Gm-APD LiDAR Single-Source Data Self-Guided: Obtaining High-Resolution Depth Map

Boyi Song, Jianfeng Sun ⓘ, *Member, IEEE*, Yongji Yan ⓘ, Rui He, Xin Zhou, and Wei Lu

Abstract—Geiger-mode avalanche photodiode (Gm-APD) array LiDAR has become the current research focus due to its sensitive response, high precision, and easy integration. However, due to the limitations of the fabrication process and manufacturing cost, the images collected by Gm-APD array LiDAR have very serious image low-resolution problems. Since the intensity map and depth map of the target can be obtained simultaneously by relying on single-source data from Gm-APD array LiDAR, we propose a Gm-APD LiDAR single-source data self-guided method. We propose to first perform Gm-APD LiDAR intensity map superresolution, and then use the processed intensity map with the corresponding depth map for guided superresolution. The advantage of this process is that instead of using high-resolution (HR) imaging devices from different domains, it relies only on single-source data from Gm-APD LiDAR to obtain HR depth map, thus eliminating the need for additional image registration work and providing wider applicability. We investigate the feasibility of the proposed single-source data processing method and evaluate our method on the real Gm-APD LiDAR single-source data with an average peak signal-to-noise ratio of 42.21, which better preserves the original distance information of the targets while providing visually sharper outputs.

Index Terms—Geiger-mode avalanche photodiode (Gm-APD) LiDAR, noise suppression, self-guided method, single-source data.

I. INTRODUCTION

SINGLE-PHOTON counting LiDAR has become the current research focus for many remote sensing applications requiring three-dimensional mapping [1], [2]. Due to the sensitive response, high precision, and low echo energy requirements of single-photon detectors, single-photon counting LiDAR-based depth imaging has emerged as a candidate technology for collecting target intensity and depth information in challenging environments [3]. Compared to the traditional scanning LiDAR systems, Geiger-mode avalanche photodiode (Gm-APD) LiDAR with single-photon detection has the capability of both single-photon detection and single-frame depth imaging, which is more advantageous for acquiring depth information of weak signal or long-distance targets.

A. Related Works

Depth maps provide three-dimensional information that directly characterizes the geometry of the target surface and are widely used in many remote sensing applications [4], [5], [6], [7], [8]. However, depth sensors, such as LiDAR, suffer from very serious image low-resolution (LR) problems because depth sensors can only provide depth maps of limited resolution under suboptimal conditions due to manufacturing process limitations and manufacturing costs. Given that it is too costly to solve the image LR problems from hardware, from the perspective of software and algorithm, based on the input LR depth maps, superresolving the undersampling information of LR depth sensors during data acquisition to obtain the high-resolution (HR) depth maps [9], [10], [11], [12], [13] and achieving depth map superresolution (DSR) has become the current research focus in various fields, such as image processing and computer vision [14]. Depth maps can be collected by LiDAR depth imaging, computer stereo vision imaging [15], [16], but the depth maps collected by LiDAR are generated from real-world three-dimensional points [14] and generally contain less textures and more sharp boundaries, so DSR for LiDAR is a challenging task. In addition, it is difficult to supersolve the undersampling information based only on the existing depth information because LR depth maps have almost no texture, which is why computer vision currently uses RGB-D composite data for guided superresolution to process depth maps.

Although numerous algorithms have been proposed for DSR and presented impressive performance, existing DSR is generally focused on RGB-D datasets [9], [13], [17], [18], [19], [20]. As the sharp boundaries and intricate details in DSR are difficult to recover, color-image-guided methods have been introduced to solve this problem [9], [21], [22], [23], [24]. We argue that the following factors need to be considered when using guided superresolution for DSR tasks in realistic devices. First, accurate image registration is a prerequisite for DSR tasks when using multisource data (e.g., RGB-D datasets) for guided superresolution. However, for realistic devices, depth maps obtained from depth sensors are always in LR and cannot match the resolution of conventional camera images, so accurate image registration is challenging for realistic devices. Second, RGB images contain rich detail and color information, so multiscale structure guidance and color guidance of the target can be performed. However, the image obtained by realistic devices (e.g., LiDAR) is usually a single-band grey-scale image with extremely monotonous pixel grey-scale compared to the HR
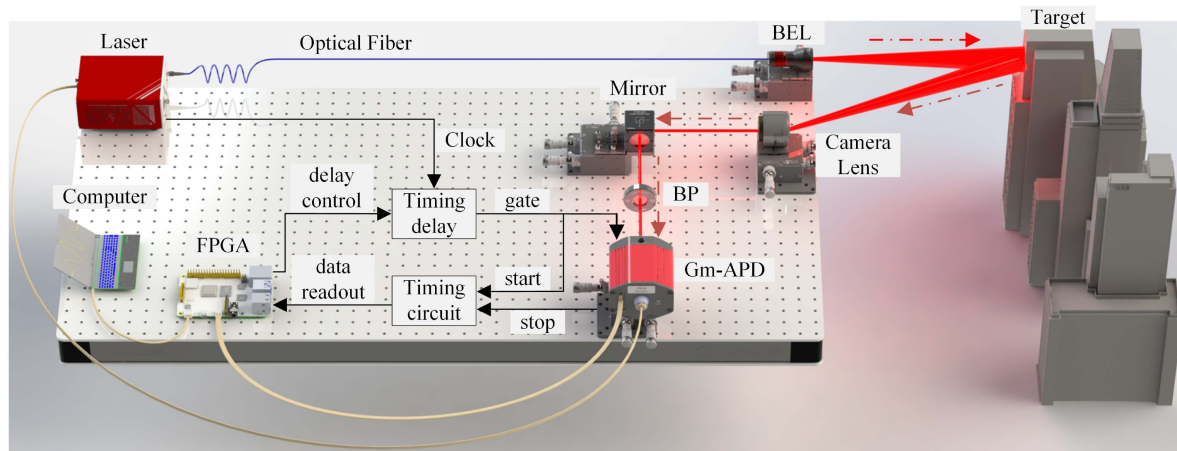
Fig. 1. Signal acquisition equipment system of Gm-APD array LiDAR. BEL: Beam expander lens. BP: Bandpass filter at 1064 nm with linewidth of 1.0 nm.

conventional camera image, so the imaging characteristics of realistic devices must be considered.

B. Motivations and Contributions

For the Gm-APD array LiDAR, the following device imaging analysis we performed. The Gm-APD array LiDAR uses the photon time-of-flight method to collect depth information of the target [25], and uses the time-to-digital converter (TDC) for timing [26], [27]. Each pixel in the Gm-APD array is given the time-of-flight recorded by the TDC at the corresponding TDC trigger frequency. The TDC trigger frequency corresponds to the pixel value of the corresponding pixel in the Gm-APD LiDAR intensity map, and the flight time recorded by TDC corresponds to the pixel value of the corresponding pixel in the Gm-APD LiDAR depth map. The resulting intensity map matches the depth map pixel by pixel, so no additional image registration work is required. Based on this characteristic of Gm-APD LiDAR, this article proposes the Gm-APD LiDAR single-source data self-guided method: Gm-APD LiDAR composed of 64×64 pixels is used for target imaging, then Gm-APD LiDAR intensity map superresolution is performed, and then the processed intensity map with the corresponding depth map is used for guided superresolution. This method has the advantage of relying only on single-source data from Gm-APD LiDAR to obtain HR depth map, and thus has a wider applicability, whereas our method provides the outputs of HR data that better preserve the original distance information.

In this article, the Gm-APD LiDAR single-source data processing method is presented. By analyzing the single-source data of Gm-APD LiDAR, after obtaining the HR intensity map by enhanced superresolution generative adversarial networks (ESRGANs), the HR depth map is simultaneously obtained by the parallel two-branches network based on neighbor pixels guidance, and this method eliminates the guidance error due to the accuracy of image registration. Finally, testing on the data collected in our outfield experiments, the average peak signal-to-noise ratio (PSNR) reached 42.21, demonstrating the feasibility of obtaining HR intensity map and depth map simultaneously

using only single-source data from 64×64 pixels Gm-APD array LiDAR photographic imaging.

