5.2

	R@1	R@5	R@10	R@20	Time per Query
Cosplace					
Tokyo	65.1	79.7	86.0	89.5	
SuperPoint+LightGlue	82.9	86.7	87.9	89.5	3.55
LoFTR	84.8	87.9	88.6	89.5	3.73
SuperGlue	82.5	86.3	88.6	89.5	1.44
Sf	63.1	74.8	78.6	81.4	
SuperPoint+LightGlue	77.7	80.4	80.8	81.4	3.52
LoFTR	77.4	79.7	80.6	81.4	3.74
SuperGlue	76.7	79.4	80.5	81.4	1.41
Svox_night	33.3	51.5	59.1	67.7	
SuperPoint+LightGlue	60.6	65.1	66.3	67.6	3.52

LoFTR	61.4	65.4	66.2	67.6	3.69
SuperGlue	59.4	64.5	65.9	67.6	1.46
Svox_day	62.3	78.5	84.5	88.8	
SuperPoint+LightGlue	83.8	86.9	87.9	88.8	3.52
LoFTR	85.1	87.1	87.9	88.8	3.70
SuperGlue	81.3	85.9	87.4	88.8	1.47

	R@1	R@5	R@10	R@20	Time per Query
NetVLAD					
Tokyo	49.8	62.5	70.5	78.7	
SuperPoint+LightGlue	68.3	72.1	73.7	78.7	3.45
LoFTR	68.3	72.7	73.7	78.7	3.63
SuperGlue	69.5	72.7	74.9	78.7	1.27
Sf	27.2	43.8	50.3	56.2	

SuperPoint+LightGlue	53.2	55.4	55.8	56.2	3.44
LoFTR	53.6	55.3	55.8	56.2	3.62
SuperGlue	52.4	55.3	55.8	56.2	1.23
Svox_night	8.0	17.4	23.1	29.6	
SuperPoint+LightGlue	25.4	27.1	28.4	29.5	3.45
LoFTR	25.5	28.1	28.6	29.5	3.59
SuperGlue	24.5	26.6	28.1	29.5	1.30
Svox_day	35.4	52.7	58.8	65.8	
SuperPoint+LightGlue	61.7	63.1	64.2	65.8	3.45
LoFTR	61.6	63.6	64.5	65.8	3.60
SuperGlue	60.4	62.9	64.2	65.8	1.33

	R@1	R@5	R@10	R@20	Time per Query
MixVPR					

Tokyo	78.1	89.5	92.4	93.7	
SuperPoint+LightGlue	89.2	92.4	92.7	93.7	3.40
LoFTR	89.8	92.4	93.7	93.7	3.58
SuperGlue	86.7	91.7	93.3	93.7	1.26
Sf	70.2	79.0	81.3	83.9	
SuperPoint+LightGlue	81.4	83.3	83.7	83.9	3.40
LoFTR	80.1	82.9	83.7	83.9	3.58
SuperGlue	80.1	82.7	83.2	83.9	1.23
Svox_night	62.9	79.8	84.1	88.0	
SuperPoint+LightGlue	81.9	86.3	87.1	88.0	3.41
LoFTR	82.5	86.8	87.6	88.0	3.54
SuperGlue	81.9	86.5	87.5	88.0	1.31
Svox_day	85.4	93.0	94.7	95.9	
SuperPoint+LightGlue	91.7	95.0	95.6	95.9	3.41
LoFTR	93.4	95.0	95.3	95.9	3.56

SuperGlue	89.8	94.0	95.2	95.9	1.33

	R@1	R@5	R@10	R@20	Time per Query
MegaLoc					
Tokyo	94.9	97.5	98.4	99.0	
SuperPoint+LightGlue	94.3	98.4	99.0	99.0	3.57
LoFTR	94.9	97.5	98.4	99.0	3.78
SuperGlue	93.7	97.5	98.4	99.0	1.25
Sf	86.7	89.9	90.8	91.3	
SuperPoint+LightGlue	86.8	90.2	90.7	91.3	3.58
LoFTR	85.6	89.3	90.6	91.3	3.78
SuperGlue	85.7	89.4	90.2	91.3	1.22
Svox_night	95.1	98.1	98.8	99.1	
SuperPoint+LightGlue	90.8	97.4	98.8	99.1	3.59

LoFTR	92.3	97.9	98.8	99.1	3.73
SuperGlue	90.0	97.6	98.5	99.1	1.33
Svox_day	96.5	99.4	99.6	99.6	
SuperPoint+LightGlue	95.8	99.3	99.5	99.6	3.58
LoFTR	96.7	99.4	99.5	99.6	3.76
SuperGlue	93.9	98.9	99.6	99.6	1.33

6.2

For each query q, we compute the number of inliers correspondence when we run the matching model between the query image and the top-1 retrieved database image (from the VPR model).

