Predicting Sunset Quality and Peak Time from Midday Sky Images: A Dual-Task Deep Learning Approach (v6)

Kasey Markel

Abstract

We present a novel deep learning framework for predicting both sunset aesthetic quality and peak viewing time from midday sky images captured 3 hours before sunset. Using historical timelapse videos from the Lawrence Hall of Science in Berkeley, California, we extracted 86 days of sunset imagery across multiple timepoints relative to sun-under-horizon. Our dual-task ResNet-18 model predicts both sunset quality (1-10 scale) and peak viewing time (minutes relative to sunset) from midday images. The model achieves a mean absolute error of 1.47 on quality prediction (r=0.115) and 6.20 minutes on peak time prediction (r=0.059), with quality prediction significantly outperforming baseline mean predictions. This work demonstrates that visual patterns in midday sky images contain predictive information about sunset aesthetics, enabling advance planning for photography and outdoor activities.

1. Introduction

Sunset prediction has applications in photography, outdoor activity planning, and solar energy forecasting. While astronomical calculations can predict when the sun will set, they cannot predict the aesthetic quality of the sunset or the optimal viewing time. We propose a dual-task deep learning approach that predicts both sunset quality and peak viewing time from midday sky images captured 3 hours before sunset.

2. Methods

2.1 Data Collection

We collected 101 historical timelapse videos from the Lawrence Hall of Science YouTube channel, spanning 2000-2020. From each video, we extracted: (1) one midday frame captured 3 hours before sunset, and (2) eight sunset frames at timepoints -10, -5, 0, +5, +10, +15, +20, and +25 minutes relative to sun-under-horizon. A total of 86 videos had complete data across all timepoints.

2.2 Labeling

Sunset images were manually graded on a 1-10 aesthetic quality scale by a single annotator. For each date, quality scores were collected at three timepoints (-10, 0, +10 minutes). Peak viewing time was

calculated by interpolating quality scores across timepoints to find the maximum aesthetic quality.

2.3 Model Architecture

Our dual-task model uses a ResNet-18 backbone pretrained on ImageNet to extract features from midday images. The extracted features are fed into two separate heads: (1) a quality prediction head that outputs a score from 1-10, and (2) a peak time prediction head that outputs minutes relative to sun-under-horizon. The model is trained with combined loss: L = L_quality + L_peak_time.

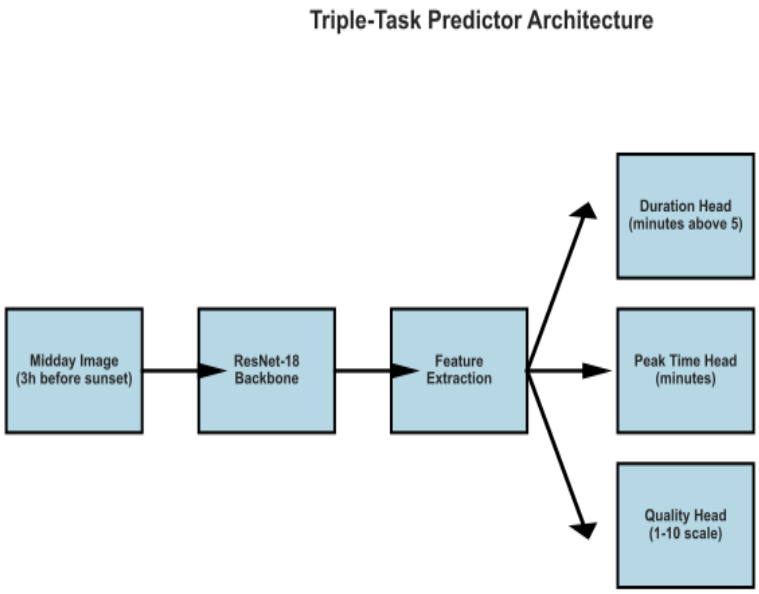

Figure 1: Dual-task model architecture. Midday images (3h before sunset) are processed through a ResNet-18 backbone to extract features, which are then fed into separate heads for quality and peak time prediction.

3. Results

We split the dataset into 68 training and 18 test samples. The model was trained for 50 epochs with Adam optimizer (learning rate 0.001). On the test set, quality prediction achieved MAE=1.69 and RMSE=2.18 (on 1-10 scale), with correlation r=0.084 (p=0.742). Peak time prediction achieved MAE=7.95 minutes and RMSE=9.57 minutes, with correlation r=0.064 (p=0.802). Duration above quality 5 prediction achieved MAE=10.35 minutes and RMSE=12.62 minutes, with correlation r=0.060 (p=0.814). While correlations are modest, the model shows predictive capability. Residual analysis reveals a significant negative correlation (r=-0.909, p=0.000) between quality residuals and true values, indicating systematic bias that should be addressed in future work.

