Adapting Existing Colorization Models to Small Datasets: GAN Approach

Using Generative Adversarial Networks to Post-process Baseline Results

Alexander Sukennyk
Computer Science
Cleveland State University
Cleveland, Ohio
a.sukennyk@vikes.csuohio.edu

***Abstract*— Coloring images with deep learning frameworks is a uniquely challenging task which has no standard solution and many problem interpretations. Recent papers have been published reporting successes using approaches which frame colorization as a multimodal problem, with complex loss functions but relatively simple layering techniques. With the complexity of these models' training due to their custom loss functions, it is desirable to have a system which can transform pretrained colorization models to more accurately predict color in smaller datasets with unique or specific color distributions—tropical animals, historical footage, and others. This paper proposes and explores a GAN-based approach which directly transforms the *ab* color layer output of a pretrained colorization model using reference images before joining it with its light layer *L*.**

Keywords—GAN; Colorization; Pix2Pix; Model Adaptation

I. Introduction

The human mind trivializes grayscale colorization, as even untrained observers could reasonably color objects from memory. Some color choices are rigidly true—grass must be green, the sky must be blue; some color choices can be variable—cars can be many colors, but some are more likely than others. Historical photos are often recolored by artists using logical inferences to accurately discern what the colors must have been. The MHSWA has a collection of WWII photos with professional color restoration work for reference [1].

While it is easy for a human to guess expected colors in a gray image, it is challenging to define such a problem for automation with deep learning techniques. Other works refer to the process of extracting color from grayscale images as 'hallucination' [2, 3]. As mentioned, several color options may be plausible for a given object or set of pixels. In considering that each grayscale image has a single associated ground truth, there is a unique challenge in creating a model which can hallucinate *probable* data, rather than *exact* data. When generating an objective function—how do we translate the concept of 'plausible' results to a mathematical formulation? The experiments in this report will be built upon the work of Zhang et al. in *Colorful Image Colorization* (CIC) [2]. Their model was trained with an objective function that considers the color probability distribution for each pixel, rather than rigidly comparing it to the true color in training. Due to the model's success, it has been chosen as the baseline for the work in this report.

Section 2. discusses related work in colorizing grayscale images. The methodology of the CIC paper is described in Section 2.B. Section 3 provides an overview of improvements made to the given model. Section 4 describes the dataset used. Section 5 provides quantitative and qualitative results for our experiments. Sections 6, 7 summarize findings and propose future work which could prove beneficial for this niche.

II. Related Work

A. Historic Methods of Colorization

Larsson et al. [3] refer to 3 types of colorization models that have been developed: scribble-based, transfer, and automatic direct prediction.

Scribble-based methods require human color tagging to specify what color is desired in a given region. These solutions perform as a pick-and-paint coloring book for human labeling, which may be fun, but ultimately still require significant labor hours for large datasets for training or testing. This method appeared in papers from 2003 to 2007.

Transfer-based methods rely on abundant reference images from which to select colors. The models identify similar regions on which to draw the sampled colors. This method automates the color selection portion of Scribble-based methods, dramatically reducing the time cost of preparing datasets for training and testing. The reference images themselves must still be manually selected, to some degree. These methods struggle to consistently produce accurate results, oftentimes mismatching colors or disregarding fine details. This style of model was a significant improvement to other work at the time, being featured in several papers between 2005 and 2011.

The works of Larsson et al. and Zhang et al. both fall under the automatic direct prediction category, with substantial differences between them. Their results measure similarly by

