

人体姿态引导的鲁棒生成式隐写^{*}

张清华^{1,2}, 黄方军^{1,2}, 马 宾³

¹(中山大学 网络空间安全学院, 广东 深圳 518107)

²(广东省信息安全技术重点实验室, 广东 广州 510006)

³(算力互联网与信息安全教育部重点实验室 (齐鲁工业大学 (山东省科学院)), 山东 济南 250353)

通信作者: 黄方军, E-mail: huangfj@mail.sysu.edu.cn



摘 要: 传统图像隐写术通过将秘密信息嵌入载体图像中实现信息隐蔽传输, 在信息安全和数据通信领域发挥着重要作用. 然而, 信息嵌入过程不可避免地会修改载体图像, 容易被隐写分析工具检测到. 相比之下, 生成式图像隐写术利用生成模型直接从秘密信息生成隐写图像, 从而避免了修改载体图像的问题. 但现有生成式图像隐写方法在面对各种攻击尤其是几何攻击时普遍存在鲁棒性不足的缺陷, 隐藏的信息易受各种攻击破坏而无法有效提取. 为此, 提出一种人体姿态引导的生成式隐写 (generative steganography guided by human posture, GSHP) 方法. 该方法的核心思想是将秘密信息映射为人体姿态特征, 再将此特征输入生成模型以生成隐写图像. 在信息提取阶段, 通过人体姿态检测算法识别隐写图像中的人体姿态, 进而还原出秘密信息. 由于人体姿态固有的结构稳定性, GSHP 对各种攻击表现出了良好的鲁棒性. 广泛的实验也充分证明了 GSHP 在安全性和鲁棒性方面的优势.

关键词: 生成式隐写; 鲁棒隐写; 人体姿态; 几何不变性; 扩散模型

中图法分类号: TP309

中文引用格式: 张清华, 黄方军, 马宾. 人体姿态引导的鲁棒生成式隐写. 软件学报. <http://www.jos.org.cn/1000-9825/7662.htm>

英文引用格式: Zhang QH, Huang FJ, Ma B. Robust Generative Steganography Guided by Human Posture. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7662.htm>

Robust Generative Steganography Guided by Human Posture

ZHANG Qing-Hua^{1,2}, HUANG Fang-Jun^{1,2}, MA Bin³

¹(School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen 518107, China)

²(Guangdong Provincial Key Laboratory of Information Security Technology, Guangzhou 510006, China)

³(Key Laboratory of Computing Power Network and Information Security (Qilu University of Technology (Shandong Academy of Sciences)), Ministry of Education, Jinan 250353, China)

Abstract: Traditional image steganography achieves covert information transmission by embedding secret information into cover images, playing a crucial role in information security and data communication. However, the information embedding process inevitably alters the cover image, making it more susceptible to detection by steganalysis tools. In contrast, generative image steganography directly synthesizes stego images from secret information using generative models, avoiding modifications to cover images. Nevertheless, existing generative image steganography methods generally suffer from insufficient robustness against various attacks, particularly geometric attacks, often resulting in damage to the hidden information and failure to effectively extract it. To address this issue, this study proposes a generative steganography method guided by human posture (GSHP). The core idea is to map secret information into human posture features, which are then fed into a generative model to produce stego images. During the information extraction stage, a human posture detection algorithm is used to identify the posture in the stego image, thus recovering the hidden information. Owing to the inherent structural

* 基金项目: 国家自然科学基金 (U2336208, 62472454, 62272255); 深圳市科技计划 (JCYJ20250604175534044); 算力互联网与信息安全教育部重点实验室开放课题 (2024ZD022)

收稿时间: 2025-10-18; 修改时间: 2025-12-23, 2026-02-09; 采用时间: 2026-03-04; jos 在线出版时间: 2026-06-03

stability of human posture, GSHP exhibits good robustness against various attacks. Extensive experiments also fully demonstrate the advantages of GSHP in terms of security and robustness.

Key words: generative steganography; robust steganography; human posture; geometric invariance; diffusion model

在信息安全与数据通信领域, 图像隐写术作为实现秘密信息隐蔽传输的关键技术, 长期以来备受关注. 传统图像隐写术通过特定的算法, 将秘密信息巧妙地嵌入载体图像之中. 其核心原理是利用载体图像在视觉感知上的冗余性, 在不显著影响图像视觉质量的前提下, 将秘密信息嵌入图像的像素值^[1-3]、频率域系数^[4-6]等信息中. 然而, 传统图像隐写术存在一个难以回避的缺陷, 即信息嵌入不可避免地会对载体图像进行修改. 尽管这些修改通常被设计得非常细微, 可能在视觉上不可感知, 但仍然会导致载体图像的某些统计特性发生变化, 例如像素值的分布、相邻像素的相关性、频率域的能量分布等. 隐写分析工具^[7-9]正是通过检测这些统计差异, 来判断图像中是否隐藏了秘密信息, 从而对传统隐写术的安全性造成威胁.

为了突破传统隐写术依赖载体图像修改的局限性, 生成式图像隐写技术应运而生. 该技术借助生成模型直接从秘密信息生成隐写图像. 其核心思想是将秘密信息编码为生成模型的条件特征^[10-15]、潜变量^[16-22]或者采样概率^[23,24], 然后通过生成模型输出包含秘密信息的图像. 这种方法无需对已有的载体图像进行修改, 从根本上避免了因修改载体图像而留下隐写痕迹, 大大降低了被隐写分析工具检测到的可能性, 为信息隐蔽传输提供了新的思路.

相较于将秘密信息直接编码为生成模型的潜变量或干预其采样概率, 基于特征映射的生成式图像隐写通常将秘密信息转换为图像语义标签或结构特征, 作为附加条件输入生成模型, 从而生成符合条件要求的隐写图像. 刘明明等人^[10]提出了一种基于生成对抗网络的生成式隐写方法, 该方法将秘密信息映射为类标签以驱动辅助分类器生成对抗网络 (auxiliary classifier generative adversarial network, ACGAN) 生成隐写图像. 在接收端利用判别器对隐写图像进行分类, 进而提取秘密信息. Zhang 等人^[25]同样基于 ACGAN 提出使用类标签生成隐写图像的方法. 发送者可以通过字典将秘密信息映射到分类标签中, 然后使用标签生成隐写图像. 接收者可以通过模型识别图像标签, 获取秘密信息. 随后, Zhang 等人^[11]设计了一种合成语义隐写生成对抗网络 (synthetic semantics stego generative adversarial network, SSS-GAN). 发送方通过建立秘密信息和语义类别信息之间的映射关系, 然后利用类标签在生成模型上生成伪真实图像. 接收方可以通过分类器网络识别类标签, 以恢复秘密信息. Xue 等人^[26]提出一种基于多域图像转化的无载体图像隐写框架, 称为 MDI (message drives image). 该框架在转化的目标标签和秘密信息之间建立映射关系, 然后根据目标标签将载体图像转化为隐写图像. 在信息提取阶段, 接收方使用多标签分类器从隐写图像中提取标签, 并根据映射规则恢复秘密信息. 以上生成式隐写方法通过建立秘密信息和分类标签的映射关系实现, 在鲁棒性上都存在一定的缺陷.

Chen 等人^[12]提出了一种基于图像选择和星生成对抗网络 (star generative adversarial network, StarGAN) 的无载体隐写方法, 提高了隐写方法的容量和鲁棒性. 该方法使用传统的无载体隐写方法从数据库中选择一张自然图像来表示秘密信息的第 1 部分, 然后在剩余秘密信息和人脸属性之间建立映射, 最后利用 StarGAN 生成具有映射关系的隐写图像. Liu 等人^[27]提出了一种用于隐写的图像解纠缠自动编码器 (image disentanglement autoencoder for steganography, IDEAS). 该方法将图像分为结构和纹理的两种表示形式, 然后利用结构表示的稳定性来提高隐写方法的鲁棒性, 同时通过随机纹理表示来增加合成多样性, 以增强隐写方法的安全性. Sun 等人^[14]提出了一种基于引导特征的生成式图像隐写, 并提出了一种秘密融合算法来解决引导特征提取错误的问题. 接收方可以直接从合成图像中提取引导特征, 进而提取秘密信息. Zhou 等人^[13]提出了一种基于自动生成语义对象轮廓的生成隐写框架. 该隐写框架将秘密信息编码为语义对象轮廓, 然后将轮廓转换为相应的图像以进行隐蔽通信. 由于对象轮廓具有良好的稳定性, 该方法对多种攻击都表现出良好的鲁棒性. Rehman 等人^[15]提出了一种基于动漫角色生成的无载体信息隐藏方法, 该方法将秘密信息转换为动漫角色的属性标签集, 并使用该标签集直接生成动漫角色, 具有良好的容量和鲁棒性.

