Sensor Data-Driven Road Geometry Extraction and Correction

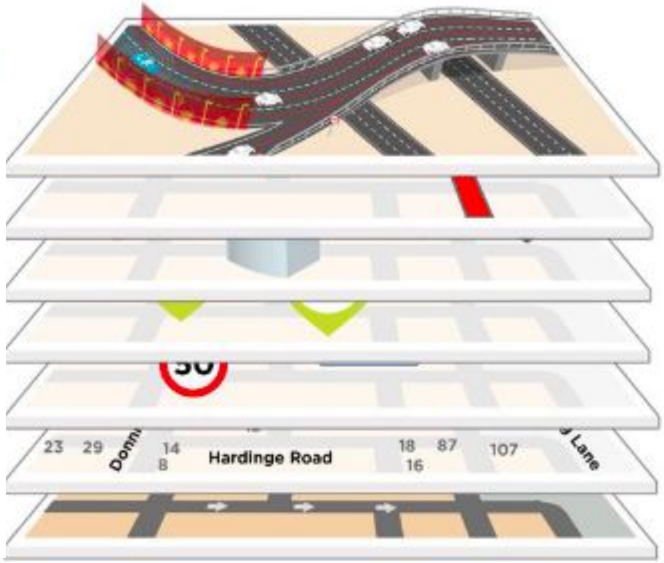
Soojung Hong

March 14. 2023

Map

1 Ranking algorithm

Automated driving
Driver assistance
Visualization
Search
Navigation
Geocoding
Routing



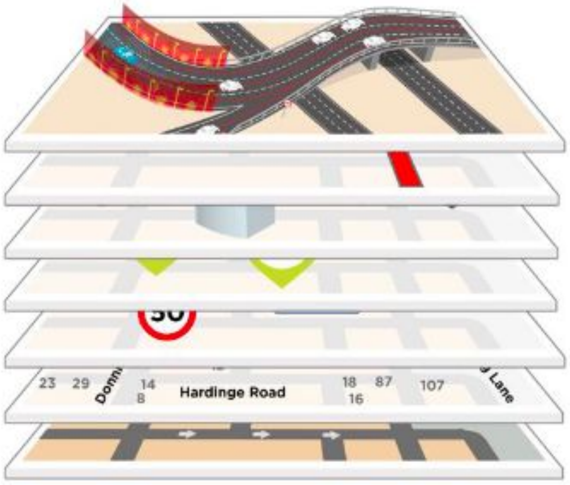
HD & RoadDNA
Advanced driver assistance
3D
POI
Traffic signs
Addressing
Road network

2 Place Category Prediction using GNN

3 Road Geometry Extraction

Image credit : TomTom

Road Network Layers



Road Network

Image credit : HERE Technologies

Data Collection for Road Network

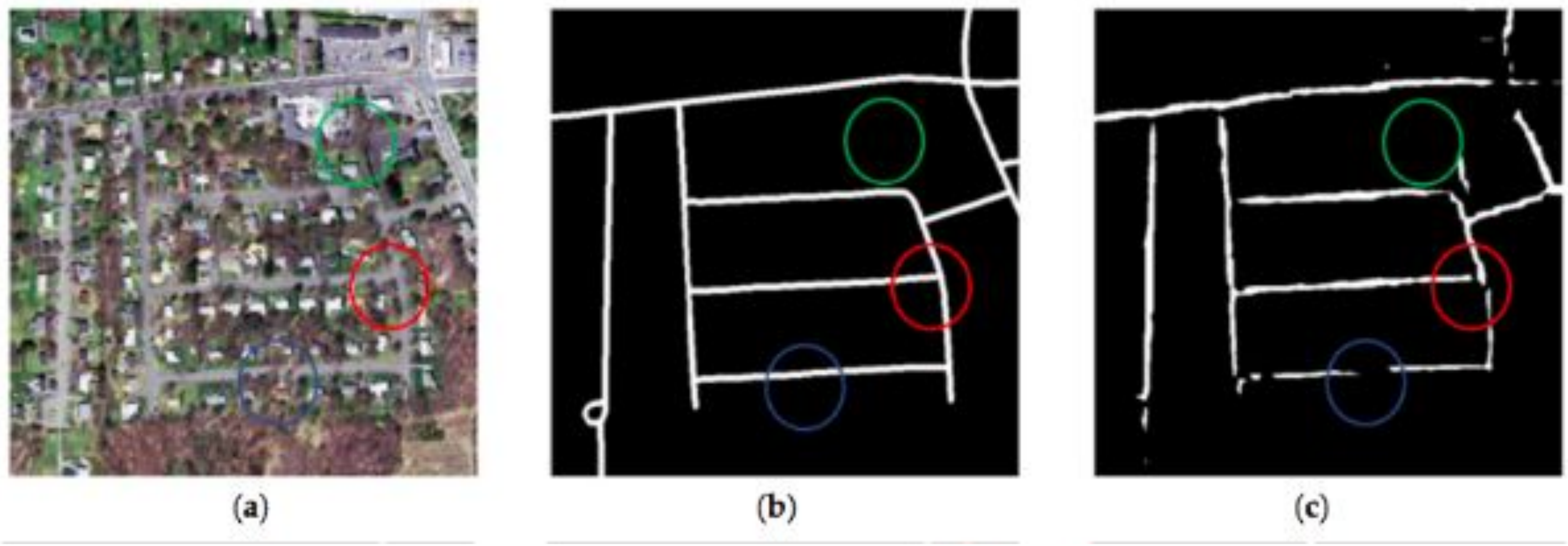
Satellite (e.g NASA Landsat 8) or Aircraft

Radar, Lidar, Image data collection

Aerial Image

Vehicle location data (GPS data, Probe data)

Road Detection using Semantic Segmentation

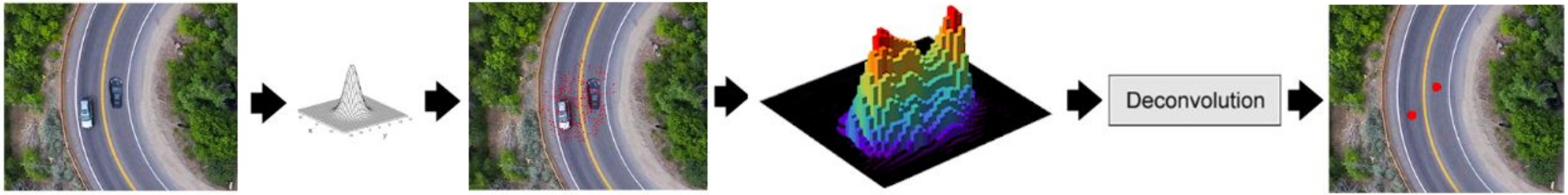


Input remote sensing image | Label (Ground Truth) | Predicted Road by DNN

GapLoss: A Loss Function for Semantic Segmentation of Roads in Remote Sensing Images
Wei Yuan 1,2,* and Wenbo Xu 3

