Distortion and Uncertainty Aware Loss for Panoramic Depth Completion

Zhiqiang Yan [1] **Xiang Li** [1] **Kun Wang** [1] **Shuo Chen** [2] **Jun Li** [1] **Jian Yang** [1]
{yanzq,xiang.li.implus,kunwang,junli,csjyang}@njust.edu.cn, shuo.chen.ya@riken.jp

Abstract

Standard MSE or MAE loss function is commonly used in limited field-of-vision depth completion, treating each pixel equally under a basic assumption that all pixels have same contribution during optimization. Recently, with the rapid rise of panoramic photography, panoramic depth completion (PDC) has raised increasing attention in 3D computer vision. However, the assumption is inapplicable to panoramic data due to its latitude-wise **distortion** and high **uncertainty** nearby textures and edges. To handle these challenges, we propose distortion and uncertainty aware loss (DUL) that consists of a distortion-aware loss and an uncertainty-aware loss. The distortion-aware loss is designed to tackle the panoramic distortion caused by equirectangular projection, whose coordinate transformation relation is used to adaptively calculate the weight of the latitude-wise distortion, distributing uneven importance instead of the equal treatment for each pixel. The uncertainty-aware loss is presented to handle the inaccuracy in non-smooth regions. Specifically, we characterize uncertainty into PDC solutions under Bayesian deep learning framework, where a novel consistent uncertainty estimation constraint is designed to learn the consistency between multiple uncertainty maps of a single panorama. This consistency constraint allows model to produce more precise uncertainty estimation that is robust to feature deformation. Extensive experiments show the superiority of our method over standard loss functions, reaching the state of the art.

[1]PCALab, School of Computer Science and Engineering, Nanjing University of Science and Technology, China [2]RIKEN Center for Advanced Intelligence Project, Japan. Correspondence to: Shuo Chen <shuo.chen.ya@riken.jp>, Jun Li <junli@njust.edu.cn>.

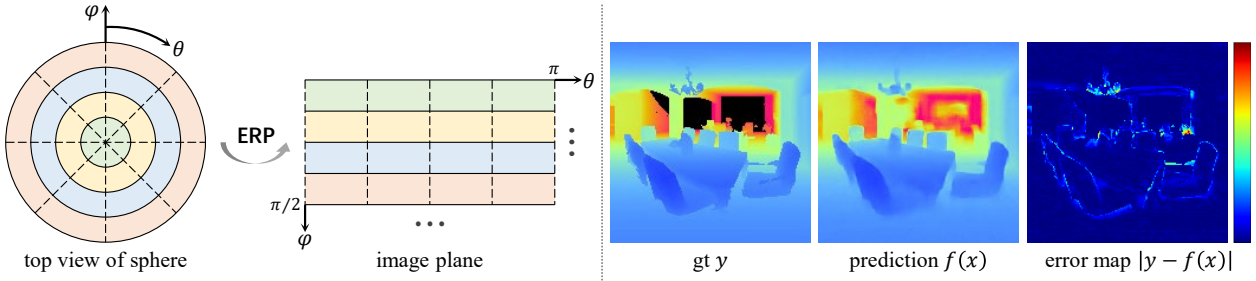
Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

1. Introduction

Depth completion aims to recover dense depth from the sparse one and its corresponding perspective color image with narrow field of vision (FoV). Plenty of works (Cheng et al., 2020; Tang et al., 2020; Lin et al., 2022) have been proposed to tackle this task. With the advent of panoramic cameras, predicting depth from the 360° full-FoV color image becomes a fashionable trend (Wang et al., 2020; Sun et al., 2021).Recently, instead of directly estimating depth from pure 360° full-FoV image, panoramic depth completion (PDC) task has been raised in M^3PT (Yan et al., 2022) where the sparse depth can be facilitated to generate much more accurate prediction. Generally, to produce precise depth estimations, most of these works employ *MSE* loss to optimize their networks. However, it is well-known that such loss distributes same weight to each pixel regardless of their uneven importance in an image, which is especially not appropriate for panoramic data in two aspects:

Latitude-wise distortion in panoramas. For panoramic depth perception (Zioulis et al., 2018; Yan et al., 2022), the most commonly used data format is the equirectangular projection (ERP) image (Pintore et al., 2021; Zhuang et al., 2022) from spherical coordinate to plane coordinate, which is shown in Fig. 1(a). We observe that the area of grid is gradually increasing from the poles to the equator. However, ERP maps each grid to image plane with an equivalent area, indicating that the closer the grid is to poles, the more pixels are interpolated during the projection process. Hence, the distortion of the ERP panoramic image gets severe step by step from the equator to poles. Besides, most objects are located mainly near the equator, *i.e.*, the middle region of the image, while the two poles primarily contain the ceiling and floor. Therefore, traditional loss distributing same weight to each pixel is no longer suitable for PDC.

High uncertainty nearby non-smooth regions. It is usually difficult for networks to predict accurate depth with high uncertainty near edge, occlusion, blur, outlier, etc (Eldesokey et al., 2020; Park et al., 2020; Zhu et al., 2022; Jin et al., 2020). Moreover, the panoramic data introduces latitude-wise distortion which makes those regions more irregular, negatively leading to even worse depth prediction. Fig. 1(b) demonstrates that non-smooth regions (*e.g.*, edges



(a) Equirectangular projection (ERP)

(b) Visualizations of depth and error

Figure 1. Illustrations of ERP and depth error. (a): $\varphi \in [0, \pi]$ is the polar angle (latitude) and $\theta \in [0, 2\pi]$ is the azimuth angle (longitude). For convenience, in the image plane we show only half of the whole. (b): We visualize the error map between ground-truth depth and predicted depth reconstructed by M^{3}PT (Yan et al., 2022) on Matterport3D (Albanis et al., 2021), all of which are cropped for visibility.

of wall and chair) are not estimated as well as smooth regions (*e.g.*, surfaces of floor, wall, and ceiling). The error map in Fig. 1(b) further shows that depth in non-smooth regions is harder to recover than smooth regions, *i.e.*, the importance distribution of different pixels is uneven. Consequently, it is suggested to focus on learning those hard pixels with high uncertainty adaptively.

In this paper, we design new loss functions to address the above issues. **For latitude-wise distortion**, we propose distortion-aware loss to distribute different pixels with uneven weights. Specifically, we calculate the weight coefficient for each pixel based on the fundamental of ERP, *i.e.*, the mapping relationship from spherical coordinate to plane coordinate, where the latitude plays a key role. In this way, a latitude-wise weighted matrix is produced to redistribute the importance of each pixel. As a result, the proposed distortion-aware loss function mitigates the negative effects of the latitude-wise distortion and the unbalanced distribution of objects in panoramas. **For high uncertainty**, we present uncertainty-aware loss to facilitate depth recovery near non-smooth regions. Concretely, we first introduce the mean and variance Bayesian framework (Kendall & Gal, 2017; Choi et al., 2019; Eldesokey et al., 2020) into PDC, which simultaneously predicts dense depth (mean) and uncertainty (variance). Importantly and differently, we further propose to learn the *similarity* of multiple uncertainty maps of a single panorama, *i.e.*, the consistency across different uncertainty estimations from the same panorama. This consistency constraint brings more precise uncertainty estimation that is robust to feature deformation, while alleviating its learning difficulty. Finally, we organically merge the distortion-aware and uncertainty-aware loss functions to boost the panoramic depth recovery and 3D reconstruction. In summary, our main contributions are:

- Distortion-aware loss (DAL). We propose to learn the weighted loss adaptively based on ERP principle, which effectively mitigates the negative impacts of the inherent distortion in panoramic data.
- Uncertainty-aware loss (UAL) with consistent uncertainty estimation (CUE) constraint. Under Bayesian deep learning framework which characterizes uncertainty into PDC solutions, we present to estimate consistent uncertainty that markedly boosts the model.
- Universality of DUL. The DAL and UAL, together termed DUL, can be easily deployed in existing PDC networks for improvement and hardly increase additional overhead. Experimental results show the superiority of DUL than standard *MSE* or *MAE* loss function.

