Image Representation Learning by Color Quantization Optimization

Érico Marco D. A. Pereira, Jefersson A. dos Santos
Department of Computer Science
Federal University of Minas Gerais
31270-901 - Belo Horizonte - MG - Brazil
Emails: emarco.pereira@dcc.ufmg.br, jefersson@dcc.ufmg.br

***Abstract*—The state-of-art methods of representation learning, based on Deep Neural Networks, present serious drawbacks regarding usage complexity and resources consumption, leaving space for simpler alternatives. We proposed two approaches of a Representation Learning method which aims to provide more effective and compact image representations by optimizing the colour quantization for the image domain. Our hypothesis is that changes in the quantization affect the description quality of the features enabling representation improvements. We evaluated the method performing experiments for the task of Content-Based Image Retrieval on eight known datasets. The results showed that the first approach, focused on representation effectiveness, produced representations that outperforms the baseline in all the tested scenarios. And the second, focused on compactness, was able to produce superior results maintaining or even reducing the dimensionality and representations until 25% smaller that presented statistically equivalent performance.**

***Keywords*—Representation Learning; Color Quantization; CBIR; Genetic Algorithm; Feature Extraction;**

I. Introduction

It is known that the way data is represented can highly influence the performance of machine learning methods in visual pattern recognition tasks, such as Content-Based Image Retrieval [1], Object Detection [2], Remote Sensing [3] and Image Classification [4]. That being said, Representation Learning [5], which consists on the process of using pattern recognition algorithms to find representations optimized for a given data domain and/or task at focus, has become a tendency.

The current state-of-art methods for representation learning, which are based on Deep Learning [6] techniques, in many cases present a considerable gain in description effectiveness. However, the use of these methods presents serious drawbacks, such as the difficulties in properly exploring its range of parameters and possible architectures, the superior computational time spent its training and the big amount of data required to produce efficient learning models, leaving space for alternatives.

The possible methodologies for representation learning can be classified into two main approaches: those that learn representations from a feature set provided by a hand-crafted extractor and those that completely compose new ones without any prior feature extraction. Following the later approach, complex multi-layered learning processes as the ones executed by deep neural networks are not always needed in order to produce representative features. Depending on the scenario, the improvement of existent representations is already enough to fairly solve the task.

Few years ago, before the arising of Deep Neural Networks, hand-crafted feature extractors were used in order to compose image representations [7]–[9]. Among them, the BIC [10] achieved prominent results, being in several cases more effective than all its competitors but the Deep Nets [11], [12] and also faster at computing representations. This behaviour states BIC features as promising candidates for undergoing a feature leaning process and providing good results.

Examining the extraction procedure of this method, a fact comes to observation: it uses a fixed RGB colour-space uniformly quantized in 4 tonalities for each axis. According to Stehling et. al [13], this configuration was chosen due the achievement of good results in a majority of tested scenarios and the compatibility with other feature extraction methods which rely on the same colour scheme. However, it raises the question whether a different quantization could provide better representations.

The use of a different colour quantization, specially one adapted to the current image domain instead of a predefined one, could allow the enhancement of convenient image features and the suppression of others. Since the representations are based on colour histograms, the enhancement and detailing of colours that favour the closeness of similar images and the distinction of different ones, according to the task criteria, would provide the composition of more representative features and, consequently, improvements on the task performance. Furthermore, a domain-oriented quantization allows the discard of the less contributing tonalities resulting in a possible reduction of the representation size.

This work proposes an approach of representation learning in order to improve the description effectiveness of an existent feature extractor by exploring a particular characteristic of the current image context, its colour distribution. Our hypothesis is that changing the colour quantization affects the description quality of the features in the sense that a tonality configuration optimized for a given domain could produce more effective and compact image representations.

II. Related Work

Representation Learning: In the last decade, several feature learning techniques were developed for raw image data [14]–

[20]. Approaches regarding deep belief nets [14], denoising autoencoders [15], deep Boltzmann machines [16], convolutional deep belief networks [17], K-Means based feature learning [18], hierarchical matching pursuit [19] and sparse coding [20] address this purpose.

