Concordia University
Department of Computer Science & Software Engineering

COMP 472
Artificial Intelligence (Winter 2024)
Education MindA.I.lytics
Report: Part I

Submitted To
Dr. René Witte

Team: AK_16
https://github.com/RevoSam/COMP472

We certify that this submission is the original work of members of the group and meets the Faculty's Expectations of Originality

Sami Najim

Sami Najim
21289640
Data Specialist

Fida Inthesar

Fida Inthesar
40132564
Training Specialist

Atul Gupta

Atul Gupta
40188227
Evaluation Specialist

Dated: March 10, 2024

Contents

Emotions

Our team decided to analyze the emotions: *Neutral, Focused, Surprised, Happy*

Datasets

Our team believed that a great dataset for facial emotion recognition should balance diversity in expressions, backgrounds, and participants. Controlled environments, frontal angles, and diverse demographics were essential in dataset selection. Each dataset presented unique challenges and strengths, contributing to our model's robustness and effectiveness. Careful consideration of factors such as bias and ethical considerations was paramount to ensure the model's applicability across varied scenarios and demographics. A summary of all datasets can be found in Table 1.

Our first dataset is the AMFD dataset [1], containing approximately 200 images evenly distributed across the emotion groups. Each participant's contribution of both neutral and happy expressions, along with consistent frontal poses against a neutral white background, helps to eliminate external influences. This dataset was reviewed minimally since its goal of capturing facial emotions aligns with our objectives.

The CK dataset [2,3] consists of 500 images portraying neutral, happy, surprised, and focused expressions. All images are in black and white, offering a consistent aesthetic yet requiring careful review to remove slight redundancies. The distinction between neutral and focused expressions presents a significant challenge, necessitating meticulous curation during model training. However, CK's diverse set of expressions captured in a controlled environment makes it an invaluable resource for training models to recognize emotions effectively.

The FFHQ dataset [4], characterized by emotions like happiness, neutrality, and surprise, adds complexity with its diverse backgrounds. The abundance of neutral and surprised expressions, compared to happy ones, complements what we lacked in other datasets. The inclusion of non-frontal poses and an imbalanced distribution of emotions may challenge the model's adaptability to real-world scenarios, highlighting the need for adaptability to different poses and backgrounds to enhance the model's robustness.

The OSF Indian Database [5], featuring 180 participants, showcases happy, focused, and surprising emotions against a white background, adding diversity to the dataset. The controlled environment simplifies emotion verification, and the clear portrayal of expressions strengthens the dataset. However, its specific demographic focus necessitates broader population considerations to prevent model bias toward specific ethnic characteristics.

The TFEID Asian database [6], characterized by close-up-colored images with minimal background, offers a concise and focused dataset for facial emotion recognition. The minimal background interference enhances emotion clarity, creating an ideal training set for our project. However, like the Indian database, we considered potential biases due to its regional specificity to ensure the model's generalizability across diverse demographics.

The Real and Fake Face Detection dataset [7], encompassing all emotion groups, presents significant challenges in filtering from a larger database with cropped pictures of individuals' faces. Due to the uncontrolled environment and factors like odd angles, many images were removed from the dataset. Despite these challenges, its real-world scenarios highlight its relevance, but the curation process is crucial for ensuring model effectiveness and avoiding bias during data selection.

Lastly, the TDP dataset [8], featuring around fifteen participants expressing happiness, surprise, and neutrality against varying backgrounds, offers a small-scale yet diverse dataset. It provides a glimpse into real-world scenarios without compromising our training set goals, despite its smaller size.

Dataset	Sample Image	Licensing
AMFD		Creative Commons Attribution 4.0 International Public License
CK		Custom non commerical
FFHQ		Creative Commons BY 2.0, Creative Commons BY-NC 2.0, Public Domain Mark 1.0, Public Domain CC0 1.0, or U.S. Government

OSF		Apache License 2.0
Real and Fake Face Detection		<u>CC BY-NC-SA 4.0</u>
TFEID		Others (Not provided)
Training Data Profacial [TDP]		Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Table 1: Overview of different datasets used.

