

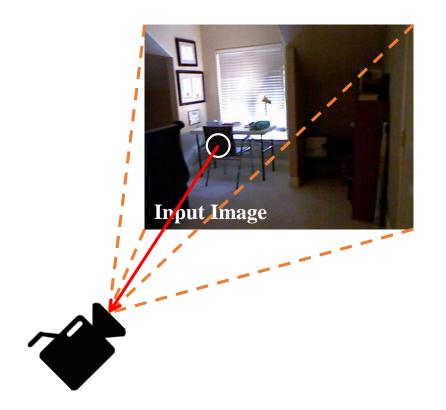
## DCDepth: Progressive Monocular Depth Estimation in Discrete Cosine Domain

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

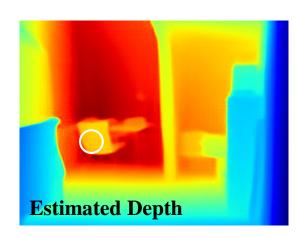




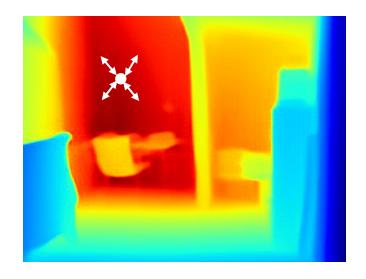
### **Monocular Depth Estimation**

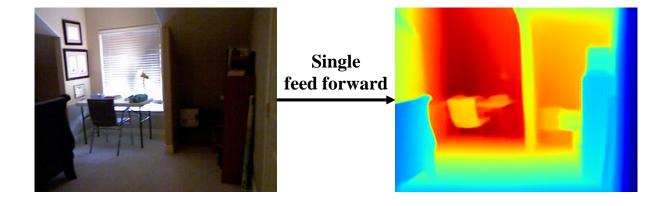






#### **Limitations of Existing Works**





• Unable to model the local correlations.

Hard to manage complex scenes.

#### **Introduction to 2D Discrete Cosine Transform (DCT)**

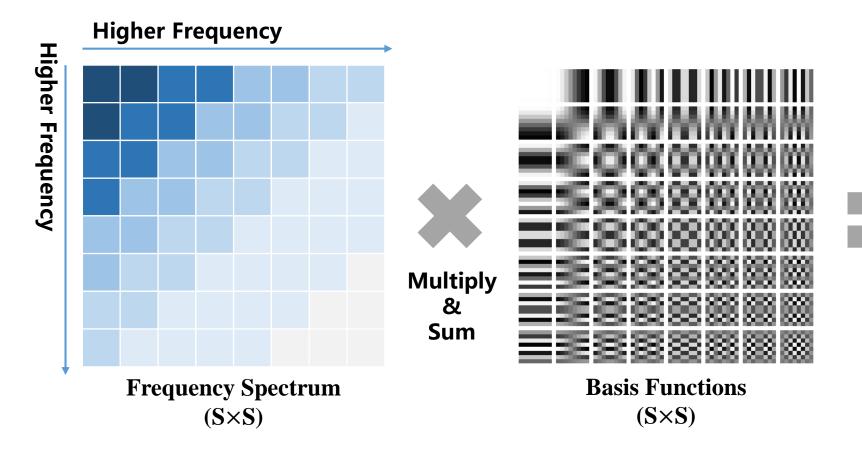
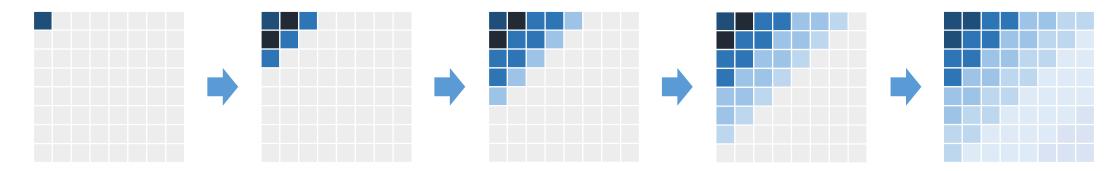




Image (S×S)