Border Interior Classification (BIC): BIC is a simple and fast feature extractor which computes a image representation composed by two colour histograms: one for border pixels and other for interior pixels. This pixel-wise classification occurs according to a 4-pixel neighbourhood criteria: when the four immediate neighbours (right, left, top, bottom) present the exact same colour as the pixel in analysis, it is labelled as interior, otherwise, as border. At the end of the computation process, the histograms undergo two normalizations: division by the maximum value, for image dimension invariance, and a transformation according to a discrete logarithmic function, aiming to smooth major discrepancies. This algorithm obtained good results in previous works for web image retrieval [7] and for remote sensing image classification [8].

Colour Quantization: Some works developed quantization leaning using evolutive heuristics for Image Segmentation [21]. Scheunders [22] treats the quantization problem as global image segmentation and proposes an optimal mean squared quantizer and a hybrid technique combining optimal quantization with a Genetic Algorithm modelling [23]. Further, the same author [22] presents a genetic c-means clustering algorithm (GCMA), which is a hybrid technique combining the c-means clustering algorithm (CMA) with Genetic Algorithm. Lastly, Omran et al. [24] developed colour image quantization algorithm based on a combination of Particle Swarm Optimization (PSO) and K-means clustering.

Regarding the effects of colour quantization on image representations, Ponti et al. [25] approached the colour quantization as a pre-processing step of feature extraction. They applied four different quantization models over three image datasets and showed that distinct quantizations can produce very different results in terms of accuracy and dimensionality.

III. Methodology

In order to learn a colour quantization optimized for the given image domain we proposed the method described by the process presented on Figure 1. The main phases of the method are described as follows.

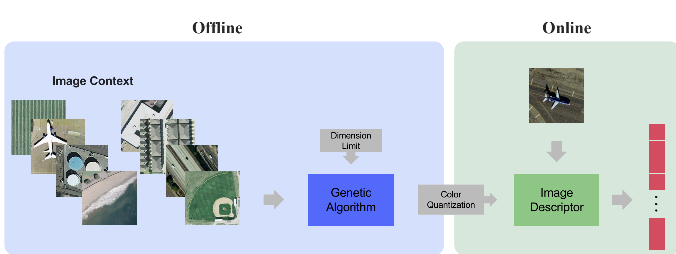


Fig. 1. Proposed Method Process: First we use Genetic Algorithm to search a optimized colour quantization, then the resultant configuration is incorporated in the feature extractor to generate improved image descriptions.

A. Quantization Learning

In order to find a quantization that would provide a superior power of description and compactness for the image representations generated for a given image context, we opted by perform a optimization process provided by the Genetic Algorithm [23]. GA is a bio-inspired optimization heuristic that mimics natural genetic evolution to search the optimal in a solution space. It provides a fairly chance of reaching a global optimum by starting with multiple random search points and considering several candidate solutions simultaneously. Consequently, it represents a fair alternative to an exhaustive search strategy, which would be infeasible given the amount of possible quantizations.

According to this optimization algorithm, an individual corresponds to a representation of a potential solution to the problem that is being analysed. In our case, each individual represents a possible colour quantization. During the evolution process, which is based on a survival-of-the-fittest fashion, these individuals are gradually manipulated and selected, according to the established optimization criteria, in a iterative procedure until the stopping condition be satisfied. At this point, the expected result is an evolved individual that encodes a quantization by which the improved representations will be generated.

Individual Modelling: In our GA modelling, each quantization is represented in an individual by the following manner. The individual takes reference from the widest possible quantization, i. e., the one that has the maximum number of tonalities in each colour axis (8 in our case), and aggregates its intervals according to the configuration specified by the respective individual as it is detailed in Figure 2.

