CyEDA: CYCLE-OBJECT EDGE CONSISTENCY DOMAIN ADAPTATION

Jing Chong Beh [1] Kam Woh Ng [2] Jie Long Kew [1] Che-Tsung Lin [3]
Chee Seng Chan [1] Shang-Hong Lai [4, 5] Christopher Zach [3]

[1] CISiP, Faculty of Comp. Sci. and Info. Tech., Universiti Malaya, Kuala Lumpur, Malaysia [2] CVSSP, University of Surrey, Guildford, U.K.
[3] Dept. of Electrical Engineering, Chalmers University of Technology, Sweden [4] Microsoft AI R&D Center, Taiwan
[5] Dept. of Computer Science, National Tsing Hua University, Taiwan

ICIP 2022

Introduction

Images are often taken under sub-optimal lighting conditions. [1]

Adapt object detection model from normal light (or day time) domain to low light (or nigh time) domain.

[2]

[1]

(a) back-lit (b) non-uniform illumination

(c) weak lighting (d) extremely low light

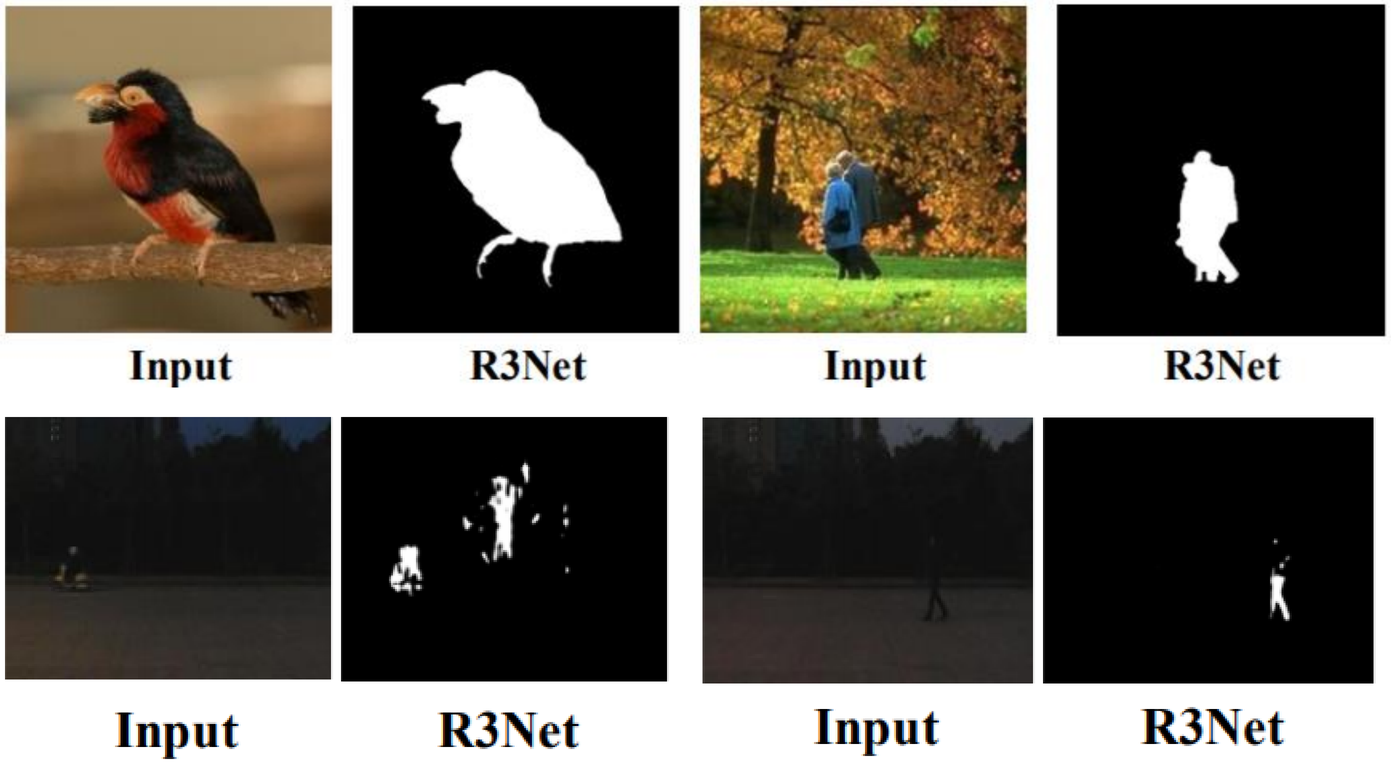
(e) colored light (f) boosted noise

1. Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, & Chen Change Loy. (2021). Lighting the Darkness in the Deep Learning Era.
2. Wenjing Wang, Wenhan Yang, & Jiaying Liu. (2021). HLA-Face: Joint High-Low Adaptation for Low Light Face Detection.

