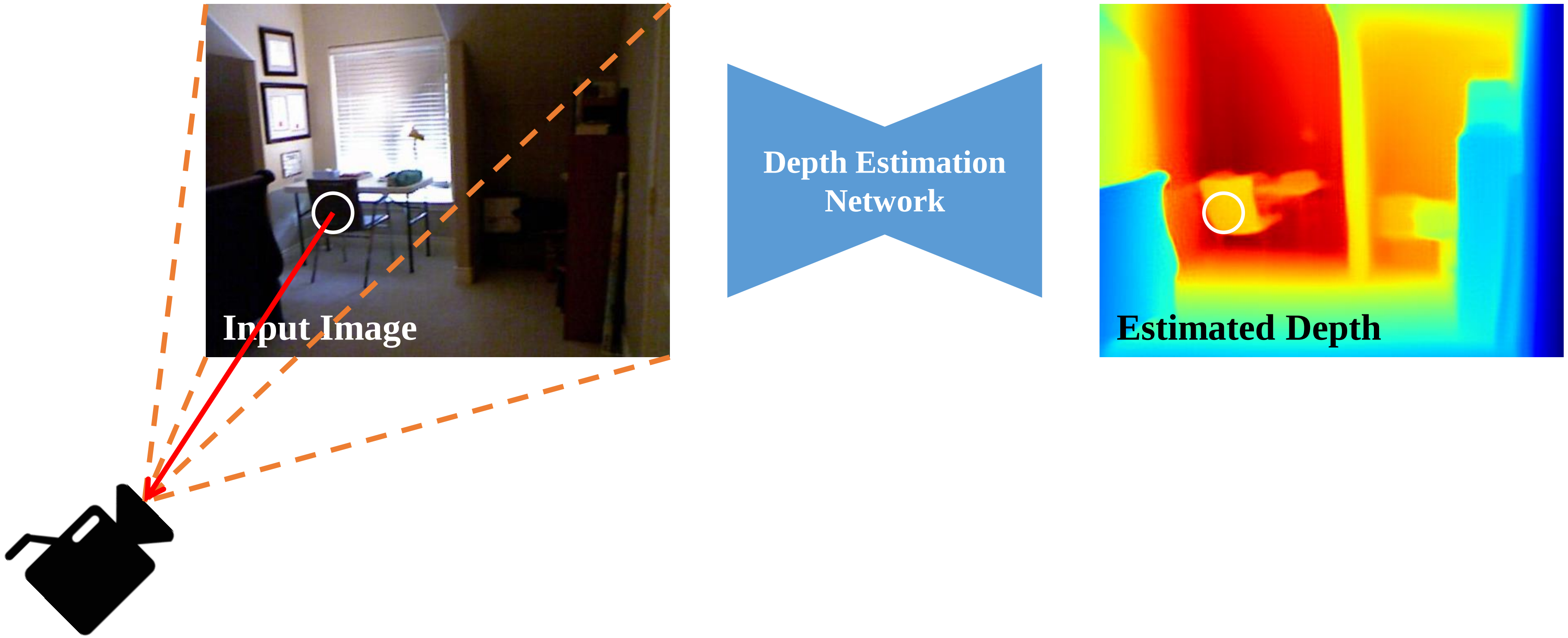
DCDepth: Progressive Monocular Depth Estimation in Discrete Cosine Domain

Kun Wang, Zhiqiang Yan, Junkai Fan, Wanlu Zhu, Xiang Li, Jun Li and Jian Yang

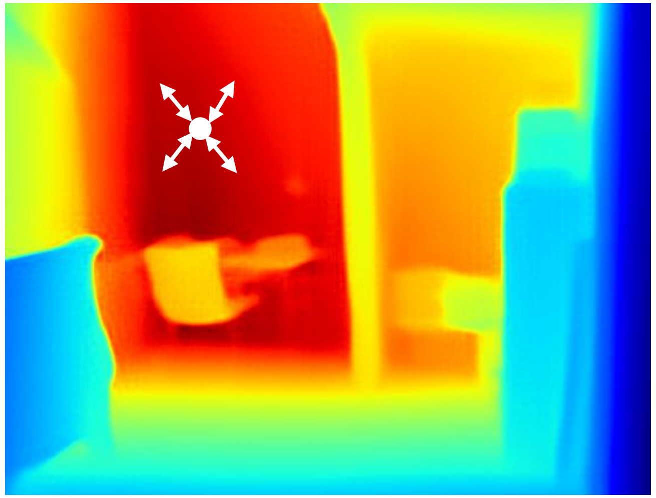
南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

