Multi-modal Characteristic Guided Depth Completion Network

Yongjin Lee†, Seokjun Park†, Beomgu Kang, and HyunWook Park

Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
{dydwls462, diamondpark, almona, hwpark}@kaist.ac.kr

Abstract. Depth completion techniques fuse sparse depth map from LiDAR with color image to generate accurate dense depth map. Typically, multi-modal techniques utilize complementary characteristics of each modality, overcoming the limited information from a single modality. Especially in the depth completion, LiDAR data has relatively dense depth information for objects in the near distance but lacks the information of distant object and its boundary. On the other hand, color image has dense information for objects even in the far distance including the object boundary. Thus, the complementary characteristics of the two modalities are well suited for fusion, and many depth completion studies have proposed fusion networks to address the sparsity of LiDAR data. However, the previous fusion networks tend to simply concatenate the two-modality data and rely on deep neural network to extract useful features, not considering the inherited characteristics of each modality. To enable the effective modality-aware fusion, we propose a confidence guidance module (CGM) that estimates confidence maps which emphasizes salient region for each modality. In experiment, we showed that the confidence map for LiDAR data focused on near area and object surface, while those for color image focused on distant area and object boundary. Also, we propose a shallow feature fusion module (SFFM) to combine two types of input modality. Furthermore, a parallel refinement stage for each modality is proposed to reduce the computation time. Our results showed that the proposed model showed much faster computation time and competitive performance compared to the top-ranked models on the KITTI depth completion online leaderboard.

1 Introduction

Depth information is important in computer vision for various applications such as autonomous driving, and 3D reconstruction. For depth measurement, Light Detection and Ranging (LiDAR) sensors are commonly used, which measure accurate depth information. However, the LiDAR sensor provides the limited amount of valid depth points due to the hardware limitations. For example,

†These authors contributed equally.

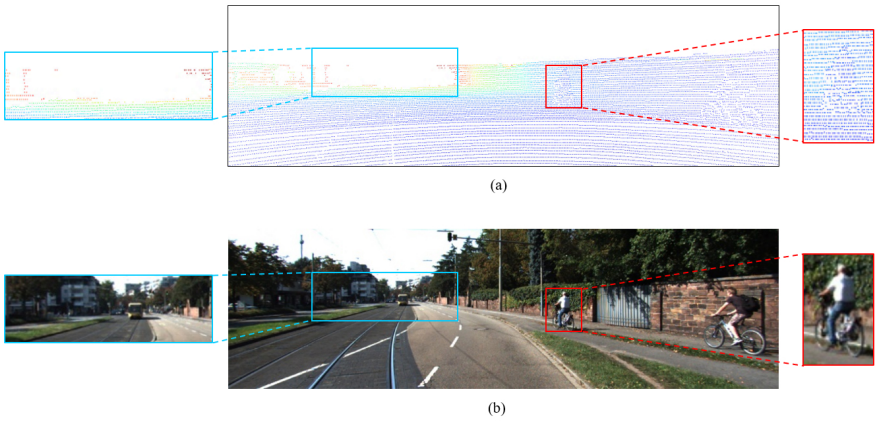

Fig. 1. Example of LiDAR data and color image showing their characteristic. The blue box, which represents distant area, shows that the LiDAR data has sparse depth information while the color image has dense color information. Also, the red box, which represents object, shows that the LiDAR data has depth information for object surface but hard to recognize object boundary. On the other hand, the object boundary can be easily recognized in color image.

the projected depth map of point cloud data measured by the Velodyne HDL-64E has a density of approximately 4% compared to the color image, which is insufficient for high-level applications such as autonomous driving [1].

To address the sparsity of depth data, which is a fundamental problem of the LiDAR sensor, color images can provide good complementary information. The two modalities, LiDAR data and color image, are completely complementary to each other. As shown in Fig. 1(a), the LiDAR data has depth information, but lacks the data points for distant area (blue) and object boundary (red), respectively. On the other hand, the color image counterpart part has dense color information. Therefore, the color image can complement the sparse part of LiDAR data to predict a dense depth map.

