

# Doppler-aware Odometry from FMCW Scanning Radar

Fraser Rennie, David Williams, Paul Newman and Daniele De Martini

Oxford Robotics Institute, Dept. Engineering Science, University of Oxford, UK.

fraser.rennie@exeter.ox.ac.uk, {dw,pnewman,daniele}@robots.ox.ac.uk

**Abstract**—This work explores Doppler information from a millimetre-Wave (mm-W) Frequency-Modulated Continuous-Wave (FMCW) scanning radar to make odometry estimation more robust and accurate. Firstly, doppler information is added to the scan masking process to enhance correlative scan matching. Secondly, we train a Neural Network (NN) for regressing forward velocity directly from a *single* radar scan; we fuse this estimate with the correlative scan matching estimate and show improved robustness to bad estimates caused by challenging environment geometries, e.g. narrow tunnels. We test our method with a novel custom dataset which is released with this work at <https://ori.ox.ac.uk/publications/datasets>.

**Index Terms**—radar odometry, doppler, navigation, dataset

## I. INTRODUCTION

Autonomous vehicles start to play a more important and widespread role in transportation as more and more countries are amending road regulations to allow their deployment<sup>1</sup>.

Reliable ego-motion estimation and localisation are fundamental components of autonomous navigation, answering the question of where the vehicle is located and how it moves in the environment. As considered deployment scenarios become more challenging, the detection methods and the sensors collecting data about a vehicle’s surroundings must advance. Currently, the primary sensors used by autonomous vehicles are cameras and LiDAR: while these traditional sensors may perform adequately under favourable conditions, their effectiveness can be significantly compromised by factors such as rain, snow, fog and low light. In contrast, radar offers distinct advantages, including a superior range and invariance to adverse weather and low light conditions [1].

We explore exploiting Doppler information in radar scans for odometry. We first evaluate its utility for masking distractors in radar scans before motion estimation through cross-correlation. We then rely solely on Doppler for motion estimation by training a regressor directly on single scans.

Although Doppler information is usually delivered from radar sensors together with intensity, no available datasets provide it for Frequency-Modulated Continuous-Wave (FMCW) scanning radars [1], [2], [3], [4]. Thus, we collected a custom dataset – an example image is shown in

This project was supported by EPSRC Programme Grant “From Sensing to Collaboration” (EP/V000748/1). We also thank the teams from Oxbotica and Navtech Radar for their invaluable help in collecting the dataset and support on this work.

<sup>1</sup><https://www.gov.uk/government/news/self-driving-revolution-to-boost-economy-and-improve-road-safety>

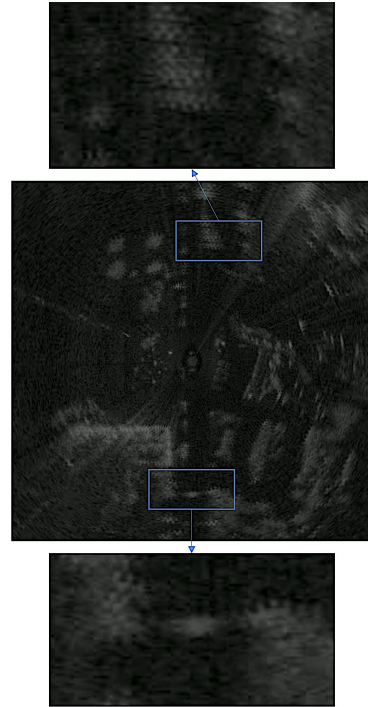


Fig. 1: Radar scan from the RDD dataset. In the right portion, two regions extracted show the “zig-zag” pattern caused by the alternating modulation patterns – in conjunction with the ego-vehicle speed. The top depicts a static object which is affected uniformly by it, while, at the bottom, a chase car with the same speed as the ego-vehicle is unaffected.

Figure 1 – which we used to train and evaluate our approach. On such a dataset, we demonstrate how Doppler information can improve the masking approach *and* generate reliable forward-motion estimation when used for direct regression, even in feature-poor scenery such as narrow tunnels.

To summarise, the main contributions of this work are:

- Use of Doppler information for the masking process;
- Single-scan longitudinal relative-pose estimation;
- A novel FMCW radar dataset containing Doppler information, released alongside this work.

## II. RELATED WORK

FMCW scanning radar has seen an increasing interest in the past years, from odometry and localisation [5], [6], [7], [2] to semantic segmentation [8], [9], detection [4], path planning [10], [11] and cross-modal localisation [12], [13]. As this work discusses odometry, we will focus on this topic.

Odometry can be classified into two main groups, direct and indirect: the first uses direct data collected from the sensors whereas the second requires preprocessing, usually in the form of landmark detection. Both families of algorithms have been explored for FMCW scanning Radar Odometry (RO).

