Unsupervised Selective Rank Fusion on Content-Based Image Retrieval

Lucas Pascotti Valem[1], Daniel Carlos Guimarães Pedronette
Department of Statistics, Applied Math. and Computing
São Paulo State University (UNESP)
Rio Claro, SP - Brazil
{lucas.valem,daniel.pedronette}@unesp.br

***Abstract*—Mainly due to the evolution of technologies to store and share images, the growth of image collections have been remarkable for years. Therefore, developing effective methods to index and retrieve such extensive available visual information is indispensable. The CBIR (Content-Based Image Retrieval) systems are one of the main solutions for image retrieval tasks. These systems are mainly supported by the use of different visual descriptors and machine learning methods. Despite the relevant advances in the area, mainly driven by deep learning technologies, accurately computing the similarity between images remains a complex task in various scenarios due to the well known semantic gap problem. As distinct features produce complementary ranking results with different effectiveness performance, a promising solution consists in combining them. However, how to decide which visual features to combine is a very challenging task. This work proposes three novel methods for selecting and combining ranked lists by estimating their effectiveness in an unsupervised way. The approaches were evaluated in five different image collections and several descriptors, achieving results comparable or superior to the state-of-the-art in most of the evaluated scenarios.**

I. Introduction

In the modern world, the task of searching images through their visual content is essential, mostly due to the huge amount of data available. The Content-Based Image Retrieval (CBIR) systems are originally based on the use of descriptors [5], [6] for encoding and retrieving images based on visual properties. In fact, a myriad of different descriptors are available, often providing distinct and complementary results even for a same query [7]. Different categories of descriptors may be more appropriate to certain problems than others. A classic and simple example that represents this matter is when we compare the image of an orange to the one of a lemon [7]. By considering a shape descriptor, the score of similarity between them tends to be very high. Although, if a color descriptor is employed, a very low similarity is obtained.

Recently, features based on deep convolutional neural networks [8], [9] have been achieving very promising results in many diverse scenarios. Most of the networks are often trained in the ImageNet dataset [8] and fine-tuned to the task being addressed. However, besides the very high results achieved, there are still scenarios where researchers find out indicatives that the effectiveness of these systems can be further improved when considering other types of visual descriptors [10].

[1]This work relates to a M.Sc. dissertation [1].

In the recent years, different methods have been proposed with the intent of fusing results from distinct visual features [10]–[13]. Most of them still require labels or user intervention to achieve effective results, and only few works have addressed aspects of selection without labeled data [14]–[18]. Despite the fact that fusion strategies in unsupervised scenarios have achieved significant gains in terms of effectiveness [16]–[18], there are still relevant open research challenges. In most of the cases, the fusion approaches are static and the selection of descriptors is *ad hoc* [16]–[18], once there is no dynamic methodology for selecting the visual descriptors to be combined.

Our objective is to investigate and propose novel methods based on unsupervised learning approaches for dynamic selection and combination of rankings obtained through different visual descriptors or retrieval systems, without the need of any training data or user intervention. Considering the diversity of results obtained from different descriptors, we assume that the best combinations are provided from results of high effectiveness and low correlation. In such situations, the combination can exploit the complementarity and generate results of higher quality.

However, selecting the best rankers that fit this criteria in unsupervised scenarios involves multiple research challenges, including: (*i*) find the optimal way to determine the rankers with low correlation and high complementarity; (*ii*) find the most adequate way to measure the effectiveness of the evaluated rankers; (*iii*) estimate the effectiveness of rankers in unsupervised scenarios; (*iv*) fuse the selected rankers exploiting their complementarity and increasing the effectiveness results.

For fusion, there are two main categories of methods: early and late fusion [19], [20]. Early fusion consists in any kind of combination directly applied in the feature vectors, which is generally the primary structure obtained from a feature extractor. By contrast, late fusion approaches are done in a higher level of the retrieval process being performed in any type of structure obtained from the original feature vectors like similarity matrices or ranked lists.

In this work, three unsupervised methods are proposed for selection and fusion of ranked lists. Each of them consider different strategies for this task and belong to the late fusion category. The Unsupervised Selective Rank Fusion [2] (USRF) applies different pairwise measures (more specifically correlation and effectiveness estimations) in order to apply weights

