Secondly, the sensor data can include many different sources of noise which is especially true when it comes to radar. In order for the model to learn all these effects,

Requirement 1.2 (R1.2): *The trained ISM must be capable to learn from big data.*

Third, the trained ISM should be obtained as resource efficient as possible. Therefore, the following requirement arises

Requirement 1.3 (R1.3): *Minimize the amount of manpower, work hours and equipment needed to create the trained ISM.*

Also, the data-driven ISM should be able to run in parallel with the already existing geo ISM on the hardware of production-ready vehicles in real-time. The sensor with the highest capture frequency in the NuScenes dataset is the lidar sensor with 20 Hz. To provide some room to perform other computations, it is proposed to aim for a 100 Hz inference time of the trained ISM. To emulate these hardware restrictions, the requirement can be formulated as follows

Requirement 1.4 (R1.4): *The trained ISM shall be executable with 100 Hz on a single core of a CPU (Intel Core Processor i7-10750H).*

For this work, the width and height of input and output grid maps shall be 128×128 cells for an area of 40 m $\times$ 40 m. This is an acceptable range for low speed scenarios like parking. Additionally, the resolution of 31.25 cm per cell is satisfactory for the trained ISM, as it is mainly used to enhance the geo ISM. Thus, leading to the following requirement

Requirement 1.5 (R1.5): *The input and output grid maps shall cover an area of* 40 *m* $\times$ 40 *m with* 128×128 *cells.*

Finally, on the theoretical side, the model shall estimate the evidential classes as defined in 2.1, leading to the following requirements

Requirement 1.6 (R1.6): *The predicted output should capture the amount of free, occupied and unknown information.*

Requirement 1.7 (R1.7): *The unknown mass should be an inverse measure for the overall information content capturing both uncertainty and lack of information.*

work which does not specifically scale the IDM based on range, angle and ego vehicle velocity.

The second problem revolves around the enrichment of free space. Here, the SotA proposes to define additional free space between the maximum absolute angles of the radar's FoV cone and its closest detection. This assumes high certainty of the sensor towards the edges of the FoV cone in order to make the implication of free space based on absence of detections. The remaining literature, however, does not support this assumption, which can be seen by the scaling factor G_φ in Eq. 2-14 which reduces the IDM's certainty towards the FoV's edges in angular direction. Thus, the need arises to provide a free space enrichment method which better suits the sensor certainty

Research Question 2 (RQ2): *Define a procedure to enrich the baseline IDM ray casting free space predictions in regions lacking detections in a way that strictly increases the mIoU between radar- and lidar-based occupancy maps.*

3.2.2 Research Needs for deep, evidential ISMs

Given the target and baseline model, the creation of trained ISMs can be studied in more detail. Here, the current SotA approaches all tackle the creation of trained ISMs by applying CNNs in the form of UNets with skip connections. While the interpretation of the input varies between it being an image or some kind of multi-channel BEV grid map, CNNs are an appropriate choice for they are designed to leverage spatial context from matrix-like data. Moreover, through breakthroughs like Dropout for regularization, BatchNorm for normalization and skip connections for conservation of information, current CNN models can be designed with increased number of stacked layers. This results in increased modeling capacities allowing them to capture the information from big amounts of data. Thus, the application of UNets as the de facto standard model already suffices R1.1 and R1.2.

Regarding the resource efficiency during inference, the literature only discloses run time information on GPUs. Therefore, a rough architecture search shall be conducted for UNets with skip connections and ResNet layers to answer the following question

Research Question 3 (RQ3): *How should the amount of filters of a UNet architecture be chosen to maximize performance while sufficing the run time requirement as stated in R1.4?*

Additionally, the majority of deep ISMs in the literature model the problem in the probabilistic framework which does not suffice R1.6. On the evidential side, the problem is either modeled as a three class classification or a regression task, both of which suffice

R1.6. However, to the best of the authors knowledge, non of the published methods model dynamic objects by distributing mass equally to the free and occupied class. Training on dynamic object targets disqualifies the standard classification approach, since the dynamic object targets are not represented as a one-hot encoding. On the other hand, regression problems can cope with continuous targets. Thus, the training of deep ISMs will be defined as a regression problem in this work. However, no prior work has provided a thorough comparison of deep ISM's capabilities to estimate evidential masses given different sensor inputs. Which is especially true for the dynamic class which is in most of the literature not modeled for occupancy mapping. Thus, the following research questions arise

