Abstract—We learn, in an unsupervised way, an embedding from sequences of radar images that is suitable for solving the place recognition problem with complex radar data. Our method is based on invariant instance feature learning but is tailored for the task of re-localisation by exploiting for data augmentation the temporal successivity of data as collected by a mobile platform moving through the scene smoothly. We experiment across two prominent urban radar datasets totalling over 400 km of driving and show that we achieve a new radar place recognition state-of-the-art. Specifically, the proposed system proves correct for 98.38% of the queries that it is presented with over a challenging re-localisation sequence, using only the single nearest neighbour in the learned metric space. We also find that our learned model shows better understanding of out-of-lane loop closures at arbitrary orientation than non-learned radar scan descriptors.

Index Terms—Radar, Place Recognition, Deep Learning, Unsupervised Learning, Autonomous Vehicles

I. INTRODUCTION

For autonomous vehicles to drive safely at speed, in inclement weather, or in wide-open spaces, very robust sensing is required. Thus, the interest in scanning Frequency-Modulated Continuous-Wave (FMCW) radar for place recognition and localisation is required. However, all of the approaches presented thus far which are learned from data are also supervised. As more datasets featuring this modality become available, unsupervised techniques will allow us to learn from copious unlabelled measurements. Nevertheless, there are interesting sensor characteristics that must be overcome in order to fully exploit this technology for mapping and localisation tasks. These include artefacts and speckle noise, sensor configuration, perceptual ambiguity and aliasing, and road layout and driving scenarios.

Fig. 1 illustrates a sequence of temporally successive radar scans used to train our system.

As in previous work [2], [3], we do not pursue fine-grained localisation with pose refinement but are interested in learning robust representations of radar imagery, which are quick to compute and compare and which will in any case bolster the performance of downstream pose estimation systems. In our work, we exploit the fact that the sensory records collected by our vehicle are strongly sequential within an otherwise unsupervised learning framework. Specifically, we leverage temporally successive complex radar data as “augmentations” of instances in mini-batches in order to learn robust and invariant radar scan embeddings. We exploit commonality in the sensing horizons in every direction around the robot of scans captured shortly after one another in radar videos. This starts from significant overlap (yellow on red) and degrades to minimal overlap (blue on red). The primary effect of this strategy is that our learned network is presented with more real data than typical contrastive learning approaches, leveraging the “noises” in radar measurements (c.f. artefacts and speckle noise, above). Indeed, maximally overlapping scans are a rich source of training data for neural networks, as opposed to totally synthesised contrastive content. Minimally overlapping scans are useful in learning a metric space – we should ensure that their embeddings are well separated (c.f. perceptual ambiguity and aliasing, above). This work is an extension of our initial investigation in [3], where we additionally boost performance by measuring place-to-place distance as the distributional similarity of stochastic embeddings generated by unfreezing dropout layers during inference.

Our principal contributions are as follows:
1) Mini-batch strategies which exploit the temporal successivity of radar scans gathered from the scanning sensor,
2) Augmentation operations for learning invariance to rotation,
3) Boosting performance by using active dropout layers
during inference and a distributional similarity metric
which accounts for uncertainty due to the stochasticity of
the radar measurement process as well as learned weights.

Incidentally, we perform the first tests of a place recognition
system across multiple radar datasets (c.f. sensor configuration,
above), and comment on the suitability of extant metrics and
the presentation of a more nuanced interpretation decomposed
into common driving scenarios (c.f. road layout and driving
scenarios, above).

II. RELATED WORK

Radar place recognition and localisation has received some
interest, albeit not as much as the related task of motion
estimation. Kim et al [4] show that radar can outperform
LiDAR in place recognition using a non-learned rotationally-
invariant ring-key descriptor. Săftescu et al [2] present the
first supervised deep approach, adapting CNNs for equiv-
ariance and invariance to azimuthal perturbations. Barnes
and Posner [5] learn keypoints which are useful for motion
estimation and localisation simultaneously. Gadd et al [6]
leverage rotationally-invariant representations in a sequence-
based localisation system. De Martini et al [7] fine-tune the
results from the embedding space with pointcloud registration
techniques. Wang et al learn metric localisation directly along
with self-attention [8].

