DEXA 2024
Conferences & Workshops
The 35th DEXA Conferences and Workshops
26-28 August 2024
Naples, Italy

R-DiP: Re-ranking based Diffusion Pre-computation for Image Retrieval

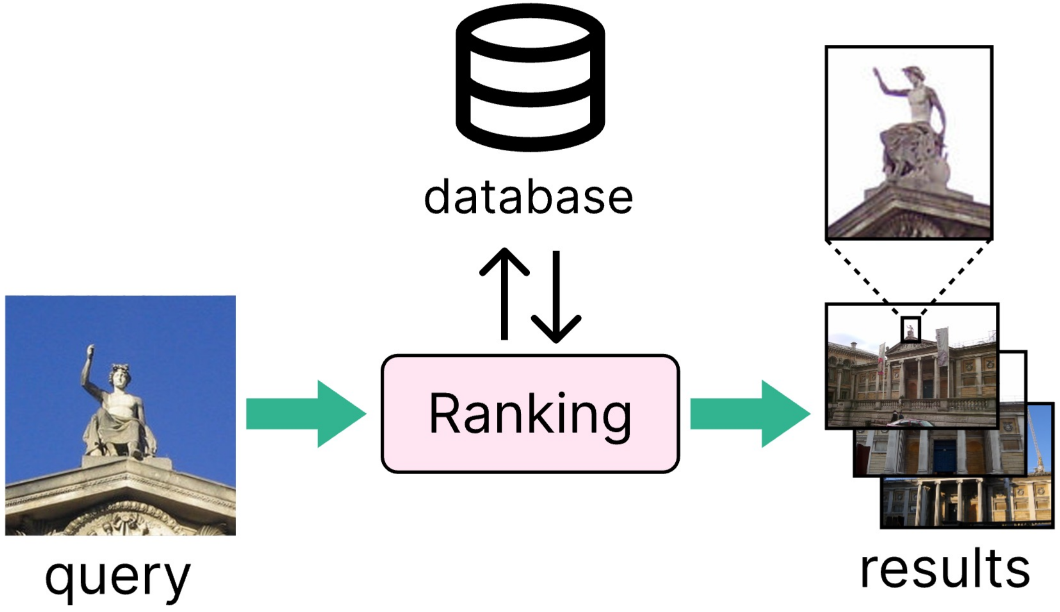
Tatsuya Kato, **Takahiro Komamizu**, and Ichiro Ide

Nagoya University

Contents Science Lab.

Content-Based Image Retrieval（CBIR）

- Goal: find images containing
  a specified object in a query image



- Efficiency-Effectiveness Trade-off

		Efficiency	Effectiveness
Existing Approaches	NN-based re-ranking	×	○
	Offline Diffusion	○	△

How can we take the both advantages in a retrieval framework?

Difficulties of CBIR

- Different **Angle** – Diffusion is strong to this



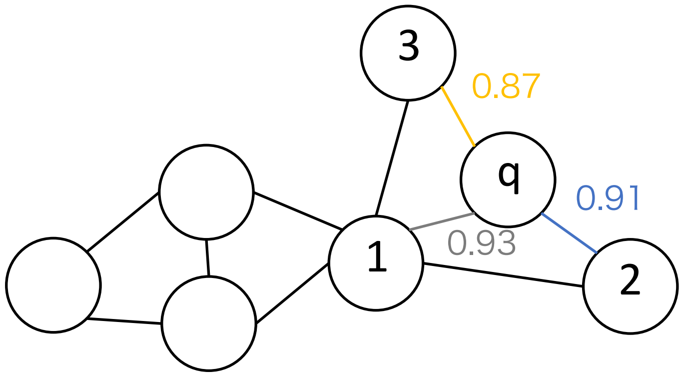

- Different **Viewpoint** – NN-based re-ranking is strong to this




Related Work: Diffusion [9]

1. Graph Construction

- Each node is connected to k nearest neighbors based on image similarities
- Nodes are query and database images

