

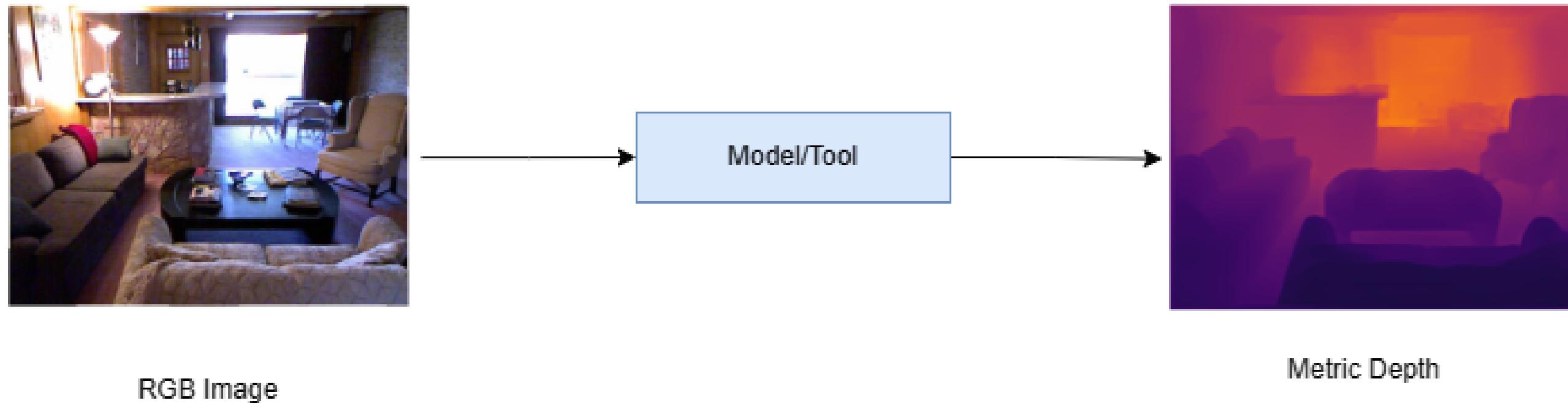
TPA 10: True Depth Estimation from Indoor Scenes

Presenters : Sujay S | Amruth Vardhan

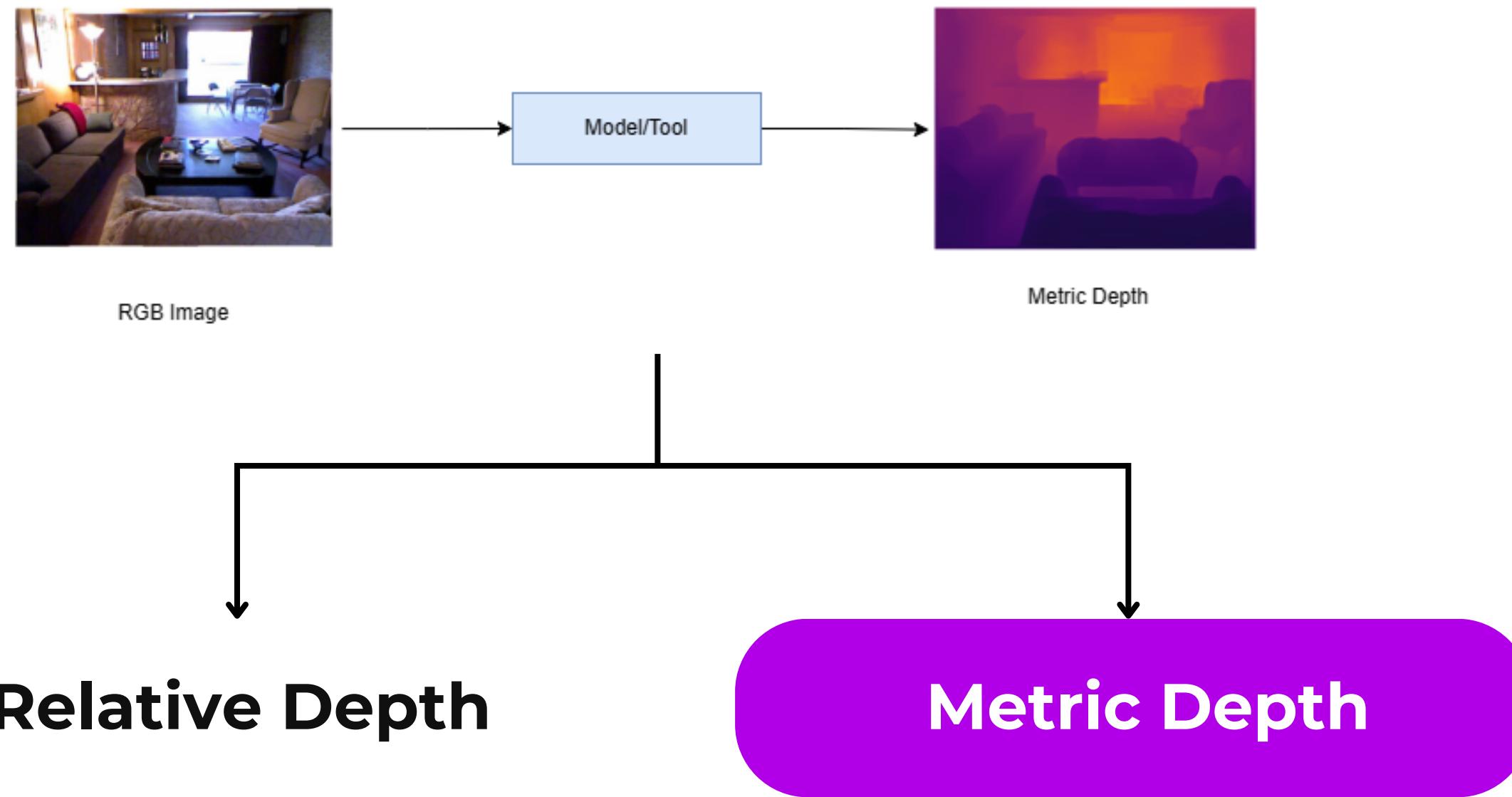


Problem Statement

Design a model that can estimate real-world distances for every pixel in the input image from the camera sensor



Depth Estimation



ZoeDepth, Adabins, Localbins, Pixelformer

Applications

01

Autonomous Vehicles and Robotics

: Monocular depth estimation is crucial for autonomous navigation systems in vehicles and robots. It helps in understanding the environment, detecting obstacles, and planning safe paths by estimating the distance of objects from the vehicle.

02

3D Reconstruction and Modeling:

