

4.5.4 Discussion

Based on the observations in Section 4.5.2, first, the deep ICMs shall be discussed. Here, using the C_{RGB} input encoding resulted in the worst performance. This might be another indicator to proof the flaws in using homography over MonoDepth or the other deep learning based methods mentioned in the state-of-the-art review in Section 2.3.2 to obtain environment perception from camera images in the BEV projection. However, it needs to be mentioned that the decreased performance might be due to the fact that the network's capacity has been adjusted in Section 4.2 for the R_1 input. Thus, it might be possible to eliminate some of the deficits using a bigger model and more data.

When comparing the C_{RGB} with the C_S input encoding, it can be seen that the semantic information leads to an overall better performance. This, again, might be due to the fact that the network capacity is not big enough to fully learn the mapping based on the camera image input. The biggest performance gap, however, is between the true positive rates of dynamic predictions which shows that including semantic information helps in identifying dynamic objects. Another prominent observation is the increased false rate of $p(\tilde{f}|d)$ showing the confusion of homography-based ISMs to distinguish free from dynamic space. A possible explanation for this characteristic might be the shape distortion of non-flat objects like cars during the homography projection leading to free space in areas where the camera captures space beneath the car. A qualitative example showing the distortion is depicted in the upper left white box in Scene A of Fig. 4-15. Eventually, it needs to be mentioned that, up to a certain degree, the decreased performance of the deep IC_SM can be explained by the errors of the used semantic segmentation model. Since the model has not been retrained, its performance is limited for some classes like the only partially correct classified bus depicted in the upper right white box in Scene A of Fig. 4-15.

For the case of the deep IC_DM, objects can be detected way more clearly as compared to the homography variants. However, the model is confused whether the object is indeed static or dynamic. Here, the implicit assumption to only utilize the height information in C_D might not suffice to provide this information. Here, enhancing C_D with semantic information, e.g. using an additional channel, at least for potentially dynamic classes like e.g. pedestrians or vehicles might lead to a better distinction.

~~Static?~~ Eventually, moving to the geo and deep ILMs as discussed in Section 4.5.3. Here, the geo ILM's free predictions provide almost perfect overlap with the ground-truth maps', showing that the mapping process does not alter the free space much. However, the occupied space shows lots of free space ray breaches which can also be verified qualitatively. This either shows deficits of the chosen parameterization (e.g. opening angle of cone φ_{\triangleleft} might be too small) or hint to potential improvements of the

5.1.3 Discussion of the Verification of Combination Rule Choices

The above results demonstrate the claimed properties of all combination rules and, thus, verify the choices in Section 3.3.10. With regards to RQ11 to disable the influence of the deep ISM, the results for the lower-bounded Yager rule clearly show the restricting properties of the fusion to suffice a lower bound on the unknown mass. However, to fully verify the robustness of the fusion approaches and investigate additional cases not covered in the above simulations, the proposed approach is applied on real data in the following sections.

~~real~~

5.2 Analysis of redundant Information in deep ISMs

Welche Daten werden verwendet?

This section focuses on the experimental verification of H2 which concerns the accumulation of temporal redundancy in deep ISM mapping.

5.2.1 Setup of Redundancy Analysis in deep ISMs

To verify H2, occupancy maps are created in five ways. The baseline is the direct accumulation using Yager's rule, abbreviated as "Yager". This is evaluated for the deep as well as the geo ISM. Second, a naïve solution is evaluated which moves all free and occupied mass to unknown for predictions with $m_u \geq 0.9$ to hinder small predictions to accumulate over time ("Yager + cutoff"). Third, the method as proposed in Section 3.3.9 is used to reduce the redundancy of the deep ISM predictions before accumulation with Yager's rule. The so reduced predictions are abbreviated with "deep, red.". Eventually, the fusion using the deep ISM with and without redundancy reduction respectively with the geo ISMs by accumulating the predictions using Yager's rule is evaluated. Here, the fusion is investigated to analyze whether the deep ISM indeed overwrites most of the geo ISM's predictions as suggested in Section 3.3.10.

