

DEEP LEARNING AND ITS APPLICATIONS

PROJECT PRESENTATION ON VIDEO COLORIZATION

GROUP-09

Mukul Jangid, Ritwik Saha, Hrushikesh Sarode



June 8, 2019

- 1 Problem Statement
- 2 Main Idea
- 3 Datasets
- 4 Basic Model
- 5 Basic Model Output
- 6 Bad Model
- 7 Bad Model Output
- 8 Desired Model
- 9 Desired Model Output
- 10 Conclusion

Problem Statement

Given black and white video, we are trying to output a color video by relating a frame with its preceding frames.

Main Idea

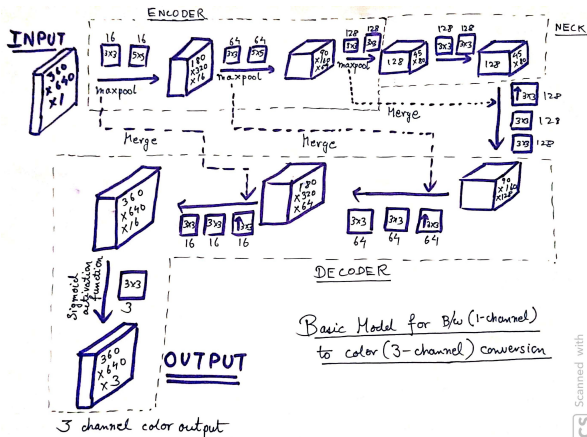
- If we apply the network directly on black and white images, it could not relate to the video frames.
videos, the frames are related to the previous frames, the direct approach would not detect that relation.
- We are trying to keep that relation between the frames so that we will get better results.

Datasets

- Tom and Jerry videos from YouTube
- Shape of image (360,640)

Basic Model

- 1 channel input, 3 channel output
- Considers only current black and white frame

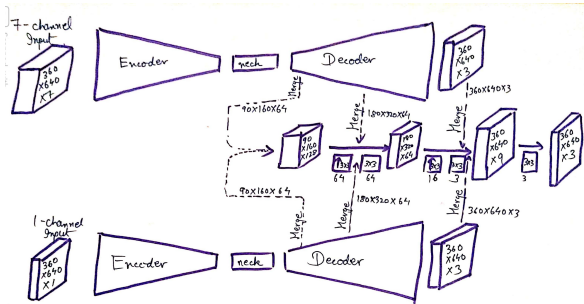


Basic Model Output

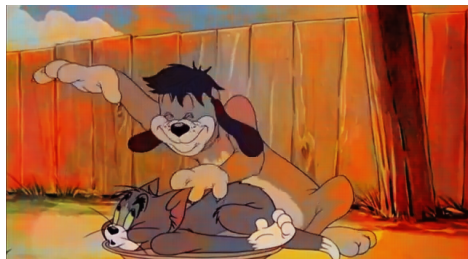
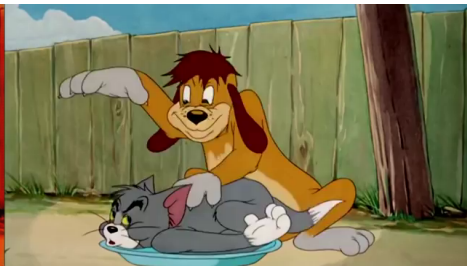


Bad Model

- 7 channel input, 3 channel output
- Considers previous 2 colored frame and current black and white frame
- Kept basic model non trainable



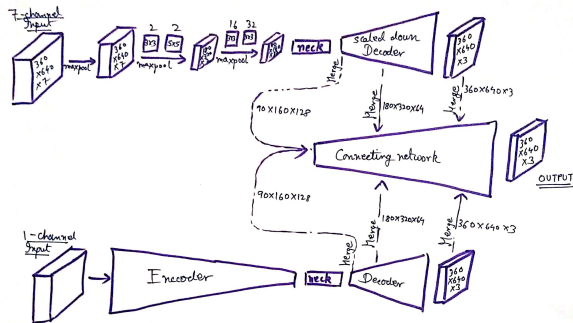
Bad Model Output (frame 1 and frame 2)



Desired Model

- 7ch \rightarrow 3ch model + basic model
- Kept basic model non-trainable

Desired Model Architecture



Desired Model Output



Conclusion

- **MSE with GT:**
 - Basic Model : 21.7358
 - Desired Model : 12.7757
- **MSE with Itself(frame by frame):**
 - Basic Model : 244.7051
 - Desired Model : 238.2823
- Learning of Final Model was done by downgrading the 7-channel input and its overall network so that the spatial features can be reconstructed from the 1-channel input and temporal relationship can be inferred from the upper network. This work of combining the features is done by an intermediate network.