Deep Open-Set Segmentation in Visual Learning

Ian M. Nunes and Marcus Poggi
Department of Informatics
Pontifícia Universidade Católica
Rio de Janeiro, Brazil
inunes,poggi@inf.puc-rio.br

Hugo Oliveira
Institute of Mathematics and Statistics
Universidade de São Paulo
São Paulo, Brazil
oliveirahugo@usp.br

Matheus B. Pereira
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
matheuspereira@dcc.ufmg.br

Jefersson A. dos Santos
Computing Science and Mathematics
University of Stirling
Stirling, Scotland, UK
jefersson@ieee.org

***Abstract*—Collecting samples that exhaust all possible classes for real-world tasks is usually hard or even impossible due to many different factors. In a realistic/feasible scenario, methods should be aware that the training data is incomplete and not all knowledge is available. In this scenario, in test time, developed methods should be able to identify the unknown samples while correctly executing the proposed task to the known classes. Open-Set Recognition and Semantic Segmentation models emerge to handle this sort of scenario for visual recognition and dense labeling tasks, respectively. In this work, we propose a novel taxonomy aiming to organize the literature and provide an understanding of the theoretical trends that guided the existing approaches which may influence future methods. Moreover, we also provide the first systematic review of open-set semantic segmentation methods.**

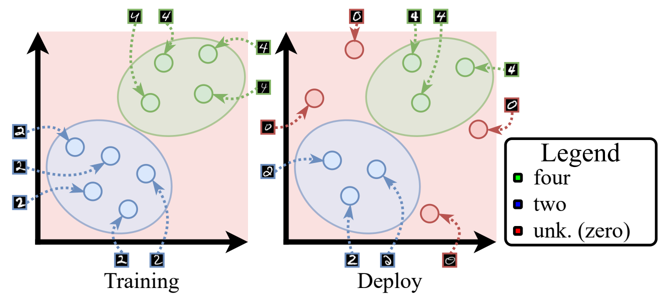
***Index Terms*—open-set, semantic segmentation, open-set recognition, open-set segmentation, deep learning, neural network**

I. Introduction

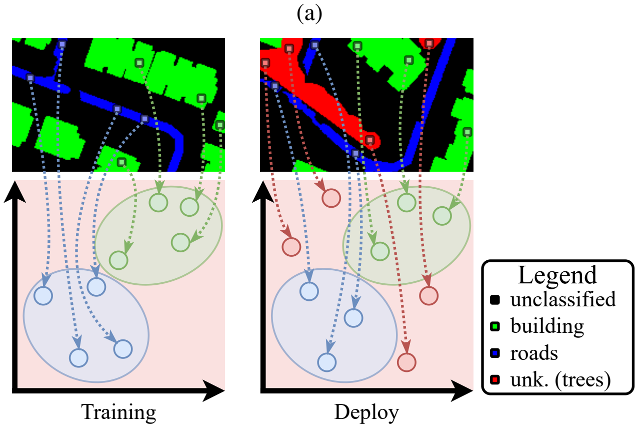
During the last decade, the automation of visual recognition tasks have reached human level standards in many domains [19], [31], [35], [36]. Convolutional Neural Networks (CNNs) [23] shifted the main limitation of visual recognition from the lack of representation capability of shallow features to the amount of labeled training data in a dataset/domain. Closed-set tasks in CNNs and related network architectures, such as classification, detection, or segmentation assume that the training and testing label spaces are the same [30]. This scenario is not compatible with the majority of real-world problems, since the tasks are limited due to the difficulty of collecting labeled samples that exhaust all possible classes.

As stated by Scheirer *et al.* [29], an open-set scenario happens when unknown samples can appear in the prediction phase, meaning that at training time not all possible classes are known. Applying this definition to a classification problem, a new task called Open-Set Recognition (OSR) emerges.

The authors would like to thank CAPES, CNPq (312102/2017-8, 424700/2018-2, and 311395/2018-0), FAPEMIG, FAPESP (grants #2020/06744-5 and #2015/22308-2), and Serrapilheira Institute (grant #R-2011-37776).



(a)



(b)

Fig. 1: Difference between training and deployment phases in OSR (a) and OSS (b) scenarios. Red circle samples (for OSR) or red pixels (for OSS) represent samples unknown in training.

