CVPR'21 Workshop "Vision for All Seasons: Adverse Weather and Lighting Conditions"

Workshop at CVPR21 <link>

1

Keynote - Prof. Wolfram Burgard: Exploiting Knowledge from Multiple Modalities for Robust Perception

1.A

Idea of Using Thermal Cameras for Transfering Capabilities from Day Time to Night Time

Exploiting Knowledge from Multiple Modalities for Robust Perception

Wolfram Burgard

Joint work with: Jannik Zürn, Johan Vertens, Kshitij Sirohi, Rohit Mohan, Abhinav Valada …

AiS Autonomous Intelligent Systems

Motivation

- We want to minimizie labeling efforts
  - New tasks
  - Domain transfer

- In this talk: **using multi-modal setups**

Robust Semantic Segmentation by Domain Adaption

HeatNet: Bridging the Day-Night Domain Gap in Semantic Segmentation with Thermal Images

Johan Vertens*, Jannik Zürn*, and Wolfram Burgard

Abstract— **The majority of learning-based semantic segmentation methods are optimized for daytime scenarios and favorable lighting conditions. Real-world driving scenarios, however, entail adverse environmental conditions such as nighttime illumination or glare which remain a challenge for existing approaches. In this work, we propose a multimodal semantic segmentation model that can be applied during daytime and nighttime. To this end, besides RGB images, we leverage thermal images, making our network significantly more robust. We avoid the expensive annotation of nighttime images by leveraging an existing daytime RGB-dataset and propose a teacher-student training approach that transfers the dataset's knowledge to the nighttime domain. We further employ a domain adaptation method to align the learned feature spaces across the domains and propose a novel two-stage training scheme. Furthermore, due to a lack of thermal data for autonomous driving, we present a new dataset comprising over 20,000 time-synchronized and aligned RGB-thermal image pairs. In this context, we also present a novel target-less calibration method that allows for automatic robust extrinsic and intrinsic thermal camera calibration. Among others, we employ our new dataset to show state-of-the-art results for nighttime semantic segmentation.**

Fig. 1. Our multimodal segmentation network leverages both nighttime and daytime images. We transfer relevant knowledge from a large-scale unimodal daytime dataset for semantic segmentation with a teacher model to our multimodal HeatNet and simultaneously adapt our model to the nighttime domain by unsupervised domain adaptation.

I. INTRODUCTION

Robust and accurate semantic segmentation of urban scenes is one of the enabling technologies for autonomous driving in complex and cluttered driving scenarios. Recent years have shown great progress in RGB image segmentation for autonomous driving [36], [5], which were predominantly demonstrated in favorable daytime illumination conditions. While the reported results demonstrate high accuracies on benchmark datasets [5], [18], these models tend to generalize poorly to adverse weather conditions and low illumination levels present at nighttime. This constraint becomes especially apparent in rural areas where artificial lighting is weak or scarce. In autonomous driving, to ensure safety and situation awareness, robust perception in these conditions is a vital prerequisite.

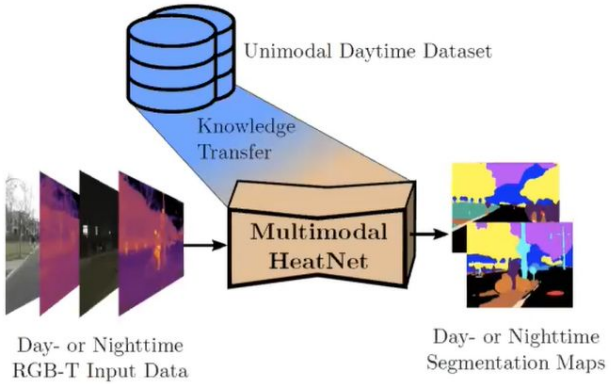
Transfer learning and domain adaptation approaches aim at narrowing the domain gap between a source domain, where supervised learning from labelled data is possible, to a target domain, where labelled data is either sparse or not available. Such approaches, as demonstrated in [28] or [35], allow to adapt a given segmentation model to a different domain. These approaches, however, do not leverage a complementary modality such as thermal infrared images that can contain more relevant information to solve a given task is certain environmental conditions than a single modality would provide.

In order to perform similarly well in challenging illumination conditions, it is beneficial for autonomous vehicles to leverage modalities complementary to RGB images [29], [30]. Encouraged by prior work in thermal image processing for object detection [31], object tracking [14], and semantic segmentation [9], [25], we investigate leveraging thermal images for nighttime semantic segmentation of urban scenes. Thermal images contain accurate thermal radiation measurements with a high spatial density. Furthermore, thermal radiation is much less influenced by sunlight illumination changes and is less sensitive to adversary conditions. Existing RGB-thermal datasets for semantic image segmentation such as [9] are not as large-scale as their RGB-only counterparts. Thus, models trained on such datasets generalize poorly to challenging real-world scenarios.

In contrast to previous works, we utilize a semantic segmentation network for RGB daytime images as a teacher model to provide labels for the RGB daytime images in our dataset. We project the thermal images into the viewpoint of the RGB camera images using extrinsic and intrinsic camera parameters that we determine using our novel target-less camera calibration approach. Afterwards, we can reuse labels from this teacher model to train a multimodal semantic segmentation network on our daytime RGB-thermal image pairs. In order to encourage day-night invariant segmentation of scenes, we simultaneously train a feature discriminator

*These authors contributed equally. All authors are with the University of Freiburg, Germany. Wolfram Burgard is also with the Toyota Research Institute, Los Altos, USA. Corresponding author: vertens@informatik.uni-freiburg.de

HeatNet: Bridging the Day-Night Domain Gap in Semantic Segmentation with Thermal Images. Johan Vertens, Jannik Zürn and Wolfram Burgard, IROS 2020



INSIGHTS

- IRIS Published.

- Focused on - how we can actually leverage - the thermal cameras - for better semantic segmentation, in particular, - for situations - where normal RGB cameras face difficulty - like the low light vision.

- Labelling nighttime images is very difficult for semantic segmentation task.

Motivation

Semantic segmentation prediction during nighttime using a conventional CNN trained on publicly available datasets

Labeling night-time images is extremely painful!

INSIGHTS

- The idea of this work is - to leverage daytime semantic segmentation AND another modality - to bootstrap semantic segmentation for nighttime.

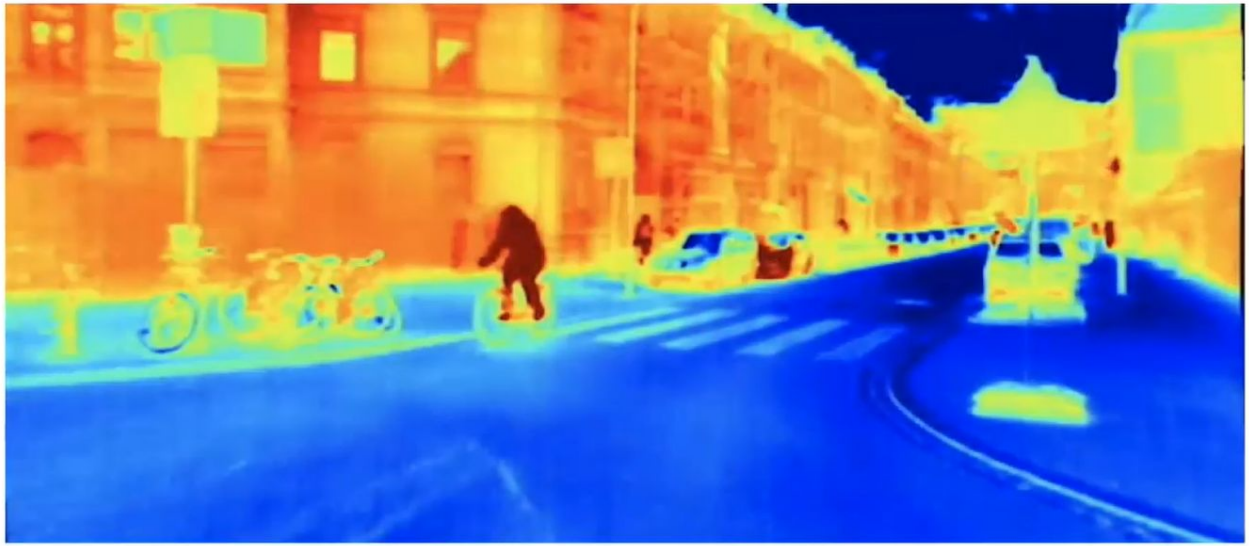
- NEXT SLIDE - A Typical Night Scene

- AND THEN - Overlaying with Thermal Infrared Images.

