

Off the Radar: Uncertainty-Aware Radar Place Recognition with Introspective Querying and Map Maintenance

Jianhao Yuan, Paul Newman, Matthew Gadd

Mobile Robotics Group, Department of Engineering Science, University of Oxford, UK

jianhao.yuan@oxfordrobotics.institute, {pnewman,mattgadd}@robots.ox.ac.uk

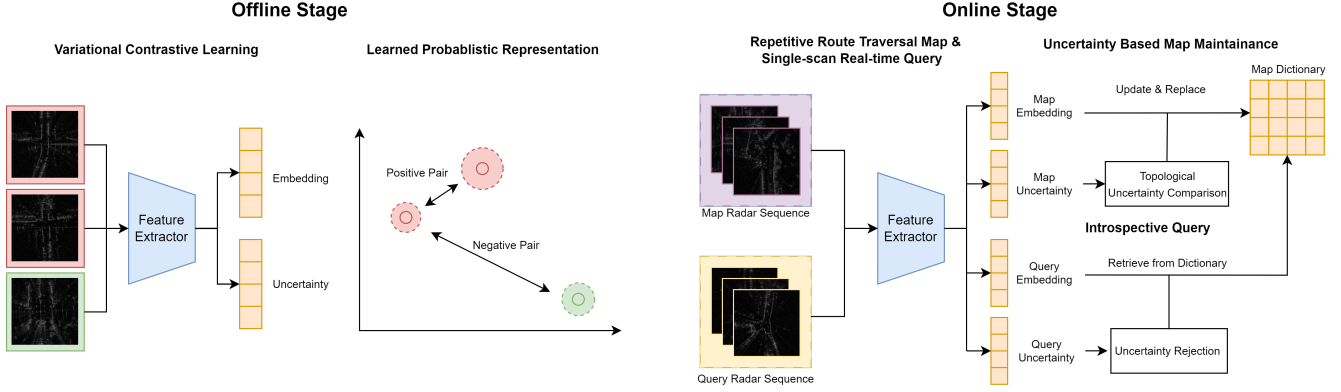


Fig. 1: Overview of system pipeline. In the offline stage, we use a variational contrastive learning framework (see Sec. III-A for full details) to learn a hidden embedding space with estimated uncertainty such that radar scans from similar topological locations are close to each other, and vice versa. In the online stage, we develop two uncertainty-based mechanisms to handle sequentially collected radar scans for inference and map construction. For repetitive traversal of the same routes, we actively maintain an integrated map dictionary by replacing highly uncertain scans with more certain ones. For query scans with low uncertainty, we retrieve the matching map scans from the dictionary based on metric space distance. In contrast, we reject prediction for high-uncertainty scans.

Abstract—Localisation with Frequency-Modulated Continuous-Wave (FMCW) radar has gained increasing interest due to its inherent resistance to challenging environments. However, complex artefacts of the radar measurement process require appropriate uncertainty estimation – to ensure the safe and reliable application of this promising sensor modality. In this work, we propose a multi-session map management system which constructs the “best” maps for further localisation based on learned variance properties in an embedding space. Using the same variance properties, we also propose a new way to introspectively reject localisation queries that are likely to be incorrect. For this, we apply robust noise-aware metric learning, which both *leverages* the short-timescale variability of radar data along a driven path (for data augmentation) and *predicts* the downstream uncertainty in metric-space-based place recognition. We prove the effectiveness of our method over extensive cross-validated tests of the *Oxford Radar RobotCar* and *MulRan* dataset. In this, we outperform the current state-of-the-art in radar place recognition and other uncertainty-aware methods when using only single nearest-neighbour queries. We also show consistent performance *increases* when rejecting queries based on uncertainty over a difficult test environment, which we did not observe for a competing uncertainty-aware place recognition system.

Index Terms—Radar, Place Recognition, Deep Learning, Uncertainty Estimation, Autonomous Vehicles, Robotics

I. INTRODUCTION

Place recognition and localisation are important tasks in the field of robotics and autonomous systems, as they enable a system to understand and navigate its environment.

Traditional vision-based methods for place recognition are often vulnerable to changes in environmental conditions such as lighting, weather, and occlusion, leading to performance degradation [1]. To address this issue, there has been increasing interest in the use of FMCW radar as a robust sensor substitute for such adversarial environments. Existing works have demonstrated the effectiveness of FMCW radar place recognition with hand-crafted [2]–[6] and learning-based feature extraction approaches [7]–[15]. Despite the success of existing works, the deployment of these methods in safety-critical applications such as autonomous driving is still limited by the lack of calibrated uncertainty estimation. While there have been works [16]–[19] that address the problem in similar areas such as image retrieval or visual place recognition, which leverage both Bayesian and learning-based approaches, there is no previous work specifically targeting uncertainty estimation for radar place recognition.

In this area, it is important to consider: (1) safety requires uncertainty estimation to be well-calibrated to false positive rate in order to enable *introspective* rejection; (2) real-time deployment requires *fast* single-scan uncertainty-based inference capability; (3) repetitive route traversal in long-term autonomy requires online *continual* map maintenance.

While the Variational Autoencoder (VAE) [20] is usually used for generative tasks, its probabilistic latent space can serve as an effective metric space representation for place recognition [21]–[23] and allow prior assumption on the data

noise distribution, which also gives a normalized aleatoric uncertainty estimation.

Thus, in this paper, to achieve reliable and safe deployment of FMCW radar in autonomous driving, we leverage a variational contrastive learning framework and propose a unified uncertainty-aware radar place recognition method as shown in Fig. 1. Our key contributions are as follows:

- 1) Uncertainty-aware metric learning framework for radar place recognition.
- 2) Introspective query mechanism based on false positive calibrated uncertainty estimation for real-time autonomous driving deployment.
- 3) Online recursive map maintenance and improvement mechanism for repeated traversals of changing environments in long-term autonomy.

