# LỜI MỞ ĐẦU

Học sâu là một nhánh lớn của ngành học máy dựa trên tập hợp các thuật toán để cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao, bằng cách sử dụng nhiều tầng xử lý phức tạp, hoặc bao gồm nhiều biến đổi phi tuyến. Trong những năm gần đây, nó đã phát triển một cách rất nhanh chóng và được ứng dụng trong nhiều lĩnh vực của công nghệ thông tin để giải quyết các vấn đề thực tế trong cuộc sống. Với mỗi vấn đề đặc trưng cho mỗi bài toán, các nhà nghiên cứu đã tìm ra các mô hình mạng nơ-ron sâu (Deep Neural Networks), mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi quy (RNN), … nhằm xử lý các bài toán trong lĩnh vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên, dịch máy, nhận dạng âm thanh, xử lý tiếng nói, sinh trắc học, …

Ngày nay những thành tựu trong lĩnh vực thị giác máy tính cũng như xử lý ngôn ngữ tự nhiên đã cho phép chúng ta có thể xử lý các vấn đề phức tạp như phát hiện và nhận dạng câu, chữ trong hình ảnh tự nhiên. Thành tựu đó là bước đầu cho việc xử lý và hỗ trợ cho con người có thể hiểu được những ngôn ngữ phức tạp khó có thể viết ra ví dụ kiểu chữ tượng hình trong một số ngôn ngữ như chữ Hán, chữ Kanji (tiếng Nhật).

Việc tò mò về nội dung của các câu chữ Hán Nôm được khắc trên các cột, cổng, … trong đền chùa, các khu di tích lịch sử đã thôi thúc tôi tìm hiểu và xây dựng ứng dụng nhận dạng và dịch nghĩa chúng. Tôi quyết định đưa việc giải quyết bài toán này làm đề tài đồ án tốt nghiệp.

# CHƯƠNG 1: ĐẶT VẤN ĐỀ

* 1. Giới thiệu bài toán

Khi vào bất cứ đền, chùa nào ta đều thấy những bức hoành phi, câu đối bằng chữ Hán, một số gia đình cũng sắm hoành phi câu đối làm đẹp và sang cho ngôi nhà của mình. Nhưng đa số khách và chủ nhà vẫn chưa hiểu hết được các bức hoành phi, câu đối ấy có ý nghĩa gì, bởi chúng đều được viết bằng chữ Hán. Hoành phi là một tấm biển gỗ có hình thứ trình bày theo chiều ngang, treo trên cao bên ngoài các gian thờ tại đình chùa, từ đường, … Nội dung của hoành phi thường thể hiện lòng tôn kính với tổ tông, bề trên, thần thánh,… Hoành phi được treo ở những nơi tôn nghiêm vừa mang tính chất nghi lễ lại vừa có giá trị nghệ thuật. Câu đối thuộc thể loại văn biền ngẫu gồm hai vế đối nhau nhằm biểu thị một ý chí, quan điểm tình cảm của tác giả trước một sự việc, một hiện tượng nào đó trong đời sống xã hội. Câu đối có nguồn gốc từ Trung Quốc và được xem là tinh hóa của chữ Hán, người Trung Quốc quan niệm: “Nếu thơ văn là tinh hoa của chữ nghĩa thì câu đối là tinh hoa của tinh hoa”. Không những thế, câu đối còn thể hiện trí thông minh, sự nhanh nhạy và phong cách tài tử trong việc đối đáp ứng xử. Câu đối tuy không dài, không nhiều chữ như những bài văn bài thơ, chỉ có hai vế nhưng vẫn thể hiện được ý tưởng, quan điểm một cách rõ ràng, cô đọng và súc tích. Mỗi câu đối được coi là một tác phẩm nghệ thuật, qui mô của tác phẩm tuy nhỏ, nhưng sức chứa, sức gợi rất lớn. Ngôn từ của câu đối được cân nhắc chọn lọc, âm điệu nhịp nhàng, kết cấu chặt chẽ, nhiều câu tài tình đến mức người đọc người nghe cảm thấy kỳ thú một cách bất ngờ khi nó bật ra ý tưởng mới lạ. Do ảnh hưởng của văn hóa Trung Quốc câu đối cũng là một trong những thể loại văn học của Việt Nam và một số nước châu Á.

