ForkGAN with Single Rainy Night Images:

Leveraging the *RumiGAN* to See into the Rainy Night

Project Page

Github Code

Seri Lee @cma lab

Problem Setting

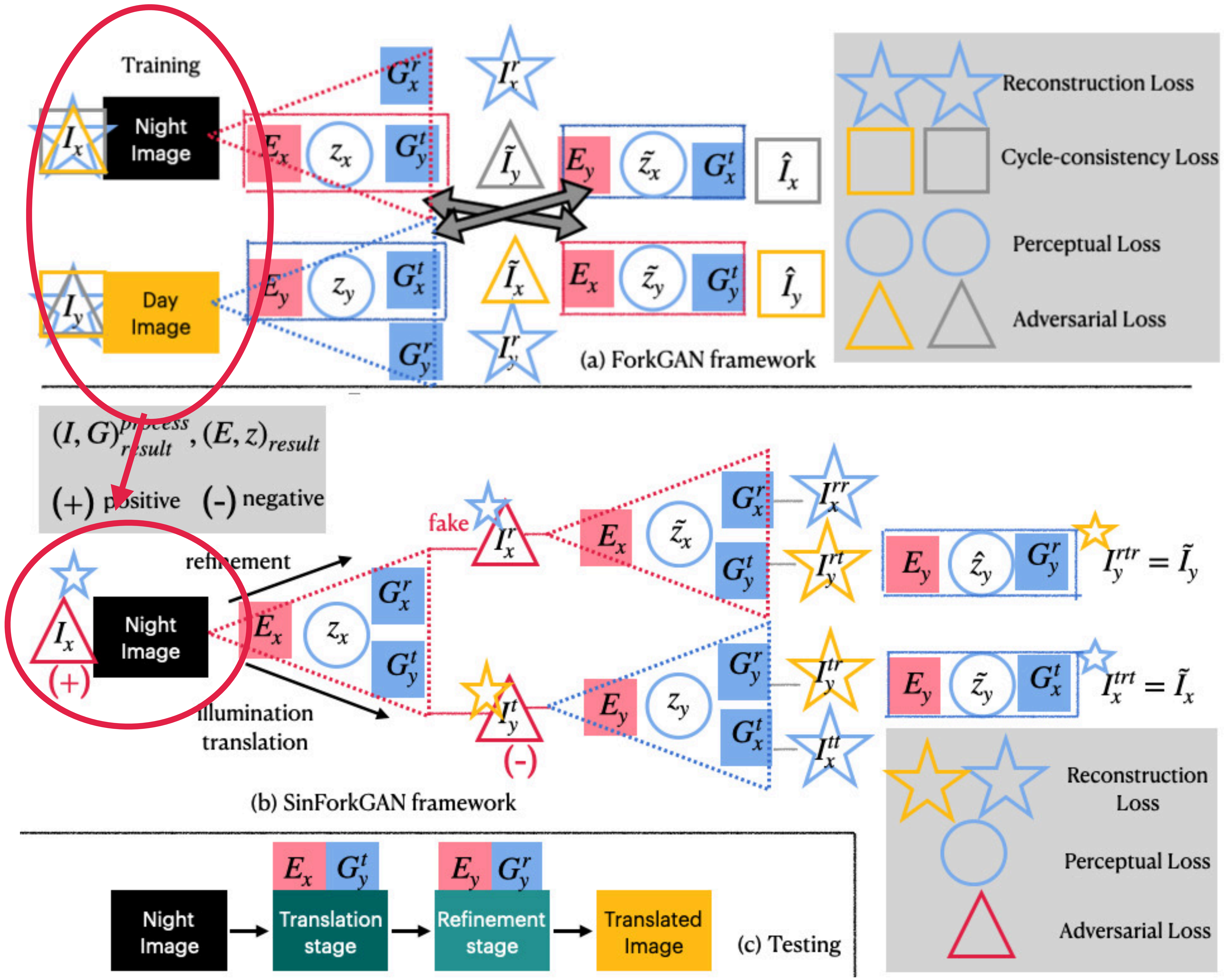
The Problem with ForkGAN

- ForkGAN: task-agnostic unsupervised image translation method that boost the performance of multiple vision tasks in adverse weather conditions represented as "rainy night"
- key idea: fork-shape generator with one encoder and two decoders
- ★ requires some sort of division between nighttime and daytime images in the training set

Improving ForkGAN

ForkGAN meets *the Rumi framework*

- *the Rumi framework*: provide GAN positive data and negative data
- Generator is tasked with learning the distribution of only the positive ones by simultaneously learning to avoid the negative ones
- The discriminator learns to bin the samples into one of three classes: (1) Positives (2) Negatives (3) Fakes



Related Work

The *RumiGAN* Formulation

- Metric learning and self-supervised representation learning involves splitting the data into positive and negative samples for discriminative learning
- triplet loss: the distance of the target from the positive class is minimized, while that from the negative class is maximized
- *RumiGAN framework* provides GAN not only positive data it must learn to model but also negative samples it must learn to avoid
- *Rumi-SGAN:* $L_D^S = -\left(\alpha^+ \mathbb{E}_{x\sim p_d^+}[\log D(x)] + \mathbb{E}_{x\sim p_g}[\log(1 - D(x))] + \alpha^- \mathbb{E}_{x\sim p_d^-}[\log(1 - D(x))]\right)$
- $L_G^S(D^*(x)) = -L_S^D(D^*(x)) + \lambda_p\left(\int_X p_g(x)\,dx - 1\right) + \int_X \mu_p(x) p_g(x)\,dx$
- *Rumi-LSGAN:* $L_D^{LS} = \beta^+ \mathbb{E}[(D(x) - b^+)^2] + \beta^- \mathbb{E}[D(x) - b^-)^2] + \mathbb{E}[(D(x) - a)^2]$

