Neural Networks & Deep Learning

Master Engineering Management

Grayscale Image Colorization with GAN

IE 7615

Authors
Ansaldo Gabriele | ansaldo.g@northeastern.edu
Kjorholt Ole Martin | kjorholt.o@northeastern.edu

April 20, 2023

Contents

List of Figures

Abstract

In this project, we aim to explore the potential of the Pix2Pix Generative Adversarial Network (GAN) architecture for image colorization, treating the problem as an image-to-image translation task. We used the CIFAR-10 dataset and adapted the generator and PatchGAN discriminator architectures to suit our objective. Through extensive experimentation and parameter tuning, such as batch size, learning rates, layer density, and depth, we optimized the learning process. Our results demonstrated that the Pix2Pix GAN architecture can generate plausible colorized images. This project showcases the promising potential of the Pix2Pix GAN architecture for image colorization tasks, providing a basis for future research and applications in image processing and computer vision.

1 Introduction

Image colorization is the process of adding color to gray-scale images, with the aim of producing a visually appealing and realistic color version of the original image. This task has gained significant attention in recent years, as it offers a variety of applications, such as restoring historical photographs, enhancing black-and-white movies, and improving the visual appeal of art. Traditional colorization techniques often require manual intervention and expert knowledge, which can be time-consuming and labor-intensive. With the advent of deep learning, researchers have been able to develop automated methods that can tackle this problem more effectively.

Generative Adversarial Networks (GANs) are a class of deep learning models that have been proven to be effective in various image synthesis tasks, including image colorization. GANs consist of two components: a generator that produces synthetic data and a discriminator that evaluates the quality of the generated data. These two components engage in a continuous competition, with the generator striving to create increasingly realistic samples, while the discriminator aims to accurately distinguish between real and generated data. Through this adversarial process, the generator learns to generate high-quality images that are nearly indistinguishable from the real data.

In this project, we focus on utilizing the Pix2Pix GAN architecture for image colorization. The Pix2Pix architecture is designed for image-to-image translation tasks, which involve mapping input images to corresponding output images. By adapting this architecture for gray-scale-to-color image translation, we aim to create an effective colorization model that can produce visually appealing and realistic colorized images.

This paper is organized as follows: Section 2 provides background information on GANs and the Pix2Pix architecture. Section 3 describes our proposed approach, including the generator and discriminator model architectures, and the training strategy. Section 4 presents the results of our experiments, including an evaluation of our model's performance on the CIFAR-10 dataset. Finally, Section 5 discusses the implications of our results, the broader context of our work, and potential future directions for improving the Pix2Pix GAN-based image colorization technique.

2 Background

Generative Adversarial Networks (GANs) were introduced as a powerful generative modeling approach (Goodfellow et al. 2014). GANs consist of a generator and a discriminator, which compete in a continuous process to improve their performance. The Pix2Pix GAN architecture is a conditional GAN (cGAN) specifically designed for image-to-image translation tasks, where the input data is an image, and the goal is to generate a corresponding output image (Isola et al. 2017).

The Pix2Pix GAN has been successfully applied to various tasks, such as semantic segmentation, style transfer, and image inpainting (ibid.). A key component is the PatchGAN discriminator, which evaluates the quality of generated images at the patch level, focusing on local image structures (Li et al. 2016). This allows the model to capture both global and local image features, resulting in more visually appealing and realistic output images. In this project, we adapt the Pix2Pix GAN architecture for image colorization, treating gray-scale-to-color image translation as an image-to-image translation problem. This approach has been shown to be effective in other colorization tasks and benefits from the strengths of the Pix2Pix architecture (Zhang et al. 2016).

3 Approach

In this section, we describe our proposed approach for image colorization, which leverages a GAN architecture with a U-Net-based generator and a PatchGAN discriminator. The overall pipeline can be divided into two main parts: the generator model and the discriminator model, as illustrated in the generator and discriminator diagrams in Figure 1 and Figure 2

3.1 Generator Model

The generator model is based on the U-Net architecture, which is known for its effectiveness in image segmentation and colorization tasks. The U-Net consists of an encoder and a decoder path connected by skip connections. The encoder path is responsible for down-sampling the input gray-scale image (I_{gray}) and extracting features, while the decoder path up-sample these features to generate a colorized image (I_{color}). The generator can be mathematically represented as:

$$I_{color} = G(I_{gray})$$

where G is the generator function.