Motivation

Thermal infrared images exhibit small domain gap between day- and nighttime

Motivation



Thermal infrared images exhibit small domain gap between day- and nighttime

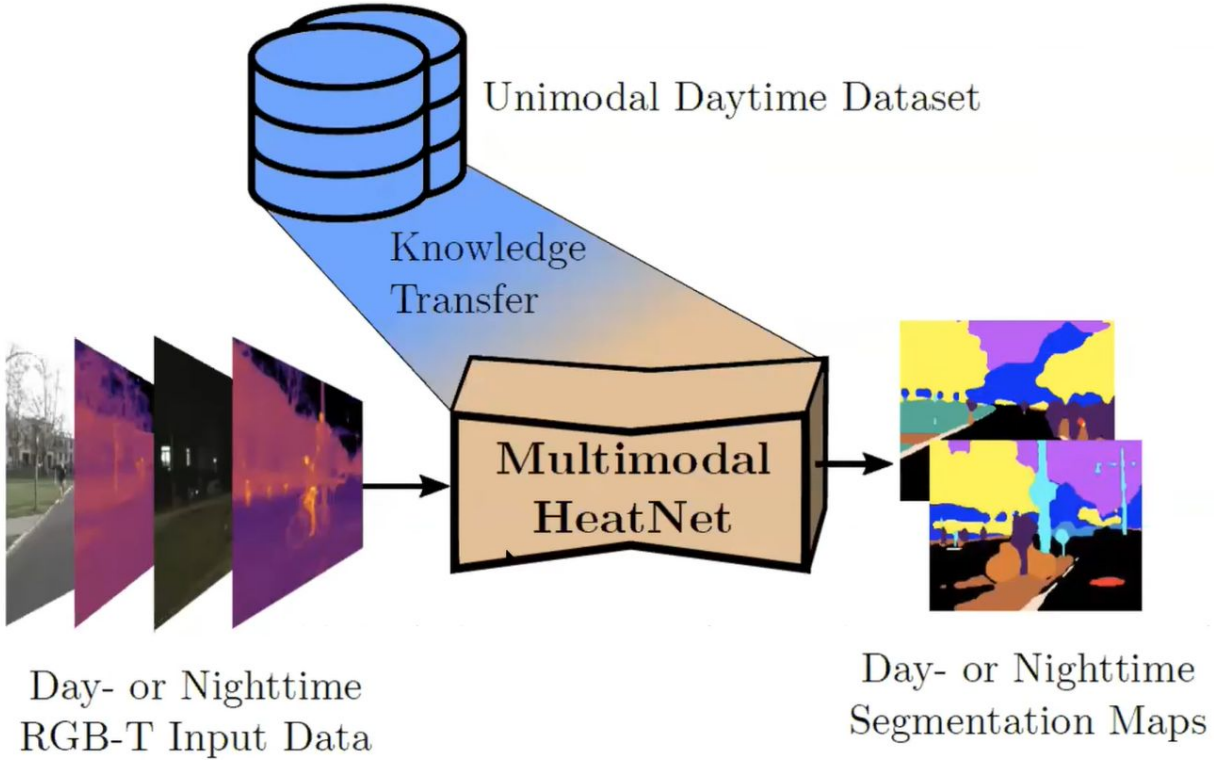
INSIGHTS

- THERMAL INFRARED IMAGES exhibit small domain gap between day and night time.

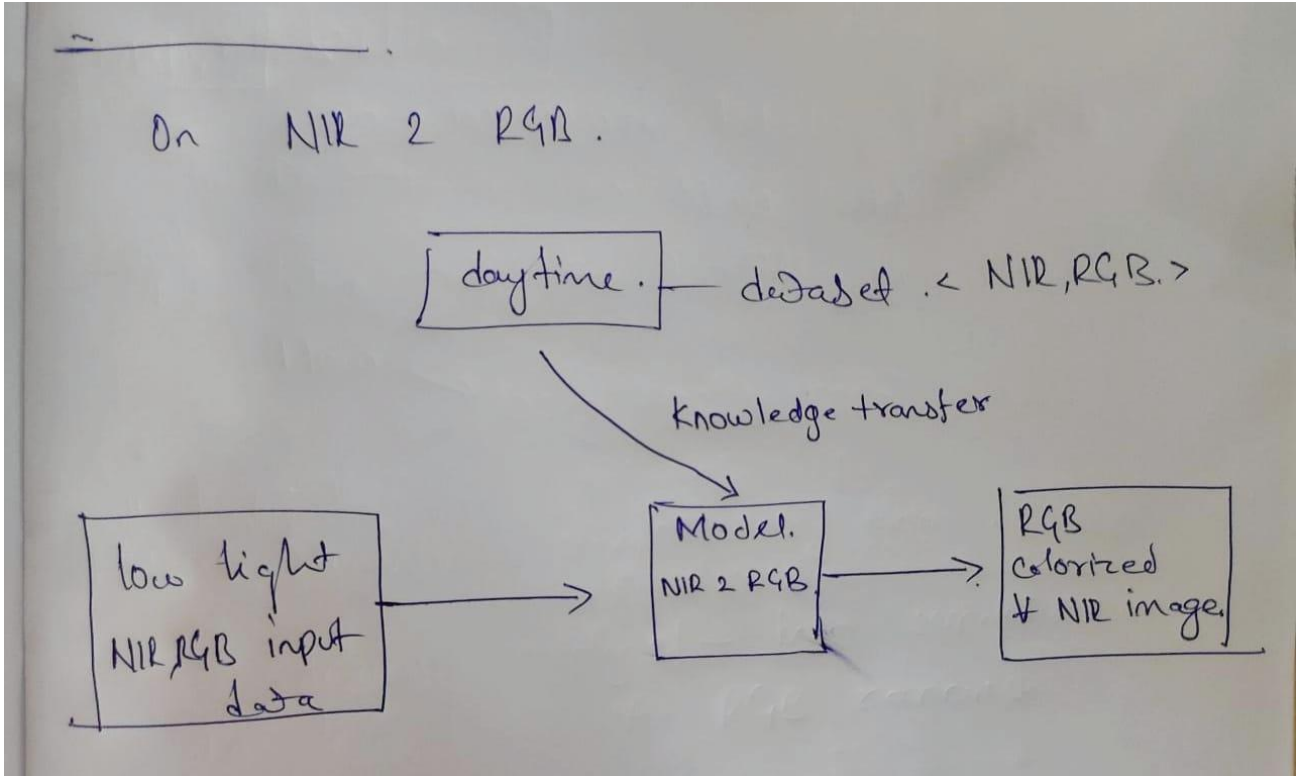
- **NEXT SLIDE :** Shows the Approach - To Train a Multimodal HeatMap - That takes RGB - Thermal Images - AND - Creates Semantic Segmentation Map - For Day & Night time Images.

- At the same time, - have to leverage all the knowledge we have - for Daytime Semantic Segmentation.

