Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 23, Iss. 1, pp. 33–48, 2025
https://doi.org/10.21528/lnlm-vol23-no1-art3

A SURVEY ON OPEN WORLD LEARNING FOR IMAGE SEGMENTATION: DEFINITIONS, CHALLENGES, AND DIRECTIONS

Anderson Brilhador
André Eugênio Lazzaretti
Heitor Silvério Lopes

Federal University of Technology - Paraná
Av. Sete de Setembro, 3165, Curitiba, 80230-901, Paraná, Brazil.
{brilhador,lazzaretti,hslopes}@utfpr.edu.br

Resumo – Este artigo apresenta uma pesquisa sobre aprendizado de mundo aberto no contexto de segmentação de imagens, além de analisar um subconjunto de tarefas associadas que exibem características desejáveis para aplicações no mundo real, como direção autônoma, inspeção industrial, diagnóstico médico, sensoriamento remoto, entre outras. O objetivo é identificar as principais abordagens, desafios e lacunas nesse campo de pesquisa. Por meio de um procedimento rigoroso, foram selecionados 39 artigos para análise publicados entre 2013 e 2023. Em seguida, foi conduzida uma revisão para examinar esses estudos, e três principais questões de pesquisa foram levantadas. Após uma análise extensa, os resultados indicam que o aprendizado de mundo aberto para segmentação de imagens só foi explorado recentemente e parece ser um campo promissor de pesquisa para os próximos anos. Com base nessa revisão, fornecemos direções potenciais e lacunas a serem exploradas em trabalhos futuros sobre este tópico.

Palavras-chave – mundo aberto, conjunto aberto, aprendizagem incremental, segmentação de imagens, revisão, pesquisa.

Abstract – This article provides a survey of open-world learning in the context of image segmentation and a subset of associated tasks that exhibit desirable characteristics for real-world applications, such as autonomous driving, industrial inspection, medical diagnosis, remote sensing, among others. The objective is to identify the main approaches, challenges, and gaps in this research field. Through a rigorous procedure, 39 articles published between 2013 and 2023 were selected for analysis. Then, a review was conducted to examine these documents, and three main research questions were posed. After an extensive analysis, results indicate that open-world learning for image segmentation has been explored only recently, and it seems to be a promising field of research for the upcoming years. Established on this reviewed literature, we provide the potential directions and open gaps for future works on this topic.

Keywords – open-world, open-set, incremental learning, image segmentation, review, survey.

1 Introduction

Image segmentation is one of the main problems in computer vision (CV). Essentially, it is defined as the ability to subdivide an image into regions or objects that compose it [1]. However, with recent advances in deep learning, the image segmentation task has been strongly boosted [2]. Currently, image segmentation can be divided into three main types, as shown in Figure 1. First, semantic segmentation (SS) [3] aims to classify each pixel in the image labeled with a corresponding class. Instance segmentation (IS) [4] additionally distinguishes between different instances of the same class label, e.g., multiple car instances shown in Figure 1(c). Lastly, panoptic segmentation (PS) [5] can be briefly defined as combining the two previous tasks.

(a) Input image. (b) Semantic segmentation. (c) Instance segmentation. (d) Panoptic segmentation.

Figure 1: Examples of the input image and ground truth of the corresponding image segmentation tasks. Images from Mapillary Vista [6].

The information extracted by these tasks is essential and serves as a basis for applications in different fields such as autonomous driving [7], robot navigation [8], medical diagnostics [9], remote sensing [10], industrial inspection [11], industry

4.0 [12], to name a few. These applications are usually built within a closed and static set perception, where all categories are known a priori [13, 14].

A closed-set perception (CSP) system cannot update its knowledge base according to what it has seen. It is limited to particular scenarios and needs to be retrained when its knowledge becomes outdated. Furthermore, mislabeling in CSP can lead to potentially harmful situations in safety-critical applications, such as autonomous driving [15]. For instance, an autonomous car may cause a severe accident due to incorrect recognition of another vehicle, a person, or other objects on the road.

A possible way to circumvent the eventual drawbacks of using CSP systems in real-world environments is to build recognition systems based on dynamic and open-set perception (OSP). Such systems assume that applications must deal with new classes of objects that arrive regularly in the real-world. Then, they recognize these new classes not previously seen in the training step of the algorithm and, subsequently, they learn how to identify these new classes efficiently. Ideally, a model based on OSP is capable of evolving incrementally, without the need to train the model from scratch to preserve the knowledge already learned [16].

Open-world learning (OWL) is a problem that unifies the need to recognize and learn new classes over time while preserving the already known knowledge [13]. Due to the importance of these learning capabilities in real-world scenarios, studies related to this issue have received significant attention in recent years. Nevertheless, the vast majority of such studies partially solve the OWL problem, i.e., handle the recognition of new classes [17], or lead the incremental learning of new classes [18]. However, they do not unify the both learning issues as it is expected for OWL applications.

Due to the recent introduction of the OWL problem [13], the lack of in-depth studies is even more evident in the context of image segmentation, which usually has solutions based on methods initially proposed for image classification. For instance, [3] adapted convolutional neural networks intended for image classification [19–21] into a fully convolutional network for the semantic segmentation, and recently, [22] adapted the transformer architecture [23] for the same task. Although image segmentation models can be built from image classification models, there are differences in the tasks. Therefore, it is infeasible to directly adopt an OWL image classification approach to the segmentation task. In addition, image segmentation serves as the basis for different real-world applications, making it essential to explore OWL in this area of research, even with advances in other areas, such as image classification [24] and object detection [25].

In fact, OWL for image segmentation is far from the maturity already achieved in other branches of CV and machine learning (ML). Therefore, surveys about the state-of-the-art are still missing. To the best of our knowledge, this is the first review explicitly focused on advancing OWL for image segmentation tasks. There are already surveys related to OWL, such as [26] and [27]. However, both are focused on solutions for the image classification task. The present work also pushes the boundaries of OWL, looking for related ML tasks that could contribute for advancing this important research area. In this work, we also point out the main differences between tasks and OWL. Keeping this in mind, we tried to map the leading solutions and related challenges to guide future research development. Therefore, the main contributions of this work are:

a) A formal definition semantic segmentation, instance segmentation, panoptic segmentation, and open-world learning;

b) A clear description of the differences between OWL and related fields, such as: continual learning, open-set recognition, incremental learning, and open-set domain adaptation;

c) A description of the current research scenario, with charts of papers distribution per year, per task, and per application domain, as well as possible research directions;

d) A taxonomy of the leading approaches linked to the OWL in the context of image segmentation. A timeline with landmarks in OWL studies for image segmentation is also presented;

e) The main challenges related to the OWL in the context of image segmentation are pointed out;

The remaining of this work is organized as follows. Section 2 presents some theoretical aspects related to the SS, IS, and PS tasks to enable a full understanding of the OWL problem. Section 3 describes in detail the research methodology used to select the works analyzed in this review. Section 4 presents the answers to the research questions defined in the previous section. Section 5 discusses the open problems and possible solutions for the development of future works. Section 6 summarizes the paper and presents general conclusions.

2 Theoretical Aspects

This section provides a formal definition of image segmentation tasks studied in this review and defines the OWL problem later. Such definitions are essential for understanding the issues and approaches discussed in the following sections. Moreover, due to the recent introduction of the OWL problem, few works focus on its formalism. Therefore, we unify and complement the existing definitions so as to strengthen these concepts for future research development. In addition, we describe the differences between OWL and related fields.

