不同面部遮挡范围对换脸检测的影响

邹煜 缪志成 张雨禾 周致行

江苏警官学院；江苏南京；210031；

摘要：随着深度学习技术的迅猛发展，人像伪造技术如深度伪造（Deepfake）变得越来越普遍，对个人隐私和社会稳定构成了严重威胁。本研究聚焦于伪造人像检测领域，以面部遮挡物为切入点，重点探究在口罩、墨镜、帽子等常见遮挡物覆盖下，经过深伪技术处理的图片在现有检测软件识别下的检出率。通过构建包含多种遮挡情境的换脸图片数据集，并量化分析了不同遮挡条件下换脸图片的检测成功率。研究发现，面部遮挡物对检测结果有显著影响，部分遮挡情境下检测软件的识别准确率大幅下降，表明现有检测技术在面对复杂遮挡场景时存在局限性。本研究为优化伪造人像检测算法提供了新的视角和数据支持，也为应对日益复杂的伪造图像问题提供了理论依据。

关键词：深度学习；人像伪造检测；面部遮挡

DOI：10.69979/3041-0673.24.12.053

随着深度学习技术的迅猛发展，深度伪造（Deepfake）技术已经达到了前所未有的逼真程度。这种技术利用生成对抗网络（GAN）等算法，生成高度逼真的人脸图像和视频，已广泛应用于娱乐、广告、教育等多个领域，但其潜在的风险也日益凸显，对个人隐私、社会稳定及国家安全构成了严重威胁。为了减少深度伪造技术的负面影响，营造良好网络生态环境，国家互联网信息办公室于 2019 年 12 月 15 日发布《网络信息内容生态治理规定》，提出对网络信息发布内容的监管要求，明确规定发布内容的用途、使用范围、传播路径等需要清晰标记。于 2023 年 1 月 10 日施行《互联网信息服务深度合成管理规定》，明确了监管对象为深度合成内容，落实深度合成内容提供者的主体责任，不得发布法律法规禁止内容的信息，不得侵害他人的合法权益，更不能引起恐慌，破坏社会秩序。

不法分子利用深度伪造技术制造虚假新闻、诽谤他人、进行诈骗等活动，给社会带来了极大的负面影响，因而对其检测显得尤为重要。李旭嵘[1]对深伪技术和检测方法做了详细综述，如表 1 所示。对于深伪视频的检测，李梓楷[2]团队提出了一种基于帧间量化参数强度值的检测方法，认为可通过对视频帧间量化参数强度值的分析来分辨真伪。在司法实践中，陆璟妍[3]提出了基于细粒度图像分类的图像鉴定方法。除了常规的深伪与检测互相攻防，卫霞[4]提出了一种基于区块链技术对抗深度伪造的方法。然而，由于伪造技术的不断进化和多样化，现有的检测方法仍面临诸多挑战，如泛化性不足、对新型伪造手段的适应性差等问题。

在传统人像比对中，面部遮挡对比对结果有较大影响。在现实生活中，面部遮挡一般有两种情况：物体遮挡和阴影（强光）遮挡。其中物体遮挡最为常见，如口罩遮、围巾、墨镜等引起的遮挡。相对于阴影（强光）遮挡尚可通过算法来恢复部分面部信息，物体遮挡造成的面部信息缺失是不可逆的。由此，面部遮挡对于人像比对造成了巨大挑战。

本文以面部遮挡物为切入点，重点探究在口罩、墨镜、帽子等常见遮挡物覆盖下，经过深伪技术处理的图片在现有检测软件识别下的检出率。构建包含多种遮挡情境的换脸图片数据集，并量化分析了不同遮挡条件下换脸图片的检测成功率。本文的研究具有重要的理论意义和应用价值，将为维护网络空间的安全和稳定提供有力保障。

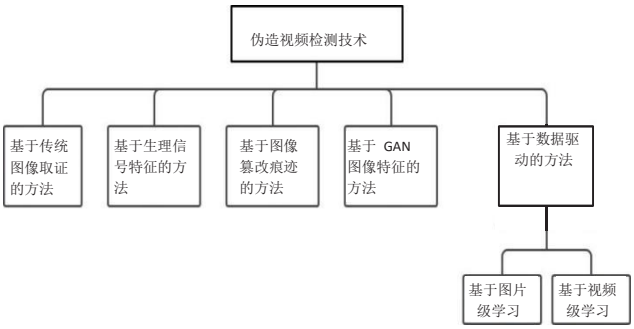

表 1　对深度伪造检测技术的分类

1 基于重绘贴合度检测伪造人像图片方法

1.1 Stable Diffusion 重绘强度值的选取

本次实验采用 Stable Diffusion 生成的图片，实验中采用了 Stable Diffusion 的外接 ReActor 插件，这是一个 Stable Diffusion 的扩展，允许用户从参考

