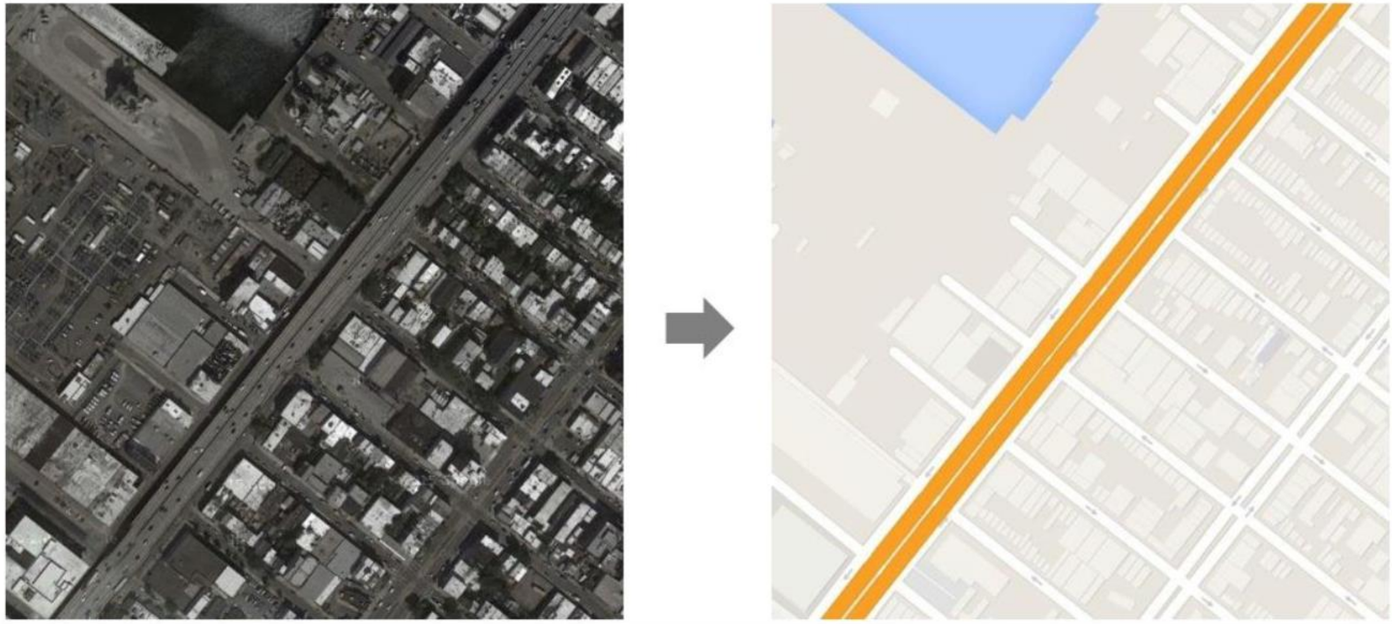
SMAPGAN: Generative Adversarial Network-Based Semi-Supervised Styled Map Tile Generation Method

Xu Chen, Songqiang Chen, Tian Xu, Bangguo Yin, Jian Peng,
Xiaoming Mei, and Haifeng Li Member, IEEE

