A Rainbow from Shades of Gray: Video Colourization

Group 18

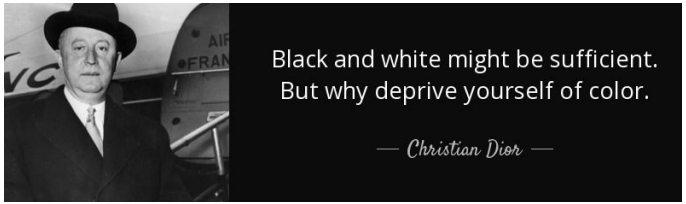
Abhishek Agarwal Abhishek Maiti Surabhi S Nath

Github

Problem Statement

To develop an end-to-end framework for meaningful and consistent colourization of black and white videos

- **Colour is a characteristic of human visual perception**
- **We are naturally receptive to colour and light intensities**
- **Black and white representations deny us of a very meaningful and significant feature - Colour**
- **Aim to *colour the past* and *bring it to life***



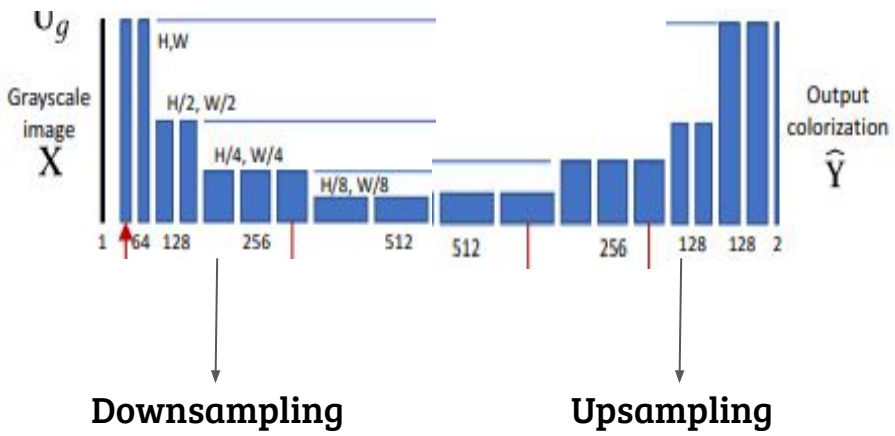
State of the Art

Image Colourization	Video Colourization

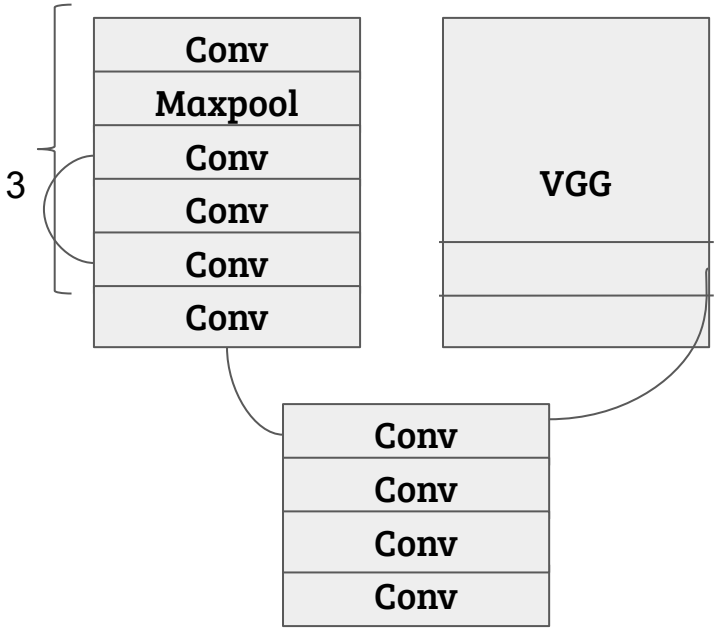
• Zhang et. al.
  ○ PSNR (dB): 27.85±0.13
  ○ AMT Fooling Rate: 30.04% ± 1.80
• Isola et. al.
  ○ AUC: 67.3%
• Larrson et. al.
  ○ RMSE: 0.299
	• Thomas et. al.
	  ○ Accuracy: 68%
	• Meyer et. al.
	  ○ PSNR (Averaged over 10 frames): 43.64