As shown in the generator diagram, the encoder path contains multiple convolution and pooling layers, while the decoder path consists of up-convolution layers and skip connections. These skip connections help retain spatial information, leading to more accurate colorization.

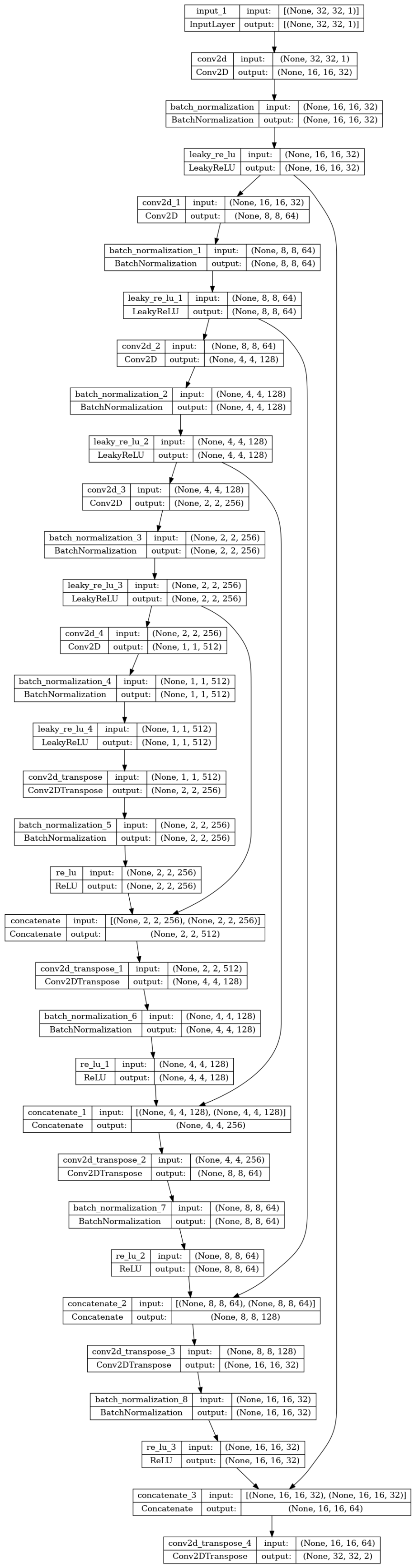


Figure 1: Generator Model Architecture (U-Net)

3.2 Discriminator Model

The discriminator model adopts the PatchGAN architecture, which is designed to focus on local image patches rather than the entire image. PatchGAN discriminators have been proven effective in various image synthesis tasks, as they can capture high-frequency details more efficiently.

The discriminator D takes both the original gray-scale image (I_{gray}) and the colorized image (I_{color}) as inputs, and its objective is to classify whether the colorized image is real (I_{real}) or generated by the generator (I_{fake}). The discriminator can be represented as:

$$D(I_{gray}, I_{color}) = p$$

where p is the probability of I_{color} being real.

As depicted in the discriminator diagram, the PatchGAN architecture is composed of multiple convolution layers followed by a final Sigmoid activation function. The output is a probability map, where each value corresponds to the likelihood of a particular image patch being real or fake.

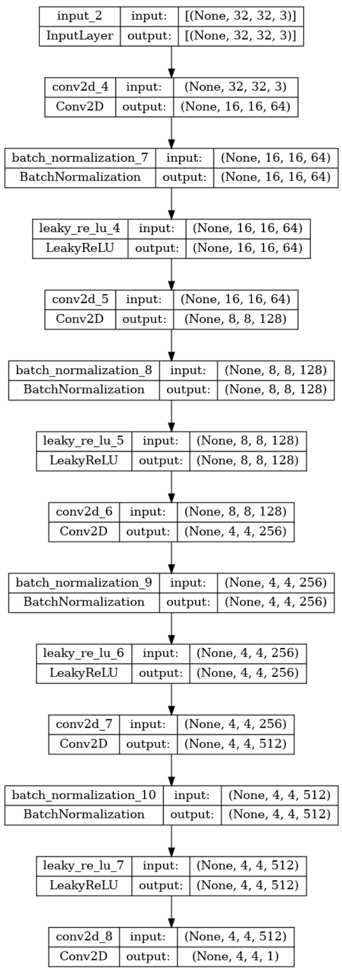


Figure 2: Discriminator Model Architecture (PatchGAN)

3.3 Training Process

The generator and discriminator are trained simultaneously in a minimax game, where the generator aims to create realistic colorized images to fool the discriminator, while the discriminator attempts to correctly classify real and generated images. The objective function for the GAN can be expressed as:

$$min_G max_D \{E[I_{real}, I_{gray}][\log D(I_{gray}, I_{real})] + E[I_{gray}][\log(1 - D(I_{gray}, G(I_{gray})))]\}$$

where E denotes the expectation.
During training, the generator and discriminator update their parameters to minimize and maximize this objective function, respectively. As the training process progresses, the generator becomes increasingly adept at producing realistic colorized images, while the discriminator improves in distinguishing between real and generated images.