Outline

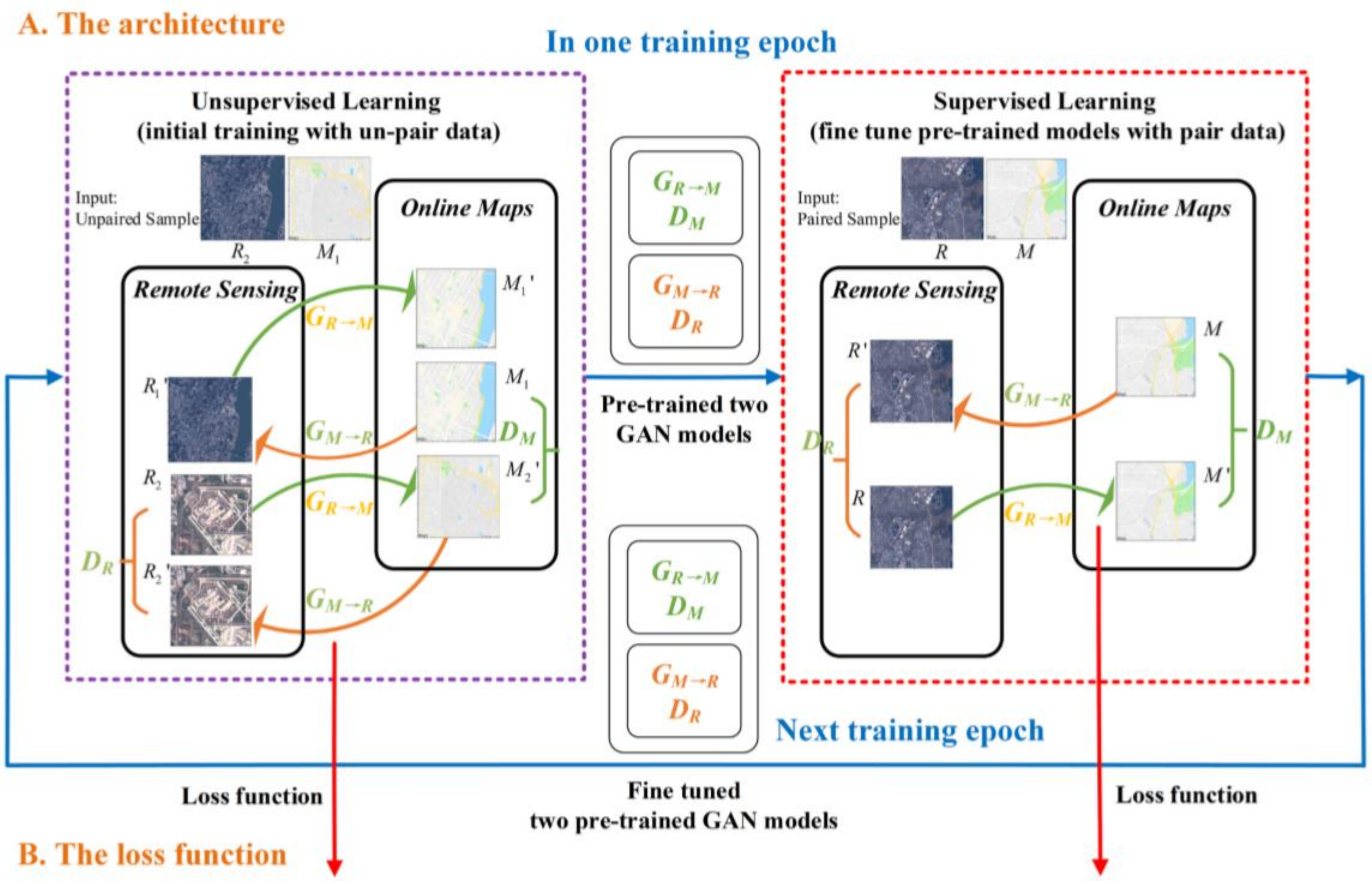
- Challenge
  - ➢ The lack of paired samples consisting of a remote sensing image and its corresponding map tile
  - ➢ Ignored the topological relationship among objects, which is vital in cartography
  - ➢ The metric for evaluating the quality of generated map tiles is lacking.



Outline

- Contribution
  - ➢ proposed a novel styled map tiles generation model named SMAPGAN
  - ➢ designed an image gradient L1 loss and an image gradient structure loss.
  - ➢ proposed a full-reference image quality metric, the edge structural similarity index (ESSI)
- Experiments
- Ablation Study

SMAPGAN Framework



$G_{R \to M}$ is the generator producing map tiles from remote sensing images

D_M is the discriminator that discriminates against generated map tiles

$G_{M \to R}$ is the generator producing remote sensing images from styled map tiles

