Generating Historical Maps from Online Maps

Zekun Li
University of Southern California
zekunl@usc.edu

ABSTRACT

This paper proposes an automatic system to generate a large amount of data for the training of text detection systems for historical maps. The system takes online maps as input and learns a *conditional* GAN model, to generate realistic historical map images from existing geographic datasets. Then the system uses the generated images as the base map and inserts synthetic text. Since the system has the control of text content, font style, and location, the system can obtain ground truth information (minimum bounding boxes) of the synthetic text. To overcome the challenge of content mismatch, the proposed system uses a novel loss function to encourage the generation of historical cartographic symbols in the foreground areas and discourage the generation in the background. The final output is a set of images resembling historical maps and the minimum bounding boxes around text regions on the images as annotations.

CCS CONCEPTS

• **Information systems → Geographic information systems**.

KEYWORDS

Generative Adversarial Networks, Historical Map Processing

1 INTRODUCTION

Historical map is an important data source for understanding city evolution and human activities. Many machine learning algorithms have been developed to extract valuable information from historical maps [1, 4]. One problem that hinders the automatic data extraction from the historical maps is the lack of training data. Existing approaches either manually annotate a set of data or use data from other domains for training. The first method is time-consuming and labor-intensive, while the second method usually suffers from the issue of domain mismatch. For example, if we want to train a text detection system for historical maps and use the real-life images for training, it is often the case that the system fails to detect single characters. The reason is that real-life images rarely contain single-character, as most of them are words.

Generative adversarial networks (GAN) [3] can perform style transfer to help generate synthetic data in another data domain. The major challenge in generating historical map images from existing geographic datasets is the content mismatch. For most of the existing GAN models, such as Pix2Pix [3] and CycleGAN [6], the object outlines in the training data for both styles should be

(a)
(b)
(c)


Figure 1: Comparison of open street map (a), historical map from National Library of Scotland (b). In (c), white region is the foreground and gray region is the background.

about the same. While in our case, the outline for modern maps and historical maps varies significantly. The variation is majorly due to the differences in the level of details (i.e., data generalization). Lines and shapes that do not appear on online maps sometimes appear in historical maps. Another reason for the content mismatch is the occurrence of human activities during the years: people build new roads and structures while old ones become abandoned. For example, Figures 1 (a) and (b) show the differences between OpenStreetMap (OSM) and historical maps from Ordnance Survey maps depicting the same location but from a different time.

2 METHODOLOGY

We build a conditional generative adversarial network, based on Pix2Pix [3], to generate synthetic historical map images. We jointly train a discriminator and a generator that take OMS maps as input and produce the corresponding maps in historical style. The generator is responsible for the synthetic image generation, and the discriminator is responsible for differentiating the real image patches from the generated patches. The novel contribution is that different from Pix2Pix, which generates the whole image at once, the proposed network separates the image generation step into two pieces: **foreground generation** and **background generation**. The two-piece process ensures the resulting image to have a clean background while still have strokes on the foreground.

2.1 Foreground and Background Separation

One important step in pre-processing is the the separation of foreground and background for OSM images. We first divide the input OSM image x into small tiles of size $S \times S$ pixels, and suppose there are $M \times N$ tiles. Let *Mask* be an indicator function that applies on each tile $B_{i,j}$ where $i = \{1 \dots M\}$ and $j = \{1 \dots N\}$. (Note that B_{ij} is of the dimension $S \times S$ pixels.) The *Mask* can be defined as following.

$$Mask(B_{i,j}) = \begin{cases} \mathbf{I}_{S\times S} & |max(B_{ij}) - min(B_{ij})| \leq \delta \\ \mathbf{0}_{S\times S} & otherwise \end{cases} \tag{1}$$

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGSPATIAL '19, November 5–8, 2019, Chicago, IL, USA
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6909-1/19/11.
https://doi.org/10.1145/3347146.3363463

where $max(B_{ij})$ extracts the maximum color intensity within tile B_{ij}, and similarly for $min(B_{ij})$. δ is a small value in range 0-255 and we used 5 in the experiments. Define $Mask(x)$ as the concatenation of $Mask(B_{i,j})$ for tiles B_{ij} that are inside x.

$$Mask(x) = Concat_{\forall (i \in \{1 \ldots M\}, j \in \{1 \ldots N\})}(Mask_(Bi, j)) \tag{2}$$

Figure 1(c) shows the computed mask. For the foreground regions, the generated content should be similar to historical map, while for the background regions, the generated content should be close to the historical map background color. Then it comes to the question: what should be the background color?

Although all the Ordnance Survey historical maps have a yellowish tone, the shade could vary a lot due to the lighting conditions. To determine the background color, we obtain the intensity value that appears most frequently for each channel and take that value as the mode intensity. When combining the color intensity from all three channels, we could obtain a color value that most of the pixels look like in the map (i.e., the background color).