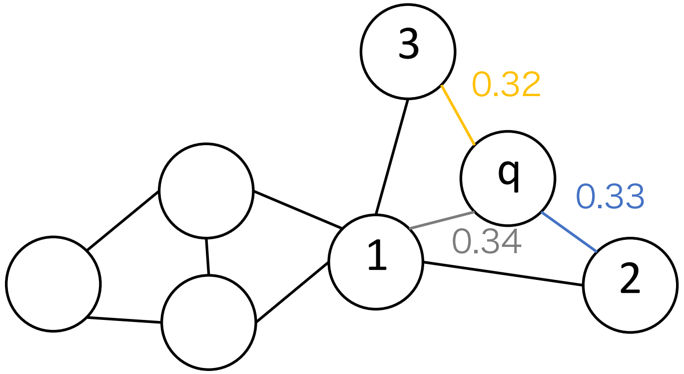


2. Graph Normalization

- Normalize similarities By

  $S = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$

  D is diagonal matrix where $d_{ii} = \sum_{j=1}^{N} a_{ij}$

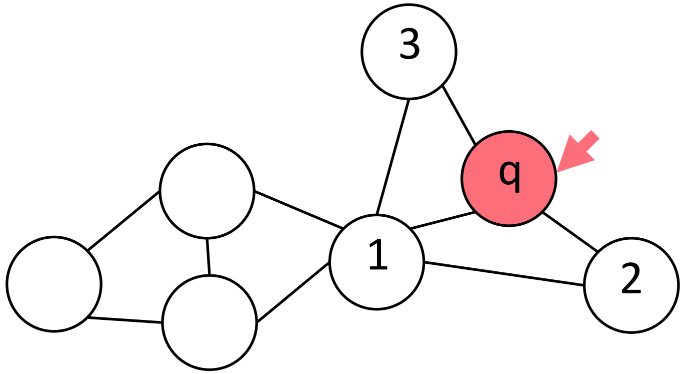


[9] Iscen, A., et al.: Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations. Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 926–935. (2017)

Related Work: Diffusion [9]

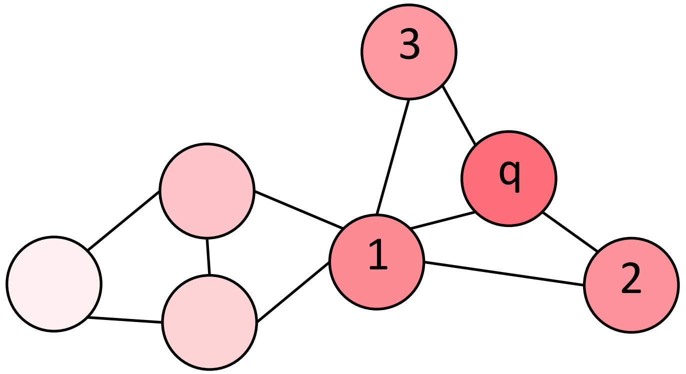
3. Initialization

- State of the ranking scores $f^t = \left[{f_q^t}^\top, {f_d^t}^\top\right]^\top \in \mathbb{R}^{1+n}$
- Initialization: $f_q^0 = 1, f_d^0 = 0 \ \rightarrow f^0 = [1, 0, 0, \dots, 0]^\top$



4. Random Walk: $f^{t+1} = \alpha S f^t + (1-\alpha) f^0$

- Closed form solution $f^\infty = (1-\alpha)(\boldsymbol{I} - \alpha S) f^0$



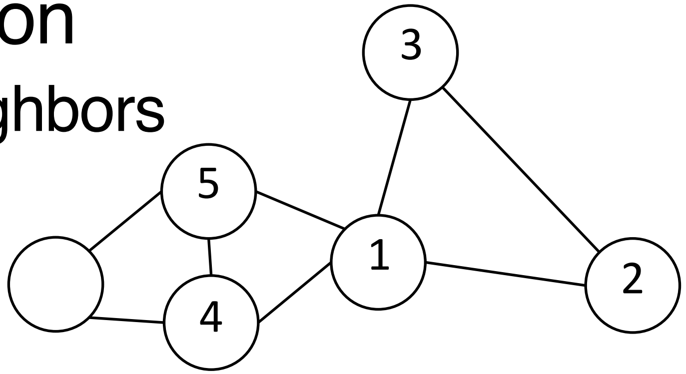
[9] Iscen, A., et al.: Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations. Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 926–935. (2017)

Related Work: Offline Diffusion [29] (Offline)

Idea: Similarity calculation on the database images is pre-computed.

