

Unsupervised Effectiveness Estimation Measure Based on Rank Correlation for Image Retrieval: Supplemental Document

1. ADDITIONAL ILLUSTRATIONS

Figure S1 illustrates the process of Query Performance Prediction (QPP) in a Content-based Image Retrieval (CBIR) system. Given a dataset of images, a descriptor extracts a feature for each image. Different feature extraction models can be used, from hand-crafted (e.g., color, texture, shape) to deep learning ones (e.g., Convolutional Neural Networks and Vision Transformers). The descriptor also computes the distance between these images originating ranked lists. This work employed the Euclidean distance. A ranked list is a structure that, for a given query image, sorts the most similar images in descending order of similarity. These ranked lists provide different effectiveness based on the extraction model used. Given a ranked list, the QPP approach predicts a score that estimates its quality. To compute the QPP for a descriptor, each image in the dataset is treated as a query. The average result then serves as an effectiveness estimation for that descriptor. The higher the score, the better the expected effectiveness.

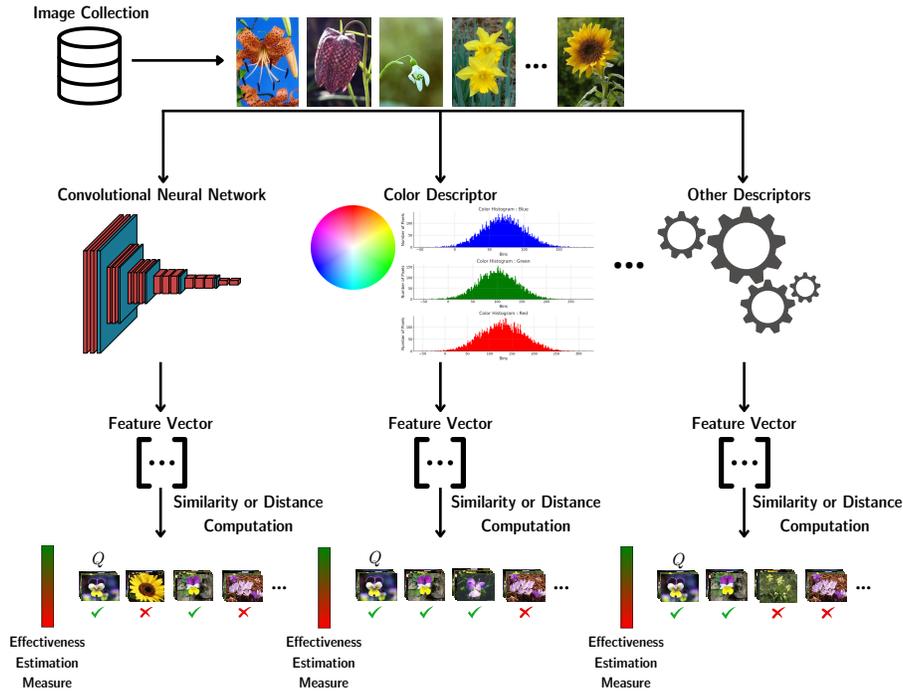


Fig. S1. Workflow illustrating the steps for computing an effectiveness estimation (QPP) measure for image retrieval.

2. ADDITIONAL RESULTS

Table S1 presents an expanded version of the results for the Flowers [1] and Core5k [2] datasets, including additional descriptors and $\alpha = 0.80$ for the Acc. JacMax measure. The Pearson correlation between the Mean Average Precision (MAP) and the effectiveness estimation measures was reported for the proposed approach and three baselines: Authority [3], Reciprocal [4], and Deep Rank Noise Estimator (DRNE) [5]. Note that DRNE values are only available for the Flowers dataset, as the DRNE paper did not include results for the Core5k dataset. Gray highlights the best value per row, and bold highlights the best value of the proposed measure (Acc. JacMax). The descriptors are sorted by their original MAP, demonstrating that higher correlation values

are typically found in descriptors with high effectiveness and vice versa. It is noteworthy that the proposed approach achieved the best results in most cases. It also demonstrates that Acc. JacMax is robust to variations in the parameter α , with $\alpha = 0.95$ and $\alpha = 0.90$ likely being close to the optimal choice in most situations.

Table S1. Pearson correlation (between MAP and QPP, the higher the better) for our proposed approach compared to baselines on **Flowers and Corel5k datasets**. Gray highlights the best value per row, and bold highlights the best value of the proposed measure (Acc. JacMax).

