Through the Lens of Doubt: Robust and Efficient Uncertainty Estimation for Visual Place Recognition

Emily Miller[1], Michael Milford[2], Muhammad Burhan Hafez[1], SD Ramchurn[1] and Shoaib Ehsan[1,3]

***Abstract*—Visual Place Recognition (VPR) enables robots and autonomous vehicles to identify previously visited locations by matching current observations against a database of known places. However, VPR systems face significant challenges when deployed across varying visual environments, lighting conditions, seasonal changes, and viewpoints changes. Failure-critical VPR applications, such as loop closure detection in simultaneous localization and mapping (SLAM) pipelines, require robust estimation of place matching uncertainty. We propose three training-free uncertainty metrics that estimate prediction confidence by analyzing inherent statistical patterns in similarity scores from any existing VPR method. Similarity Distribution (SD) quantifies match distinctiveness by measuring score separation between candidates; Ratio Spread (RS) evaluates competitive ambiguity among top-scoring locations; and Statistical Uncertainty (SU) is a combination of SD and RS that provides a unified metric that generalizes across datasets and VPR methods without requiring validation data to select the optimal metric. All three metrics operate without additional model training, architectural modifications, or computationally expensive geometric verification. Comprehensive evaluation across nine state-of-the-art VPR methods and six benchmark datasets confirms that our metrics excel at discriminating between correct and incorrect VPR matches, and consistently outperform existing approaches while maintaining negligible computational overhead, making it deployable for real-time robotic applications across varied environmental conditions with improved precision-recall performance.**

***Index Terms*—Visual Place Recognition, Uncertainty Estimation, Localization**

I. Introduction

Visual Place Recognition (VPR) serves as a fundamental component in robotic localization and autonomous navigation systems [1], enabling robots to determine their position by matching current observations with a database of georeferenced images [2]. Despite significant advances in descriptor-based methods [3], real-world VPR deployment in failure-critical applications like autonomous vehicles, remains challenging due to inadequate uncertainty quantification. Mission-critical systems such as loop closure detection in SLAM pipelines [4] require robust confidence estimates to prevent catastrophic failures when VPR systems encounter perceptual aliasing, seasonal variations, and repetitive structures across different environmental conditions and camera viewpoints.

[1]E. Miller, M. Burhan Hafez, S. D. Ramchurn and S. Ehsan are with the School of Electronics and Computer Science, University of Southampton, United Kingdom (email: em3g20@soton.ac.uk; burhan.hafez@soton.ac.uk; sdr1@soton.ac.uk, s.ehsan@soton.ac.uk)

[2]M. Milford is with the School of Electronics and Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia (email: michael.milford@qut.edu.au)

[3]S. Ehsan is also with the School of Computer Science and Electronic Engineering, University of Essex, United Kingdom (email: sehsan@essex.ac.uk)

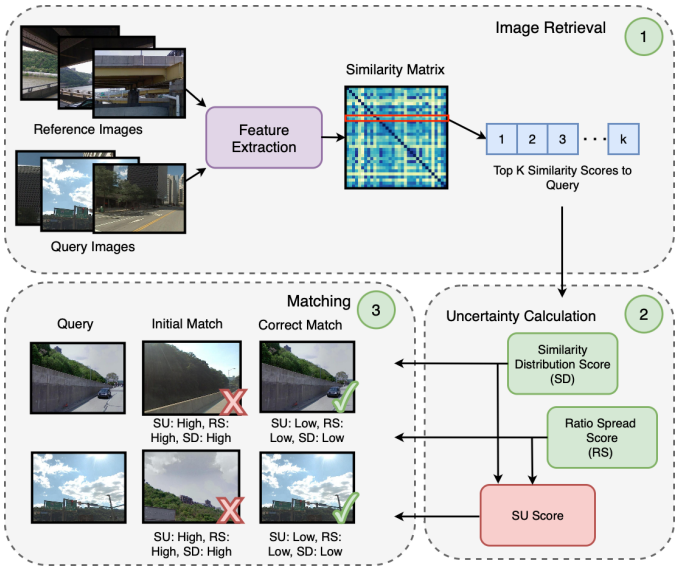

Fig. 1. Overview of our training-free uncertainty estimation metrics integrated into the VPR pipeline, operating directly on similarity scores to provide reliable confidence estimates without requiring model modifications. Shown in Stage 3 in the diagram, examples of the initial match being wrong and having high uncertainty and the following image, which is correct, is the least uncertain image from the corresponding top-k on the Pittsburgh250k dataset using Eigenplaces.

Current uncertainty estimation methods face practical limitations that prevent their use in failure-intolerant applications. Retrieval-based approaches such as L2 distance [5] and Perceptual Aliasing Score [6] show inconsistent performance across datasets and environments, with metric selection depending on the deployment context. Deep Uncertainty Estimation methods like Bayesian Triplet Loss [7] and STUN [8] require costly retraining for each new VPR architecture [4], while Geometric Verification methods using SIFT [9], DELF [10], and SuperPoint [11] are computationally prohibitive for real-time operation. Selecting the most reliable metric typically requires validation data that may be unavailable in unseen environments, posing risks in safety-critical systems where incorrect high-confidence matches can lead to catastrophic failures [12].

This work introduces training-free uncertainty estimation metrics that operate directly on VPR similarity scores without requiring retraining or architectural modifications. These metrics are derived from distributional analysis of top-k matches: Similarity Distribution (SD) measures match distinctiveness by quantifying how clearly the top candidate separates from competing matches through the gap between highest and median similarity scores, while Ratio Spread (RS) evaluates competitive ambiguity by analyzing score clustering among top candidates. We further develop another metric, Statistical

Uncertainty (SU) score, which is a combination of SD and RS, that eliminates the need for validation data to determine which individual metric performs best for specific VPR methods or datasets. These metrics maintain robust performance across diverse deployment scenarios while preserving the computational efficiency required for real-time operation.