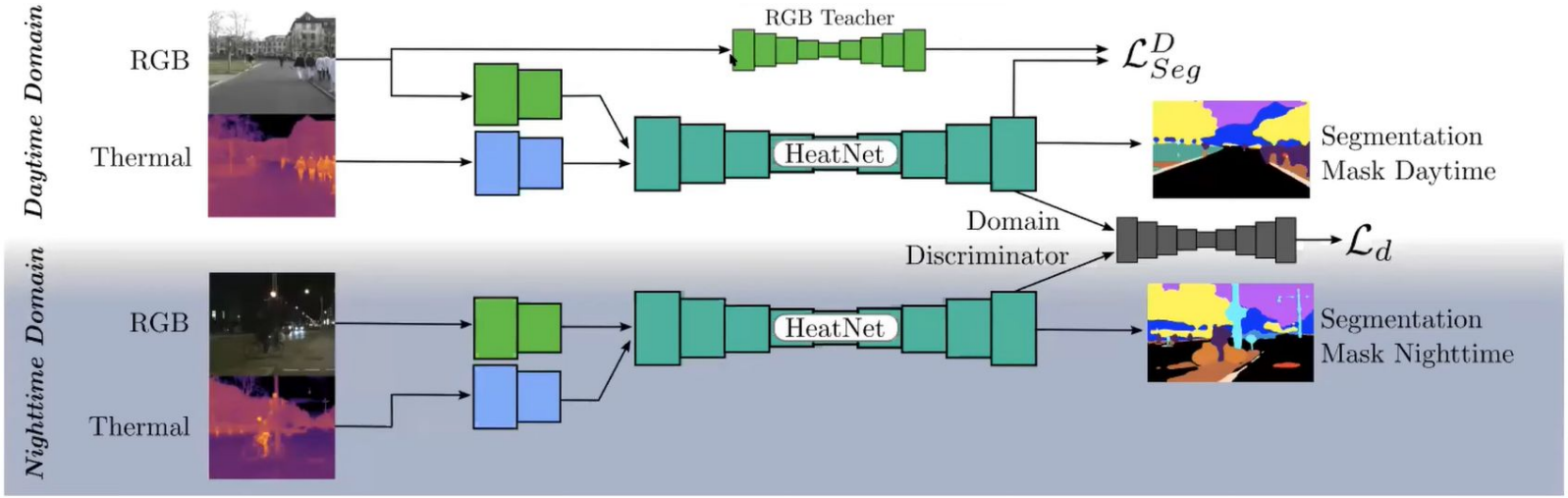
Approach


Unimodal Daytime Dataset
Knowledge Transfer
Multimodal HeatNet
Day- or Nighttime RGB-T Input Data
Day- or Nighttime Segmentation Maps

INSIGHTS


On NIR 2 RGB.
daytime.
dataset .< NIR, RGB. >
knowledge transfer
low light
NIR, RGB input
data
Model.
NIR 2 RGB
RGB
Colorized
of NIR image.

Approach



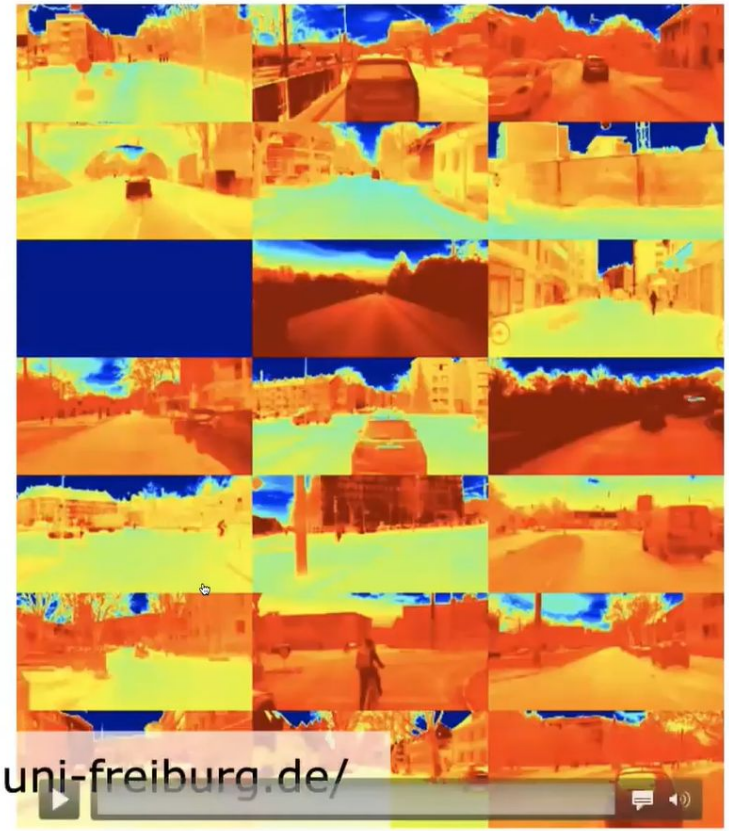
$$\mathcal{L}_{p_1} = \mathcal{L}_s^D + \lambda[0 - C(S_N)]^2,$$

$$\mathcal{L}_{p_2} = \frac{1}{HW} \sum_{h,w} \begin{cases} [0 - C(S_X)]^2, & \text{if } X = D \\ [1 - C(S_X)]^2, & \text{if } X = N \end{cases}$$

Dataset

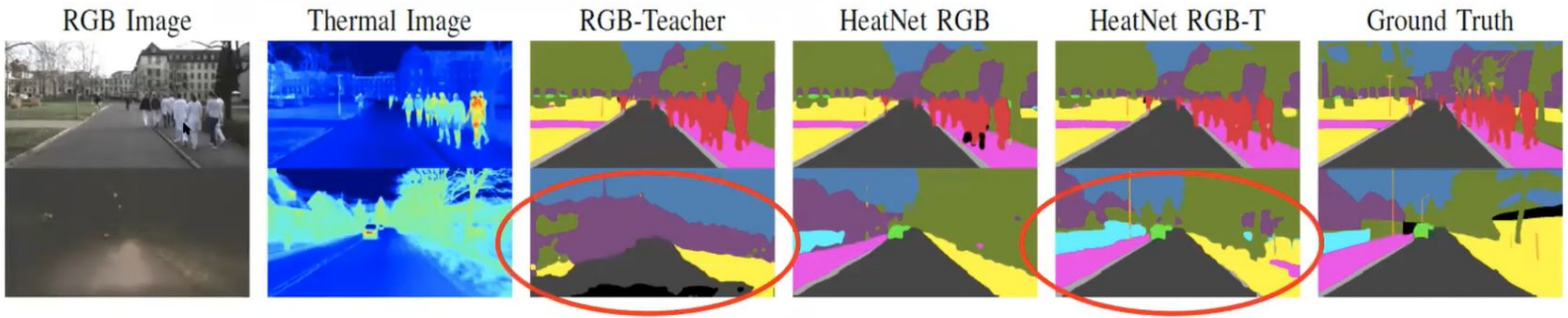
- Freiburg Thermal Dataset
- Five day- and three nighttime collections
- Multiple seasons
- 12,000 daytime images
- 8,000 nighttime images
- GPS and IMU data
- LiDAR point clouds
- 64 evaluation images