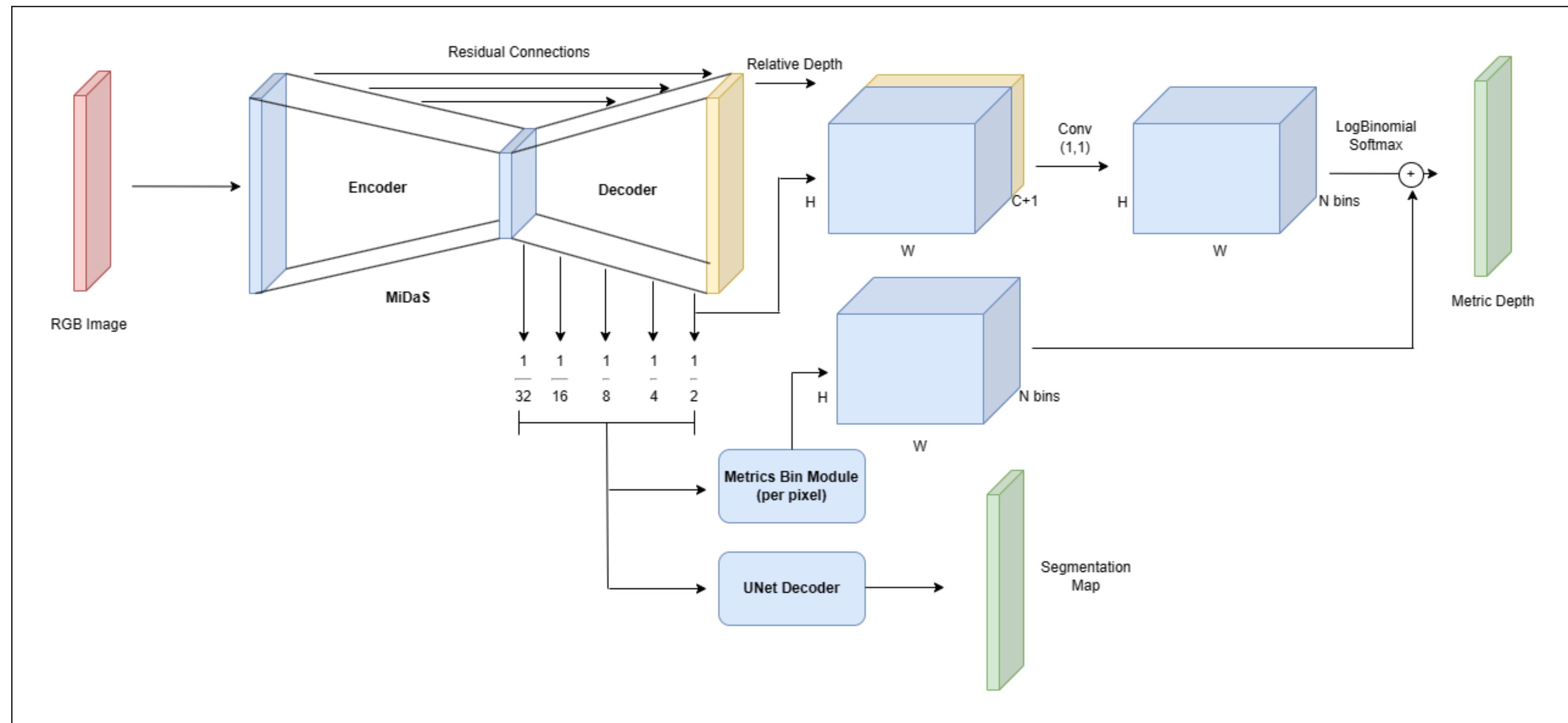
Monocular depth estimation contributes to creating 3D models of scenes or objects from 2D images. It helps in applications such as cultural heritage preservation, architecture, and virtual museum experiences.

03

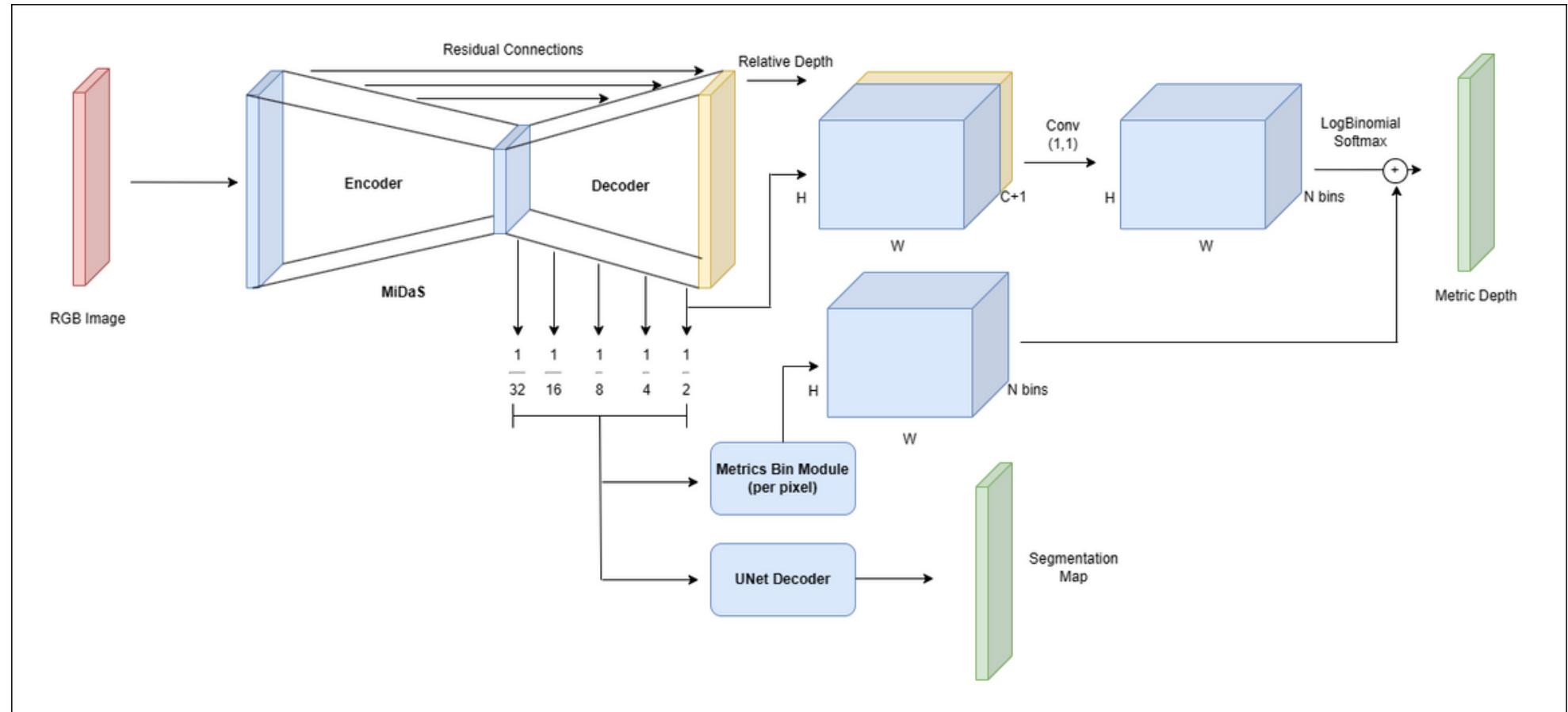
Medical Imaging and Healthcare

In medical imaging, estimating depth can aid in analyzing and segmenting organs or tumors, assisting in surgery planning, and providing depth cues for various diagnostic purposes.

