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

Text2Autochrome: Text guided autochrome synthesis using generative models

P. J. Kühn[1], S.N.Sinha[1], D.Nguyen[1,2], R.Horst[1,3], A.Kuijper[1,2] D. W. Fellner[1,2]

[1]Fraunhofer IGD, [2]TU Darmstadt,
[3]RheinMain University of Applied Sciences

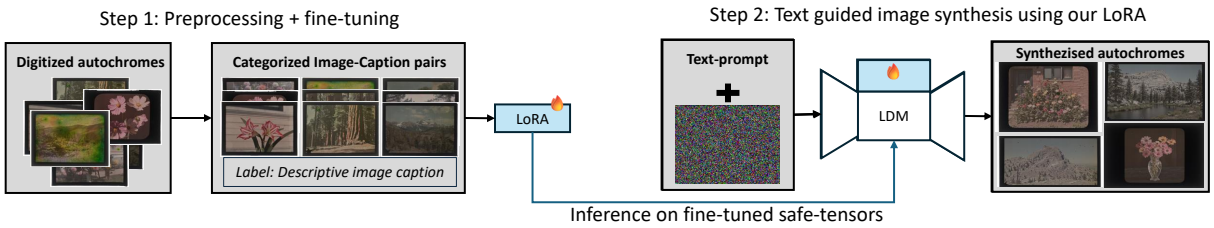

Figure 1: ***Step 1: Model Adaptation.*** *The digitized Harold Taylor autochrome collection is initially preprocessed. Subsequently, Low-Rank Adaptation (LoRA) techniques are utilized to fine-tune a Latent Diffusion Model (LDM) specifically on this dataset.* ***Step 2: Image Synthesis.*** *Conditioned on a descriptive textual prompt and an initial random noise vector, the LoRA model generates high-fidelity images designed to exhibit the characteristic visual style of autochrome photography.*

Abstract

Autochrome is an early color photography technique that is highly sensitive and prone to deterioration, limiting their public display. A limited collection of digitized autochromes exists, often with defects due to their fragile nature. We applied generative AI methods, specifically Low-Rank Adaptation (LoRA), to fine-tune diffusion models, enabling efficient use of computational resources. Our curated dataset of vintage digitized autochromes showcased various styles and served as the basis for training the LoRA model, resulting in the generation of digitized autochromes that preserved the original color filter effects and characteristic granularity. By leveraging generative AI, we can utilize the multi-modal capabilities of the model, allowing each user to generate images through concept-based prompts. This approach empowers users to creatively interact with the model, producing personalized images while maintaining the historical color fidelity and structure of autochromes. This capability also enables us to generate defect-free autochromes, which can be utilized for synthetic training in autochrome restoration efforts. We evaluated our approach using the CLIPScore metric for quantitative similarity and conducted a user study for qualitative feedback on the generated images. Our results show that the fine-tuned LoRA model effectively captures the essence of autochromes, producing visually appealing images that respect the historical aesthetic. Considering the potential for misinterpretation and ethical concerns surrounding text-to-image methods using deep learning with historical photographs, we are committed to enhancing transparency by releasing our model weights and training datasets, thereby empowering the community to better understand, evaluate, and address these important issues. Further we release an interactive demo together with the fine-tuned weights available via huggingface.

CCS Concepts

• ***Computing methodologies*** → ***Neural networks;***

1. Introduction

Autochromes, as one of the first commercially successful color photography [LG13] processes (developed by the Lumière brothers in the early 1900s), have been studied primarily from historical, technical, and preservation perspectives rather than perception testing. Autochromes are renowned for several distinctive properties that contribute to their unique aesthetic and historical significance. Firstly, they feature a characteristic color palette characterized by soft, muted colors, which are often accompanied by a visible grain structure resulting from the potato starch color fil-

DOI: 10.2312/dh.20253061

ter mosaic employed in their production. This grain structure not only adds to the textural quality of the images but also influences the overall perception of color. Additionally, autochromes exhibit a limited dynamic range when compared to modern photographic techniques, which impacts the tonal variation present in the images. Autochromes, once the leading format in early 20th-century photography, gradually fell out of favor as Kodachrome emerged, effectively addressing many of the problems associated with autochromes. Today, the production of autochromes is rare, and only digitized versions are displayed in exhibitions due to their susceptibility to light damage and various defects. Despite this decline, the soft colors and grainy texture characteristic of autochromes offer a unique retro aesthetic that resonates with contemporary audiences. Our innovative approach harnesses the power of large language models (LLMs) to generate images that evoke this nostalgic style, bringing the charm of autochromes back to life. Additionally, we can create synthetic data that simulates common autochrome defects, such as greening, oranging, and emulsion cracking. This capability allows us to develop machine learning methods aimed at repairing these defects, enhancing the quality of the generated images. Moreover, our approach can be integrated with advanced architectures like ControlNet [ZRA23], which provides the potential for consistent conditional autochrome generation. This consistency is vital for preserving authenticity and care in visual arts. Furthermore, these images possess unique luminosity and transparency qualities, enhancing their visual appeal when viewed under optimal lighting conditions. Collectively, these attributes render autochromes a fascinating subject of study in the context of historical photography and image restoration. In this paper we explore the application of diffusion-based models fine-tuned using Low-Rank Adaptation (LoRA) techniques on autochrome images from the curated Harold Taylor collection. By leveraging the strengths of diffusion models and the efficiency of LoRA, we aim to enhance the generation and quality of historical color images. Additionally, we seek to utilize these autochromes to further improve our defect correction models, thereby enabling more effective restoration and enhancement of archival images. Our approach not only preserves the intricate details of the original dataset, but also demonstrates the potential of advanced generative techniques to address challenges within image correction and restoration. However, it is constrained by the underlying LLM's ability to generate scene coherence based on text prompts, meaning we primarily reproduce the aesthetic aspects of autochrome photography rather than fully capturing the complexity of the original scenes. In summary our contributions are the following:

- We fine-tune a diffusion model through LoRA techniques to effectively generate authentic autochromes.
- We utilize the openly available Harold Taylor collection [Tay] as our dataset. This collection will be systematically categorized to facilitate the training process, ensuring that the data is organized and representative of various autochrome characteristics necessary for model fine-tuning.
- We present quantitative results to evaluate the performance of our model using the CLIPScore.
- We have a qualitative observation of representative generated autochromes, demonstrating the capabilities of the fine-tuned model's generative potential.
- We conduct a user study to collect subjective feedback on the generated autochromes, providing a comprehensive understanding of the effectiveness and authenticity of the model.
- We publish a demo of our model along with the fine-tuned weights to facilitate further research and application which can be found here:

2. Related Work

We now discuss previous work from relevant topics, such as generative modeling, fine-tuning strategies for generative models, the selected dataset and recent works on autochrome digitizations.