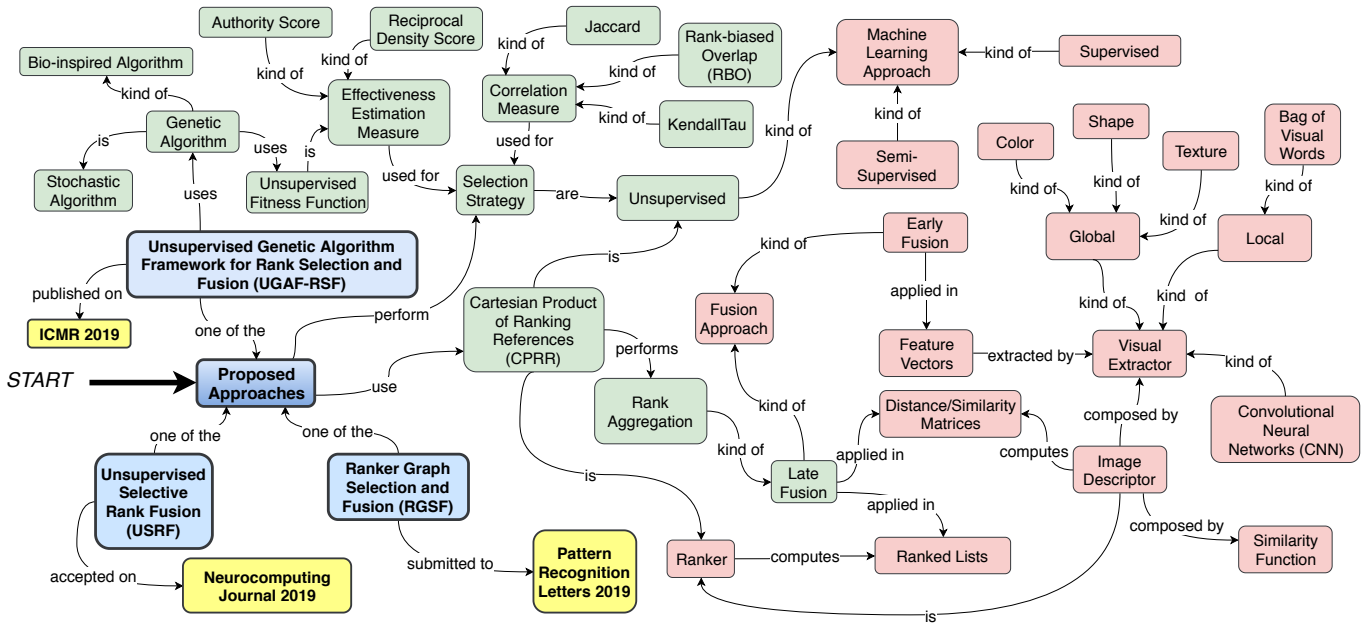

Fig. 1. Main concepts, contributions, and publications of this Master's work [2]–[4].

for pairs of rankers. From the intersection and union of pairs, the approach expands the idea for selecting combinations of any arbitrary size. Another proposed approach is the Ranker Graph Selection and Fusion [3] (RGSF) that uses a weighted graph where each node represents a ranker with a certain effectiveness estimation and the vertexes the correlation between them. The connected components are considered to represent the possible combinations that are candidates to be fused. One challenging task is to keep the diversity while selecting the best combination along multiple iterations. With this in mind, we also proposed an Unsupervised Genetic Algorithm Framework for Rank Selection and Fusion [4] (UGAF-RSF) which aims in selecting and fusing rankers through genetic algorithms by using an unsupervised fitness function.

Figure 1 presents a diagram that illustrates the different concepts and contributions involved in this work and the relations between them. In blue, the main methods proposed in this work; in green, the concepts used to address the contributions; in red, the other subjects related to the research; and in yellow, the associated publications. The proposed methods were evaluated in various public datasets, commonly used as benchmark in image retrieval tasks. The conducted experimental evaluation considered around 40 different features, including global, local and deep features. Significant effectiveness results were obtained, with gains up to +55% in relation to the highest effective isolated feature. The proposed methods were also compared to other state-of-the-art approaches, achieving comparable or superior results on different datasets.

The remaining of this paper is organized as follows: Section II addresses the image retrieval model and the problem definition. While Section III details the methods proposed and the main contributions of this work, Section IV shows and discusses the obtained results. Finally, Section V presents the conclusions, publications, and future work originated from this Master's work.

II. Image Retrieval Model and Problem Definition

This section formally defines the rank model used along the paper. Let $\mathcal{C}$=$\{x_1, x_2, \ldots, x_N\}$ be an image collection, where N denotes the collection size. Let us consider a retrieval task where, given a query image, returns a list of images from the collection $\mathcal{C}$.

Formally, given a query image x_q, a ranker denoted by R_j computes a ranked list τ_q=$(x_1, x_2, \ldots, x_k)$ in response to the query. The ranked list τ_q can be defined as a permutation of the k-neighborhood set $\mathcal{N}(q)$, which contains the k most similar images to image x_q in the collection $\mathcal{C}$. The permutation τ_q is a bijection from the set $\mathcal{N}(q)$ onto the set $[k] = \{1, 2, \ldots, k\}$. The $\tau_q(i)$ notation denotes the position (or rank) of image x_i in the ranked list τ_q.

