

## Article

# Ten Years of Active Learning Techniques and Object Detection: A Systematic Review

Dibet Garcia <sup>1,\*</sup>, João Carias <sup>1</sup>, Telmo Adão <sup>2,3</sup>, Rui Jesus <sup>4</sup>, Antonio Cunha <sup>2,5</sup> and Luis G. Magalhães <sup>3</sup>

- <sup>1</sup> Centro de Computação Gráfica—CCG/zgdv, University of Minho, Campus de Azurém, Edifício 14, 4800-058 Guimarães, Portugal; joao.carias@cpg.pt
- <sup>2</sup> Department of Engineering, School of Sciences and Technology, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal; telmoadao@utad.pt (T.A.); acunha@utad.pt (A.C.)
- <sup>3</sup> ALGORITMI Research Centre/LASI, University of Minho, 4800-058 Guimarães, Portugal; Imagalhaes@dsi.uminho.pt
- <sup>4</sup> Department of Computer Science, Faculty of Computer Science, Campus Elviña S/N, University of A Coruña, E-15071 A Coruña, Spain
- <sup>5</sup> INESC TEC—Institute for Systems and Computer Engineering, Technology and Science, Rua Doutor Roberto Frias, 4200-465 Porto, Portugal
- \* Correspondence: dibetg@gmail.com; Tel.: +351-932-170-505

**Abstract:** Object detection (OD) coupled with active learning (AL) has emerged as a powerful synergy in the field of computer vision, harnessing the capabilities of machine learning (ML) to automatically identify and perform image-based objects localisation while actively engaging human expertise to iteratively enhance model performance and foster machine-based knowledge expansion. Their prior success, demonstrated in a wide range of fields (e.g., industry and medicine), motivated this work, in which a comprehensive and systematic review of OD and AL techniques was carried out, considering reputed technical/scientific publication databases—such as ScienceDirect, IEEE, PubMed, and arXiv—and a temporal range between 2010 and December 2022. The primary inclusion criterion for papers in this review was the application of AL techniques for OD tasks, regardless of the field of application. A total of 852 articles were analysed, and 60 articles were included after full screening. Among the remaining ones, relevant topics such as AL sampling strategies used for OD tasks and groups categorisation can be found, along with details regarding the deep neural network architectures employed, application domains, and approaches used to blend learning techniques with those sampling strategies. Furthermore, an analysis of the geographical distribution of OD researchers across the globe and their affiliated organisations was conducted, providing a comprehensive overview of the research landscape in this field. Finally, promising research opportunities to enhance the AL process were identified, including the development of novel sampling strategies and their integration with different learning techniques.

**Keywords:** active learning; sampling strategies; acquisition function; object detection; score; confidence; uncertainty; diversity; aggregation



**Citation:** Garcia, D.; Carias, J.; Adão, T.; Jesus, R.; Cunha, A.; Magalhães, L.G. Ten Years of Active Learning Techniques and Object Detection: A Systematic Review. *Appl. Sci.* **2023**, *13*, 10667. <https://doi.org/10.3390/app131910667>

Academic Editor: Oscar Reinoso García

Received: 8 May 2023

Revised: 17 September 2023

Accepted: 18 September 2023

Published: 25 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

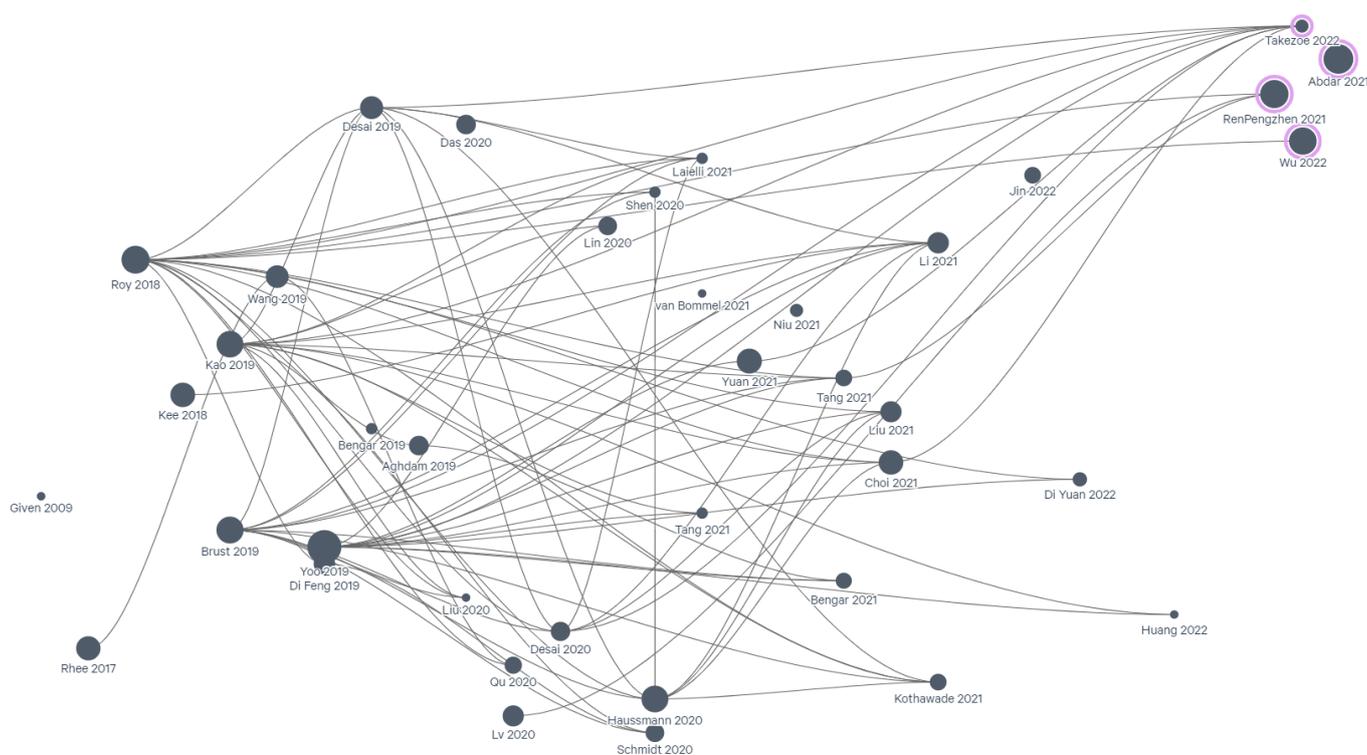
## 1. Introduction

Artificial Intelligence (AI) is becoming increasingly present in many quotidian and professional contexts (e.g., cybersecurity activities [1], medical applications [2], industrial procedures [3], etc.). Within the field of AI are included the machine learning (ML) algorithms that continuously improve as more data are considered. When these algorithms are implemented as multi-layer neural networks, they become part of the deep learning (DL) branch of AI [4]. This inference strategy has proven to be highly effective in solving numerous computer vision challenges when trained with vast amounts of data. However, the demands in terms of data quantity and variability also pose non-negligible challenges associated with this class of algorithms, which are especially relevant in areas such as the

medical field or the industry, wherein the availability of human resources is not compatible with the laborious and time-consuming operations that are required to collect and screen balanced, qualitative, and representative sets of problem-oriented training examples, which also usually involve burdensome computation when it comes to building high-performance AI models.

Considering the previously identified issue, it is important to achieve the integration of AI algorithms and methods capable of optimising the human interaction required for the construction of suitable datasets, as well as computational resources, without penalising the expected models' learning ability, especially in the context of object detection (OD) problems, wherein the identification of regions of interest per image is crucial for a proper computational training. A valid way to overcome these challenges lies in the application of AL techniques [5–8], which focuses on the acquisition of a large and diverse set of examples, while reducing the time needed by researchers/professionals to screen the data. Therefore, the underlying operational and computational optimisation potential associated with AL sets this topic as a relevant one for the sustainable implementation and deployment of AI/ML/DL.

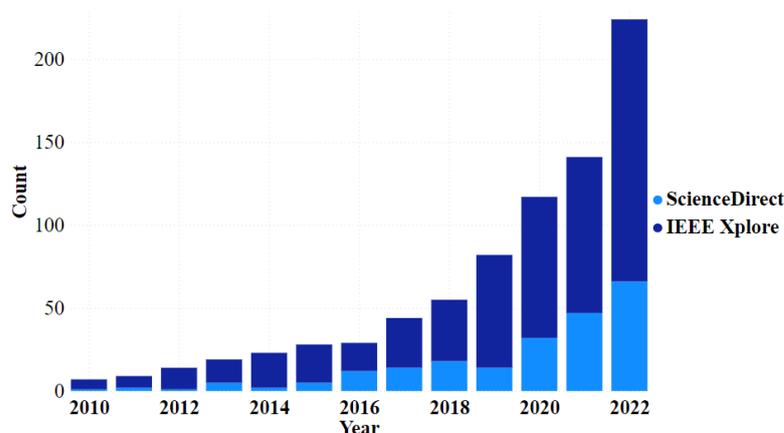
As such, in this work, we present a systematic review of the AL strategies focused on the iterative labelling step applied in the training of deep OD neural networks, covering a topic that seems to be scarcely addressed in the literature [9]. Furthermore, previous reviews [8,10–13] are more oriented to other AI areas such as natural language processing, speech processing, and computer vision, but without going deeper into OD, showing only a few related papers. Additionally, as can be observed in the graph depicted in Figure 1, which explores the relations between previous reviews and the other papers included in this work concerning AL and OD, there is an apparent lack of association between the two terms.



**Figure 1.** Relation between the articles included in this review and previous published reviews. The diagram was constructed using the Litmaps tool available online in <https://www.litmaps.com/> (accessed on 11 July 2023). The four at the top/right are the previous reviews.

On the other hand, the growth of the interest around topics such as “active learning”, “object detection”, “acquisition function”, “query strategy”, “uncertainty”, “diversity”, and

“sampling strategy” by the scientific community over the last years is perceptible, as shown by Figure 2.



**Figure 2.** AL/OD topic interest growth, from 2010—wherein works such as [14] were among a scarce set of proposals—to 2022, in which a much wider diversity of contributions, for example, [15–18], can be found. Results obtained using the IEEE Xplore and the Science Direct platforms with the terms “active learning” AND “object detection” AND (“acquisition function” OR “query strategy” OR “uncertainty” OR “diversity” OR “sampling strategy”).

Therefore, the main purpose of this work is to provide an overview of the technical/scientific contributions combining AL and OD that have been proposed over the past ten years.

With this work, AL synergy with OD is profiled in terms of tendencies and several contributions are summed up, seeking to answer to the following main research questions (RQ):

- (RQ1) What is the geographical distribution of researchers, in terms of countries and organisations?
- (RQ2) In which domains have the application of AL for OD tasks been reported?
- (RQ3) What are the most commonly used deep OD neural network (NN) architectures?
- (RQ4) What sample strategies of AL have been proposed and how are they grouped?
- (RQ5) How are they combined with other learning techniques?
- (RQ6) What are the strengths, weaknesses, opportunities, and threats (SWOT) associated with the implementation of AL methods for OD in computer vision?

In addition to summarising the literature-based contributions that apply AL in OD tasks, other important actions with a motivational essence are taken in this review proposal:

1. To survey the methods capable of computing per-image global scores, as alternatives to address the limitations of confidence-based methods in scoring samples for the query strategy in OD tasks [10];
2. To identify the main approaches dedicated to reduce the efforts required to mark and label an OD dataset;
3. To point out the main strategies to perform sampling, along with their corresponding strengths and weaknesses;
4. Finally, to identify the challenges and opportunities that are currently associated to AL methods for OD tasks.

The rest of this work is organised as follows: Section 2 consists of a contextualisation of the main addressed topics; then, in Section 3, the methodology for carrying out this review in a systematic style is presented; afterwards, Section 4 addresses, with AL/OD as the central research topic, the question posed previously (related to geographic distribution of the authors, application domains, most used artificial neural networks architectures, active

learning sampling strategies, etc.); Section 5 elaborates on these results and conclusions are drawn in Section 6.

## 2. Contextualisation

Depending on the task, DL algorithms (applied to computer vision field) can be divided into three categories: classification, segmentation, and object detection [19].

For conciseness purposes, and to focus the subject of interest of this work, only the applications of AL for OD tasks are addressed.

