

Superpixel Segmentation Effect on Hierarchical GNN applied to Image Classification

João Pedro de Melo Murta, João Pedro O. Batisteli, Silvio Jamil F. Guimarães, Zenilton K. G. Patrocínio Jr

Laboratory of Image and Multimedia Data Science (IMScience)

Pontifical Catholic University of Minas Gerais (PUC Minas), Belo Horizonte, Brazil

Email: jmurta@sga.pucminas.br, joao.batisteli@sga.pucminas.br, sjamil@pucminas.br, zenilton@pucminas.br

Abstract—In the image processing field, Graph Neural Networks (GNNs) are employed to learn from the graph structure to its full potential. Before model training, the data transformation converts the image into a graph structure. In this sense, many superpixel algorithms such as SLIC, DISF, SCALP, and ODISF have been adopted for reducing complexity and creating meaningful pixel regions based on some criteria. As these algorithms offer distinct characteristics and superpixel aspects, this work-in-progress presents a study on the performance of the HiErarchical Layered Multigraph Network (HELMNet) according to the variation of superpixel methods used to generate the base graph of the HiErarchical Layered Multigraph (HELM) representation with images from the STL-10 dataset. An attention function called Region Graph Readout (RGR), used for classification, is also set to guide the clarification of the divergent results. Furthermore, this study aims to highlight the nonexistence of a general rule for selecting a superpixel method given a specific visual task.

I. INTRODUCTION

Graph Neural Networks (GNNs) [1] are powerful tools for image classification tasks with data represented in the form of graphs. This revolutionary model can operate directly with data in the non-Euclidean space and harness contextual information in graph structures with an information diffusion mechanism. Since its proposal, GNNs have been applied to diverse problem domains such as image classification [2]–[8], object detection [9], scene graph generation [10], and visual question answering [11].

Converting the image data into a graph structure is a delicate matter and imposes many aspects to consider. One of these aspects is the computational cost. As image resolution increases, the cost of processing each pixel as a graph node grows substantially, imposing significant limitations on this direct approach. Therefore, to handle larger images, most methods rely on superpixel algorithms to generate meaningful segmentations with fewer elements, treating each superpixel as a graph node.

However, despite considerable diversity in superpixel techniques, the rapid progress in this area has led to difficulty in catching up with new proposals and a lack of standardization in terms of comparisons with state-of-the-art methods. As a consequence, most works in literature are limited to a simple yet classical superpixel segmentation method called Simple Linear Iterative Clustering (SLIC) [12]. Also explored in recent works [3]–[5], the seed-based Dynamic and Iterative

Spanning Forest (DISF) [13] has proven to be a relevant variation.

Hence, initiatives to diminish this gap in literature have been conducted in [14], [15]. A complete taxonomy categorizing 59 open-source superpixel algorithms within 10 classes based on their respective processing steps is proposed in [15]. It is observed that the SLIC and DISF algorithms fall in the neighborhood and path-based categories, respectively, indicating the need for further evaluation of different methods and taxonomic classes, especially to understand how their different outputs influence GNN training effectiveness for image classification tasks. This arises as a necessity because there is no clear guideline for selecting the segmentation method, as methods may perform differently on distinct tasks and may even be influenced by the nature of the images in the training dataset [15].

To enhance the effectiveness of graph-based image representations, hierarchical graph structures have demonstrated performance gains in image classification tasks [3]–[5], [16], as they enrich representations derived from a base graph, typically constructed using superpixel methods. They allow for multi-scale analysis and the encoding of relationships between regions at different levels of abstraction. A significant limitation in most hierarchical methods, however, is their inability to differentiate between spatial and hierarchical edges and thus leverage properly the particular contextual information they encode.

In this study, we focus on the HiErarchical Layered Multigraph Network (HELMNet) [5], a GNN model designed to effectively leverage hierarchical information through multigraph representations, termed HiErarchical Layered Multigraphs (HELM). HELM is particularly noteworthy as it explicitly encodes both spatial and hierarchical edges as distinct edge types, enabling the model to separately exploit information across scales and within regions at the same scale.

