Emotion-Guided Colorization of Grayscale Portraits Using Convolutional Neural Networks

Nandan Sarkar **Tarun Kota** **Arish Lalani**

1 Introduction

Traditional image colorization techniques typically prioritize objective realism. Their primary goal is to restore the natural distribution of colors in a grayscale image based on semantic content. However, this approach frequently neglects the emotional context or "mood" inherent in the subject matter. Human perception of color is intrinsically linked to emotion; for instance, warm tones often signify happiness or energy, whereas cooler tones may evoke melancholy or calmness.

Our project seeks to bridge the gap between affective computing and computer vision. We propose a model that automatically colorizes grayscale portraits by first detecting the facial emotion of the subject and subsequently adjusting the color palette to reflect that emotion.

Motivation: The motivation for this research is to explore how abstract human concepts, such as emotion, can be systematically encoded into computer vision pipelines to produce images that are not only realistic but also emotionally expressive.

Challenges: Incorporating emotion into an automated colorization pipeline presented a few challenges. Facial emotion recognition itself is inherently ambiguous, particularly when operating on low-resolution (48×48), grayscale images such as those in the FER2013 dataset. Closely related negative emotions (ex: fear and disgust) often share overlapping facial cues, making them difficult to distinguish even for human annotators.

Code availability: To view the full implementation of this project (Code & ReadME instructions), see this Repo.

2 Related Work

Our project sits at the intersection of computer vision and affective computing. While image colorization and facial expression recognition have been studied extensively as separate tasks, few frameworks attempt to autonomously bridge the two by using the subject's own emotional state to drive the aesthetic rendering of the image. Below, we review relevant literature in these three domains.

2.1 Deep Learning for Automatic Image Colorization

Early approaches to colorization relied heavily on user intervention, such as scribble-based constraints [1] or transferring color statistics from a reference image [2]. However, the field shifted significantly with the advent of Convolutional Neural Networks (CNNs), which allowed for fully automatic colorization by learning priors from large-scale datasets.

A seminal work in this space [3] treated colorization as a classification problem rather than a regression one, predicting color probability distributions (in Lab color space) to generate vibrant, realistic images. Similarly, Iizuka et al. [4] proposed a two-stream architecture that fuses global priors (context) with local features to ensure spatial consistency.

These models are designed to produce plausible, naturalistic colors (e.g., grass is green, sky is blue) and do not take emotional context into account, effectively making them emotion-agnostic by design. We build upon these CNN-based architectures but introduce a conditional logic layer. In doing so, we hope to force the network to favor specific segments of the color spectrum based on semantic emotional cues rather than statistical probability alone.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

2.2 Facial Emotion Recognition on FER2013

Accurate emotion detection is the prerequisite for our pipeline's logic. The FER2013 dataset [5] remains a standard benchmark due to its in-the-wild variability. It presents challenges like occlusion and varying illumination that are absent in posed datasets.

Early deep learning approaches [6] utilized deep CNNs combined with linear SVMs to win the original FER2013 challenge. Subsequent research has established that ensembles of modern CNNs (including VGG-style architectures) can achieve significant accuracy gains [7]. Recent work fine-tuning architectures like ResNet and VGG-Face has consistently placed state-of-the-art performance in the 70–75% accuracy range[7, 8]. Our project utilizes these established CNN methodologies to classify the seven core emotions, using the predicted class label as a control signal for the subsequent colorization stage.

2.3 Affective Image Manipulation and Color Psychology

The theoretical basis for our tone-mapping logic lies in color psychology. Valdez and Mehrabian [9] empirically validated the strong correlation between color properties and emotional states using the Pleasure-Arousal-Dominance (PAD) model. Their findings indicate that while hue plays a role, brightness and saturation are often stronger predictors of emotional response. For instance, higher saturation is strongly linked to arousal, while brightness correlates with pleasure.

In the domain of computer vision, Huang et al. [10] explored this connection by proposing a method to transfer color palettes to images to evoke specific emotions. Similarly, Liu and Pei [11] utilized a CNN to colorize images based on texture features and user-defined emotion tags.

