Synthetic Map Generation to Provide Unlimited Training Data for Historical Map Text Detection

Zekun Li[1], Runyu Guan[2], Qianmu Yu[2], Yao-Yi Chiang[1], Craig A. Knoblock[2]

[1] University of Minnesota
[2] University of Southern California

Motivation

- **Automatically reading text** from map images could speed up map interpretation
- Many existing text detection models are trained with data from **different domains** (e.g. ICDAR scene images)
- Gathering **training data** for historical maps is important, while manual annotation takes a lot of time and effort

Sample images from ICDAR 15 dataset

Sample image patches from David Rumsey historical map collection

Related Work

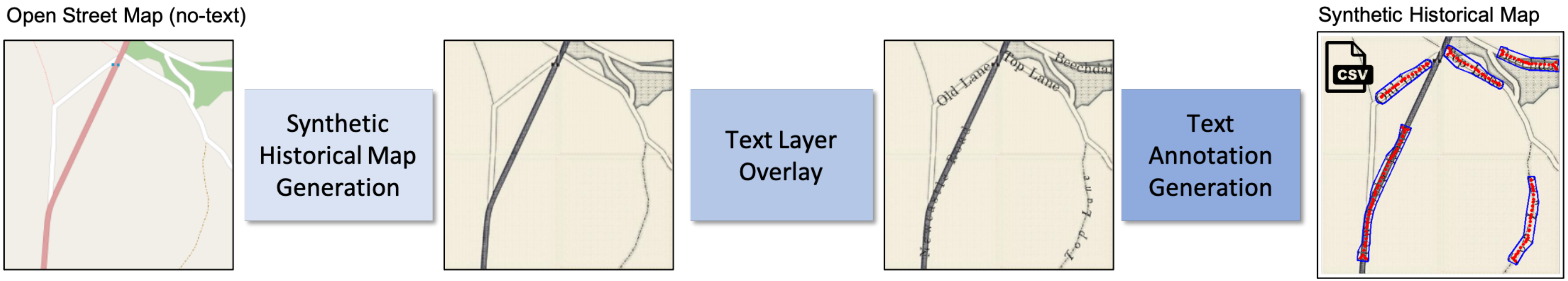
- Text Detection Datasets
  - International Conference on Document Analysis and Recognition **(ICDAR)**
    - Has been releasing text detection datasets since 2011
    - Before 2015: focused scene images and born-digital images [Karatzas et al. (2013)]
    - After 2015: incidental scene images [Karatzas et al. (2015)][Nayef et al. (2019)]
  - **COCO-Text** [Veit et al. (2016)]
    - Machine-printed and hand-written text, multi-lingual
  - **SCUT-CTW1500** [Liu et al. (2019)]& **Total Text** [Ch’ng et al. (2019)] :
    - Contain curved text instances
  - **David Rumsey** Map Collection [Weinman et al. (2019)]:
    - Historical map images
    - Dataset that we use for evaluation in this paper

Related Work

- Synthetic Data Generation
  - **SynthText** [Gupta et al. (2016)]
    - Large scale dataset with 800K images
    - Segmentation-based method for text label placement
  - **UnrealText** [Long et al. (2020)]
    - 600K images with 12 million word instances
    - Utilize Unreal 3D graphics engine for placing text on object surfaces

