Celeb-DF++：一个用于通用法证的大规模挑战性视频 DeepFake 基准

Yuezun Li, Delong Zhu, Xinjie Cui, Siwei Lyu, *Fellow, IEEE*

***Abstract*—AI 技术的迅速发展显著增加了网上流传的 DeepFake 视频的多样性，给可泛化的取证工作带来了紧迫的挑战，即使用单一模型检测广泛的未曾见过的 DeepFake 类型。应对这一挑战需要的数据集不仅要大规模，而且还要具备伪造的多样性。然而，大多数现有数据集尽管规模大，但只包含有限种类的伪造类型，使其不足以开发通用的检测方法。因此，我们在之前的 Celeb-DF 数据集基础上，介绍了 Celeb-DF++，这是一个新的大规模且具有挑战性的视频 DeepFake 基准，专注于可泛化的取证挑战。Celeb-DF++ 覆盖了三种常见的伪造场景：面部交换（FS）、面部重演（FR）和说话脸（TF）。每个场景都包含大量通过总共 22 种最新 DeepFake 方法生成的高质量伪造视频。这些方法在架构、生成流程和目标面部区域上有所不同，涵盖了野外最常见的 DeepFake 案例。我们还介绍了用于评估 24 种最新检测方法的泛化能力的评估协议，突出了现有检测方法的局限性以及我们新数据集的难度。该数据集已在 https://github.com/OUC-VAS/Celeb-DF-PP 发布。**

***Index Terms*—DeepFake Benchmark, Generalizable Forensics**

I. 介绍

深度伪造 指的是基于人工智能的面部合成技术，这些技术可以轻易地创造出具有高度真实感的伪造视频。考虑到面部与身份之间的紧密联系，精心制作的 DeepFakes 可以制造出个人参与从未发生事件的幻觉，从而引发严重的政治、社会、金融和法律问题 [1]–[3] 。

为应对 DeepFakes，近年来开发了许多检测方法 *e.g.*, [4]–[16] ，这需要大规模数据集进行训练和评估。迄今为止，已发布了许多数据集，如 UADFV [7] 、DeepFake-TIMIT [17] 、FaceForenscics++ [18] 、DFD [19] 、DFDC [20] 、Celeb-DF [21] 等。虽然这些数据集极大地加速了 DeepFake 检测的进展，但它们是利用早期的 DeepFake 方法构建的，因而在伪造类型上常常缺乏多样性。

近年来，生成模型（*e.g.*, GAN [22] , Diffusion model [23] ）的迅速发展导致了 DeepFake 方法的发展激增，大大增加了其多样性。随着架构的不断优化和生成策略的演变，DeepFake 变得越来越多样化，更令人担忧的是，其中许多细节可能是未知的。这种情况对 DeepFake 检测提出了一个实际而紧迫的挑战，称为可泛化的取证：检测器是否能有效识别各种未知的 DeepFake?（见图 1 ）。然而，由于现有数据集中伪造的多样性有限，它们在支持能在实际应用中泛化良好的检测方法的发展方面显得不足。这突显了构建一个大规模且多样化的数据集以推动 DeepFake 取证进展的关键需求。

Yuezun Li is corresponding author.
Yuezun Li, Delong Zhu, and Xinjie Cui are with the School of Computer Science and Technology, Ocean University of China, Qingdao, China. Email: liyuezun@ouc.edu.cn; { zhudelong;cuixinjie } @stu.ouc.edu.cn.
Siwei Lyu is with the University at Buffalo, SUNY, USA. Email: siweilyu@buffalo.edu.

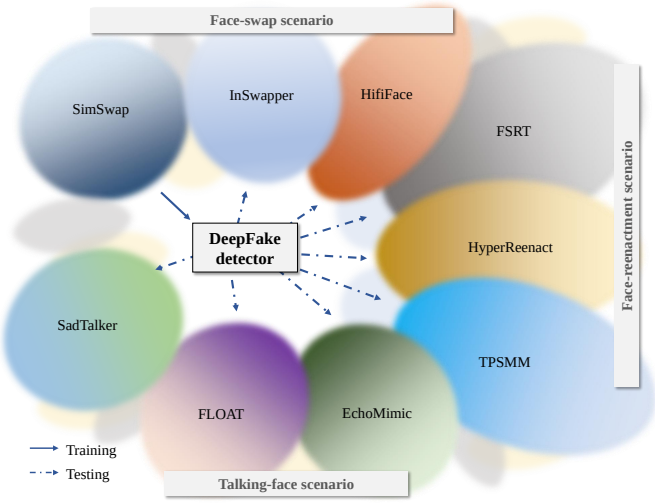

Fig. 1. 提出的 Celeb-DF++ 基准概述。该基准的动机是深度伪造检测中通用取证的需要。

为了解决这个差距，我们引入了 Celeb-DF++ ，其中包含大量不同的 DeepFake 方法，分别涵盖三个常见的场景：换脸（FS）、面部重现（FR）和说话脸（TF）[1] 。在 FS 场景中，来源个体的面部被目标个体的合成面部所替代，同时保持一致的面部特征。FR 涉及生成由来源个体的行为驱动的目标个体的新视频，以确保行为的一致性。TF 利用音频输入生成目标个体的合成视频中的同步唇部运动。对于每个场景，我们采用了一套多样的最先进的 DeepFake 方法，FS 8 种，FR 7 种，TF 7 种，共覆盖 22 种具代表性的方法。这些方法在架构、生成流程和被操控的面部区域上各不相同，有效地模拟了现实媒体中遇到的多样性。

Celeb-DF++ 是在我们早期的 Celeb-DF 数据集 [21] 基础上构建的，该数据集包含 59 名不同性别、年龄和种族的名人。真实视频集保持不变，包括 590 个在 YouTube 上公开可获得的视频。我们使用大量先进和最新的 DeepFake 方法，将 DeepFake 视频扩展到 53196 个，每个视频平均时长 10 秒，总计超过 15 百万帧。遵循先前工作中广泛使用的评估设置 [13], [15], [24] ，我们展示了所提议基准的增加难度。该基准的详细信息如表 I 所示。

为了全面衡量 DeepFake 检测方法的泛化能力，我们依托 Celeb-DF++ 基准建立了三个具有挑战性的评估协议：泛化伪造评估（GF-eval）、跨质量泛化伪造评估（GFQ-eval）以及跨数据集泛化伪造评估（GFD-eval）。GF-eval 在所有三种伪造场景中评估检测表现，反映了伪造多样且

[1]全脸合成和脸部编辑也是常见的 DeepFake 形式。然而，由于它们专注于图像处理且难以在多个帧之间保持时间一致性，它们很少应用于视频。鉴于视频通常更具欺骗性和影响力，我们在我们的基准测试中排除这些类型。

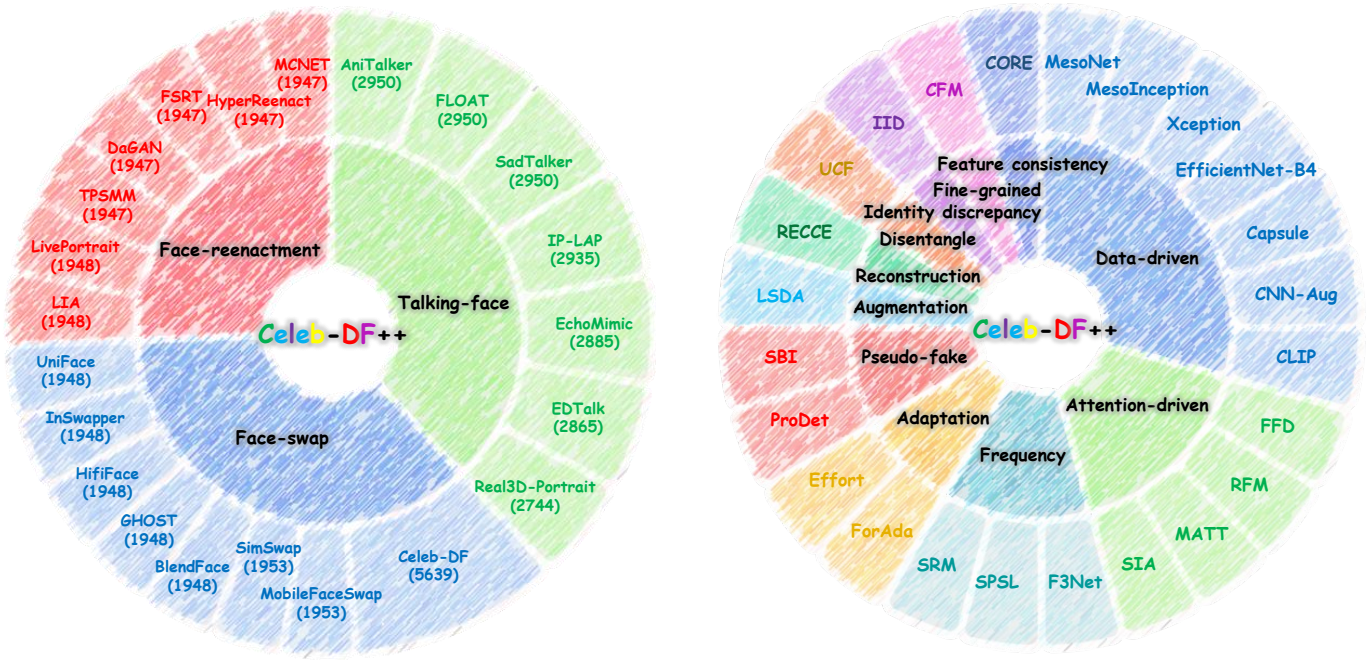

Fig. 2. The sunburst chart highlights that our Celeb-DF++ includes a wide range of both DeepFake methods and assessed detection methods. See Table I, Table II and Table III for details.