In doing so we outperform the previous radar place recognition state-of-the-art [8] and show that our learned uncertainty is more suitable for query rejection than a previous approach [17] developed for vision but tested here in radar.

II. RELATED WORK

A. Radar Place Recognition

Place recognition can be viewed as a query and retrieval process, where a map is constructed with a dictionary of radar scans. For every new query scan, the algorithm needs to retrieve a scan from the dictionary such that samples from a similar topological location are close to each other, and vice versa. Traditional hand-crafted feature extraction methods, such as correlative scan matching [2], graph matching [3], [4], and scan context descriptor [6], have demonstrated the effectiveness of using radar perception for place recognition and localisation. Recently, learning-based methods have shown impressive performance. Supervised metric learning methods [7], [9]–[13], which exploit rotational invariance and spatial-temporal consistency in radar scans, have demonstrated remarkable performance. Then, Gadd *et al* [8] achieved comparable performance to supervised methods using contrastive learning. Moreover, multi-modal methods incorporating additional modalities, such as LIDAR points, through direct joint learning [15] or point-cloud registration [24], are also explored to aid radar-based recognition. Among numerous metric learning methods, the VAE structure is extensively explored. It uses variational inference to approximate the posterior distribution of latent representation and thus gives a probabilistic estimation on embedding with consideration of data noise. Lin *et al* [25] use VAE in metric learning for the visual classification task. Burnett *et al* [26] use a VAE-like structure with factor graph optimisation for radar odometry.

Inspired by previous works of the variational metric learning approach of Lin *et al* [25] and the contrastive learning approach of Gadd *et al* [8], we form the basis of our representation learning framework in a contrastive manner with a VAE-like structure.

B. Uncertainty Estimation

Uncertainty estimation aims to quantify the model prediction confidence, which is a crucial component in the

safety-critical application for the model to introspectively reject low confidence predictions. One line of work in similar areas, such as image retrieval and place recognition, uses the Bayesian approach such as Monte Carlo Dropout [16] to estimate the epistemic uncertainty, which, however, usually is very computationally expensive and hinders the deployment in real-time autonomous driving. Another line of work directly learns the uncertainty from data to estimate the aleatoric uncertainty. However, previous approaches such as PTL [19], BPE [18], and TAH [27] are restricted to specific pairwise or triplet loss, which lacks flexibility in offline system design. Later STUN [17] uses a knowledge distillation framework to break such constraints that adapt to arbitrary metric learning loss functions. Moreover, STUN [17] use a dynamic binning strategy that requires a large batch of test samples to determine a meaningful uncertainty rejection threshold at inference time. Here, only the relative uncertainty within a batch can be exploited. However, this is not suitable for real-time place recognition deployment as the entire test sequence is not known a priori. Thus, to tackle this problem, we develop a static binning strategy by leveraging the VAE variance prediction, such that the uncertainty estimated is compared against a pre-defined prior distribution.

C. Uncertainty-aware Radar Localisation

Although uncertainty estimation in radar place recognition tasks is still largely under-explored, there have been attempts to develop introspective radar systems for odometry. Adolfsson *et al* [28] present a radar system that verifies loop closure candidates introspectively by combining multiple place-recognition techniques, including tightly-coupled place similarity and odometry uncertainty search, creating loop descriptors from origin-shifted scans, and delaying loop selection until after verification. In contrast, our system does not rely on a secondary odometry estimation system, as uncertainty is inherent to the representation we learn for place recognition. Aldera *et al* [29] focus only on odometry, and use inertial sensors to learn an uncertainty classifier, which we do not.

III. METHOD

In this section, we present the proposed modules for the variational uncertainty-aware radar place recognition method including metric space learning (Sec. III-A), mapping (Sec. III-B), and querying (Sec. III-C). We explain how these modules are integrated to form a unified system.

A. Variational Contrastive Learning

The work in this section is both a key enabler of our core contributions, map maintenance and introspective querying – see below, but is also a novel integration of deep variational metric learning (DVML) [25] with radar place recognition [8], [14] and a new way to characterise uncertainty in place recognition. Indeed, we show in Sec. V that learning uncertainty in this way yields more calibrated introspection than in self-teaching uncertainty estimation (STUN) [17].

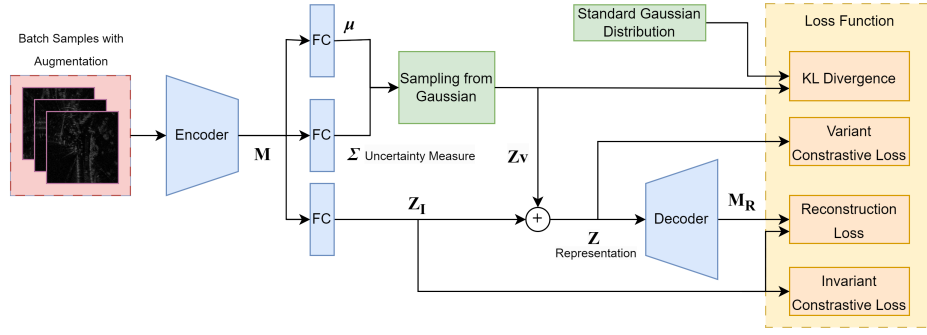


Fig. 2: Overview of Variational Contrastive Learning Framework, based on [25]. A metric space is learnt through an encoder-decoder structure with two reparameterised parts: a deterministic embedding for recognition and a set of parameter modelling multivariate Gaussian distribution with its variance serving as the uncertainty measure. The overall learning is jointly driven by both reconstruction and contrastive loss to ensure an informative and discriminative hidden representation for radar scans.

