COLOURIZING HISTORIC ARCHAEOLOGICAL SITES USING GENERATIVE ADVERSARIAL NETWORKS

Anusha Fatima Alam
Student# 1009056539
anusha.alam@mail.utoronto.ca

Marwan Ismail
Student# 1008773506
marwanmohamed.ismail@mail.utoronto.ca

Mateusz Kazimierczak
Student# 1008886421
mateusz.kazimierczak@mail.utoronto.ca

Sebastian Pigeon
Student# 1008801101
sebastian.pigeon@mail.utoronto.ca

ABSTRACT

The purpose of this report is to detail the final results of the historic archaeological site image colourization project. Archaeological images from the "a_amphitheatre" categories of the MIT Places 365 Dataset (Zhou et al., a) are used to train an adversarial GAN model to infer the A and B channels of black and white images in the CIELAB colour space, which are then converted to the RGB colour space. The performance of the GAN is compared to the baseline model. Important concerns consider cultural bias, ethical dilemmas, and ownership. All code, data, models, and loss graphs can be found in Google Drive.

—-Total Pages: 9

1 INTRODUCTION

Ever since coloured photography was introduced in 1890's, it has become widely popular for its ability to capture the richness and vibrancy of the world in full spectrum (Horiuchi). From fostering artistic expression to providing enhanced visualization, image colourization plays a critical role in preserving, understanding and appreciating cultural heritage and collective history (Horiuchi). In an archaeological context, historical black and white images fail to capture the visual richness, texture and diversity of artifacts and structures. As such, colourization of these images, like the one displayed in Figure 1, seems to offer a promising solution to restore this contextual information. The objective of this project is to use generative adversarial networks (GANs) to colourize black and white images of historic archaeological sites. At the moment, the most reliable method of

Figure 1: Image of the Colosseum in Rome, Italy (Amp). This is an example of the archaeological images for which colourization with GANs will be applied.

image colourization is extensive hand labour and thorough research of the true historical colours of the clothes, scenery, and time of day, among other factors. Complicated statistical models, such as that from Horiuchi, are not completely autonomous, as they must still be initialized with appropriate pixel values. Given Deep Learning techniques like GANs have previously proven successful in generating images from noise (Goodfellow et al.), it would be a reasonable approach to use Deep Learning in generating a coloured image of archaeological images given a black and white input.

2 ILLUSTRATION

This project uses GANs to interpolate the true CIELAB colour of a black and white archaeological image, hence increasing the time efficiency and accuracy of current colourization techniques. Figure 2 shows how the model is being trained while Figure 3 shows how the model is being used for inference. The novelty of our approach consists of a highly tailored set of training data, as well as a new context for the use of the model which provides strong advantages for educational and visualization purposes in the domain of ancient architecture.

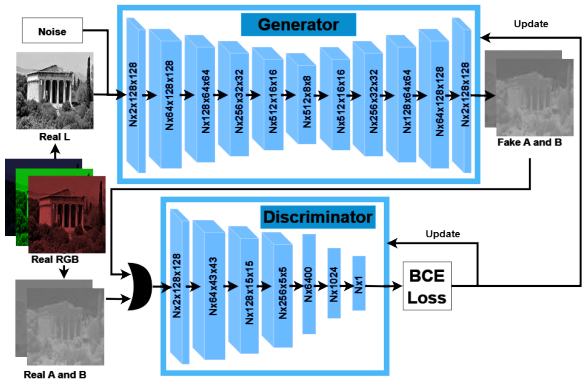


Figure 2: GAN training model of image colourization. Each block represents the layer's size. Each arrow represents the operation (convolution, flattening, fully connected, activation, batch normalization, among others, as described later in this report). First, a real RGB image from the data set is converted to LAB and split into its components. The L component is passed into the generator along with noise to ensure random variability in the results, where it creates fake A and B components of the image. Both the real and fake AB components are passed into the discriminator one at a time (hence the OR gate), which must decide if the image is real or fake. Based on its prediction, the generator and the discriminator are updated appropriately, so the generator can create more realistic fake images and the discriminator can better discern between the real and fake images.

