

# Look Around You: Sequence-based Radar Place Recognition with Learned Rotational Invariance

Matthew Gadd, Daniele De Martini, and Paul Newman

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

{mattgadd, danielle, pnnewman}@robots.ox.ac.uk

**Abstract**—This paper details an application which yields significant improvements to the adeptness of place recognition with Frequency-Modulated Continuous-Wave radar – a commercially promising sensor poised for exploitation in mobile autonomy. We show how a rotationally-invariant metric embedding for radar scans can be integrated into sequence-based trajectory matching systems typically applied to videos taken by visual sensors. Due to the complete horizontal field of view inherent to the radar scan formation process, we show how this off-the-shelf sequence-based trajectory matching system can be manipulated to detect place matches when the vehicle is travelling down a previously visited stretch of road in the opposite direction. We demonstrate the efficacy of the approach on 26 km of challenging urban driving taken from the largest radar-focused urban autonomy dataset released to date – showing a boost of 30 % in recall at high levels of precision over a nearest neighbour approach.

**Index Terms**—radar, localisation, place recognition, deep learning, metric learning

## I. INTRODUCTION

In order for autonomous vehicles to travel safely at higher speeds or operate in wide-open spaces where there is a dearth of distinct features, a new level of robust sensing is required. Frequency-Modulated Continuous-Wave (FMCW) radar satisfies these requirements, thriving in all environmental conditions (rain, snow, dust, fog, or direct sunlight), providing a 360° view of the scene, and detecting targets at ranges of up to hundreds of metres with centimetre-scale precision. Indeed, there is a burgeoning interest in exploiting FMCW radar to enable robust mobile autonomy, including ego-motion estimation [1]–[6], localisation [7], [8], and scene understanding [9].

Figure 1 shows an overview of the pipeline proposed in this paper which extends our recent work in extremely robust radar-only place recognition [7] in which a metric space for embedding polar radar scans was learned, facilitating topological localisation using nearest neighbour (NN) matching. We show that this learned metric space can be leveraged within a sequenced-based topological localisation framework to bolster matching performance by both mitigating visual similarities that are caused by the planarity of the sensor and failures due to sudden obstruction in dynamic environments. Due to the complete horizontal field of view (FoV) of the radar scan formation process, we show how the off-the-shelf sequence-based trajectory matching system can be manipulated to allow us to detect place matches when the vehicle is travelling down a previously visited stretch of road in the opposite direction.

This paper proceeds by reviewing related literature in Sec-

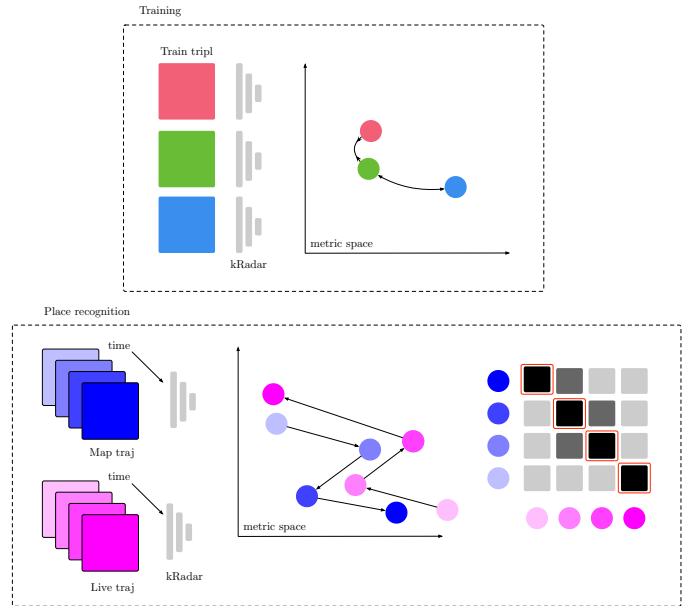


Figure 1: An overview of our pipeline. The offline stages include *enforcing* a metric space by training a Fully Convolutional Neural Network (FCNN) which takes polar radar scans as input, and *encoding* a trajectory of scans (the map) by forward passes through this network (c.f. Section III-B). The online stages involve *inference* to represent the place the robot currently finds itself within in terms of the learned knowledge and *querying* the space (c.f. Section III-A) which – in contrast to our prior work – involves a search for coherent sequences of matches rather than a globally closest frame in the embedding space.

