The art of restoration for incorrect lane geometry object images by image inpainting model

Soojung Hong*
HERE Technologies
Zurich, Switzerland
`soojung.hong@here.com`

Abstract

We present a systematic image correction method that detects wrong pixel areas in lane geometry images and corrects them. We experimented image inpainting models that we improved the restored image quality significantly. Our model adopted the deep dilated convolution layers in Generative Adversarial Networks (GAN) model. The automatically corrected areas in images show that our approach can restore any shape of lane geometry object without losing the overall contextual continuity.

1 Introduction

Object detection models have been trained with open image data such as MS-COCO or Google Open Images. The images and their labels in open image set are clean and correct. However, the training images in unusual problem domains are often dirty and incorrect. For instance, the lane geometry objects such as lane boundary lines or road centerlines are detected from aerial images by object detection models. The detected lane geometry objects are core components in any location services such as Advanced driver-assistance system (ADAS) or driving guidance of self-driving car (e.g., Figure 1A, 1B). However, the aerial images capture casted shadow on roads created by trees or high buildings. The tunnels hide roads completely, thus it is impossible to detect any lane lines within tunnels from image. These inherent obstacles in aerial images introduce incorrect line segments in lane geometry object images (e.g., Figure 1C, 1D).

2 Problem statement

The prior works to correct the wrong lane geometry in images were often done by interpolation methods using statistical approximation [2, 5]. Since interpolation methods depend on assumptions that roads are flat and curvatures are constant curvature [14], constructing any unprecedented complex shape of lane geometry is still challenging problem. In this paper, we propose to use image inpainting model to correct the wrong lane geometry objects. Our proposed model can learn various shapes of lane geometry objects regardless whether the objects are linear or nonlinear. The model can be generalized to predict any complex lane geometry object in images. (Figure 2).

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

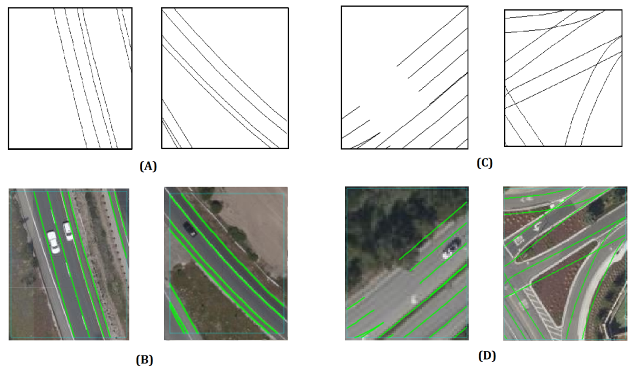

Figure 1: Example of lane geometry object images (A) and their projection on map (B). The incorrect lane geometry object images (C) and their projection on map (D). Left image of (C) contains missing lane line segments. Right image of (C) has misaligned lane boundary lines.

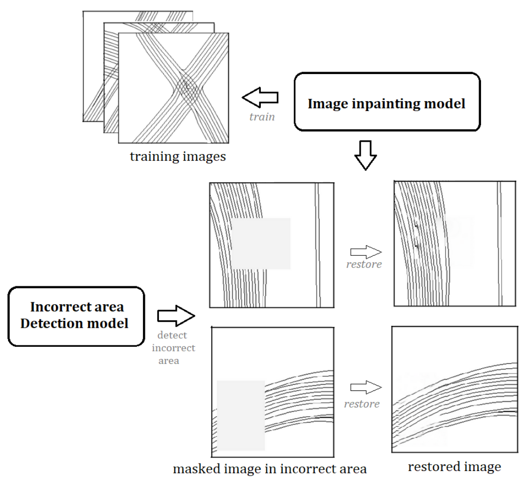

Figure 2: The process of image correction from model training to lane geometry objects restoration

3 Solution for lane geometry object images

3.1 model

Image inpainting models fill in missing pixels of photographic images to make the completed images look realistic. Context Encoder [1], Globally and Locally Consistent Image Completion (GLCIC) [4], Generative Image Inpainting with Contextual Attention [9] and Free-Form Image Inpainting with Gated Convolution [13] have shown meaningful achievement in image inpainting problems. However, these models are not optimal to restore non-photographic images or images from closed image dataset such as lane geometry object images. Generative Image Inpainting with Contextual Attention [9] and Free-Form Image Inpainting with Gated Convolution [13] failed to restore any lane geometry. Context Encoder [1] and vanilla GLCIC constructed very blurry lane lines (e.g., Figure 4A). We extended the architecture of vanilla GLCIC [4] and experimented various GAN loss functions to achieve high image quality in restored images. Our model architecture has 8 dilated convolution layers with increasing dilation factor in the middle layer of the model. It made our model can learn the geometric relationship between the geometric shape of inpainting area and lane shape in far distant in images. Unlike photographic images in open image dataset, the restored lane objects need to be harmonized with not only near distanced objects but also those in far distance in images. Figure 3 shows the model architecture for lane geometry objects inpainting model. We implemented an incorrect area detection model (in Figure 2) which detects the wrong area in lane geometry images. Once the wrong area is identified with its area index, the mask is generated and put on the area to be corrected. The incorrect area detection model is a type of multi-label classifier. The detail specification about incorrect area detection model is out of scope in this paper.