Datasets	Descriptors	Original MAP	Baselines			Acc. JacMax (Ours)				
			Auth. [3]	Recip. [4]	DRNE [5]	$\alpha = 1$	$\alpha = 0.99$	$\alpha = 0.95$	$\alpha = 0.90$	$\alpha = 0.80$
Flowers	CNN-FBResNet [6]	52.56%	0.73744	0.67153	0.79920	0.79971	0.81825	0.85141	0.84601	0.80387
	CNN-ResNeXt [7]	51.91%	0.76568	0.66525	0.79265	0.81018	0.82847	0.86458	0.86125	0.81776
	CNN-ResNet [6]	51.83%	0.72981	0.63672	0.79903	0.78955	0.80859	0.84203	0.83306	0.77900
	CNN-Xception [8]	47.31%	0.74365	0.64060	0.76958	0.77405	0.79445	0.84027	0.84046	0.79514
	CNN-AlexNet [9]	46.04%	0.46586	0.35353	0.63521	0.57404	0.61838	0.69974	0.70822	0.66950
	CNN-SENet [10]	43.16%	0.58722	0.57195	0.63076	0.62609	0.64647	0.68539	0.67083	0.61455
	CNN-InceptRN [11]	42.20%	0.62725	0.53364	0.55041	0.66704	0.67726	0.69233	0.67876	0.63954
	CNN-BnVGGNet [12]	41.87%	0.48524	0.36175	0.63133	0.59161	0.62697	0.69856	0.70519	0.66306
	CNN-NASNetLg [13]	40.74%	0.63091	0.55103	0.54974	0.65723	0.66711	0.67673	0.65818	0.61168
	CNN-VGGNet [12]	39.05%	0.50498	0.32844	0.63850	0.61308	0.65163	0.72334	0.72956	0.68988
	SIFT [14]	28.47%	0.34815	0.31624	0.48026	0.44652	0.48238	0.55495	0.56620	0.54478
	SPJCD [15, 16]	22.56%	0.27962	0.24767	0.33553	0.33251	0.35104	0.40788	0.43425	0.43471
	SPCEDD [16, 17]	21.94%	0.31110	0.26055	0.34731	0.36155	0.38084	0.43446	0.45146	0.43906
	COMO [18]	21.83%	0.10506	0.08213	0.25892	0.15345	0.18135	0.25700	0.28658	0.29025
	SPFCTH [16, 19]	21.73%	0.19618	0.18878	0.26632	0.24354	0.26151	0.31856	0.35007	0.35978
	JCD [15]	21.43%	0.15319	0.11306	0.24018	0.18774	0.20256	0.24853	0.27256	0.27445
	FCTH [19]	20.56%	0.18428	0.13488	0.23862	0.21393	0.22587	0.26483	0.28575	0.28750
	CEDD [17]	20.37%	0.13077	0.10192	0.20104	0.15773	0.17221	0.21587	0.23871	0.24583
	SPACC [16, 20]	19.20%	0.07436	0.03312	0.20229	0.14722	0.18318	0.29400	0.34779	0.36244
	ACC [20]	18.99%	0.03264	0.02153	0.28373	0.11935	0.15978	0.27446	0.32457	0.33329
	PHOG [16, 21]	15.47%	0.33586	0.33548	0.37418	0.33371	0.34181	0.37008	0.39126	0.40979
	EHD [22]	12.56%	0.03510	0.06457	0.20214	0.07389	0.09381	0.17720	0.25335	0.32491
	SPLBP [16, 23]	11.26%	0.06942	0.07869	0.14425	0.08773	0.10019	0.14908	0.19222	0.23649
	LBP [23]	10.34%	0.01482	0.02083	0.07323	0.02800	0.03239	0.05366	0.07740	0.10508