The works [14], [15] demonstrated the potential of FMCW scanning radar for odometry. Such indirect-odometry methods focus on resilient landmark extraction – based on shape similarity and a gradient-based feature detector respectively – and a robust segment-based matching algorithm to deliver RO competitive with vision and LiDAR. [9] applied a learnt method to reject landmarks appearing on unwanted distractor artefacts, speeding up the sequent scan-matching process.

Other RO indirect methods include Cell Averaging CFAR (CA-CFAR) [16], Bounded False Alarm Rate (BFAR) [17] and Conservative Filtering for Efficient and Accurate Radar Odometry (CFEAR) [18]; these methods focus on improving the feature extraction through hand-crafted landmark detection algorithms. Other methods learn feature detectors from the data: [19] uses a U-Net network to predict a map of landmark locations, scores and descriptions from odometry annotations. In [20], the learning approach is pushed even further by training the network unsupervised, jointly estimating landmarks and relative poses.

Differently, direct methods estimate odometry directly from the raw scan. Our method belongs to this family; in particular, we base our scan-matching procedure on [21], [22]. [21] employs a Neural Network (NN) to generate a mask then used to filter out distractors from the radar scans. The relative pose is then estimated using correlative scan matching and, through supervision from odometry directly, the masking network is trained. [22] improves this method's time and memory performances by exploiting the Fourier Transform's properties. Our approach applies the same technique as [22], where we tailor the network to use Doppler information as an additional source.

Doppler effect has been explored for FMCW scanning radars in [23], where the authors evaluated its effect on radar odometry. Here, the authors propose a compensation method to correct return displacements based on their estimated velocity. As the data did not contain multiple modulations or direct Doppler measurement, the speed needed to be estimated from multiple scans and the effect of the correction on the final accuracy was negligible.

Doppler information has also been exploited in the literature to solve odometry. For instance, [24] exploits landmark velocities to assist a radar-based Simultaneous Localisation And Mapping (SLAM) pipeline; yet the estimates are given as a pointcloud, and the raw data is not accessed directly. Doppler has also been explored in a recent work on FMCW LiDAR [25], where the Doppler velocity measurements are used to assist relative pose estimations.

### III. PRELIMINARIES AND DATASET

The sensor we employ is the Navtech CTS350-X, a FMCW scanning radar. The sensor spins at 4 Hz, transmitting a continuous radio wave; however, the radio wave frequency

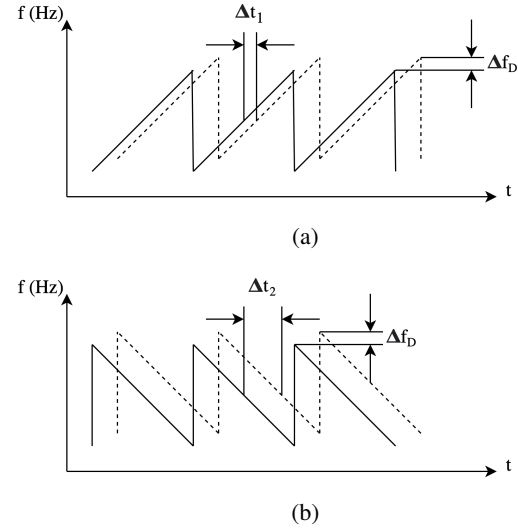


Fig. 2: FMCW Modulation patterns: (a) a transmitted sawtooth wave (solid line) is reflected and distorted by the object to produce the reflection wave (dotted line) used to estimate the distance of the reading; (b) a reverse sawtooth wave transmitted onto the same objects can be compared to (a) to extract Doppler information.

is modulated, i.e. changed throughout the operation, allowing for object-distance measurements by comparing the frequency of the transmitted and returned waves.

In this system, the echo signal's frequency has a delay of  $\Delta t$  called a *runtime shift*, calculated via the change in frequency, also called the *beat frequency*. This is shown in Figure 2a, where  $\Delta f_1$  is the beat frequency.

The distance  $R$  to a given object is then given by:

$$R = \frac{c_0 |\Delta t|}{2} = \frac{c_0 |\Delta f|}{2 \left( \frac{df}{dt} \right)} \quad (1)$$

where  $\frac{df}{dt}$  is the gradient of the sawtooth wave. Dividing the beat frequency by the gradient gives the runtime shift,  $\Delta t$ , which then allows for a direct calculation of the range of the object from which the echo is transmitted.

#### A. Doppler Effect

If an object is non-stationary, the term  $\Delta f$  becomes influenced by the object's speed towards the radar sensor, affecting the perceived range estimation due to the Doppler effect introduced into the system. The beat frequency,  $\Delta f$ , will then contain a frequency shift due to the target's range, runtime shift  $\Delta t$ , and a Doppler frequency shift,  $\Delta f_D$ , which is shown as an error in the range reading.

