

# Handwritten Arabic Text Recognition

Project Milestone 1



# Outline

- Summary of proposal
- Problem statement
- Related work
- Progress
- Comparison
- To do next
- Team members contribution



# Summary of proposal

- **Problem statement:** Handwritten Arabic Text Recognition.
- **Available Datasets:** IFN/ENIT, AlexU-Word and **KHATT**.
- **Selected Dataset:** KHATT dataset because It is the largest dataset, provides text lines images, not just words, has a simple ground truth and it is a well-organized dataset and divided into train, validation and test sets.
- **Available Models:** discussed in Related Work.
- **Selected Model:** HTR-Flor Model which is used to recognize English text with cursive letters like Arabic text.
- **Evaluation Metrics:** character error rate (CER) and word error rate (WER).
- **Graduation Project:** An OCR system to read medical prescriptions.



# Problem Statement

## Arabic Handwriting Text Recognition

A challenging problem !

- More complex compared to handwritten Latin text recognition
- Cursive script and joined writing
- Same character variations
- Large number of words
- Variations in font style



# Related Work

[1] KHATT: a Deep Learning Benchmark on Arabic Script

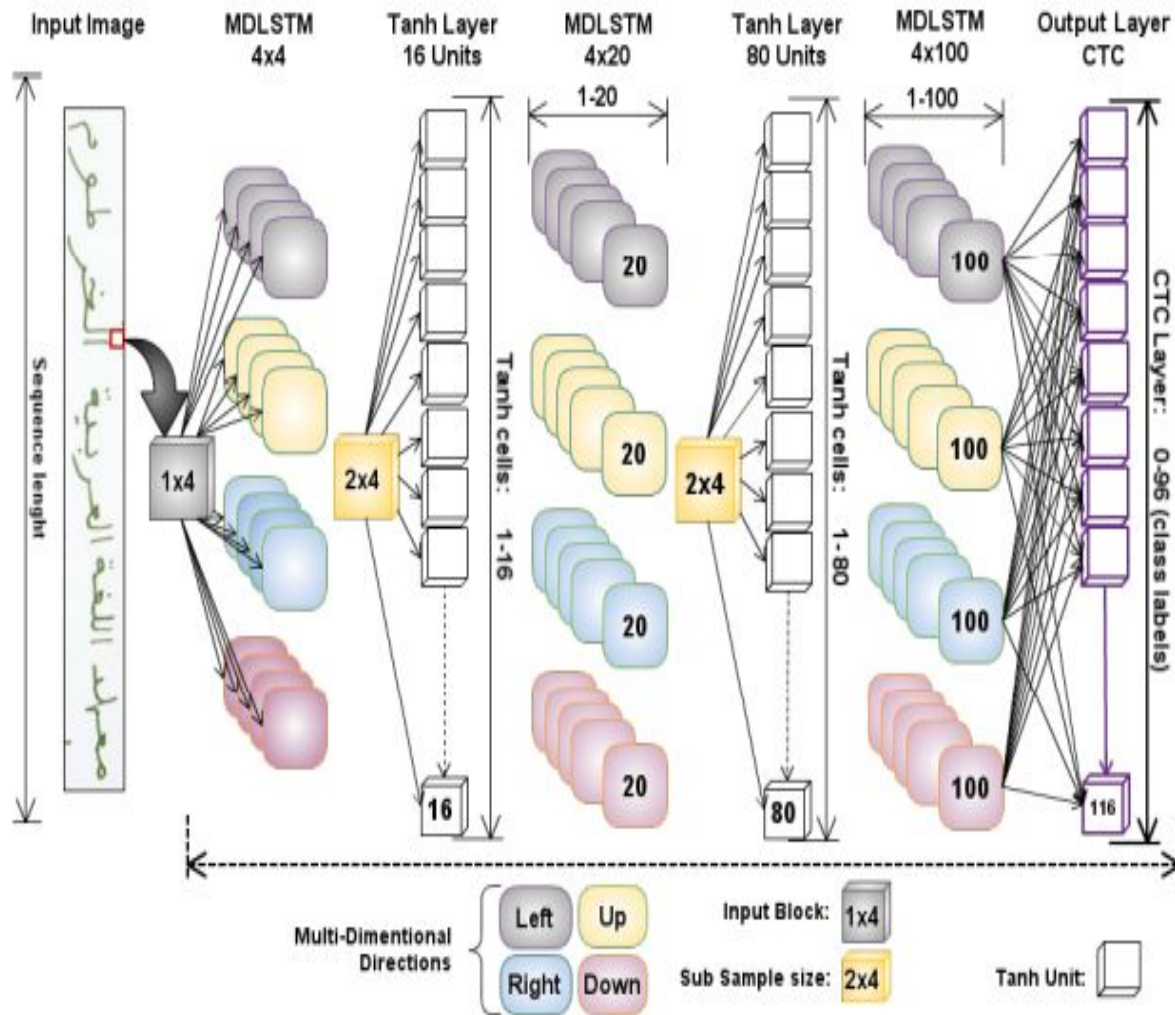
## Model Architecture

- Input Layer.
- 3 MD-LSTM with Tanh layer.
- Output Layer (CTC).

- In the notation MD-LSTM-4x4, the first 4 represents the number of directions to be scan, while the 20 represents the number of LSTM units.

- In the output layer (0-95) units represents number of classes(number of characters) + 96 unit for blank space.

- Recognition rate was 75.8%

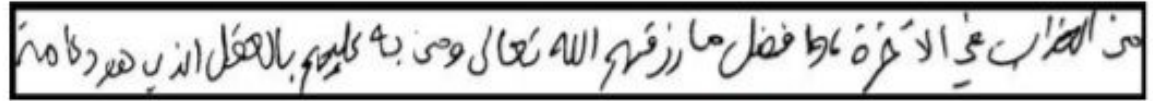


## Related Work cont.

[2] A Deep Learning based Arabic Script Recognition System: Benchmark on KHAT

Same model used in the previous paper are used here but with adding data-augmentation techniques as they added 4 extra samples for each original KHATT instance and which make the results to be improved.

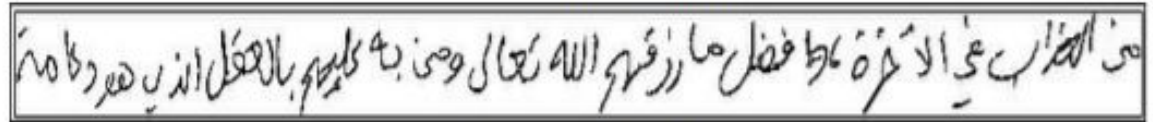
**Recognition rate increased to 80.02%**



a) Original sample.



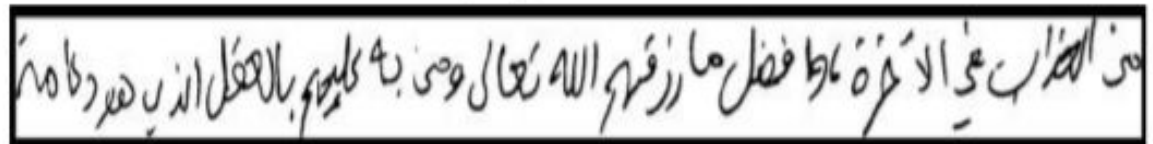
b) Blur sample.



c) Contours of text-line.



d) Edge enhancement.



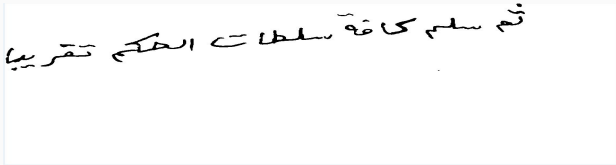
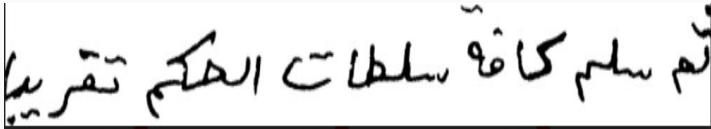
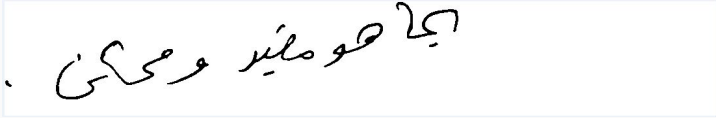
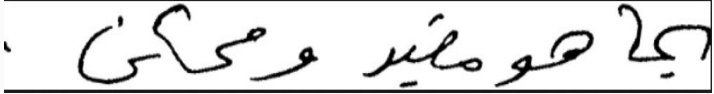
e) Bleed through effect.

Figure 2. The resulted samples text-lines after Data-augmentation.

