remote sensing

MDPI

Article

Depth Information Precise Completion-GAN: A Precisely Guided Method for Completing Ill Regions in Depth Maps

Ren Qian [1], Wenfeng Qiu [1], Wenbang Yang [1], Jianhua Li [1], Yun Wu [1], Renyang Feng [2], Xinan Wang [3,*] and Yong Zhao [1,3]

[1] College of Computer Science and Technology, Guizhou University, Guiyang 550025, China; qrenqr@163.com (R.Q.); summit@gdmu.edu.cn (W.Q.); cse.wbyang20@gzu.edu.cn (W.Y.); huarzail@163.com (J.L.); wuyun_v@126.com (Y.W.); yongzhao@pkusz.edu.cn (Y.Z.)
[2] School of Information, Guizhou University of Finance and Economics, Guiyang 550031, China; renyan_feng@163.com
[3] School of Electronic and Computer Engineering, Shenzhen Graduate School of Peking University, Shenzhen 518055, China
* Correspondence: anxinwang@pku.edu.cn

Abstract: In the depth map obtained through binocular stereo matching, there are many ill regions due to reasons such as lighting or occlusion. These ill regions cannot be accurately obtained due to the lack of information required for matching. Since the completion model based on Gan generates random results, it cannot accurately complete the depth map. Therefore, it is necessary to accurately complete the depth map according to reality. To address this issue, this paper proposes a depth information precise completion GAN (DIPC-GAN) that effectively uses the Guid layer normalization (GuidLN) module to guide the model for precise completion by utilizing depth edges. GuidLN flexibly adjusts the weights of the guiding conditions based on intermediate results, allowing modules to accurately and effectively incorporate the guiding information. The model employs multiscale discriminators to discriminate results of different resolutions at different generator stages, enhancing the generator's grasp of overall image and detail information. Additionally, this paper proposes Attention-ResBlock, which enables all ResBlocks in each task module of the GAN-based multitask model to focus on their own task by sharing a mask. Even when the ill regions are large, the model can effectively complement the missing details in these regions. Additionally, the multiscale discriminator in the model enhances the generator's robustness. Finally, the proposed task-specific residual module can effectively focus different subnetworks of a multitask model on their respective tasks. The model has shown good repair results on datasets, including artificial, real, and remote sensing images. The final experimental results showed that the model's REL and RMSE decreased by 9.3% and 9.7%, respectively, compared to RDFGan.

Keywords: ill regions; stereo matching; GAN; depth map repair

Citation: Qian, R.; Qiu, W.; Yang, W.; Li, J.; Wu, Y.; Feng, R.; Wang, X.; Zhao, Y. Depth Information Precise Completion-GAN: A Precisely Guided Method for Completing Ill Regions in Depth Maps. *Remote Sens.* **2023**, *15*, 3686. https://doi.org/10.3390/rs15143686

Academic Editors: Javaan Chahl, Huajian Liu, Asanka Perera and Ali Al-Naji

Received: 18 May 2023
Revised: 14 July 2023
Accepted: 19 July 2023
Published: 24 July 2023

1. Introduction

In recent years, there has been significant improvement in the accuracy of 3D matching algorithms in the field of 3D matching. However, there are still issues with the depth values of ill regions. These ill regions can be categorized into occluded regions, reflective regions, and low-texture regions.

Occluded regions occur when certain regions in a scene cannot be simultaneously captured by all cameras in a multiview 3D matching algorithm. This results in a lack of corresponding information and prevents the accurate generation of stereoscopic vision using the 3D matching algorithm. Reflective regions arise due to lighting conditions that destroy information in a specific region, making it difficult to find matching information for that area. The lack of available corresponding information hinders the accurate matching process. Low-texture regions occur when there is insufficient information available for

successful matching. These regions lack the necessary features or distinctive patterns required for effective matching. All these regions, which suffer from a lack of matching information due to various reasons, are collectively referred to as ill regions.

Repairing ill regions in the depth map can significantly enhance the accuracy of the overall depth map. Currently, the accuracy of 3D matching algorithms that match incorrect depth values is mostly concentrated in ill regions. Consequently, the overall accuracy of the depth map obtained by the algorithm depends on the accuracy of ill regions. Therefore, addressing and rectifying these ill regions is crucial for improving the accuracy of the depth map in 3D matching algorithms.

Enhancing the accuracy of 3D matching algorithms is particularly essential in the industry. The inadequacy of 3D matching accuracy has previously resulted in serious accidents, as exemplified by the Tesla Autopilot accident in 2016, which caused the driver's death. An investigation revealed that the 3D matching algorithm failed to identify the truck correctly, leading to incorrect decisions by the autonomous driving system. Hence, in industrial applications, the accuracy of 3D matching algorithms holds tremendous importance.

Currently, there are many problems with existing depth map completion algorithms for completing large ill regions. The main problem is that the repaired ill regions lack detailed information, especially when the ill region is too large, resulting in incorrect or random results. Therefore, it is necessary to use specific guidance information when repairing ill regions to avoid repairing errors caused by the lack of detailed information in the ill region.

Satellite depth maps have been widely used in fields such as Earth observation, natural disaster warning, and urban planning. Although researchers have proposed a series of depth estimation algorithms, such as stereoscopic matching, optical flow, and deep learning-based methods, and made progress, there are still serious ill regions in the depth maps obtained from satellite images, which limit the accuracy and precision of depth estimation. These ill regions mainly refer to areas such as rivers and lakes, where the different reflection strengths and angles of light due to the different satellite photographing positions make it impossible to accurately match the corresponding images. In addition, buildings, hills, and trees in the images also generate occlusion areas due to the different photographing positions and angles. Although there are already some models available for processing these problem areas and improving the accuracy and precision of depth estimation, the results are not satisfactory. We proposed depthFillGan, a model that repairs ill regions in depth maps by using depth edges as guidance information. This model can effectively utilize the guidance information to fill in the lost details in the ill regions, thus improving the reliability of the completion of depth maps.

There are many problems with existing depth map completion algorithms for completing large ill regions. The main problem is that the repaired ill regions need more detailed information, especially when the ill region is too large, resulting in incorrect or random results. Therefore, it is necessary to use specific guidance information when repairing ill regions to avoid repairing errors caused by the lack of detailed information in the ill region.

Satellite depth maps have been widely used in fields such as Earth observation, natural disaster warning, and urban planning. Although researchers have proposed a series of depth estimation algorithms, such as stereoscopic matching, optical flow, and deep learning-based methods, and made progress, there are still serious ill regions in the depth maps obtained from satellite images, which limit the accuracy and precision of depth estimation. These ill regions mainly refer to areas such as rivers and lakes, where the different reflection strengths and angles of light due to the different satellite photographing positions make it impossible to match the corresponding images accurately. In addition, buildings, hills, and trees in the images generate occlusion areas due to the different photographing positions and angles. Although some models are already available for processing these problem areas and improving the accuracy and precision of depth estimation, the results could be more satisfactory. We proposed depthFillGan, a model that repairs ill regions in depth maps

using depth edges as guidance information. This model can effectively utilize the guidance information to fill in the lost details in the ill regions, thus improving the reliability of the completion of depth maps.

