Catalyst

Multilingual Text detection in Video/Images

Lokesh Nandanwar (lokeshkvn)
lokeshnandanwar150@gmail.com

Data Fest

Catalyst

My background

Lokesh Nandanwar

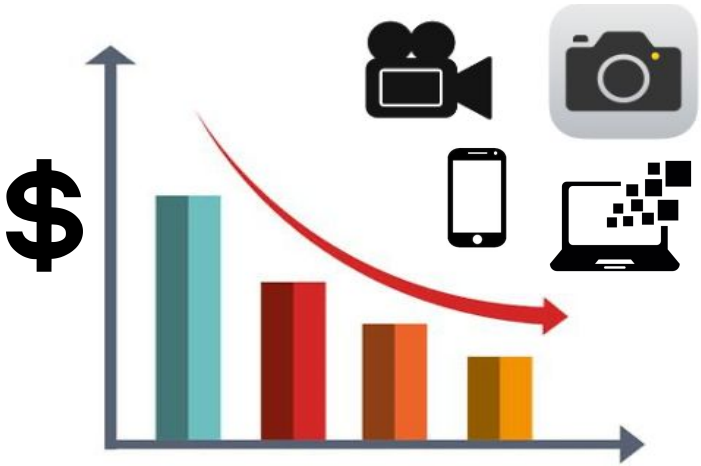
LinkedIn: lokeshkvn
Github: lokeshkvn

- Current :
  - M.Sc. Student (Computer Vision and Image Processing)
  - Graduate Research Assistant, University of Malaya, Malaysia
- Belt and Road Award 2019, China Students Service Outsourcing Innovation and Entrepreneurship Competition
- Winner - Smart India Hackathon 2019
- Student Developer, Google Summer of Code 2018
- 1 Journal Publication in **Expert Systems with Applications 2020.**
- 6 Conference Publications (2 in ICPR 2020, 2 in ICPRAI 2020, 1 in DAS 2020, 1 in ICACCP 2019).

Data Fest

Catalyst

Introduction



- Everyday new multimedia tools and devices are released with **low prices**.
- Digital data storage and access facility is **cheaper and freely* available.**

- Increase in **digital lifestyle** of people
- Increase in Social Media, News broadcast, Internet and digital content **usage**.

Need to automate Text Detection/Recognition for many useful purposes!

