Learning on graphs and hierarchies

Raquel Almeida*
ImScience and IRISA
PUC Minas and Université de Rennes 1
Brazil and France
raquel.almeida.685026@sga.pucminas.br

Ewa Kijak
Linkmedia, IRISA
Université de Rennes 1
Rennes, France
ewa.kijak@irisa.fr

Simon Malinowski
Linkmedia, IRISA
Université de Rennes 1
Rennes, France
simon.malinowski@irisa.fr

Silvio Jamil F. Guimarães
ImScience
PUC Minas
Belo Horizonte, Brazil
0000-0001-8522-2056

***Abstract*—Hierarchies, as described in mathematical morphology, represent nested regions of interest that facilitate high-level analysis and provide mechanisms for coherent data organization. Represented as hierarchical trees, they have formalisms intersecting with graph theory and applications that can be conveniently generalized. However, due to the deterministic algorithms, the multiform representations, and the absence of a direct way to evaluate the hierarchical structure, it is hard to insert hierarchical information into a learning framework and benefit from the recent advances in the field. This work aims to create a learning framework that can operate with hierarchical data and is agnostic to the input and the application. The idea is to study ways to transform the data to a regular representation required by most learning models while preserving the rich information in the hierarchical structure. The methods in this study use edge-weighted image graphs and hierarchical trees as input, evaluating different proposals on the edge detection and segmentation tasks. The model of choice is the Random Forest, a fast, inspectable, scalable method. The experiments in this work are an outline of the original study in the related Ph.D. thesis. They demonstrate that it is possible to create a learning framework dependent only on the hierarchical data that performs well in multiple tasks.**

I. Introduction

Hierarchies are an inherent property composing several elements in real life, relating to how we naturally perceive patterns, scenes, and movement [1]. According to [2], there is a pattern identifier in the core of our visual perception, operating hierarchically to recognize parts, objects, and abstract concepts simultaneously. The perceptual hierarchy mimicking our ability to perceive reality's intrinsic nature is difficult to translate to computer models. But, in visual media processing, mathematical morphology has an edge in defining, creating, and manipulating hierarchies.

The hierarchies formulated in mathematical morphology [3] use the non-linear geometric space generalizing the set theory of complete lattices [4]. Hierarchical methods represent nested regions of interest and provide merging operations to build more semantically significant objects from lower-level instances. In multimedia processing, the lower-level regions consider the media's building blocks, such as pixels, voxels, and frequency. The semantically significant objects built from those regions help perform more complex tasks such as object proposal, semantic contour, and semantic segmentation [5].

. *Work related to the Ph.D. thesis entitled *Learning on graphs and hierarchies* available on the database of theses and dissertations of PUC Minas and Université de Rennes 1.

However, most hierarchical methods require thorough pre-processing of the data [6], [7] and strategies to deal with issues like over/under-partitioning of the space [8], [9] or selecting an ideal number of regions [10]. Therefore, it is difficult to generalize a successful approach to similar media and tasks.

For a generalization in terms of the media, most challenges regard the characterization of the information, in which the media data presents different characteristics and the building blocks have different connotations. These differences in form and connotation eventually become limiting factors and models created to solve a problem could only deal with that particular data type. In terms of task, the generalization is challenging due to the lack of a measure assessing the quality of a hierarchy, which requires an empirical refinement through a series of trial-and-error fittings for a particular application.

Furthermore, a framework operating on hierarchies presents some considerable additional challenges since the methods are deterministic, the product is multiform (different sizes, components, and interpretations) and the same data could create multiple hierarchical structures depending on the operators. Therefore, applying the morphological hierarchies in an agnostic learning framework requires a strategy to overcome the generalization issues and the deterministic, the quality assessment, and the heterogeneous aspects.

We argue that it is possible to directly insert the hierarchical structures in a learning framework and benefit from the embedded information to create a model for visual tasks that is agnostic to the media and task. The main goal is to design a learning framework that can operate directly on the hierarchical data and in doing so, it must deal with the generalization challenges and place a strategy to conform the hierarchical information to a learning framework.

We hypothesize that hierarchical representations contain valuable information embedded in their structures for a generic learning framework (**Hypothesis 1**). For this purpose, understanding the media's building blocks relation at the low level is critical to group them into homogeneous regions and the task definition should not impose assumptions on the data source.

Visual data, such as images and videos, are organized data structures, and information such as color, spatial distance, or variance defines homogeneity. An appropriate representation is ideally shared among most media types and is capable to retain the information presented in the media. Graphs are structures used to represent objects, and the primary concern in graph theory is how these objects are interconnected. They can depict many data and carry information about the objects in their components, including from different domains. In this sense, despite their differences, multimedia data share the form once modeled as graphs. And although defining homogeneous regions and their connotations are particular for each media, the grouping strategy and their storage in the hierarchical structure follow the same rules. Furthermore, one way to represent hierarchical data is as hierarchical trees. Therefore, both graphs and hierarchies have formalisms intersecting with graph theory and applications that can be conveniently generalized.