2.1 Semantic segmentation

In this work, we define SS as the classification of the content of a scene at the pixel level, in which each pixel of the image is labeled with a corresponding class that represents the category of the pixel (Figure 1(b)). Formally, this task can be defined as follows:

Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 23, Iss. 1, pp. 33–48, 2025
https://doi.org/10.21528/lnlm-vol23-no1-art3 © Brazilian Society on Computational Intelligence

Definition 1. *Given an input image and a set of semantic classes $\mathcal{L} = \{0, 1, 2, ..., N-1\}$, where $N \in \mathbb{N}$ is the number of classes of interest. The SS algorithm must assign to each pixel i of the image a specific class of $\mathcal{L}$.*

Furthermore, SS does not distinguish among instances of a particular semantic class. It is only possible to know which classes are present in a scene and not count the number of objects deferring to a specific class. This restriction can be advantageous for some applications (e.g., remote sensing) that are not concerned with accounting for the number of unique class instances in the image.

2.2 Instance segmentation

IS expands the SS, introducing the capacity to distinguish between instances of a particular semantic class. The aim is to label and identify instances in an image at the pixel level. To do this, a pixel level mask will be produced for each identified instance containing the pixel class category and, also, a unique identifier that distinguishes object instances of a given class. This is shown in Figure 1(c) where more than one instance related to the car category is usually identified in an urban setting. In this case, each car found in the image receives an unique value, thus enabling to count the number of car instances in the image. IS can be formally defined as follows:

Definition 2. *Given a predefined set of semantic classes $\mathcal{L} = \{0, 1, 2, ..., N-1\}$ and the cluster set of an instance $Z = \{0, 1, 2, ..., C-1\}$, where $C \in \mathbb{N}$ is the cardinality, an IS algorithm must map each pixel i of an image to a pair (l_i, z_i), where l_i represents the semantic class of the pixel i, and z_i represents its instance id. An instance is a set of pixels that corresponds to a single element of a given class.*

Finally, one can emphasize that IS does not consider classes of uncountable objects (e.g., sky and ocean), which are generally assigned to the background class.

2.3 Panoptic segmentation

Due to the restriction of only labeling the semantic classes of the pixels of SS and the inability of the IS to segment the uncountable classes, the PS was recently proposed [5]. It aims at performing a complete segmentation as SS, while exclusively counting instances as IS. To simplify this fusion, PS considers two types of class categories present in the image: *stuff* and *things*. *Stuff* is related to uncountable regions, like the sky, the pavement, and the water, while *things* include all countable objects such as cars, people, and animals. The output of the panoptic is a dense slice containing the semantic class of each pixel and a multi-mask slice for each countable instance that receives a unique id value. The result of this segmentation is shown Figure 1(d). Such a complete segmentation may help to increase the understanding of the scene. According to [28], PS can be defined as follows:

Definition 3. *PS algorithm must map each pixel i of an image to a pair (l_i, z_i), where l_i represents the semantic class of the pixel i and z_i represents its id of instance. The set of semantic labels consists of the subsets things ($\mathcal{L}^{things}$) and stuff ($\mathcal{L}^{stuff}$), such that $\mathcal{L} = \mathcal{L}^{stuff} \cup \mathcal{L}^{things}$ and $\mathcal{L}^{stuff} \cap \mathcal{L}^{things} = \emptyset$. For each pixel in the image, a class l_i and z_i are assigned. When a pixel is labeled with $l_i \in \mathcal{L}^{stuff}$, the value of z_i is irrelevant, as for all classes of stuff, pixels belong to the same instance (e.g., the same sky). Meanwhile, all pixels with the same assignment l_i, z_i where $l_i \in \mathcal{L}^{things}$ belong to the same instance (e.g., the same car).*

Finally, the selection of stuff and things is a design choice left to the creators of the dataset and may vary across application domains [28].

2.4 Open-world learning

Traditional ML methods are usually trained using supervised learning, where a training algorithm consists of learning an objective function that maps an input to a given output based on a labeled dataset. This training dataset consists of an input object and a corresponding output value. This approach works well in CSP, controlled environments where the application domain is well known. However, supervised learning algorithms in dynamic scenarios suffer from the presence and emergence of new classes – that is, classes that are not learned by the algorithm in the training stage. In this situation, the algorithms must learn new classes and develop the ability to identify them after being deployed. This ability to recognize and learn new things has been defined in the literature as OWL [13].

OWL is not limited to supervised learning. It can be widely defined as the ability of a model to perform a specific learning task while identifying new things, and then incrementally learning the new things without completely retraining the model. OWL can be found in different learning scenarios and paradigms. For example, a natural language processing system might recognize a new word that it does not know and then learn by looking up its meaning in the dictionary. If it cannot find the word, the system can call a human agent to describe and learn it. Due to these characteristics, OWL can be categorized as a form of self-motivated learning. That is, the learner is curious to explore new territories and learn new things. The key to OWL is to recognize something that was never seen or learned before. Without this ability, it is not possible to learn or explore new things, other than being instructed by a human user or an external system, which is not ideal for a true autonomous intelligent system [16].

Due to the recent introduction of OWL, few works seek to formalize this problem [13, 16, 27, 29]. In this scenario, the literature lacks new formal definitions to strengthen knowledge about this problem and expands the works related to OWL.

However, before defining OWL, it is necessary to formally determine which class categories can be present along this learning process. The categories of basic recognition classes suggested by [30, 31], which were initially adapted from [32], are shown in Table 1.

Table 1: The four basic categories of classes in OWL. Adapted from [30] and [31].

Category	Description
Known known classes (KKCs)	Classes with distinctly labeled positive training samples (also serving as negative samples for other KKCs).
Known unknown classes (KUCs)	Labeled negative samples, not necessarily grouped into meaningful categories, such as background [33] and universum [34] * classes.
Unknown known classes (UKCs)	Classes without available samples in training, but with available side information (e.g., semantic/attribute data).
Unknown unknown classes (UUCs)	Classes without any available information during training.

* The universum [34] classes represent the samples that do not belong to either class of interest for the specific learning problem.

Inspired by the works [13, 16, 27, 29], we formalize a general definition of OWL applied to SS, IS, and PS:

Definition 4. *Let a predefined set of KKCs*[1] $\mathcal{L}_{known} = \{0, 1, 2, ..., N-1\}$ *and the cluster set of an instance* $Z = \{0, 1, 2, ..., C-1\}$*, where* $C \in \mathbb{N}$ *is the cardinality. The objective of the classifier is to map each pixel* i *of an image to a pair* (l_i, z_i)*, where* l_i *represents a* $\mathcal{L}_{known}$ *or an UUCs. If* l_i *belongs to an UUCs, then* z_i *will receive a single value. Otherwise,* z_i *receives its instance id referent to the* $\mathcal{L}_{known}$*. Notice that for SS, zi is ignored, and for PS, exclusively applicable to* $\mathcal{L}^{stuff}$ *and ignored to* $\mathcal{L}^{things}$*.*

Hence, OWL can be accomplished in three steps:

1. At a given time t, a multi-class classification model M_t is built by a learner based on all classes known prior to t, on the labeled training data with $\mathcal{L}_{known}$. M_t is able to classify each test sample $\mathcal{L}^t_{known}$ or reject them as UUC, and place them in a given rejection set R_e. The R_e can include more than one UUC.
2. At this point, the system (or a human user) can identify the hidden classes c in R_e to collect and prepare a training set of these UUCs. In general, hidden classes are interest objects for a given application that are unknown by the M_t.
3. The existing model M_t is partially fitted based on the updated training dataset (previous data + newly identified data) to learn to identify c while preserving the already learned knowledge of the previous classes. At the end of the adjustment, the model M_t is updated to a new model M_{t+c}.

2.5 Tasks related to Open-world learning

OWL is one of the central learning paradigms that enable the deployment and maintenance of ML algorithms in a real-world environment. In CV and, more specifically, in image segmentation, OWL is still under-explored, as will be shown later in this review. Nevertheless, we decided to include other related tasks with desirable characteristics in this review to discover insights that can accelerate the construction of viable applications for an open-world environment. These tasks will be briefly described below. In addition, we define the main differences between them and OWL according to our perception.

