Colorization of B/W Images & Videos Using Convolutional Neural Networks (CNN)

Submitted in partial fulfillment of the requirements

of the degree of

Bachelors of Engineering

in

Electronics and Telecommunication

by

Gori Shakeel (16ET15)

Kagdi Arbaz Altaf (16ET17)

Shaikh Touhid Alam (16ET29)

Sharma Nikhil (16ET30)

Under the guidance of

Dr. Mujib Abbas Tamboli

Department of Electronics and Telecommunication Engineering

Anjuman-I-Islam's Kalsekar Technical Campus

Sector 16, New Panvel, Navi Mumbai

University of Mumbai

2019-2020

CERTIFICATE

Department of Electronics and Telecommunication Engineering
Anjuman-I-Islam's Kalsekar Technical Campus
Sector 16, New Panvel , Navi Mumbai
University of Mumbai

This is to certify that the project entitled **Colorization of B/W Images & Videos Using Convolutional Neural Networks (CNN)** is a bonafide work of **Gori Shakeel (16ET15), Kagdi Arbaz Altaf (16ET17), Shaikh Touhid Alam (16ET29), Sharma Nikhil (16ET30)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of Bachelor of Engineering in Department of Electronics and Telecommunication Engineering.

Supervisor | Examiner

Head of Department | Director

Project Report Approval for Bachelor of Engineering

This project entitled **"Colorization of B/W Images & Videos Using Convolutional Neural Networks (CNN)" by Gori Shakeel (16ET15), Kagdi Arbaz Altaf (16ET17), Shaikh Touhid Alam (16ET29), Sharma Nikhil (16ET30)** is approved for the degree of **Bachelor of Engineering in Electronics and Telecommunication**.

Examiner

....

Supervisor

....

Date:

Place:

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

……………………

Gori Shakeel

(16ET15)

……………………

Kagdi Arbaz Altaf

(16ET17)

……………………

Shaikh Touhid Alam

(16ET29)

……………………

Sharma Nikhil

(16ET30)

Acknowledgments

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

We are highly indebted to **Dr. Mujib Abbas Tamboli** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express my gratitude towards my parents & Staff of Anjuman-I-Islam's Kalsekar Technical Campus for their kind co-operation and encouragement which help me in completion of this project.

We My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

Gori Shakeel (16ET15)
Kagdi Arbaz Altaf (16ET17)
Shaikh Touhid Alam (16ET29)
Sharma Nikhil (16ET30)

Abstract

The proposed approach present a novel technique to automatically colorize grayscale images that combines both global priors and local image features. Based on Convolutional Neural Networks, our deep network features a fusion layer that allows us to elegantly merge local information dependent on small image patches with global priors computed using the entire image. The entire framework, including the global and local priors as well as the colorization model, is trained in an end-to-end fashion. Furthermore, our architecture can process images of any resolution, unlike most existing approaches based on CNN. We leverage an existing large-scale scene classification database to train our model, exploiting the class labels of the dataset to more efficiently and discriminatively learn the global priors. We validate our approach with a user study and compare against the state of the art, where we show significant improvements. Furthermore, we demonstrate our method extensively on many different types of images, including black-and-white photography from over a hundred years ago, and show realistic colorization.

The final classification-based model we build generates colorized images that are significantly more aesthetically-pleasing than those created by the baseline regression-based model, demonstrating the viability of our methodology and revealing promising avenues for future work.

Contents

Project I Approval for Bachelor of Engineering ii
Declaration iii
Table of Contents vi
List of Figures vii
List of Tables vii
Keywords vii

1 Introduction **1**

1.1 Statement of Project 1

1.1.1 Project Architecture 1

1.1.2 Motivation 2

1.2 Objective and Scope 2

2 Literature Review **3**

2.1 Paper Title 3

2.1.1 Weaknesses 4

2.1.2 How to Overcome 4

3 Technical Details **5**

3.1 Methodology 5

3.2 Project Requirements 10

3.2.1 Software Requirements 10

3.2.2 Hardware Requirements 10

4 Market Potential **11**

4.1 Market Potential of Project 11

4.2 Competitive Advantages of Project 12

5 Working Phase **13**

5.1 Building Datasets. 13

5.2 Model Building and Training. 14

5.3 Model Testing and Validation… 16
5.4 Model Interface. 17
5.5 Difficulties Faced and Lesson Learnt. 18
5.6 Result. 19

6 Conclusion and Future Scope **21**
6.1 Conclusion 21
6.2 Future Scope. 22

References **23**

List of Figures

1.1 Color Inconsistency 4
1.2 Overview of Model 6
1.3 Baseline Regression 9
1.4 Classification Model. 10

List of Tables

FIGURE NO.	FIGURE IMPLICATION	PAGE
1	Global Features Network	06
2	Low level Features Network	07
3	Mid-level Features Network	08
4	Colorization Network	08

Keywords

Keywords : CNN, Convolutional Neural Network, Computer Vision, Neural Network, Deep Learning, Image Processing, Colorization.