TABLE I
DEEPFAKE 视频基准的比较。我们强调，CELEB-DF++ 更为多样，包括 22 种不同的 DEEPFAKE 方法，涵盖了脸部替换（FS）、脸部重演（FR）和说话脸（TF）场景，并使用 24 种检测器进行全面的最新评估，包括 5 种 2024 年后发布的最新检测器。

Dataset	Venue	Modality		DeepFake Model				DeepFake Method				# Real	# Fake	Assessed Detectors			Source
		Visual	Audio	AE	GAN	DM	Other[2]	# Total	# FS	# FR	# TF			# Total	Pre'24	Post'24	
DF-TIMIT [17]	ArXiv'18	✓			✓			2	2	-	-	320	640	4	4	-	Link
DFDCP [25]	ArXiv'19	✓			✓			2	2	-	-	1,131	4,113	3	3	-	Link
UADFV [7]	ICASSP'19	✓					✓	1	1	-	-	49	49	6	6	-	Link
FF++ [18]	ICCV'19	✓		✓			✓	4	2	2	-	1,000	4,000	6	6	-	Link
DFD [19]	-	✓		-	-	-	-	5	5	-	-	363	3,068	-	-	-	Link
DFDC [20]	ArXiv'20	✓		✓	✓		✓	8	6	-	2	23,654	104,500	5	5	-	Link
DFFD [26]	CVPR'20	✓		✓	✓		✓	7	7	-	-	1,000	3,000	10	10	-	Link
DForensics-1.0 [27]	CVPR'20	✓		✓				1	1	-	-	50,000	10,000	5	5	-	Link
WildDeepfake [28]	MM'20	✓		-	-	-	-	-	-	-	-	3,805	3,509	15	15	-	Link
OpenForensics[3] [29]	ICCV'21	✓			✓			1	1	-	-	45,473	70,325	12	12	-	Link
KoDF [30]	ICCV'21	✓		✓	✓		✓	6	3	1	2	62,166	175,776	1	1	-	Link
FFIW [31]	CVPR'21	✓		✓	✓			3	3	-	-	10,000	10,000	9	9	-	Link
FakeAVCeleb [32]	NeurIPS'21	✓	✓	✓	✓		✓	4	2	-	2	500	19,500	8	8	-	Link
DFDM [33]	ICIP'22	✓		✓				5	5	-	-	590	6,450	9	9	-	Link
DF-Platter [34]	CVPR'23	✓		✓	✓			3	3	-	-	764	132,496	6	6	-	Link
DefakeAVMiT [35]	TIFS'23	✓	✓	✓	✓	✓		5	2	-	3	540	6,480	21	21	-	Link
AV-Deepfake1M [36]	MM'24	✓	✓		✓			1	-	-	1	286,721	860,039	22	22	-	Link
THBench [37]	ArXiv'25	✓	✓		✓	✓		6	-	-	6	2,312	2,984	4	2	2	Link
(Ours) Celeb-DF [21]	CVPR'20	✓		✓				1	1	-	-	590	5,639	13	13	-	Link
(Ours) Celeb-DF++	-	✓	✓	✓	✓	✓	✓	22	8	7	7	590	53,196	24	19	5	Link

可能未曾见过的真实世界条件。GFQ-eval 通过在不同视频压缩级别下评估检测性能增加了挑战。这与在线视频通常受到压缩，且这种压缩可能掩盖伪造痕迹的现实情况一致，提高了具有泛化能力的法医分析的难度。GFD-eval 评估检测方法在不同数据集间的泛化能力，模拟了训练和测试数据来自不同来源的实际情况。

使用这些评估协议，我们进行了大量实验，实验中使用了 24 种最新的检测方法。与现有的数据集相比，我们的评估包含了更多的最新方法，突出了其全面性和及时性。结果显示，具有普遍适应性的 DeepFake 检测仍然是一个未解决的挑战，需要持续的研究和创新。

II. 相关工作

A. 深度伪造生成

原始 DeepFake 。DeepFake 这个术语出现于 2017 年，最初是指面部交换伪造技术。典型的流程包括从源视频中提取面部，并将其输入基于自编码器的生成模型，这些模型合成与目标个体具有相同面部属性（如方向和表情）的面部。过去几年中，面部交换技术已得到充分发展，促使许多开源工具的发布，*e.g.*， [38]–[44] 。虽然其方法相对简单，但这些工具用户友好，性能可靠，且已获得广泛普及。目前，网上流传的大量 DeepFake 视频就是利用这些工具制作的，形成了当前的面部交换伪造格局。

随着生成模型的快速进化，换脸技术变得越来越高效和有效，不断改善生成质量。与此同时，许多其他类型的媒体伪造也随之出现。其中，人脸重演和谈话人脸最近吸引

了相当大的关注。与换脸不同，人脸重演是基于驱动源生成整个视频帧，而谈话人脸则基于给定的音频轨生成唇部运动，甚至面部表情。其他伪造类别，如面部属性编辑和风格转移，相对而言不那么具有欺骗性，因为它们只专注于操纵图像，并且难以在视频帧之间保持一致性。有趣的是，生成模型还促成了基于音频的伪造，例如合成从未说出这些话的目标个体的语音。因此，DeepFake 的概念已经扩展到代表一系列基于 AI 的伪造场景。

B. DeepFake 检测

常规检测。为了识别 DeepFakes，在短时间内提出了大量的检测方法 [6]–[9], [11], [12], [16], [26] 。这些方法大多数基于深度神经网络（DNNs），利用其强大的特征学习能力。根据这些方法旨在提取的特征类型，可以将其大致分为几类。一种经典类别侧重于检测物理或生理特征中的不一致性，比如异常的眨眼 [6] 、不规则的头部姿态运动 [7] 或心跳节律的中断 [16] 。另一种常见的方向尝试捕捉生成过程引入的伪造伪迹，包括在空间或频率域中的混合伪迹 [8], [9] 、生成伪迹 [11], [45] 和时间伪迹 [46], [47] 。第三种研究线设计专门的网络架构和训练目标，以直接从数据中学习判别性特征。这些包括结合注意力机制 [12] 、对比学习 [48] 、增强特征提取模块 [26] 等策略。

泛化检测。尽管检测方法展示了有前景的结果，但在检测未见过的伪造时仍然困难，这限制了其实用性。许多努力致力于通过提升检测的泛化性来解决这个问题。一个有效的解决方案是创建多样的伪造训练人脸，通过模拟在真实 DeepFakes 中常见的伪造特征 [15], [49]–[51] 。另一种研究方向集中于从已知的训练样本中学习可泛化的伪造特征，使用诸如对抗训练 [52] 、特征解缠 [13] 、自监督学习 [53] 等策略。尽管这些方法提高了泛化性，现有数据集缺乏多样性阻碍了该领域的进一步发展。

C. DeepFake 数据集

深度伪造检测方法的训练和评估需要数据集。早期的深度伪造数据集由于最初深度伪造方法的能力不足，通常规模有限。一些最早公开可用的数据集，如 UADFV [6] 和 DeepFake-TIMIT (DF-TIMIT) [17] ，仅包含使用早期面部交换方法生成的少量深度伪造视频 [38], [40] 。人脸取证 ++ (FF++) [18] 的发布标志着大规模深度伪造数据集的出现，包括数千个真实视频和使用各种面部交换处理生成的相同数量的深度伪造视频。后来，谷歌 & JigSaw 发布了 深度伪造检测 (DFD) 数据集 [19] ，其特点是更大规模的深度伪造视频集合，这些视频来自于同意的演员。

随着 DeepFakes 的近期进展，数据集不断演变，其规模和多样性显著增加。Facebook（现为 Meta）启动了一个 DeepFake 检测挑战 (DFDC) ，并发布了一个相应的数据集 [20] 。最初，一个预览版本 (DFDCP) 被发布，在挑战完成后，完整的数据集被公开。该数据集包含了数十万段视频剪辑，特邀演员并使用了各种基于 GAN 和非学习的 DeepFake 方法。其他主流数据集如 深度前馈密度 [26] ，深度取证-1.0 (DForensics-1.0) [27] ，和 野生 Deepfake [28] ，按时发布，规模较大，涵盖了更多样化的场景和更广泛的面部表情。我们之前的工作 Celeb-DF 数据集 [21] ，包括大规模和挑战性的名人 DeepFake 视频，通过精心策划的生成管道实现了增强的视觉质量。此外，一些专门的数据集被创建来解决特定的挑战，比如 KoDF 数据集 [30] 中缺乏亚洲主体的问题， DFDM 数据集 [33] 中的模型归因问题， TalkingHeadBench (THBench) 数据集 [37] 中的讲话头检测问题，以及 开放法证学 数据集 [29] ，FFIW 数据集 [31] 和 DF-Platter 数据集 [34] 中的多脸伪造检测问题。

TABLE II
DEEPFAKE METHODS USED IN CELEB-DF++ BENCHMARK.

Scenario	DeepFake Method	Venue	Code
FS	Celeb-DF [21]	CVPR'20	Link
	SimSwap [54]	MM'20	Link
	InSwapper [55]	-	Link
	HifiFace [56]	IJCAI'21	Link
	GHOST [57]	Access'22	Link
	UniFace [58]	ECCV'22	Link
	MobileFaceSwap [59]	AAAI'22	Link
	BlendFace [60]	ICCV'23	Link
FR	DaGAN [61]	CVPR'22	Link
	TPSMM [62]	CVPR'22	Link
	MCNET [63]	ICCV'23	Link
	HyperReenact [64]	ICCV'23	Link
	LIA [65]	TPAMI'24	Link
	FSRT [66]	CVPR'24	Link
	LivePortrait [67]	ArXiv'24	Link
TF	SadTalker [68]	CVPR'23	Link
	IP-LAP [69]	CVPR'23	Link
	AniTalker [70]	MM'24	Link
	EDTalk [71]	ECCV'24	Link
	Real3D-Portrait [72]	ICLR'24	Link
	EchoMimic [73]	AAAI'25	Link
	FLOAT [74]	ICCV'25	Link

除了单模态数据集外，一些多模态数据集已经出现，结合了视觉和音频信息，例如 FakeAVCeleb 数据集 [32] 、DefakeAVMiT [35] 和 AV-Deepfake1M 数据集 [36] 。

我们强调，虽然现有的数据集提供了有前景的数据规模，但它们在 DeepFake 方法上缺乏足够的多样性。其中大多数只考虑了单一的伪造场景，如换脸，这限制了它们有效评估检测方法泛化能力的能力。详细比较如表 I 所示。

III. CELEB-DF++ 基准测试

Celeb-DF++ 是从我们先前的 Celeb-DF 数据集 [21] 扩展而来的，具备更多的多样性，包括 22 种不同的 DeepFake 方法，这些方法涵盖了面部交换 (FS)、面部再现 (FR) 和动态人脸 (TF) 场景。此外，我们使用 24 个检测器进行了全面的最新评估，其中包括 5 个在 2024 年后发布的且未在现有数据集中考虑的新检测器。现有数据集的汇总如表 I 所示。

A. 重访 Celeb-DF 数据集

Celeb-DF 数据集包括 590 个真实视频和 5,639 个 DeepFake 视频（对应超过两百万个视频帧）。所有视频的平均长度约为 13 秒，标准帧率为 30 帧每秒。真实视频从公开的 YouTube 视频中选择，对应的采访对象是 59 位名人，他们在性别、年龄和族群上分布多样 [4] 。真实视频中的 56.8% 个对象是男性，43.2% 个是女性。8.5% 个年龄在 60 岁及以上，30.5% 个年龄在 50 到 60 岁之间，26.6%

[2]Other category includes traditional graphics-based approaches [41], [75], neural rendering technique [76], and 3DMM–based generation methods [56], [77].

