.

TABLE I: Brief details of major deepfake datasets

Year	Dataset	#Real	#Fake	#Subject	Side-face?	Demography?	Public?
2018	CelebA-HQ [6]	30K	30K	6217	Minimal	-	Yes
2019	FaceForensics++ [7]	1000	4000	977	Limited (varied angles)	-	Yes
2019	Celeb-DF [8]	590	5639	59	No (frontal celebrity deepfakes)	-	Yes
2020	DFDC [9]	23654	104500	960	Limited (primarily frontal)	-	Yes
2020	DeeperForensics [10]	48475	11000	100	Minimal (diverse expressions)	-	Yes
2020	KoDF [11]	62166	175776	403	Limited (self-recorded)	Korean	-
2021	FaceSynthesis [12]	100K	100K	334K	Limited	-	Yes
2021	OpenForensics [13]	45473	70325	–	Limited (multi-face)	-	Yes
2022	FMFCC-V [14]	44290	38102	83	Minimal	Asian	Yes
2023	DF-Platter [15]	764	132496	454	Limited (occlusions, multi-face)	Indian	Yes
2024	INDIFACE [16]	404	1668	58	Limited	Indian	Yes

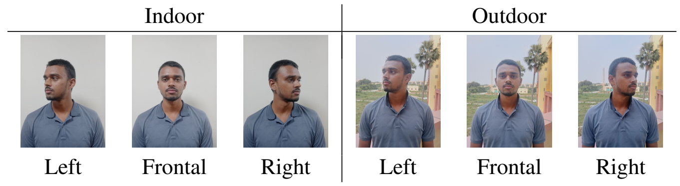

Fig. 1: Genuine data samples of a subject showing left, frontal, right side faces

(iii) Perspectives on generative models: We discuss broader implications of generative models in security-sensitive applications, highlighting their impact on identity verification, surveillance, and forensic investigations.
(iv) Future research directions: We propose recommendations for improving DF synthesis and detection, including the need for more diverse training datasets, adaptive generative models, multimodal detection approaches, and side-face-specific forensic techniques.

By addressing these critical gaps, we aim to advance the understanding of DF vulnerabilities in non-frontal perspectives and contribute to the development of more secure and reliable AI-driven media applications. Our work underscores the importance of addressing the challenges posed by side-face DFs, particularly in the context of ethnic diversity and real-world environmental conditions, to ensure the responsible deployment of generative AI technologies, while advancing the accuracy of DF technology and the resilience of AI systems against evolving spoofing techniques.

The remainder of this paper is organized as follows. Section II reviews relevant literature, and Section III details the dataset utilized, focusing on side-faces. Then Section IV explores challenges of deepfake generators. Section V examines the performance and limitations of existing deepfake detectors, followed by a discussion of key observations and potential future research directions. Section VI concludes the paper.

II. Brief Literature Review

DF generation has progressed from early computer graphics to deep learning frameworks. Traditional methods, such as key feature matching [17], struggled with pose, illumination, and occlusion variations. VAEs have significantly improved latent space modeling, leading to automated DF generation, with further advancements such as CVAEs [18] and VQ-VAEs [19] enhancing resolution and realism. GANs revolutionized image synthesis, with models like pix2pix [20] and CycleGAN [21] improving identity extraction and attribute integration. Advanced techniques, including FSGAN [22] and SimSwap++ [23], refined face-swapping, while high-resolution approaches like StyleIPSB [24] incorporated 3D priors for improved realism. Occlusion remained a challenge, partially addressed by FSGAN [22]. Recently, diffusion models have emerged as strong alternatives to GANs. Models such as DiffFace [25] and DiffSwap [26] achieve temporally consistent video DFs. Hybrid architectures integrating GANs with diffusion processes [27] further enhance synthesis.

As DF generation techniques grow more sophisticated, DF detection has become a critical research area. Different strategies are employed: some methods focus on spatial, frame-level characteristics, analyze texture/ lighting/ facial alignment, sometimes use efficient architectures like BNNs [28]. Others utilize frequency-based approaches, analyze wavelet/ Fourier transforms to detect high-frequency artifacts, as seen in Wavelet-CLIP [29]. To improve detection accuracy, multimodal techniques leverage diverse data sources [27].

