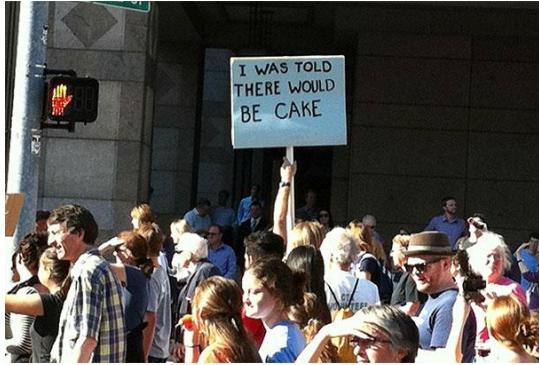


You Only Scale Once

Efficient Text Detection Using Adaptive Scaling

Elad Richardson, Tomer Ronen, Niv Geron, Zach Avraham, Alon Palombo and Yaniv Azar

So what is text detection?



ICDAR13

ICDAR15

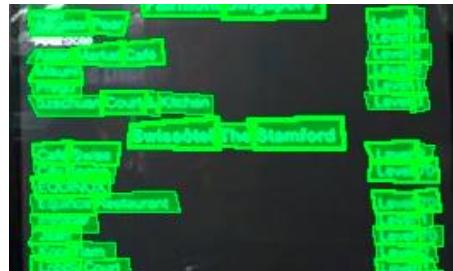
ICDAR17

What makes it difficult?

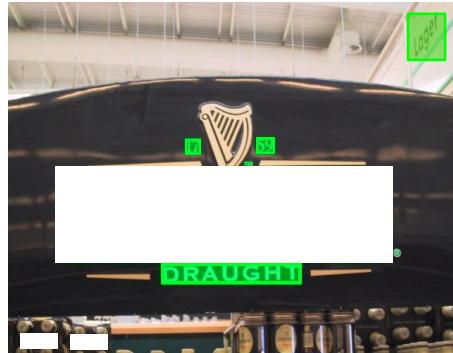
Scale Variability



Density



Irregular Shapes



ICDAR13



VIA 9GAG.COM

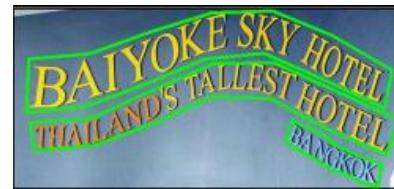
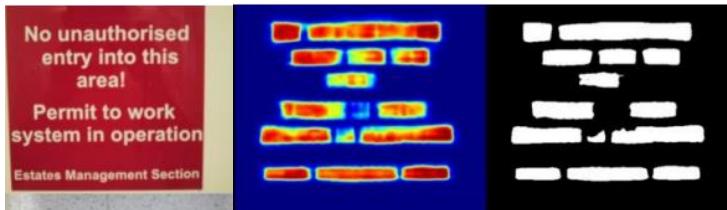
ICDAR15



ICDAR17

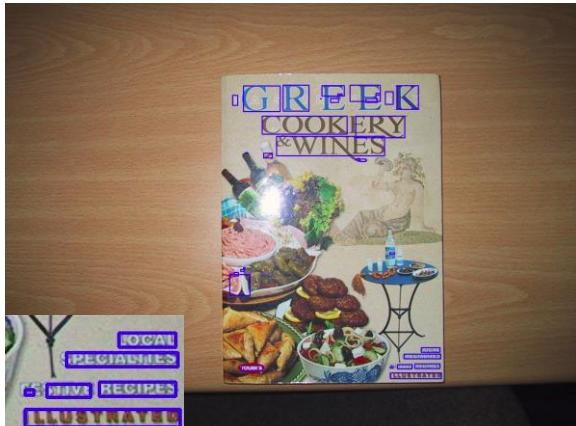
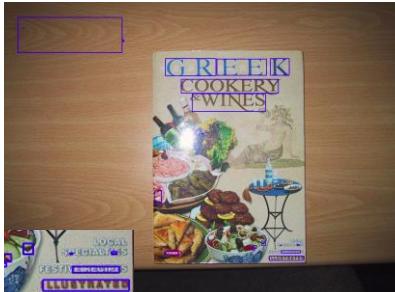
So how do we solve it?

