BILINGUAL PUBLISHING GROUP

Journal of Electronic & Information Systems
https://journals.bilpubgroup.com/index.php/jeis

ARTICLE

Enhancing Semantic Segmentation through Reinforced Active Learning: Combating Dataset Imbalances and Bolstering Annotation Efficiency

*Dong Han, Huong Pham, Samuel Cheng**

School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, 74135, USA

ABSTRACT

This research addresses the challenges of training large semantic segmentation models for image analysis, focusing on expediting the annotation process and mitigating imbalanced datasets. In the context of imbalanced datasets, biases related to age and gender in clinical contexts and skewed representation in natural images can affect model performance. Strategies to mitigate these biases are explored to enhance efficiency and accuracy in semantic segmentation analysis. An in-depth exploration of various reinforced active learning methodologies for image segmentation is conducted, optimizing precision and efficiency across diverse domains. The proposed framework integrates Dueling Deep Q-Networks (DQN), Prioritized Experience Replay, Noisy Networks, and Emphasizing Recent Experience. Extensive experimentation and evaluation of diverse datasets reveal both improvements and limitations associated with various approaches in terms of overall accuracy and efficiency. This research contributes to the expansion of reinforced active learning methodologies for image segmentation, paving the way for more sophisticated and precise segmentation algorithms across diverse domains. The findings emphasize the need for a careful balance between exploration and exploitation strategies in reinforcement learning for effective image segmentation.

Keywords: Semantic segmentation; Active learning; Reinforcement learning

1. Introduction

Semantic segmentation involves assigning a class label to each pixel within an image, effectively dividing the image into segments that carry semantic

*CORRESPONDING AUTHOR:
Samuel Cheng, School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, 74135, USA; Email: samuel.cheng@ou.edu

ARTICLE INFO
Received: 1 November 2023 | Revised: 25 November 2023 | Accepted: 27 November 2023 | Published Online: 4 December 2023
DOI: https://doi.org/10.30564/jeis.v5i2.6063

CITATION
Han, D., Pham, H., Cheng, S., 2023. Enhancing Semantic Segmentation through Reinforced Active Learning: Combating Dataset Imbalances and Bolstering Annotation Efficiency. Journal of Electronic & Information Systems. 5(2): 45-60. DOI: https://doi.org/10.30564/jeis.v5i2.6063

meaning. Unlike image classification, which assigns a single class label to the entire image, semantic segmentation is a more granular task, amounting to pixel-level classification [1]. Over the past few years, the computer vision community has heavily relied on effective deep neural networks (DNNs) designed for semantic segmentation, as evidenced by recent research [2-11]. These efficient DNNs are characterized by their low computational demands and quick inference times [12], and their widespread adoption has significantly influenced applications in various fields such as autonomous driving [13,14], semantic segmentation enables precise scene understanding, allowing the vehicle to identify and differentiate between various objects on the road, such as pedestrians, vehicles, traffic signs, and obstacles. This technology aids in real-time decision-making, helping the vehicle navigate complex environments and ensure the safety of passengers and pedestrians; robot manipulation [15,16], for robots to interact intelligently with their environment, they require a comprehensive understanding of the objects and structures in their surroundings. Semantic segmentation facilitates this by enabling robots to identify and differentiate between different objects and their corresponding spatial relationships. This capability is crucial for tasks such as object manipulation, navigation in dynamic environments, and human-robot collaboration, enhancing the overall efficiency and safety of robotic systems; and biomedical image analysis [17,18], this technology assists healthcare professionals in diagnosing diseases, monitoring the progression of conditions, and planning effective treatment strategies. By providing detailed insights into complex biological structures, semantic segmentation contributes to advancements in medical research, patient care, and disease management. where advanced computer vision systems are paramount. For these models to operate effectively, however, they typically rely on a substantial volume of pixel-level annotations, a process that often necessitates expensive human labor.

Semantic segmentation datasets, with their pixel-wise annotations for each image, have been instrumental in advancing computer vision tasks. However, they are not without their limitations. Here are some of the key limitations associated with semantic segmentation datasets: 1) Extensive pixel-level annotations. Semantic segmentation models often require a large volume of accurately labeled training data, where each pixel in the image is assigned a corresponding class label. This process demands meticulous and precise annotations, which can be time-consuming and resource-intensive. Obtaining such detailed annotations for diverse datasets can be challenging, particularly for complex scenes with numerous objects and intricate boundaries. 2) Labor-intensive annotation process. The pixel-wise annotation process for semantic segmentation datasets is a labor-intensive task, often requiring significant human effort and time [19]. This manual labeling process is labor-intensive, time-consuming, and can be prone to human error, especially when dealing with large datasets. As a result, the creation of high-quality annotated datasets requires significant human resources and can be a bottleneck in the development of accurate and robust semantic segmentation models. 3) Data imbalance and variability. Semantic segmentation datasets may suffer from data imbalance and variability in the distribution of classes within the dataset. Certain classes may be underrepresented, leading to biased model predictions and reduced performance in specific classes. Handling such data imbalance and variability is crucial to ensure that the model can generalize effectively across different scenarios and accurately segment diverse objects in various contexts. 4) Generalization and robustness. Semantic segmentation models must be capable of generalizing well to unseen data and diverse environments. Achieving robust performance across different lighting conditions, viewpoints, and environmental changes remains a significant challenge. Ensuring that the model can accurately segment objects in various real-world scenarios is essential for its practical deployment in applications such as autonomous driving, robotics, and biomedical image analysis.