The subset of related tasks belonging to OWL:

- **Continual Learning (CL)**: It is the ability of a ML algorithm to learn from a data stream continuously. In practice, the algorithm should learn autonomously to recognize new data as it arrives. Actually, OWL is considered a subarea of CL, and the main difference between them is that in OWL the learning step of new classes can be left to a second stage, and it is only necessary to identify the UUCs as soon as the data arrives.
- **Open-Set Recognition (OSR)**: It is the ability of a ML algorithm to recognize classes of unknown samples. OSR can be considered a subarea of OWL which, besides recognizing UUCs, also aims to identify these classes.
- **Incremental Learning (IL)**: It is the ability of a ML algorithm to learn to identify new classes (UUCs). It is also considered a subarea of OWL, since incremental learning is one of the steps present in OWL.

The subset of related tasks with desirable characteristics to OWL:

- **Open-Set Domain Adaptation (OSDA)**: It is the ability of a ML algorithm to adapt to a new data domain that differs from the previously learned data distribution, while also dealing with and recognizing new classes. However, OWL does not consider domain adaptation (DA) in its tasks, although, this adaptation can be crucial for replicating already trained algorithms in different data domains.

[1] Include KUCs when results in models contain an explicit "other class" [31].

3 Research Methodology

To obtain an overview of the research in OWL applied to image segmentation tasks, a literature review was done. We used the method proposed by [35] with a few variations. The reviewing protocol is divided into three steps: (1) Definition of research questions; (2) Search strategy; and (3) Inclusion and exclusion criteria. These steps are described as follows.

3.1 Definition of research questions

The primary goal of this study is to map and categorize the state-of-the-art OWL applied to image segmentation tasks, synthesizing the advantages and disadvantages of the existing approaches. The idea is to find gaps and to provide several insights that can then help researchers to develop new solutions [36]. The research questions are described in Table 2.

Table 2: Research questions.

Research Question	Motivation
What is the distribution of papers per year, application field, and tasks related to open-world learning in the image segmentation context?	The answer to this question allows analyzing how the academic community approaches the Open-World Image Segmentation problem.
Which are the leading solutions to the problems linked to open-world learning in the image segmentation context?	The answer to this question should indicate what solutions exist, which evidence supports these different solutions, and how they relate.
Which are the main challenges reported by the selected works?	The answer to these questions helps to explore the main difficulties that arise when using OWL in Image Segmentation and depicts the advantages and limitations of existing solutions.

3.2 Search strategy

The first step of our protocol is to define sources of information and carefully check the available data. The following databases and search engines were selected for their popularity and common use in the area of Computer Science: ACM digital library, IEEE Xplore, Science Direct, Springer Link, Scopus, Web of Science, and Google Scholar.

The next step is the definition of the terms or keywords used in the searches. This review aims to summarize the literature on OWL applied to image segmentation. In addition to OWL, the tasks related to OWL were mixed with the image segmentation tasks described in Section 2, as terms used in this search. Thus, the following rule to form the search queries was established: a combination of an OWL term [2] and an image segmentation task [3], e.g., "*open-set recognition*" and "*semantic segmentation*". Finally, we define the bounding of the query terms, respecting the recommendations and limitations of each platform:

1. Publication date in the range January 2013 to December 2023;
2. Search performed on the title, abstract, and author keywords;
3. Combination of OWL and image segmentation terms.

3.3 Selection Criteria

It is fundamental for a review to refine the publications found as a result of the previous step, aiming at selecting only the most relevant publications for this study. To complete this goal, we defined a screening process that has the following steps:

Step 1. Remove all duplicate publications indexed in more than one information source.

Step 2. Review each publication's title, abstract, and keywords to apply the inclusion/exclusion criteria defined in Table 3. A study is selected when it meets all inclusion criteria and no exclusion criteria. Skimming and scanning reading techniques were used on the full text when the information in the title, abstract, and keywords is insufficient to evaluate the inclusion and exclusion criteria.

Step 3. Apply the quality assessment of the publications. This step is divided into two parts: (a) Select the studies published in classified publishers in the highest quartile (Q1) of the CiteScore 2020[4]. We chose this metric because it unifies the citation impact of conferences and journals, allowing a more concise selection among studies. (b) According to the eight quality criteria shown in Table 3, we selected the articles that satisfied 75% of the issues. For this process, each question could obtain one of three possible answers with its respective score according to the following criteria: (i) contemplates

[2] Open-world learning terms: open-set recognition, continuous learning, incremental learning, and open-set domain adaptation.
[3] Image segmentation terms: semantic segmentation, instance segmentation, and panoptic segmentation.
[4] https://www.elsevier.com/solutions/scopus/how-scopus-works/metrics/citescore

the criteria = 1, (ii) partially covers the criteria = 0.5, and (iii) criteria not satisfied = 0.0. Except for the first question presented in Table 3, which is an exclusionary question, studies that do not satisfy this criterion are ignored by this review. However, the score for this question is also considered in the final criteria evaluation.

Table 3: List of the Inclusion, Exclusion, and Quality Criteria.

Inclusion Criteria
Papers published during the period between 2013 and 2023
Papers published/in-press in a journal, conference, or magazines
Exclusion Criteria
Papers not written in English
Papers with electronically unavailable full text or not freely accessible through the standard university proxy services.
Papers that do not use one of the open-world terms over image segmentation.
Quality Criteria Questions
Did this study select any open-world terms applied directly to the image segmentation task, and were they sufficiently described in the study?
Do the authors clearly define the problem or improvement they address?
Are the contributions of the research specified clearly?
Does the study provide a sufficient explanation of its methodology?
Do the authors straightforwardly present the results?
Is the conclusion of the study based on the evidence found in the experiments?
Is the dataset publicly and freely available?
Is the source code publicly and freely available?

Among the quality criteria presented in Table 3, we emphasize the importance of making the datasets and source code used in the studies publicly and freely available. This fact allows the replication of the experiments, thus boosting further development. Access verification was done using the information available in the papers. After the steps of search and filtering, the 39 papers were selected. In the following Sections, these works will be analyzed in depth to extract the relevant information.

4 Answers to the research questions

After applying the methods mentioned in the previous Section, a total of 39 papers were selected. This Section presents the answers to the research questions (Table 2) extracted from the selected papers.

4.1 What is the distribution of papers per year, application field, and tasks related to open-world learning in the image segmentation context?

To investigate the research trends, we inspected the publication year of selected papers. Figure 2(a) shows the distribution of studies from January 2011 to December 2023. There is an increasing trend in the number of studies published in the last five years related to OWL. This fact suggests that studies in this area are receiving increasing interest from the research community. We foresee that the next big challenge for researchers is to improve the viability of the approaches for real environments.

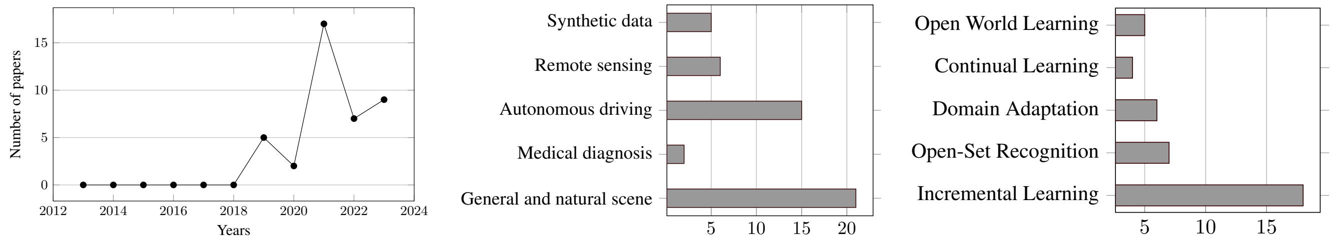
Image segmentation is a multidisciplinary research area and a fundamental step in different image processing applications. In this review, we map these application areas, aiming to indicate which are linked to OWL and are more relevant to future research.