Our main contributions are:

- Three training-free uncertainty estimation metrics applicable to any VPR method: Similarity Distribution, Ratio Spread, and their combination (SU score) provides a unified metric that generalizes across VPR architectures and datasets without requiring validation data.
- We conduct a comprehensive two-stage evaluation encompassing nine VPR methods across six benchmark datasets. In the first stage, we assess the predictive capability of the proposed uncertainty metrics using AUC-PR, demonstrating their superior ability to identify matching failures. In the second stage, we evaluate their practical impact on end-to-end VPR performance by showing they measurably improve end VPR system performance, boosting precision at any given recall level when used as a rejection mechanism.

The rest of the paper is organized as follows. Section II reviews related work in VPR and uncertainty estimation. Section III introduces three uncertainty estimation metrics. Section IV outlines experimental setup. Results and analysis are given in Section V. Ablation studies are provided in Section VI. Finally, Section VII draws conclusions.

II. Related Work

A. Visual Place Recognition

Visual Place Recognition (VPR) has evolved from handcrafted local features like SIFT [9] and probabilistic frameworks such as FAB-MAP [13] to deep learning approaches that achieve superior robustness through automatically learned hierarchical representations [3]. Modern VPR architectures employ feature aggregation layers that compress spatial feature maps into compact global descriptors. NetVLAD [14] introduced differentiable aggregation using learnable cluster centers, while subsequent methods like MixVPR [15] incorporate multi-scale feature mixing and TransVPR [16] uses Transformer-based self-attention mechanisms. These architectures are typically trained using Siamese or Triplet networks with metric learning objectives [17], [18], enabled by large-scale datasets like GSV-Cities [19].

B. Uncertainty Estimation in Visual Place Recognition

Robust uncertainty estimation proves crucial for safety-critical applications, where high-confidence incorrect matches from VPR systems can lead to catastrophic failures such as corrupted maps in Simultaneous Localization and Mapping (SLAM) pipelines [1], [20]. Several works have demonstrated that rejecting predictions based on confidence estimates can significantly improve accuracy and prevent such failures [21], [22], motivating the development of specialized uncertainty quantification methods for VPR systems. These VPR-specific methods draw inspiration from the broader deep learning community, where uncertainty is typically categorized as aleatoric (inherent data noise) and epistemic (model ignorance due to limited training data) [23]. General techniques to capture this include Monte Carlo Dropout [24], which approximates Bayesian inference by running multiple forward passes with active dropout layers, and Deep Ensembles [25], which train multiple models to measure prediction variance. Other approaches modify the model's output to directly predict a probability distribution instead of a point estimate [26]. However, these methods often impose significant computational overhead or require architectural modifications and extensive retraining, making them impractical for seamless integration with pre-existing, real-time VPR pipelines [27]. Our work bypasses these limitations by operating directly on the output similarity scores of any VPR model.

VPR-specific uncertainty estimation methods address these challenges through four main categories. Retrieval-based Uncertainty Estimation (RUE) methods estimate uncertainty directly from similarity scores, such as L2 distance or Perceptual Aliasing (PA) score, which is computed as the ratio between first and second nearest neighbor distances [5], [6], [22]. While computationally efficient and training-free, RUE methods suffer from overconfidence in perceptual aliasing scenarios [22].

Data-driven Uncertainty Estimation (DUE) methods learn to infer uncertainty from image content, typically modeling aleatoric uncertainty. Approaches like Bayesian Triplet Loss [7] and self-supervised methods such as STUN [8] exemplify this category but require specialized training and exhibit sensitivity to domain shifts across different environments.

Geometric Verification (GV) approaches assess spatial consistency between query and database images through local feature matching [9]. Methods employing local features like DELF [10] or SuperPoint [11] with RANSAC-based verification achieve high robustness but incur computational costs prohibitive for real-time applications.

Spatial Uncertainty Estimation (SUE) methods [22], [28] infer uncertainty from the spatial distribution of retrieved poses. Though effective in dense mapping scenarios, SUE performance depends on database quality and fails to detect visual ambiguity where geographically distinct locations appear similar. Additionally, SUE requires spatial pose information that may be unavailable at runtime.

Our metrics address these limitations by providing training-free uncertainty estimation that operates directly on similarity scores without model modifications, additional training, or external spatial information, while maintaining computational efficiency for real-time deployment.

Beyond VPR, similar principles for uncertainty estimation have been explored in object recognition [29], semantic segmentation [30], and aerial navigation [31], where predictive uncertainty guides decision confidence and safety assessment. Our metrics extend these ideas to the retrieval-based VPR setting, emphasizing computational efficiency and training-free deployment.

III. Proposed Uncertainty Estimation Metrics

Uncertainty in VPR arises from two primary sources, each evident in the distribution of similarity scores. The first is perceptual aliasing, where visually similar but geographically distinct locations lead the system to assign high similarity scores to incorrect matches, producing tightly clustered high scores among the top candidates. The second is weak distinctiveness, which occurs when even correct matches achieve similarity scores only slightly higher than competitors due to variations in illumination, weather, or viewpoint.

We capture these distributional patterns by analyzing two key factors: the local competition among top-ranked candidates and the global separation between the best match and the remaining ones. This allows for training-free uncertainty estimation applicable to any VPR method.

Our uncertainty metrics analyze the distributional properties of top-k similarity scores from any VPR method, as illustrated in Figure 1. We introduce three uncertainty metrics: two individual metrics capturing complementary aspects of uncertainty, and a unified metric that integrates both to provide robust performance without requiring validation data for metric selection.

A. Ratio Spread (RS)

It measures how strongly alternative candidates compete with the top-ranked match by evaluating how closely their similarity scores approach the highest score. It is calculated as the average ratio of each subsequent candidate's score (from rank 2 to k) to the top score, reflecting the degree of competition among the most similar matches.

$$RS = \frac{1}{k-1} \sum_{i=2}^{k} \frac{s_i}{s_1}$$

Where s_i is the similarity score of the $i-th$ ranked place in descending order.