Let's now gather a radar reading of the same azimuth with a different modulation pattern, as in Figure 2b. This new modulation pattern allows for a comparison between range errors as the received waves have been influenced in opposite directions due to the opposite sawtooth gradients.

The Doppler frequency is calculated via:

$$\Delta f = \frac{\Delta v}{c} f_e \quad (2)$$

where  $\Delta v$  is the relative radial velocity between the vehicle and object,  $c$  is the light speed and  $f_e$  the emitted frequency.

While rotating, the radar sensor inspects one angular portion (azimuth,  $\alpha$ ) of space at a time. A complete scan is composed of  $N_\alpha = 400$  azimuths. The sensor has been set up to utilise the two sawtooth waves alternately, standard for odd and reversed for even azimuths. This approach approximates gathering the same environment with the two modalities, as the azimuths overlap. This results in polar radar images of shape  $N_\alpha \times N_b = 400 \times 3600$ , where  $N_b$  is the number of bins, each with a size of 4.38 cm, which can be transformed in cartesian coordinates, as shown in Figure 3.

### B. Radar Doppler Dataset (RDD)

No available datasets include scanning FMCW radar with Doppler information. For this reason, we propose the Radar Doppler Dataset (RDD), a novel dataset containing both structured and unstructured environments to stretch-test the contribution of Doppler information to motion estimation.

Figure 1 depicts one image from the proposed dataset. Notably, the car is moving, and a chase car follows it at roughly the same speed. Clear “zigzag” artefacts directly result from the two different perceived distances of the objects: the environment is affected by the artefacts depending on the angle of perception, as the Doppler shift depends on the direction of movement. The chase car, instead, is practically unaffected as its speed matches the ego vehicle.

Various routes were collected to maximise the coverage of Doppler effect artefacts, including driving:

- at 10 mph towards and away from a wall;
- along a 4.5 km stretch of a geometrically degraded road;
- around the feature-rich Culham Science Centre.

The total driving amounts to 24.7 km, for a duration of 59 min. Figure 3 shows examples of feature-poor and feature-rich scans. In particular, Figure 3b shows a very challenging scene for standard RO pipelines, as the narrow tunnel is practically unaffected visibly by the vehicle’s movement, and so odometry is likely to fail.

The dataset is released alongside other sensor readings, especially cameras and LiDARs. Importantly, RTK GPS is used to create the ground-truth annotations exploited in this work to learn to estimate odometry.

## IV. METHODOLOGY

Here, we introduce the correlative scan-matching approach to retrieve the ego-motion between two radar scans. We then describe how we integrate Doppler information and fuse such estimate with a pure Doppler-based velocity estimation.

### A. Correlative scan-matching

Let  $Z_{t-1}$  and  $Z_t$  be two consecutive radar scans in cartesian form in  $\mathbb{R}^{n \times n}$ , where  $n$  is the dimension in pixels of the images. Our goal is to retrieve a rigid-body transformation  $[R|t] \in \text{SO}(2)$  – where  $R$  is a rotation of an angle  $\theta$  and  $t$  a translation  $[t_x, t_y]$  – such as  $Z_t \approx T_{[R|t]}(Z_{t-1})$ , where  $T_{[R|t]}$  is an affine transformation following  $R$  and  $t$ . Such

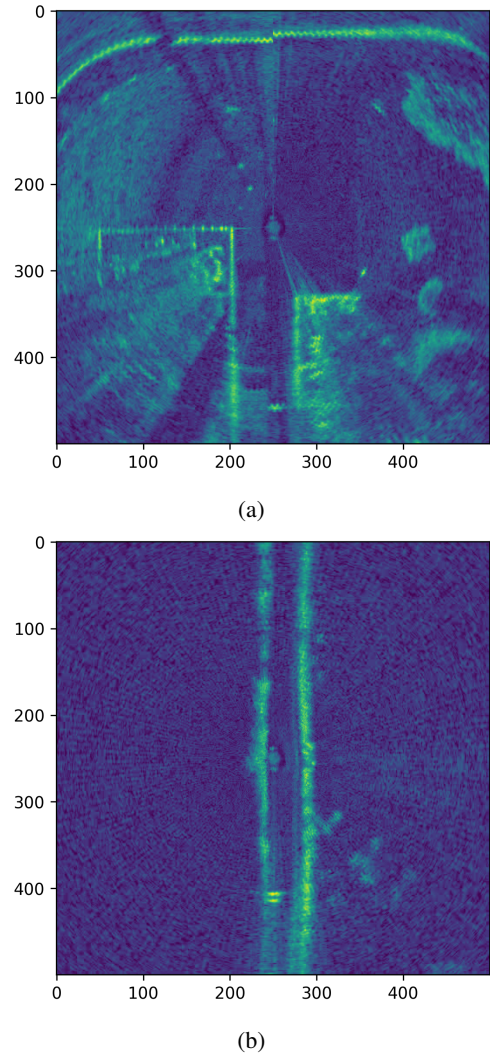


Fig. 3: Radar scans in the RDD show (a) structured and (b) unstructured scenes.

an expression is approximate as the radar signals suffer appearance change due to noises and artefacts.