tion II. Section III describes our approach for a more canny use of a metric space in which polar radar scans are embedded. We describe in Section IV details for implementation, evaluation, and our dataset. Section V discusses results from such an evaluation. Sections VI and VII summarise the findings and suggest further avenues for investigation.

## II. RELATED WORK

Recent work has shown the promise of FMCW radar for robust place recognition [7], [10] and metric localisation [6]. None of these methods account for temporal effects in the radar measurement stream.

SeqSLAM [11] and its variants have been extremely successful at tackling large-scale, robust place recognition with

video imagery in the last decade. Progress along these lines has included automatic scaling for viewpoint invariance [12], probabilistic adjustments to the search technique [13], and dealing with challenging visual appearance change using Generative Adversarial Networks (GANs) [14].

The work presented in this paper is most closely influenced by the use of feature embeddings learned by training Convolutional Neural Networks (CNNs) [15], omnidirectional cameras [16], and Light Detection and Ranging (LiDAR) [17] within the SeqSLAM framework.

### III. METHODOLOGY

Broadly, our method can be summarised as leveraging very recent results in Deep Learning (DL) techniques which provide good metric embeddings for the global location of radar scans within a robust sequence-based trajectory matching system. We begin the discussion with a brief overview of the baseline SeqSLAM algorithm, followed by a light description of the learned metric space, and concluded by an application which unifies these systems – the main contribution of this paper.

#### A. Overview of SeqSLAM

Our implementation of the proposed system is based on an open-source, publicly available port of the original algorithm<sup>1</sup>.

Incoming images are preprocessed by downsampling (to thumbnail resolution) and patch normalisation. A difference matrix is constructed storing the euclidean distance between all image pairs. This difference matrix is then contrast enhanced. Examples of these matrices can be seen in ???. For more detail, a good summary is available in [18].

When looking for a match to a query image, SeqSLAM sweeps through the contrast-enhanced difference matrix to find the best matching sequence of adjacent frames.

In the experiments (c.f. Sections IV and V) we refer to this baseline search as SEQSLAM.

#### B. Overview of Kidnapped Radar

To learn filters and cluster centres which help distinguish polar radar images for place recognition we use NetVLAD [19] with VGG-16 [20] as a front-end feature extractor – both popularly applied to the place recognition problem. Importantly, we make alterations such that the network invariant to the orientation of input radar scans, including: circular padding [21], anti-aliasing blurring [22] azimuth-wise max-pooling.

To enforce the metric space, we perform online triplet mining and apply the triplet loss described in [23]. Loop closure labels are taken from a ground truth dataset (c.f. Section IV).

The interested reader is referred to [7] for more detail.

In the experiments (c.f. Sections IV and V) we refer to representations obtained in this manner as KRADAR.

#### C. Sequence-based Radar Place Recognition

We replace the image preprocessing step with inference on the network described in Section III-B, resulting in radar scan