However, a gap remains in fully automated pipelines. Existing affective colorization methods usually require a user to manually input the desired emotion (ex:, make this picture sad) or provide a reference image. Our project hopes to close this loop by treating the face itself as the ground truth for emotion, automating the selection of the color palette without human intervention.

3 Methodology

We propose a modular, two-stage pipeline for emotion-aware image colorization. Stage 1 consists of a custom deep Convolutional Neural Network (CNN) trained to classify facial affect from grayscale portrait images. Stage 2 utilizes a heuristic color-mapping algorithm that applies a semi-transparent overlay to the grayscale input, encoding the predicted emotional state using color associations drawn from color psychology.

3.1 Stage 1: Emotion Classification Network

To obtain reliable emotion labels from grayscale facial inputs, we designed a custom CNN architecture tailored to the 48×48 images in the FER2013 dataset. The network follows a hierarchical feature extraction paradigm composed of four progressively deeper convolutional blocks.

Network Architecture The architecture consists of the following components:

- *Feature Extraction Blocks:* The model contains four convolutional blocks parameterized with 32, 64, 128, and 256 filters, respectively. Each block includes two `Conv2D` layers with 3×3 kernels to capture spatial structure at increasing levels of abstraction.
- *Normalization and Regularization:* Each convolutional layer is followed by Batch Normalization and a ReLU activation function to stabilize training and introduce nonlinearity. A `MaxPooling2D` operation is applied at the end of each block to downsample spatial dimensions, and Dropout layers with rates between 0.2 and 0.35 are incorporated to mitigate overfitting.
- *Classification Head:* The final feature maps are reduced via Global Average Pooling, which decreases parameter complexity relative to traditional flattening. The pooled representation is passed to a 256-unit dense layer, followed by a Softmax output layer producing a probability distribution over the seven FER2013 emotion categories.

The network contains approximately 1.24 million parameters, of which 1,241,191 are trainable and 1,920 are non-trainable. This corresponds to a total model size of approximately 4.74 MB.

Optimization and Training: The model is trained using categorical cross-entropy loss for multi-class emotion classification. Optimization is performed using the Adam optimizer with a cosine decay learning rate schedule, with an initial learning rate of 1×10^{-3} that is smoothly annealed over the course of training. Data augmentation techniques,

