2025 19th International Conference on Automatic Face and Gesture Recognition (FG)

IndicSideFace: A Dataset for Advancing Deepfake Detection on Side-Face Perspectives of Indian Subjects

Anurag Deo[1], Aditya Bangar[2], Chandranath Adak[1], Rahul Verma[1], Deepak Nagar[1], Zahid Akhtar[3], Soumya Dutta[2], Soumi Chattopadhyay[4], Sukalpa Chanda[5]
[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
[5] Østfold University College, Norway–1757

***Abstract*— The rapid advancement of generative models and their misuse have made deepfake detection a crucial area of research. However, existing datasets and detection techniques predominantly focus on frontal-face perspectives, leaving side-face views largely underexplored. To bridge this gap, we present IndicSideFace, a novel dataset specifically curated for advancing deepfake detection on side-face perspectives of Indian subjects. This dataset encompasses a diverse range of side-face angles, varying lighting conditions, and demographic attributes, providing a comprehensive benchmark for evaluating detection algorithms. Our experiments using state-of-the-art models highlight the unique challenges posed by side-face deepfakes, such as partial facial feature visibility and uncommon head poses. The findings reveal significant limitations in existing detection approaches when applied to side-face perspectives, underscoring the need for specialized solutions. With IndicSideFace, we aim to strengthen the resilience of deepfake detectors and stimulate further research in this critical yet underexplored domain.**

I. Introduction

The advancement of AI and generative deep neural networks has revolutionized digital content creation, enabling realistic deepfake media that alter facial appearances, voices, and actions in videos, images, and audio [38]. While deepfakes offer benefits in entertainment, content enhancement, and education, their misuse for misinformation, fraud, and identity manipulation underscores the urgent need for robust detection mechanisms [2].

Despite significant advancements in deepfake detection, the majority of research efforts and publicly available datasets have concentrated on frontal-face perspectives, leaving side-face views underexplored [30]. This gap in research is critical, as deepfake manipulations of side-face views introduce distinct challenges that differ from frontal-face alterations. The visibility of facial features in side-face perspectives is often limited due to occlusions, uncommon head poses, and variations in environmental factors such as lighting conditions [34]. These characteristics make side-face deepfakes harder to detect using existing methods, which primarily rely on full facial visibility and symmetry-based analysis. Furthermore, side-face deepfakes pose significant risks in scenarios where frontal views are not available, such as in surveillance footage, candid photography, or social media images captured from different angles. Since current deepfake detection models are predominantly trained on frontal perspectives, their ability to generalize to side-face scenarios remains inadequate, leading to an increased risk of undetected manipulations in real-world applications [25].

From a technical perspective, distinguishing between indoor and outdoor side-face deepfakes is essential due to the considerable variations in lighting, background complexity, and environmental occlusions. Indoor environments typically feature controlled lighting conditions, which minimize extreme shadows and reflections, making it easier to identify inconsistencies in deepfake manipulations. However, artificial lighting used indoors can also create uniform illumination, allowing deepfake models to blend synthetic features more seamlessly, thereby reducing detection effectiveness. In contrast, outdoor environments introduce additional complexities, such as natural lighting variations, dynamic shadows, and fluctuating illumination levels throughout the day. Outdoor settings also present unpredictable occlusions, such as obstacles partially covering the face or interference from moving objects, which can obscure facial details. Moreover, factors like motion blur and depth-of-field variations add further challenges to deepfake detection models, particularly those trained predominantly on controlled indoor datasets. By incorporating a diverse range of indoor and outdoor images, IndicSideFace ensures that detection models are better equipped to handle real-world variations, ultimately improving their robustness and generalizability. In the Indian context, the emergence and rapid dissemination of deepfake technology presents unique challenges, particularly concerning misinformation, political propaganda, and digital fraud. With social media shaping public discourse, deepfakes can manipulate opinions, incite unrest, and harm reputations. The rise of AI-driven authentication, including biometric verification in banking, governance, and law enforcement, further heightens security concerns. Deepfake-based identity fraud threatens national security and legal integrity.

To bridge the gap in deepfake detection research, we introduce IndicSideFace, a novel dataset explicitly designed to advance side-face deepfake detection, particularly for Indian subjects. This dataset encompasses a broad spectrum of side-face images captured under varying conditions, including different angles, illumination settings (both indoor and outdoor), and demographic variations. By providing

979-8-3315-5341-8/25/$31.00 ©2025 IEEE

a well-structured benchmark, IndicSideFace facilitates the evaluation of deepfake detection algorithms under realistic conditions, enabling a more comprehensive understanding of their limitations and potential improvements. Through extensive experimental analyses leveraging state-of-the-art deepfake detection techniques, we assess the performance of existing approaches in handling side-face deepfakes. Our findings reveal that current detection frameworks struggle with side-face perspectives, underscoring the necessity for specialized detection solutions that can effectively mitigate the risks associated with side-face deepfake manipulations.

By creating IndicSideFace dataset, we aim to foster further research in this underexplored domain, encouraging the development of more resilient and adaptive deepfake detection methods. Our primary contributions include: ***(a)*** creation of a high-quality, curated dataset comprising both real and deepfake side-face images, ***(b)*** an in-depth evaluation of existing deepfake detection techniques on side-face perspectives to highlight their limitations, and ***(c)*** valuable insights into the challenges posed by side-face deepfakes, along with recommendations for future advancements in detection methodologies. This work aims at strengthening digital security and ensuring the authenticity of visual content in an increasingly AI-driven world.

The rest of the paper is organized as follows. Section II provides a concise overview of the existing literature. The following Section III describes the proposed dataset, IndicSideFace, including details on the generators utilized for synthetic fake image generation. Section IV briefly introduces the detectors employed for benchmarking, and Section V presents the experimental setups and discusses the results. Finally, Section VI concludes this paper.

II. Brief Literature Review

This section enlists some popular publicly available deepfake datasets, off-the-shelf sythetic fake image generators, and deepfake detectors [38].

A. Deepfake Datasets

We begin by summarizing some popular publicly available deepfake datasets, as outlined in Table I.

1) FaceForensics++ (FF++) [35]: This benchmark dataset includes 1000 genuine videos, each manipulated using four automated face forgery techniques: Deepfakes (DF), FaceSwap (FS), Face2Face (F2F), and Neural Textures (NT), resulting in 4000 fake videos. DF uses autoencoders for face replacement, FS relies on landmark-based graphics, F2F transfers expressions while retaining identity, and NT modifies mouth movements using GAN-based rendering. Although this dataset offers high-quality manipulations, it lacks diversity in environmental settings and subject demographics, limiting its effectiveness in real-world scenarios.

2) Celeb-DF [28]: This dataset contains 590 authentic and 5639 deepfake videos featuring 59 celebrities, with a balanced representation across gender, age, and ethnicity. The fake samples are generated using an improved synthesis pipeline that enhances temporal coherence and color alignment through Kalman filtering and data augmentation. Despite these advancements, the dataset predominantly includes Western celebrities, which may limit its applicability for deepfake detection across diverse demographic groups.