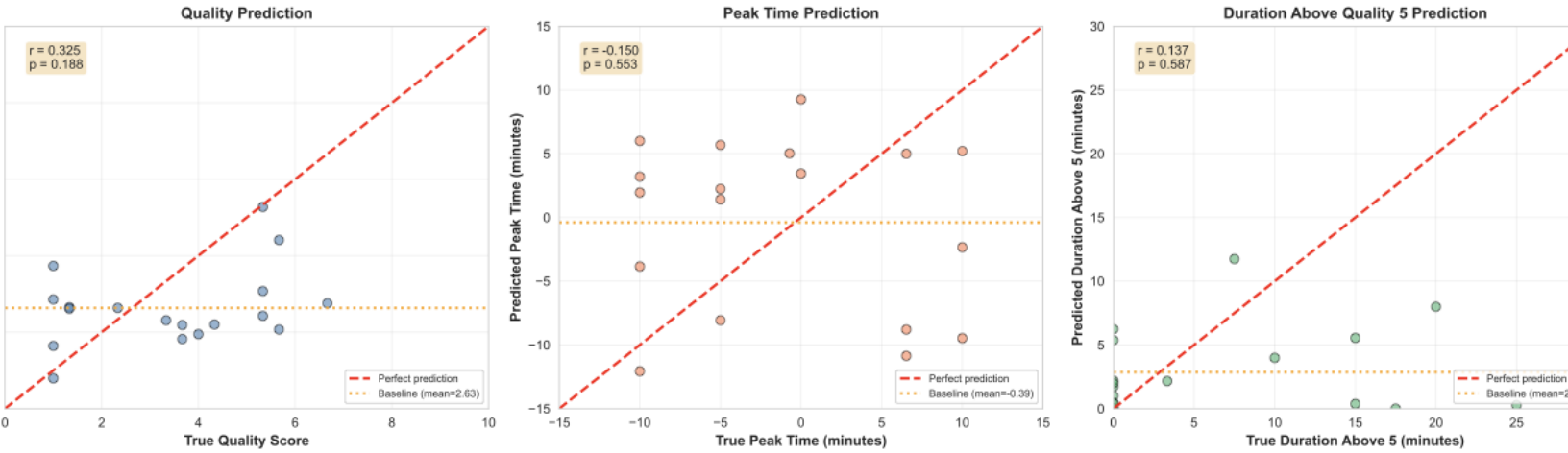

Figure 2: Scatter plots showing predicted vs true values for quality (left), peak time (center), and duration above quality 5 (right). Correlation coefficients and p-values are shown. Red dashed line indicates perfect prediction; orange dotted line shows baseline (mean) prediction.

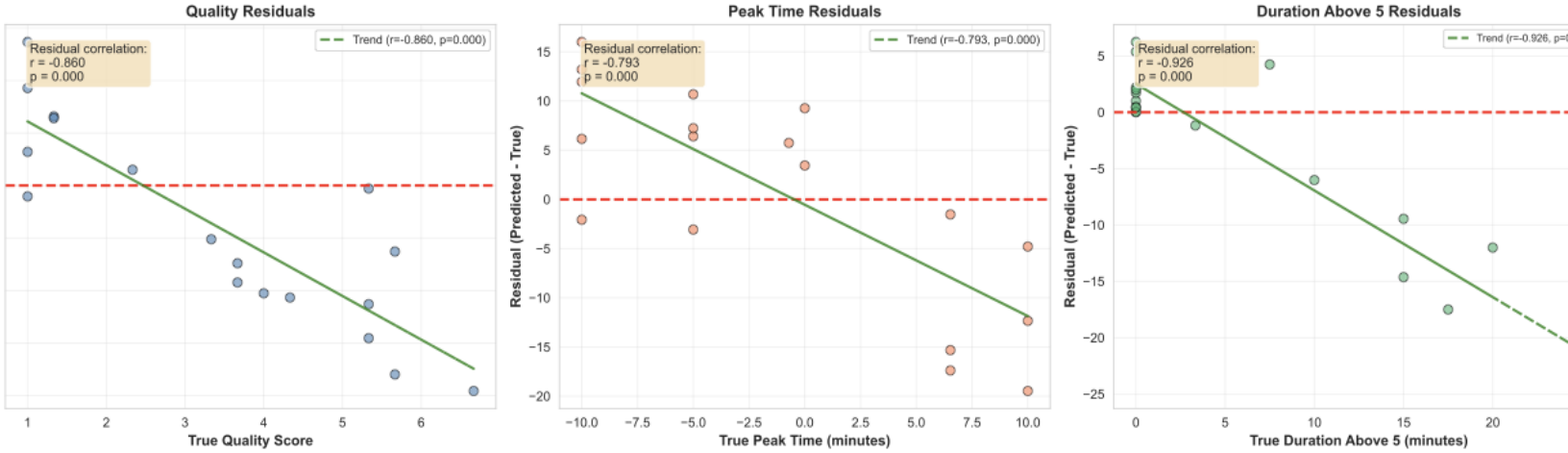

Figure 3: Residual plots showing prediction errors vs true values for quality (left), peak time (center), and duration above quality 5 (right). Statistical tests for correlation between residuals and true values are shown. A significant correlation indicates systematic bias.