Motivations

1. Images captured under insufficient illumination worsen the performance of machine vision tasks which may cause potential risks in surveillance video analysis and night time autonomous driving. [2]

[3]



2. *Wenjing Wang, Wenhan Yang, & Jiaying Liu. (2021). HLA-Face: Joint High-Low Adaptation for Low Light Face Detection.*
3. Xin Xu, Shiqin Wang, Zheng Wang, Xiaolong Zhang, & Ruimin Hu. (2020). Exploring Image Enhancement for Salient Object Detection in Low Light Images.

Motivations

1. Images captured under insufficient illumination worsen the performance of machine vision tasks which may cause potential risks in surveillance video analysis and night time autonomous driving. [2]
2. Most of the baseline I2I translation models were focused on migrating styles or attributes onto the entire images which result in unrealistic translation when images are content rich while state-of-the-art instance level I2I translation model such as DUNIT integrated detection subnet in the architecture make it not end-to-end trainable.

[5]

Figure 5: Qualitative comparison on Sunny to Night. We show, from left to right, the input image in the source domain, the result of cycleGAN [45] and UNIT [25], and random outputs from MUNIT [12], DRIT [22] and DUNIT (ours), respectively.

4. *Zhiqiang Shen, Mingyang Huang, Jianping Shi, Xiangyang Xue, and Thomas S. Huang. Towards instance-level imageto-image translation. CoRR, abs/1905.01744, 2019.*
5. *Deblina Bhattacharjee, Seungryong Kim, Guillaume Vizier, and Mathieu Salzmann, “Dunit: Detection-based unsupervised image-to-image translation,” in CVPR, 2020, pp. 4786–4795.*

Contributions

In this work, we propose a novel method namely $CyEDA$ to perform global level domain adaptation that can preserve image contents without any pre-trained networks integration or annotation labels.

Specifically, we introduce

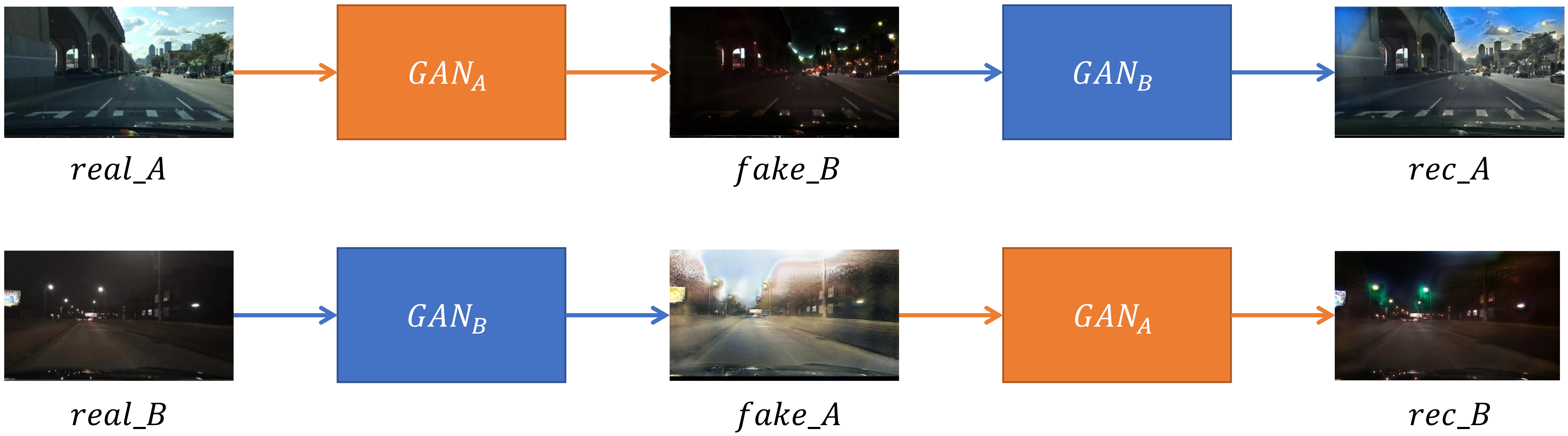
1. blending masks and
2. cycle-object edge consistency loss

