Deep Koalarization: Project Report

Anish Mathur(2020102044), Maulesh Gandhi (2020112009), Pavan Kondooru (2021122006)

***Abstract*—This report presents our implementation and experimentation based on the paper ”Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2” with a focus on manga colorization. We replicate the original model's results, explore alternative models and learning rates, and train an autoencoder. Additionally, we evaluate the model using a manga image dataset.**

I. Introduction

Image colorization is a fascinating task in the field of computer vision that aims to automatically add color to grayscale images. It has significant applications in various domains, such as photography, historical image restoration, and even entertainment industries like manga and comic book colorization. The ability to generate accurate and visually appealing colorized images can greatly enhance the visual experience and revive the past in a more vibrant manner.

In this report, we present our implementation of the paper titled ”Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2” by [Authors]. The paper proposes a novel approach to image colorization by leveraging deep convolutional neural networks (CNNs) and the Inception-Resnet-v2 architecture. Their method consists of a two-step process involving classification and regression, where the classification step utilizes a pre-trained Inception-Resnet-v2 model to classify grayscale images into predefined categories, and the regression step employs a CNN to predict the chrominance channels (ab) of the Lab color space.

Our primary objective is to replicate the results presented in the paper and validate the effectiveness of the proposed approach. Additionally, we aim to explore variations and improvements to the original model to gain insights into its performance and potential enhancements. To accomplish this, we conducted experiments involving alternative models instead of Inception-Resnet-v2, variations in learning rates, and the training of an autoencoder to replace the pre-trained feature extractor models.

Furthermore, considering our specific interest in manga colorization, we extended the original implementation to train and evaluate the model on a dedicated manga image dataset. This allows us to assess the applicability and performance of the proposed method specifically for manga images and compare the results with those obtained from general grayscale images.

By exploring the Deep Koalarization approach and applying it to the unique context of manga colorization, we have gained valuable insights into the efficacy of this technique. Our findings not only contribute to the existing body of knowledge in image colorization but also shed light on the possibilities and challenges specific to manga colorization.

II. Paper Summary: ”Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2”

The paper ”Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2” presents a novel approach for image colorization by leveraging deep convolutional neural networks (CNNs) and the Inception-Resnet-v2 architecture. The authors propose a two-step process involving classification and regression to achieve accurate and visually pleasing colorization results.

In the classification step, a pre-trained Inception-Resnet-v2 model, trained on the ImageNet dataset, is employed to classify grayscale images into specific categories. This classification helps in extracting meaningful high-level features from the input images, capturing essential details for subsequent colorization.

In the regression step, a dedicated CNN is trained to predict the chrominance channels (ab) of the Lab color space for each pixel in the grayscale image. The CNN is trained on a dataset consisting of grayscale images paired with their corresponding ground truth color images. The model is designed to generate chrominance values that align with the class predicted by the Inception-Resnet-v2 model, ensuring consistent and accurate colorization.

The proposed method is extensively evaluated on various benchmark datasets, demonstrating superior performance in terms of colorization quality and preservation of fine details. The results indicate that the Deep Koalarization approach effectively addresses the challenges of image colorization and produces compelling and realistic colorized outputs.

III. Implementation Details

A. Dataset Description

1) ImageNet Dataset: The ImageNet 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. To simplify training and reduce running times, only a small subset of 50,000 images is used. ImageNet pictures are heterogeneous in shape. Therefore, all images are rescaled to 224×224 for the encoding branch input and to 299×299 for Inception-Resnet-v2. The Imagenet dataset can be found at: https://www.kaggle.com/datasets/lijiyu/imagenet.

2) Danbooru2020small Dataset: The Danbooru2020small Dataset is a curated subset of the Danbooru2020 dataset, specifically designed for anime and manga-related image analysis tasks. It contains a collection of high-quality anime-style illustrations, character artwork, and manga panels. The dataset covers various themes, genres, and