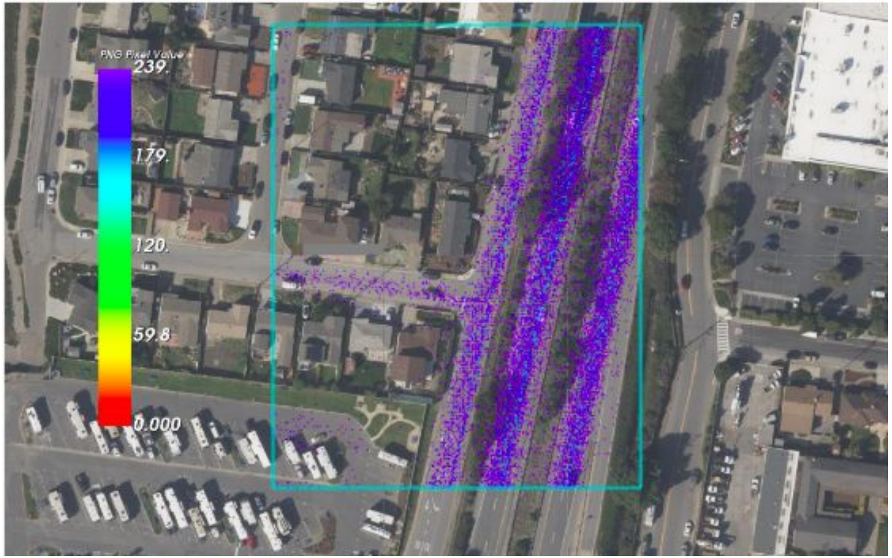
Prior art for deriving road geometry from probe (sensor) data

- K-means trajectory clustering
- Trajectory merging
- Kernel density Estimation (KDE)
- Principal Curves (using moving direction, vehicle speed, position data)



Lane center restoration from probe data

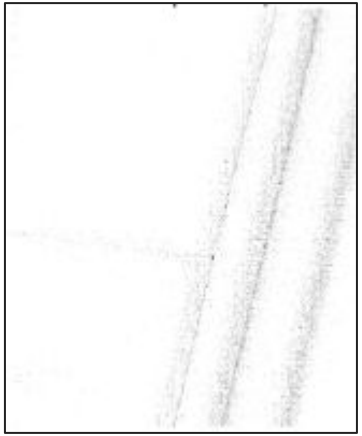
Re-think : Road geometry creation as Deblurring process



Deblurring

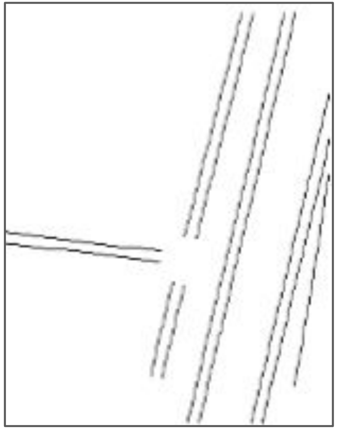
Road geometry extraction using DeblurGAN

Geo coordinate (Latitude, Longitude)
~~Elevation~~
~~Velocity~~
~~Direction~~



Input

DeblurGAN



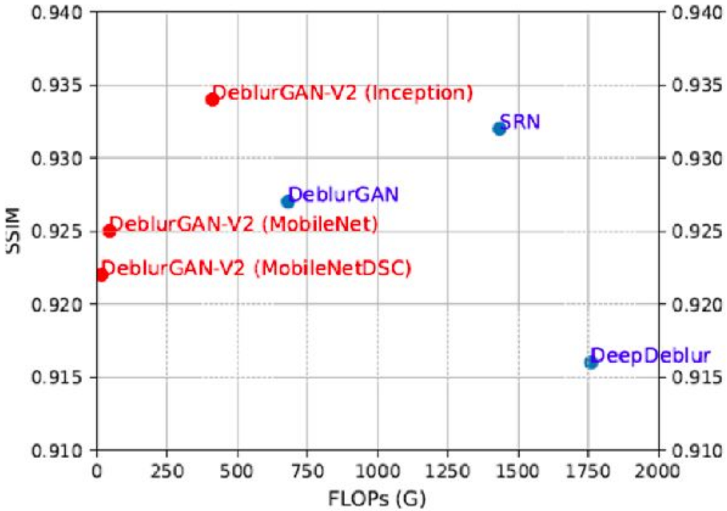
Output

Deblurred image example by DeblurGAN

Blurred Image

Deblurred Image by DeblurGAN

Sharp Image

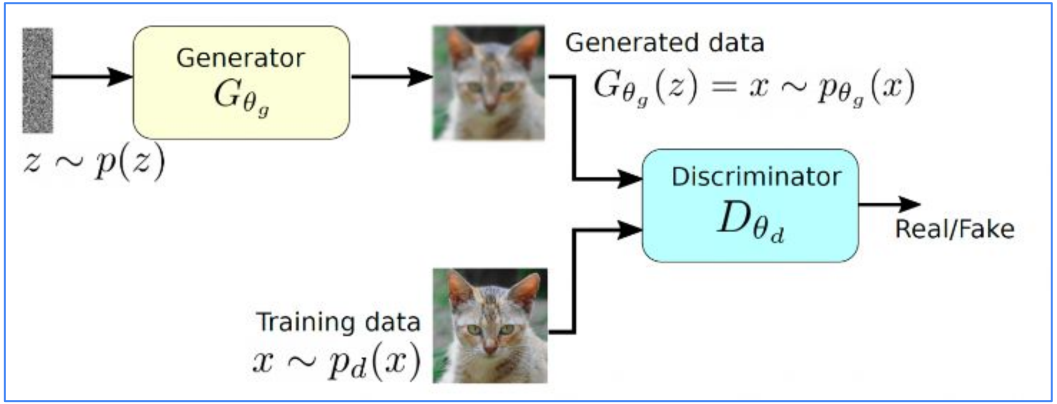


✓ Performance measure :

SSIM : structural similarity index measure

FLOPs : floating point operations per second

Generative Adversarial Networks (GAN)



$$\min_G \max_D \underset{x \sim \mathbb{P}_r}{\mathbb{E}} [\log(D(x))] + \underset{\tilde{x} \sim \mathbb{P}_g}{\mathbb{E}} [\log(1 - D(\tilde{x}))]$$

Discriminators try to **maximize**

(i.e. D(x) – the probability that real image is real is high)

D(G(z)) is the probability that fake image is real become low

(close to 0)

Generators try to **minimize**

(i.e. try to maximize the discriminator's

output for fake instance , D(G(z)), is

probably real (close to 1)

DeblurGAN ver1 architecture

CNN learns a residual correction IR to the blurred image IB, so IS = IB + IR.

