

# Open-set semantic segmentation for remote sensing images

Ian Nunes<sup>\*†‡</sup>, Hugo Oliveira<sup>§</sup> and Marcus Poggi<sup>‡</sup>

<sup>†</sup>Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil

Email: ian.nunes@ibge.gov.br

<sup>§</sup>Departamento de Informática

Universidade Federal de Viçosa, Viçosa, Brazil

<sup>‡</sup>Department of Informatics

Pontifícia Universidade Católica, Rio de Janeiro, Brazil

**Abstract**—Collecting samples that exhaust all possible classes for real-world tasks is usually difficult or impossible due to many different factors. In a realistic/feasible scenario, methods should be aware that the training data is incomplete and that not all knowledge is available. Therefore all developed methods should be able to identify the unknown samples while correctly executing the proposed task to the known classes in the tests phase. Open-Set Recognition and Semantic Segmentation models emerge to handle this kind of scenario for, respectively, visual recognition and dense labeling tasks. Initially, this work proposes a novel taxonomy aiming to organize the literature and provide an understanding of the theoretical trends that guided the existing approaches that may influence future methods. This work also proposes two distinct techniques to perform open-set semantic segmentation. First, a method called Open Gaussian Mixture of Models (OpenGMM) extends the Open Principal Component Scoring (OpenPCS) framework using a Gaussian Mixture of Models to model the distribution of pixels for each class in a multimodal manner. Second, the Conditional Reconstruction for Open-set Semantic Segmentation (CoReSeg) method tackles the issue using class-conditioned reconstruction of the input images according to their pixel-wise mask. The third proposed approach is a general post-processing procedure that uses superpixels to enforce highly homogeneous regions to behave equally, rectifying erroneously classified pixels within these regions. We also proposed a novel superpixel generation method called Fusing Superpixels for Semantic Consistency (FuSC). All proposed approaches produce better semantic consistency and outperformed state-of-the-art baseline methods on Vaihingen and Potsdam ISPRS dataset.

The official implementation of all proposed approaches is available at <https://github.com/iannunes>.

## I. INTRODUCTION

Closed-set visual learning tasks are limited due to the difficulty of collecting labeled or classified training samples that exhaust all possible classes in the real world. The expected scenario in real-world problems is Open with objects of classes not seen during training that may be submitted to the model during the deployment phase [1].

Incomplete knowledge of the world during training or the existence of unknown samples during inference is a major unsolved challenge. So-called Open Set Recognition (OSR) tasks have caught the interest of the research community

with multiple methods recently proposed for classification problems [2]–[6]. However, few publications have dealt with open visual tasks other than scene/object classification, such as segmentation or object detection.

Semantic Segmentation is often defined as a classification problem concerned with attributing a class prediction to each individual pixel. The Open version of Semantic Segmentation is called Open-set Semantic Segmentation (OSS) [7]. It refers to the set of algorithms that address the identification of pixels of unknown or out-of-distribution (OOD) classes at inference time while correctly classifying pixels of the known classes (KCC) learned in training [7]. OSS is an inherently harder problem due to its dense labeling nature compared to Open-set classification. It is hard to learn open-set semantic segmentation precisely in real-world scenarios [8]. This fact may explain why there is still a gap in the literature with only a handful of articles tackling the issue [5].

In this work, we focus on finding methods that improve segmentation results and semantic consistency in OSS tasks. Improving the semantic consistency allows the deployment in real-world scenarios, as models with poor semantic consistency can result in avoidable errors and render them unusable.

### A. Contributions

This thesis presents five contributions as follows:

- 1) a systematic mapping of the literature for OSS with the proposal of a taxonomy to organize the related literature;
- 2) an OSS method called OpenGMM;
- 3) a novel end-to-end fully convolutional method for OSS called CoReSeg;
- 4) a general superpixel post-processing technique for OSS; and
- 5) a novel superpixel generation algorithm called FuSC.

We propose a taxonomy to organize and assist in better understanding the existing literature and current trends in deep open-set segmentation [9]. The systematic mapping extended the proposed taxonomy and was published with a more comprehensive set of papers [10].

Our two OSS methods improved baseline quantitative results and semantic consistency. OpenGMM [11] is a modification of the previously proposed OpenPCS [7], replacing PCA

\*This work is related to a Ph.D. thesis.

by GMM to represent the compressed feature space. CoReSeg is fully a novel end-to-end fully convolutional method [12] for OSS based on pixelwise conditional reconstruction.

The general superpixel post-processing strategy improved the quantitative results and the semantic consistency of all tested OSS. At last, post-processing with FuSC [11] improved the robustness of hyperparameter selection while producing better, more stable, and reliable results.