SCD [24]	10.25%	0.25619	0.10035	0.05702	0.27886	0.27424	0.25204	0.22785	0.19472	
Average	29.35%	0.35219	0.29657	0.42006	0.40273	0.42323	0.47388	0.48928	0.47708	
Corel5k	CNN-ResNet [6]	64.86%	0.84968	0.81174	—	0.84720	0.85176	0.83879	0.80230	0.73602
	CNN-FBResNet [6]	64.25%	0.85457	0.80835	—	0.85130	0.85604	0.84728	0.81674	0.75870
	CNN-InceptRN [11]	61.31%	0.80549	0.81055	—	0.79166	0.79247	0.76302	0.71192	0.64129
	CNN-ResNeXt [7]	62.45%	0.87020	0.82547	—	0.86710	0.87148	0.85955	0.82470	0.76234
	CNN-SENet [10]	57.10%	0.78353	0.82476	—	0.78396	0.78806	0.76699	0.72028	0.65412
	CNN-Xception [8]	54.60%	0.87838	0.84928	—	0.87071	0.87368	0.85464	0.81626	0.75233
	CNN-NASNetLg [13]	53.78%	0.81245	0.84444	—	0.82735	0.83450	0.81993	0.77578	0.71019
	CNN-BnVGGNet [12]	52.82%	0.82058	0.78008	—	0.83687	0.84614	0.85010	0.82639	0.77086
	CNN-VGGNet [12]	47.99%	0.80709	0.77088	—	0.82886	0.83869	0.83907	0.81165	0.75505
	CNN-AlexNet [9]	37.89%	0.67625	0.72093	—	0.73486	0.75387	0.76378	0.72285	0.64486
	SPCEDD [16, 17]	28.98%	0.68519	0.67809	—	0.73346	0.74562	0.76306	0.74719	0.69659
	SPJCD [15, 16]	28.30%	0.68950	0.69116	—	0.73979	0.75044	0.76490	0.74741	0.69283
	SPFCTH [16, 19]	26.68%	0.64710	0.64830	—	0.70244	0.71438	0.73341	0.72064	0.67298
	JCD [15]	25.01%	0.76185	0.70122	—	0.81234	0.81735	0.81204	0.78133	0.71018
	FCTH [19]	24.19%	0.74485	0.68045	—	0.79078	0.79604	0.79296	0.76289	0.69111
	SPACC [16, 20]	24.11%	0.65464	0.55990	—	0.70161	0.71784	0.73828	0.72216	0.67593
	ACC [20]	23.70%	0.72589	0.64916	—	0.77260	0.78447	0.79393	0.76618	0.69778
	CEDD [17]	23.29%	0.70728	0.63540	—	0.76847	0.77598	0.77307	0.73785	0.65865
	COMO [18]	21.30%	0.65908	0.56197	—	0.70781	0.71315	0.70198	0.66650	0.59797
	EHD [22]	17.03%	0.60975	0.57463	—	0.64772	0.65978	0.67819	0.66838	0.62481
	PHOG [16, 21]	16.07%	0.41810	0.38642	—	0.45532	0.48034	0.54862	0.57446	0.57009
	SPLBP [16, 23]	15.74%	0.49790	0.49686	—	0.55106	0.56733	0.60112	0.60245	0.57407
	LBP [23]	15.13%	0.49217	0.37729	—	0.54693	0.55824	0.57333	0.56229	0.52252
	SCD [24]	14.91%	0.50915	0.41427	—	0.53831	0.54961	0.56755	0.55750	0.51647
SIFT [14]	12.97%	0.52773	0.51391	—	0.37515	0.40413	0.45880	0.47267	0.47119	
Average	34.98%	0.69954	0.66462	—	0.72335	0.73366	0.74018	0.71675	0.66236	

