OMNIVISION™

August 2022

Algorithm Tech Sharing

Virtual & Physical CVPR'22 Conference

OVT Singapore Algorithm Computer Vision Team

OVT Computer Vision Algorithm Team

OUR GOAL: deliver optimized, light-weight Computer Vision Solutions to our customers, by integrating our technologies into our sensor and/or chipset

Technologies

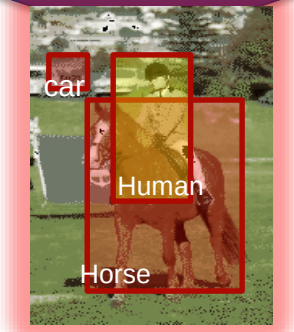
Sensor Business

Machine Learning

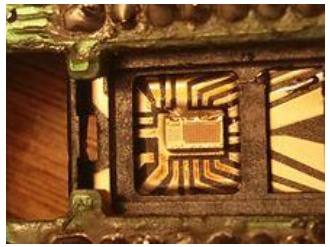
Video Analytic

3D Computer Vision

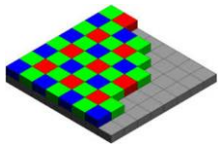
Object Recognition



Face Analysis

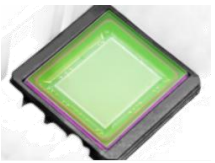


Smart (AI) Sensor



CIS

+

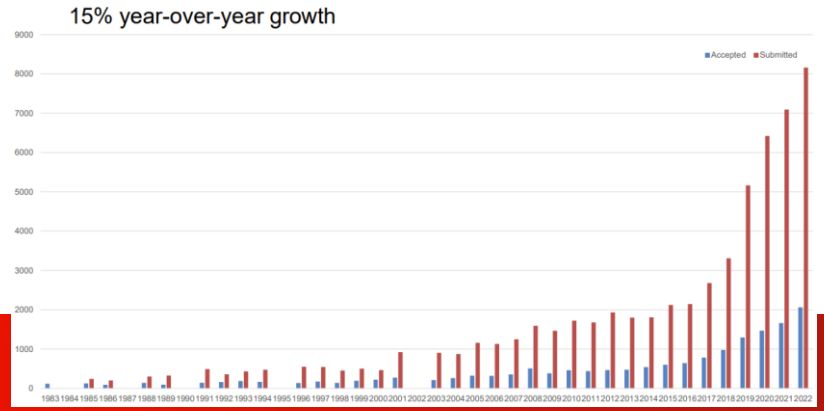


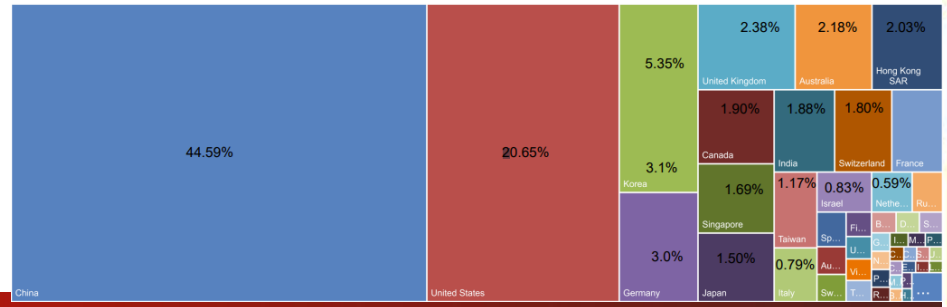
EVS

*Over the past **9 years**, computer vision team has built up **strong competencies** in various CV areas by using Machine Learning*

General Info about CVPR'22

- 74 workshops, 29 tutorials, 17 demos
- 9,800+ (5,500+ in person, 4,300+ virtual) attendees
- 70 sponsors (~$2M in sponsorship & exhibition)
- Main conference
  - **2,064** accepted papers from total **8,161** valid submissions
    - 25.3% acceptance rate, **244** oral papers among all accepted papers
  - 3 keynote speech a) understanding visual appearance from micron to global scale; b) toward integrative AI with CV; c) learning to see the human way
  - 1 panel session – Embodied Computer Vision





Best Papers Review

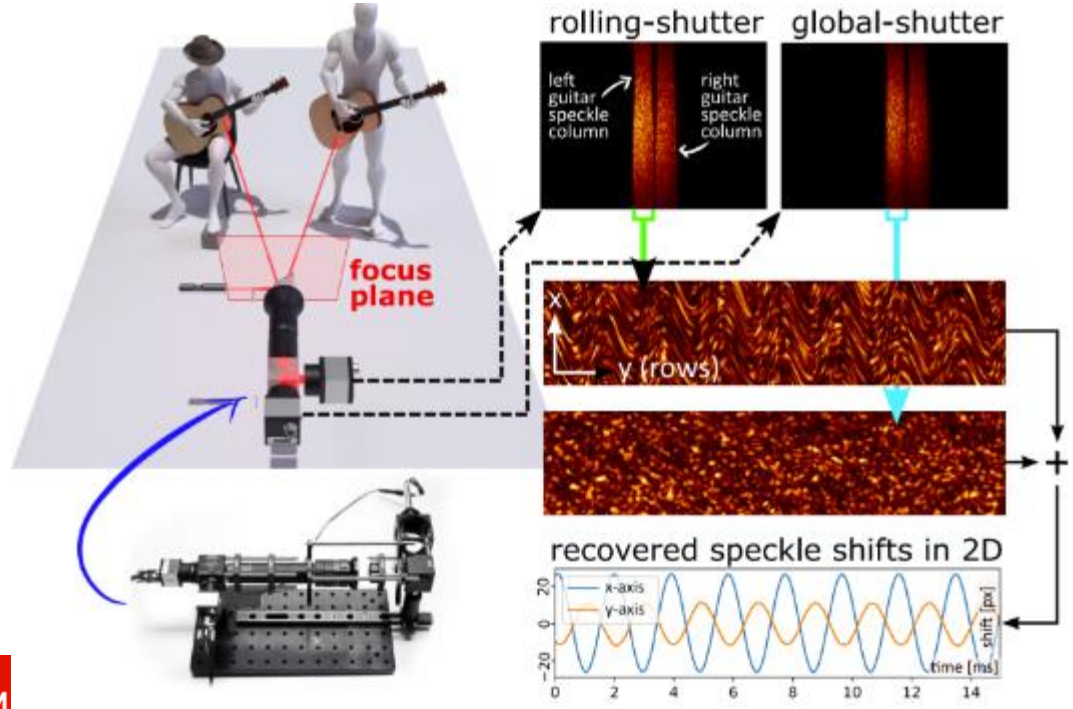
Zhongyang HUANG

CVPR'22 Best Papers

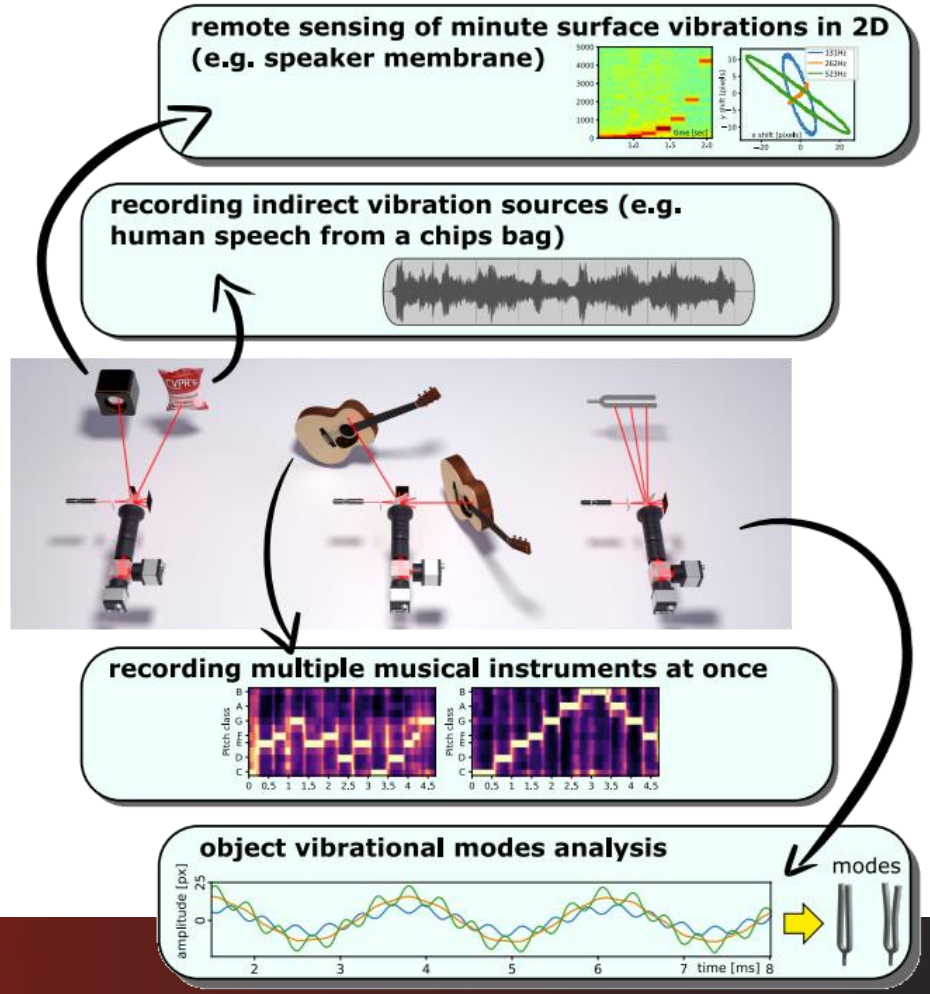
- *Dual-Shutter Optical Vibration Sensing*, CMU (Best Paper Honorable Mention)

"Seeing" sound in a novel way

- A novel **bandwidth-efficient 2D** optical vibration sensing approach
- Allows sensing **low-amplitude high-frequency** surface vibrations
- For **multiple scene points at once**
- **High-speed** sensing (63kHz), using **"slow" cameras** (e.g. 130Hz)
- Able to sense **non-static objects** (e.g. guitar played by a musician)

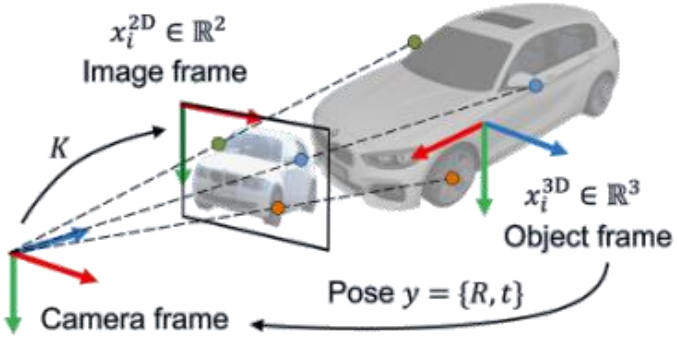


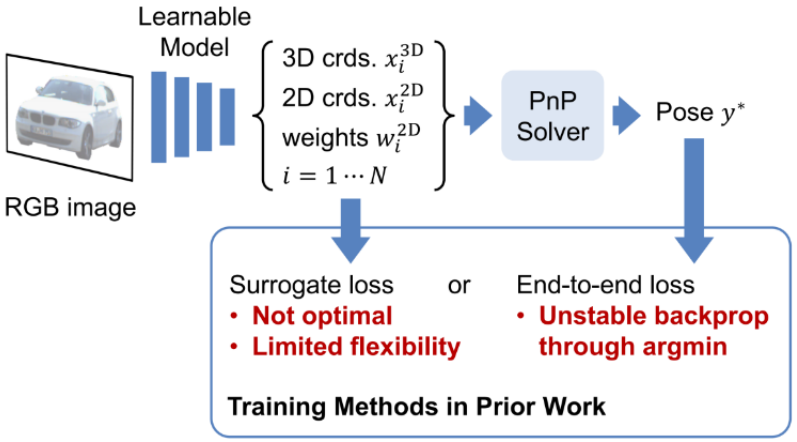
Applications



CVPR'22 Best Student Papers

- *EPro-PnP: Generalized End-to-End Probabilistic Perspective-n-Points for Monocular Object Pose Estimation,* Tongji Uni., Alibaba
  - **Object Pose Estimation** by solving the Perspective-n-Point problem