### 2.1. Evolution of the Object Detection Architectures

Object detection algorithms’ main purpose lies in accurately identifying and locating objects in images or videos, making it a valuable practical technology in the real world. OD algorithms allow the recognition of objects in an image and have increasingly received attention from researchers in the last ten years. In 2012, AlexNet [20] was a game changer for the field of computational inference (see Figure 3). This architecture was capable of generating models for detecting objects in an image and, for each object, output its location, a predefined label, and a prediction confidence value. Mainly, they are grouped into two categories [21,22]—single-stage detectors (e.g., YOLO versions [23–26], SSD [27], RetinaNet [28], RefineDet [29], EfficientDet [30]) or two-stage detectors (e.g., R-CNN [31], Fast R-CNN [32], Faster R-CNN [33]).

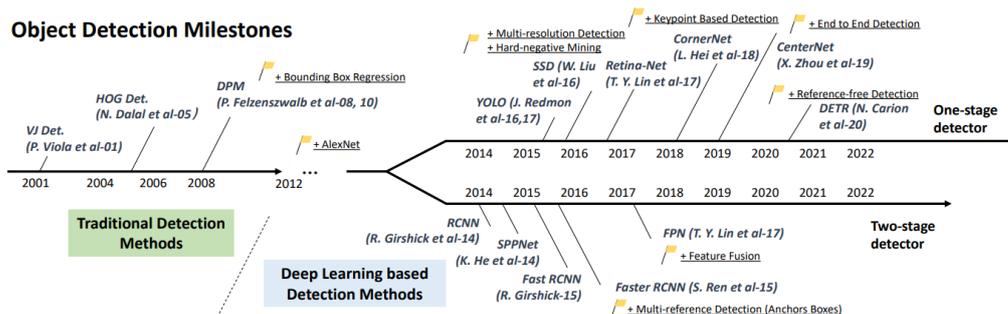


Figure 3. Object detection road map since 2001. Image taken from [21]. Similar evolution was presented in [22].

As OD algorithms belong to the DL family, a large amount of data is required for the learning process, and special attention needs to be given to the quality of data. For example, it is well known that the performance of OD algorithms is improved when the training uses large amounts of labelled data.

The conventional procedure when using supervised DL can be subdivided in the following operational steps: (i) collect a dataset as large as possible; (ii) assign labels to images, pixels, or regions; (iii) train models on that dataset; and (iv) place the model to production. So far, this strategy has worked with different degrees of success, albeit with high human and computational resource costs. Another type of approach, known as active learning, consists in using the knowledge learned by the model (from a small dataset) to filter and rank the images provided to the specialists during the annotation task. It is a powerful technique for improving data efficiency (i.e., avoiding significant overlap between data instances [34]), as it allows supervised learning algorithms to achieve the desired performance using a small training set at time, diluting the annotation efforts by more but lighter sessions.

### 2.2. Active Learning Scenarios

The idea of AL is to enable the selection of fewer training instances (e.g., images or patches) to achieve a higher performance when they are labelled by an oracle (i.e., human expert). It is an iterative process, repeated over time, that combines data sampling and model training. The number of learning cycles or iterations  $K$  is defined according to the

following equation  $K = B/b$ , where  $B$  is the total budget for labelling and  $b$  is the labelling budget per cycle, also known in some works as the batch size. Another condition that allows stopping the iteration process is related to a performance goal, which is calculated via inference on the validation and test sets that are selected in the beginning of the process and kept until it finishes.

The choice of AL technique to be applied depends on the type of application. Generally, there are three main scenarios in AL, as discussed in [11,35]:

- Membership query synthesis: this scenario involves generating new unlabelled instances to be queried. The AL algorithm assesses the usefulness of creating new images that need to be labelled and added to the annotated image dataset.
- Stream-based selective sampling: in this scenario, unlabelled instances are observed one at a time, and query selection is performed sequentially. The algorithm determines whether each instance should be selected for labelling or not.
- Pool-based sampling: this scenario involves generating a query by evaluating the entire “pool” of unlabelled instances. Since deep neural network training is typically batch-based, there is a need for Batch Active Learning (BAL) techniques. BAL selects multiple instances from the pool to be labelled [9,36,37]. In [8], it is also referred to as Batch Mode Deep Active Learning (BMDAL), which is considered one of the most efficient strategies for OD.

The pool-based scenario assumes three pools of images or patches: (i) unlabelled ( $U_0$ ), (ii) selected ( $S_0$ ), and (iii) labelled ( $L_0$ ). On each iteration  $I_{(0...K)}$ , a set of images  $b$  from  $U_i$  is included in  $S_i$  according to an acquisition function  $f$ , and the images in  $S_i$  are labelled by the external oracle and included into  $L_{(i+1)}$ . Also, models are trained using the  $L_i$  set of images and producing AI models  $M_i$  that can be used as part of the acquisition function  $f(U_0, M_i)$ . At the end of each iteration [35]:

$$L_{(i+1)} = S_i \cup L_i \quad (1)$$

$$U_{(i+1)} = U_i - S_i. \quad (2)$$

The process ends when a desired performance is achieved, or the annotation budget is spent. Pool-based sampling strategies are very time-consuming since they require the evaluation and ranking of all the instances of unlabelled data in order to select the top  $N$  samples for model training [38]. To reduce the processing time needed to rank all the instances, some methods apply a sub-pooling strategy [39].

### 2.3. Active Learning in Deep Learning

Active learning has been successful in applications involving image classification [39,40], segmentation [41], pose estimation [7], and visual tracking [42] while its use in other tasks, such as OD, needs more attention [36]. Active learning for OD aims to decrease labelling costs by selecting the most interesting samples—depending on a certain criteria—from the unlabelled collection, aiming to improve the training effectiveness for improved detection models, allowing to reach suitable performances using a less demanding number of examples, in a particular dataset. Techniques include estimating scores either by statistical methods [36] or by using learning strategies [43]. Scores can also be computed at different levels (e.g., pixel, image [39,44], box [45,46], or using combinations of the former elements [47,48]). In the methods found in literature, a wide range of sample selection criteria were explored, such as: uncertainty, diversity, density [49], rareness [15,50], or a mixture of them to query instances according to a previously computed score.

Active learning has been applied in various fields and for several purposes, where the main task is to detect objects. In medical applications, AL methods have used AI models to suggest which images to annotate, to reduce the workload and to optimise the available budget of specialists [41,51]. Particularly, in digital pathology, AL can be used to annotate whole slide images (WSI) in an efficient way, since pathologists receive a reduced set of

samples or patches from the set of unlabelled images. Military applications, such as aircraft detection on very high-resolution imagery, are using AL too [6], to tackle the laborious nature of the tasks involved in the annotation of WSI. In the field of autonomous driving, AL is used to annotate trajectory time-series data and discover new unknown classes [52], to locate 3D objects from 2D image detector [53], to estimate sample uncertainties [54] or diversity [16] for 2D or 3D scenarios, to perform video OD in road scenes [9], and to label portion of the dataset on a device—federated learning—avoiding the transmission of the information [34] through communication networks, among other purposes. Also, other applications benefit from AL, such as defect detection [55] and virtual reality-based pose estimation [7].

Applying traditional AL techniques when using DL for OD poses some challenges [51]. An issue reported by authors results from the use of softmax as a score to select unlabelled instances when the current knowledge of the network can be used to select informative samples [47]. Another cause is that the models learn from many features that may not be representative, decreasing the inference capacity. Another identified cause relies in the evidence that DL algorithms perform better on batch samples than traditional techniques do on an instance-by-instance basis. Also, there are additional issues, related to the training process, to overcome [56]:

- Dataset annotation for OD requires a bounding box and the respective category for each object in the image, making the annotation a labour-intensive task compared with classification or action recognition;
- It is challenging to identify images with real potential to improve performance;
- Finally, there are outliers and noisy instances that can affect the final performance.

As such, there is still space for improvements in the scope of AL sampling strategies focused on OD, to more effectively reduce labelling costs while aiming at potential top-performances, which may vary in accordance with the adopted artificial network architecture, by selecting the most suitable samples from the unlabelled sets of data.

The next chapter will address the methodology used to search, collect, and process the scientific documents considered in this review.

### 3. Research Methodology

With the intention to produce a systematic review of AL methods used in combination with image-based OD, a preliminary search was carried out to obtain an overview of the current state of the scientific literature regarding the highlighted topic. Five recent AL reviews [8,10–13] were identified in that process, along with titles, abstracts, and contents that were used to elaborate the search terms. The human-in-the-loop term was excluded from the search criteria, as it is a much wider concept than the scope defined for this paper. The most recent review found [13] was published in November 2022 and addresses the applications of AL in several computer vision tasks, including OD, but with a superficial contribution. In contrast, this review tightens the subject mesh, but explores, in more detail, some relevant aspects within the scope of AL-OD combination (e.g., methods capable of computing per-image global scores, approaches dedicated to reducing the efforts required to mark and label datasets, suitable sampling strategies, as well as the main challenges and opportunities to be addressed).

Therefore, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) procedure (Figure 4), the next step consisted of performing an extensive search in databases and open-access repositories, including Science Direct, IEEE Xplore, arXiv, and PubMed.

The search criteria combined the following keywords: “active learning” AND “object detection” AND (“acquisition function” OR “query strategy” OR “uncertainty” OR “diversity” OR “sampling strategy”). Then, documents that cited any of the initially included studies as well as their own references were considered. However, no extra articles fulfilling the inclusion criteria were found. In terms of numbers, 852 records were initially collected, from which four repeated papers were excluded and 848 records were screened, resulting

in 60 full-text documents that were considered for the final set of papers to use in this review. The data collected to keep track of the documents during the whole process were as follows:

- Report: authors, institutions, year, and source of the publications.
- Study: active learning scenarios, active learning query strategies, object detection architectures, application fields, metrics, and active learning categories.

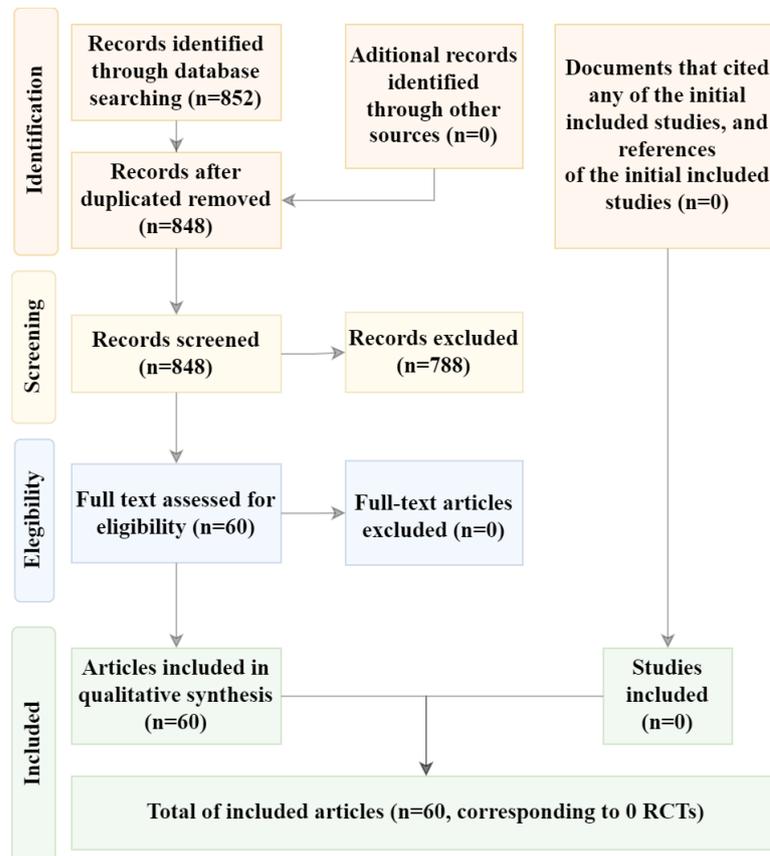


Figure 4. PRISMA flow diagram to collect and screen data from articles.

Table 1 presents the inclusion and exclusion criteria for the article selection.

Table 1. Inclusion and exclusion criteria for article selection.

Inclusion Criteria	Exclusion Criteria
Published between 2010 (first article available) and December 2022.	Published before 2010.
Papers from peer-reviewed journals and conferences.	Published as editorial, poster, newsletters, letter to editor, interviews, commentaries or abstract only.
Published in English.	Published in a different language than English.
Application of AL in the computer vision task of OD.	Applications of AL on segmentation or classification.
Application of AL in 2D or 3D environment.	

#### 4. Active Learning and Object Detection: A Literature-Driven Multi-Perspective Analysis

This section addresses the different perspectives tied to the research questions regarding AL and OD formerly presented in the introduction section, including the geographical distribution of the researchers, applications domains, deep neural networks, query strategies taxonomy, and informativeness-based methods.