The evaluation is performed using only the accumulated radar detections of the recent 20 timesteps (R_{20}) since the focus of this work lies on radar occupancy mapping. Additionally, it is assumed that the effects are similar using other sensor modalities. The geo ISM method is used as described in Section 3.3.3 and ShiftNet, as described in Section 3.3.7, is used as a deep ISM.

The metrics used are the normed confusion matrix (see Section 3.3.8) and the mIoU. Here, the metrics are evaluated separately in all of the area in a 20m vicinity around the ego vehicle trajectory ("whole mapped area") and in an area of 15 pixels around occupied ground truth pixels in the reference occupancy maps. Here, the mIoU is only evaluated around the occupied borders to quantify the cleanliness. The reference maps are created by accumulating the geo ILM using Yager's rule. Two examples of

~~ILM~~

the evaluation areas can be found in 5-3. The scenes mapped are solely taken from the test set which was not used during training.

5.2.2 Results of Redundancy Analysis in deep ISMs

The findings in Fig. 5-2 show that, even though, the geo IR₂₀M's occupancy map has the least true positive rates overall, it also has by far the least false rates. More specifically, the deep IR₂₀M without redundancy reduction produce about four times and the one with the reduction about twice the amount of false occupied predictions in the whole mapped area. Around the boundaries, this ratio even goes up slightly. For free false rates, the ratio of geo IR₂₀M maps is about half of the other variants. On the other hand, the geo IR₂₀M maps only reach about half the true occupied and about two third of the free positive rates. Also, the sparse nature of the geo IR₂₀M can be seen for the unknown class where the accumulation of the deep IR₂₀M results in ten times the false rates both for free and occupied. This is also reflected by the "cleanliness" of boundaries as measured by the mIoU where this variant reaches the best score in the occupied and unknown classes and is clearly shown in Fig. 5-3 for both scenes.

Next, scores of the accumulated deep IR₂₀M reach the highest true and false rates for occupied predictions especially in the unknown regions which can be seen in all of the marked boxes in Fig. 5-3. This also leads to the lowest mIoU scores in the boundary region. When fusing the predictions with the geo IR₂₀M, the scores do only minimally change hinting to a domination of the deep IR₂₀M over the geometric one. The qualitative results verify this domination by barely showing any signs of the geo IR₂₀M influence. By looking closely in the upper white box in scene B some occupied detections of the geo IR₂₀M can be found.

The naïve solution to cutoff low probable predictions $m_u > 0.9$ and set them to unknown slightly reduces the true and false rates of occupied predictions. However, in unknown areas, the amount of occupied predictions is more drastically reduced. Interestingly, the false and true rates of free space slightly increase through the cutoff. These findings can also be verified in Fig. 5-3 by comparing the accumulated deep IR₂₀M with and without cutoff. Here, the most significant change can be seen by a reduction in occupied predictions in the unknown areas. In some cases (white box in scene A), free space replaces the former occupied areas.

Applying redundancy reduction before accumulation has a similar effect than cutoff. The occupied rates decline while the free rates are increased. However, in the unknown area, both the free and occupied rates are reduced in contrast to cutoff where the free rate is increased.

false rate only slightly worsens. In the boundary area, the overall improvement leads to reaching the best score for free space and a close performance in the other classes compared to the geo ILM map. These findings are also qualitatively verified in Fig. 5-9 in Section 5.4.2.

