Hochschule für Technik
und Wirtschaft Berlin
University of Applied Sciences

Text Detection in Natural Scene Images using CRAFT and EAST

Henry Febrian

Module: *Computergrafik und Bildverarbeitung*

Lecturer: Erik Rodner

September XX, 2021

Abstract

To be edited at a later date.

Contents

1 Introduction

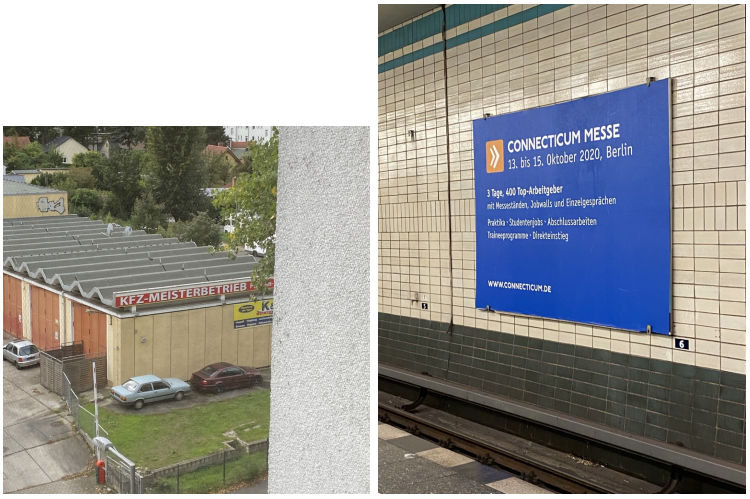
Text is arguably one of the most essential form of communication. According to Long et al. (2020):

> *"As the written form of human languages, text makes it feasible to reliably and effectively spread or acquire information across time and space. In this sense, text constitutes the cornerstone of human civilization."* (Long et al. 2020)

In the modern world, text as a medium of communication is not only consumed by humans but has claimed its place in the world of technology. However, text detection in natural scene images is proven to be challenging. Compared to detecting text on handwritten materials, the randomness of a natural scene is a big hurdle to overcome. This paper aims to test and evaluate the two proposed methods on their performance in detecting natural-scene texts. It begins by observing the interference found on natural scenes followed by introducing and giving an overview of CRAFT and EAST. It then presents an overview on the how text is detected using the two methods, along with an explanation of the evaluation dataset. Subsequently, the performance review of the algorithm will be presented, which is obtained by testing the preferred method on the aforementioned dataset.

2 Challenges of Natural Text Detection

Natural scene images could be classified as images which are taken in uncontrolled environments, with any device ranging from smartphones to professional cameras. These images are snapshots of things in the real world.



(a) A scenery (b) A billboard

Figure 1: Samples of natural scene images.

The random nature of the real world, combined with the diversity of available devices introduced some factors which make natural text detection a greater

challenge than detecting structured text in documents. Mancas-Thillou & Gosselin (2007) mentioned some conditions that are found in natural scene which may significiantly impact text detection procedure. They are:

- Raw sensor image and sensor noise
- Viewing angle
- Blur
- Lighting
- Resolution and Aliasing

Although devices have evolved to a point where we do not have to worry too much about sensor noise and resolution in a casual setting, the other factors remain the same. Blur, lighting, viewing angle, and other factors not mentioned in the cited paper, such as weather, are still going to provide a challenge to text detection.

3 Overview of the Methods

3.1 Overview of EAST

According to Zhou et al. (2017), the key component of EAST is a neural network model, which is trained to directly predict the existence of text instances and their geometries from full images (Zhou et al. 2017). Hence, the abbreviation EAST: Efficient and Accurate Scene Text Detector.

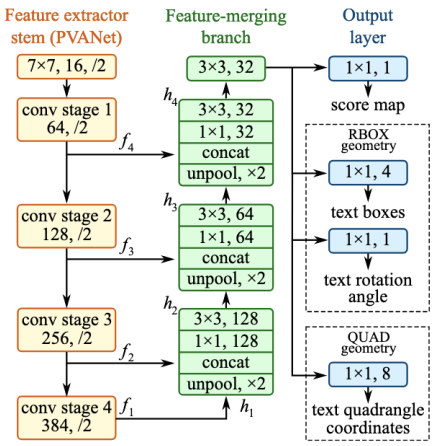


Figure 2: Structure of EAST, adopted from Zhou et al. (2017)

3.2 Overview of CRAFT

CRAFT stands for Character Region Awareness for Text Detection. According to Baek et al. (2019), CRAFT is a novel text detector which localizes the individual character regions and links the detected characters to a text instance (Baek et al. 2019).