Data Fest

Catalyst

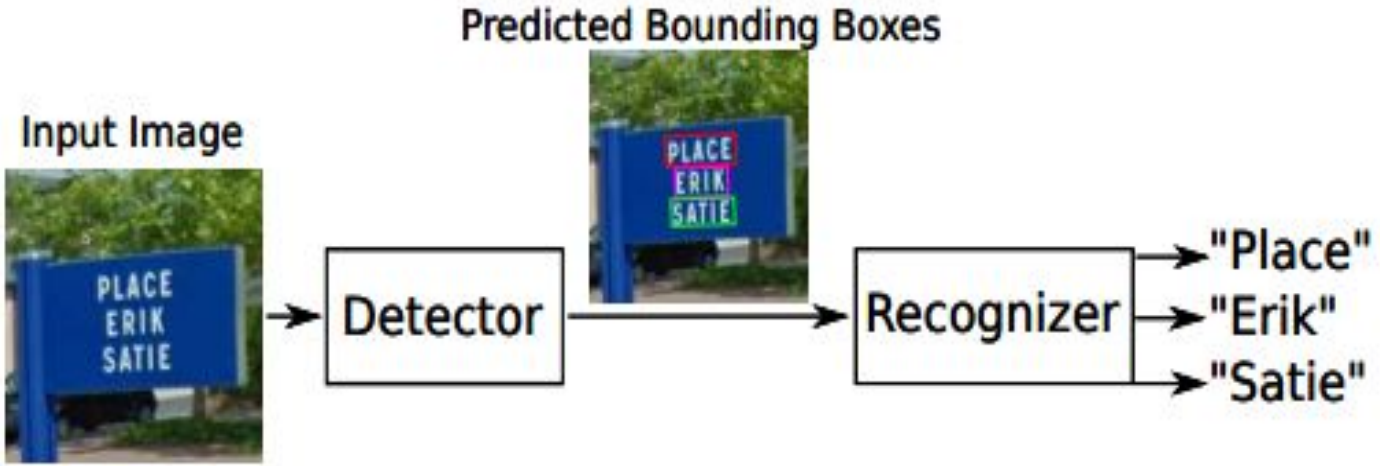
Motivation

- Widely studied in **pattern recognition fields also known as automatic text detection and recognition (OCR)**
- OCR has started from isolated character recognition and evolved to printed/ handwriting document recognition.
- Recently, embedded texts in videos and natural scenes have received increasing attention due to crucial information about the media content.
- Extracting text from content is a **non-trivial task** due to many challenges.

Data Fest

Catalyst

Complete flow



Data Fest

Catalyst

Why Text Detection/Recognition?

- Annotating the images and video through **Captions**.
  - Better indexing and retrieval at semantic level.
- Recognizing **signs** in driver assisted systems.
- Providing **scene information** to visually impaired people.
- **Events extraction** from sports, news broadcast, etc.
- **Tracing and watching** the persons.
  - Marathons, Exhibitions, processions, etc.

Data Fest

Catalyst

Our Focus: Text Detection

- Process of **detecting** the text present in the image, followed by surrounding it with a rectangular bounding box.
- The image is **segmented** into multiple segments of texts.
- Each segment is a **connected component of pixels** with similar characteristics (Characters).

Data Fest

Catalyst

Objectives

- So far, the methods have **focused only on some languages** such as English, Latin and Chinese.
- For a language like **Hindi, Russian, Arabic, etc** which is also used by more than one billion people around the world, the literature is limited to very few studies.
- This presentation aims to tackle the challenges in **Multilingual Text detection** in Video/Images with the help of Catalyst framework.

Data Fest

Catalyst

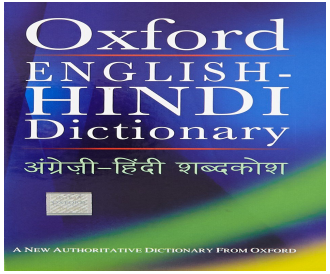
Challenges

Uneven Lighting

Scene Complexity

Uneven Color

Blurring/degradation



Multilingual

Arbitrary Shaped text

Data Fest

Catalyst

Existing Methods and Drawbacks

Some Recent State of the Art Methods

1) **FOTS (Fast oriented text spotting)** : Weak backbone (Resnet) and segmentation head [CVPR 2018] *(https://arxiv.org/abs/1801.01671)*
2) **PSENet(Progressive Scale Expansion Network)** : Weak Backbone and decoder [CVPR 2019] *(https://arxiv.org/abs/1806.02559)*
3) **CRAFT(Character Region Awareness for Text detection)** : Weak feature extractor (VGG-Unet) [CVPR 2019] *(https://arxiv.org/abs/1904.01941)*
4) **DB-Net (Differential Binarization Network)** : Weak backbone (Resnet) [AAAI 2020] *(https://arxiv.org/abs/1911.08947)*