As illustrated in Fig. 2, we adopt a VAE structure to disentangle the radar scan embedding into a noise-induced variant part Z_V , which captures the variance of prediction-irrelevant uncertainty sources, and a semantic invariant part Z_I , for the essential features of the scene representation. The variant part is later sampled from a prior multivariate isotropic Gaussian distribution and added to the invariance part to form the overall representation $Z = Z_I + Z_V$. The variant output is directly used as the uncertainty measure.

The assumption is we only consider the aleatoric uncertainty in model prediction caused by the inherent ambiguity and randomness in data as the main source of uncertainty [30]. In particular for radar scanning, this is likely due to speckle noise, saturation and temporary occlusion. The standard metric learning approaches, regardless of the loss function chosen, tend to enforce identical embeddings between positive sample pairs while ignoring the underlying variances between them. However, this can lead to the model being insensitive to minor features and overfitting to the training distribution. Thus, to model the noise variance, we use the extra probabilistic variance output in the VAE structure to estimate the aleatoric uncertainty.

To establish such a noise-aware representation for radar perception, we use four loss functions to guide the overall training. We choose to demonstrate the benefit of our specific method of learning uncertainty in the setting of losses which feature increased numbers of *negative* examples – this having been proven in STUN [17] to perform best (i.e. quadruplet, with 2 negatives). In this area, the state-of-the-art in radar place recognition [8] already uses a loss with many (i.e. more than 2) negative samples – so, we extend upon this one. For a fair comparison, we compare the uncertainty mechanism of STUN with ours under the same loss regime – see further experimental details in Sec. IV-B.

(1) **Invariant contrastive loss** on the deterministic representation Z_I to disentangle the task-independent noise from radar semantics such that the invariant embedding contains sufficient causal information; and

(2) **Variant contrastive loss** between on the overall representation Z to establish a meaningful metric space. Both

contrastive losses take the following form in Eq. (1).

$$L_{Con} = - \sum_i^m \log P(i | \hat{\mathbf{x}}_i) - \sum_i^m \sum_{j \neq i}^m \log(1 - P(i | \mathbf{x}_j)) \quad (1)$$

where a batch consists of m samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ and synthetically rotated temporal proximal frames augmentations $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_m\}$ using a “spinning” strategy [8], which is simply rotation augmentation, for rotation invariance. We aim to maximise the probability that an augmented sample \hat{x}_i is recognised as the original instance x_i while minimising the probability of the reversed case. The probability of recognition is approximated as Eq. (2).

$$P(i | \mathbf{x}_j) = \frac{\exp\left(\frac{\mathbf{z}_i^T \mathbf{z}_j}{\tau}\right)}{\sum_{k=1}^m \exp\left(\frac{\mathbf{z}_k^T \mathbf{z}_j}{\tau}\right)}, j \neq i \quad (2)$$

where the embeddings \mathbf{z} are either Z_I or Z as in 1) and 2).

(3) **Kullback–Leibler (KL) divergence** between the learned Gaussian distribution and a standard isotropic multivariate Gaussian, which is our prior assumption on the data noise. This ensures an identical distribution across all the sample noises and provides a static reference for the absolute value of variant output.

$$D_{KL} = \sum_{z \in Z_V} \mathcal{N}(\mu, \Sigma) \log \left(\frac{\mathcal{N}(\mu, \Sigma)}{\mathcal{N}(\mathbf{0}, \mathbf{I})} \right) \quad (3)$$

(4) **Reconstruction loss** between the extract feature map M and decoder output M_R , which forces the overall representation Z to contain sufficient information in the original radar scans for reconstruction. However, instead of pixel-level radar scan reconstruction, we only reconstruct a lower dimensional feature map to reduce the computational cost in the decoding process.

$$L_{Rec} = \|M_R - M\|_2 \quad (4)$$

While the vanilla VAE structure driven by only KL divergence and reconstruction loss also provides latent variance, it is considered unreliable for uncertainty estimation [31] due to its well-known problem of posterior collapse [32] and vanishing variance [33]. We confirm this experimentally in Sec. V.

Such ineffectiveness is mainly due to the imbalance of two losses during the training process: when the KL divergence dominates, the latent space posterior is forced to equal the prior whereas when the reconstruction loss dominates, the latent variance is pushed to zero. In our method, however, we achieve more stable training by introducing the variant contrastive losses as an extra regulariser, where the variance is driven to keep a robust boundary among clustering centres in metric space. Hence, we obtain a more reliable latent space variance reflecting the underlying aleatoric uncertainty of radar perception. We show that this VAE-like uncertainty mechanism *alongside* a metric learning objective is more well-calibrated to place recognition than STUN [17], vanilla VAEs [20], MC Dropout [16], etc, which we prove experimentally in Sec. V.

B. Continual Map Maintenance

Continual Map maintenance is an important function of the online system since we aim to fully exploit the scanning data obtained during a lifetime of autonomous vehicle operation and improve the map in a recursive manner. The process of merging a new radar scan into the parent map, consisting of scans from previous traversals, is illustrated in Algorithm 1. Each radar scan is represented by both a hidden representation μ and an uncertainty measure U . In the merging process, we search for matched positive scans for each new scan with topological distance under a threshold D_t . If the new scan has a lower uncertainty then it will be integrated into the parent map and replace the matched scan, otherwise, it will be discarded.

