Change Detection in Automotive Radar based Occupancy Maps using Siamese Networks

Harihara Bharathy Swaminathan
Semantic Information Systems Group
Osnabrück University
Osnabrück, Germany
hswaminathan@uos.de

Aron Sommer
Aptiv Services
Deutschland GmbH
Wuppertal, Germany
aron.sommer@aptiv.com

Uri Iurgel
Aptiv Services
Deutschland GmbH
Wuppertal, Germany
uri.iurgel@aptiv.com

Andreas Becker
Faculty of Information Technology
Fachhochschule Dortmund
Dortmund, Germany
andreas.becker@fh-dortmund.de

Martin Atzmüller
Semantic Information Systems Group
Osnabrück University, and DFKI
Osnabrück, Germany
martin.atzmueller@uos.de

***Abstract*—In this paper we present a deep learning based approach for detecting changes or deviations in the context of radar based occupancy grid maps. Specifically, we propose a convolutional neural network (CNN) based architecture to identify spatial changes. As a reference map of the environment, we use occupancy maps generated using detections obtained from automotive radar sensors fitted to the corners of a test-vehicle. For the purpose of similarity learning, a siamese architecture is used. The network is trained with occupancy maps of highway and urban scenes captured over a period of time around the city of Wuppertal, Germany, focusing on construction zones on the road. As per the initial evaluations, the siamese network is able to classify images with construction zones as changes from non-changes i.e. images without construction zones.**

***Index Terms*—Change detection, Automotive radars, Occupancy maps, Siamese networks, Similarity learning**

I. Introduction

Along with perception sensors such as lidar, camera and radar, maps are crucial for estimating precise position [1], [2], path planning [3], and navigation [4]. An ideal map should represent important features in the immediate environment and also be up-to-date for ensuring safety and reliability of the autonomous vehicle. Hence, the task of detecting changes or deviations from a previously known map using perception sensors is an on-going research topic in the domain of autonomous driving. In addition, identifying relevant changes from irrelevant and apparent changes is still a challenge [5]. While apparent changes are caused due to sensor characteristics and prevailing circumstances such as traffic, parked vehicles are considered as part of irrelevant changes for this work.

In this paper, we propose a technique to detect deviations from a previously known sensor map of an environment with the help of automotive radars attached to four corners of a test vehicle. As a popular example of a deviation from a known map, we consider construction zones as a relevant change due to their steady increase in number every year in both highway and urban scenes. A deviation from the known map could, e.g., be a construction zone that has appeared on the highway last month. An ideal classifier should be able to detect changes such as appearance of new and disappearance of old construction zones in the environment around a vehicle. We propose a convolutional neural network (CNN) [6] based deep learning architecture, using siamese networks [7] for similarity learning on specific occupancy maps constructed via data preparation and augmentation from radar information. As an important assumption for the experiments, the vehicle is considered to have precisely localized itself with an accuracy of up to a few centimeters.

Our main contributions are summarized as follows:

1) We propose a siamese network based deep learning architecture to extract features from radar sensor based occupancy grid maps that are spatial representation of static environment at different points in time. This deep learning classifier is trained using a cross-entropy loss function to detect changes from non-changes.
2) We present a data preparation and augmentation approach for obtaining occupancy maps suitable for deep learning based methods, specifically to create instances of disapparing and appearing construction zones.
3) We provide a comprehensive evaluation, focusing on highway and urban scenes, both for training and testing, where the measurements were taken over a time period of one year. As part of the evaluation, we present the performance of the classifier in detecting both appearance and disappearance of construction zones.

The rest of the paper is structured as follows: Section II discusses related work. After that, Section III presents our proposed network architecture and change detection dataset generation process. Next, Section IV discusses our experimental results in detail. Finally, Section V concludes with a summary and interesting directions for future research.

This work was supported by Aptiv Services Deutschland GmbH.