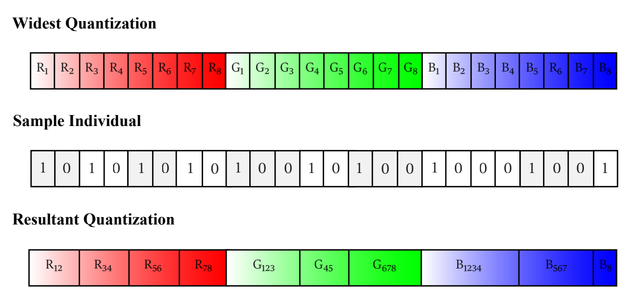


Fig. 2. Our modelling implements each individual as a binary array, being one value for each colour tonality interval. If a interval has its respective bit as set, it has its own position in the produced quantization, otherwise, it is aggregated to the immediate previous interval.

B. Image Description

On the second phase, the learnt quantization is used with the feature extractor algorithm in order to produce an image representation consequent from this quantization. In order to do that, it was necessary to implement a slightly modified version of the feature extractor, that incorporates the capacity of generating representations according to a specified colour quantization. The equations 1, 2 and 3, where N is the maximum colour axis size and Ind the quantization individual, show how to calculate the new R, G and B values for each pixel.

$$R_{new} = \left(\sum_{i=0}^{r} Ind[i]\right) * \frac{|R_{axis}|}{256} \qquad where \quad r = R * \frac{N}{256}; \quad |R_{axis}| = \sum_{i=0}^{N} Ind[i] \quad (1)$$

$$G_{new} = \left(\sum_{j=N}^{g+N} Ind[j]\right) * \frac{|G_{axis}|}{256} \qquad where \quad g = G * \frac{N}{256}; \quad |G_{axis}| = \sum_{i=N}^{2N} Ind[j] \quad (2)$$

$$B_{new} = \left(\sum_{k=2N}^{k+2N} Ind[k]\right) * \frac{|B_{axis}|}{256} \qquad where \quad b = B * \frac{N}{256}; |B_{axis}| = \sum_{k=2N}^{3N} Ind[k] \quad (3)$$

IV. Experimental Setup

In order to evaluate the proposed approaches for the method we conducted experiments relying on the BIC descriptor algorithm using eight different image datasets. The details about the experiments are presented as follows.

- *Task:* The proposed method was trained optimizing quantizations intending to solve the problem of Content-Based Image Retrieval (CBIR). The adopted process intended to solve this task basically consists on describing the whole image set and computing one similarity ranking by Manhattan distance (L1) for each one over all of them and then measuring the rankings quality. For this measurement, the image class is adopted as similarity criteria, as many images of the same class of the image in comparison remains in the top, better is the ranking.
- *Datasets:* The experiments were executed over a set of eight image datasets: UCMerced Land-use [26] and Brazilian Coffee Scenes [12], were initially created for Remote Sensing purposes, and the remaining, Coil-100 [27], Corel-1566 [28], Corel-3906 [28], ETH-80 [29], MSRCORID [30] and FRUITS [31], are intended for tasks of CBIR.
- *Baseline:* Since our goal is to propose a method capable of producing improved representations from already defined feature extraction, in order to measure the representations performances, the most suitable baseline is feature extractor itself, BIC, committed to the same experimental process although using its original colour quantization.
- *Metrics:* In order to evaluate the produced representations in the task of CBIR, the metric P@10, which means precision over the top 10 results, was used to measure each ranking performance and its average over all the images in the case of a whole dataset. The reason for choosing this metric is based on the fact that, usually on applications of the referred problem, the user gives prior attention for a small group of the top results.
- *Parameters:* For the GA in the quantization learning phase of the method on the Non-Limited approach were used the parameters: 200 individuals for 200 generations, 60% probability as two-point cross-over, 40% probability for one-point mutation, a tournament of 5 individuals and 1% of elitism. For the Limited approach the same parameters were used except that were used on 150 generations probably due the occurrence of an earlier convergence consequent of the discarded individuals of dimension out of the limit. As fitness function, we opted for FFP4 [32], which penalizes the misplaced results on the ranking by a quadratic factor and consequently tends to prioritize individuals that produce better quality rankings.
- *Experimental Protocol:* We carried out all experiments following a 5-fold cross-validation protocol. As a consequence, the datasets was almost equally divided into five non-overlapping folds and the approaches were executed 5 times using 80% of the images as training set and 20% as test set. Specifically for the Limited approach of the proposed method, were conducted five experiments according to the same protocol with different dimension limits, named sizes 128, 96, 64, 32 and 16, composing reductions of respectively 0%, 25%, 50%, 75% and 12.5%.
- *Execution Environment:* All experiments were performed on a 64 bits Intel Xeon E5-2673 v3 machine with 16 cores, 2.4 GHz of clock and 32GB of RAM memory. Ubuntu 14.04 LTS was used as operating system.