- Anchor-based methods dominate the Object Detection world
 - Widely used for Text Detection as well
- But Semantic Segmentation is getting pretty common
 - Works pretty good for text (Tagging is tight)
 - Better support for rotations and irregular text
- Tends to connect words
 - Solved using different regularizations



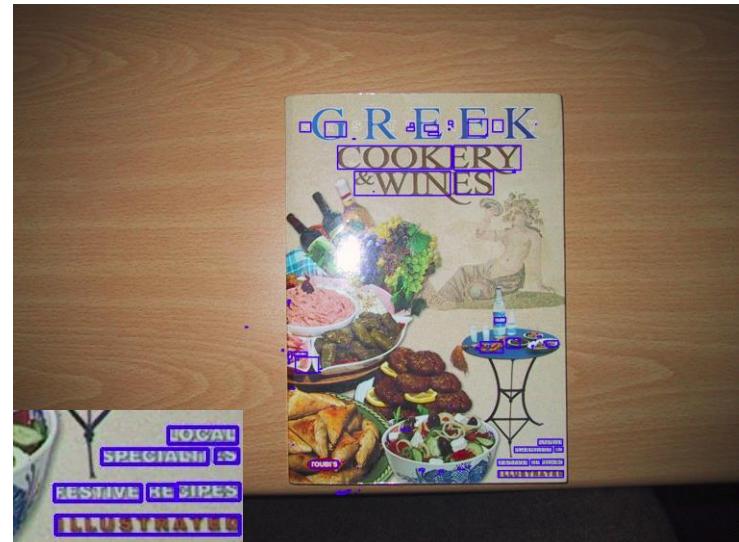
How can we improve our baseline?

- Simply run Multi-Scale
 - Using predefined scales (0.5, 0.7, 1, 1.4, 2)
 - Each scale captures different text regions
- But takes much more time
 - Not practical in many scenarios 😔



