Domain Adaptive Object Detection

Yunlong MENG

2023-02-08

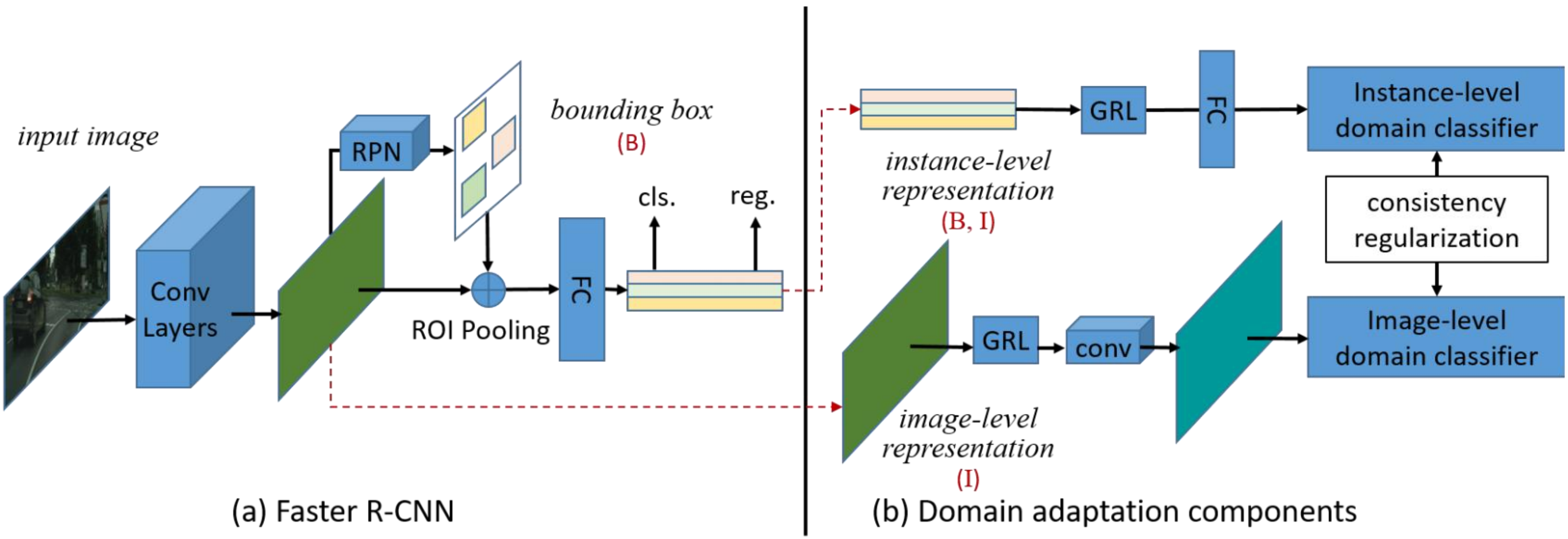
Outline

- **Discrepancy-based domain adaptive object detection**
  - Diminish the domain shift by finetuning the deep network based detection model with ***labeled*** or ***unlabeled*** target data
- **Adversarial-based domain adaptive object detection**
  - Utilize domain discriminators and conduct adversarial training to encourage domain confusion between ***source*** domain and ***target*** domain
- **Reconstruction-based domain adaptive object detection**
  - Improve the domain adaptive object detection by reconstructing the ***source*** samples or ***target*** samples
- **Hybrid domain adaptive object detection**
  - Use **two or more above domain adaptation mechanisms** simultaneously to achieve superior performance
- **Other domain adaptive object detection**
  - Graph-induced protory alignment
  - Categorical regularization
  - ......

Cross domain object detection problem formulation

- Labelled source set $\mathcal{S}$
- Unlabelled source set $\mathcal{T}$
- $\mathcal{S}$ share the same classes with $\mathcal{T}$
- Train a detector with promising accuracy in $\mathcal{T}$

Domain adaptive Faster RCNN



- **Joint adaptation**
  - Image-level adaptation
  - Instance-level adaptation
- **Consistency regularization**

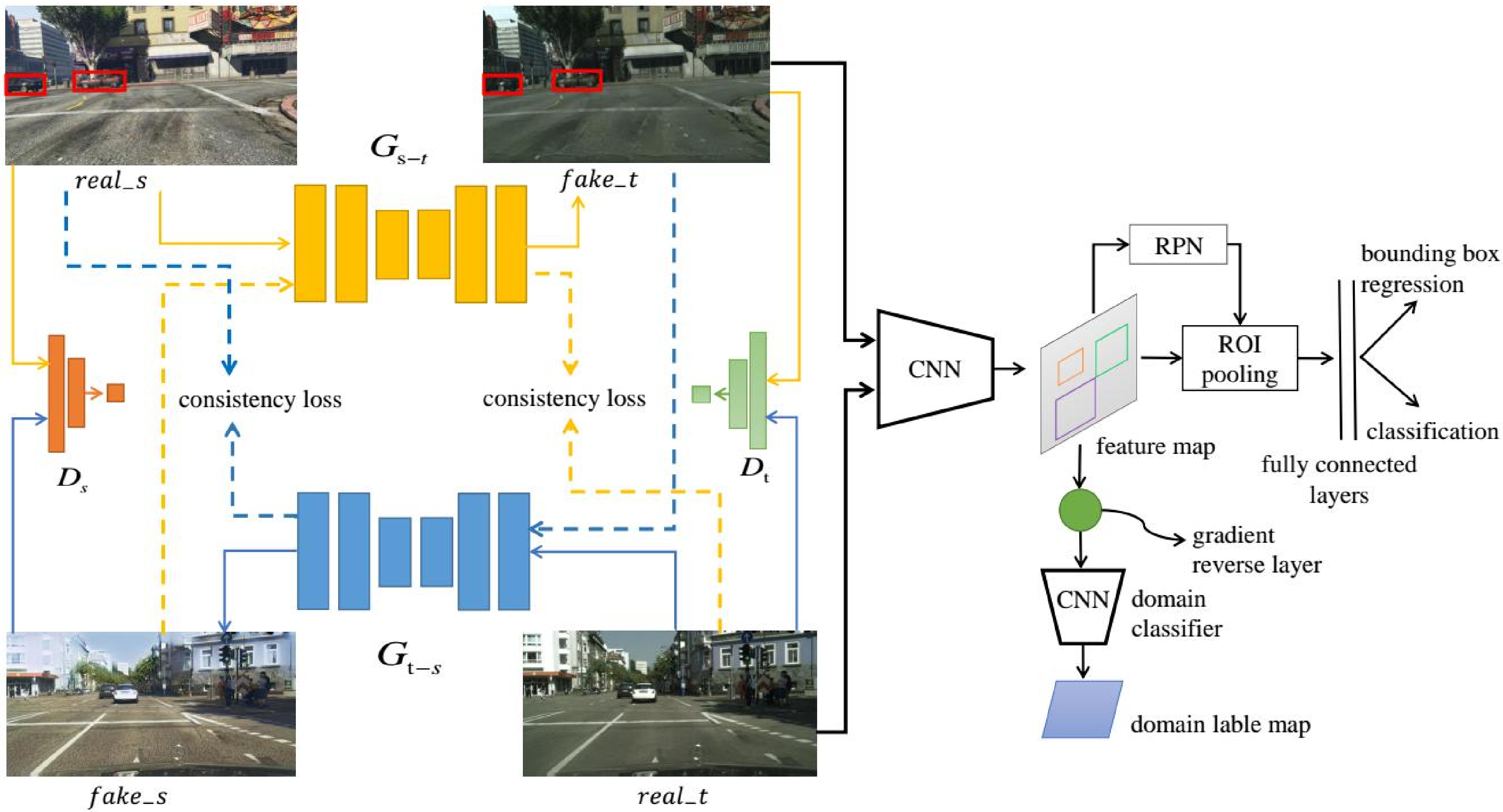
Quantative results of Domain adaptive Faster RCNN on *Foggy Cityscapes* task

	img	ins	cons	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Faster R-CNN				17.8	23.6	27.1	11.9	23.8	9.1	14.4	22.8	18.8
Ours	✓			22.9	30.7	39.0	20.1	27.5	17.7	21.4	25.9	25.7
		✓		23.6	30.6	38.6	20.8	**40.5**	12.8	17.1	26.1	26.3
	✓	✓		24.2	**31.2**	39.1	19.1	36.2	19.2	17.1	27.0	26.6
	✓	✓	✓	**25.0**	31.0	**40.5**	**22.1**	35.3	**20.2**	**20.0**	**27.1**	**27.6**

Chen *et al.* CVPR 2018

Hybrid domain adaptive object detection framework

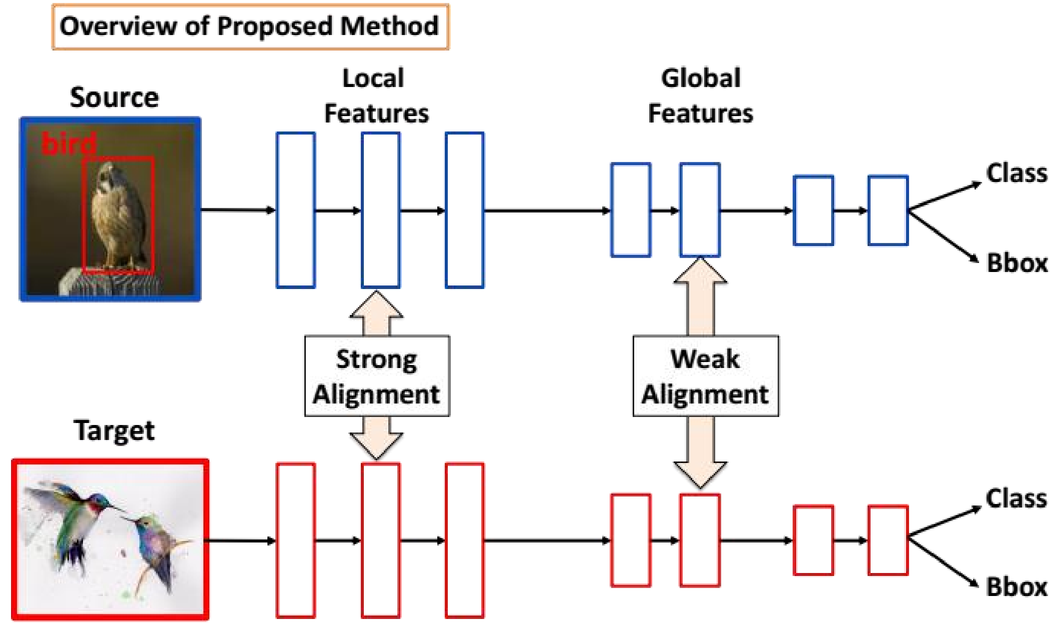
- **Pixel-level domain adaptation (PDA)**
- **Feature-level adaptation (FDA)**

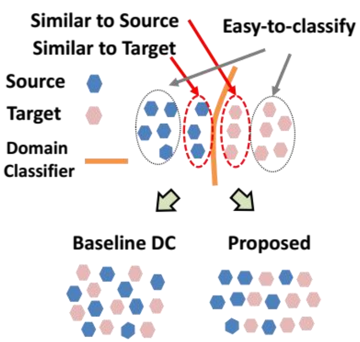


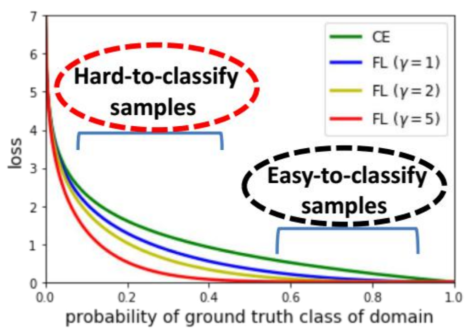
Shan *et al.* Neurocomputing 2018

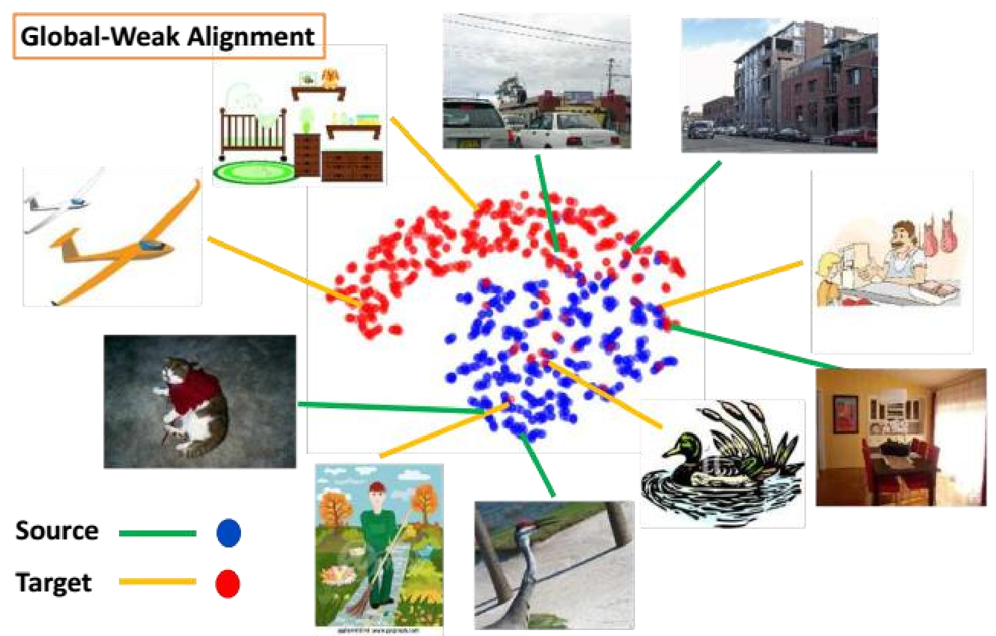
Strong-weak alignment domain adaptive Faster RCNN

- Strong-weak models learns domain-invariant features
    - Strongly aligned at the local patch level
    - Weakly (partially) aligned at the global scene level



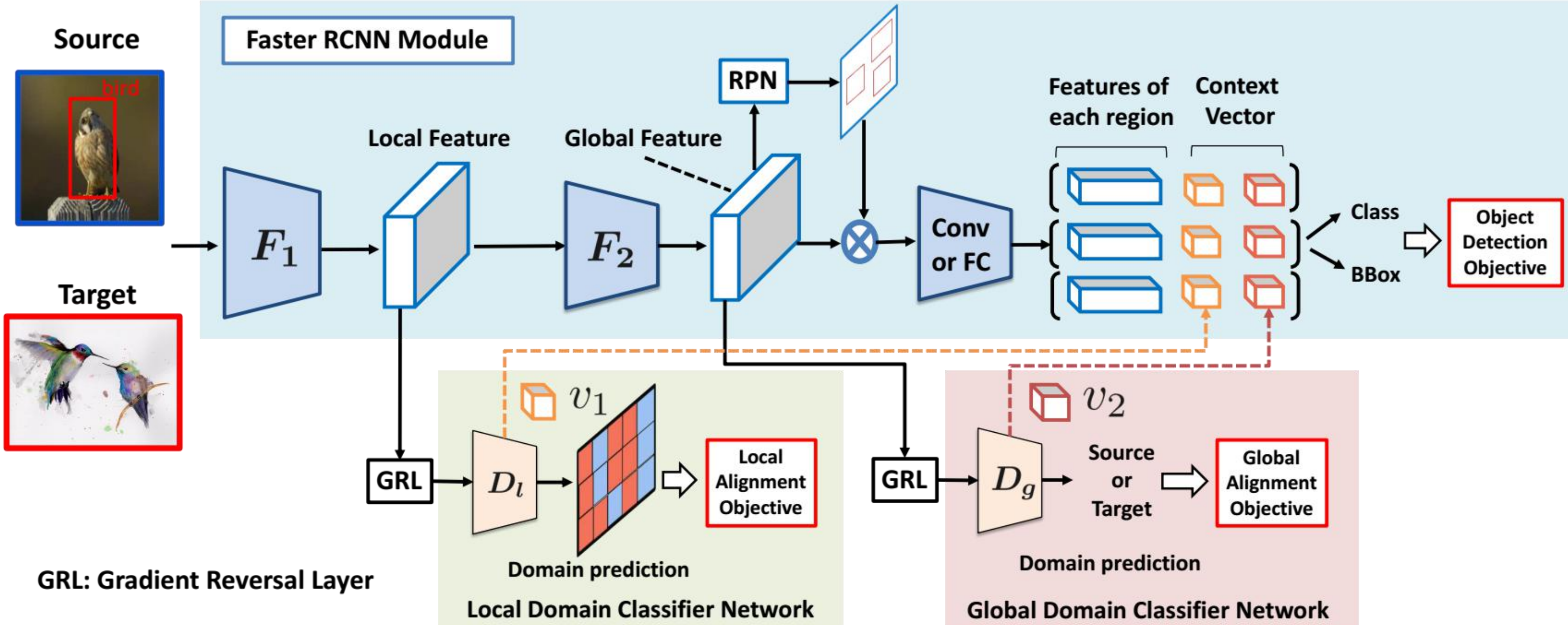






Saito *et al.* CVPR 2019

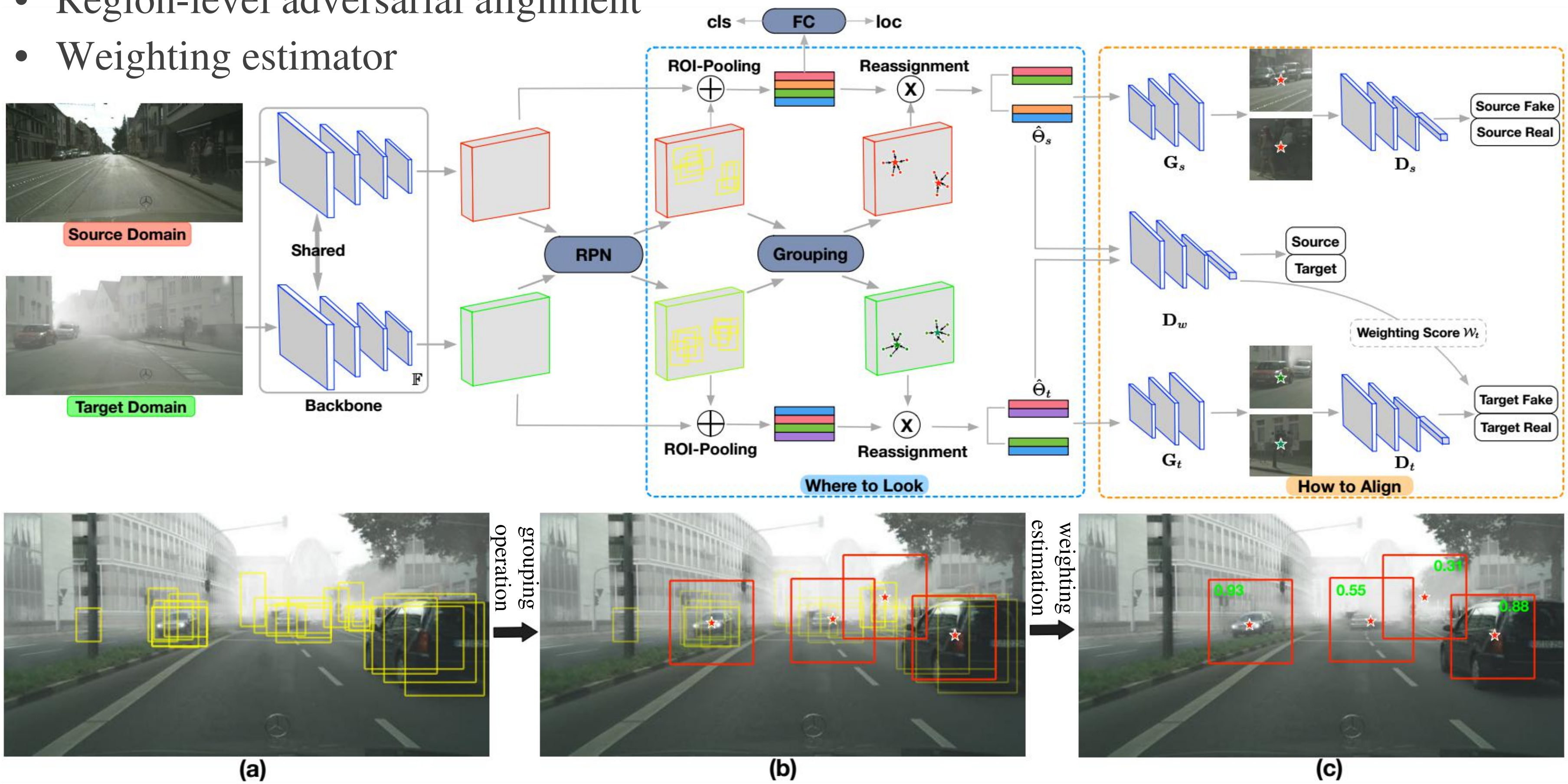
Strong-weak alignment domain adaptive Faster RCNN



Saito *et al.* CVPR 2019

Networks (SCDA)

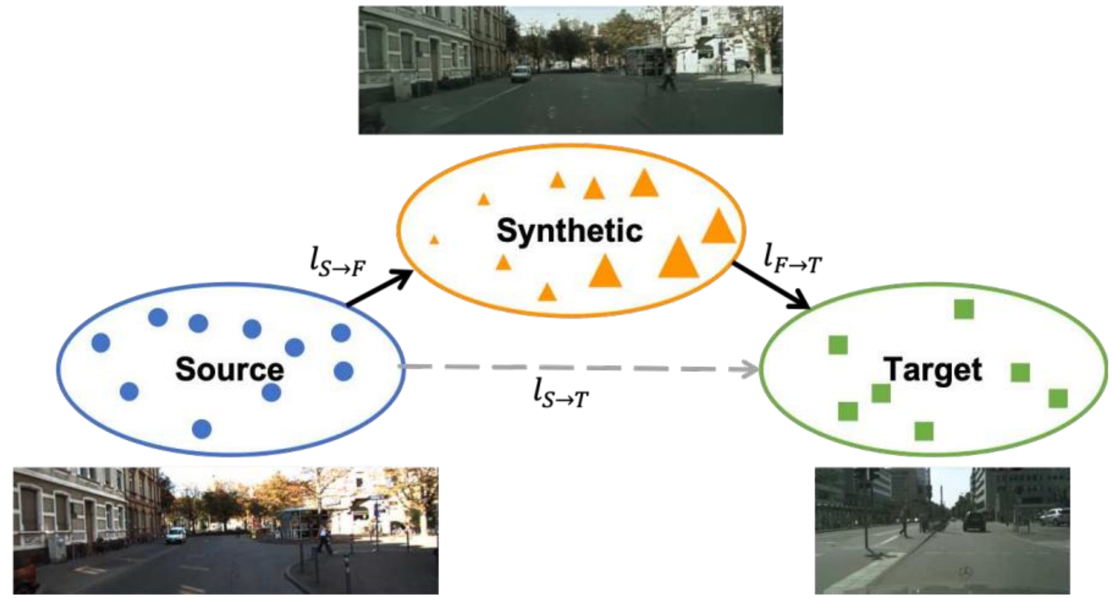
- **Region mining (Where to look)**
  - Grouping
  - Feature assignment
- **Adjusted region-level alignment (How to align)**
  - Region-level adversarial alignment
  - Weighting estimator