Dataset

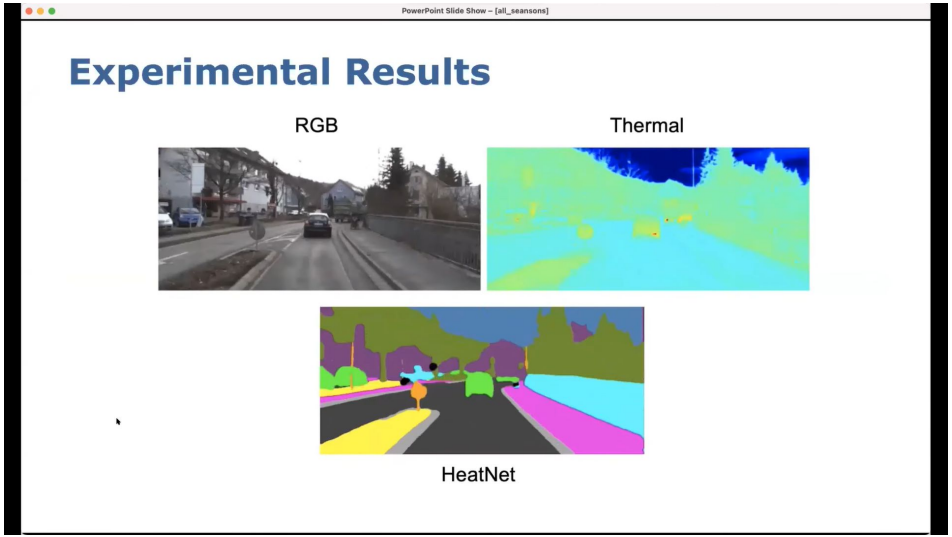


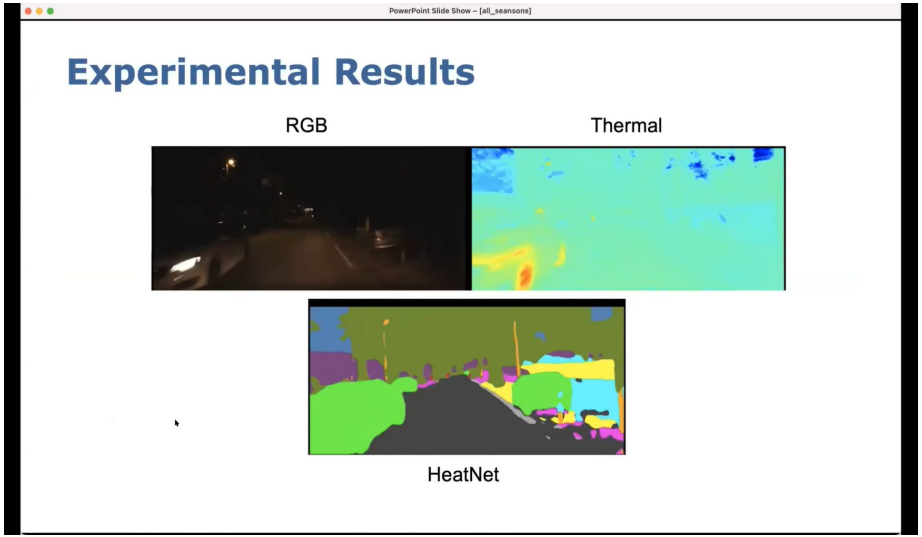
http://thermal.cs.uni-freiburg.de/

Experimental Results



DAYTIME AND NIGHT TIME





INSIGHTS

Instead of Daytime and Nighttime, we can do for without rain and in the rain.

Quantitative Evaluation

Train On	Test On	Model	RGB	T	Road	Sidewalk	Building	Curb	Fence	Pole	Vegetation	Terrain	Sky	Person	Car	Bicycle	Mean
MF	MF	MFNet [9]	✓	✓	-	-	-	-	-	-	-	-	-	58.9	65.9	42.9	55.9
		RTFNet-50 [25]	✓	✓	-	-	-	-	-	-	-	-	-	**67.8**	**86.3**	**58.2**	**70.7**
		HeatNet	✓	✓	-	-	-	-	-	-	-	-	-	56.4	68.8	33.9	53.0
	FR-T Day/Night	MFNet [9]	✓	✓	-	-	-	-	-	-	-	-	-	42.8	27.0	24.5	31.4
		RTFNet-50 [25]	✓	✓	-	-	-	-	-	-	-	-	-	**63.2**	61.5	51.3	58.6
		HeatNet	✓	✓	86.7	57.5	67.7	46.4	41.5	43.8	57.9	44.1	63.7	63.1	**85.6**	**58.2**	59.7
FR-T	MF	HeatNet	✓	✓	-	-	-	-	-	-	-	-	-	51.6	61.8	30.2	47.9
(Vistas)	FR-T Day	RGB Teacher	✓	✗	**89.7**	**67.0**	73.8	56.9	48.8	53.8	73.8	62.8	84.3	72.0	90.1	60.4	69.4
FR-T		HeatNet	✓	✓	89.4	65.6	**74.8**	**59.7**	**52.9**	**54.3**	**74.1**	**65.1**	**84.5**	**74.0**	**91.2**	**64.1**	**70.8**
FR-T	FR-T Night	Thermal Teacher	✗	✓	84.9	60.5	**65.5**	43.1	31.8	38.1	51.8	40.1	72.6	49.6	**87.1**	**56.9**	57.0
(Vistas)		RGB Teacher	✓	✗	76.3	22.6	53.4	10.8	14.1	31.6	10.4	13.5	47.7	28.0	74.3	45.2	35.7
FR-T		HeatNet	✓	✓	**86.4**	**60.9**	65.4	**45.5**	**35.5**	**42.0**	**52.5**	**52.3**	**73.9**	**54.9**	85.7	53.3	**59.0**
FR-T	FR-T Day/Night	HeatNet	✓	✓	**87.9**	**63.3**	**70.1**	**52.6**	**44.2**	**48.2**	**63.3**	**58.9**	**79.2**	**64.5**	**88.5**	**58.7**	**64.9**
FR-T		HeatNet RGB-only	✓	✗	82.7	56.0	66.0	45.3	34.0	37.8	58.4	49.5	71.0	54.4	84.2	57.4	58.0
(Vistas)	BDD Night [34]	RGB Teacher	✓	✗	68.8	21.5	32.9	-	0.0	12.3	11.5	6.6	**27.2**	24.5	40.4	-	24.6
FR-T		HeatNet RGB-only	✓	✗	**87.1**	**40.0**	**50.2**	-	**25.9**	**22.9**	**12.8**	**8.5**	25.0	**27.4**	**68.3**	-	**36.8**

2

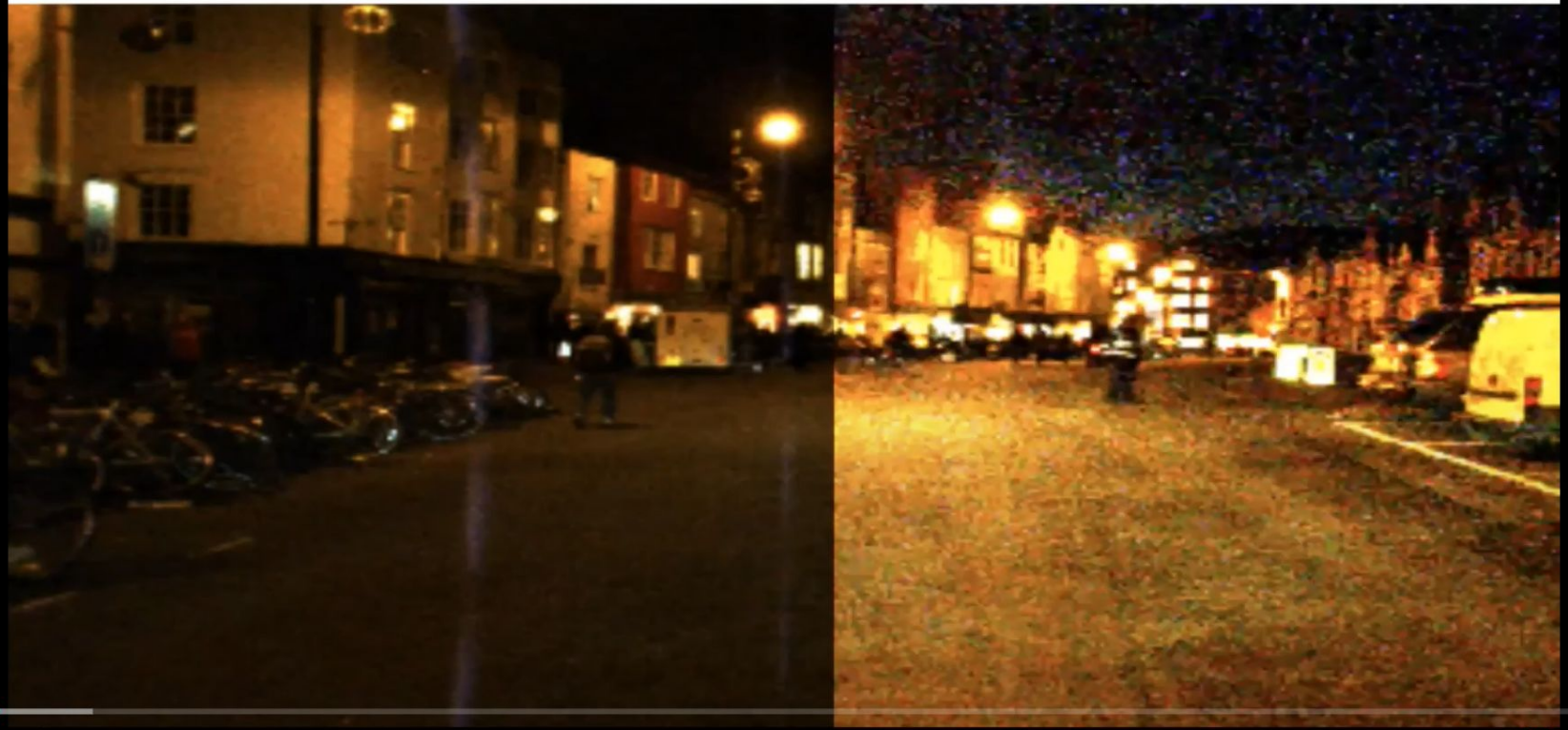
NIGHT IMAGES: WHAT INFORMATION WE CAN EXTRACT & ENHANCE?

