2024 年 秋 季学期研究生课程考核

（读书报告、研究报告）

考 核 科 目: 学 术 写 作 与 学 术 规 范

学生所在院（系）: 未 来 技 术 学 院

学 生 所 在 学 科: 计 算 机 科 学 与 技 术

学 号: 2 0 2 1 1 1 2 3 9 7

学 生 类 别: 本 科 生

考 核 结 果 阅 卷 人

题 目：深度伪造人脸视频检测方法综述

摘要：随着深度伪造技术的迅猛发展，深度伪造人脸视频的检测已成为信息安全领域的重要研究方向。传统的视频伪造检测方法已无法有效应对基于生成对抗网络和扩散模型的深度伪造技术。本文首先回顾了当前视频和音频伪造技术的发展，包括人脸交换、面部属性操控、说话人脸合成等技术。接着，总结了目前深度伪造检测数据集的评价指标，分析了深度伪造检测领域的最新进展，重点关注了基于单帧图像、视频时空特征以及多模态融合的检测方法，并对现有检测技术的优缺点进行了总结。本文还聚焦了目前该领域的热点和难点问题，尤其是如何通过摆脱监督学习的限制来提高模型的泛化能力。最后，本文提出了深度伪造人脸检测领域未来的研究方向，旨在为更有效的深度伪造检测方法提供参考。

关键词：深度伪造；音视频伪造；伪造检测；多媒体取证；信息安全

目录

一、研究背景、目的及意义

1.1 研究背景

随着互联网的广泛应用和移动通信技术的飞速发展，互联网社交软件已经融入到人民群众的日常生活中，成为人们进行交流和传播信息的重要平台。中国互联网络信息中心统计显示[1]，截至 2024 年 6 月，我国网民规模近 11 亿人，较 2023 年 12 月增长 742 万人，互联网普及率达 78.0%。与此同时，抖音、快手等网络视听类应用的用户量超越即时通信应用，已成为我国互联网第一大类应用[2]。

图 1-1 2024 年用户规模前五的互联网应用[1]（单位：亿人）

在视频内容日益丰富的同时，伪造视频的数量也在迅速增加，给社会带来了信息真实性的挑战。传统的视频伪造技术基于 Adobe Photoshop 和 Primiere 等软件，通过人工编辑的方式对图像内容和音频信息进行篡改，这种伪造技术效率低下，且容易被人察觉。面对传统视频伪造技术，研究人员提出了一系列检测方法[3]–[5] ，这些方法在应对传统伪造技术时表现良好，具有极高的准确性。

随着深度学习技术的飞速发展，以 Deepfakes[6]、Faceswap[7]和文本语音生成为代表的新型视频和音频合成技术逐渐兴起，这些技术通过生成对抗网络和扩散模型等深度学习技术，生成高度逼真的视频和音频，并在社交媒体和在线视频平台等领域广泛传播。这些技术在影视娱乐、教育培训和宣传营销等方面生成了丰富且优质的内容，提供了创新性的应用和新机遇，图 1-2 展示了近期深度合成技术在多个领域的应用。

刘强东 AI 数字人直播带货[8]

天津大学推出虚拟 AI 担任助教[9]

抖音 AI 说话人脸视频生成技术[10]

图 1-2 深度合成技术应用举例

然而，这些技术同样可以用于对视频和音频的恶意伪造，其恶意使用对社会构成了重大威胁，导致了包括欺诈、诽谤和虚假信息传播等一系列问题。例如，不法分子利用智能 AI 换脸和伪造音频技术实施电信诈骗等新型犯罪，导致人民财产安全受到严重破坏。此外，在持续至今的俄乌冲突中，社交媒体上不断出现乌克兰总统泽连斯基的伪造投降视频，误导了民众认知。图 1-3 展示了近期深度伪造技术引发的社会安全问题案例。

不雅视频人物换脸[11]

虚假政治新闻传播[12]

新型电信诈骗[13]

图 1-3 深度伪造技术引发的社会安全问题案例

1.2 研究目的

为了应对深度伪造视频和音频技术带来的潜在安全风险，不仅要求社交媒体平台制定相应的管理策略和国家层面的法规政策[14]，更需要研发有效的深度伪造检测手段，以提升安全防范能力。传统图像伪造检测技术在处理深度伪造时的局限性愈发明显，在检测的准确性方面严重下降[15],[16]。针对复杂的深度伪造技术，迫切需要研究深度伪造检测方法。因此，本文的研究目的在于，调研当前流行的深度伪造技术，总结公开的人脸伪造检测数据集和评价指标，对深度伪造检测方法进行分类和综述，并对未来的研究方向进行展望。