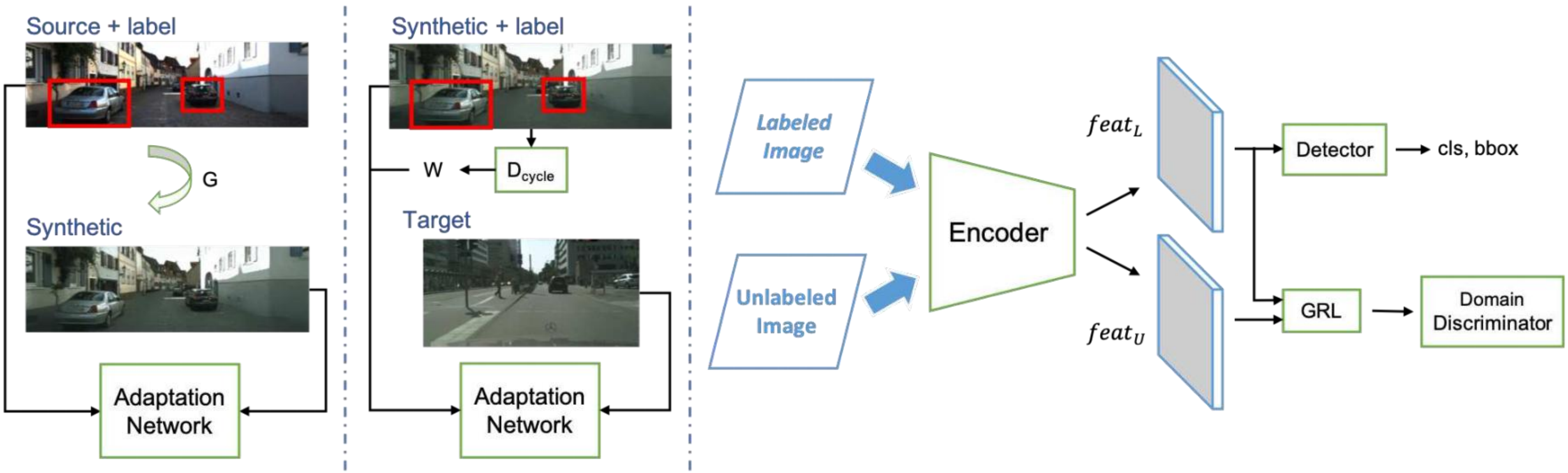


Zhu *et al.* CVPR 2019

Progressive domain adaption

- **Adaptation in feature space**
- **Progressive adaptation**
  - Intermediate domain
    - CycleGAN
  - Adaptation
  - Weighted supervision

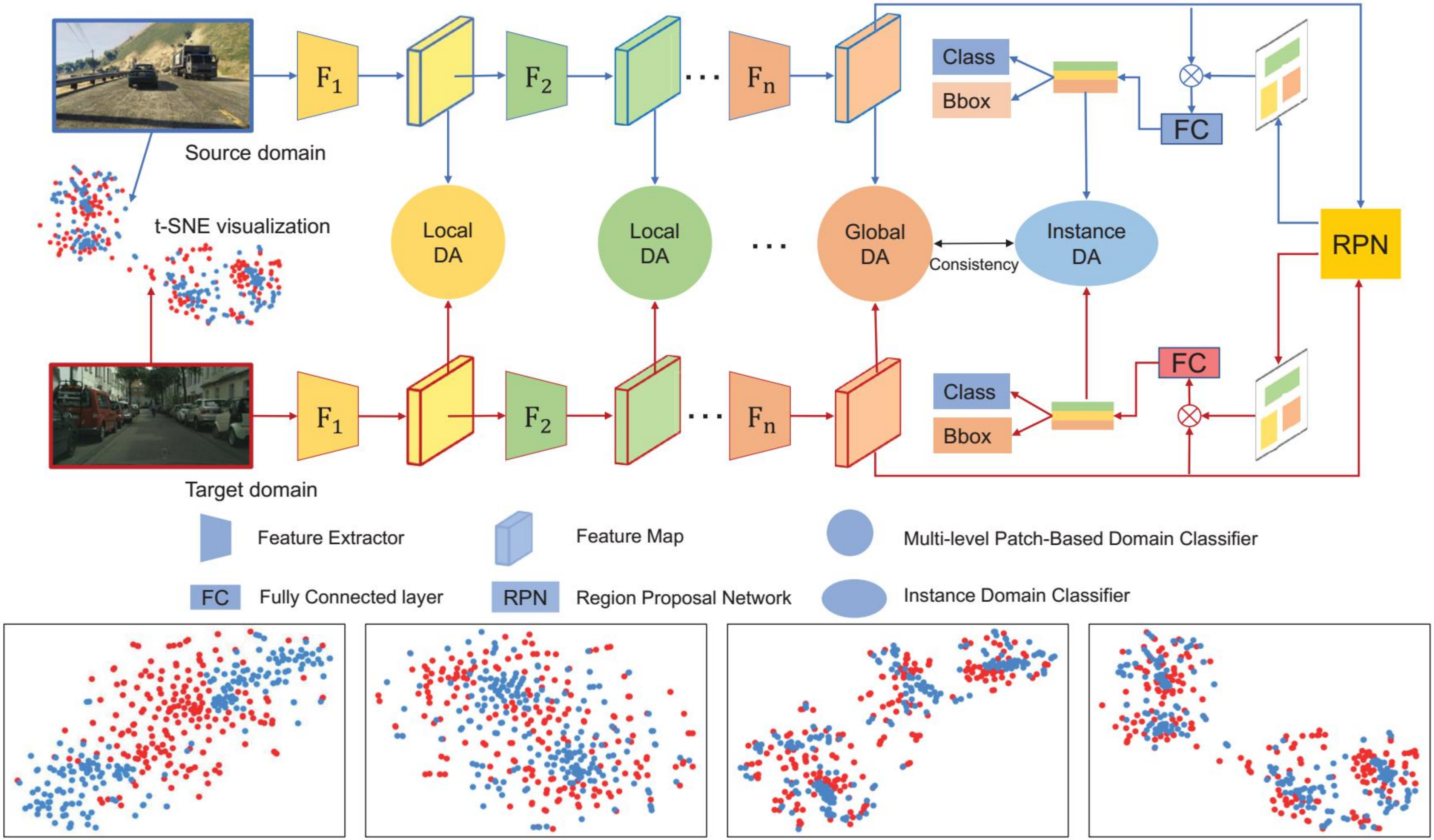




a) Stage one b) Stage two c) Adaptation network

Hsu *et al.* CVPR 2019 workshop

Multi-level domain adaptive learning for cross domain object detection

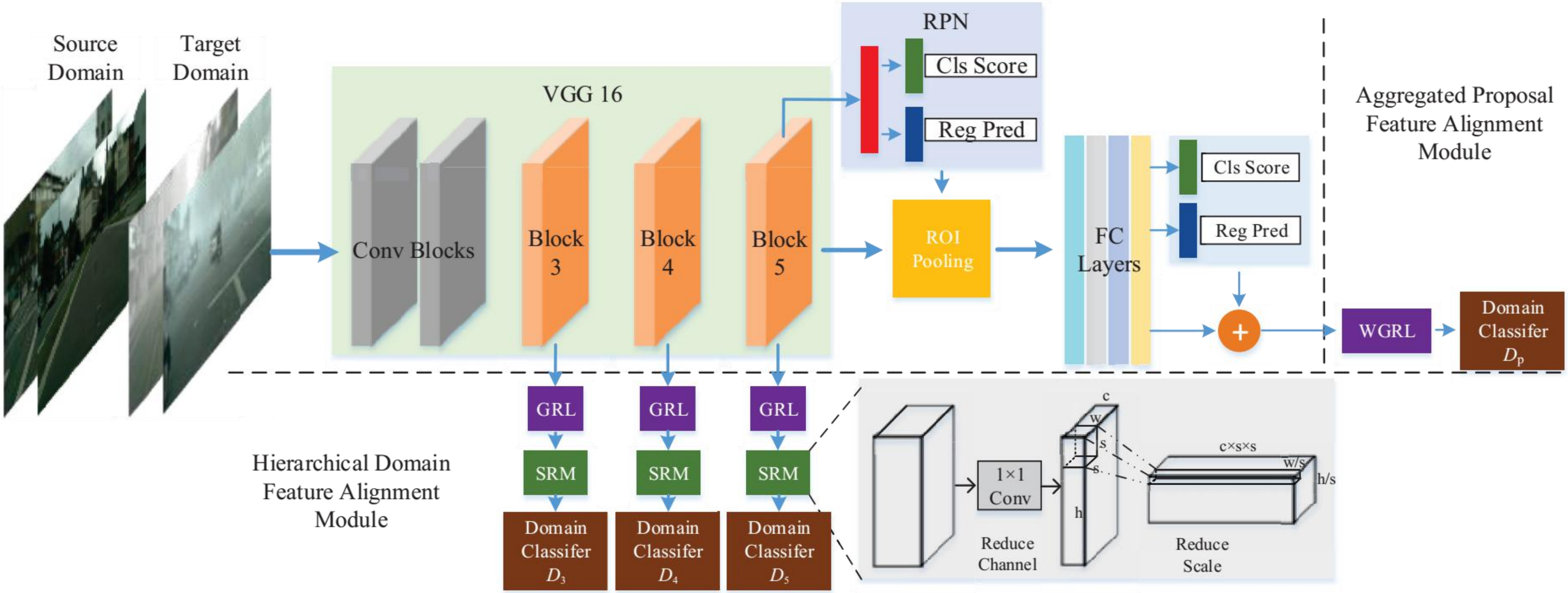


(a) Local Features (DA) (b) Local Features (Proposed) (c) Global Features (DA) (d) Global Features (Proposed)

Xie *et al.* ICCV 2019 Workshop

Multi-adversarial Faster RCNN for cross domain object detection

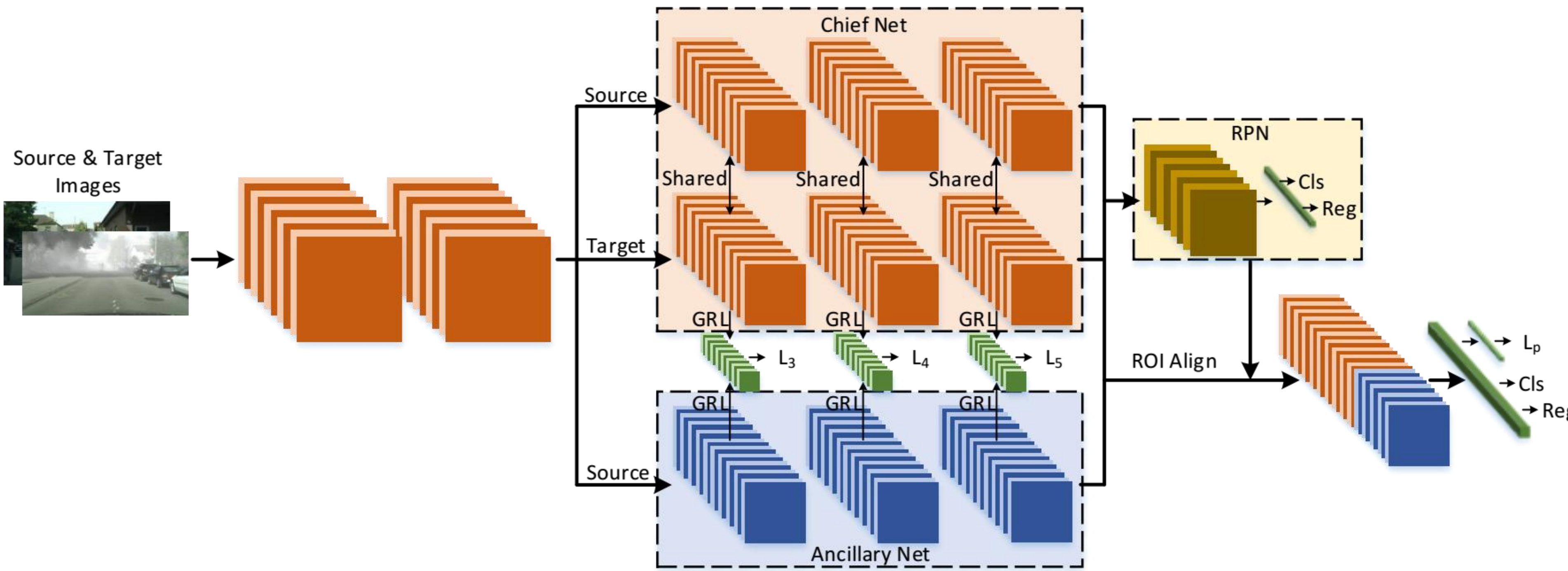
- **Hierarchical domain feature alignment**
- **Aggregated proposal feature alignment**



He *et al.* ICCV 2019

Asymmetric tri-way Faster RCNN for cross domain object detection

- **Chief net → bound the domain discrepancy d_A**
  - Global domain alignment with domain adversarial confusion
  - Local domain alignment with domain adversarial confusion
- **Ancillary net → bound the empirical risk ε_S**
  - Detection loss for the chief net is reused for the supervision of the ancilalry net



He *et al.* ECCV 2020

Asymmetric tri-way Faster RCNN for cross domain object detection

- quantitative and qualitative results

Quantative and qulitative results on *Cityscapes* → *FoggyCityscapes* cross domain object detection task

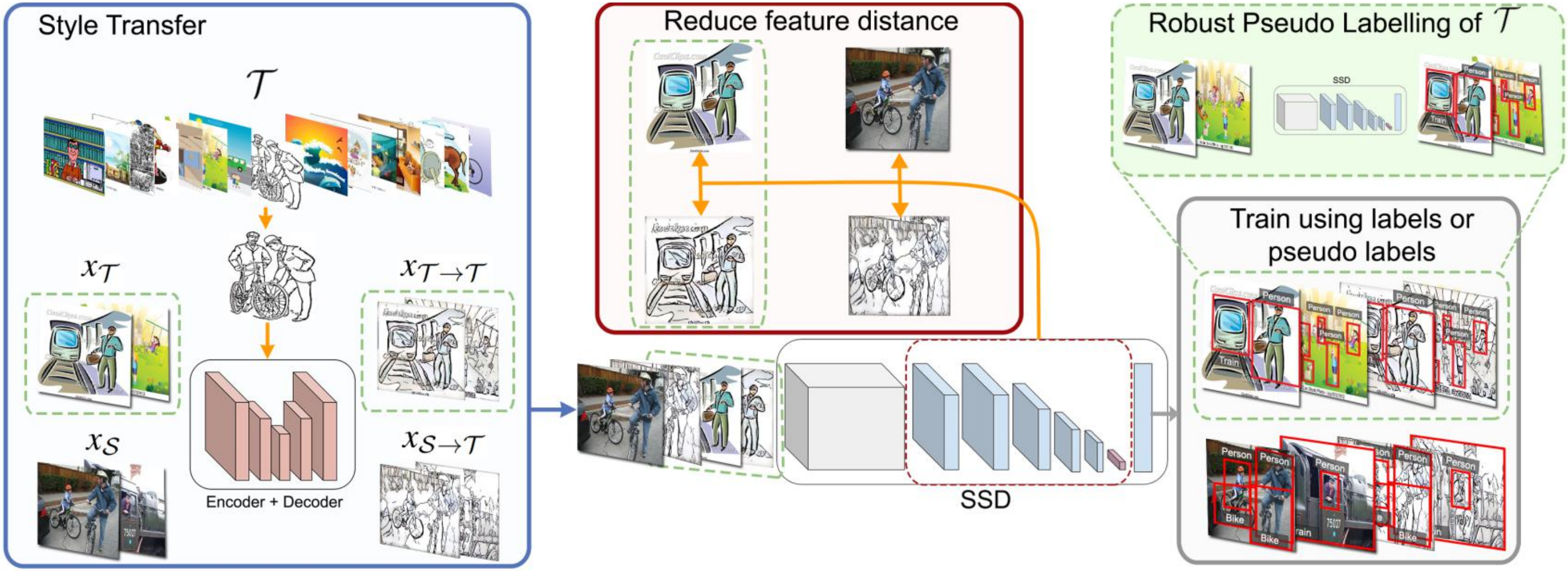
Methods	person	rider	car	truck	bus	train	mcycle	bcycle	mAP
Faster-RCNN	24.1	33.1	34.3	4.1	22.3	3.0	15.3	26.5	20.3
DAF(CVPR'18) [4]	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
MAF(ICCV'19) [17]	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
Strong-Weak [31]	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
D&match [22]	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
NL /w res101 [20]	**35.1**	42.2	49.2	30.1	45.3	27.0	26.9	36.0	36.5
SCL [35]	31.6	44.0	44.8	**30.4**	41.8	**40.7**	**33.6**	36.2	37.9
ATF(1-block)	33.3	43.6	44.6	24.3	39.6	10.5	27.2	35.6	32.3
ATF(2-blocks)	34.0	46.0	49.1	26.4	**46.5**	14.7	30.7	37.5	35.6
ATF(ours)	34.6	**47.0**	**50.0**	23.7	43.3	38.7	33.4	**38.8**	**38.7**

DAF(CVPR'2018) MAF(ICCV'2019) ATF(Ours) Ground Truth

He *et al.* ECCV 2020

Style consistency domain adaptive object detection

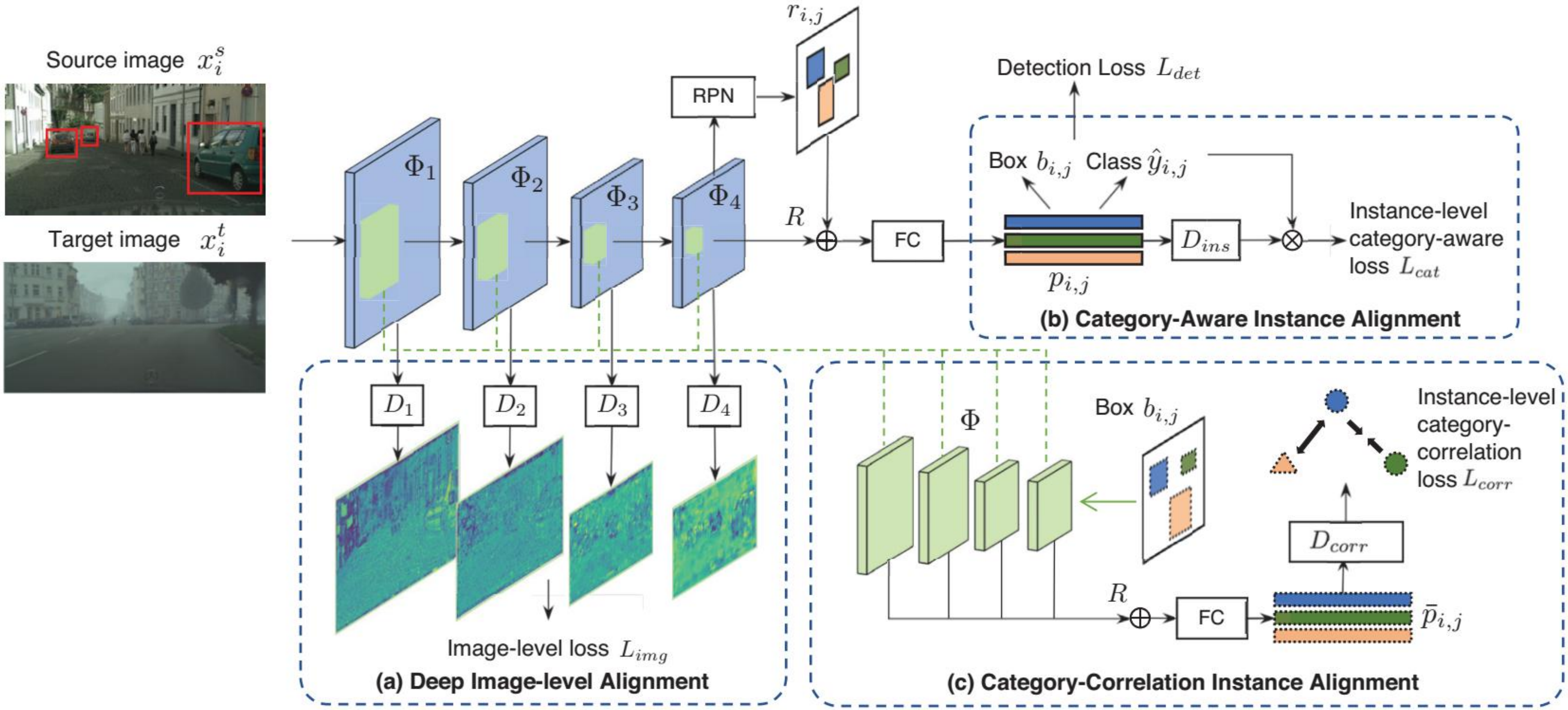
- **Style transfer with feature consistency**
  - Enforcing feature consistency
- **Robust pseudo labelling**
  - Positive examples sampling
  - Negative examples sampling
  - Combine pseudo labels and style transfer



Chen *et al.* CVPR 2020

Image-Instance Full Alignment Networks (iFAN)

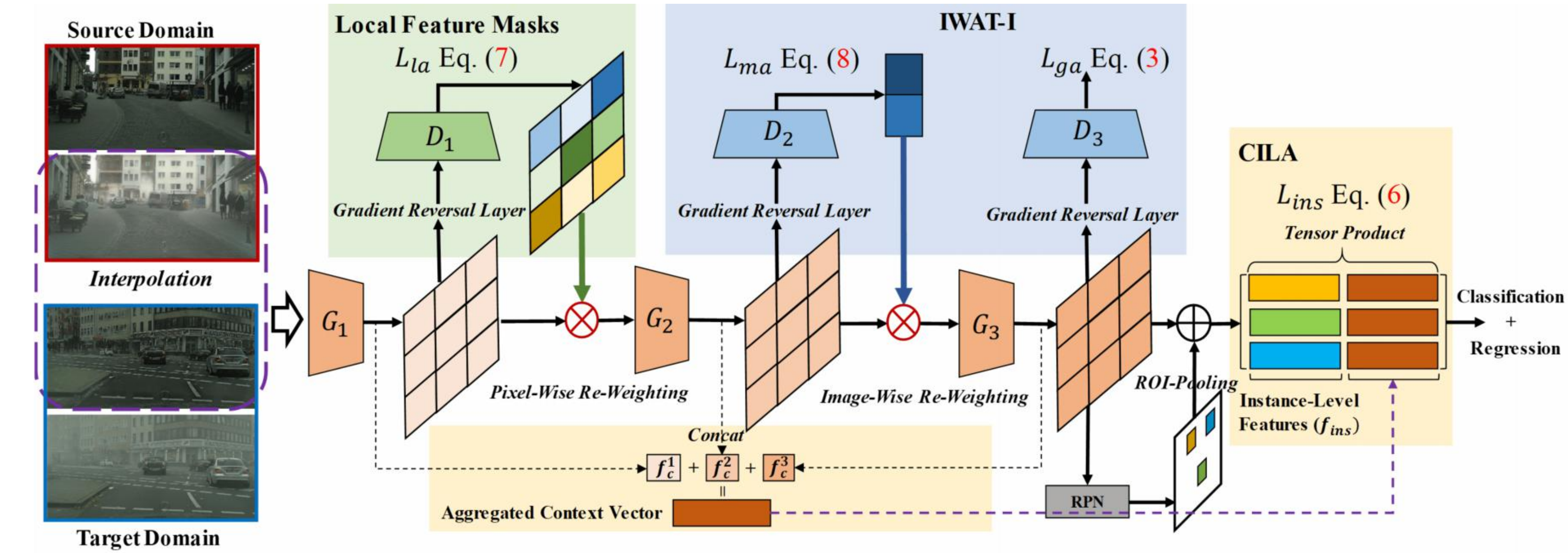
- **Deep image-level alignment**
- **Full instance-level alignment**
  - Category-agnostic instance alignment
  - Category-aware instance alignment
  - Category-corrleation instance alignment