TABLE I: Existing major deepfake datasets

Dataset	#Genuine	#Fake	#Subjects	Demography?	Side-face?
FaceForensics++ [35]	1000	4000	977	–	Limited (varied angles)
Celeb-DF [28]	590	5639	59	–	No (frontal celebrity deepfakes)
DFDC [10]	23654	104500	960	–	Limited (primarily frontal)
DeepForensics [20]	50000	10000	100	–	Minimal (diverse expressions)
INDIFACE [22]	404	1668	58	Indian	Limited
KoDF [23]	62166	175776	403	Korean	Limited (self-recorded)
DF-Platter [32]	764	132496	454	Indian	Limited (multi-face)

3) DeepFake Detection Challenge Dataset (DFDC) [10]: This dataset comprises 23654 real and 104500 fake videos involving 960 subjects spanning diverse ethnicities, age groups, and genders, captured under varied environmental conditions. It features multiple manipulation techniques, including DFAE, MM/NN face swap, NTH, FSGAN, and StyleGAN. Although DFDC offers substantial real-world variability, its emphasis on Western subjects and predominantly front-facing videos may hinder the effectiveness of detection models on Indian demographics and side-face scenarios.

4) DeeperForensics-1.0 [20]: This dataset comprises 10000 real and 50000 fake videos generated from 100 subjects. Face reenactment and swapping were performed using DF-VAE, which enables precise disentanglement of pose and texture. Despite its scale, the dataset was collected under controlled conditions and primarily uses synthetic distortions, limiting its ability to reflect real-world variability and authentic manipulation artifacts.

5) INDIFACE [22]: This dataset addresses the underrepresentation of Indian demographics in deepfake datasets. It features 404 real and 1668 fake videos generated using SimSwap and Ghost. It captures diverse Indian faces with variations in skin tone, facial structure, and cultural backgrounds, and includes real-world perturbations such as Gaussian blur and brightness changes. Fine-tuning on INDIFACE improves detection performance, highlighting the need for demographic-specific datasets.

6) Korean DeepFake Detection Dataset (KoDF) [23]: This dataset comprises 62166 real and 175776 fake videos aimed at improving representation of Korean subjects. It employs six manipulation methods: FaceSwap, DeepFaceLab, FSGAN, FOMM, ATFHP, and Wav2Lip. It ensures quality through carefully curated real clips and validated synthetic videos, filling gaps in subject representation across benchmarks.

7) DF-Platter [32]: This dataset tackles real-world challenges such as low resolution, occlusion, and multiple faces. It comprises 764 real and 132496 fake videos generated using FSGAN, FaceShifter, and FaceSwap. It emphasizes Indian ethnicity with a balanced gender and age distribution. While it strengthens evaluation under practical conditions, it includes limited coverage of side-face perspectives.

B. Generators

Deepfake generation has rapidly advanced with the emergence of sophisticated generative models, primarily pow-

ered by GANs, VAEs, and more recently, diffusion-based architectures. These models aim to synthesize or manipulate facial identities, expressions, and movements in videos or images while ensuring high visual realism and temporal coherence [16]. Early approaches like faceswap and deepfakes relied on autoencoder frameworks for face replacement in video frames. Subsequent models, such as Face2Face [41] and Neural Textures [42], introduced real-time facial reenactment and texture-based rendering. More advanced architectures like FaceShifter [26] and FSGAN [33] addressed robustness to occlusions, identity mismatches, and pose variations by incorporating identity-aware synthesis and reenactment pipelines. DeepFaceLab [30] further popularized customizable deepfake creation with modular autoencoder-based pipelines, while SimSwap [7] unified identity encoding and style-based generation for arbitrary face swaps. Recent innovations, such as FaceDancer [34] attempted to enhance face swapping performance under non-frontal head poses by employing attention-based feature fusion and pose-aware regularization. Disentangled representation learning has become prominent in improving control and realism. RelGAN [45] separated semantic and structural information, enabling targeted manipulation. In attribute editing, models like TUSLT [44] used CLIP-based supervision for multi-attribute transformation in StyleGAN latent space, while AU-EditNet [21] incorporated cross-branch interaction for improved disentanglement of identity and expression features. DreamSalon [29] introduced a two-stage diffusion framework enabling fine-grained attribute editing while preserving contextual identity. The recent shift toward diffusion-based models offers improved video consistency and realism. Latent flow diffusion [6] enhanced temporal coherence by modeling optical flow in latent space and integrating frequency and spatial cues. Multi-modal techniques [39], [49] further enriched generation by combining facial landmarks, audio, and motion features. Prototype-driven methods [48] improved generalization using shared latent representations across identities.

Despite significant progress, challenges remain in handling extreme non-frontal head poses, fast motion, and generalizing to unseen identities, particularly in real-world or demographically diverse scenarios. Reducing artifacts and maintaining authenticity continue to drive ongoing research in deepfake generation.

C. Detectors

Deepfake detection has become an essential research area in response to the increasing realism of generative models. Initial approaches focused on visual artifacts such as inconsistent blinking [27] and abnormal head poses [47], but these became less effective as synthesis techniques evolved. To overcome these limitations, deep learning-based methods were introduced. MesoNet [1] proposed a compact CNN that captured mesoscopic features, while XceptionNet-based detectors [35] achieved strong results through transfer learning on large-scale datasets like FaceForensics++. Temporal approaches, including LSTM-based [17] and recurrent convolutional architectures [36], modeled spatio-temporal inconsistencies across frames, improving performance on manipulated videos. Beyond RGB-based detection, frequency-domain techniques [12], [14] analyzed spectral discrepancies that often revealed generative artifacts not visible in the spatial domain. Transformer-based models such as TransForensics [19] leveraged self-attention mechanisms to capture fine-grained manipulations. Additionally, multi-modal methods combining facial cues with audio [31] or biological signals [9] offered improved robustness under challenging conditions.

Despite these advancements, empirical evidence suggests that most detectors are trained and evaluated predominantly on near-frontal facial inputs. As a result, they struggle to generalize to profile or side-face views, where modern deepfakes remain highly convincing. This poses a serious limitation for real-world deployment.

III. Proposed Dataset: IndicSideFace

This section presents the proposed IndicSideFace dataset, outlining the collection process of real images and the manipulation techniques used to generate fakes.

A. Genuine Data Collection

To capture the diversity of India's population, we selected 164 individuals of Indian origin from various regions, encompassing a broad spectrum of skin tones and facial features. This selection aimed to represent a wide range of ethnic and cultural backgrounds, ensuring the dataset is both comprehensive and inclusive. The gender distribution in this dataset includes 26 females, and 138 males. Additionally, 43 out of the 164 subjects wore glasses.

We collected 6 genuine images from an individual wearing daily attire (refer to top row of Fig. 1 for an example). Three of these images were captured indoors, showcasing frontal, left, and right side face views. The remaining three images were taken outdoors, providing a variety of background contexts. In total, the dataset comprises 984 $(= 6 \times 164)$ genuine face images. The images were primarily captured upto the bust to ensure that the person's head, face, and shoulders were clearly visible. However, for added naturalness and diversity, some images included slight cropping of the head and side portions. To further enhance the diversity of our dataset, we utilized various smartphone cameras, with a majority of the images taken using the subjects' personal smartphones.

B. Synthetic Fake Data Generation

We generated fake face images synthetically using five different identity-swapping tools, and one attribute manipulation tool, as discussed below.