1.3 研究意义

如图 1-4 所示，本文的研究意义包含理论意义、现实意义和时代意义三个方面。

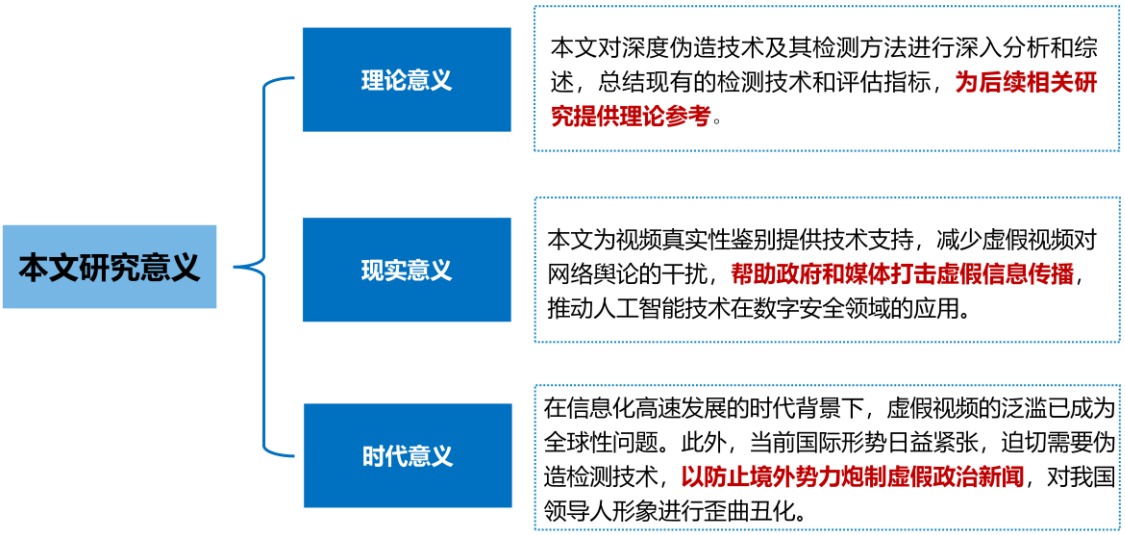

图 1-4 本文的研究意义

二、国内外研究现状分析

2.1 深度伪造技术

目前的深度伪造技术复杂多样，可大致分为视频伪造和音频伪造两类，如图 2-1 所示。

其中视频深度伪造可以分为人脸交换、面部属性操纵、面部重演、全脸合成和说话人脸合成。人脸交换通过将一个视频中的人脸替换为另一个人的面部图像，生成全新的人物面部视频；面部属性操控涉及对面部特征（如胡须、皱纹等）的修改，使得原始人物的面部属性发生变化；面部重演将一个人的面部表情或动作迁移到另一个人的面部；全脸合成对于整个人物面部，包括眼睛、鼻子、嘴巴、肤色、面部表情等各个部分，均进行伪造合成；说话人脸合成则通过使人物的面部动作（尤其是嘴唇运动）与语音内容同步，生成高度逼真的假视频[17]–[19]。

音频深度伪造则主要包括文本到语音合成（Text to Speech，TTS）和语音转换（Voice Conversion，VC）[20],[21]，文本到语音合成通过输入文本信息，生成与原始文本内容相匹配的语音，常用于智能助手、语音导航等应用；语音转换将一个人的声音转换成另一个人的声音，而不改变音频原本的内容。

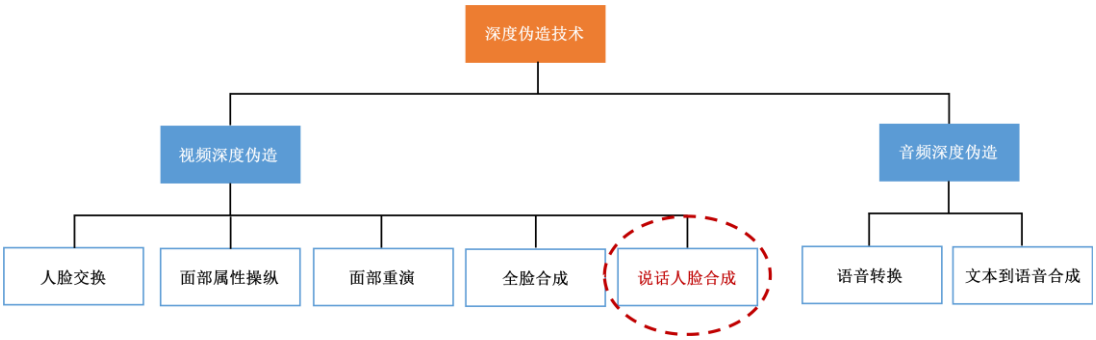

图 2-1 常见的深度伪造技术

说话人脸合成是视频深度伪造中危害程度极大的一种伪造方法[17],[18],[22]，特别是它能够与语音转换和文本到语音合成等音频伪造技术相结合，使目标人物用他/她自己的声音，说出从未说出过的话。因此，本课题主要针对这类深度伪造技术进行检测。如图 2-2 所示，伪造者首先根据本文通过音频伪造技术生成特定的音频内容，说话人脸合成技术根据音频内容对视频中的人物嘴唇动作进行篡改，使其与音频内容相匹配。

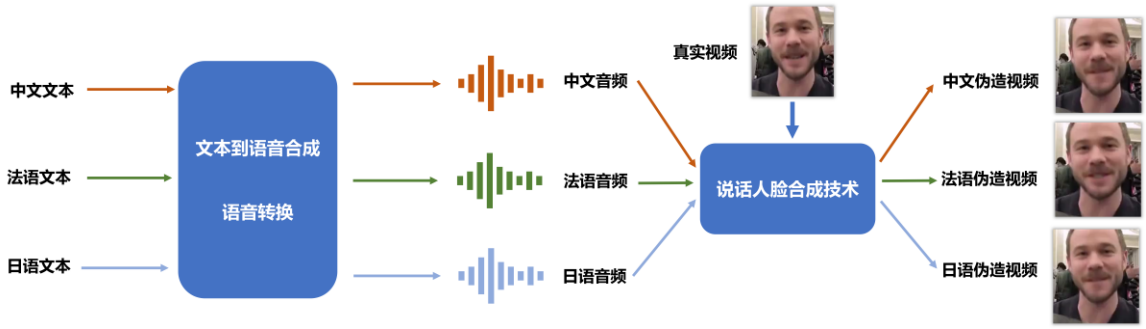

图 2-2 音频伪造技术与说话人脸合成技术相结合的示例

当前主要的说话人脸合成技术复杂，且随着研究的进展而持续改进，根据技术的可以分为以下 4 类：音频/文本驱动的方法，多模态条件化的方法，基于扩散模型的方法和基于 3D 模型的方法，表 2-1 对这 4 类方法进行了总结。