Existing depth map completion algorithms face numerous challenges when completing large ill regions. The main issue stems from the fact that the reconstruction of these ill regions necessitates more detailed information, particularly when dealing with large-scale ill regions. That often leads to incorrect or random outcomes. To circumvent this problem, it becomes imperative to utilize specific guidance information during the repair process to avoid errors resulting from the lack of detailed information within the ill region.

Satellite depth maps have found widespread application in various fields such as Earth observation, natural disaster warning, and urban planning. Despite the introduction of various depth estimation algorithms, including stereoscopic matching, optical flow, and deep learning-based methods, significant ill regions remain within the depth maps obtained from satellite imagery. Consequently, the accuracy of depth estimation are limited. These ill regions primarily manifest in areas like rivers and lakes, where the varying reflection strengths and angles of light in satellite images impede accurate image matching. Furthermore, buildings, hills, and trees in these images create occlusion areas due to the differing photographing positions and angles. While certain models have been proposed to address these problematic areas and enhance depth estimation accuracy and precision, the results have been suboptimal.

We propose a model called depth information precise completion GAN (DIPC-GAN), which repairs ill regions in depth maps by utilizing depth edges as guidance information. This model effectively leverages this guidance information to fill in the missing details within the ill regions, thereby improving the reliability of depth map completion. Figure 1 demonstrates the model's efficacy.

Figure 1 illustrates the limitations of RDFGAN in effectively utilizing guidance information to repair diseased areas within reflective regions like rivers and lakes. On the contrary, our model successfully utilizes the provided guidance information to repair reflective regions effectively. Similarly, RDFGAN encounters difficulties when attempting to rectify occlusion areas in regions obstructed by buildings due to its limited ability to leverage guidance information efficiently. As a result, the repairs in these areas are less accurate, as the model struggles to determine whether it should repair the occlusion area based on the depth information of the roof or the ground. In contrast, our model excels in precisely repairing occlusion areas in such regions.

This article makes the following contributions:

- By using a GAN network for the precise correction of deep images, this article proposes a multiscale discriminator GAN network for repairing stereoscopic images. The multiscale discriminator can distinguish between different scales of results, allowing the generator to learn overall features more effectively and strengthening the generator's learning of details at different scales.
- We propose the GuidLN module, which effectively introduces guidance information to guide the generator. To accurately repair large ill regions, we design the GuidLN to effectively utilize deep edges as guidance information to enable the generator to perform precise repair on ill regions.
- We propose an attention module that effectively enables the generator and discriminator subnetworks of the GAN model to adapt better to their respective tasks. Specifically, in the multiscale discriminator network proposed in this article, the attention module can enhance the performance of different discriminators by enabling them to focus on the most relevant features for their specific task.

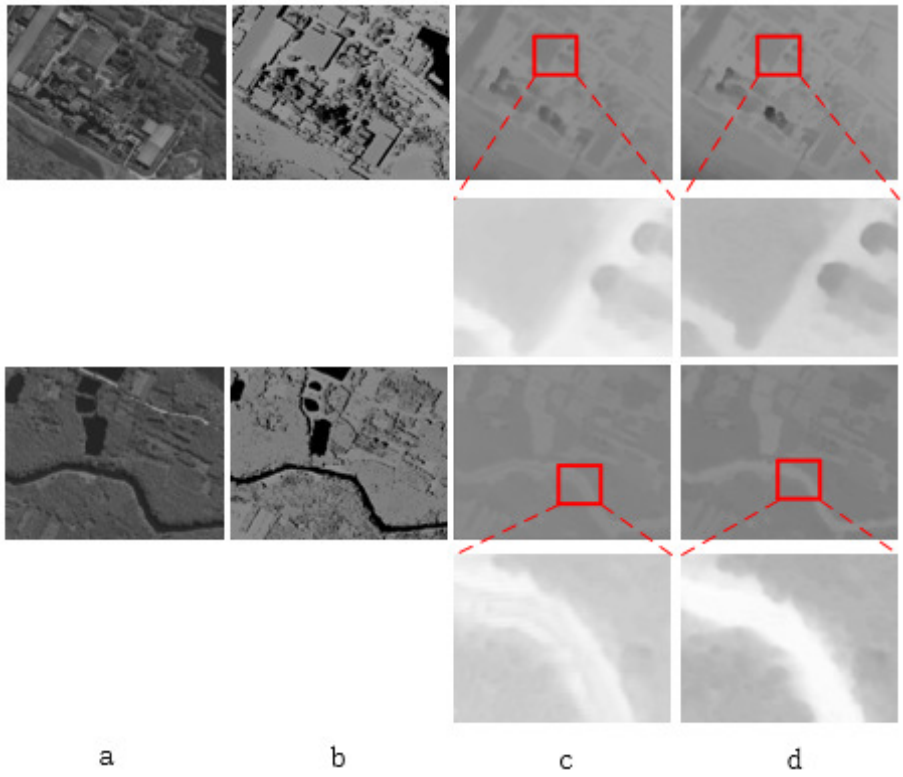

Figure 1. In the figure, (**a**) is the satellite reference image, and (**b**) is the depth map of the reference image, which contains a large number of ill regions, including reflective rivers, lakes, areas blocked by buildings and trees, etc., which are marked as black. (**c**) is the depth map of the repaired image by the RDFGAN model, and (**d**) is the depth map of the repaired image by our model. It can be seen that both the reflective rivers and lakes and the occlusion areas caused by buildings and trees have been well repaired.

2. Related Work

The current research focus on depth map completion technology is mainly on completing sparse depth maps into dense depth maps. Dense depth maps can be obtained from sparse images or point clouds obtained by a lidar.

Traditional methods use matrix interpolation to complete depth maps. In the literature, [1] uses semantic segmentation relationships to interpolate the matrix, while [2] uses constraints on the edges of sparse and dense depth maps to complete the interpolation and restoration. The authors of [3] use RGB images to restore sparse depth maps, which have rich information. Although this is too redundant for depth map restoration work, it is more effective compared to previous methods.

Although there are many image completion algorithms, there are still few algorithms for depth map completion. In non-deep learning methods, [4] selects noncovered points near the covered points to replace the covered points. The authors of [5] use a video's previous and future frames to complete the depth map. The authors of [6] improved a random forest model, which iteratively inputs RGB images and depth maps until the ill regions disappear. The authors of [1] use the relationship between semantic segmentation to interpolate the matrix, and the authors of [3] use RGB image information to repair sparse depth maps.

With the continuous development of neural networks and deep learning, image repair and completion using deep learning techniques is also a hot research topic in image restoration. However, research on depth map completion is still insufficient, such as [7], where the authors use RGB images to guide the completion of sparse visual difference maps into dense visual difference maps. The authors of [8] use traditional image processing methods to fill the sparse depth maps in the KITTI dataset.