1) Identity Swapping Tools:

a) Ghost [15]: This framework is a one-shot transfer method for face swapping, effectively combining identity transfer and attribute preservation. It features an identity encoder, a U-Net-based attribute encoder, an AAD generator for combining identity and attribute vectors, and a multiscale

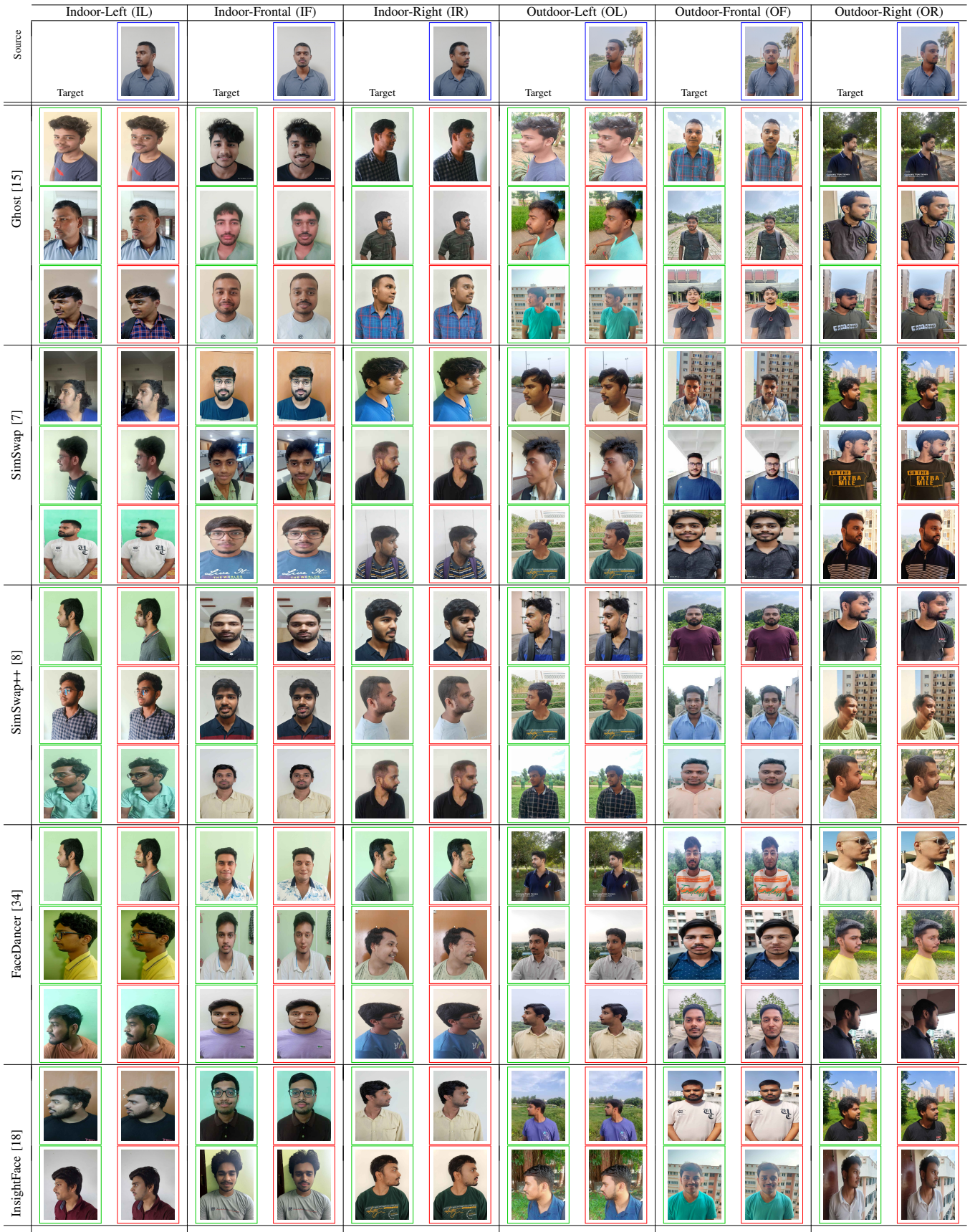

Fig. 1: Synthetic *fake* images of a male subject in IndicSideFace generated by identity swapping tools [7], [8], [15], [18], [34]. Source and corresponding target face images are *genuine*. (Best viewed in color)

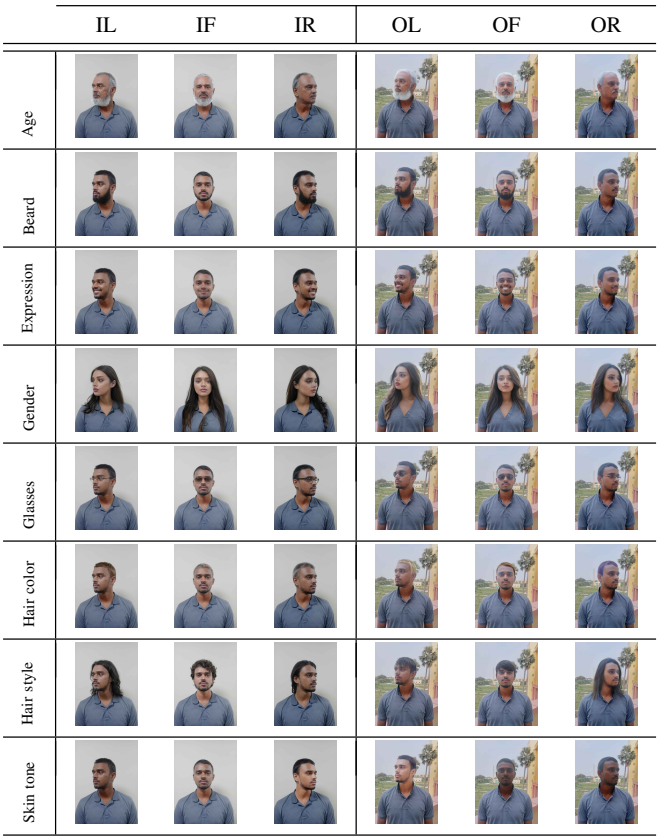

Fig. 2: Attribute manipulated *fake* images of the same source images of Fig. 1 in IndicSideFace.

discriminator for ensuring image quality. Enhanced by a custom loss function, including reconstruction, attribute, identity, adversarial, and eye loss, it maintains gaze consistency in swaps. The model addresses shape mismatches through landmark tracking and ensures stability via bounding box smoothing. Pretrained on the VGG-Face2 dataset [5], Ghost was used to generate 2952 $(= 984 \times 3)$ fake images by pairing each of the 984 genuine face images with 3 different randomly chosen target images in this study.

b) SimSwap [7]: This simple face-swapping framework, designed for high fidelity and generalization to arbitrary identities, transfers the identity of a source face onto a target face while preserving the target's key attributes, such as facial expressions and gaze direction. This is achieved through the *ID injection module* that embeds source identity information at the feature level, and the *weak feature matching loss* that implicitly ensures facial attribute preservation. Here also, it was pretrained using the VGG-Face2 [5], and each of the 984 genuine images leveraged 3 different target images to produce 2952 fake images.

c) SimSwap++ [8]: It is an upgraded version of SimSwap [7] for face identity editing, which introduces *conditional dynamic convolution* that enhances efficiency by enabling anisotropic processing and injection with lower complexity, and *morphable knowledge distillation*, a heterogeneous teacher-student framework that maximally retains the teacher's knowledge while reducing model complexity through structure re-parameterization. Similar to the above strategy, we generated 2952 fake images utilizing SimSwap++, pretrained on VGGFace2-HQ [8].

