

Cinematic Color and Colorization

FRANKIE EDER, VIVIEN NGUYEN, and CECILIA ZHANG, University of California, Berkeley, USA

CCS Concepts: • Computing methodologies → Computational photography.

Additional Key Words and Phrases: computational-photography

ACM Reference Format:

Frankie Eder, Vivien Nguyen, and Cecilia Zhang. 2020. Cinematic Color and Colorization. *ACM Trans. Graph.* 1, 1 (February 2020), 7 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

There are many things that set professional-grade movies, or *cinematic* videos, apart from videos captured by everyday users, or *casual* videos. The equipment, narrative, visual framing, sound quality, and more all contribute to the sense that cinematic videos are more artistic and higher quality. One critical editing step in the movie-making pipeline is *color grading*, the process of editing the colors in a video to achieve a desired visual appearance, whether for correctness or aesthetic reasons.

It is well known that cinematographers may carefully select colors to provoke an emotional response, provide visual clarity, or serve a narrative purpose. We also have the sense, however, that the simple *set* of colors used in cinema is distinguishable both from the colors of casual videos and among different cinematic genres.

In this project, we analyze how color is used in cinema and explore methods to transfer color between cinematic and casual videos.

2 CINEMATIC VIDEO DATASET

To ensure that we perform our analyses on films typically considered cinematic, we collect a dataset of 2003 movies trailer from the top 1137 rated movies from The Movie Database (TMDb). We consider trailers an accurate estimation of the characteristics of color in the entire film, and perform some pre-processing steps to remove extraneous artifacts in our analysis (further details in section 5.1). Notably, our analyses only include live-action films; any animated films (i.e. films have one of their genres as animated) are excluded from our analyses. We assume that all of these trailers and films were made with professionals overseeing the visuals of the film to achieve a "cinematic" look. We also collect a smaller dataset of 18 casual videos, which we occasionally compare our cinematic videos to as a baseline. In contrast to cinematic videos, casual videos are

Authors' address: Frankie Eder, frankieeder@berkeley.edu; Vivien Nguyen, vivien.nguyen@berkeley.edu; Cecilia Zhang, cecilia77@berkeley.edu, University of California, Berkeley, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

0730-0301/2020/2-ART \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

assumed to not have any professional oversight when it comes to image and color quality.

3 COLOR REPRESENTATIONS AND VISUALIZATION

To get a sense of how color is used in videos, we experimented with different methods of visualization. Different color representations and types of plots can reveal different information. In general, we used RGB to perform visualizations, and CIE LAB when performing color transformations and calculations.

3.1 CIE Chromaticity

We simply plot points from the image directly on top of the CIE chromaticity diagram to get a sense of color distribution. We quantized the frames in order to avoid cluttering the diagram.

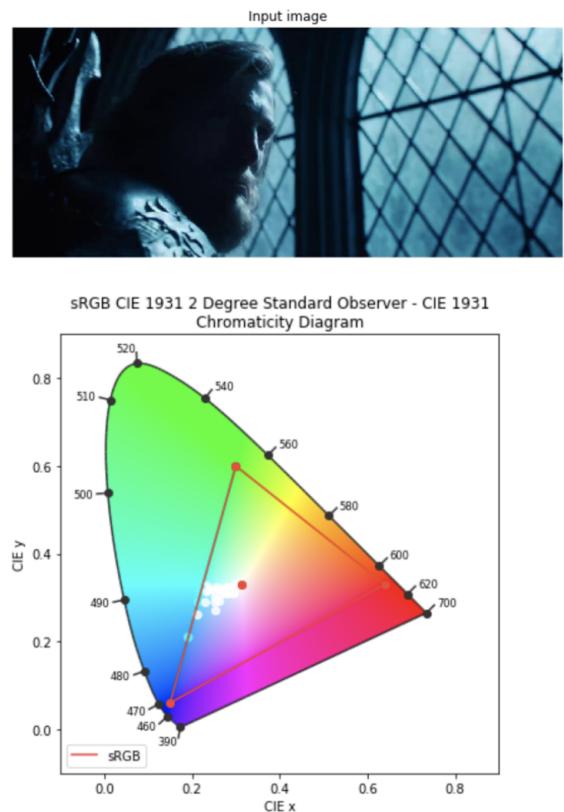


Fig. 1. Plotting colors from the quantized image gives a sense of color distribution in the image.