D_R is the discriminator to discriminate against generated remote sensing images

SMAPGAN Framework

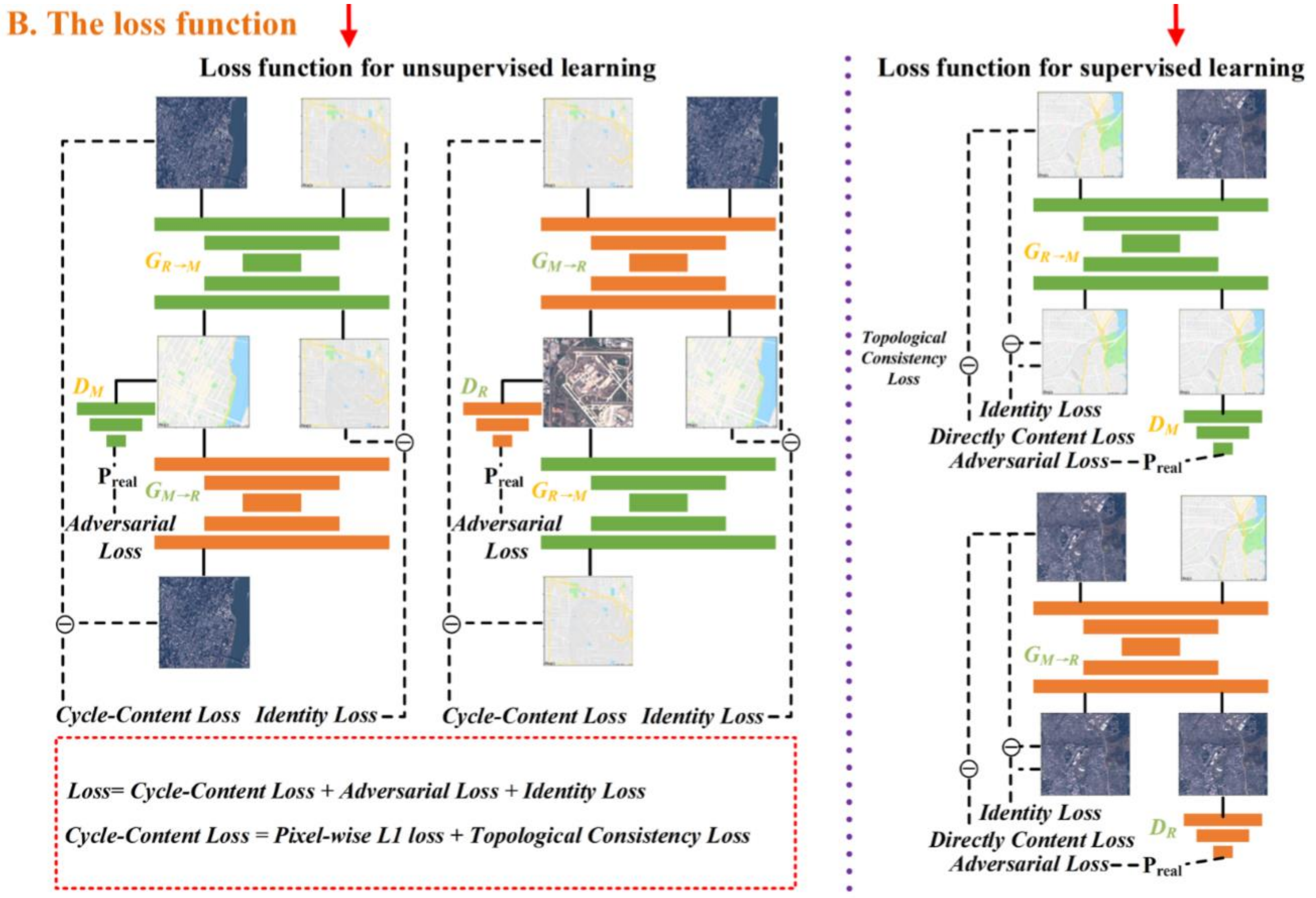


Fig. 1: A SMAPGAN Method Overview; B Unsupervised learning process and supervised learning process in detail.

Loss Function in Semi-Supervised Learning

$$\mathcal{L} = \lambda_{ctn} \ \mathcal{L}_{ctn} + \lambda_{adv} \ \mathcal{L}_{adv} + \lambda_{idt} \ \mathcal{L}_{idt}$$

Content Loss

Adversarial Loss

Identity Loss

RSI-OM-CyC losses

$$\mathcal{L}_{ctn} = \mathcal{L}_{ctn}^{R\to M\to R} + \mathcal{L}_{ctn}^{M\to R\to M} + \mathcal{L}_{ctn}^{M\to R} + \mathcal{L}_{ctn}^{R\to M}$$

Direct-Content losses

Content Loss

$$\mathcal{L}_{ctn} = \mathcal{L}_{ctn}^{R\to M\to R} + \mathcal{L}_{ctn}^{M\to R\to M} + \mathcal{L}_{ctn}^{M\to R} + \mathcal{L}_{ctn}^{R\to M}$$

$$\mathcal{L}_{ctn}^{R\to M\to R} = \lambda_{L1u}\mathcal{L}_{L1} = \lambda_{L1u}\mathbb{E}_{x_R\sim p}[\|G_{M\to R}(G_{R\to M}(x_R) - x_R\|_1]$$

$$\mathcal{L}_{ctn}^{M\to R\to M} = \lambda_{L1u}\mathcal{L}_{L1} + \mathcal{L}_{gral1} + \mathcal{L}_{grastr}$$
$$= \lambda_{L1u}\mathbb{E}_{x_M\sim p}[\|G_{R\to M}(G_{M\to R}(x_M)) - x_M\|_1]$$
$$+ \mathbb{E}_{x_M\sim p}[\|\mathcal{G}(G_{R\to M}(G_{M\to R}(x_M))) - \mathcal{G}(x_M)\|_1]$$
$$+ \mathbb{E}_{x_M\sim p}[2 - \frac{1}{N-1}\sum_{j=0}^{N-2}\frac{|\sigma_{\mathcal{G}_j(x_M)\mathcal{G}_j(G_{R\to M}(G_{M\to R}(x_M)))}| + C_1}{|\sigma_{\mathcal{G}_j(x_M)}\sigma_{\mathcal{G}_j(G_{R\to M}(G_{M\to R}(x_M)))}| + C_1} - \frac{1}{M-1}\sum_{i=0}^{M-2}\frac{|\sigma_{\mathcal{G}_i(x_M)\mathcal{G}_i(G_{R\to M}(G_{M\to R}(x_M)))}| + C_2}{|\sigma_{\mathcal{G}_i(x_M)}\sigma_{\mathcal{G}_i(G_{R\to M}(G_{M\to R}(x_M)))}| + C_2}]$$