2.1. Generative Modeling

Generative Adversarial Networks (GANs): The foundational work that introduced this framework is the paper "Generative Adversarial Nets" by Goodfellow et al. [GPAM*14]. This publication detailed the adversarial training mechanism and demonstrated its potential for learning generative models without explicitly defining the data distribution. One of the primary advantages of GANs is their ability to generate high-quality and realistic samples, often surpassing the capabilities of other generative models available at the time [NM21]. The fundamental idea of adversarial training proved to be highly influential, leading to a proliferation of GAN variants like conditional GAN (cGAN) [SC21], CircleGAN [MO14] and StyleGAN [KLA19]. These advancements aimed to address the limitations of the original framework and expand the applicability of GANs to a wider range of tasks and data modalities.

Denoising Diffusion Probabilistic Models (DDPMs) emerged as another powerful class of generative model, drawing inspiration from principles of non-equilibrium thermodynamics. Ho et al. [HJA20] work demonstrated the capability of DDPMs to achieve high-quality image synthesis results, often surpassing those of GANs. One of the notable strengths of DDPMs is their training stability, which contrasts with the often unstable training dynamics of GANs. Furthermore, DDPMs are known for their ability to generate diverse and high-fidelity samples, capturing intricate details of the data distribution [EYB*23], [LMSX24], [WHP24]. However, a primary limitation of early DDPMs was the slow sampling process. Generating a single sample typically required a large number of denoising steps, making the inference process computationally intensive [UA24].

Latent Diffusion Models (LDMs) represent a significant advancement in the evolution of diffusion models, primarily addressing the computational challenges associated with operating in the high-dimensional pixel space. Rombach et al. [RBL*22] demonstrated that by operating in the latent space, it was possible to achieve a near-optimal balance between reducing computational complexity and preserving the visual fidelity of the generated images. A key architectural innovation introduced in this paper was the incorporation of cross-attention [LCW*21] layers into the model. These layers enable the conditioning of the diffusion process on various types of input, such as text-prompts or bounding boxes, facilitating tasks like text-to-image synthesis and image inpainting

which have emerged as powerful tools for different, image synthesis tasks [SCS*22], [RDN*22], [NDR*22]. Podell et. al. [PEL*23] then scaled this approach by introducing larger model backbones in the work SDXL to achieve higher image resolution outputs.

Flow Matching: The concept of Flow Matching was introduced by Lipman et al. [LCBH*23]. This work presented a novel training objective that allows for the use of general Gaussian probability paths, which includes diffusion paths as a specific instance. The research found that employing Flow Matching with diffusion paths resulted in a more robust and stable alternative for training diffusion models. The approach has demonstrated its effectiveness across various domains, including image generation [EKB*24], video synthesis [JSL*25], and even modeling biological structures [LK25].

Adversarial Diffusion Distillation (ADD): Sauer et. al. [SLBR23] designed training methodology to enable efficient sampling from large-scale diffusion models in a remarkably small number of steps, typically just 1 to 4, while preserving a high level of image quality [SLBR23]. The core of ADD lies in its innovative combination of Score Distillation Sampling (SDS) [AKS24] and adversarial training. An adversarial loss is incorporated to ensure that the images generated in this highly accelerated regime maintain high fidelity and visual appeal.

Fast High-Resolution Image Synthesis with Latent Adversarial Diffusion Distillation (LADD) builds upon the principles of ADD to achieve even greater efficiency and scalability [SBD*24]. LADD presents an innovative distillation method that directly addresses and overcomes the limitations encountered in pixel-based approaches like ADD, especially when aiming for high-resolution image synthesis. The most significant innovation in LADD is its operation within the latent space of pre-trained diffusion models. This strategic shift simplifies the training process and significantly enhances performance, particularly for generating high-resolution images with multiple aspect ratios.

Low-Rank Adaptation (LoRA): The technique was proposed by Hu et al. [JSW*22]. The research demonstrated that fine-tuning updates to the weight matrices of LLMs often have a low "intrinsic rank." By approximating these updates with low-rank matrices, LoRA can drastically reduce the number of trainable parameters. LoRA has also proven to be a valuable tool for efficient super-resolution using diffusion models. AdaptSR, a LoRA framework, efficiently repurposes bicubic-trained super-resolution models for real-world tasks [KMT25]. By selectively updating only lightweight LoRA layers while keeping the pre-trained backbone intact. This efficient adaptation not only reduces memory and compute requirements but also outperforms GAN and diffusion-based Super Resolution (SR) methods while training significantly fewer parameters. Furthermore, LoRA plays a crucial role in multi-concept customization within diffusion models [YCZ*24]. LoRA-Composer is a training-free framework designed for seamlessly integrating multiple LoRAs to enhance the harmony among different concepts within generated images [YWP*24]. To further enhance the efficiency of diffusion models, Guo et. al. [GLD*24] have explored integral Low-Rank Adaptation of quantized diffusion models. IntLoRA proposes using integer-type low-rank parameters to adapt quantized diffusion models. By operating in integer arithmetic, IntLoRA offers advantages in terms of reduced memory usage during fine-tuning, lower storage requirements, and hardware-friendly acceleration during inference.