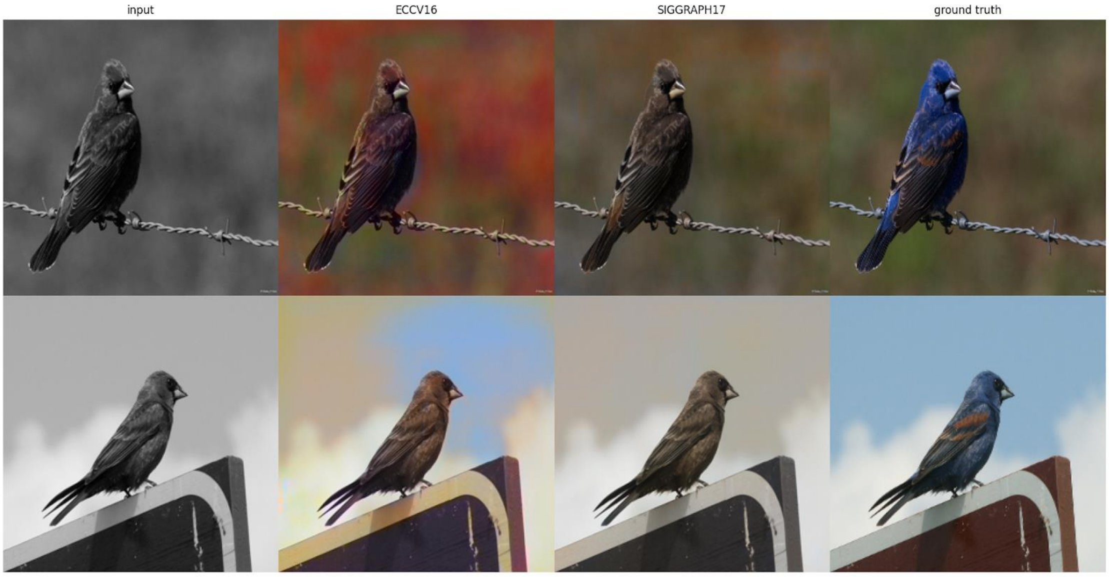

Figure 1 – Results of *Colorful Image Colorization* [2] model when passed two training images (bird species--Blue Grosbeak) from the CUB_200_2011 dataset [6].

quantitative metrics, but the CIC paper presents a model which has higher qualitative appeal, fooling more quizzed participants than the Larsson et. al. fully convolutional method [2]. Due to this advantage, as well as access to pretrained models, the CIC method has been selected for this project.

B. *Colorful Image Colorization*

The model created by Zhang et al. was generated by respecting the multimodal problem style of color prediction, an approach they credit to Charpait et al. in their ECCV 2008 paper—*Automatic image colorization via multimodal predictions* [4]. The model predicts a color probability range for each pixel, selecting the rarest possibility to encourage color into their model. According to their paper, Zhang et al. claim “this encourages our model to exploit the full diversity of the large-scale data on which it is trained.” [2].

The convolutional layers of the model are a set of standard layer expansions with normalization and dilated steps between, as shown in Figure 2. The input is the grayscale image as a lightness layer *L*, which is later combined with the generated *ab* color layer to recreate the images in the CIELAB or ‘*Lab*’ color space. The natural approach to calculating loss within these images is a summed Euclidean loss L2, but this alone will not account for the multimodality of the presented problem. Instead, Zhang et al. proposed an objective function which maps an input image set to a probability distribution of possible colors at given light levels. When training using this objective function, the model is able to assign probable colors to each pixel within a given context, selecting an option which is uncommon, and hence not low in saturation. This results in more vibrant, plausible, output images. In crowdsourced real vs. fake tests comparing two images, participants were fooled 32.3% of the time [2], an improvement to the 27.2% achieved by Larsson et al. [3] (compared to 50% max signifying total confusion).

III. Improvements

The model generated in CIC is groundbreaking and hallucinates realistic possible color images, but it is by definition incapable of generating the exact colors found in ground truth images where featured objects have diverse or improbable colors. In addition to struggling with artificial or man-made objects with uncharacteristic texture or color combinations, vibrantly colored animals–birds, fish, especially of tropical origins–are unable to be accurately colored. This issue is shown in Figure 1.

The first phase of improvements for this task is implementing a GAN model for post-processing the generated images to more closely match ground truths in a small dataset. In this project, the GAN was implemented closely following the Pix2Pix method by Isola et al. [5]. This involved creating a U-Net style generator, pictured in Figure 3, as well as a fully convolutional discriminator. With appropriate parameter setup, this method is adequate at changing a 3-layer image from one style to another. In this case, it will be used to translate from one pair of *ab* color input layers to another. Thus, provided sufficiently abundant training images, this method would effectively self-sample colors to transfer to the newly generated image.

