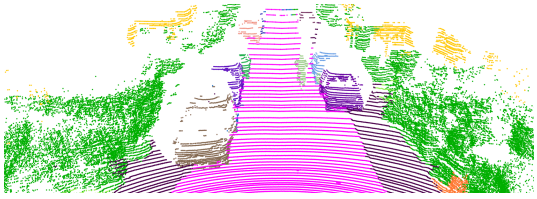
2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

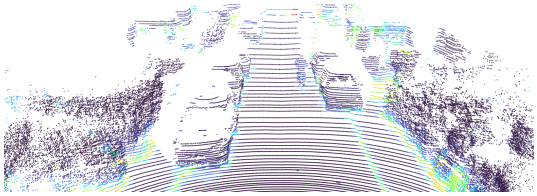
Uncertainty-aware LiDAR Panoptic Segmentation

Kshitij Sirohi[1], Sajad Marvi[1], Daniel Büscher[1] and Wolfram Burgard[2]

***Abstract*— Modern autonomous systems often rely on LiDAR scanners, in particular for autonomous driving scenarios. In this context, reliable scene understanding is indispensable. Conventional learning-based methods generally try to achieve maximum performance for this task, while neglecting a proper estimation of the associated uncertainties. In this work, we introduce a novel approach for solving the task of uncertainty-aware panoptic segmentation using LiDAR point clouds. Our proposed EvLPSNet network is the first to solve this task efficiently in a sampling-free manner. It aims to predict per-point semantic and instance segmentations, together with per-point uncertainty estimates. Moreover, it incorporates methods that utilize the uncertainties to improve the segmentation performance. We provide several strong baselines combining state-of-the-art LiDAR panoptic segmentation networks with sampling-free uncertainty estimation techniques. Extensive evaluations show that we achieve the best performance on uncertainty-aware panoptic segmentation quality and calibration compared to these baselines. We make our code available at: `https://github.com/kshitij3112/EvLPSNet`**



(a) LiDAR panoptic segmentation



(b) LiDAR panoptic uncertainties

Fig. 1: Panoptic segmentation and associated uncertainties as predicted by our EvLPSNet for the SemanticKITTI validation dataset.

I. Introduction

A perception system capable of providing comprehensive and reliable scene understanding is crucial for the safe operation of an autonomous vehicle. The recently introduced panoptic segmentation [1] combines the semantic segmentation of *stuff* and instance segmentation of *thing* classes into a single task. This facilitates the evaluation of the overall accuracy, which is crucial for a holistic scene understanding. In practice, however, the performance can only be evaluated on a limited dataset, while the real-world consists of scenarios and objects possibly not present in the dataset. Therefore, in addition to an evidence signal, a reliable uncertainty estimate is crucial for safety-critical applications, such as autonomous driving. Hence, the task of uncertainty-aware panoptic segmentation [2] for a unified evaluation of panoptic segmentation and uncertainty estimation offers a better potential for deployment. Our method aims to solve this task for LiDAR point-clouds, as illustrated in Fig. 1.

The regular grid structure of images allows a number of works on the panoptic segmentation to take advantage of recent advances in deep learning, in particular using convolutional neural networks (CNNs) [3], [4]. On the other hand, the irregular, sparse and unordered structure of LiDAR point clouds posed unique challenges. However, LiDARs provide an illumination-independent accurate geometric description of the environment, yielding a great advantage over images. This motivated recent works for panoptic segmentation of LiDAR point clouds, represented in various ways, such as range images [5], [6], [7], 3D voxels [8], birds-eye-views (BEVs) [9], or direct points [10]. These methods are generally classified into proposal-based [5] and proposal-free [11].

Conventional CNN-based methods, utilizing the softmax operation, typically show overconfidence in their predictions [12]. On the other hand, popular sampling-based methods for uncertainty estimate methods, such as Monte Carlo dropout [13], and Bayesian neural networks (BNNs) [14], are time and memory intense hence not suitable for real-time applications. Therefore, there is a recent interest in sampling-free methods for uncertainty estimation, such as evidential deep learning [12], predicting uncertainties in a single pass. However, most of these works for classification or segmentation deal with the image domain. Hence, to the best of our knowledge, there is still no existing approach to provide sampling-free point-wise uncertainty estimates for the panoptic segmentation of LiDAR point clouds.