Data Cleaning

Data cleaning involved multiple steps to ensure standardization in quality and size. We manually inspected each image to ensure its quality, eliminating pictures that were corrupted or noisy. By prioritizing image quality from the outset, we negated the need for adjustments in brightness, contrast, or angle. To exclude children, we removed all images of individuals suspected to be under 18 years of age.

Our quality standards required extra time to find datasets meeting our criteria, as many images were unfit for our project. Finding sufficient images for each emotion category proved challenging, with some emotions more difficult to source online. To avoid unwanted biases from an imbalance of images, we set aside excess pictures as potential backups, striving to maintain a similar number of images in each category. Initially planning to use only black and white images, we ultimately chose not to alter colors due to the lack of a color criterion in the assignment, preserving the option to grayscale pictures later using the PIL library if necessary.

With regards to the size of the images, we excluded images that were cropped too closely to the individual's face, showed the whole body, were not rectangular in shape, or the image contained too few pixels that enlargements would distort the facial expressions. To ensure consistency, we ran a python script to gather information pertaining to the size and distribution of the images (see table 2).

Width	Height	Count		Width	Height	Count
600	600	519		2464	3280	3
1024	1024	508		2304	1120	3
640	490	331		1472	3264	3
1080	1445	300		3888	3888	3
2444	1718	215		1072	1009	1
480	600	114		497	600	1
640	480	21		954	1093	1
2448	3264	15		570	588	1
720	480	10		943	529	1
1944	2592	9		904	736	1
1280	720	6		790	849	1
1448	3216	3		784	849	1
3880	5184	3		783	808	1
3456	4608	3		739	769	1
2592	4608	3		696	786	1
2556	3408	3				

Table 2: Distribution of number of images per image size

Based on preliminary examination of the width and height of our images, we noticed that the range of sizes of the images proved to be quite varied. We would need to rescale our images to achieve uniformity; however, we wanted to minimize potential distortion of the images. Given that scaling down images does not affect the quality of the images, we agreed that 640 x 640 is the preferred size as the bulk of our images fell in that range.

The largest of the images, which represent a small percentage of the overall datasets, were pre-processed either manually in photoshop, in an additional script for resizing, or a combination of both techniques. Once those images were resized, we regrouped them with remaining images which were then resized to 640 x 640 via the script we wrote using PIL library. Table 2 shows these three stages in which the large images go through: Original, Preprocessed, and Final.

Original	Preprocessed	Final
1080 x 1445 pixels	1080 x 1080 pixels	640 x 640 pixels

Table 3: Various stages of processing large images

Labeling

After determining which datasets met our criteria, we established definitions for happy, surprised, neutral, and focused expressions. Five of the seven datasets utilized in the project were pre labeller. Even though most datasets came pre-labeled, they lacked sufficient images for our project, necessitating manual labeling for the Real and Fake Face Detection Dataset and the FFHQ Dataset for a significant portion of our work.

Identifying happy expressions was the simplest, characterized by upward turning lips, sometimes showing teeth, and often accompanied by crinkling at the eyes' outer corners and slightly raised eyebrows, creating an open facial expression. Surprised expressions were more challenging to find, marked by widened eyes,

eyebrows raised towards the hairline, and horizontal forehead wrinkles, with subjects typically having their mouths open to varying extents. These expressions were grouped together if they shared any of these characteristics. Neutral expressions were more readily available in pre-labeled datasets, defined by a relaxed facial demeanor with eyes focused directly on the camera or a distant object. Distinguishing focused expressions proved most difficult; we sought signs of engagement different from neutral expressions, such as tension around the eyes, a slight frown or squint, a forward-leaning posture, or intense eye gaze. Some neutral images were reclassified as focused upon closer examination.

For manual labeling of the Real and Fake Face Detection Dataset and the FFHQ Dataset, we utilized Labelbox[9], creating individual accounts to upload images. We developed an ontology within Labelbox to categorize images into our four selected emotions, opting for radio buttons over dropdown lists for quick labeling via hotkeys. This process, supported by Labelbox's efficient UI and shortcuts, allowed us to label images in approximately four seconds each. Labelbox's JSON file export feature facilitated organizing images into correctly labeled folders through a script we developed with guidance from Labelbox's documentation[10]. The main limitation encountered was the usage cap on free Labelbox accounts, preventing its use for the extensive FFHQ Dataset.