尽管以上生成式隐写方法致力于提高自身的鲁棒性, 但是在面临各类攻击尤其是几何攻击时, 由于隐写图像的结构特征容易遭到破坏, 进而可能影响秘密信息的提取. 针对现有生成式图像隐写存在的鲁棒性不足问题, 本文

提出一种人体姿态引导的生成式隐写 (generative steganography guided by human posture, GSHP) 方法. 该方法的核心在于设计了一种新的映射机制, 将秘密信息编码为特定的人体姿态特征. 随后, 将编码后的姿态特征作为条件输入, 引导预训练的生成模型生成包含目标姿态的人体图像, 即隐写图像. 在信息提取阶段, 为了恢复隐藏的秘密信息, 需要利用人体姿态检测算法识别隐写图像中人体的关键点位置及其空间关系, 从而重建出图像中的人体姿态特征. 最后, 利用在嵌入阶段设定的编码规则将人体姿态特征转换回原始的秘密信息. 人体姿态是一种具有丰富表达能力和较强稳定性的结构特征. 一方面, 人体姿态能够在不显著影响图像整体视觉效果的前提下进行多种变化, 进而承载丰富的秘密信息; 另一方面, 人体姿态的节点位置和肢体结构具有较强的抗干扰能力, 在面对常见的图像攻击时, 能够保持稳定. 因此, GSHP 能够在保证隐写隐蔽性的同时, 增强对各类攻击的鲁棒性, 为信息的安全可靠传输提供有效的解决方案.

本文第 1 节介绍姿态引导人物图像生成和人体姿态识别的相关工作. 第 2 节从映射构建、图像生成和信息提取这 3 个阶段对 GSHP 方法进行介绍. 第 3 节通过实验分析 GSHP 的视觉质量、安全性和鲁棒性. 第 4 节总结全文.

1 相关工作

本文所提方法主要基于姿态引导人物图像生成和人体姿态估计, 下面就相关工作和基本理论予以介绍.

1.1 姿态引导人物图像生成

姿态引导人物图像生成技术致力于根据目标姿态生成逼真的人物影像, 目前已成为计算机视觉领域的研究热点. Ma 等人^[28]提出的经典方法将条件图像与目标姿态输入生成网络, 先产生带有目标姿势的初始模糊图像, 再通过对抗训练优化生成结果, 最终得到逼真影像. Siarohin 等人^[29]在生成对抗网络 (generative adversarial network, GAN) 架构中引入可变形跳跃连接, 有效解决了姿态差异导致的像素错位问题, 同时提出最近邻损失函数以实现生成图像与目标图像的细节匹配. Esser 等人^[30]提出一种条件 U-Net 架构, 可用于形状引导的图像生成. 该方法以外观变分自编码器 (variational autoencoder, VAE) 的潜在表示为条件约束, 支持端到端训练且无需依赖同一对象的多姿态或多外观样本. Zhu 等人^[31]设计的姿态转移 GAN 通过一系列姿势注意力转移模块, 实现了外观一致性更优的人物姿态迁移效果. Men 等人^[32]提出的属性分解 GAN 则将人体属性作为独立编码嵌入潜空间, 通过风格表示的混合与插值操作实现属性的连续可控生成.

然而, 现有方法仍存在显著局限性: 基于 GAN 的生成模型易出现纹理变形与身体结构失真, 且样本多样性不足; Esser 等人^[30]提出的条件 U-Net 架构虽无需多姿态样本即可实现端到端训练, 但生成的图像存在细节模糊问题. 为突破上述瓶颈, Zhang 等人^[33]提出一种用于大型预训练的文本到图像扩散模型的条件控制方法, 称为 ControlNet. 该方法通过锁定预训练扩散模型的核心参数并复制其编码层以创建可训练副本, 然后通过权重初始化为零的卷积层将可训练副本与原始锁定模型连接, 确保训练初期不向大模型深层特征引入有害噪声, 并保护主干网络免受损害. 通过这种方式, ControlNet 在保留大模型生成能力的同时实现了对目标姿态的高精度引导. 鉴于 ControlNet 在可控图像生成任务中的卓越性能, 本文将其引入姿态引导的生成式隐写框架, 以生成高视觉质量的隐写图像.

1.2 人体姿态估计

人体姿态估计旨在从图像或视频中定位人体关键点并构建姿态表示, 近年来广受研究者关注. 早期研究聚焦于单人姿态估计, 其目标是在单人图像中定位人体关键点位置, 相关方法主要分为回归法和热图法两类. 回归法^[34,35]采用端到端框架学习从输入图像到关键点坐标或人体模型参数的直接映射, 热图法^[36,37]则通过预测身体部位与关节的概率分布实现位置估计并以热图作为监督信号.

相比之下, 多人姿态估计覆盖更广泛的应用场景, 但也面临更高的技术挑战, 尤其是关键点与不同个体的匹配问题, 其研究主要分为自上而下和自下而上两种范式. 自上而下法^[38,39]先利用预训练人体检测器从图像中提取个体边界框, 再对每个框内目标独立应用单人姿态估计模型以组合得到多人姿态; 自下而上法^[40,41]先检测图像中所有关键点, 再通过关联算法将关键点分组到不同个体. 由于无需逐人检测, 其计算效率通常高于自上而下法.

Cao 等人^[41]提出的 OpenPose 作为典型的自下而上框架,引入非参数化的部件亲和场 (part affinity field, PAF) 显式编码肢体位置和方向信息,通过多阶段网络联合学习关键点检测与关联,有效提升了多人姿态估计的速度和精度. 鉴于 OpenPose 在速度与精度上的显著优势,本文采用该框架从隐写图像中提取人体姿态信息,进而用于秘密信息的恢复.

2 隐写方法

针对现有生成式图像隐写存在的鲁棒性不足问题,本文提出一种人体姿态引导的生成式隐写方法. 该方法将不同的秘密信息映射为具有不同人体姿态特征的隐写图像,主要包含 3 个阶段,即映射构建、图像生成和信息提取. 在映射构建阶段,秘密信息被编码为特定的人体姿态特征;在图像生成阶段,利用编码后的姿态特征引导预训练的生成模型生成包含目标姿态的人物图像,即隐写图像;在信息提取阶段,利用人体姿态检测算法识别隐写图像中的人体姿态特征,然后根据已构建的映射规则将人体姿态特征转换回原始的秘密信息. 具体隐写流程如图 1 所示.

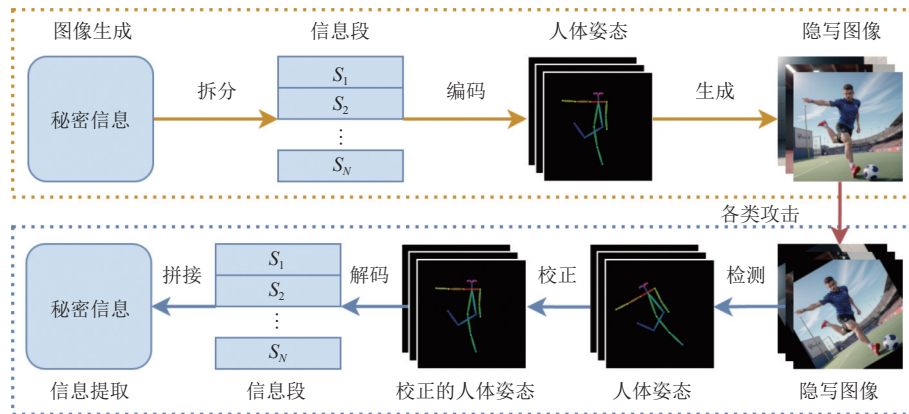


图 1 人体姿态引导的生成式隐写流程

2.1 映射构建

不同的人体姿态检测算法得到的人体姿态的关键点数量不同,关键点数量越多对于人体姿态的描述就越精细,越少对人体姿态的描述就越概括. 本文将秘密信息编码为人体姿态以进行生成式隐写,为了让得到的人体姿态在面对各种攻击时具有良好的鲁棒性,使用的关键点不宜过于精细. 本研究中,我们使用了 OpenPose^[41]中 18 个关键点的方案,包括面部的 5 个关键点(鼻、右眼、左眼、右耳、左耳)、躯干的 5 个关键点(颈、右肩、左肩、右臀、左臀)、上肢的 4 个关键点(右肘、左肘、右腕、左腕)和下肢的 4 个关键点(右膝、左膝、右踝、左踝),如图 2 所示.

具体的人体姿态分类规则如表 1 所示. 表 1 第 1 列将人体划分为 4 个区块,分别为躯干、面部、上肢和下肢. 第 2 列展示了每个区块包含的人体部位. 第 3 列展示了每个部位可以活动的总范围,以角度表示,具体的角度计算方法我们在本节后续逐个讨论. 第 4、5 列展示每个部位可能的状态及对应的角度范围. 本文将每个人体部位可以活动的总范围平均划分为 3 部分,每部分对应一种状态,同一部位的所有状态对应的角度范围是等大的. 发送方在构造人体姿态时,根据每个部位的状态在其对应的角度范围内随机选择一个角度;接收方在检测人体姿态时,根据每个部位所在的角度范围确定对应的状态.

