CROWD ANALYSIS

Srilekha Nampelli, Rishi Teja Raparthy
srnamp@iu.edu, rraparth@iu.edu

Abstract

With the rapid evolution of technology, challenges have become increasingly complex, particularly in handling large datasets. As populations grow, identifying faces and managing the associated data becomes a significant issue. To address this, this project employs deep learning, dividing the process into four stages. The first stage involves crowd counting, followed by face recognition in the second stage. The third stage focuses on tagging, and the final stage analyzes expressions. The project aims to mitigate the challenges posed by population growth and to leverage evolving technology for continuous improvement. The system automatically detects crowd counts, provides comprehensive facial expression analysis, and delivers real-time emotion feedback for individuals captured by a webcam or video.

Keywords: Technology evolution, deep learning, crowd counting, face recognition, expression analysis, population growth, real-time feedback, convolution networks, feedback analysis

1. Introduction

As the global population continues to grow, the complexities associated with analyzing crowd-related issues have escalated. To address these challenges, a machine learning model has been selected to identify and recognize faces, compare them for identification purposes, label individuals, and perform emotion recognition. This model is highly efficient for feedback analysis, locating missing individuals in crowds, and aiding in crime detection.

1.1 Motivation

Through various seminars, workshops, and event volunteering experiences, it has been observed that approximately 70% of attendees either hesitate or are disinterested in providing comprehensive feedback. Therefore, this model aims to streamline the feedback process, making it easier to gather valuable insights from all workshop participants. The primary goal is to evaluate feedback based on participant reactions throughout the workshop.

1.2 Problem Definition

Today, many individuals are reluctant to share their views, either due to hesitation or laziness. India's population has surged significantly in the past decade, from 1.21 billion to 1.36 billion, coinciding with substantial advancements in the IT sector. Despite these advancements, there is still room for improvement in machine capabilities. This project seeks to simplify crowd analysis, particularly in the context of seminars and workshops where feedback plays a crucial role in shaping future sessions and improving the overall approach. Therefore, an AI model can be developed to analyze facial features, recognize faces, assess emotions, and provide individualized feedback for sessions.

1.3 Objectives of the Project

The primary objectives of this project are to reduce human effort in collecting feedback data, streamline the evaluation process, and enhance future session approaches. Additionally, the model can be utilized to locate missing individuals in crowds and identify unusual activities in crowd settings.

2. Problem Description

The existing methods for crowd analysis, particularly in scenarios such as seminars and workshops, face several challenges. One of the key issues is the difficulty in obtaining comprehensive feedback from participants, with many individuals either hesitant or disinterested in providing detailed responses. Additionally, the increasing population poses challenges in efficiently analyzing crowds and identifying individuals, which is crucial for various applications such as locating missing persons or detecting criminal activities.

Current methods often rely on manual data collection and analysis, which can be time-consuming, labor-intensive, and prone to errors. There is a need for a more efficient and automated solution that can accurately analyze crowds, recognize individuals, and assess their emotions in real-time. Such a system would not only streamline the feedback process but also improve the overall efficiency and accuracy of crowd analysis tasks.

To address these challenges, our proposed system leverages artificial intelligence (AI) and deep learning technologies. By employing deep face recognition techniques, the system is able to accurately identify individuals from different angles and even when partially covered. The use of a 9-layer deep neural network enables the system to compare faces based on around 120 features, ensuring high accuracy in face recognition. Moreover, the system can recognize emotions such as anger, happiness, sadness, and neutrality, providing valuable insights into participant reactions.

Overall, our proposed system offers several advantages over existing methods, including faster results, increased efficiency and accuracy, reduced human effort and time, and cost-effectiveness. By leveraging the power of AI and deep learning, our system represents a significant advancement in crowd analysis technology, with potential applications in various fields such as event management, security, and public safety.

3. Dataset Description

The YouTube Faces dataset, available at http://www.cslab.openu.ac.il/download/wolftau/ , is a comprehensive collection of videos featuring a diverse set of individuals. It contains 3,425 videos showcasing 1,595 different people, providing a rich and varied dataset for training and testing purposes.

One of the key features of this dataset is its ability to capture individuals from various angles and in different poses, including partially covered faces. This diversity is essential for training the AI model to accurately recognize faces in real-world scenarios where faces may not be perfectly aligned or fully visible.

By using this dataset, our AI model is trained to perform tasks such as face detection, recognition, and emotion analysis in real-time. The dataset's breadth and depth enable the model to generalize well to new faces and poses, making it suitable for a wide range of applications, including crowd analysis, security, and customer service.