For the FFHQ Dataset, needing only focused or surprised images, we divided the dataset into thirds to optimize efficiency, starting at different points. This strategy enabled rapid collection of focused images, while surprised expressions took longer to identify, sometimes yielding up to 40 images in a single folder. We manually reviewed the dataset, removing unfit images.

Upon processing and categorizing each dataset, we uploaded the data to GitHub, keeping datasets in separate folders with subfolders for clear identification and citation of each image.

Dataset Visualization

The visualization of our dataset played a crucial role in understanding its structure, diversity, and potential biases. Utilizing Python's Matplotlib library, we conducted a comprehensive analysis to visualize class distributions, sample images, and pixel intensity distributions. These visualizations provided us with valuable insights, enabling informed decisions for further data preprocessing and augmentation strategies.

Class distribution analysis highlighted the need for balancing, sample images showcased the dataset's diversity and flagged potential issues, and pixel intensity distributions shed light on variations in image quality and lighting conditions. Together, these visualizations guided our approach to creating a robust and well-prepared

dataset for training our emotion recognition model, ensuring that it is not only diverse but also representative and balanced across different classes.

A. Class Distribution

To assess the balance of our dataset, we plotted the number of images in each emotional category (happy, surprised, neutral, and focused) using a bar graph **(figure 1)**. This visualization highlighted disparities in class representation, revealing that some emotions were overrepresented while others were underrepresented. For instance, 'happy' images were more abundant, reflecting easier availability or identification, whereas 'focused' expressions were less common, indicating a potential need for targeted data augmentation or collection to balance the dataset.

Based on the analysis we augmented the class representation to maintain uniformity across all classes by removing abundant images of overrepresented classes. The exact numbers on each bar provided a clear, quantifiable measure of the class distribution, which was instrumental in planning our next steps for dataset augmentation and ensuring that our model training would not be biased towards more frequently represented classes.

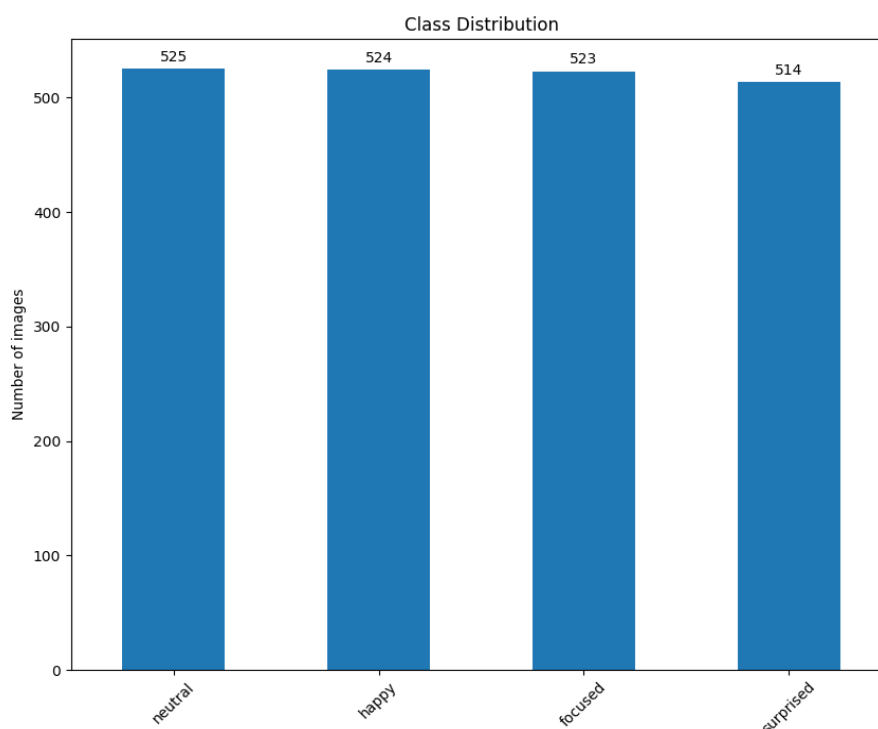| Initial Distribution		Final Distribution	
neutral	1078	neutral	525
happy	588	happy	524
focused	544	focused	523
surprised	500	surprised	514

