Evaluating Photo Aesthetics Using MobileNetV3 and Feature Engineering

Frank Zhao¹, Scott Xu¹
Department of Computer Science, Stanford University

Project Overview
- For a professional photographer, sorting through thousands of images can be time-consuming and inefficient. Thus, automatic image aesthetics assessment can be a game changer.
- It can also be applied in a wide variety of applications, such as image thumbnailing and automatic editing.
- Pass original image through MobileNetV3 - Large.
- Different from previous work, we also pass the image's saliency map generated from a pretrained model through MobileNetV3 - Small.

Datasets & Metrics
- Uses AVA (Aesthetics Visual Analysis) dataset.
- 255,000 images rated manually based on aesthetic qualities.
- The image rating ranges from 1 to 10 with 10 being the highest. The mean rating is around 5.5.
- Use 5 as cutoff score. Images with mean rating less than 5 are "bad" and higher than 5 are "good".

Methods & Model Architecture
- Color histogram and HOG as augmentation before 2 fully connected layers.
- Achieves 77.2% accuracy in binary classification.
- Problem Statement: takes an mage and its auxiliary features to predict a discrete probability distribution over the rating 1-10.
- Main contribution: Feature Engineering, an Aggregation of MobileNetV3 networks, and demonstration of some important applications.

Results
- Our model converges in 45 epochs, compared to 85 epochs for baseline.

Evaluation & Applications
- Integrated gradients method in Captum.
- Color histogram and HOG not so useful.
- Saliency map more useful.
- In future work, get rid of histogram and HOG, and replace the original image's network to Inception to compare.

Effect of Image Adjustments on Aesthetic Scoring
- Inverted U-shaped: if contrast is too low or too high, predicted aesthetics scores decline.
- Similar shape for brightness.
- Monotonic increase for sharpness.
- Use best adjustment factor in each to auto-enhance image.

Automatic Cropping
- High separability.
- Model is expressive.
- Maybe 2 FC layers is not necessary.

Future work: compare with 1 FC layer architecture.

Integrated gradients method in Captum.