

Automotive mmWave Spinning Radar Place Recognition with Spatially Gated Feature-Correlation Representation

Saimunur Rahman¹, Sagun Singh Shrestha¹, Abdelwahed Khamis^{1,2}, Peyman Moghadam^{1,2}

¹CSIRO Robotics, CSIRO, Australia `firstname.lastname@csiro.au`

²Queensland University of Technology, Australia `firstname.lastname@qut.edu.au`

Abstract

Automotive spinning FMCW radar provides dense, 360° sensing and remains reliable under poor illumination and adverse weather, making it well-suited to autonomous navigation. Place recognition uses these observations to identify previously visited locations for re-localization and long-term navigation. However, heading changes appear as circular shifts in the polar radar representation, and conventional global aggregation can lose relationships among radar responses that are important for distinguishing similar places. We propose SGCA-Net, a spinning radar place recognition framework that combines rotation-robust feature extraction with Spatially Gated Correlation Aggregation (SGCA). SGCA learns spatial weights to reduce the influence of unstable and ambiguous radar regions, while aggregating pairwise correlations among local responses to preserve informative feature relationships. Experiments on the MulRan dataset show that SGCA-Net consistently outperforms SOTA methods across urban, campus, and open-road environments, while remaining robust to substantial heading variation. Evaluation on the HeRCULES dataset further demonstrates that SGCA-Net generalizes to unseen environments and radar sensors without fine-tuning.

1. Introduction

Place recognition (PR) determines whether a current observation corresponds to a previously visited place and is central to long-term autonomous robot navigation, loop closure, and re-localization [1, 2]. PR is commonly formulated as a retrieval problem, where observations are encoded as compact global descriptors and a query is matched to a database using nearest-neighbour search, with candidate matches subsequently refined through a re-ranking method [3]. The effectiveness of this retrieval formulation therefore depends on descriptors that remain consistent across revisits while providing sufficient discrimination be-

tween geographically distinct locations [4]. The ability to construct such robust descriptors depends strongly on the sensing modality. Cameras [5, 6, 7] and LiDAR [2, 8] have been widely adopted for PR, while automotive mmWave frequency-modulated continuous-wave (FMCW) radar offers complementary robustness under poor illumination and adverse weather conditions, including rain, fog, and dust [9, 10, 11, 12]. Among automotive radar configurations, spinning FMCW radar is particularly suitable for global descriptor-based PR [13, 14, 15, 16] because it provides dense, 360° azimuth-range measurements, in contrast to the sparser and more limited field-of-view observations typically produced by 4D automotive radar sensors [17, 18, 19].

A spinning FMCW radar scan is formed by mechanically rotating the radar antenna over azimuth while recording reflected signal power over range, producing a 360° azimuth-range polar representation [9, 12]. While this representation provides dense coverage of the surrounding environment, it introduces several challenges for place recognition. Changes in vehicle heading appear as circular shifts along the azimuth dimension, while the measured radar responses vary with reflector geometry, material properties, incidence angle, multipath, sidelobes, and range [11, 10, 9]. In addition, dynamic objects, occlusions, and transient clutter can alter the spatial distribution of returns between revisits. Consequently, observations of geographically distinct places may contain similar local radar responses, while observations of the same place may exhibit substantial variation across traversals [13, 14, 15, 16]. An effective radar descriptor must therefore preserve repeatable spatial structure across revisits while reducing the influence of unstable and ambiguous responses.

Existing radar PR methods have primarily addressed heading variation through radar-specific convolutional processing, circular operations, transform-domain representations, or rotation-aware descriptor design [13, 14, 15, 16]. However, the subsequent aggregation of local radar features into a compact global representation remains comparatively less explored. First-order aggregation methods, including

average, max, and generalized mean (GeM) pooling, primarily summarize feature presence or response magnitude. Learned aggregation approaches, such as NetVLAD [4] and its radar variants [14, 16], provide more expressive representations by aggregating local features, but largely characterize the distribution of individual responses. For spinning radar, this can discard informative co-occurrence structure across the scan, where place identity may depend on joint response patterns rather than isolated feature strength. A global descriptor should therefore preserve not only salient individual responses, but also relationships among responses that remain consistent across revisits.