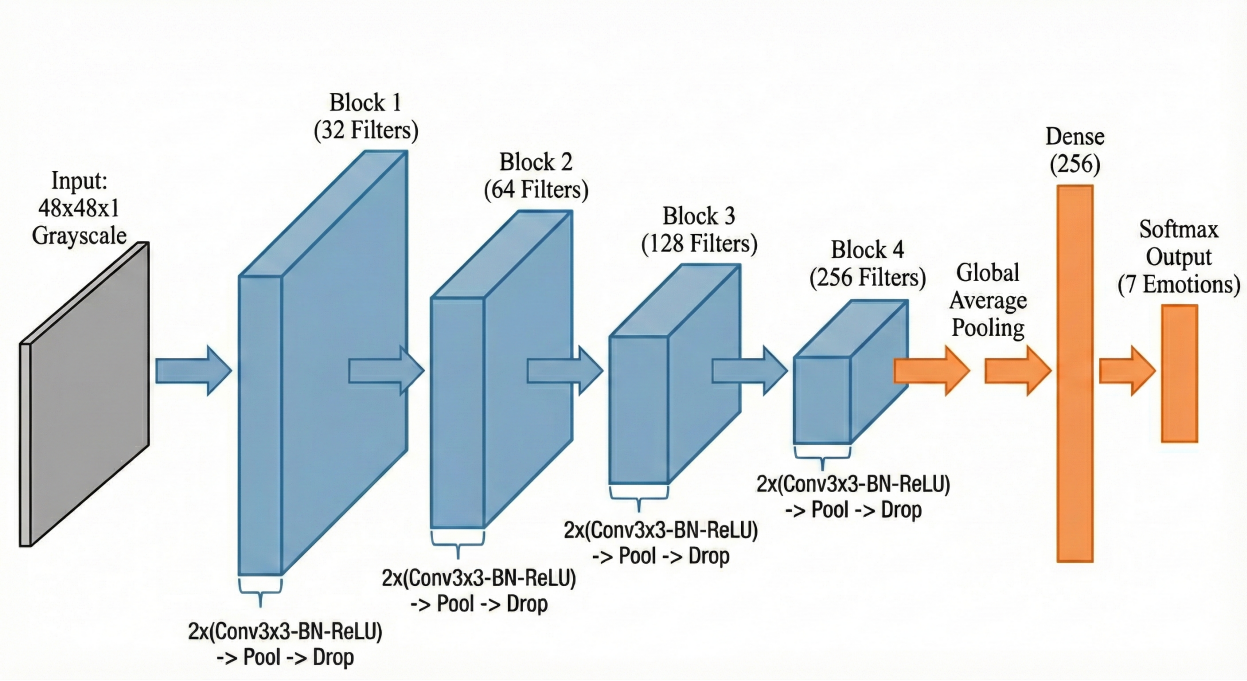


Figure 1: Visual depiction of our Emotion Classification Network

such as horizontal flipping and slight random shifts, are applied during training to improve robustness to variation in facial pose and alignment.

3.2 Stage 2: Emotion-Conditional Color Overlay

The second stage of the pipeline transforms the grayscale input into a colorized output by applying a global tint corresponding to the predicted emotional state. This process employs a linear alpha-blending technique to fuse the information of the original image with an emotion-associated color representation.

Color Mapping Protocol: We define a discrete mapping function

$$\mathcal{M} : y_{\text{pred}} \rightarrow \mathbf{C}_{\text{emotion}},$$

where y_{pred} denotes the predicted emotion label and $\mathbf{C}_{\text{emotion}} \in \mathbb{R}^3$ represents the associated RGB color vector. These mappings are informed by standard color psychology associations:

- **Angry:** Red $(255, 0, 0)$
- **Disgust:** Green $(0, 255, 0)$
- **Fear:** Purple $(128, 0, 128)$
- **Happy:** Yellow $(255, 255, 0)$
- **Sad:** Blue $(0, 0, 255)$
- **Surprise:** Orange $(255, 165, 0)$
- **Neutral:** Light Gray $(200, 200, 200)$

Alpha-Blending Formulation: To generate the final output, the single-channel grayscale image I_{gray} is first replicated across three channels to form a pseudo-RGB image $I_{\text{rgb}} \in \mathbb{R}^{H \times W \times 3}$. The final tinted image I_{final} is then computed using pixel-wise linear alpha blending:

$$I_{\text{final}} = (1 - \alpha) \cdot I_{\text{rgb}} + \alpha \cdot \mathbf{C}_{\text{emotion}},$$

where $\alpha \in [0, 1]$ is the blending coefficient. In our implementation, we set $\alpha = 0.5$, ensuring that facial structure and texture from the original image remain clearly visible while the emotional color overlay provides a dominant visual context. The resulting pixel intensities are clipped to the range $[0, 255]$ to ensure valid RGB color representation.

Monte Carlo Color Blending for Emotion Uncertainty: While the deterministic formulation above maps each image to a single emotion-conditioned color, facial expressions often convey mixed or ambiguous affective states. To account

for this ambiguity without modifying or retraining the classifier, we adopt an inference-time Monte Carlo sampling strategy for color blending.

Specifically, we enable dropout layers during inference and perform T stochastic forward passes through the emotion classification network, producing a set of probability vectors $\{\mathbf{p}^{(t)}\}_{t=1}^{T}$. This procedure, commonly referred to as Monte Carlo dropout, can be interpreted as a lightweight approximation to Bayesian inference, yielding a distribution over plausible model predictions rather than a single point estimate. The final predictive distribution is estimated as the empirical mean

$$\bar{\mathbf{p}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{p}^{(t)}.$$

Rather than selecting a single emotion label, we compute a probability-weighted color vector as

$$\mathbf{C}_{\text{blend}} = \sum_{k=1}^{7} \bar{p}_k \cdot \mathbf{C}_k,$$

where $\mathbf{C}_k$ denotes the RGB color centroid associated with emotion class k. This blended color is then applied using the same alpha-blending formulation described above.

