

FACULTY **OF ELECTRICAL** ENGINEERING CTU IN PRAGUE



https://github.com/klarajanouskova/textdetection-recognition-PGT

Summary

- A novel method for automatic generation of pseudo ground truth data (less than 4.1 % error rate) from images with weak annotations
- Exploiting existing text metadata as weak annotations, ie. product databases
- The method works with an arbitrary detector/recognizer
- The accuracy of the best-performing open-source recognition model is improved by 3.7 % on average on 7 benchmarks
- Recognition performance in weakly-supervised domain adaptation improved by 24.5 %
- a single model surpasses/matches the performance of different second best methods on multiple dataset
- Pseudo ground truth for the Amazon Book Covers dataset publicly available



Images with text metadata are common - our method exploits images accompanied by a list of texts that have a high probability of appearing in the image. Such texts are called weak annotations.



Weak annotation: The Sherlock Holmes Museum, 221B Baker Street

> Words in weak annotation

Nords missing

An image from London, 221B Baker Street, in front of the Sherlock Holmes Museum. Weak annotation is obtained ie. from google nearby search.

Motivation

Text Recognition - Real World Data and Where to Find Them

Klára Janoušková, Jiří Matas Visual Recognition Group, Dept. of Cybernetics FEE, Czech Technical University in Prague Czech Republic

Method

The **core idea** is to associate the output of the detector and the recognizer with weak annotations. The resulting set of text detections with weak label transcriptions is referred to as pseudo-ground truth (PGT). The PGT is added to the training set to finetune the recognizer, whose output in turn updates the PGT.



PGT - pseudo ground truth DET - detector OCR - recognizer F - fully-annotated dataset D - weakly annotated dataset Di - dataset with PGT

The PGT generation algorithm:

Algorithm 0: PGT-GEN **Input:** *I*, *O*, *G* - Input image, E2E output, weak annotations **Output:** PGT P := AssignWeak(O, G); $PGT := \{\};$ foreach $(bb, tt, g) \in P$ do $(bb_f, tt_f) = FindOptimalBox(I, bb, tt, g);$ if $IsPGT(tt_f, g)$ then $PGT = PGT \cup \{(bb_f, g)\};$ enc return PGT

Amazon Book Covers dataset (ABC)



(UT)

Uber-Text dataset



The accuracy of PGT verified on 500 samples from each dataset:

False positives

- 4.1 % on Uber-Text
- 2.0 % on Amazon Book Covers

Experiments: PGT analysis

PGT instances generated: 1 594 333 (ABC)



113 810 (UT)

Lluis Gómez, Dimosthenis Karatzas Computer Vision Center Universitat Autonoma de Barcelona Spain

The **AssignWeak** method is implemented as a mutually nearest weak label assignment in a bipartite graph, where detections (and their text transcriptions) and weak labels become pgt candidates if they are mutually nearest, using the edit distance metric:.



The **FindOptimalBox** searches the neighbourhood of the original box for the bounding box minimizing the edit distance from the weak label.

Example: P GROCER the assigned weak label is 'GROCERY' "P GROCE" the edit distance is minimized by GROCER GROCERY extending the bounding box to the right "P GROCER" "P GROCERY" and shrinking on the left GROCERY <u>GROCERY</u> " GROCERY" **"GROCERY"**

PGT images:

with artistic fonts, blurred texts, perspective transformation.





https://arxiv.org/abs/2007.03098