II. GM-APD ARRAY LiDAR WORKING MECHANISM

The signal acquisition equipment system of Gm-APD array LiDAR is shown in Fig. 1. The laser emits periodic pulses to the target. The 64×64 pixels Gm-APD camera is used to receive the echo photons. Each pixel of the Gm-APD camera is single-photon sensitive with a time resolution of 1 ns. Field-programmable gate array is used to control the sending of commands and the transmission of collection data. The collection target data are displayed on the computer [26].

A. Data Acquisition

The data used in this article are all collected in our outfield experiments. The collection area is located in Harbin, China, and the collection targets are urban high-rise buildings.

Because the Gm-APD camera we used only has a 64×64 pixels array, too far away from the urban high-rise buildings would not be able to capture enough detail information about the target (e.g., doors, windows, etc.), a distance as far away from the target as possible but still able to capture enough detail information needs to be selected. During the data collection process, based on the actual situation, we selected the detection targets to be approximately 500 m away from the LiDAR system.

We use the LiDAR with center wavelength of 1064 nm and use the photon time-of-flight method to collect depth information of the target [25], [28].

B. Signal Analysis

Given that Gm-APD can only respond to the presence of echo photons, namely, the detector obtains the target signal with an output of 1 or 0, it cannot capture intensity information of the target by the amplitude of the output pulse, such as the linear APD. Therefore, we use the accumulated detection target signal restoration method based on the Poisson probability response model and the center-of-mass algorithm to obtain the
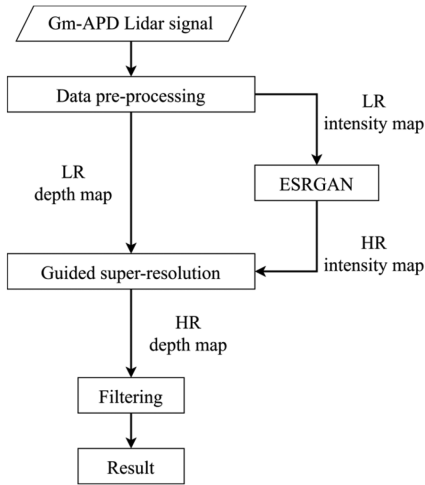
Fig. 2.   Overall framework for Gm-APD LiDAR single-source data self-guided method.

Fig. 3. k-nearest neighbor correction for Gm-APD LiDAR depth maps.

depth information and intensity information of the target [29]. Compared with the HR conventional camera using linear APD, the Gm-APD LiDAR intensity map obtained from a single-band pulse echo does not respond well to the subtle differences in the target intensity. Therefore, the poor information characteristics of the Gm-APD LiDAR intensity map need to be considered in guided superresolution. In addition, the accumulated detection acquires intensity map and depth map of the target, the common feature of which is that each pixel in the Gm-APD array gets a corresponding time bin. The flight time recorded by TDC with the corresponding TDC trigger frequency can be obtained according to the time bin. Where the TDC trigger frequency corresponds to the pixel value of the corresponding pixel in the Gm-APD LiDAR intensity map, and the flight time recorded by TDC corresponds to the pixel value of the corresponding pixel in the Gm-APD LiDAR depth map. The resulting intensity map matches the depth map pixel by pixel, so no additional image registration work is required. This correspondence between the Gm-APD LiDAR intensity map and depth map eliminates the guidance error due to the accuracy of image registration.

III. THEORETICAL MODEL FOR SINGLE-SOURCE DATA SELF-GUIDED

Given that the intensity map and depth map obtained by Gm-APD LiDAR do not require any image registration work, we propose the overall framework, as shown in Fig. 2, for Gm-APD LiDAR single-source data self-guided method. In this method, after preprocessing the data, ESRGAN is first used to process the Gm-APD LiDAR intensity map [30]. The processed intensity map is then used with the corresponding depth map for guided superresolution. In addition, a filtering algorithm is developed to remove the bad pixels from the obtained HR depth map.

A. Data Preprocessing

Due to the sensitive response, high accuracy, and low echo energy requirements of single-photon detectors, abnormal noise
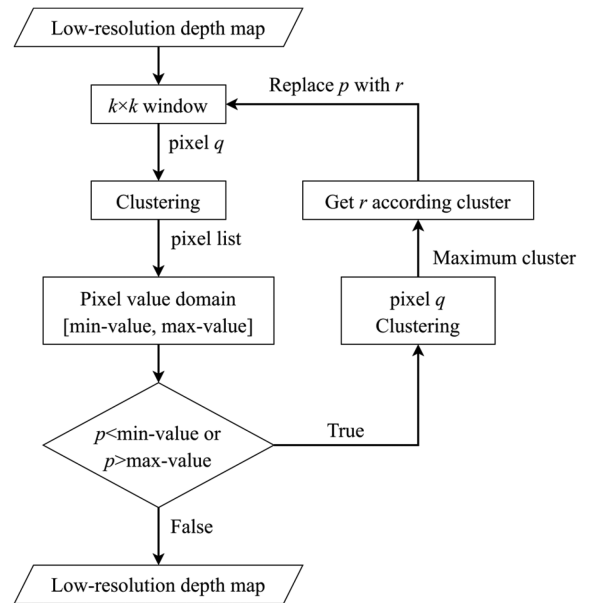
suppression is a prerequisite for Gm-APD LiDAR to achieve precise sensing. In its data processing, it is particularly important to algorithmically separate the signal and noise photons, so some efficient single-photon depth and reflectivity estimation algorithms have been proposed. However, even with relevant signal processing algorithms, some abnormal noise is still present in the extracted Gm-APD LiDAR images, and the LR depth maps of the network inputs are used as training labels. It is obvious that the abnormal noise causes distance errors in guided super-resolution, so it is necessary to suppress the noise in Gm-APD LiDAR images.

For LiDAR depth maps, from the visual perception of the observer, a pixel should be considered as abnormal noise if its pseudocolor (the value of the pixel) is clearly different from its neighboring pixels. Therefore, we propose the k-nearest neighbor correction algorithm shown in Fig. 3: pixel p is selected to form a $k \times k$ window with its neighbor pixels q. All pixels q are clustered to obtain the distance value domain of each window, and pixel p is replaced if the value of pixel p exceeds the distance value domain of the window in which it is located. For the replacement value r of pixel p, we argue that although weighted computation is widely used in intensity maps (e.g., bilinear, bicubic, etc.), weighted computation cannot be used in the preprocessing of depth map to obtain a distance value. Therefore, a reasonable replacement value for a distance value can only be selected in the $k \times k$ window. Similarly, all pixels q are clustered and the replacement value r is selected in the cluster with the highest number of pixels.

In the practical application of the k-nearest neighbor correction algorithm, a two-dimensional operator that slides on the depth map similar to the filtering process has been designed, and the fixed value of 0 is used for the boundary fill of the depth map, after which the pixels in the $k \times k$ window are one-dimensionally expanded before the process shown in Fig. 3. At the end, the
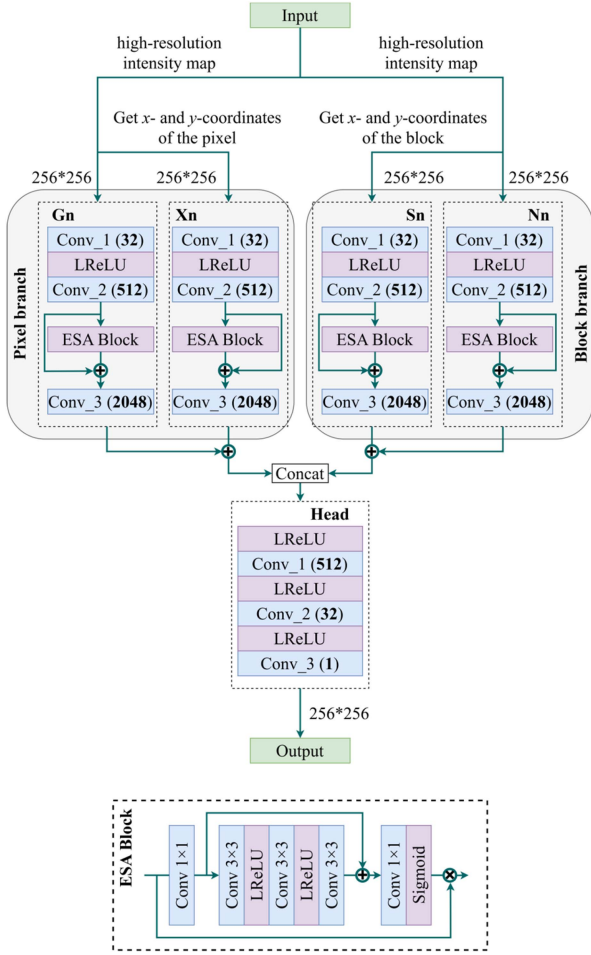
Fig. 4. Pipeline of the proposed guided superresolution.