照片中复制脸部到使用 Stable Diffusion 生成的图像中。

在 ReActor 插件中，“重绘幅度”（Redrawing Strength）是一个重要的参数，它影响着换脸过程中图像的生成效果，其作用如下：1. 控制细节保留程度：重绘幅度决定了在换脸过程中，原图中的细节被保留的程度。重绘幅度越小，保留的原图细节越多，换脸后的图像与原图在非面部区域的相似度越高。2.影响图像质量：如果重绘幅度设置得过低（小于 0.25），可能会导致图片边缘模糊的问题。而如果设置得过高，则可能会导致换脸后的图像与原图差异较大，失去原图的一些细节特征。3.平衡相似度与差异：生成与原图相似度较高的图片时，重绘幅度应保持一个较低的值，而如果需要生成更精密的图片，不保留原图过多的细节，应提高重绘幅度。因算法问题，当重绘幅度过高时，会出现图像变形的情况，直观表现为图中文字扭曲，边界模糊。如果要对运用 Stable Diffusion 进行换脸的图片进行研究，就要首先找到合适的重绘幅度区间。

表 2 是对每种重绘幅度带来效果的测试。当重绘幅度大于 0.5 时图形已明显扭曲，不予考虑。

表 2 重绘幅度与检出率关系

重绘幅度	0	0.1	0.2	0.3	0.4	0.5
检出率	100%	100%	100%	2.3%	0%	0%

图 1 真伪检测结果突变

从表中可以看出，重绘幅度变化对于真伪检测的影响是突变式的。从 0.2 到 0.3 这个区间，检测成功率急速下跌，如图 1 所示。在测试中，重绘幅度为 0.3 时，篡改指标已经很少少于 2。重绘幅度较低时，能很好地保留原本人脸特征，直观上肉眼难以分辨真伪。当重绘幅度较高时，图像又会过度扭曲，从而展现出肉眼就能轻易分辨的不和谐之处。在

在重绘幅度较低时，脸部模型被较为完好地移动到目标人脸上，因而会显得有些僵硬。表现在技术层面就是光照角度不同，边缘伪影不符合真实情况等等。在重绘幅度逐渐升高的过程中，换脸算法在处理脸部细节时更为精细，部分面部表情、照明等被“偏转”，包括面部表情、光照条件和肤色等，使得生成的脸部与原始图像在视觉上更加一致，有效消除了生成图像的混合伪影，使其难以被检测软件识别出不一致性。高重绘幅度下，面部修复技术被更多使用，如 GFPGAN（Generative Facial Priors with Adaptive Network），这种技术能够对换脸后的图像进行细节上的优化和修复，进一步提升换脸图像的真实度。除了边界处理，还有对局部身份相关特征的增强，如嘴唇、眉毛等区域的精细调整，这样的局部优化有助于保持身份的一致性，使得换脸结果更加真实。此外频域融合（Frequency Domain Fusion）也被应用到该领域，如 DCT 融合和小波融合，将图像从像素域转换到频率域进行处理，通过融合不同频率分量的信息来增强图像质量。这种融合方法可以在不损失图像细节的情况下提高换脸的真实性，增加检测难度。

2 口罩对面部 AI 换脸识别的影响

面部遮挡物是 AI 换脸过程中难以克服的问题。AI 换脸技术需要精确地识别和匹配脸部特征点，脸部遮挡物如眼镜、帽子、手等会干扰这一过程，增加技术实现的难度。为了解决遮挡物问题，研究者们开发了专门的网络结构，如 Face Shifter 中的 HEAR-Net，专注于解决目标图像脸部遮挡问题，并进一步优化换脸效果。尽管 AI 换脸技术在逼真度和高清化方面取得了进展，但在处理遮挡物、光影变化、表情自然度等方面仍存在技术瓶颈。因此，可以尝试在这一方面探查出图像真伪。以下是在面部有遮挡的情况下进行的真伪测试。

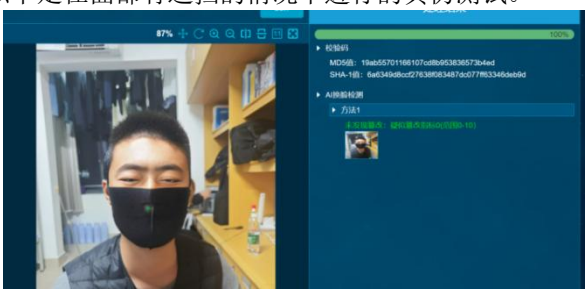

图 2 遮挡物为口罩

在面部遮挡物为口罩时，如图 2 所示，由于口罩遮挡了面部的关键特征，如鼻子和嘴巴，传统的特征提取方法如主成分分析（PCA）、线性判别分析（LDA）、局部二值模式（LBP）等在这种情况下部分失效，导致特征提取困难。此时，软件自动检测已经难以满足需求。在测试中，由于特征量过少，所有图片均无法通过一键检测的方式检测出来。