Recently, multi-modal depth completion networks have been proposed by developing architecture of neural network to extract effective fused features, such as feature extraction with the learned affinity among neighboring pixels [2–5], two-stage framework [6, 7], a cross guidance module [8], and a content-dependent and spatially-variant kernel from color images [9]. These novel neural network architectures were specialized for extracting useful features related to each modality. However, the complementary characteristic of each modality was remained to be explored in the fusion step. To further improve the fusion process, attention and confidence-based approaches have been developed. The estimated attention and confidence maps were used to guide the extracted features [10–12]. Although these methods takes advantage of the complementary characteristic by using respective attention and confidence maps, the inherited characteristics of each modality was not fully considered.

Table 1. Characteristics of LiDAR data and color image.

	LiDAR data	Color image
Depth information	O	X
Distant area information	Sparse	Dense
Object information	Object surface	Object surface and boundary

The 2D depth map shares the property of extremely unbalanced distribution of structures in image space resulted from the perspective projection, because 3D point cloud data is projected on the 2D images. Close objects have a large area in the image plane with sufficient depth points, whereas distant objects have a small area with insufficient depth points. The data distribution of the depth map from LiDAR data was taken into consideration [13, 14]. However, these modality characteristic based depth completion networks only considered the characteristic of LiDAR data.

In this study, we propose a multi-modal characteristic guided depth completion network that considers both characteristic of LiDAR data and color image, which are represented in Table 1. The proposed two-stage depth completion network predicts a dense coarse depth map in the first stage and refines it in the second stage. The confidence guidance module (CGM) is applied to the second stage to estimate confidence maps that represents salient region for each modality. To consider the multi-modal characteristic, the Sobel filter is utilized to detect the object boundary in the CGM and we show that the confidence maps are well predicted according to the properties of each modality. We also propose the shallow feature fusion module (SFFM) applied to combine two types of input modality with the sparsity invariant CNN (SI-Conv [1]). The color-refinement (CR) layer and depth refinement (DR) layer, each of which refines the depth maps in the second stage, are implemented in parallel to reduce the computation time. The final depth map is obtained by combining two depth maps from the refinement layers using the corresponding confidence maps.

To summarize, our contributions are as follows.

- We propose a multi-modal characteristic guided depth completion network to fully consider the both characteristic of LiDAR data and color image. The CGM plays an important role to estimate confidence map for each modality. The confidence map for LiDAR data focuses on the near area and object surface while those for color image focuses on the distant area and object boundary.
- We propose a simple and efficient combining module, SFFM, for sparse depth map by using sparsity invariant CNN. In the ablation study, the contribution of SFFM to performance improvement was shown.
- Our model showed much faster computation time and competitive performance compared to the top-ranked models on the KITTI depth completion online leaderboard by constructing the refinement layers and SFFM in parallel.

2 Related work

We briefly review previous studies on depth completion grouped by three types: conventional neural network based, attention- and confidence-map based, and modality characteristic based approach.

2.1 Conventional neural network based approach

As with deep learning based models, most depth completion studies focus on developing the architecture of neural network to improve performance. Cheng *et al.* [2] and Cheng *et al.* [3] proposed a simple and efficient linear propagation model, convolutional spatial propagation network (CSPN), to address blur of the result depth map. CSPN learned the affinity among neighboring pixels to refine the initial estimated depth map. As CSPN was successfully applied to depth completion, Park *et al.* [4] and Cheng *et al.* [5] further improved CSPN by proposing non-local spatial propagation network and CSPN++, respectively. However, CSPN methods suffer from slow computation time.

Xu *et al.* [6] proposed a unified CNN framework that consisted of prediction and refinement network. The prediction network estimated coarse depth, surface normal and confidence map for LiDAR data, and then diffusion refinement module aggregated the predicted maps to obtain the final results. Similarly, Liu *et al.* [7] designed a two-stage residual learning framework consisting of sparse-to-coarse and coarse-to-fine. In the sparse-to-coarse stage, the coarse dense depth map was obtained and combined with the features from the color image by channel shuffle. The energy-based fusion was implemented in the coarse-to-fine stage.