Supporting the paper’s findings, the results were organized into bar graphs to facilitate comparison between our approach and baseline methods. Figure S2 presents the Pearson correlation between MAP and the QPP approaches for the MPEG-7 (Figure S2a) and Brodatz (Figure S2b) datasets. The red bar represents the Acc. JacMax, showing the results for the best α value reported in the tables. The results were reported for the Regression for Query Performance Prediction Framework (RQPPF) [25] considering its two variants: “RQPPF + A” and “RQPPF + R”, which considered Authority and Reciprocal for computing its meta-features, respectively. For MPEG-7 [26], although the Accumulated Jaccard Max did not yield the best results, it remains highly competitive among the baseline measures, suggesting its potential for shape descriptors. Conversely, the Brodatz [27] dataset highlights the superior performance of the proposed measure, underscoring its effectiveness for texture descriptors.

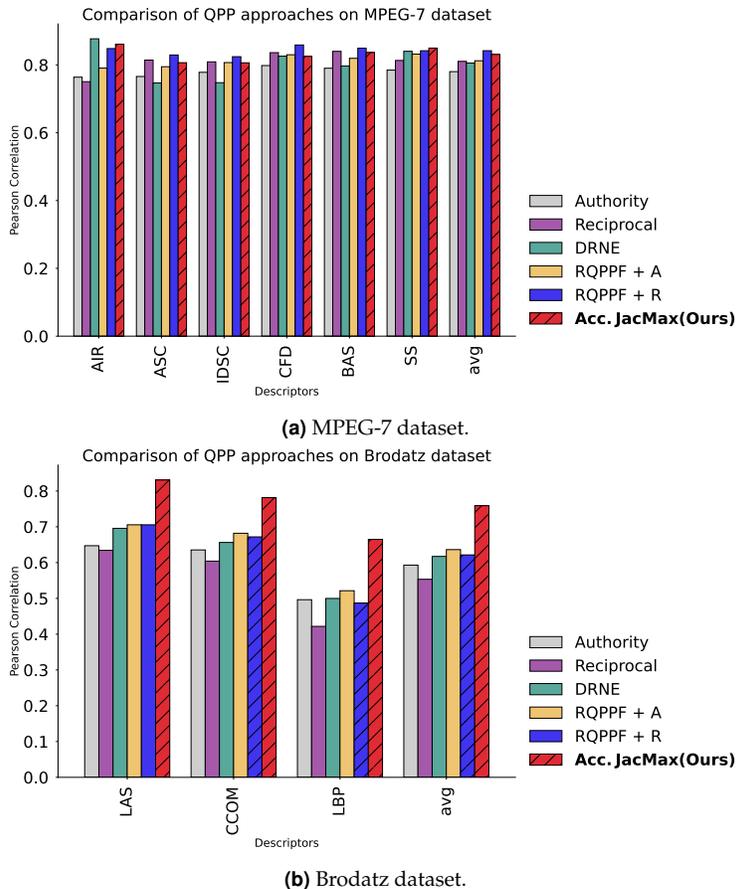


Fig. S2. Bar graphs showing the Pearson correlation between MAP and various QPP approaches, including baselines and our proposed measure (Acc. JacMax).

Aiming to visually compare the correlation of the proposed approach with MAP concerning baselines, Figures S3, S4, and S5 present visualizations for the datasets Flowers [1], Corel5k [2], and Soccer [28], respectively. For each dataset, there is a plot for (a) Authority [3]; (b) Reciprocal [4]; and (c) our proposed Acc. JacMax. In each graph, each point corresponds to a ranked list. The MAP values of the ranked lists are presented on the x -axis, and the effectiveness estimation values are on the y -axis. All the measures exhibit an approximately linear and positive behavior. The higher the correlation, the more linear the behavior tends to be, which is indicated by the Pearson correlation. In all cases, our proposed measure outperformed the baselines. For the Corel5k and Soccer datasets, represented by Figures S4 and S5, despite the Authority Score achieving similar results to the Accumulated JaccardMax, our measure yielded better results than Authority, considering the Pearson correlation value.

