MFFI: Multi-Dimensional Face Forgery Image Dataset for Real-World Scenarios

Changtao Miao*
Yi Zhang*
Man Luo*
Ant Group
Hangzhou, China

Weiwei Feng
Kaiyuan Zheng
Ant Group
Hangzhou, China

Qi Chu
Tao Gong
Anhui Province Key Laboratory of Digital Security
Hefei, China

Jianshu Li†
Ant Group
Hangzhou, China

Yunfeng Diao‡
Hefei University of Technology
Hefei, China

Wei Zhou
Cardiff University
Cardiff, UK

Joey Tianyi Zhou
IHPC, Agency for Science, Technology and Research
CFAR, Agency for Science, Technology and Research
Singapore

Xiaoshuai Hao‡
Beijing Academy of Artificial Intelligence
Beijing, China

Abstract

Rapid advances in Artificial Intelligence Generated Content (AIGC) have enabled increasingly sophisticated face forgeries, posing a significant threat to social security. However, current Deepfake detection methods are limited by constraints in existing datasets, which lack the diversity necessary in real-world scenarios. Specifically, these data sets fall short in four key areas: unknown of advanced forgery techniques, variability of facial scenes, richness of real data, and degradation of real-world propagation. To address these challenges, we propose the Multi-dimensional Face Forgery Image (**MFFI**) dataset, tailored for real-world scenarios. MFFI enhances realism based on four strategic dimensions: 1) Wider Forgery Methods; 2) Varied Facial Scenes; 3) Diversified Authentic Data; 4) Multi-level Degradation Operations. MFFI integrates 50 different forgery methods and contains 1024*K* image samples. Benchmark evaluations show that MFFI outperforms existing public datasets in terms of scene complexity, cross-domain generalization capability, and detection difficulty gradients. These results validate the technical advance and practical utility of MFFI in simulating real-world conditions. The dataset and additional details are publicly available at https://github.com/inclusionConf/MFFI.

CCS Concepts

• **Computing methodologies** → **Computer vision**; *Biometrics.*

*These authors contributed equally to this paper.
†Project leader. Email:jianshu.l@antgroup.com
‡Corresponding authors. Email:diaoyunfeng@hfut.edu.cn; xshao@baai.ac.cn

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

MM '25, Dublin, Ireland

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758280

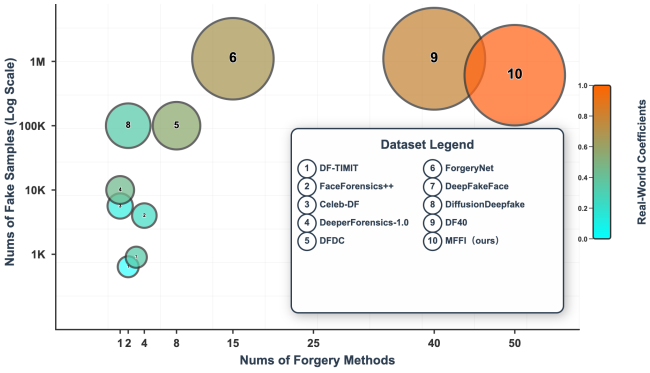

Figure 1: Comparison with other dataset in diversity and real-world realism. The *Real-World Coefficients* denote the degree of similarity between datasets and real-world scenarios.

Keywords

Deepfake, Face Forgery Detection, Real-World Scenarios

ACM Reference Format:
Changtao Miao, Yi Zhang, Man Luo, Weiwei Feng, Kaiyuan Zheng, Qi Chu, Tao Gong, Jianshu Li, Yunfeng Diao, Wei Zhou, Joey Tianyi Zhou, and Xiaoshuai Hao. 2025. MFFI: Multi-Dimensional Face Forgery Image Dataset for Real-World Scenarios. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3746027.3758280

1 Introduction

As we accelerate towards the era of Artificial General Intelligence, its core technology, Artificial Intelligence Generated Content (AIGC), demonstrates remarkable capabilities in producing hyper-realistic content, such as highly lifelike text, images, and audio-visual content [5]. While AIGC expands creative possibilities, it simultaneously presents unprecedented challenges to societal governance

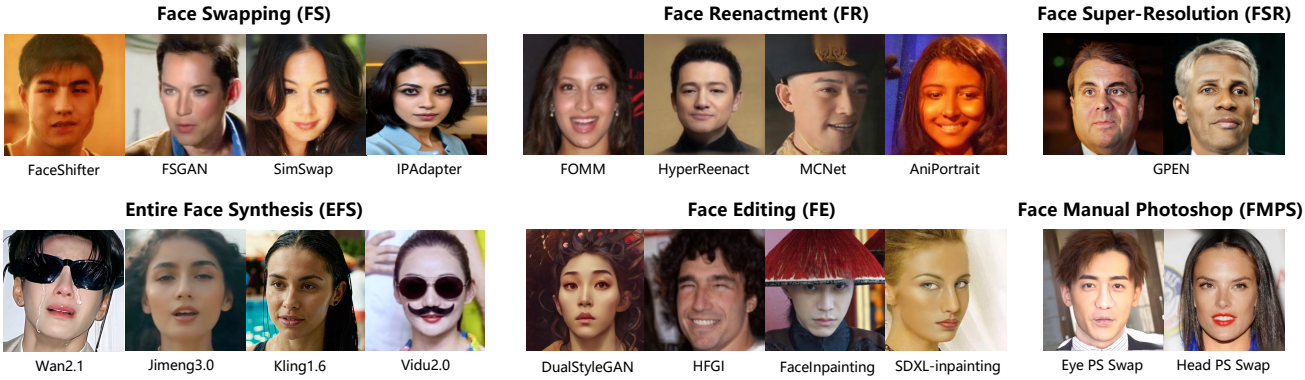

Figure 2: Examples of generated images from 6 major face forgery categories in our MFFI dataset. The categories include Face Swapping (FS, 15 types), Face Reenactment (FR, 8 types), Entire Face Synthesis (EFS, 13 types), Face Editing (FE, 6 types), Face Super-Resolution (FSR, 1 type), and Face Manual Photoshop (FMPS, 7 types).

structures. Central among these threats are deepfake-based face forgery attacks [35, 56, 76], where identity features can be precisely replicated, facial microexpressions altered, or social trust mechanisms systematically eroded. Such capabilities provide technical enablers for malicious activities such as financial fraud and political subversion. In response, face forgery detection technologies [44–47, 62, 92, 93] have emerged as a critical research domain to safeguard the integrity of digital societal infrastructures.

