

# Unsupervised Selective Rank Fusion for Content-based Image Retrieval

Lucas Pascotti Valem<sup>1</sup>, Daniel Carlos Guimarães Pedronette<sup>1</sup> (Advisor)

<sup>1</sup> Department of Statistics, Applied Math. and Computing  
São Paulo State University (UNESP)  
Rio Claro, SP - Brazil

{lucas.valem,daniel.pedronette}@unesp.br

***Abstract.** 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 features and machine learning methods. 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, especially when no training data is available. 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 scenarios.*

## 1. 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 for encoding and retrieving images based on their visual properties. In fact, a myriad of different descriptors are available, often providing distinct and complementary results even for a same query. 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. 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, learned features based on deep convolutional neural networks have been achieving very promising results in many diverse scenarios. Most of the networks are often trained in the ImageNet dataset 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, or even different network architectures.

In the recent years, different methods have been proposed with the intent of fusing results from distinct visual features. 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. Despite the fact that fusion strategies in unsupervised scenarios have achieved significant gains in terms of effectiveness, 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*, once there is no dynamic methodology for selecting the visual descriptors to be combined.

We investigated and proposed novel methods based on unsupervised learning approaches for dynamic selection and combination of different visual descriptors, without the need of any training data or user intervention. The diversity of results obtained by distinct descriptors or retrieval systems is represented through a rank model, assuming 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 a 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.

In this Master's work, three unsupervised methods are proposed for selection and fusion of ranked lists. Each of them consider different strategies for this task, being more suitable for each scenario. 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.

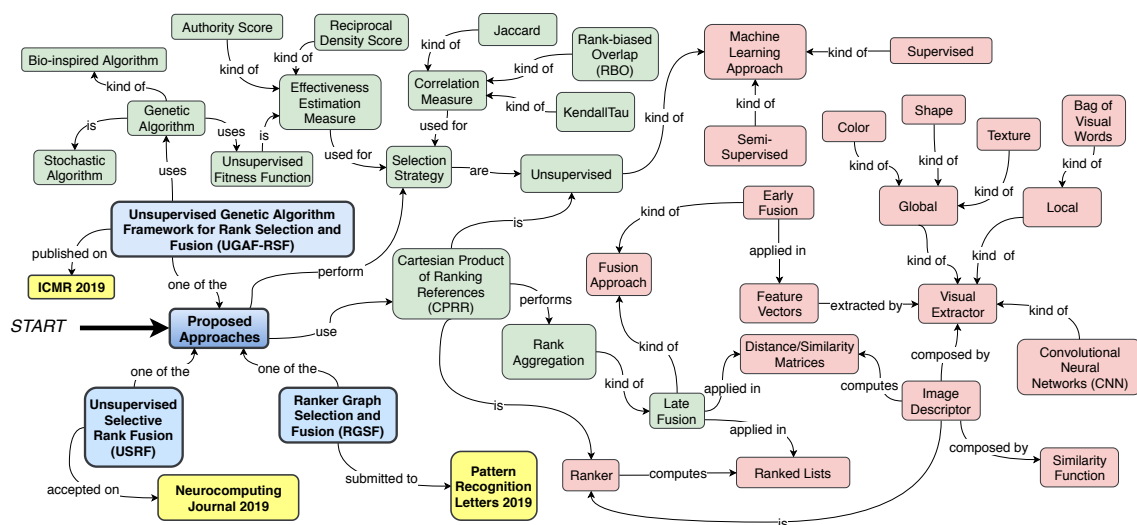
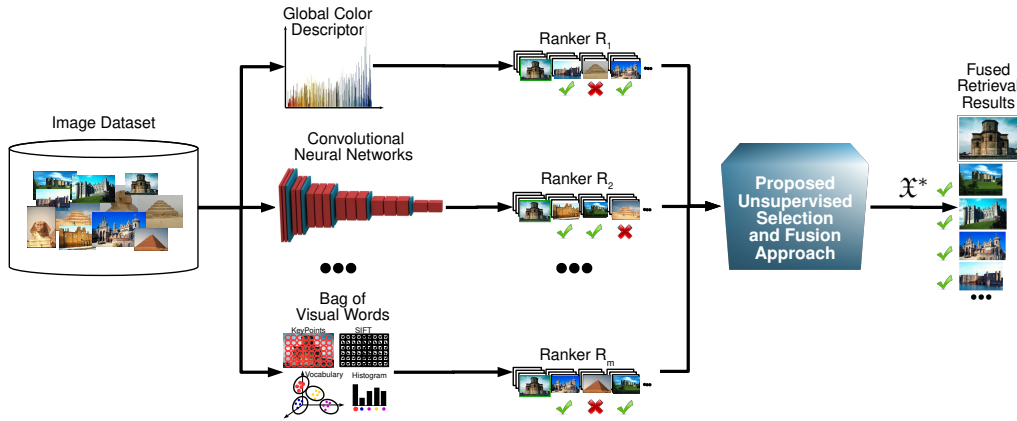


Figure 1. Main concepts, contributions, and publications of this Master's work.