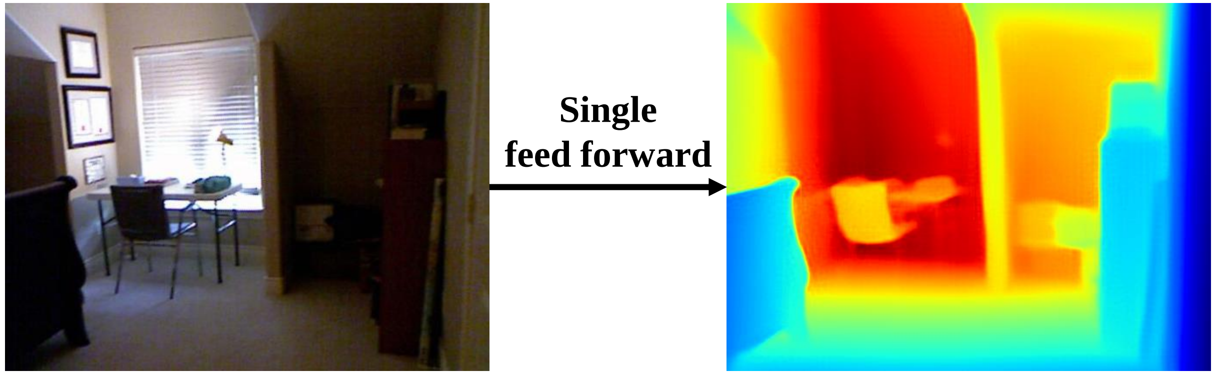
南開大學
Nankai University

Monocular Depth Estimation



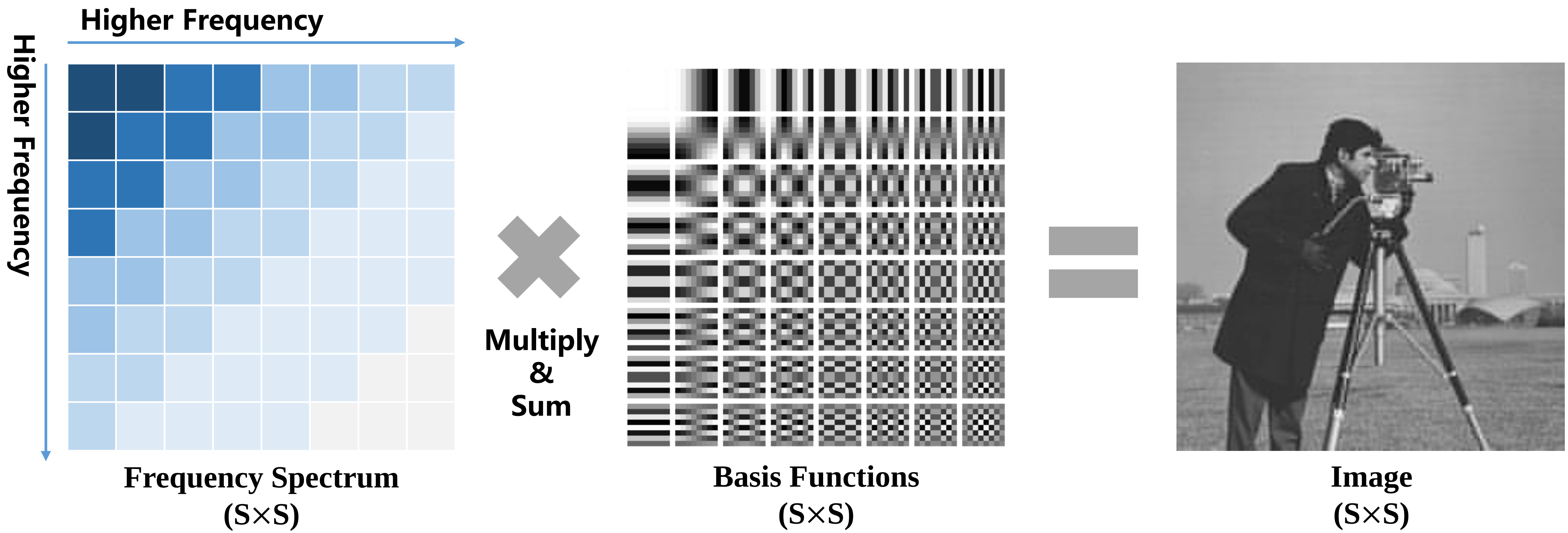
Limitations of Existing Works