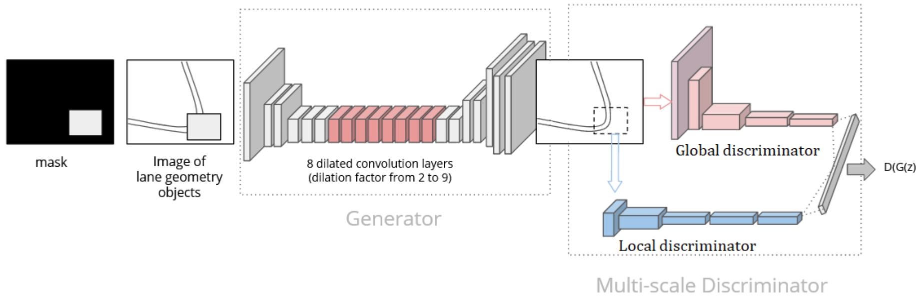

Figure 3: Architecture of lane geometry object image inpainting model

3.2 Loss function

Our generator network trains with perceptual loss (1) [6]. The completion network in vanilla GLCIC [4] used Mean Square Error (MSE) as a loss function. However, we believe that MSE loss is not optimal to tell the difference between sharp lane geometry lines and blurry lines. Perceptual loss computes the Euclidean loss (L2-loss) on the VGG19 [15] feature maps. It noticeably increased the perceptual quality of restored objects in images and perceptual loss helped the model can generate the correct texture distortion along the object lines.

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

where $\phi_{i,j}$ is the feature map obtained by the j-th convolution and before the i-th maxpooling layer within the VGG19 network [16], pretrained on ImageNet [11], $W_{i,j}$ and $H_{i,j}$ are the dimensions of the feature maps. I^O is a original correct image and I^R is input image to be restored by model.

On discriminator networks - both global discriminator and local discriminator, we adopt Relativistic Least Square GAN (RaGAN-LS) loss [7, 8, 10] (2) as a loss function instead of Binary Cross Entropy (BCE) loss. RaGAN-LS loss serves to achieve better contextual quality in restored region. The restored lane geometry lines show better line-smoothness and presented sharper lines. Besides, it made the discriminator network training faster to converge than training with BCE loss. Our model showed not only the good quality of restored images but also presented more stable learning during the training (Figure 5C).

$$\begin{aligned} L_D^{RaLSGAN} = \mathbb{E}_{x\sim p_{data(x)}}[(D(x) - \mathbb{E}_{z\sim p_{z(Z)}}D(G(z)) - 1)^2] \\ +\mathbb{E}_{z\sim p_{z(Z)}}[(D(G(z)) - \mathbb{E}_{x\sim p_{data(x)}}D(x) + 1)^2] \end{aligned} \tag{2}$$

3.3 Experiments

We experimented 5 different image inpainting models. We trained our models with 9951 lane geometry object images. The training dataset is the output images of lane detection model on aerial images in San Francisco area. The first experiment trained the model of vanilla GLCLC [4] to set the baseline model performance (Figure 5's column A shows training loss in all training phases). The second experiment was model training of our extended model architecture - the model with 8 increasing factored dilated convolution layers in the middle layer. During second experiment, we kept the same loss functions of vanilla GLCIC. The output images from the model in second experiment proved that having the 8 dilated convolution layers introduce better restored image quality. Therefore, from the third experiment, we used this deep dilated convolution layered model architecture as default model architecture. The third model we experimented used the perceptual loss (1) in the generator training and kept the BCE loss in the discriminator network. The fourth model kept using MSE loss in generator training, however we used RaGAN-LS loss (2) in discriminator training. The last model we experimented used perceptual loss (1) in generator training and RaGAN-LS loss (2) during discriminator network training. Figure 5 depicts the training loss from each experiments - baseline model training (Figure A), training loss of model with 8 dilated convolution layers (Figure B) and training loss with perceptual loss in generator and RaGAN-LS loss in discriminator (Figure C). We followed the similar training scheme as vanilla GLCIC [4] i.e., training only generator network in phase 1 and training only discriminator network in phase 2. Phase 1 and phase 2 training were served as pre-training, and it proved that pre-training is critical for successful training [4]. The phase 1 required 90000 iterations and phase 2 trained with 10000 iterations. After pre-training on each network, both generator and discriminator were trained alternately with 400000 iterations in phase 3. We trained lane geometry object images in Berlin area with same experiment settings. Compared to San Francisco dataset, the lane geometry object images in Berlin area (e.g., Figure 2) contain more lane lines with complex lane shapes. However, we obtained similar model performance as trained models with San Francisco' lane geometry images.