This work aims to expand the tests performed in [5] by training the HELMNet model with HELM representations created from distinct superpixel techniques, while preserving the original parameters, and observing how its behavior changes. The Region Graph Readout (RGR), proposed in [4], is additionally used to investigate the significance of each scale within the graph representation. This could provide deeper insights into HELMNet’s classification accuracy across data

generated by different superpixel techniques.

The remainder of this paper is organized as follows. Section II provides context for this research with a commentary on the related works. In Section IV, we outline the experimental scenario and justify the method selection for the initial evaluation. The current state of the research is covered in Section V, followed by a glimpse at the preliminary findings in VI. Finally, Section VII draws some conclusions and points towards our next steps.

II. RELATED WORKS

A variety of procedures can be applied to represent images in the form of graphs. As each method has its advantages and downsides, a decision whether to use one or another should be made based on a series of criteria and depending on the specific image task. For example, the fixed-size patches approach categorizes each patch as a node in the graph, effectively framing local image regions [17], [18]. At the other extreme, each singular pixel of an image may as well be considered as a node, resulting in dense graphs with spatial relationships encoded to a higher level of information [19].

A predominant technique in literature is the application of superpixel segmentation algorithms. Each superpixel is then treated as a node in the graph [2], [3], [6], [7], [16], [20], [21]. Although the definition lacks consensus, a superpixel consists of a union of individual pixels that have a similar characteristic value, such as color, texture, and spatial proximity, and can be created following a variety of clustering strategies.

Research efforts to benchmark and classify these methods within taxonomic classes have been conducted in [14], [15], softening the lack of experimental parameters between methods. However, in the context of image classification with Graph Neural Networks (GNNs), the performance documentation for most of these methods is non-existent, and many works are limited to the application of the SLIC method [6], [7], [16], [20], [21], thanks to its broad availability and fast execution [12].

Presenting notoriety among more recent works [3]–[5], the DISF superpixel algorithm [13] is evaluated, showing improved performance when compared to SLIC. Another approach, the multiscale WaveMesh, was proposed and evaluated with a GNN model for image graph classification in [7]. It introduces a compelling conceptual premise by systematically determining the number and size of superpixels based on the image content. In [7], WaveMesh imposed improvements on image classification in comparison to SLIC as well.

Graph Neural Networks capable of interpreting hierarchical representations are equally important for this work [5], [16], [18], [21]. These models can process multi-scale graph structures, enabling the exploitation of information across different levels of granularity. A distinctive characteristic of some of these models is their differentiation between spatial edges connecting nodes within the same scale and hierarchical edges linking nodes across different scales [5], [16]. This dual-edge architecture allows the network to maintain spatial coherence

while enabling cross-scale information propagation, resulting in more comprehensive image understanding.

This work aims to integrate superpixel segmentation, hierarchical representation, and GNN-based models for image recognition using HELMNet [5] on the STL-10 dataset. The HELMNet is trained with HELM representations that leverage the watershed-by-area [22] hierarchy to create a multi-graph structure, where each level corresponds to a different segmentation, forming a comprehensive hierarchical representation of the input image.

III. BACKGROUND

A. HELM – HiErarchical Layered Multigraph

The HELM construction begins with an input image, which is segmented into superpixels. Each superpixel is represented as a node in the graph, and edges are added to connect spatially neighboring superpixels, forming the base Region Adjacency Graph (RAG). This base graph is then processed through a hierarchical segmentation method, producing a hierarchy of partitions $\mathcal{H} = (\mathbf{P}_0, \dots, \mathbf{P}_k)$ where \mathbf{P}_{i-1} is a refinement of \mathbf{P}_i (i.e., every element or region in \mathbf{P}_{i-1} is entirely contained within an element of \mathbf{P}_i) for any i in $\{1, \dots, k\}$, and \mathbf{P}_k is the set of superpixels. Horizontal cuts \mathcal{C} are made within the hierarchy, and for each cut $c_j \in \mathcal{C}$, a RAG is built using the regions \mathcal{R}_j from the corresponding partition \mathbf{P}_j . The set of all regions from every partition is denoted by $\mathcal{R} = \bigcup_{i=0, \dots, |\mathcal{C}|} \mathcal{R}_j$.

