A Simple yet Universal Framework for Depth Completion

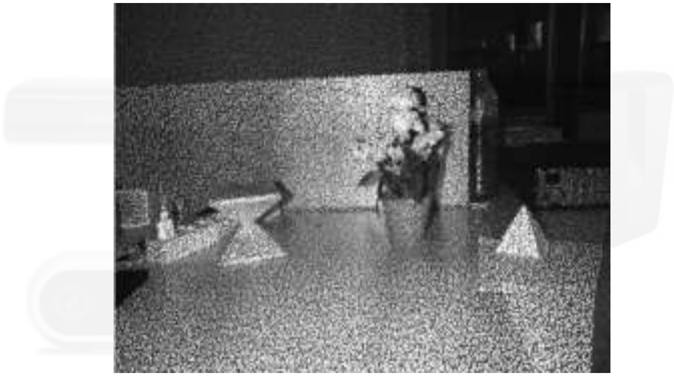
Jin-Hwi Park, Hae-Gon Jeon

Depth Perception with Sensors

Requiring 3D information with depth sensors.

Commercial Depth Sensors

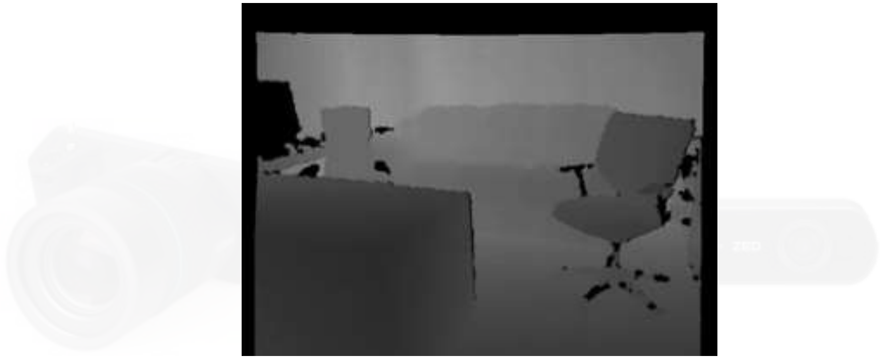
Limitations: Sensors hardly provide dense 3D information in real-time



Structured Light

LiDAR

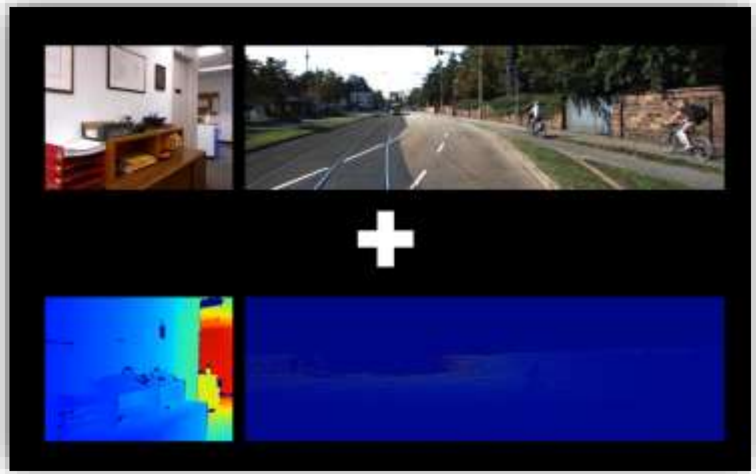
ToF (Time-of-Flight)



Stereo Camera

Depth Completion

Depth Estimation with Sparse Measurement and corresponding RGB image



Apple iPhone & iPad

Microsoft Kinect & Velodyne LiDAR

*video source: developer.apple.com/videos/play/wwdc2020

Commercial Depth Sensors

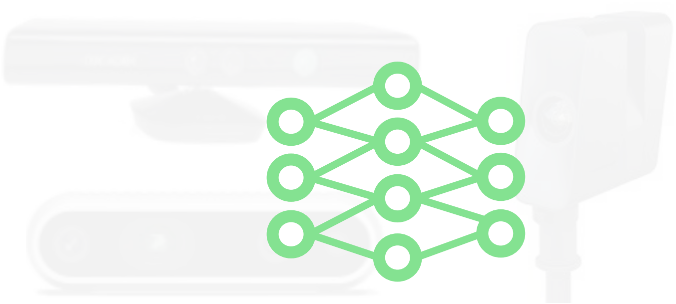
Structured Light

LiDAR

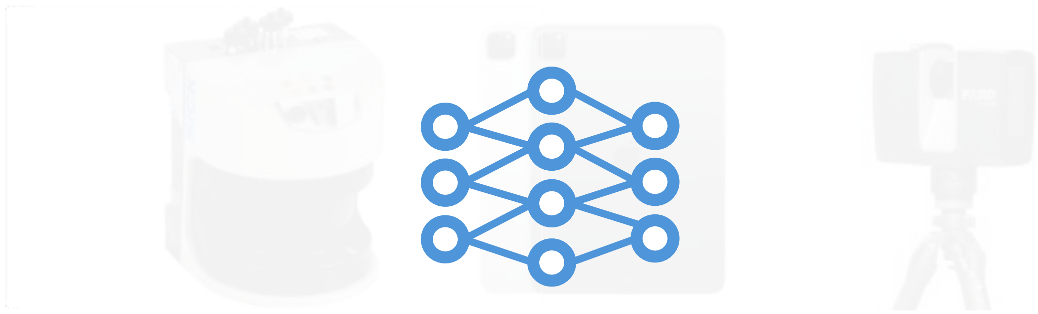
ToF (Time-of-Flight)