A query is considered **hard** if: inliers_top1 < t

If hard (inliers_top1 < t) → we rerank using the best-inlier candidate among top-K

If easy (inliers_top1 >= t) → we keep retrieval top-1 (no reranking)

Low inliers: The model is uncertain (the query is "hard"). Trigger the re-ranking process (checking Top-20) to find a better match.

High inliers: The model is confident. Trust the Top-1 result and skip the expensive re-ranking of the other 19 candidates.

Hard = low inlier count with the first retrieved match

Thresholding:

Train DataSets:

VPR Model	Image Matching Model	Type	Threshold	R@1	R@5	R@10	R@20	Reranked	Time
SVOX_Night									
MegaLoc		Retrieval		91.17	95.73	97.44	98.43		
	LoFTR	Rerank		87.46	95.58	97.29	98.43		
		Adaptive	5	91.17	95.73	97.44	98.43	0	0
	SuperGlue	Rerank		86.89	95.87	97.72	98.43		
		Adaptive	6.10	92.45	96.15	97.72	98.43	10.1	1.33
NetVLAD		Retrieval		3.13	8.97	12.11	17.81		
	LoFTR	Rerank		13.68	15.24	16.24	17.81		
		Adaptive	406	13.68	15.24	16.24	17.81	99.9	3.725
	SuperGlue	Rerank		13.25	15.10	16.52	17.81		
		Adaptive	15	13.39	14.96	16.38	17.81	97.2	1.292

SVOX_Sun									
MegaLoc		Retrieval		94.52	97.89	89.31	98.60		
	LoFTR	Rerank		93.26	96.49	97.75	98.60		
		Adaptive	2	94.52	97.89	98.31	98.60	0	0
	SuperGlue	Rerank		90.59	96.21	97.19	98.60		
		Adaptive	0	94.52	97.89	98.31	98.60	0	0
NetVLAD		Retrieval		27.11	40.73	46.21	51.40		
	LoFTR	Rerank		48.74	50	50.84	51.40		
		Adaptive	134	48.74	50	50.84	51.40	82.3	3.07
	SuperGlue	Rerank		46.91	49.86	50.70	51.40		
		Adaptive	44.12	46.91	49.86	50.70	51.40	87.5	1.164

Validation DataSets:

VPR Model	Image Matching Model	Type	Threshold	R@1	R@5	R@10	R@20	Reranked	Time
SF									
MegaLoc		Retrieval		97.26	98.92	99.22	99.41		
	LoFTR	Rerank		92.89	97.34	98.59	99.41		
		Adaptive	5	97.26	98.92	99.22	99.41	0	0
	SuperGlue	Rerank		91.33	97.72	98.87	99.41		
		Adaptive	0	97.26	98.92	99.22	99.41	0	0
NetVLAD		Retrieval		53.22	67.55	73.33	78.98		
	LoFTR	Rerank		72.85	75.82	77.36	78.98		
		Adaptive	132	72.98	75.82	77.36	78.98	62.4%	2.328
	SuperGlue	Rerank		70.81	75.42	77.10	78.98		
		Adaptive	19	70.84	75.39	77.09	78.98	64.9%	0.864

Test DataSets:

Thresholds To test:

```
{
    (netvlal, loftr): 132,
    (netvlad, superglue): 19,
    (megaloc, loftr): 5,
    (megaloc, superglue): 6
}
```

