
OPTICAL CHARACTER RECOGNITION AND TRANSLATION FOR HINDI INDIC TEXT

2018115049 -- Kevin Paul

2018115109 – Srijevarankesh

2018115134 – Vishwa Kumar

CRAFT ALGORITHM

What is CRAFT?

The framework, referred to as **CRAFT** for **Character Region Awareness For Text** detection, is designed with a convolutional neural network producing the character region score and affinity score. The region score is used to localize individual characters in the image, and the affinity score is used to group each character into a single instance. To compensate for the lack of character-level annotations, we propose a weakly supervised learning framework that estimates character level ground truths in existing real word-level datasets.

Why do we plan to use CRAFT?

We have seen Abstract Scene text detection methods based on neural networks emerging recently and these methods have shown promising results. Previous methods trained with rigid word-level bounding boxes exhibit limitations in representing the text region in an arbitrary shape. In this Creative and Innovative Project, we plan to use a scene text detection method to effectively detect text area by exploring each character and affinity between characters. To overcome the lack of individual character level annotations, the proposed framework exploits both the given character level annotations for images and the estimated character-level ground-truths for real images acquired by the learned interim model. In order

to estimate affinity between characters, the network is trained with the newly proposed representation for affinity.

Extensive experiments on lot of images, including the datasets which contain highly curved texts in natural images, demonstrate that our character-level text detection significantly outperforms the state-of-the-art detectors. According to the results, our proposed method guarantees high flexibility in detecting complicated scene text images, such as weirdly-oriented, curved, or deformed texts.

How does CRAFT work?

A fully convolutional network architecture with batch normalization is adopted as the backbone. Our model has skip connections in the decoding part, which is similar to U-net in that it aggregates low level features. The final output has two channels as score maps: the region scores and the affinity scores.

This dataset was trained with different learning mechanisms as

- Ground Truth Label Generation
- Weakly-Supervised Learning



This algorithm was trained with extensive datasets extending to multiple languages such as Quadrilateral-type datasets (ICDARs, and MSRATD500) and Polygon-type datasets (TotalText, CTW-1500) to help develop an accuracy model convenient to suit our technical aspects.

The proposed method provides the character region score and the character affinity score that, together, fully cover various text shapes in a bottom-up manner. Since real datasets provided with character-level annotations are rare, we proposed a weakly supervised learning method that generates pseudo-ground truths from an interim model. CRAFT shows state-of-the-art performances on most public datasets and demonstrates generalization ability by showing these performances without fine-tuning.

Where did we refer on the technicalities and workings of the algorithm?

We have referred the paper on the CRAFT algorithm by the engineers who built it and wrote a research paper termed as :

Character Region Awareness for Text Detection

Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee_

From Clova AI Research, NAVER Corp.