UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

A Component-based Systematic Review of Pool-based Active Learning for 2D Object Detection

Mohammed Yasin[1*], Md Sah Hj Salam[2], Tarmizi Adam[3]
Faculty of Computing,
Universiti Teknologi Malaysia
81310, UTM Johor Bahru, Johor, Malaysia
Email: mohammedyasin@graduate.utm.my[1]; sah@utm.my[2]; tarmizi.adam@utm.my[3]

Submitted: 21/4/2024. Revised edition: 29/7/2024. Accepted: 12/8/2024. Published online: 25/11/2024
DOI: https://doi.org/10.11113/ijic.v14n2.448

***Abstract*—Active learning can help reduce the labelling burden in training deep object detection models by strategically selecting the most informative data points for labelling from unlabelled data. Different approaches to this task have been proposed in recent years, with differences often found in their approaches to the specific components. Consequently, this study breaks down active learning methods into their individual components and examines the diverse strategies associated with each intending to shed light on areas that have not yet been explored in depth. Doing so, the study offers valuable insights into potential research directions in this area.**

***Keywords*—Active learning, object detection, data-efficient learning, systematic review, neural networks**

I. Introduction

Deep learning models require large amounts of labelled data to generalize well. However, labelling costs often make gathering such large amounts of labelled data too expensive and economically infeasible. This traditional approach to deep learning falls under passive learning, where the datapoints to label are hand-picked by humans. It does not take into account the model's current knowledge, hence why it is passive. As an alternative, active learning is a method of learning that puts the model in the loop of selecting the best datapoints for labelling, incorporating the model's current knowledge into the selection process.

There are two broad approaches to active learning: sampling-based approaches and query synthesis approaches. Sampling-based approaches can be divided into two types, each having its own advantages and disadvantages: pool-based active learning and stream-based active learning. Pool-based active learning provides a way to choose the most informative data points to label from a pool unlabelled data such the model has the highest gain in performance. This is as opposed to stream-based active learning where data is being streamed continuously, such as in online learning scenarios. Meanwhile, query synthesis involves not selecting but generating synthetic data through the learner that ought to be labelled. Although active learning has seen a lot of use in the domain of image classification, active learning for object detection still largely remains an active area of research, particularly due to the nature of object detection algorithms. In this study, the literature concerning active learning for 2D object detection is reviewed and then analyzed based on their components. The division into components sets it apart from other systematic review studies in this area.

II. Research Questions

As highlighted previously, the review will analyze the components of active-learning methods proposed, and hence the study will be answering the following research questions:

1. What are the different components of active learning methods for object detection?
2. What components of active learning methods lack diversity in terms of approaches?

III. Research Methodology

For this study, relevant works in active learning for 2D object detectors were discovered through the use of keyword searches in prominent scholarly databases. The keywords used in this study are listed in Table I.

As depicted in the table, the search terms for active learning methods are combined with the search terms for object detection so that the methods that are only in the domain of object detection are extracted. Note that no special keyword to limit search results to 2D object detection was used at this level as it is not consistently referenced across the literature. Similarly, pool-based active learning was also not specified among the keywords. Furthermore, the search was limited to abstract as any work that made significant contribution to the above ought to have these keywords in their abstract. Finally, the search was conducted in: Web of Science, Scopus and arXiv.

TABLE I. BOOLEAN GROUPING OF SEARCH TERMS

Search Terms Group	Keywords	Operation with next string
Active learning	active learning	AND
Object detection	object detection OR object detector OR object detectors	

TABLE II. NUMBER OF RESULTS RETURNED BY EACH DATABASE

Search	Database	No. of papers found	Chosen based on inclusion/exclusion criteria
Active learning methods for object detection	Web of Science	137	23
	arXiv	400	
	Scopus	244	

TABLE III. INCLUSION AND EXCLUSION CRITERIA FOR SCREENING

Inclusion Criteria	• Must be in the domain of 2D object detection. • Must be a general method. • Must have extensive benchmark results of their method on at least Microsoft Common Objects in Context (MS COCO) and/or PASCAL VOC. • Must be from a reputable journal or conference. • Must be a theoretical paper proposing a new method.
Exclusion Criteria	• Must not be in 3D object detection. • Must not be primarily a weakly-supervised, semi-supervised, unsupervised method. • Must not simply applying an existing method to a particular problem.