A correlative scan-matching approach can find the optimum  $[R^*|t^*]$  as the transformation that maximises the correlation between the two signals  $Z_{t-1}$  and  $Z_t$ , such as:

$$[R^*|t^*] = \arg \max_{[R|t] \in \text{SO}(2)} T_{[R|t]}(Z_{t-1}) \star Z_t \quad (3)$$

where  $\star$  is the correlation operation. We follow the work of [22], which decouples the correlation operation for the rotation retrieval  $R^*$  and the translation  $t^*$  by exploiting the intrinsic translational invariance of the Fourier Transform. This procedure has huge advantages in run-time performances and memory efficiency compared to the original work [21] which performed an exhaustive brute-force search.

Importantly, the system is based on the assumption that  $Z_t \approx T_{[R|t]}(Z_{t-1})$ ; yet, this is challenged by the very nature of radar, where noise, occlusions or multi-path reflections can make the two scans, although taken close spatially, very different visually. To mitigate such an issue, [21], [22]

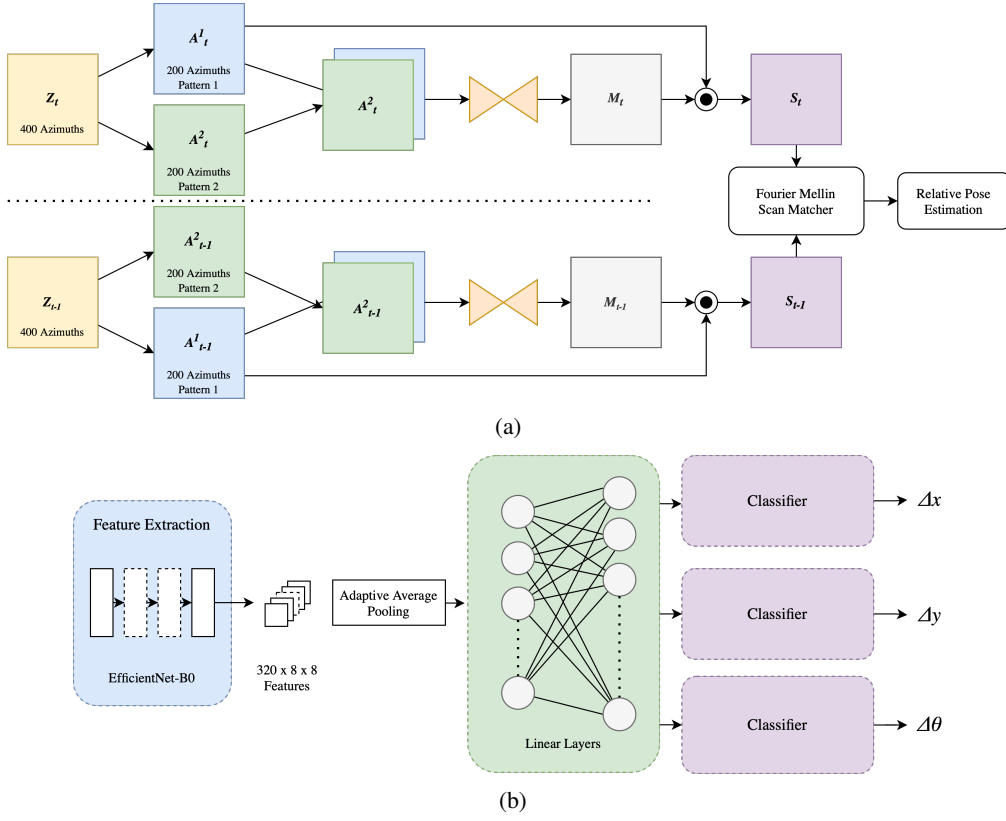


Fig. 4: (a) Our Doppler-aware scan matching generates a mask from the 2-channel images of the two modulations for both time steps. It is then applied to the modulation-pattern-1 scans via element-wise product. The resultant scans are used in the scan-matching process. (b) DAVE-NN is composed of a feature extractor and three regression heads. These are implemented to solve regression through an intermediate classification step.

introduce a learnt masking process to eliminate portions of the scans that would make the motion estimation inaccurate.