The same definition can also be applied to each pixel in an image, extending the traditional semantic segmentation problem to Open-Set Segmentation (OSS). OSS refers to the set of algorithms that identifies pixels of unknown or out-of-distribution (OOD) classes at inference time, while still correctly classifying pixels of the known classes learned in training [26]. Figure 1 illustrates the OSR and OSS tasks.

The open-set tasks have caught the research community's interest with multiple recently proposed methods for open-set recognition (OSR) problems [2], [8], [16], [27], [30]. However, only a few works tackle the problem for different visual tasks, such as segmentation or object detection [17]. OSS is an inherently harder problem due to its dense labeling nature compared to Open-Set Recognition or Classification. Thus, in real-world scenarios, it is harder to perform open-set semantic segmentation precisely [5]. This may explain why there is still a gap in the literature with only a handful of works tackling the issue [8].

Understanding and organizing the literature on any area is a challenging task that can help researchers to place their work among the many existing methods. It is also useful to provide an overview of the research area for newcomers and for the following works. Considering the lack of a more structured organization for the OSR and OSS literature, in this work, we propose a taxonomy for deep learning open-set recognition and segmentation, helping to organize the literature by classifying existing methods according to their characteristics. Furthermore, the taxonomy allows to identify the main emerging trends that may serve as the base for future approaches. To the best of our knowledge, this is the first work that reviews the OSS methods available in the literature. We focused our study on deep learning-based methods, therefore shallow approaches will not be considered.

The remaining of this paper is organized as follows. In Section II we introduce the literature reviewing process and the proposed taxonomy. Section III discusses the most representative OSR papers according to the proposed taxonomy, while Section IV reviews, analyzes, and categorizes the OSS works. At last, Section V concludes the paper discussing some of the most promising trends in OSR and OSS, while also presenting possible future research directions in the field.

II. Reviewing Process and Taxonomy

Aiming to systematize the choice and analysis of works on OSS, we followed the methodology from the literature of systematic mapping [21] as how to conduct an organized review process. The following subsection will explain the procedure.

A. Systematic Review Methodology

Since this research focuses on deep learning methods for OSS, the defined search terms were: 1) "*segmentation*"; 2) "(*open-set* OR *open set* OR *openset* OR *open-world* OR *open world*)"; "(*deep learning* OR *neural network*)".

We selected three digital libraries/search engines to gather comprehensive results: Google Scholar[1], Scopus[2], and Web of Science[3]. We defined only one search string for Scopus and Web of Science since they allow for structured search strings, as presented in Table I. The table also shows the two less restrictive search strings defined for Google Scholar.

[1]https://scholar.google.com.br/

[2]https://www.scopus.com/search/form.uri?display=basic#basic

[3]https://www.webofscience.com/wos/woscc/basic-search

Database	Search	Results
Web of Science	"segmentation" AND ("open set" OR "open-set" OR "openset" OR "open world" OR "open-world") AND ("neural network" OR "deep learning")	16
Scopus	"segmentation" AND ("open set" OR "open-set" OR "openset" OR "open world" OR "open-world") AND ("neural network" OR "deep learning")	36
Google Scholar	open-set segmentation	33
Google Scholar	open-world segmentation	27

TABLE I: Table presents the used queries for each search engine and the number of results returned.

Besides the search results for OSS, multiple relevant OSR works were manually included in the mapping, as the majority of OSS methods are adapted from the OSR literature. Section III presents an overview of the OSR literature according to the taxonomy presented in Section II-B. We highlight that the review on the OSR literature was thought to be representative rather than extensive, considering that the amount of OSR papers is considerably higher than the ones for OSS. We aimed to reach the representativity by including the seminal articles for each category from the taxonomy, along with the ones also found during the OSS paper search. Thus, the selection of works for the OSR task is not necessarily fully complete, differently from the OSS review. Yet, the review on OSR papers is necessary, as we describe OSR methods in order to introduce OSS later.

Since the total number of articles is relatively small, we considered the union of all results and manually excluded the following types of works: surveys, thesis and dissertations, submitted and rejected works, works describing frameworks used on competitions, works in which the main task is other than segmentation or recognition, and works not focusing on images. We refined the search of OSS methods to 71 works after duplicate removal, further reducing this number to 24 papers after applying the exclusion criteria, with 15 being focused on OSS and 9 dealing with OSR tasks. Figure 3 shows the distribution of the works by year of publication and the growing interest in OSR and OSS from the research community.