978-83-956020-9-2 ©2024 Warsaw University of Technology
IRS-2024, Wrocław, Poland

II. Related Work

Image comparison techniques have been extensively studied for various use cases ranging from signature [8] and face verification [9] to change detection in remote sensing applications [10]. Identifying deviations from a previously known map is highly relevant for autonomous driving with an increased use of maps [11]–[13] as a sensor for autonomous vehicles.

Bromley et al. were first to propose a CNN based architecture for verifying signatures [8]. This "Siamese" artificial neural network [7] consists of two sub-networks to extract features from two images using a contrastive loss function. Then, the cosine distance or some distance metric such as euclidean [9] between the extracted signature feature vectors is determined. During training, the weight shared between the two CNN sub networks are then adjusted in a way that the distance between feature vectors produced by similar images is less than a chosen margin. In the remote sensing domain, Mou et al. proposed a CNN [10] for identifying similar image patches of very high resolution (VHR) optical and synthetic aperture radar (SAR) images of urban scenes. The task is to learn a similarity function for image patches obtained from multiple sensors. This network architecture is called "pseudo-siamese" as the input is obtained from two different sensors.

As an extension of the aforementioned works into the autonomous driving domain, Drost et al. [11] proposed a classifier based on a siamese architecture for validating HD map elements against occupancy grid maps generated using a combination of perception sensors such as lidar and radar. Training was performed on pseudo-labels where map deviations are artificially generated by matching map data from different time stamps and GPS positions than sensor data. It was also suggested that the classifier can be refined through training on actual map deviations rather than artificially generated samples. Lambert and Hays [12] presented the first dataset for the change detection task named as "Trust, but Verify (TbV)" along with several learning based architectures for HD map change detection problems. However, the examples of HD map deviations in this dataset such as lane marking color change, deletion of crosswalk and change from lane boundary marking from double-solid yellow to single-solid yellow can only be detected using cameras. In contrast, we focus mainly on use cases such as construction zone appearances which can be detected using radar sensors. In our previous work, we proposed a spatial change detection technique for detecting changes in the position of an isolated semi-static pole [13] involving statistical features that are extracted manually. In this paper, we expand our experiments to real world scenes containing construction zones and propose a siamese network architecture based change detection classifier, trained only on occupancy grid based radar sensor maps without using any artificial map deviations.

Overall, to the best of authors' knowledge, a deep learning based change detection classifier trained only on radar based occupancy maps of real world scenes without the use of artificial data has not yet been proposed in literature before.

III. Methodology

A. Network Architecture

The network architecture consists of two CNN sub networks as a backbone and a decision head with fully connected layers for comparing the extracted feature vectors. For the two CNN backbones, we use the ResNet-18 [14] feature extractor. For the fully connected layers, the feature vectors from the CNN backbones are concatenated as input. The fully connected layers have an output size of 256 and 1 followed by a sigmoid activation unit as final layer. The network is trained using a binary cross-entropy (BCE) loss function.

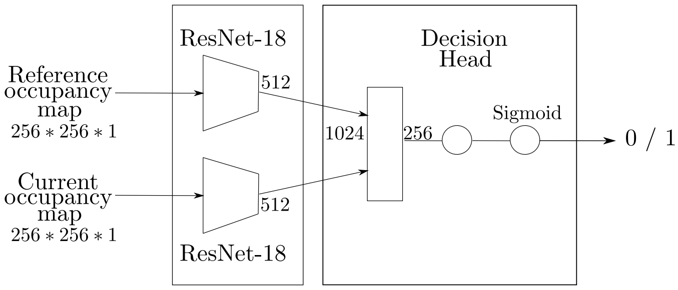


Fig. 1: Siamese network architecture

Each branch of the applied ResNet-18 takes a single channel image of size $256 * 256$ pixels as input. Before feeding to the neural network, images are normalized in such a way that all the pixel values are scaled to a range of -1 to 1. As an output from the ResNet18 network, we obtain a feature vector of length 512 for an image. The outputs from the two CNN branches are concatenated to produce a feature vector of length 1024 for each image pair given to the CNN backbone. The first fully connected layer of the decision head consists of 256 output units followed by activation using the ReLu function. Finally, these features are fed to a single neuron followed by activation using a sigmoid function.