However, current deepfake detection methods [30, 39] exhibit inadequate generalization ability to real-world scenarios due to limitations in training data quality and diversity. This shortfall arises primarily from the multi-dimensional complexity of forged samples in authentic environments: (1) **Unknown of Advanced Forgery Techniques**: attackers frequently use emerging forgery techniques that are unseen during training [48]; (2) **Variability of Facial Scenes**: facial images often appear under varying environmental conditions and composition styles, which hinder the extraction of invariant features [85]; (3) **Richness of Real Data**: fake faces in real-world may originate from highly divergent source distribution [13], leading to domain generalization gaps for models trained on narrow datasets; (4) **Degradation of Real-world Propagation**: transmission over social platforms often degrade input quality via adversarial perturbations, color distortions, and compression-induced blurring [11, 12, 83, 89]. While existing Deepfake datasets address subsets of these challenges, a unified benchmark integrating all four core dimensions remains unexplored. For example, FF++ [56] and Celeb-DF [35] are constrained by the singularity of generation techniques and the limitations of the original data. DFDC [13] relies on 8 traditional manipulation methods, lacking coverage of advanced techniques like diffusion models [19]. The latest DF40 [49] increases the generation diversity to 40 methods but reuses real videos from FF++ and Celeb-DF, limiting the expansion of the source domain. Moreover, these datasets largely overlook real-world transmission artifacts such as color shifts, compression, blurring, and adversarial perturbations. Therefore, developing a comprehensive face forgery dataset that integrates all four core dimensions is both urgent and essential to effectively address real-world challenges.

To address these challenges, we introduce the Multi-dimensional Face Forgery Image (**MFFI**) dataset, which designed to enhance the adaptability of face forgery detection models in real-world scenarios. MFFI systematically covers all the four complex forgery dimensions that may exist in realistic settings, with core innovations in the following dimensions:

Wider Forgery Methods: Beyond traditional Deepfake techniques, we innovatively incorporate image super-resolution and manual Photoshop. In particular, we integrate commercial tools just released in 2025 (e.g., Wan2.1 [67], Vidu2.0 [66], Kling1.6 [29], and Jimeng2.0 [23]) to ensure comprehensive coverage of the latest techniques. Overall, MFFI incorporates up to 50 diverse forgery methods, surpassing current datasets by a big margin.

Varied Facial Scenes: This dataset comprehensively covers different skin tones, age groups, pose angles, and real shooting scenarios including occlusions, complex lighting, and diverse backgrounds, fully simulating the complexity of facial images in the real world.

Diversified Authentic Data: We integrate real facial data from four distinct sources, breaking away from the single-source collection mode of existing datasets to ensure diversity and authenticity in the foundational data.

Multi-level Degradation Operations: To simulate potential degradation processes during data propagation, we randomly embed multi-level interference factors including blur processing, noise interference, color distortion, geometric deformation, sharpening artifacts, compression distortion, and adversarial perturbations.

A detailed comparison with current face forgery datasets are shown in Table 1. To our best knowledge, MFFI is the first face forgery dataset to uniformly consider all the four strategic dimensions (WFM, VFS, DAD, and MDO) in the real-world scenarios. Through this multi-dimensional data construction strategy, we achieve a dataset encompassing 50 diverse forgery techniques and $1024K$ samples (including both real and forged faces) to simulate forged faces in real-world environments. To validate the practical value of our dataset, we conduct extensive generalization and robustness experiments to analyze the performance of existing detection methods.

Table 1: Comparison of previous face forgery datasets. MFFI surpasses any other dataset in Wider Forgery Methods (WFM), Varied Facial Scenes (VFS), Diversified Authentic Data (DAD), and Multi-level Degradation Operations (MDO). The *Latest Face Foregry* represents the latest forgery method in this dataset. The *Fake Samples* represents the number of forged samples. Each forged video or image is considered as one sample.

Dataset	Publication	Latest Face Forgery	Wider Forgery Methods (WFM)								VFS	DAD	MDO
			Methods	FS	FR	EFS	FE	FSR	FMPS	Fake Smaples			
DF-TIMIT [31]	ArXiv'18	faceswap-GAN [57] (2018)	2	2	-	-	-	-	-	640	-	-	-
FaceForensics++ [56]	ICCV'19	NeuralTextures (2019)	4	2	2	-	-	-	-	4,000	-	-	-
Celeb-DF [35]	CVPR'20	Unknown	1	1	-	-	-	-	-	5,639	-	-	-
DeeperForensics-1.0 [22]	CVPR'20	DF-VAE [22] (2020)	1	1	-	-	-	-	-	10K	-	-	✓
DFDC [13]	ArXiv'20	StyleGAN [26] (2018)	8	5	1	2	-	-	-	0.1M+	-	✓	-
ForgeryNet [18]	CVPR'21	StarGANv2 [27] (2020)	15	6	4	2	3	-	-	1.1M+	-	-	-
FakeAVCeleb [28]	NeurIPS'21	Wav2Lip [53] (2021)	4	2	2	-	-	-	-	9,500	-	-	-
DF3 [24]	TMM'22	StyleGAN3 [25] (2021)	6	-	-	6	-	-	-	15k+	-	-	-
DeepFakeFace [60]	ArXiv'23	Stable-Diffusion [55] (2021)	3	1	-	2	-	-	-	90K	-	-	-
DiffusionDeepfake [3]	ArXiv'24	Stable-Diffusion [55] (2021)	2	-	-	2	-	-	-	0.1M+	-	-	-
DF40 [76]	NeurIPS' 24	PixArt-α [6] (2024)	40	10	13	12	5	-	-	1.1M+	-	-	-
MFFI (ours)	-	**Jimeng3.0 [23] (2025)**	**50**	15	8	13	6	1	7	0.7M+	✓	✓	✓

By addressing the critical challenge face forgery in real-world scenarios, our MFFI dataset aims to drive advancements in the development of robust, real-world-oriented face forgery detection methodologies. To do this end, the MFFI dataset served as the core benchmark for the Global Multimedia Deepfake Detection Challenge [1] held in Kaggle [84], attracting 1500 teams globally.