Figure 2(b) shows that general and nature scenes represent approximately 42.86% of the application domains the selected papers explored. In general, datasets from this domain depict everyday situations that humans easily understand. The main challenge when using these datasets is the recognition of different objects in distinct scenarios, not focusing on a specific task, such as medical diagnosis or robot navigation. On the one hand, segmenting these daily images is challenging due to the wide variations in classes, objects, distances between objects, different points of view, lighting conditions, etc [37]. Notwithstanding, such challenges make this domain excellent for evaluating OWL methods, which can be applied later to more specific application domains.

Additionally, this review found four other application domains. In medical diagnosis, the main tasks associated with this domain are IL and DA, which seek to reduce the dependence of predictive models on large amounts of data for training methods due to the high costs of labeling data in medical imaging [38, 39]. In remote sensing, OSR, IL, and CL stand out in this domain. The explanation to attention is to find UUCs at the inference time in geographic mapping applications. The size of the mapped areas in a real scenario increases the chances of finding this unknown data [17, 40].

However, unlike remote sensing, autonomous driving applications require higher safety requirements since any failure in object recognition can cause disastrous accidents [15]. Furthermore, due to data acquisition limitations, synthetic data have been frequently used for OWL tasks but fulfill a complementary function for real data, especially for the DA task [41].

Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 23, Iss. 1, pp. 33–48, 2025
https://doi.org/10.21528/lnlm-vol23-no1-art3



(a) Publications per year. (b) Publications per application domain. (c) Publications per task.

Figure 2: Distribution of selected papers.

In conclusion, regarding the other extreme of Figure 2(c), few studies handled the challenge of OWL in their experiments. Although the concept of CL is not relatively new, only recently, with advances in methodologies based on deep neural networks, has it been possible to solve highly complex problems in computing, such as image segmentation in CSP. Thus, we understand that one of the following challenges to be overcome by scientific studies is to break the limitations of works focused on a CSP of classes since most applications are embedded in natural environments and suffer from the high variability of information in this scenario. The tasks not mentioned are related to this challenge and have only gotten more attention in the last few years, as shown in Figure 2(a).

4.2 Which are the leading solutions to the problems linked to open-world learning in the image segmentation context?

This review evaluated and categorized the selected studies to define the main approaches. We emphasize that these approaches were extracted from selected studies and might not represent all the approaches in the literature. Table 4 presents a summary of the leading approaches.

The fact that features are fundamental to the success of ML algorithms is not a novelty. In this sense, analyzing the leading approaches in Table 4 also indicates the importance of the characteristics to implement tasks related to OWL. Furthermore, it demonstrates that features are essential for developing strategies for exploring the unknown, contradicting a possible irrelevance with the arrival of the era of deep learning models that generally work on a "black box" structure and perform expressive results [3, 22]. This is for a better understanding of the importance of features for OWL tasks. Next, the way the approaches handle features will be discussed.

First, the metric learning strategy proposed by [14] aims to direct the feature extraction process to obtain a well-controlled latent space, maximizing inter-class spacing and minimizing intra-class spacing based on a distance metric and resulting in boosting the identification and learning of UUCs. The clustering, anomaly map, and likelihood approaches proposed for OSR also emphasize the extracted features. Clustering needs to find patterns among features to group KKCs and distinguish UUCs. Meanwhile, the anomaly map and likelihood functions are built on a set of previously extracted features and must be able to identify KKCs and reject UUCs. In both cases, unrepresentative features resulted in ineffective clusters and functions for the segmentation process of KKCs and UUCs.

Second, in IL, the strategies seek to preserve the structure of features already learned to protect the knowledge learned. They partially freeze the model layers and later apply fine-tuning or knowledge distillation. On the other hand, in dealing with OSDA, the idea is to modify or create new features to adapt to target domains while recognizing KKCs and identifying UUCs. Hence, it was concluded that features are crucial for implementing approaches aimed at OWL and should receive significant attention in developing future studies.

Moreover, in Figure 3, we underscore the notable advancements to OWL in the context of image segmentation. These strides were delineated based on the contributions of each study toward propelling the development of OWL within image segmentation. Additionally, the impact of each survey on the scholarly landscape was considered, considering the number of citations garnered by each work.

Upon observing Figure 3, it becomes apparent that a considerable portion of the milestones is associated with the inception of open-world learning, with many landmarks revolving around the proposition of novel tasks and approaches applied to image segmentation. These observations signify a substantial journey ahead for exploring OWL within the context of image segmentation by the research community.

4.3 Which are the main challenges reported by the selected works?

This section describes the main challenges related to OWL reported by the studies selected by the review:

4.3.1 Training data acquisition

Overall, the primary challenge highlighted by the selected studies revolves around the acquisition of images for training deep neural networks. This challenge stems from the remarkable success attained by deep learning models across various computer vision (CV) tasks, including scene understanding, object detection, image segmentation, and image captioning, among others.

Table 4: Taxonomy and a brief description of the leading approaches per task.

Task	Leading approaches	Description
Open World Learning	Vision-language learning [42–44]	It uses token/embedding information in the pixel-level or object-level learned by vision-language models to segment KKC and UUC objects
Continual Learning	Embedding Learning [45, 46]	Aim to learn how to convert high-dimensional data (e.g., in this case, classes of objects present in images) into low-dimensional data
	Knowledge distillation [47–51]	It is an approach based on knowledge transfer. It distills knowledge from previously trained models to new models by freezing parts of the network structure and using loss functions to minimize the probability divergence between models
	Search and data generation [52]	The strategy is to reproduce the knowledge learned through data generation networks that recreate the previous training data, or the information from the previous data is fetched on the internet to complement the training dataset
Open Set Recognition	Likelihoods [17, 53, 54]	This method builds models based on closed-set data distribution and uses its discriminator as an open-set likelihood function to recognize UUCs
	Anomaly Map [14, 55, 56]	This approach builds a pixel-level probability map based on the weights/outputs/prototypes/embeddings from the segmentation model, in which values over/below a set threshold are identified as UUCs
	Clustering [57–59]	The strategy consists of identifying patterns and shared characteristics among the closed-set data distribution to perform the grouping of classes. UUCs are identified through their dissimilarity from KKC groups
Incremental Learning	Fine-tunning [60]	It is a form of knowledge transfer that consists of freezing parts of the previously learned model and training only some layers with the new classes
	Knowledge destillation [18, 40, 61–67]	As described above for CL
	Endpoints Weight Fusion [68]	It is an incremental learning strategy that complements Knowledge distillation. In this strategy, the knowledge of the old model is merged with the model that holds the new knowledge in a dynamic fusion, reinforcing the memory of the old classes in the current segmentation.
	Weakly learning [69, 70]	The approach consists of pre-trained a model using expensive pixel-wise annotations. In the following steps, the model is updated to segment new classes provided image-level labels without access to old data
	Few-shot [14]	Use a few samples to learn the new class
Open Set Domain Adaptation	Data Generation [41, 71, 72]	Models based on this strategy rely on a model's ability to learn to classify a joint distribution of multidomain samples. The inclusion of a data generation method aims to alleviate the difference between domains, producing intermediate images or generating images from previously processed embedding
	Meta-Learning [39, 73, 74]	Meta-learning-based methods seek to optimize the parameters of a trained model to adjust to a new task quickly. This fit can be based on external data, metrics, or classifier performance
	Hybrid [75, 76]	It consists of a combination of data generation and meta-learning strategies

