计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210900146

人脸伪造检测泛化性方法综述

董　琳[1]　黄丽清[1,2,3]　叶　锋[1,2,3]　黄添强[1,2,3]　翁　彬[1,2,3]　徐　超[1,2,3]
1 福建师范大学计算机与网络空间安全学院　福州 350117
2 数字福建大数据安全技术研究所　福州 350117
3 福建省公共服务大数据挖掘与应用工程技术研究中心　福州 350117
(donglin_c@163.com)

摘　要　深度学习技术的快速发展为深度伪造的研究提供了强有力的工具,人眼越来越难区分伪造视频图像的真假。伪造的视频图像会对社会生活造成巨大的负面影响,如:金融欺诈、假新闻传播、人身欺凌等。目前,基于深度学习的假脸检测技术在多个基准数据库(如FaceForensics++)上已经达到了较高的准确率,但在跨数据库上的检测精度远低于源数据库内的检测精度,即许多检测方法难以推广到不同的或未知的伪造类型上。专注于基于深度学习的人脸伪造检测方法泛化性研究,首先对伪造检测常用的数据库进行简单介绍和比较;其次从数据、特征和学习策略3个方面对视频图像篡改检测方法的泛化性进行分类总结和分析;最后讨论未来人脸篡改检测泛化性的发展方向和挑战。

关键词:人脸伪造检测;视频图像篡改;泛化性;媒体取证;视频图像分类
中图法分类号　TP309

Survey on Generalization Methods of Face Forgery Detection

DONG Lin[1],HUANG Li-qing[1,2,3],YE Feng[1,2,3],HUANG Tian-qiang[1,2,3],WENG Bin[1,2,3] and XU Chao[1,2,3]
1 College of Computer and Cyber Security,Fujian Normal University,Fuzhou 350117,China
2 Digital Fujian Institute of Big Data Security Technology,Fuzhou 350117,China
3 Fujian Provincial Engineering Research Center of Big Data Analysis and Application,Fuzhou 350117,China

Abstract　The rapid development of deep learning technology provides powerful tools for the research of deepfake. Forged videos and images are more and more difficult for human eyes to distinguish between real and fake. Videos and images on the internet may have a huge negative impact on social life,such as financial fraud,the spread of fake news,and personal bullying. At present,the fake face detection technology based on deep learning has reached a high accuracy on multiple benchmark databases such as FaceForensics++,but the detection accuracy on cross-databases is much lower than accuracy on the source database,that is,it is difficult for many detection methods to generalize to different types of forgeries,or unknown types of forgeries,which also motivates more scholars to focus on generalization methods. The generalization research of face forgery detection focuses on methods based on deep learning. Firstly,the commonly used datasets including real-world datasets and multi-task datasets for forgery detection are discussed and compared. Secondly,it classifies and summarizes the generalization of video and image tampering detection from three aspects:data,features,and learning strategies. The data refers to data augmentation in deepfake detection. The features include single-domain features such as frequency domain features and multi-domain features. The learning strategies consist of transfer learning,multi-task learning,meta-learning,and incremental learning. And the advantages and shortcomings of three different types are analyzed. Finally,the future development direction and challenges of face tampering detection generalization are discussed.

Keywords　Face forgery detection,Video image tampering,Generalization,Media forensics,Video image classification

到稿日期:2021-09-16　返修日期:2021-10-25
基金项目:国家重点科研专项基金(2018YFC1505805);国家自然科学基金(62072106);福建省科技计划创新战略研究项目(2020R0178,2021R0041);福建省教育厅项目(JT180078)
This work was supported by the National Key R&D Program Special Fund(2018YFC1505805),National Natural Science Foundation of China(62072106),Science and Technology Plan Innovation Strategy Research of Fujian Province,China(2020R0178,2021R0041) and Education Department Program of Fujian Province,China(JT180078).
通信作者:叶锋(yefeng@fjnu.edu.cn)

1　引言

人脸伪造指通过相关操作技术生成非真实的人脸,可以分为以下4种类型。1)全脸伪造(entire face synthesis),即合成一个完全不存在的人脸[1-3],经常使用基于生成对抗网络(Generative Adversarial Network,GAN)[4]的方法[1,5]合成。2)人脸交换(face swap),是将一个人的脸换到另一个人的脸上,这也是最常见的伪造方式。该方法包括基于计算机图形学的技术 FaceSwap[6]、基于深度学习的技术 DeepFakes[7]、FaceShifter[8]以及基于GAN的技术[9]。3)表情交换(expression swap),也称面部重现(face reenactment),是将一个人的面部表情转移到另一个人脸上,常用方法除了基于GAN的方法[10]之外,还有 Face2Face[11]和 NeuralTextures[12]方法。4)属性操作(Attribute Manipulation),即只更改一个或多个属性[13],如是否戴眼镜、肤色、年龄、有无刘海等,常使用基于GAN的技术,如 Stgan[10],StarGAN[14]和 AttGAN[15]。

最流行的伪造方法是人脸交换,统称 DeepFakes,通常以(源,目标)的图像对作为操作单元,通过将目标人脸的面部叠加到源人脸中来进行身份的改变。DeepFakes 视频最初是由自动编码器生成,但自动编码器往往无法重构细节,导致生成的假脸比较模糊。后来的 DeepFakes 视频越来越多地由GAN生成,生成的效果也明显更好。各种深度神经网络的发展也促生了更好的篡改方法,目前的许多篡改人脸技术达到了几乎以假乱真的程度。本文中使用"deepfake"指代假脸合成技术或由该技术合成的视频图像。

伪造技术的发展对于许多行业都是有益的,如影视和游戏,其可以为这些行业节约大量的成本。然而,伪造技术是一把双刃剑,其滥用不可避免地会导致安全隐患和信任危机,特别是许多公众人物的图片和视频很容易从网络上获取;而且许多开源人脸操作软件和应用程序,如 FaceApp[16]的使用,使得人脸篡改变得极其容易操作。只需要一张图像或者一个视频,任何人都可以合成假视频。第一个深度伪造视频出现于2017年,一位名人的脸被换成了一位色情演员的脸。同年,Suwajanakorn 等[17]伪造了奥巴马演讲的视频,其中奥巴马的发言与真实内容不符。深度伪造技术的不当使用,不仅会影响人们的正常生活,甚至会造成社会政治的混乱。因此,强有力的假脸检测方法是极有必要的。深度伪造检测依据检测对象的不同可分为图像检测、视频检测以及音频检测等,本文专注于分析图像和视频检测方面的泛化方法。

根据检测依据的不同,假脸检测可以分为:1)基于传统图像取证的方法,即采用传统的信号处理方法,利用频域特征和统计特征进行分析,如设备指纹、篡改痕迹、图像噪声等;2)基于生理特征的方法,如眨眼[18]、心脏跳动[19]、头部姿态变化[20-21]等;3)基于深度学习的方法,使用深度学习模型学习真实人脸和篡改后的人脸之间的差异。随着假图像合成得越来越逼真,更多的检测方法开始探索基于不同特征和先进的网络架构的深度学习方法。Yu 等[22]提出利用 GAN 指纹来检测假图像。文献[23]利用自注意力机制来获取图像的全局信息。文献[24]基于纹理不变性进行真假预测。Xception 网络[25]在假脸检测上表现出了优异的性能,许多检测器的骨干网络都使用 Xception。文献[26]设计了胶囊网络对 VGG 网络提取的特征进行分类。对于视频检测,经常使用循环神经网络[27-28]和光流[29]来表示时间信息,学习帧序列之间的不一致性等。文献[30]使用卷积视觉 transformer 进行 deepfake 视频检测。

虽然深度学习的方法取得了较高的检测准确率,但大部分检测方法只专注于提高数据库内检测精度,而没有明确考虑泛化能力。随着伪造技术的更新换代,假脸的种类在不断增加,而且大多数伪造品来源于网络,其合成技术也未知。针对每一类生成方法来训练特定的模型是耗时且代价昂贵的,因此设计出能够在不同的假脸数据集上表现优异的模型或方法才是假脸检测未来需要努力的重要方向。

目前,专注于提高泛化性的检测方法层出不穷。将各种先进方法的泛化性进行对比分析,可以更系统地了解目前假脸伪造检测泛化性的发展状况,以更好地认识不同方法的优缺点以及该方法对泛化性的作用,有助于在此基础上提出性能更好、泛化性更强的检测方案。文中重点对基于深度学习的泛化性假脸检测方法进行总结和对比,包括每种方法的特点、所用数据集和泛化性能等。检测方法的详细分类如图1所示。

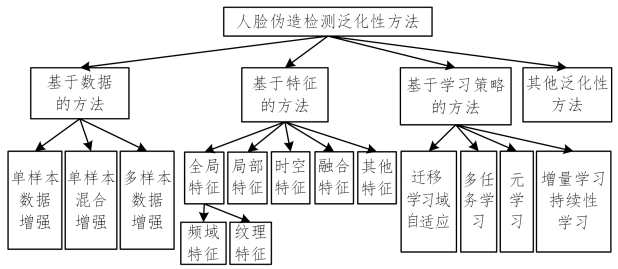

图1　人脸伪造检测泛化性方法分类

Fig. 1　Classification of generalized methods for face forgery detection

2　假脸检测数据集

随着假脸检测的发展,促生了许多公开可用的数据集,越来越多高质量数据集的出现反过来促进了假脸检测技术的快速发展和性能的持续提升。本节首先对常用的公开数据集的主要细节进行介绍,包括数据集生成方法、视频数量以及其优缺点等;然后对上述数据集进行对比分析。

2.1　低质量数据集

UADFV 数据集由 Yang 等[20]于2019年提出,是最早的公共数据集之一。其中真实视频和假视频[18]各有49个,典型分辨率为294×500像素,每个视频的平均长度大约为11 s。

Deepfake-TIMIT(TIMIT)数据集[31]于2018年被提出,是第一个 GAN 版本的 deepfake 数据集。TIMIT 数据集中的原始视频来自 VidTIMIT 数据集[32],从中选择了32位受试者,每位受试者10个视频。TIMIT 数据集中的假视频使用基于 GAN 的人脸交换算法[33]生成,每个原始视频对应合成两种视频质量的假视频,分别是低质量(Low Quality,LQ)和高质量(High Quality,HQ)。TIMIT 数据集共包含320个真实视频和640个假视频。

为了标准化假脸检测方法的评估，Rossler 等[34]于 2019 年提出了面部操作检测的基准。其中 FaceForenscis++(FF++)数据集分别使用基于计算机图形学的方法 Face2Face、FaceSwap 和基于深度学习的方法 DeepFakes、NeuralTextures 对每一个原始视频进行操作，以生成篡改视频。DeepFakes 与 FaceSwap 属于换脸伪造，Face2Face 与 NeuralTextures 属于表情伪造[35]。2020 年新增了一种换脸方法 FaceShifter，该方法首先基于对抗网络生成高保真人脸，然后还原面部遮挡。FF++数据集中的原始视频来自 YouTube，共1 000 个。为了模拟互联网上的视频压缩方式，原始视频和每种篡改方法生成的视频都包含 3 种视频质量输出：原始(Raw)，高质量(HQ，固定量化参数为 23)，低质量(LQ，固定量化参数为 40)。该数据集不仅提供了数据标签，还提供了指示一个像素是否被修改的掩码，可用于伪造检测和定位伪造区域。

图 2 给出了 FF++数据集中不同压缩率下的真实图像和对应的伪造图像。图 2 的 3 行分别对应压缩率为 0，23 和 40 的图片，其中不同方法对真实图像的操作区域并不相同。前 3 种合成方法容易产生伪影，而相对来说，NeuralTextures 和 FaceShifter 合成的图像具有更好的视觉质量。此外，可以看出，压缩率越高，图像的质量就越差，因此更难检测。

图 2 FaceForenscis++数据集示例

Fig. 2 Examples of FaceForenscis++ dataset

2.2 高质量数据集

2019 年 9 月，谷歌公司发表了 DeepFake Detection(DFD)数据集[36]，其中包含由 28 个付费演员拍摄的 363 个真实视频和 3068 个篡改视频。每个人表达不同的表情状态，如开心或愤怒，且该数据集采用了多种人脸交换方法来合成假视频。

Li 等[37]提出了一个大规模的具有挑战性的假脸视频数据集 Celeb-DF。Celeb-DF(v2)包含 890 个真实视频和 5 639 个高质量的假视频，真实视频中有 590 个视频来源于 YouTube。真实视频种类多样，具有不同性别、年龄、眉毛和发色等 40 种属性。假视频由改进的 deepfake 算法合成，相比于之前的视频而言，其整体具有更好的视觉质量。