Stereo Camera

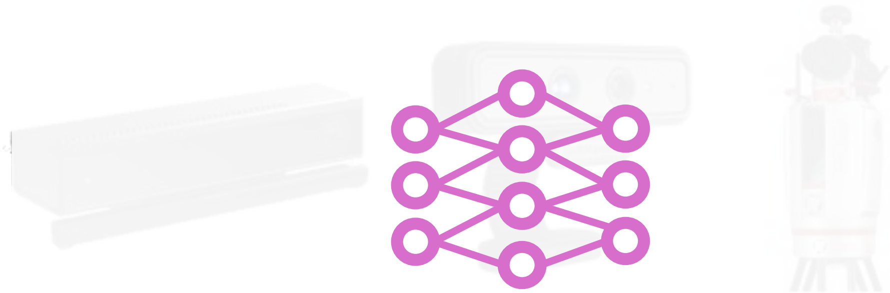
Commercial Depth Sensors



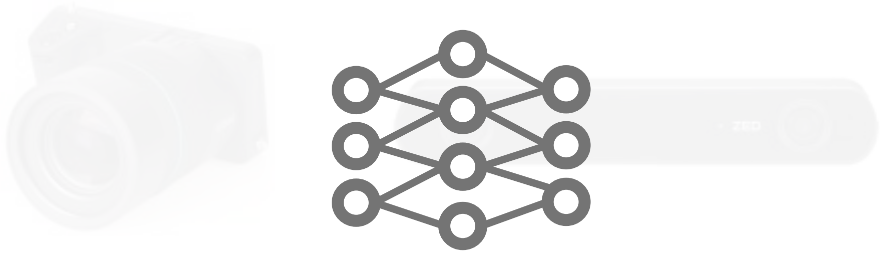
Model for Structured Light



Model for LiDAR



Model for ToF (Time-of-Flight)



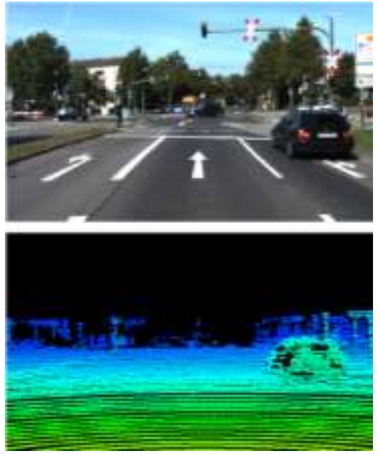
Model for Stereo Camera

Commercial Depth Sensors

Accuracy ↓

Range ↓ Density ↑

Structured Light



Accuracy ↑

Range ↑ Density ↓

LiDAR

Accuracy ↑

Range ↓ Density ↓

ToF (Time-of-Flight)