Approach

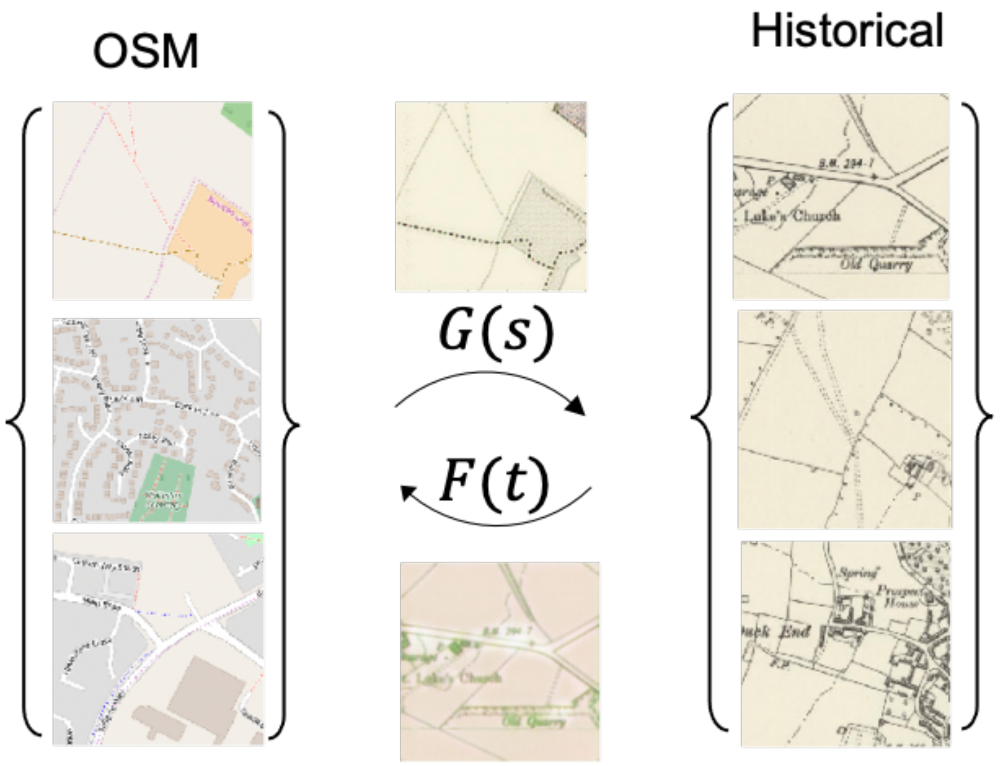
- We propose to generate **synthetic historical** maps images to aid the training of text detection models
- General Idea
  - Generate **synthetic map background** without any text labels
  - Automatically **place text labels** and generate ground-truth **annotation**



Synthetic Map Generation

- Public map sources such as OpenStreetMap (OSM) contain **separable** map layers: rasterized map **background** layer and **text** layer
  - Use clean (no text) OSM map background to produce style-transferred historical map
- Generative Adversarial Networks (GAN) can efficiently perform **style transfer**
  - CycleGAN model for the map style transfer – this can be replaced with other models as well
- Target map style
  - **Ordnance Survey** 6-inch map during year 1888-1913

Synthetic Map Generation - CycleGAN



Cycle Consistency Loss:

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

Adversarial Loss:

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$

Sample Synthetic Map Images

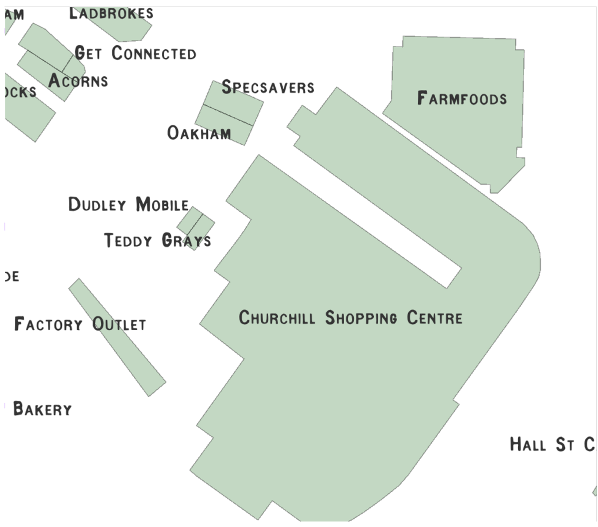
input OSM map tiles

output synthetic historical map tiles

Text Layer Overlay

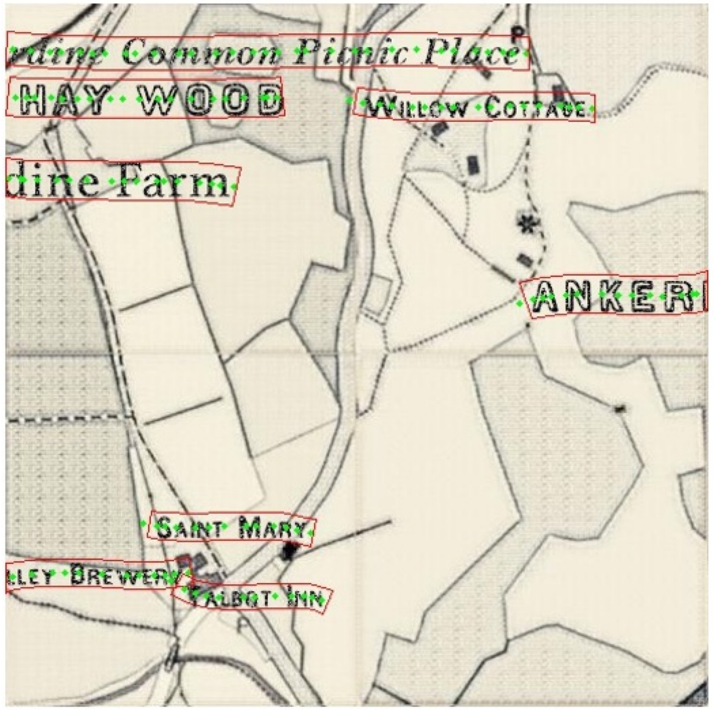
- Three types of geometries (QGIS PAL API for automatic label placement)
  - **Polygon**: text label on the polygon
  - **Line**: text label on the line
  - **Point**: text label near the point
- Font style and size
  - Roughly three levels of **font sizes**: large, medium and small
  - Collected 16 **historical** font styles
  - Same type of geo features have the **same** font size and style

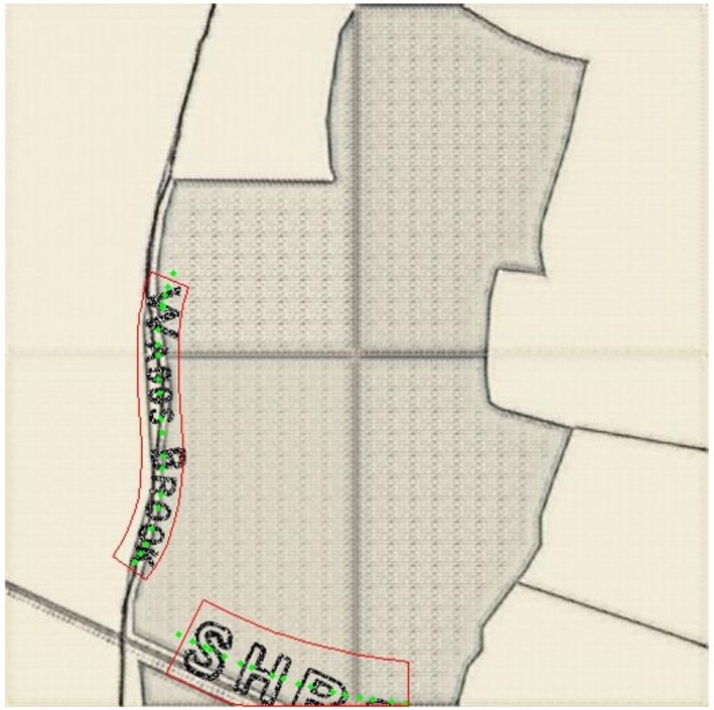
Groups	Size (pt)	Geo Features
Large	[60,80]	canal, city, county, town, village waterfall, wetland, island
Med.	[35,45]	airfield, airport, allotment, archaeological battlefield, camp site, cliff, dock, farmland farm, forest, fort, hamlet, nature reserve reservoir, ruins, vineyard, rail river, stream
Small	[20,30]	others (e.g. streets)