Research Question 4 (RQ4): *To which extend, as measured by the normed confusion matrix (see Section 3.3.8), are the proposed deep ISMs capable to estimate the position of dynamic, free and occupied areas given radar, camera and lidar data respectively?*

Research Question 5 (RQ5): *To which extent, as measured by the normed confusion matrix (see Section 3.3.8), does the capability of the proposed deep ISMs to estimate the position of dynamic, free and occupied areas given camera and lidar data respectively change, when radar information is added?*

Regarding the encoding of radar detections for deep ISMs, the literature proposes to either encode the positions of a single timestep or use the temporally accumulated point cloud respectively projected into BEV. However, to the best of the author's knowledge, no investigation on how to encode the motion information of the detections has been conducted. Therefore, a rough investigation distinguishing three approaches shall be conducted which is formulated in the following research question

Research Question 6 (RQ6): *To which extent, as measured by the normed confusion matrix (see Section 3.3.8), does the capability of the proposed deep ISMs to estimate the position of dynamic, free and occupied areas change choosing the input to be BEV projected radar detections...*

- *...of a single timestep encoding the dynamic detections with half intensity?*
- *...accumulated over a certain time horizon only encoding the dynamic detections with half intensity of the latest timestep?*
- *...accumulated over a certain time horizon linearly reducing the intensity of dynamic detections over time starting at half intensity?*

Looking at the uncertainty estimation, neither the classification nor the regression approaches for deep evidential ISMs do explicitly handle occurring aleatoric uncertainties,

letting us arrive at the following hypothesis.

Hypothesis 1 (H1): *In case of occurring aleatoric uncertainty, the current state of the art deep ISMs distribute the mass evenly into the free and occupied class rather than shifting it to the unknown class.*

In case H1 holds, these models would, thus, lack the possibility to distinguish between dynamic objects (per definition of evidential occupancy mapping indicated by mass equally distributed into free and occupied class) and regions of high uncertainty. Additionally, the unknown class cannot be used as a measure of information content, since some of the uncertainty is distributed into the free and occupied classes. This behavior would violate the requirements R1.7, hence, raising the following research questions.

Research Question 7 (RQ7): *How can a deep, evidential ISM be defined to separate conflicting mass due to aleatoric uncertainty into the unknown class while leaving conflicting mass due to dynamic objects untouched.*

3.2.3 Research Needs for Usage of deep, evidential ISMs as Priors in Occupancy Mapping

With regards to occupancy mapping with deep ISMs, not much literature is available. To the best of the authors knowledge, the only instances of occupancy mapping with deep ISMs use the standard Bayes filtering approach which does not cover evidential models. Thus, the first step consists in analyzing the characteristics and identifying short comings when applying the proposed evidential, deep ISMs for occupancy mapping.

First, in contrast to geo ISMs which only provide estimates in regions directly affected by data, deep ISMs additionally perform interpolations in intermediate regions and even extrapolations in regions further away from data. To illustrate the issue arising from this behavior, consider the example depicted in Fig. 3-1. Here, a scenario is shown in which the ego vehicle only partially observes a wall to its left-hand side for the first two time steps. Based on the majority of observations captured in the training dataset, the network might tend to extrapolate the wall as rectangular. In the standard occupancy formulation, this information is treated as independent and, thus, accumulated. When the vehicle finally obtains measurements of the wall's contour in the former occluded area, the extrapolation might have already be accumulated to high certainty. Therefore, many estimates based on measurements of this area would have to be accumulated to correct the assigned training data bias. In a similar way, this effect can also lead to overwriting area with predictions close to data with later occurring extrapolations. This thought experiment leads to the following hypothesis.