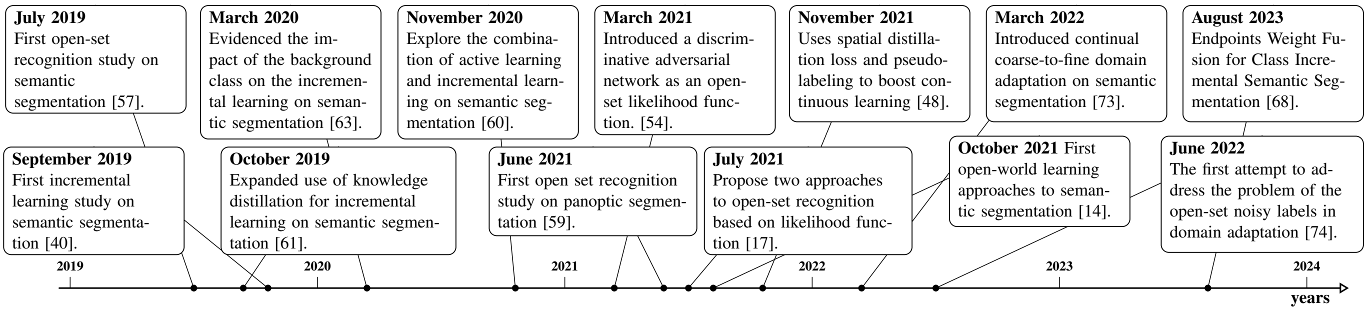


Figure 3: Recent timeline of landmarks in OWL studies for image segmentation.

Conversely, one of the predominant bottlenecks for current deep learning methods lies in their reliance on extensive training data to achieve meaningful performance [45].

However, the collection of labeled data presents a formidable hurdle for image segmentation, where annotations are conducted at the pixel level and often necessitate manual intervention by a human annotator. While common imaging scenarios may not require domain expertise for annotation, specific domains such as medical diagnostic imaging and remote sensing demand the involvement of subject matter experts, whose time is both scarce and invaluable, for reliable annotation. The annotation process becomes further challenging in multiclass scenarios and multidimensional data formats that transcend 2D dimensions, such as 3D images and Point Clouds, thereby increasing the annotation workload per image. The entire annotation process is also susceptible to errors due to the inherent human factor. Various methodologies, including Active Learning (AL) and Shot Learning (SL) tasks, aim to alleviate the annotation burden.

4.3.2 Closed-set perception

ML methods, particularly deep learning, have demonstrated significant efficacy across various computer vision (CV) tasks. However, these methods are typically trained on a static and Closed-Set Problem (CSP), assuming the availability of all data during training. This assumption proves invalid for real-world applications such as autonomous driving and remote sensing, where new data may emerge unpredictably during testing and application. Moreover, in high-risk applications like autonomous driving, reliance on a CSP can lead to catastrophic consequences due to erroneous attributions stemming from unknown data [15].

Another critical limitation is that these systems are static and incapable of updating their knowledge base in response to observed data. They are often constrained to specific scenarios and necessitate retraining after a certain period [14]. Additionally, storing massive amounts of data for training or model adjustment may not always be feasible [40]. Consequently, researchers advocate for developing dynamic systems that recognize and learn from new data.

In this context, the importance of tasks such as OSR, IL, CL, and Object OWL, as described in Section 2.5, is underscored. These tasks are pivotal in enabling ML systems to adapt to dynamic environments and effectively handle unknown data.

4.3.3 Catastrophic Forgetting

Due to the need to build dynamic systems to accommodate new classes encountered over time incrementally, some approaches are already being researched [18, 40, 62, 63, 65]. However, when exposing the model to a continuous data stream, it will suffer from the problem of Catastrophic Forgetting (CF) [77].

This problem consists of completely and abruptly forgetting previously learned knowledge when trying to learn new classes. This fact causes a significant decrease in the performance related to the previously learned classes due to changes in the parameters of the model [66]. CF presets are a real challenge for CL and OWL applications based on deep learning methods, especially when storing previously seen data is not allowed for privacy reasons [62]. Thus, one of the main challenges for building CL and OWL models is overcoming catastrophic forgetting.

4.3.4 Semantic shift

The semantic shift is caused by classes that influence model predictions due to their semantic similarity with other classes. These semantic aspects, such as color, texture, and shape, directly impact learning labels in deep neural networks of image segmentation. It makes distinguishing similar classes challenging, causing misclassification of pixels. The semantic shift is especially noticeable for open-world recognition for image segmentation in two situations.

First, it involves the background class, typically used in image segmentation tasks to denote pixels not assigned to any existing object category. Since this class may contain invisible classes, it often includes a substantial amount of semantic information that can negatively impact segmentation models. This semantic shift, known as background shift, has garnered attention in incremental learning methods [64, 66], which may exacerbate catastrophic forgetting. However, its potential impact in open-set recognition remains an area that requires further exploration, given that semantic information similar to UUCs may be present within the background class [78, 79].

Second, due to the large number of possible classes in a real scenario, it is natural to find semantic information of KKCs similar to UUCs [78, 79]. This similarity increases the difficulty of models in distinguishing between classes. For instance, while a person class may be easily distinguished from an animal class by human observers, human hair and animal hair may be similar, making it challenging for models to classify these parts correctly. This similarity can also be expanded to objects like sofas and rugs. Therefore, alleviating the semantic shift in open-set recognition is a field to be investigated.

4.3.5 Domain shift

In an ideal scenario, the test images should be captured under the same conditions as the training images so that the trained model does not suffer performance degradation due to possible domain discrepancies [80]. However, in the context of a real-world application, data samples often come from different domains. These domain changes may be related to different conditions when training and test images are acquired, such as climate changes, lighting variations, regional and cultural characteristics, and other image capture devices.

Although data in the source domain (training) and the target domain (test) may share the same semantic concepts at a high level, they are significantly different within the sample feature space, negatively impacting the performance of models trained in a given domain over time when being tested in other domains. This degradation problem can be avoided by collecting enough training data to cover all possible input distributions that may occur during testing. However, as discussed earlier, manual collection and annotation costs make this process unfeasible.

4.3.6 Computational power limitations

The last challenge pointed out by this review is the computational limitations for deploying applications in real environments. It is well known that deep neural networks demand a lot of computational resources. In many cases, it requires hours or days of processing to optimize the parameters of a classification or segmentation model. However, real-world applications such as autonomous driving and remote sensing are usually deployed on hardware with limited computing power, such as onboard computers and drones.

This problem could be solved by capturing the data and classifying the information in a second step. Most applications have only a few seconds to make some decisions based on the classification performed by the computational model. For example, seconds of delay in autonomous driving applications can result in disastrous accidents [15]. Nevertheless, this factor received little attention from the selected studies, which may be related to the recent introduction of most tasks studied in this review. We believe this is related to research strategies, due to this short maturity period. For instance, studies prioritize maximizing the performance of the models while computational costs are left in the background.

These computational resources can be left aside in offline applications that do not require real-time processing, but in real-world applications, it is a relevant factor in enabling the implementation of models created in real environments. Finally, with advances in technology, these computational limitations are naturally reduced. However, reducing the use of computing resources also has an economic impact by reducing production and maintenance costs.

5 Future Research Directions

This section aims to discuss the present open problems and future research directions that would be interesting to follow. Based on the research carried out and selected by this review.

5.1 Open Set Recognition

OSR is a crucial step in building an OWL system, as it helps recognize UUCs in data distributions. Traditional ML methods struggle to handle new data due to the presence of UUCs, leading to incorrect results. OSR approaches aim to incorporate this capability into ML algorithms but face numerous challenges, particularly in image segmentation tasks.