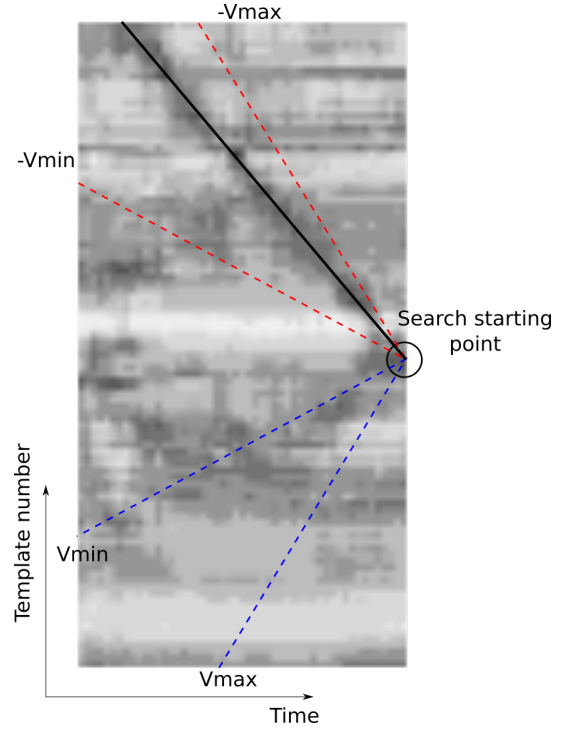


Figure 2: The off-the-shelf sequence matching SeqSLAM system is manipulated in this paper to facilitate backwards Loop Closure Detection (LCD). This is achieved by mirroring the set,  $v_{min} < v < v_{max}$  (blue), of trajectories considered – also considering  $-v_{max} < v < -v_{min}$  (red). Importantly, this is not a useful adjustment under a naïve application of SeqSLAM to radar images and is only beneficial if a rotationally invariant representation is used to construct the difference matrices.

descriptors of size 4096.

The difference matrix is obtained by calculating the Euclidean distance between every pair of embeddings taken from places along the reference and live trajectories in a window of length  $W$ . This distance matrix is then locally contrast enhanced in sections of length  $R$ .

When searching for a match to a query image, we perform a sweep through this contrast-enhanced difference matrix to find the best matching sequence of frames based on the sum of sequence differences. In order to be able to detect matches in reverse, this procedure is repeated with a time-reversed live trajectory – this method would not be applicable to narrow FoV cameras but is appropriate here as the radar has a 360° FoV.

A simple visualisation of the process is shown in Figure 2. In this case, the forward search (blue lines) is mirrored to perform a backwards search (red lines), which results in the selection of the best match (solid black line). In the experiments (c.f. Sections IV and V) we refer to this modified search as LAY (“Look Around You”).

This procedure is performed for each template, on which a threshold is applied to select the best matches. Section V

<sup>1</sup><https://github.com/tmadl/pySeqSLAM>

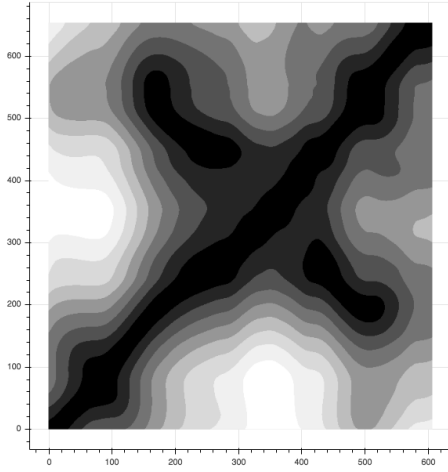


Figure 3: Visualisation of a ground truth  $SE(2)$  matrix showing the Euclidean distance between the global positions that pairs of radar scans were captured at. Each trajectory pair is associated with such a matrix. Values in these matrices are scaled from distant (white) to nearby (black). In this region of the dataset, the vehicle revisits the same stretch of the route in the opposite direction – visible as the contours perpendicular to the main diagonal.

discusses the application of the threshold and reports the results in comparison to the original SeqSLAM approach; in particular Figure 4 shows visual examples of the discussed methodology.

#### IV. EXPERIMENTAL SETUP

This section details our experimental design in obtaining the results to follow in Section V.

##### A. Vehicle and radar specifications

Data was collected using the *Oxford RobotCar* platform [24]. The vehicle, as described in the *Oxford Radar RobotCar Dataset* [25], is fitted with a CTS350-X Navtech FMCW scanning radar.