通常而言,一幅图像可以从任意角度呈现人体姿态,但当人体躯干以左侧或右侧呈现时(即人体躯干的左侧或右侧面向屏幕时),人体左侧和右侧的关键点容易重叠;当人体躯干以背面呈现时,面部的关键点难以识别,从而影响图像的生成与检测. 因此,我们将人体躯干正向面对屏幕作为基准用于构建映射,如图 3 所示. 人体各个关键点

从躯干向外延伸, 其角度受到各个关节的活动范围限制, 都可以表示为相对于躯干的角度. 为了避免躯干的角度变化增加其他关键点构造和检测的复杂性, 我们让躯干处于竖直状态, 躯干的关键点不用于编码信息. 将左右臀中点到左右肩中点的连接线作为躯干的中心线, 让躯干的中心线与水平线 (箭头指向屏幕右侧) 的角度为 90° .

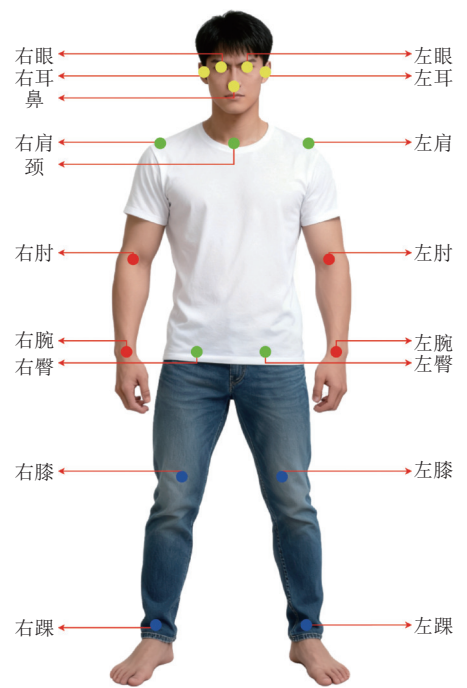


图 2 人体姿态关键点

表 1 人体姿态分类规则

区块	部位	总范围	状态	角度范围
躯干	颈、右肩、左肩、右臀、左臀	—	竖直	—
面部	鼻、右眼、左眼、右耳、左耳	$[45^{\circ}, 135^{\circ}]$	竖直	$[75^{\circ}, 105^{\circ})$
			右倾	$[105^{\circ}, 135^{\circ}]$
			左倾	$[45^{\circ}, 75^{\circ})$
	右肘	$[0^{\circ}, 360^{\circ})$	水平	$[120^{\circ}, 240^{\circ})$
			上旋	$[0^{\circ}, 120^{\circ})$
			下旋	$[240^{\circ}, 360^{\circ})$
	左肘	$[0^{\circ}, 360^{\circ})$	水平	$[0^{\circ}, 60^{\circ}) \cup [300^{\circ}, 360^{\circ})$
			上旋	$[60^{\circ}, 180^{\circ})$
			下旋	$[180^{\circ}, 300^{\circ})$
	右腕	$[-150^{\circ}, 150^{\circ}]$ (相对于右肘的角度)	平行	$[-50^{\circ}, 50^{\circ})$
			逆时针旋转	$[50^{\circ}, 150^{\circ}]$
			顺时针旋转	$[-150^{\circ}, -50^{\circ})$
下肢	左腕	$[-150^{\circ}, 150^{\circ}]$ (相对于左肘的角度)	平行	$[-50^{\circ}, 50^{\circ})$
			逆时针旋转	$[50^{\circ}, 150^{\circ}]$
			顺时针旋转	$[-150^{\circ}, -50^{\circ})$
	右膝	$[180^{\circ}, 360^{\circ}]$	竖直	$[240^{\circ}, 300^{\circ})$
			右旋	$[180^{\circ}, 240^{\circ})$
			左旋	$[300^{\circ}, 360^{\circ}]$

表 1 人体姿态分类规则 (续)

区块	部位	总范围	状态	角度范围
下肢	左膝	[180°, 360°]	竖直	[240°, 300°]
			右旋	[180°, 240°]
			左旋	[300°, 360°]
	右踝	[-150°, 150°] (相对于右膝的角度)	平行	[-50°, 50°]
			逆时针旋转	[50°, 150°]
			顺时针旋转	[-150°, -50°]
	左踝	[-150°, 150°] (相对于左膝的角度)	平行	[-50°, 50°]
			逆时针旋转	[50°, 150°]
			顺时针旋转	[-150°, -50°]

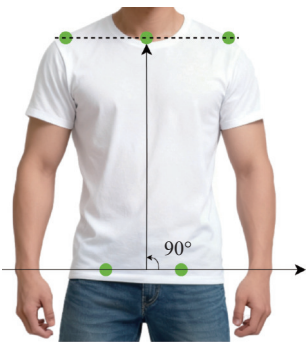


图 3 躯干关键点的分布与角度

当人体躯干正向面对屏幕时,人体面部的活动范围有竖直、右倾和左倾这 3 种状态,如图 4 所示. 竖直状态是指颈到双目中点的连接线与水平线 (箭头向右) 的夹角在 [75°, 105°] 范围内,如图 4(b) 所示; 右倾状态是指夹角在 [105°, 135°] 范围内,如图 4(a) 所示; 左倾状态是指夹角在 [45°, 75°] 范围内,如图 4(c) 所示. 我们可以根据面部所有关键点 (即鼻、右眼、左眼、右耳、左耳) 确定面部状态,并将其编码为一个三进制数,例如面部右倾对应三进制数 0, 面部竖直对应三进制数 1, 面部左倾对应三进制数 2.

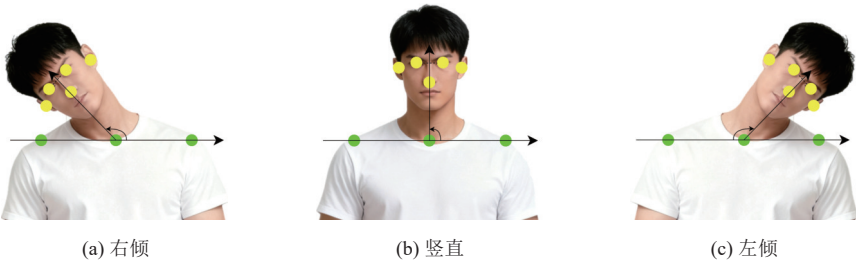


图 4 面部关键点的分布与角度

上肢中的右肘和左肘的活动范围相似且对称,可分为水平、上旋和下旋这 3 种状态,如图 5 所示. 以右肘为例,水平状态是指右肩到右肘的连接线与水平线 (箭头向右) 的夹角在 [120°, 240°] 范围内,如图 5(b) 所示; 上旋状态是指夹角在 [0°, 120°] 范围内,如图 5(a) 所示; 下旋状态是指夹角在 [240°, 360°] 范围内,如图 5(c) 所示. 右肘和左肘可分别编码为一个三进制数,例如右肘上旋对应三进制数 0, 右肘水平对应三进制数 1, 右肘下旋对应三进制数 2, 左肘同理.

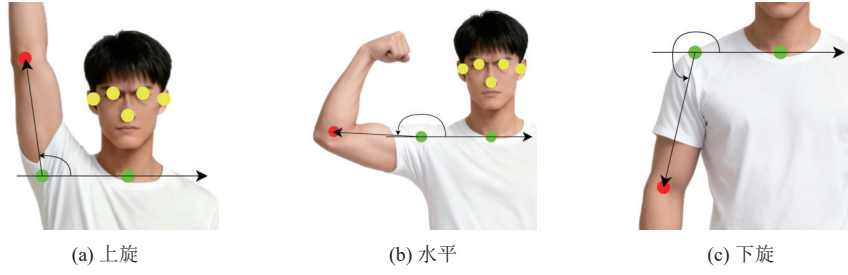


图5 右肘关键点的分布与角度

上肢中的右腕和左腕的活动范围受到右肘和左肘的影响, 相对于肘, 可分为平行、逆时针旋转和顺时针旋转这3种状态, 如图6所示. 以右腕为例, 平行状态是指右肘到右腕的连接线与右肩到右肘的连接线的夹角在 $[-50^\circ, 50^\circ]$ 范围内, 如图6(b)所示; 逆时针旋转状态是指夹角在 $[50^\circ, 150^\circ]$ 范围内, 如图6(a)所示; 顺时针旋转状态是指夹角在 $[-150^\circ, -50^\circ]$ 范围内, 如图6(c)所示. 右腕和左腕可分别编码为一个三进制数, 例如右腕逆时针旋转对应三进制数0, 右腕平行对应三进制数1, 右腕顺时针旋转对应三进制数2, 左腕同理.

