COSE474-2024F: Final Project Proposal "Development of a Deep Learning-Based Facial Recognition Model Using Pretrained Networks "

Youngmin Kim

1. Introduction

To address these issues, recent approaches use pretrained models like CLIP (Contrastive Language-Image Pretraining), which learns from large-scale image-text pairs and provides general representations applicable to tasks like facial recognition without requiring task-specific training data. CLIP's ability to understand both visual and textual data makes it highly adaptable and effective in real-world scenarios with limited training data.

In this paper, we propose a facial recognition model that leverages CLIP's pretrained capabilities to handle common challenges in facial recognition, such as lighting and occlusion. We also explore the potential benefits of fine-tuning CLIP to further improve performance in face recognition tasks.

2. Problem definition & challenges

In this paper, we aim to build a facial recognition model that: 1. Can accurately recognize faces in diverse environments without needing extensive training data. 2. Utilizes CLIP, a pretrained vision-language model, to reduce the need for task-specific data while maintaining high performance.
And there are three challenges we can think of. 1. Adapting CLIP for Facial Recognition: CLIP is not specifically designed for facial recognition, so fine-tuning it for precise face identification is a challenge. 2. Handling Facial Variations: Faces in real-world settings vary due to lighting, expressions, and occlusions. Ensuring CLIP can handle these variations effectively is crucial. 3. Limited Data: Task-specific facial recognition data may be limited. We need to ensure that CLIP performs well even with minimal labeled data.

3. Related Works

Vision-Language Models for Recognition Tasks: The application of vision-language models like CLIP in facial recognition tasks has recently gained attention. Studies utilizing CLIP have demonstrated its remarkable generalization capabilities when applied to facial recognition, particularly excelling in zero-shot learning and data-scarce scenarios. However, existing research has primarily focused on CLIP's generalization performance, with limited optimization specifically for facial recognition tasks. .

4. Datasets

In this study, we use publicly available benchmark datasets to evaluate the performance of our facial recognition model. These datasets offer diverse facial images under various conditions, such as lighting and expressions, ensuring the model's generalizability and robustness. Using open datasets also allows for fair comparisons with existing models while saving time and resources by eliminating the need to create a new dataset.

5. State-of-the-art methods and baselines

Recent facial recognition models like FaceNet have achieved over 99% accuracy on datasets like LFW, relying on large labeled data and complex architectures. However, their performance decreases in challenging conditions like occlusions or varying lighting. As a baseline, we use the CLIP model without fine-tuning to evaluate its performance on facial recognition tasks. We also compare it with a ResNet-based model to assess improvements in data efficiency and generalization.

6. Schedule & Roles

Research & Literature Review (2024-10-14 ̃2024-10-20)

Data Collection & Preprocessing (2024-10-21 ̃2024-10-27)

Model Development (2024-10-28 ̃2024-11-10)

Evaluation & Analysis (2024-11-11 ̃2024-11-17)

Report Writing & Final Submission (2024-11-18 ̃2024-11-24)