1. Graph Construction and Normalization

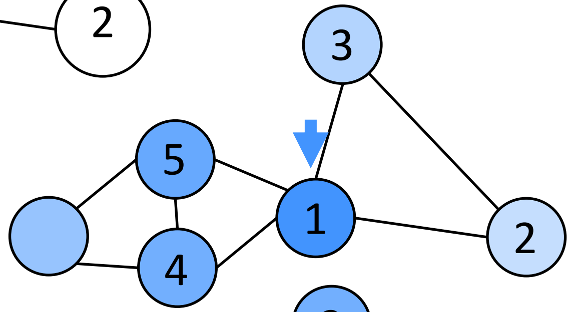
- Each node is connected to k nearest neighbors based on image similarities
- Nodes are only database images



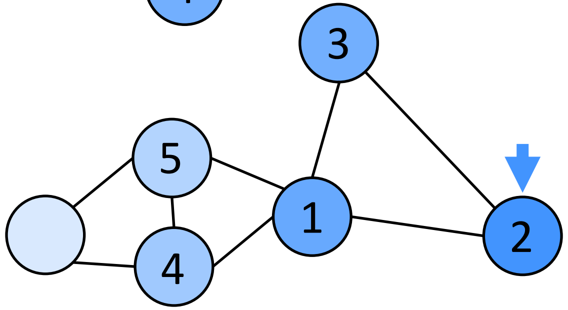
2. Perform diffusion from each node, repeatedly

- Treating each node as the query
- The offline result is $C = [c_1, c_2, \dots, c_n]^\top, c_i \in \mathbb{R}^n$

$c_1 =$



$c_2 =$

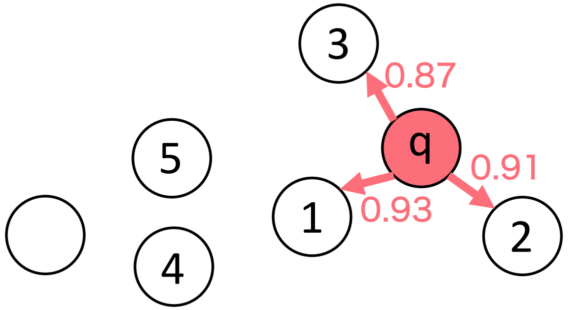


[29] Yang, F., et al.: Efficient image retrieval via decoupling diffusion into online and offline processing. Proc. 33rd AAAI Conf. on Artificial Intelligence (AAAI), pp. 9087–9094. AAAI Press (2019)

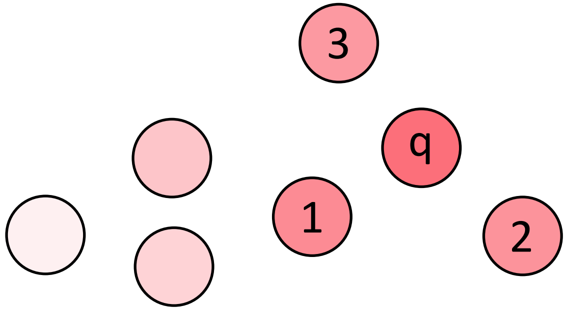
Related Work: Offline Diffusion [29] (Online)

<u>Idea: Linear combination of pre-computed and similarities to query image</u>

1. Perform k-NN search
   - $S = [s_1, s_2, \dots, s_n], s_i = \begin{cases} \mathrm{sim}(q, i) & \text{if } i \in k\mathrm{NN} \\ 0 & else \end{cases}$