Existing DF datasets primarily focus on frontal-face synthesis, with limited attention to side-face views. As shown in Table I, datasets such as FaceForensics++ [7], Celeb-DF [8], DFDC [9], and DeeperForensics [10] contain extensive manipulated and real videos, but lack substantial side-profile data, which is crucial for improving model robustness. The predominance of frontal and near-frontal views restricts DF detection and generation models from generalizing effectively across diverse facial angles. Additionally, datasets such as KoDF [11] and FMFCC-V [14] focus on specific ethnic groups. The need for Indian face datasets is evident, as current datasets provide limited representation of this demographic. INDIFACE [16] focuses on Indian subjects but remains small, highlighting the need for a larger dataset with side-profile Indian faces. While DF-Platter [15] provides Indian representation, but lacks side-profile data, indicating the need for improved datasets for DF generation and detection models across diverse populations and facial orientations.

III. Dataset with Side-Face

We used the IndicSideFace dataset [33], which contains both genuine and fake side-face images suitable for our empirical analysis. Below, we provide brief details of this dataset.

Genuine Dataset: The genuine dataset consists of images collected from 164 individuals, with each participant contributing 6 viewpoint images: two left-side, two frontal, and two right-side images. For each viewpoint, one image was captured indoors and the other outdoors. Thus, the dataset includes 6 categories of side-face images: left_indoor (LI), frontal_indoor (FI), right_indoor (RI), left_outdoor (LO), frontal_outdoor (FO), and right_outdoor (RO). As a result, the dataset comprises a total of 984 $(= 164 \times 6)$ images. A sample set of 6 side-face category images obtained from a single subject is presented in Fig. 1. Among the participants, 26 were female and 138 were male. Out of 164 individuals, 43 wore glasses.

Fig. 2: Synthetic *fake* images (marked by red colored boxes) of a male subject generated by identity swapping tools [3], [23], [30]–[32]. Source and corresponding target face images are *genuine*, and enclosed in blue and green boxes, respectively. (Best viewed in color)

Fake Dataset: For fake image generation, some major off-the-shelf generators were utilized. Five distinct identity swapping tools, Ghost [30], SimSwap [31], SimSwap++ [23], FaceDancer [3], and InsightFace [32], along with one attribute manipulation tool, FaceApp [34] were used.

For each genuine source image from six viewpoints, 3 target images were utilized in each of Ghost [30], SimSwap [31], SimSwap++ [23], and FaceDancer [3], while 2 target images were used in InsightFace [32]. Consequently, the identity-swapping tools generated a total of 84 ($= 6 \times 3 \times 4 + 6 \times 2$) fake images per subject. Additionally, for each genuine image of six viewpoints, 8 fake images were produced using eight distinct attribute filters—age, beard, expression, gender, glasses, hair color, hair style, and skin tone, available in FaceApp [34]. As a result, the attribute manipulation tool contributed 48 ($= 6 \times 8$) fake images per subject. In total, each subject contributed 132 ($= 84 + 48$) fake images, leading to an overall dataset of 21648 ($= 132 \times 164$) fake images across 164 subjects.

IV. Challenges in Generators

Despite advances in DF generations [3], [23], [30]–[32], [34], existing generators face below challenges in side-view perspectives and a variety of environmental conditions, affecting the realism and authenticity of synthetic images:

(i) Boundary artifacts: Boundary distortion, particularly unnatural blending at the interface between synthesized and original regions, is a key DF artifact. It becomes more evident in high-contrast settings and degrades realism.

(ii) Texture and skin tone inconsistencies: DF generation struggles with consistent skin texture, leading to artifacts like unnatural smoothness variations (blurring/roughness) and tone discrepancies. These inconsistencies are evident from side perspectives due to difficulties in blending regions.