Algorithm 1 Map Maintenance

Input: Parent Map: $M = \{(\mu_i, U_i) | i \in [0, n]\}$, Map GPS $\{D_i\} | i \in [0, N]$, New Scan: (μ_j, U_j) , Scan GPS D_j .
Output: Updated Parent Map: $\underline{M} = \{(\mu_i, U_i) | i \in [0, n]\}$
Find Matched Scan Index:
1: $\{\Delta D_i | i \in [0, n]\} \leftarrow \text{Sort}(\|D_i - D_j\|)$
2: **if** $\Delta D_i < D_t$ **then**
3: $\{U_1, U_2, \dots, U_m\} \leftarrow \text{GetUncertainty}(M, \Delta D_i)$
4: **end if**
Uncertainty Comparison and Replacement:
5: **for** $i = 0$ to m **do**
6: **if** $(U_j > U_i)$ **then**
7: $(\mu_i, U_i) \leftarrow (\mu_j, U_j)$
8: **end if**
9: **end for**
10: **return** \underline{M}

As shown in Fig. 3, The replacement of higher uncertainty scans will remove noisy scans, which potentially lead to incorrect prediction, such as saturation and range uncertainty.

By iteratively performing the maintenance process for all obtained scans, we can gradually enhance the quality of the integrated parent map. Thus, the maintenance algorithm can serve as an effective online deployment strategy as it continually exploits multiple experiences of the same route traversal to boost recognition performance and, in the

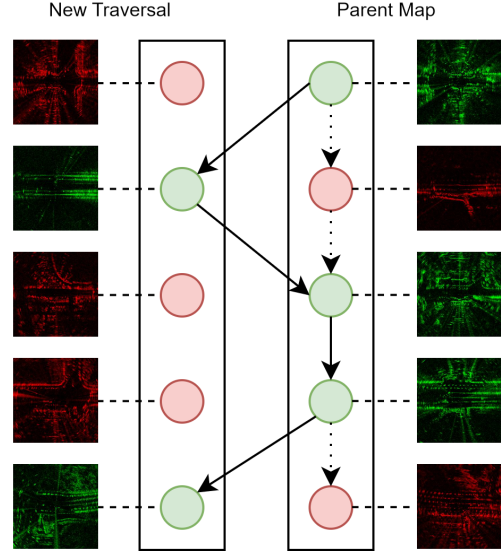


Fig. 3: Illustration of Map Maintenance. The red and green nodes each represent radar scans with higher and lower uncertainty respectively. We always maintain a parent map as the localisation reference consisting of only scans with the lowest uncertainty at each location.

meanwhile, preserve a *constant parent map size* resulting in budgeted computation and storage cost.

In order to facilitate this map maintenance, we currently use GPS, but it is important to note that this step is purely offline (place recognition is performed entirely online, with radar alone) and that this offline step may be readily be replaced by whatever underlying map representation the place recognition queries are issued against, i.e. one built by radar-only mapping and motion estimation [3], [12], [28], [34]. It is also important to note that our results in Sec. V show that even without this map maintenance, our uncertainty-aware representation of Sec. III-A outperforms previous methods, but that this method is capable of significantly boosting localisation performance by being careful about the learned uncertainty of map contents over many repeat traversals.

C. Introspective Query Rejection

During the inference, we use the Euclidean distance between the query and scanned map in a metric space as a measure of similarity. The top k pairs with the highest similarity serve as the prediction result. To further enhance recognition performance and achieve introspective prediction, we also develop an uncertainty rejection mechanism.

Static Thresholding Since the aleatoric uncertainty is measured against a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, the estimated variances in all dimensions are close to 1. Thus, we can use two hyperparameters Δ and N to fully define the scale and resolution of uncertainty rejection respectively. The resultant thresholds T are defined as follows:

$$T = \{(1 - \Delta) + n \times \frac{2\Delta}{N} | n \in [0, N]\} \quad (5)$$

Given a scan with m dimension latent variance Σ , we average across all the dimensions and obtain a scalar uncertainty measure $U = \frac{1}{m} \sum_i \Sigma_i$.

Prediction Rejection At inference time, we perform introspective query rejection, where the query scans with a higher variance than a defined threshold will be rejected from recognition. Existing methods, such as STUN and MC Dropout, dynamically divide the uncertainty range of batch samples into threshold levels. However, this imposes the requirement of multiple samples during inference and can result in unstable rejection performance, particularly when a small number of samples are available. In contrast, our static thresholding strategy offers sample-independent threshold levels and provides consistent single-scan uncertainty estimation and rejection. This feature is crucial for the real-time deployment of the place recognition system, as the radar scan is acquired on a frame-by-frame basis during the driving process.

IV. EXPERIMENTAL SETUP

A. Datasets

We evaluate the performance using two standard benchmarks. (1) *Oxford Radar RobotCar* [35]: It contains radar scans from 30 traversals of the same route around Oxford city centre equivalent to approximately 300 km driving. We follow the train-test split of Săftescu *et al* [7]. An unseen route representing forward and backward traversal of both urban and vegetation environments is used for testing, and all the rest of the routes are used for training. We use five test sequences¹ where two arbitrary sequences can form a mapping and localisation pair. We report evaluation metric averages over all possible combinations². (2) *MulRan* [6]: It contains radar scans from the traversal of different routes at four locations in Daejeon equivalent to around 120 km driving. We train on all sequences from three locations, DCC, KAIST, and Sejong, then test on the same route in Riverside of two sequences, which alternate to serve for mapping and localisation in two testing trails.