These databases provide comprehensive coverage of high-quality research. Web of Science and Scopus are well-known databases indexing publications from peer-reviewed journals and conference papers, while arXiv, on the other hand, is a popular repository for preprints, many a times containing research that are still under review prior to publication. Moreover, 317 results (~40%) from the search were duplicates, as shown in Fig. 1. This significant overlap indicates that the search was exhaustive and additional databases were unnecessary.

As for the execution of the query, Web of Science and Scopus had graphical interface that allowed the required grouping of terms, but for arXiv, the API was used to perform the search as grouping boolean operations was not possible through their search engine user interface.

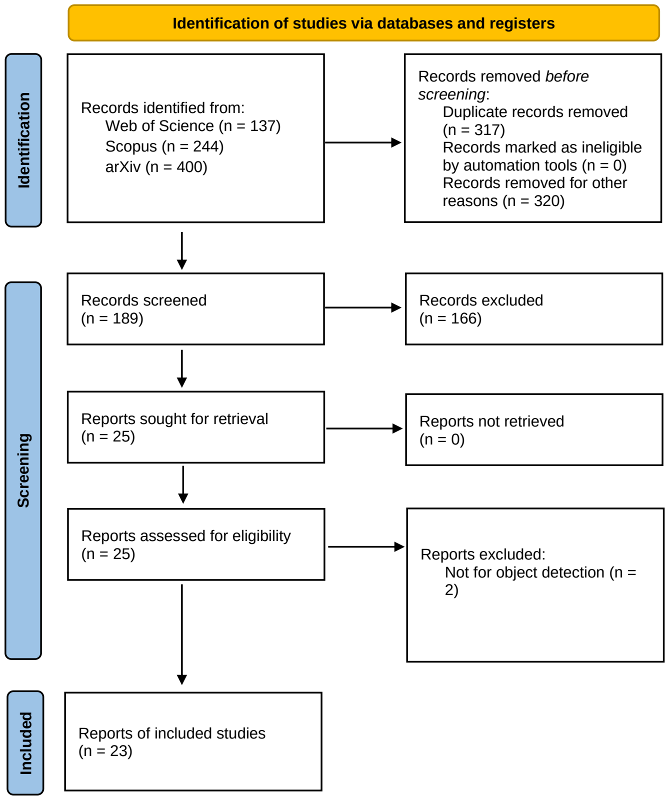

Fig. 1. PRISMA Flow Diagram

IV. LITERATURE SEARCH AND SELECTION

The number of works that resulted from using the keyword search is summarized in Table II. The works were then further filtered using the inclusion and exlusion criteria mentioned in Table III. The PRISMA diagram indicating the flow of the literature search is shown in Fig. 1.

V. ACTIVE LEARNING METHOD COMPONENT ANALYSIS

In this section, the different components of active learning methods are analyzed, discussing the notable approaches taken by the works to achieve the objectives associated with each component.

A. Image, Region and Instance Level Query Strategies

Object detection entails identifying multiple instances of objects within an image. Consequently, active learning methods for object detection often assess informativeness based on metrics obtained at the instance level. However, since labelling typically occurs for the entire image rather than individual objects, these metrics are usually aggregated to determine informativeness at the image level. Most proposed active learning methods follow this approach [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. This paradigm, however, has been challenged as being inefficient as the limited labelling budget gets wasted annotating instances that may or may not be informative.