The authors of [9] use CNN to process sparse differential images by designing a mask image to mark the sparse differential images and generate a dense image by processing the differential and mask images. The authors of [10] improve this method by modifying it, resulting in better results. The authors of [11] generate a confidence map to mark the confidence of each pixel's differential value and then generate a dense differential image through propagation. The authors of [12] obtain better dense differential images by iterating and repairing sparse differential images to dense differential images continuously. The authors of [13] extract features from RGB images and use an encoding network to extract features from sparse differential images. The features of RGB images are integrated into feature maps during the decoding process, and three different scales of encoding networks are used to fill the sparse images. The authors of [14] use normal vectors to fill sparse differential images. The authors of [15] input RGB images and sparse differential images, generate a confidence map, and use nonlocal propagation for spatial propagation to fill the differential image. Although these methods have been studied extensively, they mainly focus on completing sparse depth maps produced by devices such as lidar [16]. They need to provide deep completion for large areas.

The authors of [17]. use RGB images as input to predict object surfaces' dense polygons and occlusion information. These predictions are then combined with the original depth map through global optimization to address missing pixel issues in the original image. However, the algorithm takes a long time to execute. The authors of [18] input RGB images and depth maps into a multiscale network for learning and obtaining predicted depth maps, which are used for depth map completion. These methods directly utilize RGB images to repair depth maps introduce redundancy, resulting in less ideal repair results. The authors of [19] categorize the repair areas into 12 types and use semantic segmentation images to guide the filling of the repair areas. However, semantic segmentation images still contain redundancy because the depth of the same target in the semantic segmentation image may not be the same. The authors of [20] super-resolve low-resolution depth maps to obtain high-resolution depth maps. Then, a texture edge extraction network is used to obtain texture edges in the RGB image. Finally, the initial depth map is optimized to obtain a repaired depth map. This method uses texture edges for depth map optimization, but texture edges are not effective in guiding depth map repair, because there is still excessive redundancy.

Goodfellow [21] proposed the generator–discriminator network (GAN) in 2014, which has since been applied to numerous scenarios in digital image generation, such as image super-resolution, image editing, and so on. In particular, GAN has been applied to image restoration. The authors of [22] train a general restoration network by using a GAN to generate training data. However, the network is not used for ill region restoration. The authors of [23] propose a style transformation method for generating complete depth maps. The authors of [24] design a network with two branches: the first branch uses an encoding–decoding approach to convert sparse or incomplete depth maps into complete depth maps, while the other branch uses a GAN network to perform depth map style transformation on RGB images to generate depth maps for restoration. The authors of [25] use domain adaptation methods to design and train networks, generating geometric information or noise on synthetic datasets to mimic real datasets. The GAN network generates RGB images more consistent with real-world scenarios to assist network training. Since the accuracy of the depth maps generated through this style transformation is low, the performance of depth map completion could be improved.

Although GAN-based models can generate images that conform to the data distribution, which is acceptable for general generative tasks, there is an increasing demand for tasks that require more than random generation. Hence, there is a need to properly guide the models for completion. Many completion models currently use conditional information to guide the generation of desired results [26]. There are primarily two methods to incorporate guiding information into models. The first approach involves introducing the guiding information through the loss function. For instance, Liu et al. [27] added an

edge loss term to the overall loss function. Hegde et al. [28] proposed gradient-aware mean-squared error loss (GAMSE) that effectively harnesses edge information. The second approach involves integrating the guiding information into the model itself [29], meaning incorporating the guiding information during the model's inference process. Conditional information can also be used to guide GAN models in depth map restoration. Numerous studies focus on how to use conditional information to guide or constrain the generation results of GAN models, such as conditional self-attention [30] and domain adaptation [31]. These models incorporate guiding information into the inference process, which generally satisfies the requirements for incomplete tasks that do not require high precision. In depth map completion, they can restore minor ill regions or convert sparse disparities into dense disparities. However, when the ill regions are too large, these models cannot effectively use the guiding information to restore the lost details, thus failing to meet the requirements for depth map completion. Accurate completion is crucial for depth map restoration tasks, particularly in autonomous driving scenarios where inaccurate completion due to large ill regions could lead to serious safety hazards. To effectively reintegrate detailed information into ill regions, this paper proposes GuidLN, which can effectively incorporate detailed information between different stages of the generator. This approach effectively guides the generator to accurately restore large ill regions.

Currently, many models require collaboration between multiple tasks, typically falling into three types. The first type directly concatenates multiple tasks and then inputs them into the main network [32]. The second type merges intermediate features of different tasks [33]. The third type aggregates the results from different tasks to obtain the final result [34]. Recent approaches have introduced more complex fusion methods, such as image-guided spatially variant convolution [16] and graph propagation [35]. However, these algorithms only fuse different tasks at different stages without considering the interference that may occur between tasks, making it difficult to effectively train all tasks in the model and seriously affecting the algorithm's performance. GAN models themselves are also multitask models, typically consisting of a generator and a discriminator, each responsible for different tasks. The generator generates results, while the discriminator evaluates the reasonability of the generator's outputs. In this paper, since the GAN model uses three discriminators for three different resolutions of results, it is more heavily impacted. To address this issue, we propose Attention-ResBlock, which enables different tasks in the multitask model to effectively focus on their respective tasks, thereby ensuring that the performance of the multitask model is not affected.

3. Methods

We designed an end-to-end network, Depth-GanNet, for restoring ill regions in depth maps, such as occluded, reflective, and low-texture regions. These regions cannot be accurately obtained due to the lack of information. We will introduce Depth-GanNet in three parts: the main network, GuidLN, and AttentionResNet modules.

3.1. Depth Information Precise Completion GAN

The overall structure is shown in Figure 2, where the generator of Depth-GanNet uses a U-shaped network as the main structure of the generator network, and the input of the network is a stereoscopic image containing diseased areas. In the decoder structure of the generator, we will obtain different scale results, whose sizes are respectively 1/4 of the original size, 1/2 of the original size, and the original size of the resulting image. In the network, we design three different discriminators to distinguish these results. The input of the generator is the depth map with labeled diseased areas and the depth edge map, where the depth map is used as the input of the generator's first layer of convolutions. The depth edge map is fused as guidance information in the various convolution blocks of the generator to guide the generator to repair the depth map more effectively.

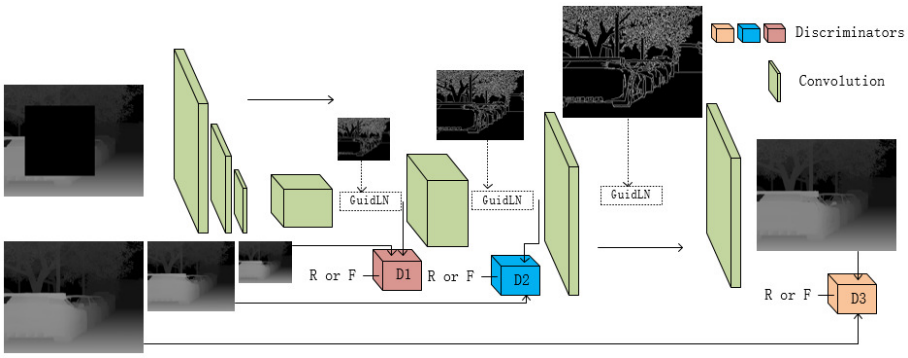

Figure 2. The overall structure of the model.