Figure 1: Class distribution bar graph

B. Sample Images

Understanding the content and quality of our dataset was furthered by presenting collections of 25 randomly chosen images for each class, arranged in a 5x5 grid **(Figures 2 - 5)**. This exercise was not only instrumental in visually assessing the diversity and consistency of image representation within each category but also highlighted any anomalies or mislabels that might have escaped earlier scrutiny.

Ensuring that these images were randomly selected upon each execution of our code allowed for a dynamic assessment of our dataset, preventing oversight of potential issues confined to a subset of the data. By resizing each sample image to a calculated ratio using our defined heuristic, we maintained consistency in visual presentation, which was essential for accurately gauging the dataset's characteristics.
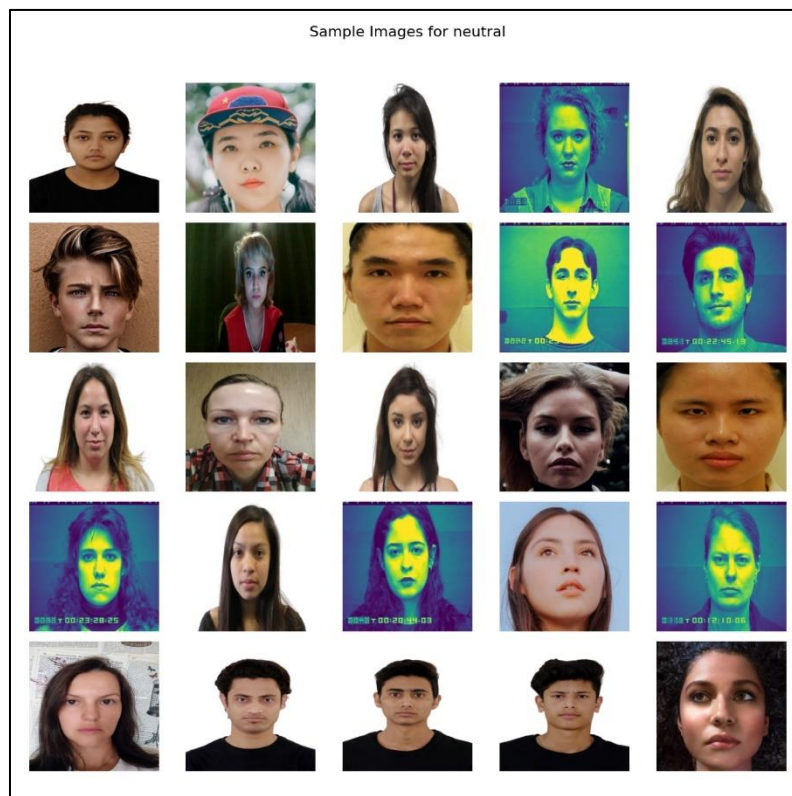