[3]Note that OpenForensics [29] only contains images, # Real and # Fake in the table refer to the number of images rather than videos.

[4]我们选择名人的面孔，因为观众对这些面孔更为熟悉，因此任何视觉瑕疵都能更容易地识别。此外，名人据说是 DeepFake 视频的主要目标。

TABLE III
评估的 DEEPFAKE 检测方法。

Detector	Venue	Architecture	Category[5]	Code
MesoNet [4]	WIFS'18	Designed CNN	Data-driven	Link
MesoInception [4]	WIFS'18	Designed CNN	Data-driven	Link
Xception [18]	ICCV'19	Xception	Data-driven	Link
EfficientNet-B4 [78]	ICML'19	EfficientNet	Data-driven	Link
Capsule [10]	ICASSP'19	CapsuleNet	Data-driven	Link
F3Net [11]	ECCV'20	Designed CNN	Frequency	Link
CNN-Aug [79]	CVPR'20	ResNet	Data-driven	Link
FFD [26]	CVPR'20	Designed CNN	Attention-driven	Link
SPSL [80]	CVPR'21	Designed CNN	Frequency	Link
SRM [81]	CVPR'21	Designed CNN	Frequency	Link
RFM [82]	CVPR'21	Designed CNN	Attention-driven	Link
MATT [24]	CVPR'21	Designed CNN	Attention-driven	Link
CLIP [83]	ICML'21	CLIP	Data-driven	Link
RECCE [5]	CVPR'22	Designed CNN	Reconstruction	Link
SBI [49]	CVPR'22	Designed CNN	Pseudo-fake	Link
CORE [84]	CVPRW'22	Designed CNN	Feature consistency	Link
SIA [12]	ECCV'22	Designed CNN	Attention-driven	Link
UCF [13]	ICCV'23	Designed CNN	Disentangle	Link
IID [85]	CVPR'23	Designed CNN	Identity discrepancy	Link
LSDA [86]	CVPR'24	Designed CNN	Augmentation	Link
CFM [45]	TIFS'24	Designed CNN	Fine-grained	Link
ProDet [51]	NeurIPS'24	Designed CNN	Pseudo-fake	Link
ForAda [15]	CVPR'25	Designed CLIP	Adaptation	Link
Effort [14]	ICML'25	Designed CLIP	Adaptation	Link

个年龄在 40 多岁，28.0% 个年龄在 30 多岁，6.4% 个年龄在 30 岁以下。5.1% 个是亚洲人，6.8% 个是非裔美国人，88.1% 个是白人。此外，真实视频在主体的面部尺寸（以像素为单位）、朝向、光照条件和背景等方面展现出大范围的变化。DeepFake 视频通过交换每对 59 个对象的面部生成。最终视频是 MPEG4.0 格式。

为了生成 DeepFake 视频，我们通过扩大合成尺寸、减少替换区域与周围环境之间的颜色不匹配、优化替换掩膜以及缓解时间闪烁，改进了初始的换脸方法 [38] 。这个数据集通常用于训练和评估 DeepFake 检测模型。然而，由于已发布的 DeepFake 方法的多样性有限，这个数据集仅涉及使用单一 DeepFake 方法的换脸场景。因此，它不足以评估 DeepFake 检测方法的泛化能力。

在 Celeb-DF++ 中，我们分别探索了三种伪造场景：面部交换（FS）、面部重演（FR）和说话人脸（TF），且每种场景均使用大量近期的 DeepFake 方法构建。表格 II 说明了每种方法的详细信息，图 **??** 展示了几个生成实例。

换脸场景。这个场景指的是面部伪造，即原始面部区域（*e.g.*，整个面部或面部器官如眼睛、嘴巴等）被一个新合成的面部区域替换，同时保留相同的行为属性。我们在这个场景中包括了 8 种方法，分别是 Celeb-DF [21] 中改进的换脸方法，以及七种额外的方法：SimSwap [54] 、InSwapper [55] 、HifiFace [56] 、GHOST [57] 、UniFace [58] 、MobileFaceSwap [59] 、BlendFace [60] 。这些方法依赖于自编码器或 GAN，并具有不同的学习策略，如 ID 保持模块（SimSwap），基于循环的工具（InSwapper），3D 形状感知身份提取器（HifiFace），融合/稳定机制（GHOST），身份一致性重建（UniFace），知识蒸馏（MobileFaceSwap），以及基于解耦的身份指导（BlendFace）。每种额外方法的数据规模均大于 1900 。

面部重演场景。在这种情况下，会生成全新的视频，其中目标个体的面部表情、动作和行为由源个体的行为驱动。该场景包括 7 种最新的方法，包括 DaGAN [61] 、TPSMM [62] 、MCNET [63] 、HyperReenact [64] 、LIA [65] 、FSRT [66] 、LivePortrait [67] 。这些方法采用了多种策略，例如基于深度的 3D 面部重建（DaGAN）、运动估计驱动的光流生成（TPSMM）、隐式身份记忆网络（MCNET）、基于超网络的图像反演（HyperReenact）、潜在空间线性运动建模（LIA）、基于 Transformer 的潜在表示学习（FSRT），以及关键点引导的可控性（LivePortrait）。每种方法生成超过 1900 个 DeepFake 视频。

谈话面孔情景。这个情景涉及使用音频输入来创建目标人物的面部视频，使嘴唇的动作和情感与音频完美对齐。为了构建这一情景，我们采用了 7 种方法，包括 SadTalker [68] 、IP-LAP [69] 、AniTalker [70] 、EDTalk [71] 、Real3D-Portrait [72] 、EchoMimic [73] 、FLOAT [74] 。这些方法在生成策略上多种多样，例如使用音频驱动的 3DMM 表情建模（SadTalker）、基于标志和外观引导的合成（IP_LAP）、通用运动表示学习（AniTalker）、组件解耦训练（EDTalk）、通过图像到平面模型的单次 3D 重建（Real3D-Portrait）、联合音频标志监督（EchoMimic），以及通过运动场匹配进行流动生成（LivePortrait）。每种方法都生成超过 2700 个 DeepFake 视频。

数据生成细节。对于面部交换和面部重演场景，我们从真实视频集中随机选择 2,000 身份对来生成 DeepFake 视频，产生 13,646 个视频，不包括之前 Celeb-DF 数据集中的视频。对于说话面孔场景，我们使用每个真实视频的第一帧作为源，并从 VoxCeleb2 数据集中随机选择 5 个音频片段 [87] 用于驱动合成，每个视频帧生成 20,279 个 DeepFake 视频。总计生成了 53,196 个 DeepFake 视频。

训练和测试划分。对于真实视频，我们遵循 Celeb-DF 的划分，选择 178 个视频。对于 DeepFake 视频，我们在脸部交换场景中每种方法随机选择 200 个视频，在脸部重新表演场景中每种方法选择 200 个视频，在说话的脸部场景中每种方法选择 300 个视频。此外，更多细节可以在项目页面找到。

我们整合了 24 种主流的 DeepFake 检测器进行评估，包括 MesoNet [4] （及其变体 MesoInception）、Xception [18] 、EfficientNet-B4 [78] 、Capsule [10] 、F3Net [11] 、CNN-Aug [79] 、FFD [26] 、SPSL [80] 、SRM [81] 、RFM [82] 、CLIP [83] 、多注意力（MATT）[24] 、RECCE [5] 、SBI [49] 、CORE [84] 、SIA [12] 、UCF [13] 、IID [85] 、LSDA [86] 、CFM [45] 、ProDet [51] 、ForensicsAdapter（ForAda）[15] ，以及 Effort [14] 。所有的 deepfake 检测器均使用其默认设置进行训练和测试。值得注意的是，大多数方法都按照 DeepfakeBench [88] 中的默认设置和预处理程序进行实施。对于最近的方法，如 CFM、ProDet、ForAda 和 Effort，我们严格按照其官方代码进行实施。表 III 中展示了一个总结。

表 I 显示了不同数据集之间评估尺度的比较。从统计数据中，我们强调我们的基准测试涉及了更多最新的检测方法，更好地反映了该领域的当前进展，并揭示了未来的潜在方向。

为了评估这些检测方法，我们使用帧级和视频级 ROC AUC 度量，这些度量通常在检测任务中采用，*e.g.*，[11], [15], [24] 。具体而言，我们为每个被测试的视频随机抽样 32 帧。对于帧级 AUC，我们获取每个被抽样帧的检测概率，并使用所有抽样帧计算 AUC 分数。对于视频级 AUC，我们首先平均每个视频内抽样帧的概率以获得视频级概率，并使用这些概率计算 AUC 分数。

为了突出这一挑战，我们使用之前的工作中常用的评估协议，对比分析了 Celeb-DF++ 与我们先前的 Celeb-DF 数据集，并使用了所有 24 种检测器 [13], [15], [24] 。在此

[5]Since detection methods often rely on a combination of multiple clues, we report only the most representative clue to define their category.

