A Hyperdimensional One Place Signature to Represent Them All: Stackable Descriptors For Visual Place Recognition

Connor Malone Somayeh Hussaini Tobias Fischer Michael Milford
QUT Centre for Robotics, Queensland University of Technology, Australia
{cj.malone, s.hussaini, tobias.fischer, michael.milford}@qut.edu.au

Abstract

Visual Place Recognition (VPR) enables coarse localization by comparing query images to a reference database of geo-tagged images. Recent breakthroughs in deep learning architectures and training regimes have led to methods with improved robustness to factors like environment appearance change, but with the downside that the required training and/or matching compute scales with the number of distinct environmental conditions encountered. Here, we propose Hyperdimensional One Place Signatures (HOPS) to simultaneously improve the performance, compute and scalability of these state-of-the-art approaches by fusing the descriptors from multiple reference sets captured under different conditions. HOPS scales to any number of environmental conditions by leveraging the Hyperdimensional Computing framework. Extensive evaluations demonstrate that our approach is highly generalizable and consistently improves recall performance across all evaluated VPR methods and datasets by large margins. Arbitrarily fusing reference images without compute penalty enables numerous other useful possibilities, three of which we demonstrate here: improved performance with reduced dimensionality descriptors, stacking synthetic images, and coarse localization to an entire traverse or environmental section.

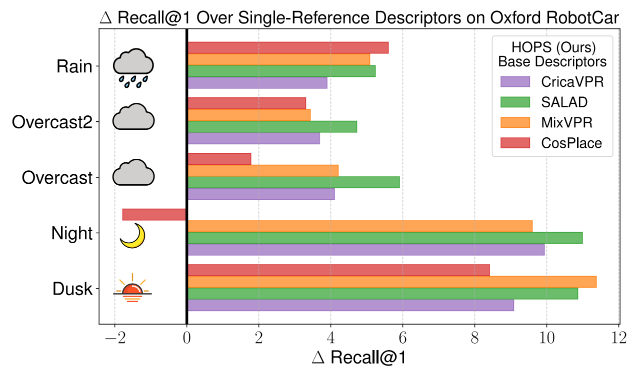

Figure 1. Here, we demonstrate near unanimous improvements to recall@1 by using our HOPS fused descriptors across multiple state-of-the-art base descriptors and query conditions. We show absolute improvement over the best recall achieved by a single reference set, using hyperdimensional computing to fuse descriptors from multiple reference sets with no dimensionality increase.

1. Introduction

Localization is a critical task in robotics [77, 85], autonomous vehicles [16, 38], and augmented reality [58, 65]. Long-term operation requires localization systems that are robust to factors like lighting, weather and dynamic scene changes — all of which significantly impact a place's appearance [72].

Visual Place Recognition (VPR) is the task of identifying previously visited places given a query image and a database of geo-tagged reference images [23, 42, 47, 68, 88]. In applications such as loop closure in Simultaneous Localization and Mapping (SLAM) [14, 21, 75], VPR is often formulated as an image retrieval problem that provides coarse localization estimates, which are then refined in a hierarchical process using feature matching approaches [50, 62, 63].

Most state-of-the-art (SOTA) VPR methods use deep learning models to represent images as feature-based descriptors [3, 10, 31, 35, 43]. While significant progress towards VPR that is robust to lighting, weather, viewpoint and other appearance changes has been made, most approaches adopt the general formulation of using a single reference set (often captured in 'ideal daytime' conditions) to perform place recognition. To further improve appearance invariance, recent deep learning methods have used multi-condition training sets [2, 80], explicit consideration of multiple instances of places captured under varying conditions to improve feature robustness [10, 43], and domain adaptation [12, 28]. Further work has attempted to consolidate separate VPR matches across multiple reference datasets [22, 49], or simply develop ever more robust feature extractors [10, 35, 43, 44, 76].

In this work, we explore an alternative approach for improving general robustness to appearance changes which does not involve computationally- and time-intensive training of a new deep learned feature extractor (see Figure 1). We instead propose Hyperdimensional One Place Signatures (HOPS)[1] to fuse VPR descriptors from the same place captured under varying conditions using the Hyperdimensional Computing (HDC) framework [37, 52] – as opposed to fusing VPR descriptors obtained by complementary techniques [51].

HOPS leverages the capability of current SOTA VPR descriptors to match images in similar domains whilst using the HDC formulation to avoid any additional training, and computational or memory costs. Importantly, HOPS is generalizable and complementary to existing SOTA VPR descriptors. We make the following contributions:

1. The first use of a Hyperdimensional Computing (HDC) framework for fusing multiple reference sets—either from different traverses of the environment, or synthetically generated using image augmentations—in VPR to improve robustness to appearance changes without increasing computation or memory requirements.
2. Extensive experiments showing the framework generalizes across several SOTA VPR methods and multiple datasets with various challenging condition changes, generally outperforming the best single reference set by large margins and achieving better performance than other multi-reference set approaches that require additional computation or memory costs.
3. An alternative operation mode with equivalent recall to baseline at significantly reduced dimensionality: in the case of high-dimensional descriptors such as SALAD [31] (8448D) and CricaVPR [43] (10752D) about a 97% and 95% reduction in feature dimensions, respectively, and for low-dimensional descriptors such as CosPlace [9] and EigenPlaces [10] (both 512D) still achieving about a 50% and 25% reduction, respectively.

The ability to combine reference images together for Visual Place Recognition to improve recognition performance without any scaling of matching compute is a powerful and versatile one. Here, our primary demonstration is in showing the performance and compute benefits in appearance-invariant place matching, through both combining multiple real-world reference traverses and synthetic imagery. Such a capability opens up many other possibilities, some of which are discussed in Section 5.

2. Related Work

2.1. Visual Place Recognition

In Visual Place Recognition (VPR), images are typically converted to high-level feature descriptors robust to appearance and viewpoint changes, allowing a query image to match the correct reference image in the feature space [47, 68]. Early VPR solutions used handcrafted feature descriptors, including global aggregation methods such as Bag of Words (BoW) [19, 70], Fischer Vectors (FV) [55, 56], Vector of Locally Aggregated Descriptors (VLAD) [5, 33], and local descriptors such as SIFT [41] and SURF [8]. With deep learning, these methods evolved into architectures such as NetVLAD [6], NetBoW [59], and NetFV [48]. Since the introduction of CNNs to VPR [6], deep learning techniques have enabled greater robustness against appearance and viewpoint changes, which include works such as DELF [54], DELG [15], DOLG [84] and SuperGlue [64].

Recent approaches address VPR challenges through spatial pooling and aggregation methods such as Generalized Mean Pooling (GeM) [60], and Conv-AP [2], innovative architectures [3], VPR-method-agnostic feature alignment procedures such as MeshVPR [11], effective training regimes [9, 10, 74], and targeted VPR-specific loss functions [39, 61]. MixVPR [3] uses CNN backbones and Feature Mixer layers to establish global relationships within feature maps. EigenPlaces [10] targets viewpoint tolerance by dividing the training dataset to form small classes with images of multiple perspectives. CosPlace [9] reformulates VPR training as a classification task by organizing data into geographically distinct classes. Generalized Contrastive Loss (GCL) [39] improves global descriptor robustness by computing graded similarity for image pairs.

Other SOTA VPR models leverage vision transformers [20, 25] for enhanced feature extraction, including DinoV2 SALAD [31] that treats descriptor aggregation as an optimal transport problem, AnyLoc [35] that also uses DinoV2 without VPR-specific fine-tuning, CricaVPR [43] that introduces cross-image correlation awareness, and BoQ [4] which learns a set of global queries, using cross-attention with local input features to derive global representations.

Other VPR approaches enhance performance using two-stage retrieval techniques, initially identifying top-k candidates using global features, and then re-ranking these candidates using local features [47]. Recent two-stage approaches include Patch-NetVLAD [26] and transformer-based methods such as TransVPR [79], ETR [87], R^2Former [89], SelaVPR [44], and EffoVPR [76]. Relevant to this work, [7] investigates how existing local features and re-ranking methods can be used to improve VPR with challenges such as night time conditions and image occlusions.