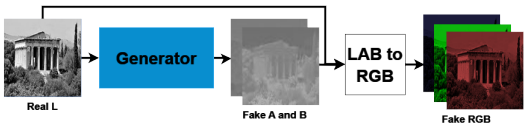


Figure 3: How to use the fully-trained GAN. The L (greyscale) channel is fed into the generator, where it will generate the AB components. The real L and fake AB components are concatenated and translated to RGB.

3 BACKGROUND AND RELATED WORK

Since the early 2000s, there has been great interest in automated image colourization. The earliest methods involve purely traditional and statistical methods (Horiuchi), but do not show great accuracy and feasibility.

3.1 DEEP LEARNING FOR IMAGE COLOURIZATION

The first Deep Learning Image colourization (DLIC) algorithm was proposed by Cheng et al.. The main limitation of this algorithm is the use of a traditional Artificial Neural Network (ANN) architecture instead of using convolutional layers which are much more powerful in image processing. With an evolving Deep Learning knowledge base and community, there are a number of techniques and algorithms that are highly effective for this task, such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Capsule Neural Networks (CapsNet), and Transformer Neural Networks (Huang et al.). Generative Adversarial Networks are used for this project.

3.2 Image Representations

Images can be represented in many ways, including, but not limited to:

1. **Black and White (Greyscale)**: images are representing with a single integer value for each pixel. This value represents the brightness of the pixel between black and white.(Gra)
2. **RGB Colour Space**: The most popular colour representation technique defined by the red, green, and blue components of an image. However, it is not ideal for Deep Learning colourization models, as it requires the prediction of unnecessary information. Zhang et al. is an example of the use of the RGB colour space for image colourization.
3. **CIELAB**: A more efficient approach is to predict only 2 channels and use the given greyscale image to reconstruct the final image (Cao et al.). CIELAB consists of 3 channels: the brightness of the pixel L (obtained from the greyscale value), the ratio between green-magenta A, and ratio between blue-yellow B.
4. **Others**: Literature (Kanan & Cottrell) has shown that various methods of converting to the greyscale colour space may affect performance of machine learning models. Examples include YUV and HSV, the use of which has shown to improve the reliability of the model, and produce images with more saturated colours (Hensman & Aizawa; Zhou et al., b).

4 Data Processing

The full data processing flowchart is shown in Figure 4.

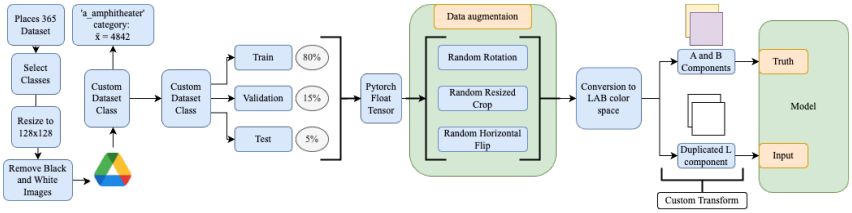


Figure 4: Flowchart depicting how the images were processed from when they are downloaded from the Places 365 dataset to when they are used to train the model.

4.1 Categories and Cleaning

The ('r_amphitheatre') category from the Places 365 dataset (Zhou et al., a) is selected because it is most relevant to ancient structures. The images are downloaded and resized from 256x256 to 128x128, with already black and white images removed from the dataset. The selected sample contains 4842 images, comprising of 3874 training images, 726 validation, and 242 testing images. This reduced dataset prevents unnecessary computational power consumption.

The final images are compressed and stored in Google Drive for easy retrieval later. A custom Dataset class is created to aid with image querying and performing data augmentation in the specific formal preparation method discussed in the next sections.