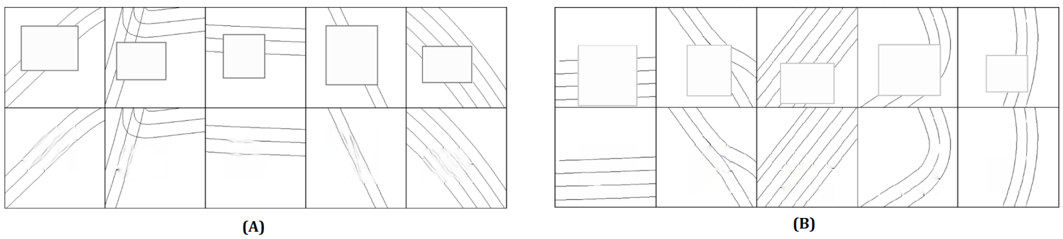

Figure 4: (A) lane geometry object images with vanilla GLCIC (B) lane geometry object images with our image inpainting model

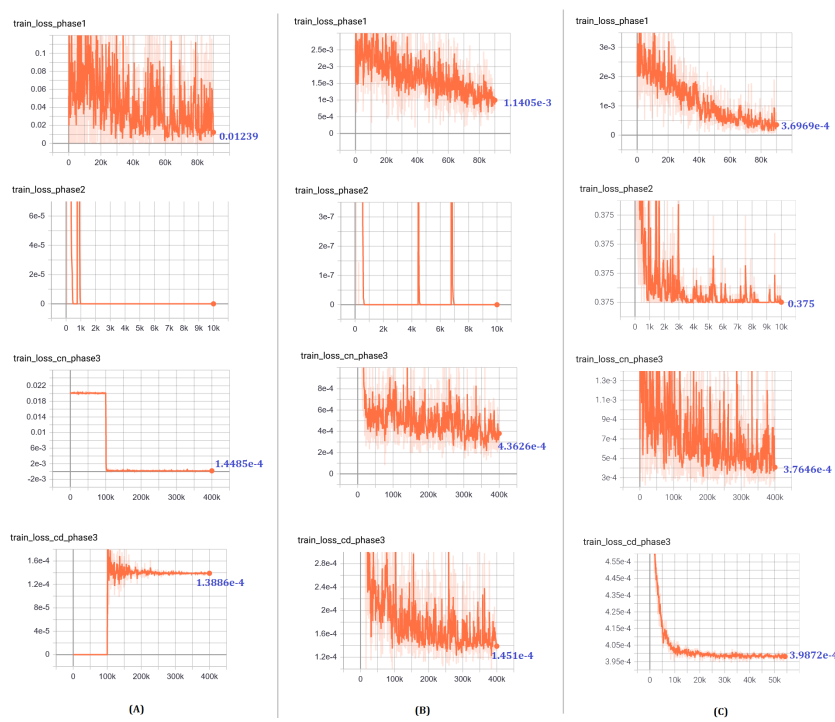

Figure 5: Training loss comparison - column A: model trainig with vanila GLCIC, column B: 8 dilated convolution layered image inpainting model, column C: Same model architecture as column B with perceptual loss and Relativistic Least Square GAN loss.

4 Conclusion

We presented a novel approach to systematically correct wrong lane geometry object images. Our approach can restore any shape of lane geometry objects; thus, it showed the generalization capability compared to correction methods with interpolation algorithms. We experimented various model architectures and examined efficient loss functions to restore pixels in incorrect lane geometry object images. The proposed model architecture and adopted loss functions prove that restored images obtain high visual quality without losing overall geometric context.

References

[1] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros, "Context Encoders: Feature Learning by Inpainting," In *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] Kendall Atkinson, "Modelling a Road Using Spline Interpolation", 2002

[3] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li, "High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis," In *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[4] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2017. Globally and Locally Consistent Image Completion. *ACM Trans. Graph.* 36, 4, Article 107, (July 2017), 14 pages.

[5] Ciroski, V. and Park, J., "Lane Line Detection by LiDAR Intensity Value Interpolation", *SAE International Journal of Advanced Practice in Mobility*, 2(2):939-944, 2020

[6] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In "*European Conference on Computer Vision* , 2016.

[7] Orest Kupyn, Tetiana Martyniuk, Junru Wu, Zhangyang Wang , DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. cite arXiv:1908.03826

[8] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734, 2018.

[9] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang, "Generative Image Inpainting with Contextual Attention," In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, 2018.

[10] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Least squares generative adversarial networks, 2016. cite arxiv:1611.04076.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* , 2009.

[12] Fisher Yu and Vladlen Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. In *International Conference on Learning Representations*. 2016

[13] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas Huang, "Free-Form Image Inpainting with Gated Convolution," In *Proceedings of International Conference on Computer Vision (ICCV)*, 2019.

[14] Antonio M. López et al., Robust lane markings detection and road geometry computation *International Journal of Automotive Technology*, June 2010

[15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[16] Orest Kupyn et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Network. arXiv:1711.07064v4. Apr 2018