北京大學
PEKING UNIVERSITY

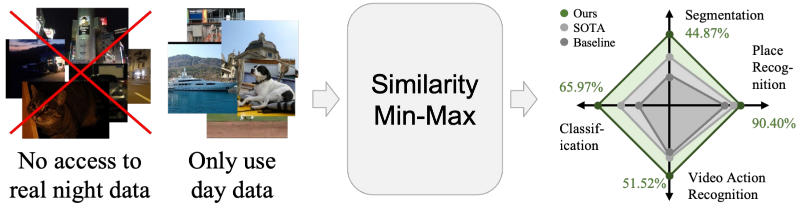
Similarity Min-Max: Zero-Shot Day-Night Domain Adaptation

Rundong Luo[1], Wenjing Wang[1], Wenhan Yang[2], Jiaying Liu[1]
[1]Peking University, [2]Peng Cheng Laboratory

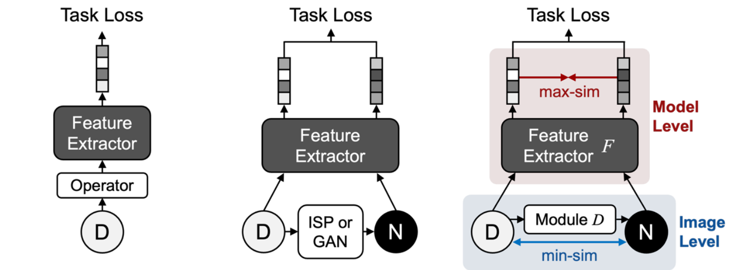
ICCV23 PARIS

Task Description

We tackle the task of Zero-Shot Day-Night Domain Adaptation (Zero-Shot Day-Night DA), *i.e.,* adapt deep models pre-trained on daytime data to nighttime domains, without any real nighttime data available.



Motivation



(a) Operator-based (b) Darkening-based (c) Our method

As shown above, existing methods on day-night DA can be generally categorized into:

- Operator-based (a): using manually defined operators ***at the model level*** to handle illumination variations, which are not adaptive to real complex scenarios.
- Darkening-based (b): transfer labeled daytime data to nighttime by GAN or reverse ISP ***at the image level***. However, GAN requires real nighttime data, while ISP is sensor-dependent.

Method

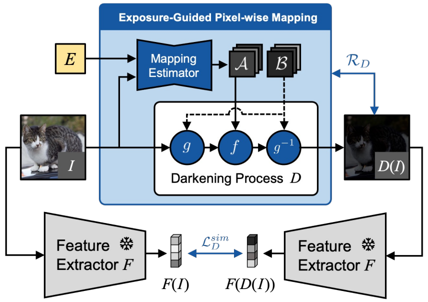
Unlike prior methods, we propose a similarity min-max framework that jointly considers model level and image level, formulated as:

$$\max_{\theta_F}\min_{\theta_D} \quad \mathrm{Sim}(F(I), F(D(I))),$$

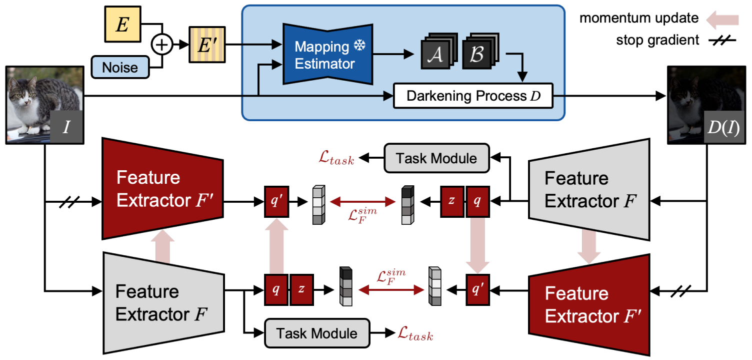
where D is the darkening module and F is the feature extractor. To prevent trivial solutions, we add regulations to (1):

$$\max_{\theta_F}\min_{\theta_D} \mathrm{Sim}(F(I), F(D(I)))+\mathcal{R}_D(\theta_D)-\mathcal{R}_F(\theta_F)$$

Right: Image-level translation. We design D to be a pixel-wise mapping controlled by two adjustment maps regularized by R_D, and L_D^{sim} is the cosine similarity.