4.1. (RQ1) Geographical Distribution

An analysis of the research groups or technological companies working in AL for OD showed that Asia (China), North America (United States of America), and Europe (Germany) are producing the main advances in this research field (e.g., [5,36,55]), as can be seen in Figure 5. Also, many large organisations have been contributing to the topic, such as Google, MIT, Microsoft, Amazon, and NVIDIA, who are all actively researching in this field (see Tables A1–A3).

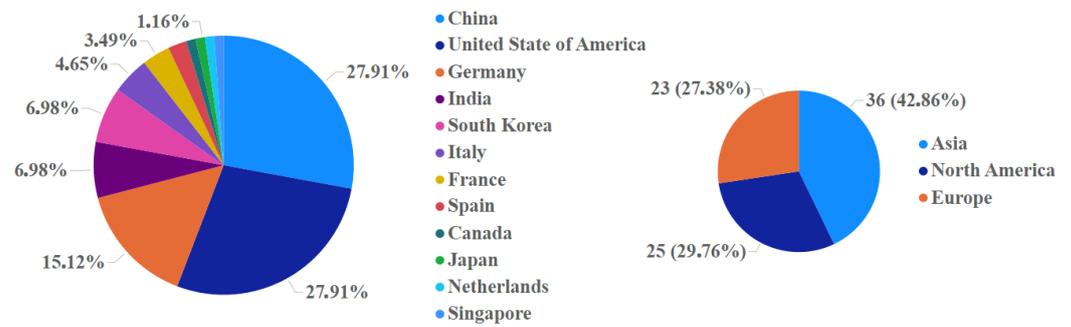


Figure 5. Count of institutions researching AL for OD, their countries (left) and geographical regions (right). The references used to create the graphs can be found at Tables A1–A3.

4.2. (RQ2) Application Domains

In this work, the areas wherein AL techniques were applied were considered for statistical purposes ( see Table 2). In the health area, four articles were found [37,57–59], making evident the lack of the use of AL to facilitate the annotation process by specialists, such as pathologists. Other applications such as document layout detection [14,17], pose estimation [7], and surface defect detection [55] remain under-explored, as well.

Table 2. References grouped by application (also indicating the public dataset of reference, whenever the information is available).

Application	3D Context	2D Context
General—PASCAL Dataset		[5,18,34,43–46,48–50,56,57,60–66]
General—MS-COCO Dataset		[5,18,34,44,48,56,57,61,64,67]
Autonomous driving, KITTI Dataset , Waymo Open Dataset , nuScenes	[15,16,53,54]	[34,36,37,46,54,57]
Healthcare		[37,57–59]
Caltech Pedestrian Detection		[47,67]
Documents (Pub-LayNet, PRImA, HJDataset, cBAD and Horae)		[14,17]
BDD100K		[47,65]
Bank		[37]
Pose Estimation		[7]
ILVRC Object Detection		[67]
SYNTHIA-AL		[9]
Sorghum and Wheat		[68]
Surface defect dataset NEU-DET		[55]
DOTA		[69]
ImageNetVID		[9]
General—ImageNet		[70]
Stanford Dog dataset		[59]
Yeast		[37]
Table-Top		[71]
iCubWorld Transformations (iCWT)		[71]
YCB-Video		[71]
Cityscapes		[65]
CityPersons		[47]

The implementation of AL experiments, specifically those that allow the comparison of sampling methods, does not require the intervention of the annotators. The use of public databases such as PASCAL [72], MS-COCO [73], or autonomous driving [74–76] simplifies the performance of experiments to compare sampling methods. On the other hand, in cases wherein representative datasets are required but the available data are rather insufficient; AL involving human-in-the-loop strategies to validate (accept or reject) machine-based annotation proposals stand out as a solution to progressively assemble sets of well-structured examples that sustainably support computational learning, allowing us to build increasingly effective inference models over time.

Even though a significant part of the works found in the literature addresses OD over 2D data (see Table 2), 3D data are also focused on by other approaches (e.g., [15,53]), in which the application of AL techniques proved to be beneficial. A method that combines 3D LiDAR and 2D data is proposed in [53], where the bounding boxes obtained by the 2D image detector are used to locate 3D objects. Afterwards, the human annotator labels only the LiDAR points with frustums. Uncertainty is estimated using three different classification predictive probabilities for comparison purposes: (i) softmax output, (ii) Monte-Carlo dropout, and (iii) deep ensembles. Later, the informativeness or Uncertainty is calculated using Shannon Entropy and Mutual Information to analyse different points of view of the epistemic uncertainty, such as predictive uncertainty and confidence in the data, respectively.

Schmidt et al. [54] propose four ensemble-based methods to estimate sample uncertainties for 2D and 3D object detection: Consensus Score, Consensus Score Variation Ratio, Region of Interest (RoI) Matching, and Sequential RoI Matching. They propose an Ensemble-based AL method with the assumption that if Ensembles for classification are better than other techniques, the same should be observable in OD problems. Also, they consider two AL strategies for OD, namely: (i) Continuous Training (which outperforms Training from Scratch) and (ii) Active Class Weighting. This work was supported by previous works [5,53]. Figure 6 presents the proposed setup.

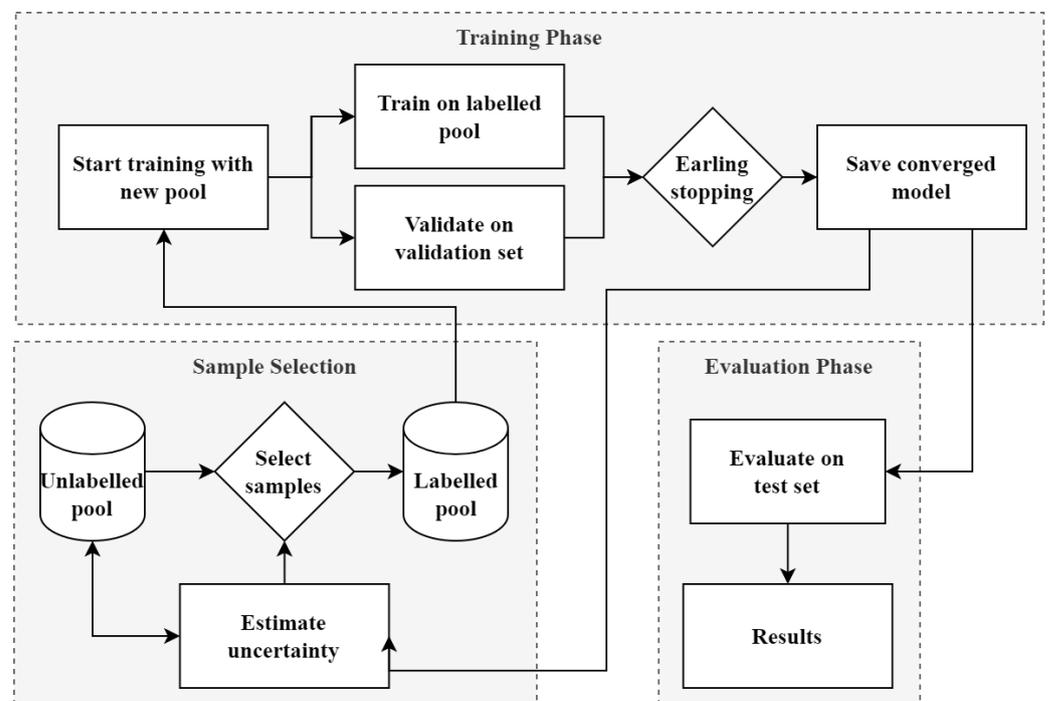


Figure 6. Active learning setup from [54].

The authors of [16] proposed the use of multimodal data (GPS, IMU, and LiDAR information) in the autonomous vehicle dataset nuScenes [76] to sample instances according to pure diversity criteria. Two diversity terms, spatial- and temporal-diversity objectives,

were proposed to achieve a better sampling performance when compared with the random and entropy-based methods.

#### 4.3. (RQ3) Deep Neural Network Architectures

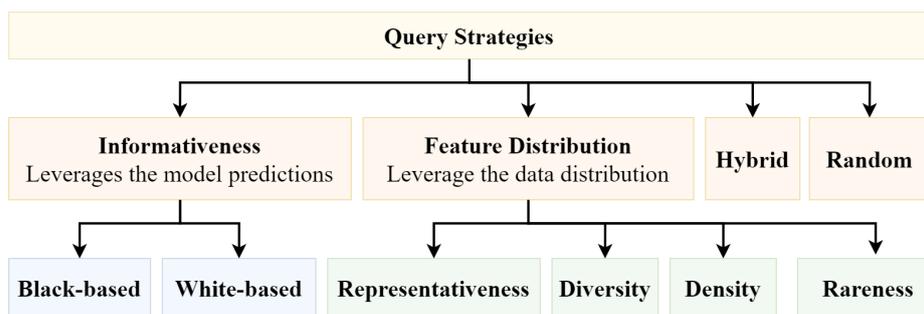
The most relevant impacts of AL methods were observed from 2014, after the releasing of architectures such as YOLO [23], Faster R-CNN [33], among others, which attracted the interest of researchers and practitioners in applying AL in OD, as can be concluded from Figure 3. Table 3 presents the most-reported deep neural network architectures reviewed in this work.

**Table 3.** References grouped by deep learning architectures.

Architecture	References
Faster R-CNN	[5,9,14,44,45,48,50,54,56,59–61,68,77]
SSD	[5,43,46,60,61,63,65,70]
YOLO versions	[34,55,57,61,62,70]
RetinaNet	[7,44,63]
CenterNet	[60,69]
R-FCN ResNet-101	[60,64]
Fast R-CNN	[18,64]
VoxelNet	[16]
EfficientDet	[60]
UNet-backbone for OD	[70]
Doc-UFCN	[17]
Not mentioned	[15,37,78]
Self-assembly network	[47,67]

#### 4.4. (RQ4) Query Strategies Taxonomy

Active learning uses query strategies (namely, relying on sampling) to present the oracle with instances to annotate, mainly based on certain properties (e.g., score, distance) calculated for each instance. In the screening process of this literature review, four criteria were found in the query strategies used in the context of OD (Figure 7).



**Figure 7.** Query strategies identified in this review.

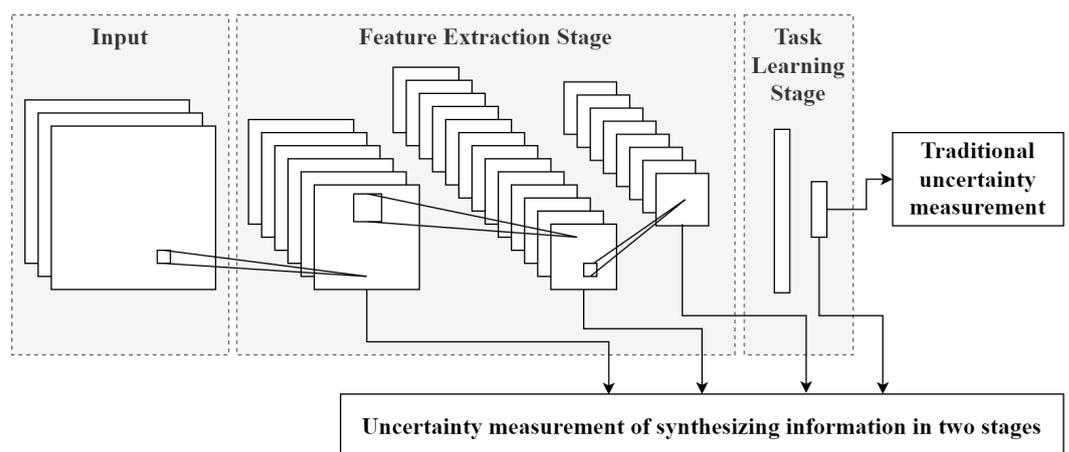
In Table 4, the articles were grouped according to the previously identified taxonomy and, also, to learning techniques (e.g., weak-supervised, semi-supervised, ensemble, Monte-Carlo dropout and Core-Set) combined with AL. Uncertainty-based strategies, in the 2D context, are the methods most explored by the technical/scientific community, followed by hybrid methods.

**Table 4.** References grouped by the AL Taxonomy. Uncertainty (U), Feature Distribution (FD), Hybrid (H), Weak-supervised (WS), Semi-supervised (SS), Ensemble (E), Monte-Carlo Dropout (MC), and Core-Set (CS).