This aspect gains prominence during the process of collecting annotated data under human supervision for the creation of either a novel dataset or the supplementation of an existing one. Mitigating the challenges entails the systematic and efficient selection of image regions warranting annotation. Active learning (AL) represents a well-established research discipline explicitly focused on this area. Its primary objective is the identification of the most informative samples for annotation, with the overarching goal of enhancing the performance of learning algorithms with a minimized data requirement, in contrast to a non-selective approach where the entire dataset undergoes indiscriminate labeling. Active learning methodologies can be broadly categorized into two main groups: (i) methodologies that integrate various manually crafted active learning strategies [20-22], and (ii) data-centric active learning approaches [23-25]. Notwithstanding the heightened cost and time associated with acquiring labels for semantic segmentation in comparison to image classification, the realm of active learning for semantic segmentation has garnered relatively less attention [26-28], primarily emphasizing the development of manually engineered strategies.

How can reinforced active learning be effectively employed to enhance semantic segmentation, specifically addressing challenges posed by dataset imbalances and improving annotation efficiency? The latest active learning techniques leveraging reinforcement learning primarily concentrate on annotating one sample at each step [29-31], progressing until a predetermined label budget is fulfilled. Inspired by the AL-RL model by Casanova et al. [32], the proposed approach expedites the annotation process by selectively choosing informative and representative images to accelerate model learning. Additionally, we tackle the issue of imbalanced datasets. For instance, in clinical contexts, biases related to age and gender can arise due to constraints on the diversity of medical image contributors. In natural images, certain categories may be significantly more abundant than others, potentially skewing the model's performance towards the most frequently represented category. We investigate strategies to mitigate these biases with the aim of enhancing efficiency and accuracy in semantic segmentation analysis.

Furthermore, we conduct an in-depth exploration of various Reinforced Active Learning methodologies for image segmentation to optimize the precision and efficiency of segmentation tasks across diverse domains. To achieve this, we implement a robust framework that integrates various Reinforcement Learning (RL) techniques, including Dueling Deep Q-Networks (DQN) [33], Prioritized Experience Replay [34], Noisy Networks [35], Emphasizing Recent Experience [36], Soft Update Target Network [37], and Adaptive Epsilon Greedy [38]. We test the proposed method in the CamVid [39] dataset. Our results illustrate both improvements and limitations associated with various approaches in terms of overall accuracy and efficiency in image segmentation tasks.

2. Related work

Active learning serves as a dedicated methodology focused on optimizing performance gains with a minimal number of labeled samples. Its primary goal is to identify the most informative samples from the unlabeled dataset, subsequently presented to an oracle, such as a human annotator, for labeling. This process effectively minimizes labeling costs while ensuring sustained performance. Active learning approaches can be classified into membership query synthesis [40,41], stream-based selective sampling [42,43], and pool-based [44] strategies, each derived from diverse application scenarios [45]. Certain methodologies amalgamate various techniques to enhance the overall performance of active learning. For example, Shui et al. [46] take into account the diversity and uncertainty of query samples, and try to discover a balance between those two approaches. Further investigation into traditional query strategies is undertaken [47]. Despite the considerable volume of existing research on active learning, it continues to grapple with the challenge of extending its applicability to high-dimensional data, such as images, text, and videos [48]. Consequently, the majority of active learning studies tend to focus on low-dimensional problems [49]. Several methodologies integrate various techniques to enhance the performance of artificial intelligence, such as leveraging the exploration-exploitation trade-off [50], on a bandit formulation [21] and reinforcement learning [51].

In recent times, there has been a growing interest in reinforcement learning as an approach to acquiring a labeling policy that directly optimizes the performance of the active learning algorithm. For example, Dhiman et al. [29] proposed an automated annotation model for Multimedia Streaming Applications (MAS) to address the existing challenges of slow speeds and inefficiencies in accessing multimedia content. By leveraging Multi-modal Active Learning (MAL) and Convolutional Recurrent Neural Network (CRNN) in tandem with Deep Reinforcement Learning (DRL), the model demonstrates superior retrieval accuracy and performance metrics. Gong et al. [52] proposed Meta Agent Teaming Active Learning (MATAL) framework that effectively minimizes the laborious efforts involved in pose annotations. Sadigh et al. [53] present an active learning method for Inverse Reinforcement Learning (IRL) that relies on human-provided preferences between two sample trajectories. In a similar vein, Kunapuli et al. [54] incorporate human expert information through preference elicitation for actions in a designated state. Ezzeddine et al. [55] integrate feedback from a human trainer, particularly in cases where the provided demonstrations are less than optimal. Liu et al. [56] utilize expert knowledge derived from oracle policies to develop a labeling policy. In contrast, Pang et al. [57] employ policy gradient methods to acquire knowledge for the function. In an alternative strategy, certain techniques aggregate all labeled data in a single comprehensive step. Contardo et al. [58] employ a bi-directional RNN to select all samples simultaneously, particularly for the task of one-shot learning. Meanwhile, Sener et al. [59] suggest choosing a batch of representative samples that maximize coverage across the entire unlabeled set.