Text Annotation Generation

- We provide following annotation information
  - Bounding polygon for each text region
  - Centerline across the text region
  - Local height (or diameter) of the text region

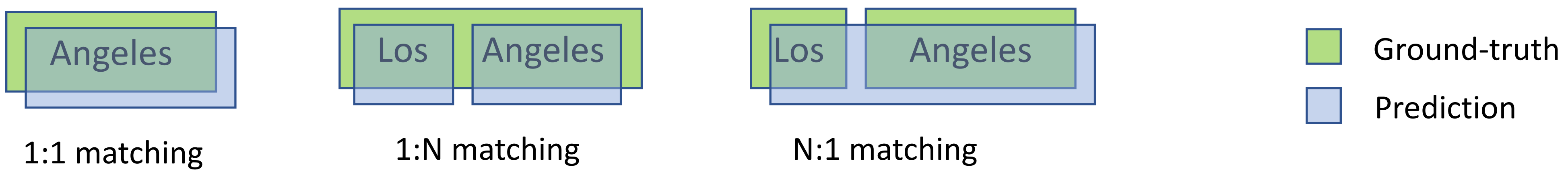




*Details for automatically computing the annotation info can be found in the paper

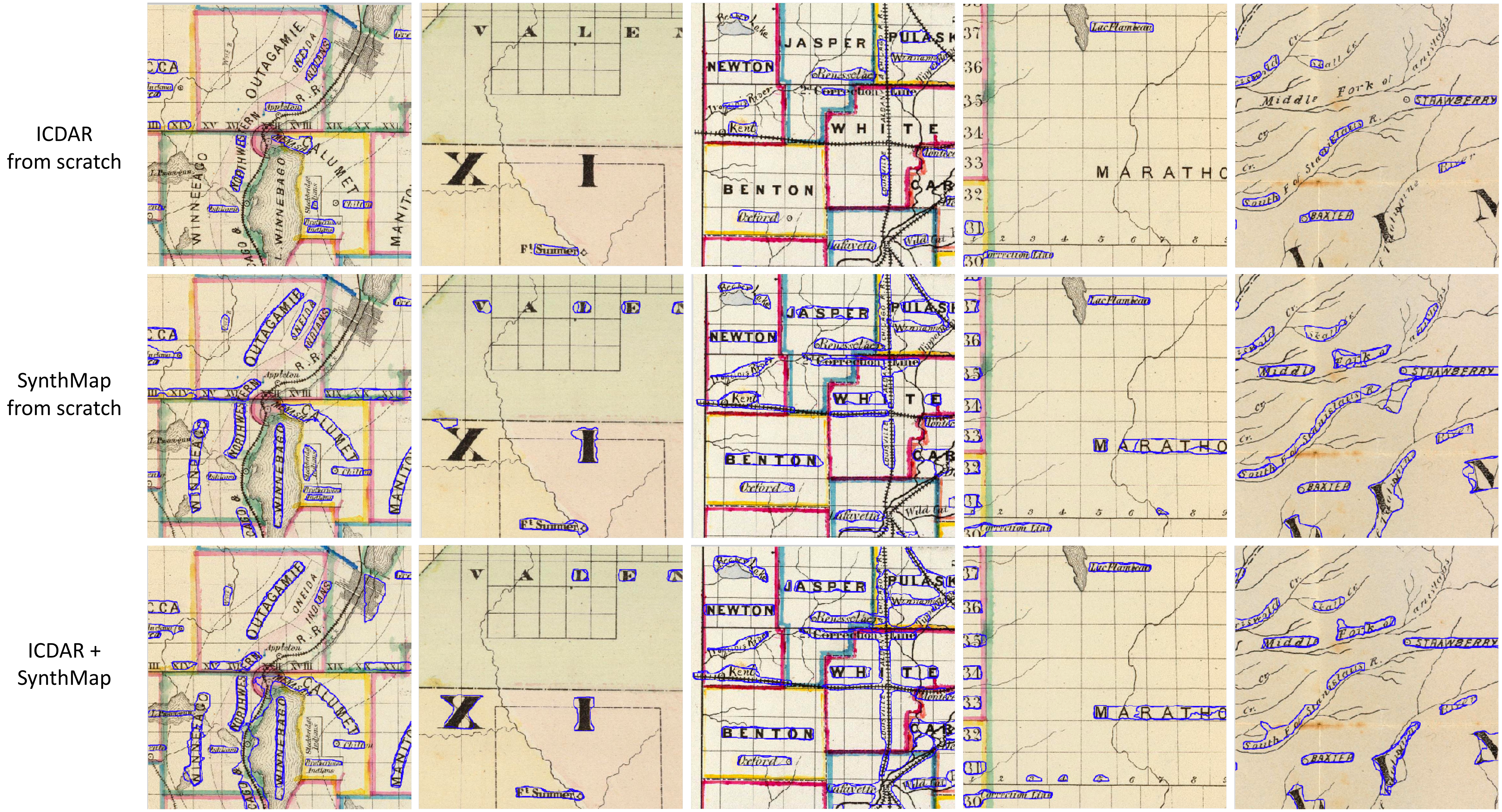
Experiment Setting

- Text Detection Model: PSE-Net [Wang et al. (2019)]
- Training Strategies (notice that we **did not** use any **real map** data for training)
  - **ICDAR:** train from scratch on ICDAR 2015 dataset only
  - **SynthMap:** train from scratch on our SynthMap dataset only
  - **ICDAR + SynthMap:** pretrain on ICDAR 2015, fine tune on SynthMap dataset
- Evaluation Metrics [Wolf et al. (2006)]
  - Considers 1:1 matching, 1:N matching and N:1 matching of GT and prediction



Experiment Results

	ICDAR2015			SynthMap			ICDAR + SynthMap		
	prec.	recall	F1	prec.	recall	F1	prec.	recall	F1
D0006	84.30%	44.80%	58.50%	68.90%	25.30%	37.00%	79.60%	27.70%	41.10%
D0017	81.10%	49.30%	61.30%	85.30%	63.40%	72.70%	88.90%	60.80%	72.20%