Unsupervised visual place recognition and image retrieval
is an emerging area of study. Merrill and Huang [9] use an au-
toencoder tasked with decoding a handcrafted HOG descriptor,
as a geometric prior. Caron et al [10] iteratively assign pseudo-
labels to images by clustering deep features periodically while
training. Zaffar et al [11] propose image areas of focus based
on entropy measures and describe the image by regional
descriptors which are convolutionally matched for viewpoint-
invariance. Bonin-Font and Burguera [12] present an approach
described as the instance and that no other instance is likewise
classified as the instance \( \hat{x}_i \), being recognised as that instance is

\[
P(i|\hat{x}_i) = \frac{\exp(\frac{f^T \hat{x}_i}{\tau})}{\sum_{k=1}^{m} \exp(\frac{f^T \hat{x}_k}{\tau})}
\]

The probability of another instance, \( x_j \), being recognised as the instance \( x_i \), is given by

\[
P(i|x_j) = \frac{\exp(\frac{f^T x_j}{\tau})}{\sum_{k=1}^{m} \exp(\frac{f^T x_k}{\tau})}, \quad j \neq i
\]

We need to enforce that the instance augmentation be classified as the instance and that no other instance is likewise
classified. The joint probability of this is given by

\[
P_1 = P(i|\hat{x}_j) \prod_{j \neq i} (1 - P(i|x_j))
\]

Therefore, over the entire batch, we seek to minimise

\[
J = - \sum_{i} \log P(i|\hat{x}_i) - \sum_{i} \sum_{j \neq i} \log(1 - P(i|x_j))
\]

B. Motion-enabled data augmentation

The baseline approach in Sec. III-A is typically applied
using synthetically generated instances of images. We can do
better than this in the place recognition task, by exploiting the
strong signal from the relationship between scans captured
close in time.

We retrieve an instance randomly by using a shuffled
sampler to draw a sample from the trajectory where each \( x_i \)
is equally likely to be drawn, as is standard in deep learning
procedures (Fig. 2, orange). We then retrieve a scan several
frames ahead of this instance, \( \hat{x}_i \), to within a maximum time
difference of \( d_{\text{min}} = 2 \text{sec} \) (Fig. 2, red).

1) Augmentation operations for learning invariance to rotation,
2) Boosting performance by using active dropout layers
during inference and a distributional similarity metric
which accounts for uncertainty due to the stochasticity of
the radar measurement process as well as learned weights.

Incidentally, we perform the first tests of a place recognition
system across multiple radar datasets (c.f. sensor configuration,
above), and comment on the suitability of extant metrics and
the presentation of a more nuanced interpretation decomposed
into common driving scenarios (c.f. road layout and driving
scenarios, above).

II. RELATED WORK

Radar place recognition and localisation has received some
interest, albeit not as much as the related task of motion
estimation. Kim et al [4] show that radar can outperform
LiDAR in place recognition using a non-learned rotationally-
invariant ring-key descriptor. Săftescu et al [2] present the
first supervised deep approach, adapting CNNs for equiv-
ariance and invariance to azimuthal perturbations. Barnes
and Posner [5] learn keypoints which are useful for motion
estimation and localisation simultaneously. Gadd et al [6]
leverage rotationally-invariant representations in a sequence-
based localisation system. De Martini et al [7] fine-tune the
results from the embedding space with pointcloud registration
techniques. Wang et al learn metric localisation directly along
with self-attention [8].

