COMP 472 - Project Part #1

Report #1

Team OB_05

Bryan Carlo Miguel 40231530

Yasser Ameer 40212780

William Nazarian 40213100

GitHub: https://github.com/WilliamNazarian/Comp472Ai.git

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

Contents

Dataset

The project began by searching for an appropriate dataset. After considering many datasets (ex. FER2013, CK+, expW - to name a few), we decided to use the **AffectNet** dataset. Some notable characteristics about this dataset is its variety in demographic, backgrounds, face shot orientations, and lighting conditions.

The dataset is already pre-labelled into eight classes as displayed in the table below:

Table 1 Total of number of Images

Class	Number of images
Anger	3218
Contempt	2871
Disgust	2477
Fear	3176
Happy	5044
Neutral	5126
Sad	3091
Surprise	4039

The dataset includes individuals across a variety of age groups, from toddlers to elderly adults. Additionally, the dataset includes individuals from multiple ethnic backgrounds, such as Asian, African, Hispanic, etc. This demographic variety closely reflects the variety of students that can be found in classrooms. Furthermore, the dataset is not just limited to frontal face shots. There are plenty of images with a ¾-profile shot. Images in the dataset also contain diverse backgrounds, with a wide range of settings and environments. This variety in the dataset reflects “real-world” situations closely, which could improve the performance of our model by making it more able to deal with noise and differences.

Despite the benefits that this variety provides, it can also pose some challenges. A potential problem would be overfitting due to the model focusing on very specific patterns, or underfitting due to the model not recognizing enough patterns. Additionally, the model may be more complex in order to handle the variability. One final potential problem may be a potential bias towards certain demographics. Regardless of these challenges, we believe that the dataset will provide a strong foundation to train a robust model with “real-world” applications, capable of dealing with noise and variances.

Provenance Information

Dataset Overview

Table 2 Source provenance:

Name	FER AffectNet
Source	Kaggle
URL(s)	https://www.kaggle.com/datasets/noamsegal/affectnet-training-data/ https://web.archive.org/web/20240205175353/http://mohammadmahoor.com/affectnet/
Author(s)	Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor, Noam Segal
Temporal Coverage Start Date	12/30/1938
Temporal Coverage End Date	12/30/2022
Date of Download	5/24/2024

Data Description

The dataset is organized as a collection of 96×96px .png files, each categorized into the eight classes mentioned before. Additionally, the dataset comes with a .csv containing index ‘i’ of the image, path of the image file in the dataset, label/class, and the PFC%s. In this case, the PFC%s represents the monochromaticity of an image. A high PFC%s (~>99%) means that an image is grayscale, while a lower PFC%s (~<80%) means that an image has plenty more RGB values. Below is a subset of the contents of the .csv file:

Table 3 CSV file:

#	Path	label	relFCs
0	anger/image0000006.jpg	surprise	0.8731421290934949
1	anger/image0000060.jpg	anger	0.852310783018639
2	anger/image0000061.jpg	anger	0.80095684657714
…	…	…	…

We decided to select images with a relFC value near 1 to ensure high-quality images overall. Some problems we faced with this dataset included having the same person in different poses within the same class. This needed to be addressed to maintain consistency. Additionally, we had to filter out irrelevant pictures and ensure that each folder contained images reflecting the correct emotions.

The relFC value was particularly helpful for us in determining if an image was clear and effectively expressed the person's emotions.

Data Collection

The images were collected by querying three major search engines using 1250 keywords in six different languages [2]. The images were collected 'in the wild', meaning that the images were not collected in a controlled and posed environment [1]. There is no information on where exactly each individual image is sourced from.

Licensing and Permissions

The dataset is licensed under the "Attribution-NonCommercial-ShareAlike 3.0 IGO (CC BY-NC-SA 3.0 IGO)" licence.

Data Cleaning

1. Techniques and Methods

To begin with, after organizing our dataset and personal images into proper directories, we needed to standardize the dataset. To do this, we created a script to resize every image to 96 x 96 pixels, ensuring uniformity in image size.

According to a research paper, converting images from RGB to grayscale can achieve better classification accuracy using genetic algorithms.[3] To optimize our accuracy, we wrote a script to transform the images into black and white. This script also performs histogram equalization to adjust the lighting. At the end of the process, the script stores the processed images in a new folder.