2.2. Multi-Reference and Fusion Approaches

Several VPR techniques focus on fusion approaches [27, 32, 69, 82, 86] or consider multiple reference sets [18, 40, 49, 78] by generating enriched reference maps that enable robots to perform long-term autonomous navigation as changes in the environment over time can be incorporated [68]. Feature fusion has been used to fuse input data from a range of

[1] https://github.com/CMalone-Jupiter/HOPS

sensors such as camera, laser and sonar [32], omnidirectional observations with a depth sensor and camera [69], and image-based and event-based camera data [27]. Feature fusion has also been used for re-ranking top-candidate matches obtained through matching global feature descriptors [82, 86].

Training using multi-condition datasets is a common way for VPR methods to achieve more invariant features [2, 80]. While not strictly using multiple reference sets, the SOTA VPR method CricaVPR even specifically incorporates correlations between images of the same place captured under varying conditions [43].

Multiple reference sets have been more explicitly used for improving place recognition performance by incrementally adapting to appearance changes [18] and using probabilistic approaches to predict the best reference set to use for a given query image [40, 49]. [78] used an efficient hashing technique to generate feature descriptors and used a data association graph to store representations from multiple reference sets, and performed place matching using an informed search. While these works [18, 40, 49, 78] have addressed the problem of multiple reference maps, an ongoing concern is the increasing storage and computational requirements with increase in the number of reference sets.

2.3. Hyperdimensional Computing Frameworks

Hyperdimensional Computing (HDC), also known as Vector Symbolic Architectures (VSA), is a brain-inspired computing framework [24, 34]. HDC is used to handle data which is represented in extremely high, or 'hyper', dimensional spaces [24]; expected to have thousands or tens of thousands dimensions. One of the key properties in such hyperdimensional spaces is that there is a high likelihood that two randomly sampled vectors will be near or 'quasi' orthogonal to one another [66]. As a result, several HDC operations can be performed to improve the computational and memory efficiency of dealing with these vectors, including bundling, binding, and permutation [24].

Of interest for this paper is bundling, which fuses sets of input vectors such that the output vector is similar to all input vectors [51]. One method for bundling which has precedence in VPR literature is an element-wise sum of the vectors [51]. The binding operation can be used to assign 'role' or 'class' information to vectors. The output of binding is not similar to the two input vectors but can be reversed to recover the input components; one implementation is through an element-wise multiplication of two vectors [51].

HDC has been used in a range of machine learning applications for learning temporal patterns such as text classification [36], addressing catastrophic forgetting in deep learning-based architectures [17], in robotics for reactive behavior learning, and object and place recognition [52], and out-of-distribution detection [81].

In the context of VPR, [53] presented the Vector Semantic Representations (VSR) image descriptor, which uses HDC to encode the appearance and semantic properties of a place, as well as the topological relationship between semantic classes. [51] presented an HDC-based framework to aggregate image descriptors from multiple different global VPR methods, or for aggregating local features and binding their image position information. [51] exploits the HDC properties of orthogonal vectors to fuse descriptors from different VPR methods – we differ from this by instead exploiting the reinforcement of features by fusing multiple reference descriptors of the same place from the same VPR method.

3. Methodology

3.1. Visual Place Recognition Formulation

We formulate Visual Place Recognition (VPR) as an image retrieval task. Given a query image of the current place and a database of geo-tagged reference images, our goal is to identify the reference image that most closely resembles the query. State-of-the-art VPR methods commonly use deep neural networks to embed images as n-dimensional feature vectors, thereby abstracting complex visual scenes into compact representations.

Formally, let $\mathbf{q} \in \mathbb{R}^n$ represent the feature vector of the query image and $\mathbf{R} = \{\mathbf{r}_i\}$ the set of geo-tagged reference vectors, with $\mathbf{r}_i \in \mathbb{R}^n$ and $|\mathbf{R}| = M$ being the number of reference images. To compute the degree of similarity between the query and each reference, we calculate a distance vector $\mathbf{d} = [d(\mathbf{q}, \mathbf{r}_1), d(\mathbf{q}, \mathbf{r}_2), \ldots, d(\mathbf{q}, \mathbf{r}_M)]$, where $d(\cdot)$ denotes the cosine distance. The estimated location is then derived by selecting the reference with the minimum distance:

$$\mathbf{r}_{\text{match}} = \arg\min_i d(\mathbf{q}, \mathbf{r}_i). \tag{1}$$

This approach critically depends on the robustness of neural network feature extractors, which must maintain discriminative power across various environmental conditions and viewpoints for each unique place. Achieving high consistency across such changes is crucial for robust and long-term VPR. However, instead of relying solely on improved feature extraction, we propose leveraging Hyperdimensional Computing (HDC) to fuse multiple reference sets into Hyperdimensional One Place Signatures (HOPS), enhancing condition invariance without altering existing VPR descriptors.

3.2. Bundling Reference Datasets

Our approach exploits the properties of high-dimensional spaces by aggregating multiple feature vectors to create a fused descriptor which is similar to all inputs. In other words, we put forward the idea that hyperdimensional feature vectors from the same place, captured under different conditions, can be combined to form a unified descriptor that remains robust against minor variations.

Formally, let $\mathbf{r}^k$ be feature vectors representing the same place under different conditions k, with an additional noise vector $\mathbf{z}$ affecting either vector. Due to quasi-orthogonality, the influence of $\mathbf{z}$ on the cosine similarity between $\mathbf{r}^l$ and $\mathbf{r}^m$ $(l \neq m)$ is negligible in high-dimensional space, preserving the similarity despite the noise. $\mathbf{r}_{\text{fused},i}$ combines K reference descriptors from the same place i across diverse conditions, allowing salient features to reinforce while diminishing transient ones:

$$\mathbf{r}_{\text{fused},i} = \sum_{k=1}^{K} \mathbf{r}_i^k. \tag{2}$$

Bundling via summing has the useful property of being able to 'stack' many reference descriptors, which is useful as new descriptors can be easily added to the fusion over time as the places are revisited. It maintains a complexity of $\mathcal{O}(M)$. *Note: HOPS fused descriptors must be L2 normalized to maintain unit norm for cosine distance calculations.*

3.3. Gaussian Random Projection

Beyond the core benefits of our HOPS approach for fusing descriptors without additional compute or memory overhead, it also enables other beneficial applications such as improved performance after dimensionality reduction operations. To demonstrate this, we use Gaussian Random Projection as a representative method in an additional experiment (Section 4.5) to project feature vectors into a lower-dimensional space. Using a random projection matrix, the Johnson-Lindenstrauss Lemma asserts that the distance between a set of points in high-dimensional space can be approximately preserved when embedding in a lower-dimensional space [1]. In this work, we use Gaussian Random Projections to evaluate the capacity for HOPS to reduce the descriptor dimensionality required to maintain performance. This is **not** done for core experimental results (Tables 1-4 and Figures 2-3).

Given a high-dimensional feature vector $\mathbf{r}_{\text{fused},i} \in \mathbb{R}^n$, the Gaussian Random Projection $\mathbf{G} \in \mathbb{R}^{o \times n}$ projects $\mathbf{r}_{\text{fused},i}$ to a lower-dimensional space $\mathbb{R}^o$ where $o \ll n$. The projection is performed using matrix multiplication:

$$\hat{\mathbf{r}}_{\text{fused},i} = \mathbf{G}\mathbf{r}_{\text{fused},i}, \tag{3}$$

where elements in $\mathbf{G}$ are sampled from a Gaussian distribution $\mathcal{N}(0, \frac{1}{n})$, and $\hat{\mathbf{r}}_{\text{fused},i} \in \mathbb{R}^o$ is the lower-dimensional representation of $\mathbf{r}_{\text{fused},i}$.