TABLE II: Sample count of IndicSideFace

		Generator	Indoor			Outdoor			
Side-face			Left	Frontal	Right	Left	Frontal	Right	**Total**
Genuine samples			164	164	164	164	164	164	**984**
Fake samples	Identity Swapping	Ghost [15]	492	492	492	492	492	492	2952
		SimSwap [7]	492	492	492	492	492	492	2952
		SimSwap++ [8]	492	492	492	492	492	492	2952
		FaceDancer [34]	492	492	492	492	492	492	2952
		InsightFace [18]	328	328	328	328	328	328	1968
	Attribute Manipulation	FaceApp [13]	1312	1312	1312	1312	1312	1312	7872
									21648

d) FaceDancer [34]: This is a single-stage framework for subject-agnostic face swapping and identity transfer, featuring two key innovations: AFFA and IFSR. The AFFA module, embedded in the decoder, adaptively fuses attribute features and identity-conditioned features without requiring additional facial segmentation. IFSR leverages intermediate features in the identity encoder to preserve critical target face attributes, such as head pose, facial expressions, lighting, and occlusions, while achieving high-fidelity identity transfer from the source face. In this paper, we engaged FaceDancer pretrained on the VGGFace2 [5] and LS3D-W [4] datasets to obtain 2952 fake images, similar to the aforementioned strategy.

e) InsightFace [18]: For synthetic image generation, we also utilize *Picsi.Ai* face-swapping service, powered by InsightFace, which integrates inswapper_cyn and inswapper_dax face-swapping models. Here, for each of the 984 genuine face images, we paired them with 2 different target images, resulting in a total of 1968 $(= 984 \times 2)$ fake images.

For all of the aforementioned identity swapping tools, source and target images are of the same side profile and gender. In Fig. 1, we present 84 $(= 14 \times 6)$ synthetic fake images of a male subject generated by the above identity swapping tools.

2) Attribute Manipulation Tool:

a) FaceApp [13]: This is one of the widely popular photo editing applications that we utilized for facial attribute manipulation. We applied 8 attribute filters, including age, beard, expression, gender, glasses, hair color, hair style, and skin tone, to each of the 984 genuine face images for generating a total of 7872 $(= 984 \times 8)$ fake images. Fig. 2 shows 48 $(= 8 \times 6)$ fake sample images obtained from this attribute manipulation tool of the same male subject shown in Fig. 1.

Table II provides a summary of the sample distribution in our dataset, IndicSideFace, which comprises a total of 984 genuine samples and 21648 fake samples from 164 subjects. Each of the 6 side face categories (i.e., IL: indoor_left, IF: indoor_frontal, IR: indoor_right, OL: outdoor_left, OF: outdoor_frontal, OR: outdoor_right) includes 164 genuine samples, and 3608 fake samples.

C. Informed Consent

Informed consent was obtained from each participant before any facial data was collected. Clear explanations were pro-

TABLE III: Performance ($\mathcal{BA}$ %) of pretrained detectors on generator-specific fake image groups paired with genuine images

		Identity Swapping					Attribute Manipulation								
Detector	**Category**	SimSwap [7]	SimSwap++ [8]	Ghost [15]	FaceDancer [34]	InsightFace [18]	Age	Beard	Expression	Gender	Glasses	Hair color	Hair style	Skin tone	**Mean**
Selim [37]	IL	37.58	39.87	40.20	41.63	51.47	48.39	48.51	48.26	51.96	41.25	39.06	48.73	48.73	45.05
	IF	45.75	45.42	43.87	39.54	51.47	51.51	52.94	49.37	50.86	49.37	50.08	48.25	46.42	48.07
	IR	42.10	36.10	41.53	43.44	53.63	51.40	51.77	51.53	50.10	48.44	48.28	48.28	50.10	47.44
	OL	44.55	44.39	44.81	46.94	52.45	50.86	51.33	47.94	48.77	49.10	51.09	51.16	47.77	48.55
	OF	47.39	46.08	47.55	45.42	50.00	43.63	51.96	45.96	39.00	48.63	51.96	50.40	51.96	47.69
	OR	45.33	39.33	44.03	47.33	49.51	50.00	50.00	50.00	50.00	48.48	45.65	46.43	48.28	47.26
	Overall	**43.78**	**41.87**	**43.67**	**44.05**	**51.42**	**49.30**	**51.09**	**48.84**	**48.45**	**47.55**	**47.69**	**48.88**	**48.88**	**47.34**
CNN-Detector [43]	IL	50.00	50.31	50.98	50.00	50.49	49.19	50.98	50.98	47.13	50.98	49.37	50.98	50.98	50.18
	IF	53.27	53.27	50.25	50.98	50.98	51.06	52.36	50.35	47.67	52.14	52.49	52.36	51.75	51.46
	IR	49.02	49.69	48.51	47.02	48.57	51.02	46.02	47.45	46.02	49.35	51.02	46.47	49.35	48.42
	OL	49.33	48.03	49.75	49.33	49.51	50.00	48.39	50.00	50.00	48.08	50.00	50.00	48.28	49.28
	OF	49.67	48.69	49.75	49.67	49.02	50.00	50.00	50.00	48.15	50.00	50.00	48.44	50.00	49.49
	OR	50.29	50.29	49.97	50.63	50.49	48.39	50.35	50.64	50.18	50.45	51.96	51.96	51.96	50.58
	Overall	**50.26**	**50.05**	**49.87**	**49.61**	**49.84**	**49.94**	**49.68**	**49.90**	**48.19**	**50.17**	**50.81**	**50.04**	**50.39**	**49.90**
ViT [46]	IL	39.18	36.04	40.03	44.59	31.18	31.47	32.59	30.32	38.98	33.74	31.52	39.96	33.76	35.64
	IF	38.55	48.28	34.71	36.13	37.96	37.38	30.22	38.05	35.90	35.99	36.67	39.41	36.15	37.34
	IR	38.41	35.18	34.85	39.03	43.67	38.32	39.15	37.03	35.27	39.54	31.81	30.86	38.84	37.07
	OL	40.54	37.32	35.03	37.21	37.83	36.10	33.04	35.89	33.05	30.26	38.01	35.45	44.81	36.50
	OF	37.31	36.19	30.96	35.40	39.39	31.63	38.87	31.33	30.08	33.27	32.47	32.46	39.15	34.50
	OR	38.11	41.17	35.00	34.05	37.78	35.24	32.84	31.74	30.62	38.71	32.38	34.05	33.28	35.00
	Overall	**38.68**	**39.03**	**35.09**	**37.74**	**37.97**	**35.02**	**34.45**	**34.06**	**33.98**	**35.25**	**33.81**	**35.37**	**37.67**	**36.01**
NPR [40]	IL	49.09	49.90	48.48	47.88	55.48	43.08	49.33	45.89	42.76	45.89	54.75	43.39	44.01	47.69
	IF	45.06	48.08	47.49	47.49	50.83	43.52	47.18	44.43	43.52	42.30	48.70	42.91	43.82	45.79
	IR	47.63	48.45	49.46	49.16	52.70	46.58	47.50	48.42	47.19	46.58	50.56	49.03	47.19	48.50
	OL	49.71	48.58	49.90	48.55	48.96	45.88	45.96	47.42	47.49	44.19	42.90	40.43	46.45	46.65
	OF	49.33	48.42	49.33	50.35	54.93	46.84	40.74	47.75	45.85	41.35	38.79	40.02	46.84	46.20
	OR	49.47	48.58	48.95	49.47	51.69	48.80	48.19	48.49	49.10	47.58	50.32	50.32	49.96	49.30
	Overall	**48.38**	**48.67**	**48.94**	**48.81**	**52.43**	**45.78**	**46.48**	**47.07**	**45.98**	**44.65**	**47.67**	**44.35**	**46.38**	**47.35**

vided regarding the objectives of the research project and the purpose of collecting their facial data. Participation was voluntary, and participants were assured that their data would be used for research purposes.

