1. Scene Text Detection via Holistic, Multi-Channel Prediction

Đặc điểm nổi bật của phương pháp:

* Có khả năng phát hiện được nhiều vùng văn bản, kể cả văn bản với các hướng khác nhau trong một ảnh tự nhiên.
* Sử dụng phương pháp tiếp cận tương tự như semantic segmentation với 2 class (text và non-text).

Phương pháp tiếp cận:

* Bài báo giải quyết vấn đề phát hiện văn bản tưởng tự như giải quyết vấn đề của bài toán phân vùng ngữ nghĩa (semantic segmentation) khác biệt ở đây là chỉ có 2 class là text và non-text (phương pháp được dựa trên nghiên cứu của Xie & Tu "Holistically-Nested Edge Detection" [1]), dự đoán dựa trên việc phân tích tổng quan toàn bộ bức ảnh cách này trái ngược với những cách tiếp cận trước đó là đưa ra dự đoán mang tính cục bộ. Vì vậy, phương pháp trước đây không sử dụng đầy đủ tất cả các thông tin trong toàn bộ bức ảnh.
* Phương pháp này đồng thời dự đoán được xác suất của vùng text, kí tự và mối quan hệ giữa các kí tự lân cận. Đặc biệt, nó có khả năng phát hiện được nhiều vùng của văn bản, các văn bản có hướng khác nhau.
* Thuật toán được đánh giá tốt hơn đáng kể so với các thuật toán trước đây.

1. Orientation Robust Text Line Detection in Natural Images

Đặc điểm nổi bật của phương pháp:

* Phát hiện đối tượng văn bản mà không phụ thuộc nhiều vào font chữ và ngôn ngữ trong hình ảnh tự nhiên.
* Sử dụng phương pháp: Maximally stable extremal regions (MSER) mục đích chính của phương pháp này là phát hiện các đặc điểm nổi bật của đối tượng trong ảnh như là độ tương phản và màu sắc so với các vùng lân cận

Phương pháp tiếp cận:

* B1: MSER extraction
* B2: Local text line hypotheses
* B3: Pairwise edges in HOCC
* B4: Results for HOCC
* B4: Results for texture classification

1. Detecting Oriented Text in Natural Images by Linking Segments

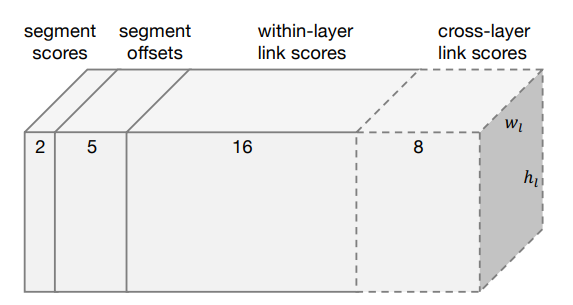
Đặc điểm nổi bật của phương pháp: phương pháp đã giải quyết tốt vấn đề phát hiện text trong hình ảnh tự nhiên, nó giải quyết tốt 2 vấn đề quan trọng mà các phương pháp trước đây chưa làm được:

* Phát hiện được các dòng chữ dài kể cả không phải các chữ cái Latin ví dụ là chữ Trung Quốc.
* Tốc độ xử lý nhanh, phù hợp với các ứng dụng realtime.

Phương pháp tiếp cận:

* Sử dụng mạng VGG16 và CNN để tách các feature

Output của bước này là tập hợp các feature có độ sâu là 31, chiều rộng và chiều cao lần lượt là w và h.



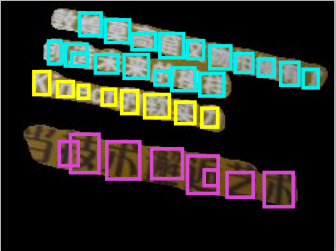
* Linking: Tạo liên kết các segment với nhau, sau khi tách được các feature sau đó kết hợp lại với nhau để dự đoán ra text location. Độ chính xác của bước này phụ thuộc nhiều vào cách lựa chọn threshhold đối với từng bộ dữ liệu khác nhau.

1. Multi-Oriented Text Detection with Fully Convolutional Networks

Ý tưởng chính của phương pháp: đầu tiền một mạng neuron(FCN) được train để dự đoán vùng text trong tổng thể của bức ảnh., sau đó là sử dụng một mạng neuron khác (FCN) để phân loại (yes or no) các kí tự trong vùng được dự đoán ở trên. Phương pháp được đánh giá là có hiệu quả khi sử dụng phát hiện với các ngôn ngữ, font chữ và các hướng văn bản khác nhau.

Các bước thực hiện:

* Text block detection: ở bước này sẽ phát hiện ra vị trí của các vùng text trong tổng thể bức ảnh. Sử dụng mô hình tùy biến trên mạng VGG-16.
* Multi-Oriented Text Line Candidate Generation: đây là bước xem xét và xác định các dòng có thể chứa text được gọi là các text block.
  + Character Components Extraction: Sử dụng MSER để trích rút ra các thành phần của kí tự (character components)



* + Orientation Estimation: ước tính góc nghiêng của dòng văn bản.
  + Text Line Candidate Generation: xác định các dòng text
* Text Line Candidates Classification: sử dụng mạng neuron có tên là Character-Centroid để phân loại, xác định chính xác các candidates ở trên có phải là dòng text hay không. Dưới đây là hình minh họa cho phương pháp Character-Centroid.