4. Experiments

This section first details the experimental setup (Section 4.1), including the datasets, underlying VPR descriptors, and metrics used to evaluate HOPS. Section 4.2 introduces two strong baseline multi-reference approaches. We then provide experimental results and analysis for place matching performance, including comparison to single-set baselines (Section 4.3), multi reference-set baselines (Section 4.4), and experiments with reduced dimensionality descriptors (Section 4.5). The section ends with studies on using image augmentations to generate multiple reference sets (Section 4.6), and dataset identification (Section 4.7).

4.1. Experimental Setup

General Setup: Throughout our experiments, we evaluate VPR performance using a single-stage image retrieval pipeline. That is, for every query descriptor, we create a ranked list from the set of reference descriptors in order from most to least similar.

Datasets: To demonstrate the applicability and robustness of our approach across diverse real-world environments and conditions, we evaluate results across three datasets [13, 46, 71], each of which contain images from a unique route captured under varying conditions. The overarching properties of these datasets include urban, suburban, and rural environments captured under various times of day, seasons, weather conditions, and dynamic elements such as structural changes, occlusions, and glare. We also evaluate on the more unstructured Google Landmarks v2 micro and Pittsburgh 250k [73] datasets in the Supplemental Material.

1) Oxford RobotCar [46]: The Oxford RobotCar Dataset contains images from 100 traverses across a route around Oxford throughout the course of a year, capturing the same places under different lighting conditions due to time of day, in changing weather conditions, and with other dynamic changes. We use six separate traverses: sunny, dusk, night, rainy, and two sets of overcast conditions, following prior works [29, 49]. Each set contains 3876 images which have been sampled at ≈1m intervals and have a direct correlation between sets.

2) Nordland [71]: The Nordland dataset is often used as a benchmark in VPR literature because it captures a large geographical area of 729km across the four seasons, including a snowy winter and seasonal changes to trees and plants. In this work, we subsample the original image sets to use 3975 images per season, all with direct correlation across sets. As typical in the literature [26], we remove stationary periods and tunnel sequences.

3) SFU Mountain [13]: The SFU Mountain Dataset provides >8 hrs of sensor data collected with a ClearPath Husky robot on trails around Burnaby Mountain, Canada. We use the following image sets: Dry, Dusk, January, Night, November, September, and Wet. We combine 'Part-A' and 'Part-B' to provide a single set with 385 images per condition.

Baseline VPR Descriptors: To validate the generalizability and applicability of our approach to SOTA VPR descriptors, we evaluate using a large selection of recent methods: CosPlace [9], EigenPlaces [10], MixVPR [3], DinoV2 SALAD [31] (referred to as SALAD from here on), CricaVPR [43], and include AnyLoc [35] and BoQ [4] in the supplemental. For MixVPR [3] and SALAD [31], we use the author provided implementations, and for other VPR descriptors, we use the VPR method evaluation repository released with EigenPlaces[2] which collates the original implementations. We also include NetVLAD [6], as implemented in the Patch-NetVLAD [26] repository, as a common benchmark still used in the literature. We re-iterate that techniques such as CricaVPR [43] are trained so that they explicitly consider the correlations between features of the same place under multiple conditions.

Evaluation Metrics: Recall@N is a metric commonly used for benchmarking VPR methods. It reports the success rate of a VPR method for retrieving the correct reference image in its top N highest ranked references with respect to similarity with the query. $N = 1$ is mathematically equivalent to the precision at 100% recall, assuming every query has a match [68]. Given the difference in sampling between datasets, we assign the following tolerances, as done in prior works [30, 57, 67, 83], for what are considered true matches: RobotCar, $\pm$ 2 images (which is equivalent to 2m); SFU-Mountain, $\pm$ 1 image; Nordland, $\pm$ 0 images (given the distance between images after subsampling).

4.2. Baseline Multi-Reference Approaches

This section introduces two baseline approaches which have explicit access to multiple reference sets at inference time.

Reference Set Pooling: A straightforward approach to leveraging multiple reference sets involves pooling all reference images into a single, larger reference set. Given K individual reference sets $\mathbf{r}^k$, this method constructs a unified set $\mathbf{r}_{\text{pooled}} = \bigcup_{k=1}^{K} \mathbf{r}^k$. During query-time matching, the distance vector $\mathbf{d}_{\text{pooled}}$ is computed by comparing the query vector $\mathbf{q}$ against each feature vector in $\mathbf{r}_{\text{pooled}}$:

$$\mathbf{d}_{\text{pooled}} = \left[d(\mathbf{q}, \mathbf{r}^1), d(\mathbf{q}, \mathbf{r}^2), \ldots, d(\mathbf{q}, \mathbf{r}_M^K) \right]. \quad (4)$$

This simple pooling strategy linearly increases the computational complexity with the number of reference sets K, resulting in an overall complexity of $\mathcal{O}(K \cdot M)$, where M represents the number of images in each reference set. This increase can significantly impact memory usage and processing time, especially in large-scale environments.

Distance Matrix Averaging: Another multi-reference baseline approach entails performing VPR separately on each reference set and then averaging the resultant distance matrices [22]. For each reference set $\mathbf{r}^k$, an independent distance vector $\mathbf{d}^k$ is computed between the query $\mathbf{q}$ and the reference vectors in $\mathbf{r}^k$:

$$\mathbf{d}^k = \left[d(\mathbf{q}, \mathbf{r}_1^k), d(\mathbf{q}, \mathbf{r}_2^k), \ldots, d(\mathbf{q}, \mathbf{r}_M^k) \right]. \quad (5)$$

Once each distance vector $\mathbf{d}^k$ has been computed, they are combined by averaging across corresponding distances, producing a final aggregated distance vector $\mathbf{d}_{\text{avg}}$:

[2]https://github.com/gmberton/VPR-methods-evaluation

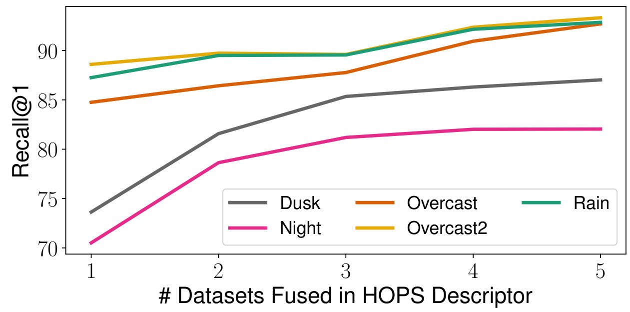

Figure 2. The above plot shows the increase in recall@1 for each Oxford RobotCar query set using our HOPS descriptors with SALAD as more reference sets are progressively fused. The final fused reference descriptors include all non-query sets.

$$\mathbf{d}_{\text{avg}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{d}^k. \quad (6)$$

This averaging approach also scales linearly in computational complexity, $\mathcal{O}(K \cdot M)$, as each reference set requires separate matching computations. However, it offers potential for parallelisation, as the VPR matching for each reference set can be executed independently, enabling efficient processing on multi-core or distributed computing systems. [22] also introduced other approaches which we compare to in the Supplemental Material, however, distance matrix averaging was reported as the highest performing.

Summary: In both baseline approaches, the increased computation and memory requirements limit scalability, particularly in applications requiring real-time performance. Nonetheless, these baseline approaches serve as useful comparisons, providing insight into the trade-offs associated with managing multiple reference sets in VPR tasks. All asserted computational complexities are empirically confirmed in the Supplemental Material.

4.3. Performance Comparisons to Single Set Baselines

First, Tables 1-3 demonstrate that using our HOPS fused descriptors provides significant performance improvements over the best single reference set baselines. For example, on the Oxford RobotCar dataset, HOPS descriptors provide significant improvements to recall, in many cases over absolute 10%, even for SOTA VPR descriptors such as SALAD and CricaVPR (Table 1). Even on the SFU dataset, where the single reference descriptors already perform strongly, our HOPS fused descriptors generally improve performance, for example from 99.0% to 100% on the Dusk query set for SALAD. For the Nordland dataset, HOPS fused descriptors increased R@1 on average by an absolute 2.9% across the 4 query sets for SALAD. For all experiments, we emphasize that the multi-reference set approaches only combine the sets which are not the query being used for evaluation. Figure 2 provides insight into how R@1 is improved incrementally with each additional dataset fused in the HOPS descriptors,

Table 1. Recall@1 on RobotCar datasets: The table is divided into single-reference and multi-reference approaches. The best single-reference result is underlined and the best multi-reference result is **bolded**. Comparisons in this table should be made vertically down columns. Importantly, our HOPS fused descriptors are near unanimously better than the best single-reference results (in **28/30** cases) **and** better than alternative multi-reference approaches the majority of the time (in **22/30** cases).