The proposed methodology uses the Random Forests [11] as a model. It is fast, simple, and scalable. Furthermore, it presents satisfactory results in multiple tasks. The main challenge in this proposal concerns the regular representation required by most machine learning algorithms. The regular representation is inherently opposed to the unconstrained nature of graphs and hierarchies. Hence, the proposed strategy is to represent the graph's components as vectors of selected attributes and assess its capability to retain the information modeled in the graphs while remaining discriminant for a task. Using a selection of graph attributes as input to the learning framework allows the formulation of a model agnostic to the media, and casting the information at the graphs' components level allows assigning each entry with a task label without imposing assumptions on the data source (**Hypothesis 2**).

However, depending on the modeling choices of the graphs, it can create a particular structured space known as a grid graph, which is close to the spatial domain of the media. Presuming generalization on a grid graph can be deceptive, and more than the structural information may be necessary for a discriminative representation. However, modeling the graphs from the hierarchical structure provides a non-regular characterization of regions with notions of order and navigation. The topology of the hierarchical structures alone could be used in a learning framework to solve multiple tasks if it preserves their semantical arrangement (**Hypothesis 3**).

To be explicit, this work does not present a multimedia application. Instead, the formulations and considerations focus only on the structures of the graphs and hierarchies. Therefore, it asses the discriminant information present in these structures while maintaining a certain level of agnosticism.

This work is organized into three main parts, one theoretical and two methodological, each addressing one hypothesis. Specifically: (i) Section II contextualize the theory of morphological hierarchies, formalizing graphs and hierarchies on the shared notation, and delimiting the problem; (ii) Section III presents a case study method for a learning framework operating on a selection of graph attributes; and (iii) Section IV presents the culmination of the proposals, expanding the concepts and strategies to the hierarchical data. After the main parts, Section V outlines the results of each proposed approach in an experimental setting and presents a discussion of the experimental investigation, and Section VI draws some conclusions and delineates some possible future work derived from this study.

II. Hierarchies and graphs

The hierarchical functions on mathematical morphology are rooted in the algebraic theory of complete lattices, modeling non-linear transformations with set operators to correlate whole sets of values [3]. In [12], the authors provided formal links between the morphological partitions and edge-weighted graphs. In this formalism, a structure could be defined as a hierarchy if it follows two **hierarchical principles** [13]: (i) causality: a particular element at one hierarchical level should be present at any consecutive level; and (ii) locality: regions must be stable when creating or removing partitions.

A. *Graph's formalism and notions*

A graph $G = (V, E)$ consists of a finite set of vertices, denoted by V, and a finite set of edges denoted by E, where $E \subseteq V \times V$. If $(u, v) \in E$ for two vertices $u, v \in V$, then u and v are adjacent vertices. The notion of vertices relates to the data's elemental components and the edges with the connections and dynamics between the parts. The set E induces a unique adjacency relation Γ on V, which associates $u \in V$ with $\Gamma(u) = \{u\} \cup \{v \in E | (u, v) \in E\}$.

An edge-weighted graph is denoted by $(G, \mathcal{F})$, in which $\mathcal{F} : V \times V \to \mathbb{R}$ is a function that weights the edges of $G = (V, E)$ and $\mathcal{F}(E)$ is the weighted map for the function $\mathcal{F}$ on the set E. The nature of $\mathcal{F}$ determines which characteristics the graph preserves, and selecting a function could be considered a similarity measure problem between two finite sets of points, where $\{w = \mathcal{F}(u, v) | (u, v) \in E\}$ is the weight w of an edge $(u, v) \in E$ that could describe the dissimilarity of u and v.

A tree is a particular case of a direct graph ($(u, v) \neq (v, u), \forall u, v \in V$). In a tree, we denote vertices as nodes. The root is the single node at the top of the tree that connects all the other nodes. From the root, every subsequent node is a child and the leaves are at the bottom with no children. From the root, each node in the path to a leaf characterizes one level, the maximum number of levels defines the tree depth, and the altitude of a node starts from the leaves and it is inversely proportional to the depth of the node.

B. *Hierarchies: from graphs to graphs*

A hierarchy operating on the edge-weighted graph defines non-gridded regions as subsets of the vertices. For G and G' the graph induced by V' is $G = (V', \epsilon)$ where $V' \subseteq V$ and $\epsilon = \{(u, v) \in E \,|\, u, v \in V'\}$. V' is a connected component of G if V' is connected for G and maximal.

A set $\mathcal{H} \subseteq \mathbb{V}$, where $\mathbb{V}$ denotes the set of all subsets on V, is a hierarchy on V if $H_1 \cap H_2 \in \{\emptyset, H_1, H_2\}$ for any two elements $H_1, H_2 \in \mathcal{H}$. These notations characterize a direct forest that portrays the hierarchy as a Hasse diagram, also known as a dendrogram representation [14]. Therefore, **a hierarchy is a graph in the form of a hierarchical tree**.

