

Text Recognition: Real World Data and Where to Find Them

ICPR 2020

Klára Janoušková



Jiří Matas



Visual Recognition Group
FEE, Czech Technical University in Prague
Czech Republic

Lluís Gómez



Dimsothenis Karatzas



Computer Vision Center
Universitat Autònoma de Barcelona
Spain

Annotating real world data is costly.
Scene text recognition training therefore relies
heavily on synthetic data...



Investment



204 NOSTAL



180



BRUNK HED



you and



heavy



180



prove



has

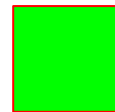
...yet images with text metadata available are common

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



Weak annotation:

The Sherlock
Holmes Museum,
221B Baker Street

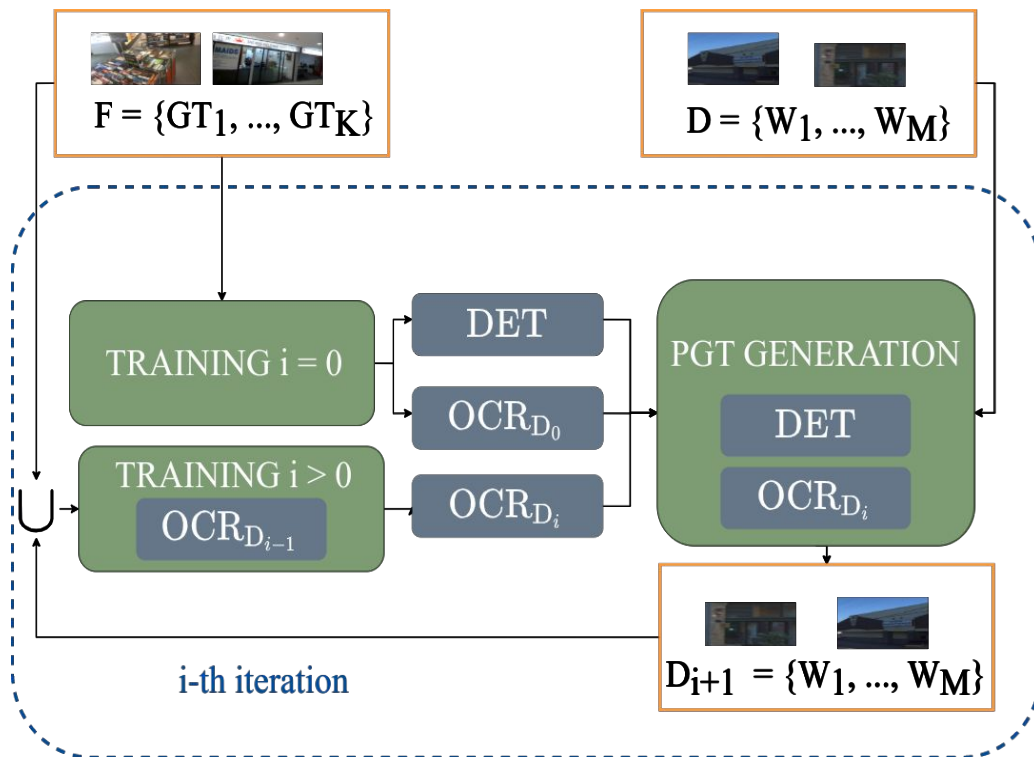


Words in weak
annotation



Words missing

The Method

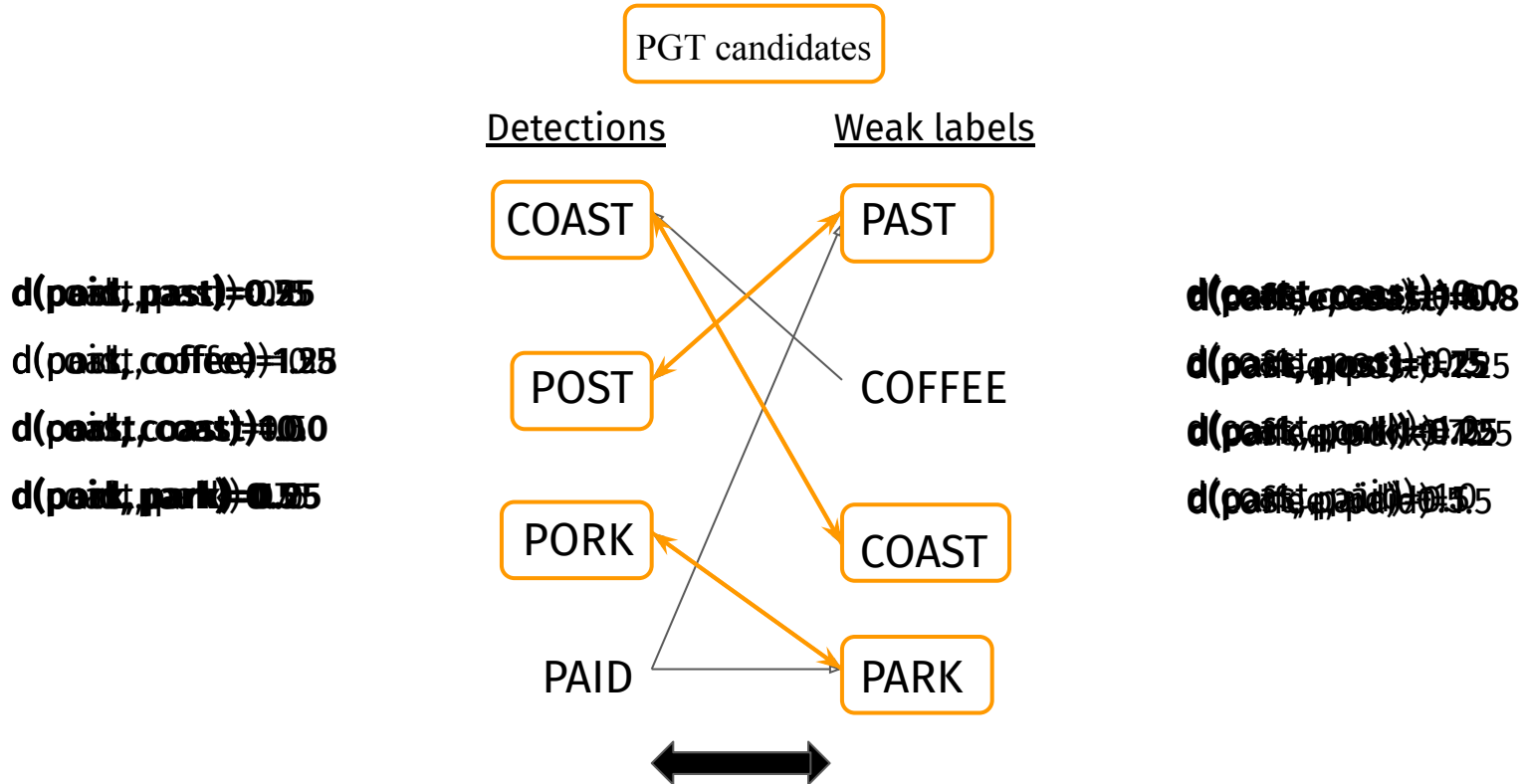


PGT - pseudo ground truth
DET - detector

OCR - recognizer
F - fully-annotated dataset

D - weakly annotated dataset
D_i - dataset with PGT

PGT Generation: Mutually-nearest Weak Label Assignment in a Bipartite Graph



PGT Generation: Edit-distance Guided Neighbourhood Search

Detected text: P GROCE

Weak label: GROCERY



"P GROCE"

"P GROCER"

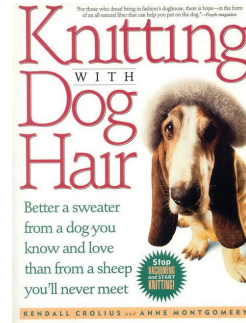
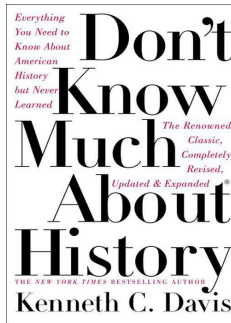
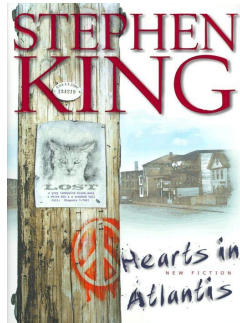
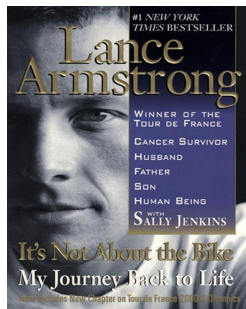
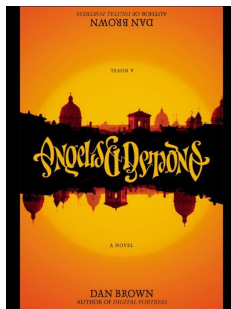
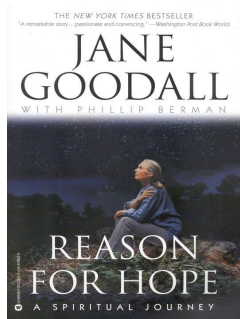
"P GROCERY"

