Computer Vision with Omnidirectional and Event Cameras

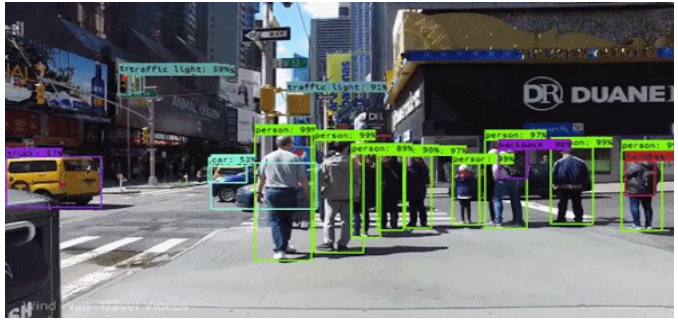
Kuk-Jin Yoon

Visual Intelligent Laboratory

KAIST

Performance Boost using Deep Learning with Frame-based Cameras

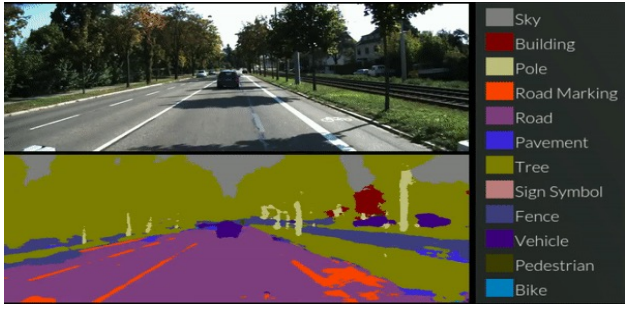
Computer Vision with Deep Learning has become the core of Autonomous Driving.



GIF source: https://github.com/abhishek1605/Detect-Me

GIF source: https://paperswithcode.com/task/depth-and-camera-motion/codeless.

Liu et al., SelFlow: Self-Supervised Learning of Optical Flow.



https://towardsdatascience.com/review-pspnet-winner-in-ilsvrc-2016-semantic-segmentation-scene-parsing-e089e5df177d

Limitations of ADAS using Conventional Frame-based Cameras (1)

Limited field-of-view

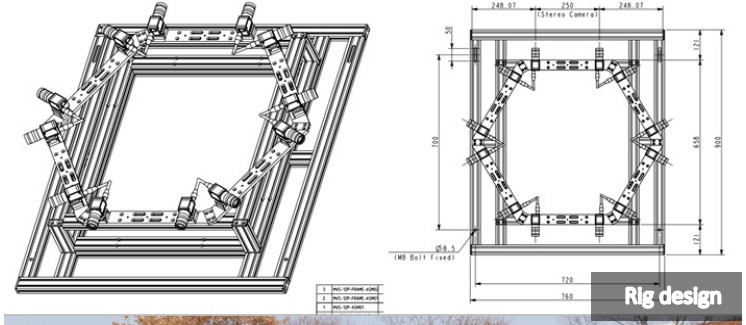
Limitations of ADAS using Conventional Frame-based Cameras (2)

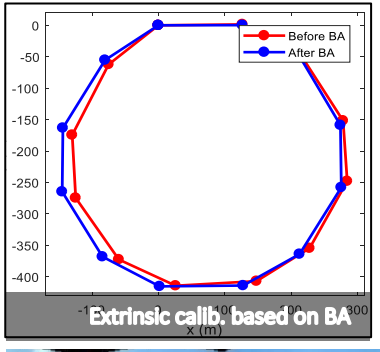
Low dynamic range and framerate

Limitations of ADAS using Conventional Frame-based Cameras (1) - Limited field-of-view

COMPUTER VISION WITH OMNIDIRECTIONAL CAMERAS

Multi-view-based ADAS





Multi-view-based ADAS

Disparity Image (SGM)

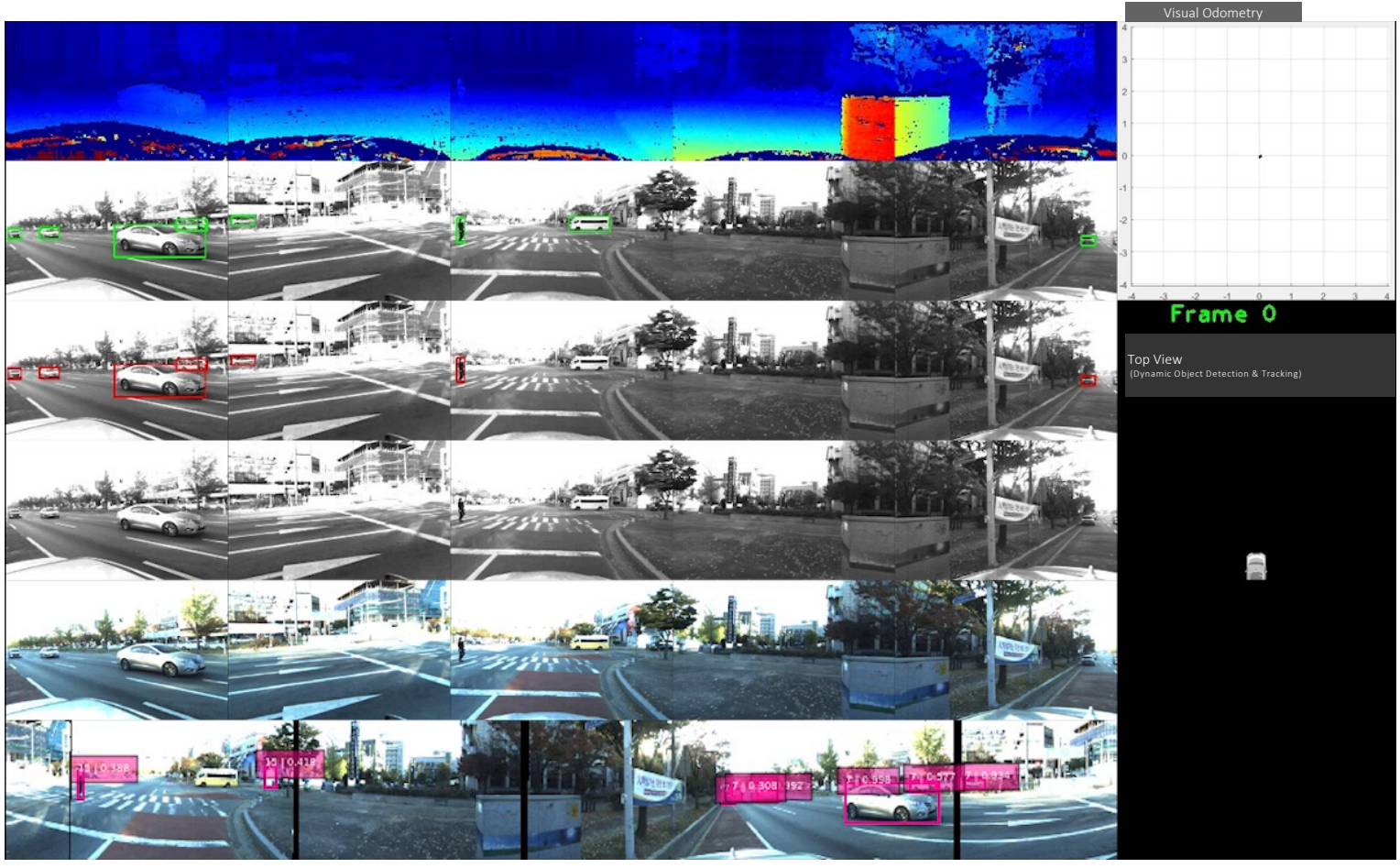
Appearance-based Object Detection

Motion-based Dynamic Object Detection

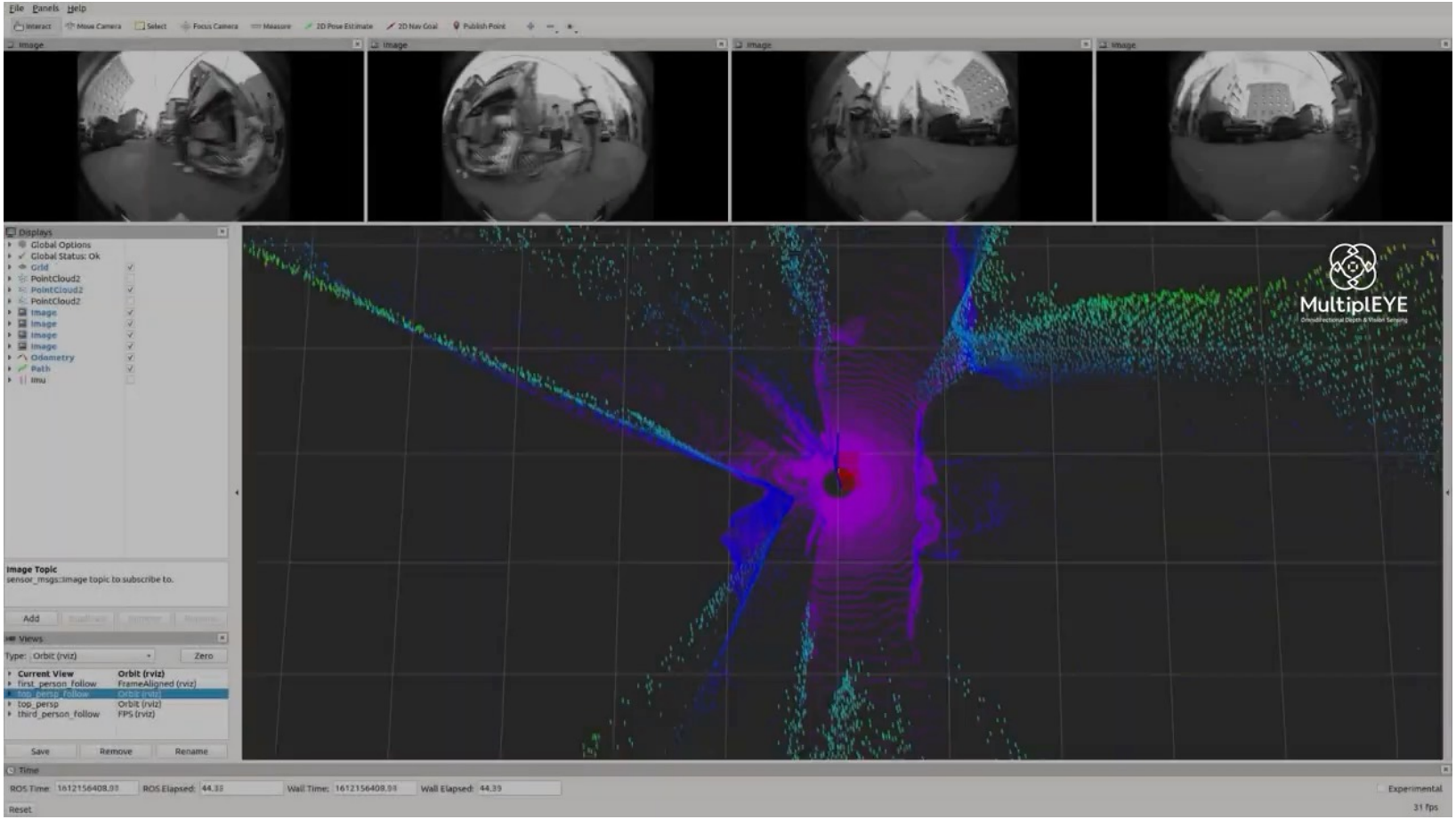
Motion-based Dynamic Object Tracking

RMN-based Object Tracking

Appearance-based Object Detection in 360º images



Multi-WFoV Camera-based ADAS



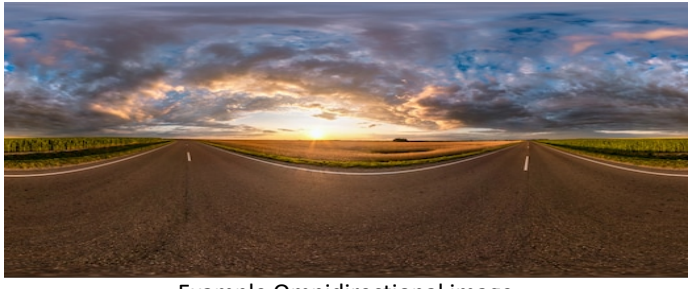
from Multipleye.co

Omni-directional Cameras

Omnidirectional images provide all-around information that can be obtained from a single viewpoint. As such, they are widely used in graphics, autonomous vehicles, drones, or robotics.

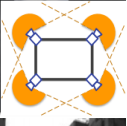
Samsung Gear 360 LG 360 Cam

Portable 360° cameras



Example Omnidirectional image

Autonomous Robots

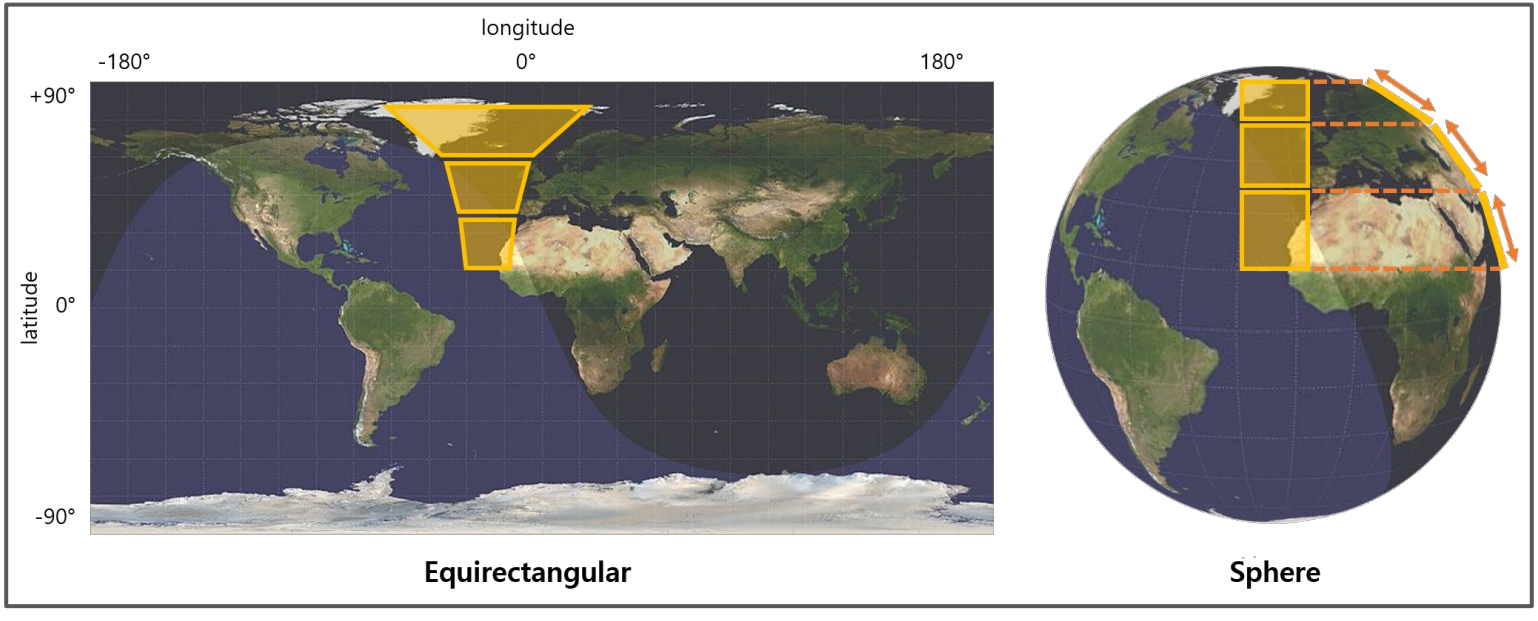


Hardware setup for acquiring grey-scale RGB and D Panorama

Representations of 360° Images

Equirectangular Projection (ERP)

- The most conventional way to represent 360° images
- Yellow squares on left and right images represent the same area.
- ERP has 2D-array data structure like other images

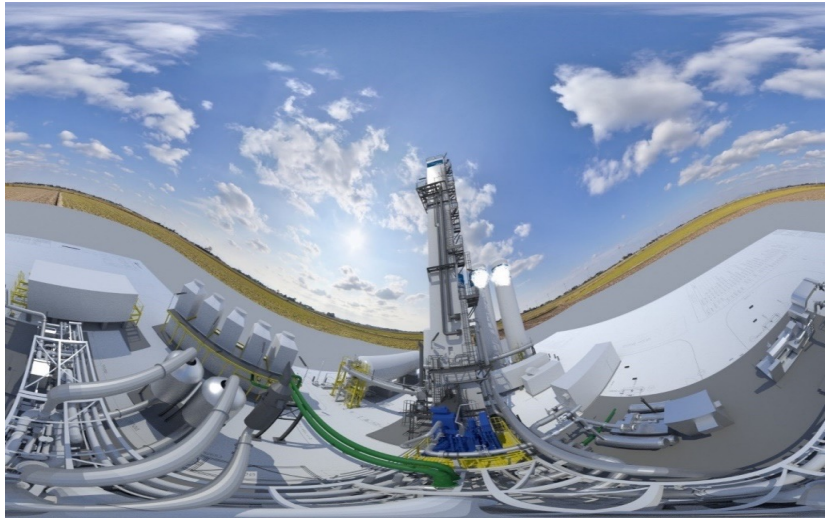


Issues of ERP-based 360° Image Representations

Disadvantages of ERP

- Content deforms around pole regions (±90 latitude)
- ERP representation also suffers from sinusoidal fluctuation when a camera is tilted.

Severe distortion around top and bottom sides



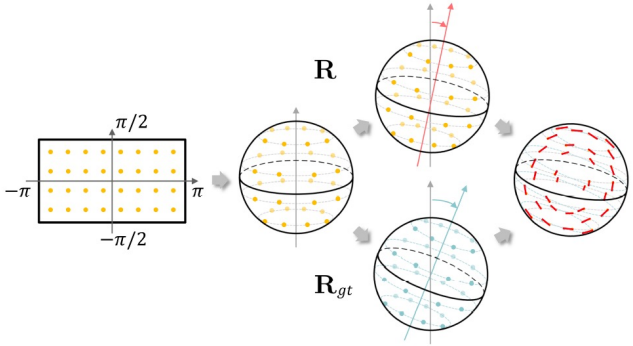
Sinusoidal fluctuation of ERP image

Those effects severely degrades performance of vision algorithm (e.g., detection, tracking…)

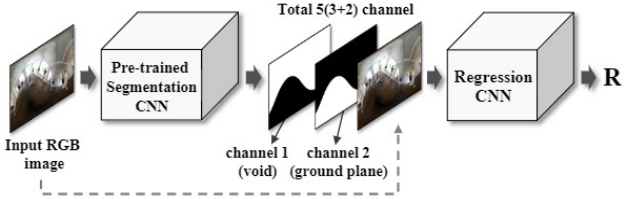
360˚ Image Upright Adjustment

Reducing the distortion of Equi-Rectangular Projection (ERP) image

- ERP shows significant fluctuation when tilting the 360˚ camera axis.
- Upright adjustment improves the object detection performance on ERP images.



Loss function using the distance on surface



CNN for upright adjustment

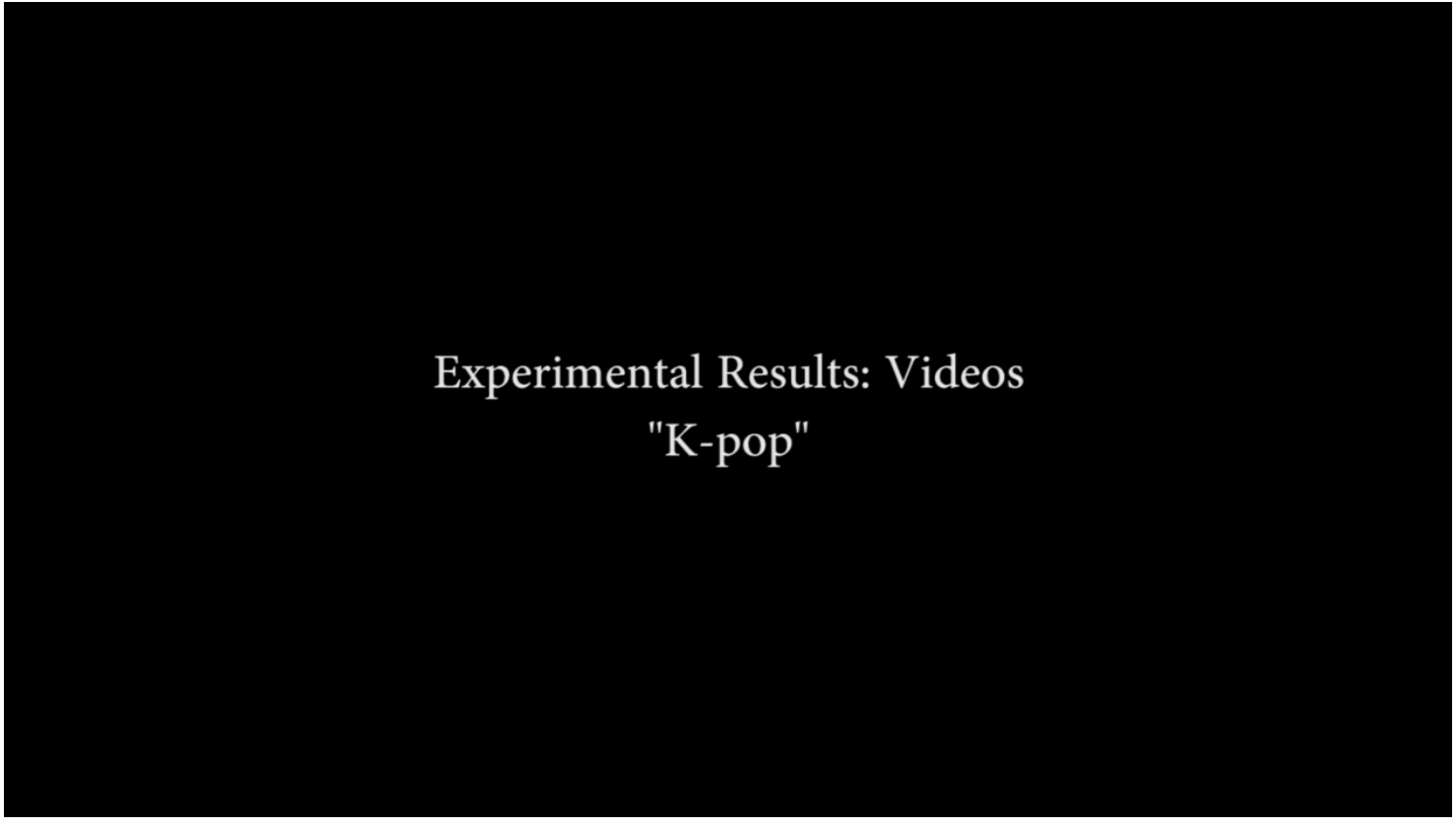
Input 360˚ images

Corrected 360˚ images

360° Image Upright Adjustment

Reducing the distortion of Equi-Rectangular Projection (ERP) image

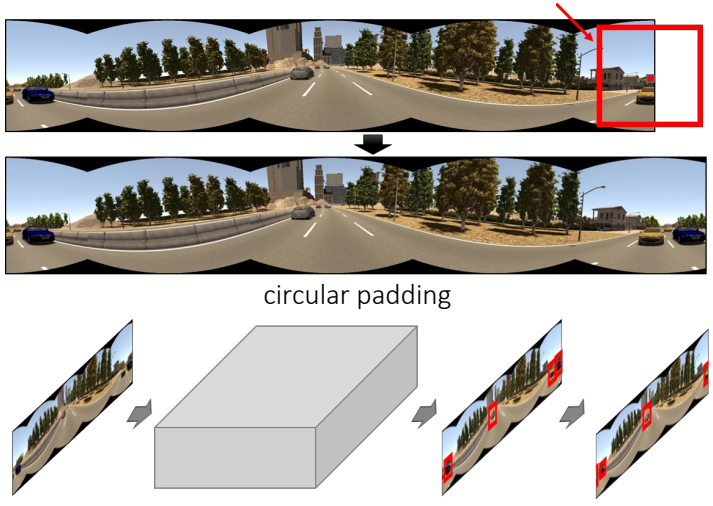
- ERP shows significant fluctuation when tilting the 360° camera axis.
- Upright adjustment improves the object detection performance on ERP images.



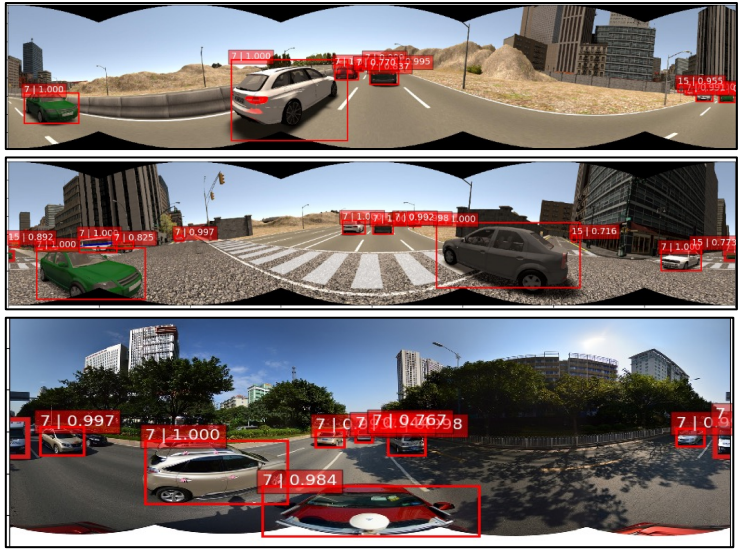
Object Detection using ERP-based 360° Image Representations

Discontinuous object detection from ERP images

- ERP has discontinuous on both side boundaries of an image.
- Detection[1] in these regions produces two different IDs.
- Additional padding in consideration of the receptive field could solve this problem.



Method for continuous object detection result



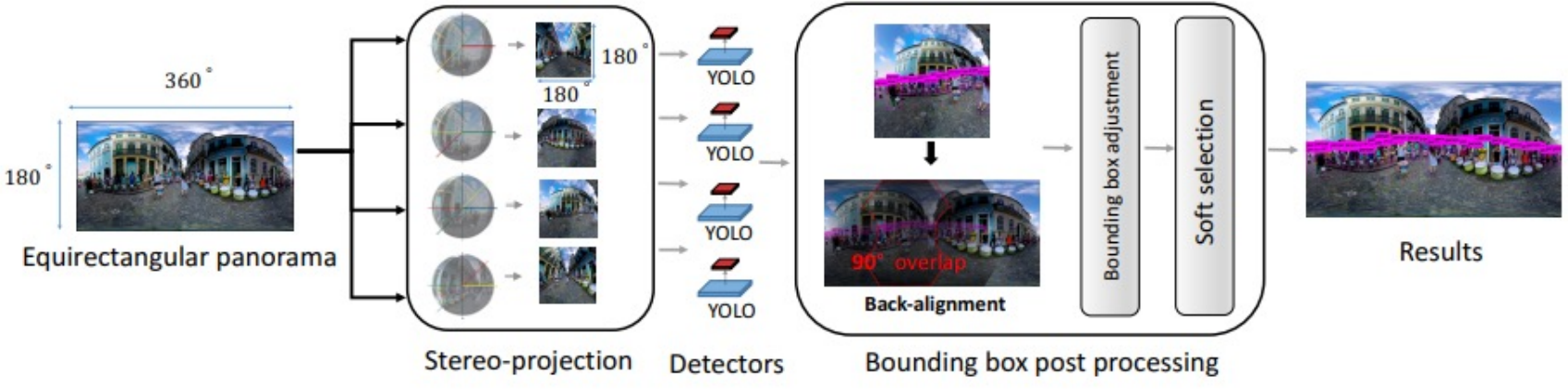
Continuous detection from ERP image

[1] Liu et al, “SSD: Single Shot MultiBox Detector”, ECCV, 2016.