(iii) Lighting and shadow mismatches: DFs often exhibit inconsistent lighting, leading to brightness and shadow mismatches across the face. Improper shadow rendering, particularly noticeable in side views, creates unnatural artifacts and aids detection, especially under non-uniform illumination.

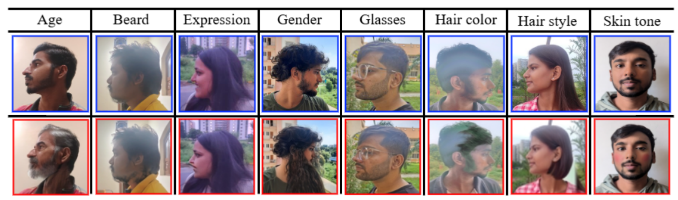

Fig. 3: Synthetic *fake* images (marked by red colored boxes) generated by attribute manipulation tool [34]. *Genuine* source images are enclosed in blue boxes.

(iv) Facial symmetry and structural distortions: Generating realistic side-face DFs is challenging due to asymmetry and geometry issues, as models distort features like ears, jawlines, and cheeks, compromising anatomical accuracy.

(v) Artifacts in facial hair and hair style Rendering: DF hair synthesis often fails, creating artifacts like unnatural hairline blending, inconsistent strand direction, and missing facial hair. These issues are more apparent in side views, with complex or curly/coarse hair types, and following extreme modifications (e.g., age/gender swaps) which can add unnatural styles/deformations.

(vi) Inaccurate representation of accessories: DF generators often produce noticeable artifacts in accessories like glasses, earrings, and headwear, causing distorted or misaligned sides. These artifacts, particularly in occluded or angled views, reduce realism of DF images and serve as indicators for detection, underscoring need for improved generation/detection.

(vii) Performance on non-Western ethnicities: DF generation models, primarily trained on Western facial features, often underperformed for Indian ethnicity and underrepresented groups. These models often fail to accurately represent culturally specific features like bindi marks or traditional jewelry.

(viii) Image quality degradation with distance from camera: Capturing subjects at a distance degrades DF realism due to resolution loss on smaller faces (blurring, pixelation, fewer

fine details).

(ix) Pose and viewpoint limitations: Identity-swapping models handle well frontal faces but struggle with extreme side-face angles and consistency when partially occluded/ rotated.

(x) Background Diversity: Lighting/shadow mismatches between retained backgrounds and synthesized faces are common identity-swap artifacts. Tools may also introduce unrealistic reflections/shadows, notably outdoors.

Some of these challenges in identity-swapped and attribute-manipulated fake images are shown in Fig.s 2, 3.

V. Experimental Analysis & Discussions

This section first analyzes experimental detector results to understand challenges faced by them under diverse indoor and outdoor lighting conditions with Indian ethnicity. From the detector perspective, we formulated the problem as a binary classification task to distinguish between genuine and fake images. We engaged three recent off-the-shelf detectors: BNN (Binary Neural Network) [28], Wavelet-CLIP [29], and DFDC (DeepFake Detection Challenge)-winner Selim [35].

We evaluated performance using Precision ($\mathcal{P}$), Recall ($\mathcal{R}$), F-measure ($\mathcal{FM}$), Accuracy ($\mathcal{A}$), and Balanced Accuracy ($\mathcal{BA}$). IndicSideFace dataset [33] poses a challenge due to class imbalance, containing 984 genuine and 21648 fake samples, where $\mathcal{BA}$ serves as a more suitable evaluation metric, as it averages the true positive rate and false positive rate, ensuring a balanced assessment of model performance.

A. Results

To ensure a comprehensive evaluation and to determine whether the detectors face challenges with synthetic fake images produced by generators, we assessed the baseline performances of the abovementioned pretrained detectors. Wavelet-CLIP [29], BNN [28], and Selim [35] detectors were pretrained on FaceForensics++ [7], DFFD [36], and DFDC [9] datasets, respectively. In our study, we employed a zero-shot evaluation strategy for pretrained detectors, where the entire IndicSideFace dataset [33] served as the test set, and the results are summarized in Table II. This table presents the performance of the detectors across six categories of side-face images: LI, FI, RI, LO, FO, RO (refer to Section III). Overall, BNN [28], Wavelet-CLIP [29], and Selim [9] achieved 53.98%, 50.44%, and 47.34% $\mathcal{BA}$, respectively. BNN performed better on frontal faces and outdoor categories, while Selim performed poorly on the LI.