## II. CONTRIBUTION 1: OSS SURVEY AND TAXONOMY

Aiming to better understand current trends, the selection of articles guided us to the following taxonomy, mapping three identified paradigms that organize the families of methods for OSR and OSS commonly found in the literature:

- 1) *Statistical modeling*: statistics of the intermediary and output activations from the networks are used to define in- and out-of-distribution samples [2], [3], [5], [7], [13]–[24], it is possible to further split it in four overlapping subdivisions according to the characteristics of the statistical modeling - which activation layers are used, the employment of Extreme Value Theory (EVT), the use of activations to represent known and unknown classes (UUC), and the output of an anomaly (entropy or probability) score;
- 2) *Reconstruction-based*: image reconstruction loss is used to model or classify OOD samples [3], [4], [12], [25] This category can be further split into two subdivisions - Conditional or not. The conditional sub-group is characterized by the employment of class conditioning as a mean of reconstructing an input image according to the desired condition, which is a strategy that tends to generate worst reconstructions for the OOD classes due to unknown adequate conditioning;
- 3) *Auxiliary data*: when known unknown samples are available, one can use them to turn a generative model for OSR/OSS into a discriminative distinction [14], [20], [26]–[29]. This category can be split into two subdivisions - Synthetic or not. The Synthetic methods use some type of generative strategy to generate OOD samples, helping to better model in- and out-of-distribution samples.

A graphical visualization of all selected papers under the respective category is shown in Figure 1.

## III. CONTRIBUTION 2: OPEN GAUSSIAN MIXTURE OF MODELS

Open Gaussian Mixture of Models (OpenGMM) processes intermediate feature maps with the last layers’ activation maps of a deep neural network. Combining the activations from earlier layers with final layers produces a tensor that fuses low and high semantic level information. The concatenated tensor may have hundreds or thousands of channels, which are known to contain redundant information [3], [31]. OpenGMM handles the concatenated tensor size and redundancy by fitting a GMM on each known-class distribution. Each GMM model computes a score tensor with the log-likelihood values for all pixels,

| Ref  | T | D  | R | A | G | S | F | P | E | SE    |
|------|---|----|---|---|---|---|---|---|---|-------|
| [2]  | R | I  | × | × | × | ✓ | × | ✓ | ✓ | M     |
| [13] | R | I  | × | × | × | ✓ | × | × | ✓ | M     |
| [14] | R | I  | × | ✓ | × | ✓ | × | × | × | M     |
| [25] | R | I  | ✓ | × | ✓ | × | × | × | × | M     |
| [4]  | R | I  | ✓ | × | ✓ | × | × | × | ✓ | M     |
| [3]  | R | I  | ✓ | × | ✓ | ✓ | ✓ | × | × | M     |
| [16] | R | I  | × | × | ✓ | ✓ | ✓ | ✓ | × | M     |
| [30] | R | I  | × | × | × | × | × | × | × | W     |
| [22] | R | RS | × | × | × | ✓ | × | × | × | W     |
| [15] | S | RS | × | × | × | ✓ | × | ✓ | ✓ | S     |
| [26] | S | I  | × | ✓ | ✓ | × | × | × | × | G     |
| [5]  | S | I  | × | × | × | ✓ | × | × | × | S,W,G |
| [7]  | S | RS | × | × | ✓ | ✓ | ✓ | ✓ | ✓ | S,G   |
| [17] | S | RS | × | × | ✓ | ✓ | ✓ | ✓ | ✓ | S,G   |
| [18] | S | I  | × | × | × | ✓ | × | × | × | S     |
| [19] | S | I  | × | × | × | ✓ | ✓ | × | × | S,G   |
| [20] | S | I  | × | ✓ | ✓ | ✓ | × | × | × | G     |
| [21] | S | I  | × | × | × | ✓ | × | × | × | W     |
| [27] | S | I  | × | ✓ | ✓ | × | ✓ | × | × | G     |
| [12] | S | RS | ✓ | × | ✓ | × | × | × | × | G     |
| [28] | S | I  | × | ✓ | × | × | ✓ | × | × | S,G   |
| [29] | S | I  | × | × | × | ✓ | × | × | × | G     |
| [23] | S | I  | × | × | × | ✓ | × | × | × | G     |
| [24] | S | I  | × | × | × | ✓ | × | × | × | G     |

TABLE I: The table shows systematic review results for OSS and the selected articles of OSR. Data is ordered by task (column T) and by publish year. Columns stand for, respectively: T - main task tackled (S - segmentation, R - recognition); D - data type (I - 2D image, RS - remote sensing image); R - if the model uses image reconstruction somehow; A - if it uses auxiliary data; G - if it uses generative modeling; S - if it uses any statistical modeling; F - if it uses the intermediate feature space to model open-set distributions; P - if the model can be used in a plug & play fashion; E - if the method uses EVT to model open-set distributions; and SE - the source of the article (M - manually included; W - Web of Science; S - Scopus; and G - Google Scholar).

which allows for the computation of a final score tensor by combining all GMM scores with the closed-set prediction. All pixels below a certain threshold in the final score are identified as unknown.

We performed tests with three different backbones as the closed-set segmentation method: DenseNet-121 (DN-121) [31], WideResNet-50 (WRN-50) [32] and U-net [33].

