Automatic Colorization of Videos

Tanvi Sahay, Ashutosh Choudhary

Problem Statement

Video colorization has been a popular part of the motion pictures as a means to modernize black-and-white movies or provide an artistic visual effect to them, to restore original color movies and videos and to integrate originally black-and-white videos into modern day color films. In this project, we attempt to automatically colorize videos by building on the present state-of-the-art image colorization architectures. We focus on rectifying two major issues encountered with video colorizations :

- loss of color consistency between subsequent frames
- desaturated colorization of individual frames.

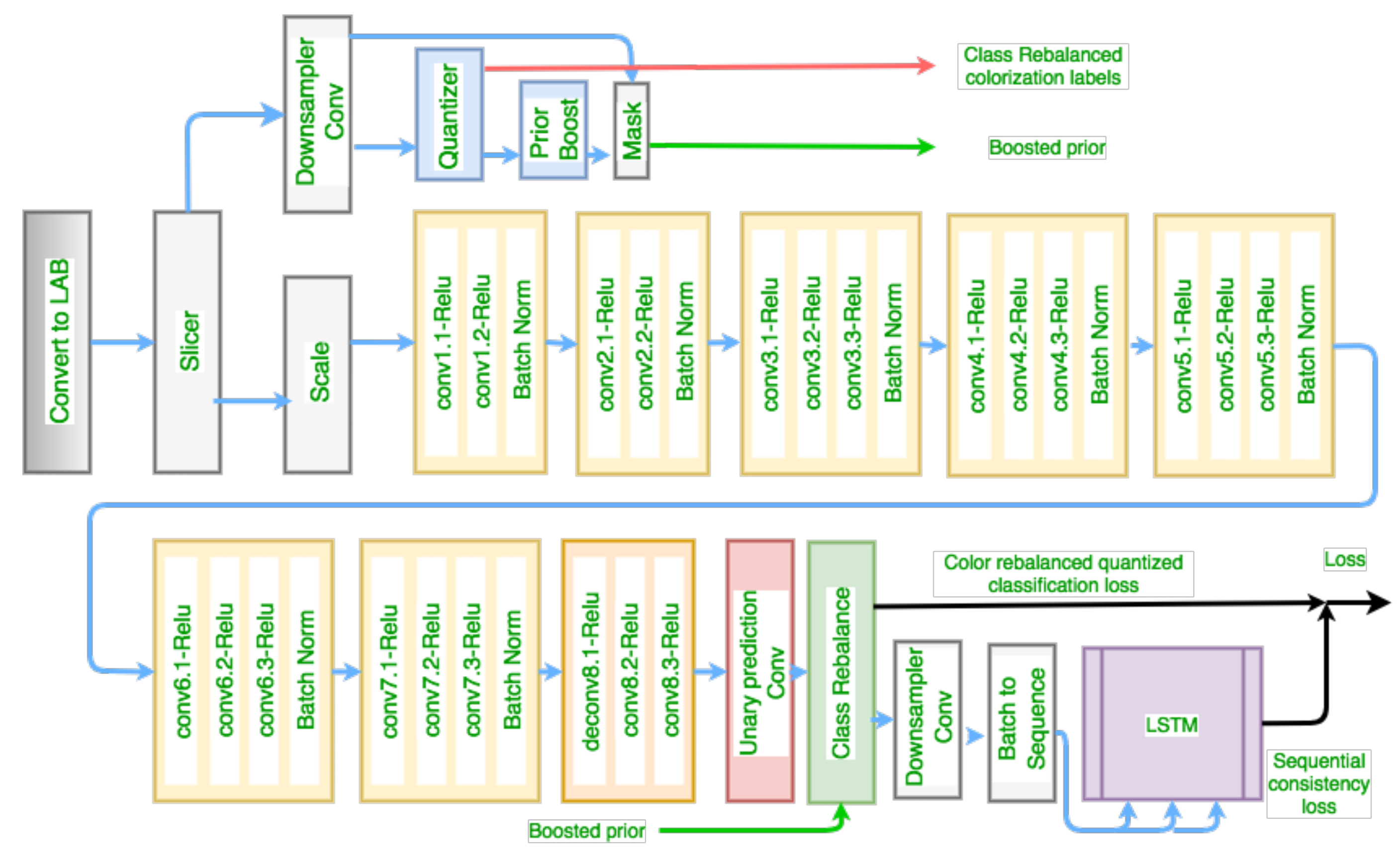
We use an LSTM to maintain color consistency between successive frames and employ a class-rebalancing loss to re-weight color predictions on the basis of their rarity.