VPR Model	Image Matching Model	Type	Threshold	R@1	R@5	R@10	R@20	Reranked	Time
Tokyo									
MegaLoc		Retrieval		94.92	97.46	98.41	99.05		
	LoFTR	Rerank		94.92	97.46	98.73	99.05		
		Adaptive	5	94.92	97.46	98.41	99.05	0	0
	SuperGlue	Rerank		94.92	97.46	98.41	99.05		
		Adaptive	6	95.24	98.10	98.73	99.05	5.4%	0.07
NetVLAD		Retrieval		49.84	62.54	70.48	78.73		
	LoFTR	Rerank		68.57	72.70	73.65	78.73		
		Adaptive	132	68.57	72.70	73.65	78.73	77.8%	2.901

	SuperGlue	Rerank		69.21	72.38	74.60	78.73		
		Adaptive	19	69.21	72.38	74.60	78.73	68.9%	0.96
SF									
MegaLoc		Retrieval		86.70	89.90	90.8-	91.30		
	LoFTR	Rerank		85.60	89.20	90.5	91.30		
		Adaptive	5	86.70	89.90	90.80	91.30	0	0
	SuperGlue	Rerank		85.90	89.4	90.4	91.3		
		Adaptive	6	86.40	89.5	90.6	91.30	7.9%	0.1
NetVLAD		Retrieval		27.20	43.80	50.30	56.20		
	LoFTR	Rerank		53.60	55.30	55.80	56.2		
		Adaptive	132	53.60	55.30	55.8	56.2	83.4%	3.111
	SuperGlue	Rerank		52.60	55.10	56	56.20		
		Adaptive	19	52.60	55.10	56	56.2	82.7%	1.1
SVOX_Night									
MegaLoc		Retrieval		95.14	98.06	98.78	99.15		

	LoFTR	Rerank		92.47	98.06	98.78	99.15		
		Adaptive	5	95.14	98.06	98.78	99.15	0	0
	SuperGlue	Rerank		90.52	97.45	98.42	99.15		
		Adaptive	6	94.65	97.69	98.66	99.15	3.6%	0.04
NetVLAD		Retrieval		8.02	17.38	23.09	29.65		
	LoFTR	Rerank		25.64	27.83	28.55	29.65		
		Adaptive	132	25.64	27.83	28.55	29.65	96.4%	3.594
	SuperGlue	Rerank		24.54	26.97	27.95	29.65		
		Adaptive	19	24.54	26.85	27.83	29.65	93.1%	1.23
SVOX_Sun									
MegaLoc		Retrieval		96.49	99.41	99.65	99.65		
	LoFTR	Rerank		96.72	99.41	99.41	99.65		
		Adaptive	5	96.49	99.41	99.65	99.65	0.1%	0.004
	SuperGlue	Rerank		94.26	98.95	99.65	99.65		
		Adaptive	6	95.90	99.30	99.65	99.65	1.5%	0.02

NetVLAD		Retrieval		35.36	52.69	58.78	65.81		
	LoFTR	Rerank		61.48	63.47	64.52	65.81		
		Adaptive	132	61.48	63.47	64.52	65.81	73.7%	2.477
	SuperGlue	Rerank		60.66	52.69	58.78	65.81		
		Adaptive	19	60.3	63.23	64.40	65.81	67%	0.89

Logistic:

To evaluate the quality of our uncertainty estimation, we compute the area under the Precision–Recall curve (AUPRC) using p_{wrong} as the uncertainty score and the event 'top-1 retrieval is wrong' as the positive class. AUPRC = 1.0 indicates perfect ranking of errors (all wrong queries receive higher uncertainty than correct ones). A random ranking achieves an AUPRC approximately equal to the prevalence of wrong queries in the dataset.

We visualize uncertainty by plotting histograms of the logistic score p_{wrong} for correct vs. wrong retrieval queries. Correct queries (y=0) concentrate near low values of p_{wrong}, while wrong queries (y=1) shift towards higher values. The degree of separation between the two distributions provides visual evidence of how well the model ranks failures by uncertainty.