2.2. Digitization and AI based approaches for autochromes

Digitization: Digitizing early regular color screen photographs like Dufaycolor, Finlay Colour, and Paget Colour plates poses challenges because standard high-resolution stitching techniques fail due to the repeating geometric patterns, creating artifacts [CM23]. Hubička et al. developed a novel method using custom 'ColorScreen' software to analyze screen patterns for precise stitching, addressing high resolution, dynamic range, distortion, and degradation issues, successfully implemented at the National Geographic Society (NGS) [CM23]. Similarly, autochromes require extremely high resolution to capture their discrete color granules correctly [Pet19]. Peterson details the NGS and DT Heritage collaboration for preservation-grade digitization of the Society's large autochrome, dufaycolor, and finlaycolor collections, highlighting workflow testing [Pet19]. Wolska describes conserving and digitizing a Polish autochrome collection, addressing fragility (e.g., detached emulsions) with methods like electrostatic film stabilization [Wol22]. Their high-resolution digitization captured multiple views, and experimental bright-field lighting was tested to reproduce the Callier effect for stereoscopic plates, finding it useful for documentation despite emphasizing defects [Wol22]. These studies highlight the necessity of specialized, collaborative approaches for these materials.

AI based approaches: Deep learning techniques, like the Channel Interaction Restoration (ChaIR) [CK23] model fine-tuned with synthetic data, show promise in restoring greening defects in autochromes, preserving image structures better than traditional methods [Kop24]. Furthermore, AI-driven monocular depth estimation can convert digitized mono-autochromes into layered images and 3D meshes for immersive VR exploration, an experience participants found enjoyable [SHC*25].

3. Methodology

Next, we discuss our dataset, the data preprocessing, the base model on which we fine tune, the LoRA parameters that were tested for training and the setup for our user study.

3.1. Dataset

The Harold Taylor collection of digitized autochromes offers a valuable glimpse into early 20th-century California through the pioneering autochrome process. Photographer Harold A. Taylor, who immigrated to California in 1896 and later settled in Coronado, captured a diverse range of subjects, including landscapes of Yosemite, numerous floral studies reflecting his involvement with the Coronado Floral Association, and scenes of coastal California. Created primarily between the 1910s and early 1930s, these fragile color transparencies represent a significant step in the history of photography. Preserved and made accessible online by the Coronado Historical Association in partnership with California Revealed, the collection provides a unique and colorful window into a bygone era of California's natural beauty and cultural landscape [Tay]. It is a public collection of about 400 different high resolution digitized autochromes. The digitizations come with a resolution of 6300x5100. Due to the nature of autochrome photographs all digitizations are placed on small glass plates. Each image has its own

(a) ***Class****: Still Life Objects |* ***Caption****: Photo of a single red rose out of its vase on a white rose detailed cloth. To the left of the vase is a pair of scissors. To the right of the rose is a photograph in an silver oval frame. The background is a warm midtone brown wall.*

(b) ***Class****: Plant |* ***Caption****: Colorized photograph on a small glass panel depicting a branch of red cherries with trees and a clear blue sky in the background.*

(c) ***Class****: Forest |* ***Caption****: Colorized photograph on a small glass panel depicting trees and greenery next to the side of a mountain. Discoloration is present across the entire image, as well as silvering. Most of the discoloration is green, with some areas that are purple, orange, and blue. There are also bubbles in the panel.*

(d) ***Class****: Forest |* ***Caption****: Colorized photograph on a small glass panel depicting a redwood forest with different sized trees, assumedly the Mariposa Grove, a giant redwood forest in Yosemite National Park. Some areas of green discoloration are on the image.*

Figure 2: *The categorized dataset comprises diverse image-caption pairs, with representative examples for illustration. Although primarily consisting of nature scenes, the dataset intentionally includes images displaying typical autochrome artifacts, notably small greening (2c) and large-scale greening degradation (2d).*

metadata file containing a caption which is describing the content depicted by the image. Images contain mostly still life nature like trees, flowers, desserts, mountains etc.

3.2. Data preprocessing

We categorize the image-caption pairs manually and came up with 13 categories: *Body Of Water, Desert, Field, Flower, Forest, Mountain, Nature, People, Plant, Still Life Objects, Structures, Tree, Waterfall*. Figure 2 illustrates representative image-caption pairs from the training dataset, showcasing its heterogeneity. The dataset encompasses high-fidelity, artifact-free examples such as 2a and 2b, alongside instances exhibiting significant degradation, including severe greening defects (2c) and localized greening spots (2d). Further we extract the caption of each image for LoRA training, giving us the label as text-prompt for the fine tuning process. Since the high resolution digitizations do not fit into our training environment memory we resize the the images from the Harold Taylor collection to 576x408 image resolution. This results in the final dataset with 420 image-prompt pairs.

3.3. Training

We finetune the distilled model FLUX.1-dev [Doc]. We make use of Low-Rank Adaptation (LoRA). LoRA is a technique used to fine-tune large pre-trained models efficiently by introducing low-rank updates to their weights. Instead of modifying the entire weight matrix, LoRA approximates the updates using two low-rank matrices, which significantly reduces the number of trainable parameters and computational costs. There are two key parameters to consider when training. The rank parameter determines the dimensionality of the low-rank matrices used in the adaptation. A higher rank allows for more expressive updates but increases the number of parameters. Conversely, a lower rank reduces the model's capacity to adapt but improves efficiency. The choice of rank is a trade-off between model performance and resource consumption. The alpha parameter is a scaling factor that controls the strength of the low-rank adaptation. It adjusts the contribution of the low-rank updates to the overall model output. A higher alpha value amplifies the impact of the adaptations, while a lower value diminishes their effect. This parameter is crucial for balancing the original model's knowledge with the learned adaptations. Figure 3 shows the training input, we fine-tune on the categorized collection of ca. 420 digitized image-caption pairs. The training was done on an A100 40GB GPU and 128GB RAM using the AdamW optimizer with a 1e-4 learning rate and a batch size of 4. We train the model with four different LoRA rank and alpha combinations shown in Table 1. All experiments are trained for 8000 steps.