Unsupervised visual place recognition and image retrieval
is an emerging area of study. Merrill and Huang [9] use an au-
toencoder tasked with decoding a handcrafted HOG descriptor,
as a geometric prior. Caron et al [10] iteratively assign pseudo-
labels to images by clustering deep features periodically while
training. Zaffar et al [11] propose image areas of focus based
on entropy measures and describe the image by regional
descriptors which are convolutionally matched for viewpoint-
invariance. Bonin-Font and Burguera [12] present an approach
described as the instance and that no other instance is likewise
classified as the instance \( \hat{x}_i \), being recognised as that instance is

\[
P(i|\hat{x}_i) = \frac{\exp(\frac{f^T \hat{x}_i}{\tau})}{\sum_{k=1}^{m} \exp(\frac{f^T \hat{x}_k}{\tau})}
\]

The probability of another instance, \( x_j \), being recognised as the instance \( x_i \), is given by

\[
P(i|x_j) = \frac{\exp(\frac{f^T x_j}{\tau})}{\sum_{k=1}^{m} \exp(\frac{f^T x_k}{\tau})}, \quad j \neq i
\]

We need to enforce that the instance augmentation be classified as the instance and that no other instance is likewise
classified. The joint probability of this is given by

\[
P_1 = P(i|\hat{x}_j) \prod_{j \neq i} (1 - P(i|x_j))
\]

Therefore, over the entire batch, we seek to minimise

\[
J = - \sum_{i} \log P(i|\hat{x}_i) - \sum_{i} \sum_{j \neq i} \log(1 - P(i|x_j))
\]

B. Motion-enabled data augmentation

The baseline approach in Sec. III-A is typically applied
using synthetically generated instances of images. We can do
better than this in the place recognition task, by exploiting the
strong signal from the relationship between scans captured
close in time.

We retrieve an instance randomly by using a shuffled
sampler to draw a sample from the trajectory where each \( x_i \)
is equally likely to be drawn, as is standard in deep learning
procedures (Fig. 2, orange). We then retrieve a scan several
frames ahead of this instance, \( \hat{x}_i \), to within a maximum time
difference of \( d_{\text{min}} = 2 \text{sec} \) (Fig. 2, red).
Fig. 2: Overview of our network architecture and the training procedure more generally — described in Sec. III.

$$\hat{x}_i = x_{i+k}, \ k \in [0, d_{\text{min}}]$$ (5)

This takes the place of the “augmentation” which in other applications would be identical to the instance $x_i$. This ensures that real data is presented to the network, rather than synthetic augmentations which would anyway be prohibitively difficult to simulate for radar sensor artefacts. This also presents the network with some translational invariance, along the route driven. This “augmentation” is further augmented by rotating it (Fig. 2, red, (5)).

$$x_i = R(x_{i+k}, r), \ k \in [0, d_{\text{min}}], \ r \in [0, A]$$ (6)

where $A$ is the number of azimuths in the polar grid and $R$ denotes a cylindrical shift of the azimuthal returns. This helps us achieve rotational invariance. True scans at similar attitudes are of course available in the dataset but would only be usable in a supervised learning framework, and so we settle for synthetic data augmentation at this point. Finally, other batch instances $x_j, j \neq i$ (Fig. 2, blue) and their augmentations $\hat{x}_j$ (Fig. 2, green, (5)) are retrieved at least as distant (in time) as $d_{\text{min}} = 2\text{sec}$ and at most as distant (in time) as $d_{\text{max}} = 6\text{sec}$ from the original instance.

$$\hat{x}_j = x_{j+k}, \ k \in [d_{\text{min}}, d_{\text{max}}]$$

$$\hat{x}_j = R(x_{j+k}, r), \ k \in [d_{\text{min}}, d_{\text{max}}], \ r \in [0, A]$$ (7)

This mitigates the (already quite small but non-zero) chance that batches may simultaneously hold instances which are in fact nearby, although randomly sampled. We also expect that because these instances are somewhere nearby each other with some data overlap that the constraint of Eq. (4) enforces a better metric space where they are well separated and are nevertheless distinct places. Sampling based on attenuated global positioning is also possible but we present this strategy in terms of sampling based on elapsed scans in the radar video as it is true to the unsupervised nature of the base system that we adapt to this task. It should be noted that the base system, Ye et al, is prone to instances in the batch being in fact of the same class, analogously to the vehicle idling in the same position in our system. The second sample at up to $d_{\text{max}}$ reduces the likelihood of this as compared to using $d_{\text{min}}$ alone. Sec. V proves the effectiveness of each sampling and data augmentation step, which are designed to counteract stationary data.