INSIGHTS

- Night Images Problem - LOW LIGHT
- **WHAT IF** we multiply the pixels with some constant values?
    - NOISY IMAGES
    - INTERESTINGLY – NOISE CAN BE DIFFERENTIATED FROM STRUCTURES

Keynote - Prof. Robby T. Tan: Night Images: What information we can extract and enhance?

Night Image Problem: Low Light



1:00 / 35:37

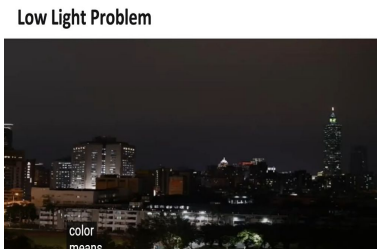
INSIGHTS

- Night Images Problem: Glow/ Glare/ Floodlight.

Night Image Problem: Uneven Light Distribution

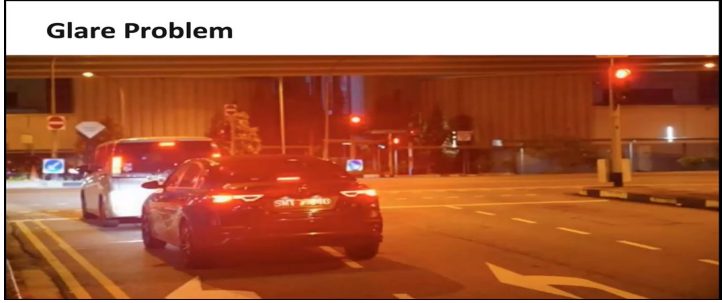
Night Image Problem: Light Colors

INSIGHTS - **NIGHT IMG ENHANCEMENT**



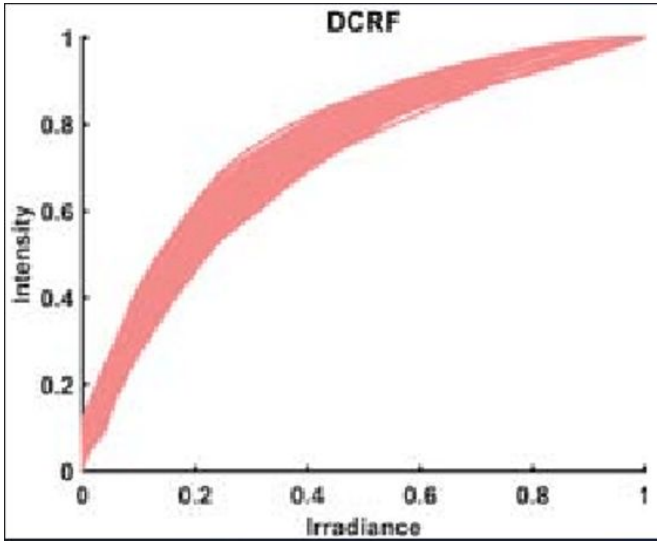
- Camera response function (CRF) relates scene irradiance to image intensities.
    - Estimation of CRF is a fundamental and necessary step in many computer vision applications such as the generation of high dynamic image, bidirectional reflectance distribution function (BRDF) estimation etc.
    - CRF is non-linear.
- Irradiance vs Radiance
    - In terms of explanation, it can be said that Radiation is the number of photons that are being emitted by a single source. Irradiation, on the other hand, is one where the radiation is falling on the surface is being calculated.
- **IDEA : BOOST the intensity, BUT have to keep the color => ESTIMATE THE CRF.**

INSIGHTS - **NIGHT IMG ENHANCEMENT**



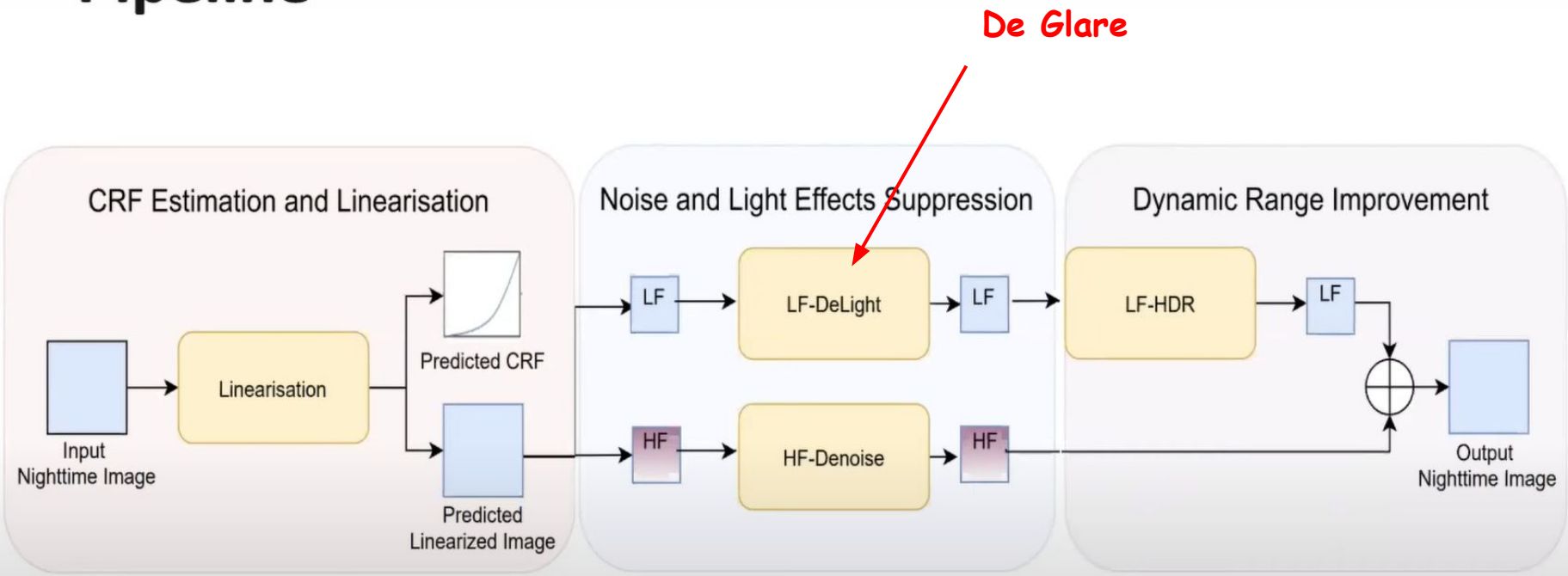
- NOT JUST BOOST THE IMAGE -
- HAVE TO SUPPRESS LIGHT IN IMGS LIKE THIS
- HAVE TO REMOVE GLARE HERE

CRF of Camera



Problem: Boosting low light regions, at the same time suppress glare and noise

Pipeline



Published in CVPR'21

HDR

- HDR stands for high dynamic range.
- Put simply, *it's the range of light and dark tones in your photos*.
- The human eye has a very high dynamic range — which is why we can see details in both shadows and highlights.