We also evaluated the detector performances on separate groups of fake images generated by each of the employed generators [3], [23], [30]–[32], [34]. For the experiments, we paired all genuine images with individual generator-specific fake image groups. The results with respect to $\mathcal{BA}$ % are summarized in Table III. From this table, for example, it can be observed that by employing the above pretrained BNN on a test dataset comprising all genuine LI images and Ghost [30]-generated fake LI images, a $\mathcal{BA}$ of 52.21% was achieved. BNN and Wavelet-CLIP achieved the highest overall $\mathcal{BA}$ on SimSwap [31]-generated synthetic images among identity-swapping tools, while Selim attained the best overall $\mathcal{BA}$ on InsightFace [32]. For attribute manipulation, BNN, Wavelet-CLIP, and Selim yieled the highest overall $\mathcal{BA}$ on gender, age, and beard filters, respectively.

The results presented in Tables II and III indicate that existing detectors struggle with side-face DFs. The primary challenges are:

(i) Pose variations: Detection performances of some detectors [28] drop significantly for non-frontal poses due to reduced facial feature visibility.

(ii) Lighting conditions: Both indoor and outdoor lighting variations cause inconsistencies in detection.

(iii) Texture and shadow mismatches: Synthetic artifacts such as unnatural blending and shadow distortions hinder detection models.

(iv) Dataset bias: Pretrained models primarily trained on frontal images struggle with diverse ethnicities and pose variations.

Enhancing DF detection robustness against these challenges requires side-face-specific forensics, diverse datasets, and multimodal detection.

B. Observations

Despite advancements, DF generation still has limitations, particularly in realistic side-face synthesis under a variety of real-world conditions (indoor/ outdoor lighting). Our analysis reveals that current models struggle to maintain realism in these perspectives, particularly for people of Indian origin. The observed issues in side-face DF synthesis and potential solutions are discussed below:

(i) Boundary artifacts and blending issues: DFs often exhibit boundary artifacts along synthetic facial edges, which are exacerbated by high-contrast backgrounds, increasing detectability.

– Advanced blending techniques, such as adaptive Poisson image editing or neural rendering-based smooth transitions, can be integrated into DF models to improve boundary realism. Additionally, incorporating context-aware background synthesis can help mitigate unnatural blending artifacts.

(ii) Texture and skin tone inconsistencies: Many DF models struggle to maintain uniform texture and skin tone, particularly in side-face perspectives. The transition between real and synthetic skin regions often appears abrupt, with noticeable variations in smoothness and color consistency.

– Enhancing texture synthesis using adversarial texture refinement networks and training models on diverse datasets with varied skin tones can help address these inconsistencies. Additionally, style-based generative models with adaptive texture blending can ensure smoother transitions.

(iii) Lighting and shadow mismatches: Lighting inconsistencies, including unnatural shadow placement and brightness mismatches, significantly impact DF realism. These issues become more evident in side-face perspectives and under varying indoor and outdoor conditions.