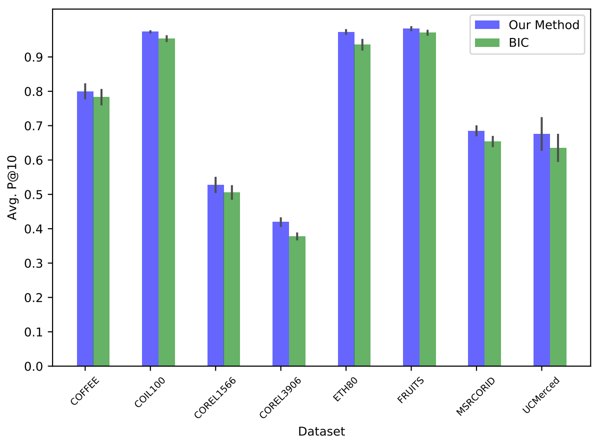
V. Results and Discussion

We propose two approaches for the described method. The first, named Non-Limited Approach, is intended to provided a quantization focused on generating representations that have the better performance as possible. The second, named Limited Approach, has the same goal, however it imposes a limitation on the representations size by giving negative fitness for the generated individuals that present dimensions over this limit. As a consequence, this later approach tends to focus on compactness. The following subsections present and discuss the experimental results and comparison between these two approaches of our method and its baseline.

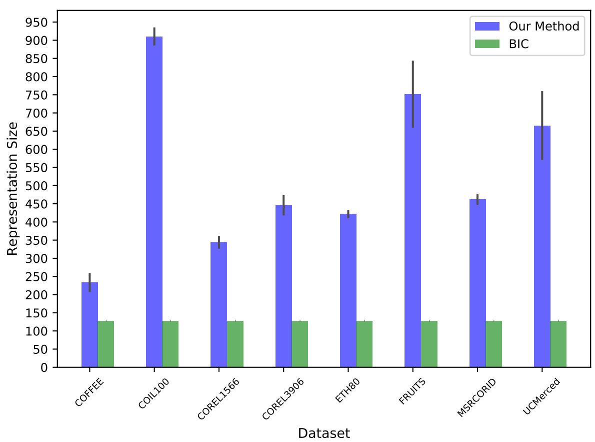
A. Non-limited Approach

Figure 3a presents the performance comparison between the Non-limited approach of our method and its baseline, in terms of performance in the described task. Considering only the mean values of avg. P@10, our method outperforms the baseline. However, given the proximity of the results, we used the Students Paired t-Test [33] to statistically verify this conclusion. According to the null hypothesis criteria of this test, which the measures are presented on Table I, it is possible to say that our method results outperforms the ones of the baseline in all datasets.

Observing the resultant feature vector dimensions in Figure 3b, the discrepancy between the two methods is easily noticeable. The representations produced by our method approach are, on average, around 300% bigger than the ones generated by the baseline. The reason for this outcome relies on the fact that the fitness function used for evaluate the genetic algorithm individuals prioritizes the representations performance on the task and does not consider any aspect related to its dimensions. That being said, it is likely that occurred a detailing of the colour tonalities generating a superior number of intervals and resulting in higher dimensions.



(a) Representation Performance



(b) Dimensionality

Fig. 3. Results comparison of the Non-limited approach of our method with BIC algorithm.

B. Limited Approach

The charts of Figure 4 show a pattern among for all datasets. The performance results of the Limited approach of our method were superior than the baseline for limits 64, 96, and 128. However, statistical tests, presented on Table I, according the Students Paired Method show an overlapping between the results for the limits 96 and 64, consequently, it is possible to declare that BIC method was outperformed only in cases of limit 128. Furthermore, the ascending behaviour of the performances suggests that as bigger the representation as superior its feature detailing level, which leads to a better representation quality.