4.2 Data Augmentation

Data augmentation is incorporated to increase model robustness by introducing regularization:

1. **Random rotation**: The images are slightly rotated from the centered position. This is appropriate because ancient pictures are not always upright.
2. **Random resize crop:** A small outer section of the image is cropped out, and the rest of the image is resized to the dimensions of the original image. This is appropriate because ancient pictures are not always the same size.
3. **Random horizontal flip:** The image is mirrored horizontally with a probability of 50%. This is appropriate because a structure should always be coloured the same irrespective of its orientation.

Figure 5 shows 3 original samples compared to one instance of the output of the data augmentation pipeline.

Figure 5: Data augmentation samples from the dataset class compared to their original counterparts.

4.2.1 Format Preparation

Lastly, the dataset class converts the images to the LAB colour space before passing them into the model. While RGB stores the Red, Green and Blue components of an image, the LAB format stores the Lightness of the image (obtained from the greyscale value), the ratio between green-magenta A, and ratio between blue-yellow B. Conversion of images from RGB to LAB is accomplished using Sci-kit Image's library function, rgb2lab. Figure 6 shows one sample image from the dataset and compares the RGB representation to the L, A, and B components.

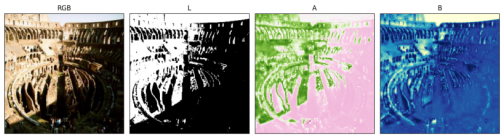


Figure 6: Comparison of the RGB version with the LAB components of an image for one sample image from the dataset.

5 Architecture

The architecture of both the generator and discriminator components of the GAN is a convolutional neural network. As shown in Figure 2 from the Illustration section, the generator creates fake A and B components from the real L component of the input image, while the discriminator must decide if the given AB components are either real from the input or fake from the generator. The number of layers and the sizes of convolution are outlined in Figure 2. N denotes the batch size.

5.1 Generator

The generator is designed with an encoder and decoder block, where the data is gradually compressed towards the central bottleneck then gradually expanded to the appropriate size of the image, making the fake A and B components. The encoder block consists of one convolution block with kernel of 3, padding of 1, and stride of 1, followed by four convolution blocks of kernel of 3, padding of 1, and stride of 2. The decoder block consists of four transposed convolution blocks each with a kernel of 4, stride of 2, and padding of 1, followed by one transpose convolution of kernel of 3, stride of 1, and padding of 1. The encoder receives a duplicated L channel input of size Nx2x128x128 and outputs an A and B channel of Nx2x128x128. All layers use ReLU and batch normalization except the final layer where there is no activation nor normalization. This model contains roughly 11,000,000 trainable parameters. Furthermore, an earlier version of the generator contained a skip connection, but through extensive testing, it was determined that skip connections reduce the model's performance in colourizing. Lastly, it was found that this asymmetric model helped reduce the vanishing gradient problem and prevents over fitting, as suggested by (Meng et al.; Siddiqua & Fan).

5.2 Discriminator

On the other hand, the discriminator is designed as a three-layer convolutional feature extractor that take in an input of Nx2x128x128, all with kernel sizes of 3, strides of 3 (for quick convergence), and padding of 1, as well as a 2-layer linear classifier to eventually give a binary "Real" or "Fake" output of Nx1. Even though Nazeri et al. suggests that tanh is the ideal function to use in this case, it was found that the sigmoid provided better results for this dataset using the nn.BCEWithLogitsloss function. All convolutional layers use batch normalization, and all layers use ReLU except the final one. This model contains 7,000,000 trainable parameters.

6 Baseline Model

The baseline model depicted in Figure 7 was created to compare with the adversarial model. It is composed of a simple convolutional autoencoder architecture with an expected input and output of Nx2x128x128. The encoder consists of a convolutional layer of kernel of 3, stride of 1, padding of 1, followed by similar convolution of stride of 2. The decoder consists of a transpose convolution of kernel of 4, stride of 2, and padding of 1, followed by a convolution of kernel of 3, stride of 1, and padding of 1 that uses the concatenated output of the first convolution and the previous layer. Figure 7 shows the change in sizes, including channel depth, and the skip connection. All layers use ReLU and batch normalization except the output layer which contains no activation function nor normalization. The model contains about 210,000 trainable parameters.