LoRA	Rank	Alpha	Steps
0	16	16	8000
1	16	32	8000
2	32	32	8000
3	32	64	8000

Table 1: *Summary of the four distinct training configurations employed in this study, each characterized by a unique combination of LoRA rank and alpha values.*

3.4. CLIPScore calculation

We generate sample images every 1000 training steps to visually verify the model's progress in learning the desired image style concept. We do this because training configurations with a higher rank and alpha tend to recreate the target domains style earlier during training but also tend to overfit earlier which results in noisy images. For the CLIP(Contrastive Language Image Pre-training) Score calculation we take models that are configured with higher rank and alpha at checkpoints with lesser training steps, this ensures

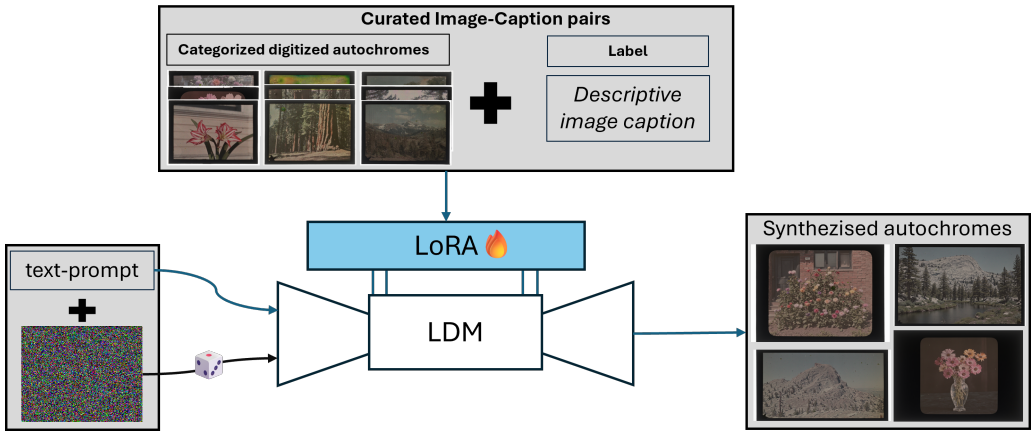

Figure 3: *The training process commenced with a curated dataset of digitized autochromes. Corresponding caption labels were extracted from the metadata files provided for each digitized instance, constituting the training input. Subsequently, the LDM Flux.1-dev underwent fine-tuning, with experimental variations in LoRA configurations. Given a descriptive text-prompt and random noise we aim to achieve optimal synthesis results.*

that the images generated by these models still depict recognizable content and not only noise. We then calculate the CLIPScore [HHF*22] of the fine-tuned models before we conduct our user-study. The score evaluates the compatibility of image-caption pairs, with higher scores indicating greater compatibility. It serves as a quantitative assessment of the qualitative notion of "compatibility." This compatibility can also be viewed as the semantic similarity between the image and the caption. Research has shown that the CLIPScore correlates well with human judgments [HHF*22]. To compute the CLIPScore, we generated 420 synthetic autochrome images. This process involved utilizing the captions associated with real digitized autochromes as input prompts. Given that these captions were previously encountered by the model during its training phase, we employed a LLM to simplify and rephrase the captions to enhance their novelty and efficacy. Additionally, we introduce variability in the aspect ratio of the generated images to enhance the diversity of our test dataset. Subsequently, we use the generated images from the best performing LoRA model and randomly select a subset from this collection for the user study. This approach ensures a representative sampling of the generated images, allowing for a comprehensive evaluation of user preferences and perceptions. By employing random selection, we mitigate potential biases and enhance the robustness of our findings, thereby facilitating a more objective assessment of the image quality and attributes as perceived by the participants.

3.5. User study preparation

Following a comprehensive evaluation of various models, we determined that LoRA **2** yields the most favorable results in terms of image generation. This assessment was conducted by a first stage evaluation of the CLIPScore. This approach allowed us to quantitatively ascertain the model's performance in replicating the unique aesthetic qualities inherent to autochromes. For the user study, synthetic autochromes were randomly sampled exclusively from the set generated by the LoRA **2** model.

4. Evaluation

Next we present our evaluation. Initially a quantitative analysis via the CLIPScore metric is done, followed by an analysis of some autochrome generations to finally discuss the extensive user study conducted exclusively with the model achieving the highest score in the first stage. During hyperparameter tuning, it was observed that training configurations employing higher LoRA ranks and alpha values achieved satisfactory style adaptation at 5000 training steps. However, extending the training duration to the initially planned 8000 steps resulted in a noticeable degradation of image quality, characterized by increased noise artifacts. Consequently, for subsequent image generation and evaluation, LoRA model checkpoints saved at the empirically determined optimal of 5000 steps were utilized. This applies for LoRA **2** and **3**.

(a) *Defect free, dark lighting conditions.* **(b)** *Defect free, bright lighting conditions.*

Figure 4: *Defect free synthetizations under different lighting conditions.*

4.1. CLIPScore discussion

Next we discuss the results of the evaluation on the CLIPScore. Table 2 shows the trained models LoRA parameters as well as the

CLIPScore results. A higher CLIPScore indicates a better performance. We notice that all scores are close together. While LoRA **0** and **1** show a good performance in adapting the style, they need more training steps and hence more time to recreate the style compared to configurations with a higher LoRA rank indicating that our strategy of using the earlier checkpoints of train configuration **2** and **3** was successful. All trained LoRA configurations revealed highly comparable performance when it comes to image generation but also show that training with higher ranks is more efficient due to the style adaptation after less training steps. The scores obtained were remarkably close, with a maximum difference of approximately one point observed between the best performing configuration **2** and the worst performing configuration **3**. This minimal variation suggests that, according to the CLIP metric which measures image-text semantic similarity, the effectiveness of the different methods in is largely equivalent. Overall, there exist differences in the CLIP-Score ranging from 33.75 to 34.93, however, they are of negligible practical significance concerning the performance.