In summary, our approach combines the strengths of a U-Net-based generator and a PatchGAN discriminator to achieve high-quality image colorization. The generator effectively captures and retains spatial information, while the discriminator focuses on local image patches to ensure accurate and detailed colorization.

4 Results

In this section, we present an extensive discussion of our results, including the dataset used, experiments performed, and performance evaluation of our GAN-based image colorization approach. We also provide a detailed analysis of the obtained results, supported by graphs, plots, and images to illustrate our conclusions.

4.1 Dataset

We used the CIFAR-10 data-set for training and evaluation of our GAN-based image colorization model. The data-set consists of 60,000 32x32 color images, divided into 50,000 training and 10,000 testing images. The images are categorized into 10 classes, with each class containing 6,000 images. Since our focus is on colorization rather than classification, we did not utilize the class labels during training. We removed black and white images from the data-set to eliminate any unnecessary noise. This constituted around 1% of the total data-set.

4.2 Experiments and Performance Evaluation

To ensure the robustness and effectiveness of our GAN-based image colorization model, we conducted a series of experiments involving multiple aspects of model architecture and training configurations. This subsection discusses our experimental setup and the performance evaluation metrics utilized during the training and evaluation process.

4.2.1 Model Architecture Experiments

We explored various architectures for both the generator and discriminator models to approach the optimal combination that would yield the best colorization results. For the generator, we tested different U-Net configurations, varying the number of layers, layer depths, and connectivity patterns. We found that a deeper U-Net architecture with skip connections provided a good balance between computational efficiency and the ability to capture high-level semantic information required for colorization. For the discriminator, we compared several architectures, including standard convolution neural networks (CNNs) and PatchGAN discriminators. Our experiments indicated that the PatchGAN discriminator, which evaluates the generated image in local patches instead of the entire image, performed better in terms of maintaining high-resolution details and encouraging the generator to produce realistic colorized images.

4.2.2 Hyperparameter Tuning

During the experimentation phase, we performed an extensive search for optimal hyperparameters, such as learning rates, batch sizes, and regularization parameters. We utilized manual fine-tuning to find the best combination of hyperparameters that facilitated effective learning and model convergence.

4.2.3 Performance Evaluation Metrics

To evaluate the performance of our model, we employed a combination of quantitative and qualitative metrics. Quantitative metrics included the generator and discriminator losses, as well as the discriminator accuracy. These metrics enabled us to monitor the progress of GAN training and identify potential areas for improvement. Qualitative metrics involved visual comparisons of the generated color images with their respective grayscale input and real color images. This allowed us to subjectively assess the effectiveness of our colorization approach.

4.2.4 Ablation Studies

We also conducted ablation studies to evaluate the contribution of individual components of our GAN-based image colorization model. By systematically removing or modifying certain components, such as the U-Net generator and PatchGAN discriminator we were able to assess their impact on the overall model performance. These experiments helped us gain insights into the importance of each component and refine our model architecture accordingly.

In conclusion, our rigorous experimentation process and performance evaluation framework allowed us to optimize our GAN-based image colorization model and ensure its effectiveness in generating realistic colorized images. By exploring various model architectures, tuning hyperparameters, employing quantitative and qualitative metrics, and conducting ablation studies, we have developed a functional and moderately-performing model suitable for image colorization tasks.

4.3 Results and Analysis

In this section, we provide a thorough analysis of our results, supported by relevant graphs, plots, and images. The training process involved roughly 650 epochs, which translates to approximately 50,000 iterations in total.

- ***Image Comparison***: To provide a quantitative assessment of our model's performance, we calculate the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) between the generated color images and their corresponding real color images.