which exploit the preservation of image objects

Methodology

Structure

Our work has a similar structure with CycleGAN [6] to leverage its structure strength which does not require paired dataset for training.



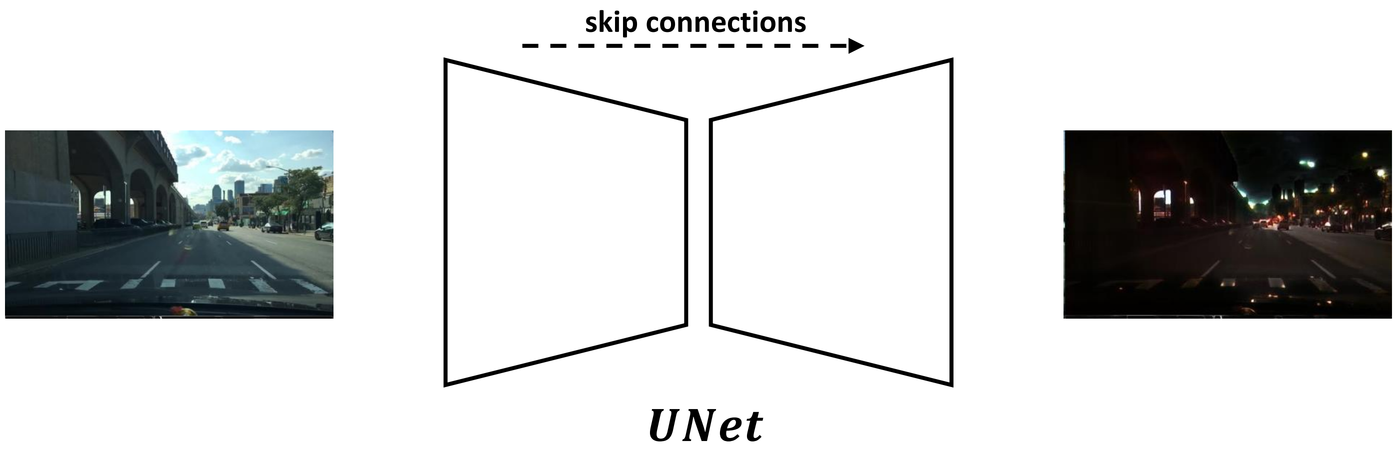
6. *Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in ICCV, 2017.*

Methodology

GAN Backbone (UNet)

We choose UNet [7] as our model GAN backbone and make modifications on it to enforce image content preservation.

Original UNet: Learn to construct whole image in target domain based on information available in hidden embedding