##### B. Ground truth database

The ground truth database is curated offline to capture the sets of nodes that are at a maximum distance (15m) from a query frame, creating a graph-structured database that yields triplets of nodes for training the representation discussed in Section III-B.

To this end, we adjust the accompanying ground truth odometry described in [25] in order to build a database of ground truth locations. We manually selected a moment during which the vehicle was stationary at a common point and trimmed each ground trace accordingly. We also aligned the ground traces by introducing a small rotational offset to account for differing attitudes.

##### C. Trajectory reservation

Each approximately 9km trajectory in the Oxford city centre was divided into three distinct portions: *train*, *valid*, and *test*.

The network is trained with ground truth topological localisations between two reserved trajectories in the *train* split.

The *test* split, upon which the results presented in Section V are based, was specifically selected to feature vehicle traversals over portions of the route in the opposite direction; data from this split are not seen by the network during training.

The results focus on a teach-and-repeat (TR) scenario, in which all remaining trajectories in the dataset are localised against a map built from the first trajectory that we did not use for learning, totalling 27 trajectory pairs (and 26km of driving) with the same map but a different localisation run.

##### D. Measuring performance

In the ground truth  $SE(2)$  database, all locations within a 15m radius of a ground truth location are considered true positives whereas those outside are considered true negatives, a more strictly imposed boundary than in [7].

Evaluation of Precision-Recall (PR) is different for the sequence- and NN-based approaches. For the NN-based approach of [7] we perform a ball search of the discretised metric space out to a varying embedding distance threshold. For the sequence-based approach advocated in this paper, we vary the minimum match score.

As useful summaries of Precision-Recall (PR) performance, we analyse area-under-curve (AUC) as well as some F-scores, including  $F_1$ ,  $F_2$ , and  $F_\beta$  with  $\beta = 0.5$  [26].

##### E. Hyperparameter tuning

To produce a fair comparison of the different configurations we can utilise to solve the topological localisation problem, we performed a hyperparameter tuning on the various algorithms; we selected two random trials and excluded them from the final evaluation. The window width for the trajectory evaluation  $W$  and the enhancement window  $R$  have been chosen through a grid search procedure. The final values are the ones which produced precision-recall curves with the highest value of precision at 80% recall.

#### V. RESULTS

This section presents instrumentation of the metrics discussed in Section IV-D.

The hyperparameter optimisation (c.f. Section IV-E) results in the parametisation of the systems for comparison as enumerated in Table I.

Figure 6 shows a family of PR curves for various methods that it is possible to perform SeqSLAM with radar data. Here, only a single trajectory pair is considered (one as the map trajectory, the other as the live trajectory). From Figure 6 it is evident that:

- 1) Performance when using the learned metric embeddings is superior to either polar or cartesian radar scans,
- 2) Sequence-based matching of trajectories outperforms NN-based searches,