2. Linear Combination: $F = S \cdot C$
   - For example,
     - $F = \begin{bmatrix} 0.93 \times c_{1,1} + 0.91 \times c_{2,1} + 0.87 \times c_{3,1} \\ 0.93 \times c_{1,2} + 0.91 \times c_{2,2} + 0.87 \times c_{3,2} \\ \vdots \\ 0.93 \times c_{1,n} + 0.91 \times c_{2,n} + 0.87 \times c_{3,n} \end{bmatrix}^{\mathsf{T}}$



[29] Yang, F., et al.: Efficient image retrieval via decoupling diffusion into online and offline processing. Proc. 33rd AAAI Conf. on Artificial Intelligence (AAAI), pp. 9087–9094. (2019)

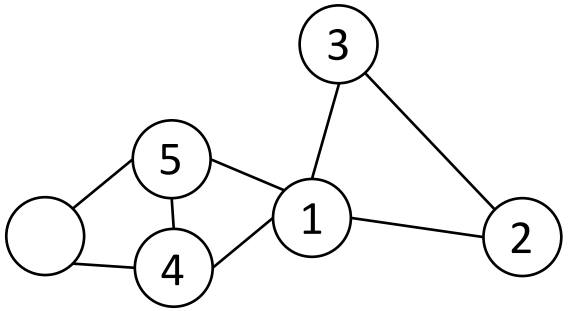
Proposed Method: R-DiP

<u>Idea: Improve Offline Diffusion to be effective.</u>

- Room for improvement in graph construction
  - Replace the simple k-NN search by highly effective re-ranking approaches
- Inherit the benefit of Offline Diffusion:
  - No additional overhead in the search process