# Progress

- **Filter KHATT dataset:**
  - Remove extra text files from ground truth to make number of images equal number of text files in the ground truth.
  - Find images that don't contain text files in the ground truth and build their ground truth by hand.
- **Applying preprocessing on KHATT dataset:**
  - prune the extra white regions that are present in Arabic text-lines.
  - de-skew some of the Arabic text-lines that are skewed using skew detection and correction techniques.
  - Resize images.

## Progress cont.

Before Preprocessing	After Preprocessing
	
	



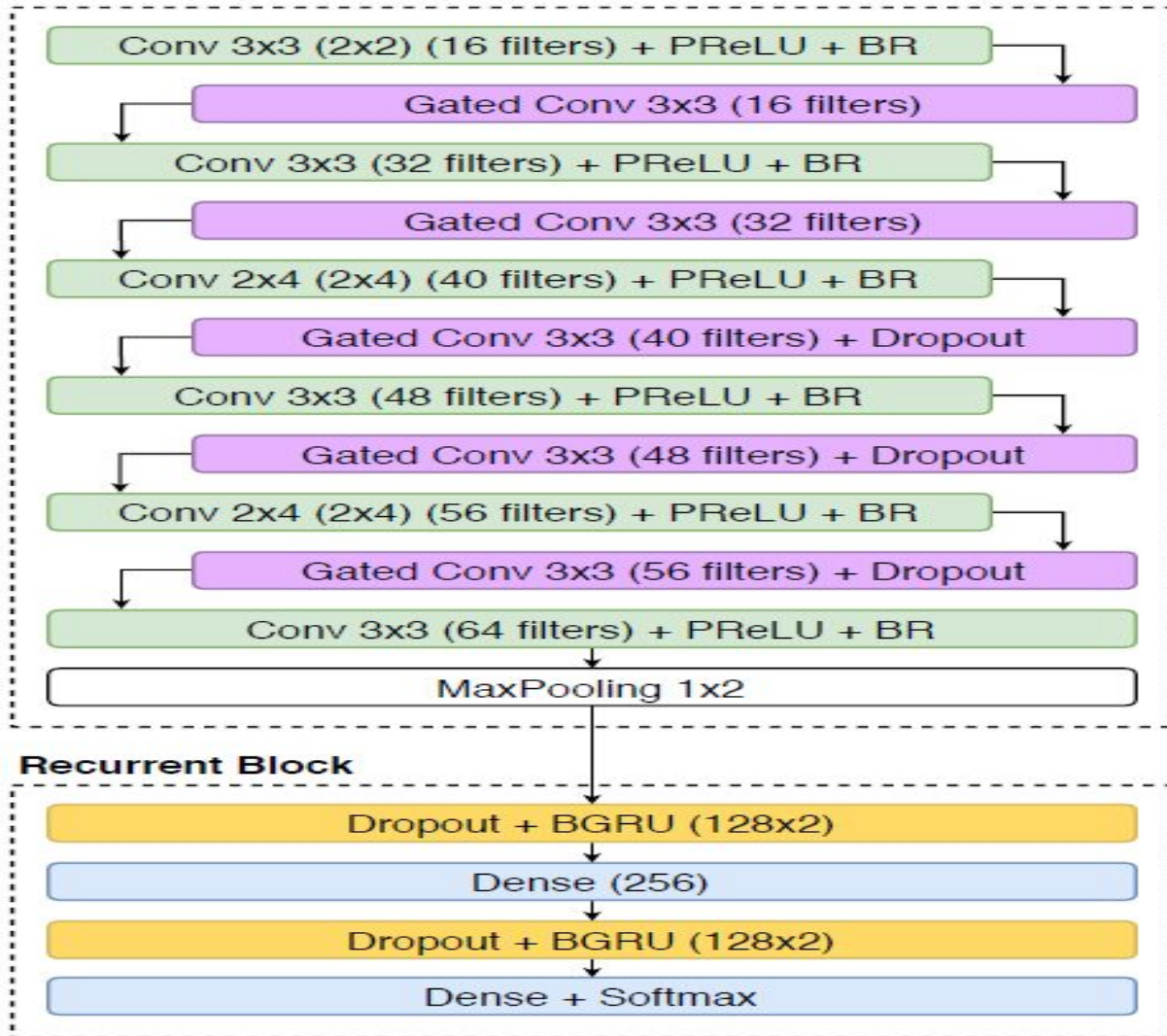
# Progress cont.

## Model 1 Architecture

- Input Layer.
- 6 CONV Layer with 5 Gated CONV Layer.
- Max-Pooling Layer.
- 2 Bidirectional GRU Layer with Dropout.
- 2 Dense Layer.