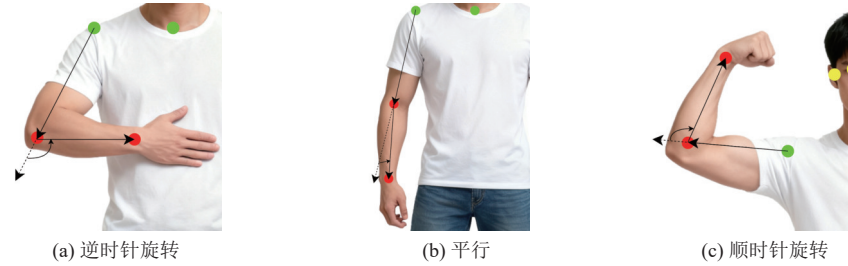


图6 右腕关键点的分布与角度

下肢的关键点与上肢的关键点类似. 下肢中的右膝和左膝的活动范围相似且对称, 可分为竖直、右旋和左旋这3种状态. 以右膝为例, 竖直状态是指右臀到右膝的连接线与水平线(箭头向右)的夹角在 $[240^\circ, 300^\circ]$ 范围内, 右旋状态是指夹角在 $[180^\circ, 240^\circ]$ 范围内, 左旋状态是指夹角在 $[300^\circ, 360^\circ]$ 范围内. 右膝和左膝可分别编码为一个三进制数, 例如右膝右旋对应三进制数0, 右膝竖直对应三进制数1, 右膝左旋对应三进制数2, 左膝同理.

下肢中的右踝和左踝的活动范围受到右膝和左膝的影响, 相对于膝, 可分为平行、逆时针旋转和顺时针旋转这3种状态. 以右踝为例, 平行状态是指右膝到右踝的连接线与右臀到右膝的连接线的夹角在 $[-50^\circ, 50^\circ]$ 范围内, 逆时针旋转状态是指夹角在 $[50^\circ, 150^\circ]$ 范围内, 顺时针旋转状态是指夹角在 $[-150^\circ, -50^\circ]$ 范围内. 右踝和左踝可分别编码为一个三进制数, 例如右踝逆时针旋转对应三进制数0, 右踝平行对应三进制数1, 右踝顺时针旋转对应三进制数2, 左踝同理.

如上所述, 每种人体姿态可以由一个9位的三进制数表示, 共有 $3^9 = 19683$ 种人体姿态. 因此, 所有的人体姿态可以表达至少14位 ($2^{14} = 16384$) 二进制信息, 即每幅图像的隐写容量为14位. 可以使用密钥生成随机种子, 然后使用随机抽样函数在19683种三进制数中任选16384种, 建立三进制数与二进制信息的映射关系, 即可得到人体姿态与二进制信息的映射关系. 考虑到单人单图场景下容量偏低的问题, 可以采取传输多幅图像或在每幅隐写图像包含多个人物等方式. 尽管也可以通过增加手指、脚趾、面部的关键点来增大特征空间, 但这些关键点的生成与检测精度较低, 会严重降低算法的鲁棒性.

2.2 图像生成

假定秘密信息序列 S 的长度为 L_s , 首先, 将秘密信息序列 S 拆分为长度为 L 的信息段 (本文中 $L = 14$), 其数量可以表示为:

$$P = \text{ceil}(L_s/L) \quad (1)$$

其中, $\text{ceil}(\cdot)$ 表示向上取整函数. 如果最后一个信息段的长度小于 L , 则在末尾补 0.

利用密钥 K_1 生成随机种子, 在 $3^9 = 19683$ 种三进制数中随机选取 2^L 种, 建立三进制数与二进制信息的映射关系. 利用该映射关系, 将每个长度为 L 的二进制信息段转化为 9 位的三进制信息段. 使用密钥 K_2 生成随机种子, 然后将三进制信息段的每一位与面部、右腕、左腕、右肘、左肘、右膝、左膝、右踝和左踝进行随机映射. 根据该映射关系和三进制数每一位的值确定上述各个部位的状态, 依据状态信息可以确定各个部位的角度.

在构造人体姿态时, 首先需要确定躯干部位各个关键点的位置. 我们让人体躯干正向面对屏幕, 并让躯干的中心线与水平线 (箭头向右) 的夹角为 90° , 如图 7 所示. 仅角度已知的情况下无法确定关键点的具体位置, 需要进一步确定关键点间的距离. 我们以左肩和右肩为参照, 令它们之间的距离为 $2d$, 其他关键点之间的距离可以设定为左肩和右肩之间距离的一定比例, 并在 $\pm 0.1d$ 范围内随机波动. 当左肩和右肩的位置确定时, 颈的位置即为左肩和右肩的中点. 令颈与左右臀中点之间的距离为 $3d$, 通过颈的位置可以确定左右臀中点的位置. 令左右臀之间的距离为 d , 通过左右臀中点的位置可以进一步确定左臀和右臀的位置. 至此, 躯干部位所有关键点的位置都确定了.

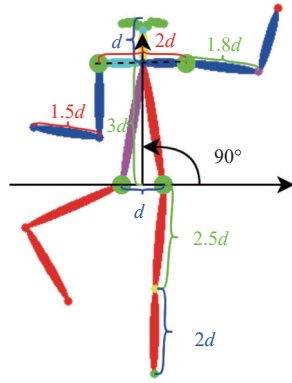


图 7 人体姿态构造示意图

在构造面部关键点时, 令颈与双目中点的距离为 d , 依据面部的角度和颈的位置可以确定双目中点的位置. 令双目之间的距离为 $0.4d$, 然后依据双目中点的位置和面部的角度可以确定双目的位置. 最后, 利用面部其他关键点与双目的相对位置关系确定面部的其他关键点的位置. 在构造右肘时, 令右肘与右肩的距离为 $1.8d$, 依据右肘的角度和右肩的位置可以确定右肘的位置. 在构造左肘时同理. 在构造右腕时, 令右腕与右肘的距离为 $1.5d$, 依据右肘的位置、右肘的角度和右腕的角度可以确定右腕的位置. 在构造左腕时同理. 令左右膝与左右臀的距离为 $2.5d$, 左右踝与左右膝的距离为 $2d$, 左右膝位置的确定与左右肘类似, 左右踝位置的确定与左右腕类似.

当所有的关键点位置确定后, 可以得到一组人体姿态特征. 最后, 将所有二进制信息段对应的人体姿态特征传入生成模型生成 P 幅带有特定姿态的人物的隐写图像. 在生成隐写图像后, 发送端通常会运行一次本地提取以进行自我验证. 如果因生成瑕疵等原因导致提取失败, 该图像将被丢弃并重新生成. 这确保了发出的每一幅隐写图像在理想条件下都是可解码的. 此外, 当单幅图像中包含多个人物时, 可以将人体姿态实例分散布置在图像的不同区域, 以避免关键部位的重叠, 从而确保每个姿态都能被独立、准确地检测.

2.3 信息提取

隐写图像在传输过程中可能受到各种扰动, 导致人物相对于图像的位置和状态发生改变, 进而影响秘密信息的提取. 在构造人体躯干关键点时, 我们将人体躯干正向面对屏幕作为基准, 并让其处于竖直状态, 避免了关键点重叠和人体姿态的混淆, 保证了人体姿态的稳定性. 因此, 在检测人体姿态时, 可以将人体躯干用于校正人体姿态. 首先, 检测每幅隐写图像中的人体姿态, 从中得到一组人体姿态关键点. 然后, 计算每组人体姿态的躯干中心线与水平线 (箭头向右) 的夹角. 若夹角不为 90° , 即人体躯干未处于竖直状态, 则旋转图像使人体躯干处于竖直状态, 如图 8 所示.

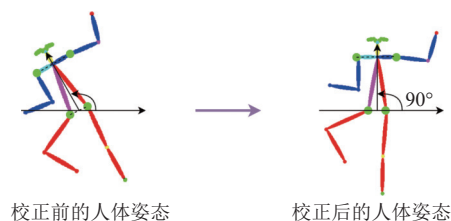


图8 人体姿态校正

对于校正后的人体姿态, 根据每个人体部位的状态所在的角度范围依次确定其状态. 通过计算颈到鼻的连接线的角度, 然后根据表1第5列的范围可以确定面部的状态; 通过计算右肩到右肘的连接线的角度, 然后根据其所处范围可以确定右肘的状态, 左肘同理; 通过计算右肘到右腕的连接线与右肩到右肘的连接线的角度, 然后根据其所处的范围可以确定右腕的状态, 左腕同理; 通过计算右臀到右膝的连接线的角度, 然后根据其所处的范围可以确定右膝的状态, 左膝同理; 通过计算右膝到右踝的连接线与右臀到右膝的连接线的角度, 然后根据其所处的范围可以确定右踝的状态, 左踝同理.

根据所有部位的状态确定其对应的三进制数, 然后利用已建立的映射关系将三进制数转化为秘密信息段. 最后, 连接所有信息段得到秘密信息.