From these hierarchical cuts and the base RAG, multiple region graphs $G_j^{\mathcal{R}} = (\mathcal{R}_j, E_j^{\mathcal{S}})$ are obtained, where $E_j^{\mathcal{S}}$ encodes spatial adjacency between regions. The final representation is defined as $\text{HELM} = (\mathcal{R}, E^{\mathcal{S}} \cup E^{\mathcal{H}})$, where $E^{\mathcal{S}}$ contains all spatial connections and $E^{\mathcal{H}}$ encodes hierarchical relations between regions across different levels of the hierarchy.

B. HELMNet – HiErarchical Layered Multigraph Network

The Hierarchical Layered Multigraph Network (HELMNet) builds on the HELM representation to efficiently process spatial and hierarchical relationships within a multigraph. It takes node features h_i , spatial edge features $w_{ij}^{\mathcal{S}}$, and hierarchical edge features $w_{ij}^{\mathcal{H}}$ as inputs, embedding them into a unified feature space via linear layers. The core of HELMNet is the Multigraph Processing (MGP) module, which independently handles two relationship types: spatial (S) and hierarchical (H). Each MGP module transforms and updates the node and edge features for both relationship types separately. This is achieved through a process of linear transformations, batch normalization, and graph convolutions.

At the end of each MGP module, the distinct node feature vectors for each relationship are fused into a unified representation using a relation fusion layer. This representation then serves as the input for the next MGP module. After three MGP modules, HELMNet applies the Region Graph Readout (RGR) method to generate a fixed-size graph representation. This representation is passed through a Classification Layer composed of two linear transformations, producing the final scene classification output. By explicitly modeling and integrating

spatial and hierarchical dependencies, HELMNet effectively captures multi-scale relational patterns in remote sensing data.

C. Region Graph Readout

The Region Graph Readout (RGR) [3], [4] is an attention mechanism capable of capturing the importance of each region graph within the hierarchical representation by reweighting graph features based on scale. This dynamic process refines the representation for classification and enables the model to focus on more semantically meaningful scales while neglecting less informative ones.

First, node features are embedded into a higher-dimensional space using a linear layer with the ReLU activation. Then, to generate a fixed-size vector for each region graph ($G_j^{\mathcal{R}}$), the features of all nodes within the same region graph are summed, as follows, $h_{G_j^{\mathcal{R}}} = \sum_{r_i \in \mathcal{R}_j} \text{ReLU}(S h_{r_i}^3 / \sqrt{|\mathcal{R}_j|}, \forall j = 0, \dots, |\mathcal{C}|$, in which $h_{r_i}^3$ stands for the features of node r_i after the third MGP module, S represents a linear transformation, and $|\mathcal{R}_j|$ denotes the size of region \mathcal{R}_j in number of nodes. To prevent imbalance due to the size of larger graphs (such as base graph), node features are normalized using the graph size norm [2].

The importance coefficients are employed to weigh and aggregate the corresponding region graph features. This produces a fixed-size feature vector, denoted as h_G , that summarizes the entire graph, $h_G = \sum_{j=0}^{|\mathcal{C}|} \hat{\sigma}(h_{G_j^{\mathcal{R}}}) \odot h_{G_j^{\mathcal{R}}}$, in which $\hat{\sigma}$ is the *softmax* function.

IV. EXPERIMENTAL SETUP

This section presents the STL-10 dataset and discusses the superpixel method selection to expand the work done in [5] with a brief commentary on the reason for choosing to use each method alongside its computational complexity, when provided by the authors.

A. The STL-10 Dataset

Stanford University’s STL-10 dataset is specialized for image classification tasks and comprises 13,000 color images with dimensions of 96×96 pixels, organized into ten distinct classes. The additional 100,000 unlabeled images provided in the dataset were not utilized in [5].

The data distribution was organized as follows: 4,500 images for training, 500 for validation, and 8,000 for testing. According to the methodology established in [2], the 500 validation images were randomly sampled from the training set to form the validation set, with this split being maintained in all experiments. A comprehensive data augmentation strategy was applied, generating 51 variants for each image in the training set.

Following the dataset split and data augmentation, the DISF method was applied to generate segmentations with target superpixel counts of 50, 185, and 250, using 2,000 initial seeds. The resulting segmentations were then employed to construct the RAG, which served as the base graph for deriving the HELM representation. Therefore, this setting will be consistently maintained when evaluating the supplementary methods listed below.