The first attempt at instance-level querying was made by [17] where they queried only for the most uncertain instances in an image and used pseudo-labels to compensate for the missing labels for instances that were excluded from query in the selected image. More recently, [18] demonstrated that existing image-level query methods can be easily surpassed by querying for the most crowded boxes, reinforcing the notion that image-level query strategies do not effectively sample the best possible dataset for training. They introduced a new box-level active learning method that combines pseudo-labels with a committee based acquisition function, significantly outperforming existing methods. However, instance level querying is prone to false negatives as querying for a portion of an image naturally leads to confusion when the queried region contains multiple overlapping objects. Arguing this point, [19] and [20] introduced region-level querying (ReAL) strategies that aim to strike a balance between coarse image-level strategies and fine-grained instance approaches. [19] advocate for querying fixed-size regions, while [20] adopt a dynamic approach where the region expands to include nearby uncertain instances. On the other hand, [21] proposed a novel decoupled localization and recognition query strategy (DeLR) that tries alleviating some of the issues associated with instance level querying by arguing that instances whose pseudo-labels have confident localization predictions likely also would have been classified correctly, and hence do not require corrections.

TABLE IV. ACTIVE LEARNING METHODS WITH THEIR COMPONENTS

Year	**Paper**	**Query Type**	**Scoring Function Type**	**Aggregation Function**	**Diversity Sampling**	**Class-balancing**	**Sampling Strategy**	**Retraining Strategy**	**Initial Set**
2018	[1]	Image	Uncertainty	Sum*	Batched sampling	Simple reweighting	Top-k	Scratch	Random
2018	[7]	Image	Uncertainty	Sum	Not used	Not used	Top-k	Scratch	Random
2019	[13]	Image	Expected model change.	Not applicable Image level	Not used	Not used	Top-k	Scratch	Random
2020	[17]	Instance	Multiple	Not applicable	Not used	Not used.	Top-k	Scratch	Random
2020	[11]	Image	Uncertainty	Average	Not used	Weighted loss	Top-k	Continuous fine-tuning	Random
2020	[4]	Image	Uncertainty*	Sum*	Core-set*	Not used	Top-k	Scratch	Random
2020	[22]	Image	Uncertainty w/ diversity	Contextual Diversity (novel)	Part of Contextual Diversity	Not used	Top-k	Scratch	Random
2021	[3]	Image	Uncertainty w/ robustness	Maximum	Not used	Not used	Top-k	Scratch	Random
2021	[15]	Image	Distribution based	Mean*	Not used	Not used	Top-k	Scratch	Random
2021	[8]	Image	Expected gain	Not applicable Image level	Modified core-set + distance	Not used.	Top-k	Scratch	Random
2021	[2]	Image	Uncertainty	Maximum	Not used	Not used	Top-k	Scratch	Random
2022	[12]	Image	Uncertainty	ENMS**	DivProto**	Part of DivProto	Top-k	Scratch	Random
2022	[14]	Image	Uncertainty w/ consistency augmentation	Maximum (min. inconsistency)	Mutual information	Part of diversity sampling	Top-k	Scratch	Random
2022	[10]	Image	Uncertainty	HUA	Not used	Not used	Top-k + filtered	Scratch	Random
2022	[23]	Image	Uncertainty - DCUS**	Sum	CCMS**	Not used	Top-k	Scratch	Random
2023	[6]	Image	Multiple uncertainty	Not applicable	Cosine similarity	Not used	W-Filter**	Scratch	W-Filter**
2023	[19]	Region	Uncertainty	Average	Not used	Frequency based reweighting	Top-k	Scratch	Random
2023	[5]	Image	Uncertainty w/ diversity	Not applicable Image level	Distance of object-level features	Not used	Top-k	Scratch	Random
2023	[18]	Instance	Uncertainty w/ consistency augmentations	Not applicable	Not used	Not used	Top-k	Scratch (VOC) Fine-tune (COCO)	Random
2023	[20]	Region	Uncertainty w/ diversity	ReAL**	Part of informativeness scoring function	Not used	Top-k	Scratch	Random
2023	[16]	Image	Uncertainty	Sum	Not used	Not used	Top-k	Scratch	Random
2023	[9]	Image	LSTM	Not applicable Image level	Not used	Not used	MAGRAL**	Scratch	Random
2023	[21]	Instance	Uncertainty	Not applicable	Not used	Not used	DeLR**	Scratch	Random

* best in ablation. ** novel method introduced by the work.