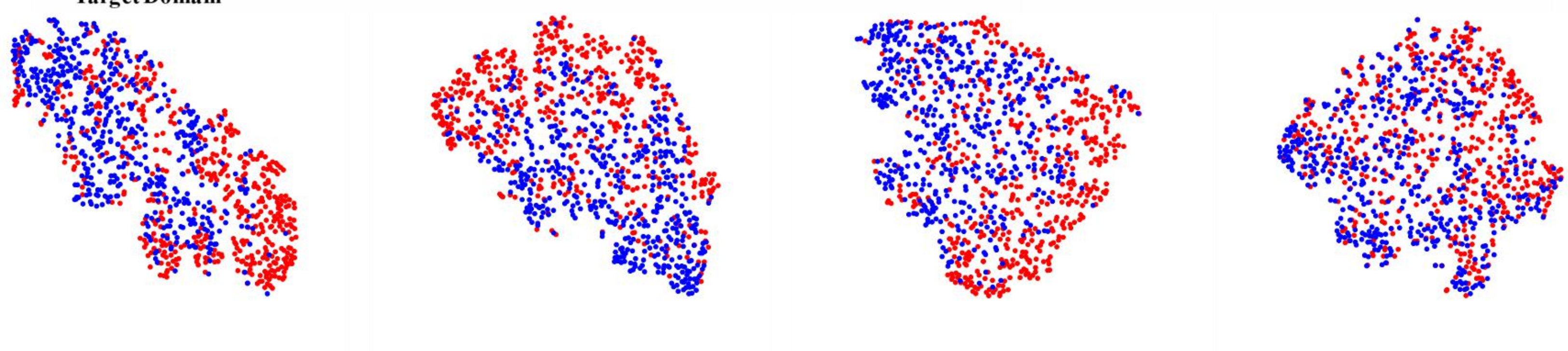


Zhuang *et al.* AAAI 2020

Hierarchical Transferability Calibration Network (HTCN)

- **Importance weighted adversarial training with input interpolation**
- **Context-aware instance-level alignment**
- **Local feature mask for semantic consistency**





(a) SWDA (b) HTCN-w/o CILA (c) HTCN-w/o Interpolation (d) HTCN

Chen *et al.* CVPR 2020

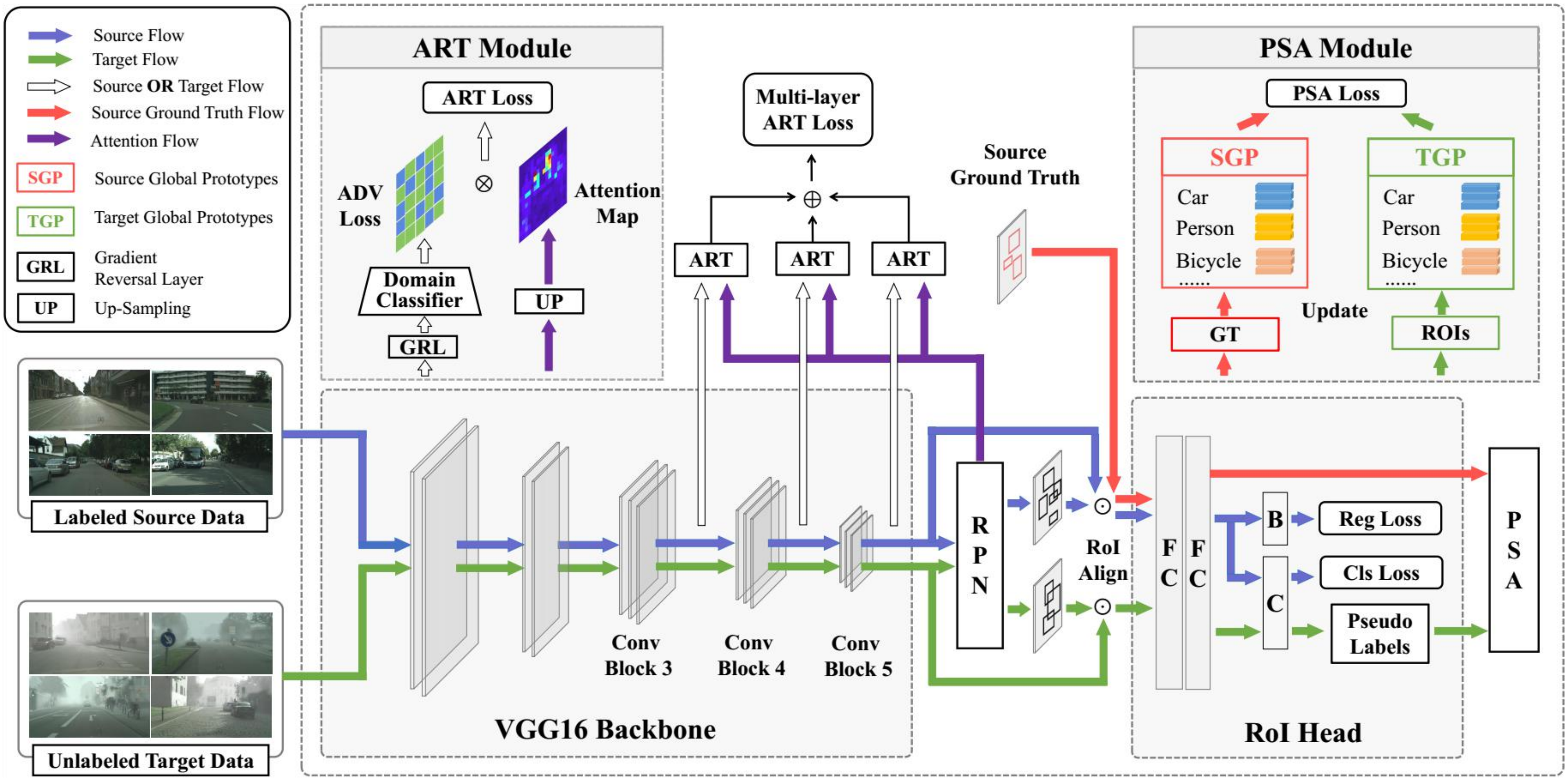
Hierarchical Transferability Calibration Network (HTCN) for cross domain object detection - qualitative results

(a) Source Only **(b) SWDA (CVPR 2019)** **(c) HTCN (Ours)**

Chen *et al.* CVPR 2020

Coarse-to-fine feature adaptation network

- **Attention-based region transfer (ART)**
- **Prototype-based semantic alignment**



Zheng *et al.* CVPR 2020

Coarse-to-fine feature adaptation network for cross domain object detection - quantitative results

Quantative results on *Cityscapes* → *FoggyCityscapes* cross domain object detection task

Cityscapes → FoggyCityscapes												
Method	Arch.	Bus	Bicycle	Car	Motor	Person	Rider	Train	Truck	**mAP**	**mAP***	**Gain**
MTOR [4]	R	38.6	35.6	44.0	28.3	30.6	41.4	40.6	21.9	35.1	26.9	8.2
RLDA [29]	I	45.3	36.0	49.2	26.9	35.1	42.2	27.0	30.0	36.5	31.9	4.6
DAF [8]	V	35.3	27.1	40.5	20.0	25.0	31.0	20.2	22.1	27.6	18.8	8.8
SCDA [74]	V	39.0	33.6	48.5	28.0	33.5	38.0	23.3	26.5	33.8	26.2	7.6
MAF [22]	V	39.9	33.9	43.9	29.2	28.2	39.5	<u>33.3</u>	23.8	34.0	18.8	15.2
SWDA [52]	V	36.2	35.3	43.5	30.0	29.9	42.3	32.6	24.5	34.3	20.3	14.0
DD-MRL [31]	V	38.4	32.2	44.3	28.4	30.8	40.5	**34.5**	27.2	34.6	17.9	16.7
MDA [68]	V	41.8	36.5	44.8	30.5	33.2	44.2	28.7	28.2	36.0	22.8	13.2
PDA [25]	V	<u>44.1</u>	35.9	**54.4**	29.1	**36.0**	45.5	25.8	24.3	36.9	19.6	<u>17.3</u>
Source Only	V	25.0	26.8	30.6	15.5	24.1	29.4	4.6	10.6	20.8	-	-
3DC (Baseline)	V	37.9	37.1	51.6	33.1	32.9	45.6	27.9	28.6	36.8	20.8	16.0
Ours w/o ART	V	41.6	35.4	51.5	**36.9**	33.5	45.2	26.6	28.2	37.4	20.8	16.6
Ours w/o PSA	V	**45.2**	<u>37.3</u>	51.8	33.3	33.9	<u>46.7</u>	25.5	<u>29.6</u>	<u>37.9</u>	20.8	17.1
Ours	V	43.2	**37.4**	<u>52.1</u>	<u>34.7</u>	<u>34.0</u>	**46.9**	29.9	**30.8**	**38.6**	20.8	**17.8**
Oracle	V	49.5	37.0	52.7	36.0	36.1	47.1	56.0	32.1	43.3	-	-

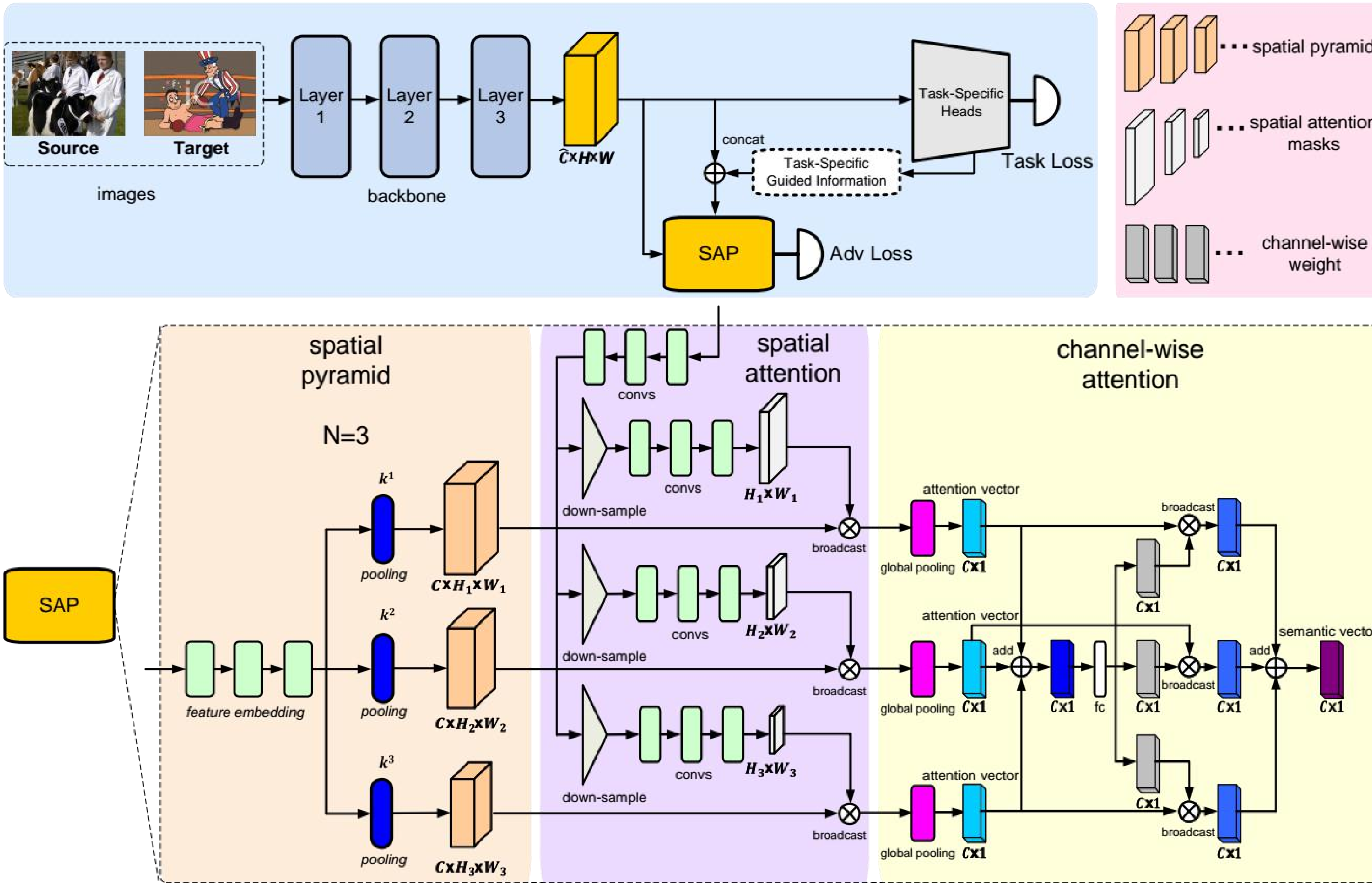
Zheng *et al.* CVPR 2020

Coarse-to-fine feature adaptation network for cross domain object detection - qualitative results

(a) Source Only (b) SWDA (c) Ours (d) Attention Map

Zheng *et al.* CVPR 2020

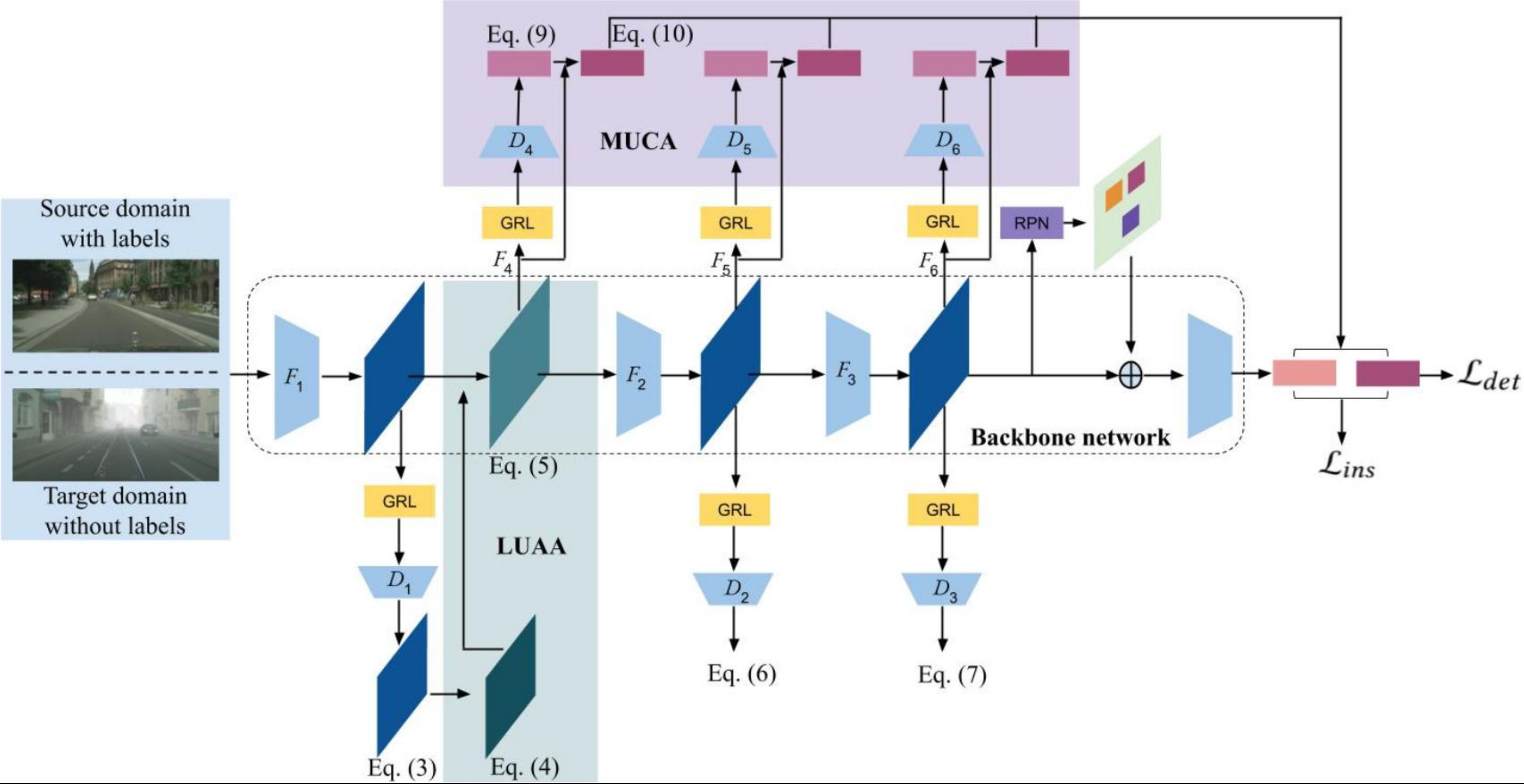
Spatial attention pyramid network



Li *et al.* ECCV 2020

Uncertainty-aware distribution alignment for cross-domain adaptive object detection

- **Uncertainty-aware attentional alignment module**
- **Multi-level uncertainty-aware context alignment moduel**



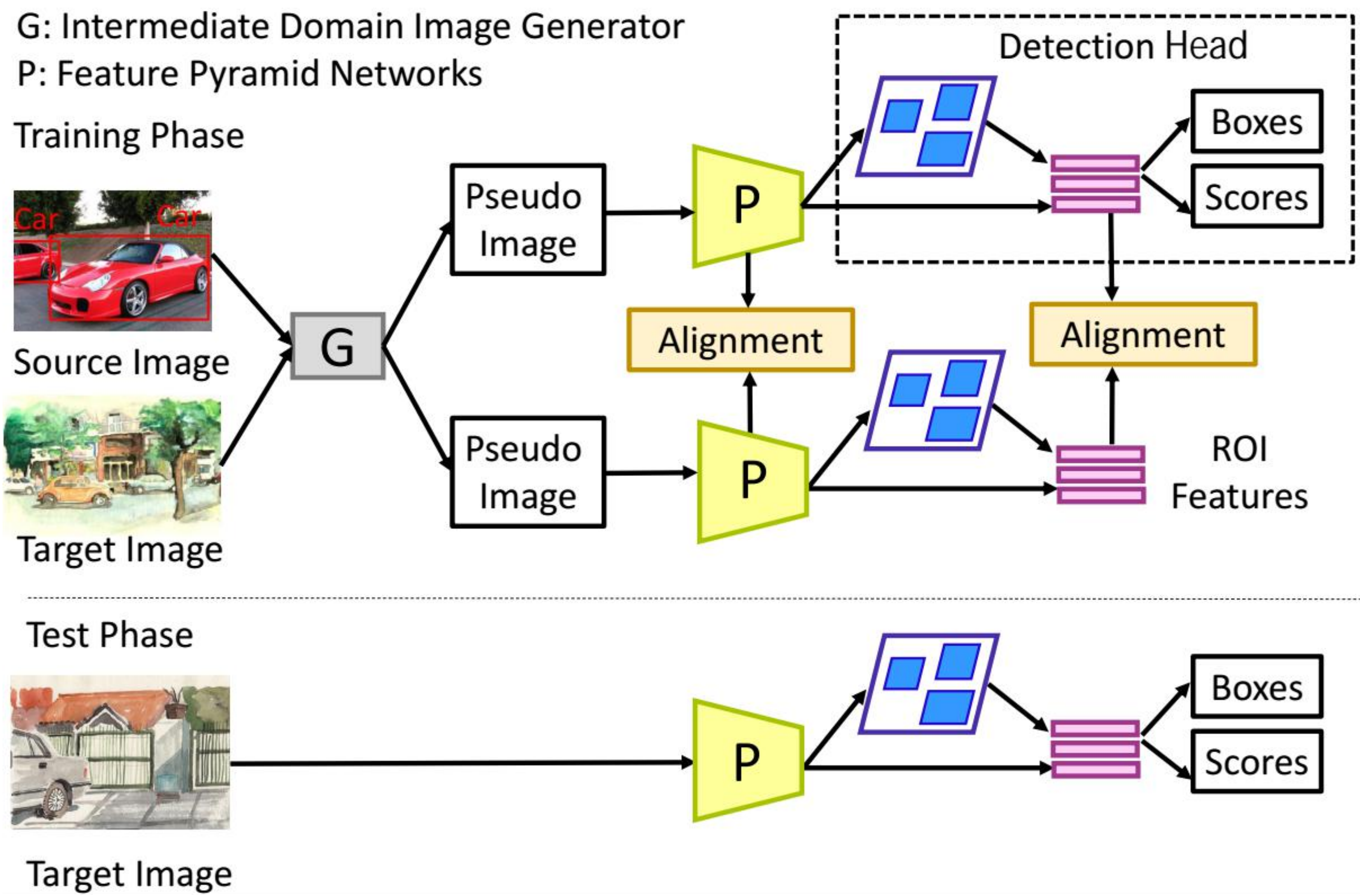
Nguyen *et al.* ACM MM 2020

Spatial attention pyramid network

Methods	Person	Rider	Car	Truck	Bus	Train	Motorbike	Bicycle	mAP
Source Only [21]	24.1	33.1	34.3	4.1	22.3	3.0	15.3	26.5	20.3
DA-Faster (CVPR' 18) [3]	25.5	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
SCDA (CVPR' 19) [33]	33.5	38.0	48.5	26.5	39.0	23.3	28.0	33.6	33.8
MAF (ICCV' 19) [12]	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34.0
SWDA (CVPR' 19) [22]	29.9	42.3	43.5	24.5	36.2	32.6	30.0	35.3	34.3
DD-MRL (CVPR' 19) [17]	30.8	40.5	44.3	27.2	38.4	34.5	28.4	32.2	34.6
ICR-CCR (CVPR' 20) [30]	32.9	43.8	49.2	27.2	45.1	36.4	30.3	34.6	37.4
HTCN (CVPR' 20) [2]	33.2	47.5	47.9	**31.6**	**47.4**	40.9	32.3	**37.1**	39.8
MEAA (Ours)	**34.2**	**48.9**	**52.4**	30.3	42.7	**46.0**	**33.2**	36.2	**40.5**
Oracle [22]	33.2	45.9	49.7	35.6	50.0	37.4	34.7	36.2	40.3

Nguyen *et al.* ACM MM 2020

Augmented Feature Alignment Networks (AFAN)



Wang *et al.* Arxiv 2021

Augmented Feature Alignment Networks (AFAN)

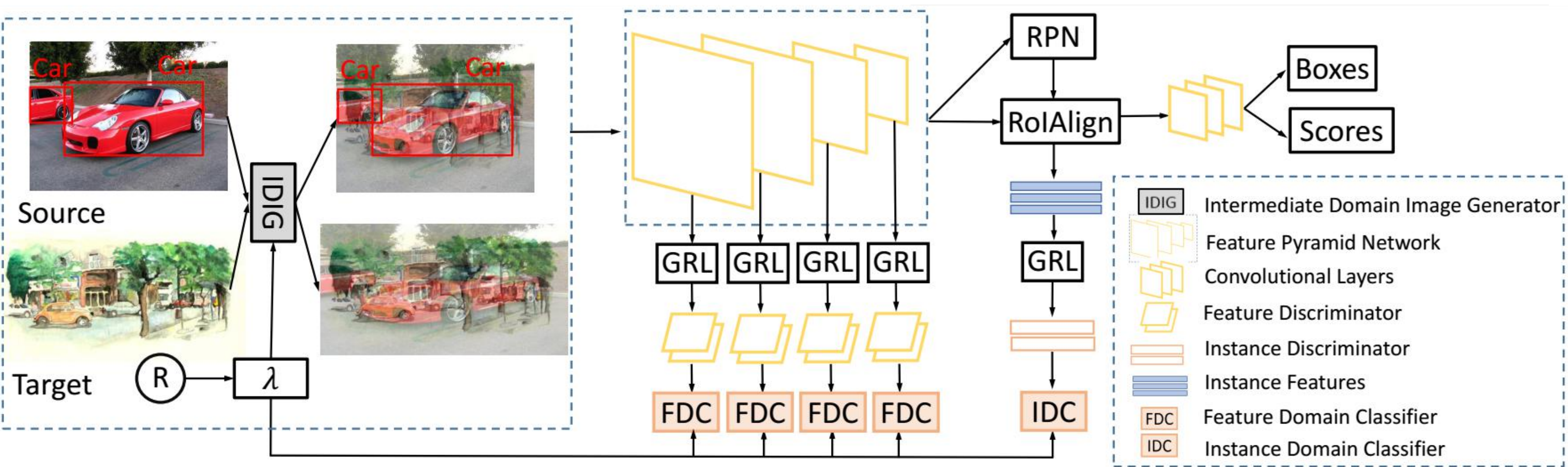
- Pseudo image generation
- Feature pyramid alignment
- Region feature alignment
- Object detection head

Algorithm 1 Training process of the AFAN.