表 2-1 说话人脸合成技术

方法类型	代表性技术	方法介绍
音频/文本驱动	Wav2Lip[23]、MakeItTalk[24]、TalkLip[25]、RADIO[26]	通过音频或文本输入生成与之同步的面部表情和口型动画，适用于简单的语音或文本到视频合成任务
多模态条件化	PC-AVS[27]、GC-AVT[28]、LipFormer[29]	结合音频、视频、文本多个模态作为输入，融合来自不同来源的信息，生成更加复杂、自然且多样化的说话人脸
基于扩散模型	DAE-Talker[30]、DreamTalk[31]、EmoTalker[32]	通过逐步去噪的生成过程，使用扩散模型生成高质量的面部动画或视频，尤其擅长细腻的图像细节生成
基于 3D 模型	DFRF[33]、AE-NeRF[34]、SyncTalk[35]	利用三维人脸模型生成面部动画，能够处理角度变化和复杂的光照条件，适用于高保真度和真实感要求高的应用

2.2 人脸伪造检测数据集和评价指标

为了训练和评估检测模型的性能，研究人员构建了一系列人脸伪造检测数据集。根据伪造技术的复杂性、数据规模和伪造模态，现有人脸伪造数据集可分为三代[36]：第一代数据集采用的伪造技术较为单一且规模较小，包括 FaceForensics++[37]、Celeb-DF-v1[38]、Celeb-DF-v2[39]、WildDeepfake[40]等；第二代数据集结合多种伪造技术，伪造质量显著提高，辨别难度增加，包括 DFDC[41]和 ForgeryNet[42]等；第三代数据集的伪造技术更为复杂，涵盖音频与视频的多模态伪造，甚至涉及语音同步和内容生成等语义层面的深度伪造，包括 KoDF[43]、FakeAVCeleb[44]、LAV-DF[45]和 DefakeAVMiT[46]等。

表 2-2 人脸伪造检测数据集

数据集名称	模态	真实数据规模	伪造数据规模	数据集特点
FaceForensics++	视频	1000	4000	涵盖多种伪造方法（如 Deepfake 和 Face2Face），广泛用于评估检测算法
Celeb-DF-v1	视频	408	795	基于 Youtube 上不同年龄、种族和性别的原创视频生成的高质量深伪数据集

续表 2-2 人脸伪造检测数据集

数据集名称	模态	真实数据规模	伪造数据规模	数据集特点
Celeb-DF-v2	视频	590	5926	Celeb-DF-v1 的优化版本，伪造人脸质量更高、更逼真，规模更大
WildDeepfake	视频	3805	3509	数据均来源于网络,内容丰富多样。视频效果更加真实,更符合真实生活场景
ForgeryNet	视频	99630	121617	目前规模最大的深度伪造数据集之一,包括时序伪造和空间伪造等伪造方法
DFDC	视频+音频	23654	104500	Facebook 举办的 Deepfake 检测竞赛数据集，来源多样且包含多种伪造技术
KoDF	视频+音频	62166	175776	包含针对韩国人的大量真实、伪造视频
FakeAVCeleb	视频+音频	500	19500	涵盖音频和视频伪造,适用于研究跨模态一致性检测任务
LAV-DF	视频+音频	36431	99873	基于内容驱动对音频和视频进行伪造
DefakeAVMiT	视频+音频	540	64800	利用了 5 种视觉生成技术和 3 种语音生成技术的多模态数据集

对检测模型进行评价时，常采用的指标包括准确率（Accuracy，ACC）、平均精度（Average Precision, AP）和 ROC 曲线下与坐标轴围成的面积（Area Under the ROC Curve, AUC）。

ACC 表示模型正确预测的样本数量占总样本数量的比例，计算方式为：

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-1}$$

其中，TP（True Positive）为正确预测为正类的样本数；TN（True Negative）为正确预测为负类的样本数；FP（False Positive）为错误地预测为正类的负类样本数；FN（False Negative）为错误地预测为负类的正类样本数。

尽管 ACC 简单直观，但在正负样本分布不均的情况下，可能无法全面反映模型的性能。因此，本课题在采用 ACC 指标的基础上，额外采用在处理数据不平衡时常用的 AP 和 AUC 作为评估指标，以更准确地衡量模型的分类能力。

AP 是精确率（Precision）-召回率（Recall）曲线下的面积，精确率和召回率的定义为：

$$Precision = \frac{TP}{TP + FP} \tag{2-2}$$

$$Recall = \frac{TP}{TP + FN} \tag{2-3}$$

AP 通过衡量模型在不同阈值下的精确率和召回率，来判断模型在不同决策阈值下的综合表现。AP 的范围为 0~1，AP 越接近 0，则表明模型的分类能力较差；AP 越接近 1，表示模型在所有阈值下均能准确区分正负样本。

AUC 是召回率-假阳性率（FPR）曲线下的面积，FPR 的定义为：

$$FPR = \frac{FP}{FP + TN} \tag{2-4}$$

AUC 的范围同样为 0~1，AUC = 0.5 表示模型没有分类能力，相当于随机猜测；AUC < 0.5 表示模型的分类性能低于随机猜测，可能模型出现了问题或标签反转；AUC 越接近 1 则表示模型的分类能力越强，可以完美区分正负样本。

2.3 人脸伪造检测方法

目前的人脸伪造检测方法可以分为三大类：一种是基于单帧图像检测方法，另一种综合考虑视频空间特征和时序特征进行检测，还有一种融合音频模态和视频模态的多模态人脸伪造检测方法。

2.3.1 基于单帧图像的人脸伪造检测方法

基于单帧图像进行检测的方法通过分析每一帧图像来识别伪造痕迹。在真实视频中，图像的相邻区域通常具有自然的连续性，而深度伪造视频中，伪造区域与真实区域通常来源于不同的图像，导致出现明显的不一致现象。因此，基于单帧图像的检测方法主要关注图像空间的不一致性，如颜色纹理、噪声指纹、视频伪影等。例如，在[37]中，作者基于 XceptionNet[47]提出了一个作为基准的人脸伪造检测模型，并通过公开数据集展示了其有效性。如图 2-3(a)，在[48]中作者提出了一种名为 Face X-ray 的检测方法，将伪造图像视为不同图像的混合，通过显示混合边界对伪造图像进行检测。如图 2-3(b)，Hao Dang 等人制作了一系列注意力图模板，利用注意力区域来处理伪造图像的特征图，而不是简单地预测伪造图像的伪造混合区域[49]。Hanqing Zhao 等人进一步将深度伪造检测问题深化为细粒度图像分类问题，在[50]中提出了多注意力区域的方法，通过数据增强机制和区域独立损失函数迫使模型关注伪造视频的不同区域。