Motivated by these observations, we introduce SGCA-Net, a learned spinning FMCW radar place recognition framework that combines a rotation-robust radar backbone with a Spatially Gated Correlation Aggregation (SGCA) descriptor. The backbone operates directly on the azimuth-range polar representation and extracts local features that are robust to heading-induced circular shifts. The SGCA module subsequently projects these features into a compact descriptor space, learns spatial keep/dustbin weights to reduce the contribution of unstable or ambiguous locations, and aggregates pairwise correlations among the reweighted local responses. Unlike conventional first-order aggregation, which mainly summarizes feature presence or response strength, SGCA therefore produces a spatially gated feature-correlation representation to preserve repeatable radar structure. Power normalization further stabilizes the representation. The contributions of this work are summarized as follows:

- We propose SGCA-Net, a spinning radar place recognition framework that combines rotation-robust feature extraction with spatially gated correlation aggregation.
- We introduce a spatially weighted feature correlation representation that captures pairwise relationships among radar responses while reducing the influence of unstable and ambiguous regions.
- We demonstrate consistent improvements across diverse MulRan environments and show that the representation learned by SGCA-Net transfers effectively to unseen sensor and environments without fine-tuning.

2. Related Work

Radar Place Recognition. Radar Place Recognition (RPR) has evolved from hand-crafted representations towards learned global features. ScanContext [20] represents a scan using a polar grid and handles heading variation through circular alignment, while Ring Key provides a compact rotation-invariant summary derived from the same representation. Although these methods are efficient and interpretable, their hand-crafted formulations offer limited

adaptability to the substantial appearance variation present in automotive radar observations. RadarLoc [15] showed that a compact radar-specific convolutional network operating directly on polar radar images can learn effective global descriptors for large-scale place recognition. Its cylindrical processing and GeM pooling provide robustness to heading variation while maintaining a lightweight architecture. More recent methods have explored alternative feature aggregation, matching, and sensing configurations. OpenRadVLAD [14] aggregates local radar features using a VLAD-based representation, while RaPlace [13] employs Radon-transform-based processing to improve robustness to changes in radar appearance. SHeRLoc [16] extends radar place recognition to heterogeneous radar sensors and includes evaluation on spinning radar observations. Despite these advances, most learned spinning radar approaches still compress their feature maps using conventional global aggregation. Consequently, descriptor formation is dominated by the strength or frequency of local responses, while interactions among feature responses receive limited attention. This is restrictive for spinning radar, where a place may be identified more reliably by recurring combinations of returns than by isolated strong measurements.

Global Feature Aggregation. Global aggregation converts local feature maps into compact descriptors for efficient place retrieval. First-order methods, including average, maximum, and generalized mean (GeM) pooling [21], summarize individual feature responses but discard relationships between feature channels. GeM is commonly used for place recognition and is adopted by RadarLoc for spinning radar PR. Learned aggregation methods such as NetVLAD [4] and OpenRadVLAD [14] provide more expressive descriptors by modelling the distribution of local features, but do not explicitly capture relationships among feature responses. This is limiting for automotive radar, where feature maps may contain empty regions, multipath, transient objects, and repeated structures that can introduce unstable or ambiguous responses.

Pairwise Feature Correlation Aggregation. Pairwise correlation representations model dependencies between feature channels through bilinear or covariance statistics. Bilinear pooling [22] and covariance-based representations [23] have shown that such interactions can improve discrimination beyond first-order aggregation. However, directly applying correlation aggregation to radar is problematic because unreliable spatial locations can influence the resulting correlation statistics. SGCA-Net addresses this by applying learned spatial gating before correlation aggregation, reducing the contribution of unstable regions while retaining informative relationships among feature responses. Matrix square-root normalization [23] is then used to stabilize the resulting representation.