The ranker R can be defined based on diverse approaches, including feature extraction or learning methods. In this paper, a feature-based approach [6] is considered, defining R as a tuple (ϵ, ρ), where $\epsilon : \mathcal{C} \rightarrow \mathbb{R}^d$ is a function that extracts a feature vector v_x from an image $x \in \mathcal{C}$; and d: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a distance function that computes the distance between two images according to their corresponding feature vectors. Formally, the distance between two images x_i, x_j is defined by $d(\epsilon(x_i), \epsilon(x_j))$. The notation $d(x_i, x_j)$ is used for readability purposes.

A ranked list can be computed by sorting images in a crescent order of distance. In terms of ranking positions we can say that, if image x_i is ranked before image x_j in the ranked list of image x_q, that is, $\tau_q(i) < \tau_q(j)$, then $d(q, i) \leq d(q, j)$. Taking every image in the collection as a query image x_q, a set of ranked lists $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ can be obtained.

Different features and distance functions give rise to different rankers which, in turn, produce distinct ranked lists. Let $\mathcal{R} = \{R_1, R_2, \ldots, R_m\}$ be a set of rankers and $R_j \in \mathcal{R}$, we denote by $\mathcal{T}_j$ the set of ranked lists produced by R_j. A ranked

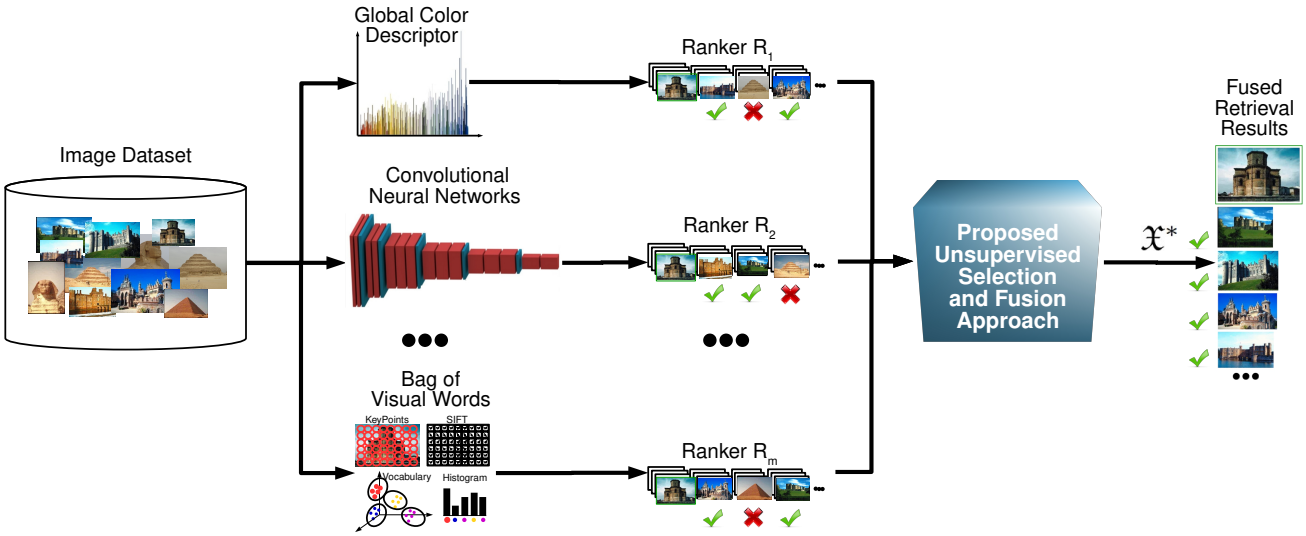

Fig. 2. The proposed workflow for rank selection and fusion in image retrieval tasks [2]–[4].

list computed by the ranker R_j in response to a query x_q is denoted by $\tau_{j,q}$.

The objective of the proposed methods is to select from the set $\mathcal{R}$ a sub-set of rankers which produces the most effective retrieval results, based on the analysis of their respective ranked lists, without the need of any labeled data. Formally, the framework can be defined by a function f_s as follows:

$$\mathfrak{X}^*_n = f_s(\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m), \tag{1}$$

where $\mathfrak{X}^*_n$ denotes the set of rankers selected by the framework for a given size n, such that $|\mathfrak{X}^*_n| = n$.

III. Proposed Methods

This work proposes three distinct methods for rank selection and fusion considering different strategies. Figure 2 illustrates the workflow with the steps that compose the proposed framework employed in this work. Given different rankers as input, provided by descriptors of different types, the proposed methods are applied in order to select and combine them in a completely unsupervised fashion. In other words, without the use of training labels or user intervention. Two different hypotheses were considered for structuring the selection strategy in the proposed methods:

1) **effectiveness:** the more effective the rankers being combined, more effective the fusion result obtained;
2) **diversity:** if two rankers are low correlated, they probably have high complementarity between their results.

To assure the effectiveness (first hypothesis), different effectiveness estimation measures were used as the Authority Score [21] and the Reciprocal Density [22]. For the diversity, different strategies were used, some approaches applied correlation measures (e.g. Jaccard, Kendallτ, Spearman). In all the methods, a fusion step is performed using an unsupervised fusion approach. The proposed approaches are very flexible, allowing the use of different methods for this task. In this work, we are using the CPRR [12] because the method is recent and presents results comparable to the state-of-the-art.