As only 24 works resulted from the combined search and exclusion criteria, all articles were read and further classified into the taxonomy. To better understand research trends in OSS using deep neural networks, we extracted the following

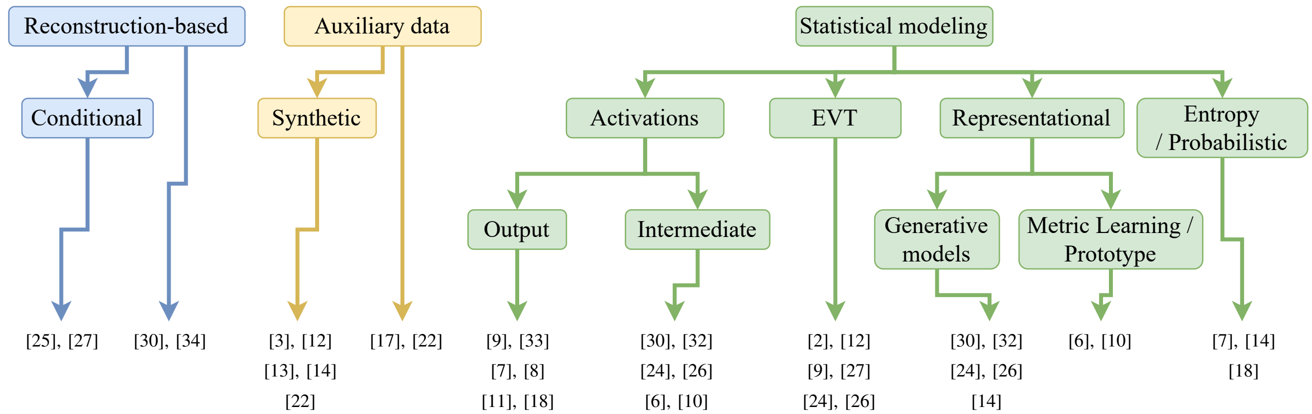


Fig. 2: Classification of the selected works under the proposed categories of the taxonomy presented in Section II. Each category can be further divided into more refined groups according to the methods' characteristics. Each method may fall under more than one group, as they are not mutually exclusive.

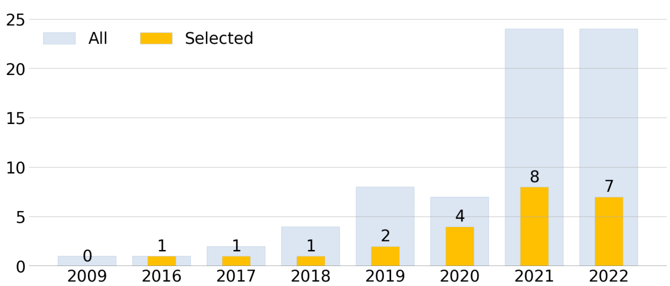


Fig. 3: The evolution in number of publications of the combined search results shown in grey and in yellow the final number of selected articles.

complementary data from all final selected articles:

1) Does the article address the open-set scenario?
2) Which is the main task addressed by the article?
3) What kind of data is used?
4) Does the method use reconstruction?
5) Does the method use auxiliary data?
6) Does the method use a generative approach?
7) Does the method use any statistical modeling?
8) Does the method use the intermediate feature space?
9) Can the method be easily adapted from the closed-set task or, in short, is the method plug & play?
10) Does the method use EVT to model OOD classes?

We compiled the Table II from the proposed questions above. Each column of the table answers one proposed question to map the architectural choices made by the authors. The proposed questions were used to map the emerging trends in literature and to help organizing the methods as to define an adequate taxonomy.

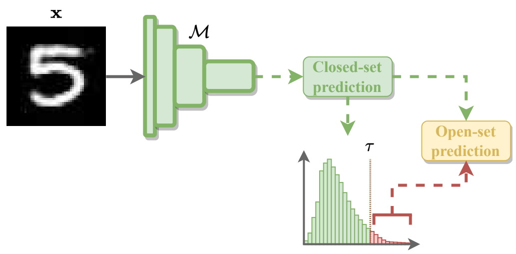
We further detail the most relevant individual works presented in Table II in Sections III and IV.

