Wednesday, 11 September 2024

Conversational AI: Speech Recognition & Synthesis

Assignment

- This work, "Speech Commands: Dataset for Limited-Vocabulary Speech Recognition", presents a dataset of 65,000 one-second utterances for limited voice command recognition tasks. The dataset has been used more as a standard testbed to validate voice recognition models and to make the creation and evaluation of small-footprint models in resource-constrained environments easier. The dataset includes 30 short words, silence, and background noise.

Summary of my work:-

- **Data Preparation:** Labels and Features: For picture labels and features, use numpy arrays.
- **One-Hot Encoding:** Use tf.keras.utils.to_categorical to convert categorical labels into a format that is one-hot encoded.
- **Creation of Datasets:**
- **TensorFlow Dataset:** Use tf.data to transform numpy arrays into TensorFlow datasets.From_tensor_slices dataset.

To improve training dynamics, apply batching and shuffling to the training dataset.

- **Model Defined:**
- **Architecture:** Construct a Sequential model using MaxPooling2D, Dense, Flatten, and Conv2D layers.
- **Output Layer**: Verify that there are the appropriate number of units (e.g., 12 units for 12 classes) in the output layer.

- **Gathering:**
- **Diminished Function:** For labels that are encoded one-hot, use categorical_crossentropy.
- **Optimizer:** Use a suitable optimizer, such as Adam, to compile the model and provide assessment metrics.
- **Instruction:**

Model Fitting: Use a model to train the model.in line with the instruction and verification along with test data