Model Architecture



Model Components



01

**ZoeDepth
Network**

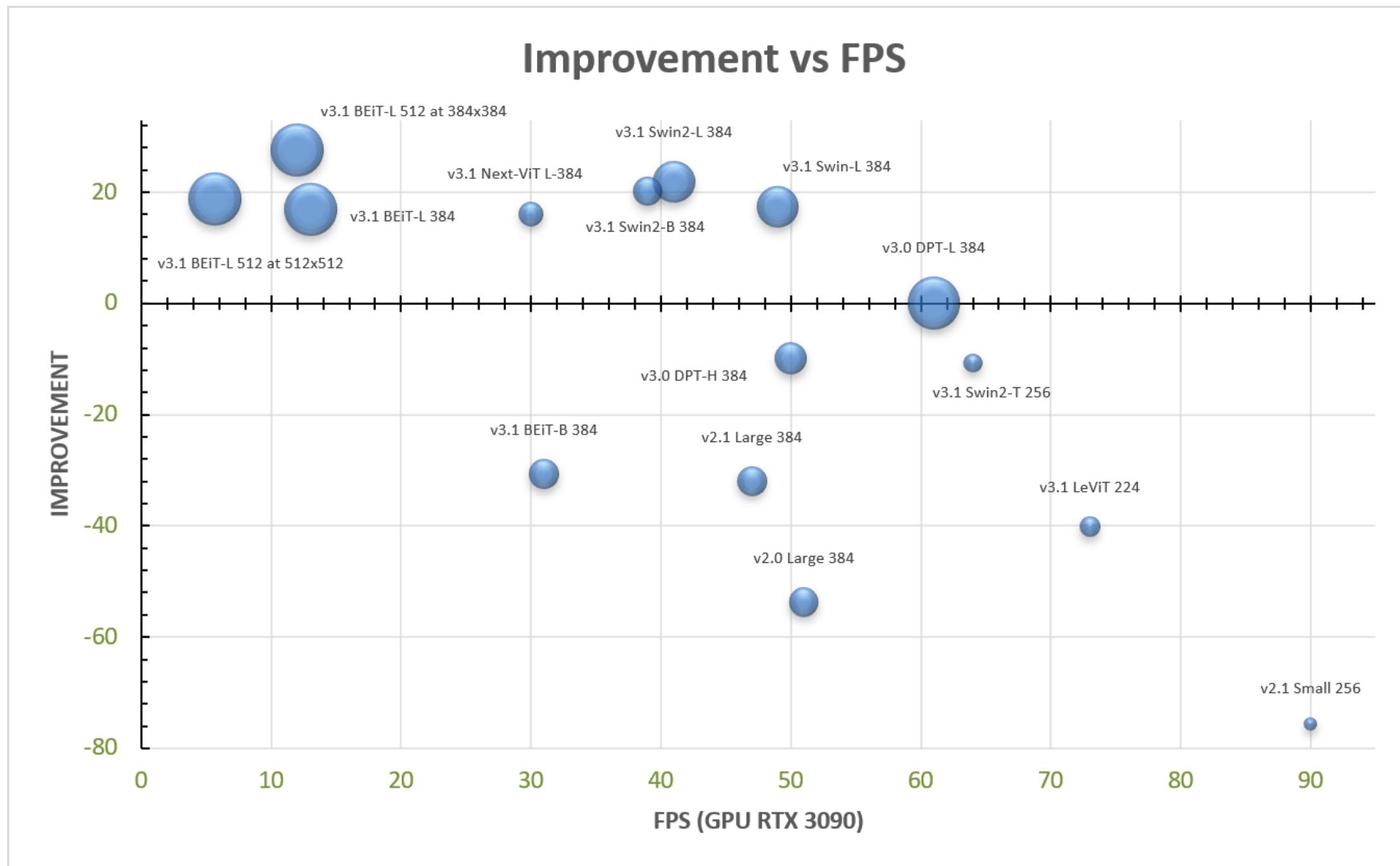
02

**Segmentation
Distillation**

(Inactive during inference)

ZoeDepth

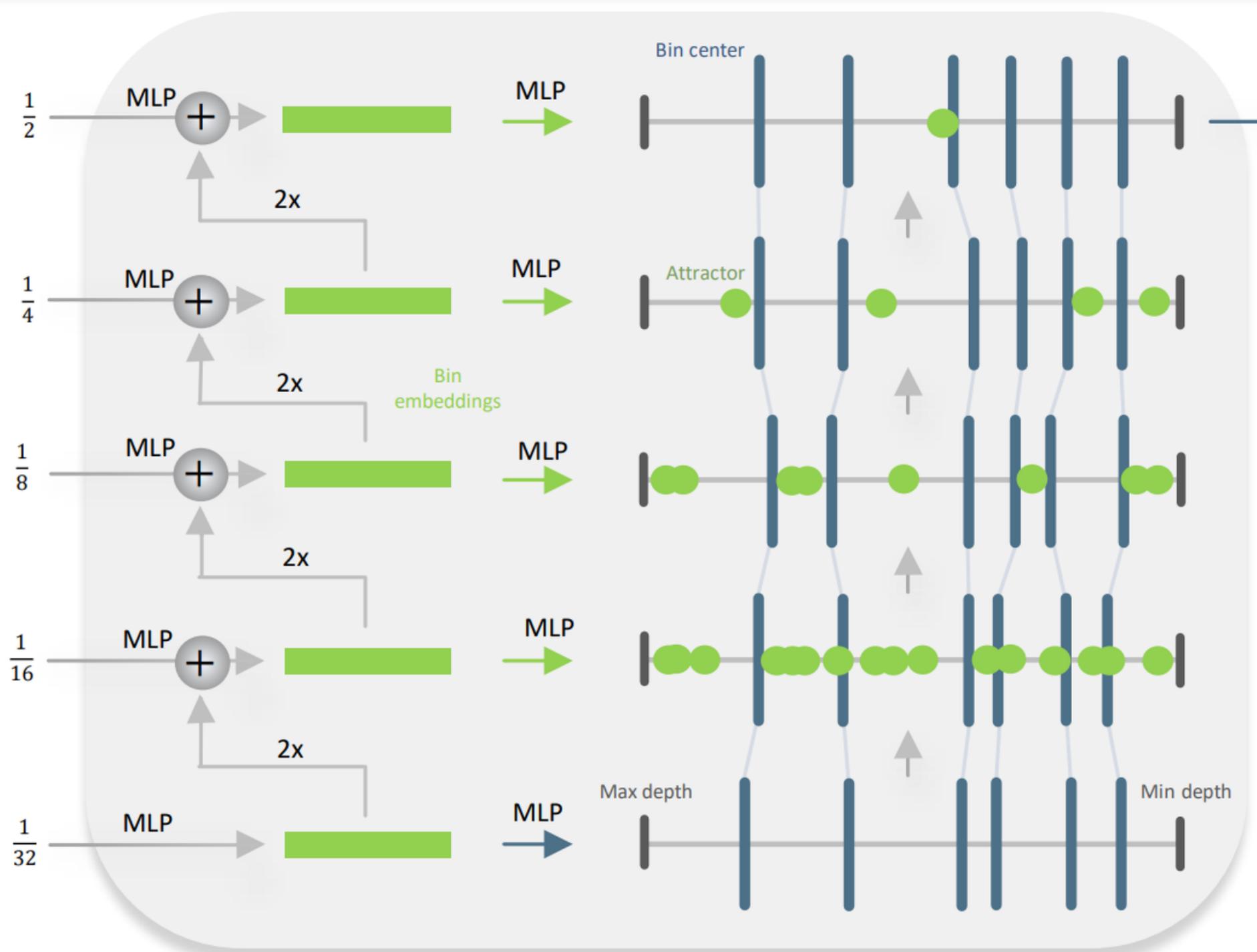
Core : SwinV2 Transformer



Model	Parameters	Image Resolution
SwinV2_B_384	102M	(384,384)

ZoeDepth

Metric Bins Module

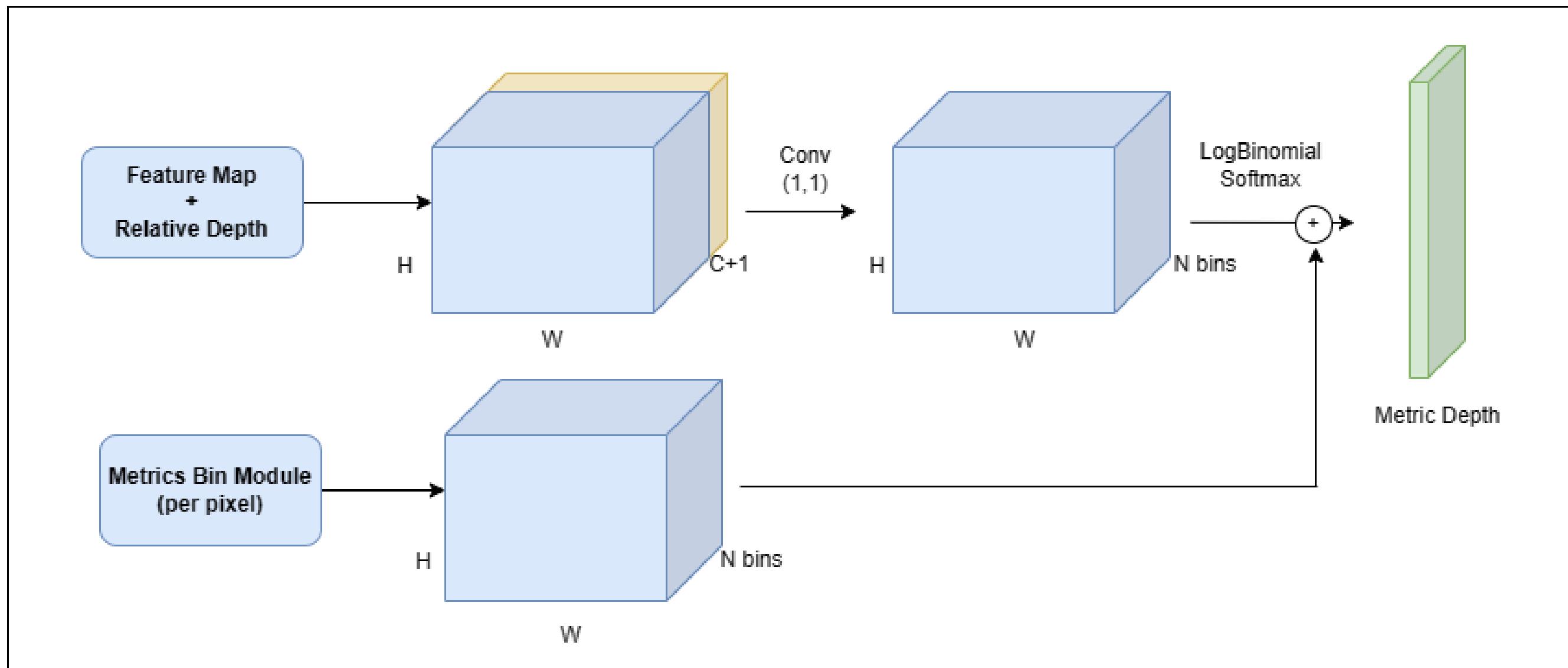


$$\Delta c_i = \sum_{k=1}^{n_a} \frac{a_k - c_i}{1 + \alpha |a_k - c_i|^\gamma}$$