Overall, active learning methods for object detection have predominantly focused on creating new image-level query strategies, but recent works have shown increased interest in proposing region-level and instance-level query methods. While it is argued that image-level query methods result in the wastage of labelling budgets, it is also true that region and instance level methods cause more ambiguity in labelling due to their unorthodox nature where the annotator has to provide labels for cropped regions of image which increases chances of partial labels and false negatives.

B. Informativeness Scoring Function

Active learning strategies utilize a scoring function to quantify the informativeness of an instance. The design of this scoring function is one of the primary elements that separates one proposed method from another. The methods used to obtain informativeness score can be classified into four types: uncertainty-based, distribution-based, expected model change methods [24] and more recently reinforcement-learning based methods [9], [22].

Most proposed active learning methods for object detection are uncertainty-based. Some of the initial uncertainty-based methods relied only on the classification uncertainty. [1] used 1 vs. 2 method comparing the confidence of the top two predictions, [17] utilized the least predicted class probability and [12] relied on entropy, [6] tested various redesigned uncertainty scoring functions alongside their novel frequency based weighting-filter (W-filter), while [22] introduced a contextual diversity metric arguing that confusion between classes in a given region provided a better measure of spatial uncertainty. However, since localization is one of the major components of object detection, many of the later uncertainty-based methods started incorporating measures of localization uncertainty into the scoring function. [7] incorporated localization uncertainty along with classification uncertainty by using localization stability and localization tightness, while [23] introduced a category-wise difficulty scaling based on a combination of class probabilities and intersection-over-union and used it reweight the calculated uncertainties. Some methods proposed calculating aleatoric and epistemic uncertainty to differentiate between the reducible and irreducible type of uncertainty, such as [2] through gaussian mixture models and [10] by adding a model evidence head inspired by evidential deep learning approaches.

Moreover, a number of methods relied on robustness or consistency to determine uncertainty, essentially constituting a sub-class of uncertainty-based methods. For example, [14] added localization inconsistency produced to the uncertainty calculation, while [3] proposed an augmentation based consistency score to measure robustness alongside uncertainty. [18] also employed augmentations, but instead used them like an input-end committee to measure disagreement and thereby quantify the uncertainty. [21] used a cleverer approach by training two different heads and using the inconsistency between their predictions as a measure of uncertainty.

Besides single model approaches, predictions from an ensemble of models trained are also used to derive the informativeness score. [4] tested various scoring functions that utilize an ensemble of models and found that mutual information with max aggregation provided the best trade-off taking into account labelling cost. [11] designed a consensus score based on average of minimum IoU for each matching Region-of-Interest (RoI) obtained from the ensemble of models, but it did not perform better than random sampling. [16] proposed a classification committee with a specialized loss to maximize the discrepancy between them in order to utilize the discrepancy as a measure of uncertainty but showed marginal improvements despite the added cost of the approach.

The second type of informativeness scoring functions that are considered distribution-based methods try to capture the distribution and thereby the information of an unlabelled set through a subset sampled from the unlabelled set. Yuan *et al.* [15] trained adversarial classifiers that are used to compute prediction uncertainty to accomplish this for object detectors.

The third type referred to as expected model change methods try to predict the gain a sample from the unlabelled set would produce if it were added to the training set. [13] were the first to propose such a method by learning to predict the loss for each image and hence the associated informativeness. Later, working on the same principle, [8] proposed predicting the expected gradient reduction and expected error reduction for each sample.

Lastly, some methods have attempted a reinforcement-learning based approach to selecting the most informative images. [22] used their aggregated contextual diversity metric as reward to train a Bi-LSTM (bi-directional long short-term memory) sampling agent. More recently, [9] argued that minimizing uncertainty does not necessarily correspond with gain in performance. They introduced MeanAP Guided Reinforced Active Learning (MAGRAL) where they used reinforcement learning-based to train an LSTM model which acted as a sampling agent to select the best the images to label, using mean average precision as the reward which they argued was a more natural objective in increasing a model's performance.