Data Fest

Catalyst

Existing Methods

PSENet

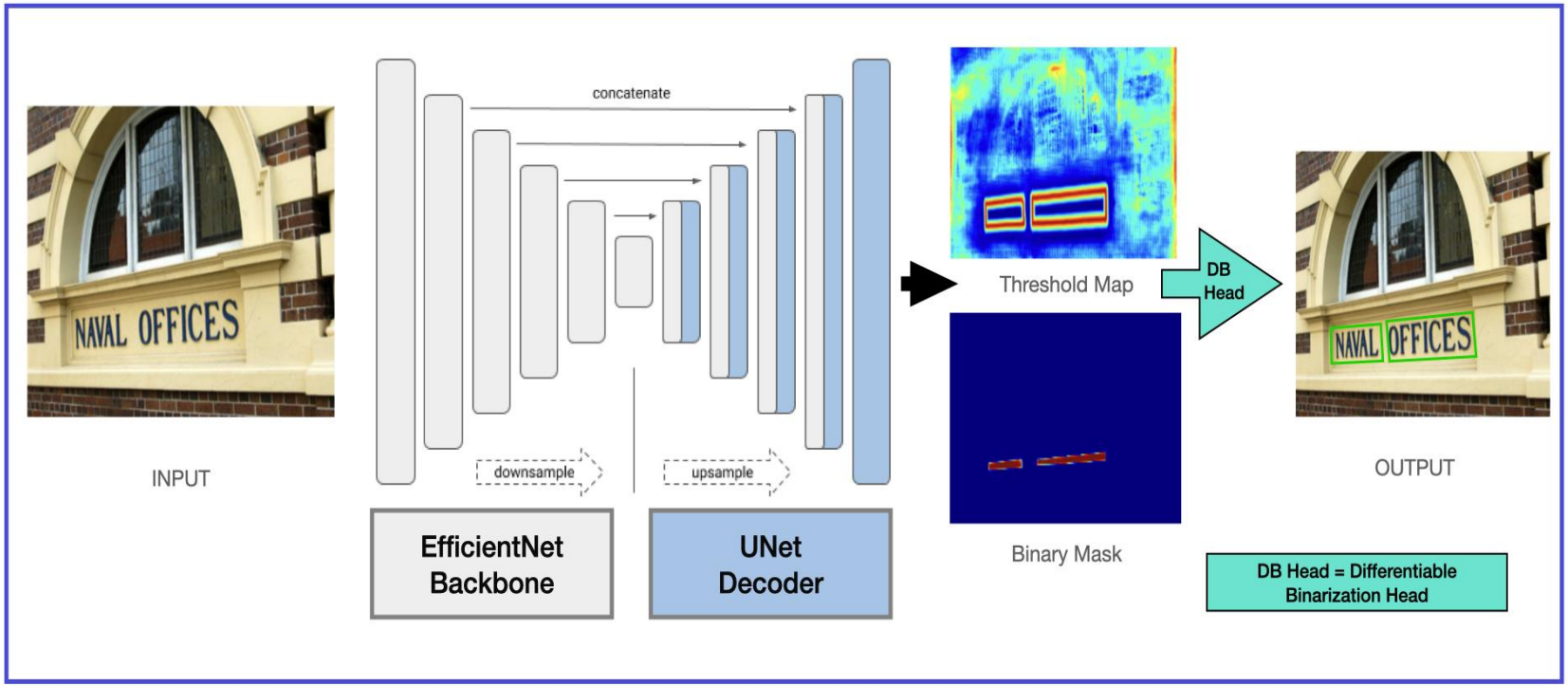
FOTS

DB-Net

Data Fest

Catalyst

Proposed Method (Eff-DB)



Data Fest

Catalyst

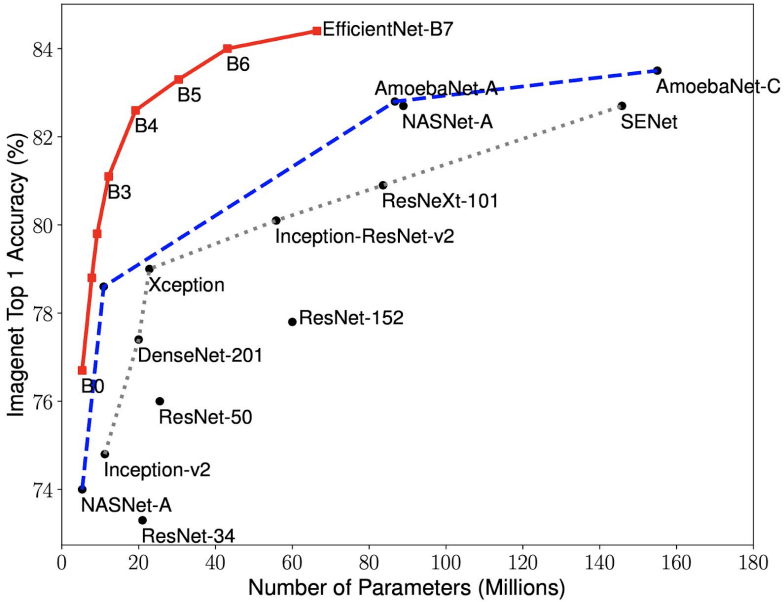
Final Result

Data Fest

Catalyst

Why EfficientNet as Backbone?