LoRA	Rank	Alpha	Steps	CLIPScore ↑
0	16	16	8000	33.75
1	16	32	8000	34.61
2	32	32	5000	34.93
3	32	64	5000	33.96

Table 2: *CLIPScores of our trained models. A higher score inidcating a better performance. Scores were calculated using OpenAIs clip-vit-base-patch16 model [RKH*21].*

4.2. Synthesized autochrome observations

A qualitative assessment of the model's generative capabilities follows. Figures 4, 5, 6 and 7, present a selection of generated images, illustrating the diversity achievable with the fine-tuned diffusion model. For instance, examples 4a and 4b depict flowers under varied illumination, demonstrating the LoRA model's ability to synthesize plausible images across distinct lighting conditions. Furthermore, images 5a and 5b showcase the model's capacity to replicate common autochrome artifacts when prompted with descriptive embeddings such as "small green defect". The model also exhibits spatial control over defect placement, as exemplified by the localization to the lower right region in 5a. The capacity to generate larger-scale defects, also characteristic of the source material, is demonstrated in 6a and 6b, which depict substantial greening and bluing artifacts, respectively; regional conditioning is also possible for these, as shown by the defect confinement in the lower portion of 6a. The model's generalization extends beyond outdoor scenes, as 7a renders a plausible indoor still life composition with multiple objects. Finally, 7b confirms the ability to generate plausible monochromatic images and incorporate specific physical defects like frame cracks. An analysis of model limitations is presented in Figure 8. A notable decrease in generative fidelity was observed for anthropomorphic subjects. Specifically, images 8a, 8b, and 8c exhibit poorly defined or distorted facial features, rendering the synthetic nature of the images readily apparent. Challenges were also encountered with complex anatomical structures in animals; for example, 8b, 8c, 8d presents a horse generated with anatomically

(a) *Greening dot, placed in the lower right part of the image.* **(b)** *Greening dots all over the image.*

Figure 5: *Realistic defects small green dot defects occurring in real autochromes. Defects can be placed in any part of the image.*

(a) *Large defect in the lower part of the image.* **(b)** *Large defect over the image.*

Figure 6: *We are able to recreate realistic defects occuring in real autochromes like greening and bluing precisely in any part of the image.*

incorrect limbs. These observed limitations are likely attributable to the composition of the training dataset (Harold Taylor collection [Tay]). While the dataset contains some images featuring human figures, it is predominantly composed of natural scenery, leading to an under representation of human and complex animal forms during fine-tuning. Finally the model architecture supports the generation of images across a range of different aspect ratios.

4.3. User Study

We conducted a user study to draw conclusions on how humans perceive our artificially created autochromes, specifically how similar our AI-based images are perceived with respect to digitized original autochromes. The study was performed as an online survey involving 90 unpaid and voluntary participants (43 male, 42 female, 1 diverse, 4 not answered) between 19 and 85 years with Ø 39.6 and SD 14.3. On a 5-point scale, with 1 indicating no expertise and 5 indicating being an expert in the field, we asked about experience with photography and autochromes. We recruited our participants from various sources, particularly distributing our links

(a) *Still life objects placed on a table.* **(b)** *Monochrome image with a cracked frame.*

Figure 7: *We observe a strong performance on still life and monochrome nature scenery.*

within the cultural heritage community. Results indicating a heterogenous background concerning photographic skills (Ø 2.9, SD 1.2) and beginner level concerning autochromes (Ø 1.5, SD 0.9).

4.3.1. Study Design

The approach within our study was to show a set of AI-generated and original digitized autochromes to users and let them rate the pictures based on a set of criteria to evaluate the output of our method.
For the questionnaire, we derived 13 items I1-I13 based on an expert interview with a curator specialized in autochrome photography, related literature [Lan23, KOZ80, LG13, RKW23], and related studies on the digital perception of autochromes [SHC*25]. They were given a 7-point Likert scale with 0 meaning full disagreement and 6 meaning full agreement. We clustered these questions into 5 criteria as follows:

- Color (Col)

  I1: The colors of the picture appear unnatural.
  I2: The colors of the picture appear somewhat muted or faded.
  I3: The colors of the picture look soft.

- Emotion (Emo)

  I4: The image triggered emotions in me.
  I5: There is something dreamlike or ethereal about the image.
  I6: The picture looks like a painting.
  I7: The picture has a nostalgic effect on me.

- Technological aspects (Tech)

  I8: The image has a grainy or textured quality.
  I9: This image reminds me of early color photography.

- Authenticity (Auth)

  I10: I found the picture authentic.
  I11: I think it's a digitised copy of a real photo.

- Historical relatedness (Hist)

  I12: The picture looks historical.
  I13: The picture looks as if it was taken a long time ago

For the study, we randomly selected 73 generated autochromes and took digitized autochromes from a base of 420. Each image showed either flowers, forests, mountains, or single trees and was classified accordingly. For each of these classes, we randomly chose 5 images per class, which were utilized within the study, making up a pool of 20 original and 20 artificially generated autochromes. To minimize biases of individual images, each participant was given 2 images from each class, one artificial and one original, at random. This resulted in each participant rating 8 images given our 13 items.

4.3.2. Discussion of the Results

We analyzed our study's results regarding perceived similarity of the images by means of a two one-sided tests (TOST) approach after Schuirmann [Sch87], also known as the equivalence test, using the TOSTER R package [LC17]. Before testing, we chose a range of differences that are not significant for the TOST of ± 0.4 points (smaller than the commonly utilized 0.5 points on a 7-point scale) and in raw bounds, to emphasize practical relevance and directly reflect the magnitude of the effect in the original scale. Statistical measures are shown in Table 3.

Table 3: *Descriptives*

	Group	N	Mean	SD	SE
Auth	*Original*	40	3.564	0.619	0.098
	Artificial	40	3.774	0.559	0.088
Col	*Original*	60	3.599	0.900	0.116
	Artificial	60	3.474	0.730	0.094
Emo	*Original*	80	2.834	0.675	0.075
	Artificial	80	2.649	0.573	0.064
Hist	*Original*	40	3.307	0.612	0.097
	Artificial	40	3.324	0.782	0.124
Tech	*Original*	40	3.433	0.607	0.096
	Artificial	40	3.769	0.596	0.094

The results of the T-Tests to check on significance within the data, with a threshold for statistical significance of 5%, are shown in Tab. 4. The table shows that all p-values of the bounds, except for the upper bounds of Auth (p=0.114) and Tech (p=0.320), show significant differences, indicating equivalence.