Queries →	Dusk	Night	Overcast	Overcast2	Rain	Dusk	Night	Overcast	Overcast2	Rain
References	**NetVLAD (4096D)**					**SALAD (8448D)**				
Sunny	25.5	9.8	68.0	79.1	73.5	73.6	70.5	84.8	88.6	87.3
Dusk	-	19.9	24.1	23.0	23.3	-	71.1	68.1	68.8	70.5
Night	27.6	-	13.6	11.6	10.6	71.7	-	66.4	63.7	66.2
Overcast	33.0	13.4	-	79.6	72.8	74.3	71.0	-	88.3	87.6
Overcast2	27.0	11.2	75.9	-	73.2	74.4	69.1	86.8	-	87.2
Rain	29.2	9.1	68.8	72.7	-	76.3	68.6	85.8	86.6	-
dMat Avg [22]	49.0	24.6	79.4	85.9	82.6	86.2	81.3	90.2	91.4	91.5
Pooling	36.9	20.3	80.2	85.8	80.0	79.9	74.8	90.0	91.8	91.2
HOPS (Ours)	**49.8**	**27.7**	**83.7**	**89.5**	**85.7**	**87.1**	**82.1**	**92.8**	**93.3**	**92.9**

Queries →	Dusk	Night	Overcast	Overcast2	Rain	Dusk	Night	Overcast	Overcast2	Rain
References	**MixVPR (4096D)**					**CosPlace (512D)**				
Sunny	69.0	50.9	86.3	91.2	88.7	44.1	14.0	78.3	86.5	84.6
Dusk	-	59.2	60.1	61.4	63.6	-	21.2	42.7	42.2	44.1
Night	64.6	-	52.2	50.3	48.3	46.5	-	28.8	27.2	26.5
Overcast	71.7	57.2	-	90.6	89.6	48.6	18.2	-	85.3	83.6
Overcast2	67.4	52.0	89.1	-	89.5	45.1	15.6	84.2	-	84.2
Rain	68.3	46.0	87.1	88.8	-	44.8	15.0	81.6	83.9	-
dMat Avg [22]	82.9	**70.0**	91.5	93.6	92.7	56.0	**22.9**	79.8	84.6	83.7
Pooling	77.1	60.1	92.0	93.7	93.4	55.9	21.3	**87.8**	**90.5**	89.9
HOPS (Ours)	**83.1**	68.8	**93.3**	**94.7**	**94.7**	**57.0**	19.4	85.9	89.8	**90.2**

Queries →	Dusk	Night	Overcast	Overcast2	Rain	Dusk	Night	Overcast	Overcast2	Rain
References	**EigenPlaces (512D)**					**CricaVPR (10752D)**				
Sunny	42.3	13.0	81.8	88.3	87.5	81.4	77.9	90.6	93.9	92.4
Dusk	-	22.7	42.0	41.9	43.5	-	77.8	77.2	79.4	80.8
Night	46.3	-	25.9	25.7	22.9	81.1	-	75.5	73.6	72.7
Overcast	48.2	19.1	-	87.9	86.9	85.7	81.0	-	93.9	93.5
Overcast2	42.7	14.0	86.5	-	86.1	84.2	77.2	92.2	-	93.1
Rain	44.3	15.2	84.8	86.2	-	85.0	75.9	92.5	92.9	-
dMat Avg [22]	**55.7**	**23.8**	84.7	88.2	86.4	94.0	89.2	95.6	96.9	96.5
Pooling	54.3	22.7	**90.0**	**92.1**	**91.1**	89.6	81.6	95.4	96.0	95.9
HOPS (Ours)	54.9	20.3	89.2	92.0	**91.1**	**94.8**	**91.0**	**96.6**	**97.5**	**97.4**

showing the maximum performance occurs with the fusion of all reference sets in this case.

There are three outlier cases where HOPS descriptors perform slightly worse than the best single reference set: using CosPlace or EigenPlaces on Oxford RobotCar Night query (1.8% and 2.4% reduction in R@1), and CosPlace on the Nordland Summer query (1.0% reduction in R@1). Though it might not be the only factor, the relatively low dimensionality of EigenPlaces and CosPlace (512D) intuitively makes them less suitable for HOPS, given that HDC principles assume vectors have thousands or tens of thousands of dimensions. Additional experiments using CosPlace, included in the Supplemental Material, indicate the style of training could also be a factor. Further investigation may provide insights into how HDC can be applied in these cases.

Figure 3 provides insights into how the HOPS fused descriptors are improving VPR performance. It shows that they are, especially for already high performance baseline methods, further reducing the metric error of place matches that are already quite close to the ground truth match. This is a different phenomenon to typical improvements in VPR where egregiously wrong matches are "corrected" by improved features to fall within the correct zone around the ground truth. We suspect the reason for this is that the stacking/fusing of multiple reference descriptors for each place is reducing the volatility of matching in the region near the ground truth location (in datasets where subsequent frames often belong to a similar spatial location), meaning the true best match is less likely to be "outmatched" by a nearby visually similar images. For VPR descriptors with lower baseline performance, such as NetVLAD, there are still a high number of large errors corrected as well.

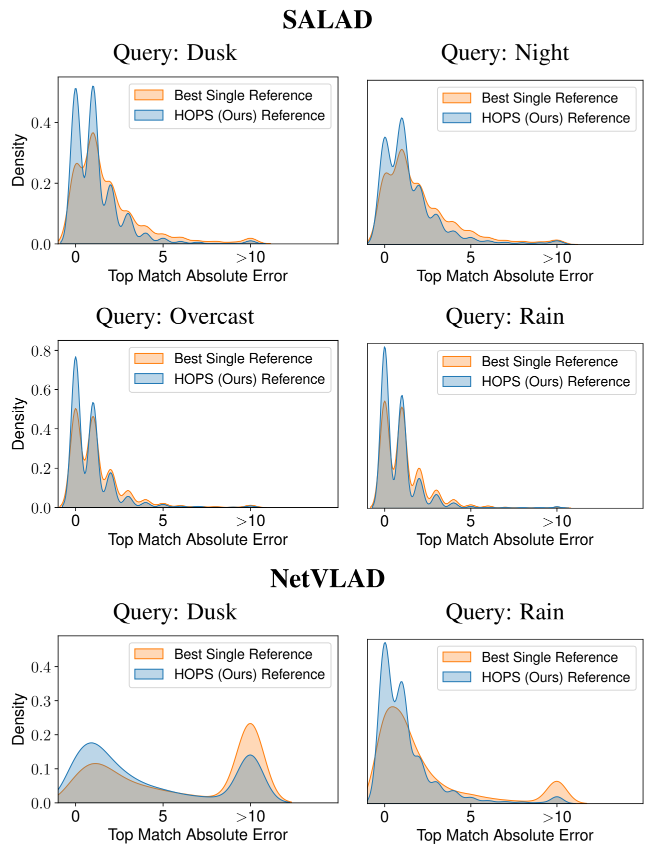

Figure 3. **Top:** Match error density plots for the top VPR match on Oxford RobotCar sets using SALAD descriptors (error measured in frames, ≈ 1m/frame for RobotCar). For already high-performing VPR descriptors, our HOPS fused descriptors are able to further reduce the error of matches that are already made in close proximity to the true match, disambiguating spatially close places. **Bottom:** For lower performing baselines, such as NetVLAD, our HOPS fused descriptors corrected a high number of large errors as well.