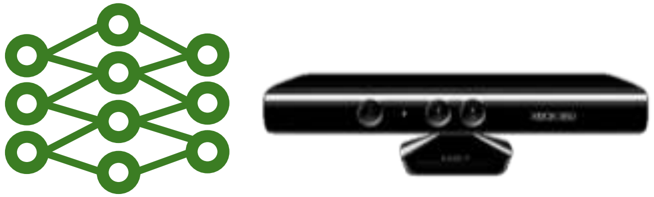
Accuracy ↓

Range ↑ Density ↑

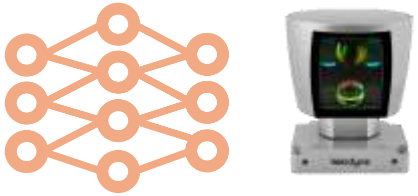
Stereo Camera

Commercial Depth Sensors

[Park et al., CVPR 2024] Exploring Sensor Bias Problem

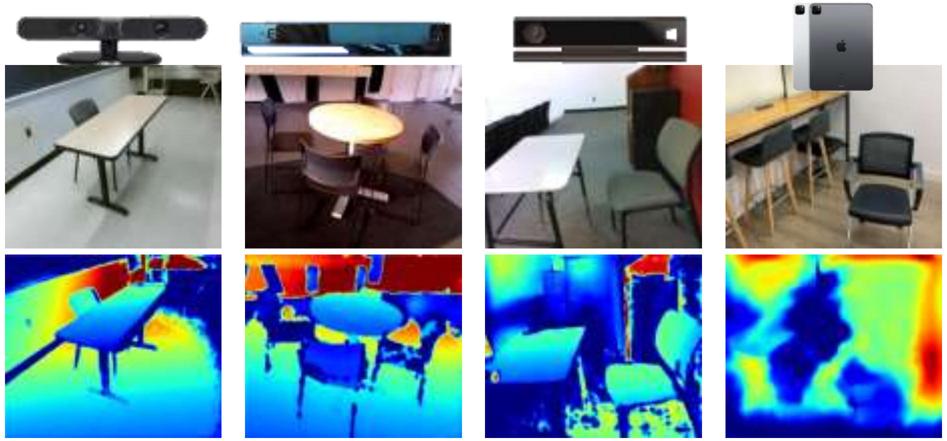


Model trained on Kinect v1

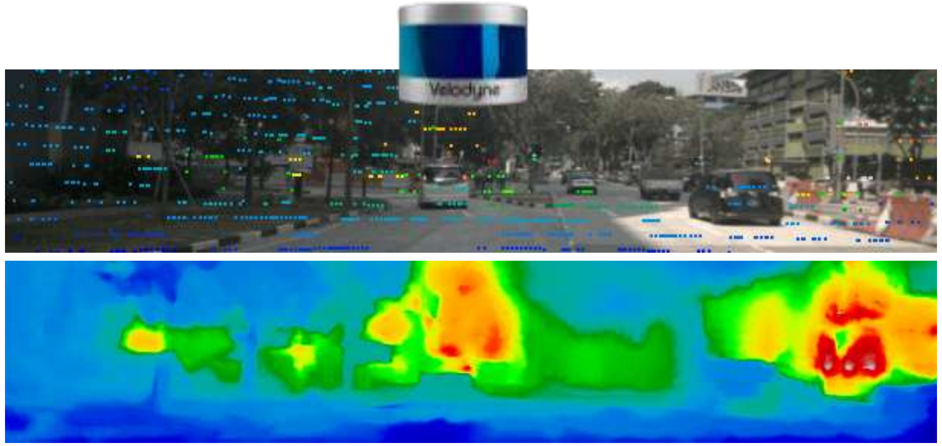


Model trained on32-Line LiDAR

Intel RealSense Asus Xtion Kinect v2 Apple iPad



32-Line LiDAR

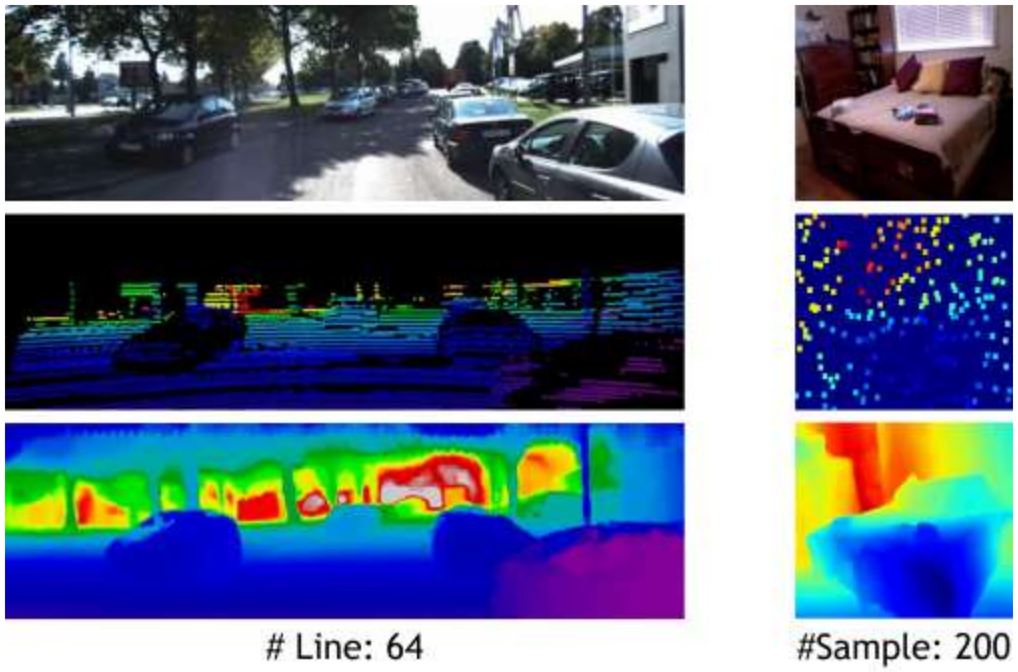


“Models are biased toward specific sensor type.”

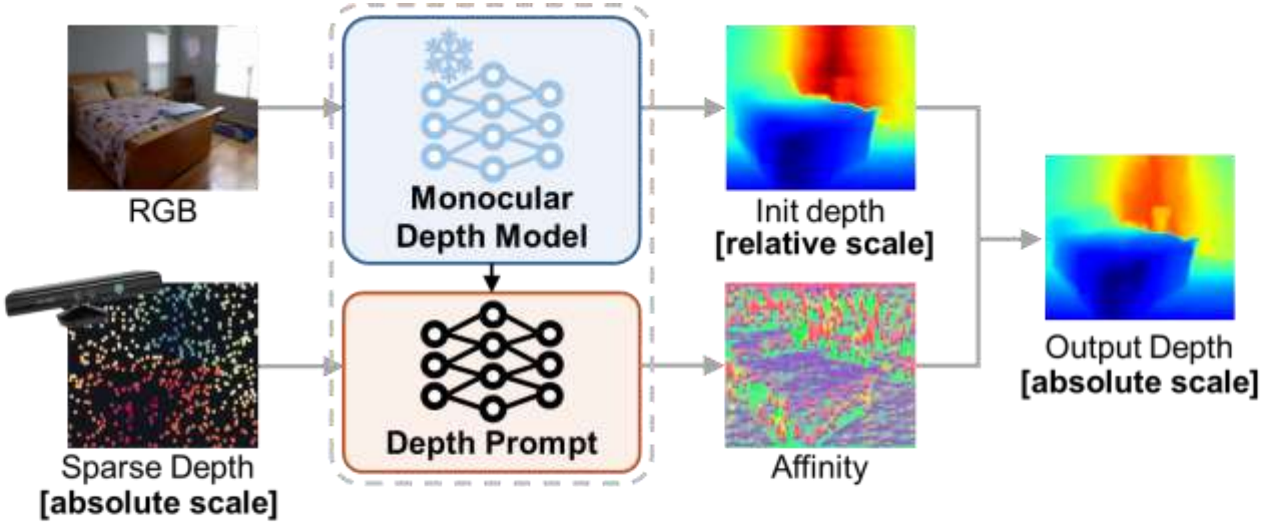
Depth Prompting for Sensor-Agnostic Depth Estimation. CPVR2024

Commercial Depth Sensors

[Park et al., CVPR 2024] Sensor-Agnostic Depth Completion



Qualitative Results (KITTI & NYU)



DepthPrompting Architecture

Depth Prompting for Sensor-Agnostic Depth Estimation. CPVR2024

Commercial Depth Sensors

[Park et al., CVPR 2024] Sensor-Agnostic Depth Completion

Structured Light

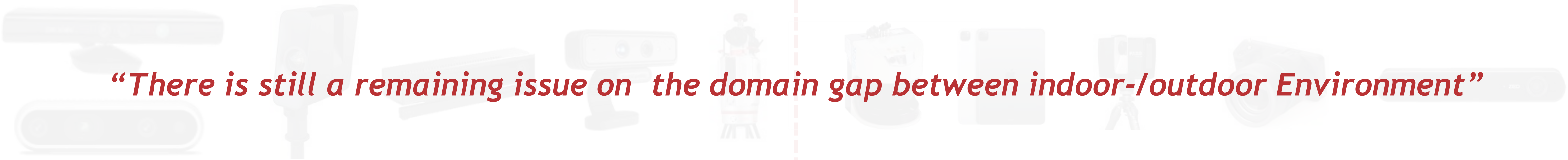
LiDAR

ToF (Time-of-Flight)