C. Aggregation Function

Unlike active learning used in image classification, there is usually an additional step in designing an acquisition function for active learning with object detectors. This comes from the fact that image classification requires image-level labels while object detectors require instance-level labels. Due to this, within an image, there can be multiple instances with different informativeness scores. These scores need to be aggregated to get an image-level score because sample selection is typically done for the whole image. That is, the whole image is selected to be labelled, not just one instance.

[1] was the first to propose sum, average and maximum as aggregation function for object detection. Later many works corroborated the effectiveness of the maximum aggregation function such as [4], [7]. [2] empirically showed that maximum aggregation for uncertainty-based methods were better than any other form of aggregation.

Although, sampling new data to annotate at the image level is the common approach as mentioned in the beginning, instance level and region-level sampling has also been shown to be viable. The obvious problem that arises in such settings is partially annotated images, as the model would still be training of on the whole image, not a particular portion of it holding the instance that was sampled. To compensate for missing labels, [17] used pseudo-labelling where missing labels are obtained from the trained model. However, since these labels may not be accurate, they also introduced a dampening factor in the loss function to give less weight to these noisy pseudo-labels in the calculation of the final loss. Meanwhile, region-level methods such as [5] and [20] crafted nuanced aggregation functions, incorporating not just uncertainties of the instances involved but also diversity by penalizing similarity with other regions in the final region level scores.

Some methods employ multi-step aggregation algorithms, usually involving some form of similarity calculation, instead of performing a naïve aggregation over all the instances in an images. [12] used an entropy-based aggregation method where they dropped any instances that were similar to the picked instance at every step, while summing the entropy to get the image-level score. [10] introduced a novel hierarchical uncertainty aggregation (HUA) method to aggregate the uncertainties by distributing the predictions and their associated uncertainty score at three different hierarchical levels, namely: class-level, scale-level and object-level. At each level, they use a trivial aggregation function to aggregate the uncertainties in those levels separately. They demonstrated their aggregation method to achieve better performance than indiscriminately using a trivial aggregation function as had so far been the case in previous works.

D. Diversity Sampling

Although the informativeness score can help sample images with informative instances, it does not guarantee diversity in the sampling. Images of a particular class may get sampled more than other classes, leading to a data bias. This can be because models can also be confidently wrong about a prediction, which means informativeness score that rely on prediction confidence may incorrectly consider an image to be uninformative. To resolve this, many works also incorporate a diversity sampling strategy to force acquisition of training data that is more diverse. [1] for example, randomly divided the unlabelled dataset into batches of 10 images each, so that sampling happens on the batch-level based on sum of informativeness score aggregated across all the images the whole batch, although they warn that this could lead to unintended side effects. Meanwhile, [8], [22] utilized diversity sampling based on k-Centre-Greedy method. [8] tested both L1 and L2 norm to calculate distance between samples and found L2 distance to be the best, which was also the distance used by [22]. [14] used mutual information to remove redundant samples. [4] found core-set to be the best strategy for diversity sampling in their ablation study, while [8] modified the core-set strategy by incorporating a distance function based on uncertainty and diversity. The aforementioned methods all calculated the distances at the image-level. On the other hand, [5] proposed calculating the distance using object-level features, while [23] utilized the object-level features to create a novel metric called Category Conditioned Matching Similarity (CCMS) that used the object-level feature similarity as a proxy for image-level similarity. Meanwhile [20], given their region-level approach, used a query-neighbour strategy to calculate similarity between different regions.

Moreover, certain works used their proposed scoring function as a metric for similarity. [22] used their class-specific confusion metric with Kullback-Leibler (KL) divergence to compute pairwise similarity, [6] made use of their W-filter to increase diversity and [12] combined their Entropy-based Non-Maximum Suppression (ENMS) to promote intra-image diversity with Diverse Prototype (DivProto) to promote intra-image diversity. The latter's method of selecting diverse images for labelling was very efficient and did not require pixel level pairwise instance comparisons which greatly reduced the time complexity.