Higher RS values indicate that several candidates achieve similarity scores close to the top match, reflecting greater uncertainty in the retrieval and a higher likelihood of perceptual aliasing. In contrast, lower RS values suggest higher confidence, as a clear margin separates the best match from its competitors. As illustrated in Figure 2, RS aggregates information from all top-k candidates rather than relying solely on the top two, making it more robust to outliers and more effective at capturing cases where multiple ambiguous matches exist.

B. Similarity Distribution (SD)

It measures the overall dispersion of similarity scores across the top-k candidates by computing the ratio of the median score to the highest score. This metric captures the global shape of the score distribution: lower SD values indicate large separations between the top match and remaining candidates, suggesting high confidence in the retrieval, while values approaching 1.0 reflect uniform score distributions where multiple candidates are equally competitive, indicating matching ambiguity. This is demonstrated in Figure 2. Unlike

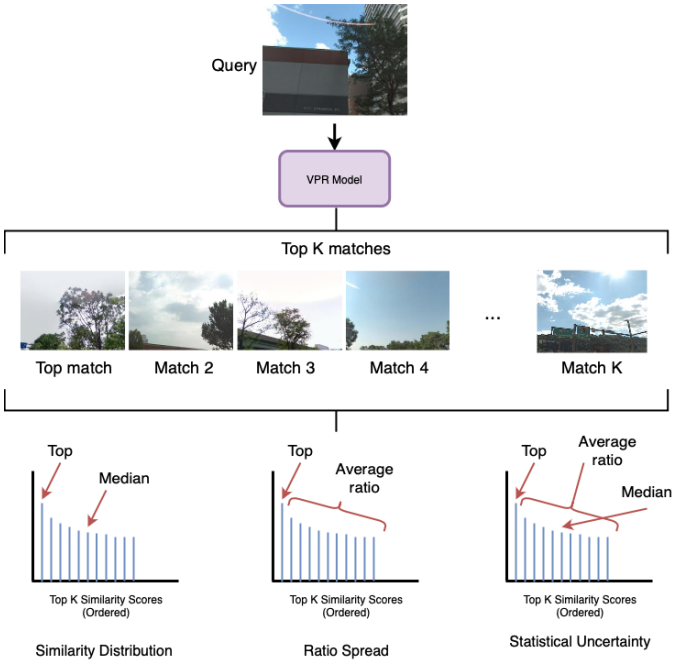

Fig. 2. Visual illustration of uncertainty metrics operating on similarity score distributions. RS captures competitive ambiguity through score clustering among top-k candidates, while SD measures match distinctiveness using the distance between maximum and median scores. SU combines both metrics to provide unified uncertainty estimation.

methods that evaluate query-reference pairs in isolation, SD analyzes the collective distribution of top candidates, enabling more effective detection of perceptual aliasing scenarios where several visually similar but potentially incorrect matches compete with the true positive.

$$SD = \frac{s_{median}}{s_{best}}$$

C. Statistical Uncertainty (SU)

It combines two complementary distributional properties of the similarity score distribution. These metrics are combined to form the final uncertainty estimate:

$$SU = \alpha \times RS + (1 - \alpha) \times SD \quad (1)$$

where α controls the relative contribution of each component. This formula captures both local competitive dynamics among top candidates and global distributional characteristics of the similarity space. Figure 3 illustrates these uncertainty metrics applied to real retrieval examples, demonstrating how low SU scores correspond to correct matches with clear separation, while high SU scores indicate matching ambiguity, as demonstrated in Figure 2. We validate the choice of weighting parameters through comprehensive ablation studies presented in Section VI.

IV. Experimental Setup

We evaluate our uncertainty metrics across diverse VPR methods and environmental conditions using AUC-PR on benchmark datasets. All uncertainty estimation techniques are applied post-hoc to pre-computed descriptors, preserving

