

# **Total-Text:**

## A Comprehensive Dataset for Scene Text Detection and Recognition

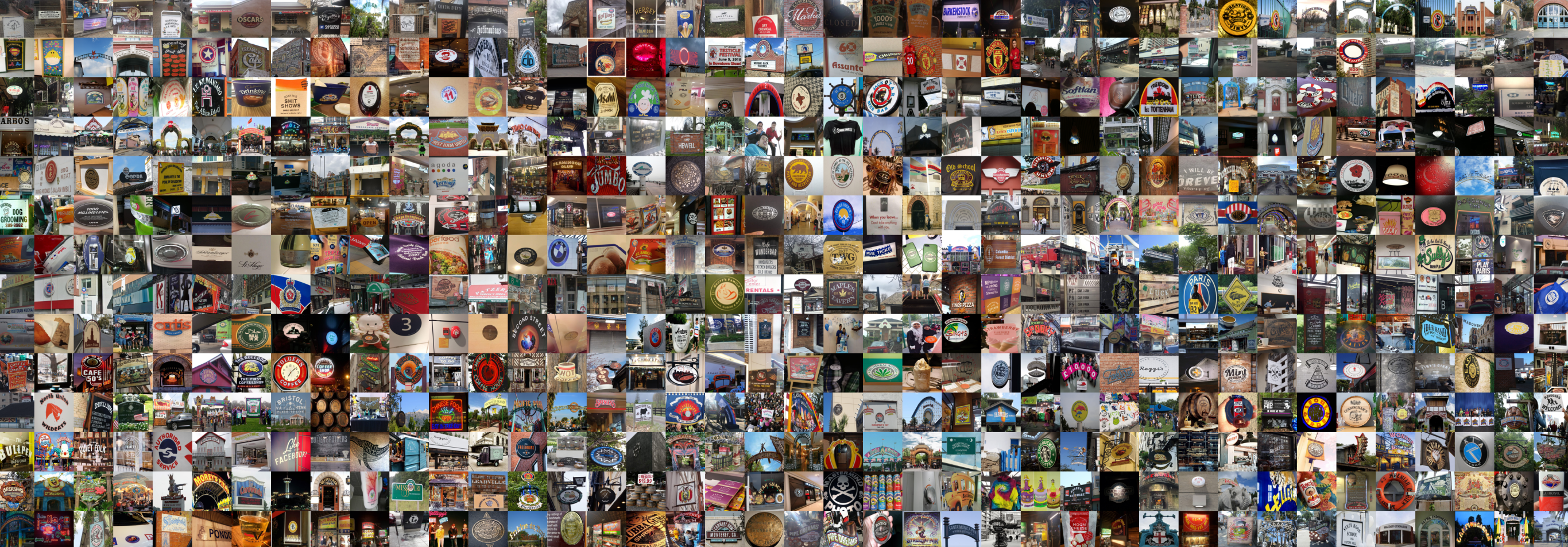
*Chee Kheng Ch'ng*

*Chee Seng Chan*

*Centre of Image & Signal Processing, Computer Science and Info. Technology  
University of Malaya, Malaysia*

*chngcheekheng@gmail.com cs.chan@um.edu.my*





<https://github.com/cs-chan/Total-Text-Dataset>





# Introduction

- **Total-Text** is the first dataset that features 3 different orientations: *horizontal, multi-oriented, and curve*.
- It has 1555 images in total, with 11459 annotated text instances.





# Motivation



- Curved text is commonly seen in real world scenery, but it has close to zero existence in existing datasets.
- As a result, text detection algorithms with curved-oriented text in consideration is rarely seen.
- With the introduction of Total-Text, we hope to spur an interest in the community of scene text understanding.



# Related Works - *Scene Text Datasets*

Datasets	No. of Images	Text Orientation	Year
ICDAR(2003-13)	462	Horizontal	2003 <small>(was then revised several times)</small>
MSRA-TD500	500	Horizontal, Multi-oriented	2012
ICDAR2015	1670	Horizontal, Multi-oriented	2015
COCO-Text	63686	Horizontal, Multi-oriented	2016
Total-Text	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*



# Related Works - *Scene Text Datasets*

Datasets	No. of Images	Text Orientation	Year
ICDAR(2003-13)	462	Horizontal	2003 (was then revised several times)
MSRA-TD500	500	Horizontal, Multi-oriented	2012
ICDAR2015	1670	Horizontal, Multi-oriented	2015
COCO-Text	63686	Horizontal, Multi-oriented	2016
Total-Text	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*





# Related Works - Scene Text Datasets

Datasets	No. of Images	Text Orientation	Year
ICDAR(2003-13)	462	Horizontal	2003 (was then revised several times)
<b>MSRA-TD500</b>	500	Horizontal, Multi-oriented	2012
ICDAR2015	1670	Horizontal, Multi-oriented	2015
COCO-Text	63686	Horizontal, Multi-oriented	2016
Total-Text	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*





# Related Works - *Scene Text Datasets*

Datasets	No. of Images	Text Orientation	Year
<b>ICDAR(2003-13)</b>	462	Horizontal	2003 (was then revised several times)
<b>MSRA-TD500</b>	500	Horizontal, Multi-oriented	2012
<b>ICDAR2015</b>	1670	Horizontal, Multi-oriented	2015
<b>COCO-Text</b>	63686	Horizontal, Multi-oriented	2016
<b>Total-Text</b>	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*

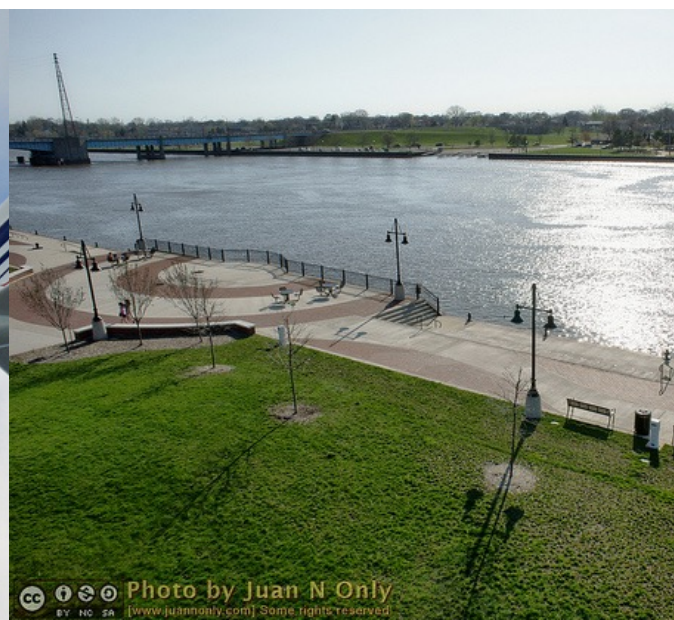




# Related Works - *Scene Text Datasets*

Datasets	No. of Images	Text Orientation	Year
<b>ICDAR(2003-13)</b>	462	Horizontal	2003 <small>(was then revised several times)</small>
<b>MSRA-TD500</b>	500	Horizontal, Multi-oriented	2012
<b>ICDAR2015</b>	1670	Horizontal, Multi-oriented	2015
<b>COCO-Text</b>	63686	Horizontal, Multi-oriented	2016
<b>Total-Text</b>	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*





# Related Works - Scene Text Datasets

Datasets	No. of Images	Text Orientation	Year
ICDAR(2003-13)	462	Horizontal	2003 (was then revised several times)
MSRA-TD500	500	Horizontal, Multi-oriented	2012
ICDAR2015	1670	Horizontal, Multi-oriented	2015
COCO-Text	63686	Horizontal, Multi-oriented	2016
<b>Total-Text</b>	1555	Horizontal, Multi-oriented, Curve	2017

*Well known scene text datasets*



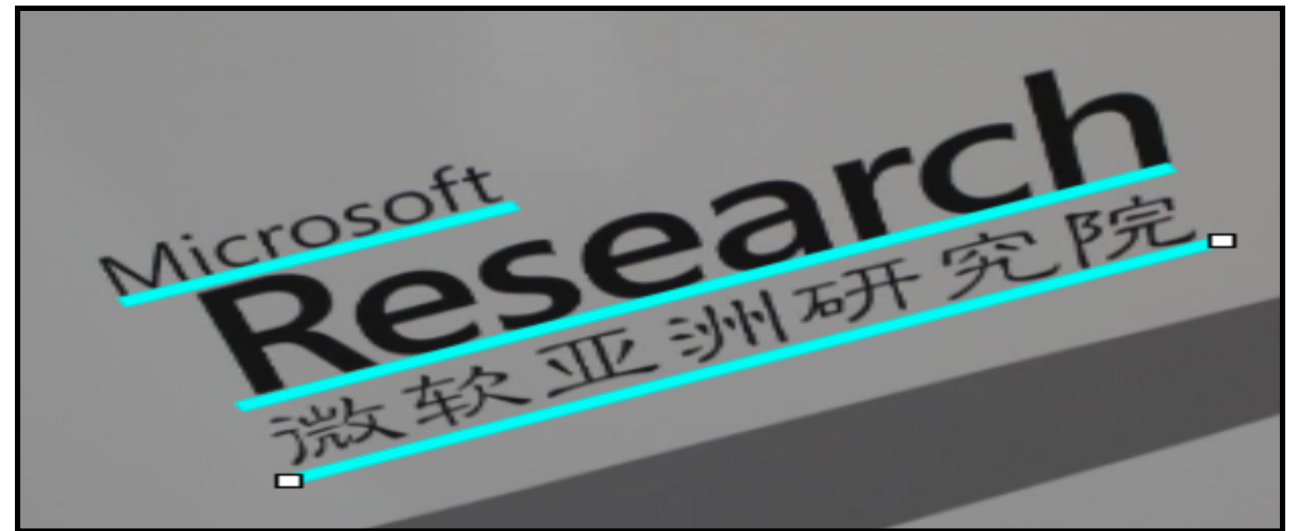


# Curve Text Observations

- Unlike horizontal and multi-oriented text, curved text cannot be connected by a *straight line*.



