Adversarial Generation of Historical Maps

Internship Defence

Dina El Zein, supervised by Loic Landrieu & Mathieu Aubry

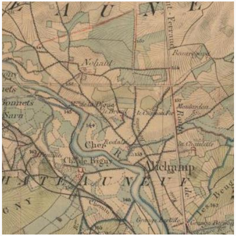
October 10, 2022

Motivation

Map Styles [1]

(a) Orthophotography

(b) Cassini XVII



(c) Etat Major ∼1840s

(d) Scan50 1950

(e) Plan IGN

Source: https://www.geoportail.gouv.fr

Motivation

Generating maps is important:

- **Disaster Relief:** legible maps from aerial images
  ➔ assists response after a disaster.
- **Cartographic Perspective:** switch between map styles
  ➔ better symbolic representation across scales.
- **Social Science Perspective:** critical analysis of cartographic choices
  ➔ informs us about semiotics from centuries ago (scale, symbols, density).
- **Historical Perspective:** deconstructing old maps
  ➔ allows for a better understanding of historical context

(a) Cassini XVII

(b) Plan IGN

Definition of the Problem

Plan IGN

(a) Orthophotography

(b) Plan IGN Map

Cassini XVII

(a) Orthophotography

(b) Cassini XVII Map

➔ Problem: low-resolution dataset

Translation: orthophotography $\mapsto$ map

(a) Orthophotography

(b) Cassini XVII Map

Create Dataset:

- 512 (2000 $\times$ 2000 pixels) aligned ortho-map data of using Geoportail[2]
- 375 aligned ortho-map data are left after cleaning

[2] https://www.geoportail.gouv.fr

Problem

Natural image style transfer does not work on maps:

- **Symbolic Representations:** discrete points, lines, or shaded areas
- **Scanning Artifacts:** crease, folds, different scanners and cartographers
- **Text:** a network can not predict named entities
- **Maps and photos are aligned (georeferencing):** not exploited by standard computer vision algorithms.

➔ Exploit the uniqueness of the problem to propose a better solution!

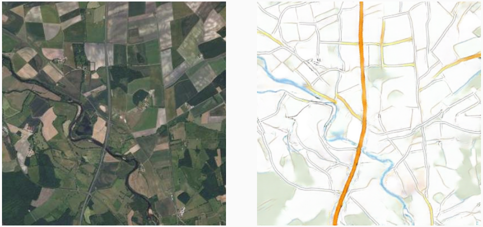
(a) Cassini XVII (b) Plan IGN (c) Cassini XVII

Contributions

We use CycleGAN as a basis for style transfer.

We modify it in two key ways:

- Handling Text: detection and masking strategy to prevent inconsistent supervisory signal.
- Alignment Supervision: exploit the natural alignment of maps/orthophotography.



(a) Orthophotography (b) Predicted Map

Text Detection

Related Work[3]

Text detection is hard on maps because:

- Diversity and Variability of Text on Maps: text font, size, and shape
- Complexity of Background: signs and grass can be indistinguishable from text
- Cursive writing: on Cassini XVII

(a) Cassini XVII

(b) Plan IGN

[3] Zhu *et al.* 2016 Frontiers of Computer Science, Lin *et al.* 2020 Archives of computational methods in engineering, Long *et al.* 2021 International Journal of Computer Vision

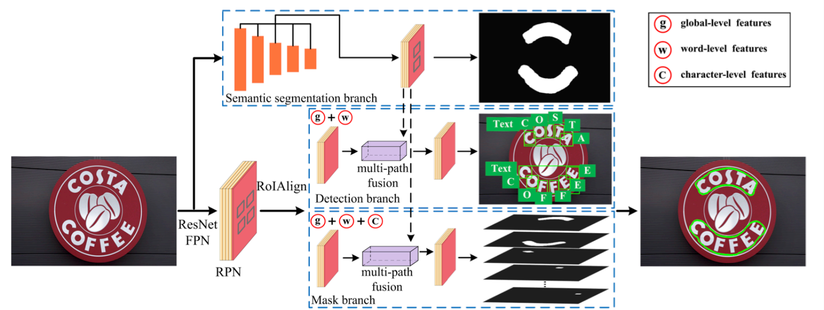


Figure 9: TextFuseNet Framework. Two-stage convolution-based architecture with multiple heads for character, word, and global level feature representations.