Stereo Camera

Depth Prompting for Sensor-Agnostic Depth Estimation. CPVR2024

Problem Definition



“There is still a remaining issue on the domain gap between indoor-/outdoor Environment”

Indoor Environment

Outdoor Environment

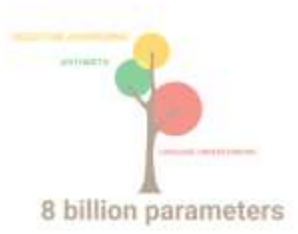
Problem Definition

Our solution is **a** ***universal framework*** *that can be …*

- Adapted for any off-the-shelf sensor
- Fine-tuned with minimal labels
- Boosted up with hyperbolic geometry and foundation knowledge

Universal Depth Completion

Idea 1. Foundational Knowledge



NLP: ChatGPT

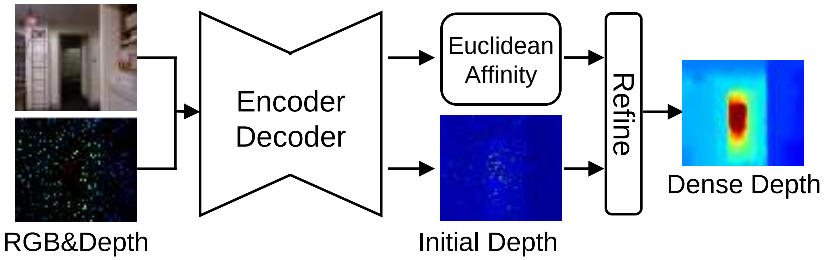
***Text & Image**: CLIP*

***Text to Video Generation**: Sora*

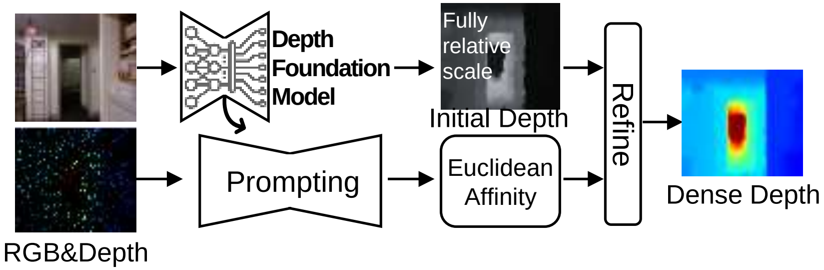
Foundation models are **large-scale pre-trained models** that learn from vast datasets, serving as a **versatile baseline** for various tasks in Natural Language Processing and Computer Vision.

Universal Depth Completion

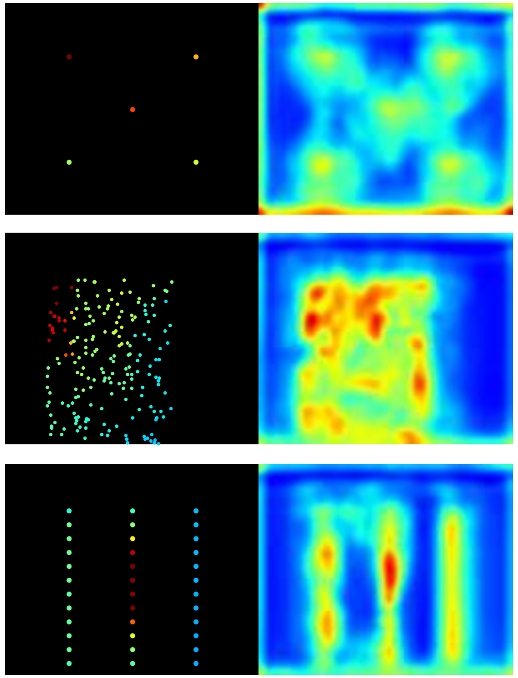
Idea 2. Prompting Depth



Conventional Model



DepthPrompting (CVPR24)



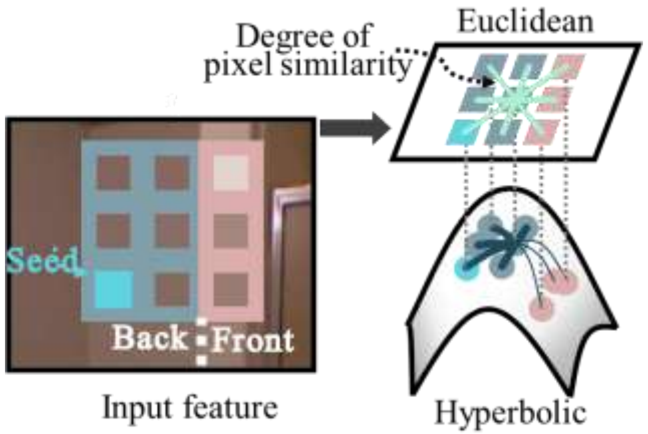
Depth input and feature visualization of prompt encoder.

Depth Prompting for Sensor-Agnostic Depth Estimation. CPVR2024

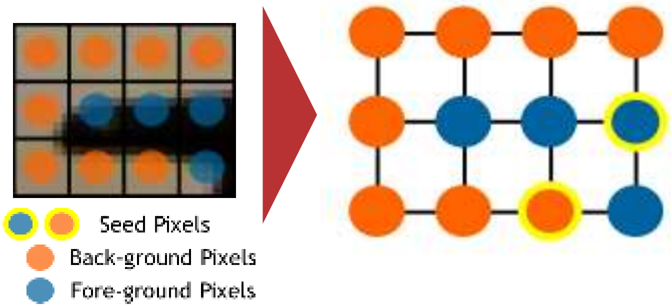
Universal Depth Completion

Idea 3. Hyperbolic Geometry

- Utilize **hyperbolic geometry** in pixel domain, which allows to construct a **hierarchical structure that serves as a continuous version of a tree.**



