# **CHƯƠNG 1: KHẢO SÁT, ĐÁNH GIÁ VÀ LỰA CHỌN PHƯƠNG PHÁP GIẢI QUYẾT BÀI TOÁN.**

* 1. **Text detection**

Sau đây là các phương pháp tiếp cận để giải quyết bài toán text detection mà em đã tìm hiểu được, trong đó 3 phương pháp đầu tiên giải quyết bài toán theo hướng toán học cổ điển mà không sử dụng đến mô hình mạng nơ-ron. Kết quả đánh giá mỗi phương pháp được so sánh với nhau trên bộ dữ liệu MSRA-TD500

* + 1. **Detecting Texts of Arbitrary Orientations in Natural Images [1].**

Thuật toán này bao gồm 4 bước:

* Component Extraction: Sử dụng *Edge detection* để tạo ra *Edge map*  làm đầu vào cho module *SWT* để tạo ra *SWT image*. Những pixel lân cận trong *SWT image* được nhóm lại với nhau để tạo thành các components.
* Component Analysis: Nhiều *Component* được lấy ra ở bước trên không thuộc bất kì vùng chứa text nào, cho nên ở bước này sẽ sử dụng bộ phân lớp đã được *train* để định nghĩa và loại bỏ các component không phải là text.
* Candidates Linking: Các *Component* được giữ lại coi như là các C*haracter Candidates*. Bước đầu tiên của candidate linking là liên kết các character candidates thành từng cặp cho 2 *Candidates* liền kề nhau nếu như chúng giống nhau về hình dạng và màu sắc. Các cặp candidates được tổng hợp lại thành các chuỗi (chains) theo kiểu đệ quy.
* Chains Analysis: Phân tích các chuỗi để xác định chính xác có thuộc vùng kí tự hay không.

Hầu hết các phương pháp tiếp cận cho bài toán này là tập trung phát hiện vị trí biên của các vùng text . Trong bài báo này chính tôi tập trung vào việc phát hiện text trong những bức ảnh thực tế, do đó gặp rất nhiều khó khăn do có nhiều biến thể của đoạn văn bản (kích thước, font chữ, góc nghiêng,…)

Phương pháp này sử dụng SWT kết hợp với các phương pháp khác để tăng đọ chính xác khi giải quyết bài toán text detection trong hình ảnh tự nhiên. Độ chính xác của phương pháp được đánh giá khác quan qua tập dữ liệu MSRA-TD500.

* + 1. **Scene Text Detection via Holistic, Multi-Channel Prediction [2].**

Ý tưởng chính của phương pháp: Ban đầu sử dụng Edge Detection để phát hiện đường biên của các đối tượng trong ảnh, sau đó việc phát hiện kí tự, dòng chữ được phát hiện bằng cách tổng hợp các dấu hiệu đặc trưng để nhận dạng.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-measure |
| **Proposed** | 0.7651 | **0.7531** | **0.7591** |
| Zhang et al. [5] | **0.83** | 0.67 | 0.74 |
| Kang et al. [7] | 0.71 | 0.62 | 0.66 |
| Yin et al. [3] | 0.71 | 0.61 | 0.66 |
| TD-Mixture [1] | 0.63 | 0.63 | 0.60 |
| TD-ICDAR [1] | 0.53 | 0.52 | 0.50 |
| Epshtein et al. [8] | 0.25 | 0.25 | 0.25 |

* + 1. **Orientation Robust Text Line Detection in Natural Scene Images [3].**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pricision | Recall | F-measure |
| **Our method** | **0.71** | 0.62 | **0.66** |
| Yao et al. [1] | 0.63 | **0.63** | 0.60 |
| Epshtein et al. [8] | 0.25 | 0.25 | 0.25 |

* + 1. **Detecting Oriented Text in Natural Images by Linking Segments [4].**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Precision | Recall | F-measure | FPS |
| Kang et al. [7] | 0.71 | 0.62 | 0.66 | - |
| Yao et al. [1] | 0.63 | 0.63 | 0.60 | 0.14 |
| Yin et al. [9] | 0.81 | 0.63 | 0.71 | 0.71 |
| Yin et al. [3] | 0.71 | 0.61 | 0.66 | 1.25 |
| Zhang et al. [5] | 0.83 | 0.67 | 0.74 | 0.48 |
| Yao et al. [2] | 0.77 | **0.75** | 0.76 | ~1.61 |
| **SegLink** | **0.86** | 0.70 | **0.77** | **8.9** |

* + 1. **Multi-Oriented Text Detection with Fully Convolutional Networks [5].**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-Measure | FPS |
| **Proposed** | **0.83** | **0.67** | **0.74** | 0.48 |
| Yin et al. [9] | 0.81 | 0.63 | 0.71 | 0.71 |
| Kang et al. [7] | 0.71 | 0.62 | 0.66 | - |
| Yin et al. [3] | 0.71 | 0.61 | 0.66 | **1.25** |
| Yao et al. [1] | 0.63 | 0.63 | 0.60 | 0.14 |