4.4. Comparisons to Multi Reference Set Baselines

With respect to the multi-reference set approaches, Tables 1-3 show that while the distance matrix averaging and pooling methods typically provide improvements over single-reference methods, HOPS descriptors provide the highest R@1 in 69 out of 90 cases. In addition, we reiterate that HOPS descriptors maintain the same computation and

Table 2. Recall@1 on Nordland datasets: See Table 1 for format conventions. Our HOPS fused descriptor outperforms the best single-reference results in **23/24** cases and the other multi-reference approaches in **18/24** cases.

Queries →	**Fall**	**Spring**	**Summer**	**Winter**	**Fall**	**Spring**	**Summer**	**Winter**	**Fall**	**Spring**	**Summer**	**Winter**	**Fall**	**Spring**	**Summer**	**Winter**	**Fall**	**Spring**	**Summer**	**Winter**	**Fall**	**Spring**	**Summer**	**Winter**
References	**NetVLAD (4096D)**				**SALAD (8448D)**				**MixVPR (4096D)**				**CosPlace (512D)**				**EigenPlaces (512D)**				**CricaVPR (10752D)**			
Fall	-	43.3	61.5	16.1	-	80.2	79.9	72.8	-	78.8	78.8	66.9	-	76.9	77.6	61.5	-	77.5	78.5	63.3	-	81.6	81.3	77.3
Spring	37.0	-	35.2	16.2	78.4	-	76.8	75.8	73.3	-	69.3	73.6	71.2	-	65.3	70.5	74.5	-	68.8	67.5	80.6	-	77.8	79.6
Summer	61.1	41.0	-	15.5	80.0	78.2	-	71.1	78.6	75.5	-	63.7	77.7	72.3	-	56.6	78.8	74.3	-	59.5	81.4	80.4	-	74.6
Winter	12.4	18.1	11.9	-	71.0	76.9	69.3	-	57.2	70.9	52.9	-	51.2	68.4	46.5	-	57.1	71.0	52.9	-	73.9	79.8	70.9	-
dMat Avg [22]	57.3	56.8	55.7	**26.6**	81.2	81.7	80.0	79.4	80.3	81.3	77.9	76.5	77.7	79.6	73.8	70.9	79.9	80.4	77.2	72.5	83.2	83.2	81.5	82.1
Pooling	63.2	50.9	62.9	18.5	81.5	81.9	80.5	77.3	80.8	81.5	**79.9**	75.8	80.4	80.3	**78.6**	**71.5**	81.0	81.1	**79.3**	68.7	82.9	**84.1**	**82.5**	81.1
HOPS (Ours)	**63.5**	**62.7**	**63.3**	25.7	**82.1**	**82.0**	**80.7**	**79.7**	**81.7**	**81.8**	79.2	**77.1**	**80.4**	**81.2**	76.6	71.3	**81.2**	**81.6**	78.6	**72.7**	**83.9**	83.8	**82.5**	**82.4**

Table 3. Recall@1 on SFU-Mountain datasets: See Table 1 for format conventions. Our HOPS fused descriptor outperforms the best single-reference results in **100%** of cases and the other multi-reference approaches in **29/36** cases.

Queries →	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**
References	**NetVLAD (4096D)**						**SALAD (8448D)**						**MixVPR (4096D)**					
Dry	-	43.9	25.5	33.0	23.6	38.4	-	99.0	92.5	96.9	94.8	96.6	-	94.3	81.6	89.9	86.8	92.0
Dusk	52.7	-	28.6	36.9	34.0	62.1	99.0	-	95.6	96.1	94.0	98.2	98.4	-	90.9	94.3	93.3	98.4
Jan	25.5	34.6	-	26.5	21.8	31.2	94.6	96.9	-	95.8	93.5	93.5	75.1	84.4	-	71.7	70.7	79.5
Nov	30.1	31.2	23.1	-	33.8	32.7	95.3	94.0	94.8	-	96.4	96.4	86.0	84.2	75.6	-	92.2	88.1
Sept	27.0	30.7	20.5	38.4	-	29.1	94.0	88.8	93.5	95.3	-	92.5	84.9	86.5	75.1	94.0	-	85.5
Wet	44.4	63.9	28.3	38.2	28.8	-	97.7	98.7	94.6	95.1	93.5	-	95.8	96.9	92.7	95.1	92.5	-
dMat Avg [22]	63.4	62.3	40.5	61.0	48.3	66.8	99.5	99.5	98.7	99.0	98.4	99.2	99.2	98.4	93.2	**99.5**	97.4	98.7
Pooling	59.7	68.8	38.2	50.9	42.6	66.2	**99.7**	99.2	98.2	98.0	97.1	**99.5**	99.2	98.4	95.3	98.7	96.6	**99.7**
HOPS (Ours)	**68.3**	**74.6**	**47.5**	**68.1**	**56.6**	**76.1**	**99.7**	**100**	**99.2**	**99.5**	**98.7**	99.2	**99.5**	**99.5**	**97.1**	**99.5**	**97.9**	99.5

Queries →	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**	**Dry**	**Dusk**	**Jan**	**Nov**	**Sept**	**Wet**
References	**CosPlace (512D)**						**EigenPlaces (512D)**						**CricaVPR (10752D)**					
Dry	-	91.7	79.2	82.9	81.0	88.6	-	92.5	83.1	87.8	87.0	93.0	-	98.7	91.9	95.8	93.0	97.9
Dusk	91.7	-	82.1	84.9	77.7	94.8	95.1	-	89.4	90.4	88.1	97.1	99.2	-	97.4	96.6	94.8	99.0
Jan	81.0	86.5	-	77.9	70.6	85.5	86.8	88.3	-	82.1	77.7	86.8	95.6	96.4	-	94.5	93.5	95.6
Nov	80.8	80.8	72.5	-	89.9	80.8	88.3	86.0	79.5	-	93.5	87.0	94.5	96.1	93.8	-	97.7	96.1
Sept	77.9	75.1	68.3	89.6	-	75.8	81.8	80.8	73.5	92.7	-	83.9	92.7	91.9	93.0	95.8	-	90.9
Wet	94.0	95.1	84.9	88.3	84.7	-	95.1	97.1	91.7	91.2	89.9	-	97.1	99.2	96.6	96.6	93.5	-
dMat Avg [22]	95.6	96.6	87.3	94.5	93.8	96.1	97.1	98.2	94.0	98.2	96.1	97.9	99.5	99.7	**98.7**	99.5	98.2	99.5
Pooling	**98.4**	98.2	93.2	96.1	**95.8**	**98.4**	**99.0**	98.4	95.3	95.6	**96.4**	**99.0**	99.5	99.5	97.9	98.4	**99.0**	99.2
HOPS (Ours)	97.9	**98.7**	**95.1**	**96.9**	95.6	98.2	98.7	**99.5**	**97.4**	**99.2**	96.1	**99.0**	**100.0**	**100.0**	**98.7**	**99.7**	**99.0**	**99.7**

memory costs as for the single-reference set approach, providing significant advantage over the pooling and averaging approaches whose computational and storage complexities increase linearly with the number of reference sets.

One can observe that the reference pooling approach is more performant for lower dimensional descriptors such as CosPlace and EigenPlaces, whereas the distance matrix averaging performs better for the other higher dimensional descriptors — as highlighted in the previous subsection, these results are intuitive given that HDC assumes high-dimensional feature vectors, but both CosPlace and EigenPlaces are relatively low-dimensional.

4.5. Reducing Dimensionality

For large scale image retrieval tasks, the size of image descriptors can have a significant effect on the computational overhead and required memory allocation. Here, we investigate the possible advantages HOPS fused descriptors have for reducing the dimensionality of existing SOTA VPR methods. That is, given a VPR descriptor and a selection of separate reference sets which achieve a certain performance, how can HOPS fused descriptors reduce dimensionality while still matching or exceeding this original performance. We used Gaussian Random Projection to reduce descriptor dimensionality in these experiments because (similarly to HOPS) it also leverages properties of high dimensional spaces (Section 3.3), however, this method could be substituted with other dimensionality reduction methods.