- Designed through **Neural Architecture Search (NAS)**
- Based on Model scaling method using **effective compound coefficient.**
- Uses low Parameters and High Accuracy, implies **less training time** and improve in real time performance.
- Superpasses state-of-the-art accuracy with up to **10x better efficiency (smaller and faster).**



Reference: https://arxiv.org/abs/1905.11946

Data Fest

Catalyst

Why UNet as Decoder?

- Initially designed for **medical image** segmentation, then adapted for **fast and precise segmentation** in most of the Computer Vision Tasks
- Most **Reliable and cost-effective** decoder for segmentation tasks.
- **Performs better or nearly same** than even the most recent architectures such as DeepLab, FPN, FCN, PSPNet, etc
- Needs **less parameters**, effectively reducing training and inference time.

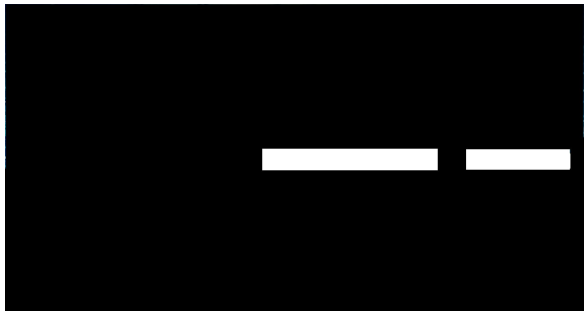
Reference: https://arxiv.org/abs/1505.04597

Data Fest

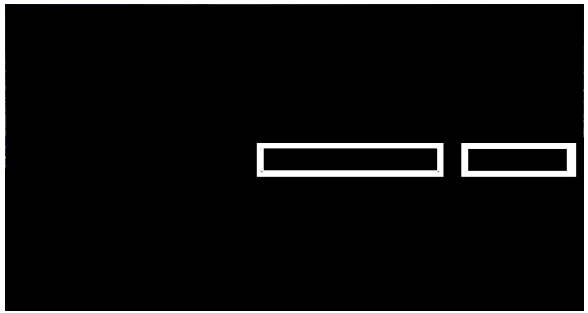
Catalyst

Binary Mask with Threshold Map!!!

Input Image



Binary Mask



Threshold(Border) Map

- Optimal strategy for Segmentation of words in case of **text lines/ crowded text (Superimposed segmentation in general).**
- Precise and accurate word wise segmentation helps in **effective recognition.**

Data Fest

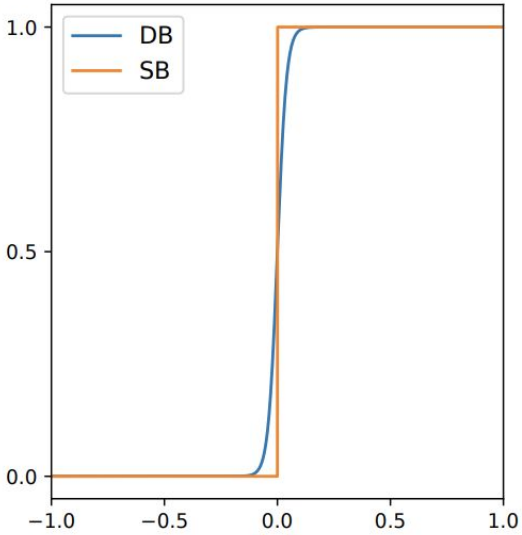
Catalyst

DB-Head

- **Differential Binarization (DB)** instead of Standard Binarization (SB).
- The major effect of DB-Head is differentiability, which makes the process of **binarization end-to-end trainable** in a CNN.