2. Related Work

2.1. Panoramic Depth Perception with Distortion

Panoramic depth perception mainly consists of panoramic depth estimation (Zioulis et al., 2018) from RGB and panoramic depth completion (Yan et al., 2022) from RGB-D, in which ERP data is commonly used. Till now, there have been many works (Tateno et al., 2018; Lee et al., 2020; Pintore et al., 2021; Zhuang et al., 2022) focusing on the inherent distortion in panoramic data caused by ERP. For example, (Lee et al., 2019) and (Lee et al., 2020) develop a new panoramic data format using icosahedral spherical polyhedron representation, which alleviates the distortion from source data. In networks, (Pintore et al., 2021) and (Sun et al., 2021) propose a pre-processing head and a post-processing tail to reduce the negative effects of distortion, respectively. (Tateno et al., 2018) and (Zhuang et al., 2022) employ distortion-aware convolutions to adaptively encode distortion. Since cubemap image dose not introduce distortion (but suffers from blurry edge), (Wang et al., 2020) and (Jiang et al., 2021) present to predict cubic depth and then project it into ERP depth for complementation. Besides, (Eder et al., 2019) and (Jin et al., 2020) utilize geometric structure to regularize distortion. Different from previous methods, based on the principle of ERP we directly adjust the loss weight of each pixel to balance distortion without changing data format or adding extra complex modules.

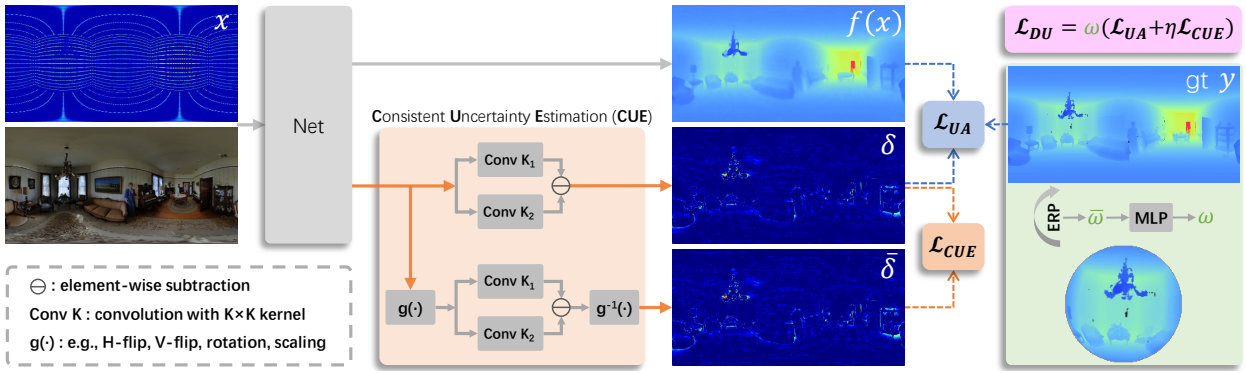


Figure 2. Overview of training PDC networks with $\mathcal{L}_{DU}$. We introduce mean and variance estimation under Bayesian framework to formulate $\mathcal{L}_{UA}$, where $\mathcal{L}_{CUE}$ is designed for more precise uncertainty. Then the distortion weight ω is based on the fundamental of ERP.

2.2. Perspective Depth Prediction with Uncertainty

Uncertainty estimation can trace back to (MacKay, 1992; Welling & Teh, 2011) with Bayesian neural networks. After that, it is very popular for deep learning networks to predict uncertainty in many computer vision tasks, such as object detection (He et al., 2019), image super-resolution (Ning et al., 2021), optical flow estimation (Ilg et al., 2018; Yang & Ramanan, 2019), and depth perception (Poggi et al., 2020; Eldesokey et al., 2020; Zhu et al., 2022), which is the most relevant to our research. Not coincidentally, they introduce mean (depth) and variance (uncertainty) estimation for robustness. Moreover, (Van Gansbeke et al., 2019; Park et al., 2020; Xu et al., 2019) present to estimate confidence which is equivalent to uncertainty for perspective depth completion. Most recently, M^{3}PT (Yan et al., 2022) proposes the new panoramic depth completion task, based on which the mean and variance system is applied in this paper. But differently, we aim to learn precise and reasonable variance from *multiple* uncertainty maps instead of *single*, based on the fundamental that the uncertainty of one scene is definite once its panoramic image is given.

3. Method

Our pipeline is shown in Fig. 2. We use x and y to denote the panoramic sparse depth and ground-truth depth, δ and $\bar{\delta}$ to represent two uncertainty maps that correspond to a same scenario. $f(\cdot)$ refers to an arbitrary PDC network.

3.1. Distortion-Aware Loss (DAL)

ERP for Panoramic Data. ERP is a special case of cylindrical equidistant projection, from the surface of a sphere (φ-θ) to a flat image (m-n), which is defined as

$$m = r\left(\theta - \theta_0\right)\cos\varphi_1, \quad n = r\left(\varphi - \varphi_0\right), \tag{1}$$

where r is the radius of sphere, θ and θ_0 denote longitude and central meridian severally. φ, φ_0, and φ_1 refer to latitude, central parallel, and standard parallel, respectively. For panoramic data, $r = 1$, $\theta_0 = 0$, and $\varphi_0 = 0$. In particular, when the standard parallel coincides with the equator, taken as $\cos\varphi_1 = 1$, the cylindrical equidistant projection degenerates into ERP, which is expressed as

$$m = \theta, \quad n = \varphi. \tag{2}$$

We can find the mapping relationship of ERP is very simple, *i.e.*, the horizontal coordinate is simply longitude with 360° FoV, and the vertical coordinate is simply latitude with 180° FoV, with no transformation or scaling applied. It is the main reason for existing panoramic photography to choose ERP as the common storage format. However, ERP is neither equivalent nor conformal, and introduces considerable distortion. Next, we focus on adjusting the loss weight to alleviate the negative impact of the distortion in networks.