*topological consistency loss

$$\mathcal{L}_{ctn} = \mathcal{L}_{ctn}^{R\to M\to R} + \mathcal{L}_{ctn}^{M\to R\to M} + \mathcal{L}_{ctn}^{M\to R} + \mathcal{L}_{ctn}^{R\to M}$$

$$\mathcal{L}_{ctn}^{M\to R} = \lambda_{L1}\mathcal{L}_{L1} = \lambda_{L1}\mathbb{E}_{x_M\sim p}[\|G_{M\to R}(x_M) - x_R\|_1]$$

$$\mathcal{L}_{ctn}^{R\to M} = \lambda_{L1}\mathcal{L}_{L1} + \mathcal{L}_{gral1} + \mathcal{L}_{grastr}$$
$$= \lambda_{L1}\mathbb{E}_{x_R\sim p}[\|G_{R\to M}(x_R) - x_M\|_1]$$
$$+ \mathbb{E}_{x_R\sim p}[\|\mathcal{G}(G_{R\to M}(x_R)) - \mathcal{G}(x_M)\|_1]$$
$$+ \mathbb{E}_{x_R\sim p}[2 - \frac{1}{N-1}\sum_{j=0}^{N-2}\frac{|\sigma_{\mathcal{G}_j(x_M)\mathcal{G}_j(G_{R\to M}(x_R))}| + C_1}{\sigma_{\mathcal{G}_j(x_M)}\sigma_{\mathcal{G}_j(G_{R\to M}(x_R))} + C_1} - \frac{1}{M-1}\sum_{i=0}^{M-2}\frac{|\sigma_{\mathcal{G}_i(x_M)\mathcal{G}_i(G_{R\to M}(x_R))}| + C_2}{\sigma_{\mathcal{G}_i(x_M)}\sigma_{\mathcal{G}_i(G_{R\to M}(x_R))} + C_2}]$$

Loss Function in Semi-Supervised Learning

$$\mathcal{L} = \lambda_{ctn}\ \mathcal{L}_{ctn} + \lambda_{adv}\ \mathcal{L}_{adv} + \lambda_{idt}\ \mathcal{L}_{idt}$$

Content Loss

Adversarial Loss

Identity Loss

$$\mathcal{L}_{adv}^{M\rightarrow R} = \mathbb{E}_{x_R\sim p}[\log D_R(x_R)] + \mathbb{E}_{x_M\sim p}[\log(1 - D_R(G_{M\rightarrow R}(x_M)))]$$

$$\mathcal{L}_{adv}^{R\rightarrow M} = \mathbb{E}_{x_M\sim p}[\log D_M(x_M)] + \mathbb{E}_{x_R\sim p}[\log(1 - D_M(G_{R\rightarrow M}(x_R)))]$$

Loss Function in Semi-Supervised Learning

$$\mathcal{L} = \lambda_{ctn}\ \mathcal{L}_{ctn} + \lambda_{adv}\ \mathcal{L}_{adv} + \lambda_{idt}\ \mathcal{L}_{idt}$$

Content Loss

Adversarial Loss

Identity Loss

$$\mathcal{L}_{idt}^{R\rightarrow M} = \mathbb{E}_{x_M \sim p}[\|G_{R\rightarrow M}(x_M) - x_M\|_1]$$

$$\mathcal{L}_{idt}^{M\rightarrow R} = \mathbb{E}_{x_R \sim p}[\|G_{M\rightarrow R}(x_R) - x_R\|_1]$$

ESSI Compared to MSE and SSIM

Edge Structural Similarity Index ⟶ pearson correlation coefficient + gradient magnitude similarity

measure the topological consistency between a generated map and a real map.

$$ESSI(G(x_R), x_M) = \frac{(|\sigma_{\mathcal{E}(G(x_R))\mathcal{E}(x_M)}| + C_1)(2\mu_{\mathcal{E}(G(x_R))}\mu_{\mathcal{E}(x_M)} + C_2)}{(\sigma_{\mathcal{E}(G(x_R))}\sigma_{\mathcal{E}(x_M)} + C_1)(\mu^2_{\mathcal{E}(G(x_R))} + \mu^2_{\mathcal{E}(x_M)} + C_2)}$$

MSE of 148.50
SSIM of 0.9531
ESSI of 0.3036

MSE of 169.41
SSIM of 0.9482
ESSI of 0.4299

MSE can measure the global pixel-wise similarity between the generated maps and their corresponding ground truth.

SSIM can measure the similarity of the luminance, contrast, and structure between generated maps and their ground truth.