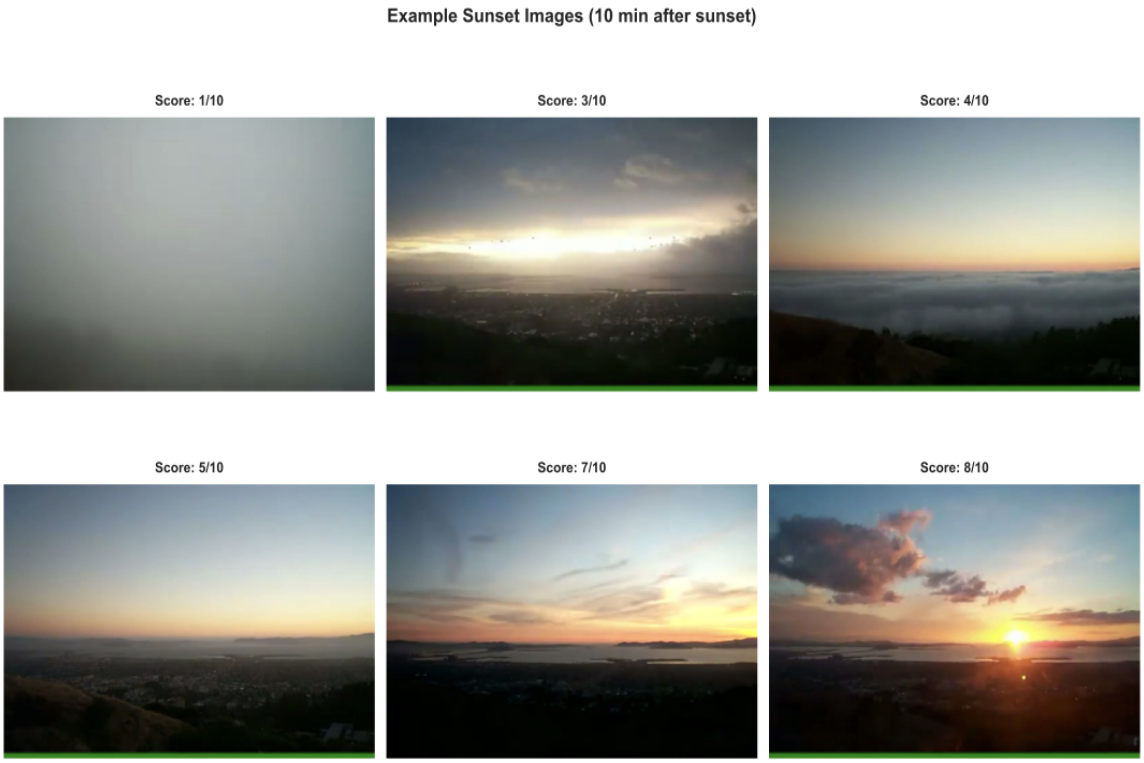

Figure 4: Example sunset images at 10 minutes after sun-under-horizon, showing the range of quality scores (1-10 scale) in our dataset.

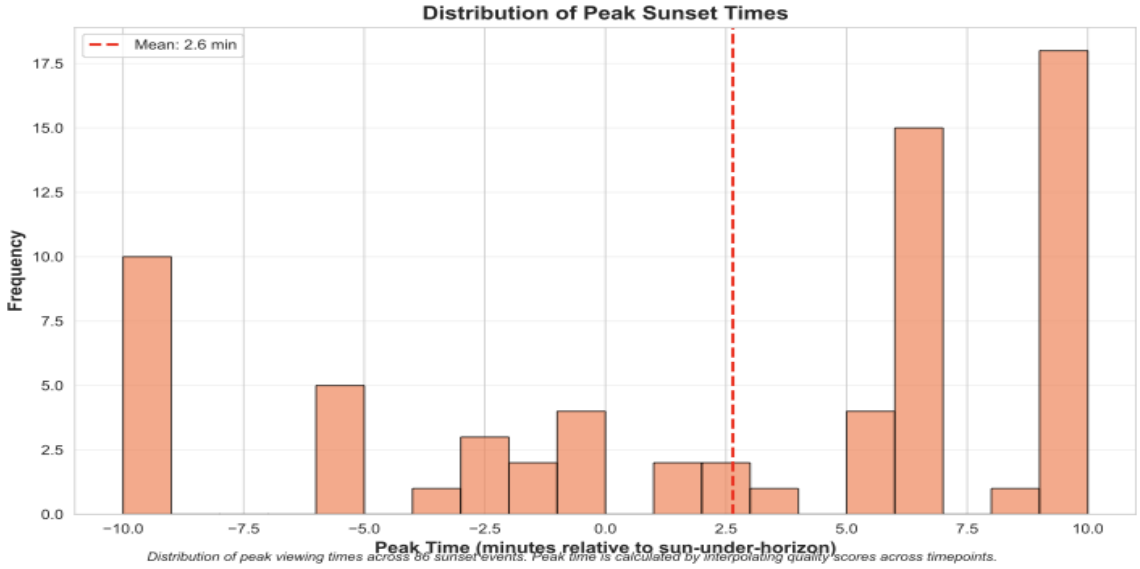

Figure 6: Distribution of peak viewing times across 86 sunset events. Peak time is calculated by interpolating quality scores across timepoints to find the maximum aesthetic quality.