TABLE II: Detector performances on indoor and outdoor side-faces

Detectors	BNN [28]							Wavelet-CLIP [29]							Selim [35]						
Category	LI	FI	RI	LO	FO	RO	**Overall**	LI	FI	RI	LO	FO	RO	**Overall**	LI	FI	RI	LO	FO	RO	**Overall**
$\mathcal{P}$ %	50.27	51.29	49.48	51.37	55.61	51.43	**51.57**	58.73	58.98	58.98	58.01	59.23	59.56	**58.92**	48.42	50.24	49.50	49.35	48.84	49.89	**49.37**
$\mathcal{R}$ %	78.43	64.71	73.47	82.91	76.47	84.31	**76.72**	98.98	98.74	99.37	97.14	97.99	98.46	**98.45**	86.18	90.25	84.67	91.22	91.45	94.52	**89.72**
$\mathcal{FM}$ %	61.26	57.22	59.13	63.43	64.39	63.88	**61.67**	73.71	73.83	74.02	72.64	73.83	74.22	**73.72**	62.01	64.54	62.47	64.04	63.67	65.3	**63.69**
$\mathcal{A}$ %	51.45	51.99	49.42	53.88	57.84	53.46	**53.01**	58.90	58.95	59.03	57.85	59.27	59.68	**58.95**	44.68	48.11	46.66	47.85	47.48	47.62	**47.07**
$\mathcal{BA}$ %	51.85	52.98	50.10	55.96	58.59	54.37	**53.98**	50.68	50.28	50.30	50.10	50.53	50.78	**50.44**	45.05	48.07	47.44	48.55	47.69	47.26	**47.34**

TABLE III: Detector performances ($\mathcal{BA}$ %) on generator-specific fake image groups paired with genuine images

		Identity Swapping					**Attribute Manipulation** [34]								
Detector	Category	Ghost [30]	SimSwap [31]	SimSwap++ [23]	FaceDancer [3]	InsightFace [32]	Age	Beard	Expression	Gender	Glasses	Hair color	Hair style	Skin tone	**Mean**
BNN [28]	LI	52.21	54.25	52.94	50.88	49.02	51.72	58.18	48.47	56.52	49.93	48.89	50.51	50.51	51.85
	FI	50.00	57.52	49.35	46.41	49.02	52.35	51.10	62.71	57.35	50.21	58.07	52.67	51.92	52.98
	RI	52.31	51.40	45.07	46.40	48.50	45.99	53.40	54.59	56.73	53.40	50.37	48.10	45.07	50.10
	LO	54.72	53.59	53.68	56.84	52.94	49.51	55.69	61.18	66.18	58.48	57.84	59.03	47.77	55.96
	FO	55.88	60.46	54.25	52.94	55.39	61.15	64.10	60.24	62.31	61.57	53.45	63.24	56.75	58.59
	RO	52.60	53.16	53.49	53.82	51.96	54.66	56.67	61.89	56.44	57.31	53.03	49.30	52.50	54.37
	Overall	**52.95**	**55.06**	**51.46**	**51.22**	**51.14**	**52.56**	**56.52**	**58.18**	**59.26**	**55.15**	**53.61**	**53.81**	**50.75**	**53.98**
Wavelet-CLIP [29]	LI	51.21	51.21	50.91	50.68	51.06	50.90	49.02	50.27	50.90	51.21	50.59	50.90	49.96	50.68
	FI	50.40	50.91	50.91	50.71	50.45	50.60	49.99	48.77	49.99	50.30	50.60	49.99	49.99	50.28
	RI	50.61	50.61	50.61	50.51	50.30	50.30	49.38	50.00	50.00	50.30	50.61	50.30	50.30	50.30
	LO	51.52	51.52	51.52	51.00	50.15	51.20	48.96	46.72	50.23	50.88	50.56	48.34	48.64	50.10
	FO	51.53	51.53	51.43	51.53	50.77	51.23	50.61	44.52	50.31	50.92	51.53	50.31	50.61	50.53
	RO	51.35	51.55	51.55	51.23	50.94	51.55	50.03	48.50	50.64	51.25	51.55	50.03	50.03	50.78
	Overall	**51.10**	**51.22**	**51.16**	**50.94**	**50.61**	**50.97**	**49.67**	**48.13**	**50.35**	**50.81**	**50.91**	**49.98**	**49.92**	**50.44**
Selim [35]	LI	40.20	37.58	39.87	41.63	51.47	48.39	48.51	48.26	51.96	41.25	39.06	48.73	48.73	45.05
	FI	43.87	45.75	45.42	39.54	51.47	51.51	52.94	49.37	50.86	49.37	50.08	48.25	46.42	48.07
	RI	41.53	42.10	36.10	43.44	53.63	51.40	51.77	51.53	50.10	48.44	48.28	48.28	50.10	47.44
	LO	44.81	44.55	44.39	46.94	52.45	50.86	51.33	47.94	48.77	49.10	51.09	51.16	47.77	48.55
	FO	47.55	47.39	46.08	45.42	50.00	43.63	51.96	45.96	39.00	48.63	51.96	50.40	51.96	47.69
	RO	44.03	45.33	39.33	47.33	49.51	50.00	50.00	50.00	50.00	48.48	45.65	46.43	48.28	47.26
	Overall	**43.67**	**43.78**	**41.87**	**44.05**	**51.42**	**49.30**	**51.09**	**48.84**	**48.45**	**47.55**	**47.69**	**48.88**	**48.88**	**47.34**