" GROCERY"

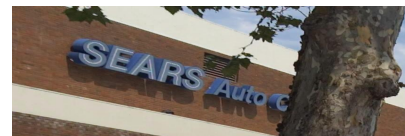
"GROCERY"

Weakly Annotated Datasets in our Experiments

Amazon Book Covers - the author and title used as weak annotations

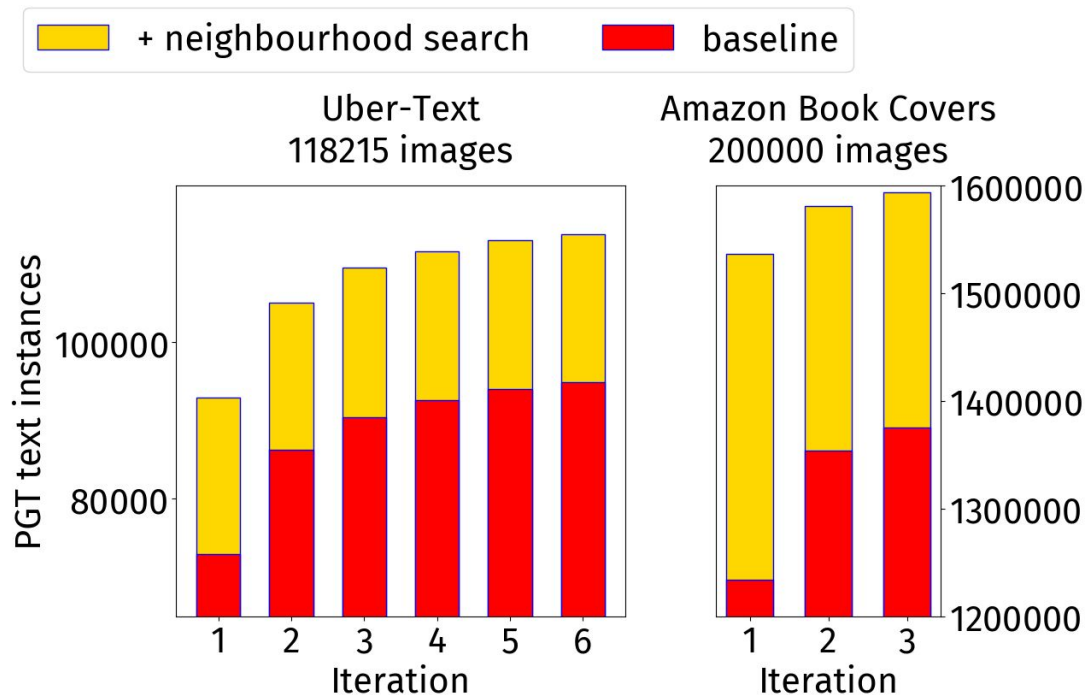


Uber-Text - fully annotated, we ignore the text location



PGT: Quantitative Results

Number of PGT text instances



False positives

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



PGT: **C**ancer
GT: CONNER



PGT: **H**ILLING SERVICES
GT: KILLING SERVICES

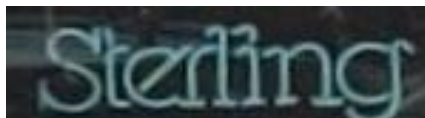