References	U	FD	H	2D	3D	WS	SS	E	MC	CS
[5,9,14,34,43,47,48,55,60,62,63,66,69]	✓			✓						
[59,68,77]	✓			✓		✓				
[56,64]	✓			✓			✓			
[46,70]	✓			✓				✓		
[17]	✓			✓					✓	
[50,78]		✓		✓						
[18]		✓		✓		✓				
[65]		✓		✓						✓
[37]			✓	✓				✓		
[71]			✓	✓		✓	✓			
[36,44,45,57,61,67]			✓	✓						
[54]	✓			✓	✓			✓		
[53]	✓				✓			✓	✓	
[15,16]		✓			✓					

4.4.1. Informativeness-Based Methods

According to [49], informativeness measures the effectiveness of an instance on reducing the uncertainty of a model. This sampling strategy identifies unlabelled items that are near a decision boundary of the current ML model, i.e., samples that are expected to be a less obvious classification for the model. The main concept is to sample labels with the highest uncertainty, which are considered to be the most informative. As such, selected samples can belong to the subset with the lowest prediction probabilities, simultaneously ensuring the presence of useful information to improve the model. In the uncertainty approach, a naive way to define uncertainty—based on black-box methods—is to sample data according to the model’s predictions on the unlabelled data [51]. In black-box methods [46], the image uncertainty is computed from the confidence returned by the softmax layer regardless of the neural network architecture. On the other hand, the white-box methods [46] focus on the network architecture and its knowledge to compute image scores (see Figure 8).



**Figure 8.** Image adapted from [8]. Black-box methods propose the use of traditional uncertainty measurement. White-box methods use the network architecture and its knowledge to compute the image score, i.e., uses information from the feature extraction stage and the prediction stage.

Uncertainty-based methods are widely used, but they are prone to the problem of data bias [60], since the selected data are hardly representative of the full unlabelled dataset.

### Black box-based methods

Classical query strategies adapted for OD, using the softmax layer output as a confidence metric, are considered black-box methods. In this case, the confidence from all objects belonging to an image is computed at the box level. Considering that the confidence of the prediction  $i$  that belongs to the class  $j$  is  $p_i^j$  and the number of classes is  $c$ , the Least Confident (LC) method [60] selects the detected object instances whose prediction is the least confident by the model for each class (see Equation (3)). Alternatively, the Margin of Confidence (MC) method [60] considers the top two confidence rated objects of the same class  $j$  with the smallest margin (see Equation (4)). Entropy-based (E) method [46] selects the instance with the highest Uncertainty of belonging to a target class. This method is more exhaustive than LC and MC for the OD task, since it considers the uncertainty of all objects belonging to the same class (see Equation (5)). Another alternative for computing the uncertainty can be explored in [62].

$$U_j = \arg \max p_{ij \in [1, \dots, c]}^j \quad (3)$$

$$U_j = \arg \max p_{ij \in [1, \dots, c]}^j - - \arg \max p_{kj \in [1, \dots, c]/i}^j \quad (4)$$

$$U_j = \sum -p_i^j \log p_{ij \in [1, \dots, c]}^j. \quad (5)$$

Regarding the design of sampling strategies, performing them for OD has a higher complexity than for classification tasks, since the impact of the background that affects the former constitutes an imbalance factor. In practice, this aspect reflects in the way that uncertainty is calculated using OD algorithms. Firstly, multiple boxes and their confidences are computed after an OD-based image processing, followed by the application of an aggregation method (see Equations (6)–(9)) combining the individual scores of each box-label pair to calculate the score of the entire image, in compliance with the recommendations of [45].

$$U(I) = \sum_i^c U_j \quad (6)$$

$$U(I) = \sum_1^c U_j / c \quad (7)$$

$$U(I) = \min_c U_j \quad (8)$$

$$U(I) = \max_c U_j. \quad (9)$$

In [34], the authors evaluate LC sampling techniques using the YOLOv5 model according to the Federated Learning approach, more specifically, using unlabelled data available from its devices, considering sum, average, and maximum as aggregation methods. While the sum method depends on the detection count per image, the average method allows the comparison between images, and the maximum method deals with noisy detection [34,62].

Another black-box strategy is the query by committee, in which the disagreement of the predictions made from a consortium of models [37] is used to select unlabelled instances and, therefore, to try to reduce the generalisation error in further training sessions. Still on the same topic, three questions were posed in [79]:

- How to build the committee?
- How to quantify the disagreement?
- How to combine the committee member responses to obtain a robust classification?

In [62], a combination between a method of AL for OD and an incremental learning scheme were proposed, to enable continuous exploration of unlabelled datasets. The authors proposed a set of uncertainty-based AL “metrics” suitable for OD algorithms,

which are computed from a DL model trained incrementally and an approach to leveraging class imbalances during sample selection over the PASCAL VOC 2012 dataset. To select unlabelled images from a pool, they compute, for each image, a detection score based on the confidence margin, considering the whole set of detections, which are aggregated as a sum into a unique image-score (Figure 9).

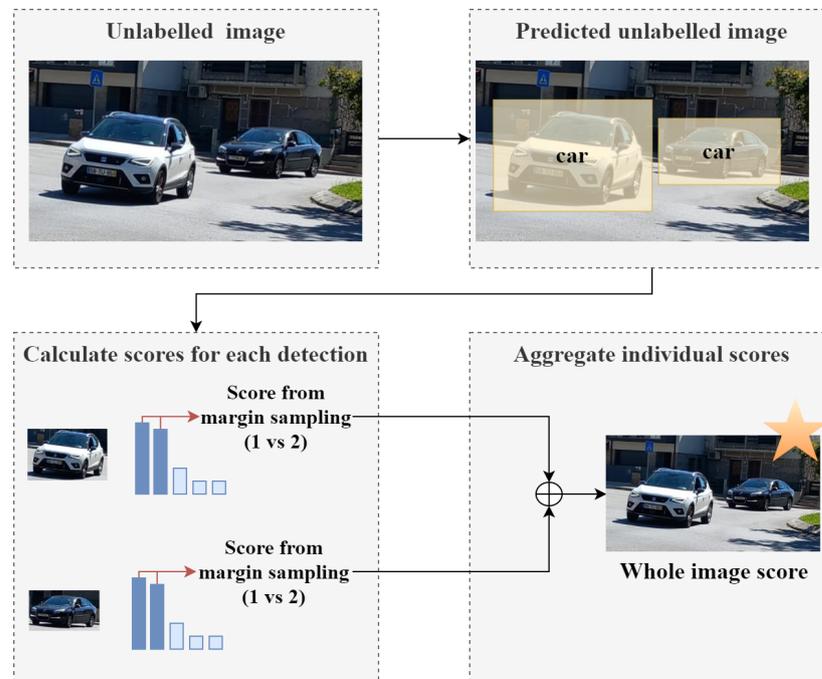


Figure 9. Margin confidence with sum aggregation function proposed in [62].

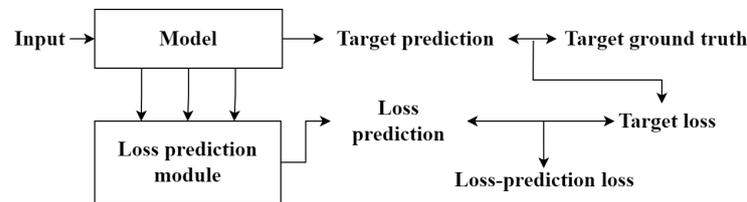
### White box-based methods

In [46], the authors proposed an AL method based on query-by-committee with a box-level scoring technique, which is calculated after the non-maximum suppression step of the Single Shot Multi-Box Detector (SSD) output. To deal with the exploration/exploitation trade-off, the authors divided the score space into bins and selected high-scored images from the top bins (exploration) and the lowest-scored images from the last bins (exploitation).

Another contribution found in the literature proposes a Multiple Instance Active Object Detection (MI-AOD) [63] to highlight informative instances in the unlabelled dataset. It explores how to use a detector (RetinaNet) trained to evaluate the Uncertainty of the unlabelled images and how the image uncertainty can be precisely estimated while filtering noisy instances using Multiple Instance Learning (MIL). More specifically, Retina-Net was modified with two discrepant instance classifiers and the prediction differential was used to learn instances uncertainty from the unlabelled set. The training focused on maximising the prediction discrepancy of the classifiers. In the second stage, MIL allows pruning images with high uncertainty. Additionally, a relationship between instance uncertainty and image uncertainty is established.

In [47], a method that aggregates pixel-level scores for each image to compute its image-level score is proposed. This score is used in a ranking process, where  $b$  (the budget) is the best-ranked image selected for annotation. The method includes a detection model capable of computing the posterior probability for each pixel in a multi-resolution approach (i.e., resizing the logits layer to obtain the same image size). At the pixel-level, the score is the sum of the difference between the entropy of mean predictions and the mean entropy of predictions for each resolution. Moreover, the image-level score is the average of max-pooled scores considering non-overlapping regions. Finally, the  $b$  images with the highest score are selected for manual annotation. In the case of video processing, the method includes a temporal smoothing of the image-level score.

Another example in this category introduces a loss prediction module attached to the SSD target model [43]. The result of this module tries to simulate the loss defined by the SSD architecture using a small configuration learned in parallel with the target model. The proposed configuration was linked with several intermediate layers, considering the model's knowledge, to predict the loss (see Figure 10).



**Figure 10.** Loss prediction module and the method to learn the loss. Image adapted from [43].

#### 4.4.2. Feature Distribution-Based Methods

In this group, it was possible to differentiate four categories: diversity-, density-, rareness-, and representativeness-based methods. According to [49], diversity-based methods measure the information overlap between instances at the feature or label dimension. It is a sampling strategy purely exploratory that aims at selecting the most dispersed samples (outliers) according to a feature distance [41,60]. The simplest diversity function is the Euclidean distance, but others can be considered in high dimensional data scenarios, such as Gaussian or Cosine functions, which are radial and angular approaches [37].

Moreover, this class of methods (feature distribution-based) anchor on the key-concept of forcing the selection of instances where there is more density (a core subset [8]), relying on feature space analysis. This avoids the selection of outliers from the unlabelled dataset and focuses the selection on a relatively dense region [37].

In the opposite direction, rareness-based method search instances that do not frequently appear in the training dataset [15,50]. Ref. [15] proposed the selection of rare instances based on density estimations from the 3D data to improve long-tail performance and, in [50], recognition of unknown instances/classes absent in the training dataset was addressed.

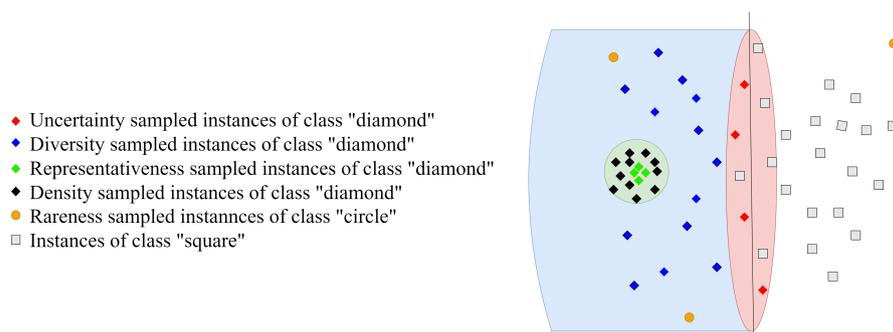
Another alternative is the representativeness-based approach, which seeks instances in the centre of the cluster of instances as in [67]. In this case, the sampling strategy quantifies the capability of an instance to represent the overall input patterns of unlabelled data [49].

Figure 11 depicts a graphical example to explain the previously mentioned concepts. The red region is known as the area where the detector has more uncertainty in selecting an instance, while the union of the blue and green region represents the zone where a higher diversity of instances can be found.

A few articles [16,78] applying a diversity sampling strategy exclusively were identified. In [78], the authors applied a two-stream active clustering method intending to deal with bias in small datasets, i.e., instances far from representative of the whole dataset. In [16], the focus was the diversity over the spatial and temporal space to achieve increased performances. Moreover, Ref. [18] focuses mainly on diversity sampling but mixed with weakly-supervised OD techniques.

#### 4.4.3. Hybrid Methods

The hybrid sampling strategy or simultaneous approaches [36,37,44,45,57,61,67] combines two or more of the previously mentioned sampling strategies. Methods in this category aim at reaching a balance between multiple strategies [8]. For example, in the combination of uncertainty- and diversity-based methods, diversity can prevent instances from overlapping or becoming redundant in queries producing better sampling results [36,67], or can overcome noisy labels [61].