B. Dataset preparation

For measurements, we used a test vehicle equipped with a high precision DGPS measurement unit, the POS LV-220 positioning system from Applanix corporation. Additionally, it is fitted with Aptiv's four corner radar sensors and an integrated lidar, camera system from Pandora. In order to create a change detection dataset, the measurements were done over a period of one year on highway and urban areas containing construction zones around the city of Wuppertal, Germany. A detailed description of the dataset is given in Table I. For highway scenarios, we selected the stretch of A46 connecting cities Wuppertal and Duesseldorf. Before its completion in year 2023, the construction zone was present on both directions of the highway. For urban scenarios, we recorded measurements around the university of Wuppertal. Each sample of the dataset consists of a bird's eye-view (BEV) image pair obtained from different measurement drives on the same route.

As per Table I, mini datasets with case numbers 1, 2, 3 and 11 contain instances of construction zones that disappeared and case numbers 8, 9, 10 and 12 contain instances of appearances. Finally, the mini datasets with case numbers 4, 5, 6 and 7 do not contain any changes between the two measurements.

For detailing the process of BEV image pairs generation, we assume measurement drives A and B which could have taken place over a period of few hours to several months. Let A refer to measurements used for generation of the reference occupancy map and let B refer to measurements used for generation of the current sensor occupancy map. The following steps have been applied to generate data samples from measurement sets A and B.

1) For a time period $t = 1s$, we consider the test-vehicle's position and orientation as P consisting of easting (e), northing (n) and heading (θ) [15] from B. Additionally, we use the distance formula to calculate the distance (d) travelled in meters during this time period (t).
2) For point P, find the nearest vehicle position and orientation (Q) from A using 2D euclidean distance formula.
3) We select radar detections in A that they were captured at the closest vehicle position Q and other nearby positions calculated using distance d. We then generate 2D map coordinates for the radar detections considering point Q as origin.
4) Similarly, we select all radar detections in B that were captured from point P and also at nearby vehicle positions calculated using distance d for occupancy map generation. We then generate 2D map coordinates for the radar detections considering point P as origin.
5) We generate 2D occupancy grid maps that depict a region of up to 50m on all sides and the vehicle pointing towards the north (top). The occupancy map generated from detections in A is the reference occupancy map and the map created from detections in B is the current sensor occupancy map. The number of cells were selected in such a way that the occupancy maps have a resolution of 0.1m.
6) Depending upon the position of construction zone related objects such as poles and walls, each data sample is labelled 0 for no change and 1 for change. For e.g. If a reference occupancy map contains no construction zone related object and the current sensor occupancy map contains construction zone related object, the sample is labeled 1.

Please note, that driving a vehicle on the same trajectory in each of the measurement drives is not possible due to a number of factors such as traffic situation and lane closure. As a result, the two measurement drives i.e. A, B could have offset in vehicle trajectory and variation in speed. As the occupancy grid represents only the static environment around the vehicle, radar detections belonging to moving objects and parked cars are filtered out during occupancy grid map creation.

C. Data Augmentation

For generating instances of disappearing construction zones, we consider measurements taken at the highway A46 from the year of 2022 as a reference. In contrast, instances of appearing construction zones are generated by considering measurements of the same region from the year of 2023 as a reference. In particular, we augmented mini dataset cases 1, 2, 3 and 11 to generate cases 8, 9, 10 and 12. This can be verified by observing the measurement dates i.e. columns 3, 4 on Table I for the aforementioned cases. We provide examples for instances of a non change (top image), a disappearing construction zone (middle image) and an appearing construction zone (bottom image) as shown in Figure 2 below.