In this paper, we designed three discriminators to improve the generator's performance, composed of residual blocks. During training, the generator generates three different scales of results during the decoding process. This paper uses three discriminators to evaluate these three scales of images, which can comprehensively improve the generator's performance. Detecting small-scale images can better evaluate the overall quality of images, while evaluating large-scale images emphasizes the evaluation of details. Using multiscale discriminators to evaluate different scales of images can comprehensively detect the quality of images generated by the generator, making the network more precise and robust. As the generator decodes the feature map at each step, it can obtain more details from the feature map, and the images generated by the generator become more detailed. Therefore, evaluating the smallest-scale images focuses on evaluating the global information of images, helping the generator better learn global information. As the scale of the images increases, the discriminators can combine the new detailed information added by the generator to evaluate them, helping the generator better learn details.

3.2. Attention-ResBlock

The residual module is widely used in various tasks due to its excellent performance. In the network structure of this article, the edge extraction network and the generator and discriminator of the GAN network use residual blocks as the basic convolutional module. To make the generator and different discriminators adapt and focus on their own tasks, we propose the AttentionResNet block, which improves the ResNet block by incorporating a double-attention layer. The layer in ResNet obtains the weight of the feature map by calculating the correlation between the input feature map and the output feature map. The modules in different tasks adapt to their tasks by sharing weights. The structure of the module is shown in Figure 3.

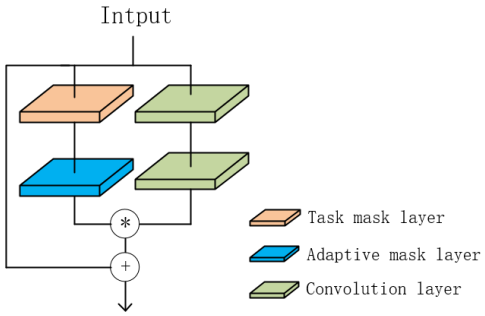

Figure 3. The structure of Attention-ResBlock.

The first layer is the task mask layer, which shares parameters with the entire subtask network, allowing the residual block to focus on the current task. For example, in the generator of the adversarial generation network described in this article, the task mask layer of each ResNet block shares parameters to focus on the image generation task of the generator. The three discriminators in the model also have a shared task mask layer

for each resolution, allowing them to focus on different tasks at different resolutions. The second layer is the adaptive mask layer, which does not share parameters with other Attention-ResBlock layers. It obtains the required weights for the block itself, allowing the Attention-ResBlock to have its weight adjustment function under the task mask layer. The calculation formula for the Attention-ResBlock is as follows:

$$y = \text{sigmoid}(weight_{m2} \cdot \text{sigmoid}(weight_{m1} \cdot \text{I})) * O \tag{1}$$

The operator "$\cdot$" represents the convolution operation, while the operator "$*$" represents the matrix multiplication operation. The parameter $weight_{m1}$ is the mask layer parameter of the module, while $weight_{m2}$ is the mask layer parameter of the residual block. The output of the right-hand two ordinary convolutions is O.

The AttentionResBlock in this article allows the generator and discriminator of the GAN network to focus more on their respective tasks. All residual blocks in the same task can learn to extract the necessary features more effectively. In the multiscale discriminator of this article, each discriminator can also focus on its discriminator task.

3.3. Guid Layer Normalization

When employing GANs to repair diseased areas, the model has the capability to utilize the surrounding information and incorporate global information in order to complete the depth of the region. However, the loss of details remains an issue, particularly in larger ill regions. Consequently, to prevent the introduction of random or incorrect fill-in information, it becomes essential to guide the network in repairing the image details of the ill regions.

Using RGB images or semantic images as guidance information can introduce redundancy, which may lead the network to perform erroneous repairs. In contrast, the utilization of depth edges as guidance information can effectively minimize redundancy. By relying on depth edges, the network can better focus on the specific details necessary for accurate repair work, resulting in improved performance and reduced likelihood of incorrect repairs.

To effectively guide the network in repairing details in a depth map, this paper proposes the GuidLayerNormalization(GuidLN) module placed after the generator network's feature extraction stage. This allows the generator to utilize guidance information better to supplement missing detail information, resulting in more accurate repair results.

As shown in Figure 4, the module accepts two inputs: the intermediate features map and the guidance map of the generator's output. Similar to the batch normalization module, σ is calculated for each channel of the feature map. The σ is used for two purposes. Firstly, it serves as the weight of the guidance information required for the feature map. Secondly, it is used for subsequent normalization of the feature map to improve the training speed of the network.

$$\omega = \frac{1}{e^{\sigma}} \tag{2}$$

This paper integrates guidance information into the feature map to guide the generator for precise completion, and then performs normalization on the feature map m to meet the required distribution for the model task.

$$y = \gamma_c \left(\frac{h_g - \mu_{c|g}}{\sigma_{c|g}} + \omega * \frac{h_x - \mu_{c|x}}{\sigma_{c|x}} \right) + \beta_c \tag{3}$$

where h is the activation value processed by the activation function in the generator, $\mu_{c|g}$ and $\sigma_{c|g}$ are the mean and variance of the guidance information in channel c, $\mu_{c|x}$ and $\sigma_{c|x}$ are the mean and variance of the feature of input in channel c.

$$\mu_c = \frac{1}{NHW} \sum_{n,y,x} h_{n,c,y,x} \tag{4}$$

$$\sigma_c = \sqrt{\frac{1}{NHW}\sum_{n,y,x}\left(\left(h_{n,c,y,x}\right)^2 - \left(\mu_c\right)^2\right)} \tag{5}$$

where γ and β are parameters designed to enhance the adaptability of the model. Unlike other models that input other information directly with the feature map into the module, we provide guidance throughout the entire generation process, allowing the guidance information to effectively guide the generation process. Since directly using guidance information would prevent the model from effectively integrating it into the feature map, we perform normalization on the guidance information and feature map information before effective fusion within a common standard range. This avoids dominant parameters being determined by data with large or small distributions, which can otherwise lead to unstable training. After fusion, the model can adaptively adjust the distribution of the feature map using the learnable γ and β parameters.

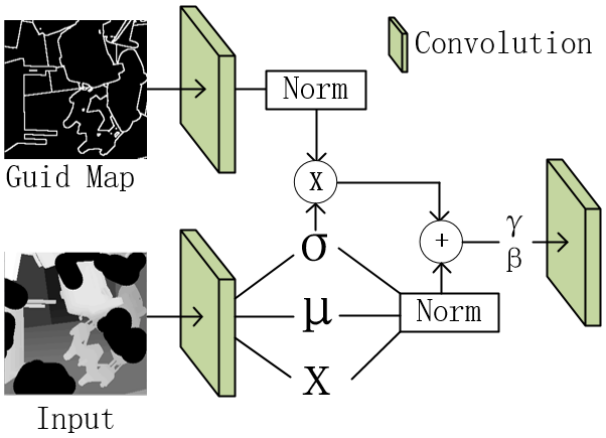

Figure 4. The structure of Guid normalization.