boundary fill value of the depth map is removed. We define $S(p, k)$ as the k-nearest neighbor of pixel p. $S(p, k)$ can be expressed as follows:

$$S(p, k) = \left\{ x \left| p - \frac{k^2 - 1}{2} \leq x \leq p + \frac{k^2 - 1}{2}, x \in N \right. \right\}.$$

We denote the clustering process of pixels q as

$$\mathcal{H}(\boldsymbol{Q}, th) = \sum \text{saturate}(th - |q_j - q_i|)$$

where $\boldsymbol{Q}$ represents the obtained cluster, th is the distance difference between maximum value and minimum value of pixels in the cluster. Based on aforementioned definitions, the k-nearest neighbor correction algorithm can be expressed as

$$p = \begin{cases} p, & \mathcal{H}(\boldsymbol{Q}, th) \geq \text{num}_1 \text{ and } p \in \text{domain} \\ r, & \mathcal{H}(\boldsymbol{Q}, th) \geq \text{num}_1 \text{ and } p \notin \text{domain} \end{cases}$$

$$r = \begin{cases} \mathcal{G}(\boldsymbol{Q}, \text{flag}), & \mathcal{H}(\boldsymbol{Q}, th) \geq \text{num}_2 \\ 0, & \mathcal{H}(\boldsymbol{Q}, th) < \text{num}_2 \end{cases}$$

$$\text{domain} = [\min(\boldsymbol{Q}) - \alpha \cdot th, \max(\boldsymbol{Q}) + \beta \cdot th] \tag{1}$$

where $\mathcal{G}(\boldsymbol{Q}, \text{flag}) = \text{flag}(\boldsymbol{Q})$. In order to retain the original distance information as much as possible, we use a relaxed

solution condition: we take num_1 to be k^2-2 and num_2 to be $(k^2$-1)/2, taking $\alpha = \beta = 1$.

In (1), $\mathcal{H}(\boldsymbol{Q}, th)$ denotes the number of pixels contained in the cluster. If the number of pixels contained in the cluster exceeds num_1, it means that p is a valid pixel, where domain denotes the distance value domain of the cluster, and pixel p is replaced if it is a valid pixel and does not belong to domain. In addition, the replacement value r is a valid value if the number of pixels contained in the cluster exceeds num_2, otherwise it is assigned the null value 0.

For LiDAR intensity map, random scattering originating from the target within the basic resolution cell generates speckle noise, which degrades the image quality of the intensity map superresolution results and affects the information extraction of the HR intensity map in the subsequent guided superresolution. Therefore, we sequentially use the k-nearest neighbor correction algorithm, bilateral filtering with a 5×5 kernel for intensity map preprocessing.

B. Guided Superresolution Module

We denote the preprocessed LR Gm-APD LiDAR intensity map as I^{LR}, the HR intensity map that we want to recover as I^{HR}, and the solution architecture using ESRGAN [30]

$$I^{\text{HR}} = F_{\text{ESRGAN}}\left(I^{\text{LR}}; \Theta\right). \tag{2}$$

The original LR Gm-APD intensity map corresponds to the depth map pixel by pixel, but the HR intensity map recovered by the superresolution network destroys this correspondence, so we perform a pixel correction of the HR intensity map to ensure that the recovered HR intensity map still maintains the pixel correspondence with the depth map. The pixel correction of the HR intensity map can be expressed as

$$p_{ij} = \begin{cases} \min\left(I^{HR}\right), & \text{upsample}\left(I^{LR}\right)_{ij} \neq 0 \text{ and } p_{ij} = 0 \\ 0, & \text{upsample}\left(I^{LR}\right)_{ij} = 0 \text{ and } p_{ij} \neq 0 \end{cases} \tag{3}$$

where we propose to use nearest neighbor to describe the up-sampling progress.

For real Gm-APD LiDAR single-source data, there is no corresponding HR reference images, so only unsupervised method can be used. We denote the LR Gm-APD LiDAR depth map as D^{LR}, the HR depth map that we want to recover as D^{HR}, where D^{HR} has the same resolution as I^{HR}, and the up-sampling factor from D^{LR} to D^{HR} is $\times k$. The recovered HR intensity map I^{HR} can represent the contour structure of the target, whereas the corresponding LR depth map D^{LR} can represent the constraint of down-sampled distance value. Therefore, drawing on the multiscale guided convolutional network, a parallel two-branches structure is used to obtain D^{HR} from I^{HR}, and the down-sampled version of D^{HR} with D^{LR} is used to forming loss function that constrains the output of the network.

For the mean square error (MSE) loss function used, it can be expressed as the pixel s in D^{LR} corresponds to $k \times k$ pixels block $s(q)$ in D^{HR}. Expressing s in the form of a pixel
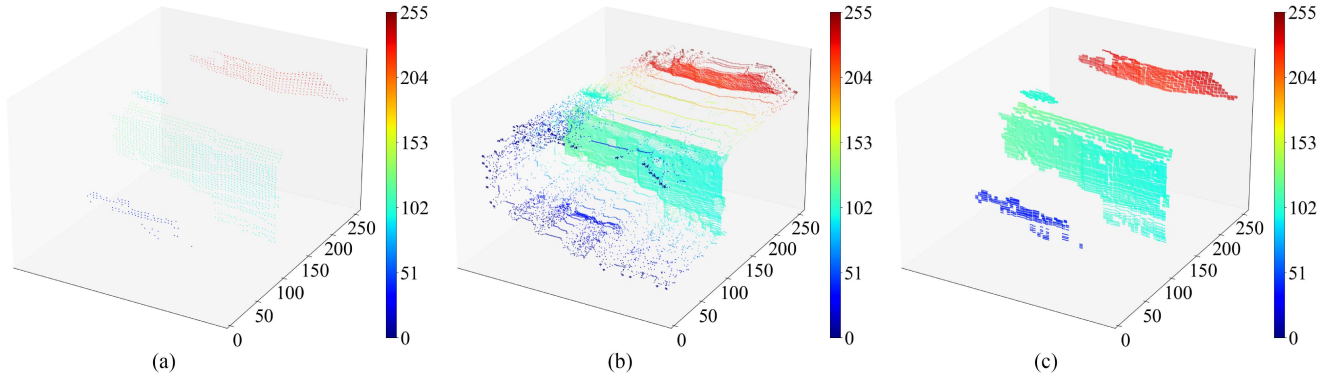
Fig. 5. (a) Original LR image. (b) Bicubic interpolation (up-sampling factor ×4). (c) Nearest neighbor (up-sampling factor ×4).
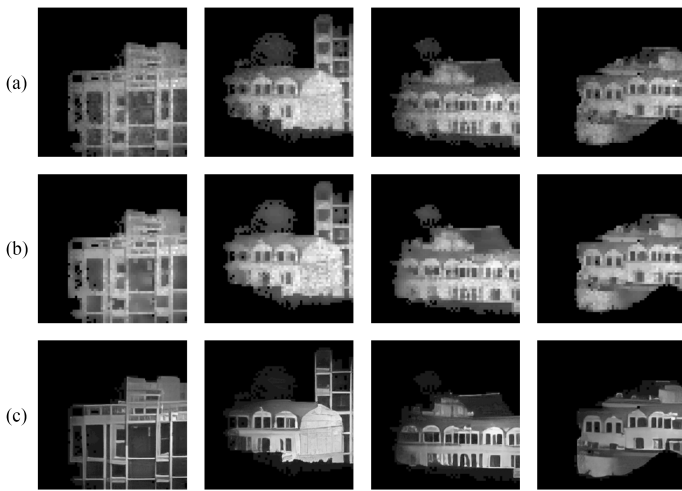