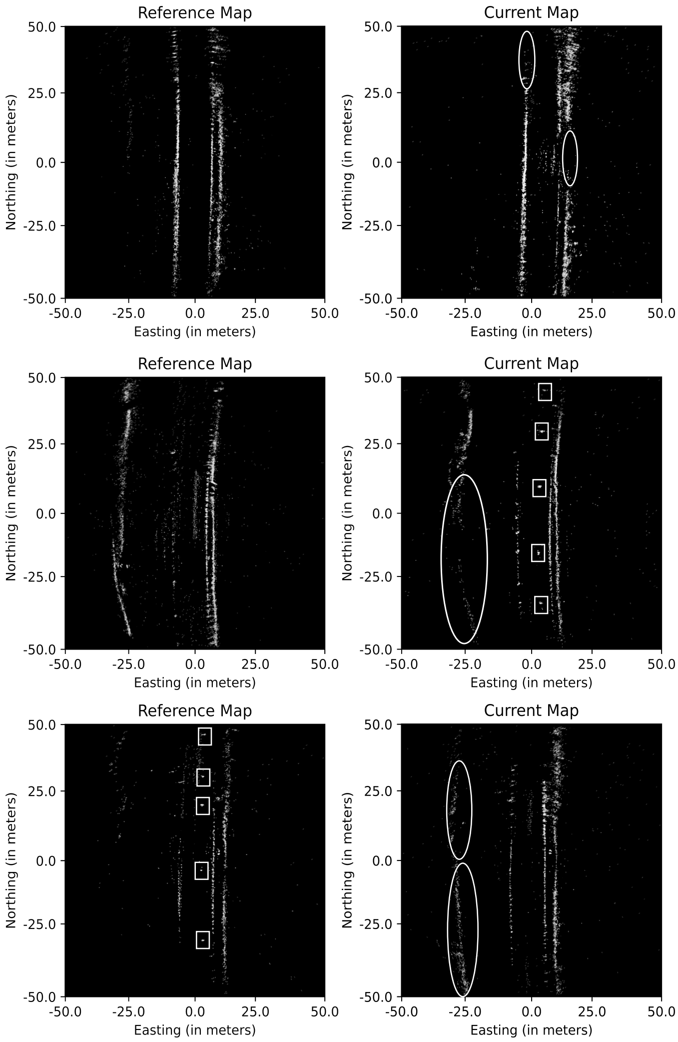


Fig. 2: Instances of a non change (top) (Label : 0), an appearing construction zone (middle) (Label: 1) and a disappearing construction zone (bottom) (Label: 1) from the change detection dataset. Relevant changes i.e. construction zones are annotated with white boxes and apparent changes are annotated in ovals. Note: Only relevant changes were taken into account for labeling.

TABLE I: Details of measurement sessions carried out over a period of one year for the purpose of change detection dataset generation with focus on construction zones.

Case	Type	Reference map	Current map	Type of change	Direction	Positives	Negatives	Total
1	Highway	Aug 2022	Oct 2023	Disappearance	Towards Duesseldorf	3300	805	4105
2	Highway	Aug 2022	Oct 2023	Disappearance	Towards Duesseldorf	3863	295	4158
3	Highway	Sep 2022	Oct 2023	Disappearance	Towards Wuppertal	1328	3441	4769
4	Highway	Oct 2023	Oct 2023	No change	Towards Wuppertal	0	12876	12876
5	Highway	Oct 2023	Oct 2023	No change	Towards Duesseldorf	0	2072	2072
6	Highway	Oct 2023	Oct 2023	No change	Towards Duesseldorf	0	1989	1989
7	Highway	Oct 2023	Oct 2023	No change	Towards Duesseldorf	0	2072	2072
8	Highway	Oct 2023	Aug 2022	Appearance	Towards Duesseldorf	4366	789	5155
9	Highway	Oct 2023	Aug 2022	Appearance	Towards Duesseldorf	5512	414	5926
10	Highway	Oct 2023	Sep 2022	Appearance	Towards Wuppertal	988	3210	4198
11	Highway	Aug 2022	Aug 2023	Disappearance	Towards Duesseldorf	6323	1029	7352
12	Highway	Aug 2023	Aug 2022	Appearance	Towards Duesseldorf	6323	1029	7352
13	Urban	Sep 2022	Jul 2023	Appearance, Disappearance	Around Uni Wuppertal	2076	7298	9374

