Shades of History: Reviving Nepal's Heritage through AI

Nitesh Rajbanshi[1] , Shlok Koirala[2], Yagya Raj Pandeya[*3]

[1] Department of Artificial Intelligence, Kathmandu University, Nepal
[2] Guru Technology Pvt. Ltd., Kathmandu, Nepal
[3] Artificial Intelligence and Smart System Research laboratory, Kathmandu University, Nepal
** Corresponding author: yagyapandeya@gmail.com*

Abstract

Historical photographs provide vital insights into Nepal's rich cultural heritage; however, most existing archival collections remain in black and white, limiting visual engagement and cultural comprehension for modern audiences. Despite advancements in AI-driven image colorization, current methods often suffer from inaccuracies in historical and cultural authenticity, highlighting a crucial research gap. This study addresses challenge by employing Conditional Generative Adversarial Networks (cGANs), leveraging a U-Net architecture with a pre-trained ResNet18 backbone. Initially trained using supervised L1 loss and subsequently refined through adversarial training, our method significantly enhances the visual authenticity and accuracy of colorized images. Quantitative assessments yielded a discriminator loss of 0.62 and generator loss of 4.42 for our best model with pretrained backbone. The resulting high-quality colorizations vividly depict historical narratives, greatly enriching the preservation and appreciation of Nepal's cultural heritage.

Keywords: *AI, Colorization, Nepal Heritage, Deep Learning*

Introduction

Historical images are an essential part of our cultural assets and provide us with glimpses into the past." There are dozens of black and white photos covering Nepal's fascinating history, offering glimpses of its culture, landscape and people. Yet the monochromatic quality of these images might limit their impact and reach, particularly among younger generations. One of the best ways to restore these ancient papers is by image colorization which adds colorful colors and a great appearance to the document.

The traditional hand-coloring processes are labor intensive and require high levels of expertise, which creates challenges to widespread preservation efforts for this content. State-of-the-art methods for image colorization, based on latest advances in artificial intelligence (AI), especially deep learning, have opened up new opportunities for fully automated colorization with high quality. When it comes to generating colored images that are natural looking and eye-catching, Generative Adversarial Networks (GANs) have been the leading approach.

AI-driven colorization has advanced, but there remain several impediments. Fine details and various textures that are observed in old images may be challenging to existing methods; therefore, a system needs to be established to capture cultural and historical context authentically. Further, there is little focused work on the applications of AI colorization in heritage protection and promotion practice in Nepal. It attempts to address these gaps by applying a GAN-based method with a pre-trained ResNet18 backbone for the colorization of historical images from Nepal in the hope that it would contribute to preserving and broadening the rich cultural heritage of the country.

Literature Review

Isola et al. (2017) explored conditional adversarial networks as a versatile approach for solving image-to-image translation tasks. Their framework not only taught effectively to map input images to desired output images but also autonomously acquired a loss function tailored for training this mapping. The release of their associated software, pix2pix, has further spurred widespread adoption and experimentation among numerous Twitter users, showcasing the system's impact on fostering artistic exploration in digital media.

Treneska et al. (2022) primarily focuses on leveraging generative adversarial networks (GANs) for image colorization for their capability to produce highly realistic colorized outputs. They propose employing conditional GANs (cGANs) specifically for this purpose and extend their findings to enhance performance in multilabel image classification and semantic segmentation tasks. Their empirical evaluations on the COCO and Pascal datasets reveal notable improvements, achieving a 5% increase in classification accuracy and a 2.5% enhancement in segmentation accuracy. These results underscore the effectiveness of image colorization with cGANs in enhancing downstream task performance without requiring additional manual annotations.

Materials and Methods

The project is centred on leveraging the power of Generative Adversarial Networks (GANs)to resurrect historical photographs.

Image Colorization

Image colorization is the process of converting black and white images into their colorful states. It is still an area of research. Much research and models have been developed but failed to conserve their color integrity. Most papers use luminance–chrominance color spaces that allow us to separate pixel intensity from the pixel color information (Treneska et al., 2022). We follow the same color space and Pix-to-Pix architecture defined by Isola et al. (2017). They use conditional Generative Adversarial Networks to train their image colorization model where the generator part consists of a UNET network, and the discriminator is PatchGAN.