Under a CSP framework, ML methods assign a KKC label to UUCs, typically labeling them with the most similar KKC. However, this similarity is contingent on factors such as the application domain, extracted features, and the available KKCs [78, 79]. OSR algorithms must generate highly representative features to delineate class distinctions and control the feature space's dimensionality to address this challenge of data similarity. Effective control of the feature space is crucial for maximizing inter-class spacing, minimizing intra-class spacing, and accommodating new classes without significant distortions.

The semantic shift caused by the background class also makes it challenging to recognize UUCs, as objects irrelevant to specific tasks are often labeled with the background class, resulting in a mixture of disorganized semantic information. These challenges exponentially increase the complexity of OWL systems over time, as new UUCs are considered KKCs, impacting the similarity between classes and latent feature space. Recent works indicate that metric learning is a way to improve OSR [81, 82] and, consequently, the OWL [14, 83].

5.2 Incremental Learning

IL is a model that continuously learns new tasks without losing previously learned knowledge. Current models face challenges such as CF [84–86], maintenance costs, and the cost of manually labeling images for training. The high volume of data is another issue, as it can be expensive to store and maintain. However, in tasks like image segmentation, the cost of labeling can overcome other challenges, such as image classification.

For instance, in medical diagnostics, specialists often need to label images, which can be costly and time-consuming [38, 87]. Erroneous labeling can also negatively impact the performance of segmentation models [88]. The high volume of data also increases the training time of ML models, which should be close to real-time in a real environment. Studies have attempted to perform IL without storing data referring to classes learned [18, 40, 61], but their capacity needs to be evaluated in more complex scenarios.

One possible solution to these challenges is SL, where the ML model performs learning on a small dataset and seeks results similar to training on a large dataset [89]. However, CSP learned models are limited in most applications in real environments. For a real-world application, combining the capabilities of IL and SL is essential. This combination, introduced in [14], requires further investigation by researchers to find an optimal solution.

In conclusion, IL is a valuable learning method that can learn new tasks without losing previously learned knowledge. However, it faces challenges such as the high volume of data and labeling costs, making it essential to find a balance between IL and SL for real-world applications while controlling CF.

5.3 Domain Adaptation

The DA problem is a significant challenge in OWL studies, as ML algorithms are exposed to various variations in the application domain that differ from the learned data distribution. To adapt to these variations, algorithms should be trained on different datasets. However, the amount of information in these datasets is often lower than in diverse environments, such as urban or rural settings. Moreover, adjusting the model to climatic conditions, [71], such as snow, can create domain shifts, making it challenging to identify KKCs with distorted semantic information and recognize and learn UUCs.

Despite the complexity of adopting DA for OWL tasks, this capability is essential for enabling the application of models in different real-world scenarios, considering different local and global situations. Introductory studies [41, 73, 74, 90, 91] aim to incorporate DA with the presence of UUCs, addressing the challenge of identifying KKCs with distorted semantic information while learning UUCs. Overall, DA is essential for enabling the application of models in different real-world scenarios.

5.4 Active Learning

AL is a promising approach to reducing data annotation costs [92], particularly for image segmentation tasks that require pixel-level labeling. AL algorithms help external agents select a more representative subset of images over a large set of unlabeled data, boosting the training of ML algorithms and reducing human efforts in the annotation process. Thus, AL can perform a fundamental role in the complete OWL process, connecting the recognition of new data and the learning process.

However, most AL methods still require high computational costs and inefficient data selection [93]. The dataset for the image segmentation task is unbalanced, with certain classes being more present than others, distorting the performance of models towards the most representative classes [94]. Therefore, introducing AL in the OWL context still needs further investigation for more representative data selection techniques and more efficient strategies.

5.5 Datasets

Datasets have a crucial role in ML, fundamental for constructing algorithms capable of classifying or segmenting the data of interest. The performance of these algorithms, particularly deep learning approaches, is linked to the quality and quantity of data available for training [95]. Labeling failures are often due to the volume of data and manual image labeling processes. Tasks like AL and SL can be implemented to reduce data dependency and focus on quality labeling.

Additionally, it is essential to note that the datasets represent a small sample of the real world, making it necessary to build ML algorithms that can learn incrementally. To do this, training forms must be structured to simulate real-world learning. Most current segmentation datasets focus on CSP, requiring studies to adapt existing datasets to simulate the open-world environment. This involves subdividing data sets into small training sets based on available object classes.

The choice of classes and the number of images should directly influence the evaluations of the algorithms, as they are necessary to build rigorously detailed training protocols that allow replication of experiments by other researchers. Some datasets are already being used for this purpose, such as the Lost and Found [96], Road Anomaly [97], and StreetHazerds [98] datasets, are leveraged to evaluate IL and SL, but they are adaptations of existing datasets. The FSS-1000 dataset [99] is built explicitly for SL, but it cannot be used to evaluate an IL method due to the small number of images per class. If the IL algorithm is combined with SL, this dataset can serve as a benchmark for this challenge.

This review reveals that current evaluation solutions for OWL algorithms are limited to adapting existing datasets for the closed world [17, 59] or anomaly segmentation [14]. These adaptations require detailed protocols for reproducibility. Building an OWL-oriented dataset is highly desirable due to its evolution and attention in recent years. Alternatives, such as synthetically creating an OWL-oriented dataset, are being explored in studies [79].

6 Conclusions

This paper is a survey on OWL tasks focused on image segmentation. It is devoted to emerging topics related to building dynamic real-world systems and covers the main tasks and works in this research area. This work stands out for pointing out the

Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 23, Iss. 1, pp. 33–48, 2025
https://doi.org/10.21528/lnlm-vol23-no1-art3

leading solutions and limitations in this field.

We formally define the problems of semantic, instance, and panoptic segmentation and the open-world learning problem in the context of image segmentation. We also compare the tasks related to OWL, describing the differences between the tasks and complementing the knowledge on the topics.

Likewise, a strict research methodology was defined to cover the contemporary literature on image segmentation and OWL. As a result of this process, this review evaluated 39 relevant studies. Thus, it extracted significant information, such as the distribution of papers published in the last ten years, the leading solutions, and the main challenges.

In the end, we discuss the extracted data and provide useful information for future research directions and open problems in the area. A general conclusion that we can draw from this study is that open-world learning for image segmentation is a very recent problem. Therefore, it remains open to waiting for reasonable solutions that would increase the viability of ML algorithms in real environments. In this way, it expects numerous works to tackle this problem in the coming years, exploring points such as metric learning, federated learning, and hybrid neural networks.

REFERÊNCIAS

[1] R. C. Gonzalez and R. C. Woods. *Digital Image Processing*. Pearson Education, Londres, Inglaterra, fourth edition, 2018.

[2] X. Liu, Z. Deng and Y. Yang. “Recent progress in semantic image segmentation”. *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1089–1106, 2019.

[3] J. Long, E. Shelhamer and T. Darrell. “Fully convolutional networks for semantic segmentation”. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[4] K. He, G. Gkioxari, P. Dollár and R. Girshick. “Mask R-CNN”. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[5] A. Kirillov, K. He, R. Girshick, C. Rother and P. Dollár. “Panoptic segmentation”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019.

[6] G. Neuhold, T. Ollmann, S. Rota Bulo and P. Kontschieder. “The mapillary vistas dataset for semantic understanding of street scenes”. In *Proc. of the IEEE International Conference on Computer Vision*, pp. 4990–4999, 2017.

[7] Y.-H. Tseng and S.-S. Jan. “Combination of computer vision detection and segmentation for autonomous driving”. In *IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 1047–1052, 2018.

[8] Y. Zhang, H. Chen, Y. He, M. Ye, X. Cai and D. Zhang. “Road segmentation for all-day outdoor robot navigation”. *Neurocomputing*, vol. 314, pp. 316–325, 2018.

