LIP READER USING DEEP LEARNING MODEL

Project Reference No.:
47S_BE_3986

Gopalan College Of Engineering And Management

Computer Science And Engineering

Student(s) Name:
1. Darshan R
2. T Sai
3. Arya Biswas
4. Varun Reddy B

Guide(s) Name:
1. Dr. R. G. Sakthivelan
2. Dr. Swathi Y

OBJECTIVES

- The objectives of the non-real-time lip reading system revolve around developing an accurate and efficient solution for transcribing spoken words from pre-recorded video data. The primary objective is to design and implement a robust feature extraction and modeling framework capable of accurately capturing and interpreting lip movements. This involves optimizing deep learning architectures for batch processing, focusing on enhancing the system's ability to extract relevant spatial-temporal features from the video frames. Additionally, the system aims to achieve high transcription accuracy by training the model on a diverse dataset of pre-recorded video clips, encompassing variations in speakers, languages, and environmental conditions. Furthermore, the objective includes deploying the developed system for batch processing of pre-recorded video data, integrating it seamlessly into existing software infrastructure. Through these objectives, the lip reading system seeks to provide a reliable and effective solution for transcription tasks in non-real- time scenarios, catering to applications such as video analysis, assistive technologies, and transcription services.

METHODOLOGY

- The methodology for developing a non-real-time lip reading system begins with data collection, where a diverse dataset of pre-recorded video clips containing individuals speaking in various languages and environmental conditions is gathered. Following this, the collected data undergoes preprocessing to enhance its quality and prepare it for feature extraction, focusing on tasks such as resizing, normalization, and denoising. Feature extraction techniques, often based on deep learning architectures like convolutional neural networks (CNNs), are then employed to capture relevant spatial-temporal patterns in the preprocessed video frames. Subsequently, a lip reading model architecture is developed, optimized for batch processing of pre-recorded video data. This model is trained on the annotated dataset, utilizing optimization algorithms to maximize accuracy and robustness. Evaluation of the trained model's performance is conducted on a separate validation dataset, assessing metrics such as accuracy. Once the model demonstrates satisfactory performance, it is deployed for batch processing of pre-recorded video data, integrating into existing software infrastructure. Finally, the deployed system is monitored for performance on new batches of data, with regular updates and maintenance to address any identified issues and improve overall effectiveness. Through this methodology, developers can systematically design and deploy a non-real-time lip reading system tailored to processing pre-recorded video data efficiently and accurately.

WORK PLAN

1. Initial Setup and Research

- Install and configure essential libraries (OpenCV, TensorFlow, imageio, matplotlib, gdown).
- Conduct literature review on existing lip reading systems and deep learning models.

2. Data Collection and Preprocessing

- Collect and download relevant video datasets.
- Implement preprocessing steps: convert images to grayscale, isolate lip regions, and perform statistical analysis.
- Develop functions for loading and splitting video data and alignment paths.

3. Model Design and Development

- Design the deep neural network architecture using Conv3D and LSTM layers.
- Configure model parameters such as ReLU activation, MaxPool3D, Adam optimizer, and orthogonal kernel initialization.

WORK PLAN

4. Training the Model

- Implement a learning rate scheduler for dynamic adjustment during training.
- Train the model using the pre-processed data, employing CTC loss function for alignment sequences.
- Use ModelCheckpoint to save model checkpoints during training.

5. Model Evaluation and Validation

- Evaluate the model on a separate validation dataset to ensure generalization and robustness.
- Calculate performance metrics: accuracy, precision, recall, and F1 score.
- Analyze results to identify areas for improvement.

6. Making Predictions

- Utilize the trained model to make predictions on new pre-processed video data.
- Process the video data and obtain text representations from the model's output.

SUMMARY OF PROGRESS (DESIGN, EXPERIMENTS, RESULTS) (Maximum 2 Slides)

DESIGN

- **Architecture**:
  - Developed a deep neural network using Conv3D and LSTM layers for temporal modeling of lip movements.
  - Implemented preprocessing steps: grayscale conversion, lip region isolation, and statistical analysis.
  - Configured model parameters: ReLU activation, MaxPool3D, Adam optimizer, and orthogonal kernel initialization.
- **Data Processing**:
  - Created functions for loading video data, converting images to grayscale, isolating lip regions, and calculating statistical values.
  - Developed functions for loading alignment data, handling silence, splitting data, and converting characters to numbers.

EXPERIMENTS

- **Training Phase**:
  - Trained the model using pre-processed data with a learning rate scheduler and CTC loss function.
  - Employed ModelCheckpoint to save model checkpoints during 97 epochs.
- **Validation Phase**:
  - Evaluated the model on a separate validation dataset.

SUMMARY OF PROGRESS (DESIGN, EXPERIMENTS, RESULTS) (Maximum 2 Slides)

- Calculated performance metrics: accuracy, precision.

- **Prediction Phase**:
  - Utilized the trained model to make predictions on new pre-processed video data.
  - Processed video data through the model to predict corresponding text or phonetic representations.

RESULTS

- **Accuracy**:
  - Achieved over 90% accuracy on test datasets.
- **Foundation for Future Work**:
  - Established a strong foundation for future developments, including real-time processing and integration with assistive devices.

KSCST

NOVELTY OF THE PROJECT

- **Pre-Processed Lip Reading System**:
  - Utilizes advanced deep learning models (Conv3D and LSTM layers) for accurate interpretation of lip movements.
  - Incorporates sophisticated preprocessing techniques (grayscale conversion, lip region isolation, statistical analysis) to enhance model input.
- **Accessibility Enhancement**:
  - Targets communication improvement for individuals with hearing impairments, offering a more inclusive solution.
  - Designed to perform well in noisy environments where traditional audio-based methods may fail.
- **Generalization Across Conditions**:
  - Robust model performance across diverse speakers, lighting conditions, and background noises.
  - Ability to generalize across multiple languages, accents, and dialects.

SCOPE FOR FURTHER DEVELOPMENT

- **Real-Time Processing**:
  - Develop capabilities for real-time lip reading to enhance immediacy and usefulness in dynamic communication scenarios.
- **Mobile and Wearable Integration**:
  - Implement the system on mobile devices or wearable technology for increased accessibility and convenience.
- **Enhanced User Interface**:
  - Create a more intuitive and user-friendly interface, incorporating feedback from extensive user testing.

INDUSTRIAL COLLABORATION

- **Partnership Opportunities**:
  - Collaborate with tech companies specializing in assistive technologies to refine and scale the system.
  - Work with hardware manufacturers to optimize the system for specific devices and environments.

- **Healthcare and Educational Institutions**:
  - Partner with hospitals, clinics, and schools for individuals with hearing impairments to deploy the system and gather real-world usage data.
- **Funding and Grants**:
  - Seek funding from organizations and government bodies focused on disability inclusion and technological innovation.

IPR SUPPORT / LEVERAGING EXISTING PATENTS

- **Review Existing Patents**:
  - Conduct a thorough review of existing patents related to lip reading systems and identify any that are relevant to the project.
- **Patent Landscaping**:
  - Perform a patent landscape analysis to understand the scope of current intellectual property and identify potential areas for improvement or differentiation.
- **IP Strategy**:
  - Develop an IP strategy that focuses on enhancing and differentiating the existing technology to carve out new patentable areas or to protect improvements made to existing systems.