2 Related Work

2.1 Face Forgery Datasets

Early datasets [35, 56] were constrained by scene singularity, using single-source domain data and limited face manipulation techniques. While subsequent datasets like DFDC [13] enhanced data diversity, they still used only a few generation techniques, failing to overcome technological coverage limitations. ForgeryNet [18] achieved large-scale dataset construction but neglected recent diffusion model-based generation paradigms. DeepFakeFace [60] and DiffusionDeepfake [3] datasets introduced cutting-edge diffusion model technologies but suffered from homogenization of forgery modes. DF40 [76] improved diversity by expanding generation techniques, but was limited by existing original data (i.e., FF++ [56] and Celeb-DF [35]) distribution constraints and overlooked real-world interference and degradation factors. To address these deficiencies, our MFFI dataset effectively resolves the aforementioned limitations by simulating real-world scenarios across four dimensions.

2.2 Face Forgery Methods

Existing face forgery research typically categorizes forgery methods into four types: Face Swapping (FS), Face Reenactment (FR), Entire Face Synthesis (EFS), and Face Editing (FE). FS and FR are the most common technologies [9, 15, 64, 65] in face forgery datasets. However, with the rapid development of generative technologies, these techniques often suffer from obsolescence and homogenization. To address this issue, our MFFI dataset incorporates 15 different FS methods and 8 diverse FR methods, including latest audio-driven technologies, such as AniPortrait [73], InstantID [70]. EFS techniques have matured significantly due to progress in diffusion models [3]. The MFFI incorporates the latest commercial text-to-video and image-to-video models to ensure technological frontier coverage, e.g., Wan2.1 [67] and Jimeng3.0 [23]. FE represents a more fine-grained form of forgery. The MFFI implements both traditional GAN-based [17] and the latest diffusion [19] model-based methods in this category. While these classifications encompass most mainstream face forgery methods, they overlook two common forms of facial manipulation: Face Super-Resolution (FSR) and Face Manual PhotoShop (FMPS). Therefore, MFFI supplements 1 FSR technique and 7 FMPS operations to enhance the dataset's comprehensiveness.

3 Dataset

3.1 Overview

The MFFI dataset addresses existing limitations through four dimensions, i.e., WFM, VFS, DAD, and MDO. First, WFM expands beyond conventional categories (FS, FR, EFS, FE) to include Face Super-Resolution (FSR) and Face Manual Photoshop (FMPS) operations. Second, VFS systematically incorporates skin tones, head poses, occlusions, backgrounds, age ranges, and illumination conditions to simulate real-world complexity. Third, DAD combines established benchmarks with newly collected celebrity face images. Fourth, MDO simulates internet transmission artifacts (compression, blurring) while incorporating adversarial attacks for environmental fidelity. To prioritize real-world complexity, MFFI focuses on image-level data provision. For video-based forgery methods, forged frames are analyzed through frame sampling. As shown in Table 2, MFFI includes Train, Valid, Test, and Test-D sets.

3.2 Wider Forgery Methods

We introduce and implement 50 diverse face forgery techniques to ensure the realization of the diversity objective. These varied forgery techniques are categorized into six main classes:

Face Swapping (FS): This technique replaces the target face with the source face while preserving the identity of the source

[1] https://www.kaggle.com/competitions/multi-ffdi

Figure 3: Fake Examples of the VFS dimension.

and the background of the target. We implement 15 different methods, including SimSwap [7], FaceShifter [33], FaceFusion [10], FSGAN [50], InfoSwap [16], Stable-Diffusion-1.5 [55], HiFiface [72], IPAdapter [81], MegaFS [91], MobileFaceSwap [75], FaceMerge [14], RAFSwap [74], e4s [34], AIM [32], I2G [87], and SBI [58].

Face Reenactment (FR): It modifies the motion or expression information of the original face using video, audio, or text-driven approaches. We implement 8 methods, including FOMM [86], MCNet [20], ArticulatedAnimation [59], DaGAN [21], HyperReenact [4], MonkeyNet [2], AniPortrait [73], and Wav2Lip [53].

Entire Face Synthesis (EFS): This category uses GANs or diffusion models to directly generate complete synthetic faces. We implement 13 techniques, including StyleGAN2 [26], StyleGAN3 [90], InstantID [70], Stable-Diffusion-XL [52], Hailuoai [63], Jimeng2.0 [23], Jimeng3.0 [23], Kling1.6 [29], Pika1.5 [43], Pixverse [51], Vidu2.0 [66], Vivago [61], and Wan2.1 [67].

Face Editing (FE): This class aims to modify specific attributes of a given facial image. We implement 6 techniques, including HFGI [71], pSp [54], FaceInpainting [40], DualStyleGAN [79], SD-Inpainting [55], and SD-XL-Inpainting [52].

Face Super-Resolution (FSR): It enhances image quality by recovering details from low-resolution fake face images. We implemented the commonly used GPEN [80] technique for this.

Face Manual Photoshop (FMPS): This category involves manual manipulation of local facial areas using Photoshop software. We implement 6 common operations, including Random Facial Component Swapping, Random Facial Component Excision, Full Face Swapping, Facial Artifact Injection, Random Region Erasure, and Background Replacement. We also apply 1 Image-Filtering operation (including Kuwahara, Bilateral, and Oil filters) to process forged faces and render them with artistic stylization, thereby simulating Photoshop-style manipulations of real-world scenarios.

All these forgery techniques were implemented based on corresponding research papers or common technical approaches, generating a rich set of forged facial images that comprehensively simulate diverse scenarios in real-world settings.