IS is sharp image, IB is blur image, IR is residual correction

global skip connection

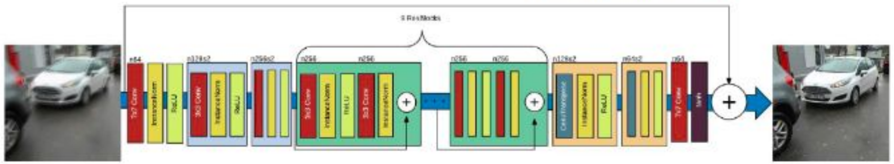


Figure 3: DeblurGAN generator architecture. DeblurGAN contains two strided convolution blocks with stride $\frac{1}{2}$, nine residual blocks [13] and two transposed convolution blocks. Each ResBlock consists of a convolution layer, instance normalization layer, and ReLU activation.

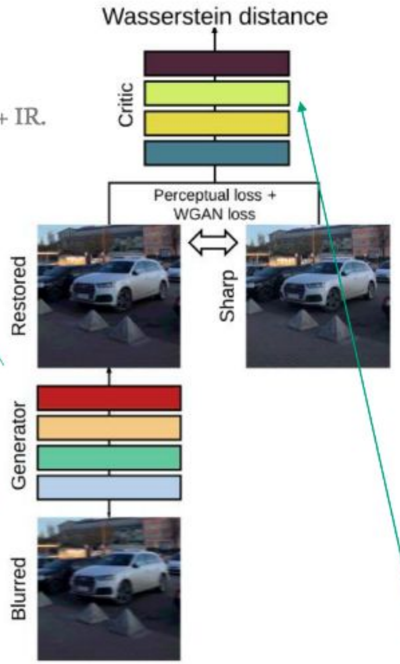
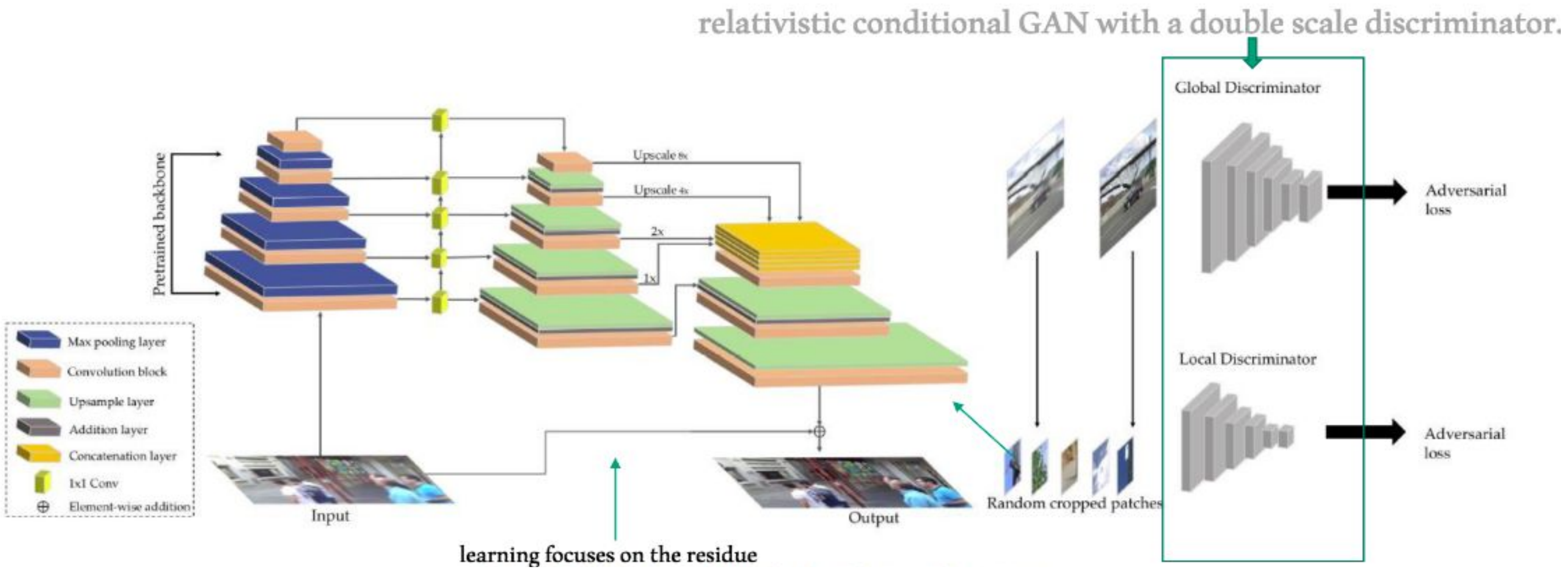


Figure 4: DeblurGAN training. The generator network

5

Critic network D architecture of critic network is identical to PatchGAN

DeblurGAN ver2 architecture



DeblurGAN ver1 Loss Function

Loss function : combination of content loss and adversarial loss

$$\underbrace{\mathcal{L} = \underbrace{\mathcal{L}_{GAN}}_{adv\ loss} + \underbrace{\lambda \cdot \mathcal{L}_X}_{content\ loss}}_{total\ loss}$$

$$L = \underbrace{\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g} [D(\tilde{\boldsymbol{x}})] - \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_r} [D(\boldsymbol{x})]}_{\text{Original critic loss}} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}} \left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2 \right]}_{\text{Our gradient penalty}}$$

Wasserstein GAN - GP

- ✓ WGAN-GP improve stability during training model
- ✓ WGAN-GP is robust to the choice of generator architecture
- ✓ Loss function helps to generate higher quality result

"content" loss function : L1 or MAE loss, L2 or MSE loss on raw pixels.