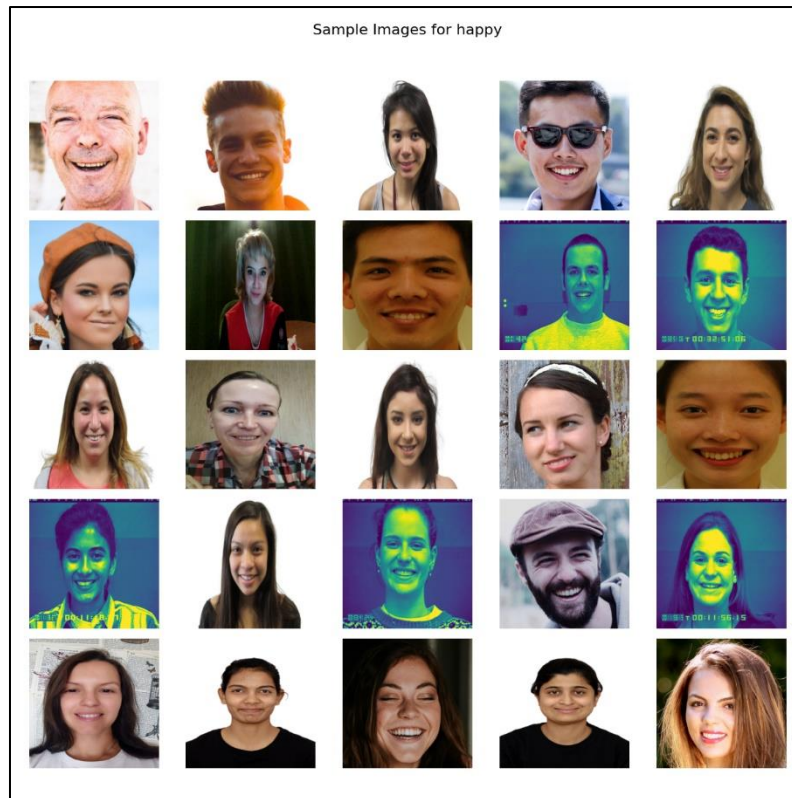Figure 2: Sample Image Grid of `Neutral` expression
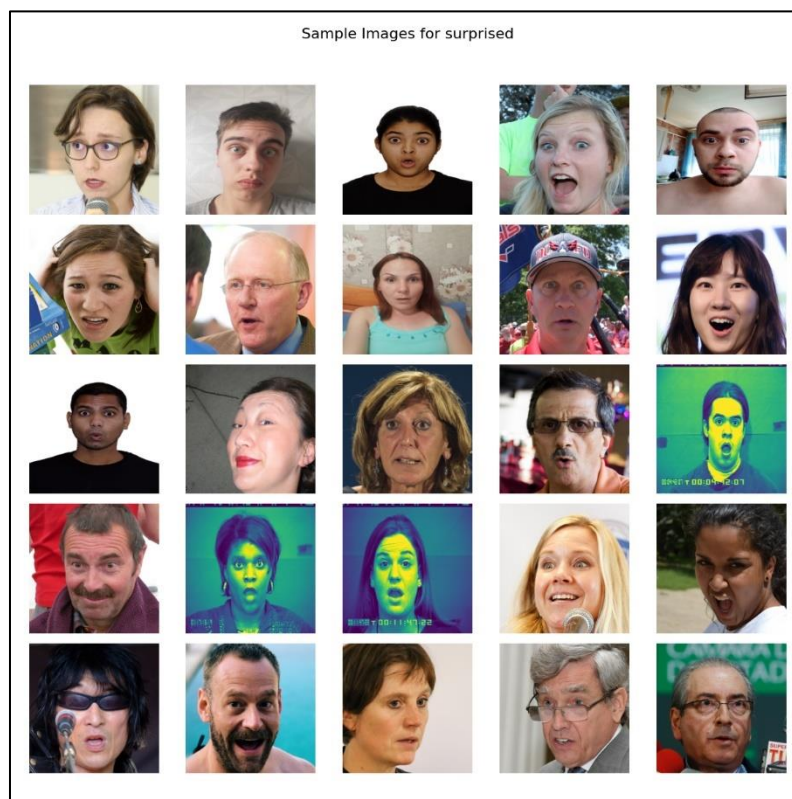
Figure 3: Sample Image Grid of `Happy` expression