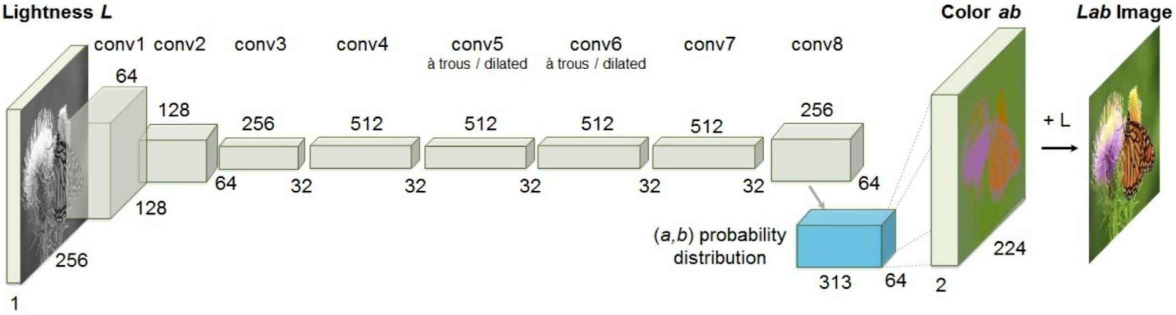

Figure 2 – Architecture of *Colorful Image Colorization* [2] model.

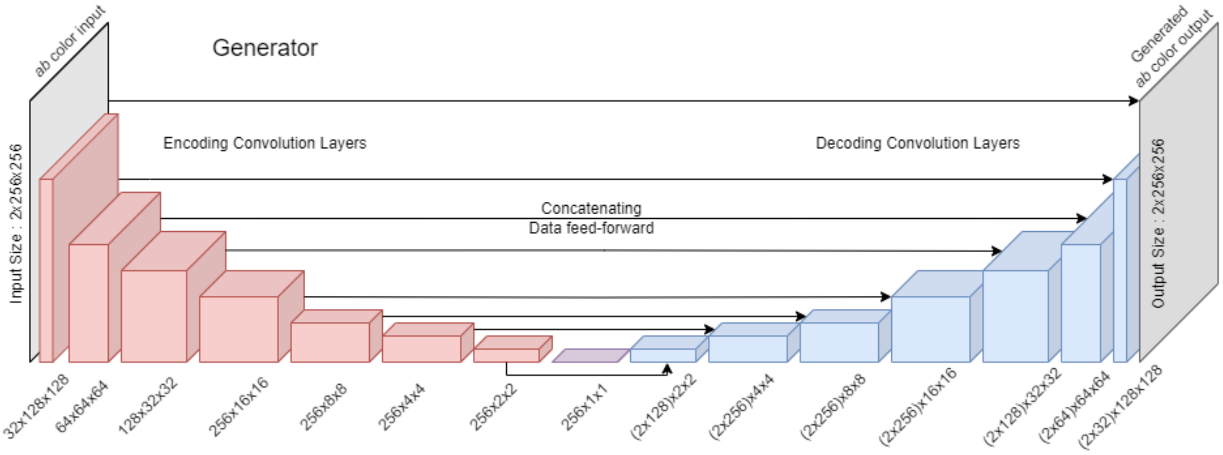

Figure 3 – Generator architecture used. Red blocks symbolize encoding convolution layers, blue symbolize decoding ones. Given sizes consider ground truth images of size 256x256. This model closely follows the U-Net architecture of the Pix2Pix method [5].

Given that our dataset provides a very small number of examples for a given class (see Section 4), the reference images must be transformed in various ways to provide additional knowledge to the GAN model. Once these transformed images are passed through the baseline CIC model, the GAN is trained on retouching their colors to more closely match the ground truth. This effectively extends the provided database and permits more knowledge gain per unique training image. In this first version it is assumed that the classification of birds is known, so uniquely colorful bird species are independently used for training and testing the effectiveness of this approach.

More improvements are possible, including segmenting objects of interest and classification extensions to this implementation. The remaining phases of improvements have not been completed within the timeframe allotted for this project, but a description of intended next steps is included within Section 7.A

IV. Dataset

The dataset chosen for experimentation is CUB_200_2011 [6], a dataset of 200 bird species with just under 12,000 images. This leaves about a mere 50 training and 10 testing images per bird species. This dataset provides a unique challenge with such a low number for each bird species: only a few effective training shots will be possible for the GAN on a given species. Hence, the colorizer must be efficient with its learning steps.

V. Experimental Results

Several loss functions were used for experimentation, with minimal differences and improvements between them. The custom loss function from the CIC Paper was not available in their provided code and proved too challenging to recreate. The Adam optimizer was used with a relatively high learning rate and no weight decay[1].