content loss can lead to the **blurry artifacts** on generated images

Instead, proposed **Perceptual loss** (a simple L2-loss,) but based on the difference of **the generated and target image CNN feature maps**

$$\mathcal{L}_X = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^S)_{x,y} - \phi_{i,j}(G_{\theta_G}(I^B))_{x,y})^2$$

Feature map by j-th convolution before ith maxpooling layer in VGG19 network

DeblurGAN ver2 Loss Function

$$L_G = 0.5 * L_p + 0.006 * L_X + 0.01 * L_{adv}$$

Pixel-space loss (simplest L1 or L2 distance)

Lp term was not included in ver 1

It helps correct color and texture distortions

Relativistic GAN loss

$$L_D^{RaLSGAN} = \mathbb{E}_{x \sim p_{data}(x)}\left[(D(x) - \mathbb{E}_{z \sim p_z(z)} D(G(z)) - 1)^2\right] + \mathbb{E}_{z \sim p_z(z)}\left[(D(G(z)) - \mathbb{E}_{x \sim p_{data}(x)} D(x) + 1)^2\right]$$

- ✓ notably faster
- ✓ more stable compared to WGAN-GP objective
- ✓ empirically the generated results possess higher perceptual quality and overall sharper outputs

Content loss

Again, using perceptual distance

(Euclidean loss on the VGG19 conv3 3 feature maps)

$$\mathcal{L}_X = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^S)_{x,y} - \phi_{i,j}(G_{\theta_G}(I^B))_{x,y})^2$$

Feature map by j-th convolution before ith maxpooling layer in VGG19 network

Evaluation Metric (quality metric)

- PSNR (Peak Signal to Noise ratio)

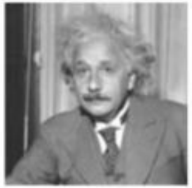
$$MSE = \frac{1}{wh}\sum_{i=0}^{h-1}\sum_{j=0}^{w-1}|I(i,j) - K(i,j)|^2$$

$$\begin{aligned} PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\ &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned}$$

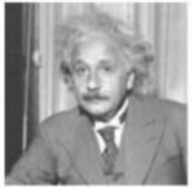
power of distorting noise

MAX_f is the maximum signal value that exists in our original "known to be good" image

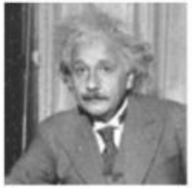
✓ strictly on numeric comparison and does not actually take into account any level of biological factors of the human vision system



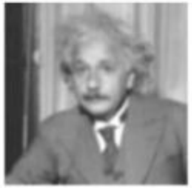
Original
SSIM=1



PSNR=26.547
SSIM=0.988



PSNR=26.547
SSIM=0.840



PSNR=26.547
SSIM=0.694

- SSIM (similarity structure Index measure)

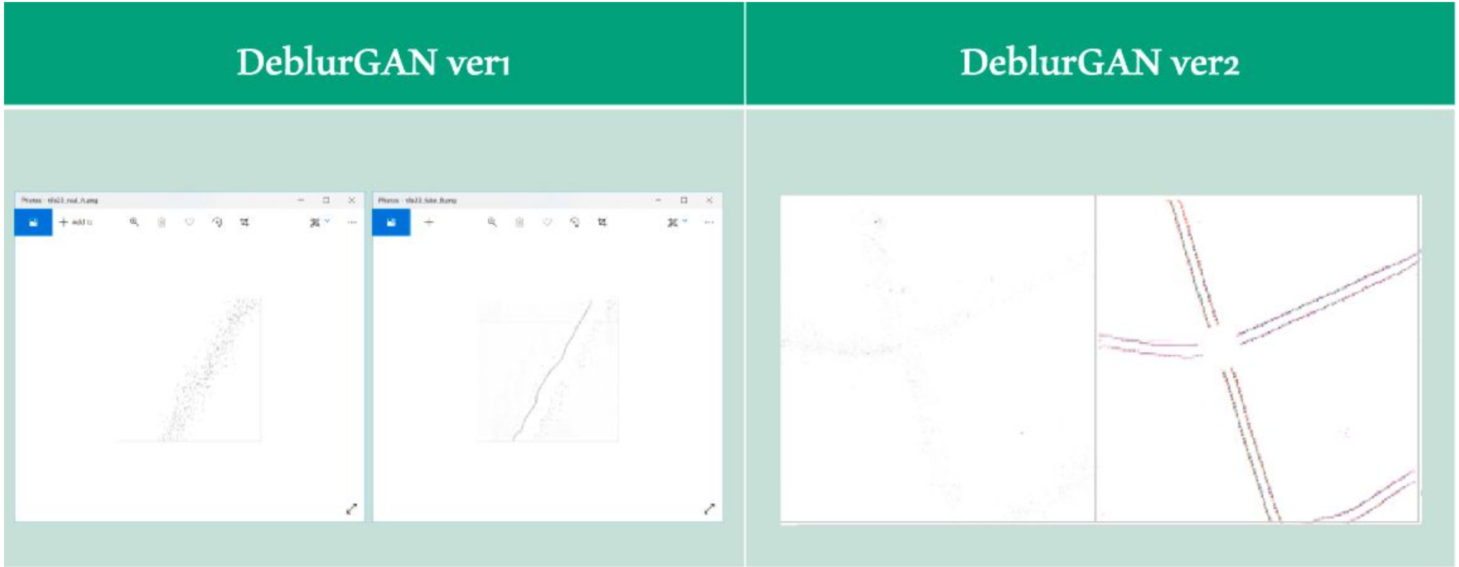
$$\text{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha} \cdot [c(\mathbf{x},\mathbf{y})]^{\beta} \cdot [s(\mathbf{x},\mathbf{y})]^{\gamma}$$

✓ **SSIM** is more perceptual metric

✓ **SSIM** is a newer measurement tool based on three factors (i.e. luminance, contrast, and structure to better suit the workings of the human visual system)