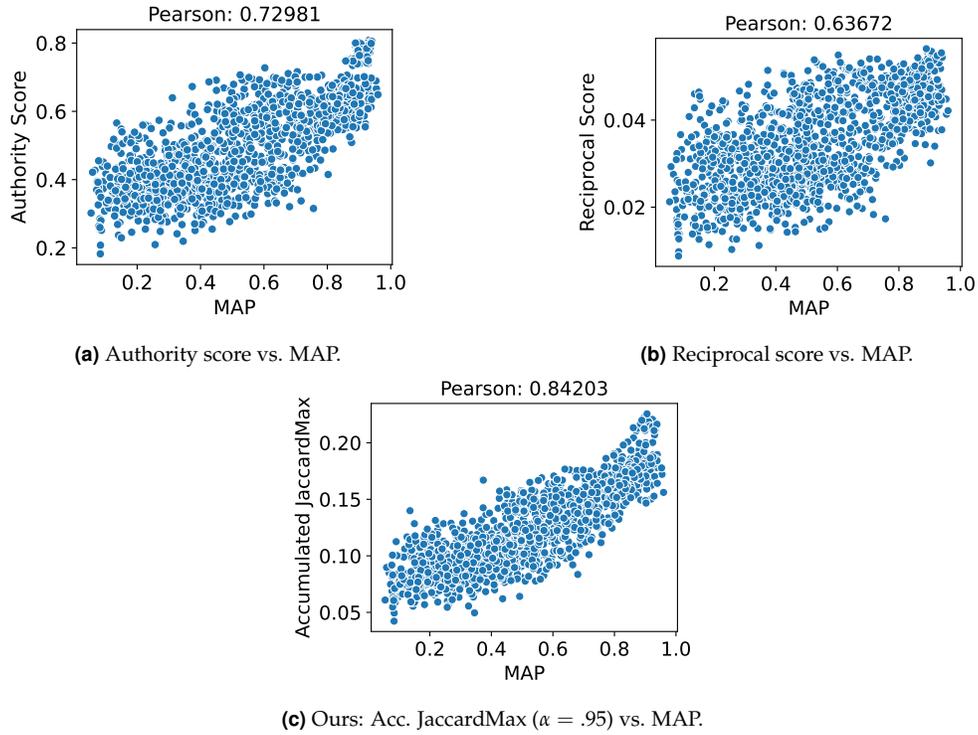


Fig. S3. Correlation of MAP and QPP measures on **Flowers dataset** with ResNet descriptor.

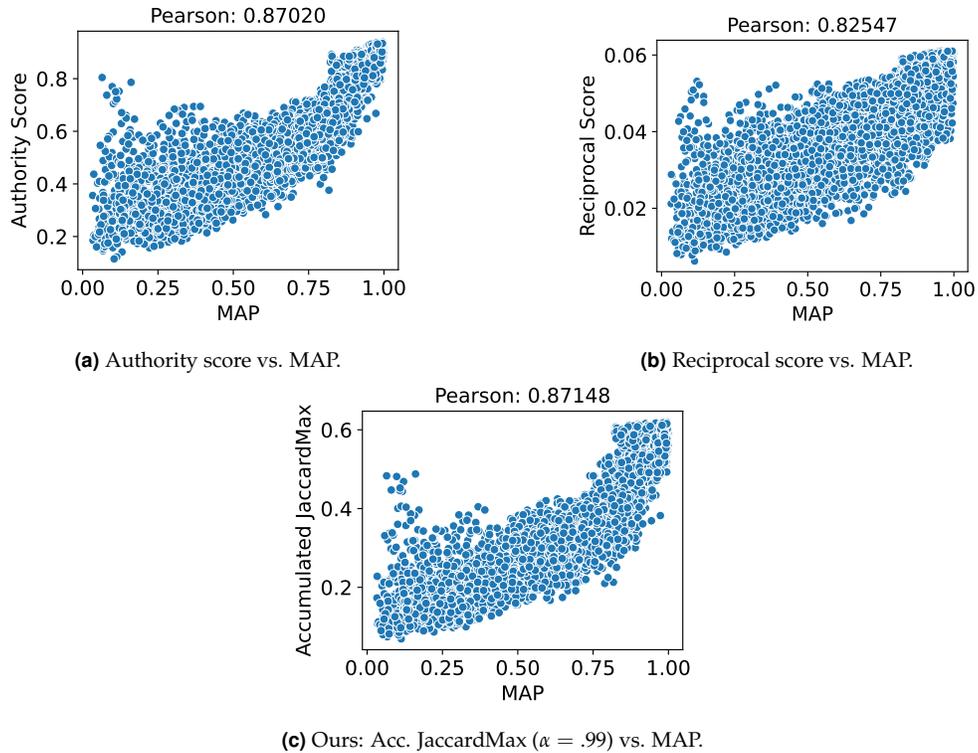


Fig. S4. Correlation of MAP and QPP measures on **Core5k dataset** with ResNeXt descriptor.