— The use of physics-based rendering models and neural relighting techniques can improve lighting consistency in DF images. By learning the spatial illumination properties from real-world samples, DF generators can produce more natural-looking lighting and shadow effects.

(iv) Facial symmetry and structural distortions: Side-face perspectives introduce challenges related to facial asymmetry, leading to distortions in ear, jawline, and cheek structures. These distortions are particularly noticeable when the subject's head is tilted or partially occluded.

— Geometric consistency constraints and 3D-aware generative models can be incorporated to preserve facial symmetry. Leveraging multi-view consistency learning can also help DF models generate structurally accurate facial features.

(v) Artifacts in facial hair and hair-styles: Rendering realistic facial hair and hair-styles remains a major challenge in DF synthesis. Inconsistent hair strand directions, missing facial hair portions, and unnatural blending of hairlines are commonly observed issues, especially for individuals with complex or coarse hair textures.

— High-fidelity hair synthesis models trained on diverse hairstyle datasets can improve DF quality. Additionally, attention-based generative adversarial networks (GANs) focusing on hair texture details can enhance realism.

(vi) Inaccurate representation of accessories: DF models often fail to accurately synthesize accessories such as glasses, earrings, and headwear. Issues include missing reflections, misaligned glasses frames, and incomplete synthesis of earrings or other objects.

— Integrating object-aware generative networks that explicitly model accessories alongside facial features can improve synthesis accuracy. Furthermore, leveraging physically based rendering techniques can ensure more realistic accessory representation.

(vii) Bias in DF generation for non-Western ethnicities: Many deepfake models are trained on datasets that predominantly feature Western facial features, leading to suboptimal performance for underrepresented ethnic groups.

— Expanding DF training datasets to include a broader range of ethnicities and cultural markers is essential. Additionally, developing ethnicity-aware generative models can ensure fairer and more accurate DF synthesis across diverse demographics.

(viii) Image quality degradation with distance from camera: DF image quality deteriorates when the subject is farther from the camera, leading to blurring, pixelation, and loss of fine-grained facial details.

— Super-resolution techniques, such as generative adversarial networks for image upscaling, can help enhance low-resolution DF images. Multi-scale training approaches can also improve the robustness of DF generation across varying distances.

(ix) Pose and viewpoint limitations: Identity-swapping DF models often struggle with extreme side-face angles, sometimes failing to generate a consistent face when the head is rotated or partially occluded.

— Pose-aware generative models trained on multi-angle datasets can improve synthesis quality for non-frontal faces. Incorporating 3D morphable models can also enhance facial consistency across different viewpoints.

(x) Environmental and background inconsistencies: DF identity-swapping often fails to modify background elements, leading to mismatches in lighting, shadows, and reflections.

— Implementing holistic scene-aware DF generation, where both the foreground and background are synthesized coherently, can improve visual realism. GAN-based scene adaptation techniques can also help reduce background mismatches.

(xi) Movement and animation: Maintaining consistent facial details and motion fluidity is challenging when the side-face is in motion, leading to unnatural transitions or visual artifacts.

— 3D facial modeling and multiview training enhance accuracy in different angles. Motion capture, temporal consistency models, and facial landmark detection help ensure smooth, fluid animations. GANs and pose-aware networks refine movement, reducing visual artifacts and ensuring natural transitions.