The remaining of this section is organized as follows: Section III-A presents a method based on pairwise measures, Section III-B presents a method based on graphs and connected components, Section III-C presents a method that proposes a genetic algorithm framework.

A. Selection through Pairwise Measures

The Unsupervised Selective Rank Fusion [2] (USRF) relies on the idea of assigning estimation scores for pairs of rankers. This score is formulated from two different equations, which were designed based on the hypotheses earlier described. Figure 3 presents the USRF workflow that illustrates the main steps of the method. Each of theses steps are described in the following subsections in their respective order: selection by effectiveness is described in Section III-A1, selection by correlation in Section III-A2, and the joint selection measure in Section III-A3. Finally, Section III-A4 describes the fusion approach applied.

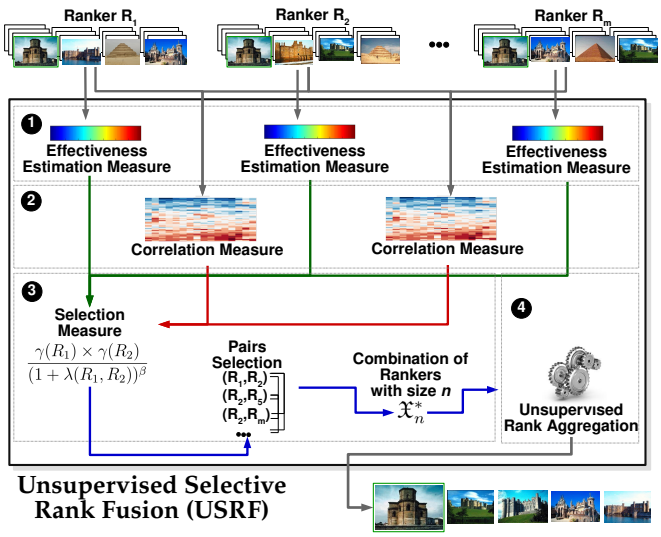

Fig. 3. Illustration of the USRF iterations [2].

1) Selection through Effectiveness Estimation: Considering the assumption that relevant combinations are given by rankers of high effectiveness, let Γ be the effectiveness selection measure applied to a pair $\{R_1, R_2\}$ and $\gamma(R_i)$ be a function that returns the effectiveness estimation (e.g. Authority Score [21], Reciprocal Density [22]) of ranked lists provided by the ranker R_i. The measure for selection through effectiveness estimation is defined as:

$$\Gamma(R_1, R_2) = \gamma(R_1) \times \gamma(R_2). \tag{2}$$

2) Selection through Correlation: Considering the assumption that rankers with low correlation provide a potential way for exploiting the complementarity of the data, let Λ be a correlation function (e.g. Jaccard, Spearman, Goodman) and $\lambda(R_1, R_2)$ be a function that returns a value of correlation (similarity in the interval $[0, 1]$) between the ranked lists offered by the rankers R_1 and R_2. The measure for selection through correlation is defined by the Equation 3.

$$\Lambda(R_1, R_2) = \frac{1}{1 + \lambda(R_1, R_2)}. \tag{3}$$

3) Joint Selection Measure: Our proposed selection measure employed by the USRF is based on the incorporation of the two earlier equations, with the objective of selecting pairs of high effectiveness and low correlation (high complementarity). The measure w_p is defined as:

$$\begin{aligned} w_p(\{R_1, R_2\}) &= \Gamma(R_1, R_2) \times \Lambda(R_1, R_2)^{\beta} \\ &= \frac{\gamma(R_1) \times \gamma(R_2)}{(1 + \lambda(R_1, R_2))^{\beta}}, \end{aligned} \tag{4}$$

where the exponent β is used with the intent of applying a weight for the correlation, which allows the use of the proposed measure in different scenarios.

One important remark is that, since the employed measure computes estimations only for pairs of rankers, we consider the intersection and union of combinations to select combinations of any arbitrary size.

4) Fusion Approach: After selecting the combination of rankers (which is denoted by $\mathfrak{X}^*$), the CPRR [12] is applied in order to fuse the rankers into a single new ranker that is expected to be more effective than the isolated descriptors provided.

B. Selection through Graphs and Connected Components

The Ranker Graph Selection and Fusion [3] (RGSF) is a method that consists in the use of a weighted graph for selecting the combinations of rankers. In this graph, each vertex corresponds to a ranker and the edge between them is weighted by a correlation measure. The vertexes are also weighted according to an effectiveness estimation measure associated to each ranker. Different from the USRF, the RGSF does not require a separate iteration for each size of combination, being more flexible.