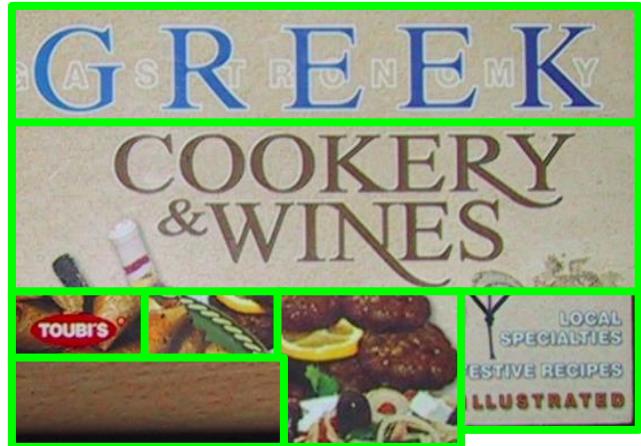
ICDAR15 Benchmark

Method	Recall	Precision	F-Score
EAST	0.73	0.83	0.78



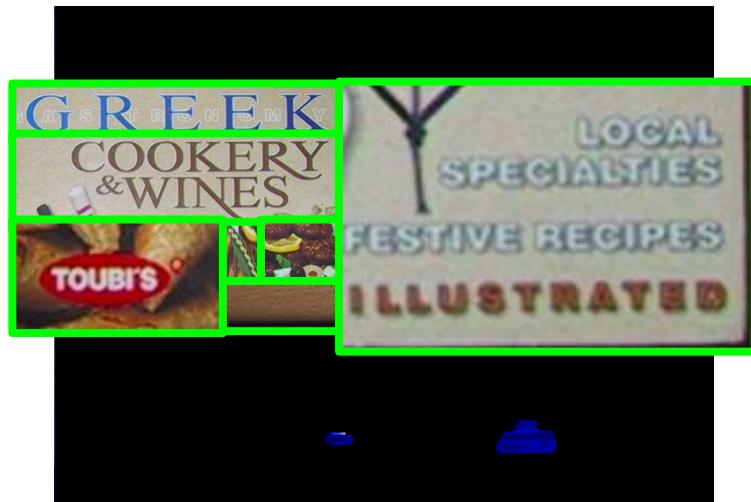
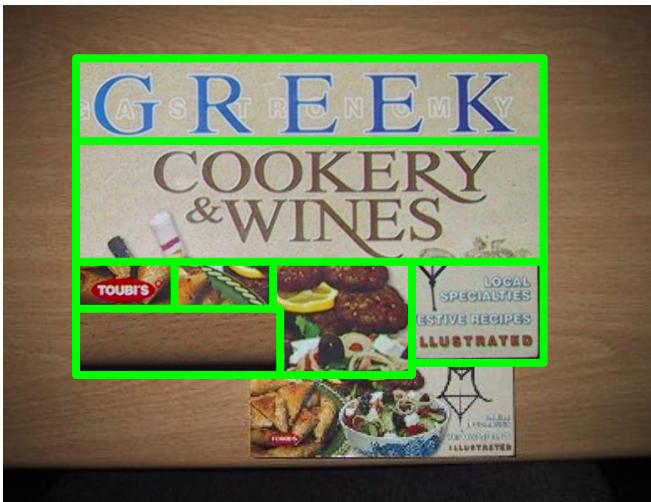
Can we do better?

- Base Scale is enough for “Coarse Detection”
- What if we filter background regions? 🤔
 - Get a compact image representation
- Apply Multi-Scale only on regions of interest



Can we do even better?

- Text detection works pretty well under the right scale
 - But we don't know the scale of each region 😔
- So why not learn it?
 - Can resize all text to the desired scale!

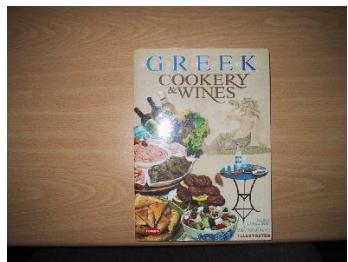


Our Solution In a Nutshell

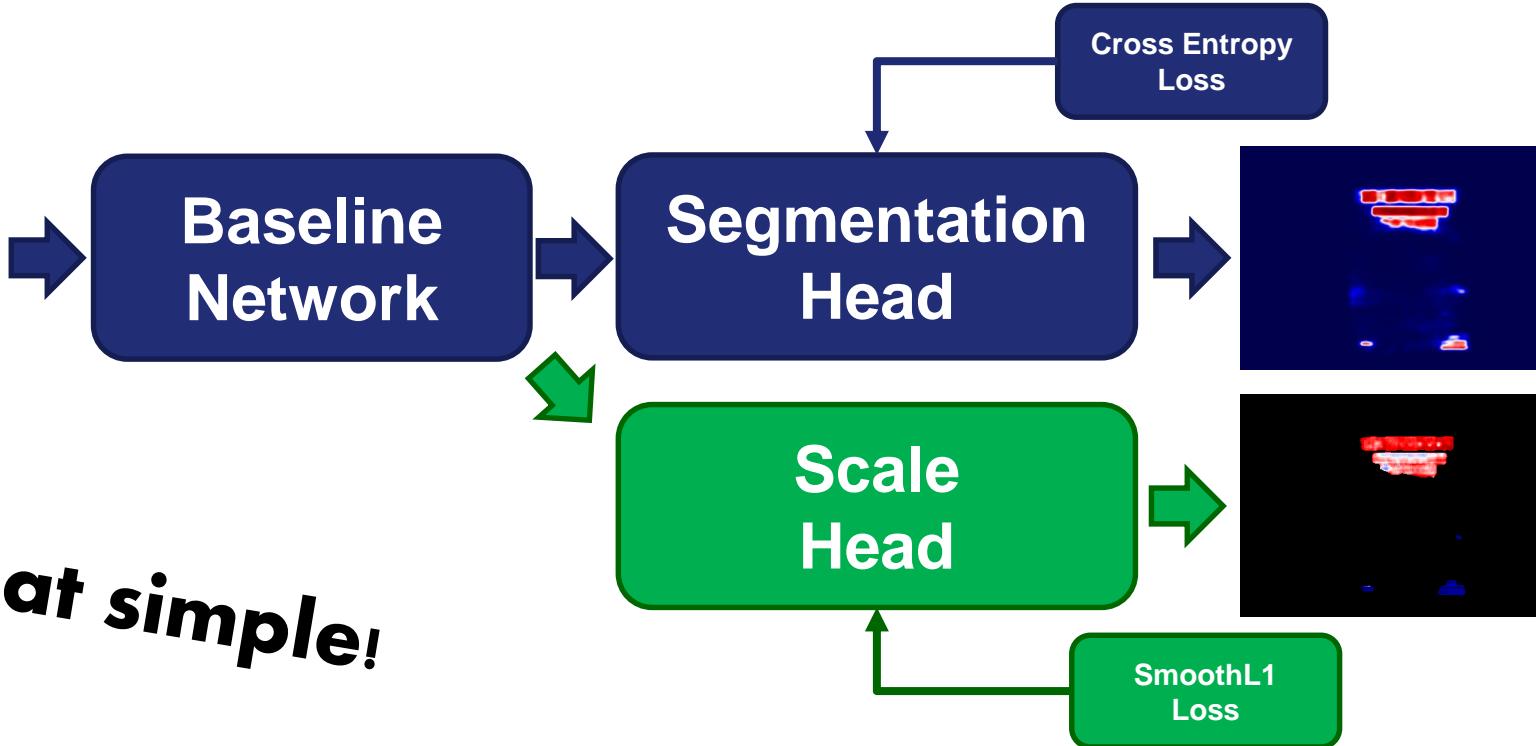
- We already do a single forward pass
 - Let's gather some information there (segmentation + scale)
- Use it to create the “optimal scale”
 - And do only one additional forward pass 🤘