Input: A batch of labeled source images $\{(x^s, y^s), \cdots\}$ from the source domain, a batch of unlabeled target images $\{x^t, \cdots\}$.

Draw λ and γ from $\mathbb{U}(0, \lambda_m)$ and $\mathbb{U}(0, 1)$, respectively.
if $\gamma > 0.5$ **then**
  $\lambda \leftarrow 0$.
end if
Generate intermediate source domain images $\{(\tilde{x}^s, y^s), \cdots\}$ and intermediate target domain images $\{\tilde{x}^t, \cdots\}$ using Equation (1).
Perform the forward pass of deep networks by feeding $\{(\tilde{x}^s, y^s), \cdots\} \cup \{\tilde{x}^t, \cdots\}$.
Calculate the feature level and proposal level discriminator losses with regard to λ.
Perform the backward pass by minimizing the loss defined in Equation (9).

Output: The updated network.

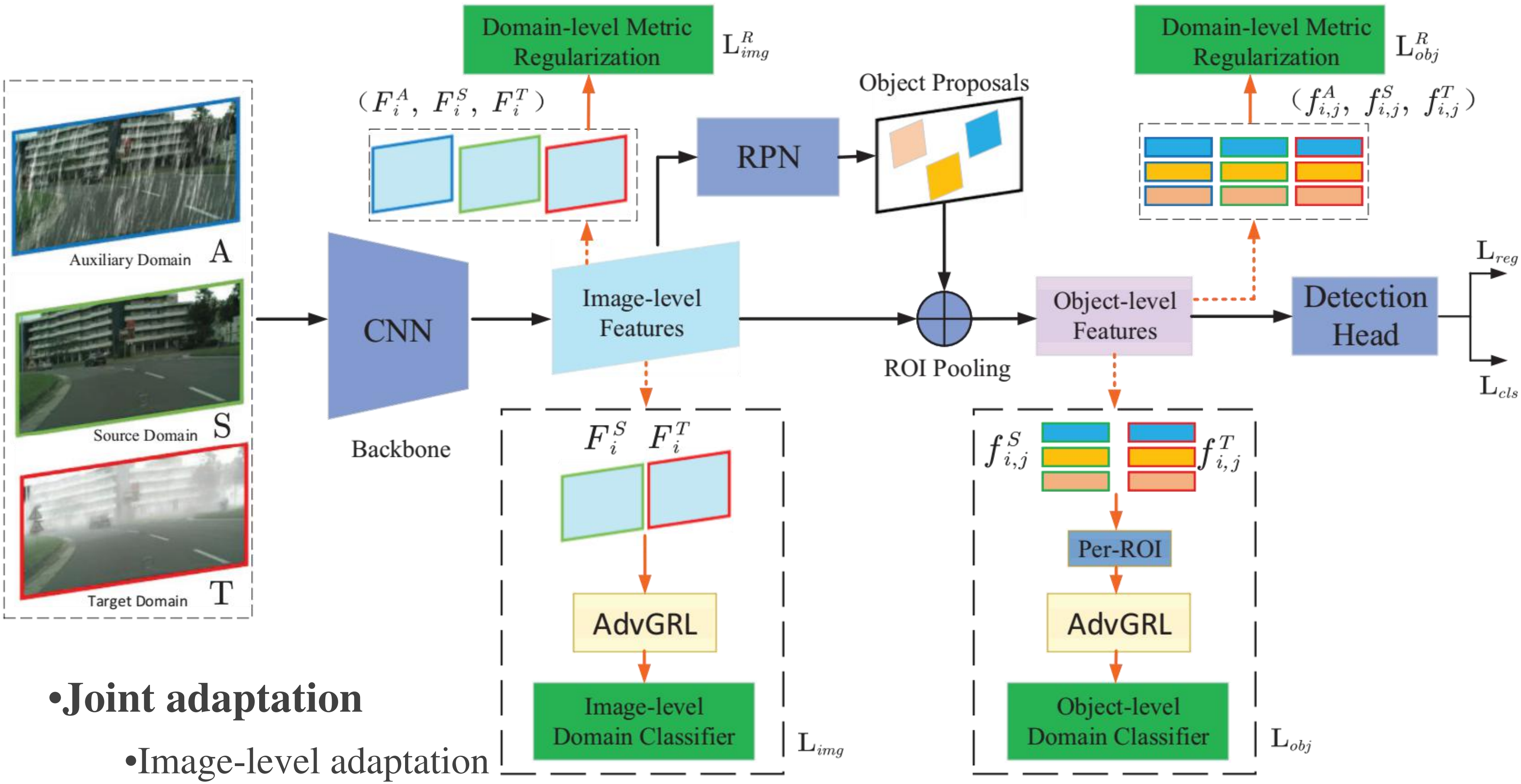


Wang *et al.* Arxiv 2021

Augmented Feature Alignment Networks (AFAN) for cross domain object detection - qualitative results

Wang *et al.* Arxiv 2021

Improved domain adaptive Faster RCNN



- **Joint adaptation**
  - Image-level adaptation
  - Object-level adaptation
- **Adversarial gradient reversal layer**
- **Domain-level metric regularization**

Li *et al.* WACV 2023

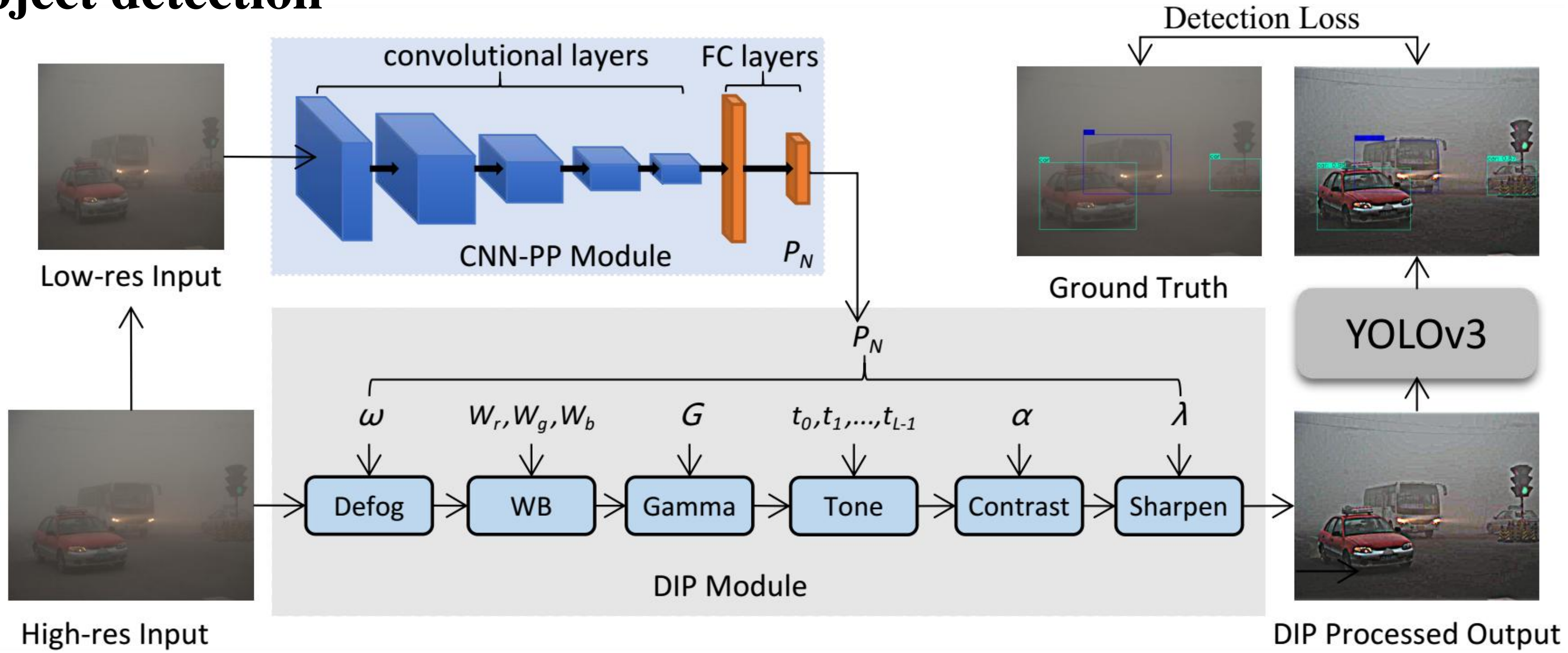
Improved domain adaptive Faster RCNN for cross domain object detection - quantitative results

Quantative results on *Cityscapes* → *FoggyCityscapes* cross domain object detection task

Methods	bus	bicycle	car	mcycle	person	rider	train	truck	mAP
SCDA-CVPR'19 [61]	39.0	33.6	48.5	28.0	33.5	38.0	23.3	26.5	33.8
DM-CVPR'19 [24]	38.4	32.2	44.3	28.4	30.8	40.5	34.5	27.2	34.6
MAF-ICCV'19 [18]	39.9	33.9	43.9	29.2	28.2	39.5	33.3	23.8	34.0
MCAR-ECCV'20 [57]	44.1	36.6	43.9	**37.4**	32.0	42.1	43.4	**31.3**	38.8
SWDA-CVPR'19 [39]	36.2	35.3	43.5	30.0	29.9	42.3	32.6	24.5	34.3
PDA-WACV'20 [21]	44.1	35.9	**54.4**	29.1	36.0	45.5	25.8	24.3	36.9
MTOR-CVPR-19 [3]	38.6	35.6	44.0	28.3	30.6	41.4	40.6	21.9	35.1
DA-Faster-CVPR'18 [5]	<u>49.8</u>	<u>39.0</u>	53.0	28.9	35.7	45.2	<u>45.4</u>	<u>30.9</u>	41.0
GPA-CVPR'20 [47]	45.7	38.7	54.1	<u>32.4</u>	32.9	**46.7**	41.1	24.7	39.5
RPN-PR-CVPR'21 [55]	43.6	36.8	50.5	29.7	33.3	45.6	42.0	30.4	39.0
UaDAN-TMM'21 [15]	49.4	38.9	53.6	32.3	**36.5**	<u>46.1</u>	42.7	28.9	<u>41.1</u>
Ours w/o Auxiliary Domain	48.4	36.7	53.5	26.1	<u>36.1</u>	45.9	39.1	29.3	40.2
Ours	**51.2**	**39.1**	<u>54.3</u>	31.6	**36.5**	**46.7**	**48.7**	30.3	**42.3**
Oracle	49.9	45.8	65.2	39.6	46.5	51.3	34.2	32.6	45.6

Li *et al.* WACV 2023

End-to-end training of IA-YOLO framework for cross domain object detection



Liu *et al.* AAAI 2022

End-to-end training of IA-YOLO framework for cross domain object detection - training procedure

---

Algorithm 1: Image-Adaptive YOLO training procedure

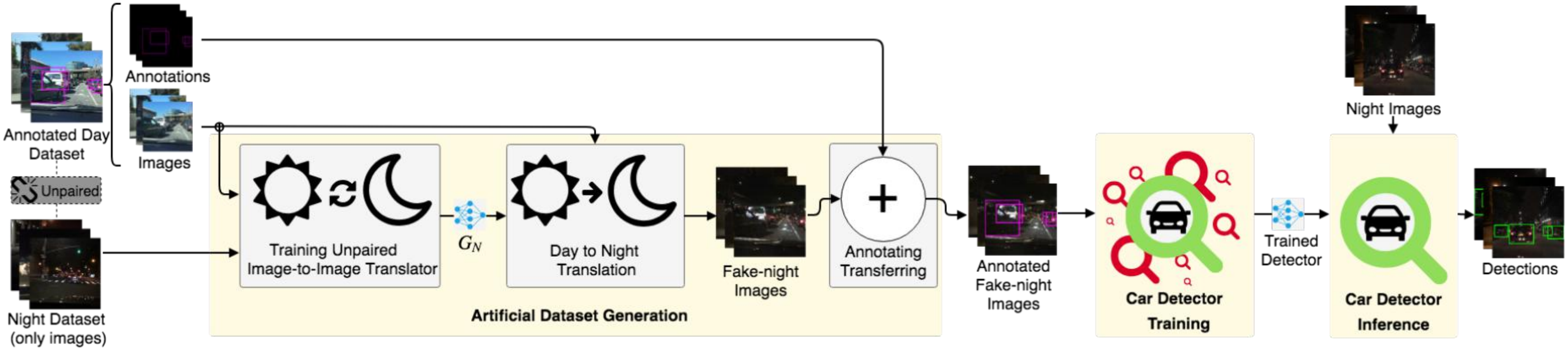
Initialize the CNN-PP network P^θ and the YOLOv3 network D^β with random weights θ and β.
Set the training stage: $num\_epochs = 80$, $batch\_size = 6$.
Prepare the normal dataset VOC_norm_trainval.
for i in num_epochs **do**
  repeat
    Take a batch images M from VOC_norm_trainval;
    for j in $batch\_size$ **do**
      if $random.randint(0, 2) > 0$ **then**
        Generate the foggy image $M(j)$ (Eq. (10, 11)), where $A = 0.5$, $k = random.randint(0, 9)$, $\beta = 0.01 * k + 0.05$ //for foggy conditions
        Generate the low-light image $M(j)$ by $f(x) = x^\gamma$, where $\gamma = random.uniform(1.5, 5)$ //for low-light conditions
      end if
    end for
    Compute DIP params by $P_N = P^\theta(image\_batch)$;
    Perform DIP filter processing by $image\_batch = DIP(image\_batch, P_N)$;
    Send $image\_batch$ to YOLOv3 network D^β;
    Update the CNN-PP network P^θ and the YOLOv3 network D^β according to the YOLOv3 detection loss.
  until all images have been fed into training models
end for

Real-world foggy condition

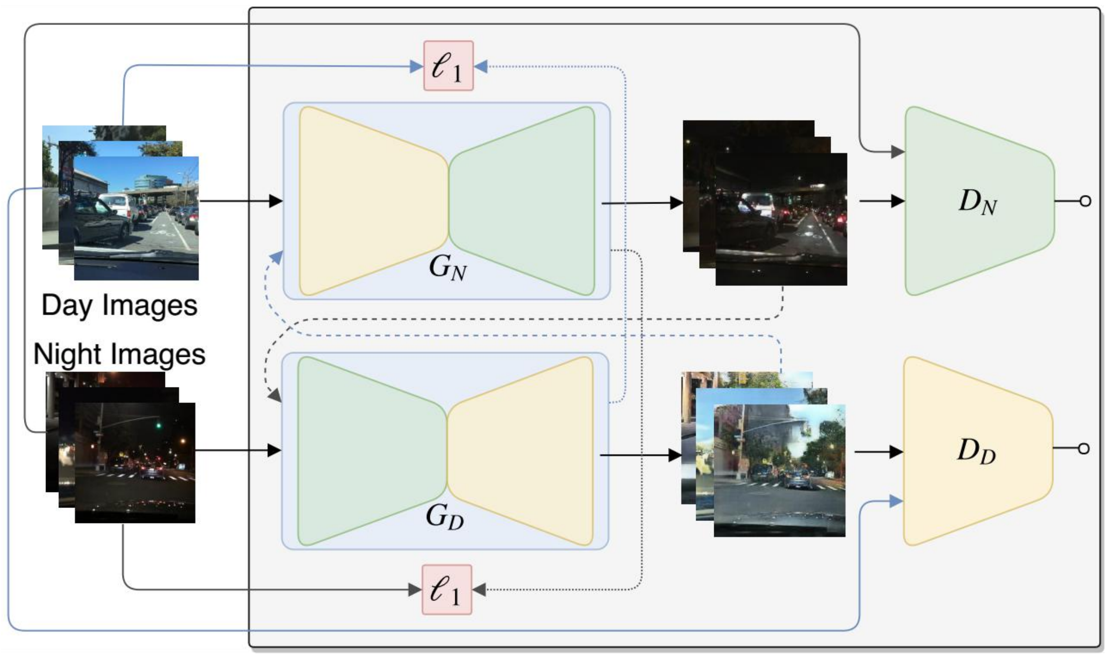
(a) YOLO II (Baseline) (b) IA-YOLO (Ours)

Liu *et al.* AAAI 2022

Unsupervised image-to-image (UI2I) translation for cross domain object detection - Entire Framework



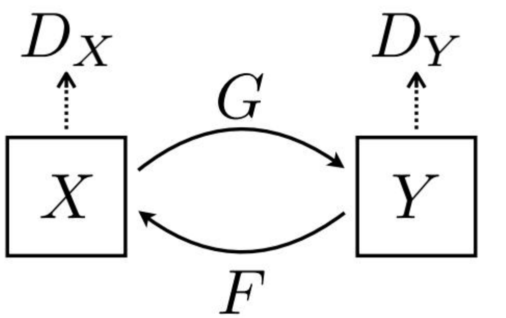
- Augmented dataset generation in ***target*** domain via UI2I translation
    - CycleGAN
- *De facto* detector training in the ***target*** domain
    - Faster RCNN
    - FCOS
    - YOLO series
    - ......
- Inference in the ***target*** domain



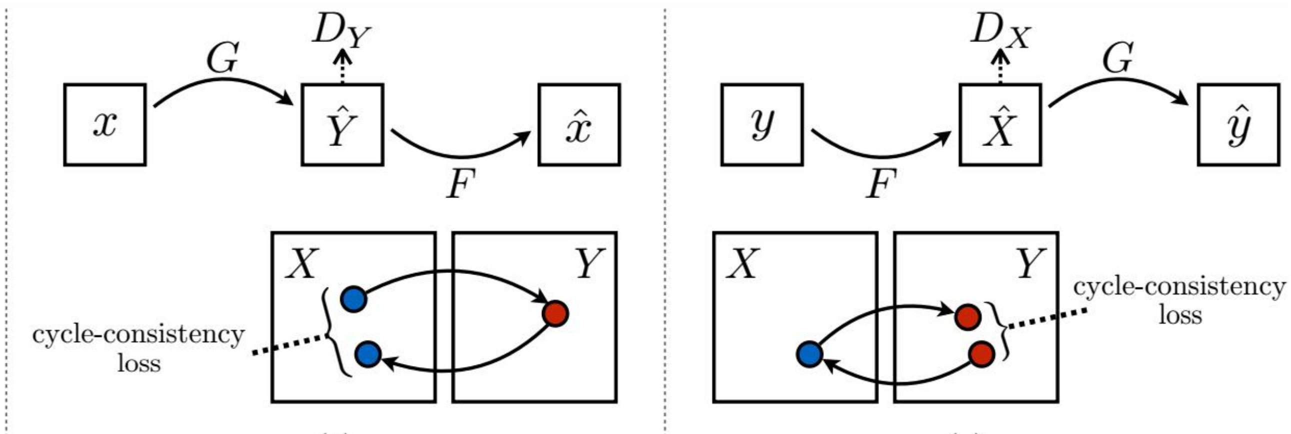
Arruda *et al.* IJCNN 2019

Cycle-consistent generative adversarial networks for unsupervised image-to-image (UI2I) translation

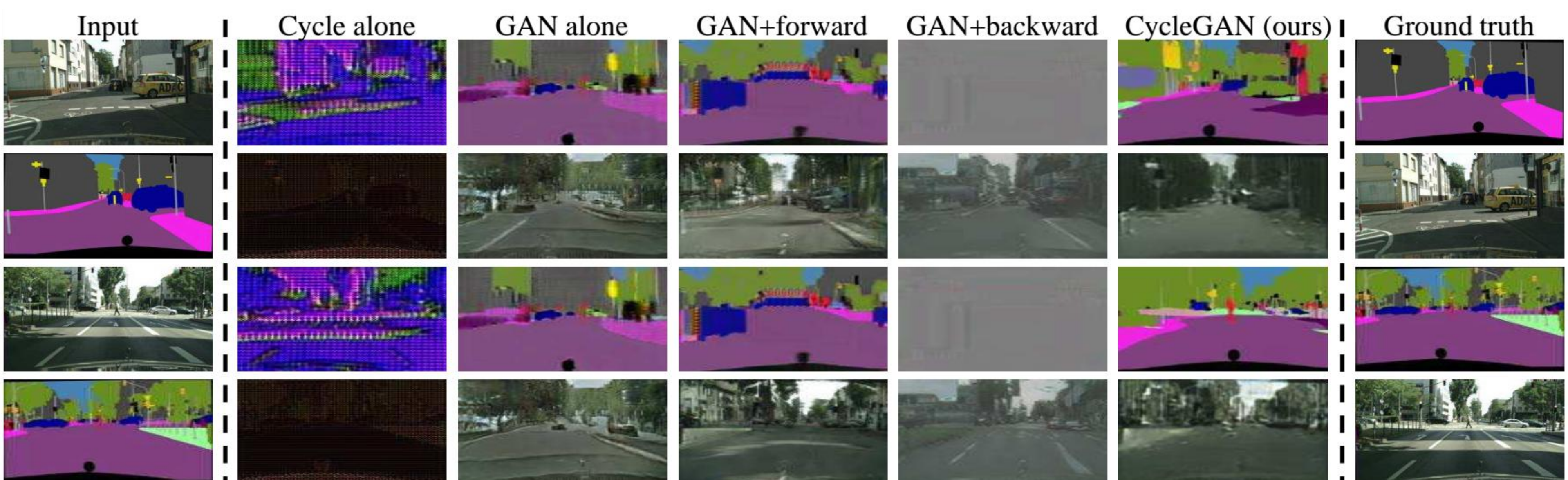
- **Adversarial loss**
- **Cycle-consistent loss**



(a)