C. Future Research Directions

To address the above challenges and improve DF generation/ detection, we propose some future research directions:

(i) Dataset diversity: Generalization is enhanced by diversifying training datasets in terms of variety of ethnicity, poses, lighting, and accessories.

(ii) Adaptive generative models: Incorporating physics-based rendering, multi-view learning, and 3D-aware GANs for more accurate DF synthesis.

(iii) Multimodal detection approaches: Combining audio, thermal imaging, and physiological signals with visual analysis to enhance DF detection.

(iv) Forensic techniques for side-face DFs: Developing specialized forensic analysis tools for detecting inconsistencies in side-face DFs, focusing on symmetry, lighting, and texture anomalies.

(v) Contextual Integration: Enhancing DF systems to better blend the manipulated subject into real-world environments, especially when generating side-face views under varied lighting and backgrounds.

By implementing these solutions, DF synthesis can be improved for side-face perspectives while detection models can be strengthened to counter emerging DF threats in security-sensitive applications.

VI. Conclusion

DF technology continues to evolve, pushing the boundaries of AI-generated media while simultaneously exposing new vulnerabilities. In this position paper, we have critically examined the limitations of state-of-the-art DF generators, particularly in side-face perspectives under diverse environmental conditions. Our analysis highlights key synthesis challenges, including boundary artifacts, lighting inconsistencies, and difficulties in rendering facial features, especially for individuals of Indian ethnicity. These shortcomings not only reduce the realism of generated images but also create exploitable weaknesses for DF detection models. Our evaluation of existing detection frameworks demonstrates that while they effectively identify frontal DFs, they often struggle with side-face inconsistencies, leading to gaps in real-world applicability. This underscores the urgent need for more robust and adaptive detection techniques capable of handling diverse facial angles, lighting conditions, and demographic variations. To address these challenges, we advocate for a multi-pronged approach: enhancing dataset diversity to improve generative model generalization, refining DF synthesis methods to reduce artifacts, and developing multimodal detection strategies that leverage spatial, temporal, and contextual cues. By strengthening both generative and forensic capabilities, we can move toward a more secure and trustworthy AI-driven media landscape. Future research should balance DF generation and detection through collaborative efforts to prevent misuse while leveraging its benefits in creative and assistive AI.

Acknowledgment

The authors heartily thank all the volunteers, and interns/ trainees/ researchers, who helped in database generation. Partial support from IHUB-NTIHAC Foundation, IIT Kanpur, under proposal no. 3092 is gratefully acknowledged.

References

[1] J. W. Seow *et al.*, "A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities," *Neurocomputing*, vol. 513, pp. 351–371, 2022.

[2] Z. Akhtar *et al.*, "Video and audio deepfake datasets and open issues in deepfake technology: being ahead of the curve," *Forensic Sciences*, vol. 4, no. 3, pp. 289–377, 2024.

[3] F. Rosberg *et al.*, "Facedancer: Pose-and occlusion-aware high fidelity face swapping," in *WACV*, 2023, pp. 3454–3463.

[4] K. Liu *et al.*, "Deepfacelab: Integrated, flexible and extensible face-swapping framework," *Pattern Recognition*, vol. 141, p. 109628, 2023.

[5] G. Li, X. Zhao, and Y. Cao, "Forensic symmetry for deepfakes," *IEEE TIFS*, vol. 18, pp. 1095–1110, 2023.

[6] H. Huang *et al.*, "IntroVAE: Introspective variational autoencoders for photographic image synthesis," *NeurIPS*, vol. 31, 2018.

[7] A. Rössler *et al.*, "Faceforensics++: Learning to detect manipulated facial images," in *ICCV*, 2019, pp. 1–11.

[8] Y. Li *et al.*, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," in *CVPR*, 2020, pp. 3207–3216.

[9] B. Dolhansky *et al.*, "The DeepFake Detection Challenge (DFDC) Dataset," *arXiv:2006.07397*, 2020.