So what does it take?

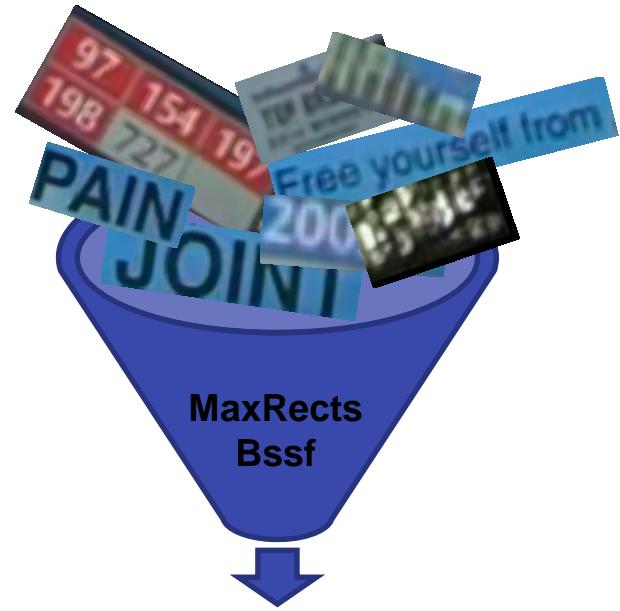


It's that simple!

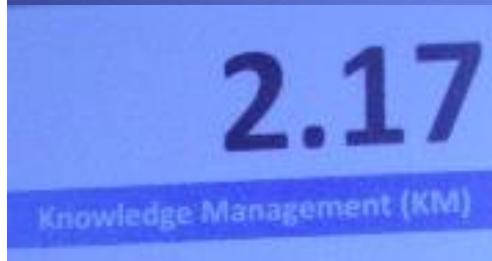
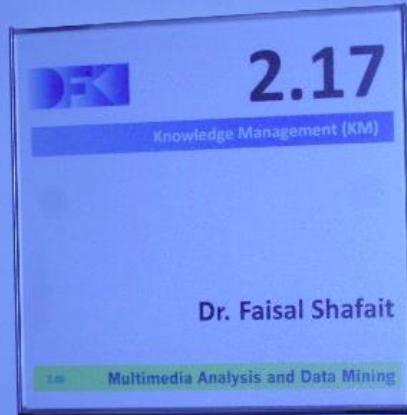


The Packing Problem

- How can we pack the blobs efficiently?
 - Use a 2D Knapsack solution
 - Specifically, the “**Maximal Rectangles Best Short Side Fit**” algorithm
- But knapsack images are not realistic
 - Add a knapsack augmentation to training!