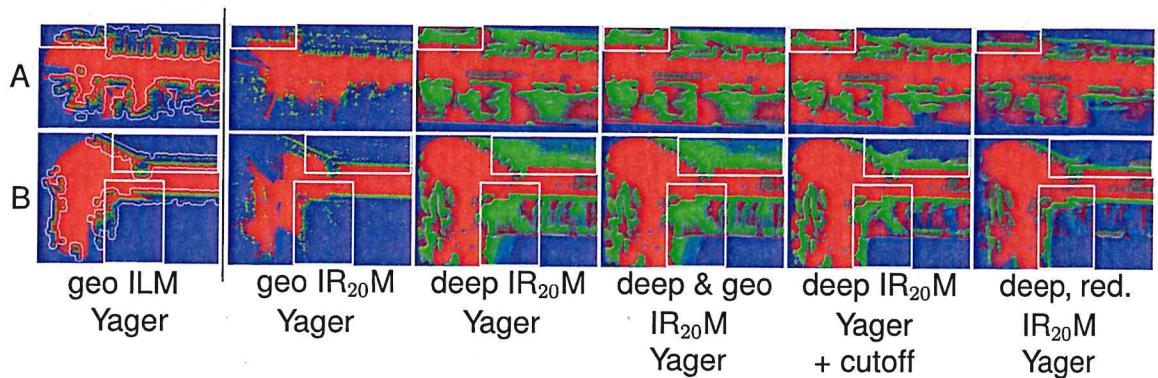


Fig. 5-3: Illustration of qualitative results for the different occupancy mapping variants for two scenes (A & B).

To further verify the workings of the redundancy reduction update, three steps of the mapping procedure are visualized in Fig. 5-4. It can be seen that in the first step, when all map cells are initialized to unknown, the complete prediction of the deep IRM is used to update the map. Afterwards, the majority of new information comes from areas that just entered the deep IRM's FoV. In the depicted case, the ego vehicle is moving from left to right and, thus, the majority of new information is at the right border (see e.g. white box on the right for $t = 10$).

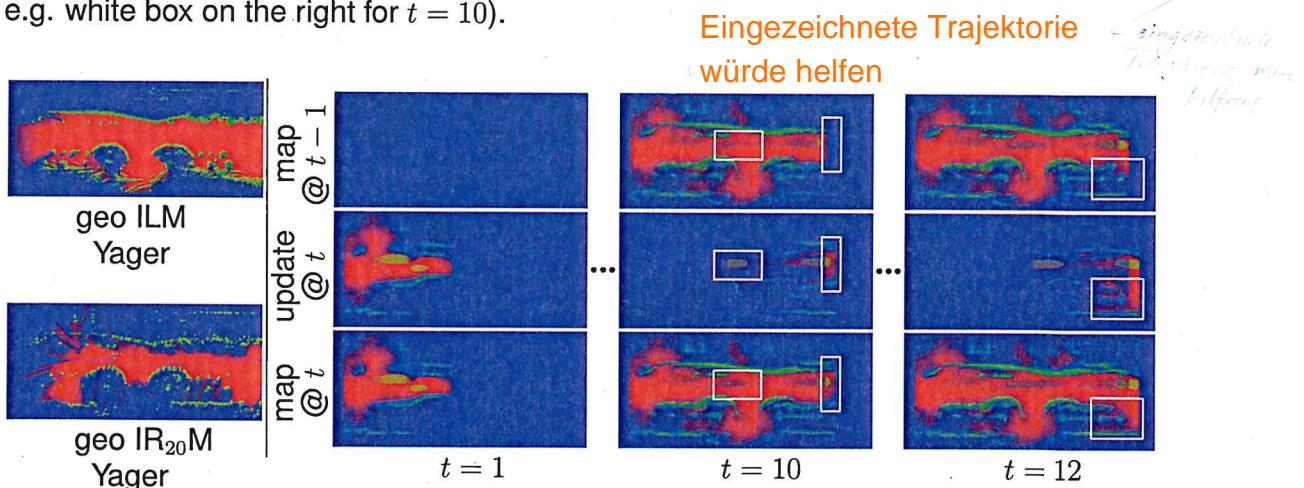


Fig. 5-4: Mapping process given redundancy reduced deep IRM updates. Left, the geo ILM and IRM map are shown for reference. Right, the first row shows the map state before the update, the second the redundancy reduced deep IRM update based on the first row's state and, finally, the updated map in the last row.

Besides from areas entering the ISM's FoV, formerly occluded that become visible are also identified to provide new information (see white box for $t = 12$). Finally, the redundancy reduction is formulated to treat regions of conflict as rich in new information. This is done to allow the map to react to changes in the environment e.g. a parked vehicles starting to move. Thus, dynamic objects are also treated as new information and remain in the update (see left white box for $t = 10$). ~~1~~ ?