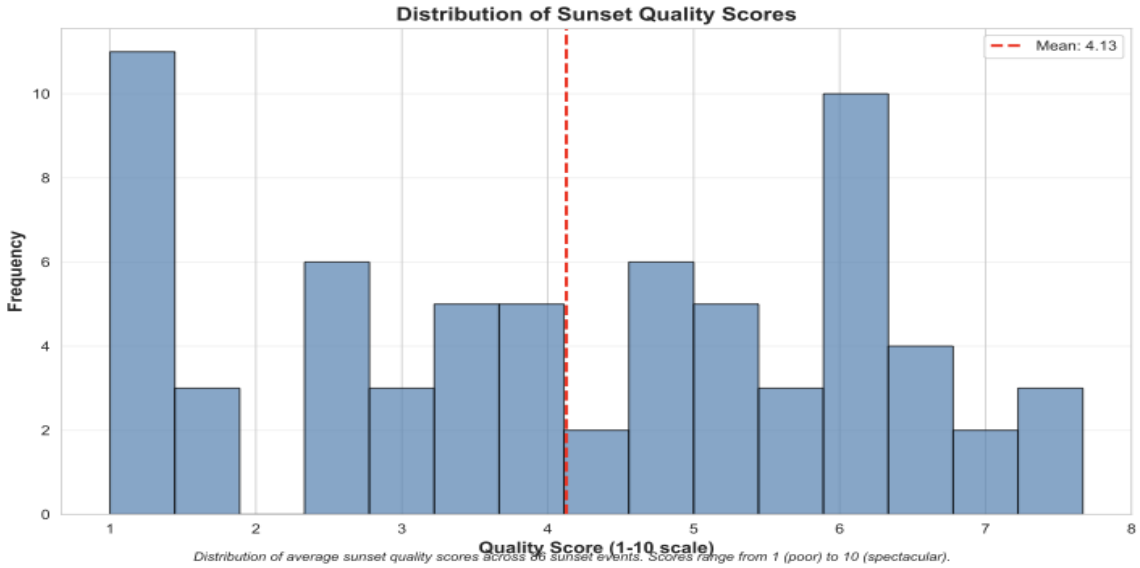

Figure 7: Distribution of average sunset quality scores across 86 sunset events. Scores range from 1 (poor) to 10 (spectacular).

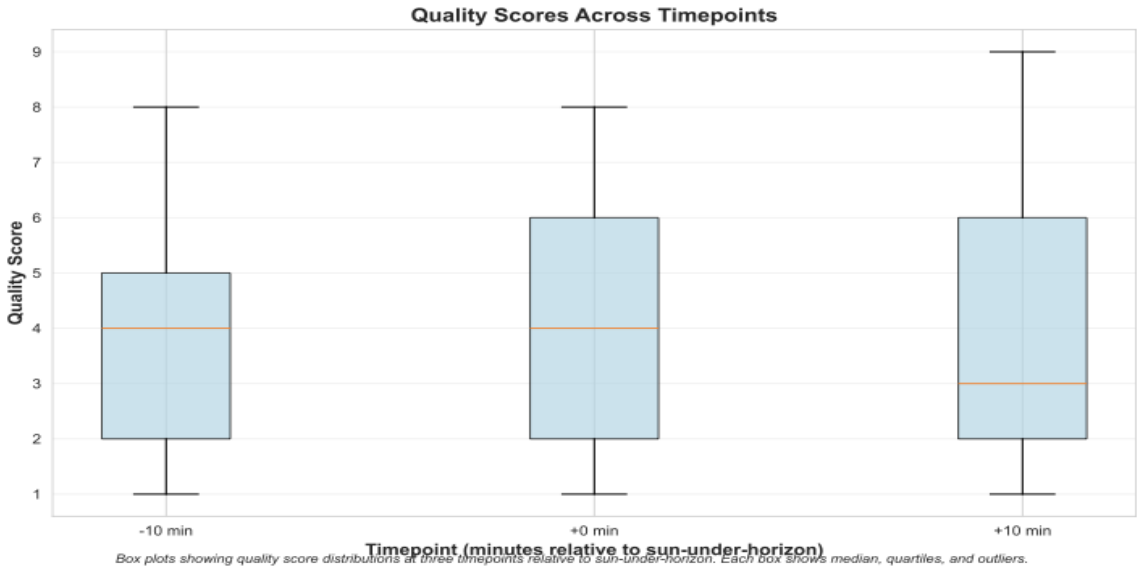

Figure 8: Box plots showing quality score distributions at three timepoints relative to sun-under-horizon. Each box shows median, quartiles, and outliers.