Object Detection using 360° Images

ERP-based methods

- Split the ERP image into small pieces that do not include severe ERP distortion.
- Analyze each piece and gather the network output to one.
- Information discontinuity between the split images degrades the performance.



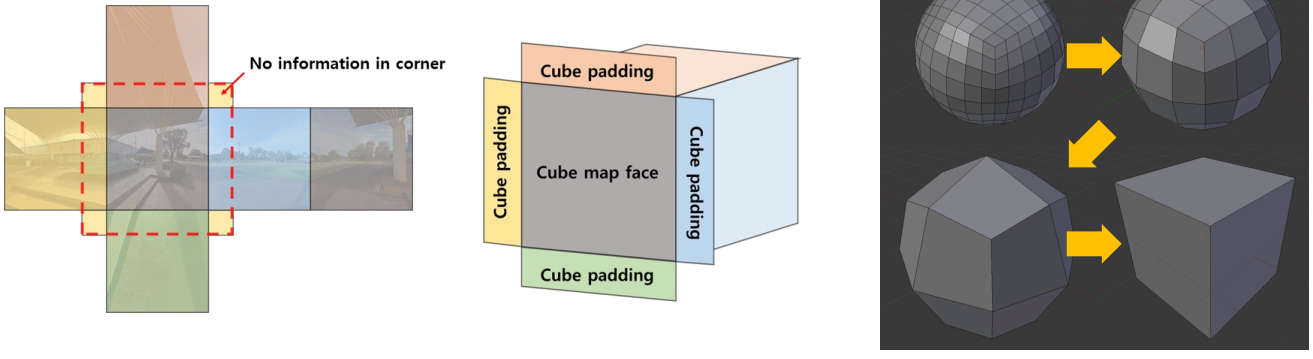
Existing method to reduce the effect of ERP distortion. Split image, analyze image, and gather the result.[2]

[2] Yang, Wenyan, et al. "Object Detection in Equirectangular Panorama." arXiv preprint arXiv:1805.08009 (2018).

Other Representations of 360° Images

Cubemap-based methods

- The CubePadding method pads the information from adjacent faces when it convolves the cube map images.
- Receptive field of convolution can spread over the other faces of the cube.
- But cube map irregularity still makes the task harder.



The concept of cube map convolution. Cube map has no severe distortion near the top and bottom sides.
CubePadding method makes it possible to convolve the cube-shaped images.[3]

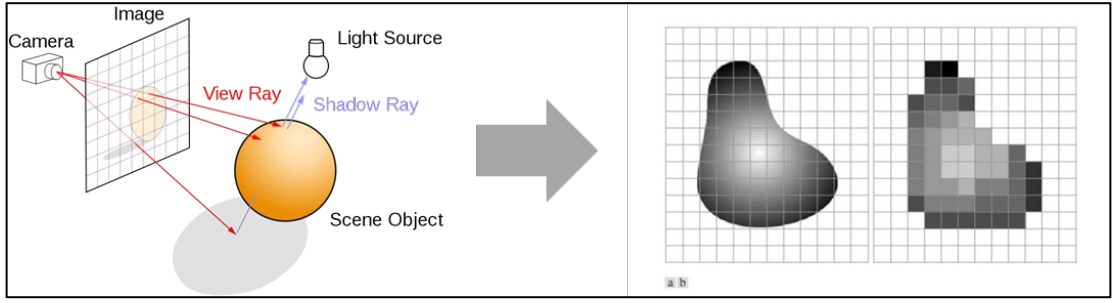
[3] Cheng, Hsien-Tzu, et al. "Cube Padding for Weakly-Supervised Saliency Prediction in 360° Videos." arXiv preprint arXiv:1806.01320 (2018).

Proposed Representation of 360° Images

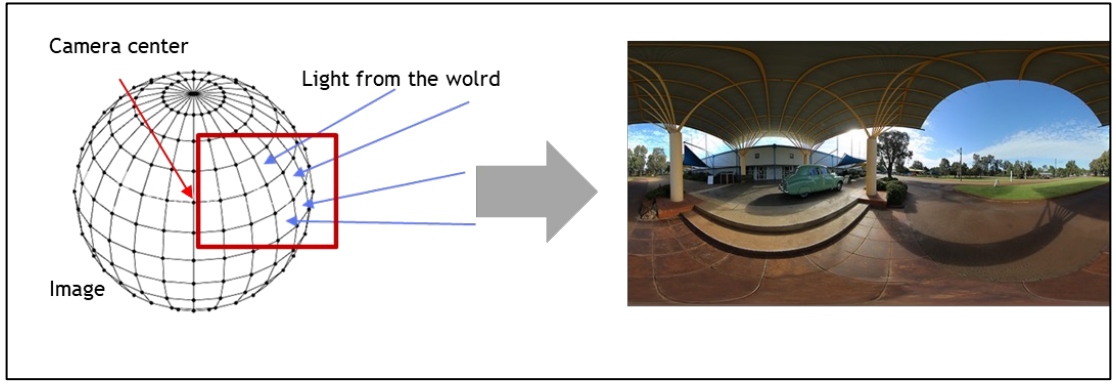
Sampling 360° Information from the world

- Assingning light from the camera's surrounding to a specific data structure.

Conventional normal field of view image (2D array data structure)



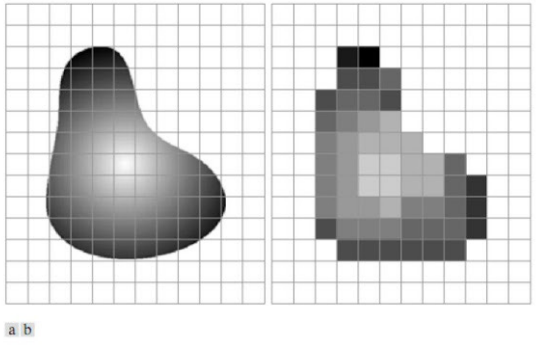
Omni-directional image (3D data structure)



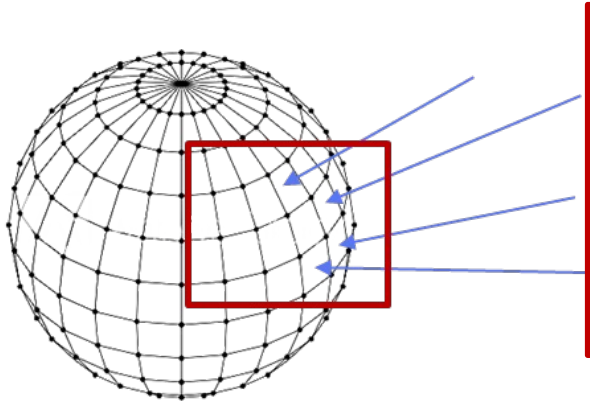
Proposed Representation of 360° Images

The way of representing 360° image: Why mesh structure?

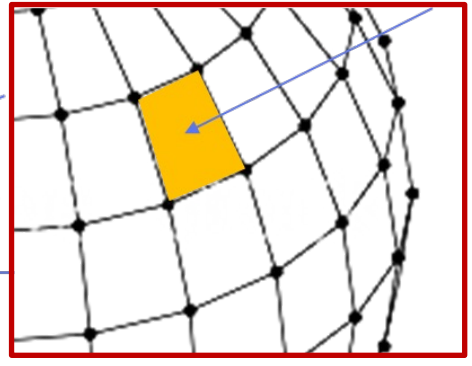
- Light from the world enters a 360° camera from all directions.
- The 360° image must have a spatial sampling method that defines image pixels.
- This spatial sampling method can be represented as a sphere mesh structure.



Sampling on 2D array



Sampling on sphere surface

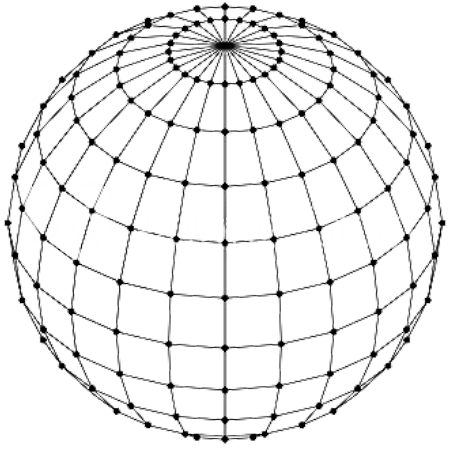


Effective area of light sample

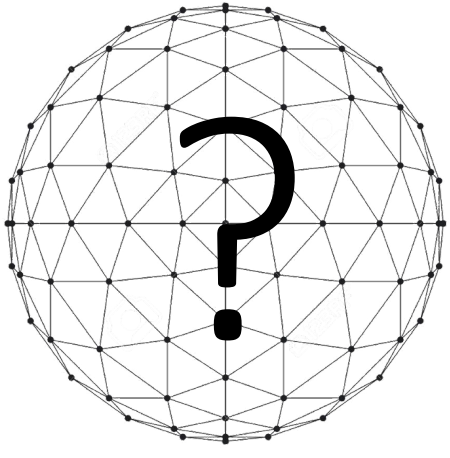
Proposed Representation of 360° Images

What is a good mesh structure to represent 360 image?

- ERP sphere mesh is a proper shape for 2D array data, but it has high irregularity at poles.
- The mesh faces need to have equiareal and equidistant properties, but that is impossible.



ERP sphere mesh

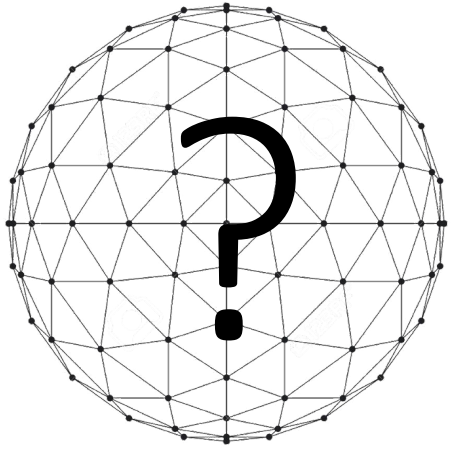


Unknown uniform mesh

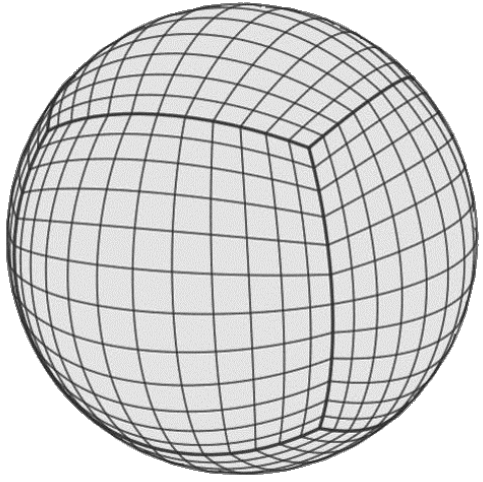
Proposed Representation of 360° Images

Generating a uniform sphere mesh

- Not only equiareal and equidistance properties, but also recursive structure is required.
- The regular polyhedrons satisfy these constraints.
- If we use cubic structure to generate sphere mesh, we could get cube map mesh structure.



Unknown uniform mesh



Cubemap mesh

Proposed Representation of 360° Images

Cube map representation of 360° images

- Cube map represents 360° images with 6-directional views.
- Usually, we set 6 directions as [Top, Bottom, Left, Right ,Front ,Back]
- No sinusoidal fluctuation and severe irregularity around top and bottom sides like in ERP

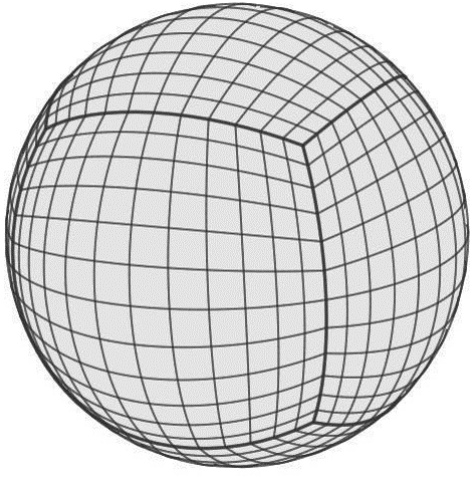
Cubemap image

ERP image

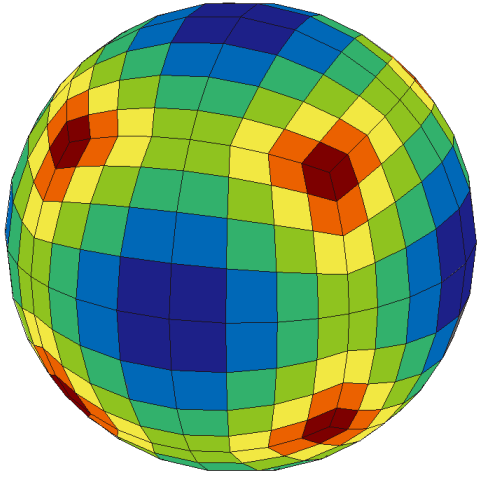
Proposed Representation of 360° Images

Cube map irregularity problem

- Cube map represents images more densely near the edge of cube(different pixel area).
- Such irregularity introduces image distortion similar to the poles of ERP representation.
- Representing 360 images using perfectly uniform mesh structure does not exist!



Cubemap irregularity

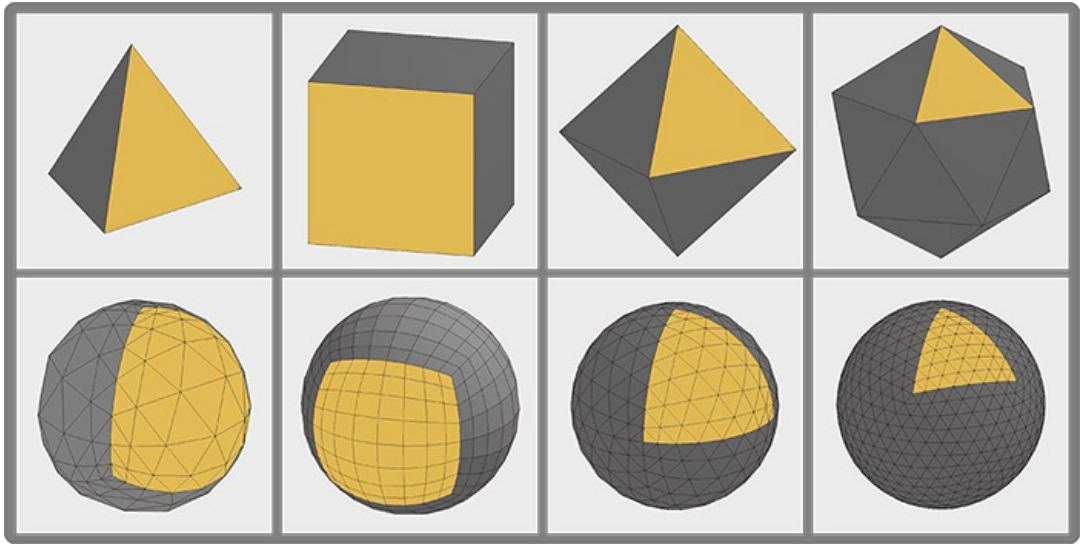


Irregularity heatmap of Cubemap

SpherePHD

Image representation using spherical polyhedrons

- Existing cube map representation has much less irregularity than ERP representation.
- But cube map still has noticeable irregularity near the edge of the cube.
- A spherical polyhedron based on an icosahedron shows very uniform pixels, which means low irregularity.



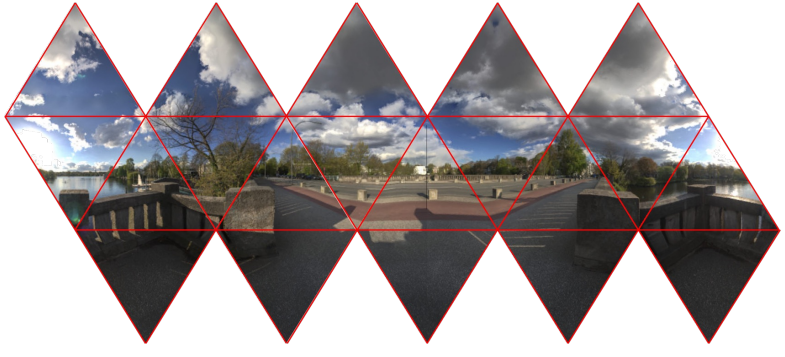
Regular spherical polyhedrons from different original regular polyhedrons

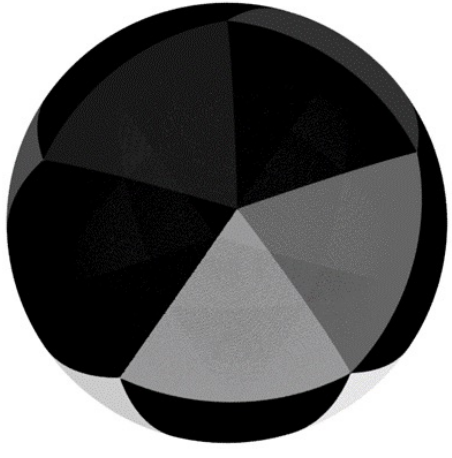
SpherePHD: Applying CNNs on 360° images with non-euclidean spherical polyhedron representation, Yeonkun Lee, Jaeseok Jeong, Jongseob Yun, Wonjune Cho, Kuk-Jin Yoon, TPAMI 2022
SpherePHD: Applying CNNs on a spherical PolyHeDron representation of 360° images, Yeonkun Lee, Jaeseok Jeong, Jongseob Yun, Wonjune Cho, Kuk-Jin Yoon, CVPR 2019

SpherePHD

Icosahedron-based 360° Image Representation (SpherePHD)

- Regular 20-sided polyhedron-based omni-directional image representation
- SpherePHD representation is a set of ray vectors equal to the number of pixels
- The advantage of using the SpherePHD representation is much less irregularity score





SpherePHD

Subdivision and Pooling

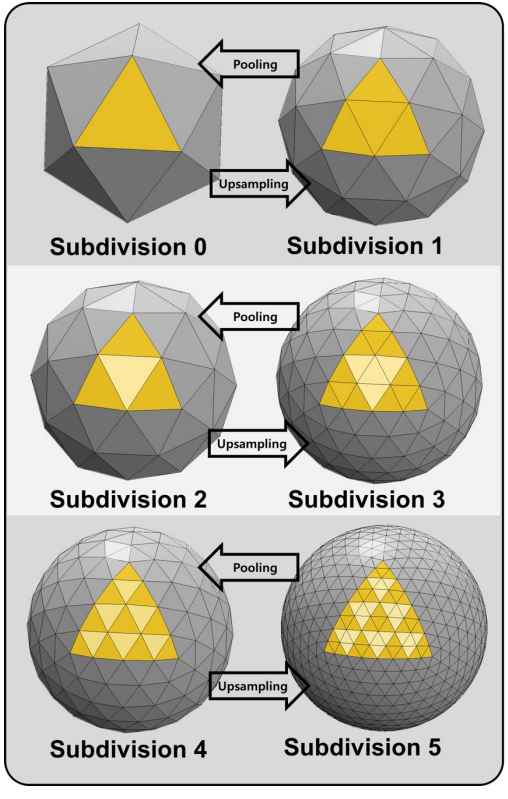
Subdivision: Recursive dividing method for generation of high resolution SpherePHD image

- Start from an icosahedron
- The higher the subdivision, the closer to the sphere shape
- One subdivision step is :

[Step 1] Divide a face into four identical faces

[Step 2] Project new vertices onto unit sphere surface

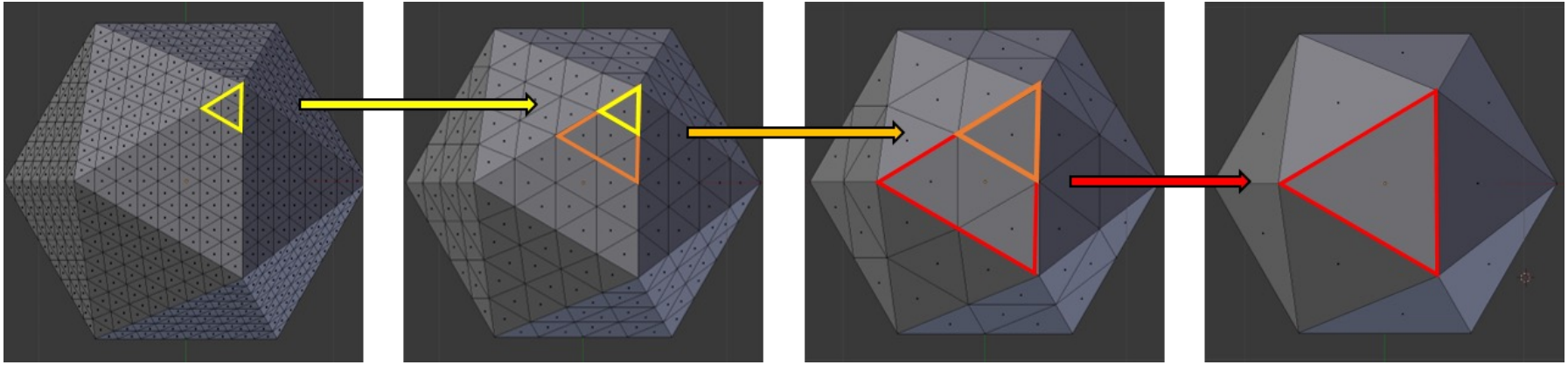
- In subdivision level 8, SpherePHD becomes similar resolution with 1600x800 ERP image.



SpherePHD

Pooling

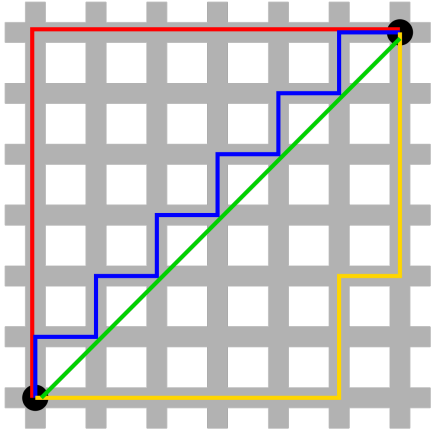
- Pooling process is an inverse process of subdivision.
- For convenience, we only use 1/4 pooling and 4 times subdivision in this work.
- Repeated pooling will eventually make the image to a base icosahedron.

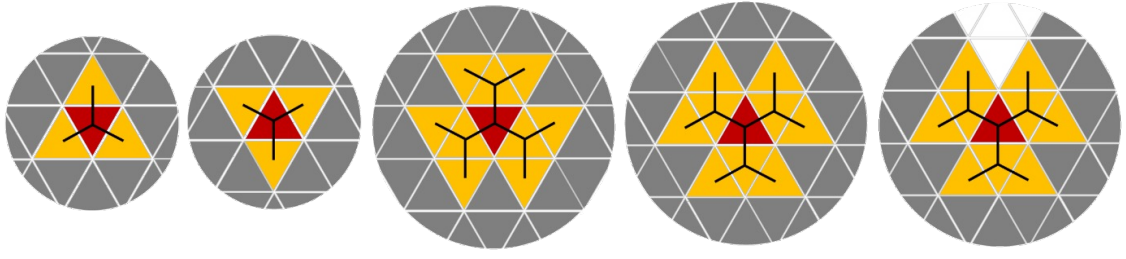


Pooling procedure of 3rd subdivision -> 2nd subdivision -> 1st subdivision -> 0 subdivision (base icosahedron)

Kernel shape design

- A kernel should be reformed in triangular mesh.
- The neighboring pixels of a kernel are selected by Manhattan distance from the center pixel.
- We use only 4-pixel and 10-pixel kernels that are corresponding to traditional 2x2 and 3x3 kernels.





Neighboring pixels in the 4-pixel kernel have 1 Manhattan distance from the center pixel and neighboring pixels in the 10-pixel kernel have 2 Manhattan distance from the center pixel

SpherePHD

Convolution

- Applying a new kernel shape based on triangular-mesh structure

Same radius kernel

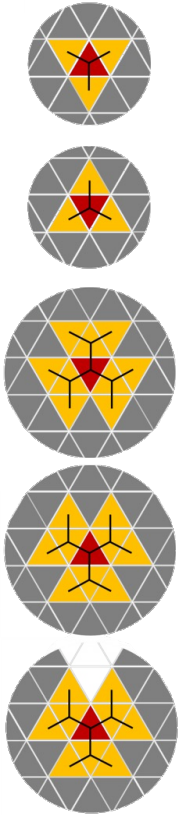
Each kernel element is defined in the same radius region.
The taxicab geometry radius from a center pixel is used.

Reference of other faces

Each element of a kernel can refer other faces of an icosahedron.
Interconnection information graph of an icosahedron is used.

Receptive field

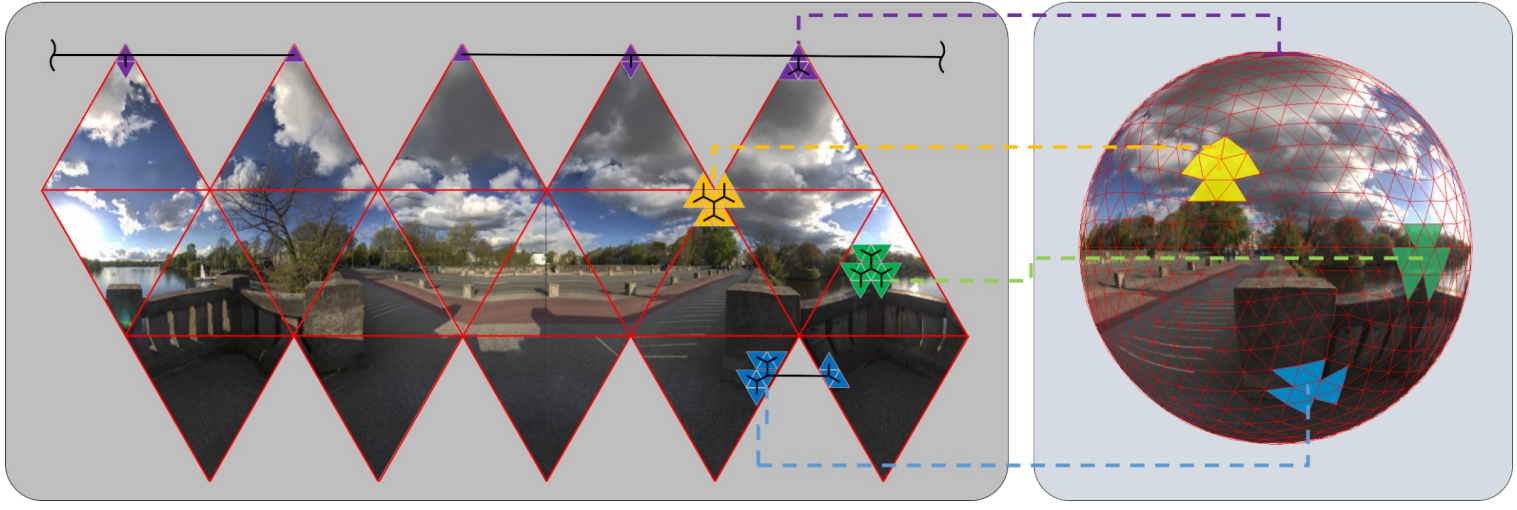
Receptive field can spread over other faces.
Spread receptive field is also defined in the taxicab geometry radius.



SpherePHD

Convolution

- The proposed kernel can refer other faces of an icosahedron.
- The proposed image representation doesn’t need to be padded before the convolution process.
- An image size is perfectly preserved after convolution, which means localization information is also preserved.