TABLE I
CHARACTERISTICS OF SUPERPIXEL METHODS TO BE TESTED.

† INDICATES THE TIME COMPLEXITY COVERS ONLY THE SEGMENTATION PROCESS, AND ‘—’ DENOTES NOT REPORTED DATA.

| Method | Taxonomic Class | Additional Data | Superpixel | Time Complexity | Target Number |
|--------|--------------------|-----------------|--|-----------------|---------------|
| DISF | Path-based | × | Boundary adherent | $O(n \log n)$ | ✓ |
| ODISF | Path-based | Saliency map | Boundary adherent, concentrated on the salient area | $O(n \log n)$ † | ✓ |
| SLIC | Neighborhood-based | × | Balanced compactness and boundary adjustment | $O(n)$ | × |
| SCALP | Neighborhood-based | Boundary map | Balanced compactness and boundary adjustment, enhanced color homogeneity | — | × |

B. Superpixel Method Selection

The three superpixel algorithms selected to extend the job initiated with DISF are distributed within two classes of the taxonomy adopted from [15]: path-based and neighborhood-based. Both DISF and its object-oriented version ODISF [23] are path-based due to their behavior of creating paths in the image following some criteria. SLIC and Superpixels with Contour Adherence using Linear Path (SCALP) [24], on the other hand, perform clustering with a spatial distance function that calculates the distance from a pixel to a reference point in the image. Table I presents a comparison of fundamental aspects of all four methods for elucidation.

The path-based strategy employed by DISF and ODISF produces regions that closely adhere to image boundaries while exhibiting low compactness and irregular shape–size distributions. This behavior can thus approximate the image’s interest regions, a characteristic that could be conceptually leveraged in image classification tasks, explaining the application of both methods. DISF was originally used due to its straightforward seed oversampling during initial processing. ODISF, in contrast, despite sharing the $O(n \log n)$ time complexity with DISF, requires a saliency map to guide the recomputation of seeds from its original segmentation [23]. This additional computational overhead can be justified by improved identification of discriminative image regions.

Inspired by the k -means clustering algorithm, SLIC is widely adopted due to its availability and $O(n)$ complexity, bounded only by the target superpixel count. Being categorized in the neighborhood-based, SLIC generates superpixels with balanced compactness and image boundary adjustment, having preferable performance with a higher number of superpixels [12].

The spatial distance function adopted in SCALP when clustering a pixel into a superpixel is enhanced by pondering image feature and contour intensity from the pixel to the region barycenter [24]. Therefore, at the cost of higher CPU use [15] and additional effort to create the preliminary boundary map, SCALP presents better delineation, color homogeneity, and compactness than SLIC. Overall, both methods generate more compact segmentations than DISF and ODISF. But while compactness appears less critical for classification tasks, we include these methods as experimental baselines with the endeavor to investigate the superiority of boundary-adherent superpixels.

V. CURRENT STATE OF THE RESEARCH

The superpixel algorithms utilized in this research are open-source and can be easily cloned from their official GitHub repositories. SLIC is readily available through the scikit-image library, while DISF provides Python integration through its Makefile configuration. In contrast, for the implementation of ODISF and SCALP algorithms, custom Python scripts were developed to handle data transmission via command-line interface, as the original authors did not provide direct integration with the Python programming language.

It should be noted that no novel methods were proposed or implemented throughout this research. The study focuses on the comparative analysis and evaluation of existing superpixel segmentation techniques rather than methodological innovation.

Following the completion of all experiments with the SLIC superpixel method, ODISF will be the next algorithm to be evaluated. The ODISF implementation is fully prepared with all necessary code components, including dataset definition with data augmentation for the training set, saliency map generation using U²Net, and segmentation pipeline with the ODISF binary ready for execution. However, there remains the need to prepare the experimental materials for SCALP evaluation, which will be conducted last in the experimental sequence.

Overall, the four selected algorithms represent only two of the superpixel method classes described in [15], and the results obtained will guide the subsequent evaluation of additional methods. This approach is adopted because, despite the existence of multiple taxonomic classes, some share common characteristics. For instance, similar to the path-based class, the dynamic-center-update class contains methods that generally provide good image delineation. Therefore, if DISF and ODISF demonstrate superior performance, it would be suitable to include tests with methods from this class rather than the boundary evolution-based class, which contains methods with more compact results similar to SLIC and SCALP from the neighborhood-based class.