TABLE II: Elements of a confusion matrix for the change detection problem.

		Truth: Change	Truth: No-change	Total
Predictions	Change	a	b	$a+b$
	No-change	c	d	$c+d$
	Total	$a+c$	$b+d$	N

D. Evaluation metrics

In real-world situations, when considering the change detection dataset, the reference map typically matches with the current sensor measurements and deviations from a reference map are rather rarely to appear. As a result, a corresponding change detection dataset features a rather large amount of non-changes (negatives) compared to instances of changes (positives). Since, accuracy can be a misleading performance metric for imbalanced datasets, we make use of the F1 scores for evaluating the predictions of the classifier. Table II shows the elements of confusion matrix for calculating the Precision and Recall scores.

Precision is defined as the number of true positives over the number of true positives plus the number of false positives as expressed by (1).

$$P = \frac{a}{a+b} \tag{1}$$

Recall is defined as the number of true positives over the number of true positives plus the number of false negatives as expressed by (2).

$$R = \frac{a}{a+c} \tag{2}$$

The F1 score is defined as the harmonic mean of precision and recall as expressed by (3).

$$F1 = 2 \cdot \frac{P \cdot R}{P+R} \tag{3}$$

IV. Experimental Results

In this section, we present the training details along with the most notable test results.

A. Training Details

We have trained all the models from scratch since no other pre-trained models were available with the input as radar based occupancy grids. For training the siamese network, we used the Adam Optimization Algorithm [16]. The learning rate is fixed at $l_r = 0.001$ and weight decay parameter $1e^{-5}$ which introduces L2 regularization to avoid overfitting. For training, we use mini-batches of 16 image pairs, the Binary Cross-Entropy (BCE) loss function as formalized in Equation (4)

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \log(p(y_i)) + (1-y_i)\cdot \log(1-p(y_i))] \tag{4}$$

Each network was trained for a total of 10 epochs. The weights were initialized using the Glorot initialization [17] method, using a uniform distribution. We selected the training and testing datasets considering the following objectives:

1) We aimed to determine the performance of a change detection classifier by training it only on image datasets depicting a disappearing construction zone i.e. cases 1, 2, 4, 5, 6 and 7.
2) Furthermore, we targeted the assessment of the performance of a change detection classifier after data augmentation to generate image datasets depicting both disappearance and appearance of construction zones. i.e. cases 1, 2, 4, 5, 6, 7, 8 and 9.
3) Finally, we applied the classifier in order to understand the ability of a network to classify unseen construction zones, the testing dataset consisted of zones which were not part of the training dataset. Furthermore, we were also interested in evaluating the network performance on an urban dataset. Therefore, we consider cases 3, 10 and 13 for testing.

B. Test Results

In this section, we compare the performances of two models A and B, trained using two different strategies. In particular, model A was trained using datasets depicting only disappearances of construction zones i.e. a total of 27,272 samples containing 73.73% negatives and 26.27% positives. On the other hand, model B was trained using datasets additionally generated through data augmentation process and depict both appearance and disappearance of construction zones i.e. a total of 38,353 samples containing 55.56% negatives and 44.43% positives were used for training model B. As testing datasets, we consider cases 3, 10 and 13, where cases 3 and 10 contain unseen construction zones in the direction of Wuppertal city on the highway A46 and case 13 contains instances of unseen construction zones located in an urban scene.