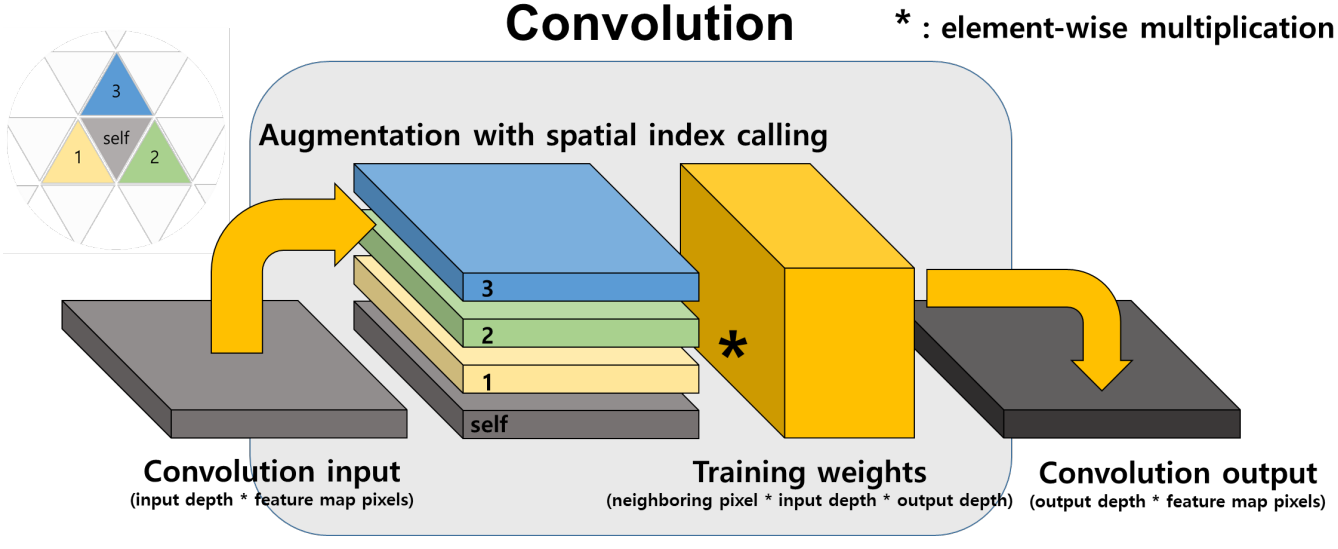


Planar view of our spherical polyhedron image representation and applied kernel shape

Implementation of Convolution for SpherePHD

GPU implementation: Convolutional layer

- All pixels of spherical polyhedron image representation have adjacent pixels information.
- Before convolution, we stack the adjacent pixels into another dimension.
- Then, element-wise multiplication of tensor means convolution.

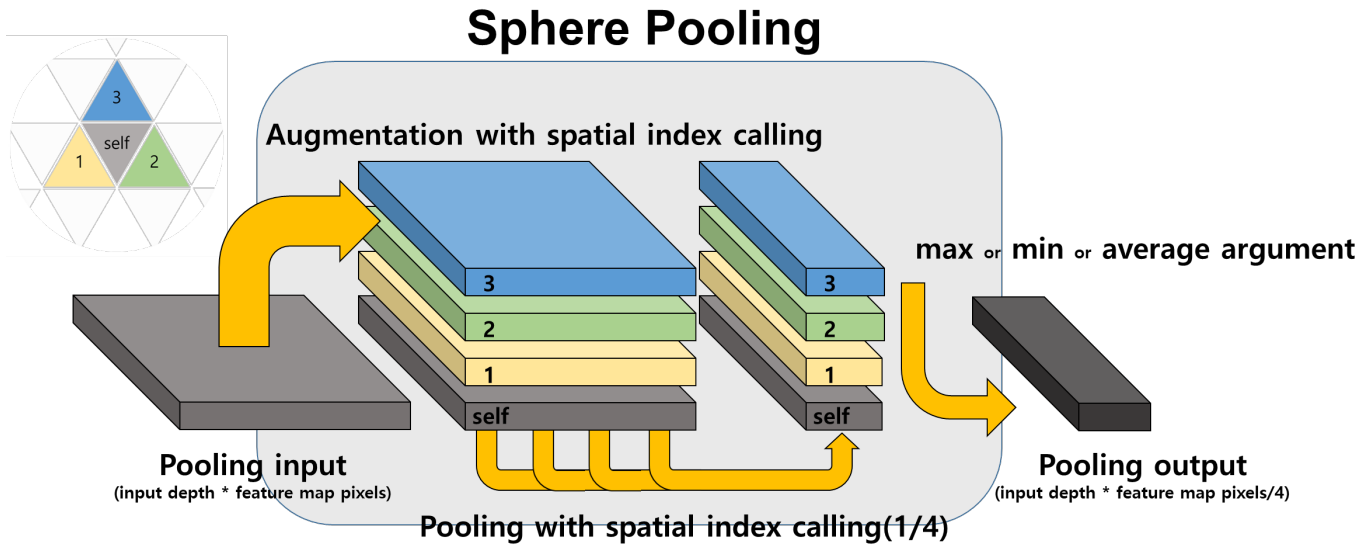


Convolution implementation in general tensor multiplication form

Implementation of Pooling for SpherePHD

GPU implementation: Pooling layer

- Find adjacent pixels information of the center-located pixel of a higher subdivision level.
- Apply the pooling method(max, min, average) and merge four pixels into one pixel.
- After the pooling layer, our image representation has a one-level lower subdivision level.

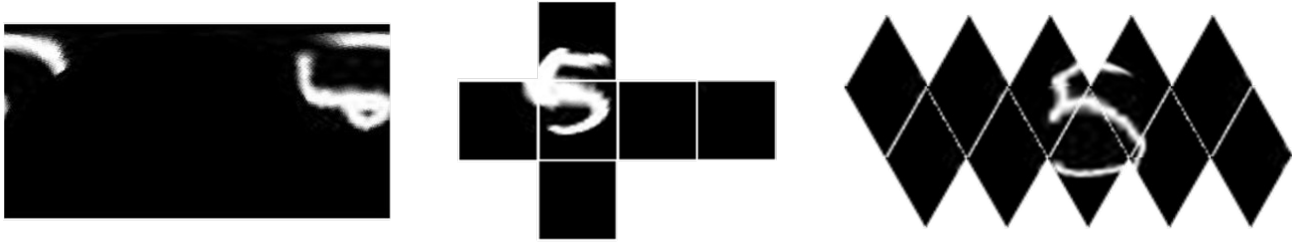


Pooling implementation using center pixel indices of spherical polyhedron representation

Experiments: Random-located MNIST Classification

Objective: Compare different image representation methods, ERP, Cubemap, Spherical polyhedron representation

- Dataset contains random-directionally projected MNIST digits.
- We want to verify the classification accuracy for different locations in terms of longitude and latitude.
- We use the shallow network with a fully convolution layer to equalize parameter usage.



(a) ERP image (b) Cubemap image (c) Spherical polyhedron image

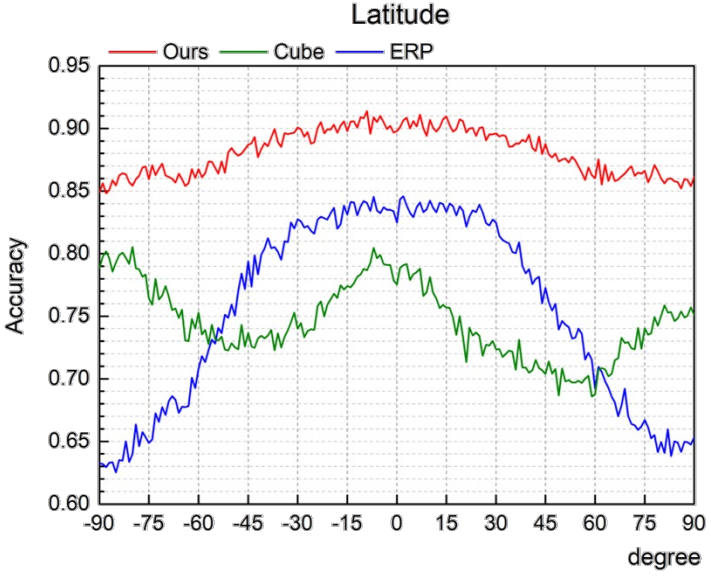
Random directional MNIST dataset of ERP, Cubemap, Spherical polyhedron representation

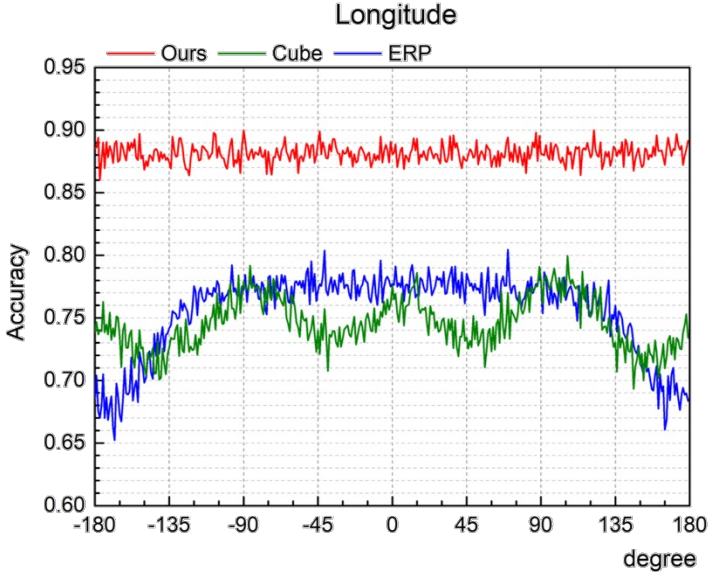
We augmented training data from 60k to 1200k with random directional projection

Results: Random-located MNIST Classification

Classification accuracy

- Classification accuracy of random location MNIST digits
- Longitude and latitude of projected digits affect classification accuracy.
- Accuracy variation is related with irregularity of input image representation.

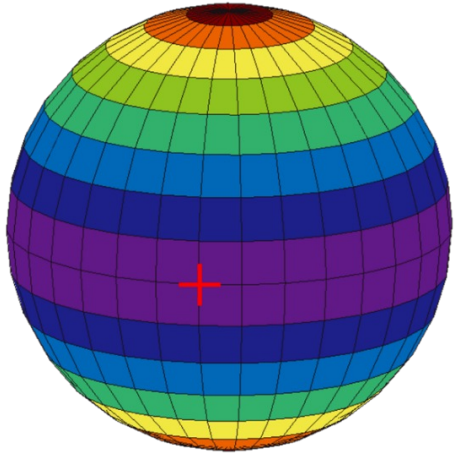


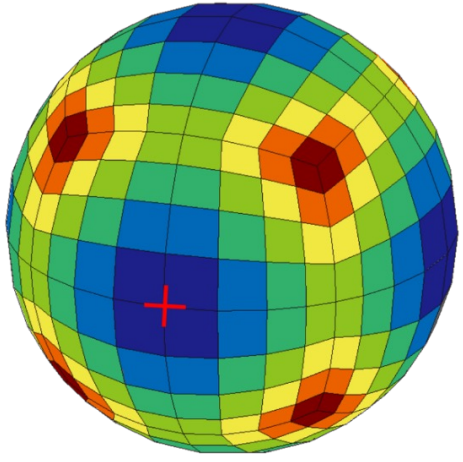


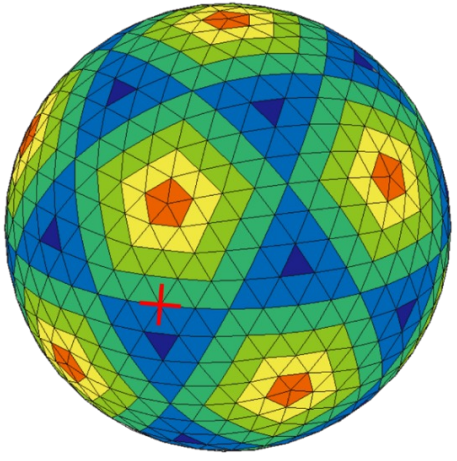
Results: Random-located MNIST Classification

Irregularity of input image representation

- The ERP image has high irregularity in polar regions(±90 latitude) and has discontinuity in the side of image (±180 longitude).
- The cubemap image has high irregularity in ±45 latitude regions and ±45, ±135 longitude regions.
- The spherical polyhedron image has irregularity in ±90 latitude regions.
- The effect of irregularity is much less in the spherical polyhedron representation.



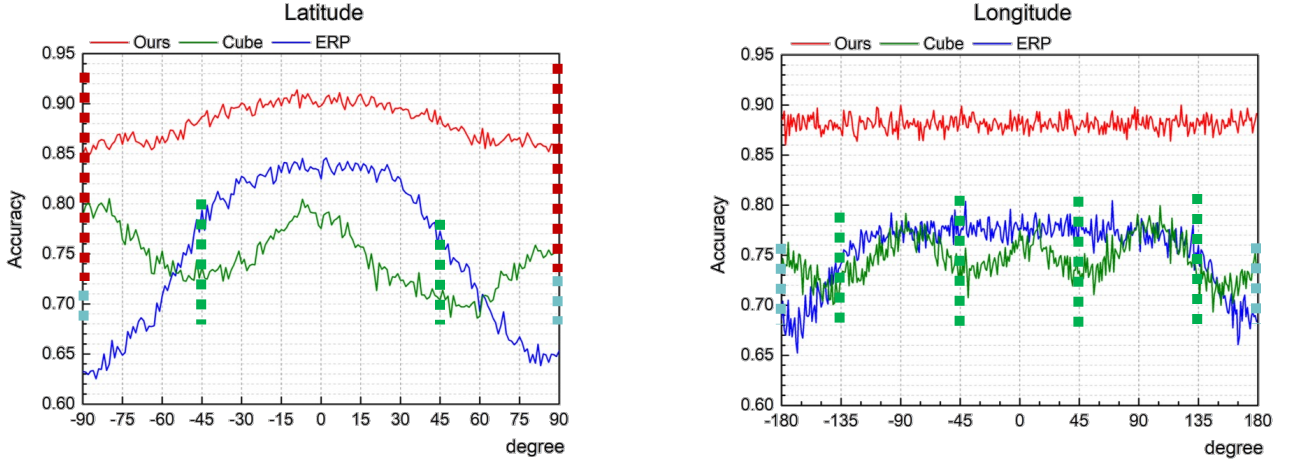




Results: Random-located MNIST Classification

Irregularity of input image representation

- The ERP image has high irregularity in polar regions(±90 latitude) and has discontinuity in the side of image (±180 longitude).
- The cubemap image has high irregularity in ±45 latitude regions and ±45, ±135 longitude regions.
- The spherical polyhedron image has irregularity in ±90 latitude regions.
- The effect of irregularity is much less in the spherical polyhedron representation.



Accuracy variation is highly related with the irregularity of the representation method

	SpherePHD	SphereNet [5]	SphericalCNN [4]
MNIST classification	88.13%	82.79%	87.57%

Experiments: SYNTHIA Vehicle Detection

Detection task with YOLO-like structure and YOLO-like loss function

- Show applicability of the proposed method to solve more complicated tasks, i.e., detection.

Tilted SYNTHIA dataset

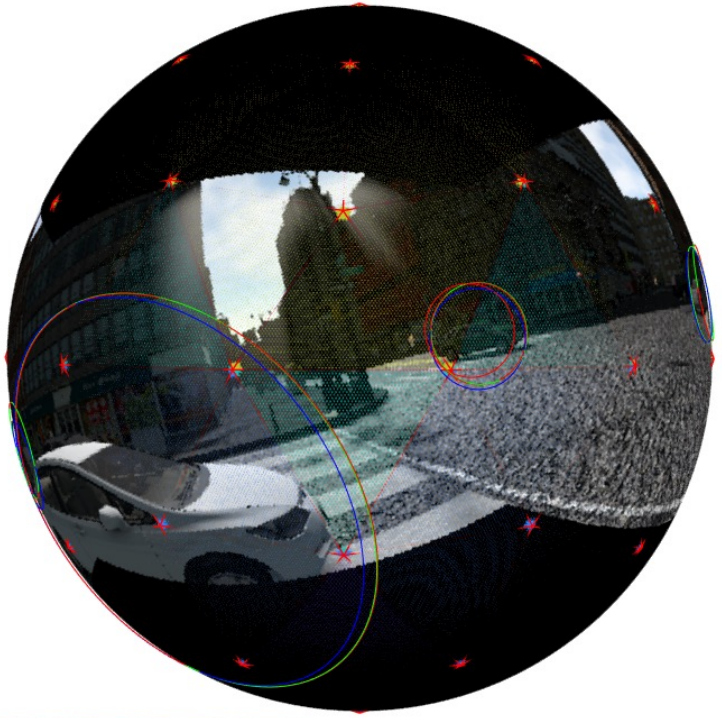
SYNTHIA is the synthetic driving sequence dataset. We tilt this dataset and annotated a car location in a circle. We made ERP and Spherical polyhedron versions of datasets.

YOLO-like structure

We used the YOLO-like structure and YOLO-like loss, which is more fit to the spherical polyhedron and ERP image.

Average precision of car detection

In network using ERP average precision is 39.87% and proposed method shows 64.52%

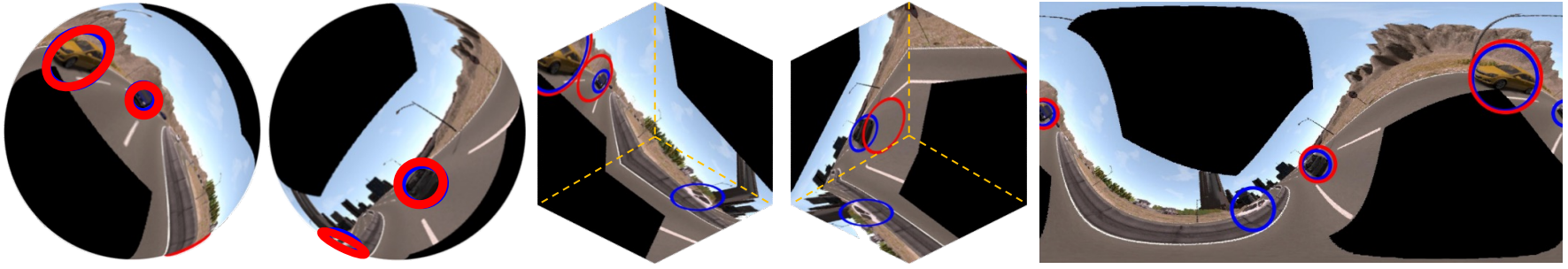


Results: SYNTHIA Vehicle Detection

Comparison with other 360° image representation methods

- In case of training with no rotated data, an ERP method shows better performance.
- In case of training with rotation-augmented data(tilted data), the ERP shows severely degraded performance.
- The proposed SpherePHD method shows better performance for general 360° images.

Red circle: Prediction, Blue circle: GT



		SpherePHD	ERP	Cubemap
Vehicle detection (avg. precision)	SYNTHIA(no rotation)	43.00	56.04	30.13
	SYNTHIA(rot.-augmented)	64.52	39.87	26.03

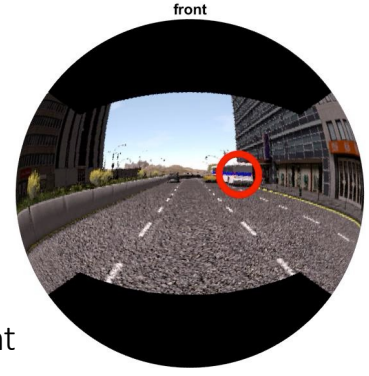
SYNTHIA vehicle detection performance depends on 360° image representation methods (%)

[4] Ros, German, et al. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." CVPR. 2016.

Results: SYNTHIA Vehicle Detection

Detection task with YOLO-like structure and YOLO-like loss function

Front



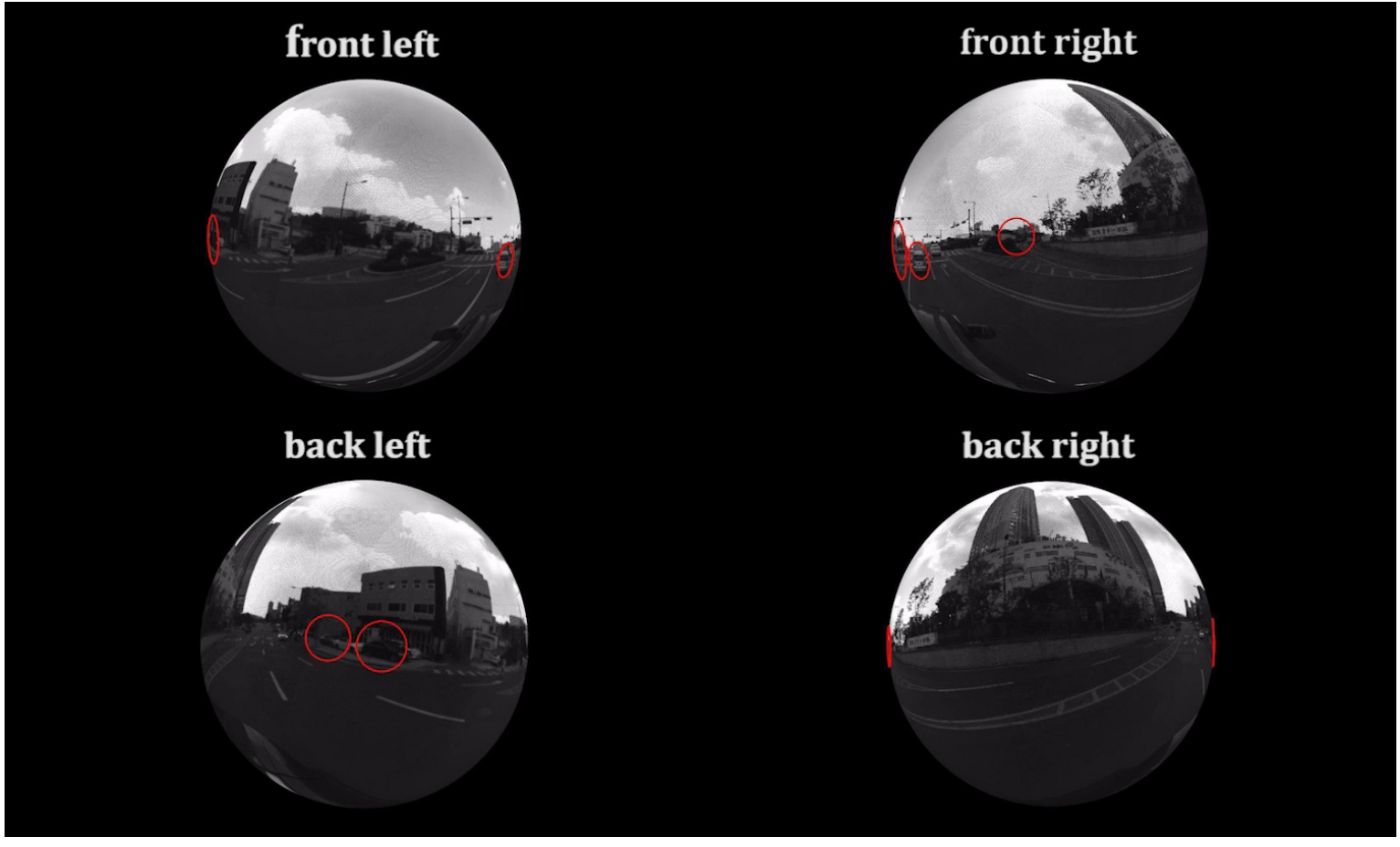
Four side views of tilted SYNTHIA
Red circle: detection result

Left

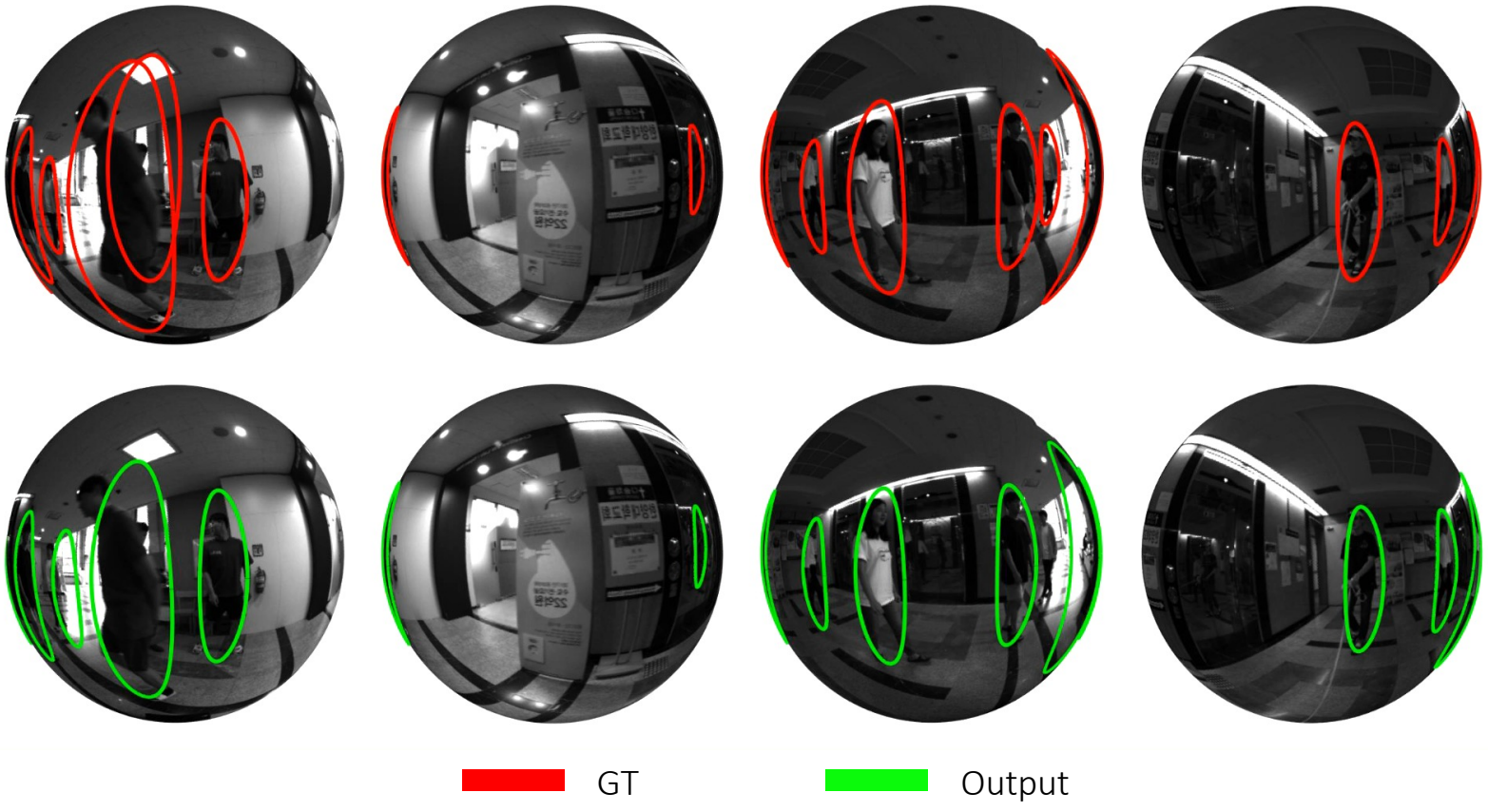
Back

Right

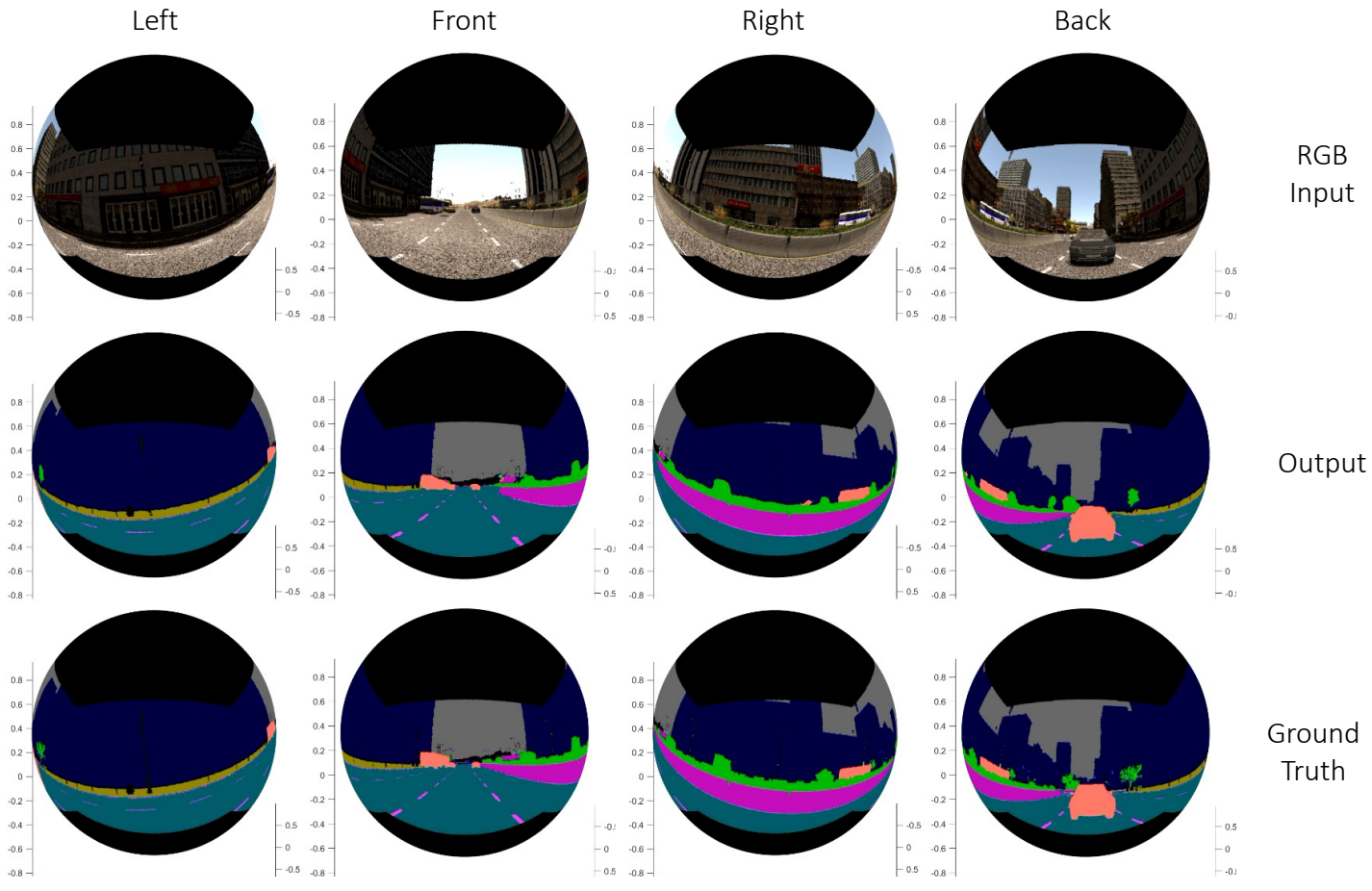
Results: Hanyang Real-world Vehicle Detection



