DIGITAL HERITAGE (2025)
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

Unsupervised Colorization and Diffusion-Based Virtual Try-On for Ottoman Heritage Preservation

Z. Akant, E. Ghazaei, and S. Balcisoy

Sabanci University, Faculty of Engineering and Natural Sciences,
Computer Science and Engineering, Istanbul, Turkiye

Abstract

Colorizing historical images and modernizing traditional attire are key to bridging past and present in digital heritage preservation. Accurate colorization improves the interpretation of old photos, while modernizing historical attire supports cultural adaptation and fashion preservation. This paper presents an unsupervised method for colorizing 19th century images using GANs, trained with a dataset from modern-historical films. By leveraging the GAN discriminator, realistic colorizations are generated without paired data, capturing the textures and authenticity of historical scenes. A diverse film-based dataset enables the model to generalize across eras. Additionally, historical clothing is segmented and transferred onto modern subjects using diffusion-based virtual try-on techniques. Together, these methods support cultural preservation by blending historical accuracy with modern representation.

CCS Concepts

• ***Computing methodologies*** → ***Image colorization; Computer vision tasks;*** *Generative adversarial networks;*

1. Introduction

The preservation and interpretation of historical images are vital for understanding cultural heritage. Many photographs from earlier periods have been captured in black-and-white, limiting the visual connection with the past. Colorization serves as a powerful tool to revitalize these images by adding colors based on historical context, thus enhancing the viewer's understanding of history. Moreover, transforming historical clothing into modern portraits bridges past and present, adapting traditional attire to contemporary contexts and offering new insights into cultural identity and heritage.

Colorizing historical images has gained significant attention in recent years due to its benefits for cultural and historical studies [HJJL22]. Modern deep learning techniques have led to remarkable progress in computer vision tasks such as image classification, detection, segmentation, and colorization. However, a unique challenge arises in historical image colorization due to the lack of ground truth data, as traditional supervised approaches rely heavily on paired datasets, which are often unavailable for many historical collections (images, videos, etc.), including those of the 19th century East Mediterranean era. Moreover, collecting annotated high-quality colorized historical images is a time-consuming and burdensome task. To address these issues, this paper proposes a novel approach to unsupervised image colorization, building upon DeOldify, a state-of-the-art deep learning model for automatic image colorization [Ant19]. In the absence of high-quality colored datasets, a custom dataset was created by extracting human frames from two modern historical videos, serving as a proxy for realistic color palettes. Additionally, the final phase of the colorization process was enhanced by incorporating a Generative Adversarial Network (GAN) [GPAM*14] discriminator, ensuring that the output images not only exhibit real colors but also align with historical contexts and visual authenticity.

Transforming historical clothing from 19th century portraits into modern styles has also become an interesting way to connect the past with the present. Advances in generative deep learning techniques [GPAM*14] have enabled remarkable progress in tasks such as image synthesis, style transfer, and segmentation. However, adapting historical attire to contemporary subjects presents unique challenges, including the absence of high-quality datasets and the inherent complexity of preserving cultural authenticity. Traditional supervised approaches, which depend on paired datasets, are impractical for such historical collections. To address these limitations, we developed a method that takes clothing from images and applies it to modern people using generative diffusion techniques, creating results that are both realistic and respectful of history. Our contributions can be summarized as follows:

- Development of a novel unsupervised colorization method tailored specifically for 19th century images, building upon DeOldify and GAN-based frameworks to address the challenges of ground truth absence and historical authenticity.