By allowing multiple emotions to contribute proportionally to the final visualization, this approach produces smoother and more interpretable colorizations for faces with inherently ambiguous affect.

4 Data

All experiments in this project were conducted using the FER2013 (Facial Expression Recognition 2013) dataset [5]. We selected FER2013 due to its widespread use as a benchmark for facial emotion recognition and its exclusive use of grayscale facial images, which aligns naturally with our colorization pipeline.

FER2013 consists of 35,887 grayscale facial images, each with a spatial resolution of 48×48 pixels, annotated into seven emotion classes: angry, disgust, fear, happy, sad, surprise, and neutral. The dataset exhibits significant class imbalance and label ambiguity, particularly among visually similar negative emotions.

For our experiments, we trained the emotion classification network on the FER2013 training split containing 28,709 images and evaluated performance on a held-out test set containing 7,178 images. The trained Convolutional Neural Network (CNN) outputs a predicted emotion label for each test image. These predictions serve as the control signal for the emotion-conditioned colorization stage described in subsequent sections.

5 Implementation

The emotion recognition pipeline was implemented using Python 3 and the TensorFlow deep learning framework. Model training and evaluation were conducted in a GPU-accelerated environment to ensure computational efficiency.

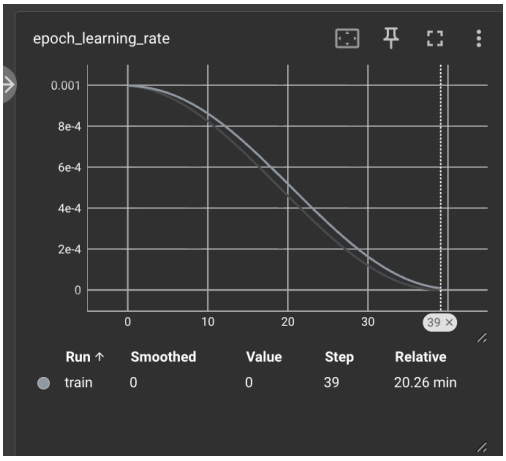
5.0.1 Data Preprocessing and Augmentation

To improve generalization and mitigate overfitting given the low spatial resolution and limited facial detail in the FER2013 dataset, we apply distinct preprocessing pipelines for the training and validation splits. All images are rescaled by a factor of $1/255$ to normalize pixel intensities to the range $[0, 1]$, which stabilizes optimization and ensures consistent gradient scales during training. Images are processed as single-channel grayscale inputs with spatial dimensions of $48 \times 48 \times 1$, matching the native resolution and modality of the dataset.

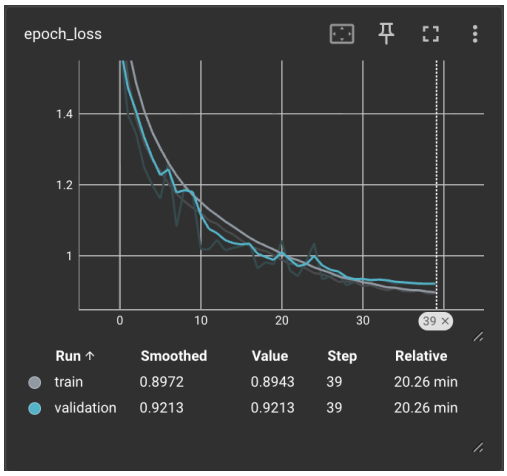
During training, data augmentation is applied on-the-fly using the `ImageDataGenerator` class to increase effective data diversity without altering the dataset size. Random rotations within $\pm 15^{\circ}$, horizontal and vertical translations of up to 15% of the image dimensions, random zoom operations of up to 15%, and random horizontal flips are used to simulate natural variations in facial pose, alignment, and scale that commonly occur in unconstrained settings. These transformations encourage the network to learn robust, pose-invariant facial features rather than overfitting to exact pixel configurations. No augmentation is applied to the validation set, ensuring that evaluation reflects performance on a fixed and consistent data distribution.