Recent active learning work has also looked at semantic segmentation [60]. Uncertainty-driven active learning identifies data samples with elevated aleatoric uncertainty. Entropy [61], which estimates uncertainty, serves as a commonly employed baseline in active learning selection. This function calculates per-pixel entropy for the predicted output and utilizes the averaged entropy as the final score. BALD [62] frequently serves as a baseline in previous studies. It is applied in segmentation by integrating dropout layers into the decoder module of the segmentation model and subsequently computing pixel-wise mutual information through multiple forwards passes. Kampffmeyer et al. [63] strive to optimize the average standard deviation of the predicted probabilities. Jain et al. [64] integrate metrics, defined by manually engineered heuristics, to promote the diversity and representativeness of labeled samples. Certain methodologies leverage unsupervised super pixel-based over segmentation [65,66], relying heavily on the precision of the super-pixel segmentation. Others concentrate on foreground-background segmentation of biomedical images [67,68], employing similarly crafted heuristics. Golestaneh et al. [69] focus on self-consistency that uses simple transformations should not change the observation in active learning for semantic segmentation. Mackowiak et al. [70] concentrate on cost-effective strategies, emphasizing that the labeling cost for an image is not uniformly treated across all images.

The advent of DQN marked a significant milestone; however, numerous constraints associated with this algorithm have surfaced, leading to the proposal of various extensions. Double DQN [71] mitigates the overestimation bias of Q-learning [72] by separating the selection and evaluation of the bootstrap action. Prioritized experience replay [34] enhances data efficiency by prioritizing more frequent replay of informative transitions. The dueling network architecture [33] aids in action generalization by independently representing state values and action advantages. Learning from multi-step bootstrap targets, as seen in A3C [73], adjusts the bias-variance trade-off and accelerates the propagation of newly observed rewards to earlier visited states. Noisy DQN [35] introduces stochastic network layers to facilitate exploration. To the best of our knowledge, our work is the first to examine an agent that integrates all the aforementioned components to the problem of active learning for semantic segmentation. Emphasizing recent experience [36] typically refers to assigning greater importance to recent observations and actions when making decisions or updating the learning model. ERE is often driven by the recognition that the environment is non-stationary, implying that the optimal policy might evolve over time.

3. Methods

3.1 Active learning with reinforcement learning for semantic segmentation

Followed by Casanova et al. [32], we use their architecture for training segmentation networks. We frame the active learning problem as a Markov decision process (MDP). The proposed process entails an iterative active learning strategy for enhancing the performance of a segmentation network, denoted as f and parameterized by θ, within a limited labeled sample budget, B. At each iteration, a query network, represented by π and parameterized by φ, selectively picks K regions from the large unlabeled set, U_t. These regions are then submitted to an oracle for labeling, subsequently augmenting the labeled set, $\mathcal{L}_t$. The segmentation network f is trained using the enriched $\mathcal{L}_t$, and its performance is evaluated based on the Intersection-over-Union (IoU) metric. This iterative procedure continues until the designated budget B is attained. By strategically selecting informative regions for labeling, this process optimizes the performance of the segmentation network, thus efficiently leveraging a limited labeled dataset to achieve superior segmentation results. This data-centric approach enables the model to acquire selection strategies purely from past active learning encounters.

In the setting, we employ four distinct data partitions. For training the query network π, we designate a portion of labeled data D_T, utilizing it for multiple iterations of the active learning process to acquire an effective acquisition function that optimizes performance within a B region budget. The evaluation of the query network takes place on a separate data split D_V. Moreover, we utilize a distinct subset D_R to generate the reward signal, which involves evaluating the segmentation network's performance on this set. Additionally, the set D_S (with D_S being not larger than D_T) is utilized for constructing the representation of the current state.

A Markov Decision Process (MDP) is defined as a tuple (S, A, r, T, γ) where S stands for a set of states; A for actions, the composed action, consisting of K sub-actions, relies on the segmentation network, along with the labeled and unlabeled sets. Each sub-action entails requesting the labeling of a particular region; r, $S \times A \rightarrow R$, for the function based on improvement in mean IoU per class of taking an action in a state; T, $S \times A \times S \rightarrow R$, for the state-transition function; and r, for the discount factor implying that a reward obtained in the future is worth a smaller amount than an immediate reward. **Figure 1** describes this training workflow. In our approach, the episode concludes upon reaching the designated budget B for labeled regions. Post-episode termination, we reset the weights of the segmentation network, denoted as f, to the initial weights θ_0, and initiate a new episode. The training process for the query policy π involves the simulation of multiple episodes, with weight updates occurring at each time step through the sampling of transitions $\{(s_t, a_t, r_{t+1}, s_{t+1})\}$from the experience replay buffer ε.