A partition $\mathbb{P}$ is a set of non-empty disjoint subsets of V, meaning that $\forall X, Y \in \mathbb{P}$, X and Y are regions, $X \cap Y = \emptyset$ if $X \neq Y$ and $\cup\{X \in \mathbb{P} = V\}$. The partition set is ordered from finer in $\mathbb{P}'$ to coarser in $\mathbb{P}''$ if any region in $\mathbb{P}'$ is present in $\mathbb{P}''$ for any $\mathbb{P}', \mathbb{P}'' \in \mathbb{P}$. The ordered relation conveys the idea of refinement, in which navigating from finer to coarser is known as region aggregation and the opposite as region splitting.

A hierarchy of partitions $\mathcal{H} = (\mathbb{P}_0, \ldots, \mathbb{P}_k)$ is a sequence of partitions on V, such that $[\mathbb{P}]_{i-1}$ is a refinement of $[\mathbb{P}]_i$ $\forall i \in \{1, \ldots, k\}$ where k is the number of levels in the hierarchy characterizing its altitude and depth. The hierarchy preserves the non-empty disjoint sets notion and the ordered relation. The union of all partitions of $\mathcal{H}$ creates the set of regions of $\mathcal{R}_{\mathcal{H}}$, and the inclusion relation induces a tree structure. The hierarchical partition tree $\mathcal{T}_{\mathcal{H}}$ is the tree representing the hierarchy $\mathcal{H} = (\mathbb{P}_0, \ldots, \mathbb{P}_k)$.

The hierarchical construction algorithms use the weights to regulate how regions are formed, the criterion to merge and create new ones, and the order to pursue. This work presents two particular hierarchical models, namely: (i) the **quasi-flat zones (QFZ)**, induced directly from the graph with altitudes ordered based on increasing values of the edge-weights [15]; and (ii) the **hierarchical watershed** with altitudes ordered based on a geometric criterion [16]. Each construction algorithm has its particular properties and interpretation of the data. However, the rules on the hierarchical principles and the ordered representation of regions create a shared space convenient for commuting from one type to another if one representation is inadequate for an application.

In a typical pipeline for an image processing task, after adequately preparing the image, the graph is created and the hierarchy is constructed according to the selected method. Once constructed, it is necessary to decide how to represent the hierarchies to be applied since most ground-truth references need a flat (*i.e.*, non-hierarchical) form for comparison. In this step resides the central problem of this work.

The **trivial approach** is a series of horizontal cuts selecting multiple independent partitions representing the hierarchy. The selection could indicate the desired number of regions portrayed on the partition (a strenuous job that goes from a single to the total number of regions) or a threshold of the hierarchical levels (crucial detail present at one hierarchical level could be merged on the subsequent levels). Furthermore, a good horizontal cut for one specific hierarchy does not guarantee that it will be ideal for another on the same dataset.

Other representation strategies include post-processing the hierarchies by flattening [17], realigning [18], or filtering the structure [19]. These strategies rely on identifying less relevant regions and re-weight or merging these regions, creating more concise representations. The problem with these approaches is that defining the region's importance is subjective and strongly related to a media type or task. Alternatively, one could search for the ideal representation with a non-horizontal cut [20], which is, by all means, a combinatorial problem.

III. Learning on graphs

The discussions about graph creation and manipulation can be made generic enough to model any data, but, for instance, we are interested in image graphs, in which the spatial connectivity of the pixels gives us a structured representation of a grid graph, close to the spatial domain and strengthened with relational aspects. For the image graph G defined on the image domain, the adjacency relation Γ between the pixels is typically obtained by a structured adjacency relation, such as 4- or 8-adjacency in a grid form, and the set of vertices $V = \{v_1, v_2, \ldots, v_N\}$ represents the N pixels of the image.

The set of functions associated with each vertex is denoted by $f : V \subset \mathbb{Z}^2 \rightarrow \mathbb{R}$. Common functions in f include low-level descriptors and variations in the color space or in the gray-scale magnitudes. The latter is notably important as the most common source to calculate the weighting function. Ideally, the weighting function could characterize similarities, and for such, the Euclidean distance is the most common, defined in E as $\boldsymbol{\mathcal{F}}_{\mathbf{euc}}(u, v) = \sqrt{(f(u) - f(v))^2}$.

The edge weights may represent the local variation around a vertex, and serve as an image gradient operator bounded by the adjacency relation. Weighting edges as an image gradient operator acts as a transformation filter on the image creating a transformed space by changing the contrast of the original image and spreading the intensity levels. As in the case of many spatial filters based on local differences, the graph-based gradient operator defined for $\mathcal{F}_{\mathbf{euc}}$ is subjected to respond strongly to noise. We expect that the attribute selection on the RF trees can mitigate this aspect and also any eventual poor topology choice while reinforcing desirable characteristics.

A. Random Forest as regularizers to edge-weighted graphs

An RF is a non-parametric machine learning method that can be used both for classification and regression. The RF predictor consists of M randomized trees. The core of the RF algorithm, as proposed by [11], is the randomization of sampled data distributed to supervise the training of independent decision trees and the aggregation of the results for the final prediction. RF is empirically successful in suppressing noise, although the statistical and mathematical properties of the procedure are still obscure [21]. Some consensus is that the randomness in RF performs as an implicit regularization process, behaving as interpolating classifiers that encourage large consistent regions and reduce the effect of noise [22].