(a)基于混合边界进行检测，白色区域为混合边界　(b)基于注意力图进行检测，白色区域为伪造区域

图 2-3 基于单帧图像的人脸伪造检测方法示例

基于单帧图像进行伪造检测的优点在于，其每次检测都只输入单帧图片，计算开销较低，适合于资源受限的环境或需要快速响应的场景，并且可以给出具体的伪造区域，具有

较高的可解释性。然而，由于单帧图像无法捕捉到视频中的动态信息，如面部表情的变化、眨眼等动态特征，这可能导致伪造视频中的动作不连续性和突变性无法有效检测。

2.3.2 基于视频时空特征的人脸伪造检测方法

由于许多深度伪造方法生成视频时是逐帧处理并拼接图像，缺乏对帧间连续性的关注，这导致生成的伪造视频可能出现抖动或突变等问题[51]（如图 2-4 所示）。因此一些研究人员开始考虑同时利用空间域和时间域的信息。例如，近期研究[52]通过三维卷积神经网络提取空间特征，并将这些特征与时间模块进行融合，从而有效地捕捉时空上下文信息。其他研究考虑了视频中人物头部姿势不一致[53],[54]或眨眼[55],[56]等因素。

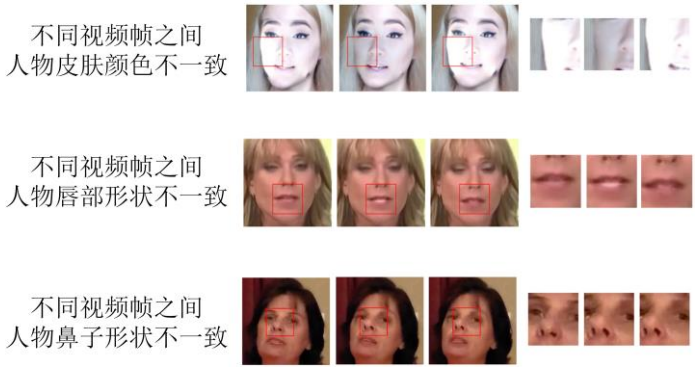

图 2-4 伪造视频帧与帧之间出现的不连续或突变等问题

自 Vision Transformer（ViT）[57]提出以来，许多基于 ViT 的伪造人脸检测方法应运而生，ViT 的网络结构如图 2-5 所示。ViT 相比于卷积神经网络能更好地理解输入序列中的时间和空间上下文关系。例如，近期研究[58]提出了一种结合卷积神经网络和 Transformer 网络的方法，以利用短时和长时的时间不一致性来对伪造视频进行检测。在[59]中，作者通过 ViT 模拟有限空间区域序列的时间上下文一致性,以全局对比方式对伪造视频分类。

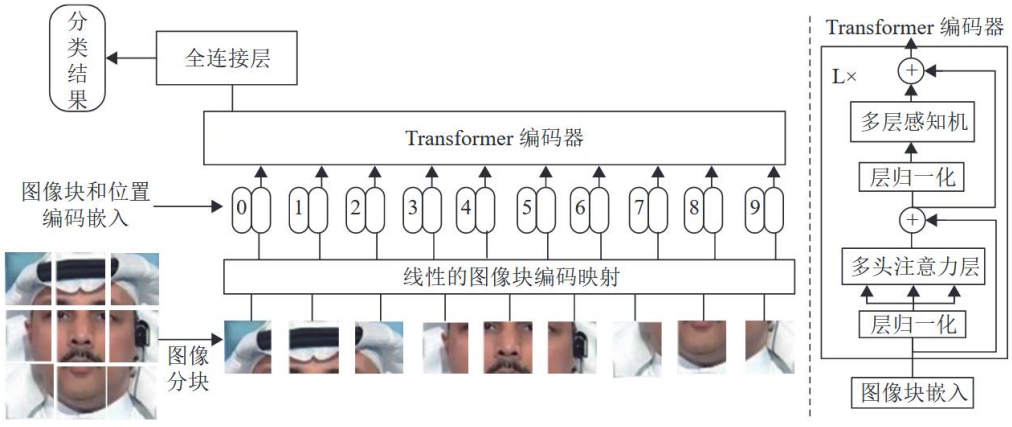

图 2-5 ViT 的网络结构

基于视频时空特征的人脸伪造检测方法的优点在于，这种方法能够捕捉视频中的动态变化，分析视频帧之间的时序关系，识别因伪造技术导致的视频帧之间的突变和不一致，

相比单帧图像检测，提高了伪造视频检测的准确性。但其缺点在于，综合考虑视频的空间和时序特征需要处理大量数据，尤其是在长时间的视频中，计算复杂度和存储需求较高，可能导致实时性差。

此外，当图像遭受到压缩、视频编解码转换和高斯模糊等破坏性操作导致分辨率降低时，基于视频单模态的人脸伪造检测方法容易受到干扰，从而导致检测准确性下降[22],[60]。为了应对这些挑战，未来的研究需要探索多模态融合方法，以提高检测的鲁棒性和准确性。