Figure 4 presents the workflow of the RGSF method. In (1) the graph is built along the iterations considering thresholds that gradually decrease. The connected components (CC) that appear along the iterations are used as the combinations to be fused. Each selected combination is denoted as $\mathfrak{X}_i$ where i denotes an index. In this method, each CC corresponds to a different combination, in such way that, each ranker that belongs to it is part of the combination.

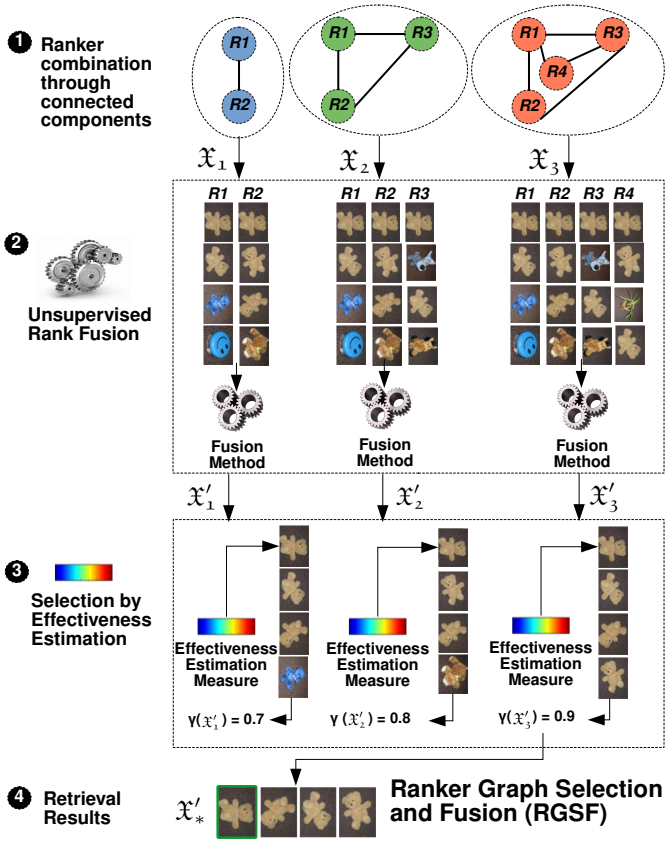

Fig. 4. Illustration of the RGSF execution steps [3].

Step (2) considers an unsupervised post-processing method (CPPR, in this case) that performs a rank-aggregation procedure for the input rankers and outputs a new set of ranking results for each combination, which is expected to be more effective than the isolated ones. For this method, we adopt the $\mathfrak{X}'_i$ notation to differentiate the combination before and after the fusion stage.

For all the executions performed, the effectiveness estimation measure (which is unsupervised) computes a score for each output ranker ($\mathfrak{X}'_i$) in (3). Finally, the $\mathfrak{X}'_i$ with the highest score is provided as output in (4), being denoted as $\mathfrak{X}'^*$.

C. Selection through Genetic Algorithm

The Unsupervised Genetic Algorithm Framework for Rank Selection and Fusion [4] (UGAF-RSF) employs a genetic algorithm approach to address the task of selecting and fusing rankers. The idea is that the most effective rankers are kept along the generations considering an unsupervised fitness measure. Different from the other methods proposed in this work, which consider correlation measures as a strategy to ensure the diversity and complementarity, this approach relies on the randomness of the genetic algorithm. Figure 5 illustrates

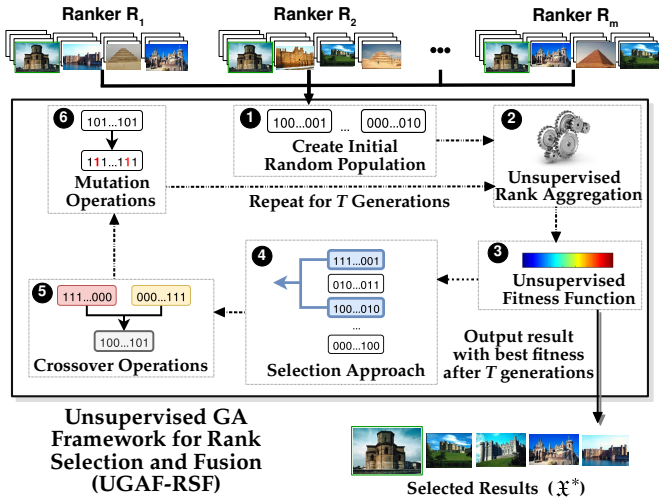

Fig. 5. Proposed framework for the genetic algorithm approach [4].

the workflow with the main steps of the framework, each of them are detailed as follows:

1) each chromosome is represented as a sequence of bits where each bit indicates whether or not a ranker is part of a combination. A random population is generated, where all the rankers have the same probability of being set to a combination;
2) a fusion step is executed for each of the chromosomes of the population considering the CPRR method (as previously described);
3) the unsupervised fitness function is applied in the fusion result of each of the chromosomes. In this framework, we considered a hybrid estimation measure as fitness function, which consists in the multiplication of Authority Score [21] and Reciprocal Density [22] ($Hybrid = Authority \times Reciprocal$);
4) a tournament selection (considering three individuals per tournament) is performed to choose the best individuals for the next stages. For this step, we also consider a small elitism rate in order to decrease the number of iterations required for the algorithm convergence;
5) each individual has a probability (crossover rate) of being used for crossover. Given two parent chromosomes, each bit of the child chromosome has the same probability of belonging to either parent. This type of crossover is called uniform crossover. We use this technique because it avoids drastic changes in the new individuals;
6) mutation operations are applied for all individuals, where each bit has a probability (mutation rate) of being flipped. Such operation is important to keep the diversity of the population along the generations.