DOI: 10.2312/dh.20253174

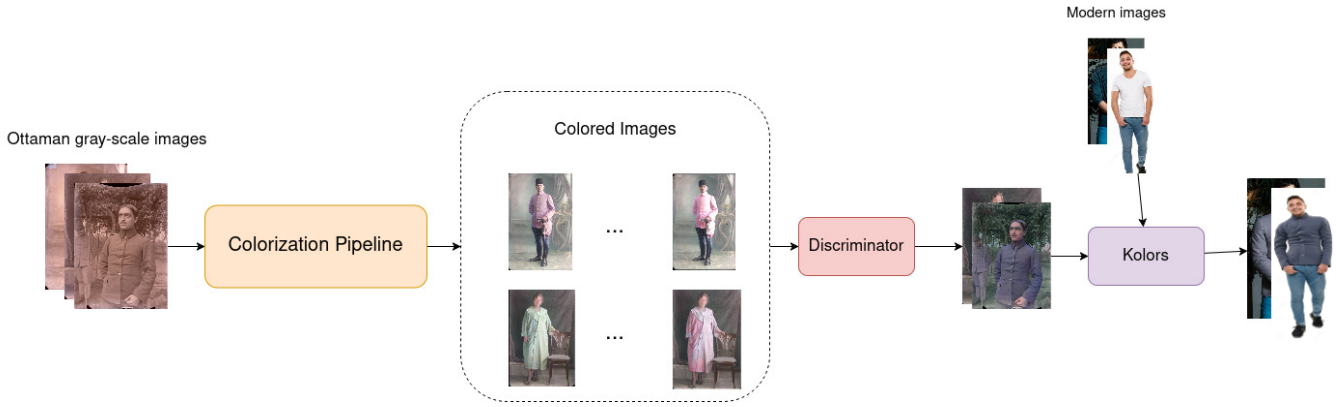

Figure 1: ***Schematic Overview of the 19th Century East Mediterranean Portrait Colorization and Clothing Transfer Pipeline:*** *This pipeline transforms grayscale 19th century portraits into colorized and modernized versions. Images are first colorized using a deep network, then refined with a discriminator for realism. Garments are segmented via a fine-tuned SegTransformer and transferred to modern subjects using a diffusion-based virtual try-on model.*

- Creation of a custom dataset derived from modern historical video footage, enabling effective training and evaluation for historical image colorization.
- Integration of a GAN discriminator to refine the colorization process, ensuring enhanced alignment with historical references.
- Introduction of a method for transferring period clothing to modern subjects using generative diffusion techniques, preserving cultural and artistic value.
- Bridging historical visuals with contemporary contexts, providing a framework to analyze and reimagine historical images with cultural respect and authenticity.

This paper is organized as follows: We start with a review of related studies in the related work section. Then, we explain our methodology, present our findings in the Results section, and conclude by highlighting the main outcomes of our wo-rk.

2. Related Work

Recent deep learning advances have improved image colorization, historical restoration, and virtual clothing transfer. This section reviews related work and their connection to our approach.

2.1. Unsupervised Colorization

Several methods have tackled grayscale image colorization without large labeled datasets. Larsson et al. [LMS16] used CNNs to predict color histograms, and Zhang et al. [ZIE16] added class-based priors for realism. GANs, like in Isola et al. [IZZE17], have also advanced colorization through image-to-image translation. Our approach builds on these by focusing on 19th century East Mediterranean photographs, fine-tuning models with domain-specific data from historical films. This enhances preservation of cultural textures and styles, improving color accuracy for heritage images beyond general methods.

2.2. Diffusion-Based Virtual Try-On

Virtual try-on has advanced with deep learning, enabling realistic clothing transfer. Diffusion-based models like [ZZX*24] use Transformers for multi-garment try-on and texture transfer, while [GSZ*23] employs exemplar-based inpainting to preserve garment coherence. Our approach segments 19th century East Mediterranean clothing and applies it to modern subjects using generative diffusion models. Unlike typical retail-focused methods, we repurpose the Kolors [Tea24] diffusion model without retraining to visualize historical attire in a contemporary context.

3. Methodology

The colorized images are refined by a Discriminator that evaluates their realism and filters out unnatural results. Next, the Kolors [Tea24] enhances the images by blending historical features with modern color styles. The final outputs preserve the historical essence of 19th century East Mediterranean photos while offering more realistic, visually appealing colors for improved interpretation.

3.1. Problem Definition

Image restoration for historical photos involves two main tasks: colorization and transformation.

Colorization $\mathcal{C}(I) \rightarrow I_c$ maps a grayscale image I to a colorized image I_c, aiming for plausible, historically accurate colors without paired ground truth, requiring unsupervised methods due to limited datasets.

Transformation $\mathcal{T}(I_h) \rightarrow I_m$ converts historical elements like clothing in I_h to modern equivalents I_m, preserving structure while handling domain differences between historical and modern styles.

3.2. DeOldify

DeOldify is an open-source deep learning model for automatic image and video colorization, leveraging GANs and CNNs to produce