PGT: 1:15,**0**00
GT: 1:15 000

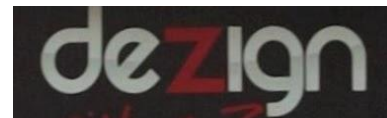
PGT Qualitative Examples



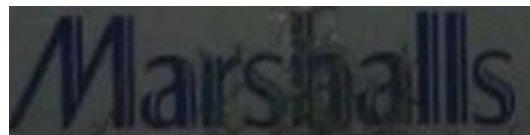
rockstar



sterling



deZign



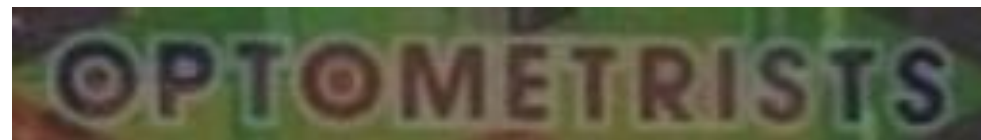
marshalls



mitre



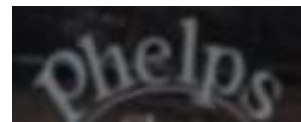
paolo



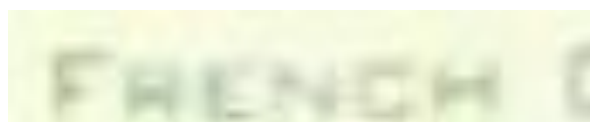
optometrists



fitness



phelps

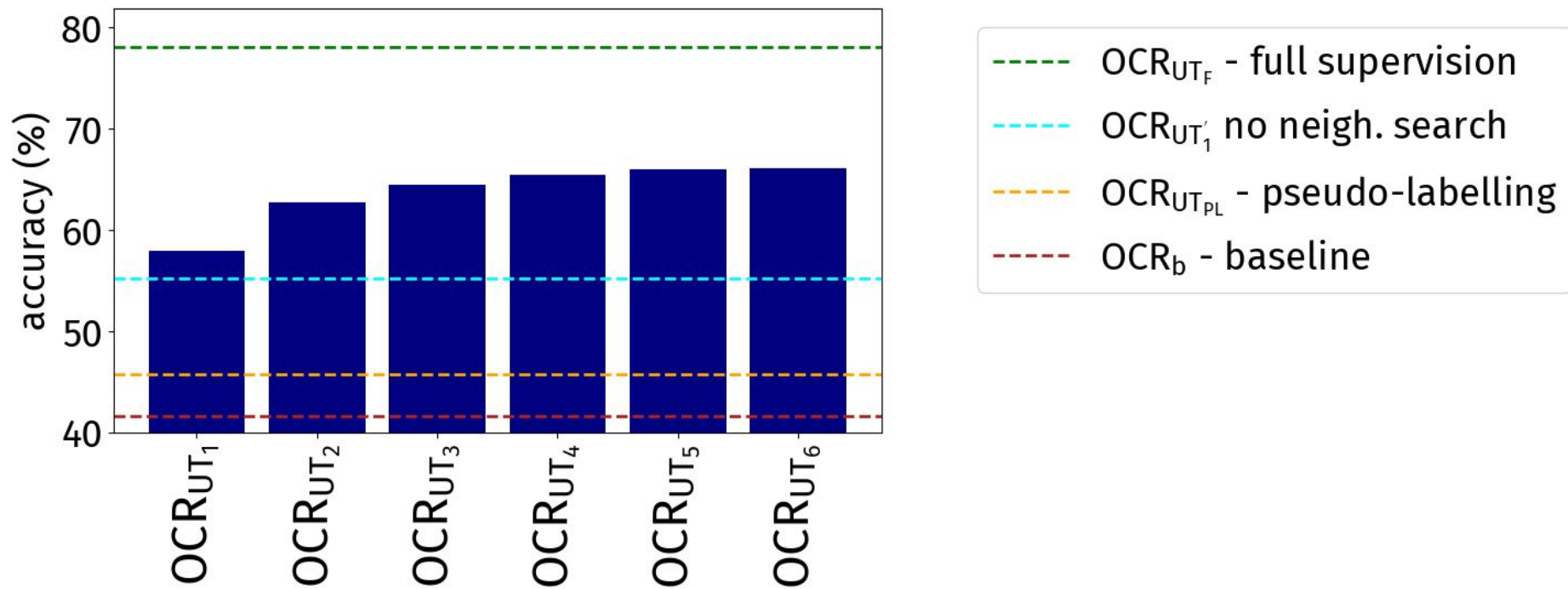


french

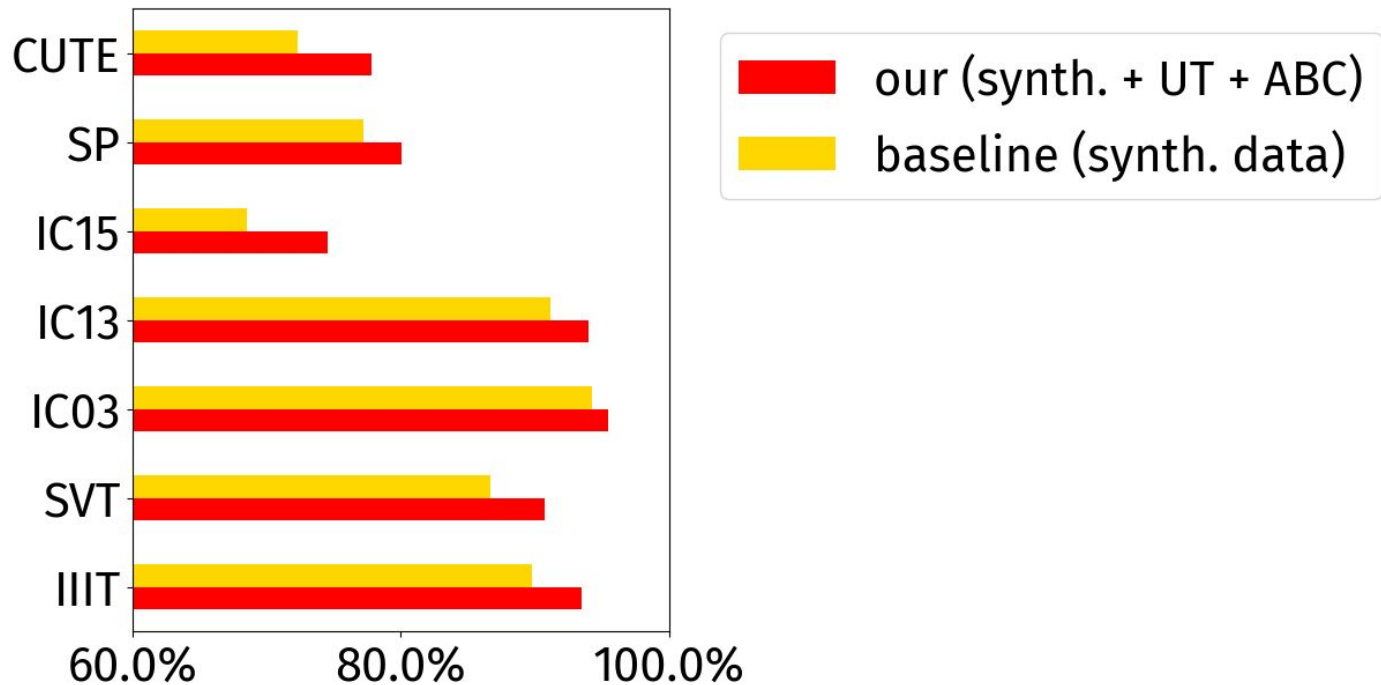


cane's

Domain Adaptation on Uber-Text: Word accuracy