Let  $G$  be an encoder-decoder network that returns a mask scoring each pixel of a radar scan with a scalar in the range  $[0, 1]$ . The masking network can be applied to  $Z_{t-1}$  and  $Z_t$  to retrieve the relative masks  $M_{t-1}$  and  $M_t$ , which can then be multiplied element-wise to the raw scans to retrieve  $S_{t-1}$  and  $S_t$ ; these latter can be used then in the scan-matching pipeline to retrieve  $[R|t]$  from eq. (3). Importantly, as the scan-matching pipeline is completely differentiable,  $G$  can be learnt from an odometry signal through  $L_1$  loss.

Whereas the original pipeline used a concatenation of  $Z_{t-1}$  and  $Z_t$  as input to  $G$ , we propose adopting a Siamese network architecture to inform the network through Doppler information; this pipeline can be seen in Figure 4a. Specifically,  $G$  receives  $Z_{t-1}$  and  $Z_t$  singularly to create  $M_{t-1}$  and  $M_t$  as 2-channel images, where the first channel contains the 200 azimuths modulated with pattern 1 and the second with pattern 2 (see Figure 2). The resulting masks are then applied to the first channel to create  $S_{t-1}$  and  $S_t$ . This approach allows the network to exploit Doppler information implicitly from the raw signals while maintaining the benefits from a fully-differentiable approach and consistency in terms of pattern modulation when cross-correlating the radar images.

### B. Doppler-Aware Velocity Estimation (DAVE-NN)

We propose to extract motion estimation directly from one radar scan containing Doppler information. Whereas the previous experiments relied on geometry matching, this approach regresses the pose  $[R|t]$  end-to-end from a *single* 2-channel Doppler radar scan, which naturally exposes the network to the “zig-zag” patterns essential to extracting odometry information.

Our Doppler-aware Velocity Estimation network (DAVE-NN) is directly inspired from [26], where the authors use an intermediate classification step to regress the head orientations of users. In the same way, we trained our network to learn odometry by using an  $L_1$  loss on the softmax-weighted average of the classification results. Differently than [26], we did not apply a classification loss as the system did not empirically achieve better results.

Notably, we decided to include all three components of the motion estimation  $t_x$ ,  $t_y$  and  $\theta$  to test the ability of the network to retrieve them from Doppler. Physically, only translation can be estimated from it, as the velocity shift caused by Doppler happens only in the radial direction. Nevertheless, we included the rotation to test if other effects, such as the first-last azimuth seam seen in Figure 3a, can lead to an orientation estimation.

### C. Estimation fusion

To combine the benefits of scan-matching and direct regression using Doppler information, we also propose an approach that fuses both methods. This is done using a complementary filter approach, where confidences are extracted from both methods by considering the sharpness of the pose estimate, which is then used to weight the translation estimate from each method. We limit ourselves in the forward dimension of motion for the estimate fusion, although the approach can be extended to the lateral direction as well.

The different estimation procedures encode their confidence in the longitudinal pose estimate differently. The scan-matching approach outputs a correlation volume  $C_{xy} \in \mathbb{R}^{n \times n}$  over possible translations ( $t_x$  and  $t_y$ ). DAVE-NN outputs a logits vector  $l_x \in \mathbb{R}^b$ , containing classification scores for all  $b$  possible values of  $t_x$ .

Firstly, the correlation volumes are reduced over the  $y$ -dimension with a max operation, followed by a bilinear interpolation to the same dimensionality as the logits, resulting in  $C_x \in \mathbb{R}^b$ . Correlation and logit values are distributed differently, so they are both normalised to have zero-mean and unit variance. The confidence values for each are calculated as  $c_S = \max[\sigma_{\tau_S}(C_x)]$  and  $c_D = \max[\sigma_{\tau_D}(l_x)]$  for scan-matching and direct regression respectively, where  $\sigma_\tau$  is the softmax function with a given temperature  $\tau$ .

In a complementary filter, the estimate weights must sum to 1; therefore, the weights are calculated as  $[w_S, w_D] = \sigma_{\tau_w}([c_S, c_D])$ . Finally,  $t_x$  is calculated as follows from the scan-matching estimate  $[t_x]_S$  and direct regression estimate  $[t_x]_D$ :

$$t_x = w_S[t_x]_S + w_D[t_x]_D \quad (4)$$

## V. EXPERIMENTAL SETUP

In this section, we report details on the implementation and training regimes and our evaluation methodology.

### A. Implementation details

Our approach is implemented using the PyTorch library, more specifically though MMSegmentation<sup>2</sup> and TIMM<sup>3</sup>, which provide a collection of popular deep-learning modules. We use an Adam optimiser for training both networks.

As backbones for the masking network  $G$  and DAVE-NN, we employed EfficientNet-B0 [27].  $G$  then decodes the embedding using an Atrous Spatial Pyramid Pooling (ASPP) head [28], while DAVE-NN uses five linear layers of sizes [512, 1024, 512, 256, 128] with ReLU activation followed by three separate heads for each variable, with sizes [256, 127]. The output is a classification score, which corresponds to the values in the ranges  $[-1, 5]$  m for  $t_x$ ,  $[-1, 1]$  m for  $t_y$  and  $[-0.5, 0.5]$  rad for  $\theta$ . For both networks, we input radar images of size  $255 \times 255$ .