VPR Model	Image Matching Model	Type	cutoff	R@1	R@5	R@10	R@20	unceraint y	Random	Reranked	Time
SF Validation DataSet											
Mega Loc	LoFTR	Rerank	0.83	92.89	97.34	98.59	99.41	0.446	0.027		
		Adaptive		97.26	98.92	99.22	99.41			0	0
	SuperGlue	Rerank	0.92	91.33	97.72	98.87	99.41	0.426	0.027		

		Adaptive		97.26	98.92	99.22	99.41			0	0
NetVLAD	LoFTR	Rerank	0.02	72.85	75.82	77.36	78.98	0.953	0.467		
		Adaptive		72.78	75.82	77.36	78.98			53.8%	2
	SuperGlue	Rerank	0.47	70.81	75.42	77.10	78.98	0.95	0.467		
		Adaptive		70.89	75.38	77.09	78.98			55.1%	0.73

VPR Model	Image Matching Model	Type	cutoff	R@1	R@5	R@10	R@20	uncerainty	Random	Reranked	Time
SF – Test DataSet											
MegaLoc	LoFTR	Rerank	0.83	85.6	89.2	90.5	91.3	0.716	0.133		
		Adaptive		86.7	89.9	90.8	91.3			0	0
	SuperGlue	Rerank	0.92	85.9	89.4	90.4	91.3	0.73	0.133		
		Adaptive		86.7	89.9	90.8	91.3			0	0
NetVLAD	LoFTR	Rerank	0.02	53.6	53.3	55.8	56.2	0.99	0.728		
		Adaptive		53.5	55.2	55.8	56.2			77.6%	2.89
	SuperGlue	Rerank	0.47	52.60	55.10	56	56.2	0.989	0.728		
		Adaptive		52.3	55.1	56	56.2			77.6%	1.03

VPR Model	Image Matching Model	Type	cutoff	R@1	R@5	R@10	R@20	unceraintу	Random	Reranked	Time
Tokyo – Test DataSet											
MegaLoc	LoFTR	Rerank	0.83	94.92	97.46	98.73	99.05	0.465	0.05		
		Adaptive		94.92	97.46	98.41	99.05			0	0
	SuperGlue	Rerank	0.92	93.65	97.46	98.73	99.05	0.516	0.5		
		Adaptive		94.92	97.46	98.41	99.05			0	0
NetVLAD	LoFTR	Rerank	0.02	68.57	72.70	73.65	78.73	0.965	0.501		
		Adaptive		68.75	72.70	73.65				59.7%	2.22
	SuperGlue	Rerank	0.47	69.21	72.38	74.6	78.73	0.983	0.501		
		Adaptive		69.21	72.38	74.6	78.73			60%	0.79

VPR Model	Image Matching Model	Type	cutoff	R@1	R@5	R@10	R@20	unceraintу	Random	Reranked	Time
Svox_night – Test DataSet											
MegaLoc	LoFTR	Rerank	0.83	92.47	98.06	98.78	99.15	0.421	0.048		
		Adaptive		95.14	98.06	98.78	99.15			0	0
	SuperGlue	Rerank	0.92	90.52	97.45	98.42	99.15	0.548	0.048		

		Adaptive		95.14	98.06	98.78	99.15			0	0
NetVLAD	LoFTR	Rerank	0.02	25.64	27.83	28.55	29.65	0.998	0.919		
		Adaptive		25.64	27.70	28.43	29.65			93.8%	3.49
	SuperGlue	Rerank	0.47	24.54	26.97	27.95	29.65	0.997	0.919		
		Adaptive		24.54	26.97	27.95	29.65			92.1%	1.22

VPR Model	Image Matching Model	Type	cutoff	R@1	R@5	R@10	R@20	uncerainty	Random	Reranked	Time
Svox_sun											
MegaLoc	LoFTR	Rerank	0.83	96.72	99.41	99.41	99.65	0.263	0.035		
		Adaptive		96.49	99.41	99.65	99.65			0.1	0.004
	SuperGlue	Rerank	0.92	94.26	98.95	99.65	99.65	0.225	0.035		
		Adaptive		96.49	99.41	99.65	99.65			0	0
NetVLAD	LoFTR	Rerank	0.02	61.48	63.47	64.52	65.81	0.995	0.646		
		Adaptive		61.12	63.47	64.52	65.81			67.8	2.52
	SuperGlue	Rerank	0.47	60.66	63.23	64.40	65.81	0.992	0.643		
		Adaptive		59.48	63	64.17	65.81			64.3%	0.85

1. **Hard Thresholding:** A binary decision rule based on an optimized integer cutoff.
2. **Logistic Regression:** A probabilistic model that predicts the likelihood of retrieval failure

Introduction

Visual Place Recognition (VPR) is the task of recognizing the location of an image based solely on its visual content, serving as a fundamental building block for autonomous driving and robotics. The standard VPR pipeline treats this as an image retrieval task: a global descriptor aggregates visual features into a vector, which is compared against a database to retrieve the top-k candidates.