In this work, we present the novel Evidential LiDAR Panoptic Segmentation Network (EvLPSNet), the first network to tackle this task, by utilizing evidential deep learning. We use the 2D polar BEV grid representation [11] for our network, facilitating fast inference times and better separability of instances. However, the projection into the grid structure leads to discretization errors, as all points in a grid cell are assigned the same prediction. We approach this issue using the 3D point information, as well as our uncertainty estimates, proposing a novel learnable uncertainty-based Query and Refinement (uQR) module. This module employs a simple point-based convolution layer to achieve point-wise

[1]Department of Computer Science at University of Freiburg, Germany. This work was financed by the Baden-Württemberg Stiftung gGmbH. [2]Department of Engineering at Technical University Nürnberg, Germany.

predictions for points selected based on their uncertainty. We also propose to utilize the predicted probabilities to create an efficient version of the k nearest neighbors algorithm (pKNN). Furthermore, we provide several baselines and evaluate their results on the task of uncertainty-aware LiDAR panoptic segmentation. In summary, our contributions are as follows:

- The novel proposal-free EvLPSNet architecture for uncertainty-aware LiDAR panoptic segmentation.
- The uQR module for refining the prediction for the most uncertain points.
- The efficient pKNN algorithm utilizing the predicted class probabilities.
- Several baselines for comparison with EvLPSNet.

II. Related Work

A. Segmentation of LiDAR Point Clouds

The release of the SemanticKITTI dataset [15] led to the emergence of many works, initially for the semantic segmentation of LiDAR point clouds. These can generally be classified based on the point cloud representations they employ, such as projected range images [16], [17], [18], 3D voxels [19], point-based [10], and BEV polar coordinates [9]. Most panoptic approaches utilize these representations as well.

Panoptic segmentation approaches can be classified as proposal-based and proposal-free. While both employ separate semantic and instance segmentation branches, the distinction lies in the latter. Proposal-based methods typically employ bounding box regression for discovering instances, such as Mask-RCNN [20] in the case of EfficientLPS [5]. On the other hand, proposal-free approaches perform clustering on the semantic prediction to obtain instance ids for objects belonging to separate instances. Panoptic-PolarNet [11] utilizes a Panoptic Deeplab-based [4] instance head to regress offsets and centers for different instances. DS-Net [21] proposes a dynamic shifting module to move instance points towards their respective center. Panoptic-PHNet [22] utilizes two different encoders, BEV and voxel-based, to encode point cloud features, followed by a KNN-transformer module to model interaction among voxels belonging to thing classes.

B. Uncertainty Estimation

Many works for estimating uncertainty in segmentation tasks employ sampling-based methods, such as Bayesian Neural Networks [14] or Monte Carlo dropout [13], [23]. However, such methods are time and memory-intensive, requiring multiple passes or sampling operations. For LiDAR point clouds, SalsaNext [17] is an uncertainty-aware semantic segmentation utilizing BNNs. Even though the network output is quick to evaluate, due to the sampling of the BNN approach the uncertainty is slow to obtain. Further, no metric is presented to quantify the calibration of the predicted uncertainty for this approach. We believe these are severe limitations for safety-critical real-time applications like autonomous driving.

The need for single-pass sampling-free uncertainty estimation motivates many works in the field [24], [25], [26]. Classical neural networks utilize softmax operations of the final logits to predict per class score or probability, which is not a reliable estimate of the network's confidence in the prediction, as shown by [12]. Guo et al. [26] propose the Temperature Scaling (TS) method to learn a logit scaling factor on the softmax operation to provide calibrated probability predictions. Other methods, such as [27], learn to separate different classes in a latent space and, based on the distance of the predicted to the nearest class feature, calculate the uncertainty.

Sensoy et al. [12] proposed evidential deep learning to provide reliable and fast uncertainty estimation with minimal changes to a network. Petek et al. [28] utilize this method to simultaneously predict semantic segmentation and bounding box regression uncertainty. Sirohi et al. [2] introduce the uncertainty-aware panoptic segmentation task and provide a sampling-free network for a unified panoptic segmentation and uncertainty for images. In our present work, we build upon this to extend the approach to LiDAR point-clouds and we provide a comprehensive quantitative analysis.

III. Technical Approach

Fig. 2 presents an overview of our network architecture. It is based on the proposal-free Panoptic-PolarNet network [11]. Our evidential semantic segmentation head and Panoptic-Deeplab based [4] instance segmentation head utilize the learned features to predict per-point semantic segmentation, semantic uncertainty, instance center and offsets. The predictions from both heads are fused to provide panoptic segmentation results. Leveraging the segmentation uncertainties, our proposed query and refine module helps to improve the prediction for points within uncertain voxels. Moreover, post-processing using our efficient probability-based KNN improves the results further.