Baseline Architecture

Architecture 1 - Classification Model

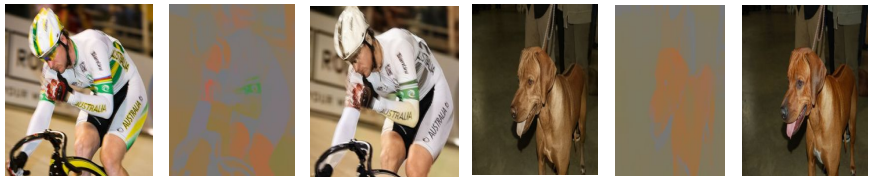


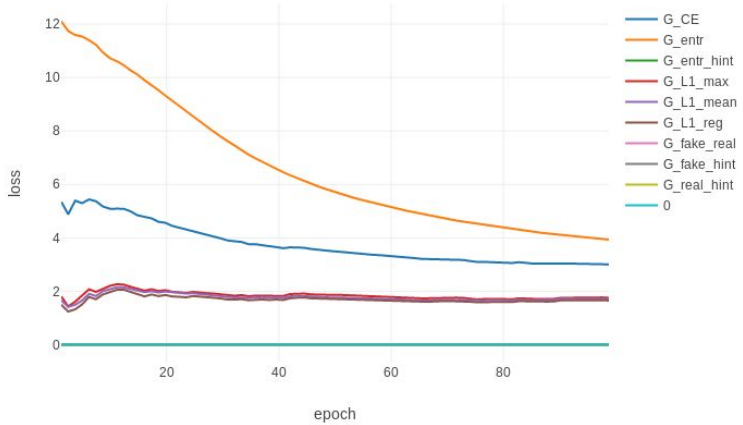
Architecture 2 - Regression Model



Results

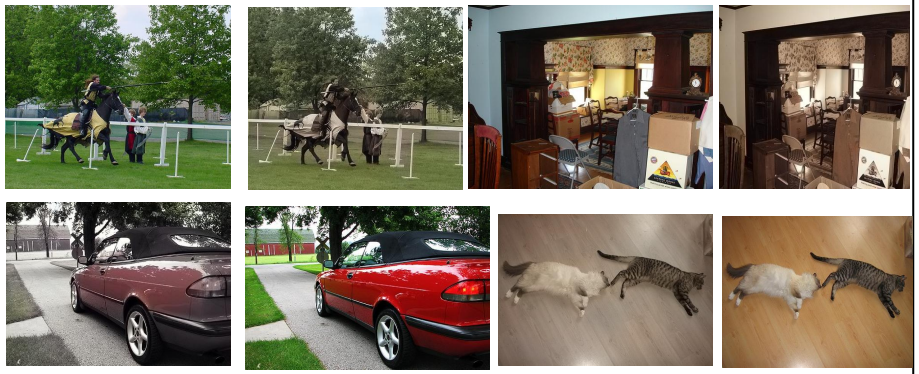
Architecture 1 - Classification Model





Ref: https://github.com/richzhang/colorization-pytorch

Architecture 2 - Regression Model

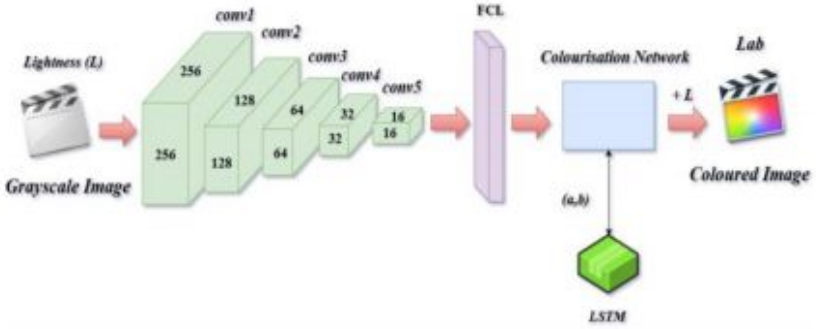


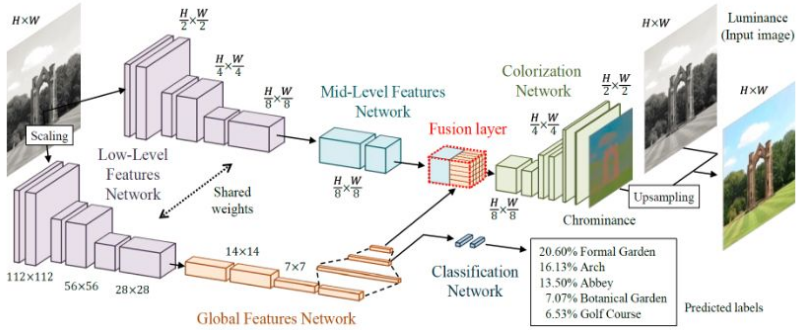
Mean L1 Norm = 132.66, RMSE = 8.23

Ref: https://github.com/PrimozGodec/ImageColorization

Planned Next Steps

- Introducing LSTM to capture dependencies among frames of video
- Improve underlying image colourization using a fusion network