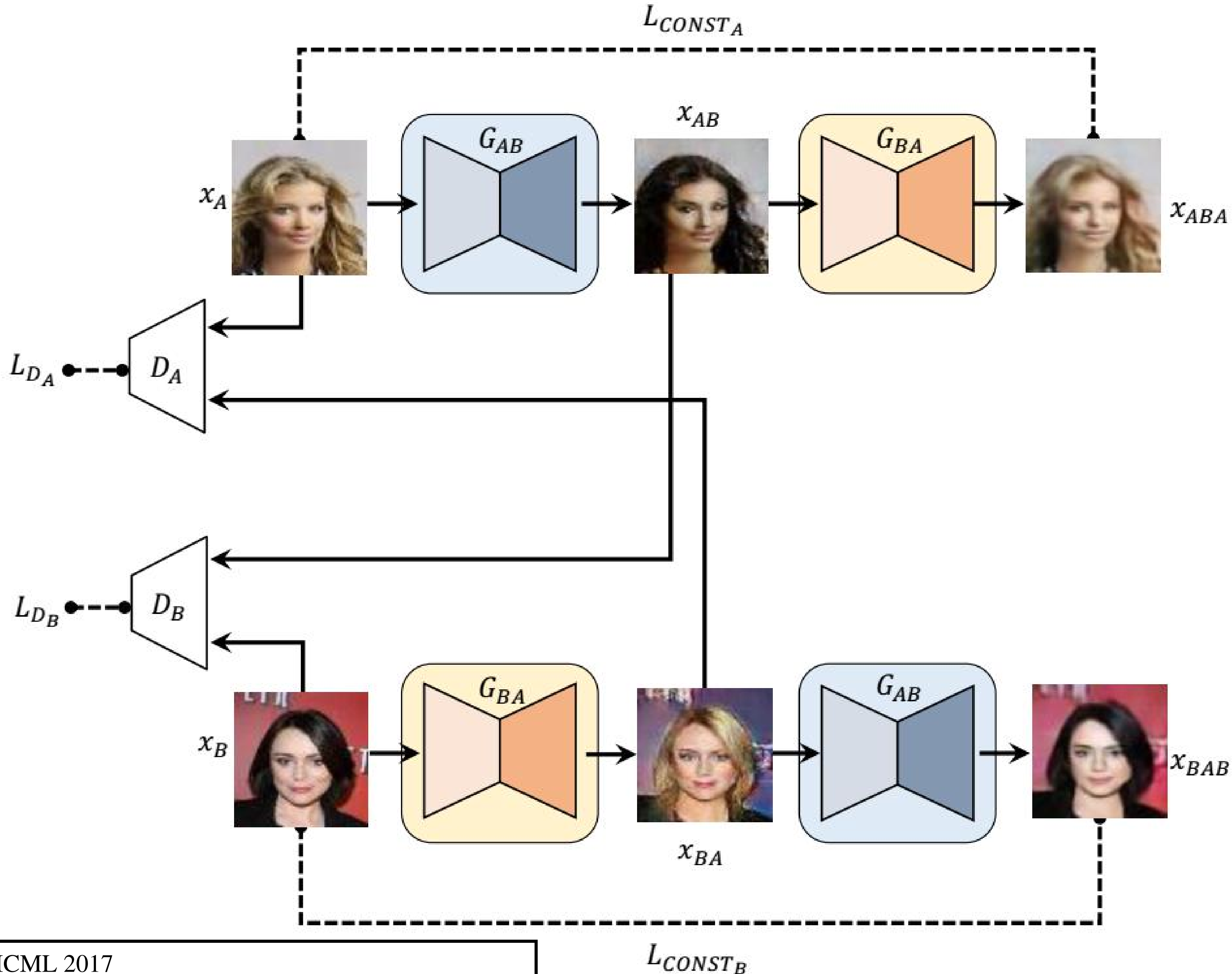


(b) (c)



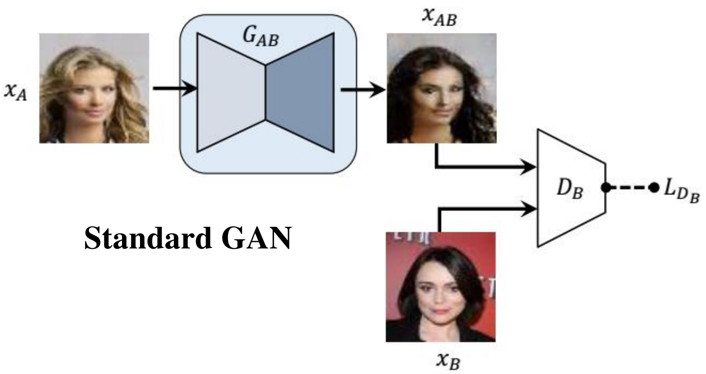
Zhu *et al.* ICCV 2017

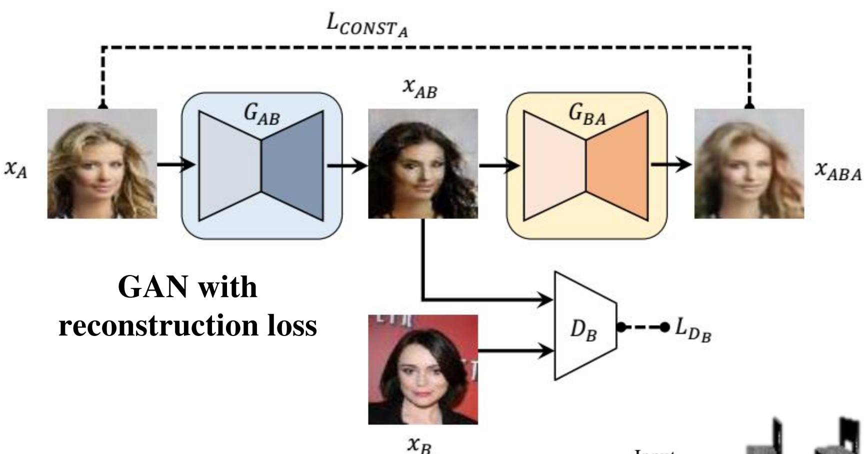
Discover relations between different domains for unpaired image-to-image (I2I) translation

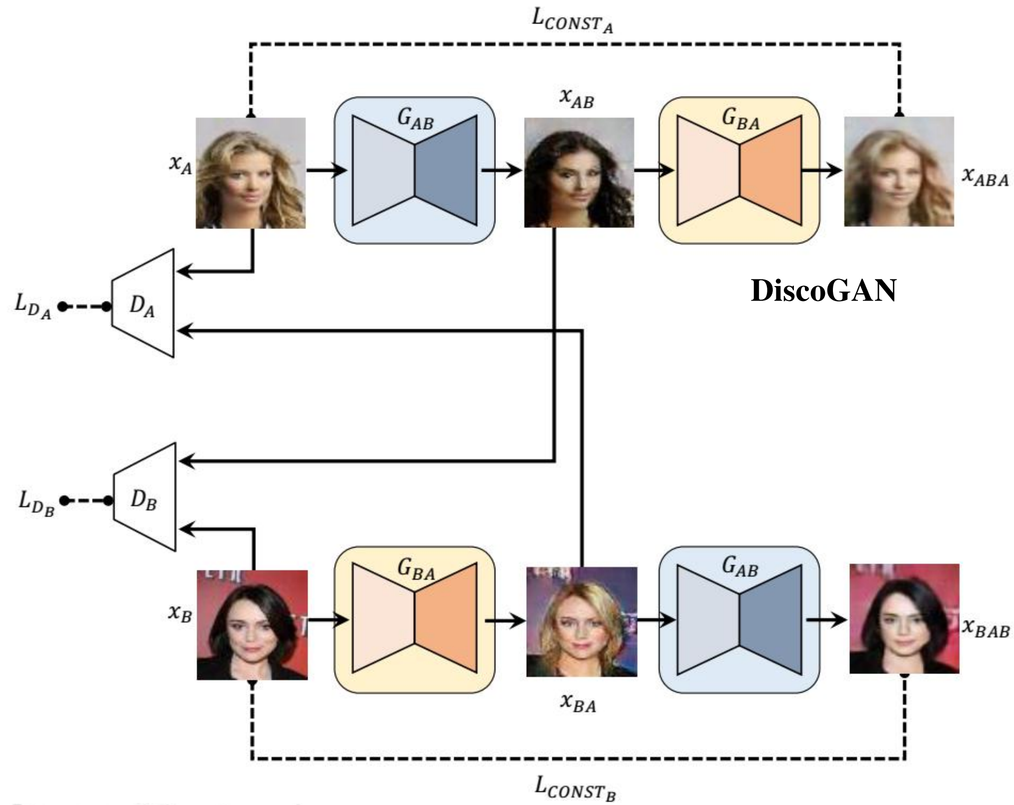


Kim *et al.* ICML 2017

Discover relations between different domains for unpaired image-to-image (I2I) translation



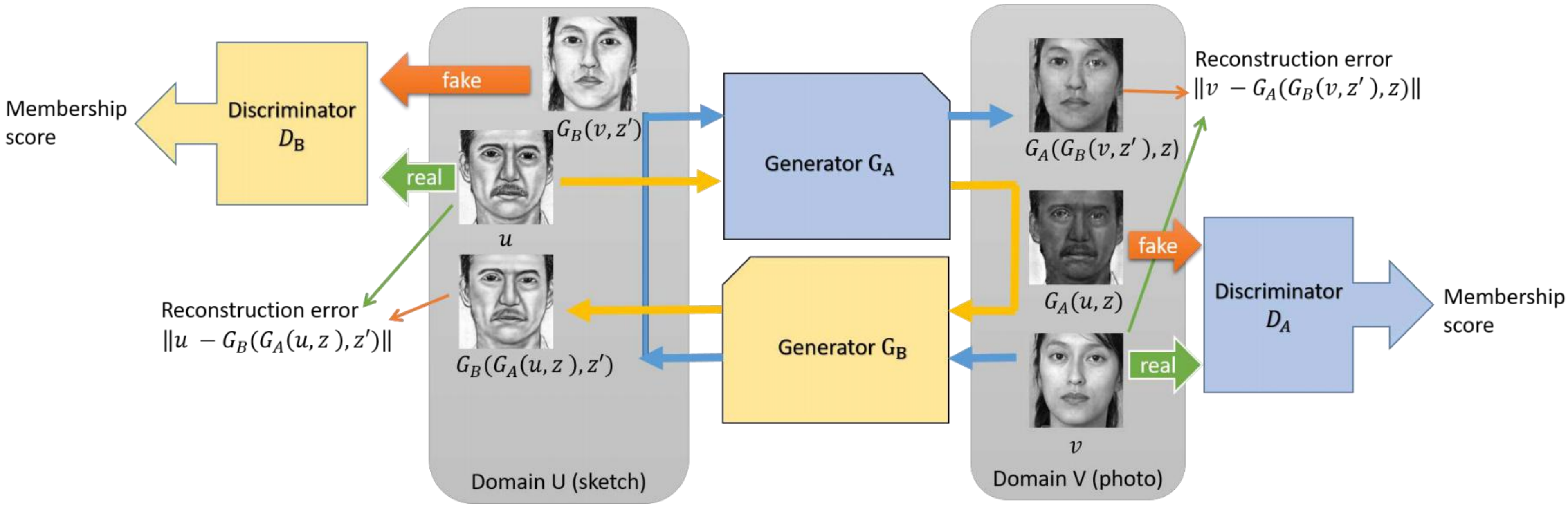




(a) Chair to Car (b) Car to Face

Kim *et al.* ICML 2017

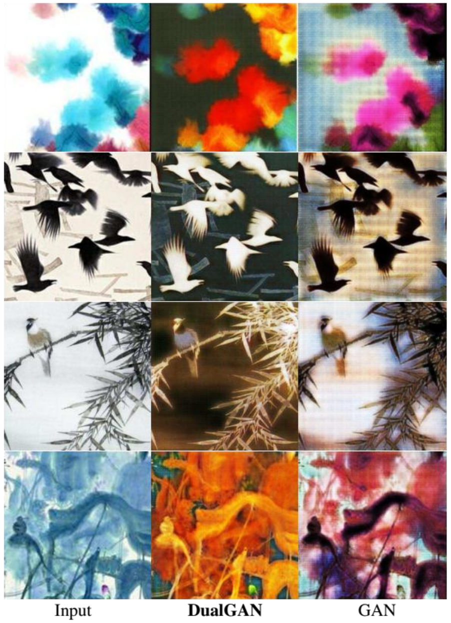
Unsupervised dual learning for image-to-image (I2I) translation



Algorithm 1 DualGAN training procedure

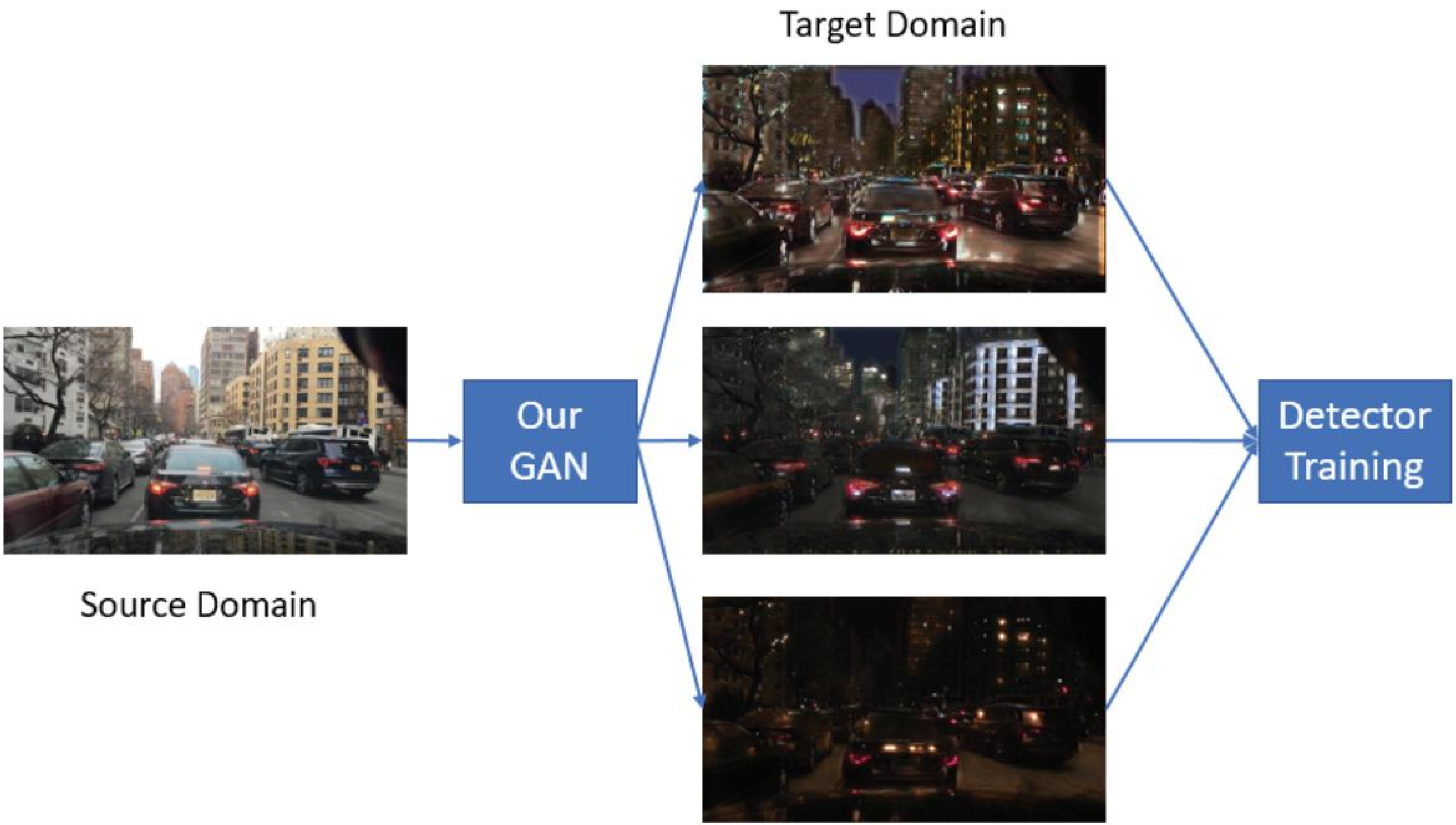
Require: Image set U, image set V, GAN A with generator parameters θ_A and discriminator parameters ω_A, GAN B with generator parameters θ_B and discriminator parameters ω_B, clipping parameter c, batch size m, and n_{critic}

```
Randomly initialize ω_i, θ_i, i ∈ {A, B}
repeat
    for t = 1, ..., n_critic do
        sample images {u^(k)}_{k=1}^m ⊆ U, {v^(k)}_{k=1}^m ⊆ V
        update ω_A to minimize (1/m) Σ_{k=1}^m l_A^d(u^(k), v^(k))
        update ω_B to minimize (1/m) Σ_{k=1}^m l_B^d(u^(k), v^(k))
        clip(ω_A, −c, c), clip(ω_B, −c, c)
    end for
    sample images {u^(k)}_{k=1}^m ⊆ U, {v^(k)}_{k=1}^m ⊆ V
    update θ_A, θ_B to minimize (1/m) Σ_{k=1}^m l^g(u^(k), v^(k))
until convergence
```



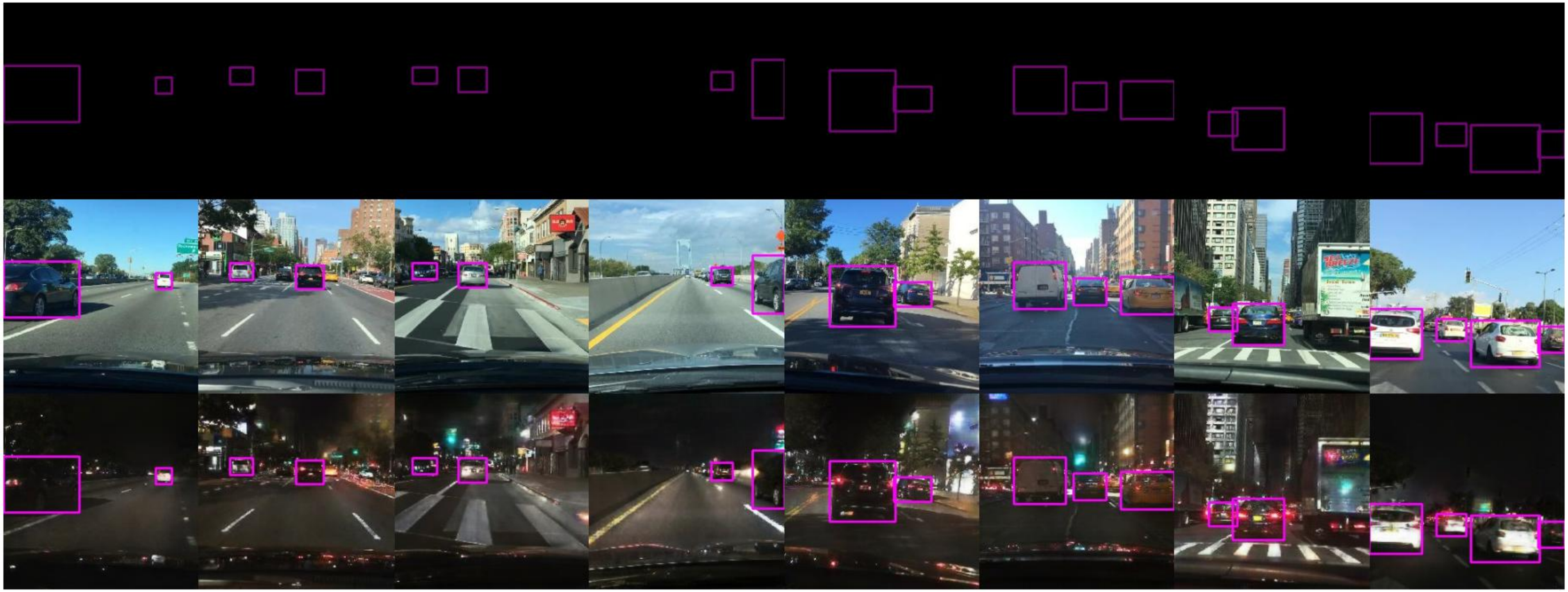
Yi *et al.* ICCV 2017

Unsupervised image-to-image (UI2I) translation for cross domain object detection - Entire Framework



Lin *et al.* AAAI 2020

Unsupervised image-to-image (UI2I) translation for cross domain object detection - annotation transfer process

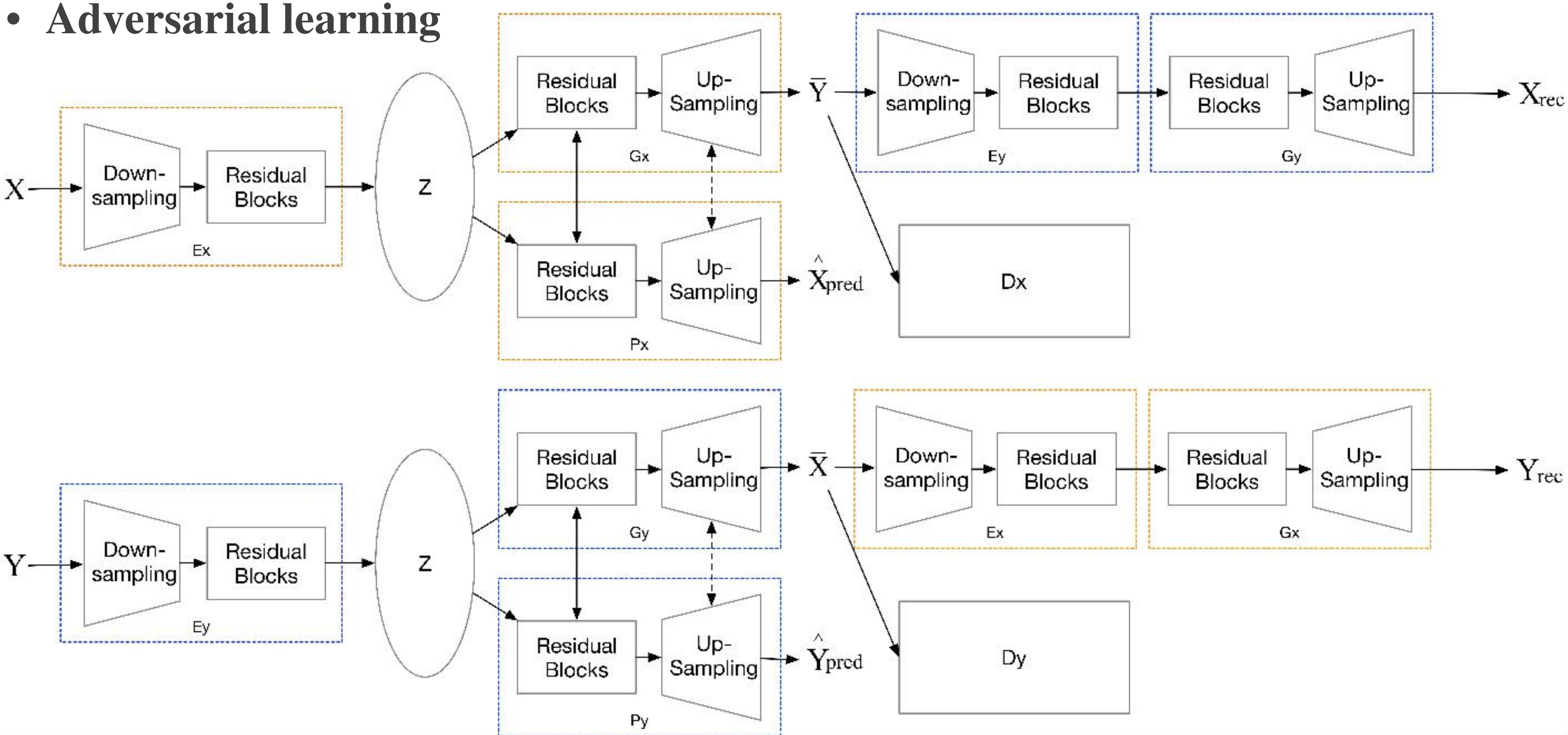


- First row: annotations for the ***source*** domain images
- Second row: annotations on the ***source*** domain images
- Third row: transfered annotations on the ***target*** domain images