Figure 3 shows the F1 scores of models A and B on the aforementioned test datasets, where an improved model performance through data augmentation can be observed with better F1 scores for model B than model A on the test datasets. In particular, model B which was trained using datasets generated through data augmentation attained an F1 score of 0.82, 0.81 and 0.38 on test datasets case 3, case 10 and case 13 respectively. Furthermore, we also present randomly selected examples computed by network B in Figure. 4.

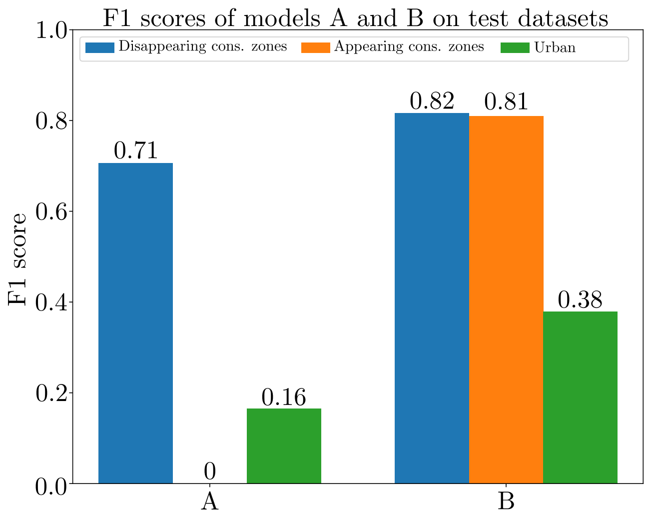


Fig. 3: F1 scores of models A and B on test datasets containing unseen disappearing (case 3), appearing (case 10) construction zones on highway A46 in the direction of Wuppertal and urban dataset (case 13) around the university of wuppertal

V. Conclusions and Future Work

In this work, we presented a deep learning based classifier for detecting changes or deviations from a radar based occupancy grid map. For training this network, we used measurements from automotive radar sensors fitted to the corners of a test vehicle. The dataset consisted of images depicting construction zones measured using the test vehicle over an year on highway and urban scenes. For the architecture, we were inspired by siamese network methods used previously for signature and face verification tasks. Through data augmentation to create instances of disappearing and appearing construction zones, we have shown that a model could perform better at detecting changes on instances of unseen construction zones from highway and urban scenes.

For future work, we aim to consider to expand this classification task by creating a map that shows the changed portions of the environment. In addition, a variety of other objects such as disappearance or appearance of poles, walls of buildings and guard rails could be considered for a more elaborate change detection dataset generation with focus exclusively on objects that are detect by automotive radar sensors. Not only occupancy maps, but HD maps belonging to feature domain are also used as reference maps in autonomous driving. However, realistic simulation of radar measurements is needed to lay the foundation for detecting changes on such HD maps using radar sensors.

References

[1] A. Pishehvari, U. Iurgel, M. Stefer, and B. Tibken, "Range-doppler registration," in *2020 5th International Conference on Control and Robotics Engineering (ICCRE)*. IEEE, 2020, pp. 169–176.

[2] H. B. Swaminathan, A. Sommer, A. Becker, and M. Atzmueller, "Performance evaluation of gnss position augmentation methods for autonomous vehicles in urban environments," *Sensors*, vol. 22, no. 21, p. 8419, 2022.

[3] Z. Jian, S. Zhang, S. Chen, X. Lv, and N. Zheng, "High-definition map combined local motion planning and obstacle avoidance for autonomous driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 2180–2186.

[4] A. Diaz-Diaz, M. Ocaña, Á. Llamazares, C. Gómez-Huélamo, P. Revenga, and L. M. Bergasa, "Hd maps: Exploiting opendrive potential for path planning and map monitoring," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 1211–1217.