This baseline differs from the generator, as it is trained non-adversarially using an MSE loss rather than the BCE loss from the discriminator. Additionally, it contains much fewer parameters and can be quickly trained. Therefore, comparing the results from this baseline clearly demonstrates whether the generator is performing better than expected.

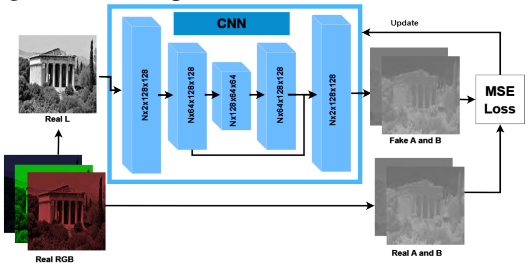


Figure 7: Baseline CNN architecture diagram. Each block represents the layer's size. Each arrow represents the operation (convolution, flattening, fully connected, activation, batch normalization).

7 Results and Discussion

Although the loss curves appear to present conflicting conclusions (as in Figure 8, for example), it is imperative to acknowledge that loss curves are not always sufficient for evaluating the performance of the GAN, specifically for the purpose of image colourization. This is primarily because the loss curves indicate the generator's ability to produce images that fool the discriminator, but they do not indicate the perceptual quality of the colourized images.

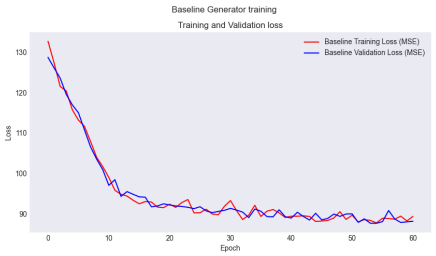
However, the goal of image colourization is to both produce visually identical images and capture the colours and finer details of the original greyscale images. Therefore, solely depending on the loss curves to judge the performance of the model is not appropriate. The quantitative results must be considered in conjunction with the qualitative interpretations of the model's outputs through visual inspection to more accurately assess the model's performance.

7.1 Quantitative Results

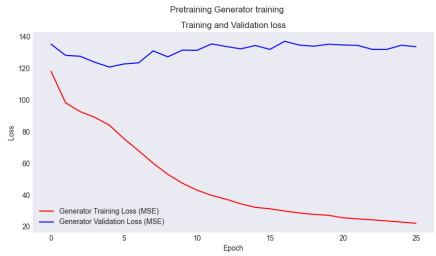
The two primary loss functions used for updating the models are nn.MSELoss (non-adversarial training) and nn.BCEWithLogitsloss (adversarial training). The training and validation losses for the baseline and primary model are presented in Figure 8. The MSELoss is chosen as it computes the difference in the true colours for each pixel in the AB channels, making it an appropriate proxy to determine the "correctness" of the interpolated colours. In contrast, BCEloss was chosen for the discriminator to perform binary classification between the real and fake images.

7.1.1 Baseline Results

The best baseline results occurred upon training continuously for about 60 epochs using Adam with learning rate of 0.0001, as seen in Figure 8a. The MSE loss for both training and validation decreases gradually over the epochs. This indicates that the model is improving in generating colourized images that are closer to the ground truth before stabilizing at a loss of about 89 and 88 for training and validation. After this stabilization, further training may not significantly reduce the errors in the colourization outputs. Thus, the baseline reached its full capacity and more parameters are needed to further train it. It is also seen that the validation and training loss curves are relatively similar to

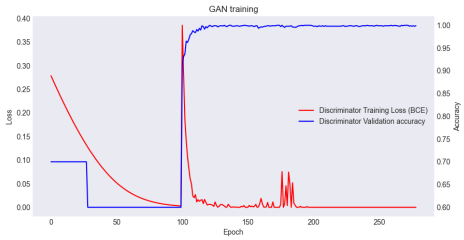


(a) **Baseline Model**: Training Loss (*in Red*) and Validation Loss (*in Blue*) both using MSELoss