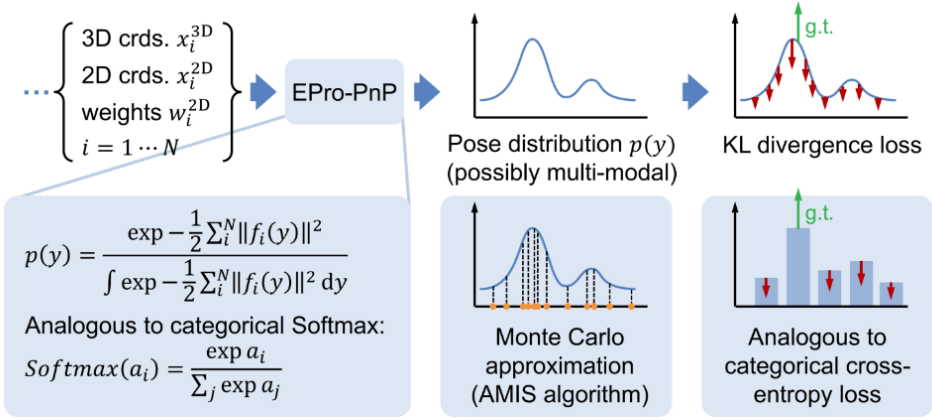




Full Inference Pipeline

Proposed Approaches

EPro-PnP Layer: outputs differentiable probability density function (PDF) of pose



Simplified Loss function

$$L = \frac{1}{2}\sum_i^N \|f_i(y_{gt})\|^2 + \log \int \exp -\frac{1}{2}\sum_{i=1}^N \|f_i(y)\|^2 \, dy$$

Gradients

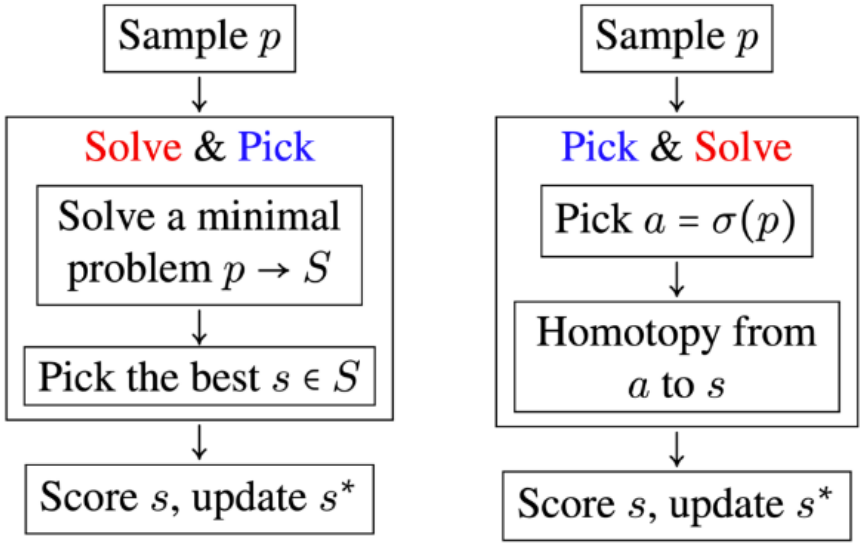
$$\frac{\partial L}{\partial (\cdot)} = \frac{\partial}{\partial (\cdot)} \frac{1}{2}\sum_i^N \|f_i(y_{gt})\|^2 - \mathop{\mathbb{E}}_{y \sim p(y)} \frac{\partial}{\partial (\cdot)} \frac{1}{2}\sum_{i=1}^N \|f_i(y)\|^2$$

Interpretation

- Minimize the reprojection error at the true pose, making the weighted points aware of reprojection **uncertainty**
- Maximize the reprojection error over the predicted pose, s.t. the points are **discriminative** to wrong poses

CVPR'22 Best Papers Honorable Mention

- *Learning to Solve Hard Minimal Problems,* ETH Zurich, Uni. of Washington, Georgia Tech, Czech Tech Uni. (Best Paper)



- *Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields,* Harvard Uni., Google (Best Student Paper Honorable Mention)

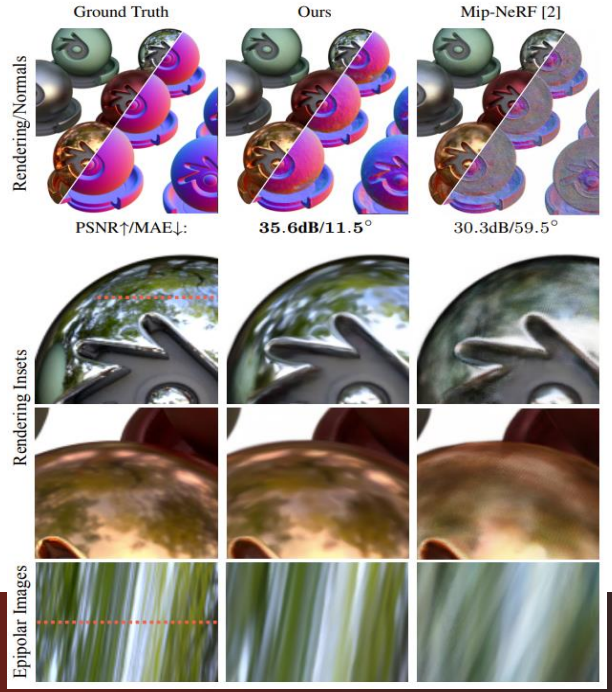
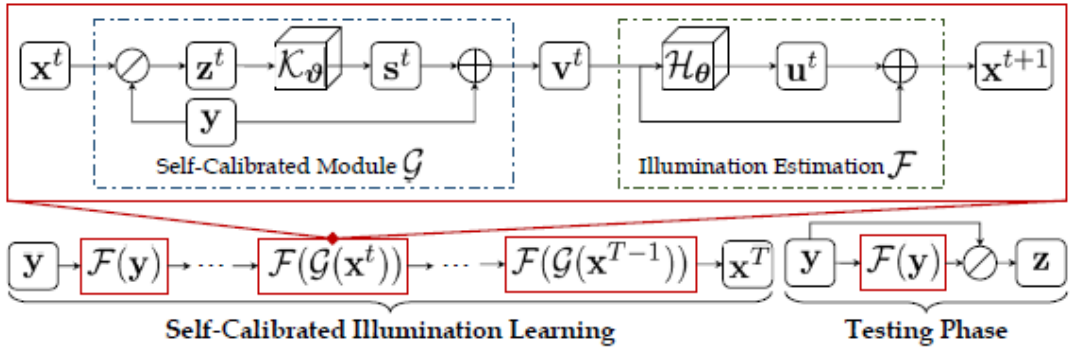


Image Quality Enhancement

Yuting ZHOU, Yan YE, Shimiao LI

Image Quality Enhancement (I)

- *Toward Fast, Flexible, and Robust Low-Light Image Enhancement*, DUT (Oral)
  - Contribution
    - established a lightweight yet effective framework
    - Self-Calibrated Illumination (SCI) to deal with different low-light real-world scenarios
  - Idea

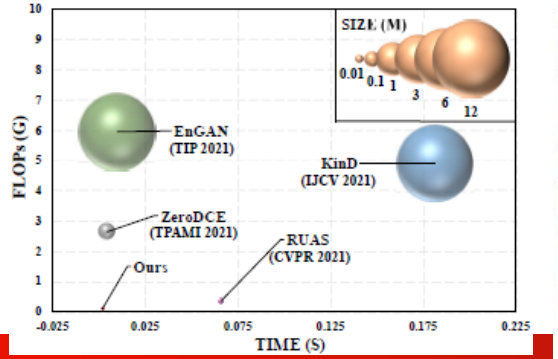


$$\mathcal{F}(x^t): \begin{cases} u^t = \mathcal{H}_\theta(x^t), x^0 = y \\ x^{t+1} = x^t + u^t \end{cases}$$

$$\mathcal{G}(x^t): \begin{cases} z^t = y \oslash x^t \\ s^t = \mathcal{K}_\theta(z^t) \\ v^t = y + s^t \end{cases}$$

$$Loss: \alpha \sum_{t=1}^{T} \|x^t - (y + s^{t-1})\|^2 + \beta \sum_{i=1}^{N} \sum_{j \epsilon \mathcal{N}(i)} \omega_{i,j} |x_i^t - x_j^t|$$

  - Results



Dataset	Metrics	*Recent Traditional Methods*			*Supervised Learning Methods*				*Unsupervised Learning Methods*				
		LECARM	SDD	STAR	RetinexNet	FIDE	DRBN	KinD	EnGAN	SSIENet	ZeroDCE	RUAS	Ours
MIT	PSNR↑	17.5993	**19.5241**	17.6464	13.7444	17.1902	17.5910	17.0935	16.7682	10.1396	16.6114	18.5372	**20.4459**
	SSIM↑	0.8556	**0.8690**	0.7793	0.7394	0.7853	0.7840	0.8307	0.8346	0.6456	0.8144	0.8642	**0.8934**
	DE↑	6.8069	6.8253	6.3677	6.2850	6.6543	6.5914	6.7233	**7.0382**	6.3879	6.2116	6.9068	**7.0429**
	EME↑	8.8779	8.6987	5.9128	9.1800	8.4146	7.4620	8.5482	7.9499	5.3423	7.8658	**10.6396**	**10.9627**
	LOE↓	613.2689	505.2951	**70.5651**	1812.853	264.4661	705.2620	500.6578	812.9041	646.9047	508.2960	579.0181	**273.3409**
	NIQE↓	4.3627	4.6477	4.2611	4.5289	5.2720	4.8166	4.2658	**3.9997**	5.2792	4.0933	4.1754	**3.9630**
LSRW	PSNR↑	15.4747	14.6694	14.6080	15.9062	**17.6694**	16.1497	16.4717	16.3106	**16.7380**	15.8337	14.4372	15.0168
	SSIM↑	0.4635	0.5061	0.5039	0.3725	**0.5485**	**0.5422**	0.4929	0.4697	0.4873	0.4664	0.4276	0.4846
	DE↑	5.9980	6.7307	6.4943	6.9392	6.8745	7.2051	**7.0368**	6.6692	**7.0988**	6.8729	5.6056	6.5524
	EME↑	**24.4089**	8.5431	9.4636	14.6119	5.6885	9.9968	12.0881	22.2345	9.3801	20.8010	23.5139	**24.9625**
	LOE↓	**34.1438**	296.0794	**103.2322**	591.2793	194.7405	755.1283	379.8994	248.1947	261.2802	219.1284	357.4125	280.8935
	NIQE↓	3.8189	5.6401	3.7537	4.1479	4.3277	4.5500	**3.6636**	3.7754	4.0631	3.7183	4.1687	**3.6590**

Image Quality Enhancement (II)

- *A Lightweight Network for High Dynamic Range Imaging*, Uni. of Adelaide (workshop)
  - Problem to solve
    - solve ghosting artifacts occurring in multi-frame HDR reconstruction on dynamic scene or hand-held camera

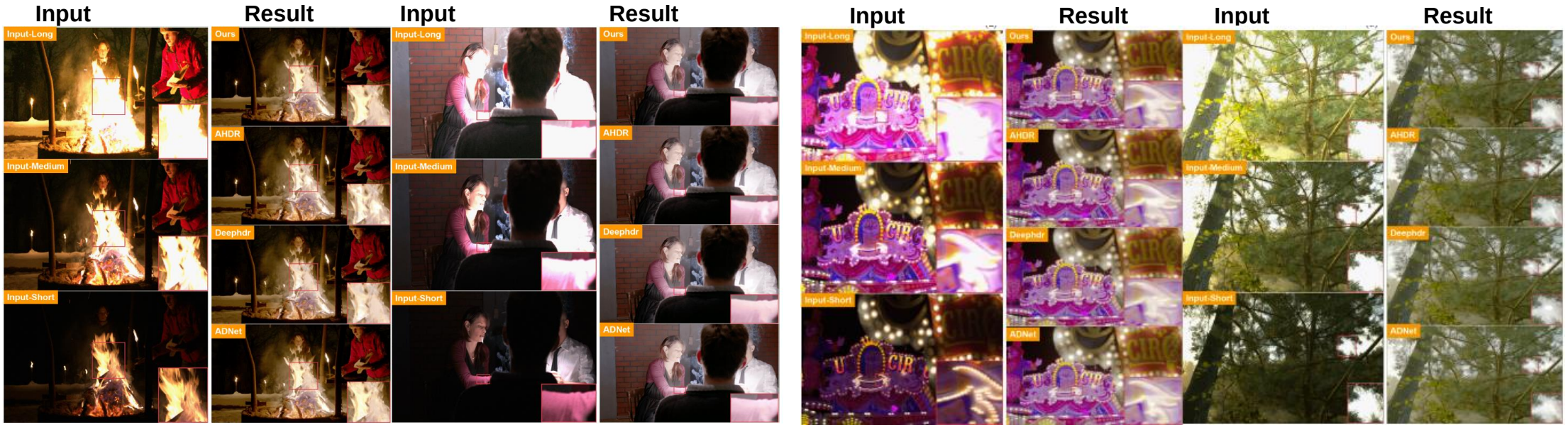


Figure 1: Visual comparison on the testing data. The LDR images are shown on the left. The proposed network can produce a high-quality HDR image.