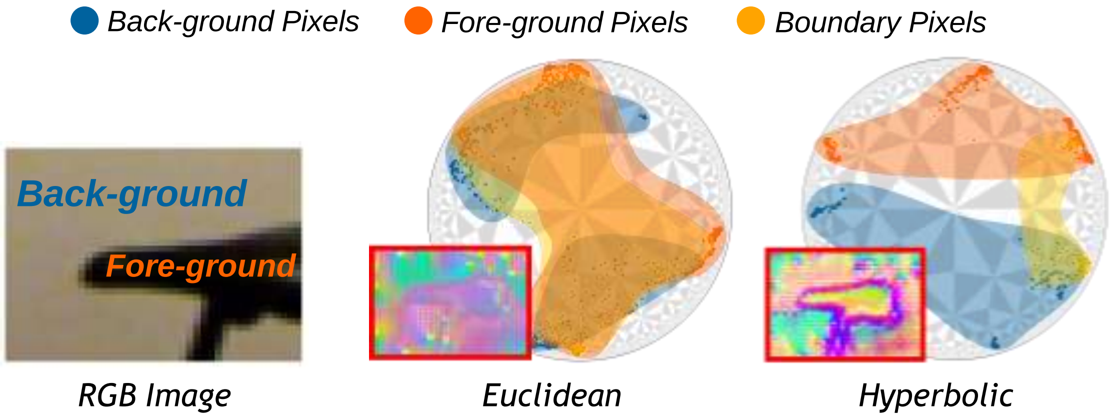
Hyperbolic embedding of pixel

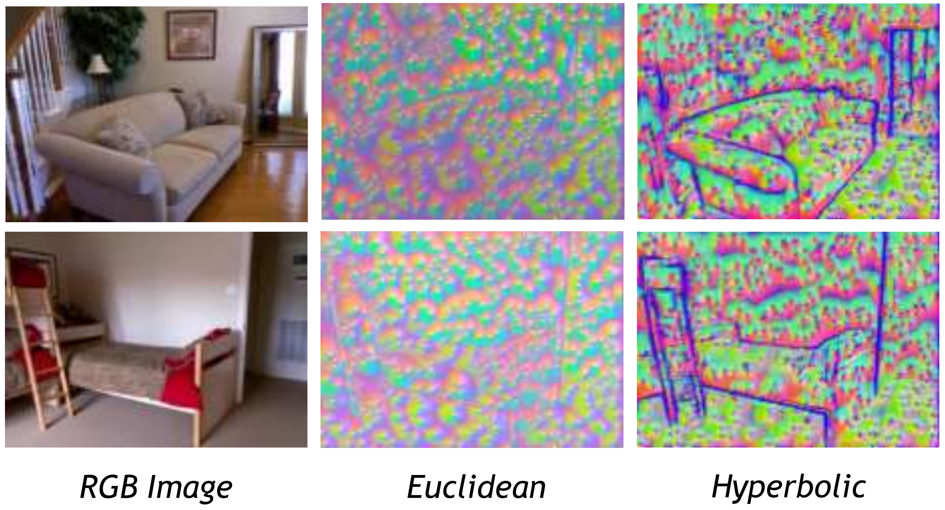


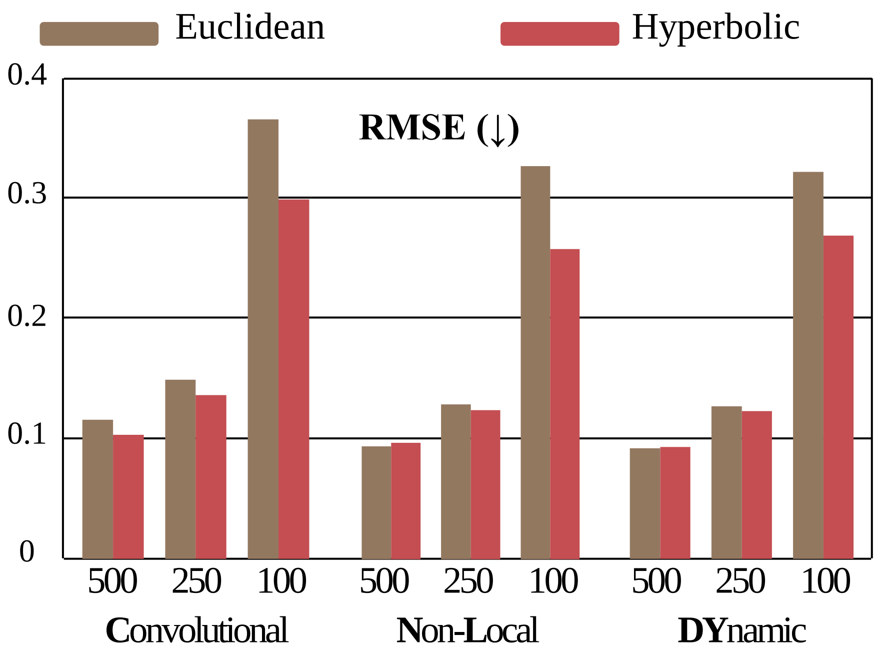
Local Connection -> Hierarchical Structure Graph

Universal Depth Completion

Idea 3. Hyperbolic Geometry





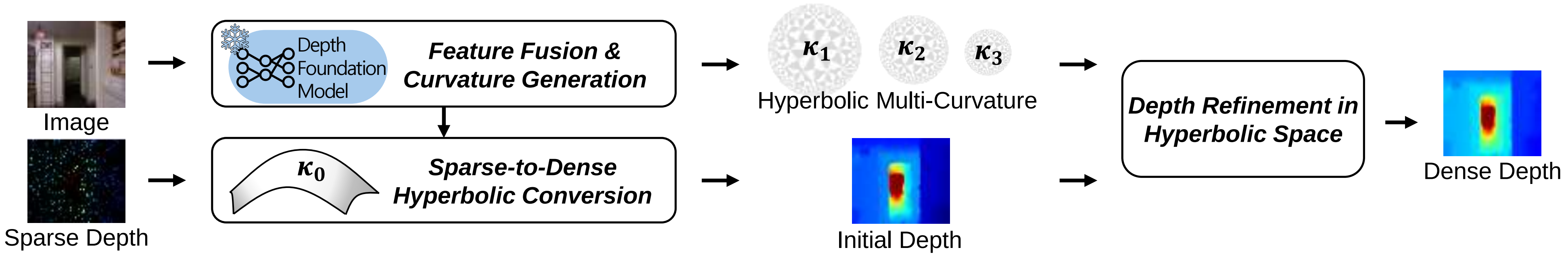


Comparison results for various spatial propagation schemes (Convolutional (**C**-SPN), Non-Local (**NL**-SPN), and DYnamic (**DY**-SPN)) w.r.t. the number of samples.