Experiment

Dataset A

Aerial photograph ↔ Maps

Dataset B

Simple styled maps ↔ Target styled maps

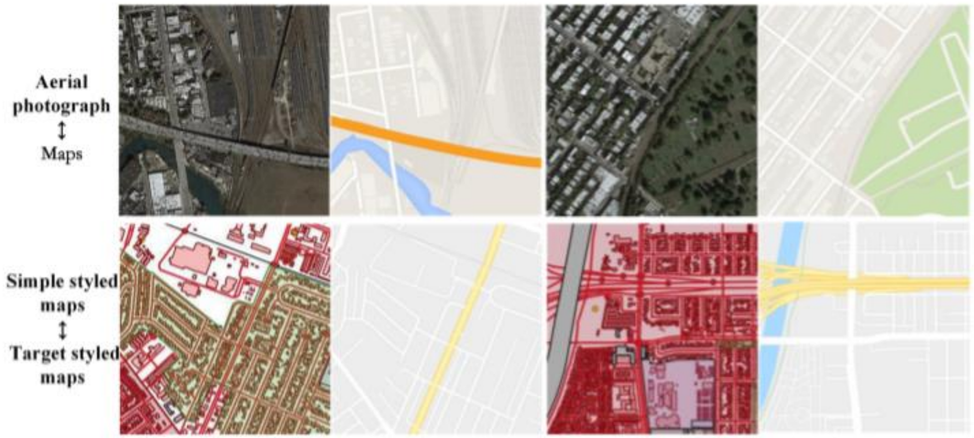


TABLE I: Training and Testing Time on Different Datasets

	amount of training samples	amount of testing samples	training time (h:m:s)	testing time (h:m:s)
A-10%	1768	1030	12:18:24	0:02:45
A-50%	2240	1030	14:36:30	0:03:04
B-10%	840	580	5:53:38	0:01:21
B-50%	1064	580	6:51:41	0:01:29

Experiment

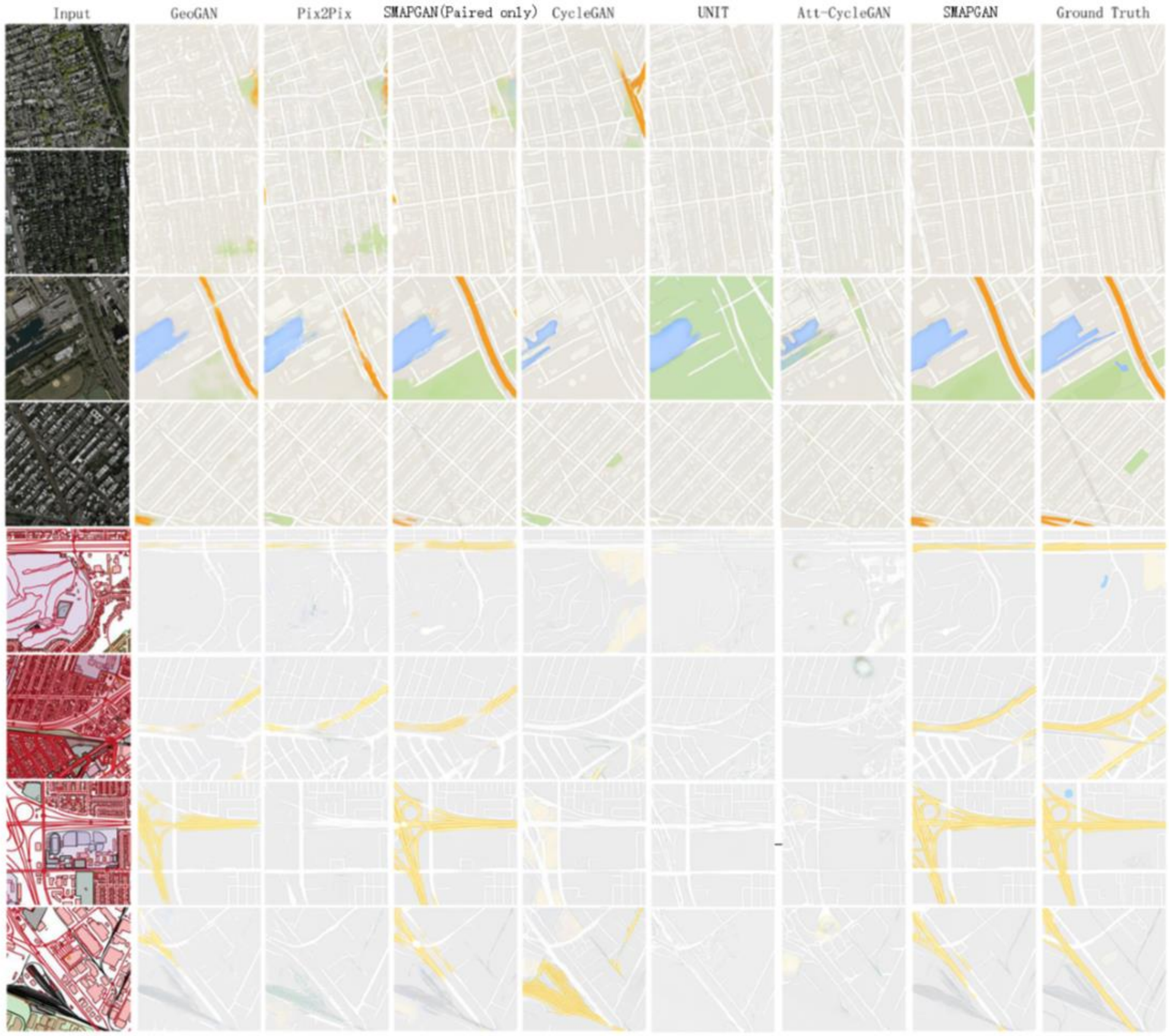


TABLE II: MSE (the smaller the better)