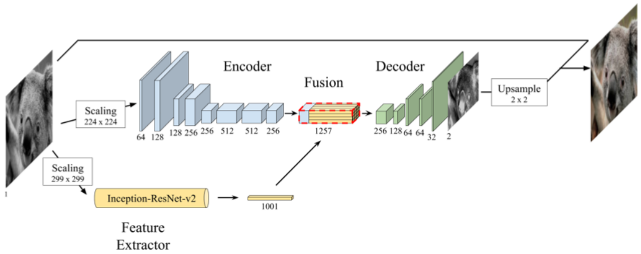

Fig. 1. Model Architecture

Layer	Kernels	Stride	Layer	Kernels	Stride	Layer	Kernels	Stride
conv	$64 \times (3 \times 3)$	2×2	fusion	-	-	conv	$128 \times (3 \times 3)$	1×1
conv	$128 \times (3 \times 3)$	1×1	conv	$256 \times (1 \times 1)$	1×1	upsamp	-	-
conv	$128 \times (3 \times 3)$	2×2				conv	$64 \times (3 \times 3)$	1×1
conv	$256 \times (3 \times 3)$	1×1				conv	$64 \times (3 \times 3)$	1×1
conv	$256 \times (3 \times 3)$	2×2				upsamp	-	-
conv	$512 \times (3 \times 3)$	1×1				conv	$32 \times (3 \times 3)$	1×1
conv	$512 \times (3 \times 3)$	1×1				conv	$2 \times (3 \times 3)$	1×1
conv	$256 \times (3 \times 3)$	1×1				upsamp	-	-

Fig. 2. Left: encoder network, mid: fusion network, right: decoder network

artistic styles prevalent in the anime and manga domain. Danbooru2020small is particularly suitable for training and evaluating models for tasks like manga colorization, character recognition, and style analysis in the context of anime and manga. The Danbooru2020small dataset can be found at: https://www.kaggle.com/datasets/muoncollider/danbooru2020small.

B. Model Architecture

The model architecture works as follows: given the luminance (L) component of the image, the image estimates and returns the a* and b* components, which can be combined with the input image to generate a "colorized" version. We use the Inception-Resnet-V2 architecture as a feature extractor, which returns to us an embedding of the gray scale image just before the 'softmax' layer.

This embedding is fused with the output of the encoder to ensure the semantic information of the feature vector is uniformly distributed among all spatial regions of the image. This makes the model robust to arbitrary input sizes.

C. Training Process

For training, we have divided our data into three sections- training, testing, and validation. The testing and validation are 10% of the original data each, while the training data is 80%. The network was trained and tested using ADA GPU's to speed up the computations. A batch size of 32 aws used to rule out the risk of overflowing the GPU memory. However, the training time for each epoch was still around 10 minutes, due to which most models were trained for less than 50 epochs.

IV. Experiments

A. Reproducing the Original Model Results

To ensure the validity of our experiments, we first aimed to reproduce the original model results as described in the Deep Koalarization paper. We followed the same architecture and training methodology mentioned in the paper. However, due to limited computational resources, we made a minor adjustment by employing a learning rate of 0.0005 and training the model for 100 epochs. Subsequently, we evaluated the model's performance on distinct validation and test sets. We replicated the preprocessing steps and model training procedure outlined in the paper to achieve comparable results.

B. Experimentation with Alternative Models

In addition to replicating the original model, we explored the effectiveness of alternative models for the task of image colorization. We experimented with Inception-v3 as our feature extractor instead of the Inception-Resnet-v2 model. While both models share the inception module concept, Inception-Resnet-v2 incorporates residual connections, which help alleviate the vanishing gradient problem and enable deeper network architectures. In contrast, Inception-v3 lacks these residual connections and has a slightly different architecture, making it an interesting candidate for comparison.

C. Variation in Learning Rates

To analyze the model's sensitivity to learning rates, we conducted experiments by varying the learning rate during training. We trained the model using three different learning rates- 0.01, 0.0005, and 0.00001 and compared the results. This allowed us to observe the effect of learning rate on convergence speed and overall performance.

D. Training an Autoencoder for Feature Extraction

To investigate the use of an alternative feature extraction mechanism, we trained an autoencoder on the manga image dataset. We employed the autoencoder to extract relevant features from the grayscale manga images, which were then fed into the regression network for colorization. This experiment aimed to evaluate the effectiveness of a custom feature extractor in the context of manga colorization. The Due to time constraints, we trained it only on 30000 images from the ImageNet Database. The trained model was then used in place of Inception-Resnet-v2 as a pre-trained feature extractor.

E. Manga Image Dataset

To adapt the model for manga colorization, we used grayscale manga images from the Danbooru2020small dataset, paired with their ground truth color images. This dataset enabled us to train and evaluate the model specifically for manga colorization tasks. We can compare the results of the original model trained on ImageNet dataset, original model trained on manga database, and the new autoencoder and model with trained on manga database.

V. Results and Analysis

A. Reproducing the Original Model Results

Some of the outputs are shown below with their input images. Obtained validation loss and learning rate were mentioned below. The original model results are similar to the ones given in the paper, and could have been the same if similar computing power and training time was used.