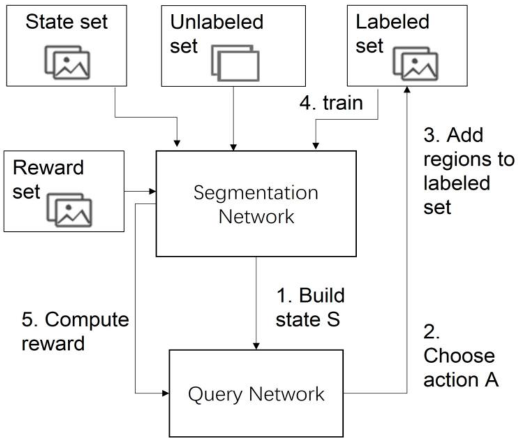

Figure 1. The overall workflow of active learning with reinforcement learning in semantic segmentation.

3.2 Extensions to DQN

The evolution of Deep Q-Networks (DQN) has given rise to several significant extensions, each addressing specific limitations and enhancing the algorithm's overall performance. Individually, each of these algorithms leads to significant performance enhancements. Given that they tackle fundamentally different issues and share a common framework, there is a plausible opportunity for their integration. We suggest six extensions, each designed to overcome a specific limitation, contributing to an overall improvement in performance. To maintain a manageable selection size, we have chosen extensions that address distinct concerns.

Double deep Q-Learning

Following the prior work [32], we set double DQN as our baseline architecture. One issue with the DQN algorithm is that it tends to overestimate the true rewards, leading to inflated Q-values. To address this, the Double DQN algorithm [71] introduces a modification to the Bellman equation used in DQN. Instead of using the same equation, the action selection and action evaluation are decoupled in the following way:

$$Q(s,a;\theta) = r + \gamma Q\left(s^{'}, argmax_{a^{'}}\, Q\left(s^{'}, a^{'}; \theta\right); \theta^{'}\right) \quad (1)$$

Here, the main neural network, θ, determines the best next action, $a^{'}$, while the target network is used to evaluate this action and compute its Q-value. This simple change has been shown to reduce over-estimations and lead to better final policies.

Dueling deep Q-Learning

The Dueling DQN algorithm introduced by Wang et al. [33] seeks to improve upon traditional DQN by decomposing the Q-values into two separate components: the value function, $V(s)$, and the advantage function, A(s, a). The value function represents the expected reward for a given state, s, while the advantage function reflects the relative advantage of taking a particular action, a, compared to other actions. By combining these two functions, it is possible to compute the full Q-values for each state-action pair.

To implement this decomposition, the Dueling DQN algorithm introduces a neural network with two separate output layers, one for the value function and one for the advantage function. These outputs are then combined to produce the final Q-values. This modification allows the network to learn more efficiently in situations where the exact values of individual actions are not as important, as it can focus on learning the value function for the state.

Prioritized experience replay

The proposition by Schaul et al. [34] in 2015 introduces a resolution termed prioritized experience replay (PER). This approach involves the utilization of an added data structure that maintains the priority of each transition. Subsequently, experiences are sampled based on their respective priorities.

$$P(i) = \frac{p_i^{\alpha}}{\sum_k p_k^{\alpha}} \quad (2)$$

The hyperparameter alpha determines the degree of sampling bias desired. The priorities correspond to the temporal difference error of the agent during its most recent training on that particular experience. This strategy enables the agent to emphasize learning from its less accurate predictions, thereby refining its weak areas and significantly improving sample efficiency. New transitions are incorporated into the replay buffer with the highest priority, introducing a bias toward recent transitions. It is essential to recognize that stochastic transitions may also receive preference, even when there is limited remaining knowledge to be gained from them.

Emphasizing recent experience

Primarily conceived to expedite the convergence speed of Soft Actor Critic (SAC) [36], this methodology can conceivably be extended to a wide array of algorithms and tasks that inherently profit from accelerated learning of recent experiences, particularly those involving multiple components. The fundamental concept entails, during the parameter update phase, sampling the initial mini batch from the entire dataset within the replay buffer. Subsequently, for each subsequent mini batch, the sampling range is gradually narrowed, enabling a more pronounced focus on recent data points. This scheme revolves around two fundamental aspects: (i) a heightened sampling frequency for more recent data, and (ii) a systematic arrangement of updates ensuring that older data does not overwrite the more recent ones. The introduction of Experience Replay Emphasis (ERE) establishes a straightforward yet effective sampling technique that enables the agent to prioritize recent transitions without disregarding previously learned policies.