Proposed Method

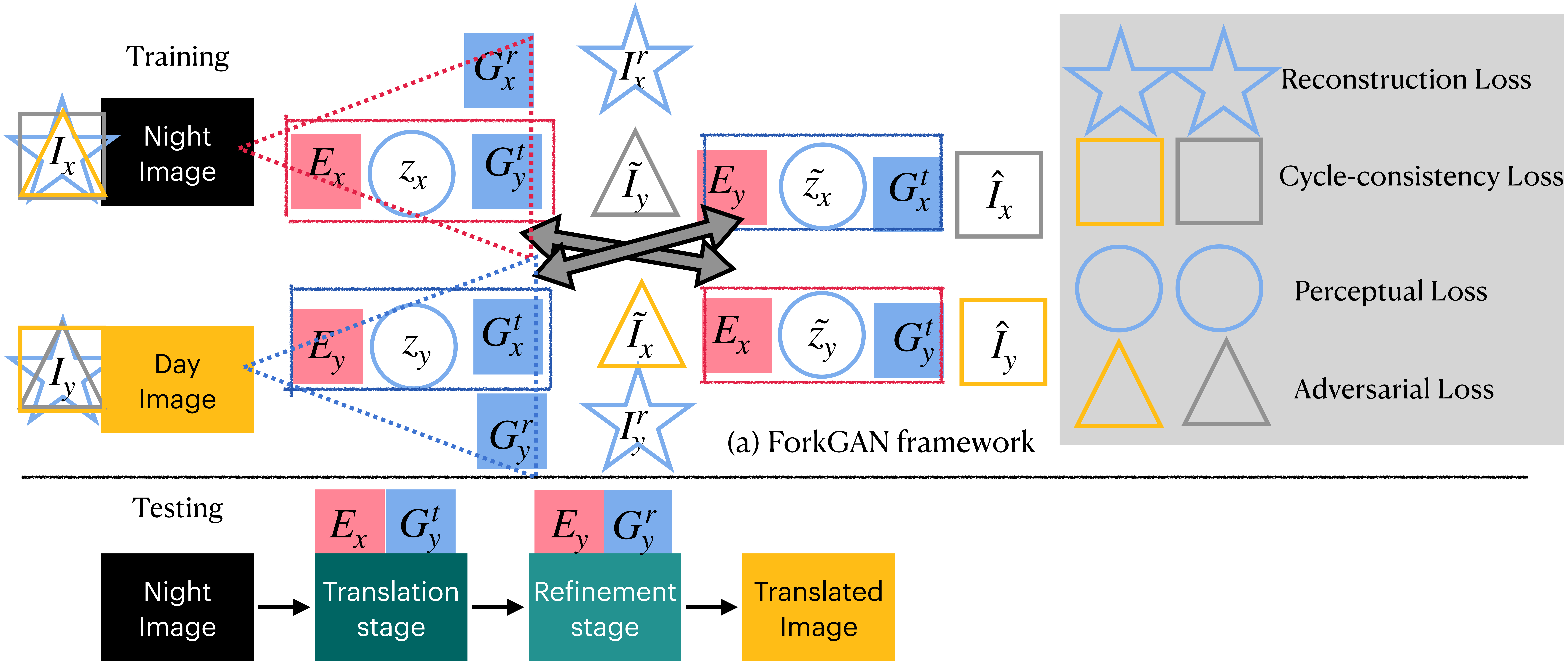
SinForkGAN

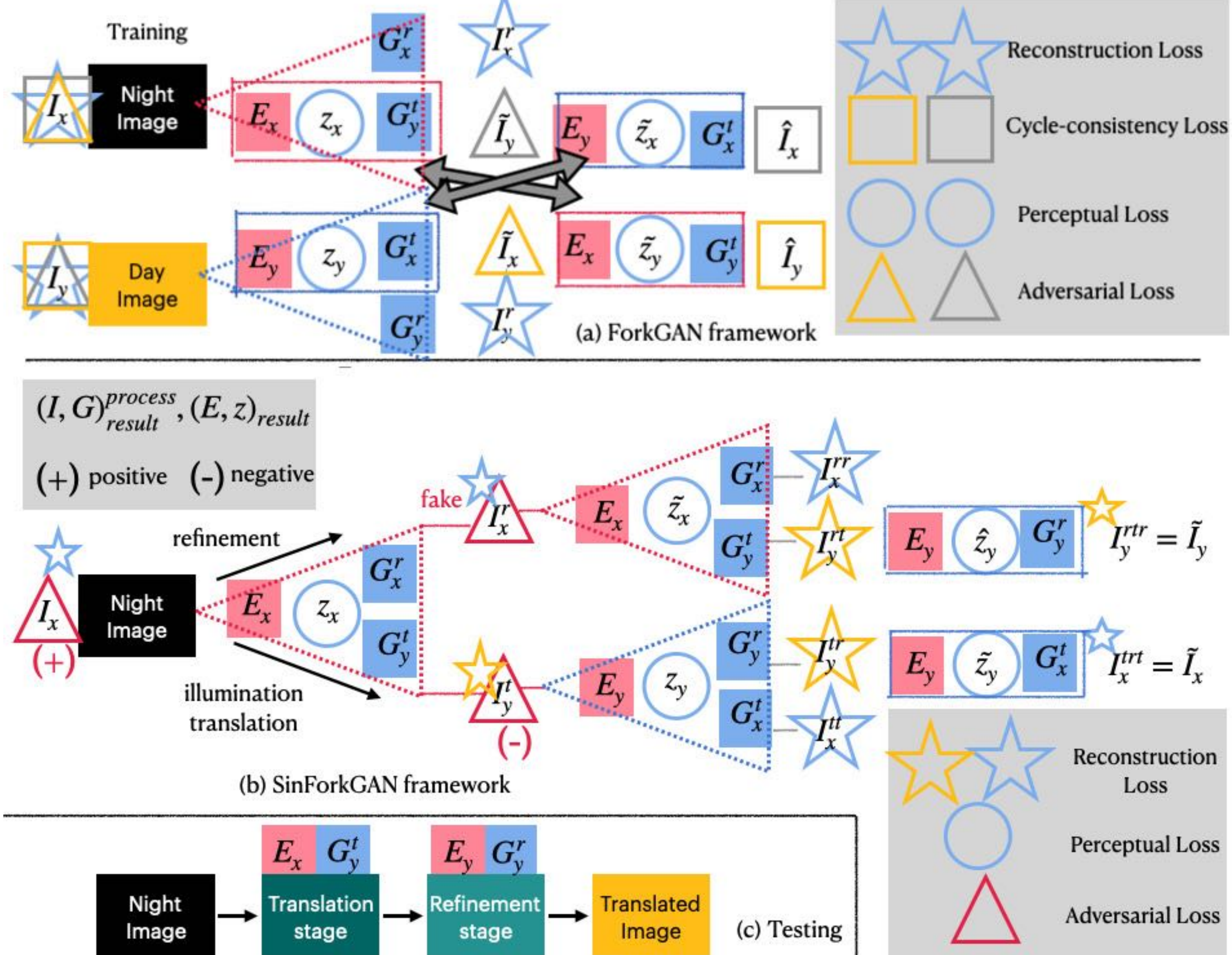
Project Page

Github Code

Seri Lee @cma lab

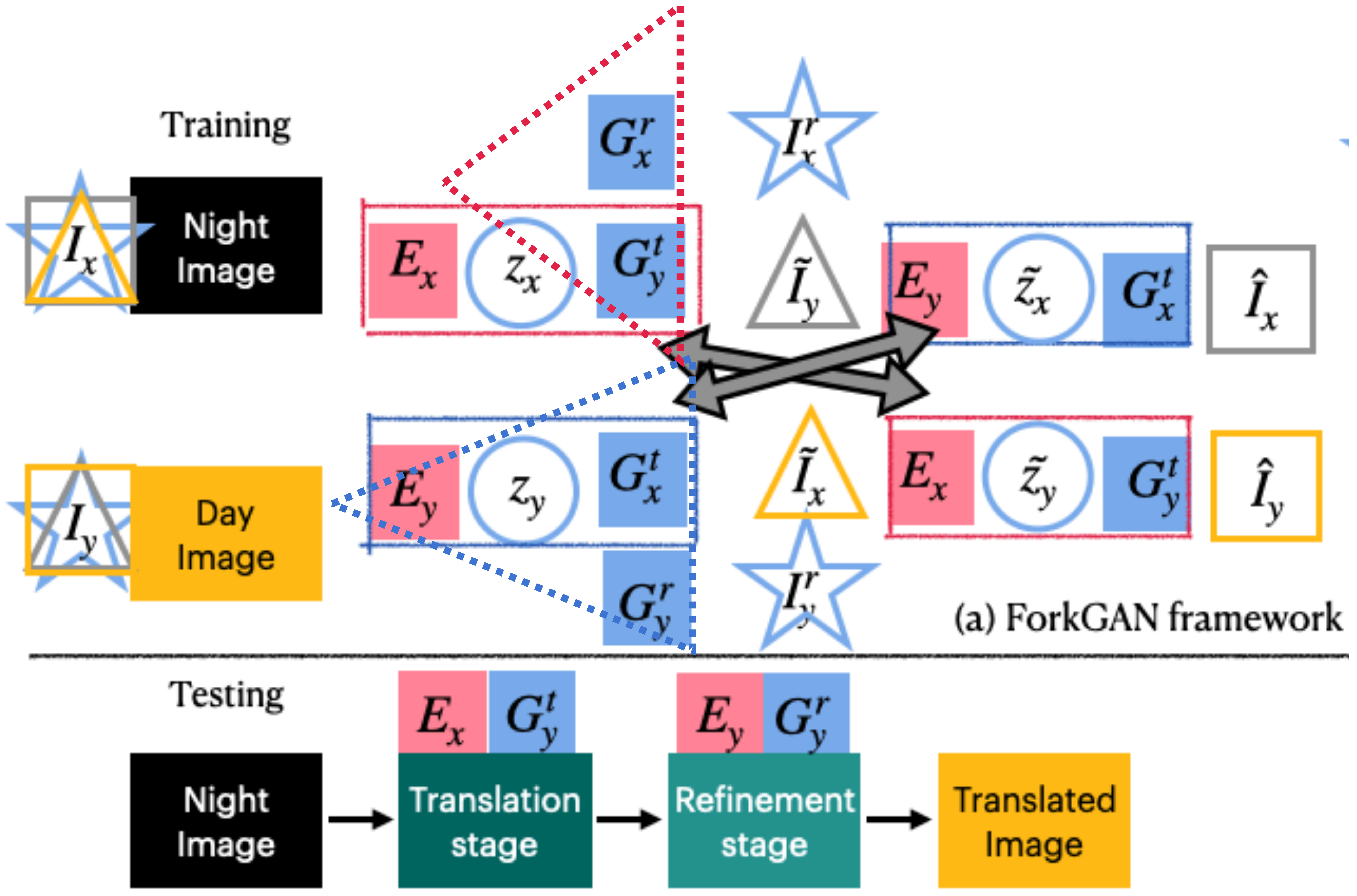
Revisiting the ForkGAN Framework




Training
Night Image
Day Image
Reconstruction Loss
Cycle-consistency Loss
Perceptual Loss
Adversarial Loss
(a) ForkGAN framework
$(I, G)^{process}_{result}, (E, z)_{result}$
(+) positive (-) negative
fake
refinement
illumination translation
Night Image
(+)
(-)
$I_y^{rtr} = \tilde{I}_y$
$I_x^{trt} = \tilde{I}_x$
Reconstruction Loss
Perceptual Loss
Adversarial Loss
(b) SinForkGAN framework
Night Image
Translation stage
Refinement stage
Translated Image
(c) Testing

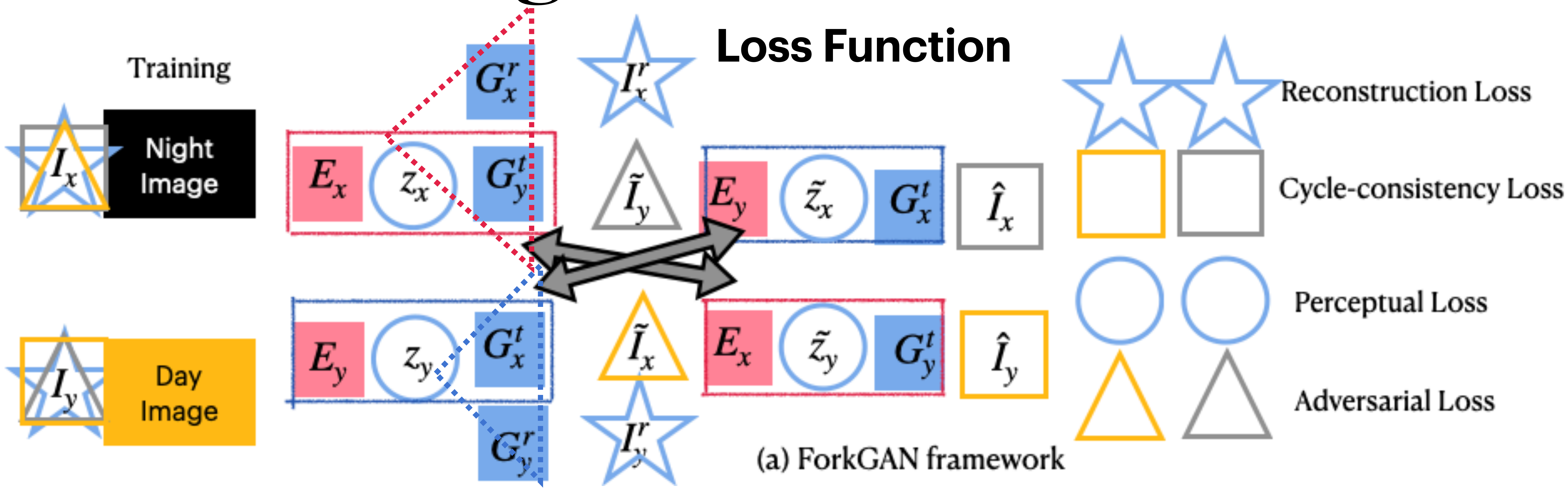
Revisiting ForkGAN Model

$$Y = F_{DOWN}(f_d(I)) = F_{DOWN}(J)$$



- Two separate modules: night-to-day translation module and day-to-night translation module
- Images are fed into encoder and the domain-invariant representation is obtained
- Reconstruction decoder aims to synthesize the original image
- Translation decoder generates translated image
- Adversarial loss to distinguish between the random real image and the translated iamge
- Cycle-consistency loss between the original and final translated-translated image
- 2 stage translation and refinement procedure at inference