Supervised Training: CRF

- CRF Loss:

$$\mathcal{L}_{\text{mse}} = \|\hat{\mathbf{g}} - \mathbf{g}^{gt}\|_2 \qquad \hat{\mathbf{g}} = \mathbf{g_0} + \sum_{i=1}^{11} \mathbf{h}_i \mathbf{c}_i$$

- Image-Linearization Loss:

$$\mathcal{L}_{\text{lin}} = \|\hat{\mathbf{g}}(\mathbf{Z}') - \mathbf{g}^{gt}(\mathbf{Z}')\|_1$$

Supervised Training: HDR

- For decomposing the linearized image to low and high frequency layers, we employ: 'Fast end-to-end trainable guided filter', CVPR'18.
- HDR Loss:

$$\mathcal{L}_{\text{HDR}} = \left\| \frac{\log(1+\mu\hat{\mathbf{X}})}{\log(1+\mu)} - \frac{\log(1+\mu\mathbf{X}^{gt})}{\log(1+\mu)} \right\|_1$$

$$\hat{\mathbf{X}} = \frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{HrLF}_{\mathbf{X}k} + \mathbf{DnHF}_{\mathbf{X}k}\right)$$

Loss function is based on μ-law (used for tone mapping)

X is an HDR image; K is the number of filters

Unsupervised Test-Time Training: CRF

- CRF-Monotonicity Loss:

$$\mathcal{L}_{\text{mon}} = \sum_{t=0}^{1} H\left(-\frac{\partial \hat{\mathbf{g}}(t)}{\partial t}\right)$$

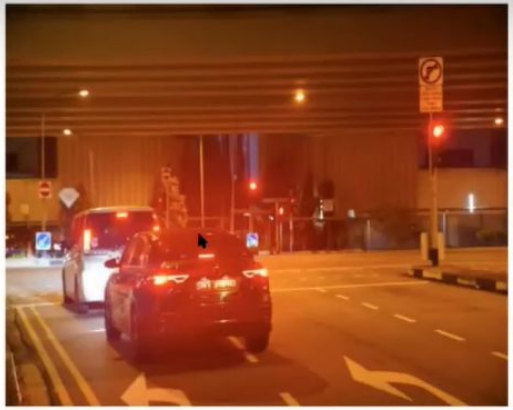
- CRF-Pixel-Linearization Loss:

$$\mathcal{L}_{es} = \sum_{i=1}^{S}\left(\frac{|\mathbf{n}_{\mathbf{Y}es}^{\min} - \mathbf{n}_{\mathbf{Y}es}^{\max}| \times |\mathbf{n}_{\mathbf{Y}es}^{\min} - \mathbf{n}_{\mathbf{Y}es}^{\mathrm{i}}|}{|\mathbf{n}_{\mathbf{Y}es}^{\min} - \mathbf{n}_{\mathbf{Y}es}^{\max}|}\right),$$

$$\mathcal{L}_{\text{distlin}} = \sum_{e=1}^{E}\left(\sum_{s=1}^{S}(\mathcal{L}_{es})\right),$$

E = #patches; Each patch has a size of S x S

Light Effect Suppression:



(a) Input

(b) LF Map

(d) LF Map (w/o l. e.)

(e) Output (w/o l. e.)

Results:

Input

Our Method

Ground-Truth

HDRCNN [2]

SingleHDR [10]

DrTMO [3]

Results:

Input

Our Method

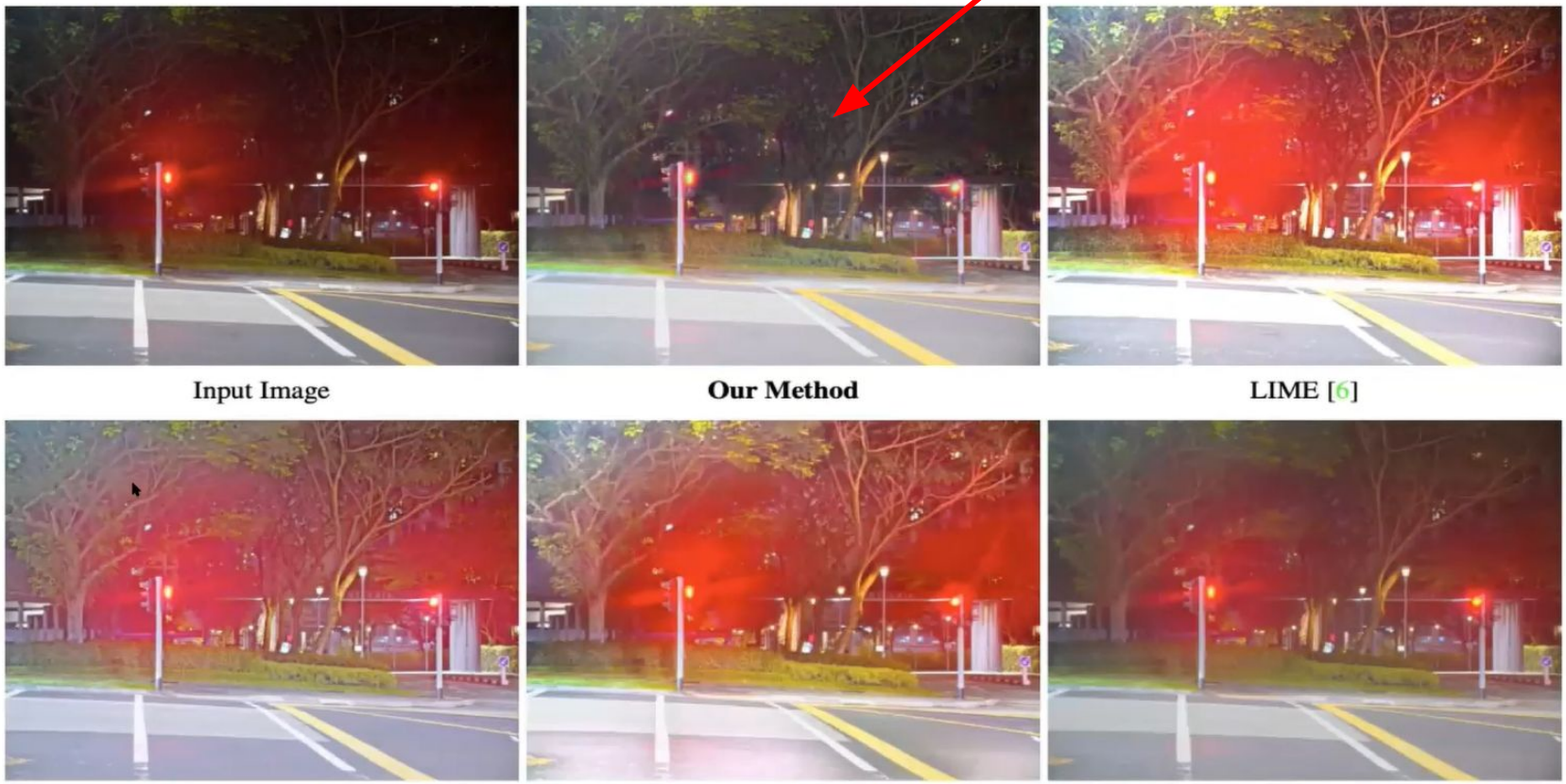
LIME [6]

ZeroDCE [5]

EnlightenGAN [7]

SingleHDR [10]

Results:


GLARE
Input Image
Our Method
LIME [6]
ZeroDCE [5]
EnlightenGAN [7]
SingleHDR [10]

Publications:

- Nighttime Haze Removal with Glow and Multiple Light Colors ICCV'15
- Nighttime Defogging Using High-Low Frequency Decomposition and Grayscale-Color Networks, ECCV'20
- Single-Image Camera Response Function Using Prediction Consistency and Gradual Refinement, ACCV'20
- Nighttime Stereo Depth Estimation using Joint Translation-Stereo Learning: Light Effects and Uninformative Regions, 3DV'20
- Nighttime Visibility Enhancement by Increasing the Dynamic Range and Suppression of Light Effects, CVPR'21.