DeepFake Detection Challenge(DFDC)数据集包含两个版本，一个是预览版本[38]，另一个是完整数据集版本[39]。最初的 DFDC 预览数据集由 Facebook 于 2019 年首次提出，用于 deepfake 检测的 Kaggle 挑战赛，使用两种面部修改算法，包括由 66 位付费演员组成的 1 131 个真实视频和 4 113 个假视频。预览数据集分布大致为 74%女性，26%男性；68%白种人，20%非裔美国人，9%东亚裔，3%南亚人。完整数据集于 2020 年公布，该数据集使用 8 种面部修改算法，其数据量更大，包含真假的完整数据集的大小超过 471 GB。

2.3 真实世界数据集

WildDeepfake[40]由 707 个 deepfake 视频中提取的 7 314 个人脸序列组成，视频全部来源于互联网，是一个小数据集。该数据集只发布人脸序列，而不是视频。

DeeperForensics-1.0(DF-1.0)[41]是一个用于真实世界人脸伪造检测的大型数据集，发布于 2020 年。其中真实视频 50 000 个，篡改视频 10 000 个，篡改视频是在 FF++数据集中的源视频的基础上合成的。为了解决低视觉质量问题，特别设计了高保真的人脸交换方法(DeepFake Variational Auto-Encoder，DF-VAE)。图 3 给出了除 FF++数据集之外的其他常用的假脸检测数据集的示例，每个数据集的示例包括两组由真实图像和其对应的篡改图像组成的图像对。

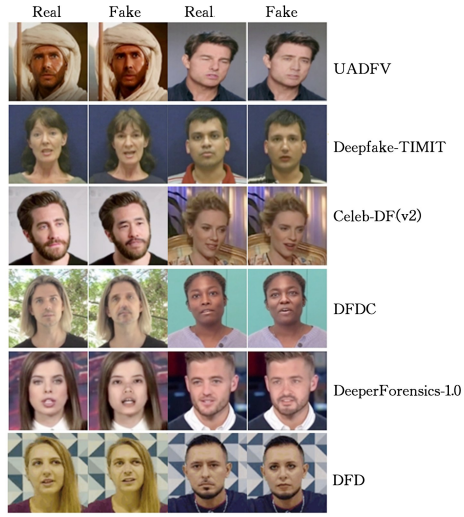

图 3 常用 deepfake 数据集示例

Fig. 3 Examples of commonly used deepfake datasets

2.4 多任务数据集

$FFIW_{10K}$[42]是第一个多人脸的大规模数据集，包含真实视频和假视频各 10 000 个，平均每个视频 12 s。该数据集每一帧的身份数量从 1 到 15 个不等，平均每帧包含 3 张人脸。假视频由 3 种人脸交换方法 DeepFaceLab[43]，FSGAN[9]，FaceSwap 之一合成，并设计了篡改视频质量评估网络来过滤低质量视频，最终的篡改视频达到了肉眼难以分辨的效果。该数据集不仅提供视频级标签，还提供了人脸级标签，可用于多示例人脸伪造检测和伪造人脸定位。

ForgeryNet[44]包含 99 630 个真实视频和 121 617 个伪造视频，假视频的合成方法高达 15 种，保证了数据集的多样性。更重要的是，该数据集在图像和视频级具有更加细粒度的标注，能满足多种伪造检测任务的需要，包括图像伪造分类、视频伪造分类、空间伪造定位和时间伪造定位。图像伪造分类又分为二分类(真/假)、分类(真/假且身份交换/假且身份保留)和 n 分类(真/15 种伪造方法)。

Le 等[45]提出了一个自然场景的多人脸数据集 OpenFo-

rensics，其中包含 45 473 个真实视频和 70 325 个假视频。该数据集提供的监督信息是人脸的边界框、分割掩码和伪造边界等，可用于多人脸伪造检测和分割。为了增强数据集的挑战，在数据上应用了各种扰动：颜色处理（亮度变化、灰度转换等），边缘操作，逐块失真（像素化、颜色量化等），图像损坏（JPEG 压缩、噪声等），卷积掩码变换（高斯模糊、锐化等），以及外部影响（雾、云、太阳等）。

2.5　数据集比较

表 1 列出了上文提到的 deepfake 检测数据集的数据量、合成方法、数据来源以及每种数据集包含的独特身份数量等。除了 OpenForensics 数据集是图像数据集之外，其他的数据集均报告其视频数量。

表 1　常见 deepfake 检测数据集

Table 1　Common deepfake detection datasets

Dataset	Resolution	Real video/image		Tampered video/image		Identity
		Quantity	Source	Quantity	Synthesis method/s	
UADFV	294×500	49	YouTube	49	FaceApp	49
DF-TIMIT	64×64 128×128	320	VidTIMIT	LQ-320 HQ-320	Face swap	43
FF++	1 280×720	raw-1 000 LQ-1 000 HQ-1 000	YouTube	YouTube	Face2Face FaceSwap DeepFakes NeuralTextures FaceShifter	−
DFD	−	363	Actor	3 068	Deepfake	28
DFDC preview	256×256	1 131	Actor	4 113	Face swap	66
DFDC	256×256	23 564	Actor	104 500	DF-128 DF-256 MM/NN NTH[46] FSGAN StyleGAN[1] Refinement Audio swaps[47]	960
Celeb-DF(v2)	256×256	890	YouTube	5 639	Improved deepfake	59+
WildDeepfake	−	3 805	Internet	3 509	Internet	−
DF-1.0	1 920×1 080	50 000	Actor	10 000	FaceSwap	100
$FFIW_{10K}$	480 or higher	10 000	YouTube	10 000	DeepFaceLab FSGAN FaceSwap	−
ForgeryNet	240 p-1 080 p	99 630	Face datasets[48-51]	121 617	15 types[8,9,43,52-59]	5 400+
OpenForensics	512×512	45 473	Google Open Images[60]	70 325	GAN[61-62]	−

由于较早的 UADFV 和 TIMIT 数据集数量较少，合成假视频的方法单一且视频质量不高，因此最近的伪造检测方法已经不再将这两种数据集作为比较的基准数据集。如图 1 和图 2 所示，相比于最新的数据集，FF++的假视频合成痕迹明显，但由于该数据集中包含不同的视频质量且采用了不同种类的合成算法，因此其仍被许多篡改检测方法作为性能评估的基准数据集之一。

早期数据集有以下缺点：1）视觉质量低；2）可见伪影；3）合成方法单一；4）数量少；5）多为室内场景。随着伪造检测方法的进步，其在低质量数据集上的检测精度已经接近饱和，因此需要更大规模的高质量数据集。高质量数据集使用了多种合成方法，通常具有更好的视觉质量，且数据集规模较大。Celeb-DF 数据集使用改进的 deepfake 算法弥补了人脸的低分辨率和颜色不一致等缺陷。DFDC 数据集没有互联网上的视频，均为演员拍摄，包含室内和室外设置，以及在各种真实照明环境下的视频，且生成数据集的方式多样化。

真实世界数据集更加接近现实的场景设置，不仅合成视频质量高，而且合成视频的方法具有多样性。DF-1.0 在样本和录像过程具有丰富的多样性，如肤色、相机角度、表情和头部姿势等，并且考虑了图像失真的情况，以适应现实世界的各种变化。虽然 DF-1.0 假视频数量较多，但其假视频是由 1 000 个独特身份的视频进行增广后得到的。而 DFDC 数据集中的每个假视频都是由独特的源和目标交换得到的，因此包含了更多的身份信息。目前许多数据集的每个视频中只存在一个身份，即一个视频中只有一张人脸。DFDC 中有 5%～10%的数据中包含多张人脸。WildDeepfake 数据集在一个场景中可能超过 10 个人。$FFIW_{10K}$ 提供了可用于多人脸检测和定位的高质量数据集。ForgeryNet 和 OpenForensics 数据集则可用于更加多样的检测任务。

3　人脸伪造检测泛化性

本节主要从数据、特征和学习策略 3 个方面介绍具有泛化性的 deepfake 检测方法。并对上述各类检测方法进行总结。基于数据、特征和学习策略的泛化方法如表 2－表 4 所列，其中包括每种方法的模型、特征、数据集及性能。每个工作的模型主要指该方法使用的骨干网络，而不是完整的模型架构。表中的实验数据取自相应论文的结果，由于部分方法的泛化实验使用的数据集较多，性能

部分仅报告了跨数据集实验的部分结果。

3.1 基于数据的泛化性检测方法

人脸篡改算法的快速发展造成了假脸的多样性。不同的合成算法专注不同的篡改问题，如人脸交换[8-9,11,63-64]，生成的假脸也会暴露与该算法相关的缺陷。对于数据驱动的检测模型来说，通过数据预处理来提高假脸检测的泛化性是有效提高模型泛化性的一个重要环节。在 deepfake 中，常见的数据处理方法是数据增强（data augmentation），也就是图像增广，其通常用于模型训练过程。数据增强通过随机改变图像产生相似但不相同的训练样本，可以增加数据集的数量和多样性，提高模型的鲁棒性，并解决样本不均衡的问题。这种对样本的随机改变可以降低模型对某些属性的依赖，避免过拟合，从而提高模型的泛化能力。

假脸检测所用的数据增强可分为单样本数据增强和多样本数据增强两大类。单样本数据增强操作只针对单张图；而多样本数据增强则在多张图片的基础上进行变换，如两张图片的随机裁剪和拼接。单样本数据增强包括两大方面：几何变换（翻转、旋转、缩放、裁剪、变形等）和像素变换（模糊、噪声、对比度、亮度、色彩抖动、擦除、填充等）。在实际训练时，经常叠加多个数据增强方法同时使用。多样本数据增强有 Mixup[65] 和 CutMix[66]。Mixup 通过对混合图像进行插值来生成新样本；而 CutMix 是分区域按像素值填充，新样本的标签按照图像混合比例来分配。上述数据增强是在原有数据的基础上进行变换，属于有监督的数据增强。无监督数据增强不在原始样本上进行处理，主要有 2 种方式：1）是生成数据分布与原始训练数据分布一致的数据，多使用基于 GAN 的方法；2）是自动增强策略，即通过模型学习出最适合当前任务的数据增强组合方式，如 AutoAugment[67] 和改进方法 RandAugment[68]。

许多方法关注特定的操作技术，如 deepfake 风格或者逼真合成[1]，但这些方法只代表了应用广泛的两个技术集合，文献[69]致力于探索一种更加通用的图像取证方法来检测 CNN 生成的图像，证明了通过细致的预处理、后处理和数据增强，特别是数据增强，一个只在特定 CNN 生成器（ProGAN）上训练的模型可以泛化到其他 CNN 生成的数据集。Zhou 等[42,70]提出了自己的高质量数据集，其泛化性实验均在不同测试集组合上进行，证明了在高质量的多样化数据集上训练的模型通常具有更好的泛化性。

3.1.1 单样本数据增强

由于计算资源和生成时间的限制，早期的 deepfake 算法只能合成固定大小的人脸图像，且必须经过仿射变换来匹配源人脸。由于扭曲后的人脸面积与周围环境不一致，这种扭曲会留下明显的伪影，而这些伪影可以作为假脸检测的线索。Li 等[71]认为利用 deepfake 算法生成假数据过于耗时且需要资源，因此直接模拟仿射变换扭曲人脸的过程来简化假数据的生成。为增加数据的多样性，改变了颜色信息：亮度、失真、对比度和锐度。此外，通过改变仿射变换的形状来模拟不同的 deepfake 后处理操作。该方法使用 4 种卷积神经网络（Convolutional Neural Network，CNN）在 UADFV 和 TIMIT 的两个视频质量上进行评估，都取得了不错的效果。但可以很明显地发现，该处理方法在 TIMIT 高质量视频上的检测性能明显下降。

为了研究数据预处理的有效性，Hulzebosch 等[72]在模型 Xception 和 ForensicTransfer[73] 上评估了不同条件下的检测性能，包括跨模型、跨数据和后处理 3 个现实场景。实验结果表明，在不同的实验场景下，3 类预处理方法（高通滤波器、共现矩阵和颜色转换）会不同程度地提高模型的性能，但没有一种单一类型的预处理方法可以在多个场景中提高模型性能。专注于检测 GAN 生成的伪造图像，Xuan 等[74]在训练阶段对真实图像和假图像进行高斯模糊或高斯噪声等图像预处理，迫使模型学习更加内在的分类特征。He 等[75]通过下采样以及超分辨率、着色和去噪等数据增强方式将真实图像进行再合成，以训练再合成器捕获各种视觉模式来提取鲁棒的特征。