To use the RF implicit regularization process with the local variation representation of the edge-weighted graph, we propose to use the information on the graph edges and vertices to represent the graph on the framework. We represent the regular input of the RF as $\mathcal{D}_n = ((\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n))$ with $n \subseteq |V|$ samples of vertices of the edge-weighted graph $(G, \mathcal{F})$, each represented as a vector $\mathbf{X} \in \mathbb{R}^p$ and label $\mathbf{Y}$. In our application, edge-weighted graphs are created from images, each vertice thus corresponds to a pixel. $\mathbf{X}$ is a vector with dimension $p = |\mathbf{G_{att}}|$ for $\mathbf{G_{att}}$ representing a set of selected attributes of the vertices of $(G, \mathcal{F})$. In this work, the selected attributes belong to two categories:

- **vertex attributes ($\mathbf{X}_V$)**, belonging to the set of vertices functions f. Each $v \in V$ is mapped into a set of low-level color descriptors proposed in [23].
- **edge weights ($\mathbf{X}_{\mathcal{F}}$)**, for a given vertex v and all its adjacent vertices, it is represented by the set of edge weights between them. Therefore $\mathbf{X}_{\mathcal{F}} = \{\mathcal{F}_{\mathbf{euc}}(u, v) \mid \forall u \in$

$\Gamma(v)\}$. In this work, we go further into the immediate neighbors of v and include also the neighbors in the adjacency of the immediate neighbors. Therefore,

$$\mathbf{X}_{\mathcal{F}} = \{\mathcal{F}_{\mathbf{euc}}(u, v), \mathcal{F}_{\mathbf{euc}}(w, u)\}$$

for all $u \in \Gamma(v)$ and $\forall w \in \Gamma(u)$.

We thus end with $\mathbf{X} = \mathbf{G_{att}} = \{\mathbf{X}_V \mathbf{X}_{\mathcal{F}}\}$, by concatenating the two sets of selected attributes.

To obtain gradients, RF is trained on an edge detection task. Because each vertex of the graph is created from a pixel of the image with a unique label on the ground truth, all the n entries $\mathbf{X}$ have a unique discrete label $\mathbf{Y} \in \{0, 1\}$ on the task of edge detection. At inference, all vertices of a test graph are subjected to the estimations of the RF The final estimated value, obtained by averaging the M estimates is thus taken as a confidence value that a certain vertex $\mathbf{X}$ indeed represents an edge. These estimated values for vertices are mapped back to the image coordinates as an intensity value to create the image gradient. We called this method **graph-based image gradient operator (GIG)** and the produced gradients depict firm contours of the objects and other characteristics such as minor components, textures, and large uniform regions.

IV. Learning on hierarchical attributes

This section presents a learning framework formulated on the structural components of the hierarchies and a regular representation of the structure attributes. We present two strategies for selecting attributes from the hierarchical structures: (i) a regular representation selecting topological properties from the hierarchical trees; and (ii) regional features deduced from the hierarchies and their conjoined graph.

A. Topological attributes

A hierarchical tree $\mathcal{T}_{\mathcal{H}}$ representing the hierarchy of partitions $\mathcal{H} = (\mathbb{P}_0, \ldots, \mathbb{P}_k)$ created from the edge-weighted graph $(G, \mathcal{F})$ has a set of nodes $\mathcal{N}$. The depth d_n of a node $n \in \mathcal{N}$ is its number of parents. At the bottom of this tree, there is a collection of leaves $\mathcal{L}$ representing the partition $\mathbb{P}_0$, where $\mathbb{P}_0 = \{[\mathbb{P}]_v \mid \forall v \in V\}$ and each $l \in \mathcal{L}$ corresponds to a $v \in V$. The proposed representation depicts each leaf $l \in \mathcal{L}$ as a vector $\mathbf{T}_l$ of selected attributes. The selection corresponds to one of the following attributes: (i) **Altitude**: the value inversely proportional to the depth of the node; and (ii) **Area**: sum of the number of leaves on the subtree rooted on the node.

The selected attribute is computed for all parents of l. Each leaf has a variable number of parents; therefore, the dimension p_t of the vector $\mathbf{T}_l$ is standardized by the maximum depth in all $\mathcal{T}_{\mathcal{H}}$ computed for a dataset. Also, the leaves with a set of parents smaller than the maximum depth receive a padding value of -1.

The semantical meaning is kept by representing the parents of a leaf node in the *order* they appear transversing the hierarchical tree. The order could be *ascending* (from leaf to root) or *descending* (from root to leaf). Early experiments showed that essential attributes occur at the initial positions of the feature vector and are favored by the RF model during training. Therefore, we use the *ascending order* in this work.