IV. Benchmarking

To benchmark our IndicSideFace dataset, we employed four state-of-the-art off-the-shelf baseline detectors to identify fake images, as detailed below:

(i) Selim (DFDC Winner): Selim, the winner of the DeepFake Detection Challenge (DFDC) [10], leverages EfficientNet-B7 as its encoder and employs multitask cascaded CNN as the detector for its superior speed and memory efficiency [37]. The model's performance is finetuned by optimizing input sizes, while its generalization capability is bolstered through advanced data augmentation techniques, including isotropic scaling and dropout-based variations.

(ii) NPR-based Detector: Up-sampling operators in CNN-based generators induce local pixel interdependencies, resulting in generalized forgery artifacts. Neighboring Pixel Relationships (NPR) [40] effectively characterizes and detects these structural artifacts in GAN- and diffusion-generated images.

(iii) CNN-Detector: This detector [43] operates on the principle that, with proper pre- and post-processing and data augmentation, a standard image classifier trained on a single CNN generator can effectively generalize to unseen samples.

(iv) ViT-based Detector: Vision Transformer (ViT) [11] has demonstrated well performance in classification tasks by utilizing self-attention mechanisms to capture long-range dependencies and global context. Recognizing its capability to detect subtle forgery artifacts and generalize across datasets, we engage ViT for deepfake detection [46] to evaluate its effectiveness in identifying spatial and structural inconsistencies.

TABLE IV: Performance of pretrained detectors

Detector	**Category**	$\mathcal{P}$ %	$\mathcal{R}$ %	$\mathcal{FM}$ %	$\mathcal{A}$ %	$\mathcal{BA}$ %
Selim [37]	IL	48.42	86.18	62.00	44.68	45.05
	IF	50.24	90.25	64.55	48.11	48.07
	IR	49.50	84.67	62.48	46.66	47.44
	OL	49.35	91.22	64.05	47.85	48.55
	OF	48.84	91.45	63.68	47.48	47.69
	OR	49.89	94.52	65.31	47.62	47.26
	Overall	**49.37**	**89.72**	**63.68**	**47.07**	**47.34**
CNN-Detector [43]	IL	50.86	98.41	67.06	51.00	50.18
	IF	51.75	95.07	67.02	52.52	51.46
	IR	50.03	94.81	65.50	49.14	48.42
	OL	49.65	98.57	66.04	49.21	49.28
	OF	49.68	98.99	66.16	49.35	49.49
	OR	51.32	97.24	67.18	51.48	50.58
	Overall	**50.55**	**97.18**	**66.49**	**50.45**	**49.90**
ViT [46]	IL	53.69	17.62	26.53	21.87	35.64
	IF	55.63	12.70	20.68	22.65	37.34
	IR	56.54	11.62	19.28	25.15	37.07
	OL	33.83	12.02	17.74	29.67	36.50
	OF	49.10	12.24	19.60	23.80	34.50
	OR	55.92	11.30	18.80	21.71	35.00
	Overall	**50.79**	**12.92**	**20.44**	**24.14**	**36.01**
NPR [40]	IL	52.36	50.70	51.52	51.83	47.69
	IF	50.47	48.79	49.62	49.68	45.79
	IR	53.64	51.43	52.51	52.53	48.50
	OL	52.62	47.62	50.00	49.11	46.65
	OF	52.88	48.26	50.46	49.63	46.20
	OR	55.92	50.84	53.26	53.28	49.30
	Overall	**52.98**	**49.61**	**51.23**	**51.01**	**47.35**

V. Experiments and Discussions

This section presents and analyzes the experimental results. The problem was formulated as a binary classification task from the perspective of detectors to distinguish between genuine and fake images. We conducted benchmarking using both pretrained detectors and those fine-tuned on our IndicSideFace dataset via transfer learning.

A. Evaluation Metrics

We evaluated performance using the following metrics: Precision ($\mathcal{P}$), Recall ($\mathcal{R}$), F-measure ($\mathcal{FM}$), Accuracy ($\mathcal{A}$), and Balanced Accuracy ($\mathcal{BA}$). Our IndicSideFace dataset poses a challenge due to class imbalance, containing 984 genuine and 21648 fake samples. In such scenarios, $\mathcal{BA}$ serves as a more

TABLE V: Performance ($\mathcal{BA}$ %) of transfer learning-based detectors on generator-specific fake image groups paired with genuine images

