# Project Documentation

---

# Blnk Task

---

**Arabic Optical Character Recognition**

**By: Mohamed Sherif El-Boraie**

# Blnk Company Machine Learning Project Documentation

## Table of content

# Introduction

As a dedicated and diligent Machine Learning Engineer, I embarked on an ambitious project at Blnk Company that involved Optical Character Recognition (OCR) for Arabic text. This documentation serves as a testament to the immense effort, dedication, and learning I poured into this endeavor.

I recognized the significance of OCR technology, especially when applied to the complexities of the Arabic language. My journey began with the recognition of the challenges posed by unprocessed OCR tasks, particularly for Arabic text. These challenges included the presence of two distinct datasets: one containing blurred images and the other comprising sharpened images.

# Problem Statement

The core problem statement of this project was to design and implement a state-of-the-art OCR system for Arabic text recognition. This undertaking demanded a multifaceted approach, including the following key aspects:

*OCR for Arabic Text:* Addressing the unique complexities of Arabic script in OCR tasks.

Dual Datasets: Handling and making meaningful use of two datasets, one consisting of blurred images and the other containing sharpened images.

*CTC Loss and Encoding/Decoding:* Gaining proficiency in Connectionist Temporal Classification (CTC) loss and its encoding/decoding techniques.

*Image Enhancement:* Exploring methods to enhance the quality of blurred images, with a specific focus on deep deblurring.

*Segmentation:* Tackling the challenge of segmenting images that contain two sentences with different fonts.

*Model Generalization:* Developing a model that can adapt to varying datasets, thereby ensuring its versatility.

# Data Overview

*Dataset*

The project leveraged two distinct datasets: one comprising blurred images and the other containing sharpened images with 8120 images. However, due to computational constraints and to maintain a focus on the primary objective of building a robust OCR model, I chose to concentrate primarily on sharpened images during the project's course.

*Data Preprocessing*

The critical step of data preprocessing encompassed the following operations:

In my pursuit of improving the quality of our dataset, I explored two distinct approaches: deep deblurring and data augmentation using Generative Adversarial Networks (GANs).

Deep Deblurring:

Deep deblurring is a sophisticated image processing technique aimed at removing blur and restoring sharpness to images. It involves the use of intricate convolutional neural networks (CNNs) and requires training on a large dataset containing both blurred and sharp images. While this approach holds immense promise in enhancing image quality, it comes with a substantial computational cost.

In practice, dealing with deep deblurring in environments like Google Colab and Jupyter, which rely on limited computational resources, proved to be a time-consuming endeavor. Even with a reasonably capable CPU, applying the model demanded a significant amount of time. Given the primary focus of our task, which centers around building the OCR model, the computational expense associated with deep deblurring was deemed impractical. Consequently, I made the strategic decision to primarily work with sharpened images, as our core objective is to develop a robust OCR model rather than perfect image quality.

Data Augmentation and Thresholding:

An alternative approach to enhancing image quality is through data augmentation using GANs. However, after careful consideration, I opted for a different route—thresholding. Thresholding is a technique that involves setting a specific threshold value and classifying pixels in an image as either foreground or background based on their intensity. In my experiments, I found that thresholding offered a more efficient and suitable choice for our dataset.

This strategic decision allows us to allocate our computational resources more effectively, focusing on the core challenge of OCR model development. By streamlining our approach, we can optimize the model's performance and achieve our primary goal of accurate Arabic text recognition.

Grayscaling: Conversion of the images to grayscale to simplify and standardize the data.

Thresholding: Application of both Binary and NiBlack thresholding techniques for image preprocessing. Surprisingly, Binary Thresholding showed superior results for the dataset.

Grayscaling was an essential preprocessing step to reduce dimensionality and simplify the input for subsequent model layers.

Two thresholding techniques were applied: Binary Thresholding and NiBlack Thresholding, with Binary Thresholding exhibiting better compatibility with the dataset.

# Model Architecture

Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM)

The heart of the OCR model architecture was a fusion of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) layers. Key architectural elements included:

Convolutional Layers: Employed for feature extraction.

Max-Pooling Layers: Applied for down sampling feature maps.

LSTM Layers: Incorporated to capture sequential information.

CTC Loss Layer: Facilitated sequence-to-sequence alignment for effective OCR.

# Training the Model

*Training Process*

The model training process involved several crucial steps:

Optimizer: Adam optimizer was employed for model training.

Early Stopping: Early stopping mechanisms were implemented to prevent overfitting.

TensorBoard: TensorBoard was utilized for real-time monitoring and visualization of the training process.

Limited Epochs: Due to computational constraints, a limited number of epochs were executed during training.

*Training Results*

The outcomes of model training were noteworthy:

Training Loss: It consistently decreased with each epoch.

Validation Loss: Similar to training loss, validation loss also exhibited a decreasing trend, signifying model convergence.

Training Time: The model training process was computationally expensive, necessitating a restriction on the number of epochs.

*Model Evaluation*

The model evaluation phase involved decoding the CTC loss predictions and comparing them against ground truth labels. Various evaluation metrics such as accuracy, precision, and recall were employed to comprehensively assess the model's performance.

*Conclusion*

In conclusion, this OCR project for Arabic text recognition has been a significant learning journey. Despite the challenges posed by limited computational resources, I have made substantial progress in building a robust OCR model. There is immense potential for further improvements, particularly with additional computational resources and more advanced image enhancement and segmentation techniques.

# Future Improvements

Future enhancements and explorations for this project may encompass:

- Exploring advanced image enhancement techniques to enhance OCR accuracy.
- Investigating and implementing more robust segmentation methods to handle varying datasets.