4. Experiments

4.1. Loss Function

The $L1$ loss and the $L2$ loss are the most widely used loss functions, but they have advantages and disadvantages. The $L1$ loss is relatively stable in the early stages of training, and noise in the dataset has a smaller impact on network training. However, it can cause difficult convergence in the later stages of training. On the other hand, the $L2$ loss is relatively smooth in the later stages of training, but it is more affected by noise in the early stages of training.

The smooth_{L1} loss function, developed on top of these two losses, enjoys both advantages. It is stable in the early and middle stages of training and smooth in the later stages, making it easier for the network to converge. Therefore, we adopt the smooth_{L2} loss function as the basis for designing the loss function of our model.

The output of the discriminator in this experiment is a tensor. The discriminator includes loss functions for both positive and negative examples. The loss function for negative examples of discriminator is shown in Formula (6):

$$\mathcal{L}_{DF} = \text{smooth}_{L1}(F - D(Y_i, x_i)) = \begin{cases} 0.5 * |F - D(Y_i, x_i)|^2, & \text{if } |F - D(Y_i, x_i)| < 1 \\ |F - D(Y_i, x_i)| - 0.5, & \text{otherwise} \end{cases} \tag{6}$$

The loss function for positive examples of discriminators is shown in Formula (7):

$$\mathcal{L}_{DR} = \text{smooth}_{L1}(R - D(Y_i, Y_i)) = \begin{cases} 0.5 * |F - D(Y_i, Y_i)|^2, & \text{if } |F - D(Y_i, Y_i)| < 1 \\ |F - D(Y_i, Y_i)| - 0.5, & \text{otherwise} \end{cases} \tag{7}$$

The total function of discriminators is shown in Formula (8):

$$\mathcal{L}_D = 0.5 * (\mathcal{L}_{DF} + \mathcal{L}_{DR}) \tag{8}$$

where F is a tensor corresponding to the negative examples, R is a tensor corresponding to the positive examples, and Y_i is the groundtruth.

The generator loss function contains two parts: the loss relative to the groundtruth and the loss caused by the discriminators.

$$\mathcal{L}_G = \text{smooth}_{L1}(Y_i - G(x_i)) = \begin{cases} 0.5 * |Y_i - G(x_i)|^2, & \text{if } |Y_i - G(x_i)| < 1 \\ |Y_i - G(x_i)| - 0.5, & \text{otherwise} \end{cases} \tag{9}$$

The total function of generator is shown in Formula (8):

$$\mathcal{L} = \mathcal{L}_G + \sum_{i=1}^{n} \lambda_i \mathcal{L}_{Di} \tag{10}$$

This paper presents a GAN-based depth map completion model that utilizes three different scales of discriminator. The generator's loss function is denoted as L_G, while the loss functions of the different discriminators are L_{D1}, L_{D2}, and L_{D3}, respectively. Each discriminator has its own set of hyperparameters: $\lambda_1 = 0.3$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.7$, respectively.

4.2. Dataset

The WHU-Stereo dataset is a public dataset collected by the GF-7 satellite from China, which contains 1757 sets of images with a resolution of 1024 × 1024. Each set of images contains a left image, a right image, and a disparity map. Although the disparity map is dense (each pixel contains a disparity value), it does not have accurate disparity values in ill regions. The dataset covers a variety of scenes, including cities, mountains, and rivers.

Sceneflow is an artificially generated dataset. Since it is generated in an ideal condition, it does not contain any noise. The dataset contains a variety of scenes, including both indoors and outdoors, as well as road scenes. It contains 354,540 training sets and 4370 testing sets with a resolution of 960 × 540. Each set of images contains a left image, a right image, and a disparity map, where the disparity map is dense.

KITTI is a real-world scene dataset with a large amount of noise. The dataset includes KITTI2012 and KITTI2015, mainly used for highway driving scenarios. KITTI2015 contains 200 sets of images for training with a resolution of 1242 × 375. KITTI2012 contains 194 sets of training images. Each set of images used in the dataset contains a left image, a right image, and a disparity map. The disparity map in the dataset is obtained from Lidar, so it is a sparse disparity map.

4.3. Metric

All the metrics for evaluation are shown as follows:

$$\text{RMSE(mm)} : \sqrt{\frac{1}{v}\sum_{x}\left(\hat{h}_x - h_x\right)^2} \tag{11}$$

$$\text{MAE(mm)} : \frac{1}{v}\sum_{x}\left|\hat{h}_x - h_x\right| \tag{12}$$

$$\text{iRMSE(1/km)} : \sqrt{\frac{1}{v}\sum_{x}\left(\frac{1}{\hat{h}_x} - \frac{1}{h_x}\right)^2} \tag{13}$$

$$\text{iMAE(1/km)} : \frac{1}{v}\sum_{x}\left|\frac{1}{\hat{h}_x} - \frac{1}{h_x}\right| \tag{14}$$

$$\delta_\tau : \max\left(\frac{h_x}{\hat{h}_x} - \frac{\hat{h}_x}{h_x}\right) < \tau, \tau \in \left\{1.25, 1.25^2, 1.25^3\right\} \tag{15}$$

4.4. Results

We compared recent models used for image or depth map completion in our experiments, including NLSPN [15], PRR [36], ACMNet [35], DySPN [37], UARes [38], and RDFGAN [24]. As shown in Figure 5, the RDFGAN model uses the original image as guidance. However, the excessive redundancy in the original image cannot accurately guide the model to repair the missing details. In contrast to RDFGAN, the model proposed in this paper uses the deep edge map instead of the original image as guidance information. Therefore, during the generation of results, we effectively avoid the excessive redundancy brought by the original image, which is prone to misleading the model into generating incorrect results. At the same time, our multiscale detector can also provide a comprehensive evaluation of the generator's results at different scales, thereby improving the generator's performance.