NIGHTTIME HAZE REMOVAL WITH GLOW AND MULTIPLE LIGHT COLORS

- 2015 IEEE International Conference on Computer Vision.
- This paper focuses on dehazing nighttime images.

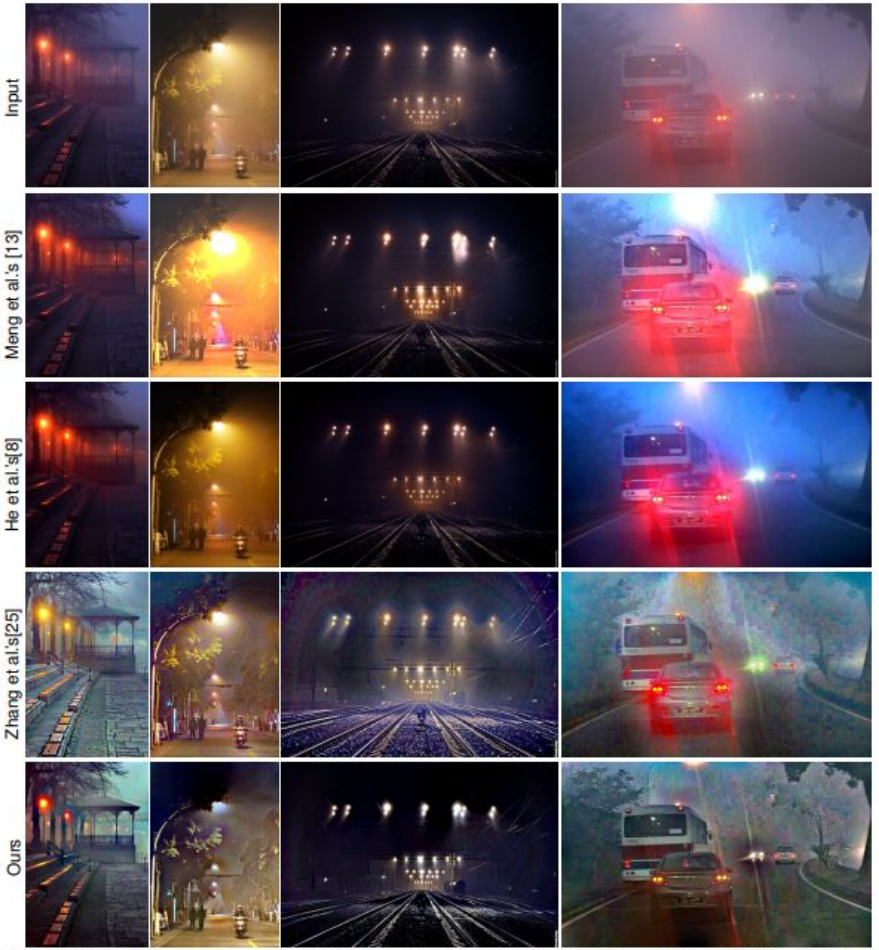


Figure 10. The qualitative comparisons of Meng et al.'s method [13], He et al.'s method [8], Zhang et al.'s method [25], and ours using various nighttime images.

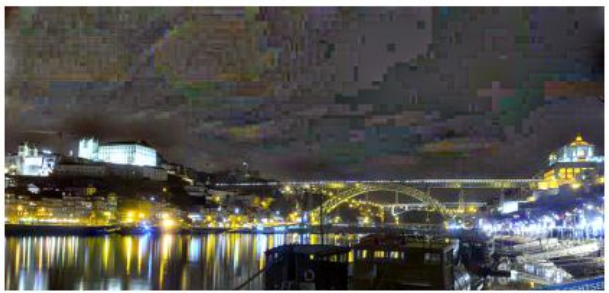
NIGHTTIME DEFOGGING USING HIGH-LOW FREQUENCY DECOMPOSITION AND GRAYSCALE-COLOR NETWORKS

- We address the problem of nighttime defogging from a single image by introducing a framework consisting of two modules: grayscale and color modules.