Đến nay, chúng ta vẫn còn thấy những câu đối , hoành phi thường được treo ở những ngôi nhà cổ, đền, chùa, miếu, nhà thờ họ, cổng làng, đình làng. Hiện nay đa số người được giao trách nhiệm trông coi các đền, chùa, khu di tích… không biết chữ Hán, nên khi giải thích cho du khách ý nghĩa các hoành phi, câu đối đều lúng túng, hoặc giải thích không chuẩn xác. Vậy những hoành phi câu đối được thể hiện rất công phu ấy nhằm mục đích gì? Phải chăng chỉ làm sang, để thỏa mãn ý tưởng, nhận thức của những nhà xây dựng? Đôi khi những nhà hảo tâm hiến tặng cũng không nghĩ đến việc dịch câu đối đó cho người đọc dễ hiểu, nên chưa phục vụ đa số quần chúng nhân dân, chưa giúp cho người dân có thể hiểu, cảm, đồng điệu và thấm được cái hồn trong mỗi câu đối đó.

Từ những lý do đó tôi quyết định tìm hiểu và xây dựng một ứng dụng giúp mọi người có thể hiểu được nội dung ý nghĩa của các câu đối, hoành phi giúp cho mọi người thỏa mãn tính tò mò, nhu cầu tìm hiểu văn hóa của cha ông ta ngày xưa. Để thực hiện điều đó, đồ án của tôi cần được nghiên cứu giải quyết 3 bài toán chính là : Phát hiện phần nội dung văn bản trong ảnh (Text Detection ), Nhận dạng phần nội dung văn bản (Text Regcognition) và dịch sang Tiếng Việt (Translation).

* 1. Mục đích và bố cục của đồ án

Mục đích của đồ án:

* Xây dựng ứng dụng di động có tính năng hỗ trợ người dùng chụp ảnh, dịch nghĩa các câu từ tiếng Hán sang tiếng Việt.

Bố cục của đồ án được chia thành 3 phần:

* Phần 1: Tìm hiểu các cách tiếp cận, hướng giải quyết cho bài toán Text Detection, Text Regconition và Translation. Tìm hiểu các ý tưởng chính của các công trình nghiên cứu đã có, lý thuyết thuật toán, độ chính xác, ưu nhược điểm của mỗi phương pháp.
* Phần 2: Đánh giá, lựa chọn phương pháp phù hợp sau đó đưa vào xây dựng ứng dụng thử nghiệm. Đánh giá tổng quan việc sử dụng các phương pháp sau đó quyết định chọn phương pháp phù hợp với bài toán, ngữ cảnh hiện tại. Nêu cụ thể quá trình code, các bước tiến hành, khó khăn gặp phải, thành tựu đạt được trong quá trình thử nghiệm và xây dựng ứng dụng. Chia ra làm 2 phần: mô hình Deep Learning và xây dựng ứng dụng Android.
* Phần 3: Kết luận, đưa ra hướng phát triển.Nêu ra những thành quả đạt được, khó khăn gặp phải, cách giải quyết vấn đề đó, Những phần chưa làm được, định hướng phát triển ứng dụng trong tương lai.

# **CHƯƠNG 2: CÁC CÁCH TIẾP CẬN ĐỂ GIẢI QUYẾT BÀI TOÁN**

* 1. **Bài toán Text Detection**

Việc phát hiện nội dung văn bản trong hình ảnh tự nhiên là một thách thức trong lĩnh vực thị giác máy tính (computer vision). Bên cạnh đó, việc phát hiện văn bản là chữ tượng hình cũng gặp nhiều khó khăn hơn so với việc phát hiện văn bản chữ Latinh. Do đặc thù của bài toán đang cần giải quyết với chữ Hán nên tôi sử dụng bộ dữ liệu MSRA-TD500 vì các lý do: thứ nhất đây là bộ dữ liệu tiếng Hán, thứ 2 bộ dữ liệu chủ yếu là ảnh được chụp thực tế, tự nhiên nó giúp bao phủ được các vấn đề thường gặp khi phát hiện văn bản trong hình ảnh tự nhiên.

|  |  |
| --- | --- |
|  |  |
|  |  |

*Hình 1: Một số ảnh trong bộ MSRA-TD500*

Sau đây là các phương pháp tiếp cận để giải quyết bài toán text detection mà tôi đã tìm hiểu được, trong đó 3 phương pháp đầu tiên giải quyết bài toán theo hướng toán học cổ điển 4 phương pháp tiếp theo sử dụng các 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.*

Phương pháp tiếp cận:

Phương pháp này sử dụng *Stroke Width Transform* (SWT) [] kết hợp với các phương pháp khác để tăng độ chính xác khi giải quyết bài toán Text Detection trong hình ảnh tự nhiên. Hầu hết các phương pháp tiếp cận cho bài toán này là tập trung phát hiện vị trí biên của các vùng text . Phương pháp này đưa ra hướng giải quyết vấn đề phát hiện nội dung trong ảnh tự nhiên nhưng chưa giải quyết được các vấn đề thường gặp như kích thức văn bản khác nhau, font chữ, góc nghiêng, …

Các bước tiến hành:

* Bước 1: Component Extraction

Sử dụng Edge detection để tạo ra Edge Map làm đầu vào cho module SWT tạo ra SWT image. Những pixel lân cận trong SWT image được nhóm lại với nhau để tạo thành các *Components*.

* Bước 2: Component Analysis

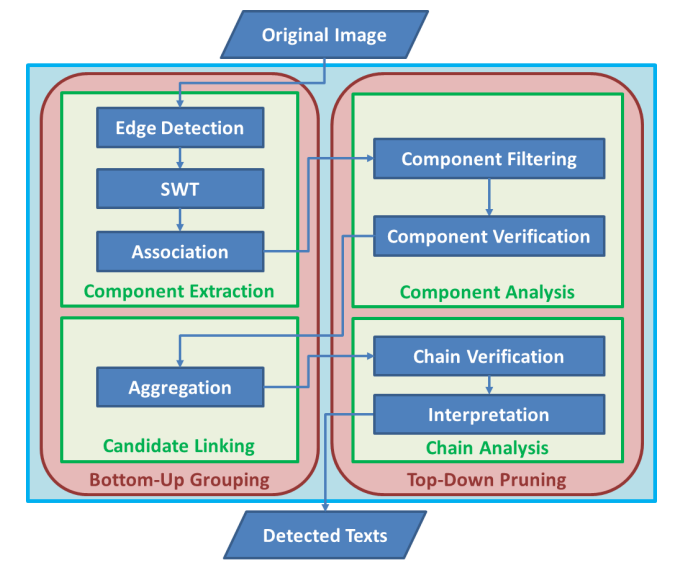
Có nhiều *Components* được lấy ra ở bước trên không thuộc bất kì vùng chứa text nào, cho nên ở bước này sẽ sử dụng bộ phân lớp đã được huấn luyệnđể định nghĩa và loại bỏ các *Components* không phải là văn bản.

* Bước 3: Candidates Linking

Các *Components* được giữ lại được coi như là các C*haracter Candidates*. Bước đầu tiên của Candidate Linking sẽ liên kết các *Character Candidates* thành từng cặp cho nếu hai *Candidates* liền kề nhau giống nhau về hình dạng và màu sắc. Các cặp *Candidates* được tổng hợp lại thành các chuỗi (*Chains*) theo kiểu đệ quy.

* Bước 4: Chains Analysis

Phân tích các chuỗi để xác định chính xác *Chains* có thuộc vùng văn bản hay không.



*Hình 2: Minh họa các bước phương pháp Detecting Texts of Arbitrary Orientations in Natural Images*

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

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 nhãn (text và non-text).
* Thuật toán được đánh giá tốt hơn đáng kể so với các thuật toán trước đây.

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 hai class là text và non-text (phương pháp được dựa trên nghiên cứu của Xie & Tu "Holistically-Nested Edge Detection" [15]), 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 tức là 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.

Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |
|  |  |  |  |

*Bảng 1: Scene Text Detection via Holistic, Multi-Channel Prediction trên bộ dữ liệu MSRA-TD500*

* + 1. *Orientation Robust Text Line Detection in Natural Scene 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:

* Bước 1: MSER extraction
* Bước 2: Local text line hypotheses
* Bước 3: Pairwise edges in HOCC
* Bước 4: Results for HOCC
* Bước 5: Results for texture classification

Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |
|  |  |  |  |

*Bảng 2: Orientation Robust Text Line Detection in Natural Scene Images trên bộ dữ liệu MSRA-TD500*