The results of Figure 5 show that the generated quantizations almost exhausted the feature detailing by producing representations that reached or stayed very close to the dimension upper-bound. This is possibly a consequence of the optimization strategy of the method which is guided by the task performance. A fitness function that also considers the feature vector dimension would likely favour the generation of smaller representations under the same limit.

The presented results prove our hypothesis that it is possible to find an quantization optimized for a given domain that could provide an improved representation effectiveness and compactness. According to Figure 5, the results of limit 128, which present the same representation size as the ones of the BIC extractor, outperformed the retrieval quality of the baseline. Some results of limit 96 were even further presenting better performance with a smaller feature vector leading to conclude the possibility of improvements in performance and compactness simultaneously on the same quantization. Other results of limit 96 and 64 were statically tied with the baseline demonstrating the possibility of a significant reduction of the description size, until 50% in this case, but maintaining similar performance. Lastly, results of limit 32 and 16, performed badly for all datasets, showing the occurrence of loss in representation quality at a linear decay.

VI. Conclusions

We proposed two approaches of a representation learning method which intends to provide more effective and compact image representations by optimizing the colour quantization for the image domain. We performed experiments on eight different image datasets comparing the results with a predefined quantization approach in terms of performance on the task of CBIR and representations dimensionality.

The first approach, produced representations that outperform the performance of the baseline by a small percentage and presented a two times higher dimensionality. The second approach, which imposes a limitation on the representation dimension, presented results that show improvements on performance for the same dimensionality (128 bins), results that performed better even reducing the dimensionality in 25%, and also others that reduced the representation size until 50% but maintained statistically equivalent performance. Finally, the later approach also had results that imposed a reduction of more than 75% but presented poor performance showing the existence of a lower-bound for lossless compactness and that representations quality declines linearly with the limit.

At the end, the results prove the hypothesis, for the tested scenarios, that it was possible to produce more effective and compact fitness by exploring a colour quantization optimized for the image domain. Moreover, we remain at the end with a method capable of improve already existent feature extraction methods by providing descriptions more effective in terms of representation quality and more compact according to a parametric upper bound.

As future work we plan experimenting on approaches that use fitness functions that consider both effectiveness and compactness in the optimization process as the way of softening the dimensionality increasing. Furthermore, we aim to analyse how the presented approaches behave using different feature extractors and performing over other pattern recognition tasks. As long as the hypothesis were confirmed in these different scenarios, we consider scaling a similar optimization processes for use in GPUs aiming the possibility of providing an alternative for Deep Learning approaches.

ACKNOWLEDGMENT

This work was financed by CNPq (grant 449638/2014-6), CAPES, and Fapemig (APQ-00768-14). We also thank Microsoft Azure for the Research Grant.

REFERENCES

[1] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[3] S. M. Davis, D. A. Landgrebe, T. L. Phillips, P. H. Swain, R. M. Hoffer, J. C. Lindenlaub, and L. F. Silva, "Remote sensing: the quantitative approach," *New York, McGraw-Hill International Book Co., 1978. 405 p.*, 1978.

[4] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.

[5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] O. A. Penatti, E. Valle, and R. d. S. Torres, "Comparative study of global color and texture descriptors for web image retrieval," *Journal of Visual Communication and Image Representation*, 2012.

[8] J. A. dos Santos, O. A. B. Penatti, and R. da Silva Torres, "Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification." *VISAPP (2)*, pp. 203–208, 2010.

[9] O. A. B. Penatti and R. d. S. Torres, "Color descriptors for web image retrieval: A comparative study," *2008 XXI Brazilian Symposium on Computer Graphics and Image Processing*, pp. 163–170, 2008.

[10] R. O. Stehling, M. A. Nascimento, and A. X. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," *Proceedings of the eleventh international conference on Information and knowledge management*, pp. 102–109, 2002.

[11] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539 – 556, 2017.

[12] O. A. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015.