7. *Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, pp. 234–241.*

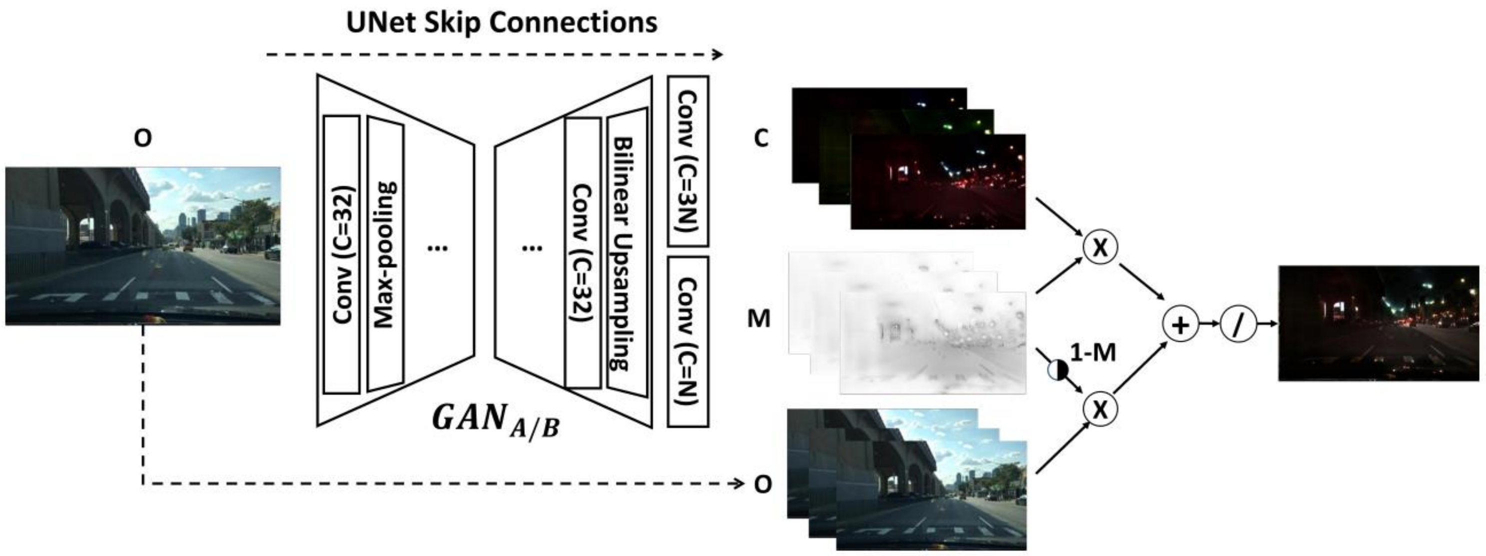
Methodology

GAN Backbone (UNet)

We choose UNet [7] as our model GAN backbone and make modifications on it to enforce image content preservation.

Original UNet: Learn to construct whole image in target domain based on information available in hidden embedding

Mask UNet : Learn to make color changes to every pixel of original image



7. *Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, pp. 234–241.*

Methodology

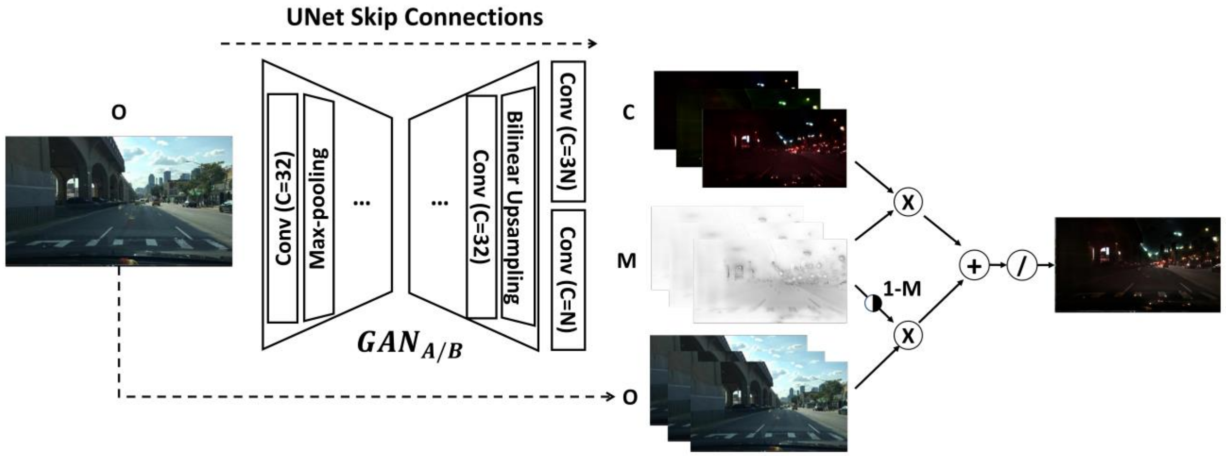
Mask UNet

We modify the last layer of UNet [7] into two separate blocks:

i. a convolution layer that yields $3N$ channels of output followed by $tanh$ activation layer (predicted color changes, C)

ii. a convolution layer that yields N channels of output followed by $sigmoid$ activation layer (predicted degree of color changes / mask, M)

Translated image is computed by summing the weighted mask of color changes and the inverted weighted mask of the original image and normalized by number of mask N