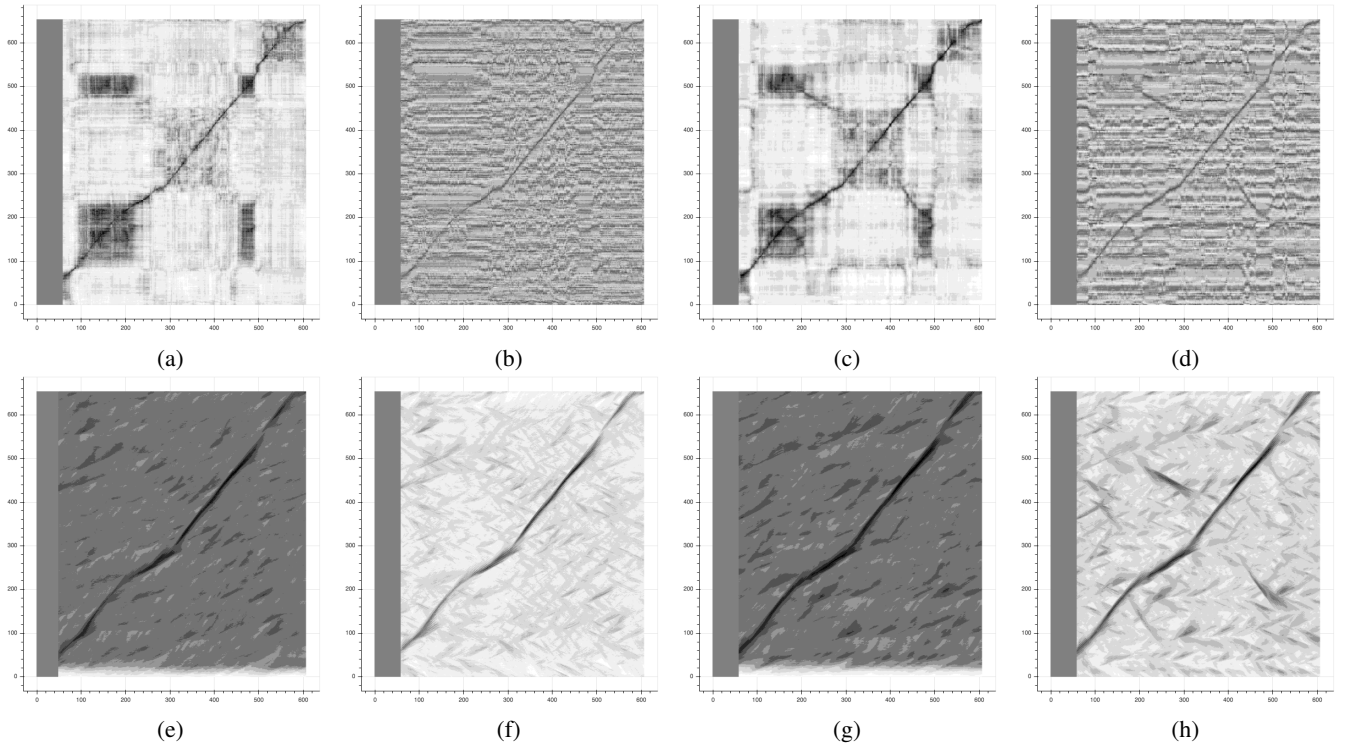


Figure 4: Difference matrices upon which the SeqSLAM variants perform trajectory searches (top row) and relative match-score-matrices (bottom row). These are constructed matching the representation of radar scans in two trajectories (rows-versus-columns for each matrix) – VGG-16/NETVLAD on the left side (a, b, e and f) and KRADAR on the right side (c, d, g and h). (a) and (c) are the difference matrices before enhancement – directly used by the NN search in [7] – and (b) and (d) are the respective enhanced form – on which SeqSLAM performs its searches. (e) and (f) are constructed using embeddings inferred by VGG-16/NETVLAD in the enhanced form (b), while (g) and (h) use embeddings inferred by KRADAR in the enhanced form (d). (e) and (g) are computed by using the forward-style match score method employed by standard SeqSLAM; in contrast, (f) and (h) employ the proposed backward-style match score method. All match-score matrices are not defined for the first window of columns as SeqSLAM must fill a buffer of frames before any matching is possible.

Representation	Search	$R$	$W$
VGG-16/NETVLAD	SEQSLAM	34	50
KRADAR	SEQSLAM	37	60
VGG-16/NETVLAD	LAY	31	60
KRADAR	LAY	37	60

Table I: Hyperparameter summary as results of the hyperparameter grid-search optimisation.

- 3) Performance when using the baseline architecture is outstripped by the rotationally-invariant modifications, and
- 4) Performance when using the modified search algorithm is boosted.

Observation 1 can be attributed to the fact that the learned representation is designed to encode only knowledge concerning place, whereas the imagery is subject to sensor artefacts. Observation 2 can be attributed to perceptual aliasing along straight, canyon-like sections of an urban trajectory being

mitigated. Observation 3 can be attributed to the rotationally-invariant architecture itself. Observation 4 can be attributed to the ability of the adjusted search to detect loop closures in reverse.