Image Source: Ye *et al.* 2020, IJCAI

TESTR: Text Spotting Transformers [5]

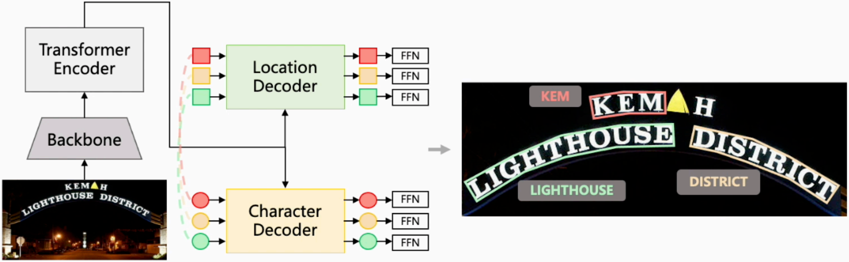


Figure 10: TESTR architecture. A single-encoder dual-decoder framework that jointly performs text detection and recognition.

Image Source: Zhang *et al.* 2022, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

Experimental Setup

Evaluation Setup: training process is divided into two stages:

- models pre-trained on synthetic data and fine-tuned on real-world images.

Cassini XVII Annotation: hand-annotated 5 images using VGG [6] for evaluation

Post-Processing Procedure: predicted text boxes are merged if:

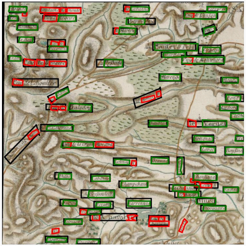
- the boxes are close and aligned.

(a) Before

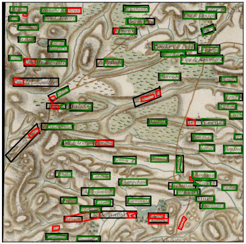
(b) After

[6] VGG Image Annotator Software, Dutta *et al.* 2016

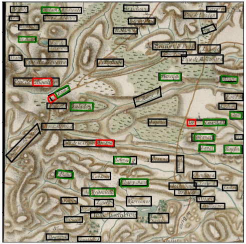
Qualitative Results



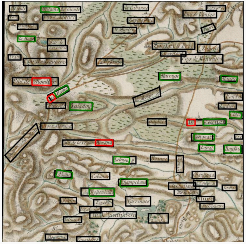
(a) TESTR



(b) TESTR + PostProc



(c) TexTFuseNet



(d) TexTFuseNet + PostProc

Figure 12: **annotations**, correct predictions, incorrect predictions

TESTR outperforms TextFuseNet

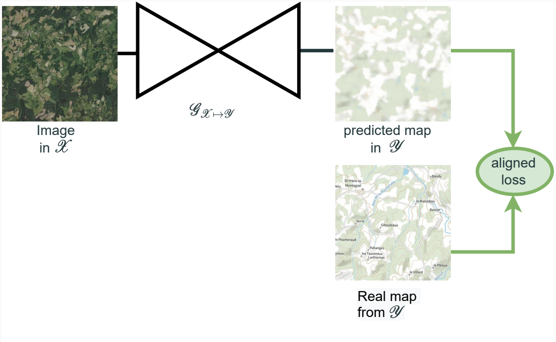
Table 1: **Text Detection Quantitative Results.** Comparison between TFN and TESTR on text detection for the Cassini dataset.

Model	Recall	Precision	F1-score
TFN	29	76	43
TFN + post proc.	30	**79**	43
TESTR	82	56	66
TESTR + post proc.	**86**	71	**78**

Recall is more important than Precision!

