Ain Shams University

Faculty of Engineering

Computer and Systems Department

CSE 616: Neural Networks and their applications

Project Report

Name	**Abdelrahman Ibrahim Yassin**
Code	**1902394**

Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2

Abstract. We review some of the most recent approaches to colorize gray-scale images using deep learning methods. Inspired by these, we propose a model which combines a deep Convolutional Neural Network trained from scratch with high-level features extracted from the InceptionResNet-v2 pre-trained model. Thanks to its fully convolutional architecture, our encoder-decoder model can process images of any size and aspect ratio. Other than presenting the training results, we assess the "public acceptance" of the generated images by means of a user study. Finally, we present a carousel of applications on different types of images, such as historical photographs.

1 Introduction

Coloring gray-scale images can have a big impact in a wide variety of domains, for instance, re-master of historical images and improvement of surveillance feeds. The information content of a gray-scale image is rather limited, thus adding the color components can provide more insights about its semantics. In the context of deep learning, models such as Inception [1], ResNet [2] or VGG [3] are usually trained using colored image datasets. When applying these networks on grayscale images, a prior colorization step can help improve the results. However, designing and implementing an effective and reliable system that automates this process still remains nowadays as a challenging task. The difficulty increases even more if we aim at fooling the human eye.

In this regard, we propose a model that is able to colorize images to a certain extent, combining a deep Convolutional Neural Network architecture and the latest released Inception model to this date, namely Inception-ResNet-v2 [4], which is based on Inception v3 [1] and Microsoft's ResNet [2,5]. While the deep CNN is trained from scratch, Inception-ResNet-v2 is used as a high-level feature

extractor which provides information about the image contents that can help their colorization.

Due to time constraints, the size of the training dataset is considerably small, which leads to our model being restricted to a limited variety of images. Nevertheless, our results investigate some approaches carried out by other researchers and validates the possibility to automate the colorization process.

2 Architecture

Our model owes its architecture to [16]: given the luminance component of an image, the model estimates its a*b* components and combines them with the input to obtain the final estimate of the colored image. Instead of training a feature extraction branch from scratch, we make use of an Inception-ResNetv2 network (referred to as Inception hereafter) and retrieve an embedding of the gray-scale image from its last layer. The network architecture we propose is illustrated in Fig. 1.

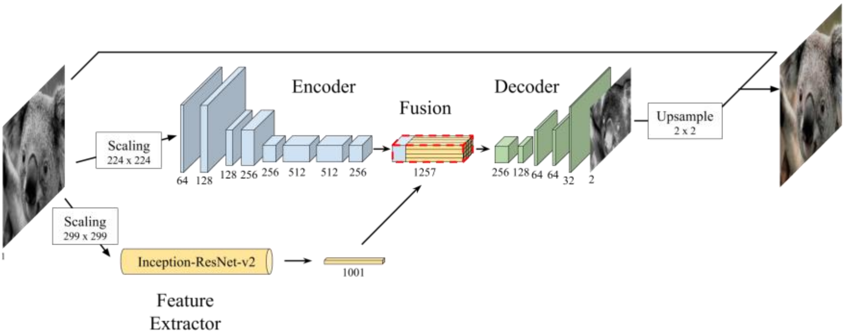


Fig.1: An overview of the model architecture.

The network is logically divided into four main components. The encoding and the feature extraction components obtain mid and high-level features, respectively, which are then merged in the fusion layer. Finally, the decoder uses these features to estimate the output.

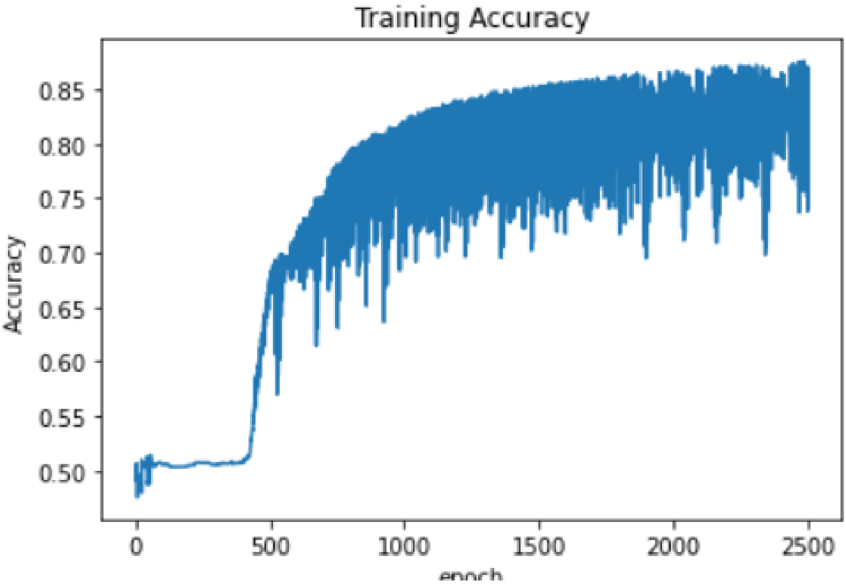
3 Experiments and Discussion

As important as the network's architecture itself is the choice of the dataset. In the majority of the approaches to automatic image recoloring so far, ImageNet has been extensively used [17][18]. Besides, ImageNet's impressive size (more than

14,000,000 images), extensive documentation and free access makes it appealing for our purpose. The dataset is composed of millions of pictures within a wide variety of sets. In particular, it is based on the *name* nodes contained in the word dataset WordNet. In order to simplify training and reduce running times, only a small subset of approximately 60,000 images is used.