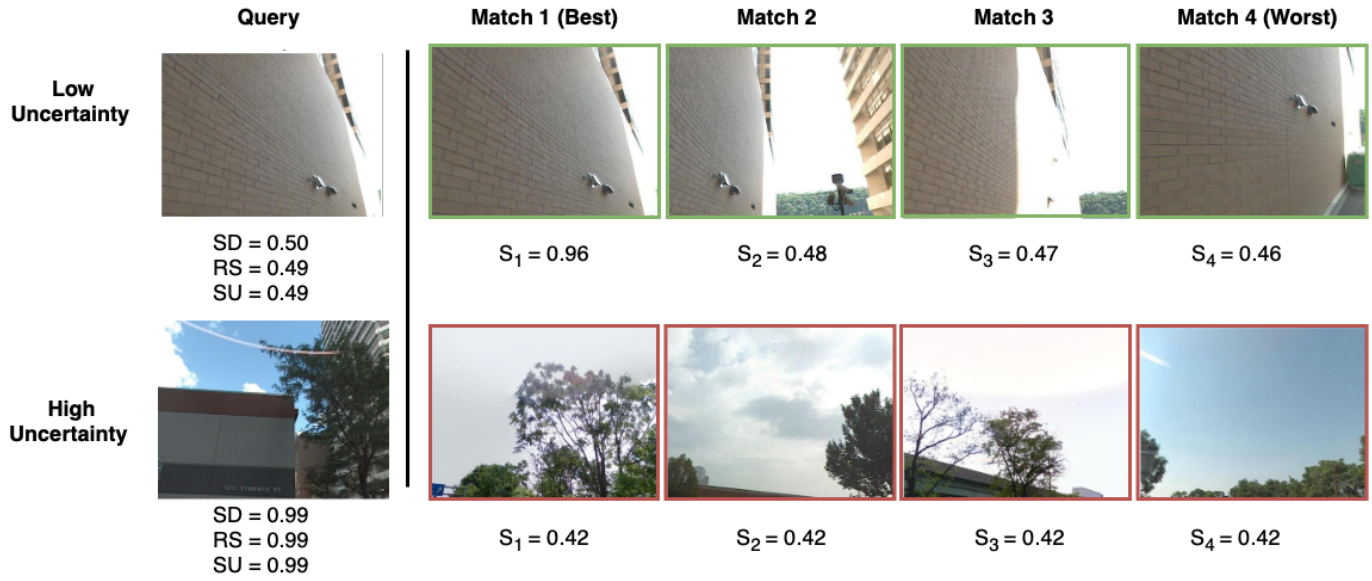

Fig. 3. Visual examples of the top 4 matches for two queries using APGeM on the Pittsburgh250k dataset. Each retrieved image displays its similarity score, with the RS, SD and SU values shown below each query image. The top row demonstrates a correct retrieval with low uncertainty, while the bottom row shows incorrect matches with high uncertainty.

the original performance of each VPR system at its native resolution.

A. Datasets and Methods

We evaluate nine VPR methods: Eigenplaces [32], Cosplace [33], ConvAP [34], MixVPR [15], AP GeM [35], SALAD [36], AnyLoc [37], CricaVPR [38], and NetVLAD [14] across six benchmark datasets: Pittsburgh 250k [39], St. Lucia [40], Nordland [41], Amstertime [42], Eynsham [43], and Tokyo24/7 [44]. These datasets exhibit environmental diversity (seasonal changes, illumination variations, weather) and geometric challenges (viewpoint changes, perceptual aliasing, repetitive structures). Dataset details and ground-truths are provided in [45].

B. Evaluation Metrics and Baselines

To ensure a fair and interpretable evaluation, we assess the uncertainty estimators independently from the retrieval backbone, thereby isolating their ability to rank predictions by confidence. We evaluate two separate but complementary aspects:

Predictor Quality: How well does each uncertainty metric identify incorrect matches? This is measured using the Area Under the Precision-Recall Curve (AUC-PR). For this evaluation, we treat the problem as a binary classification task: ”correct match” vs. ”incorrect match,” where the uncertainty score acts as the classifier’s output. A high AUC-PR indicates the metric successfully assigns low uncertainty to correct matches and high uncertainty to incorrect ones. This directly evaluates the quality of the confidence estimate itself [26], [27], [46].

End-to-End VPR Impact: Does using the uncertainty metric to reject predictions improve the VPR system’s real-world performance? We measure this by generating precision-recall curves for the overall VPR system. By setting a threshold on the uncertainty score and rejecting all predictions deemed too uncertain, we can analyze the trade-off between precision and recall. A superior uncertainty metric will allow the VPR system to achieve higher precision at any given level of recall. Notably, this evaluation pertains exclusively to the uncertainty estimation module (Stage 3 in Figure 1), rather than the feature extraction or retrieval components of the system.

We compare against representative methods from each category: 1) Retrieval-based: L2 Distance [5] and Perceptual Aliasing (PA) [6]; 2) Spatial: Spatial Uncertainty Estimation (SUE) [22]; 3) Learning-based: Bayesian Triplet Loss (BTL) [7], implemented following the original paper’s pseudo-code and trained on Pittsburgh250k; 4) Geometric Verification: For SIFT [9] and DELF [10], features are extracted and then matched using a nearest-neighbor search with a traditional Lowe’s ratio test for matching, while for SuperPoint [11], its extracted features are matched using the SuperGlue deep-learning-based matcher [47]. Following the matching stage, the RANSAC algorithm is employed in all three pipelines to identify the set of inlier matches. The final uncertainty value is then quantified as the negative of the total inlier count ($U_{GV} = -c_{inliers}$).

C. Configuration

Our metrics analyze similarity scores from the top 10 ($k = 10$) retrieved candidates using cosine distance via Faiss [48]. The SU score combines RS and SD with equal weighting ($\alpha = 0.5$). Experiments run on NVIDIA A100 GPUs with batch size 1 to simulate real-time robotic deployment. Results are averaged across three independent runs using standard dataset splits.

V. Results and Discussion

This section presents the comprehensive evaluation of our uncertainty metrics. We analyze their overall performance,

computational efficiency, and ability to generalize across different VPR systems.

A. Overall Performance Analysis

Table I evaluates predictor quality—how well each uncertainty metric distinguishes correct from incorrect matches, as measured by AUC-PR. A score closer to 1.0 indicates a more effective predictor, this measures the quality of the confidence estimates themselves.

Our three metrics consistently outperform existing retrieval-based approaches (L2, PA), learning-based methods (BTL) and spatial methods (SUE), achieving higher AUC-PR scores across nearly all VPR architectures. In particular, SU achieves the best overall results with a mean AUC-PR of 0.92, demonstrating robust uncertainty estimation across diverse visual environments and architectures. This strong intrinsic performance in identifying bad matches is the foundation for improving downstream VPR reliability.

B. Cross-Method Generalization

Figure 4 illustrates the performance gains of our metrics across different VPR architectures and datasets. Each heatmap cell shows the difference in AUC-PR relative to baseline uncertainty methods.

Figure 4 visualizes the performance gains (in AUC-PR) of our metrics relative to baseline uncertainty methods across different VPR architectures and datasets. Our metrics consistently demonstrate positive gains, particularly on challenging datasets like Amstertime (viewpoint changes) and Nordland (seasonal variation). This confirms the robustness of our statistical approach across architectures, from CNN-based (NetVLAD) to modern Transformer-based models (Eigenplaces, SALAD). While geometric verification with SuperPoint achieves high accuracy, its computational cost is prohibitive for real-time use. Our metrics provide a superior balance of accuracy and speed, establishing a practical solution for VPR uncertainty estimation.