Figure 4 shows representative results using CosPlace, MixVPR, SALAD, and CricaVPR on the RobotCar Dusk dataset (see the Supplementary Material for full results). Our

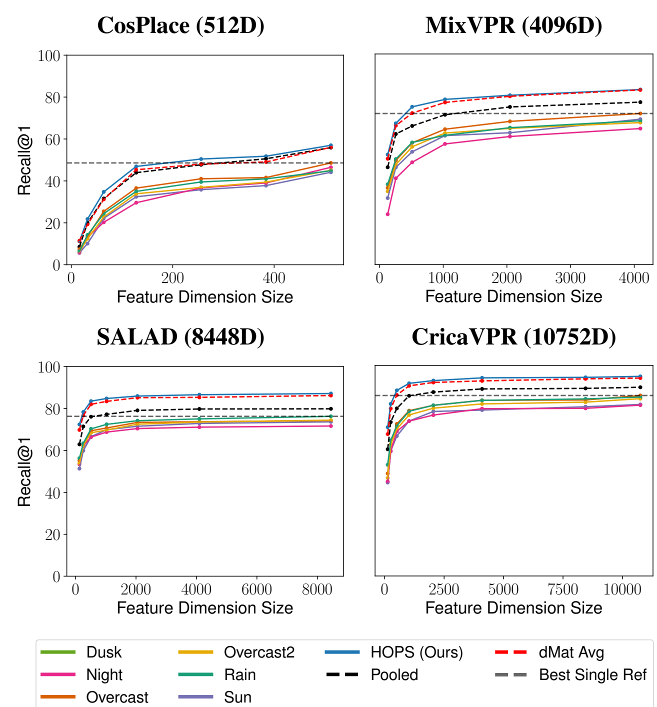

Figure 4. Recall@1 performance for different VPR descriptors across the Oxford RobotCar Dusk set as dimensionality is reduced using Gaussian Random Projection. Our HOPS fused descriptors are able to maintain the highest R@1, allowing for an alternative use where descriptor dimensionality can be reduced by up to 97% while exceeding the best single-reference performance at full-size.

Table 4. Recall@1 on RobotCar datasets Using Synthetic Changes

Queries → References ↓	DinoV2 SALAD (8448D)				
	Dusk	Night	Overcast	Overcast2	Rain
Sunny	73.6	70.5	84.8	88.6	87.3
Synthetic Dark [45]	70.9	68.4	73.2	77.7	77.4
Poisson Noise	64.3	60.6	77.1	80.8	79.9
Downsample-Upsample	68.8	67.0	80.1	83.2	82.7
dMat Avg [22]	75.8	**73.2**	82.8	87.3	86.5
Pooling	73.5	69.4	**84.5**	88.5	86.8
HOPS (Ours)	**76.1**	72.7	84.2	**88.8**	**87.7**

proposed HOPS fused descriptors exceed the performance of the best full-sized single-reference results with a much smaller descriptor size; about a 50% and 95% reduction for CosPlace and CricaVPR, respectively, *i.e.* a recall of 85.7% for CricaVPR can either be obtained using the 10752D original descriptor or our 512D reduced-dimension fused HOPS descriptor. Our HOPS fused approach and single-reference approaches follow a similar trend, with performance gradually being more affected by dimensionality reduction before a sudden drop off in R@1 – importantly, HOPS maintains the highest R@1 values across all descriptor dimensions.

4.6. Substituting Synthetic Image Augmentations

So far, we have explored multi-reference VPR approaches with the assumption that multiple reference sets have been collected from real-world data. However, here we show that multiple reference sets can also be created by synthetically augmenting a single reference dataset. This is one possible way to enable the use of our HOPS fused descriptors in single-reference scenarios.

Table 4 shows a proof of concept study where image augmentations such as synthetic darkening of an image (generated using [45]), the application of Poisson noise, and downsampling and re-upsampling an image are used to exploit some of the performance benefits of HOPS fused descriptors without requiring *real* multiple reference traverses.

For the RobotCar Dusk and Night sets, HOPS fused descriptors using the synthetic condition changes improve R@1 by absolute 2.5% and 2.2% respectively over the best single-reference results. We note that while we improve performance on average by 1.0%, in the Overcast query the performance reduces slightly by 0.6%. More results are included in the Supplemental Material.

4.7. Dataset Identification

Here we provide a brief investigation into another possible application of descriptor fusing via hyperdimensional computing: identifying in which environment one is located based on a single descriptor, *i.e.* all reference descriptors of a dataset are fused into a *single overall dataset descriptor*. Individual query descriptors from each of the datasets (and not from any reference set), can then be compared to these dataset descriptors to determine which dataset the query is from. By using all available non-query sets for each dataset and fusing them, this results in dataset identification with an accuracy $>99.7\%$ for all datasets. Full details can be found in the Supplemental Material.

5. Conclusion

This paper investigated how reference sets captured under varying conditions can be fused with minimal compute and storage overhead using a hyperdimensional computing framework, to improve VPR performance under appearance change. Through an extensive set of experiments, we demonstrated that our HOPS fused descriptors improve recall@1 over the best single-reference results for several multi-condition datasets and SOTA VPR methods. We also showed that while other multi-reference approaches also improve over the single-reference case, our HOPS fused descriptors are generally the highest performing whilst also avoiding the computation and memory costs incurred in these other multi-reference approaches. This research further highlights the potential of the HDC framework for improving VPR, which is complementary to ongoing research efforts on extracting more invariant place features.

Multiple reference sets can be obtained both from real world sensory data but also from synthetically generated image transformations, especially when multiple reference sets are not available: we demonstrated the performance achievement of the latter when fusing descriptors from multiple image augmentations of a single reference set.

Finally this research also explored how to reduce computation and memory costs for real-time deployment without sacrificing performance: HOPS fused descriptors can maintain the same performance as the best single-reference results whilst reducing descriptor dimensionality by up to an order of magnitude. We also demonstrated how the HDC framework can be used to create whole dataset descriptors which can be used for identifying which dataset a query is from.

Future work can further improve both the capability and efficiency of HOPS descriptors by deeper investigating the effect of bundling on features and by exploring whether HOPS fused descriptors can be used to train more robust feature extractors. This could include investigating how well HOPS descriptors maintain fine-grained features. The work here primarily investigated the combination of multiple reference images from the *same* location: preliminary investigation has also indicated that it is possible to stack together reference imagery from completely different datasets with no computational and minimal performance penalty, providing the possibility for highly compressible encoding of many maps into a single representation.

Acknowledgements. This research was partially supported and funded by the QUT Centre for Robotics, ARC Laureate Fellowship FL210100156 to MM, and ARC DECRA Fellowship DE240100149 to TF.

References

[1] Dimitris Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003. 4

[2] Amar Ali-bey, Brahim Chaib-draa, and Philippe Giguère. Gsv-cities: Toward appropriate supervised visual place recognition. *Neurocomputing*, 2022. 1, 2, 3

[3] Amar Ali-Bey, Brahim Chaib-Draa, and Philippe Giguere. Mixvpr: Feature mixing for visual place recognition. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2998–3007, 2023. 1, 2, 4

[4] Amar Ali-bey, Brahim Chaib-draa, and Philippe Giguère. Boq: A place is worth a bag of learnable queries. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17794–17803, 2024. 2, 4

[5] Relja Arandjelovic and Andrew Zisserman. All about VLAD. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1578–1585, 2013. 2

[6] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5297–5307, 2016. 2, 5

[7] Giovanni Barbarani, Mohamad Mostafa, Hajali Bayramov, Gabriele Trivigno, Gabriele Berton, Carlo Masone, and Barbara Caputo. Are local features all you need for cross-domain visual place recognition? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6155–6165, 2023. 2

[8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 2

[9] Gabriele Berton, Carlo Masone, and Barbara Caputo. Rethinking visual geo-localization for large-scale applications. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4878–4888, 2022. 2, 4