		Identity Swapping					Attribute Manipulation								
Detector	**Category**	SimSwap [7]	SimSwap++ [8]	Ghost [15]	FaceDancer [34]	InsightFace [18]	Age	Beard	Expression	Gender	Glasses	Hair color	Hair style	Skin tone	**Mean**
Selim [37]	IL	87.48	94.13	74.94	77.46	82.77	89.42	82.33	85.32	89.93	87.35	84.21	83.55	90.09	85.31
	IF	95.65	89.01	78.11	76.08	91.29	89.01	80.41	81.82	85.61	86.91	79.16	74.90	82.99	83.92
	IR	88.79	95.15	77.52	79.48	90.50	84.59	79.41	83.96	84.70	87.04	82.62	85.47	86.26	85.04
	OL	89.00	89.60	72.78	80.15	89.84	89.41	85.83	85.71	85.74	88.12	83.99	77.18	88.40	85.06
	OF	93.43	90.04	81.54	76.56	88.31	87.65	85.13	80.03	86.97	89.21	82.86	80.60	84.56	85.15
	OR	93.22	93.21	71.21	75.39	86.55	88.13	77.68	86.49	83.33	85.71	87.07	82.21	85.03	84.25
	Overall	**91.26**	**91.86**	**76.02**	**77.52**	**88.21**	**88.04**	**81.80**	**83.89**	**86.05**	**87.39**	**83.32**	**80.65**	**86.22**	**84.79**
CNN-Detector [43]	IL	98.69	92.16	70.59	56.51	86.76	91.88	84.99	75.82	90.31	86.80	95.79	94.81	87.67	85.60
	IF	96.73	93.14	68.38	81.37	89.71	84.65	84.16	81.62	83.46	84.38	90.00	87.47	69.65	84.21
	IR	97.63	92.52	74.36	65.70	89.39	89.53	80.54	80.87	86.94	84.80	98.98	87.29	90.92	86.11
	OL	88.54	93.46	64.68	56.20	85.29	93.75	89.97	72.75	85.66	85.48	92.37	91.35	82.32	83.22
	OF	97.06	93.14	68.87	71.90	87.25	85.54	80.32	85.06	91.50	89.71	93.90	90.41	84.75	86.11
	OR	97.00	93.14	69.02	56.51	89.71	91.70	80.93	74.38	91.53	77.63	96.85	90.20	90.91	84.58
	Overall	**95.94**	**92.93**	**69.32**	**64.70**	**88.02**	**89.51**	**83.49**	**78.42**	**88.23**	**84.80**	**94.65**	**90.26**	**84.37**	**84.97**
ViT [46]	IL	87.54	78.38	94.20	95.54	92.88	75.07	87.65	93.37	86.73	75.12	95.77	68.86	84.82	85.84
	IF	78.47	82.86	93.70	77.56	84.48	70.47	83.63	81.64	81.24	68.82	87.39	61.77	80.60	79.43
	IR	90.76	80.32	92.53	90.79	91.37	70.34	82.66	91.97	85.39	79.56	91.32	88.34	82.99	86.03
	OL	92.27	74.50	94.10	93.90	89.97	71.33	84.70	97.15	94.27	90.40	94.17	84.83	83.90	88.11
	OF	87.42	86.91	93.19	88.10	87.84	82.29	89.14	92.52	86.48	87.51	93.19	84.12	81.38	87.70
	OR	92.97	77.32	93.40	93.92	93.67	60.00	87.16	96.27	67.16	91.79	90.20	83.58	80.27	85.21
	Overall	**88.24**	**80.05**	**93.52**	**89.97**	**90.04**	**71.58**	**85.82**	**92.15**	**83.55**	**82.20**	**92.01**	**78.58**	**82.33**	**85.39**
NPR [40]	IL	50.01	50.03	50.01	50.02	50.48	51.29	54.48	50.96	51.29	50.96	64.66	51.94	50.96	52.08
	IF	50.00	50.01	50.02	50.00	50.64	50.68	53.29	50.37	50.68	50.65	66.72	51.40	50.36	51.91
	IR	50.00	50.04	50.00	50.01	49.86	50.65	51.74	50.35	50.35	50.33	57.89	52.30	50.33	51.07
	OL	49.80	50.22	50.11	50.23	49.92	50.34	54.26	50.65	51.63	50.32	56.13	51.01	50.32	51.15
	OF	50.00	50.55	50.44	50.44	52.03	52.15	55.04	52.43	53.73	51.42	55.74	53.64	52.45	52.31
	OR	50.01	50.34	50.66	49.85	49.46	52.07	53.42	50.47	51.16	50.40	59.52	52.36	51.08	51.60
	Overall	**49.97**	**50.20**	**50.21**	**50.09**	**50.40**	**51.20**	**53.70**	**50.87**	**51.47**	**50.68**	**60.11**	**52.11**	**50.92**	**51.69**

suitable evaluation metric, as it averages the true positive rate and false positive rate, ensuring a balanced assessment of model performance [3].

B. Baseline Performance of Pretrained Detectors

To ensure a comprehensive evaluation, we begin by assessing the baseline performance of off-the-shelf pretrained detectors, including Selim [37], NPR-based detector [40], CNN-detector [43], and ViT-based detector [46]. These detectors were pretrained on diverse datasets: Selim [37] on DFDC dataset [10], ViT [46] on OpenForensics [24], and CNN-detector [43] and NPR [40] on ForenSynths dataset[1].

In our study, we employed a zero-shot evaluation strategy for pretrained detectors, meaning that no IndicSideFace data was used during training. The entire IndicSideFace dataset served as the test set, and the results are summarized in Table IV. This table presents the performance of the detectors across six categories: indoor_left (IL), indoor_frontal (IF), indoor_right (IR), outdoor_left (OL), outdoor_frontal (OF), and outdoor_right (OR). Overall, Selim [37], CNN-detector [43], ViT [46], and NPR [40] achieved 47.34%, 49.90%, 36.01%, and 47.35% $\mathcal{BA}$, respectively.

We evaluated the detector performances on separate groups of fake images generated by each of the employed generators [7], [8], [13], [15], [18], [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 Selim [37] on a test dataset comprising all genuine IL images and SimSwap [7]-generated fake IL images, a $\mathcal{BA}$ of 37.58% was achieved.

[1]https://github.com/chuangchuangtan/NPR-DeepfakeDetection, https://github.com/peterwang512/CNNDetection, Accessed: 2025-05-10

TABLE VI: Performance of detectors with transfer learning

Detector	**Category**	$\mathcal{P}$ %	$\mathcal{R}$ %	$\mathcal{FM}$ %	$\mathcal{A}$ %	$\mathcal{BA}$ %
Selim [37]	IL	84.17	86.57	85.35	85.28	85.31
	IF	84.39	84.30	84.31	83.94	83.92
	IR	84.65	84.80	84.68	85.06	85.04
	OL	86.56	84.87	85.63	84.98	85.06
	OF	84.01	85.76	84.85	85.13	85.15
	OR	84.48	84.87	84.61	84.32	84.25
	Overall	**84.71**	**85.19**	**84.90**	**84.79**	**84.79**
CNN-Detector [43]	IL	88.76	88.54	88.65	89.39	85.60
	IF	85.89	83.65	84.76	87.26	84.21
	IR	86.83	88.40	87.61	88.57	86.11
	OL	85.50	87.70	86.59	87.54	83.22
	OF	87.44	87.60	87.52	89.01	86.11
	OR	85.93	88.31	87.10	87.48	84.58
	Overall	**86.73**	**87.37**	**87.04**	**88.21**	**84.97**
ViT [46]	IL	89.26	86.35	87.78	86.74	85.84
	IF	84.17	82.88	83.52	81.52	79.43
	IR	88.70	87.32	88.00	87.09	86.03
	OL	91.65	88.81	90.21	89.01	88.11
	OF	89.08	91.13	90.09	88.96	87.70
	OR	87.14	87.13	87.14	86.01	85.21
	Overall	**88.33**	**87.27**	**87.79**	**86.56**	**85.39**
NPR [40]	IL	61.57	49.90	55.12	63.32	52.08
	IF	61.78	50.40	55.51	59.77	51.91
	IR	60.95	49.03	54.35	57.95	51.07
	OL	60.26	50.44	54.91	57.66	51.15
	OF	62.10	53.40	57.43	61.51	52.31
	OR	61.60	51.61	56.17	59.85	51.60
	Overall	**61.38**	**50.80**	**55.58**	**60.01**	**51.69**

C. Performance of Transfer Learning-based Detectors

We explored the impact of transfer learning by retraining the detectors, adapting them to the specific characteristics of our IndicSideFace dataset. For Selim [37], CNN-detector [43], and NPR [40], we unfroze the last three layers and applied transfer learning for retraining. In the case of ViT [46], the entire MLP head was retrained. We utilized 60% of the IndicSideFace dataset for retraining and the remaining 40% for testing. The results presented here are based on this test set. From Table VI, we can observe that Selim [37], CNN-detector [43], ViT [46], and NPR [40] obtained overall 84.79%, 84.97%, 85.39%, and 51.69% $\mathcal{BA}$, respectively.