European Conference on Computer Vision
ECVV 2020

Nighttime foggy Image

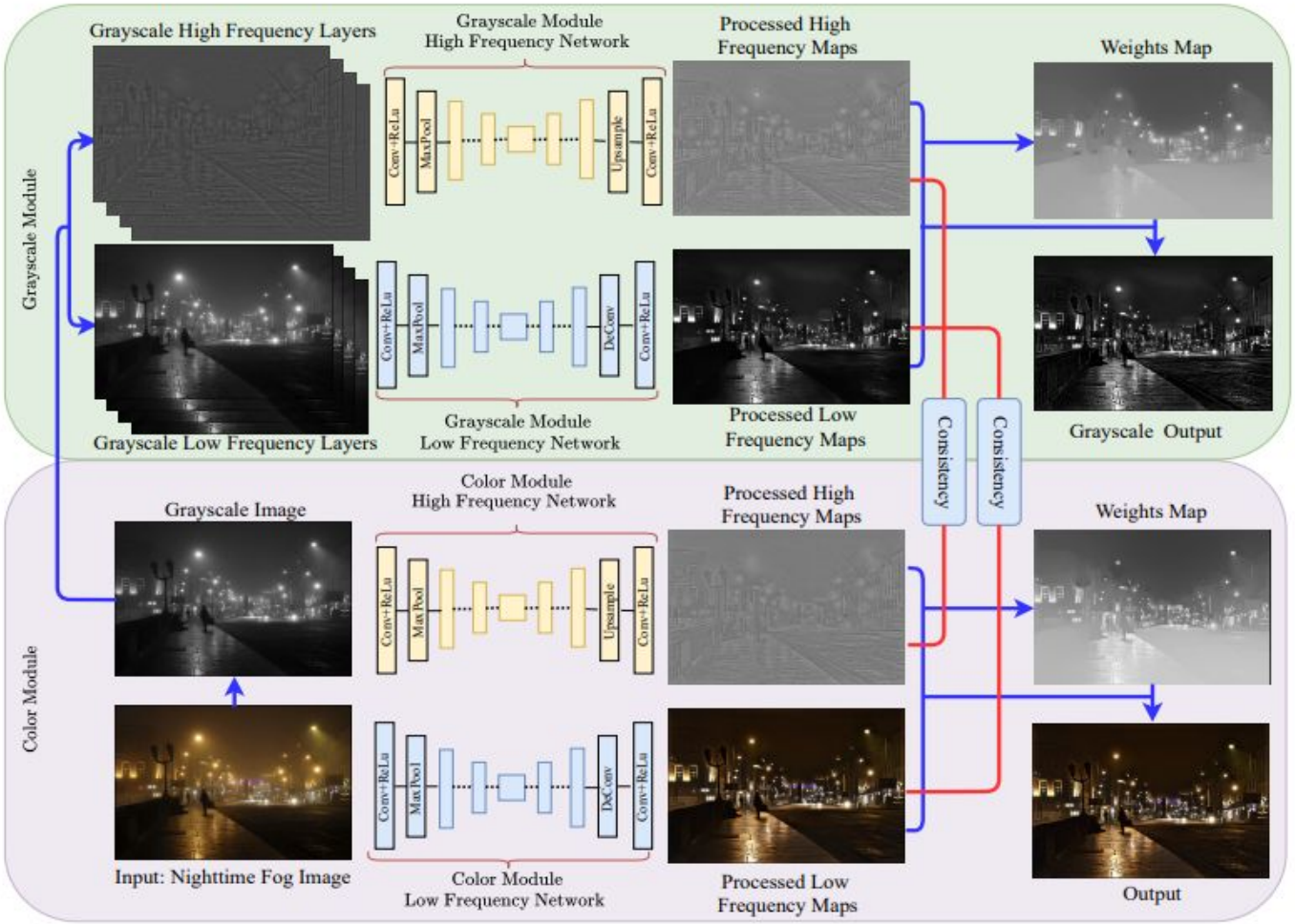
Our Result



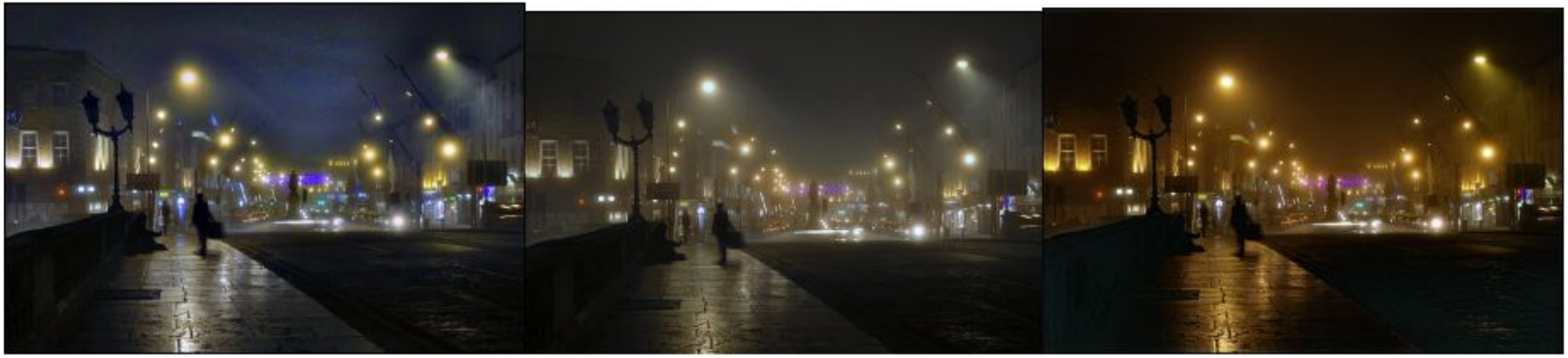
Li et al. [17]

Zhang et al. [26]

MODEL PIPELINE



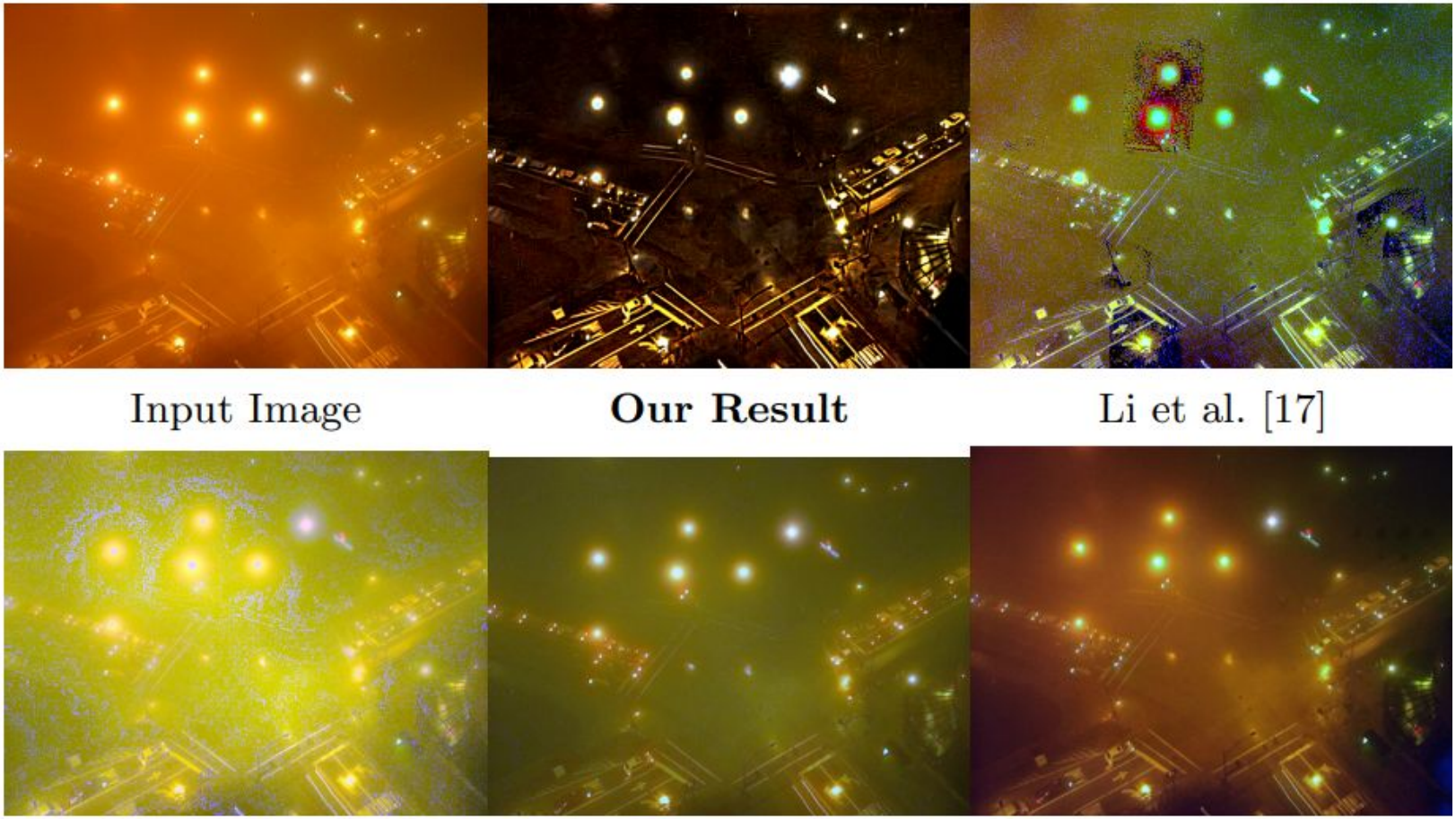
Input Image | **Our Result** | Li et al. [17]



Zhang et al. [26] | Ancuti et al. [1] | EPDN [21]

Input Image | **Our Result** | Li et al. [17]

Zhang et al. [26] | Ancuti et al. [1] | EPDN [21]



Input Image | **Our Result** | Li et al. [17]

Zhang et al. [26] | Ancuti et al. [1] | EPDN [21]

<LINK2PAPER>

6 Conclusion

We have introduced a learning-based nighttime defogging method. To our knowledge, this is the first time, a deep learning-based method is dedicated to handle nighttime defogging problem. To achieve our goal, we design grayscale and color modules, which rely mainly on the high/low frequency layers to enhance textures and at the same time suppress glow, fog and noise. Due to the lack of paired real ground-truths, our training process employs both paired synthetic data and unpaired real data. For this, we introduce new consistency losses between the outputs of the grayscale and color modules. Experimental results and evaluations, both quantitative and qualitative, show the effectiveness of our method.