3.3 Varied Facial Scenes

To further enhance the facial scenes, we filtered the aforementioned authentic and forged face images across six dimensions, specifically: 1) **Ethnicity:** Covering Caucasian, African, Indian, and Asian populations. 2) **Age:** Spanning from 0 to 85 years old. 3) **Pose:** Including frontal, profile, downward, and upward views. 4) **Occlusion:** Considering scenarios with sunglasses, face masks, hoodies, and bangs. 5) **Background:** Distinguishing between indoor and outdoor scenes. 6) **Lighting:** Incorporating conditions of strong light, low light, and colored lighting.

As illustrated in Figure 3, this comprehensive approach to facial scene design captures the diversity and complexity of real-world facial data, providing robust support for the diversity and realism of the dataset.

Table 2: Composition of real/fake class and forgery types in the proposed MFFI dataset. The *Test-D* denotes the test set processed according to MDO.

Sample Type		Train	Valid	Test	Test-D	Total
Real	CelebA [38]	100K	-	-	-	100K
	RFW [42, 69]	-	60K	25K	25K	110K
	CASIA-WebFace [82]	-	-	25K	25K	50K
	Chinese Celeb	-	-	30K	30K	60K
Fake	FS	96K	53K	21K	21K	191K
	FR	90K	2K	18K	18K	128K
	EFS	9K	-	14K	14K	37K
	FE	16K	2K	15.2K	15.2K	48.4K
	FSR	198K	23K	4.8K	4.8K	230.6
	FMPS	15K	-	27K	27K	69K
Total		524K	140K	180K	180K	**1024K**

3.4 Diversified Authentic Data

In face forgery datasets, the diversity of authentic facial data is crucial. However, previous datasets often lacked in this aspect, potentially leading to biases in model training. To construct a dataset better aligned with real-world scenarios, we select three distinctive real-world face recognition and a self-collected celebrity datasets:

CelebA [38]: A collection of celebrity images from the internet, featuring 200K images across 10K identities, providing a rich sample size and diverse identities.

Racial Faces in the Wild (RFW) [69]: Meticulously curated to represent four racial groups - Caucasian, Indian, Asian, and African. Each racial subset comprises 10K images spanning 3K identities, ensuring balanced ethnic representation. To further mitigate racial bias, we incorporate the extended variants of the RFW dataset: the BUPT-GlobalFace and BUPT-BalancedFace datasets [42].

CASIA-WebFace [82]: A large-scale face dataset with about 500K facial images across 10K identities, offering enhanced sample quantity and identity coverage.

Chinese Celeb: To address the under-representation of Chinese internet facial data, we self-collect Chinese celebrity faces containing 30K images and about 1.5K identities, further enhancing geographical diversity.

3.5 Multi-level Degradation Operations

To simulate the degradation of the image during transmission, we randomly applied various disturbance operations, primarily categorized into conventional disturbances and deep adversarial attacks.

For conventional disturbances, we employ common perturbation techniques including uniform blur, Gaussian blur, uniform noise,

Table 3: Intra-dataset evaluation with the SOTAs on MFFI datasets.

Types	Models	Test				Test-D			
		ACC	AUC	EER	AP	ACC	AUC	EER	AP
Spatial Detector	Xception [8]	0.7683	0.8522	0.2320	0.8855	0.6470	0.7075	0.3550	0.7727
	RFM [68]	0.7588	0.8487	0.2336	0.8853	0.6560	0.7203	0.3441	0.7869
Frequency Detector	SRM [41]	0.7619	0.8765	0.2105	0.9146	0.5524	0.5801	0.4307	0.6590
	SPSL [37]	0.7492	0.8455	0.2396	0.8772	0.6305	0.7144	0.3486	0.7763

Gaussian noise, sharpening, compression, and geometric transformations. These operations are widely observed in data transformation processes across social networks on the Internet, effectively mimicking real-world scenarios.

In terms of deep adversarial attacks, to further enhance the realism of forgery scenarios and counter potential malicious attackers, we simulate adversarial attack strategies. Specifically, we generate perturbations using PatchAttack [78], a black-box attack method based on reinforcement learning. This approach induces model misclassifications by superimposing small textured patches onto images, thereby increasing the difficulty of face forgery detection.

To align with real-world evaluation requirements, we apply these degradation operations exclusively to the test set (i.e., Test-D). By incorporating both conventional disturbances and advanced adversarial attacks, our methodology provides a comprehensive framework for assessing the robustness of face forgery detection models under diverse and challenging conditions.

4 Evaluations

4.1 Experimental Setup

In this study, all preprocessing and training codebases strictly adhere to the experimental setup of DeepfakeBench [77]. We select the classic Xception [8] as our baseline model and incorporate three SOTA detectors: SRM [41], SPSL [37], and RFM [68], all of which utilize Xception as their backbone network. This configuration enables a comprehensive analysis of traditional detection models' performance on our proposed MFFI dataset, facilitating a comparison of detection efficacy between MFFI and other datasets. The algorithmic implementation details and training hyperparameters are consistent with the publicly available settings in DeepfakeBench [77], ensuring reproducibility and consistency of results.

4.2 Evaluation Protocols and Metrics

To comprehensively evaluate the face forgery detection methods on the proposed MFFI dataset, we establish three evaluation protocols: 1) Intra-dataset; 2) Cross-dataset; 3) Zero-shot detection of MLLM. Following the DeepfakeBench [77], we report ACC, AUC, EER, AP as metrics of the detection models.

4.3 Intra-dataset Evaluation

To evaluate the overall performance of MFFI, we designed intra-dataset evaluation experiments where models were trained on the MFFI training set and tested on both the standard test set (Test) and degradation-simulated test set (Test-D). As shown in Table 3, on the standard Test set, the baseline Xception [8] model achieved comparable performance to state-of-the-art detectors (RFM [68], SRM [41], SPSL [37]), particularly attaining the highest accuracy (ACC = 0.7683) among all methods. On the real-world degradation-embedded Test-D set, spatial-domain detectors (Xception [8], RFM [68]) demonstrated superior robustness compared to frequency-domain counterparts (SRM [41], SPSL [37]), with SRM [41] exhibiting the most significant ACC degradation (ΔACC = -0.2095). This indicates that the introduced degradation operations disproportionately disrupt frequency-domain features of the SRM model, highlighting the increased detection challenges of the MFFI dataset relative to existing benchmarks.