5.2.3 Discussion of Redundancy Analysis in deep ISMs

The experiment in Section 5.2.2 shows the problem of uncontrolled accumulation of masses in occluded areas. This property leads to fully overwriting the geo IR₂₀Ms predictions during mapping. This makes the use of the original deep IR₂₀M in fusion with the standard Yager rule prohibitive.

Applying cutoff reduces the distribution of occupied mass in unknown areas more compared to visible. This proves H2 by showing that small biases are being accumulated. Here, the cutoff might have been chosen too small since, instead of occupied, now free mass is being accumulated in occluded areas (see Fig. 5-3 white box in scene A and free rate in unknown area in Fig. 5-2). However, by further increasing the cutoff threshold, only high certain estimates can be conserved leading to too much loss in information of the deep ISM predictions.

Using redundancy reduction, the low certain estimates can still be included in the mapping. But, they can only contribute up to their certainty and, thus, not overwrite areas by sheer number. Additionally, no hyperparameters are introduced. Among the deep IR₂₀M based maps, this leads to the best edge preservation even coming close to the geo IR₂₀M map with regards to mIoU (see Fig. 5-2). It is also qualitatively verified in Fig. 5-4 that only conflicting estimates (e.g. dynamic objects or falsely predicted regions) and areas with lower unknown mass (e.g. formerly completely unknown areas) are updated. This verifies the properties claimed in Section 3.3.10 of the lower-bounded Yager rule. The benefits of the redundancy reduction are further verified by the results of its fusion with the geo IRM. The improvements of the scores clearly shows the influence of the geo IRM is not negated but instead utilized. The qualitative and quantitative verification of the redundancy removal approach described in Section 3.3.9 clearly prove the problem of temporal redundancy accumulation, as claimed in H2. Furthermore, the benefits of the proposed countermeasure prove the choice of mutual information and, based on it, the discount factor to be valid. Thus, verifying the proposed answers of Section 3.3.9 to RQ8 and 9.

Since the problem indeed seems to originate from the accumulation of false predictions in occluded areas, another solution might be to directly train the deep ISM to predict the

geo ILM and not the mapped geo ILM patches. This solution has not been considered from the start since the work's aim is to initialize as much space around the vehicle as possible to increase the convergence speed. Also, it might be possible to train the deep ISM to, based on a given state of a map, predict either the best next occupancy state or the next update. This, however, is highly non-trivial since the network has to be trained to cope with the maps created by a number of sensor models including itself.

~~einbaubare ZSL~~
Even though the redundancy reduced $IR_{20}M$ map comes close to the edge performance, it still lacked behind the geo $IR_{20}M$ map's mIoU. Thus, in Section 5.4, it is analyzed how the inter- & extrapolation properties of a deep ISM can be combined with the edge accuracy of a geo ISM during mapping.

5.3 Comparison of deep ISM Occupancy Maps given different Sensor Modalities

In this section, RQ12 shall be tackled concerning the comparison of occupancy mapping results given the proposed deep ISM with different inputs.

5.3.1 Setup of Deep ISM Maps Comparison

The setup for comparing occupancy maps given different sensor modalities is the same as described in 5.2.1. The only difference is that for this experiment the fusion method is chosen to be the redundancy reduced accumulation with Yager's rule for all variants (abbrev. as "deep, red."). The compared ISMs are all trained ShiftNets while the inputs are MonoDepth (C_D), lidar (L) and their respective combinations with the accumulated radar point cloud of the last 20 steps (R_{20}).

5.3.2 Results of deep ISM Maps Comparison

The quantitative results show a similar line of improvement as shown for the ISM comparison in Section 4.5 and are overall consistent in the whole as well as the boundary area. More specifically, the MonoDepth input lacks behind the lidar input in all classes. While the false occupied rate is doubled for the lidar resulting in a 1% increase, the true occupied rate is more than three times higher providing an increase of almost 50%. This improvement can be clearly verified qualitatively comparing the sharpness of occupied boundaries between MonoDepth and lidar maps. Moreover, adding radar information provides a further improvement in true rates. This is especially true for the MonoDepth case, showing that providing measured depth over estimated depth improves the occupancy predictions. For the false rates, radar provides consistent improvement for the free class while slightly decreasing the performance for the occupied case.