B. Regional attributes

The second strategy uses a set of regional attributes created on the conjoined graph by the hierarchical structure. Formally, each node $n \in \mathcal{N}$ represents a region $\mathcal{R}_n$ that is the union of all regions on the subtree τ_n rooted on the node n. A cut is a partition $\mathbb{P}$ of V made of regions of $\mathcal{H}$, where a horizontal cut is a partition $\mathbb{P} = \mathbb{P}_i$ for $i \in \{0, \ldots, k\}$ for all k altitude levels on the tree. A horizontal cut by altitude levels defines the partition by a threshold σ on its altitude values. Two regions $\mathcal{R}$ and $\mathcal{R}'$ are in the same region $\mathcal{R}_n$ if n is their lowest common ancestor that have $\texttt{alt}_n > \sigma$.

Consider β as a series of altitude levels to cut the hierarchy. The proposed representation depicts each leaf $l \in \mathcal{L}$ as a vector $\mathbf{R}_l$ of size $|\beta|$. At each position of this vector, there is a cut $\mathbb{P}_\sigma$ for $\sigma \in \beta$. Thus, the leaf l is represented by a selected regional attribute for the region $\mathcal{R}_n$ where n is the lowest parent of l whose $\texttt{alt}_n > \sigma$. The selection corresponds to one of the following: (i) **Contour strength**: The contour of a node is the number of edges on the conjoined weighted graph shared among the regions merged by a node. The contour strength is the average of edge weights on the contour; and (ii) **Gaussian**: Estimates the Gaussian distribution of leaf weights in the region $\mathcal{R}_n$ defined by the node n. The function returns the mean and the variance. The leaf weights could be defined for any attribute or set of attributes (on which one could calculate the covariance). Here, they are the sum of the weights of the edges comprising the vertice equivalent of the leaf.

The selected attribute is computed for all regions created by the cut $\sigma \in \beta$, and the ordered representation is preserved on the cut despite not representing every possible region in the hierarchy. It is proposed to select only a few steps in the normalized altitudes creating a reduced set of features guaranteed to be present in all hierarchical types.

The procedure for test instances in both proposed representations takes the regular representation of each hierarchy in the test set and individually subjects them to the RF estimations without the labels.

V. Experiments, main results, and discussion

This section outlines the main results obtained with the trivial, the graph, the topological, and the regional approaches in two image tasks: edge detection and segmentation. The graph approach compares three representations: (i) onlyColor: with only vertex attributes; (ii) GIG-Edge: with only weight values; and (iii) GIG: with both categories of attributes. The pipeline (Fig. 1) is formulated on the structural components of the graphs and hierarchies and the regular representation of the structures uses the discussed selection of attributes.

The edge detection dataset is the Berkeley Segmentation Dataset and Benchmark (BSDS500 [24], illustrated in Fig. 2). It contains 500 (200 train, 100 validation, and 200 test) natural images, presenting complicated/high-contrast patterns, occluded objects, and objects indistinguishable from the background by color. Each image has multiple labels performed by different annotators; thus, we performed a majority vote to obtain a single label. For segmentation, the Birds [25]

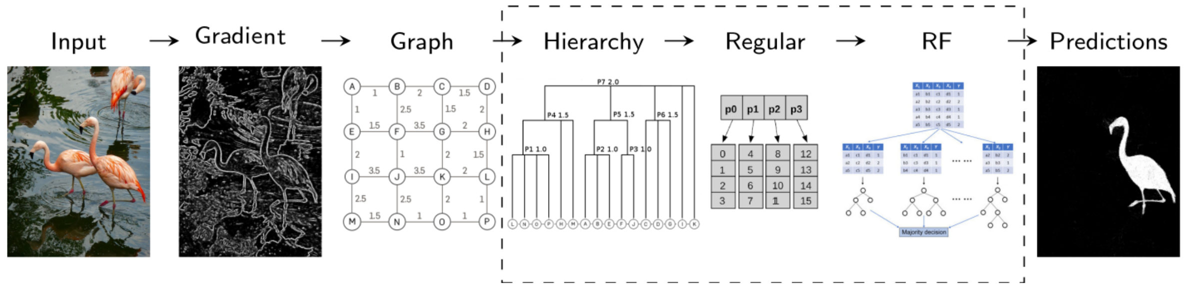

Fig. 1: Illustration of the framework from the input image to the Random forest predictions performing the task. First, it computes the gradient for each image in the dataset. Then, it calculates the edge-weighted graphs, here illustrated with the $4-$adjacency relation, and the hierarchical approach constructs the hierarchies from the graphs. The next step creates a regular representation from the selected attributes to serve as input for the Random Forest model. The regular input for the training set includes the associated label: the unique discrete label on the task for each component. During the test, the Random Forest subject each leaf of the test hierarchies to prediction, where the estimated values are mapped back to the image coordinates for evaluation.

(illustrated in Fig 1) is a binary segmentation public dataset. It contains 50 images of birds with manual annotations and no official train/test sets division (we randomly selected a 35/15 train/test split). The challenging images usually portray the birds close to a body of water, with areas of high-intensity lights and annotations covering only one leading object, despite the presence of multiple similar objects.