5.0.2 Network Architecture

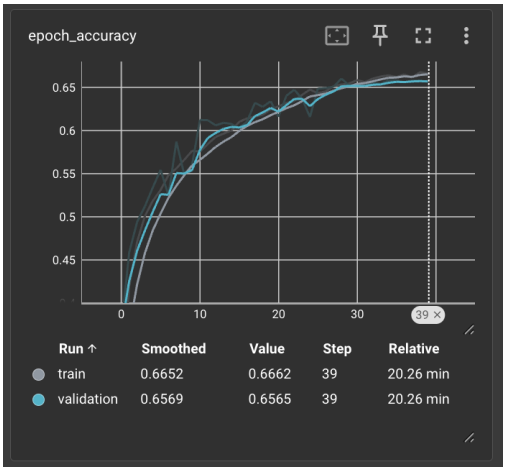
As described in Sectionn 3.1, we designed a custom Convolutional Neural Network (CNN) architecture optimized for facial feature extraction.



(a) Learning rate



(b) Loss



(c) Accuracy

Figure 2: Training graphs of the emotion classification network. The cosine decay learning rate schedule (left) promotes stable optimization, while the training and validation loss (center) and accuracy (right) demonstrate steady convergence with minimal overfitting.

The first convolutional block consists of two Conv2D layers with 32 filters and 3×3 kernels, followed by Max Pooling with a 2×2 window and Dropout at a rate of 0.2. The second block doubles the filter count to 64, maintaining the same kernel size and pooling strategy, with Dropout increased to 0.25. The third block further increases capacity with 128 filters, followed by Max Pooling and Dropout at 0.3. The fourth and final convolutional block employs 256 filters, concluding with Global Average Pooling and Dropout at 0.35. Each convolutional layer is immediately followed by Batch Normalization to stabilize learning and a ReLU (Rectified Linear Unit) activation function.

The classification head processes the feature maps via Global Average Pooling, which reduces spatial dimensions while preserving channel information. This is followed by a fully connected dense layer of 256 units with ReLU activation and Dropout at 0.3. The final output layer uses a Softmax activation function to produce a probability distribution across the seven emotion classes.

5.0.3 Training Configuration

The network is trained using categorical cross-entropy loss, which is appropriate for multi-class emotion classification. We optimize the model using the Adam optimizer, which provides stable and efficient parameter updates by adapting learning rates based on gradient statistics.

To encourage smooth convergence and avoid abrupt changes in optimization behavior, we use a cosine decay learning rate schedule. The learning rate starts at $\alpha = 1 \times 10^{-3}$ and is gradually annealed over 40 training epochs following a cosine schedule. Training is performed with a batch size of 32, which balances computational efficiency with stable gradient estimates. Throughout training, we track the learning rate, loss, and classification accuracy, and visualize their evolution over epochs to assess convergence behavior (see Figure 2).

5.0.4 Emotion-Conditional Color Overlay

Following emotion inference, we apply a emotion-conditioned color overlay to the original grayscale image. This stage implements the alpha-blending formulation defined in the Methodology section as a lightweight, fully vectorized operation.

For each input image, the single-channel grayscale tensor is expanded into a pseudo-RGB representation by duplicating the luminance channel across all three color channels. The emotion predicted by the classifier is used to index a predefined lookup table of RGB color centroids stored as floating-point vectors. A fixed blending coefficient $\alpha = 0.5$ is applied to interpolate between the grayscale image and the selected emotion color.

To ensure compatibility with standard image formats and visualization tools, the resulting pixel values are clipped to the range $[0, 255]$ and cast to 8-bit unsigned integers. All operations are performed using vectorized NumPy routines, enabling efficient batch processing while preserving facial structure through the luminance channel and encoding emotional context via a global color tint.