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

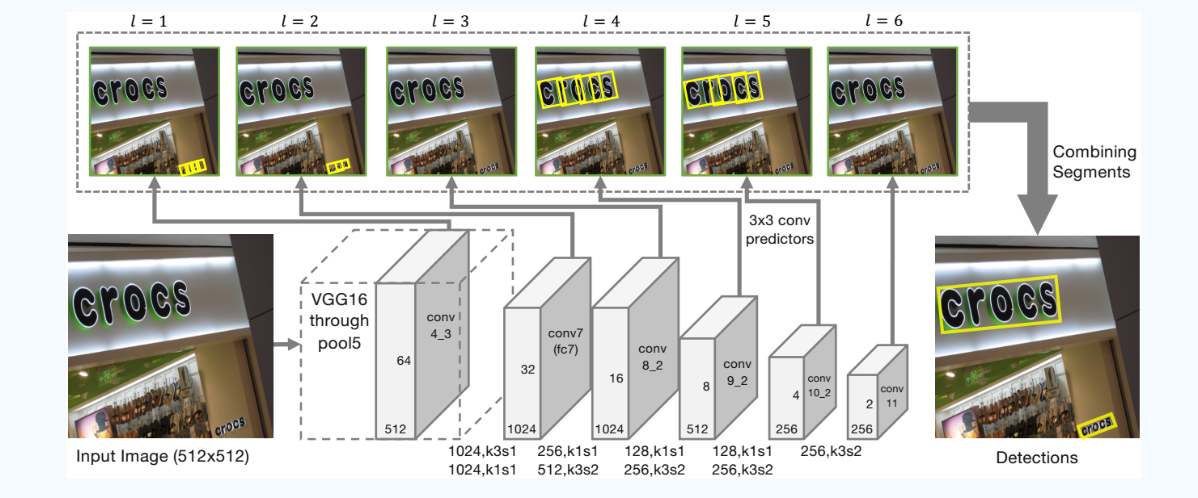
Đặc điểm nổi bật của phương pháp: phương pháp đã giải quyết tốt vấn đề phát hiện văn bản trong hình ảnh tự nhiên, nó giải quyết tốt hai vấn đề quan trọng mà các phương pháp trước đây chưa làm được đó là:

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

Khái niệm cơ bản:

* Segment: là một hộp giới hạn bao gồm một phần tử được định nghĩa bởi 5 thông số (x, y, w, h, θ).
* Link: một liên kết dùng để kết nối các segment liền kề, cho biết các segment thuộc cùng một dòng văn bản.
* Word: là một vùng bao phủ toàn bộ văn bản, có thể chứa một hoặc nhiều segment.

Các bước tiến hành:

*Hình 3: Minh họa mô hình mạng sử dụng cho Detecting Oriented Text in Natural Images by Linking Segments*

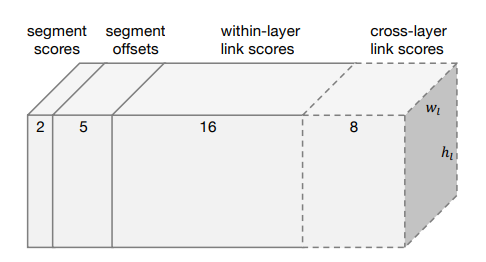
* Bước 1:

Các *Segment* (hộp màu vàng) và các link (không được vẽ ở đây) được phát hiện bởi các lớp convolution (đánh số từ 1 đến 6) mỗi convolution đại diện cho việc phát hiện segment với các kích thước khác nhau. Sau đó nó được kết hợp với nhau bằng phương pháp combining segments (ở bước 2) tạo thành các word.

Ở đây, sự khác biệt của mạng so với mạng SSD gốc là :

* + Thông tin về góc được thêm vào, thay vì 4 giá trị như mạng SSD thì ở đây có 5 giá trị (x, y, w, h, θ).
  + Chỉ sử dụng duy nhất một kích thước của defaut box.
  + Thay đổi lớp pooling cuối cùng của mạng SSD trở thành conv11 để detect links.

Output của bước 1: ứng với mỗi conv



*Hình 4:*

* Segment score có 2 giá trị là text hoặc non-text.
* Segment offset có 5 giá trị là (x, y, w, h, θ).
* 2 tầng còn lại là **within-layer link score** và **cross-layer link score** chứa giá trị thể hiện sự liên kết giữa các segment.
* Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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** |
|  |  |  |  |  |

* Segment score có 2 giá trị là text hoặc non-text.
* Segment offset có 5 giá trị là (x, y, w, h, θ).
* 2 tầng còn lại là **within-layer link score** và **cross-layer link score** chứa giá trị thể hiện sự liên kết giữa các segment.
  + 1. *Multi-Oriented Text Detection with Fully Convolutional Networks.*