Chapter 1

Introduction

1.1 Statement of Project

Traditional Colorization Techniques Requires Significant Human Intervention & Lots of Manual Data Feeding. However, obtained output is not Satisfactory. We Attempt to build a model using Multiple Convolutional Neural Networks and Fuse them together to perform final Colorization & Classification. Proposed Model Easily Outperforms the Traditional Approach. Proposed Model overcomes the limitations by previous model.

1.2 Project Architecture

Proposed approach is based on deep Convolutional Neural Networks that have been proven able to learn complex mappings from large amounts of training data. the network is formed by several subcomponents that form a Directed Acyclic Graph (DAG) and contain important discrepancies with widely-used standard models. In particular, our model:

• Can process images of any resolution,

• Incorporates global image priors for local predictions, and

• Can directly transfer the style of an image into the colorization of another.

It consists of four main components: a low-level features network, a mid-level features network, a global features network, and a colorization network. The components are all tightly coupled and trained in an end-to-end fashion. The output of our model is the chrominance of the image which is fused with the luminance to form the output image.

Proposed approach uses a regression-based model similar to the model described in as our baseline. this architecture is defined as comprising a “summarizing”, encoding process on the left side followed by a “creating”, decoding process on the right side. The architecture of the leftmost column of layers is inherited from a portion of the VGG16 network. During this “summarizing” process, the size (height and width) of the feature map shrinks while the depth increases. As the model forwards its input deeper into the network, it learns a rich collection of higher-order abstract features.

1.3 Motivation

Automated colorization of black and white images has been subject to much research within the computer vision and machine learning communities. Beyond simply being fascinating from an aesthetics and artificial intelligence perspective, such capability has broad practical applications ranging from video restoration to image enhancement for improved interpretability.

Neural Networks is currently in its exploratory stage lots of research is carried out to find a practical application of Neural Networks and Computer Vision for greater benefits and paving a new way towards technological advancement.

1.4 Objective

- Current model will only be able to colorize images that share common properties with those in the training set. if the input images is insignificant of those in training images, the models fails.
- To overcome limitations of both Research Papers.
- To build a model that gives the colorized output irrespective of similarities between input images & training images.

1.4 Scope

- The proposed work will unlock the possibilities for future research in convolutional neural network. For Image classifications with high accuracy it can also be used for crime scene detection for appropriate evidence.
- Implementation of convolutional Neural Network for image classification. It can also be used for diseases detection with accuracy. Our proposed project will serve as a base for further research in the field.
- Exploring the aspects of applications of Neural Networks more precisely Convolutional Neural Networks. which will be valuable for understanding more effective applications in future.

Chapter 2

Literature Review

2.1 Paper Title: A) Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification.

Introduction:

This research paper present a novel technique to automatically colorize grayscale images that combines both global priors and local image features. The Traditional colorization requires significant user interaction whether in the form of placing numerous color scribbles, looking at related images, or performing segmentation. In this paper, we instead propose a fully automated data-driven approach for colorization of grayscale images. This approach is based on Convolutional Neural Networks, which have a strong capacity for learning. This model consists of four main components: a low-level features network, a mid-level features network, a global features network, and a colorization network. Conceptually, these networks function as follows: First, a common set of shared low-level features are extracted from the image. Using these features, a set of global image features and mid-level image features are computed. Then, the mid- level and the global features are both fused by our proposed "fusion layer" and used as the input to a colorization network that outputs the final chrominance map. We train and evaluate our model on a large-scale scene database.

2.1.1 Weaknesses:

- The main weakness of the method lies in the fact that it is data driven and thus will only be able to colorize images that share common properties with those in the training set.

2.1.2 How to Overcome:

- To overcome the weakness (1) we will train our model with a very large diverse set of both indoor and outdoor scene images.
- To overcome the limitations baseline regression is used from [2]

2.2 Project Title: B) Image Colorization with Deep Convolutional Neural Networks.

Introduction:

This research paper present a convolutional-neural-network-based system that faithfully colorizes black and white photographic images without direct human assistance. In this a statistical-learning-driven approach is taken for solving the problem. In this method they design and 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.

2.2.1 Weaknesses:

- The main challenge this model faces is inconsistency in colors within individual objects.

Figure 7. Sample outputs exhibiting color inconsistency issue.

- The another weakness is that in a man made objects with a large intra domain color variation are generally more challenging for this model.

2.2.2 How to overcome:

- To overcome this model is trained with both global and local features.

Chapter 3

Technical Details

3.1 Methodology

Proposed approach is based on deep Convolutional Neural Networks [Krizhevsky et al. 2012] that have been proven able to learn complex mappings from large amounts of training data. Our network is formed by several subcomponents that form a Directed Acyclic Graph (DAG) and contain important discrepancies with widely-used standard models. In particular, our model:

- can process images of any resolution,
- incorporates global image priors for local predictions, and
- can directly transfer the style of an image into the colorization of another

It consists of four main components: a low-level features network, a mid-level features network, a global features network, and a colorization network. The components are all tightly coupled and trained in an end-to-end fashion. The output of our model is the chrominance of the image which is fused with the luminance to form the output image.