*Horizontal text from ICDAR 2013*



*Multi-oriented text from MSRA-500*



*Curved text from Total-Text*



# Curve Text Observations

- Unlike horizontal and multi-oriented text, curved text cannot be connected by a *straight line*.
- Curved text in our dataset range from *slightly* to *extremely curved*.





# Curve Text Observations

- Unlike horizontal and multi-oriented text, curved text cannot be connected by a *straight line*.
- Curved text in our dataset range from *slightly to extremely curved*.
- Majority of them exist in the shape of symmetric arc.





# Orientation Assumptions

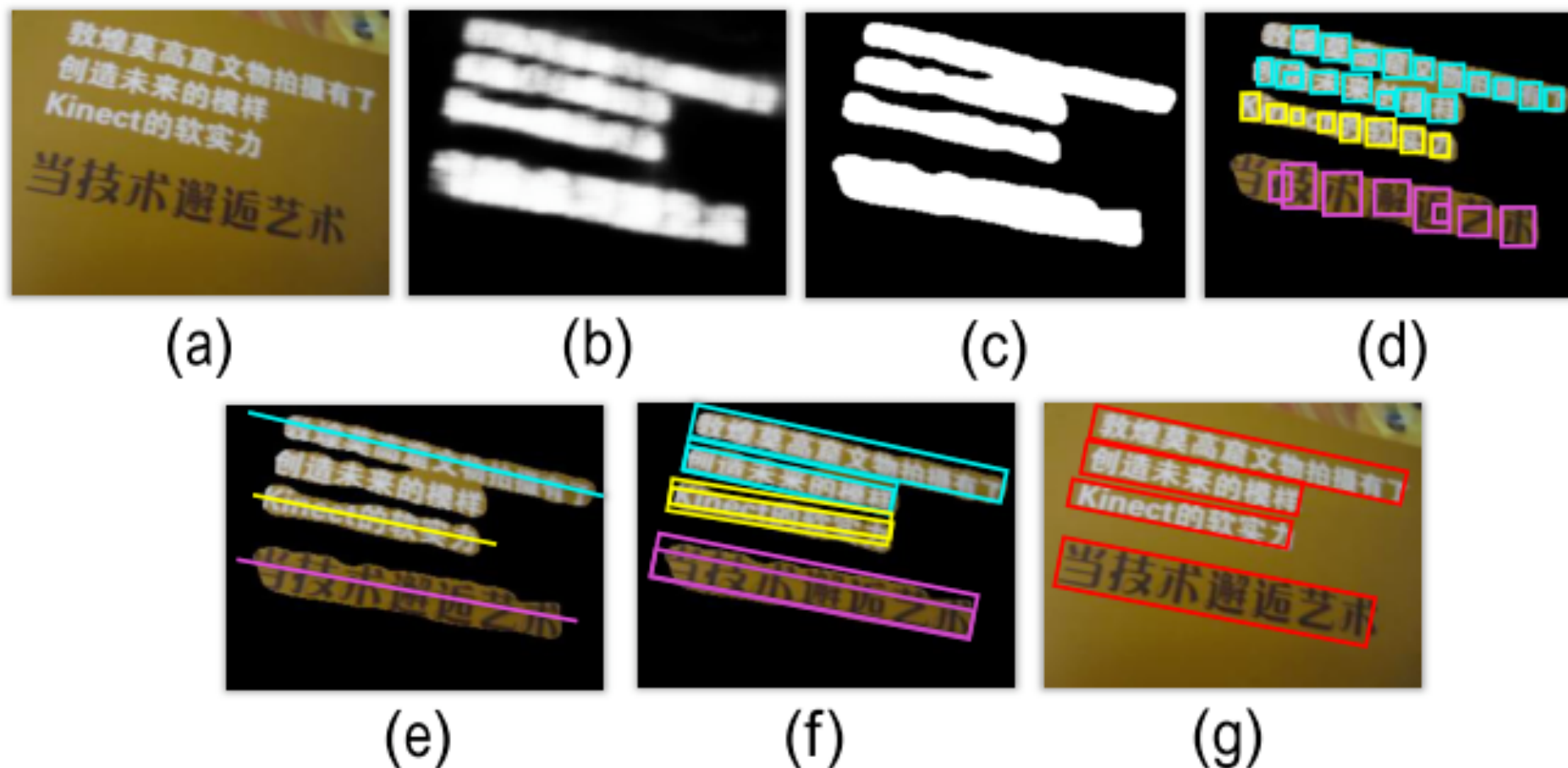
- Many text detection algorithms have rules and assumptions on text orientations.
- We took a closer look into some of these assumptions and see how it would fits the observation that we have made on curved text.
- Specifically, we looked into two high performing multi-oriented scene text detection systems that have reported their result on MSRA-TD500.



# Orientation Assumptions

*Multi-Oriented Text Detection with Fully Convolutional Networks, 2016*  
[1]

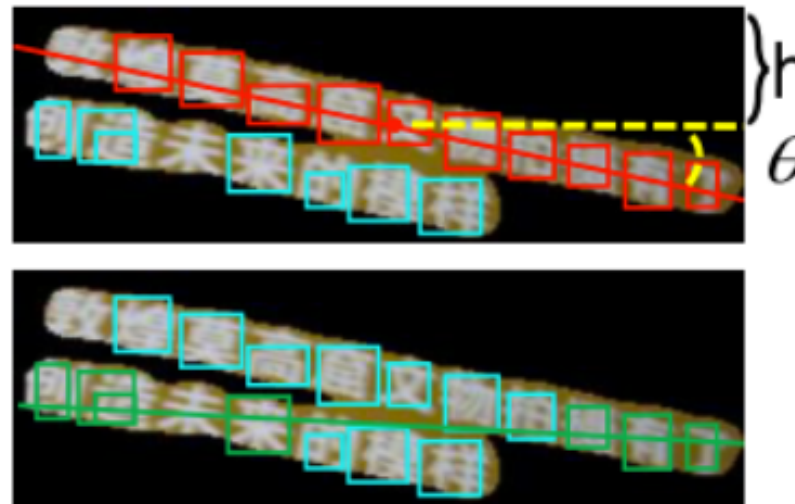
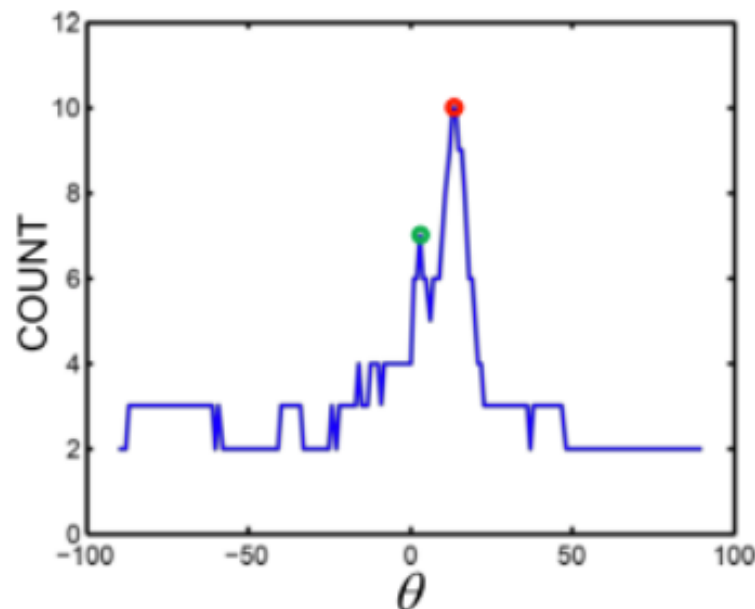
- Uses **FCN** to spot for **text regions**; **MSER** to spot for **character candidates**.
- Combining both **global** (text regions) and **local** (characters) to infer **text line candidates**.