Table II provides further evidence for these findings by aggregating PR-related statistics over the entirety of the dataset discussed in Section IV-C – the map trajectory is kept constant and the live trajectory varies over forays spanning a month of urban driving.

While it is clear that we outperform NN techniques presented in [7], the F-scores in Table II present a mixed result when comparing the standard SeqSLAM search and the modified search discussed in Section III-C. However, consider Figure 5. Here, the structure of the backwards loop closures is discovered more readily by the backwards search.

It is important to remember when inspecting the results shown in Table II and Figure 6 that the data in this part of the route (c.f. Section IV-C) is *unseen* by the network during training, and particularly challenging. This is a necessary analysis of the generalisation of learned place recognition methods but is not a requirement when deploying the learned

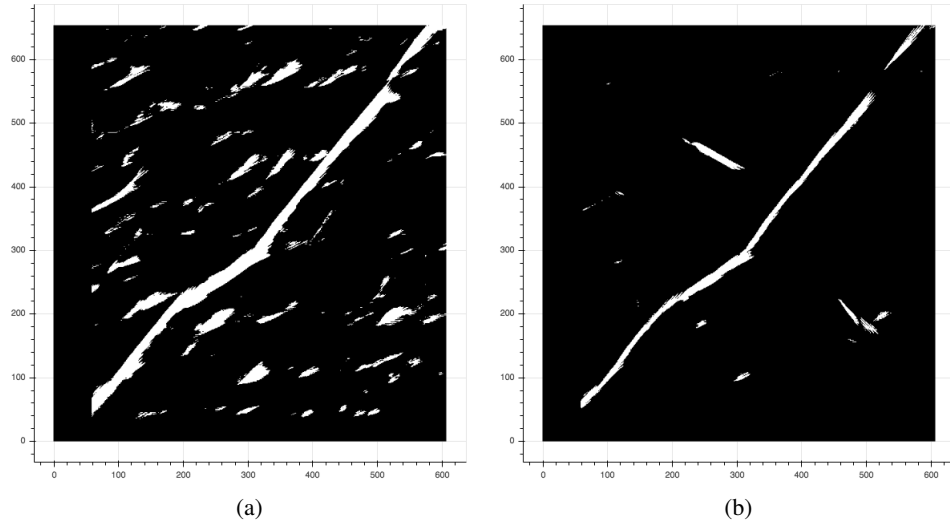


Figure 5: Binarised match score matrices for the (a) baseline and (b) mirrored SeqSLAM variants. The threshold applied for binarisation is higher for (a) (kRADAR,SeqSLAM) than for (kRADAR,LAY). This is in order to qualitatively show that even when increasing numbers of potential matches are allowed in SeqSLAM (high recall), the true backwards loop closures are not featured. For LAY (right), they are. From these it is evident that the tailored changes to the fundamental SeqSLAM search strategy are better suited to discover loop closures as the vehicle revisits the same route section with opposing orientation – a common scenario in structured, urban driving.

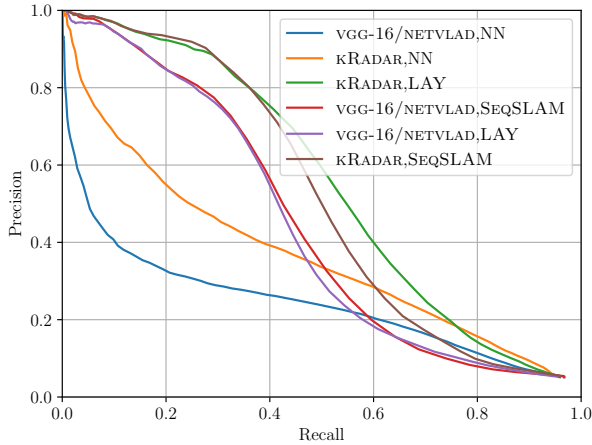


Figure 6: PR curves showing the benefit of, firstly, using learned metric embeddings as opposed to radar scans directly and, secondly, the tailored changes to the baseline SeqSLAM search algorithm when performing sequence-based radar place recognition.

knowledge in TR modes of autonomy.