Both datasets encompass data collected from a CTS350-X Navtech FMCW scanning radar. This particular radar system is devoid of Doppler information and is mounted on a platform in such a manner that its axis of rotation is perpendicular to the driving surface. The operating frequency of this radar system is situated within the range of 76 GHz to 77 GHz, which enables it to generate up to 3768 range readings with a resolution of 4.38 cm. The total range of the radar system is 165 m, and each range reading corresponds to one of the 400 azimuth readings, which have a resolution of 0.9° degrees. The radar system is characterized by a scan rotation rate of 4 Hz.

B. Baselines

Benchmarking the recognition performance is done by comparison to several existing methods, including the vanilla

VAE [20], the state-of-the-art radar place recognition method by Gadd *et al* [8] (referred as BCRadar), and the non-learning-based method RingKey [5] (one part of ScanContext, without rotation refinement). Additionally, the performance is compared against MC Dropout [16], and STUN [17], which serve as uncertainty-aware place recognition baselines. We note that our testing of BCRadar uses a different list of dataset/sequence combinations than in [8].

Ablation study In order to evaluate the effectiveness of our proposed Introspective Query (Q) and Map Maintenance (M) modules, we conduct an ablation study by comparing different variants of our method, denoted as OURS(O/M/Q/QM), which are as follows

- 1) O: No map maintenance, no introspective querying
- 2) M: Map maintenance only
- 3) Q: Introspective querying only
- 4) QM: Both map maintenance and introspective querying

Specifically, we compare the performance of O against M for recognition performance and Q against QM for uncertainty estimation performance. In each group of the variant comparison, we perform independent mapping inference over multiple localisation sequences and report the average metric, or merge multiple localisation sequences and report a single aggregate metric.

Common settings To ensure a fair comparison, we adopt a common batch contrastive loss [36] as for all metric learning-based methods, thus enabling a training regime with a consistent loss function across the benchmarking.

C. Implementation details

Scan settings For all methods, we transform polar radar scans of with $A = 400$ azimuths and $B = 3768$ bins of size 4.38 cm to Cartesian scans with side-length $W = 256$ and bin size 0.5 m.

Training hyperparameters We use a VGG-16 [37] as the backbone feature extractor with a linear layer to project the extracted feature to a lower embedding dimension of $d = 128$. We train all baselines for 10 epochs in *Oxford Radar RobotCar* and 15 epochs in *MulRan* respectively³, with a learning rate of $1e-5$, batch size of 8. In adapting the batch criterion of [36], we use a negative margin in the embedding space of 0.1 and temperature τ of 1.

D. Evaluation Metric

To assess the place recognition performance, we use Recall@N (R@N) metric, which is the accuracy of the localisation by determining whether at least one candidate amongst N candidates is close to the ground truth, as indicated by GPS/INS. This is particularly important to safety assurance in the autonomous driving application because it reflects the system's calibration to false negatives rate. We also use Average Precision (AP) to measure the mean precision across all recall levels. Finally, we employ F-scores with $\beta = 2/1/0.5$ to assign the increasing level of importance to recall over precision as an aggregate metric assessing overall recognition performance. For all metrics,

¹2019-01-18-14-46-59-radar-oxford-10k,
2019-01-18-15-20-12-radar-oxford-10k,
2019-01-18-14-14-42-radar-oxford-10k,
2019-01-14-14-15-12-radar-oxford-10k,
2019-01-16-14-15-33-radar-oxford-10k

²there are in total 20 pairs, 10 unique pairs of sequences, but the assignment of reference/query being important for introspective query rejection

³STUN is finetuned for another 5 epochs for knowledge distillation

TABLE I: Oxford Radar RobotCar Recognition Performance. The best and second-best results are bold and underlined, respectively.

	R@N=1/5/10			F-1/0.5/2			AP
BCRadar	85.39	88.59	90.52	0.89	0.83	0.75	0.95
VAE	48.35	62.88	69.81	0.56	0.48	0.56	0.70
RingKey	83.68	86.49	88.32	0.85	0.81	0.71	0.94
STUN	89.49	92.23	93.71	0.92	0.87	0.76	0.98
MC Dropout	88.51	91.57	93.24	0.91	0.85	0.76	0.97
OURS(O)	<u>90.46</u>	<u>93.76</u>	<u>95.22</u>	<u>0.93</u>	<u>0.87</u>	<u>0.77</u>	<u>0.98</u>
OURS(M)	93.67	96.24	97.39	0.95	0.91	0.78	0.99

we impose a boundary of 25m such that all radar scans within it are considered positive pairs to a mapping scan. Similarly, a 50m lower boundary defines the corresponding true negative pairs.

Moreover, to assess the uncertainty estimation performance. We use Recall@RR , where we perform introspective query rejection and evaluate $\text{Recall@N}=1$ over different uncertainty threshold levels – rejecting all queries of scans with uncertainty greater than a threshold. We thus reject between 0% to 100% of queries. A reliable uncertainty measure should reflect the confidence level of the model’s predictions, and the rejection of low-confidence scans should result in improved recognition performance (i.e. we should not reject certain, good results). Our goal is to achieve higher improvement in recognition performance with a lower rate of rejections.

V. RESULTS

A. Place Recognition Performance

As demonstrated in Tab. I of the *Oxford Radar RobotCar* experiments, our methods, utilizing only the metric-learning module, achieve the highest performance across all the metrics. Specifically, in terms of Recall@1 , our approach OURS(O) showcases the efficacy of the variance-disentangled representation learned through the variational contrastive learning framework, resulting in superior 90.46% recognition performance. This is further supported with the *MulRan* experiment results demonstrated in Tab. II, where our method outperforms all the other methods on Recall@1 and aggregate F-scores and AP. Although VAE outperforms our method in Recall@5/10 in the *MulRan* experiment, the best F-1/0.5/2 and AP of our method in both settings indicate an overall more accurate and robust recognition performance with high precision and recall. Furthermore, VAE uncertainty is shown below (Sec. V-B) to be poor for introspective query rejection, while our learned uncertainty is useful in both settings.