# Orientation Assumptions

- The algorithm tries to determine the best  $h$  and  $\theta$  for the **straight line** to hit the most character candidates.



$$\theta_r = \arg \max_{\theta} \max_h \Phi(\theta, h)$$



# Orientation Assumptions

- The algorithm tries to determine the best  $h$  and  $\theta$  for the **straight line** to hit the most character candidates.
- A straight line would miss characters arranged in curve orientation.

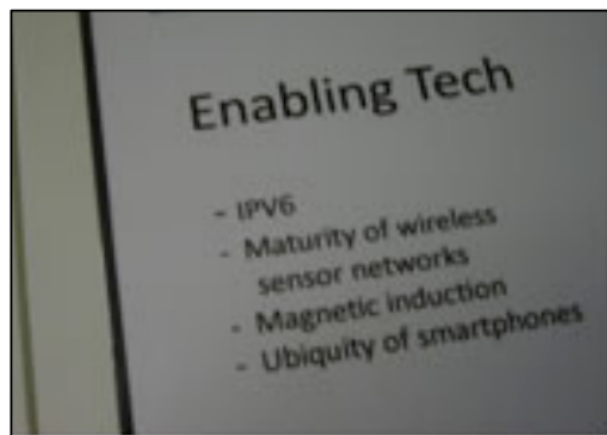




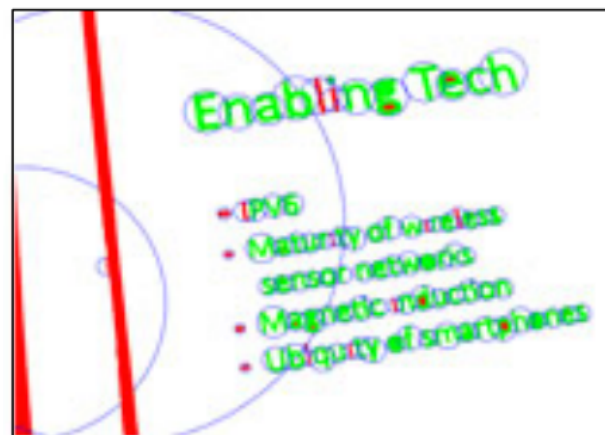
# Orientation Assumptions

*Multi-Orientation Scene Text Detection with Adaptive Clustering, 2015[2]*

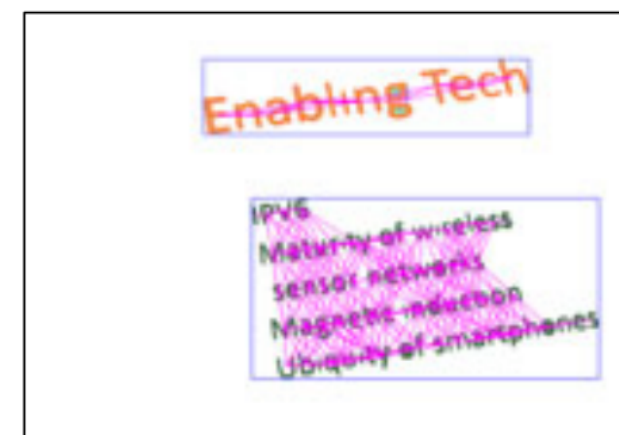
- Generate text line candidates by grouping character candidates with *similar* **color, stroke width, location differences, and orientation.**



(a) Original image



(b) Character candidates



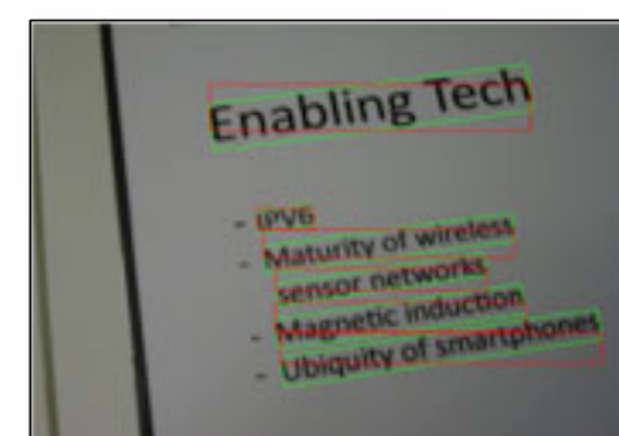
(c) Text candidates by morphology



(d) Text candidates by orientation



(e) Text candidates by projection



(f) Detection results



# Orientation Assumptions

- Character candidates within a single word can have multiple **orientation variations**.





# Orientation Assumptions



(a) Yin *et al.* [1] (red bounding box) and Huang *et al.* [2] (blue bounding box)

(b) Shi *et al.* [3]

1. Yin *et al.* (2014) "Robust Text Detection in Natural Scene Images" *T-PAMI*, vol. 36(5), pp. 970-983.
2. Huang *et al.* "Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees", *ECCV 2014*.
3. Shi *et al.* "Detecting Oriented Text in Natural Images by Linking Segments", *CVPR 2017*.



# Dataset Statistics

- Train set: 1255 images, Test set: 300 images
- All 11459 text instances can be broken down into the table below.

	Train	Test
Curved	3936	971
Horizontal	2841	744
Multi-oriented	2487	480
Total	9264	2195

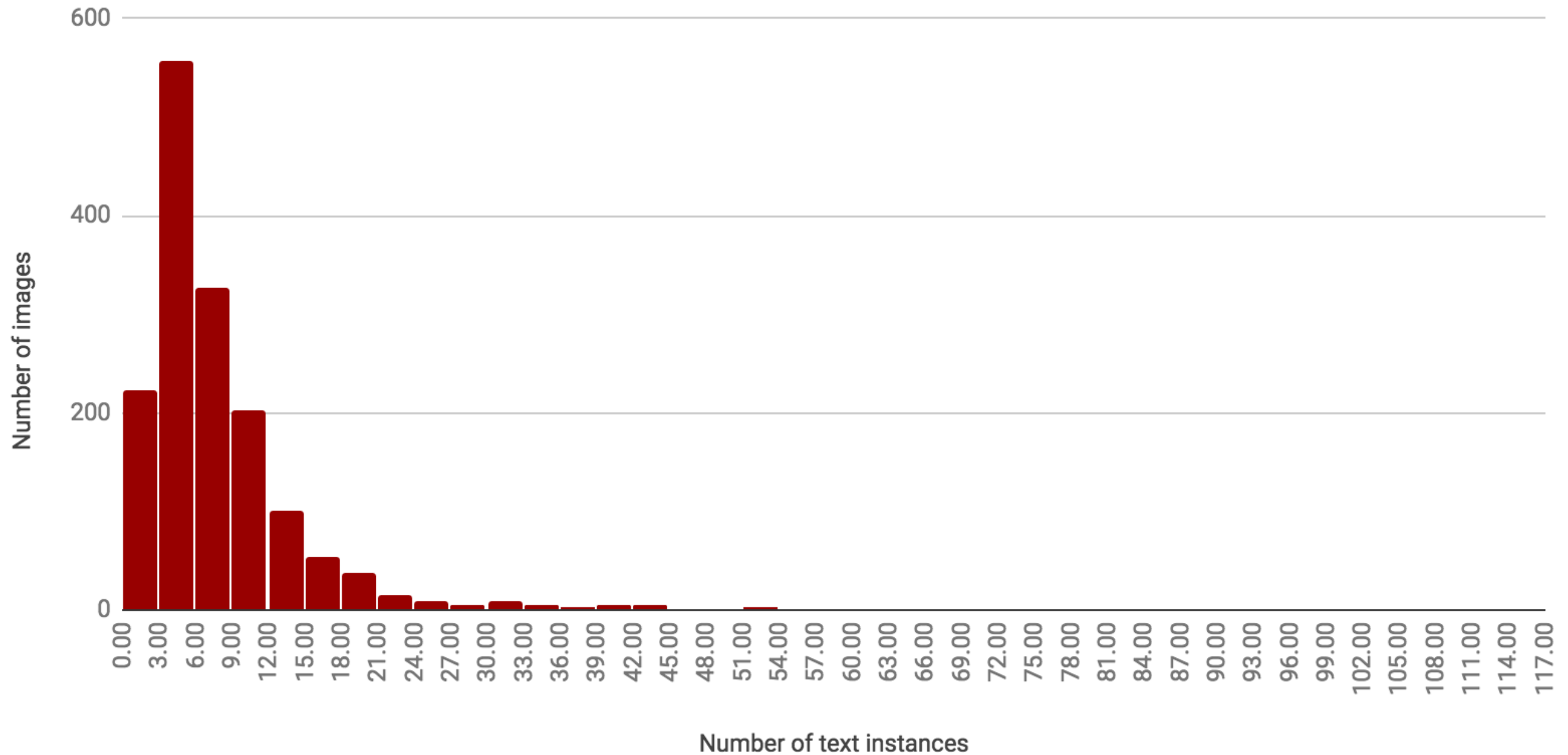


# Dataset Statistics

Datasets	No. of images	No. of Text Instances	Text instances per image
<i>ICDAR2013</i>	462	1943	4.2
<i>ICDAR2015</i>	1670	11886	7.12
<i>MSRA-TD500</i>	500	1719	3.4
<i>COCO-Text</i>	63686	173589	2.73
<b><i>Total-Text</i></b>	1555	11459	7.37