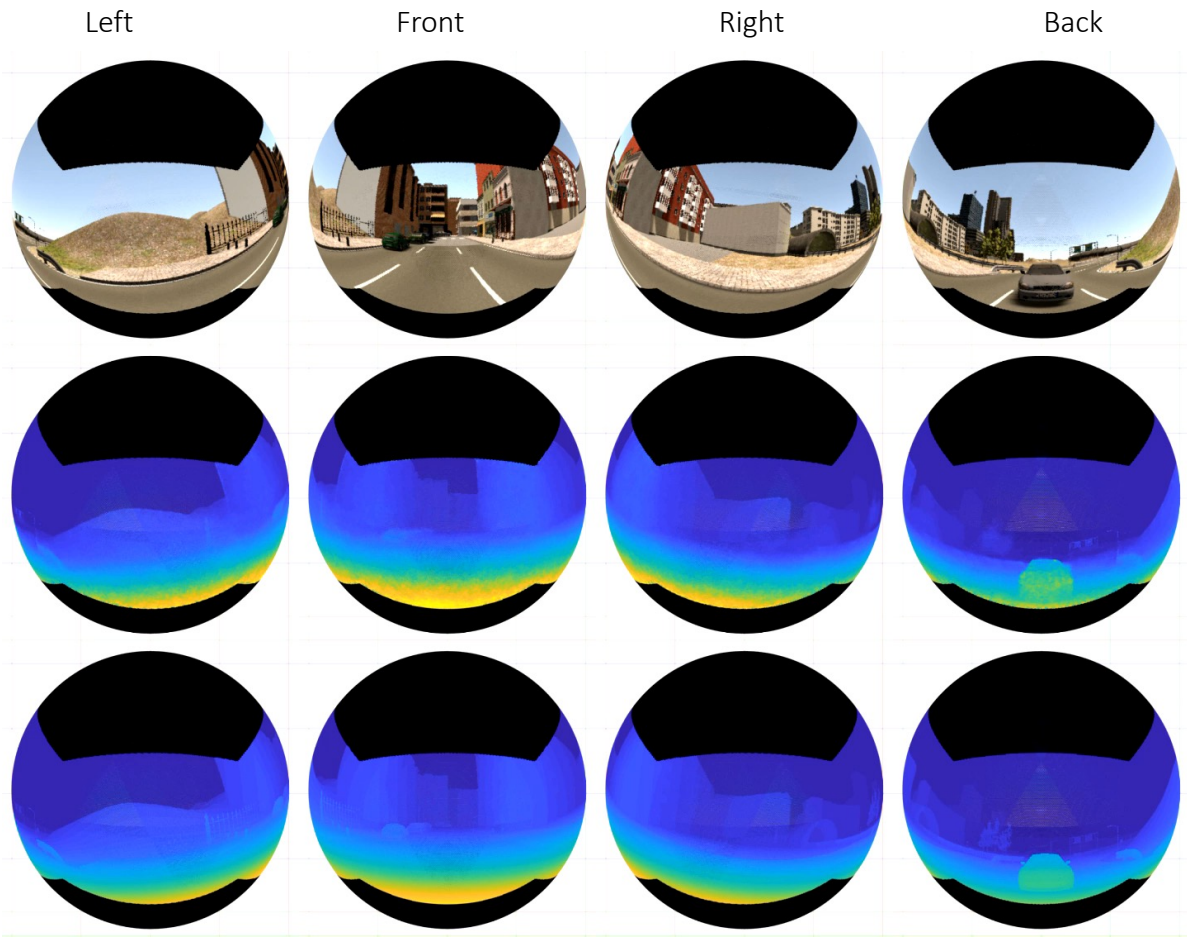
Results: Hanyang Real-world Human Detection



Results: SYNTHIA Semantic Segmentation



Results: SYNTHIA Depth Estimation



MORE APPLICATIONS USING OMNIDIRECTIONAL CAMERAS

(1) Saliency Estimation and Viewpoint Selection for Automatic Video Generation

(2) Perspective and 360°-image-based Stereo-Monocular Depth Estimation

1) Wide Field of View Information

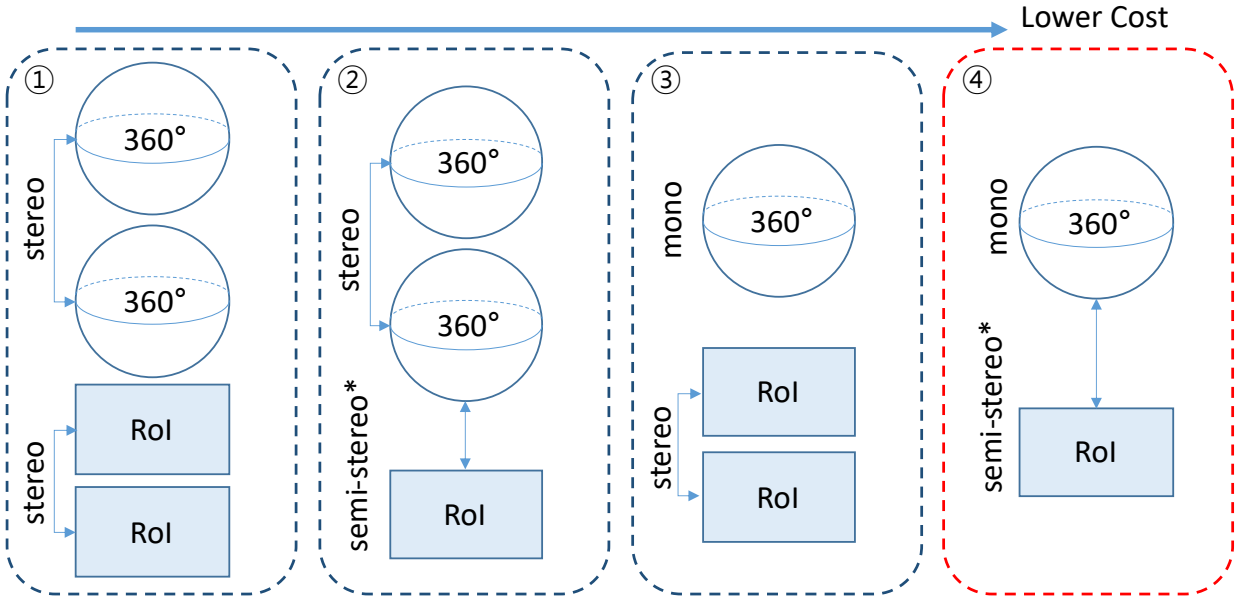
: 360° image can provide omnidirectional information

2) Clear Information in details

: With in narrow field of view image, detail information can be preserved

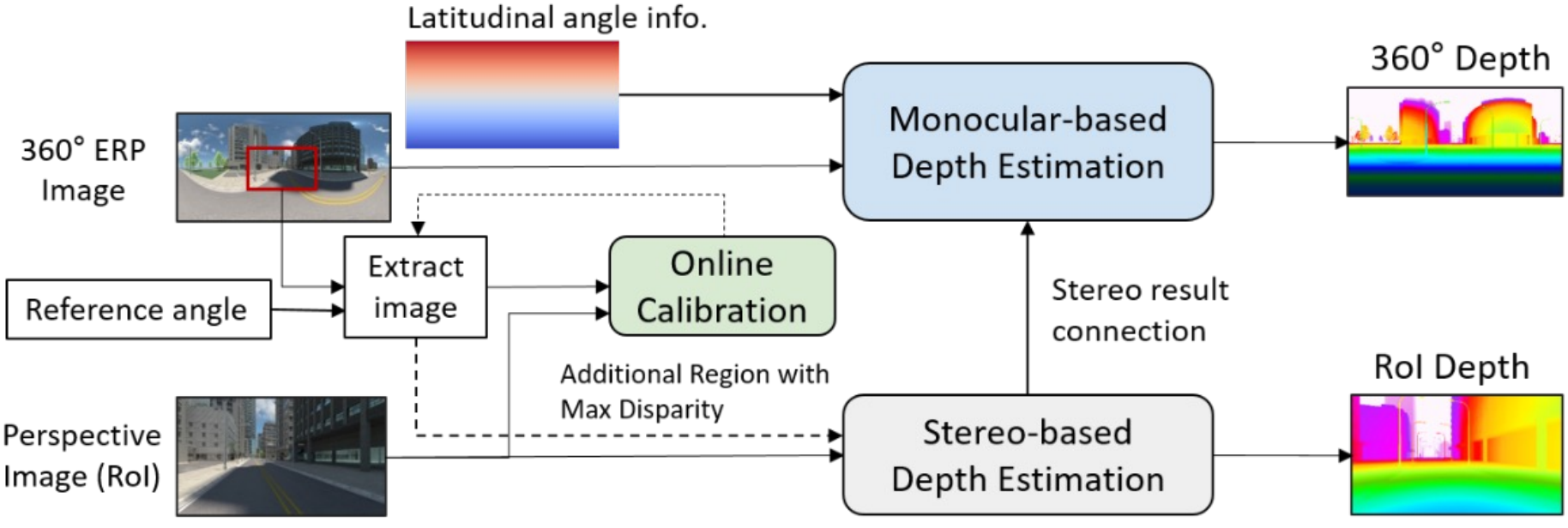
Possible Camera Configurations

- With stereo or monocular solution of depth, 4 configurations of system is introduced



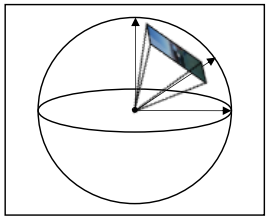
* : stereo with extracted image from 360° image

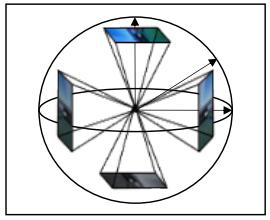
(2) Perspective and 360° ERP image-based Stereo-Monocular Depth Estimation

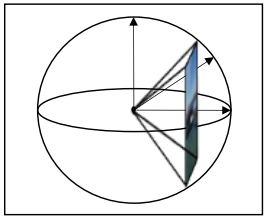


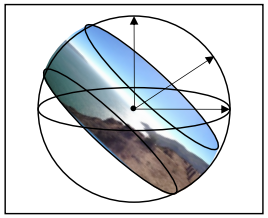
(3) RGBD Panorama Synthesis with Limited RGB and Depth Information

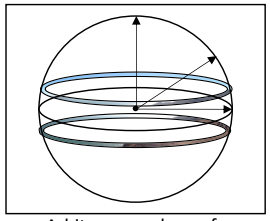
Synthesize RGBD panorama, conditioned on images from NFoV cameras and depth information from mobile depth sensors in various sensor configurations.

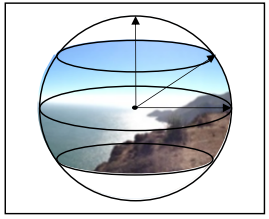

Arbitrary camera orientation

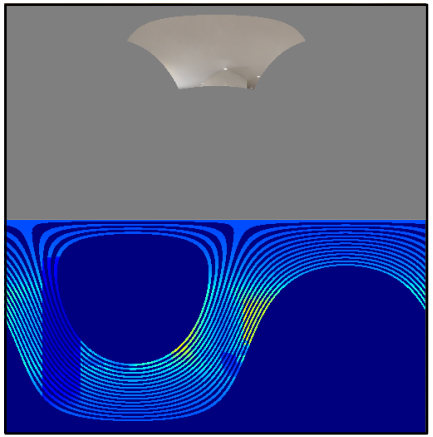

Arbitrary numbers of viewpoints

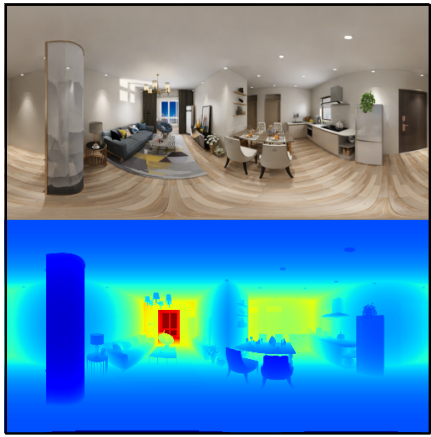

Arbitrary camera field-of-view

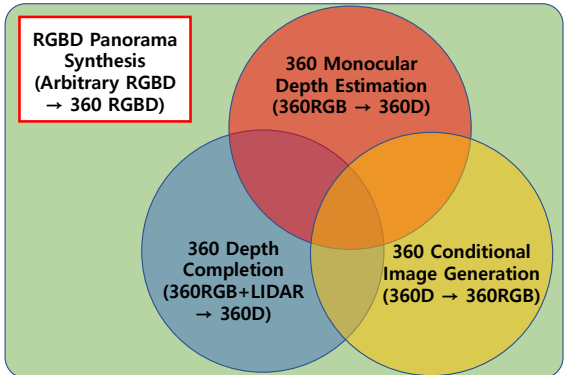

Arbitrary LIDAR orientation


Arbitrary numbers of LIDAR channels


Arbitrary LIDAR field-of-view


RGBD information


RGBD panorama

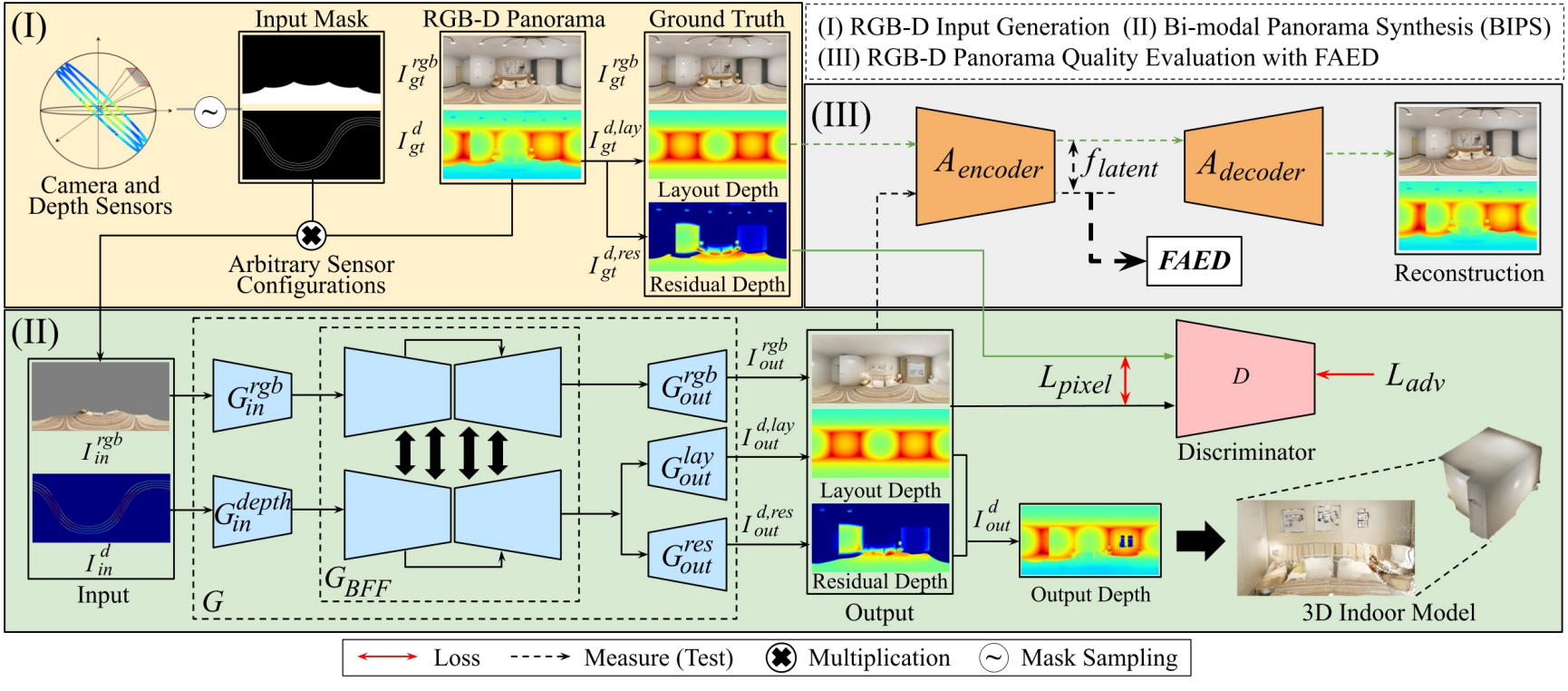


(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning

(Ⅰ) RGB-D input generation with various sensor configurations

(Ⅱ) Bi-modal panorama synthesis with bi-modal feature fusion generator and residual depth-aided learning

(Ⅲ) RGB-D panorama quality evaluation with FAED

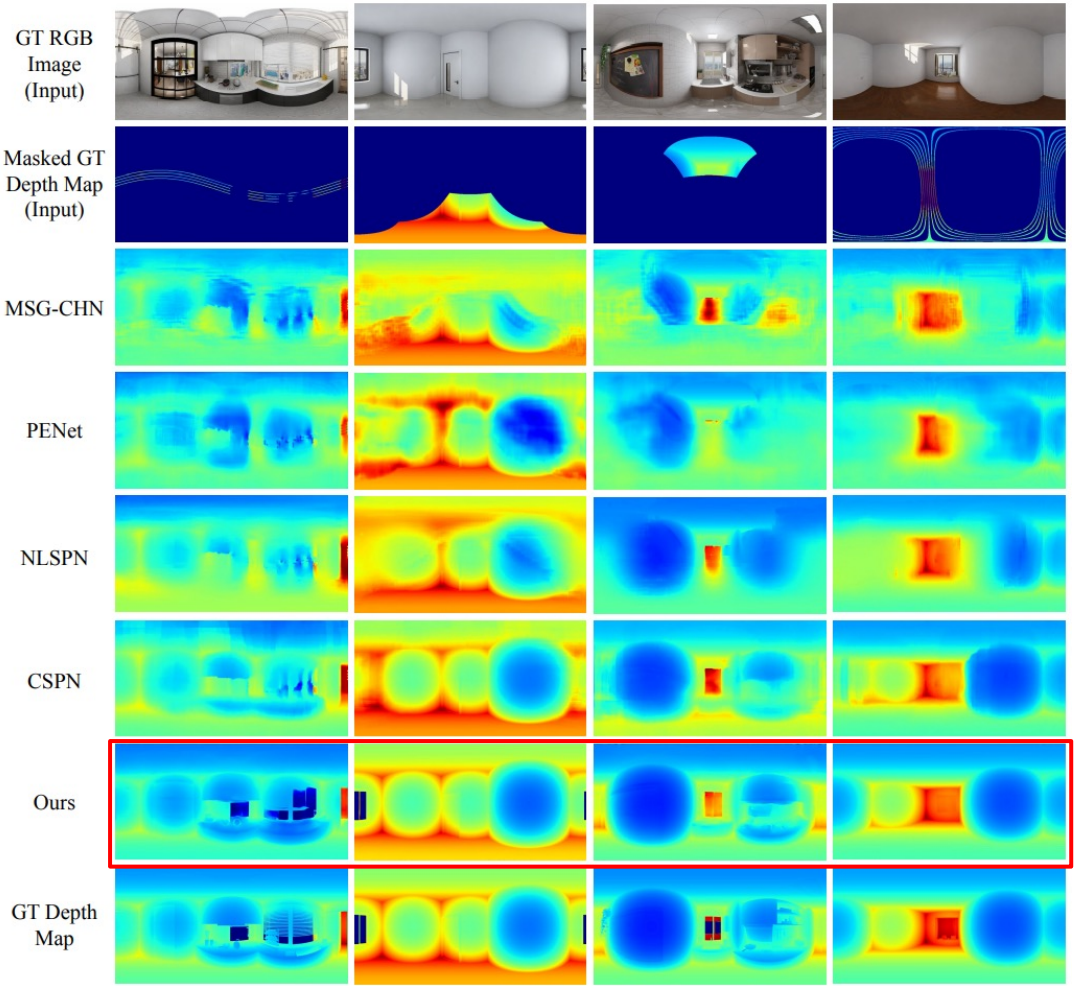


(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning

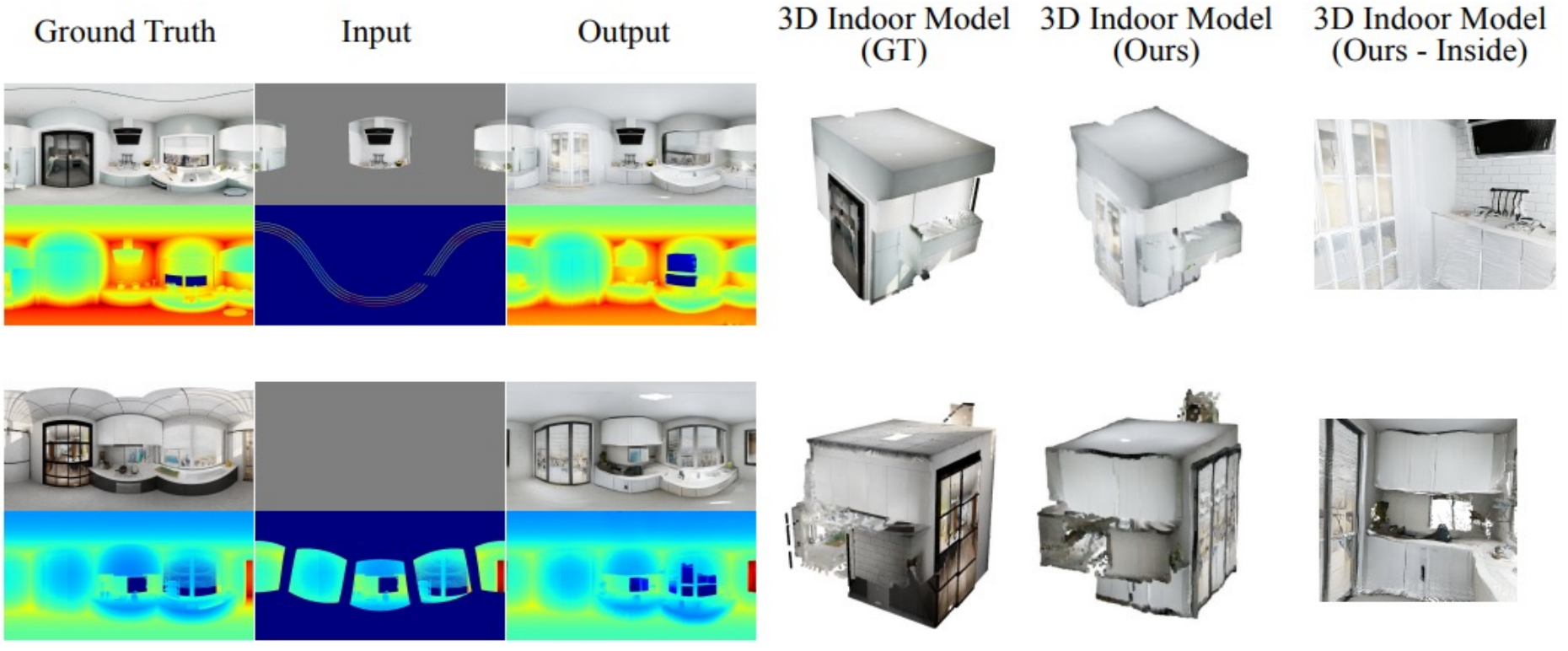
RGB Panorama Synthesis Results

(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning

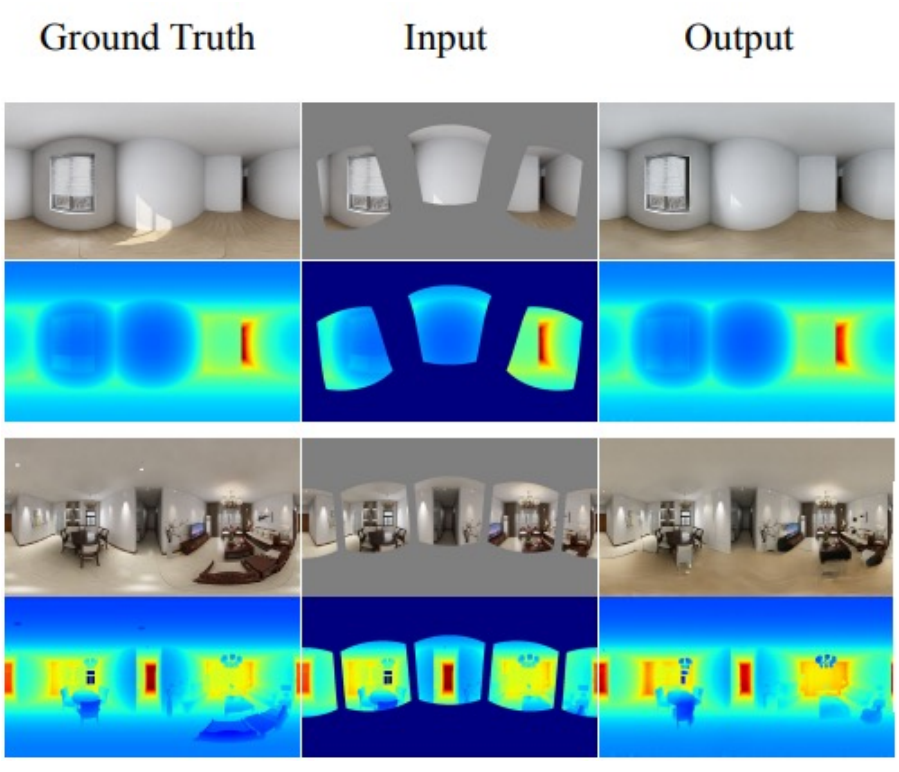
Depth Panorama Synthesis Results



(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning



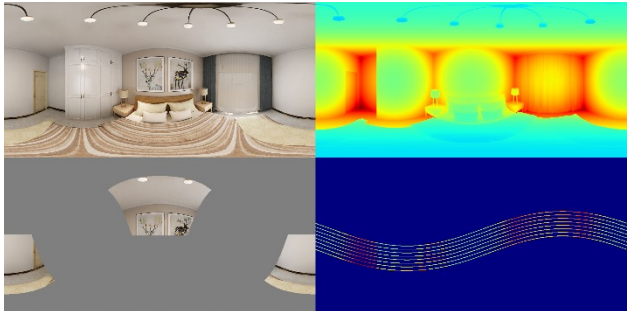
(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning



(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning

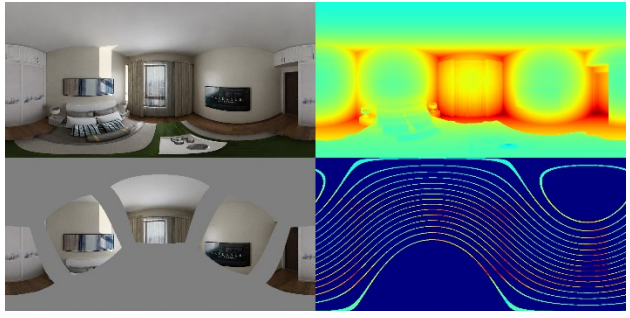
GT

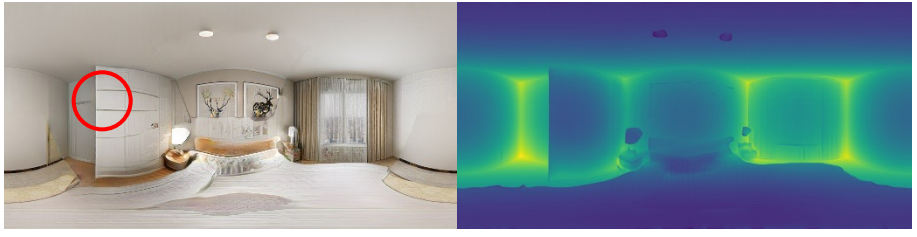
Input

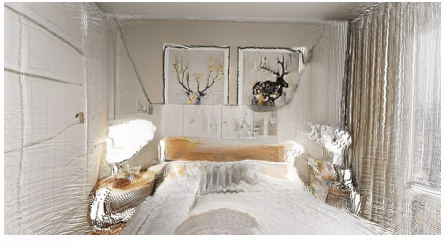


GT

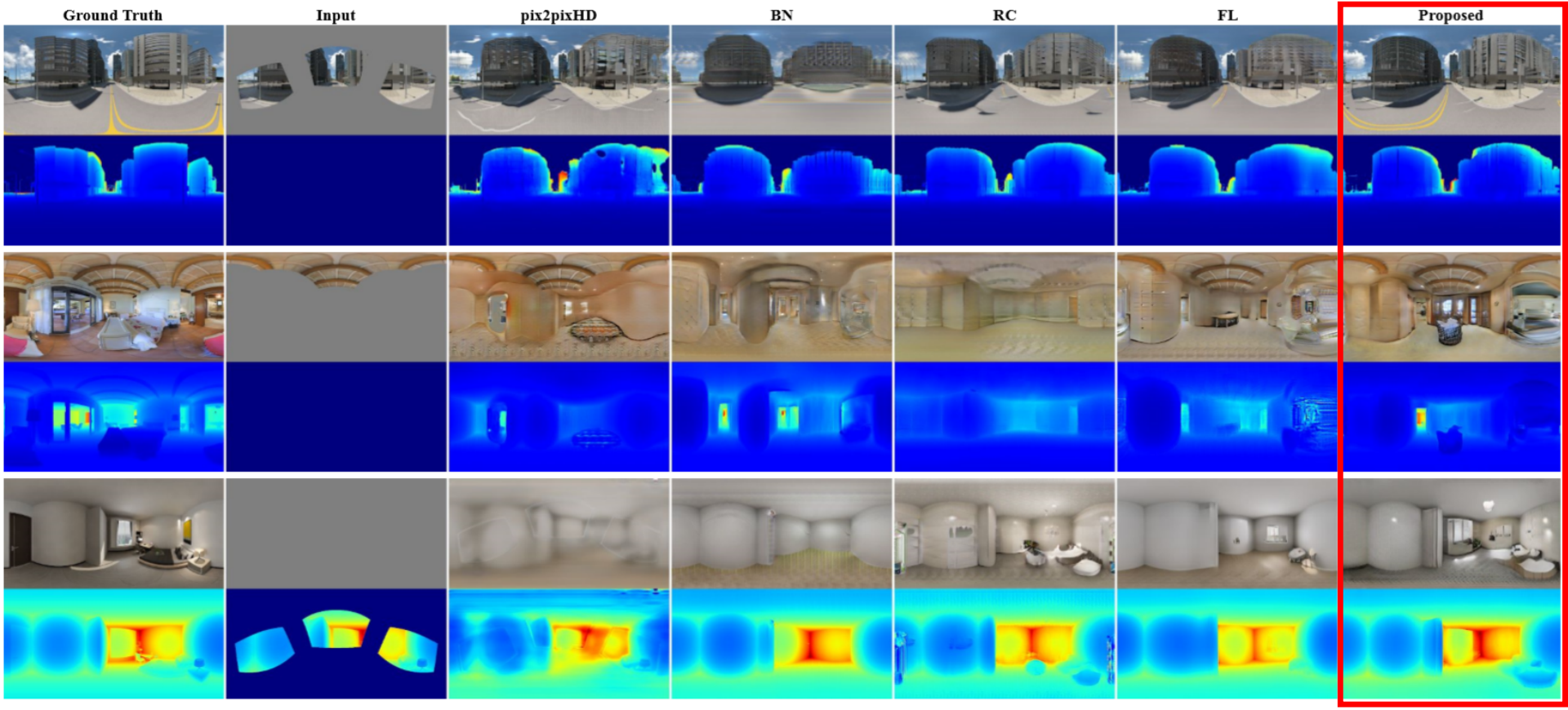
Input







(3) Bi-modal Indoor Panorama Synthesis via Residual Depth-aided Adversarial Learning



(4) SphereSR: 360° Image Super-Resolution

Panoramic images & Super-Resolution

- Panoramic images have the advantage of being able to contain information from all directions in one image, so they are practically used in various fields of real life.
- However, since panoramic images have a wide field of view (FOV) within a limited resolution, they contain more information compared to conventional perspective images.
- Therefore, Super-Resolution is an essential task for panoramic images that contain a lot of information for a limited resolution.

Road environment drone shooting Image
(Fisheye)

Indoor environment CCTV image
(Fisheye)

Google Street View image
(ERP image)

SphereSR: 360◦ Image Super-Resolution with Arbitrary Projection via Continuous Spherical Image Representation, Youngho Yoon, Inchul Chung, Lin Wang, Kuk-Jin Yoon, CVPR 2022 (oral)

(4) SphereSR: 360° Image Super-Resolution

Various projection types of panoramic images

- Currently, studies on super-resolution of panorama image are mostly conducted on Super-Resolution for ERP image, which is one of various types of panorama image.
- The final output is usually an ERP image, a network that can generate panoramic image formats with various expression methods has not yet been developed.
- For this reason, bilinear/bicubic sampling must be used to convert the original panoramic image to another image format.

Wield FOV perspective Image

Fisheye Image

Stereographic Projection

Equi-rectangular Image

https://www.360cities.net/image/cole-militaire-champ-de-mars-eiffel-tower-paris-france-iii

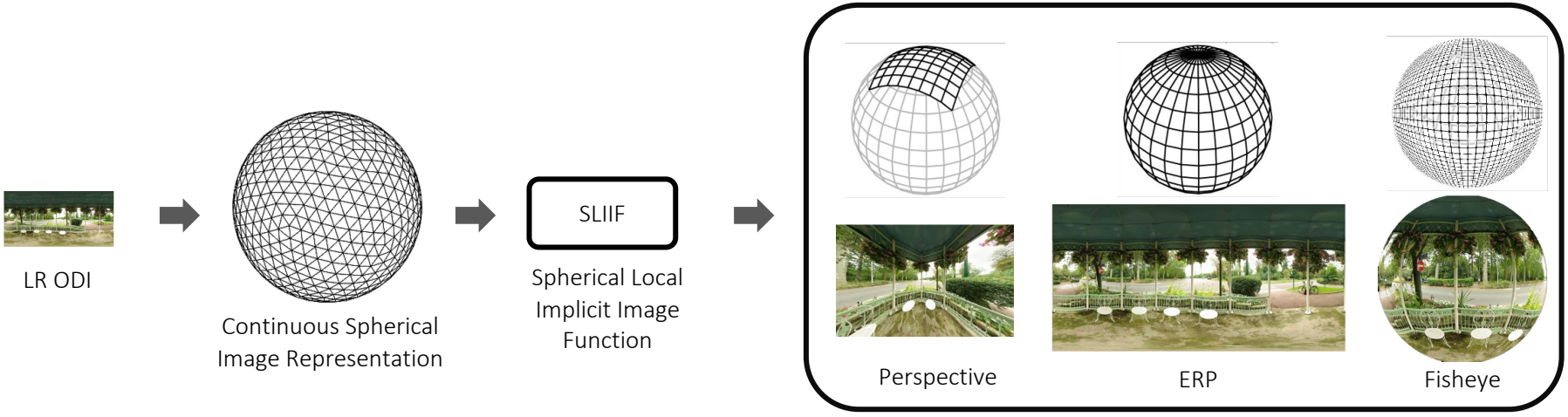
(4) SphereSR: 360° Image Super-Resolution

Key Ideas

- We propose a novel framework with the goal of super resolving an LR 360° image to an HR image with an arbitrary projection type via continuous spherical image representation.

- By using the spherical image representation, we can extract features efficiently on a spherical surface composed of uniform faces.

- By introducing the continuous image representation on the unit sphere, we can convert spherical images flexibly into various projection types (ERP, perspective, and fisheye projection).



Super resolved ODI with Arbitrary Projection Type and Scale

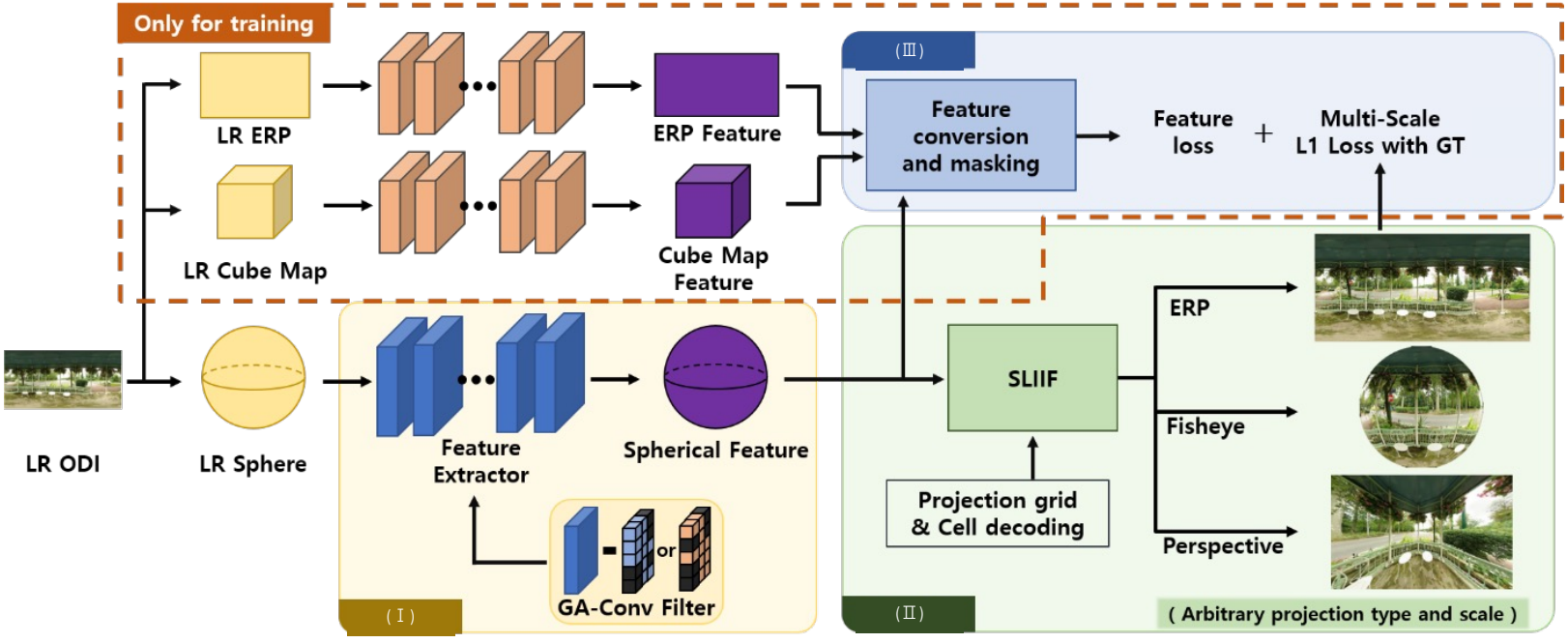
(4) SphereSR: 360° Image Super-Resolution

SphereSR framework

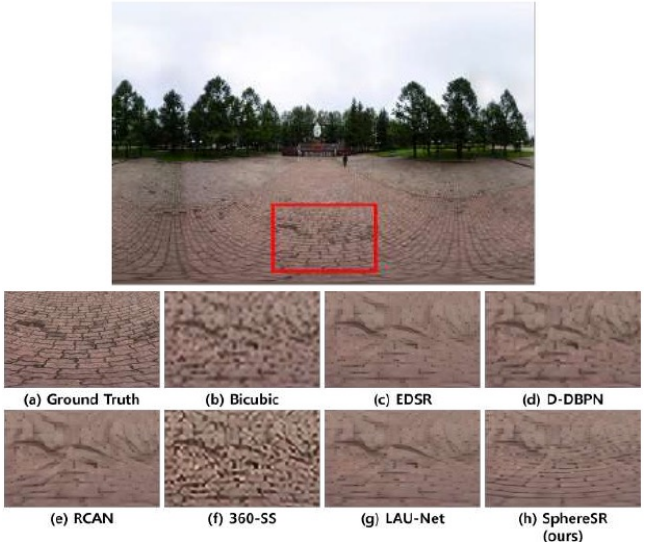
(Ⅰ) Feature extraction Method for spherical images

(Ⅱ) Spherical Local Implicit Image Function(SLIIF) which predicts RGB values through the extracted features

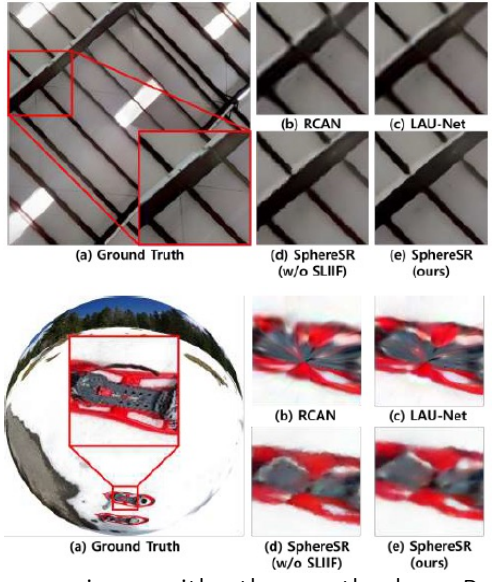
(Ⅲ) Feature Loss to obtain support from features of other projection types (ERP & Cube Map)



(4) SphereSR: 360° Image Super-Resolution



Qualitative comparisons with other methods on ERP projection

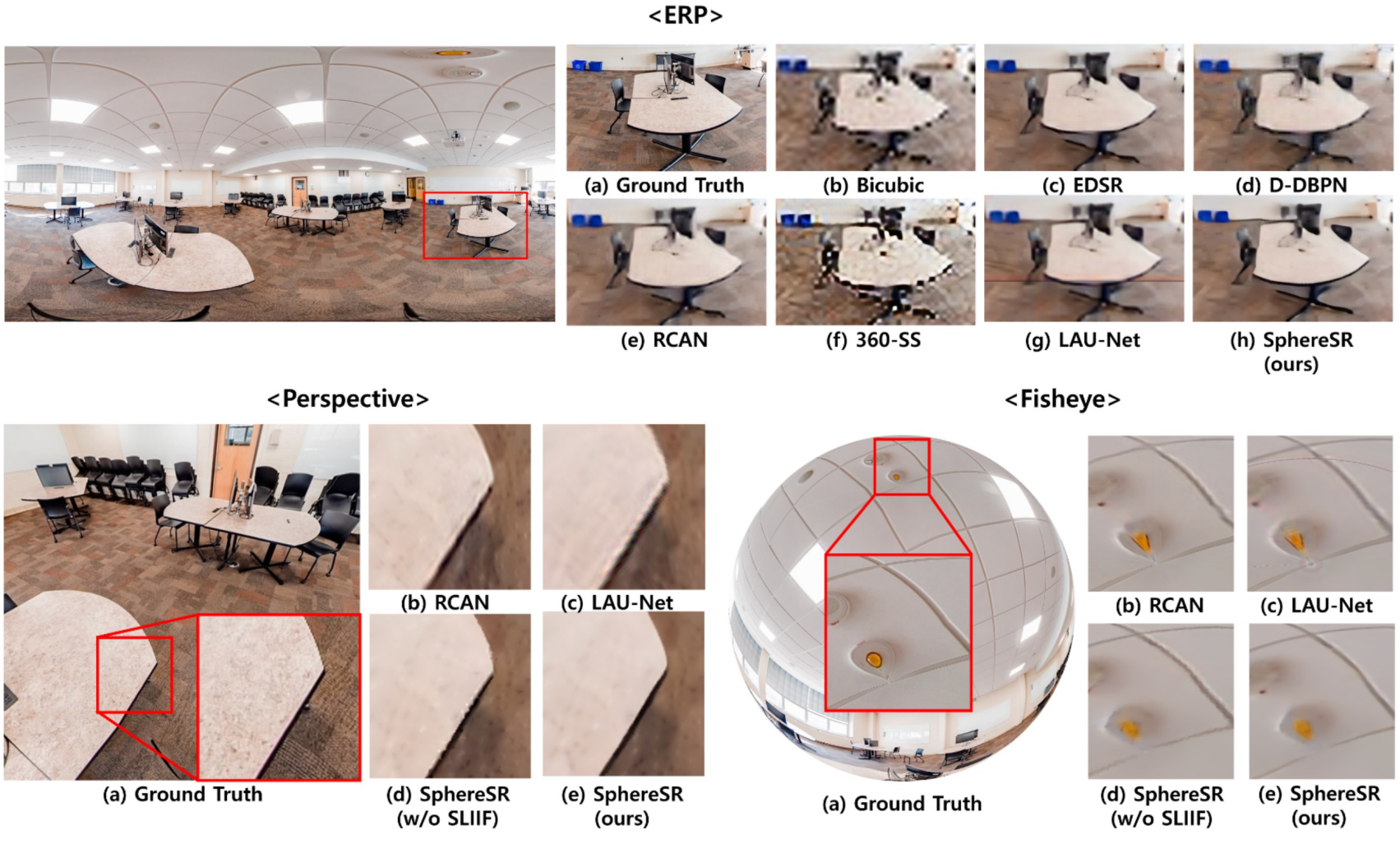


Qualitative comparisons with other methods on Perspective and Fisheye projections

Qualitative Results

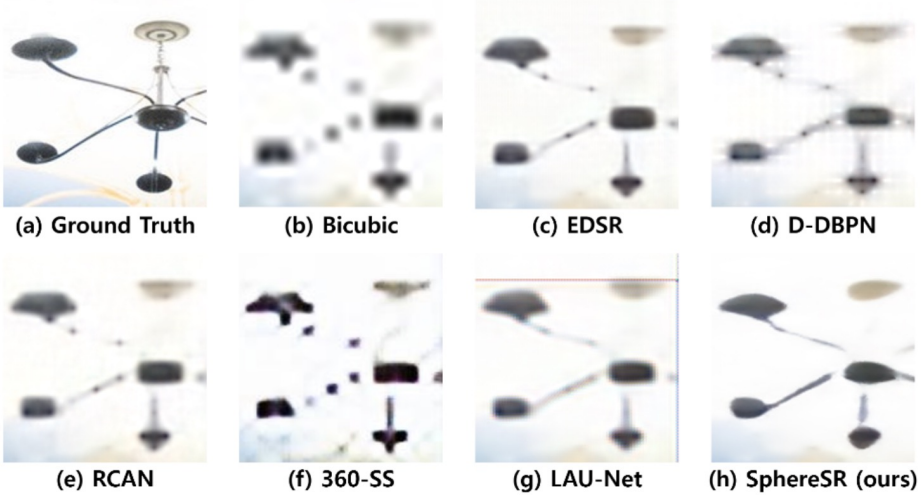
- In ERP projection results, SphereSR reconstructs clear textures and more accurate structures, while other compared methods suffer from the problems of blurred edges or distorted structures.
- In perspective projection results, SphereSR reconstructs clear straight lines and textures rather than other compared methods.
- In fisheye projection results, other methods generate inappropriate textures with several lines rushing to the south pole. On the other hand, SphereSR doesn't make this problem.

(4) SphereSR: 360° Image Super-Resolution



(4) SphereSR: 360° Image Super-Resolution

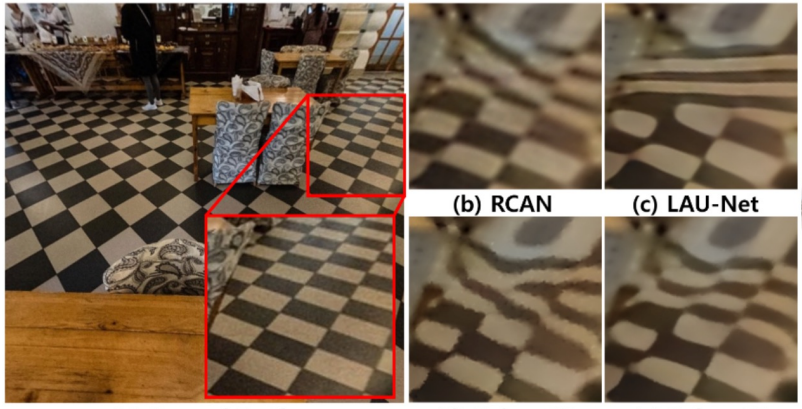
<ERP>



(a) Ground Truth (b) Bicubic (c) EDSR (d) D-DBPN

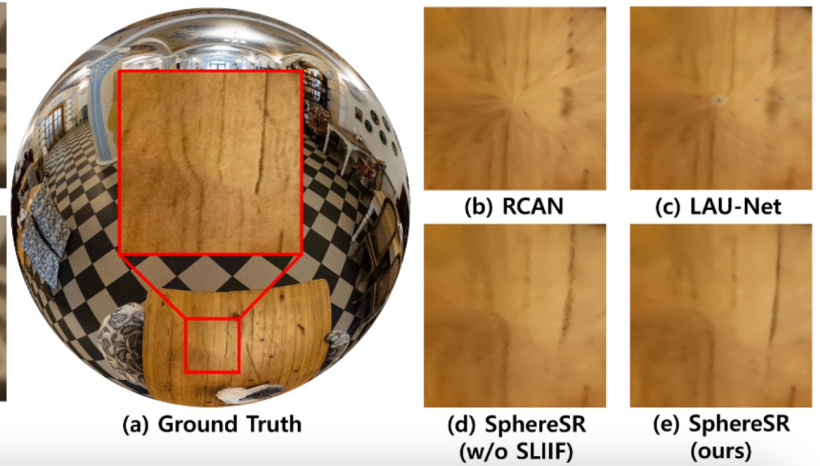
(e) RCAN (f) 360-SS (g) LAU-Net (h) SphereSR (ours)

<Perspective>



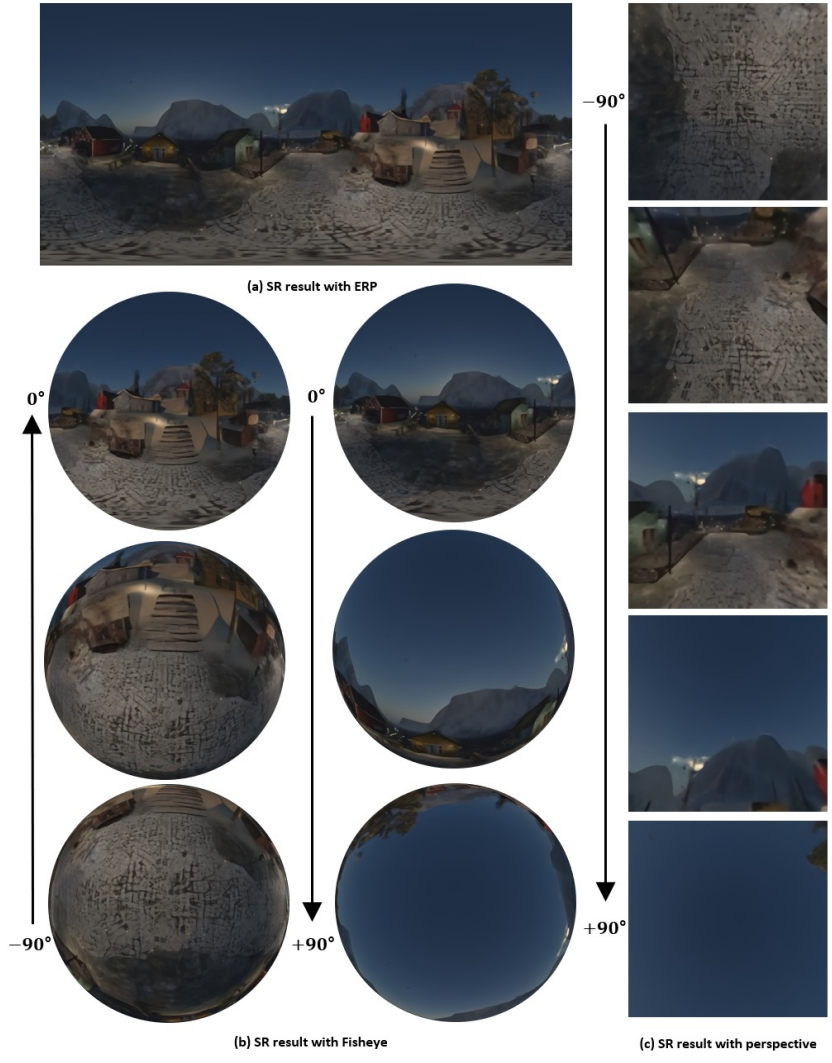
(a) Ground Truth (b) RCAN (c) LAU-Net (d) SphereSR (w/o SLIIF) (e) SphereSR (ours)

<Fisheye>



(a) Ground Truth (b) RCAN (c) LAU-Net (d) SphereSR (w/o SLIIF) (e) SphereSR (ours)