## 2. Main Contributions

This work proposes unsupervised methods for rank selection and fusion. 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.



**Figure 2. Proposed workflow for rank selection and fusion in image retrieval.**

The three proposed methods employ distinct strategies, which are described in the next sub-sections.

### 2.1. Selection through Pairwise Measures

The Unsupervised Selective Rank Fusion [Valem and Pedronette 2020b] (USRF) relies on the idea of assigning estimation scores for pairs of rankers. It applies different pairwise measures (more specifically correlation and effectiveness estimations) in order to apply weights for pairs of rankers. From the intersection and union of pairs, the approach expands the idea for selecting combinations of any arbitrary size.

### 2.2. Selection through Graphs and Connected Components

The Ranker Graph Selection and Fusion [Valem and Pedronette 2020a] (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. Figure 3 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  $\mathcal{X}_i$  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.

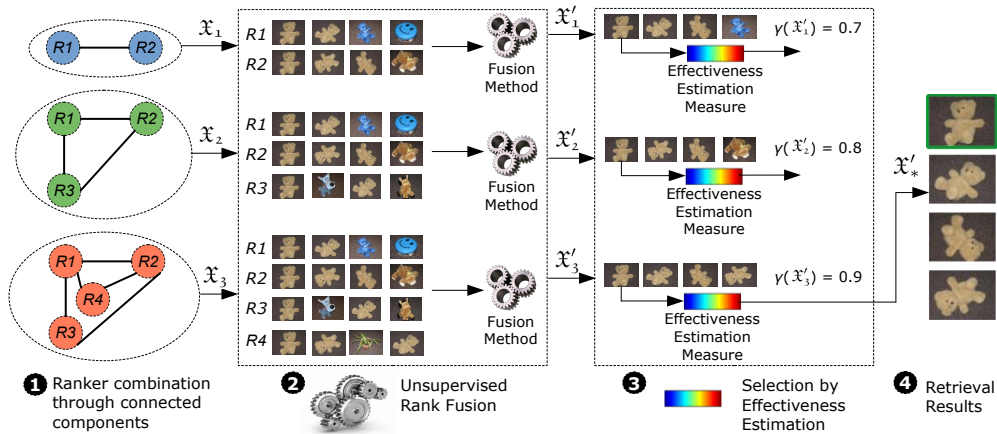
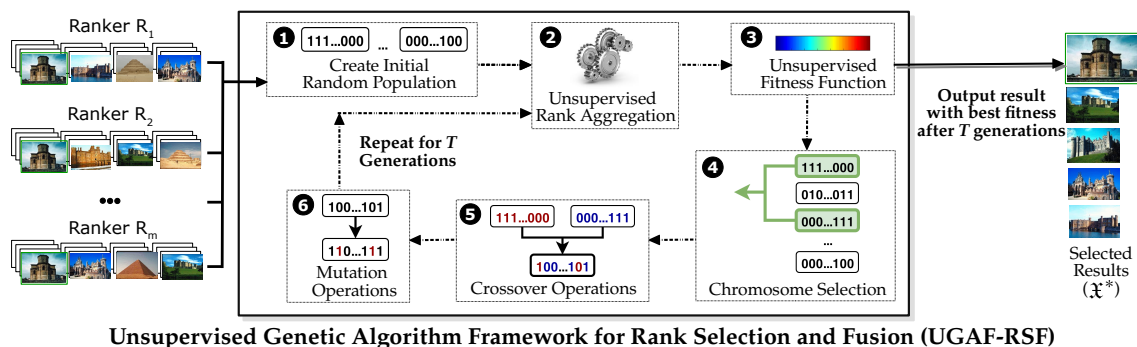


Figure 3. Illustration of the RGSF execution steps.

### 2.3. Selection through Genetic Algorithm

The Unsupervised Genetic Algorithm Framework for Rank Selection and Fusion [Valem and Pedronette 2019c] (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 4 illustrates the workflow with the main steps of the framework.



Unsupervised Genetic Algorithm Framework for Rank Selection and Fusion (UGAF-RSF)

Figure 4. Proposed framework for the genetic algorithm approach.

## 3. Results, Conclusions, 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. Figure 5 presents a visual example, where 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, which indicates that the fused result has solved the issues of the isolated rankers.