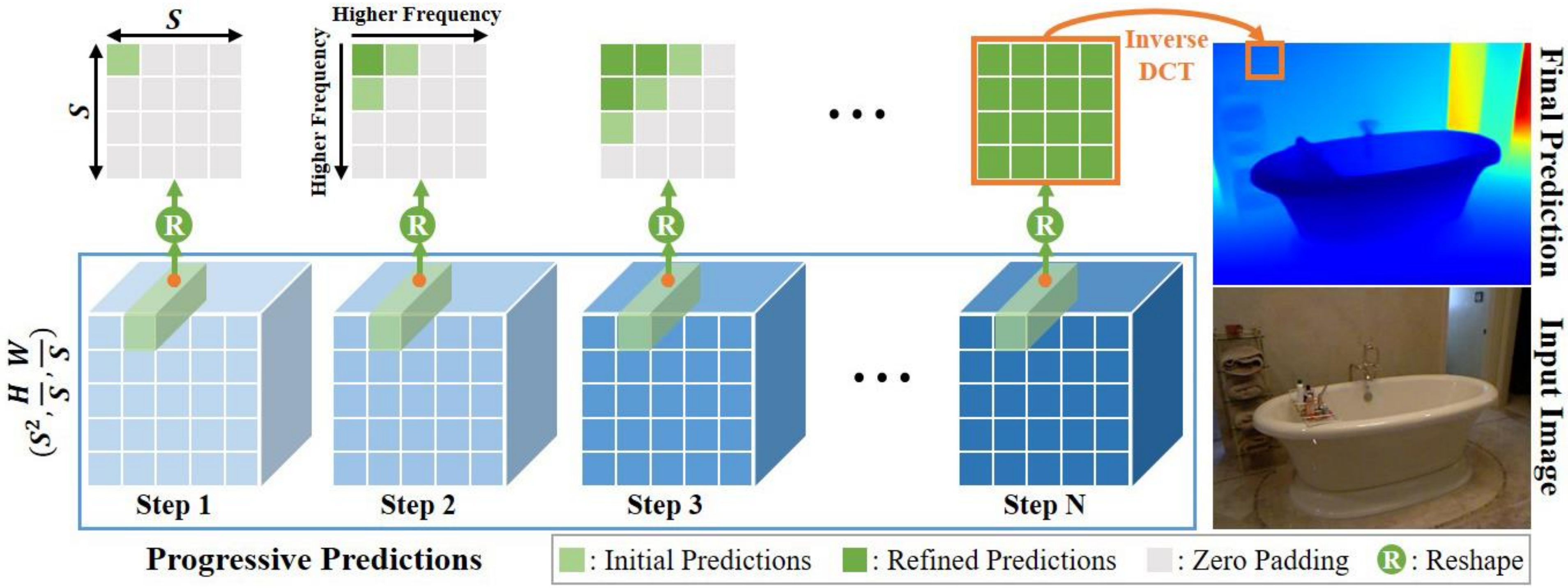
- **Unable to model the local correlations.**
- **Hard to manage complex scenes.**

Introduction to 2D Discrete Cosine Transform (DCT)