Performance comparison

Model	PSNR	SSIM
DeblurGAN ver1	21.75	0.903
DeblurGAN ver2	**23.56**	**0.9217**

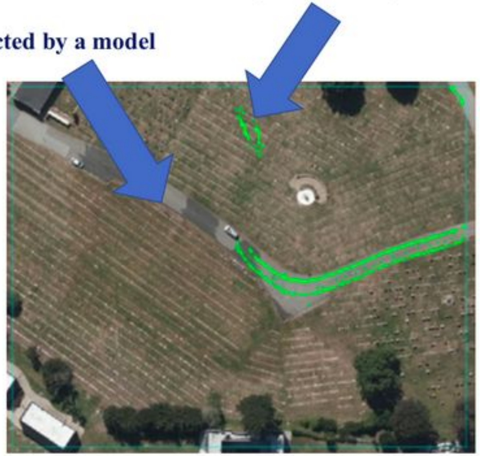


Missing or Wrong geometry

- Example 1 : Detected lane centerlines

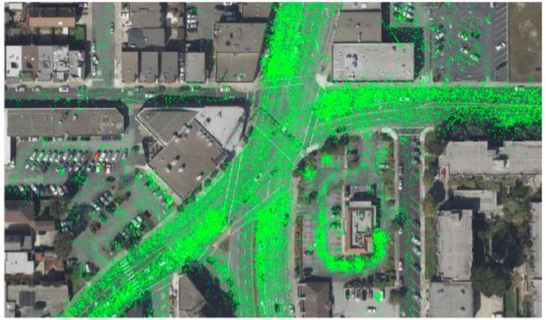
DeblurGAN model

Tile number : 5185912319934464



Detected lane centerlines on map

Why? Hint



Road geometry correction by Image Inpainting model

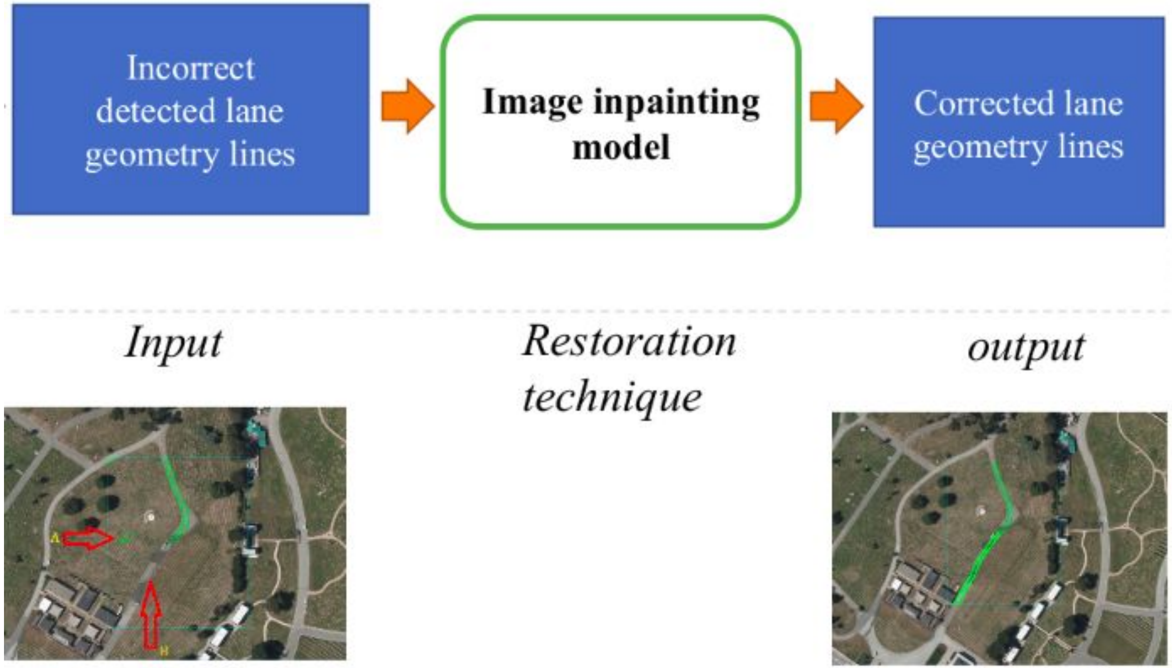
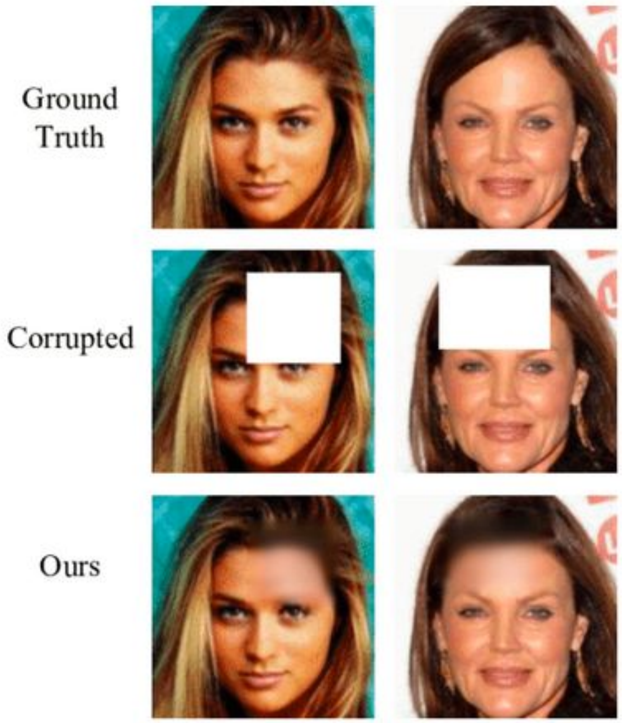
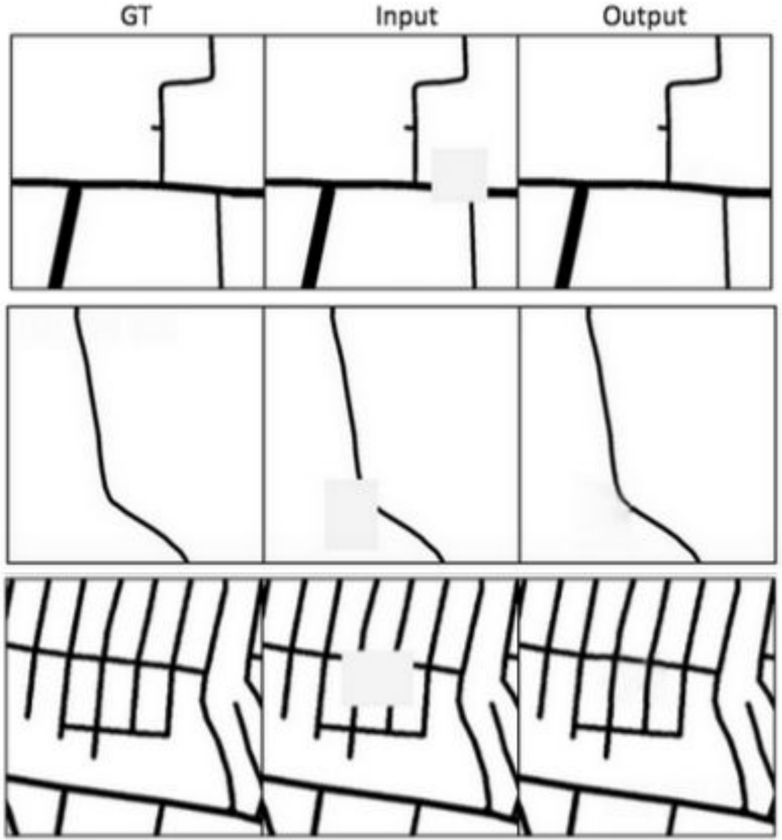
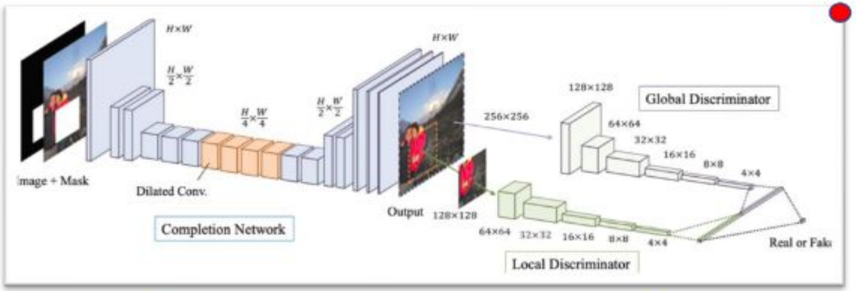