3.2 3D RGB Plots

We plot a three dimensional plot with the axes corresponding to R,G, and B channels. For cinematic videos, this demonstrates that colors are dense (smoothly varying) and use a small amount of the

total space (not highly varying in RGB space). This represents the general aesthetic goals of color grading for cinema.

We noted that points tended to cluster along the diagonal representing grey-values. See Figure 2.

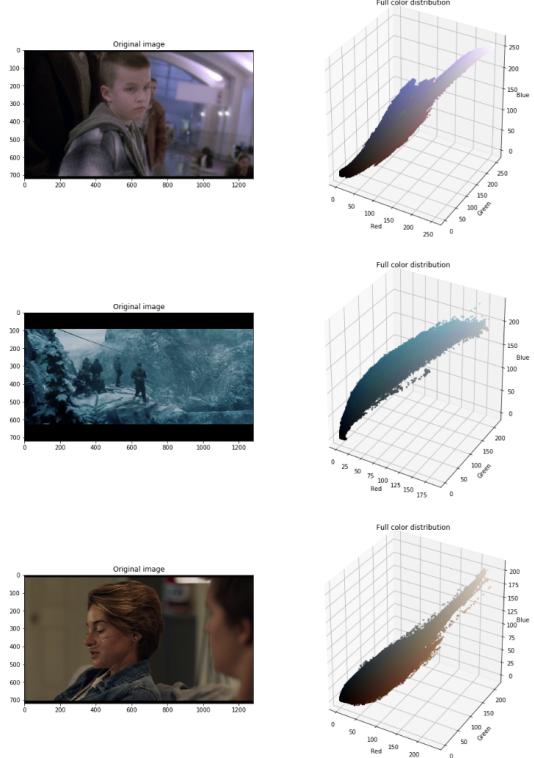


Fig. 2. Plotting in 3D RGB gives a sense of color shape and distribution and allows us to see levels of darkness and brightness.

We also created a version of the 3D RGB plot that was a histogram, where the plotted point size corresponds to the amount of pixels in the frame of that color.

3.3 Log-chroma Histogram

We ultimately use the log-chroma histogram, as described in [insert citation] to visualize color for most of this project. The histogram allows us to quickly and clearly see which colors are present or dominant in a frame or aggregated over an entire trailer or genre. We can also use it to visualize which colors have the most variation (section 4.3.2).

To create the histogram, we follow the steps from [Barron 2015]. Let R_i, G_i, B_i be the raw RGB values for a particular pixel, i . Compute

$$u_i = \log\left(\frac{G_i}{R_i}\right)$$

and

$$v_i = \log\left(\frac{G_i}{B_i}\right)$$

. After calculating this for all pixels, we take the square root of the L1-normalized values and generate a 2D histogram for u and v using

a $(50, 50)$ grid of bins. To recover a visualization, we set the green channel equal to the counts of the histogram, c . We then set the red channel equal to the counts of the histogram divided by the exponentiated u edges, and the blue channel equal to the counts divided by exponentiated v edges.

$$g = c, r = \frac{c}{\exp(edges_u)}, b = \frac{c}{\exp(edges_v)}$$

We clip values to $[0,1]$ and re-normalize the visualization to the max of the *non-grey value points*. As noted with the 3D RGB visualization, the high presence of white, grey, and black values in the frames would make it too difficult to see any of the colors in the histogram if we normalized to the overall maximum value.

See Figure 3 for results.

4 COLOR ANALYSIS

If we ultimately want to make progress on automatic color grading, we must first understand how color is used in cinema across frame, trailer, and genre levels.

4.1 Segmentation

One way of understanding how color is used within a particular frame is by segmenting the frames. We experimented with two segmentation methods: semantic segmentation, where the image is labeled based on concepts like 'sky', 'grass', or 'person', and color segmentation, where the image is labeled based on color names like 'blue', 'green', or 'red'. We can then use simple transforms to transfer colors from one frame to another. By modulating the colors in this way, we can visually see how colors differ between types of video (cinematic vs casual) and between genres (for example, Romance vs Thriller).