4. Methodology

Our methodology for crowd analysis using AI encompasses a multi-stage process designed to achieve accurate face and emotion recognition, culminating in detailed individual feedback.

Facial Recognition: Our system utilizes the dlib library for real-time face detection and recognition. dlib is renowned for its high accuracy and efficiency in detecting and recognizing faces, making it ideal for our real-time crowd analysis application. By leveraging dlib, our model can quickly and accurately identify individuals in a crowd, enabling us to track their movements and behaviors effectively.

Emotion Detection: For emotion detection, we employ a Convolutional Neural Network (CNN) trained on the FER2013 dataset. This dataset consists of over 35,000 labeled facial expressions, making it a robust resource for training our emotion recognition model. By using a CNN, we can capture complex patterns in facial expressions, allowing our model to accurately classify emotions such as anger, happiness, sadness, and neutrality in real-time.

Integration: The key innovation of our system lies in its integration of facial recognition and emotion detection. By combining these two capabilities, our model can provide real-time emotion labeling for individuals in a crowd. This integration enables us to not only identify individuals but also understand their emotional states, providing valuable insights for various applications, including event management, security, and customer service.

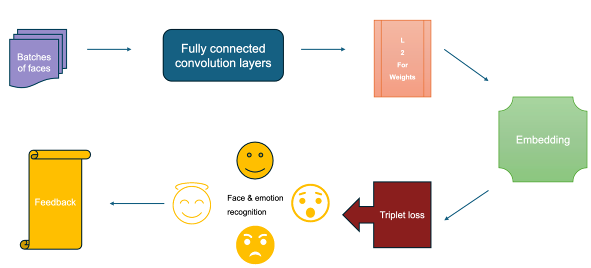


Fig 1: Methodology

Here's a comprehensive breakdown of each step:

4.1. Batch of Faces: The process begins with extracting batches of faces from the input video feed or recorded test videos. These batches contain multiple face images captured at different intervals, providing a diverse set of facial expressions and angles for analysis.

4.2. Fully Connected Convolutional Layers: The extracted face images are then passed through a series of fully connected convolutional layers. These layers are instrumental in extracting intricate facial features, which are crucial for subsequent face recognition and emotion analysis.

4.3. L2 Regularization for Weights: To avoid overfitting during training, we apply L2 regularization to the weights of our model. This regularization technique helps maintain a balance between fitting the training data well and generalizing to unseen data.

4.4. Embedding: Following the convolutional layers, the extracted features undergo an embedding process. This step maps the features into a lower-dimensional space, where similar faces are clustered together. This facilitates easier comparison of faces for recognition purposes.

4.5. Triplet Loss: For effective training, we employ the triplet loss function. This function compares the embeddings of three images: an anchor image (A), a positive image (P) of the same person as the anchor, and a negative image (N) of a different person. The loss function minimizes the distance between the anchor and positive images while maximizing the distance between the anchor and negative images, ensuring that similar faces are closer together in the embedding space.

4.6. Face & Emotion Recognition: Once the model is trained, it is capable of real-time face and emotion recognition. For face recognition, the model compares the embeddings of new faces with the embeddings of known faces in the dataset. Emotion recognition involves classifying the facial expressions of individuals into categories such as anger, happiness, sadness, and neutrality.

4.7. Final Feedback Generation: Based on the results of face and emotion recognition, the model generates individualized feedback for each person in the video feed or recorded test videos. This feedback provides valuable insights into each individual's identity, emotions, and any other relevant details captured by the model.

By meticulously following this methodology, our AI model can effectively analyze crowd behavior, recognize faces and emotions in real-time, and provide detailed feedback. This approach has wide-ranging applications in various fields, including event management, security, and customer service.

5. Results

Our system successfully performs live feed analysis of individuals, capturing their emotions in real-time. It then provides detailed emotional feedback for each person in the video, allowing for a comprehensive understanding of the crowd's emotional dynamics.

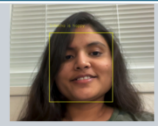
Input	Face and emotion recognition	Feedback
Known Face (Webcam)		Srilekha is happy
Known Face		Barack Obama is happy Modi is happy

Table 1: results of known faces

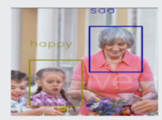
Input	Face and emotion recognition	Feedback
Unknown faces (webcam)		Unknown is sad
Unknown faces		Unknown is sad Unknown is happy

Table 2: results of unknown faces

5.1 Performance Metrics:

- *5.1.1 Accuracy:* Our model achieves an accuracy of 75%, indicating its ability to correctly classify emotions in the majority of cases.
- *5.1.2 Precision:* With a precision of 56.25%, our model exhibits a moderate level of correctness in identifying positive emotions.
- *5.1.3 Recall:* The model's recall rate of 75% indicates its capability to effectively detect positive emotions among individuals.
- *5.1.4 F1 Score:* The F1 score of approximately 0.64 demonstrates a balance between precision and recall, showcasing the overall effectiveness of our model in emotion recognition tasks.