7. *Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015, pp. 234–241.*

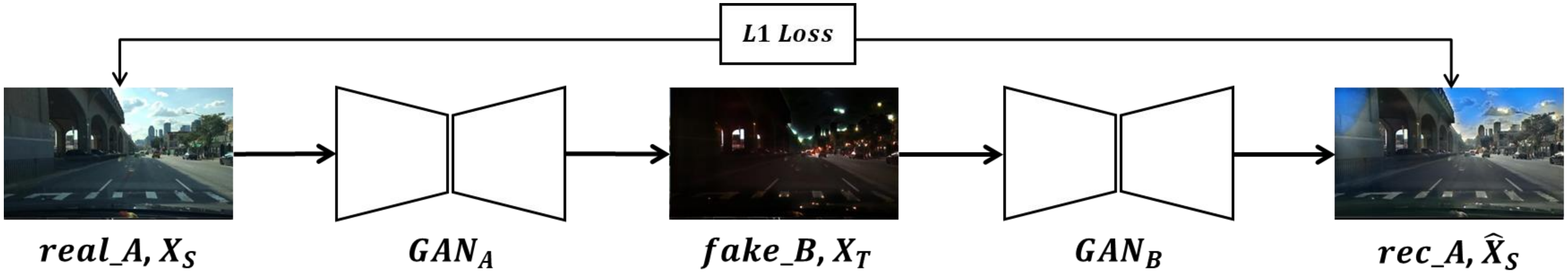
Methodology

Cycle-Object Edge Consistency Loss

CycleGAN learns to hide information in the network to satisfy L1 cycle consistency requirement. [8]

We argue that the cycle consistency loss should only enforce the preservation of objects in the images instead of every pixel details.

We enforce L1 consistency between edge of real and reconstructed images instead of the image itself.



8. *Casey Chu, Andrey Zhmoginov, and Mark Sandler, "Cyclegan, a master of steganography," arXiv preprint arXiv:1712.02950, 2017.*

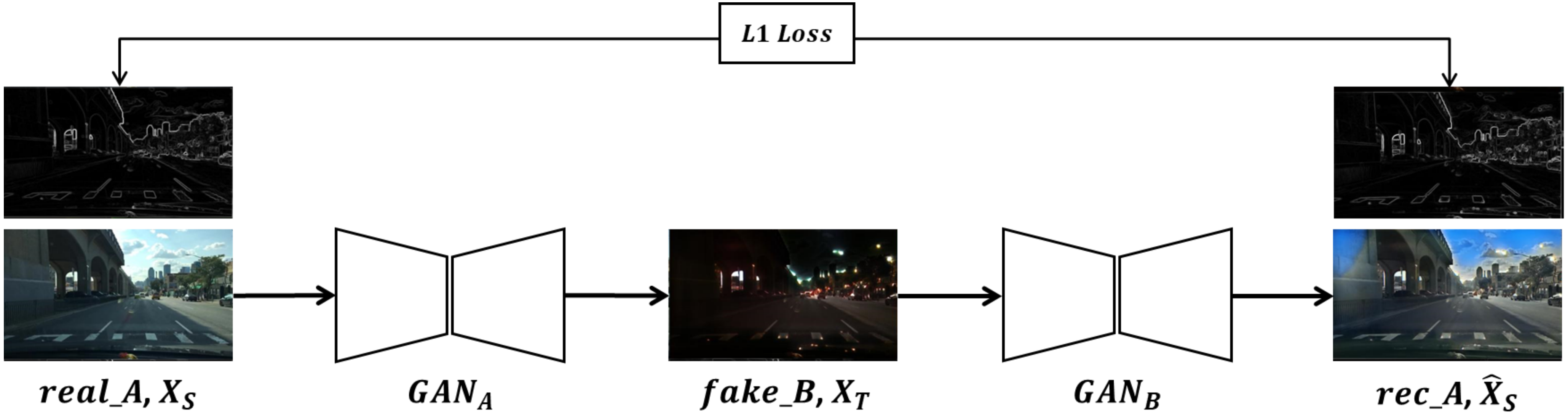
Methodology

Cycle-Object Edge Consistency Loss

CycleGAN learns to hide information in the network to satisfy L1 cycle consistency requirement. [8]

We argue that the cycle consistency loss should only enforce the preservation of objects in the images instead of every pixel details.