Readers should notice that adapting any pretrained closed-set semantic segmentation network to the OpenGMM framework is relatively quick and straightforward and does not require retraining the neural network. The only trainable component in our framework is the GMM to fit into the data, which is considerably faster than retraining a neural network. The plug-and-play characteristic of the method is a great advantage when considering the problem of adapting the solution to real-world applications and novel domains.

## IV. CONTRIBUTION 3: CONDITIONAL RECONSTRUCTION FOR OPEN-SET SEMANTIC SEGMENTATION

Conditional Reconstruction training aims at reconstructing the image that serves as input to the closed-set semantic

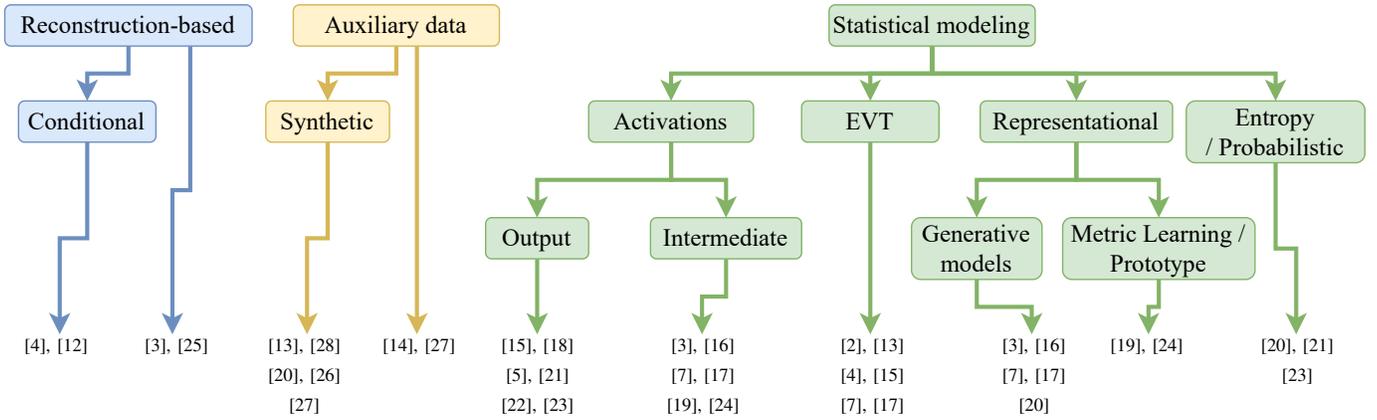


Fig. 1: Classification of the selected works under the proposed taxonomy. Each category can be further divided into more refined groups according to the methods’ characteristics. Each method may fall under more than one group, as they are not mutually exclusive.

segmentation block of the framework from its latent representation. The reconstruction is guided by a conditioning input, which, in the case of a semantic segmentation task, is comprised of a mask providing a class for each pixel. The conditional reconstruction block of the framework can be seen as an auto-encoder where the conditioning layers are the encoder, and the reconstruction layers are the decoder. Figure 3 shows how these different parts of the training are connected.

Aiming to enforce the conditioning, the encoder from the conditional reconstruction block applies a transformation to the intermediate features from the frozen Closed Set Encoder layers ( $e_i$ ). The result of this transformation is then used as input on the corresponding layer of the reconstruction decoder ( $d_i$ ). In Figure 3, this process is represented by the  $f_i(e_i)$  blocks and it is performed for both match and non-match conditioning masks. The transformation responsible for the conditioning is the FiLM method proposed by [34] extended to work in a pixel-wise problem. More specifically, to use the pixel-wise FiLM, the conditional reconstruction decoder is composed of two auxiliary encoders:  $\beta$  and  $\gamma$ . Both encoders have the same shape as  $e_i$  and  $d_i$ . To apply the transformation, we perform the following operation  $\gamma_i \odot e_i + \beta_i$ , where  $\beta_i$  and  $\gamma_i$  are the  $i^{th}$  blocks of the conditional reconstruction encoder and  $e_i$  is the output of the  $i^{th}$  block from the Closed Set Encoder. This procedure allows us to perform pixel-wise FiLM conditioning on  $e_i$ .

During deploy – shown in Figure 4 – we cannot provide match and non-match masks for the conditional encoder, as the labels for these samples are not available. So, to define which pixels are known and unknown CoReSeg tries to condition every pixel for each known class.

The input image is processed by the closed-set semantic segmentation block, generating a closed-set prediction. Then, the reconstruction decoder is conditioned with  $K$  masks, with  $K$  being the number of known classes, where all pixels from the mask  $m_k$  are set as the class  $k$ . Each one of these masks will provide a reconstructed output where all

pixels were conditioned by the class  $k$ , and the corresponding reconstruction loss can be calculated from the input image for all of them.