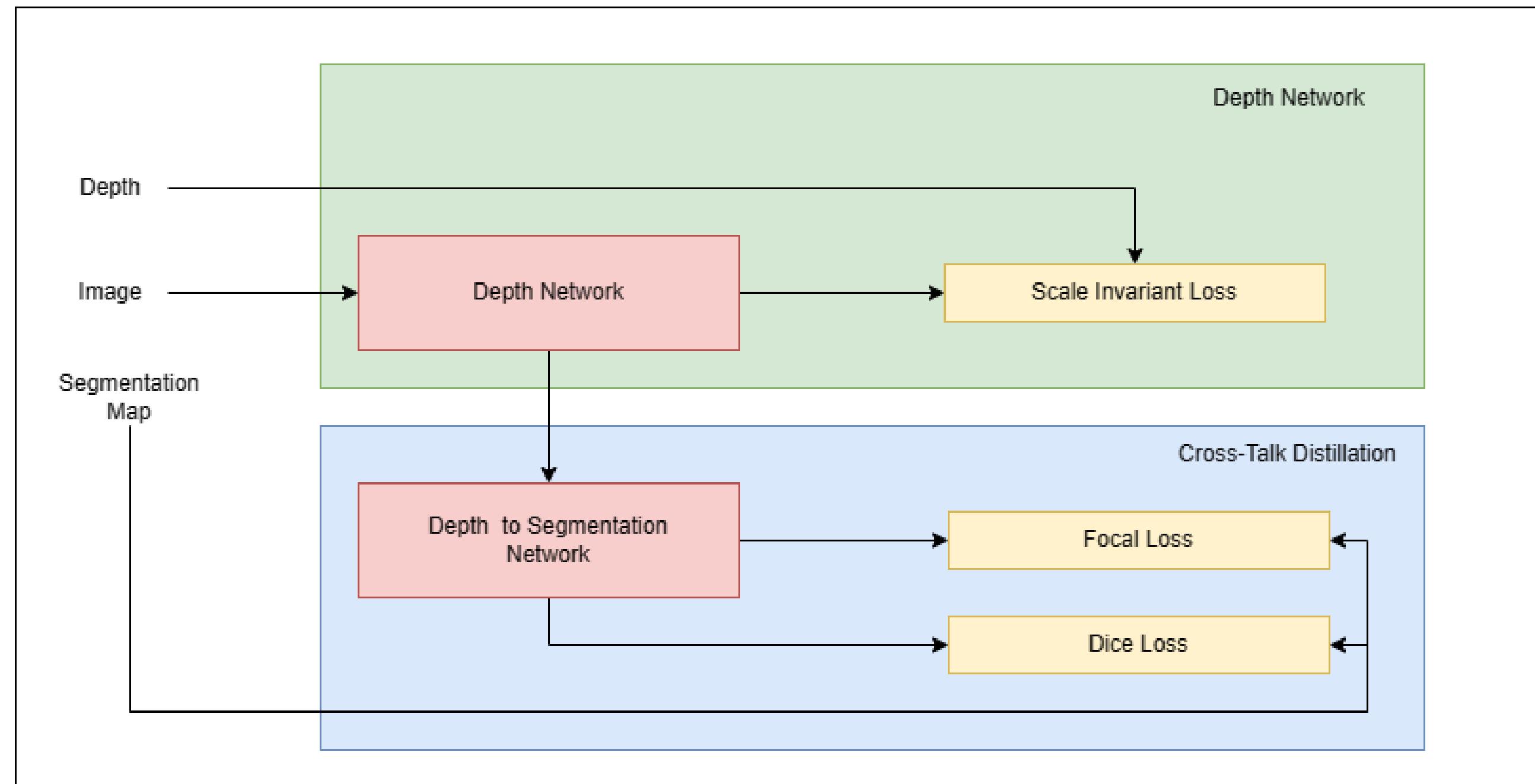
Hyperparameters for Bin Metric Module			
n_bins	bin_embedding_dim	bin_centers_type	n_attractors
64	128	softplus	[16, 8, 4, 1]
attractor_alpha	attractor_gamma	attractor_kind	attractor_type
1000	2	mean	inv