3.1.2 单样本混合增强

文献[76]为训练学生模型，加入了数据增强（Cutmix、JPEG 压缩、高斯模糊和随机水平翻转），实验中虽然在未增强情况下源数据集的 AUC（Area Under Curve）（即 ROC 曲线下的面积）更高，但是目标数据集上 AUC 的增加，说明数据增强可以防止过拟合以及灾难性遗忘问题的发生。社交媒体上的人脸图像大多经过 JPEG 压缩、缩放、高斯模糊等后处理操作，Guo 等[77]发现具有混合参数的图像操作使检测器能够学习更多的鉴别特征，从而提高其泛化能力。Li 等[78]除了采用常用的数据增强，如水平翻转、旋转、缩放和 JPEG 压缩，还采用了基于人脸特征点的 Cutout[79] 方法来进行数据增强，在基于卷积 LSTM（Long Short-Term Memory）的 deepfake 视频检测中抑制模型学习特定人脸，在 DFDC 数据集训练后的模型在 FF++数据集上取得了较高的准确率。

3.1.3 多样本数据增强

实际上，并不是所有的伪造技术都是已知的或者数据充足的，因此需要根据小样本或者单样本学习数据的完整分布。Yang 等[80]提出了一种新颖的域自适应的框架，可以在特定人脸操作分布的单个示例的情况下有效训练检测器，并在检测真假脸方面取得了优越的性能。该方法首先使用预训练的 StyleGAN 模型在大量人脸上学习人脸图像的一般概率分布，然后根据给出的特定分布的单一数据调整模型权重。为生成相同分布的图像，引入了一种风格混合技术 style-mixing，将目标域的低级统计信息从目标人脸转移到随机生成的人脸，由此便可以随机生成无数的人脸。

现实世界的假视频通常会经过压缩、裁剪大小等后处理，处理后的图像可能会丢失原有的特征。因此，提出一个通用的假脸检测框架就需要考虑图像可能会遭受的预处理和后处理等操作。在 CNN 架构上的数据增强实验表明，数据增强有助于提高 CNN 模型对跨数据集检测的泛化能力，但并不是所有的数据增强都是有益的，探索每种类型的数据增强的有效性是提高检测性能的一个重要步骤[81]。通常情况下，单一的数据增强方法难以提高检测器的通用性，而组合的增强方法会显著提高其泛化性。对于不同篡改方法生成的数据，有效的数据增强组合也可能是不同的，因此需要探索自适应

的数据增强组合方法。

3.2　基于特征的泛化性检测方法

不同类型的特征，如RGB特征、频域特征和时间特征等，学习到的信息是不同的。RGB特征和频域特征常用于图像检测，而时间特征是视频检测必不可少的。本节主要分析在图像篡改检测中各种类型的特征对泛化性的不同作用，以及如何有效地融合多个特征。此外，也介绍了在视频篡改检测中如何捕捉时间信息和时空特征并融合的方法。基于特征的泛化性deepfake检测方法，可以分为基于全局特征、局部特征、时空特征和融合特征等方法。

3.2.1　全局特征

全局特征可进一步分为频域特征和纹理特征。

(1)频域特征。除了学习原始图像中的RGB特征之外，频域特征常被用作假脸检测的重要线索，有些在RGB域表现不明显的篡改信息可以很容易地在频域中被识别出来。大部分方法采用滤波器，如离散余弦变换(Discrete Cosine Transform，DCT)或者傅里叶变换(fourier transform)将RGB图像转换为频域图像来放大高频伪影；同时，为了匹配自然图像的平移不变性和局部一致性，再将特征图从频域转换到RGB颜色空间。

Qian等[82]提出频率感知图像分解和局部频率统计信息2种提取频域特征的方法，只利用频域特征就在FF++的低质量数据集上取得了优异的检测精度，但没有明确考虑泛化性。Yu等[83]专注于相机成像过程和人工智能操作过程中的通道差分图像(channel difference image)和光谱图像(spectrum image)中的内在线索，利用Octave Convolution[84]有效挖掘频域信息。针对人脸伪造过程中普遍存在的上采样操作，Liu等[85]发现上采样会造成频域显著变化，特别是在相位谱中。因此，文献[85]结合空间图像和相位谱来捕获上采样伪影，通过丢弃许多卷积层来抑制高级语义特征，以获取更丰富的浅层纹理特征。相比于文献[20，26，34，71，82，86-90]方法，该方法在从FF++到Celeb-DF的跨数据库实验上取得了更好的泛化性结果。

(2)纹理特征。纹理特征代表了图像对应物体的表面性质，如图像纹理的粗细、稠密等特性。常见的纹理特征一般使用局部二值模式(Local Binary Pattern，LBP)、灰度共生矩阵(Gray Level Co-occurrence Matrix，GLCM)和小波变换(Wavelet Transform，WT)等来提取。

文献[91]对基于纹理(LBP)和基于CNN方法的检测性能进行了比较，包括已知数据集和未知数据集，得出的结论是基于纹理和基于CNN的方法都不能应对未知攻击的挑战。然而，近年来，许多关于纹理的检测方法被陆续提出，而且其中有些方法是专门针对假脸检测泛化性的。大纹理信息对图像失真更加稳健，对来自不同GAN的人脸图像也更加具有不变性。Liu等[92]发现真实图像与合成图像在纹理上有明显的差异，因此设计了一种新的体系结构Gram-Net，将Gram矩阵作为全局描述符来捕获全局纹理特征，提高了CNN在检测GANs生成的假脸方面的鲁棒性和泛化能力。Zhao等[93]在注意力图的引导下，聚合了浅层纹理特征和高级语义特征，但FF++上训练的模型在Celeb-DF上的*AUC*值只有67.44%。上述方法说明，纹理信息可以作为假脸检测的重要特征之一，其关键在于如何提取纹理信息以及如何将纹理特征与网络进行更好的融合。

3.2.2　局部特征

篡改图像常常是由不同的图像来源合成的，因此篡改图像的块与块之间存在着不一致性，其可以通过块之间的相似度分数来度量。由于大部分的篡改检测是需要人脸拼接的，因此文献[94]利用图像统计的不一致性来检测图像中是否存在伪造边界。该方法在具有混合伪影的伪造检测中表现出了较好的泛化性，但可以看出，该方法明显不适用于没有进行图片混合的伪造，如全脸伪造等。同样地，文献[95]将视频换脸视为特殊的拼接篡改问题，利用图像分割逐像素地估计篡改区域，然后将篡改区域和人脸框的交并比作为是否发生换脸的依据。相比Xception和ResNet-50[96]等网络，该方法的跨库检测性能更好，但其实验所采用的数据库通常为低质量数据库。

Zhou等[86]基于三重态损失[97]，使用数据驱动的方式来细化隐写特征，确保来自同一图像的块在嵌入空间更接近，而来自不同图像的块之间的距离较远。Zhao等[98]认为伪造图像是来自多个源的块的组合，由此提出了一种块一致性学习方法，通过测量图像块之间的一致性来学习，其具有更好的可解释性。该方法在FF++和DFD数据集之间表现出了超越文献[94]方法的极好的泛化性。Shang等[99]关注不同层次的被操纵区域和原始区域之间的内在关系，分别提取像素关系和区域关系，以检测假图像中的不一致性。虽然在大多数伪造帧中都存在篡改区域与源图像区域的不一致性，通过检测图像中的不一致性具有一定的通用性，但是有些篡改方法合成的人脸是完全不存在的，即整张脸均是合成的，此时如果仍以块不一致性作为检测假脸的依据就会失败。

假脸篡改通常只对真实人脸的某一部分进行伪造，因此可以对人脸进行分块检测。Yu等[100]利用可分离卷积神经网络(separable convolutional neural network)来提取特征，使用图像分割将人脸图像分块，并对整体和每一块分别进行检测，最后投票得出预测类别。相比MesoNet[8]和Capsule[26]网络，该方法的泛化性最好。Chen等[101]将一张人脸图像分割为6个语义片段，包括局部(眼睛、鼻子和嘴巴)和全局(背景、人脸和原图像)语义区域，基于局部注意力分别进行预测，最后通过语义注意力模块为每个片段赋予不同的权重并联合预测其真假。其在FF++数据集内部和GAN生成图像之间的泛化性能很好，最高可达到99.95%的精度。基于deepfake的细粒度性质和空间局部性特征，Du等[102]提出了局部感知自动编码器来减小泛化差距，使用额外的像素级伪造掩膜进行正则化，以学习有意义的内在伪造表示，但其在未知伪造方法上的检测精度不超过70%。

3.2.3　时空特征

时间特征常与空间特征结合在一起，用于视频伪造检测。如何挖掘和利用操作视频的时间特征仍然是一个有待解决的问题。Sun等[103]提出了一个高效而鲁棒的框架LRNet，其

基于精确的几何特征进行时间建模来检测deepfake视频。该方法并不是将图像作为特征输入网络，而是将校准后的人脸特征点转换为特征向量输入双流循环神经网络中进行分类。在FF++和Celeb-DF数据集的测试表明，该方法的抗压缩和噪声的鲁棒性都不错，但其泛化性一般。光流可以表示视频运动的速度和方向，也常用于假脸检测中表示视频帧的运动模式。Caldelli等[104]专注于交叉伪造，提出利用光流场来识别一个视频序列的时间结构中可能存在的运动差异。文献[78]则采用经典的卷积LSTM结构来捕获时空信息。不同于图像检测的双流设计，文献[88]在网络后期对融合后的双流特征采用了双向长短期记忆(Bilateral-directional Long Short-Term Memory，Bi-directional LSTM)架构来学习视频序列之间的时间特征。在FF++数据集上训练后，相比文献[71]方法和基于Xception架构的方法，该方法在Celeb-DF上的泛化能力表现更好。文献[105]基于三维卷积神经网络建模时空特征，捕捉不同deepfake之间的相似性，增强了模型的泛化能力，但其在人脸交换方法上的泛化性较差。文献[106]结合空间和时间信息，使用动态原型，即非自然运动和时间伪影作为一种视觉解释形式，学习潜在空间中时间不一致的原型表示，然后根据测试视频的动态原型与学习的动态原型之间的相似性进行预测。Gu等[107]设计了时空不连续性学习模块(Spatial-Temporal Inconsistency Learning，STIL)分别对deepfake视频中的单帧和连续帧之间的不一致进行联合学习，以获得更全面的表示。在FF++数据集上训练、在Celeb-DF数据集上测试时，该模型取得了优于Xception和Capsule等方法的结果。

3.2.4 融合特征

单域能学习到的特征有限，因此越来越多研究者开始从多域信息中进行联合学习。文献[101]提出双流网络，分别学习RGB域的颜色信息和多频段 中的频域信息，通过高斯拉普拉斯算子在抑制高级人脸内容的同时放大伪影。针对视频换脸伪造，Han等[108]提出了基于Inception3D(I3D)网络的口部与眼部的双流检测方法，针对伪造视频中容易出现的眨眼不自然和口型拟合问题，分别检测眼部和口部篡改痕迹。但在Celeb-DF上训练的模型，在其他数据集上测试时其泛化能力一般，其原因在于不能保证网络学习到的仅为伪造痕迹，而没有受到非相关信息(语义、背景等)的影响，这也说明了基于生物信号的方法并不适用于高质量数据集之间的泛化。文献[109]采用基于3D CNNS的模型来学习伪造视频中的时空不一致性。相比CNN+LSTM架构，I3D网络能够获取更加全面的信息，其泛化性能也更好。为了避免模型过于关注特定的篡改证据，并实现鲁棒的篡改检测，Zhou等[86]提出了一种双流网络架构来捕获篡改伪影和局部噪声残差证据。第一个流是基于GoogleNet的分类流；第二个流是基于块的三重态网络，用于捕获局部噪声残差，以确保来自同一图像的两个块在嵌入空间距离更近，而来自不同图像的块之间距离更远。其在换脸应用SwapMe和FaceSwap生成的数据集上均表现出了较好的泛化性。文献[110]通过将人脸图像和UV纹理图[111-112]分别输入到Xception来提取图像特征，三维的纹理图可以确保面部信息更加完整。