Data Collection

We scrape the internet for our data collection. Sites like Pinterest, Twitter posts, Facebook and Instagram posts were some of the sites we used to scrape data. Our prime target was colorizing old and vintage Nepalese photos for citation photos from the Rana regime, Prithvi Narayan Shah's great conquest, or frosty and foggy photos in people's storerooms and boxes. These photos are unorganized assets, which were not available on the internet. However, some photographic sites played an instrumental role in finalising our dataset. We have listed the sites in the Bibliography and Dataset section. Among them, archivenepal was the most conducive resource in this project.

We created two datasets and used a standard celeb_A dataset. One of the prepared datasets contained images that purely represent Nepalese contemporary society along with cultural, historical and political assets. These datasets contained 8666 image samples. The second dataset was contaminated by samples of high-quality scenery images, making it a total of 15,433 sample datasets. The celeb_A dataset contained 202,599 image samples. Among them, we picked up 20,000 random samples for training.

Also, we have 1,515 black-and-white images collected from different sources as test data.

Data Source	Instances
archive Nepal, and Internet	8666
Scenery dataset	15433
Celeb_A	20000
Total	44099

Table 1: Data Information

Data Preprocessing

As per the data preprocessing step, first convert RGB colorspace into CIELAB (LAB) colorspace. The L component stands for perceptual lightness with a range [0, 100], meaning that it is the grayscale element. The A component represents the color position between red and green, while the B component represents the color position between blue and yellow; both have ranges [−128, 127].

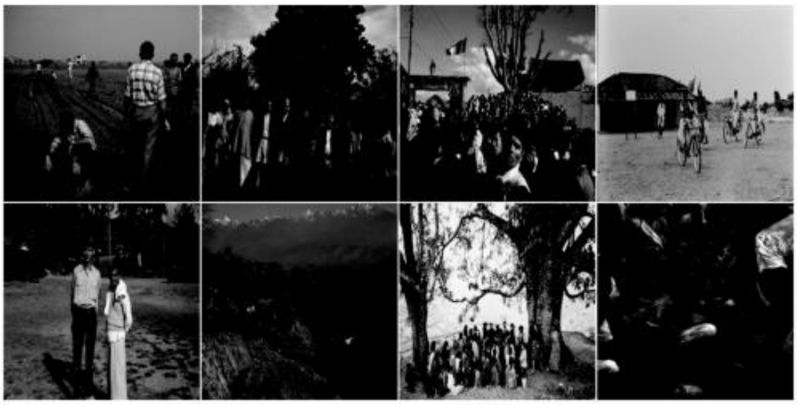


Figure 1: L channel visualization

The intuition is that a luminance–chrominance color space is needed to separate the intensity from the color information. The CIELAB (Lab) is one such color space used to describe all visible colors by the human eye. It was created to represent color changes in the same way as humans do. It means that a numeric change corresponds to a similar perceived difference in color. Space has little correlation between its three components. The L channel is used as an input to the model, while the 'A' and 'B' channels are the target values.

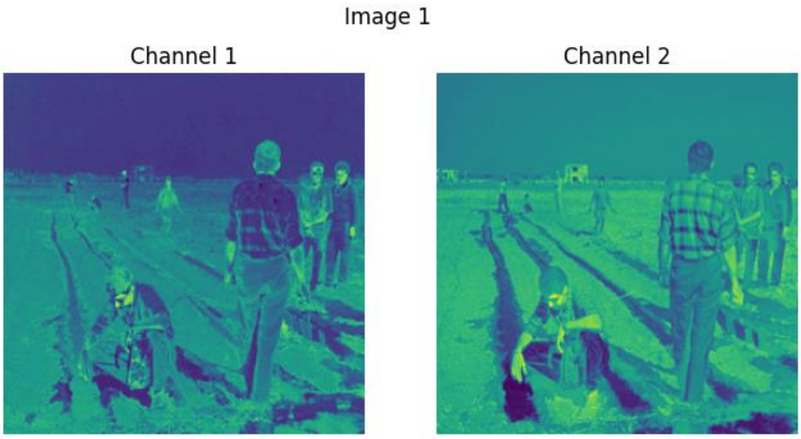
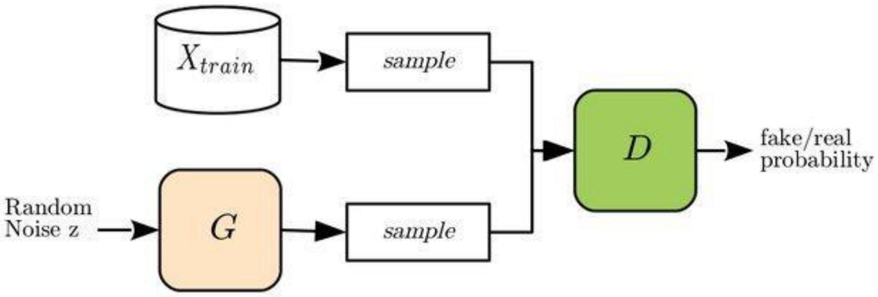