Results on Benchmarks: Word Accuracy



Character set: [abcdefghijklmnopqrstuvwxyz
0123456789!"#\$%&'()*+,-./:;<=>?@[\\]^_`{|}~]

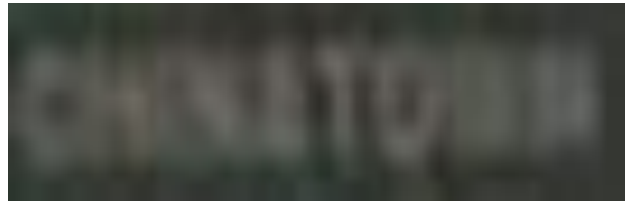
Recognition after Training with PGT: Improvement



brioffole
→ barbeque



jow
→ join



chantronn
→ chinetown



rark
→ park



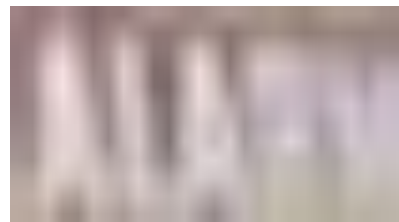
guice
→ guide



literama
→ cinerama



dopert
→ expert



alaet
→ alarm

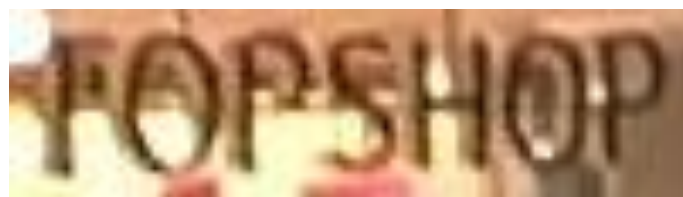
Recognition after Training with PGT: Deterioration