Perceiving Distortion in ERP. Proceeding from reality, there are two factors that prevent standard loss functions from better performance. (1) The area of spherical grid that is composed of interlaced longitude and latitude progressively increases from the poles to the equator, but ERP maps them to image plane with same area. Consequently, more inaccurate pixels are interpolated in a grid as it is closer to poles. (2) Most objects are mainly distributed near middle regions of panoramas instead of ceiling and floor. Hence, *we propose to distribute uneven importance instead of the equal treatment to every pixel by calculating the weight of latitude-wise distortion based on spherical surface integral.*

For simplicity, as shown in Fig. 1(a), we move the plane coordinate center to the top-left corner of image, and modify the intervals of φ and θ from $[-\frac{\pi}{2}, \frac{\pi}{2}]$, $[-\pi, \pi]$ to $[0, \pi]$, $[0, 2\pi]$, respectively. The weight of one pixel surface integral to the total area is formulated as

$$\begin{aligned}\bar{\omega}\left(\theta, \varphi\right) &= \frac{k}{4\pi r^2}\int_{\theta}^{\theta+\Delta\theta} r d\theta \int_{\varphi}^{\varphi+\Delta\varphi} r\sin\varphi d\varphi \\ &= \mu\left[\cos\varphi - \cos\left(\varphi + \Delta\varphi\right)\right],\end{aligned} \tag{3}$$

where $\mu = \frac{k\Delta\theta}{4\pi}$, k is a hyper parameter, and $\Delta\varphi$ can be 1° in practice. It is observed that the weight is not relevant to

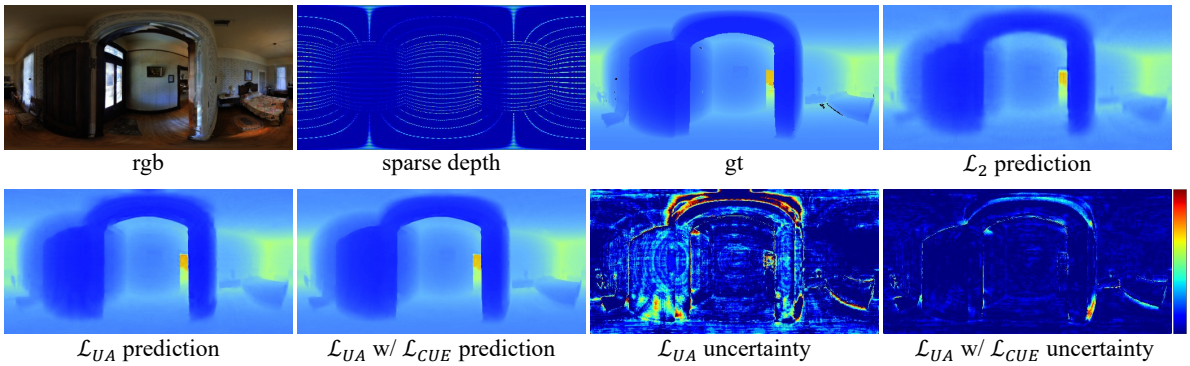


Figure 3. Visual comparisons of panoramic depth completion results based on M^{3}PT (Yan et al., 2022) using different loss functions.

θ, indicating that the distortion occurs only along latitude direction, which is the specific characteristic of ERP. As illustrated in Fig. 1, assume i is one pixel of the image and (a, b) is its coordinate, I_w and I_h are the image width and height respectively. Then the transformation relationship between spherical coordinate and plane coordinate is

$$\theta = \frac{a}{I_w} 2\pi, \quad \varphi = \frac{b}{I_h} \pi. \tag{4}$$

Substituting Eq. (4) into Eq. (3), we yield

$$\bar{\omega}_i = \bar{\omega}\left(a, b\right) = \mu \left[\cos\left(\frac{b\pi}{I_h}\right) - \cos\left(\frac{b\pi}{I_h} + \Delta\varphi\right) \right]. \tag{5}$$

Based on $\bar{\omega}$, we employ two Multi-Layer Perceptrons (MLP, denoted as $m(\cdot)$) with residual connection to generate the adaptive distortion-aware weight ω, obtaining

$$\omega_i = \bar{\omega}_i + m\left(\bar{\omega}_i\right). \tag{6}$$

After yielding ω_i, we then multiply it by standard loss functions. Taking *MSE* as an example, the DAL is defined as

$$\mathcal{L}_{DA} = \omega \mathcal{L}_2 = \frac{1}{N} \sum_{i=1}^{N} \omega_i \left(y_i - f\left(x_i\right)\right)^2, \tag{7}$$

where N is the valid pixel set of the ground-truth depth y.

3.2. Uncertainty-Aware Loss (UAL)

Optimization with Mean and Variance. For PDC, existing deep learning networks (Yan et al., 2022) mainly devote into predicting dense depth (mean) only, leading to not very accurate results near non-smooth regions that are illustrated in Fig. 1(b). To tackle this issue, we introduce uncertainty (variance) estimation framework (Kendall & Gal, 2017) into 360° panoramas. As explored in the framework, Bayesian deep learning tools make it possible that modeling uncertainty in computer vision. In general, there are two major kinds of uncertainty, namely aleatoric and epistemic. The aleatoric uncertainty captures noise distributed in the observation data while the epistemic uncertainty depicts the uncertainty in the model. As shown in the left of Fig. 2, since the sparse depth x is generated by sensor scanning with equal angles (Yan et al., 2022; Uhrig et al., 2017), resulting in uneven depth distribution, we opt the aleatoric uncertainty δ and apply it into a PDC network $f(\cdot)$. Then we can formulate the observation model as

$$y_i = f\left(x_i\right) + \tau \delta_i, \tag{8}$$

where τ represents the Gaussian prior distribution with zero mean and unit variance.

Without the uncertainty δ, we hope the dense depth prediction $f(x)$ can infinitely approximate y in PDC networks, whose procession can be described by maximizing the posterior probability $p(y_i|x_i)$. With the uncertainty δ, the joint posterior probability is denoted as $p(y_i, \delta_i|x_i)$ [1].

$$L_{UA} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left(y_i - f\left(x_i\right)\right)^2}{{\delta_i}^2} + 2 \log {\delta_i}^2. \tag{9}$$

This loss is similar with the aleatoric uncertainty loss presented in (Kendall & Gal, 2017). But differently, δ_i in our model describes the uncertainty measure that encodes the observed noise variance and the consistent uncertainty estimation, while in (Kendall & Gal, 2017), it is the variance of the noise. Fig. 3 shows that the $\mathcal{L}_{UA}$ depth prediction possesses more reasonable visual effects than $\mathcal{L}_2$ depth result, *e.g.*, edges of bed and chair.

Consistent Uncertainty Estimation (CUE). Uncertainty has been studied in many depth related works (Van Gansbeke et al., 2019; Zhu et al., 2022), most of which employ convolutions to predict a single uncertainty map. However, as we know that the uncertainty map itself is hard to model since it usually lies near edge, outlier, etc (Eldesokey et al.,

[1] $p(y_i, \delta_i|x_i) = p(\delta_i|x_i)p(y_i|\delta_i, x_i)$, where we use Jeffrey's prior $p(\delta_i|x_i) \approx 1/\delta_i$ (Figueiredo, 2001) to model the likelihood of $p(\delta_i|x_i)$. $p(y_i|\delta_i, x_i)$ belongs to Gaussian distribution that can be formulated as a maximum likelihood problem $p(y_i|\delta_i, x_i) = 1/\sqrt{2\pi}\delta_i \exp(-(y_i - f(x_i))^2/2{\delta_i}^2)$. By combining the $p(\delta_i|x_i)$ prior and the $p(y_i|\delta_i, x_i)$ likelihood, then taking negative log operation, we obtain UAL in Eq. (9).