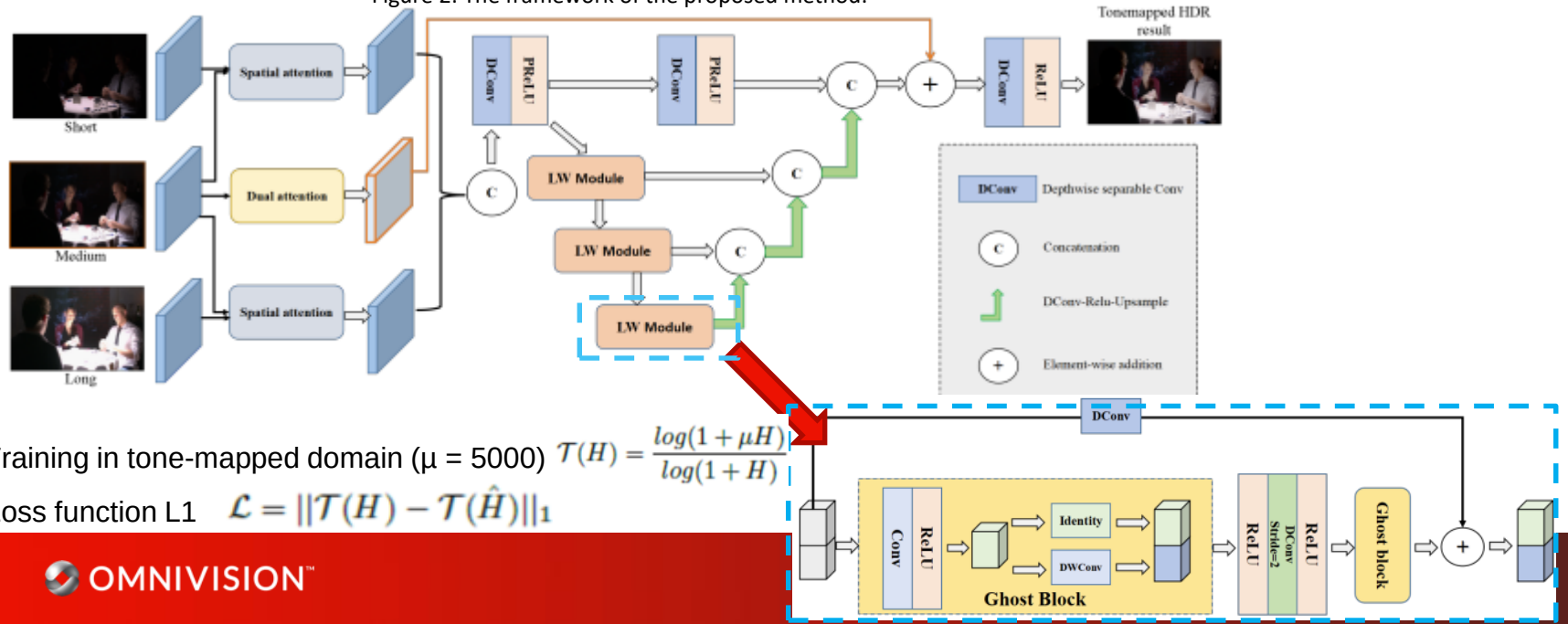
Image Quality Enhancement (II)

- *A Lightweight Network for High Dynamic Range Imaging*, Uni. of Adelaide (workshop)
  - Architecture and training process
    - Attention network: removing misaligned regions between no-reference & reference imgs
    - Dual attention model: highlighting the useful regions of the reference image, which forces the fusion network to learn the details of degenerated regions with a residual
    - A lightweight (LW) module with fewer parameters

* Spatial attention: $A_i = a_i(F_i, F_r).$

* Dual attention (spatial and channel): $A_2 = ca(sa(F_2)),$

Figure 2: The framework of the proposed method.



* Training in tone-mapped domain (μ = 5000) $\mathcal{T}(H) = \frac{log(1+\mu H)}{log(1+H)}$

* Loss function L1 $\mathcal{L} = ||\mathcal{T}(H) - \mathcal{T}(\hat{H})||_1$

Figure 3. Attention maps of the spatial block. The first row shows 3 LDR images, the second and third rows are attention maps of Long and Short frame.

Image Quality Enhancement (II)

- *A Lightweight Network for High Dynamic Range Imaging*, Uni. of Adelaide (workshop)
  - Experiments and Results
    - Dataset: 149 scenes as validation, 1,345 scenes as training (from NTIRE2022 HDR Challenge)
    - Evaluation Metrics: PSNR, PSNR-μ

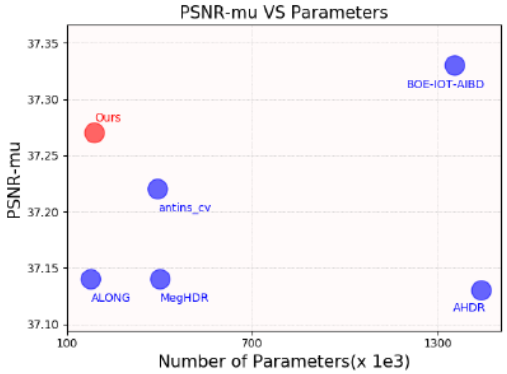


Table 1. Compared with SOTA methods. * denotes the model size in 200 GMACs.

Model	PSNR	PSNR-μ	GMACs	Para.
DeepHDR	37.57	31.93	1983.38	16606339
AHDRNet	38.94	32.60	2916.92	1441283
ADNet	39.73	32.88	6249.43	3132773
DeepHDR*	36.16	31.21	180.79	1096706
AHDRNet*	36.32	31.26	186.17	90679
Ours	39.29	32.73	156.12	188992

Table 2. Ablation study on the network structure.

Model	PSNR	PSNR-μ	GMACs	Para.
Baseline	36.32	31.26	186.17	90679
Model1	37.60	31.49	187.78	91511
Model2	38.31	32.22	150.80	188992
Ours	39.29	32.73	156.12	188992

- Baseline : small version of AHDRNet
- Model1 : add dual attention to baseline
- Model2 : encoder-decoder structure
- Ours: full model of HUNet

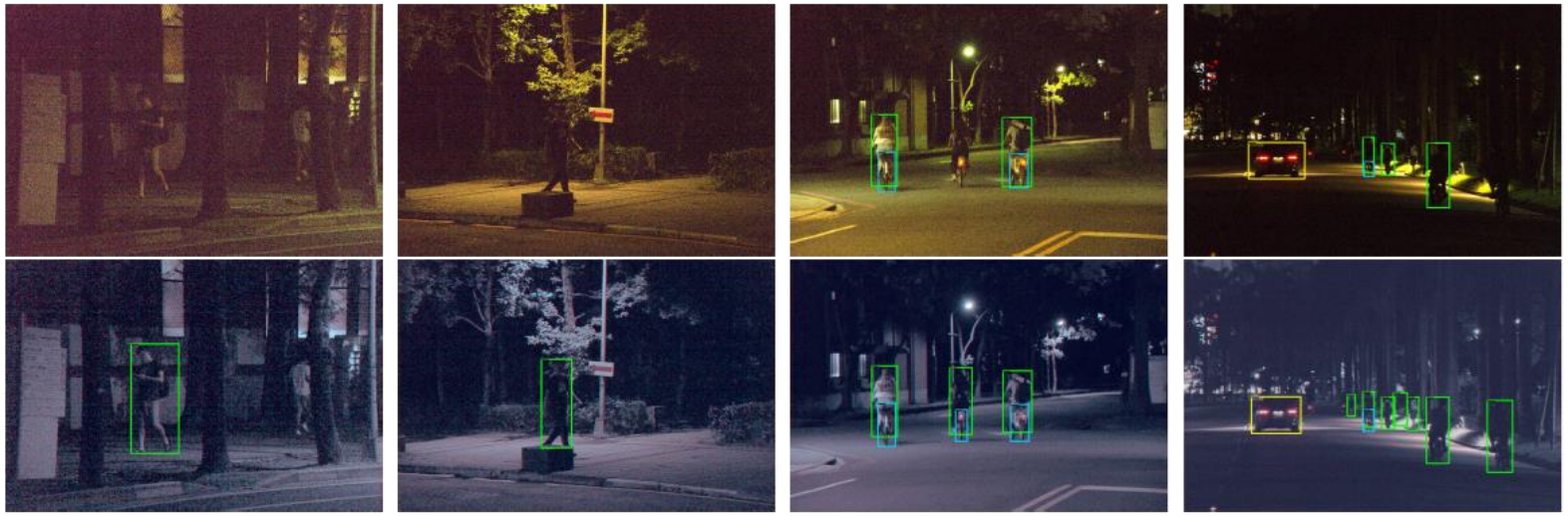
  - Conclusion
    - A hybrid network for HDR image de-ghosting
    - Dual attention module to highlight details of useful regions
    - Lightweight but better performance than prior work

Image Quality Enhancement (III)

- *GenISP: Neural ISP for Low-Light Machine Cognition*, NTU (workshop)
  - Key idea

- **low-light image restoration and enhancement** with an OD dataset
- **optimal for machine cognition** rather than human visual system
- **trained under the guidance of a pre-trained detector** wo assumptions about how the **enhanced** image should look
- no need for **paired** low- and normal-light ground-truth data for training
- Improves pre-trained detector performance

- *Make us rethink about the relationship btw neural ISP & object detector*
- *The strategy might be applied to other domain adaptation problems for OD*

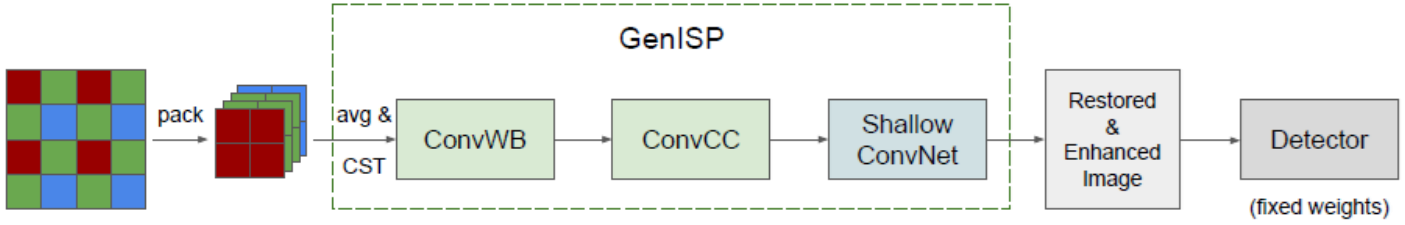


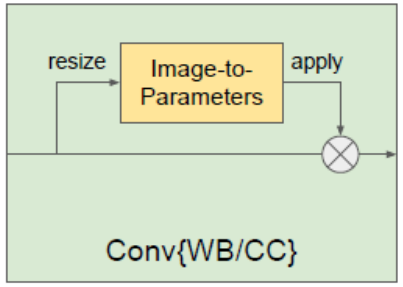
Traditional Minimal ISP + Pre-Trained Detector