4.1.1 Semantic Segmentation. We attempted to use pre-trained DeepLab models, one trained on Cityscapes and one trained on COCO Things and Stuff. As expected, the model trained on Cityscapes provided high quality segmentation on outdoor images resembling those in the dataset created for self-driving related applications, but poorly on anything that did not fit that paradigm. Though the model trained on COCO Things and Stuff was more generalizable and contained many more labels, we were still dissatisfied with the segmentation results. We believe this is at least partially to do with the artistic editing that is applied to cinematic frames, and also the model's difficulty with artifacts like motion blur. We note that we did not try using any model explicitly trained for segmentation with videos.

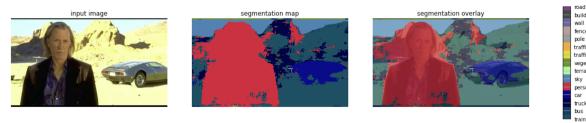


Fig. 4. Example of segmentation using DeepLab trained on Cityscapes

4.1.2 Color Segmentation. We decided to try a different method of segmentation: segmenting by color. We re-implemented the color naming model from [Heer and Stone 2012], which allows us to select



Fig. 3. 2D log-chroma histograms from different genres. Each image represents the aggregation of colors from a single trailer that is labeled with the corresponding genre

all blue, green, red, etc. pixels from an image. The color naming model includes over 150 color names.

In brief, the color naming model uses the CIELAB values for each pixel in the image and performs a transformation utilizing values from their data driven model. At look-up, you can provide a threshold and a name to search for pixels corresponding to a particular color name.

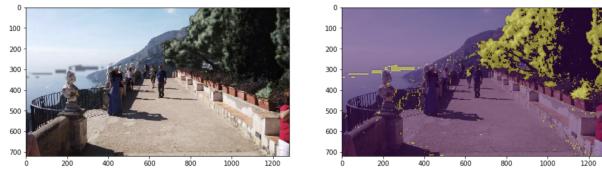


Fig. 5. Example of segmentation using color naming model

4.1.3 Color Transfer. Color transfer is the process of transforming the colors in a *source* image in order to more closely resemble the colors in a *target* image.

We see that global color transfers are generally undesirable. Though it effectively transfers colors to certain regions, it modifies other regions of the image too intensively.



Fig. 6. Transferring colors from target to source, via histogram matching.

We use the color segmentation method and a simple affine transformation to transfer colors from one frame to another. We loop through each color in the naming model, and for each color:

- (1) We locate the pixels corresponding to this name in a *downsampled* version of the image (in order to reduce noise and for tractable computation).
- (2) We sample 20 pixels from both the target and the source.

- (3) We solve a linear least squares approximation, $Ax = b$ to determine the affine transformation matrix x that will allow us to transfer colors from the source (represented by A) into colors from the target (represented by b). There are certainly other elegant or complex ways to achieving this.
- (4) We obtain the mask in the *full-resolution* source image for this particular color name and apply a Gaussian blur to avoid edge artifacts.
- (5) Using the mask and the transformation matrix, we apply the transform onto the correct pixels. We average the new values into the existing ones in order to avoid additional artifacts.



Fig. 7. Transferring colors from target to source, via global or segmentation based affine transforms..

4.2 Temporal

As time is an integral component of film, we decided to also explore the characteristics of color over time. Before we proceed, we define some terminology based off of standard definitions used in filmmaking. A *frame* f is a still, three-channel image from a video. We define a frame pair $p = (f_i, f_{i+1})$ as a pair of subsequent frames. A *shot* $s = (f_1, f_2, \dots, f_n)$ is an ordered collection of frames that were recorded continuously (and lie between two subsequent *cuts*). We formally define a *cut* as a frame pair that include the last frame of one shot and the first frame of the following shot. This lets us understand each trailer in our dataset as an ordered collection of shots $t = (s_1, \dots, s_n)$. We use PySceneDetect to detect every cut in each trailer of our dataset. We run three separate analyses comparing log-chroma histograms over time in order to investigate (Sections 4.3.1 - 4.3.3), and discuss in Section 4.3.4.

4.2.1 Local Variation by Frame. Our simplest analysis consisted of comparing log-chroma histograms between all frame pairs. For the

log-chroma histograms of each frame in a frame pair $p = (f_i, f_{i+1})$, which we will call $L(f_i)$ and $L(f_{i+1})$, we use the correlation coefficient r (between corresponding bins) as a metric for how similar the color is in between frame f_i and f_{i+1} . We can analyze the distribution of the correlation coefficient r among frame pairs to determine the typical variation in color over time. Figure 4 shows these distributions for select subsets of all pairs successive frames in our dataset.