Geometric verification techniques achieve strong accuracy but are computationally expensive, making them unsuitable for real-time applications. Retrieval-based baselines (L2, PA) are efficient but less reliable across architectures. Learning-based methods (BTL, SUE) either under perform or require retraining. In contrast, our metrics achieve both superior accuracy and real-time inference speed, establishing a more practical solution for VPR uncertainty estimation.

C. Computational Efficiency

The proposed metrics are exceptionally efficient, requiring less than 0.03 ms per query (see Table I). This is orders of magnitude faster than geometric verification methods like SuperPoint (84 ms) and makes our approach suitable for real-time robotic applications where computational resources are limited. This efficiency is achieved by performing lightweight statistical analysis directly on pre-computed similarity scores, avoiding any feature extraction or matching overhead.

D. Training-Free Advantage and Practical Deployment

The training-free nature of our metrics ensures immediate applicability to any VPR system without fine-tuning, simplifying integration across different camera setups and environments. Figure 5 provides qualitative examples from the Pittsburgh250k dataset, illustrating how our metrics correctly identify high uncertainty in ambiguous scenes (e.g., repetitive building facades) and low uncertainty in distinctive, correct matches.

VI. Ablation Study

This ablation study examines the effect of different α weighting combinations in the SU equation and evaluates the robustness of the chosen top-k value.

A. SU Alpha Weighting Study

We evaluated the SU score's performance by varying $\alpha \in [0.0, 1.0]$ for the APGeM and EigenPlaces methods, with results shown in Figure 6. While performance is stable for most datasets, challenging ones like Amstertime and Nordland show preferences for either SD or RS, depending on the VPR method used. This highlights a practical challenge: the optimal individual metric can change with the deployment context. Setting $\alpha = 0.5$ provides a robust default that balances both metrics, delivering consistently strong performance without requiring validation data for tuning. It prevents significant performance drops and generalizes well across diverse conditions.

B. K-Value Robustness Analysis

This ablation study examines the influence of the top-k candidates used in our uncertainty metrics, evaluated on the Pittsburgh250k dataset across multiple VPR methods, with findings that generalize consistently to other datasets.

We analyzed the influence of the number of top candidates (k) on the Pittsburgh250k dataset. As shown in Figure 7, both SD and RS show stable performance for $k \geq 5$, with performance stabilizing around $k = 10$ and plateauing beyond that. This robustness allows for smaller k values in resource-constrained settings without compromising estimation quality and avoids the need for method-specific tuning.

VII. Conclusions and Future Work

This paper introduced three training-free uncertainty metrics for VPR: Ratio Spread (RS), Similarity Distribution (SD), and their unified combination, Statistical Uncertainty (SU). Our key contribution is a method-agnostic approach that requires no training, architectural modifications, or validation data for metric selection, while adding negligible computational overhead ($<$0.03 ms per query).

Our comprehensive two-stage evaluation demonstrated that SU not only excels at identifying incorrect matches (0.92 mean AUC-PR) but also translates this capability into tangible improvements for the end VPR system, enhancing precision in applications like loop-closure detection. While these metrics are highly effective, their reliance on similarity scores means performance may degrade in extremely repetitive scenes

Uncertainty Metric	Eigenplaces	Cosplace	ConvAP	MixVPR	APGeM	SALAD	AnyLoc	Crica	NetVLAD	Mean	Time
Baseline Model (BM)	0.91	0.9	0.87	0.91	0.62	0.94	0.81	0.87	0.71	0.84	–
BM + (DUE) BTL	0.24	0.25	0.17	0.17	0.13	0.33	0.26	0.13	0.10	0.20	0.52
BM + (RUE) L2	0.89	0.90	0.85	0.90	0.62	0.94	0.82	0.87	0.71	0.83	0.01
BM + (RUE) PA Score	0.93	0.93	0.89	0.92	0.74	**0.97**	0.88	0.93	0.78	0.88	0.01
BM + (RUE) RS	0.94	**0.95**	**0.92**	**0.94**	0.80	0.96	0.90	0.94	0.82	0.91	0.01
BM + (RUE) SD	**0.95**	0.94	**0.92**	**0.94**	**0.82**	0.96	0.90	0.93	**0.83**	0.91	0.02
BM + (RUE) SU	**0.95**	**0.95**	**0.92**	**0.94**	**0.82**	**0.97**	**0.91**	**0.96**	**0.83**	**0.92**	0.03
BM + SUE	0.68	0.68	0.71	0.81	0.48	0.88	0.73	0.88	0.53	0.71	0.05
BM + (GV) SIFT	0.72	0.72	0.67	0.66	0.38	0.83	0.66	0.72	0.42	0.64	139.86
BM + (GV) DELF	0.79	0.78	0.75	0.85	0.72	0.93	0.81	0.87	0.74	0.80	78.54
BM + (GV) SuperPoint	**0.95**	**0.95**	**0.98**	**0.95**	**0.93**	0.96	**0.94**	0.95	**0.94**	**0.95**	84.11

TABLE I

TABLE 1. A COMPARISON OF UNCERTAINTY METRICS FOR VISUAL PLACE RECOGNITION (VPR) METHODS, EVALUATED BY THEIR AUC-PR SCORES AVERAGED ACROSS ALL DATASETS. HIGHER SCORES INDICATE BETTER PERFORMANCE, WITH THE BEST RESULTS IN BOLD. THE TABLE INCLUDES A BASELINE (TOP ROW), COMPUTATIONALLY INTENSIVE GEOMETRIC VERIFICATION (GV) METHODS FOR COMPARISON (BOTTOM ROWS), AND THE AVERAGE COMPUTATION TIME TO CALCULATE UNCERTAINTY PER QUERY (LAST COLUMN).