ZoeDepth

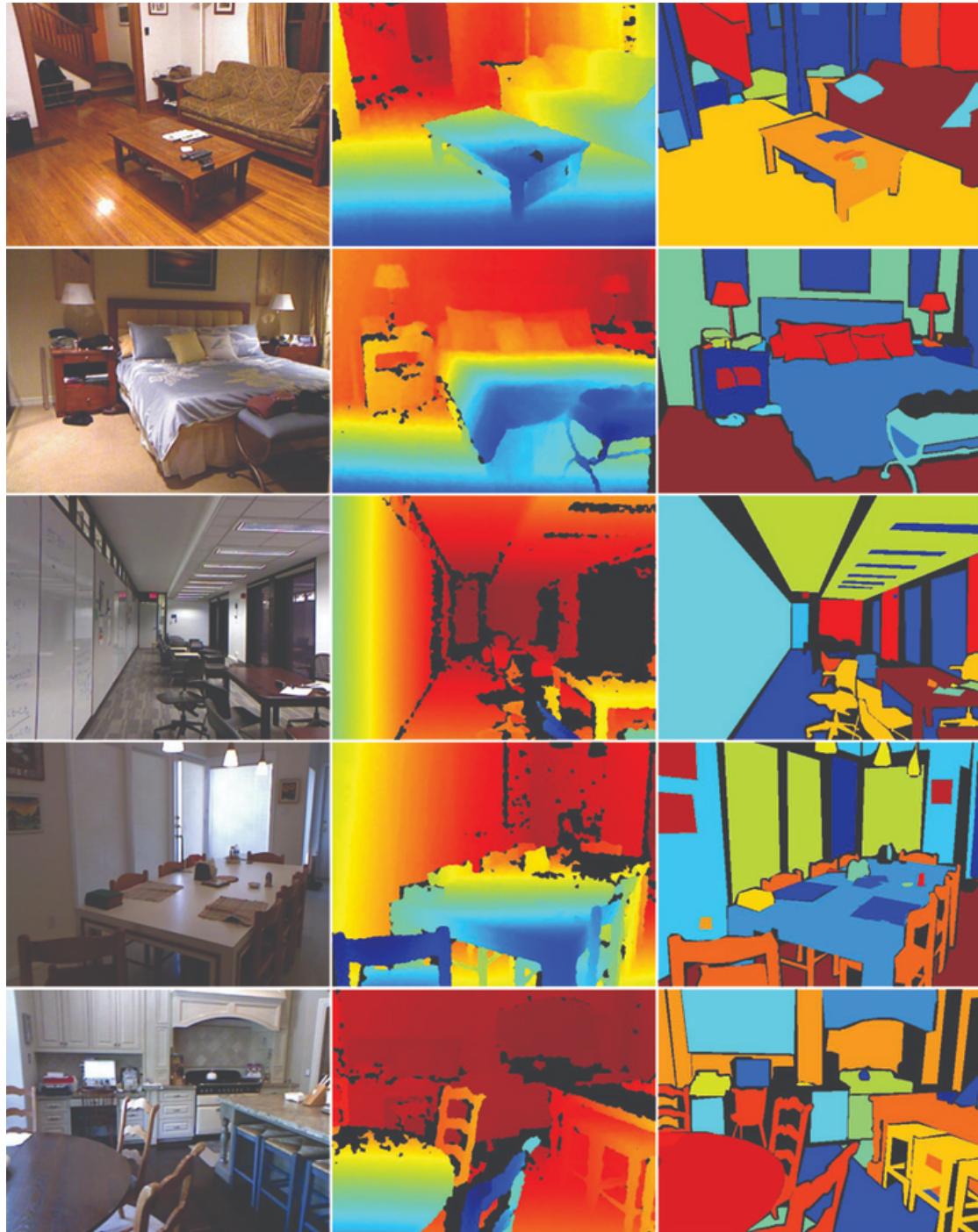
Metric Depth Head



Segmentation Distillation



Dataset



NYU V2 Dataset

01

Indoor Scenes

02

Resolution: 640 x 480

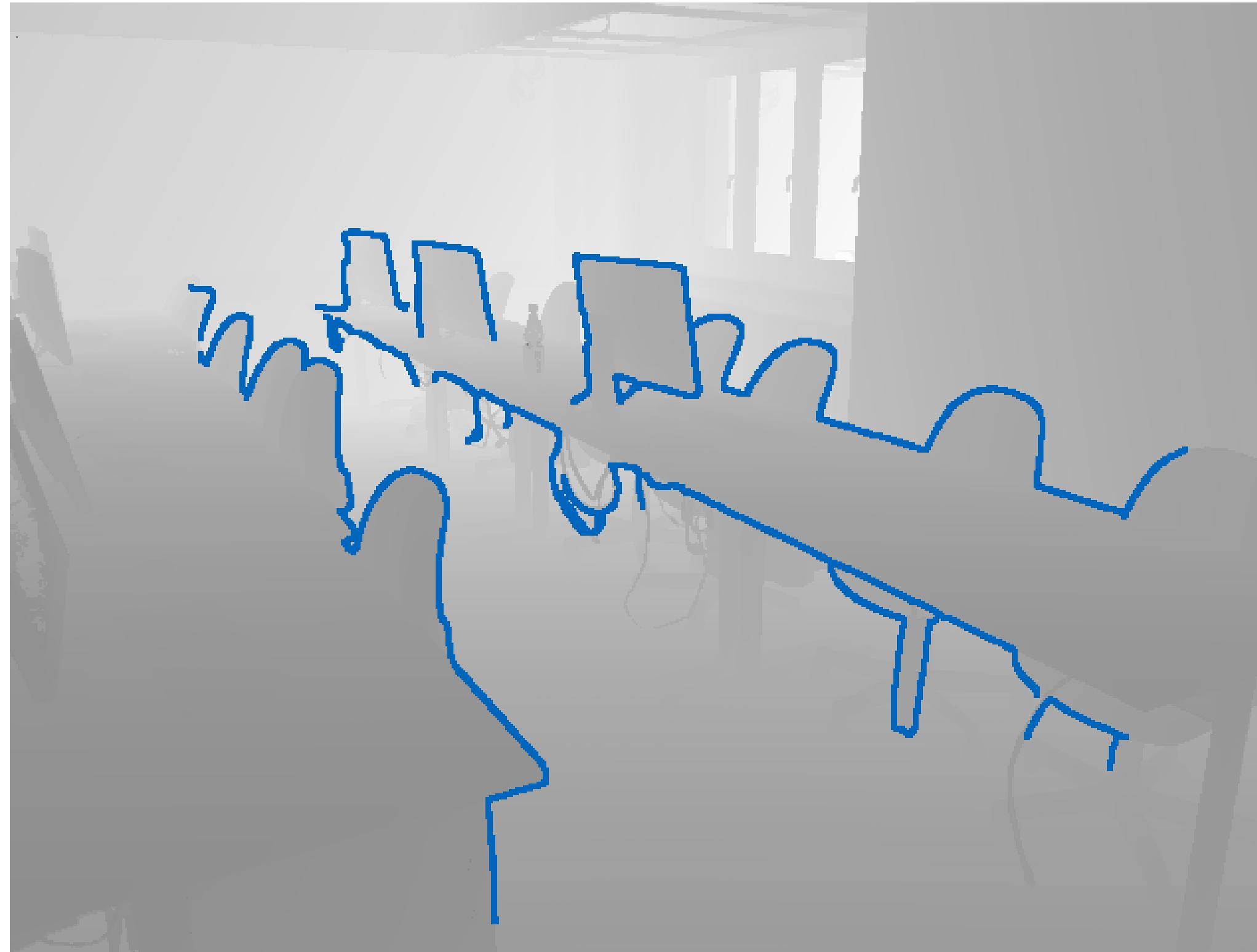
03

Range: [0,10m]

04

1449 labelled pairs

Dataset



iBims-1 Dataset

- 01 Indoor Scenes
- 02 Resolution: 640 x 480
- 03 Range: [0,10m]
- 04 100 labelled pairs

LOSS



Depth Loss

Scale Invariant Loss



Segmentation Loss

Focal Loss + Dice Loss

$$L_{final} = \lambda_{dis} \cdot L_{segmentation} + L_{pixel}$$

Training

01

Input Image Resize (384 x 384)