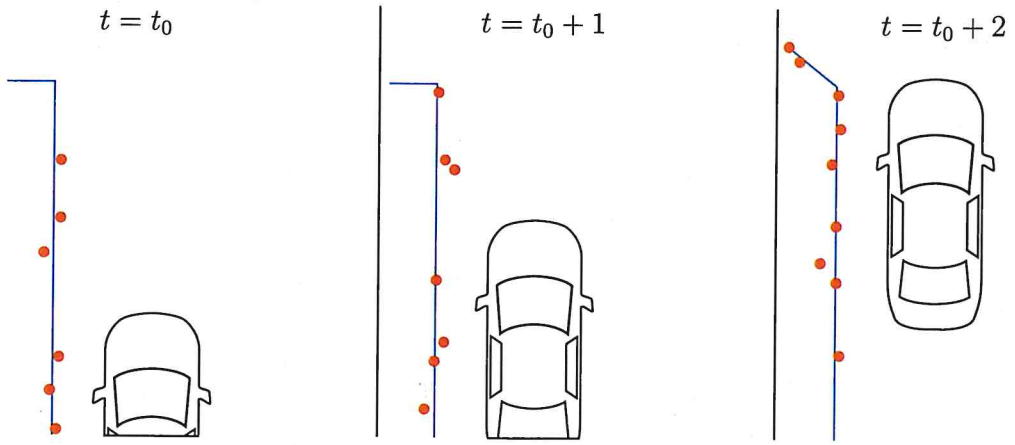


Fig. 3-1: Illustration of informational dependence between deep ISM predictions over time on the example of dataset bias. Here, the ego vehicle (black) drives along a wall and obtains radar measurements (orange) over three time steps. In each time step, the contour of the wall is estimated (blue).

Hypothesis 2 (H2): *Due to inter- and extrapolation in areas not directly measured, deep ISMs contain informational dependence between time steps. This leads to accumulation of bias and or falsification of previously correct assigned areas when their estimates are fused into the occupancy maps using a combination rule that assumes informational independence.*

In case H2 holds, the literature in Section 2.4.2 suggests to either remove the redundancy before combination or adapt the combination rule itself to account for the redundancy. This work will focus on removing the redundancy beforehand using Eq. 2-25 since the Yager and Dempster combination rule, as defined in Section 2.4.1, provide well studied, often used fusion methods for evidential occupancy mapping. To remove the redundancy, half of the approaches in the literature focus on defining a constant redundancy weighting between sensor modalities. This is, however, not applicable for the setup in this work, since there is only one sensor modality and the dependency is on the environment.

The other half proposes approaches to measure the amount of information in each to be fused mass and compare them using the mutual information, as stated in Eq. 2-30. However, no general procedure is proposed to measure the mutual information in signals. Thus, the following questions emerges

Research Question 8 (RQ8): *How can the mutual information be measured to asses the informational redundancy in evidential occupancy mapping?*

only dataset sufficing these requirements is the publicly available NuScenes dataset [CAE20], which can be seen in the dataset comparison provided in Table 1 in the paper.

The dataset is separated into 1000 so called scenes each containing the data of roughly a 20 second drive during which data from each modality is recorded, which is referred to as sensor sweeps. Additionally, so called samples are defined every 0.5 seconds containing annotations like bounding boxes and semantics for all sensor modalities. To enable comparability, the train-val-test split is predefined by the NuScenes creators.

Regarding the sensor setup, six cameras three of which face front left, center, right and three of which face back left, center, right are installed. Each camera captures 1.4MP images at a rate of 12 Hz. Additionally, five 77 GHz Frequency-Modulated Continuous-Wave (FMCW) radars are mounted to the car's front left, center, right and back left and right. To obtain ground-truth depth information, a 32 beam spinning lidar is mounted to the roofs center that captures frames with 20 Hz. The pose information is based on a fusion of lidar odometry, Global Positioning System (GPS) and Inertial Measurement Unit (IMU). For more details, the reader is referred to Table 2 in the paper.

3.3 Overview of Methodology

In this section, a framework is presented to address the research questions defined in Section 3.2.3. The framework extends the standard evidential occupancy mapping pipeline [PAG96] to incorporate estimates of a data-driven ISM, as shown in Fig. 3-2. The incorporation of the deep ISMs information is realized by fusing the estimates directly into the map rather than first fusing it with the geo ISM's estimate. This is done, in order to enable an asynchronous fusion of information into the map, which allows for differing execution times of the ISMs. In order to tackle the problem of temporal redundancy, as mentioned in H2, a specific fusion approach is defined for the deep ISM. This general procedure is detailed in the subsequent subsections as follows.