In addition, Lee *et al.* [8] designed a cross guidance module for multi-modal feature fusion, propagating with intersection of the features from different modality. Zhao *et al.* [15] applied a graph structure to extract multi-modal representation. Ma *et al.* [16] proposed an autoencoder network with self-supervised training framework. Tang *et al.* [9] estimated content-dependent and spatially-variant kernels from color images, where the kernels weights were applied to sparse depth map.

Although these methods have the novel architectures to assemble the multi-modal information by concatenating the LiDAR data and color image, the complementary characteristic of each modalities was not directly used. Therefore, the methods lacks the rationale how the information of each modality is used.

2.2 Attention- and confidence-map based approach

Van *et al.* [10] designed a confidence map based depth completion model that extracted global and local information from LiDAR map and RGB image by estimating confidence maps for each global and local branch. Then the global and local features were weighted by their respective confidence map to predict dense depth map. Similarly, Qiu *et al.* [11] considered surface normal as intermediate representation and fused the color image and surface normal with learned

attention maps. An additionally confidence map was predicted for LiDAR data to handle mixed LiDAR signals near foreground boundary. In addition, Hu *et al.* [12] proposed color-dominant and depth-dominant branches, and then combined the results of each branch with confidence map.

Like the above models, multi-branch networks adopting attention and confidence maps have shown high performance improvement. However, the extremely unbalanced distribution of structures resulted from the perspective projection in both modalities were not considered. Therefore, a more effective method for utilizing the property of modality is necessary.

2.3 Modality characteristic based approach

LiDAR data and color image have their own characteristic because of the unique properties of sensors. Recent studies posed and addressed the problem of the extremely unbalanced distribution of structures. Li *et al.* [13] argued that most of the LiDAR data was distributed within a distance of 20 meters, and the variance of depth for distant object farther than 30 meters was quite large. Based on the claim, they proposed a multi-scale guided cascade hourglass network, considering the unbalanced data distribution for effective fusion of two different types of data. They extracted multi-scale features from color image and predicted multi-scale depth map to represent the different data distributions. Also, Lee *et al.* [14] changed the regression task to the classification task for depth completion by considering data distribution of the depth map. They separated the depth map and color image into multiple planes along the channel axis, and applied channel-wise guided image filtering to achieve accurate depth plane classification results. Although the modality characteristic based approaches showed the improved performance and properly addressed the unbalanced distribution problem, they did not consider the property of color image. Therefore, a more effective method that considers both properties of the LiDAR data and color image is necessary.

3 METHODOLOGY

3.1 Overall Network Architecture

The entire architecture of the proposed model is described in Fig. 2. Note that all encoder-decoder blocks, which are coarse prediction, color refinement, and depth refinement, have same network architecture containing residual blocks as shown in Fig. 3. Our model is a two-stage network. In the first stage, a coarse dense depth map called first depth map is predicted from a color image and a sparse depth map as follows:

$$D_c = CP(SFFM(I_c, I_s)) \tag{1}$$

where D_c denotes the coarse dense depth map from the first stage, I_c denotes the input color image, I_s denotes the input sparse depth map, the CP is the