[13] R. d. O. Stehling, "Recuperação por conteudo em grandes coleções de imagens heterogeneas," Ph.D. dissertation, Unicamp, 2002.

[14] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[15] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

[16] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," *Artificial Intelligence and Statistics*, pp. 448–455, 2009.

[17] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," *Proceedings of the 26th annual international conference on machine learning*, pp. 609–616, 2009.

[18] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 921–928, 2011.

[19] L. Bo, X. Ren, and D. Fox, "Hierarchical matching pursuit for image classification: Architecture and fast algorithms," *Advances in neural information processing systems*, pp. 2115–2123, 2011.

[20] K. Yu, Y. Lin, and J. Lafferty, "Learning image representations from the pixel level via hierarchical sparse coding," *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 1713–1720, 2011.

[21] L. Luccheseyz and S. Mitray, "Color image segmentation: A state-of-the-art survey," *Proceedings of the Indian National Science Academy (INSA-A)*, vol. 67, no. 2, pp. 207–221, 2001.

[22] P. Scheunders, "A genetic lloyd-max image quantization algorithm," *Pattern Recognition Letters*, vol. 17, no. 5, pp. 547–556, 1996.

[23] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning, 1989," *Reading: Addison-Wesley*, 1989.

[24] M. G. Omran, A. P. Engelbrecht, and A. Salman, "A color image quantization algorithm based on particle swarm optimization," *Informatica*, vol. 29, no. 3, 2005.

[25] M. Ponti, T. S. Nazaré, and G. S. Thumé, "Image quantization as a dimensionality reduction procedure in color and texture feature extraction," *Neurocomputing*, vol. 173, pp. 385–396, 2016.

[26] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pp. 270–279, 2010.

[27] S. A. Nene, S. K. Nayar, H. Murase *et al.*, "Columbia object image library (coil-20)," 1996.

[28] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 9, pp. 947–963, 2001.

[29] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, pp. II–409, 2003.

[30] A. Criminisi, "Microsoft research cambridge object recognition image database," 2004.

[31] A. Rocha, D. C. Hauagge, J. Wainer, and S. Goldenstein, "Automatic fruit and vegetable classification from images," *Computers and Electronics in Agriculture*, vol. 70, no. 1, pp. 96–104, 2010.

[32] F. A. Faria, R. T. Calumby, and R. da Silva Torres, "Recod at imageclef 2011: Medical modality classification using genetic programming." *CLEF (Notebook Papers/Labs/Workshop)*, pp. 1–8, 2011.

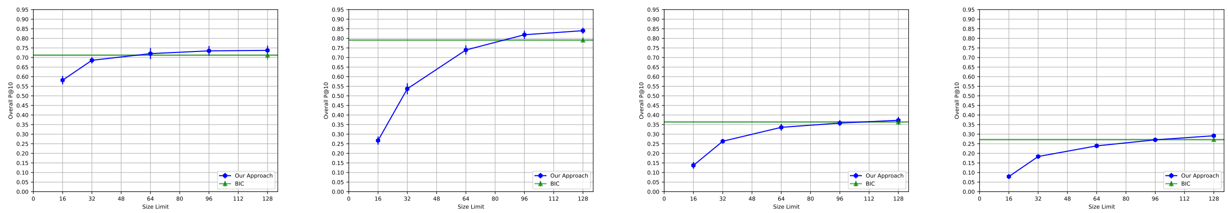
[33] T. K. Kim, "T test as a parametric statistic," *Korean journal of anesthesiology*, vol. 68, no. 6, pp. 540–546, 2015.

TABLE I
P-VALUES OF STUDENTS PAIRED T-TEST BETWEEN OUR METHOD APPROACHES AND THE BASELINE.