(b) **Pre-trained Generator**: Training Loss (*in Red*) and Validation Loss (*in Blue*) both using MSELoss

(c) **Trained GAN - Generator**: Training BCELoss (*in Red*) and Validation MSELoss (*in Blue*)



(d) **Trained GAN Discriminator**: Training BCELoss (*in Blue*) and Validation Accuracy (*in Red*)

Figure 8: Training and Validation Curves for the baseline, pre-trained generator, and trained GAN (generator and discriminator) model. Note that all the training is conducted with batch size of 32.

each other, which is likely due to the skip connection and not a proper colourization (see Qualitative Results Section).

7.1.2 Pre-Trained Generator Results

Even with very simple discriminators, the generator was not initially able to perform better than the baseline when trained adversarially. As such, the model was pre-trained with MSE loss on 27 epochs using Adam with learning rate of 0.0001, as seen in Figure 8b. Through non-adversarial training, the generator can colourize archaeological images with an MSE loss as low as 22 after only 25 epochs on the training dataset. It is also seen that the training and validation curves diverge sufficiently as the number of epochs increases where the training loss is significantly lower than the validation loss, which indicates the model is overfitting. Normally, this would be a sign that the model has too large of a capacity relative to the data set size. However, since this is intended to be a generative model, the validation MSE loss is not expected to reduce since the interpolated colour may not always match the exact colour but can still fool a discriminator. Therefore, the validation MSE loss fluctuating at roughly 130 does not indicate that the model is performing poorly. However, as the curve flattens after 25 epochs adversarial training was implemented at this point to avoid the MSE averaging effects, starting with this pre-trained model instead of random weights.

7.1.3 GAN Model Results

Before interpreting the results, first the choice of loss is explained. As is well-known, GAN loss plots are usually not very insightful and difficult to interpret. As a such, a unique set of losses was chosen to monitor model performance. As can be seen in Figure 8c, the generator's training BCE loss indicates how well it can fool the discriminator while the validation MSE loss indicates how well it can generalize. On the other hand, as seen in Figure 8d, the training BCE loss indicates how well it can determine between real and fake images while the validation accuracy provides an intuitive idea of the number of correct predictions.

To understand the curves in Figures 8c and 8d, it is important to note that the GAN was trained in three phases, all using stochastic gradient descent with no momentum as this provided the best results, using the learning rates from Table 1. Phase 1 showed minimal changes to the generator losses, with the sharp decrease to about 20 MSE validation loss due to further hyperparameter tuning

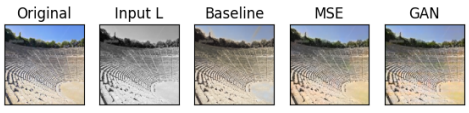
Phase	Epoch Range	Generator Learning Rate	Discriminator Learing Rate
1	0 - 100	0.0001	0.0001
2	100 - 200	0.0001	0.00001
3	200 - 250	0.00001	0.00001

Table 1: Summary of the three phases of training the GAN with learning rates for each model in the phase.

of the pre-trained model. Phase 1 also ensured that the discriminator's loss decreased enough to start learning properly, given the generator was pre-trained. Phase 2's learning rates were chosen such that the generator can fool the discriminator, as seen by the rising curve in Figure 8c, while the discriminator can determine the difference between the real and fake accurately, as seen by the sharp increase in accuracy in 8d. Lastly, Phase 3's learning rates ensured that the validation MSE loss of the generator stopped from rising, as seen by the plateau at about 50, while the noise in the discriminator reduced, as seen by the flatter lines. This ensures that the generator can provide less noisy yet creative colourization. As the generator's BCE loss finished increasing exponentially and the discriminator's loss plateaued at 0.999, this shows that the primary model has reached its capacity and would thus provide reasonable qualitative results that exceed expectations, as explained in the following section.