Deep Network

Deep networks are neural networks that are formed by many layers. These networks serve to predict continuous values from a given input. They consist of layers that realize a function of the form:

$$y = \tilde{\sigma}(b + Wx), \quad \ldots\ldots\ldots\ldots (1)$$

where $x \in Rn$ and $y \in Rm$ are the input and output of the layer, respectively, W is an m-by-n matrix of weights, $b \in Rm$ is a bias vector, and $\tilde{\sigma} : Rm \rightarrow Rm$ is a non-linear transfer function $\sigma : R \rightarrow R$ applied component wise. Both the weights and the bias are learnt through back-propagation [Rumelhartetal.1986], which consists of using the chain rule to propagate a loss to update the parameters. The loss consists in the error between the prediction of the network and the training data ground truth.

Fusing Global and Local Features for Colorization

We use a novel approach to fuse both global and local features together. The global features act as an image prior on the local features to indicate what type of image the input is. For example, if the global features indicate that it is an indoor image, the local features will be biased to not attempt to add sky colors or grass colors to the image, but instead will add colors suitable for furniture. We intertwine both a global image feature network, similar to those that compete in image classification tasks, with a fully convolutional neural network that colorizes the image. In order to improve the model efficiency, both networks use a number of common shared low-level features.

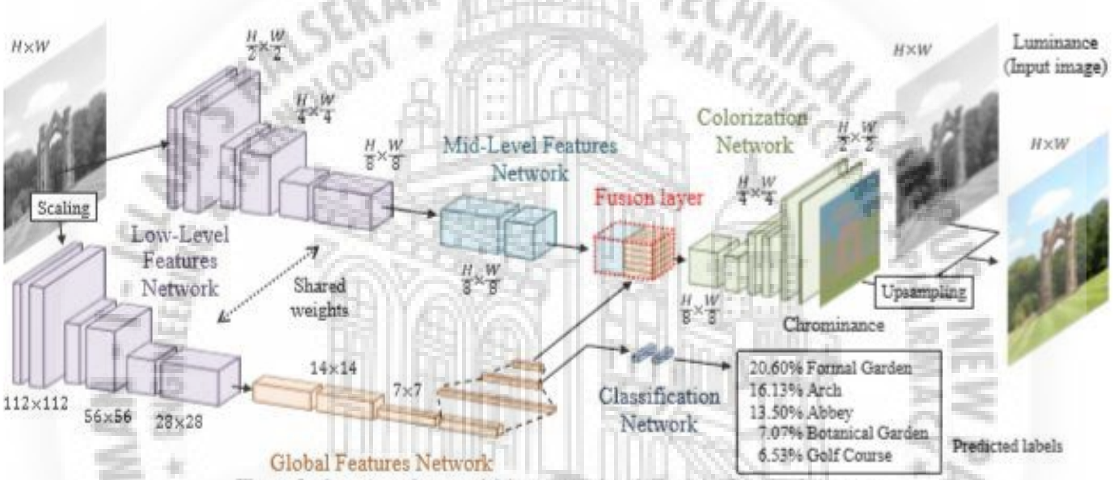


Figure 2: *Overview of our model for automatic colorization of grayscale images.*

Global Image Features

The global image features are obtained by further processing the low-level features with four convolutional layers followed by three fully-connected layers. This results in a 256-dimensional vector representation of the image. The full details of the global image features network can be seen in Table 1-(b). Note that due to the nature of the linear layers in this network, it requires the input of the low-level features network to be of fixed size of 224×224 pixels. However, this limitation does not affect the full approach.

Type	Kernel	Strides	Outputs
Conv	3x3	2x2	512
Conv	3x3	1x1	512
Conv	3x3	2x2	512

Conv	3x3	1x1	512

Low Level Features

A 6-layer Convolutional Neural Network obtains low-level features directly from the input image. The convolution filter bank the network represents are shared to feed both the global features network and the mid-level features network. This is similar to the weight sharing in Siamese networks [Bromleyetal.1994], however, in our approach only a subset of the full network is shared. Instead of using max-pooling layers to reduce the size of the feature maps, we use convolution layers with increased strides. This is also important for increasing the spatial support of each layer. A stride of 2 indicates that instead of computing Eq. (2) for consecutive pixels, every other pixel is computed. If padding is added to the layer, the output is effectively half the size of the input layer. This can be used to replace the max-pooling layers while maintaining performance [Springenbergetal.2015]. Weuse3×3convolutionkernels exclusively and a padding of 1×1 to ensure the output is the same size (or half if using a stride of 2) as the input.

Type	Kernel	Strides	Outputs
Conv	3x3	2x2	64
Conv	3x3	1x1	128
Conv	3x3	2x2	128
Conv	3x3	1x1	256
Conv	3x3	2x2	256
Conv	3x3	1x1	512

Mid-Level Features

The mid-level features are obtained by processing the low-level features further with two convolutional layers. The output is bottle

necked from the original 512-channel low-level features to 256channel mid-level features. Note that unlike the global image features, the low-level and mid-level features networks are fully convolutional networks, such that the output is a scaled version of the input.