Then, for each pixel the minimum error for  $k \in \{1, 2, \dots, K\}$  is computed and selected – where  $\{1, 2, \dots, K\}$  is the set of known classes. Pixels that were conditioned to the right class yield a small minimum error, while unknown pixels result in higher error values for each one of the reconstructions, since none of them match the right expected class. At last, a threshold operation defines which pixels are known and unknown. We use error quantiles to set thresholds and find the best performance for the model. If the minimum reconstruction loss of a pixel is below the threshold, its class is deemed as known and set to the closed-set predicted output, and otherwise, it is set as unknown.

## V. CONTRIBUTION 4 AND 5: SUPERPIXEL POST-PROCESSING AND FUSC

Superpixels are commonly employed before or during the segmentation process [35]–[39]. In general, when employed as post-processing, the input image is used to generate the SPS and apply it somehow in the output prediction. This procedure produces more consistent borders among objects and tends to improve semantic consistency for the final segmentation. Following the literature, in the present work, we employ superpixels as a post-processing step applied to the scores returned by the OSS algorithms (i.e. reconstruction error, PCA/GMM likelihood, entropy, heat-map, etc.).

The final superpixel segmentation (SPS) reflects its generation characteristics. We can see in Figure 5 an illustrative example of two SPS that present different characteristics and may represent better different scenarios. The SLIC algorithm could better represent textures, while the FZ algorithm could better identify borders, but none of the single SPS could represent the underlying image properly. Figure 5 also compares the single SPSs with the Fusing Superpixels for the Semantic Consistency method proposed in the next section that

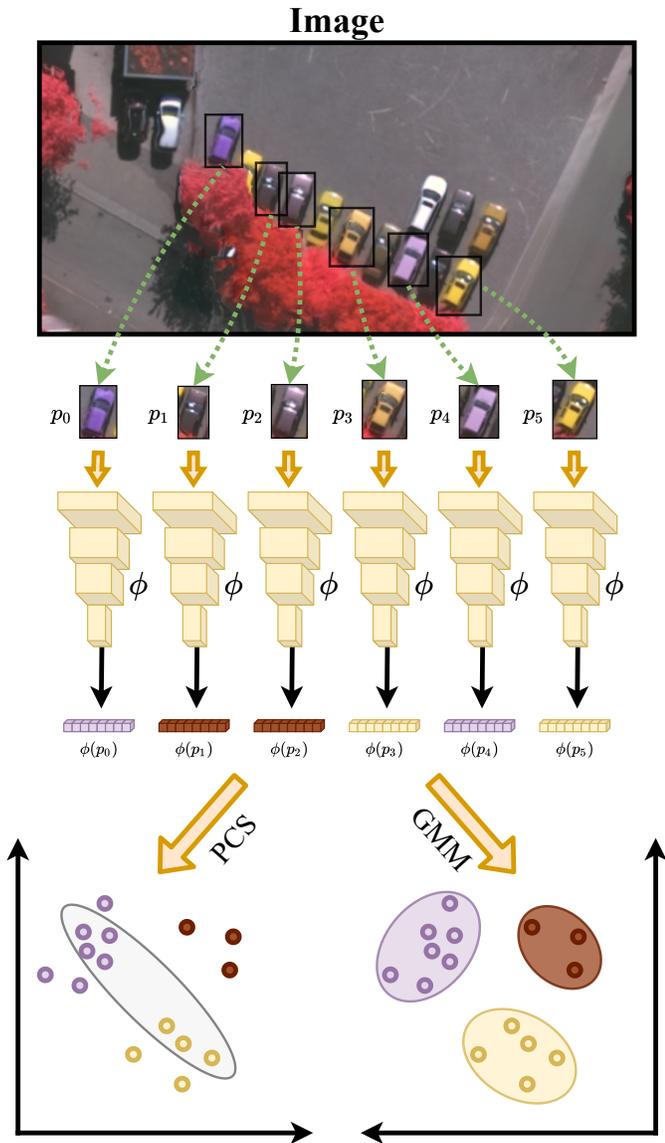


Fig. 2: The figure shows an example of how different objects can be represented by distinct data distributions. Due to the multimodal representation capability, GMM is better suited for representing real-world data than OpenPCS [7].

produces improvement when compared to the single SPSs. The theoretical complexity of the proposed procedure is linear to the number of pixels in the image ( $O(n)$  with  $n$  the number of pixels). Hence its use as post-processing is not computationally expensive and can be coupled with OSS methods to improve the quality of the final produced segmentation prediction.