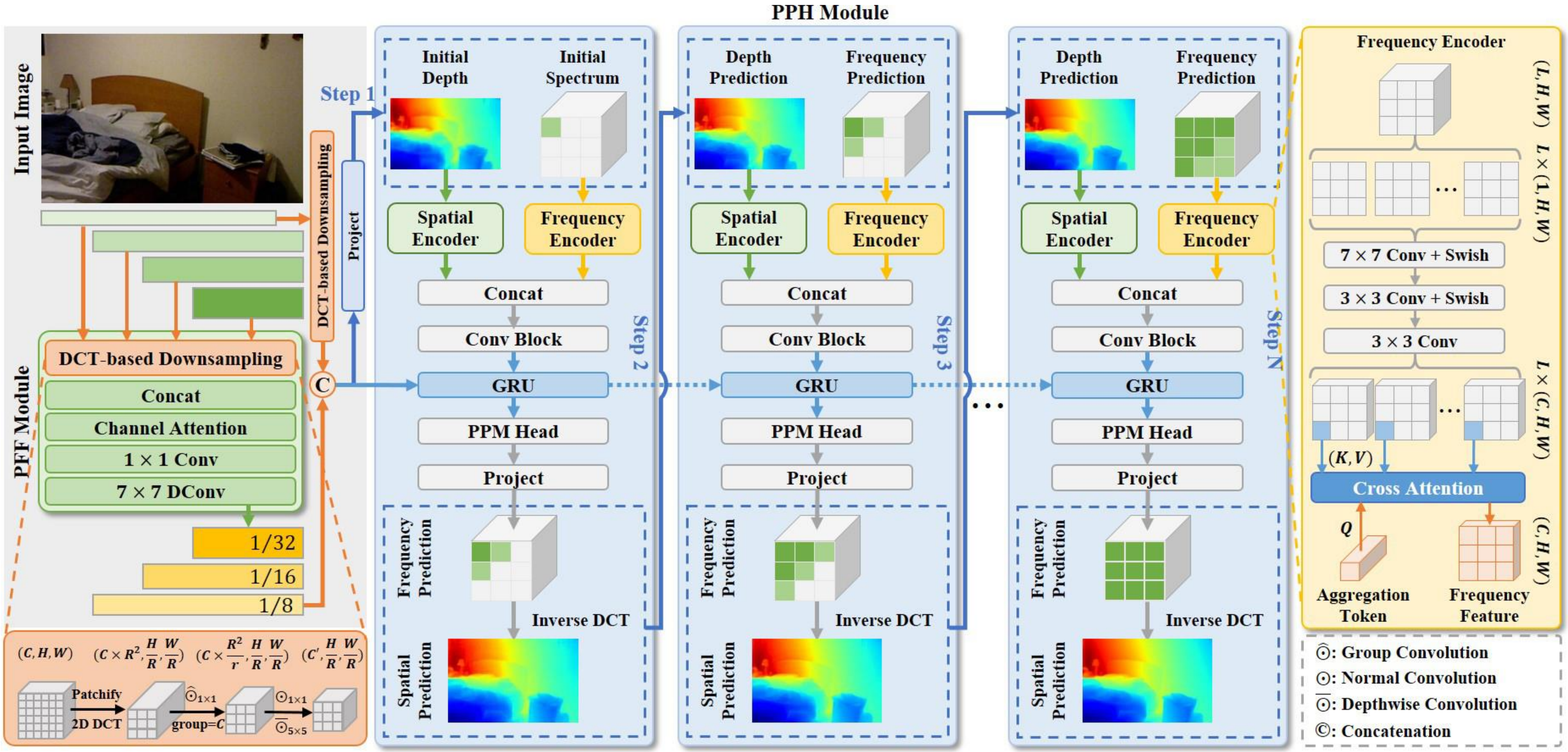


Introduction to 2D Discrete Cosine Transform (DCT)

Progressive Estimation Scheme



Network Architecture



Training Loss

- **Scale-Invariant Log Loss**

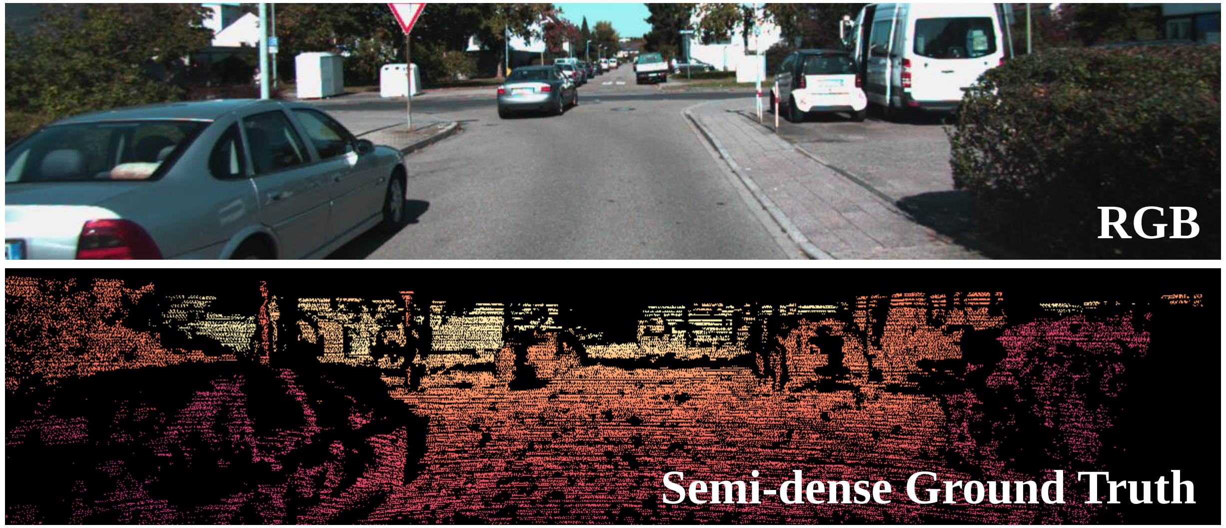
$$L_d = \alpha \cdot \sum_{i=1}^{N} \beta^{N-i} \sqrt{\frac{1}{M} \sum d_i^2 - \frac{\lambda}{M^2} (\sum d_i)^2},$$

