**Image Color Conversion using Deep Neural Network**

**Team Tensors:**

Lalitha Amudalapally 012565540

Sowndhariya Nandarajkuamar 013761410

Yee Zhian Liew 013708240

**Base Paper :** <http://cs231n.stanford.edu/reports/2016/pdfs/219_Report.pdf>

**Introduction**

In image colorization, our goal is to produce a colored image given a grayscale input image. This problem is challenging because it is multimodal -- a single grayscale image may correspond to many plausible colored images. As a result, traditional models often relied on significant user input alongside a grayscale image.

Recently, deep neural networks have shown remarkable success in automatic image colorization going from grayscale to color with no additional human input. This success may in part be due to their ability to capture and use semantic information (i.e. what the image actually is) in colorization, although we are not yet sure what exactly makes these types of models perform so well.



Grey Image Target Colorized Image



For the dataset, we are using publicly available series of 1000 photos from open dataset and deep learning net website. The link is shown as below: .<https://skymind.ai/wiki/open-datasets> and <http://deeplearning.net/datasets/>. We are planning to split 70% in the training set and 30% in the the test set. Since we have not much of dataset, we might need to give up on validate set for now. These photos are each 256\*256 pixels and includes faces, photography, painting, animals and objects.To help with generalization, we also performed various transformations including zooms, flips and shears to prevent overfitting.

**Related Works**

In this paper, the author has build a convolutional neural network (CNN) that accepts a black-and-white image as an input and generates a colorized version of the image as its output. The system generates its output based solely on images it has “learned from” in the past, with no further human intervention.

During training time, our program reads images of pixel dimension 256 × 256 and 3 channels corresponding to red, green, and blue in the RGB color space. The images are converted to CIELUV color space. The black and white luminance L channel is fed to the model as input. The U and V channels are extracted as the target values.

During testing time, the model accepts a 256×256×1 black and white image. It generates two arrays, each of dimension 256 × 256 × 1, corresponding to the U and V channels of the CIELUV color space. The three channels are then concatenated together to form the CIELUV representation of the predicted image.

The author have successfully demonstrated that the potential of using deep convolutional neural networks to convert black and white images into color images. In short, the author have used classification model and a baseline regression-based model to make a comparison. What the author can concluded was the classification model did better than regression-based model as it produced much more aesthetically-pleasing images. Thus, the author is going to continue using classification model for further development. The author has listed one of the issue which is color inconsistency. He tried to consider incorporating segmentation to enforce uniformity in color within segments. Also, he recommended to utilize post-processing schemes such as total variation minimization and conditional random fields to achieve a similar end. Finally, he suggested in redesigning the system around an adversarial network. The author believed this kind of system may yield improved results, since instead of focusing on minimizing the cross-entropy loss on a per pixel basis, the system would learn to generate pictures that compare well with real-world images. Based on the output result, the author believed that his model would be a prime candidate for being the generator in such an adversarial network.

The author did mentioned that the main challenge of their model faces color is inconsistent for individual objects. The author illustrated one of the picture of a man wear a sweatshirt. The author is focusing on the sweatshirt’s color as the colored parts is red and other parts of it grey. In common sense, we can imagine a sweatshirt being red or grey as a whole. The author was hoping to improve their model by making one color prediction on each pixel, and hopefully the close-by pixels have similar color assignment.

The author has experimented with applying a smoothing technique which is called as Gaussian smoothing on the class scores in order to solve the issue. This technique has a better performance compared to the previous model. Unfortunately,it has addressed another new issue. For using this technique, visual noise is occured along the object edges. The author address this issue as under-colored problem. He mentioned that the reason the model was not able to give a good prediction for the sweater because of the wide range of color choices. Upon close examination, the author noticed that the model even painted part of it slightly green.

**Future Work**

In the future, there are many ways in which we could improve our model. It is better to train on more images by using deep neural network method as deep learning is meant to train on large dataset. We also can try on larger images as it would most likely lead to more interesting results.

The first thing we are going to do is to explore different types of loss functions. These loss function could better reflect the goal of our model, which was to produce realistic images. Moreover, new approaches would improve on the currently denatured images by penalizing the model less for picking the incorrect color.

We would also look into color rebalancing. Color prediction is inherently multimodal in that many objects can take on several plausible colorization. The next step, we are going to apply various deep learning algorithms such as Logistic Regression, R-CNN, GAN to compare the accuracy. We are going to decide using which algorithm best suites by confusion matrix and bleu values results. Then, we are implementing them in AUTOML H2o.ai for searching which model gives us the best accuracy. Finally, we will be applying these concepts to videos and accounting for consistent colorization across multiple frames.

**References**

1. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift.arXiv preprint arXiv:1502.03167, 2015.

2 .D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

3. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages

1097–1105, 2012.

4. A.OlivaandA.Torralba.Modeling the shape of the scene:Aholistic representation of the spatial envelope. International journal of computer vision, 42(3):145–175, 2001.

5. A. Olmos et al. A biologically inspired algorithm for the recovery of shading and reflectance images. Perception,33(12):1463–1473, 2004.

6. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh,S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3):211–252, 2015.