A. Network Architecture

We project the LiDAR points into a polar BEV grid utilizing the encoder design proposed by PolarNet [9]. First, the points (represented in 3D polar coordinates) are grouped according to their location in a 2D polar BEV grid. The grid has the dimensions of $H \times W = 480 \times 360$, where H corresponds to the range and W to the heading angle. Then, for each grid cell, the corresponding points are encoded using a simplified PointNet [29]. This is followed by a max pooling operation to calculate the feature vector for every 2D grid cell and to create a fixed-size grid representation of $W \times H \times F$, where $F = 512$ is our number of feature channels.

The subsequent encoder-decoder network utilizes the U-net [30] architecture. The semantic and instance segmentation branches share its first three decoder layers, while the remaining layers are separate. The instance segmentation regresses the instance center heatmap and the instance offsets on the BEV grid.

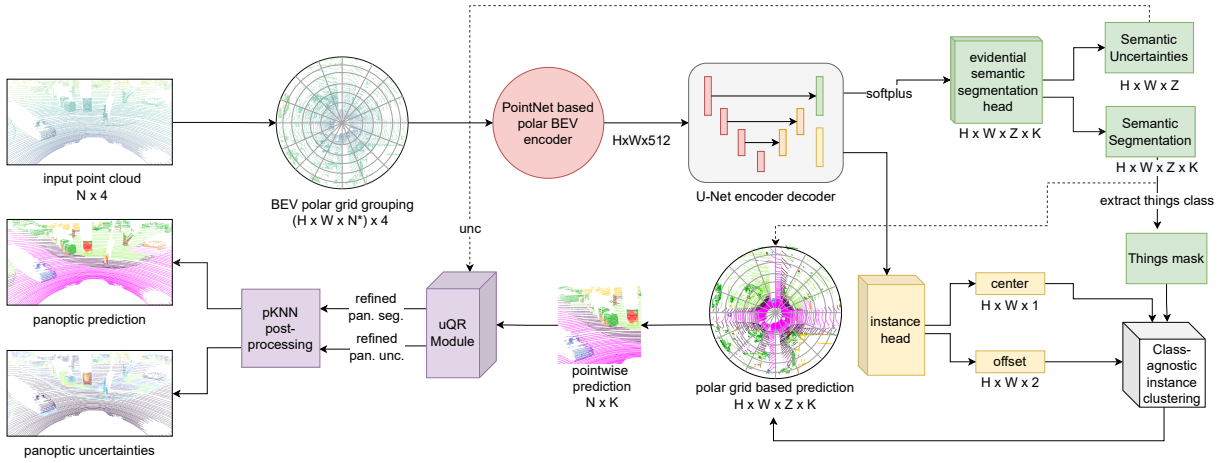


Fig. 2: Overview of our EvLPSNet architecture.

B. Evidential Semantic Segmentation

We employ evidential learning [12] for providing voxel-level semantic segmentation with calibrated uncertainty estimation. Here, the voxels are in polar coordinates with dimension $H \times W \times Z$, where $Z = 32$ corresponds the vertical segmentation of the BEV grid. Our network is based on Panoptic-PolarNet [11].

To incorporate the uncertainty estimate, we first add a softplus activation function on the network's final logits. This works as the evidence signal, essentially signifying the evidence or weight collected for a specific class. For the per-point multinomial classification, instead of treating the per class prediction as a single-value estimate, we employ the Dirichlet distribution [12]. For each point i, the parameters of the Dirichlet distribution are $\alpha = [\alpha^1, ..., \alpha^K]$. Here, K is the number of classes and $\alpha_i^k = \text{softplus}(l_i^k) + 1$ for class k and logit output l_i^k. As in [2] we calculate the associated probabilities p_i and uncertainties u_i as:

$$p_i^k = \alpha_i^k / S_i \tag{1}$$

$$u_i = K / S_i, \tag{2}$$

with $S_i = \sum_{k=1}^{K} \alpha_i^k$.

Our semantic segmentation head is trained using the type-II maximum likelihood variant of the loss and the KL term, provided by [12], to regularize the evidence magnitude in order to predict high uncertainty for wrongly predicted classes. The loss is given by:

$$\mathcal{L}_{\text{sem}} = \mathcal{L}_{\log}^{\text{s}} + \lambda_t \mathcal{L}_{\text{KL}}^{\text{s}}, \tag{3}$$

During training we increase λ_t first linearly and then keep it constant by $\lambda_t = 0.065 \times \min\{1, t/(20I)\}$. Here, I is the number of iterations per epoch and t is the current iteration. The log loss is given as:

$$\mathcal{L}_{\log}^{\text{s}} = \sum_{i=1}^{N} \sum_{k=1}^{K} o_i^k \log(S_i / \alpha_i^c). \tag{4}$$

Here, the total number of voxels is $N = W \cdot H \cdot Z$, and the one-hot encoded vector o is unity for the ground truth class and zero otherwise.