* + 1. **Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation [6]**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Precision | Recall | F-measure | FPS |
| TD-Mixture [1] | 0.63 | 0.63 | 0.60 | - |
| TD-ICDAR [1] | 0.53 | 0.52 | 0.50 | - |
| Kang et al. [7] | 0.71 | 0.62 | 0.66 | - |
| Zhang et al. [5] | 0.83 | 0.67 | 0.74 | 0.48 |
| Yao et al. [2] | 0.77 | 0.75 | 0.76 | ~1.61 |
| EAST [10] | 0.82 | 0.62 | 0.7 | 6.52 |
| EAST+ [10] | 0.81 | 0.67 | 0.76 | **13.2** |
| SegLink [4] | 0.86 | 0.70 | 0.77 | 8.9 |
| **ours** | **0.87** | **0.76** | **0.82** | 5.7 |

*Ghi chú: + là viết tắt của mạng nơ-ron không bao gồm VGG16*

* + 1. **EAST: An Efficient and Accurate Scene Text Detector**

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Precision | Recall | F-measure |
| Ours+PVANET2x | **0.8728** | **0.6743** | **0.7608** |
| Ours + PVANET | 0.8356 | 0.6713 | 0.7445 |
| Ours + VGG16 | 0.8167 | 0.6160 | 0.7023 |

**Đánh giá tổng hợp:**

* Tất cả số liệu trên được đánh giá dựa trên bộ dữ liệu MSRA-TD500
* Phương pháp có Precision, Recall, F1 lớn nhất là phương pháp 1.6:

P = 87.6%, R = 76.2%, F1 = 81.5%

* Phương pháp nhanh nhất là EAST+:

FPS = 13.2

* Theo đánh giá cá nhân: phương pháp phù hợp nhất để giải quyết bài toán hiện tại là phương pháp: EAST+ và phương pháp SegLink do tốc độ xử lý nhanh, độ chính xác tương đối cao, và mô hình không quá phức tạp.
  1. **Text recognition**

Sau đây là các phương pháp tiếp cận đã tìm hiểu được để xử lí bài toán Text Recognition, các bài báo đều thực nghiệm qua bộ dữ liệu: CASIA-OLHWDB.

* Drawing and Recognizing Chinese Characters with Recurrent Neural Network
* Learning Spatial-Semantic Context with Fully Convolutional Recurrent Network for Online Handwritten Chinese Text Recognition
* Stroke Sequence-Dependent Deep Convolutional Neural Network for Online Handwritten Chinese Character Recognition
* High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps

# **TÀI LIỆU THAM KHẢO**

[1] <http://pages.ucsd.edu/~ztu/publication/cvpr12_textdetection.pdf>

Cong Yao, Xiang Bai, Wenyu Liu, Yi Ma, and Zhuowen Tu, “Detecting texts of arbitrary orientations in natural images,” in CVPR’12, 2012, pp. 1083–1090.

[2] <https://arxiv.org/pdf/1606.09002.pdf>

C. Yao, X. Bai, N. Sang, X. Zhou, S. Zhou, and Z. Cao. Scene text detection via holistic, multi-channel prediction. CoRR, abs/1606.09002, 2016.

[3]<https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Kang_Orientation_Robust_Text_2014_CVPR_paper.pdf>

X. C. Yin, X. Yin, K. Huang, and H. Hao. Robust text detection in natural scene images. IEEE Trans. on PAMI, 36(5):970–983, 2014.

[4] <https://arxiv.org/pdf/1703.06520.pdf>

[5] <https://arxiv.org/pdf/1604.04018.pdf>

Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai. Multioriented text detection with fully convolutional networks. In Proc. of CVPR, 2015.

[6] <https://arxiv.org/pdf/1802.08948.pdf>

[7] <https://research.ijcaonline.org/etc2016/number2/etc2016263.pdf>

L. Kang, Y. Li, and D. Doermann. Orientation robust text line detection in natural images. In Proc. of CVPR, 2014.

[8] <http://cmp.felk.cvut.cz/~cernyad2/TextCaptchaPdf/Detecting%20Text%20in%20Natural%20Scenes%20with%20Stroke%20Width%20Transform.pdf>

B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In Proc. of CVPR, 2010.

[9] <https://research.ijcaonline.org/etc2016/number2/etc2016263.pdf>

X. Yin, W. Pei, J. Zhang, and H. Hao. Multi-orientation scene text detection with adaptive clustering. PAMI, 37(9):1930–1937, 2015

[10] <https://arxiv.org/pdf/1704.03155.pdf>

X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and

J. Liang. East: An efficient and accurate scene text detector. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017

[11] <https://arxiv.org/pdf/1606.06539.pdf>

[12] <http://discovery.ucl.ac.uk/1569458/1/TPAMI-2016-08-0656-R2.pdf>

[13] <https://arxiv.org/pdf/1610.04057.pdf>

[14] <https://arxiv.org/pdf/1505.04925.pdf>