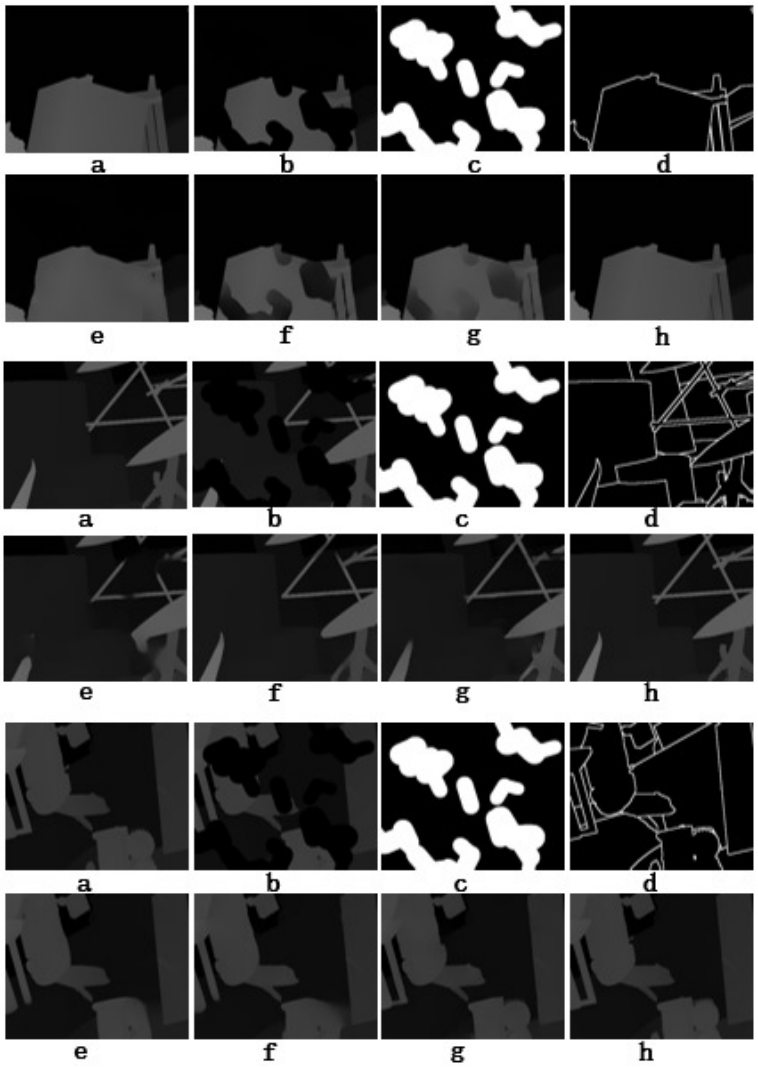

Figure 5. The visualization results of each algorithm on the SceneFlow dataset, where (**a**) is the ground truth, (**b**) is the disparity map after generating ill regions randomly, (**c**) is the label of the ill regions, (**d**) is the deep edge map, (**e**) is the result of the NLSPN model, (**f**) is the result of the ACMNet model, (**g**) is the result of the RDFGAN model, and (**h**) is the result of the model proposed in this paper.

Other models, such as ACMNet, cannot effectively repair lost details, and therefore lose details during the repair process and generate incorrect repair results. In contrast to these models, our model uses the GuildLN module to effectively introduce guidance information into the generator so that it can be more effectively completed in areas with much missing information and repair ill regions in the deep map.

As shown in Table 1, our algorithm achieved better results on the SceneFlow dataset, with a reduction of 36.1% and 9.3% in REL compared to NLSPN and RDFGAN, and a reduction of 32.3% and 9.7% in RSME compared to NLSPN and RDFGAN.

Table 1. The result on Sceneflow.

Method	MAE	iMAE	iRMSE (m)	REL	RMSE (m)	$\sigma_{1.25}$	$\sigma_{1.25}{}^2$	$\sigma_{1.25}{}^3$
NLSPN	145.68	0.9	1.94	0.061	0.288	91.5	95.6	97.5
PRR	127.89	0.81	1.68	0.054	0.257	92.3	96.3	97.9
ACMNet	108.54	0.78	1.42	0.051	0.261	92.8	96.6	98.3
DySPN	112.95	0.82	1.46	0.053	0.243	93.4	97.8	98.7
UARes	93.58	0.71	1.39	0.047	0.224	93.6	97.5	98.5
RDFGAN	89.21	0.68	1.33	0.043	0.216	94.2	98.2	98.8
Ours	81.32	0.63	1.26	0.039	0.17	96.3	98.6	99.1

Figure 6 shows the repair results of various models on the KITTI dataset. It can be seen that our model achieves better repair results. When the ill region contains multiple targets, the NLSPN and RDFGAN models cannot effectively obtain detailed information, so they cannot accurately distinguish the depths of various targets and generate incorrect repair results. However, our model can effectively utilize guidance information to supplement the missing detailed information in the ill region. Our model can still achieve good repair results, even when the ill region is large.

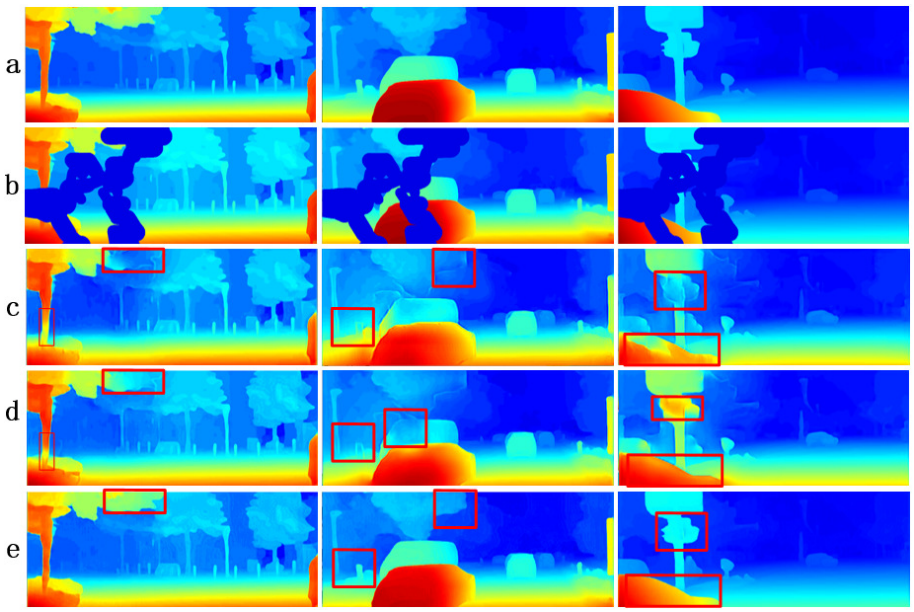

Figure 6. In the figure, row (**a**) is the groundtruth, row (**b**) is a randomly generated ill region, row (**c**) is the result diagram of the NLSPN model, row (**d**) is the result diagram of the RDFGAN model, and row (**e**) is the result diagram of our model.

Due to the unavoidable noise influence on the real dataset, as shown in Table 2, all models have a decline in performance on the KITTI dataset. However, our model is less influenced by noise compared to other algorithms, since the multiscale discriminator can distinguish results at different scales. The small-scale discriminator helps the generator to grasp the overall structure of the image, while the large-scale discriminator helps the generator to reduce the impact of noise while maintaining the overall image structure.

Table 2. The result on KITTI.

Method	MAE	RMSE (mm)	iMAE	iRMSE
NLSPN	209.46	898.13	1.08	2.23
PRR	183.32	762.25	1.03	2.12
ACMNet	157.13	620.36	0.94	1.98
DySPN	162.89	572.62	0.91	1.88
UARes	153.34	531.86	0.88	1.73
RDFGan	145.63	518.18	0.85	1.69
Ours	128.21	490.37	0.81	1.65

Figure 7 demonstrates that our model achieves good repair results for ill regions when repairing the WHU-Stereo dataset. Our model can effectively repair lost details in the images, such as the depth information of lost trees and blurred river banks.