[9] H. Guo and D. Yang. “PRDNet: medical image segmentation based on parallel residual and dilated network”. *Measurement*, vol. 173, pp. 108661, 2021.

[10] R. Kemker, C. Salvaggio and C. Kanan. “Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning”. *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 60–77, 2018.

[11] X. Tao, D. Zhang, W. Ma, X. Liu and D. Xu. “Automatic metallic surface defect detection and recognition with convolutional neural networks”. *Applied Sciences*, vol. 8, no. 9, pp. 1575, 2018.

[12] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee and S. J. Lee. “Automated defect inspection system for metal surfaces based on deep learning and data augmentation”. *Journal of Manufacturing Systems*, vol. 55, pp. 317–324, 2020.

[13] A. Bendale and T. Boult. “Towards open world recognition”. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1893–1902, 2015.

[14] J. Cen, P. Yun, J. Cai, M. Y. Wang and M. Liu. “Deep metric learning for open world semantic segmentation”. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 15333–15342, 2021.

[15] D. Bozhinoski, D. Di Ruscio, I. Malavolta, P. Pelliccione and I. Crnkovic. “Safety for mobile robotic systems: A systematic mapping study from a software engineering perspective”. *Journal of Systems and Software*, vol. 151, pp. 150–179, 2019.

[16] Z. Chen and B. Liu. *Lifelong machine learning*. Williston, USA, 2018.

[17] 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.

[18] L. Shan, W. Wang, K. Lv and B. Luo. “Class-incremental learning for semantic segmentation in aerial imagery via distillation in all aspects”. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[19] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*, 2014.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. "Going deeper with convolutions". In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

[21] A. Krizhevsky, I. Sutskever and G. E. Hinton. "Imagenet classification with deep convolutional neural networks". *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[22] R. Strudel, R. Garcia, I. Laptev and C. Schmid. "Segmenter: Transformer for semantic segmentation". In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7262–7272, 2021.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin. "Attention is all you need". In *Proc. of the 30^{th} International Conference on Advances in Neural Information Processing Systems*, edited by I. Guyon *et al.*, 2017.

[24] H. Xu, B. Liu, L. Shu and P. Yu. "Open-world learning and application to product classification". In *Proc. The World Wide Web Conference*, pp. 3413–3419, 2019.

[25] K. J. Joseph, S. Khan, F. S. Khan and V. N. Balasubramanian. "Towards open world object detection". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5830–5840, 2021.

[26] M. Mundt, Y. W. Hong, I. Pliushch and V. Ramesh. "A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning". *ArXiv*, vol. abs/2009, 01797, 2020.

[27] J. Parmar, S. S. Chouhan, V. Raychoudhury and S. S. Rathore. "Open-world machine learning: applications, challenges, and opportunities". *ACM Computing Surveys*, 2021.

[28] A. Kirillov. "Exploring aspects of image segmentation: Diversity, global reasoning, and panoptic formulation". Ph.D. thesis, Heidelberg University, 2018.

[29] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota and T. E. Boult. "Toward open set recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013.

[30] W. J. Scheirer, L. P. Jain and T. E. Boult. "Probability models for open set recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2317–2324, 2014.

[31] 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.

[32] A. R. Naylor. "Known knowns, known unknowns and unknown unknowns: A 2010 update on carotid artery disease". *The Surgeon*, vol. 8, no. 2, pp. 79–86, 2010.

[33] A. R. Dhamija, M. Günther and T. E. Boult. "Reducing network agnostophobia". In *Proc. of the 32^{nd} International Conference on Neural Information Processing Systems*, p. 9175–9186, 2018.

[34] J. Weston, R. Collobert, F. Sinz, L. Bottou and V. Vapnik. "Inference with the Universum". In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 1009–1016, 2006.

[35] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman. "Systematic literature reviews in software engineering–a systematic literature review". *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009.

[36] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee. "A design science research methodology for information systems research". *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.

[37] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick. "Microsoft COCO: Common objects in context". In *Proc. European Conference on Computer Vision*, pp. 740–755, 2014.

[38] H. Tang, X. Liu, S. Sun, X. Yan and X. Xie. "Recurrent mask refinement for few-shot medical image segmentation". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 3918–3928, 2021.

[39] P. Zhang, J. Li, Y. Wang and J. Pan. "Domain adaptation for medical image segmentation: a meta-learning method". *Journal of Imaging*, vol. 7, no. 2, pp. 31, 2021.

[40] O. Tasar, Y. Tarabalka and P. Alliez. "Incremental learning for semantic segmentation of large-scale remote sensing data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3524–3537, 2019.

[41] M. Bucher, T.-H. Vu, M. Cord and P. Pérez. "Handling new target classes in semantic segmentation with domain adaptation". *Computer Vision and Image Understanding*, vol. 212, pp. 103258, 2021.

[42] J. Wu, Y. Jiang, B. Yan, H. Lu, Z. Yuan and P. Luo. "Exploring transformers for open-world instance segmentation". In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6611–6621, 2023.

[43] K. Cai, P. Ren, Y. Zhu, H. Xu, J. Liu, C. Li, G. Wang and X. Liang. "Mixreorg: Cross-modal mixed patch reorganization is a good mask learner for open-world semantic segmentation". In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1196–1205, 2023.

[44] J. Cha, J. Mun and B. Roh. "Learning to generate text-grounded mask for open-world semantic segmentation from only image-text pairs". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11165–11174, 2023.

[45] M. Siam, B. N. Oreshkin and M. Jagersand. "AMP: Adaptive masked proxies for few-shot segmentation". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 5249–5258, 2019.

[46] U. Michieli and P. Zanuttigh. "Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1114–1124, 2021.

[47] Y. Feng, X. Sun, W. Diao, J. Li, X. Gao and K. Fu. "Continual learning with structured inheritance for semantic segmentation in aerial imagery". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2021.

[48] A. Douillard, Y. Chen, A. Dapogny and M. Cord. "PLOP: Learning without forgetting for continual semantic segmentation". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4040–4050, 2021.

[49] C.-B. Zhang, J.-W. Xiao, X. Liu, Y.-C. Chen and M.-M. Cheng. "Representation Compensation Networks for Continual Semantic Segmentation". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7053–7064, 2022.

[50] M. H. Phan, S. L. Phung, L. Tran-Thanh, A. Bouzerdoum *et al.*. "Class Similarity Weighted Knowledge Distillation for Continual Semantic Segmentation". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16866–16875, 2022.

[51] F. Cermelli, M. Cord and A. Douillard. "Comformer: Continual learning in semantic and panoptic segmentation". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3010–3020, 2023.

[52] A. Maracani, U. Michieli, M. Toldo and P. Zanuttigh. "Recall: Replay-based continual learning in semantic segmentation". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 7026–7035, 2021.

[53] 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.

[54] S. Kong and D. Ramanan. "OpenGAN: Open-set recognition via open data generation". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 813–822, 2021.

[55] P. Bevandić, I. Krešo, M. Oršić and S. Šegvić. "Dense open-set recognition based on training with noisy negative images". *Image and vision computing*, vol. 124, pp. 104490, 2022.

[56] J. Li and Q. Dong. "Open-set semantic segmentation for point clouds via adversarial prototype framework". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9425–9434, 2023.

[57] Y. Nakajima, B. Kang, H. Saito and K. Kitani. "Incremental class discovery for semantic segmentation with RGBD sensing". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 972–981, 2019.

[58] M. Siam, A. Kendall and M. Jagersand. "Video class agnostic segmentation benchmark for autonomous driving". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2825–2834, 2021.

[59] J. Hwang, S. W. Oh, J.-Y. Lee and B. Han. "Exemplar-based open-set panoptic segmentation network". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1175–1184, 2021.