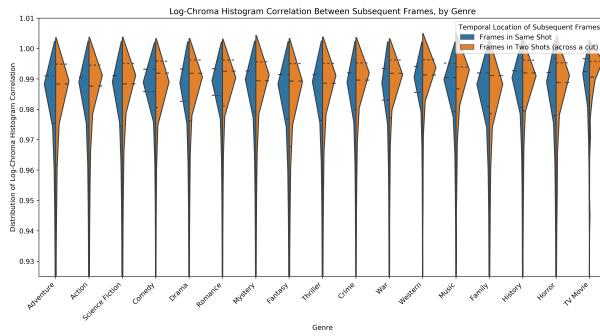


Fig. 8. Distribution of log-chroma histogram correlation between frame pairs, individually calculated and plotted for all frame pairs within different genres. Within each genre, we also provide a distinction between frame pairs that are within the same shot, and frame pairs that are in two different shots, defining a cut. These violin plots contain IQR estimates depicted by dotted lines, and the distributions are smoothed with a KDE bandwidth of 0.01.

There are three main conclusions that we can draw from these distributions:

- (1) Color between frame pairs is highly correlated, with $r \approx 0.99$.
- (2) The correlation in color between successive frames within shots is similar to the correlation of color between those that are in a cut.
- (3) As there are (again) no significant distinctions between genres, both of the above conclusions hold regardless of genre.

4.2.2 Local Variation by Shot. Our subsequent analysis consists of comparing log-chroma histograms between each successive shot. To do this, we calculate the *mean log-chroma histogram* $\bar{L}(s)$ by calculating the log-chroma histogram for each $f \in s$ and performing an arithmetic mean between all corresponding bins. Specifically, for a shot s with n frames f_1, \dots, f_n , all bins (u, v) in $\bar{L}(f)$ must satisfy the following equation:

$$\bar{L}(s)(u, v) = \frac{\sum_{f \in s} L(f)(u, v)}{n} \quad (1)$$

This definition allows us to compare the color over time in a similar manner as Section 4.3.1, but instead treating subsequent shots (s_i, s_{i+1}) as we did frame pairs and using the mean log-chroma histogram as we did the traditional single-frame log-chroma histogram. The results of this analysis are depicted in Figure 5.

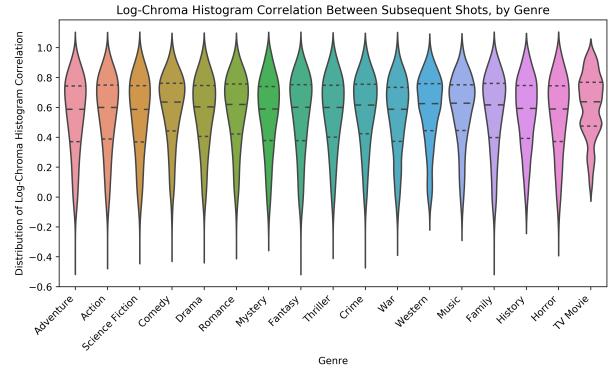


Fig. 9. Distribution of mean log-chroma histogram correlation between successive shots, individually calculated and plotted for all successive shots within different genres. These violin plots contain IQR estimates depicted by dotted lines, and the distributions are smoothed with a KDE bandwidth of 0.2.

When comparing this with the results of our earlier section, we can conclude the following:

- (1) Color between successive shots is significantly correlated, with $r \approx 0.6$.
- (2) The correlation in color between successive shots is lower than the correlation between successive frames.
- (3) As there are (again) no significant distinctions between genres, both of the above conclusions hold regardless of genre.

4.2.3 Global Variation by Shot. To gain an even more holistic view of color changes over time, we compare color between all shots in our trailers using our *standard deviation log-chroma graph* (SDLCG). An SDLCG, $L_\sigma(t)$, is calculated from a trailer t , i.e. a set of k shots, $t = (s_1, s_2, \dots, s_k)$. In short, L_σ computes the element-wise standard deviation (σ) of each bin (u, v) across the mean log-chroma histograms of all shots s_i in the trailer t . More formally, each bin (u, v) in $L_\sigma(t)$ is calculated by:

$$L_\sigma(t)(u, v) = \sigma(\bar{L}(s)(u, v) \forall s \in t) \quad (2)$$

Where $\sigma(x_1, \dots, x_n)$ is the traditional standard deviation calculation of the observations x_1, \dots, x_n .