Fig. 6. (a) Different targets with speckle noise. (b) Qualitative results of the intensity map preprocessing. (c) Qualitative results of the pixel correction.

average as

$$s = \frac{1}{k^2} \sum_{p \in s(q)} D_p^{\text{HR}}. \qquad (4)$$

We expect to obtain an estimate D^{HR} of the HR Gm-APD LiDAR depth map, given D^{LR} and I^{HR}. In this case, the functional relationship between D^{HR} and I^{HR} is unknown, so we denote the function as

$$D^{\text{HR}} = f\left(I^{\text{HR}}; \theta\right). \qquad (5)$$

According to (4) and (5), the objective of the guided superresolution is as follows:

$$\hat{\theta} = \arg\min_{\theta} \mathcal{L}\left(f(I^{\text{HR}}; \theta)_{s(q)}, s\right) + \lambda\Phi(\theta) \qquad (6)$$

where $\mathcal{L}(f(I^{\text{HR}}; \theta)_{s(q)}, s)$ represents the loss function between the generated HR depth map D^{HR} and the ground truth LR depth map D^{LR}, $\Phi(\theta)$ is the regularization term, and λ is the tradeoff parameter.

We expect similar areas at different locations in I^{HR} to have as much trainability as possible, so we drawing on the vision

transformer model to split I^{HR} into multiple patches based on a learnable approach, and use absolute location encoding to characterize the location information of each patch

$$\hat{\theta} = \arg\min_{\theta} \mathcal{L}\left(f\left(I^{\text{HR}}, x, y; \theta\right)_{s(q)}, s\right) + \lambda\Phi(\theta). \qquad (7)$$

The data characteristics of the Gm-APD LiDAR should be considered when using the Gm-APD LiDAR intensity map with the corresponding depth map for guiding superresolution. Based on the results of accumulated detection to obtain target information discussed in Section II-B: compared to linear APD, which captures intensity information by the amplitude of the output pulse, Gm-APD LiDAR does not respond well to the subtle differences in the target intensity. Moreover, the intensity of a point of the target is based to some extent on the reflectivity of the target in the used band, so the pixel grayscale of the Gm-APD LiDAR intensity map obtained from a single-band pulse echo is extremely monotonous compared to the HR conventional camera image.

The monotonicity of the pixel grayscale obviously reduces the solvability of (7). As mentioned earlier: we expect similar areas at different locations in I^{HR} to have as much trainability as possible. And the pixel s in D^{LR} corresponds to $k \times k$ pixels block $s(q)$ in D^{HR}, so we use a parallel two-branches structure with the block location encoding of pixel as the second branch. The purpose of using parallel two-branches structure is to enhance the solvability of (7) and, thus, suppress the cluttered three-dimensional points in the obtained HR depth map.

To this end, the objective of acquiring HR Gm-APD LiDAR depth map is as follows:

$$\hat{\theta} = \arg\min_{\theta} \mathcal{L}\left(f\begin{pmatrix} F_{\text{ESRGAN}}\left(I^{\text{LR}}; \Theta\right), \\ x, y, x_{s(q)}, y_{s(q)}; \theta \\ D_s^{\text{LR}} \end{pmatrix}_{s(q)}\right) + \lambda\Phi(\theta)$$
$$(8)$$

where the loss function is chosen as MSE loss, and for the regularization term, L2-penalty of the network weights is used to combat overfitting.

As shown in Fig. 4, according to the discussed theoretical model: our network can be unfolded to two branches, the pixel branch implemented by a multilayer convolutional network
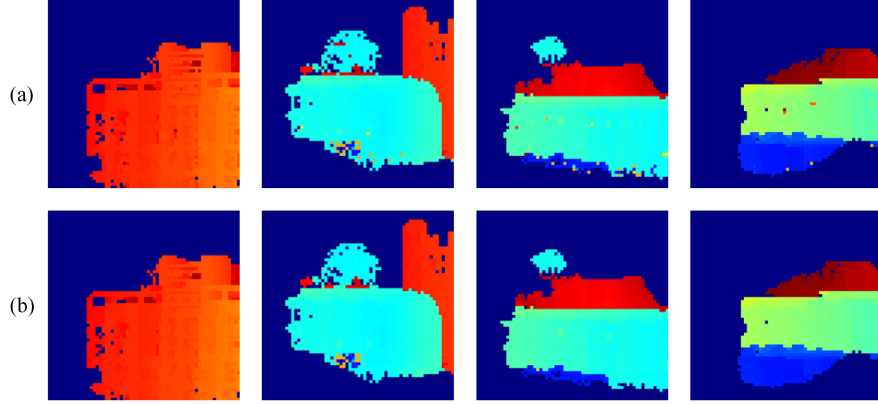
Fig. 7. (a) Different targets with abnormal noise. (b) Qualitative results of the k-nearest neighbor correction algorithm.

using 1×1 kernels and the block branch implemented by a multilayer convolutional network using 3×3 kernels. For the branch n, its input vectors and the feature extraction operation of the branch is defined as $\mathbf{I}^n$ and $\mathcal{F}_n(\cdot)$, respectively. $\mathbf{I}^n$ covers a block $\mathbf{S}(q)^n$ of $k \times k$ pixels. The output of the pixel branch and block branch is given by

$$
\begin{aligned}
\mathbf{O}^k_{\text{Pixel}} &= \mathcal{F}_{\text{Gn}}\left(\mathbf{I}^{\text{Gn}}\right) + \mathcal{F}_{\text{Xn}}\left(\mathbf{I}^{\text{Xn}}\right) \\
\mathbf{O}^k_{\text{Block}} &= \mathcal{F}_{\text{Nn}}\left(\mathbf{S}(q)^{\text{Nn}}\right) + \mathcal{F}_{\text{Sn}}\left(\mathbf{S}(q)^{\text{Sn}}\right).
\end{aligned}
\tag{9}
$$

Based on aforementioned definitions, the output of our network for any $\mathbf{I}^n$ and $\mathbf{S}(q)^n$ can be expressed as

$$
\mathbf{O}^k_{\text{Head}} = \mathcal{F}_{\text{Head}}\left(\mathcal{F}_{\text{concat}}\left(\mathbf{O}^k_{\text{Pixel}}, \mathbf{O}^k_{\text{Block}}\right)\right)
\tag{10}
$$

where the number of output feature maps for each convolution layer is given in Fig. 4, and the ESA block employed uses a dimensionality reduction factor of $\times 4$.

C. Filtering Algorithm

In the field of LiDAR depth imaging, filtering can be used to distinguish foreground and background [31]. As shown in Fig. 9(c), there are a few cluttered three-dimensional points in the targets, and these cluttered three-dimensional points are the bad pixels in depth map, which can interfere with the distance information of the targets. The bad pixels in depth map are the pixels that have excessive deviation from the ground truth. It is defined in the filtering algorithm of this article as the pixels that are beyond the distance value domain of the original depth map. As shown in (4), the depth value in the corresponding pixels block $s(q)$ fluctuates around s when the pixel s is expressed in the form of pixel average. When the obtained depth value is beyond the minimum or maximum value of the original depth map, the obtained depth value is not credible due to the lack of distance calibration. Therefore, referring to the filtering method of LiDAR data processing, we design a filtering algorithm to remove the bad pixels from the obtained HR depth map. We denote the filtering algorithm as

$$
p = \begin{cases} p, & p \in \left[\min\left(D^{\text{LR}}\right) - \alpha, \max\left(D^{\text{LR}}\right) + \beta\right] \\ r, & p \notin \left[\min\left(D^{\text{LR}}\right) - \alpha, \max\left(D^{\text{LR}}\right) + \beta\right] \end{cases}.
\tag{11}
$$

IV. EXPERIMENTS AND RESULTS

In this section, we describe the image quality assessment methods used and implement all the proposed methods using *Python* and *PyTorch*. In addition, we use the *Nvidia GeForce RTX 4090* for the calculations.