D0041	71.90%	48.90%	58.20%	71.70%	70.80%	71.20%	74.90%	72.35%	73.60%
D0042	81.28%	34.18%	47.75%	75.88%	48.65%	58.83%	77.86%	55.75%	64.39%
D0079	45.30%	4.20%	7.70%	40.50%	20.60%	27.30%	31.30%	13.60%	19.00%
D0089	83.10%	49.90%	62.40%	75.90%	44.60%	56.20%	69.30%	48.10%	56.80%
D0090	82.80%	55.80%	66.70%	90.60%	63.40%	74.60%	91.00%	70.30%	79.30%
D0117	89.75%	56.13%	68.90%	72.72%	55.55%	62.95%	82.40%	55.12%	66.02%
D5005	88.55%	57.68%	69.57%	78.38%	54.60%	64.23%	82.55%	57.95%	67.87%
All	**82.76%**	**45.00%**	**57.32%**	**74.90%**	**51.73%**	**60.62%**	**78.51%**	**55.25%**	**64.25%**


ICDAR from scratch
SynthMap from scratch
ICDAR + SynthMap

Conclusion and Future Work

- We proposed a pipeline to generate unlimited amount of training data for map text detection algorithms
- We use style transfer model to convert OSM maps to historical style
- Text labels are placed according to the underlying geo feature locations
- This approach can be extended to generate synthetic data for word linking and road delineation

Demo and Code

- Live Demo

https://zekun-li.github.io/side-by-side/

- Code released

https://github.com/zekun-li/generate_synthetic_historical_maps