COMP-SCI 5588
Data Science Capstone
Week6 Handson

Professor: Dr Yugyung Lee
Term name: Bug Killers
Feb 28, 2025

Content

Links

GitHub: https://github.com/nanxuanhui/DSCapstone.git
Demo and data: https://drive.google.com/drive/folders/1RtW4V2NCIfo-9MSLoPdLQqdnaWesD8VR?usp=sharing

Individual Contributions

1. Team Members

Name	Contributions
Saniya Pandita	Frontend Developer: Developed the user interface using Streamlit, ensuring an intuitive and user-friendly design.
Jayadithya Nalajala	Project Lead: Stable Diffusion for synthetic data generation and augmentation.
Sai Jahnavi Devabhakthuni	NLP Engineer: Integration of Wav2Vec2 for speech emotion recognition.
Hui Jin	DL Engineer: Implementation of YOLOv5 for real-time object detection.

Introduction

This report provides an extensive overview of the progress made in Week 6 of the Data Science Capstone project. This week, our primary focus was on integrating advanced computer vision, speech emotion recognition, and generative AI techniques into our existing pipeline. In addition, we enhanced data processing techniques and refined model deployment strategies.

Techniques Implemented and Impact

1. Object Detection with YOLOv5

- Integrated YOLOv5 into the system for real-time object detection and tracking.
- Loaded custom-trained models using PyTorch Hub for improved accuracy.
- Designed an automated pipeline for retrieving and processing the latest experiment results for better evaluation.
- Impact: This feature is crucial for applications such as elderly monitoring, fall detection, and home security, allowing real-time alerts and automated responses based on detected objects or human activities.

2. **Speech Emotion Recognition with Wav2Vec2**
    - Leveraged Hugging Face’s wav2vec2-lg-xlsr-en-speech-emotion-recognition model for speech-based emotion analysis.
    - Developed an end-to-end pipeline to process speech input, classify emotions, and provide meaningful insights.
    - Impact: The ability to analyze speech emotions enables early detection of stress, anxiety, or emotional distress, improving mental health monitoring for elderly individuals or healthcare applications.
3. **Synthetic Data Generation with Stable Diffusion**
    - Implemented Stable Diffusion (CompVis/stable-diffusion-v1-4) to generate high-quality synthetic images.
    - Optimized inference settings to minimize computational overhead while maintaining image fidelity.
    - Impact: Synthetic data augmentation helps in training AI models with diverse datasets, reducing biases and improving generalization across different scenarios.

Functionalities Implemented

1. **Real-Time Object Detection**
    - YOLOv5 model deployed for detecting objects and human activity in real-time.
    - Developed an automated method to retrieve the latest experiment results for continuous evaluation.
2. **Speech Emotion Analysis**
    - Integrated a speech-to-emotion analysis pipeline to classify emotions such as neutral, happy, sad, angry, or fearful.
    - This can be combined with conversational AI to enhance human-computer interaction.
3. **Generative AI for Data Augmentation**
    - Stable Diffusion generates synthetic images based on textual descriptions.
    - Supports custom augmentation pipelines for training AI models with synthetic datasets.

Results and Findings

1. **YOLOv5 Model Performance**
    - Successfully integrated and tested the YOLOv5 model for real-time detection.
    - Automated experiment tracking improved evaluation efficiency and debugging.
    - Achieved high accuracy with reduced latency by optimizing model inference..
2. **Speech Emotion Recognition**
    - Achieved accurate classification of emotions in controlled environments.

- Performance decreased slightly in noisy conditions, necessitating further noise filtering techniques.

3. **Synthetic Data Generation**
   - Stable Diffusion generated high-resolution images with varying levels of realism.
   - Computational overhead was observed, suggesting further optimization.

Challenges

1. **Model Latency and Computational Overhead**
   - Running YOLOv5 and Stable Diffusion on standard hardware led to slow inference times.
   - Solution: Investigating model quantization, pruning, and acceleration using TensorRT.
2. **Emotion Model Accuracy in Noisy Environments**
   - Background noise reduced the accuracy of speech emotion recognition.
   - Solution: Implementing noise reduction and filtering techniques before feeding data to the model.
3. **Managing Large Model Dependencies**
   - Dependencies like diffusers and transformers increased deployment complexity.
   - Solution: Using Docker to containerize dependencies and simplify deployment.
4. **Model Latency**
   - The Whisper model had high latency for large audio files.
   - Solution: Optimized the model by truncating long audio files and processing them in chunks.

Future Improvements

- Optimize YOLOv5 inference using model pruning, TensorRT acceleration, and FPGA deployment.
- Enhance noise robustness in speech emotion recognition through advanced signal processing techniques.
- Expand generative AI capabilities by fine-tuning Stable Diffusion for domain-specific synthetic data generation.
- Implement AI-driven decision-making by combining vision, audio, and generative AI into a unified intelligent system.