IV. Results

This section presents a brief overview of the results obtained in this work. Our approaches were compared with other state-of-the-art methods on the Holidays and UKBench datasets, presented in Tables I and II, respectively. Notice that our approaches present effectiveness results superior or comparable to the baselines in the majority of the scenarios.

TABLE I
State-of-the-art on the Holidays [23] dataset (MAP).

MAP for state-of-the-art methods				
Jégou *et al.* [23]	Tolias *et al.* [24]	Paulin *et al.* [25]	Qin *et al.* [26]	Zheng *et al.* [27]
75.07%	82.20%	82.90%	84.40%	85.20%
Sun *et al.* [28]	Zheng *et al.* [29]	Pedronette et al. [30]	Li *et al.* [31]	Liu et al. [32]
85.50%	85.80%	86.19%	89.20%	90.89%

Proposed Methods		
USRF [2]	**RGSF [3]**	**UGAF-RSF [4]**
90.51%	**89.56%**	**86.99% ± 1.1297**

TABLE II
State-of-the-art on the UKBench [33] dataset (N-S Score).

N-S Score for state-of-the-art methods					
Zheng et al. [34]	Wang et al. [35]	Sun et al. [28]	Paulin *et al.* [25]	Zhang et al. [18]	Zheng et al. [17]
3.57	3.68	3.76	3.76	3.83	3.84
Bai et al. [36]	Xie et al. [37]	Liu et al. [32]	Pedronette et al. [30]	Bai et al. [38]	
3.86	3.89	3.92	3.93	3.94	

Proposed Methods		
USRF [2]	**RGSF [3]**	**UGAF-RSF [4]**
3.94	**3.79**	**3.95 ± 0.3206**

Figure 6 presents examples of visual results for three different queries on the UKBench dataset. Each row shows the results for the queries that were selected as part of the combination. In order to facilitate the visualization, query images are shown with green borders and incorrect results with red borders. The last row shows the fusion result. For this example, it can be noticed that the fused result has solved the issues of the isolated rankers, which indicates that the selection was effective and the complementarity between the results was correctly exploited in the fusion stage.

The methods were also evaluated considering their efficiency. Table III presents the execution time for each method and dataset. Notice that RGSF is the faster and UGAF-RSF is the slowest among the three. This is due to the number of fusions that the genetic algorithm requires to compute along the generations. Besides the slower times for the UGAF-RSF, it is still much faster than computing all the possible combinations.

TABLE III
Execution times for the proposed methods.

Image Dataset	Mean Execution Time (seconds)		
	USRF [2]	**RGSF [3]**	**UGAF-RSF [4]**
Flowers [39]	1094.56 ± 6.81	752.16 ± 4.81	1871.24 ± 62.39
Corel5k [40]	8116.09 ± 29.79	3295.16 ± 6.05	46539.31 ± 502.22
UKBench [33]	882.95 ± 10.97	247.67 ± 1.71	6685.59 ± 336.61
Holidays [23]	417.49 ± 0.85	104.478 ± 0.40	6559.08 ± 230.11

For a broad view of the obtained results in each case, Table IV introduces a joint analysis that considers both ef-

Fig. 6. Three query examples illustrating the impact of selection and combination on the UKBench dataset [2]–[4].

fectiveness and efficiency. The methods are ranked according to their results in each dataset. In general, USRF and RGSF are comparable being among the best ranked in terms of effectiveness. Although, regarding efficiency, RGSF is the faster in all the considered scenarios. Considering both aspects, in general, USRF seems to be the most balanced choice. However, the ideal method still depends on the user requirements and constraints for each case.

TABLE IV
EFFECTIVENESS AND EFFICIENCY ANALYSIS OF THE PROPOSED METHODS.