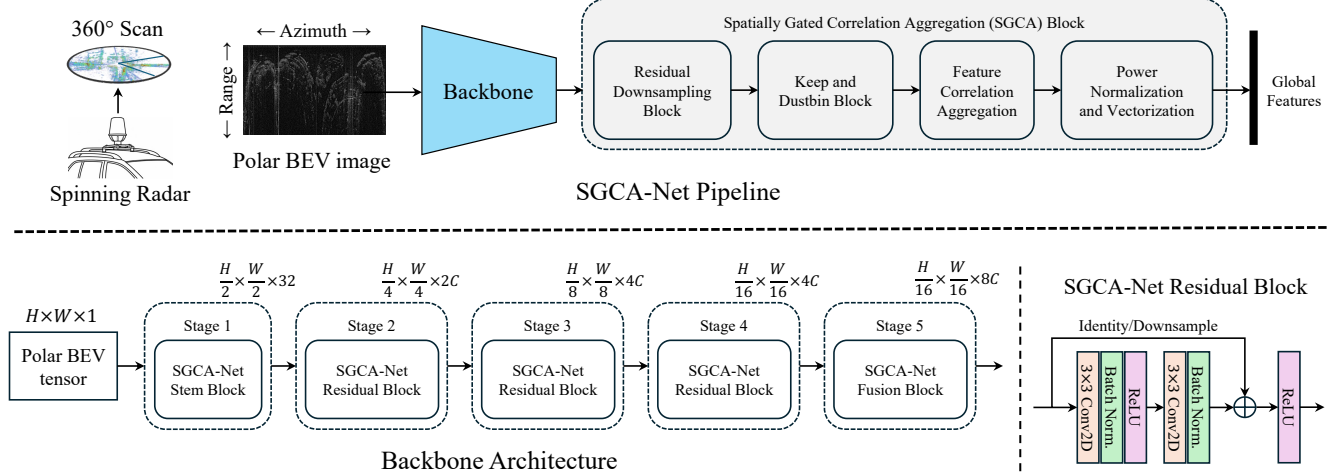


Figure 1. Overview of SGCA-Net. The input is a polar BEV radar image, with azimuth and range as image axes. The image is processed by a stem block and residual backbone stages that reduce spatial resolution while increasing feature channels. The resulting feature map is passed to the SGCA block, which performs residual downsampling, keep-and-dustbin selection, feature-correlation aggregation, power normalization, and vectorization to produce a global descriptor.

3. Methodology

SGCA-Net is a deep metric learning framework for spinning FMCW radar place recognition. Given a polar radar image $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$, the network extracts a dense feature map using a rotation-robust radar backbone and converts it into a compact global descriptor through the proposed SGCA pooling module, as illustrated in Fig. 1. The complete descriptor extraction pipeline is expressed as $\mathbf{d} = f(\mathbf{X}) = \Psi(\Phi(\mathbf{X}))$, where $\Phi(\cdot)$ denotes the backbone, $\Psi(\cdot)$ denotes the SGCA module, and $\mathbf{d} \in \mathbb{R}^D$ is the final descriptor. The network is trained end-to-end with a rotation-robust metric learning objective, described in Sec. 3.3

3.1. Radar Polar Image Backbone

The backbone converts the range-azimuth radar scan $\mathbf{X} \in \mathbb{R}^{H \times W \times 1}$ into a dense feature map for SGCA (shown in Fig. 1). It consists of a convolutional stem, four residual stages, and a top-down fusion block. We denote a cylindrically padded $k \times k$ 2D convolution with stride s by $\mathcal{C}_{k,s}(\cdot)$, batch normalization by $\mathcal{B}(\cdot)$, ReLU by $\sigma(\cdot)$, and transposed convolution by $\mathcal{T}_{k,s}(\cdot)$.

Since spinning radar measurements are periodic in azimuth, conventional zero padding introduces an artificial discontinuity at the 0° and 360° boundaries. We therefore apply cylindrical padding along the azimuth dimension while retaining zero padding along range:

$$\tilde{\mathbf{X}}(r, c) = \begin{cases} \mathbf{X}(r, c + W), & c < 0, \\ \mathbf{X}(r, c), & 0 \leq c < W, \\ \mathbf{X}(r, c - W), & c \geq W. \end{cases} \quad (1)$$

This allows convolutional kernels near the azimuth boundary to operate on physically continuous neighbourhoods,

improving robustness to heading-induced circular shifts. The stem applies a 5×5 stride-2 convolution, batch normalization, and ReLU, giving $\mathbf{F}_0 = \sigma(\mathcal{B}(\mathcal{C}_{5,2}(\mathbf{X})))$, where $\mathbf{F}_0 \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 32}$. The stem output is processed by four residual stages. Each residual block contains two 3×3 convolutions and a shortcut connection, as shown in bottom right of Fig. 1. For an input feature map $\mathbf{U}$, the residual mapping is

$$\mathcal{R}(\mathbf{U}) = \mathcal{B}(\mathcal{C}_{3,1}(\sigma(\mathcal{B}(\mathcal{C}_{3,s}(\mathbf{U}))))), \quad (2)$$

and the block output is $\mathbf{Y} = \sigma(\mathcal{R}(\mathbf{U}) + \mathcal{S}(\mathbf{U}))$, where

$$\mathcal{S}(\mathbf{U}) = \begin{cases} \mathbf{U} & \text{if dimensions are unchanged,} \\ \mathcal{B}(\mathcal{C}_{1,2}(\mathbf{U})), & \text{otherwise.} \end{cases} \quad (3)$$

The projection shortcut is used when spatial downsampling or channel expansion is required. The four stages produce

$$\begin{aligned} \mathbf{F}_1 &\in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 32}, & \mathbf{F}_2 &\in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 64}, \\ \mathbf{F}_3 &\in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 128}, & \mathbf{F}_4 &\in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times 256}. \end{aligned} \quad (4)$$

The fusion block combines the deepest feature map with the intermediate feature map $\mathbf{F}_3$. The deepest feature is projected and upsampled as

$$\hat{\mathbf{P}}_3 = \mathcal{T}_{2,2}(\mathcal{C}_{1,1}(\mathbf{F}_4)), \quad \hat{\mathbf{P}}_3 \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 256}. \quad (5)$$

The lateral feature is projected as $\mathbf{L}_3 = \mathcal{C}_{1,1}(\mathbf{F}_3)$, with $\mathbf{L}_3 \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 256}$. The fused backbone output is

$$\mathbf{F} = \hat{\mathbf{P}}_3 + \mathbf{L}_3, \quad \mathbf{F} \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 256}. \quad (6)$$

This fused tensor is passed to the SGCA module to form the final global descriptor. For notational simplicity, we define the spatial dimensions of the fused backbone feature map as $h = H/16$ and $w = W/16$, such that $\mathbf{F} \in \mathbb{R}^{h \times w \times 256}$.

