SLiDE: Self-supervised LiDAR De-snowing through Reconstruction Difficulty (Supplementary Material)

Gwangtak Bae¹, Byungjun Kim¹, Seongyong Ahn², Jihong Min², and Inwook Shim³

¹Seoul National University,²Agency for Defense Development,³Inha University {tak3452,peterbj95,seongyong.ahn,happymin77}@gmail.com, iwshim@inha.ac.kr

In the supplementary material, we present more details on the data synthesis method used for quantitative evaluations in Section 1. Then, we demonstrate more qualitative results on the synthetic data in Section 3 and on the real-world data in Section 4.

1 Details on Data Synthesis Method

Here, we present detailed information on our data synthesis method which is introduced to enable extensive quantitative evaluations. Real-world snowy point clouds and clean point clouds are captured by using our stationary data capturing system. Then, snow points are picked out by comparing noisy point clouds and clean point clouds. After augmenting the collected snow points, snow points are synthesized into other road scene point clouds captured under favorable weather conditions.

1.1 Data Capturing System

We set up a data capturing system that consists of a LiDAR sensor (Velodyne 'HDL-32E') and a control computer in a rainproof ruggedized case. Our system captures data periodically. To capture weather effects only, the system is placed in the controlled outdoor environment where people are not allowed to come into the scene. Fig. 1 shows our data capturing system. LiDAR is installed at a height of 1.7m which is similar to the height of LiDAR mounted on the top of autonomous vehicles.

1.2 Data Acquisition

Noisy LiDAR point cloud sets (*Noise-Set*) are captured in snowy weather conditions, and clean point cloud sets (*Clean-Set*) are captured in favorable weather conditions. In most cases, the detected snow noise point is located between a sensor origin and a background point; that is, we cannot detect the background point if the noise point is placed on the same ray of the LiDAR. The range image representation inherently involves this ray direction property, and we can directly pick out noise points by comparing range images of *Noise-Set* and *Clean-Set*. We generate a reference range image from *Clean-Set* of 200 range images by selecting the minimum range value of each sensor ray. Then, snow noise points are collected from *Noise-Set* of 7, 687 range images. 2 G. Bae et al.



Fig. 1: Our stationary data capturing system.



Fig. 2: Data annotation process. A label map L^N is assigned by comparing a noisy range image R^N (a) and the reference range image R^F (b). Collected noise points are depicted as the red points in (c).

1.3 Data Annotation

To distinguish noise points from background points, we compare the range values between the reference range image $R^F \in \mathbb{R}^{n \times m}$ and each range image $R^N \in \mathbb{R}^{n \times m}$ of *Noise-Set*, where *n* and *m* indicate the height and width of the range image. Then, we assign label information as follows:

$$L_{(u,v)}^{N} \leftarrow \begin{cases} \mathbf{N} \text{ (Noise)}, & \text{if } R_{(u,v)}^{F} \ge R_{(u,v)}^{N} + \tau \\ \mathbf{C} \text{ (Clean)}, & \text{else} \end{cases}$$
(1)

where $L^N \in \mathbb{R}^{n \times m}$ indicates a label map of R^N and τ is a margin for sensing errors of LiDAR. In the case of scanning an empty space (e.g., sky), since no valid background information is projected to $R^F_{(u,v)}$, we always assign the 'Noise' label to $L^N_{(u,v)}$. Through this process, we collect real-world noise points from each noisy point cloud in snowy weather, as seen in Fig. 2.



Fig. 3: The number of data obtained for each noise level.



Fig. 4: Process of generating a synthesized scene R^S . Noise points in R^N are synthesized into a road scene R^B .

1.4 Data Augmentation

The label map L^N only covers a part of the entire scanning area of LiDAR because our data capturing system has a limited horizontal Field-of-View as seen in Fig. 1. To supplement this limited scanning area, we need to augment the annotated noise points in \mathbb{R}^N . To this end, we generate the augmented noise points by combining multiple \mathbb{R}^N s belonging to the same noise level. The noise level is decided based on the density of noise points, following the criterion used in the Canadian Adverse Driving Condition Dataset [5]: Light $(0.0 - 0.6m^{-3})$, Medium $(0.3 - 0.6m^{-3})$, Heavy $(0.6 - 0.9m^{-3})$, and Extreme (above $0.9m^{-3})$. Fig. 3 shows the number of data obtained for each noise level.

1.5 Data Synthesis

We now generate synthetic snowy scenes by injecting the augmented snow noise points into road scene point cloud sets taken in clean weather (*Base-Set*). When synthesizing the augmented snow noise points into *Base-Set*, we have to consider scene structures of point clouds in *Base-Set*. If snow points are injected naively, unrealistic results may occur (e.g., snowflakes inside a car). To prevent this issue, a clean point in *Base-Set* is replaced by a noise point as follows:

$$R_{(u,v)}^{S} = \begin{cases} R_{(u,v)}^{N}, & \text{if } R_{(u,v)}^{N} \le R_{(u,v)}^{max} \text{ and } L_{(u,v)}^{N} = \mathbf{N} \text{ (Noise)} \\ R_{(u,v)}^{B}, & \text{else} \end{cases}$$
(2)



Fig. 5: Synthesized scenes of each noise level.

where $R^B \in \mathbb{R}^{n \times m}$ is a range image in *Base-Set*, and $R^S \in \mathbb{R}^{n \times m}$ is a synthesized range image. R^{max} is the maximum detectable range of noise points which is introduced for a realistic synthesis [1,4] by considering scene structures of R^B as follows:

$$R_{(u,v)}^{max} = \min(\frac{-ln(\frac{n}{I_{(u,v)}^{B}+g})}{2*\beta}, R_{(u,v)}^{B}),$$
(3)

given the received laser intensity $I^B_{(u,v)} \in \mathbb{R}^{n \times m}$, the adaptive laser gain g, the atmospheric extinction coefficient β , and the detectable noise floor n.