Revisiting the ForkGAN Framework



(a) ForkGAN framework

$$L(E, G^r, G^t) = L_{adv} + L_{cls} + L_{per} + \gamma L_{cyc} + \epsilon L_{rec}$$

extract domain-invariant content

$$L_{adv} = \mathbb{E}[\log D(I_x)] + \mathbb{E}[\log(1 - D(\tilde{I}_x))]$$

$$L_{per} = \tau\left(\sum_{n=1}^{N} \lambda_n \|\Phi_n(\tilde{z}_x) - \Phi_n(z_x)\|_1\right)$$

pixel-level L1 norm

$$(\gamma = \epsilon = 10)$$

Experiments

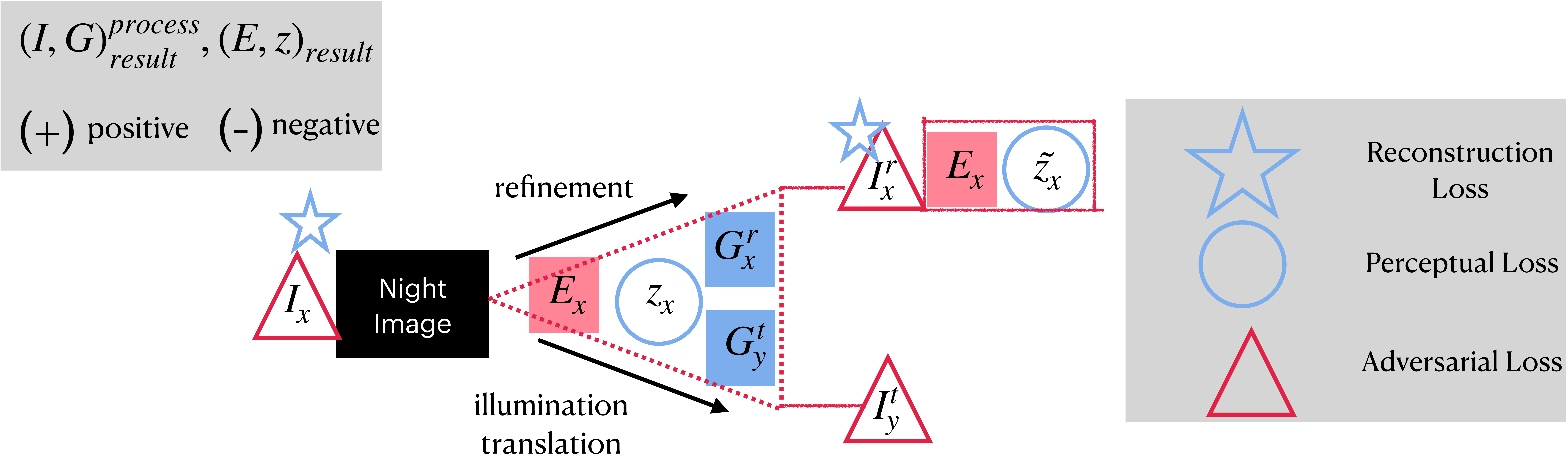
Ablation Study

Project Page

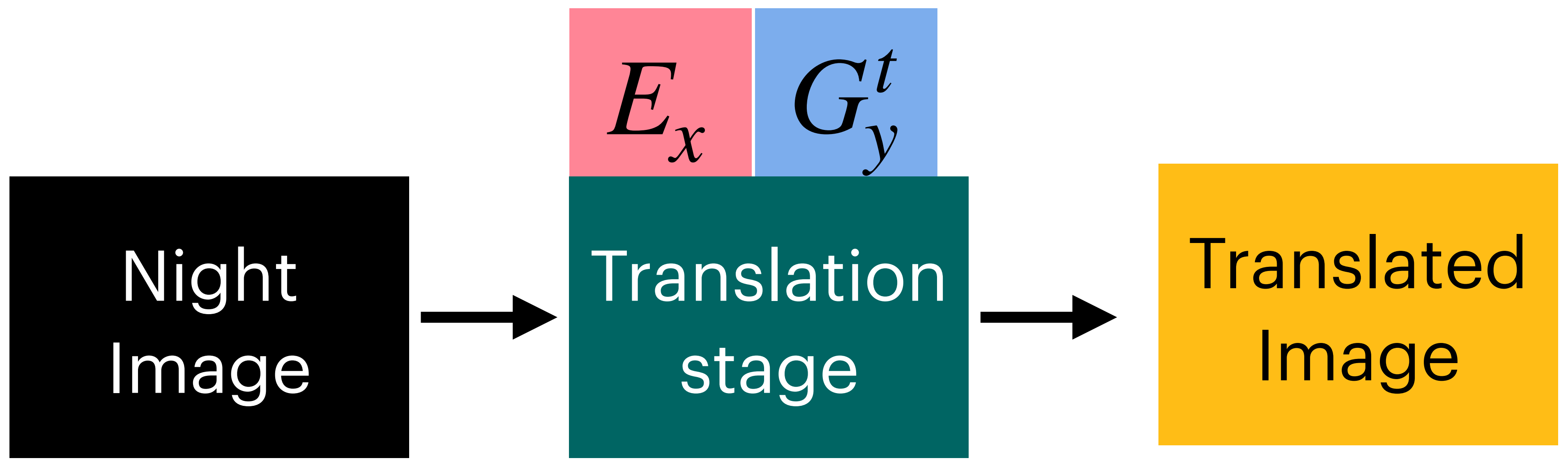
Github Code

Seri Lee @cma lab

The Baseline ForkGAN framework



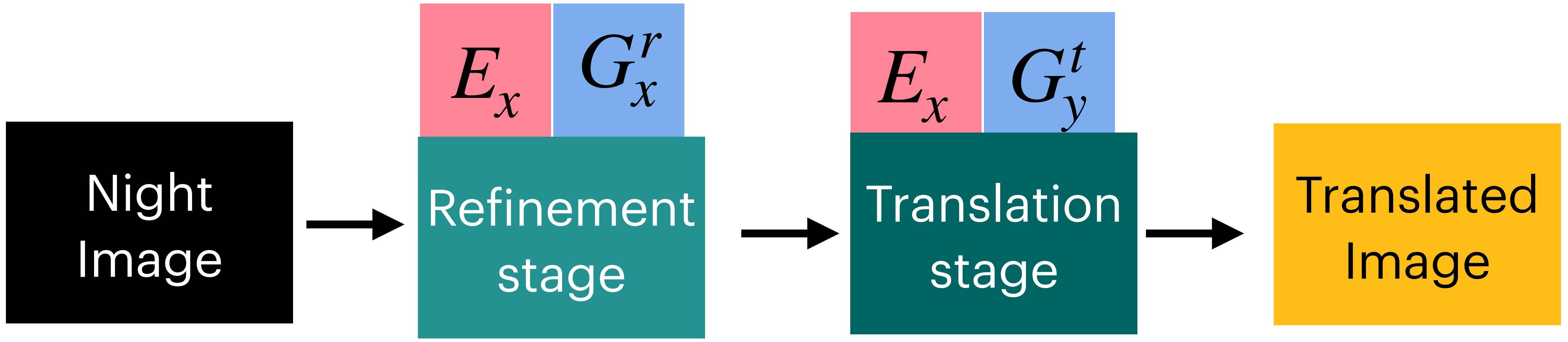
(b) Baseline SinForkGAN



(c) Testing

The Reduced SinForkGAN framework

(b) SinForkGAN framework



(c) Testing

The Proposed SinForkGAN Framework

(b) SinForkGAN framework

(c) Testing

The Proposed SinForkGAN Framework

Loss Function

- unique adversarial loss leveraging the RumiGAN framework
- modified perceptual loss and reconstruction loss according to adapted structure

$$L(E, G^r, G^t) = L_{adv} + \beta L_{per} + \gamma L_{rec}$$

$$L_{rec} = (\sum_{w \in W} \sum_{v \in V} \|w - v\|_1)$$

$$W = \{I_x, I_x^r, I_x^{rr}, I_x^{tt}, \tilde{I}_x\}$$

$$V = \{I_y^t, I_y^{rt}, I_y^{tr}, \tilde{I}_y\}$$

pixel-level L1 norm between two elements in

$$L_{adv} = -[\alpha^+ \mathbb{E}[\log D(I_x)]$$
$$+ \mathbb{E}[\log(1 - D(I_x^r)] + \alpha^- \mathbb{E}[\log(1 - D(I_y^t))]]$$

$$L_{per} = \tau(\sum_{x,y \sim U} \sum_{n=1}^{N} \|\Phi_n(x) - \Phi_n(y)\|_1$$

$$U = \{z_x, \tilde{z}_x, z_y, \tilde{z}_y, \hat{z}_y\}$$

The Proposed SinForkGAN Framework

- $\alpha^- \geq \alpha^+ - 1, \alpha^+ \in [0,1]$, $\alpha^+ = 0.2, \alpha^- = -0.8$
- $(\epsilon = 1, \beta = 1, \gamma = 1)$
- dilated residual networks for the generator to adopt a large receptive field to alleviate occlusion issues
- multi-scale discriminator architecture to improve the ability to distinguish the fake, positive, and negative images