TABLE IV
帧级别 AUC（%）评估结果。所有模型都在 FF++ (HQ) 上训练并在其他数据集上测试。

Detector	Venue	Celeb-DF [21]		Celeb-DF++
		v1	v2	
MesoNet [4]	WIFS'18	49.5	53.1	48.3
MesoInception [4]	WIFS'18	69.3	65.0	64.0
Xception [18]	ICCV'19	75.4	74.0	73.7
EfficientNet-B4 [78]	ICML'19	76.7	75.0	70.6
Capsule [10]	ICASSP'19	77.3	76.5	70.9
F3Net [11]	ECCV'20	75.0	72.9	70.2
CNN-Aug [79]	CVPR'20	71.5	67.5	61.7
FFD [26]	CVPR'20	70.9	68.7	67.1
SPSL [80]	CVPR'21	80.4	72.9	68.4
SRM [81]	CVPR'21	76.5	76.0	72.6
RFM [82]	CVPR'21	79.3	76.4	70.4
MATT [24]	CVPR'21	75.5	72.0	67.0
CLIP [83]	ICML'21	85.7	82.7	75.1
RECCE [5]	CVPR'22	73.4	74.1	75.5
SBI [49]	CVPR'22	69.1	74.8	70.9
CORE [84]	CVPRW'22	71.8	74.1	70.4
SIA [12]	ECCV'22	81.5	72.6	65.0
UCF [13]	ICCV'23	81.1	77.2	72.4
IID [85]	CVPR'23	73.0	74.7	71.4
LSDA [86]	CVPR'24	74.7	73.7	70.0
CFM [45]	TIFS'24	83.6	81.1	73.3
ProDet [51]	NeurIPS'24	87.6	84.2	69.2
ForAda [15]	CVPR'25	91.4	89.9	71.7
Effort [14]	ICML'25	86.4	86.8	80.8
Average	-	76.5	74.8	69.6

TABLE V
视频级 AUC（%）评估结果。所有模型均在 FF++ (HQ) 上训练，并在其他数据集上进行测试。

Detector	Venue	Celeb-DF [21]		Celeb-DF++
		v1	v2	
MesoNet [4]	WIFS'18	49.4	53.2	48.0
MesoInception [4]	WIFS'18	74.2	70.2	68.3
Xception [18]	ICCV'19	81.0	81.6	79.1
EfficientNet-B4 [78]	ICML'19	81.5	80.8	73.8
Capsule [10]	ICASSP'19	82.9	83.4	75.4
F3Net [11]	ECCV'20	81.1	78.9	73.8
CNN-Aug [79]	CVPR'20	79.2	74.2	66.2
FFD [26]	CVPR'20	76.1	74.2	70.9
SPSL [80]	CVPR'21	85.0	79.9	74.4
SRM [81]	CVPR'21	83.6	84.0	79.4
RFM [82]	CVPR'21	85.5	82.6	74.4
MATT [24]	CVPR'21	79.2	76.0	74.8
CLIP [83]	ICML'21	89.4	88.2	79.2
RECCE [5]	CVPR'22	81.5	82.3	80.8
SBI [49]	CVPR'22	71.2	79.4	73.4
CORE [84]	CVPRW'22	76.5	80.9	74.9
SIA [12]	ECCV'22	86.3	79.1	70.0
UCF [13]	ICCV'23	86.1	83.7	76.1
IID [85]	CVPR'23	77.2	80.7	75.6
LSDA [86]	CVPR'24	79.2	77.7	72.7
CFM [45]	TIFS'24	88.9	87.5	76.5
ProDet [51]	NeurIPS'24	94.5	92.6	73.6
ForAda [15]	CVPR'25	96.9	95.7	75.1
Effort [14]	ICML'25	92.7	93.8	85.1
Average	-	81.6	80.9	73.8

设置中，检测器使用 FF++ (HQ) 数据集 [18] 进行训练，然后直接在其他数据集上进行测试。为了全面起见，我们报告了帧级和视频级的 ROC AUC 分数。

表 IV 和 V 显示了 Celeb-DF [21] 和 Celeb-DF++ 预览版（v1）和正式版（v2）的帧级和视频级评估结果。注意，Celeb-DF++ 包含比 Celeb-DF 更为多样的 DeepFake 方法。为了评估 Celeb-DF++，对于一个特定的检测方法，我们使用同一组真实视频和相应的 DeepFake 视频计算每个 DeepFake 方法的 AUC 得分，然后平均这些得分以获得最终的 AUC 表现。结果显示，与 Celeb-DF 相比，所有的 DeepFake 检测器在 Celeb-DF++ 上的性能显著下降，帧级 AUC 平均下降约为 5.2%，视频级 AUC 平均下降约为 7.1%。这些发现突显了 Celeb-DF++ 数据集的挑战性增加。

为了全面评估 DeepFake 检测方法的泛化能力，我们分别描述了三个评估协议：广义伪造评估（GF-eval）、跨质量广义伪造评估（GFQ-eval）以及跨数据集广义伪造评估（GFD-eval）。

协议 # 1 (GF-eval)。在这个协议下，所有的检测方法都仅在 Face-swap 场景中的 Celeb-DF 上进行训练，并在所有其他 DeepFake 方法上测试，包括 Face-swap, Face-reenactment, 和 Talking-face 场景。

由于一些检测方法尚未完全发布其训练代码，我们选择了八个具有代表性的方法，并在这个协议下重新训练它们。如表 VI 和表 VII 所示，这些检测方法平均仅达到大约 71.7% 和 72.1%，分别显示了它们在不同 DeepFake 方法之间的泛化能力的局限性。即使在相同场景中，不同类型 DeepFake 方法的检测性能也显著下降，突显出对方法特定伪影的敏感性。注意，Effort 实现了最高的平均帧级 AUC 为 83.0% 和视频级 AUC 为 84.4%，显示出良好的场景内性能。然而，当应用于其他场景时，其性能显著下降，表明其在跨场景检测能力上的局限性。这些发现强调了此评估协议的价值，并为提高检测方法的泛化能力提供了指导性见解。

协议 # 2 (GFQ-eval)。在现实世界的社交媒体平台中，视频内容在上传和下载过程中经常会经历不同程度的有损压缩。这些压缩可能会降低其视觉质量，从而影响 DeepFake 检测器的性能。为了模拟这种情况，我们使用 FFmpeg 工具 [89] 对 DeepFake 视频进行不同因素的压缩。具体而言，我们采用 H.264 编码标准并配置两种压缩级别：c35 和 c45，分别对应中等和高强度压缩。所有其他实验配置均遵循 GF-eval 协议。

表 VIII，IX，X 和 XI 显示了现有检测器在两种压缩级别下的帧级和视频级性能。结果显示出明显的性能下降：在 c35 压缩下，帧级性能平均下降了 3.5%，而在更强的 c45 压缩下，平均下降增加到 4.4%，而在视频级别，下降分别为 2.2% 和 7.5%。这些发现证实了更重的压缩可以更有效地遮蔽伪造痕迹，从而使检测更加困难。使用这一协议突显了现有检测方法在面对现实场景中常见压缩时的有限泛化能力。

协议 # 3 (GFD-eval)。除了在 GFQ-eval 中考虑跨质量外，我们还研究了在跨数据集场景中检测方法的性能。这既实用又具有挑战性，因为它反映了训练和测试数据来自不同来源的现实应用情况。由于所有检测器都提供经过 FF++ (HQ) 训练的权重，我们采用 **??** 节中描述的常用设置，在该设置中，每个检测器都在 FF++ (HQ) 数据集上训练，并在 Celeb-DF++ 的各个 DeepFake 方法上进行测试。

表格 XII 和表格 XIII 分别报告了帧级和视频级的 AUC 得分。由于 FF++ (HQ) 主要包含面部交换的 DeepFakes，所有检测器在 Celeb-DF 的面部交换情景中表现优于面部重演和说话面孔情景。然而，与 GF-eval 中的结果相比，由于 FF++ 和 Celeb-DF++ 之间的领域差异，该协议的结果显著下降，这突显了开发一种可以普遍适用于数据域转换的 DeepFake 检测器的必要性。

在本文中，我们引入了 Celeb-DF++，这是一个多样化且具有挑战性的 DeepFake 大规模基准，专门用于 DeepFake

TABLE VI
协议 # 1 (GF-EVAL)：帧级 AUC (%) 结果。所有检测器均使用 CELEB-DF 进行训练，并在 CELEB-DF++ 中测试其他 DEEPFAKE 方法。第一名和第二名的性能分别用粗体和 UNDERSCORE 突出显示。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	82.6	56.4	70.5	70.9	81.7	58.6	64.2	70.0	86.1	89.8	77.7	65.9	78.4	79.5	76.4	54.2	76.7	79.2	62.7	79.3	56.6	72.3
RFM [82]	84.0	57.0	71.7	70.6	84.4	62.0	63.5	67.6	81.2	90.6	74.2	67.2	77.2	77.6	71.6	48.8	71.0	76.7	65.2	73.3	54.8	71.0
CLIP [83]	84.5	74.3	72.8	77.2	86.8	71.5	66.2	65.5	76.7	80.2	72.1	54.8	70.6	72.2	62.7	46.2	67.4	62.5	63.9	70.2	53.7	69.1
SIA [12]	82.2	58.2	70.8	70.5	85.8	63.5	63.1	68.7	78.9	86.5	71.1	67.9	75.5	76.4	70.5	50.5	71.1	71.4	62.9	74.8	56.4	70.3
UCF [13]	79.8	59.2	63.3	64.9	85.0	56.9	60.3	63.6	79.8	88.3	69.2	62.9	72.2	73.2	68.4	44.9	65.6	71.5	65.6	73.2	50.2	67.5
IID [85]	80.9	51.7	70.2	67.6	84.1	54.8	61.2	67.9	82.6	91.4	75.6	59.6	73.1	74.1	72.4	49.3	69.1	75.4	58.6	76.9	46.7	68.7
ProDet [51]	82.1	57.9	76.8	74.6	83.5	65.1	70.0	66.9	79.7	90.0	74.5	68.5	75.9	75.9	64.9	56.4	61.7	72.3	71.3	77.8	54.5	71.4
Effort [14]	97.9	92.2	95.2	95.4	96.8	90.2	97.0	76.6	87.2	96.9	80.9	69.7	80.1	80.2	74.6	59.6	83.5	83.1	65.2	84.1	57.5	83.0
Average	84.3	63.4	73.9	73.0	86.0	65.3	68.2	68.4	81.5	89.2	74.4	64.6	75.4	76.1	70.2	51.2	70.8	74.0	64.4	76.2	53.8	71.7