由以上检测结果可知，口罩覆盖对寻常面部 AI 换脸识别的影响极大。面部识别技术主要依赖于分析人脸的面部特征，如鼻子、嘴巴等部位的形状、大小、位置等信息来进行身份识别。当人脸被口罩遮挡时，这些关

键特征中的大部分被隐藏，从而大大增加了识别的难度。戴口罩的图像更频繁地导致算法无法处理人脸，技术上称为“注册失败或模板缺失”（FTE）。这意味着该算法不能很好地提取出一张脸的特征，不能进行有效的比较。

3 墨镜对面部 AI 换脸识别的影响

墨镜对人脸识别、视觉跟踪和表情分析带来显著挑战，尤其影响面部特征点定位。眼部因其对称性、稳定的两眼间距及低灰度值，一直是研究重点。然而，墨镜遮挡面部上半部分，严重影响识别系统的准确性。实验表明，在一对多人脸识别中，添加墨镜会显著降低识别率。

在深度伪造检测中，300 张伪造图像的检出率为 55%，比口罩遮挡情况下更高。这表明眼部在伪造检测中的重要性低于面部识别。墨镜遮挡关键特征，使基于眼部的伪造检测能力受限。在换脸检测中，眼睛的形态、大小、间距及眼神变化是重要判断依据，而墨镜阻碍系统获取这些特征，影响检测准确性。

此外，实验中还观察到，部分伪造图像中的墨镜色调变浅、透明度升高，而这些图像的检测成功率相对更高。这一现象进一步验证了上述推测，透明度的提升使得更多眼部细节显露，从而增加了检测系统发现伪造破绽的可能性。

相比之下，墨镜的检测比口罩检测更具挑战性。口罩的遮挡范围固定，对面部特征的影响较为单一，经过对抗性训练后，换脸检测系统的识别率可以显著提高。而墨镜的款式、颜色、形状多种多样，并且与面部融合程度较高，使得检测难度进一步加大。尤其是当墨镜完全遮挡眼部时，系统无法利用眼部特征进行辅助判断，导致换脸检测的复杂度显著增加，提升了检测系统的难度。

4 帽子对面部 AI 换脸识别的影响

相较于口罩遮挡的面部下半部分，帽子可能会遮挡住上半部分面部信息，如眉毛和部分额头，使得眉毛的形状和走向等特征无法被准确提取，进而影响识别结果。在定量检测中，使用帽子作为遮挡物的换脸图片，检出率有 27%。该检出率大于口罩组，而小于墨镜组。推测上半部分的头发、眉毛，在检测中所占权重低于眼部区域。此外，帽子遮挡住了头发区域，很多易出现破绽的头发过渡区域均不可见，欺骗性更强，难以通过表象看出真伪。

5 结论

本文深入探究了不同面部遮挡物对面部 AI 换脸识别的影响。研究发现：帽子遮挡下的面部特征保留程度较高，特征融合相对容易，且检测技术的适应性较好，使得检出率处于中间水平；口罩遮挡导致面部特征信息缺失较为严重，特征融合难度较大，且训练数据相对不足，导致检出率最低；墨镜遮挡了关键的眼部区域，特征融合难度最大，检测技术的应用受到限制，但正因为其遮挡的显著性和特征融合的困难，使得换脸后的异常更容易被检测系统识别，检出率最高。

本文研究也存在一定局限性：本文针对 Stable Diffusion 平台合成的深度伪造人像图片，对其他伪造方式的人像图片检测效果尚需验证。此外，本文方法受图片清晰度影响大。尽管如此，探求面部不同面部遮挡范围对换脸检测的影响，可以帮助检验人员了解面部区域所占权重，本文的研究为换脸检测技术的发展提供了实证依据和理论支持。

参考文献

[1]李旭嵘,纪守领,吴春明,等.深度伪造与检测技术综述[J].软件学报,2021,32(02):496-518.
[2]李梓楷，王宇飞，廖广军，等.基于帧间量化参数强度值检测深度伪造人像视频[J].刑事技术,2024,49(01):1-10.
[3]陆璟妍.基于细粒度图像分类的深度伪造视频鉴定研究[D].华东政法大学,2022.
[4]卫霞,白国柱,张文俊.基于区块链技术对抗深度伪造现状研究[J].信息安全研究,2021,7(07):615-620.
基金项目：江苏警官学院大学生创新创业训练计划项目《基于深度学习框架的人像伪造研究》；项目编号：202410329069Y。江苏省高等学校基础科学（自然科学）研究面上项目《基于深度学习的图像伪造检验研究》；项目编号：23KJB620002。痕迹科学与技术公安部重点实验室开放课题《深度伪造人像特征提取研究》；项目编号：2023FMKFKT05。
作者简介：邹煜（2004—），男，汉，江苏常州人，本科在读，单位：江苏警官学院，研究方向：刑事科学技术。