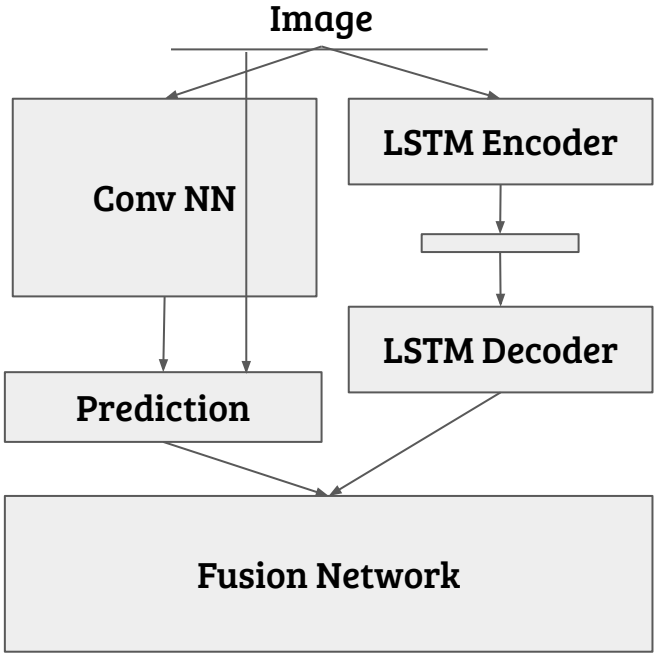


FURTHER EXPERIMENTS

We performed the following three different approaches:

1. Use the above image colourization architectures to colour videos frame-wise
2. Coded from scratch to integrate LSTM with CNN for temporal consistency
3. Utilized a Fusion-based Network for video colourization

New Architecture 1 - built from scratch



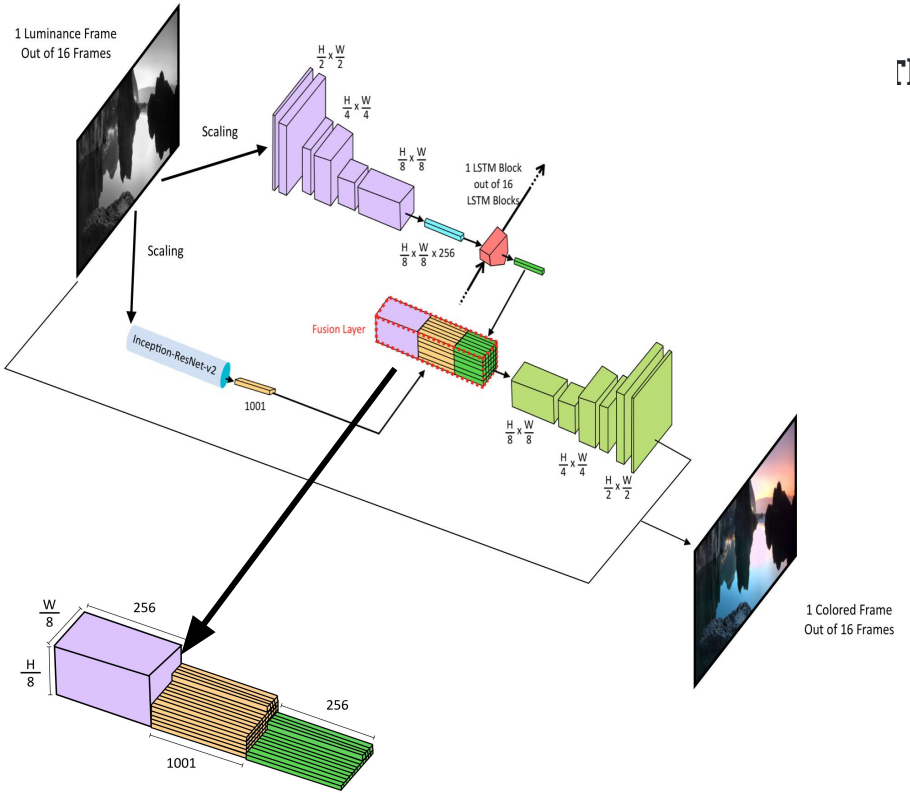
Expected output

CNN output

LSTM output

Due to limited GPU resources, training video frames on this network was a challenge and did not result in the expected outcomes

