

一种用于监视和侦察中面部属性操作和重建的深度学习方法

Anees Nashath Shaik
Department of Networks
and Digital Media
Kingston University
London, United Kingdom
A.Shaik@kingston.ac.uk

Barbara Villarini
School of Computer Science
and Engineering
University of Westminster
London, United Kingdom
b.villarini@westminster.ac.uk

Vasileios Argyriou
Department of Networks
and Digital Media
Kingston University
London, United Kingdom
vasileios.argyriou@kingston.ac.uk

Abstract—监控系统在安全和侦察中发挥关键作用，但其性能常因图像和视频质量较差而受到影响，导致面部识别的准确性降低。此外，现有的基于 AI 的面部分析模型存在与肤色变化和部分遮挡面部相关的偏差，使其在多样化的真实场景中有效性下降。这些挑战是数据局限性和不平衡的结果，现有的训练数据集缺乏足够的多样性，导致面部识别性能不公正和不可靠。为了解决这些问题，我们提出了一种数据驱动的平台，通过生成合成训练数据来增强监控能力，以补偿数据集偏差。我们的方法利用深度学习的面部属性操控和重构技术，通过自动编码器和生成对抗网络（GANs）创建多样化且高质量的面部数据集。此外，我们的系统整合了图像增强模块，提高了监控录像中低分辨率或被遮挡面部的清晰度。我们使用 CelebA 数据集评估我们的方法，证明所提平台增强了训练数据的多样性和模型的公平性。这项工作有助于减少 AI 面部分析中的偏差，并在具有挑战性的环境中提高监控精度，从而推动更加公平和可靠的安全应用。

I. 介绍

面部分析是现代监控和侦察系统的基本组成部分，能够实现生物特征识别 [1], [2]、安全监控和法医学调查。然而，由于图像和视频质量较低，这些系统在实际场景中往往难以有效运行，图像和视频质量的下降可能是由于光线条件差、运动模糊、压缩伪影以及遮挡 [3] 所导致的。此外，基于人工智能的面部识别模型在不同肤色和部分遮挡的面部上表现出性能差异，限制了它们在多样化环境中的可靠性 [4], [5]。这些挑战主要源于数据限制和偏见，因为许多现有的数据集缺乏各种人口群体和面部条件的充分代表性。因此，在这类数据集上训练的监控模型可能无法有效推广，导致不公平和不准确的识别。为克服这些限制，我们提出一个数据驱动的平台，旨在生成多样化和高质量的训练数据，以补偿偏见和数据集不平衡。我们的方法利用基于深度学习的面部属性操控和重构，使用自动编码器和生成对抗网络（GAN）以合成肤色、遮挡及其他面部特征的现实变化。具体而言，所提出的系统包括三个关键组件：(1) 一个使用 Wasserstein GAN 与梯度惩罚 (WGAN-GP) 来生成多样肤色同时保留身份的肤色修改模型，(2) 一个基于编码-解码 GAN 架构的眼镜去除模型，以提高被遮挡面部的识别准确性，(3) 一个基于 GAN 的图像增强模块，结合风格迁移技术来优化退化的监控片段。我们使用 CelebA [6] 评估了我们的方法，证明所提出的平台可以提高训练数据的多样性和模型的公平性，从而显著改善识别性能，尤其是在代表性不足的人群和具有挑战性的面部条件。像 [7] 和 [8] 这样的研究突出了数据集多样性对模型公平性的影响，进一步支持了我们的方法。通过解决数据偏差和图像质量问题，我们的平台为开发更加公平和有效的 AI 驱动的监控系统做出了贡献。本文的其余部分结构

如下：第二部分回顾了面部识别中的偏差以及基于深度学习的图像增强的相关工作。第三部分详细介绍了我们的方法，包括数据集增强、GAN 架构以及图像增强技术。第四部分介绍了实验结果和性能分析。最后，第五部分讨论了主要发现和未来研究方向，强调了我们的方法对提高 AI 驱动监控系统的公平性和有效性的影响。

II. 相关工作

深度学习在面部分析中的进步已经大大改进了监控和生物识别系统。然而，现有文献强调了与数据偏差、低质量图像和遮挡处理相关的关键限制。本节回顾了关于面部图像增强的深度学习技术、AI 驱动识别中公平性以及数据增强策略的先前的研究。