(4) SphereSR: 360° Image Super-Resolution



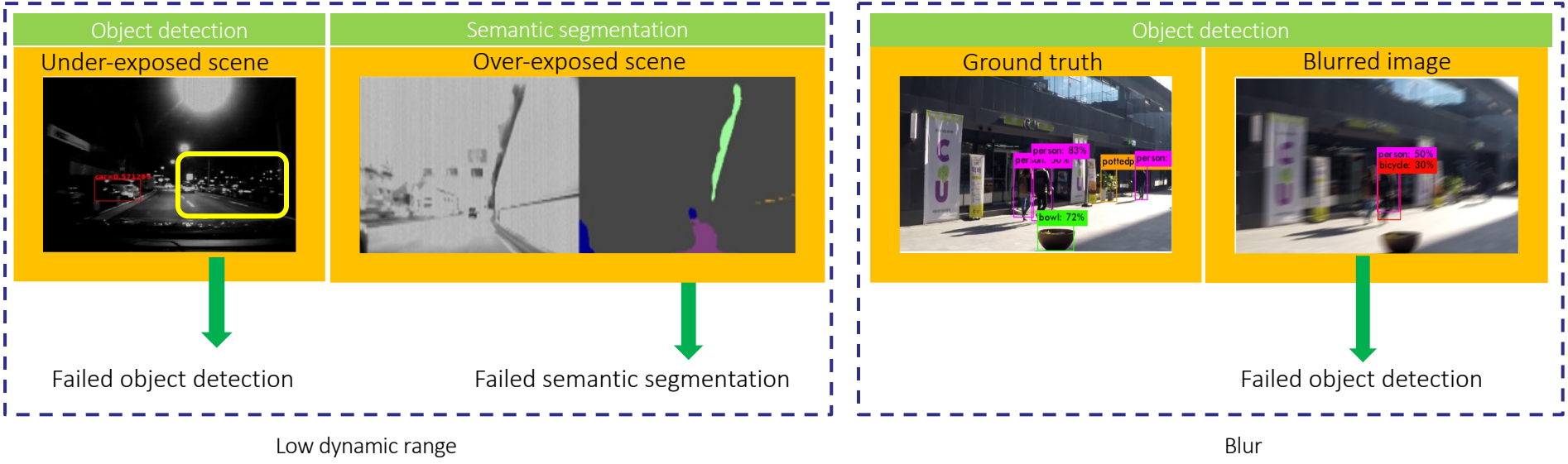
Limitations of ADAS using Conventional Frame-based Cameras (2) - Low dynamic range and framerate

COMPUTER VISION WITH EVENT CAMERAS

Computer Vision using Frame-based Cameras

Open challenges

- Low dynamic range (low contrast in too bright or too dark scenes)
- Motion blur



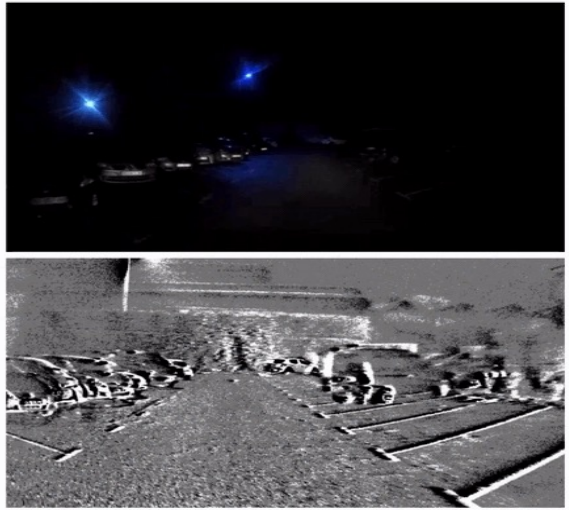
Chen et al., Pseudo-labels for Supervised Learning on Dynamic Vision Sensor Data, Applied to Object Detection under Ego-motion, CVPRW, 2018

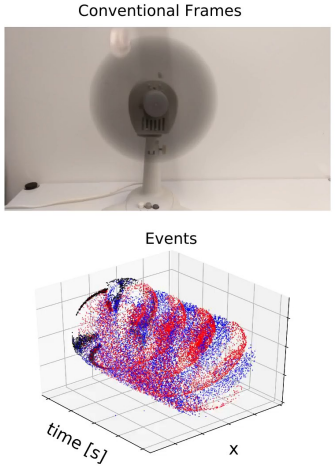
Kupyn et al., DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks, CVPR, 2018.

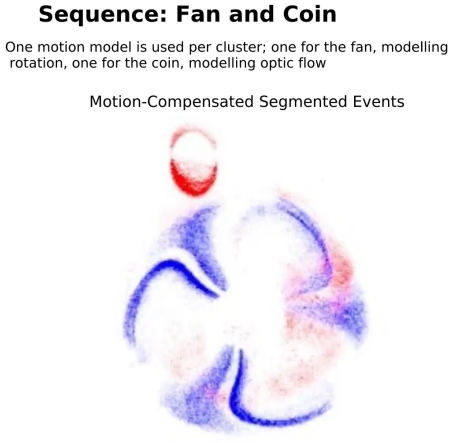
Tackling These Challenges

New camera model is needed.

- Event camera is one of the novel sensors.







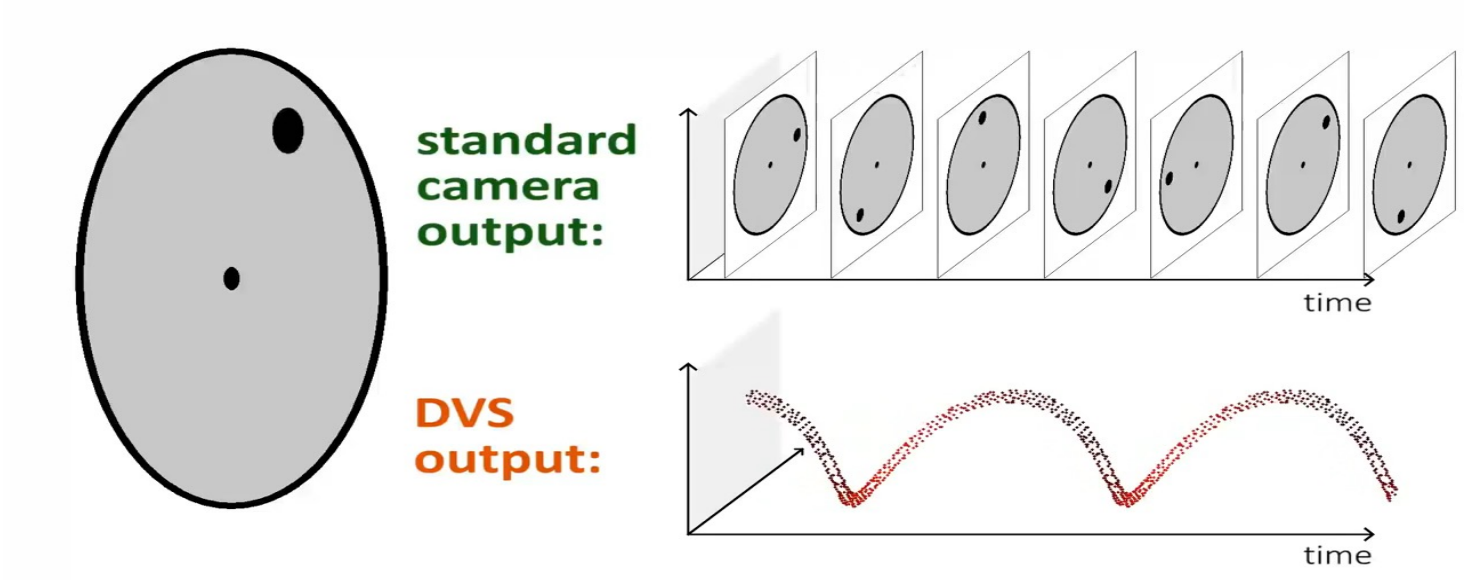
Event cameras can help to handle these challenges!

GIF source: https://blog.scarpellini.dev/posts/introduction.
Video source: Stoffregen et al., Motion Segmentation by Motion Compensation, ICCV'19.

Neuromorphic Cameras

Traditional camera vs Neuromorphic camera

The event camera is not restricted to specific times and report the events as they are created without the motion blur affect. It also does not report events if the scene does not change.



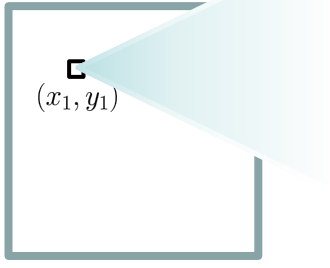
Mueggler, E., Huber, B., & Scaramuzza, D. Event-based, 6-DOF pose tracking for high-speed maneuvers IROS 2014 IEEE/RSJ International Conference on (pp. 2761-2768).

Neuromorphic Cameras

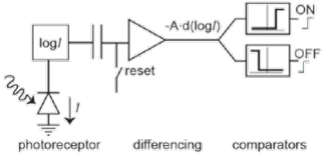
Event (data) stream

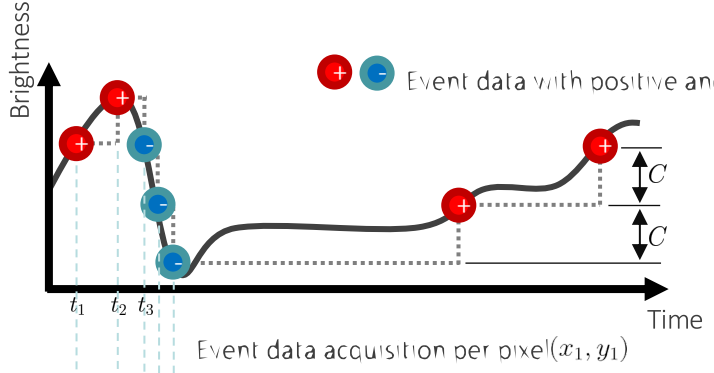
The event stream of an event camera includes the location of an event plus the precise time and sign information.

$\Delta \log I \geq C$



N x N
Photoreceptors





Event data type

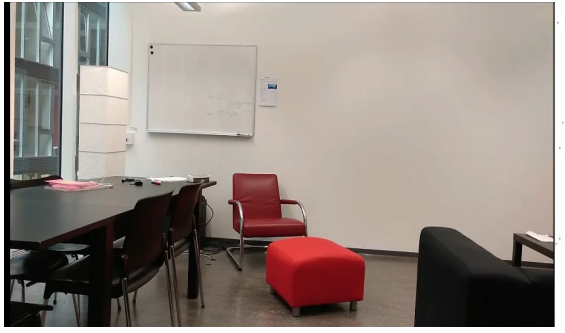
$[(x_1, y_1), t_1, +]$

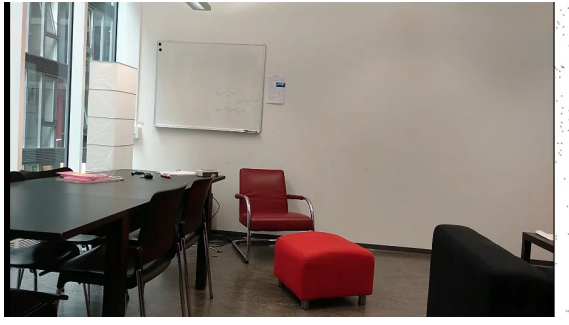
$[(x_1, y_1), t_3, -]$

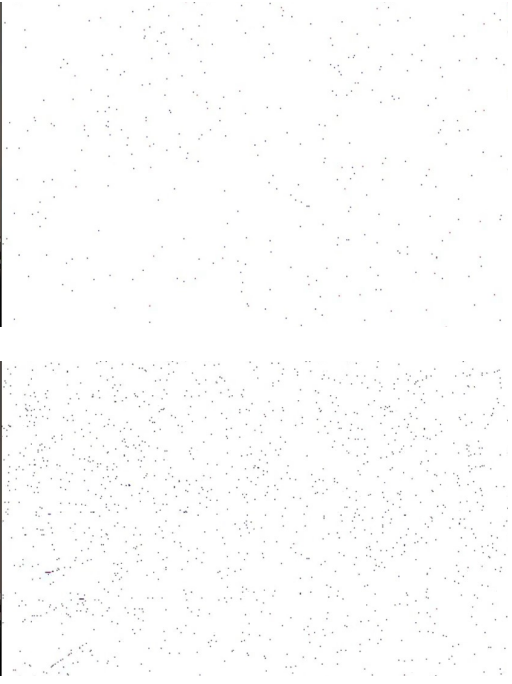
What are event cameras?

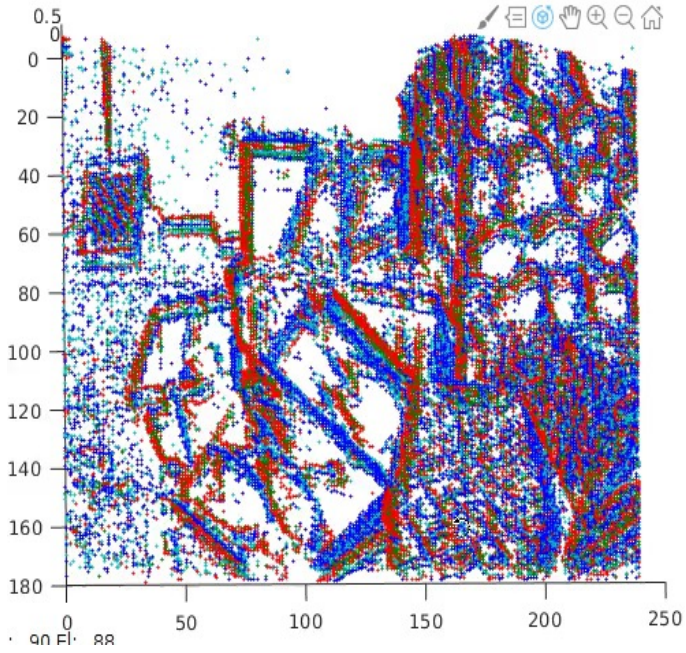
What do events look like?

An event camera reports the +/- events (right) when there is relative motion in comparison to a traditional camera (left).







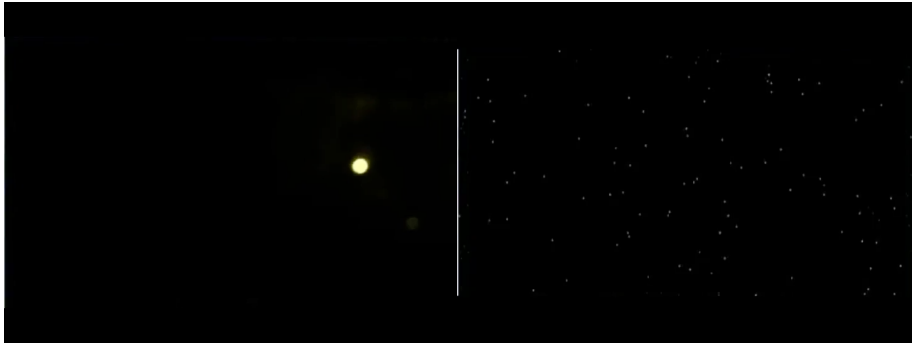


Video referred to http://rpg.ifi.uzh.ch/people_scaramuzza.html

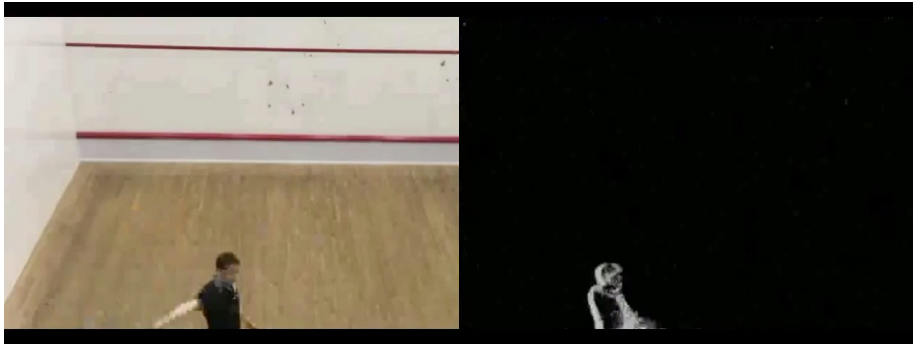
What do events offer?

Vision even with large illumination changes

Motion in extreme low light conditions



High-speed motion detection capability



Efficient detection of dynamic objects

Neuromorphic Cameras

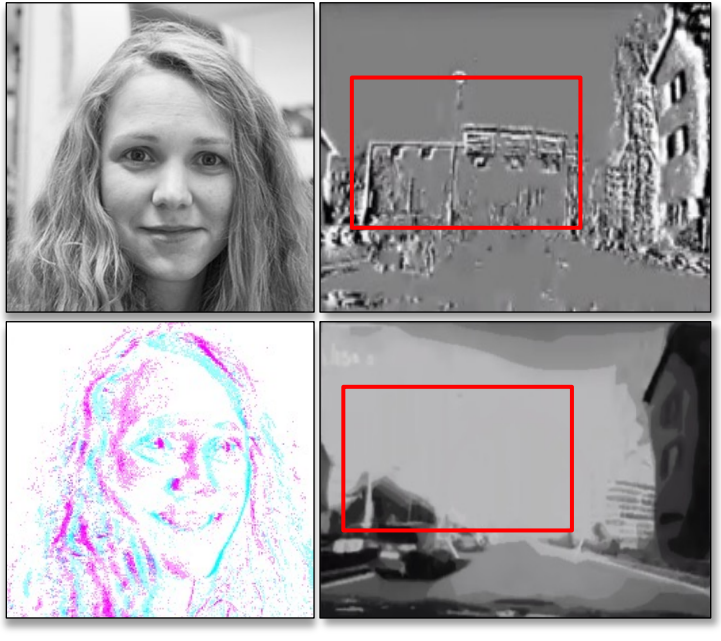
Advantages and disadvantages of event cameras

Advantages

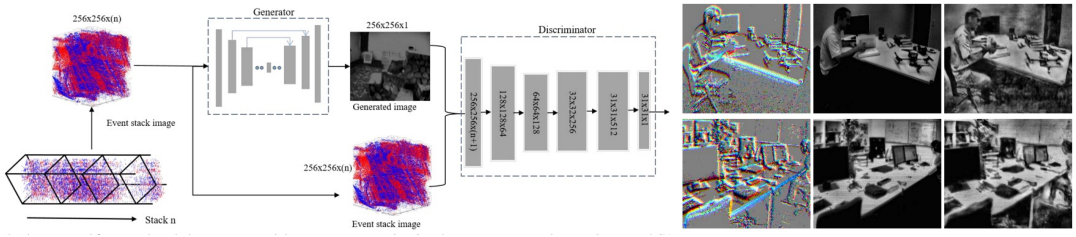
- Low-latency (~1 micro-second)
- High-dynamic range (120 dB instead of 60 dB)
- Very low bandwidth
  (only intensity changes are transmitted): ~200Kb/s
- Low storage capacity, processing time, and power

Disadvantages

- Requires totally new vision algorithms
- No intensity information
  (only binary intensity changes)
- Very low image resolution: 128x128 pixels



Images from Inivation.com

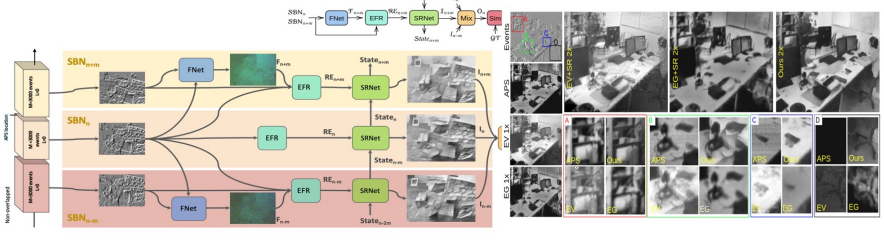


"Event-based High Dynamic Range Image and Very High Frame Rate Video Generation using Conditional Generative Adversarial Networks," CVPR 2019 → IJCV 2021

Lin Wang, S. Mohammad Mostafavi I., and Kuk-Jin Yoon

(1) SUPERVISED
Image Reconstruction from Events

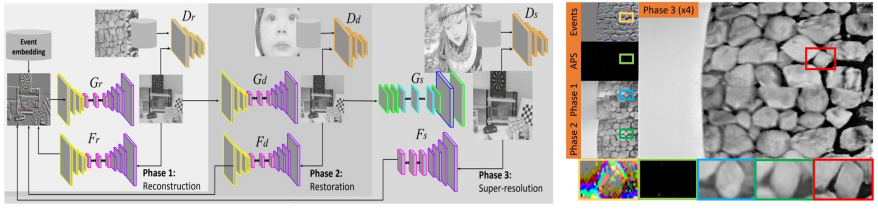
(2) SUPERVISED
Image Reconstruction + Super-Resolution from Events



"Learning to Super Resolve Intensity Images from Events," CVPR 2020 → IEEE TPAMI 2022

S. Mohammad Mostafavi I., Jonghyun Choi, Kuk-Jin Yoon

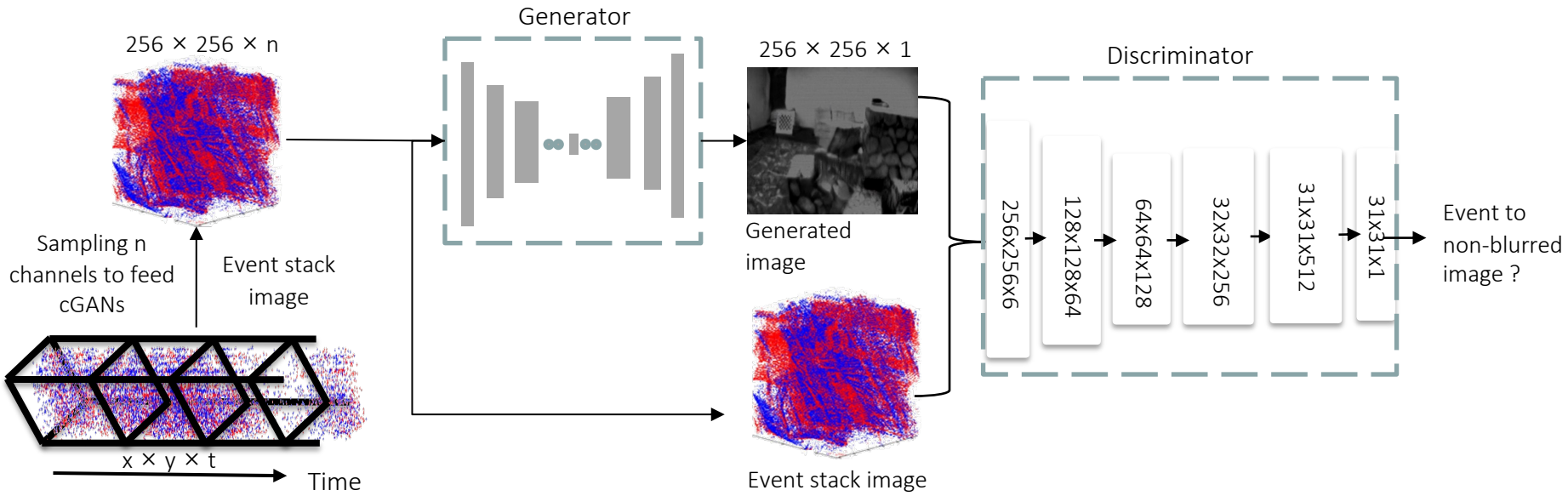
(3) UNSUPERVISED
Image Reconstruction + Super-Resolution from Events



"EventSR: From Asynchronous Events to Image Reconstruction, Restoration, and Super-Resolution via End-to-End Adversarial Learning," CVPR 2020 → IEEE TPAMI 2022

Lin Wang, Tae-Kyun Kim, Kuk-Jin Yoon

(1) SUPERVISED Image Reconstruction from Events



- Learning to Reconstruct HDR Images from Events, with Applications to Depth and Flow Prediction, Mohammad Mostafavi, Lin Wang, Kuk-Jin Yoon, IJCV 2021
- Event-based High Dynamic Range Image and Very High Frame Rate Video Generation using Conditional Generative Adversarial Networks, Lin Wang, S. Mohammad Mostafavi I., Yo-Sung Ho, Kuk-Jin Yoon, CVPR 2019

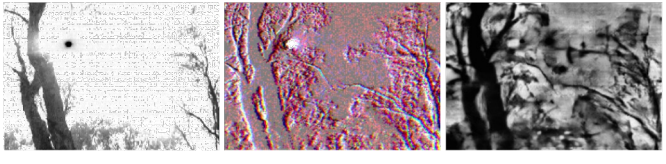
(1) SUPERVISED Image Reconstruction from Events

Real-world HDR scenes

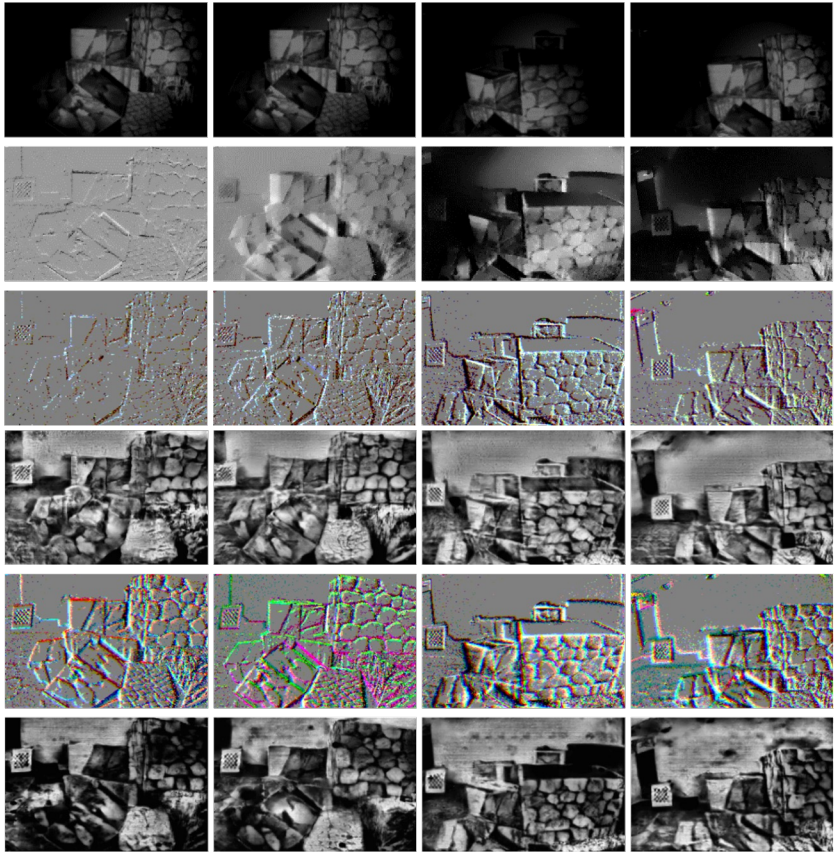
SBT vs SBE vs Bardow's in HDR imaging scene with low light

Extreme case

Directly viewing the sun

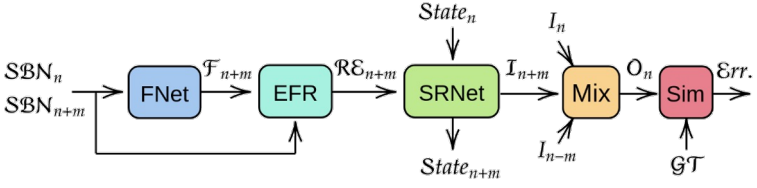


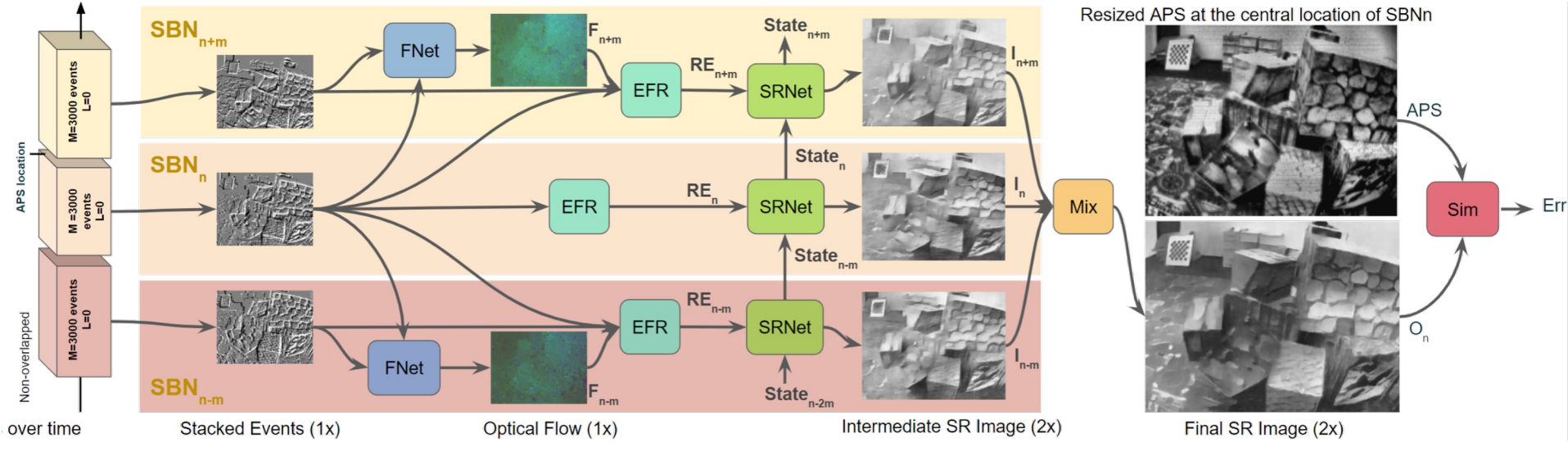
APS | Stack | Our result



Bardow, P., Davison, A. J., & Leutenegger, S. Simultaneous optical flow and intensity estimation from an event camera. CVPR 2016 (pp. 884-892).