2.3.3 基于音视频多模态的人脸伪造检测方法

现如今的深度伪造视频往往包含音频伪造和视频伪造，研究人员尝试结合音频信息，利用音频和视频模态之间的关联进行深度伪造检测。如图 2-6，近期研究[60]指出，真实视频中的特定嘴形与声音具有匹配性，可以通过检测视频中的音素和唇形不匹配来辨别伪造视频，作者重点关注音素 M、B、P，在真实视频中，当人发出以上音素时，唇形是闭合的，而伪造视频则缺乏这一特征。在[61]中，作者利用视觉和听觉模态之间的语义一致性和时间同步关系，提取视频特征、音频特征以及二者的时间同步特征，作为检测模型的输入，实现了对音频和视频二者真实性的检测。为了防止公众人物遭受深度伪造视频技术的攻击，Davide 等在[62]中在真实视频上通过对比学习，提取公众人物的音频和视频的一致性特征，将待检测视频的特征与真实特征进行比对，实现了良好的检测效果。近期研究[63]提出了一个统一的多模态框架 Multimodaltrace，该框架通过音频和视频模态的独立混合和联合混合方式进行提取和处理特征，最终通过多标签分类器对视频进行检测。

图 2-6 音素及其对应的唇部动作

基于音视频多模态的人脸伪造检测方法的优点是，深度伪造视频通常会出现视频与音频的不一致，特别是在面部表情与语音内容之间，通过结合音频和视频，能够更好地发现这些不一致性，可以有效弥补单一模态的局限性，提高检测的全面性和准确性。

其缺点在于，多模态方法需要同时处理视频和音频数据，这增加了数据的存储和处理复杂度。此外，需要确保音视频的同步性，以便有效地进行融合分析。多模态融合模型通常较为复杂，需要处理不同模态之间的特征提取和融合问题，设计合适的融合策略也是该类方法的一个难点。

三、目前该领域的热点和难点问题

尽管许多现有伪造检测模型在训练集内的检测性能较好，由于训练方法受限于在伪造检测数据集上对模型通过监督学习的方式进行训练，导致模型的泛化能力受到限制[64]–[66]。模型一旦遇到新的数据集或训练集中未出现过的伪造方法时，往往会出现性能下降的问题。

因此，目前伪造检测领域的热点问题是如何突破监督学习的限制，以提升模型的跨伪

造类型泛化能力和跨数据集泛化能力。

一种热点解决方法是在大量真实视频上对模型进行音频表征学习和视频表征学习，通过特定的预训练任务，使模型学习到视频和音频的深层次特征，并将学习到的特征应用于下游任务。何恺明等人在[67]提出了一种名为掩蔽自动编码器（masked autoencoders，MAE）的 ViT 自监督预训练方法，其核心思想是随机屏蔽输入图像的部分区域，以重建丢失的像素为预训练任务，使编码器学习到图像的特征表示。此后，出现了一些基于 MAE 框架的自监督方法。例如，AV-MAE[68]将 MAE 扩展到了视频和音频多模态，编码器生成融合特征，用于对音频和视频解码。CAV-MAE[69]则在音频和视频特征之间施加对比学习损失函数，使模型能够学习到具有内在一致性的音频和视频的特征。

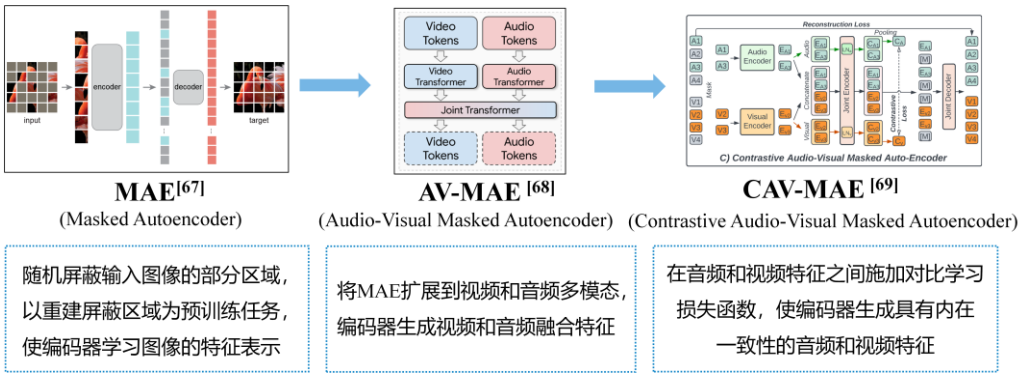

图 3-1 基于 MAE 架构的一系列自监督方法

许多研究人员将伪造检测任务视为音频视频表征学习的一种下游任务，并基于不同的上游任务，提取音频特征和视频特征。例如，LipForensics[22]通过在视觉语音识别任务上对模型进行预训练，学习自然嘴部运动的特征表示，并将其迁移到伪造检测任务中，通过识别嘴部运动的不规则性来检测伪造视频。还有研究人员指出，真实音视频中的音频和视频存在时间不同步性，但应遵循概率分布[70]。为此，作者在大量真实视频数据上训练了一个自回归模型，以学习真实音频和视频的时间对应关系。该模型能够预测待检测音视频中的音频与视频之间的不匹配程度，从而判断是否为伪造。如图 3-2，近期研究[71]基于 MAE 的预训练任务，对一张图片中的面部区域进行遮挡掩蔽，通过重建被掩蔽的面部区域，让编码器学习面部的深层次特征，用于对伪造人脸视频进行检测。

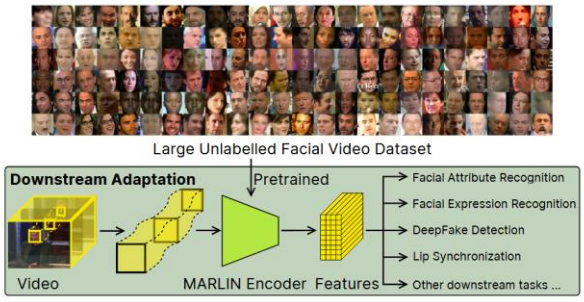

图 3-2 MAE 预训练任务应用于下游的伪造视频检测任务[71]