Type	Kernel	Strides	Outputs
Conv	3x3	2x2	512
Conv	3x3	1x1	256

Colorization Network

Once the features are fused, they are processed by a set of convolutions and up sampling layers, the latter which consist of simply up sampling the input by using the nearest neighbor technique so that the output is twice as wide and twice as tall. These layers are alternated until the output is half the size of the original input. The output layer of the colorization network consists of a convolutional layer with a Sigmoid transfer function that out puts the chrominance of the input grayscale image. The architecture can be seen in Table 1-(d). Finally, the computed chrominance is combined with the input intensity image to produce the resulting color image. In order to train the network, we use the Mean Square Error (MSE) criterion. Given a color image for training, we convert the image to grayscale and CIE L*a*b* color space. The input of the model is the grayscale image while the target output is the a*b* components of the CIE L*a*b* color space. The a*b* components are globally normalized so they lie in the [0,1] range of the Sigmoid transfer function. We then scale the target output to the size of the output of the colorization network and compute the MSE between the output and target output as the loss. This loss is then back-propagated through all the networks (global features, mid-level features and low-level features) to update all the parameters of the model.

Type	Kernel	Strides	Outputs
Fusion	----	----	256
Conv	3x3	1x1	128
Up sample	----	----	128
Conv	3x3	1x1	64
Conv	3x3	1x1	64
Up sample	----	----	64
Conv	3x3	1x1	32

Output	3x3	1x1	2

Baseline Regression Model

A regression-based model similar to the model described in as our baseline. Figure 2 shows the structure of this baseline model. We describe this architecture as comprising a "summarizing", encoding process on the left side followed by a "creating", decoding process on the right side The architecture of the leftmost column of layers is inherited from a portion of the VGG16 network. During this "summarizing" process, the size (height and width) of the feature map shrinks while the depth increases. As the model forwards its input deeper into the network, it learns a rich collection of higher-order abstract features.

For the objective function in our system, we considered several loss functions. We began by using the vanilla `2 loss function. Later ,we moved onto deriving a loss function from the Huber penalty function, which is defined as

$$`L(u) = \{u^2 \quad |u| < M$$
$$M(2|u|-M) \quad |u| > M\}$$

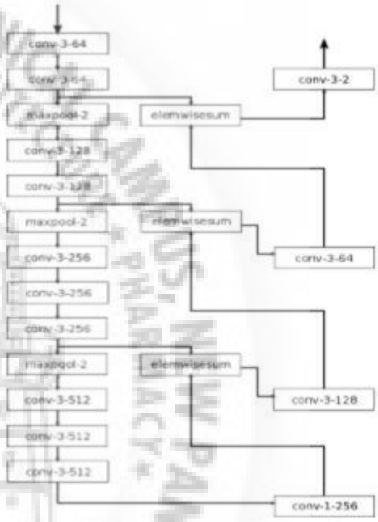
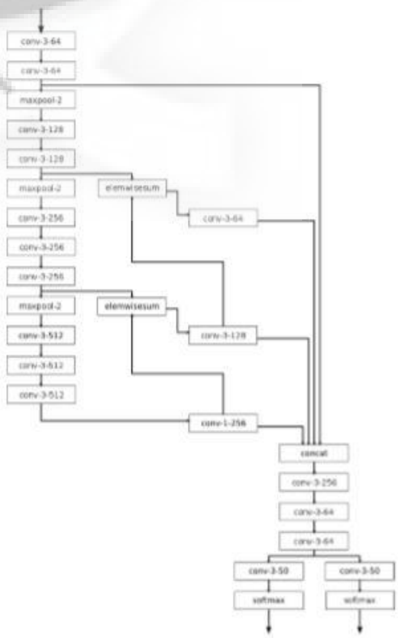


Figure 2. Regression network schematic.

Final Classification Model

Above Figure depicts a schematic of our final classification model. The regression model suffers from adimmingproblem because it minimizes some variant of the `p norm, which motivates the model to choose an average or intermediate color when multiple distinct color choices are possible. To address this issue, we remodeled our problem as a classification problem. In order to perform classification on continuous data,we must discretize the domain. The targets U and V from the CIELUV color space take on values in the interval [−100,100].



3.2 Project Requirements

Proposed project requires millions of images for training the model. which requires high computational power and it is not feasible with CPU.

3.2.1 Software Requirements

- Anaconda (Spyder)
- Caffe Framework
- Tensorflow
- Keras
- MIT Places Scene Dataset

3.2.2 Hardware Requirement

- Personal Computer with GPU
- Intel Movidius Neural Compute Stick 2

Chapter 4

Market Potential

4.1 Market Potential of Project

Market potential is the entire size of the market for a product at a specific time. It represents the upper limits of the market for a product. Market potential is usually measured either by sales value or sales volume

Colorization of Black and White Images will be beneficial in today's platform as it would have multiple applications in various domains. The technique could be used anywhere from usage at school or in companies etc. Google, Samsung and Adobe are also implementing their versions of colorization model based on AI technology. Once the model is ready and operational it can be used in Smartphones and tablets, it can also help in retrieving information from old images.