(2) SUPERVISED Image Reconstruction + Super-Resolution from Events



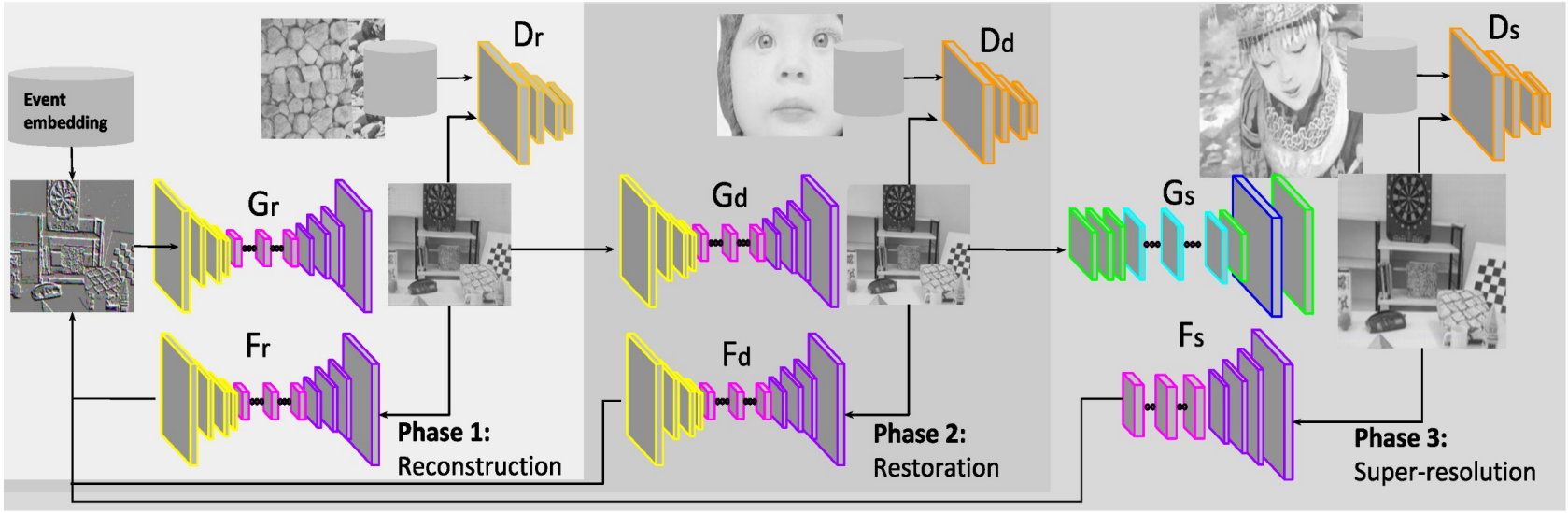


- E2SRI: Learning to Super-Resolve Intensity Images from Events, Sayed Mohammad Mostafaviisfahani, Yeongwoo Nam, Jonghyun Choi, Kuk-Jin Yoon, IEEE TPAMI 2022
- Learning to Super Resolve Intensity Images From Events, S Mohammad Mostafavi I, Jonghyun Choi, Kuk-Jin Yoon, CVPR 2020

(2) SUPERVISED Image Reconstruction + Super-Resolution from Events

(3) UNSUPERVISED Image Reconstruction + Super-Resolution from Events

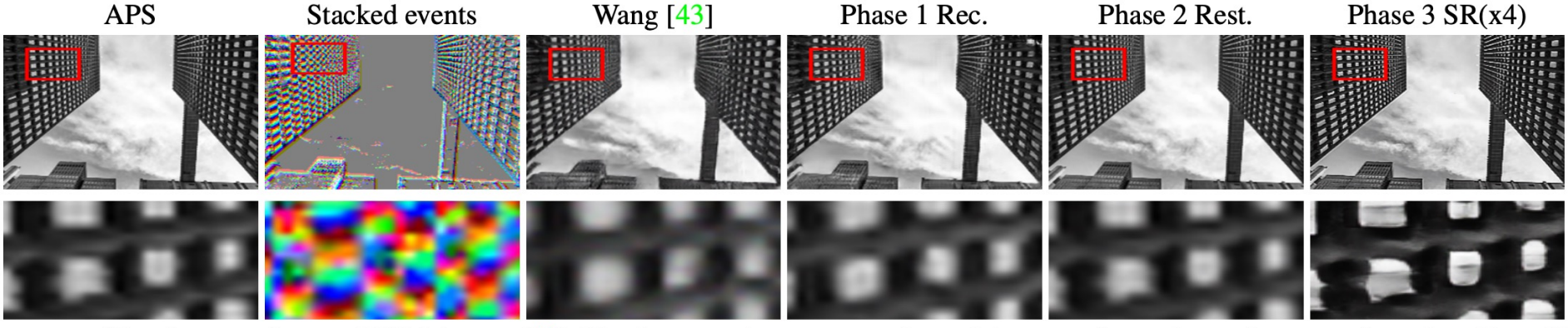
- End-to-end learning
  - Reconstructing, restoring and super-resolving intensity images from LR events in an end-to-end manner.
  - Our method is almost unsupervised, deploying adversarial learning.



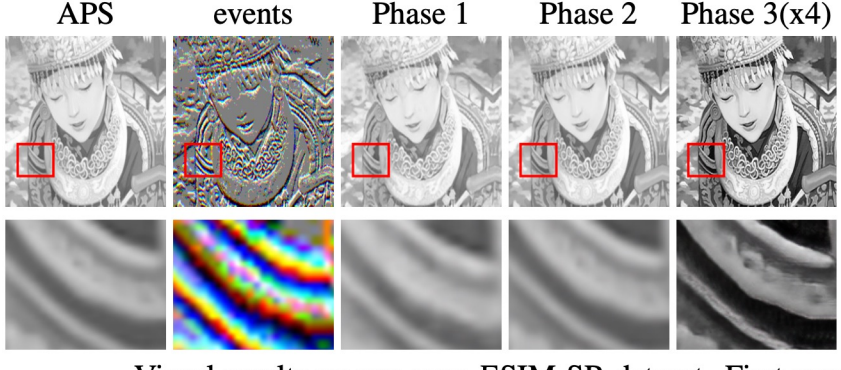
- Joint Framework for Single Image Reconstruction and Super-resolution with an Event Camera, Lin Wang, Tae-Kyun Kim, Kuk-Jin Yoon, IEEE TPAMI 2022
- EventSR: From Asynchronous Events to Image Reconstruction, Restoration, and Super-Resolution via End-to-End Adversarial Learning, Lin Wang, Tae-Kyun Kim, Kuk-Jin Yoon, CVPR 2020

(3) UNSUPERVISED Image Reconstruction + Super-Resolution from Events

- Evaluation on event simulator dataset for quantitative comparison



Visual comparison on ESIM dataset [43]. The first row shows our results and the second row shows the cropped patches. EventSR achieves similar performance regarding phase 1 and better results in phase 2. *See more visual comparisons in the supplementary material.*

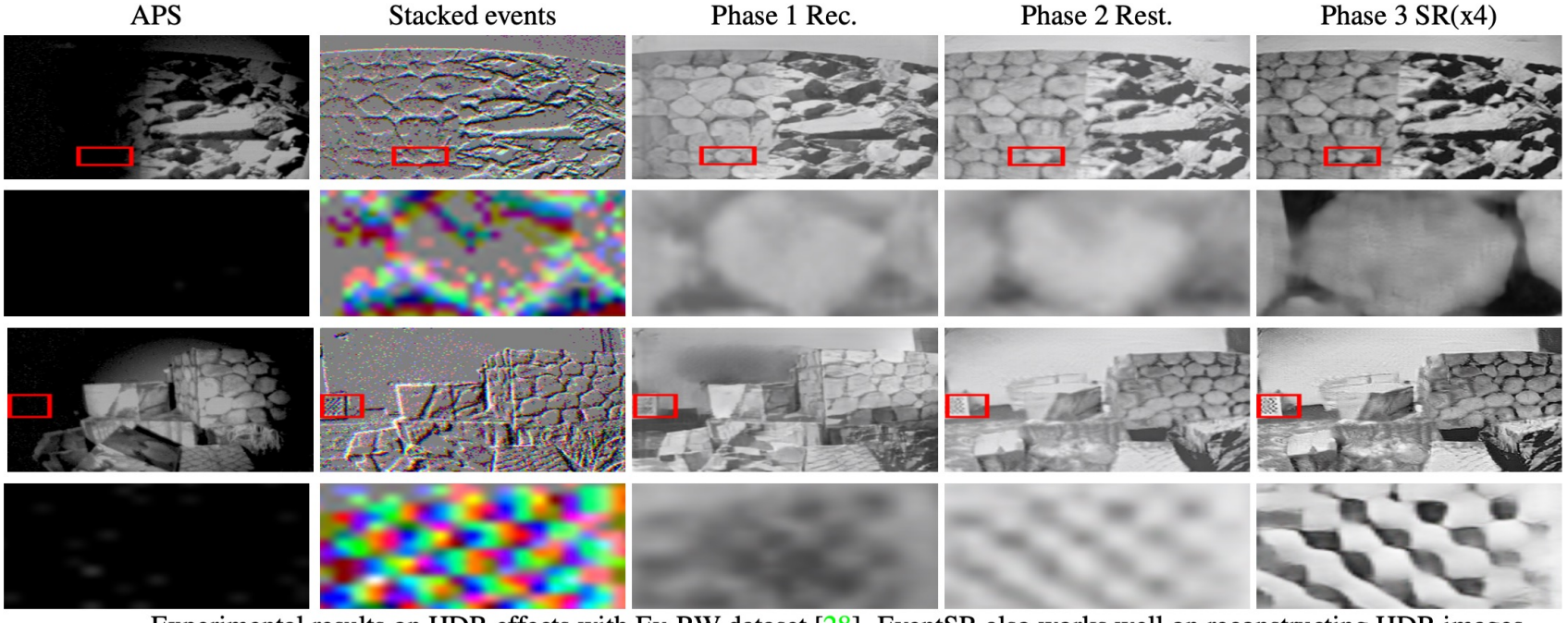


Visual results on our open ESIM-SR dataset. First row shows our results and the second row shows the cropped patches.

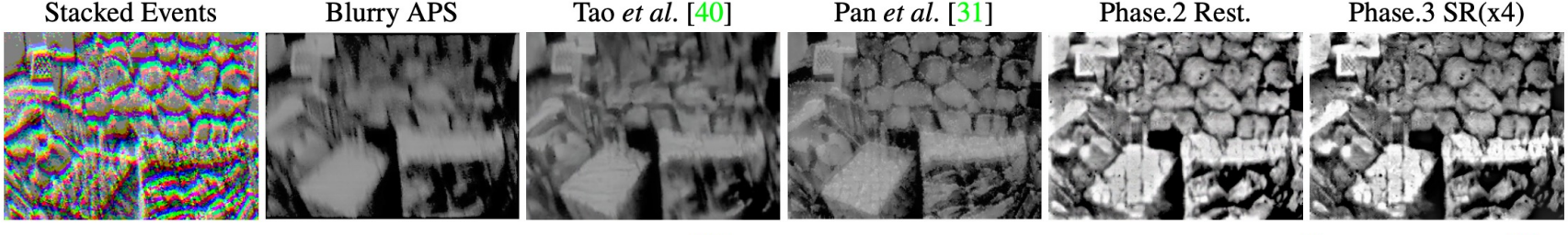
Quantitative comparison of phase 1 and phase 2 with [43] (supervised) based on dataset [43]. Our phase 1 achieves comparable results with [43] and phase 2 achieves much better results.

	PSNR (↑)	FSIM (↑)	SSIM (↑)
Wang [43]($n = 1$)	20.51±2.86	0.81±0.09	0.67±0.20
Wang [43]($n = 3$)	24.87±3.15	0.87±0.06	0.79±0.12
Ours-Rec ($n = 3$)	23.26±3.60	0.85±0.09	0.78±0.24
Ours-Rest ($n = 3$)	**26.75**±2.85	**0.89**±0.05	**0.81**±0.23

(3) UNSUPERVISED Image Reconstruction + Super-Resolution from Events



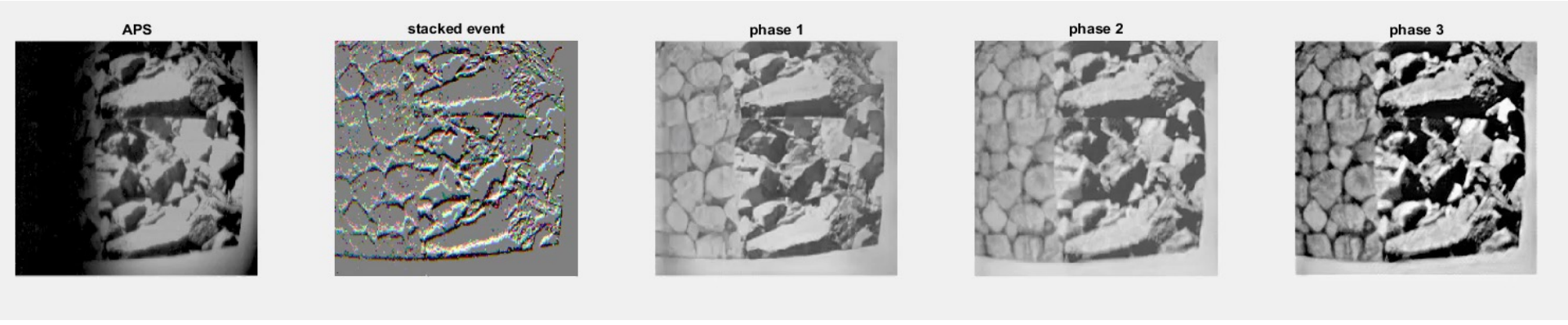
Experimental results on HDR effects with Ev-RW dataset [28]. EventSR also works well on reconstructing HDR images.



Qualitative results for motion blur on RW dataset [28]. EventSR achieves better quality than Tao *et al.* [40] and Pan *et al.* [31].

(3) UNSUPERVISED Image Reconstruction + Super-Resolution from Events

- High dynamic range image generation



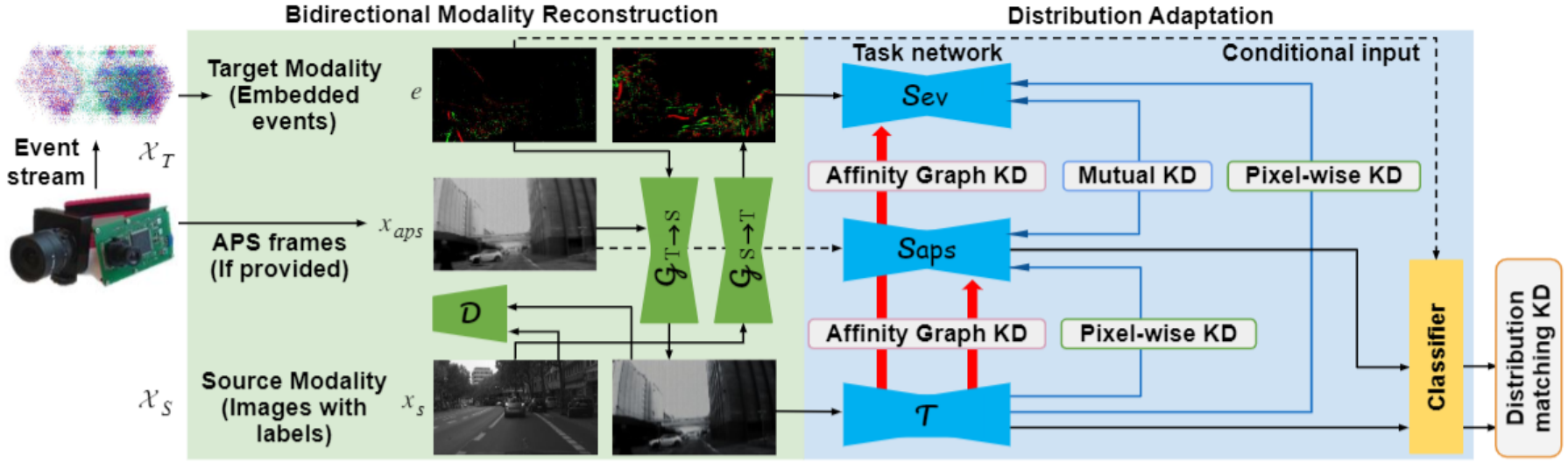
- Robustness against motion blur

(4) EvDistill: Asynchronous Events to End-task Learning via Bidirectional Reconstruction-guided Cross-modal KD

Overview

- End-to-end bidirectional modality reconstruction, which can be removed after training, adding no additional computation cost.

- Knowledge transfer to event-based end-task learning
  - The teacher network transfer knowledge to APS and event student networks.

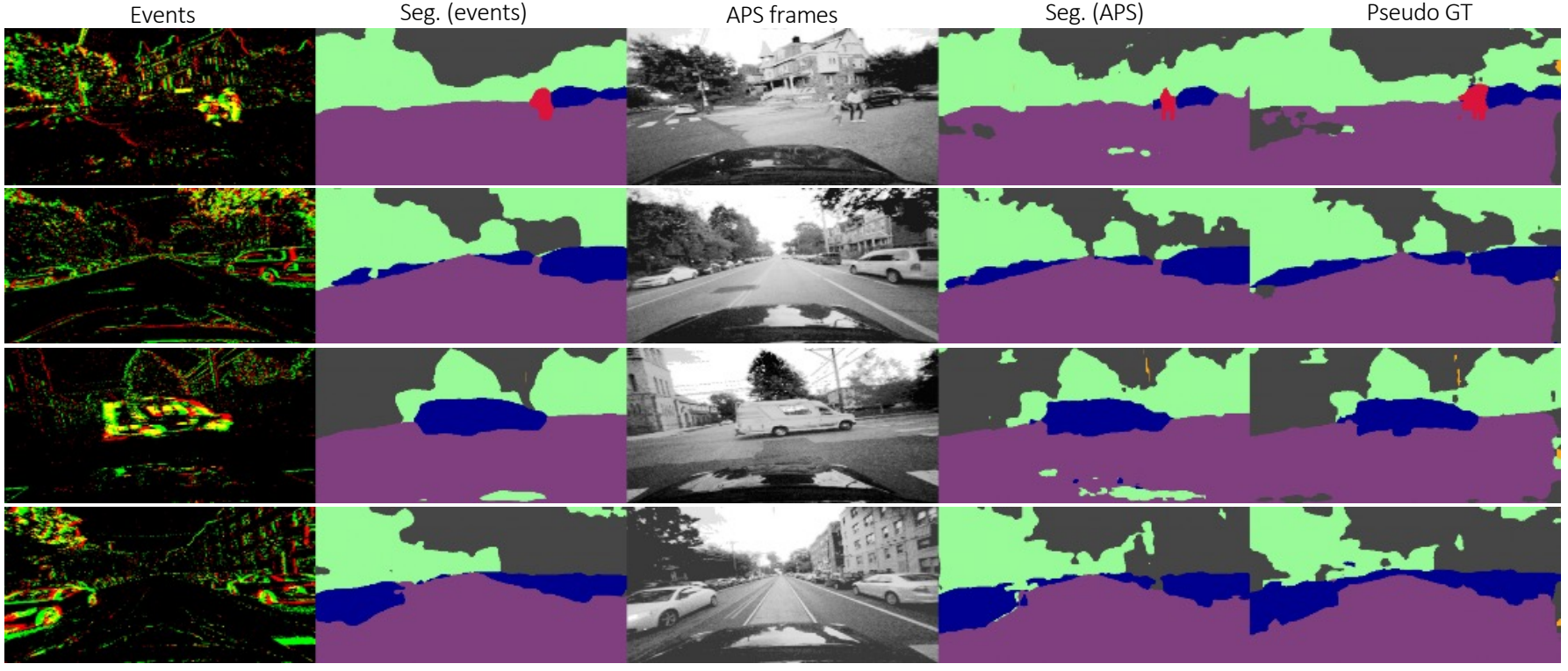


- Evdistill: Asynchronous Events to End-task Learning via Bidirectional Reconstruction-guided Cross-modal Knowledge Distillation, Lin Wang, Yujeong Chae, Sung-Hoon Yoon, Tae-Kyun Kim, Kuk-Jin Yoon, CVPR 2021

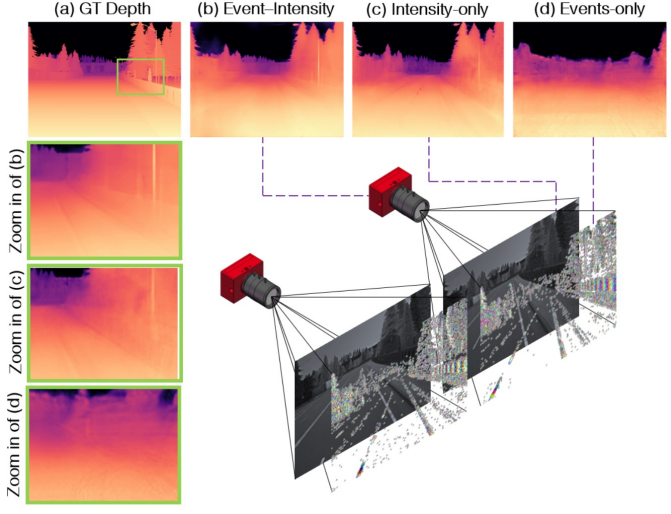
(4) EvDistill: Asynchronous Events to End-task Learning via Bidirectional Reconstruction-guided Cross-modal KD

MVSEC dataset

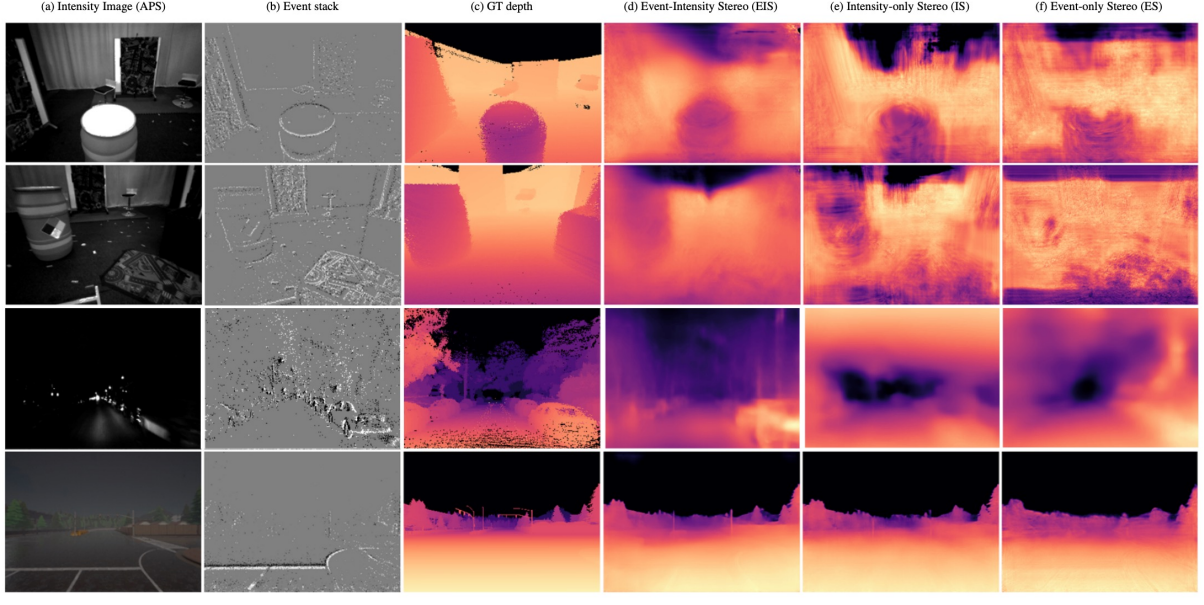
- Qualitative results (Day2 sequence).



(5) Event-Intensity Stereo: Estimating Depth by the Best of Both Worlds



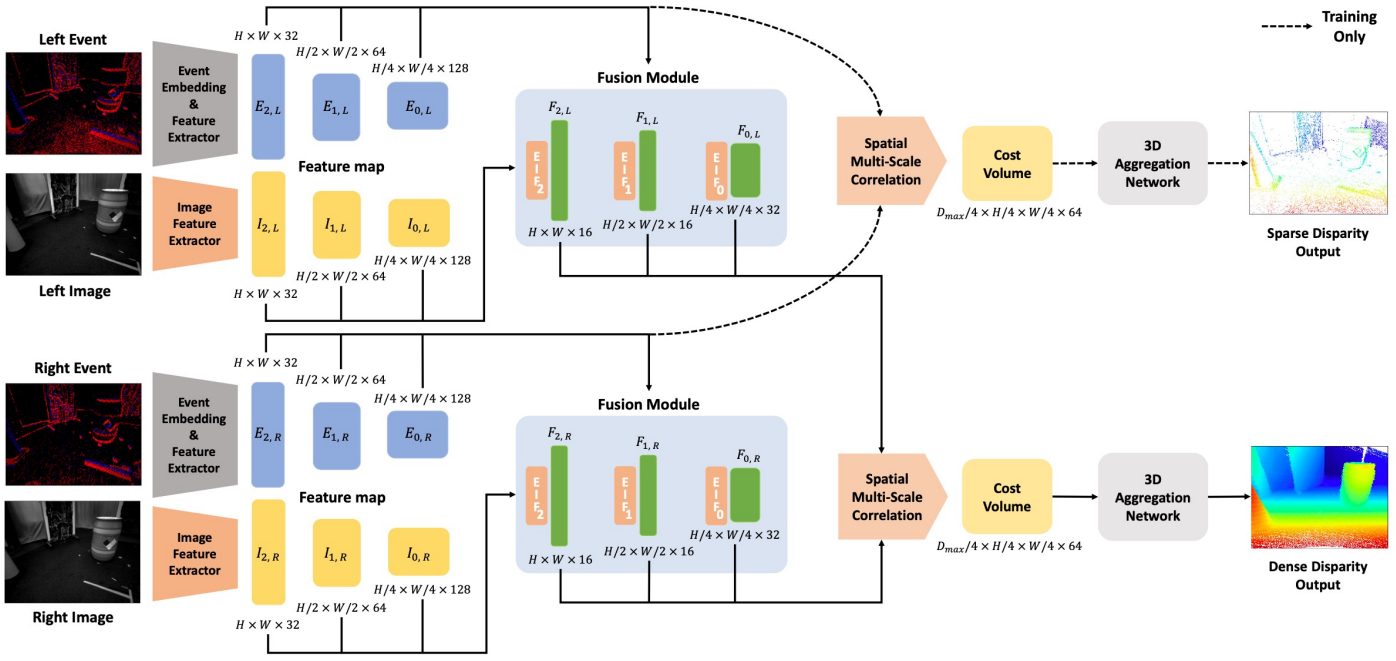
Estimating dense depth using our event-intensity stereo depth estimation framework. Our end-to-end network can estimate depth from the combination of Event-Intensity Stereo (b), Intensity-only stereo (c), or Event-only stereo (d) pairs. Using event-intensity stereo, we can reach higher quality depth in comparison to event-only or intensity-only inputs, as it can surpass the shortcomings of each source while gathering the best from them.