The pipeline takes the colored images and computes the graph gradient (GIG [26]) without any additional preprocessing. Next, a structured grid obtains the adjacency relation 4-adjacency. Each vertex is associated with a Euclidean distance on the gradient magnitudes for the weighting function. In the hierarchical approach, the hierarchy construction explores the QFZ and the hierarchical watershed using the number of parents (WATER-PAR) as topological criterion [27]. We do not perform any additional post-processing, such as filtering, realigning, or balancing the hierarchical levels.

All representations are aggregated using Random Forests (RF) with the parameters set as 500 trees in the forest. The trivial approach does not involve a learning step and the experiments explored a large range of parameters for the cut, in which the best are: (i) 1000 regions for QFZ and 60 for WATER-PAR using the cut by the number of regions; and (ii) threshold at 0.22 for QFZ and 0.53 for WATER-PAR using the horizontal cuts by threshold. The BSDS500 dataset proposes an evaluation system, which takes an edge map at multiple threshold values computing the precision-recall $F1-$score at all threshold values. The results are then presented in terms of the optimal dataset scale (best threshold representing most of the images). In the Birds dataset, the pipeline considers an RF classifier where predictions for each component on the binary segmentation labels are directly mapped back to the image space. The evaluation metric use the Jaccard similarity coefficient score as the metric, which is equivalent to the precision-recall $F1-$score on binary sets.

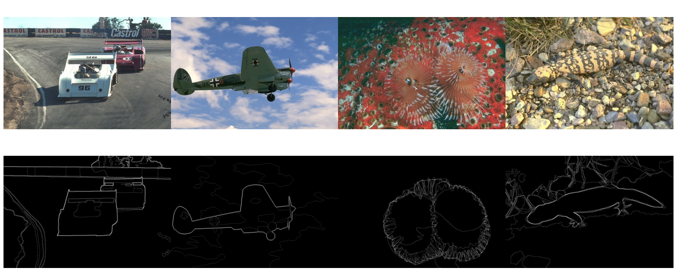

Fig. 2: BSDS500 dataset sampled images with their respective boundary ground truths. It contains colored natural images presenting complicated patterns, occluded objects, main objects indistinguishable from the background by color, and objects with patterns of high contrast. Each image contains multiple labels where line intensities indicate the annotators' agreement.

A. Quantitative analysis

Table I shows the main results for the proposed strategies. While the results with the trivial approach on the BSDS500 dataset are considerably worst compared with the other strategies, they are presented to establish a baseline, not to say that hierarchical structures are ineffectual for the edge detection task. On the contrary, many hierarchical proposals in this dataset present competitive results [20], [28]–[30] given a proper strategy to improve or filter the hierarchical contours. As for the segmentation task with Birds, the illumination conditions on the images create a scenario that is very challenging for many of the best image processing methods. With the hierarchical methods, the algorithms will create similar partitions for the many objects portrayed in the images, while only one is

TABLE I: Quantitative comparison of the results obtained in all datasets for the compared approaches. $F1-$score for best dataset scale for the BSDS500 and average Jaccard score for Birds. Emphasizes the best scores per approach variation and red emphasis the best score per dataset.

		BSDS		**Birds**	
Graph	GIG	**0.65**		**0.29**	
	GIG-Edge	0.64		0.28	
	onlyColor	0.61		0.27	
Trivial	Hierarchy	Threshold	Regions	Threshold	Regions
	QFZ	0.26	**0.28**	**0.14**	0.05
	WATER-PAR	0.24	**0.53**	**0.28**	0.24
Topological	Hierarchy	Altitude	Area	Altitude	Area
	QFZ	**0.60**	0.52	0.30	**0.37**
	WATER-PAR	**0.63**	0.54	0.32	**0.41**
Regional	Hierarchy	Contour	Gaussian	Contour	Gaussian
	QFZ	0.63	**0.67**	**0.53**	0.51
	WATER-PAR	0.63	**0.65**	**0.71**	0.64

considered a valid answer. The graph representations improve the results from the trivial approach in the BSDS500 dataset since all considerations regarding the hierarchical levels are removed from the label attribution before training.

The topological strategy improves the results for almost all hierarchical types for all datasets (except for WATERPAR with altitudes in Birds) when compared with the trivial approach. The additional benefit is that it does not require an empirical search on the hierarchical levels and regions for evaluation. Furthermore, the topological approach presents best results than the trivial and the graph in the Birds dataset. In edge detection, the graph and the topological perform better than using only the color features, with the GIG approach performing better than the best on the topological strategy. Regarding the topological attributes, the altitudes perform better on the edge detection and the area on the segmentation, which matches the task goals with the attributes' properties.

The regional strategy presents the best results in all datasets. Even for the challenging Birds, there is at least one attribute for all hierarchical types that give a satisfactory result. The Gaussian presents, in general, superior results on the different tasks. Because the Gaussian attribute quantifies the region distribution on the hierarchical trees, it assimilates the representation with the task. Future applications of this strategy may consider the hierarchical type that most agree with the objectives and use the Gaussian attribute for the representation.