- Gated CONV Layer → This gated mechanism, uses all input features ( $x$ ) to perform a sigmoid activation ( $s$ ) and the result is a pointwise multiplication between input (original features) and output features ( $y = s(x) \cdot x$ ). Gated-CNN extracts the features from the images first before enter them to the recurrent layers.

- Output Layer contains 87 units: 86 units which represent number of classes (number of characters) + 1 unit for class blank space.

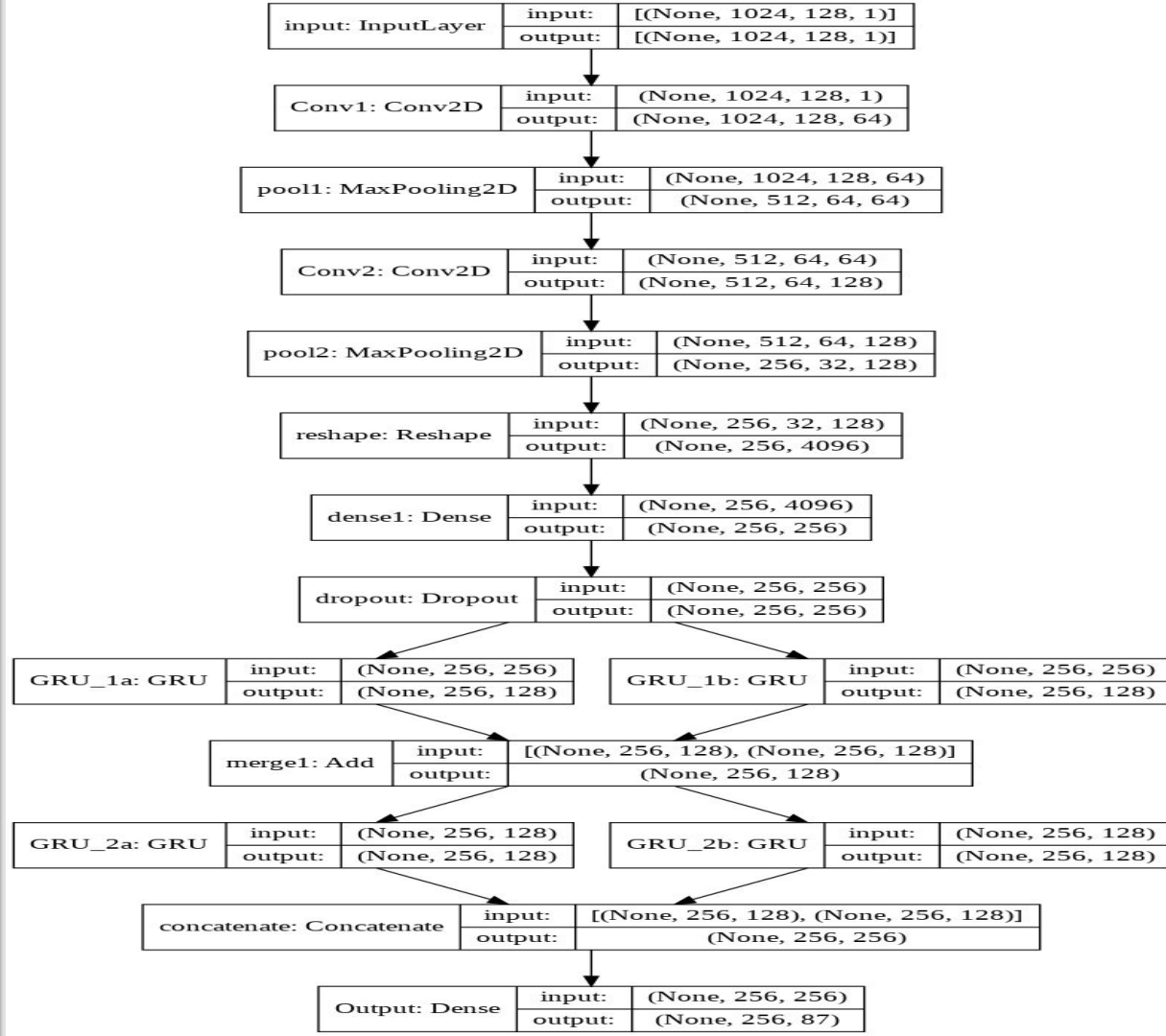


# Progress cont.

## Model 2 Architecture

- Input Layer.
- 2 CONV Layer with Max-Pooling.
- Dense Layer with Dropout.
- 2 GRU Layer.
- Add Layer.
- 2 GRU Layer.
- Concatenation Layer.
- Dense Layer (Output Layer).