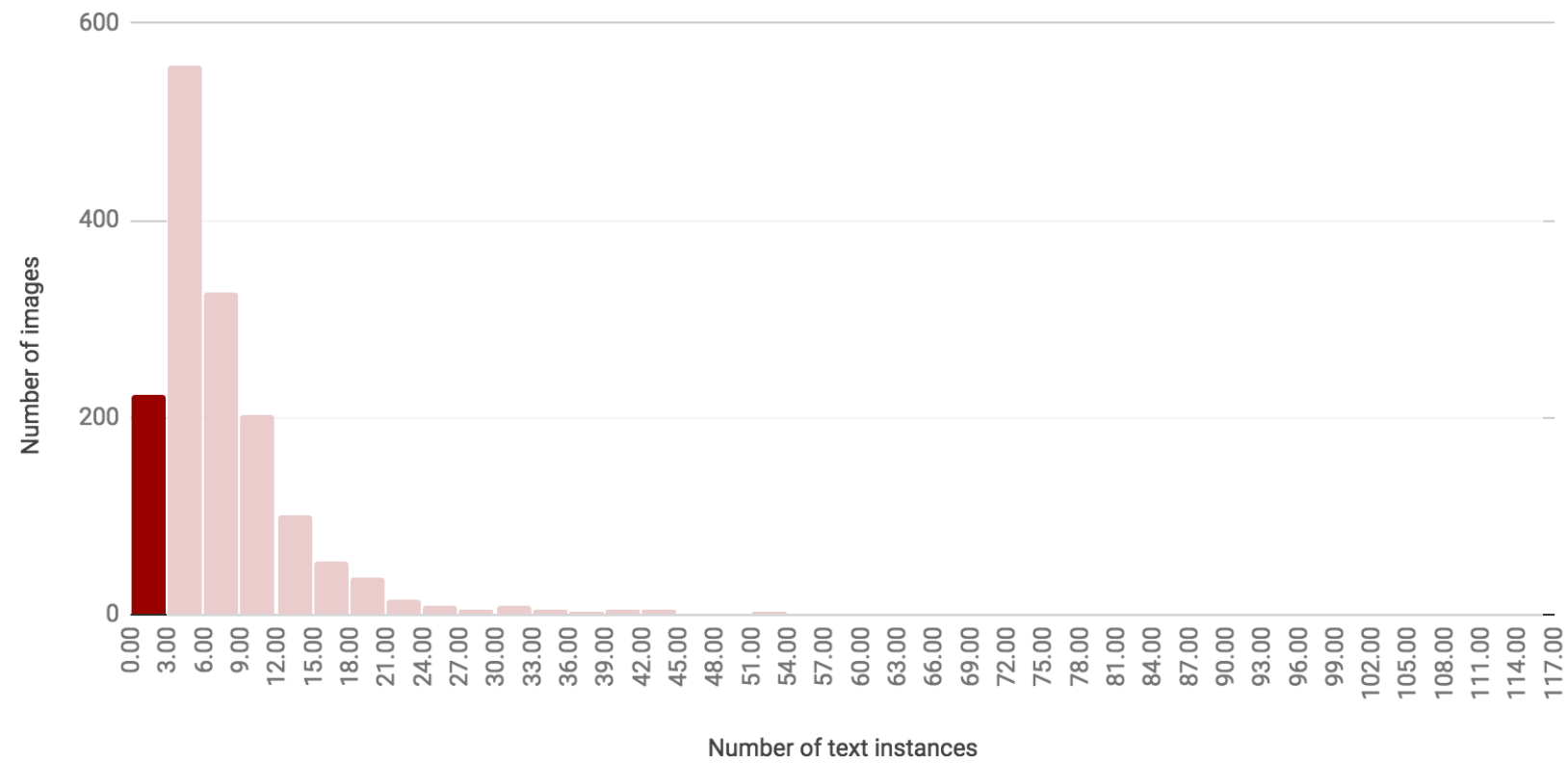
## Distribution of Text Instances Across All Images of Total-Text