B. Taxonomy

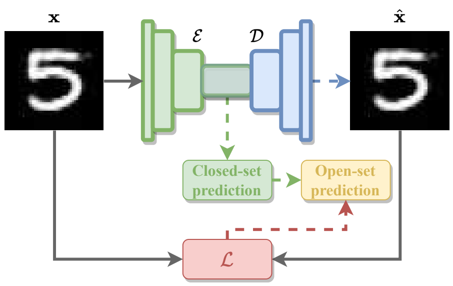
Aiming to better understand the trends, the selection of articles guided us to the following taxonomy, mapping three identified paradigms that organize the families of methods for OSR and OSS commonly found in the literature:

1) *Statistical modeling*: statistics of the intermediary and output activations from the networks are used to define in- and out-of-distribution samples [2], [6]–[12], [14], [17], [18], [24], [26], [30], [32], [33], as illustrated in Figure 4a. This is a broader category than the next two, and as such, 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, and the output of an anomaly (entropy or probability) score;
2) *Reconstruction-based*: image reconstruction loss is used to model or classify OOD samples [25], [27], [30], [34], as shown in Figure 4b. 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 [3], [13]–[15], [17], [22]. This pipeline is shown in Figure 4c. 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.

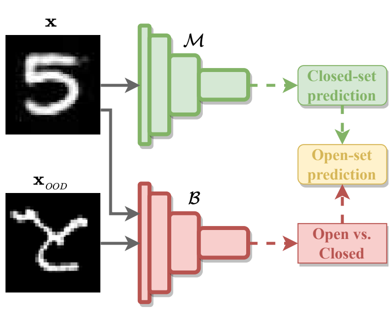
A graphical visualization of all selected papers under the respective category is shown in Figure 2.



(a)



(b)



(c)

Fig. 4: Schematics for the presented taxonomy: statistical modeling (a); reconstruction-based (b); auxiliary data (c).

Ref	T	D	R	A	G	S	F	P	E	SE
[2]	R	I	✗	✗	✗	✓	✗	✓	✓	M
[12]	R	I	✗	✗	✓	✓	✗	✗	✓	M
[17]	R	I	✗	✓	✗	✓	✗	✗	✗	M
[34]	R	I	✓	✗	✓	✗	✗	✗	✗	M
[27]	R	I	✓	✗	✓	✗	✗	✗	✓	M
[30]	R	I	✓	✗	✓	✓	✓	✗	✗	M
[32]	R	I	✗	✗	✓	✓	✓	✓	✗	M
[4]	R	I	✗	✗	✗	✗	✗	✗	✗	W
[11]	R	RS	✗	✗	✗	✓	✗	✗	✗	W
[9]	S	RS	✗	✗	✗	✓	✗	✓	✓	S
[13]	S	I	✗	✓	✓	✗	✗	✗	✗	G
[8]	S	I	✗	✗	✗	✓	✗	✗	✗	S,W,G
[26]	S	RS	✗	✗	✓	✓	✓	✓	✓	S,G
[24]	S	RS	✗	✗	✓	✓	✓	✓	✓	S,G
[33]	S	I	✗	✗	✗	✓	✗	✗	✗	S
[6]	S	I	✗	✗	✗	✓	✓	✗	✗	S,G
[14]	S	I	✗	✓	✓	✓	✗	✗	✗	G
[7]	S	I	✗	✗	✗	✓	✗	✗	✗	W
[22]	S	I	✗	✓	✓	✗	✓	✗	✗	G
[25]	S	RS	✓	✗	✓	✗	✗	✗	✗	G
[3]	S	I	✗	✓	✗	✗	✓	✗	✗	S,G
[15]	S	I	✗	✗	✗	✓	✗	✗	✗	G
[18]	S	I	✗	✗	✗	✓	✗	✗	✗	G
[10]	S	I	✗	✗	✗	✓	✗	✗	✗	G

TABLE II: 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).

III. Open-set Recognition

This section will present the seminal manually chosen articles for each category of the taxonomy. They represent well the examples of methods that fall upon the proposed categories and can be considered the base of more recent approaches.