[5] E. J. Parelius, "A review of deep-learning methods for change detection in multispectral remote sensing images," *Remote Sensing*, vol. 15, 2023.

[6] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.

[7] Y. Li, C. P. Chen, and T. Zhang, "A survey on siamese network: Methodologies, applications, and opportunities," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, pp. 994–1014, 2022.

[8] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[9] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 539–546.

[10] L. Mou, M. Schmitt, Y. Wang, and X. X. Zhu, "A cnn for the identification of corresponding patches in sar and optical imagery of urban scenes," in *2017 Joint Urban Remote Sensing Event (JURSE)*. IEEE, 2017, pp. 1–4.

[11] F. Drost, L. Parolini, and S. Schneider, "Siamese networks for online map validation in autonomous driving," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 57–62.

[12] J. Lambert and J. Hays, "Trust, but verify: Cross-modality fusion for hd map change detection," *Proc. NeurIPS 2021, Track on Datasets and Benchmarks*, 2022.

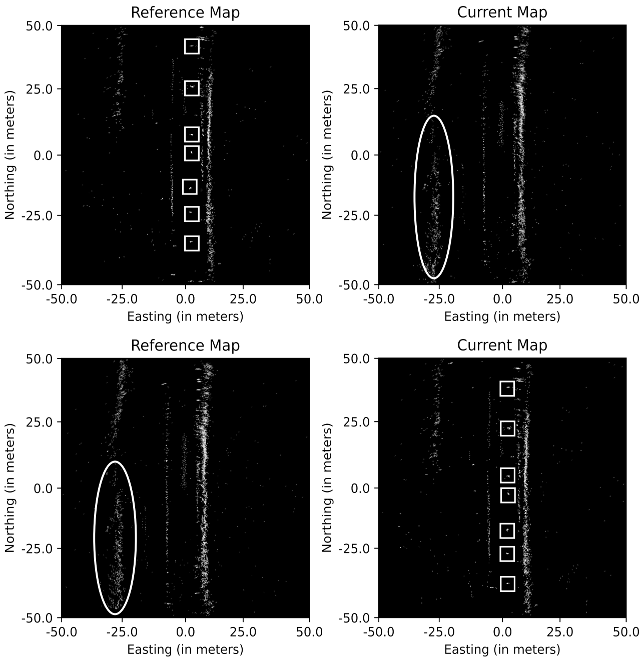
[13] H. B. Swaminathan, A. Sommer, A. Becker, and M. Atzmueller, "Spatial change detection using automotive radar," in *Proc. IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 4066–4071.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

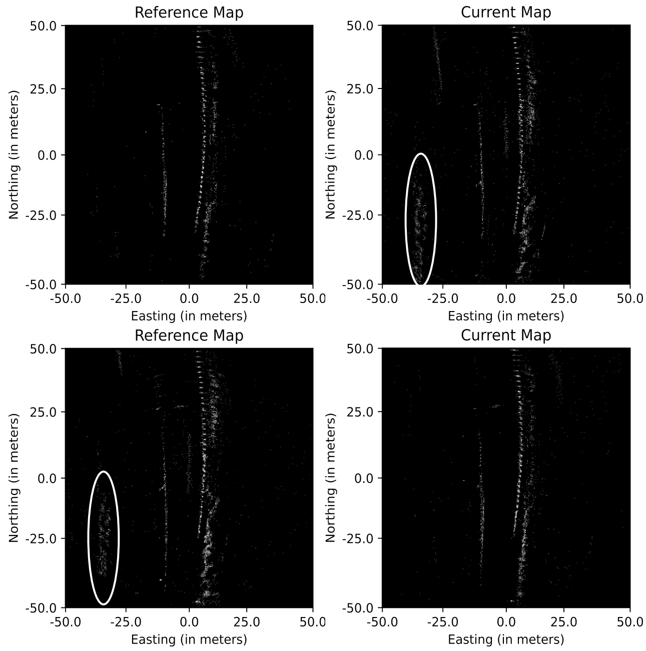
[15] L. M. Bugayevskiy and J. Snyder, *Map projections: A reference manual*. CRC Press, 1995.