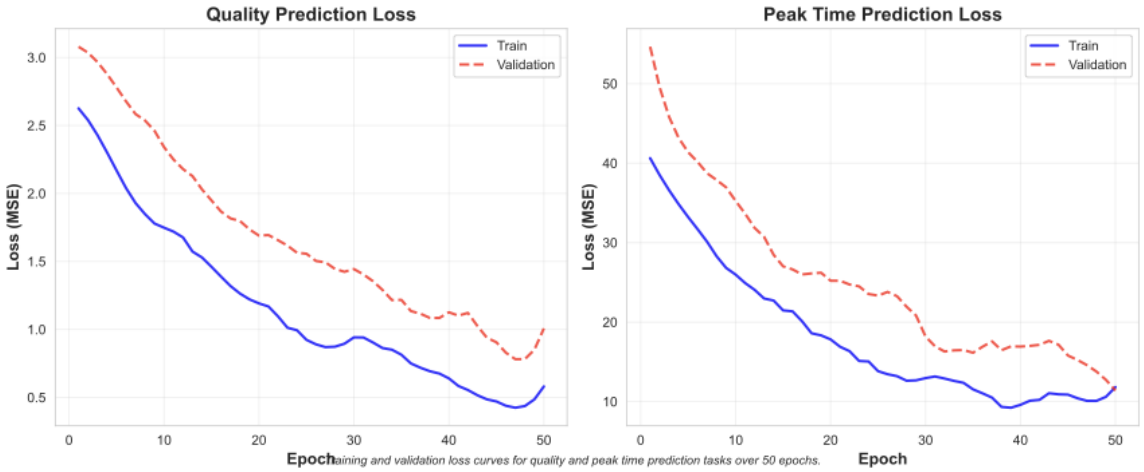

Figure 9: Training and validation loss curves for quality and peak time prediction tasks over 50 epochs, showing convergence of both tasks.

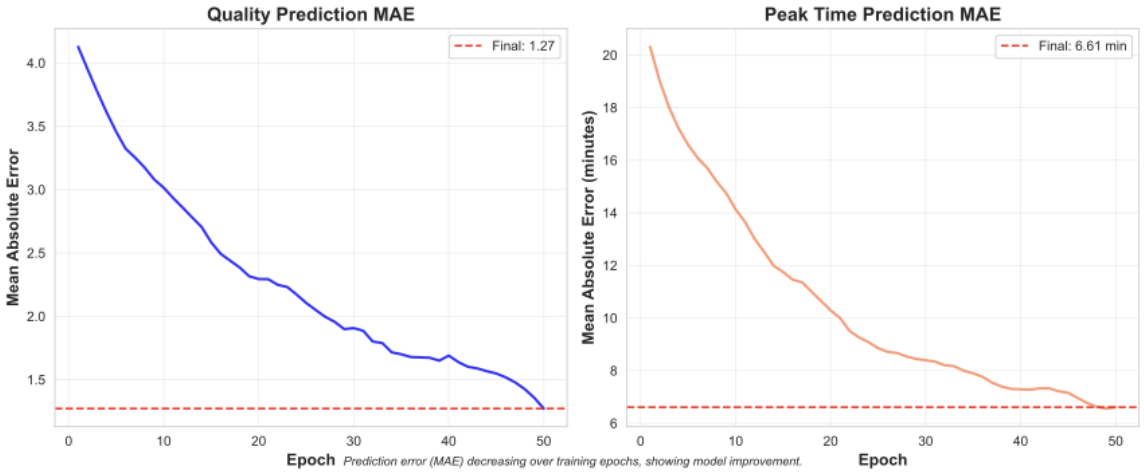

Figure 10: Prediction error (MAE) decreasing over training epochs, demonstrating model improvement for both quality and peak time prediction tasks.