[width=0.5]00_pipeline.pdf

Fig. 1. 整体提出的流程由多个阶段组成

对面部识别模型中的偏见已经进行了广泛研究，研究表明，由于数据集的不平衡，模型通常在某些人口统计群体上表现更好 [9]。诸如 [10] 的研究表明，许多非人口统计特征也会强烈影响识别性能，例如饰品、发型和颜色、脸型或面部异常，这强调了多样化训练数据以减少面部特征分类差异的必要性。我们的方法基于这些努力，通过开发一个数据生成管道来扩大数据集多样性，并在不同肤色和遮挡场景下提高识别准确性。

低质量的监控视频对人脸识别构成了重大挑战。传统方法，包括超分辨率技术 [11] 和去噪算法 [12]，虽然显示出一定的改进，但在不同条件下的泛化能力上仍然存在困难。基于 GAN 的增强模型的最新进展，如超分辨率 GAN (SRGAN) [13]，在精细化面部细节并保留关键特征方面表现出了更优异的性能。此外，AnimeGANv3 [14] 已成功应用于风格迁移和增强，提高了面部图像的清晰度和定义，特别是在传统增强技术失效的情况下。

A. 基于深度学习的人脸属性操作

生成模型，特别是自动编码器和 GAN，已被广泛应用于面部属性操作。比如，[15] 利用 CycleGAN 和 Pix2Pix 架构进行图像到图像翻译，实现精确修改，同时保持身份一致性。此外，使用 StyleGAN [16], [17], [18] 的面部属性编辑方法提供了对特征如肤色和遮挡处理的细粒度控制。近期研究还展示了 AnimeGANv3 在生成真实变换方面的有效性，这可以应用于面部属性的修改和增强。我们提出的系统通过将属性操作与自适应数据增强框架集成，扩展了这些方法，确保 AI 驱动的监控系统的稳健性和公平性。

III. 方法论

提出的方法旨在通过实现一个结构化的数据增强和操作流程来解决偏差和低质量面部数据集的局限性。该方法包括多个阶段，如图 II 所示。本节概述了每个阶段中使用的技术细节和深度学习技术。第一步涉及分析数据集中的偏差，例如缺失的肤色和被遮挡的面部特征（如眼镜、口罩）。这是通过特征提取模型实现的，利用预训练的 ResNet 来识别代表性不足的属性。提取的特征用于为数据增强和生成的下一个阶段提供信息。

[width=0.4]skin_tone_model.pdf

Fig. 2. 皮肤模型结构

[width=0.3]ERGAN.pdf

Fig. 3. ERGAN 架构具有编码器-解码器、跳跃连接、注意力模块和身份损失

第二步旨在生成合成数据。该过程分为两个关键组成部分，一个是皮肤模型，另一个是眼镜去除模型。

1) 皮肤模型架构：GAN 在图像到图像的转换方面表现优异，但常常面临诸如消失梯度和模式崩溃的不稳定性问题。为了应对这些挑战，带梯度惩罚的 Wasserstein GAN (WGAN-GP) [19] 通过梯度惩罚强制执行 Lipschitz 约束，从而提高训练的稳定性。本研究应用 WGAN-GP 来修改人脸图像中的肤色，同时保留面部身份。所提出的模型由生成器 G 和评论家 D （标准 GAN 中的鉴别器）构成，通过对抗训练。给定一个输入人脸图像 x 和随机噪声 z ，生成器产生具有改变肤色的修改版本 $G(x, z)$ ：

$$x' = G(x, z), \quad (1)$$

，其中 x' 保留了 x 的结构特征，同时修改了皮肤色素。评论者 D 通过估计其分布之间的 Wasserstein 距离来区分真实图像 x 和生成图像 x' 。与标准 GAN 中使用的传统 Jensen-Shannon 散度不同，WGAN-GP 最小化了真实 (P_r) 和生成 (P_g) 分布之间的 Wasserstein-1 距离 $W(P_r, P_g)$ ：

$$W(P_r, P_g) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{x' \sim P_g}[D(x')], \quad (2)$$