In our experiments, we found that the performance stagnates after some point due to a high number of empty voxels. On the other hand, if we only use the occupied voxels (N = number of occupied voxels in Eq. (4)), the uncertainty estimation results are not calibrated. Hence, we first train the network with all voxels, and then, after performance convergence, we train for some epochs with only the occupied voxels. This improves both the segmentation performance and uncertainty calibration.

C. Instance Segmentation

Similar to [11], we base our instance segmentation head on Panoptic-Deeplab [4]. The instance segmentation head consists of separate center prediction and offset prediction heads. The former predicts the likelihood of each grid cell being the center of an instance, while the latter predicts a 2D offset for each grid cell to its center in polar coordinates. We encode the ground truth heatmap as a 2D Gaussian around the center of each instance.

Note that the instance predictions operate in the 2D BEV domain, which allows the application of well-researched and fast 2D convolution operations. Another benefit of the BEV is the easy separation of objects that are close/overlapping in the heading or elevation angles, which is not the case in the range projection.

D. Panoptic Segmentation

To obtain the panoptic segmentation we are following [11], by first extracting the top k centers of the instance segmentation after applying non-maximum suppression. Next, we utilize the semantic segmentation to create a foreground mask, where at least one of the Z predictions in a BEV cell belongs to a thing class. We group the objects and assign an instance id I in the foreground mask based on their distance to the nearest of the k centers. Then, we assign the instance id I to the thing class predictions in the semantic segmentation output. Finally, we assign the instance class label based on a majority voting [11] utilizing the evidential probabilities

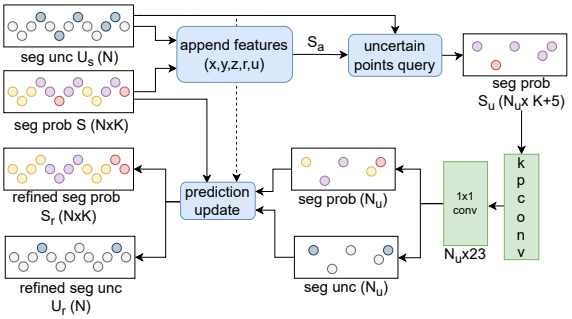


Fig. 3: Our uncertainty-based query and refinement module (uQR).

within the same instance group. All the points belonging to the stuff class get their class label from the semantic segmentation.

E. Uncertainty-based Query and Refinement module

Our uncertainty-based query and refinement module (uQR) leverages the predicted uncertainties to counter the discretization errors due to the BEV grid structure. We select the top $N = 20$k most uncertain points and pass them to our uQR module to actively improve the segmentation quality in an efficient way. An overview of the module is shown in Fig. 3. While the semantic segmentation head makes voxel-wise predictions (which are simply transferred to the corresponding points), the uQR module does refined point-wise predictions.

For N points and K classes, let $S \in \mathbb{R}^{N\times K}$ be the semantic segmentation probabilities calculated using Eq. (1) and $U_s \in \mathbb{R}^N$ be the associated uncertainties from Eq. (2). First, we append each point by its x, y, z coordinates, remission value, and uncertainty value to obtain $S_a \in \mathbb{R}^{N\times K+5}$ with a feature size of $K + 5$. Then we create a subset $S_u \in \mathbb{R}^{N_u\times K+5}$ of the N_u most uncertain points and pass these to the KPconv-based network [10] for refining. KPconv utilizes point-based convolution in 3D space by capturing contextual information from its neighbors. We only utilize one KPConv layer, followed by ReLU and a final classifier layer to obtain the refined predictions and uncertainties as $S_r \in \mathbb{R}^{N\times K}$ and $U_r \in \mathbb{R}^N$.

F. Efficient probability-based KNN

In this section, we devise our efficient probability-based k nearest neighbors (pKNN) approach. Post-processing methods based on KNNs can be employed to cluster the instances or improve the segmentation quality. However, the speed of execution is a typical limitation of KNNs. We try to decrease the execution time by limiting the number of points requiring post-processing by taking advantage of the predicted probabilities. We first select points with a probability (Eq. (1)) below a certain threshold. Then we find its k nearest neighbors in the whole point cloud, followed by a majority voting to decide the final label. We choose $k = 5$ and a probability threshold of 0.4. If a label is transferred from a neighboring point, then the corresponding uncertainty is transferred as well.

