# **PhotoOCR: Reading Text in Uncontrolled Conditions**

1. **Giới thiệu:** PhotoOCR là một hệ thống được sử dụng cho việc trích xuất văn bản từ ảnh.

* Cách tiếp cận được sử dụng là khả năng nhận viện văn bản trong nhiều điều kiện hình ảnh khác nhau, nơi mà các hệ thống OCR truyền thống đem lại kết quả không tốt, đáng chú ý là những hình ảnh với độ mờ đáng kể; độ phân giải, tương phản thấp; xuất hiện nhiễu ảnh,...
* Được thiết kế như một hệ thống OCR hoàn chỉnh với việc tận dụng những kĩ thuật computer vision hiện đại, những tiến bộ đáng kể trong thị trường học máy và những mô hình ngôn ngữ lớn (language modelling).
* Mô hình mạng neural sử dụng cho bài toán character classifier được huấn luyện trên 2 triệu labels, cùng với đó hơn 1 nghìn tỷ tokens cũng được sử dụng cho language modelling,...

Với việc sử dụng tập dữ liệu chất lượng, cũng như xây dựng được một hệ thống hoàn chỉnh thì PhotoOCR thể hiện sự vượt trội hơn hẳn so với các cách tiếp cận trước đó, giảm được phân nửa tỷ lệ lỗi, được sử dụng ở nhiều ứng dụng tại GOOGLE.

1. **Thiết kế hệ thống:** tương tự như các hệ thống OCR khác với mô hình pipeline bao gồm nhiều modules được thể hiện như trình tự dưới đây:



* Bắt đầu bằng việc phát hiện văn bản trên hình ảnh đầu vào.Trả về các candidate region chứa các dòng văn bản riêng lẻ. Những kết quả trả về đó sẽ tiếp tục được xử lý bởi module kế tiếp trong pipeline.



* Chia các thành phần dòng văn bản được phát hiện ở bước trên thành các phân đoạn.



* Phân loại kí tự (character classifier) trên từng vị trí đã được phân đoạn tại module Character segmentation



1. **Các phương pháp được sử dụng**

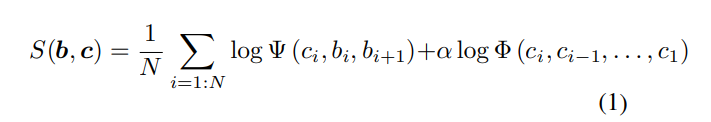
**3.1. Text detection**

* Kêt hợp đầu ra của 3 cách tiếp cận khác nhau về nhận diện văn bản
* **Object detection using a boosted cascade of simple features** : http://www.egr.unlv.edu/~b1morris/ecg782/slides/viola\_jones\_face\_detection.pdf
* **Text localization in natural scene images based on conditional random field:** https://daneshyari.com/article/preview/6863992.pdf
* **Anisotropic Gaussian filters**
* Chia các candidate region thành các dòng riêng lẻ và điều chỉnh hướng theo chiều ngang để thuận lợi cho việc xử lý ở module tiếp theo.

**3.2. Segmentation**

* Kết hợp đầu ra của hai phương pháp phân đoạn khác nhau để cải thiện recall (tỷ lệ giữa các điểm positive thực được nhận đúng trên tổng điểm positive thực)
* **Niblack binarization:** Cách tiếp cận này đêm lại hiệu quả cao đối với những hình ảnh không phức tạp, nhưng có thể thất bại với những hình ảnh mờ, độ phân giải thấp, background phức tạp,...
* **Binary sliding window classifier:** phát hiện các điểm phân đoạn giữa các kí tự + **Thuật toán HOG(Histogram of oriented gradient)**

* Giai đoạn phân đoạn xuất ra một vectơ B chứa các vị trí của các điểm phân đoạn được phát hiện, bao gồm các điểm bắt đầu và điểm cuối của text detection box
* **Beam Search:** maximize score function ở dưới

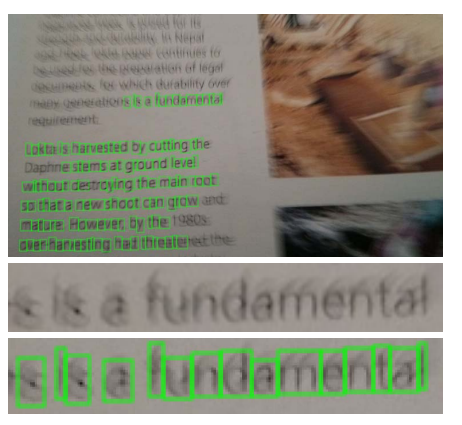


**3.3. Character Classifier**

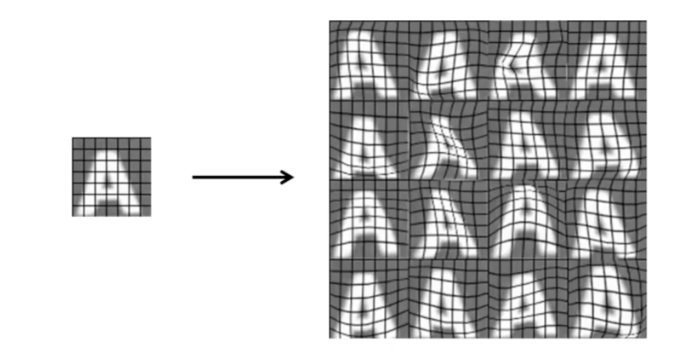
* **Languge Model:** standard ngram + 4-grams approach for language modelling. (mô hình đại diện cho những kiến thức đã biết về một ngôn ngữ, những kiến thức ấy có thể là những từ, chuỗi các từ có thể có hay mức độ thường xuyên mà chúng xuất hiện.)

**3.4 Training:**

* **Self-supervised training data:** xây dựng bộ dữ liệu huấn luyện lớn hơn



Tạo thêm dữ liệu mới bằng cách phóng to vùng ảnh từ dữ liệu có sẵn

****

Làm biến dạng dữ liệu có sẵn

1. **Source code:**