For further analysis using TOST, Tab. 5 shows the 90% confidence intervals for both raw and Cohen's d. The results are visualized in Fig. 9, that shows except for Auth and Tech, the entire ranges of the respective confidence intervals of the mean differences lie within prespecified range of indifference (± 0.4), so that we conclude the TOST with 95% confidence that our participants rated the criteria Col, Emo, and Hist as equivalent. Auth overlaps only slightly, whereas Tech overlaps more, indicating that our participants did not rate these criteria of our artificially generated autochromes as equivalent to the original ones.

4.3.3. Qualitative Feedback and Limitations of the Study

Since this is pioneering research in the field of evaluating autochromes in modern days, there were no standardized questionnaires we could use that were specific to autochromes. As a result, standardization of our questions and criteria, even though discussed

(a) (b) (c) (d)

Figure 8: *Examination of edge cases highlighted the model's limitations. Synthesis of realistic human subjects, especially those with direct camera gaze (8a, 8b, 8c), and the generation of anatomically correct animals presented significant challenges (8b, 8c, 8d).*

Table 4: *T-Tests results rounded to three decimal places.*

	Statistic	t	df	p
Auth	T-Test	−1.598	78.000	0.114
	Upper bound	1.435	78.000	0.078
	Lower bound	−4.631	78.000	< 0.001
Col	T-Test	0.832	118.000	0.407
	Upper bound	3.505	118.000	< 0.001
	Lower bound	−1.841	118.000	0.034
Emo	T-Test	1.867	158.000	0.064
	Upper bound	5.910	158.000	< 0.001
	Lower bound	−2.176	158.000	0.016
Hist	T-Test	−0.110	78.000	0.913
	Upper bound	2.437	78.000	0.009
	Lower bound	−2.656	78.000	0.005
Tech	T-Test	−2.505	78.000	0.014
	Upper bound	0.470	78.000	0.320
	Lower bound	−5.480	78.000	< 0.001

Table 5: *Equivalence Bounds*

				90% Confidence Interval	
	Bounds type	Low	High	Lower	Upper
Auth	Cohen's d	−0.678	0.678	−0.736	0.009
	Raw	−0.400	0.400	−0.430	0.009
Col	Cohen's d	−0.488	0.488	−0.150	0.373
	Raw	−0.400	0.400	−0.124	0.373
Emo	Cohen's d	−0.639	0.639	0.034	0.348
	Raw	−0.400	0.400	0.021	0.348
Hist	Cohen's d	−0.569	0.569	−0.397	0.244
	Raw	−0.400	0.400	−0.279	0.244
Tech	Cohen's d	−0.665	0.665	−0.946	−0.113
	Raw	−0.400	0.400	−0.561	−0.113

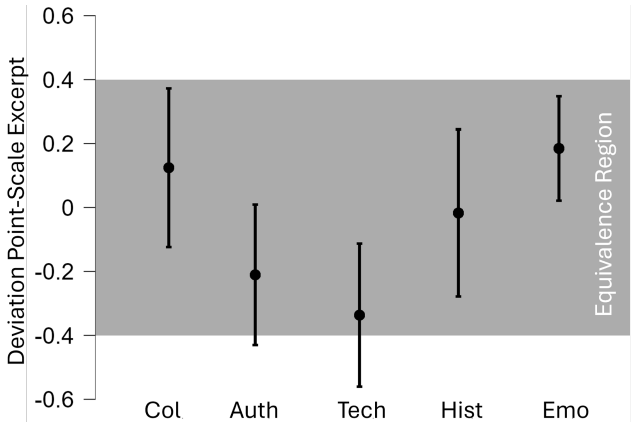

Figure 9: *Chart visualizing the equivalence analysis based on a Bland-Altman plot, using raw bounds specification.*

with and agreed by experts and further supported by literature and related work, would strengthen the study. In this matter, we provided the possibility to give us feedback in the form of free-text after filling out the questionnaire. One expert participant gave us the opinion, that 'Factors such as the focus or whether some lines would be drawn straighter in reality often played a role in the authenticity of the photos'.

Another participant communicated a general displeasure for AI-generated photographs, even though it was not mentioned within the study that some of the images were AI-generated. 'It's possible 10 or 15 years ago I would assume most of these were real before the internet was flooded with AI imagery and such.' This might indicate a bias specifically from more advanced and expert participants towards AI-generated autochromes. However, we could not verify a trend based on our data.

One participant identified as an expert in photography without any experience with autochromes rated Col-, Hist-, and Emo-related items rather high in both AI and original autochromes, but Tech- and Auth- related questions significantly higher for the original autochromes. This also supports the quantitative analysis of our user study. This participant stated that 'Most of the time, it felt like an Instagram filter to intentionally make them look old. The edges were too perfect and the star/pentagram shaped cut in the bottom left corner was annoying, as it was making it obvious that this is

processed/manipulated and not the real one scanned. I think the convincing ones should have more "damaged" edges - in terms of physical wearing, not in terms of grain and speckle.' This provide hints on how to improve our methods that should be addressed in future work. Finally, our study also had some limitations, particularly, when it comes to the inspection of visual quality and characteristics of images within an online survey, the display device that our participants used matters. We captured the type of device (smartphone, computer screen, tablet, or other) that was used to explore possible differences. We found that 54 participants used computer screens, 34 utilized a smartphone, and 2 specified their device as 'other'. We could not find a difference of answers between these conditions.

4.4. Summary of the Findings

Overall, our evaluation illustrates LoRA's performance through quantitative analysis, qualitative assessments, and user perceptions. The quantitative results indicated that configurations with higher LoRA ranks achieved efficient style adaptation in fewer training steps while maintaining comparable CLIPScores across models, suggesting robust performance. Qualitative assessments demonstrated the model's ability to generate diverse and plausible images, particularly in natural scenes, although it struggled with anthropomorphic subjects. User study results revealed that participants rated the color, emotion, and historical relatedness of AI-generated images as equivalent to original autochromes, while authenticity and technological aspects showed less equivalence.