C. Measuring place similarity

In performing place recognition by neighbourhood searches in the learned metric space — i.e. in constructing difference matrices such as those shown in Fig. 3, for a fixed embedding function we typically use Euclidean or cosine distance metrics. In this work, in order to account for uncertainty in the model weights, stochasticity of the radar measurement process, and perceptual ambiguity of self-similar scenes we produce a family of embeddings for each map and query frame during inference. This family of embeddings is available by activating dropout layers during inference. We then use as a distance metric the Kullback-Liebler divergence between two normal distributions which are located at the mean and scaled by the variance of each embedding family. While we only explore Kullback-Liebler, other divergence measures — such as Mahalanobis — are also appropriate here.

D. Training details and parameters

Polar radar scans of with $A = 400$ azimuths and $B = 3768$ bins of size $4.38\text{cm}$ are converted to Cartesian scans with side-length $W = 256$ and bin size $0.5\text{m}$. Feature vectors at the output of our networks are length $d = 4096$. We train each variant for 10 epochs, using a learning rate of $3e-4$, batch size of 12. We collect $T = 24$ samples for distributional parameters during forward passes with active dropout. For comparing against the non-learned descriptor of Kim et al [4], we use the reported ring-key parameters which downsample the radar scan to $120 \times 40$ before averaging over azimuths to form a $40 \times 1$ feature vector.

IV. Experimental Setup

A. Datasets

We train and test our method with various urban radar data.

1) Oxford: These data are collected in the Oxford Radar RobotCar Dataset [17]. Train-test splits are as per [2], testing on hidden data featuring backwards traversals, vegetation, and ambiguous urban canyons2. We do, however, train on most data available from [5] (barring anything from the test split) – totalling 30 forays, or about $280\text{km}$ of driving. This is a similar quantity of training data as reported in [5].

2) MulRan: These data are collected in the MulRan dataset [4]. We train using all but one of these sequences. The test sequence4 is chosen as it features revisits in the opposite direction as well as lane-level differences. The data used totals approximately $124\text{km}$ of driving.

---

22019-01-18-14-46-59-radar-oxford-10k for mapping and 2019-01-18-15-20-12-radar-oxford-10k for localisation
3DCC, KAIST, Sejong
4Riverside
Fig. 3: (a) Ground truth matrix, (b) Distance matrix for embeddings learned by our top-performing model, vTR2, (c) match matrix for Recall@P=80%, (d) match matrix for Recall@N=25. In (c) and (d), true positives are shown green while false positives are shown red.

(a) Recall@P metrics for the Oxford dataset.
(b) Recall@N metrics for the Oxford dataset.

Fig. 4: Evaluation of Recall@P and Recall@N performance for batch strategies over the Oxford Radar RobotCar Dataset.

B. Benchmarking

In addition to the distributional similarity (denoted KL), our ablation study examines the performance gain available from the motion-enabled strategies described in Sec. III-B above, namely:

- vR or “spin augmenting”: where augmentations are produced from instances themselves and synthetic rotations,
- vT or “video sampling”: where augmentations are taken from temporally proximal frames in the radar sequence, but not synthetically rotated,
- vTR or “video sampling and spin augmenting”: temporally proximal frames which are also synthetically rotated, and finally
- vTR2: or “video sampling and spin augmenting with batch pairing”, where other instances in the batch are ensured to be true negatives.

Note the baseline of [19], entails augmenting only the instance itself, closest in spirit to vR.