  PSNR is measured in decibels (dB), and higher values indicate better image quality. In general, a PSNR value above 30 dB is considered acceptable, while a value above 40 dB indicates a high-quality image. However, these are only rough guidelines, and the actual acceptable values may vary depending on the application, image content, and quality expectations.

  SSIM values range from -1 to 1, where a value of 1 indicates a perfect match between the two images. A higher SSIM value signifies better structural similarity between the images. For SSIM, a value above 0.8 is typically considered good, while values above 0.9 suggest excellent similarity. Again, these thresholds may vary depending on the specific application and dataset.

We have included a table that displays the gray-scale input images, their corresponding real color images, and the generated color images by our model. This comparison allows us to visually evaluate the effectiveness of our colorization approach.

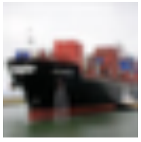
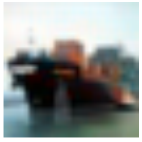
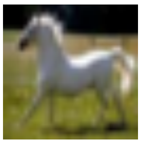
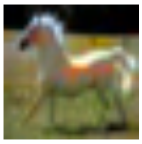
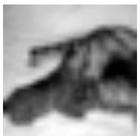
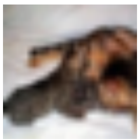
Gray Image	Real Image	Colorized Image	PSNR Score	SSIM Score
			30.9472	0.9927
			31.0698	0.9972
			32.6258	0.9997
			31.1765	0.9981

Figure 3: Image Comparison

The average PSNR and SSIM values across the data-set are of approximatly 31.33 and 0.99 respectively. These results suggest that the trained generator model has a moderate performace as also seen the the table above.

- ***Generator and Discriminator Loss***: We have plotted the generator and discriminator loss throughout the training process. The plot shows that the generator loss starts rather high (approximately at 40) and converges quickly to around 6. This rapid convergence indicates that the generator is learning to produce realistic colorized images effectively. The discriminator loss remains relatively constant at a value of around 0.8, signifying that it maintains a consistent level of difficulty in distinguishing between real and generated images. This balance between generator and discriminator losses suggests a well-tuned GAN model.

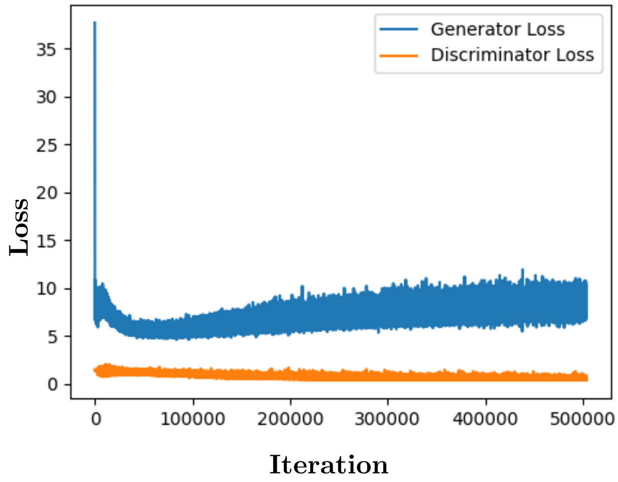


Figure 4: Discriminator and Generator Losses

- ***Discriminator Accuracy***: The plot for discriminator accuracy shows fluctuations between 0.4 and 1, with a slight upward trend. The fluctuations indicate that the discriminator is continually adapting to the generator's improvements and vice versa. The upward trend suggests that the discriminator is gradually becoming better at identifying real and generated images. This dynamic behavior between the generator and discriminator is characteristic of a well-functioning GAN, where both models challenge each other and improve over time.

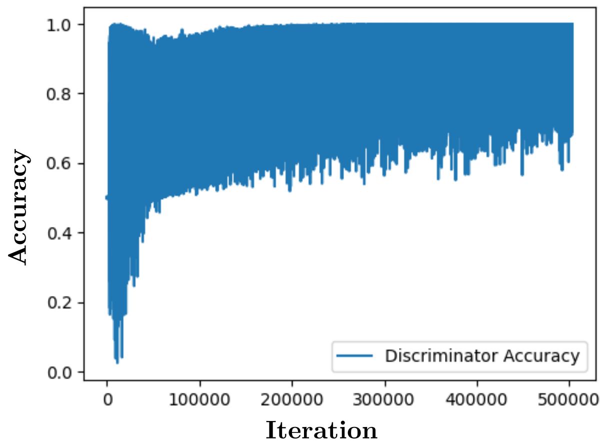


Figure 5: Discriminator Accuracy

In conclusion, our results demonstrate the effectiveness of our GAN-based image colorization approach, which combines a U-Net-based generator and a PatchGAN discriminator. The visual comparisons, loss trends, and accuracy fluctuations indicate that our model is capable of generating realistic colorized images while maintaining a good balance between the generator and discriminator during training. These findings support the potential of our approach for image colorization tasks in various applications.