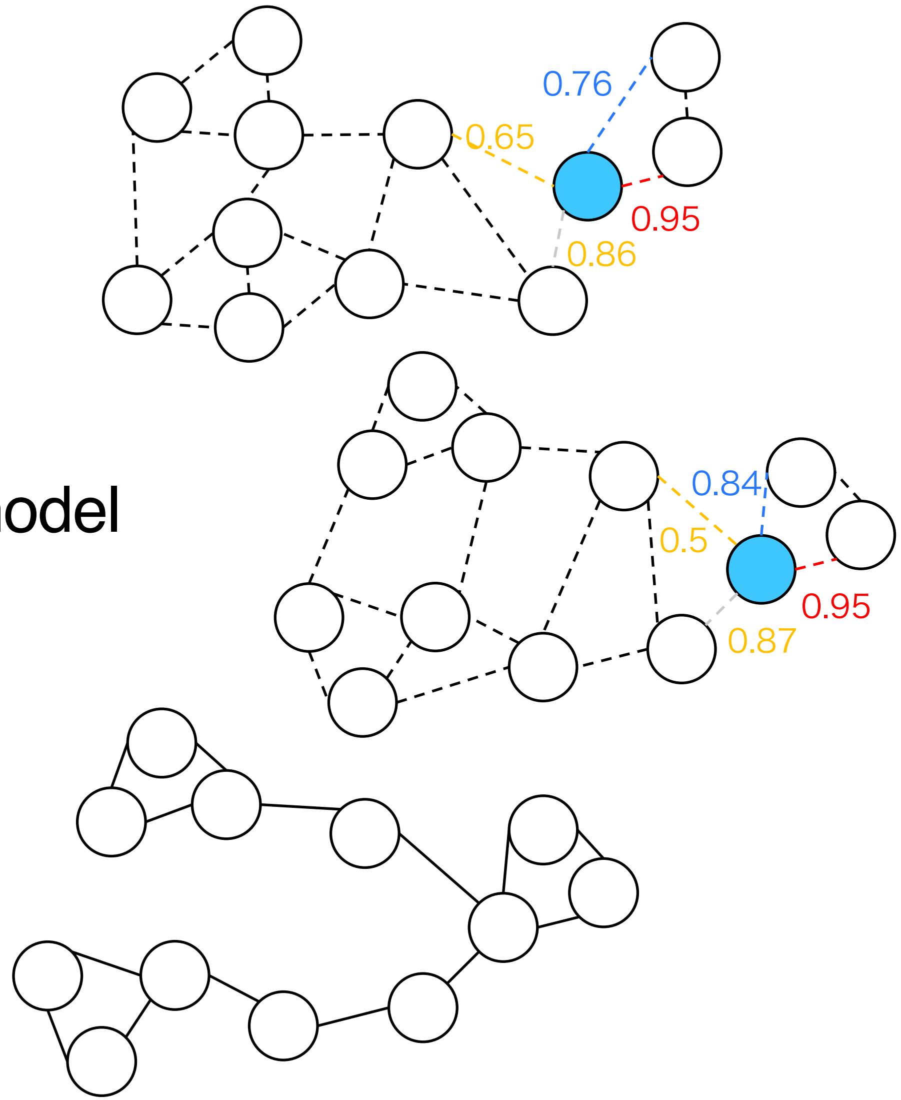
Offline Diffusion

1. Graph Construction and Normalization
   - Each node is connected to k nearest neighbors based on image similarities
   - Nodes are only database images



Offline Step of R-DiP

1. Perform k_m-NN search
   - Search k_m nearest neighbors from each node based on image similarities
2. Re-calculate similarities
   - By more sophisticated NN-based re-ranking model
3. Graph Construction
   - Re-construct the k_d-NN graph ($k_m \geq k_d$)
4. The offline process is
   the same as Offline Diffusion



Wrap-up of the R-DiP – an extensible framework

R-DiP: Offline-Diffusion + Re-ranking Model(s)

- Efficient and Effective
  - Efficiency: Should be comparable to Offline Diffusion.
  - Effectiveness: Should be improved through the similarity re-calculation.
- Flexibility to use any (one or more) similarity re-calculation model
  - Re-ranking model-agnostic
  - Models can be selected based on the objective
    - e.g., focusing on local shapes of objects, overall sight of the images, etc.
  - (future direction) Potentially multiple models can be used simultaneously.
    - e.g., selective to users, or automatic query intent estimation-based model selection.

Experimental Setup

- Implementation of R-DiP: Offline Diffusion [29] + SuperGlobal [21]
- Comparison Methods
  - Offline Diffusion: Verify the similarity re-calculation improves its effectiveness
  - SuperGlobal (SOTA re-ranking method): Verify improving efficiency
- Datasets: ROxford5k, RParis6k (Historical buildings / landmarks) [20]
  - # Database images: 5,000 (ROxford5k), 6,000 (RParis6k)
  - # Query images: 70 for each dataset
  - + R1M Distractor set (about 1 million images): Similar domain images
- Evaluation metric: mean Average Precision (mAP)

[21] Shao, S., et al.: Global features are all you need for image retrieval and reranking. Proc. 19th IEEE/CVF Int. Conf. on Computer Vision (ICCV), pp. 11036–11046. (2023)

[20] Radenovic, F., et al.: Revisiting Oxford and Paris: Large-scale image retrieval benchmarking. Proc. 2018 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 5706– 5715. (2018)