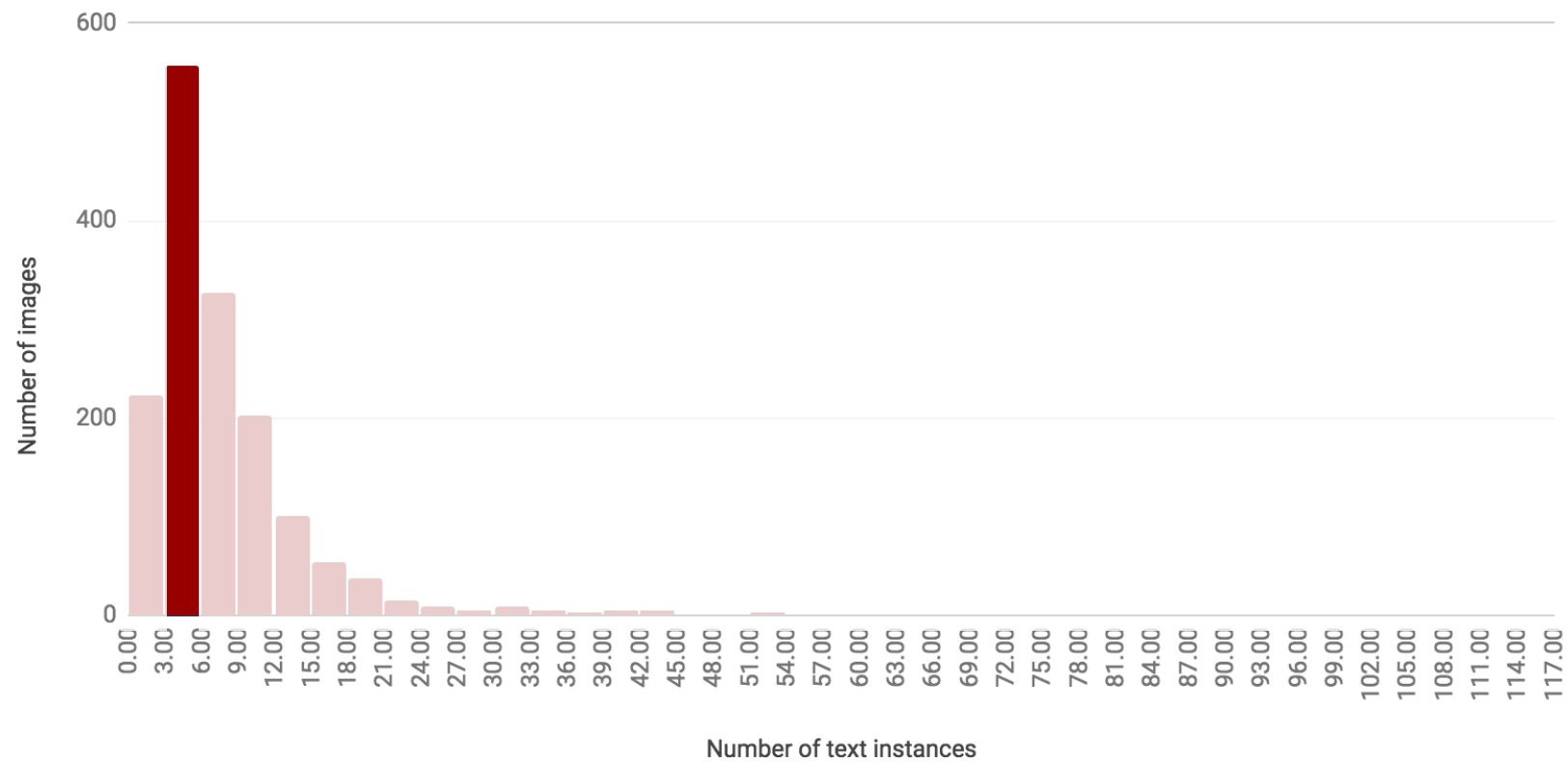
# Dataset Statistics

Distribution of Text Instances Across All Images of Total-Text



# Dataset Statistics

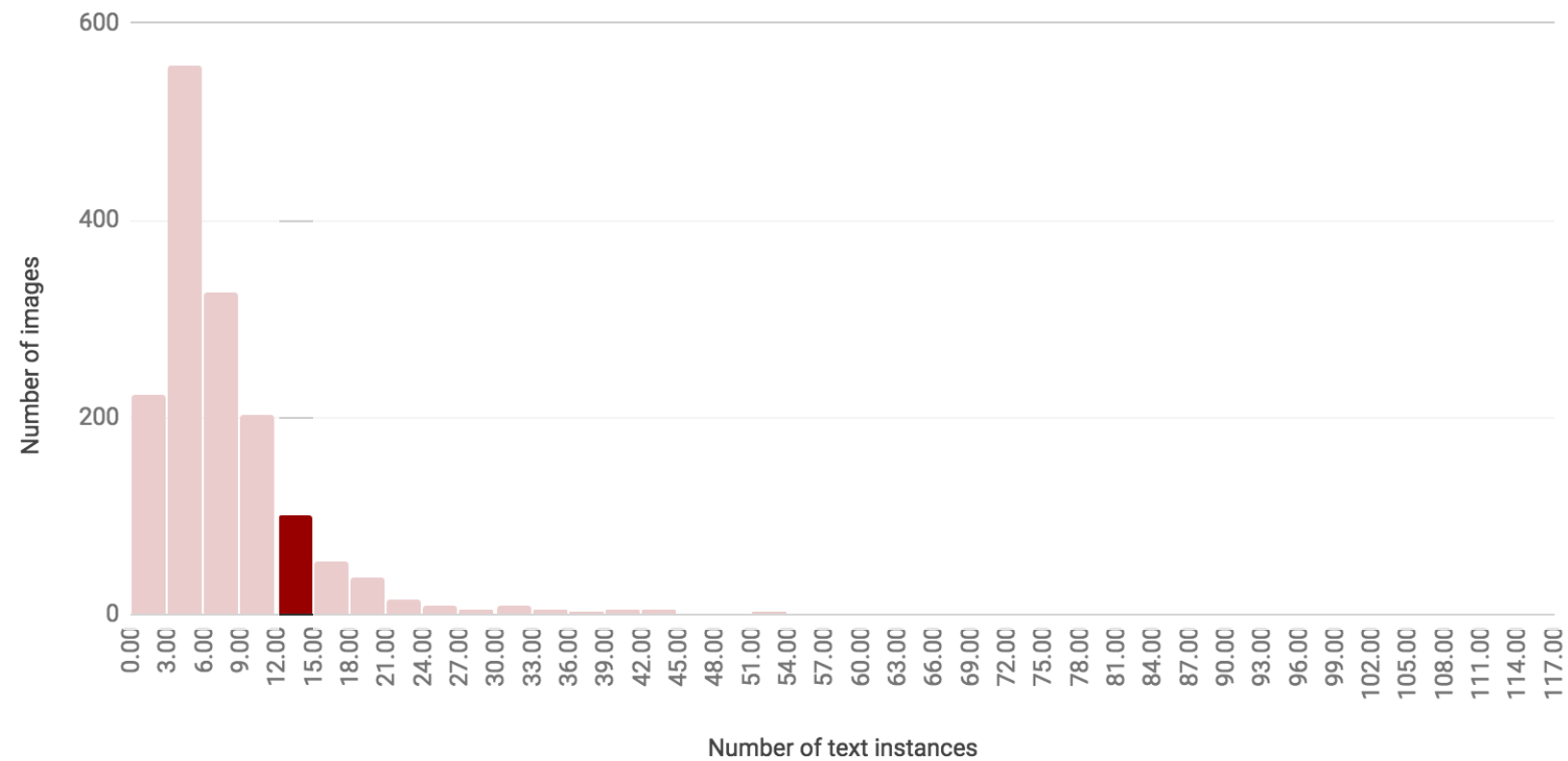
Distribution of Text Instances Across All Images of Total-Text





# Dataset Statistics

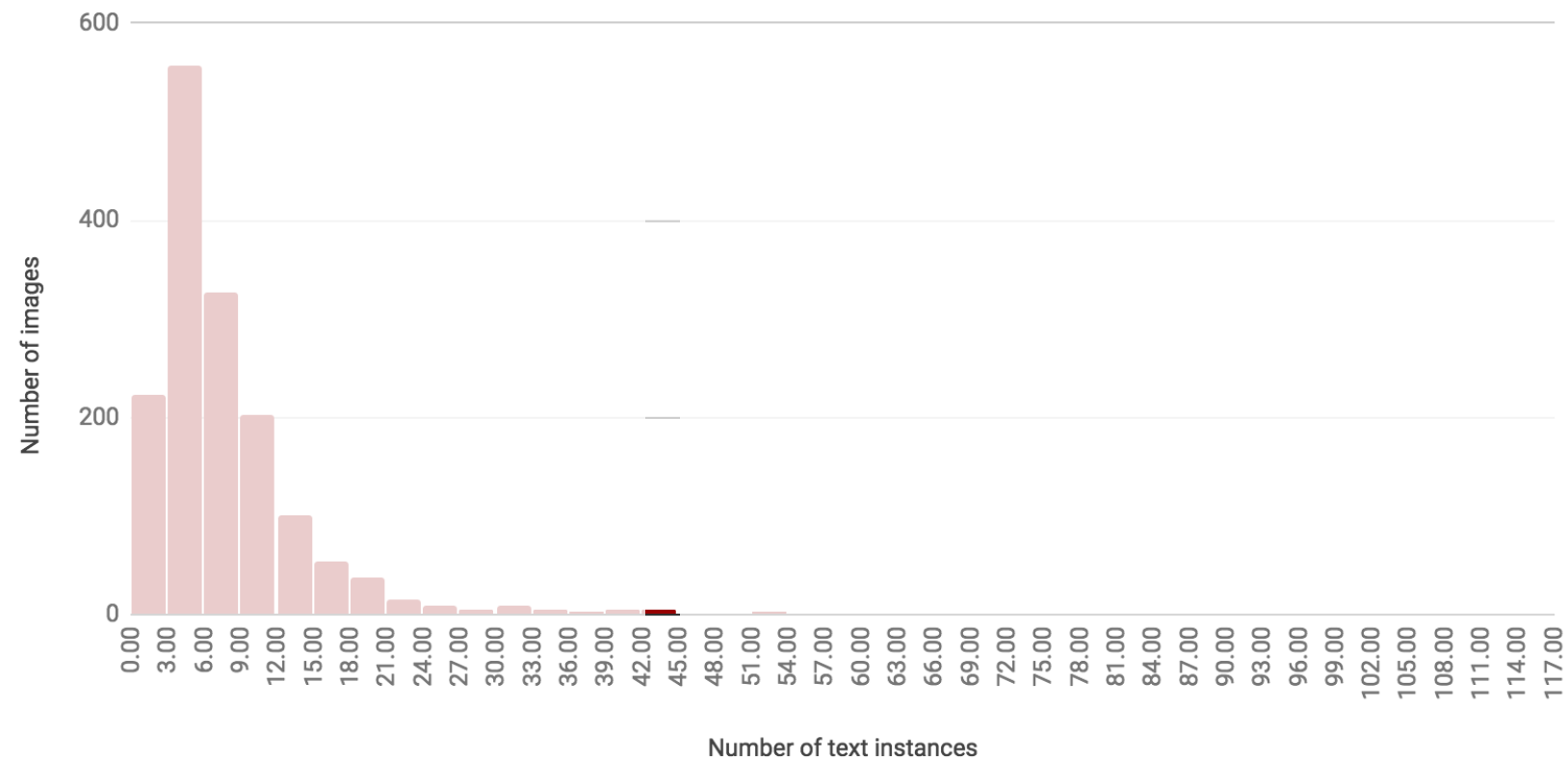
Distribution of Text Instances Across All Images of Total-Text