在频域中，特别是在中高频带，真实人脸与操作过的人脸有明显不同。Li等[113]提出了一种频率感知鉴别特征学习框架，并特别设计了自适应频域特征生成模块，以数据驱动的方式学习更有区分度的特征。该方法对RGB特征和频域特征进行了融合来共同挖掘伪影，虽然优于同类型的大部分方法，但其泛化性差。对于伪造图像，RGB信息有助于定位异常纹理，而频率信息则放大了微妙的操纵伪影。文献[114]特别设计了一个RGB-频域注意力模块来融合RGB和频域的信息，从而丰富了局部特征。Luo等[115]提出利用高频噪声来提高泛化能力，并特别设计了一个双交叉模态注意力(dual cross-modality attention)，将SRM[116]提取的多尺度高频特征与从RGB中提取的低频纹理特征进行融合。与先进的检测方法Xception和Face X-ray相比，该方法在FF++内部跨合成方法测试以及从FF++到其他4种高质量数据集(DFD，DFDC，Celeb-DF，DF-1.0)上的跨库测试上都取得了更好的效果。Wang等[70]采用了与文献[82]中相似的处理方法，首先使用DCT将输入图像从RGB域转为频域，然后通过可学习的滤波器来获取低、中、高频带信息，最后通过反DCT转换到RGB域。该方法同样采用RGB和频域的双流融合，不同的是其融合方法受到了自注意力(self-attention)的启发，采用query-key-value机制。

3.2.5 其他特征

除了频域特征常作为辅助特征进行假脸检测之外，光流特征也常用于表示视频帧之间的运动大小和方向。文献[117]基于改进的Xception网络和多级双向LSTM，结合边缘特征和光流图组成边缘流图来补充RGB通道的特征。实验结果证明，中期特征融合方法在FF++数据集到DFDC数据集的跨域实验上取得了较好的性能。Wang等[118]提出了一种基于运动特征的视频取证方法，通过比较提取的运动模式和真实视频的运动模式的异常来识别伪造视频。该方法是完全可解释的，且是与视频内容无关的，因此其具有抗视频压缩和噪声的鲁棒性。通过对视频中多个特定空间位置的时间运动进行建模，增强不同视频内容的通用性，该方法在FF++数据集内部表现出了较好的泛化性。唇部取证也被用于deepfake视频检测。在伪造视频中，可以观察到连续帧之间的唇部运动常常是不连续的，因此可以作为伪造检测的线索。文献[119]使用预训练的唇语提取器来学习唇部运动的特征，然后基于唇部运动的不规则性，通过多尺度时间卷积网络来检测假视频。

3.3 基于学习策略的泛化性检测方法

面对层出不穷的假脸，检测方法不仅要专注于研究如何处理数据、有效利用特征，更要探索如何训练数据。由不同篡改方法合成的假脸通常具有不同的数据分布，因此如何使训练模型将一个数据集中学习的知识应用到另一个数据集是很重要的。本节介绍假脸检测方法中常用的几种学习策略，包括迁移学习、多任务学习、元学习和增量学习等，每种学习策略的特点如图4所示。

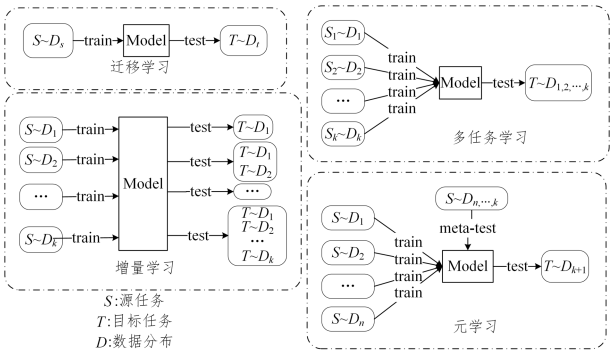

图 4　学习策略对比

Fig. 4　Comparison of learning strategies

3.3.1　迁移学习/域自适应

迁移学习(transfer learning)就是在源任务上进行训练，然后将训练好的模型在目标任务上进行微调(fine tuning)。迁移学习前后的数据分布不同，源任务与目标任务相似。域自适应(domain adaptation)是迁移学习的特殊情况，指源域和目标域的数据分布不同，且源任务和目标任务相同。在假脸检测中，源任务与目标任务相同，但是数据分布可能不同，提高泛化性的重点是如何从不同的数据中学到篡改特征。因此，可以认为迁移学习等同于域自适应。对于假脸检测而言，迁移学习的重点是如何在较少样本的情况下将模型从一个域迁移到另一个域，且模型在两个域都能表现出较好的性能。

文献[73]第一次在媒体取证的背景下，解决在特定操作下训练的网络可以扩展到类似操作上的问题，并提出了一种新的基于编码器-解码器的表示学习方法，以弱监督的方式将有用的信息约束在潜在空间，提高检测器在零样本学习(one-shot learning)和少样本学习(few-shot learning)情况下的域适应能力。这种通过微调进行域适应的方法，虽然在一定程度上提高了泛化性，但随着数据的增多，模型在旧数据上的表现会越来越差，也就是灾难性遗忘(catastrophic forgetting)[120]。文献[121]提出了一种新的迁移学习方法来解决人脸伪造检测背景下的零样本和少样本迁移问题，引入了一种基于多模态分布的训练方法，并结合空间混合增强策略，进一步提高了跨域泛化。

Jeon 等[76]提出了一种有效检测 GAN 图像的可迁移的框架 T-GD，该框架由教师模型和学生模型组成，性能优于 ForensicTransfer。Tariq 等[122]提出了基于卷积 LSTM 的残差网络，结合迁移学习的方法来提高检测的通用性。在此基础上，文献[123]分别对单域学习、合并学习和迁移学习进行实验，发现迁移学习在防御域外攻击和开放域攻击方面都表现出了令人满意的精度，证明了迁移学习在提高假脸检测方面的有效性和可行性。Lee 等[124]提出了一种基于迁移学习的残差自动编码器结构，采用多级迁移的方法，在只使用 50 个真实帧和 50 个假帧的情况下，其性能明显超越基线[25,73,125]的性能，同时在未知数据集上也实现了良好的泛化。文献[78]将模型从 DFDC 迁移到 FF++时，使用 FF++的完整数据集进行微调，在 FF++数据上实现了较好的检测性能。Yu 等[83]通过最小化不同操作方法在特征分布上的散度来降低分布偏差，以此进行域对齐。

3.3.2　多任务学习

多任务学习(multi-task learning)指一个模型同时学习多个任务，每个任务数据分布不同，多个任务通常存在相关性。在假脸检测中，多任务学习通常设计为在预测人脸为真或假的同时预测假脸的篡改区域，通过细粒度的分类来提高假脸检测的准确性。

不同于只进行二分类的检测方法，Nguyen 等[89]使用多任务学习来同时检测被操纵的图像和视频并定位被操纵的区域；同时还设计了一个编码器和一个 Y 形解码器，激活的编码特征用于分类，编码器的一支用于分割被操纵区域，另一支用于重构输入。实验结果表明，分类、分割和重构任务之间的信息共享提高了网络的整体性能。对于未知的攻击方法，该方法通过少量的数据微调来实现，但随着数据集的持续增长，其检测性能会逐渐下降。文献[2]采用基于注意力的方法定位被操纵的区域，并在预测的掩码图的基础上进行二分类，但其泛化性一般。

除了定位假脸中具体的篡改区域外，假脸检测领域的多任务学习还包括适用于多人脸伪造检测的多示例学习(multiple instance learning)。在 deepfake 中，多示例学习通常将视频看作包，人脸看作一个示例，结合示例知识和包级标签进行预测。一个假视频中可能存在多张人脸，但并不是每张人脸都被篡改过。多示例学习可以在分类的同时定位假视频中的假脸。

文献[90]基于视频级标签进行训练，通过时空编码器对时间和空间的不一致性进行建模，并对多个示例结果集成以进行预测。该方法在视频级和帧级的数据库内测试中取得了优于文献[34]等方法的结果。文献[42]首先获取多尺度短期特征和长期特征，然后基于注意力进行包特征聚合，最后基于稀疏正则化选择可能被操作的人脸示例。相较于文献[90]方法，文献[42]方法在视频级检测上具有明显的优势，并取得了从 FF++数据集到 Celeb-DF 和 DFDC Preview 数据集较好的泛化性。

3.3.3　元学习

元学习(meta-learning)是一种学会如何去学习的方法，旨在通过少量的训练样本来解决新的任务。元学习的特点使其非常适合日渐增加攻击类型的假脸检测任务。针对传统二分类伪造检测方法无法很好地检测未知攻击算法的缺点，Sun 等[126]提出了基于权重的算法，即对不同域的人脸分配不同的权重，并结合类内紧凑损失，以此来提高跨域的泛化性。该方法在不同域的数据集上表现出了泛化性，但其检测精度并不高。

3.3.4　增量学习/可持续性学习

Marra 等[127]提出了基于增量学习(incremental learning)来对 GAN 生成的图像进行检测和分类，使用 iCaRL[128]的类增量学习方法，将由不同 GAN 架构生成的图像看作一个新类别，并设计了两种基于 iCaRL 算法的多任务版本，以联合解决分类和检测问题。随着 GAN 类别的增多，与基线方法相比，该方法的检测精度最高且下降缓慢，在新类别 Style-

GAN 上也得到了较好的准确性。针对迁移学习中可能存在的灾难性遗忘问题，Jeon 等[76]设计了基于 L^2-SP 的正则化方法，将源数据集中预训练的权重作为起始点，分别对卷积层和分类层进行 L2 约束。Kim 等[129]采用知识蒸馏和持续性学习(continual learning)的方法，结合表示学习，在师生模型架构中通过学生损失、蒸馏损失和表示损失的约束来尽量减少在持续学习新的 deepfake 任务时的灾难性遗忘。

3.4 其他泛化性检测方法

目前的许多篡改方法都是对成对的图像进行换脸或表情篡改，因此会改变被操作者的身份。针对这种情况，Dong 等[130]提出了一种基于身份驱动的算法，但其需要知道待检测图像的身份，也就是需要提供额外的真实参考图像作为先验信息。该方法采用轻量级网络框架 MobileNet[131]，其只使用真实图像进行训练，来学习遮挡后的外部人脸的特征。该方法在 JPEG 压缩、改变大小、噪音等操作下的性能仍优于文献[94]方法的检测性能，表现出了优异的鲁棒性。

现实中的篡改图像在上传到网络之前，大都会进行大小调整或者 JPEG 压缩等图像处理，这种处理可能会失去假图像的判别特征。对此，Tanaka 等[132]研究了一种鲁棒的哈希方法，通过比较查询图像与参考图像之间的汉明距离来评估其相似度，进而判定查询图像是真或假。

Transformer[133]作为一种处理序列化信息的优秀模型，最初多用于自然语言处理领域(Natural Language Processing，NLP)，最近越来越多地开始用于计算机视觉(Computer Vision，CV)，特别是图像分类等。在假脸检测任务上，Transformer 不仅可以处理视频帧之间的差异，还可以对图片块进行编解码学习。Miao 等[134]使用 Transformer 结构对块间关系进行编码来提高检测的泛化性和鲁棒性，并允许在没有任何明确监督的情况下学习局部伪造特征，其在 Celeb-DF(v1)上的泛化性优于文献[94]的方法，在 Celeb-DF(v2)上的泛化性优于文献[88]的方法。文献[110]以 video transformer 为骨干网络，加入段嵌入以区分人脸特征和纹理特征。

表 2 列出了基于数据处理的泛化性方法，并介绍了每种方法所用的模型、数据增强方式、使用的数据集以及泛化性能。Wang 等[69]报告了使用不同的数据增强方法，包括无增强、模糊、JPEG 压缩、模糊与 JPEG 压缩(0.5)、模糊与 JPEG 压缩(0.1)时，模型在 10 种测试数据集下的平均精度均值(mean Average Precision，mAP)，训练集为 ProGAN。文献[72]分别在一阶导数滤波器、三阶导数滤波器、共现矩阵和颜色空间转换这 4 种数据增强的条件下测试了 Xception 和 ForensicTransfer 模型在 $StyleGAN_{CAHQ}$ 到 $StyleGAN_{FFHQ}$ 的泛化性能，表中报告的结果只检测伪造数据而不包括真实数据的准确率。Xuan 等[74]将采用高斯模糊或高斯噪声进行处理的训练集得到的模型分别记为 M_{GB} 和 M_{GN}，通过总体的准确率(Accuracy，ACC)、真阳率(True Positive Rate，TPR)和假阳率(True Negative Rate，TNR)来衡量模型性能。He 等[75]在进行跨域检测时，使用 ProGAN 或 StyleGAN 在 CelebA-HQ 上生成的数据作为训练集，将 StyleGAN 或 StyleGAN2 在 FFHQ 上生成的数据作为测试集。表中列出了在 3 种不同的重新合成方法下，每个模型在测试数据集上的平均分类准确率。在同时应用 4 种数据增强的情况下，文献[76]的模型从 PGGAN 到 StyleGAN2 的泛化 *AUC* 值为 98.13%。Guo 等[77]使用自己构建的数据集 HFF 和公开数据集 FF++进行实验，在测试其他类型的人脸图像，如 GB-mix，MED-mix，GC-mix，JP2-mix 和 SC-mix 时，其平均准确率为 95.17%。文献[78]展示了在 FF++这 3 种视频质量下的 4 种算法生成的伪造视频和真实视频混合数据集上的检测准确率。文献[80]在少样本学习的情况下，在 DF 和 ProGAN 数据集上的平均检测精度分别为 93.4%和 62.1%。Fung 等[135]分别在 FF++，UADFV 和 Celeb-DF 数据集上训练模型，并在其他 2 种数据集上进行泛化性测试，表中报告了模型在另外两种未知数据集上的 AUC 值。