2. Challenges and Solutions

This method involved using OpenCV to adjust the brightness and color of the images, ensuring they were consistently prepared for our project. Using OpenCV in the Anaconda environment was a bit difficult for us, as we had no prior experience with it. We had to learn how to use OpenCV and integrate it into our workflow.

3. Example Images

Here is some of the examples of image for each category we have done:

Table 4 example of images:

Before	After
	
	
	

Data Labeling

1. Method Used

One of the emotions that the team wasn't able to find was engagement and focus. To address this, we had to select a bunch of images from a chosen dataset, specifically from the contempt and surprise folders, and use the suggested website LabelBox. We went through around 1500 images to pick about 500 that seemed to depict engaged and focused facial emotions.

LabelBox provided several useful features that allowed us to verify and label the images we wanted to use. We could then export the labeled data in a JSON file, enabling us to extract the image file names and update our CSV file, changing "contempt" to "engagement." This helped us build a well-structured dataset.

With this data, we were able to create a Python script that filtered the images into the correct categories, ultimately giving us a new folder containing only images for the engaged and focused emotion.

Finding a dataset with decent-sized images wasn't easy. Most of the images were 96 x 96, which doesn't provide the best resolution possible. For this reason, we decided to manually review all the chosen data to ensure they had good resolution, reducing ambiguity in later phases of the project.

The dataset we chose mostly contained images with a single person, which is crucial for our project's accuracy and functionality. We needed to ensure that all images in the dataset featured only one person and did not include multiple faces. Manual labeling in LabelBox helped us achieve this requirement.

2. Challenges faced.

Regarding our chosen dataset, we were able to find everything we needed within a single dataset that had several folders representing different classes of emotions. This meant we didn't face any issues with handling multiple datasets or mapping those classes.

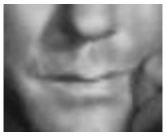
One challenge we faced during the labeling phase was that the criteria for identifying engaged facial emotions differed from those for other emotions. For example, we labeled an image from the contempt folder as an engaged image.

Figure 1 Engaged/Focused

We had to ensure that the engaged facial expressions didn't resemble other classes, such as neutral faces. Many of the engaged images gave the impression of being neutral, which could cause inaccuracies later in the project. It was crucial for us to accurately label these emotions to maintain the integrity of our data.

We also needed to determine how the AI would recognize an emotion and establish the criteria for this recognition. For example, engaged facial expressions might be identified by specific features such as slightly raised eyebrows, focused eyes, and subtle mouth movements, distinguishing them from neutral expressions. Establishing clear criteria like this is essential for accurate emotion recognition.

Table 5 Area of focus:

Face Part	
Mouth	
Hand gesture	
eyes/ forehead/ eyebrows	

Dataset Visualization

1. Class Distribution

This part of the report was conducted using some Python scripts that can be found in a folder in the GitHub repository or in the zip file called "scripts."

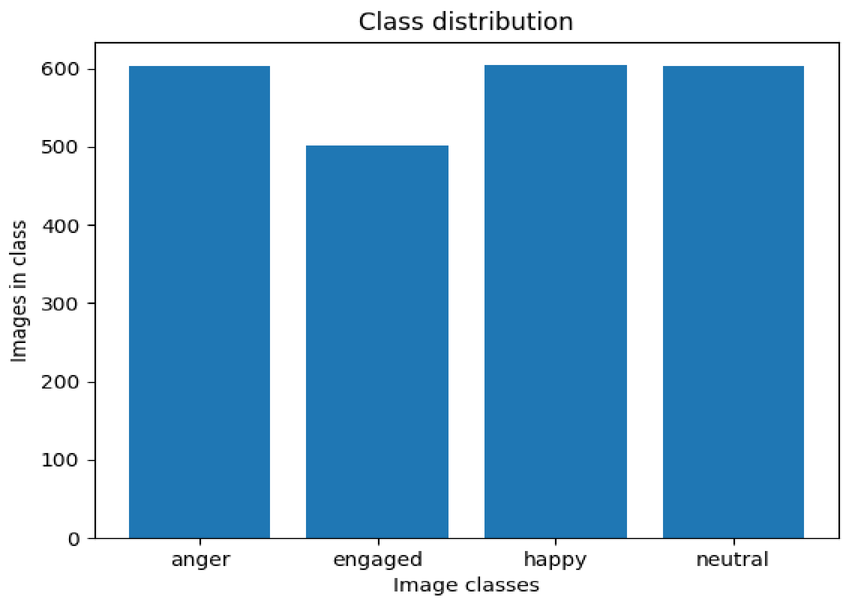


Figure 2 Class distribution.