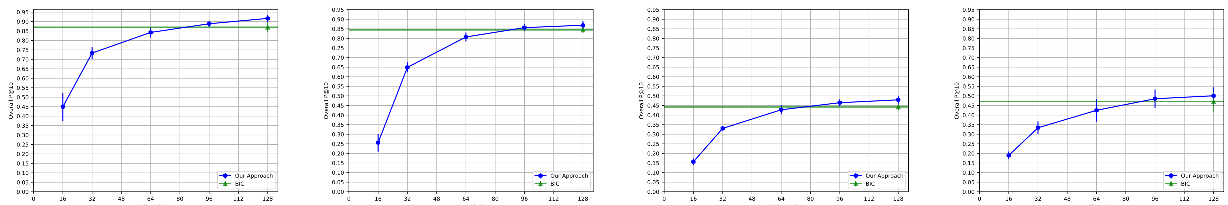
Dataset	Limit 16	Limit 32	Limit 64	Limit 96	Limit 128	NLA
Brazilian Coffee	0.000016	0.011904	0.422338	0.001282	0.000183	0.0001
COIL100	0.000004	0.000059	0.007223	0.026657	0.000867	0.0046
COREL1566	0.000025	0.000199	0.012336	0.334188	0.084463	0.0014
COREL3906	0.000007	0.000004	0.000308	0.779525	0.004303	0.0000
ETH80	0.000207	0.000260	0.082828	0.254271	0.006830	0.0027
Tropical Fruits	0.000027	0.000386	0.007135	0.241805	0.040865	0.0022
MSRCORID	0.000048	0.000343	0.098357	0.066615	0.000308	0.0004
UCMerced	0.000092	0.001865	0.000396	0.134587	0.023366	0.0052

Superior to BIC
Not Independent
Inferior to BIC

Considering null hypothesis rejected when p-value < 0.1, the green values correspond to the cases in which the method outperformed the baseline, the red ones in which it lost, and the blue ones in which the independence was not confirmed.

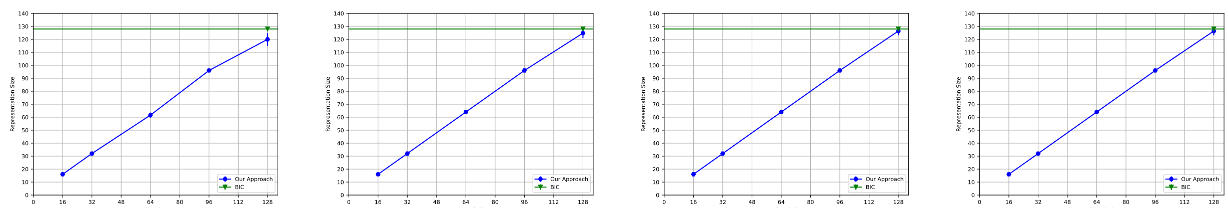


(a) Brazilian Coffee Scenes (b) Coil-100 (c) Corel-1566 (d) Corel-3906

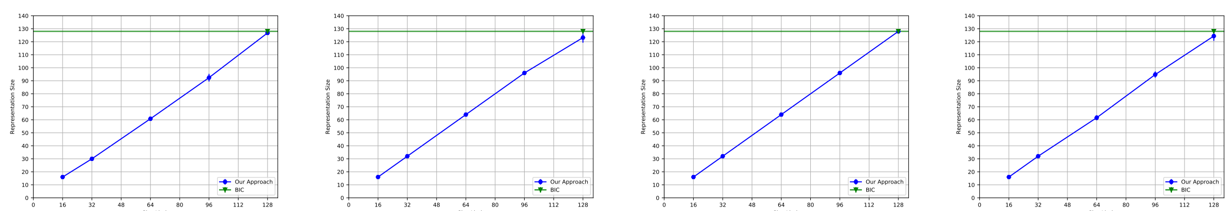


(e) ETH-80 (f) FRUITS (g) MSRCORID (h) UCMerced Land-use

Fig. 4. Results comparison of efficiency of the Limited approach of our method with BIC algorithm considering the dimension limits of 16, 32, 64, 96, 128.



(a) Brazilian Coffee Scenes (b) Coil-100 (c) Corel-1566 (d) Corel-3906



(e) ETH-80 (f) FRUITS (g) MSRCORID (h) UCMerced Land-use

Fig. 5. Results comparison of compactness of the Limited approach of our method with BIC algorithm considering the dimension limits of 16, 32, 64, 96, 128.