02

**Trained for 100 epochs
with Batch Size of 3
(NYU V2)**

03

**Adam Optimiser with Cyclic LR
Scheduler**

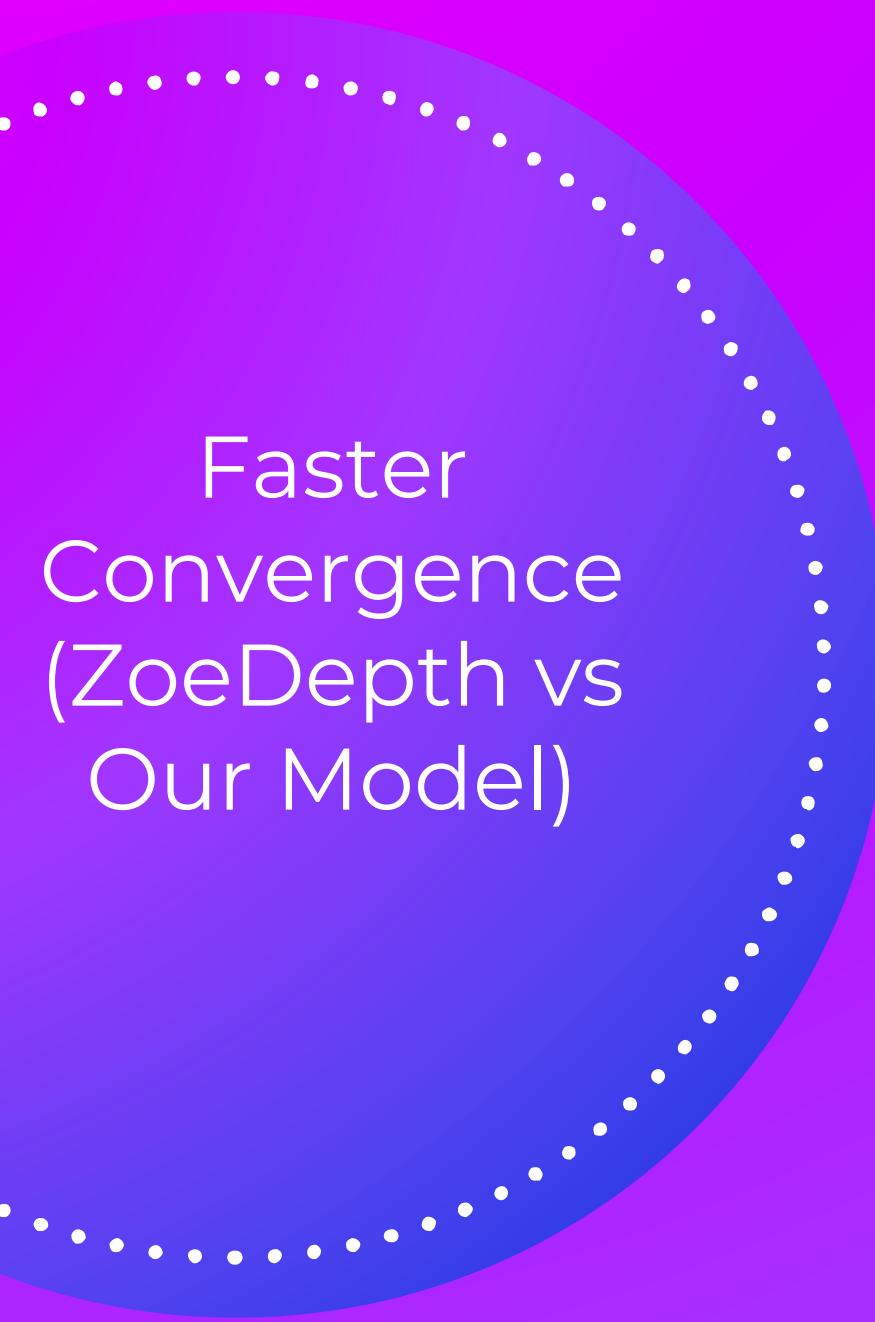
04

Validation every 25% of an epoch

05

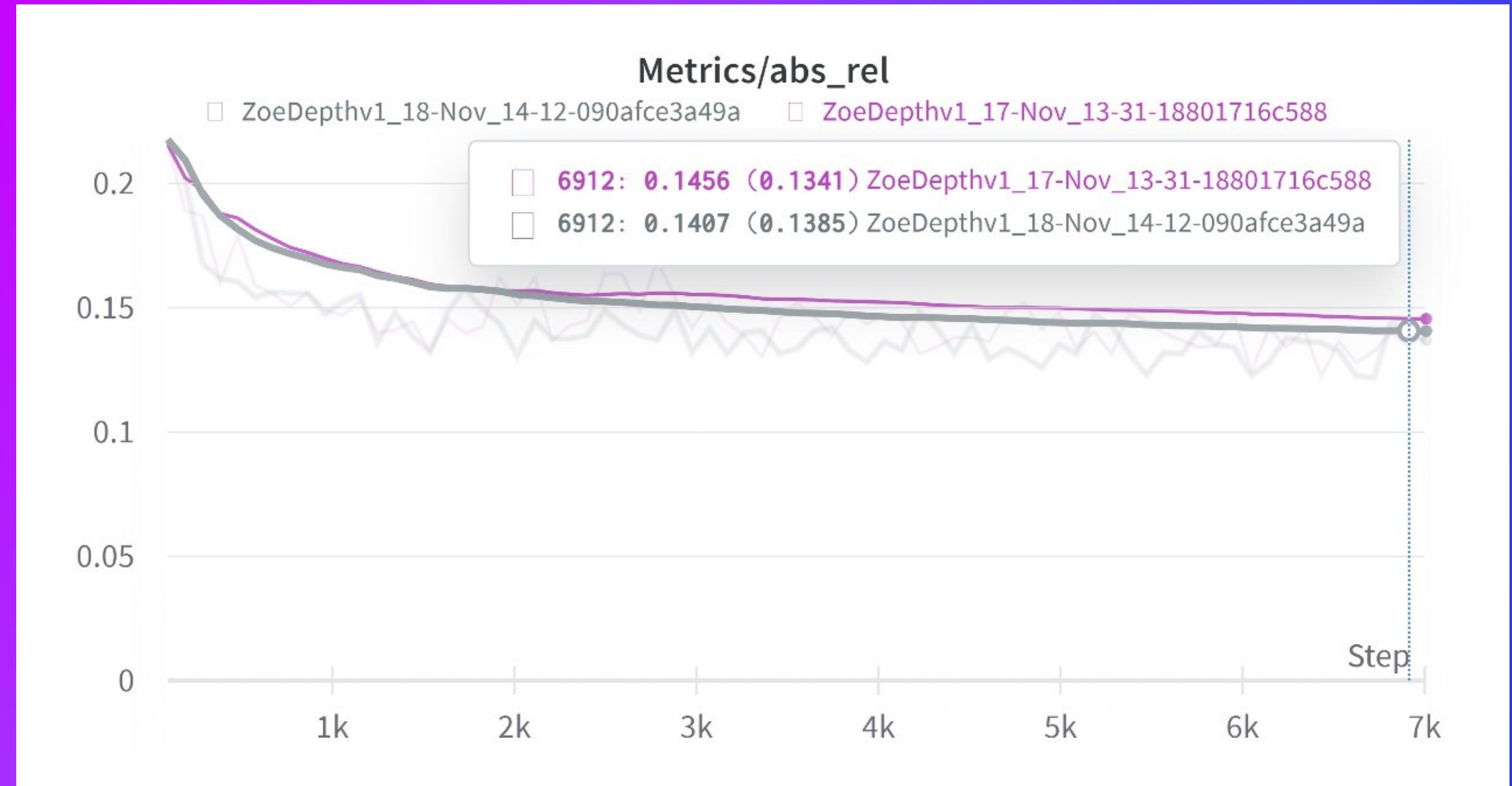
Trained on NVIDIA T4 GPU (16 GB)

Results



Faster
Convergence
(ZoeDepth vs
Our Model)

Grey - Our model; Pink-ZoeDepth



Note: This plot was obtained from weights and biases log during training run

NYU V2

Evaluation
Metrics
(Quantitative
Comparison)

Model	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	$REL \downarrow$	$RMSE \downarrow$	$\log_{10} \downarrow$
Adabins[11]	0.903	0.984	0.997	0.103	0.364	0.044
NeWCRFs[12]	0.922	0.992	0.998	0.095	0.334	0.041
ZoeDepth (S2-B)[1]	0.927	0.992	0.999	0.090	0.313	0.038
ZoeDepth (B-L)[1]	0.955	0.995	0.999	0.075	0.270	0.032
Our model	0.980	0.998	0.999	0.048	0.2	0.021

iBims

Evaluation
Metrics
(Quantitative
Comparison)

Model	$\delta_1 \uparrow$	$\delta_2 \uparrow$	$\delta_3 \uparrow$	REL \downarrow	RMSE \downarrow	$\log_{10} \downarrow$
Adabins[11]	0.555	0.873	0.960	0.212	0.901	0.107
NeWCRFs[12]	0.548	0.884	0.979	0.206	0.861	0.102
ZoeDepth (B-L)[1]	0.658	0.947	0.985	0.169	0.711	0.083
Our model	0.669	0.924	0.972	0.182	0.922	0.086