To improve localization accuracy, modern pipelines often incorporate a refinement step, where local feature matching models (e.g., LoFTR, SuperGlue, and SuperPoint+LightGlue) re-rank the candidates based on geometric consistency. While this two-stage approach significantly boosts performance, it introduces a critical bottleneck: local matching is computationally expensive.

Motivation: The Performance-Efficiency Trade-off Our preliminary analysis revealed a severe efficiency gap in the standard pipeline. For example, adding LoFTR re-ranking to a NetVLAD baseline on the San Francisco dataset improved Recall@1 by **+26.4 (from 27.2 to 53.6)**, but increased the inference latency to approximately **3.6 seconds per query**. Conversely, applying "blind" re-ranking to robust models like MegaLoc often yields diminishing returns or even performance degradation (dropping from 86.7 to 85.7 Recall@1), wasting computational resources on queries that were already correct.

In this work, we propose an **Adaptive Re-ranking Strategy** to address this trade-off. Instead of processing every query, our system utilizes the **geometric inlier count** to assess retrieval confidence. We investigate two decision mechanisms: **Hard Thresholding** and **Logistic Regression**. Our results demonstrate that the Logistic Regression approach not only optimizes efficiency—reducing re-ranking volume compared to thresholding—but also provides the uncertainty estimation required for safety-critical applications.

Methodology

VPR Pipeline Components:

Following the standard protocol, we employed the following models:

Global Retrievers: We utilized **NetVLAD** as a standard baseline and **MegaLoc** as a robust, state-of-the-art retriever for cross-domain scenarios.

Local Matchers: We applied **LoFTR** (detector-free) and **SuperGlue** (graph-based) to re-rank the top-20 retrieved candidates.

Metric: We utilized the number of **geometric inliers** as the primary metric for re-ranking and confidence estimation

Adaptive Re-ranking Strategies:

To solve the efficiency problem, we implemented an adaptive policy that triggers re-ranking only when the global retrieval is deemed "uncertain."

Strategy 1: Hard Thresholding

We define an integer threshold T to make a binary decision and apply the policy:

If $Inliers_{Top1} < \mathrm{T} \Rightarrow$ Re-rank. Else keep retrieval.

T controls the compute–accuracy trade-off: increasing T reranks a larger fraction of queries (higher cost), while decreasing T reranks fewer queries (lower cost).

We determine T by sweeping candidate values on the **SF-XS validation Dataset**, selecting the threshold that maximizes Recall@1; we choose the smallest T to minimize the fraction of re-ranked queries.

Strategy 2: Logistic Regression

Instead of a hand-chosen inlier cutoff, we train a **logistic regression** model to estimate the probability that the top-1 retrieval is wrong:

$$P_{wrong} = P(top1\ incorrect \mid\ Inliers_{Top1}\ , margin$$

The system triggers re-ranking when:

$$P_{wrong} > P_0$$

where the cutoff $p_0$is chosen on validation to maximize accuracy while minimizing the fraction reranked (cost).

This method serves two purposes:

1. Adaptive Efficiency: It learns a non-linear decision boundary that optimizes the trade-off between accuracy and computational cost.

2. Uncertainty Quantification: Unlike the binary threshold, the regressor outputs a continuous confidence score. As noted in safety-critical contexts, this probability serves as a proxy for uncertainty, allowing the system to quantify *how* reliable a prediction is.

Experimental Results

We evaluated our strategies on the SF-XS (Urban), Tokyo (Urban), and SVOX (Sun/Night) datasets

Baseline Performance: section 5.2

We evaluated our VPR + re-ranking pipeline on **SF-XS** (urban), **Tokyo** (urban), and **SVOX** (Sun/Night cross-domain). We report Recall@K (R@1, R@5, R@10, R@20; K=20). Our baseline results show that re-ranking is highly beneficial for weak retrievers (**NetVLAD**) but can be unnecessary or harmful for strong retrievers (**MegaLoc**), motivating an adaptive policy rather than "always re-rank."