ImageNet pictures are heterogeneous in shape, therefore all images in the training set are rescaled to 200 × 200 for the encoding branch input and to 299 × 299 for Inception. Each image gets stretched or shrunk as needed, but its aspect ratio is preserved by adding a white padding if needed.

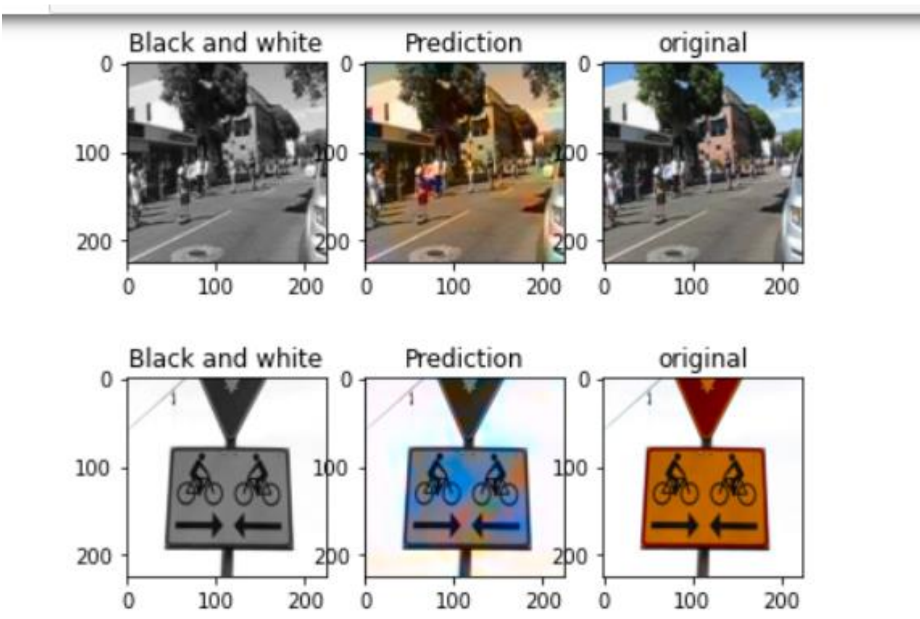
3.1 Results



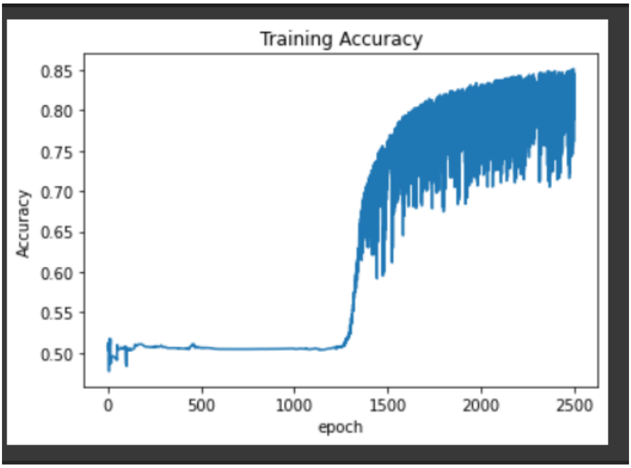
The previous figures show the training results on the previous mentioned dataset. The Training was done with the following parameters:

- Batch Size = **5**
- Epochs = **2500**

And in the following figure we can see the output of the model on the some of the image in the test data



After changing the batch size to **10**, we got the following results:



References

1. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. CoRR **abs/1512.00567** (2015)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR **abs/1512.03385** (2015)
3. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR **abs/1409.1556** (2014)
4. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning. CoRR **abs/1602.07261** (2016)
5. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. CoRR **abs/1603.05027** (2016)
6. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: ACM Transactions on Graphics (TOG). Volume 21., ACM (2002) 277–280
7. Ironi, R., Cohen-Or, D., Lischinski, D.: Colorization by example. In: Rendering Techniques, Citeseer (2005) 201–210
8. Charpiat, G., Hofmann, M., Scho¨lkopf, B.: Automatic image colorization via multimodal predictions. Computer Vision–ECCV 2008 (2008) 126–139
9. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM Transactions on Graphics (ToG). Volume 23., ACM (2004) 689–694
10. Qu, Y., Wong, T.T., Heng, P.A.: Manga colorization. In: ACM Transactions on Graphics (TOG). Volume 25., ACM (2006) 1214–1220
11. Sy`kora, D., Dingliana, J., Collins, S.: Lazybrush: Flexible painting tool for handdrawn cartoons. In: Computer Graphics Forum. Volume 28., Wiley Online Library (2009) 599–608
12. Deshpande, A., Rock, J., Forsyth, D.: Learning large-scale automatic image colorization. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 567–575
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105
14. Cheng, Z., Yang, Q., Sheng, B.: Deep colorization. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 415–423
15. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE (2008) 1–8
16. Iizuka, S., Simo-Serra, E., Ishikawa, H.: 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) (2016) 110
17. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: European Conference on Computer Vision, Springer (2016) 649–666
18. Larsson, G., Maire, M., Shakhnarovich, G.: Learning representations for automatic colorization. In: European Conference on Computer Vision, Springer (2016) 577– 593