Project proposal

Topic: Scene Text Detection

I.Introduction

Scene Text Detection is a task to detect text regions in the complex background and label them with bounding boxes.

II. Project details

1. Dataset

500 ảnh cung cấp bởi BK.AI: [link](https://drive.google.com/drive/folders/1WNOk3EMSgawdbrHeLiO7N_WcapgmbdmV)

2000 ảnh cung cấp bởi VinAI: [link](https://drive.google.com/file/d/1UUQhNvzgpZy7zXBFQp0Qox-BBjunZ0ml/view)

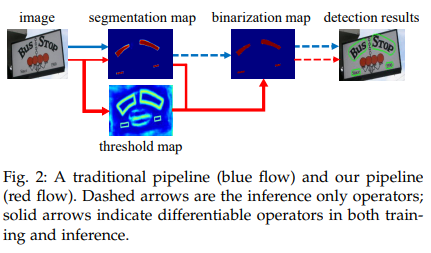
Evaluation script: [link](https://github.com/SakuraRiven/EAST)

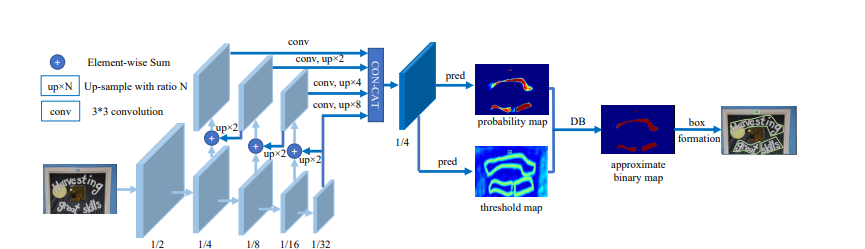
1. Model and paper

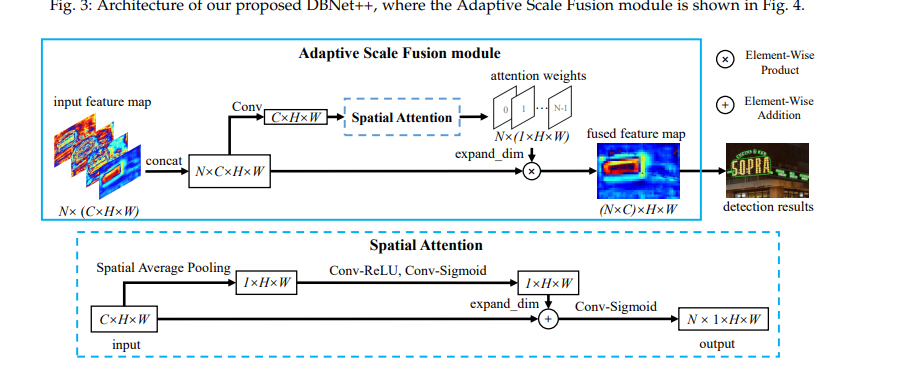
2.1. DBNet (Nguyễn Duy Khánhh)

https://arxiv.org/pdf/2202.10304v1.pdf

Recently, segmentation-based methods are quite popular in scene text detection, as the segmentation results can more accurately describe scene text of various shapes such as curve text. However, the post-processing of binarization is essential for segmentation-based detection, which converts probability maps produced by a segmentation method into bounding boxes/regions of text. In this paper, we propose a module named Differentiable Binarization (DB), which can perform the binarization process in a segmentation network. Optimized along with a DB module, a segmentation network can adaptively set the thresholds for binarization, which not only simplifies the post-processing but also enhances the performance of text detection



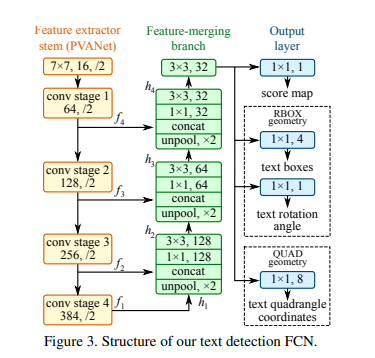




2.2. EAST ( Nguyễn Đức Quyết + Nguyễn Hữu Tuấn Duy)

In this work, we propose a simple yet powerful pipeline that yields fast and accurate text detection in natural scenes. The pipeline directly predicts words or text lines of arbitrary orientations and quadrilateral shapes in full images, eliminating unnecessary intermediate steps (e.g., candidate aggregation and word partitioning), with a single neural network

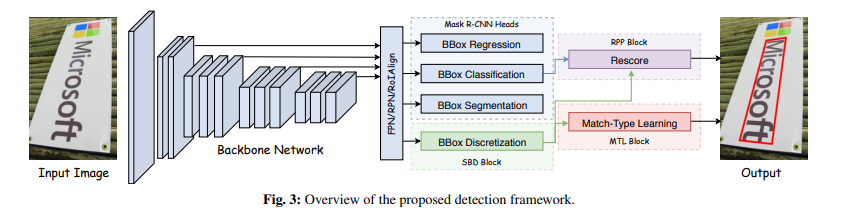
<https://arxiv.org/pdf/1704.03155v2.pdf>