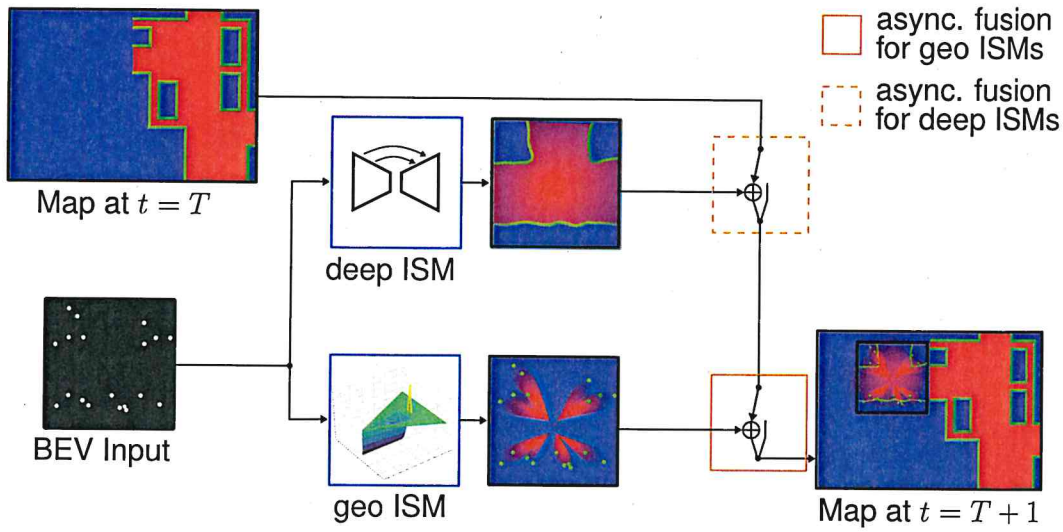


Fig. 3-2: Structural overview of the proposed framework, showing how the BEV input is transformed both by the deep and geo ISM into occupancy estimates. These estimates are then fused into the occupancy map using a specialized method to fuse the deep ISM that accounts for the temporal redundancy.

First, the creation of ground-truth lidar and baseline radar occupancy maps is discussed. Here, a method is shown in Section 3.3.1 to interpolate the annotations of dynamic objects for the higher frequency intermediate lidar detections. This is followed by the definition of the geo ILM and IRM used in this work in Section 3.3.2 and 3.3.3. Eventually, a method is described in 3.3.4 to maximize the overlap between the ground-truth lidar and baseline radar occupancy maps.

The definition of baseline and ground-truth data is followed by defining the deep ISM. Starting with the description of the different inputs and the target inspected in this work in Section 3.3.5, followed by the architecture search in Section 3.3.6, the investigated methods to account for aleatoric uncertainty in Section 3.3.7 and concluded by a description of the used metric in Section 3.3.8. The specific choice of the fusion methods for both geo and deep ISMs are detailed in Section 3.3.9 and 3.3.10.

3.3.1 Method to provide dynamic Information for Lidar Sweeps

To obtain the dynamic information for lidar sweeps, the sample's bounding boxes of dynamic objects are interpolated for intermediate sweeps and all lidar detections intersecting with the boxes are marked as dynamic. Here, the interpolations are obtained as follows. First, corresponding bounding boxes of dynamic objects are identified by their track ids for two subsequent samples. To avoid singularities, the bounding box poses of each identified pair are first transformed into the temporally first bounding box's coordinate frame. Next, a third degree polynomial is used to interpolate the poses. The polynomial is defined to intersects with the positions of the provided poses.

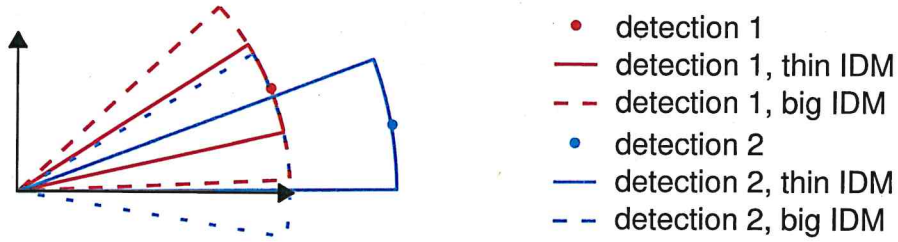


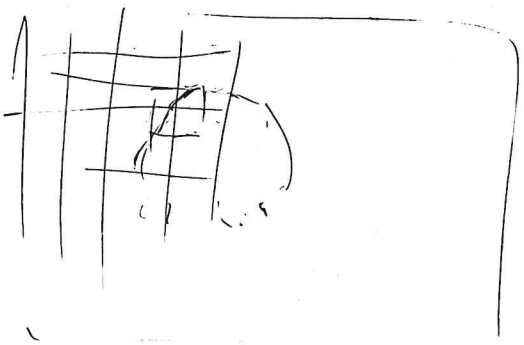
Fig. 3-4: Illustration of the proposed IDM's contours being applied on two detections. It shows the big and thing IDMs being cast for each detection. Also, it shows that the big IDM cone's range for detection two (stripped blue) is limited by the first detection. On the other hand, the thin IDM of the second detection is, for the given scenario, able to assign free space up to the second detection.