Adaptive epsilon greedy

The epsilon-greedy technique serves as a means to strike a balance between exploration and exploitation in the process of training reinforcement learning policies. For instance, when epsilon is set to 0.3, the output action is randomly chosen from the action space with a probability of 0.3, and with a probability of 0.7, the output action is selected greedily based on argmax (Q).

A refined version of the epsilon-greedy method is referred to as the Adaptive-epsilon-greedy approach [38]. In this approach, for instance, the policy is trained over N epochs/episodes, a value contingent upon the specific problem. Initially, the algorithm sets epsilon to p_{init} (e.g., $p_{\text{init}} = 0.6$), gradually reducing it to reach $\epsilon = p_{\text{end}}$ (e.g., $p_{\text{end}} = 0.1$) over a designated number of training epochs/episodes (n_{step}). Primarily, during the initial training phase, the model is granted increased exploration freedom with a higher probability (e.g., $p_{\text{init}} = 0.6$), followed by a gradual epsilon decrease at a rate r over the training epochs/episodes, adhering to the subsequent formula:

$$r = max\left(\frac{N - n_{\text{step}}}{N}, 0\right) \tag{3}$$

$$\epsilon \leftarrow (p_{\text{init}} - p_{\text{end}})r + p_{\text{end}} \tag{4}$$

This adaptable approach concludes with a notably low exploration probability, p_{end}, after n_{step}, thereby facilitating a transition towards an increased emphasis on exploitation (i.e., a greedier approach) during the latter stages of the training process. Despite this shift, a minimal exploration probability persists, ensuring the ability to explore even as the policy nears convergence.

Noisy network

Noisy networks are often utilized instead of the epsilon-greedy method to promote more effective and dynamic exploration during training. Unlike the epsilon-greedy approach, which only adjusts the exploration probability, noisy networks introduce stochasticity directly into the network's parameters, enabling a more nuanced and continuous exploration process. The Noisy Network [35] introduces a novel concept of a noisy linear layer, integrating both deterministic and noisy components.

$$y = (b + Wx) + \left(b_{\text{noisy}} \odot \epsilon^b + \left(W_{\text{noisy}} \odot \epsilon^w\right)x\right) \tag{5}$$

where ϵ^w and ϵ^b are random variables, and $\odot$ denotes the element-wise product. With time, the network can gradually disregard the noisy stream, albeit at varying rates across distinct regions of the state space, thereby enabling state-specific exploration with a form of intrinsic self-annealing. This dynamic exploration strategy allows for a more fine-grained balance between exploration and exploitation, facilitating improved learning efficiency and adaptability in complex environments.

Soft update for target network

The soft update target network is a key concept in the field of deep reinforcement learning [37]. It refers to a technique used to stabilize and improve the training of deep neural networks in reinforcement learning tasks. Unlike hard updates, which involve periodically copying the parameters of the main network to the target network, soft updates gradually blend the parameters of the target network towards those of the main network. This process helps to mitigate the issue of drastic changes in the target network, which can lead to instability during the learning process. The value of τ is used. In the paper, it proposed an algorithm called DPG. They used τ = 0.001. The target network is updated as follows.

$$\theta^{target} = \theta \times \tau + \theta^{target} \times (1 - \tau) \tag{6}$$

Due to the small value of the parameter τ, the target network smoothly adjusts towards the Q-network's value. To ensure the noticeable impact of this adjustment, frequent updates are required. By employing a soft update strategy, the target network can more smoothly track the changes in the main network, enabling a more stable and effective learning process. This technique has proven to be particularly useful in complex reinforcement learning tasks where maintaining stability during the training phase is crucial for achieving optimal performance.

4. Results

This section describes the setup of our experiments, including the dataset, evaluation methods, and the baselines for comparison.

4.1 Experimental setup

Data collections

The dataset we primarily used in our experiments is the Cambridge-driving Labeled Video Database or CamVid [34] with samples shown in **Figure 2**. This public dataset comprises 360 × 480 street scene color images or frames, each annotated with ground truth semantic labels for pixels across 32 classes. The images were captured from a moving automobile using high-resolution cameras placed on the streets, allowing for the observation of various objects. In this study, our focus is on the segmentation of 11 key classes. Given the urban context, the Road, Building, and Sky classes collectively constitute most frame pixels, accounting for approximately 15.81% and 27.35%, and appearing in almost all frames. Other significant classes, such as Car and Pedestrian, are consistently present throughout the frame sequence but occupy smaller portions, approximately 3.93% and 0.64%, respectively. Additionally, our analysis includes other classes depicted in **Table 1**. The accompanying table illustrates a significant imbalance among the different classes, a challenge that we address in our study. The video sequences were shot during the daytime and at dusk, where the objects in the scene can still be recognized but appear darker than in other sequences. Daylight sequences were captured in sunny weather conditions, featuring mixed urban and residential surroundings.

For a fair comparison between different methods, the segmentation networks of all methods have been pre-trained on the GTA dataset [74], which comprises extensive synthetic images with pixel-level semantic annotations. These images are generated through the open-world video game Grand Theft Auto 5, depicting scenes from a car perspective within virtual cities designed in an American style, like our primary testing dataset introduced earlier.