Results!





Multimedia Analysis and Data Mining

Knowledge Management (KM)

Dr. 2.17

Shafait

Faisal

Multimedia Analysis and Data Mining

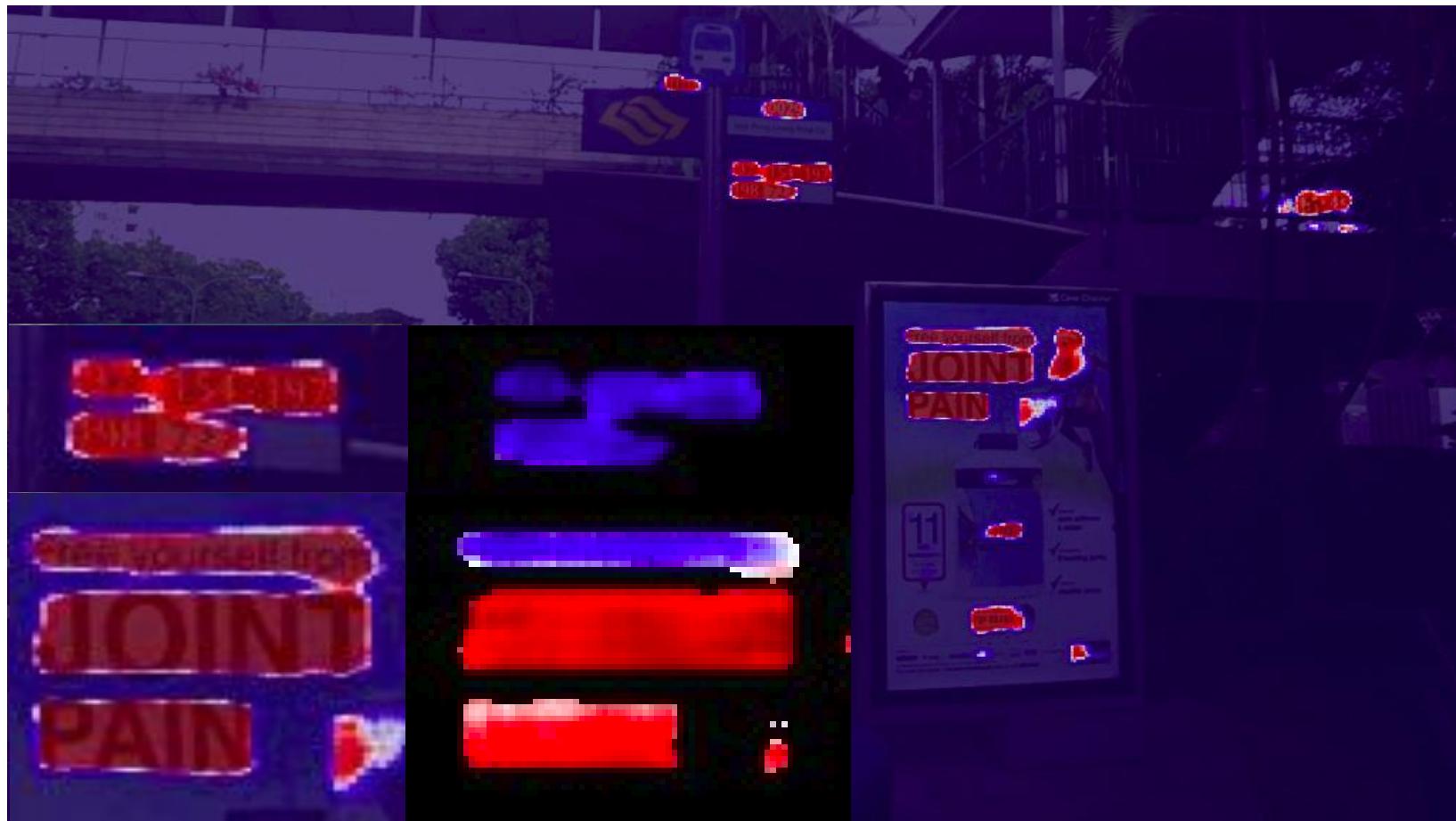
Knowledge Management (KM)