E. Class Balancing

Apart from diversity sampling, some works also introduced class balancing techniques to specifically address performance issues for minority classes. [1] used inverse of the posterior to weight the informativeness score and [19] also used a similar strategy to reweight the uncertainty score. [11] introduced active class weighting at the loss level taking the ratio of total labels to that of the class in concern. Similarly, [8] argued that the loss has more relevance to improving the mean-average precision (mAP) than ratio of class labels, and therefore applied a loss-based weighting to acquisition function. [12] used class balancing to improve inter-class diversity by giving more sampling quota to the minority classes. [6] used their frequency-based W-filter to weight the informativeness score.

All in all, explicit class balancing techniques has only been employed by a handful of works. It could also be argued that explicit class balancing may not be required considering the use of diversity sampling which may be able to take care of the underrepresented classes. For example, [14] managed to alleviate the unbalanced distribution as part of their diversity sampling strategy using mutual information, hence avoiding the need to craft a special class balancing strategy.

F. Sampling Strategies

Once the unlabelled data has been scored and ranked by an acquisition function, there are multiple approaches that could be taken for sampling the ranked unlabelled data. [4] tested top-K, top-third, top-K/2-bottom-K/2, bottom-K, and they found top-K to be the best strategy and bottom-K the worst. Although, most work utilize the top-K strategy, some recent works have also proposed more nuanced sampling strategies. [10] showed that including a certain ratio of filtered out images besides top-K had a slight improvement in performance. [6] sampled images from each grouped frequency domain through their W-filter strategy. [9], owing to their reinforcement learning-based approach, used their trained LSTM-based sampling agent to sample the best images to label.

On the other hand, [21] took an unorthodox approach and developed a more intricate querying strategy. They first obtained the pseudo-labels and then performed a decoupled query to the oracle. The first query verified the localization of the pseudo-labels, and then the second query used the first that information to again query to verify the classes of the pseudo-labels whose localization and classification scores were both uncertain. Their intuition behind the decoupling was that pseudo-labels with high localization scores need not be queried for class labels as they found the likelihood of the predicted class being correct was high.

G. Retraining Strategy

The literature was divided as to whether each active learning cycle should train a new model or fine-tune the model from the last cycle. [11] found that fine-tuning was better than training from the scratch. However, [24] argued training from scratch is better than fine-tuning and tried to compensate for it using knowledge distillation to distil knowledge from the model in the last active learning cycle. But their test was performed on image classification, while [11] specifically obtained that result on object detection. The difference in the findings could be explained by the work of [25] where the authors challenged the assumption that the last trained model was better than models from earlier active learning cycles in estimating the uncertainty score and demonstrated that assumption to be false. They attributed this to the fact that deep learning models suffer from example forgetting which can lead to degraded performance on a particular class in the next cycle. They use correct inconsistency score to keep track of samples that are prone to forgetting and determine the best model from previous active learning cycles based on mean of correct inconsistency and distil knowledge from that using pseudo-labels. Even though this is a plausible explanation, the work by [26] suggests that warm-starting is fundamentally problematic.

H. Initial Set

Most the works on active learning for object detection start with a random initial set. [6], however, argued that initial set has a significant impact on the effectiveness of active learning as they influence the informativeness score which often come from the trained model's predictions. To resolve this, they propose a weighting filter that use frequency domain information to weight the samples in terms of diversity. They argue that samples with diverse frequencies are more informative as they carry more edge information. On the contrary, [2] found that initial set had minimal impact on the performance of their active learning strategy.

To reconcile the two, one could argue that the effect of initial set would depend on the acquisition function used. An ideal acquisition function should not be affected by bad model predictions.

VI. MAPPING OF LITERATURE

The works selected for the study have been concisely mapped based on their approach to each of the components of active learning in Table IV. This provides a overview of the analysis done which would be referenced to answer the research questions of this study.

VII. DISCUSSION

This section addresses the research questions posed by the study, drawing on the literature reviewed.