TABLE VII
协议 # 1 (GF-EVAL): 视频级别 AUC (%) 结果。所有检测器均在 CELEB-DF 上训练，并在 CELEB-DF++ 中的其他 DEEPFAKE 方法上进行测试。表现排名前一和前二的结果用粗体和 UNDERSCORE 标出。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	86.2	52.8	70.7	70.8	85.7	56.9	66.3	70.9	90.0	92.2	81.5	66.4	82.1	83.0	75.4	47.5	75.9	79.6	62.3	75.6	46.9	72.3
RFM [82]	86.5	54.8	71.0	71.5	86.9	62.2	64.6	65.9	83.1	91.9	75.6	66.6	78.7	78.5	67.7	43.3	68.0	75.3	63.2	67.5	45.5	69.9
CLIP [83]	87.3	75.9	74.0	79.0	89.0	72.4	68.4	63.4	77.1	80.3	71.5	52.6	69.4	70.7	57.3	39.7	63.9	60.7	61.0	63.8	45.6	67.8
SIA [12]	85.5	57.6	72.0	71.3	89.1	62.8	63.7	69.0	80.7	88.6	73.1	68.2	77.1	78.0	69.2	44.9	70.5	71.1	61.9	72.0	48.9	70.2
UCF [13]	85.4	56.9	62.6	67.2	90.2	56.5	62.1	67.3	86.2	91.6	74.7	61.6	78.5	78.9	66.0	38.5	66.6	74.2	66.0	67.8	39.6	68.5
IID [85]	84.9	53.1	73.5	70.6	88.5	56.5	63.1	69.7	86.2	95.1	78.4	60.8	76.2	77.4	73.6	48.2	70.3	77.8	58.2	77.2	41.2	70.5
ProDet [51]	88.5	58.7	80.2	78.4	87.7	66.0	73.8	68.0	84.9	93.2	78.7	71.1	79.1	79.0	63.4	53.8	61.7	73.9	71.6	76.4	49.1	73.2
Effort [14]	99.0	93.7	96.7	97.1	98.2	92.1	98.2	81.8	90.7	97.6	86.1	76.7	85.5	85.2	73.0	54.6	83.9	86.8	67.5	79.3	48.8	84.4
Average	87.9	62.9	75.1	75.7	89.4	65.7	70.0	69.5	84.9	91.3	77.5	65.5	78.3	78.8	68.2	46.3	70.1	74.9	64.0	72.5	45.7	72.1

TABLE VIII
协议 # 2 (GFQ-EVAL): 帧级别的 AUC (%) 结果。所有检测器都使用 CELEB-DF 进行训练，并在 CELEB-DF++ 中采用 C35 压缩对其他 DEEPFAKE 方法进行测试。最好的前两种性能分别用粗体和 UNDERSCORE 标出。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	76.8	62.5	64.1	67.4	76.3	62.0	64.6	70.8	81.8	68.5	73.0	68.8	75.7	76.1	72.8	59.5	68.7	75.5	62.6	80.9	63.5	70.1
RFM [82]	77.4	65.0	65.1	67.3	80.4	66.1	64.3	68.7	77.6	73.9	70.1	69.5	74.5	74.2	68.7	52.0	59.0	72.9	61.6	75.5	60.5	68.8
CLIP [83]	81.6	73.3	67.2	72.2	83.9	67.0	67.9	66.9	75.9	75.3	69.8	56.8	70.6	71.2	63.1	44.2	64.8	60.1	60.6	69.3	56.2	67.5
SIA [12]	75.4	61.5	64.2	67.5	80.0	65.7	66.0	67.8	75.1	72.8	67.8	70.0	72.4	72.2	66.0	56.2	58.5	67.0	60.7	76.8	60.7	67.8
UCF [13]	70.7	58.8	57.4	60.1	75.4	57.2	59.1	58.2	71.3	69.2	60.0	63.3	64.2	65.2	61.4	44.1	52.2	65.6	56.6	70.9	50.2	61.5
IID [85]	75.5	57.5	61.0	63.5	76.1	56.5	60.5	65.5	77.3	67.2	70.0	61.0	68.0	68.5	63.4	45.4	52.2	64.8	54.9	74.3	50.4	63.5
ProDet [51]	77.5	60.6	66.1	68.3	79.6	66.3	67.1	64.8	74.4	80.5	67.6	65.7	70.5	70.3	62.7	53.6	55.9	68.2	64.4	75.1	53.5	67.3
Effort [14]	95.4	90.6	89.8	90.0	94.0	87.3	92.6	75.6	83.1	92.1	74.9	68.6	77.1	76.3	71.8	53.2	75.2	75.3	61.0	77.1	57.1	79.0
Average	78.8	66.2	66.9	69.5	80.7	66.0	67.8	67.3	77.1	74.9	69.2	65.5	71.6	71.8	66.2	51.0	60.8	68.7	60.3	75.0	56.5	68.2

检测中的通用法医分析。基于我们早期的 Celeb-DF 数据集，Celeb-DF++ 纳入了更广泛的最近 DeepFake 方法，涵盖了三种常见的伪造场景：面部交换 (FS)、面部再现 (FR) 和说话面 (TF)。每种场景包含大量高质量的伪造视频，分别使用 8 种、7 种和 7 种不同的 DeepFake 方法生成。此外，我们描述了三种用于测量检测方法普遍性的新的评估协议。与现有的基准相比，Celeb-DF++ 研究了更广泛的最新检测方法，涵盖了 19 种经典方法和自 2024 年引入的 5 种最先进的方法。实验结果强调，通用的 DeepFake 检测仍然是一项非常具有挑战性的任务，同时也展示了我们新基准的难度。

REFERENCES

[1] F. Folorunsho and B. F. Boamah, “Deepfake technology and its impact: ethical considerations, societal disruptions, and se-

TABLE IX
协议 # 2 (GFQ-EVAL)：视频级 AUC (%) 结果。所有检测器都使用 CELEB-DF 进行训练，并在 CELEB-DF++ 中用 C35 压缩测试其他 DEEPFAKE 方法。排名前一和前二的性能用加粗和 UNDERSCORE 标出。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	82.5	63.1	65.1	70.7	78.8	58.5	67.4	74.6	83.4	67.7	76.0	71.7	76.7	80.0	72.0	54.1	74.2	76.7	61.6	78.9	57.0	71.0
RFM [82]	83.0	68.0	66.2	70.3	81.9	66.7	66.7	69.3	80.4	72.3	71.4	70.4	75.0	77.2	63.5	47.1	62.1	74.8	60.6	73.0	53.8	69.2
CLIP [83]	86.5	77.1	69.8	75.0	84.3	65.7	70.7	65.2	74.9	74.3	68.5	54.2	67.7	69.7	56.7	36.9	64.7	59.3	56.7	62.0	46.9	66.0
SIA [12]	81.1	62.2	66.0	70.1	80.9	64.0	66.8	69.8	76.9	69.7	68.6	69.6	70.8	74.0	64.6	51.4	60.9	69.2	59.1	73.2	53.3	67.7
UCF [13]	81.9	64.8	60.6	68.5	82.6	57.4	66.5	70.2	82.7	73.1	70.4	67.0	73.6	78.0	66.6	44.5	70.3	75.8	60.3	72.5	46.9	68.3
IID [85]	81.2	61.9	63.9	67.1	80.3	58.3	64.3	69.5	80.4	70.4	73.3	63.9	68.9	72.4	64.2	45.6	55.4	71.3	53.8	74.9	45.5	66.0
ProDet [51]	87.1	65.7	71.9	74.2	84.6	70.1	72.9	69.9	82.1	83.6	74.6	68.9	75.1	77.3	60.5	52.7	60.0	71.6	66.0	78.1	51.5	71.4
Effort [14]	97.6	92.3	90.8	91.8	93.6	86.4	92.7	79.2	84.1	92.9	78.6	72.8	78.9	80.1	69.9	45.9	75.6	78.5	61.6	71.4	48.2	79.2
Average	85.1	69.4	69.3	73.5	83.4	65.9	71.0	71.0	80.6	75.5	72.7	67.3	73.3	76.1	64.8	47.3	65.4	72.2	60.0	73.0	50.4	69.9

TABLE X
协议 # 2 (GFQ-EVAL)：帧级 AUC (%) 结果。所有检测器都使用 CELEB-DF 训练，并在压缩为 C45 的 CELEB-DF++ 中测试其他 DEEPFAKE 方法。顶级和次顶级性能用粗体和 UNDERSCORE 标出。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	68.6	61.5	64.2	65.5	71.5	64.0	60.9	69.1	72.1	56.8	66.7	70.9	70.0	73.6	56.2	48.8	63.9	59.8	64.8	72.2	60.3	64.8
RFM [82]	68.3	63.6	62.5	64.5	74.4	65.6	63.6	69.5	71.6	60.9	67.9	72.0	70.5	73.4	60.9	44.9	55.6	58.2	60.4	68.8	61.9	64.7
CLIP [83]	74.9	69.6	62.7	68.7	77.7	60.5	66.5	71.2	74.7	67.5	72.3	62.7	72.7	73.3	59.9	42.2	59.5	49.7	62.8	65.5	61.3	65.5
SIA [12]	68.3	60.3	58.5	63.0	71.6	64.7	63.3	64.5	66.6	62.6	61.2	71.6	69.7	70.4	59.0	48.3	54.1	53.9	59.6	71.5	60.3	63.0
UCF [13]	64.8	55.9	55.9	59.0	66.5	56.7	52.4	59.3	59.9	53.7	58.0	64.2	57.2	63.2	53.3	43.6	48.2	53.4	54.4	64.5	49.3	56.8
IID [85]	70.5	61.9	61.6	65.3	73.8	62.3	58.1	66.2	70.7	54.8	68.0	65.5	63.9	70.2	55.5	41.0	43.9	51.2	57.5	68.7	53.2	61.1
ProDet [51]	69.5	58.2	61.7	64.4	73.4	65.9	61.5	66.8	68.5	66.7	65.7	67.8	70.2	68.9	61.0	45.0	56.8	59.6	68.0	65.8	58.6	64.0
Effort [14]	84.3	83.3	81.3	81.8	85.4	79.3	84.2	66.7	69.5	80.4	63.8	67.8	69.0	71.8	64.1	41.9	64.2	61.5	55.0	62.4	55.3	70.1
Average	71.1	64.3	63.6	66.5	74.3	64.9	63.8	66.7	69.2	62.9	65.5	67.8	67.9	70.6	58.7	44.5	55.8	55.9	60.3	67.4	57.5	63.8