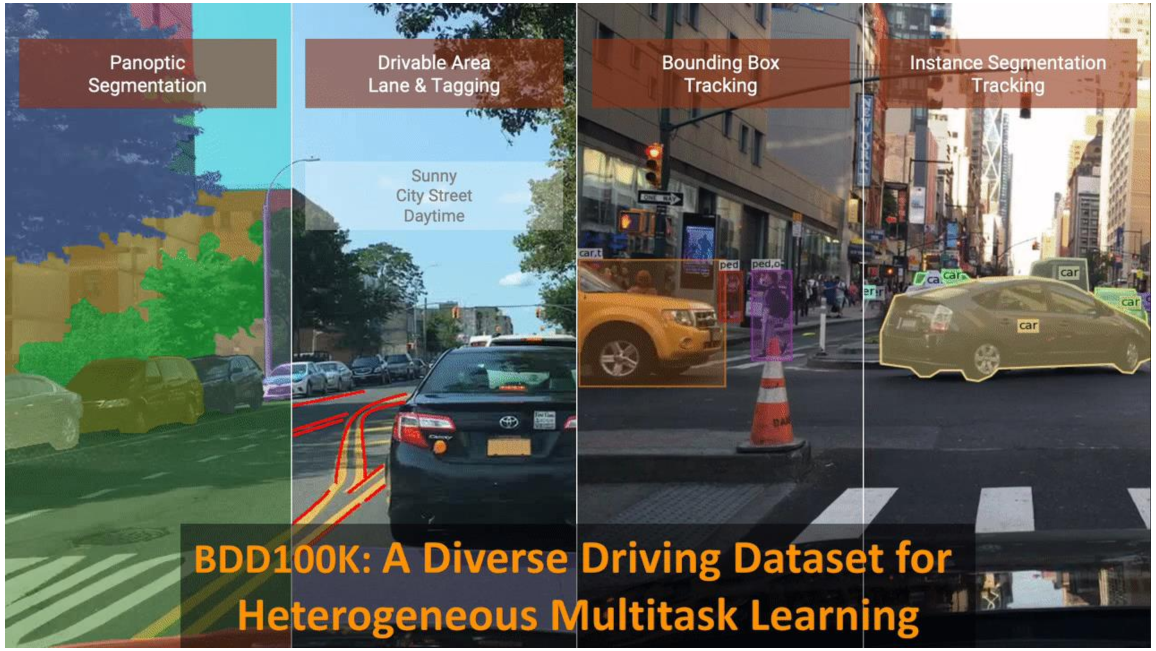
We enforce L1 consistency between edge of real and reconstructed images instead of the image itself.



8. Casey Chu, Andrey Zhmoginov, and Mark Sandler, "Cyclegan, a master of steganography," arXiv preprint arXiv:1712.02950, 2017.

Benchmark Dataset

BDD100k [9] and GTA [10]



9. Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in CVPR, 2020, pp. 2633–2642.
10. *Stephan R. Richter, Zeeshan Hayder, and Vladlen Koltun, "Playing for benchmarks," in ICCV, 2017, pp. 2232–2241.*