A. Evaluation Metrics

For LiDAR depth maps, we prefer that the more perfect quality image should contain more information about the target while preserving the ground truth of the target as much as possible. Therefore, when we use evaluation metrics to analyze the performance of the method, we focus more on the information content of the obtained HR Gm-APD LiDAR depth map and the ground truth of the target.

Given that the data used in this article are all collected in our outfield experiments, there is a lack of reference images when using evaluation metrics to analyze the performance of the method. However, blind image quality assessment obviously cannot meet our focused evaluation needs, so the following assessment methods were used for analysis: *PSNR*, mutual information (*MI*) [32], and percentage of bad pixels (*PBPs*) [33]. They were defined by (12) to (14)

$$
\text{PSNR}\left(I, \hat{I}\right) = 10 \cdot \log_{10}\left(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}\left(I(i) - \hat{I}(i)\right)^2}\right).
\tag{12}
$$

We consider that the original Gm-APD LiDAR depth map can represent the true distance to the target, so interpolation up-sampling of the original LR Gm-APD LiDAR depth map can be used as the ground truth image I. It is obvious that the visual perception of image I is completely different from the reconstructed image $\hat{I}$. However, *PSNR* only cares about the differences between corresponding pixels instead of visual perception [34], it allows the evaluation of the ground truth in evaluating the reconstruction quality of the Gm-APD LiDAR depth map. And interpolation is as discussed in Section III-A: the weighted calculation cannot be used in the preprocessing of Gm-APD LiDAR depth map to obtain a distance value, only a distance
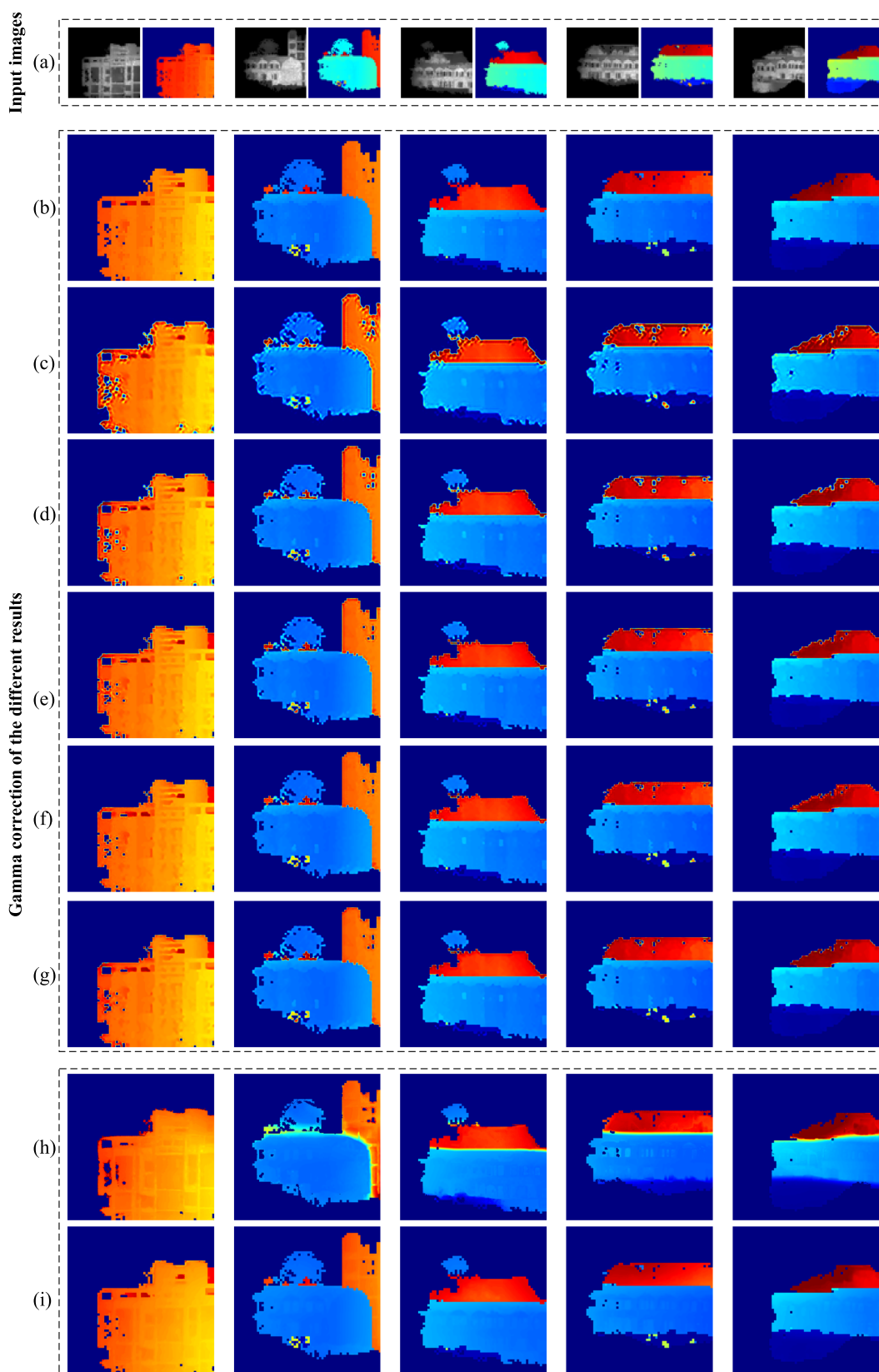
Fig. 8. Qualitative results of our network (up-sampling factor ×4). (a) LR intensity maps with depth maps for network input. (b) Nearest neighbor. (c) DJIF [21]. (d) MSGNet [9]. (e) FEAG [22]. (f) FDSR [24]. (g) GDSR-DCTNet [23]. (h) Remove the network outputs of the block branch. (i) Retaining the network outputs of the block branch.
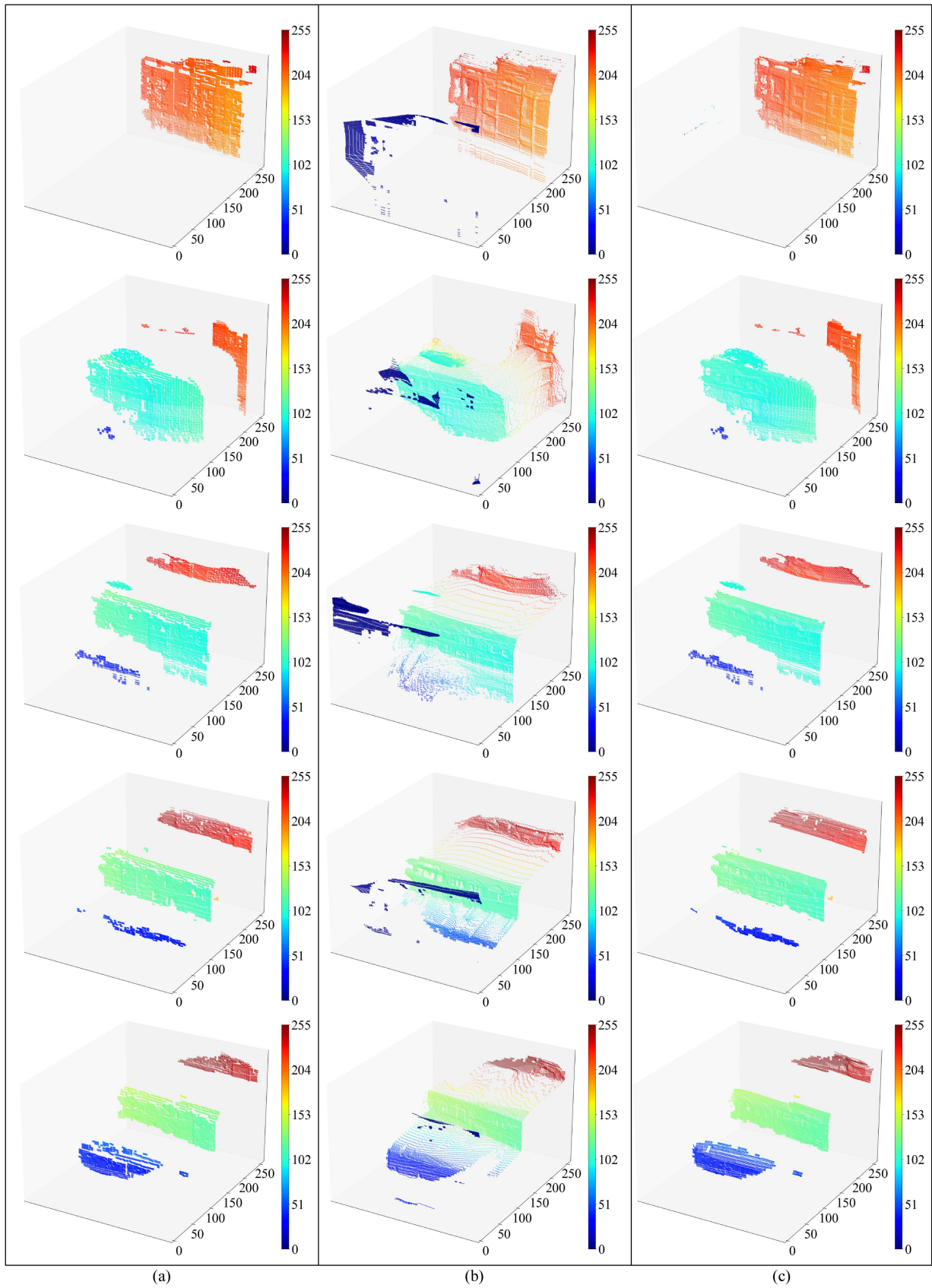
Fig. 9. Three-dimensional scatter of the qualitative results. (a) Nearest neighbor. (b) Remove the network outputs of the block branch. (c) Retaining the network outputs of the block branch.