4.4 Cross-dataset Evaluation.

Cross-dataset testing serves as a standard evaluation protocol in face forgery detection to evaluate the generalization of the model on unseen datasets. As shown in Table 4, the SRM [41] and SPSL [37] models exhibit suboptimal performance in cross-dataset scenarios, particularly in the highly challenging Test-D dataset. In particular, we observe that SRM [41] trained FF++ (C23) [56] demonstrates superior generalizability, with trained counterparts in DF40-FS [76] achieving an improvement in accuracy of 0.0974. Although DF40-FS [76] encompasses more diverse forgery techniques compared to FF++ (C23) [56], it fails to deliver expected performance gains, potentially due to insufficient coverage in diversified authentic data and varied facial scenes. These findings validate that our proposed MFFI dataset achieves superior diversity and authenticity in real-world alignment, establishing a more challenging yet practical benchmark for face forgery detection tasks.

Furthermore, we conducted cross-dataset evaluations on widely adopted unseen test sets in face forgery detection: CDF-V1 [35], CDF-V2 [35], DFD [56], and DFDC [13]. Detection models were trained on the MFFI training set and tested on these unknown benchmarks. As shown in Table 5, the Xception [8] model and SOTA detectors demonstrated comparable generalization performance across all datasets. This indicates that the MFFI training set effectively enhances detection models' generalization capabilities, enabling robust responses to unseen forgery challenges in real-world scenarios.

4.5 Zero-shot Evaluation of MLLM

With the continuous advancement of multi-modal large models, we conducted zero-shot evaluation experiments on the MFFI benchmark to assess their performance in face forgery detection. The results are visualized in Figure 4. We observe that the Llava-1.5 [36] model exhibits abnormally high TN and FN values, indicating

Table 4: Cross-datasets evaluation with the SOTAs. The models train on DF40-FS [76] or FF++(C23) [56], then they test on Test and Test-D sets of ours MFFI datasets.

Models	Train Set	Test				Test-D			
		ACC	AUC	EER	AP	ACC	AUC	EER	AP
Xception [8]	DF40-FS [76]	0.4840	0.4571	0.5309	0.5588	0.4987	0.4741	0.5136	0.5648
RFM [68]		0.5285	0.4772	0.5219	0.5763	0.5332	0.4718	0.5226	0.5678
SRM [41]		0.5067	0.4540	0.5368	0.5591	0.5134	0.4697	0.5236	0.5620
SPSL [37]		0.4919	0.4584	0.5323	0.5707	0.5106	0.4795	0.5134	0.5748
Xception [8]	FF++ (C23) [56]	0.5527	0.6249	0.4059	0.6905	0.5118	0.5118	0.4591	0.6418
RFM [68]		0.5282	0.6175	0.4161	0.6837	0.5169	0.5804	0.4412	0.6463
SRM [41]		0.6041	0.6552	0.3930	0.7148	0.5662	0.5914	0.4370	0.6564
SPSL [37]		0.5536	0.5564	0.4619	0.6416	0.5356	0.5429	0.4737	0.6227

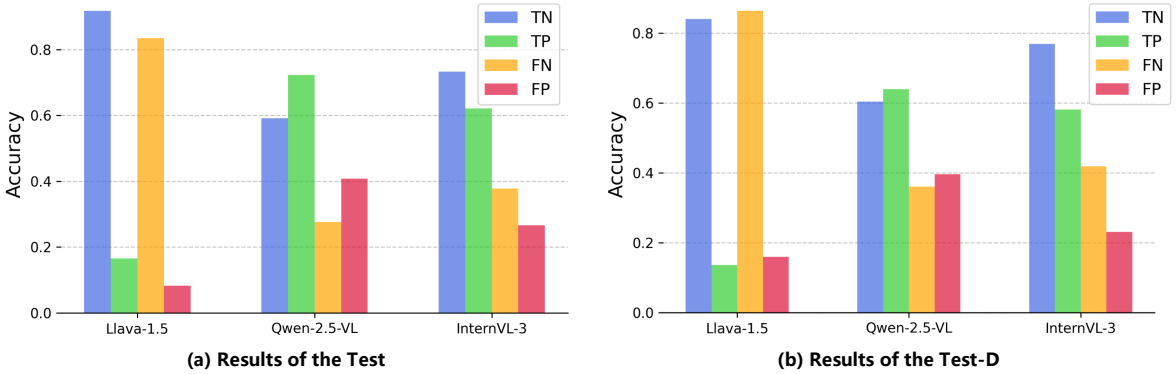

Figure 4: Zero-shot testing performance of MLLM on the MFFI Test and Test-D sets. True Negative (TN): Correctly identified fake samples. True Positive (TP): Correctly identified real samples. False Negative (FN): Real samples misclassified as fake. False Positive (FP): Fake samples misclassified as real.

Table 5: Cross-datasets evaluation with the SOTAs. The models are trained on our MFFI, then they are tested on other datasets. The metric is AUC.

Moedels	CDF-V1 [35]	CDF-V2 [35]	DFD [56]	DFDC [13]	Avg.
Xception [8]	0.6933	0.6695	0.7117	0.5523	0.6567
RFM [68]	0.6452	0.6346	0.6872	0.5375	0.6261
SRM [41]	0.6932	0.5599	0.5921	0.5259	0.5928
SPSL [37]	0.7006	0.7030	0.7535	0.5942	0.6878

a strong bias toward classifying all samples as forged. In contrast, Qwen-2.5-VL [1] and InternVL-3 [88] demonstrate relatively balanced performance, yet their overall accuracy (ACC) remains below 0.68. This underperformance, compared to specialized small models like SRM, fails to reflect the expected advantages of large models. These findings highlight that our MFFI dataset retains significant real-world challenges even for advanced multi-modal large models.