Evaluated Criteria	Rank		
	1st	2nd	3rd
Effectiveness Evaluation			
Best MAP on Flowers [39]	USRF	RGSF	UGAF-RSF
Best MAP on Corel5k [40]	UGAF-RSF	USRF	RGSF
Best MAP on UKBench [33]	UGAF-RSF	USRF	RGSF
Best MAP on Holidays [23]	USRF	RGSF	UGAF-RSF
Efficiency Evaluation			
Exec. Time on Flowers [39]	RGSF	USRF	UGAF-RSF
Exec. Time on Corel5k [40]	RGSF	USRF	UGAF-RSF
Exec. Time on UKBench [33]	RGSF	USRF	UGAF-RSF
Exec. Time on Holidays [23]	RGSF	USRF	UGAF-RSF

V. CONCLUSIONS, PUBLICATIONS, AND FUTURE WORK

In this work, we have presented three different unsupervised approaches for selecting and fusing different rankers. The methods were evaluated on image retrieval datasets considering rankers provided by diverse descriptors covering different categories (global, local, and deep learning). We achieved results comparable or better to the state-of-the-art in the majority of the cases.

This paper focuses on the results and methods proposed in [2]–[4], which composes the main body of the Master's dissertation [1]. In addition, the Master's work was concluded in two years, resulting in various contributions and eight papers published or submitted along the period:

- four international conference papers [4], [41]–[43];
- three international journal papers [2], [44], [45];
- one submission to an international journal paper [3].

Most of the publications have the student as the first author, except [42] and [45], which are collaborations. While [42] evaluates late fusion rank aggregation approaches for video interestingness prediction, [45] proposes an unsupervised manifold ranking algorithm based on hypergraphs for multimedia retrieval tasks. An extension of one of the student's undergraduate work is presented in [41], which exploits the use of supervised measures to select the best rankers. A semi-supervised classifier based on correlation graphs and majority vote strategies was proposed in [43], originated from the Master's Artificial Intelligence course. In [44], different kNN sets and correlation measures were exploited for re-ranking and rank fusion tasks.

As future work, there are multiple research topics that can be exploited: (*i*) use and evaluation of other methods for the fusion stage; (*ii*) development of new estimation measures based on different strategies, not only in the analysis of the reciprocal neighborhood of the ranked lists; (*iii*) investigation of effectiveness estimation measures as fitness functions in genetic programming scenarios, where the chromosomes are represented as trees; (*iv*) there is also the possibility of applying the proposed methods in diverse and multimodal retrieval scenarios (e.g. sound, video, text).

ACKNOWLEDGMENT

The authors are grateful to the Coordination for the Improvement of Higher Education Personnel - CAPES, São Paulo Research Foundation - FAPESP (grant #2017/02091-4, #2018/15597-6, and #2017/25908-6), the Brazilian National Council for Scientific and Technological Development - CNPq (grant #308194/2017-9), and Petrobras (grant #2017/00285-6).

REFERENCES

[1] L. P. Valem and D. C. G. Pedronette, "Combinação Seletiva Não Supervisionada de Listas Ranqueadas Aplicada à Busca de Imagens pelo Conteúdo." 2019, Dissertation (M.Sc. in Computer Science), UNESP (Universidade Estadual Paulista Júlio de Mesquita Filho), Rio Claro, São Paulo, Brazil.

[2] ——, "Unsupervised selective rank fusion for image retrieval tasks," *Neurocomputing*, 2019 (Accept with Minor Revision).

[3] ——, "Graph-based selective rank fusion for unsupervised image retrieval," *Pattern Recognition Letters*, 2019 (Submitted).

[4] ——, "An unsupervised genetic algorithm framework for rank selection and fusion on image retrieval," in *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, ser. ICMR '19. New York, NY, USA: ACM, 2019, pp. 58–62.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 5:1–5:60, 2008.

[6] R. D. S. Torres and A. X. Falcão, "Content-based image retrieval: Theory and applications," *Revista de Informática Teórica e Aplicada*, vol. 13, pp. 161–185, 2006.

[7] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: A comprehensive overview," vol. 37, no. Supplement C, 2017, pp. 50 – 60.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12, 2012, pp. 1097–1105.

[9] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[10] F. F. Faria, A. Veloso, H. M. Almeida, E. Valle, R. d. S. Torres, M. A. Gonçalves, and W. Meira, Jr., "Learning to rank for content-based image retrieval," in *Proceedings of the International Conference on Multimedia Information Retrieval*, ser. MIR '10, 2010, pp. 285–294.

[11] L. P. Valem and D. C. G. Pedronette, "Unsupervised selective rank fusion for image retrieval tasks," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '17. New York, NY, USA: ACM, 2017, pp. 107–111.

[12] L. P. Valem, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through Cartesian product of ranking references," *Pattern Recognition Letters*, vol. 114, pp. 41 – 52, 2018.

[13] J. A. Vargas Muñoz, R. da Silva Torres, and M. A. Gonçalves, "A soft computing approach for learning to aggregate rankings," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, ser. CIKM '15, 2015, pp. 83–92.

[14] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proceedings of the 18th International Conference on Neural Information Processing Systems*, ser. NIPS'05, 2005, pp. 507–514.

[15] Z. Zhao and H. Liu, "Spectral feature selection for supervised and unsupervised learning," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07, 2007, pp. 1151–1157.