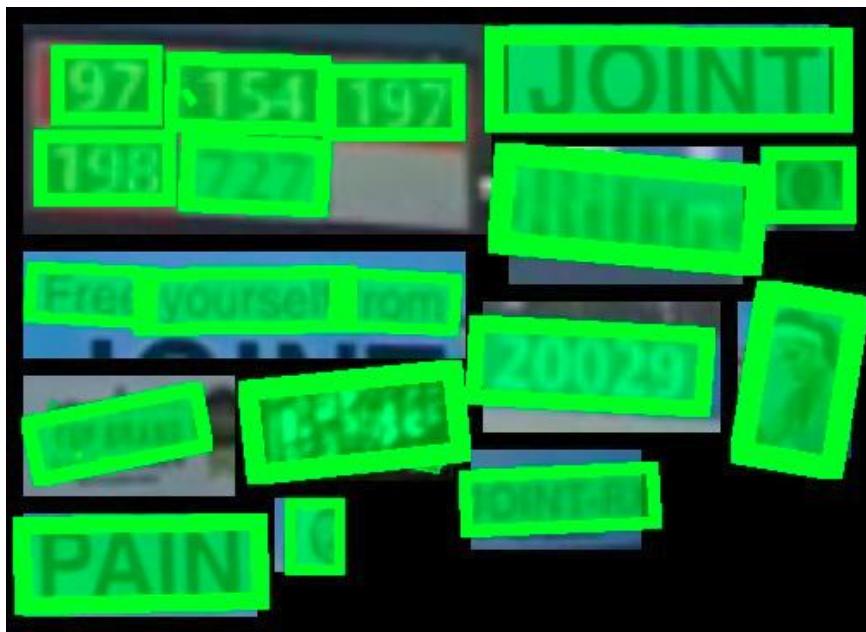
Dr. 2.17

Shafait

Faisal











How does it compare?

You Only Scale Once



Multi-Scale



Method	Recall	Precision	F-Score
Baseline	41.47%	46.09%	43.81%
Ours	59.12%	51.25%	54.91%
Multiscale	59.7%	30.56%	40.42%

Conclusion

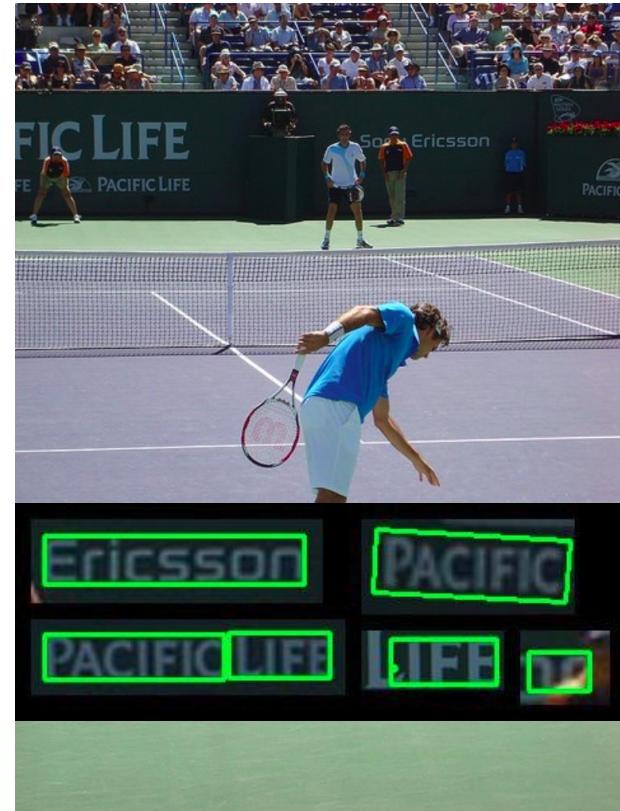
- A simple technique to boost single-scale methods
 - Without the overhead of running in multiscale 🤘
- Easily applied on top of any text detection algorithm
 - And possible general object detection
- Code available soon 😊



ICDAR13



ICDAR15



ICDAR17

Thanks!