Learning Affinity with Hyperbolic Representation for Spatial Propagation. ICML2023

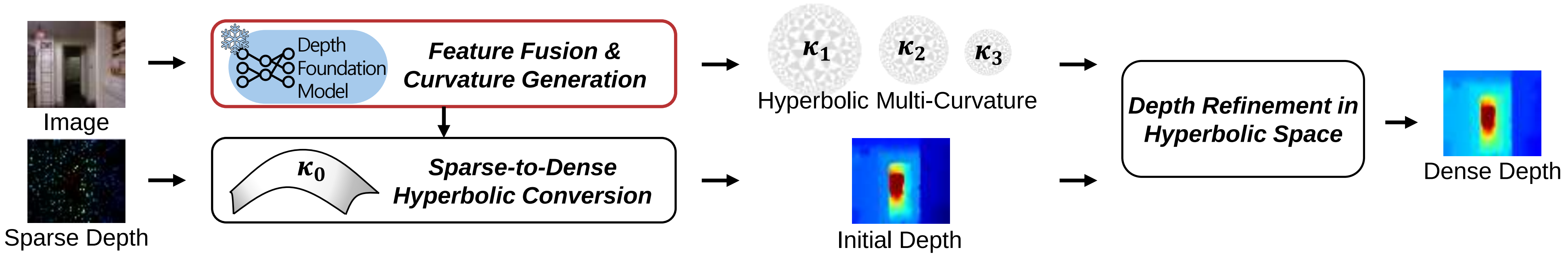
Universal Depth Completion

Methodology: UniDC



Universal Depth Completion

*Methodology: **UniDC***



Stage-① Multi-scale Feature Fusion & Hyperbolic Curvature Generation.

$E = f_{foundation}(I)$ .... Feature Extraction from Depth Foundation Model

$\boldsymbol{For}\ E_l\ \boldsymbol{in}\ E\ \boldsymbol{do} \quad (l = 0, \dots, L-1)$

$\quad E_{l+1}^M = f_l^{fusion}(E_l^M, E_{l+1})$ .... Multi-Scale Feature Aggregation

$\boldsymbol{End\ for}$

$\kappa = \mathcal{C}(E_L^M)$ .... Curvature Generation

I: RGB image

$f_{foundation}$: depth foundation model

$E_l \in E$: intermediate features

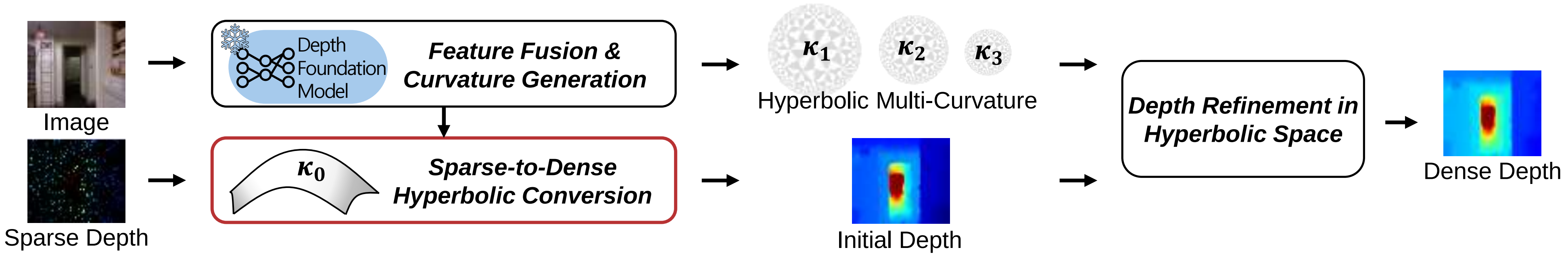
E_L^M: aggregated feature

$\mathcal{C}$: hyperbolic curvature Generator

κ: hyperbolic curvature

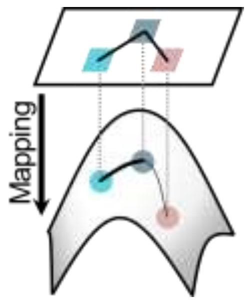
Universal Depth Completion

Methodology: UniDC



Stage-② Sparse-to-Dense Conversion based on Hyperbolic Features.