B. Discussion

The great incentive to center the considerations towards graph processing is that they are critical for hierarchical analyses, and machine learning operating on graphs provides a form to create an agnostic model regarding the media type. Machine learning on graphs is a topic of great interest due to its autonomy, the multiple possibilities of applications, and the capacity to represent multivariate information.

The hierarchical structure provides a non-regular characterization of regions with notions of order and navigation without needing many parametrizations other than those offered by the already modeled edge-weighted graph. They introduce a semantic interpretation into media processing through meaningful partitioning of the perceptual space. Hierarchical operators are idempotent and provide a consistent data organization.

By keeping the formulation on the structures, the proposed framework evades decisions at the media level. It avoids any feature extracted from the media and only uses the information on the hierarchical tree and their conjoined graph. Also, it does not select any particular region that better suits an application. Instead, the entire structure is represented in a vectorial form that preserves its semantical arrangement. Furthermore, the task label attribution is performed at the leaf level at the bottom of the tree; therefore, each leaf has a unique discrete label and does not demand any considerations specific to a task. Similar methods in the literature use attributes and regions defined in the hierarchies to gather features from the media for the learning model [31]–[33].

VI. Conclusion

The main goal of this work was to design a generic learning framework that could operate on hierarchical data, dealing with the generalization challenges in media and tasks and placing a strategy to conform the hierarchical data to a learning framework. Hierarchies are rich structures that could model a myriad of data. It facilitates the analysis of complex problems in multiple domains. However, they require careful consideration, and parsing the structures can be challenging and limit their applications. Graphs are dynamic structures for modeling multimedia, but like hierarchies, they require thoughtful considerations when applied in a machine learning framework. Using the information on the graph edges and vertices is a viable method to represent a graph in a learning framework. It allows controlling the representation size and selecting the information depicted considering the type of graph, its proximity to the original data, and the expected results. Furthermore, representing the graphs at the vertex level allows maintaining the analyses on the discrete space.

The thesis demonstrated that it is possible to create a learning framework dependent only on the hierarchical data that performs well in multiple tasks with different models. It created and delivered a learning framework operating directly on the hierarchical structure, avoiding any feature extracted from the media and only using the information on the hierarchical tree and graph while preserving its semantical arrangement.

A. Perspectives and future work

The applications and experiments developed in this study were all performed on the image space because it allows an easy visual inspection of the result's quality. An application in another media type could take the same considerations of attribute selection since once the media is modeled as a hierarchy or graph, they will all share the same rules in that space. If generalization is not a concern in future applications, one could use the proposed strategies to transpose the structure to the vectorial space while taking the appropriate measures to improve the results on a media-specific task. One possible direction for future work on the already proposed strategies is to combine the attributes selected from the graphs and hierarchies in the same or different categories of features, enriching the information presented to the machine learning model. Another possibility is to apply a strategy to reduce the structure prior to the attribute selection or employ a data reduction strategy.

B. Main contributions

We list the main contributions in terms of publications: (i) Learning framework on graph attributes for image processing (Published in SIBGRAPI'22 - Awarded as best paper [34]) (ii) Extended formalism on graphs attributes exploring more extensive input areas through region adjacency graphs and changes driven by the model mechanics (Published in PRL [26]). (iii) Learning framework operating directly on the hierarchical data, focusing the formulations solely on the structural components of the hierarchies (submitted).

ACKNOWLEDGMENT

The authors thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – (PQ 310075/2019-0 and Grant 407242/2021-0), Fundação de Amparo a Pesquisa do Estado de Minas Gerais – FAPEMIG – (Grants PPM-00006-18, APQ-01079-23), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant STIC-AmSUD TRANSFORM 88881. 14325 8/2017-01, grant COFECUB 88887.191730/2018-00 and funding Finance Code 001), PUC Minas and INRIA under the project *Learning on graph-based hierarchical methods for image and multimedia data*.

REFERENCES

[1] D. Marr, *Vision: a computational investigation into the human representation and processing of visual information*, M. Acock, Ed. MIT Press, 1982.

[2] R. Kurzweil, *How to create a mind: the secret of human thought revealed*, B. Giffords and R. Ottewell, Eds. Penguin Books, 2013.

[3] L. Najman and H. Talbot, *Mathematical morphology: from theory to applications*, 1st ed., L. Najman and H. Talbot, Eds. John Wiley & Sons, Inc., 2013.

[4] J. Serra, "A lattice approach to image segmentation," vol. 24, pp. 83–130.

[5] P. Bosilj, E. Kijak, and S. Lefèvre, "Partition and inclusion hierarchies of images: a comprehensive survey," *Journal of Imaging*, vol. 4, p. 33, 2018.

[6] M. Clément, C. Kurtz, and L. Wendling, "Learning spatial relations and shapes for structural object description and scene recognition," *Pattern Recognition*, vol. 84, pp. 197–210, 2018.

[7] T. T. Nguyen, P. Krishnakumari, S. C. Calvert, H. L. Vu, and H. van Lint, "Feature extraction and clustering analysis of highway congestion," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 238–258, 2019.