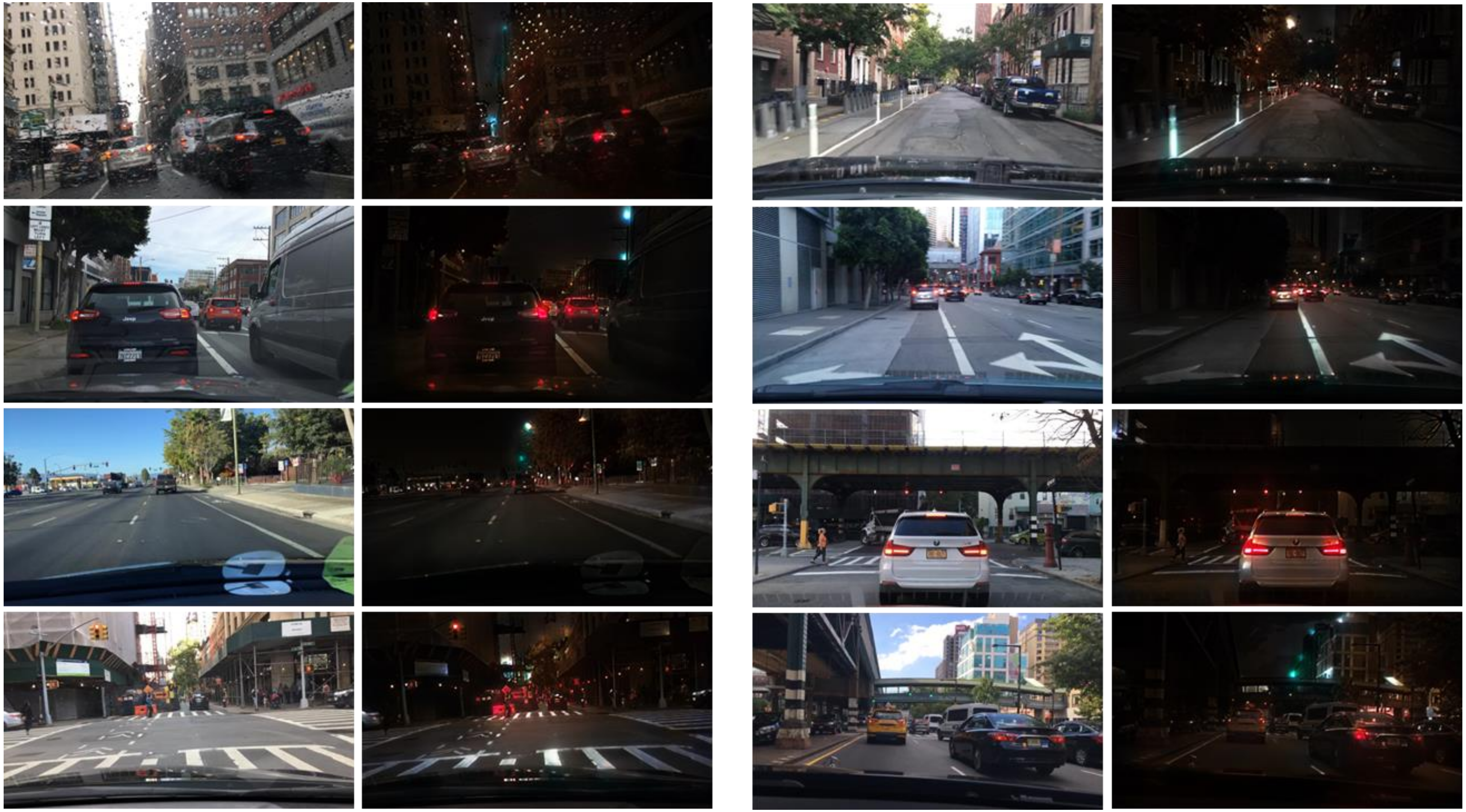
Experiment Results

First row: GTA dataset. Second row: BDD100k

FID score comparison of our approach with SOTAs on GTA (val-night and val-day day-to-night) and BDD100k (det-val-night and det-val-day day-to-night) datasets.

GAN model	annotation?	GTA	BDD100k
MUNIT [16]	No	1.066	2.461
NICE-GAN [15]	No	2.466	1.913
AugGAN [5]	Yes	0.825	0.332
Multimodal AugGAN [6]	Yes	1.023	0.496
CyEDA (This work)	No	**0.737**	**0.297**

BDD100k day-to-night Results



More Comparison

Compared to results presented in DUNIT [5] paper (using same experiment setting) as DUNIT open sourced code is incomplete

CycleGAN

UNIT

MUNIT

DRIT

DUNIT

This work (CyEDA)

5. *Deblina Bhattacharjee, Seungryong Kim, Guillaume Vizier, and Mathieu Salzmann, "Dunit: Detection-based unsupervised image-to-image translation," in CVPR, 2020, pp. 4786–4795.*

Domain Adaptation

Does the generated night time images helpful in training object detection model?