Image Inpainting Methods





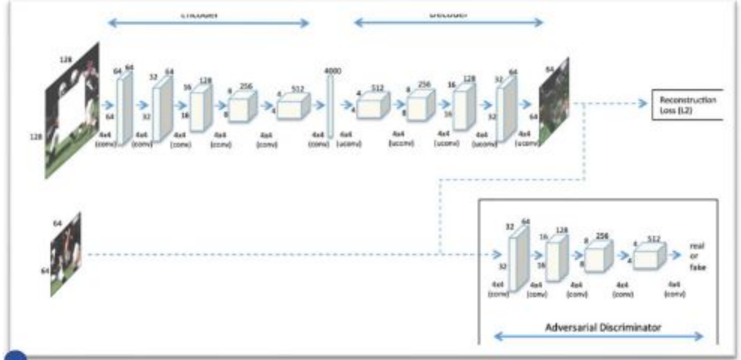


DeepFill v1 (Generative Image Inpainting with contextual attention) 2018

Context Encoder (1st GAN-based inpainting, 2016)

GLCIC (Globally and Locally Consistent Image Completion) 2017

DeepFill v2 (Free form image inpainting with Gated Convolution)



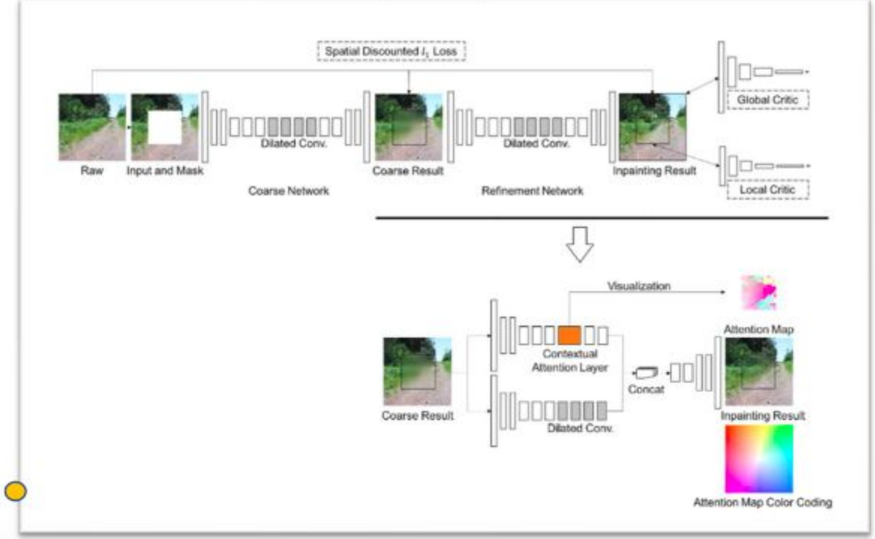


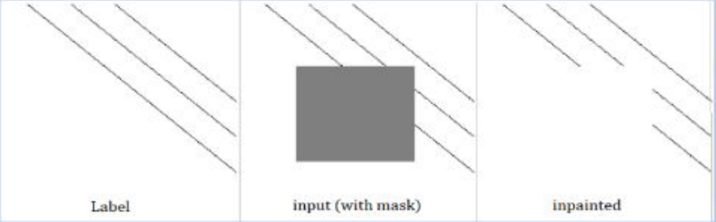
Image Inpainting Model	Image Inpainting Result

Context Encoder

- Data : San Francisco road centerline labels
- 6544 label images
- Epoch 700

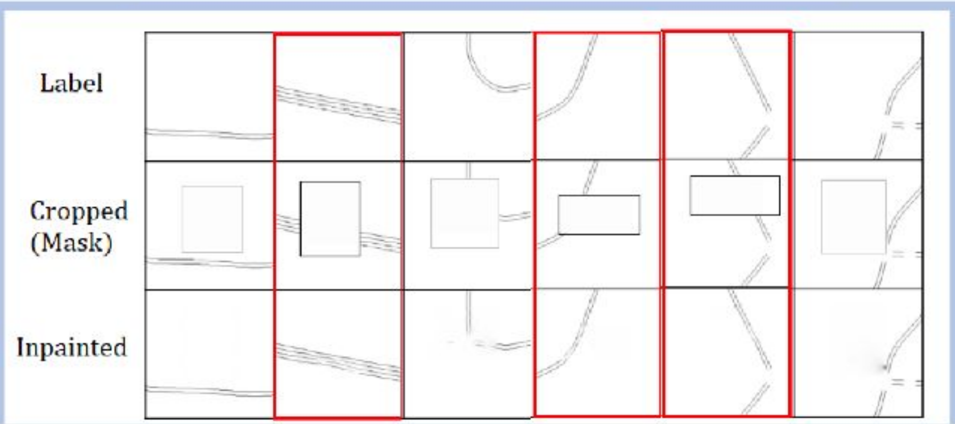
DeepFill ver2

- Data : San Francisco road centerline labels
- 6544 label images
- Epoch : 100000