Adversarial Maps Generation

Encoder-Decoder Architecture

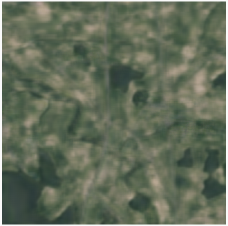


$$L_{\text{align}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}) = \mathop{\mathbb{E}}_{x,y\sim\mathcal{X}\times\mathcal{Y}}\left[\|\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x) - y\|_1\right]$$

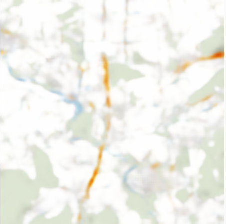
Encoder-Decoder Failure Result

(a) Image in $\mathcal{X}$

(b) Real map from $\mathcal{Y}$

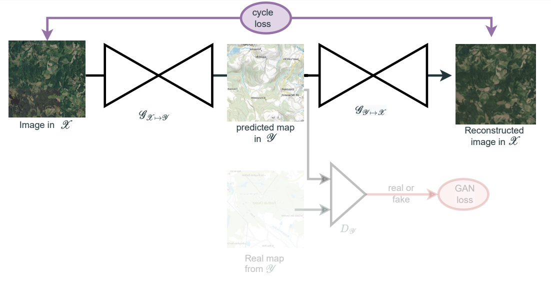


(c) Predicted image in $\mathcal{X}$



(d) Predicted map in $\mathcal{Y}$

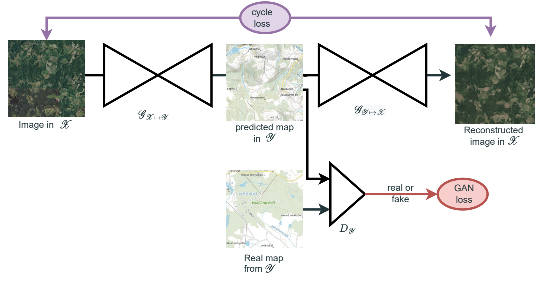
Failure: predicted images and maps are blurry; ambiguous/inconsistent supervision, loss induces a poor perceptual quality



Key Idea:

- use cycle consistency for aligned supervision

$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}\left[\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1\right].$$



Key Idea:

- use cycle consistency for aligned supervision

$$\mathcal{L}_{\mathrm{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}\left[\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1\right] .$$

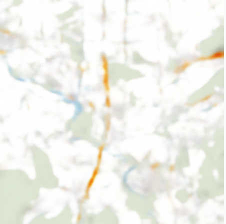
- use GAN loss to create good looking predictions

$$\mathcal{L}_{\mathrm{GAN}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, D_{\mathcal{Y}}) = \mathop{\mathbb{E}}_{y\sim\mathcal{Y}}\left[\log D_{\mathcal{Y}}(y)\right] + \mathop{\mathbb{E}}_{x\sim\mathcal{X}}\left[\log(1 - D_{\mathcal{Y}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)))\right] .$$