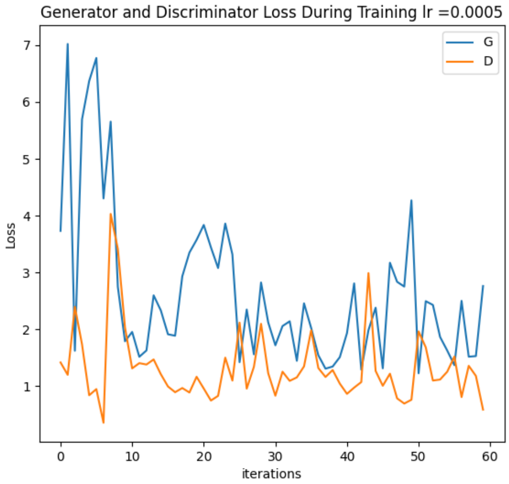

Figure 4 – Generator and Discriminator loss over 60 iterations.

Due to the nature of the GAN model, the Generator and Discriminator losses behave sporadically and do not smoothly converge to some minimum. The combative nature of optimizing two opposing models will result in sporadic and inversely correlated progressions in loss minimizing. This effect is apparent in Figure 4 above.

The best achieved results are shown in Figure 5. It is clear that these images were not colored in a way believable to human observers, hence no quantitative measurements were taken using any outsourced human-input testing. Particularly in the second example, it is apparent that some colorization has been learned by the model, but ultimately many more training iterations are necessary to further improve results. Changes to methodology would likely achieve more effective results with the same number of iterations. Section 7 will discuss what can be done to benefit the model further.

[1] learning_rate = 0.0005 chosen experimentally, betas = (0.5, 0.999) per original GAN paper [7]

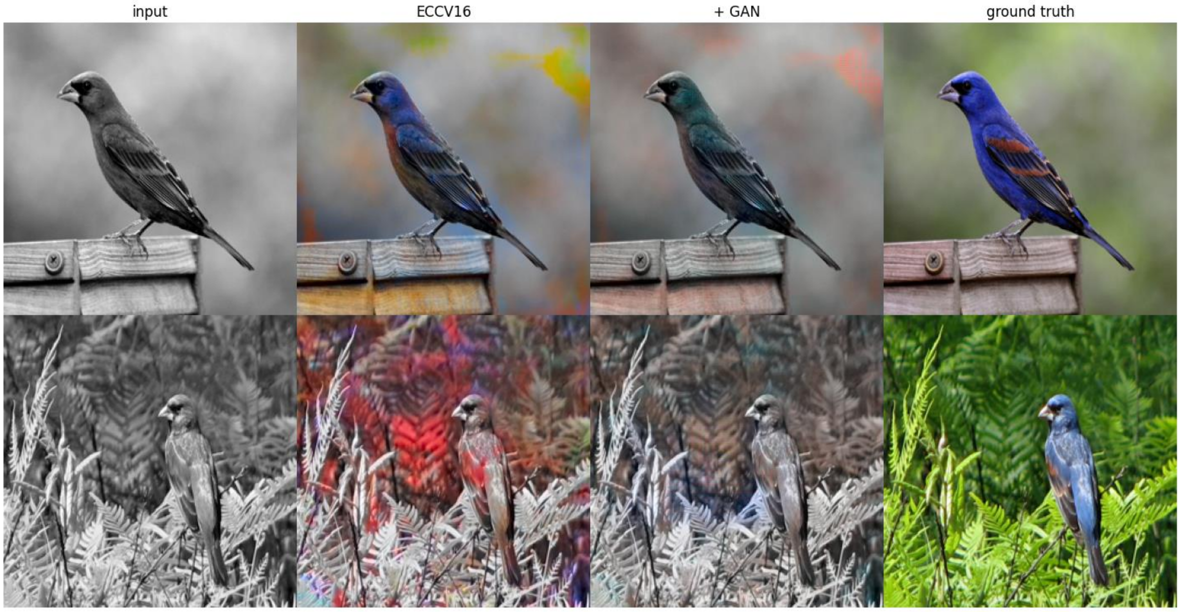

Figure 5 – Results of experiment using two training images (bird species--Blue Grosbeak)

VI. Conclusion

For the purpose of adapting existing models to specialized datasets, this approach shows promise. The current results may be unremarkable, and will not fool any human observer, but the splashes of blue visible within the generated images are evidence that the model has learned to transfer at least one key color from the training images. With some further adjustments, a viable system to adapt pretrained colorization models can be created. GAN model effectiveness appears to be limited for few-shot learning, but has provided minorly successful results even on an exceptionally small scale of training data.