本文认为，可以将[71]的工作扩展到音频、视频多模态，以音频、视频相互重建作为上游任务，来进行音视频信息表征学习。在下游伪造人脸检测任务中，融合从上游任务中学

到的音频与视频特征，作为检测模型的输入，从而更准确地辨别视频的真伪。通过这种上游任务，模型能够学习到音频模态和视频模态之间更深层次的特征。

四、未来的研究展望

虽然已经有不少研究在伪造说话人脸检测领域取得了显著进展，但仍有许多问题亟待解决。本部分将展望伪造说话人脸检测技术的未来研究方向，讨论该领域的潜在发展趋势以及可能出现的挑战和机遇。

4.1 提高检测模型的实时性

尽管现有方法在研究环境中表现优秀，但要将这些技术应用于实际场景，还需要实现实时性和低延迟的需求。未来的研究可以关注提升深度伪造检测模型的计算效率，尤其是在高分辨率视频和大规模数据集上的实时处理能力。低延迟响应不仅对在线内容监控至关重要，也对防止恶意伪造具有重要意义。以具有较高准确率的在线检测平台 Sightengine 为例，对于一条长度为 45 秒的视频，检测时间长达 2 分钟，如果能够实现检测模型的实时性，将极大提高伪造视频的检测效率。

4.2 提高检测模型的可解释性

目前，深度伪造检测方法在可解释性方面往往存在不足，这导致了检测结果的信任度较低。尤其在法律、新闻等敏感领域，这一问题尤为突出，因为这些领域对检测结果的准确性和可靠性有着极高的要求。如果检测模型的判断过程无法被理解或解释，使用者往往会对其结果产生怀疑，进而影响模型的实际应用。

为了应对这一挑战，未来的研究可以着重探索如何提高深度伪造检测模型的可解释性。一个重要的方向是开发能够提供透明决策过程的模型，帮助用户理解模型是如何做出判定的。例如，研究者可以尝试结合可解释的机器学习技术，设计一些能够生成可视化或易于理解的判别依据的模型，使得用户不仅能看到“真伪”标签的输出，还能看到哪些特征或数据驱动了这一判断。这种可解释性不仅可以增强模型的透明度，还能增强用户对模型判断的信任。

4.3 提高检测模型的跨语言能力

当前的许多深度伪造说话人脸检测方法主要针对英语伪造视频进行检测，缺乏跨语言的普适性。随着全球化的深入，未来的研究应着力提升检测方法的跨语言和跨文化适应能力，确保伪造检测模型能够在多种语言和文化环境下有效进行检测。

4.4 应对来自对抗样本的攻击

随着伪造技术的持续进步，对抗攻击能够通过对输入数据进行精细扰动，使检测模型进行错误判断，因此当模型的训练数据中没有包含攻击样本时，会导致模型的检测性能急剧下降[72]-[74]。因此，提高检测模型对对抗样本的鲁棒性，是深度伪造检测领域未来研究中的一个关键方向。

面对这一挑战，研究者们可以从多个角度入手，探索抗对抗攻击的深度学习模型。例如，通过引入对抗训练机制，利用对抗样本增强模型的训练过程，使模型能够在面对伪造样本的干扰时，仍然保持高效的识别能力。此外，还可以考虑设计更加鲁棒的特征提取算法，强化模型对输入数据扰动的敏感性，以提高其鲁棒性。

五、结论

本文首先回顾了当前流行的视频音频伪造技术、公开的伪造检测数据集以及评价指标，并分析了深度伪造检测领域的最新进展。具体来说，本文重点关注了基于单帧图像、视频时空特征以及多模态融合的检测方法，并对现有检测技术的优缺点进行了总结。本文还聚焦了目前该领域的热点和难点问题，尤其是如何通过摆脱监督学习的限制来提高模型的泛化能力。虽然现有方法在检测精度方面取得了显著成果，但仍存在许多亟待解决的问题，包括检测模型的实时性、可解释性、跨语言能力和应对对抗样本攻击的能力。综上所述，深度伪造人脸视频检测是一个充满挑战但也极具研究价值的领域，随着技术的不断进步和创新，未来将会涌现出更多高效、准确的检测方法，助力信息安全和社会稳定。

六、学完本课程后的收获和体会

作为一名本科生，我在《学术写作与学术规范》这门课程中的收获和体会，可以分为两个部分，分别是在课堂小组展示中的收获和通过学习 mooc 内容的收获。

6.1 课堂小组展示中的收获

在课题小组展示中，通过参与小班讨论，我了解到了不同学术主题的写作和投稿经验，特别是有学术投稿经历的学长和学姐们分享了他们的实际案例和体会。这不仅使我对学术写作有了更为清晰的认知，还帮助我理解了如何从实际出发进行论文写作，如何规避常见的错误，并总结出有用的经验。

此外，通过点评典型的学位论文和经典的中英文学术论文，我对学术论文的结构、语言表达和研究方法有了更加全面的认识。在这些点评过程中，我学会了如何有理有据地分析一篇论文的优缺点，并结合实际案例给出改进意见。这对于我未来的学术写作和论文评审有着非常大的帮助。

通过这些展示和讨论，我不仅提升了自己的学术写作能力，还学会了如何与他人合作进行深入的学术探讨，这对我未来的科研和学习是一次非常有益的经验积累。

6.2 通过学习 mooc 内容的收获

通过学习 MOOC 的内容，我对学术研究的各个环节有了更深入的理解，尤其是在选题、文献检索、论文写作和学术规范方面取得了显著的收获。

首先，在选题和研究的模块中，我学习到了如何根据自己的兴趣和学术领域选择合适的课题。课程强调了选题的基本原则，如创新性、可行性和学术价值，这让我在未来的研究中能够更好地规划我的研究方向。在文献检索与阅读的学习中，我掌握了如何高效检索