When averaged (again, element-wise) among all trailers in the desired subset of our dataset, we can estimate a more generalized metric of color variation within that subset. Like in Section 4.3.2, to do this, we define a mean standard deviation log-chroma graph (MSDLCG), that, for a set of j trailers $T = t_1, t_2, \dots, t_j$ and all bins (u, v) satisfy:

$$\bar{L}_\sigma(T)(i, j) = \frac{\sum_{t \in T} L_\sigma(t)(u, v)}{j} \quad (3)$$

In Figure 6, we can see that visualizing this structure allows us to draw conclusions about the variation of color between different genres. This allows us to draw the following conclusions about variation of colors in cinema:

- (1) All genres vary most at the wait point, implying that luminance is the primary color component that varies in film.

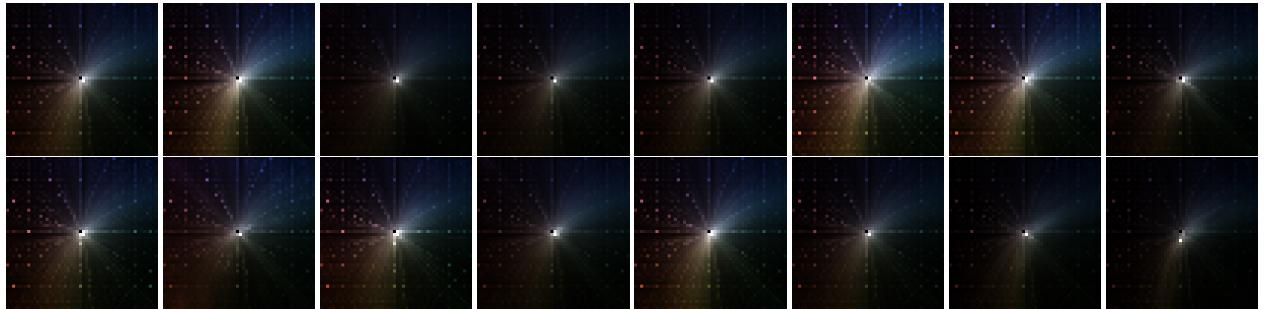


Fig. 10. Visualizations of the MSDLCGs from different genres in our dataset, which compares the variation of colors between different genres. The genres visualized, in order from top left to bottom right (in row-major order), are: Action, Adventure, Comedy, Crime, Drama, Family, Fantasy, History, Horror, Music, Mystery, Romance, Science Fiction, Thriller, War, and Western. Note that we visualize our standard deviation log-chroma graphs in the same manner that we visualize our log-chroma histograms; however, the brighter regions no longer correspond to the count of the particular color, but to our estimation of variation as described in equation 3.

- (2) Variation in color is distinct between some genres of film. For instance, War films have much more variation in its luminance in comparison to other colors, and Westerns have much more variation in blues and yellows between shots.
- (3) Many genres have similar color variation between shots. This is because each trailer is multi-labelled with genres, many movies fall into both of these genres. Some of these genres are highly correlated. For instance, Action and Adventure films have similar variation because it is rare for movies to be classified as one of genre and not the other.

We can repeat this analysis to compare our casual and cinematic data. We display the results in Figure 7.

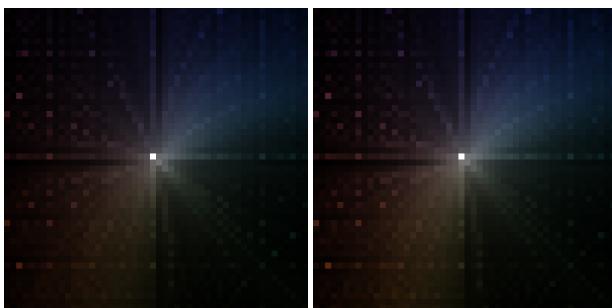


Fig. 11. Visualizations of the MSDLCGs of all of our cinematic videos (left) and all of our casual videos (right). Note that we visualize our standard deviation log-chroma graphs in the same manner that we visualize our log-chroma histograms. However, note that the brighter regions no longer correspond to the count of the particular color, but to our estimation of variation as described in equation 3.