Model\Dataset	A-10%		A-50%		B-10%		B-50%	
	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance
GeoGAN	449.9124	275.8598	193.0430	129.7763	91.8676	53.7895	70.8694	42.5808
Pix2Pix	382.4626	253.9473	296.8432	205.6530	111.4474	68.0716	116.1450	67.0667
SMAPGAN(paired)	323.7949	199.8037	174.6797	116.9807	94.2073	55.3836	71.5514	43.8748
CycleGAN	449.9124	275.8598	439.5733	265.6958	173.4133	110.6720	167.6997	104.2759
UNIT	525.7523	305.2930	483.8580	277.7960	142.5751	90.8529	127.1820	71.7548
Att-CycleGAN	524.8167	299.4017	501.5179	291.2331	165.3915	114.7556	148.6484	98.4731
SMAPGAN	***385.5125***	***226.1506***	***176.3012***	***116.7916***	***100.4296***	***58.0107***	***75.2226***	***44.3301***

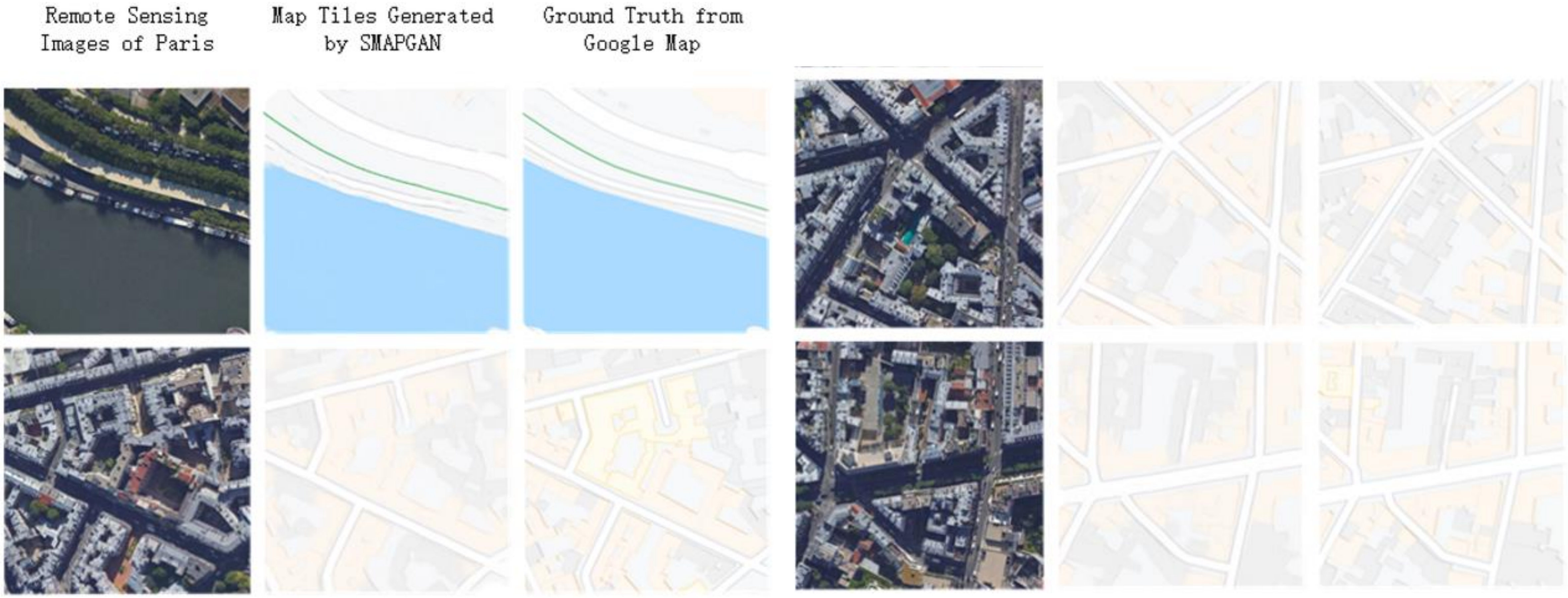
TABLE III: SSIM (the closer to 1 the better)

Model\Dataset	A-10%		A-50%		B-10%		B-50%	
	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance
GeoGAN	0.7363	0.7534	0.7889	0.8008	0.8408	0.8483	0.8606	0.8667
Pix2Pix	0.6657	0.6862	0.6998	0.7243	0.8029	0.8108	0.8249	0.8318
SMAPGAN(paired)	0.7419	0.7572	0.7915	0.8029	0.8432	0.8506	0.8644	0.8698
CycleGAN	0.7140	0.7273	0.7166	0.7287	0.6802	0.6880	0.6893	0.6968
UNIT	0.6390	0.6534	0.7003	0.7155	0.7160	0.7217	0.8039	0.8105
Att-CycleGAN	0.6568	0.6709	0.6708	0.6836	0.6270	0.6318	0.6710	0.6760
SMAPGAN	***0.7515***	***0.7651***	***0.7993***	***0.8105***	***0.8443***	***0.8511***	***0.8665***	***0.8720***

TABLE IV: ESSI (the closer to 1 the better)