[16] D. C. G. Pedronette and R. da S. Torres, "A correlation graph approach for unsupervised manifold learning in image retrieval tasks," *Neurocomputing*, vol. 208, no. Sup C, pp. 66 – 79, 2016.

[17] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *CVPR*, 2015.

[18] S. Zhang, M. Yang, T. Cour, K. Yu, and D. Metaxas, "Query specific rank fusion for image retrieval," *IEEE TPAMI*, vol. 37, no. 4, pp. 803–815, April 2015.

[19] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '05, 2005, pp. 399–402.

[20] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: a survey," *Multimedia Systems*, vol. 16, no. 6, pp. 345–379, Nov 2010.

[21] 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 and Vision Computing*, vol. 32, no. 2, pp. 120 – 130, 2014.

[22] D. C. G. Pedronette and R. d. S. Torres, "Unsupervised effectiveness estimation for image retrieval using reciprocal rank information," in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, 2015, pp. 321–328.

[23] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *European Conference on Computer Vision*, ser. ECCV '08, 2008, pp. 304–317.

[24] G. Tolias, Y. Avrithis, and H. Jégou, "To aggregate or not to aggregate: Selective match kernels for image search," in *IEEE International Conference on Computer Vision (ICCV'2013)*, Dec 2013, pp. 1401–1408.

[25] M. Paulin, J. Mairal, M. Douze, Z. Harchaoui, F. Perronnin, and C. Schmid, "Convolutional patch representations for image retrieval: An unsupervised approach," *Int. Journal of Computer Vision*, 2017.

[26] D. Qin, C. Wengert, and L. V. Gool, "Query adaptive similarity for large scale object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2013)*, June 2013, pp. 1610–1617.

[27] L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image retrieval," *IEEE Transactions on Image Processing (TIP)*, vol. 23, no. 8, pp. 3368–3380, 2014.

[28] S. Sun, Y. Li, W. Zhou, Q. Tian, and H. Li, "Local residual similarity for image re-ranking," *Information Sciences*, vol. 417, no. Sup. C, pp. 143 – 153, 2017.

[29] L. Zheng, S. Wang, Z. Liu, and Q. Tian, "Packing and padding: Coupled multi-index for accurate image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2014)*, June 2014, pp. 1947–1954.

[30] D. C. G. Pedronette, F. M. F. Gonçalves, and I. R. Guilherme, "Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks," *Pattern Recognition*, vol. 75, pp. 161 – 174, 2018.

[31] X. Li, M. Larson, and A. Hanjalic, "Pairwise geometric matching for large-scale object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2015)*, June 2015, pp. 5153–5161.

[32] Z. Liu, S. Wang, L. Zheng, and Q. Tian, "Robust imagegraph: Rank-level feature fusion for image search," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3128–3141, 2017.

[33] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2006)*, vol. 2, 2006, pp. 2161–2168.

[34] L. Zheng, S. Wang, and Q. Tian, "Lp-norm idf for scalable image retrieval," *IEEE TIP*, vol. 23, no. 8, pp. 3604–3617, Aug 2014.

[35] B. Wang, J. Jiang, W. Wang, Z.-H. Zhou, and Z. Tu, "Unsupervised metric fusion by cross diffusion," in *CVPR*, 2012, pp. 3013 –3020.

[36] S. Bai and X. Bai, "Sparse contextual activation for efficient visual re-ranking," *IEEE Trans. on Image Processing (TIP)*, vol. 25, no. 3, pp. 1056–1069, 2016.

[37] L. Xie, R. Hong, B. Zhang, and Q. Tian, "Image classification and retrieval are one," in *ACM ICMR'2015*, 2015, pp. 3–10.

[38] S. Bai, X. Bai, Q. Tian, and L. J. Latecki, "Regularized diffusion process for visual retrieval," in *Conf. on Artificial Intelligence (AAAI)*, 2017, pp. 3967–3973.

[39] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1447–1454.

[40] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188 – 198, 2013.

[41] L. P. Valem and D. C. G. Pedronette, "Selection and combination of unsupervised learning methods for image retrieval," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, ser. CBMI '17, 2017, pp. 27:1–27:6.

[42] J. Almeida, L. P. Valem, and D. C. G. Pedronette, "A rank aggregation framework for video interestingness prediction," in *Proceedings of the 9th International Conference on Image Analysis and Processing*, ser. ICIAP '17, 2017.

[43] L. P. Valem, D. C. G. Pedronette, F. Breve, and I. R. Guilherme, "Manifold correlation graph for semi-supervised learning," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–7.

[44] L. P. Valem, C. R. D. Oliveira, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through rank correlation and knn sets," in *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 4. New York, NY, USA: ACM, Oct. 2018, pp. 80:1–80:23.

[45] D. C. G. Pedronette, L. P. Valem, J. Almeida, and R. da Silva Torres, "Multimedia retrieval through unsupervised hypergraph-based manifold ranking," *IEEE Transactions on Image Processing (TIP)*, 2019, On-line, to appear.