APS360 PROJECT PROPOSAL: PICTURE COLOURIZATION

Claire He
Student# 1010255698
claire.he@mail.utoronto.ca

Huda Imran
Student# 1009910606
huda.imran@mail.utoronto.ca

Manha Siddiqua
Student# 1010315375
manha.siddiqua@mail.utoronto.ca

Jihoon You
Student# 1009796801
jihoon.you@mail.utoronto.ca

ABSTRACT

Our project proposes to develop a deep learning system to automatically colourize greyscale images using CNN-based encoder-decoder architectures with generative adversarial training networks (GAN). This method will convert images to LAB colour space, using the L channel (lightness) as input and then predict the a,b channels (colour information). The system will be trained on large datasets to learn the relationship between greyscale patterns and realistic colour distributions. —-Total Pages: 5

1 INTRODUCTION

Photography is a very powerful tool utilized by humanity to capture, record, and preserve information and memories. However, it was not until the 1970s that the use of colour photography became popularized (Artsper Magazine, 2023). Currently, there are still many historical and personal photos from that time that are still in black and white, creating a visual disconnect and an inaccurate presentation of what the past was like for people to learn from now. In this project, we propose using deep learning algorithms to implement the colourization of greyscale images. Our goal is for our project to be able to convert black-and-white images into colour by estimating the RGB colours for the image (Anwar et al., 2025). The image colourizer should utilize convolutional neural networks (CNNs) to learn colour and patterns to apply to black and white photos. Machine learning is the appropriate tool to do this due to CNNs ability for image detection and breaking up an image in layers that detect different features, building a deeper understanding of the image to choose appropriate colouring (Mandić et al., 2024). Through this, machine learning tools can complete the colourization process in mere seconds, compared to the time it takes to colourize manually.

2 ILLUSTRATION

Our approach converts greyscale images to colour by predicting missing colour information using deep learning techniques. The overall system's structure can be seen in Figure 1. Our colourization system uses a simple yet effective CNN-based approach to transform greyscale images into colour. The process begins by converting the input image to the LAB colour space and extracting the L channel. Our trained CNN model then predicts the corresponding colour ab channels by learning patterns from a dataset of coloured images. These predicted ab channels are combined with the original L channel and converted back to the RGB space to produce the final colourized result.

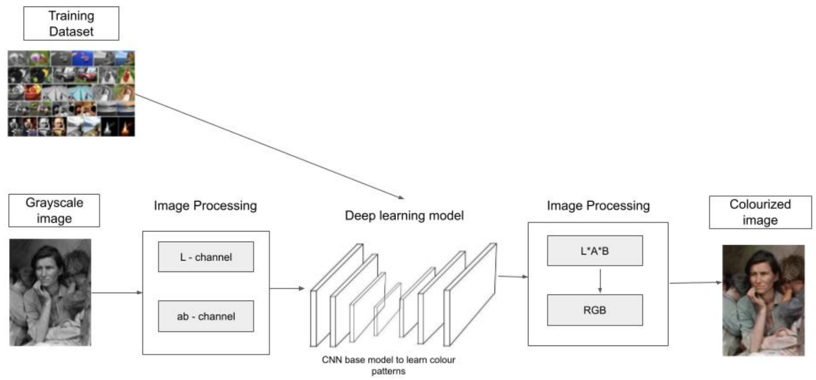

Figure 1: Our overall project structure, which consists of the path from greyscale input to colourized output through our deep learning model

3 Background and Related Work

Image colourization is fundamentally a type of image classification and regression problem, where the goal is to predict possible colour values for each pixel from a black and white image. For some background knowledge, CNNs are usually the standard approach for this task because they are ideal in supervised image classification (Richards, 2022). Expanding on that point, CNNs process images through layers that use different filters (or building blocks, usually called kernels) to analyze the image and automatically detect and learn spatial hierarchies of features. These kernels can be optimized during training to ”minimize the difference between outputs and ground truth labels through an optimization algorithm called backpropagation” (Yamashita et al., 2018).

A significant amount of research has been conducted in the field of image colourization, with various approaches, several contributions include:

1. User-guided colourization methods. Early image colourization models often relied on user interaction to colour an image. Levin et al. (2004) demonstrated how this method allows the user to colourize the image by annotating the greyscale images using coloured strokes or scribbles. The colours are then propagated throughout the image using optimization techniques.

2. Goodfellow et al. (2014) introduced a new way to train generative models, Generative Adversarial Networks (GAN), where two neural networks, a generator and a discriminator, work to train and create realistic data. For example, Pix2Pix was a model developed by Isola et al. (2017) which used GANs for image colourization, from greyscale inputs, where the generator predicts a possible LAB colour scheme, and the discriminator checks for realism, to create the best output. This idea helps us work towards our project and the goal of image colourization, with accurate and vibrant colours.