**Figure 5. Impact of the selection and combination of our approach.**

Among the contributions, this text focuses on the results and methods proposed in [Valem and Pedronette 2020b], [Valem and Pedronette 2020a], and [Valem and Pedronette 2019c], which compose the main body of the Master's dissertation [Valem and Pedronette 2019b]. In addition, the Master's work was concluded in two years, resulting in various contributions and nine papers published along the period. The publications are listed below with the respective *Qualis* (old/new).

*Journal Papers:*

- **Pattern Recognition Letters** (*Qualis A1/A2*) [Valem and Pedronette 2020a];
- **Neurocomputing** (*Qualis A1/A2*) [Valem and Pedronette 2020b];
- **IEEE Transactions on Image Processing** (*Qualis A1/A1*) [Pedronette et al. 2019];
- **ACM TOMM** (*Qualis A1/A2*) [Valem et al. 2018].

*Conference Papers:*

- **SIBGRAPI WTD** (*Qualis B1/A4*) [Valem and Pedronette 2019a];
- **ICMR** (*Qualis A1/A2*) [Valem and Pedronette 2019c];
- **IJCNN** (*Qualis A1/A2*) [Valem et al. 2018];
- **CBMI** (*Qualis B1/B1*) [Valem and Pedronette 2017];
- **ICIAP** (*Qualis B1/A3*) [Almeida et al. 2017].

Most of the publications have the student as the first author, except [Almeida et al. 2017] and [Pedronette et al. 2019], which are collaborations. While [Almeida et al. 2017] evaluates rank aggregation approaches for video interestingness prediction, [Pedronette et al. 2019] 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 [Valem and Pedronette 2017], 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 [Valem et al. 2018], originated from the Master's Artificial Intelligence course. In [Valem et al. 2018], different kNN sets and correlation measures were exploited for re-ranking and rank fusion tasks. This work also received the **Best Master's Thesis Award in the Workshop of Theses and Dissertations (WTD)** of the Conference on Graphics, Patterns and Images (SIBGRAPI 2019) [Valem and Pedronette 2019a], one of the most important conferences in this research field.

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 multi-modal retrieval scenarios (e.g. sound, video, text).

## Acknowledgment

Authors thank CAPES, FAPESP (grant #2017/02091-4, #2018/15597-6, and #2017/25908-6), CNPq (grant #308194/2017-9), and Petrobras (grant #2017/00285-6).

## References

- Almeida, J., Valem, L. P., and Pedronette, D. C. G. (2017). A rank aggregation framework for video interestingness prediction. In *Image Analysis and Processing - ICIAP 2017*, pages 3–14. Springer International Publishing.
- Pedronette, D. C. G., Valem, L. P., Almeida, J., and da Silva Torres, R. (2019). Multimedia retrieval through unsupervised hypergraph-based manifold ranking. *IEEE Transactions on Image Processing*, 28(12):5824–5838.
- Valem, L. and Pedronette, D. (2019a). Unsupervised selective rank fusion on content-based image retrieval. In *Anais Estendidos da XXXII Conference on Graphics, Patterns and Images*, pages 63–69, Porto Alegre, RS, Brasil. SBC.
- Valem, L. P., Oliveira, C. R. D., Pedronette, D. C. G., and Almeida, J. (2018). Unsupervised similarity learning through rank correlation and knn sets. In *ACM Trans. Multimedia Comput. Commun. Appl.*, volume 14, pages 80:1–80:23, New York, NY, USA. ACM.
- Valem, L. P. and Pedronette, D. C. G. (2017). Selection and combination of unsupervised learning methods for image retrieval. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, CBMI '17*, pages 27:1–27:6.
- Valem, L. P. and Pedronette, D. C. G. (2019b). Combinação Seletiva Não Supervisionada de Listas Ranqueadas Aplicada à Busca de Imagens pelo Conteúdo. Dissertation (M.Sc. in Computer Science), UNESP (Universidade Estadual Paulista Júlio de Mesquita Filho), Rio Claro, São Paulo, Brazil.
- Valem, L. P. and Pedronette, D. C. G. (2019c). An unsupervised genetic algorithm framework for rank selection and fusion on image retrieval. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR '19*, pages 58–62, New York, NY, USA. ACM.
- Valem, L. P. and Pedronette, D. C. G. (2020a). Graph-based selective rank fusion for unsupervised image retrieval. *Pattern Recognition Letters*, 135:82 – 89.
- Valem, L. P. and Pedronette, D. C. G. (2020b). Unsupervised selective rank fusion for image retrieval tasks. *Neurocomputing*, 377:182 – 199.
- Valem, L. P., Pedronette, D. C. G., Breve, F., and Guilherme, I. R. (2018). Manifold correlation graph for semi-supervised learning. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7.