Previous methods for image colorization either:

1. Relied on significant human interaction and annotation
2. Produced desaturated colorization

An Earlier project was based on a research work developed at the University of California, Berkeley by Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful Image Colorization where the idea was to develop a fully automatic approach that will generate realistic colorizations of Black & White (B&W) photos and by extension, videos. As explained in the original paper, the authors, embraced the underlying uncertainty of the problem by posing it as a classification task using class-rebalancing at training time to increase the diversity of colors in the result. The Artificial Intelligent (AI) approach is implemented as a feed-forward pass in a CNN (" Convolutional Neural Network") at test time and is trained on over a million color images.

Researchers from University of California, Berkeley developed an interactive deep learning-based app that makes it easy to accurately colorize a black and white image in minutes. Building on the researcher's previous work of a convolutional neural network automatically adding color to black and white photos, their new app uses the same process, but with the addition of user-guided clues and hints to produce more realistic results.

4.2 Competitive Advantages of Project.

The testing performed by Model has demonstrated the efficiency and potential of using deep convolutional neural networks to colorize black and white images. In particular, it has been seen that formulating the task as a classification problem can yield colorized images that are much more aesthetically-pleasing than those generated by a baseline regression-based model, and thus shows much promise for further development.

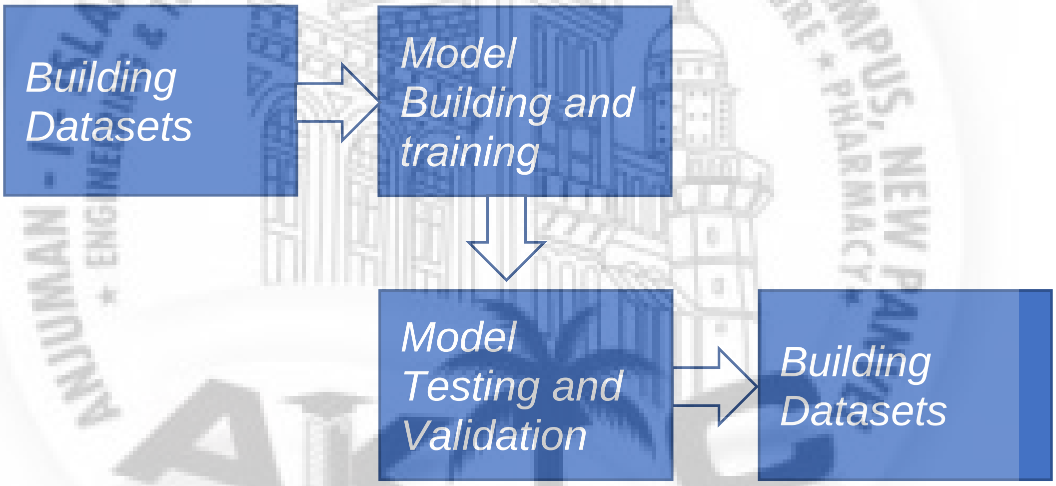
This Project therefore lays a solid foundation for future work. Moving forward, several avenues for improving our current system have been noticed. To address the issue of color in consistency, we can consider incorporating segmentation to enforce uniformity in color within segments. Post-processing schemes can also be utilized, such as total variation minimization and conditional random fields to achieve a similar end.

Finally, it could be interesting to apply colorization techniques to video sequences, which could potentially re-master old documentaries. This, of course, would require adapting the network architecture to accommodate temporal coherence between subsequent frames.

Chapter 5

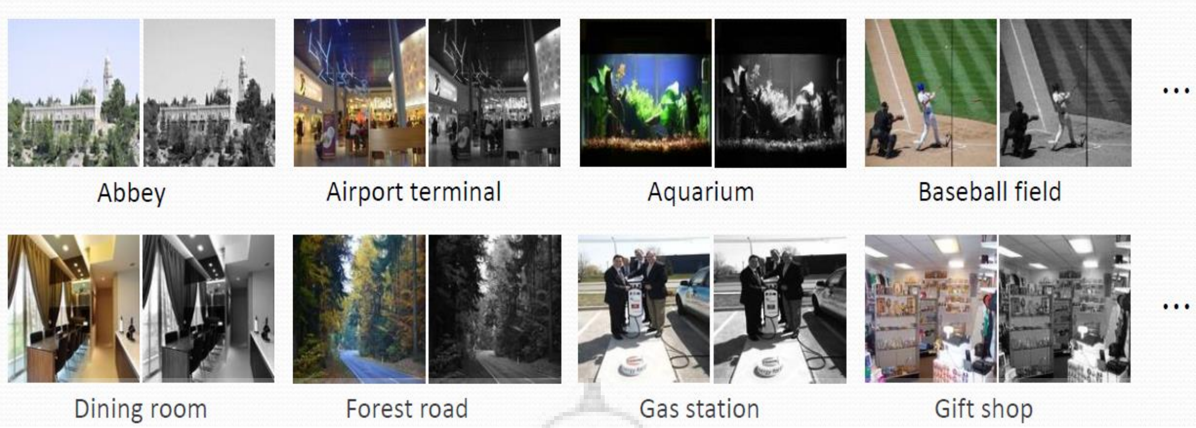
Working Phase

The working phase is divided into four phases. In the first phase we will collect all the datasets from different sources and the data sources are in large numbers. After collecting the datasets we will create a model in such a way that it will give result correctly without any errors. After that we will train our model according to the result. Then we will test the model with different datasets and after that we will interface the model. Below is the flow chart explaining the working phase.