For determining the temperature values it is observed that, in most scenes, scan-matching is typically more accurate, therefore the temperature values are weighted towards this:

<sup>2</sup>github.com/open-mmlab/mms Segmentation

<sup>3</sup>github.com/rwightman/pytorch-image-models

$\tau_S = 1$ ,  $\tau_D = 1.2$ . In order to reject pose estimation failure,  $\tau_w$  is set to 0.0001, thereby severely down-weighting unconfident estimates.

### B. Datasets

In conjunction with RDD, we use the Radar RobotCar Dataset Radar RobotCar Dataset (RRCD) [3] to validate our masking pipeline and understand the challenging nature of RDD. For RRCD, we employ the training and validation routes designed in [29] and will report the results on the test split of the run 2019-01-10-12-32-52. Instead, results on RDD are reported on the test split of the run log\_13, composed of a rural driving scene with moderate traffic. This run includes a narrow tunnel of trees – shown in Figure 3b – which challenges geometrical scan matches.

### C. Baselines and Metrics

To evaluate our Doppler-aware scan-matching method, 2-channel-doppler, and the fused system, fused, we compare them against four baselines:

- 1) raw: this baseline does not apply masking to the radar scans.
- 2) no-doppler-400: this baseline mimics the original work [22] with the updated masking network to verify the goodness of the results.
- 3) no-doppler-200: the same network is now applied to scans with a reduced azimuth number: for RRCD we downsample the azimuth direction by two and for RDD we extract only the azimuths with one modulation. The rationale behind this baseline is to understand how much Doppler information does benefit the process.
- 4) full-scan: this baseline directly applies the masking network and cross-correlation process to the radar scan without splitting the two modalities into separate channels. This baseline shows how a more naive approach does not benefit the system.

We use the odometry metrics proposed for the KITTI odometry benchmark [30], i.e. rotational and translational errors, as the main indicators of the accuracy of the odometry algorithms.

## VI. EXPERIMENTAL RESULTS

Here, we discuss the results of this work.

### A. Reduced Azimuth number

Table I shows the results of our pipeline applied to the RRCD to validate our segmentation network. Indeed, the original work employed a UNet, while we used a DeepLabv3 network [28]. Our system reaches comparable odometry results with the original implementation, even when the number of azimuths is halved to resemble the RDD. On this note, we can see how the same network, trained and tested on the RDD, reaches far worse results, highlighting how challenging the new dataset is for a geometric scan-matching process. We blame two main factors: the reduced training dataset compared to the RRCD and the feature-poor rural scenery, as exemplified in Figure 3b.

| Model                        | Rot. error °/km | Trans. Error % |
|------------------------------|-----------------|----------------|
| Original results [22] – RRCD | 6.30            | 2.00           |
| raw – RRCD                   | 17.46           | 2.42           |
| no-doppler-400 – RRCD        | 8.06            | 1.38           |
| no-doppler-200 – RRCD        | 5.56            | 0.91           |
| raw – RDD                    | 34.83           | 58.35          |
| no-doppler-200 – RDD         | 32.19           | 28.03          |

TABLE I: KITTI metrics to validate our approach, evaluating the test sets of the RRCD and RDD. We achieve similar performances to the reported accuracy values of the original work [22], even with half the number of azimuths, to allow direct comparison to the RDD. Here, we see stark worse results due to the fewer training data points and the more challenging scenery.

| Model             | Rot. Error °/km | Trans. Error |
|-------------------|-----------------|--------------|
| raw               | 34.83           | 58.35        |
| no-doppler-200    | 32.19           | 28.03        |
| full-scan         | 35.96           | 19.20        |
| 2-channel-doppler | 35.83           | 5.98         |
| fused             | 35.83           | 4.59         |

TABLE II: KITTI metrics to evaluate the benefits of using Doppler information. We can see how the usage of Doppler can enhance the accuracy results, but only if used taking into consideration the nature of the scan-matching process.

### B. Doppler-aware masking network selection

Table II shows the results of using Doppler information during motion estimation. We can see how using Doppler information naively does not bring any benefit, while our approach reduced the translation error three-fold. The estimates from the proposed approach are shown in Figure 5a against using the raw scans with no Doppler information. The benefit is clear: the forward motion estimation using raw scans could not follow the ground truth for most of the run, while the proposed approach exhibits high accuracy. Nevertheless, many failure points are present, due to the inability of the geometric approach to deal with ambiguous scenery, as the narrow tunnel depicted in Figure 3b.