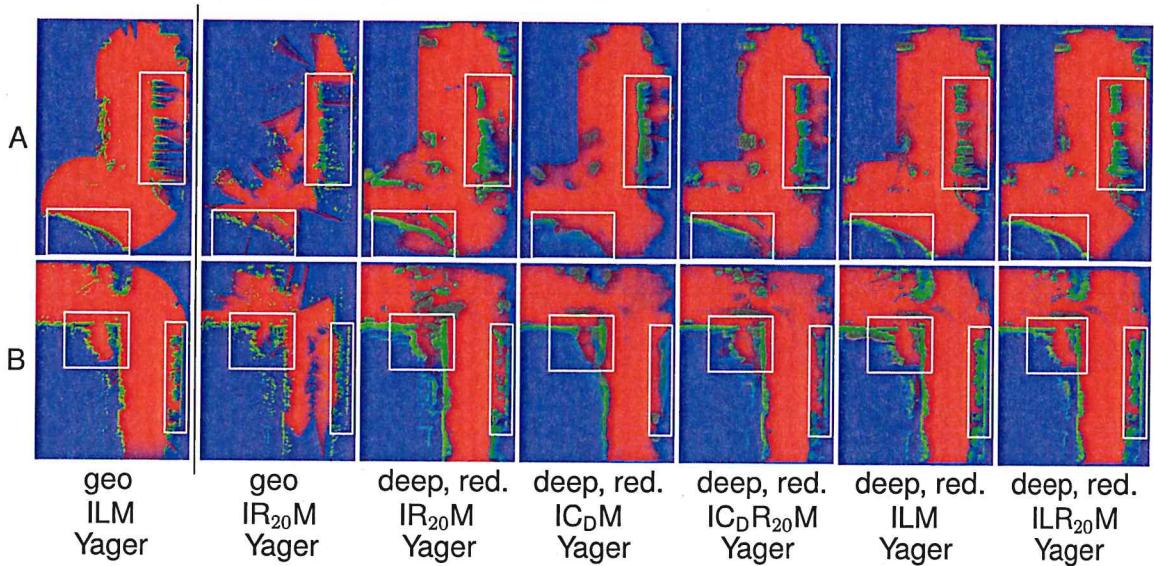


Fig. 5-6: Qualitative results of the occupancy maps created using different ISMs. The abbreviations are explained in Section 5.3.1.

nebenwirk?

The analysis in this section provides a quantitative as well as qualitative comparison of the occupancy maps created the proposed deep ISM using different sensor input, thus, answering RQ12.

5.4 Analysis of deep ISM Priors for Occupancy Mapping

In this section, it shall be analyzed how the high accuracy of geo ISM around edges can be combined with the inter- & extrapolation properties of deep ISMs by initializing the map using the deep and finetuning it with the geo ISM.

5.4.1 Setup of deep ISM Priors Analysis

The experiments are performed using the same sensor inputs, test sets and ground truth as described in Section 5.2.1 for the same reasons. The methods compared create maps by accumulating

- the geo IRM using Yager's rule,
- the redundancy reduced deep and geo IRM combined with Yager's rule and $m_u = 0$ for the deep IRM and
- the fusion of the reduced deep IRM with $m_u = 0.3$ using Yager while the geo IRM is fused using Yager for all cells with $m_u < 0.3$ and YaDer elsewhere

The considerations and definitions for this approach are detailed in Section 3.3.10.