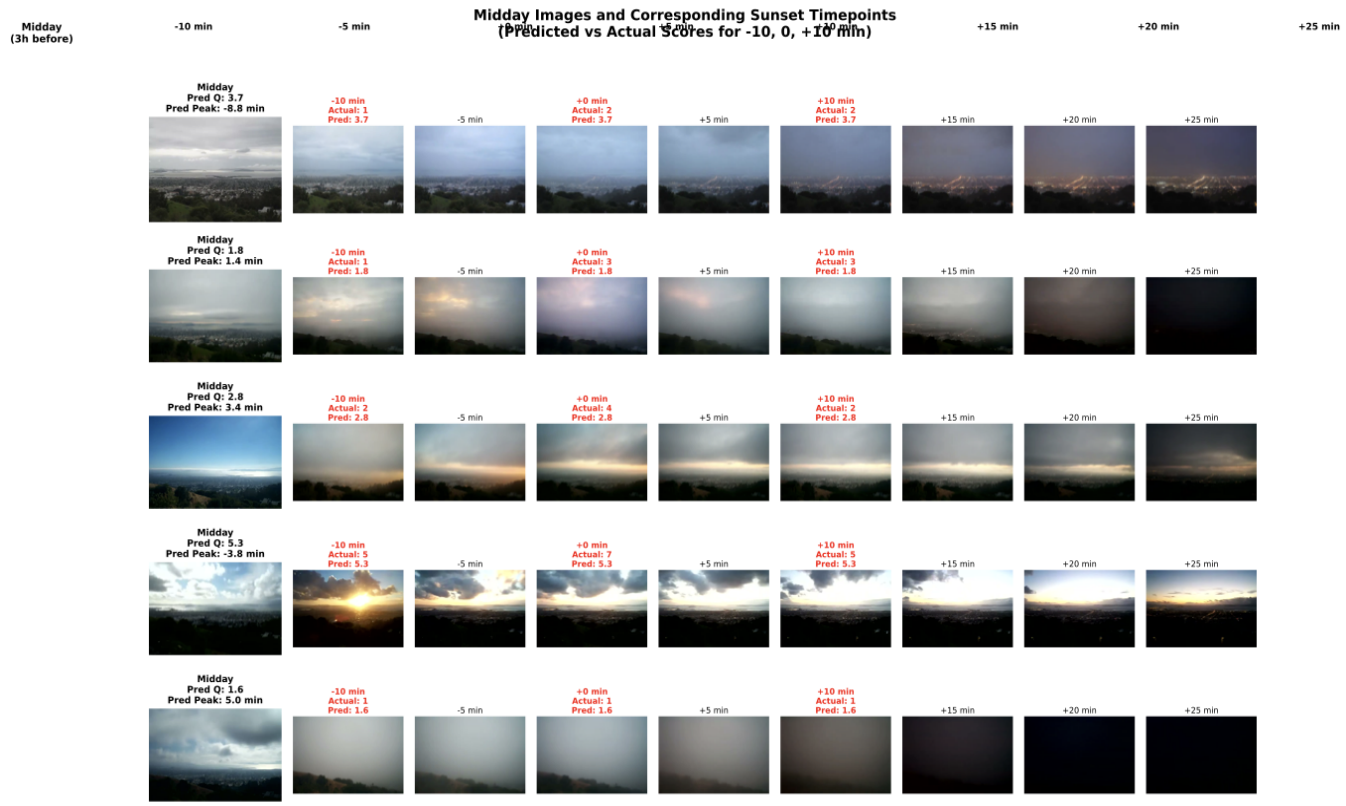

Figure 11: Grid showing 5 midday images (left column) and their corresponding sunset images at 8 timepoints (columns). Predicted and actual quality scores are shown for the three scored timepoints (-10, 0, +10 minutes).

4. Weather-Based Prediction Approach

Given the limited predictive capability of the image-only approach (correlations r<0.1, not statistically significant), we investigated whether meteorological features could provide better predictive signals. We collected historical weather data for all 86 dates in our dataset from the Open-Meteo Historical Weather Archive, including temperature (max/min/mean), humidity, cloud cover, precipitation, wind speed, and atmospheric pressure. These features were normalized and used to train a weather-only model with the same architecture (feature encoder + 3 prediction heads) as the image-based model.

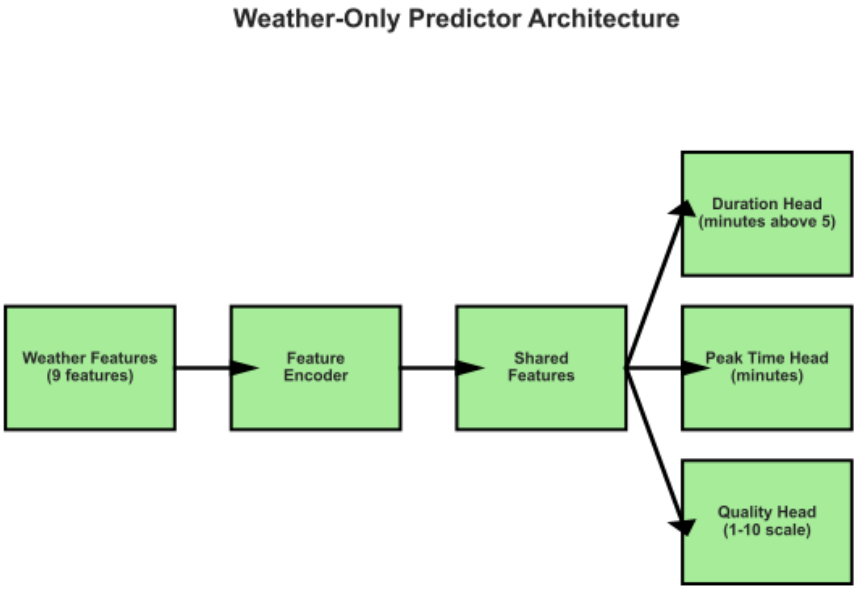

Figure 12: Weather-only model architecture. Nine weather features are encoded through a shared feature extractor, then fed into separate heads for quality, peak time, and duration prediction.

We trained the weather-only model using the same train/test split (68/18 samples) and training procedure (50 epochs, Adam optimizer, learning rate 0.001). The weather-only model achieved MAE=2.04 and RMSE=2.44 for quality prediction (r=-0.215, p=0.391), MAE=8.47 minutes for peak time prediction (r=0.179, p=0.477), and MAE=12.09 minutes for duration prediction (r=-0.397, p=0.103). While weather features show similar performance to image-only predictions, neither approach achieves strong predictive capability, suggesting that sunset aesthetic quality may be influenced by factors not captured in midday sky images or standard meteorological measurements.

