

Text detection in screen images with a Convolutional Neural Network

Dominik Moritz¹

¹ University of Washington

DOI: [10.21105/joss.00235](https://doi.org/10.21105/joss.00235)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Licence

Authors of JOSS papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

The repository contains a set of scripts to implement text detection from screen images. The idea is that we use a Convolutional Neural Network (CNN) (Le Cun et al. 1990) to predict a heatmap of the probability of text in an image. The network outputs a heatmap for text with 64×64 pixels and is implemented in Darknet (Redmon 2013–2016). To train the network, we use a set of pairs of images and training labels. We obtain the training data by extracting figures with embedded text from research papers in PDF form and generated pixel masks from them.

With the code, we also provide a dataset of around 500K labeled images extracted from 1M papers from arXiv and the ACL anthology.

References

Le Cun, B Boser, John S Denker, D Henderson, Richard E Howard, W Hubbard, and Lawrence D Jackel. 1990. “Handwritten Digit Recognition with a Back-Propagation Network.” In *Advances in Neural Information Processing Systems*. Citeseer.

Redmon, Joseph. 2013–2016. “Darknet: Open Source Neural Networks in c.” <http://pjreddie.com/darknet/>.