Results

Visual ablation study on Dark Zurich dataset (ICCV 2019)

(a) Input image

(b) ForkGAN

(c) baseline

(d) reduced SinForkGAN

(e) Proposed Method

Experiments

Downstream Task, Dataset and Implementation Detail

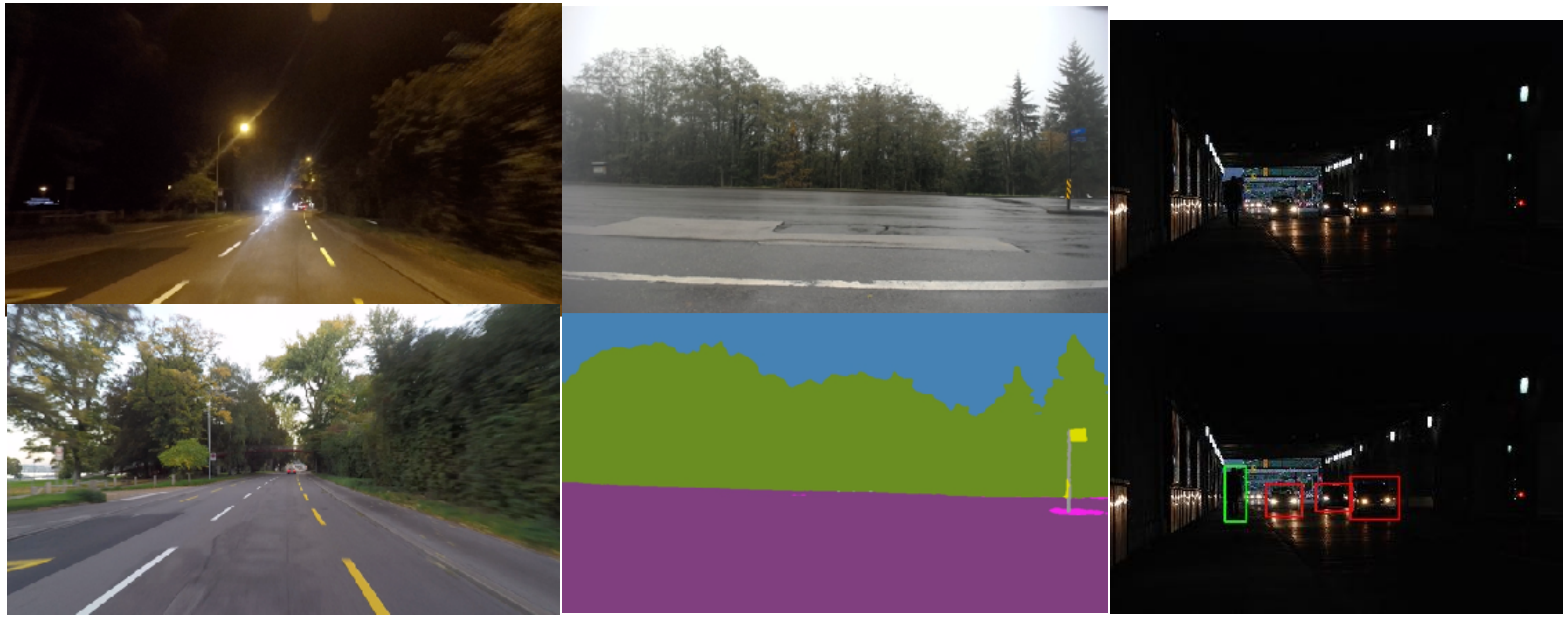
Project Page

Github Code

Seri Lee @cma lab

Experiments

Vision Task Metrics



- Localization: SIFT interest point matching
- Semantic Segmentation: Intersection-over-Union(IoU) (overlap between predicted segmentation map and the ground truth)
- Object Detection: mean average precision (mAP)

Experiments

Datasets Including Adverse Weather Conditions

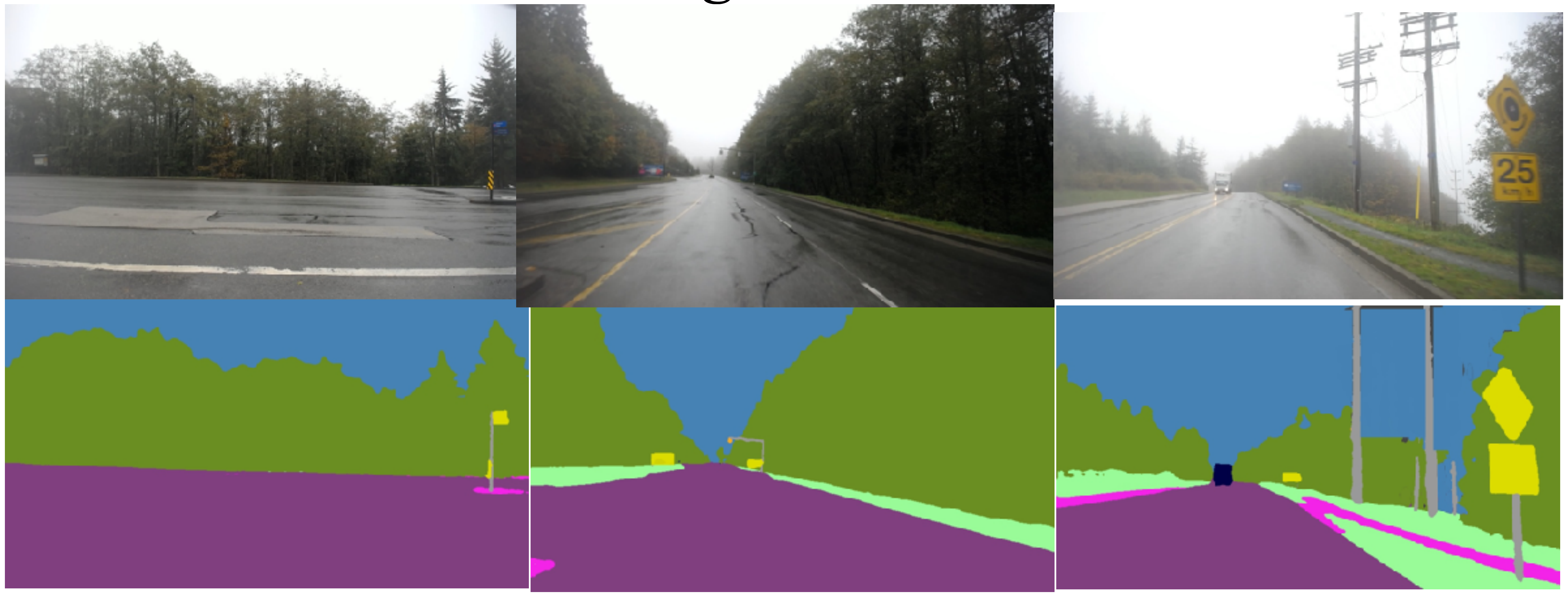
- Dark Zurich Dataset (ICCV 2019): 2,416 nighttime images along with the respective GPS coordinates of the camera for each image used to construct cross-time-of-day correspondences for evaluation on localization task

Nighttime images (the first row) and corresponding daytime images (the second row) for SIFT feature matching

Experiments

Datasets Including Adverse Weather Conditions

- RaidaR (CVPR 2021): a rich annotated dataset of rainy street scenes, 5,000 images provide semantic segmentations and 3,658 provide object instance segmentations used for evaluation on semantic segmentation task