Figure 6 illustrates the use of FuSC to merge two different superpixel segmentations. As shown in the figure, the final SPS respects both segmentations’ borders, and each segment represents better the underlying region. FuSC is agnostic to the SPS algorithm, being applicable to any set of distinct superpixel algorithms. However, in practice, using more than two algorithms yields exceedingly small segments, motivating

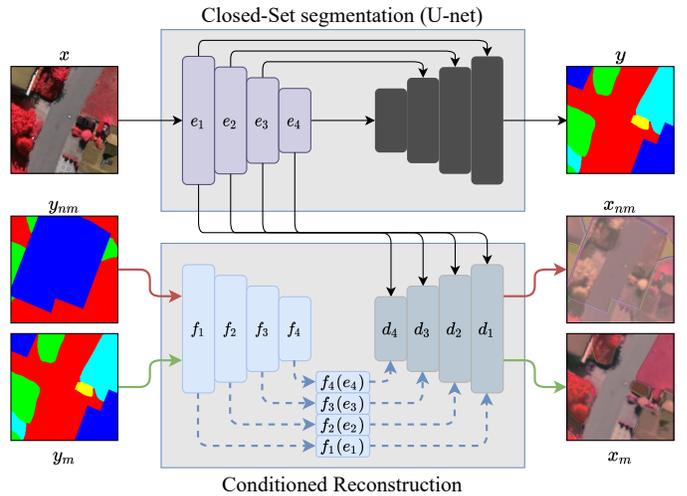


Fig. 3: Training schema where  $e_i$  denotes a layer on the closed-set encoder,  $d_i$  denotes a layer on the reconstruction decoder, and  $f_i$  denotes a simplified FiLM conditioning layer that has two encoders  $\beta$  and  $\gamma$ . The model is trained to reconstruct each image with matching and non-matching masks as a way of enforcing the conditioning with good (match) and poor (non-match) representations of the original image.

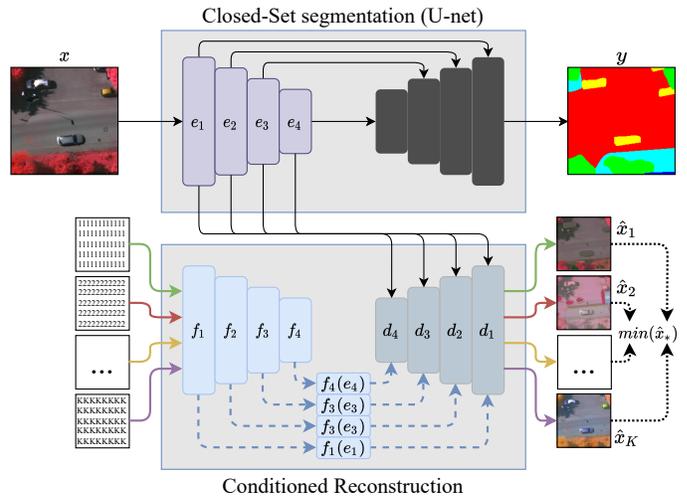


Fig. 4: The figure shows the “Deploy” schema where  $e_i$  denotes a layer on the closed-set encoder,  $d_i$  denotes a layer on the reconstruction decoder, and  $f_i$  denotes a simplified FiLM conditioning layer that has two encoders  $\beta$  and  $\gamma$ .

our experiments to focus only on pairs of algorithms.

## VI. CONCLUSION

This work proposed and described two distinct methods for open-set semantic segmentation: 1) OpenGMM as an extension of a known baseline method called OpenPCS [7] and 2) a novel end-to-end fully convolutional model called

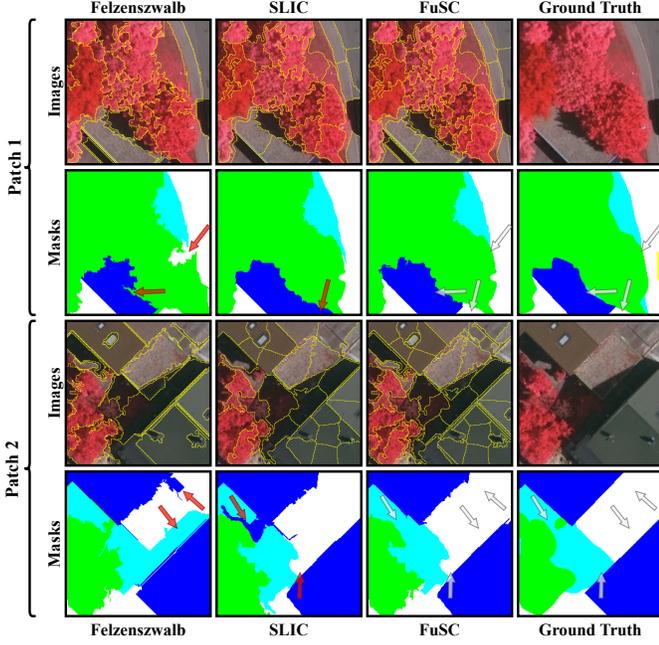


Fig. 5: The figure shows the comparison of the resulting segmentation from two SPS algorithms (Felzenszwalb and SLIC) and our proposed fusion algorithm, FuSC. The first and third rows show the input image superimposed with the superpixel segments and the second and fourth rows depict the closer class fit of each segment according to the real labels. Red arrows indicate areas where class boundaries failed when using one single SPS algorithm, while gray arrows point to these same regions fixed after applying the FuSC algorithm.