2. Pixel Intensity Distribution

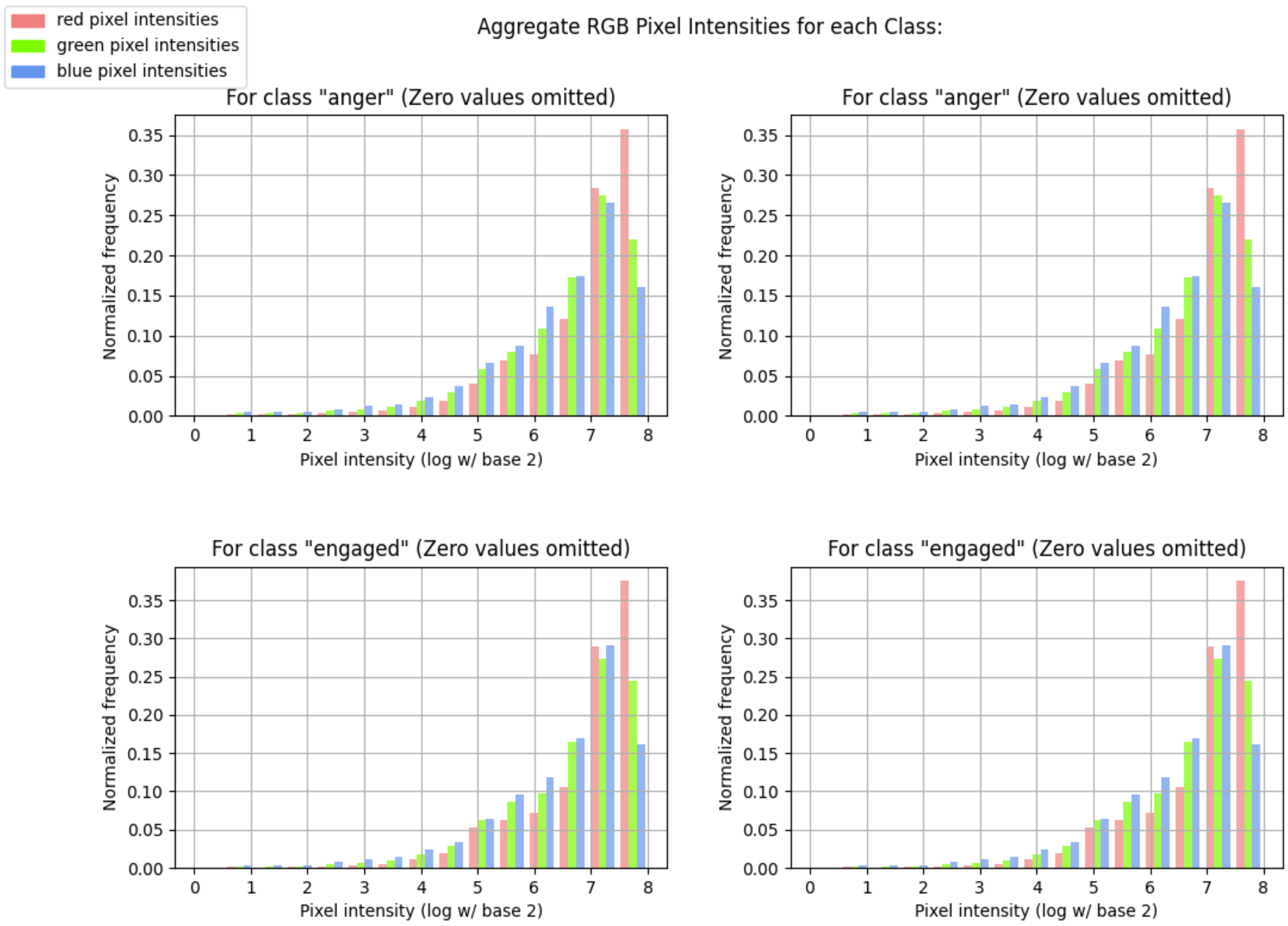


Figure 3 Pixel Intensity Distribution