IV. Experimental Evaluation

Our network's performance is evaluated on the challenging SemanticKITTI [31] dataset. The dataset consists of 43,551 LiDAR scans distributed over 21 driving sequences. Sequences 00 to 10 are used for training, except the sequence 08 which is reserved for validation, and sequences 11 to 21 are for testing. The dataset provides point-wise annotation for 20 classes, out of which 8 are *thing* classes and contain unique instance ids. As the test set annotations are not provided, we evaluate the uncertainty-aware panoptic segmentation on the validation set.

A. Baselines

We aim to tackle the task of uncertainty-aware LiDAR panoptic segmentation for autonomous driving scenarios. Hence, the uncertainty estimation method should be sampling-free and not create extra computation overhead. The temperature scaling (TS) [26] and evidential learning (Ev) [12] qualifies both criteria. For panoptic segmentation, we choose the proposal-based EfficientLPS [5] and the proposal-free Panoptic-PolarNet [11]. We train these networks to the best setting provided by the official codes. However, for a fair comparison we do not use pseudo labels for EfficientLPS. First, we evaluate both original networks without any uncertainty estimation method involved. Then, for temperature scaling, we add a scaling parameter to the semantic segmentation logits, freeze the networks and train the scaling parameter until it converges on the validation set, as suggested by the original authors. Finally, we train our network utilizing evidential learning, our proposed uncertainty-based query and refinement module (uQR), and our efficient probability-based KNN (pKNN) post-processing.

B. Training Procedure

We discretize the space into the grid size of $480 \times 360 \times 32$ polar voxels within the range $r \in [3, 50]$ m and height $z \in [-3, 1.5]$ m w.r.t. the LiDAR scanner. For non-maximum suppression, we use kernel size of 5 with 0.1 threshold and select top $k = 100$ centers similar to [11]. We train the network for 50 epochs on single NVIDIA TITAN RTX GPU with a batch size of 3. We use Adam optimizer with step learning rate with an initial value of 0.01 and a drop by a factor of 10 at epoch 40 and 45.

We apply instance augmentation and random data flipping along the x and y axes as suggested by [11]. However, we do not utilize their Self Adversarial Pruning (SAP) since it requires two forward passes per iteration, slowing down the training significantly. In contrast, we utilize Lovász Evidential loss [2] ($\mathcal{L}_{\text{lev}}$) for the last five epochs, which significantly improves the performance. Further, we employ the evidential loss for semantic segmentation $\mathcal{L}_{\text{sem}}$ from Eq. (4), MSE loss ($\mathcal{L}_{\text{h}}$) for the center heatmap and L1 loss ($\mathcal{L}_{\text{o}}$) for the offset. The overall loss is

$$\mathcal{L} = \mathcal{L}_{\text{sem}} + \lambda_{\text{h}}\mathcal{L}_{\text{h}} + \lambda_{\text{o}}\mathcal{L}_{\text{o}} + \lambda_{\text{lev}}\mathcal{L}_{\text{lev}}, \tag{5}$$

where $\lambda_{\text{h}} = 100$, $\lambda_{\text{o}} = 10$ and $\lambda_{\text{lev}} = 1$ for the last five epochs and $\lambda_{\text{lev}} = 0$ otherwise. After the main network, we

train the uQR module with $\mathcal{L}_{\text{sem}} + \mathcal{L}_{\text{lev}}$ for 15 epochs with the Adam optimizer and a learning rate of 0.0001.

C. Metrics

We employ the uncertainty-aware Panoptic Quality (uPQ) and panoptic Expected Calibration Error (pECE) metrics, as proposed in [2], to evaluate the performance of our network. The metrics were utilized for evaluating performance for images where each instance consists of many densely organized pixels. However, LiDAR point clouds are sparse, and separate instances generally contain fewer points, which can lead to biases in the metrics. Hence, we adjusted the metrics to fit the sparse LiDAR points setting.

First, we search for the unique matching pairs between the ground truth and the prediction having IoU > 0.5. For each matching pair we calculate the average accuracy (acc) and confidence (conf). Similar to [2], we calculate the confidence as $\text{conf}_i = 1 - u_i$, where u_i is the predicted uncertainty for point i. Further, we define $\text{acc} = 1$ if the predicted class and instance id matches the ground truth and $\text{acc} = 0$ otherwise. Then, we average the accuracy in $J = 10$ bins of confidence (from 0 to 1) for each class separately over the full dataset. This is opposed to [2], where this calculation was done separately for each instance. Finally, we calculate uECE_k for each class k, pECE and uPQ as:

$$\text{uECE}_k = \sum_{j=1}^{J} \frac{|B_j|}{N} \left|\text{acc}(B_j) - \text{conf}(B_j)\right| \tag{6}$$

$$\text{pECE} = \frac{1}{K} \sum_{k} \text{uECE}_k \tag{7}$$

$$\text{uPQ} = (1 - \text{pECE})\text{PQ}. \tag{8}$$

where $|B_j|$ is the number of points in bin j, and $\text{acc}(B_j)$ and $\text{conf}(B_j)$ are the average accuracy and confidence for bin B_j. In addition, we provide separate results for the *thing* and *stuff* classes. Further, we provide the mean Intersection over Union (mIoU) and the semantic uECE to evaluate the semantic segmentation and uncertainty estimation performance.