TABLE XI
协议 # 2 (GFQ-EVAL)：视频级 AUC (%) 结果。所有检测器均使用 CELEB-DF 进行训练，并在 CELEB-DF++ 中的其他 DEEPFAKE 方法上进行测试，压缩为 C45。最佳性能和第二性能均通过粗体和 UNDERSCORE 标记。

	Face-swap (FS)							Face-reenactment (FR)							Talking-face (TF)							
Detector	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
Xception [18]	77.6	55.0	56.6	62.6	61.6	54.4	54.9	72.1	68.7	51.0	73.0	72.8	66.1	67.6	49.9	45.6	69.7	61.4	66.3	70.8	56.4	62.6
RFM [82]	76.3	60.1	57.5	62.0	63.2	57.6	56.4	63.3	65.7	54.4	66.3	66.8	64.5	65.7	58.0	42.7	58.6	56.7	62.4	69.8	54.8	61.1
CLIP [83]	83.4	69.7	63.8	67.0	73.7	58.6	66.2	68.9	69.8	60.5	68.2	60.8	66.2	68.0	46.4	39.4	60.8	52.8	60.9	59.7	45.8	62.4
SIA [12]	76.0	54.2	56.5	65.3	63.8	54.5	56.7	68.8	63.5	58.4	67.1	70.5	62.8	64.8	51.4	48.8	55.3	58.2	60.5	69.8	53.3	61.0
UCF [13]	74.6	55.4	53.9	59.9	62.2	56.1	50.7	65.8	63.2	45.4	63.6	66.1	60.4	62.0	46.5	43.8	61.0	55.6	58.5	66.1	45.8	57.9
IID [85]	80.4	56.8	58.0	63.3	64.3	52.5	55.3	66.7	66.3	52.4	70.4	68.9	57.1	62.5	45.2	40.4	44.8	52.6	58.9	69.8	49.7	58.9
ProDet [51]	80.2	57.6	61.8	63.8	68.9	61.1	61.5	69.9	68.5	70.5	69.8	69.0	67.1	67.6	51.2	45.1	61.7	59.9	69.6	74.7	56.6	64.6
Effort [14]	91.6	82.5	79.2	84.0	82.9	74.9	83.3	70.7	67.7	82.8	68.8	70.9	69.5	68.5	60.8	42.8	64.3	66.4	59.8	64.2	48.5	70.7
Average	80.0	61.4	60.9	66.0	67.6	58.7	60.6	68.3	66.7	59.4	68.4	68.2	64.2	65.8	51.2	43.6	59.5	58.0	62.1	68.1	51.4	62.4

curity threats in ai-generated media," *International Journal of Information Technology and Management Information Systems*, 2025.

[2] A. Busacca and M. A. Monaca, "Deepfake: Creation, purpose, risks," *Innovations and Economic and Social Changes due to Artificial Intelligence: The State of the Art*, 2023.

[3] M. Tahraoui, C. Krätzer, and J. Dittmann, "Defending informational sovereignty by detecting deepfakes: Risks and opportunities of an ai-based detector for deepfake-based disinformation and illegal activities," in *Weizenbaum Conference Practicing Sovereignty: Interventions for Open Digital Futures*, 2023.

[4] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *IEEE International Workshop on Information Forensics and Security*, 2018.

[5] J. Cao, C. Ma, T. Yao, S. Chen, S. Ding, and X. Yang, "End-to-end reconstruction-classification learning for face forgery detection," in *IEEE Conference on Computer Vision and Pattern*

TABLE XII
协议 # 3 (GFD-EVAL)：帧级别 AUC（%）结果。所有检测器均在 FF++ (HQ) 上进行训练，并在 CELEB-DF++ 中的所有 DEEPFAKE 方法上进行测试。表现排名前一和前二的结果分别用粗体和 UNDERSCORE 标出。

	Face-swap (FS)								Face-reenactment (FR)							Talking-face (TF)							
Detector	Celeb-DF [21]	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
MesoNet [4]	53.1	57.3	44.5	46.4	50.6	46.4	45.0	48.8	41.9	46.8	44.7	56.5	38.6	42.8	43.1	46.3	51.2	64.7	41.6	51.3	49.1	43.9	47.9
MesoInception [4]	65.0	62.7	56.3	63.5	67.6	61.7	60.0	63.9	63.6	67.4	58.9	62.7	56.8	68.3	68.0	55.5	63.3	64.0	54.4	80.0	76.8	63.1	63.8
Xception [18]	74.0	73.8	56.0	78.3	81.6	71.1	54.3	74.4	70.2	82.2	77.1	79.3	59.0	79.8	80.9	75.5	68.7	84.4	64.2	84.6	89.1	57.9	73.5
EfficientNet-B4 [78]	75.0	73.3	56.9	73.2	73.4	65.5	50.3	68.4	64.2	76.0	70.6	75.0	55.8	74.0	75.9	66.3	69.1	83.6	67.5	80.8	86.9	58.5	70.0
Capsule [10]	76.5	67.1	52.9	72.9	72.5	66.7	57.1	69.1	67.9	79.7	74.8	72.9	61.8	76.6	77.4	63.8	68.2	72.3	61.7	86.7	90.0	63.4	70.5
F3Net [11]	72.9	71.5	51.2	73.8	76.5	67.8	51.3	68.1	66.4	80.2	64.4	80.9	56.0	78.9	80.2	69.9	67.9	83.3	55.1	83.9	82.5	56.8	70.0
CNN-Aug [79]	67.5	66.9	57.0	67.5	65.4	64.3	62.4	64.3	58.7	68.5	59.0	61.1	53.6	61.4	62.4	52.2	58.0	54.3	53.7	71.9	69.9	59.1	61.8
FFD [26]	68.7	67.2	52.7	70.7	73.2	65.8	47.1	60.4	64.6	74.4	64.4	73.3	51.7	73.2	73.7	71.0	59.8	79.7	63.0	80.6	79.0	52.4	66.7
SPSL [80]	72.9	72.1	59.8	69.5	73.4	72.2	53.9	65.6	62.3	70.4	77.7	65.9	53.5	68.3	68.2	69.2	69.0	72.1	65.3	79.1	83.1	51.7	68.0
SRM [81]	76.0	71.7	55.1	76.9	82.4	69.5	58.4	68.8	71.6	79.5	74.8	80.4	64.7	79.0	79.5	67.7	68.8	80.1	58.2	84.6	88.1	59.6	72.5
RFM [82]	76.4	81.0	63.2	75.3	82.4	72.5	49.7	76.8	62.4	76.0	70.3	72.5	60.8	73.9	75.4	67.8	65.2	80.2	60.6	77.4	76.2	51.6	70.3
MATT [24]	72.0	68.2	53.1	68.8	74.4	65.6	58.1	60.1	65.5	72.7	68.6	63.2	57.1	68.8	69.9	73.7	61.5	79.5	47.2	76.3	75.8	65.2	66.6
CLIP [83]	82.7	74.2	65.3	75.5	81.3	74.6	70.7	69.4	75.3	84.1	68.4	81.1	56.6	80.7	82.3	74.9	71.5	79.3	55.0	89.6	81.5	73.4	74.9
RECCE [5]	74.1	76.9	62.3	77.2	84.4	66.7	60.0	73.1	76.3	87.4	76.6	84.3	65.9	84.9	85.4	74.7	70.2	86.7	54.1	88.7	89.6	63.8	75.6
SBI [49]	74.8	85.1	57.0	79.2	83.5	73.3	59.7	62.5	68.8	75.3	65.6	78.6	65.7	75.6	77.9	53.6	64.2	84.7	48.0	84.0	86.5	62.7	71.2
CORE [84]	74.1	73.8	55.5	75.6	83.3	72.4	53.5	70.4	63.6	76.3	73.6	76.3	56.1	76.4	76.7	69.8	63.3	86.2	52.6	80.3	86.1	51.7	70.3
SIA [12]	72.5	62.2	55.1	65.3	65.7	65.9	57.2	63.4	62.1	68.4	61.6	61.2	50.1	65.5	66.8	65.7	65.1	69.0	61.0	75.9	75.5	61.2	64.4
UCF [13]	77.2	75.5	57.0	76.8	79.9	73.6	55.6	70.5	66.9	81.8	72.8	80.3	56.7	79.0	79.8	73.0	65.4	84.4	63.9	81.0	85.3	52.7	72.2
IID [85]	74.7	76.7	58.2	75.2	79.5	70.8	57.1	68.8	69.3	83.2	71.4	80.8	62.3	80.8	81.8	68.5	63.8	80.1	50.5	83.8	82.1	57.2	71.7
LSDA [86]	73.7	77.4	60.1	78.3	78.0	64.2	51.0	71.3	63.3	78.7	66.2	76.0	61.7	75.3	76.4	63.9	67.1	78.2	58.2	77.8	85.2	55.8	69.9
CFM [45]	81.1	80.3	64.4	76.5	78.3	69.9	57.5	77.2	66.4	79.6	66.2	80.8	63.2	81.0	82.5	69.0	71.4	84.9	50.7	85.0	84.5	60.8	73.2
ProDet [51]	84.2	83.8	55.5	79.0	81.9	81.3	52.1	73.7	51.0	64.5	68.8	67.9	54.7	63.6	66.1	61.0	66.1	81.5	66.4	76.0	86.8	46.3	68.7
ForAda [15]	89.9	84.1	67.5	83.3	87.8	82.1	70.4	86.4	61.6	65.2	71.8	69.4	56.0	69.7	67.3	68.6	58.9	84.3	56.4	72.5	73.8	53.7	71.9
Effort [14]	86.8	84.6	67.0	80.8	83.7	77.1	67.6	89.7	75.8	83.1	86.2	83.3	59.6	84.7	83.7	81.9	78.1	93.8	78.2	84.9	92.3	65.8	80.4
Average	74.8	73.6	57.7	73.3	76.7	69.2	56.7	69.4	65.0	75.1	68.9	73.5	57.4	73.4	74.2	66.8	65.7	78.8	57.8	79.9	81.5	57.8	69.4