3. In 2015, Ronneberger et al. (2015) developed a CNN for more precise image segmentation called U-Net, which utilizes an encoder-decoder structure. The encoding layers compress the black-and-white image to capture important features, and the decoding layers reconstruct the image back to its original size to restore and add colour (Nazeri et al., 2018). U-Net uniquely incorporates skip connections, which help the network retain important details like patterns, making the colourized output more accurate to the original (Nazeri et al., 2018).

4. Another approach is through the use of Stacked Convolutional Auto-Encoders (CAEs), which stacks several Convolutional Auto-Encoders with each layer receiving its input from the layer below. Each CAE learns a different feature of the original input. A CNN can be

initialized by a trained CAE stack to have improved performance of the model compared to a standard CNN (Masci et al., 2011).

5. On the other hand, other previous works explored developing fully automatic colourization models that require no human input. For example, Gupta et al. (2012) proposed an example-based method where the user provides a reference colour image similar to the target greyscale image. The model then transfers colours from the reference image to the greyscale image using superpixel-based feature matching. This automatically identifies corresponding regions between greyscale and colour images.

4 Data Processing

The advantage of such a project is the abundance of data sources, as any image can be used as data. We will use a 50,000 image subset of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) image classification and localization dataset. In addition, we will also use an 800 image dataset of "Random Unsplash Images" (similar to stock photos), improving data diversity and to test the generalizability of our model. Both of these datasets can be found on Kaggle (Kaggle.com, 2022), (Kaggle.com, 2021).

The ImageNet LSVRC dataset has been previously used in some of the most well-known automatic image colourization models, such as those presented by Zhang et al. (2016) and Larsson et al. (2016), which can provide an interesting point of comparison further down the line regarding our model's performance. Data augmentation through random image transformations (flips/rotations) and the addition of noise are other preprocessing steps we are considering, including, to simulate real-world inconsistencies and obtain a more diverse dataset (Mumuni & Mumuni, 2022).

For both datasets, we will resize all images to 572x572, to standardize input dimensions, and convert all images from the RGB colour space to the CIELAB colour space. Here, we will remove all images that are already monochromatic, as the ab channel values are to be used as ground truth for supervised learning.

We will use 70% of dataset 1 (ILSVRC) for training and 20% for validation. For testing, we will use the remaining 10% of dataset 1 and the entirety of dataset 2.

5 Architecture

Our proposed image colourization network will be based on U-Net, a Convolutional Neural Network with an Encoder-Decoder structure (Ronneberger et al., 2015). Other CNN architectures have been successfully employed in the task of image colourization; most prominently, VGG-style networks, such as by Zhang et al. (2016) and Larsson et al. (2016), have resulted in highly realistic images. However, U-Net-based image colourization networks improve on the results of those networks, resulting in greater detail and clearer boundaries between objects, as noted by Wang et al. (2022), who employed a modified version of U-Net, titled "CU-Net".

The U-Net architecture has two main components: a "contracting path" and an "expanding path" (Figure 2). The contracting path follows a typical CNN structure, with convolution, max pooling, and ReLu layers, while the expanding path up-samples the compressed data back to the original image dimensions, while also predicting pixel colours (Ronneberger et al., 2015). In addition, U-Net employs "skip connections" that connect information from the contracting path (i.e. about boundaries, shapes, textures, etc.) to the expanding path, improving accuracy (Wang et al., 2022).

6 Baseline Model

We decided to use a simple CNN-based encoder and decoder as our baseline colourization model because CNNs are well-suited for image data. They take advantage of two key principles: locality, where nearby pixels are often related, and translation invariance, which allows the network to recognize features regardless of their position it is in Kamtziridis (2022). These properties make CNNs ideal for our task of image colourization. For our model, we will take a greyscale L channel (from the LAB colour space) as the input and predict the missing a and b colour channels, giving us the