表 2 基于数据的泛化性方法

Table 2 Data-based generalization methods

Work	Model	Data augmentation method/s	Dataset	Performance/%
Wang et al.[69]	ResNet-50	Gaussian blur JPEG blur	ProGAN StyleGAN，BigGAN CycleGAN，StarGAN GauGAN，CRN[136] IMLE[137]，SITD[138] SAN[139]，DF	mAP No aug：90.1 Blur only：84.4 JPEG only：93.0 Blur+JPEG(0.5)：90.8 Blur+JPEG(0.1)：92.6
Hulzebosch et al.[72]	Xception(X) ForensicTransfer (FT)	High-pass filtering Co-occurrence matrix Color conversion	$StyleGAN_{CAHQ}$→ $StyleGAN_{FFHQ}$	ACC X：37.8/62.1/31.2/44.7 FT：90.2/87.8/32.4/46.8
Xuan et al.[74]	CNN	Gaussian blur Gaussian noise	PGGAN[5]→ WGAN-GP[140]	ACC/TPR/TNR MGB：68.07/93.08/43.06 MGN：68.28/94.65/41.91
He et al.[75]	Residual network	Super resolution(SR) Colorization(C) Gaussian noise(D)	{ProGAN，StyleGAN→StyleGAN， StyleGAN2[53]}	ACC SR：78.3 SR+C：87.0 SR+D：87.4
Jeon et al.[76]	EffificientNet-b0	Intra-class Cutmix JPEG compression Gaussian blur Random horizontal flip	PGGAN→ StyleGAN2	AUC98.13

（续表）

Work	Model	Data augmentation method/s	Dataset	Performance/%
Guo et al.[77]	AMTEN	JPEG compression Scaling Gaussian blur Mean filtering Median filtering	HFF FF++	ACC 95.17
Li et al.[78]	CNN+LSTM	Cutout Horizontal flip Rotation Scaling JPEG compression	DFDC→FF++	ACC Raw:99.53 C23:97.79 C40:94.08
Yang et al.[80]	ResNet-50	Style-mixing	DF ProGAN[5]	Average Precision 93.4 62.1
Fung et al.[135]	DeepFakeUCL	Random crop Color jittering Gray scale Flip	FF++ UADFV Celeb-DF	AUC 67.5/56.8 56.2/64.8 58.9/85.6

表3列出了基于特征的泛化性检测方法，包括每种方法使用的模型、特点、泛化性实验所用的数据集和泛化性能。文献[83]将表3中的10种数据集重新组合为8组训练集和测试集不相交的组合，在8组实验数据集上分别测试模型性能。文献[86]在SwapMe和FaceSwap数据集上分别进行训练，并在另一个数据集上测试泛化性。文献[88]分别给出了该模型在帧级和视频级上对FF++到Celeb-DF上的AUC值和正确接受率（True Acceptance Rate，TAR）。文献[92]在CelebA-HQ上应用StyleGAN和PGGAN生成数据集，并测试了2个数据集在不同数据增强下的泛化性，表中给出了每组数据集下的平均准确率。Li等[94]在FF++数据集内部进行泛化性测试，并给出了Face X-ray在其他3种数据集上测试的AUC值。文献[100]对FF++数据库和其他数据库均进行了泛化性实验，表中只给出了FF++数据集到其他数据集的泛化性结果。Chen等[101]在FF++数据集内部和基于GAN的数据集上均进行了泛化性测试，但表中只列出了在FF++数据集上的实验结果。以FS泛化性为例，该模型先在DF，F2F和NT上进行训练，然后在FS上进行测试。文献[104]在FF++的c40数据集上进行了跨伪造方法的实验，并以直方图的形式进行对比，具体数值未给出。文献[105]对FF++数据集中的5种方法进行组合并测试其单类精度，由于数据复杂，表中并未给出结果。文献[117]中的模型XceptionNet*在中期进行双流融合时实现了最好的效果，表中给出了采用中期融合时模型对DFDC-mini真实数据、伪造数据和整体数据的泛化性结果。文献[118]对FF++数据集的4类数据进行了泛化性实验，在单一方法上进行训练，然后在其他3种数据集上进行测试。

表3　基于特征的泛化性方法

Table 3　Feature-based generalization methods

Work	Model	Feature/s	Dataset	Performance/%
Zhou et al.[43]	CNN	Multi-instance learning Attention	FF++→Celeb-DF/ DFDC preview	AUC 74.1/78.3
Wang et al.[70]	M2TR	Frequency domain Mutil-scale transformer	FF++→Celeb-DF/SR-DF SR-DF→FF++/Celeb-DF	ACC 65.7/62.6 77.9/80.7
Li et al.[78]	CNN+LSTM	Spatio-temporal information	DFDC→FF++	ACC Raw:99.53 C23:97.79 C40:94.08
Yu et al.[83]	OctResNet-34	Attention-based feature fusion	StyleGAN StyleGAN2 ExperGAN GANimation[141] HomoInterpGAN[142] CycleGAN StarGAN STGAN FaceSwap DeepFakes	ACC 97.2/98.92/98.59/ 97.65/97.26/98.13/ 98.09/97.19
Liu et al.[85]	Xception	Phase spectrum	FF++→Celeb-DF	AUC 76.88
Zhou et al.[86]	GoogeLeNet[143]	Patch inconsistency	SwapMe FaceSwap	AUC 82.9 85.4

(续表)

Work	Model	Feature/s	Dataset	Performance/%
Masi et al.[88]	CNN+LSTM	Frequency domain Color domain	FF++→Celeb-DF	AUC/$TAR_{10\%}$ Frame level:73.41/32.22 Video level:76.65/39.70
Liu et al.[92]	Gram-Net	Global texture	StyleGAN→PGGAN PGGAN→StyleGAN	ACC 89.26 87.52
Zhao et al.[93]	EffificientNet-b4	Texture enhancement Attention	FF++→DFDC	ACC 67.44
Li et al.[94]	HRNet	Blending boundary	DF F2F FS NT	AUC 97.64/98.00/97.97 99.03/98.64/98.14 99.10/98.16/96.66 99.27/98.43/97.85 99.17/98.57/98.21
Hu et al.[95]	DNN	Splicing tampering	DFD→TIMIT/FF++(c0)/ FF++(c23)/FFW[91]	Average error rate 15.9/7.9/11.4/20.2
Zhao et al.[98]	ResNet-34	Patch-wise consistency	FF++→DFD	AUC 99.07
Shang et al.[99]	PRRNet	Pixel correlation Regional correlation	FF++(raw→HQ/LQ)	ACC 84.70/56.42 88.73/50.83 91.75/63.03 77.72/53.80
Yu et al.[100]	CNN	Block detection	FF++→ DeepFaceLab/ StyleGAN	ACC 80.5/84.9 ACC 80/81
Chen et al.[101]	VGG-19	Global and local features	FS DF F2F NT	ACC 94.47 99.95 99.94 99.84
Du et al.[102]	autoencoder	Local feature	F2F→FS StarGAN→Glow G&L[144]→ContextAtten[145]	ACC 68.06 62.11 64.42
Sun et al.[103]	LRNet	Geometric feature Spatial feature	FF++→UADFV/ Celeb-DF	AUC 98.5/56.9
Caldelli et al.[104]	CNN	Optical flow	FF++	—
Ganiyusufoglua et al.[105]	CNN+RNN	Spatio-temporal feature	FF++	—
Trinh et al.[106]	DPNet	Spatio-temporal feature Inconsistency	Celeb-DF→DFD/DF-1.0/ Celeb-DF	AUC 92.44/90.80/68.20 TAR 76.21/75.67/25.88 EER 16.21/17.30/37.08
Gu et al.[107]	CNN	Spatio-temporal inconsistency	FF++→Celeb-DF	AUC 75.58
Han et al.[108]	Inception3D	Eye and mouth area	FF++→DFDC/DFD/FF++	ACC 52.29/59.63/73.39 60.45/63.64/68.64
Xing et al.[109]	3D CNNS	RGB Optical flow	FF++→Celeb-DF	ACC:83 AUC:88
Khan et al.[110]	Video transformer	UV texture map	{FS,DF}→F2F, NT,DFD,DFDC	ACC 88.57 51.42 93.27 91.69
Chen et al.[114]	CNN	Patch similarity RGB-Frequency domain attention	FF++→Celeb-DF/ DFDC/DFD	AUC 78.62/76.53/89.24 EER 29.67/32.41/20.32
Luo et al.[115]	Xception	High frequency noise Attention	FF++→DFD/DFDC/ Celeb-DF/DF-1.0	AUC 91.9/79.7/79.4/73.8
Chintha et al.[117]	XceptionNet	Edges and optical flow	FF++→ DFDC-mini	ACC 73.33(real)/86.81(fake)/ 81.29(all)
Wang et al.[118]	CNN	Local motion consistency	DF F2F FS NT	ACC 92.15/93.45/95.85 84.25/76.75/84.95 70.30/64.85/81.65 76.20/65.15/77.85
Haliassos et al.[119]	ResNet-18	Mouth movement	FF++→Celeb-DF(v2)/ DFDC/DF-1.0	AUC 82.4/73.5/97.6

表4对基于学习策略的检测方法进行了总结,包括每种方法使用的模型、相关的学习策略、使用的数据集和泛化性能。文献[73]合成了自己的数据集,其中Inpainting数据集是在ImageNet[146]数据集上应用不同的图像修复算法[144-145]分别得到的源数据集和目标数据集。Jeon等[76]分别以非人脸图像Bedroom和Bird为源数据集训练模型,并在其他基于GAN的人脸数据集上进行测试。Nguyen等[89]报告了模型在面对未知攻击时的分类准确率和等错误率(Equal Error Rate,EER)。文献[145]采用了一级迁移和二级迁移的学习策略,表中列出了采用二级迁移策略时模型在新数据集上的准确率。基于增量学习的方法[127],在不同的内存预算的情况下,测试了检测器对5组GAN图像的检测性能。根据正则化参数的不同,检测器分为两个模型MT-SC和MT-MC,表中列出了两个模型在内存预算依次为0,128,256,512,1024和∞时,对最后一种GAN训练后的检测准确率。Kim等[129]在基于GAN图像的任务1中训练模型CoReD,然后运用持续性学习的策略学习任务2、任务3、任务4。表中列出了CoReD在学习4个任务之后在FF++两种压缩质量数据集上的准确率。