**Figure 11.** Eligible samples to be selected according to the sampling method. Uncertainty (red instances), Diversity (blue instances), Representativeness (green instances), Density (black instances inside the green circle), and Rareness (orange instances) and sampling examples. In this example, three classes (circle, square and diamond) are a representation of a 2D feature space.

In a few other works [61,67], the combination of AL and Semi-Supervised Learning (SSL) was addressed. Following a similar topic, Ref. [67] presented a refinement based in batch-mode, which is known as Active Semi-Supervised Learning (ASSL). The proposed method consists of firstly selecting a random set of samples from a pool and, then, adopting an efficient and flexible collaborative sampling strategy. The method integrates informative (least confidence and margin confidence sampling) and diversity (k-means) criteria from the concept of AL, as well as the confidence criterion from SSL. The former methods of SSL enabled the selection of the samples with the highest confidence (highest agreement among classifiers), considering that if samples are in the same cluster, the probability that they belong to the same class is high. The authors propose a detection model with a softmax classifier to compute the confidence scores over  $K$  classes. This allows the selection of samples with the highest Uncertainty with a probability close to 0.5 at the decision boundary, defining an upper and lower range.

In [36], the authors train object detectors to compare some query methods from the unlabelled dataset, considering two stages: scoring and sampling. The informativeness measured in the scoring stage results in one score per image. In the sampling stage, the computed score (uncertainty measure) was used to select a batch of images without ignoring sample diversity. The authors empirically evaluated several scoring functions (entropy, mutual information, gradient of the output layer and bounding boxes with confidence) and two score aggregation methods (maximum and average). To evaluate diversity, the method converts unlabelled samples into feature vectors and computes similarity distance (i.e., Euclidean and Cosine Distances). K-Means, Core-Set, and Sparse Modelling were used as sampling strategies to avoid selecting highly similar instances. The result of the extensive evaluation was that mutual information with a maximum seems to offer the best trade-off between mAP and samples count.

In [44], the authors focus on fine-grained sampling for AL. They proposed a method that generated queries requesting bounding box labels, selecting the most informative subset rather than entire images on a pool-based scenario. The AL framework uses Random Sampling, Mean Classification Uncertainty (i.e., choose images with high mean classification uncertainty for querying), and Coreset Greedy Diversity Sampling to solve classification problems based on image-level feature representations, demonstrating savings in labelling efforts, while ensuring a suitable OD performance. The solution was tested on two standard datasets: PASCAL VOC-12 and MS-COCO.

In [37], a Batch Active Learning (BAL) method was developed, proposing a query-by-committee (QBC) for uncertainty sampling (i.e., selecting instances closer to the decision boundary) using the random forest distance function. The method computes (i) the entropy, (ii) the disagreement margin between the class with the highest estimated confidence and the second highest of an instance, and (iii) the difference in the two classes with the largest vote proportions for each class. The method includes diversity and density sampling to

ensure that the selection of instances is placed on feature space areas with high density, thus avoiding the selection of outliers from the unlabelled dataset. They tested three distance measures: Euclidean distance, Gaussian kernel distance, cosine for angular diversity and random forest dissimilarity. Each batch has a size of five elements, and the method uses 20 rounds of queries.

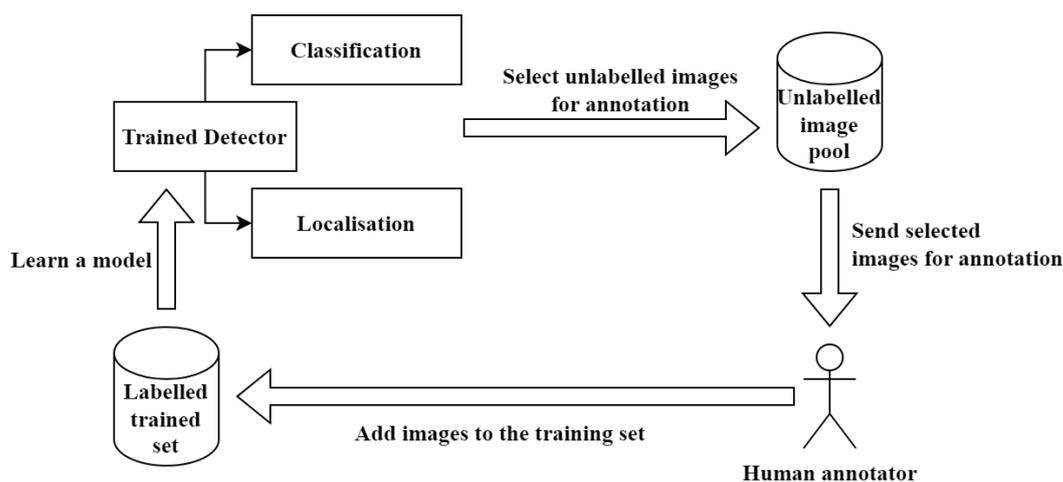
#### 4.4.4. Random Method

The random sampling strategy is mostly used for comparison purposes, grounded on to the principle that AL-based methods have the potential to perform better (e.g., [5,48,52,60]). This method uses a random function with uniform distribution over the interval  $[0, 1]$ , and it consists in selecting an image from the pool uniformly random.

### 4.5. (RQ5) Other Techniques

#### 4.5.1. Combining Localisation and Classification

There are works demonstrating that an object detector's localisation and classification output should be considered in AL. For example, two metrics were proposed in [5]. The former, localisation tightness, uses the intersection over union (IOU) ratio between the region proposal and the final prediction to select images with inconsistency between the classification and the localisation. The second, localisation stability, focuses on the variation of predicted localisation behind images corrupted by noise (Figure 12). It measures the predicted bounding boxes variation when noise is added to the original image.



**Figure 12.** Overview of the AL algorithm proposed in [5].

In subsequent works [48], the uncertainty was efficiently estimated, making a single forward pass in a single model and considering both the classification and detection information (Figure 13).

Additionally, instead of a single value prediction for each convolutional neural network output, the authors designed an output layer to predict the probability distribution of the whole image and introduce a new loss function in the training of the Gaussian mixture model (GMM), increasing the overall performance (its code is available at <https://github.com/NVlabs/AL-MDN> accessed on 20 July 2023).

Document layout detection is addressed using OD in [14]. In this case, AL selects regions considered the most ambiguous by the object detector. The perturbation-based scoring step considers object position and category to evaluate the prediction ambiguities of the DL model. Additionally, the authors proposed a semi-automatic correction algorithm, which rectifies out of selection predictions, with minor supervision, based on prior knowledge of layout structures, i.e., identifying duplicated and false negative predictions.

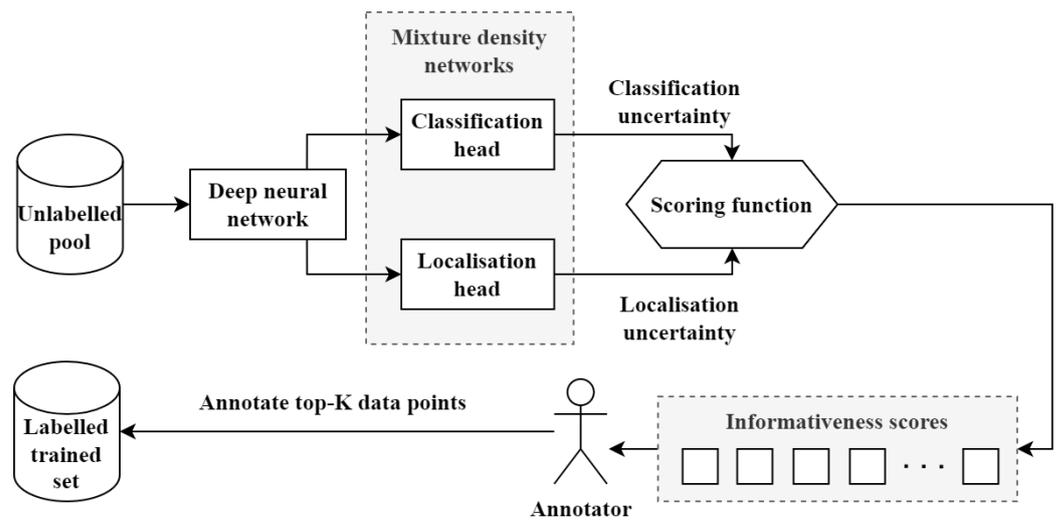


Figure 13. Aggregation of the localisation and classification approach proposed in [48].

#### 4.5.2. Monte-Carlo Dropout-Based Approaches

The Monte-Carlo dropout technique utilises multiple passes with different dropout masks, and the average of its softmax outputs can be used as the output of the classes. Monte-Carlo dropout, or stochastic regularisation techniques, were used in [8] to implement variational inference and calculate the posterior distribution of network predictions. Also, in [17], the authors proposed two confidence estimators, in a document analysis application: Dropout Average Precision and Dropout Object Variance, for the OD task.

#### 4.5.3. Weakly Supervised-Based Approaches

AL and weak supervision can be combined to reduce costs in the annotation process, as investigated in [77], which proposed a framework that is depicted by Figure 14, and which includes a module that decides when to switch from a weak (object centre annotation) to a strong (bounding box annotation) supervision method, complying with the requirements of the training process. The maximum margin, least confident, and average entropy are used as uncertainty query strategies. In [68], a modified version of the previous framework was proposed, introducing a weakly labelled pool into the AL workflow. Recently, in [18], the authors proposed the acquisition function “box-in-box” (BiB) to detect failure cases of the weakly-supervised detector. The method aims to correct mistaken inferences of the weakly-supervised detector.

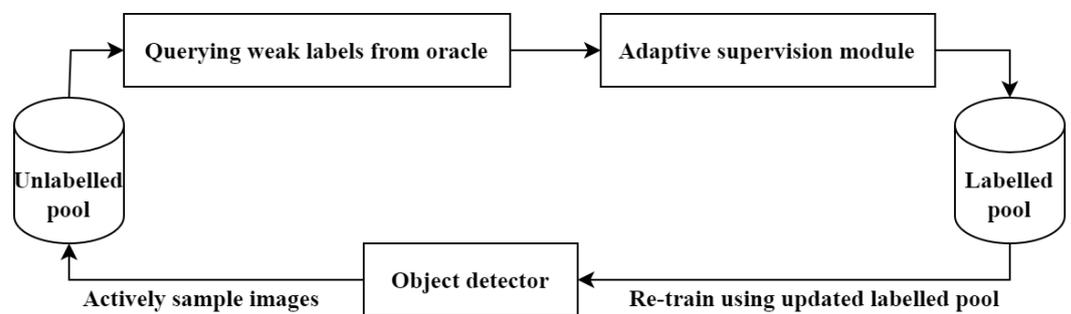


Figure 14. Proposed combination between AL and weak supervision from [77].

Another work based on weak supervision proposes annotating images with double predictions, when their IOU is higher than 0.5. The object instance with the higher Uncertainty is the pair with the higher IOU. For online OD [71], two pool- and stream-based scenarios were evaluated using a hybrid sampling strategy. Generally, pool-based is the most used, and stream-based is more suitable for online OD in scenarios involving robotics.

#### 4.5.4. Ensemble-Based Approaches

Several works mentioned the inclusion of ensembles usage as part of AL heuristics [37,53,54]. In [70], the importance of training in a subset of the data is demonstrated, resulting in better performances. The authors evaluate uncertainty strategies such as entropy, mutual information, variation ratios, and error count in an ensemble-based approach.

#### 4.5.5. Other Proposals

A method for AL applied to OD in video, using temporal coherence, was presented in [9], which proposes an acquisition function called Temporal Coherence that resorts to an object tracker between consecutive frames.

For remote sensing applications, [69] propose a new weighted classification regression (WCR) uncertainty sampling method, which consists of selecting images according to its size, using regression and adapting the classification uncertainty according to its informativeness. Also, the proposed method deals with class imbalance by considering the objects count by class, thus guaranteeing a better performance on classes with fewer instances. The least confident metric ( $U_c$ ) for classification is calculated using the average number of objects for each class in one image and the class ratio. For regression, the Uncertainty ( $U_r$ ) is calculated, taking into consideration the size of the bounding box of the detected object. Finally, the image uncertainty is calculated as  $U_s = \sum(U_c \cdot U_r)$ .