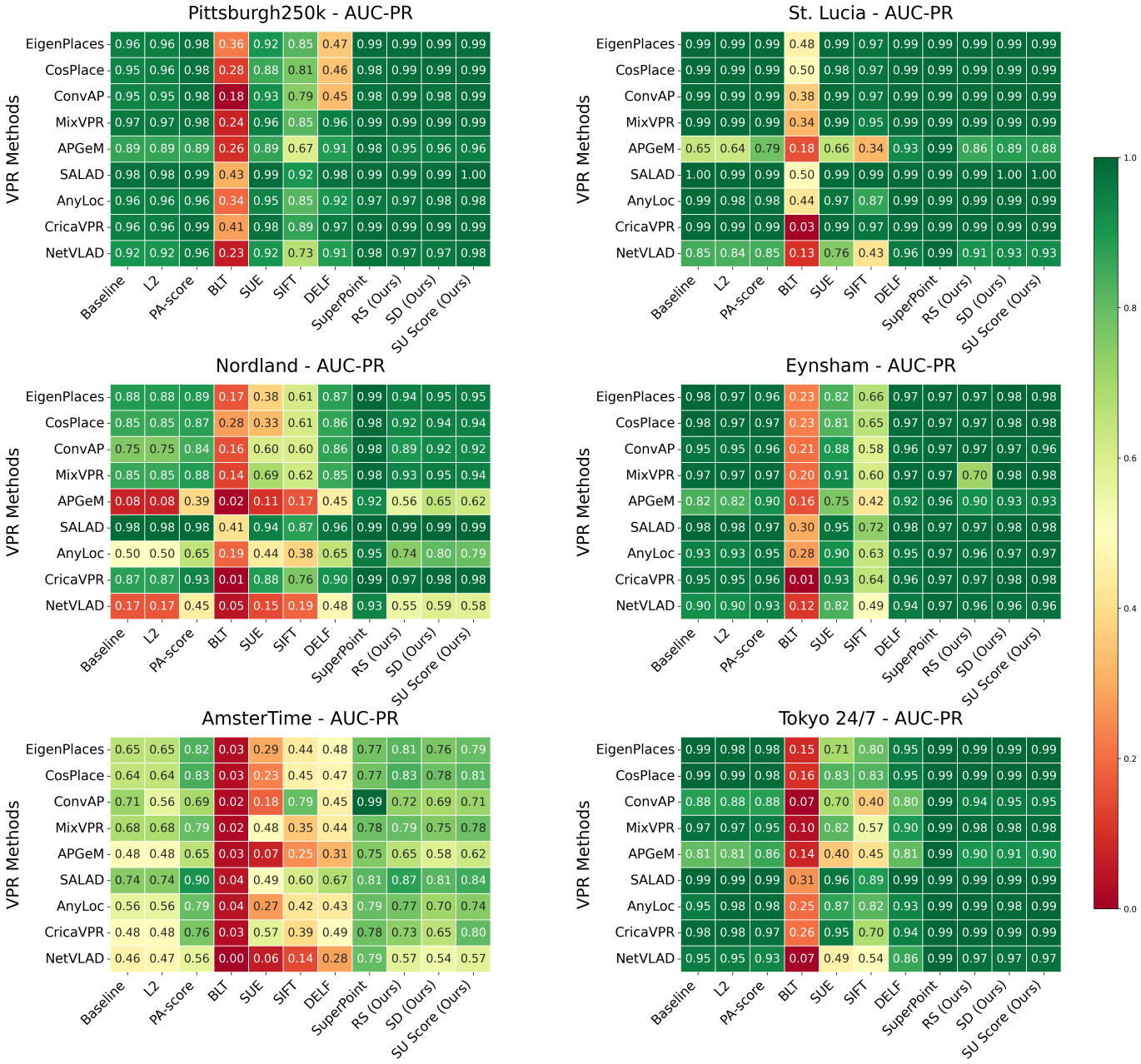

Fig. 4. Performance difference (AUC-PR) for each VPR and uncertainty metric combination across datasets.

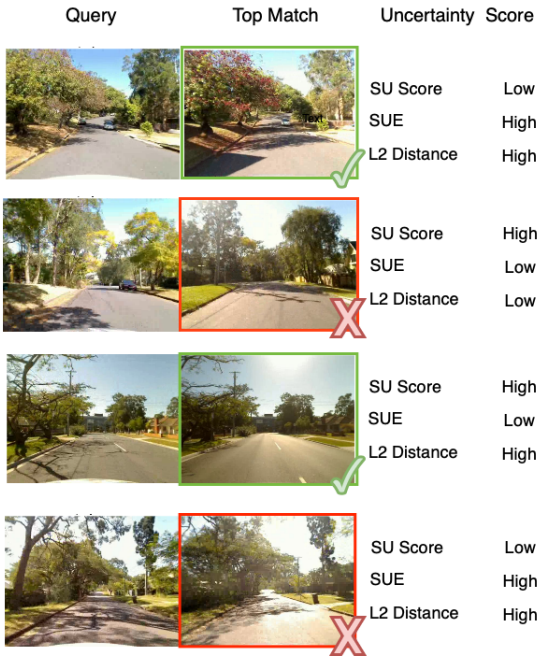

Fig. 5. Qualitative comparison on the Pittsburgh250k dataset. Top: correct uncertainty identification; bottom: failure cases. Query (left) and top-ranked retrieval (right) pairs shown.

where geometric cues are necessary. Future work will explore lightweight integration of geometric or semantic information to further improve robustness and investigate adaptive weighting of the RS and SD components based on score distributions.

REFERENCES

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[2] 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, 2016.