7.2 Qualitative Results

Figure 9 displays the input and output results from the baseline, pre-trained (MSE), and adversarially trained (GAN) models. Looking at the model outputs, we can observe that the GAN performs better than the baseline and the pre-trained model for many of images in the training, validation, and testing dataset.



(a) **Training Dataset**: Image Sample A

(b) **Training Dataset**: Image Sample B

(c) **Validation Dataset**: Image Sample A

(d) **Testing Dataset**: Image Sample A

(e) **Testing Dataset**: Image Sample B

(f) **Testing Dataset**: Image Sample C

Figure 9: Performance of the baseline and primary model on sample images of the training, validation and testing datasets. The original RGB, the greyscale image (input L) are presented to compare with the Baseline, non-adversarially pre-trained (MSE), and adversarially trained (GAN) model.

7.2.1 Comparison of Vibrancy

Since the baseline concatenates an earlier layer (where there is more gray), the output from the baseline looks slightly more colourized than the original L input but still appears monochromatic relative to the original RGB image and the generator output. As such, the baseline fails to capture the vibrancy and large spectrum of colours which results in duller images, suggesting that the baseline model cannot properly infer colour brightness of the different channels. Thus, both the non-adversarial and adversarial generators perform better than the baseline in this aspect.

7.2.2 Comparison of Object and Boundary Detection

Furthermore, the generator appears to colourize trees and structural element better than the baseline, which colourizes these objects with the same colour, as shown in Figure 9c. As well, the baseline colourizes yellowish-brown between the trees and the sky, as in Figure 9e, suggesting that the baseline model cannot properly distinguish boundaries between objects. In contrast, the non-adversarial and adversarial generators in the same Figure properly colourize these sections, thus demonstrating superior performance to the baseline in this aspect as well.

7.2.3 Comparison of Colour Diversity Identification

Pertaining to more diverse ranges of colour, the adversarially trained generator outperforms both the non-adversarial generator and the baseline, which allows it to better generalize the true colours and finer boundary details of unseen images. At closer inspection of the provided images, it can be observed that adversarially trained generator yields slightly brighter images with a higher saturation of colour (matching the initial input image), but also reduces the monochromaic noise in images such as the centre of Figures 9d and 9f. Notice the shells of the amphitheatre in the MSE are incorrectly yellow while they are correctly white in the GAN. This is most likely due to the GAN's design, which was made to overcome the averaging effect of the MSE loss, as it does not necessarily produce the exact colours, but is able to fool the discriminator. Lastly, it was found that the adversarial generator colourizes trees and the sky well, but struggles with the grey concrete or rocks because it seeks to minimize MSE loss on more vibrant objects and considers duller objects to already be minimzed.

8 Evaluate the Model on New Data

To ensure the results are a good representation of the GAN model's performance on new data, a testing data set of 242 images was set aside as unseen data, as shown on the testing dataset in Figures 9d, 9e, and 9f. This data are neither used for tuning the hyperparameters nor validating the model.

Given that the GAN is a generative model that must infer colour from a greyscale image, it is expected that the model to identify features such as a sky, amphitheaters, and trees, and colourize them blue (or white with clouds), yellow or brown or red, and green, respectively. In other words, the GAN is not expected to reproduce the colours of the original image pixel for pixel, as this is an almost impossible tasks, even for humans. As such, the GAN model performed remarkably well on the testing dataset and successfully colourized the greyscale input by correctly identifying crisp boundaries and colourizing either side with the expected colour. Furthermore, as discussed previously, it is also reasonable to conclude that the GAN is performing better than the pre-trained (MSE) model and much better than the baseline, both quantitatively and qualitatively.

Given the generative nature of the model, the GAN appears to colourize images with slightly more vibrancy than even the original RGB image. For example in Figure 9d, it is obvious that the staircase in the original RGB is slightly dull and more red than the GAN output. Similarly, in Figure 9e the left hand bottom corner of the original image is less yellow and less bright than the GAN counterpart. As the GAN model is designed to bring back life into the images, this is an unexpected feature that exceeds the team's expectations.