To deal with the problem of data bias on uncertainty-based AL methods, the Weighting filter (W-filter), custom-made for OD, is proposed in [60]. This method (see Figure 15) filters similar images from the training dataset in the first epoch guaranteeing better performance and stability in the learning process. Also, some classical uncertainty methods are adapted for the OD task. W-filter focuses on high-frequency analysis (i.e., edges of objects in images) and the removal of similar ones. After the first epoch, the method reduces the bias in the training dataset by removing images with high similarity in the frequency domain.

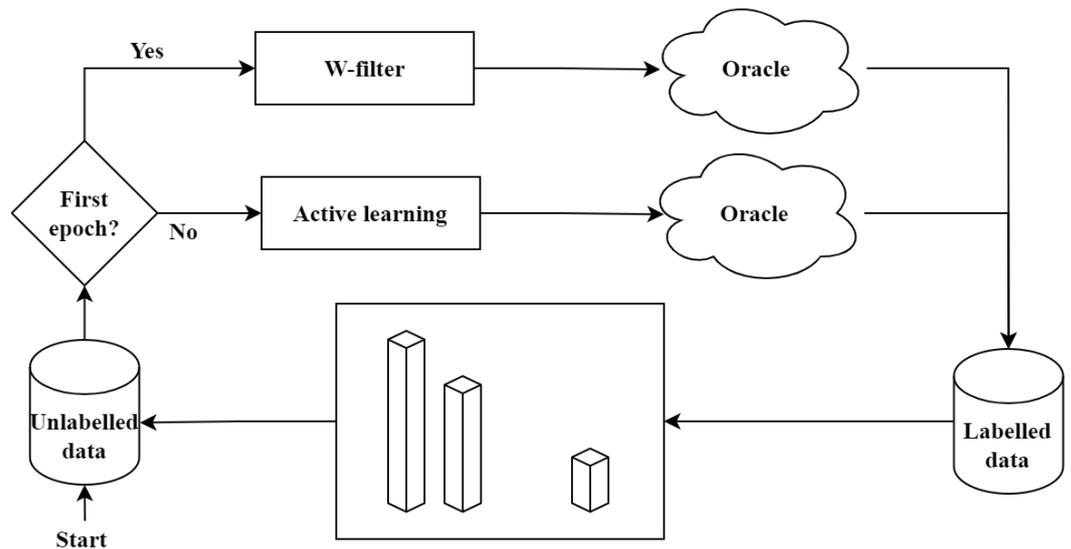


Figure 15. Procedure comparison between W-filter and other methods in the first epoch, retrieved from [60].

In [55], the authors proposed an AL uncertainty-based framework, which includes a method to measure the sampling scale of the different defects in a defect detection context. Another way to select unlabelled images is according to the detector’s confidence [56,64]. Instances with high-confidence prediction can be automatically pseudo-labelled via a self-learning process. If instances have low confidence, the images are manually annotated after being selected by the AL strategy. The authors consider an image annotation between different oracles with ambiguities to have a noisy sample and proposed a suitable method to deal with it via the switchable criteria. For instance, in [64], the authors proposed

a Self-Supervised Sample Mining approach for OD, while in [56], the active sampling mining (ASM) framework was chosen instead. Also, a Contextual Diversity approach was proposed in [65], presenting two AL frameworks—the first based on a core-set strategy and the second based on reinforcement learning.

#### 4.6. (RQ6) SWOT Analysis

Finally, this section presents a SWOT analysis of the AL techniques for OD task in computer vision, summarised as follows:

- Strengths
  - Efficient way to annotate unlabelled samples;
  - Reduction of the overlapping between labelled instances;
  - Control over the available resources in the annotation process by creating a budget planning by iteration.
- Weakness
  - The possibility of selecting an inappropriate aggregation method for global image scoring;
  - The exclusive use of uncertainty-based acquisition function may cause bias in the dataset;
  - The implementation of an AL strategy is a hard task, with high computational cost, involving several iterations of model training and sample selection.
- Opportunities
  - To define strategies for improving ML models over time, through transfer learning of previous training sessions;
  - To improve the annotation interfaces, e.g., with the inclusion of voice commands;
  - To evaluate, in real-time, the specialist's visual attention during the annotation process;
  - To build thematic/contextual gamification strategies that could make turn the annotation process—that is usually dull and requires several sessions—in a more appealing task;
  - To explore Explainable Artificial Intelligence (XAI) approaches to increase trust in the approaches;
- Threats
  - Legal and ethical obligations for data handling (e.g., anonymisation, user authorisation) in an era wherein AI regulation is still under construction;
  - Setback challenges in collecting data from various research centres or organisations, not only in terms of access but also regarding format standardisation;

To summarise, there is a wide range of strategies that have been proposed by the scientific community to combine AL and OD towards AI models optimisation, and which will be discussed in the next section.

## 5. Discussion

In computer vision applications, more specifically the ones focusing on object detection, annotation is usually a burdensome and time consuming process, due to the need to outline the bounding box and assign the corresponding label. The application of AL techniques offers advantages in the annotation process of large datasets—such as effort rationing/reduction combined with approaches that can speed up the production of accurate models with an agenda for continuous improvement in mind, involving annotators with high expertise. Based on the literature collected for this review, and considering the field of application, one can infer that AL techniques can be particularly relevant in areas such as healthcare, wherein a more intensive exploration activity seems to taking place.

### 5.1. (RQ1) Geographical Distribution

The analysis results showed a concentration of countries in the northern hemisphere, more specifically, in North America (United States of America), Europe (Germany), and Asia (China). Most of the countries in this group are highly developed, which leaves them in a position to invest more resources in this field of research, which is particularly demanding because of the following main reasons:

- To have AI in a continuous and sustainable production environment, it requires access to processing capabilities (i.e., hardware and/or services) that are still costly, despite the increasing affordability of consumer-grade AI-capable graphical cards for the general public, which has been empowering indie AI prototyping over time;
- It demands highly qualified human resources that are usually associated with significant efforts in terms of salary (the average paycheque of a computer vision engineer working in the United States and Germany is around, respectively, USD 165 K\$ (<https://www.talent.com/salary?job=computer+vision+engineer> accessed in 21 July 2023) and EUR 69 K\$ ([https://www.glassdoor.com/Salaries/germany-computer-vision-engineer-salary-SRCH\\_IL.0,7\\_IN96\\_KO8,32.htm](https://www.glassdoor.com/Salaries/germany-computer-vision-engineer-salary-SRCH_IL.0,7_IN96_KO8,32.htm) accessed in 21 July 2023), in 2023).

On the other hand, the most advanced companies and research centres in the field of AI are also interested in continuing research around AL and exploring its potential benefits further [14,36,50,59,78].

### 5.2. (RQ2) Application Domains

Generally, in an object detection task, multiple target objects may appear in a single image. Therefore, sampling strategies should deal with each object's information (bounding box, class, and confidence)—which contrasts with sampling strategies for classification that only need to address confidence. If an appropriate aggregation method is not considered (e.g., maximum focus on outliers, average weight of all objects equally, sum is dependent on the number of objects), the sampling may perform poorly as a relevant contributor to smoothing annotation efforts, as well as to promote the construction of increasingly robust models. Multiple alternatives include the aggregation of information from multiple sources such as: pixel-level, box-level [45] or image-level [44,47], which have been useful for both uncertainty-based and diversity-based methods, and more specifically to guarantee the reduction of labelled instances overlapping, while optimising data contributions aiming at models with increasingly improved performance.

AL strategies can be designed with a focus on issues that reduce the performance of object detection methods. Examples include class-imbalance, poor diversity, rareness, or outlier instances. Feature distribution-based methods compensate for the lack of exploration of uncertainty-based methods. Clustering algorithms, for example, have been widely applied to optimise criteria such as diversity [49], representativeness [67], rarity [15,50], and density [16,78]. However, the proposed methods do not always fully explore their potential by focusing on an exclusive criteria—among the previously referenced ones—when considering several complementary sampling strategies simultaneously is possible [36,37,44,45,57,61,67].

Moreover, object detection applications are concerned with the location and classification of objects in an image. The classifiers used for uncertainty estimation ignore the location information [48]. Therefore, a good instance selection method for object detection must consider the convolutional neural network's ability to classify and locate, even in conditions wherein it is challenging to determine object-location informativeness [5]. The acquisition function should be able to measure these couple of features out of a neural network to produce reliable instance rankings. The possibility of unsuccessfully combining two or more sampling strategies should be considered a weakness. Applying one type of sampling strategy (e.g., uncertainty) tends to select similar images [6], resulting in an insufficient approach [37].

The application of an AL strategy may be regulated for several reasons. As for any validation process involving humans, the oracle verification process must insure the inclusion of intra- and inter-group variability. Also, it is important to avoid situations wherein annotators spend or exceed the expected budget before checking the entire stream-based selective sampling scenario. With these issues in mind, AL approaches allow for not only the reduction of human effort, optimising the annotation process, but also for increasing the ability to manage the aspects regarding the variability of the verification process, as well as the monitoring of the available annotation budget along the operationalisation steps [6].

Another conclusion drawn from this work, which can be also observed in Table 2, relates to the predominance of data dimension addressed in AL, which is mostly 2D. This observation may be corroborated by the following factor: data with other spectral and spatial dimensionalities (3D, multi/hyperspectral, etc.) are usually more demanding and expensive to acquire—this is even more aggravated if precision is required—since costly devices, preparation, and/or pre-processing steps are often involved, as well. Having said that, great and interesting opportunities to keep exploring the application of AL and OD can be glimpsed and pursued.

### 5.3. (RQ3) Deep Neural Network Architectures

The OD approaches most explored seem to be those that emerged from the beginning (year 2015) such as YOLO versions [23–26], SSD [27], and Faster R-CNN [33].

Among the OD approaches utilisation, there is, at least, an apparent tendency/preference for one-stage detectors, instead of two-stage. Since the former group, without the region proposal layer, usually involves shorter processing times, the use of the algorithms in its umbrella can be considered as more suitable for real-time scenarios. Also, derived architectures are simpler and have fewer hyper-parameters and parameters to manage and fine-tune, which facilitates the implementation of sampling strategies using intermediate layers.

### 5.4. (RQ4) Query Strategies Taxonomy

It was noted that there is an apparent balance between the two uncertainty categories, i.e., black- and white-based methods. Initially, the scientific community has tried to extend existing knowledge in AL to the problem of OD with acceptable results.

Other researchers argue that white-based methods will allow us to achieve top results in the performance of deep neural network models comparatively to black-box approaches, where multiple query samples can be adapted for OD tasks, such as GANs-inspired techniques [43,80] or loss prediction in the training stage [43]. Moreover, uncertainty- and feature distribution-based methods can focus on visual data structure. These approaches proved to be very useful but other types of descriptors, such as morphological ones, have not been sufficiently explored (e.g., area, perimeter, or circularity).

### 5.5. (RQ5) Combining with Other Learning Techniques

It is reported that MC-dropout and Deep Ensembles, when used to estimate Uncertainty, show slightly better results compared to using a single softmax output [53]. It is also reported that the Ensemble method outperforms the MC-dropout and the Core-Set in multiple applications [39]. One advantage of ensembles is that they do not need to alter the deep neural network structure when compared to the MC-dropout method. Additional studies related to scaling the number of models in ensemble-based methods [70] are necessary for the context of AL. Notwithstanding, ensemble- and MC-dropout-based seem to be open research lines in the OD field.

### 5.6. (RQ6) Research Opportunities

Public tools, such as benchmark frameworks, would allow the research community to validate its work with an impartial comparison of the published AL strategies in public datasets. However, most available tools are commercial and seem to mainly

focus on uncertainty-based methods, which are, in theory, less demanding in terms of implementation.

Therefore, pursuing the goal of obtain higher performance in instance sampling is still a valid challenge for AL researching community, which keeps designing novel methods to find instances more informative or representative for OD applications, to increase the overall performance known for the currently proposed methods. However, other criteria, such as usability and interpretability, are ignored by focusing on performance.

In the case of interpretability [81], it would be interesting to develop AL methods that consider this criterion as it would increase the trust of the end users in the results obtained by the deep learning-based system. For usability, the emergence of methodologies for multi-oracle annotation and validation to deal with the inter- and intra- observer variability should be considered. Also, it would be interesting to include voice- and image-based technologies for assessing motivation or voice control to speed-up the annotation process. Gamification would be another approach to involving annotators in a more interactive and competitive environment, that could, simultaneously, reduce the dullness of the task.

Additionally, other research opportunities, such as investing in one-shot AL frameworks [41,82], fusing weak supervision techniques [68,77], and classifiers ensembles [54] remain an open exploration path. Furthermore, there are only a few studies comparing continuous AL memory-based training to training from scratch, which could be interesting for certain applications focusing on training times reduction.