[16] D. Kinga, J. B. Adam *et al.*, "A method for stochastic optimization," in *International conference on learning representations (ICLR)*, vol. 5. San Diego, California;, 2015, p. 6.

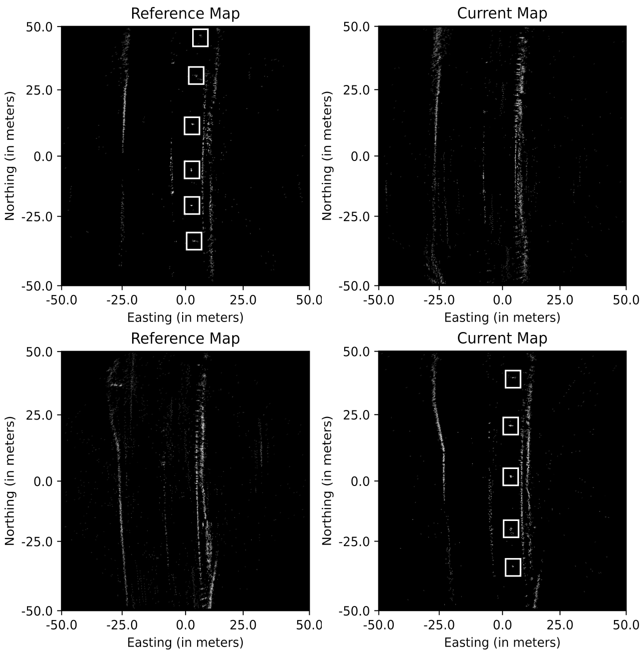
[17] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.



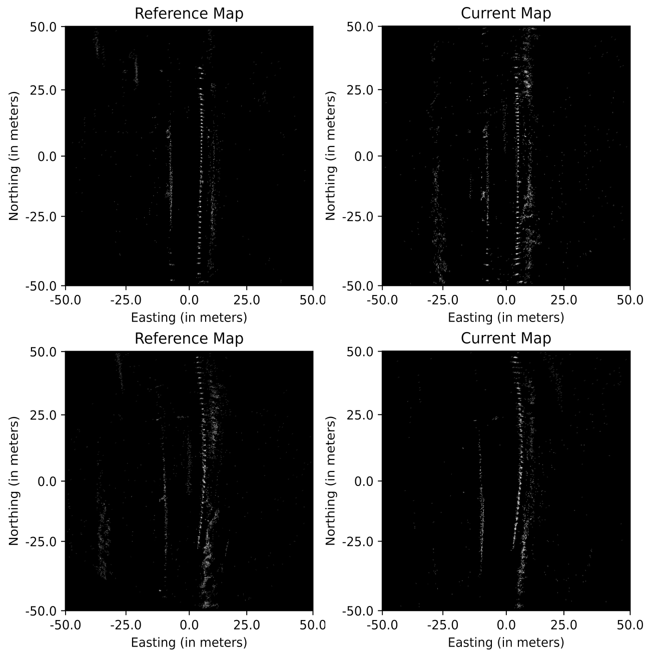
(a) True Positives



(b) False Positives



(c) False Negatives



(d) True Negatives

Fig. 4: Randomly selected predictions of Model B on datasets containing disappearing and appearing construction zones on the highway. The classifier is able to identify instances of changes as seen in Figure 4a and non changes as seen in Figure 4d. Regarding False positive predictions shown in Figure 4b, we clearly observe that apparent changes (annotated in ovals) are a reason for misclassification. False negative predictions are mainly due to examples with specific situations which are relatively difficult to classify, i.e. construction zones not clearly visible and incorrect labels.