3 实验分析

为了验证提出的 GSHP 的优势, 本节我们将讨论 GSHP 的视觉质量、安全性和鲁棒性. 为了确保生成的图像接近真实风格, 我们使用 Stable Diffusion 1.5 的微调模型 epiCRealism V4^[42]生成隐写图像, 其默认的生成图像大小为 512×512. 具体的实验结果与分析如下.

3.1 视觉质量

GSHP 可以通过提示引导在生成模型上生成各种姿态的人物图像, 且生成的图像具有良好的视觉质量. 样例图像如图9所示, 其中第1行表示生成的人物图像, 第2行表示人物图像对应的人体姿态. 为了进一步量化生成图像的视觉质量, 我们使用 NIQE^[43]、BRISQUE^[44]和 PIQE^[45]这类无参考图像质量评估指标对图像视觉质量进行定量分析. 实验中, 对所有可能的 19683 种姿态每种生成 10 幅图像, 共生成 196830 幅图像. 所有生成图像的视觉质量评估结果平均值如表2所示. 生成图像的 NIQE、BRISQUE 和 PIQE 表现良好, 这表明生成的图像具有良好的视觉质量.



图9 生成的样例图像

表2 生成图像的视觉质量

评估标准	NIQE	BRISQUE	PIQE
范围	<2.5 优秀	<20 优秀	<20 优秀
	2.5-5 良好	20-40 良好	20-40 良好
	5-7.5 一般	40-60 一般	40-60 一般
	>7.5 较差	>60 较差	>60 较差
结果	3.9482	35.5197	39.0345

3.2 安全性

随着图像生成技术的不断发展,生成图像在我们的日常生活中越来越普遍,传输生成图像的行为不会让人感到异常,这为生成式隐写提供了良好的伪装环境.由于GSHP得到的隐写图像是直接生成模型生成的,未对已生成的图像进行任何修改,并且按照映射规则,隐写图像中每个人体部位在其活动范围内出现在各个角度的概率是相同的,这与随机生成的任意姿态的图像具有相同的统计特性,从理论上保证了GSHP生成的隐写图像无法被隐写分析工具检测.

为验证GSHP的安全性,我们使用XuNet^[7]、YeNet^[8]、SRNet^[46]和SiaStegNet^[47]等当前主流隐写分析工具对GSHP进行检测.使用GSHP随机生成10000幅隐写图像,并在人体姿态的活动范围内随机生成10000幅普通图像.训练集、验证集和测试集分别包含8000、1000和1000幅图像.不同隐写分析器的检测错误率见表3,可以看到结果都在0.5附近波动,这表明上述隐写分析器无法正确区分普通生成图像和隐写图像,证明了GSHP的安全性.

表3 不同隐写分析方法的检测错误率

隐写分析工具	XuNet	YeNet	SRNet	SiaStegNet
检测错误率	0.4955	0.4920	0.4975	0.5005

此外,我们采用双样本Kolmogorov-Smirnov(KS)检验,从统计层面进一步验证GSHP生成的隐写图像和普通姿态引导生成图像在人体姿态角度分布上的一致性.该方法是一种非参数检验方法,用于判断两组独立样本是否来自同一个连续分布.通过比较样本累积分布函数的最大差异实现统计推断,并以 p 值衡量两组样本分布差异的统计显著性.我们将隐写图像和普通生成图像对应的人体姿态角度数据作为两组独立样本,随机生成10000幅隐写图像和10000幅普通图像,并对所有图像中人体各部位的角度进行双样本KS检验,得到的检验结果如表4所示.结果显示,所有部位的 p 值均远大于0.05,表明两组图像的人体姿态在各个部位的角度分布上不存在显著统计差异.

表4 隐写图像和普通姿态引导生成图像中人体各个部位角度的双样本KS检验结果

部位	面部	右肘	左肘	右腕	左腕	右膝	左膝	右踝	左踝
p 值	0.8629	0.5457	0.7912	0.2582	0.2582	0.7912	0.9783	0.6042	0.2888

算法实施过程中我们使用密钥 K_1 生成随机种子,在所有位数为 $B=9$ 的三进制数集中随机选取 2^L ($L=14$)个三进制信息段,并依次将其与长度为14的二进制信息段建立映射关系.同时使用密钥 K_2 生成随机种子,然后将三进制信息段的每一位与面部、右腕、左腕、右肘、左肘、右膝、左膝、右踝和左踝进行随机映射.由于 B 位三进制信息段与 L 位二进制信息段之间、三进制信息段中的每一位与身体各个部位(即面部、右腕、左腕、右肘、左肘、右膝、左膝、右踝和左踝)之间均为随机映射,因此对GSHP进行穷举攻击的计算复杂度为:

$$N = C_{3^B}^{2^L} \cdot A_B^B = \frac{(3^9)! \cdot 9!}{(2^{14})! \cdot (3^9 - 2^{14})!} \quad (2)$$

其中, A_x^y ($x \geq y$) 表示从 x 个不同样本中有序选取 y 个样本的排列数, C_x^y ($x \geq y$) 表示从 x 个不同样本中选取 y 个样本的组合数, $x!$ 表示 x 的阶乘.攻击GSHP的计算复杂度将达到 10^{3867} 数量级,所以难以从隐写图像中推断秘密信息.

若要进一步提在每一幅隐写图像中的秘密信息传输数量,可以在每一幅隐写图像中包含多个人物信息.我们在构建映射时分别使用密钥 K_1 和 K_2 生成与隐写图像中人物数量相同的随机种子.此时,假定一幅隐写图像中

人物的数量为 H , 攻击 GSHP 的计算复杂度可以表示为:

$$N = (C_{3^b}^{2^L} \cdot A_B^B)^H = \left(\frac{(3^9)! \cdot 9!}{(2^{14})! \cdot (3^9 - 2^{14})!} \right)^H \quad (3)$$

此时, 攻击 GSHP 的计算复杂度将达到 10^{3867H} 数量级, 并且随着 H 的增加, N 将呈指数级增长, 从隐写图像中推断秘密信息将更加困难.

3.3 鲁棒性

在传输隐写图像时, 各种干扰可能影响秘密信息的提取精度. 因此, 对于生成式隐写的鲁棒性研究受到广泛关注. 由于人体姿态固有的结构稳定性, GSHP 对各种攻击表现出了优异的鲁棒性. 本节使用不同攻击下的提取精度验证隐写方法的鲁棒性. 假定秘密信息为 $S = [s_1, s_2, \dots, s_{L_S}]$, 提取的秘密信息为 $S^* = [s_1^*, s_2^*, \dots, s_{L_S}^*]$, 其中 L_S 表示信息长度, 则秘密信息的提取精度表示为:

$$EA = \frac{1}{L_S} \sum_{i=1}^{L_S} (1 - s_i \oplus s_i^*) \quad (4)$$

其中, $\oplus$ 表示异或运算.

实验中我们对隐写图像进行了常规图像处理攻击以及几何攻击, 具体的攻击及其参数如表 5 所示, 攻击后的样例图像如图 10 所示.

表 5 图像攻击类型和参数

攻击类型	参数
常规攻击	JPEG压缩
	质量因子: 10%, 30%, 50%, 70%, 90%
	高斯滤波
	窗口大小: 3×3, 5×5, 7×7
	中值滤波
	窗口大小: 3×3, 5×5, 7×7
几何攻击	高斯噪声
	均值: 0; 方差: 0.001, 0.003, 0.005
	椒盐噪声
	密度: 0.001, 0.003, 0.005
	乘性噪声
	均值: 0; 方差: 0.001, 0.003, 0.005
几何攻击	缩放
	比例: 0.25, 0.5, 0.75, 1.5, 2
	旋转
	角度: 10°, 30°, 50°, 70°, 90°
	平移
	距离: [16, 10, 32, 20, 40, 25], [80, 50], [160, 100]
几何攻击	中心裁剪
	比例: 5%, 10%, 15%, 20%
	边缘裁剪
	比例: 5%, 10%, 15%, 20%
几何攻击	随机裁剪
	比例: 5%, 10%, 15%, 20%



图 10 攻击后的样例图像

不同攻击下的提取精度如图 11 所示. 由图 11 的结果可知, 在面对常规攻击时, 提取精度均超过 0.95, 展现出良好的鲁棒性. 当图像遭受旋转攻击时, 从隐写图像中检测到的人体姿态将偏离其原始映射值, 且这种偏差会随旋转角度的增大而进一步扩大. 在人体姿态构造过程中, 若所选角度恰好位于关键点不同状态对应角度范围的阈值附近, 即使微小的角度偏移也可能使该角度越过阈值边界, 导致其映射的秘密信息无法正确提取. 从实验结果可知, 随着旋转角度的增大, 信息提取精度呈现逐渐下降趋势, 但始终高于 0.94. 对于裁剪攻击和平移攻击, 当关键点所在区域被遮挡时, 人体姿态检测算法通常可以根据未遮挡的部分进行预测补全. 然而, 一旦遮挡范围过大 (如表 5 中偏移量为 [160, 100] 像素的平移攻击和比例为 20% 的裁剪攻击), 算法将无法有效预测, 导致关键点缺失, 进而造成该关键点所映射的信息提取失败.