[10] Gabriele Berton, Gabriele Trivigno, Barbara Caputo, and Carlo Masone. Eigenplaces: Training viewpoint robust models for visual place recognition. In *IEEE/CVF International Conference on Computer Vision*, pages 11080–11090, 2023. 1, 2, 4

[11] Gabriele Berton, Lorenz Junglas, Riccardo Zaccone, Thomas Pollok, Barbara Caputo, and Carlo Masone. MeshVPR: Citywide Visual Place Recognition Using 3D Meshes. In *European Conference on Computer Vision*, 2024. 2

[12] Gabriele Moreno Berton, Valerio Paolicelli, Carlo Masone, and Barbara Caputo. Adaptive-attentive geolocalization from few queries: A hybrid approach. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2918–2927, 2021. 1

[13] Jake Bruce, Jens Wawerla, and Richard Vaughan. The SFU mountain dataset: Semi-structured woodland trails under changing environmental conditions. In *Workshop on Visual Place Recognition in Changing Environments, IEEE International Conference on Robotics and Automation*, 2015. 4

[14] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016. 1

[15] Bingyi Cao, André Araujo, and Jack Sim. Unifying deep local and global features for image search. In *The European Conference on Computer Vision*, pages 726–743, 2020. 2

[16] Athanasios Chalvatzaras, Ioannis Pratikakis, and Angelos A Amanatiadis. A survey on map-based localization techniques for autonomous vehicles. *IEEE Transactions on Intelligent Vehicles*, 8(2):1574–1596, 2022. 1

[17] Brian Cheung, Alexander Terekhov, Yubei Chen, Pulkit Agrawal, and Bruno Olshausen. Superposition of many models into one. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[18] Winston Churchill and Paul Newman. Practice makes perfect? managing and leveraging visual experiences for lifelong navigation. In *IEEE International Conference on Robotics and Automation*, pages 4525–4532, 2012. 2, 3

[19] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, The European Conference on Computer Vision*, pages 1–2, 2004. 2

[20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 2

[21] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine*, 13(2):99–110, 2006. 1

[22] Tobias Fischer and Michael Milford. Event-based visual place recognition with ensembles of temporal windows. *IEEE Robotics and Automation Letters*, 5(4):6924–6931, 2020. 1, 5, 6, 7, 8

[23] Sourav Garg, Tobias Fischer, and Michael Milford. Where is your place, visual place recognition? In *International Joint Conferences on Artificial Intelligence*, pages 4416–4425, 2021. 1

[24] Lulu Ge and Keshab K Parhi. Classification using hyperdimensional computing: A review. *IEEE Circuits and Systems Magazine*, 20(2):30–47, 2020. 3

[25] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):87–110, 2022. 2

[26] Stephen Hausler, Sourav Garg, Ming Xu, Michael Milford, and Tobias Fischer. Patch-NetVLAD: Multi-scale fusion of locally-global descriptors for place recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2, 4, 5

[27] Kuanxu Hou, Delei Kong, Junjie Jiang, Hao Zhuang, Xinjie Huang, and Zheng Fang. Fe-fusion-vpr: Attention-based multi-scale network architecture for visual place recognition by fusing frames and events. *IEEE Robotics and Automation Letters*, 8(6):3526–3533, 2023. 2, 3

[28] Hanjiang Hu, Zhijian Qiao, Ming Cheng, Zhe Liu, and Hesheng Wang. Dasgil: Domain adaptation for semantic and geometric-aware image-based localization. *IEEE Transactions on Image Processing*, 30:1342–1353, 2020. 1

[29] Somayeh Hussaini, Michael Milford, and Tobias Fischer. Spiking neural networks for visual place recognition via weighted neuronal assignments. *IEEE Robotics and Automation Letters*, 7(2):4094–4101, 2022. 4

[30] Somayeh Hussaini, Michael Milford, and Tobias Fischer. Applications of spiking neural networks in visual place recognition. *arXiv preprint arXiv:2311.13186*, 2023. 5

[31] Sergio Izquierdo and Javier Civera. Optimal transport aggregation for visual place recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17658–17668, 2024. 1, 2, 4

[32] Adam Jacobson, Zetao Chen, and Michael Milford. Autonomous multisensor calibration and closed-loop fusion for slam. *Journal of Field Robotics*, 32(1):85–122, 2015. 2, 3

[33] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. Aggregating local descriptors into a compact image representation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3304–3311, 2010. 2

[34] Geethan Karunaratne, Manuel Le Gallo, Giovanni Cherubini, Luca Benini, Abbas Rahimi, and Abu Sebastian. In-memory hyperdimensional computing. *Nature Electronics*, 3(6):327–337, 2020. 3

[35] Nikhil Keetha, Avneesh Mishra, Jay Karhade, Krishna Murthy Jatavallabhula, Sebastian Scherer, Madhava Krishna, and Sourav Garg. Anyloc: Towards universal visual place recognition. *IEEE Robotics and Automation Letters*, 2023. 1, 2, 4

[36] Denis Kleyko, Abbas Rahimi, Dmitri A Rachkovskij, Evgeny Osipov, and Jan M Rabaey. Classification and recall with binary hyperdimensional computing: Tradeoffs in choice of density and mapping characteristics. *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):5880–5898, 2018. 3

[37] Denis Kleyko, Dmitri Rachkovskij, Evgeny Osipov, and Abbas Rahimi. A survey on hyperdimensional computing aka vector symbolic architectures, part ii: Applications, cognitive models, and challenges. *ACM Computing Surveys*, 55(9): 1–52, 2023. 2

[38] Debasis Kumar and Naveed Muhammad. A survey on localization for autonomous vehicles. *IEEE Access*, 2023. 1

[39] María Leyva-Vallina, Nicola Strisciuglio, and Nicolai Petkov. Data-efficient large scale place recognition with graded similarity supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23487–23496, 2023. 2

[40] Chris Linegar, Winston Churchill, and Paul Newman. Work smart, not hard: Recalling relevant experiences for vast-scale but time-constrained localisation. In *IEEE International Conference on Robotics and Automation*, pages 90–97, 2015. 2, 3

[41] D.G. Lowe. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*, pages 1150–1157 vol.2, 1999. 2

[42] Stephanie Lowry, Niko Sünderhauf, Paul Newman, John J Leonard, David Cox, Peter Corke, and Michael J Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics*, 32(1):1–19, 2015. 1

[43] Feng Lu, Xiangyuan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, and Chun Yuan. CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16772–16782, 2024. 1, 2, 3, 4, 5

[44] Feng Lu, Lijun Zhang, Xiangyuan Lan, Shuting Dong, Yaowei Wang, and Chun Yuan. Towards seamless adaptation of pre-trained models for visual place recognition. In *The International Conference on Learning Representations*, 2024. 1, 2

[45] Rundong Luo, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Similarity min-max: Zero-shot day-night domain adaptation. In *IEEE/CVF International Conference on Computer Vision*, pages 8104–8114, 2023. 8

[46] Will Maddern, Geoff Pascoe, Chris Linegar, and Paul Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 4

[47] Carlo Masone and Barbara Caputo. A survey on deep visual place recognition. *IEEE Access*, 9:19516–19547, 2021. 1, 2

[48] Antoine Miech, Ivan Laptev, and Josef Sivic. Learnable pooling with context gating for video classification. *arXiv preprint arXiv:1706.06905*, 2017. 2

[49] Timothy L Molloy, Tobias Fischer, Michael Milford, and Girish N Nair. Intelligent reference curation for visual place recognition via bayesian selective fusion. *IEEE Robotics and Automation Letters*, 6(2):588–595, 2020. 1, 2, 3, 4

[50] AC Murillo, Carlos Sagüés, José Jesús Guerrero, Toon Goedemé, Tinne Tuytelaars, and Luc Van Gool. From omnidirectional images to hierarchical localization. *IEEE Robotics and Autonomous Systems*, 55(5):372–382, 2007. 1

[51] Peer Neubert and Stefan Schubert. Hyperdimensional computing as a framework for systematic aggregation of image descriptors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16938–16947, 2021. 2, 3