[3] S. Schubert, P. Neubert, S. Garg, M. Milford, and T. Fischer, "Visual place recognition: A tutorial [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 31, no. 3, p. 139–153, Sep. 2024. [Online]. Available: http://dx.doi.org/10.1109/MRA.2023.3310859

[4] S. Garg, T. Fischer, and M. Milford, "Where is your place, visual place recognition?" in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, ser. IJCAI-2021. International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, p. 4416–4425. [Online]. Available: http://dx.doi.org/10.24963/ijcai.2021/603

[5] N. Piasco, D. Sidibé, C. Demonceaux, and V. Gouet-Brunet, "A survey on Visual-Based Localization: On the benefit of heterogeneous data," *Pattern Recognition*, vol. 74, pp. 90 – 109, Feb. 2018. [Online]. Available: https://hal.science/hal-01744680

[6] S. Hausler, T. Fischer, and M. Milford, "Unsupervised complementary-aware multi-process fusion for visual place recognition," 2021. [Online]. Available: https://arxiv.org/abs/2112.04701

[7] F. Warburg, M. Jørgensen, J. Civera, and S. Hauberg, "Bayesian triplet loss: Uncertainty quantification in image retrieval," 2021. [Online]. Available: https://arxiv.org/abs/2011.12663

[8] K. Cai, C. X. Lu, and X. Huang, "Stun: Self-teaching uncertainty estimation for place recognition," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 6614–6621.

[9] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–, 11 2004.

[10] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," 2018. [Online]. Available: https://arxiv.org/abs/1612.06321

[11] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 337–33 712.

[12] 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," 2025. [Online]. Available: https://arxiv.org/abs/2405.04812

[13] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Rob. Res.*, vol. 27, no. 6, pp. 647–665, Jun. 2008.

[14] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," 2016. [Online]. Available: https://arxiv.org/abs/1511.07247

[15] A. Ali-bey, B. Chaib-draa, and P. Giguère, "Mixvpr: Feature mixing for visual place recognition," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2023, pp. 2998–3007.

[16] R. Wang, Y. Shen, W. Zuo, S. Zhou, and N. Zheng, "Transvpr: Transformer-based place recognition with multi-level attention aggregation," 2022. [Online]. Available: https://arxiv.org/abs/2201.02001

[17] M. Leyva-Vallina, N. Strisciuglio, and N. Petkov, "Generalized contrastive optimization of siamese networks for place recognition," 2023. [Online]. Available: https://arxiv.org/abs/2103.06638

[18] J. Revaud, J. Almazan, R. Rezende, and C. D. Souza, "Learning with average precision: Training image retrieval with a listwise loss," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019.

[19] A. Ali-bey, B. Chaib-draa, and P. Giguère, "Gsv-cities: Toward appropriate supervised visual place recognition," *Neurocomputing*, vol. 513, p. 194–203, Nov. 2022. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2022.09.127

[20] H.-M. Won, J. Lee, and J. Oh, "Localization meets uncertainty: Uncertainty-aware multi-modal localization," 2025. [Online]. Available: https://arxiv.org/abs/2504.07677

[21] S. Agarwal, J. Yuan, P. Newman, D. De Martini, and M. Gadd, "Bayesian radar cosplace: Directly estimating location uncertainty in radar place recognition," *IET Radar, Sonar & Navigation*, vol. 19, 03 2025.

[22] M. Zaffar, L. Nan, and J. F. P. Kooij, "On the estimation of image-matching uncertainty in visual place recognition," 2024. [Online]. Available: https://arxiv.org/abs/2404.00546

[23] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 5580–5590.

[24] M. Hasan, A. Khosravi, I. Hossain, A. Rahman, and S. Nahavandi, "Controlled dropout for uncertainty estimation," 2022. [Online]. Available: https://arxiv.org/abs/2205.03109

[25] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," 2017. [Online]. Available: https://arxiv.org/abs/1612.01474

[26] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," 2018. [Online]. Available: https://arxiv.org/abs/1802.10501

[27] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," 2018. [Online]. Available: https://arxiv.org/abs/1610.02136

[28] N. Pion, M. Humenberger, G. Csurka, Y. Cabon, and T. Sattler, "Benchmarking image retrieval for visual localization," 2020. [Online]. Available: https://arxiv.org/abs/2011.11946

[29] F. Neha, D. Bhati, D. K. Shukla, and M. Amiruzzaman, "From classical techniques to convolution-based models: A review of object detection algorithms," 2024. [Online]. Available: https://arxiv.org/abs/2412.05252

[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2014. [Online]. Available: https://arxiv.org/abs/1311.2524

[31] J. Lee, T. Miyanishi, S. Kurita, K. Sakamoto, D. Azuma, Y. Matsuo, and N. Inoue, "Citynav: A large-scale dataset for real-world aerial navigation," 2025. [Online]. Available: https://arxiv.org/abs/2406.14240

[32] G. Berton, G. Trivigno, B. Caputo, and C. Masone, "Eigenplaces: Training viewpoint robust models for visual place recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 11 080–11 090.

[33] G. Berton, C. Masone, and B. Caputo, "Rethinking visual geo-localization for large-scale applications," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 4878–4888.