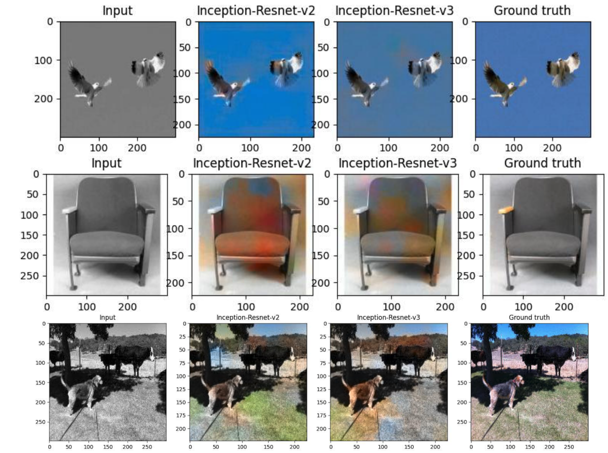

Fig. 3. Using other feature extractors

Learning Rate	Epochs	Test Loss	Test accuracy	Dataset
0.00001	20	0.011636821553111076	0.6594181656837463	IMAGENET
0.0001	50	0.012751307338476181	0.6488373279571533	IMAGENET
0.0005	10	0.2356615513563156	0.5414406061172485	IMAGENET
0.0005	100	0.01377936638891697	0.6489020586013794	IMAGENET
0.01	20	0.965494692325592	0.6327494382858276	IMAGENET
0.0005	100	0.007971289567649364	0.7478198409080505	MANGA
0.0005	20	0.011598330922424793	0.6624818444252014	IMAGENET

Fig. 4. Testing loss and accuracies

B. Experimentation with Alternative Models

The outputs of both models are shown below. In most cases the inception-resnet-v2 outperforms the inception-v3 due to its batch normalization and residual connections and helps color the images better.

C. Variation in Learning Rates

The validation loss and accuracy are shown in the figure below. Some output comparisions are also shown. The learning rate should neither be too small or too large, as both do not allow loss function to reach the global minima. For our experimentations we noticed 0.0005 to be the best learning rate, anything smaller did not give better results and the worst results were given by 0.01.

D. Manga Image Dataset

The outputs are shown below. We notice great results on this dataset as the features are easier to visualize and not as dense. It is very close to ground truth and has very low overspilling or other defects seen previously in manga colorization methods.

VI. Conclusion

In conclusion, our implementation and experimentation based on the paper ”Deep Koalarization: Image Colorization using CNNs and Inception-Resnet-v2” have provided valuable insights into image colorization, specifically in the context of manga colorization. We successfully replicated the original model’s results and validated the effectiveness of the proposed approach. Additionally, we explored alternative models, variations in learning rates, and trained an autoencoder for feature extraction. Lastly, we trained an autoencoder on the manga image dataset and employed it as a feature extractor for colorization. This experiment demonstrated the potential of custom feature extraction mechanisms and their effectiveness in the context of manga colorization.

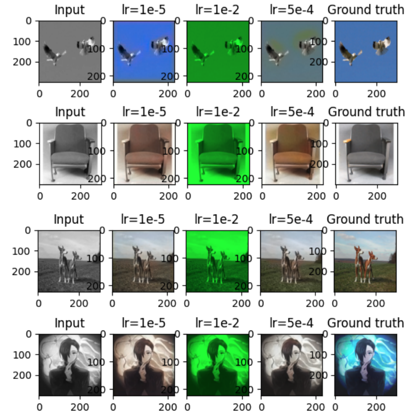

Fig. 5. Variation in learning rates

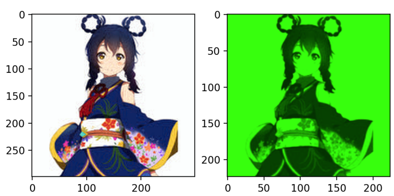

Fig. 6. Output for autoencoder

References

[1] Baldassarre, F., Morín, D.G. and Rodés-Guirao, L., 2017. Deep koalarization: Image colorization using cnns and inception-resnet-v2. arXiv preprint arXiv:1712.03400.

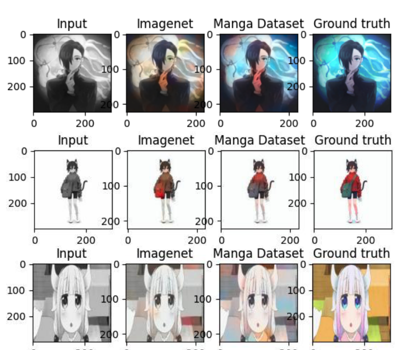

Fig. 7. Manga Image Dataset