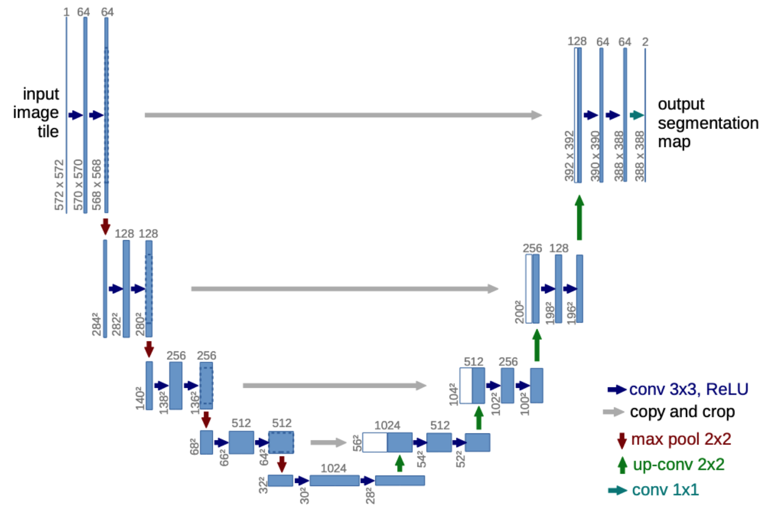

Figure 2: A general overview of the U-Net architecture (Ronneberger et al., 2015)

output. As shown in Figure 3, specifically, the encoder uses a few conventional layers to extract features from the greyscale image. The decoder then uses upsampling and more conventional layers to rebuild the colour data.

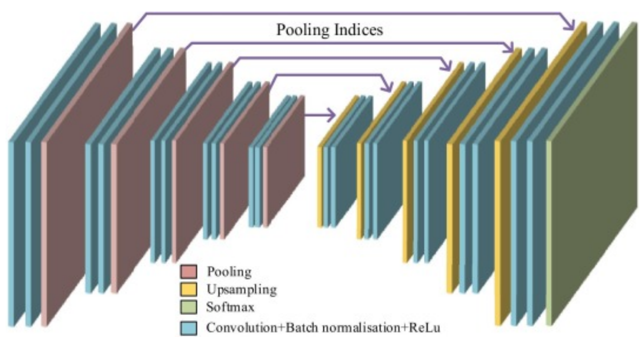

Figure 3: A visual representation of a CNN encoder-decoder for image decomposition Ümit Budak et al. (2020)

7 Ethical Considerations

Many ethical concerns need to be considered for image colourization. One major issue is the probability of false or incorrect colourization of historical photos, resulting in the distortion of historical events and the perpetuation of discrimination. If the training data we feed into the neural network contains biases, the model might reproduce those biases, creating discriminatory or insensitive outcomes (Carleton University, 2024). There is also the ethical issue of the corruption of artists' creative vision, as many artists may not want their images to be turned to colour, as it may ruin their initial intent with the original form (Liu et al., 2022).

8 Project Plan

As a team we will hold weekly meetings to work on the project together, and also to check on the progress of the team. The team will also be in constant contact via text on the project's group chat, and will use the git repository to update any work they have. We will use our branches and merge our work together, using pull. We will distribute tasks regularly among the team, and work on project deliverables together as seen in Figure 4.

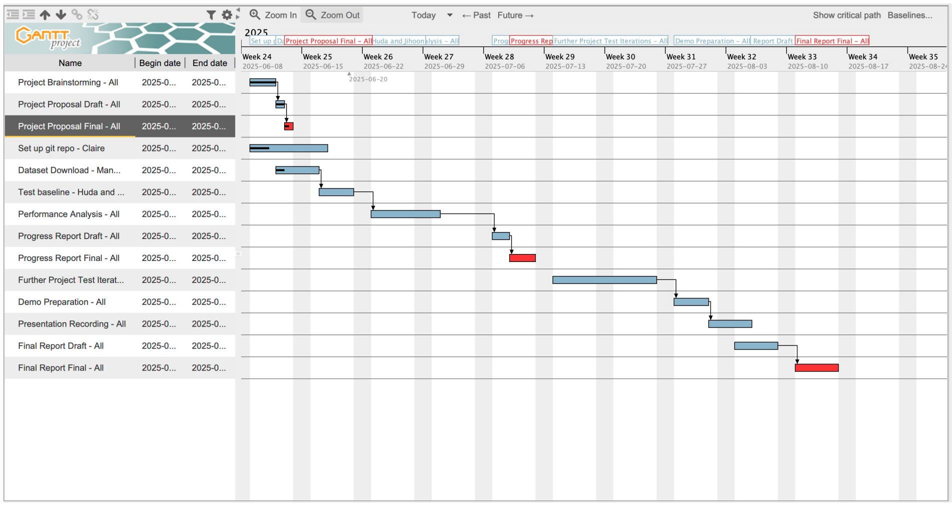

Figure 4: Using Gantt Chart, we have created a project plan that includes task distribution, deadlines, and progress checks.

9 Risk Register

In order for our project to be a success, we have to consider possible risk factors, their possibilities, and the proposed mitigation strategies we plan on implementing as provided in Table 1.