3. Sample Image

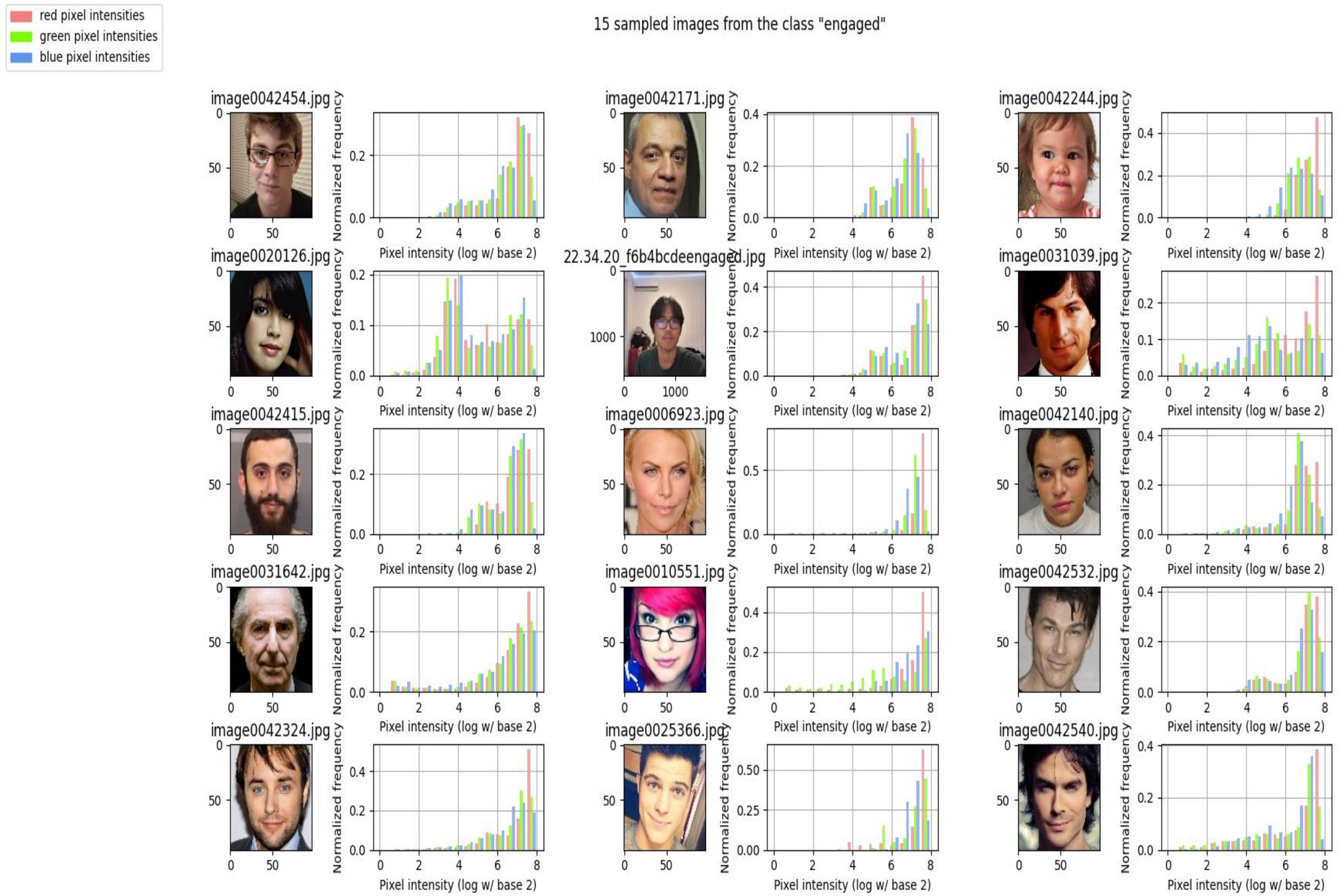


Figure 4 Sample Images Engaged.