表4　基于学习策略的泛化性方法

Table 4　Learning strategies-based generalization methods

Work	Model	Generalization Strategy	Dataset	Performance/%
Dang et al.[2]	Xception	Multi-task learning	DFFD→UADFV/Celeb-DF UADFV→Celeb-DF {UADFV,DFFD}→Celeb-DF	AUC 84.2/64.4 57.1 71.2
Zhou et al.[42]	CNN	Multiple instance learning	FF++→DFDC Preview,Celeb-DF	AUC 74.1 78.3
Cozzolino et al.[73]	autoencoder	Few-shot learning Transfer learning	ProGAN→CycleGAN CycleGAN→StyleGAN StarGAN→Glow[147] Inpainting Face2Face→FaceSwap	ACC 85.00 90.53 82.05 70.62 72.57
Jeon et al.[76]	EfficientNet-b0	Transfer learning	BiedroomPGGAN/ StarGAN/StyleGAN/ StyleGAN2/Bird Bird→PGGAN/ StarGAN/StyleGAN/ StyleGAN2/Bedroom	AUC 86.25/88.08/90.47/90.25/90.15 87.80/78.32/98.49/98.49/97.63
Nguyen et al.[89]	autoencoder	Multi-task learning	{DF,F2F}→FS	ACC:83.71 EER:15.07
Aneja et al.[121]	ResNet-18	Zero/Few-shot learning	FF++→DFD/AIF/ Dessa/Celeb-DF	ACC 81.21/60.79/74.28/68.83
Tariq et al.[123]	CNN+LSTM (CLRNet)	Merged learning Transfer learning	{NT→DF,FS,F2F}→DFW	F1 score 93.86±0.2
Lee et al.[124]	autoencoder	Transfer learning	{DF→FS}→F2F {DF→F2F}→FS {F2F→DF}→FS {F2F→FS}→DF {FS→DF}→F2F {FS→F2F}→DF	ACC 89.4 95.9 98.3 97.1 87.8 95.5
Sun et al.[126]	CNN	Meta-learning	{F2F,FS,NT}→DF {DF,FS,NT}→F2F {DF,F2F,NT}→FS {DF,F2F,FS}→NT	ACC(c23/c40) 85.6/69.1 65.6/65.7 54.9/62.5 65.3/58.5
Marra et al.[127]	CNN	Incremental learning	CycleGAN ProGAN-GP(256) ProGAN(1024) Glow StarGAN	ACC 69.15/92.80/92.42/96.37/97.22/97.76 67.71/86.47/93.50/95.36/94.5/ 99.37
Kim et al.[129]	CNN	Continual learning Representation learning Knowledge distillation	ProGAN StyleGAN StarGAN CelebA-HQ FF++	ACC HQ:91.77 LQ:80.55 Average:86.16

4 实验分析

本节在统一标准下对假脸检测常用的分类网络进行泛化性实验，通过对不同网络检测性能的比较，观察不同方法的检测性能，便于选择泛化性更好的检测框架。用于比较的模型有 ResNet-50，Xception 和 EfficientNet-b0，数据集使用 FF++。ResNet-50 采用经典的残差网络，而 Xception 被认为是目前最好的图像分类模型。EfficientNet-b0 于 2019 年被提出后，越来越多地被用于假脸检测的骨干网络中来提取特征，该网络表现出了并不逊色于 Xception 的性能。由于 FF++中有多种篡改方法的数据集，因此在 FF++的 5 种不同篡改方法之间也进行了泛化性实验。

4.1 实验设置

在 FF++内部数据集的实验中，对 FF++中的 5 种篡改方法生成的数据分别进行了检测，但数据集只包含低质量真实视频和篡改视频，即压缩率为 40 的视频。为了减少帧距过短造成数据集中图像的相似度太高的问题，其中每个真实视频和假视频使用 MTCNN 库每隔 5 帧提取一次人脸，每个视频最多提取 20 帧。数据集分为训练集、验证集和测试集，根据标准比例 720∶140∶140 进行划分。将非人脸图片筛选之后，最终得到的数据集数量如表 5 所列。在训练或测试时，将真实数据分别与每种篡改方法的假数据组合。

表 5 FF++数据集实验数据

Table 5 Experimental data of FF++ dataset

Dataset	Number of videos	Number of images		
		Training set	Validation set	Test set
Real	1000	14316	2796	2796
DF	1000	14297	2790	2793
FS	1000	14326	2793	2787
F2F	1000	14293	2781	2796
NT	1000	14327	2797	2796
FSh	1000	14296	2796	2796

所有模型都在 ImageNet 数据集上进行预训练，预训练参数作为初始化参数，且在之后的训练过程中不冻结参数。裁剪后的人脸在输入模型时统一调整大小为 299×299，然后进行归一化，其均值和方差都为 0.5。模型训练时，使用 Adam 优化器，批大小为 32，初始学习率为 0.001，每经过 5 轮更新为原来的 0.5 倍，损失函数使用交叉熵损失。每个模型均训练 50 轮，选择验证集准确度最高轮次的模型进行测试。

4.2 基于 FF++数据集的假脸检测泛化性

目前，FF++数据集共包含 5 种篡改方法，分别是 DeepFakes(DF)，FaceSwap(FS)，Face2Face(F2F)，NeuralTextures(NT)和 FaceShifter(FSh)。FF++作为经典的假脸检测数据集，其数量较多且合成方法多样。在研究 FF++数据集内部的泛化性时，训练与测试的过程中均不采用数据增强。

上述 3 个模型在相同的设置中进行了 FF++数据集不同篡改方法之间的泛化性实验。每个模型在每种方法上分别进行训练，然后直接在其他数据集上进行零样本测试，评估结果如表 6 所列。训练集与测试集相同时的检测结果加粗标示，3 个模型在同一数据组设置下平均准确率最高的用下划线标出。

表 6 基于 FF++数据集的泛化(准确率)

Table 6 Generalization accuracy based on FF++ dataset

(单位：%)

Model	Training set	Test set					
		DF	FS	F2F	NT	FSh	Average
Xception	DF	**93.88**	54.00	51.31	52.99	55.47	61.53
	FS	56.99	**89.47**	50.36	48.00	52.41	59.45
	F2F	55.16	51.96	**86.25**	54.31	51.22	59.78
	NT	53.37	49.17	51.95	**97.41**	77.07	65.79
	FSh	57.09	51.23	50.70	57.17	**93.38**	61.91
ResNet-50	DF	**90.34**	54.18	51.25	52.99	55.97	60.95
	FS	62.64	**85.00**	51.56	47.62	53.11	59.99
	F2F	54.54	53.05	**81.89**	51.54	50.52	58.31
	NT	51.17	49.38	51.47	**94.67**	76.14	64.57
	FSh	57.09	51.60	50.70	60.68	**93.38**	62.69
Efficient-Net-b0	DF	**94.58**	55.56	51.34	52.63	55.19	61.86
	FS	61.82	**92.14**	50.59	48.67	51.11	60.87
	F2F	56.00	52.37	**85.89**	52.67	51.02	59.59
	NT	50.89	48.74	50.88	**96.84**	76.38	64.75
	FSh	57.99	52.34	51.18	60.68	**93.30**	63.10

可以看到，Xception，ResNet-50 和 EfficientNet-b0 这 3 个模型在已知数据集上测试时均可以得到不错的结果。但不可避免地，在未知数据集上进行训练时，其准确率通常较差，甚至会出现低于 50%的准确率。在没有数据增强和特征处理等方法的辅助下，这几种常用的假脸检测分类器的泛化性较差，但这为提出更好的检测方案提供了对比基线。基于 3 种网络训练的模型中，当源数据集为 NT 时，3 种检测模型在其他篡改方法上的检测精度都是最高的，平均分别为 65.79%，64.57%和 64.75%。通过比较不同模型在同一训练集下的准确率可以发现，ResNet-50 的性能普遍较差，只有当模型在 NT 或 F2F 上训练后，Xception 的泛化性才能优于 EfficientNet-b0，而在其他情况下，EfficientNet-b0 保持最高的平均准确率。除此之外，EfficientNet-b0 具有比 Xception 更少的参数，训练时间也更短。

4.3 基于数据增强的假脸检测泛化性

通过上节实验可以发现，EfficientNet-b0 在域内和域外的检测性能均优于 Xception 和 ResNet-50，因此本节选择 EfficientNet-b0 进行测试。基于 EfficientNet-b0 模型的数据增强实验，目的在于研究各种数据增强方法对模型泛化能力的影响以及探索有益于提高 deepfake 检测泛化性和鲁棒性的数据增强方法。本实验所用数据集与训练参数均遵循 4.1 节的设置，不同之处在于训练时使用了数据增强。所用的数据增强方法有竖直翻转(Vertical Flip，VF)、随机旋转(Random Rotation，RR)、颜色变换(亮度 Bright，B)以及组合数据增强(VF+RR+B)等。基于 EfficientNet-b0 的数据增强实验结果见表 7，粗体表示数据增强后准确率提升的结果。从表 7 中的实验可以看出，只应用单个数据增强就能在一定程度上提高模型在 FF++数据集之间的泛化性。例如，应用亮度变化后，EfficientNet-b0 模型在 F2F 数据库内测试中的精度提高了 2.27%，在 DF 到 F2F 的跨库测试中的精度提高了 2.06%。相较于几何变换中的翻转和旋转，改变亮度更有益于提升模型性能。而将上述 3 种数据增强组合起来训练的模型的检测性能相较于基于单个数据增强的训练模型，性能有进一步的提高，特别是当源数据集为 FaceShifter 时，该模型

在其他数据集上的准确率均有显著提升。但也可以看出，应用数据增强后，并不能全面提高检测器在所有数据集上的准确率，甚至可能会降低其在部分数据集上的检测性能。

表 7　基于 EfficientNet-b0 数据增强（准确率）

Table 7　Accuracy of EfficientNet-b0 based data augmentation

（单位：%）

Data augmentation	Training set	Test set				
		DF	FS	F2F	NT	FSh
VF	DF	93.04	**56.19**	51.31	51.14	54.76
	FS	**63.68**	**92.44**	**50.66**	**48.70**	**52.38**
	F2F	**55.22**	**54.25**	**86.05**	**53.81**	**52.38**
	NT	50.74	**48.85**	50.50	**97.21**	75.72
	FSh	57.52	51.41	**51.74**	**61.12**	92.87
RR	DF	**94.76**	**56.94**	50.99	51.23	54.24
	FS	60.46	90.78	49.95	**49.11**	**51.67**
	F2F	55.11	51.60	**85.94**	52.13	50.00
	NT	**50.92**	**49.06**	**51.59**	93.37	76.00
	FSh	57.11	**53.22**	**51.40**	59.14	**93.92**
B	DF	93.97	**55.56**	**53.40**	51.95	**55.47**
	FS	59.19	92.01	**51.22**	**49.02**	**51.65**
	F2F	55.50	**53.16**	**88.16**	**53.20**	**51.43**
	NT	**50.98**	**48.95**	**51.16**	95.73	75.57
	FSh	**58.65**	51.98	50.98	**59.59**	**93.35**
VF＋RR＋B	DF	94.53	**56.44**	**51.82**	50.72	**55.24**
	FS	**63.25**	90.10	**50.77**	48.66	**53.38**
	F2F	55.38	**52.55**	85.03	**53.15**	**53.32**
	NT	**51.14**	**49.90**	**51.13**	**97.34**	76.23
	FSh	**58.65**	**53.05**	**52.04**	**60.86**	**93.47**

从表 7 中可以得出以下结论：1）数据增强在提升检测器在某些数据集上准确率的同时，也会降低检测器在其他数据集上的准确率；2）特定的篡改方法适合特定的数据增强方法；3）整体而言，组合数据增强比单一的数据增强更能提高模型在假脸检测上的泛化性。

5　展望

目前，许多优秀的检测方法已经可以在单个数据集内达到很高的精度，这主要得益于训练集与测试集是同分布，模型在大量数据上训练之后会学到特定操作方法的特征。当训练数据与测试数据不同分布时，检测器的性能就会显著下降，说明检测器无法对抗不同数据分布的假脸视频图像。面对生成技术的快速发展，合成方法在不断变化，合成假脸的质量也在不断提升。

如何提高检测器在真实世界中的泛化性以及鲁棒性都是未来的一大挑战。这就需要考虑现实世界中可能会发生的影响检测的因素，如图像压缩、裁剪、分辨率变化等。当前的许多检测模型都没有考虑在不同压缩率和不同视频分辨率下的泛化问题，而不同的压缩率[148]和分辨率在视频图像中是很常见的。在现实场景中，互联网上的图像或视频通常会经过压缩，而通过压缩之后，假视频中的伪造痕迹会变淡，甚至可能会丢失用于判别真假图像的重要特征，从而增加检测难度。

本文对泛化性方法的总结集中于检测方法，而主动防御也是阻止 deepfake 泛滥的一种重要方式。主动防御通常采用添加水印或者噪声等预处理的方式，使图片或视频难以被篡改或降低篡改后的效果，防止生成以假乱真的伪造品。除了图片伪造和视频伪造之外，语音伪造和全身伪造也是常见的伪造方式，但针对语音伪造和全身伪造的研究并不多。当前针对语音伪造的检测开始从传统信号处理深入到深度学习方法，但更多的仍是针对特定的攻击类型。文献[149]通过在线频率掩码增强和大边距损失函数（Large Margin Cosine Loss Function，LMCL）迫使神经网络学习更强大的特征嵌入，从而提高其泛化性。