Before proceeding, it should be noted that these algorithms were chosen based on theoretical grounds presented in [15] and in their respective papers [12], [13], [23], [24]. Given their varied performance across the Birds, Insects, Sky, ECSSD, and NYUV2 datasets, their application to STL-10 may yield unpredictable results. Additionally, their inherent characteristics may not align well with HELM’s hierarchical properties, which could impact the final performance.

VI. PRELIMINARY RESULTS AND OBSERVATIONS

As shown in Table II, HELMNet₁₈₅ achieved a significant increase of 5.27% in average accuracy over HIGSI₁₈₅ when using SLIC as the superpixel method. However, both models still underperform compared to those trained on graphs generated with DISF. This result suggests two hypotheses: (i) the lower compactness provided by the DISF method is more suitable for image classification, as it allows a finer alignment with image boundaries; and (ii) even with an overall lower

TABLE II
COMPARISON WITH STATE-OF-THE-ART GNN METHODS ON THE STL-10 DATASET. † INDICATES THE TARGET NUMBER OF NODES BECAUSE THE AVERAGE VALUE WAS NOT REPORTED, ‘—’ DENOTES NOT REPORTED DATA, AND ‘×’ STANDS FOR NOT APPLICABLE.

| Model | # params | Superpixel Method | # nodes | # edges | Avg. Accuracy (± std dev.) | # Region Graphs |
|----------------------------|----------|-------------------|---------|---------|-------------------------------|--------------------|
| GCN [8] | — | SLIC | 400† | — | 0.5210 (±0.004) | × |
| HIGSI ₅₀ [4] | 105,226 | DISF | 69.77 | 491.40 | 0.5608 (±0.006) | 6.74 |
| HIGSI ₁₈₅ [4] | 105,226 | DISF | 260.36 | 2231.07 | 0.5847 (±0.009) | 12.36 |
| HIGSI ₂₅₀ [4] | 105,226 | DISF | 353.53 | 3150.68 | 0.5815 (±0.004) | 14.33 |
| HIGSI ₅₀ [4] | 105,226 | SLIC | 71.25 | 491.24 | 0.5151 (±0.009) | 7.34 |
| HIGSI ₁₈₅ [4] | 105,226 | SLIC | 280.94 | 2368.54 | 0.5469 (±0.008) | 13.76 |
| HIGSI ₂₅₀ [4] | 105,226 | SLIC | 366.52 | 3197.53 | 0.5481 (±0.012) | 15.51 |
| HELMNet ₅₀ [5] | 483,658 | DISF | 78.15 | 3150.68 | 0.5941 (±0.006) | 6.74 |
| HELMNet ₁₈₅ [5] | 483,658 | DISF | 288.32 | 3150.68 | 0.6307 (±0.013) | 12.36 |
| HELMNet ₂₅₀ [5] | 483,658 | DISF | 382.66 | 3490.68 | 0.6419 (±0.006) | 14.61 |
| HELMNet ₅₀ | 483,658 | SLIC | 163.74 | 1382.28 | 0.5176 (±0.004) | 9.35 |
| HELMNet ₁₈₅ | 483,658 | SLIC | 321.89 | 2798.82 | 0.5757 (±0.005) | 14.18 |
| HELMNet ₂₅₀ | 483,658 | SLIC | 386.27 | 3423.02 | 0.5727 (±0.010) | 15.50 |

performance with SLIC superpixels, satisfactory gains are still observed when treating spatial and hierarchical connections as distinct edge types, an assignment HELMNet excels in.

HELMNet₅₀ and HELMNet₂₅₀ further support these hypotheses, leading to additional insights. With a target of 50 superpixels, a slight increase in average accuracy can be observed (hypothesis ii). However, as widely reported, SLIC performs poorly with a small number of superpixels since it prioritizes compactness over boundary adherence (hypothesis i) [12], [15]. Increasing the number of SLIC superpixels to 250 results in a more pronounced accuracy gain compared to HIGSI₂₅₀, attributable not only to the strengths of HELMNet (hypothesis ii) but also to improved alignment with image boundaries. Nevertheless, in comparison to HELMNet₁₈₅, in the SLIC setup, the additional superpixels did not yield further improvements with HELMNet₂₅₀.