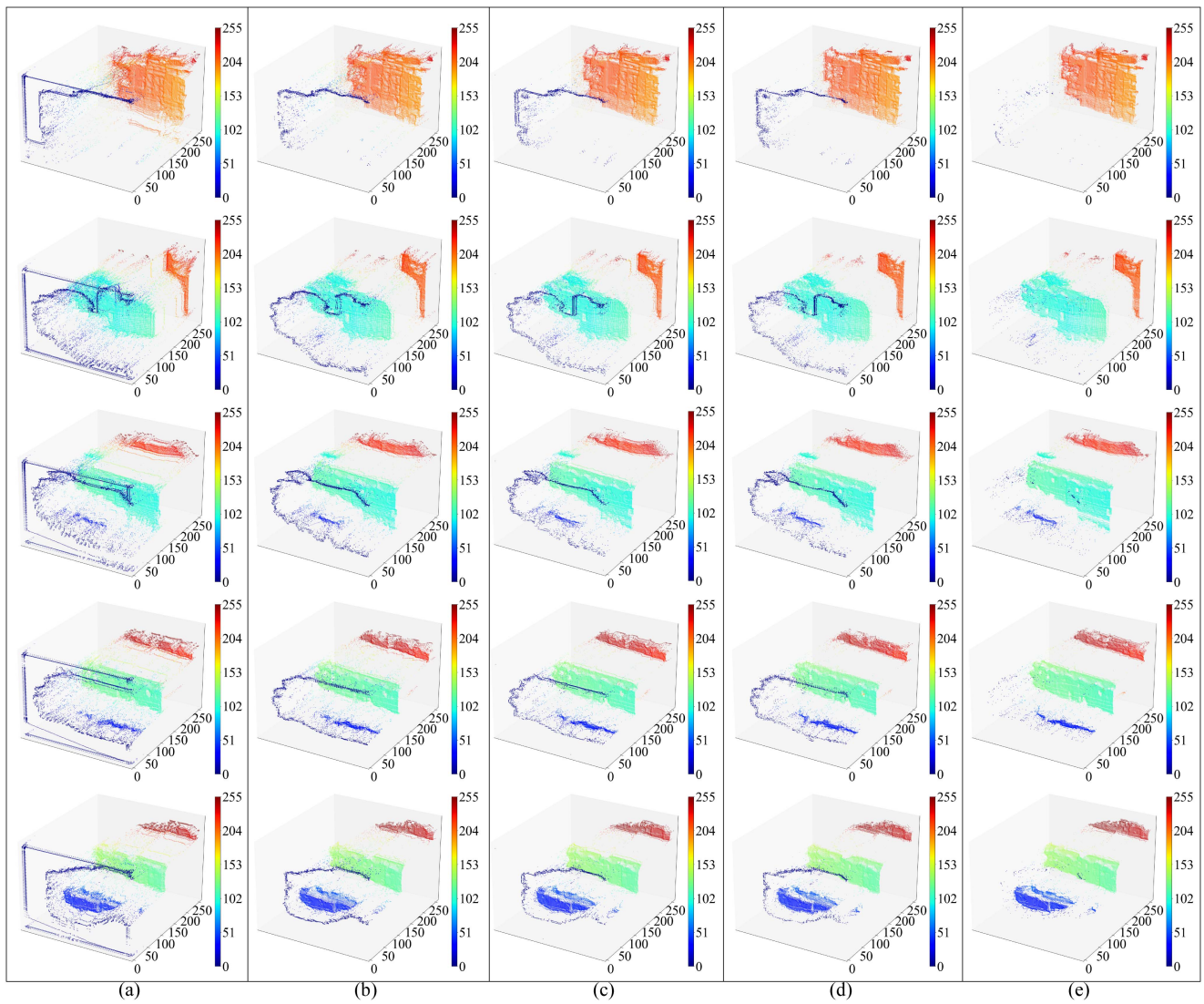Fig. 10. Three-dimensional scatter of the qualitative results. (a) DJIF [21]. (b) MSGNet [9]. (c) FEAG [22]. (d) FDSR [24]. (e) GDSR-DCTNet [23].
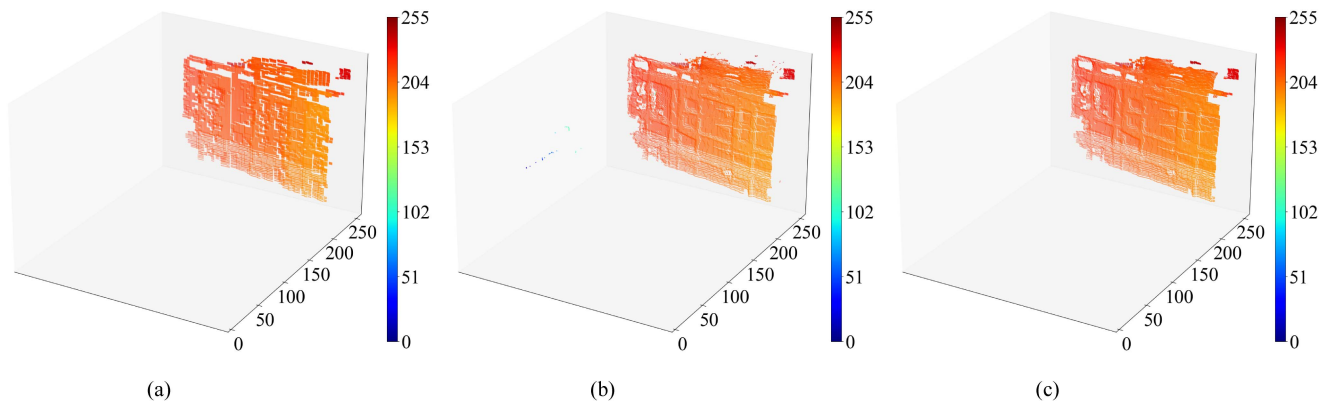

Fig. 11. (a) Nearest neighbor. (b) Target with cluttered three-dimensional points. (c) Qualitative results of the filtering algorithm.