[8] K. Nandy, P. R. Gudla, R. Amundsen, K. J. Meaburn, T. Misteli, and S. J. Lockett, "Supervised learning framework for screening nuclei in tissue sections," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE*. Boston, MA, USA: IEEE, 2011, pp. 5989–5992.

[9] G. Zwettler and W. Backfrieder, "Evolution strategy classification utilizing meta features and domain-specific statistical a priori models for fully-automated and entire segmentation of medical datasets in 3d radiology," in *International Conference on Computing and Communications Technologies*. Chennai, India: IEEE, 2015, pp. 12–18.

[10] F. Meyer, "Hierarchies of partitions and morphological segmentation," in *Scale-Space and Morphology in Computer Vision*. Springer Berlin Heidelberg, 2001, vol. 2106, pp. 161–182.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[12] J. Cousty, L. Najman, and B. Perret, "Constructive links between some morphological hierarchies on edge-weighted graphs," in *Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer Berlin Heidelberg, pp. 86–97.

[13] L. Guigues, J. P. Cocquerez, and H. L. Men, "Scale-sets image analysis," vol. 68, pp. 289–317.

[14] R. R. Sokal and J. Rohlf, "The comparison of dendrograms by objective methods," vol. 11, pp. 33–40, 1962.

[15] J. Cousty, L. Najman, Y. Kenmochi, and S. Guimarães, "Hierarchical segmentations with graphs: quasi-flat zones, minimum spanning trees, and saliency maps," *Journal of Mathematical Imaging and Vision*, vol. 60, pp. 479–502, 2018.

[16] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Computational Imaging and Vision*. Springer, 1994, vol. 2, pp. 69–76.

[17] C. Xu, S. Whitt, and J. J. Corso, "Flattening supervoxel hierarchies by the uniform entropy slice," in *IEEE International Conference on Computer Vision*. IEEE, 2013, pp. 2240–2247.

[18] M. M. Adão, S. J. F. Guimarães, and Z. K. G. P. Jr, "Learning to realign hierarchy for image segmentation," vol. 133, pp. 287–294.

[19] B. Perret, J. Cousty, S. J. F. Guimarães, Y. Kenmochi, and L. Najman, "Removing non-significant regions in hierarchical clustering and segmentation," *Pattern Recognition Letters*, vol. 128, pp. 433–439, 2019.

[20] P. Arbelaez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 328–335.

[21] E. Scornet, "Random forests and kernel methods," *IEEE Transactions on Information Theory*, vol. 62, pp. 1485–1500, 2016.

[22] A. Wyner, M. Olson, J. Bleich, and D. Mease, "Explaining the success of adaboost and random forests as interpolating classifiers," *The Journal of Machine Learning Research*, vol. 18, pp. 1558–1590, 2017.

[23] P. Dollar, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *Procedings of the British Machine Vision Conference*. British Machine Vision Association, 2010, pp. 68.1–68.11.

[24] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision*. IEEE Comput. Soc, 2001, pp. 416–423.

[25] L. A. C. Mansilla and P. A. V. Miranda, "Oriented image foresting transform segmentation: connectivity constraints with adjustable width," in *Conference on Graphics, Patterns and Images*. IEEE, 2016, pp. 289–296.

[26] R. Almeida, E. Kijak, S. Malinowski, Z. K. P. Jr, A. A. Araújo, and S. J. Guimarães, "Graph-based image gradients aggregated with random forests," *Pattern Recognition Letters*, 2022.

[27] B. Perret, J. Cousty, S. J. F. Guimaraes, and D. S. Maia, "Evaluation of hierarchical watersheds," *IEEE Transactions on Image Processing*, vol. 27, pp. 1676–1688, 2018.

[28] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "From contours to regions: An empirical evaluation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2294–2301.

[29] K.-K. Maninis, J. Pont-Tuset, P. Arbelaez, and L. V. Gool, "Convolutional oriented boundaries: from image segmentation to high-level tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 819–833, 2018.

[30] C. J. Taylor, "Towards fast and accurate segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 1916–1922.

[31] E. Grossiord, H. Talbot, N. Passat, M. Meignan, and L. Najman, "Automated 3d lymphoma lesion segmentation from pet/ct characteristics," in *International Symposium on Biomedical Imaging*. IEEE, 2017, pp. 174–178.

[32] Z. Hu, T. Shi, C. Wang, Q. Li, and G. Wu, "Scale-sets image classification with hierarchical sample enriching and automatic scale selection," *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, p. 102605, 2021.

[33] F. J. A. Padilla, B. Romaniuk, B. Naegel, S. Servagi-Vernat, D. Morland, D. Papathanassiou, and N. Passat, "Random walkers on morphological trees: a segmentation paradigm," *Pattern Recognition Letters*, vol. 141, pp. 16–22, 2021.

[34] R. Almeida, Z. K. G. Patrocínio Jr., A. d. A. Araújo, E. Kijak, S. Malinowski, and S. J. F. Guimarães, "Descriptive image gradient from edge-weighted image graph and random forests," in *34th Conference on Graphics, Patterns and Images*, 2021, pp. 338–345.