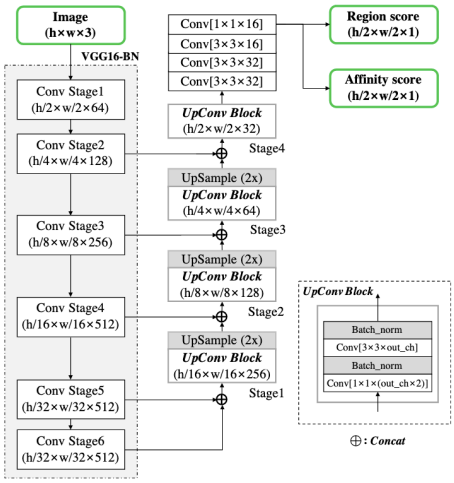


Figure 3: Structure of Craft, adopted from Baek et al. (2019)

4 Methodology

asddssafs

5 Results and Evaluation

asdasdssada

6 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

7 Section about quotations

In this section, an example for a literal quotation is given.

> *"A persona is a rich picture of an imaginary person who represents your core user group."* (Baek et al. 2019)

Sometimes you might want to make use of the authors name within the text. Before, we used the command `citep{}`, which creates the brackets around author name and year. You can also use the `cite` command like this:

Dix et al. (2004) defined the concept of persona as follows:

> *"A persona is a rich picture of an imaginary person who represents your core user group."* (Zhou et al. 2017)

You may notice, that this increases the readability of the text.

According to APA format[1] there are some rules, when and how to include page numbers, when referring to literature.

> *"Include page numbers for any citations in the text of your paper that include direct quotations or refer to a specific part of the work you are referencing. Direct quotations must include a page number as part of the citation. The quoted material should be followed by a citation in parentheses that gives the author's name, the year in which the work was published, and the page number from which the quoted material appears."* (Hall 2013)

Check out the example and recommendations of Hall (2013) on `http://www.ehow.com/how_5689799_cite-numbers-apa-format.html`. In LaTeXyou can include the pages very easy. For example:

Baddeley & Hitch (1974, p. 86) stated:

> *"We hope that our preliminary attempts to begin answering the question will convince the reader, not necessarily that our views are correct, but that the question was and is well worth asking"* (Baddeley & Hitch 1974, p. 86)

Note that in the first reference, we used `citet[]{}` in order to have brackets just around year and page number; later we used `citep[]{}`.

8 Section about references within the document

If you want to refer to you own chapters, figures, tables or the like, you can make use of the `ref{}` command, for example:

- section 7 on page 5

8.1 Subsection within Foundations

Lorem ipsum dolor sit amet, consetetur Mancas-Thillou & Gosselin (2007) sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

[1] American Psychological Association (APA)

8.2 Another subsection within Foundations

Lorem ipsum dolor sit amet, consetetur sadipscing Rosebrock (2018) elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

References

Baddeley & Hitch 1974
BADDELEY, Alan D.; HITCH, Graham: Working memory. In: *Psychology of learning and motivation* (1974), pages 47–89

Baek et al. 2019
BAEK, Youngmin; LEE, Bado; HAN, Dongyoon; YUN, Sangdoo; LEE, Hwalsuk: *Character Region Awareness for Text Detection*. https://arxiv.org/pdf/1904.01941.pdf. Version: 2019. – Last accessed 14 September 2021

Dix et al. 2004
DIX, Alan; FINLAY, Janet E.; ABOWD, Gregory D.; BEALE, Russell: *Human-Computer Interaction*. 3. Pearson Education Limited, Essex, England, 2004. – ISBN 0-13-046109-1

Hall 2013
HALL, Shane: *How to Cite Page Numbers in APA Format*. http://www.ehow.com/how_5689799_cite-numbers-apa-format.html. Version: 2013. – Last accessed 16 July 2013

Long et al. 2020
LONG, Shangbang; HE, Xin; YAO, Cong: *Scene Text Detection and Recognition: The Deep Learning Era*. https://arxiv.org/pdf/1811.04256.pdf. Version: 2020. – Last accessed 7 September 2021

Mancas-Thillou & Gosselin 2007
MANCAS-THILLOU, Céline; GOSSELIN, Bernard: *Natural Scene Text Understanding*. https://pdfs.semanticscholar.org/c6de/0b0485e4ebc6df300669969ed743b86b511e.pdf?_ga=2.82453078.1125900540.1632067547-1687845596.1632067547. Version: 2007. – Last accessed 20 September 2021

Rosebrock 2018
ROSEBROCK, Adrian: *OpenCV Text Detection (EAST Text Detector)*. https://www.pyimagesearch.com/2018/08/20/opencv-text-detection-east-text-detector/. Version: 2018. – Last updated July 7, 2021; Last accessed 21 September 2021

Zhou et al. 2017
ZHOU, Xinyu; YAO, Cong; WEN, He; WANG, Yuzhi; ZHOU, Shuchang; HE, Weiran; LIANG, Jiajun: *EAST: An Efficient and Accurate Scene Text Detector*. https://arxiv.org/pdf/1704.03155.pdf. Version: 2017. – Last accessed 14 September 2021