We also compare our method via direct experimentation to a non-learned radar scan descriptor using the method of Kim et al [4] as it is readily implementable and not learned from data. To embed the radar scan, this method resizes it and averages within range-bins and over azimuths to form a feature vector. We refer to these results as SC. Finally, we also discuss performance figures in comparison to results reported in Săftescu et al [2] and Barnes and Posner [5].

For completeness, we note that we use either VGG-19 [20] or ResNet-152 [21] as front-end feature extractors.

C. Performance assessment

For performance assessment, we employ both Recall@P – where predicted positives are within a varying embedding distance threshold of the query – and Recall@N – where a query is considered correctly localised if at least one database candidate is close in space as measured by ground truth (in our case, GPS/INS). Note that the latter is a more lenient measure of performance and that we employ no pose refinement to boost results, as in [4], [7]. We specifically impose – for both Oxford and MulRan datasets – for the first metric (e.g. Figs. 4(a) and 6) boundaries of 25 m and 50 m, inside of which predicted matches are considered true positives, outside of which false positives, as per [2]. The second metric (e.g. Figs. 4(b) and 7 to 9) supports only a single boundary, which we set to 25 m, the more difficult of the two. For precision-recall results, we also present aggregate metrics, including F-scores ($F_0.5$, $F_1$, and $F_2$) and area-under-curve (AUC).

D. Loop closure event types

Fig. 5 shows decomposition of loop closure events by proposed types. We argue for the importance of a qualitative distinction of performance for matches arising from:

1) teach-and-repeat, where the taught path is closely followed (in the case of the Oxford and MulRan datasets where the vehicle is under manual control there may be minor intra-lane variation), and
2) revisiting, where the same place is visited on more than one occasion and at arbitrary attitude (in this example, backwards) during a foray. Specifically, the results presented in Sec. V mask out from the embedding distance matrix the true positives of all other
decomposition categories. We refer to the decompositions as \text{rpt} and \text{rev}. Therefore, to isolate e.g. \text{rev} in Fig. 8, we use Fig. 5(a) (\text{rpt}) to ignore ground truth and predicted matches in Figs. 3(a) and 3(b).

Fig. 5: True positives corresponding to ground truth, Fig. 3(a). We consider distinct types of loop closures events and how they are and should be interpreted and understood in radar place recognition – \textit{teach-and-repeat} is shown in (a), revisiting is shown in (b).

![Fig. 5](image)

(a) True positives corresponding to ground truth, Fig. 3(a).
(b) Revisiting is shown in (a), revisiting is shown in (b).

Fig. 6: Evaluation of aggregates of precision-recall as well as Recall@N performance for batch strategies over \textit{MulRun}.

![Fig. 6](image)

(a) Aggregate precision-recall area-under-curve (AUC) metrics for the \textit{MulRun} dataset.
(b) F-Score metrics for the \textit{MulRun} dataset.

V. RESULTS

Fig. 4(a) shows the Recall@P metrics for the \textit{Oxford} dataset. We achieve from 11.72\% up to as much as 16.34\% recall for precisions of 99\% to 95\%. This is compared to 4.26\% to 7.08\% for the most naive constrastive learning approach, \text{vr}. This corresponds to \text{F}_1, \text{F}_2, and \text{F}_{0.5} scores of 0.60, 0.55, and 0.63, respectively. Here, we have outperformed the supervised embedding learning of Săftescu \textit{et al} [2], where on this same test split the corresponding scores were reported as 0.53, 0.48, and 0.60.

Fig. 4(b) shows the Recall@N metrics for \textit{Oxford}. Here, we notice when only using a single database candidate (Recall@N=1), we correctly localise 98.38\% of the time (vTR2) as compared to 73.13\% for the baseline, \text{vR}. This is also superior to the results reported in Barnes and Posner [5], approximately 97\%.