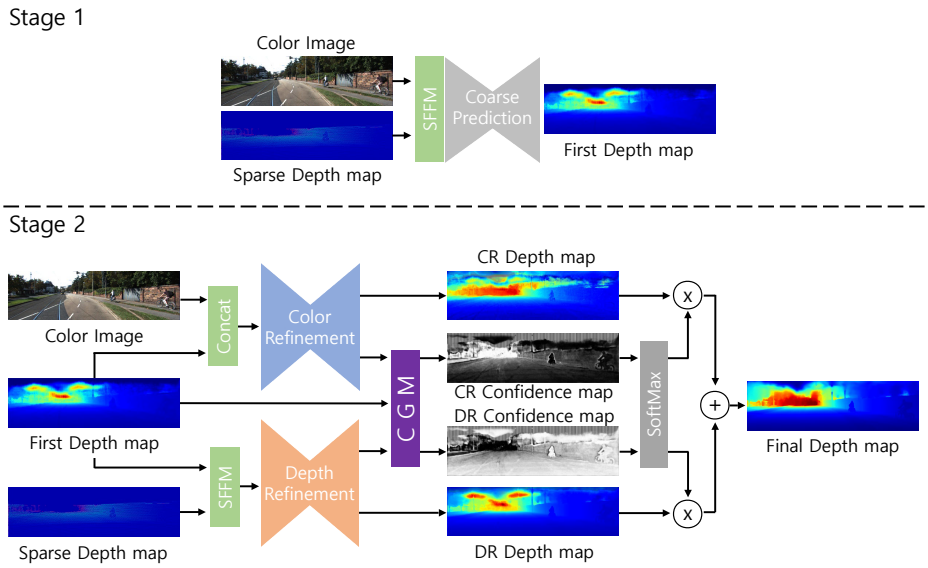

Fig. 2. Overall diagram of the proposed two-stage model. The coarse dense depth map is obtained in the first stage by coarse-prediction (CP) layer, and refined by color-refinement (CR) layer and depth-refinement (DR) layer in second stage. Shallow feature fusion module (SFFM) is applied to fuse the sparse depth map and dense data

coarse-prediction layer in Fig. 2, and the SFFM is the proposed feature fusion module.

In the second stage, the color-refinement (CR) and depth-refinement (DR) layers refine the coarse dense depth map with the color image and the sparse depth map respectively, and predict initial confidence maps at the same time, which can be written as:

$$(D_{cr},\ C_{ic}) = CR(D_c, I_c) \tag{2}$$

$$(D_{dr},\ C_{id}) = DR(SFFM(D_c, I_s)) \tag{3}$$

where D_{cr} denotes the dense depth map from the CR layer, C_{ic} denotes the initial confidence map from the CR layer, D_{dr} denotes the dense depth map from the DR layer, C_{id} denotes the initial confidence map from the DR layer.

The estimated initial confidence maps, C_{ic} and C_{id}, are refined to represent the characteristic of each modality through the confidence guidance module (CGM). The CGM receives two initial confidence maps and first depth map, and outputs CR confidence map and DR confidence map, which can be written as:

$$(C_{cr},\ C_{dr}) = CGM(D_c, C_{ic}, C_{id}) \tag{4}$$

where C_{cr} denotes the CR confidence map, and C_{dr} denotes the DR confidence map.

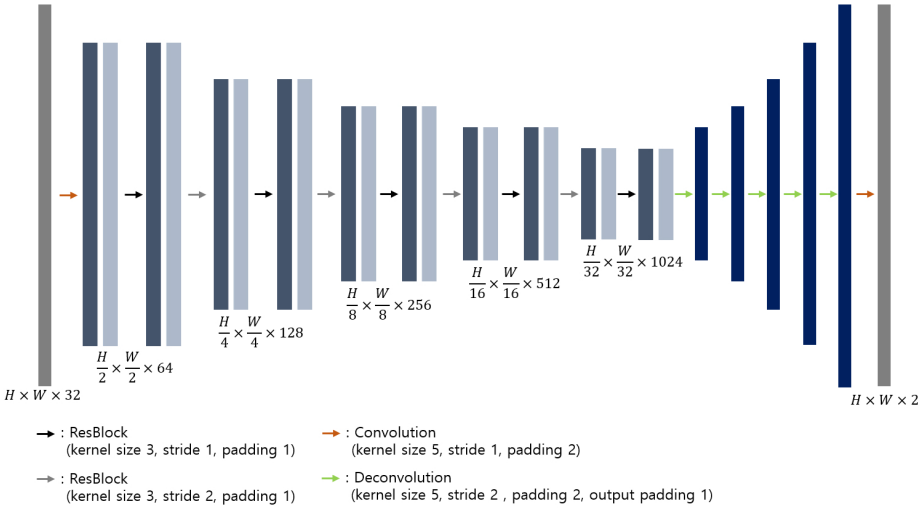

Fig. 3. Detailed architecture for coarse prediction, color refinement, and depth refinement layers

The CR and DR layers do not need to predict accurate depth maps for all regions. Each layer refines the exclusive region by using the confidence map that can effectively make use of the distinctive characteristics of different input data spaces. The two layers are complementary, and the final depth map is obtained through the fusion of the depth maps using the confidence map, which can be written as:

$$D_f(u,v) = \frac{e^{C_{cr}(u,v)} \cdot D_{cr}(u,v) + e^{C_{dr}(u,v)} \cdot D_{dr}(u,v)}{e^{C_{cr}(u,v)} + e^{C_{dr}(u,v)}} \tag{5}$$

where (u, v) denotes a pixel, and D_f denotes the final depth map.