Arruda *et al.* IJCNN 2019

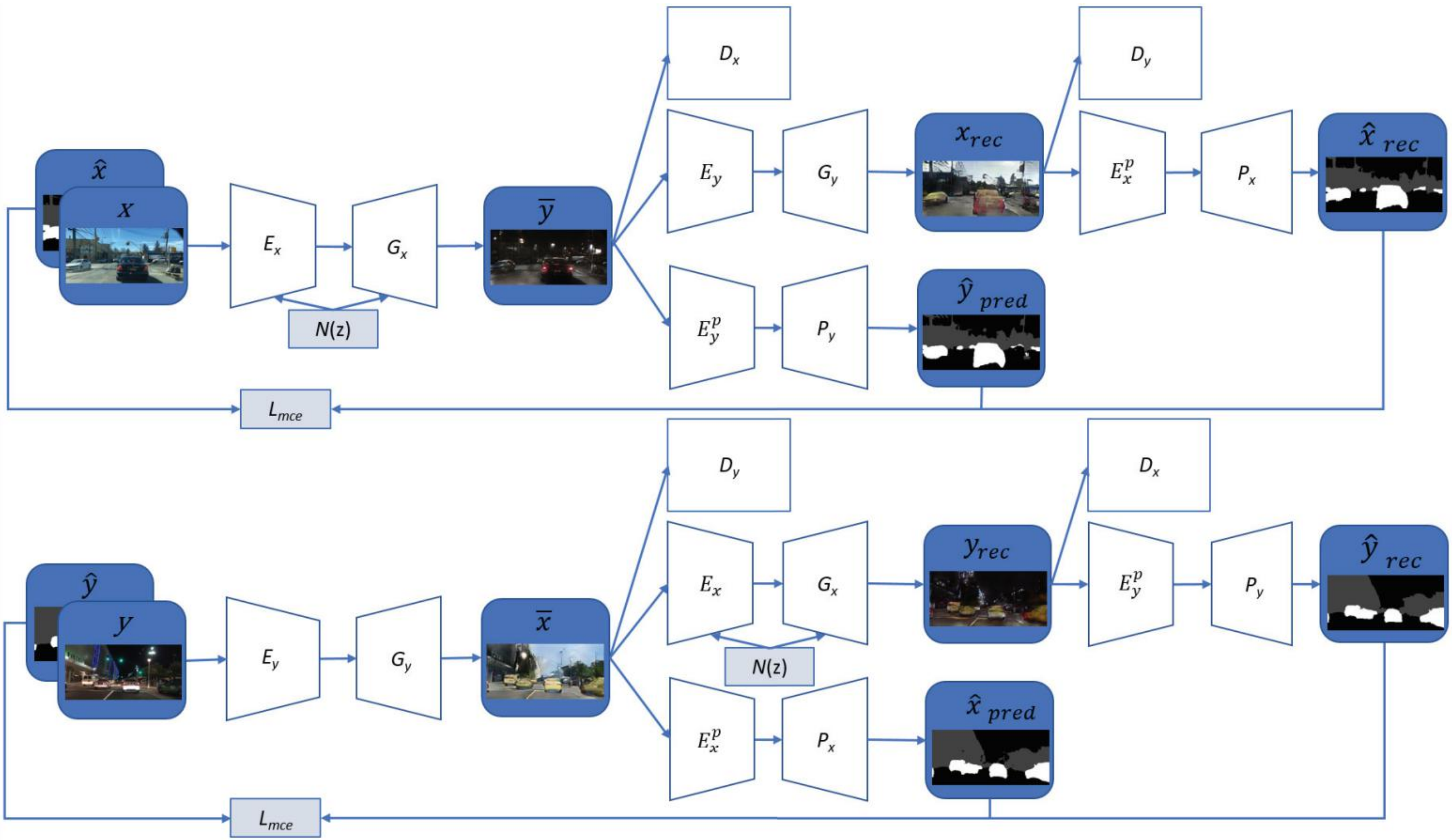
AugGAN for structure-consistent multimodal unsupervised image-to-image translation

- **Structure-aware encoding and segmentation subtask**
- **Weight sharing for multi-task network**
- **Cycle-consistency**
- **Adversarial learning**



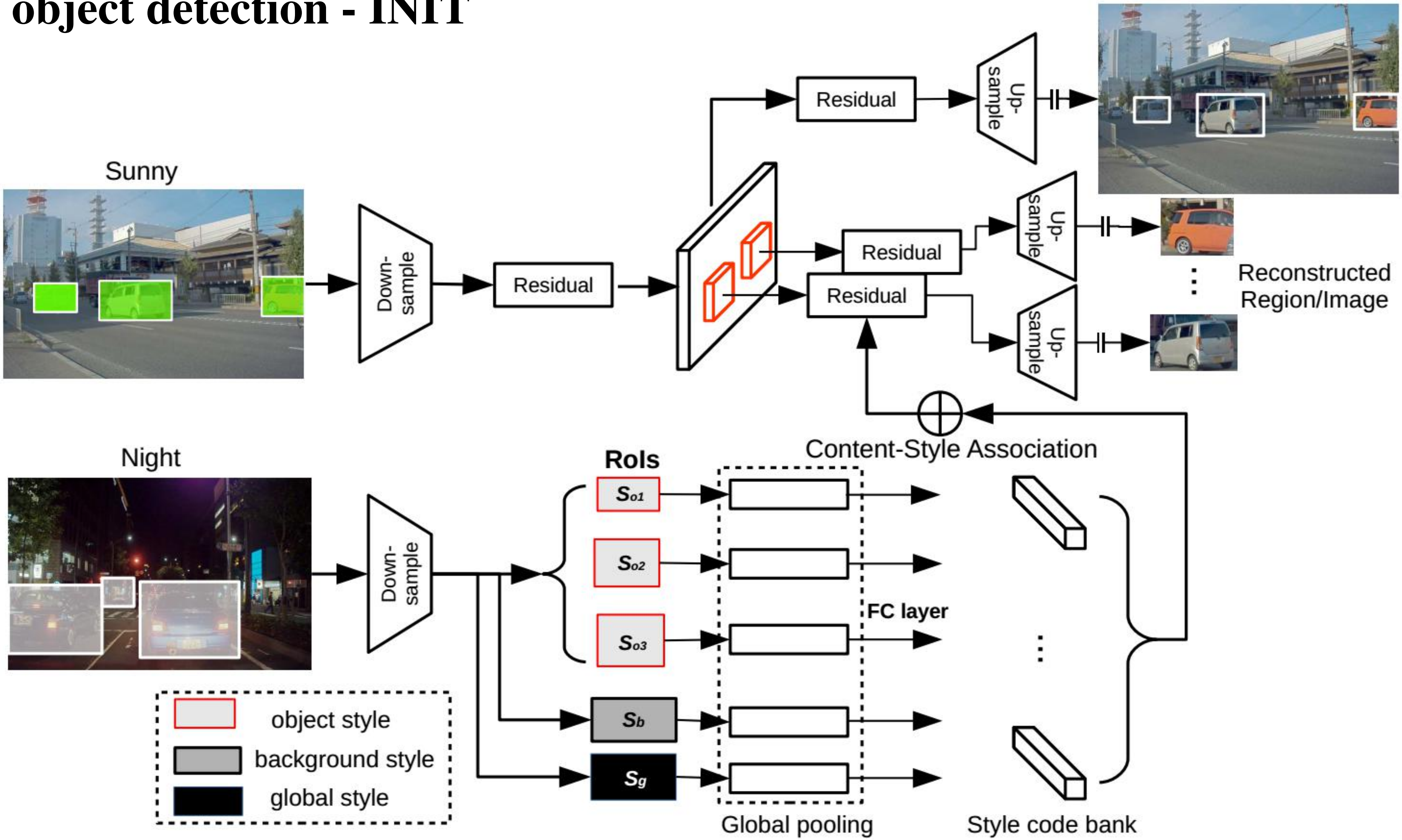
Huang *et al.* ECCV 2018

Multimodal AugGAN for structure-consistent multimodal unsupervised image-to-image translation



Lin *et al.* AAAI 2020

Unsupervised image-to-image (UI2I) translation for cross domain object detection - INIT



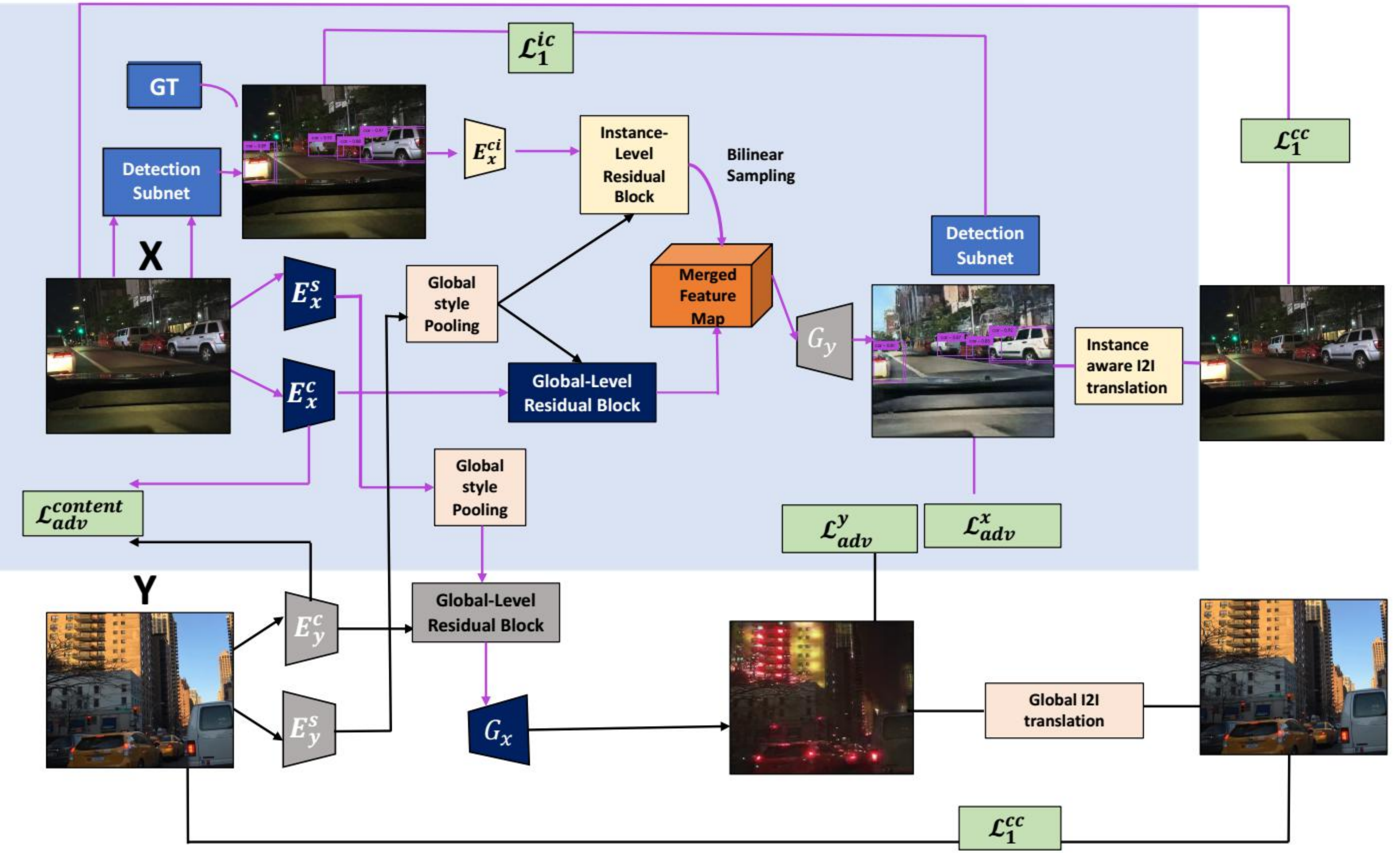
Shen *et al.* CVPR 2019

INIT for cross domain object detection - qualitative results

Sunny Night Cloudy Rainy

Shen *et al.* CVPR 2019

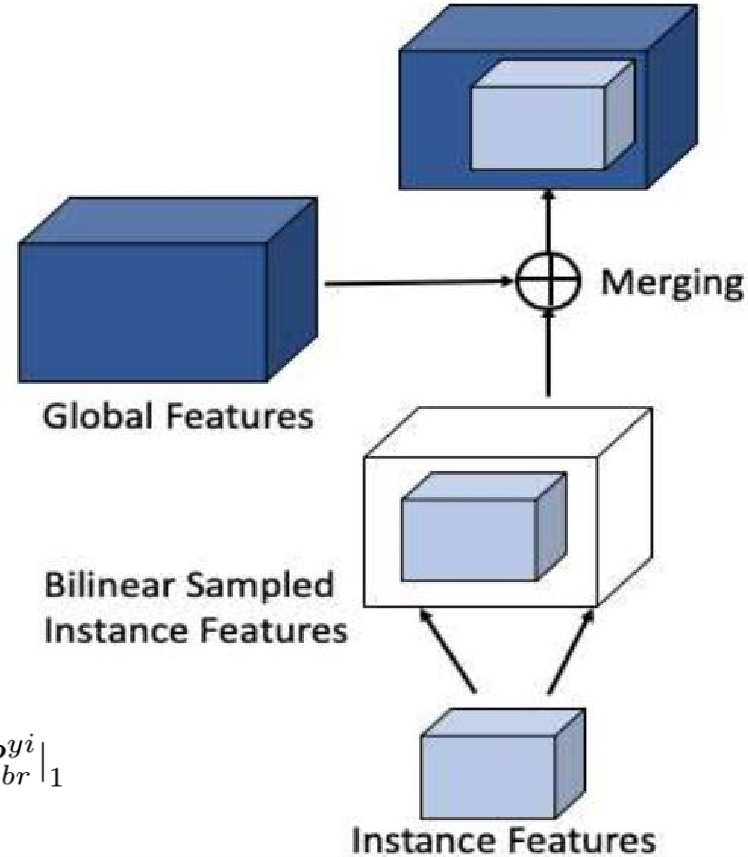
Unsupervised image-to-image (UI2I) translation for cross domain object detection - DUNIT



Bhattacharjee *et al.* CVPR 2020

Unsupervised image-to-image (UI2I) translation for cross domain object detection - DUNIT

- **Image-to-image translation module**
  - Content adversarial loss
  - Domain adversarial loss
  - Cross-cycle consistency loss
  - Self-reconstruction loss
  - KL loss
  - Latent regression loss
- **Object detection module**
  - Instance consistency loss
    - $\mathcal{L}_1^{ic} = \sum_{i|\hat{y}_i^x=1\wedge\hat{y}_i^y=1} |P_{tl}^{xi} - P_{tl}^{yi}|_1 + \sum_{i|\hat{y}_i^x=1\wedge\hat{y}_i^y=1} |P_{br}^{xi} - P_{br}^{yi}|_1$



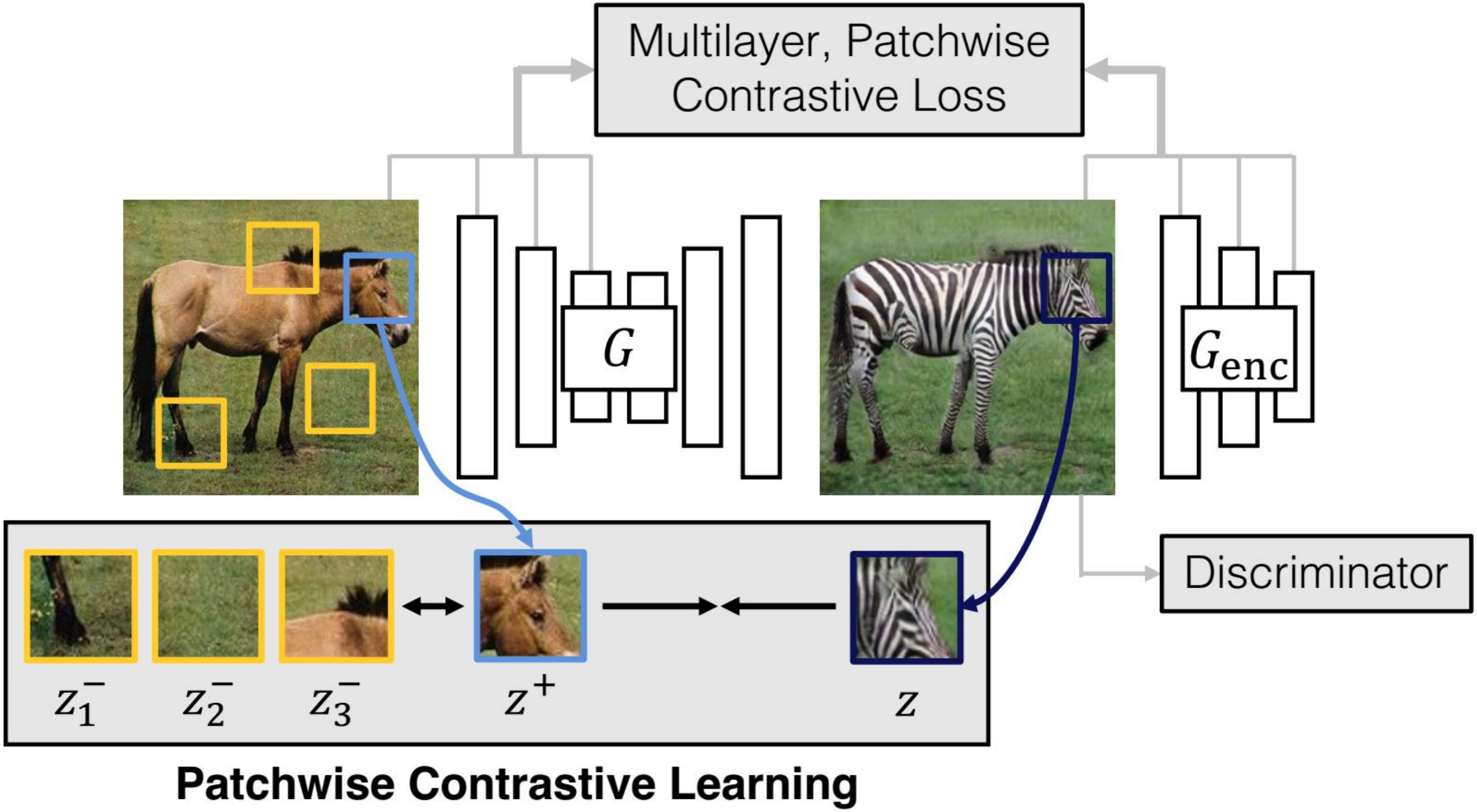
Bhattacharjee *et al.* CVPR 2020

DUNIT for unsupervised image-to-image (UI2I) translation and cross domain object detection - qualitative results

Input | Style Image | MUNIT | DRIT | DUNIT

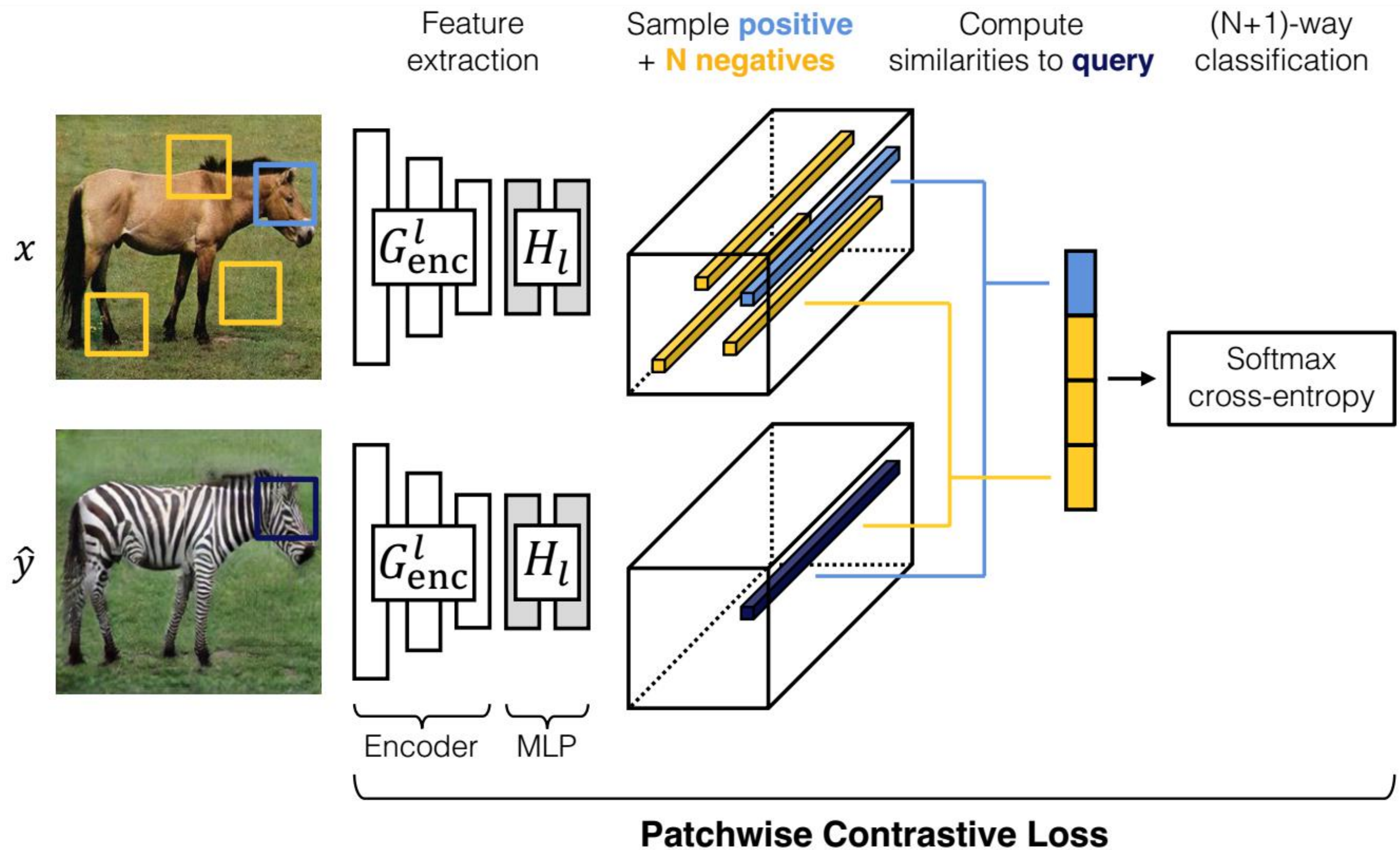
Bhattacharjee *et al.* CVPR 2020

Patchwise contrastive learning for one-sided unsupervised image-to-image translation (CUT)