We also assessed the transfer learning-based detector performances on generator-specific fake image groups paired with genuine images. For retraining the detectors, we utilized 60% of the generator-specific fake image groups along with 60% genuine images of IndicSideFace. The remaining 40% of the generator-specific fake images and 40% of genuine images were used for testing, and the results are detailed in Table V expressed in terms of $\mathcal{BA}$. For instance, this table shows that applying the above transfer learning-based Selim [37] to a test dataset comprising 40% genuine IL images and SimSwap [7]-generated fake IL images yielded 87.48% $\mathcal{BA}$.

D. Cross Dataset Evaluation

To assess the generalization capability of the detectors, we conducted cross-dataset evaluations on IndicSideFace using two experimental setups: *ES-C1*, and *ES-C2*. We first partitioned genuine images of 164 subjects of IndicSideFace into two equal distinct groups: group-1 and group-2, each consisted of 82 $(= 164/2)$ subjects, contributing a total of 492 $(= 82 \times 6)$ genuine images.

• *ES-C1:* The detectors were retrained using 13776 fake images of IndicSideFace dataset, generated via identity swapping, and 492 genuine images from above group-1. They were then tested on 7872 attribute-manipulated fake images from IndicSideFace, and 492 genuine images from group-2.

• *ES-C2:* This setup swapped the training and testing sets of *ES-C1*. Specifically, the detectors were retrained using 7872 fake images created through attribute manipulation, and tested on 13776 fake images generated via identity swapping. Additionally, the training set contained 492 genuine images from group-2, while the test set included 492 genuine images from group-1.

In *ES-C1* and *ES-C2* also, Selim [37], CNN-detector [43], and NPR [40] engaged transfer learning by unfreezing the last three layers of each model for retraining, whereas ViT [46] retrained the entire MLP head.

The results, summarized in Table VII, highlight the generalization performance of the detectors on test sets of *ES-C1* and *ES-C2*. Among all detectors engaged here, ViT [46] achieved the highest $\mathcal{BA}$, while Selim [37] recorded the lowest.

TABLE VII: Performance of detectors on cross dataset

Experimental Setup	Detector	$\mathcal{P}$ %	$\mathcal{R}$ %	$\mathcal{FM}$ %	$\mathcal{A}$ %	$\mathcal{BA}$ %
ES-C1	Selim [37]	88.82	96.22	92.37	85.87	49.59
	CNN-Detector [43]	90.70	96.55	93.53	88.13	58.64
	ViT [46]	91.56	99.99	95.59	91.76	61.60
	NPR [40]	89.13	84.39	86.69	86.67	50.43
ES-C2	Selim [37]	93.05	82.53	87.48	77.94	48.17
	CNN-Detector [43]	94.29	97.58	95.90	92.22	57.50
	ViT [46]	85.78	81.23	83.45	83.25	83.30
	NPR [40]	89.29	84.24	86.69	86.69	50.00

E. Observations

We observed that pretrained deepfake detectors struggled to identify manipulations in side-face images, as shown in Tables III, IV. The $\mathcal{BA}$ in this scenario ranged from only 36.01% to 49.90%, indicating a significant performance drop compared to only its near-frontal counterparts. This suggests that existing models, primarily trained on frontal views, are not adequately equipped to generalize across diverse head poses. However, when these detectors were fine-tuned using transfer learning on the IndicSideFace dataset, we observed a substantial performance improvement (refer to Tables V, VI). The $\mathcal{BA}$ increased to a range of 51.69% to 85.39%, demonstrating the value of domain-specific retraining. This highlights the importance of pose-diverse training data in building robust deepfake detectors suitable for real-world applications.

VI. Conclusion

Existing deepfake detection research has predominantly focused on frontal-face perspectives, leaving side-face manipulations largely unexplored. In this work, we introduced IndicSideFace, a novel dataset explicitly designed to advance deepfake detection for side-face perspectives of Indian subjects, incorporating diverse lighting conditions (indoor and outdoor). IndicSideFace consists of 984 genuine images and 21648 fake images generated using five identity-swapping tools and one attribute manipulation tool. Our benchmarking experiments with state-of-the-art deepfake detection models revealed a critical limitation: existing approaches struggle to detect side-face deepfakes effectively, particularly under varying lighting conditions. These results emphasize the necessity for fine-tuning or developing specialized models to improve detection performance on Indian side-face deepfakes. This initial study highlights an urgent need for dedicated research in this domain. Future work will focus on expanding the dataset with additional manipulation techniques, increasing its scale, and developing more robust deepfake detection methodologies tailored for side-face perspectives.

ACKNOWLEDGMENTS

The authors sincerely thank all the volunteers, and interns/ trainees/ researchers, who helped in database generation. Special thanks are extended to Ashutosh Parihar (IIT Indore), Anand Suralkar (IIT Indore), and Vishal Kumar (IIT Kanpur). Partial support from IHUB-NTIHAC Foundation, IIT Kanpur, under proposal no. 3092 is gratefully acknowledged.

ETHICAL IMPACT STATEMENT

1. Did you read the Ethical Impact Statement Guidelines document? Yes
2. Is it clear that all studies and procedures described in the paper were approved (or exempted) by a valid ethical review board? Alternatively, is a valid and sufficient justification provided for why the oversight of an ethical review board was not required? Yes
3. Does the ethical impact statement provide a clear, complete, and balanced discussion of the potential risks of individual harm and negative societal impacts associated with the research? Note that this includes harm to research participants as well as harm to other

individuals that may be affected by the use, misuse, or misunderstanding of the research. Yes

4. Does the ethical impact statement describe reasonable, valid, and sufficient use of risk-mitigation strategies by the authors to lessen these potential risks? Alternatively, if relevant strategies were not used, is a valid and sufficient justification for this provided? Yes
5. Does the ethical impact statement provide a valid and sufficient justification for how/why the potential risks of the research are outweighed by the risk-mitigation strategies and potential benefits of the research? Note that papers with serious potential risks that are not outweighed by risk-mitigation strategies and potential benefits may be rejected. Yes
6. If the paper involves human subjects, are all of the following sub-boxes checked?
   - 6a. Does the main paper describe whether/how informed consent and/or assent were obtained from participants? If consent and/or assent were fully or partially obtained, were the methods used to do so valid? If not fully obtained, does the ethical impact statement provide a valid and sufficient justification for this? Yes
   - 6b. Does the main paper state whether the participants explicitly consented to the use of their data in the manner described in the paper? For example, if the data was or will be shared with third parties, does it state that the participants explicitly agreed to this sharing? If some uses were not explicitly consented to, does the ethical impact statement provide a valid and sufficient justification for this? Yes
   - 6c. Does the main paper explain whether/how participants were compensated? If participants were compensated, does the ethical impact statement provide a valid and sufficient justification for the form and amount of compensation provided? Yes
   - 6d. If the research involves any special or vulnerable populations (e.g., minors, elderly individuals, prisoners, refugees, and migrants, individuals with disabilities, individuals with mental illness, or patients in medical settings), does the ethical impact statement provides a valid and sufficient explanation of how the rights, well-being and autonomy of such individuals were safeguarded in the research? N.A.

REFERENCES

[1] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen. MesoNet: a compact facial video forgery detection network. In *WIFS*, pages 1–7, 2018.

[2] Z. Akhtar, T. L. Pendyala, and V. S. Athmakuri. Video and audio deepfake datasets and open issues in deepfake technology: being ahead of the curve. *Forensic Sciences*, 4(3):289–377, 2024.