GLCIC
(Globally and Locally Consistent Image Completion)

- Data : San Francisco road centerline labels
- 6544 label images
- Epoch : 90000



Model Enhancement : From GLCIC To GLCRC+L

Globally and **L**ocally **C**onsistent **I**mage **C**ompletion **(GLCIC)**

Vanilla (Baseline) Model

Globally and **L**ocally **C**onsistent **R**oad map **C**ompletion **(GLCRC)**

Architectural Improvement

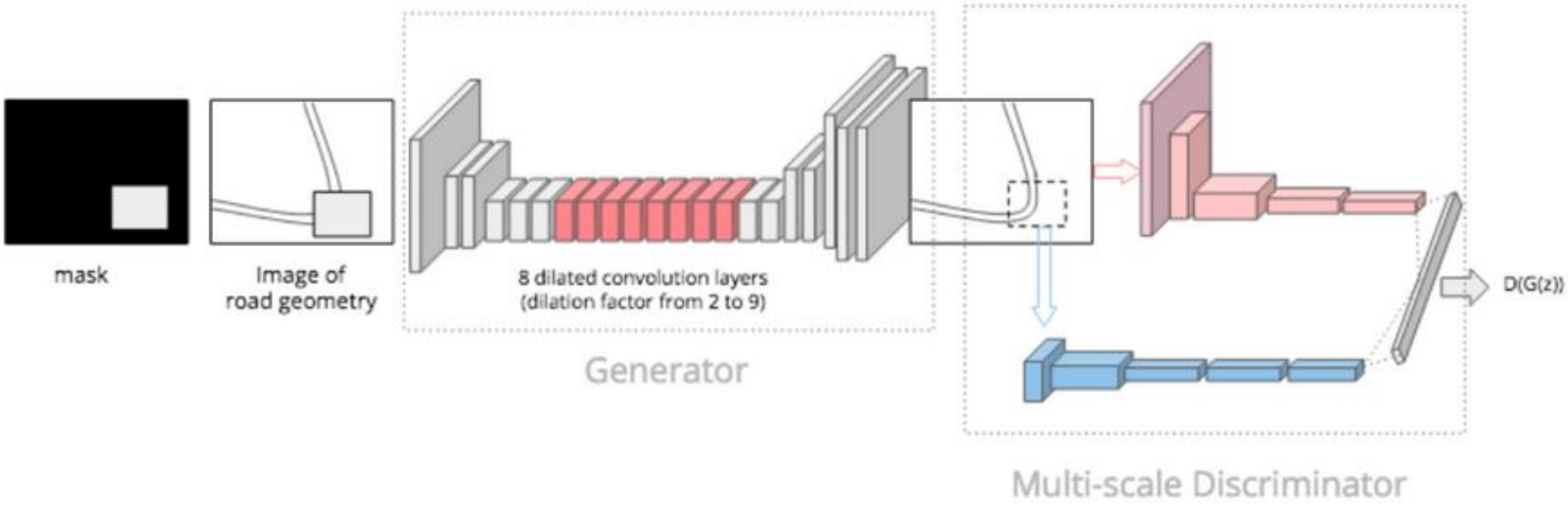
Globally and **L**ocally **C**onsistent **R**oad map **C**ompletion **(GLCRC+L)**

Architectural Improvement
&
Quality enhancement
with new loss functions

Globally and locally consistent image completion. ACM Trans. Graph., 36(4):107:1–107:14, 2017

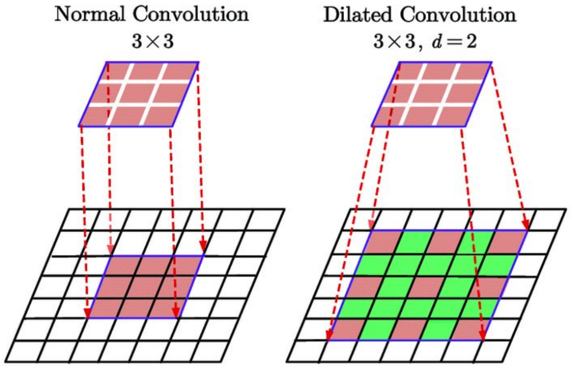
Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa.

Architecture of GLCRC



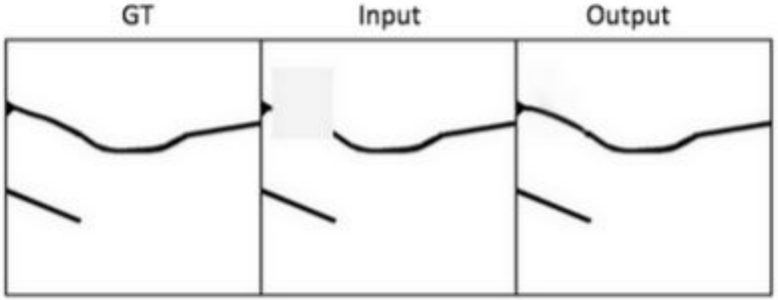
GLCRC : Globally and **L**ocally **C**onsistent **R**oad map **C**ompletion

Why deeper Dilated Convolution layers?



Standard and Dilated Convolution

vs



Loss function

- Loss Function for Generator : Avoiding blurriness via Perceptual loss

$$\mathrm{MSE} = \frac{1}{MN}\sum_{n=0}^{M}\sum_{m=1}^{N}\left[\hat{g}(n,m) - g(n,m)\right]^2$$

$$\mathcal{L}_P = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^O)_{x,y} - \phi_{i,j}(G_{\theta_G}(I^R))_{x,y})^2$$

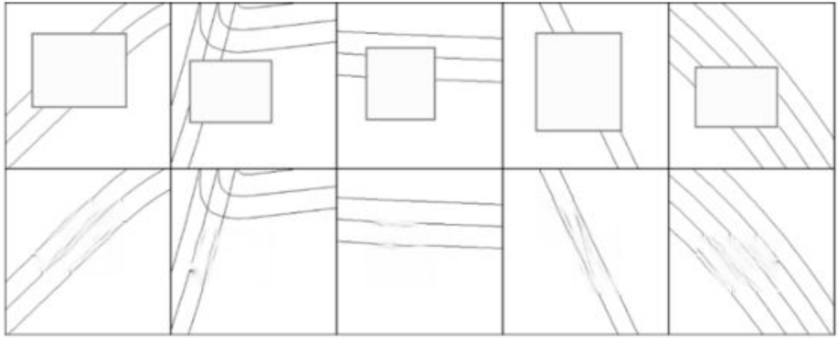
- Loss Function for Discriminator : Increase the sharpness of line via Relativistic Least Square GAN loss

$$L = -\sum_{i\in I}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right)$$