Examples of the data synthesis process and results are described in Fig 4 and 5, respectively. Note that we use the same road scene R^B to highlight the differences of each noise level. In our experiments, we do not re-use the same road scene to generate R^S . We utilize the Nuscenes dataset [2] as *Base-Set* to generate a number of synthetic snowy scenes.

2 Details on the Semi-supervised Extension

We proposed and evaluated three different weighting functions for combining the supervised and the self-supervised loss functions. In this section, the weighting functions are explained. Firstly, weighting functions for $Ramp\ up/down$ are as follows,

$$w_{self,ramp} = \begin{cases} 0, & t < T_s \\ \exp(-5(1 - \frac{t - T_s}{T_u - T_s})^2), & T_s \le t < T_u \\ 1, & T_u \le t < T_d \\ \exp(-12.5(1 - \frac{T_e - t}{T_e - T_d})^2), & T_d \le t < T_e \end{cases}$$
(4)

$$w_{sup,ramp} = \begin{cases} 0, & v \in I_s \\ 1, & T_s \le t < T_e \end{cases}$$
(5)

where w_{self} and w_{sup} are the weights for the self-supervised loss and the supervised loss, respectively. t is the current epoch. Ramp up starts at T_s and ends at T_u . Ramp down starts at T_d and ends at T_e . Secondly, the weighting functions for *Pretrain* are as follows,

$$w_{self,pretrain} = \begin{cases} 1, & t < T_s \\ 0, & T_s \le t < T_e \end{cases}$$
(6)

$$w_{sup,pretrain} = 1 - w_{self,pretrain}.$$
 (7)

Lastly, the weighting functions for Smooth transfer are as follows,

$$w_{self,smooth} = \begin{cases} 1, & t < T_s \\ \exp(-12.5(\frac{t-T_s}{T_e-T_s})^2), & T_s \le t < T_e \end{cases}$$
(8)

$$w_{sup,smooth} = 1 - w_{self,smooth}.$$
 (9)

3 Additional Qualitative Results on Synthetic Data

More qualitative results are provided for comprehensive evaluations. Our selfsupervised de-snowing method is compared with the state-of-the-art label-free method, DROR [3] and the state-of-the-art supervised method, WeatherNet [4]. Fig. 6 - 9 presents qualitative comparisons on a light, medium, heavy, and extreme snow scenes, respectively.

4 Additional Qualitative Results on Real-world Data

More qualitative results on real-world data are presented for comprehensive evaluations. In Fig. 10 and Fig. 11, our self-supervised de-snowing method is compared with the state-of-the-art label-free method, DROR [3], which also does not use any point-wise annotation. The supervised method, WeatherNet [4], cannot be used in this scenario since the collected data do not contain point-wise annotations.



Fig. 6: Additional qualitative comparisons on a *light snow* scene of the synthesized snow noise data. The first row shows all points with their prediction results (red: true positive, green: false positive, gray: true negative, yellow: false negative). The second row shows de-snowed point clouds.



Fig. 7: Additional qualitative comparisons on a *medium snow* scene of the synthesized snow noise data. The first row shows all points with their prediction results (red: true positive, green: false positive, gray: true negative, yellow: false negative). The second row shows de-snowed point clouds.

SLiDE: Self-supervised LiDAR De-snowing through Reconstruction Difficulty

7



Fig. 8: Additional qualitative comparisons on a *heavy snow* scene of the synthesized snow noise data. The first row shows all points with their prediction results (red: true positive, green: false positive, gray: true negative, yellow: false negative). The second row shows de-snowed point clouds.



Fig. 9: Additional qualitative comparisons on a *extreme snow* scene of the synthesized snow noise data. The first row shows all points with their prediction results (red: true positive, green: false positive, gray: true negative, yellow: false negative). The second row shows de-snowed point clouds.

8 G. Bae et al.



Fig. 10: Additional qualitative comparisons on the real-world snowy weather data (red: positive, gray and blue: negative).



Fig. 11: Additional qualitative comparisons on the real-world snowy weather data (red: positive, gray and blue: negative).

References

- Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., Heide, F.: Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 11682–11692 (2020)
- Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 11621–11631 (2020)
- 3. Charron, N., Phillips, S., Waslander, S.L.: De-noising of lidar point clouds corrupted by snowfall. In: Conference on Computer and Robot Vision. pp. 254–261 (2018)
- Heinzler, R., Piewak, F., Schindler, P., Stork, W.: Cnn-based lidar point cloud denoising in adverse weather. IEEE Robotics and Automation Letters 5(2), 2514–2521 (2020)
- Pitropov, M., Garcia, D.E., Rebello, J., Smart, M., Wang, C., Czarnecki, K., Waslander, S.: Canadian adverse driving conditions dataset. The International Journal of Robotics Research (2020)