The baseline evaluation confirms two regimes:

- **Weak retrieval regime (NetVLAD):** re-ranking substantially increases performance (e.g., SF improves from **27.2%** retrieval-only to **53.6%** after LoFTR re-ranking).
- **Strong retrieval regime (MegaLoc):** blind re-ranking does not consistently help and may degrade performance (e.g., on SF, MegaLoc drops under re-ranking compared to retrieval-only).

These results justify adaptive re-ranking: allocate matching only to **“hard/uncertain”** queries rather than spending compute on cases where retrieval is already correct.

Adaptive Strategy Evaluation:

We implement two adaptive gating strategies. Both use the **inliers of the top-1 retrieved candidate** as a confidence cue: low inliers indicate weak geometric agreement and higher probability of failure.

Hard Thresholding (inlier cutoff T): We sweep thresholds on validation and select the Tthat maximizes Recall@1 (ties broken by lower rerank fraction).

Dataset effect on the threshold. The optimal threshold depends on dataset difficulty and domain:

- **NetVLAD + LoFTR (SF validation):** best threshold is $T = 132$, yielding **Adaptive R@1 = 72.98** while re-ranking **62.4%** of queries (estimated **2.32 s/query** vs **3.73 s/query** for always re-ranking).
  Interpretation: NetVLAD produces many ambiguous top-1 retrievals, so a higher cutoff triggers re-ranking more often.
- **MegaLoc + SuperGlue (SF validation):** best threshold collapses to **no reranking** (0% reranked), reflecting that MegaLoc is already confident on easy validation data.

Generalization to difficult domain (SVOX Night). On **SVOX Night**, MegaLoc retrieval-only is very strong, and blind re-ranking can significantly reduce accuracy. Using a conservative threshold ($\mathrm{T} = 6$) triggers re-ranking only for a small fraction of queries and mitigates the degradation introduced by always re-ranking, while keeping the compute overhead minimal.

Logistic Regression: Instead of using a fixed inlier threshold, we train a logistic regression model to estimate the probability that the top-1 retrieval is wrong:

- **MegaLoc (SF test):** the model learns that re-ranking is unnecessary or harmful, and therefore triggers re-ranking for **0%** of queries, preserving retrieval performance while avoiding matching cost.
- **NetVLAD (SF test):** the model triggers re-ranking only for uncertain queries, maintaining performance close to always re-ranking while reducing computation.

Uncertainty evaluation (PR curves and AUPRC): We evaluate the quality of p_{wrong} using Precision–Recall curves and **AUPRC**, comparing against a **random baseline equal to the positive rate** (fraction of wrong queries). On the SF test split, MegaLoc has a low positive rate (**0.133**) while NetVLAD has a high positive rate (**0.728**). Despite these different operating regimes, logistic gating produces informative uncertainty estimates: for NetVLAD, AUPRC is near **0.99**, indicating excellent separation between correct and wrong queries; for MegaLoc, AUPRC is around **0.72**, still substantially above the random baseline.

Comparison (NetVLAD + LoFTR, SF test): Hard thresholding ($T = 132$) re-ranked **83.4%** of queries to obtain **R@1 = 53.6**, whereas logistic achieved a comparable **R@1 = 53.5** while re-ranking only **77.6%** of queries. For LoFTR (≈3.73 s/query if always applied), adaptive time is proportional to the reranked fraction (≈2.89 s/query for 77.6%). For SuperGlue (≈1.33s/query if always applied), the same reranked fraction corresponds to ≈1.03 s/query, highlighting that the absolute time savings depend on the chosen local matcher.

Conclusion

Both adaptive strategies reduce unnecessary re-ranking relative to always re-ranking. **Hard thresholding** is simple and effective but is sensitive to dataset/domain shifts and may over-trigger on weak retrievers. **Logistic regression** consistently identifies a more cost-effective operating point (similar Recall@K with fewer queries reranked), and provides a continuous uncertainty score p_{wrong} that can be evaluated via PR/AUPRC and is useful for safety-critical decision-making.