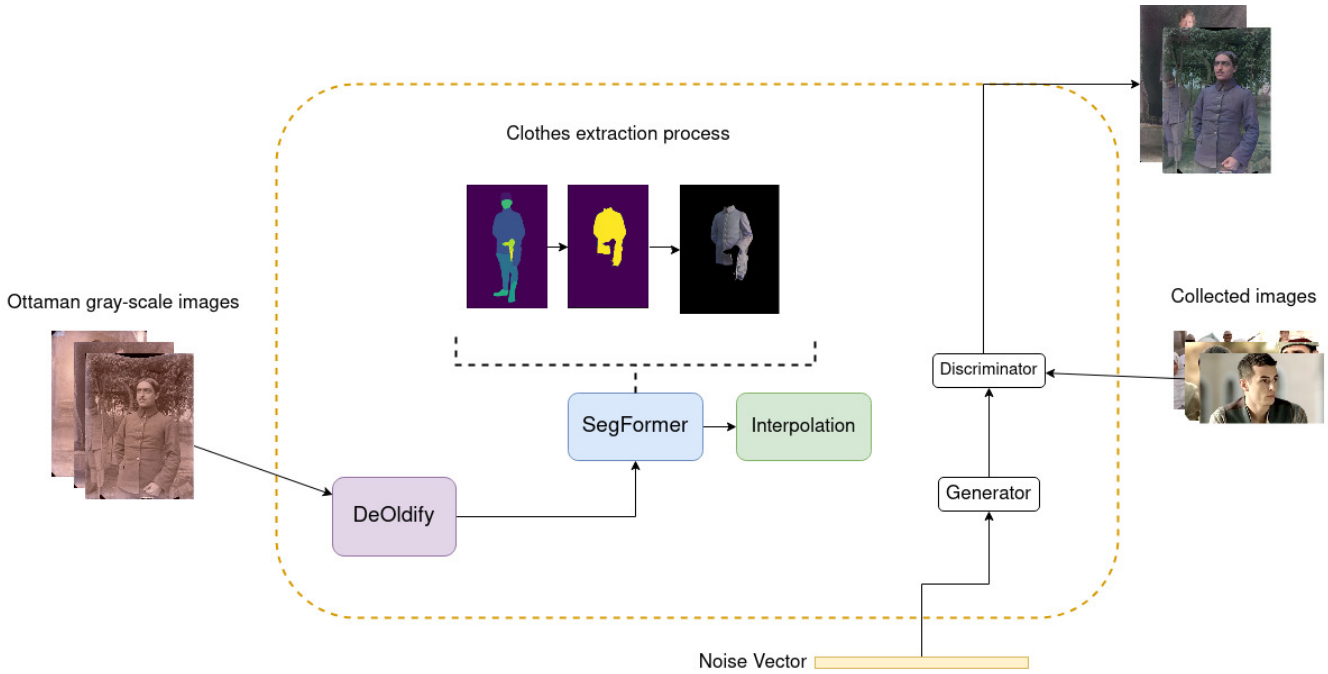

Figure 2: *The figure illustrates the pipeline for colorizing 19th century East Mediterranean grayscale images [Arcnd]. Grayscale inputs are first processed by DeOldify, a pretrained GAN model preserving facial and scene details but less effective on historical clothing. The colorized results are then refined by a discriminator network that filters out unrealistic outputs based on realistic examples from the collected dataset.*

high-quality, realistic results.
Trained on a large dataset, it employs perceptual and adversarial loss functions to enhance color realism, making it ideal for historical media restoration. In this paper, DeOldify's transferred weights are frozen, as it effectively colorizes subtle details like faces but struggles with clothing due to the absence of 19th century East Mediterranean images in its training data [Ant19].

3.3. SegFormer

SegFormer [XZL*21] is a state-of-the-art semantic segmentation model that combines transformer architectures [VSP*17] with a lightweight, hierarchical design to capture local and global context effectively. It is used to extract clothing parts from images and generate varied colors by interpolating between generated images $\mathbf{I}_{gen} \in \mathbb{R}^{H\times W\times C}$ and reference RGB images $\mathbf{I}_{rgb}$. The interpolation is defined as:

$$\mathbf{I}_{interp} = \alpha \cdot \mathbf{I}_{gen} + (1-\alpha)\cdot \mathbf{I}_{rgb},$$

where $\alpha \in [0,1]$ is the interpolation factor.

3.4. Generative Adversarial Networks

GAN [GPAM*14] consists of two neural networks, a generator G and a discriminator D, which are trained simultaneously in a minimax game. The generator G aims to create realistic data samples (e.g., images), while the discriminator D attempts to distinguish between real samples from the training dataset and fake samples generated by G.

The discriminator is designed to classify inputs as either real (from the true data distribution) or fake (generated by the GAN's generator). In the context of image generation tasks, the discriminator learns to evaluate the realism of generated images based on features such as color distribution, texture, and overall coherence. By extracting the discriminator from the GAN and using it independently, it can serve as a powerful tool for distinguishing high-quality, realistic images from less plausible ones.

3.5. Clothing Segmentation

The first step is to extract clothing from historical portraits. A machine learning model called `SegFormer_B2_clothes` [Mat22] is utilized to do this. This model is part of a family of Vision Transformers, designed to work well with images. The `SegFormer_B2_clothes` model was fine-tuned using the ATR dataset, which contains labels for various clothing and accessory categories. The base model is `SegFormer` [XWY*21]. The labels in this dataset include:

Although `SegFormer_B2_clothes` was fine-tuned on modern fashion datasets, we adapt it to segment garments from colorized historical portraits. This cross-domain application demonstrates that existing segmentation models, when combined with culturally aware colorization, can be effectively repurposed for digital heritage tasks.