Quantitative comparison among different stereo methods based on their input sources. Our Event-Intensity Stereo (EIS) method (d), utilizes the Intensity (a) and event stacks (b) to estimate more accurate detailed depth in comparison to Intensity-only Stereo (IS) (e), or Event-only Stereo (ES) (f), methods. Data are from the real-world MVSEC [49] and simulated ECAR dataset (last row only).

- Event-Intensity Stereo: Estimating Depth by the Best of Both Worlds, Mohammad Mostafavi, Kuk-Jin Yoon, Jonghyun Choi, ICCV 2021

(5) Event-image Fusion Stereo using Cross-modality Feature Propagation

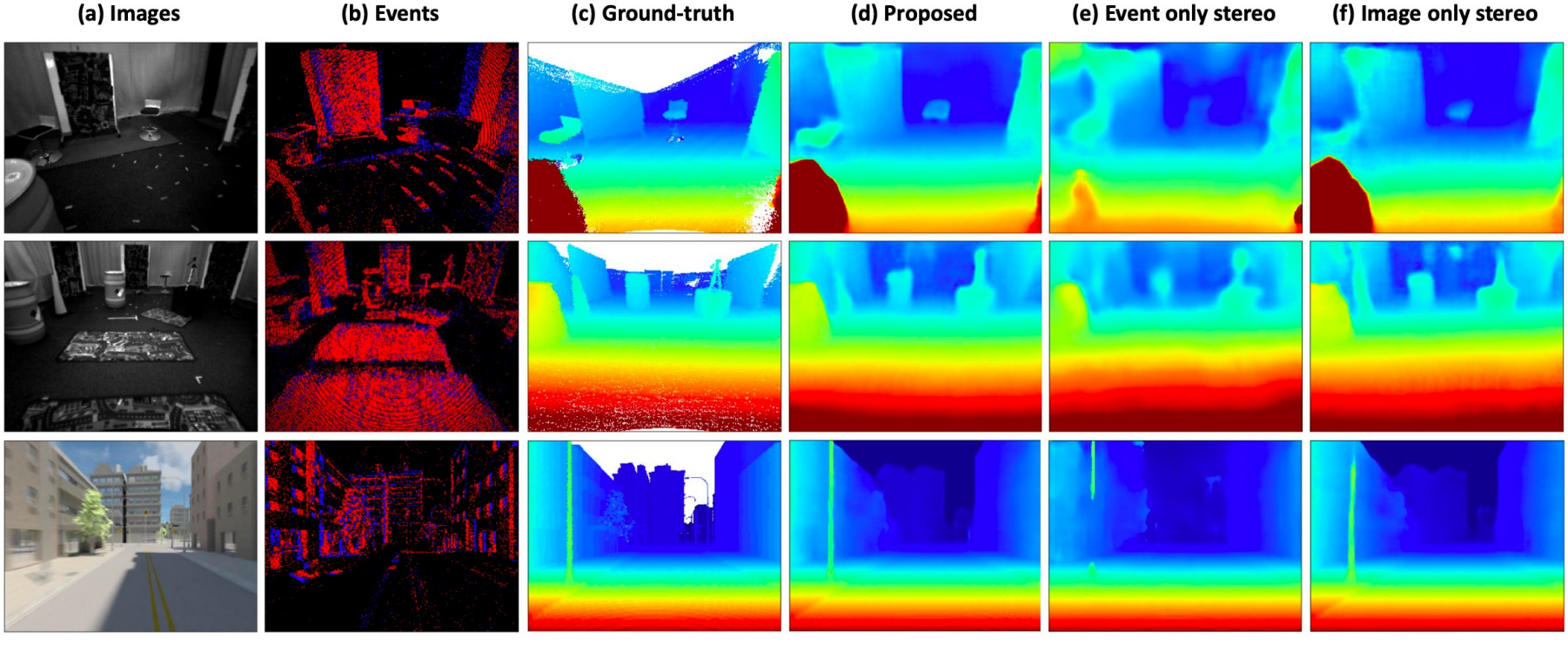


Overall framework of the proposed method. The proposed network employs both images and events from a stereo event camera as inputs. The event stream data are passed through an embedding module to enable the use of a CNN. The image feature extractor shares weights between the image feature extractor, which is also true for event feature extractors. The features of various scales are passed through the fusion and correlation modules to generate a cost volume. $F_{2,L}, F_{1,L}, F_{0,L}$ represent the multi-scale features of the left output of the fusion module, and $F_{2,R}, F_{1,R}, F_{0,R}$ represent the multi-scale features of the right output of the fusion module. The cost volume generated from the spatial multi-scale correlation is passed through a 3D aggregation network for dense disparity extraction. The dotted line is solely used to train the model, and it creates a sparse disparity using the event features alone.

- Event-image Fusion Stereo using Cross-modality Feature Propagation, Hoonhee Cho and Kuk-Jin Yoon, AAAI 2022

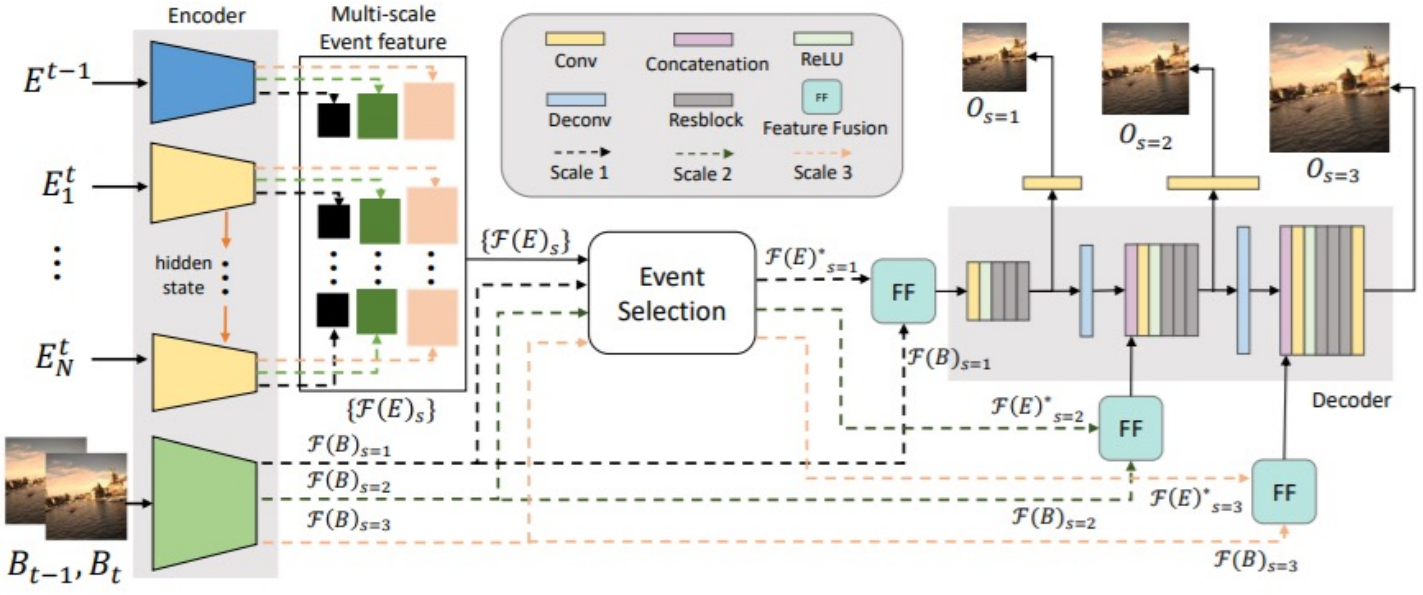
87

(5) Event-image Fusion Stereo using Cross-modality Feature Propagation



Qualitative comparison of the proposed method with an event-based method and a frame-based method. The first two rows are split 1 and split 3 from the MVSEC (real-world) dataset, and the third row is the RGB frame-based synthetic dataset, respectively. In (a) and (b), we visualize the image and the 15,000 most recent events from the left camera. Note that (e) and (f) are the results of (Tulyakov et al. 2019) and (Guo et al. 2019), respectively. Our proposed method (d) utilizes an image with the corresponding events.

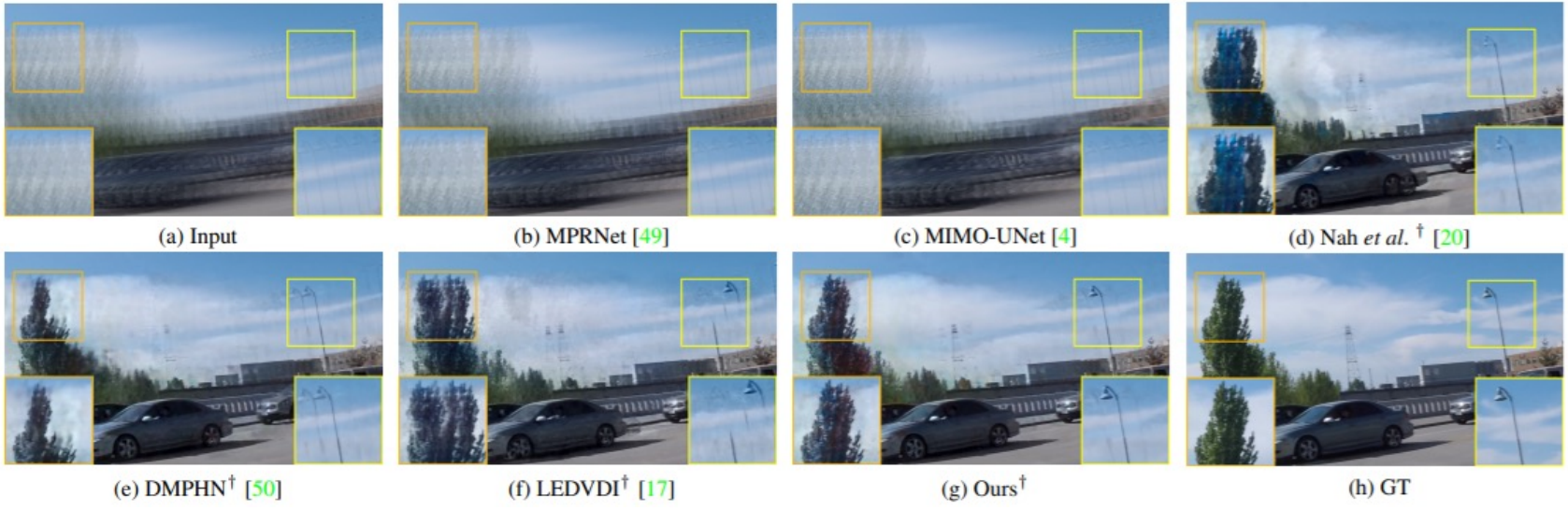
(6) Event-guided Deblurring of Videos



The overall framework of our methods

- The proposed frameworks consists of two major components: event-selection and feature fusion.
- For event selection, we first encode the embedded events via an RNN-based encoding network. Then we propose an novel ETES module to select the beneficial event features without any supervision.
- Finally, we propose a new feature-fusion module that effectively exploits the complementary information of events and frame.

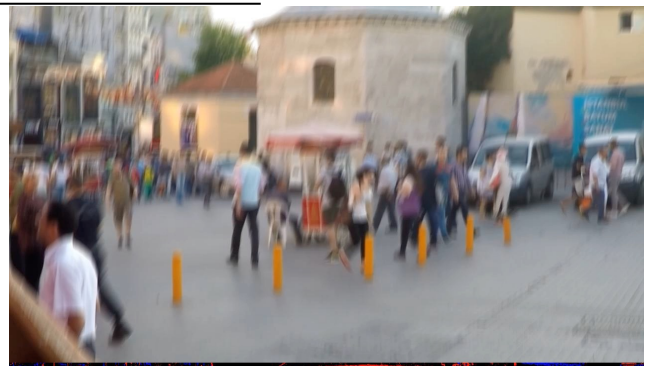
(6) Event-guided Deblurring of Videos



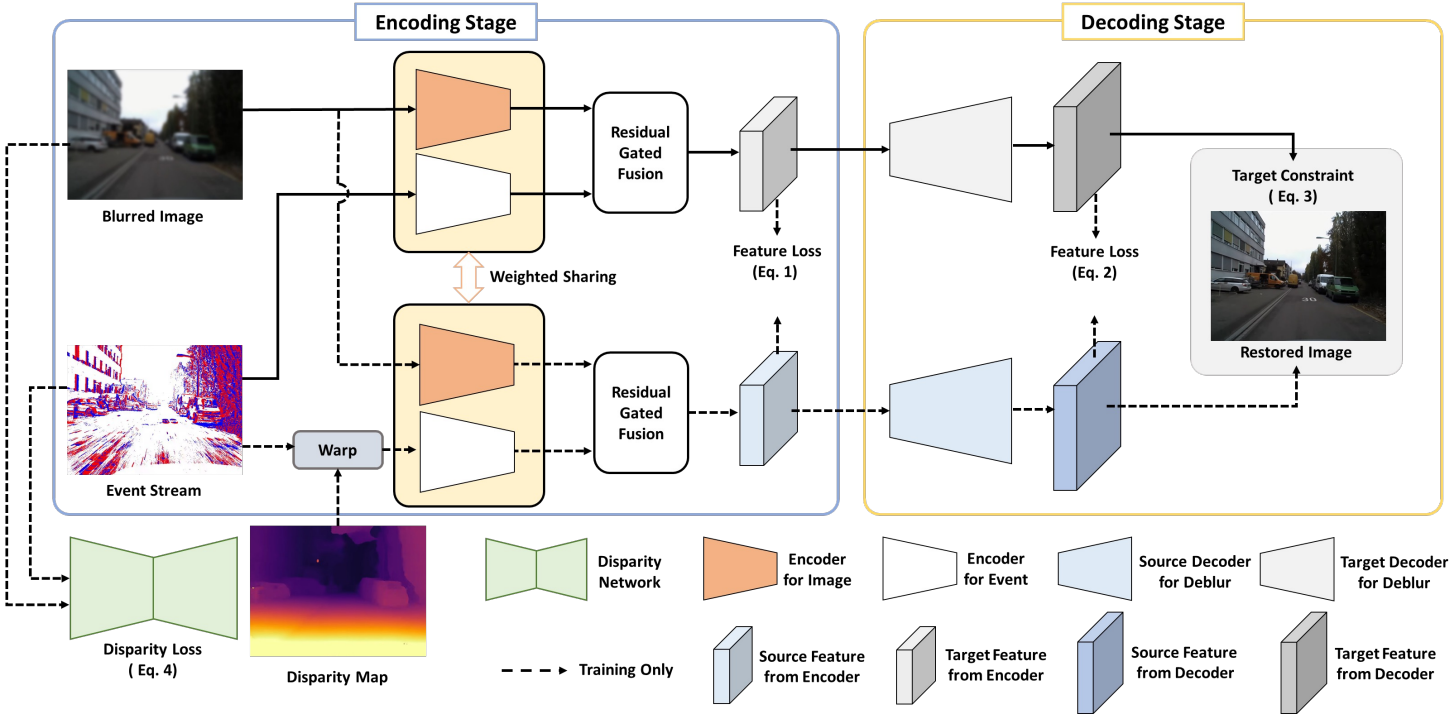
Qualitative comparison between our method and state-of-the art methods

- our methods can restore sharper texture details than the previous event-guided methods under extreme blur conditions.
- In particular, only our methods can restore the sophisticated tree structure and thin street lamp.

(6) Event-guided Deblurring of Videos



(7) Image Deblurring with Non-coaxial Event Camera via Distillation Learning



The overall framework of DDNC

- The DDNC is the unified framework consisting of a target network (student) and a source network (teacher).
- The target network is given an event stream non per-pixel aligned with the image while the source network is provided the event stream per-pixel aligned through a disparity estimation network.

(7) Image Deblurring with Non-coaxial Event Camera via Distillation Learning

Qualitative comparisons with other methods

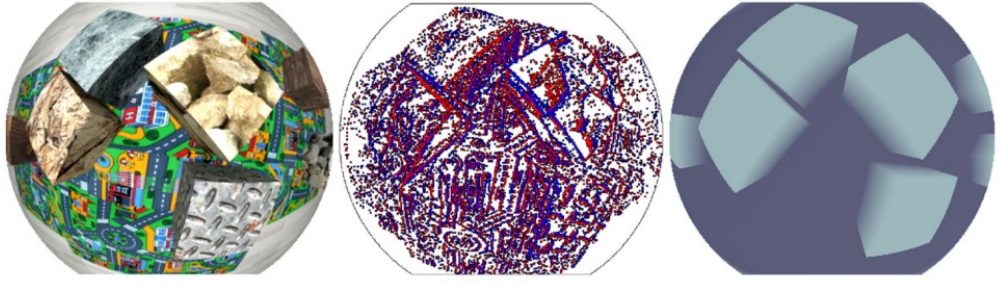
Qualitative Results

- Compared to the previous state-of-the-art frame-based deblurring methods, the proposed method generates much sharper images guided by the events stream.
- In particular, it is excellent for recovering information lost in the degradation process, and it is effective for areas with a lot of blur.
- Even in a detailed area such as text, it can be seen that artifacts are significantly less compared to the existing method.

(8) EOMVS: Event-Based Omnidirectional Multi-View Stereo

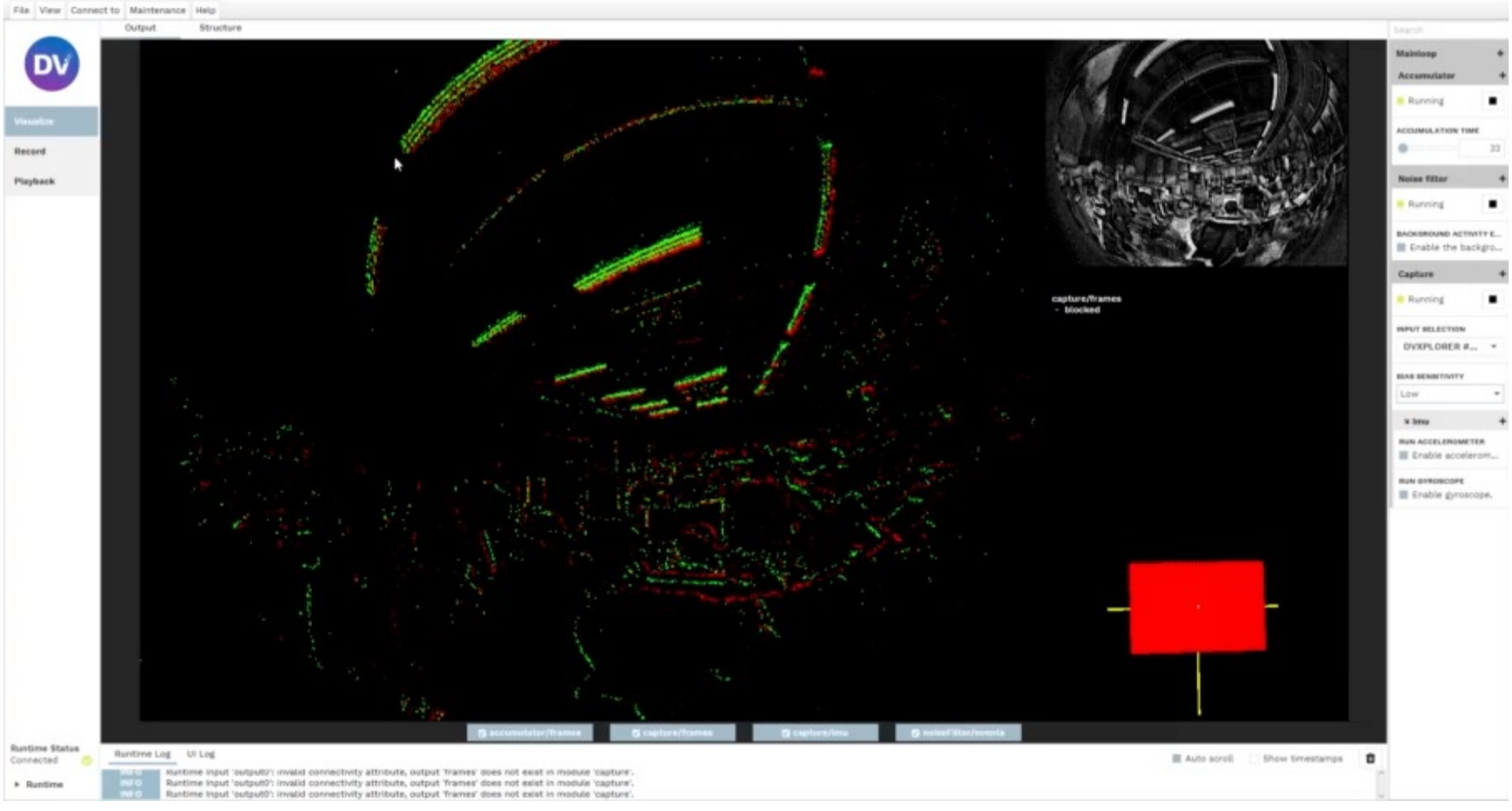
Proposing a new multi-view stereo method, called EOMVS, to reconstruct a 3D scene with a wide view using the event data captured by omnidirectional event cameras.

Omnidirectional event camera setup consisting of a DVXplorer event camera (rear) and an Entaniya Fisheye lens (front).

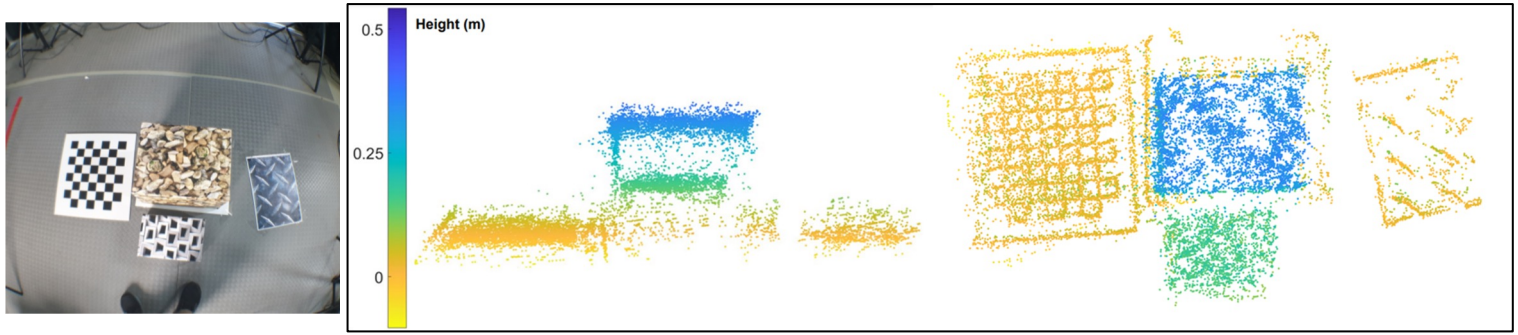


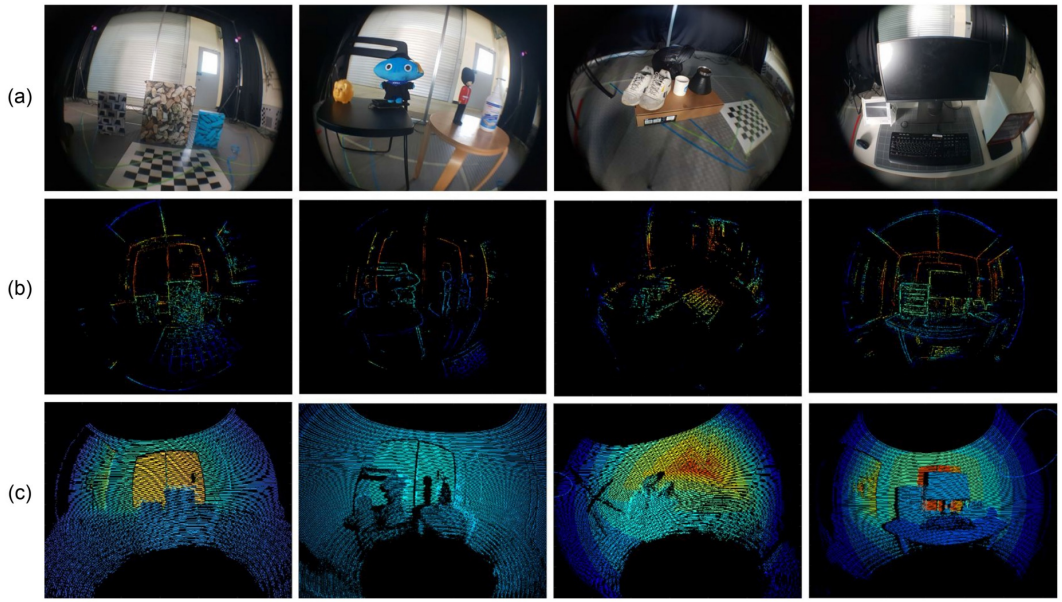
Synthetic omnidirectional event dataset made from Blender. **Left:** RGB image with a high frame rate. **Center:** the event data generated by ESIM. **Right:** ground-truth depth map used for evaluation. (Note: all data are obtained from the top view. The point of view of the camera is toward the ground.).

- EOMVS: Event-Based Omnidirectional Multi-View Stereo, Hoonhee Cho, Jaeseok Jeong, Kuk-Jin Yoon, IEEE RA-L 2021


File
View
Connect to
Maintenance
Help
Output
Structure
Visualize
Record
Playback
capture/frames
- blocked
Mainloop
Accumulator
Running
ACCUMULATION TIME
Noise filter
Running
Capture
Running
INPUT SELECTION
DVXPLORER #...
BIAS SENSITIVITY
Low
RUN ACCELEROMETER
Enable accelerom...
RUN GYROSCOPE
Enable gyroscope.
accumulator/frames
capture/frames
capture/imu
noisefilter/events
Runtime Status
Connected
Runtime
Runtime Log
UI Log
Auto scroll
Show timestamps
Runtime input 'output0\': invalid connectivity attribute, output 'frames' does not exist in module 'capture'.
Runtime input 'output0\': invalid connectivity attribute, output 'frames' does not exist in module 'capture'.
Runtime input 'output0\': invalid connectivity attribute, output 'frames' does not exist in module 'capture'.

(8) EOMVS: Event-Based Omnidirectional Multi-View Stereo





Real-world dataset. **(a)** Real scenes in which the real-world dataset was created. **(b)** Semi-dense depth map obtained by our method for the real-world dataset. Blue represents nearby objects and red represents distant objects. **(c)** Ground-truth depth map of the real-world dataset. (Note: **(a)** was taken using a separate camera that is not aligned with the omnidirectional event camera and is only shown for visualizing the scene.

Thank you for your attention.

QnA