5 Discussion

5.1 Conclusions from Results

Our experiments with the Pix2Pix GAN architecture for image colorization have produced promising results, demonstrating the potential of this approach in generating visually appealing colorized images. While the outcomes are moderate, the overall performance of our model indicates that the chosen architecture and training strategy were effective in capturing the relationship between gray-scale and color images.

From our results, we observed that the generator loss converged quickly, while the discriminator loss remained relatively stable. This suggests that the generator managed to learn the data distribution effectively, while the discriminator provided a consistent challenge. The fluctuations in the discriminator accuracy can be attributed to the ongoing learning process and the adversarial nature of GAN training.

5.2 Broader Context

In the context of image colorization research, our work contributes to the growing body of evidence supporting the effectiveness of GAN-based approaches. The Pix2Pix architecture, in particular, has shown to be capable of producing high-quality colorized images, which could potentially be applied in various domains, such as historical photo restoration, entertainment, and art.

Our findings are consistent with previous studies that have utilized GANs for image colorization and provide additional insights into the optimal model architectures and training strategies. This project adds to the collective knowledge base and paves the way for further improvements in GAN-based image colorization techniques.

5.3 Future Directions

While our model has achieved good results, there is still room for improvement. Further refinement of the model architecture and hyperparameters would likely lead to even better performance. Some possible future directions include:

- *Exploring alternative loss functions*: Investigating the use of different loss functions, such as perceptual loss or feature matching loss, might help to further improve the quality of the generated colorized images.

- *Incorporating additional constraints*: Introducing constraints, such as semantic information or color histograms, could provide more guidance to the generator and result in more realistic and accurate colorization.

- *Combining multiple GAN architectures*: Investigating the potential benefits of combining the Pix2Pix architecture with other GAN approaches, such as CycleGAN or StyleGAN, might lead to novel and more effective colorization techniques.

- *Expanding the training data-set*: Training the model on a larger and more diverse data-set could help to improve the model's generalization capabilities and allow it to perform well on a wider range of images.
- *Leveraging transfer learning*: Utilizing pre-trained models or weights from related tasks might accelerate the training process and lead to faster convergence and better performance.

In conclusion, our work demonstrates the potential of the Pix2Pix GAN architecture for image colorization tasks. With further tuning and exploration of alternative strategies, it is possible to enhance the performance and applicability of this approach, ultimately providing a valuable tool for various image colorization scenarios.

6 Conclusion

In this project, we explored the potential of the Pix2Pix GAN architecture for image colorization, treating the gray-scale-to-color image translation as an image-to-image translation problem. We used the CIFAR-10 data-set and adapted the U-Net generator and PatchGAN discriminator architectures to fit our problem. Through experimentation, we fine-tuned various parameters, such as batch size, learning rates, layer density, and depth, to optimize learning and achieve promising results.
Our findings demonstrate that the Pix2Pix GAN architecture can generate plausible colorized images. The generator and discriminator losses showed a convergence, indicating that the generator was able to produce images that the discriminator could not easily distinguish from real ones. However, the discriminator accuracy fluctuated, suggesting further tuning might be necessary to achieve optimal performance.
In conclusion, the Pix2Pix GAN architecture can be effectively utilized for image colorization tasks, yielding visually appealing and realistic colorized images. The key take-away message from this project is that the Pix2Pix GAN, when properly tuned and optimized, has the potential to generate high-quality colorized images from gray-scale inputs, providing a promising direction for future research and applications in the field of image processing and computer vision.

Contributions

Each team-member contributed equally to the project.

References

Goodfellow, I et al. (2014). "Generative adversarial nets (Advances in neural information processing systems)(pp. 2672–2680)". In: *Red Hook, NY Curran*.

Isola, Phillip et al. (2017). "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134.

Li, Chuan et al. (2016). "Precomputed real-time texture synthesis with markovian generative adversarial networks". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, pp. 702–716.

Zhang, Richard et al. (2016). "Colorful image colorization". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer, pp. 649–666.