# Dataset Statistics

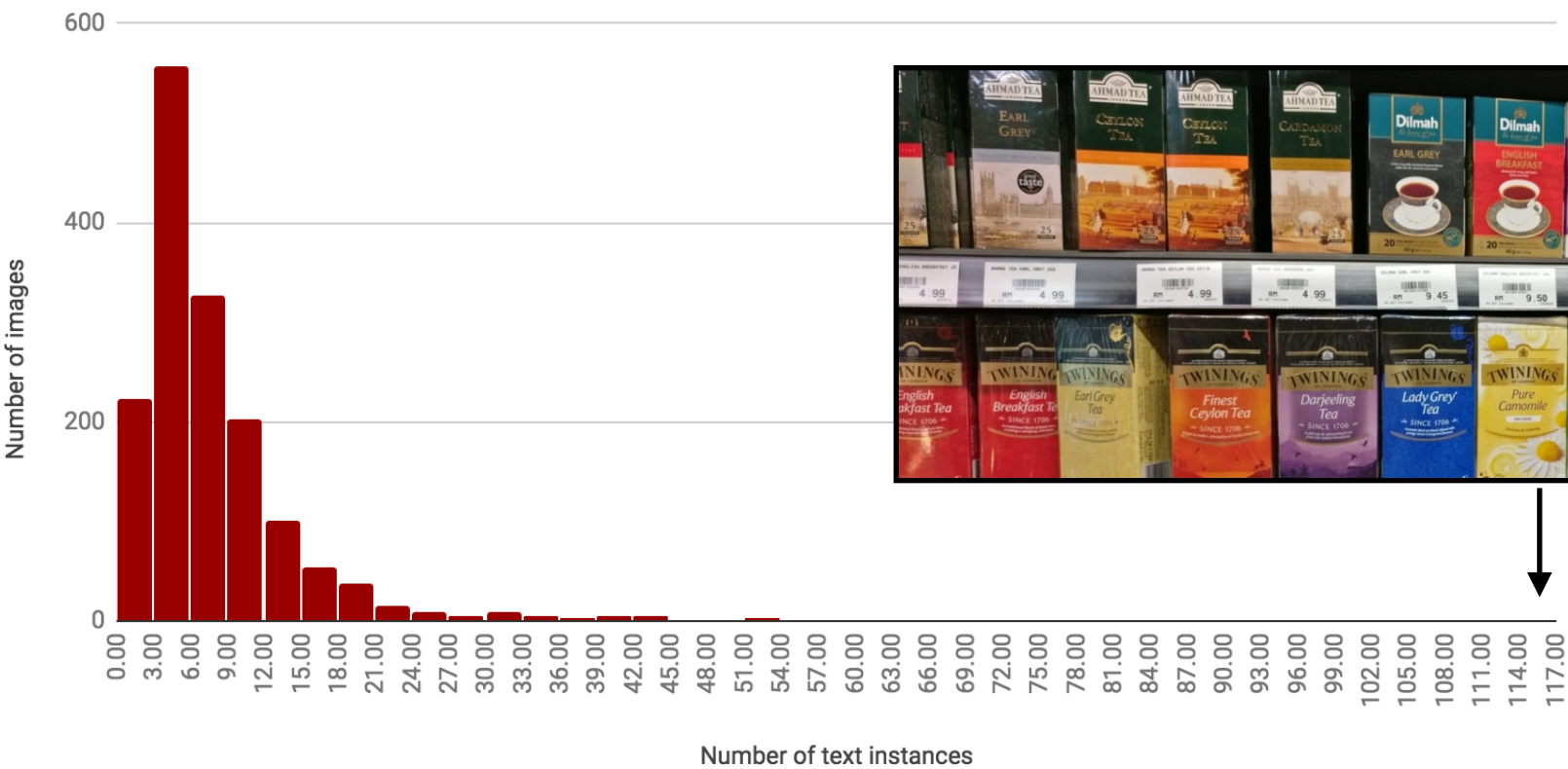
Distribution of Text Instances Across All Images of Total-Text