一些用于 deepfake 检测的传统技术越来越多地被深度学习技术代替，但传统技术仍有其独特的优势。例如，Jia 等[150]基于 mini-batch 策略提出了一个深度数字水印框架来提高模型面对 JPEG 压缩的鲁棒性。相比于卷积神经网络提取的特征，传统的数模算法提取的特征更加具有可解释性。本文主要分析基于深度学习方法的帧内和帧间检测方法，Bao 等[151]提出了基于视频完整性的检测方法，包括基于区块链、数字水印和视频指纹的检测方法。未来需要的不仅是更通用和鲁棒的伪造检测方法，还需要可以满足多样化检测的数据集。

结束语　基于 GAN 的篡改技术的进步使假视频在互联网上泛滥，对政治、社会和生活造成了一定的负面影响，因此辨别伪造视频日益迫切。本文从 3 个角度来探索提高假脸检测模型泛化性的方法，分别是数据、特征和学习策略，具体包括其特点以及对泛化性的影响。同时，本文还通过对比了 Xception，ResNet-50 和 EfficientNet-b0 这 3 种常用的 CNN 分类器在 FF＋＋数据集中的泛化性，发现大多数模型在相同篡改方法上表现出了高精度，但在未知篡改方法上效果较差。通过实验分析，得到了如下结论：从数据方面来看，合适的数据增强方法会在一定程度上提升模型的泛化能力。从特征方面来看，真脸与假脸不仅在 RGB 域中会表现出差异，在频域中也会表现出更明显的不同。为了更好地区分真实人脸与伪造人脸，需要探索与伪造方法无关的内在特征，如视频帧内或帧间的不一致性。多种特征结合使用常常会提高分类器的泛化性，因此如何选择和融合不同域的特征是未来需要研究的一个方向。从学习策略的角度来看，假脸检测作为分类任务，可以与其他先进的训练策略相结合，如迁移学习、可持续性学习和元学习等。

参 考 文 献

[1] KARRAS T，LAINE S，AILA T. A style-based generator architecture for generative adversarial networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[2] DANG H，LIU F，STEHOUWER J，et al. On the detection of digital face manipulation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020：5780-5789.

[3] NEVES J C，TOLOSANA R，VERA-RODRIGUEZ R，et al. GANprintR：Improved fakes and evaluation of the state-of-the-art in face manipulation detection [J]. IEEE Journal of Selected Topics in Signal Processing，2020.

[4] GOODFELLOW I，OUFET-ABADIE J，MIRZA M，et al. Generative adversarial nets[C]// Neural Information Processing

Systems (NeuralIPS'14). 2014:2672-2680.

[5] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation[C]// Proceedings of the International Conference on Learning Representations (ICLR). 2018.

[6] FaceSwap [EB/OL]. (2018-10-29) [2021-10-12]. https://github.com/MarekKowalski/FaceSwap.

[7] DeepFakes[EB/OL]. (2018-10-29)[2021-10-12]. https://github.com/deepfakes/faceswap.

[8] LI L Z, BAO J, YANG H, et al. FaceShifter: Towards high fidelity and occlusion aware face swapping[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[9] NIRKIN Y, KELLER Y, HASSNER T. Fsgan: Subject agnostic face swapping and reenactment[C]// IEEE International Conference on Computer Vision (ICCV). 2019:7184-7193.

[10] LIU M, DING Y K, XIA M, et al. Stgan: A unified selective transfer network for arbitrary image attribute editing[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019:3673-3682.

[11] THIES J, ZOLLHÖFER M, STAMMINGER M, et al. Face2Face: Real-time face capture and reenactment of RGB videos [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016:2387-2395.

[12] THIES J, ZOLLHÖFER M, NIEßNER M. Deferred neural rendering: Image synthesis using neural textures [J]. ACM Transactions on Graphics (TOG), 2019, 38(4):1-12.

[13] GONZALEZ-SOSA E, FIERREZ J, VERA-RODRIGUEZ R, et al. Facial soft biometrics for recognition in the wild: Recent works, annotation and COTS evaluation[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(8):2001-2014.

[14] CHOI Y, CHOI M, KIM M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.

[15] HE Z L, ZUO W M, KAN M, et al. AttGAN: Facial attribute editing by only changing what you want[J]. IEEE Transactions on Image Processing, 2019, 28(11):5464-5478.

[16] FakeApp[EB/OL]. (2018-09-01)[2021-10-12]. https://www.fakeapp.com.

[17] SUWAJANAKORN S, SEITZ S M, KEMELMACHER-SHLIZERMAN I. Synthesizing obama: learning lip sync from audio [J]. ACM Transations on Grapgics(TOG), 2017, 36(4):95. 1-95. 13.

[18] LI Y, CHING M C, LYU S. In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking[C]// 2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, 2018:1-7.

[19] FERNANDES S, RAJ S, ORTIZ E, et al. Predicting heart rate variations of deepfake videos using neural ode[C]// IEEE International Conference on Computer Vision Workshops. 2019:1721-1729.

[20] YANG X, LI Y, LYU S. Exposing deep fakes using inconsistent head poses[C]// Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019:8261-8265.

[21] AGARWAL S, FARIDL H. Protecting world leaders against deep fakes[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019:38-45.

[22] YU N, DAVIS L, FRITZ M. Attributing Fake Images to GANs: Learning and analyzing GAN fingerprints[C]// IEEE International Conference on Computer Vision (ICCV). 2019:7556-7566.

[23] MI Z J, JIANG X H, SUN T F, et al. Gan-generated image detection with self-attention mechanism against gan generator defect[J]. IEEE Journal of Selected Topics in Signal Processing, 2020.

[24] SUN X W, WU B T, CHEN W. Identifying invariant texture violation for robust deepfake detection[J]. arXiv:2012. 10580, 2020.

[25] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[26] NGUYEN H H, YAMAGISHI J, ECHIZEN I. Capsule-forensics: Using capsule networks to detect forged images and videos [C]// Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019:2307-2311.

[27] GUERA D, DELP J. Deepfake video detection using recurrent neural networks[C]// 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2018:1-6.

[28] SABIR E, CHENG J, JAISWAL A, et al. Recurrent convolutional strategies for face manipulation detection in videos[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019:80-87.

[29] AMERINI I, GALTERI L, CALDELLI R, et al. Deepfake video detection through optical flow based CNN[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019:1205-1207.

[30] WODAJO D, ATNAFU S. Deepfake video detection using convolutional vision transformer [J]. arXiv:2102. 11126, 2021.

[31] KORSHUNOV P, MARCEL S. Deepfakes: a new threat to face recognition? assessment and detection[J]. arXiv:1812. 08685, 2018.

[32] VidTIMIT Audio-Video Dataset[EB/OL]. [2021-10-12]. http://conradsanderson. id. au/vidtimit.

[33] Faceswap-GAN[EB/OL]. (2018-08-27) [2021-10-12]. https://github. com/shaoanlu/faceswap-GAN.

[34] ROSSLER A, COZZOLINO D, VERDOLIVA L, et al. FaceForensics++: Learning to detect manipulated facial images[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019:1-11.

[35] LI X R, JI S L, WU C M, et al. Survey on deepfakes and detection Techniques [J]. Journal of Software, 2021, 32(2):496-518.

[36] Google AI blog. Contributing data to deepfake detection research [EB/OL]. [2021-10-21]. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html.

[37] LI Y, SUN P, QI H G, et al. Celeb-DF: A large-scale challenging dataset for deepfake forensics[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020.

[38] DOLHANSKY B, HOWES R, PFLAUM B, et al. The deepfake detection challenge (dfdc) preview dataset [J]. arXiv: 1910.08854, 2019.

[39] DOLHANSKY B, BITTON J, PFLAUM B, et al. The DeepFake Detection Challenge (DFDC) dataset [J]. arXiv: 2006.07397, 2020.

[40] ZI B J, CHANG M H, CHEN J J, et al. WildDeepfake: A challenging real-world dataset for deepfake detection[J]. ACM MM, 2020.

[41] JIANG L M, LI R, WU W, et al. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020: 2886-2895.

[42] ZHOU T F, WANG W G, LIANG Z Y, et al. Face forensics in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.

[43] PEROV I, GAO D H, CHERVONIY N, et al. DeepFaceLab: A simple, flexible and extensible face swapping framework[J]. arXiv: 2005.05535, 2020.

[44] HE Y N, GAN B, CHEN S Y, et al. ForgeryNet: A versatile benchmark for comprehensive forgery analysis[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.

[45] LE T N, NGUYEN H H, YAMAGISHI J, et al. OpenForensics: Large-Scale challenging dataset for multi-face forgery detection and segmentation in-the-wild [J]. arXiv: 2107.14480, 2021.

[46] ZAKHAROV E, SHYSHEYA A, BURKOV E, et al. Few-shot adversarial learning of realistic neural talking head models[C]// IEEE International Conference on Computer Vision (ICCV). 2019.

[47] POLYAK A, WOLF L, TAIGMAN Y. TTS skins: Speaker conversion via asr[J]. arXiv: 1904.08983, 2019.

[48] CAO H W, COOPER D G, KEUTMANN M K, et al. Crema-d: Crowd-sourced emotional multimodal actors dataset[J]. IEEE Transactions on Affective Computing, 2014, 5(4): 377-390.

[49] LIVINGSTONE S R, RUSSO F A. The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial andvocal expressions in North American English[J]. PLOSONE, 2018, 13(5): e0196391.

[50] CHUNG J S, NAGRANI A, ZISSERMAN A. Voxceleb2: Deep speaker recognition[J]. arXiv: 1806.05622, 2018.

[51] EPHRAT A, MOSSERI I, LANG O, et al. Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation[J]. arXiv: 1804.03619, 2018.

[52] SIAROHIN A, LATHUILIÈRE S, TULYAKOV S, et al. First order motion model for image animation[C]// Neural Information Processing Systems (NeuralIPS'19). 2019.

[53] KARRAS T, LAINE S, AITTALA M, et al. Analyzing and improving the image quality of stylegan[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[54] CHEN L, MADDOX R K, DUAN Z Y, et al. Hierarchical cross-modal talking face generation with dynamic pixel-wise loss[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[55] CHOI Y, UH Y, YOO J, et al. Stargan v2: Diverse image synthesis for multiple domains[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020: 188-197.

[56] FRIED O, TEWARI A, ZOLLHOFER M, et al. Text-based editing of talking-head video[J]. ACM Transactions on Graphics (TOG), 2019, 38(4): 1-14.

[57] DENG Y, YANG J L, CHEN D, et al. Disentangled and controllable face image generation via 3d imitative-contrastive learning [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[58] JO Y, PARK J. Sc-fegan: Face editing generative adversarial network with user's sketch and color[C]// ICCV. 2019.

[59] LEE C H, LIU Z W, WU L Y, et al. Maskgan: Towards diverse and interactive facial image manipulation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[60] KUZNETSOVA A, ROM H, ALLDRIN N, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale [J]. International Journal of Computer Vision, 2020.

[61] SHEN Y J, GU J J, TANG X O, et al. Interpreting the latent space of gans for semantic face editing[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[62] PIDHORSKYI S, ADJEROH D, DORETTO G. Adversarial latent autoencoders[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.

[63] NIRKIN Y, MASI I, TUAN A T, et al. On face segmentation, face swapping, and face perception[C]// 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, 2018: 98-105.

[64] BLANZ V, SCHERBAUM K, VETTER T, et al. Exchanging faces in images[C]// EuroGraphics. 2004.

[65] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[C]// Proceedings of the International Conference on Learning Representations (ICLR). 2018.

[66] YUN S, HAN D, OH S J, et al. CutMix: Regularization strategy to train strong classifiers with localization features [C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2019.

[67] CUBUK E D, ZOPH B, MAN D, et al. AutoAugment: Learning augmentation policies from data[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[68] CUBUK E D, ZOPH B, SHLENS J, et al. RandAugment: Practical automated data augmentation with a reduced search space [J]. arXiv: 1909.13719, 2019.