TABLE XIII
协议 # 3 (GFD-EVAL)：视频级别的 AUC（%）结果。所有检测器均在 FF++ (HQ) 上训练，并在 CELEB-DF++ 中的所有 DEEPFAKE 方法上测试。前两名的性能分别用粗体和 UNDERSCORE 突出显示。

	Face-swap (FS)								Face-reenactment (FR)							Talking-face (TF)							
Detector	Celeb-DF [21]	BlendFace [60]	GHOST [57]	HifiFace [56]	InSwapper [55]	MobileFaceSwap [59]	SimSwap [54]	UniFace [58]	DaGAN [61]	FSRT [66]	HyperReenact [64]	LIA [65]	LivePortrait [67]	MCNET [63]	TPSMM [62]	AniTalker [70]	EchoMimic [73]	EDTalk [71]	FLOAT [74]	IP-LAP [69]	Real3DPortrait [72]	SadTalker [68]	Average
MesoNet [4]	53.2	58.0	44.3	46.2	50.7	46.2	44.7	48.8	41.3	46.3	44.0	56.5	37.4	42.0	42.4	45.7	51.0	65.1	40.9	50.8	48.4	43.3	47.6
MesoInception [4]	70.2	67.0	58.7	68.0	73.3	66.1	63.5	68.4	67.8	73.0	62.4	66.7	60.2	73.8	73.5	56.9	67.2	69.0	55.3	87.0	82.3	67.4	68.1
Xception [18]	81.6	80.1	57.5	85.0	88.1	78.8	55.2	81.0	76.3	89.3	85.3	86.7	61.6	87.4	88.4	80.5	72.6	90.6	67.4	91.8	93.8	59.4	79.0
EfficientNet-B4 [78]	80.8	78.3	56.5	78.0	77.6	68.5	48.2	72.1	66.3	80.4	75.0	79.9	55.4	78.8	81.4	68.7	71.1	89.1	69.6	86.9	90.7	57.8	73.2
Capsule [10]	83.5	71.5	52.6	78.3	76.8	71.4	59.2	73.7	72.3	84.7	81.1	78.3	65.9	82.0	83.3	66.7	71.9	77.3	64.0	93.1	95.1	66.2	75.0
F3Net [11]	78.9	76.1	47.8	79.6	81.9	72.3	50.9	71.7	71.0	86.5	68.0	87.1	56.3	85.2	86.4	73.1	70.9	88.3	54.4	91.0	85.8	55.9	73.6
CNN-Aug [79]	74.2	72.8	61.1	73.5	71.0	70.1	67.3	69.6	62.9	75.2	64.3	65.7	56.5	66.1	67.5	53.7	60.9	57.0	55.2	78.3	75.4	62.9	66.4
FFD [26]	74.2	71.8	53.0	76.0	78.2	71.1	45.4	62.6	68.5	80.6	69.0	79.7	51.8	79.4	79.9	75.3	61.1	85.5	64.6	87.5	84.1	50.5	70.4
SPSL [80]	79.9	79.0	63.6	76.2	80.6	79.8	55.3	71.2	67.4	78.2	85.7	72.3	55.1	75.9	76.0	75.1	75.7	79.8	69.5	87.6	90.2	53.4	74.0
SRM [81]	84.0	77.9	56.6	83.7	88.5	77.9	62.6	74.4	79.8	89.9	81.6	89.2	69.2	88.5	88.7	74.8	73.2	87.5	64.1	94.0	95.1	63.4	79.3
RFM [82]	82.6	87.2	65.6	85.5	87.8	78.6	48.6	82.7	65.7	82.0	75.1	77.9	62.9	79.6	81.3	70.9	68.4	85.3	61.8	84.0	79.4	49.6	74.7
MATT [24]	76.0	73.9	53.3	75.4	84.2	68.3	64.2	64.5	75.9	85.0	75.2	74.7	64.7	80.9	82.7	80.7	68.3	90.4	43.4	87.7	89.4	78.2	74.4
CLIP [83]	88.2	79.4	68.3	80.3	86.6	80.1	75.6	73.6	80.1	89.7	73.1	86.0	56.9	85.9	87.4	79.2	74.5	84.3	55.4	94.6	84.6	76.0	79.1
RECCE [5]	82.3	83.6	65.3	83.6	90.0	72.6	62.4	78.4	82.7	94.1	84.0	90.7	70.0	92.2	92.1	79.7	74.4	92.2	54.6	95.3	94.2	66.0	80.9
SBI [49]	79.4	89.9	57.2	84.4	88.3	77.5	61.2	66.0	72.2	79.9	64.4	84.8	68.8	79.8	82.3	50.9	63.9	88.0	45.1	89.2	88.5	61.4	73.8
CORE [84]	80.9	79.6	56.5	81.7	88.6	79.7	54.6	75.6	67.5	82.5	79.7	82.4	57.9	83.2	83.6	73.8	66.2	91.7	53.2	87.6	91.0	51.9	75.0
SIA [12]	79.2	66.9	57.1	70.3	70.8	71.0	60.0	67.4	66.5	74.6	66.1	65.8	51.2	71.2	73.0	69.7	70.1	74.3	64.5	84.2	82.5	65.8	69.2
UCF [13]	83.7	80.6	55.4	82.1	85.0	78.5	56.6	74.8	71.2	87.7	78.9	85.9	58.0	85.5	86.0	76.5	67.6	88.8	64.2	88.3	88.7	49.6	76.1
IID [85]	80.8	82.8	60.0	80.3	84.5	76.0	58.6	73.4	74.2	89.4	77.3	86.9	65.4	87.2	88.1	71.8	66.0	85.6	49.1	89.9	86.3	57.3	76.0
LSDA [86]	77.6	81.3	60.8	82.1	81.1	66.4	50.7	74.4	65.6	82.9	69.9	80.2	64.3	79.4	80.3	65.9	69.9	81.8	58.5	82.4	88.9	55.5	72.7
CFM [45]	87.5	85.5	65.3	80.8	82.4	74.3	57.8	82.6	68.0	83.9	70.6	87.0	64.0	86.6	88.7	70.0	74.1	88.7	48.6	91.2	87.4	58.7	76.5
ProDet [51]	92.6	90.3	56.2	87.2	89.9	90.1	52.1	80.7	50.6	69.2	75.4	72.8	56.1	67.6	70.6	62.8	68.8	88.0	69.8	83.1	91.9	43.6	73.2
ForAda [15]	95.7	90.1	71.7	89.5	92.7	89.5	75.3	92.4	63.1	67.6	76.0	72.7	56.9	73.3	70.1	71.3	59.8	89.1	57.8	76.6	77.2	51.1	75.4
Effort [14]	93.8	90.7	72.1	87.0	90.2	84.7	72.0	96.4	79.4	88.1	91.5	88.1	61.5	90.0	88.5	85.0	81.8	96.8	83.2	90.2	95.8	62.2	85.0
Average	80.9	78.9	59.0	78.9	82.0	74.6	58.4	74.0	69.0	80.9	73.9	79.1	59.5	79.2	80.1	69.9	68.7	83.9	58.9	86.3	86.1	58.6	73.7

Recognition, 2022.

[6] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing AI generated fake face videos by detecting eye blinking," in *IEEE International Workshop on Information Forensics and Security*, 2018.