Table 1: Potential risks, their likelihood, and mitigation strategies

Risk	**Possibility**	**Mitigation Strategy**
Team member drops the course	5 percent	the rest of the members have enough experience and can divide the workload among each other.
It is hard for the team to meet and work on the project	10 percent	Team members can regularly communicate via text, and also have regular team meetings every week for project updates.
Model training takes longer than expected	30 percent	We can use different hyperparameters, and pre-train different sections of our model related to semantic classification and segmentation.
Image output is saturated or inaccurate	30 percent	Training on diverse datasets can help the model learn a wide range of colour mapping. The use of GANs will also allow for accurate and vibrant colours.

10 Link to Repository

Colab Notebook Link:
Click here to view Colab page

REFERENCES

Saeed Anwar, Muhammad Tahir, Chongyi Li, Ajmal Mian, Fahad Shahbaz Khan, and Abdul Wahab Muzaffar. Image colorization: A survey and dataset. *Information Fusion*, 114:102720, 2025. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2024.102720. URL https://www.sciencedirect.com/science/article/pii/S1566253524004986.

Artsper Magazine. A closer look at the history of color photography, 2023. URL https://blog.artsper.com/en/a-closer-look/color-photography/.

Carleton University. Reflection on the ethics of colorizing with ai. https://hh2024w.amason.sites.carleton.edu/week-4-spatial-humanities/reflection-on-the-ethics-of-colorizing-with-ai/, 2024. Accessed: 2025-06-13.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014. URL https://papers.nips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf.

Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong. Image colorization using similar images. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pp. 369–378, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310895. doi: 10.1145/2393347.2393402. URL https://doi.org/10.1145/2393347.2393402.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5975, 2017.

Kaggle.com. Unsplash random images collection, 2021. URL https://www.kaggle.com/datasets/lprdosmil/unsplash-random-images-collection.

Kaggle.com. Imagenet-1k-valid, 2022. URL https://www.kaggle.com/datasets/sautkin/imagenet1kvalid.

George Kamtziridis. Building an image colorization neural network — part 3: Convolutional neural networks. Medium, 2022. URL https://medium.com/@geokam/building-an-image-colorization-neural-network-part-3-convolutional-neural-networks

Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. *Computer Vision - ECCV 2016*, 2016. doi: 10.48550/arXiv.1603.06668.

Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23(3):689–694, August 2004. ISSN 0730-0301. doi: 10.1145/1015706.1015780. URL https://doi.org/10.1145/1015706.1015780.

Siyang Liu, Arjun Taluja, and David Huang. Image colorization using generative adversarial networks and transformers. Technical Report CS231n Final Report, Stanford University, 2022. URL https://cs231n.stanford.edu/reports/2022/pdfs/109.pdf. Accessed: 2025-06-13.

Lidija Mandić, Jesenka Pibernik, Maja Kurečić, and Ana Cmrk. Artificial intelligence or man: Colorization of black and white photographs. 11 2024. doi: 10.24867/GRID-2024-p34.

Jonathan Masci, Ueli Meier, Dan Cireşan, and Jurgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning – ICANN 2011*, Lecture Notes in Computer Science. Springer, 2011. URL https://people.idsia.ch/~ciresan/data/icann2011.pdf.

Alhassan Mumuni and Fuseini Mumuni. Data augmentation: A comprehensive survey of modern approaches. *Array*, 2022. doi: 10.1016/j.array.2022.100258.

Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi. Image colorization using generative adversarial networks. In *International Conference on Articulated Motion and Deformable Objects (AMDO)*, pp. 85–94. Springer, 2018. URL `https://www.imaginglab.ca/data/NaNgEb2018AMDO.pdf`.

John A. Richards. *Supervised Classification Techniques*, pp. 263–367. Springer International Publishing, Cham, 2022. ISBN 978-3-030-82327-6. doi: 10.1007/978-3-030-82327-6_8. URL `https://doi.org/10.1007/978-3-030-82327-6_8`.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, 2015. doi: 10.48550/arXiv.1505.04597.

Na Wang, Guo-Dong Chen, and Ying Tian. Image colorization algorithm based on deep learning. *Symmetry*, 2022. doi: 10.3390/sym14112295.

Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4): 611–629, 2018. doi: 10.1007/s13244-018-0639-9. URL `https://doi.org/10.1007/s13244-018-0639-9`.

Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *Computer Vision - ECCV 2016*, 2016. doi: 10.48550/arXiv.1603.08511.

Ümit Budak, Yanhui Guo, Erkan Tanyildizi, and Abdulkadir Şengür. Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation. *Medical Hypotheses*, 134:109431, 2020. ISSN 0306-9877. doi: https://doi.org/10.1016/j.mehy.2019.109431. URL `https://www.sciencedirect.com/science/article/pii/S0306987719309971`.

,