Figure 2: ab channel visualization



Generative Adversarial Network

Figure 3 Generative Adversarial Network

During training, the generator iteratively enhances the ability to create realistic outputs by attempting to deceive the discriminator. Simultaneously, the discriminator refines its capability to distinguish real from synthetic data. This adversarial process results in:

- A generator that effectively captures the underlying distribution of data to produce realistic outputs.
- A discriminator is proficient in identifying subtle discrepancies between authentic and generated data.

Conditional GANs extend this framework by conditioning the network on black-and-white images to generate colorized outputs, thus facilitating the restoration of historical photographs.

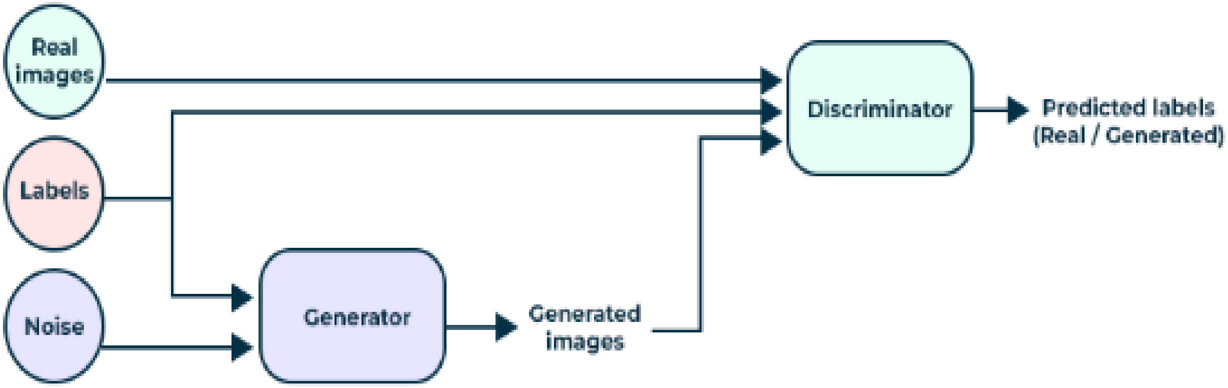


Figure 4 Conditional GAN

Objective Function

The objective functions used for training conditional generative adversarial networks are as follows:

$$V(G,D) = E_{x,y \sim p_{data}(x)}[log\ D\ (x,y)] + E_{y,z \sim p_z(z)}\left[log\ \left(1 - D\big(x, G(x,z)\big)\right)\right] \quad (1)$$

The generator G tries to minimize the objective function while discriminator D tries to maximize it, where x is the grayscale image input (L channel) and y is the output for color channels (ab). D (x, y) is the discriminator model that classifies the given x, y as true labels i.e., represents real data loss. G (x, z) is the conditional generator model that takes x (the L channel) as the input condition and z as a gaussian noise distribution resulting in prediction of later two channels. D (x, G (x, z)) now is the discriminator function to classify fake labels, i.e., fake data loss. To test the importance of conditioning the discriminator, we also compare an unconditional variant in which the discriminator does not observe x.

$$V(G,D) = E_{x,y \sim p_{data}(x)}[log\ D\ (y)] + E_{y,z \sim p_z(z)}\left[log\ \left(1 - D\big(G(x,z)\big)\right)\right] \quad (2)$$

Another loss function is required to tune the generator only. For this we explore both L1 and L2 loss i.e., pixel-wise loss calculation. At the end, we go with the L1 loss function as it encourages less blurring.

$$L(G) = E_{x,y,z}[||y - G(x,z)||] \quad (3)$$

So, our final objective for the generator is:

$$G * = arg\, min(G)\, max(D)\, V\,(G, D)\ +\ \lambda\, L(G) \tag{4}$$

This enables the Generator an extra edge to trick the Discriminator whereas the Discriminator's tasks remain unchanged.

UNET

The Pix2Pix architecture employs UNET architecture for the generator and PatchGan for the Discriminator section.U-Net architecture allows low-level information to shortcut across the network, i.e. UNET progressively downsamples the image, until a bottleneck, after which the process is reversed, and the image is upsampled to its original size. Skip connections are also added to facilitate the flow of low-level information through the network.