TABLE I
COMPARISON OF EVALUATION METRICS RESULTS AT ×4 UP-SAMPLING FACTOR

×4 (*PSNR*)	Target 1	Target 2	Target 3	Target 4	Target 5	Mean
DJIF [21]	24.94	25.45	26.65	26.34	28.04	26.29
MSGNet [9]	27.86	28.57	30.01	28.41	30.25	29.02
FEAG [22]	32.98	30.75	33.54	31.60	34.45	32.67
FDSR [24]	**36.48**	32.37	35.00	32.91	35.61	34.48
GDSR-DCTNet [23]	34.34	31.33	33.27	32.50	35.6	33.41
Removing block branch	25.09	24.27	28.43	24.97	28.71	26.30
Retaining block branch	35.99	**43.87**	**44.51**	**43.12**	**43.56**	**42.21**

Note: All images of the *PSNR*.
The best values are highlighted by bold.

TABLE II
COMPARISON OF EVALUATION METRICS RESULTS AT ×4 UP-SAMPLING FACTOR

×4 (*MI*)	Target 1	Target 2	Target 3	Target 4	Target 5	Mean
Nearest Neighbor	0.841	0.846	0.964	0.942	0.893	0.897
DJIF [21]	0.598	0.689	0.584	0.654	0.867	0.679
MSGNet [9]	0.591	0.676	0.579	0.648	0.878	0.674
FEAG [22]	0.586	0.667	0.574	0.645	0.883	0.671
FDSR [24]	0.592	0.674	0.577	0.649	0.876	0.674
GDSR-DCTNet [23]	0.586	0.666	0.572	0.647	0.891	0.672
Removing block branch	0.616	0.677	0.598	0.663	0.880	0.687
Retaining block branch	**0.876**	**0.890**	**1.009**	**0.988**	**0.908**	**0.934**

Note: All images of the *MI*.
The best values are highlighted by bold.

TABLE III
COMPARISON OF EVALUATION METRICS RESULTS AT ×4 UP-SAMPLING FACTOR

×4 (*PBP*)	Target 1	Target 2	Target 3	Target 4	Target 5	Mean
DJIF [21]	36.9	53.1	38.5	41.8	26.7	39.4
MSGNet [9]	23.9	27.2	20.5	21.5	12.9	21.2
FEAG [22]	18.1	24.1	16.8	16.1	9.5	16.9
FDSR [24]	**8.7**	14.5	10.0	11.1	8.0	10.5
GDSR-DCTNet [23]	10.9	16.1	10.9	10.0	**5.3**	10.6
Removing block branch	26.5	32.7	18.5	34.5	21.1	26.7
Retaining block branch	10.4	**5.3**	**3.6**	**6.3**	6.9	**6.5**

Note: All images of the *PBP* (in %).
The best values are highlighted by bold.

value can be selected from the original depth map, so as shown in Fig. 5(c), we propose to use nearest neighbor to describe the up-sampling progress.

Considering that the more perfect quality image should contain more information about the target, we use *MI* to evaluate the information richness of the target in obtained HR LiDAR depth map. For the calculation of *MI*, we consider the processed Gm-APD LiDAR intensity map with the corresponding depth map for guided superresolution. The target information in intensity map is fused into depth map, which increases the target information content of depth map. Therefore, we use the interpolation up-sampling depth map as source image *A* and the processed intensity map as source image *B*

$$MI_{K,F} = \sum_{k,f} P(k,f) \log_2 \frac{P(k,f)}{P(k)P(f)}$$

$$MI_{A,B,F} = MI_{A,F} + MI_{B,F}. \tag{13}$$

PBP is defined as follows:

$$\text{PBP}\left(I, \hat{I}\right) = \frac{1}{N} \sum_i \left[\left| I(i) - \hat{I}(i) \right| > \delta \right] \tag{14}$$

where δ is a depth error tolerance, we take $\delta = 5$ pixel.

B. Results of Data Preprocessing

Fig. 6 shows the qualitative results of the pixel correction of the HR Gm-APD LiDAR intensity map recovered by the superresolution network. As shown in Fig. 6(c), we have ensured that the recovered HR intensity map still maintains the pixel correspondence with the depth map.

Fig. 7 shows the qualitative results of the proposed *k*-nearest neighbor correction algorithm on the data collected in outfield experiments. As shown in Fig. 7(a), there is some abnormal noise in the extracted Gm-APD LiDAR depth map. Obviously, the proposed method can suppress the abnormal noise in depth map.

C. Results of Guided Superresolution Module

Fig. 8 shows the qualitative results of our network on the data collected in outfield experiments, and we compare the results of ablation experiments with retaining versus removing the block branch. We evaluate the performance of Gm-APD LiDAR single-source data self-guided method on different urban high-rise building targets. In this case, the network input Gm-APD LiDAR intensity maps with the corresponding depth maps are both 64×64 pixels, and the network output depth maps are 256×256 pixels. To better illustrate the differences in visual perception of the different results, we use the gamma correction with a gamma factor of 1.5 to process the depth maps obtained by the different methods. It is obvious that the proposed method can obtain more visually perceptional results and retaining the block branch can better retain the original distance information of the targets. As shown in Fig. 8(i), the targets structure and detail information in the obtained HR Gm-APD LiDAR depth maps can be well recovered. Meanwhile, the results of the ablation experiments demonstrate the effectiveness of retaining the block branch.

DSR on RGB-D data has the corresponding HR ground truth as reference images, thus allowing the accurate perceptual assessment by color comparison with the ground truth. However, the data collected in outfield experiments do not have any reference images, so as shown in Fig. 9, we illustrate the three-dimensional scatter corresponding to Fig. 8(b), (h), and (i). Displaying the depth map as three-dimensional scatter can show the three-dimensional distance information of the target more clearly. As shown in Fig. 9(b), there are cluttered three-dimensional points in the targets, which are close to the bicubic result shown in Fig. 5(b). Obviously, these cluttered three-dimensional points do not serve as a reference for the distance information of the targets, because the depth maps collected by LiDAR are generated from real-world three-dimensional points, and the real distance information of the targets should be close to the nearest neighbor results shown in Fig. 9(a). Therefore, as shown in Fig. 9(c), the HR depth maps obtained by retaining the block branch can better retain the original distance information of the target.

To better illustrate the differences in the results of different methods, we show the corresponding three-dimensional scatter from Fig. 8(c)–(g), as shown in Fig. 10. In the three-dimensional scatter, the three-dimensional distance information of the target can be visualized.

According to the evaluation metrics discussed in Section IV-A, the quantitative results of evaluation metrics for the $\times 4$ up-sampling factor are shown in Tables I to III. As shown in Tables I–III, the proposed Gm-APD LiDAR single-source data self-guided method can obtain HR depth maps with richer information content, which is consistent with the visual perception shown in Figs. 8(i) and 9(c). And results of retaining the block branch obtain better *PSNR* results, i.e., the original distance information of the targets can be better retained.

D. Results of Filtering Algorithm

The results of our proposed filtering algorithm are shown in Fig. 11(c). In practical applications, we take $\alpha = \beta = 0$ to remove the bad pixels that are beyond the distance value domain of the original LiDAR depth map. Due to the pixels that exceed the distance value domain of the original depth map lack the constraint of the distance value, we believe that the distance value of the obtained HR depth map should be within the range of the distance value domain.

V. CONCLUSION

In this article, we propose the Gm-APD LiDAR single-source data self-guided method using the data characteristics of Gm-APD LiDAR. The advantage of this method is that it relies only on the single-source data of Gm-APD LiDAR to obtain HR intensity map and depth map simultaneously, and thus has wider applicability.