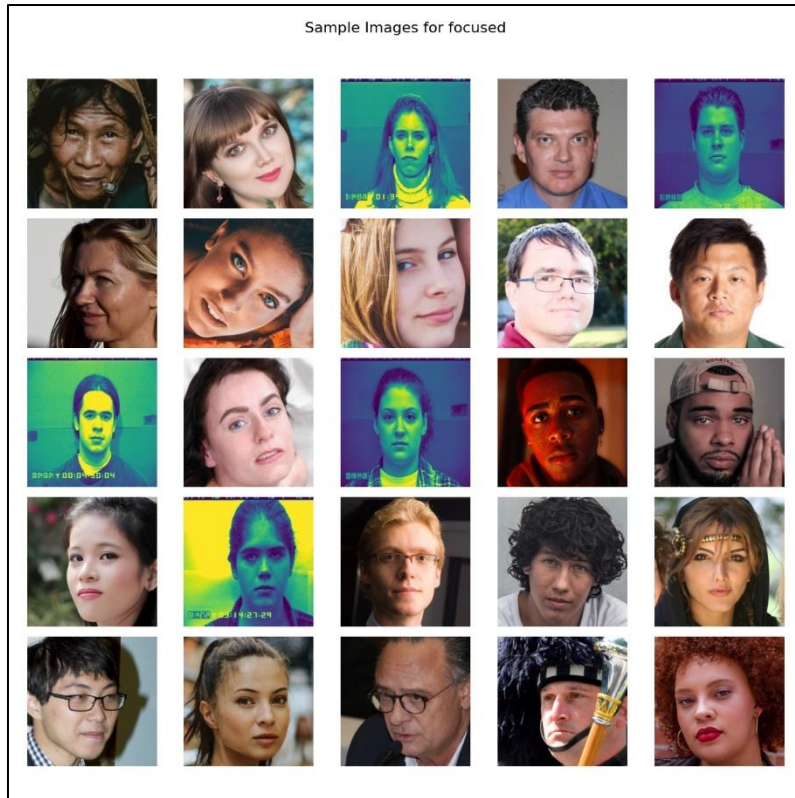Figure 4: Sample Image Grid of `Surprised` expression

Figure 5: Sample Image Grid of `Focused` expression

C. Pixel Intensity Distribution

The analysis of pixel intensity distributions for a randomly chosen subset of 25 images from each class provided insights into the dataset's lighting conditions and overall image quality **(Figures 6-9)**. We plotted histograms for each image, overlaying the intensity distributions for the **Red, Geen, and Blue** channels in color images, which was instrumental in identifying variations in lighting conditions across our dataset.