2020; Jin et al., 2020). *Considering a basic fact that the uncertainty is relatively definite once a panorama is given, i.e., edge and outlier in this known panoramic picture are fixed. The uncertainty maps estimated by the same network should be similar. As a result, different from previous methods, we present to learn the consistency between multiple uncertainty maps of a single panorama.*

Overall, as shown in Fig. 2, given an arbitrary PDC network, we embed our lightweight CUE at its end as an overhead. CUE outputs the uncertainty map δ which is combined with the dense depth $f(x)$ for joint optimization in Eq. (9). Specifically, CUE first takes as input the output feature base of the PDC network. Next, the base is fed into two branches, both of which conduct subtraction between two convolutions (Xu et al., 2020) followed by a ReLU activation function. In one branch, taking *two uncertainty maps* as an example, the base is transformed by $g(\cdot)$ and the inverse $g^{-1}(\cdot)$, *e.g.*, flipped horizontally twice to produce another uncertainty map $\bar{\delta}$. During this process, we utilize subtraction of two convolutions with different receptive fields (*e.g.*, $K_1 = 1, K_2 = 3$) to extract the high-frequency part, which is proved effective (Xu et al., 2020; Yu et al., 2020). On the other hand, it is widely known that uncertainty is difficult to predict as it usually appears near edges. Therefore, besides the hard constraint of Eq. 9, we also employ the subtraction to highlight the uncertainty. Finally, we perform $\mathcal{L}_{CUE}$ to learn the consistency between δ and $\bar{\delta}$, urging the model to acquire more precise uncertainty that is robust to feature deformation, and simultaneously alleviating its learning difficulty. The $\mathcal{L}_{CUE}$ [2] is formulated as

$$\mathcal{L}_{CUE} = \frac{1}{N}\sum_{i=1}^{N}\left(\delta_i - \bar{\delta}_i\right)^2. \tag{10}$$

As evidenced in Fig. 3, we discover that $\mathcal{L}_{CUE}$ produces more transparent uncertainty estimation. Obviously, there are fewer uncertain regions in $\mathcal{L}_{CUE}$ than $\mathcal{L}_{UA}$ uncertainty, contributing to more accurate depth recovery.

3.3. Distortion and Uncertainty Aware Loss (DUL)

We observe that $\mathcal{L}_{DA}$ and $\mathcal{L}_{UA}$ can be unified into a general form, termed joint distortion and uncertainty aware loss, which can be defined as

$$\begin{aligned}\mathcal{L}_{DU} &= \omega\left(\mathcal{L}_{UA} + \eta\mathcal{L}_{CUE}\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{\omega_i\left(y_i - f\left(x_i\right)\right)^2}{{\delta_i}^2} + 2\omega_i\log{\delta_i}^2\\ &+ \frac{1}{N}\sum_{i=1}^{N}\eta\omega_i\left(\delta_i - \bar{\delta}_i\right)^2,\end{aligned} \tag{11}$$

where η is a hyper parameter. The first term is data term and the second term is its inherent regularizer, both of which are inferred by the mean and variance estimation under Bayesian framework. The third term is also a regularizer.

Properties of DUL. (1) Degeneration. **(i)** Without distortion, such that the distortion-aware weight meets $\omega_i = 1$, indicating that the network treats every pixel equally. Consequently, the joint loss function $\mathcal{L}_{DU}$ will degenerate into the uncertainty-aware loss $\mathcal{L}_{UA}$. **(ii)** Without uncertainty, such that the two uncertainty representations satisfy $\delta_i = \bar{\delta}_i = 1$. In this case, the model pays no additional attention to the optimization near non-smooth regions. As a result, $\mathcal{L}_{DU}$ will degenerate into the distortion-aware loss $\mathcal{L}_{DA}$. **(iii)** Without distortion and uncertainty, we yield $\omega_i = 1$ and $\delta_i = \bar{\delta}_i = 1$, then $\mathcal{L}_{DU}$ turns into the standard *MSE* loss function. **(2) Variation tendency.** As discussed in Sec. 3.2, panoramic data carries inherent noise caused by sensors. Assume that the model has been well studied the real data distribution ($\left(y_i - f(x_i)\right)^2 \to 0$) but always with a slight error that can be treated as the noise. We try to counteract the loss caused by this noise via $\left(y_i - f(x_i)\right)^2/{\delta_i}^2$. However, for one thing, the model tends to predict an infinite ${\delta_i}^2$ to minimize the loss function. In consequence, the second term is involved for balance. For another thing, the ${\delta_i}^2$ itself is hard to estimation but the better performance of the model heavily depends on more accurate ${\delta_i}^2$. Thus, we introduce the third term to give ${\delta_i}^2$ strong and meticulous constraint. **(3) Easy training.** Upon baselines, only two steps are needed for training. **(i)** Replacing the raw loss with $\mathcal{L}_{DU}$. **(ii)** Adding the very lightweight CUE and MLP to estimate uncertainty and distortion weight. Consequently, our training and testing times are almost the same with the baselines. In addition, the time-consuming variance estimation in Bayesian learning for computer vision tasks is mainly caused by the *two-stage* training strategy (Ning et al., 2021; Zhu et al., 2022), *i.e.*, estimating uncertainty first and then predicting target. Differently, our method is trained end-to-end in *one-stage* manner and thus it is time-saving.

4. Experiment

4.1. Experimental Settings

Datasets and Metrics. Following M^{3}PT (Yan et al., 2022), we train the model on Matterport3D (Albanis et al., 2021) and 3D60 (Zioulis et al., 2019) datasets with 512×256 resolution. Matterport3D is composed of 7,907 RGB-D panoramas, 5,636 for training and 1,527 for testing. For 3D60, there are 6,669 RGB-D pairs for training and 1,831 for testing. We use seven standard metrics for evaluation, *i.e.*, mean absolute error (MAE(mm)), root mean square error of linear measures (RMSE (mm)), mean relative error (MRE), root mean square error of log measures (Log), and σ_t which denotes the percentage of predicted pixels whose

[2] Assume that there are M uncertainty maps, j denotes one of them. Then we yield $\mathcal{L}_{CUE} = \frac{1}{N(M-1)}\sum_{j=1}^{M-1}\sum_{i=1}^{N}\left(\delta_i^j - \delta_i^{j+1}\right)^2$.