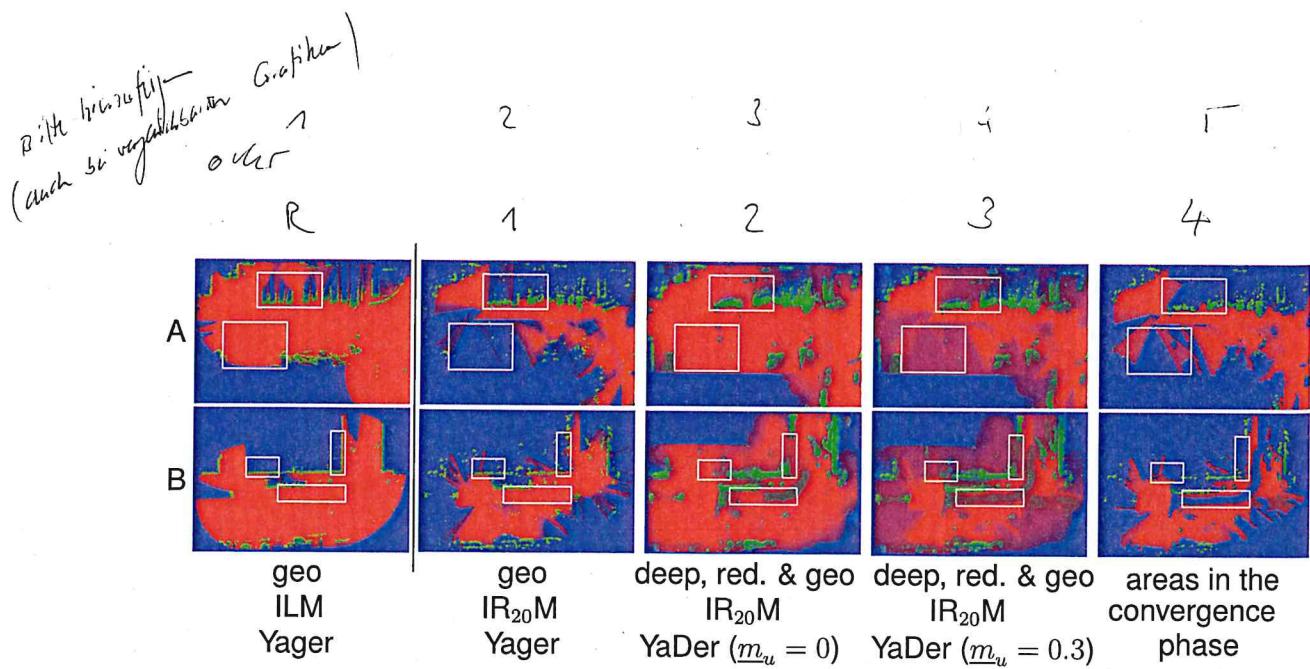


Fig. 5-9: Qualitative results of three mapping approaches (column 1-4) for two scenes with the geo ILM ground-truth map in the first column. Additionally, the 4th column shows the result of filtering out all non-converged pixels of the 3rd column's map. This is done by setting all pixels with $m_u > \underline{m}_u = 0.3$ to $m_u = 1$.

5.4.3 Discussion of deep ISM Priors Analysis

The above shown experimental results on real world sensor data together with the verification based on simulation in Section 5.1 show that the method devised in Section 3.3.10 is indeed capable to restrict the fusion of an ISM to a lower bound given a grace interval to allow for numerical instability. Thus providing an answer to RQ11. It has also been shown that the majority of the geo IRM map's correct predictions remain intact. The found falsification is to be expected since it is still a fusion approach. If both models result in similar errors, e.g. due to incorrect modeling, missing detections, false dynamic state of a detection etc., these errors are amplified resulting in a falsification. It is, however, shown that the falsification is reduced using the deep IRM as a prior, again demonstrating the restrictive influence of the deep ISM purely for initialization.

~~Satz!~~

CNNs are designed to utilize the spatial coherence in data, which can be seen e.g. in the white box of scene A in Fig. 4-13 where the shape of moving vehicles is being estimated solely on a single detection point and the street layout. Eventually, deep CNNs are capable to be trained on multi-class classification tasks, sufficing R1.6, thus meeting all the preliminary requirements to chose the model.

Given the usage of deep CNNs to model the ISM, an architecture search similar as proposed in [RAD20] is performed to answer RQ3 considering the best architecture for the given requirements for input & ouput dimension R1.5 and the requirement for inference speed R1.4. Even though the effects described in [RAD20] are verified in the experiments leading to a good trade-off between accuracy and speed, the search is only conducted on a small subset of the proposed parameter range in [RAD20]. Also, the search is only conducted for radar inputs R_1 since this work focuses on radar ISMs and it is the most challenging radar input encoding. However, these restrictions might cause limited performance for other sensor modalities or radar encodings. Thus, it might make sense to revisit the architecture search when specifying the deep ISMs based on other sensors or given other input/outputs dimensions.