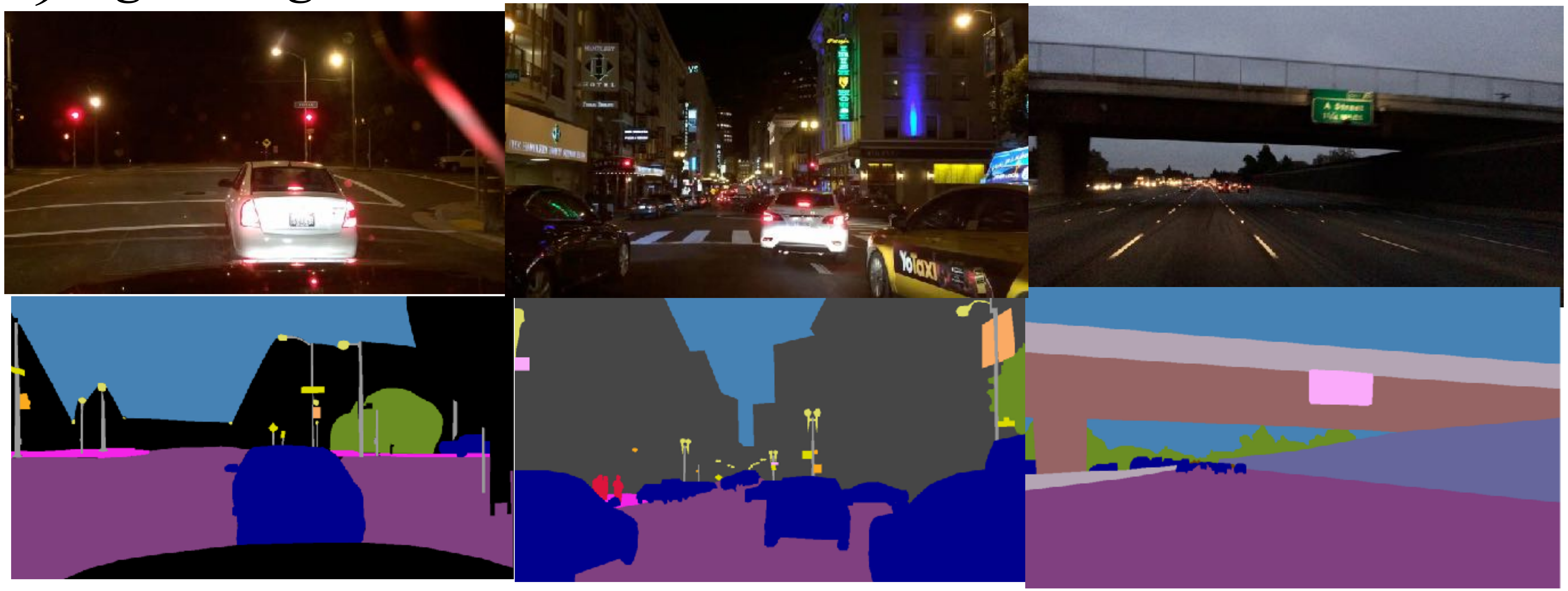


Original images of rainy street scenes (the first row) and corresponding color-coded labels (the second row) for semantic segmentation task

Experiments

Datasets Including Adverse Weather Conditions

- BDD100K (CVPR 2017): 100,000 video clips in multiple cities, multiple weathers and multiple times of day (27,971 night images for training) for semantic segmentation task (3,929 night images for evaluation)



Original images (the first row) and corresponding color-coded labels (the second row) for semantic segmentation task

Experiments

ExDark (CVIU 2019): Exclusively Dark Image Dataset

- ExDark (CVIU 2018): 7,363 low-light images from very low-light environments to twilight with 12 object classes annotated on local object bounding boxes for object detection task

Original low-light images (left) and visualization of local object bounding boxes annotations (right) for object detection task

Experiments

Image Localization/Retrieval Task

Project Page

Github Code

Seri Lee @cma lab

Results

Dark Zurich (ICCV 2019) image translation results achieved by ForkGAN, SinForkGAN

(a) Input nighttime image

(b) ForkGAN translation

SinForkGAN preserves better detail

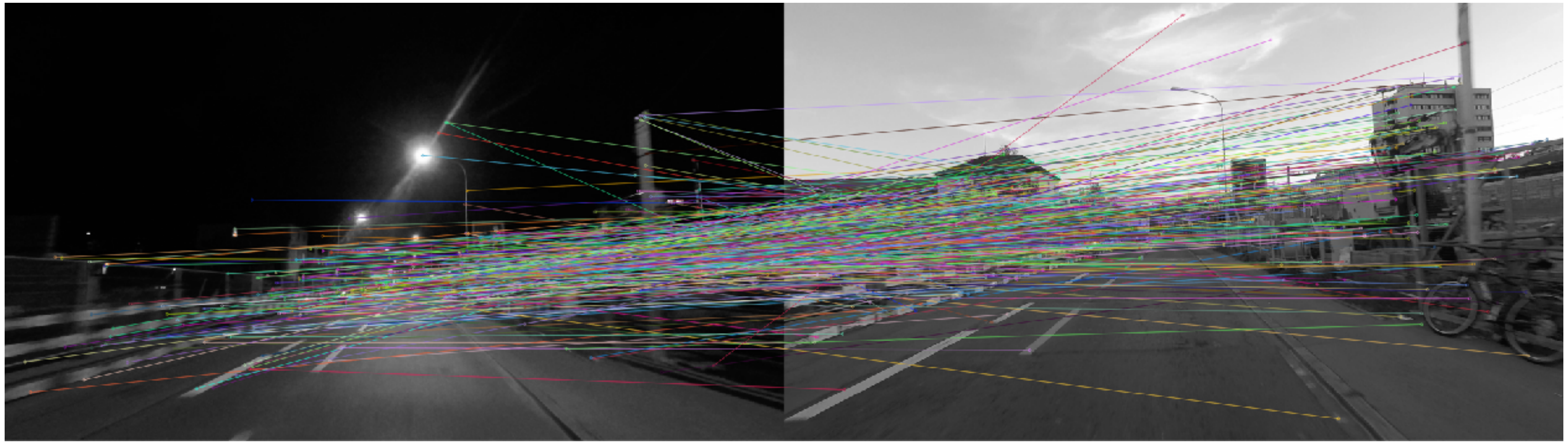
(c) Ground truth daytime image

(d) SinForkGAN translation

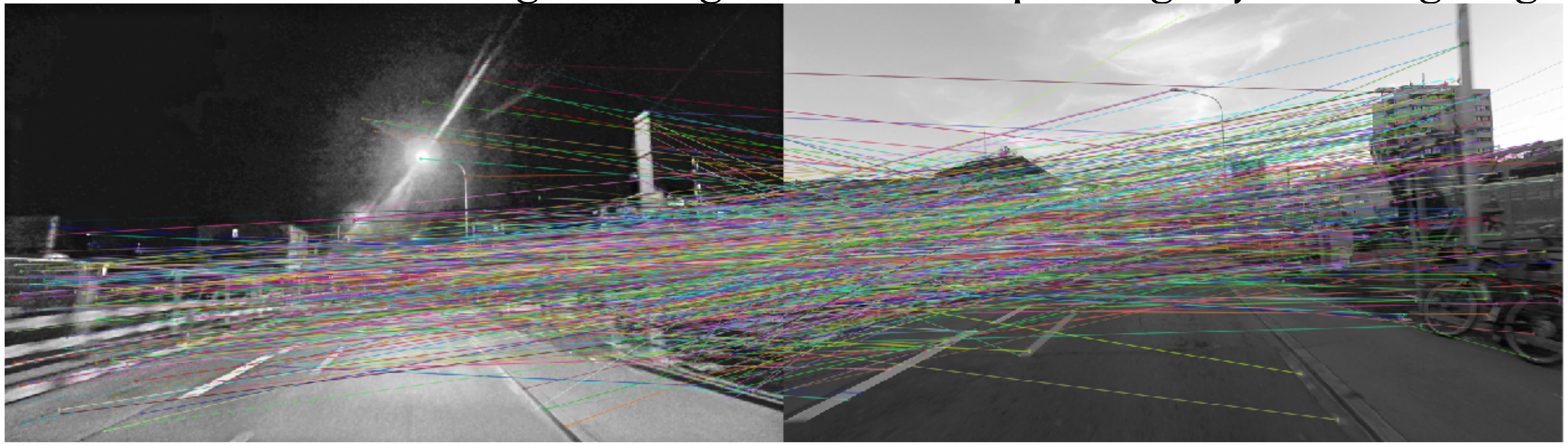
Results

Dark Zurich (ICCV 2019) SIFT feature matching demo

much denser matching points



(e) Without SinForkGAN (Original image (left) - Corresponding daytime image (right))



(f) With SinForkGAN Module (Translated image (left) - Corresponding daytime image (right))

Experiments

Localization by SIFT point matching

- compute the number of SIFT matching points between the translated images and the corresponding natural daytime images
- ForkGAN++ model tested on Dark Zurich dataset (2,416 nighttime images)
- by improving SIFT matching, it can benefit place recognition and visual localization

Method	Original	ToDayGAN
SIFT	8.12	11.8

Method	ForkGAN	SinForkGAN
SIFT	12.1	12.3

Experiments

Semantic Segmentation Task

[Project Page]
[Github Code]
[Seri Lee @cma lab]

Experiments

Semantic Segmentation

- used pretrained Deeplab-v3 model on Cityscapes dataset (no nighttime images) (https://github.com/srihari-humbarwadi/DeepLabV3 Plus-Tensorflow2.0)
- SinForkGAN model tested on BDD100K dataset (137 nighttime images)
- Visual results for PASCAL VOC 2012 + RaidaR dataset
- IOU metric between segmentation outputs and corresponding segmentation ground truth