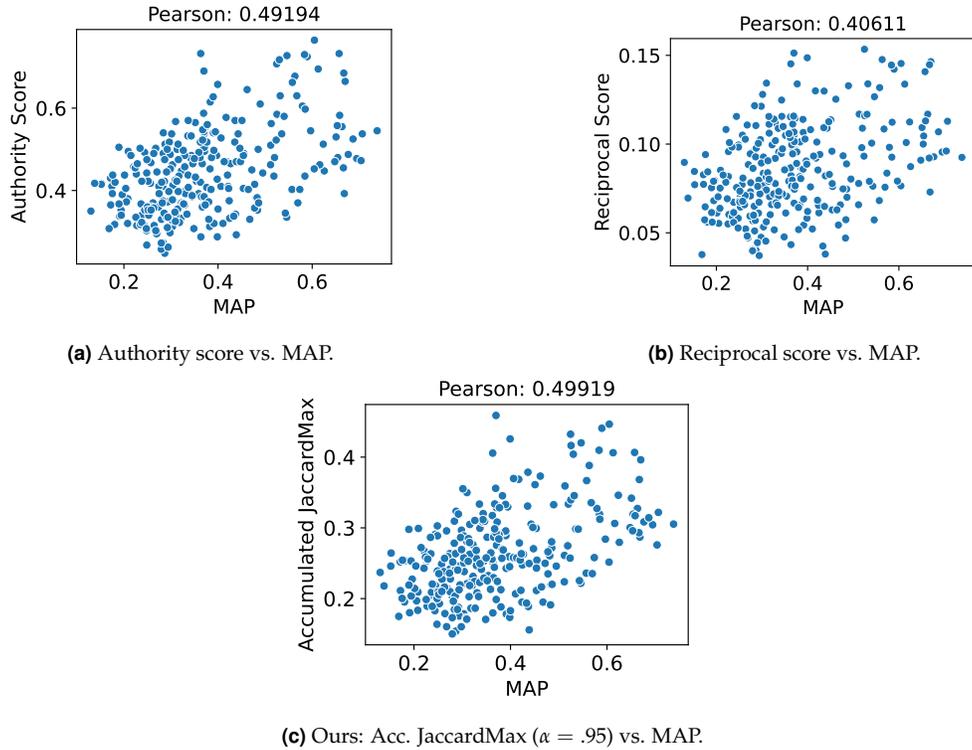


Fig. S5. Correlation of MAP and QPP measures on **Soccer dataset** with ACC descriptor.

Figure S6 presents ranked lists to illustrate the prediction of the Accumulated JaccardMax measure compared to the MAP. The examples are shown for Flowers [1] (Figure S6a) and Soccer [28] (Figure S6b) datasets. Query images are highlighted in green, while images from different classes than the query are highlighted in red. For each dataset, a good query and a bad query were selected. Notice that the proposed measure exhibits a higher score for lists with a higher MAP, and vice versa. However, the scale of the values varies depending on the dataset.

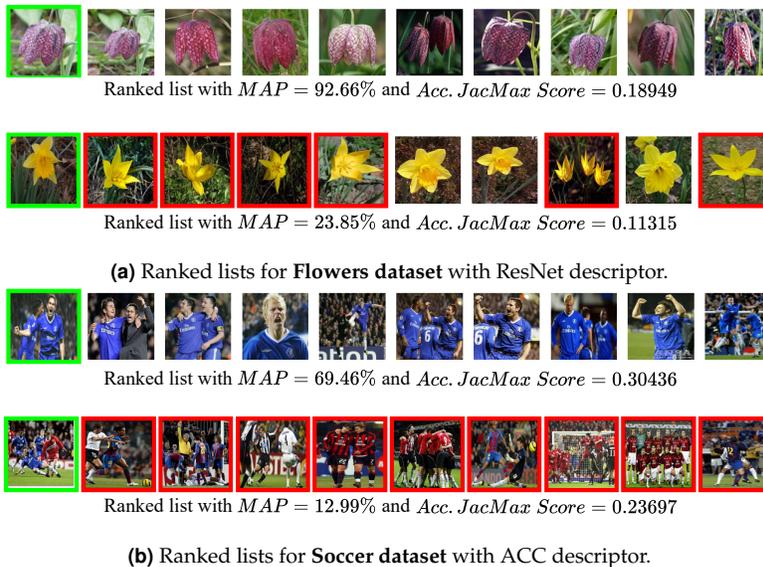


Fig. S6. Examples of ranked lists (good and bad queries) along with their Mean Average Precision (MAP) values and Accumulated JaccardMax scores.

3. CODE AND IMPLEMENTATION

The code of the proposed approach (Accumulated Jaccard Max) will be made available upon acceptance of the manuscript. Regarding baselines, the Authority [3] and Reciprocal [4] were executed considering a publicly available implementation [29]: [UDLF/USRF/blob/main/effectiveness_estimation_functions.py](https://github.com/UDLF/USRF/blob/main/effectiveness_estimation_functions.py).