3.2. Spatially Gated Correlation Aggregation

The fused backbone output $\mathbf{F} \in \mathbb{R}^{h \times w \times 256}$ is first projected into a compact feature space for descriptor construction. Let $\mathcal{P}_s(\cdot)$, $\mathcal{P}_1(\cdot)$, and $\mathcal{P}_2(\cdot)$ denote 1×1 projection convolutions, and let $\rho(\cdot)$ denote the GELU activation. The projected feature map $\mathbf{Z} \in \mathbb{R}^{h \times w \times 64}$ is computed as

$$\mathbf{Z} = \mathcal{B}(\mathcal{P}_s(\mathbf{F})) + \mathcal{B}(\mathcal{P}_2(\rho(\mathcal{B}(\mathcal{P}_1(\mathbf{F}))))). \quad (7)$$

This residual bottleneck reduces the channel dimension from 256 to 64 while preserving backbone information through the residual branch, producing a feature basis tailored for descriptor formation.

A learned keep/dustbin gate then estimates the spatial importance of each projected local feature. A 1×1 convolution predicts two logits at every spatial location,

$$\mathbf{Q} = \mathcal{G}(\mathbf{Z}), \quad \mathbf{Q} \in \mathbb{R}^{h \times w \times 2}, \quad (8)$$

where $\mathcal{G}(\cdot)$ denotes the gate projection. Applying a softmax over the two channels yields the keep probability α_j for each flattened spatial location. These probabilities are normalized across the $K = hw$ spatial locations to obtain the spatial weights:

$$a_j = \frac{\alpha_j}{\sum_{\ell=1}^K \alpha_\ell}, \quad a_j \geq 0, \quad \sum_{j=1}^K a_j = 1. \quad (9)$$

The resulting weights increase the contribution of informative radar regions while reducing the influence of unstable or ambiguous responses. The ungated variant is recovered by setting $a_j = 1/K$. The projected feature map is then flattened into local descriptors $\{\mathbf{z}_j\}_{j=1}^K$, where $\mathbf{z}_j \in \mathbb{R}^{64}$. The weighted feature mean is $\boldsymbol{\mu} = \sum_{j=1}^K a_j \mathbf{z}_j$, and the spatially weighted feature-correlation representation is computed as

$$\boldsymbol{\Sigma} = \sum_{j=1}^K a_j (\mathbf{z}_j - \boldsymbol{\mu})(\mathbf{z}_j - \boldsymbol{\mu})^\top. \quad (10)$$

Although Eq. (10) is mathematically a weighted covariance matrix, its entries capture pairwise relationships between projected feature channels and therefore preserve co-occurrence structure that is discarded by first-order pooling. The spatial weighting further limits the contribution of unstable radar regions to the resulting representation. The weighted covariance matrix also has a useful structural property for subsequent matrix normalization.

Proposition 1. *If $a_j \geq 0$ for all j and $\sum_{j=1}^K a_j = 1$, then the matrix $\boldsymbol{\Sigma}$ defined in Eq. (10) is symmetric positive semidefinite.*

Proof. Symmetry follows directly from Eq. (10). For any vector $\mathbf{v} \in \mathbb{R}^{64}$,

$$\mathbf{v}^\top \boldsymbol{\Sigma} \mathbf{v} = \sum_{j=1}^K a_j [\mathbf{v}^\top (\mathbf{z}_j - \boldsymbol{\mu})]^2 \geq 0,$$

since $a_j \geq 0$. Hence, $\boldsymbol{\Sigma}$ is positive semidefinite. $\square$

This property enables stable matrix-function normalization. To improve numerical conditioning, trace normalization is first applied as $t = \max(\text{tr}(\boldsymbol{\Sigma}), \epsilon)$ and $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}/t$. Matrix square-root normalization is then computed using the Newton–Schulz iteration [23], and the normalized matrix is rescaled as $\mathbf{M} = \bar{\boldsymbol{\Sigma}}^{1/2} \sqrt{t}$. The final global descriptor is obtained by vectorizing the upper-triangular entries,

$$\mathbf{d} = \text{vecu}(\mathbf{M}). \quad (11)$$

Since $\mathbf{M} \in \mathbb{R}^{64 \times 64}$ is symmetric, vectorizing its upper-triangular entries produces a $D = 2080$ -dimensional global descriptor.