Although the differences are subtle when placed adjacently, we can draw the following conclusions about cinematic videos in comparison to casual ones:

- (1) There is slightly higher variation in the color and slightly lower variation in the luminance of casual videos. This is likely due to auto-color-correction and auto-exposure settings that we would expect from devices that record casual video.

- (2) There is slightly higher variation in the luminance and slightly lower variation in the luminance of cinematic videos. This is what we would expect given that lighting, exposure, and color conditions are much more controlled on a professional film set than in the casual setting.

4.2.4 Discussion. While on first glance the conclusions of the previous sections might seem contradictory, they are reasonable when considering that in the cinematic setting, lighting, exposure, and color are controlled by a variety of film-making professionals at many stages of image capture and processing. Aggregating our conclusions:

- (1) The temporal characteristics of color are different between cinematic and casual videos.
- (2) Cinematic Videos have a significant correlation frame to frame and shot to shot. This is what we would expect given that lighting, exposure, and color are highly controlled by professionals in the cinematic setting.
- (3) Although cinematic Videos have a significant correlation frame to frame and shot to shot, a broader analysis shows that there may be more significant variation patterns when viewing all shots in a film. These patterns typically vary by genre.

4.3 Clustering

We also wanted to run some simple experiments based on clustering the trailers to see some general trends in color usage.

We compute and visualize 2D log chroma histograms for trailers based on their genre in Figure 3.

5 DATA-DRIVEN GENRE CLASSIFICATION

Both temporal color analysis and clustering results suggest that there is a correlation between cinematic colors and genres. We use a data-driven approach to quantitatively study how color features alone are associated with video genres.

5.1 Data Processing

We perform several data processing steps for our data-driven models.

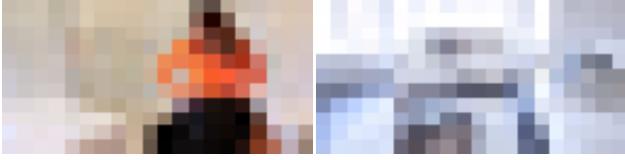


Fig. 12. Two example downsampled frames used to train the genre classification model.

- (1) We sample every 10 frames from each trailer.
- (2) We remove the first and last 5 seconds of each trailer, to avoid title screens and credits.
- (3) We crop black borders resulting from varying aspect ratios off each frame.
- (4) We 85-15 split the trailers for training and validation. From each trailer, we sample 10 sets of 20 random frames. Each set becomes an input for training or inference.

5.2 Experiments

We train two classification models that vary in how color feature is represented: histogram-only, downsampled color frames and gray-scale frames. Histogram-only gets rid of entirely the spatial structure and semantics of video frames. This model shows how color alone can infer the type of cinematic videos. Downsampled color frames (blocky frames) preserves a coarse spatial structure and semantics of the video. For example, the semantics of sky indicates blue colors are more frequently seen in the upper half. Below are descriptions on the data representation for each of the two models:

- (1) Histogram: we compute the same log-chroma histogram representation ?? used for previous color analysis, giving two histograms for the u, v channels. We also compute the histogram of the luminance channel in the LAB space and concatenate it with the log-chroma histogram. We use 128 bins per histogram
- (2) Downsampled color frames: we downsample the frames and then upsample using nearest neighbor interpolation to remove the semantics that are not color-related. Figure 12 shows 2 example downsampled frames used for training

Each training instance contains 20 frames randomly sampled from a single trailer video. For the histogram model, we use a multi-layer perceptron that has 4 intermediate layers, each with a Relu activation, and the final layer is followed by a softmax activation. For the blocky model, we use a Unet architecture, similarly with a softmax following the last layer. We apply a multi-label cross entropy loss, and train until convergence using Adam optimizer.

We maintain a balanced number of video trailers per class by randomly sampling 100 video per class label. However, that each video trailer has multiple labels, for example, Thriller is oftentimes also categorized as Action, and thus the number of video trailers per class label ends up to be not exactly the same. Below in Figure 13 we plot a correlation matrix of all the labels in the dataset, from which we can see there are certain classes that are highly correlated, e.g. Thriller and Action.

