IMAGE COLOURIZATION VIA CONVOLUTIONAL NEURAL NETWORKS AND DEEP LEARNING

Youssef Fikry
Student# 1005678901
youssef.fikry@mail.utoronto.ca

Harkirpa Kaur
Student# 1011242479
harkirpa.kaur@mail.utoronto.ca

Peter Leong
Student# 1005678901
peter.leong@mail.utoronto.ca

Thulasi Thavarajah
Student# 10115358424
t.thavarajah@mail.utoronto.ca

ABSTRACT

This template should be used for all your project related reports in APS360 course.
– Write an abstract for your project here. Please review the **First Course Tutorial** for a quick start —-Total Pages: 3

1 INTRODUCTION

While the colour photography processes first emerged in the 1890s, colour photography did not become widely accessible until the 1970s Science & Museum (2020). As a result, most historic photographs are black and white and lack the visual richness that modern viewers are accustomed to. In addition, individuals who have their cataracts removed as a part of vision restoration processes have shown to struggle with identifying objects in black and white images Vogelsang et al. (2024), making most historic photographs inaccessible to them. This project aims to use deep learning to automatically colorize black and white images, with the goal of restoring visual information and making historical imagery more accessible for all audiences. Traditional, non deep-learning based colorization methods often produce desaturated results and rely heavily on human guidance Cheng et al. (2016), making them non-scalable. Conversely, deep networks such as CNNs effectively capture spatial and semantic features and produce realistic colourized images without user interaction Zhang et al. (2016), making deep learning an ideal approach for image colorization.

1.1 BACKGROUND & RELATED WORK

The challenge of image colourization has been approached in a wide variety of ways. Even within solutions involving deep learning techniques, there numerous unique design choices. One grouping method Zěger et al. (2021) results in five categories: simple colourization neural networks, user-guided colourization neural networks, diverse colourization neural networks, multi-path colourization neural networks, and examplar-based neural networks.

Simple colourization neural networks use feedforward CNNs to map grayscale inputs to colour outputs. One of the foremost solutions proposed by Zhang et al. (2016) used a fully convolutional network to predict the a and b channels of the CIELAB colour space from grayscale images. Their architecture is composed of several convulational layers, each followed by a ReLU activation function and a batch normailization layer.

User-guided colourization neural networks use user input to guide the colourization process. One such solution Zhang et al. (2017) uses a fully convolutional network to predict the a and b channels of the CIELAB colour space from grayscale images, but also takes user input in the form of user-provided colour scribbles. The network is trained to minimize the difference between the predicted and user-provided colours, allowing it to learn to colourize images in a way that is consistent with the user's input.

Diverse colourization networks produce multiple colourization outputs for a given grayscale input. One such solution Vitoria et al. (2020) uses a generative adversarial network (GAN) to produce multiple colourization outputs for a given grayscale input. The GAN is trained to minimize the difference between the predicted and ground truth colours, allowing it to learn to produce diverse colourization outputs.

Multi-path colourization neural networks differentiate features at different scales. Iizuka et al. (2016) proposed a multi-path colourization neural network that uses multiple convolutional layers to extract features at different scales. The network is trained to minimize the difference between the predicted and ground truth colours, allowing it to learn to produce colourization outputs that are consistent with the features at different scales.

Exemplar-based neural networks use a set of exemplar images to guide the colourization process. In Su et al. (2020), example images are used to trasnfer the colour to the target image. Each instance is ouput to two different colourization networks which fuse to yield the final result. This group of solutions is easier to implement, as learning to colourize instances is significantly easier than learning to colourize an entire image.

2 METHODOLOGY

2.1 DATA PROCESSING

The project's image dataset was compiled using individual datasets from Kaggle.com, an online platform that provides access to real-world datasets and a community for data scientists. To train a model that can generalize to a broader range of images, the final dataset for this project is comprised of three categories: human, animal, and scenic. All datasets chosen have a public domain license.

2.1.1 REPURPOSING ONLINE DATASETS

The image dataset was chosen because its original purpose is for human detection (Elmenshawii, 2023). As a result, the dataset contains a diverse range of 17,300 images of people in different environments.

The animal image dataset was originally for image classification as it contains 5,400 images of 90 different animals (Banerjee, 2022). For this project's purpose, this dataset is ideal as it results in a diverse set of images with an equal distribution of the number of images per animal.

The scenic image dataset from Rougetet (2020) contains 4,319 images of a variety of landscapes that encompass a large breadth of colour palettes, which can impact how well the final model performs.