When developing and applying AL, there are also challenges to consider and address. For each iteration of this process, the workflow needs more interaction with experts and more computing resources [41]. White methods, based on multi-models (like classifier assemblies which increase the number of parameters [48]), have a high computing cost, while black methods may not need many parameters. Meanwhile, keeping track of the detection model's performance seems to still be a challenge that can be addressed with entirely novel management strategies or by producing adaptations and combinations of existing ones.

Moreover, collecting data for annotation from various institutions remains a barrier to be tackled [83]. This review shows that Europe is one of the main regions where the researchers are working with greater intensity on this subject. However, law-related challenges still need to be worked out, to find a more suitable balance between data access, consent to use personal information, respective anonymisation strategies, etc.

Despite the challenges in implementing AL for OD techniques, its numerous strengths make it a compelling field of research. AL has demonstrated its effectiveness for the annotation of unlabelled instances, allowing researchers to achieve comparable performance with significantly less annotated data and training time compared to using the entire dataset. By obtaining a dataset with more representative instances, AL also reduces the inclusion of redundant samples already present in the training dataset.

## 6. Conclusions

The primary objective of this work was to present a collection of trends associated with the research and development of AL methods, with a specific focus on OD. This topic has been steadily gaining attention from both the scientific community and professionals across diverse fields, spanning from medicine to industry. The increasing interest is evident from the growing number of publications highlighted at the beginning of this paper.

In this work, several viewpoints related to AL for OD were addressed, including the geographical distribution of research, prominent application domains, frequently used datasets, DL-based architectures, as well as a comprehensive inventory of recent AL methods and algorithms designed for iteratively sampling instances from non-annotated datasets.

Through the review of the literature, it became evident that AL applied to OD research tasks is predominantly conducted in countries such as China, the United States of America, Germany, India, South Korea, Italy, France, Spain, Japan, Canada, the Netherlands, and Singapore. As for datasets, public ones like PASCAL, MSCOCO, and KITTI have been ex-

tensively used in general studies related to artificial intelligence, while healthcare solutions, surface defect detection, pose estimation, and document analysis remain lesser-addressed domains. Regarding the methods, it seems that the ones based on Ensembles or Monte-Carlo-dropout have shown the best results, overall. Additionally, it was observed that AL strategies that achieve a proper balance between uncertainty and diversity in OD tasks can contribute to the development of more robust methods.

As it stands, the community's primary focus has been on the efficiency and relevance of AL solutions in various real-world applications, with less emphasis on interpretability—for example, by extending XAI capabilities—or multi-oracle annotation and validation; therefore, sharpening the knowledge on those under-explored approaches may be a plausible path for future research. Furthermore, there are other opportunities to continue exploring AL applied to OD, specifically in terms of monitoring annotator performance. This extends beyond biophysical insights and includes prompt-based answer analysis for assessing factors like operator focus, attention, and fatigue. Additionally, there is potential for developing more suitable interfaces and interaction strategies, including the possibility of incorporating gamification techniques.

**Author Contributions:** Funding acquisition, T.A.; Investigation, D.G.; Methodology, D.G., A.C. and L.G.M.; Supervision, A.C. and L.G.M.; Validation, A.C., L.G.M. and T.A.; Writing—original draft, D.G. and J.C.; Writing—review & editing, D.G., T.A., J.C., R.J., A.C. and L.G.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the RRP—Recovery and Resilience Plan and the European NextGeneration EU Funds, within the scope of the Mobilizing Agendas for Business Innovation, under reference C644937233-00000047 and by the Vine&Wine Portugal Project, co-financed by the RRP—Recovery and Resilience Plan and the European NextGeneration EU Funds, within the scope of the Mobilizing Agendas for Reindustrialization, under reference C644866286-00000011.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Authors express their sincere gratitude to Rui Almeida, Lina Carvalho and Vitor Manuel Leitão de Sousa from the Faculty of Medicine of the University of Coimbra, who actively participated in all discussions of this research, enriching the depth and breadth of our study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AL	Active Learning
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
OD	Object Detection
RQ	Research Question
SWOT	Strengths, Weaknesses, Opportunities, and Threats
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
mAP	Mean Average Precision
WSI	Whole Slide Image
SSL	Semi-Supervised Learnings
ASSL	Active Semi-Supervised Learning
FPN	Feature Pyramid Network
PCA	Principal Component Analysis
DADA	Diversity-Aware Data Acquisition
MI-AOD	Multiple Instance Active Object Detection
MIL	Multiple Instance Learning

LC	Least Confidence
MC	Margin of Confidence
GMM	Gaussian Mixture Model
IOU	Intersection over union
YOLO	You Look Only Once
BAL	Batch Active Learning
WCR	Weighted Classification Regression
RoI	Region of Interest
ASM	Active Sampling Mining
BMDAL	Batch Mode Deep Active Learning
WCR	Weighted Classification Regression
XAI	Explainable Artificial Intelligence

## Appendix A. Location of the Researchers by Geographical Regions

Table A1. Asia.

Country	Research Groups/Institutions/References
China	University of Shanghai for Science and Technology [60] Harbin Institute of Technology, Harbin [69] Beijing Orient Institute of Measurement and Test, Beijing [69] Shanghai Normal University [60] Shanghai Jiao Tong University [60] Fudan university [41] NVIDIA [36] Shanghai Key Laboratory [41] College of Computer Science and Technology, Zhejiang University [45] Binhai Industrial Technology Research Institute of Zhejiang University, Tianjin [45] School of Information Science and Technology, Nantong University, Nantong [45] University of Chinese Academy of Sciences, Beijing [63] Noah's Ark Lab, Huawei Technologies, Shenzhen [63] Tsinghua University, Beijing [63] Nanjing University of Aeronautics and Astronautics, Nanjing [57] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing [57] The State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing [57] Megvii Research Nanjing, Megvii Technology, Nanjing [57] The State Key Lab of Precision Measuring Technology and Instruments, Tianjin University, Tianjin [55] School of Data and Computer Science, Sun Yat-sen University, Guangzhou [49,64] SenseTime Research [64] The Hong Kong Polytechnic University, Hong Kong [49,64] South China University of Technology [16] Chengdu University of Information Technology [16]
India	Indian Institute of Technology, Hyderabad [44], [68,77] Indian Institute of Technology Kanpur [46,66] Amazon Core Machine Learning [66] IIIT-Delhi [65] Flixstock Inc. [65] Indian Institute of Technology Delhi [65]
Japan	The University of Tokyo [68,77]
Singapore	I2R, A-STAR [16]

**Table A1.** *Cont.*

Country	Research Groups/Institutions/References
South Korea	Seul National University [48] Computer and Information Engineering Department, Inha University [61,67] Lunit Inc., Seoul [43] KAIST, Daejeon [43] Korea University [59] Hankik University of Foreign Studies [59]

**Table A2.** North America.

Country	Research Groups/Institutions/References
United State of America	Mitsubishi Electric Research Laboratories [5] University of California, Santa Barbara [5] Department of Industrial Engineering, Arizona State University [37] Department of Industrial & Manufacturing Engineering, Pennsylvania State University [37] Department of Biomedical Informatics, Arizona State University [37] NVIDIA [36,70] MIT [78] Google [78] University of Massachusetts, Boston [78] Broad Institute [78] Comcast Research [78] New York University [78] Harvard University [14,78] Waymo LLC., Mountain View CA [15] University of California, Los Angeles [58] University of Kansas Medical Center [58] UCLA David Geen School of Medicine [58] Baylor College of Medicine [58] Inria and DI/ENS (ENS-PSL, CNRS, Inria) [18] Valeo.ai [18] Center for Data Science, New York University [18] University of Texas at Dallas [50] Microsoft Research [50] Center for Brains, Minds and Machines, MIT, Cambridge, MA [59]
Canada	University of Waterloo [14]

**Table A3.** Europe.

Country	Research Groups/Institutions/References
France	Preligens [6] TEKLIA, Paris [17] LITIS, Normandie University, Rowen [17]
Germany	Technical University of Munich [48,54] BMW Autonomous Driving Campus, Unterschleissheim, [54] NVIDIA [36,48] Max Planck Institute for Intelligent Systems, Tübingen [36] Computer Vision Group, Friedrich Schiller University Jena, [62] Michael Stifel Center Jena [62] Robert Bosch GmbH, Driver Assistance Systems and Automated Driving, Renningen [53] School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology [53] Stockholm, Sweden, Control and Microtechnology, Ulm University, Ulm, [53] Institute of Robotics and Mechatronics, German Aerospace Center (DLR) [7] Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT) [7] Robotics and Mechatronics Laboratory, University of Twente (UT) [7] Chair of Computer Vision and Artificial Intelligence, Technical University of Munich (TUM) [7]

Table A3. Cont.

Country	Research Groups/Institutions/References
Italy	Humanoid Sensing and Perception, Istituto Italiano di Tecnologia, Genoa [71] Laboratory for Computational and Statistical Learning, IIT, Genoa [71] MaLGA & DIBRIS, Università degli Studi di Genova, Genoa [71] iCub Tech, Istituto Italiano di Tecnologia, Genoa [71]
Spain	Computer Vision Center (CVC), Univ. Autònoma de Barcelona (UAB) [9,47] Computer Science Dpt., Univ. Autònoma de Barcelona (UAB) [47]
Netherlands	University of Twente [34]

## References

1. Khandelwal, Y.; Bhargava, R. Spam Filtering Using AI. *Artif. Intell. Data Min. Approaches Secur. Fram.* **2021**, 87–99. [CrossRef]
2. Gonzalez, D.G.; Carias, J.; Castilla, Y.C.; Rodrigues, J.; Adão, T.; Jesus, R.; Magalhães, L.G.M.; de Sousa, V.M.L.; Carvalho, L.; Almeida, R.; et al. Evaluating Rotation Invariant Strategies for Mitosis Detection Through YOLO Algorithms. In Proceedings of the Wireless Mobile Communication and Healthcare, Virtual Event, 30 November–2 December 2022; Cunha, A., Garcia, N.M., Marx Gómez, J., Pereira, S., Eds.; Springer: Cham, Switzerland, 2023; pp. 24–33.
3. Adão, T.; Gonzalez, D.; Castilla, Y.C.; Pérez, J.; Shahrabadi, S.; Sousa, N.; Guevara, M.; Magalhães, L.G. Using deep learning to detect the presence/absence of defects on leather: On the way to build an industry-driven approach. *J. Phys. Conf. Ser.* **2022**, 2224, 012009. [CrossRef]
4. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, 8, 1–74. [CrossRef]
5. Kao, C.; Lee, T.; Sen, P.; Liu, M. Localization-Aware Active Learning for Object Detection. In Proceedings of the Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018.
6. Goupilleau, A.; Ceillier, T.; Corbineau, M. Active learning for object detection in high-resolution satellite images. *arXiv* **2021**, arXiv:2101.02480.
7. Lee, J.; Balachandran, R.; Kondak, K.; Coelho, A.; De Stefano, M.; Humt, M.; Feng, J.; Asfour, T.; Triebel, R. Virtual Reality via Object Pose Estimation and Active Learning: Realizing Telepresence Robots with Aerial Manipulation Capabilities. *arXiv* **2022**, arXiv:2210.09678.
8. Ren, P.; Xiao, Y.; Chang, X.; Huang, P.; Li, Z.; Chen, X.; Wang, X. A Survey of Deep Active Learning. *ACM Comput. Surv.* **2020**, 54, 1–40. [CrossRef]
9. Bengar, J.Z.; Gonzalez-Garcia, A.; Villalonga, G.; Raducanu, B.; Aghdam, H.H.; Mozerov, M.; López, A.M.; van de Weijer, J. Temporal Coherence for Active Learning in Videos. *arXiv* **2019**, arXiv:1908.11757.
10. Wu, X.; Xiao, L.; Sun, Y.; Zhang, J.; Ma, T.; He, L. A survey of human-in-the-loop for machine learning. *Future Gener. Comput. Syst.* **2022**, 135, 364–381. [CrossRef]
11. Budd, S.; Robinson, E.C.; Kainz, B. A Survey on Active Learning and Human-in-the-Loop Deep Learning for Medical Image Analysis. *arXiv* **2019**, arXiv:1910.02923.
12. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R.; et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion* **2021**, 76, 243–297. [CrossRef]
13. Takezoe, R.; et al. Deep Active Learning for Computer Vision: Past and Future. *arXiv* **2022**. [CrossRef]
14. Shen, Z.; Zhao, J.; Dell, M.; Yu, Y.; Li, W. OLALA: Object-Level Active Learning Based Layout Annotation. *arXiv* **2020**, arXiv:2010.01762.
15. Jiang, C.M.; Najibi, M.; Qi, C.R.; Zhou, Y.; Anguelov, D. Improving the Intra-class Long-tail in 3D Detection via Rare Example Mining. *arXiv* **2022**, arXiv:cs.CV/2210.08375.
16. Liang, Z.; Xu, X.; Deng, S.; Cai, L.; Jiang, T.; Jia, K. Exploring Diversity-based Active Learning for 3D Object Detection in Autonomous Driving. *arXiv* **2022**, arXiv:2205.07708.
17. Boillet, M.; Kermorvant, C.; Paquet, T. Confidence Estimation for Object Detection in Document Images. *arXiv* **2022**. [CrossRef]
18. Vo, H.V.; Siméoni, O.; Gidaris, S.; Bursuc, A.; Pérez, P.; Ponce, J. Active Learning Strategies for Weakly-supervised Object Detection. *arXiv* **2022**. [CrossRef]
19. Salvi, M.; Acharya, U.R.; Molinari, F.; Meiburger, K.M. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Comput. Biol. Med.* **2021**, 128, 104129. [CrossRef]
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, 60, 84–90. [CrossRef]
21. Zou, Z.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *arXiv* **2019**, arXiv:1905.05055.