Another problem connected with the sparsity is the low coverage of free space. In this work, the method proposed in [PRO18] is altered to find an answer for RQ2 by instead casting the original rays again but with a much wider opening angle. This assumes that the manufacturer's prefiltering of the raw radar signal is done in away to preserve detections belonging to the object closest to the sensor. The justification for this assumption is based on the fact that radars are applied in safety critical collision avoidance systems. The second iteration of wider IDM rays shall only be used to enrich the free space and, thus, use $M_O = 0$. Additionally, it shall be fused with the first iterations IRM using Dempster's combination rule. An algorithmic summary of the geo IRM in pseudo code is shown in Alg. 2.

Algorithm 2: Geo IRM

Hyperparameters: $M_F{}^{(\text{thin})}, M_F{}^{(\text{big})}, \varphi_{\sphericalangle}{}^{(\text{thin})}, \varphi_{\sphericalangle}{}^{(\text{big})}, M_O, M_D, T$

```
function castFreeSpace
    // see Eq. 3-1 with M_O = M_D = 0
initialize all cells in the IRM image to m_u = 1
transform previous T radar point clouds of all radars to current IRM image
for (r_det, φ_det) in currentDetections do
    // range at which the free space of the ray is stopped
    (r_det, φ_det) = getClosestDetectionInsideConeArea(φ_center, φ_◁, r_max)
    // cast thin and big free space rays
    castFreeSpaceRay(φ_center, r_det, φ_◁^(thin), M_F^(thin))
    DempsterRule(castFreeSpaceRay(φ_center, r_det, φ_◁^(big), M_F^(big)))
    if (detection is static) then
        assignCellValue((r_det, φ_det), [0, M_O, 1 − M_O])
    else
        assignCellValue((r_det, φ_det), [M_D, M_D, 1 − 2M_D])
```



4 Deep ISM Experiments

As explained in Section 3.2.4, the experiments in this work will be conducted based on the NuScenes dataset. Here, the train-val-test split as proposed in NuScenes is used, while some of the scenes have been removed. Specifically, all scenes tagged with "night" or "difficult lighting" have been filtered out since they are relatively rare and thus, the networks with camera inputs could not properly adapt. Additionally, some scenes contain little to no ego vehicle movement (e.g. ego vehicle waiting at a red traffic light) which are great scenarios for tracking tasks but largely violate the static environment assumption in occupancy mapping. Thus, only scenes in which the ego vehicle moved at least 20 m are considered.

To obtain denser measurements for occupancy mapping, the sweeps are used to create the occupancy mapping dataset. Here, the sensor modality with the fewest sweeps per scene is identified and chosen as reference. Next, the temporally closest sweeps of the remaining sensors towards the reference are processed. Afterwards, sensor-dependent procedures are applied to create the different baselines, inputs and targets for the investigated geo and deep ISMs in form of a 128×128 grid map centered around the hind axle of the ego vehicle and spanning an area of 40×40 m^2.

4.1 Parameterization of geo ILM and IRM

This section details the comparison of different approaches to adapt lidar to radar BEV information. First, the sensor characteristics in the chosen dataset will be listed together with the applied methods to adapt the lidar. Afterwards, the different lidar filtering approaches will be evaluated based on the overlap between the lidar and radar maps in the mapped areas quantified by the mIoU score.

4.1.1 Experimental Setup

In the following, evidential occupancy maps of scenes as defined in the NuScenes dataset are created based on the geo IRM and ILM as defined in Section 3.3.2 and 3.3.3. Because of the large space of parameter configurations, the search is only conducted on the first 10% of the available training scenes. To fix the parameters and identify the best performing ISM variant, a two stepped approach is proposed.

The first steps consists of estimating the best parameters for the geo ILM and IRM. Since no additional occupancy ground truth is available to tune both the lidar and radar models, a temporary ILM is manually configured by the author to provide a reference. This reference is further used to perform a parameter grid search for each of the geo IRM variants mentioned in Section 3.3.4. Here, the parameters, if not fixed for the