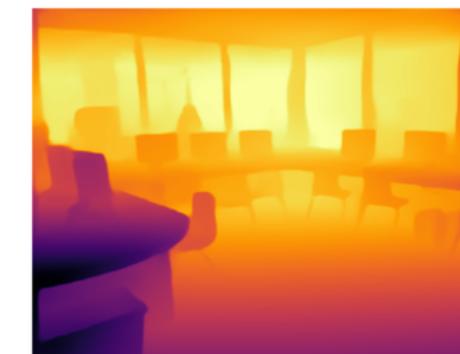
NYU V2

Plots

RGB Image



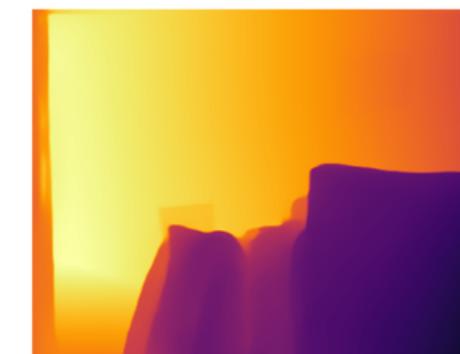
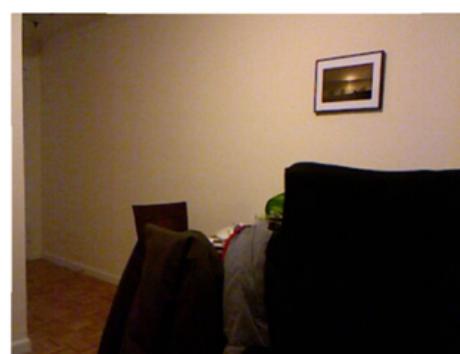
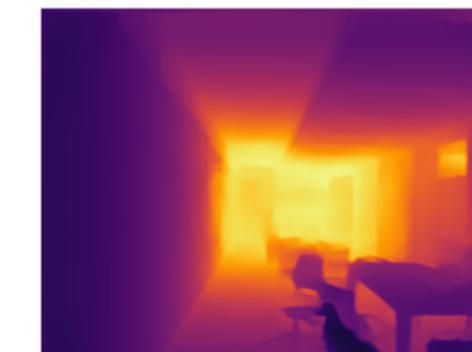
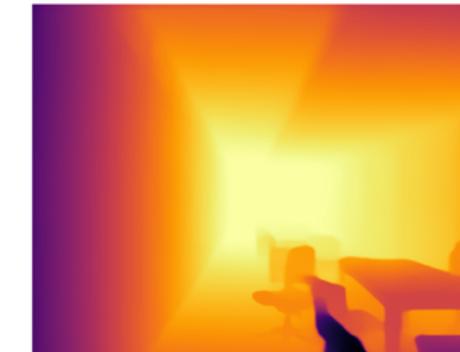
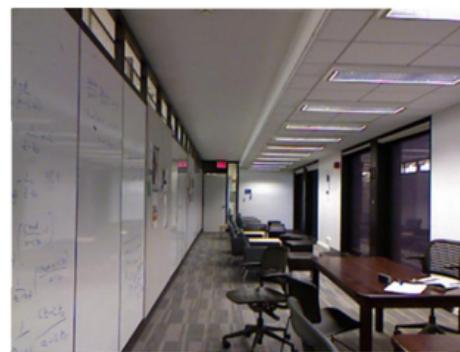
Relative Depth



Metric Depth -
Ground Truth



Metric Depth -
Prediction



NYU V2

Plots

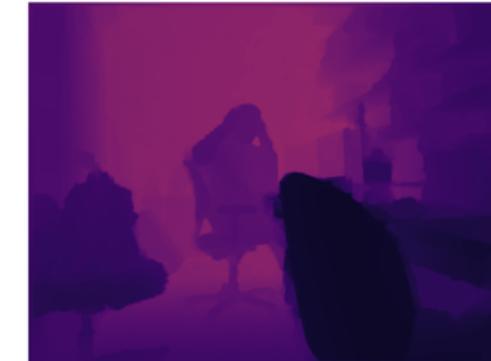
RGB Image



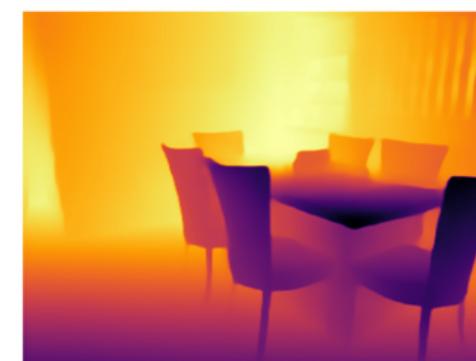
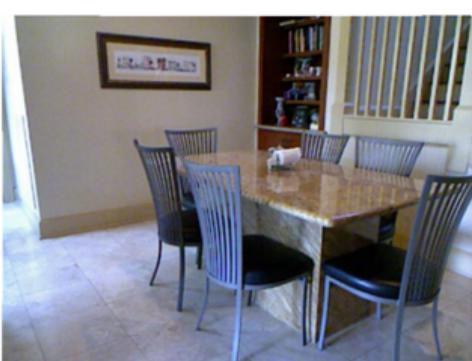
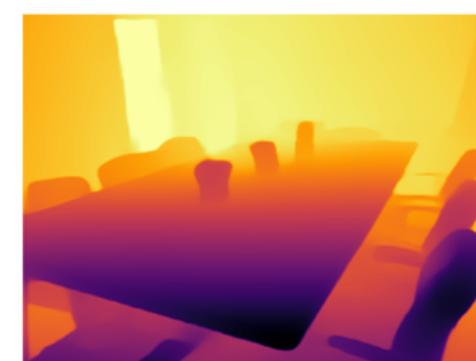
Relative Depth



Metric Depth -
Ground Truth



Metric Depth -
Prediction



iBims

Plots

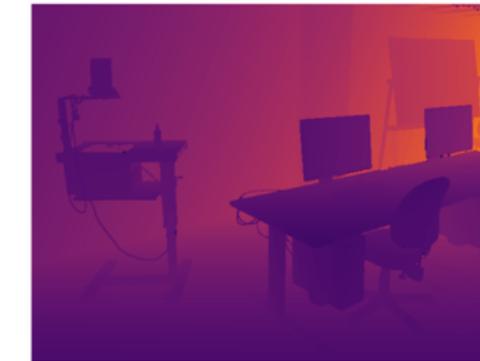
RGB Image



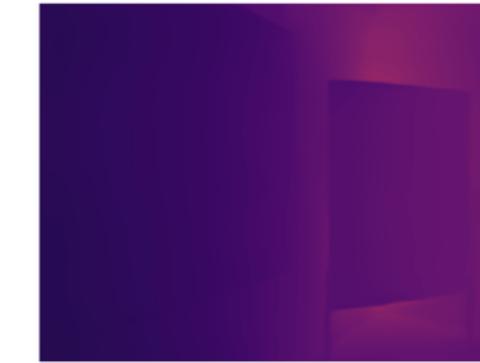
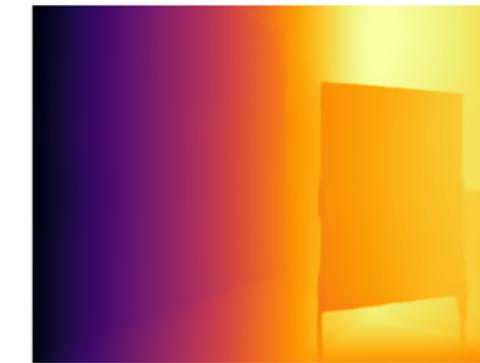
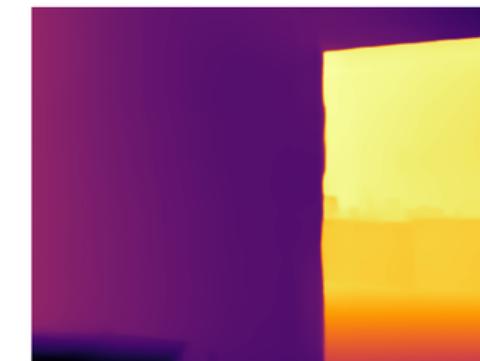
Relative Depth



Metric Depth -
Ground Truth



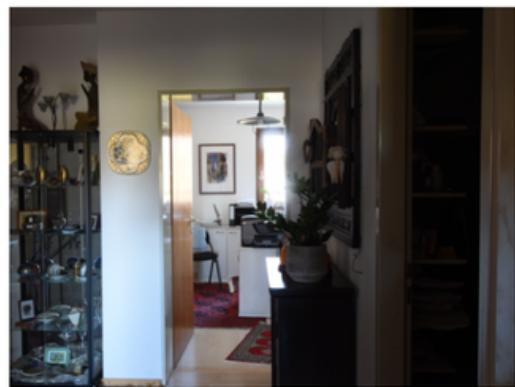
Metric Depth -
Prediction



iBims

Plots

RGB Image



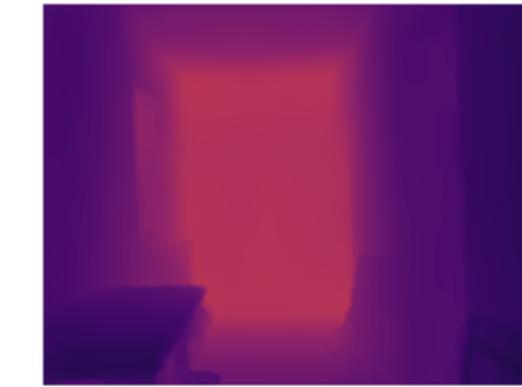
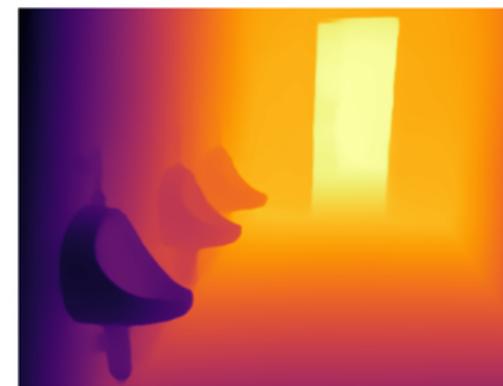
Relative Depth