Model\Dataset	A-10%		A-50%		B-10%		B-50%	
	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance	RGBmean	Luminance
GeoGAN	0.1758	0.186	0.2541	0.2661	0.3961	0.3995	0.4464	0.4502
Pix2Pix	0.1370	0.1475	0.1879	0.2047	0.3229	0.3277	0.3617	0.3673
SMAPGAN(paired)	0.2162	0.2278	0.2916	0.3071	0.4076	0.4124	0.4557	0.4605
CycleGAN	0.2025	0.2179	0.2084	0.2215	0.0150	0.0185	0.0253	0.0294
UNIT	0.0881	0.0939	0.1703	0.1787	0.1572	0.1594	0.3196	0.3199
Att-CycleGAN	0.1418	0.1509	0.0881	0.1660	0.0730	0.0725	0.0309	0.0328
SMAPGAN	***0.2390***	***0.2553***	***0.3043***	***0.3234***	***0.4270***	***0.4289***	***0.4600***	***0.4652***

Experiment

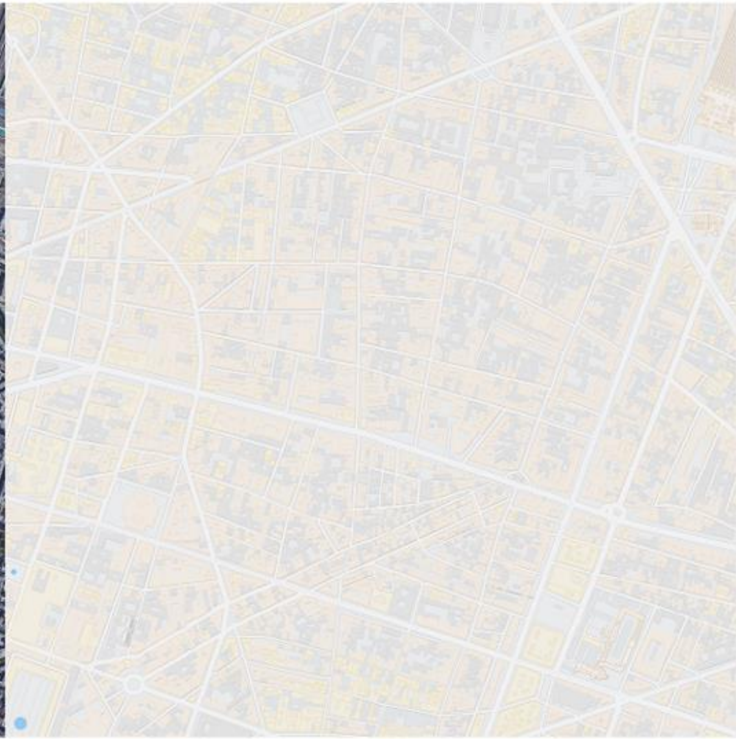


Remote Sensing Images of Paris
Map Tiles Generated by SMAPGAN
Ground Truth from Google Map

Experiment

Remote Sensing Image of Downtown in Paris
(4096*4096)

Google Map of Downtown in Paris
(4096*4096)



Generated Map Tile by SMAPGAN of Downtown in Paris
(2048*2048)

Ablation Study

$$\mathcal{L} = \lambda_{ctn}\ \mathcal{L}_{ctn} + \lambda_{adv}\ \mathcal{L}_{adv} + \lambda_{idt}\ \mathcal{L}_{idt}$$

TABLE VI: Comparison of objective metrics of L1 loss's ablation study.

Component\Metric	MSE	SSIM	ESSI
Full	83.77	0.8105	0.3234
L1	95.80	0.7915	0.2956
L1&GraStr	***83.97***	***0.8105***	***0.3227***
L1&GraL1	90.11	0.7942	0.3034

TABLE VII: Comparison of the objective metrics of image gradient L1 loss's ablation study.

Component\Metric	MSE	SSIM	ESSI
Full	83.77	0.8105	0.3234
GraL1	210.39	0.7300	0.2379
GraStr& GraL1	193.66	0.7999	0.3229
L1&GraL1	***90.11***	***0.7942***	***0.3034***

TABLE VIII: Comparison of the objective metrics of the image gradient structure loss's ablation study.

Component\Metric	MSE	SSIM	ESSI
Full	83.77	0.8105	0.3234
GraStr	162.81	0.7996	0.3222
GraStr&GraL1	193.66	0.7999	0.3229
L1&GraStr	***83.97***	***0.8105***	***0.3227***