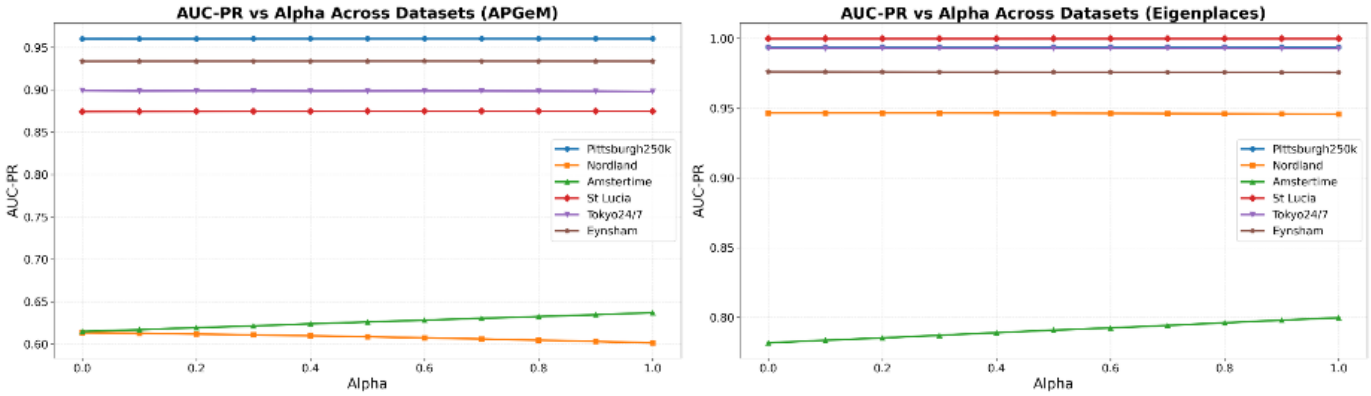

Fig. 6. AUC-PR performance as a function of α weighting parameter (Equation 1) across multiple datasets. (Left) Results using APGeM as the VPR method. (Right) Results using EigenPlaces as the VPR method. Each line represents a different dataset, showing how the combined uncertainty metric SU performs across varying contributions of RS and SD components.

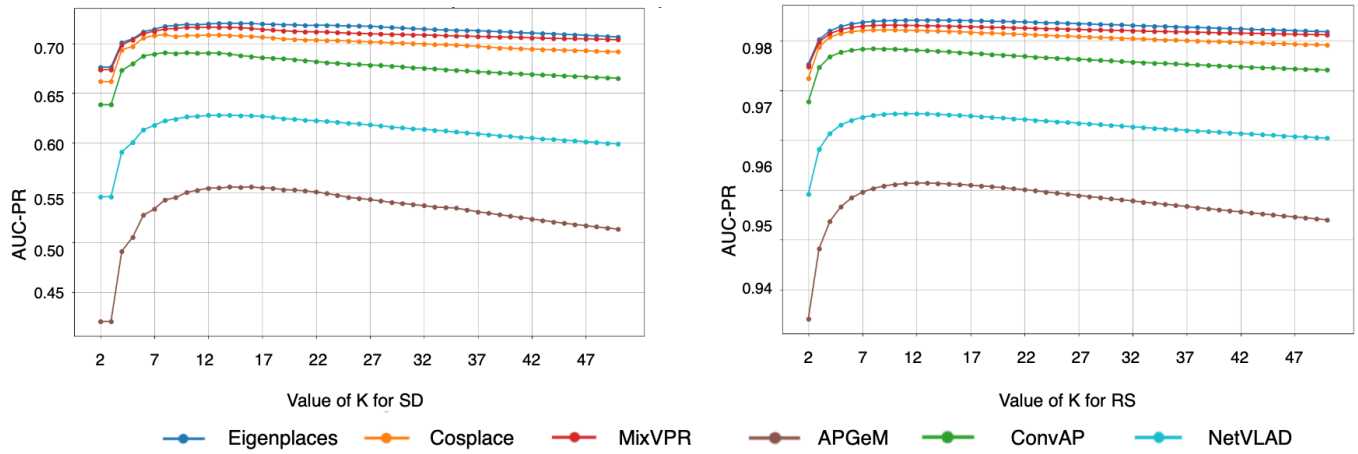

Fig. 7. Comparison of SD and RS uncertainty metrics across k values for VPR methods on Pittsburgh250k. AUC-PR results show both SD and RS provide more consistent performance across k values.

[34] A. Ali-bey, B. Chaib-draa, and P. Giguère, "Gsv-cities: Toward appropriate supervised visual place recognition," *Neurocomput.*, vol. 513, no. C, p. 194–203, Nov. 2022. [Online]. Available: https://doi.org/10.1016/j.neucom.2022.09.127

[35] 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.

[36] S. Izquierdo and J. Civera, "Optimal transport aggregation for visual place recognition," 2024. [Online]. Available: https://arxiv.org/abs/2311.15937

[37] N. Keetha, A. Mishra, J. Karhade, K. M. Jatavallabhula, S. Scherer, M. Krishna, and S. Garg, "Anyloc: Towards universal visual place recognition," 2023. [Online]. Available: https://arxiv.org/abs/2308.00688

[38] F. Lu, X. Lan, L. Zhang, D. Jiang, Y. Wang, and C. Yuan, "Cricavpr: Cross-image correlation-aware representation learning for visual place recognition," 2024. [Online]. Available: https://arxiv.org/abs/2402.19231

[39] A. Torii, J. Sivic, M. Okutomi, and T. Pajdla, "Visual place recognition with repetitive structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[40] M. J. Milford and G. F. Wyeth, "Mapping a suburb with a single camera using a biologically inspired slam system," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1038–1053, 2008.

[41] S. Skrede, "Nordland dataset," 2013, accessed: October 16, 2025. [Online]. Available: https://bit.ly/2QVBOym

[42] B. Yildiz, S. Khademi, R. M. Siebes, and J. van Gemert, "Amstertime: A visual place recognition benchmark dataset for severe domain shift," 2022. [Online]. Available: https://arxiv.org/abs/2203.16291

[43] M. Cummins and P. Newman, "Highly scalable appearance-only slam - fab-map 2.0," 06 2009.

[44] A. Torii, R. Arandjelović, J. Sivic, M. Okutomi, and T. Pajdla, "24/7 place recognition by view synthesis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 257–271, 2018.

[45] G. Berton, C. Masone, V. Paolicelli, and B. Caputo, "Viewpoint invariant dense matching for visual geolocalization," 09 2021.

[46] M. Zaffar, S. Garg, M. Milford, J. Kooij, D. Flynn, K. McDonald-Maier, and S. Ehsan, "Vpr-bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change," *International Journal of Computer Vision*, vol. 129, no. 7, p. 2136–2174, May 2021. [Online]. Available: http://dx.doi.org/10.1007/s11263-021-01469-5

[47] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *CVPR*, 2020. [Online]. Available: https://arxiv.org/abs/1911.11763

[48] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou, "The faiss library," 2024.