5 Conclusion, Board Impact, and Limitation

Conclusion: This study introduces the Multi-dimensional Face Forgery Image (**MFFI**) dataset to address real-world face forgery detection challenges. By integrating four dimensions: Wider Forgery Methods, Varied Facial Scenes, Diversified Authentic Data, and Multi-level Degradation Operations, our MFFI effectively bridges critical gaps in scene complexity and diversity compared to existing datasets. The dataset encompasses 50 distinct forgery methods and over $1024K$ image samples, demonstrating significant scale advantages. Experimental results show that MFFI outperforms state-of-the-art benchmarks in scene complexity, cross-domain generalization, and gradient-based detection difficulty. These findings validate MFFI's technical innovation and practical value in advancing deepfake detection technologies.

Broader Impact: The release of MFFI provides the research community with a real-world-aligned benchmark for face forgery detection while promoting the development of robust detection techniques to safeguard societal trust. The collection of raw data strictly adheres to source dataset licenses and regulatory guidelines, with usage agreements required during subsequent open-sourcing to ensure privacy protection and standardized data utilization.

Limitation: Despite its large-scale and diverse coverage, the current version is limited to binary labels and excluded text annotations, which remains a critical domain. Future work will incorporate fine-grained forged text annotations to facilitate the development of interpretable forgery detectors based on multi-modal large models.

Acknowledgments

This work was supported by the Ant Group Postdoctoral Programme. This work was also supported by the National Research Foundation Singapore, under its AI Singapore Programme (AISG Award No: AISG3-RP-2024-033), Fundamental Research Funds for the Central Universities (No. PA2025IISL0113, JZ2025HGTB0227), and National Natural Science Foundation of China (No. 62302139).

References

[1] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. 2025. Qwen2.5-VL Technical Report. *arXiv preprint arXiv:2502.13923* (2025).

[2] Diponkor Bala, Md Shamim Hossain, Mohammad Alamgir Hossain, Md Ibrahim Abdullah, Md Mizanur Rahman, Balachandran Manavalan, Naijie Gu, Mohammad S Islam, and Zhangjin Huang. 2023. MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification. *Neural Networks* 161 (2023), 757–775.

[3] Chaitali Bhattacharyya, Hanxiao Wang, Feng Zhang, Sungho Kim, and Xiatian Zhu. 2024. Diffusion deepfake. *arXiv preprint arXiv:2404.01579* (2024).

[4] Stella Bounareli, Christos Tzelepis, Vasileios Argyriou, Ioannis Patras, and Georgios Tzimiropoulos. 2023. Hyperreenact: one-shot reenactment via jointly learning to refine and retarget faces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7149–7159.

[5] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip Yu, and Lichao Sun. 2025. A survey of ai-generated content (aigc). *Comput. Surveys* 57, 5 (2025), 1–38.

[6] Junsong Chen, YU Jincheng, GE Chongjian, Lewei Yao, Enze Xie, Zhongdao Wang, James Kwok, Ping Luo, Huchuan Lu, and Zhenguo Li. 2024. PixArt-alpha: Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis. In *The Twelfth International Conference on Learning Representations*.

[7] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. 2020. Simswap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM international conference on multimedia*. 2003–2011.

[8] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.

[9] DeepFakes. 2019. https://github.com/deepfakes/.

[10] DeepFakes. 2023. https://github.com/facefusion.

[11] Yunfeng Diao, Baiqi Wu, Ruixuan Zhang, Ajian Liu, Xingxing Wei, Meng Wang, and He Wang. 2025. TASAR: Transfer-based Attack on Skeletal Action Recognition. In *The International Conference on Learning Representations (ICLR)*.

[12] Yunfeng Diao, Naixin Zhai, Changtao Miao, Zitong Yu, Xingxing Wei, Xun Yang, and Meng Wang. 2024. Vulnerabilities in ai-generated image detection: The challenge of adversarial attacks. *arXiv preprint arXiv:2407.20836* (2024).

[13] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. 2020. The deepfake detection challenge (dfdc) dataset. *arXiv preprint arXiv:2006.07397* (2020).

[14] Paul Dubois. 2025. Face Merge 2. https://github.com/pauldubois98/FaceMerge2.

[15] FaceSwap. 2019. https://github.com/MarekKowalski/FaceSwap.

[16] Gege Gao, Huaibo Huang, Chaoyou Fu, Zhaoyang Li, and Ran He. 2021. Information bottleneck disentanglement for identity swapping. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3404–3413.

[17] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[18] Yinan He, Bei Gan, Siyu Chen, Yichun Zhou, Guojun Yin, Luchuan Song, Lu Sheng, Jing Shao, and Ziwei Liu. 2021. Forgerynet: A versatile benchmark for comprehensive forgery analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4360–4369.

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[20] Fa-Ting Hong and Dan Xu. 2023. Implicit identity representation conditioned memory compensation network for talking head video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 23062–23072.

[21] Fa-Ting Hong, Longhao Zhang, Li Shen, and Dan Xu. 2022. Depth-aware generative adversarial network for talking head video generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3397–3406.

[22] Liming Jiang, Ren Li, Wayne Wu, Chen Qian, and Chen Change Loy. 2020. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2889–2898.

[23] Jimeng Team. 2025. Jimeng3.0. https://jimeng.jianying.com/ai-tool/video/generate. Accessed: 2025-05-30.

[24] Yan Ju, Shan Jia, Jialing Cai, Haiying Guan, and Siwei Lyu. 2023. Glff: Global and local feature fusion for ai-synthesized image detection. *IEEE Transactions on Multimedia* 26 (2023), 4073–4085.

[25] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *Advances in neural information processing systems* 34 (2021), 852–863.

[26] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[27] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8110–8119.

[28] Hasam Khalid, Shahroz Tariq, Minha Kim, and Simon S Woo. 2021. FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

[29] Kling Team. 2025. Kling1.6. https://klingai.com/global/. Accessed: 2025-05-30.

[30] Chenqi Kong, Baoliang Chen, Haoliang Li, Shiqi Wang, Anderson Rocha, and Sam Kwong. 2022. Detect and locate: Exposing face manipulation by semantic-and noise-level telltales. *IEEE Transactions on Information Forensics and Security* 17 (2022), 1741–1756.

[31] Pavel Korshunov and Sébastien Marcel. 2018. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685* (2018).