D. Quantitative Results

Tab. I presents the experimental results on the SemanticKITTI validation dataset. Our proposed EvLPSNet achieves the best score for the overall uPQ, signifying its superior performance for the uncertainty-aware panoptic segmentation task, as well for the pECE and uECE, signifying the most accurate uncertainty calibration of all methods. Interestingly, our network is also superior to the Panoptic-PolarNet in PQ^{st} and mIoU, which we attribute to our proposed uQR and pKNN modules for refining segmentation results utilizing the uncertainties and probabilities, see Sec. IV-F. Comparing the overall PQ, our approach is slightly worse than Panoptic-PolarNet. However, given its well calibrated uncertainty estimation, our network recovers the performance on uPQ. We observe mixed results for temperature scaling (TS). It generally improves uECE, but pECE is only improved for EfficientLPS, while it is worsened for Panoptic-PolarNet. This is similar to [2], where TS was able to improve uECE, but the advantage almost vanishes for the panoptic pECE.

E. Qualitative Results

Fig. 4 presents qualitative results: the predicted panoptic segmentation and uncertainties, and the corresponding error maps. For example, in Fig. 4a, we depict a misclassified vehicle (marked by the box). We observe that the wrongly classified points (see error map) are well represented with high predicted uncertainties in the network output. Similarly, in Fig. 4b, a sidewalk (dark magenta) is wrongly predicted as drivable road (light pink), but high uncertainties are predicted here as well. Generally, the uncertainty prediction is strongly correlated with the error map, validating the prediction quality.

F. Ablation Studies

1) Refinement Modules: We present the quantitative analysis of our proposed uQR and pKNN modules in Tab. II for the overall performance and for selected classes. The motivation behind both modules is to utilize uncertainties to improve the segmentation performance, specifically of points suffering from discretization errors. This affects in particular smaller objects that consist of fewer points, such as persons, poles, and traffic signs.

Model M1 leaves out the pKNN and uQR modules compared to our final approach. In model M2, we incorporate the pKNN module, which leads to a gain in PQ for all mentioned classes, with the most gain for the person class. The overall uPQ, PQ and mIoU are slightly improved as well, however, pECE and uECE are slightly worsened. Similarly, in model M3 we incorporate the uQR module, improving PQ and mIoU, but slightly worsening uPQ, pECE and uECE. Both results, for M2 and M3, signify a loss in calibration quality, but a gain in segmentation performance. Model M4 is our final model, incorporating both the pKNN and uQR modules. We observe that the contributions from both modules roughly add up, in particular a significant gain in PQ for most of the small classes can be seen.

We further present calibration curves for the models M1 and M4 in Fig. 5. It can be observed that the application of our refinement modules renders the network slightly overconfident. In conclusion, our final M4 model shows the best segmentation performance, however, our M1 model shows the most accurate uncertainty calibration.

2) Efficiency of pKNN: The efficiency of our pKNN module is showcased in Fig. 6. In our experiment, we increment the probability threshold that is used to select the points for applying the KNN post-processing. We report the increase in PQ and in runtime of the network. Here, we used a subset of the validation set. The results compare to a base runtime of about 0.09 seconds per frame without pKNN. We observe that the PQ and runtime increase with the threshold, which is expected due to an increase in the number of selected points. However, the PQ saturates for a threshold of about 0.4, our final value. For a faster evaluation one can lower the threshold to e.g. 0.2 for a modest performance gain with low increase in runtime.

TABLE I: Performance values in % on the SemanticKITTI validation set. Lower values are better for ↓, and larger values otherwise.