#### Introduction to 2D Discrete Cosine Transform (DCT)

#### : Zero Padding





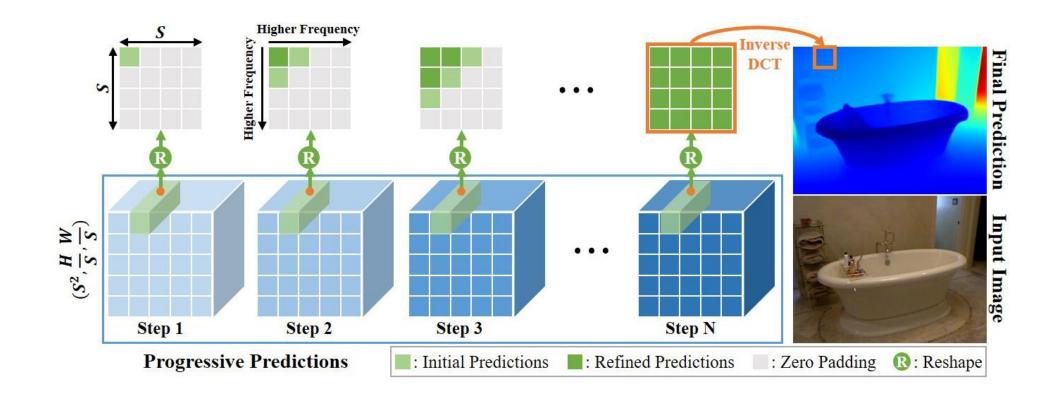




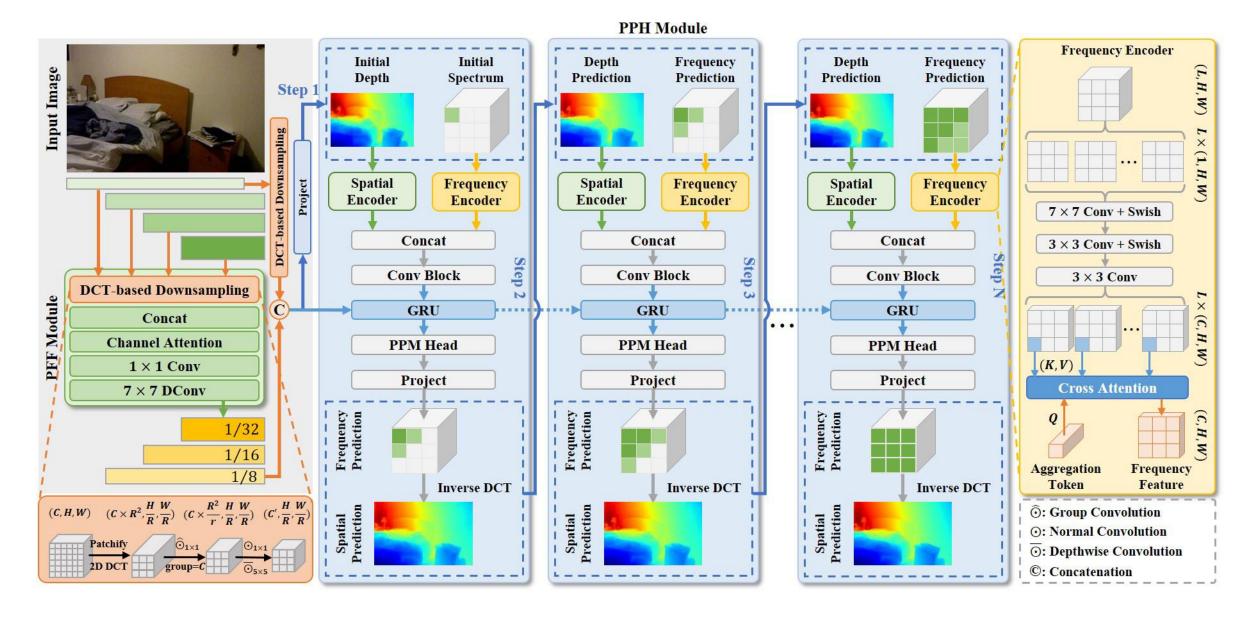




#### **Progressive Estimation Scheme**



#### **Network Architecture**



#### **Training Loss**

#### Scale-Invariant Log Loss

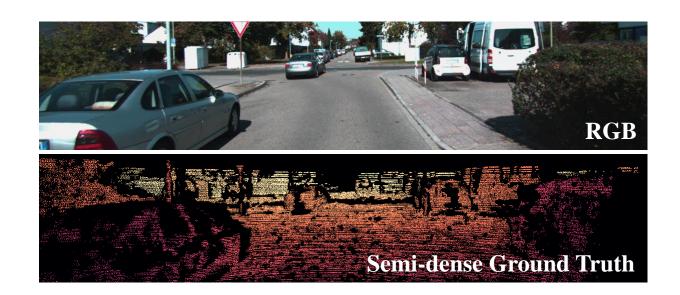
$$L_d = \alpha \cdot \sum_{i=1}^N \beta^{N-i} \sqrt{\frac{1}{M} \sum_i d_i^2 - \frac{\lambda}{M^2} (\sum_i d_i)^2},$$