22. Sharma, V.; Mir, R.N. A comprehensive and systematic look up into deep learning based object detection techniques: A review. *Comput. Sci. Rev.* **2020**, *38*, 100301. [[CrossRef](#)]
23. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
24. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. *arXiv* **2016**, arXiv:1612.08242.
25. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
26. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
27. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Springer International Publishing: Amsterdam, The Netherlands, 2016; pp. 21–37. [[CrossRef](#)]
28. Lin, T.Y.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002.
29. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-Shot Refinement Neural Network for Object Detection. *arXiv* **2017**, arXiv:1711.06897.
30. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. *arXiv* **2019**, arXiv:1911.09070.
31. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [[CrossRef](#)]
32. Girshick, R.B. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
33. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [[CrossRef](#)]
34. van Bommel, J.R. Active Learning during Federated Learning for Object Detection. Bachelor’s Thesis, University of Twente, Enschede, The Netherlands, 2021.
35. Han, J.; Kang, S. Active learning with missing values considering imputation uncertainty. *Knowl.-Based Syst.* **2021**, *224*, 107079. [[CrossRef](#)]
36. Haussmann, E.; Fenzi, M.; Chitta, K.; Ivanecky, J.; Xu, H.; Roy, D.; Mittel, A.; Koumchatzky, N.; Farabet, C.; Alvarez, J.M. Scalable Active Learning for Object Detection. *arXiv* **2020**, arXiv:2004.04699.
37. Kee, S.; del Castillo, E.; Runger, G. Query-by-committee improvement with diversity and density in batch active learning. *Inf. Sci.* **2018**, *454–455*, 401–418. [[CrossRef](#)]
38. Wang, L.; Hu, X.; Yuan, B.; Lu, J. Active learning via query synthesis and nearest neighbour search. *Neurocomputing* **2015**, *147*, 426–434. [[CrossRef](#)]
39. Meirelles, A.L.S.; Kurc, T.; Saltz, J.; Teodoro, G. Effective active learning in digital pathology: A case study in tumor infiltrating lymphocytes. *Comput. Methods Programs Biomed.* **2022**, *220*, 106828. [[CrossRef](#)] [[PubMed](#)]
40. Shen, Y.; Song, Y.; Wu, C.H.; Kuo, C.C.J. TBAL: Two-stage batch-mode active learning for image classification. *Signal Process. Image Commun.* **2022**, *106*, 116731. [[CrossRef](#)]
41. Jin, Q.; Yuan, M.; Qiao, Q.; Song, Z. One-shot active learning for image segmentation via contrastive learning and diversity-based sampling. *Knowl.-Based Syst.* **2022**, *241*, 108278. [[CrossRef](#)]
42. Cui, Z.; Lu, N.; Wang, W. Pseudo loss active learning for deep visual tracking. *Pattern Recognit.* **2022**, *130*, 108773. [[CrossRef](#)]
43. Yoo, D.; Kweon, I.S. Learning Loss for Active Learning. *arXiv* **2019**, arXiv:1905.0.
44. Vikas Desai, S.; Balasubramanian, V.N. Towards Fine-grained Sampling for Active Learning in Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Washington, DC, USA, 4–19 June 2020; pp. 4010–4014. [[CrossRef](#)]
45. Li, Y.; Fan, B.; Zhang, W.; Ding, W.; Yin, J. Deep active learning for object detection. *Inf. Sci.* **2021**, *579*, 418–433. [[CrossRef](#)]
46. Roy, S.; Unmesh, A.; Nambodiri, V.P. Deep Active Learning for Object Detection. In Proceedings of the British Machine Vision Conference 2018, BMVC 2018, Northumbria University, Newcastle, UK, 3–6 September 2018; p. 91.
47. Aghdam, H.H.; Gonzalez-Garcia, A.; van de Weijer, J.; López, A.M. Active Learning for Deep Detection Neural Networks. *arXiv* **2019**, arXiv:1911.09168, [[1911.09168](#)].
48. Choi, J.; Elezi, I.; Lee, H.; Farabet, C.; Alvarez, J.M. Active Learning for Deep Object Detection via Probabilistic Modeling. *arXiv* **2021**, arXiv:2103.16130.
49. Gui, X.; Lu, X.; Yu, G. Cost-effective Batch-mode Multi-label Active Learning. *Neurocomputing* **2021**, *463*, 355–367. [[CrossRef](#)]
50. Kothawade, S.; Chopra, S.; Ghosh, S.; Iyer, R. Active Data Discovery: Mining Unknown Data using Submodular Information Measures. *arXiv* **2022**, arXiv:cs.CV/2206.08566.
51. Carse, J.; McKenna, S. Active Learning for Patch-Based Digital Pathology Using Convolutional Neural Networks to Reduce Annotation Costs. In *Digital Pathology*; Reyes-Aldasoro, C.C., Janowczyk, A., Veta, M., Bankhead, P., Sirinukunwattana, K., Eds.; Springer: Cham, Switzerland, 2019; pp. 20–27.
52. Jarl, S.; Aronsson, L.; Rahrovani, S.; Chehrehgani, M.H. Active learning of driving scenario trajectories. *Eng. Appl. Artif. Intell.* **2022**, *113*, 104972. [[CrossRef](#)]

53. Feng, D.; Wei, X.; Rosenbaum, L.; Maki, A.; Dietmayer, K. Deep Active Learning for Efficient Training of a LiDAR 3D Object Detector. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 667–674. [[CrossRef](#)]
54. Schmidt, S.; Rao, Q.; Tatsch, J.; Knoll, A. Advanced Active Learning Strategies for Object Detection. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 871–876. [[CrossRef](#)]
55. Lv, X.; Duan, F.; Jiang, J.J.; Fu, X.; Gan, L. Deep Active Learning for Surface Defect Detection. *Sensors* **2020**, *20*, 1650. [[CrossRef](#)] [[PubMed](#)]
56. Wang, K.; Lin, L.; Yan, X.; Chen, Z.; Zhang, D.; Zhang, L. Cost-Effective Object Detection: Active Sample Mining With Switchable Selection Criteria. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 834–850. [[CrossRef](#)]
57. Tang, Y.P.; Wei, X.S.; Zhao, B.; Huang, S.J. QBox: Partial Transfer Learning With Active Querying for Object Detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 3058–3070. [[CrossRef](#)]
58. Gu, H.; Haeri, M.; Ni, S.; Williams, C.K.; Zarrin-Khameh, N.; Magaki, S.; Chen, X.A. Detecting Mitoses with a Convolutional Neural Network for MIDOG 2022 Challenge. *arXiv* **2022**. [[CrossRef](#)]
59. Yun, J.B.; Oh, J.; Yun, I.D. Gradually Applying Weakly Supervised and Active Learning for Mass Detection in Breast Ultrasound Images. *arXiv* **2020**, arXiv:2008.08416.
60. Huang, W.; Sun, S.; Lin, X.; Zhang, D.; Ma, L. Deep active learning with Weighting filter for object detection. *Displays* **2022**, *76*, 102282. [[CrossRef](#)]
61. Kyun, S.D.; Ahmed, M.U.; Rhee, P.K. Incremental Deep Learning for Robust Object Detection in Unknown Cluttered Environments. *arXiv* **2018**, arXiv:1810.10323.
62. Brust, C.A.; Käding, C.; Denzler, J. Active Learning for Deep Object Detection. *arXiv* **2018**, arXiv:1809.09875.
63. Yuan, T.; Chen, Z.; Luo, P.; Liu, X.; Jiang, Y.; Qiu, Q. Multiple Instance Active Learning for Object Detection. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 5326–5335.
64. Wang, K.; Yan, X.; Zhang, D.; Zhang, L.; Lin, L. Towards Human-Machine Cooperation: Self-supervised Sample Mining for Object Detection. *arXiv* **2018**, arXiv:1803.09867.
65. Agarwal, S.; Arora, H.; Anand, S.; Arora, C. Contextual Diversity for Active Learning. *arXiv* **2020**, arXiv:2008.05723.
66. Roy, S.; Namboodiri, V.P.; Biswas, A.K. Active learning with version spaces for object detection. *arXiv* **2016**, arXiv:1611.07285.
67. Rhee, P.K.; Erdene, E.; Kyun, S.D.; Ahmed, M.U.; Jin, S. Active and semi-supervised learning for object detection with imperfect data. *Cogn. Syst. Res.* **2017**, *45*, 109–123. [[CrossRef](#)]
68. Chandra, A.L.; Desai, S.V.; Balasubramanian, V.N.; Ninomiya, S.; Guo, W. Active Learning with Weak Supervision for Cost-Effective Panicle Detection in Cereal Crops. *arXiv* **2019**, arXiv:1910.01789.
69. Qu, Z.; Du, J.; Cao, Y.; Guan, Q.; Zhao, P. Deep Active Learning for Remote Sensing Object Detection. *arXiv* **2020**, arXiv:2003.08793.
70. Chitta, K.; Alvarez, J.M.; Haussmann, E.; Farabet, C. Training Data Subset Search with Ensemble Active Learning. *arXiv* **2019**, arXiv:1905.12737.
71. Maiettini, E.; Becattini, F.; Papi, F.; Seidenari, L.; Bagdanov, A.D. From Handheld to Unconstrained Object Detection: A Weakly-supervised On-line Learning Approach. *arXiv* **2020**, arXiv:2012.14345.
72. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [[CrossRef](#)]
73. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
74. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361. [[CrossRef](#)]
75. Sun, P.; Kretschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; et al. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. *arXiv* **2019**, arXiv:1912.04838.
76. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A multimodal dataset for autonomous driving. *arXiv* **2019**, arXiv:1903.11027.
77. Desai, S.V.; Chandra, A.L.; Guo, W.; Ninomiya, S.; Balasubramanian, V.N. An Adaptive Supervision Framework for Active Learning in Object Detection. *arXiv* **2019**, arXiv:1908.02454.
78. Lin, Z.; Wei, D.; Jang, W.D.; Zhou, S.; Chen, X.; Wang, X.; Schalek, R.; Berger, D.; Matejek, B.; Kametsky, L.; et al. Two Stream Active Query Suggestion for Active Learning in Connectomics. In Proceedings of the Computer Vision—ECCV 2020, Glasgow, UK, 23–28 August 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 103–120.
79. Vandoni, J.; Aldea, E.; le Hégarat-Mascle, S. Evidential query-by-committee active learning for pedestrian detection in high-density crowds. *Int. J. Approx. Reason.* **2019**, *104*, 166–184. [[CrossRef](#)]
80. Gissin, D.; Shalev-Shwartz, S. Discriminative Active Learning. *arXiv* **2019**, arXiv:1907.06347.
81. Chattopadhyay, A.; Sarkar, A.; Howlader, P.; Balasubramanian, V.N. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 839–847. [[CrossRef](#)]

82. Yang, Y.; Loog, M. Single shot active learning using pseudo annotators. *Pattern Recognit.* **2019**, *89*, 22–31. [[CrossRef](#)]
83. Korzynska, A.; Roszkowiak, L.; Zak, J.; Siemion, K. A review of current systems for annotation of cell and tissue images in digital pathology. *Biocybern. Biomed. Eng.* **2021**, *41*, 1436–1453. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.