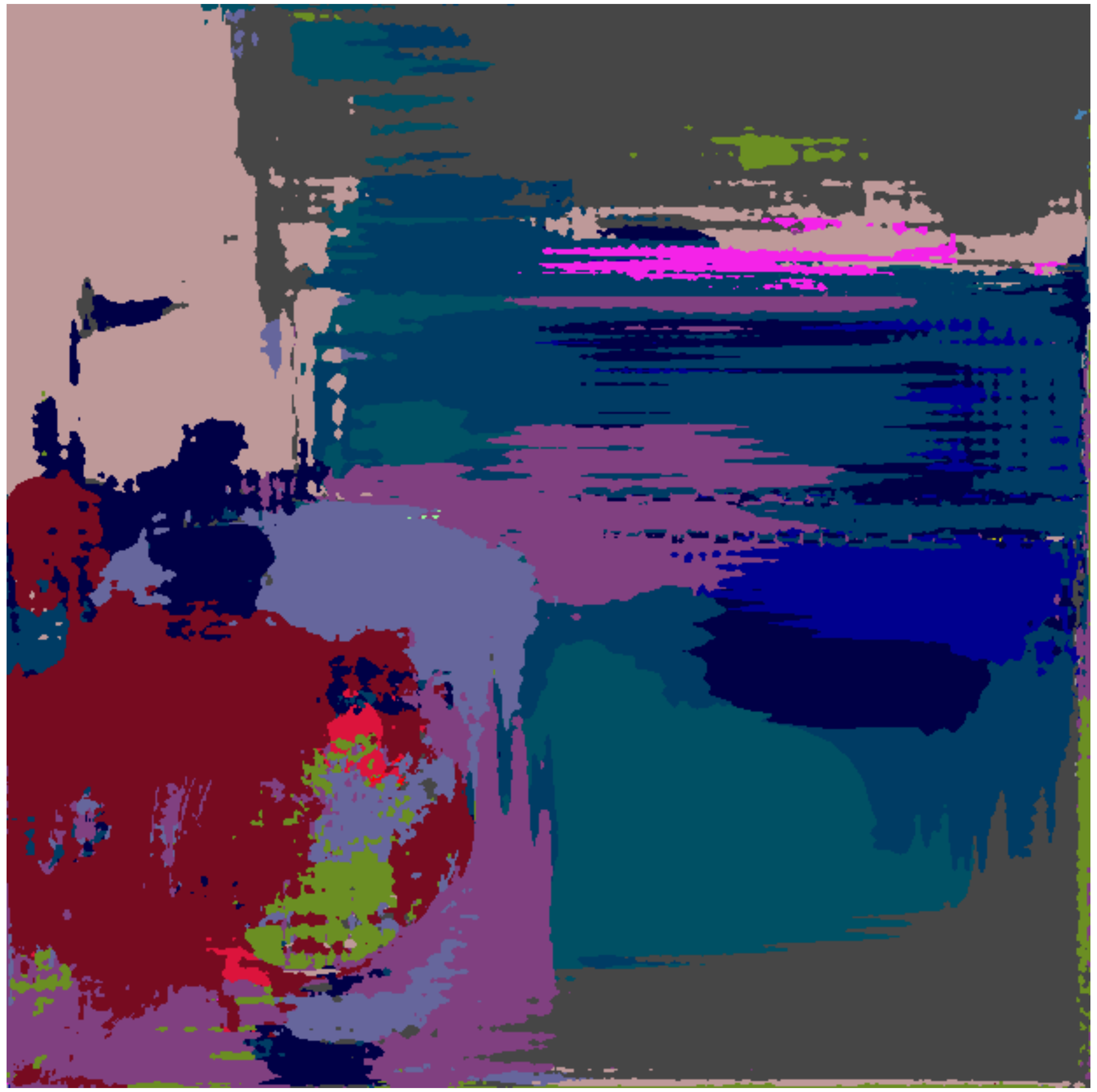
Method	Original	ToDayGAN
mIoU (%)	36.4	41.2