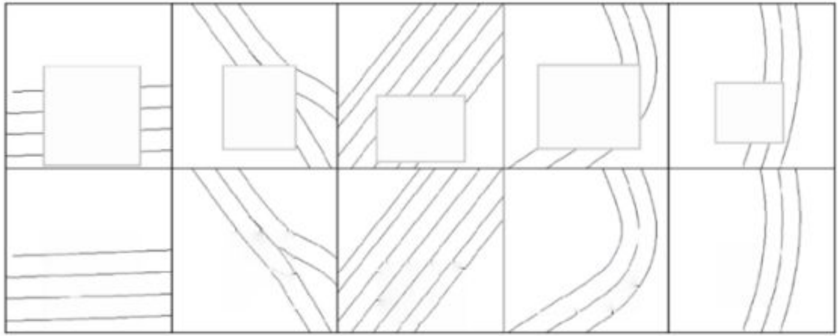
$$\mathcal{L}_D^{\mathrm{RaLSGAN}} = \mathbb{E}_{x\sim p_{\mathrm{data}}}[(D(x) - \mathbb{E}_{z\sim p_z}D(G(z)) - 1)^2] + \mathbb{E}_{z\sim p_z}[(D(G(z)) - \mathbb{E}_{x\sim p_{\mathrm{data}}}D(x) + 1)^2].$$

two greyscale images which are flattened to 1d arrays: $y = (y_1, y_2, \ldots, y_n)$ and $\hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$ with pixel values in $[0, 1]$

Loss function improves quality



(a) GLCIC with MSE Loss



(b) GLCIC with Perceptual Loss

* Mask area
is the target area to correct

Performance Metric

$$\text{Correctness} = \frac{TP}{TP + FN}$$

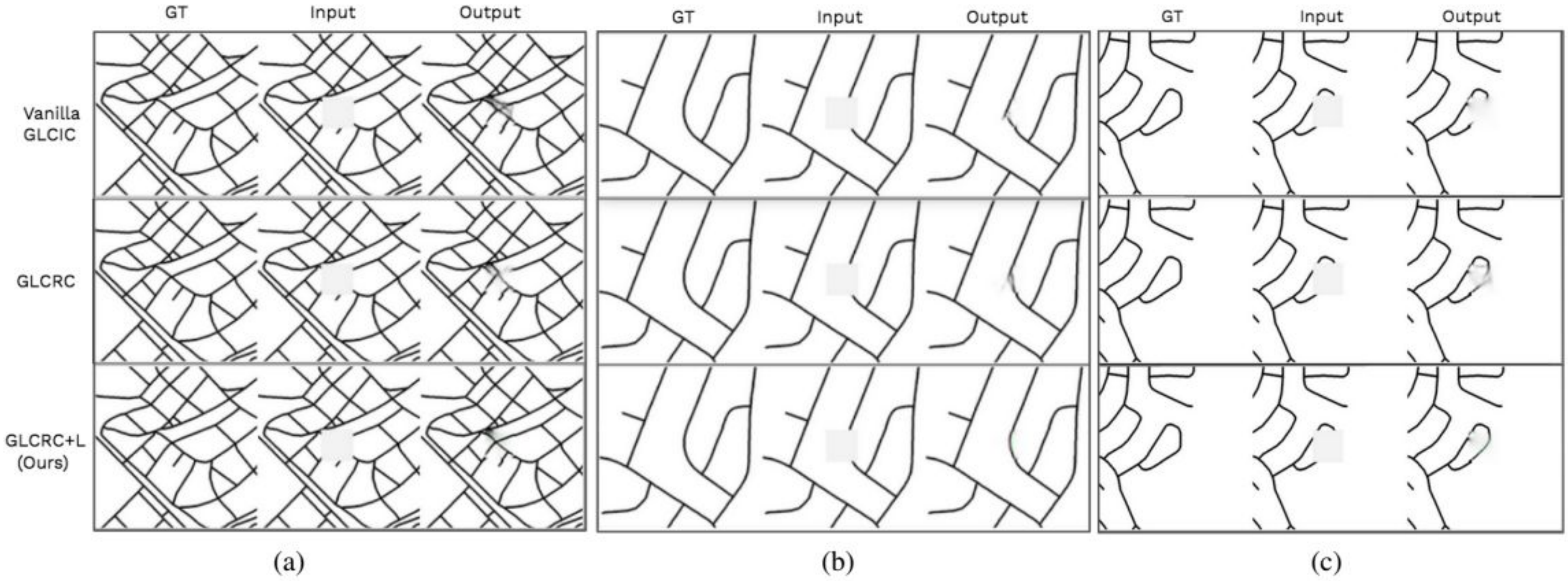
$$\text{Completeness} = \frac{TP}{TP + FP}$$

$$\text{Quality} = \frac{TP}{TP + FN + FP}$$

Method	Correctness	Completeness	Quality
Vanilla GLCIC	0.787	0.803	0.664
GLCRC	0.789	0.811	0.668
GLCRC+L (Ours)	**0.795**	**0.831**	**0.671**

Road type	Method	Correctness	Completeness	Quality
Straight	Vanilla GLCIC	0.787	0.786	0.649
	GLCRC	0.750	0.806	0.635
	GLCRC+L (Ours)	**0.894**	**0.898**	**0.811**
Curvy	Vanilla GLCIC	**0.762**	0.757	**0.613**
	GLCRC	0.723	**0.789**	0.606
	GLCRC+L (Ours)	0.754	0.766	**0.613**
T-junction	Vanilla GLCIC	0.775	0.788	0.642
	GLCRC	0.785	0.792	0.651
	GLCRC+L (Ours)	**0.842**	**0.849**	**0.733**
Intersection	Vanilla GLCIC	0.775	0.788	0.642
	GLCRC	0.785	0.792	0.651
	GLCRC+L (Ours)	**0.786**	**0.793**	**0.652**

Road Geometry Inpainting Results



Summary

- ❑ Road geometry derivation using probe data (GPS data)
- ❑ Geometry derivation via deblurring method (DeblurGAN)
- ❑ Post-processing for faulty road geometry correction using Image Inpainting
- ❑ GLCRC-L : Architectural and new loss adoption
- ❑ Single data source is not enough for road geometry derivation

Caveat : Research project (not applied in production, yet!)