Failure of CycleGan on Historical Maps

(a) Orthophotography

(b) Real map



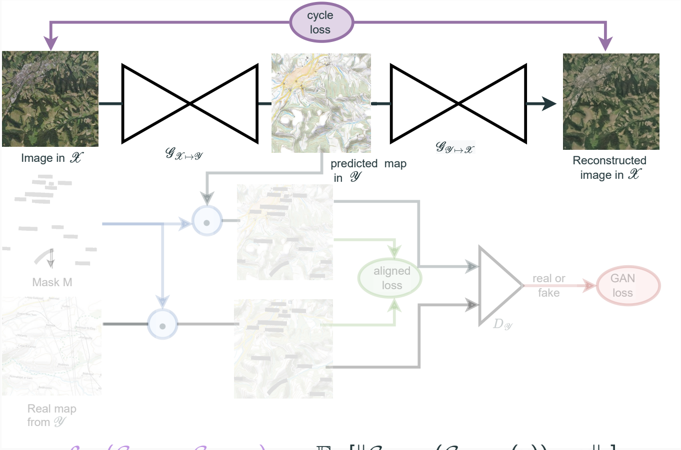
(c) Enc-Dec baseline

(d) Predicted photo

(e) Predicted map

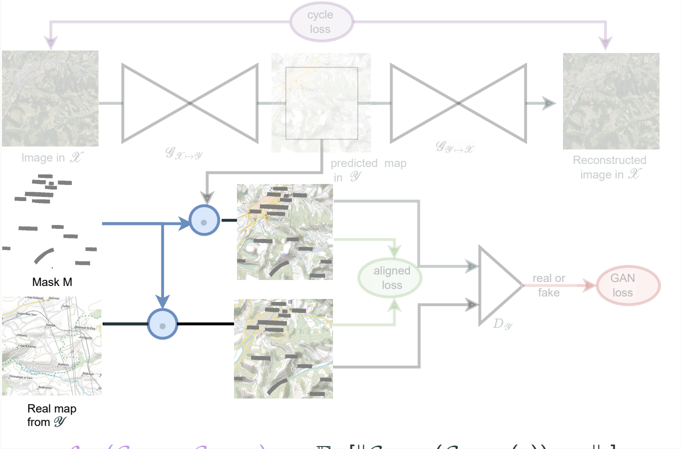
Failure: generate fake text, poor colors

Adversarial Map Generation



$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}\left[\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1\right]$$

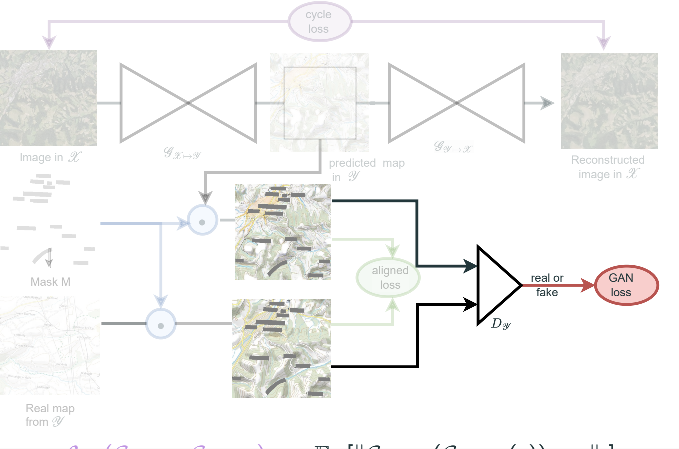
Adversarial Map Generation



$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}} [\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1]$$

Masking both real and predicted maps

Adversarial Map Generation

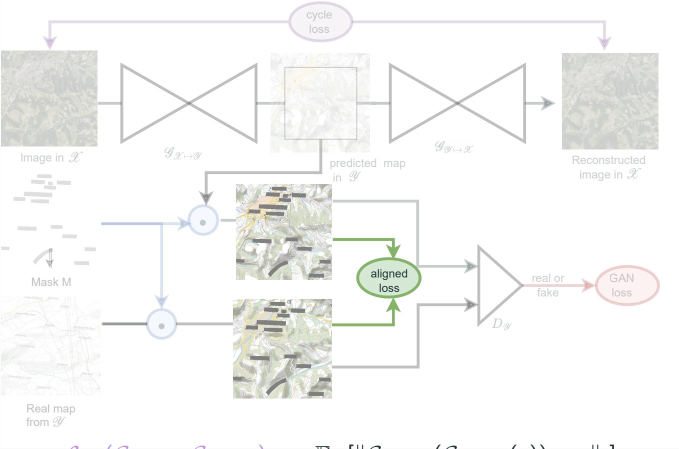


$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1]$$

Masking both real and predicted maps

$$\mathcal{L}_{\text{GAN}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, D_{\mathcal{Y}}, \mathcal{X}, \mathcal{Y}) = \mathop{\mathbb{E}}_{y\sim\mathcal{Y}}[\log D_{\mathcal{Y}}(M\odot y)] + \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[\log(1 - D_{\mathcal{Y}}(M\odot\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)))] \ .$$