### C. Doppler-only motion estimation

Figure 5b show the qualitative results for applying DAVE-NN to estimate odometry in the test set. Firstly we notice how the forward motion estimation does follow the ground truth, although underestimating it slightly; nevertheless, most importantly, no catastrophic failures are present, as happens during scan matching. Instead, the angular estimation is not working; as expected, the Doppler information cannot assist in estimating the rotation and the other artefacts are not sufficient either. For this reason, we do not report any metrics as they would be unfair to the system.

### D. Estimation fusion

Using an estimate of the uncertainty to fuse the odometry estimation for the forward motion improves the results in terms of failure mitigation. Indeed, Figure 6 shows how all

the catastrophic failures have been removed, while Table II shows how the fused approach – where only  $f_x$  is fused – has slightly lower accuracy than masking alone. Indeed, DAVE-NN gives less accurate results overall, which, when integrated over time, lead to slightly worse metrics regardless of the improved robustness.

## VII. CONCLUSIONS

We studied the beneficial effect that Doppler information contained in FMCW scanning radar scans can have on RO. Here, we propose to treat Doppler as an additional channel in helping a masking function clean radar data before applying correlative scan matching to recover the transformation between two scans. Moreover, we propose a purely-regression approach to extract translational estimates from *single* scans with Doppler information. The first approach improved accuracy compared to the same input without Doppler information, whereas the second showed high robustness to challenging scenes, such as narrow tunnels. We finally fuse these two pieces of information using intrinsic uncertainties to maximise robustness and accuracy. Our methodology is validated on a custom dataset, RDD, composed of FMCW radar scans with Doppler information and odometry ground-truth signal, which is released alongside this work.

## REFERENCES

- [1] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung, *et al.*, “Boreas: A multi-season autonomous driving dataset,” *The International Journal of Robotics Research*, 2023.
- [2] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, “Mulran: Multimodal range dataset for urban place recognition,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [3] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, “The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [4] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, “Radiate: A radar dataset for automotive perception in bad weather,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [5] M. Gadd, D. De Martini, and P. Newman, “Contrastive learning for unsupervised radar place recognition,” in *International Conference on Advanced Robotics (ICAR)*, 2021.
- [6] D. De Martini, M. Gadd, and P. Newman, “Kradar++: Coarse-to-fine fmcw scanning radar localisation,” *Sensors*, 2020.
- [7] Ș. Săftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, “Kidnapped radar: Topological radar localisation using rotationally-invariant metric learning,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [8] P. Kaul, D. De Martini, M. Gadd, and P. Newman, “Rss-net: Weakly-supervised multi-class semantic segmentation with fmcw radar,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2020.
- [9] R. Aldera, D. D. Martini, M. Gadd, and P. Newman, “Fast radar motion estimation with a learnt focus of attention using weak supervision,” in *International Conference on Robotics and Automation (ICRA)*, 2019.
- [10] M. Broome, M. Gadd, D. De Martini, and P. Newman, “On the road: Route proposal from radar self-supervised by fuzzy lidar traversability,” *AI*, vol. 1, no. 4, pp. 558–585, 2020.
- [11] D. Williams, D. De Martini, M. Gadd, L. Marchegiani, and P. Newman, “Keep off the grass: Permissible driving routes from radar with weak audio supervision,” in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2020.
- [12] T. Y. Tang, D. De Martini, S. Wu, and P. Newman, “Self-supervised learning for using overhead imagery as maps in outdoor range sensor localization,” *The International Journal of Robotics Research*, 2021.