VII. Future Work

Given more time for completion, many more steps can be taken to further improve this methodology, including expanding it to accompany multiple classifiers. The next two planned phases are described below, followed by additional possible applications for this method.

A. Incomplete Work

Due to a unique difficulty with provided preprocessing methods from the CIC model, spatial transformations to extend the learnability of the dataset was not completed within the time frame. Before the next steps are attempted, this must be completed if work is to continue such a limited dataset.

The second phase of improvements is to build upon the GAN method by first preprocessing the images and segmenting the birds, separating them from the background. This is to be done with an RCNN model or other segmentation method, labeling distinct birds in an image and separating them from the background. Given that a sufficiently accurate trace of the birds is generated, the GAN will be able to independently recolor the birds without affecting the background. Due to the nature of the *Lab* image encoding approach, this should only minimally affect the appearance of light and shadow on the subject(s) in relation to the background. Before attempting this phase, it would be beneficial to test the value of this improvement by using ground truth bounding boxes and segmentations to pass exclusively the pixels associated with the birds into the GAN stage. This would confirm the viability of segmentation for isolating the improvements and leaving the potentially well generated background alone. This background could also be isolated and 'touched up' with another model, if desired.

The final phase of improvements is to use RCNN for classification, or in tandem with another classifier model, to label the observed bird(s) from the grayscale image. This would permit training the GAN implementation on all available species in a single model. Given a classification, the GAN would be able to infer what color pallet should be used for the observed bird. This would permit generalized models to be built which handle several classifications of a single type of object—birds, lizards, fish, etc.—as well as colorize multiple types within the same image without any color cross-bleeding between them. In the best of cases, with a sufficiently large dataset, this method could improve the re-colorization process itself—specializing the convolutional layers enough to cross-reference between breeds, recognizing distinct body parts of the birds, such as their bellies, wings, beaks, heads, and other parts which can generally contain color differences.

B. Alternative Applications

It is important to note that while this problem may not appear practical, this model has feasible use in restoring low-information (grayscale) photos and videos. The database utilized in this project could be used to train this model for the purpose of recoloring gray trail camera footage from remote visual-sensing locations in national parks.

Additionally, this work could be used for the purpose of colorizing historical wartime photos while respecting ground truth color as to more accurately depict vehicle paint, uniforms, and flag colors. Similar work is done by Jin et al. [JIN] in a paper titled Focusing on Persons.

VIII. References

[1] "WWII Restored Photos – MHSWA." https://www.mhswa.org.au/wwii-photo-montage/ (accessed May 07, 2024).J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] R. Zhang, P. Isola, and A. A. Efros, 'Colorful Image Colorization', arXiv [cs.CV]. 2016.

[3] G. Larsson, M. Maire, and G. Shakhnarovich, 'Learning Representations for Automatic Colorization', *arXiv [cs.CV]*. 2017.

[4] Charpiat, G., Hofmann, M., Scholkopf, B.: Automatic image colorization via multimodal predictions. In: Computer Vision–ECCV 2008. Springer (2008) 126–139

[5] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, 'Image-to-Image Translation with Conditional Adversarial Networks', *arXiv [cs.CV]*. 2018.

[6] "Perona Lab - CUB-200-2011." *Www.vision.caltech.edu*, 2011, www.vision.caltech.edu/datasets/cub_200_2011/.

[7] Goodfellow, Ian, et al. *Generative Adversarial Nets*. 2014.

IX. Course-related Afterword

Due to unfortunate circumstances, this paper and the associated code were 'hastily' completed in less than a week. This is due to a late shift in topic as well as several limitations in data access. Having less than a week for work, the Dataset used here is the most readily accessible example where colorization needs to be accurate to the ground truth for specific classes. In discussion of future work, I mention that this method could be applied for a quick improvement to results on historical images such as the ones in the MHMD database – Found at https://github.com/BestiVictory/MHMD. Unfortunately, I was unable to get access to this database within the last week. Hence, colorful birds seemed the most reasonable alternative here. I hope that I get access to this dataset over the summer so that I can retest my methodology in a different, and maybe more practical, context.