5. Conclusion

Our approach demonstrates the use of diffusion-based models fine-tuned with Low-Rank Adaptation (LoRA) to generate autochromes from textual descriptions, capturing the essence of historical color photography. Since autochromes are rare and fragile, often not displayed publicly, our method allows for the creation of synthetic images that reflect the aesthetic qualities of original autochromes today. The features in these generated scenes depend on the underlying language model, leading to unique interpretations of the text. Despite some discrepancies, the color and aesthetics of the generated images are strikingly similar to traditional autochromes, offering a visual experience that honors the historical significance of this photographic technique.

User feedback indicates that participants perceive the color, emotion, and historical context of AI-generated autochromes as comparable to original images; however, authenticity and technological aspects require further refinement. Furthermore, our method can be utilized to create synthetic data and simulate defects in autochromes, thereby enhancing AI-based techniques for the restoration of digitized autochromes. We contribute to the community by publicly releasing both the fine-tuned model parameters and an interactive web-based demonstration, thereby enabling further research and application in digital heritage contexts.

References

[AKS24] ALLDIECK T., KOLOTOUROS N., SMINCHISESCU C.: Score distillation sampling with learned manifold corrective, 2024. URL: https://arxiv.org/abs/2401.05293, arXiv:2401.05293. 3

[CK23] CUI Y., KNOLL A.: Exploring the potential of channel interactions for image restoration. *Knowledge-Based Systems 282* (2023), 111156. URL: https://www.sciencedirect.com/science/article/pii/S0950705123009061, doi:https://doi.org/10.1016/j.knosys.2023.111156. 3

[CM23] CHERIBINI F., MUCO A.: *Color and Colorimetry. Multidisciplinary Contributions. Vol. XVIII A*. Research culture and science books. Gruppo del Colore - Associazione Italiana Colore, IT, Dec. 2023. URL: https://doi.org/10.23738/RCASB.009, doi:10.23738/RCASB.009. 3

[Doc] DOCKHORN T.:. URL: https://huggingface.co/black-forest-labs/FLUX.1-dev. 4

[EKB*24] ESSER P., KULAL S., BLATTMANN A., ENTEZARI R., MÜLLER J., SAINI H., LEVI Y., LORENZ D., SAUER A., BOESEL F., PODELL D., DOCKHORN T., ENGLISH Z., LACEY K., GOODWIN A., MAREK Y., ROMBACH R.: Scaling rectified flow transformers for high-resolution image synthesis, 2024. URL: https://arxiv.org/abs/2403.03206, arXiv:2403.03206. 3

[EYB*23] ESCHWEILER D., YILMAZ R., BAUMANN M., LAUBE I., ROY R., JOSE A., BRÜCKNER D., STEGMAIER J.: Denoising diffusion probabilistic models for generation of realistic fully-annotated microscopy image data sets, 2023. URL: https://arxiv.org/abs/2301.10227, arXiv:2301.10227. 2

[GLD*24] GUO H., LI Y., DAI T., XIA S.-T., BENINI L.: Intlora: Integral low-rank adaptation of quantized diffusion models, 2024. URL: https://arxiv.org/abs/2410.21759, arXiv:2410.21759. 3

[GPAM*14] GOODFELLOW I. J., POUGET-ABADIE J., MIRZA M., XU B., WARDE-FARLEY D., OZAIR S., COURVILLE A., BENGIO Y.: Generative adversarial networks, 2014. URL: https://arxiv.org/abs/1406.2661, arXiv:1406.2661. 2

[HHF*22] HESSEL J., HOLTZMAN A., FORBES M., BRAS R. L., CHOI Y.: Clipscore: A reference-free evaluation metric for image captioning, 2022. URL: https://arxiv.org/abs/2104.08718, arXiv:2104.08718. 5

[HJA20] HO J., JAIN A., ABBEEL P.: Denoising diffusion probabilistic models, 2020. URL: https://arxiv.org/abs/2006.11239, arXiv:2006.11239. 2

[JSL*25] JIN Y., SUN Z., LI N., XU K., XU K., JIANG H., ZHUANG N., HUANG Q., SONG Y., MU Y., LIN Z.: Pyramidal flow matching for efficient video generative modeling, 2025. URL: https://arxiv.org/abs/2410.05954, arXiv:2410.05954. 3

[JSW*22] J.HU E., SHEN Y., WALLIS P., ALLEN-ZHU Z., LI Y., WANG L., CHEN W.: Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations* (2022). URL: https://openreview.net/forum?id=nZeVKeeFYf9. 3

[KLA19] KARRAS T., LAINE S., AILA T.: A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019). 2

[KMT25] KORKMAZ C., MEHTA N., TIMOFTE R.: Adaptsr: Low-rank adaptation for efficient and scalable real-world super-resolution, 2025. URL: https://arxiv.org/abs/2503.07748, arXiv:2503.07748. 3

[Kop24] KOPPE J.: Restoration of old digitized autochrome images using deep learning techniques, 2024. URL: https://publica.fraunhofer.de/handle/publica/481389. 3

[KOZ80] KOZLOFF M.: Autochromes; the bouquet of lighted air. *Artforum* (1980). 7

[Lan23] LANGFORD C.: *Color Mania: Photographing the World in Autochrome*. National Geographic Books, 2023. 7

[LC17] LAKENS D., CALDWELL A.: Toster: Two one-sided tests (tost) equivalence testing. *R package version 0.2 5* (2017), 624. 7

[LCBH*23] LIPMAN Y., CHEN R. T. Q., BEN-HAMU H., NICKEL M., LE M.: Flow matching for generative modeling, 2023. URL: https://arxiv.org/abs/2210.02747, arXiv:2210.02747. 3

[LCW*21] LIN H., CHENG X., WU X., YANG F., SHEN D., WANG Z., SONG Q., YUAN W.: Cat: Cross attention in vision transformer, 2021. URL: https://arxiv.org/abs/2106.05786, arXiv:2106.05786. 2

[LG13] LAVÉDRINE B., GANDOLFO J.-P.: *The Lumière autochrome: history, technology, and preservation*. Getty Publications, 2013. 1, 7

[LK25] LEE J. S., KIM P. M.: Flowpacker: protein side-chain packing with torsional flow matching. *Bioinformatics 41*, 3 (01 2025), btaf010. URL: https://doi.org/10.1093/bioinformatics/btaf010, arXiv:https://academic.oup.com/bioinformatics/article-pdf/41/3/btaf010/61396781/btaf010.pdf, doi:10.1093/bioinformatics/btaf010.