1. Multi Oriented Scene Text Detection via Corner Localization and Region Segmentation

Điểm nổi bật của phương pháp: Khác với những phương pháp trước đây, phương pháp này kết hợp giữa object detection và sematic segmentation. Phương pháp dựa trên 2 cơ sở lý thuyết:

* Một hình chữ nhật có thể xác định được bằng các đỉnh.
* Phân vùng ngữ nghĩa có hiệu quả trong việc xác định vùng chứa text.

Phương pháp tiếp cận: Mạng neuron của phương pháp là một mạng *fully convolutional* đóng vai trò: trích rút đặc trưng (feature extraction), phát hiện góc (corner detection), và xác định vị trí vùng text (position-sensitive segmentation).

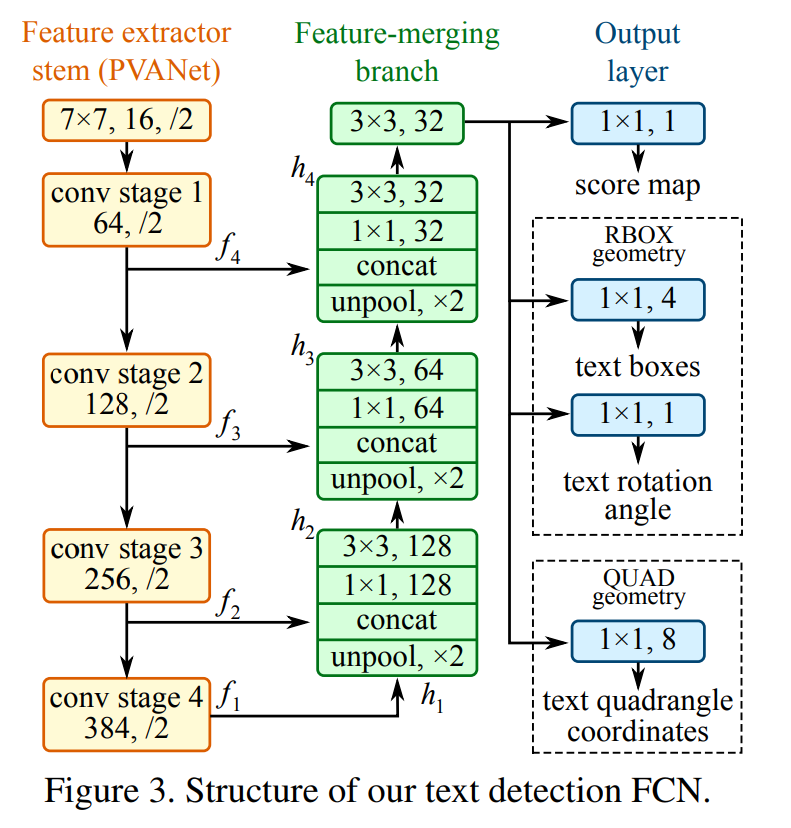
* Feature extraction: phần khung của mô hình là pre-trained của model VGG16.
* Corner detection: Dựa theo lý thuyết sử dụng SSD và DSSD phương pháp phát hiện góc bằng defaut boxes.
* Position-sensitive segmentation: Lấy ý tưởng từ phương pháp InstanceFNC, sử dụng position-sensitive segmentation để tạo ra segmentation maps.

1. EAST: An Efficient and Accurate Scene Text Detector

Điểm nổi bật:

* Mang lại tốc độ phát hiện nhanh và chính xác do dùng mạng neuron mà không qua nhiều bước như các phương pháp truyền thống khác.
* Ở phương pháp này, đưa ra một mạng deep FCN trực tiếp đưa ra kết quả cuối cùng. Phương pháp sử dụng một mạng neuron không quá phức tạp do được lược bỏ những thành phần không cần thiết đã vượt qua các phương pháp trước đó cả về độ chính xác lẫn tốc độ xử lý.

Mạng neuron:



Trên đây là sơ đồ biểu diễn các thành phần của mạng bao gồm 3 thành phần:

* Feature extractor stem: là một *convolutional network* pre-trained trên tập dữ liệu ImageNet với các tầng conv và pooling xen kẽ nhau
* Feature-merging branch: là bước merge các feature map lại với nhau (cần đọc kĩ lại để hiểu phần công thức này).
* Output layer: đây là tầng cuối cùng của mạng bao gồm: text boxes, góc nghiêng và tọa độ của text boxes.

[1] S. Xie and Z. Tu. Holistically-nested edge detection. In Proc. of ICCV, 2015.

<https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Xie_Holistically-Nested_Edge_Detection_ICCV_2015_paper.pdf>

[2] Jiri Matas, Ondrej Chum, Martin Urban, and Tomas Pa- ´ jdla, “Robust wide-baseline stereo from maximally stable extremal regions,” Image and vision computing, vol. 22, no. 10, pp. 761–767, 2004.

<http://cmp.felk.cvut.cz/~matas/papers/matas-bmvc02.pdf>