[60] Y. Lin, G. Vosselman, Y. Cao and M. Y. Yang. "Active and incremental learning for semantic ALS point cloud segmentation". *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 73–92, 2020.

[61] U. Michieli and P. Zanuttigh. "Incremental learning techniques for semantic segmentation". In *Proc. of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1–8, 2019.

[62] U. Michieli and P. Zanuttigh. "Knowledge distillation for incremental learning in semantic segmentation". *Computer Vision and Image Understanding*, vol. 205, pp. 103167, 2021.

[63] F. Cermelli, M. Mancini, S. R. Bulo, E. Ricci and B. Caputo. "Modeling the background for incremental learning in semantic segmentation". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9233–9242, 2020.

[64] F. Cermelli, M. Mancini, S. R. Bulo, E. Ricci and B. Caputo. "Modeling the background for incremental and weakly-supervised semantic segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 10099–10113, 2021.

[65] W. He, X. Wang, L. Wang, Y. Huang, Z. Yang, X. Yao, X. Zhao, L. Ju, L. Wu, L. Wu, H. Lu and Z. Ge. "Incremental learning for exudate and hemorrhage segmentation on fundus images". *Information Fusion*, vol. 73, pp. 157–164, 2021.

[66] J. Li, X. Sun, W. Diao, P. Wang, Y. Feng, X. Lu and G. Xu. "Class-incremental learning network for small objects enhancing of semantic segmentation in aerial imagery". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–20, 2021.

[67] G. Yang, E. Fini, D. Xu, P. Rota, M. Ding, M. Nabi, X. Alameda-Pineda and E. Ricci. "Uncertainty-aware contrastive distillation for incremental semantic segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2567–2581, 2022.

[68] J.-W. Xiao, C.-B. Zhang, J. Feng, X. Liu, J. van de Weijer and M.-M. Cheng. "Endpoints weight fusion for class incremental semantic segmentation". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7204–7213, 2023.

[69] F. Cermelli, D. Fontanel, A. Tavera, M. Ciccone and B. Caputo. "Incremental learning in semantic segmentation from image labels". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4371–4381, 2022.

[70] Y.-H. Hsieh, G.-S. Chen, S.-X. Cai, T.-Y. Wei, H.-F. Yang and C.-S. Chen. "Class-incremental continual learning for instance segmentation with image-level weak supervision". In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1250–1261, 2023.

[71] Ö. Erkent and C. Laugier. "Semantic segmentation with unsupervised domain adaptation under varying weather conditions for autonomous vehicles". *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3580–3587, 2020.

[72] J. Wang and J. Jiang. "Learning across Tasks for Zero-Shot Domain Adaptation from a Single Source Domain". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[73] D. Shenaj, F. Barbato, U. Michieli and P. Zanuttigh. "Continual coarse-to-fine domain adaptation in semantic segmentation". *Image and Vision Computing*, vol. 121, pp. 104426, 2022.

[74] X. Guo, J. Liu, T. Liu and Y. Yuan. "SimT: Handling Open-set Noise for Domain Adaptive Semantic Segmentation". In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7032–7041, 2022.

[75] Z. Wu, X. Wang, J. E. Gonzalez, T. Goldstein and L. S. Davis. "ACE: Adapting to changing environments for semantic segmentation". In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 2121–2130, 2019.

[76] K. Li, L. Yu and P.-A. Heng. "Domain-incremental Cardiac Image Segmentation with Style-oriented Replay and Domain-sensitive Feature Whitening". *IEEE Transactions on Medical Imaging*, 2022.

[77] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*. "Overcoming catastrophic forgetting in neural networks". *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[78] Z. Cui, W. Longshi and R. Wang. "Open set semantic segmentation with statistical test and adaptive threshold". In *Proc. of the IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2020.

[79] A. Brilhador, 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*, pp. 1–8, Joinville, Brazil, 2021. SBIC.

[80] A. Gretton, A. Smola, J. Huang, M. Schmittfull, K. Borgwardt and B. Schölkopf. "Covariate shift by kernel mean matching". In *Dataset Shift in Machine Learning*, edited by J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer and N. D. Lawrence, pp. 131–160. MIT Press, 2009.

[81] M. Gutoski, A. E. Lazzaretti and H. S. Lopes. "Deep metric learning for open-set human action recognition in videos". *Neural Computing and Applications*, vol. 33, pp. 1207–1220, 2021.

[82] A. D. S. Inácio, M. Gutoski, A. E. Lazzaretti and H. S. Lopes. "OSVidCap: A framework for the simultaneous recognition and description of concurrent actions in videos in an open-set scenario". *IEEE Access*, vol. 9, pp. 137029–137041, 2021.

[83] M. Gutoski, A. E. Lazzaretti and H. S. Lopes. "Incremental human action recognition with dual memory". *Image and Vision Computing*, vol. 116, pp. 104313, 2021.

[84] A. Robins. "Catastrophic forgetting, rehearsal and pseudorehearsal". *Connection Science*, vol. 7, no. 2, pp. 123–146, 1995.

[85] S. Thrun. “Lifelong learning algorithms”. In *Learning to learn*, pp. 181–209. 1998.

[86] R. M. French. “Catastrophic forgetting in connectionist networks”. *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 128–135, 1999.

[87] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab and C. Wachinger. “‘Squeeze & excite’guided few-shot segmentation of volumetric images”. *Medical Image Analysis*, vol. 59, pp. 101587, 2020.

[88] Z. Zhang, Q. Peng, S. Fu, W. Wang, Y.-M. Cheung, Y. Zhao, S. Yu and X. You. “A componentwise approach to weakly supervised semantic segmentation using dual-feedback network”. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[89] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr and T. M. Hospedales. “Learning to compare: Relation network for few-shot learning”. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1199–1208, 2018.

[90] P. Panareda Busto and J. Gall. “Open set domain adaptation”. In *Proc. of the IEEE International Conference on Computer Vision*, pp. 754–763, 2017.

[91] H. Liu, Z. Cao, M. Long, J. Wang and Q. Yang. “Separate to adapt: Open set domain adaptation via progressive separation”. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2927–2936, 2019.

[92] S. Budd, E. C. Robinson and B. Kainz. “A survey on active learning and human-in-the-loop deep learning for medical image analysis”. *Medical Image Analysis*, vol. 71, pp. 102062, 2021.

[93] Z. Zhao, Z. Zeng, K. Xu, C. Chen and C. Guan. “DSAL: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation”. *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3744–3751, 2021.

[94] U. A. Usmani, J. Watada, J. Jaafar, I. A. Aziz and A. Roy. “A reinforced active learning algorithm for semantic segmentation in complex imaging”. *IEEE Access*, vol. 9, pp. 168415–168432, 2021.

[95] R. Sheikh, A. Milioto, P. Lottes, C. Stachniss, M. Bennewitz and T. Schultz. “Gradient and log-based active learning for semantic segmentation of crop and weed for agricultural robots”. In *Proc. IEEE International Conference on Robotics and Automation*, pp. 1350–1356, 2020.

[96] P. Pinggera, S. Ramos, S. Gehrig, U. Franke, C. Rother and R. Mester. “Lost and found: detecting small road hazards for self-driving vehicles”. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1099–1106, 2016.

[97] K. Lis, K. Nakka, P. Fua and M. Salzmann. “Detecting the unexpected via image resynthesis”. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, pp. 2152–2161, 2019.

[98] D. Hendrycks, S. Basart, M. Mazeika, M. Mostajabi, J. Steinhardt and D. Song. “Scaling out-of-distribution detection for real-world settings”. *arXiv preprint arXiv:1911.11132*, 2019.

[99] X. Li, T. Wei, Y. P. Chen, Y.-W. Tai and C.-K. Tang. “FSS-1000: A 1000-class dataset for few-shot segmentation”. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2869–2878, 2020.