3.2 Shallow Feature Fusion Module (SFFM)

The SFFM extracts the features, which is robust to the depth validity. The point cloud data is generated from the rotation of the LiDAR sensor, and the sparse depth map is generated by the projection of this point cloud data. Therefore, there is randomness in the validity of the sparse depth map even for the same scene. Also, since invalid pixels are encoded as zero values in the projected sparse depth map, when conventional convolution is used, it may be difficult to learn the kernel depending on the local density of valid pixels.

To solve this problem, we proposed the SFFM. The SFFM consists of parallel convolutional layers. For the sparse depth map, features invariant to the scale according to the validity of pixels are extracted using the sparsity invariant CNNs (SI-Conv [1]), and the color image is extracted by conventional convolution.

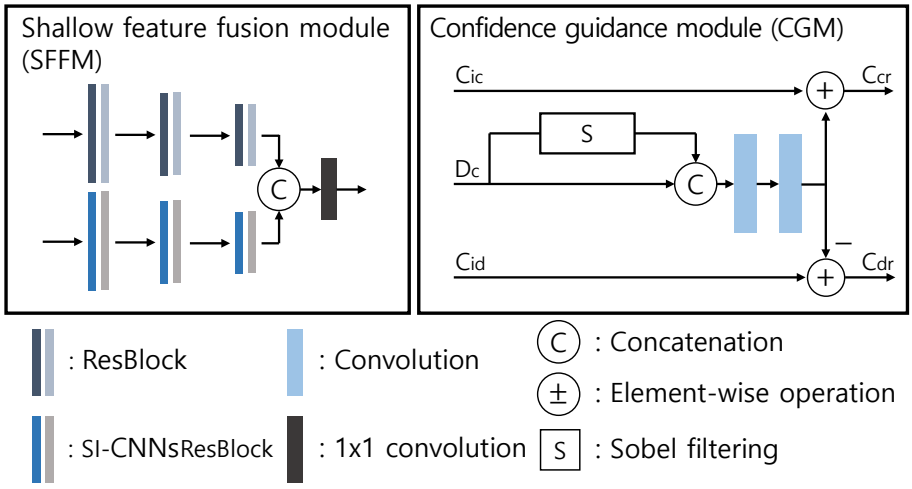

Fig. 4. Detailed architectures of SFFM and CGM.

Depending on the application of the SI-Conv, a high-density feature map can be extracted from the sparse depth map. To consider the remaining invalid pixels, the final feature map is extracted using 1×1 convolution on the concatenated features of dense color image and the high-density feature map. In this case, 1×1 convolution only needs to consider two cases, valid or invalid. SFFM makes it possible to effectively combine two data using only a small number of parameters.

3.3 Confidence Guidance Module (CGM)

A color image and a sparse depth map are the input signals of the CR layer and the DR layer, respectively, but there is no guarantee that each information will be used effectively. CGM is proposed to fully utilize the different characteristics of both data. The color image has dense data, so it has enough information about distant area and object boundaries. On the other hand, the sparse depth map is sparse but accurate, so it is useful to refine near area and object surface, which have many depth measurements. Therefore, the CR confidence map should have large values on distant area and object boundaries while the DR confidence map should have large values on near area and object surface. It makes that the two-modality data are fully utilized for depth completion.

CGM obtains the distance information of each pixel from the first depth map, and then obtains information about object boundaries from the first depth map by applying the Sobel filter[17], which are shown in Fig. 5. The sum of the two maps can adjust the confidence map. However, to reduce the scale difference, the final guidance map is obtained using concatenation and positive convolution. Finally, the guidance map has high values for object boundaries and distant

pixels. This map is added to the CR initial confidence map and subtracted from the DR initial confidence map, to become two confidence maps.