3.3. Rotation-Robust Triplet Loss

To improve robustness to heading variation, each training scan is circularly shifted along the azimuth axis before descriptor extraction. Let $\mathcal{Q}_\delta(\cdot)$ denote an azimuthal roll by δ , where δ is sampled uniformly over rotations up to 180° . For training sample i , the descriptor is therefore $\mathbf{d}_i = f(\mathcal{Q}_{\delta_i}(\mathbf{X}_i))$. For each anchor i in a mini-batch, batch-hard mining selects the hardest positive and hardest negative:

$$p_i^* = \arg \max_{p \in \mathcal{P}(i)} \|\mathbf{d}_i - \mathbf{d}_p\|_2, \quad n_i^* = \arg \min_{n \in \mathcal{N}(i)} \|\mathbf{d}_i - \mathbf{d}_n\|_2. \quad (12)$$

The training objective is

$$\mathcal{L}_{\text{triplet}} = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} [\|\mathbf{d}_i - \mathbf{d}_{p_i^*}\|_2 - \|\mathbf{d}_i - \mathbf{d}_{n_i^*}\|_2 + m]_+, \quad (13)$$

where $\mathcal{A}$ denotes the set of anchors in the mini-batch, $\mathcal{P}(i)$ and $\mathcal{N}(i)$ are the positive and negative sets for anchor i , $[x]_+ = \max(0, x)$, and $m = 0.2$ is the triplet margin.

4. Experimental Setup

Dataset and Evaluation Environments. We evaluate SGCA-Net on the MulRan dataset [12], which provides 360° radar measurements collected using a Navtech CIR204-H FMCW radar with a maximum range of 200 m, an angular resolution of 0.9° , and a radial resolution of 0.06 m. Each scan is represented as a polar image with

Table 1. Recall@1% Radar Place Recognition on the MulRan [12] evaluation sequences.

Method	Sejong		KAIST		Riverside		Mean	
	5 m	10 m	5 m	10 m	5 m	10 m	5 m	10 m
Ring Key [20]	50.3	59.4	80.5	83.8	49.7	59.5	60.2	67.6
ScanContext [20]	86.8	87.9	93.5	94.6	67.1	77.2	82.5	86.6
VGG-16+NetVLAD [15]	78.9	93.8	88.9	93.7	61.3	83.4	76.4	90.3
RadarLoc [15]	92.9	98.8	95.9	98.8	74.4	92.3	87.7	96.6
SGCA-Net (Ours)	97.4	99.9	97.5	99.2	78.9	95.2	91.3	98.1

400 azimuth bins and 3,360 range bins and is downsampled to 384×128 pixels following [15]. Evaluation is conducted across the Sejong, KAIST, and Riverside environments, which span urban, campus, and open-road conditions with varying scene structure, traffic, and landmark density.

Training and Evaluation Protocol. We follow the training and evaluation splits introduced in [15]. Training uses predefined regions from the Sejong 01 and 02 traversals, yielding 13,611 radar scans after pre-processing. Evaluation is conducted on geographically disjoint regions from Sejong, KAIST, and Riverside, using traversal 02 as the reference database and traversal 01 as the query sequence. This results in 1,366/1,337 database/query scans for Sejong, 3,209/3,420 for KAIST, and 2,117/2,043 for Riverside.

Evaluation Metric. Performance is measured using Recall@1 (R@1). For each query, database scans are ranked by Euclidean distance between global descriptors, and retrieval is considered correct when the top-ranked scan lies within a specified distance of the ground-truth position. We report R@1 at 5 m and 10 m for comparison with prior work, and additionally at 3 m to assess finer place discrimination. Results are reported for each environment and as their unweighted mean.

Implementation Details. SGCA-Net is implemented in PyTorch and trained using AdamW with an initial learning rate of 1×10^{-4} and a batch size of 256. A step-based learning-rate schedule and random erasing are used during training. Unless otherwise stated, all compared variants use the same training data, optimization settings, and evaluation protocol.

5. Results

5.1. Comparison with Existing Methods

We compare SGCA-Net with representative hand-crafted and learned radar place recognition methods evaluated under the same MulRan protocol. Ring Key [20] is a compact, rotation-invariant summary derived from ScanContext, whereas ScanContext [20] retains the full hand-crafted polar representation and accounts for heading changes through circular alignment.