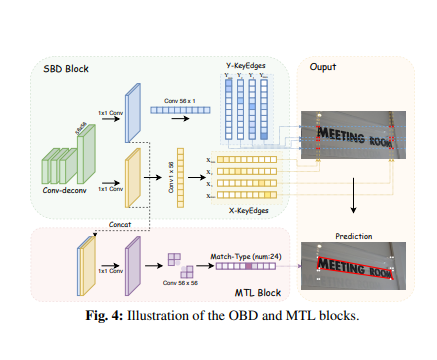


2.3. SBD (Nguyễn Phương Uyên + Nguyễn Phúc Thành)

<https://arxiv.org/pdf/1912.09629v3.pdf>

Multi-orientation scene text detection has recently gained significant research attention. Previous methods directly predict words or text lines, typically by using quadrilateral shapes. However, many of these methods neglect the significance of consistent labeling, which is important for maintaining a stable training process, especially when it comprises a large amount of data. Here we solve this problem by proposing a new method, Orderless Box Discretization (OBD), which first discretizes the quadrilateral box into several key edges containing all potential horizontal and vertical positions. To decode accurate vertex positions, a simple yet effective matching procedure is proposed for reconstructing the quadrilateral bounding boxes. Our method solves the ambiguity issue, which has a significant impact on the learning process

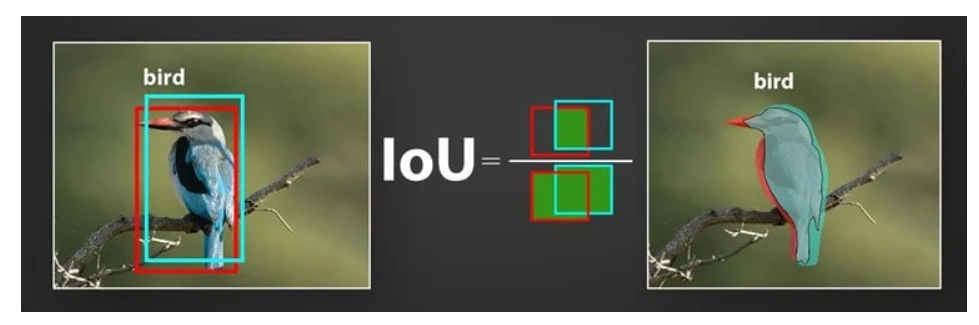




1. Metric

1. Intersection Over Union

Intersection over Union (IoU) is a metric that quantifies the degree of overlap between two regions. We have already discussed IoU meaning in-depth in our previous article Intersection over Union in Object Detection and Segmentation. Check out the article to know more about how IoU metric is designed and implemented.

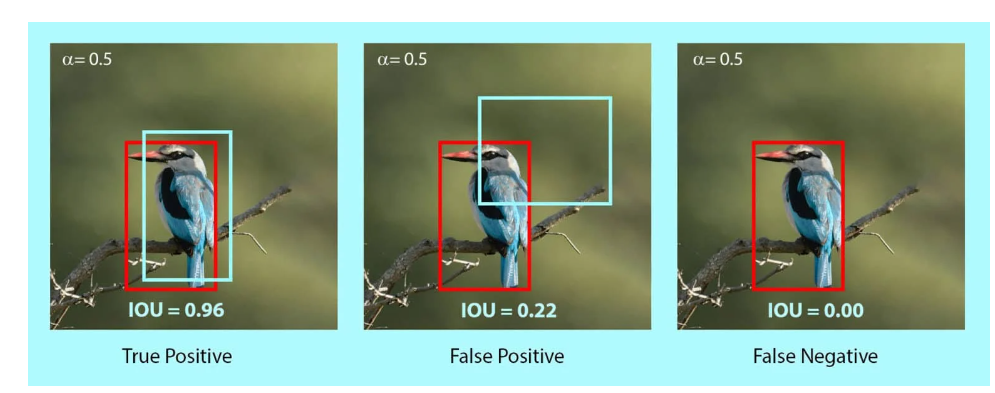


As mentioned above, IoU metric evaluates the correctness of a prediction. The value ranges from 0 to 1. With the help of IoU threshold value, we can decide whether a prediction is True Positive, False Positive, or False Negative. Confused with the terms? Interestingly, these terms collectively form the Confusion Matrix.

2. Confusion Matrix

To understand the Confusion Matrix, let’s take an example of a classification problem. A model has to identify whether there are hot dogs in the image. The predictions can be either correct or incorrect. The following combination of predictions is possible based on the output and the actual image.

Similarly, these terms apply to object detection and segmentation as well. However, the exact meanings are not the same. In object detection, the correctness of the prediction (TP, FP, or FN) is decided with the help of IoU threshold. Whereas in image segmentation, it is decided by referencing the Ground Truth pixels. Ground truth meaning known objects.



3. Precision

Precision measures the proportion of predicted positives that are actually correct. If you are wondering how to calculate precision, it is simply the True Positives out of total detections. Mathematically, it’s defined as follows.

P = TP/(TP + FP)

= TP / Total Predictions

The value ranges from 0 to 1.

4. Recall

Recall measures the proportion of actual positives that were predicted correctly. It is the True Positives out of all Ground Truths. Mathematically, it is defined as follows.

R = TP / (TP + FN)

= TP / Total Ground Truths

Similar to Precision, the value of Recall also ranges from 0 to 1.