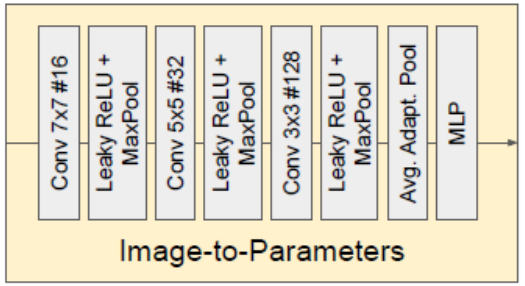
GenISP + Pre-Trained Detector

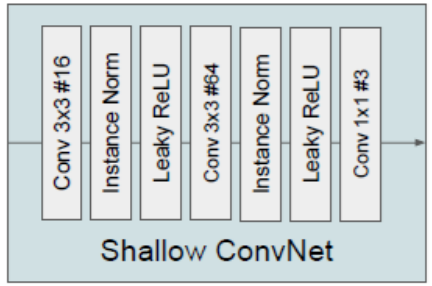
Image Quality Enhancement (III)

- *GenISP: Neural ISP for Low-Light Machine Cognition*, NTU (workshop)
  - Technical details









$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} w_{11} & 0 & 0 \\ 0 & w_{22} & 0 \\ 0 & 0 & w_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$ ConvWB

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$ ConvCC

$$L_{total} = L_{cls} + L_{reg} + \lambda_{wb} L_{wb}$$ Training Loss

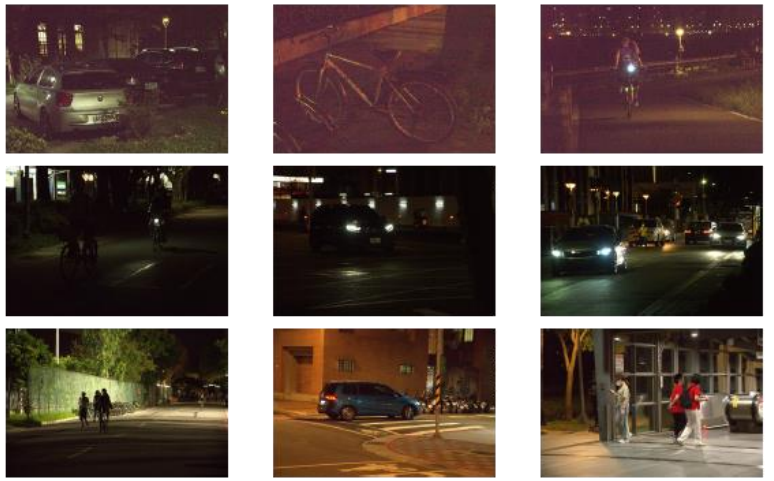
$$L_{wb} = \sum_{\forall (c_1, c_2) \in C} |J^{c_1}_{enh} - J^{c_2}_{enh}|, C = \{(R,G),(R,B),(G,B)\}$$ Gray-world hypothesis

  - a minimal ISP pipeline, two image-to-parameter modules and a shallow ConvNet
  - explicitly Color Space Transformation (CST) matrices available with raw files (instead of encoding CST implicitly)
    - helps improving the capability to generalize to unseen sensors and eliminate the need for re-training for each camera model
  - a two-stage color processing implemented by two image-to-parameters modules: ConvWB and CovCC
    - introducing expert knowledge about ISP and improving the detection results both when CST matrices are available and unavailable
  - # parameters: 0.12M, GMACs of GenISP: 3.3

Image Quality Enhancement (III)

- *GenISP: Neural ISP for Low-Light Machine Cognition*, NTU (workshop)
  - RAW-NOD dataset
    - 7K raw images collected, bbox annotations of people, bicycles and cars
    - publicly available for task-based benchmarking of future low-light image restoration and low-light object detection

Dataset	Camera Sensor	#class	#annotated images	#instance
Sony	RX100 VII	3	3.2K	18.7K
Nikon	D750	3	4K	28K



  - Result: generalizing well to unseen datasets, camera sensors, brightness levels, object detectors

Tested on	Input type	Method	Trained on	Detection eval. (↑) AP_{50}	AP_{75}	AP
Our Nikon	JPEG	Baseline (Traditional ISP)	-	41.6%	19.5%	21.2%
		Histogram Equalization	-	36.9%	16.1%	18.5%
		LIME [5]	-	41.2%	19.1%	20.7%
		Zero-DCE [4]	Our Sony	40.7%	18.5%	20.5%
		Zero-DCE [4]	SICE [1]	42.4%	19.7%	21.4%
	RAW	Lamba and Mitra [13]	SID Sony	did not generalize		
		Histogram Equalization	-	40.3%	19.0%	20.5%
		SID [2]	SID Sony	44.5%	21.8%	22.9%
		Lamba and Mitra [12]	SID Sony	45.5%	21.2%	23.0%
		Our	Our Sony	**47.0%**	**23.4%**	**24.5%**

comparison with other low-light enhancement methods

		AP (↑) Baseline	Ours
Low-light	Our Sony	21.1%	**25.7%**
	Our Nikon	22.2%	**24.5%**
	SID Sony [2]	21.2%	**22.6%**
Normal-light	PASCALRAW [21]	37.4%	**39.6%**

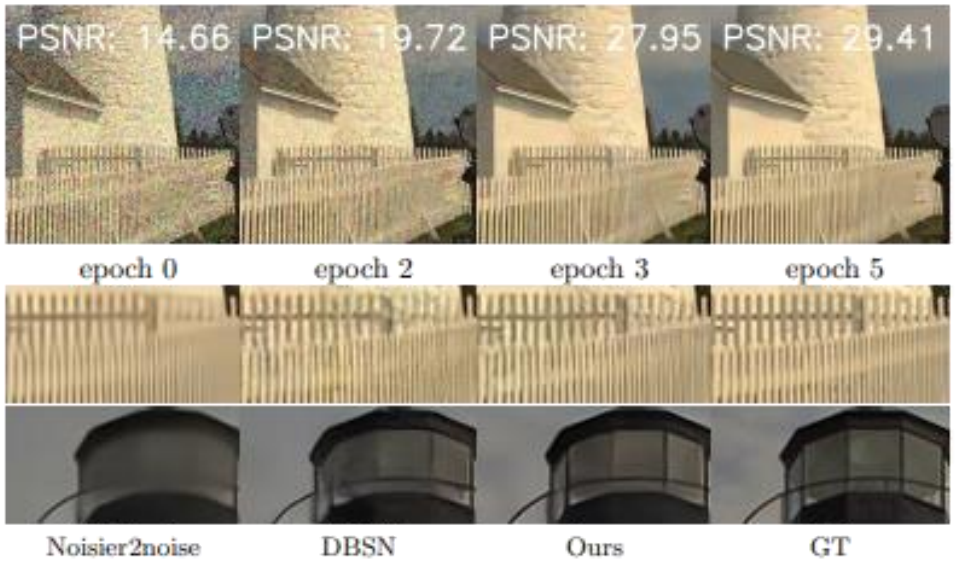
Trained on SONY RX100 dataset, cross dataset / brightness level evaluation

		AP (↑) Baseline	Ours
Two-stage	Faster R-CNN [24]	20.6%	**25.1%**
Single-stage	FCOS [29]	18.7%	**19.2%**
	RetinaNet [16]	21.1%	**25.7%**
	PAA [10]	21.8%	**25.8%**

Cross object detector evaluation

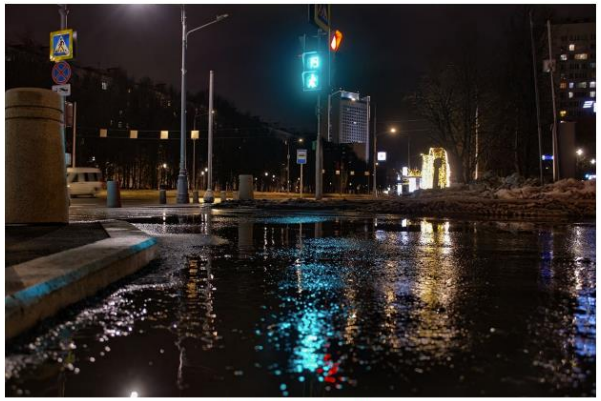
Image Quality Enhancement (IV)

- *Other Interesting Papers*
  - IDR: Self-Supervised Image Denoising via Iterative Data Refinement (CUHK, SenseTime)
    - Training using noiser-noisy dataset for denoising instead of noise-clean dataset
    - Iterative training and data generation strategy



  - Deep-FlexISP: A Three-Stage Framework for Night Photography Rendering (Xiaomi)
    - Winner of night photography rendering challenge
    - Cascaded three stage neural ISP for night photography rendering


(a) Baseline


(b) Our Deep-FlexISP

Object Recognition

Shijie XIAO, Jiawei CHEN, Wenyuan QIU

Object Detection

- *Technical Trend*
  - Backbone
    - CNN: ConvNext, RepLKNet,…
    - Transformer: MPViT, MetaFormer, Mobile-Former,…
  - Object detection
    - Faster convergence: DN-DETR, AdaMixer, SAM-DETR…
    - Better Performance: BoxR
    - Distillation: LD, FGD

DN-DETR: Accelerate DETR Training by Introducing Query DeNoising

Object Detection

- *DN-DETR: Accelerate DETR Training by Introducing Query DeNoising*, HKUST (oral)
  - Background – DETR
    - Highlights
      - Set-based prediction
      - Bipartite matching
    - Drawbacks
      - Slow convergence
      - Inexplicit meaning of query