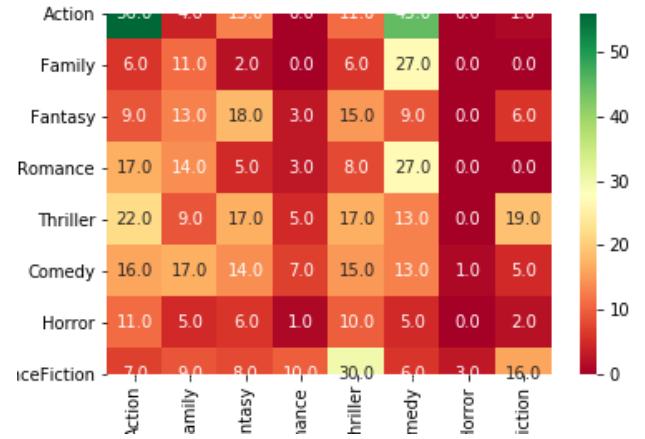


Fig. 13. Genre labels in the training data. Label pairs such as Thriller and Action, Comedy and Family, Thriller and Science Fiction are all highly correlated labels.

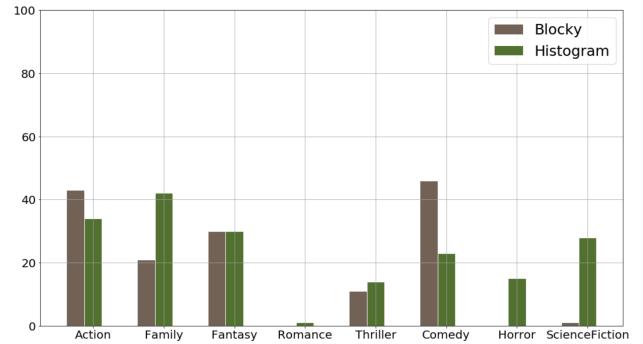


Fig. 14. Genre classification result on the two models. Both color models demonstrate that color can be informative about the video genre, even without any semantic information.

5.3 Results

Figure 14 shows the genre classification results using two different color representations: histogram and blocky downsampled frames.

We apply the trained models to the held-out test dataset. Both models achieve an average accuracy of 3 times higher than chance (12.5%), indicating that color alone is a useful signal to inform the genre of a video.

To understand the reasons about the low-accuracy class results, in particular, the Romance and Horror genres, we plot the confusion histograms in Figure 15. As indicated from the plots, Horror is confused with Action, Fantasy and Thriller, and Romance is confused with Comedy and Action. These statistics show consistency with the label correlations plotted in Figure 13.

The benefit of using color features to train a classification model is the ability to enable automatic color grading. Given a casual video to be graded and a target genre, we can freeze the trained classification model as a discriminator, and iterate over different

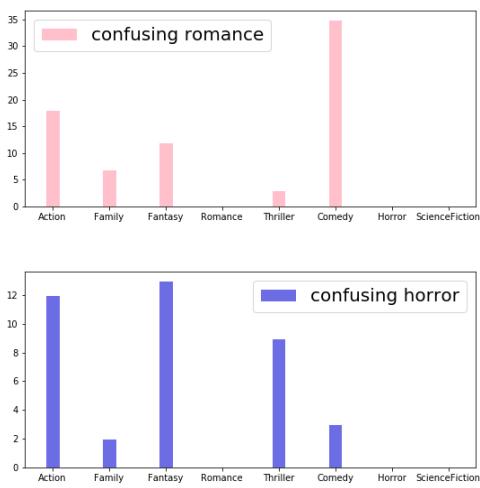


Fig. 15. Confusion histogram for Romance and Horror genre. We can see a similar pattern from the correlation matrix of the labels in the dataset plotted in Figure 13.

color transformations to optimize for the classification score of the desired class.

6 CONCLUSION

Though we originally set out to work on color transfer from cinematic to casual videos, we found there was a lot to analyze and visualize in studying color itself. We found that clustering-based, segmentation-based, and temporal-based analysis provided useful insights into understanding how color is used in cinema, and how that differs from casual videos and between genres.

We also found that color is a very strong signal for identifying genre and began some data-driven models to explore this.

We believe that this understanding and insights could be useful to automate color transfer from cinema to casual videos. For example, one could fix the weights in a model trained for genre classification, and use it as a discriminator against colorizing casual videos.

REFERENCES

- Jonathan T Barron. 2015. Convolutional Color Constancy. In *ICCV*.
 Jeffrey Heer and Maureen Stone. 2012. Color Naming Models for Color Selection, Image Editing and Palette Design. In *ACM Human Factors in Computing Systems (CHI)*.