Deep statistical models for OSR can operate either solely on the output activations of a neural network $\mathcal{M}$ [2], [4], [11] or also consider the intermediary feature representations of a closed-set classification network [30], [32], as shown in Figure 4a. An important subset of this OSR paradigm specializes in using Extreme Value Theory (EVT) for detecting OOD samples. An example of this case is the traditional OpenMax algorithm [2], which adds an "unknown" output class and estimates the probability of the input images to each of the $C+1$ classes, where C is the number of known categories. EVT, despite being a robust theoretical framework to work with long-tailed distributions and anomaly detection, is usually limited to work directly on logits, not being able to adapt to include intermediate feature representations. Vendramini *et al.* [32] showed how simple generative models (i.e. principal component analysis or gaussian mixtures) surpassed the performance of OpenMax considerably in multiple traditional OSR scenarios by introducing information from the middle layers of a CNN $\mathcal{M}$.

Following a rather distinct paradigm, reconstruction-based strategies (Figure 4b) [27], [30], [34] leverage reconstruction error from autoencoding networks (e.g. autoencoders and their variants) in order to delineate the boundary between known and unknown samples. These strategies rely on the reconstruction error from known classes being smaller than reconstruction errors from unknown classes, as only known samples are seen during training. Multiple works [27], [34] repurpose the closed-set classification encoder $\mathcal{E}$ and attach a trainable decoder $\mathcal{D}$ to try to reconstruct the input image for the known classes. Class Conditioned Auto-Encoders (C2AE) [27] exemplify the reconstruction paradigm for OSR quite well by merging a closed-set classification encoder $\mathcal{E}$ pretrained on the known classes with an upsampling decoder $\mathcal{D}$ for reconstruction. $\mathcal{E}$ works both to classify among known classes and to compress the representation of the input samples into an embedding that can be reverted to an approximation of the input space by $\mathcal{D}$. In this strategy, wrongly labeled samples are purposely fed to the network to enforce that it is able to only reconstruct samples correctly conditioned to the input label.

Conditional Gaussian Distribution Learning (CGDL) [30] is a variation of the traditional reconstrucion-based pipeline that couples the reconstruction loss with a Kullback Leibler (KL) constraint on network activations – effectively working as a cascaded Variational AutoEncoder (VAE) [20]. The KL divergence is used during the training phase in order to enforce simpler gaussian bottleneck embeddings before the reconstruction. CGDL is framed into statistical and reconstruction-based due to the use of both reconstruction loss and the KL-divergence.

At last, the third OSR strategy uses known unknown samples as auxiliary data to ensure that the model learns to differentiate between known $\mathbf{x}$ and unknown samples $\mathbf{x}_{OOD}$. This OSR paradigm leverages a known set of unknown samples – henceforth known as the support set – to transform the usually unsupervised generative modeling of OSR into a supervised discriminative process. For instance, G-OpenMax [12] employs a Generative Adversarial Network (GAN) $\mathcal{B}$ trained on OOD data to learn how to discriminate the known classes (classified through the closed-set branch $\mathcal{M}$) from synthetic samples. In the same direction, Outlier Exposure [17] model uses different datasets as OOD samples in the open *vs.* closed branch $\mathcal{B}$ to learn how to discriminate the known distribution from others. An example of this class of methods can be seen in Figure 4c.

IV. Open-Set Semantic Segmentation

Statistical Modeling is the most common background structure used by OSS methods and the usage varies from method to method. Principal Component Analysis (PCA), Gaussian Mixture of Models (GMM), entropy, or probability are used to generate anomaly scores from intermediate features or final layers of networks to distinguish and characterize OOD via threshold, as seen in [7], [8], [14], [15], [18], [24], [26]. OpenPCS [26] and OpenPCS++ [24] use PCA to reduce the dimensionality, generating a representation of the stacked intermediate features and the final layers. A threshold is employed in the resulting log-likelihood to identify OOD pixels. A great advantage of both OpenPCS and OpenPCS++ is the "plug & play" characteristic, which allows a fast adaptation of the method and the use in either new datasets or different closed-set backbones. Another related work, [8] applied a statistical test to the produced entropy-uncertainty map to determine if any area is unknown. Other representational strategies employed are Metric Learning and Prototyping, as in [6], [10], where the calculated distance between representations and each sample is used to define which are OOD. The use of the Extreme Value Theory (EVT) to model the final score or loss distribution and to separate OOD objects from the known objects is employed by [9], [24], [26]. A different approach uses probability sampling to balance sample selection and improve the learning for the method [33].

Reconstruction-based strategies are employed in only one method in OSS. In general, reconstruction-based methods use the reconstruction loss to identify OOD pixels. The only reconstruction method [25] found in our search uses conditional reconstruction to identify OOD pixels. In training, the method learns to reconstruct pixels conditioned to their class, and in testing, all pixels are conditioned to all known classes. The ones from unknown samples tend to present higher reconstruction loss values, thus being set as OOD by a threshold.