[LMSX24] LIU Z., MA C., SHE W., XIE M.: Biomedical image segmentation using denoising diffusion probabilistic models: A comprehensive review and analysis. *Applied Sciences 14*, 2 (2024). URL: https://www.mdpi.com/2076-3417/14/2/632, doi:10.3390/app14020632. 2

[MO14] MIRZA M., OSINDERO S.: Conditional generative adversarial nets, 2014. URL: https://arxiv.org/abs/1411.1784, arXiv:1411.1784. 2

[NDR*22] NICHOL A., DHARIWAL P., RAMESH A., SHYAM P., MISHKIN P., MCGREW B., SUTSKEVER I., CHEN M.: Glide: Towards photorealistic image generation and editing with text-guided diffusion models, 2022. URL: https://arxiv.org/abs/2112.10741, arXiv:2112.10741. 3

[NM21] NGWENDUNA K. S., MBUVHA R.: Alleviating class imbalance in actuarial applications using generative adversarial networks. *Risks 9*, 3 (2021). URL: https://www.mdpi.com/2227-9091/9/3/49, doi:10.3390/risks9030049. 2

[PEL*23] PODELL D., ENGLISH Z., LACEY K., BLATTMANN A., DOCKHORN T., MÜLLER J., PENNA J., ROMBACH R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023. URL: https://arxiv.org/abs/2307.01952, arXiv:2307.01952. 3

[Pet19] PETERSON D.: Preserving the national geographic society's autochrome collection, 2019. 11.04.2025. URL: https://heritage-digitaltransitions.com/preserving-the-national-geographic-societys-autochrome-collection/. 3

[RBL*22] ROMBACH R., BLATTMANN A., LORENZ D., ESSER P., OMMER B.: High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2022). URL: https://arxiv.org/abs/2112.10752. 2

[RDN*22] RAMESH A., DHARIWAL P., NICHOL A., CHU C., CHEN M.: Hierarchical text-conditional image generation with clip latents, 2022. URL: https://arxiv.org/abs/2204.06125, arXiv:2204.06125. 3

[RKH*21] RADFORD A., KIM J. W., HALLACY C., RAMESH A., GOH G., AGARWAL S., SASTRY G., ASKELL A., MISHKIN P., CLARK J., KRUEGER G., SUTSKEVER I.: Learning transferable visual models from natural language supervision, 2021. URL: https://arxiv.org/abs/2103.00020, arXiv:2103.00020. 6

[RKW23] REITTER-KOLLMANN M., WEIDINGER A.: *Autochrome: The Fascination of Early Colour Photography*. teNeues Publishing Group, 2023. 7

[SBD*24] SAUER A., BOESEL F., DOCKHORN T., BLATTMANN A., ESSER P., ROMBACH R.: Fast high-resolution image synthesis with latent adversarial diffusion distillation. pp. 1–11. doi:10.1145/3680528.3687625. 3

[SC21] SHIM W., CHO M.: Circlegan: Generative adversarial learning across spherical circles, 2021. URL: https://arxiv.org/abs/2011.12486, arXiv:2011.12486. 2

[Sch87] SCHUIRMANN D. J.: A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics 15* (1987), 657–680. 7

[SCS*22] SAHARIA C., CHAN W., SAXENA S., LIT L., WHANG J., DENTON E., GHASEMIPOUR S. K. S., AYAN B. K., MAHDAVI S. S., GONTIJO-LOPES R., SALIMANS T., HO J., FLEET D. J., NOROUZI M.: Photorealistic text-to-image diffusion models with deep language understanding. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2022), NIPS '22, Curran Associates Inc.

[SHC*25] SINHA S. N., HORST R., CIORTAN I.-M., GRAF H., KUIJPER A., WEINMANN M.: Immersiveautochrome: Enhanced digitized mono-autochrome viewing through deep learning and virtual reality. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2025), IEEE. 3, 7

[SLBR23] SAUER A., LORENZ D., BLATTMANN A., ROMBACH R.: Adversarial diffusion distillation, 2023. URL: https://arxiv.org/abs/2311.17042, arXiv:2311.17042. 3

[Tay] TAYLOR H. A.: Challenge. URL: https://californiarevealed.org/search?search_api_fulltext=&f[0]=search_page_series_title:Harold%20Taylor%20Slide%20Collection. 2, 3, 6

[UA24] ULHAQ A., AKHTAR N.: Efficient diffusion models for vision: A survey, 2024. URL: https://arxiv.org/abs/2210.09292, arXiv:2210.09292. 2

[WHP24] WANG X., HE Z., PENG X.: Artificial-intelligence-generated content with diffusion models: A literature review. *Mathematics 12*, 7 (2024). URL: https://www.mdpi.com/2227-7390/12/7/977, doi:10.3390/math12070977. 2

[Wol22] WOLSKA A.: Jan zdzislaw wlodek's autochromes: Digitizing from an interdisciplinary perspective. *Muzealnictwo 63* (08 2022), 112–119. doi:10.5604/01.3001.0015.9745. 3

[YCZ*24] YANG M., CHEN J., ZHANG Y., LIU J., ZHANG J., MA Q., VERMA H., ZHANG Q., ZHOU M., KING I., YING R.: Low-rank adaptation for foundation models: A comprehensive review, 2024. URL: https://arxiv.org/abs/2501.00365, arXiv:2501.00365. 3

[YWP*24] YANG Y., WANG W., PENG L., SONG C., CHEN Y., LI H., YANG X., LU Q., CAI D., WU B., LIU W.: Lora-composer: Leveraging low-rank adaptation for multi-concept customization in training-free diffusion models, 2024. URL: https://arxiv.org/abs/2403.11627, arXiv:2403.11627. 3

[ZRA23] ZHANG L., RAO A., AGRAWALA M.: Adding conditional control to text-to-image diffusion models, 2023. URL: https://arxiv.org/abs/2302.05543, arXiv:2302.05543. 2