Method	uPQ	PQ	pECE ↓	uPQ^{Th}	PQ^{Th}	$pECE^{Th}$ ↓	uPQ^{St}	PQ^{St}	$pECE^{St}$ ↓	uECE ↓	mIOU
EfficientLPS	48.7	57.1	14.6	52.6	59.9	**12.1**	45.9	**54.9**	16.4	16.5	62.3
EfficientLPS + TS	49.6	56.9	12.9	48.5	59.8	18.8	50.2	54.9	8.7	7.7	62.1
Panoptic-PolarNet	48.8	**58.5**	16.6	**53.1**	**65.7**	19.1	45.4	53.3	14.7	13.2	63.2
Panoptic-PolarNet + TS	48.1	**58.5**	17.7	51.4	65.7	21.8	45.4	53.3	14.7	10.5	63.3
EvLPSNet	**51.4**	58.0	**11.5**	52.7	62.7	15.9	**50.1**	54.6	**8.2**	**7.1**	**64.0**

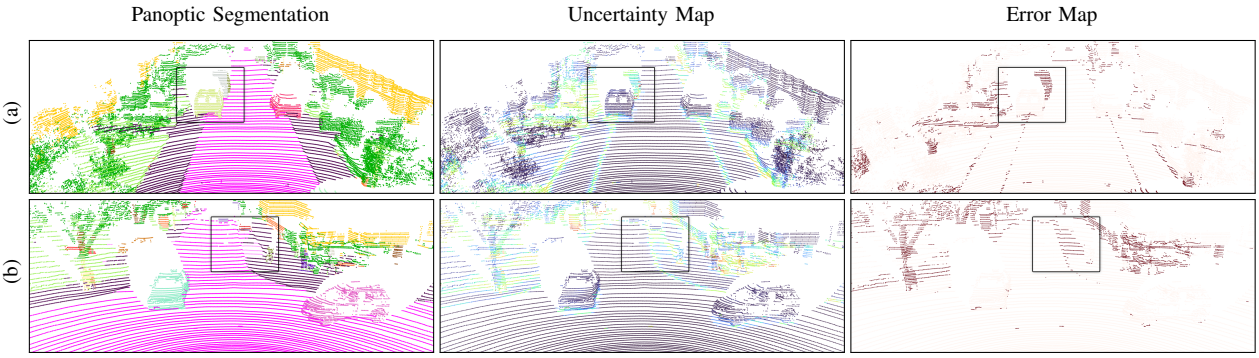


Fig. 4: Qualitative results for the uncertainty-aware panoptic segmentation by our EvLPSNet for two scans of the SemanticKITTI validation set. Red regions in the uncertainty map depict high predicted uncertainty, and dark regions in the error map depict misclassified points.

TABLE II: Class-wise PQ values in % on the SemanticKITTI validation set. Lower values are better for ↓, and larger values otherwise.

Model	pKNN	uQR	car	bicycle	m.cycle	person	bicyclist	m.cyclist	trunk	pole	tr. sign	uPQ	PQ	pECE ↓	uECE ↓	mIoU
M1	✗	✗	87.9	51.8	59.0	55.0	83.5	44.1	45.5	55.9	48.6	51.4	57.3	**10.4**	**2.7**	63.5
M2	✓	✗	88.0	52.1	60.0	57.3	84.1	**44.7**	45.9	56.0	49.0	**51.5**	57.7	10.6	2.8	63.7
M3	✗	✓	**88.1**	53.3	59.6	55.3	84.1	44.2	47.8	**56.3**	49.9	51.2	57.8	11.3	3.9	63.9
M4	✓	✓	**88.1**	**53.4**	**60.5**	**57.5**	**84.6**	**44.7**	**47.9**	56.2	**49.9**	51.4	**58.0**	11.5	7.1	**64.0**

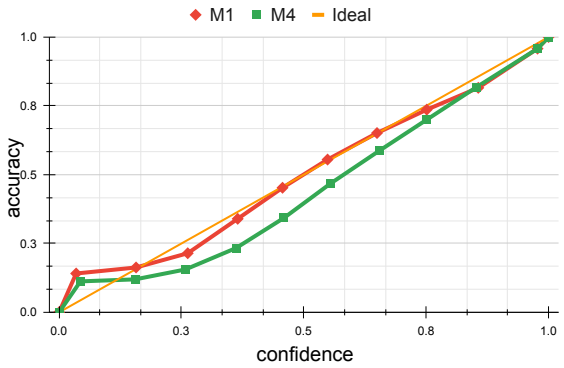


Fig. 5: Calibration curves for our M1 and M4 models.

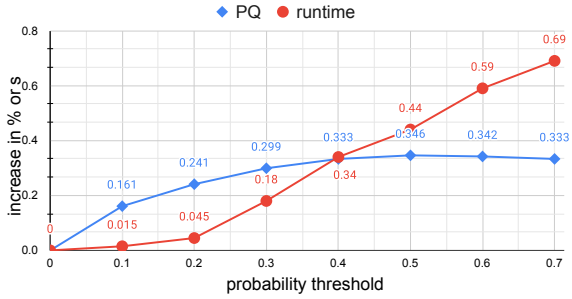


Fig. 6: PQ gain and runtime versus pKNN probability threshold.