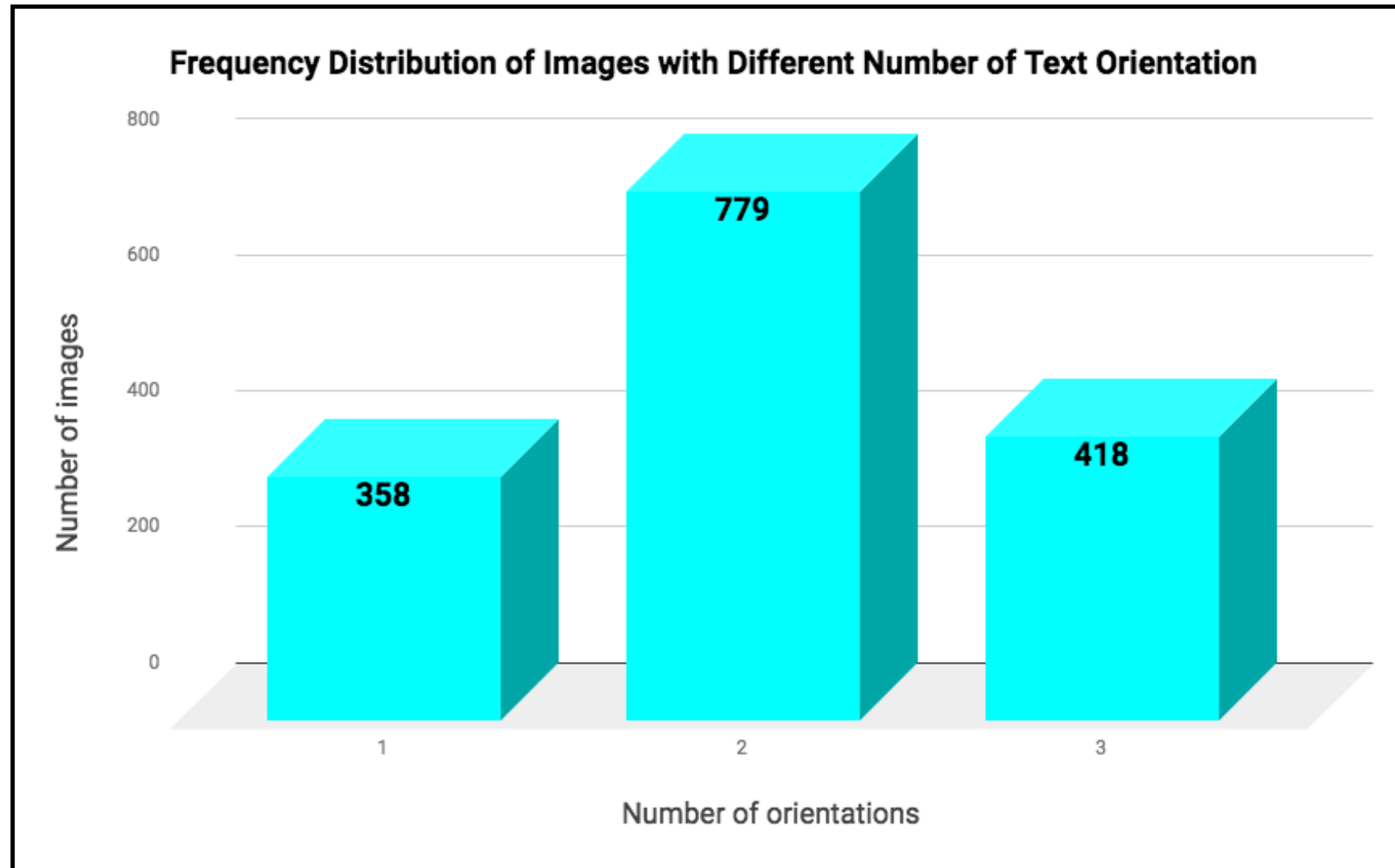
# Dataset Statistics

Distribution of Text Instances Across All Images of Total-Text



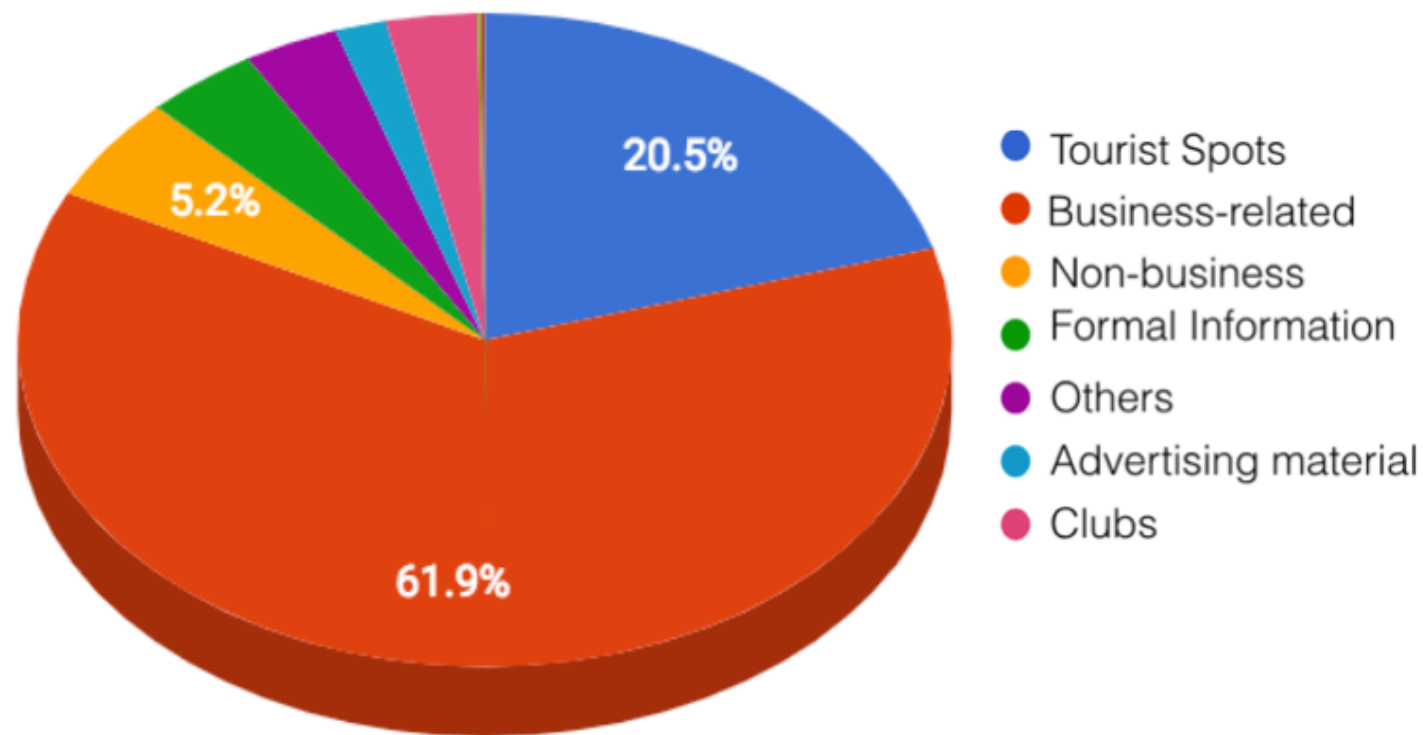


# Dataset Statistics





# Dataset Statistics - *Where do we see curve text?*

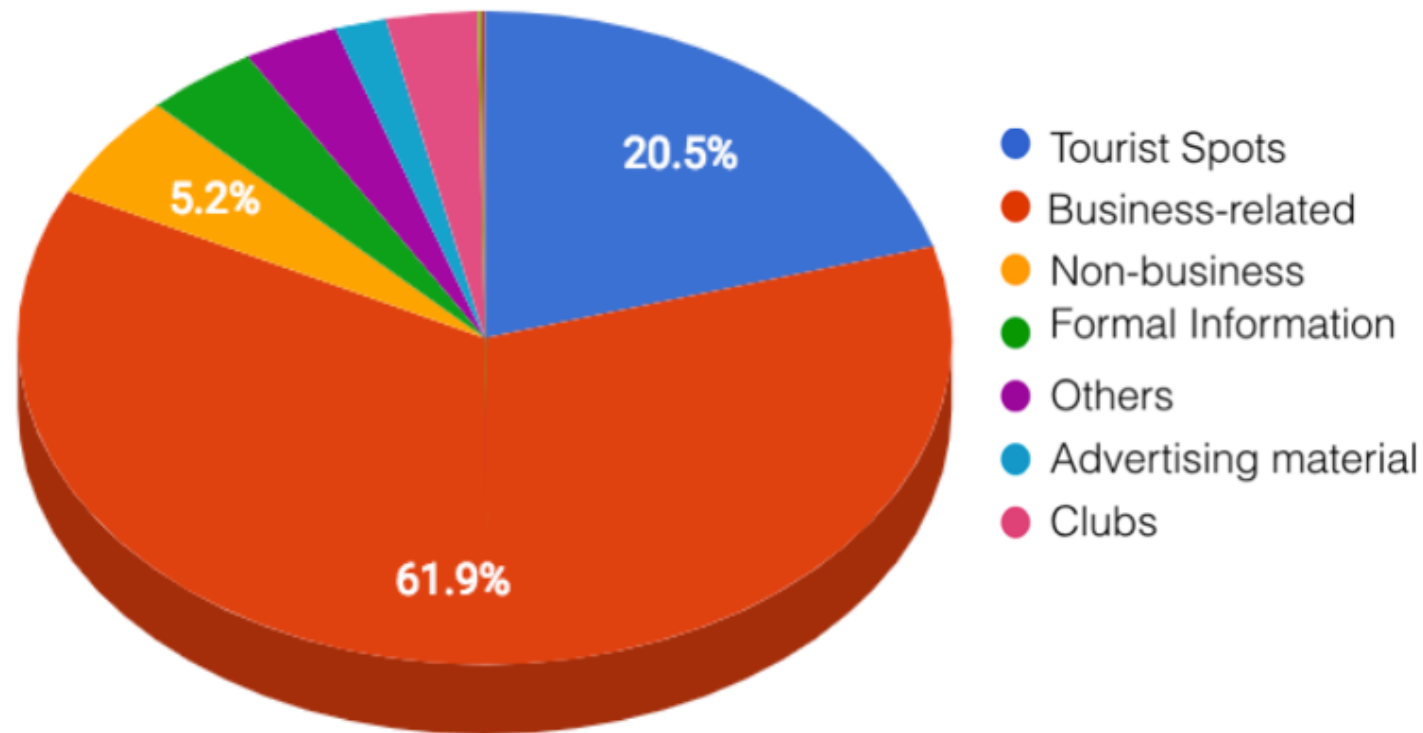


- 61.9% of them are business related.





# Dataset Statistics - *Where do we see curve text?*

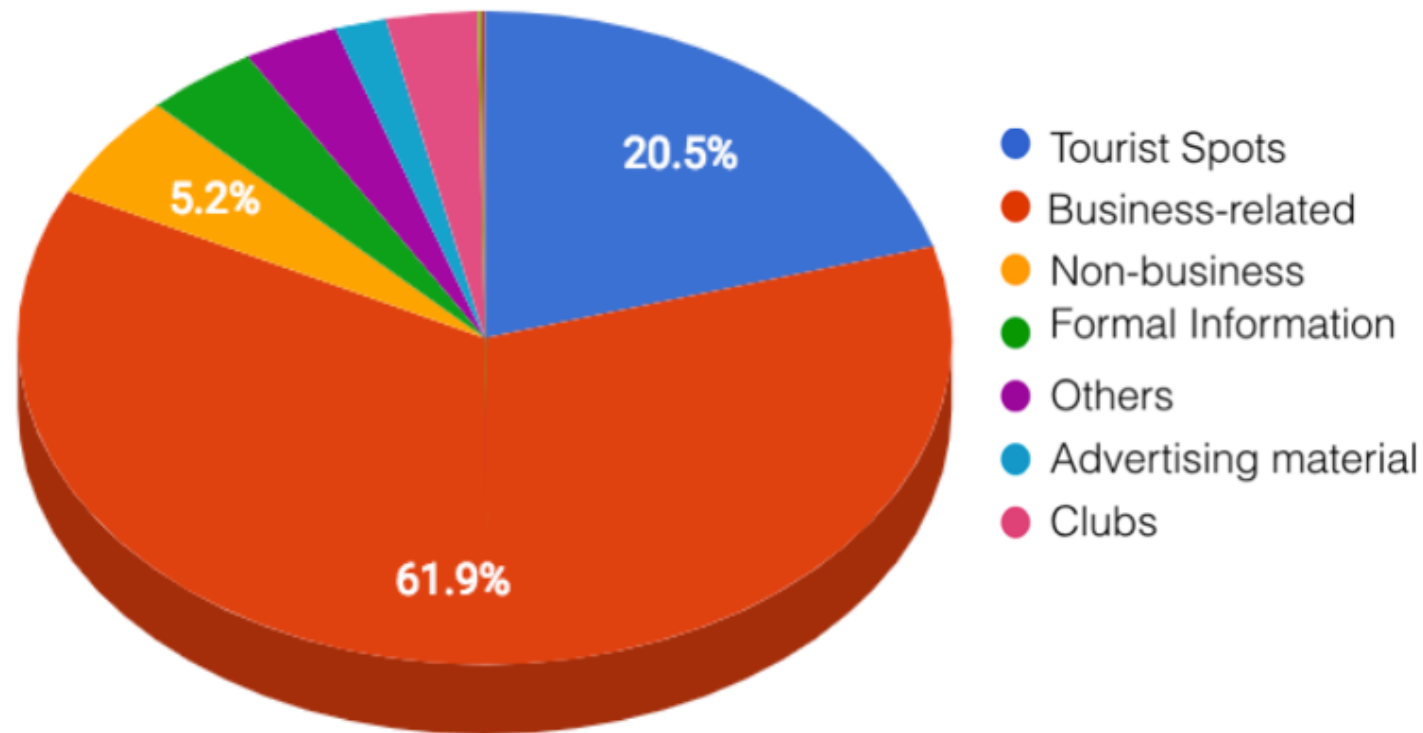


- 20.5% of them appeared in tourist spots.





# Dataset Statistics - *Where do we see curve text?*



- The other 17.5% consists of formal information, club logos, and others.





# Challenging Sceneries



*Images with low contrast and complex background*



# Challenging Sceneries



*Images with small font size*



# Challenging Sceneries



*Images with perspective distortion*



# Challenging Sceneries



*Images with uneven illumination*



# Annotation





# Annotation



- Conventional axis-aligned bounding box doesn't fit curved text well.





- Conventional axis-aligned bounding box doesn't fit curved text well.
- We propose polygon-shaped groundtruth, it fits text instances with all orientation variety.