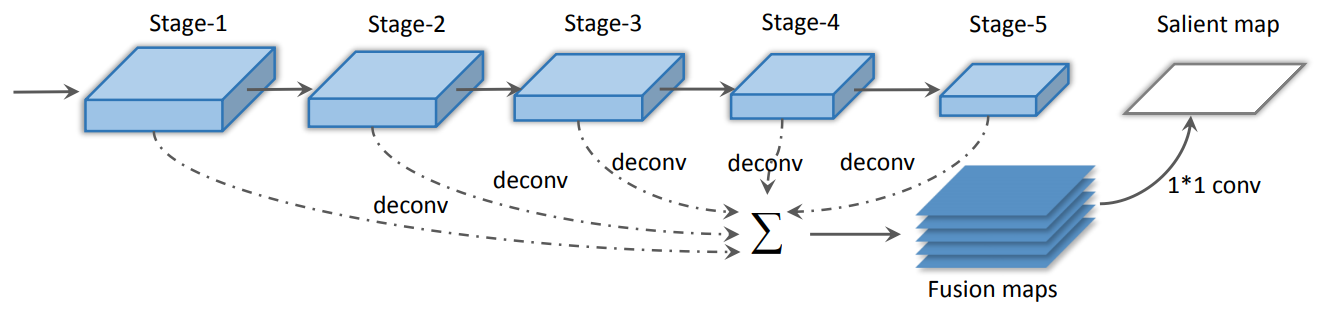
Ý tưởng chính của phương pháp:

* Đầu tiên một mạng neuron (FCN) được train để dự đoán vùng text trong tổng thể của bức ảnh. Sau đó là sử dụng phương pháp khác 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.

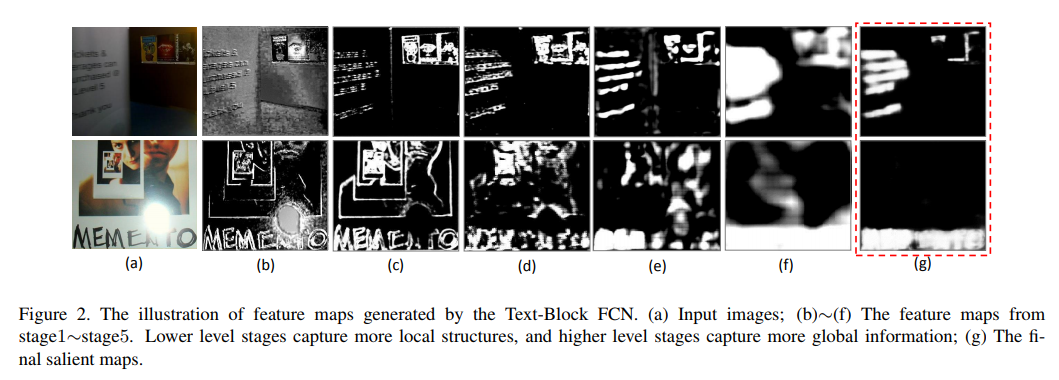
Các bước thực hiện: gồm 3 bước:

Bước 1: Text block detection:

* Text-Block FCN Network (convert the VGG 16)



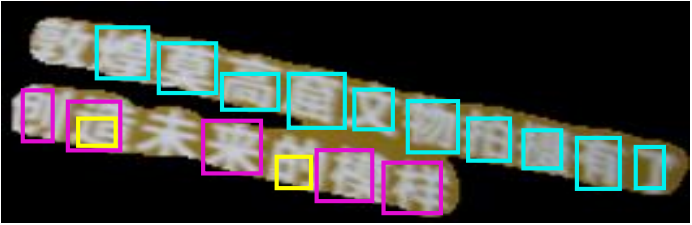
Mô tả phương pháp:



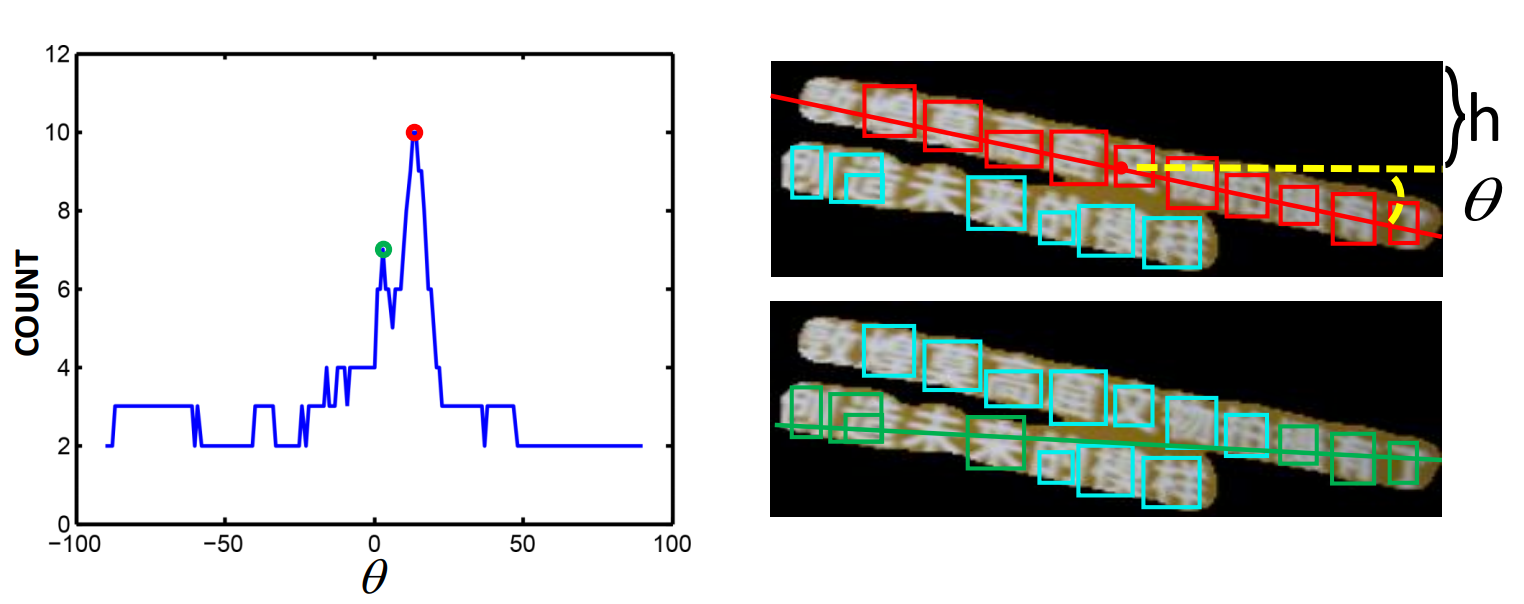
Các vùng text được phát hiện ở bước 1 được gọi là các **Text Block.**

Bước 2: Text Line Candidate Generation:

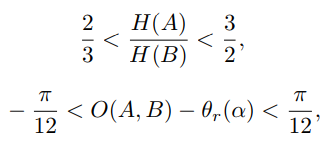
* Trích xuất kí tự: bước này sử dụng phương pháp MSER để trích xuất ra các kí tự từ **Text Box.**



* Phát hiện hướng văn bản:
  + Vẽ một đường thẳng đi qua các hộp kí tự được phát hiện ở trên
  + Hướng của văn bản được xác định bởi một đường thẳng đi qua nhiều hộp kí tự nhất.
  + Công thức xác định hướng của văn bản:
  + Mô tả phương pháp:

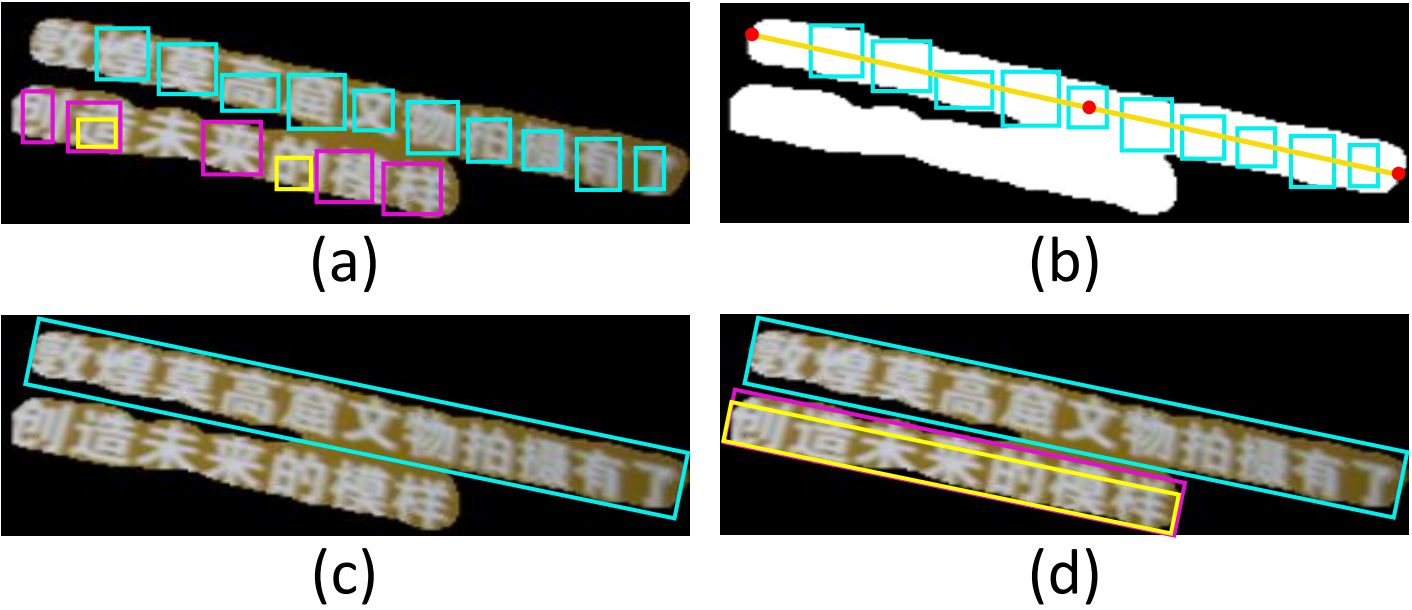


* Tạo ra **Text Line:**
  + Đầu tiên: các kí tự được nhóm lại với nhau bằng điều kiện sau:



Trong đó:

* + - H(A), H(B) là chiều cao của hộp A, B.
    - O(A, B) là góc của đường thẳng nối tâm A, B với phương ngang.
    - θr là góc của text line được ước tính ở trên.
  + Mỗi nhóm sẽ tạo ra được 1 Bounding Box (c) nó được coi là các **Text Line candidates** (d).



Bước 3: Text Line Candidates Classification:

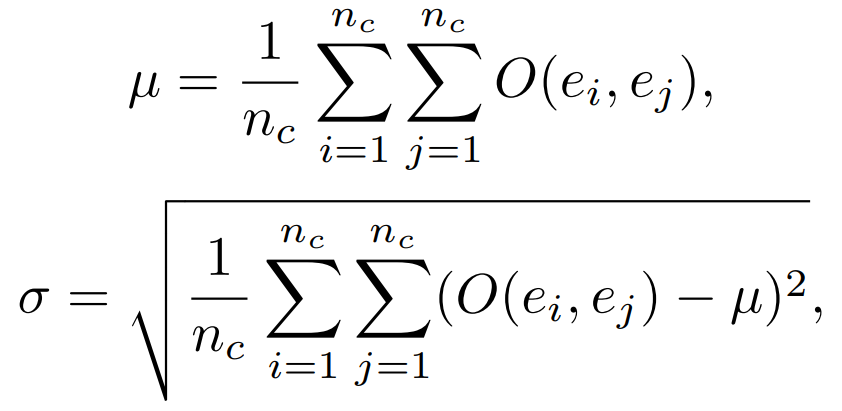
* Mục đích của bước này là loại bỏ các TextLine candidates sai.
* Để làm điều đó paper sử dụng model “Character-Centroid FCN”
* Mô tả Input-Output của model:
  + Input: **Text Line Candidates**
  + Output: The probability maps of character centroids



* Tiêu chí đánh giá output:

Với mỗi output bao gồm: ɛ = {(ei, si)}

* + Tiêu chí *Intensity Criterion:* là false Text Line Candidate nếu thoản mãn 1 trong 2 điều kiện
    - Số lương *Character Centroids < 2*
    - Giá trị trung bình xác suất < 0.6 (trong đó savg = )
  + Tiêu chí *Geometric Criterion:* sự xuất hiện của các centroids thường xấp xỉ là một đường thẳng



Trong đó:

* + - O(ei, ej) là góc giữa 2 centroids i và j.

Các TextLine candidates được cho là true nếu thỏa mãn:

Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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*

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.

Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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*

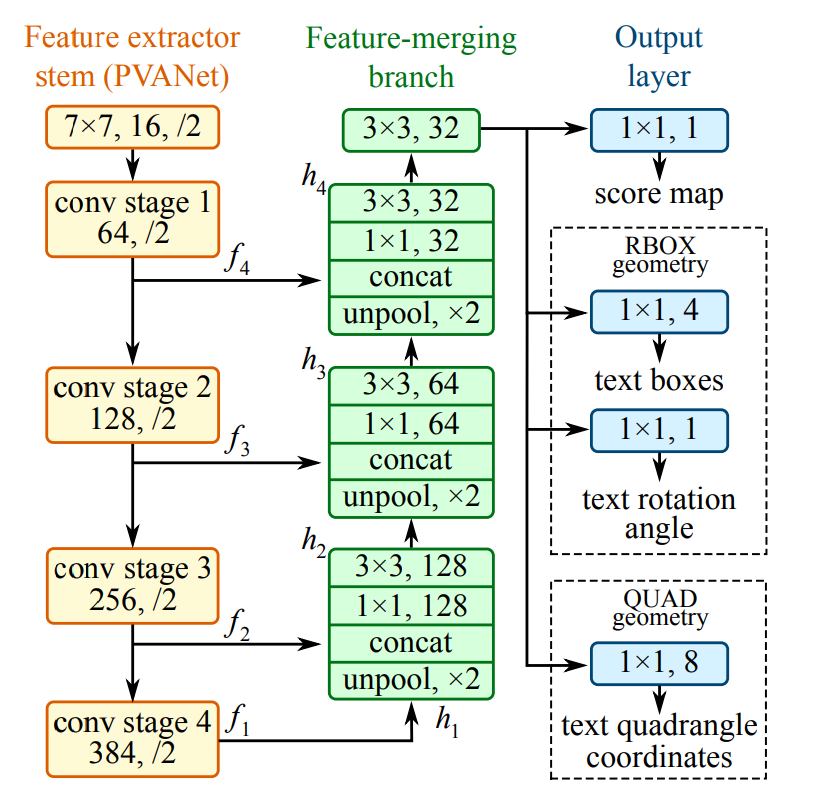
Đặc điểm nổi bật của phương pháp:

* 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:

Dưới 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.
* Output layer: đây là tầng cuối cùng của mạng bao gồm: text boxes, góc nghiêng và tọa độ của text boxes.



Bảng so sánh kết quả so với các phương pháp khác:

|  |  |  |  |
| --- | --- | --- | --- |
| 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.

**Đánh giá ưu, nhược điểm của phương pháp đã tìm hiểu:**

Các phương pháp truyền thống:

Ưu điểm:

* Không tốn thời gian xây dựng mô hình, traning,…
* Xác định vị trí của đối tượng dựa vào thuật toán và các phương pháp xử lý ảnh.

Nhược điểm:

* Tốc độ xử lý, phát hiện đối tượng chậm, không phù hợp cho ứng dụng realtime.
* Độ chính xác không cao khi áp dụng trên bộ dữ liệu là ảnh tự nhiên do background phức tạp, kích thước và font chữ không tuân theo quy tắc nhất định.

Các phương pháp mới, có xử dụng mạng neuron.

Ưu điểm:

* Độ chính xác khá cao, phát hiện linh hoạt được đối tượng trong tự nhiên nơi có background phức tạp, font chữ và kích thước tùy biến.
* Tốc độ xử lý cao, áp dụng được trong các mô hình, ứng dụng realtime.

Nhược điểm:

* Hầu hết các mô hình tương đối phức tạp. Tốn nhiều thời gian tìm hiểu, xây dựng mô hình, training.
  1. Bài toán 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
  1. Bài toán Translation

Đến nay đã có rất nhiều công trình nghiên cứu cũng như cách giải quyết bài toán dịch máy như: liệt kê ra các phương pháp ở đây

Nhưng ở đây tôi tập trung vào phương pháp Neural Machine Translation 🡪 đưa ra lý thuyết của nó ở đây

Kết Luận: Mình đã lựa chọn phương pháp nào cho bài toán nào.

# CHƯƠNG 3: QUÁ TRÌNH THỬ NGHIỆM VÀ XÂY DỰNG ỨNG DỤNG

Đưa ra sơ đồ tổng quan của hệ thống: sau đó đi sâu vào từng phần

* 1. Mô hình translation
     1. Quá trình xử lý, làm sạch dữ liệu
     2. Quá trình build model
     3. Quá trình training
     4. Quá trình testing
     5. Quá trình đánh giá mô hình
  2. Xây dựng app android

# CHƯƠNG 4: KẾT LUẬN VÀ ĐỊNH HƯỚNG PHÁT TRIỂN