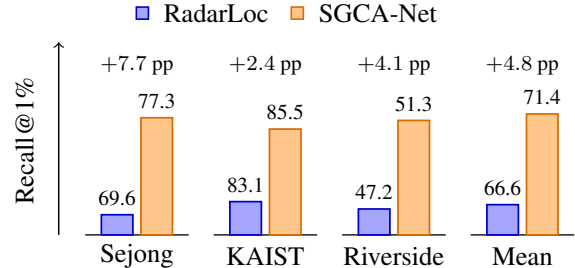


Figure 2. R@1 at the stricter 3 m localization threshold. Values above the bars report R@1, while annotations show the absolute improvement of SGCA-Net over RadarLoc in percentage points.

VGG-16+NetVLAD [15] combines a comparatively large general-purpose convolutional backbone with learned VLAD aggregation. RadarLoc [15] instead uses a compact radar-specific convolutional backbone followed by GeM pooling. Although more recent radar place recognition methods have been proposed, many use different sensor configurations, datasets, or evaluation protocols; we therefore use the methods reported under this common benchmark for direct comparison. Tab. 1 reports Recall@1 at the standard 5 m and 10 m thresholds.

SGCA-Net achieves the highest R@1 on every sequence and at both localization thresholds. Compared with RadarLoc, the mean R@1 increases from 87.7% to 91.3% at 5 m and from 96.6% to 98.1% at 10 m. SGCA-Net also substantially outperforms VGG-16+NetVLAD despite using a smaller backbone. This suggests that the improvement is not obtained through backbone capacity, but through rotation-robust feature extraction and SGCA representation. The sequence-wise gains remain consistent across different operating conditions. On Sejong, SGCA-Net improves the 5 m Recall@1 from 92.9% to 97.4% despite changes between dense urban and highway-like sections. On KAIST, performance increases from 95.9% to 97.5% in the presence of repeated campus layouts. The improvement is also maintained on Riverside, where sparse persistent landmarks and repetitive geometry make place discrimination more difficult: R@1 increases from 74.4% to 78.9% at 5 m and from 92.3% to 95.2% at 10 m.

We additionally evaluate performance at the stricter 3 m

Table 2. Ablation study of the proposed SGCA block on MulRan. Entries report R@1% at 3 m, 5 m, and 10 m. Mean is computed across the Sejong, KAIST, and Riverside sequences. The full SGCA-Net configuration is highlighted.

Method	Sejong			KAIST			Riverside			Mean		
	3 m	5 m	10 m	3 m	5 m	10 m	3 m	5 m	10 m	3 m	5 m	10 m
SGCA-Net w/ GeM pooling	69.6	93.0	98.8	83.1	95.8	98.8	47.2	74.3	92.3	66.6	87.7	96.6
SGCA-Net w/o gate	73.9	95.5	99.5	84.2	96.5	98.9	50.1	76.1	94.3	69.4	89.4	97.6
SGCA-Net w/ standard projection	75.0	96.3	99.5	84.0	96.6	99.0	49.5	77.5	94.2	69.5	90.1	97.6
SGCA-Net	77.3	97.4	99.9	85.5	97.5	99.2	51.3	78.9	95.2	71.4	91.3	98.1

threshold, as shown in Fig. 2. Since this threshold was not reported in the original RadarLoc study, we reproduce the RadarLoc results using its public implementation. SGCA-Net improves R@1 across all three evaluation environments, with gains of 7.7%, 2.4%, and 4.1% on Sejong, KAIST, and Riverside, respectively. Overall, the mean R@1 increases from 66.6% to 71.4%. These results show that SGCA-Net provides reliable discrimination when geographically close places produce similar radar observations.

5.2. Ablation Analysis

We analyse the contribution of the SGCA-Net components and evaluate robustness to heading variation.

SGCA-Net Component Ablation. We conduct an ablation study to quantify the contribution of each SGCA-Net component, as reported in Tab. 2. The GeM variant replaces the SGCA block with generalized mean pooling [21], a widely used global aggregation method also adopted by RadarLoc. This variant highlights the effect of replacing GeM with SGCA while retaining the same radar-specific backbone and training configuration. The ungated variant retains feature-correlation aggregation but assigns uniform weights to all feature-map locations. The standard-projection variant includes the learned keep/dustbin gate but replaces the residual bottleneck projection with a single projection layer. All variants use the same training data, optimization settings, and evaluation protocol.

Replacing GeM pooling with ungated feature-correlation aggregation increases the mean R@1 from 66.6% to 69.4% at 3 m, from 87.7% to 89.4% at 5 m, and from 96.6% to 97.6% at 10 m. These results demonstrate the benefit of modelling feature correlations even when all feature-map locations are weighted uniformly. Adding the keep/dustbin gate and residual bottleneck projection yields further improvements. The complete SGCA-Net achieves the highest mean R@1 at all thresholds, reaching 71.4%, 91.3%, and 98.1% at 3 m, 5 m, and 10 m, respectively.

Robustness to Heading Variation. We evaluate SGCA-Net under random query rotations of up to 180°, following the rotation-enabled setting used by RadarLoc. As reported

Table 3. Mean R@1% under random query rotations of up to 180°.