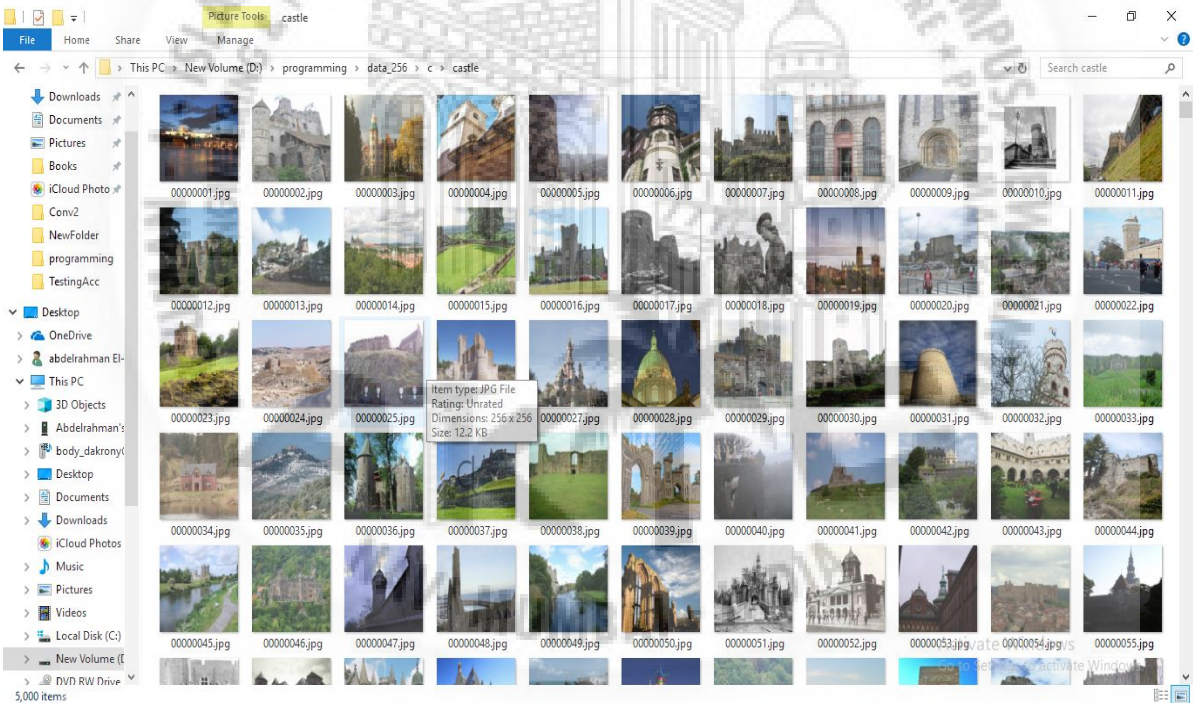


5.1 Building Datasets

The building datasets for this model is taken from the MIT Places Scene Dataset. The data sets are in millions in number . The datasets are nothing but the images in the size of 256 x 256 RGB. The length of the datasets is 24GB.



In the dataset there are different different images such as the image of roads , sea ,gardens , towers etc. But in our model from millions of datasets (Images) we have used only 15 thousands pictures



5.2 Model Building and Training

The model is consist of different networks (the networks are nothing but parts of the model) namely

1. Low Level Features Network
2. Mid Level Features Network

3. Global Features Network
4. Colorization Network
5. Classification Network

Now we will see the different networks one by one

1. **Low Level Features Network**

The low level features network consist of a 6 layer Convolutional Neural Network obtains directly from the input image. The layers are shared to feed both the global features network and the mid level features network. In this approach only a subset of the full network is shared.

2. **Mid-Level Features Network**

The mid level features are obtained by processing the low level features further with two layer convolutional layers. The output is bottlenecked from the original 512-channel low level features to 256 channel mid level features . The output is scaled version of input unlike the low level features network.