Another open question regarding hierarchical representations is the ideal number of scales within the hierarchy. A deeper hierarchy may indicate limitations of the HELM model in merging similar regions, whereas fewer scales may result in insufficient contextual information for an effective attention distribution by the RGR function. To address this, we should investigate the HELM representations across DISF, SLIC, ODISF, and SCALP. At this stage of our work-in-progress, we have observed that HELM tends to produce hierarchies with fewer layers when the base graph is segmented with DISF compared to SLIC (see the # Region Graphs column in Table II). This occurs because SLIC generates more compact superpixels with less constrained color homogeneity. Consequently, this pattern appears to hinder HELM’s region-merging process, requiring additional steps to reach the root node of the hierarchy, which represents the complete image.

This observation on the hierarchy’s depth, however, is not necessarily related to the superior classification performed by HELMNet when the origin of the hierarchy is a DISF segmentation. It can, at least, indicate that HELM creates better representations across scales for classification purposes when the base graph contains boundary-adherent superpixels with color homogeneity. Further conclusions can only be drawn

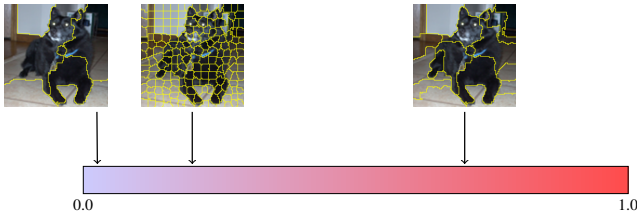


Fig. 1. Distribution of attention between scales of the HELMNet.

when HELM representations derived from segmentations of additional superpixel methods are analyzed in the continuation of this work-in-progress.

To illustrate how the RGR operates, Figure 1 presents an example of the attention distribution across different scales with base graphs originating from SLIC. The scale with the highest attention value is intermediary; the RGR was found to be most meaningful due to its better approximation of the discriminative region. However, merging more regions exhibited loss of delineation precision, and the received attention got close to zero with a value smaller than the one received by the base RAG with the largest number of regions. To aid visualization, insignificant attention values were omitted.

VII. CONCLUSION

This work analyzes the behavior of various superpixel methods used to generate a base graph for representing images. It also tries to identify the one that, when applied in HELM, yields the best overall performance of the HELMNet classification model.

Just as HELMNet presents improvements over other hierarchical models, such as HIGSI by considering spatial and hierarchical edges as different types and performing customized extraction of what they have to offer. Moreover, preliminary results indicate that HELMNet using SLIC, though inferior to the HELMNet + DISF combination, still outperforms HIGSI + SLIC due to proper treatment of hierarchical representation peculiarities.

Here, the RGR function, beyond its main role of model decision-making by re-weighting and finding the most representative scales, will also help us understand the performance differences of each superpixel segmentation method and comprehend how their inherent characteristics affect HELM creation. One point we must address in the future is the ideal number of scales in a hierarchy for image classification, which remains an open question. Many scales may represent difficulty for HELM in merging regions due to a lack of context, but on the other hand, a few regions may cause indecision in RGR due to insufficient information for analysis.

Finally, after results with SLIC, DISF, ODISF, and SCALP are obtained and properly interpreted, we should include other methods in the study chosen based on our findings. To match HELM's properties, for example, methods classified within the hierarchy-based class would signify a prudent conceptual choice to be thought into this research in the future.

ACKNOWLEDGMENT

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (CAPES STIC-AMSUD 88887.878869/2023-00, 23-STIC-10 and Finance Code 001), the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Grants 407242/2021-0, 306573/2022-9, 442950/2023-3) and Fundação de Apoio à Pesquisa do Estado de Minas Gerais – FAPEMIG (Grant APQ-01079-23, APQ-05058-23 and PCE-00301-25).