Method	3 m	5 m	10 m
RadarLoc	65.9	87.4	96.5
SGCA-Net	70.6	91.1	98.1

Table 4. Transfer evaluation on the HeRCULES [25] spinning-radar splits used in [16]. Both methods are trained only on MulRan and evaluated without fine-tuning. AR@1 denotes the mean R@1 across the four session pairs.

Method	Sports Complex		Library		AR@1
	01→02	01→03	01→02	01→03	
RadarLoc [15]	0.965	0.938	0.989	0.867	0.940
SGCA-Net	0.986	0.973	0.993	0.880	0.958

in Tab. 3, SGCA-Net improves the mean R@1 over RadarLoc from 65.9% to 70.6% at 3 m, from 87.4% to 91.1% at 5 m, and from 96.5% to 98.1% at 10 m. The consistent improvements show that the SGCA components strengthen place discrimination while retaining robustness to substantial heading changes.

5.3. Transfer to Unseen Radar Environments

Generalization to unseen environments remains a challenge for place recognition, with prior work investigating test-time adaptation to compensate for domain shift [24]. Here, we instead examine whether the representation learned by SGCA-Net transfers beyond the MulRan training environments without adaptation or fine-tuning. transfers beyond the MulRan training environments. Following the spinning radar evaluation protocol used in [16], SGCA-Net is trained only on MulRan and evaluated without fine-tuning on the Sports Complex and Library sequences of HeRCULES [25]. RadarLoc [15] is evaluated under the same setting to provide a consistent reference. The session pairs include clear-to-night/cloudy and clear-to-snowy changes. We report R@1 for each pair, while AR@1 denotes the mean R@1 across all four pairs.

As shown in Tab. 4, SGCA-Net outperforms RadarLoc on all four HeRCULES session pairs and increases AR@1

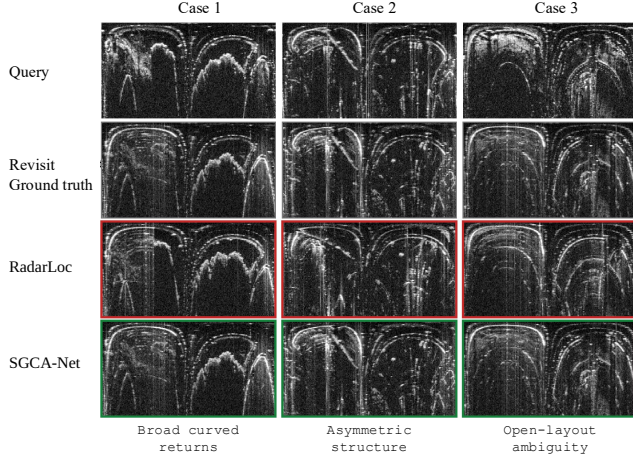


Figure 3. Representative retrieval results under the 3 m criterion. For each case, the rows show the query, the ground-truth revisit, and the top-ranked results from RadarLoc and SGCA-Net. Green and red borders denote correct and incorrect retrievals, respectively. RadarLoc is confused by structurally similar alternatives, whereas SGCA-Net retrieves the correct revisit.

from 0.940 to 0.958. The largest improvement occurs on Sports Complex 01→03, where $R@1$ increases from 0.938 to 0.973. These results suggest that the representation learned by SGCA-Net retains its advantage when transferred to previously unseen environments and conditions without additional fine-tuning.

5.4. Qualitative Analysis

Fig. 3 presents representative Sejong examples for which RadarLoc returns an incorrect top-ranked match, whereas SGCA-Net identifies the correct place under the 3 m criterion. The selected cases include broad curved returns, asymmetric scene structure, and open-road environments with limited distinctive landmarks. In each case, geographically different locations produce similar overall radar-response distributions, making them difficult to distinguish using conventional global aggregation.

SGCA-Net correctly recognizes all three places, suggesting that spatially gated feature-correlation aggregation preserves scene relationships that help distinguish structurally similar observations. These examples are consistent with the improvements observed at the stricter 3 m threshold, where accurate recognition requires finer discrimination between nearby and visually similar places.

6. Conclusion and Future Work

This paper introduced SGCA-Net, a spinning radar place recognition framework that combines rotation-robust feature extraction with spatially gated feature-correlation aggregation. By modelling pairwise relationships among radar features while reducing the influence of unstable and

ambiguous regions, SGCA-Net provides a more discriminative global representation than conventional first-order aggregation. Experiments on MulRan demonstrate consistent improvements over representative RPR methods across diverse environments and under substantial heading variation. Evaluation on HeRCULES further demonstrates that SGCA-Net generalizes to unseen environments and radar sensors without fine-tuning. Future work will investigate more computationally efficient feature-correlation representations and extend the evaluation to additional radar sensors, datasets, and long-term environmental changes. We also plan to explore uncertainty-aware spatial weighting to better account for transient and unreliable radar responses.