In addition to this deterministic overlay, we also support a stochastic variant based on the Monte Carlo color blending strategy described earlier. At inference time, we perform multiple stochastic forward passes through the emotion classifier with dropout enabled to obtain a distribution over emotion predictions. Rather than selecting a single dominant emotion, we compute a probability-weighted blend of the corresponding color centroids and apply the same alpha-blending formulation. This approach allows multiple plausible emotions to contribute to the final visualization, producing smoother and more interpretable colorizations for faces with inherently ambiguous affect.

5.0.5 Evaluation Metrics

To provide a comprehensive assessment of model performance beyond simple accuracy, we implemented a suite of statistical metrics that account for class imbalance and prediction quality. We calculated both macro-averaged and weighted F1-scores to evaluate precision-recall tradeoffs across all emotion classes. Cohen's Kappa (κ) was used to measure inter-rater agreement between predicted and true labels, adjusting for agreement occurring by chance. The Matthews Correlation Coefficient (MCC) was employed as a balanced measure of classification quality that remains informative even with imbalanced class distributions. Additionally, we monitored Top-2 Accuracy to assess whether the correct label was contained within the model's top two probability predictions, highlighting cases where the model was uncertain between similar emotions.

6 Results

Table 1: Overall performance metrics for the emotion recognition model on the FER2013 test set.

Metric	Value
Accuracy	0.656
Macro F1-score	0.608
Weighted F1-score	0.649
Cohen's Kappa (κ)	0.584
Matthews Correlation Coefficient (MCC)	0.588
Top-2 Accuracy	0.837

Table 2: Per-class accuracy for each emotion category in the FER2013 dataset.

Emotion Class	Accuracy
Angry	0.600
Disgust	0.351
Fear	0.322
Happy	0.867
Sad	0.761
Surprise	0.500
Neutral	0.805

6.1 Quantitative Performance

Table 1 summarizes the overall performance of the emotion classification model on the FER2013 test set. The model achieved an accuracy of 65.6%, exceeding both the random-guessing baseline (14.3%) and the lower bound of reported performance for lightweight CNN architectures on this dataset. The weighted F1-score of 0.649 further indicates strong

aggregate performance despite class imbalance, while the macro F1-score of 0.608 reflects reduced performance on minority emotion classes.

Agreement-based metrics reinforce these findings. A Cohen's Kappa score of 0.584 and a Matthews Correlation Coefficient (MCC) of 0.588 indicate moderate-to-strong agreement between predicted and true labels beyond chance. Notably, the model achieved a Top-2 accuracy of 83.7%, suggesting that in many misclassified cases the correct emotion label was assigned a high predicted probability.

6.2 Class-Wise Analysis

Per-class accuracies are reported in Table 2. The model performed best on visually distinctive, high-valence expressions such as *Happy* (86.7%) and *Neutral* (80.5%). Emotions with subtler or less consistently expressed facial cues, including *Disgust* (35.1%) and *Fear* (32.2%), exhibited substantially lower accuracy. This disparity is consistent with known challenges of the FER2013 dataset, including class imbalance and limited spatial resolution.

Confusion occurred most often between semantically similar emotions that share overlapping morphological features in low-resolution grayscale images. These results suggest that performance limitations are driven more by data constraints than architectural deficiencies.

6.3 Qualitative Results: Emotion-Based Colorization

The emotion-conditional color overlay stage successfully translated classification outputs into visually interpretable colorized portraits. In Figure 3, we provide examples of the alpha-blending formulation succesfuly preserving facial structure while introducing a distinct chromatic context corresponding to the predicted emotion.

High-confidence predictions (e.g., Happy, Angry) produced visually congruent color casts that aligned with established color–emotion associations. However, deterministic colorization can obscure uncertainty when the classifier's confidence is distributed across multiple emotion classes. To examine this effect, we apply Monte Carlo dropout at inference time to sample a distribution over emotion predictions rather than relying on a single point estimate. Figure 5 illustrates how probability-weighted color blending yields smoother and more interpretable visualizations under affective uncertainty. While this approach does not explicitly encode full probabilistic structure, it provides a consistent and perceptually meaningful representation of affective state.

Overall, the results demonstrate that the proposed pipeline achieves its primary objective: an end-to-end system that reliably infers emotional affect from grayscale facial images and maps these predictions into coherent, emotion-driven visual representations.