#### • Frequency Regularization Term

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

#### • Smoothness Regularization Term

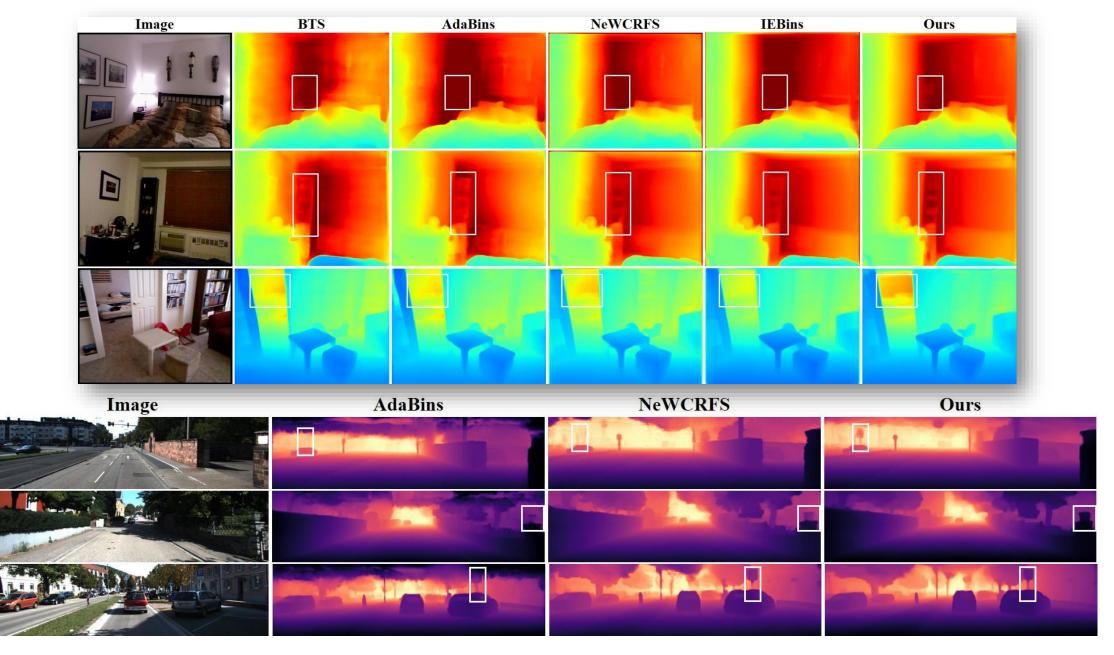
$$L_s = |\partial_x \hat{\mathcal{D}}| \cdot e^{-|\partial_x I_t|} + |\partial_y \hat{\mathcal{D}}| \cdot e^{-|\partial_y I_t|},$$