Method	ForkGAN	SinForkGAN
mIoU (%)	41.9	42.3

Experiments

Failure Cases of Deeplab-v3 model tested upon PASCAL VOC 2012 dataset

(a) Original Nighttime image



(b) Deeplab-v3 Segmentation Result

Experiments

Deeplab-v3 model tested on PASCAL VOC 2012 dataset After ForkGAN++ Module

(a) Original Nighttime image (b) After SinForkGAN (c) Final Segmentation Result

Experiments

Failure Cases of Deeplab-v3 model tested on PASCAL VOC 2012 dataset

(a) Original test image

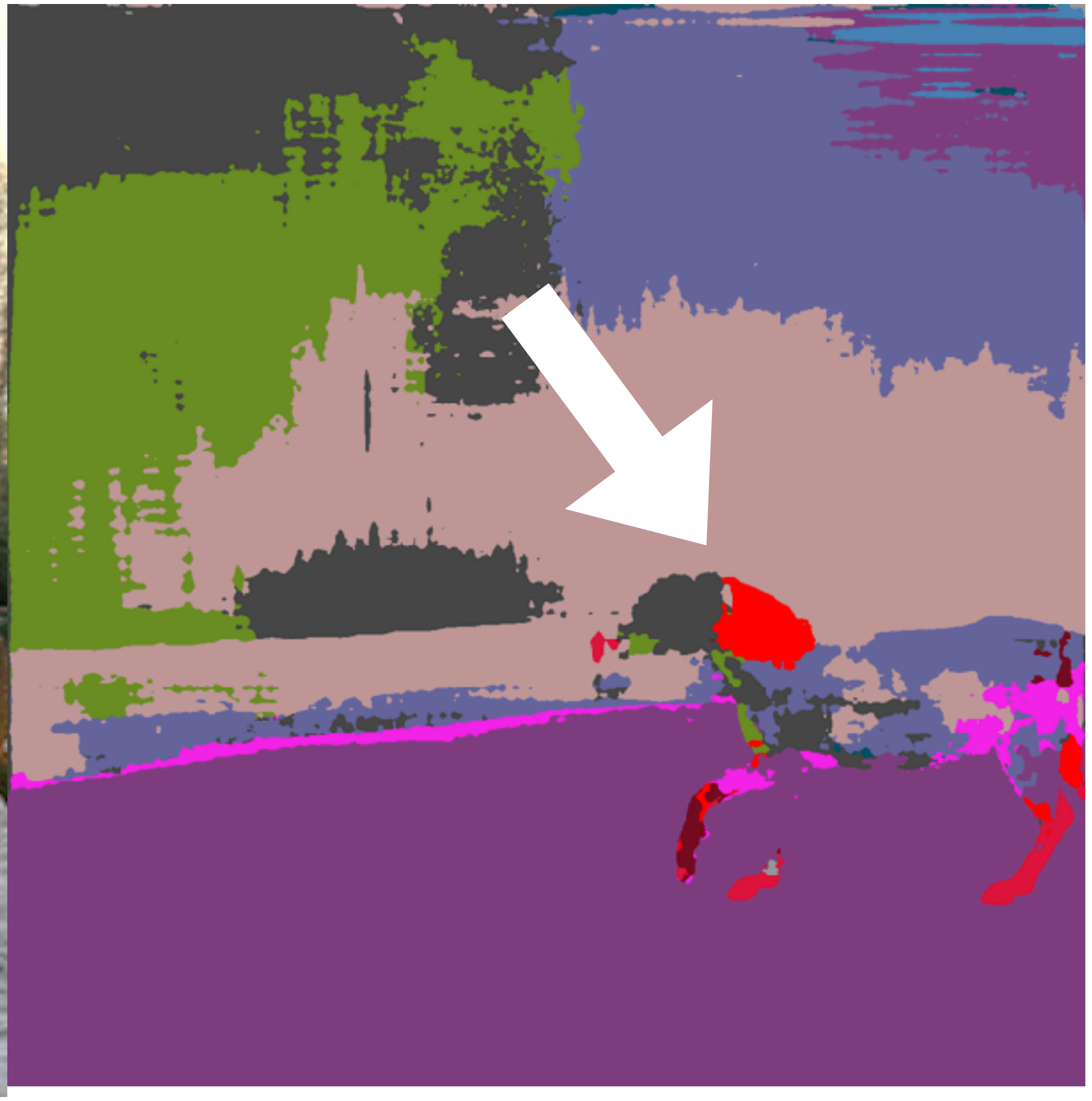
(b) Deeplab-v3 segmentation result

Experiments

Deeplab-v3 model on PASCAL VOC 2012 dataset After ForkGAN++ Module

(a) Original test image

(b) After SinForkGAN

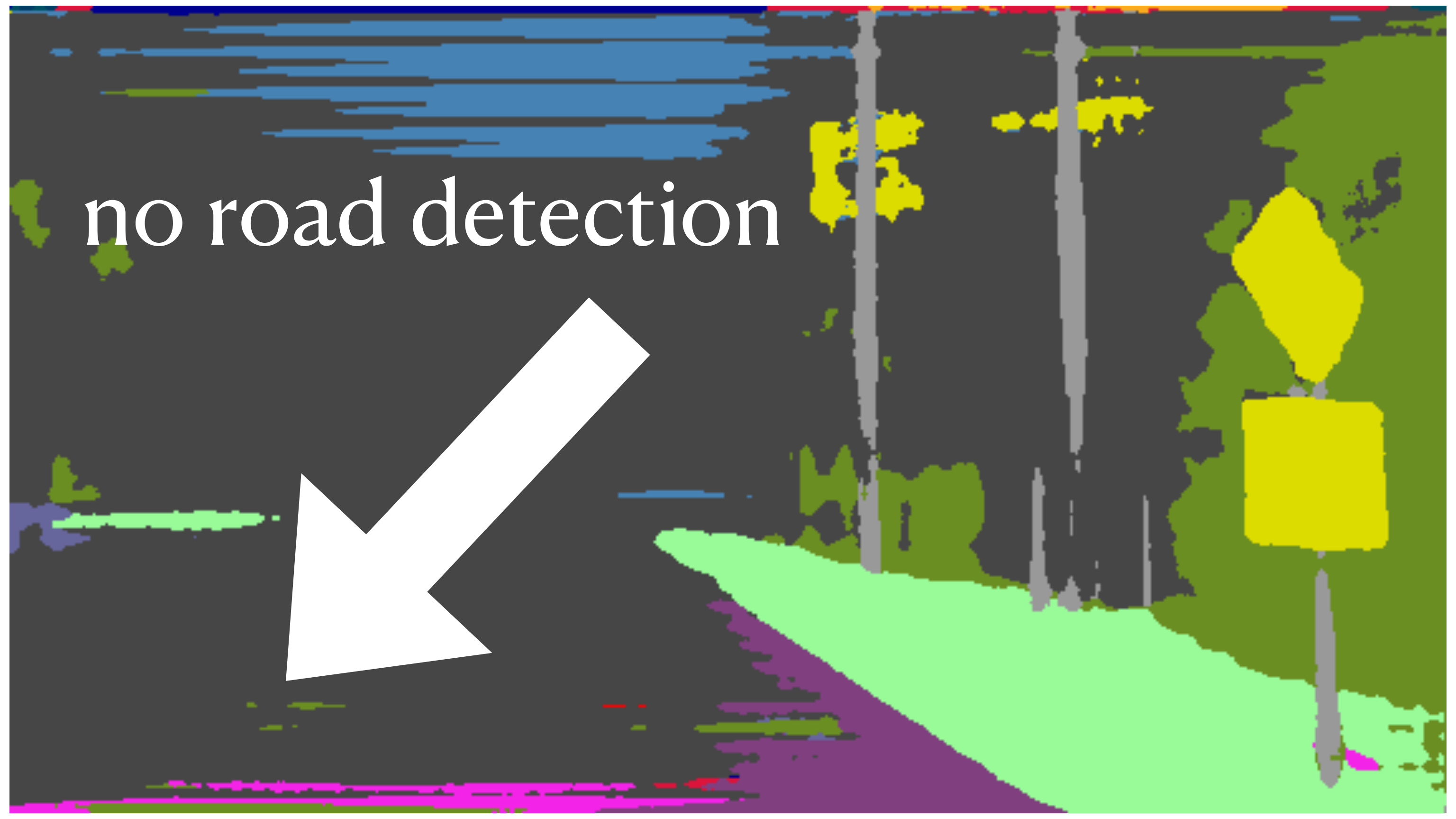


(c) Final Segmentation Result

Results

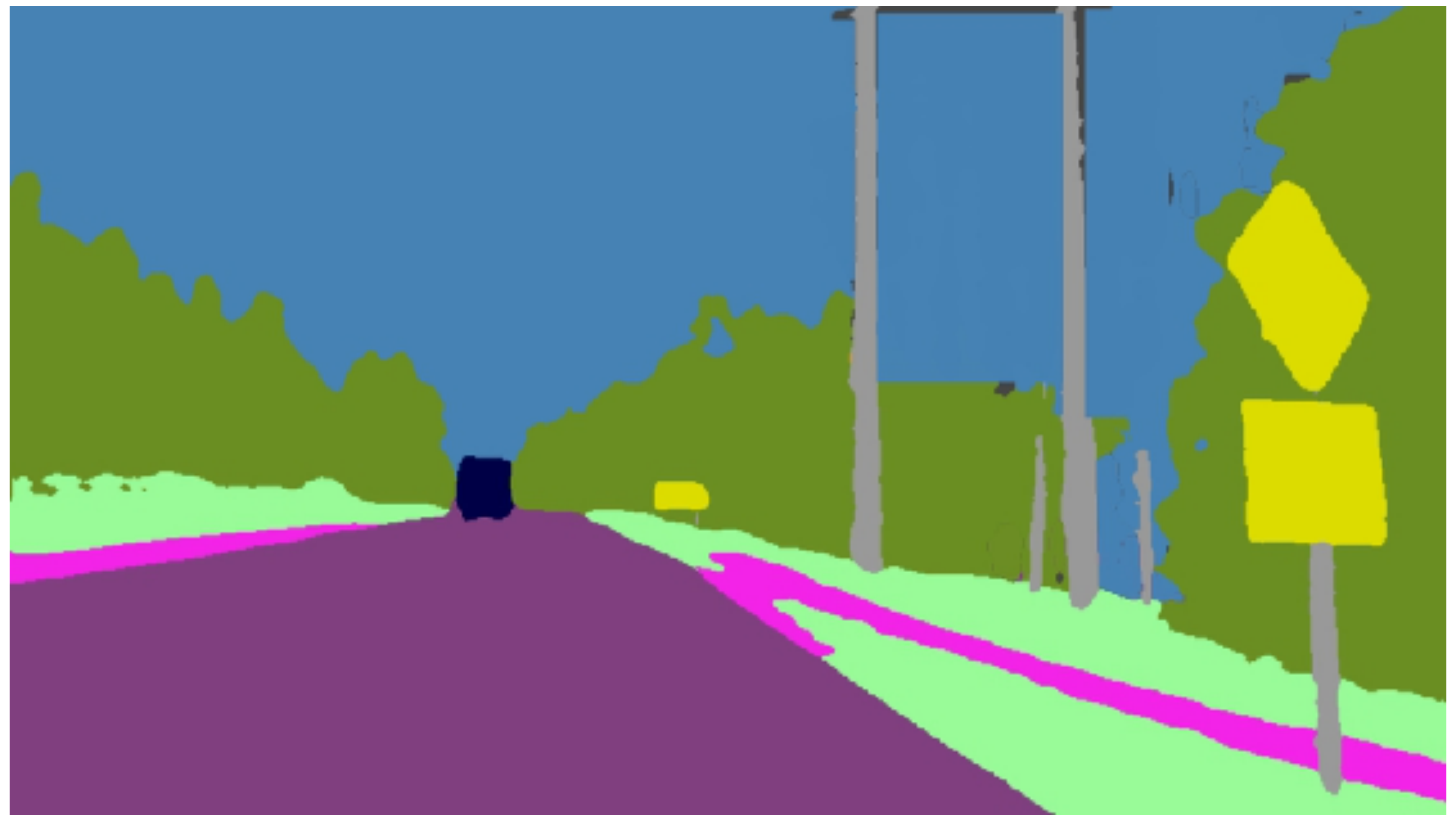
RaidaR (CVPR 2021) test image semantic segmentation results