This visualization was crucial for understanding how different lighting conditions might affect model training and performance. For instance, images with predominantly low-intensity values might be underexposed, potentially necessitating brightness normalization during preprocessing. Conversely, images skewed towards high-intensity values could be overexposed, similarly affecting model perception.
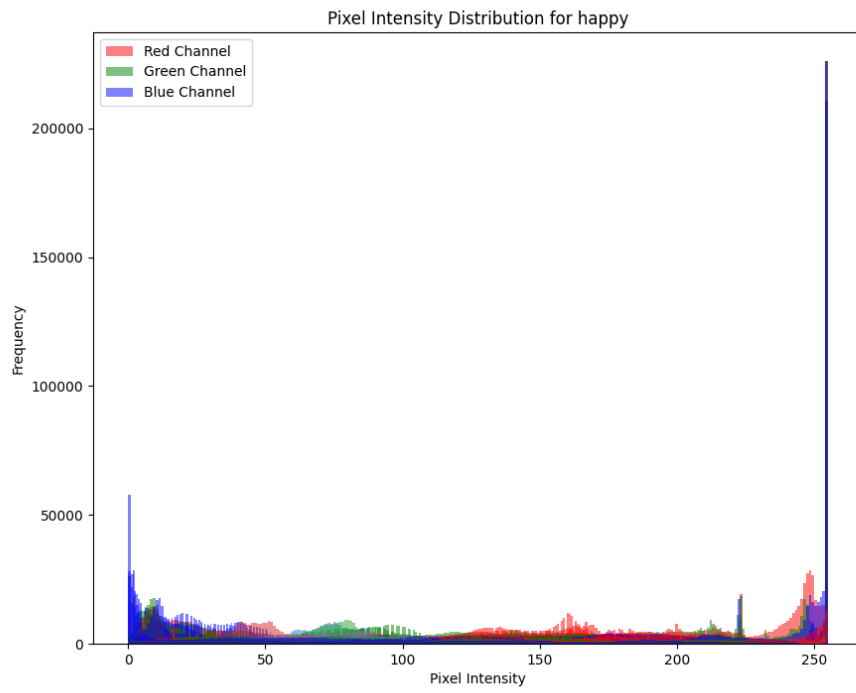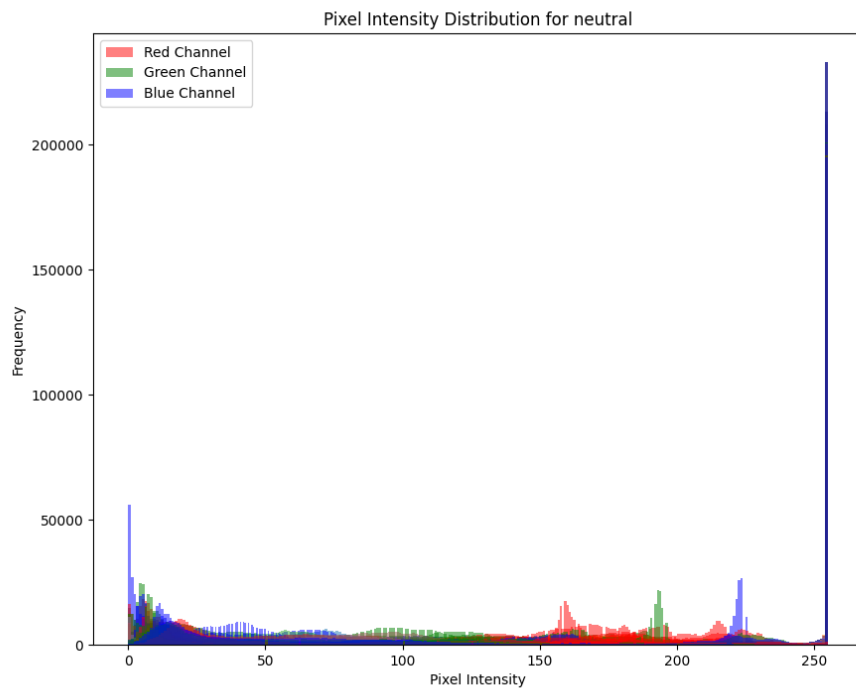
Figure 6: Pixel Intensity Distribution - Happy