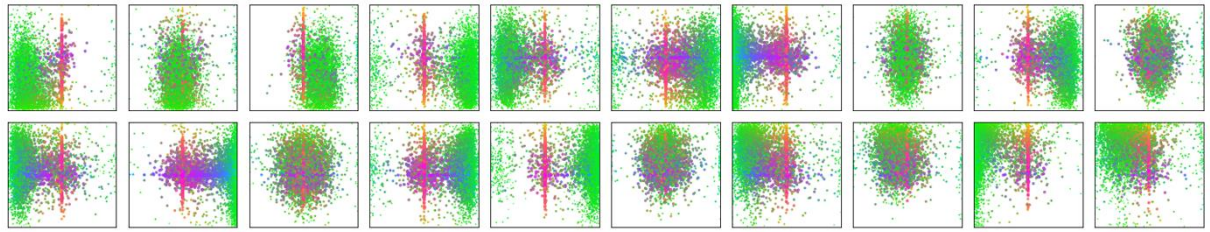
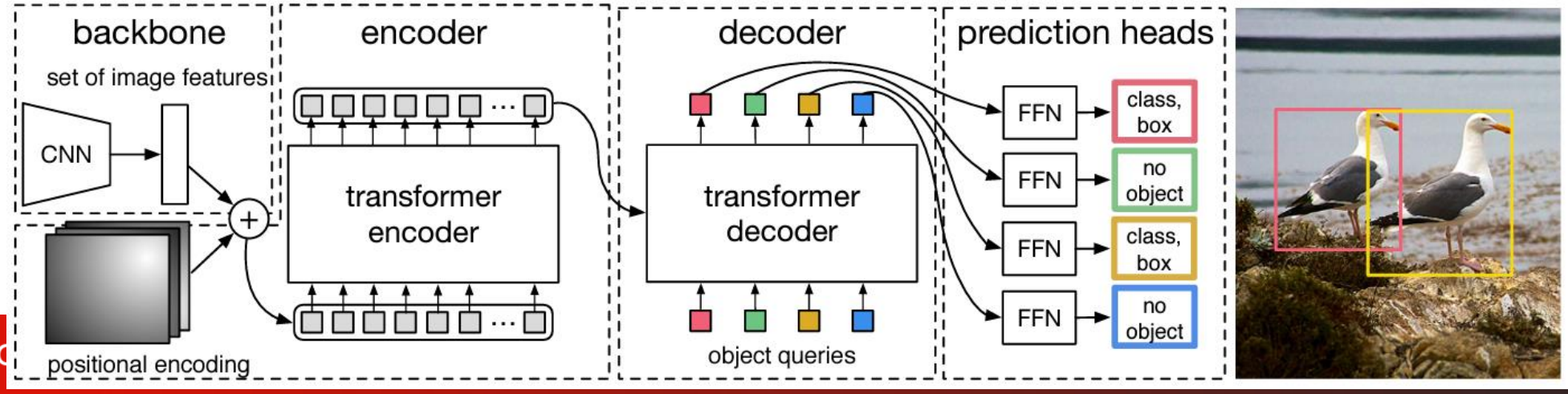
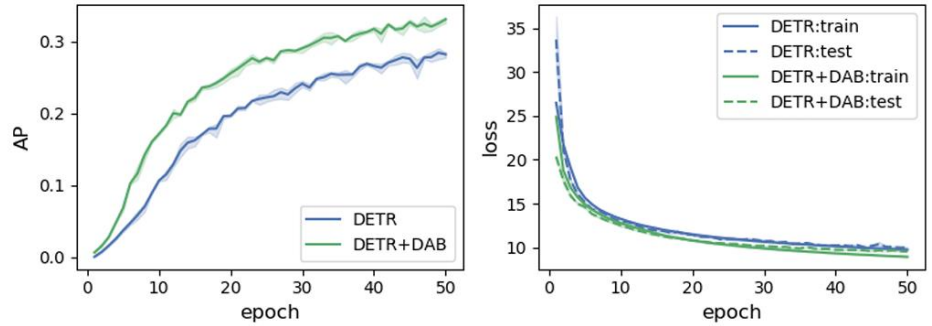


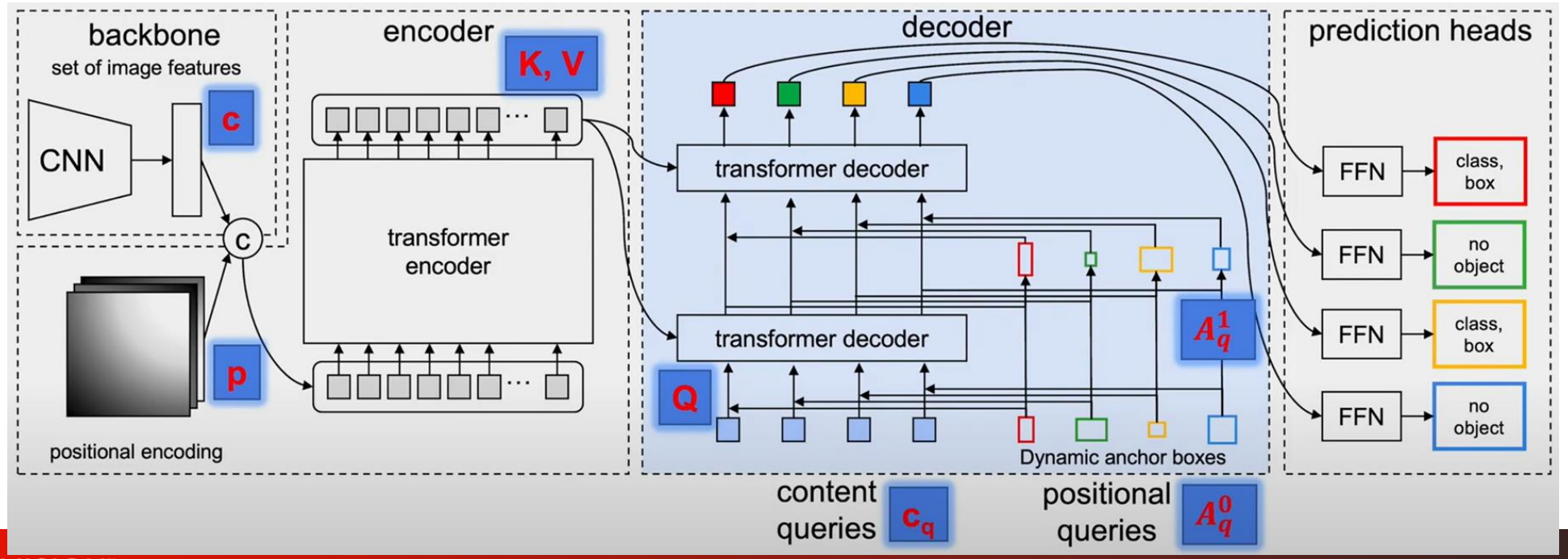
Fig. 7: Visualization of all box predictions on all images from COCO 2017 val set for 20 out of total $N = 100$ prediction slots in DETR decoder. Each box prediction is represented as a point with the coordinates of its center in the 1-by-1 square normalized by each image size. The points are color-coded so that green color corresponds to small boxes, red to large horizontal boxes and blue to large vertical boxes. We observe that each slot learns to specialize on certain areas and box sizes with several operating modes. We note that almost all slots have a mode of predicting large image-wide boxes that are common in COCO dataset.



Object Detection

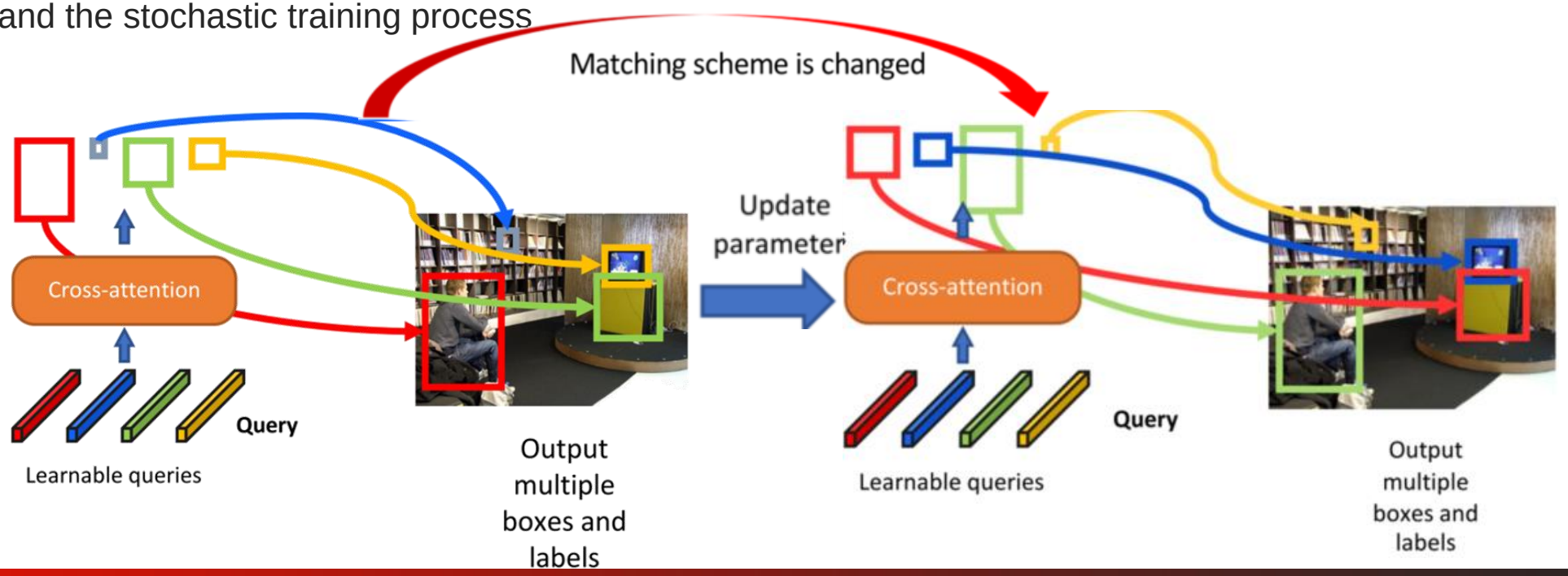
- *DN-DETR: Accelerate DETR Training by Introducing Query DeNoising*, HKUST (oral)
  - Background – DAB-DETR
    - Highlights
      - Queries are formulated as dynamic anchor boxes
      - Faster convergence and better performance





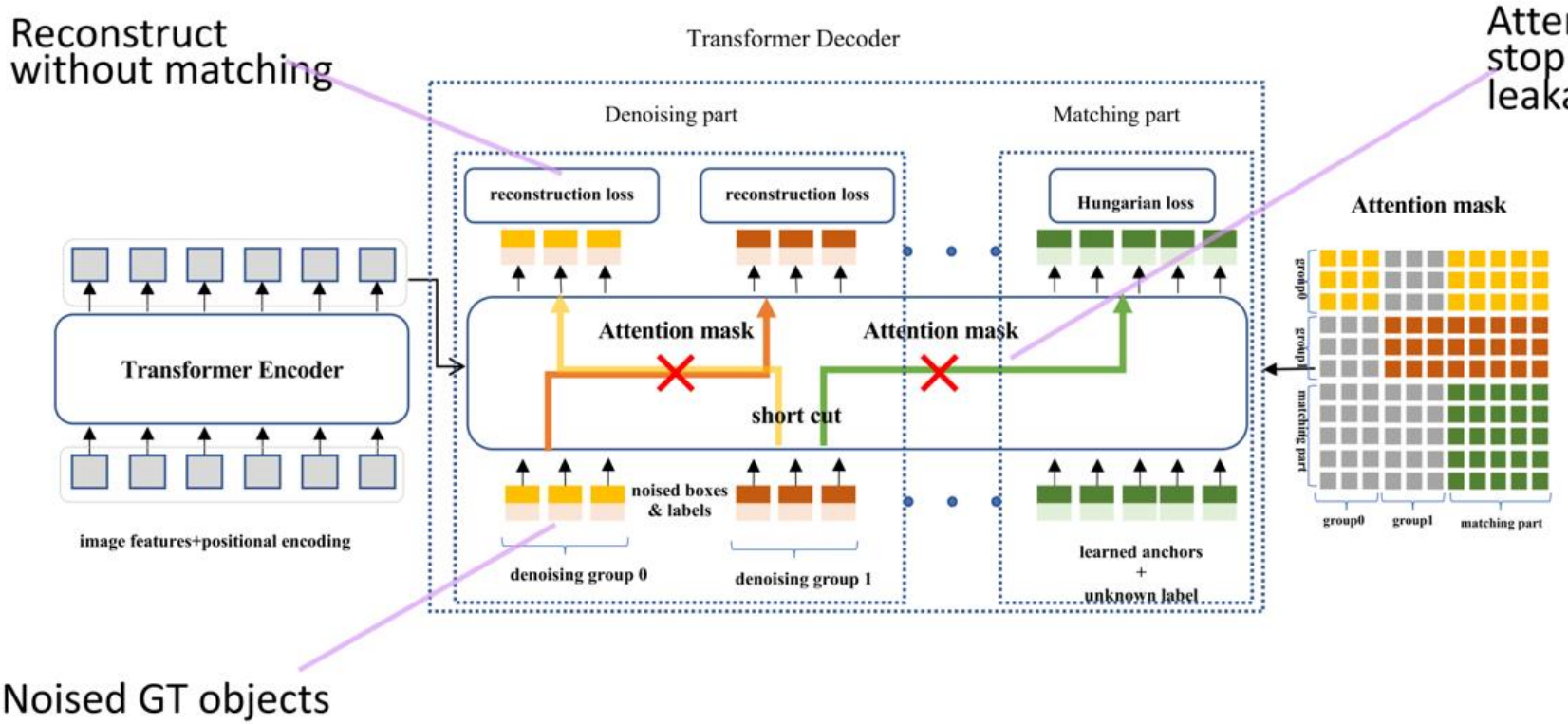
Object Detection

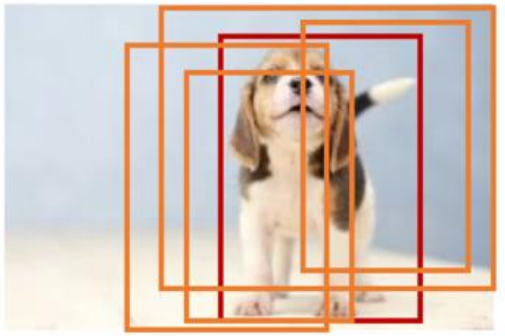
- *DN-DETR: Accelerate DETR Training by Introducing Query DeNoising*, HKUST (oral)
  - Motivation of DN-DETR
    - Possible reason of DETR's slow convergence
      - DETR adopts Hungarian matching so solve the matching problem between predicted objects and ground truth objects.
      - Ground truth assignment is a dynamic process in DETR, which may cause instability issue due to its discrete bipartite matching and the stochastic training process