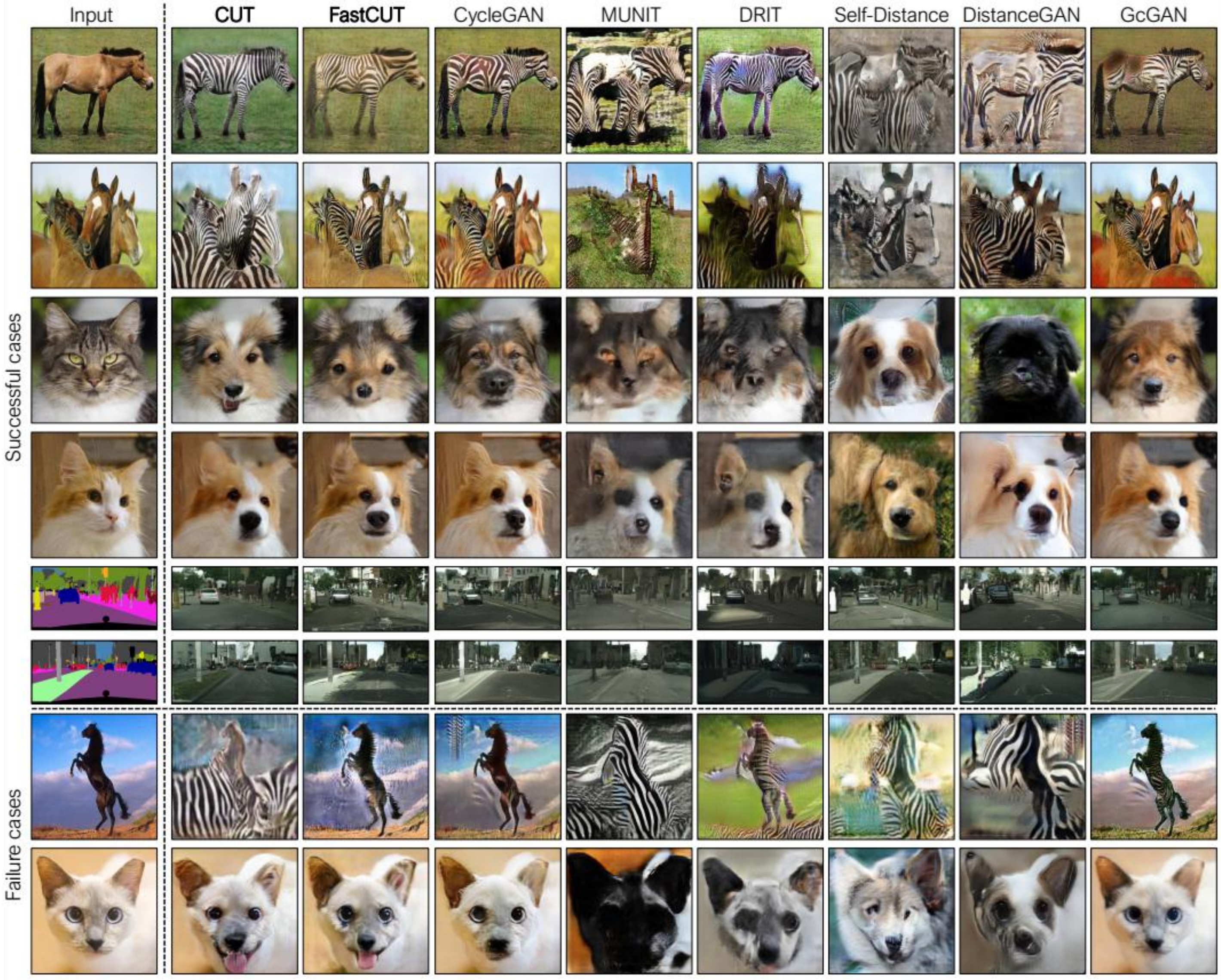


Park *et al.* ECCV 2020

Patchwise contrastive loss

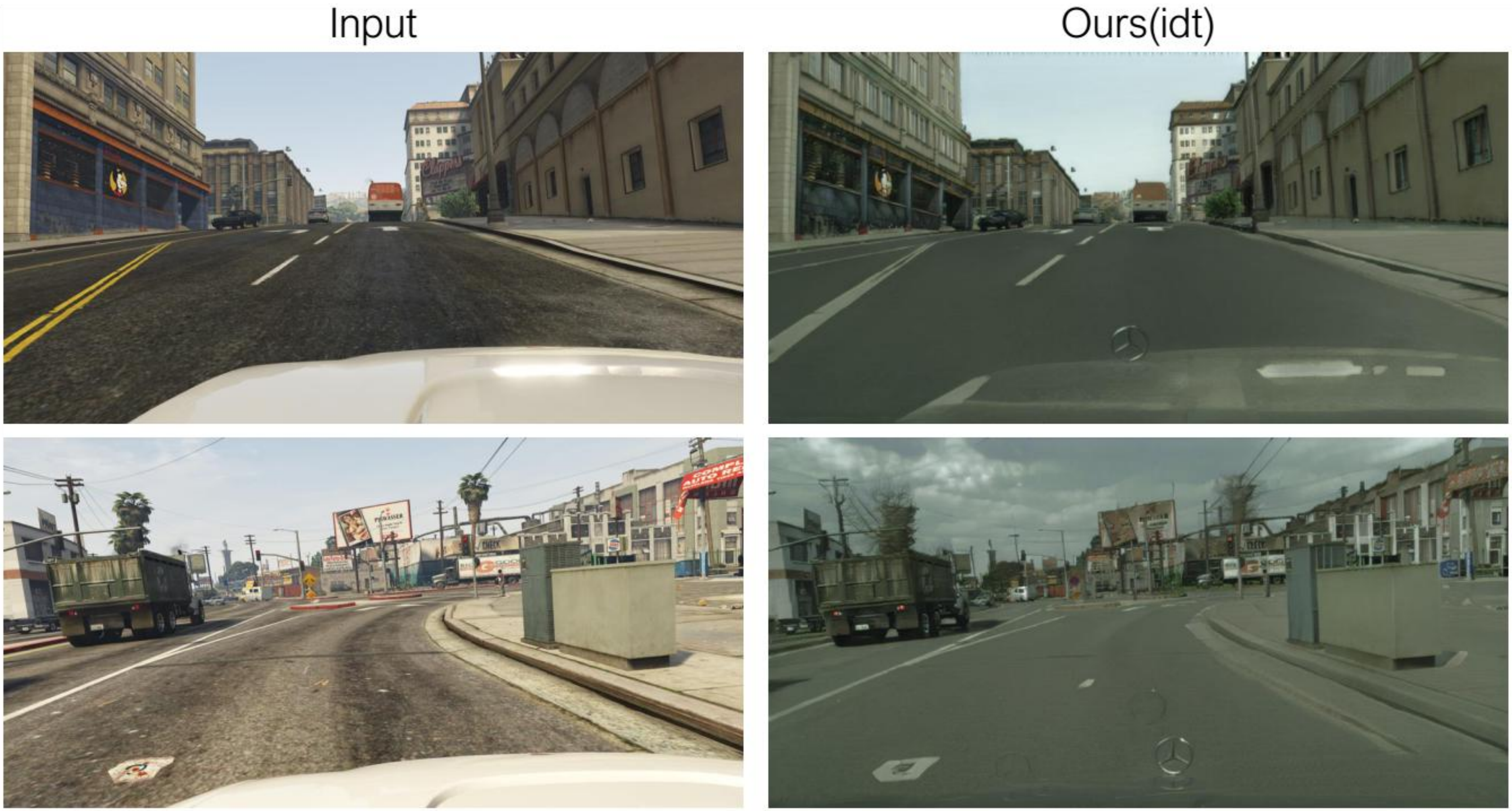


Contrastive learning for one-sided unsupervised image-to-image translation - qualitative results



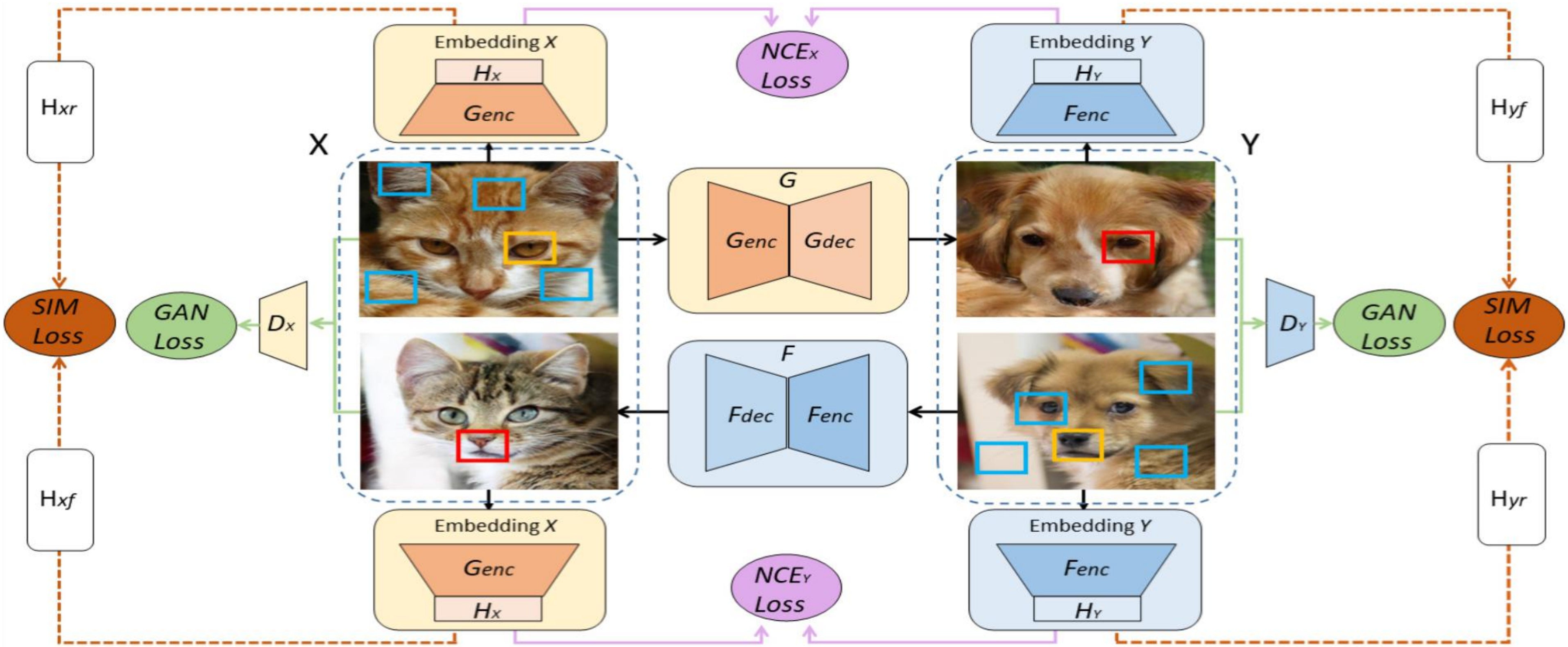
Park *et al.* ECCV 2020

Contrastive learning for one-sided unsupervised image-to-image translation (CUT) - high resolution images synthesis



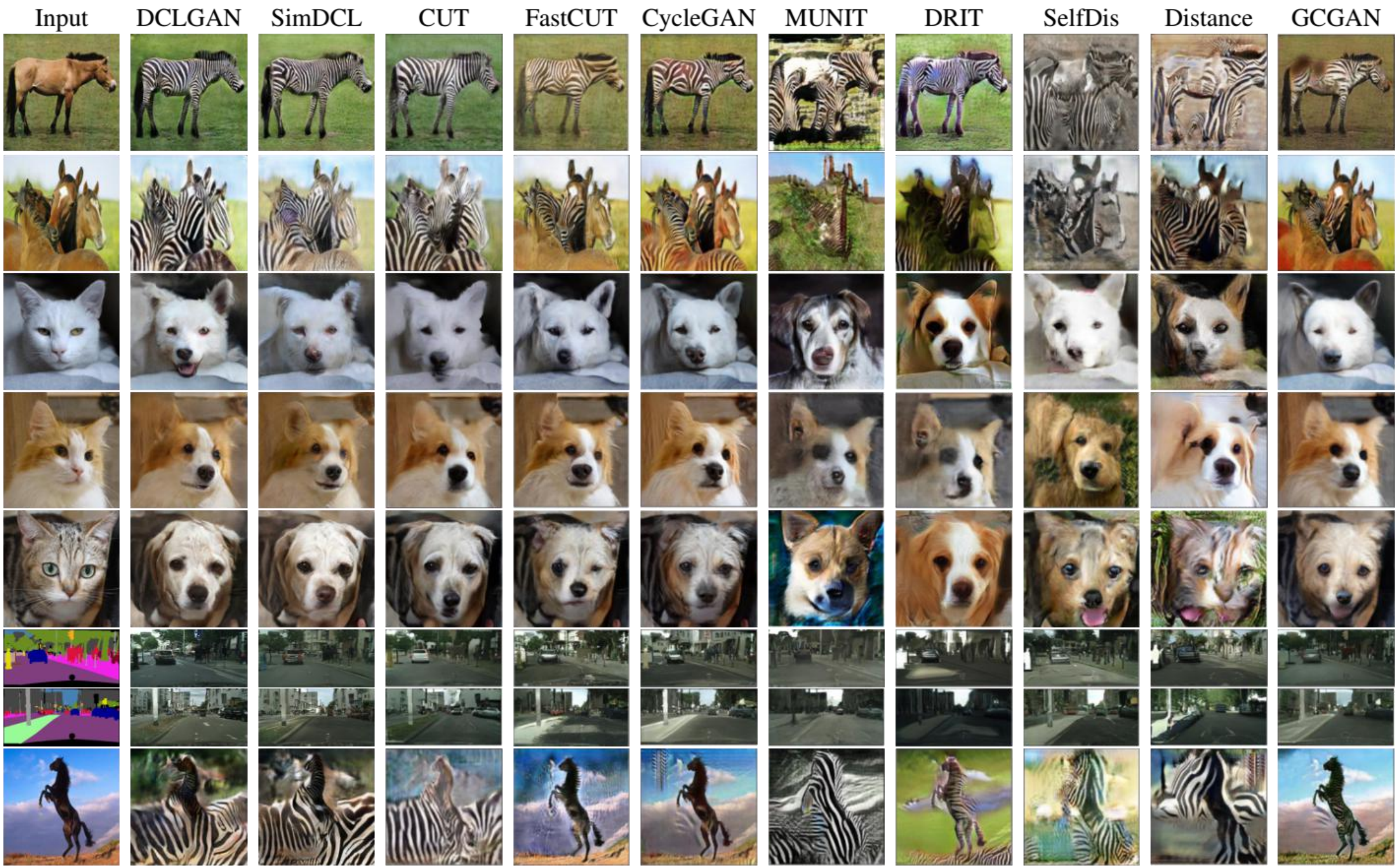
Dual contrastive learning for unsupervised image-to-image (UI2I) translation

- **Patch-based multilayer contrastive learning**
- **Similarity loss**
- **Identity loss**
  - Encourage the mapping to preserve color composition between the input and output



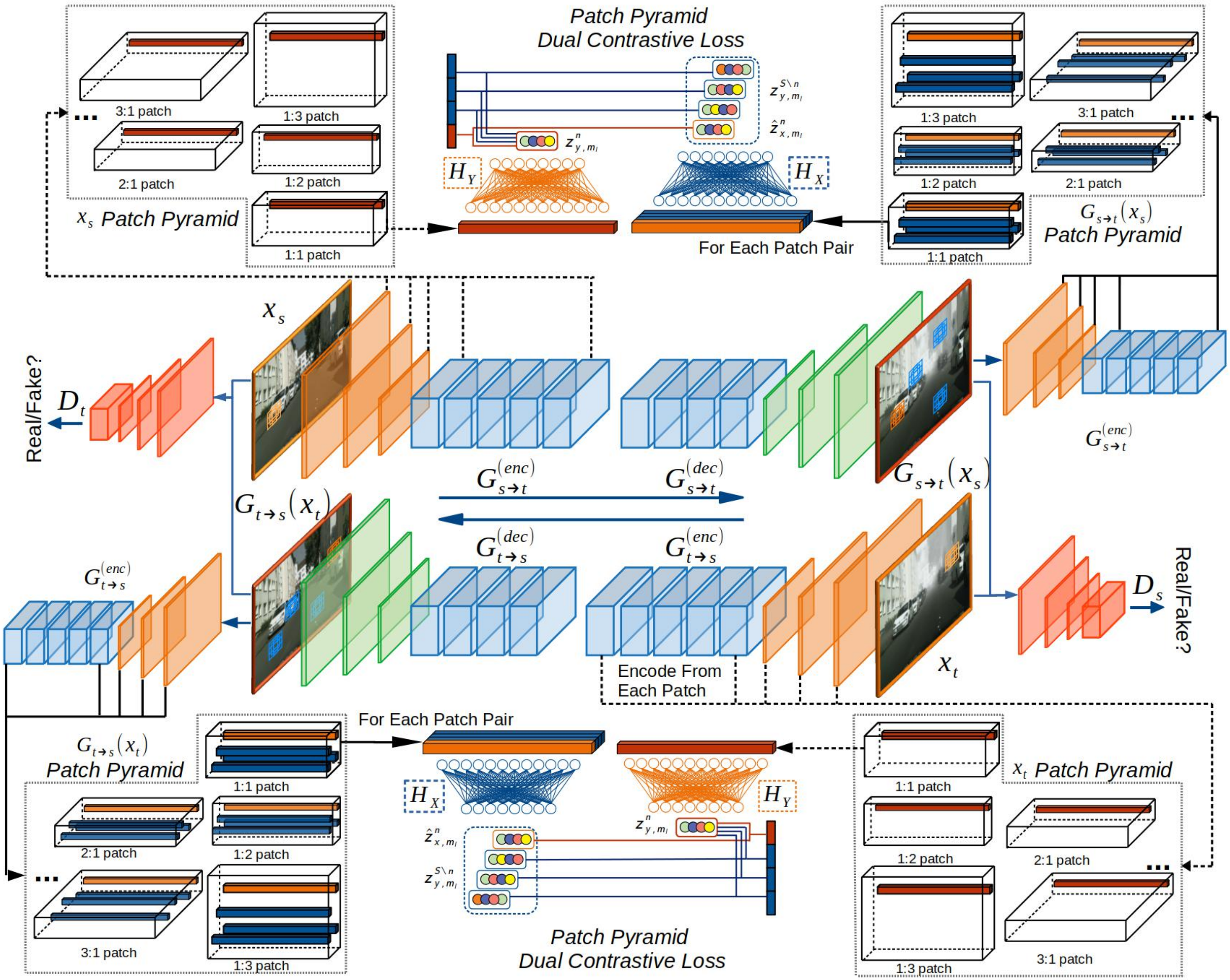
Han *et al.* CVPR 2021 workshop

Dual contrastive learning for unsupervised image-to-image (UI2I) translation - qualitative results



Han *et al.* CVPR 2021 workshop

Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation



Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation - objectives

- **Adversarial loss**
  - Apply least-square adversarial loss to encourage the generated images to produce visually similar images as the target domain
    - $\mathcal{L}_{adv}(G_{s\to t}, D_t, X, Y) = \mathbb{E}_{y\sim Y}\{(D_t(y)-1)^2\} + \mathbb{E}_{x\sim X}\left\{\left(D_t(G_{s\to t}(x))\right)^2\right\}$
- **Patch pyramid dual contrastive learning (PPCL)**
  - Employ *PatchPyramidNCE* loss to match the corresponding input-output patch pair for each single patch
    - $\mathcal{L}_{PatchPyramidNCE_X}(G_{s\to t}, H_X, H_Y, X) = \mathbb{E}_{x\sim X}\sum_{m=1}^{M}\sum_{l=1}^{L}\sum_{s=1}^{S_{m_l}}\rho(z_{m_l}^s, \hat{z}_{m_l}^s, z_{m_l}^{S\backslash s})$
    - $\mathcal{L}_{PatchPyramidNCE_Y}(G_{t\to s}, H_X, H_Y, Y) = \mathbb{E}_{y\sim Y}\sum_{m=1}^{M}\sum_{l=1}^{L}\sum_{s=1}^{S_{m_l}}\rho(z_{m_l}^s, \hat{z}_{m_l}^s, z_{m_l}^{S\backslash s})$
      - $\rho\left(z_{m_l}^s, \hat{z}_{m_l}^s, z_{m_l}^{S\backslash s}\right) = \log\left[\frac{\exp(\frac{z_{m_l}^s\cdot\hat{z}_{m_l}^s}{\tau})}{\exp\left(\frac{z_{m_l}^s\cdot\hat{z}_{m_l}^s}{\tau}\right)+\sum_{n=1,n\neq s}^{S_{m_l}}\exp\left(\frac{z_{m_l}^s\cdot\hat{z}_{m_l}^{S\backslash s}}{\tau}\right)}\right]$
        - Extracted features: $z_{m_l}^s = H_X^{(l)} G_{s\to t}^{enc\,(l)}(P_m(x_s))$
        - Positive patch: $\hat{z}_{m_l}^s = H_X^{(l)} G_{t\to s}^{enc\,(l)}\left(G_{s\to t}(P_m(x_s))\right)$
        - Negative patches: $z_{m_l}^{S\backslash s}$
- **Identity loss**
  - Further enforce the generators to produce content-preserved and visually similar images
    - $\mathcal{L}_{idt}(G_{s\to t}, G_{t\to s}) = \mathbb{E}_{x\sim X}\{\|G_{t\to s}(x) - x\|_1\} + \mathbb{E}_{y\sim Y}\{\|D_{s\to t}(y) - y\|_1\}$

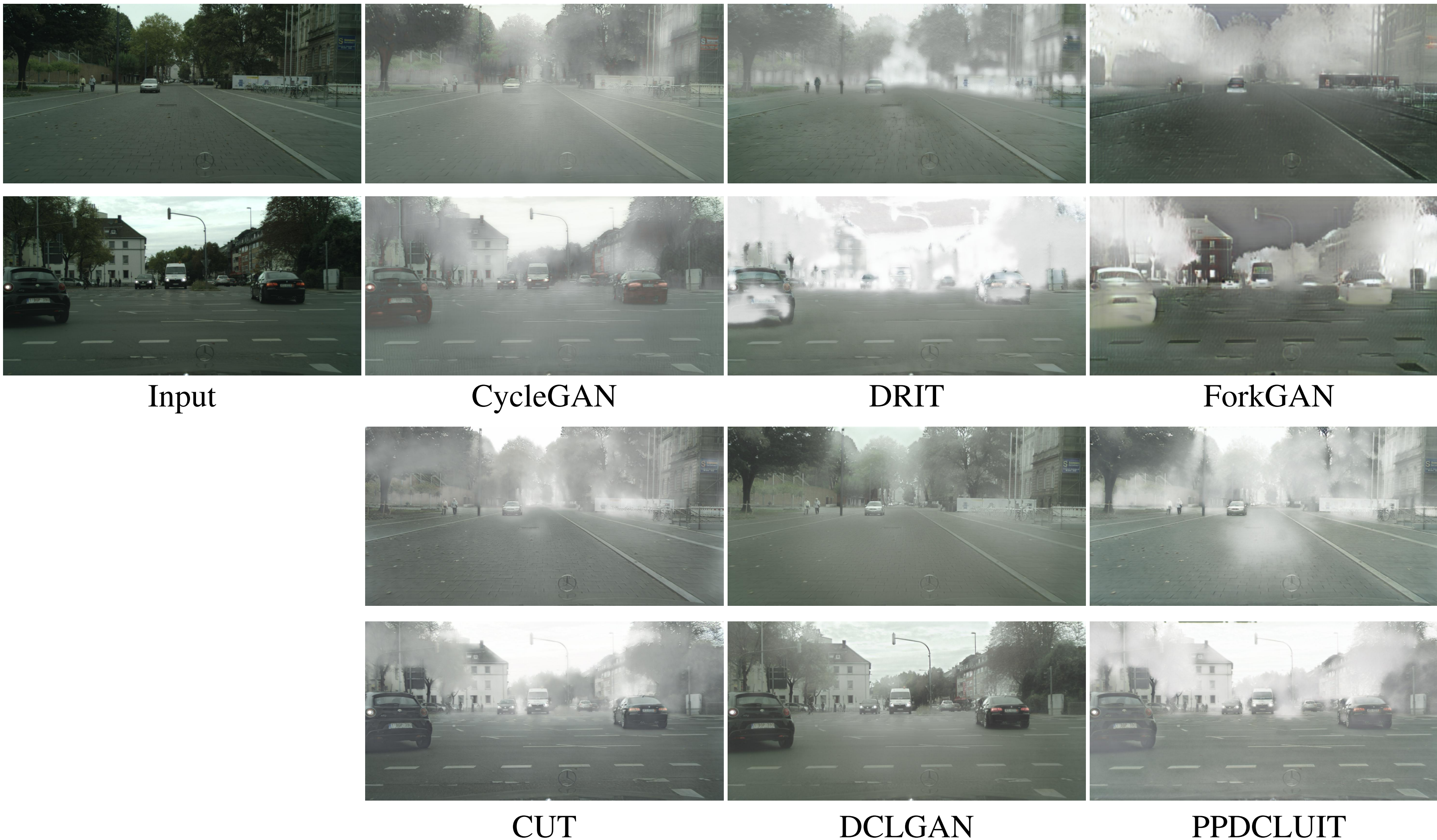
Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation - overall objectives