[32] Jianshu Li, Man Luo, Jian Liu, Tao Chen, Chengjie Wang, Ziwei Liu, Shuo Liu, Kewei Yang, Xuning Shao, Kang Chen, et al. 2022. Multi-Forgery Detection Challenge 2022: Push the Frontier of Unconstrained and Diverse Forgery Detection. *arXiv preprint arXiv:2207.13505* (2022).

[33] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. 2019. Faceshifter: Towards high fidelity and occlusion aware face swapping. *arXiv preprint arXiv:1912.13457* (2019).

[34] Maomao Li, Ge Yuan, Cairong Wang, Zhian Liu, Yong Zhang, Yongwei Nie, Jue Wang, and Dong Xu. 2023. E4S: Fine-grained Face Swapping via Editing With Regional GAN Inversion. *arXiv preprint arXiv:2310.15081* (2023).

[35] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. 2020. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3207–3216.

[36] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. Visual Instruction Tuning.

[37] Honggu Liu, Xiaodan Li, Wenbo Zhou, Yuefeng Chen, Yuan He, Hui Xue, Weiming Zhang, and Nenghai Yu. 2021. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 772–781.

[38] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2018. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August* 15, 2018 (2018), 11.

[39] Anwei Luo, Chenqi Kong, Jiwu Huang, Yongjian Hu, Xiangui Kang, and Alex C Kot. 2023. Beyond the prior forgery knowledge: Mining critical clues for general face forgery detection. *IEEE Transactions on Information Forensics and Security* 19 (2023), 1168–1182.

[40] Wuyang Luo, Su Yang, and Weishan Zhang. 2023. Reference-guided large-scale face inpainting with identity and texture control. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 10 (2023), 5498–5509.

[41] Yuchen Luo, Yong Zhang, Junchi Yan, and Wei Liu. 2021. Generalizing face forgery detection with high-frequency features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16317–16326.

[42] Wang Mei and Deng Weihong. 2020. Mitigating bias in face recognition using skewness-aware reinforcement learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9322–9331.

[43] Mellis, Inc. 2024. Pika (Version 1.5). https://pika.art/. Accessed: May 30, 2025. Version 1.5 reportedly available around October 2024..

[44] Changtao Miao, Qi Chu, Weihai Li, Suichan Li, Zhentao Tan, Wanyi Zhuang, and Nenghai Yu. 2022. Learning Forgery Region-aware and ID-independent Features for Face Manipulation Detection. *IEEE TBIOM* 4, 1 (2022), 71–84.

[45] Changtao Miao, Qi Chu, Zhentao Tan, Zhenchao Jin, Wanyi Zhuang, Yue Wu, Bin Liu, Honggang Hu, and Nenghai Yu. 2023. Multi-spectral Class Center Network for Face Manipulation Detection and Localization. *arXiv preprint arXiv:2305.10794* (2023).

[46] Changtao Miao, Zichang Tan, Qi Chu, Huan Liu, Honggang Hu, and Nenghai Yu. 2023. F 2 Trans: High-Frequency Fine-Grained Transformer for Face Forgery Detection. *IEEE Transactions on Information Forensics and Security* 18 (2023), 1039–1051.

[47] Changtao Miao, Zichang Tan, Qi Chu, Nenghai Yu, and Guodong Guo. 2022. Hierarchical frequency-assisted interactive networks for face manipulation detection. *IEEE Transactions on Information Forensics and Security* 17 (2022), 3008–3021.

[48] Aakash Varma Nadimpalli and Ajita Rattani. 2022. On improving cross-dataset generalization of deepfake detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 91–99.

[49] Kartik Narayan, Harsh Agarwal, Kartik Thakral, Surbhi Mittal, Mayank Vatsa, and Richa Singh. 2023. DF-Platter: Multi-Face Heterogeneous Deepfake Dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9739–9748.

[50] Yuval Nirkin, Yosi Keller, and Tal Hassner. 2019. Fsgan: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE/CVF international conference on computer vision*. 7184–7193.

[51] PixVerse. 2025. PixVerse AI Video Generator. https://platform.pixverse.ai/. Accessed: May 30, 2025.

[52] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2024. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In *The Twelfth International Conference on Learning Representations*.

[53] KR Prajwal, Rudrabha Mukhopadhyay, Vinay P Namboodiri, and CV Jawahar. 2020. A lip sync expert is all you need for speech to lip generation in the wild. In *Proceedings of the 28th ACM international conference on multimedia*. 484–492.

[54] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2287–2296.

[55] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[56] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. 2019. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1–11.

[57] shaoanlu. 2019. faceswap-GAN. https://github.com/shaoanlu/faceswap-GAN. Accessed: 2025-05-30.

[58] Kaede Shiohara and Toshihiko Yamasaki. 2022. Detecting deepfakes with self-blended images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18720–18729.

[59] Aliaksandr Siarohin, Oliver J Woodford, Jian Ren, Menglei Chai, and Sergey Tulyakov. 2021. Motion representations for articulated animation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13653–13662.

[60] Haixu Song, Shiyu Huang, Yinpeng Dong, and Wei-Wei Tu. 2023. Robustness and generalizability of deepfake detection: A study with diffusion models. *arXiv preprint arXiv:2309.02218* (2023).

[61] Sparking Innovations Limited. 2025. Vivago AI Video Generation. https://vivago.ai/video-generation. Accessed: May 30, 2025.

[62] Zichang Tan, Zhichao Yang, Changtao Miao, and Guodong Guo. 2022. Transformer-based feature compensation and aggregation for deepfake detection. *IEEE Signal Processing Letters* 29 (2022), 2183–2187.

[63] Hailuo AI Team. 2025. Hailuo AI: Idea to Visual - I2V-01-Live Model. https://www.hailuoai.com.

[64] Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2019. Deferred neural rendering: Image synthesis using neural textures. *Acm Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.

[65] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2016. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2387–2395.

[66] Vidu Team. 2025. Vidu2.0. https://www.vidu.com/. Accessed: 2025-05-30.

[67] Wan Team. 2025. Wan2.1. https://wan.video/. Accessed: 2025-05-30.

[68] Chengrui Wang and Weihong Deng. 2021. Representative Forgery Mining for Fake Face Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14923–14932.