[10] L. Jiang and o. Li, "DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection," in *CVPR*, 2020.

[11] P. Kwon and o. You, "KoDF: A Large-scale Korean DeepFake Detection Dataset," in *ICCV*, 2021, pp. 10 744–10 753.

[12] E. Wood *et al.*, "Fake it till you make it: face analysis in the wild using synthetic data alone," in *ICCV*, 2021, pp. 3681–3691.

[13] T.-N. Le *et al.*, "Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild," in *ICCV*, 2021, pp. 10 117–10 127.

[14] G. Li *et al.*, "FMFCC-V: An Asian Large-Scale Challenging Dataset for DeepFake Detection," in *ACM IH-MMSec*, 2022, pp. 7–18.

[15] K. Narayan and o. Agarwal, "DF-Platter: Multi-Face Heterogeneous Deepfake Dataset," in *CVPR*, 2023, pp. 9739–9748.

[16] K. Kuckreja *et al.*, "INDIFACE: Illuminating India's Deepfake Landscape with a Comprehensive Synthetic Dataset," in *FG*, 2024, pp. 1–9.

[17] D. Bitouk *et al.*, "Face swapping: automatically replacing faces in photographs," *ACM Trans. Graph.*, vol. 8, no. 3, pp. 1–8, 2008.

[18] K. Sohn and o. Yan, "Learning structured output representation using deep conditional generative models," in *NeurIPS*, vol. 28, 2015.

[19] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *NeurIPS*, vol. 30, 2017.

[20] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[21] J.-Y. Zhu *et al.*, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232.

[22] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject agnostic face swapping and reenactment," in *Proc. ICCV*, 2019, pp. 7184–71 932.

[23] X. Chen *et al.*, "SimSwap++: Towards faster and high-quality identity swapping," *IEEE TPAMI*, vol. 46, no. 1, pp. 576–592, 2023.

[24] D. Jiang and o. Song, "Styleipsb: Identity-preserving semantic basis of stylegan for high fidelity face swapping," in *CVPR*, 2023, pp. 352–361.

[25] K. Kim *et al.*, "DiffFace: Diffusion-based face swapping with facial guidance," *arXiv:2212.13344*, 2022.

[26] W. Zhao and o. Rao, "Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion," *CVPR*, pp. 8568–8577, 2023.

[27] T. Oorloff and o. Koppisetti, "AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection," in *CVPR*, 2024, pp. 27 102–27 112.

[28] R. Lanzino *et al.*, "Faster than lies: Real-time deepfake detection using binary neural networks," in *CVPR Workshops*, 2024, pp. 3771–3780.

[29] L. B. Baru *et al.*, “Wavelet-driven generalizable framework for deepfake face forgery detection,” *arXiv:2409.18301*, 2024.

[30] A. Groshev *et al.*, “Ghost—a new face swap approach for image and video domains,” *IEEE Access*, vol. 10, pp. 83 452–83 462, 2022.

[31] R. Chen, X. Chen, B. Ni, and Y. Ge, “Simswap: An efficient framework for high fidelity face swapping,” in *ACM MM*, 2020, pp. 2003–2011.

[32] J. Guo, J. Deng *et al.*, “InsightFace: 2D and 3D Face Analysis Project,” https://github.com/deepinsight/insightface, Accessed: 11 June 2025.

[33] A. Deo *et al.*, “IndicSideFace: A Dataset for Advancing Deepfake Detection on Side-Face Perspectives of Indian Subjects,” in *FG*. IEEE, 2025, pp. 1–10.

[34] FaceApp Technology Ltd., “FaceApp,” https://www.faceapp.com, Accessed: 11 June 2025.

[35] S. Seferbekov, “DeepFake Detection (DFDC) Solution,” https://github.com/selimsef/dfdc_deepfake_challenge, Accessed: 11 June 2025.

[36] H. Dang *et al.*, “ On the Detection of Digital Face Manipulation ,” in *CVPR*. IEEE, 2020, pp. 5780–5789.