- **Frequency Regularization Term**

$$L_f = \sum (\epsilon^{u+v} - 1) \cdot |f_{u,v}|,$$

- **Smoothness Regularization Term**

$$L_s = |\partial_x \hat{\mathcal{D}}| \cdot e^{-|\partial_x I_t|} + |\partial_y \hat{\mathcal{D}}| \cdot e^{-|\partial_y I_t|},$$



Quantitative Result

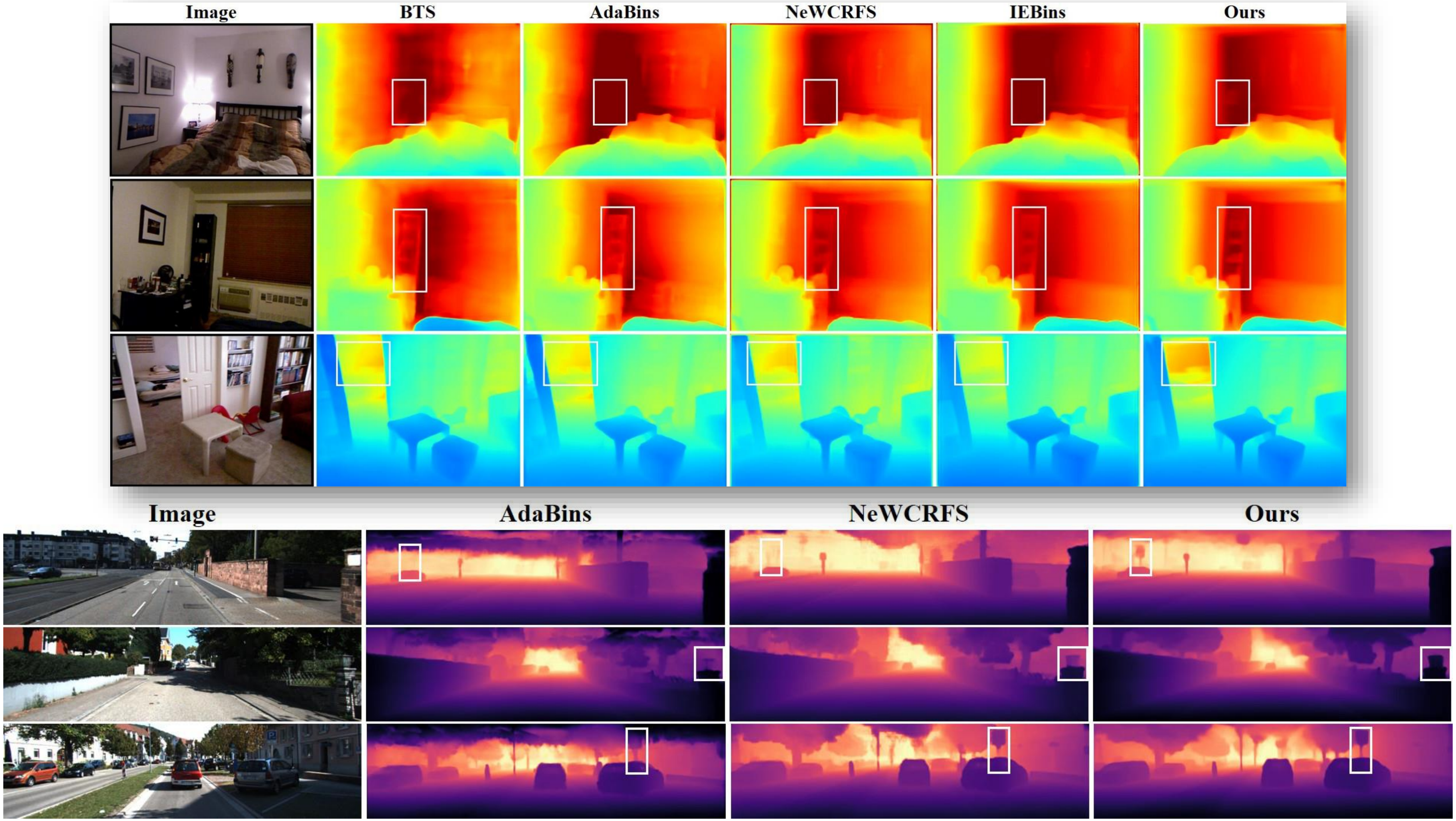
NYU-Depth-V2

Method	Backbone	Abs Rel ↓	Sq Rel ↓	RMSE ↓	$\log_{10}$ ↓	$\delta < 1.25$ ↑	$\delta < 1.25^2$ ↑	$\delta < 1.25^3$ ↑
DORN [12]	ResNet-101	0.115	–	0.509	0.051	0.828	0.965	0.992
VNL [55]	ResNet-101	0.108	–	0.416	0.048	0.875	0.976	0.994
BTS [17]	DenseNet-161	0.110	0.066	0.392	0.047	0.885	0.978	0.994
ASNDepth [24]	HRNet-48	0.101	–	0.377	0.044	0.890	0.982	0.996
TransDepth [54]	R-50+ViT-B/16	0.106	–	0.365	0.045	0.900	0.983	0.996
AdaBins [3]	E-B5+mini-ViT	0.103	–	0.364	0.044	0.903	0.984	0.997
LocalBins [4]	E-B5	0.099	–	0.357	0.042	0.907	0.987	**0.998**
NeWCRFS [58]	Swin-Large	0.095	0.045	0.334	0.041	0.922	**0.992**	**0.998**
BinsFormer [20]	Swin-Large	0.094	–	0.330	0.040	0.925	0.989	0.997
PixelFormer [1]	Swin-Large	0.090	–	0.322	0.039	0.929	0.991	**0.998**
IEBins [34]	Swin-Large	0.087	0.040	0.314	0.038	0.936	**0.992**	**0.998**
MG-Depth [21]	Swin-Large	0.087	–	0.311	–	0.933	–	–