[69] WANG S Y, WANG O, ZHANG R, et al. CNN-generated ima-

ges are surprisingly easy to spot for now[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020:8692-8701.
[70] WANG J, WU Z X, CHEN J J, et al. M2TR: Multi-modal multi-scale transformers for deepfakedetection[J]. arXiv:2104.09770, 2021.
[71] LI Y Z, LYU S W. Exposing deepfake videos by detecting face warping artifacts[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2019:46-52.
[72] HULZEBOSCH N, IBRAHIMI S, WORRING M. Detecting CNN-generated facial images in real-world scenarios[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.
[73] COZZOLINO D, THIES J, RÖSSLER A, et al. ForensicTransfer: Weakly-supervised domain adaptation for forgery detection [J]. arXiv:1812.02510, 2018.
[74] XUAN X H, PENG B, WANG W, et al. On the generalization of GAN image forensics[C]// Proceedings of the Chinese Conference on Biometric Recognition. 2019:134-141.
[75] HE Y, YU N, KEUPER M, et al. Beyong the spectrum: Detecting deepfakes via reSynthesis[C]// International Joint Conference on Artificial Intelligence(IJCAI). 2021.
[76] JEON H, BANG Y, KIM J, et al. T-GD: Transferable GAN-generated images detection framework[C]// ICML. 2020.
[77] GUO Z Q, YANG G B, CHEN J Y, et al. Fake face detection via adaptive manipulation traces extraction network [J]. arXiv: 2005.04945, 2020.
[78] LI Y Q, BAI T. Deepfake detection method in videos based on convolutional LSTM [J]. Information Technology and Network Security, 2021, 40(40):28-32.
[79] DEVRIES T, TAYLOR G W. Improved Regularization of Convolutional Neural Networks with Cutout [J]. arXiv: 1708.04552, 2017.
[80] YANG C, LIM S N. One-shot domain adaptation for face generation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.
[81] BONDI L, CANNAS E D, BESTAGINI P, et al. Training strategies and data augmentations in CNN-based deepfake video detection [J]. arXiv:2011.07792, 2020.
[82] QIAN Y Y, YIN G J, SHENG L, et al. Thinking in frequency: Face forgery detection by mining frequency-aware clues[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2020.
[83] YU Y, NI R G, ZHAO Y. Mining Generalized Features for Detecting AI-Manipulated Fake Faces[J]. arXiv:2010.14129, 2020.
[84] CHEN Y P, FAN H Q, XU B, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2019.
[85] LIU H G, LI X D, ZHOU W B, et al. Spatial-Phase shallow learning: Rethinking Face Forgery Detection in Frequency Domain[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.
[86] ZHOU P, HAN X T, MORARIU V I, et al. Two-stream neural networks for tampered face detection[C]// IEEE Conferenceon Computer Vision and Pattern Recognition Workshops. 2017: 1831-1839.
[87] AFCHAR D, NOZICK V, YAMAGISHI J, et al. Mesonet: a compact facial video forgery detection network[C]// IEEE International Workshop on Information Forensics and Security (WIFS'18). 2018:1-7.
[88] MASI I, KILLEKAR A, MASCARENHAS R M, et al. Two-branch recurrent network for isolating deepfakes in videos[C]// 16th European Conference on Computer Vision (ECCV). 2020.
[89] NGUYEN H H, FANG F M, YAMAGISHI J, et al. Multi-task learning for detecting and segmenting manipulated facial images and videos[C]// Proceedings of the IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). 2019.
[90] LI X D, LANG Y N, CHEN Y F, et al. Sharp multiple instance learning for deepfake video detection[C]// Proceedings of the 28th ACM International Conference on Multimedia. 2020:1864-1872.
[91] KHODABAKHSH A, RAMACHANDRA A, RAJA A, et al. Fake face detection methods: Can they be generalized[C]// 2018 International Conference of the Biometrics Special Interest Group. Darmstadt, Germany, 2018:1-6.
[92] LIU Z Z, QI X J, TORR P. Global texture enhancement for fake face detection in the wild[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.
[93] ZHAO H Q, ZHOU W B, CHEN D D, et al. Multi-attentional deepfake detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.
[94] LI L, BAO J, ZHANG T, et al. Face x-ray for more general face forgery detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2020.
[95] HU Y J, GAO Y F, LIU B B, et al. Deepfake videos detection based on image segmentation with deep neural networks[J]. Journal of Electronics & Information Technology, 2021, 43(1): 162-170.
[96] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[97] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015: 815-823.
[98] ZHAO T C, XU X, XU M Z, et al. Learning to recognize patch-wise consistency for deepfake detection[J]. arXiv:2012.09311, 2020.
[99] SHANG Z H, XIE H T, ZHA Z J, et al. PRRNet: Pixel-Region relation network for face forgery detection [J]. Pattern Recognition, 2021, 116:107950.
[100] YU C M, CHANG C T, TI Y W. Detecting deepfake-forged contents with separable convolutional neural network and image segmentation[J]. arXiv:1912.12184, 2019.

[101]CHEN Z H,YANG H. Attentive semantic exploring for manipulated face detection[C]// IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP). 2021:1985-1989.

[102]DU M,PENTYALA S,LI Y N,et al. Towards generalizable deepfake detection with locality-aware AutoEncoder[C]// Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20). 2020.

[103]SUN Z,HAN Y J,HUA Z Y,et al. Improving the efficiency and robustness of deepfakes detection through precise geometric features[C]// IIEEE Conference on Computer Vision and Pattern Recognition. 2021.

[104]CALDELLI R,GALTERI L,AMERINI I,et al. Optical flow based CNN for detection of unlearnt deepfake manipulations [J]. Pattern Recognition Letters,2021,146(10):31-37.

[105]GANIYUSUFOGLUA I,NGÔ L M,SAVOV N,et al. Spatio-temporal features for generalized detection of deepfake videos [C]// Computer Vision and Image Understanding. 2020.

[106]TRINH L,TSANG M,RAMBHATLA S,et al. Interpretable and trust worthy deepfake detection via dynamic prototypes [C]// 2021 IEEE Winter Conference on Applications of Computer Vision (WACV'21). 2021.

[107]GU Z H,CHEN Y,YAO T P,et al. Spatiotemporal Inconsistency Learning for DeepFake Video Detection [J]. arXiv: 2109.01860,2021.

[108]HAN Y C,HUA G,ZHANG H J. Inception3D net based video RE forgery detection jointly exploiting eye and mouth areas[J]. Journal of Signal Processing,2021,37(4):567-577.

[109]XING H,LI M. Deepfake Video Detection Based on 3D Convolutional Neural Networks [J]. Computer Science,2021,48(7):86-92.

[110]KHAN S A,DAI H. Video Transformer for Deepfake Detection with Incremental Learning[C]// Proceedings of the 29th ACM International Conference on Multimedia(MM '21). ACM,2021.

[111]GUO J Z,ZHU X G,LEI Z. 3DDFA[EB/OL]. [2021-10-12]. https://github.com/ cleardusk/3DDFA.

[112]GUO J Z,ZHU X G,YANG Y,et al. Towards fast,accurate and stable 3D dense face alignment[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2020.

[113]LI J M,XIE H T,LI J H,et al. Frequency-aware discriminative feature learning supervised by single-center loss for face forgerydetection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.

[114]CHEN S,YAO T P,CHEN Y,et al. Local Relation learning for face forgery detection[C]// AAAI. 2021.

[115]LUO Y C,ZHANG Y,YAN J C,et al. Generalizing face forgery detection with high-frequency features[J]. arXiv: 2103.12376,2021.

[116]FRIDRICH J,KODOVSKY J. Rich models for steganalysis of digital images [J]. IEEE Transactions on Information Forensics and Security,2012,7(3):868-882.

[117]CHINTHA A,RAO A,SOHRAWARDI S,et al. Leveraging edges and optical flow on faces for deepfake detection[C]// 2020 IEEE International Joint Conference on Biometrics (IJCB). 2020.

[118]WANG G X,ZHOU J H,WU Y. Exposing deep-faked videos by anomalous co-motion pattern detection [J]. arXiv:2008.04848,2020.

[119]HALIASSOS A,VOUGIOUKAS K,PETRIDIS S,et al. Lips Don't Lie:A generalisable and robust approach to face forgery detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2021.

[120]KIRKPATRICK J,PASCANU R,RABINOWITZ N,et al. Overcoming catastrophic forgetting in neural networks [J]. PNAS,2017,114(13):3521-3526.

[121]ANEJA S,NIEßNER M. Generalized zero and few-shot transfer for facial forgery detection [J]. arXiv:2006.11863,2020.

[122]TARIQ S,LEE S,WOO S S. A convolutional LSTM based residual network for deepfake video detection [J]. arXiv: 2009.07480,2020.

[123]TARIQ S,LEE S,WOO S S. One detector to rule them all:Towards a general deepfake attack detection framework[C]// Proceedingsof the Web Conference. 2021.

[124]LEE S,TARIQ S,KIM J,et al. TAR:Generalized forensic framework to detect deepfakes using weakly supervised learning[C]// IFIP-SEC. 2021.

[125]TARIQ S,LEE S,KIM H,et al. Gan is a friend or foe?:a framework to detect various fake face images[C]// Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. ACM,2019:1296-1303.

[126]SUN K,LIU H,YE Q X,et al. Domain general face forgery detection by learning to weight[C]// AAAI Conference on Artificial Intelligence. 2021:2638-2646.

[127]MARRA F,SALTORI C,BOATO G,et al. Incremental learning for the detection andclassification of GAN-generated images [J]. arXiv:1910.01568,2019.

[128]REBUFFI S A,KOLESNIKOV A,SPERL G,et al. iCaRL:Incremental classifier and representation learning[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2017:2001-2010.

[129]KIM M,TARIQ S,WOO S S. CoReD:Generalizing fake media detection with continual representation using distillation[C]// 29th ACM International Conference on Multimedia (ACMMM '21). 2021.

[130]DONG X Y,BAO J M,CHEN D D,et al. Identity-driven deepfake detection[J]. arXiv:2012.03930,2020.

[131]HOWARD A G,ZHU M L,CHEN B,et al. Mobilenets:Efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861,2017.

[132]TANAKA M,KIYA H. Fake-image detection with robust hashing[C]// 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech). 2021.

[133]VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]// NIPS. 2017:5998-6008.

[134]MIAO C T, CHU Q, LI W H, et al. Towards generalizable and robust face manipulation detection via bag-of-local-feature[J]. arXiv:2103.07915, 2021.

[135]FUNG S, LU X Q, ZHANG C, et al. DeepfakeUCL: Deepfake detection via unsupervised contrastive learning[C]// The annual International Joint Conference on Neural Networks (IJCNN). 2021.

[136]CHEN Q F, KOLTUN V. Photographic image synthesis with cascaded refinement networks[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017.

[137]LI K, ZHANG T H, MALIK J. Diverse image synthesis from semantic layouts via conditional IMLE[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2019.

[138]CHEN C, CHEN Q F, XU J, et al. Learning to see in the dark [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[139]DAI T, CAI J R, ZHANG Y B, et al. Second-order attention network for single image super-resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[140]GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans[C]// Neural Information Processing Systems. 2017:5767-5777.

[141]PUMAROLA A, AGUDO A, MARTINEZ A M, et al. Ganimation: Anatomically aware facial animation from a single image [C]// Proceedings of the European conference on computer vision (ECCV). 2018:818-833.

[142]CHEN Y C, XU X G, TIAN Z T, et al. Homomorphic latent space interpolation for unpaired image-to-image translation [C]// IEEE Conference on Computer Vision and Pattern Recognition. 2019:2408-2416.

[143]ZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016:2818-2826.

[144]IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J]. ACM Transactions on Graphics (TOG), 2017, 36(4):107.1-107.14.

[145]YU J H, LIN Z, YANG J M, et al. Generative image inpainting with contextual attention[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[146]RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[147]KINGMA D P, DHARIWAL P. Glow: generative flow with invertible 1×1 convolutions[C]// NIPS. 2018:10236-10245.

[148]LI X R, YU K. A Deepfakes detection technique based on two-stream network[J]. Journal of Cyber Security, 2020, 5(2):84-91.

[149]CHEN T, KUMAR A, NAGARSHETH P, et al. Generalization of audio deepfake detection[C]// Proceedings of the Odyssey 2020 Speaker and Language Recognition Workshop. 2020:132-137.

[150]JIA Z Y, FANG H, ZHANG W M. MBRS: Enhancing robustness of DNN-based watermarking by mini-batch of real and simulated JPEG compression[C]// 29th ACM International Conference on Multimedia. 2021.

[151]BAO Y X, LU T L, DU Y H. Overview of Deepfake Video Detection Technology [J]. Computer Science, 2020, 47(9):283-292.

DONG Lin, born in 1999, postgraduate, is a member of China Computer Federation. Her main research interests include computer vision and multi-media security.

YE Feng, born in 1978, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include computer vision and artificial intelligence.

(责任编辑:柯颖)