However, there are differences that, while not detracting from the success of the model in both colourizing the images by capturing colours and the finer details, should still be considered. The performance of any model can always be further maximized. For example, the sky in the GAN output is less blue than the original RGB of figure 9e. Overall, the performance of the GAN exceeds the team's expectation for the problem being solved in colourizing archaeological images, as it can provide a reasonable colour scheme to new data as would a human having only the black and white image available while identifying the key features of the image with the appropriate colour.

9 Ethical Considerations

Despite its diverse applications in colourizing black and white films, portraits, and historical scenes, image colourization presents three main concerns that must be acknowledged to ensure the ethical use of such a model. These concerns arise primarily from the choice of images for the dataset. As such, a well-diversified data set is chosen with 4842 archaeological images.

Firstly, using Deep Learning models to colourize an image may introduce cultural and artistic bias if there exists a bias in the training dataset. This leads to inaccuracies in the model and misinterpretations of the image content. Therefore, selecting a diversified dataset is fundamental to ensuring that the dataset is well-representative of the the intended purpose. Nevertheless, image colourization may still be contextually inappropriate for culturally-sensitive or historical images in archaeology.

Additionally, since the chosen dataset does not contain many images with faces, the model can inaccurately or insensitively colourize humans in the archaeological images if the model contains biases or lacks information related to human facial features or skin complexion. This inadvertently perpetuates harmful stereotypes and can potentially exacerbate issues regarding racial profiling and discrimination, especially when the training data is not diverse enough to account for all ethnic backgrounds.

Image colourization presents questions regarding the ownership and copyright of the colourized images as the process of doing so may challenge and violate the intellectual property of the original creators. As such, it is imperative to respect the rights of the original creators and photographers of the images. Similarly, if the original creators did not intend for their work to be colourized, doing so without their consent or giving them credit may infringe on their creative expression and artistic intent.

10 Project Difficulty Relative to Quality

Ultimately, using a GAN model to build an image colourizer is a very challenging project and is beyond the scope of the lab requirements and academic syllabus, for the following five reasons. Nevertheless, even with this higher difficulty, the team managed to provide results beyond expectations.

First, GANs were only covered in one lecture, after the bulk of the project was completed, and GANs were limited to one tutorial on speech generations unrelated to image colourization. Second, none of the team members had domain expertise with GANs before, which required a lot of time commitment and self-learning throughout the project. Extensive research was conducted on choosing the architecture, building the training loop, and interpreting the loss functions in a general GAN context. Thus, the team learned how to successfully apply GANs to this project, as image colourization on amphitheaters is not a documented topic.

Third, since the team used the CIELAB colour space, research was conducted to ensure that RGB images were converted to CIELAB and back. However, due to the complex nature of this transformation, the generator did not always produce images within a range displayable by the conversion functions. As such, extensive debugging of the image processing and visualization severely slowed down the training process, which was needed to see how the model is performing qualitatively.

Fourth, to the same effect, extensive experimentation was done for tuning hyperparameters and improving architecture design. These include extensive research and trial and error with at least 40 different model architectures with varying layer depth, kernel size, number of skip connections, type of skip connections, type of activation function, among other model parameters. Further hyperparameter tuning was completed on each model at least once, with better performing models being further tuned. This included changing the learning rate, batch size, number of epochs, optimizer, and learning sequence, as discussed briefly in the Quantitative Results section. Furthermore, due to initally poor results, the model was pre-trained, which included investigating transfer learning with ResNet on greyscale images, as well as concluding that pre-training the project's generator yielded the best results.

Lastly, due to the large dataset and limited GPU, the model was trained very slowly throughout the semester using team members' GPU runtimes. This extended the training process on an already time-consuming GAN model, which may have prevented better tuning.

Therefore, the above reasons clearly demonstrate project's difficulty and effort required to produce high-quality results. Due to the team's diligence and dedication to this project, the results were above the team's and the course's expectation as described in the previous sections, successfully colourizing never-before-seen historical images.