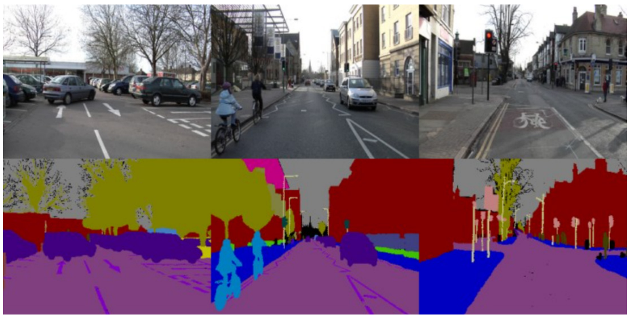

Figure 2. Labeled frames from the video at 1 Hz.

Table 1. Statistics for each class used in this study: "%" shows the ratio of pixels and "Occurrence" shows the number of occurrences over all images.

Class name	Percentage	Occurrence
Road	27.3	701
Building	22.7	687
Sky	15.8	699
Tree	10.4	636
Sidewalk	6.33	672
Car	3.4	643
Column_Pole	0.98	698
Fence	1.43	363
Pedestrian	0.64	640
Bicyclist	0.53	365
Sign_Symbol	0.12	416

Data collection instruments

The CamVid dataset was captured using a digital film camera under fixed conditions without auto zooming, focus, or adjustments during the collecting process. The camera's focus was adjusted to infinity, and both gain and shutter speed were fixed. At the outset, the aperture was widened to its maximum extent while ensuring that white objects in the scene did not become overexposed.

Data analysis

In a real-world scenario where we have unlabeled data, it's possible to involve a human annotator to label the necessary dataset based on active learning recommendations. Nevertheless, in this paper, as a proof of concept, we opted to work with fully labelled data and selectively concealed portions of it to assess the active learning algorithm's performance on the segmentation task.

The dataset we primarily used in our experiments is CamVid [39] discussed earlier. The training, validation, and test sets consist of 370, 104, and 234 images, respectively. In the training set, we used 100 labelled images for building D_T to train the DQN network for several episodes and learn a good acquisition function that maximizes performance with a budget of B regions. A set of 10 images or D_S is used to construct the state representation. For the baseline evaluation set D_V, we utilized 260 images. D_S has a similar class distribution to D_T to represent it. The original validation set was used to build D_R. The dataset's test set was employed to train the model to obtain the final segmentation results. Each image was divided into K regions (in this case, K = 24) with a resolution of 80 × 90. For implementation, we executed 5 different runs with random seeds to calculate the mean and standard deviation. Horizontal flips and random crops of 224 × 224 were applied for data augmentation.

Evaluation

We trained the active learning agent on DT with approximately 0.5 k regions to learn the selection of regions that would improve performance in data-scarce scenarios. Subsequently, we evaluated the model using DV, where the model could access an increasing number of images within different fixed budgets. Once the fixed budget was reached, the segmentation network was trained with LT until it met the early stopping condition in DR. The segmentation network f for all algorithms was pre-trained with the GTA dataset [74], a synthetic dataset, and DT. Finally, we measured the segmentation model's performance on the CamVid test set using the Intersection over Union (IoU) score.

Hardware usage

The models were trained using a single NVIDIA RTX A5000 GPU with 24 GB of VRAM. Training the active learning agents took approximately 18 hours for 5 runs, and training the segmentation models to test the active learning algorithm required a total of 8 hours for 5 runs at each of the 6 budgets.

4.2 Experiment results

In **Figures 3 and 4**, we compare various methods across increasing budgets of labeled 128 × 128 pixel regions. The x-axis, labeled as "Budget", represents the additional number of regions in thousands and the percentage of utilized unlabeled data. The plots include means and standard deviations of 5 runs. The segmentation network utilized in these methods has been pre-trained with the GTA dataset and part of their respective target datasets. The dashed line represents 96% of the best performance (Intersection Over Union) achieved by the segmentation network trained with all available labels. Given that the performance of the preceding work [32] surpasses that of the other baseline models, we will adopt it as the new baseline model for comparisons with other methods.

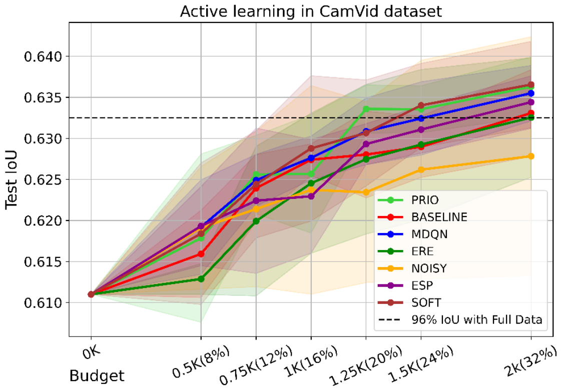

Figure 3. Comparisons of various active learning methods.