RQ1: What are the different components of active learning methods for object detection?

The analysis in Section V identifies several key components of active learning methods. These include a query strategy, an informativeness scoring function, and an aggregation strategy, particularly for image-level query methods. A final sampling strategy is used to glue all these components together to produce the dataset to be labelled. Additionally, some approaches incorporate a diversity sampling function to enhance the variety of acquired data, while a handful also use class-balancing techniques to enhance the representation of minority classes. Lastly, few works have also attempted different training strategies (e.g., training from scratch or fine-tuning a pre-existing model), and one of the works tried an alternative method to sample the initial dataset, and both of these could also be considered components of active learning methods.

RQ2: What components of active learning methods lack diversity in terms of the approaches?

It can be seen from Table IV that most approaches perform query at the image level and adopt an uncertainty based scoring function in one form or another to estimate informativeness. However, the recent work by [18] put into question the efficacy of image level methods. Similarly, [9] challenged the underlying assumption with using uncertainty as a metric of informativeness as it does not necessarily correlate with increase mean average precision, the primary metric of the benchmarks. Moreover, very few works were found to have attempted different sampling strategies and class-balancing techniques, the former mostly being due to the observation that in most cases top K sampling is optimal in obtaining the best possible dataset. On the other hand, the limited focus on dedicated class-balancing may be attributed to the assumption that the acquisition function automatically compensates for underperforming minority classes. The same also could be said about diversity sampling.

Besides these, most works have not explored alternative strategies to randomly sampling the initial set since most scoring functions require the presence of a partially trained model to work which does not become available until the end of the first cycle of training. Similarly, it can also be seen that most works discard the previously trained model and train a new model from scratch in the subsequent cycles, particularly due to the warm-starting problem highlighted in [26].

VIII. CONCLUSIONS

Although various active learning methods have been proposed for object detection, certain components have seen less diverse efforts in terms of introducing novelty. Future work in this area could focus on proposing better methods and strategies for these components. Particularly, more work is required to combat the false negative problem found in region and instance query based methods that reduce the appeal of such approaches despite their promising results in benchmarks. Likewise, methods that can leverage the trained model from the previous cycle rather than discarding it entirely could potentially enhance efficiency, especially if they can address the challenges associated with warm start-ups. Not to mention the significant reduction in training time and cost that could be yielded by such methods. Likewise, and perhaps the most challenging among them all, developing a strategy to assess the informativeness of samples without relying on a trained model from the initial cycle could mitigate issues stemming from random initial sets.

Acknowledgements

This research received no external funding.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper

References

[1] C.-A. Brust, C. Käding, and J. Denzler. (2018). Active learning for deep object detection. *arXiv*. Doi: 10.48550/arXiv.1809.09875.

[2] J. Choi, I. Elezi, H.-J. Lee, C. Farabet, and J. M. Alvarez. (2021). Active learning for deep object detection via probabilistic modeling. *arXiv*. Doi: 10.48550/arXiv.2103.16130.

[3] I. Elezi, Z. Yu, A. Anandkumar, L. Leal-Taixe, and J. M. Alvarez. (2021). Not all labels are equal: Rationalizing the labeling costs for training object detection. *arXiv*. Doi: 10.48550/arXiv.2106.11921.

[4] E. Haussmann *et al.* (2020). Scalable active learning for object detection. *arXiv*. Doi: 10.48550/arXiv.2004.04699.

[5] A. Hekimoglu, A. Brucker, A. K. Kayali, M. Schmidt, and A. Marcos-Ramiro. (2023). Active learning for object detection with non-redundant informative sampling. *arXiv*. http://arxiv.org/abs/2307.08414.

[6] W. Huang, S. Sun, X. Lin, D. Zhang, and L. Ma. (2023). Deep active learning with Weighting filter for object detection. *Displays*, *76*, 102282. Doi: 10.1016/j.displa.2022.102282.

[7] C.-C. Kao, T.-Y. Lee, P. Sen, and M.-Y. Liu. (2018). Localization-aware active learning for object detection. *arXiv*. Doi: 10.48550/arXiv.1801.05124.