Note the fidelity of the embedding space for vTR2, Fig. 3(b), as compared to the ground truth matrix, Fig. 3(a). Differently to Figs. 4(a) and 4(b), Fig. 3(c) shows the true and false positives for Recall@P=80\%. Fig. 3(d) shows such for Recall@N=25, a similar number of database candidates used as in the work of Kim \textit{et al} [4] as well as Barnes and Posner [5]. In both, we consider it reasonable to expect a downstream process to disambiguate \textit{some} false matches in order to retain various types of loop closure. We see that rotational invariance is well understood by the network trained as we propose.

![Fig. 7](image)

Fig. 7: Evaluation of Recall@N for the VGG-19 and ResNet-152 architectures over the \textit{Oxford} dataset.

Figs. 6(a) and 6(b) show the corresponding precision-recall aggregates as well as Recall@N for the \textit{MulRun} dataset. Scores on this dataset are typically lower than for \textit{Oxford}, which we attribute to lane-level differences as well as the quantity of training data available. Nevertheless, improvements from \text{vr} to vTR2 are clearly demonstrated on this challenging dataset, validating each step of the batch strategy in our unsupervised approach. Indeed, we see a similar incremental improvement in \text{F}_{0.5}, \text{F}_1, and \text{F}_2 from \text{vr} to \text{vT}, then to vTR and up to vTR2. We also plot AUC metrics with a similar improvement with strategy of 0.07 to 0.24.

Fig. 7 shows the performance over the \textit{Oxford} dataset for the two feature extraction networks. Considering Recall@1, we have only 95.96\% recall for ResNet-152 as compared to 98.38\% for VGG-19. On this basis alone, and considering that the gains in the region of Recall@6 to Recall@11 are much more slight, VGG-19 is the preferred architecture.

For comparison against the non-learned method of Kim \textit{et al} [4], we list the precision-recall metrics. These are such that our method achieves \text{F}_1, \text{F}_2, and \text{F}_{0.5} scores of 0.54, 0.50, and 0.54, respectively. This is compared to corresponding scores of 0.41, 0.37, and 0.45 for SC.

These results are further decomposed as per Sec. IV-D in Fig. 8 and correspond to the \text{rpt} matches. Our method achieves 17.78\% recall with Recall@1 for \text{rev} loop closures as compared to 7.47\% for SC. We consider that SC is translationally sensitive to out-of-lane driving, rather than being rotationally suspect. This shows that our method understands to an extent out-of-lane driving, which we attribute to our learning of rotational invariance and trajectory sampling. As mentioned in Sec. III-B above, out-of-lane robustness could be easily learned in a weakly supervised setting, using
of-the-art, showing the great promise of unsupervised techniques over published methods which are candidates for the current state-of-the-art. In tests over two challenging urban and difficult scenery by measuring distributional similarity between families of radar scan embeddings. In future work, we facilitate out-of-lane loop closures. Finally, we account for the variable view of the radar sensor and performing random rotations to further augment the data. We bolster the applicability of an invariant instance feature learning approach by learning to radar place recognition. We compare to Kim et al [4] as compared to [11].

**Fig. 8:** Evaluation on Oxford of (a) precision-recall as compared to Kim et al [4] and (b) Recall@1 for decompositions of loop closure event types – localising along the taught path (rpt) and out-of-lane (rev).

**Fig. 9:** Evaluation of precision-recall for embedding families with active dropout and distributional similarity.

GPS/INS as a pseudo groundtruth. We expect that methods for tackling this in the unsupervised setting could be a focus of future work in this area.

Fig. 9 shows the boosts to performance available from the distributional similarity proposed in Sec. III-C. Here, we use estimated uncertainty to enable or disable putative localisation results. Using a family of embeddings and distributional similarity boosts our performance to $F_1$, $F_2$, and $F_{0.5}$ scores of 0.65, 0.57, and 0.67, respectively, from 0.61, 0.55, and 0.63. Correspondingly, we achieve 17.70% for Recall@99% as compared to 11.73% and 44.54% for Recall@80% as compared to 35.49%.