3. **Global Features Network**

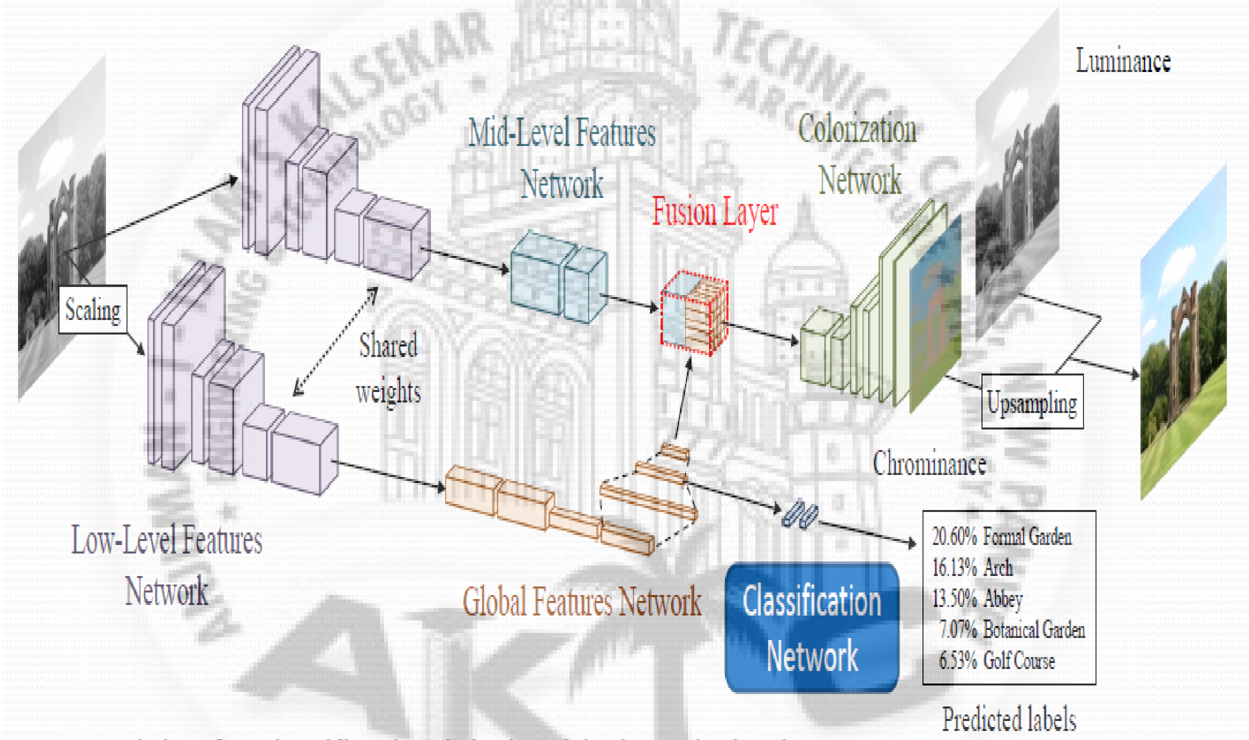
The global image features are obtained by further processing the low features with four convolutional layers followed by three fully connected layers. The results in a 256 dimensional vector representation of the image. Due to the nature of the linear layers in the network , it requires the input of the low level features network to be of fixed size of 224 x 224 pixels.

4. **Colorization Network**

Once the features are fused , they are processed by a set of convolutions and upsampling layers , the letter which consist of simply upsampling the input by using the nearest neighbor technique so that the output is twice as wide and twice as tall. The output layer of colorization network consist of convolutional layer with a Sigmoid transfer function that outputs the chrominance of the input gray scale image. And finally the compound chrominace image is combined with the input intensity image to produce the resulting color image.

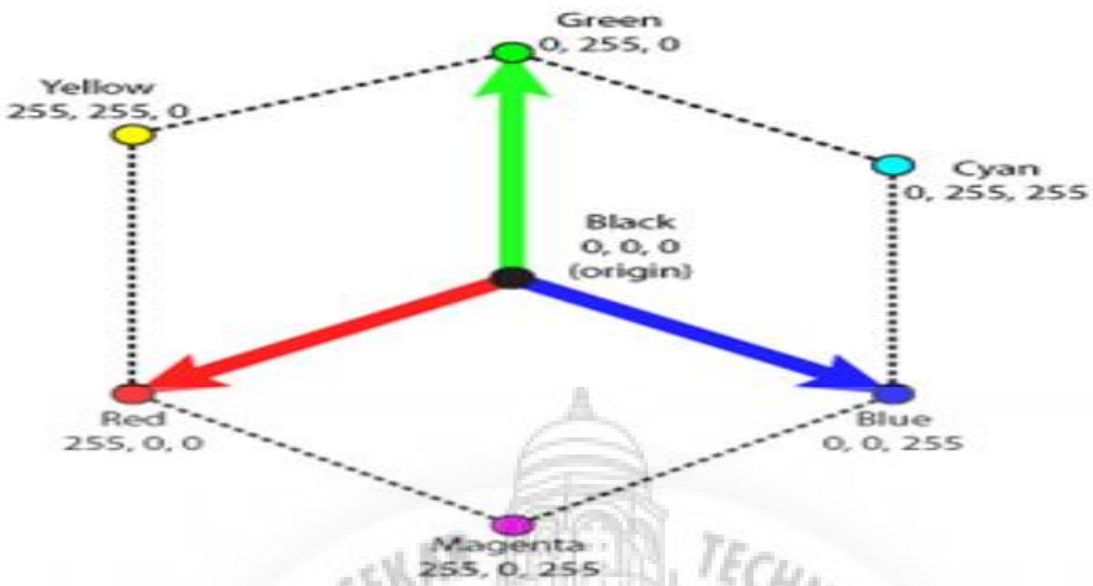
5. Classification Network

The classification network will classify the image into different parts for example if we took an image it has different objects such as sky , gardens , pillars etc. Thus the classification network will classify the this object and will try to colorize this one by one using the datasets files then it will predict the labels.