The takeaway message is that we have improved the recall at good precision levels by about 30 % by applying sequence-based place recognition techniques to our learned metric space.

## VI. CONCLUSION

We have presented an application of recent advances in learning representations for imagery obtained by radar scan

formation to sequence-based place recognition. The proposed system is based on a manipulation of off-the-shelf SeqSLAM with prudent adjustments made taking into account the complete sweep made by scanning radar sensors. We have further proven the utility of our rotationally invariant architecture – a crucial enabling factor of our SeqSLAM variant. Crucially, we achieve a boost of 30 % in recall at high levels of precision over our previously published nearest neighbour approach.

## VII. FUTURE WORK

In the future we plan to retrain and test the system on the all-weather platform described in [27], a significant factor in the development of which was to explore applications of FMCW radar to mobile autonomy in challenging, unstructured environments. We also plan to integrate the system presented in this paper with our mapping and localisation pipeline which is built atop of the scan-matching algorithm of [28], [29].

## ACKNOWLEDGMENT

This project is supported by the Assuring Autonomy International Programme, a partnership between Lloyd’s Register Foundation and the University of York as well as UK EPSRC programme grant EP/M019918/1. We would also like to thank our partners at Navtech radar.

## REFERENCES

- [1] S. H. Cen and P. Newman, “Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.
- [2] S. H. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *2019 International Conference*

Representation	Search	AUC	max $F_1$	max $F_2$	max $F_{0.5}$	$R_{P=60\%}$	$R_{P=80\%}$
VGG-16/NETVLAD	NN	0.26	0.34	0.37	0.29	0.03	0.00
KRADAR	NN	0.37	0.41	0.40	0.41	0.16	0.06
VGG-16/NETVLAD	SeqSLAM	0.47	0.49	0.37	0.60	0.40	0.31
KRADAR	SeqSLAM	0.52	<b>0.53</b>	0.41	<b>0.64</b>	0.45	<b>0.37</b>
VGG-16/NETVLAD	LAY	0.46	0.48	0.36	0.60	0.39	0.32
KRADAR	LAY	<b>0.53</b>	<b>0.53</b>	<b>0.42</b>	0.62	<b>0.46</b>	0.36

Table II: Summary statistics for various radar-only SeqSLAM techniques (including representation of the radar frame and style of search) as aggregated over a month of urban driving. All quantities are expressed as a mean value. As discussed in Section IV-D, the requirement imposed on matches (as true/false positives/negatives) is more strict than that presented in [7] with consequentially worse performance than previously published for VGG-16/NETVLAD KRADAR, and NN systems. The key message of this paper is that sequence-based exploitation of these learned metric embeddings (middle and bottom rows) is beneficial in comparison to NN matching in a discretised search space (top two rows).