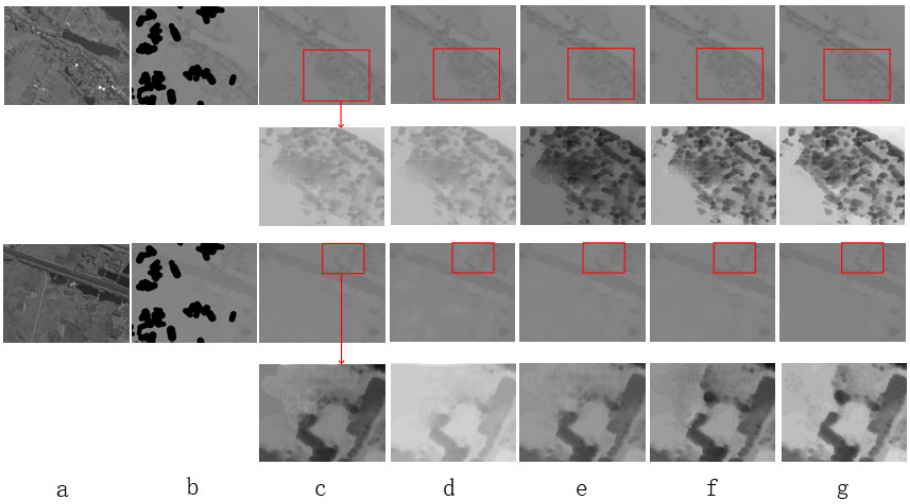

Figure 7. The (**a**) column shows a remote sensing image, the (**b**) column displays the disparity map of randomly generated defective regions, the (**c**) column exhibits the results of the NLSPN model, the (**d**) column displays the results of the PRR model, the (**e**) column shows the results of the UARes model, the (**f**) column displays the results of the RDFGAN model, and the (**g**) column shows the results of our algorithm. As the disparity map of remote sensing images is not visually prominent, we magnified the defective regions in this experiment and adjusted the contrast of the images to display the missed details by each model in defective regions more clearly.

4.5. Ablation Experiment

4.5.1. The Impact of GuidLN

Using concatenate to directly integrate guidance information into the repair process does not effectively utilize repair information, as the generator in the model cannot adapt the strength of the guidance information according to intermediate results information according to intermediate results as shown in Table 3. However, GuidLN can effectively integrate guidance information within a unified range.

Table 3. The result of ablation experiment.

Method	GildLN	Discri	AttenRes	Sceneflow		WHU-Stereo	
				MAE	RMSE	MAE	RMSE
Ours-N	✓			0.105	0.223	0.315	0.832
Ours-GA	✓		✓	0.088	0.212	0.288	0.805
Ours-GD	✓	✓		0.065	0.188	0.254	0.723
Ours-DA		✓	✓	0.071	0.198	0.271	0.785
Ours	✓	✓	✓	0.057	0.17	0.236	0.712

As shown in Figure 8, when the guidance information is not integrated with GuidLN, the model cannot achieve accurate repair when the repair area is large. Without true guidance information to guide the model during repair, the model can only perform random and incorrect repairs on the repair area. When the repair area is smaller, the lost detailed information is also smaller, so the impact of not integrating guidance information is not significant. However, when the repair area is large, there are many lost details in the repair area, so the results of the random repair by the model generate significant errors.

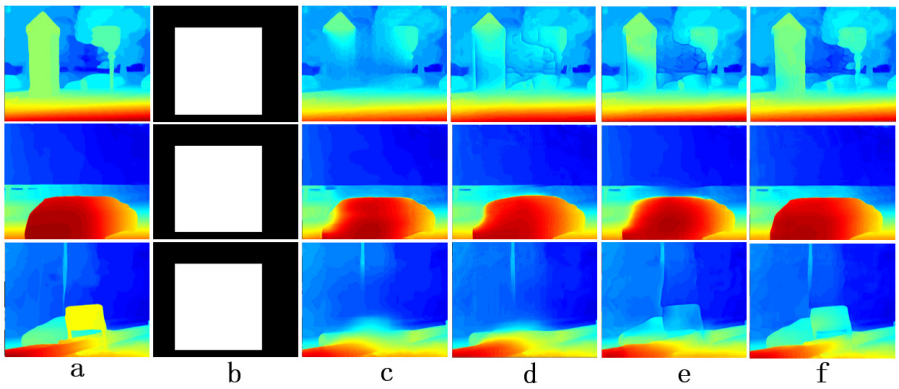

Figure 8. The (**a**) column shows a ground truth of disparity map, the (**b**) column masks a large area of repairable areas, the (**c**) column exhibits the results of the Ours-DA model, the (**d**) column displays the results of the Ours-GA model, the (**e**) column shows the results of the Ours-GD model, the (**f**) column displays the results of the complete model.

4.5.2. The Impact of Multiscale Discriminators

This paper compares the impact of using a single discriminator and multiple discriminators on the network. When using a single discriminator, we discriminate the results generated by the generator at the highest resolution. It can be seen from the results that various indicators of the model are affected. Without a low-resolution discriminator, the model lacks control over the overall structure, resulting in errors in the overall structure of objects in the results, which subsequently affects the results. Although the model incorporates detailed information as guiding information, it lacks accurate discrimination of detailed information due to the single discriminator's limitations in discerning details. Therefore, the final restoration results show that the model still cannot accurately complete the ill regions.

4.5.3. The Impact of Attention-ResBlock

When regular ResNet blocks are used instead of Attention-ResBlocks, the model's performance is seen to be compromised. That is because with regular ResNet blocks, all subtask modules in the network employ the same ResNet block structure. Consequently, the ResNet blocks in individual subtask modules cannot share task-specific information, and the subtask modules cannot fully focus on their respective tasks. That significantly impacts the performance of the model proposed in this paper. From Figure 8, it can be observed that although the restored results do not severely lose detailed information, the results in terms of detail completion are unsatisfactory.

5. Conclutions

In the problem of depth map restoration, where the lack of detailed information due to large ill regions results in deviations between the restored depth map and the real scene, this paper proposes the DIPC-GAN model, which can accurately restore the ill regions in the depth map. When the ill regions lack detailed information, our model effectively integrates depth edges as guiding information to restore the missing details in the ill regions accurately. To enhance the comprehensive learning of the generator for both the overall and detailed depth map, we use multiple discriminators to discriminate the

results generated by the generator at different resolutions, thereby improving the generator's performance. For GAN-based multitask networks, we propose Attention-ResBlock, which allows different subnetworks of the model to focus on their respective tasks, thereby enhancing the network's overall performance. Compared with recent generative restoration models, our model achieves good results in the task of depth map restoration, especially when the ill regions are large, where it can accurately restore the ill regions.