[8] Y. Li, B. Fan, W. Zhang, W. Ding, and J. Yin. (2021). Deep active learning for object detection. *Information Sciences*, 579, 418-433. Doi: 10.1016/j.ins.2021.08.019.

[9] Z. Liang, X. Zeng, R. Zhao, and P. Luo. 2023. MeanAP-guided reinforced active learning for object detection. *arXiv*. Doi: 10.48550/arXiv.2310.08387.

[10] Y. Park, W. Choi, S. Kim, D.-J. Han, and J. Moon. (2023). Active learning for object detection with evidential deep learning and hierarchical uncertainty aggregation. presented at the eleventh international conference on learning representations. https://openreview.net/forum?id=MnEjsw-vj-X.

[11] S. Schmidt, Q. Rao, J. Tatsch, and A. Knoll. (2020). Advanced active learning strategies for object detection. *2020 IEEE Intelligent Vehicles Symposium (IV)*. 871-876. Doi: 10.1109/IV47402.2020.9304565.

[12] J. Wu, J. Chen, and D. Huang. (2022). Entropy-based active learning for object detection with progressive diversity constraint. *arXiv*. Doi: 10.48550/arXiv.2204.07965.

[13] D. Yoo and I. S. Kweon. (2019). Learning loss for active learning. Doi: 10.48550/ARXIV.1905.03677.

[14] W. Yu, S. Zhu, T. Yang, and C. Chen. (2022). Consistency-based active learning for object detection. *arXiv*. Doi: 10.48550/arXiv.2103.10374.

[15] T. Yuan *et al.* (2021). Multiple instance active learning for object detection. arXiv. Doi: 10.48550/arXiv.2104.02324.

[16] L. Zhao, B. Li, and X. Wei. (2023). Classification committee for active deep object detection. *arXiv*. Doi: 10.48550/arXiv.2308.08476.

[17] S. Vikas Desai and V. N. Balasubramanian. (2020). Towards fine-grained sampling for active learning in object detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 4010-4014. Doi: 10.1109/CVPRW50498.2020.00470.

[18] M. Lyu *et al.* (2023). Box-level active detection. *arXiv*. Doi: 10.48550/arXiv.2303.13089.

[19] Y.-S. Liou, T.-H. Wu, J.-F. Yeh, W.-C. Chen, and W. H. Hsu. (2023). MuRAL: Multi-scale region-based active learning for object detection. arXiv. Doi: 10.48550/arXiv.2303.16637.

[20] M. Laielli *et al.* (2022). Region-level active detector learning. *arXiv*. 17. Doi: 10.48550/arXiv.2108.09186.

[21] Y. Zhang, Y. Deng, X. Zhang, J. Li, R. C. Qiu, and Q. Tian. (2023). DeLR: Active learning for detection with decoupled localization and recognition query. *arXiv*. Doi: 10.48550/arXiv.2312.16931.

[22] S. Agarwal, H. Arora, S. Anand, and C. Arora. (2020). Contextual diversity for active learning. *arXiv*. 13. Doi: 10.48550/arXiv.2008.05723.

[23] C. Yang, L. Huang, and E. J. Crowley. (2023). Plug and Play active learning for object detection. *arXiv*. http://arxiv.org/abs/2211.11612

[24] J. Yun, B. Kim, and J. Kim. (2020). Weight decay scheduling and knowledge distillation for active learning. *Computer Vision – ECCV 2020*. A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Lecture Notes in Computer Science. Cham: Springer International Publishing. 431-447. Doi: 10.1007/978-3-030-58574-7_26.

[25] B. Kwak, Y. Kim, Y. J. Kim, S. Hwang, and J. Yeo. 2022. TrustAL: Trustworthy active learning using knowledge distillation. arXiv. Doi: 10.48550/arXiv.2201.11661.

[26] J. T. Ash and R. P. Adams. 2023. On warm-starting neural network training. *arXiv.org*. https://arxiv.org/abs/1910.08475v3.