V. Conclusions

In this work we proposed EvLPSNet, a novel proposal-free approach for solving the task of uncertainty-aware LiDAR panoptic segmentation. It is the first network to simultaneously predict panoptic segmentation and uncertainties of LiDAR point clouds in a single forward pass. To this end, we demonstrated an effective way of utilizing evidential deep learning for our uncertainty-aware semantic segmentation head. We further proposed an uncertainty-based query and refining (uQR) module to leverage and improve the segmentation of points that suffer from discretization errors. Moreover, our pKNN module showcased how probabilities can be beneficial to reduce the runtime of KNN clustering methods, while maintaining the gain in performance. Our network achieves the best performance on the uncertainty-aware panoptic segmentation performance and calibration metrics, uPQ and pECE, respectively. We hope our work will motivate future works in holistic and reliable 3D scene understanding using LiDAR point clouds.

REFERENCES

[1] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9404–9413.

[2] K. Sirohi, S. Marvi, D. Büscher, and W. Burgard, "Uncertainty-aware panoptic segmentation," *arXiv preprint arXiv:2206.14554*, 2022.

[3] R. Mohan and A. Valada, "Efficientps: Efficient panoptic segmentation," *International Journal of Computer Vision*, vol. 129, no. 5, pp. 1551–1579, 2021.

[4] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 475–12 485.

[5] K. Sirohi, R. Mohan, D. Büscher, W. Burgard, and A. Valada, "Efficientlps: Efficient lidar panoptic segmentation," *IEEE Transactions on Robotics*, 2021.

[6] A. Milioto, J. Behley, C. McCool, and C. Stachniss, "Lidar panoptic segmentation for autonomous driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 8505–8512.

[7] L. T. Triess, D. Peter, C. B. Rist, and J. M. Zöllner, "Scan-based semantic segmentation of lidar point clouds: An experimental study," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1116–1121.

[8] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[9] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "Polarnet: An improved grid representation for online lidar point clouds semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9601–9610.

[10] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411–6420.

[11] Z. Zhou, Y. Zhang, and H. Foroosh, "Panoptic-polarnet: Proposal-free lidar point cloud panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13 194–13 203.

[12] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *Advances in neural information processing systems*, vol. 31, 2018.

[13] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[14] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *arXiv preprint arXiv:1511.02680*, 2015.

[15] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[16] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet++: Fast and accurate lidar semantic segmentation," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 4213–4220.

[17] T. Cortinhal, G. Tzelepis, and E. Erdal Aksoy, "Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds," in *International Symposium on Visual Computing*. Springer, 2020, pp. 207–222.

[18] D. Kochanov, F. K. Nejadasl, and O. Booij, "Kprnet: Improving projection-based lidar semantic segmentation," *arXiv preprint arXiv:2007.12668*, 2020.

[19] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching efficient 3d architectures with sparse point-voxel convolution," in *European conference on computer vision*. Springer, 2020, pp. 685–702.

[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[21] Y. Zhao, X. Zhang, and X. Huang, "A divide-and-merge point cloud clustering algorithm for lidar panoptic segmentation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7029–7035.

[22] J. Li, X. He, Y. Wen, Y. Gao, X. Cheng, and D. Zhang, "Panoptic-phnet: Towards real-time and high-precision lidar panoptic segmentation via clustering pseudo heatmap," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 809–11 818.

[23] P.-Y. Huang, W.-T. Hsu, C.-Y. Chiu, T.-F. Wu, and M. Sun, "Efficient uncertainty estimation for semantic segmentation in videos," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 520–535.

[24] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal, "Deep deterministic uncertainty: A simple baseline," *arXiv e-prints*, pp. arXiv–2102, 2021.

[25] J. Postels, M. Segu, T. Sun, L. Van Gool, F. Yu, and F. Tombari, "On the practicality of deterministic epistemic uncertainty," *arXiv preprint arXiv:2107.00649*, 2021.

[26] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International conference on machine learning*. PMLR, 2017, pp. 1321–1330.

[27] Y. Li and J. Košecká, "Uncertainty aware proposal segmentation for unknown object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 241–250.

[28] K. Petek, K. Sirohi, D. Büscher, and W. Burgard, "Robust monocular localization in sparse hd maps leveraging multi-task uncertainty estimation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4163–4169.

[29] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[31] J. Behley, A. Milioto, and C. Stachniss, "A benchmark for lidar-based panoptic segmentation based on kitti," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 13 596–13 603.