Evaluation of Effectiveness — ROxford5k

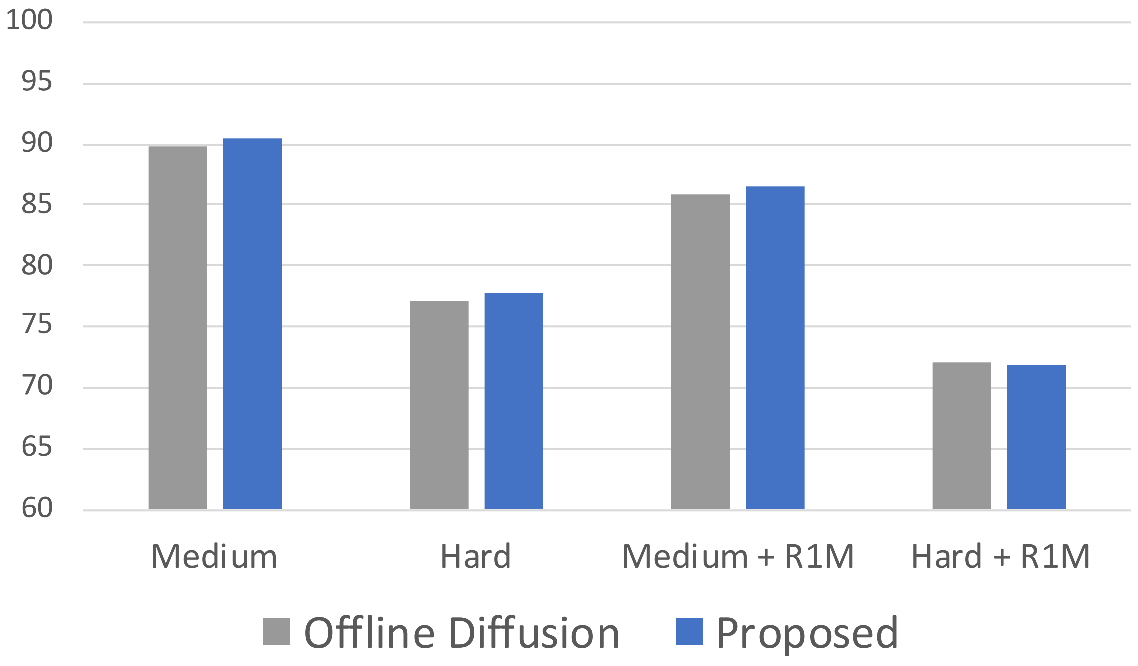
[vs. Offline Diffusion]
Mostly more effective

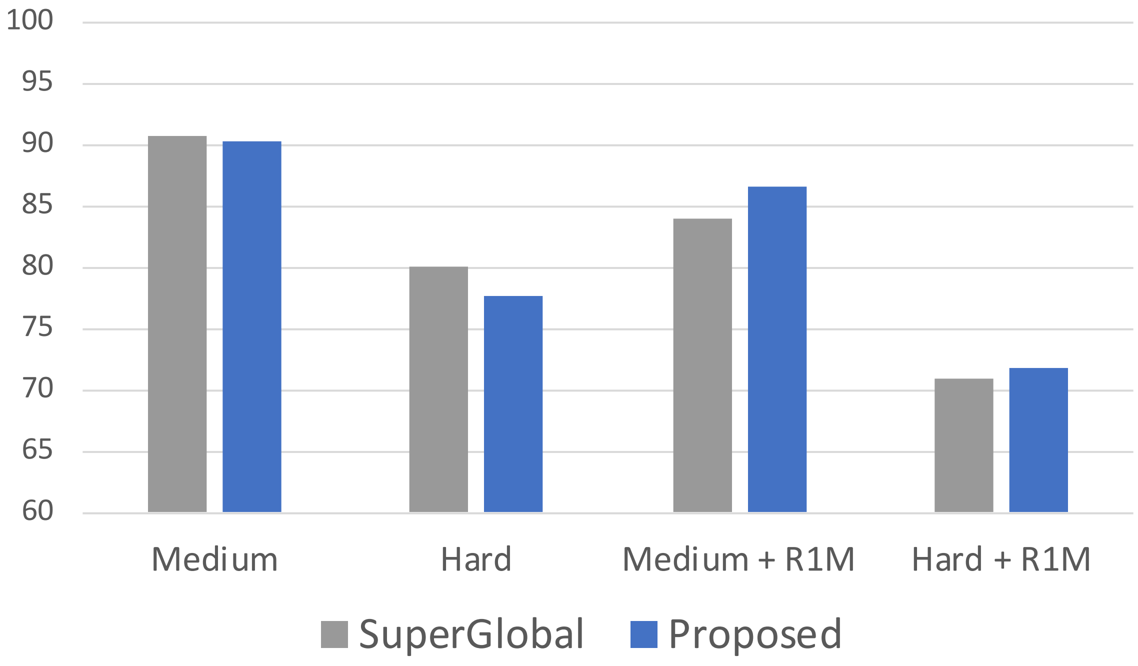
-- Confirms the similarity re-calculation improves Offline Diffusion

[vs. SuperGlobal] More effective on large-scale datasets (+R1M)

-- Suggests robustness to the noisy-but-realistic datasets

mean Average Precision (mAP)