Generator Architecture:

- A trained ResNet18 model serves as the backbone for the generator's down-sampling path. It is extended by the U-Net architecture.
- We use Resnet for utilizing ResNet18's feature extraction capabilities and tailoring them to picture colorization requirements, where the objective is to anticipate color (ab channels) based on grayscale input (L channel).

Pretraining Method:

- Use the Weight for high-level feature extraction, learned from ImageNet
- Then, the upsampling (decoder) layers are trained using L1 loss in a supervised manner.

PatchGAN (Markovian Discriminator)

PatchGAN motivates GAN to only model high-frequency structures. It tries to classify if each Nunn patch of the image is true or fake. It is run convolutionally across all the samples, averaging the values to return the final input.

Deeplabv3 Generator

Developed for segmentation tasks, we utilized this model to explore its generational power. The DeeLapv3 model uses Atrous Convolution Layers and Atrous Spatial Pyramid Pooling Layers to conserve local information of the image. We hypothesize that preserving local features of the image may lead to the color conservation of the original image. As in the UNET + PatchGAN model, the results are promising but fail to conserve the original color.

Results

In this study, we comprehensively explored the application of Conditional Generative Adversarial Networks (cGANs) for colorizing historical photographs. The experiments utilized our custom-prepared datasets, specifically representing Nepalese heritage imagery, along with the standard CelebA dataset for Qualitative Visual Comparison

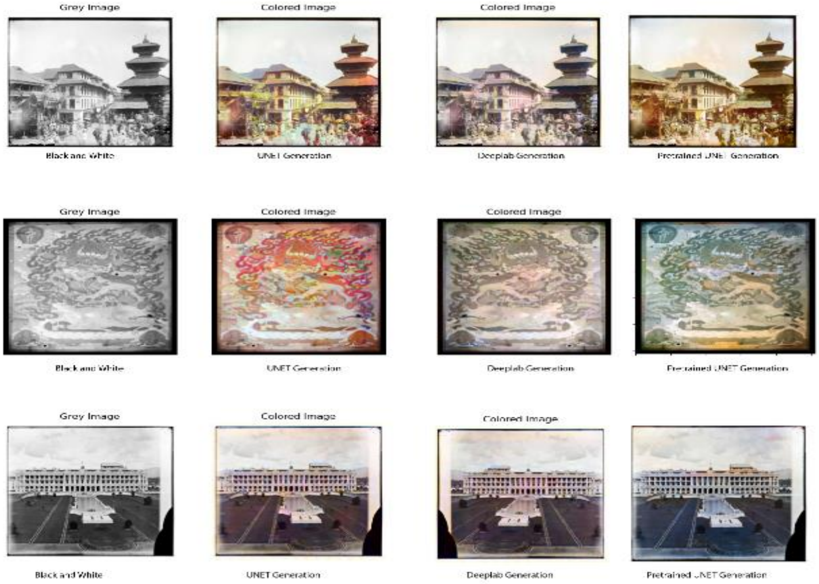


Figure 12: Result comparison of three different models. Black And White (left), UNET scratch (middle, left), Deeplab Generator (middle, right), and UNET Pretrained (right)

In a qualitative evaluation, the Deeplabv3 model erred in creating deficient natural color tones as well as historical authenticity (see Figure 12). The UNET trained from scratch exhibited significant enhancements appealing to the eye and attaining color harmony though it had hiccups with some memory images which are more complex, with human faces and fine details. UNET+PatchGAN with a pre-trained ResNet18 backbone showed great benefits; it consistently produced visually realistic images with culturally appropriate colors and, of course, fine historical details preserved (see Figure 12). This underlines the need for utilizing pre-trained features to create images with adequate accuracy in the task of colorizing image

Quantitative Evaluation

Table 2:

Model	Discriminator Loss	Generator Loss	PSNR (dB)
UNET+PatchGAN(Scratch)	0.3402	6.5659	19.234
Deeplabv3	0.0030	24.537	
UNET+PatchGAN (Pre-trained ResNet18)	0.619	4.4222	17.14
	2		2
			26.013

Discriminator Loss, Generator Loss, and PSNR scores for evaluated models

The results clearly indicate that the pre-trained UNET+PatchGAN model vastly outperforms Deeplabv3. Even though it returns slightly higher discriminator loss than Deeplabv3, the much lower generator loss and especially higher PSNR (26.013 dB) prove that pre-trained UNET yields colorized images with far superior visual quality and lower noise but preserving important details in images.