相关文献，并运用一些检索技巧来快速找到有价值的资料。课程还讲解了如何阅读和梳理文献，如何撰写综述性文章，这对我的文献调研能力提升帮助巨大。开题报告和中期报告的撰写课程则让我对如何规范地进行研究报告撰写有了明确的认识。通过学习开题和中期报告的写作技巧，我能够更加清晰地表达研究思路、方法和进度。在学位论文写作部分，我了解了学位论文的整体结构和各个部分的写作要点，特别是如何撰写摘要、引言、实验部分和结论等。这些内容为我今后撰写学术论文提供了很好的框架和指导。最后，学术规范与学术道德的课程让我认识到学术诚信的重要性，特别是在学术引文和论文写作中，如何避免抄袭和剽窃，如何正确引用他人的研究成果，这对我今后的学术道路有着极大的启发。

综上所述，通过学习这些 MOOC 课程，我不仅掌握了系统的学术写作技巧，还提高了自己在学术研究中的规范性和道德意识，为未来的研究工作奠定了坚实的基础。

七、参考文献

[1] https://www.cnnic.cn/n4/2024/0829/c88-11065.html.

[2] https://www.cnii.com.cn/rmydb/202303/t20230330_458506.html.

[3] Amerini I, Ballan L, Caldelli R, 等. A sift-based forensic method for copy–move attack detection and transformation recovery[J]. IEEE transactions on information forensics and security, 2011, 6(3): 1099-1110.

[4] Agarwal R, Verma O P. An efficient copy move forgery detection using deep learning feature extraction and matching algorithm[J]. Multimedia Tools and Applications, 2020, 79(11-12): 7355-7376.

[5] Barni M, Phan Q T, Tondi B. Copy Move Source-Target Disambiguation Through Multi-Branch CNNs[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 1825-1840.

[6] Kietzmann J, Lee L W, McCarthy I P, 等. Deepfakes: Trick or treat?[J]. Business Horizons, 2020, 63(2): 135-146.

[7] Walczyna T, Piotrowski Z. Quick Overview of Face Swap Deep Fakes[J]. Applied Sciences, 2023, 13(11): 6711.

[8] https://www.sohu.com/a/www.sohu.com/a/772380328_121124358.

[9] http://news.enorth.com.cn/system/2024/04/03/055824429.shtml.

[10] https://baijiahao.baidu.com/s?id=1805161055691666482&wfr=spider&for=pc.

[11] https://baijiahao.baidu.com/s?id=1794379566132922666&wfr=spider&for=pc.

[12] https://www.163.com/dy/article/H2PD6NL0051492T3.html.

[13] https://news.cctv.com/2024/07/24/ARTI2X4oEhscKg0anidk4Sve240724.shtml.

[14] https://www.court.gov.cn/jianshe/xiangqing/435431.html.

[15] Ciftci U A, Demir I, Yin L. Fakecatcher: Detection of synthetic portrait videos using biological signals[J]. IEEE transactions on pattern analysis and machine intelligence, 2020.

[16] Li Y. Exposing deepfake videos by detecting face warping artif acts[J]. arXiv preprint arXiv:1811.00656, 2018.

[17] Masood M, Nawaz M, Malik K M, et al. Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward[J]. Applied Intelligence, 2023, 53(4): 3974-4026.

[18] Pei G, Zhang J, Hu M, et al. Deepfake Generation and Detection: A Benchmark and Survey[EB/OL]. (2024-03-26)[2024-09-27]. https://arxiv.org/abs/2403.17881v4.

[19] Rana M S, Nobi M N, Murali B, et al. Deepfake Detection: A Systematic Literature Review[J]. IEEE Access, 2022, 10: 25494-25513.

[20] 谢元坤, 程皓楠, 叶龙. 深度伪造音频检测综述[J]. 中国传媒大学学报(自然科学版), 2024, 31(3): 26-33.

[21] Bevinamarad P R, Shirldonkar M S. Audio forgery detection techniques: Present and past review[C]//2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184). IEEE, 2020: 613-618.

[22] Haliassos A, Vougioukas K, Petridis S, et al. Lips Don't Lie: A Generalisable and Robust Approach To Face Forgery Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 5039-5049.

[23] Prajwal K R, Mukhopadhyay R, Namboodiri V P, 等. A lip sync expert is all you need for speech to lip generation in the wild[C]//Proceedings of the 28th ACM international conference on multimedia. 2020: 484-492.

[24] Zhou Y, Han X, Shechtman E, et al. MakeltTalk: speaker-aware talking-head animation[J]. ACM Transactions on Graphics, 2020, 39(6): 1-15.

[25] Wang J, Qian X, Zhang M, 等. Seeing What You Said: Talking Face Generation Guided by a Lip Reading Expert[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 14653-14662.

[26] Lee D, Kim C, Yu S, 等. RADIO: Reference-Agnostic Dubbing Video Synthesis[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, 2024: 4156-4166.

[27] Zhou H, Sun Y, Wu W, 等. Pose-controllable talking face generation by implicitly modularized audio-visual representation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 4176-4186.

[28] Liang B, Pan Y, Guo Z, et al. Expressive Talking Head Generation With Granular Audio-Visual Control[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3387-3396.

[29] Wang J, Zhao K, Zhang S, et al. LipFormer: High-Fidelity and Generalizable Talking Face Generation With a Pre-Learned Facial Codebook[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 13844-13853.

[30] DAE-Talker: High Fidelity Speech-Driven Talking Face Generation with Diffusion Autoencoder | Proceedings of the 31st ACM International Conference on Multimedia[EB/OL]. [2024-11-14]. https://dl.acm.org/doi/abs/10.1145/3581783.3613753.

[31] Ma Y, Zhang S, Wang J, 等. DreamTalk: When Emotional Talking Head Generation Meets Diffusion Probabilistic Models[A]. arXiv, 2024.