Metric Depth -
Ground Truth



Metric Depth -
Prediction

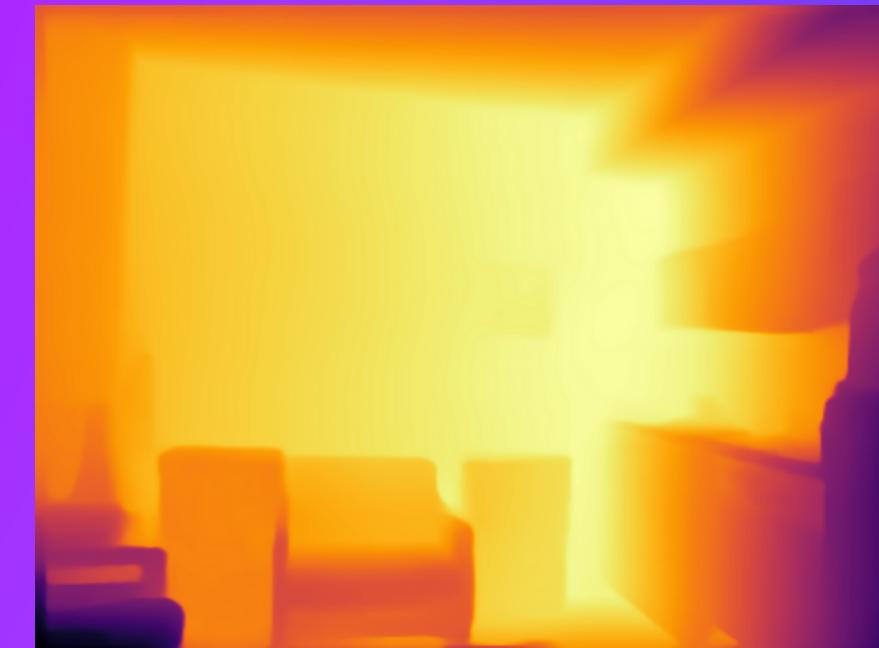


Test Images

RGB Image



Relative Depth



Metric Depth

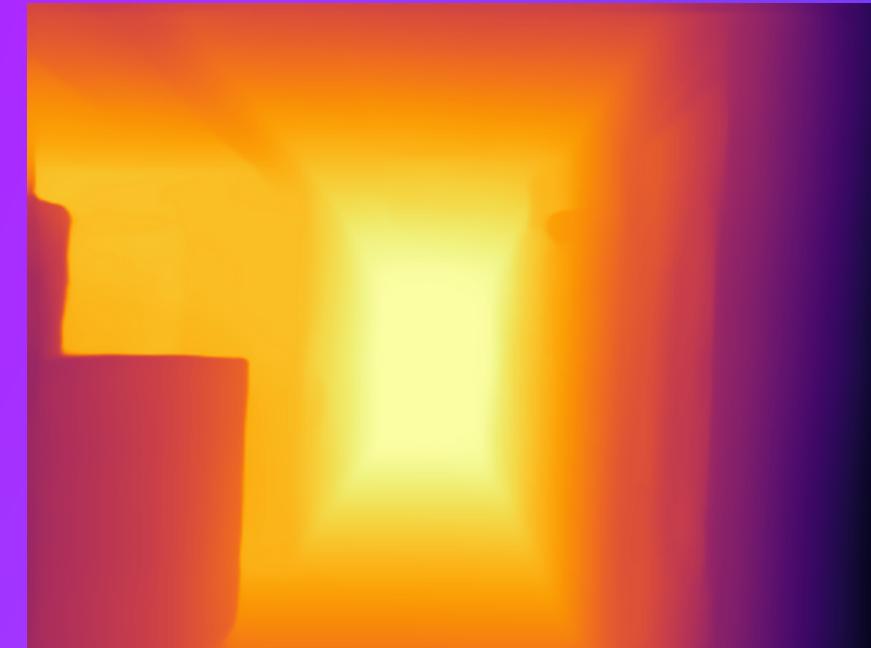


Test Images

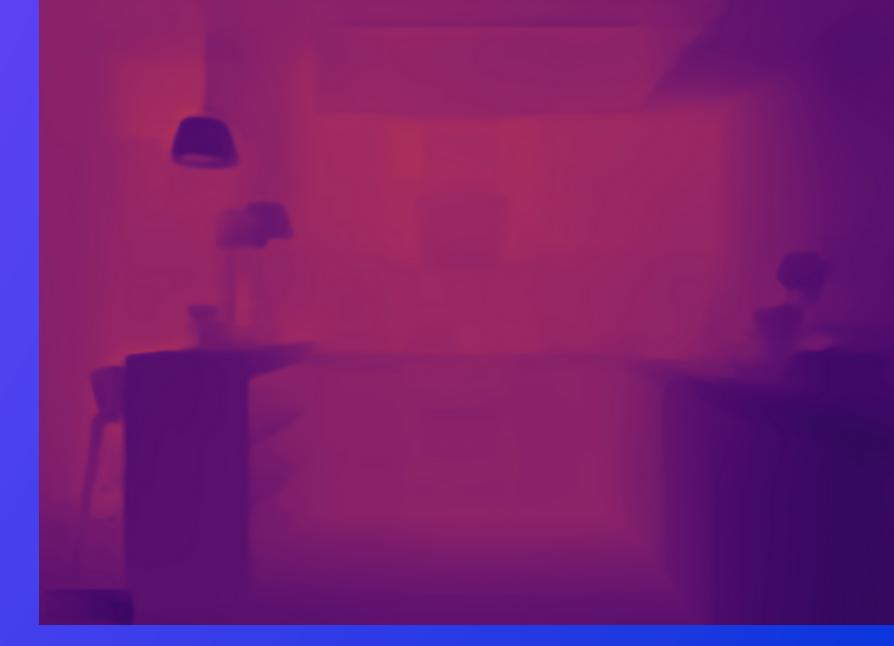
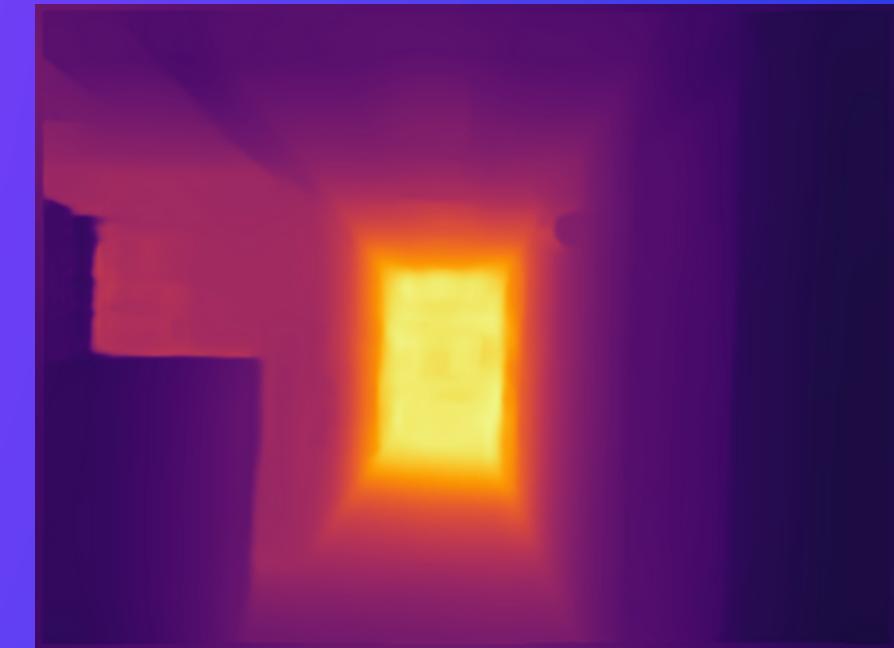
RGB Image



Relative Depth



Metric Depth



Questions?