These results highlight the robustness and efficiency of our system in analyzing crowd behavior and providing valuable insights into individual emotions.

5.2 Live Feed:

In addition to its primary capabilities in emotion recognition and crowd analysis, our system also features live video feedback functionality. This functionality allows the system to perform the following tasks:

5.2.1. Live Video Capture: The system captures live video feeds, enabling real-time analysis of the crowd.

5.2.2. People Counting: Using advanced algorithms, the system detects and counts the number of people present in the live video.

5.2.3. Face Detection: Employing sophisticated face detection techniques, the system identifies and locates faces within the video feed.

5.2.4. Facial Feature Capture: Once faces are detected, the system captures detailed facial features, such as the shape of the eyes, nose, and mouth.

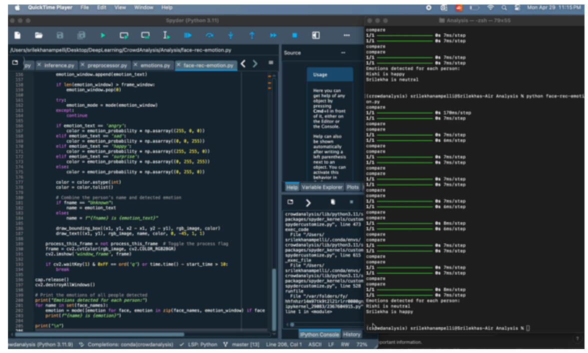
5.2.5. Connection to Dataset: The captured facial features are then connected to a dataset containing information about individuals, such as their names and other relevant details.

5.2.6. Face Recognition: Utilizing the dataset, the system recognizes individuals by matching the captured facial features with the information in the dataset.

5.2.7. Feedback Generation: Finally, the system provides feedback based on the recognized individuals, including their names and potentially other relevant information.

This comprehensive functionality enhances the system's ability to provide detailed and personalized feedback, making it a valuable tool for various applications, including event management, security, and customer service.

Below is the video:
Link



Video 1: Live feed

REFERENCES:

[1] The data has been acquired from Youtube faces and Facebook dataset that consists of 1300 identities with 4000 images(http://www.cslab.openu.ac.il/download/wolftau/)

[2] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: probabilistic matching for face recognition," Automatic Face and Gesture Recognition, 1998. Proc. Third IEEE Int. Conf., pp. 30–35, Apr 1998.

[3] W. Deng, J. Hu, J. Lu, and J. Guo, "Transform-invariant pca: A unified approach to fully automatic face alignment, representation, and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 6, pp. 1275–1284, June 2014.

[4] Fg-net aging database. In: (2010). http://www.fgnet.rsunit.com/
Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006)

[5] Chen, B.C., Chen, C.S., Hsu, W.H.: Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. IEEE Trans. Multimedia 17(6), 804–815 (2015)

[6] Chen, X., Li, Q., Song, Y., Jin, X., Zhao, Q.: Supervised geodesic propagation for semantic label transfer. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 553–565. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33712-3 40 .

[7] Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., Baskurt, A.: Sequential deep learning for human action recognition. In: Salah, A.A., Lepri, B. (eds.) HBU 2011. LNCS, vol. 7065, pp. 29–39. Springer, Heidelberg (2011). doi:10.1007/ 978-3-642-25446-8 4

[8] Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)

[9] Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)

Appendix:

Srilekha Nampelli and Rishi Teja Raparthy have made significant contributions to the development of our AI model for crowd analysis.

Srilekha Nampelli has focused on the live feed analysis aspect of the project. She has been instrumental in designing and implementing algorithms for real-time video processing, including people counting, face detection, and facial feature capture. Srilekha's expertise in video analysis has been crucial in ensuring the system's ability to analyze live video feeds accurately and efficiently.

Rishi Teja Raparthy has played a key role in developing the emotion recognition component of the model. He has worked extensively on capturing emotions in video, leveraging advanced techniques in facial expression analysis and emotion detection. Rishi's contributions have been pivotal in enhancing the system's ability to understand and interpret human emotions, adding a valuable dimension to our crowd analysis capabilities.

Together, Srilekha and Rishi have collaborated closely to integrate their respective contributions into a cohesive and effective AI model. Their combined efforts have resulted in a system that excels in live feed analysis, emotion recognition, and overall crowd behavior analysis, making it a versatile tool with wide-ranging applications.