Data	Method	Error Metric ↓				Accuracy Metric ↑			Reference
		RMSE	MAE	MRE	Log	σ_1	σ_2	σ_3	
Matterport3D	UniFuse (Jiang et al., 2021)	229.1	95.2	0.0475	0.0381	0.9710	0.9924	0.9970	ICRA 2021
	HoHo-R (Sun et al., 2021)	199.2	75.0	0.0355	0.0311	0.9806	0.9945	0.9977	CVPR 2021
	HoHo-H (Sun et al., 2021)	215.5	85.7	0.0406	0.0337	0.9772	0.9938	0.9975	CVPR 2021
	PENet (Hu et al., 2021)	248.0	91.5	0.0493	0.0350	0.9728	0.9935	0.9970	ICRA 2021
	GuideNet (Tang et al., 2020)	192.9	87.2	0.0438	0.0327	0.9806	0.9948	0.9981	TIP 2021
	360Depth (Rey-Area et al., 2021)	185.3	68.8	0.0302	0.0285	0.9833	0.9942	0.9980	CVPR 2022
	M^{3}PT (Yan et al., 2022)	138.9	36.2	0.0164	0.0193	0.9927	0.9976	0.9990	ECCV 2022
	Ours	**114.9**	**35.3**	**0.0162**	**0.0174**	**0.9947**	**0.9983**	**0.9993**	–
3D60	UniFuse (Jiang et al., 2021)	215.6	94.1	0.0446	0.0342	0.9749	0.9947	0.9984	ICRA 2021
	HoHo-R (Sun et al., 2021)	196.9	75.6	0.0338	0.0294	0.9818	0.9954	0.9983	CVPR 2021
	HoHo-H (Sun et al., 2021)	205.8	81.9	0.0376	0.0317	0.9788	0.9947	0.9981	CVPR 2021
	PENet (Hu et al., 2021)	233.9	120.3	0.0680	0.0321	0.9743	0.9926	0.9980	ICRA 2021
	GuideNet (Hu et al., 2021)	239.3	144.2	0.0689	0.0418	0.9711	0.9954	0.9987	TIP 2021
	360Depth (Rey-Area et al., 2021)	225.4	93.7	0.0677	0.0315	0.9782	0.9936	0.9985	CVPR 2022
	M^{3}PT (Yan et al., 2022)	127.2	34.1	**0.0144**	0.0165	0.9944	0.9985	0.9995	ECCV 2022
	Ours	**102.8**	**31.9**	**0.0144**	**0.0142**	**0.9963**	**0.9991**	**0.9996**	–

Table 1. **Quantitative comparisons**. HoHo-R/H: HoHoNet employs ResNet (He et al., 2016)/HardNet (Chao et al., 2019) as backbone.

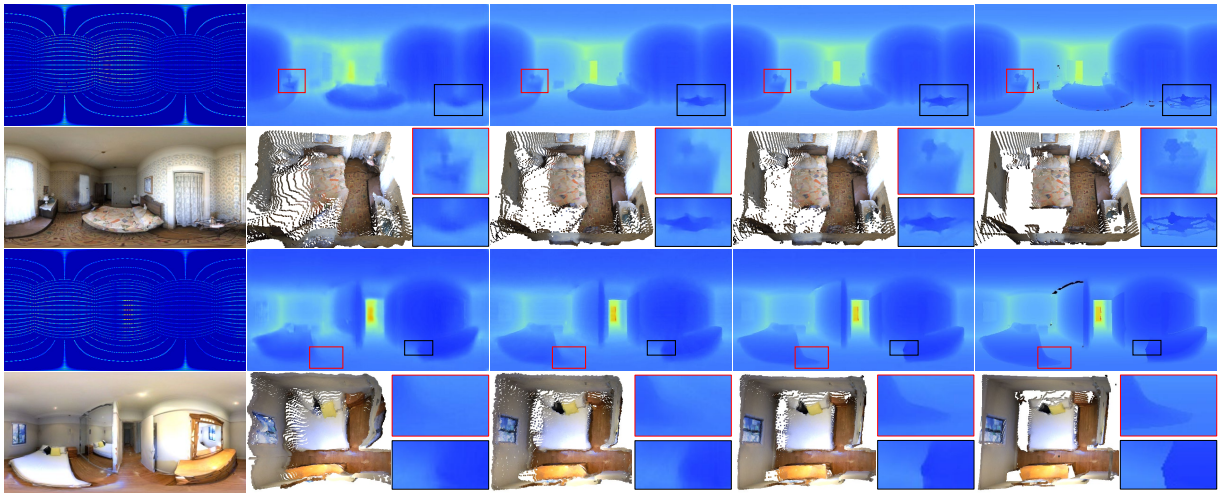


Figure 4. **Qualitative comparisons** of depth and 3D reconstruction with SoTA approaches on Matterport3D and 3D60.

relative error is $<1.25^t$ $(t = 1, 2, 3)$.

Training Settings. The whole training process is implemented on Pytorch with a single NVIDIA TITAN V GPU. AdamW optimizer is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and weight decay 0.05. We train the model for 80 epoches with batch size 16 and initial learning rate 5×10^{-4}, which drops by half every 20 epoches. Color jittering and random horizontal flip are used. μ and η are 80 and 0.5 respectively.

4.2. Comparisons with SoTA Methods

Our approach is compared with the following three categories of SoTA methods: (i) Dynamic convolution based GuideNet (Tang et al., 2020); (ii) Pre-training based M^{3}PT (Yan et al., 2022); (iii) multiple projections based UniFuse (Jiang et al., 2021), 360Depth (Rey-Area et al., 2021); and regular models HoHoNet (Sun et al., 2021), PENet (Hu et al., 2021). Tab. 1 and Fig. 4 present the quantitative and qualitative results, respectively. From Tab. 1 we can observe that, comparing with other complex solutions, our method obtains the lowest errors and the highest accuracy among all works on the two benchmarks, *e.g.*, averagely surpassing the best M^{3}PT by **11.9%** in Log and **18.2%** in RMSE, which is the primary metric for depth completion. It is worth noting that, our very lightweight loss function can be easily deployed with little extra overhead (see our Appendix C), including training/testing time and GPU memory usage. As demonstrated in Fig. 4, our approach can recover more clear and sharper object edges than other methods. The 3D reconstruction results further show the superiority of the proposed loss functions. For example, shapes of wall, desk, and bed are more complete and closer to the depth ground-truths.

Baseline	Loss	Matterport3D					3D60				
		RMSE	MAE	MRE	Log	σ_1	RMSE	MAE	MRE	Log	σ_1
M^{3}PT (Yan et al., 2022)	raw	138.9	36.2	0.0164	0.0193	0.9927	127.2	34.1	**0.0144**	0.0165	0.9944
	$\mathcal{L}_{DU}$	**114.9**	**35.3**	**0.0162**	**0.0174**	**0.9947**	**102.8**	**31.9**	**0.0144**	**0.0142**	**0.9963**
GuideNet (Tang et al., 2020)	raw	192.9	87.2	0.0438	0.0327	0.9806	239.3	144.2	0.0689	0.0418	0.9711
	$\mathcal{L}_{DU}$	**164.8**	**79.5**	**0.0370**	**0.0266**	**0.9886**	**192.3**	**121.0**	**0.0575**	**0.0363**	**0.9760**
HoHo-R (Sun et al., 2021)	raw	199.2	75.0	0.0355	0.0311	0.9806	196.9	75.6	0.0338	0.0294	0.9818
	$\mathcal{L}_{DU}$	**168.8**	**70.1**	**0.0340**	**0.0293**	**0.9865**	**168.3**	**67.6**	**0.0322**	**0.0281**	**0.9876**

Table 2. **Ablations of the total loss function $\mathcal{L}_{DU}$ using different baselines**, including M^{3}PT, GuideNet, and HoHo-R.

