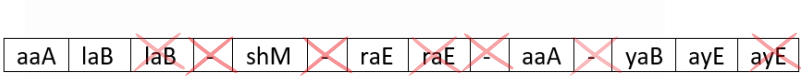
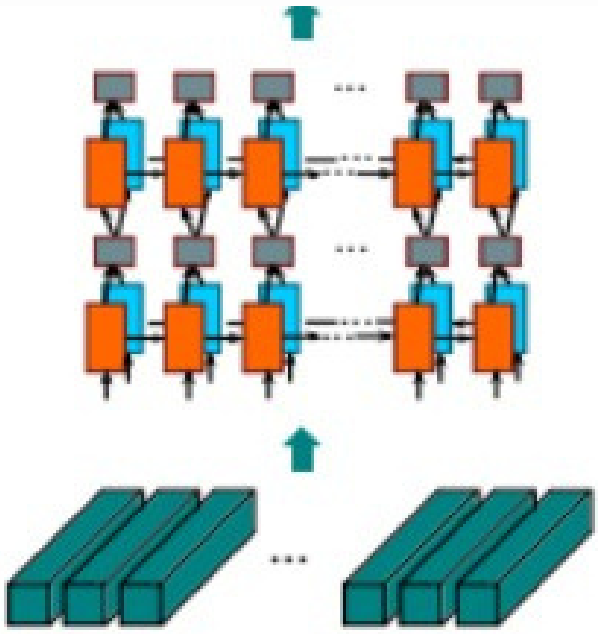
Recognize Handwritten Arabic Texts by using a combined architecture : CNN+RNN with CTC

Transcription Layer (CTC)

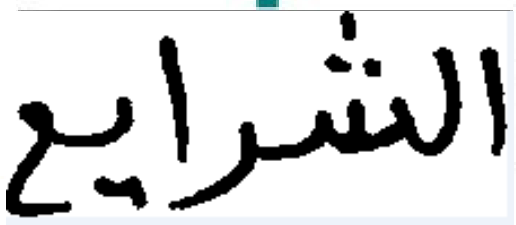
aaA|laB|shM|raE|aaA|yaB|ayE| الشرايع



RNN Layer



CNN Layer



Model Architecture

The number of layers in each part of the model are as follows:

Convolutional Neural Network (CNN):

- Convolutional Layer 1: 1 layer (Conv2D)
- Convolutional Layer 2: 1 layer (Conv2D)
- Max Pooling Layer 1: 1 layer (MaxPooling2D)
- Max Pooling Layer 2: 1 layer (MaxPooling2D)

Recurrent Neural Network (RNN):

- Bidirectional LSTM Layer: 1 layer (Bidirectional LSTM)

Other Layers:

- Reshape Layer: 1 layer (Reshape)
- Dense Layer: 1 layer (Dense)
- Output Layer: 1 layer (Dense)

Layer (type)	Output Shape	Param #	Connected to
the_input (InputLayer)	[(None, 300, 100)]	0	[]
lambda (Lambda)	(None, 300, 100, 1)	0	['the_input[0][0]']
conv1 (Conv2D)	(None, 300, 100, 16)	160	['lambda[0][0]']
pool1 (MaxPooling2D)	(None, 150, 50, 16)	0	['conv1[0][0]']
conv2 (Conv2D)	(None, 150, 50, 16)	2320	['pool1[0][0]']
pool2 (MaxPooling2D)	(None, 75, 25, 16)	0	['conv2[0][0]']
reshape (Reshape)	(None, 75, 400)	0	['pool2[0][0]']
dense (Dense)	(None, 75, 32)	12832	['reshape[0][0]']
biLSTM (Bidirectional)	(None, 75, 1024)	2232320	['dense[0][0]']
softmax (Dense)	(None, 75, 160)	164000	['biLSTM[0][0]']
the_labels (InputLayer)	[(None, 17)]	0	[]
input_length (InputLayer)	[(None, 1)]	0	[]
label_length (InputLayer)	[(None, 1)]	0	[]
ctc (Lambda)	(None, 1)	0	['softmax[0][0]', 'the_labels[0][0]', 'input_length[0][0]', 'label_length[0][0]']