REFERENCES

1. M.-E. Nilsback and A. Zisserman, "A Visual Vocabulary for Flower Classification," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, vol. 2 (IEEE, New York, NY, USA, 2006), pp. 1447–1454.
2. G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognit.* **46**, 188–198 (2013).
3. D. C. G. Pedronette, O. A. Penatti, and R. da S. Torres, "Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks," *Image Vis. Comput.* **32**, 120 – 130 (2014).
4. D. C. G. Pedronette, O. A. B. Penatti, R. T. Calumby, and R. da Silva Torres, "Unsupervised distance learning by reciprocal knn distance for image retrieval," in *International Conference on Multimedia Retrieval, ICMR '14, Glasgow, United Kingdom - April 01 - 04, 2014*, M. S. Kankanhalli, S. M. Rüger, R. Manmatha, et al., eds. (ACM, 2014), p. 345.
5. L. P. Valem and D. C. G. a. Pedronette, "A denoising convolutional neural network for self-supervised rank effectiveness estimation on image retrieval," in *Proceedings of the 2021 International Conference on Multimedia Retrieval*, (Association for Computing Machinery, New York, NY, USA, 2021), ICMR '21, p. 294–302.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), pp. 770–778.
7. S. Xie, R. Girshick, P. Dollár, et al., "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 5987–5995.
8. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), pp. 1800–1807.
9. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25 F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds. (Curran Associates, Inc., 2012).
10. J. Hu, L. Shen, S. Albanie, et al., "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis Mach. Intell.* **42**, 2011–2023 (2020).
11. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. of the 31st AAAI Conference on Artificial Intelligence*, (AAAI Press, 2017), AAAI'17, p. 4278–4284.
12. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, (2015), pp. 730–734.
13. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), pp. 8697–8710.
14. D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2 (1999), pp. 1150–1157 vol.2.
15. K. Zagoris, S. A. Chatzichristofis, N. Papamarkos, and Y. S. Boutalis, "Automatic Image Annotation and Retrieval Using the Joint Composite Descriptor," in *2010 14th Panhellenic Conference on Informatics*, (2010), pp. 143–147.
16. M. Lux, "Content based image retrieval with LIRe," in *Proceedings of the 19th ACM international conference on Multimedia*, (ACM, Scottsdale Arizona USA, 2011), pp. 735–738.
17. S. A. Chatzichristofis and Y. S. Boutalis, "Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Computer Vision Systems*, A. Gasteratos, M. Vincze, and J. K. Tsotsos, eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 312–322.
18. S. A. Vassou, N. Anagnostopoulos, A. Amanatiadis, et al., "CoMo: A Compact Composite Moment-Based Descriptor for Image Retrieval," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, (ACM, Florence Italy, 2017), pp. 1–5.

19. S. A. Chatzichristofis and Y. S. Boutalis, "FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval," in *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, (2008), pp. 191–196. ISSN: 2158-5881.
20. Jing Huang, S. Kumar, M. Mitra, *et al.*, "Image indexing using color correlograms," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (IEEE Comput. Soc, San Juan, Puerto Rico, 1997), pp. 762–768.
21. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1 (2005), pp. 886–893 vol. 1.
22. B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits Syst. for Video Technol.* **11**, 703–715 (2001).
23. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis Mach. Intell.* **24**, 971–987 (2002).
24. L. Cieplinski, "Mpeg-7 color descriptors and their applications," in *Computer Analysis of Images and Patterns*, W. Skarbek, ed. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001), pp. 11–20.
25. L. Valem, V. Pereira-Ferrero, and D. Pedronette, "Self-supervised regression for query performance prediction on image retrieval," in *2023 IEEE Sixth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, (IEEE Computer Society, Los Alamitos, CA, USA, 2023), pp. 95–98.
26. L. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 1 (2000), pp. 424–429 vol.1. ISSN: 1063-6919.
27. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover books on art, graphic art, handicrafts (Dover Publications, 1966).
28. J. Van De Weijer and C. Schmid, "Coloring Local Feature Extraction," in *Computer Vision – ECCV 2006*, vol. 3952 A. Leonardis, H. Bischof, and A. Pinz, eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006), pp. 334–348.
29. L. P. Valem and D. C. G. Pedronette, "Unsupervised selective rank fusion for image retrieval tasks," *Neurocomputing* **377**, 182–199 (2020).