[3] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *ICPR*, pages 3121–3124, 2010.

[4] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In *ICCV*, pages 1021–1030, 2017.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *FG*, pages 67–74. IEEE, 2018.

[6] A. Chandra, A. A, S. Das, and A. Das. Latent flow diffusion for deepfake video generation. In *CVPR Workshops*, pages 3781–3790, 2024.

[7] R. Chen, X. Chen, B. Ni, and Y. Ge. SimSwap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2003–2011, 2020.

[8] X. Chen, B. Ni, Y. Liu, N. Liu, Z. Zeng, and H. Wang. SimSwap++: Towards faster and high-quality identity swapping. *IEEE TPAMI*, 2023.

[9] U. A. Ciftci, I. Demir, and L. Yin. FakeCatcher: Detection of synthetic portrait videos using biological signals. *IEEE TPAMI*, 2020.

[10] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer. The DeepFake Detection Challenge (DFDC) Dataset. `https://www.kaggle.com/c/deepfake-detection-challenge`. *arXiv:2006.07397*, 2020. Accessed: 2025-05-10.

[11] A. Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

[12] R. Durall, M. Keuper, and J. Keuper. Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In *CVPR*, pages 7890–7899, 2020.

[13] FaceApp Technology Ltd. FaceApp. `https://www.faceapp.com`. Accessed: 2025-05-10.

[14] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz. Leveraging frequency analysis for deep fake image recognition. In *ICML*, pages 3247–3258, 2020.

[15] A. Groshev, A. Maltseva, D. Chesakov, A. Kuznetsov, and D. Dimitrov. Ghost—a new face swap approach for image and video domains. *IEEE Access*, 10:83452–83462, 2022.

[16] L. Guarnera, O. Giudice, and S. Battiato. Mastering deepfake detection: A cutting-edge approach to distinguish gan and diffusion-model images. *ACM TOMM*, 20(11):1–24, 2024.

[17] D. Güera and E. J. Delp. Deepfake video detection using recurrent neural networks. In *AVSS*, pages 1–6. IEEE, 2018.

[18] J. Guo, J. Deng, et al. InsightFace: 2D and 3D Face Analysis Project. `https://github.com/deepinsight/insightface`, `https://www.picsi.ai`. Accessed: 2025-05-10.

[19] J. Hao, Z. Zhang, S. Yang, D. Xie, and S. Pu. TransForensics: image forgery localization with dense self-attention. In *ICCV*, pages 15055–15064, 2021.

[20] L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy. DeeperForensics-1.0: A large-scale dataset for real-world face forgery detection. In *CVPR*, pages 2889–2898, 2020.

[21] S. Jin, Z. Wang, L. Wang, P. Liu, N. Bi, and T. Nguyen. AUEditNet: Dual-Branch Facial Action Unit Intensity Manipulation with Implicit Disentanglement. In *CVPR*, pages 2104–2113, 2024.

[22] K. Kuckreja, X. Hoque, N. Poddar, S. Reddy, A. Dhall, and A. Das. INDIFACE: Illuminating India's Deepfake Landscape with a Comprehensive Synthetic Dataset. In *FG*, pages 1–9. IEEE, 2024.

[23] P. Kwon, J. You, G. Nam, S. Park, and G. Chae. KoDF: A large-scale Korean deepfake detection dataset. In *ICCV*, pages 10744–10753, 2021.

[24] T.-N. Le, H. H. Nguyen, J. Yamagishi, and I. Echizen. OpenForensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild. In *ICCV*, pages 10117–10127, 2021.

[25] G. Li, X. Zhao, and Y. Cao. Forensic symmetry for deepfakes. *IEEE TIFS*, 18:1095–1110, 2023.

[26] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen. Advancing high fidelity identity swapping for forgery detection. In *CVPR*, pages 5074–5083, 2020.

[27] Y. Li, M.-C. Chang, and S. Lyu. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In *WIFS*, pages 1–7, 2018.

[28] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu. Celeb-DF: A large-scale challenging dataset for deepfake forensics. In *CVPR*, pages 3207–3216, 2020.

[29] H. Lin. Dreamsalon: A staged diffusion framework for preserving identity-context in editable face generation. In *CVPR*, pages 8589–8598, 2024.

[30] K. Liu et al. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *Pattern Recognition*, 141:109628, 2023.

[31] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha. Emotions don't lie: An audio-visual deepfake detection method using affective cues. In *ACM MM*, pages 2823–2832, 2020.

[32] K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, and R. Singh. DF-Platter: Multi-face heterogeneous deepfake dataset. In *CVPR*, pages 9739–9748, 2023.

[33] Y. Nirkin, Y. Keller, and T. Hassner. FSGAN: Subject Agnostic Face

Swapping and Reenactment. In *ICCV*, pages 7184–7193, 2019.
[34] F. Rosberg, E. E. Aksoy, F. Alonso-Fernandez, and C. Englund. FaceDancer: Pose-and occlusion-aware high fidelity face swapping. In *WACV*, pages 3454–3463, 2023.
[35] A. Rössler et al. Faceforensics++: Learning to detect manipulated facial images. In *ICCV*, pages 1–11, 2019.
[36] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan. Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI)*, 3(1):80–87, 2019.
[37] S. Seferbekov. DeepFake Detection (DFDC) Solution. `https://github.com/selimsef/dfdc_deepfake_challenge`, 2020. Accessed: 2025-05-10.
[38] J. W. Seow, M. K. Lim, R. C. Phan, and J. K. Liu. A comprehensive overview of deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, 513:351–371, 2022.
[39] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe. First order motion model for image animation. *NeurIPS*, 32, 2019.
[40] C. Tan, Y. Zhao, S. Wei, G. Gu, P. Liu, and Y. Wei. Rethinking the up-sampling operations in CNN-based generative network for generalizable deepfake detection. In *CVPR*, pages 28130–28139, 2024.
[41] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner. Face2Face: Real-time face capture and reenactment of rgb videos. In *CVPR*, pages 2387–2395, 2016.
[42] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. Deferred neural rendering: Image synthesis using neural textures. In *ACM TOG*, volume 38, pages 1–12, 2019.
[43] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros. CNN-generated images are surprisingly easy to spot... for now. In *CVPR*, pages 8695–8704, 2020.
[44] X. Wei, Z. Xu, C. Liu, S. Wu, Z. Yu, and H. S. Wong. Text-guided unsupervised latent transformation for multi-attribute image manipulation. In *CVPR*, pages 19285–19294, 2023.
[45] P.-W. Wu, Y.-J. Lin, C.-H. Chang, E. Y. Chang, and S.-W. Liao. Relgan: Multi-domain image-to-image translation via relative attributes. In *ICCV*, pages 5914–5922, 2019.
[46] Wvolf. ViT Deepfake Detection. `https://huggingface.co/Wvolf/ViT_Deepfake_Detection`. Accessed: 2025-05-10.
[47] X. Yang, Y. Li, and S. Lyu. Exposing deep fakes using inconsistent head poses. In *ICASSP*, pages 8261–8265. IEEE, 2019.
[48] E. Zakharov, A. Shysheya, E. Burkov, and V. Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In *ICCV*, pages 9459–9468, 2019.
[49] H. Zhou et al. Pose-controllable talking face generation by implicitly modularized audio-visual representation. In *CVPR*, pages 4176–4186, 2021.