In detail, **Figure 3** illustrates the performance of various methods, including Prioritized Experience Replay (PRIO) [34], the reproduced DQN baseline (BASELINE) [32], Dueling Deep Q-network (MDQN) [33], Emphasizing Recent Experiences (ERE) [36], and Noisy Network (NOISY) [35], Adaptive Epsilon Greedy (ESP) [38], Soft Update for Target Network [37]. It's important to note that at 1.5 k regions, the performance of some methods exceeds 96% of the maximum achieved with fully supervised training (having access to all labels). In these experiments, the NOISY model performs the worst, suggesting that acquiring new labels does not provide significant additional information to the model. PRIO and SOFT outperform the other methods, including the baseline, in all budget scenarios, except for the 1K case for the PRIO method. They achieve this without overfitting the training model, while the other methods yield similar results. This suggests that effective active learning, through selective labeling or additional information, can assist the segmentation model in avoiding local minima and achieving better performance.

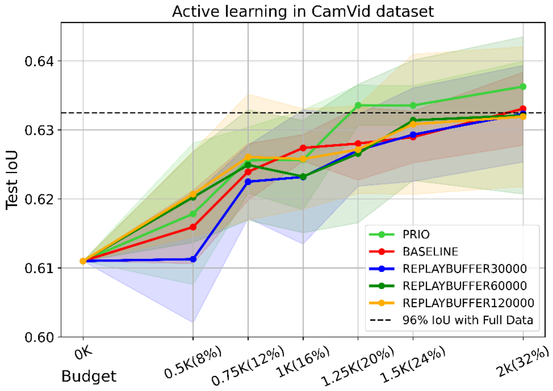

Figure 4. Compare the PRIO method with varying replay buffer sizes.

In **Figure 4**, a comparison is made between the baseline and PRIOR methods, considering different replay buffer pool sizes: 600, 30,000, 60,000, and 120,000. Performance remains relatively stable for both 60,000 and 120,000. Interestingly, PRIOR, despite having a smaller replay buffer (around 600 compared to 30,000, 60,000, and 120,000), outperforms the others by a significant margin.

4.3 Discussion

Implications and suggestions

In this section, we examine the primary findings from the experiments. As the experiment results show, implementing a soft target network update can improve performance. This is due to the smoother tracking of changes in the main network by the target network through the use of soft updates. Consequently, a more stable learning process is achieved, and the algorithm is able to converge to a better policy, resulting in improved performance reflected in better results and more reliable Q-value estimations. The Dueling DQN algorithm improves upon earlier models by decoupling value and advantage functions in Q-value estimation. This allows for better recognition of action importance in varying states, thereby enhancing learning and generalization. The resulting architecture enables more effective comprehension of state value and action advantage, leading to improved action selection and overall performance. Moreover, this division reduces the variability in learned action values, stabilizes learning, and provides more precise estimates. This solves the problems of overestimation or underestimation of Q-values. Prioritized Experience Replay (PER) improves reinforcement learning by enhancing sample efficiency, stabilizing the learning process, and promoting effective exploration of the state space. PER prioritizes experiences based on higher probabilities for transitions with greater TD errors, accelerating convergence and fostering efficient learning. Emphasizing rare events also aids agents in handling critical scenarios adeptly. The learning process stability is ensured by policy updates that efficiently lead to rapid learning and enhanced convergence. Emphasizing transitions with high learning potential aids the thorough exploration of the state space, resulting in improved overall performance and decision-making by the agent. PER usage reduces bias from uniform sampling and mitigates high variance issues, resulting in more accurate and stable Q-value updates. This enhances the learning process and improves performance in various reinforcement learning tasks.

Our experiments showed that certain techniques, like incorporating noisy networks to encourage diverse segmentation strategies, had a negative impact on segmentation performance by introducing instability in the learning process, resulting in decreased accuracy in certain scenarios. In intricate settings with sparse rewards or high-dimensional state spaces, challenges arise where adjusting the exploration rate alone may prove ineffective. Poor adjustment of the exploration rate adaptation and disregard for specific learning dynamics can disturb the balance between exploration and exploitation, despite the use of adaptable epsilon-greedy approaches, creating difficulties in obtaining the ideal exploration-exploitation equilibrium, especially in some settings. Overemphasizing recent experiences in a DQN can hinder reinforcement learning performance by reducing sample efficiency, impeding generalization, introducing increased variance, and destabilizing the learning process. It also limits exploration across the state space, restricting the discovery of critical, infrequently encountered states and thereby compromising the agent's convergence to an optimal policy. Thus, in order to improve performance, it is crucial to achieve a balance between prioritizing recent experiences and maintaining a varied set of samples that support efficient learning and exploration across the entire state space.

Limitations and future work

We have illustrated the successful integration of various enhancements into the DQN, enhancing the semantic segmentation model to achieve state-of-the-art performance. Furthermore, our findings indicate that certain components within the integrated algorithm yield distinct performance advantages. While numerous algorithmic components couldn't be incorporated in this study, they stand as promising candidates for future experiments involving integrated agents. Below, we discuss several of these potential candidates.