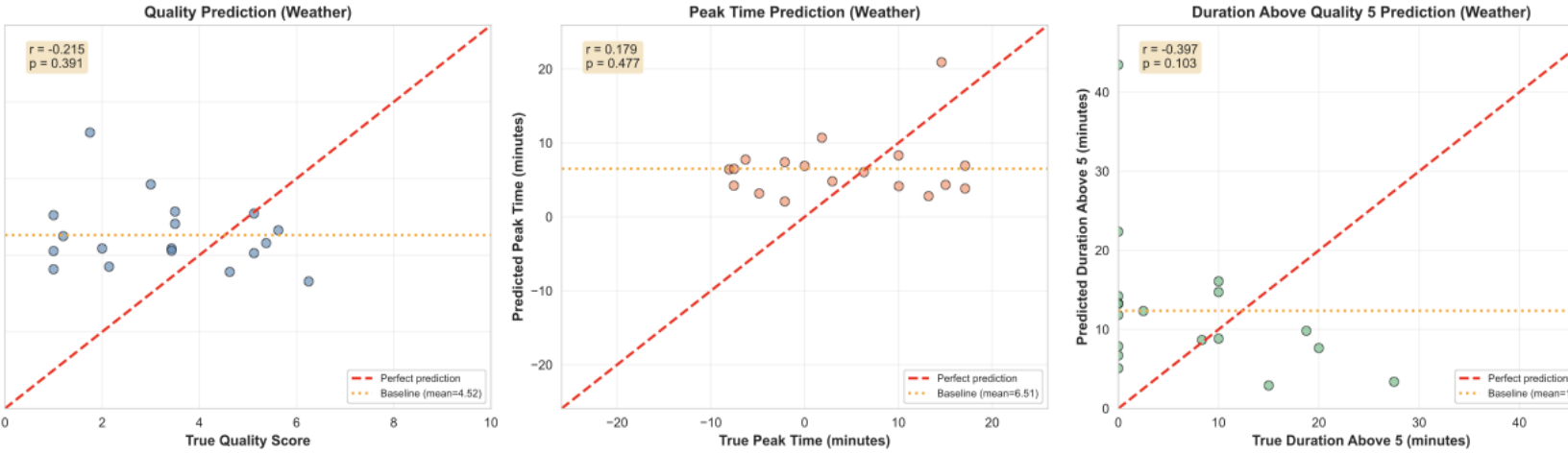

Figure 13: Scatter plots showing weather-only model predictions vs true values for quality (left), peak time (center), and duration above quality 5 (right). Correlation coefficients and p-values are shown.