- on *Robotics and Automation (ICRA)*. IEEE, 2019, pp. 298–304.
- [3] R. Aldera, D. De Martini, M. Gadd, and P. Newman, “Fast Radar Motion Estimation with a Learnt Focus of Attention using Weak Supervision,” in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 1190–1196.
- [4] R. Aldera, D. De Martini, M. Gadd, and P. Newman, “What Could Go Wrong? Introspective Radar Odometry in Challenging Environments,” in *IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, October 2019.
- [5] 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.
- [6] D. Barnes and I. Posner, “Under the Radar: Learning to Predict Robust Keypoints for Odometry Estimation and Metric Localisation in Radar,” *arXiv preprint arXiv: 2001.10789*, 2020. [Online]. Available: <https://arxiv.org/abs/2001.10789>
- [7] Ş. Săftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, “Kidnapped Radar: Topological Radar Localisation using Rotationally-Invariant Metric Learning,” *arXiv preprint arXiv: 2001.09438*, 2020. [Online]. Available: <https://arxiv.org/abs/2001.09438>
- [8] T. Y. Tang, D. De Martini, D. Barnes, and P. Newman, “RSL-Net: Localising in Satellite Images From a Radar on the Ground,” *arXiv preprint arXiv:2001.03233*, 2020.
- [9] R. Weston, S. Cen, P. Newman, and I. Posner, “Probably Unknown: Deep Inverse Sensor Modelling Radar,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5446–5452.
- [10] 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.
- [11] M. J. Milford and G. F. Wyeth, “Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights,” in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1643–1649.
- [12] E. Pepperell, P. I. Corke, and M. J. Milford, “Automatic image scaling for place recognition in changing environments,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1118–1124.
- [13] P. Hansen and B. Browning, “Visual place recognition using hmm sequence matching,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 4549–4555.
- [14] Y. Latif, R. Garg, M. Milford, and I. Reid, “Addressing Challenging Place Recognition Tasks using Generative Adversarial Networks,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2349–2355.
- [15] B. Dongdong, W. Chaqun, B. Zhang, Y. Xiaodong, Y. Xuejun *et al.*, “Cnn feature boosted seqslam for real-time loop closure detection,” *Chinese Journal of Electronics*, vol. 27, no. 3, pp. 488–499, 2018.
- [16] R. Cheng, K. Wang, S. Lin, W. Hu, K. Yang, X. Huang, H. Li, D. Sun, and J. Bai, “Panoramic annular localizer: Tackling the variation challenges of outdoor localization using panoramic annular images and active deep descriptors,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 920–925.
- [17] P. Yin, Y. He, L. Xu, Y. Peng, J. Han, and W. Xu, “Synchronous adversarial feature learning for lidar based loop closure detection,” in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 234–239.
- [18] N. Sünderhauf, P. Neubert, and P. Protzel, “Are we there yet? challenging seqslam on a 3000 km journey across all four seasons,” in *Proc. of Workshop on Long-Term Autonomy, IEEE International Conference on Robotics and Automation (ICRA)*, 2013, p. 2013.
- [19] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “Netvlad: Cnn architecture for weakly supervised place recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.
- [20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [21] T.-H. Wang, H.-J. Huang, J.-T. Lin, C.-W. Hu, K.-H. Zeng, and M. Sun, “Omnidirectional CNN for Visual Place Recognition and Navigation,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2341–2348.
- [22] R. Zhang, “Making Convolutional Networks Shift-Invariant Again,” in *International Conference on Machine Learning (ICML)*, 2019, pp. 7324–7334.
- [23] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
- [24] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 Year, 1000km: The Oxford RobotCar Dataset,” *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.
- [25] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, “The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset,” *arXiv preprint arXiv: 1909.01300*, 2019. [Online]. Available: <https://arxiv.org/pdf/1909.01300>
- [26] J. A. Pino, “Modern information retrieval. ricardo baeza-yates y berthier ribeiro-neto addison wesley hario, england, 1999,” 1999.
- [27] S. Kyberd, J. Attias, P. Get, P. Murcutt, C. Prahacs, M. Towilson, S. Venn, A. Vasconcelos, M. Gadd, D. De Martini, and P. Newman, “The Hulk: Design and Development of a Weather-proof Vehicle for Long-term Autonomy in Outdoor Environments,” in *International Conference on Field and Service Robotics (FSR)*, Tokyo, Japan, August 2019.
- [28] S. H. Cen and P. Newman, “Precise ego-motion estimation with millimeter-wave radar under diverse and challenging conditions,” *Proceedings of the 2018 IEEE International Conference on Robotics and Automation*, 2018.
- [29] S. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Montreal, Canada, 2019.