- Output Layer contains 87 units: 86 units which represent number of classes (number of characters) + 1 unit for class blank space.

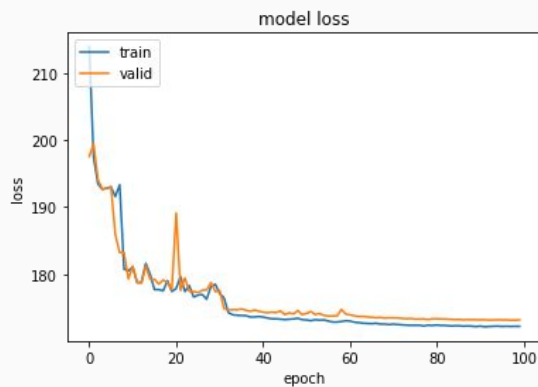


# Comparison

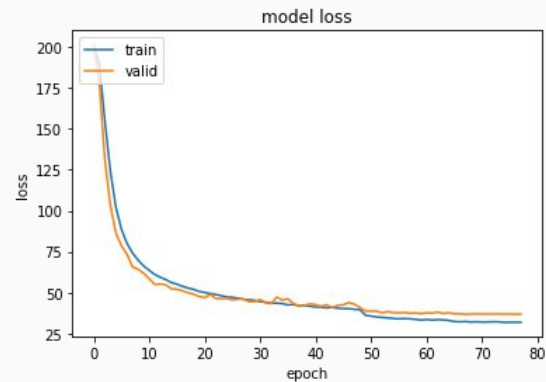
	Model 1	Model 2
Number of parameters	919,527	1,640,279
Total time	3:09:23	1:23:26
Time per epoch	0:01:54	0:01:04
Train loss	172.22	32.3
Best validation loss	173.15	36.74
CER	1.0	0.1893
WER	1.0	0.5981

# Comparison cont.

Model 1



Model 2



## Comparison cont.

	<b>NOE to get best result</b>	<b>CER (%)</b>	<b>WER(%)</b>	<b>Recognition Rate (%)</b>
<b>Related Work (1)</b>	70	19.98%	—	80.02%
<b>Our Progress (2)</b>	68	18.93%	59.81%	81.07%

NOE → number of epochs

CER → character error rate

# Results

مكتسب بالتجارب والا ادب، وله غريزة مكنونه في الانسان

TE\_L: مكتسب بالتجارب والأدب. وله غريزة مكنونة في الانسان  
TE\_P: مكتسب بالتجارب والادب وله عريزه مكنونه في الانسان

ومفتاح كل سعادته، فليس لاصد غني عن العقل، والعقل

TE\_L: ومفتاح كل سعادته. فليس لأحد غني عن العقل . والعقل  
TE\_P: ومفتاح كل سعادته. فليس لاحد. غني عن العقل. والعقل

# To do next

- try to tune the hyper parameters of the model to get better results on KHATT dataset
- Use attention layer in model 2, we assume that we will get more improvement result.



# Team Members contribution

## *Merna El-Refaie (57)*

- Preprocess KHATT dataset
- Train Model 1

## *Neveen S.Nagy (61)*

- Filter KHATT dataset
- Train Model 2

## *We Both*

- Implement and build Model 1
- Search for other models to get best result





# References

- [1] KHATT: A Deep Learning Benchmark on Arabic Script
- [2] A Deep Learning based Arabic Script Recognition System: Benchmark on KHAT
- [3] HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition
- [4] Handwritten-text-recognition Model 1
- [5] Image OCR Model 2



# Notebooks

[1] Filter KHATT Dataset

[2] Preprocess KHATT Dataset

[3] FlorModel\_ArabicText --> Model 1

[4] HTRModel\_ArabicText --> Model 2



# Any Questions?!



# Thanks!

Contact us:

**Merna El-Refaie**

- [mernamustafa1997@gamil.com](mailto:mernamustafa1997@gamil.com)
- [eng-merna.mustafa1621@alexu.edu.eg](mailto:eng-merna.mustafa1621@alexu.edu.eg)

---

**Neveen S.Nagy**

- [neveensamirnagy@gmail.com](mailto:neveensamirnagy@gmail.com)
- [es-neveensamirnagy@alexu.edu.eg](mailto:es-neveensamirnagy@alexu.edu.eg)