[69] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. 2019. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In *Proceedings of the ieee/cvf international conference on computer vision*. 692–702.

[70] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, Anthony Chen, Huaxia Li, Xu Tang, and Yao Hu. 2024. Instantid: Zero-shot identity-preserving generation in seconds. *arXiv preprint arXiv:2401.07519* (2024).

[71] Tengfei Wang, Yong Zhang, Yanbo Fan, Jue Wang, and Qifeng Chen. 2022. High-fidelity gan inversion for image attribute editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11379–11388.

[72] Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji. 2021. Hififace: 3d shape and semantic prior guided high fidelity face swapping. *arXiv preprint arXiv:2106.09965* (2021).

[73] Huawei Wei, Zejun Yang, and Zhisheng Wang. 2024. Aniportrait: Audio-driven synthesis of photorealistic portrait animation. *arXiv preprint arXiv:2403.17694* (2024).

[74] Chao Xu, Jiangning Zhang, Miao Hua, Qian He, Zili Yi, and Yong Liu. 2022. Region-aware face swapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7632–7641.

[75] Zhiliang Xu, Zhibin Hong, Changxing Ding, Zhen Zhu, Junyu Han, Jingtuo Liu, and Errui Ding. 2022. Mobilefaceswap: A lightweight framework for video face swapping. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 2973–2981.

[76] Zhiyuan Yan, Taiping Yao, Shen Chen, Yandan Zhao, Xinghe Fu, Junwei Zhu, Donghao Luo, Chengjie Wang, Shouhong Ding, Yunsheng Wu, et al. 2024. DF40: Toward Next-Generation Deepfake Detection. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

[77] Zhiyuan Yan, Yong Zhang, Xinhang Yuan, Siwei Lyu, and Baoyuan Wu. 2023. DeepfakeBench: A Comprehensive Benchmark of Deepfake Detection. In *Advances in Neural Information Processing Systems*. 4534–4565.

[78] Chenglin Yang, Adam Kortylewski, Cihang Xie, Yinzhi Cao, and Alan Yuille. 2020. Patchattack: A black-box texture-based attack with reinforcement learning. In *European Conference on Computer Vision*. Springer, 681–698.

[79] Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. 2022. Pastiche master: Exemplar-based high-resolution portrait style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7693–7702.

[80] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. 2021. Gan prior embedded network for blind face restoration in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 672–681.

[81] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. 2023. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721* (2023).

[82] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. 2014. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923* (2014).

[83] Ruixuan Zhang, He Wang, Zhengyu Zhao, Zhiqing Guo, Xun Yang, Yunfeng Diao, and Meng Wang. 2025. Adversarially Robust AI-Generated Image Detection for Free: An Information Theoretic Perspective. *arXiv preprint arXiv:2505.22604* (2025).

[84] Yi Zhang, Weize Gao, Changtao Miao, Man Luo, Jianshu Li, Wenzhong Deng, Zhe Li, Bingyu Hu, Weibin Yao, Wenbo Zhou, et al. 2024. Inclusion 2024 Global Multimedia Deepfake Detection: Towards Multi-dimensional Facial Forgery Detection. *arXiv preprint arXiv:2412.20833* (2024).

[85] Yaning Zhang, Zitong Yu, Tianyi Wang, Xiaobin Huang, Linlin Shen, Zan Gao, and Jianfeng Ren. 2024. Genface: A large-scale fine-grained face forgery benchmark and cross appearance-edge learning. *IEEE Transactions on Information Forensics and Security* (2024).

[86] Yi Zhang, Youjun Zhao, Yuhang Wen, Zixuan Tang, Xinhua Xu, and Mengyuan Liu. 2021. Facial prior based first order motion model for micro-expression generation. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4755–4759.

[87] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia. 2021. Learning Self-Consistency for Deepfake Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 15023–15033.

[88] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, Zhangwei Gao, Erfei Cui, Xuehui Wang, Yue Cao, Yangzhou Liu, Xingguang Wei, Hongjie Zhang, Haomin Wang, Weiye Xu, Hao Li, Jiahao Wang, Nianchen Deng, Songze Li, Yinan He, Tan Jiang, Jiapeng Luo, Yi Wang, Conghui He, Botian Shi, Xingcheng Zhang, Wenqi Shao, Junjun He, Yingtong Xiong, Wenwen Qu, Peng Sun, Penglong Jiao, Han Lv, Lijun Wu, Kaipeng Zhang, Huipeng Deng, Jiaye Ge, Kai Chen, Limin Wang, Min Dou, Lewei Lu, Xizhou Zhu, Tong Lu, Dahua Lin, Yu Qiao, Jifeng Dai, and Wenhai Wang. 2025. InternVL3: Exploring Advanced Training and Test-Time Recipes for Open-Source Multimodal Models. arXiv:2504.10479 [cs.CV] https://arxiv.org/abs/2504.10479

[89] Mingjian Zhu, Hanting Chen, Qiangyu Yan, Xudong Huang, Guanyu Lin, Wei Li, Zhijun Tu, Hailin Hu, Jie Hu, and Yunhe Wang. 2023. Genimage: A million-scale benchmark for detecting ai-generated image. *Advances in Neural Information Processing Systems* 36 (2023), 77771–77782.

[90] Tianlei Zhu, Junqi Chen, Renzhe Zhu, and Gaurav Gupta. 2023. StyleGAN3: generative networks for improving the equivariance of translation and rotation. *arXiv preprint arXiv:2307.03898* (2023).

[91] Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu, and Zhenan Sun. 2021. One shot face swapping on megapixels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4834–4844.

[92] Wanyi Zhuang, Qi Chu, Zhentao Tan, Qiankun Liu, Haojie Yuan, Changtao Miao, Zixiang Luo, and Nenghai Yu. 2022. UIA-ViT: Unsupervised inconsistency-aware method based on vision transformer for face forgery detection. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*. Springer, 391–407.

[93] Wanyi Zhuang, Qi Chu, Haojie Yuan, Changtao Miao, Bin Liu, and Nenghai Yu. 2022. Towards intrinsic common discriminative features learning for face forgery detection using adversarial learning. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.