3.4 Loss Function

The ground truth depth map is semi-dense and invalid pixels are represented as 0. Therefore, the loss is defined only for the valid pixels by calculating the mean squared error (MSE) between the final depth map and the ground truth map as follow:

$$L_{final} = \frac{1}{|V|} \sum_{(u,v)\in V} \left\| (D_{gt}(u,v) - D'_f(u,v)) \right\|^2 \tag{6}$$

where V denotes the set of valid pixels, $D'_f(u,v)$ denotes the final depth map at pixel (u,v) and $D_{gt}(u,v)$ denotes the ground truth depth map at pixel (u,v).

To train the network more stable, the loss for the first stage depth map was also computed in the early epochs as follows:

$$L_{first} = \frac{1}{|V|} \sum_{(u,v)\in V} \left\| (D_{gt}(u,v) - D_c(u,v)) \right\|^2 \tag{7}$$

where $D_c(u,v)$ denotes the coarse depth map called first depth map at pixel (u,v).

The overall loss can be written as:

$$L_{total} = C_{first} \times L_{first} + L_{final} \tag{8}$$

where C_{first} is a hyper-parameter of 0.3 at the first epoch and reduces to 0 at 5^{th} epoch

4 EXPERIMENTS

4.1 Experimental setup

KITTI depth completion dataset: The KITTI depth completion dataset is a large real-world street view dataset captured for autonomous driving research [1], [18]. It provides sparse depth maps of 3D point cloud data and corresponding color images. The sparse depth maps have a valid pixel density of approximately 4% and the ground truth depth maps have a density of 16% compared to the color images ([1]). This dataset contains 86K training set, 1K validation set, and 1K test set without ground truth. KITTI receives the predicted depth maps of the test set and provides the evaluation results.

Implementation details: We trained our network on two NVIDIA TITAN RTX GPUs with batch size of 8 for 25 epochs. We used the ADAM optimizer [19] with $\beta_1 = 0.9, \beta_2 = 0.99$ and the weight decay of 10^{-6}. The learning rate started at 0.001 and was halved for every 5 epochs. For data augmentation, color jittering and horizontal random flipping were adopted.

Evaluation metrics: We adopt commonly used metrics for comparison study, including the inverse root mean squared error (iRMSE [1/km]), the inverse mean absolute error (iMAE [1/km]), the root mean squared error (RMSE [mm]), the mean absolute error (MAE [mm]) and the runtime ([s]).

4.2 Comparison with state-of-the-art methods

We evaluated the proposed model on the KITTI depth completion test set. Table 2 shows the comparison results with other top ranked methods. The proposed model shows the fastest runtime, and shows similar performance to SoTA model, PENet [12], and higher than other top-ranked methods on RMSE, which is the most important metric in depth completion. Moreover, our model shows higher performance than SoTA model in iRMSE and iMAE.

Table 2. Comparison with state-of-the-art methods on the KITTI Depth Completion test set.

Method	iRMSE	iMAE	RMSE	MAE	Runtime
CrossGuidance [8]	2.73	1.33	807.42	253.98	0.2 s
PwP [6]	2.42	1.13	777.05	235.17	0.1 s
DeepLiDAR [11]	2.56	1.15	758.38	226.50	0.07s
CSPN++ [5]	**2.07**	**0.90**	743.69	209.28	0.2 s
ACMNet [15]	2.08	0.90	744.91	**206.09**	0.08 s
GuideNet [9]	2.25	0.99	736.24	218.83	0.14 s
FCFR-Net [7]	2.20	0.98	735.81	217.15	0.13 s
PENet [12]	2.17	0.94	**730.08**	210.55	0.032s
Ours	2.11	0.92	733.69	211.15	**0.015 s**

Table 3. Ablation studies on the KITTI depth completion validation set. B: basic two-stage model, CR and DR: the second stage of B is replaced with the CR and DR layers.