Discussion

Our study aimed to restore and enhance historical photographs of Nepal through AI to colorize and rehabilitate historic black-and-white photos of Nepal using Conditional Generative Adversarial Networks (cGANs) coupled with a pre-trained ResNet18 backbone. Quantitative metrics including PSNR and discriminator-generator losses confirmed that using a pre-trained UNET+PatchGAN performed significantly better than alternative models. Qualitative assessments based on participant surveys further validated the model's ability to produce colorizations that were verifiably historically accurate and visually appealing.

While we were able to achieve promising results with our approach, we also encountered some limitations. The model worked great with landscape and scenery images but struggled to accurately colorize human portraits or anything with detailed finer details. This problem is in line with previous research noting the difficulty in correctly capturing realistic human skin hues and refined textures. Future iterations also might use semantic segmentation, perhaps using techniques like Deeplabv3 (Chen et al., 2018) to better handle object level color accuracy and historic authenticity.

Additionally, GANs remain challenging to train (based on varying discriminator/generator losses across models). Methods introduced by Salimans et al. (2016) and Heusel et al. (2017) feature matching and two-time-scale update rules, would serve to alleviate training instabilities, allowing the model to converge more reliably.

Our qualitative user study indicated strong preference for outputs generated by the pre-trained UNET approach— essentially, the very same result we were to further deliver, suggesting the very effectiveness of transfer learning in preserving cultural authenticity. Future studies need to expand datasets, setting the accent particularly upon more wide-ranging historical contexts and more wide-ranging types of subjects. Collaborating with historians or cultural experts to validate historical accuracy will further enhance, and multiply by the factor of, reliability and credibility of AI-based colorization approaches.

Conclusion

In this study, we have proven that deep learning, more specifically Conditional GAN implemented using pre-trained UNET architecture, has greatly improved the accuracy and realism of historical photograph colorization. In this variation nuances are captured ideally, therefore, increasing accessibility and understanding of the cultural heritage of Nepal among an audience of the current millennium. Great challenges remain to be met concerning human portraits and fine-detail colorization.

Besides semantic segmentation, adaptive color mappings, and the availability of larger and more varied datasets, further studies involving collaboration with experts in history and culture are encouraged. Working on these aspects will benefit in strengthening the potential in artificial intelligence-based historical photo restoration thus becoming a valuable element in preserving heritage.

References

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/tpami.2017.2699184

Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of Plant-Leaf diseases using CNN and Transfer-Learning approach. *Electronics*, *10*(12), 1388 https://doi.org/10.3390/electronics10121388

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6626–6637). Curran Associates. https://arxiv.org/pdf/1706.08500

Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. *CVPR*, 5967–5976. https://doi.org/10.1109/cvpr.2017.632

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. In *Advances in Neural Information Processing Systems* (Vol. 29, pp. 2234–2242). Curran Associates. https://arxiv.org/pdf/1606.03498

Treneska, S., Zdravevski, E., Pires, I. M., Lameski, P., & Gievska, S. (2022). GAN-Based Image colorization for Self-Supervised Visual Feature Learning. *Sensors*, *22*(4), 1599 https://doi.org/10.3390/s22041599

Datasets

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). *CelebA dataset* [Dataset]. Kaggle. https://www.kaggle.com/datasets/zuozhaorui/celeba?select=img_align_celeba

Arnaud58. (2020). *Scenery dataset* [Dataset]. Kaggle. https://www.kaggle.com/datasets/arnaud58/landscape-pictures

Online Image Sources

Archive Nepal. (2024). *Historical photographs of Nepal*. https://www.archivenepal.org/home

Getty Images. (2024). *Nepal old men stock photos*. https://www.gettyimages.com/photos/nepal-old-men

Nepali Times. (2024). *Making Nepal's history* colorful. https://nepalitimes.com/here-now/making-nepal-s-history-colorful

Pinterest. (2024). *Vintage Nepal photographs*. https://www.pinterest.com/subenjabegu/vintage-nepal/

Depositphotos. (2024). *Old Nepal stock photos*. https://depositphotos.com/photos/old-nepal.html

Alamy. (2024). *Vintage Nepal photographs*. https://www.alamy.com/stock-photo/vintage-nepal.html?sortBy=relevant

iStockphoto. (2024). *Panoramic view of Kathmandu, Nepal*. https://www.istockphoto.com/photos/panoramic-view-of-kathmandu-nepal

Nepal 8th Wonder. (2024). *Old photos of Nepal*. https://nepal8thwonder.com/old-photos-of-nepal