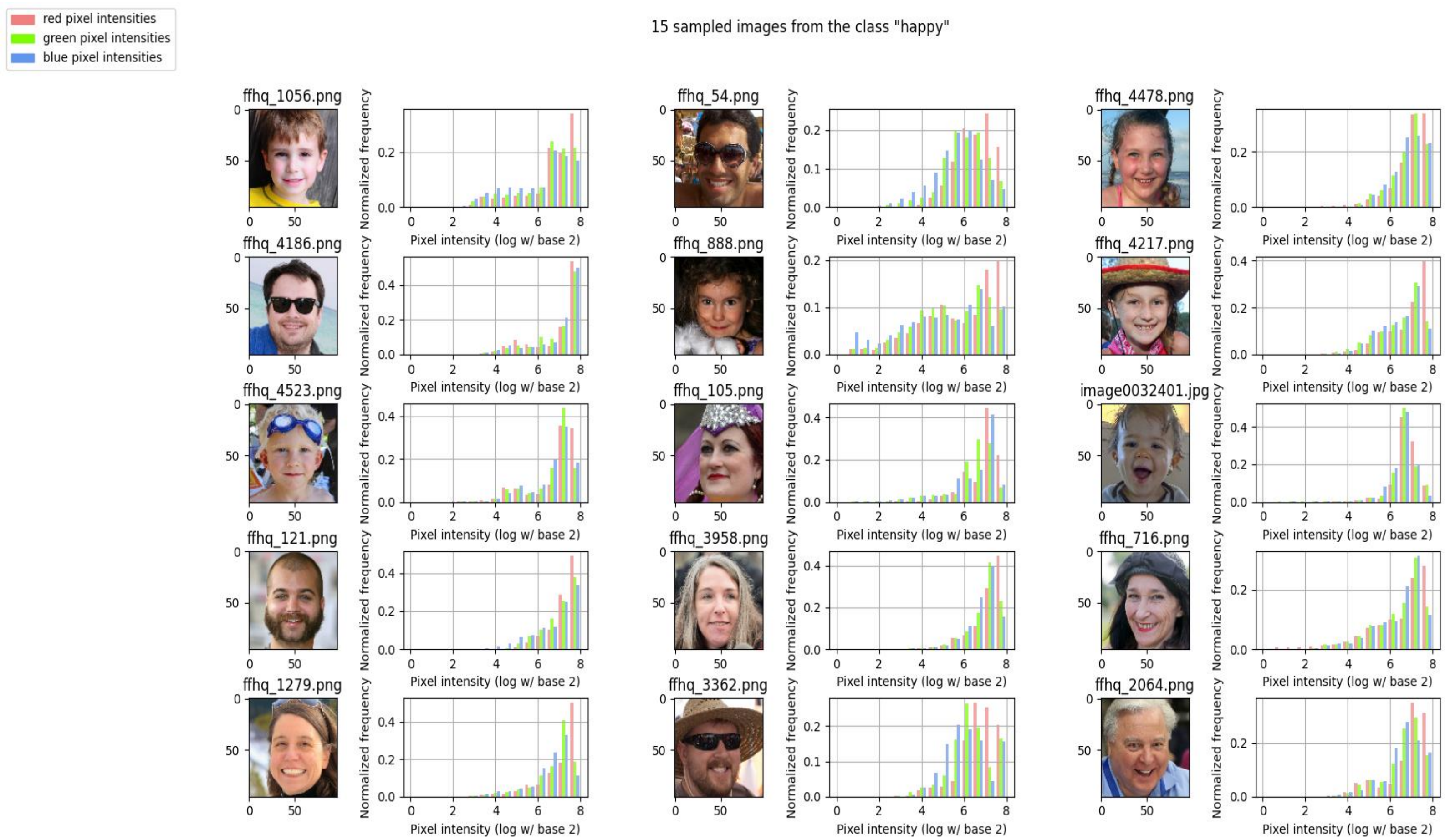


Figure 5 Sample Images Happy

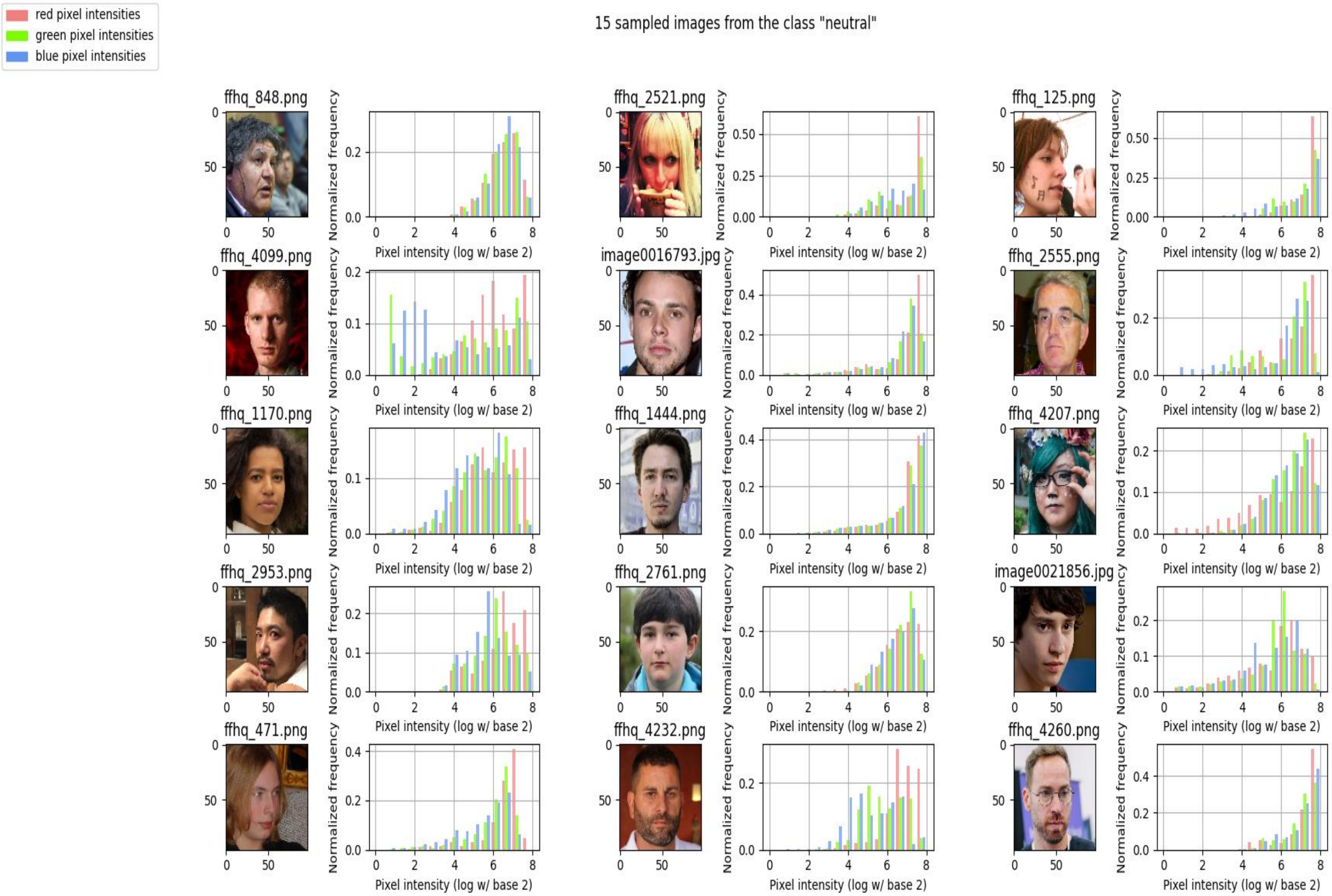


Figure 6 Sample Images Neutral

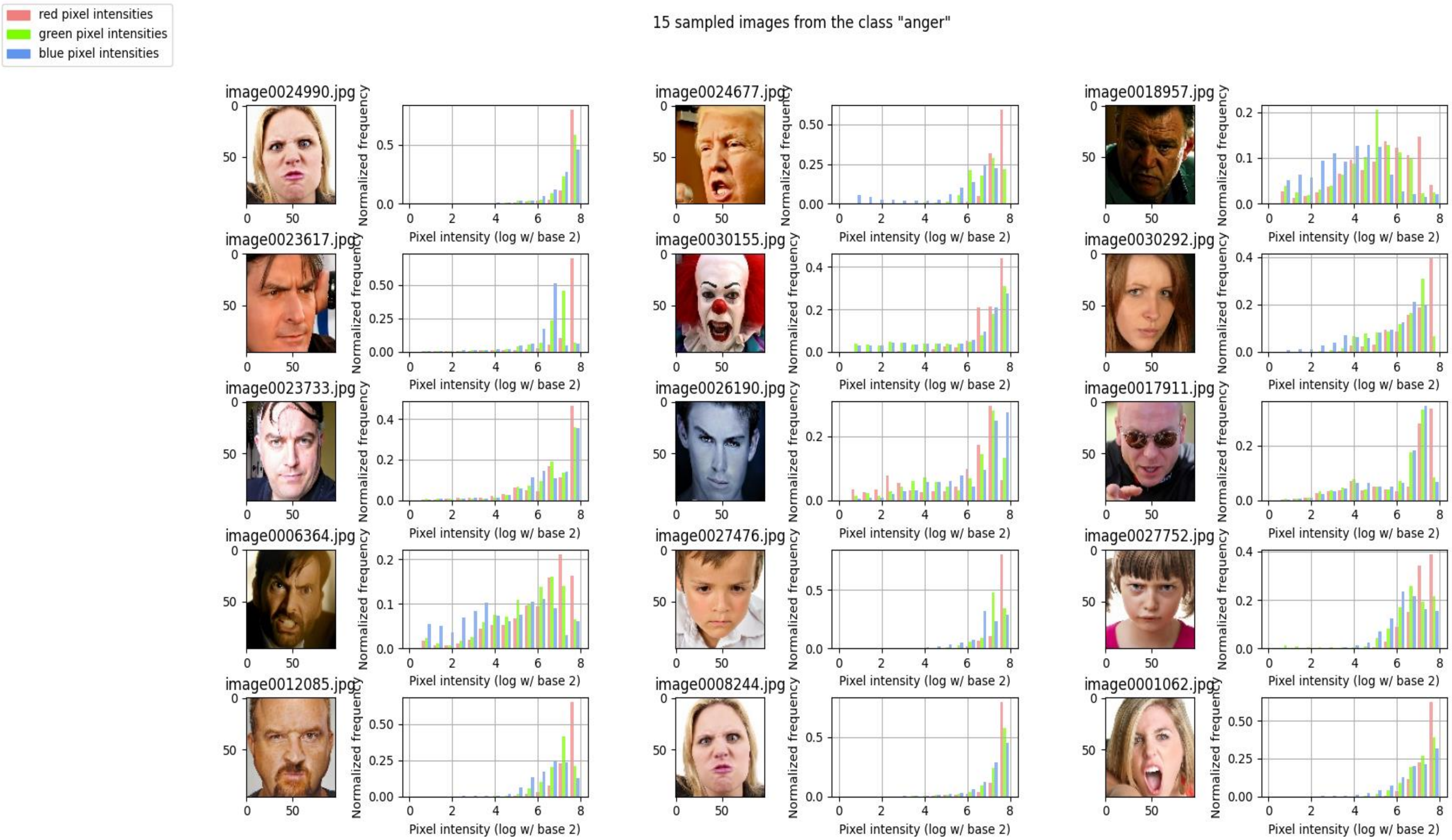


Figure 7 Sample Images Anger

References

[1] AffectNet: A database for facial expression ... - Mohammad Mahoormohammadmahoor.com/wp-content/uploads/2017/08/affectnet_onecolumn-2.PDF, http://mohammadmahoor.com/wp-content/uploads/2017/08/AffectNet_oneColumn-2.pdf (accessed May 31, 2024).

[2] Mohammad H. Mahoor, "AffectNet" mohammadmahoor.com/affectnet/.

https://web.archive.org/web/20240205175353/http://mohammadmahoor.com/affectnet/ (accessed May. 29, 20024).

[3] (PDF) optimizing the color-to-grayscale conversion for Image Classification,

https://www.researchgate.net/publication/282730506_Optimizing_the_Color-to-Grayscale_Conversion_for_Image_Classification (accessed May 31, 2024).