mink
→ mark



tape
→ tapl



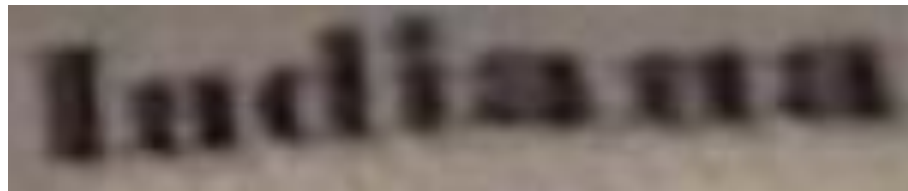
topshop
→ forshop



tqbu
→ tqlu

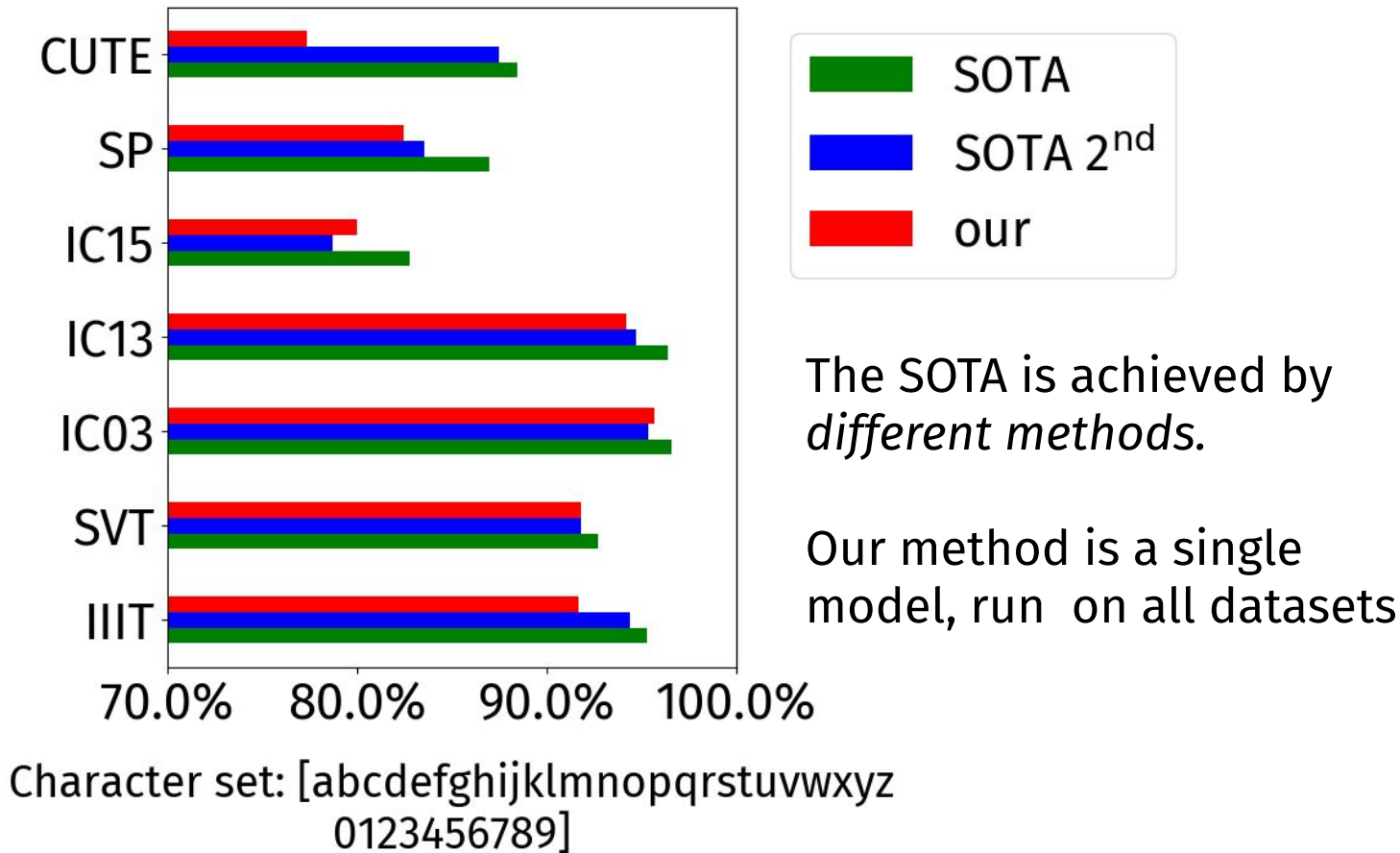


arlboro
→ arljoro



indiana
→ ludiana

Benchmarks: Comparison with State of the Art



Thank you



<https://arxiv.org/abs/2007.03098>