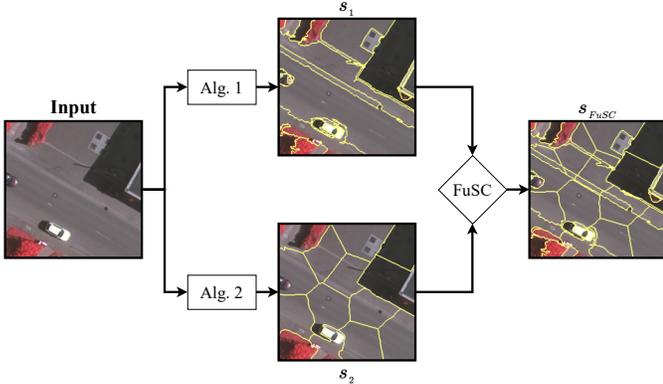


Fig. 6: The figure shows a toy example illustrating the workflow to merge two different superpixel segmentations (SPS). First, the input image  $x$  is processed by 2 different superpixel segmentation algorithms (Alg. 1 and Alg. 2). Then the generated segmentations  $s_1$  and  $s_2$  are merged into the final segmentation  $s_{FuSC}$  using the merging procedure described in Algorithm 1.

**Algorithm 1** Pseudo-algorithm for the FuSC procedure and the auxiliary procedure of joining segmentations. The complexity of the procedure is pseudo-polynomial with respect to the number of pixels in the image and the minimum size of the superpixel.

---

**Require:** scores ▷ pixel-wise array  
**Require:** segments ▷ list of segments

```

1: procedure JOIN_SEGMENTATIONS( $seg1, seg2$ )
2:   joint = []
3:   for  $s1 \in seg1$  do ▷ Selecting  $s2 \in seg2$  where  $s2 \cap s1 \neq \emptyset$ 
4:     for  $s2 \in seg2.OVERLAP\_SEGMENTS(s1)$  do
5:       overlap_area =  $s1 \cap s2$ 
6:       joint.ADD_NEW_SEGMENT(overlap_area)
7:     end for
8:   end for ▷ secure that the labels are connected and sequential
9:   joint ← connected_sequential_labels(joint)
10:  return joint
11: end procedure
12:
13: procedure FuSC( $seg1, seg2$ )
14:  joint = JOIN_SEGMENTATIONS( $seg1, seg2$ )
15:  for  $s \in joint$  do
16:    if  $s.size < min\_size$  then
17:      closest = CLOSEST_NEIGHBOR( $s, joint$ )
18:      joint = MERGE_SEGMENTS( $joint, s, closest$ )
19:    end if
20:  end for
21:  return joint
22: end procedure

```

---

| Method      | UUCs |     |     |     |     | Avg. AUROC  | Post proc. |
|-------------|------|-----|-----|-----|-----|-------------|------------|
|             | 0    | 1   | 2   | 3   | 4   |             |            |
| CoReSeg     | .91  | .81 | .72 | .79 | .65 | .777 ± .097 | ✓          |
| CoReSeg     | .89  | .77 | .69 | .74 | .63 | .742 ± .097 | -          |
| CoReSeg+Att | .87  | .94 | .73 | .75 | .79 | .815 ± .086 | ✓          |
| CoReSeg+Att | .84  | .90 | .69 | .72 | .72 | .774 ± .092 | -          |
| OpenGMM     | .88  | .74 | .63 | .59 | .72 | .713 ± .112 | ✓          |
| OpenGMM     | .84  | .74 | .62 | .56 | .68 | .690 ± .108 | -          |
| OpenPCS     | .86  | .67 | .63 | .63 | .59 | .675 ± .106 | ✓          |
| OpenPCS     | .81  | .65 | .62 | .59 | .56 | .649 ± .097 | -          |
| OpenPCS++   | .59  | .57 | .51 | .48 | .51 | .533 ± .045 | -          |
| OpenPCS++   | .59  | .57 | .50 | .48 | .52 | .532 ± .047 | ✓          |

TABLE II: The table shows the AUROC for the base open-set prediction obtained by the combination of U-net with or without attention to the method for the Vaihingen dataset. Each backbone-method pair compares the performance of the base open-set prediction with the best and the worst post-processing configuration results. The UUCs number stands for 0 - impervious surfaces; 1 - building; 2 - low vegetation; 3 - high vegetation; and 4 - car. All backbones uses the CBAM attention mechanism.