[7] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using in-

consistent head poses," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.
[8] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in *IEEE Winter Applications of Computer Vision Workshops*, 2019.
[9] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
[10] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.
[11] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *European Conference on Computer Vision*, 2020.
[12] K. Sun, H. Liu, T. Yao, X. Sun, S. Chen, S. Ding, and R. Ji, "An information theoretic approach for attention-driven face forgery detection," in *European Conference on Computer Vision*, 2022.
[13] Z. Yan, Y. Zhang, Y. Fan, and B. Wu, "Ucf: Uncovering common features for generalizable deepfake detection," in *IEEE International Conference on Computer Vision*, 2023.
[14] Z. Yan, J. Wang, Z. Wang, P. Jin, K.-Y. Zhang, S. Chen, T. Yao, S. Ding, B. Wu, and L. Yuan, "Orthogonal subspace decomposition for generalizable ai-generated image detection," in *International Conference on Machine Learning*, 2025.
[15] X. Cui, Y. Li, A. Luo, J. Zhou, and J. Dong, "Forensics adapter: Adapting clip for generalizable face forgery detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
[16] H. Qi, Q. Guo, F. Juefei-Xu, X. Xie, L. Ma, W. Feng, Y. Liu, and J. Zhao, "Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms," in *ACM International Conference on Multimedia*, 2020.
[17] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," *arXiv preprint arXiv:1812.08685*, 2018.
[18] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *IEEE International Conference on Computer Vision*, 2019.
[19] N. Dufour, A. Gully, P. Karlsson, A. V. Vorbyov, T. Leung, J. Childs, and C. Bregler, "Deepfakes detection dataset by google & jigsaw," 2019.
[20] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The deepfake detection challenge (dfdc) dataset," *arXiv preprint arXiv:2006.07397*, 2020.
[21] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Conference on Neural Information Processing Systems*, 2014.
[23] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Conference on Neural Information Processing Systems*, 2020.
[24] H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, "Multi-attentional deepfake detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
[25] B. Dolhansky, "The dee pfake detection challenge (dfdc) pre view dataset," *arXiv preprint arXiv:1910.08854*, 2019.
[26] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the detection of digital face manipulation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
[27] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
[28] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in *ACM International Conference on Multimedia*, 2020.
[29] 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 *IEEE International Conference on Computer Vision*, 2021.
[30] P. Kwon, J. You, G. Nam, S. Park, and G. Chae, "Kodf: A large-scale korean deepfake detection dataset," in *IEEE International Conference on Computer Vision*, 2021.
[31] T. Zhou, W. Wang, Z. Liang, and J. Shen, "Face forensics in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
[32] H. Khalid, S. Tariq, M. Kim, and S. S. Woo, "Fakeavceleb: A novel audio-video multimodal deepfake dataset," in *Conference on Neural Information Processing Systems*, 2021.
[33] S. Jia, X. Li, and S. Lyu, "Model attribution of face-swap deepfake videos," in *IEEE International Conference on Image Processing*, 2022.
[34] K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, and R. Singh, "Df-platter: Multi-face heterogeneous deepfake dataset," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023.
[35] W. Yang, X. Zhou, Z. Chen, B. Guo, Z. Ba, Z. Xia, X. Cao, and K. Ren, "Avoid-df: Audio-visual joint learning for detecting deepfake," *IEEE Transactions on Information Forensics and Security*, 2023.
[36] Z. Cai, S. Ghosh, A. P. Adatia, M. Hayat, A. Dhall, T. Gedeon, and K. Stefanov, "Av-deepfake1m: A large-scale llm-driven audio-visual deepfake dataset," in *ACM International Conference on Multimedia*, 2024.
[37] X. Xiong, P. Patel, Q. Fan, A. Wadhwa, S. Selvam, X. Guo, L. Qi, X. Liu, and R. Sengupta, "Talkingheadbench: A multimodal benchmark & analysis of talking-head deepfake detection," *arXiv preprint arXiv:2505.24866*, 2025.
[38] "FakeApp," https://www.malavida.com/en/soft/fakeapp.
[39] "DFaker," https://github.com/dfaker/df.
[40] "faceswap-GAN," https://github.com/shaoanlu/faceswap-GAN.
[41] "faceswap," https://github.com/deepfakes/faceswap.
[42] "DeepFaceLab," https://github.com/iperov/DeepFaceLab.
[43] "FaceFusion," https://github.com/facefusion/facefusion.
[44] "3D FaceSwap," https://github.com/MarekKowalski/FaceSwap.
[45] A. Luo, C. Kong, J. Huang, Y. Hu, X. Kang, and A. C. Kot, "Beyond the prior forgery knowledge: Mining critical clues for general face forgery detection," *IEEE Transactions on Information Forensics and Security*, 2023.
[46] Z. Guo, Y. Liu, J. Zhang, H. Zheng, and S. Shan, "Face forgery video detection via temporal forgery cue unraveling," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
[47] Y. Yu, R. Ni, S. Yang, Y. Ni, Y. Zhao, and A. C. Kot, "Mining generalized multi-timescale inconsistency for detecting deepfake videos," *International Journal of Computer Vision*, 2025.
[48] K. Sun, T. Yao, S. Chen, S. Ding, J. Li, and R. Ji, "Dual contrastive learning for general face forgery detection," in *AAAI conference on artificial intelligence*, 2022.
[49] K. Shiohara and T. Yamasaki, "Detecting deepfakes with self-blended images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
[50] H. Li, J. Zhou, Y. Li, B. Wu, B. Li, and J. Dong, "Freqblender: Enhancing deepfake detection by blending frequency knowledge," in *Conference on Neural Information Processing Systems*, 2024.
[51] J. Cheng, Z. Yan, Y. Zhang, Y. Luo, Z. Wang, and C. Li, "Can we leave deepfake data behind in training deepfake detector?" in *Conference on Neural Information Processing Systems*, 2024.
[52] L. Chen, Y. Zhang, Y. Song, L. Liu, and J. Wang, "Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
[53] M. Li, X. Li, K. Yu, C. Deng, H. Huang, F. Mao, H. Xue, and M. Li, "Spatio-temporal catcher: A self-supervised transformer for deepfake video detection," in *ACM International Conference on Multimedia*, 2023.
[54] R. Chen, X. Chen, B. Ni, and Y. Ge, "Simswap: An efficient framework for high fidelity face swapping," in *ACM International Conference on Multimedia*, 2020.
[55] "InSwapper," https://github.com/haofanwang/inswapper.
[56] Y. Wang, X. Chen, J. Zhu, W. Chu, Y. Tai, C. Wang, J. Li, Y. Wu, F. Huang, and R. Ji, "Hififace: 3d shape and semantic prior guided high fidelity face swapping," in *International Joint Conference on Artificial Intelligence*, 2021.
[57] A. Groshev, A. Maltseva, D. Chesakov, A. Kuznetsov, and D. Dimitrov, "Ghost—a new face swap approach for image and video domains," *IEEE Access*, 2022.

[58] C. Xu, J. Zhang, Y. Han, G. Tian, X. Zeng, Y. Tai, Y. Wang, C. Wang, and Y. Liu, "Designing one unified framework for high-fidelity face reenactment and swapping," in *European Conference on Computer Vision*, 2022.
[59] Z. Xu, Z. Hong, C. Ding, Z. Zhu, J. Han, J. Liu, and E. Ding, "Mobilefaceswap: A lightweight framework for video face swapping," in *AAAI Conference on Artificial Intelligence*, 2022.
[60] K. Shiohara, X. Yang, and T. Taketomi, "Blendface: Redesigning identity encoders for face-swapping," in *IEEE International Conference on Computer Vision*, 2023.
[61] F.-T. Hong, L. Zhang, L. Shen, and D. Xu, "Depth-aware generative adversarial network for talking head video generation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
[62] J. Zhao and H. Zhang, "Thin-plate spline motion model for image animation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
[63] F.-T. Hong and D. Xu, "Implicit identity representation conditioned memory compensation network for talking head video generation," in *IEEE International Conference on Computer Vision*, 2023.
[64] S. Bounareli, C. Tzelepis, V. Argyriou, I. Patras, and G. Tzimiropoulos, "Hyperreenact: one-shot reenactment via jointly learning to refine and retarget faces," in *IEEE International Conference on Computer Vision*, 2023.
[65] Y. Wang, D. Yang, F. Bremond, and A. Dantcheva, "Lia: Latent image animator," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
[66] A. Rochow, M. Schwarz, and S. Behnke, "Fsrt: Facial scene representation transformer for face reenactment from factorized appearance head-pose and facial expression features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2024.
[67] J. Guo, D. Zhang, X. Liu, Z. Zhong, Y. Zhang, P. Wan, and D. Zhang, "Liveportrait: Efficient portrait animation with stitching and retargeting control," *arXiv preprint arXiv:2407.03168*, 2024.
[68] W. Zhang, X. Cun, X. Wang, Y. Zhang, X. Shen, Y. Guo, Y. Shan, and F. Wang, "Sadtalker: Learning realistic 3d motion coefficients for stylized audio-driven single image talking face animation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023.
[69] W. Zhong, C. Fang, Y. Cai, P. Wei, G. Zhao, L. Lin, and G. Li, "Identity-preserving talking face generation with landmark and appearance priors," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023.
[70] T. Liu, F. Chen, S. Fan, C. Du, Q. Chen, X. Chen, and K. Yu, "Anitalker: animate vivid and diverse talking faces through identity-decoupled facial motion encoding," in *ACM International Conference on Multimedia*, 2024.
[71] S. Tan, B. Ji, M. Bi, and Y. Pan, "Edtalk: Efficient disentanglement for emotional talking head synthesis," in *European Conference on Computer Vision*, 2024.
[72] Z. Ye, T. Zhong, Y. Ren, J. Yang, W. Li, J. Huang, Z. Jiang, J. He, R. Huang, J. Liu, C. Zhang, X. Yin, Z. Ma, and Z. Zhao, "Real3d-portrait: One-shot realistic 3d talking portrait synthesis," in *International Conference on Learning Representations*, 2024.
[73] Z. Chen, J. Cao, Z. Chen, Y. Li, and C. Ma, "Echomimic: Lifelike audio-driven portrait animations through editable landmark conditions," in *AAAI Conference on Artificial Intelligence*, 2025.
[74] T. Ki, D. Min, and G. Chae, "Float: Generative motion latent flow matching for audio-driven talking portrait," in *IEEE International Conference on Computer Vision*, 2025.
[75] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2018.
[76] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics*, 2019.
[77] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
[78] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019.
[79] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Cnn-generated images are surprisingly easy to spot... for now," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
[80] H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, "Spatial-phase shallow learning: rethinking face forgery detection in frequency domain," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
[81] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing face forgery detection with high-frequency features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
[82] C. Wang and W. Deng, "Representative forgery mining for fake face detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
[83] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, 2021.
[84] Y. Ni, D. Meng, C. Yu, C. Quan, D. Ren, and Y. Zhao, "Core: Consistent representation learning for face forgery detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2022.
[85] B. Huang, Z. Wang, J. Yang, J. Ai, Q. Zou, Q. Wang, and D. Ye, "Implicit identity driven deepfake face swapping detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023.
[86] Z. Yan, Y. Luo, S. Lyu, Q. Liu, and B. Wu, "Transcending forgery specificity with latent space augmentation for generalizable deepfake detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2024.
[87] J. S. Chung, A. Nagrani, and A. Zisserman, "Voxceleb2: Deep speaker recognition," in *Conference of the International Speech Communication Association*, 2018.
[88] Z. Yan, Y. Zhang, X. Yuan, S. Lyu, and B. Wu, "Deepfakebench: A comprehensive benchmark of deepfake detection," in *Conference on Neural Information Processing Systems*, 2023.
[89] S. Tomar, "Converting video formats with ffmpeg," *Linux Journal*, 2006.