New Architecture 2



The architecture has four basic parts -

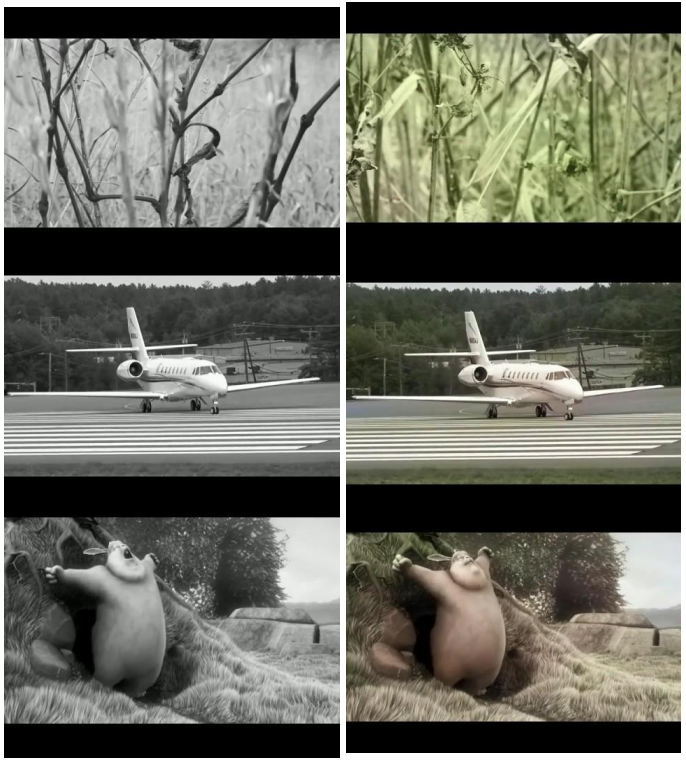
- A time distributed CNN encoder
- A time distributed CNN decoder
- A fusion layer
- A high-level feature extractor (Inception-ResNet-v2)
- An LSTM to extract temporal features