Moreover, by further utilizing the Continual Map Maintenance in *Oxford Radar RobotCar*, we are able to further improve the Recall@1 to 93.67%, outperforming the current state-of-the-art method, STUN, by a margin of 4.18%. This further proves the effectiveness of learnt variance as a valid uncertainty measure and the strategy of uncertainty-based map integration in improving place recognition performance.

TABLE II: MulRan Recognition Performance. Results reported following the same practice as Tab. I.

	R@N=1/5/10			F-1/0.5/2			AP
BCRadar	32.23	37.81	40.70	0.36	0.31	0.36	0.43
VAE	36.47	46.54	52.58	<u>0.41</u>	<u>0.35</u>	<u>0.42</u>	<u>0.50</u>
RingKey	27.53	29.34	30.17	0.30	0.27	0.30	0.36
MC Dropout	32.90	37.50	40.50	0.35	0.32	0.34	0.43
STUN	26.76	31.15	34.04	0.31	0.26	0.31	0.36
OURS(O)	37.50	<u>43.54</u>	<u>47.57</u>	0.44	0.37	0.43	0.51

B. Uncertainty Estimation Performance

The change in recognition performance, specifically the Recall@1 , with an increasing percentage of rejected uncertain queries, is illustrated in Fig. 4 for the *Oxford Radar RobotCar* experiment and in Fig. 5 for the *MulRan* experiment. Notably, our method is the only one to exhibit a consistent improvement in recognition performance as the rejection rate of uncertain queries increases in both experimental settings – *Oxford Radar RobotCar* and *MulRan* – whereas STUN increases for *Oxford Radar RobotCar* but not *MulRan* and MC Dropout increases for *MulRan* but not *Oxford Radar RobotCar*. For *Oxford Radar RobotCar* in Fig. 4, it is interesting to note that Ours(QM) and Ours(Q) but outperform the initial Recall@1 at a lower rejection rate than STUN does – about 40% for Ours(QM) and Ours(Q) but a much more comprehensive rejection of 70% than for STUN. This result serves as evidence of the robustness and effectiveness of our uncertainty measure.

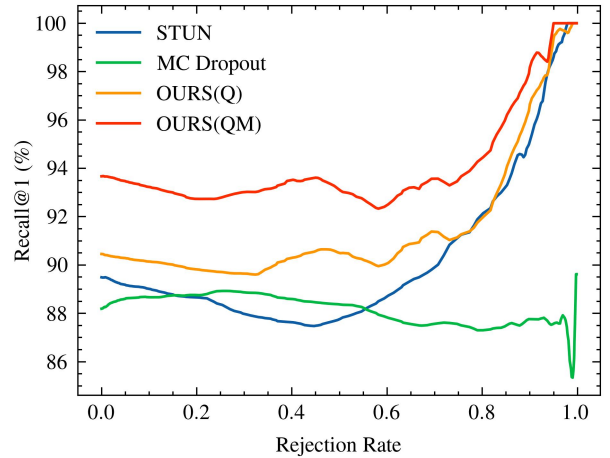


Fig. 4: *Oxford Radar RobotCar* introspective query rejection performance. Recall@1 increases/decreases as the percentage of uncertain query rejected increases. VAE’s performance is not visualised as it is was much lower in comparison to other methods (specifically (48.42/48.08/18.48)% for $\text{Recall@RR}=0.1/0.5/1.0$).

In the *MulRan* experiments, OURS(Q) is the only method that consistently and smoothly improves the Recall@RR metric as the rejection rate increases. Compared to VAE and STUN, which also estimate aleatoric uncertainty like our method, OURS(Q) achieves an improvement of +(1.32/3.02/8.46)% on $\text{Recall@RR}=0.1/0.2/0.5$, while VAE and STUN show a decline of -(3.79/5.24/8.80)% and -(2.97/4.16/6.30)%, respectively. Although in the *Ox-*

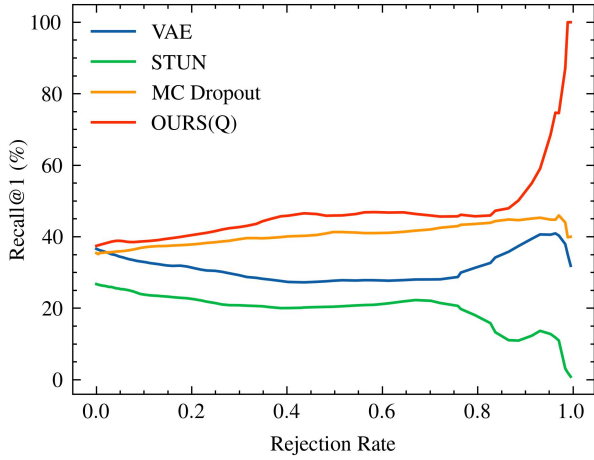


Fig. 5: Mulran introspective query rejection performance reported the same format as in Fig. 4.

ford Radar RobotCar experiments, all methods initially experience fluctuations in performance as the rejection rate increases, ultimately, OURS(Q/QM) outperforms both VAE and STUN with a higher Recall@RR metric at all rejection rates. The superior performance of OURS over VAE suggests the effectiveness of using a regularizing invariant discriminative loss to address the issues of posterior collapse and vanishing variance in the vanilla VAE structure. Furthermore, the better performance of OURS compared to STUN indicates that our variational contrastive learning framework can be more effective than knowledge distillation in learning a reliable variance for uncertainty estimation.