[32] Zhang B, Zhang X, Cheng N, 等. EmoTalker: Emotionally Editable Talking Face Generation via Diffusion Model[C]//ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024: 8276-8280.

[33] Shen S, Li W, Zhu Z, 等. Learning Dynamic Facial Radiance Fields for Few-Shot Talking Head Synthesis[A]//arXiv e-prints. 2022.

[34] Kim M, Ko J, Cho K, 等. AE-NeRF: Auto-Encoding Neural Radiance Fields for 3D-Aware Object Manipulation[A]. arXiv, 2022.

[35] Peng Z, Hu W, Shi Y, et al. SyncTalk: The Devil is in the Synchronization for Talking Head Synthesis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 666-676.

[36] 赖志茂, 章云, 李东. 基于 Transformer 的人脸深度伪造检测技术综述[J]. 广东工业大学学报, 2023, 40(6): 155-167.

[37] Rossler A, Cozzolino D, Verdoliva L, 等. FaceForensics++: Learning to Detect Manipulated Facial Images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1-11.

[38] Li Y, Yang X, Sun P, 等. Celeb-df: A large-scale challenging dataset for deepfake forensics[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 3207-3216.

[39] Li Y, Yang X, Sun P, 等. Celeb-df (v2): a new dataset for deepfake forensics [j][J]. arXiv preprint arXiv, 2019.

[40] Zi B, Chang M, Chen J, et al. WildDeepfake: A Challenging Real-World Dataset for Deepfake Detection[A]. arXiv, 2024.

[41] Dolhansky B, Bitton J, Pflaum B, 等. The DeepFake Detection Challenge (DFDC) Dataset[A]. arXiv, 2020.

[42] He Y, Gan B, Chen S, et al. ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 4360-4369.

[43] Kwon P, You J, Nam G, et al. KoDF: A Large-Scale Korean DeepFake Detection Dataset[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10744-10753.

[44] Khalid H, Tariq S, Kim M, 等. FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset[A]. arXiv, 2022.

[45] Glitch in the matrix: A large scale benchmark for content driven audio–visual forgery detection and localization[J]. Computer Vision and Image Understanding, 2023, 236: 103818.

[46] Yang W, Zhou X, Chen Z, 等. AVoiD-DF: Audio-Visual Joint Learning for Detecting Deepfake[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 2015-2029.

[47] Chollet F. Xception: Deep Learning With Depthwise Separable Convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251-1258.

[48] Li L, Bao J, Zhang T, 等. Face X-Ray for More General Face Forgery Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5001-5010.

[49] Dang H, Liu F, Stehouwer J, et al. On the Detection of Digital Face Manipulation[J].

[50] Zhao H, Zhou W, Chen D, et al. Multi-Attentional Deepfake Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2185-2194.

[51] 尚志华. 视频人脸深度伪造检测关键技术研究[D]. 中国科学技术大学, 2024.

[52] Zhao X, Yu Y, Ni R, 等. Exploring Complementarity of Global and Local Spatiotemporal Information for Fake Face Video Detection[C]//ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2022: 2884-2888.

[53] Yang X, Li Y, Lyu S. Exposing Deep Fakes Using Inconsistent Head Poses[C]//ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019: 8261-8265.

[54] Lutz K, Bassett R. DeepFake Detection with Inconsistent Head Poses: Reproducibility and Analysis[A]. arXiv, 2021.

[55] Li Y, Chang M C, Lyu S. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking[C]//2018 IEEE International Workshop on Information Forensics and Security (WIFS). 2018: 1-7.

[56] DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern | IEEE Journals & Magazine | IEEE Xplore[EB/OL]. [2024-11-14].

[57] Dosovitskiy A, Beyer L, Kolesnikov A, 等. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[A]. arXiv, 2021.

[58] Zheng Y, Bao J, Chen D, et al. Exploring Temporal Coherence for More General Video Face Forgery Detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 15044-15054.

[59] Guan J, Zhou H, Hong Z, et al. Delving into Sequential Patches for Deepfake Detection[J]. Advances in Neural Information Processing Systems, 2022, 35: 4517-4530.

[60] Agarwal S, Farid H, Fried O, et al. Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 2814-2822.

[61] Zhou Y, Lim S N. Joint Audio-Visual Deepfake Detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 14800-14809.

[62] Cozzolino D, Pianese A, Nießner M, et al. Audio-Visual Person-of-Interest DeepFake Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 943-952.

[63] Raza M A, Malik K M. Multimodaltrace: Deepfake Detection Using Audiovisual Representation Learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 993-1000.

[64] Nadimpalli A V, Rattani A. On improving cross-dataset generalization of deepfake detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 91-99.

[65] Chen L, Zhang Y, Song Y, 等. Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 18710-18719.

[66] Lin L, He X, Ju Y, 等. Preserving fairness generalization in deepfake detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16815-16825.

[67] He K, Chen X, Xie S, et al. Masked Autoencoders Are Scalable Vision Learners[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 16000-16009.

[68] Georgescu M I, Fonseca E, Ionescu R T, et al. Audiovisual Masked Autoencoders[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 16144-16154.

[69] Gong Y, Rouditchenko A, Liu A H, et al. Contrastive Audio-Visual Masked Autoencoder[A]. arXiv, 2023.

[70] Feng C, Chen Z, Owens A. Self-Supervised Video Forensics by Audio-Visual Anomaly Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10491-10503.

[71] Cai Z, Ghosh S, Stefanov K, et al. MARLIN: Masked Autoencoder for Facial Video Representation LearnINg[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 1493-1504.

[72] Hussain S, Neekhara P, Jere M, et al. Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021: 3348-3357.

[73] Gandhi A, Jain S. Adversarial perturbations fool deepfake detectors[C]//2020 International joint conference on neural networks (IJCNN). IEEE, 2020: 1-8.

[74] Neekhara P, Dolhansky B, Bitton J, 等. Adversarial threats to deepfake detection: A practical perspective[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 923-932.