Bottom: Model-level adaptation. We freeze D and train F by the BYOL non-contrastive loss (L_F^{sim}) and task-specific loss (L_{task})



Experiments

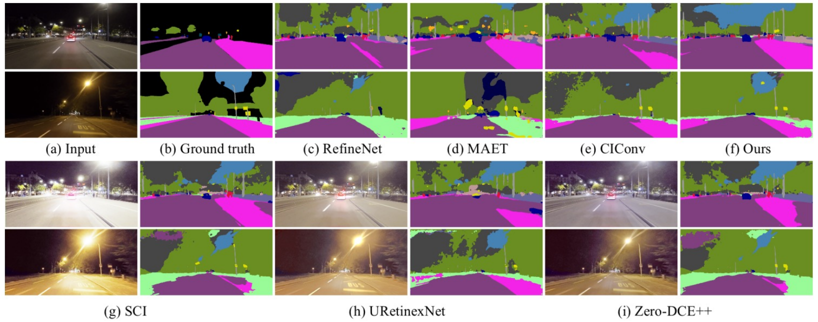
We evaluate our method on four nighttime downstream tasks: image classification, semantic segmentation, visual place recognition, and video action recognition.

I. Quantitative results for low-light image classification on CODaN

Method	Top-1 (%)
ResNet-18 [18]	53.32
Low-Light Enhancement	
EnlightenGAN [23]	56.68
LEDNet [63]	57.40
Zero-DCE++ [30]	57.96
RUAS [33]	58.36
SCI [34]	58.68
URetinexNet [56]	58.72
Domain Generalization	
MixStyle [62]	53.12
IRM [1]	54.52
AdaBN [31]	54.25
Zero-Shot Day-Night Domain Adaptation	
MAET† [8]	56.48
CIConv [29]	60.32
Ours	**65.87**

II. Quantitative results for nighttime semantic segmentation

Method	Nighttime Driving	Dark-Zurich
RefineNet [32]	34.3	30.6
Low-Light Enhancement		
EnlightenGAN [23]	25.2	24.9
Zero-DCE++ [30]	32.7	28.3
RUAS [33]	25.1	23.4
SCI [34]	28.6	25.7
URetinexNet [56]	28.1	24.0
LEDNet [63]	27.6	26.6
Domain Generalization		
AdaBN [31]	37.2	31.1
RobustNet [6]	33.0	34.5
SAN-SAW [38]	28.1	16.0
Zero-Shot Day-Night Domain Adaptation		
MAET [8]	28.1	26.4
CIConv [29]	41.2	34.5
Ours	**44.9**	**40.2**



(a) Input (b) Ground truth (c) RefineNet (d) MAET (e) CIConv (f) Ours
(g) SCI (h) URetinexNet (i) Zero-DCE++

III. Qualitative results for nighttime semantic segmentation

Method	Top-1 (%)
I3D [3]	47.02
Low-Light Video Enhancement	
StableLLVE [59]	45.08
SMOID [22]	47.27
SGZ [61]	46.42
Domain Generalization & Zero-Shot Day-Night Domain Adaptation	
AdaBN [31]	46.17
Ours	**51.52**

Method	mAP (%)
Zero-Shot Day-Night Domain Adaptation	
EdgeMAC [42]	75.9
U-Net jointly [21]	79.8
GeM [43]	85.0
CIConv-GeM [29]	88.3
Ours-GeM	**90.4**
Day-Night Domain Adaptation (night images are available for training)	
U-Net jointly [21]	86.5
EdgeMAC + CLAHE [21]	90.5
EdgeMAC + U-Net jointly [21]	90.0

IV & V. Quantitative results for low-light video action recognition on ARID, and visual place recognition on Tokyo 24/7.