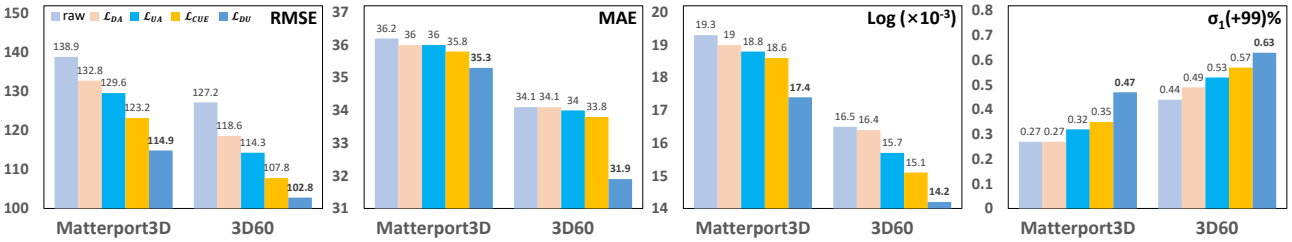


Figure 5. **Ablations of each component of $\mathcal{L}_{DU}$ using the same baseline** M^{3}PT. Here our $\mathcal{L}_{CUE}$ is deployed together with $\mathcal{L}_{UA}$.

Type of δ	RMSE	Δ	MAE	Δ
C, –	129.6	0.0	36.0	0.0
C, HF	**123.2**	**6.4** $\downarrow$	**35.8**	**0.2** $\downarrow$
C, VF	123.5	6.1 $\downarrow$	35.9	0.1 $\downarrow$
HF, VF	124.7	4.9 $\downarrow$	35.9	0.1 $\downarrow$

Table 3. **Ablations of the type of the uncertainty map δ in $\mathcal{L}_{CUE}$** employing M^{3}PT on Matterport3D.

Number of δ	RMSE	Δ	MAE	Δ
1 (C)	129.6	0.0	36.0	0.0
2 (C, HF)	123.2	6.4 $\downarrow$	35.8	0.2 $\downarrow$
3 (C, HF, VF)	**123.0**	**6.6** $\downarrow$	35.8	0.2 $\downarrow$
4 (C, HF, VF, R)	**123.0**	**6.6** $\downarrow$	**35.7**	**0.3** $\downarrow$

Table 4. **Ablations of the number of the uncertainty map δ in $\mathcal{L}_{CUE}$** employing M^{3}PT on Matterport3D. R: rotation.

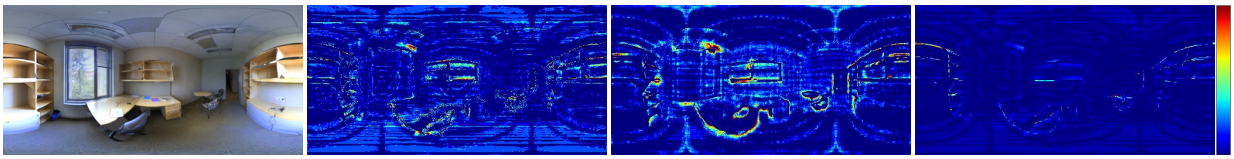
4.3. Ablation Studies

Since only M^{3}PT (Yan et al., 2022) is specifically designed for the *panoramic depth completion* task, we select two other kinds of related works as baselines, *i.e.*, SoTA *perspective depth completion* method GuideNet (Tang et al., 2020) and SoTA *panoramic depth estimation* approach HoHo-R (Sun et al., 2021). HoHo-R employs a single branch with ResNet (He et al., 2016) as backbone, while GuideNet conducts dual branches, based on which M^{3}PT uses multi-modal masked pre-training strategy to refine.

Ablations of the total loss function $\mathcal{L}_{DU}$ using different baselines. As reported in Tab. 2, $\mathcal{L}_{DU}$ dramatically improves all three baselines. For example, the RMSE of HoHo-R-$\mathcal{L}_{DU}$ is $30.1mm$ lower than that of HoHo-R-raw. In addition, based on M^{3}PT, we discover that $\mathcal{L}_{DU}$ performs not very well in MRE. We conclude that, $\mathcal{L}_2$ in M^{3}PT is more sensitive to large depth values than to the small. Hence, long-range depth error can be better optimized. However, MRE is sensitive to close-range depth error. As a result, variations on $\mathcal{L}_2$ can not markedly reduce MRE.

Ablations of each component of $\mathcal{L}_{DU}$ using the same baseline. Based on M^{3}PT, from Fig. 5 we can observe **(1)** $\mathcal{L}_{DA}$ partly improves the baseline, *e.g.*, the RMSE is reduced by $5.9mm$ and $8.4mm$ on two datasets respectively. **(2)** $\mathcal{L}_{UA}$ slightly performs better than $\mathcal{L}_{DA}$ and further reduces the errors and increase the accuracy. **(3)** Based on $\mathcal{L}_{UA}$, we introduce $\mathcal{L}_{CUE}$ to learn more precise uncertainty maps. Evidently, it is superior to $\mathcal{L}_{UA}$ in all metrics and significantly benefits the model. For example, it severally reduces RMSE by $15.7mm$ and $19.4mm$ on Matterport3D and 3D60. Fig. 3 demonstrates that the $\mathcal{L}_{CUE}$ uncertainty map is more exact. **(4)** Finally, we organically combine the three loss functions termed $\mathcal{L}_{DU}$. Notably, $\mathcal{L}_{DU}$ promotes the baseline by a large margin, *e.g.*, successively decreasing RMSE by $23.9mm$ and $24.1mm$ on two datasets severally.

Ablations of the type and the number of the uncertainty map δ in $\mathcal{L}_{CUE}$. In Tab. 3, different types of the uncertainty map δ in $\mathcal{L}_{CUE}$ are generated by the pure convolution (C), horizontal flip (HF), and vertical flip (VF). It is observed that the combination of "C, HF" (default setting) slightly outperforms others. In Tab. 4, we predict different numbers of the uncertainty map δ in $\mathcal{L}_{CUE}$ to see their influences. Notably, the model achieves the best performance when the number is 4. For simplicity however, we select "Num.=2" as the default. As shown in the second and the third columns of Fig. 6, uncertainty maps of the same panorama upon pure convolutions with and without flip, are very blurry and are quite distinct from each other. It does not conform to the common sense that a given scene has relatively definite uncertainty. By introducing our CUE (Num.=2), as shown in the last column of Fig. 6, the uncertainty estimation is apparently unambiguous and contains fewer uncertain regions, contributing to more accurate 3D reconstruction results.



rgb Num.=1 Num.=1 w/ flip Num.=2

Figure 6. **Visual comparisons of different uncertainty maps** on Matterport3D. We employ convolution and flip to generate uncertainty maps. Common PDC baselines only estimate one uncertainty map (Num.=1) while our CUE produces two (Num.=2).

Method	RMSE	MAE	MRE	Log
BiFuse (Wang et al., 2020)	0.6259	0.3470	0.2048	0.1134
UniFuse (Jiang et al., 2021)	0.4941	0.2814	0.1063	0.0701
HoHo-R (Sun et al., 2021)	0.5138	0.2862	0.1488	0.0871
SliceNet (Pintore et al., 2021)	0.6133	0.3296	0.1764	0.1045
Sphere (Coors et al., 2018)	0.5212	0.3167	0.1258	0.0778
ACDNet (Zhuang et al., 2022)	0.4629	0.2670	0.1010	0.0646
Ours	**0.4307**	**0.2538**	**0.0926**	**0.0562**

Table 5. Quantitative comparisons of panoramic depth estimation on Matterport3D, whose input is a single panoramic color image.

