Task 1 Report

Visualizing CNN Activation Maps for Emotion Detection

Stuthi Keerthana Siddabattula

June 29, 2025

---

1 Introduction

This internship project focused on visualizing activation maps from a Convolutional Neural Network (CNN) trained for emotion detection. The primary goal was to understand which regions of the face activate specific filters within the network, providing interpretability to an otherwise black-box model.

2 Background

Emotion recognition plays a vital role in intelligent systems involving human interaction. CNNs are widely used for facial emotion detection due to their spatial learning ability. However, interpretability is a concern. Activation map visualization bridges this gap by showing what the model "sees" at different layers.

3 Learning Objectives

- Learn how CNNs extract and learn spatial features from facial expressions.
- Understand and extract intermediate layer outputs in a Keras CNN.
- Visualize feature maps to interpret model behavior.

4 Activities and Tasks

- Loaded a pretrained CNN model for emotion detection.
- Identified convolutional layers for feature visualization.

- Built a Keras functional model to access intermediate outputs.
- Preprocessed images (grayscale, resized, normalized).
- Visualized activation maps using Matplotlib.

5 Skills and Competencies

- **Deep Learning:** CNN architecture, layers, and filter behavior.
- **Keras/TensorFlow:** Model manipulation and intermediate output extraction.
- **Computer Vision:** Image preprocessing with OpenCV.
- **Visualization:** Plotting multidimensional filter outputs using Matplotlib.
- **Debugging:** Solving input layer access issues in Sequential models.

6 Feedback and Evidence

The activation maps provided visual proof of which facial regions were emphasized by different filters. Early layers focused on edges and contours, while deeper layers responded to more abstract patterns like mouth curves or eye positions. Screenshots and output plots validated the success of this interpretability exercise.

7 Sample Input and Output Visualizations

The left image shows the grayscale facial expression provided as input to the CNN model. The right image illustrates one of the 64 activation maps produced by the first convolutional layer, highlighting which facial regions the filter is focusing on.

8 Challenges and Solutions

Challenge 1: Sequential model did not expose `model.input`, causing an error in intermediate model construction.
Solution: Rebuilt the model using the Functional API with `Input()`.

Challenge 2: Image shape mismatch with model requirements.
Solution: Implemented consistent preprocessing: grayscale conversion, resizing to 48x48, and adding the channel dimension.

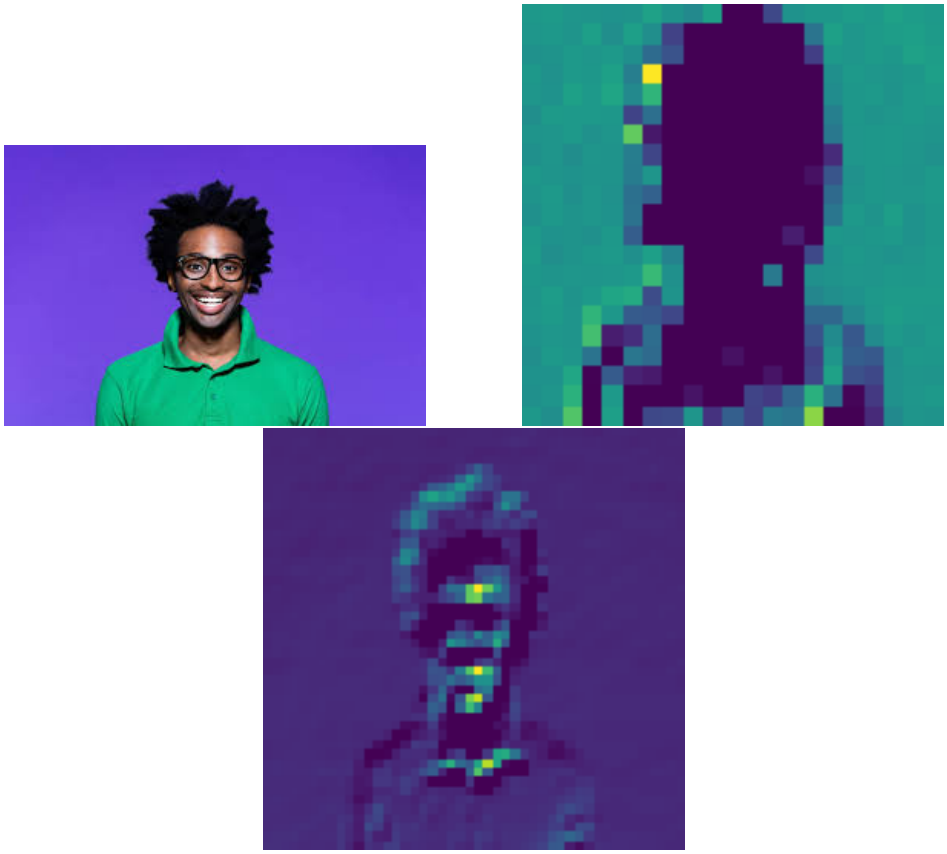

Figure 1: Left: Input grayscale facial image (48x48)

Figure 2: Right: Activation map from the first convolutional layer

Figure 3: Centre: Activation map from the 5th convolutional layer

9 Outcomes and Impact

- Gained practical experience in model interpretability.
- Improved understanding of CNN architecture and feature hierarchy.
- Built generalized code that can be reused for other CNN visualization tasks.

10 Conclusion

This internship task significantly enhanced my understanding of CNNs and their inner workings. Visualizing activation maps gave me critical insight into how deep learning models make decisions, improving both my technical skills and my ability to build interpretable AI systems.

Submitted by: Stuthi Keerthana Siddabattula