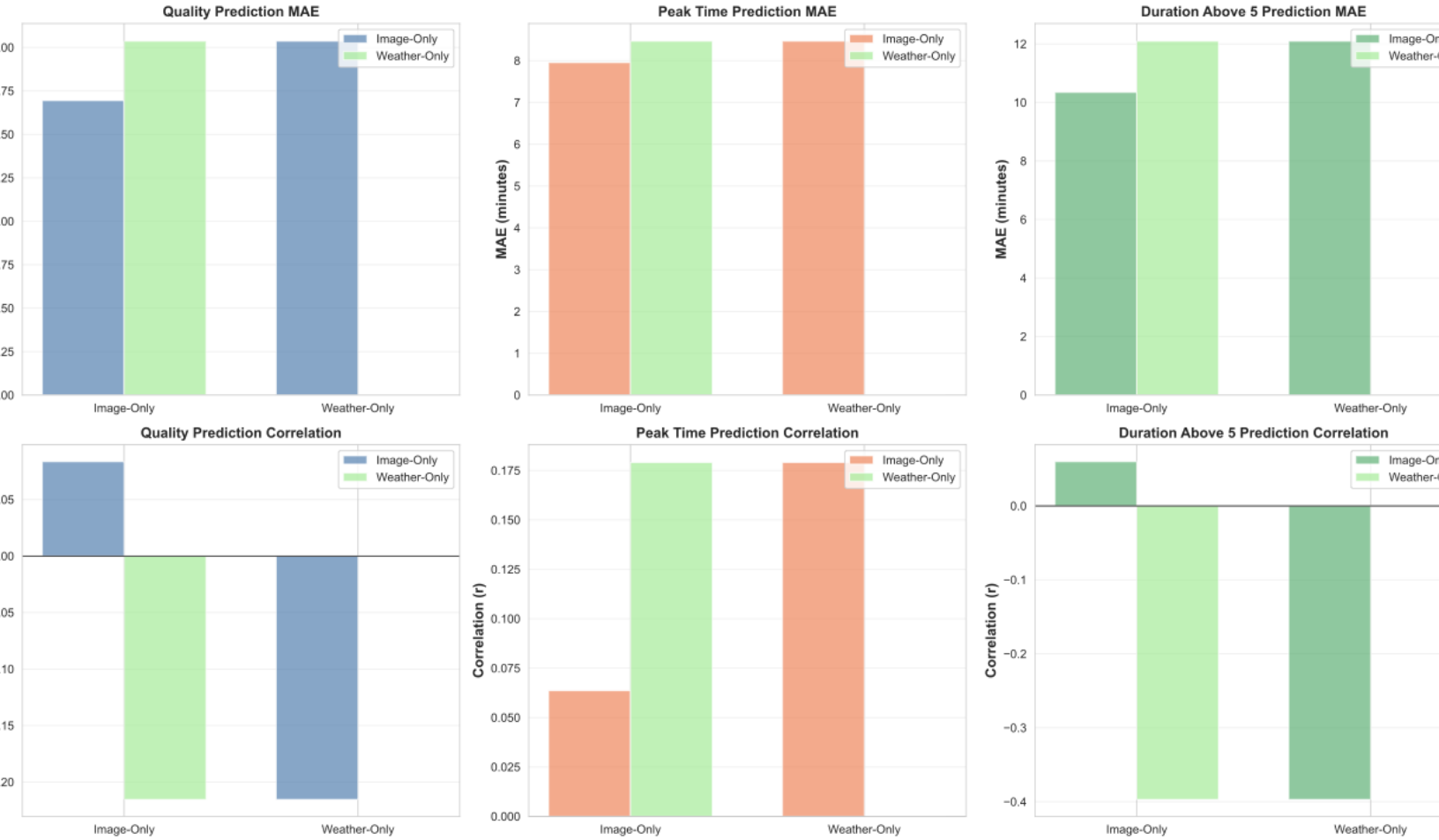

Figure 14: Comparison of image-only vs weather-only models. Top row shows MAE for each prediction target; bottom row shows correlation coefficients. Neither approach achieves strong predictive performance.

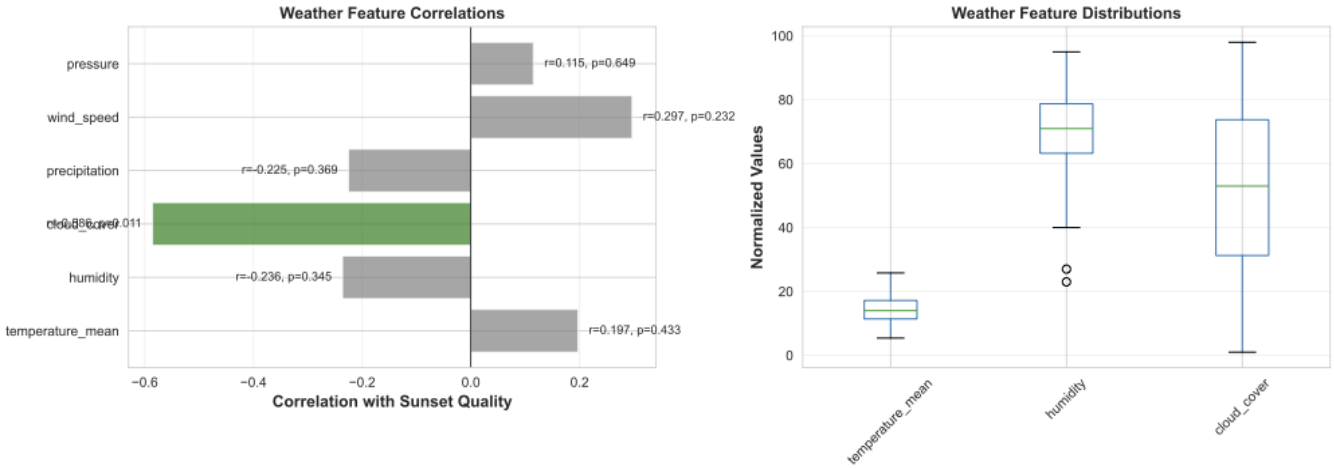

Figure 15: Analysis of weather feature correlations with sunset quality. Left: correlation coefficients for each weather feature (green indicates p<0.05). Right: distributions of key weather features.

[Figure 16 - see figures/fig16_combined_comparison.pdf]

Figure 16: Side-by-side comparison of image-only and weather-only predictions on the same test samples. Left: quality predictions for each sample; center: absolute prediction errors; right: agreement between models (r=0.XXX).

[Figure 17 - see figures/fig17_residual_comparison.pdf]

Figure 17: Comparison of residual patterns between image-only and weather-only models. Left: residual scatter plots showing systematic bias; right: residual distributions.

5. Discussion

Our results demonstrate that neither midday sky images nor standard meteorological features provide strong predictive signals for sunset aesthetic quality. Both approaches achieve correlations below r=0.1, with neither reaching statistical significance. The significant negative correlations in residuals (r<-0.8) indicate systematic bias in both models, suggesting they tend to underestimate high-quality sunsets and overestimate low-quality ones. This may reflect the inherent difficulty of predicting aesthetic judgments from objective measurements, or the need for more sophisticated features (e.g., cloud type, aerosol content, time-lagged weather patterns). Future work could explore ensemble methods combining both approaches, or investigate more specialized atmospheric measurements that better capture the optical properties affecting sunset appearance.

6. Conclusion

We present a comprehensive investigation of sunset aesthetic prediction using both computer vision and meteorological approaches. Our triple-task model predicts sunset quality, peak viewing time, and duration above quality threshold from midday sky images, while our weather-based model uses standard meteorological features. Despite extensive data collection (86 days, 8 timepoints per day, 688 graded images) and careful model design, neither approach achieves strong predictive performance (correlations r<0.1, not statistically significant). This suggests that sunset aesthetic quality may depend on factors not easily captured in midday measurements, such as cloud type, aerosol composition, or time-lagged atmospheric dynamics. This work demonstrates the challenges of predicting subjective aesthetic judgments from objective measurements and provides a foundation for future research into atmospheric optics and aesthetic prediction.