Method	RMSE	MAE	iRMSE	iMAE
SConv (Uhrig et al., 2017)	1601.33	481.27	4.94	1.78
ADNN (Chodosh et al., 2018)	1325.37	439.48	-	-
NCNN (Eldesokey et al., 2018)	1268.22	360.28	4.67	1.52
S2D (Ma et al., 2019)	954.36	288.64	**3.21**	1.35
NConv (Eldesokey et al., 2020)	954.34	258.68	3.40	1.17
Ours	**943.48**	**256.46**	3.27	**1.13**

Table 6. Quantitative comparisons of perspective depth completion on KITTI (Uhrig et al., 2017) benchmark.

4.4. Generalization Capabilities

Double-modal to single-modal: panoramic depth estimation. The panoramic depth completion task focuses on *RGB-D* double-modal data, whilst the panoramic depth estimation task pays attention to *RGB* single-modal data. Since panoramic RGB images also suffer from distortion and uncertainty, we employ the proposed distortion and uncertainty aware loss $\mathcal{L}_{DU}$ to see its generalization capability on the panoramic depth estimation task. The results are listed in Tab. 5 and the baseline is ACDNet (Zhuang et al., 2022). Note that the dataset used in Tab. 5 is the same with that in ACDNet. From the table we can observe that, our $\mathcal{L}_{DU}$ consistently improves the baseline, achieving superior or competitive performance on Matterport3D dataset. For example, our $\mathcal{L}_{DA}$ and $\mathcal{L}_{UA}$ averagely reduce RMSE and Log by **6.96%** and **13.00%**, respectively.

Panoramic to perspective: perspective depth completion. Different from *panoramic* data, *perspective* data does not contain distortion. Hence, we perform the uncertainty-aware loss $\mathcal{L}_{UA}$ with consistent uncertainty estimation (CUE) module, to capture more precise uncertainty representation to benefit the perspective depth completion task. The results are reported in Tab. 6 and the baseline is NConv (Eldesokey et al., 2020). We can find that, our $\mathcal{L}_{UA}$ with CUE congruously improves the baseline on KITTI depth completion benchmark, *e.g.*, the RMSE is reduced from $954.34mm$ to $943.48mm$, about $11mm$ improvement. Those numerical results indicate that our DUL generalizes well.

5. Conclusion

In this paper, we proposed the joint distortion and uncertainty loss (DUL) for panoramic depth completion task, which suffered from the inherent distortion and the high uncertainty all along. DUL could be specialized into the distortion-aware loss (DAL) and the uncertainty-aware loss (UAL), where DAL encouraged to distribute each pixel with uneven importance to mitigate the negative effect of the inherent distortion, and UAL enabled networks to predict precise depth via modeling uncertainty. Furthermore, based on UAL we presented to learn the consistency between different uncertainty maps of the same panorama by introducing the consistent uncertainty estimation (CUE) module, aiming to acquire more accurate uncertainty representation. Extensive experiments validated the effectiveness of DUL.

Limitation discussion: **(1)** In Eqs. 5 and 6, the distortion-aware weight is learned based on the mapping relationship between spherical and planer coordinates via MLP layers. Our initial goal of employing MLP is to build a residual connection and then fine-tune the fixed weight. However, the explainability of the MLP is not very clear. What we want to do next is to learn a prior, which simultaneously considers the mapping relationship and object distribution, to enhance the explainability whilst balancing the contribution of each pixel. **(2)** We hope our DUL design could provide some inspiration for other related research areas. Thus, in future we will test the effectiveness of our method on the most popular monocular perspective depth estimation task.

Acknowledgements

This work was supported by the National Science Fund of China (U1713208, 62072242), and the Young Scientists Fund of the National Natural Science Foundation of China (62206134). Note that the PCA Lab is associated with, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, China.

References

Albanis, G., Zioulis, N., Drakoulis, P., Gkitsas, V., Sterzentsenko, V., Alvarez, F., Zarpalas, D., and Daras, P. Pano3d: A holistic benchmark and a solid baseline for 360° depth estimation. In *CVPRW*, pp. 3722–3732. IEEE, 2021.

Chao, P., Kao, C.-Y., Ruan, Y.-S., Huang, C.-H., and Lin, Y.-L. Hardnet: A low memory traffic network. In *ICCV*, pp. 3552–3561, 2019.

Cheng, X., Wang, P., Guan, C., and Yang, R. Cspn++: Learning context and resource aware convolutional spatial propagation networks for depth completion. In *AAAI*, pp. 10615–10622, 2020.

Chodosh, N., Wang, C., and Lucey, S. Deep convolutional compressed sensing for lidar depth completion. In *ACCV*, pp. 499–513. Springer, 2018.

Choi, J., Chun, D., Kim, H., and Lee, H.-J. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *ICCV*, pp. 502–511, 2019.

Coors, B., Condurache, A. P., and Geiger, A. Spherenet: Learning spherical representations for detection and classification in omnidirectional images. In *ECCV*, pp. 518–533, 2018.

Eder, M., Moulon, P., and Guan, L. Pano popups: Indoor 3d reconstruction with a plane-aware network. In *3DV*, pp. 76–84. IEEE, 2019.

Eldesokey, A., Felsberg, M., and Khan, F. S. Propagating confidences through cnns for sparse data regression. In *BMVC*, 2018.

Eldesokey, A., Felsberg, M., Holmquist, K., and Persson, M. Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end. In *CVPR*, pp. 12014–12023, 2020.

Figueiredo, M. Adaptive sparseness using jeffreys prior. *NeurIPS*, 14, 2001.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *CVPR*, pp. 770–778, 2016.

He, Y., Zhu, C., Wang, J., Savvides, M., and Zhang, X. Bounding box regression with uncertainty for accurate object detection. In *CVPR*, pp. 2888–2897, 2019.

Hu, M., Wang, S., Li, B., Ning, S., Fan, L., and Gong, X. Penet: Towards precise and efficient image guided depth completion. In *ICRA*, 2021.

Ilg, E., Cicek, O., Galesso, S., Klein, A., Makansi, O., Hutter, F., and Brox, T. Uncertainty estimates and multi-hypotheses networks for optical flow. In *ECCV*, pp. 652–667, 2018.

Jiang, H., Sheng, Z., Zhu, S., Dong, Z., and Huang, R. Unifuse: Unidirectional fusion for 360 panorama depth estimation. *IEEE Robotics and Automation Letters*, 6(2): 1519–1526, 2021.

Jin, L., Xu, Y., Zheng, J., Zhang, J., Tang, R., Xu, S., Yu, J., and Gao, S. Geometric structure based and regularized depth estimation from 360 indoor imagery. In *CVPR*, pp. 889–898, 2020.

Kendall, A. and Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? *NeurIPS*, 30, 2017.

Lee, Y., Jeong, J., Yun, J., Cho, W., and Yoon, K.-J. Spherephd: Applying cnns on a spherical polyhedron representation of 360deg images. In *CVPR*, pp. 9181–9189, 2019.

Lee, Y., Jeong, J., Yun, J., Cho, W., and Yoon, K.-J. Spherephd: Applying cnns on 360° images with non-euclidean spherical polyhedron representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

Lin, Y., Cheng, T., Zhong, Q., Zhou, W., and Yang, H. Dynamic spatial propagation network for depth completion. In *AAAI*, 2022.