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k(P_{i,j} - T_{i,j})}}$$

- The differentiable binarization with adaptive thresholds help to differentiate text regions from the background and also to **separate text instances which are closely jointed.**



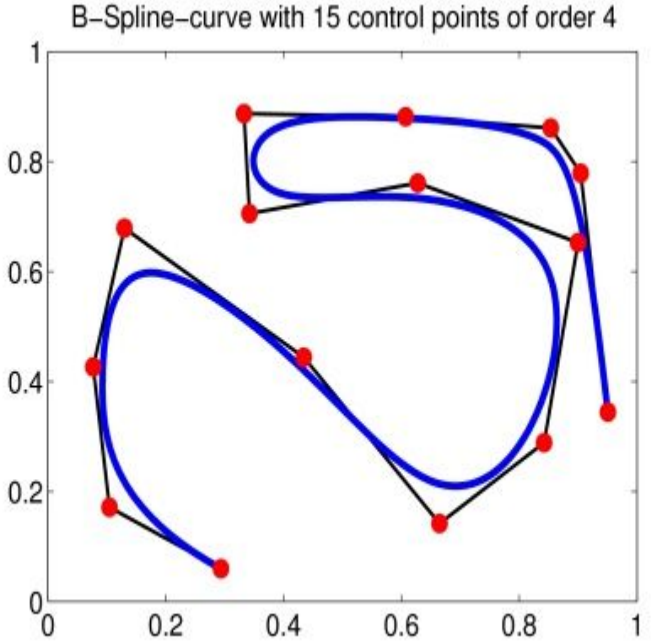
Reference: https://arxiv.org/abs/1911.08947

Data Fest

Catalyst

B-Spline Curve Fitting

- We explore B-Spline polygon curve fitting to fix **the accurate and tight bounding boxes for arbitrary oriented text lines**,
- It fixes **smooth and accurate bounding box** for arbitrary oriented text lines.

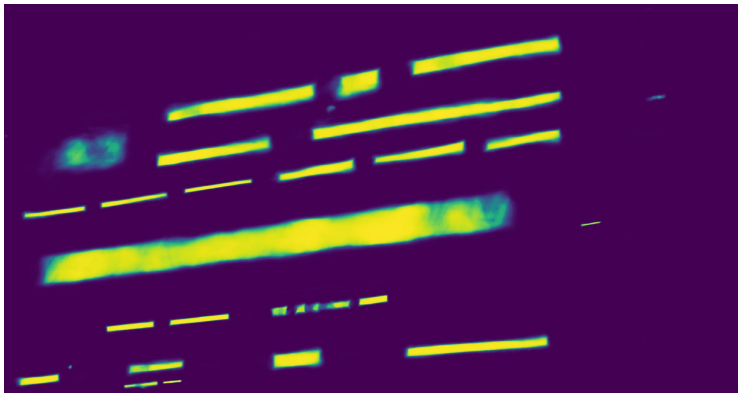


Reference: https://en.wikipedia.org/wiki/B-spline

Data Fest

Catalyst

Some more results



Data Fest

Catalyst

Frameworks and Codes

- DL Framework: **Pytorch using Catalyst**
- Encoder and Decoder: **Segmentation Models Pytorch**
- Segmentation Head (DB-Head): **Real-time Scene Text Detection with Differentiable Binarization**
- Datasets: **ICDAR MLT 2019**, **ArT 2019**

PYTORCH

Segmentation Models

ICDAR 2019

15th International Conference on Document Analysis and Recognition

20-25 September 2019 | International Convention Centre Sydney, Australia

Data Fest

Catalyst

ICPR 2020: AcTiV Competition

- **Competition on Superimposed Text Detection and Recognition in Arabic News Video Frames** (hold within the framework of the 25th **International Conference on Pattern Recognition** (ICPR2020), Milano- Italy 13 -18 September 2020)
- Results on Complete Public test dataset from competition (*Rankings Awaited):

Precision	99.4378 (513 images)
Recall	91.8398 (513 images)
F-measure	95.4879

Reference: https://diuf.unifr.ch/main/diva/AcTiVComp/

Catalyst

Summary

- Multilingual Text detection is a **non-trivial task** and is the need of the today's digital world.
- Choose **encoder, decoder and head** wisely while building architecture in any computer vision task.
- Always **validate the methodology** with real world use cases such as competitions, Hackathons, etc.
- ***Catalyst is all you need*** to work effectively and productively in any ML/DL projects.

Catalyst

Thank You!

Questions???