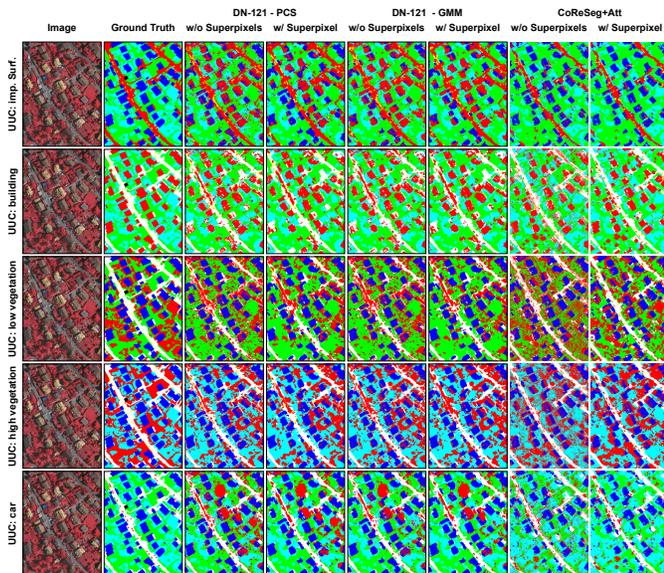


Fig. 7: Open-set segmentation predictions obtained using the best hyperparameter configuration for OpenPCS, OpenGMM, and CoReSeg+Attention for one test image of the Vaihingen dataset with all tested Unknown Unknown Classes (UUCs). Also, results with and without post-processing are presented on the right of the base prediction. The exhibited superpixel segmentation algorithm configuration used for post-processing is the best one for each method on average. The used colors are: white for *impervious surfaces*; dark blue for *building*; light blue for *low vegetation*; green for *high vegetation*; yellow for *car*; and red for the OOD pixels.

CoReSeg [12]. Besides that, this work proposed a general post-processing technique with a new superpixel merging procedure called FuSC.

To evaluate the performance of the proposed methods, we performed exploratory tests on remote sensing image datasets and extensive quantitative and qualitative experimental evaluation comparing the proposed approaches with established literature baselines.

The two proposed methods for OSS improved the baseline results and showed better semantic consistency. Output scores of four distinct OSS methods – OpenPCS [7], OpenPCS++ [17], OpenGMM, and CoReSeg – were post-processed using FuSC producing a refined open-set prediction that consistently improved the quantitative results and semantic consistency.

Table II presents the results achieved for the Vaihingen dataset where OpenGMM improved the baseline results, but CoReSeg established new state-of-the-art results for both datasets. To the author’s knowledge, CoReSeg is the first fully convolutional end-to-end method used to perform open-set segmentation in remote sensing images in literature.

Post-processing using FuSC produced more consistent and stabler results, varying less among the different tested configurations. Suggesting that FuSC is less sensitive to hyperparameter selection, the final results for FuSC performed better on

average than individual superpixel algorithms. Within the final results, post-processing with FuSC configuration produced the best overall results. Figure 7 shows the qualitative improvement achieved by each method for all the tested unknown classes.

We also published the first survey on Open-set Semantic Segmentation proposing a novel taxonomy [9] aiming to organize the literature and provide an understanding of the theoretical trends that guided the existing approaches that may influence future methods. We then extended the systematic mapping [10] with a more comprehensive set of papers improving the proposed taxonomy and adding complementary information.

## VII. PUBLICATIONS

The CoReSeg method proposed in this thesis was published in the 29th IEEE International Conference on Image Processing - ICIP2022 [12]. OpenGMM and FuSC was published in IEEE Access journal [11]. The taxonomy proposed in this thesis was published at the 35th Conference on Graphics, Patterns, and Images - SIBGRAPI2022 [9], and was extended and published in a special issue of the journal Computer and Graphics [10].

## ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

I would also like to thank Professor Jefersson A. dos Santos and Matheus Pereira for their support, help and commitment in the development of this thesis.