7. Acknowledgments

The authors would like to acknowledge support from the CSIRO’s Generative AI for Radar Self-Supervised Learning project in partnership with Boeing.

References

- [1] P. Yin, J. Jiao, S. Zhao, L. Xu, G. Huang, H. Choset, S. Scherer, and J. Han, “General place recognition survey: Towards real-world autonomy,” *IEEE Transactions on Robotics*, 2025. 1
- [2] J. Knights, K. Vidanapathirana, M. Ramezani, S. Sridharan, C. Fookes, and P. Moghadam, “Wild-places: A large-scale dataset for lidar place recognition in unstructured natural environments,” in *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023, pp. 11 322–11 328. 1
- [3] K. Vidanapathirana, P. Moghadam, S. Sridharan, and C. Fookes, “Spectral Geometric Verification: Re-Ranking Point Cloud Retrieval for Metric Localization,” *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2494–2501, 2023. 1
- [4] R. Arandjelović, 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. 1, 2
- [5] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, “Visual place recognition: A survey,” *IEEE transactions on robotics*, vol. 32, no. 1, pp. 1–19, 2015. 1
- [6] S. Hausler, S. Garg, M. Xu, M. Milford, and T. Fischer, “Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 141–14 152. 1
- [7] S. Hausler and P. Moghadam, “Pair-vpr: Place-aware pre-training and contrastive pair classification for visual place recognition with vision transformers,” *IEEE Robotics and Automation Letters*, 2025. 1

- [8] K. Vidanapathirana, M. Ramezani, P. Moghadam, S. Sridharan, and C. Fookes, "Logg3d-net: Locally guided global descriptor learning for 3d place recognition," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2215–2221. 1
- [9] 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 *2020 IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 6433–6438. 1
- [10] K. Burnett, Y. Wu, D. J. Yoon, A. P. Schoellig, and T. D. Barfoot, "Are we ready for radar to replace lidar in all-weather mapping and localization?" *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 328–10 335, 2022. 1
- [11] 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*. IEEE, 2020, pp. 5164–5170. 1
- [12] G. Kim, Y. S. Park, Y. Cho, J. Jeong, and A. Kim, "Mulran: Multimodal range dataset for urban place recognition," in *2020 IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 6246–6253. 1, 4, 5
- [13] H. Jang, M. Jung, and A. Kim, "Raplac: Place recognition for imaging radar using radon transform and mutable threshold," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2023, pp. 11 194–11 201. 1, 2
- [14] M. Gadd and P. Newman, "Open-radvlad: Fast and robust radar place recognition," in *2024 IEEE Radar Conference*. IEEE, 2024, pp. 1–6. 1, 2
- [15] J. Komorowski, M. Wysoczańska, and T. Trzcinski, "Large-scale topological radar localization using learned descriptors," in *International Conference on Neural Information Processing*. Springer, 2021, pp. 451–462. 1, 2, 5, 6
- [16] H. Kim, M. Jung, W. Yang, and A. Kim, "Sherloc: Synchronized heterogeneous radar place recognition for cross-modal localization," *IEEE Robotics and Automation Letters*, vol. 10, no. 12, pp. 13 264–13 271, 2025. 1, 2, 6
- [17] K. Cai, B. Wang, and C. X. Lu, "AutoPlace: Robust place recognition with single-chip automotive radar," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2222–2228. 1
- [18] D. Casado Herraiez, L. Chang, M. Zeller, L. Wiesmann, J. Behley, M. Heidingsfeld, and C. Stachniss, "SPR: Single-scan radar place recognition," *IEEE Robotics and Automation Letters*, vol. 9, no. 10, pp. 9079–9086, 2024. 1
- [19] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, p. 4208, 2022. 1
- [20] 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)*. IEEE, 2018, pp. 4802–4809. 2, 5
- [21] 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, 2019. 2, 6
- [22] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact bilinear pooling," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 317–326. 2
- [23] P. Li, J. Xie, Q. Wang, and Z. Gao, "Towards faster training of global covariance pooling networks by iterative matrix square root normalization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 947–955. 2, 4
- [24] J. Knights, S. Hausler, S. Sridharan, C. Fookes, and P. Moghadam, "GeoAdapt: Self-Supervised Test-Time Adaptation in LiDAR Place Recognition Using Geometric Priors," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 915–922, 2024. 6
- [25] H. Kim, M. Jung, C. Noh, S. Jung, H. Song, W. Yang, H. Jang, and A. Kim, "Hercules: Heterogeneous radar dataset in complex urban environment for multi-session radar slam," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 4649–4656. 6