2.1.2 CLEANING UP THE DATASETS

All the images from each dataset were extracted from their original folder organizations and relocated into a folder corresponding to their category. Due to the disparity in the size of the three datasets, each dataset was reduced to exactly 4200 images using Python's `random.sample` function with the seed set to 42. The images were then renamed in accordance to their respective categories (ex. human_0001.jpg).

2.1.3 FORMATTING THE DATA

This project requires a unique dataset with black and white images, along with their coloured counterparts. To format the cleaned up data and create this dataset, PyTorch's `torchvision.transforms` library was utiltized. The three sets of images were converted to 256 x 256 pixel ground truth images using the `Resize` transform and then transferred into a folder corresponding to their category. Following this, `Grayscale` transform with a output channel of 1 was used to convert the 256 x 256 pixel colour images to 256 x 256 pixel grayscale images to feed into the model.

2.1.4 SPLITTING DATASET INTO TRAINING, VALIDATION AND TEST SETS

To create the training, validation and test sets, a ratio of 70:15:15 was chosen. write steps

2.1.5 The Final Dataset

The final dataset is composed of 12,600 pairs of color and their corresponding grayscale images. Each category (human, animal, and scenic) contains 4,200 image pairs. There are 8,820 pairs in the training set, and the validation and test sets each contain 1,890 image pairs. An even distribution of each category was maintained in the training, validation and test sets.

2.2 Architecture

2.3 Baseline Model

3 Ethical Considerations

The dataset being used is public, so there are no copyright or consent issues. However, the dataset may contain racial or demographic imbalances, which could cause the model to generalize poorly or be biased towards specific skin tones. This may result in racially inaccurate or culturally insensitive outputs. A similar behaviour may be observed with animals, where a lack of diversity in breeds or fur colours in the dataset can result in misleading results. If the outputs produced by the model are used in educational contexts or in breed identification, they can contribute to misinformation. Furthermore, since the model results are plausible but cannot be verified, there is a risk that users may overtrust the outputs in sensitive contexts.

4 Project Plan

Our high-level project planning has been completed using a Gantt chart seen below. It contains many of the detailed tasks we have completed and will complete in the future. This easily shows who is responsible for each task, the time it will take to complete, and the dependencies between tasks. Generally, our internal deadlines have been set at least a few days prior to the official deadlines to ensure we have time to review our work and make any necessary changes. In the event that a new member joins the team, they can easily refer to this chart to see what responsibilites they can take on and what tasks are still available. Our Gantt chart can be found in Appendix A.

4.1 Team Communication & Coordination

We intend to use Discord as our primary communication platorm, Google Drive for file sharing (Colab notebooks, datasets, etc.), and GitHub for version control. To ensure we do not overwrite each other's work, we will use branches for each feature or task, and merge them into the main branch once they are complete. We will also have weekly meetings to discuss our progress, any issues we are facing, and plan for any action items for the subsequent meeting. Currently, we intend to meet on Saturday mornings at 8:00 AM EST as some of our team members are in different time zones.

5 Risk Register

References

Sourav Banerjee. Animal image dataset (90 different animals), 2022. URL `https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals`.

Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization, 2016. URL `https://arxiv.org/abs/1605.00075`.

Fares Elmenshawii. Human dataset, 2023. URL `https://www.kaggle.com/datasets/fareselmenshawii/human-dataset/data`.

Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (TOG)*, 35(4):110:1–110:11, July 2016. ISSN 0730-0301. doi: 10.1145/2897824.2925974. URL https://doi.org/10.1145/2897824.2925974.

Arnaud Rougetet. Landscape pictures, 2020. URL https://www.kaggle.com/datasets/arnaud58/landscape-pictures.

National Science and Media Museum. A short history of colour photography, Jul 2020. URL https://www.scienceandmediamuseum.org.uk/objects-and-stories/history-colour-photography.

Jheng-Wei Su, Hung-Kuo Chu, and Jia-Bin Huang. Instance-aware image colorization, 2020. URL https://arxiv.org/abs/2005.10825.

Patricia Vitoria, Lara Raad, and Coloma Ballester. ChromaGAN: Adversarial Picture Colorization with Semantic Class Distribution. In *Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 2434–2443, Los Alamitos, CA, USA, March 2020. IEEE Computer Society. doi: 10.1109/WACV45572.2020.9093389. URL https://doi.ieeecomputersociety.org/10.1109/WACV45572.2020.9093389.