~~Why?~~

Next in Section 4.3, the optimized architecture is used to train three variants of deep ISMs both on R_1 and R_{20} (see definitions in Section 3.3.5) to analyze the influence of added information. The first, referred to as SoftNet, is treated as the baseline using a standard softmax activation and cross entropy loss. The results verify the suggested property of SoftNet to model uncertainty between classes by distributing mass to equal portions, as formulated in H1. In case of uncertainty between free and occupied class, this leads to assigning conflicting mass violating both R1.7 and R1.8. As a solution, DirNet and ShiftNet are analyzed both using a specific mechanism to identify aleatoric uncertainty and moving it to the unknown class. The results show that indeed both methods are capable of solving the problem to a certain extent, better sufficing the requirement of the unknown mass as inverse measure for information (R1.7) and the conflicting mass being an indicator for dynamic objects (R1.8). Here, the ShiftNet variant is chosen for further considerations since it provided overall better results around the boundaries of occupied areas and higher free space scores.

Afterwards in Section 4.4, to fully analyze the effects of different radar input encodings, as formulated in RQ6, the ShiftNet results for R_1 and R_{20} inputs from the uncertainty analysis are further compared with the results of a ShiftNet trained on $R_{20|1}$ inputs (see definitions in Section 3.3.5). It can be seen that the scores overall improve in all but the dynamic class by accumulating static detections over time for $R_{20|1}$ as compared to R_1 . These scores can additionally be improved by accumulating dynamic detections over time with a decay on their intensify, as for R_{20} . However, also for R_{20} the quantitative scores suggest an increase of false occupied predictions around a factor

6 Summary and Outlook

In this chapter, the experimental findings from Chapter 4 and 5 are summarized and compared against the research questions and hypothesis defined in Section 3.1 and 3.2. Based on this summary, open questions are derived based on which future research directions are being proposed.

6.1 Summary and Outlook of geo ILM and IRM

To create the targets for the free, occupied and unknown class to train the deep ISMs, the usage of occupancy map patches created by accumulating a geo ILM over time is proposed. This geo ILM needs to only be manually defined and tuned once. Thus, it suffices R1.3 to minimize manual labor. It is further tuned for the dimensions defined in R1.5 so that the map patches as well as all models trained on them meet the requirement. However, since the lidar sensor does not measure motion information, the manually annotated labels for dynamic objects are overlayed. This can be further optimized e.g. by clustering the lidar points and determining the cluster's motion state through radar measurements or by using a Doppler lidar [MA19].

Using the preliminarily tuned geo ILM maps as reference, three variants of geo IRMs are compared to analyze how to enhance the free space prediction coverage without worsening the other classes, as formulated in RQ2. Here, both adding accumulating detections over time and casting additional free space cones with bigger opening angle lead to improvements consistent over all classes. However, there is still no free space coverage in regions without detections (see Fig. 5-9 lower white box of scene A). Hence, further improvements might contain casting free space rays in regions whenever no detections are provided within a certain opening angle.

Using the best performing geo IRM variant as reference, the initially tuned geo ILM is analyzed with regards to the best ground-plane-removal technique to provide maximal overlap with the geo IRM, as formulated in RQ1. Here, removing all street, sidewalk and terrain detections of the lidar sensor provides about the same overlap as a threshold-based removal with a cutoff at 0.5 m. To further suffice R1.3 to use minimal manual labor, the threshold-based removal is being favored.

6.2 Summary and Outlook of deep, evidential ISMs

Given the ground-truth occupancy maps and a reference geo IRM, the trained ISM models can be analyzed. Here, deep CNN-based models are favored over other machine learning approaches since they have been proven in the literature and in this work to be capable of utilizing large amounts of data, sufficing R1.2 use big data. Also,