## REFERENCES

- [1] C. Geng, S.-j. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3614–3631, 2020.
- [2] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1563–1572.
- [3] X. Sun, Z. Yang, C. Zhang, K.-V. Ling, and G. Peng, "Conditional gaussian distribution learning for open set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 480–13 489.
- [4] P. Oza and V. M. Patel, "C2ae: Class conditioned auto-encoder for open-set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2307–2316.
- [5] Z. Cui, W. Longshi, and R. Wang, "Open set semantic segmentation with statistical test and adaptive threshold," in *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2020, pp. 1–6.
- [6] Y. Guo, G. Camporese, W. Yang, A. Sperduti, and L. Ballan, "Conditional variational capsule network for open set recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 103–111.
- [7] H. Oliveira, C. Silva, G. L. Machado, K. Nogueira, and J. A. dos Santos, "Fully convolutional open set segmentation," *Machine Learning*, pp. 1–52, 2021.
- [8] A. Brilhador, M. Gutoski, A. E. Lazzaretti, and H. S. . Lopes, "A comparative study for open set semantic segmentation methods," in *Anais do 15 Congresso Brasileiro de Inteligência Computacional*, C. J. A. B. Filho, H. V. Siqueira, D. D. Ferreira, D. W. Bertol, and R. C. L. ao de Oliveira, Eds. SC: SBIC, 2021, pp. 1–8.
- [9] I. Nunes, H. Oliveira, M. B. Pereira, J. A. d. Santos, and M. Poggi, "Deep open-set segmentation in visual learning," in *Proceedings... Conference on Graphics, Patterns and Images*, 35. (SIBGRAP), 2022. [Online]. Available: <http://urlib.net/ibi/8JMKD3MGPEW34M/47MJCTH>
- [10] I. Nunes, C. Laranjeira, H. Oliveira, and J. A. dos Santos, "A systematic review on open-set segmentation," *Computers & Graphics*, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0097849323001218>
- [11] I. Nunes, M. B. Pereira, H. Oliveira, and J. A. Dos Santos, "Fusc: Fusing superpixels for improved semantic consistency," *IEEE Access*, 2024.
- [12] I. Nunes, M. B. Pereira, H. Oliveira, J. A. dos Santos, and M. Poggi, "Conditional reconstruction for open-set semantic segmentation," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 946–950.
- [13] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi, "Generative openmax for multi-class open set classification," *arXiv preprint arXiv:1707.07418*, 2017.
- [14] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep anomaly detection with outlier exposure," *arXiv preprint arXiv:1812.04606*, 2018.
- [15] C. C. da Silva, K. Nogueira, H. N. Oliveira, and J. A. dos Santos, "Towards open-set semantic segmentation of aerial images," in *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*. IEEE, 2020, pp. 16–21.
- [16] M. Vendramini, H. Oliveira, A. Machado, and J. A. dos Santos, "Opening Deep Neural Networks With Generative Models," in *ICIP*. IEEE, 2021, pp. 1314–1318.
- [17] J. A. C. Martinez, H. Oliveira, J. A. dos Santos, and R. Q. Feitosa, "Open set semantic segmentation for multitemporal crop recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [18] S. Yan, J. Zhou, J. Xie, S. Zhang, and X. He, "An em framework for online incremental learning of semantic segmentation," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3052–3060.
- [19] J. Cen, P. Yun, J. Cai, M. Y. Wang, and M. Liu, "Deep metric learning for open world semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 333–15 342.
- [20] M. Grcić, P. Bevandić, and S. Šegvić, "Dense anomaly detection by robust learning on synthetic negative data," *arXiv preprint arXiv:2112.12833*, 2021.
- [21] R. Chan, M. Rottmann, and H. Gottschalk, "Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation," in *Proceedings of the IEEE/cvf international conference on computer vision*, 2021, pp. 5128–5137.
- [22] J. Gawlikowski, S. Saha, A. Kruspe, and X. X. Zhu, "An advanced dirichlet prior network for out-of-distribution detection in remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.
- [23] J. Hong, W. Li, J. Han, J. Zheng, P. Fang, M. Harandi, and L. Petersson, "Goss: Towards generalized open-set semantic segmentation," *arXiv preprint arXiv:2203.12116*, 2022.
- [24] H. Dong, Z. Chen, M. Yuan, Y. Xie, J. Zhao, F. Yu, B. Dong, and L. Zhang, "Region-aware metric learning for open world semantic segmentation via meta-channel aggregation," *arXiv preprint arXiv:2205.08083*, 2022.
- [25] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4016–4025.
- [26] M. Grcić, P. Bevandić, and S. Šegvić, "Dense open-set recognition with synthetic outliers generated by real nvp," *arXiv preprint arXiv:2011.11094*, 2020.
- [27] S. Kong and D. Ramanan, "Opengan: Open-set recognition via open data generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 813–822.
- [28] P. Bevandić, I. Krešo, M. Oršić, and S. Šegvić, "Dense open-set recognition based on training with noisy negative images," *Image and Vision Computing*, p. 104490, 2022.
- [29] M. Grcić, P. Bevandić, and S. Šegvić, "Densehybrid: Hybrid anomaly detection for dense open-set recognition," *arXiv preprint arXiv:2207.02606*, 2022.
- [30] R. Bharadwaj, G. Jaswal, A. Nigam, and K. Tiwari, "Mobile based human identification using forehead creases: Application and assessment under covid-19 masked face scenarios," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 3693–3701.
- [31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [32] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *British Machine Vision Conference 2016*. British Machine Vision Association, 2016.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [34] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.
- [35] J. Ji, X. Lu, M. Luo, M. Yin, Q. Miao, and X. Liu, "Parallel fully convolutional network for semantic segmentation," *IEEE Access*, vol. 9, pp. 673–682, 2020.
- [36] L. Melas-Kyriazi and A. K. Manrai, "Pixmatch: Unsupervised domain adaptation via pixelwise consistency training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 435–12 445.
- [37] J. Kang, Z. Wang, R. Zhu, X. Sun, R. Fernandez-Beltran, and A. Plaza, "Picoco: Pixelwise contrast and consistency learning for semisupervised building footprint segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10 548–10 559, 2021.
- [38] R. Ratajczak, C. Crispim, B. Fervers, E. Faure, and L. Tougne, "Semantic segmentation post-processing with colored pairwise potentials and deep edges," in *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2020, pp. 1–6.
- [39] H. Zhang, K. Jiang, Y. Zhang, Q. Li, C. Xia, and X. Chen, "Discriminative feature learning for video semantic segmentation," in *2014 International Conference on Virtual Reality and Visualization*. IEEE, 2014, pp. 321–326.