Models	iRMSE	iMAE	RMSE	MAE
B	2.29	0.98	779.68	224.91
CR and DR	2.17	0.93	769.28	213.30
CR and DR + SFFM	2.17	0.91	764.93	212.71
CR and DR + SFFM + CGM	**2.17**	**0.91**	**759.90**	**209.25**

4.3 Ablation studies

In this section, we conducted ablation studies on the KITTI validation dataset to verify the effectiveness of the proposed model. The experimental results are shown in Table 3. B is a basic two-stage model, which predicts a first depth map from the concatenated input of a color image and a sparse depth map in first stage, and predicts a final depth map from the concatenated input of a first depth map, a color image and a sparse depth map. The CR and DR replace the

Fig. 5. The middle results of CGM. (a) The reference color images, (b) first depth maps form stage 1, and (c) feature maps after applying the Sobel filter. (c) Sobel-filtered features have high intensity on the pixel of object boundary and distant area, meaning that the Sobel filter is an essential factor in CGM to represent the characteristics of each modality.

second stage of the basic two-stage model. Each encoder-decoder takes a first depth map concatenated with a color image or a sparse depth map. The results show the CR and DR layers archives significant improvement in all the metrics, and both of the SFFM and the CGM also gives a performance improvement.

4.4 Analysis for predicted confidence map

We analyzed the predicted confidence map to verify that the proposed model properly utilized the characteristic of the two modality. In Fig. 5, the Sobel filter plays a important role to highlight the distant area and object boundary, where color image can complement LiDAR data by using the dense color information. With the first depth map through Sobel filter and the initial confidence maps from the CR and DR layers, the CGM outputs CR and DR confidence maps which are shown in Fig. 6. The CR confidence map highlights on the distant area and object boundary, where specialized for the dense color image, and the DR confidence map highlights on the near area and object surface, where specialize for the LiDAR data. It means that the proposed model properly utilizes the two modality inputs according to each characteristic.

Also, Fig. 7 shows that the results of multi-modal characteristic based depth completion network. The final depth map, Fig.7 (g), which is the weighted sum of CR depth map (Fig. 7 (e)) and DR depth map(Fig. 7 (f)), is similar to CR depth map for the distant area and object boundary, while it is similar to DR depth map for the near area and object surface.

Fig. 6. Qualitative result of the predicted CR and DR confidence maps. The CR confidence map focuses on the distant area and object boundary, while the DR confidence map focuses on the near area and object surface. It shows that the proposed model properly utilizes each modality input according to its characteristic.

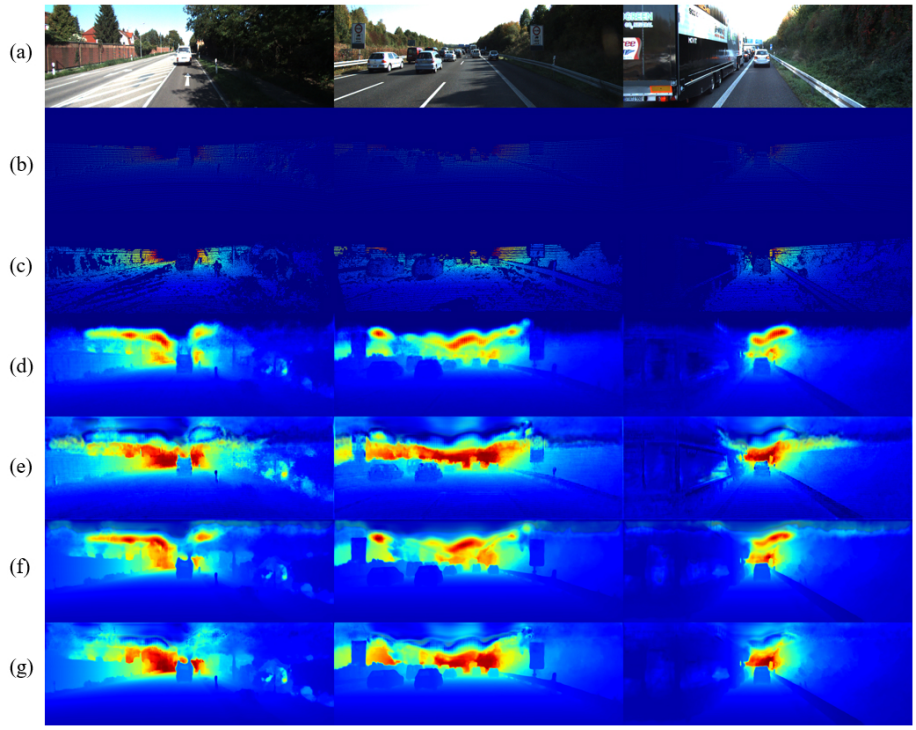

Fig. 7. Qualitative results on the KITTI depth completion validation dataset. (a) color images, (b) sparse depth maps, (c) ground truth depth maps, (d) first depth maps from CP layer, (e) CR depth maps from the CR layer, (f) DR depth maps from the DR layer, and (g) final depth maps.