#### **Quantitative Result**

	Method	Backbone	Abs Rel↓	Sq Rel↓	RMSE ↓	$\log_{10} \downarrow$	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$
	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
7	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
NYU-Depth-V2	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
			II						
			II.			"			
	Method	Backbone	Abs Rel↓	Sq Rel↓	RMSE↓	RMSE log ↓	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$
	Method DORN [12]		Abs Rel ↓ 0.072	Sq Rel ↓ 0.307	RMSE↓ 2.727	RMSE log ↓ 0.120	$\begin{array}{ c c c c c }\hline & \delta < 1.25 \uparrow \\ \hline & 0.932 \\ \hline \end{array}$	$\delta < 1.25^2 \uparrow$ $0.984$	
		Backbone					<u> </u>		$\delta < 1.25^3 \uparrow$
en	DORN [12]	Backbone ResNet-101	0.072		2.727	0.120	0.932	0.984	$\delta < 1.25^3 \uparrow$ $0.994$
igen	DORN [12] VNL [55] BTS [17] TransDepth [54]	Backbone  ResNet-101  ResNet-101	0.072 0.072	0.307	2.727 3.258	0.120 0.117	0.932 0.938	0.984 0.990	
Eigen	DORN [12] VNL [55] BTS [17]	Backbone  ResNet-101  ResNet-101  DenseNet-161	0.072 0.072 0.060	0.307 - 0.249	2.727 3.258 2.798	0.120 0.117 0.096	0.932 0.938 0.955	0.984 0.990 0.993	$\delta < 1.25^{3} \uparrow$ $0.994$ $0.998$ $0.998$
T-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3] P3Depth [26]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16	0.072 0.072 0.060 0.064	0.307 - 0.249 0.252	2.727 3.258 2.798 2.755 2.360 2.842	0.120 0.117 0.096 0.098 0.088 0.103	0.932 0.938 0.955 0.956 0.964 0.953	0.984 0.990 0.993 0.994	$\begin{array}{c} \delta < 1.25^{3} \uparrow \\ \hline 0.994 \\ 0.998 \\ \hline 0.998 \\ \textbf{0.999} \\ \textbf{0.999} \\ \textbf{0.999} \\ \underline{0.998} \\ \end{array}$
lTI-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16 E-B5+mini-ViT	0.072 0.072 0.060 0.064 0.058	0.307 - 0.249 0.252 0.190	2.727 3.258 2.798 2.755 2.360	0.120 0.117 0.096 0.098 0.088	0.932 0.938 0.955 0.956 0.964	0.984 0.990 0.993 0.994 0.995	$\begin{array}{c} \delta < 1.25^3 \uparrow \\ \hline 0.994 \\ \underline{0.998} \\ \underline{0.998} \\ \textbf{0.999} \\ \textbf{0.999} \\ \textbf{0.999} \end{array}$
ITTI-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3] P3Depth [26] NeWCRFS [58] BinsFormer [20]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16 E-B5+mini-ViT ResNet-101	0.072 0.072 0.060 0.064 0.058 0.071	0.307 - 0.249 0.252 0.190 0.270	2.727 3.258 2.798 2.755 2.360 2.842	0.120 0.117 0.096 0.098 0.088 0.103	0.932 0.938 0.955 0.956 0.964 0.953	0.984 0.990 0.993 0.994 <u>0.995</u> 0.993	$\begin{array}{c} \delta < 1.25^3 \uparrow \\ \hline 0.994 \\ \underline{0.998} \\ 0.998 \\ \hline 0.999 \\ 0.999 \\ \hline 0.998 \\ \hline 0.999 \\ 0.999 \\ \end{array}$
KITTI-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3] P3Depth [26] NeWCRFS [58] BinsFormer [20] PixelFormer [1]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16 E-B5+mini-ViT ResNet-101 Swin-Large	0.072 0.072 0.060 0.064 0.058 0.071 0.052 0.052 0.051	0.307 - 0.249 0.252 0.190 0.270 0.155	2.727 3.258 2.798 2.755 2.360 2.842 2.129	0.120 0.117 0.096 0.098 0.088 0.103 0.079 0.079	0.932 0.938 0.955 0.956 0.964 0.953 0.974 0.974	0.984 0.990 0.993 0.994 0.995 0.993	$\begin{array}{c} \delta < 1.25^{3} \uparrow \\ \hline 0.994 \\ 0.998 \\ 0.998 \\ \hline 0.999 \\ 0.999 \\ \hline 0.998 \\ \hline 0.999 \\ 0.999 \\ 0.999 \\ \end{array}$
KITTI-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3] P3Depth [26] NeWCRFS [58] BinsFormer [20] PixelFormer [1] VA-DepthNet [22]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16 E-B5+mini-ViT ResNet-101 Swin-Large Swin-Large Swin-Large Swin-Large	0.072 0.072 0.060 0.064 0.058 0.071 0.052 0.052 0.051 <b>0.050</b>	0.307 - 0.249 0.252 0.190 0.270 0.155 0.151 0.149 0.148	2.727 3.258 2.798 2.755 2.360 2.842 2.129 2.096 2.081 2.093	0.120 0.117 0.096 0.098 0.088 0.103 0.079 0.079 0.077 <b>0.076</b>	0.932 0.938 0.955 0.956 0.964 0.953 0.974 0.974 0.976 <b>0.977</b>	0.984 0.990 0.993 0.994 <u>0.995</u> 0.993 <b>0.997</b> <b>0.997</b> <b>0.997</b>	$\begin{array}{c} \delta < 1.25^3 \uparrow \\ \hline 0.994 \\ \underline{0.998} \\ 0.998 \\ \hline 0.999 \\ 0.999 \\ \underline{0.998} \\ 0.999 \\ 0.999 \\ 0.999 \\ 0.999 \\ 0.999 \\ 0.999 \\ \end{array}$
KITTI-Eigen	DORN [12] VNL [55] BTS [17] TransDepth [54] AdaBins [3] P3Depth [26] NeWCRFS [58] BinsFormer [20] PixelFormer [1]	Backbone  ResNet-101 ResNet-101 DenseNet-161 R-50+ViT-B/16 E-B5+mini-ViT ResNet-101 Swin-Large Swin-Large Swin-Large	0.072 0.072 0.060 0.064 0.058 0.071 0.052 0.052 0.051	0.307 - 0.249 0.252 0.190 0.270 0.155 0.151 0.149	2.727 3.258 2.798 2.755 2.360 2.842 2.129 2.096 2.081	0.120 0.117 0.096 0.098 0.088 0.103 0.079 0.079	0.932 0.938 0.955 0.956 0.964 0.953 0.974 0.974	0.984 0.990 0.993 0.994 0.995 0.993 0.997 0.997	$\begin{array}{c} \delta < 1.25^{3} \uparrow \\ \hline 0.994 \\ 0.998 \\ 0.998 \\ \hline 0.999 \\ 0.999 \\ \hline 0.998 \\ \hline 0.999 \\ 0.999 \\ 0.999 \\ \end{array}$

#### **Qualitative Result**





# Thank You!