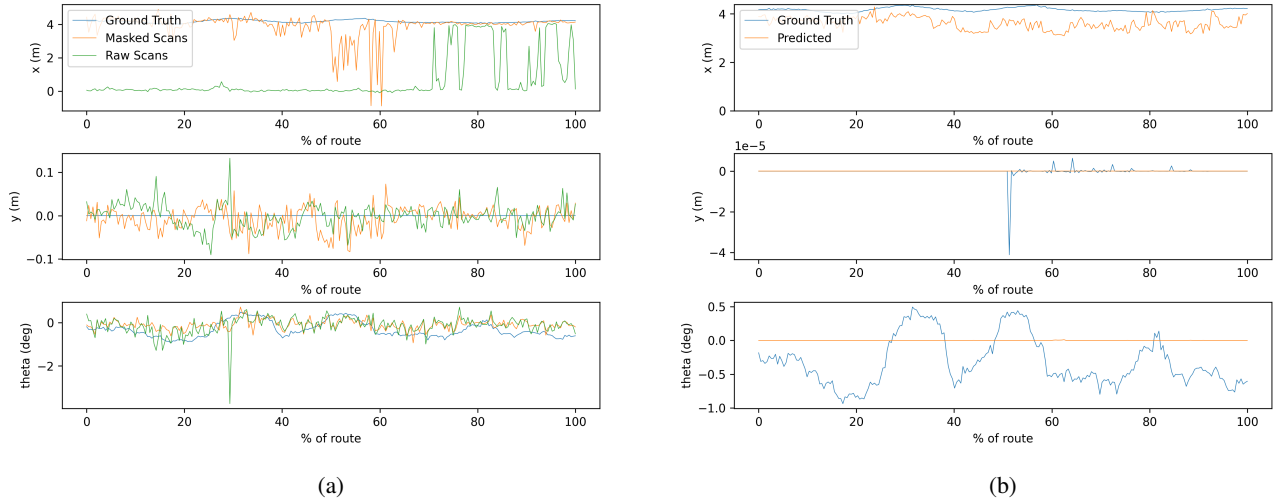


Fig. 5: Odometry signals for the test set of the RDD for our (a) Doppler-aware masking network compared with using only raw scans and (b) DAVE-NN. The masking process is clearly improving the cross-correlation odometry estimation; yet, many failure points are present in the route (around 50% through it). This is due to feature-poor scenery as in Figure 3b. Instead, DAVE-NN recovers lower-quality odometry estimations in terms of accuracy but does not present any clear failure point. On the other hand, as expected, it cannot estimate angular displacements.

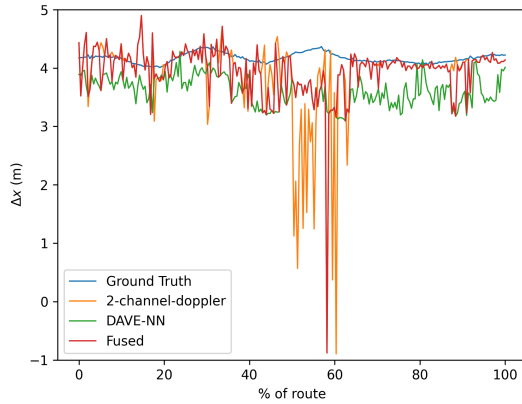


Fig. 6: Odometry signal for the two separate systems and fused together. It is noticeable how all the catastrophic failures have been removed.

- [13] T. Y. Tang, D. De Martini, and P. Newman, “Point-based metric and topological localisation between lidar and overhead imagery,” *Autonomous Robots (AURO)*, 2023.
- [14] S. H. Cen and P. Newman, “Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [15] S. H. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *International Conference on Robotics and Automation (ICRA)*, 2019.
- [16] M. Adams, E. Jose, and B.-N. Vo, *Robotic Navigation and Mapping with Radar*. 2012.
- [17] A. Alhashimi, D. Adolfsson, M. Magnusson, H. Andreasson, and A. J. Lilienthal, “Bfar-bounded false alarm rate detector for improved radar odometry estimation,” *arXiv preprint arXiv:2109.09669*, 2021.
- [18] D. Adolfsson, M. Magnusson, A. W. Alhashimi, A. J. Lilienthal, and H. Andreasson, “CFEAR radarodometry - conservative filtering for

- efficient and accurate radar odometry,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.
- [19] D. Barnes and I. Posner, “Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar,” in *IEEE international conference on robotics and automation (ICRA)*, 2020.
- [20] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, “Radar odometry combining probabilistic estimation and unsupervised feature learning,” in *Robotics: Science and Systems (RSS)*, 2021.
- [21] D. Barnes, R. Weston, and I. Posner, “Masking by moving: Learning distraction-free radar odometry from pose information,” in *Conference on Robot Learning (CoRL)*, 2019.
- [22] R. Weston, M. Gadd, D. D. Martini, P. Newman, and I. Posner, “Fast-mbyim: Leveraging translational invariance of the fourier transform for efficient and accurate radar odometry,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.
- [23] K. Burnett, A. P. Schoellig, and T. D. Barfoot, “Do we need to compensate for motion distortion and doppler effects in spinning radar navigation?,” *IEEE Robotics and Automation Letters*, 2021.
- [24] Y. Zhuang, B. Wang, J. Huai, and M. Li, “4d iriom: 4d imaging radar inertial odometry and mapping,” *IEEE Robotics and Automation Letters*, 2023.
- [25] Y. Wu, D. J. Yoon, K. Burnett, S. Kammel, Y. Chen, H. Vhavle, and T. D. Barfoot, “Picking up speed: Continuous-time lidar-only odometry using doppler velocity measurements,” *IEEE Robotics and Automation Letters*, 2022.
- [26] N. Ruiz, E. Chong, and J. M. Rehg, “Fine-grained head pose estimation without keypoints,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018.
- [27] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, 2019.
- [28] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [29] T. Y. Tang, D. De Martini, D. Barnes, and P. Newman, “Rsl-net: Localising in satellite images from a radar on the ground,” *IEEE Robotics and Automation Letters*, 2020.
- [30] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *IEEE conference on computer vision and pattern recognition*, 2012.