On the other hand, compared with MC Dropout, which estimates the epistemic uncertainty resultant from biased data and model misspecification [30], although it has a higher increase in Recall@1 at the early stage of rejection in *Oxford Radar RobotCar* experiment, its performance is generally lower than ours and fails to achieve greater improvement as the rejection rate increases further. Also, in *MulRan* experiments, it fails to produce reliable uncertainty estimates, as indicated by the general decreasing trend in Recall@RR. These results suggest that aleatoric uncertainty plays a more significant role in causing mispredictions in place recognition than epistemic uncertainty.

Finally, comparing OURS(Q) and OURS(QM) in *Oxford Radar RobotCar* experiment, we observe a similar change in Recall@RR pattern while a considerable gap exists between them. This suggests that the Introspective Query and Map Maintenance mechanisms independently contribute to the place recognition system and that each mechanism exploits the uncertainty measure in an indispensable way.

VI. DISCUSSION

A. Qualitative Analysis and Visualisation

To qualitatively assess the source of uncertainty in radar perception, we provide a visual comparison of the high/low uncertainty samples from both datasets estimated with our method. As shown in Fig. 6, the high-uncertainty radar scans usually demonstrate heavy motion blur and sparse undetected regions, while the low-uncertainty scans usually

contain distinct features with a stronger intensity across the histogram.

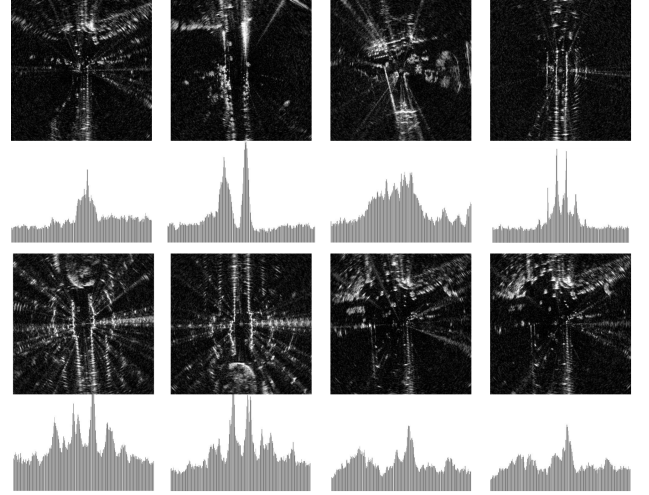


Fig. 6: Visualisation of Radar Scans with Different Levels of Uncertainty. The four examples on the left are from the *Oxford Radar RobotCar Dataset* while the four examples on the right are from *MulRan*. We show samples with the Top-10 highest (top) / lowest (bottom) uncertainty. The radar scan is displayed in a Cartesian coordinate with enhanced contrast. The histogram below each image shows the Ring Key descriptor [5] extracted feature of the intensity across all the azimuths.

This further supports our hypothesis on the source of uncertainty in radar perception and serves as qualitative evidence of the effectiveness of our uncertainty measure that captures this data noise.

B. Dataset Difficulty: Uncertain Environments

In our benchmark experiments, we observed a considerable discrepancy in the recognition performance between the two datasets. The difficulty of *MulRan* as compared to *Oxford Radar RobotCar* for both place recognition and metric localisation is already well-documented in [8], [38]. We posit that the scale of available training data can be a plausible cause. The training set in *Oxford Radar RobotCar* comprises over 300 km driving experiences, while the *MulRan* dataset includes only around 120 km. However, also consider the drop in performance for the *non-learned* Ring Key descriptor method. This suggests potential inherent indistinguishable features in radar scene perception. For instance, we found that environments with sparse open areas usually lead to identical scan and suboptimal recognition performance. We include results on this dataset to exhibit what happens to our system and the various baselines in these cases of high uncertainty. As shown in Fig. 5, our learned uncertainty is the most useful in these difficult scenarios.

VII. CONCLUSION

We have presented a novel application of uncertainty estimation to learned radar place recognition. We bolster the performance of an invariant instance feature learning

approach with a noise-aware latent representation with variational learning. We apply this to (1) continual map maintenance for online system deployment in long-term autonomy, where reference sequences are constructed from low-uncertainty scans only; and (2) introspective query rejection with static thresholding for sequential single-scan real-time inference, where our uncertainty gives the system the ability to understand when it is likely to be incorrect. In tests over two sizeable public radar datasets, we outperform the previous state-of-the-art in radar place recognition, and other uncertainty estimation techniques.

ACKNOWLEDGEMENTS

This work was supported by EPSRC Programme Grant “From Sensing to Collaboration” (EP/V000748/1). We would also like to thank our partners at Navtech Radar.