Auxiliary data had three different usages mapped in this study. The first is with the use of synthetic images [13], [14], [22]. The method proposed by [13] employs synthetic negative patches added to images that simultaneously achieve uniform discriminative prediction and high inlier likelihood. Also, Jensen-Shannon divergence was employed in both training and inference instead of the Kullback–Leibler (KL) divergence. The Jensen-Shannon divergence mildly penalizes high confidence predictions in comparison to KL-divergence. The OpenGAN method [22] learns a robust open-vs-closed discriminator that serves as open-set likelihood. The discriminator is trained with fake (synthetic) data from a generator and real open training examples as an outlier exposure (OE) strategy. As the GAN objective is not a realistic reconstruction, both generator and discriminator use the features of the closed set network. This enables readily modifying closed-set systems for open-set recognition.

The OR strategy employed by OpenGAN is the second mapped usage of auxiliary data. More specifically, OpenGAN uses synthetic data and OE together to enhance the discriminative ability of the model. Finally, the third mapped usage is a strategy that randomly replaces a small crop of the input image with some OOD mini-patch [3], [13], [14]. In [3], the mini-patch is a random crop of a real image of the same size but with a different distribution. In [13], [14], the mini-patch is synthetic. For this kind of approach to work, the ground truths must be equally modified including the unknown class to the added mini-patch area and the model is trained to differentiate in- and out-of-distribution, along with correctly identifying OOD pixels.

V. Conclusion

In general, OSS methods are based on an OSR counterpart. Hence, our proposed taxonomy works for both tasks since the fields share similar strategies. Reconstruction-based methods might be an ongoing trend for OSS since some of the more robust methods in OSR rely on reconstruction [27], [30], [34]. Our systematic review only found one method for OSS that uses reconstruction [25], which means that this type of strategy is still in its earlier steps in spite of the strong results compared to the other baseline OSS methods.

Methods to improve the semantic consistency of the segmentation seem relevant for OSS, particularly as boundaries across objects from distinct classes tend to present larger segmentation errors than their centers. Thus, post-processing schemes capable of mitigating the lack of confidence in border regions between different objects may improve OSS results, since many open segmentation strategies employ some confidence or anomaly score to identify OOD pixels. In this direction, techniques like visual attention modules, conditional random fields, and superpixel post-processing are promising alternatives to be explored in future works.

Finally, zero-shot and few-shot tasks overlap open-set tasks, since the knowledge in these scenarios is inherently incomplete during training and the method may need to handle samples of unknown classes during deploy, possibly even using some online learning strategy. Developments in the literatures of OSR/OSS and zero-/few-shot learning [6], [28], [37] seems to walk towards each other, possibly resulting in future deep Open World [1] approaches.

References

[1] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *CVPR*, pages 1893–1902, 2015.

[2] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *CVPR*, pages 1563–1572, 2016.

[3] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Dense open-set recognition based on training with noisy negative images. *Image and Vision Computing*, page 104490, 2022.

[4] Rohit Bharadwaj, Gaurav Jaswal, Aditya Nigam, and Kamlesh Tiwari. Mobile based human identification using forehead creases: Application and assessment under covid-19 masked face scenarios. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3693–3701, 2022.

[5] Anderson Brilhador, Matheus Gutoski, André Eugênio Lazzaretti, and Heitor Silvério Lopes. A comparative study for open set semantic segmentation methods. In *Congresso Brasileiro de Inteligência Computacional*, pages 1–8, Joinville, SC, 2021. SBIC.

[6] Jun Cen, Peng Yun, Junhao Cai, Michael Yu Wang, and Ming Liu. Deep metric learning for open world semantic segmentation. In *ICCV*, pages 15333–15342, 2021.

[7] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *ICCV*, pages 5128–5137, 2021.

[8] Zhiying Cui, Wu Longshi, and Ruixuan Wang. Open set semantic segmentation with statistical test and adaptive threshold. In *International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2020.

[9] Caio CV da Silva, Keiller Nogueira, Hugo N Oliveira, and Jefersson A dos Santos. Towards open-set semantic segmentation of aerial images. In *Latin American GRSS & ISPRS Remote Sensing Conference*, pages 16–21. IEEE, 2020.

