What is deep learning-based OCR?

OCR has become very popular nowadays and has been adopted by several industries for faster text data reading from images. While solutions like contour detection, image classification, connected component analysis, etc. are used for documents that have comparable text size and font, ideal lighting conditions, good image quality, etc., such methods are not effective for irregular, heterogeneous text often called wild text or scene text. This text could be from a car’s license plate, house number plate, poorly scanned documents (with no predefined conditions), etc. For this, Deep Learning solutions are used. Using DL for OCR is a three-step process and these steps are:

1. Preprocessing: OCR is not an easy problem, at least not as easy as we think it to be. Extracting text data from digital images/documents is still fine. But when it comes to scanned or phone-clicked images things change. Real-world images are not always clicked/scanned in ideal conditions, they can have noise, blur, skewness, etc. That needs to be handled before applying the DL models to them. For this reason, image preprocessing is required to tackle these issues.

2. Text Detection/Localization: At this stage models like Mask-RCNN, East Text Detector, YoloV5, SSD, etc. are used that locates the text in images. These models usually create bounding boxes (square/rectangle boxes) over each text identified in the image or a document.

3. Text Recognition: Once the text location is identified, each bounding box is sent to the text recognition model which is usually a combination of RNNs, CNNs, and Attention networks. The final output from these models is the text extracted from the documents. Some open-source text recognition models like Tesseract, MMOCR, etc. can help you gain good accuracy.