$H_i = \exp_0^{\kappa}(E_{L,i}^{M}), H_j = \exp_0^{\kappa}(E_{L,j}^{M})$ .... Hyperbolic Embedding



i: pixel coordinate

$Dist_{hyp}$: hyperbolic distance

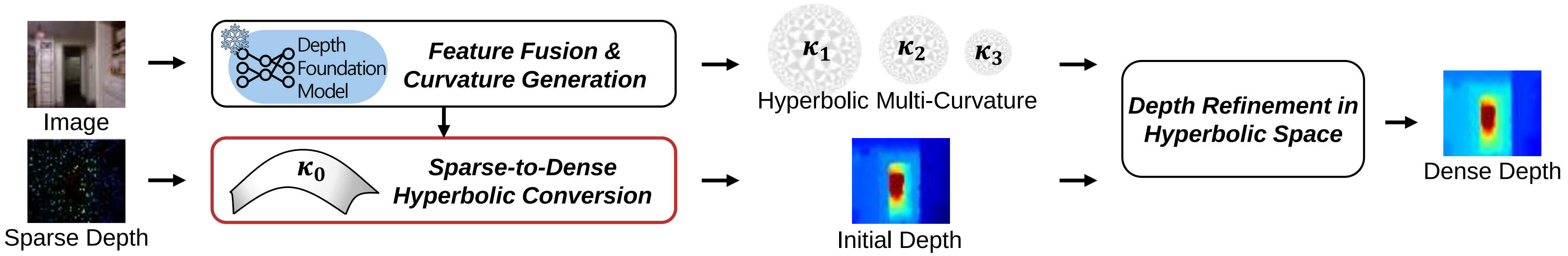
$Dist_{euc}$: Euclidean distance

S: Sparse Depth

$N(i)$: neighboring pixels

Universal Depth Completion

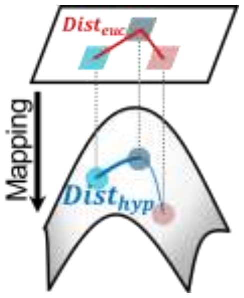
*Methodology: **UniDC***



Stage-② Sparse-to-Dense Conversion based on Hyperbolic Features.

$H_i = \exp_0^{\kappa}(E^M_{L,i}), H_j = \exp_0^{\kappa}(E^M_{L,j})$ .... Hyperbolic Embedding

$w_{ij} = \mathcal{P}(\mathbf{Dist}_{hyp}(H_i, H_j), \mathbf{Dist}_{euc}(E^M_{L,i}, E^M_{L,j}))$ .... Hyperbolic Bilateral Kernel



i: pixel coordinate

$Dist_{hyp}$: hyperbolic distance

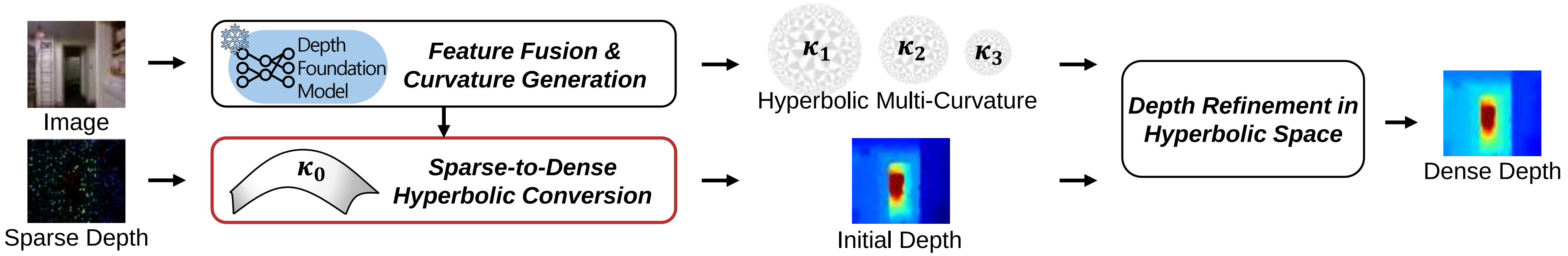
$Dist_{euc}$: Euclidean distance

S: Sparse Depth

$N(i)$: neighboring pixels

Universal Depth Completion

*Methodology: **UniDC***



Stage-② Sparse-to-Dense Conversion based on Hyperbolic Features.

$H_i = \exp_0^{\kappa}(E_{L,i}^{M}), H_j = \exp_0^{\kappa}(E_{L,j}^{M})$ .... Hyperbolic Embedding

$w_{ij} = \mathcal{P}(Dist_{hyp}(H_i, H_j), Dist_{euc}(E_{L,i}^{M}, E_{L,j}^{M})$ .... Hyperbolic Bilateral Kernel