Appendix for topological consistency loss Gradient L1 Loss

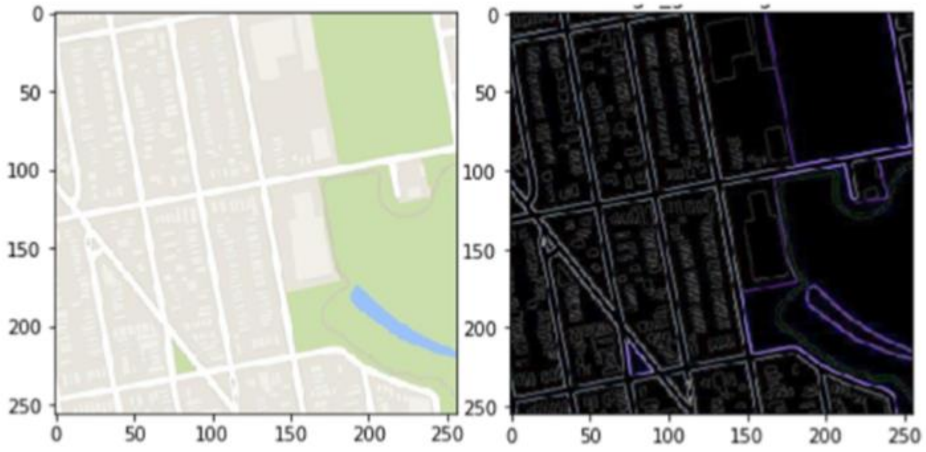


Fig. 3: A 256×256 Google map tile and its 255×255 gradient map.

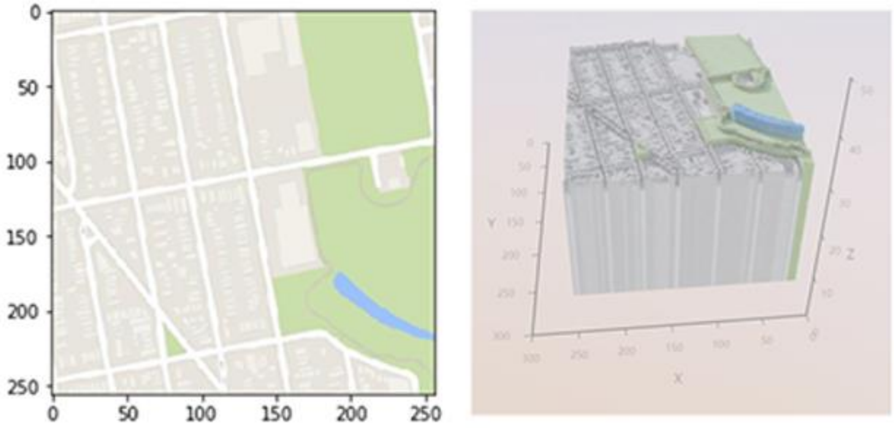


Fig. 2: A 256×256 Google map tile and its 3d map, in which the z-axis is the luminance.

a gradient map can preserve the topological relationships of objects and the differences among adjoining objects.

$$\mathcal{L}_{gral1} = \mathbb{E}_{x_R \sim p}[\|\mathcal{G}(G_{R \to M}(x_R)) - \mathcal{G}(x_M)\|_1]$$

$$g_x(i,j) = f(i,j+1) - f(i,j)$$

$$g_y(i,j) = f(i+1,j) - f(i,j)$$

$$g(i,j) = \sqrt[2]{g_x(i,j)^2 + g_y(i,j)^2}$$

$\mathcal{G}$ consisting of $g(i,j)$

Appendix for topological consistency loss

Gradient Structure Loss

Pearson correlation coefficient

$$\rho(\mathcal{G}_j(x_M), \mathcal{G}_j(G_{R\to M}(x_R))) = \frac{\sigma_{\mathcal{G}_j(x_M)\mathcal{G}_j(G_{R\to M}(x_R))}}{\sigma_{\mathcal{G}_j(x_M)}\sigma_{\mathcal{G}_j(G_{R\to M}(x_R))}}$$

$$\rho(\mathcal{G}_j(x_M), \mathcal{G}_j(G_{R\to M}(x_R)))_{mean} = \frac{1}{N-1}\sum_{j=0}^{N-2}\frac{\sigma_{\mathcal{G}_j(x_M)\mathcal{G}_j(G_{R\to M}(x_R))}}{\sigma_{\mathcal{G}_j(x_M)}\sigma_{\mathcal{G}_j(G_{R\to M}(x_R))}}$$

$$\rho(\mathcal{G}_i(x_M), \mathcal{G}_i(G_{R\to M}(x_R)))_{mean} = \frac{1}{M-1}\sum_{i=0}^{M-2}\frac{\sigma_{\mathcal{G}_i(x_M)\mathcal{G}_i(G_{R\to M}(x_R))}}{\sigma_{\mathcal{G}_i(x_M)}\sigma_{\mathcal{G}_i(G_{R\to M}(x_R))}}$$

$$\mathcal{L}_{grastr} = \mathbb{E}_{x_R\sim p}[2 - \frac{1}{N-1}\sum_{j=0}^{N-2}\frac{|\sigma_{\mathcal{G}_j(x_M)\mathcal{G}_j(G_{R\to M}(x_R))}| + C_1}{\sigma_{\mathcal{G}_j(x_M)}\sigma_{\mathcal{G}_j(G_{R\to M}(x_R))} + C_1} - \frac{1}{M-1}\sum_{i=0}^{M-2}\frac{|\sigma_{\mathcal{G}_i(x_M)\mathcal{G}_i(G_{R\to M}(x_R))}| + C_2}{\sigma_{\mathcal{G}_i(x_M)}\sigma_{\mathcal{G}_i(G_{R\to M}(x_R))} + C_2}]$$