Figure 3: *This figure compares colorization methods on 19th century East Mediterranean photos. From left to right: grayscale input, DeOldify, Zhang et al., and our method. DeOldify tends toward stylized tones; Zhang et al. improves color diversity but lacks detail. Our model, trained on historical data, achieves more realistic, texture-rich, and historically accurate results.*

Hat	Sunglass	Upper-clothes
Skirt	Pants	Dress
Belt	Left-shoe	Right-shoe
Bag	Scarf	

Table 1: *Clothing Segmentation Labels*

3.6. Diffusion-Based Virtual Try-On

Kolors [Tea24] is a U-Net–based latent diffusion model by Kuaishou Technology, optimized for high-fidelity text-to-image synthesis. It uses large-scale image-text datasets and novel noise scheduling to enhance high-resolution output consistency. Features like bucketed sampling enable flexible aspect ratios with efficient computation. While designed for general use, Kolors excels in virtual try-on tasks, realistically rendering intricate 19th century East Mediterranean textiles on modern subjects. We apply Kolors directly to segmented, colorized historical garments without retraining, creatively repurposing it for cultural heritage visualization.

3.7. Results

We compared our method with DeOldify and Zhang et al.'s models. DeOldify produces soft, pastel colors but lacks historical accuracy, while Zhang et al. improve color variety but lose fine texture details. Our GAN-based model, trained on a dataset from historical films, achieves more realistic and historically accurate colorizations, especially for detailed 19th century East Mediterranean clothing. Using the Kolors [Tea24] diffusion model, we transferred segmented garments onto modern subjects, demonstrating the potential of combining colorization with virtual try-on for heritage preservation and fashion adaptation. Overall, our approach effectively colorizes historical images and supports cultural applications.

4. Conclusion

This study introduces a novel unsupervised approach using transfer learning and a GAN discriminator to colorize and transform 19th century images. Segmentation enables adapting historical clothing to modern styles while preserving historical accuracy. Results show realistic colorizations. Future work may develop interactive platforms for immersive cultural heritage experiences using AI.

References

[Ant19] ANTIC J.: Deoldify: A deep learning-based project for image and video colorization. https://github.com/jantic/DeOldify, 2019. 1, 3

[Arcnd] ARCHIVE S. R.: Ottoman bank individual collection, n.d. URL: https://archives.saltresearch.org/handle/123456789/2503. 3

[GPAM*14] GOODFELLOW I., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)* (2014), pp. 2672–2680. 1, 3

[GSZ*23] GOU J., SUN S., ZHANG J., SI J., QIAN C., ZHANG L.: Taming the power of diffusion models for high-quality virtual try-on with appearance flow. In *Proceedings of the 31st ACM International Conference on Multimedia* (2023). 2

[HJJL22] HUANG S., JIN X., JIANG Q., LIU L.: Deep learning for image colorization: Current and future prospects. *Engineering Applications of Artificial Intelligence 114* (2022), 105006. 1

[IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 1125–1134. 2

[LMS16] LARSSON G., MAIRE M., SHAKHNAROVICH G.: Learning representations for automatic colorization. In *European Conference on Computer Vision (ECCV)* (2016), Springer, pp. 577–593. 2

[Mat22] MATTMDJAGA: Segformer b2 for clothing segmentation. https://huggingface.co/mattmdjaga/segformer_b2_clothes, 2022. 3

[Tea24] TEAM K. T. K.: Kolors: Effective training of diffusion model for photorealistic text-to-image synthesis, 2024. URL: https://github.com/Kwai-Kolors/Kolors/blob/master/imgs/Kolors_paper.pdf. 2, 4

[VSP*17] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L., POLOSUKHIN I.: Attention is all you need.(nips), 2017. *arXiv preprint arXiv:1706.03762 10* (2017), S0140525X16001837. 3

[XWY*21] XIE E., WANG W., YU Z., ANANDKUMAR A., ALVAREZ J. M., LUO P.: Segformer: Simple and efficient design for semantic segmentation with transformers. *CoRR abs/2105.15203* (2021). URL: https://arxiv.org/abs/2105.15203, arXiv:2105.15203. 3

[XZL*21] XIE E., ZHANG J., LI X., LI Y., SONG S., LIU M., BAI X., YANG J., LING H.: Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems (NeurIPS)* (2021). 3

[ZIE16] ZHANG R., ISOLA P., EFROS A. A.: Colorful image colorization. In *European Conference on Computer Vision (ECCV)* (2016), Springer, pp. 649–666. 2

[ZZX*24] ZHENG J., ZHAO F., XU Y., DONG X., LIANG X.: Viton-dit: Learning in-the-wild video try-on from human dance videos via diffusion transformers, 05 2024. doi:10.48550/arXiv.2405.18326. 2