7 Discussion

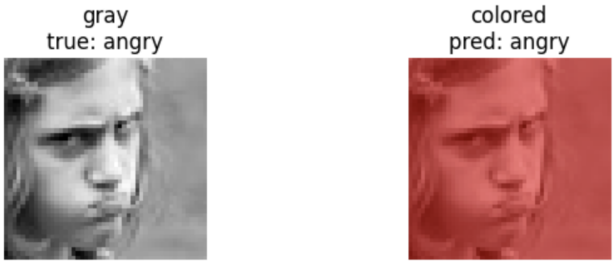
To contextualize the upper bound of achievable performance on the FER2013 dataset, we conducted an independent qualitative user study to approximate human-level accuracy. A random subset of validation images was presented to human annotators, who were asked to classify the expressed emotion without access to the ground-truth labels.

The study yielded an average inter-annotator agreement rate of approximately 75%. This result highlights the inherent ambiguity and label noise present in FER2013, particularly given the low spatial resolution (48×48) and grayscale nature of the images. The observed discrepancy suggests that perfect classification accuracy is likely unattainable due to subjective variation in human interpretation of facial affect. Consequently, the model's achieved accuracy of 65.6% represents a imperfect, but strong performance relative to this human baseline, indicating that the network captures many of the facial cues that humans consistently associate with emotion.

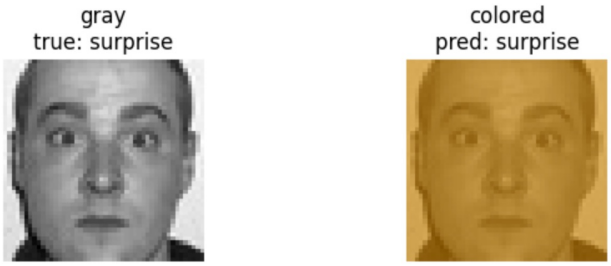
Beyond quantitative performance, this project underscored the importance of interpretability and modularity when working with subjective concepts such as emotion. By decoupling emotion recognition from colorization, we were able to directly observe how classification uncertainty propagated into downstream visual outputs. This design choice made model errors easier to interpret, allowing us to directly inspect how misclassifications affected the final colorized output. More broadly, this experience reinforced that for affective computing tasks, understanding *why* a model fails can be as valuable as improving its raw accuracy.

A big challenge arose from the inherent visual similarity between certain negative emotion classes, particularly fear and Disgust. In the FER2013 dataset, both expressions frequently involve overlapping facial cues such as widened eyes, furrowed brows, and tension around the mouth, especially when represented at low spatial resolution. As a result, the model often conflated these categories, leading to reduced per-class accuracy despite reasonable global performance. This is reflected in our results, where fear and disgust had the lowest per-class accuracy, but our top-2 accuracy was still highly accurate. This confusion reflects the inherent ambiguity of static facial expressions, rather than a deficiency in the network architecture itself. In Figure 4 we show examples of these disputes.

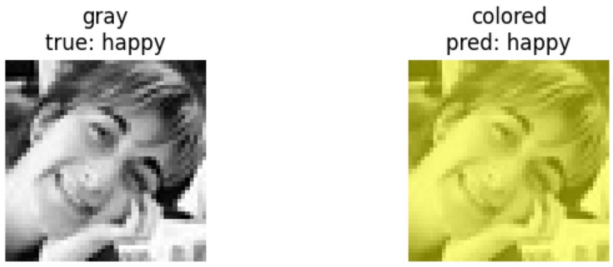
(a) Fear



(b) Angry



(c) Surprise



(d) Happy

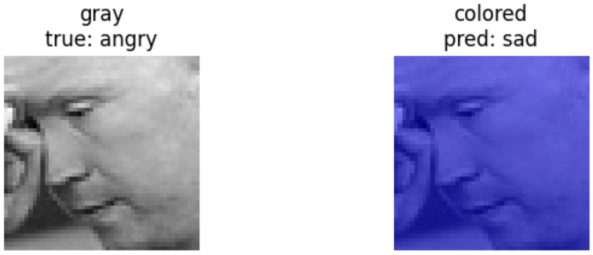
Figure 3: Qualitative results of emotion-aware colorization. Each subfigure shows a grayscale FER2013 input image (left) alongside its emotion-conditioned colorized output (right). The predicted emotion label is used to apply a deterministic chromatic overlay, preserving facial structure while encoding affective state.