Adversarial Map Generation



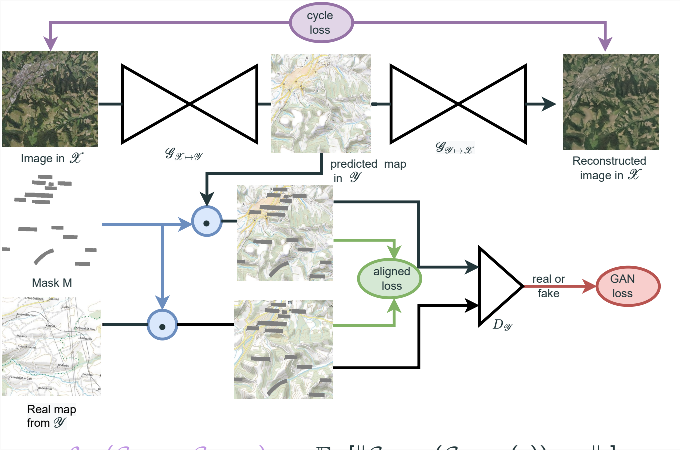
$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1]$$

Masking both real and predicted maps

$$\mathcal{L}_{\text{GAN}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, D_{\mathcal{Y}}, \mathcal{X}, \mathcal{Y}) = \mathop{\mathbb{E}}_{y\sim\mathcal{Y}}[\log D_{\mathcal{Y}}(M\odot y)] + \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[\log(1 - D_{\mathcal{Y}}(M\odot\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)))] \ .$$

$$L_{\text{align}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}) = \mathop{\mathbb{E}}_{x,y\sim\mathcal{X}\times\mathcal{Y}}[\|M\odot\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x) - M\odot y\|_1]$$

Adversarial Map Generation



$$\mathcal{L}_{\text{cyc}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, \mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}) = \mathop{\mathbb{E}}_{x\sim\mathcal{X}} [\|\mathcal{G}_{\mathcal{Y}\mapsto\mathcal{X}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)) - x\|_1]$$

Masking both real and predicted maps

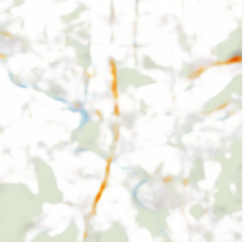
$$\mathcal{L}_{\text{GAN}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}, D_{\mathcal{Y}}, \mathcal{X}, \mathcal{Y}) = \mathop{\mathbb{E}}_{y\sim\mathcal{Y}} [\log D_{\mathcal{Y}}(M\odot y)] + \mathop{\mathbb{E}}_{x\sim\mathcal{X}} [\log(1 - D_{\mathcal{Y}}(M\odot\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x)))] \ .$$

$$L_{\text{align}}(\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}) = \mathop{\mathbb{E}}_{x,y\sim\mathcal{X}\times\mathcal{Y}} [\|M\odot\mathcal{G}_{\mathcal{X}\mapsto\mathcal{Y}}(x) - M\odot y\|_1]$$

Adversarial Map Generation Success Results

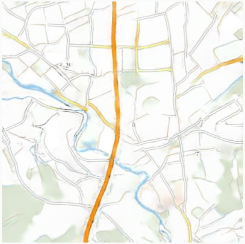
(a) Image in $\mathcal{X}$

(b) Real map from $\mathcal{Y}$



(c) Enc-Dec Baseline

(d) Predicted image in $\mathcal{X}$



(e) Predicted map in $\mathcal{Y}$

(f) CycleGAN Baseline

Conclusion

Conclusion

- Benchmark state-of-the-art text detection models and evaluated them on hand-crafted dataset of historical maps.
- Modified CycleGAN in two key ways: handle text and supervise the alignment.
- Generated maps that are useful!

Limitation to this work: no quantitative evaluation

Thank you!

Thank you for your attention!