Dataset

- Train, Val, Test videos: Google YouTube 8M dataset. Category: “Movieclips”
- Legacy videos: Google YouTube 8M dataset. Category: “black-and-white”
- Number of videos: 60 total. 48 train. 6 val. 6 test. (80-10-10)%
- Video lengths: 3 minutes. 24 frames per sec.
- Number of images (frames): 259200 total, 207360 train, 25920 val, 25920 test
- LSTM sequence length: 6

[1] G. Larsson, M. Maire, and G. Shakhnarovich. Learning representations for automatic colorization. CoRR, abs/1603.06668, 2016.
[2] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. CoRR, abs/1603.08511, 2016

Model and Experiments



- Modified VGG-16 to map from a grayscale input to a dist. over quantized color value outputs along with class rebalancing based on pixel color rarity.
- Each conv layer refers to a block of 2 or 3 repeated conv and ReLU layers, followed by a BatchNorm layer. The net has no pool layers.
- LSTM takes 6 sequential frames to predict next frame.
- Loss is weighted average of LSTM sequence loss and class rebalance loss.

$$\mathbf{L}_{cl}(\hat{\mathbf{Z}},\mathbf{Z}) = -\sum_{h,w} v(\mathbf{Z}_{h,w}) \sum_{q} \mathbf{Z}_{h,w,q} log(\hat{\mathbf{Z}}_{h,w,q}),\ where\ \mathbf{Z} = \mathcal{H}_{gt}^{-1}(\mathbf{Y})$$

$$\mathbf{L}_{seq} = -\sum_{j} y_j log(e^{\mathbf{S}_j} / \sum_{k} e^{\mathbf{S}_k}) \qquad \mathbf{L} = \lambda \mathbf{L}_{cl} + (1-\lambda)\mathbf{L}_{seq}$$

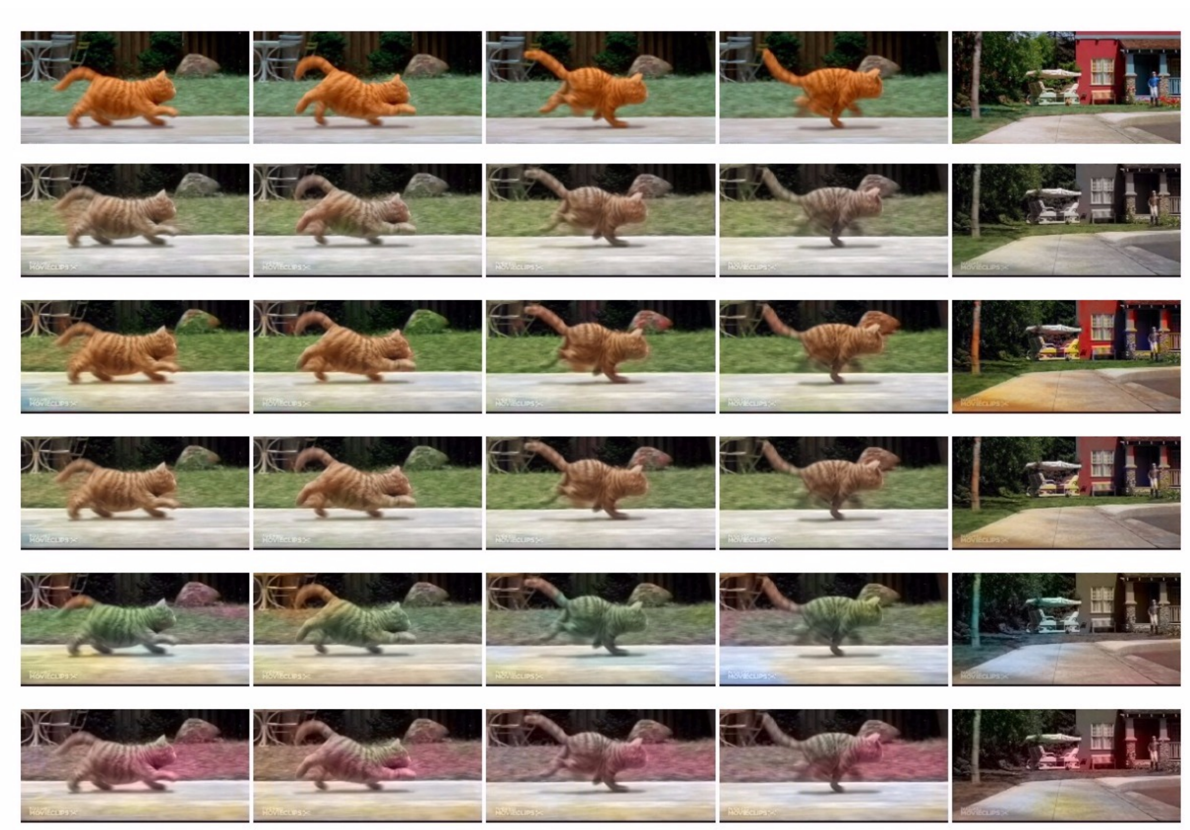
Experiments

1. Pre-trained classification on quantized color space(QCS) [1]
2. Pre-trained classification on QCS with class rebalancing. [2]
3. Ensemble of experiments 1 and 2.
4. Transfer learning and classification on QCS with class rebalancing
5. Transfer learning and classification on QCS with class rebalancing and sequential modeling (our model).

Evaluation

- Average per-pixel RMSE over the set of all frames in a single video
- Color turing test (CTT): Since, we are using class rebalancing a turing test is a better evaluation metric than RMSE.

Results and Analysis



Exp	RMSE	CTT
1	7.677	3.4
2	8.825	4.2
3	8.108	4.5
4	8.021	3.2
5	8.210	3.0

Exp1 gave the best RMSE since its learned colors were closest to ground-truth, despite being desaturated. However, overall, Exp3 performed best, giving a good comparative RMSE value and controlling the unnaturally bright colorization caused by Exp2. Exp5 despite giving good RMSE, gave inconsistent and desaturated results, suggesting need of a better loss function and more training time.