Object Detection

- *DN-DETR: Accelerate DETR Training by Introducing Query DeNoising*, HKUST (oral)
  - Solution
    - Assign Transformer decoder another task - query denoising
      - Input: groundtruth (GT) bboxes & labels with noise
      - Expected Output: GT bboxes & labels



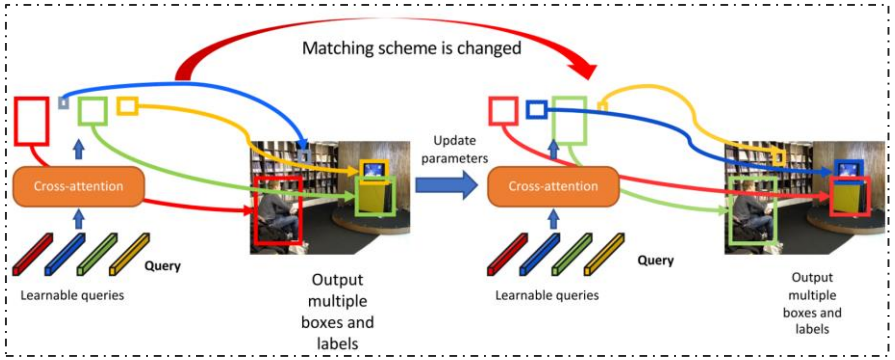


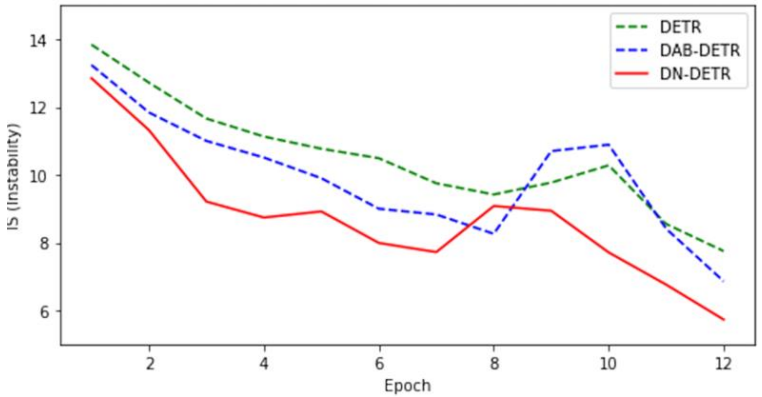
An example of noised GT objects

$$a_{ij} = \begin{cases} 1, & \text{if } j < P \times M \text{ and } \lfloor \frac{i}{M} \rfloor \neq \lfloor \frac{j}{M} \rfloor; \\ 1, & \text{if } j < P \times M \text{ and } i \geq P \times M; \\ 0, & \text{otherwise.} \end{cases}$$

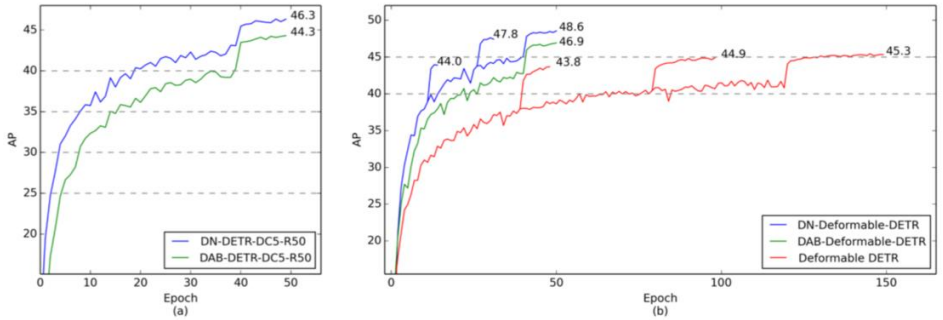
Object Detection

- *DN-DETR: Accelerate DETR Training by Introducing Query DeNoising*, HKUST (oral)
  - Results
    - Less instability during training





    - Faster convergence



    - Better performance

Model	#epochs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	GFLOPs	Params
DETR-R50 [1]	500	42.0	62.4	44.2	20.5	45.8	61.1	86	41M
Faster RCNN-FPN-R50 [15]	108	42.0	62.1	45.5	26.6	45.5	53.4	180	42M
Anchor DETR-R50 [18]	50	42.1	63.1	44.9	22.3	46.2	60.0	–	39M
Conditional DETR-R50 [12]	50	40.9	61.8	43.3	20.8	44.6	59.2	90	44M
DAB-DETR-R50 [11]	50	42.2	63.1	44.7	21.5	45.7	60.3	94	44M
DN-DETR-R50	50	**44.1(+1.9)**	64.4	46.7	22.9	48.0	63.4	94	44M
DETR-R101 [1]	500	43.5	63.8	46.4	21.9	48.0	61.8	152	60M
Faster RCNN-FPN-R101 [15]	108	44.0	63.9	47.8	27.2	48.1	56.0	246	60M
Anchor DETR-R101 [18]	50	43.5	64.3	46.6	23.2	47.7	61.4	–	58M
Conditional DETR-R101 [12]	50	42.8	63.7	46.0	21.7	46.6	60.9	156	63M
DAB-DETR-R101 [11]	50	43.5	63.9	46.6	23.6	47.3	61.5	174	63M
DN-DETR-R101	50	**45.2(+1.7)**	65.5	48.3	24.1	49.1	65.1	174	63M
DETR-DC5-R50 [1]	500	43.3	63.1	45.9	22.5	47.3	61.1	187	41M
Anchor DETR-DC5-R50 [18]	50	44.2	64.7	47.5	24.7	48.2	60.6	151	39M
Conditional DETR-DC5-R50 [12]	50	43.8	64.4	46.7	24.0	47.6	60.7	195	44M
DAB-DETR-DC5-R50 [11]	50	44.5	65.1	47.7	25.3	48.2	62.3	202	44M
DN-DETR-DC5-R50	50	**46.3(+1.8)**	66.4	49.7	26.7	50.0	64.3	202	44M
DETR-DC5-R101 [1]	500	44.9	64.7	47.7	23.7	49.5	62.3	253	60M
Anchor DETR-R101 [18]	50	45.1	65.7	48.8	25.8	49.4	61.6	–	58M
Conditional DETR-DC5-R101 [12]	50	45.0	65.5	48.4	26.1	48.9	62.8	262	63M
DAB-DETR-DC5-R101 [11]	50	45.8	65.9	49.3	27.0	49.8	63.8	282	63M
DN-DETR-DC5-R101	50	**47.3(+1.5)**	67.5	50.8	28.6	51.5	65.0	282	63M

Model	MultiScale	#epochs	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	GFLOPs	Params
Faster R50-FPN 1x [15]	✓	12	37.9	58.8	41.1	22.4	41.1	49.1	180	40M
DETR-R50 1x [1]		12	15.5	29.4	14.5	4.3	15.1	26.7	86	41M
DAB-DETR-DC5-R50 [11]		12	38.0	60.3	39.8	19.2	40.9	55.4	216	44M
DN-DETR-DC5-R50		12	**41.7(+3.7)**	61.4	44.1	21.2	45.0	60.2	216	44M
Deformable DETR-R50 1x [20]	✓	12	37.2	55.5	40.5	21.1	40.7	50.5	173	40M
Dynamic DETR-R50† 1x (without dynamic encoder)	✓	12	40.2	58.6	43.4	––	–	–	–	–
Dynamic DETR-R50† 1x [4]	✓	12	42.9	61.0	46.3	24.6	44.9	54.4	–	–
DN-Deformable-DETR-R50 [4]	✓	12	**43.4**	61.9	47.2	24.8	46.8	59.4	195	48M
DAB-DETR-DC5-R101 [11]		12	40.3	62.6	42.7	22.2	44.0	57.3	282	63M
DN-DETR-DC5-R101		12	**42.8(+2.5)**	62.9	45.7	23.3	46.6	61.3	282	63M
Faster R101 FPN [15]	✓	108	44.0	63.9	47.8	27.2	48.1	56.0	246	60M
DN-Deformable-DETR-R101	✓	12	**44.1**	62.8	47.9	26.0	47.8	61.3	275	67M

Multi-Person Pose Estimation (MPPE)

- *Technical Trend*
  - Keywords: 6D Pose, 2D/3D human pose

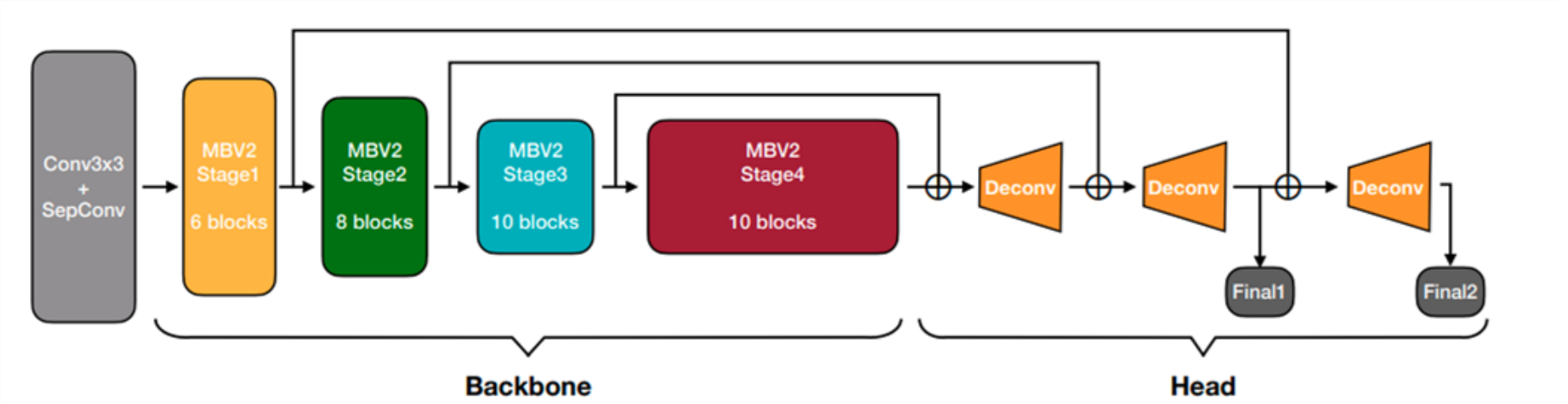
Topics	No. of Papers in CVPR22
2D/3D Human Pose Estimation	19
6D Pose Estimation	13
Lite Pose	1

Lite Pose – Efficient Architecture Design 2D for Human Pose Estimation

(A lightweight network for 2D MPPE)