REFERENCES

Amphitheatre. URL https://ancienttheatrearchive.com/glossary-term/amphitheatre/.

Grayscale Image - an overview — ScienceDirect Topics. URL https://www.sciencedirect.com/topics/engineering/grayscale-image.

Yun Cao, Zhiming Zhou, Weinan Zhang, and Yong Yu. Unsupervised Diverse Colorization via Generative Adversarial Networks. In Michelangelo Ceci, Jaakko Hollmén, Ljupčo Todorovski, Celine Vens, and Sašo Džeroski (eds.), *Machine Learning and Knowledge Discovery in Databases*, volume 10534, pp. 151–166. Springer International Publishing. ISBN 978-3-319-71248-2 978-3-319-71249-9. doi: 10.1007/978-3-319-71249-9_10. URL https://link.springer.com/10.1007/978-3-319-71249-9_10.

Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep Colorization. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 415–423. IEEE. ISBN 978-1-4673-8391-2. doi: 10.1109/ICCV.2015.55. URL http://ieeexplore.ieee.org/document/7410412/.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. doi: 10.48550/ARXIV.1406.2661. URL https://arxiv.org/abs/1406.2661.

Paulina Hensman and Kiyoharu Aizawa. cGAN-Based Manga Colorization Using a Single Training Image. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pp. 72–77. IEEE. ISBN 978-1-5386-3586-5. doi: 10.1109/ICDAR.2017.295. URL http://ieeexplore.ieee.org/document/8270240/.

Takahiko Horiuchi. Colorization algorithm using probabilistic relaxation. 22(3):197–202. ISSN 02628856. doi: 10.1016/j.imavis.2003.08.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0262885603001756.

Shanshan Huang, Xin Jin, Qian Jiang, and Li Liu. Deep learning for image colorization: Current and future prospects. 114:105006. ISSN 09521976. doi: 10.1016/j.engappai.2022.105006. URL https://linkinghub.elsevier.com/retrieve/pii/S0952197622001920.

Christopher Kanan and Garrison W. Cottrell. Color-to-Grayscale: Does the Method Matter in Image Recognition? 7(1):e29740. ISSN 1932-6203. doi: 10.1371/journal.pone.0029740. URL https://dx.plos.org/10.1371/journal.pone.0029740.

Rui Meng, Shuaidong Yin, Jianqiang Sun, Huan Hu, and Qi Zhao. scAAGA: Single cell data analysis framework using asymmetric autoencoder with gene attention. 165:107414. ISSN 00104825. doi: 10.1016/j.compbiomed.2023.107414. URL https://linkinghub.elsevier.com/retrieve/pii/S001048252300879X.

Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi. Image Colorization with Generative Adversarial Networks. doi: 10.48550/ARXIV.1803.05400. URL https://arxiv.org/abs/1803.05400.

Ayesha Siddiqua and Guoliang Fan. Asymmetric Supervised Deep Autoencoder for Depth Image based 3D Model Retrieval. In *2019 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4. IEEE. ISBN 978-1-72813-723-0. doi: 10.1109/VCIP47243.2019.8965682. URL https://ieeexplore.ieee.org/document/8965682/.

Lvmin Zhang, Chengze Li, Tien-Tsin Wong, Yi Ji, and Chunping Liu. Two-stage sketch colorization. 37(6):1–14. ISSN 0730-0301, 1557-7368. doi: 10.1145/3272127.3275090. URL https://dl.acm.org/doi/10.1145/3272127.3275090.

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Antonio Torralba, and Aude Oliva. Places: An Image Database for Deep Scene Understanding. a. doi: 10.48550/ARXIV.1610.02055. URL https://arxiv.org/abs/1610.02055.

Jinjie Zhou, Kai Hong, Tao Deng, Yuhao Wang, and Qiegen Liu. Progressive Colorization via Iterative Generative Models. 27:2054–2058, b. ISSN 1070-9908, 1558-2361. doi: 10.1109/LSP.2020.3037690. URL https://ieeexplore.ieee.org/document/9258392/.