REFERENCES

- [1] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.
- [2] V. P. Dwivedi, C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson, "Benchmarking graph neural networks," *J. Mach. Learn. Res.*, vol. 24, no. 1, Jan. 2023.
- [3] J. P. O. Batisteli, S. J. F. Guimarães, and Z. K. G. do Patrocínio Júnior, "Multi-scale image graph representation: a novel gnn approach for image classification through scale importance estimation," in *IEEE International Symposium on Multimedia (ISM)*, 2023.
- [4] J. P. O. Batisteli, S. J. F. Guimarães, and Z. K. G. Patrocínio, Jr, "Hierarchical graph neural networks with scale-aware readout for image classification," *International Journal of Semantic Computing*, vol. 18, no. 04, pp. 713–738, 2024.
- [5] J. P. O. Batisteli, S. J. F. Guimarães, and Z. K. G. do Patrocínio Júnior, "Hierarchical layered multigraph network with scale importance estimation for image classification," *Applied Soft Computing*, 2025, submitted.
- [6] P. H. Avelar, A. R. Tavares, T. L. da Silveira, C. R. Jung, and L. C. Lamb, "Superpixel image classification with graph attention networks," in *33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2020, pp. 203–209.
- [7] V. Vasudevan, M. Bassenne, M. T. Islam, and L. Xing, "Image classification using graph neural network and multiscale wavelet superpixels," *Pattern Recognition Letters*, vol. 166, pp. 89–96, 2023.
- [8] J. Rodrigues and J. Carbonera, "Graph convolutional networks for image classification: Comparing approaches for building graphs from images," in *Proceedings of the 26th International Conference on Enterprise Information Systems-Volume 1: ICEIS*, 2024, pp. 437–446.
- [9] K. Li, D. DeTone, Y. F. S. Chen, M. Vo, I. Reid, H. Rezatofighi, C. Sweeney, J. Straub, and R. Newcombe, "Odam: Object detection, association, and mapping using posed rgb video," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5998–6008.
- [10] M. Khademi and O. Schulte, "Deep generative probabilistic graph neural networks for scene graph generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 237–11 245.
- [11] W. Liang, Y. Jiang, and Z. Liu, "GraghVQA: Language-guided graph neural networks for graph-based visual question answering," in *Proceedings of the Third Workshop on Multimodal Artificial Intelligence*, Jun. 2021, pp. 79–86.
- [12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [13] F. C. Belém, S. J. F. Guimarães, and A. X. Falcão, "Superpixel segmentation using dynamic and iterative spanning forest," *IEEE Signal Processing Letters*, vol. 27, pp. 1440–1444, 2020.
- [14] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.
- [15] I. B. Barcelos, F. D. C. Belém, L. D. M. João, Z. K. G. D. Patrocínio, A. X. Falcão, and S. J. F. Guimarães, "A comprehensive review and new taxonomy on superpixel segmentation," *ACM Comput. Surv.*, vol. 56, no. 8, Apr. 2024.
- [16] B. Knyazev, X. Lin, M. Amer, and G. Taylor, "Image classification with hierarchical multigraph networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2019, pp. 223.1–223.13.

- [17] M. Munir, W. Avery, M. M. Rahman, and R. Marculescu, "GreedyViG: Dynamic axial graph construction for efficient vision GNNs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 6118–6127.
- [18] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision gnn: an image is worth graph of nodes," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022.
- [19] G. Nikolentzos, M. Thomas, A. R. Rivera, and M. Vazirgiannis, "Image classification using graph-based representations and graph neural networks," in *Complex Networks & Their Applications IX: Volume 2, Proceedings of the Ninth International Conference on Complex Networks and Their Applications*, 2021, pp. 142–153.
- [20] R. A. Cosma, L. Knobel, P. van der Linden, D. M. Knigge, and E. J. Bekkers, "Geometric superpixel representations for efficient image classification with graph neural networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 109–118.
- [21] J. Long, Z. yan, and H. chen, "A graph neural network for superpixel image classification," *Journal of Physics: Conference Series*, vol. 1871, no. 1, p. 012071, apr 2021.
- [22] J. Cousty and L. Najman, "Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts," in *ISMM*, 2011, pp. 272–283.
- [23] F. C. Belém, B. Perret, J. Cousty, S. J. F. Guimarães, and A. X. Falcão, "Towards a simple and efficient object-based superpixel delineation framework," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2021, pp. 346–353.
- [24] R. Giraud, V.-T. Ta, and N. Papadakis, "Scalp: Superpixels with contour adherence using linear path," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2374–2379.