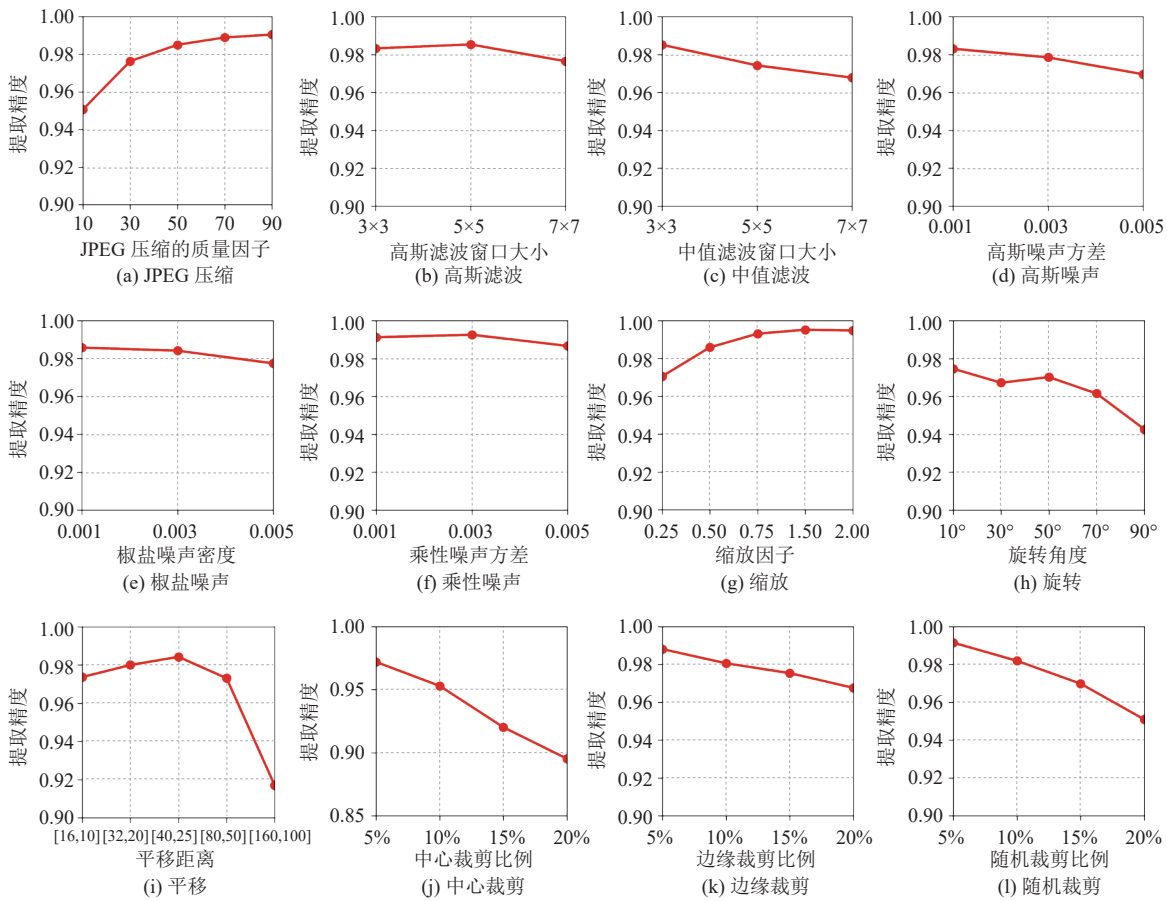


图 11 GSHP 在不同攻击下的提取精度

为量化关键点缺失对信息提取的影响, 我们进一步统计了不同关键点缺失情况下所提取的三进制数的出错位数, 结果如表 6 所示. 从表 6 数据可知, 不同关键点的缺失对信息提取的影响程度不同: 未参与映射的关键点 (如双耳、鼻、颈) 缺失不影响提取; 参与映射的部分关键点 (如双眼、双肘、双腕、双膝、双踝) 缺失仅影响 1、2 位三进制数; 而与姿态校正相关的躯干关键点 (如双肩、双臀) 缺失则会影响全部三进制数的提取. 总体而言, 本方案在面对大多数几何攻击时提取精度高于 0.9, 即使在 20% 中心裁剪攻击下也接近 0.9, 体现了良好的鲁棒性.

表 6 不同关键点缺失对信息提取的影响

缺失关键点	右眼/左眼	右耳/左耳	鼻	颈	右肩/左肩	右肘/左肘	右腕/左腕	右臀/左臀	右膝/左膝	右踝/左踝
三进制数出错位数	1	0	0	0	9	2	1	9	2	1

为了进一步验证提出的 GSHP 的鲁棒性, 我们将其与现有的无载体图像隐写方法在人体姿态图像的结果进行对比. 具体的对比方法包括 Zhang 等人^[48]、Liu 等人^[49]、Meng 等人^[50]和 Tan 等人^[51]的方法. 每种方法采用了他们在 VITON-HD 数据集^[52]、Clothing & Models 数据集^[53]和 Deep Fashion 数据集^[54]结果的均值, 如表 7 所示. 从表 7 数据可知, GSHP 的提取精度在面对椒盐噪声和高斯滤波时略低于 Zhang 等人^[48]的方法和 Liu 等人^[49]的方法, 在面对中值滤波时略低于 Liu 等人^[49]的方法, 但在这些攻击下始终优于 Meng 等人^[50]的方法和 Tan 等人^[51]的方法. 尽管 OpenPose 的人体关键点检测模块对椒盐噪声、高斯滤波与中值滤波的敏感性略高于部分对比方法, 但 GSHP 仍能维持较高的提取精度. 面对上述 3 类攻击之外的其余干扰时, GSHP 的提取精度均全面超越所有对比方法, 尤其在应对几何攻击时, 该方法相较于现有方法具有显著的优势. 这是因为本文所提出的姿态特征映射机制是基于人体各部位间的相对角度构建的, 人体的四肢、面部等关键特征与躯干之间的相对空间关系具有内在的稳定性, 即便图像经历几何变换、内容移位等攻击, 这种相对关系也不会发生本质性改变. 因此, GSHP 在应对各类导致图像内容移位的攻击时, 能够保持优异的提取精度, 展现出更强的鲁棒性.

表 7 与其他隐写方法的提取精度对比

攻击方式	Zhang等人 ^[48]	Liu等人 ^[49]	Meng等人 ^[50]	Tan等人 ^[51]	GSHP
JPEG 50%	0.9813	0.9833	0.9627	0.9700	0.9850
高斯噪声 0.001	0.9240	0.9127	0.8883	0.8906	0.9831
椒盐噪声 0.001	0.9967	0.9960	0.9617	0.9707	0.9857
乘性噪声 0.001	0.9900	0.9660	0.9593	0.9707	0.9913
高斯滤波 3×3	0.9987	0.9993	0.9667	0.9700	0.9833
中值滤波 3×3	0.9840	0.9893	0.9620	0.9607	0.9852
中心裁剪 10%	0.6953	0.4346	0.5833	0.8700	0.9530
中心裁剪 20%	0.5527	0.2927	0.4700	0.7940	0.8952
边缘裁剪 10%	0.1640	0.2260	0.8413	0.9080	0.9806
边缘裁剪 20%	0.1453	0.1753	0.8200	0.8587	0.9676
旋转 10°	0.0000	0.0001	0.7933	0.8400	0.9746
旋转 30°	0.0000	0.0000	0.5953	0.7753	0.9672
平移 [16, 10]	0.4073	0.4833	0.8410	0.8547	0.9737
平移 [32, 20]	0.1300	0.1820	0.8263	0.8313	0.9800
缩放 1.5	0.9853	0.9940	0.9770	0.9533	0.9952

注: 加粗数据表示最优结果

4 总 结

本文提出一种人体姿态引导的生成式隐写 GSHP, 该方法首先将秘密信息编码为人体姿态特征, 再将此特征输入生成模型以生成隐写图像. 在接收端, 可以通过人体姿态检测算法识别隐写图像中的人体姿态, 进而还原出秘密信息. 由于人体姿态具有丰富的表达能力和稳定的结构特性, GSHP 能够在保证隐写图像视觉质量和隐蔽性的同时, 实现对各类攻击的鲁棒性. 理论分析和实验结果表明 GSHP 生成的隐写图像能够有效避免被攻击者检测, 具有较强的安全性, 并且 GSHP 在面对各种攻击时展现出良好的鲁棒性.