NDDepth [33]	Swin-Large	0.087	0041	0.311	0.038	0.936	0.991	**0.998**
VA-DepthNet [22]	Swin-Large	0.086	**0.039**	**0.304**	**0.037**	0.937	**0.992**	**0.998**
Ours	Swin-Large	**0.085**	**0.039**	**0.304**	**0.037**	**0.940**	**0.992**	**0.998**

KITTI-Eigen

Method	Backbone	Abs Rel ↓	Sq Rel ↓	RMSE ↓	RMSE log ↓	$\delta < 1.25$ ↑	$\delta < 1.25^2$ ↑	$\delta < 1.25^3$ ↑
DORN [12]	ResNet-101	0.072	0.307	2.727	0.120	0.932	0.984	0.994
VNL [55]	ResNet-101	0.072	–	3.258	0.117	0.938	0.990	0.998
BTS [17]	DenseNet-161	0.060	0.249	2.798	0.096	0.955	0.993	0.998
TransDepth [54]	R-50+ViT-B/16	0.064	0.252	2.755	0.098	0.956	0.994	**0.999**
AdaBins [3]	E-B5+mini-ViT	0.058	0.190	2.360	0.088	0.964	0.995	**0.999**
P3Depth [26]	ResNet-101	0.071	0.270	2.842	0.103	0.953	0.993	0.998
NeWCRFS [58]	Swin-Large	0.052	0.155	2.129	0.079	0.974	**0.997**	**0.999**
BinsFormer [20]	Swin-Large	0.052	0.151	2.096	0.079	0.974	**0.997**	**0.999**
PixelFormer [1]	Swin-Large	0.051	0.149	2.081	0.077	0.976	**0.997**	**0.999**
VA-DepthNet [22]	Swin-Large	**0.050**	0.148	2.093	**0.076**	**0.977**	**0.997**	**0.999**
iDisc [27]	Swin-Large	**0.050**	**0.145**	2.067	0.077	**0.977**	**0.997**	**0.999**
Ours	Swin-Large	0.051	**0.145**	**2.044**	**0.076**	**0.977**	**0.997**	**0.999**

Qualitative Result



Thank You!