- **Full objectives**
  - $$\begin{aligned}\mathcal{L}(G_{s\to t}, G_{t\to s}, D_s, D_t, X, Y) = {} & \lambda_{idt}\mathcal{L}_{idt}(G_{s\to t}, G_{t\to s}) \\ & + \lambda_{adv}[\mathcal{L}_{adv}(G_{s\to t}, D_t, X, Y) + \mathcal{L}_{adv}(G_{t\to s}, D_s, Y, X)] \\ & + \lambda_{pp}\mathcal{L}_{PatchPyramidNCE_X}(G_{s\to t}, H_X, H_Y, X) \\ & + \lambda_{pp}\mathcal{L}_{PatchPyramidNCE_Y}(G_{t\to s}, H_X, H_Y, Y)\end{aligned}$$

Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation - quantiative results

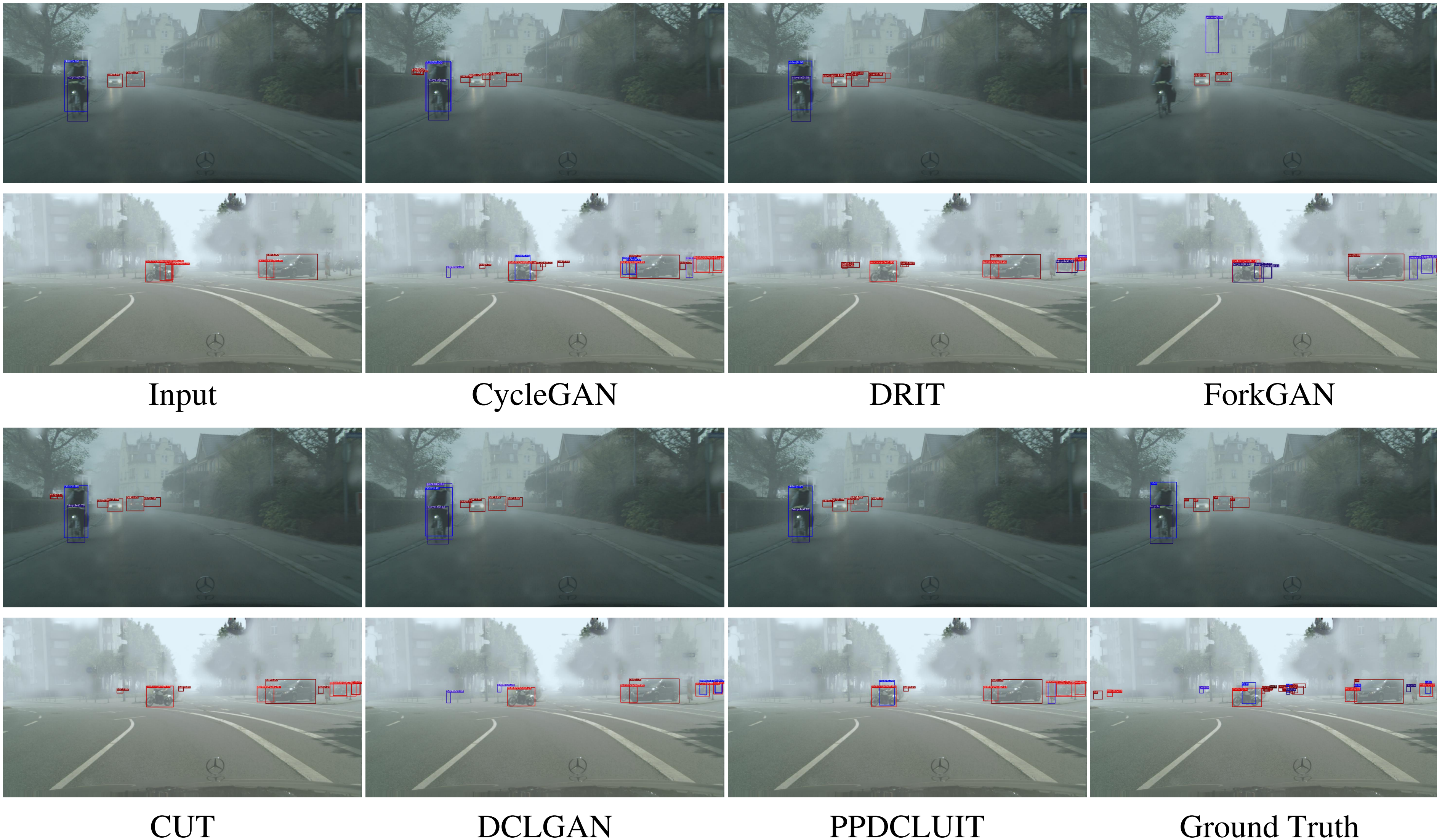
Method	FID ↓	Precision ↑	Recall ↑	Density ↑	Coverage ↑
			Cityscapes → FoggyCityscapes		
Source only	68.71	0.009	0.566	0.002	0.016
CycleGAN	45.10	**0.767**	0.304	**0.729**	0.782
CUT	42.56	0.668	0.498	0.592	**0.930**
DCLGAN	48.95	0.276	0.394	0.098	0.338
PPDCLUIT	**41.96**	0.167	**0.688**	0.053	0.340
			BDD100K Day → Night		
Source only	104.63	0.181	0.413	0.046	0.022
CycleGAN	17.90	0.763	0.594	0.828	0.729
CUT	24.17	0.711	**0.673**	0.681	0.658
DCLGAN	21.92	0.784	0.485	0.965	0.633
PPDCLUIT	**14.11**	**0.842**	0.566	**1.265**	**0.806**

Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation - qualitative results



Patch pyramid dual contrastive learning for unsupervised image-to-image (UI2I) translation - qualitative results

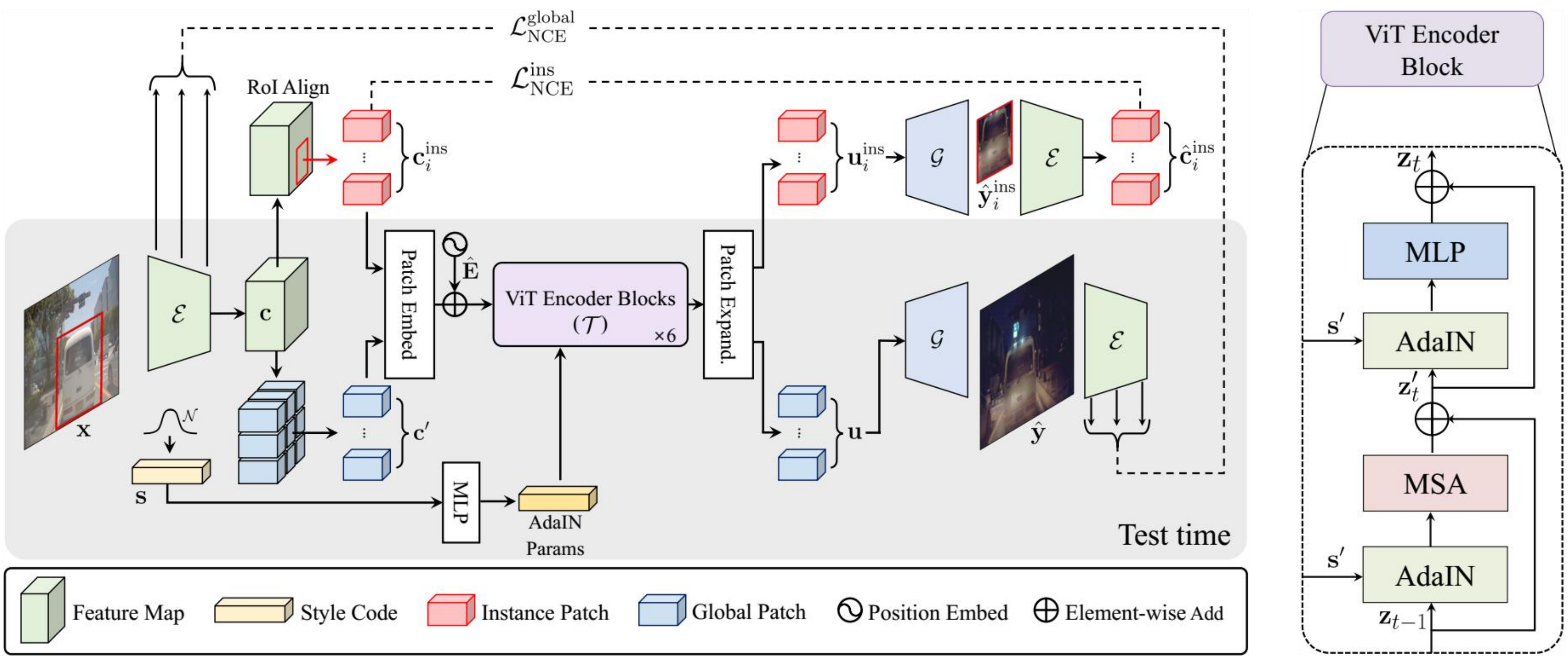
Patch pyramid dual contrastive learning for cross domain object detection - qualitative results



Patch pyramid dual contrastive learning for cross domain object detection - qualitative results

Instance-aware image-to-image translation with Transformer

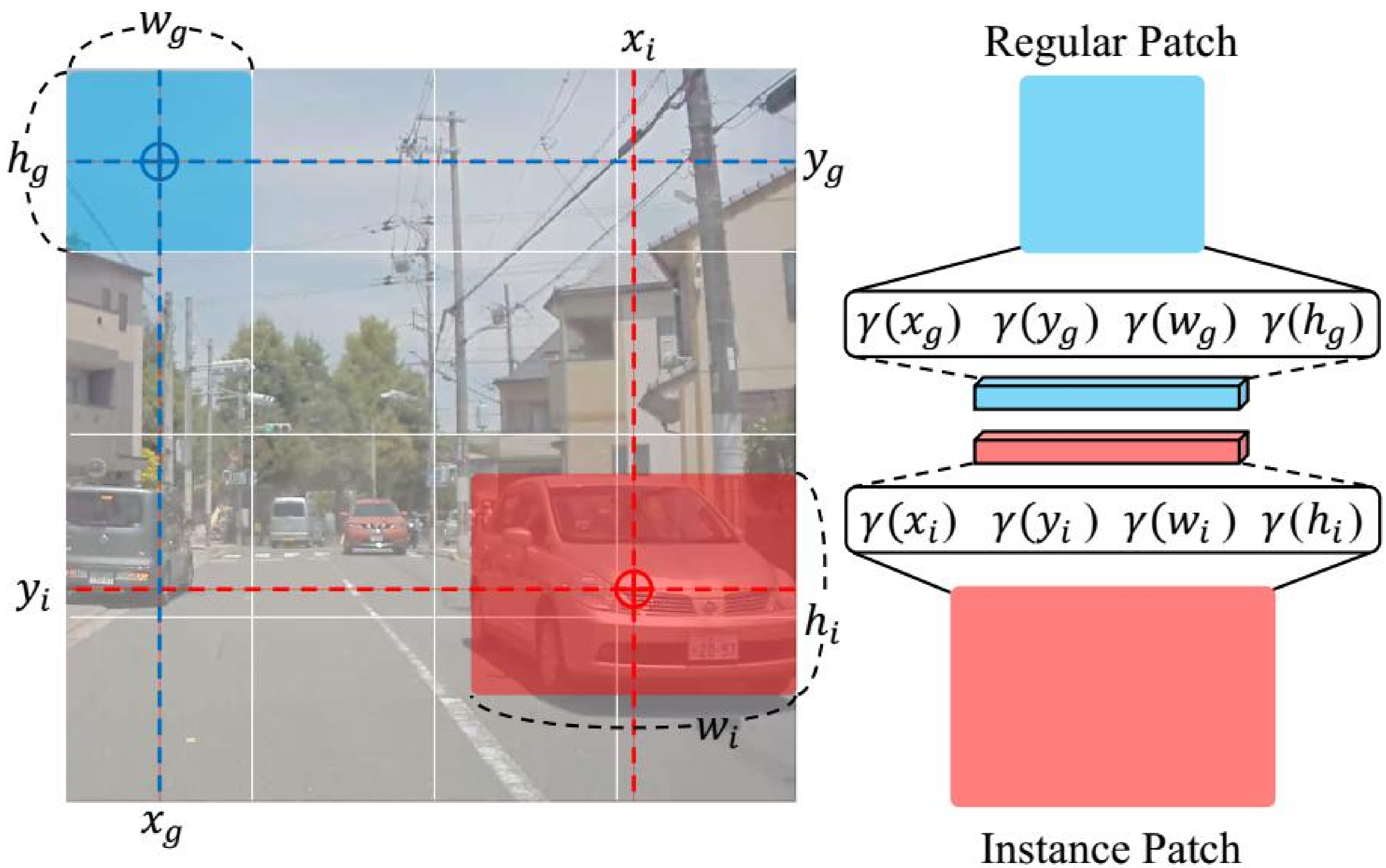
- Content and style mixing with Transformers
- Instance-aware content and style mixing
- Instance-aware position embedding



(a) Architecture (b) ViT Encoder Block

Instance-aware image-to-image translation with Transformer

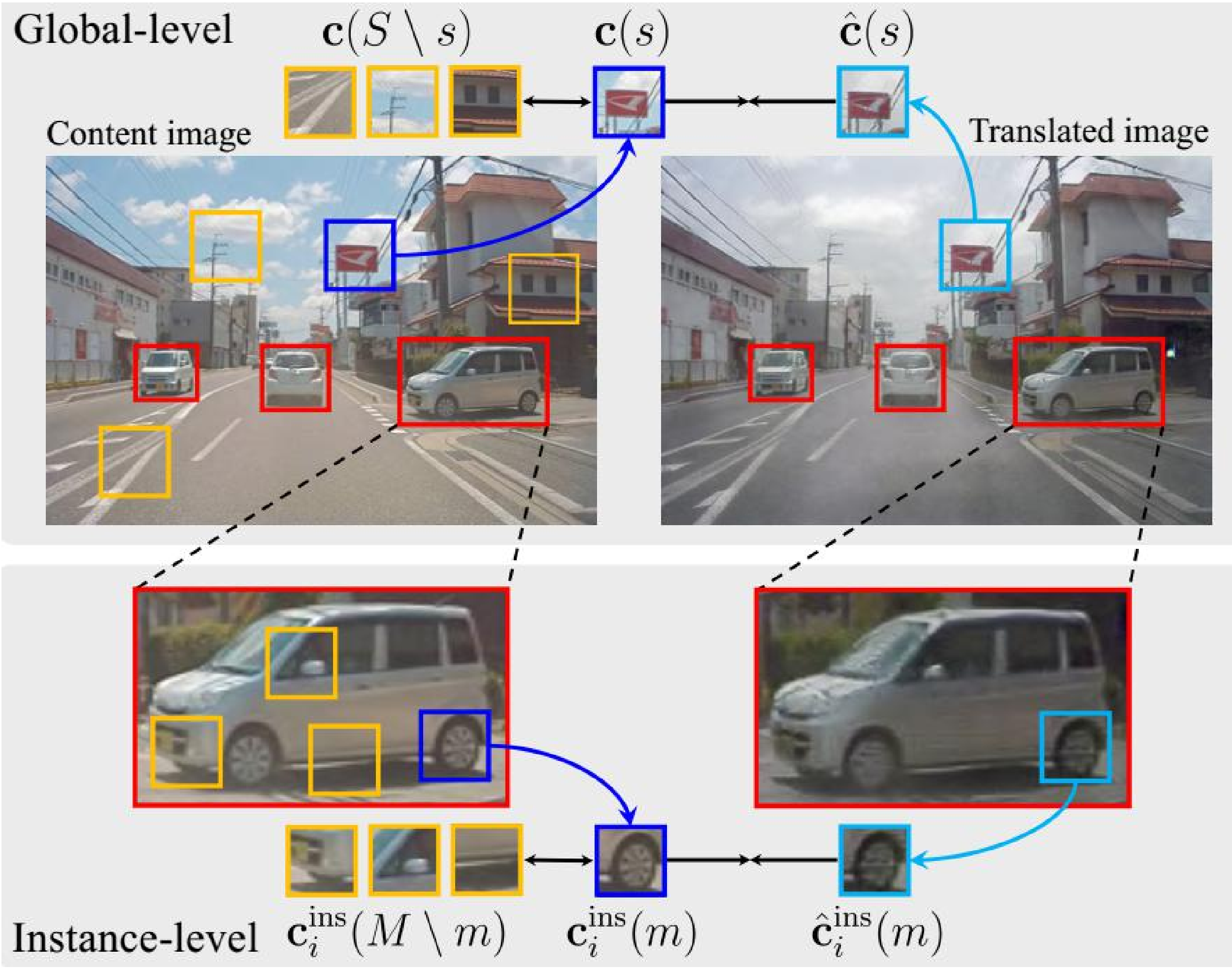
- **Building position embedding for regular patches and instance-level patches**



Kim *et al.* CVPR 2022

Instance-aware image-to-image translation with Transformer

- **Adversarial loss**
- **Global content loss**
- **Instance-level content loss**
- **Image reconstruction loss**
- **Style reconstruction loss**



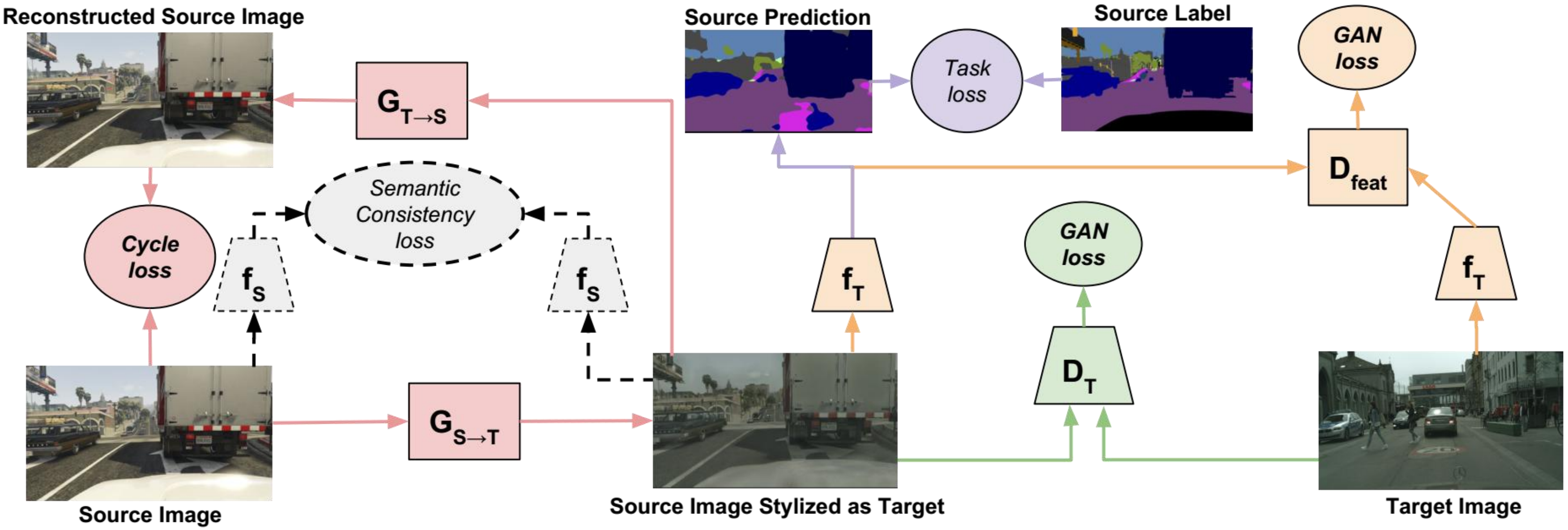
Kim *et al.* CVPR 2022

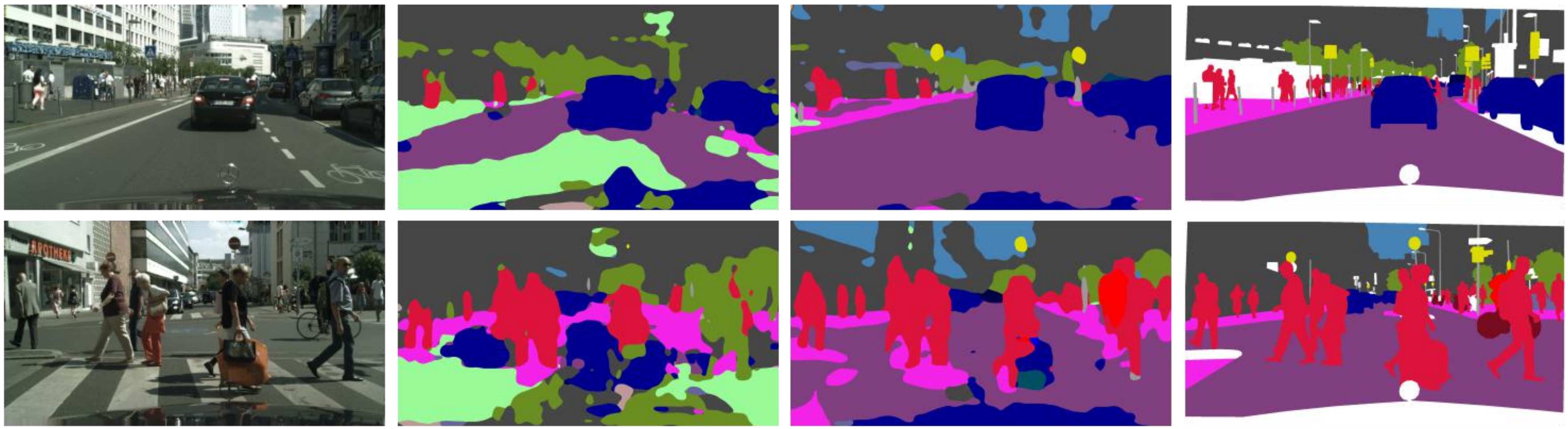
Instance-aware image-to-image translation with Transformer - qualitative results

Kim *et al.* CVPR 2022

CyCADA

- Cycle-consistent adversarial adaptation





(a) Test Image (b) Source Prediction (c) CyCADA Prediction (d) Ground Truth

Hoffman *et al.* ICML 2018

Thanks

mengunlong@matrixtime.com

Tech share plan

时间	人员	主题
2023-02-08	孟云龙	Domain adaptive object detection
2023-02-15	李煜→钱伍	OCR
2023-02-22	梁云飞	TBA
2023-03-01	孙耀状	TBA
2023-03-08	曹瑾	TBA
2023-03-15	钱伍→李煜	TBA
2023-03-23	熊鑫州	TBA
2023-03-30	周意龙	TBA
2023-04-05→2023-04-04	罗长志	TBA
2023-04-12	苏秀振	TBA
2023-04-19	……	