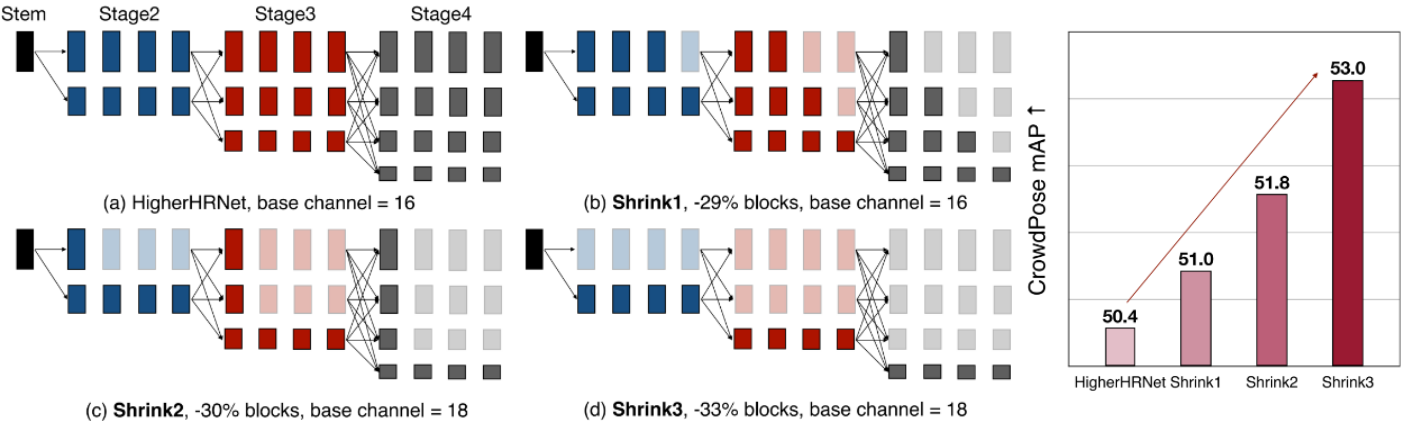
Multi-Person Pose Estimation (MPPE)

- *Lite Pose – Efficient Architecture Design 2D for Human Pose Estimation,* Tsinghua, CMU, MIT (Poster)
  - Target: real-time MPPE on edge
  - Key insights:
    - Single-branch architecture is efficient
    - Large kernel convolution is efficient
    - Light-weight fusion deconv head

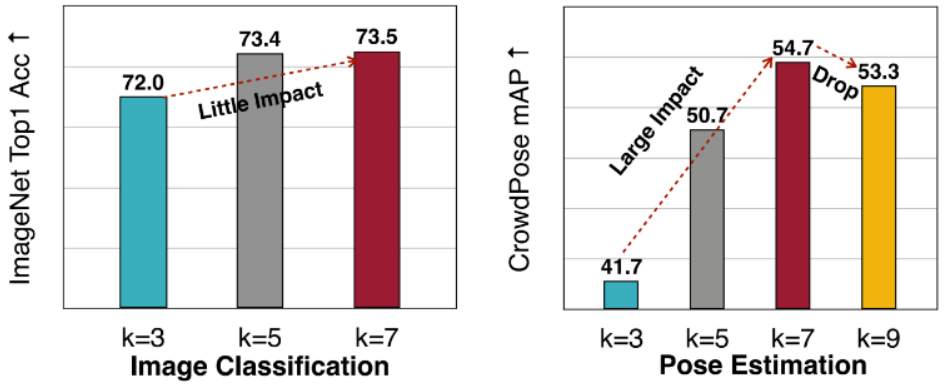


Multi-Person Pose Estimation (MPPE)

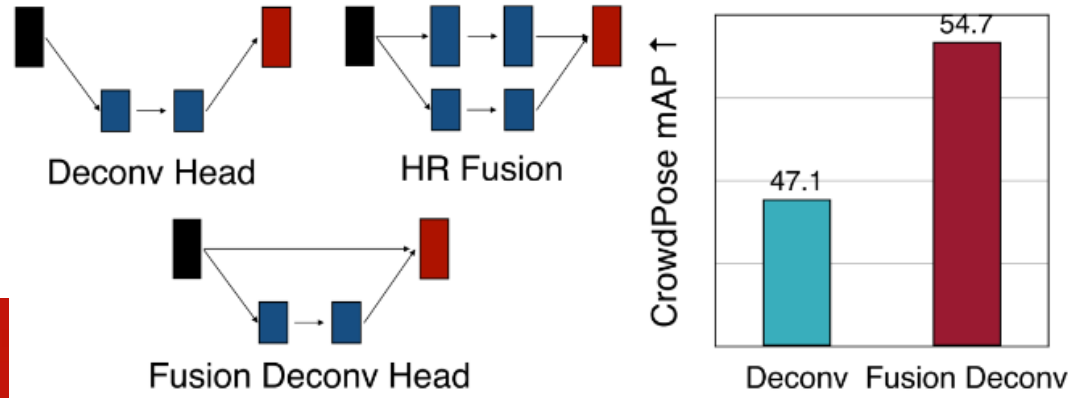
- *Lite Pose – Efficient Architecture Design 2D for Human Pose Estimation,* Tsinghua, CMU, MIT (Poster)
  - Redundancy in high-resolution branches



  - Large kernel is efficient

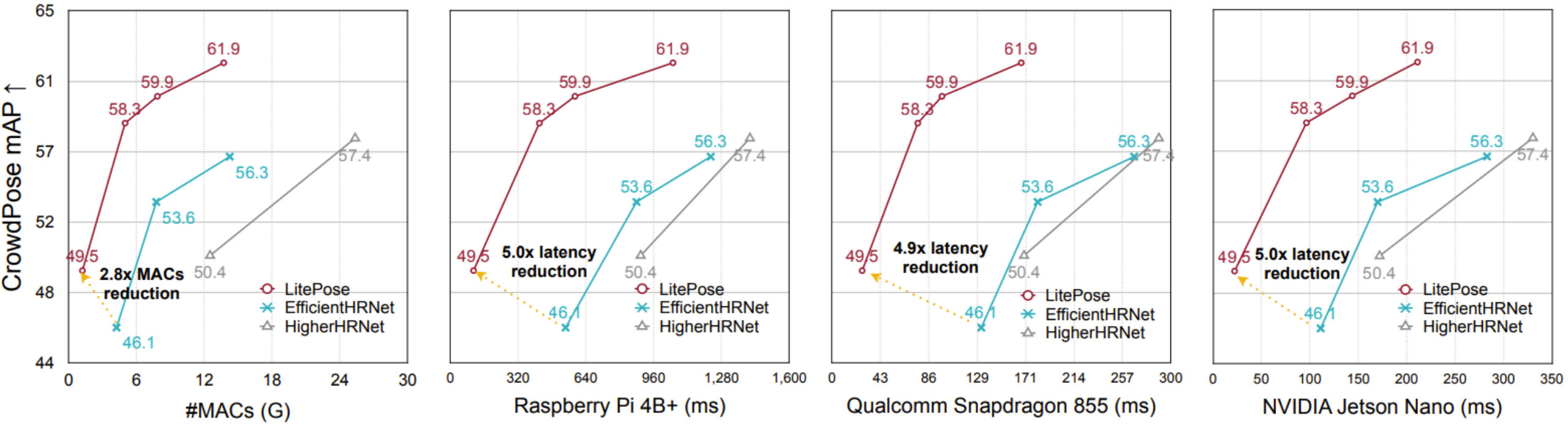


  - Light-weight fusion deconv head



Multi-Person Pose Estimation (MPPE)

- *Lite Pose – Efficient Architecture Design 2D for Human Pose Estimation,* Tsinghua, CMU, MIT (Poster)
  - Results: compare with SOTA on the CrowdPose Dataset



2.8x MACs Reduction, 5.0x Speed Up

Dataset

- *Kubric: A scalable dataset generator, Google* (Poster)
  - Why Synthetic Data?

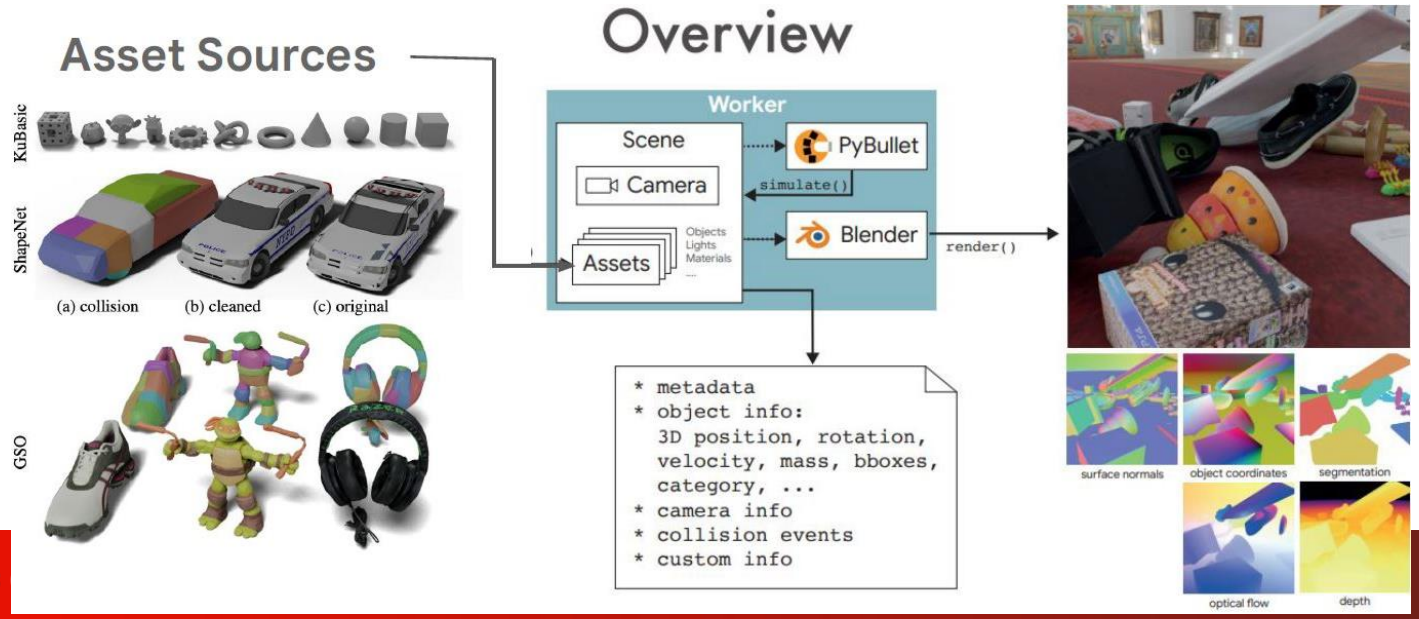
Synthetic Data	Real Data
• Cheap and Scalable	• Needs time to collect and process
• Supports rich ground-truth	• Needs to be well annotated
• Offers full control over data	• May have unbalancing issues
• No copyright issues	• May have privacy and legal concerns
• Not enough realism	• Close to the usage scenario

  - Comparison

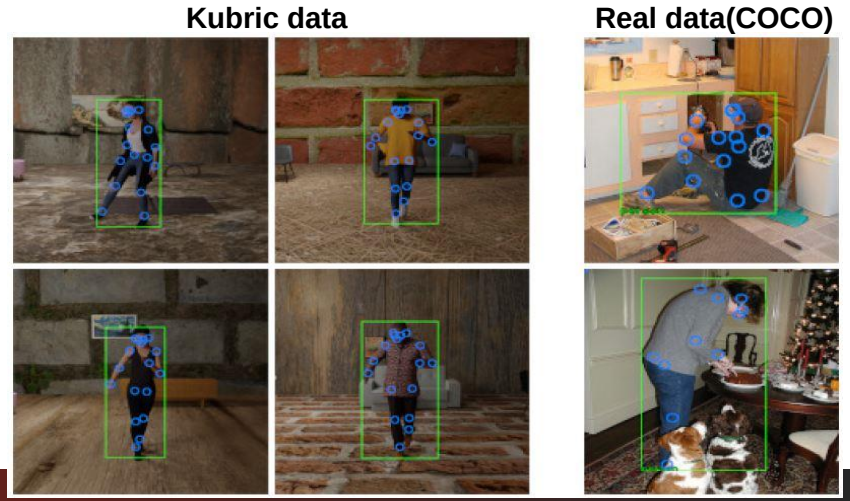
Name	Rendering	GI	Physics	Scaling	DL
Playing4Data	(Game)	×	(Game)	×	×
UnrealCV	UE4	×	UE4	✓	×
TDW	Unity	×	PhysX	✓	×
iGibson	PyRender	×	PyBullet	✓	×
Habitat	Magnum	×	Bullet	✓	×
OpenRooms	OptiX	✓	–	×	×
Omnidata	Blender	✓	–	×	✓
Blenderproc	Blender	✓	Bullet	×	×
Kubric	**Blender**	✓	**PyBullet**	✓	✓