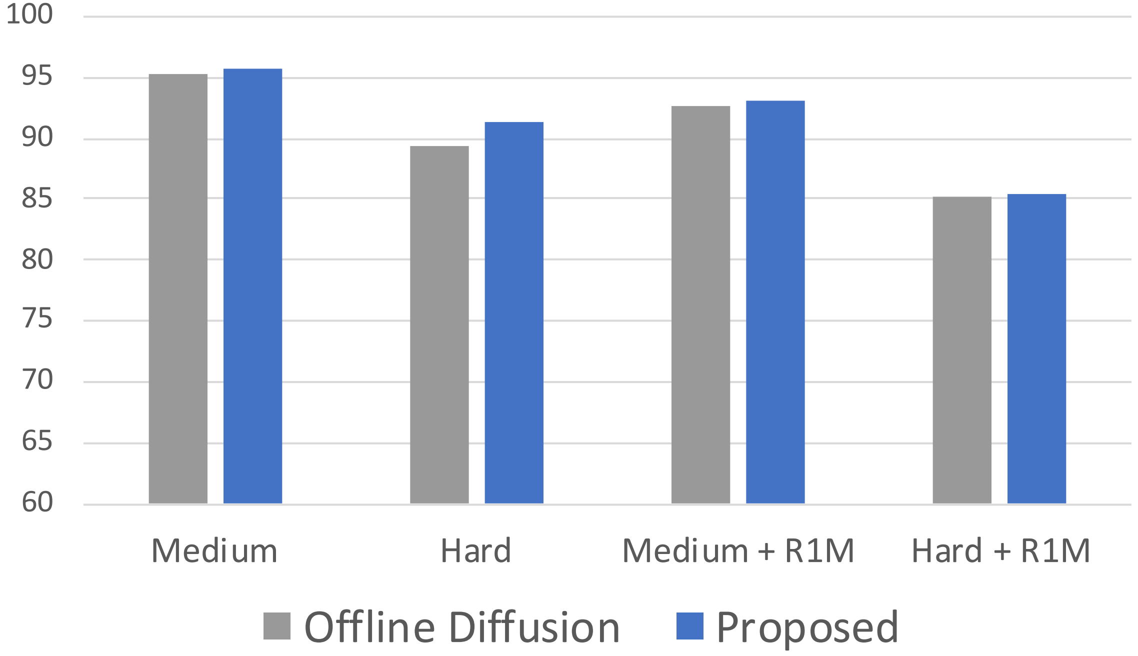


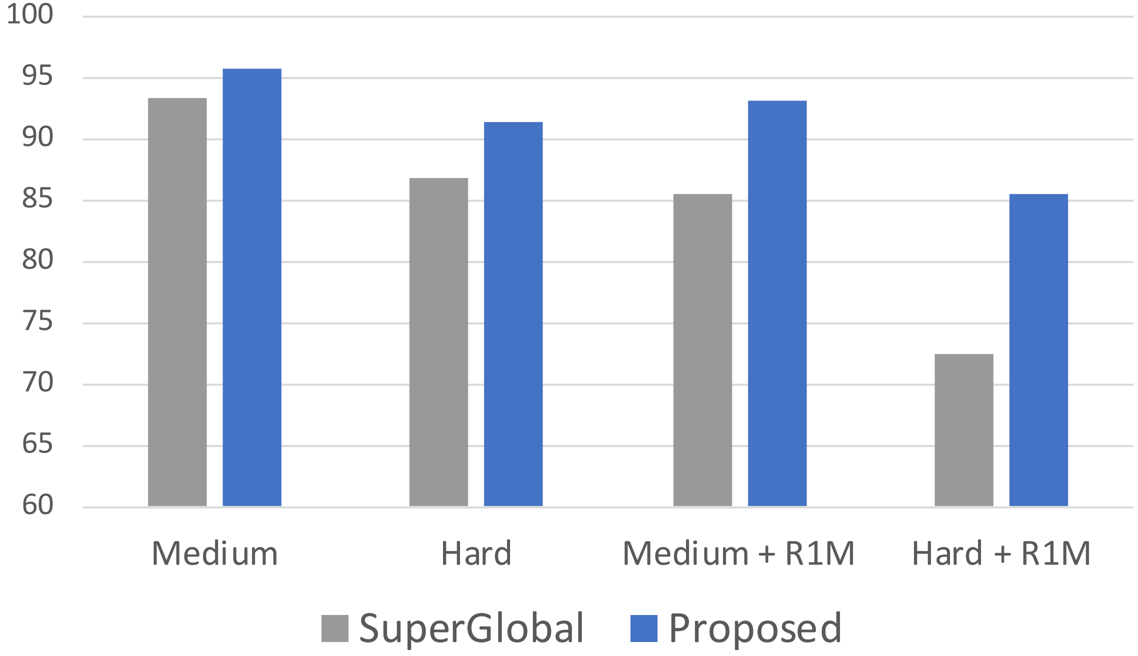


Evaluation of Effectiveness — RParis6k

- Outperforms all comparison methods
  - Significantly surpassing SuperGlobal, especially in large-scale datasets

mean Average Precision (mAP)

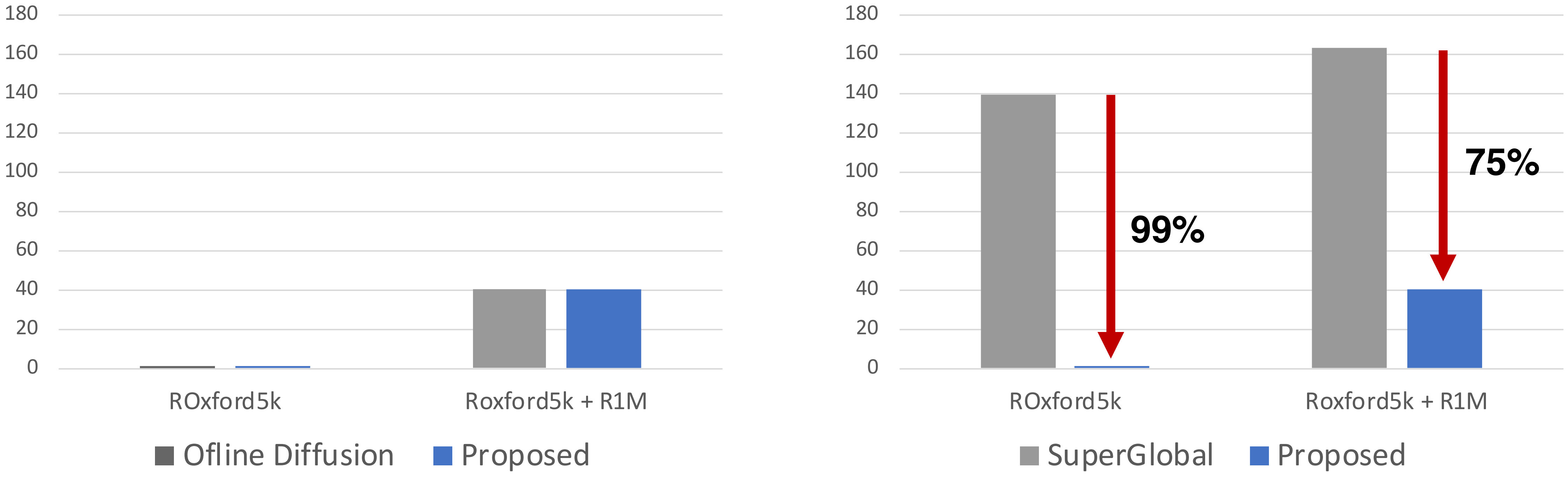




Evaluation of Efficiency

- The same efficiency as Offline Diffusion.
- 75+% reduction compared to SuperGlobal in large-scale datasets

Average retrieval time per query [ms]



Conclusion

- R-DiP: an image retrieval framework
  - Efficient and Effective
  - Flexible to use any similarity re-calculation model
  - Maintains effectiveness even with large-scale datasets
- Future Works
  - Utilizing various similarity re-calculation models
  - Integrating multiple similarity re-calculation models
  - Addressing the overhead of pre-computation with similarity re-calculation models