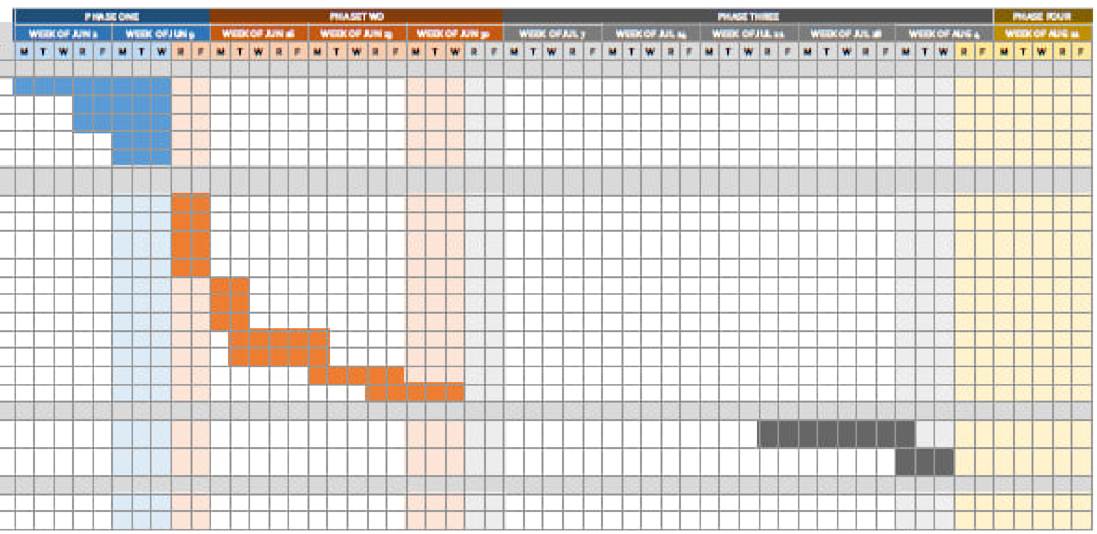
Marin Vogelsang, Lukas Vogelsang, Priti Gupta, Tapan K. Gandhi, Pragya Shah, Piyush Swami, Sharon Gilad-Gutnick, Shlomit Ben-Ami, Sidney Diamond, Suma Ganesh, and Pawan Sinha. Impact of early visual experience on later usage of color cues. *Science*, 384(6698):907–912, 2024. doi: 10.1126/science.adk9587. URL https://www.science.org/doi/abs/10.1126/science.adk9587.

Ivana Zĕger, Sonja Grgic, Josip Vuković, and Gordan Šišul. Grayscale image colorization methods: Overview and evaluation. *IEEE Access*, 9:113326–113346, 2021. doi: 10.1109/ACCESS.2021.3104515.

Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. *European conference on computer vision*, 2016.

Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. Real-time user-guided image colorization with learned deep priors, 2017. URL https://arxiv.org/abs/1705.02999.

6 Appendix A: Gantt Chart



PROJECT TITLE	APS360 - Image Colourization
DATE	[illegible]

TASK NUMBER	TASK TITLE	START DATE	DUE DATE	DURATION (DAYS)	TEAMMATE RESPONSIBLE
1	Project Proposal and Initial Research	[illegible]	[illegible]	10	
1.1	Preliminary Background Research	[illegible]	[illegible]	9	Peter
1.1.1	Baseline Model Research	[illegible]	[illegible]	6	Yousef
1.2	Dataset Characterization	[illegible]	[illegible]	8	Thulasi
1.3	Ethics Research	[illegible]	[illegible]	5	Kripa
1.5	Project Planning	[illegible]	[illegible]	6	All
2	Project Foundation Detailed Planning	[illegible]	[illegible]	19	
2.1	Detailed Background Research	[illegible]	[illegible]	1	
2.1.1	Reference Solutions	[illegible]	[illegible]	1	
2.1.2	Detailed Architecture & Hyperparameter Research	[illegible]	[illegible]	1	
2.1.3	Metrics Research	[illegible]	[illegible]	1	
2.2	Dataset Processing	[illegible]	[illegible]	1	
2.3	Detailed Architecture Design	[illegible]	[illegible]	1	
2.4	Detailed Baseline Model Design	[illegible]	[illegible]	1	
2.5	CNN Model v1.0 Implementation	[illegible]	[illegible]	6	
2.6	Baseline Model v1.0 Implementation	[illegible]	[illegible]	6	
2.7	Initial Validation Results	[illegible]	[illegible]	5	
2.8	Project Progress Report	[illegible]	[illegible]	5	
3	Model Optimization & Presentation	[illegible]	[illegible]	31	
3.7	Project Presentation Formatting (Group)	[illegible]	[illegible]	11	
3.8	Project Presentation Recording (Group)	[illegible]	[illegible]	2	
4	Final Project Report	[illegible]	[illegible]	10	
4.1					
4.2					