其中 f 是一个由判别器 D 近似的 1-Lipschitz 函数。为了强制满足 Lipschitz 约束，应用了梯度惩罚：

$$\mathcal{L}_{GP} = \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \quad (3)$$

其中 $\hat{x}$ 是实际图像和生成图像之间的插值样本：

$$\hat{x} = \epsilon x + (1 - \epsilon)x', \quad \epsilon \sim \mathcal{U}(0, 1). \quad (4)$$

用于训练评价器的总损失函数为：

$$\mathcal{L}_D = \mathbb{E}_{x' \sim P_g}[D(x')] - \mathbb{E}_{x \sim P_r}[D(x)] + \mathcal{L}_{GP}. \quad (5)$$

生成器被训练以最大化评价者的评分：

$$\mathcal{L}_G = -\mathbb{E}_{x' \sim P_g}[D(x')]. \quad (6)$$

该模型通过以下步骤迭代训练：

- 1) 对真实图像 $x \sim P_r$ 进行采样并生成伪造图像 $x' = G(x, z)$ 。
- 2) 计算评论者损失 $\mathcal{L}_D$ 并更新 D 。
- 3) 计算梯度惩罚 $\mathcal{L}_{GP}$ 并施加 Lipschitz 约束。
- 4) 通过最大化 $D(x')$ 来更新生成器。

我们使用 Adam 优化器，学习率为 $\alpha_D = 1 \times 10^{-4}$ 和 $\alpha_G = 1 \times 10^{-4}$ ，批量大小为 64。WGAN-GP 模型的详细表示在图 III 中显示。

2) 去眼镜模型：我们开发了一个眼镜移除生成对抗网络 (ERGAN)，这是一种深度学习模型，旨在去除面部图像中的眼镜，同时保留身份和面部结构。它采用编码器-解码器架构，并具有跳跃连接和注意力模块，确保生成高质量、保留身份的信息的图像。ERGAN 通过一系列卷积层 (Conv + ReLU) 处理带眼镜的输入面部图像，以提取潜在特征表示。然后通过反卷积层 (Deconv + ReLU) 将这种瓶颈表示传递，以重建无眼镜版本的人脸。该模型结合了跳跃连接，有助于保留细粒度的面部细节，以及一个专注于眼睛区域的注意力模块，以确保精确去除眼镜。此外，还有一个面部识别损失函数，以维持输入和输出图像之间的身份一致性。生成图像的真实性进一步通过使用 PatchGAN 鉴别器得到改进，该鉴别器通过应用对抗学习来优化重建人脸的质量。总体架构的表示在图 3 中提供。

[width=0.4]AnimeGANv3.pdf

Fig. 4. AnimeGANv3 架构概述，展示了生成器的结构、双输出尾部以及鉴别器反馈。

为提升监控视频的感知质量，AnimeGANv3 [14] 被纳入作为一种基于风格的增强框架。与传统的逐像素修复技术不同，AnimeGANv3 使用生成对抗网络 (GANs) 来优化面部结构、边缘和轮廓，同时减少噪声和伪影。整体架构的表示如图 4 所示。由于 AnimeGANv3 主要用于艺术化转换，因而应用领域自适应技术以使其与真实世界的监控视频相匹配。增强过程从帧提取和缩放到 256×256 像素开始，以满足兼容性。使用边缘平滑和超像素分割等预处理技术来减少噪声并在将帧输入生成网络之前保留面部结构。双尾生成器网络作为 AnimeGANv3 的主干。支撑尾生成粗粒度的风格化输出，平滑纹理并减少不必要的细节。主尾通过去除伪影并增强边缘清晰度来优化输出，确保跨帧的身份一致性。模型的一个关键组成部分是线性自适应去归一化 (LADE)，这取代了传统的归一化技术。与实例归一化不同，LADE 基于全局统计动态调整，保留基本纹理同时防止过度风格化。该架构还包括两个判别器以优化增强过程。第一个判别器 D1 评估灰度纹理，确保边缘结构和纹理保持清晰。第二个判别器 D2 评估整体帧的清晰度，防止模糊和关键细节的丢失。

A. 高质量平衡数据集

最后阶段涉及将增强和合成的图像与原始数据集整合，以创建一个平衡良好的数据集。这个高质量的数据集随后用于面部识别、属性分类和监控应用等下游任务。

TABLE I. 不同类别的 PSNR 及其标准偏差

Metric Value	Skin	Eyeglasses	Ehnanced
PSNR	29.830	29.958	29.986
PSNR std	1.771	0.964	0.752

TABLE II. 不同类别的 SSIM 及其标准差

Metric Value	Skin	Eyeglasses	Ehnanced
SSIM	0.934	0.385	0.786
SSIM std	0.100	0.195	0.034

B. 数据集和预处理

模型在大型 CelebFaces Attributes (CelebA) 数据集上进行了训练, 该数据集是一个大规模的集合, 包含超过 20 万张带有 40 种面部属性标注的名人图像 [6]。CelebA 提供了多种面部变化, 使其非常适合学习眼镜等关键特征。数据集的广泛标注使得模型能够有效地区分有无眼镜的图像, 从而促进了有针对性的特征提取。数据集被分为 70 % 用于训练, 10 % 用于评估, 20 % 用于测试。为了确保训练样本的一致性, 所有图像都被调整为 128×128 像素, 保持统一的输入尺寸。像素值被归一化到 [-1,1] 范围, 以实现稳定的训练动态。数据增强技术, 包括水平翻转和随机旋转, 提高了模型的稳健性和对面部朝向和表情变化的泛化能力。

IV. 结果和分析

皮肤模型、眼镜去除模型以及用于数据增强和风格迁移的模型通过两个关键指标对其生成的合成图像进行了评估: 峰值信噪比 (PSNR) 和结构相似性指数 (SSIM)。这些指标被选择用于评估模型重建高保真度图像的能力。

- PSNR: 衡量重建图像相对于原始图像的质量, 量化失真水平。数值越高表明重建保真度越好。

- SSIM: 评估两幅图像之间的相似性, 通常用于图像处理和计算机视觉中以评估感知图像质量。它通过比较亮度、对比度和结构信息来确定重建图像与原始图像的匹配程度。SSIM 的取值范围从 -1 到 1, 其中 1 表示完全相似, 小于 0 表示结构信息完全不同。

这些模型的表现使用 CelebA 数据集进行评估。数值结果展示在表格 I 和 II 中。由肤色修改模型生成的图像达到 29.830 的 PSNR, 而去除眼镜模型和图像增强模块生成的图像则分别获得了 29.958 和 29.986 的 PSNR。这些结果表明在不同变换中高水平的细节保留和最小的重建失真。同样, SSIM 分数显示肤色模型保持了很强的结构一致性, SSIM 达到 0.934, 而去除眼镜模型由于需要大量修改输入图像, 导致 SSIM 较低, 仅为 0.385。图像增强模块获得了 0.786 的 SSIM, 强调了其在增强面部图像和保留关键特征方面的能力。定性分析则突出了模型超越量化指标的优势。关键的面部特征如眼睛、鼻子和嘴巴得到了良好的保留, 没有出现失真。模型实现了平滑的肤色过渡和自然的肤色调整, 维持了面部的自然外观。非肤色元素如眼睛、头发及面部轮廓也得到了保留, 没有出现瑕疵。这些结果确认了我们基于深度学习的方法能够增强数据集的多样性, 提高图像的清晰度并减少面部识别模型中的偏差, 证实其在真实世界监视和侦察挑战中的鲁棒性。

V. 结论

本研究引入了一个基于深度学习的框架, 通过面部属性操控、图像增强和数据集扩充来提高以人为驱动的监控

系统的公平性和有效性。我们的方法有效地解决了数据集偏差、遮挡和低质量图像的问题, 从而构建了更鲁棒且无偏见的面部识别模型。此项工作的一个关键贡献是生成合成训练数据以平衡人口代表性, 从而减轻了基于 AI 的人脸识别中的偏差。图像增强模块的集成通过改善退化的面部图像, 进一步提高了识别准确性。定性分析证实面部特征得到了良好的保留, 确保结构完整性和自然外观。未来的工作包括扩展面部属性操作, 以整合表情和衰老效果, 增强模型的泛化能力。此外, 整合领域适应技术和实时处理能力可以改善其在各种监控环境中的适用性。总之, 该工作在减少偏差和改善图像质量上推进了以人为驱动的监控, 有助于构建更公平、更有效的面部识别系统, 用于安全和法医应用。

REFERENCES

- [1] A. R. Khadka, M. M. Oghaz, W. Matta, M. Cosentino, P. Remagnino, and V. Argyriou, "Learning how to analyse crowd behaviour using synthetic data," ser. CASA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 11–14.
- [2] V. Argyriou, M. Petrou, and S. Barsky, "Photometric stereo with an arbitrary number of illuminants," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 887–900, 2010.
- [3] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Comput. Surv.*, vol. 54, no. 10s, Sep. 2022. [Online]. Available: <https://doi.org/10.1145/3507901>
- [4] K. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer, "Issues related to face recognition accuracy varying based on race and skin tone," *IEEE Transactions on Technology and Society*, vol. 1, no. 1, pp. 8–20, 2020.
- [5] G. M. Rajegowda, Y. Spyridis, B. Villarini, and V. Argyriou, "An ai-assisted skincare routine recommendation system in xr," in *AI Technologies and Virtual Reality*, K. Nakamatsu, S. Patnaik, and R. Kountchev, Eds. Singapore: Springer Nature Singapore, 2024, pp. 381–395.
- [6] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.
- [7] T. Xu, J. White, S. Kalkan, and H. Gunes, "Investigating bias and fairness in facial expression recognition," in *Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 506–523.
- [8] H. F. Menezes, A. S. Ferreira, E. T. Pereira, and H. M. Gomes, "Bias and fairness in face detection," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2021, pp. 247–254.
- [9] V. Cherepanova, S. Reich, S. Dooley, H. Souri, J. Dickerson, M. Goldblum, and T. Goldstein, "A deep dive into dataset imbalance and bias in face identification," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023, pp. 229–247.
- [10] P. Terhörst, J. N. Kolf, M. Huber, F. Kirchbuchner, N. Damer, A. M. Moreno, J. Fierrez, and A. Kuijper, "A comprehensive study on face recognition biases beyond demographics," *IEEE Transactions on Technology and Society*, vol. 3, no. 1, pp. 16–30, 2021.
- [11] M. Farooq, M. N. Dailey, A. Mahmood, J. Moonrinta, and M. Ekpanyapong, "Human face super-resolution on poor quality surveillance video footage," *Neural Computing and Applications*, vol. 33, pp. 13 505–13 523, 2021.
- [12] S. R. Bhagwat and L. Ragha, "Empirical analysis of denoising algorithms for cctv face images," in *International Conference on Artificial Intelligence on Textile and Apparel*. Springer, 2023, pp. 255–266.
- [13] C. Song, Z. He, Y. Yu, and Z. Zhang, "Low resolution face recognition system based on esrgan," in *2021 3rd International*

Conference on Applied Machine Learning (ICAML). IEEE, 2021, pp. 76–79.

- [14] G. LIU, X. CHEN, and Z. GAO, “A novel double-tail generative adversarial network for fast photo animation,” *IEICE Transactions on Information and Systems*, vol. E107.D, no. 1, pp. 72–82, 2024.
- [15] R. K. Senapati, R. Satvika, A. Anmandla, G. Ashesh Reddy, and C. Anil Kumar, “Image-to-image translation using pix2pix gan and cycle gan,” in *International Conference on Data Intelligence and Cognitive Informatics*. Springer, 2023, pp. 573–586.
- [16] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.
- [17] S. Bounareli, C. Tzelepis, V. Argyriou, I. Patras, and G. Tzimiropoulos, “Stylemask: Disentangling the style space of stylegan2 for neural face reenactment,” in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 2023, pp. 1–8.
- [18] S. Bounareli, V. Argyriou, and G. Tzimiropoulos, “Finding directions in gan’s latent space for neural face reenactment,” *British Machine Vision Conference (BMVC)*, 2022.
- [19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” *Advances in neural information processing systems*, vol. 30, 2017.