REFERENCES

- [1] I. Gulrajani and D. Lopez-Paz, “In search of lost domain generalization,” *arXiv preprint arXiv:2007.01434*, 2020.
- [2] P. Checchin, F. G  rossier, C. Blanc, R. Chapuis, and L. Trassoudaine, “Radar scan matching slam using the fourier-mellin transform,” pp. 151–161, 2009.
- [3] S. H. Cen and P. Newman, “Radar-only ego-motion estimation in difficult settings via graph matching,” in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 298–304, IEEE, 2019.
- [4] Z. Hong, Y. Petillot, and S. Wang, “Radarslam: Radar based large-scale slam in all weathers,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5164–5170, IEEE, 2020.
- [5] G. Kim and A. Kim, “Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4802–4809, 2018.
- [6] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, “Mulran: Multimodal range dataset for urban place recognition,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [7]  . S  ftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, “Kidnapped radar: Topological radar localisation using rotationally-invariant metric learning,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4358–4364, IEEE, 2020.
- [8] M. Gadd, D. De Martini, and P. Newman, “Contrastive Learning for Unsupervised Radar Place Recognition,” in *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR)*, 2021.
- [9] J. Komorowski, M. Wysoczanska, and T. Trzcinski, “Large-scale topological radar localization using learned descriptors,” in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part II* 28, pp. 451–462, Springer, 2021.
- [10] M. Gadd, D. De Martini, and P. Newman, “Look around you: Sequence-based radar place recognition with learned rotational invariance,” in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 270–276, IEEE, 2020.
- [11] W. Wang, P. P. de Gusmo, B. Yang, A. Markham, and N. Trigoni, “RadarLoc: Learning to Relocalize in FMCW Radar,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.
- [12] D. De Martini, M. Gadd, and P. Newman, “kRadar++: Coarse-to-fine FMCW Scanning Radar Localisation,” *Sensors*, vol. 20, no. 21, p. 6002, 2020.
- [13] K. Cait, B. Wang, and C. X. Lu, “Autoplace: Robust place recognition with single-chip automotive radar,” in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2222–2228, IEEE, 2022.
- [14] M. Gadd, D. De Martini, and P. Newman, “Unsupervised Place Recognition with Deep Embedding Learning over Radar Videos,” *arXiv preprint arXiv:2106.06703*, 2021.
- [15] H. Yin, X. Xu, Y. Wang, and R. Xiong, “Radar-to-lidar: Heterogeneous place recognition via joint learning,” *Frontiers in Robotics and AI*, vol. 8, p. 661199, 2021.
- [16] Y. Gal and Z. Ghahramani, “Dropout as a bayesian approximation: Representing model uncertainty in deep learning,” in *international conference on machine learning*, pp. 1050–1059, PMLR, 2016.
- [17] K. Cai, C. X. Lu, and X. Huang, “STUN: Self-teaching uncertainty estimation for place recognition,” pp. 6614–6621, 2022.
- [18] F. Warburg, M. J  rgensen, J. Civera, and S. Hauberg, “Bayesian triplet loss: Uncertainty quantification in image retrieval,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12158–12168, 2021.
- [19] Y. Shi and A. K. Jain, “Probabilistic face embeddings,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6902–6911, 2019.
- [20] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [21] J. Oh and G. Eoh, “Variational bayesian approach to condition-invariant feature extraction for visual place recognition,” *Applied Sciences*, vol. 11, 2021.
- [22] H. Wu and M. Flierl, “Learning product codebooks using vector-quantized autoencoders for image retrieval,” in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1–5, 2019.
- [23] L. A. P. Rey, D. Jarnikov, and M. Holenderski, “Content-based image retrieval from weakly-supervised disentangled representations,” in *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.
- [24] Y. S. Park, J. Kim, and A. Kim, “Radar localization and mapping for indoor disaster environments via multi-modal registration to prior lidar map,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1307–1314, 2019.
- [25] X. Lin, Y. Duan, Q. Dong, J. Lu, and J. Zhou, “Deep variational metric learning,” in *European Conference on Computer Vision*, 2018.
- [26] K. Burnett, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, “Radar odometry combining probabilistic estimation and unsupervised feature learning,” *arXiv preprint arXiv:2105.14152*, 2021.
- [27] A. Taha, Y.-T. Chen, T. Misu, A. Shrivastava, and L. Davis, “Unsupervised data uncertainty learning in visual retrieval systems,” *arXiv preprint arXiv:1902.02586*, 2019.
- [28] D. Adolfsson, M. Karlsson, V. Kubelka, M. Magnusson, and H. Andreasson, “Tbv radar slam – trust but verify loop candidates,” *arXiv preprint arXiv:2301.04397*, 2023.
- [29] R. Aldera, D. De Martini, M. Gadd, and P. Newman, “What could go wrong? introspective radar odometry in challenging environments,” in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2835–2842, IEEE, 2019.
- [30] Y. Gal, *Uncertainty in deep learning*. PhD thesis, University of Cambridge.
- [31] J. B  tepage, L. Maystre, and M. Lalmas, “Gaussian process encoders: Vae with reliable latent-space uncertainty,” in *Machine Learning and Knowledge Discovery in Databases. Research Track*, pp. 84–99, Springer International Publishing, 2021.
- [32] J. He, D. Spokoyny, G. Neubig, and T. Berg-Kirkpatrick, “Lagging inference networks and posterior collapse in variational autoencoders,” *arXiv preprint arXiv:1901.05534*, 2019.
- [33] H. Akrami, A. Joshi, S. Aydore, and R. Leahy, “Quantile regression for uncertainty estimation in vae with applications to brain lesion detection,” in *Information Processing in Medical Imaging* (A. Feragen, S. Sommer, J. Schnabel, and M. Nielsen, eds.), pp. 689–700, 2021.
- [34] R. Weston, M. Gadd, D. De Martini, P. Newman, and I. Posner, “Fast-MbyM: Leveraging Translational Invariance of the Fourier Transform for Efficient and Accurate Radar Odometry,” in *International Conference on Robotics and Automation (ICRA)*, pp. 2186–2192, May 2022.
- [35] 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)*, (Paris), 2020.
- [36] M. Ye, X. Zhang, P. C. Yuen, and S.-F. Chang, “Unsupervised embedding learning via invariant and spreading instance feature,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6210–6219, 2019.
- [37] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [38] 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*, vol. 40, no. 12-14, pp. 1488–1509, 2021.