$D_i^{init} = \sum_{j \in N(i)} w_{ij} S_j$ .... Calculate Initial Depth

i: pixel coordinate

$Dist_{hyp}$: hyperbolic distance

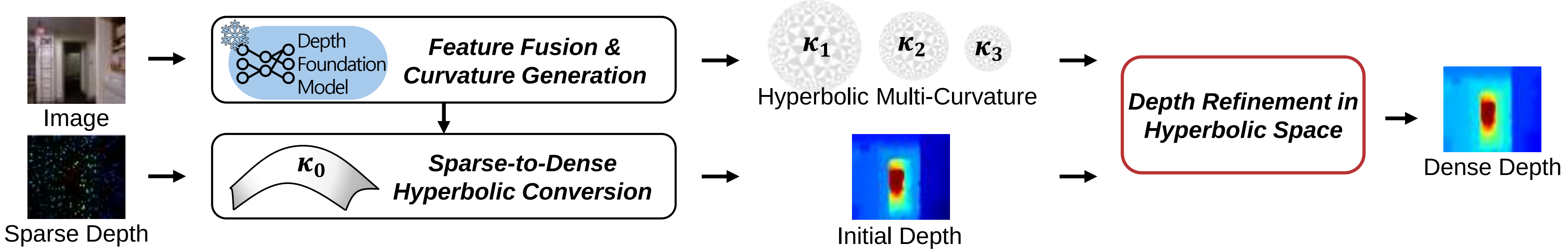
$Dist_{euc}$: Euclidean distance

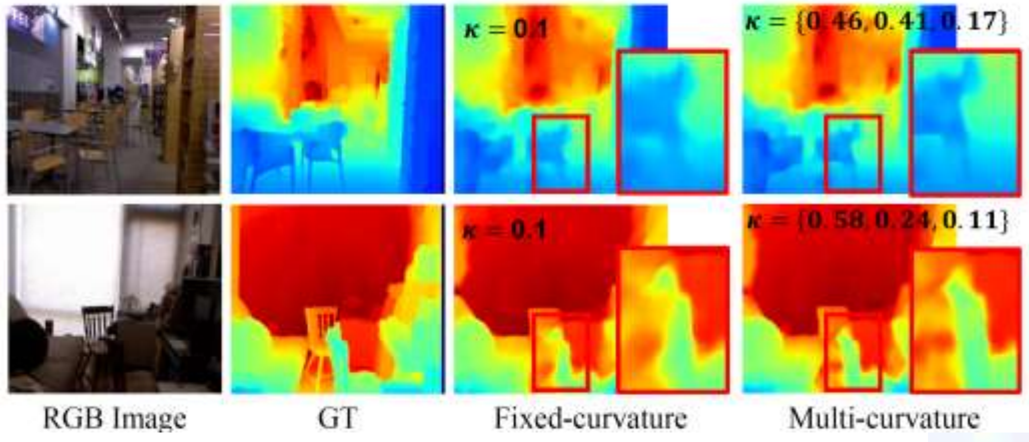
S: Sparse Depth

$N(i)$: neighboring pixels

Universal Depth Completion

*Methodology: **UniDC***





Fixed and Multi-Curvature Results

HCL: hyperbolic convolution layer

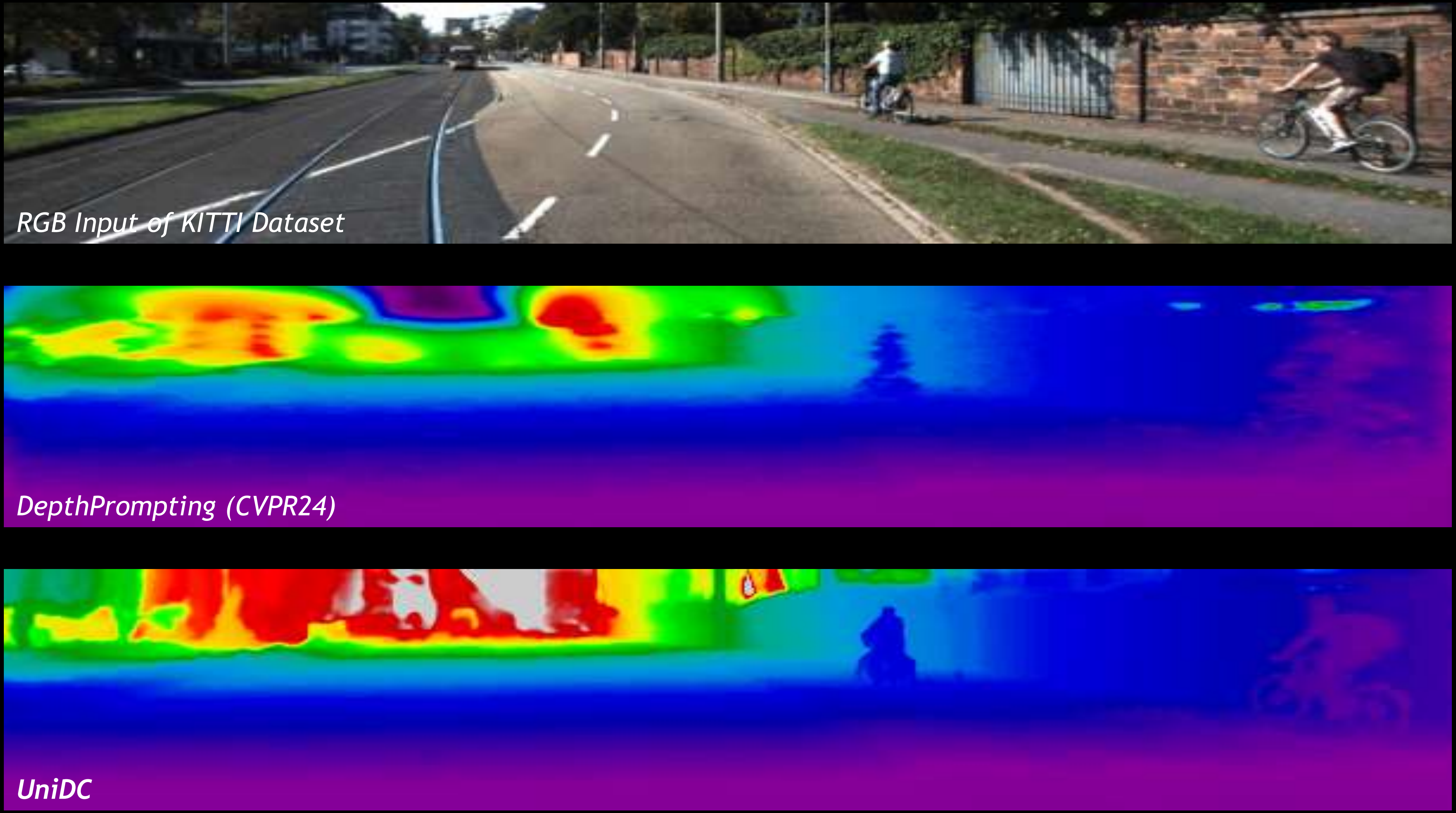
SP: Spatial Propagation [CSPN++]

$A_{\kappa_k}^{hyp}$: Hyperbolic affinity for κ_k

$\kappa_k \in \{\kappa_1, \kappa_2, \kappa_3\}$

[CSPN++] Learning context and resource aware convolutional spatial propagation networks for depth completion. AAAI2020

Experimental results of 100-shot on KITTI dataset.



Commercial Depth Sensors

Indoor Environment

Outdoor Environment

Universal Depth Completion

"Universal Depth Model
for arbitrary sensors and environments"

Summary

Our UniDC …

1. Bridges gaps across sensors and environments.
2. Leverages foundational knowledge, depth prompting, and hyperbolic geometry.
3. Delivers strong results with minimal labeled data.
4. Enables efficient, adaptable depth perception for diverse applications.

Thank you for your attention !

Look forward to any questions you may have.

- github.com/JinhwiPark/UniDC
- www.jinhwipark.com