(a) Input image



(b) Original + Deeplab-v3

(c) ToDayGAN + Deeplab-v3



(d) Ground truth

(e) ForkGAN + Deeplab-v3

(f) SinForkGAN + Deeplab-v3

Experiments

BDD100K (CVPR 2017) test image After ForkGAN++ Module

(a) Original Low-Light image

(b) After ForkGAN

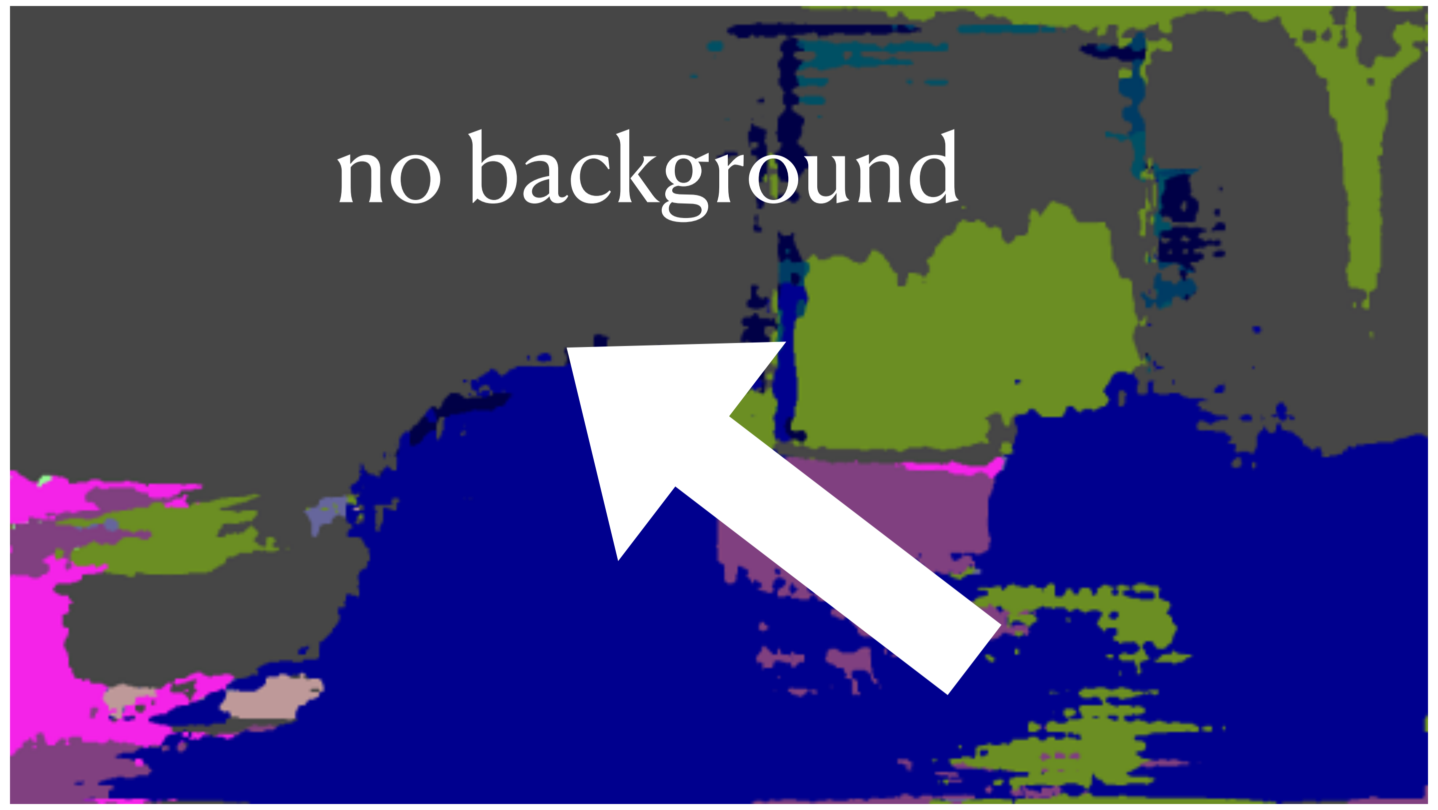
(c) After SinForkGAN

enhancing crucial detailed info

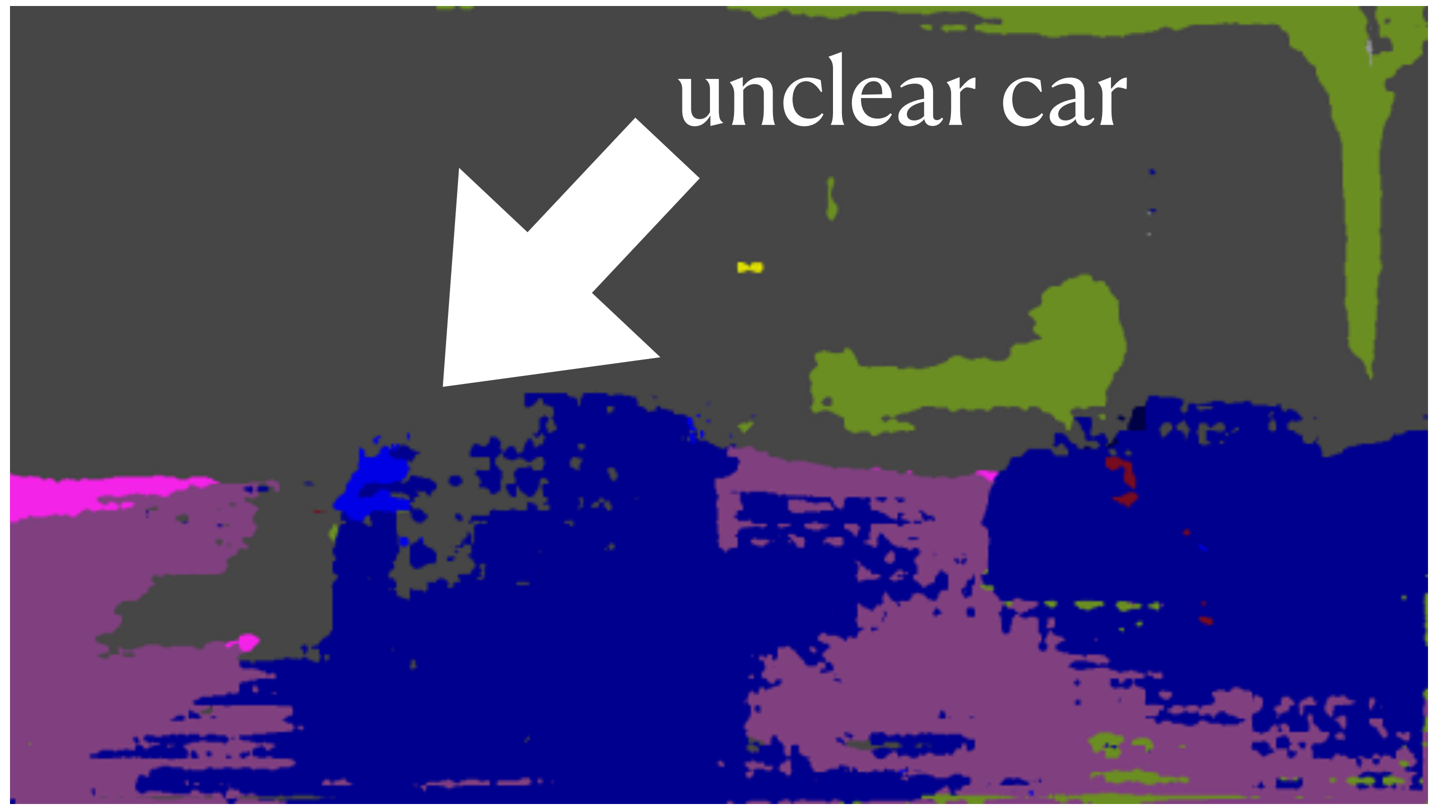
Results

BDD100K (CVPR 2017) test image semantic segmentation results


(a) Input image



(b) Original + Deeplab-v3



(c) ToDayGAN + Deeplab-v3


(d) Ground truth

(e) ForkGAN + Deeplab-v3

(f) SinForkGAN + Deeplab-v3

Experiments

Object Detection Task

Project Page

Github Code

Seri Lee @cma lab

Experiments

Object detection

- YOLOv3-tiny model pretrained on the PASCAL VOC 2007 + 2012 dataset (https://github.com/Lornatang/YOLOv3-PyTorch)
- pretrained SinForkGAN model tested on ExDark Dataset (7,363 nighttime images)

Method	Original	ToDayGAN
mAP@0.5	13.4	20.1

Method	ForkGAN	SinForkGAN
mAP@0.5	23.5	24.8

Experiments

ExDark (CVIU 2018) test image After ForkGAN++ Module

(a) Original Low-Light image

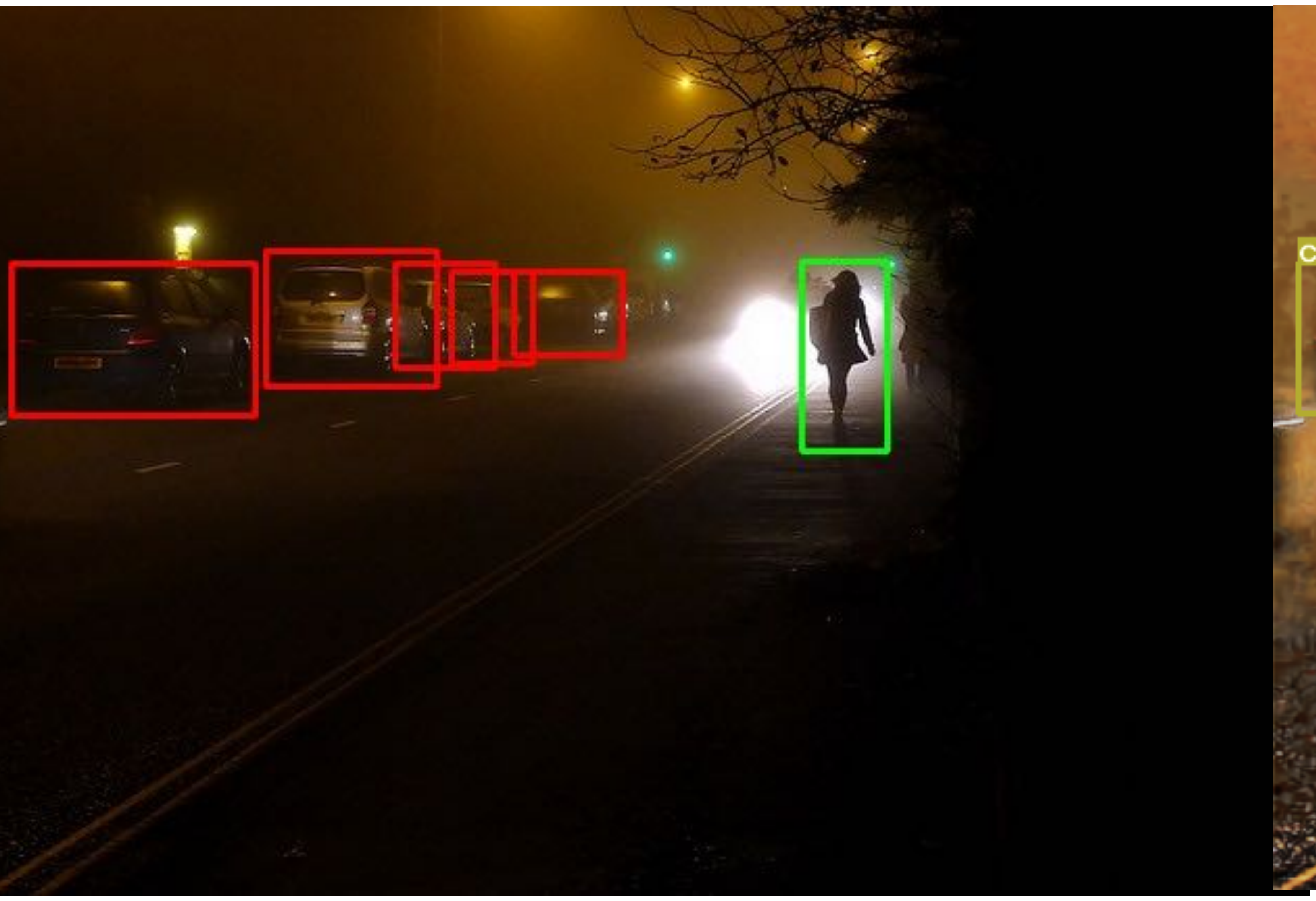
(b) After ForkGAN

(c) After SinForkGAN

improves detection of small objects

Results

Visualization of ExDark (CVIU 2018) object detection results



Ground truth

ForkGAN + YOLOv3-tiny

SinForkGAN + YOLOv3-tiny

Conclusion

and Possible Future Work

- achieve unbiased image translation without any supervision using real-world single night-time images
- works well on real-world night-time rainy images!
- can use existing daytime computer vision models
- possible future work: a multi-task learning network to share the backbone of different vision tasks
- possible future work: a one-shot end-to-end network