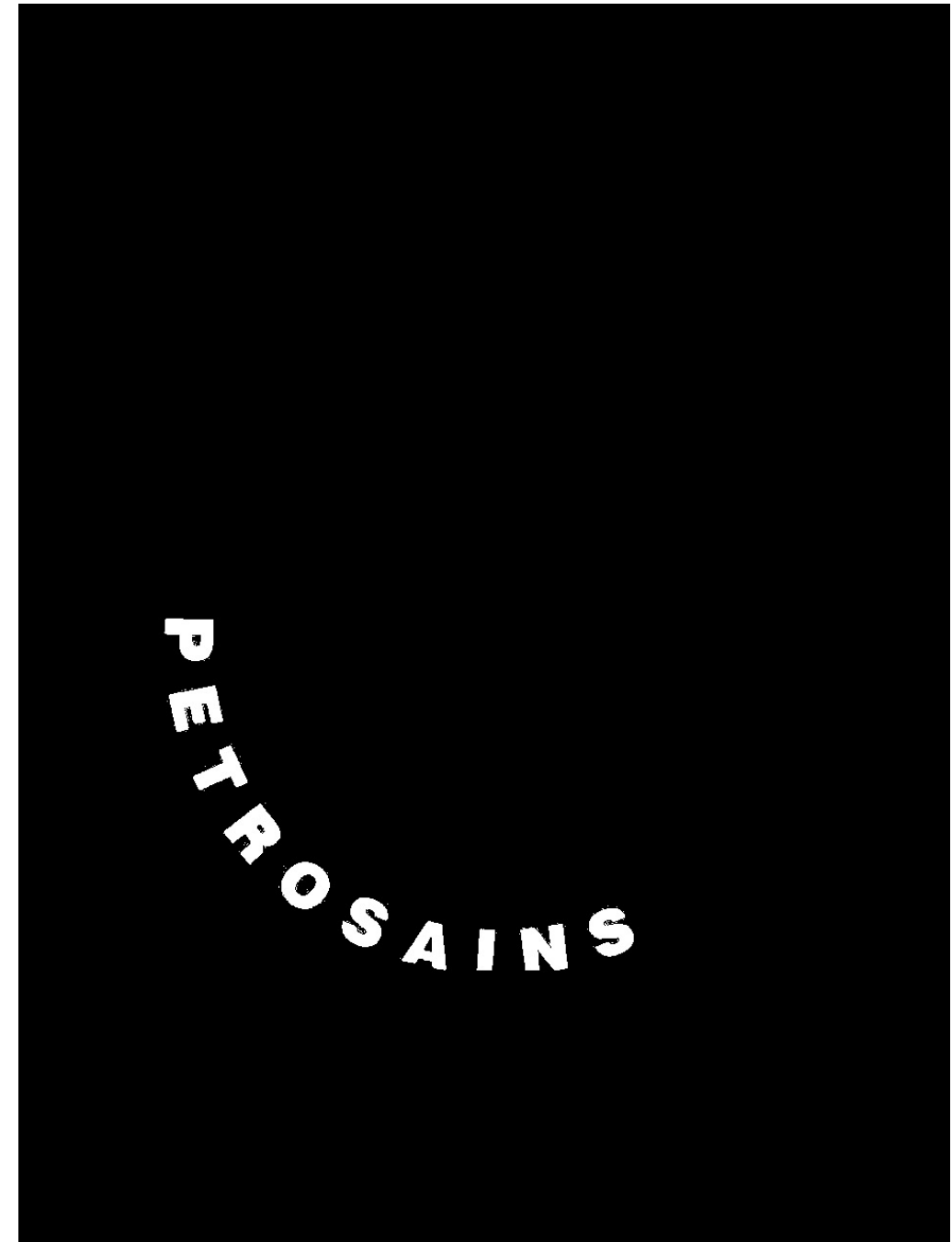
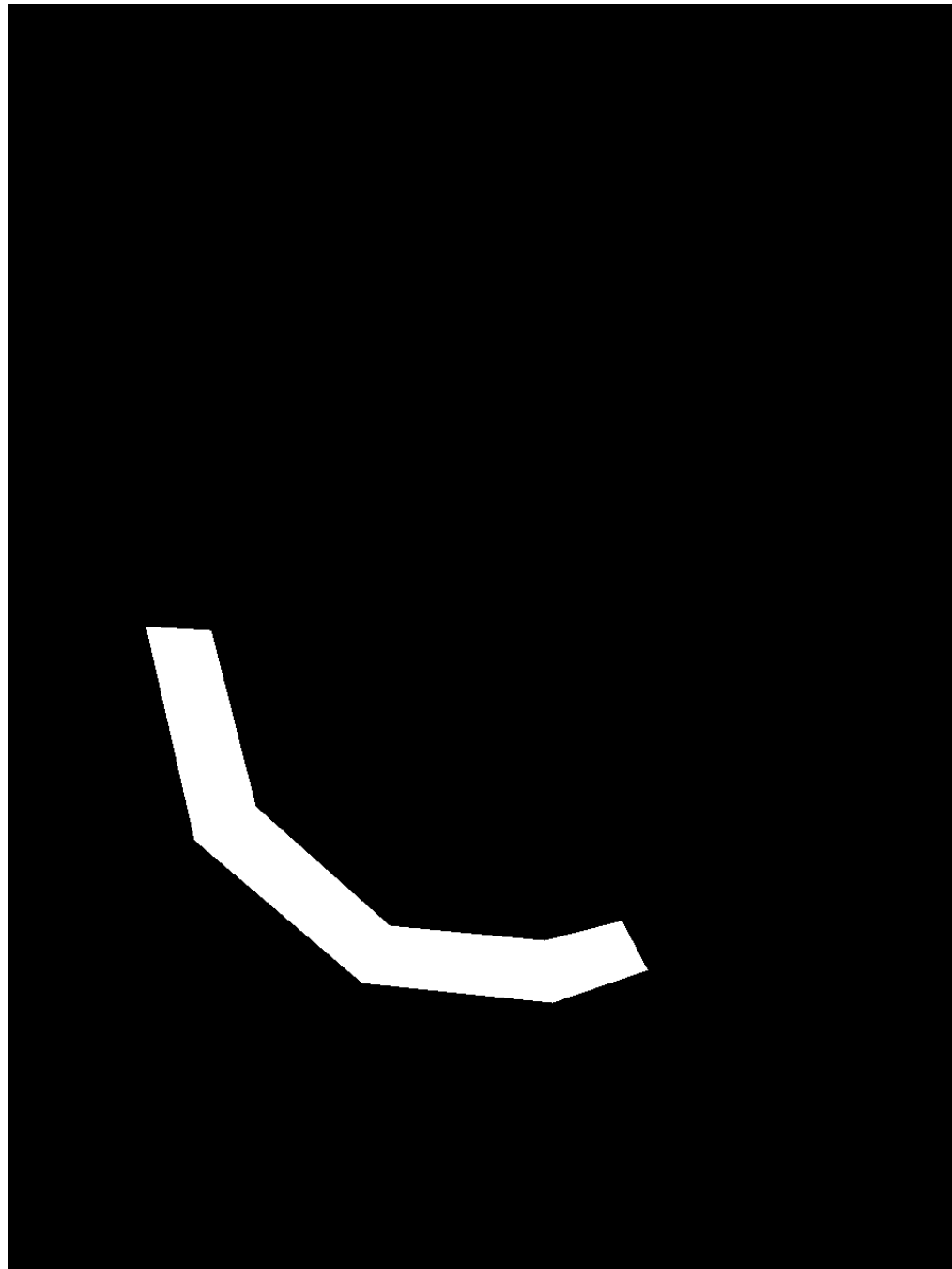
# Annotation



- X-coordinates:  
[206, 251, 386, 542, 620, 646, 550, 358, 189, 140]
- Y-coordinates:  
[633, 811, 931, 946, 926, 976, 1009, 989, 845, 629]
- Transcription:  
['PETROSAINS']
- Orientation:  
['Curve']

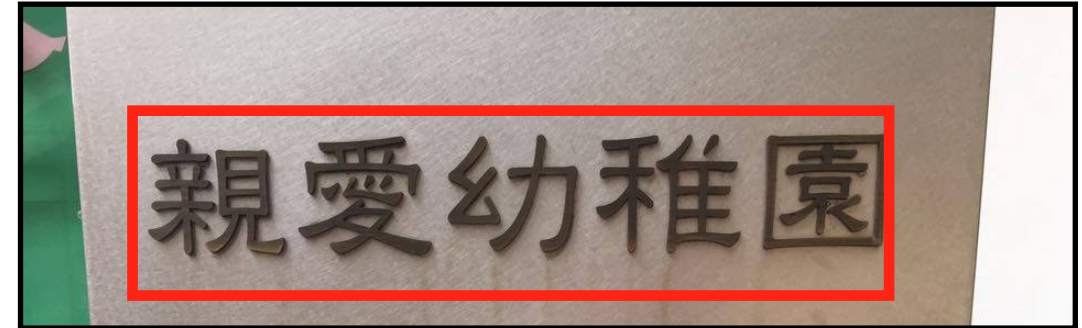


# Annotation



# Annotation

- Do not care regions:
  - Non-latin scripts





# Annotation

- Do not care regions:
  - Non-latin scripts
  - Signs and symbols



# Annotation

- Do not care regions:
  - Non-latin scripts
  - Signs and symbols
  - Illegible text by annotators





# Evaluation Protocol

- Total-Text uses the same evaluation protocol as ICDAR2013, namely *DetEval*.
- Modification was done to handle the **overlapping** between detection bounding boxes and groundtruth bounding boxes.



# Conclusion

- Total-Text is a scene text dataset that features three major text orientations.
- Polygon-shaped groundtruth is used to fit every text instances tightly.
- Annotation includes spatial location(word level), transcription, and pixel level.