Experiments

1. Train with BDD100k-det-train-night (~2k)
2. Train with BDD100k-det-train-night (~2k) + BDD100k-det-val-day day-to-night (1k)

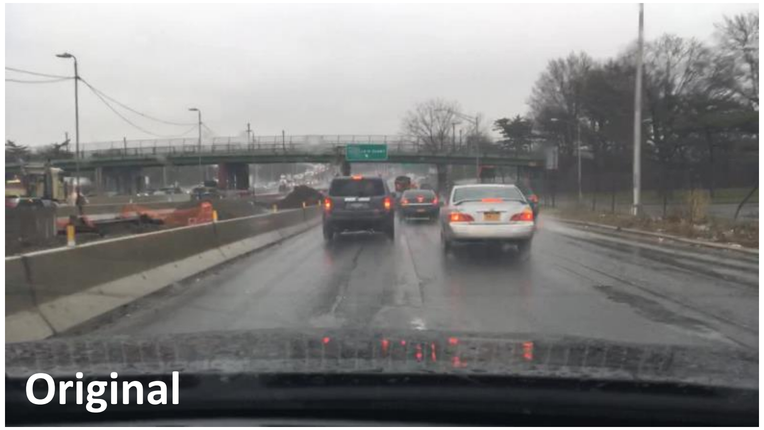
Validate on BDD100k-det-val-night

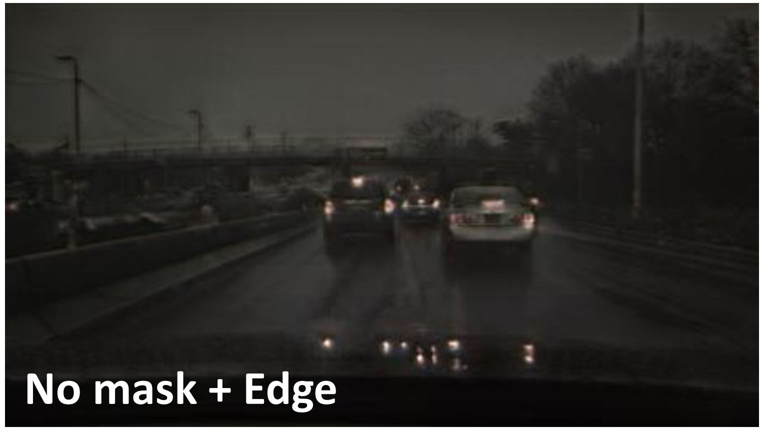
Model: pretrained Yolov5s (on COCO)

Training Dataset	mAP	AP (car)
BDD100k-det-train-night	0.444	0.627
BDD100k-det-train-night + BDD100k-det-val-day day-to-night	**0.465**	**0.644**

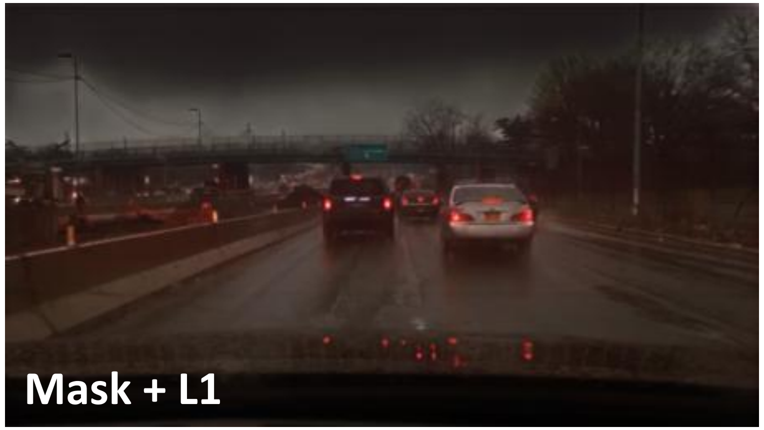
Ablation Study

Mask UNet contributes on **maintaining colour contrast** of translated image as can be seen in Fig.(c) and edge loss **removes unnecessary details** as shown in Fig.(d).





Experiment Setting	**FID score**
No mask + L1	0.551
No mask + Edge	0.579
Mask + L1	0.389
Mask + Edge (CyEDA)	0.297



Conclusion

This paper introduced an approach to retain instance-level detail when translating images to a target domain by

1. generating masks from UNet and performing color fine-tuning on original images according to the masks
2. using cycle-object edge consistency loss to remove unnecessary details and provide extra capacity for model to perform more realistic image-translation