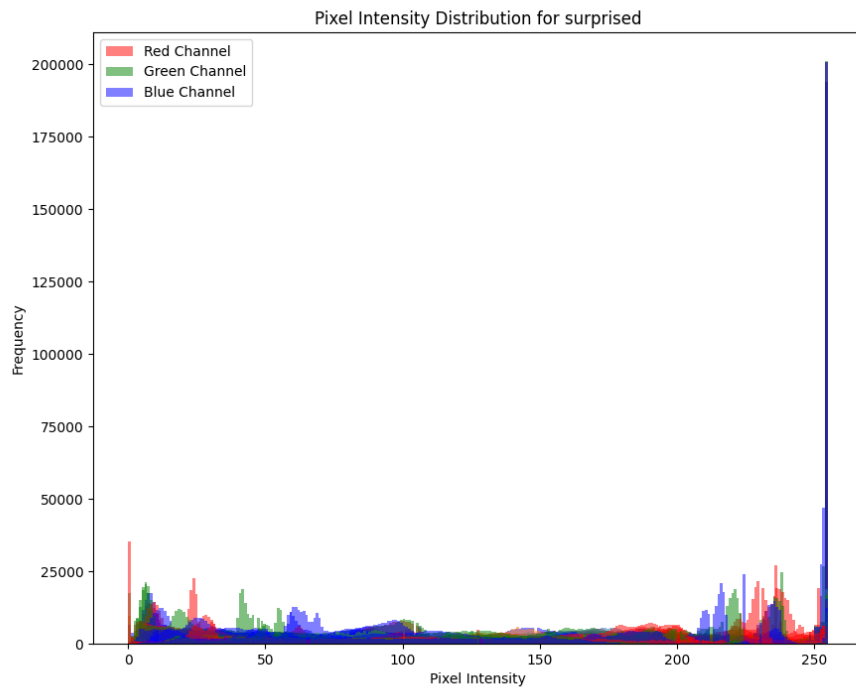Figure 7: Pixel Intensity Distribution - Neural
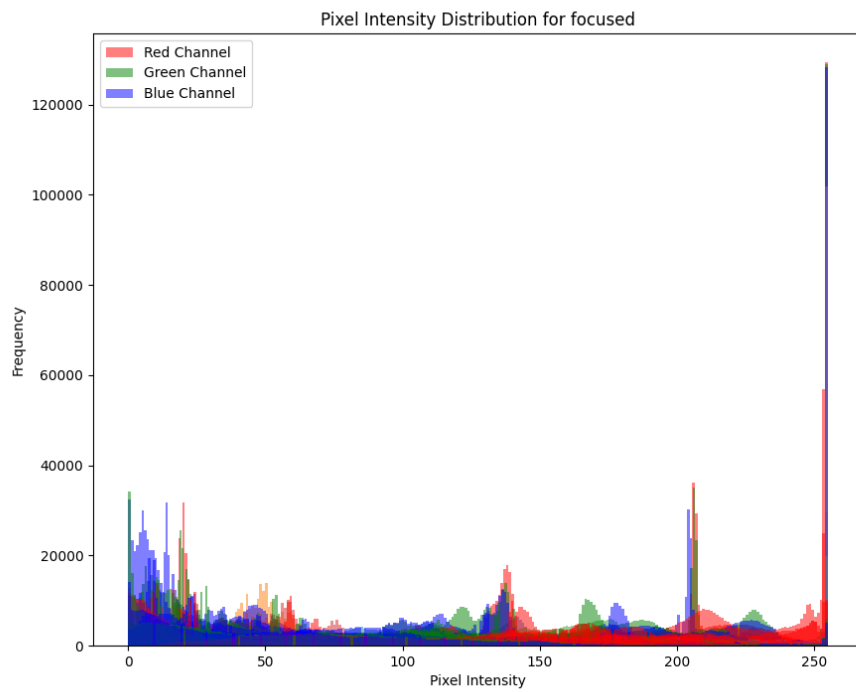
Figure 8: Pixel Intensity Distribution - Surprised

Figure 10: Pixel Intensity Distribution - Focused

References

[1]

'MULTIRACIAL FACE DATABASE', sciplab. Accessed: Mar. 03, 2024. [Online].
Available: https://jacquelinemchen.wixsite.com/sciplab/face-database

[2]

Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended
Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-
specified expression" Accessed: Mar. 03, 2024. [Online]. Available: CK+ dataset
(kaggle.com)

[3]

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010).
The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action
unit and emotion-specified expression. Proceedings of the Third International
Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010),
San Francisco, USA, 94-101. Accessed: Mar. 03, 2024. [Online]. Available: CK+
Dataset | Papers With Code

[4]

'NVlabs/ffhq-dataset'. NVIDIA Research Projects, Mar. 02, 2024. Accessed: Mar. 03,
2024. [Online]. Available: https://github.com/NVlabs/ffhq-dataset

[5]

S. Tewari, S. Mehta, and N. Srinivasan, 'IIMI Emotional Face Database', May 2023,
Accessed: Mar. 03, 2024. [Online]. Available: https://osf.io/f7zbv/

[6]

'spenceryee/CS229'. Accessed: Mar. 03, 2024. [Online].

[7]

'Real and Fake Face Detection'. Accessed: Mar. 03, 2024. [Online]. Available:
https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection
Available: https://github.com/spenceryee/CS229

[8]

'Facial Emotion Recognition Dataset'. Accessed: Mar. 03, 2024. [Online]. Available:
https://www.kaggle.com/datasets/tapakah68/facial-emotion-recognition

[9]

'Labelbox docs', Labelbox docs. Accessed: Mar. 03, 2024. [Online].
Available: https://docs.labelbox.com/

[10]

'labelbox-python/examples/basics/data_rows.ipynb at master · Labelbox/labelb ox-
python', GitHub. Accessed: Mar. 03, 2024. [Online].
Available: https://github.com/Labelbox/labelbox-python/blob/master/examples/
basics/data_rows.ipynb