2.2 Historical Map Generation

We construct a conditional-GAN model that is composed of a generator $G(x, z; \theta_g)$ and a discriminator $D(x, u; \theta_d)$ similar to Pix2Pix [3], where x is the input feature vector, z is a random noise and u could be either real data y or generated data from G. The generator tries to learn a mapping of $(x, z) \rightarrow y$ and discriminator tries to distinguish y from $G(x, z)$ with the knowledge of x. Formally, the objective is to find θ_g and θ_d that optimizes

$$\begin{aligned} \min_G \max_D \mathcal{L}(D, G) = & \mathbf{E}_{x,y}[logD(x, y)] \\ & + \mathbf{E}_{x,z}[log(1 - D(x, G(x, z)))] \end{aligned} \tag{3}$$

To generate realistic images from G that fools the discriminator D, we also need to define a loss to minimize the the distance from y and $G(x, z)$ in $L1$ distance space.

Since we separate the generation of foreground and background, we need to have two loss functions. The loss for foreground is defined as

$$\begin{aligned} \mathcal{L}_{L1-fg} &= \mathbf{E}_{x,y,z}[||y * M - G(x, z) * M||_1] \\ &= \mathbf{E}_{x,y,z}[||(y - G(x, z)) * M||_1] \end{aligned} \tag{4}$$

and the loss for background is defined as

$$\begin{aligned} \mathcal{L}_{L1-bg} &= \mathbf{E}_{x,y,z}[||C * (1 - M) - G(x, z) * (1 - M)||_1] \\ &= \mathbf{E}_{x,y,z}[||(C - G(x, z)) * (1 - M)||_1] \end{aligned} \tag{5}$$

where y is the foreground region of the historical map and C is the calculated background as in Figure 1(c).

2.3 Putting Text on the Map

Kang et al. [5] have shown that text can not be synthesized correctly through GAN models. Thus we need to add text on the synthetic map image as post-processing. We use 14 historical font styles from FontSpace [2] and use the text corpus of geo-locations provided by the National Library of Scotland (NLS) to generate locations instead of random meaningless words. In terms of the text morphosis, we generate horizontal, vertical, rotated, and curved poses.

3 PRELIMINARY RESULTS

To obtain training data for our conditional-GAN model, we downloaded historical map tiles from the Ordnance Survey website with

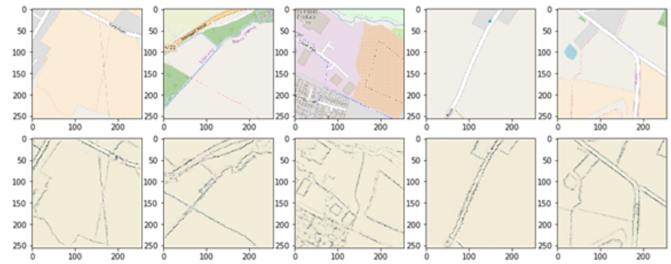

Figure 2: Visualization of our conditional-GAN model results. The First row is the model input (OSM map tiles) and the second row is the model output (historical map tiles).

the zoom-in level 16. Then we find corresponding regions on OpenStreetMap (OSM). The map tiles are all of size 256x256 pixels. After the model was trained to converge, we applied the model on the test set, which took only OSM map tiles as input and generated Ordnance-Survey-style historical maps. Figure 2 shows some sample outputs generated by our model. Afterward, we concatenated neighboring four tiles to produce map images of size 512x512.

With the above process, we finally obtained 7,482 base images. For each image, we generated 10 (image, text) pairs which basically added 10 groups of various text on one base image. For each sample, we generated a `.txt` file describing inserted text information: $(content, x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, x_c, y_c, sin, cos, w, h)$. In the end, we had a dataset with **75,000** samples, which can be used for training deep-learning based text detection and recognition system. For future work, we will use OSM text label locations instead of randomly generated locations to place the text content.

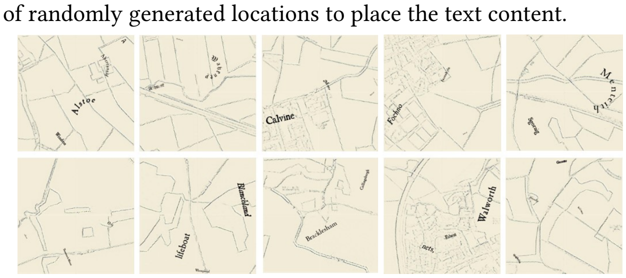

Figure 3: Sample images from the created dataset.

ACKNOWLEDGEMENT

This material is based on research supported in part by the National Science Foundation under Grant No. IIS 1563933 (to the University of Colorado at Boulder) and IIS 1564164 (to the university of Southern California), and in part by Microsoft and NVIDIA Corporation.

REFERENCES

[1] Y.-Y. Chiang and C. A. Knoblock. 2011. Recognition of multi-oriented, multi-sized, and curved text. In *ICDAR*. IEEE, 1399–1403.

[2] [n. d.] Font space. https://www.fontspace.com/. Accessed: 2019-07-30. ().

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*, 1125–1134.

[4] J. H Uhl, S. Leyk, Y.-Y. Chiang, W. Duan, and C. A Knoblock. 2018. Spatialising uncertainty in image segmentation using weakly supervised convolutional neural networks. *IET Image Processing*, 12, 11, 2084–2091.

[5] S. Gao Y. Kang and R.E. Roth. 2019. Transferring multiscale map styles using generative adversarial networks. *International Journal of Cartography*, 1–27.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2223–2232.