[10] Hexin Dong, Zifan Chen, Mingze Yuan, Yutong Xie, Jie Zhao, Fei Yu, Bin Dong, and Li Zhang. Region-aware metric learning for open world semantic segmentation via meta-channel aggregation. *arXiv preprint arXiv:2205.08083*, 2022.

[11] Jakob Gawlikowski, Sudipan Saha, Anna Kruspe, and Xiao Xiang Zhu. An advanced dirichlet prior network for out-of-distribution detection in remote sensing. *IEEE TGRS*, 60:1–19, 2022.

[12] ZongYuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017.

[13] Matej Grcić, Petra Bevandić, and Siniša Šegvić. Dense open-set recognition with synthetic outliers generated by real nvp. *arXiv preprint arXiv:2011.11094*, 2020.

[14] Matej Grcić, Petra Bevandić, and Siniša Šegvić. Dense anomaly detection by robust learning on synthetic negative data. *arXiv preprint arXiv:2112.12833*, 2021.

[15] Matej Grcić, Petra Bevandić, and Siniša Šegvić. Densehybrid: Hybrid anomaly detection for dense open-set recognition. *arXiv preprint arXiv:2207.02606*, 2022.

[16] Yunrui Guo, Guglielmo Camporese, Wenjing Yang, Alessandro Sperduti, and Lamberto Ballan. Conditional variational capsule network for open set recognition. In *ICCV*, pages 103–111, 2021.

[17] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018.

[18] Jie Hong, Weihao Li, Junlin Han, Jiyang Zheng, Pengfei Fang, Mehrtash Harandi, and Lars Petersson. Goss: Towards generalized open-set semantic segmentation. *arXiv preprint arXiv:2203.12116*, 2022.

[19] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, pages 4700–4708, 2017.

[20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[21] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. Systematic literature reviews in software engineering – a systematic literature review. *Information and Software Technology*, 51(1):7–15, 2009.

[22] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *ICCV*, pages 813–822, 2021.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *NIPS*, 25, 2012.

[24] Jorge A Chamorro Martinez, Hugo Oliveira, Jefersson A dos Santos, and Raul Queiroz Feitosa. Open set semantic segmentation for multitemporal crop recognition. *IEEE GRSL*, 19:1–5, 2021.

[25] Ian Nunes, Matheus B Pereira, Hugo Oliveira, Jefersson A dos Santos, and Marcus Poggi. Conditional reconstruction for open-set semantic segmentation. *arXiv preprint arXiv:2203.01368*, 2022.

[26] Hugo Oliveira, Caio Silva, Gabriel LS Machado, Keiller Nogueira, and Jefersson A dos Santos. Fully convolutional open set segmentation. *Machine Learning*, pages 1–52, 2021.

[27] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *CVPR*, pages 2307–2316, 2019.

[28] Kuniaki Saito, Ping Hu, Trevor Darrell, and Kate Saenko. Learning to detect every thing in an open world. *arXiv preprint arXiv:2112.01698*, 2021.

[29] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE TPAMI*, 35(7):1757–1772, 2012.

[30] Xin Sun, Zhenning Yang, Chi Zhang, Keck-Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *CVPR*, pages 13480–13489, 2020.

[31] Andrew Tao, Karan Sapra, and Bryan Catanzaro. Hierarchical multi-scale attention for semantic segmentation. *arXiv preprint arXiv:2005.10821*, 2020.

[32] Marcos Vendramini, Hugo Oliveira, Alexei Machado, and Jefersson A dos Santos. Opening Deep Neural Networks With Generative Models. In *ICIP*, pages 1314–1318. IEEE, 2021.

[33] Shipeng Yan, Jiale Zhou, Jiangwei Xie, Songyang Zhang, and Xuming He. An em framework for online incremental learning of semantic segmentation. In *ACM International Conference on Multimedia*, pages 3052–3060, 2021.

[34] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. Classification-reconstruction learning for open-set recognition. In *CVPR*, pages 4016–4025, 2019.

[35] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *British Machine Vision Conference*. British Machine Vision Association, 2016.

[36] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

[37] Chenhong Zhou, Feng Liu, Chen Gong, Tongliang Liu, Bo Han, and William Cheung. Krada: Known-region-aware domain alignment for open world semantic segmentation. *arXiv preprint arXiv:2106.06237*, 2021.