Policy-based methods directly parameterize the policy, allowing for more flexible and complex policies compared to value-based approaches like DQN. Investigating the application of policy-based methods to active learning in semantic segmentation could provide valuable insights into optimizing decision-making strategies. Actor-critic methods combine the strengths of both policy and value-based approaches by maintaining separate networks for policy and value estimation. Exploring the integration of actor-critic methods in our active learning framework may offer advantages in terms of stability and efficiency. N-step methods extend the traditional DQN by incorporating multiple consecutive states and actions. Evaluating the impact of N-step methods on active learning performance in semantic segmentation could enhance our understanding of the temporal dynamics involved. Distributional RL models the distribution of returns rather than focusing solely on expected values. Introducing distributional RL techniques into our framework may contribute to a more nuanced understanding of uncertainty and risk management in the active learning process. Imitation learning leverages expert demonstrations to guide the learning process. Integrating imitation learning into active learning for semantic segmentation could offer a valuable mechanism for initializing the model and accelerating the learning curve.

The exploration of these alternative RL techniques represents a promising direction for future research. Investigating how those methods can be tailored to the specific challenges of active learning in semantic segmentation is essential. The strengths of different RL paradigms could be harnessed by exploring the combination of these techniques in hybrid models or ensembles. Moreover, the transferability and generalization of learned policies across diverse datasets and domains require attention in future investigations.

5. Conclusions

In conclusion, our study provides a comprehensive comparison of different DQN extensions designed to improve active learning in semantic segmentation through reinforcement learning. Our primary objective is to mitigate the labour-intensive task of obtaining pixel-wise labels with human intervention. Our results demonstrate that the NOISY model performs the worst, showing rapid overfitting, therefore implying that acquiring new labels does not significantly enhance the model's information. Notably, prioritized experience replay and soft update outperform all other methods, including the baseline, in all budget scenarios. Importantly, these methods achieve superior performance without overfitting, while the other techniques yield similar results. Additionally, the comparison of the baseline and PER methods, considering different replay buffer pool sizes, indicates that PRIOR outperforms others despite having a smaller replay buffer. This emphasizes the importance of utilizing information effectively to improve the segmentation process. This highlights the effectiveness of selective labelling or including additional information to help the segmentation model avoid local minimum and achieve better performance. These findings highlight the prospect of utilizing sophisticated DQN extensions to enhance active learning in semantic segmentation, resulting in streamlined label acquisition and improved model performance.

Author Contributions

Conceptualization, D.H. and S.C.; methodology, D.H. and S.C.; investigation, D.H.; validation, D.H. and P.H.; data curation, D.H. and P.H.; writing, D.H. and P.H.; review and editing, S.C.; supervision, S.C.

Conflict of Interest

There is no conflict of interest.

Funding

This work is partially supported by the Vice President for Research and Partnerships of the University of Oklahoma, the Data Institute for Societal Challenges, and the Stephenson Cancer Center through DISC/SCC Seed Grant Award.

References

[1] Szeliski, R., 2022. Computer vision: Algorithms and applications. Springer Nature: Berlin.

[2] Li, H., Xiong, P., Fan, H., et al., 2019. DFANet: Deep Feature Aggregation for Real-Time Semantic Segmentation [Internet]. Available from: https://openaccess.thecvf.com/content_CVPR_2019/papers/Li_DFANet_Deep_Feature_Aggregation_for_Real-Time_Semantic_Segmentation_CVPR_2019_paper.pdf

[3] Liu, M., Yin, H., 2019. Feature pyramid encoding network for real-time semantic segmentation. arXiv preprint arXiv:1909.08599.
DOI: https://doi.org/10.48550/arXiv.1909.08599

[4] Li, X., You, A., Zhu, Z., et al. (editors), 2020. Semantic flow for fast and accurate scene parsing. Computer Vision-ECCV 2020: 16th European Conference; 2020 Aug 23-38; Glasgow, UK. Cham: Springer International Publishing. p. 775-793.
DOI: https://doi.org/10.1007/978-3-030-58452-8_45

[5] Yang, X., Wu, Y., Zhao, J., et al., 2020. Dense Dual-Path Network for Real-time Semantic Segmentation [Internet]. Available from: https://openaccess.thecvf.com/content/ACCV2020/papers/Yang_Dense_Dual-Path_Network_for_Real-time_Semantic_Segmentation_ACCV_2020_paper.pdf

[6] Orsic, M., Kreso, I., Bevandic, P., et al., 2019. In Defense of Pre-Trained ImageNet Architectures for Real-Time Semantic Segmentation of Road-Driving Images [Internet]. Available from: https://openaccess.thecvf.com/content_CVPR_2019/papers/Orsic_In_Defense_of_Pre-Trained_ImageNet_Architectures_for_Real-Time_Semantic_Segmentation_CVPR_2019_paper.pdf

[7] Zhang, H., Tang, W., