https://doi.org/10.2352/EI.2026.38.4.MWSF-312

Retinex-guided Relighting and Latent-Space Refinement for Realistic Diffusion-based Face Swapping

Thu Hien Le*[†] *Christophe Charrier*[†] *Emmanuel Giguet*[†] *Maxime Bérubé*[‡] *;
[†] ***Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, F-14000 Caen, France***
[‡] ***Université du Québec à Trois-Rivières, GRSF; Trois-Rivières (Québec) G8Z 4M3, Canada***

Abstract

Face swapping, or deepfake generation, remains a challenging task that requires balancing identity preservation, attribute consistency, and photorealistic realism. We propose a novel training-free, three-stage face swapping framework that improves realism by explicitly aligning illumination and skin appearance prior to diffusion-based synthesis. Our approach refines photometric consistency and skin tone while preserving facial structure and integrates seamlessly with an off-the-shelf diffusion face swapping model. Experiments on the CelebAMask-HQ dataset demonstrate significant improvements in both visual realism and attribute preservation, achieving an FID score of 7.16 compared to the baseline. The proposed method provides an efficient and robust solution for realistic face swapping under varying illumination and appearance conditions without additional model training.

Introduction

Face swapping [1, 2, 3, 4, 5, 6] has emerged as one of the most compelling recognized applications across various domains, from entertainment and social media, to data augmentation, virtual reality, and augmented reality. The goal of face swapping is to generate a synthesized facial image that faithfully contains the unique identity of a source image, while keeping other attributes such as expression, head pose, background, and lighting conditions of a target image. Achieving both high visual fidelity and robust attribute preservation remains a challenging problem, especially under unconstrained real-world conditions.

Early face swapping approaches have primarily relied on generative adversarial networks, which have demonstrated promising results but often suffer from training instability and limited generalization [7, 3, 8]. More recently, diffusion-based models have emerged as a powerful alternative, offering improved stability and superior synthesis quality [5, 9, 6]. These methods have significantly advanced the state-of-the-art in identity preservation and visual realism.

Despite these advances, diffusion-based face swapping systems still face important limitations. In particular, perceptual artifacts such as inconsistent lighting, mismatched skin tones, and imperfect blending with the target environment remain common. Moreover, improving attribute preservation typically requires redesigning conditioning mechanisms and retraining large diffusion models, which is generally computationally expensive and impractical.

In this paper, we propose a novel zero-training face swapping framework that addresses these limitations through a modular pre-processing strategy. Our approach decomposes the face swapping task into three complementary stages, explicitly aligning lighting and skin tone characteristics prior to diffusion-based synthesis. By augmenting an off-the-shelf diffusion face swapper with dedicated lighting transfer and skin refinement modules, our method significantly improves realism and attribute consistency without any model retraining.

The structure of this paper is organized as follows. We first review related work on face swapping and diffusion-based image synthesis. We then describe our proposed three-stage pipeline in the Methodology section. Experimental results are presented and compared with state-of-the-art methods, followed by a conclusion.

Related works

Face Swapping Methods

Despite the clear objective, face swapping remains a challenging problem due to the difficulty of desentangling identity representations from other facial attributes without degrading photorealistic visual quality in real-world scenarios. First approaches relied on traditional computer graphics and reenactment techniques, such as Face2Face [1], before the advent of deep models.

Learning-based face swapping methods have made significant advancements with the development of generative adversarial networks (GANs). Representative GAN-based approaches such as FaceShifter [2], SimSwap [3], and FSGAN [7] explicitly disentangle identity and attribute representations to generate realistic swaps. Despite their success, GAN-based methods often suffer from training instability, limited generalization, and artifacts caused by imperfect disentanglement.

Diffusion-based models have recently emerged as a powerful alternative [5, 6, 9]. Owing to their stable training process and strong generative capacity, diffusion models achieve superior visual quality and identity preservation. However, they may still produce perceptual artifacts under unconstrained illumination conditions, and generally require expensive retraining when modifying conditioning mechanisms.

These limitations motivate complementary strategies that enhance photometric consistency and attribute preservation without retraining large diffusion models.

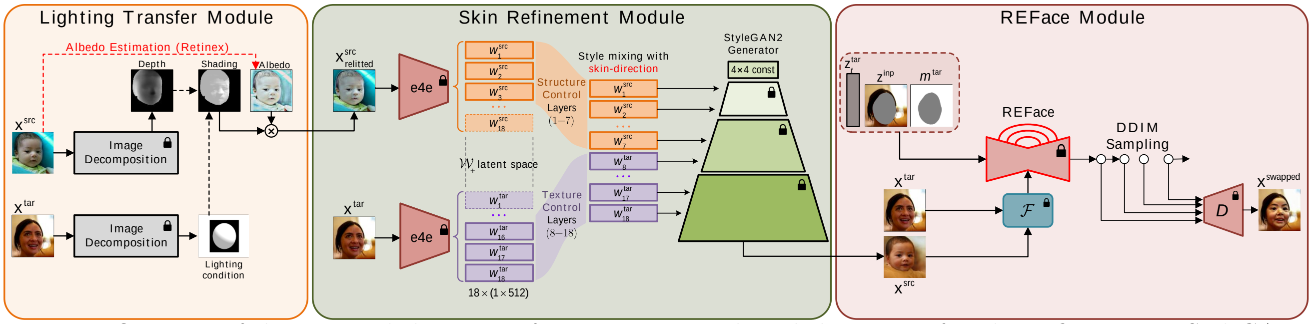


Figure 1: Overview of the proposed three-stage face swapping pipeline: lighting transfer, skin refinement in StyleGAN2 $\mathcal{W}_+$ space, and final synthesis using the REFace backbone [6].

Face Relighting and Illumination Consistency

Face relighting methods aim to improve photometric consistency by decomposing images into intrinsic components [10]. Hou et al. [11] proposed a deep intrinsic decomposition framework that estimates geometry, albedo, and lighting from a reference image following Eq. (1).

$$L_o(\mathrm{x},\omega_o) = a(\mathrm{x}).s(\mathrm{x}) \tag{1}$$

where $L_o(\mathrm{x},\omega_o)$ is the intensity of the pixel at the position (x), $a(\mathrm{x})$ and $s(\mathrm{x})$ are the albedo and the shading at that pixel, respectively. Their network predicts depth, albedo, and lighting using replicated decoders and MLPs, and generates a differentiable shadow mask based on ray-tracing principles [12]. This formulation produces shadows consistent with facial geometry, enabling physically plausible relighting.

GAN Inversion-based Image Manipulation

GAN inversion-based image manipulation aims to map an input image into the latent space of a pre-trained generative model, where semantic attributes can be edited by traversing specific directions [13, 14]. While early approaches performed inversion in the entangled $\mathcal{Z}$ space [15, 16], subsequent works favored the more expressive and disentangled $\mathcal{W}$ and $\mathcal{W}_+$ spaces introduced with StyleGAN [14], enabling finer and more controllable image edits. A key component of GAN inversion is the encoder used to project images into the latent space. Tov et al. [17] proposed *e4e*, an encoder designed to produce latent codes that closely follow the distribution learned by StyleGAN, resulting in improved reconstruction quality and editability in $\mathcal{W}_+$. Building upon this representation, style mixing [18] enables independent manipulation of coarse attributes (layers 1–7), such as structure and pose, and fine attributes (layers 8–18), including texture, color, and lighting, thus providing precise control over image synthesis.

Beyond encoder-based methods, several researches have explored direct latent space editing. InterfaceGAN [19] formulates attribute manipulation as a linear traversal in latent space along directions defined by the normal vectors of SVM decision boundaries. While this approach offers an intuitive and interpretable editing framework, it is inherently limited to attributes with well-defined binary labels and reliable external attribute predictors.

REFace: A diffusion-based framework for Face Swapping

REFace [6] is a conditional in-painting framework built upon latent diffusion models [20] for high-fidelity face swapping. It formulates the task as reconstructing a complete face from a masked target image while being guided by a composite conditioning vector. This conditioning combines identity embeddings extracted from the source face with target attributes such as pose, expression, and facial geometry, enabling explicit control over both identity transfer and attribute preservation. The in-painting formulation further allows the model to leverage contextual information from the unmasked regions, which improves structural coherence and boundary consistency.

To ensure identity preservation and perceptual realism, REFace optimizes a weighted combination of diffusion loss, identity loss, and perceptual loss. The identity loss enforces similarity between the generated face and the source identity in a deep feature space, while the perceptual loss encourages high-level semantic consistency with the target image. This multi-term objective enables REFace to achieve state-of-the-art visual fidelity and identity consistency across diverse facial poses and expressions.

Methodology

Our proposed method is a novel post-training three-stage face swapping pipeline that systematically handles different aspects of the synthesis process: illumination transfer, identity and skin attribute blending in latent space, and ultimately high-fidelity face synthesis, as depicted in Fig. 1. The *Lighting Transfer Module* first mitigates lighting inconsistencies between the source image I_{src} and the target one I_{tar}, which is a common cause of unrealistic blending. This step ensures that the source image is adapted to the target's lighting condition before the main synthesis occurs. Then the *Skin Refinement Module* manipulates texture features of the image in the disentangled space $\mathcal{W}_+$. This module performs perceptual identity transfer while ensuring accurate and natural dissemination of the target skin tone. Finally, a *face swapping backbone* REFace [6] is inherited and integrated with two previous modules in a fully-parametric-frozen scheme to synthesize high-fidelity swapped results.

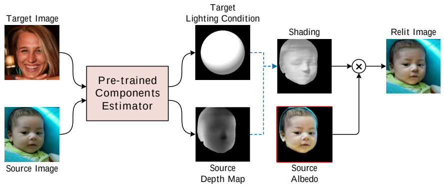


Figure 2: Diagram of the image relighting methodology.

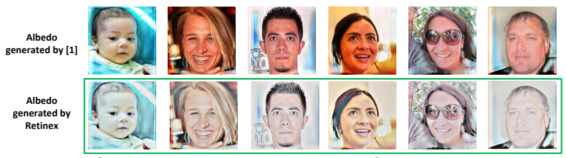


Figure 3: Qualitative comparison of albedo estimation: lighting leakage in [11] versus cleaner results from the proposed Retinex-based approach.

Lighting Transfer Module

The Lighting Transfer Module (Fig. 2) is inspired by the model introduced in [11]. It estimates the albedo $\mathbf{A}_p$ and depth map $\mathbf{D}_p$ from the source image and extracts the target lighting condition ω_t, which is then applied to compute shading based on the source geometry $\mathbf{D}_p$. The final relit image is obtained by combining this shading with the source albedo using Eq. (1).

To further reduce lighting leakage in the albedo and improve intrinsic decomposition, we adopt a Retinex-based approach [21] in the albedo estimation step. By separating reflectance and illumination components, Retinex produces cleaner albedo maps that better preserve the source's intrinsic properties, as illustrated in Fig. 3.

We investigate the impact of the relighting process by comparing face swapping outputs with and without the proposed Lighting Transfer Module. For each source-target pair, a relighting error heatmap is computed, representing the absolute pixel-wise difference between the generated and target faces. Across all test cases (Fig. 4), darker regions indicate higher errors. Results with the Lighting Transfer Module show consistently lower errors and improved lighting consistency, validating its effectiveness.

Skin Refinement Module

The aim of the *Skin Refinement Module* is to transfer the skin tone from a target image x^{tar} to a source image x^{src} while preserving the unique identity of the source. Both images are embedded into their respective latent codes within $\mathcal{W}_+$ space using a pre-trained *e4e* encoder (see middle block of Fig. 1). Style mixing [18] is then applied: the shallow layers (1–7) of the source latent code, which encode structural identity, are concatenated with the deeper layers (8–18) of the target latent code, which control fine-grained attributes such as texture and color. This hybrid latent vector is subsequently fed to the pre-trained StyleGAN2 generator to produce the skin-refined output image $x^{\text{src}'}$.

To improve the accuracy of skin tone transfer, we adopt attribute direction-based image manipulation introduced in [19]. Nevertheless, this framework was originally designed for binary, well-annotated attributes, whereas skin color is continuous and lacks explicit labels in most public datasets.

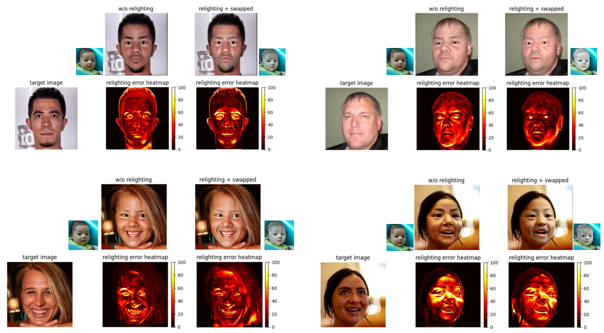


Figure 4: Relighting error heatmaps illustrating improved consistency with the proposed Lighting Transfer Module on four sample pairs.

To mismatch limits the applicability and reliability of existing approaches, motivating us to propose an extension for manipulating undefined continuous attributes.

Specifically, to identify the skin direction in the absence of explicit annotations, we adopt a proxy score. We first sample 20,000 image-latent code pairs from StyleGAN2 and apply the 3D Morphable Model (3DMM) [22] to extract attribute coefficients for each sample. We retain only the texture coefficients, denoted by $\delta \in \mathbb{R}^{80}$, and discard the remaining components. These texture coefficients serve as a high-dimensional proxy for skin tone, enabling us to sort and group images along the corresponding feature axis.

Training a linear SVM to extract the normal vector of a semantic boundary, as in InterfaceGAN [19], is not applicable in our settings due to the continuous and undefined nature of skin tone. Instead, we define the skin tone direction as a vector $\mathbf{n}$ connecting the centroids of the two clusters in latent space $\mathcal{W}_+$, obtained by applying K-means clustering in the proxy coefficient space $\mathcal{D}$.

Given the latent codes of the source z_{src} and the target z_{tar}, we obtain the skin-transferred latent code by projecting the displacement $z_{\text{tar}} - z_{\text{src}}$ onto the direction $\mathbf{n}$ and moving along this direction starting from z_{src}:

$$z_{\text{edit}} = z_{\text{src}} + \alpha_{\text{edit}} \times \mathbf{n} \quad \text{with } \alpha_{\text{edit}} = \langle z_{\text{tar}} - z_{\text{src}}, \mathbf{n} \rangle \in \mathbb{R} \quad (2)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product operator.

Fig. 5 visualizes the clustering results using the first two principal components of the texture coefficients δ, where the two clusters naturally separate faces into lighter and darker skin tone groups. The resulting direction $\mathbf{n}$ serves as a robust manipulation axis not only for skin tone adjustment but also for other continuous attributes, enabling precise and controllable edits by moving any latent code z along this direction by a signed distance α_{edit}.

Qualitative experiments confirm the effectiveness of the proposed method for skin tone manipulation (see Fig. 6). The results show that the approach consistently produces skin tones that perceptually match those of the target faces. In contrast, omitting the skin tone direction leads to intermediate color blending, with the generated results retaining noticeable influence from the source skin tone.

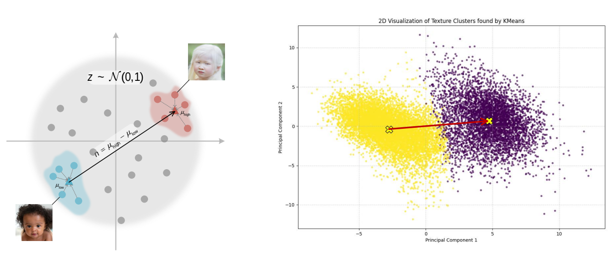


Figure 5: The idea of identifying the skin direction in our project (left) and illustration of the skin direction derived from K-Means clustered 3DMM texture coefficients (right).

Optimization-free REFace Module

The face swapping pipeline concludes by leveraging the pretrained REFace diffusion model [6]. The modified source images are directly fed into the reverse diffusion process, without any fine-tuning of either the main network or the conditioning adapter $\mathcal{F}$. Unlike prior approaches that rely on parameter updates, which introduce additional complexity and potential instability, the proposed method preserves the injected lighting and skin information without modifying the model parameters. Despite the induced domain shift, experimental results show no detrimental impact on performance. This strategy retains the strong generative priors and stability of the baseline model while aligning source and target information into a more compatible representation, enabling accurate and robust face swapping.

Experiments

Experimental Setup

Dataset Following the baseline approach [6], we evaluate our method on the CelebAMask-HQ dataset [23]. This is a large-scale and high-resolution dataset consisting of 30,000 1024×1024 face images with pixel-level segmentation masks. Without any fine-tuning, all our experiments only consider the fixed test partition of 1,000 images as designed in [6]. This test set allows us to assess the generalization capability of our proposed approach.

Evaluation Metrics The evaluation of face swapping methods encompasses multiple dimensions, reflecting various aspects of performance and fidelity. For evaluating *identity preservation* between the swapped faces and the source faces, inspired by FaceShifter [2], we employ ArcFace [24] to extract ID features from both the source and swapped images. These features are then used to compute cosine similarity, yielding the $ID\ feature\ similarity$ score, where values close to 1.0 indicate strong identity preservation.In addition, for each swapped result, we identify the most similar face among all source images and report Top-1 and Top-5 matches as the $ID\ retrieval$ score. As with $ID\ feature\ similarity$, higher values reflect better preservation of identity.

To assess the realism and quality of the swapped faces, we use the Fréchet Inception Distance (FID) [25], which

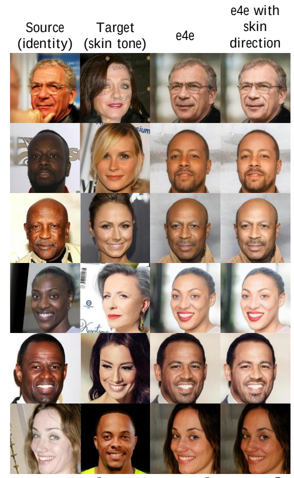


Figure 6: Comparison between skin refinement utilizing *e4e* encoder only (third column) and with skin direction approach (last column).

compares high-level statistical features between the target faces and swapped results. Lower FID values indicate that the generated faces more closely resemble real images in terms of both realism and diversity.

For evaluating fidelity and target attribute preservation, we measure *pose* and *expression* errors using HopeNet [26] and the 3DMM face reconstruction model [22]. Lower error indicates better preservation of these attributes. Additionally, the Structural Similarity Index (SSIM) [27] is used to assess the visual quality of the reconstructed area and detect potential artifacts. SSIM evaluates luminance, contrast, and structural similarity between target and swapped images, with higher values corresponding to higher fidelity and perceptual quality.

Implementation Details We implement our method using the PyTorch framework and conduct all experiments on Quadro RTX 6000 GPUs (24GB) hosted on the CRIANN server[1]. The size of input images is set to 512×512 to accommodate computational constraints. To ensure a fair and reproducible comparison with the baseline, we adopt the same publicly available ArcFace [24] face recognition model and the CLIP L14 [28] foundation model. This setup allows consistent evaluation while leveraging the computational resources provided by the CRIANN structure.

Experimental results

Our evaluation results are reported in Tab. 1. For **FID**, **Pose**, **Expression**, lower values indicate better performance, whereas higher values correspond to better performance for all other metrics. The REFace baseline returns the best performance in terms of identity preservation, achieving the highest scores for $ID\ feature\ similarity$ (**0.632**) and $ID\ retrieval$ (**96.3%** and **98.6%**). However, this high identity performance comes at a cost to other critical aspects of the face swapping task, such as facial ar-

[1]Normandy Regional Center for Computing and Digital Applications

Table 1: Quantitative comparison between the baseline and our proposed pipeline (**Ours**) on CelebAMask-HQ dataset.

Method	ID Feature Similarity ↑	ID Retrieval ↑		FID ↓	Pose ↓	Expression ↓	SSIM ↑
		Top 1 (%) ↑	Top 5 (%) ↑				
Baseline	**0.632**	**96.3**	**98.6**	7.435	3.184	0.95	0.743
Ours	0.498	81.6	90.7	**7.160**	**3.089**	**0.94**	**0.767**

Figure 7: Qualitative comparison between the whole pipeline (**Ours**) and the baseline model.

tifacts, lighting conditions, and skin tone discrepancies as illustrated in Fig. 7). By combining the lighting transfer and skin refinement modules, the proposed pipeline intentionally trades a slight reduction in identity similarity for improved realism and attribute fidelity, yielding the best overall results. As a result, our system achieves the best performance across multiple metrics: an FID of **7.160**, a pose error of **3.089**, an expression error of **0.94**, and the highest $SSIM$ score of **0.767**. These results indicate that, although the baseline better preserves identity, the proposed pipeline offers a superior trade-off by significantly enhancing perceptual quality, visual coherence and faithful attribute transfer.

Figure 7 presents qualitative examples that are consistent with the trends observed in our quantitative analysis. While the REFace baseline produces competent swaps, our method consistently generates more natural and coherent images. The key difference lies in the superior handling of lighting and skin tone, which are better matched to the target environment thanks to our preprocessing modules. Our final proposed method provides the most visually appealing and realistic results. It successfully transfers the source identity while faithfully preserving the target's expression and pose. The seamless blending and absence of major artifacts highlight the performance of the proposal.

Conclusion

This paper introduces a novel plug-and-play three-stage face swapping pipeline that requires no optimization beyond publicly available weights, enabling practical and efficient deployment. A *Lighting Transfer Module* aligns the source image to the target illumination, a *Skin Refinement Module* blends identity and skin attributes in the disentangled $\mathcal{W}_+$ latent space, and a *final synthesis stage* leverages the pretrained REFace diffusion backbone. Experimental results demonstrate that the proposed modular approach significantly improves visual realism, photometric consistency, and overall fidelity compared to the baseline, achieving a FID of 7.16. While the baseline exhibits stronger identity preservation, the proposed method attains superior perceptual quality by disentangling lighting and skin tone challenges into dedicated preprocessing modules. Without additional training or fine-tuning, the proposed framework provides a robust and effective solution for high-fidelity face swapping.

References

[1] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2387–2395, 2016.

[2] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. Faceshifter: Towards high fidelity and occlusion aware face swapping. *arXiv preprint arXiv:1912.13457*, 2019.

[3] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2003–2011, 2020.

[4] Zhian Liu, Maomao Li, Yong Zhang, Cairong Wang, Qi Zhang, Jue Wang, and Yongwei Nie. Fine-grained face swapping via regional gan inversion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8578–8587, 2023.

[5] Wenliang Zhao, Yongming Rao, Weikang Shi, Zuyan Liu, Jie Zhou, and Jiwen Lu. Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8568–8577, 2023.

[6] Sanoojan Baliah, Qinliang Lin, Shengcai Liao, Xiaodan Liang, and Muhammad Haris Khan. Realistic and efficient face swapping: A unified approach with diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1062–1071. IEEE, 2025.

[7] Yuval Nirkin, Yosi Keller, and Tal Hassner. Fsgan: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7184–7193, 2019.

[8] Yuhao Zhu, Qi Li, Jian Wang, Cheng-Zhong Xu, and Zhenan Sun. One shot face swapping on megapixels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4834–4844, 2021.

[9] Kihong Kim, Yunho Kim, Seokju Cho, Junyoung Seo, Jisu Nam, Kychul Lee, Seungryong Kim, and KwangHee Lee. Difface: Diffusion-based face swapping with facial guid-

ance. *Pattern Recognition*, 163:111451, 2025.

[10] Chen Li, Kun Zhou, and Stephen Lin. Intrinsic face image decomposition with human face priors. In *European conference on computer vision*, pages 218–233. Springer, 2014.

[11] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4217–4226, 2022.

[12] Arthur Appel. Some techniques for shading machine renderings of solids. In *Proceedings of the April 30–May 2, 1968, spring joint computer conference*, pages 37–45, 1968.

[13] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. *Advances in neural information processing systems*, 33:9841–9850, 2020.

[14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[15] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[16] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[17] Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. Designing an encoder for stylegan image manipulation. *ACM Transactions on Graphics (TOG)*, 40(4):1–14, 2021.

[18] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2287–2296, 2021.

[19] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9243–9252, 2020.

[20] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[21] Edwin H Land and John J McCann. Lightness and retinex theory. *Journal of the Optical society of America*, 61(1):1–11, 1971.

[22] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019.

[23] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[24] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[25] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[26] Nataniel Ruiz, Eunji Chong, and James M Rehg. Fine-grained head pose estimation without keypoints. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 2074–2083, 2018.

[27] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021.

Author Biography

Thu Hien LE received her BE from Hanoi University of Science and Technology (2022) and MS from Université Paris-Saclay (2025). She is currently pursuing a PhD in face swapping deepfake detection from Université de Caen Normandie.

Christophe Charrier received his PhD in Computer Science from Jean Monnet University (Saint-Étienne) in 1998. He was a Research Assistant at Laval University, Canada, before joining the Cherbourg Institute of Technology in 2001, where he became Full Professor in 2024. Since 2008, he has been a member of the GREYC laboratory and leads the SAFE research group. His research interests include image and video forensics, deepfake detection, image and video coding, quality assessment, and biometrics.

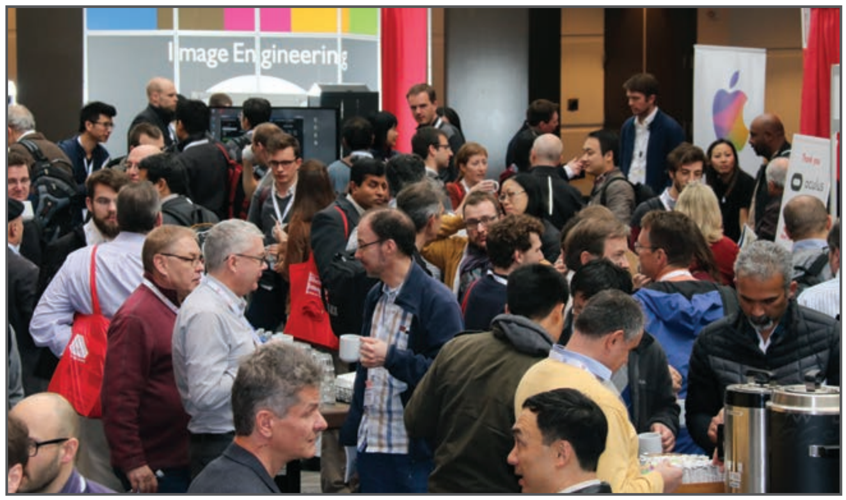
Emmanuel Giguet is a Researcher in Computer Science at CNRS (French National Center for Scientific Research). His research interests include Digital Forensics, Natural Language Processing, Document Structure Analysis, and Multimedia Forensics, with a focus on deepfake and explicit content detection. He is a member of the SAFE team at the GREYC Research Lab in Caen, France, where he leads the Digital Forensics theme. He received his PhD in 1998 and an Accreditation to Supervise Research (HDR) in 2011.

Maxime Bérubé is a Professor of Forensic Science in the Department of Chemistry, Biochemistry, and Physics at the Université du Québec à Trois-Rivières and a regular researcher with the Forensic Science Research Group. His research focuses on digital data analysis, forensic investigation, and intelligence, national security, and violent extremism, with a particular interest in the analysis of digital traces and emerging threats.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 1947 IS&T 75 2022 imaging.org