[52] Peer Neubert, Stefan Schubert, and Peter Protzel. An introduction to hyperdimensional computing for robotics. *KI-Künstliche Intelligenz*, 33(4):319–330, 2019. 2, 3

[53] Peer Neubert, Stefan Schubert, Kenny Schlegel, and Peter Protzel. Vector semantic representations as descriptors for visual place recognition. In *Robotics: Science and Systems*, pages 1–11, 2021. 3

[54] Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, and Bohyung Han. Large-scale image retrieval with attentive deep local features. In *IEEE/CVF International Conference on Computer Vision*, pages 3456–3465, 2017. 2

[55] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 2

[56] Florent Perronnin, Yan Liu, Jorge Sánchez, and Hervé Poirier. Large-scale image retrieval with compressed fisher vectors. In *Computer Society Conference on Computer Vision and Pattern Recognitionn*, pages 3384–3391, 2010. 2

[57] Tomáš Pivoňka and Libor Přeučil. On model-free re-ranking for visual place recognition with deep learned local features. *IEEE Transactions on Intelligent Vehicles*, 2024. 5

[58] Chiara Plizzari, Gabriele Goletto, Antonino Furnari, Siddhant Bansal, Francesco Ragusa, Giovanni Maria Farinella, Dima Damen, and Tatiana Tommasi. An outlook into the future of egocentric vision. *International Journal of Computer Vision*, pages 1–57, 2024. 1

[59] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Cnn image retrieval learns from bow: Unsupervised fine-tuning with hard examples. In *The European Conference on Computer Vision*, pages 3–20, 2016. 2

[60] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1655–1668, 2018. 2

[61] Jerome Revaud, Jon Almazán, Rafael S Rezende, and Cesar Roberto de Souza. Learning with average precision: Training image retrieval with a listwise loss. In *IEEE/CVF International Conference on Computer Vision*, pages 5107–5116, 2019. 2

[62] Paul-Edouard Sarlin, Frédéric Debraine, Marcin Dymczyk, Roland Siegwart, and Cesar Cadena. Leveraging deep visual descriptors for hierarchical efficient localization. In *Conference on Robot Learning*, pages 456–465, 2018. 1

[63] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12716–12725, 2019. 1

[64] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4938–4947, 2020. 2

[65] Paul-Edouard Sarlin, Mihai Dusmanu, Johannes L Schönberger, Pablo Speciale, Lukas Gruber, Viktor Larsson, Ondrej Miksik, and Marc Pollefeys. Lamar: Benchmarking localization and mapping for augmented reality. In *European Conference on Computer Vision*, pages 686–704, 2022. 1

[66] Kenny Schlegel, Peer Neubert, and Peter Protzel. A comparison of vector symbolic architectures. *Artificial Intelligence Review*, 55(6):4523–4555, 2022. 3

[67] M Sc Stefan Schubert. Visual place recognition in changing environments using additional data-inherent knowledge. *Technische Universität Chemnitz, Chemnitz*, 2023. 5

[68] Stefan Schubert, Peer Neubert, Sourav Garg, Michael Milford, and Tobias Fischer. Visual place recognition: A tutorial. *IEEE Robotics & Automation Magazine*, 2023. 1, 2, 5

[69] Sriram Siva and Hao Zhang. Omnidirectional multisensory perception fusion for long-term place recognition. In *IEEE International Conference on Robotics and Automation*, pages 5175–5181, 2018. 2, 3

[70] Sivic and Zisserman. Video google: A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision*, pages 1470–1477, 2003. 2

[71] N Sünderhauf, Peer Neubert, and Peter Protzel. Are we there yet? challenging seqslam on a 3000 km journey across all four seasons. *Workshop on Long-term Autonomy, IEEE International Conference on Robotics and Automation*, 2013. 4

[72] Carl Toft, Will Maddern, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Tomas Pajdla, et al. Long-term visual localization revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2074–2088, 2020. 1

[73] Akihiko Torii, Josef Sivic, Tomas Pajdla, and Masatoshi Okutomi. Visual place recognition with repetitive structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 883–890, 2013. 4

[74] Gabriele Trivigno, Gabriele Berton, Juan Aragon, Barbara Caputo, and Carlo Masone. Divide&classify: Fine-grained classification for city-wide visual geo-localization. In *IEEE/CVF International Conference on Computer Vision*, 2023. 2

[75] Konstantinos A Tsintotas, Loukas Bampis, and Antonios Gasteratos. The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection. *IEEE Transactions on Intelligent Transportation Systems*, 23 (11):19929–19953, 2022. 1

[76] Issar Tzachor, Boaz Lerner, Matan Levy, Michael Green, Tal Berkovitz Shalev, Gavriel Habib, Dvir Samuel, Noam Korngut Zailer, Or Shimshi, Nir Darshan, et al. EffoVPR: Effective Foundation Model Utilization for Visual Place Recognition. *arXiv preprint arXiv:2405.18065*, 2024. 1, 2

[77] Inam Ullah, Deepak Adhikari, Habib Khan, M Shahid Anwar, Shabir Ahmad, and Xiaoshan Bai. Mobile robot localization: Current challenges and future prospective. *Computer Science Review*, 53, 2024. 1

[78] Olga Vysotska and Cyrill Stachniss. Effective visual place recognition using multi-sequence maps. *IEEE Robotics and Automation Letters*, 4(2):1730–1736, 2019. 2, 3

[79] Ruotong Wang, Yanqing Shen, Weiliang Zuo, Sanping Zhou, and Nanning Zheng. Transvpr: Transformer-based place recognition with multi-level attention aggregation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13648–13657, 2022. 2

[80] Frederik Warburg, Soren Hauberg, Manuel Lopez-Antequera, Pau Gargallo, Yubin Kuang, and Javier Civera. Mapillary street-level sequences: A dataset for lifelong place recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2626–2635, 2020. 1, 3

[81] Samuel Wilson, Tobias Fischer, Niko Sünderhauf, and Feras Dayoub. Hyperdimensional feature fusion for out-of-distribution detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2644–2654, 2023. 3

[82] Zhe Xin, Xiaoguang Cui, Jixiang Zhang, Yiping Yang, and Yanqing Wang. Real-time visual place recognition based on analyzing distribution of multi-scale cnn landmarks. *Journal of Intelligent & Robotic Systems*, 94:777–792, 2019. 2, 3

[83] Ming Xu, Niko Snderhauf, and Michael Milford. Probabilistic visual place recognition for hierarchical localization. *IEEE Robotics and Automation Letters*, 6(2):311–318, 2020. 5

[84] Min Yang, Dongliang He, Miao Fan, Baorong Shi, Xuetong Xue, Fu Li, Errui Ding, and Jizhou Huang. Dolg: Single-stage image retrieval with deep orthogonal fusion of local and

global features. In *IEEE/CVF International Conference on Computer Vision*, pages 11772–11781, 2021. 2

[85] Huan Yin, Xuecheng Xu, Sha Lu, Xieyuanli Chen, Rong Xiong, Shaojie Shen, Cyrill Stachniss, and Yue Wang. A survey on global lidar localization: Challenges, advances and open problems. *International Journal of Computer Vision*, 132(8):3139–3171, 2024. 1

[86] Jun Yu, Chaoyang Zhu, Jian Zhang, Qingming Huang, and Dacheng Tao. Spatial pyramid-enhanced NetVLAD with weighted triplet loss for place recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 31(2):661–674, 2019. 2, 3

[87] Hao Zhang, Xin Chen, Heming Jing, Yingbin Zheng, Yuan Wu, and Cheng Jin. Etr: An efficient transformer for re-ranking in visual place recognition. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5665–5674, 2023. 2

[88] Xiwu Zhang, Lei Wang, and Yan Su. Visual place recognition: A survey from deep learning perspective. *Pattern Recognition*, 113:107760, 2021. 1

[89] Sijie Zhu, Linjie Yang, Chen Chen, Mubarak Shah, Xiaohui Shen, and Heng Wang. R2former: Unified retrieval and reranking transformer for place recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19370–19380, 2023. 2