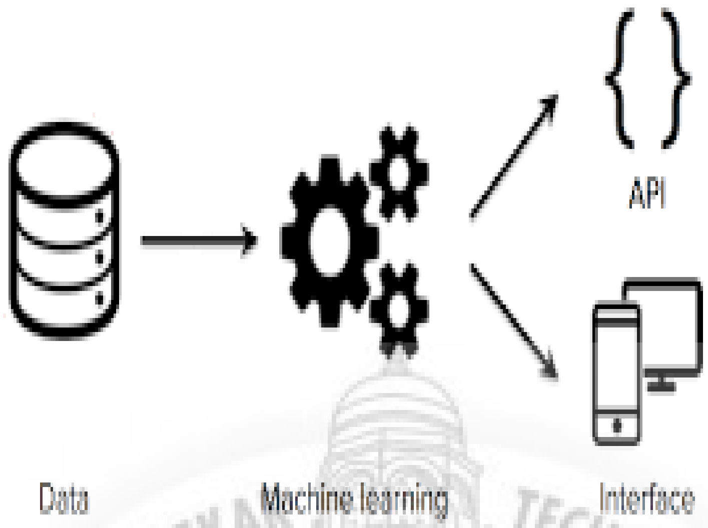
5.3 Model Testing and Validation

The accuracy is calculated by the calculating the Euclidean distance between the predicted and the ground truth . Here the color channel is represented by a dimension as you can see in the below image.



5.4 Model Interface

In this we built a web application by using Python-Flask and Micro Service Architecture . The model of this API . This application can be accessible from any device



5.5 Difficulties Faced & Lesson Learnt

- Training the model was hard because of limited hardware supply
- Working with new packages/technologies we are not familiar like tensorflow,opencv,PIL image
- Taking the Machine Learning model to production,using new technologies (Micro service architecture,Flask API)
- Learning very deep networks such as the one proposed directly from a random initialization is a very challenging task

5.6 Results

1. Below chart shows the Model Training Parameter

Experiment	Training					Results	
	Training dataset size	Optimizer	Batch Size	Learning Rate	Epochs	Test dataset size	Accuracy
Experiment 1	10000	Adam	10	0.00001	10	1000	68.96659101
Experiment 2	3500	Adam	5	0.00001	30	1000	78.6650432
Experiment 3	8500	Adam	20	0.00001	46	1000	82.59435356
Experiment 4	15000	Adam	25	0.00001	59	2000	87.12135776

2. This is the Grayscale Images which we are taking as the Input Image

3. After Colorizing the Image we got the following Output as shown below

4. And this is the Original Image as shown Below

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

The Proposed architecture for the colorization of grayscale images by fusing both global and local information. Our approach is based on convolutional neural networks and is able to perform the colorization without any user intervention. We train our model end-to-end on a large dataset for scene recognition with a joint colorization and classification loss that allows it not only to understand colors, but also adapts the colors to the context of the image, i.e., the sky color in a sunset image is not the same as in a day light image. Our architecture allows us to process images of any resolution, unlike most deep-learning frameworks. Furthermore, we show that with the same model we can do style transfer, that is, color an image using the context of another. Finally, we evaluated our model on a large diverse set of both indoor and outdoor images and showed that it can produce very credible results. We compared against the state of the art and also performed a user study that corroborates the results. Our approach runs in near real-time and has many potential applications such as automatic colorization of historical photography and movie archives.

This work therefore lays a solid foundation for future work. Moving forward, we have identified several avenues for improving our current system. To address the issue of color inconsistency, we can consider incorporating segmentation to enforce uniformity in color within segments. We can also utilize post-processing schemes such as total variation minimization and conditional random fields to achieve a similar end.

6.2 Future Scope

- The proposed work will unlock the possibilities for future research in convolutional neural network. For Image classifications with high accuracy it can also be used for crime scene detection for appropriate evidence.
- Exploring the aspects of applications of Neural Networks more precisely Convolutional Neural Networks. which will be valuable for understanding more effective applications in future.

References

[1] http://iizuka.cs.tsukuba.ac.jp/projects/colorization/en/

[2] cs231n.stanford.edu/reports/2016/pdfs/219_Report.pdf

[3] R. Dahl. Automatic colorization. http://tinyclouds.org/colorize/, 2016.

[4] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In International conference on artificial intelligence and statistics, pages 249–256, 2010.

[5] B.Hariharan, P.Arbel´aez, R.Girshick, andJ.Malik. Hypercolumns for object segmentation and fine-grained localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 447–456, 201

[6] 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.