5 CONCLUSION

The paper proposed a fast multi-modal characteristic guided depth completion network to estimate accurate dense depth maps. The proposed network has a two-stage structure including a shallow feature fusion module (SFFM), coasre-prediction (CP) layer, color-refinement (CR) and depth-refinement (DR) layers, and confidence guidance module (CGM). The first depth map from the CP layer is effectively refined in the CR and DR layers and consequently combined with the confidence map according to the multi-modal characteristic. Compared with the top-ranked models on the KITTI depth completion online leaderboard, the proposed model shows much faster computation time and competitive performance.

Acknowledgements This work was conducted by Center for Applied Research in Artificial Intelligence(CARAI) grant funded by Defense Acquisition Program Administration(DAPA) and Agency for Defense Development(ADD) (UD190031RD).

References

1. Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A.: Sparsity invariant cnns. In: 2017 international conference on 3D Vision (3DV), IEEE (2017) 11–20
2. Cheng, X., Wang, P., Yang, R.: Learning depth with convolutional spatial propagation network. IEEE transactions on pattern analysis and machine intelligence **42** (2019) 2361–2379
3. Cheng, X., Wang, P., Yang, R.: Depth estimation via affinity learned with convolutional spatial propagation network. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 103–119
4. Park, J., Joo, K., Hu, Z., Liu, C.K., Kweon, I.S.: Non-local spatial propagation network for depth completion. In: Proceedings of the European Conference on Computer Vision (ECCV). (2020) 120–136
5. Cheng, X., Wang, P., Guan, C., Yang, R.: Cspn++: Learning context and resource aware convolutional spatial propagation networks for depth completion. In: Proceedings of the AAAI Conference on Artificial Intelligence. Volume 34. (2020) 10615–10622
6. Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., Li, H.: Depth completion from sparse lidar data with depth-normal constraints. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 2811–2820
7. Liu, L., Song, X., Lyu, X., Diao, J., Wang, M., Liu, Y., Zhang, L.: Fcfr-net: Feature fusion based coarse-to-fine residual learning for monocular depth completion. arXiv preprint arXiv:2012.08270 (2020)
8. Lee, S., Lee, J., Kim, D., Kim, J.: Deep architecture with cross guidance between single image and sparse lidar data for depth completion. IEEE Access **8** (2020) 79801–79810
9. Tang, J., Tian, F.P., Feng, W., Li, J., Tan, P.: Learning guided convolutional network for depth completion. IEEE Transactions on Image Processing **30** (2020) 1116–1129

10. Van Gansbeke, W., Neven, D., De Brabandere, B., Van Gool, L.: Sparse and noisy lidar completion with rgb guidance and uncertainty. In: 2019 16th international conference on machine vision applications (MVA), IEEE (2019) 1–6
11. Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B., Pollefeys, M.: Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2019) 3313–3322
12. Hu, M., Wang, S., Li, B., Ning, S., Fan, L., Gong, X.: Penet: Towards precise and efficient image guided depth completion. In: 2021 International Conference on Robotics and Automation (ICRA), IEEE (2021) 13656–13662
13. Li, A., Yuan, Z., Ling, Y., Chi, W., Zhang, C., et al.: A multi-scale guided cascade hourglass network for depth completion. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. (2020) 32–40
14. 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. (2021) 13916–13925
15. Zhao, S., Gong, M., Fu, H., Tao, D.: Adaptive context-aware multi-modal network for depth completion. IEEE Transactions on Image Processing (2021)
16. Ma, F., Cavalheiro, G.V., Karaman, S.: Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In: 2019 International Conference on Robotics and Automation (ICRA), IEEE (2019) 3288–3295
17. Kanopoulos, N., Vasanthavada, N., Baker, R.L.: Design of an image edge detection filter using the sobel operator. IEEE Journal of solid-state circuits **23** (1988) 358–367
18. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32** (2013) 1231–1237
19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)