PROJECT REPORT

Lip Reading using Deep Learing

Padmavasan

INDEX

INTRODUCTION

The advent of deep learning has revolutionized various domains, including computer vision and natural language processing. This project, titled "Lipreading using Deep Learning," leverages the power of deep neural networks to bridge the gap between visual information from lip movements and the understanding of spoken language. By employing advanced machine learning techniques, this project aims to develop a robust and efficient system for automatic lipreading, contributing to enhanced communication accessibility.

PURPOSE

The primary objective of this project is to design and implement a deep learning model capable of accurately transcribing spoken words by analyzing visual cues from lip movements. The model will be trained on a comprehensive dataset of audio-visual samples, encompassing diverse speakers, languages, and contextual variations. The goal is to create a system that not only improves communication accessibility for individuals with hearing impairments but also finds applications in scenarios where audio information is compromised.

TRAINING

```
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv3d_3 (Conv3D)           (None, 75, 46, 140, 128   3584
                             )

 activation_3 (Activation)   (None, 75, 46, 140, 128   0
                             )

 max_pooling3d_3 (MaxPoolin  (None, 75, 23, 70, 128)   0
 g3D)

 conv3d_4 (Conv3D)           (None, 75, 23, 70, 256)   884992

 activation_4 (Activation)   (None, 75, 23, 70, 256)   0

 max_pooling3d_4 (MaxPoolin  (None, 75, 11, 35, 256)   0
 g3D)

 conv3d_5 (Conv3D)           (None, 75, 11, 35, 75)    518475

 activation_5 (Activation)   (None, 75, 11, 35, 75)    0

 max_pooling3d_5 (MaxPoolin  (None, 75, 5, 17, 75)     0
 g3D)

 time_distributed_1 (TimeDi  (None, 75, 6375)          0
 stributed)

 bidirectional_2 (Bidirecti  (None, 75, 256)           6660096
 onal)

 dropout_2 (Dropout)         (None, 75, 256)           0

 bidirectional_3 (Bidirecti  (None, 75, 256)           394240
 onal)

 dropout_3 (Dropout)         (None, 75, 256)           0

 dense_1 (Dense)             (None, 75, 41)            10537

=================================================================
Total params: 8471924 (32.32 MB)
Trainable params: 8471924 (32.32 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

```
Epoch 1/100
1/1 [==============================] - 2s 2s/step
Original: set green by j two please
Prediction: la e e e e n n
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Original: lay white by s three again
Prediction: la e e e e e n
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
450/450 [==============================] - 883s 2s/step - loss: 83.6646 - val_loss: 71.0522 - lr: 1.0000e-04
Epoch 2/100
1/1 [==============================] - 0s 212ms/step
Original: bin green in a three soon
Prediction: la e e e e oa
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Original: lay white with m zero now
Prediction: la e e e e oa
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
450/450 [==============================] - 663s 1s/step - loss: 70.3704 - val_loss: 66.3020 - lr: 1.0000e-04
Epoch 3/100
1/1 [==============================] - 0s 213ms/step
Original: lay blue in d two now
Prediction: la e e t e on
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Original: place blue by i eight now
Prediction: la e e t e on
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
450/450 [==============================] - 659s 1s/step - loss: 66.1969 - val_loss: 60.6679 - lr: 1.0000e-04
Epoch 4/100
1/1 [==============================] - 0s 203ms/step
Original: lay green with a one again
Prediction: la re t e eon
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Original: bin red sp by g two now
Prediction: la re t e eon
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

TESTING

Test on a Video

```
[ ] sample = load_data(tf.convert_to_tensor('.\\data\\s1\\bras9a.mpg'))
```

```
[ ] print('~'*100, 'REAL TEXT')
    [tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in [sample[1]]]
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ REAL TEXT
[<tf.Tensor: shape=(), dtype=string, numpy=b'bin red at s nine again'>]
```

```
[ ] yhat = model.predict(tf.expand_dims(sample[0], axis=0))
```

```
1/1 [==============================] - 1s 720ms/step
```

```
[ ] decoded = tf.keras.backend.ctc_decode(yhat, input_length=[75], greedy=True)[0][0].numpy()
```

```
[ ] print('~'*100, 'PREDICTIONS')
    [tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in decoded]
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ PREDICTIONS
[<tf.Tensor: shape=(), dtype=string, numpy=b'bin red at s nine again'>]
```

```
[ ]
```

```
[ ]
```

APPLICATIONS

Accessibility for Hearing-Impaired Individuals

The primary application is in enhancing communication accessibility for individuals with hearing impairments. The developed lipreading system can serve as a valuable tool for these individuals, providing an alternative means of understanding spoken language.

Forensic Analysis

In forensic investigations, lipreading technology can be employed to analyze video footage and aid in deciphering spoken words or conversations. This can be valuable in criminal investigations or legal proceedings.

RESULT

```
model.load_weights('models/checkpoint')
```

```
test_data = test.as_numpy_iterator()
```

```
sample = test_data.next()
```

```
yhat = model.predict(sample[0])
```

```
1/1 [==============================] - 1s 973ms/step
```

```
print('~'*100, 'REAL TEXT')
[tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in sample[1]]
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ REAL TEXT
[<tf.Tensor: shape=(), dtype=string, numpy=b'place white at x six please'>,
 <tf.Tensor: shape=(), dtype=string, numpy=b'lay blue in x four now'>]
```

```
decoded = tf.keras.backend.ctc_decode(yhat, input_length=[75,75], greedy=True)[0][0].numpy()
```

```
print('~'*100, 'PREDICTIONS')
[tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in decoded]
```

```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ PREDICTIONS
[<tf.Tensor: shape=(), dtype=string, numpy=b'place white at x six please'>,
 <tf.Tensor: shape=(), dtype=string, numpy=b'lay blue in x four now'>]
```

CONCLUSION

In conclusion, the "Lipreading using Deep Learning" project represents a significant step forward in leveraging advanced technology to enhance communication accessibility and address challenges faced by individuals with hearing impairments. The successful development and implementation of a deep learning-based lipreading system underscore the potential of artificial intelligence to bridge gaps in human communication.