References

- [1] Wang RZ, Lin CF, Lin JC. Image hiding by optimal LSB substitution and genetic algorithm. *Pattern Recognition*, 2001, 34(3): 671–683. [doi: [10.1016/S0031-3203\(00\)00015-7](https://doi.org/10.1016/S0031-3203(00)00015-7)]
- [2] Yang CH, Weng CY, Wang SJ, Sun HM. Adaptive data hiding in edge areas of images with spatial LSB domain systems. *IEEE Trans. on Information Forensics and Security*, 2008, 3(3): 488–497. [doi: [10.1109/TIFS.2008.926097](https://doi.org/10.1109/TIFS.2008.926097)]
- [3] Rustad S, Setiadi DRIM, Syukur A, Andono PN. Inverted LSB image steganography using adaptive pattern to improve imperceptibility. *Journal of King Saud University—Computer and Information Sciences*, 2022, 34(6): 3559–3568. [doi: [10.1016/j.jksuci.2020.12.017](https://doi.org/10.1016/j.jksuci.2020.12.017)]

- [4] Zhang H, Hu LT. A data hiding scheme based on multidirectional line encoding and integer wavelet transform. *Signal Processing: Image Communication*, 2019, 78: 331–344. [doi: [10.1016/j.image.2019.07.019](https://doi.org/10.1016/j.image.2019.07.019)]
- [5] Muhuri PK, Ashraf Z, Goel S. A novel image steganographic method based on integer wavelet transformation and particle swarm optimization. *Applied Soft Computing*, 2020, 92: 106257. [doi: [10.1016/j.asoc.2020.106257](https://doi.org/10.1016/j.asoc.2020.106257)]
- [6] Tang WX, Li B, Barni M, Li J, Huang JW. Improving cost learning for JPEG steganography by exploiting JPEG domain knowledge. *IEEE Trans. on Circuits and Systems for Video Technology*, 2022, 32(6): 4081–4095. [doi: [10.1109/TCSVT.2021.3115600](https://doi.org/10.1109/TCSVT.2021.3115600)]
- [7] Xu GS, Wu HZ, Shi YQ. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 2016, 23(5): 708–712. [doi: [10.1109/LSP.2016.2548421](https://doi.org/10.1109/LSP.2016.2548421)]
- [8] Ye J, Ni JQ, Yi Y. Deep learning hierarchical representations for image steganalysis. *IEEE Trans. on Information Forensics and Security*, 2017, 12(11): 2545–2557. [doi: [10.1109/TIFS.2017.2710946](https://doi.org/10.1109/TIFS.2017.2710946)]
- [9] Wei KK, Luo WQ, Liu ML. Spatial color image steganalysis based on central difference convolution and attention. *Ruan Jian Xue Bao/Journal of Software*, 2024, 35(12): 5671–5686 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/7068.htm> [doi: [10.13328/j.cnki.jos.007068](https://doi.org/10.13328/j.cnki.jos.007068)]
- [10] Liu MM, Zhang MQ, Liu J, Gao PX, Zhang YN. Coverless information hiding based on generative adversarial networks. *Journal of Applied Sciences*, 2018, 36(2): 371–382 (in Chinese with English abstract). [doi: [10.3969/j.issn.0255-8297.2018.02.015](https://doi.org/10.3969/j.issn.0255-8297.2018.02.015)]
- [11] Zhang Z, Fu GY, Ni RR, Liu J, Yang XY. A generative method for steganography by cover synthesis with auxiliary semantics. *Tsinghua Science and Technology*, 2020, 25(4): 516–527. [doi: [10.26599/TST.2019.9010027](https://doi.org/10.26599/TST.2019.9010027)]
- [12] Chen XY, Zhang ZT, Qiu AQ, Xia ZH, Xiong NN. Novel coverless steganography method based on image selection and stargan. *IEEE Trans. on Network Science and Engineering*, 2022, 9(1): 219–230. [doi: [10.1109/TNSE.2020.3041529](https://doi.org/10.1109/TNSE.2020.3041529)]
- [13] Zhou ZL, Dong XH, Meng RH, Wang MM, Yan HY, Yu KP, Choo KKR. Generative steganography via auto-generation of semantic object contours. *IEEE Trans. on Information Forensics and Security*, 2023, 18: 2751–2765. [doi: [10.1109/TIFS.2023.3268843](https://doi.org/10.1109/TIFS.2023.3268843)]
- [14] Sun YQ, Liu JY, Zhang R. A robust generative image steganography method based on guidance features in image synthesis. In: *Proc. of the 2023 IEEE Int'l Conf. on Multimedia and Expo. Brisbane: IEEE*, 2023. 55–60. [doi: [10.1109/ICME55011.2023.00018](https://doi.org/10.1109/ICME55011.2023.00018)]
- [15] Rehman HA, Bajwa UI, Raza RH, Alfarhood S, Safran M, Zhang F. Leveraging coverless image steganography to hide secret information by generating anime characters using GAN. *Expert Systems with Applications*, 2024, 248: 123420. [doi: [10.1016/j.eswa.2024.123420](https://doi.org/10.1016/j.eswa.2024.123420)]
- [16] Zhu YM, Chen F, He HJ, Chen HY. Orthogonal GAN information hiding model based on secret information driven. *Journal of Applied Sciences*, 2019, 37(5): 721–732 (in Chinese with English abstract). [doi: [10.3969/j.issn.0255-8297.2019.05.013](https://doi.org/10.3969/j.issn.0255-8297.2019.05.013)]
- [17] Zhou ZL, Su YC, Li J, Yu KP, Wu QMJ, Fu ZJ, Shi YQ. Secret-to-image reversible transformation for generative steganography. *IEEE Trans. on Dependable and Secure Computing*, 2023, 20(5): 4118–4134. [doi: [10.1109/TDSC.2022.3217661](https://doi.org/10.1109/TDSC.2022.3217661)]
- [18] Yang ZJ, Zeng K, Chen KJ, Fang H, Zhang WM, Yu NH. Gaussian shading: Provable performance-lossless image watermarking for diffusion models. In: *Proc. of the 2024 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE*, 2024. 12162–12171. [doi: [10.1109/CVPR52733.2024.01156](https://doi.org/10.1109/CVPR52733.2024.01156)]
- [19] Yang ZJ, Chen KJ, Zeng K, Zhang WM, Yu NH. Provably secure robust image steganography. *IEEE Trans. on Multimedia*, 2024, 26: 5040–5053. [doi: [10.1109/TMM.2023.3330098](https://doi.org/10.1109/TMM.2023.3330098)]
- [20] Hu XX, Li S, Ying QC, Peng WL, Zhang XP, Qian ZX. Establishing robust generative image steganography via popular stable diffusion. *IEEE Trans. on Information Forensics and Security*, 2024, 19: 8094–8108. [doi: [10.1109/TIFS.2024.3444311](https://doi.org/10.1109/TIFS.2024.3444311)]
- [21] Zhang QH, Huang FJ. Provably secure generative steganography based on adjustable orthogonal mapping. *IEEE Trans. on Dependable and Secure Computing*, 2026, 23(1): 34–48. [doi: [10.1109/TDSC.2025.3604222](https://doi.org/10.1109/TDSC.2025.3604222)]
- [22] Zhou Q, Wei P, Qian ZX, Zhang XP, Li S, Qin C. Conditional flow-based generative steganography. *IEEE Trans. on Dependable and Secure Computing*, 2025, 22(5): 5632–5647. [doi: [10.1109/TDSC.2025.3570468](https://doi.org/10.1109/TDSC.2025.3570468)]
- [23] Yang K, Chen KJ, Zhang WM, Yu NH. Provably secure generative steganography based on autoregressive model. In: *Proc. of the 17th Int'l Workshop on Digital Forensics and Watermarking. Jeju Island: Springer*, 2018. 55–68. [doi: [10.1007/978-3-030-11389-6_5](https://doi.org/10.1007/978-3-030-11389-6_5)]
- [24] Ding JY, Chen KJ, Wang YF, Zhao N, Zhang WM, Yu NH. Discop: Provably secure steganography in practice based on “distribution copies”. In: *Proc. of the 2023 IEEE Symp. on Security and Privacy. San Francisco: IEEE*, 2023. 2238–2255. [doi: [10.1109/SP46215.2023.10179287](https://doi.org/10.1109/SP46215.2023.10179287)]
- [25] Zhang Z, Fu GY, Liu J, Fu WY. Generative information hiding method based on adversarial networks. In: *Proc. of the 8th Int'l Conf. on Computer Engineering and Networks. Shanghai: Springer*, 2020. 261–270. [doi: [10.1007/978-3-030-14680-1_29](https://doi.org/10.1007/978-3-030-14680-1_29)]
- [26] Xue R, Wang YN. Message drives image: A coverless image steganography framework using multi-domain image translation. In: *Proc. of the 2021 Int'l Joint Conf. on Neural Networks. Shenzhen: IEEE*, 2021. 1–9. [doi: [10.1109/IJCNN52387.2021.9534043](https://doi.org/10.1109/IJCNN52387.2021.9534043)]
- [27] Liu XY, Ma ZP, Ma JX, Zhang J, Schaefer G, Fang H. Image disentanglement autoencoder for steganography without embedding. In:

- Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 2293–2302. [doi: [10.1109/CVPR52688.2022.00234](https://doi.org/10.1109/CVPR52688.2022.00234)]
- [28] Ma LQ, Jia X, Sun QR, Schiele B, Tuytelaars T, van Gool L. Pose guided person image generation. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 405–415.
 - [29] Siarohin A, Sangineto E, Lathuilière S, Sebe N. Deformable GANs for pose-based human image generation. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3408–3416. [doi: [10.1109/CVPR.2018.00359](https://doi.org/10.1109/CVPR.2018.00359)]
 - [30] Esser P, Sutter E. A variational U-Net for conditional appearance and shape generation. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8857–8866. [doi: [10.1109/CVPR.2018.00923](https://doi.org/10.1109/CVPR.2018.00923)]
 - [31] Zhu Z, Huang TT, Shi BG, Yu M, Wang BF, Bai X. Progressive pose attention transfer for person image generation. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2342–2351. [doi: [10.1109/CVPR.2019.00245](https://doi.org/10.1109/CVPR.2019.00245)]
 - [32] Men Y, Mao YM, Jiang YN, Ma WY, Lian ZH. Controllable person image synthesis with attribute-decomposed GAN. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5083–5092. [doi: [10.1109/CVPR42600.2020.00513](https://doi.org/10.1109/CVPR42600.2020.00513)]
 - [33] Zhang L, Rao AY, Agrawala M. Adding conditional control to text-to-image diffusion models. In: Proc. of the 2023 IEEE/CVF Int'l Conf. on Computer Vision. Paris: IEEE, 2023. 3813–3824. [doi: [10.1109/ICCV51070.2023.00355](https://doi.org/10.1109/ICCV51070.2023.00355)]
 - [34] Luvizon DC, Picard D, Tabia H. 2D/3D pose estimation and action recognition using multitask deep learning. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5137–5146. [doi: [10.1109/CVPR.2018.00539](https://doi.org/10.1109/CVPR.2018.00539)]
 - [35] Li K, Wang SJ, Zhang X, Xu YF, Xu WJ, Tu ZW. Pose recognition with cascade Transformers. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 1944–1953. [doi: [10.1109/CVPR46437.2021.00198](https://doi.org/10.1109/CVPR46437.2021.00198)]
 - [36] Wei SE, Ramakrishna V, Kanade T, Sheikh Y. Convolutional pose machines. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 4724–4732. [doi: [10.1109/CVPR.2016.511](https://doi.org/10.1109/CVPR.2016.511)]
 - [37] Li YJ, Yang S, Liu PD, Zhang SK, Wang YX, Wang ZC, Yang WK, Xia ST. SimCC: A simple coordinate classification perspective for human pose estimation. In: Proc. of the 17th European Conf. on Computer Vision. Tel Aviv: Springer, 2022. 89–106. [doi: [10.1007/978-3-031-20068-7_6](https://doi.org/10.1007/978-3-031-20068-7_6)]
 - [38] Moon G, Chang JY, Lee KM. PoseFix: Model-agnostic general human pose refinement network. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7765–7773. [doi: [10.1109/CVPR.2019.00796](https://doi.org/10.1109/CVPR.2019.00796)]
 - [39] Huang JJ, Zhu Z, Guo F, Huang G. The devil is in the details: Delving into unbiased data processing for human pose estimation. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5699–5708. [doi: [10.1109/CVPR42600.2020.00574](https://doi.org/10.1109/CVPR42600.2020.00574)]
 - [40] Cheng BW, Xiao B, Wang JD, Shi HH, Huang TS, Zhang L. HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5385–5394. [doi: [10.1109/CVPR42600.2020.00543](https://doi.org/10.1109/CVPR42600.2020.00543)]
 - [41] Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172–186. [doi: [10.1109/TPAMI.2019.2929257](https://doi.org/10.1109/TPAMI.2019.2929257)]
 - [42] Epinikion. epiCRealism. 2023. <https://civitai.com/models/25694/epicrealism>
 - [43] Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer. IEEE Signal Processing Letters, 2013, 20(3): 209–212. [doi: [10.1109/LSP.2012.2227726](https://doi.org/10.1109/LSP.2012.2227726)]
 - [44] Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. IEEE Trans. on Image Processing, 2012, 21(12): 4695–4708. [doi: [10.1109/TIP.2012.2214050](https://doi.org/10.1109/TIP.2012.2214050)]
 - [45] Venkatanath N, Praneeth D, Bh MC, Channappayya SS, Medasani SS. Blind image quality evaluation using perception based features. In: Proc. of the 21st National Conf. on Communications. Mumbai: IEEE, 2015. 1–6. [doi: [10.1109/NCC.2015.7084843](https://doi.org/10.1109/NCC.2015.7084843)]
 - [46] Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images. IEEE Trans. on Information Forensics and Security, 2019, 14(5): 1181–1193. [doi: [10.1109/TIFS.2018.2871749](https://doi.org/10.1109/TIFS.2018.2871749)]
 - [47] You WK, Zhang H, Zhao XF. A siamese CNN for image steganalysis. IEEE Trans. on Information Forensics and Security, 2021, 16: 291–306. [doi: [10.1109/TIFS.2020.3013204](https://doi.org/10.1109/TIFS.2020.3013204)]
 - [48] Zhang X, Peng F, Long M. Robust coverless image steganography based on DCT and LDA topic classification. IEEE Trans. on Multimedia, 2018, 20(12): 3223–3238. [doi: [10.1109/TMM.2018.2838334](https://doi.org/10.1109/TMM.2018.2838334)]
 - [49] Liu Q, Xiang XY, Qin JH, Tan Y, Tan JS, Luo YJ. Coverless steganography based on image retrieval of DenseNet features and DWT

- sequence mapping. Knowledge-Based Systems, 2020, 192: 105375. [doi: [10.1016/j.knsys.2019.105375](https://doi.org/10.1016/j.knsys.2019.105375)]
- [50] Meng LJ, Jiang XH, Zhang ZZ, Li ZH, Sun TF. A robust coverless image steganography based on an end-to-end hash generation model. IEEE Trans. on Circuits and Systems for Video Technology, 2023, 33(7): 3542–3558. [doi: [10.1109/TCSVT.2022.3232790](https://doi.org/10.1109/TCSVT.2022.3232790)]
- [51] Tan Y, Xiang XY, Qin JH, Tan Y. Robust coverless image steganography based on human pose estimation. Knowledge-based Systems, 2024, 296: 111873. [doi: [10.1016/j.knsys.2024.111873](https://doi.org/10.1016/j.knsys.2024.111873)]
- [52] Choi S, Park S, Lee M, Choo J. VITON-HD: High-resolution virtual try-on via misalignment-aware normalization. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 14126–14135. [doi: [10.1109/CVPR46437.2021.01391](https://doi.org/10.1109/CVPR46437.2021.01391)]
- [53] Dominic Monn. Clothing & Models. 2018. <https://www.kaggle.com/dqmonn/zalando-store-crawl>
- [54] Liu ZW, Luo P, Qiu S, Wang XG, Tang XO. DeepFashion: Powering robust clothes recognition and retrieval with rich annotations. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1096–1104. [doi: [10.1109/CVPR.2016.124](https://doi.org/10.1109/CVPR.2016.124)]

附中文参考文献

- [9] 魏康康, 骆伟祺, 刘明林. 基于中心差分卷积和注意力的空域彩色图像隐写分析. 软件学报, 2024, 35(12): 5671–5686. <http://www.jos.org.cn/1000-9825/7068.htm> [doi: [10.13328/j.cnki.jos.007068](https://doi.org/10.13328/j.cnki.jos.007068)]
- [10] 刘明明, 张敏情, 刘佳, 高培贤, 张英男. 基于生成对抗网络的无载体信息隐藏. 应用科学学报, 2018, 36(2): 371–382. [doi: [10.3969/j.issn.0255-8297.2018.02.015](https://doi.org/10.3969/j.issn.0255-8297.2018.02.015)]
- [16] 朱翌明, 陈帆, 和红杰, 陈鸿佑. 基于秘密信息驱动的正交 GAN 信息隐藏模型. 应用科学学报, 2019, 37(5): 721–732. [doi: [10.3969/j.issn.0255-8297.2019.05.013](https://doi.org/10.3969/j.issn.0255-8297.2019.05.013)]

作者简介

张清华, 博士生, 主要研究领域为信息隐藏.

黄方军, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为 AI 安全, 信息隐藏, 数字取证.

马宾, 博士, 教授, 博士生导师, 主要研究领域为信息隐藏, 多媒体安全, 图像处理.