Author Contributions: Conceptualization, R.Q.; Methodology, R.Q.; Formal analysis, W.Y.; Investigation, W.Q., J.L., Y.W., R.F. and X.W.; Writing—original draft, R.Q.; Supervision, Y.Z.; Funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the Science and Technology Planning of Shenzhen (JCYJ20180503182133411), Technology Research and Development Fund (KQTD20200820113105004), National Natural Science Foundation of China (62266011) and Science and Technology Foundation of Guizhou Province (No. ZK[2022]119).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from rgbd images. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 746–760.
2. Chiu, Y.P.; Leou, J.J.; Hsiao, H.H. Super-resolution reconstruction for kinect 3D data. In Proceedings of the 2014 IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne, VIC, Australia, 1–5 June 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 2712–2715.
3. Ma, F.; Karaman, S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 4796–4803.
4. Dumoulin, V.; Shlens, J.; Kudlur, M. A learned representation for artistic style. *arXiv* **2016**, arXiv:1610.07629.
5. Wong, A.; Cicek, S.; Soatto, S. Learning topology from synthetic data for unsupervised depth completion. *IEEE Robot. Autom. Lett.* **2021**, *6*, 1495–1502. [CrossRef]
6. Sun, J.; Lin, Q.; Zhang, X.; Dong, J.; Yu, H. Kinect depth recovery via the cooperative profit random forest algorithm. In Proceedings of the 2018 11th International Conference on Human System Interaction (HSI), Gdansk, Poland, 4–6 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 57–62.
7. Yang, Y.; Wong, A.; Soatto, S. Dense depth posterior (ddp) from single image and sparse range. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3353–3362.
8. Ku, J.; Harakeh, A.; Waslander, S.L. In defense of classical image processing: Fast depth completion on the cpu. In Proceedings of the 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 8–10 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 16–22.
9. Uhrig, J.; Schneider, N.; Schneider, L.; Franke, U.; Brox, T.; Geiger, A. Sparsity invariant cnns. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 11–20.
10. Huang, Z.; Fan, J.; Cheng, S.; Yi, S.; Wang, X.; Li, H. Hms-net: Hierarchical multi-scale sparsity-invariant network for sparse depth completion. *IEEE Trans. Image Process.* **2019**, *29*, 3429–3441. [CrossRef] [PubMed]
11. Eldesokey, A.; Felsberg, M.; Khan, F.S. Propagating confidences through cnns for sparse data regression. *arXiv* **2018**, arXiv:1805.11913.
12. Ma, F.; Cavalheiro, G.V.; Karaman, S. Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3288–3295.
13. Li, A.; Yuan, Z.; Ling, Y.; Chi, W.; Zhang, C. A multi-scale guided cascade hourglass network for depth completion. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 32–40.
14. An, P.; Fu, W.; Gao, Y.; Ma, J.; Zhang, J.; Yu, K.; Fang, B. Lambertian Model-Based Normal Guided Depth Completion for LiDAR-Camera System. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
15. Park, J.; Joo, K.; Hu, Z.; Liu, C.K.; So Kweon, I. Non-local spatial propagation network for depth completion. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 120–136.
16. Yan, Z.; Wang, K.; Li, X.; Zhang, Z.; Li, J.; Yang, J. RigNet: Repetitive image guided network for depth completion. In Proceedings of the Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 214–230.

17. Zhang, Y.; Funkhouser, T. Deep depth completion of a single rgb-d image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 175–185.
18. Jiang, L.; Xiao, S.; He, C. Kinect depth map inpainting using a multi-scale deep convolutional neural network. In Proceedings of the 2018 International Conference on Image and Graphics Processing, Hong Kong, China, 24–26 February 2018; pp. 91–95.
19. Atapour-Abarghouei, A.; Breckon, T.P. *DepthComp: Real-Time Depth Image Completion Based on Prior Semantic Scene Segmentation*; British Machine Vision Association (BMVA): Durham, UK, 2017.
20. Li, J.; Gao, W.; Wu, Y. High-quality 3d reconstruction with depth super-resolution and completion. *IEEE Access* **2019**, *7*, 19370–19381. [CrossRef]
21. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
22. Atapour-Abarghouei, A.; Akcay, S.; de La Garanderie, G.P.; Breckon, T.P. Generative adversarial framework for depth filling via wasserstein metric, cosine transform and domain transfer. *Pattern Recognit.* **2019**, *91*, 232–244. [CrossRef]
23. Baruhov, A.; Gilboa, G. Unsupervised enhancement of real-world depth images using tri-cycle gan. *arXiv* **2020**, arXiv:2001.03779.
24. Wang, H.; Wang, M.; Che, Z.; Xu, Z.; Qiao, X.; Qi, M.; Feng, F.; Tang, J. Rgb-depth fusion gan for indoor depth completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 6209–6218.
25. Lopez-Rodriguez, A.; Busam, B.; Mikolajczyk, K. Project to adapt: Domain adaptation for depth completion from noisy and sparse sensor data. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.
26. Nguyen, T.M.; Yoo, M. Wasserstein generative adversarial network for depth completion with anisotropic diffusion depth enhancement. *IEEE Access* **2022**, *10*, 6867–6877. [CrossRef]
27. Liu, L.; Liao, Y.; Wang, Y.; Geiger, A.; Liu, Y. Learning steering kernels for guided depth completion. *IEEE Trans. Image Process.* **2021**, *30*, 2850–2861. [CrossRef]
28. Hegde, G.; Pharale, T.; Jahagirdar, S.; Nargund, V.; Tabib, R.A.; Mudenagudi, U.; Vandrotti, B.; Dhiman, A. Deepdnet: Deep dense network for depth completion task. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2190–2199.
29. Hwang, S.; Lee, J.; Kim, W.J.; Woo, S.; Lee, K.; Lee, S. Lidar depth completion using color-embedded information via knowledge distillation. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 14482–14496. [CrossRef]
30. Li, Y.; Chen, X.; Wu, F.; Zha, Z.J. Linestofacephoto: Face photo generation from lines with conditional self-attention generative adversarial networks. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2323–2331.
31. Xiang, X.; Liu, D.; Yang, X.; Zhu, Y.; Shen, X.; Allebach, J.P. Adversarial open domain adaptation for sketch-to-photo synthesis. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 1434–1444.
32. Liu, T.Y.; Agrawal, P.; Chen, A.; Hong, B.W.; Wong, A. Monitored distillation for positive congruent depth completion. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 35–53.
33. Yan, Z.; Wang, K.; Li, X.; Zhang, Z.; Li, G.; Li, J.; Yang, J. Learning complementary correlations for depth super-resolution with incomplete data in real world. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. [CrossRef] [PubMed]
34. Hu, M.; Wang, S.; Li, B.; Ning, S.; Fan, L.; Gong, X. Penet: Towards precise and efficient image guided depth completion. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 13656–13662.
35. Zhao, S.; Gong, M.; Fu, H.; Tao, D. Adaptive context-aware multi-modal network for depth completion. *IEEE Trans. Image Process.* **2021**, *30*, 5264–5276. [CrossRef] [PubMed]
36. Lee, B.U.; Lee, K.; Kweon, I.S. Depth completion using plane-residual representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13916–13925.
37. Zhu, Y.; Dong, W.; Li, L.; Wu, J.; Li, X.; Shi, G. Robust depth completion with uncertainty-driven loss functions. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 3626–3634.
38. Lin, Y.; Cheng, T.; Zhong, Q.; Zhou, W.; Yang, H. Dynamic spatial propagation network for depth completion. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 1638–1646.