## VI. Conclusion

We have presented the first application of unsupervised learning to radar place recognition. We bolster the applicability of an invariant instance feature learning approach by sampling temporally proximal scans to act as augmentations of batch instances for the contrastive objective. These are further augmented by considering the full horizontal field-of-view of the radar sensor and performing random rotations to facilitate out-of-lane loop closures. Finally, we account for the uncertainty in model predictions and perceptually ambiguous and difficult scenery by measuring distributional similarity between families of radar scan embeddings. In tests over two sizeable public radar datasets, we outperform three previously published methods which are candidates for the current state-of-the-art, showing the great promise of unsupervised techniques in this area.

## Acknowledgements

This work was supported by the Assuring Autonomy International Programme, a partnership between Lloyd’s Register Foundation and the University of York as well as EPSRC Programme Grant “From Sensing to Collaboration” (EP/V000748/1). We would also like to thank our partners at Navtech Radar.

## References

[1] Q.-Y. Zhou, J. Park, and V. Koltun, “Open3D: A modern library for 3D data processing,” arXiv:1801.09847, 2018.
[2] Ş. Săftescu, M. Gadd, D. De Martini, D. Barnes, and P. Newman, “Kidnapped Radar: Topological Radar Localisation using Rotationally-Invariant Metric Learning,” in arXiv preprint arXiv:2001.09438, 2020.
[3] M. Gadd, D. De Martini, and P. Newman, “Unsupervised place recognition with deep embedding learning over radar videos,” arXiv preprint arXiv:2106.06703, 2021.
[4] 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.
[5] 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.
[6] M. Gadd, D. De Martini, and P. Newman, “Look Around You: Sequence-based Radar Place Recognition with Learned Rotational Invariance,” in IEEE/ION Position, Location and Navigation Symposium (PLANS), 2020.
[7] D. De Martini, M. Gadd, and P. Newman, “kRadar++: Coarse-to-fine FMCW Scanning Radar Localisation,” Sensors, vol. 20, no. 21, p. 6002, 2020.
[8] 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.
[9] N. Merrill and G. Huang, “Lightweight unsupervised deep loop closure,” arXiv preprint arXiv:1805.07703, 2018.
[10] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 132–149.
[11] M. Zaffa, S. Ehsan, M. Milford, and K. McDonald-Maier, “Cohog: A light-weight, compute-efficient, and training-free visual place recognition technique for changing environments,” IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 1835–1842, 2020.
[12] F. Bonin-Font and A. Burguera Burguera, “Nhethaloc: A learned global image descriptor for loop closing in underwater visual slam,” Expert Systems, vol. 38, no. 2, p. e12635, 2021.
[13] M. Leyva-Vallina, N. Strisicungli, and N. Petkov, “Place recognition in gardens by learning visual representations: data set and benchmark analysis,” in International Conference on Computer Analysis of Images and Patterns, Springer, 2019, pp. 324–335.
[14] ——, “Generalized contrastive optimization of siamese networks for place recognition,” arXiv preprint arXiv:2103.06638, 2021.
[15] F. Radenović, G. Tolias, and O. Chum, “Fine-tuning cnn image retrieval with no human annotation,” IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 7, pp. 1655–1668, 2018.
[16] P. Sermanet, C. Lynch, Y. Chebotar, J. Hsu, E. Jang, S. Schaal, S. Levine, and G. Brain, “Time-contrastive networks: Self-supervised learning from video,” in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1134–1141.
[17] D. Barnes, M. Gadd, P. Marcutt, P. Newman, and I. Posner, “The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset,” in IEEE International Conference on Robotics and Automation (ICRA), 2020.
[18] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Alhribian, S. Wang, and A. Wallace, “RADIATE: A Radar Dataset for Automotive Perception,” arXiv preprint arXiv:2010.09076, 2020.
[19] M. Ye, X. Zhang, P. C. Yuen, and S.-F. Chang, “Unsupervised Embedding Learning via Invariant and Sparse Embedding Feature,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
[20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.