A parallel two-branches structure is used to obtain the HR depth map. Retaining the block branch improves the average *PSNR* of the initial results from 26.30 to 42.21, which better retains the original distance information of the targets. The qualitative results on the data collected in outfield experiments demonstrate that the proposed Gm-APD LiDAR single-source data processing method is feasible, which is significant for promoting the application of Gm-APD LiDAR in remote sensing.

Since the advantages of LiDAR can be fully exploited by using only single-source data, we investigated the feasibility of this processing method. Due to the limitation of signal processing results of Gm-APD LiDAR (limited target information of network input), our detection targets are mainly buildings, which has some limitations. Consequently, in our future work, we will focus on combining the signal processing of Gm-APD LiDAR with the image processing network to improve the quality of network input images, and use the Gm-APD LiDAR single-source data processing method on targets, such as cars, at longer distances.

REFERENCES

[1] A. Kirmani et al., "First-photon imaging," *Science*, vol. 343, no. 6166, pp. 58–61, Jan. 2014.

[2] Z. H. Li et al., "Multi-beam single-photon-counting three-dimensional imaging LiDAR," *Opt. Express*, vol. 25, no. 9, pp. 10189–10195, May 2017.

[3] R. Tobin et al., "Three-dimensional single-photon imaging through obscurants," *Opt. Express*, vol. 27, no. 4, pp. 4590–4611, Feb. 2019.

[4] L. Liu et al., "Efficient global 2D-3D matching for camera localization in a large-scale 3D map," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2391–2400.

[5] X. Song et al., "Edge-guided depth map enhancement," in *Proc. Int. Conf. Pattern Recognit.*, 2016, pp. 2758–2763.

[6] X. B. Song et al., "ApolloCar3D: A large 3D car instance understanding benchmark for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5447–5457.

[7] X. B. Song et al., "Modeling deviations of RGB-D cameras for accurate depth map and color image registration," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 14951–14977, Jun. 2018.

[8] X. B. Song et al., "Estimation of Kinect depth confidence through self-training," *Vis. Comput.*, vol. 30, no. 6–8, pp. 855–865, Jun. 2014.

[9] T. W. Hui, C. C. Loy, and X. O. Tang, "Depth map super-resolution by deep multi-scale guidance," *Lecture Notes Comput. Sci.*, vol. 9907, pp. 353–369, 2016.

[10] G. Riegler, M. Ruther, and H. Bischof, "ATGV-Net: Accurate depth super-resolution," *Lecture Notes Comput. Sci.*, vol. 9907, pp. 268–284, 2016.

[11] X. B. Song, Y. C. Dai, and X. Y. Qin, "Deep depth super-resolution: Learning depth super-resolution using deep convolutional neural network," *Lecture Notes Comput. Sci.*, vol. 10114, pp. 360–376, 2017.

[12] X. Song, Y. Dai, and X. Qin, "Deeply supervised depth map super-resolution as novel view synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2323–2336, Aug. 2019.

[13] O. Voynov et al., "Perceptual deep depth super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5652–5662.

[14] X. B. Song et al., "Channel attention based iterative residual learning for depth map super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5630–5639.

[15] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, Oct. 1976.

[16] M. Okutomi and T. Kanade, "A multiple-base-line stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 353–363, Apr. 1993.

[17] C. Dong et al., "Learning a deep convolutional network for image super-resolution," *Lecture Notes Comput. Sci.*, vol. 8692, pp. 184–199, 2014.

[18] X. C. Hu et al., "Meta-SR: A magnification-arbitrary network for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1575–1584.

[19] Z. Li et al., "Feedback network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 3862–3871.

[20] Z. S. Liu et al., "Image super-resolution via attention based back projection networks," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 3517–3525.

[21] Y. J. Li et al., "Deep joint image filtering," *Lecture Notes Comput. Sci.*, vol. 9908, pp. 154–169, 2016.

[22] J. Y. Yang, H. Lan, X. Song, and K. Li, "Depth super-resolution via fully edge-augmented guidance," in *Proc. IEEE Vis. Commun. Image Process.*, 2017, pp. 3865–3868.

[23] Z. X. Zhao et al., "Discrete cosine transform network for guided depth map super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 9227–9236.

[24] L. Z. He et al., "Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9225–9234.

[25] F. Wang et al., "Range accuracy limitation of pulse ranging systems based on Geiger mode single-photon detectors," *Appl. Opt.*, vol. 49, no. 29, pp. 5561–5566, Oct. 2010.

[26] Y. B. Zhang et al., "Dual-parameter estimation algorithm for Gm-APD LiDAR depth imaging through smoke," *Measurement*, vol. 196, Jun. 2022, Art. no. 111269.

[27] X. Zhang et al., "Research on the detection probability curve characteristics of long-range target based on SPAD array LiDAR," *Infrared Phys. Technol.*, vol. 126, Nov. 2022, Art. no. 104325.

[28] J. Wu et al., "64 × 64 GM-APD array-based readout integrated circuit for 3D imaging applications," *Sci. China-Inf. Sci.*, vol. 62, no. 6, pp. 1–11, Jun. 2019.

[29] L. Xu et al., "Signal restoration method for restraining the range walk error of Geiger-mode avalanche photodiode LiDAR in acquiring a merged three-dimensional image," *Appl. Opt.*, vol. 56, no. 11, pp. 3059–3063, Apr. 2017.

[30] X. T. Wang et al., "ESRGAN: Enhanced super-resolution generative adversarial networks," *Lecture Notes Comput. Sci.*, vol. 11133, pp. 63–79, 2019.

[31] Y. J. Yan et al., "Extracting suburban residential building zone from airborne streak tube imaging LiDAR data," *Measurement*, vol. 199, Aug. 2022, Art. no. 111488.

[32] G. H. Qu, D. L. Zhang, and P. F. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.

[33] B. Ham, M. Cho, and J. Ponce, "Robust guided image filtering using nonconvex potentials," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 192–207, Jan. 2018.

[34] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

Boyi Song received the B.E. degree in photoelectric information science and engineering and the M.S. degree in optical engineering from Harbin Engineering University, Harbin, China, in 2017 and 2019, respectively. He is currently working toward the Ph.D. degree in electronic science and technology with the National Key Laboratory of Science and Technology on Tunable Laser, Harbin Institute of Technology.

His research interests include LiDAR data processing, deep learning, and digital image processing.

Jianfeng Sun (Member, IEEE) was born in Heilongjiang, China, in August 1978. He received the Ph.D. degree in physical electronics from the Harbin Institute of Technology, Harbin, China, in 2007.

He has been a Professor with the Harbin Institute of Technology since 2021 and the Vice Chairman of the IEEE EDS Harbin branch. He has authored or coauthored more than 40 SCI and EI academic articles, authorized five national invention patents, and applied for more than 20 patents. His main research interests include laser imaging radar and image information processing technology.

Yongji Yan received the B.E. degree in photoelectric information science and engineering from the North University of China, Taiyuan, China, in 2017, and the M.S. degree in optical engineering from Harbin Engineering University, Harbin, China, in 2019. He is currently working toward the Ph.D. degree in electronic science and Technology with the National Key Laboratory of Science and Technology on Tunable Laser, Harbin Institute of Technology.

His research interests include LiDAR data processing, deep learning, and digital image processing.

Rui He was born in 1992. He is currently working toward the Ph.D. degree in electronic science and technology with the Harbin Institute of Technology, Harbin, China.

His main research interests include single-photon LiDAR imaging and tracking technology.

Xin Zhou was born in 1990. He received the Ph.D. in physical electronics from the Harbin Institute of Technology, Harbin, China, in 2021.

He is currently a Postdoctoral Fellow with the Harbin Institute of Technology. His main research interests include the overall design of single-photon LiDAR, the scattering characteristics of single-photon targets, the improvement of detection performance of single-photon LiDAR, and the signal processing and image reconstruction of single-photon LiDAR.

Wei Lu was born in January 1967. He received the Ph.D. degree in physical electronics from the Harbin Institute of Technology, Harbin, China, in 2003.

He is currently an Associate Professor with the Department of Opto-Electronic Information Science and Technology, mainly engaged in research on LiDAR technology.