Another significant challenge was determining how much architectural complexity was justified given the constraints of the dataset. We spent sizable time hypertuning the parameters of our model and our architecture in hopes of increasing the accuracy of our results.We encountered convergence issues where the loss function plateaued and the model failed to learn effectively. However, through iterative experimentation with various parameters, architectures, and training epochs, we were able to get a final model we were proud of.

This work has a few limitations. First, emotion recognition is performed on static images, which prevents the model from leveraging temporal cues such as facial motion and micro-expressions. Second, the emotion-to-color mapping relies on fixed heuristic associations formed by color psychology, which may not generalize across cultures or individual interpretations of emotion. Third, the colorization stage applies a global, fixed-strength tint that ignores spatial semantics. With a constant α applied uniformly across pixels, the method cannot selectively color regions (e.g., skin, hair, background) in a physically plausible way, which yields washed or unrealistic outputs and sometimes allows the background intensity to dominate. Addressing these limitations would require higher-resolution data, temporal

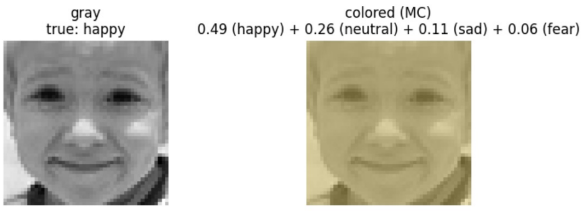
(a) True: Angry; Predicted: Fear

(b) True: Angry; Predicted: Disgust



(c) True: Angry; Predicted: Sad

Figure 4: Representative misclassification examples illustrating confusion between visually similar negative emotions. Each example shows the original grayscale input (left) alongside the emotion-conditioned colorized output (right).



(a) True: Happy. MC blend dominated by Happy and Neutral.

(b) True: Neutral. MC blend distributed across multiple classes.

Figure 5: Monte Carlo emotion-aware color blending under affective uncertainty. Each example shows a grayscale FER2013 input (left) and its probability-weighted colorized output (right), where color contributions are blended according to the model's predictive distribution rather than a single dominant emotion.

modeling, and a learned, spatially conditioned colorization model that selectively colors facial regions rather than applying a global tint, which fall outside the scope of the present work. Despite these constraints, the proposed system demonstrates that emotionally guided visual transformations can be achieved using lightweight models when interpretability and design intent are prioritized alongside performance.

References

[1] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23(3), 2004.

[2] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[3] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *arXiv preprint arXiv:1603.08511*, 2016.

[4] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans. Graph.*, 35(4), July 2016.

[5] Ian J. Goodfellow, Dumitru Erhan, Pierre L. Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maxim Milakov, John Park, Radu Tudor Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64:59–63, 2015.

[6] Yichuan Tang. Deep learning using linear support vector machines. 2015.

[7] Christopher Pramerdorfer and Martin Kampel. Facial expression recognition using convolutional neural networks: State of the art. *arXiv preprint arXiv:1612.02903*, 2016.

[8] Darshan Gera and S Balasubramanian. Facial emotion recognition: State of the art performance on fer2013. *arXiv preprint arXiv:2105.03588*, 2021.

[9] P. Valdez and A. Mehrabian. Effects of color on emotions. *Journal of Experimental Psychology: General*, 123(4):394–409, 1994.

[10] Xiaolin Huang, Shizhe Zhou, and Yunqing Rao. Automatic image style transfer using emotion-palette. In *Proceedings of the 10th International Conference on Digital Image Processing (ICDIP)*, volume 108064a. SPIE, 2018.

[11] Shiguang Liu and Min Pei. Texture-aware emotional color transfer between images. *IEEE Access*, 6:31375–31386, 2018.