Dataset

- *Kubric: A scalable dataset generator, Google* (Poster)
  - an open-source Python framework for generating photo-realistic synthetic scenes dataset
    - generates high-fidelity synthetic data with photorealistic rendering, physical collision meshes, and sophisticated 3D models
    - Builds around PyBullet physics simulator and Blender as raytracing engine
    - Supports 3 open-source 3D dataset and 1 texture dataset as asset
  - Application: Object discovery, Optical flow, Pose estimation, Object detection …



Examples in generating Pose estimation dataset



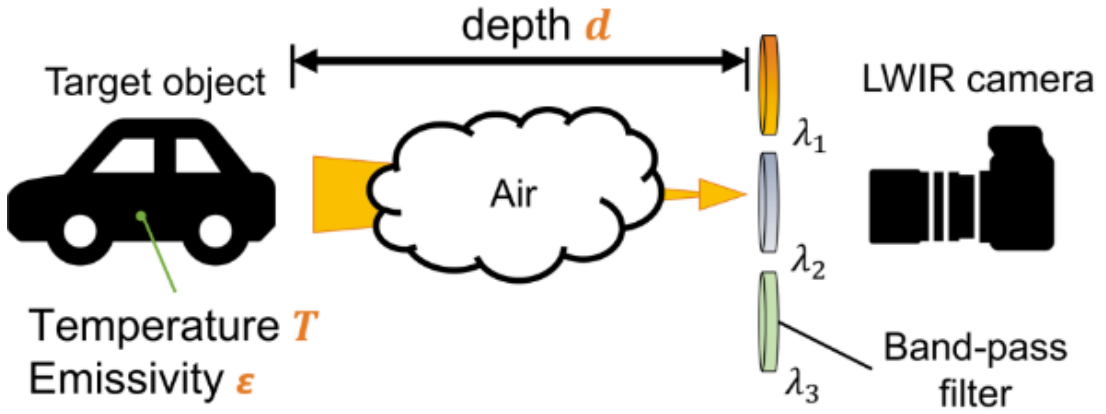
New Image Sensing Modality

Ruijiang LUO

Depth Sensing

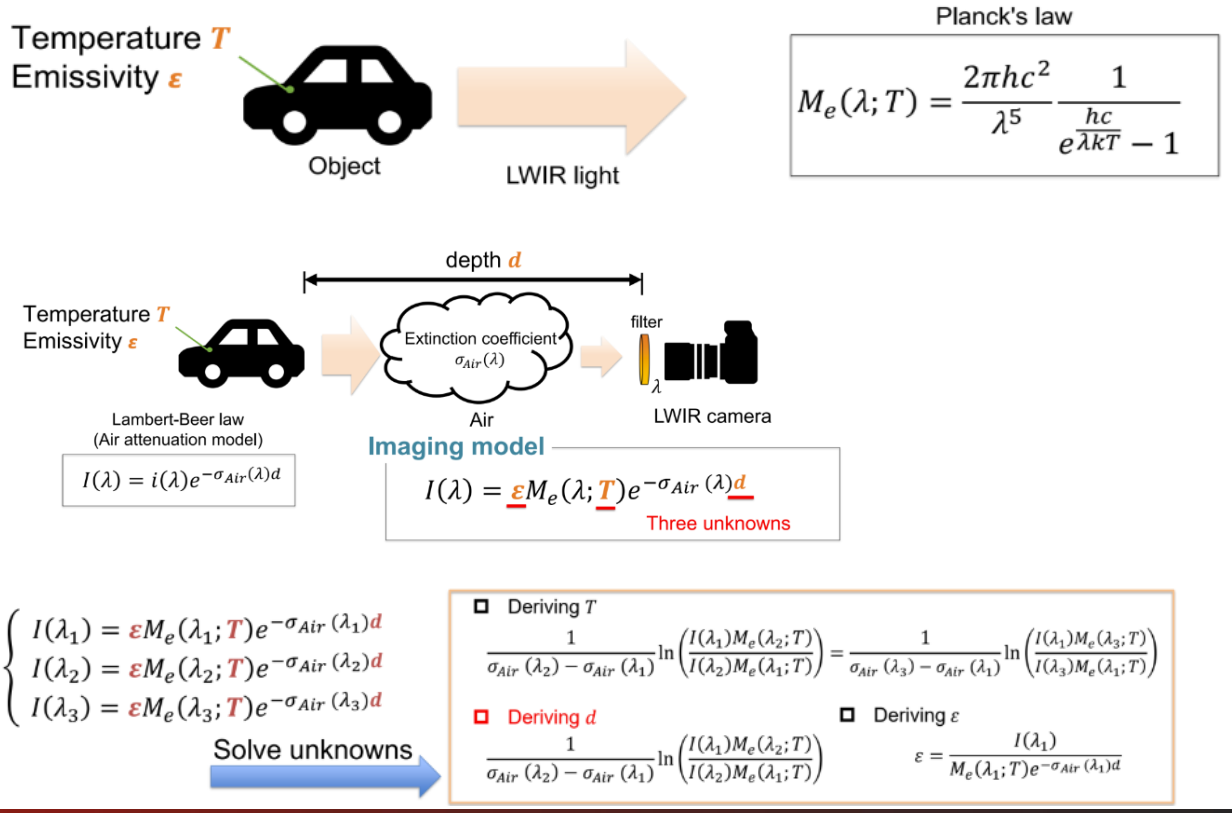
- *Shape From Thermal Radiation: Passive Ranging Using Multi-Spectral LWIR Measurements*, NAIST (poster)
  - Challenge – is passive ranging possible using a single thermal camera?
    - A novel depth sensing technique refer to physics-based cues, long wavelength infrared (LWIR)
    - Advantages
  - Key idea
    - Modeling **LWIR light attenuation** through the air
    - Multi-spectral observation enables estimate **not only depth but also temperature and emissivity**

	Stereo cameras	Structured light	LiDAR	*Ours*
Dark scene	Impossible	Possible	Possible	***Possible***
Texture independence	No	Yes	Yes	***Yes***
Far range	Baseline	Impossible	Possible	***Possible***
Outdoor	Possible	Difficult	Possible	***Possible***
Risk of interfering	Low	High	High	***Low***
Stealth measurement	Yes	No	No	***Yes***



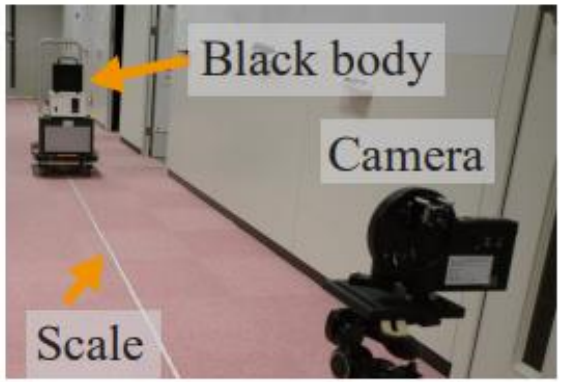
Depth Sensing

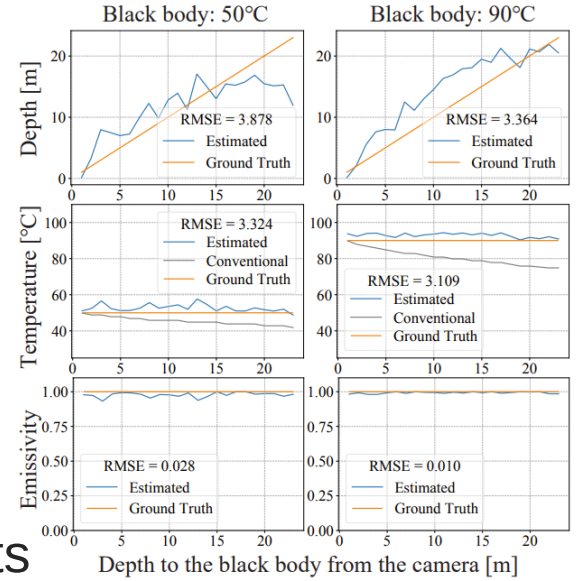
- *Shape From Thermal Radiation: Passive Ranging Using Multi-Spectral LWIR Measurements*, NAIST (poster)
  - Method
    - Thermal radiation theory
      - all object radiate LWIR light radiation according to Planck's law, depending on temperature & emissivity
    - Light attenuation and imaging model
      - Lambert-Beer law expresses transmission of the air
      - Attenuated LWIR light at a specific wavelength observed, by using a filter
    - Estimate depth
      - Derived from the difference in extinction coefficient at different wavelengths

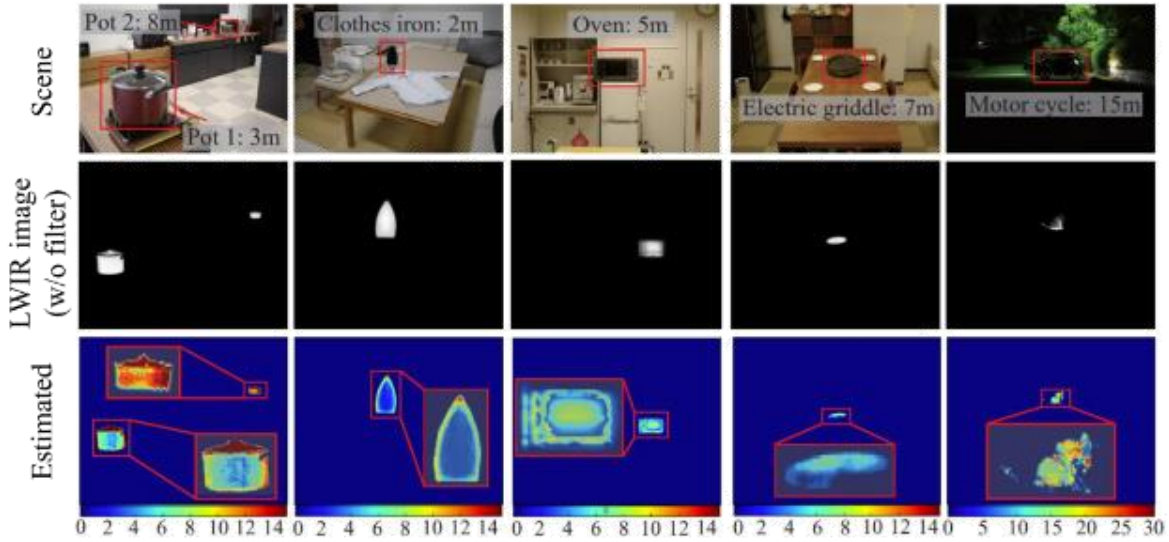


Depth Sensing

- *Shape From Thermal Radiation: Passive Ranging Using Multi-Spectral LWIR Measurements*, NAIST (poster)
  - Experimental results







  - Conclusion & comments
    - First attempt: novel & unique depth sensing approach using LWIR camera – depth, temperature, emissivity
    - Effectiveness in real-world experiments
    - Passive, texture-less, far range, and can be used even in dark scenes
    - New to computer vision and computational imaging field of research
    - Won't be easily deceived (e.g. by placing a photo in front of the camera)

OMNIVISION™

Thank You