Ref: https://github.com/ThejanW/FlowChroma

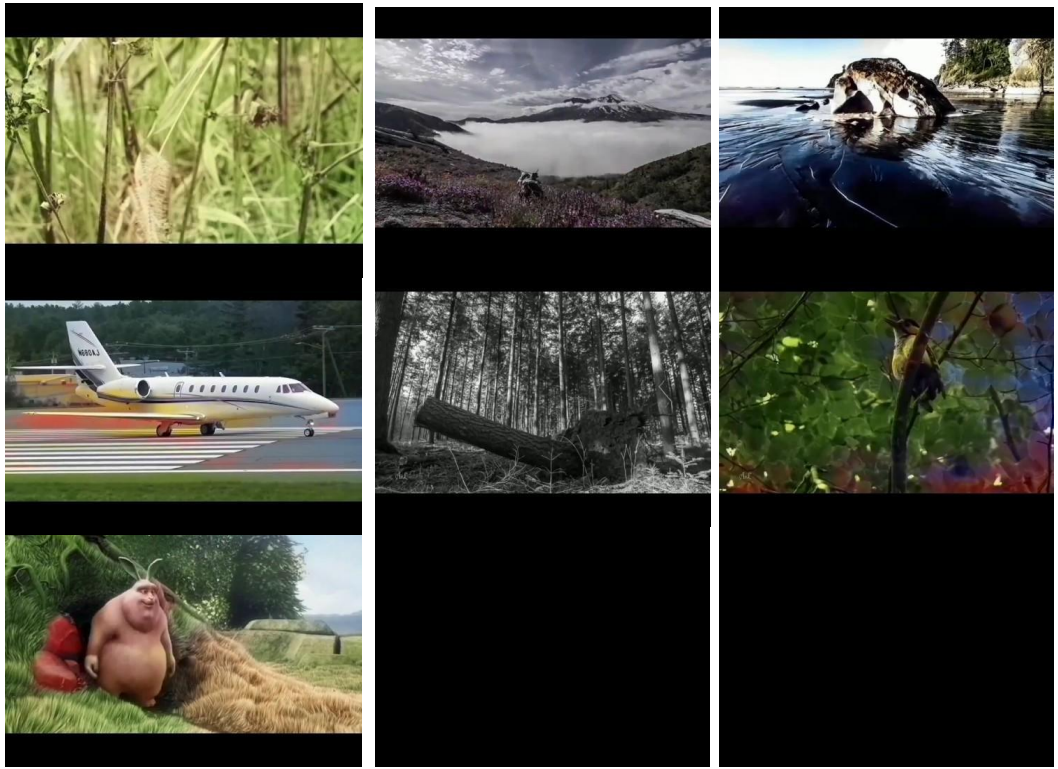
Results

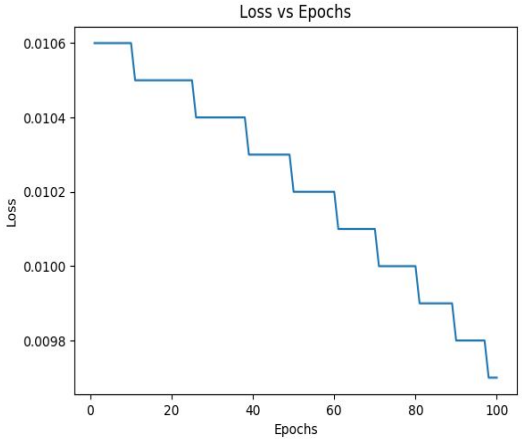
Find all results [Here]

Frame-wise Colourization



Fusion Network





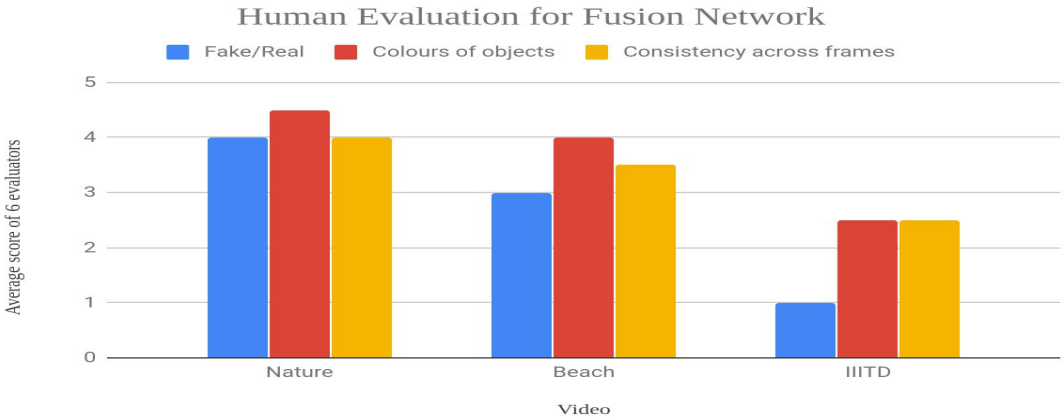
Loss plot for Fusion Network Training

```
Epoch 17/100

1/1 [==============================] - 8s 8s/step - loss: 0.0105
Epoch 18/100

1/1 [==============================] - 7s 7s/step - loss: 0.0105

Epoch 00018: saving model to checkpoints/model.hdf5
Epoch 19/100

1/1 [==============================] - 6s 6s/step - loss: 0.0105
Epoch 20/100

1/1 [==============================] - 5s 5s/step - loss: 0.0105

Epoch 00020: saving model to checkpoints/model.hdf5
```

Analysis

For evaluation, we formulated our own human evaluation metric to analyse the performance of the colourization networks.
We surveyed 6 students from our batch on the following self created metrics out of 5 for multiple videos:

- Real/Fake
- Colours of objects
- Consistency across frames



Future Potential

- We can use attention and get the weighted importance of the pixels of the previous frame to predict the current frame
- We can use GANs to train an adversarial framework to color the images
- We can combine the above two to get possibly better results

Individual Contribution

The work was evenly distributed among all of us. We all did literature survey, designed and implemented various architectures and analyzed the performance. It was a great learning experience and we thank Prof. Saket Anand for providing us this opportunity.