Ma, F., Cavalheiro, G. V., and Karaman, S. Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In *ICRA*, 2019.

MacKay, D. J. A practical bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472, 1992.

Ning, Q., Dong, W., Li, X., Wu, J., and Shi, G. Uncertainty-driven loss for single image super-resolution. *NeurIPS*, 34, 2021.

Park, J., Joo, K., Hu, Z., Liu, C.-K., and Kweon, I. S. Non-local spatial propagation network for depth completion. In *ECCV*, 2020.

Pintore, G., Agus, M., Almansa, E., Schneider, J., and Gobbetti, E. Slicenet: deep dense depth estimation from a single indoor panorama using a slice-based representation. In *CVPR*, pp. 11536–11545, 2021.

Poggi, M., Aleotti, F., Tosi, F., and Mattoccia, S. On the uncertainty of self-supervised monocular depth estimation. In *CVPR*, pp. 3227–3237, 2020.

Rey-Area, M., Yuan, M., and Richardt, C. 360monodepth: High-resolution 360° monocular depth estimation. *arXiv e-prints*, pp. arXiv–2111, 2021.

Sun, C., Sun, M., and Chen, H.-T. Hohonet: 360 indoor holistic understanding with latent horizontal features. In *CVPR*, pp. 2573–2582, 2021.

Tang, J., Tian, F.-P., Feng, W., Li, J., and Tan, P. Learning guided convolutional network for depth completion. *IEEE Transactions on Image Processing*, 30:1116–1129, 2020.

Tateno, K., Navab, N., and Tombari, F. Distortion-aware convolutional filters for dense prediction in panoramic images. In *ECCV*, pp. 707–722, 2018.

Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., and Geiger, A. Sparsity invariant cnns. In *3DV*, pp. 11–20, 2017.

Van Gansbeke, W., Neven, D., De Brabandere, B., and Van Gool, L. Sparse and noisy lidar completion with rgb guidance and uncertainty. In *MVA*, pp. 1–6, 2019.

Wang, F.-E., Yeh, Y.-H., Sun, M., Chiu, W.-C., and Tsai, Y.-H. Bifuse: Monocular 360 depth estimation via bi-projection fusion. In *CVPR*, pp. 462–471, 2020.

Welling, M. and Teh, Y. W. Bayesian learning via stochastic gradient langevin dynamics. In *ICML*, pp. 681–688. Citeseer, 2011.

Xu, K., Yang, X., Yin, B., and Lau, R. W. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pp. 2281–2290, 2020.

Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., and Li, H. Depth completion from sparse lidar data with depth-normal constraints. In *ICCV*, pp. 2811–2820, 2019.

Yan, Z., Li, X., Wang, K., Zhang, Z., Li, J., and Yang, J. Multi-modal masked pre-training for monocular panoramic depth completion. *arXiv preprint arXiv:2203.09855*, 2022.

Yang, G. and Ramanan, D. Volumetric correspondence networks for optical flow. *NeurIPS*, 32, 2019.

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., and Zhao, G. Searching central difference convolutional networks for face anti-spoofing. In *CVPR*, pp. 5295–5305, 2020.

Zhu, Y., Dong, W., Li, L., Wu, J., Li, X., and Shi, G. Robust depth completion with uncertainty-driven loss functions. In *AAAI*, 2022.

Zhuang, C., Lu, Z., Wang, Y., Xiao, J., and Wang, Y. Acdnet: Adaptively combined dilated convolution for monocular panorama depth estimation. In *AAAI*, 2022.

Zioulis, N., Karakottas, A., Zarpalas, D., and Daras, P. Omnidepth: Dense depth estimation for indoors spherical panoramas. In *ECCV*, pp. 448–465, 2018.

Zioulis, N., Karakottas, A., Zarpalas, D., Alvarez, F., and Daras, P. Spherical view synthesis for self-supervised 360 depth estimation. In *3DV*, pp. 690–699. IEEE, 2019.

A. Overview

This document provides additional technical details. Specifically, we first introduce the standard metrics in Section B. Then we conduct more ablation studies in Sections C and D.

B. Metrics

iRMSE and iMAE are calculated based on RMSE and MAE by replacing y and $f(x)$ with the inverse. The standard metrics are defined as:

$$
\begin{aligned}
&\text{– MRE}: \frac{1}{N}\sum\left|\frac{y-f(x)}{y}\right| \\
&\text{– MAE}: \frac{1}{N}\sum|y-f(x)| \\
&\text{– RMSE}: \sqrt{\frac{1}{N}\sum(y-f(x))^2} \\
&\text{– } \sigma_t: \frac{1}{N}\left|\max\left(\frac{y}{f(x)},\frac{f(x)}{y}\right)<1.25^t\right| \\
&\text{– RMSElog}: \sqrt{\frac{1}{N}\sum(\log y-\log f(x))^2}
\end{aligned}
$$

C. Ablation on Complexity

Our distortion and uncertainty aware loss $\mathcal{L}_{DU}$ introduces the very lightweight Multi-Layer Perceptron (MLP) and consistent uncertainty estimation (CUE) module to redeploy baselines when training. Here we show their complexity based on GuideNet (Tang et al., 2020). As reported in Tab. 7, the parameter with $\mathcal{L}_{DU}$ only increases from 73.536M to 73.539M, the training time is changeless, and the inference speed is only 0.0225ms slower. These facts give strong evidence that the proposed $\mathcal{L}_{DU}$ is lightweight enough.

Method	Parameter (M)	Train (h)	Test (ms)
GuideNet (Tang et al., 2020)	73.536	9.4	17.3627
+$\mathcal{L}_{DU}$	73.539	9.4	17.3852

Table 7. Complexity analysis based on GuideNet.

D. Ablation on Hyper Parameters

Ablations of μ in $\mathcal{L}_{DA}$, and η in $\mathcal{L}_{CUE}$. We severally set five values for hyper parameters μ and η in Tab. 8 and Tab. 9. Tab. 8 indicates that the model with $\mathcal{L}_{DA}$ achieves best performance when $\mu = 80$, which is the default value. Tab. 9 shows that the baseline using $\mathcal{L}_{CUE}$ performs better when $\eta = 0.1$ or 0.5. For the trade-off between RMSE and MAE, we select $\eta = 0.5$ as the default.

μ ($\mathcal{L}_{DA}$)	RMSE	Δ	MAE	Δ
20	138.5	0.0	36.2	0.0
40	136.8	1.7 ↓	36.2	0.0
80	**132.8**	**5.7** ↓	**36.0**	**0.2** ↓
120	133.3	5.2 ↓	**36.0**	**0.2** ↓
160	134.6	3.9 ↓	36.1	0.1 ↓

Table 8. **Ablations of μ in $\mathcal{L}_{DA}$** employing M^{3}PT.

η ($\mathcal{L}_{CUE}$)	RMSE	Δ	MAE	Δ
0.01	128.9	0.0	36.3	0.0
0.1	**123.1**	**5.8** ↓	36.0	0.3 ↓
0.5	123.2	5.7 ↓	**35.8**	**0.5** ↓
1	125.9	4.0 ↓	35.9	0.2 ↓
2	127.4	1.5 ↓	36.2	0.1 ↓

Table 9. **Ablations of η in $\mathcal{L}_{CUE}$** employing M^{3}PT.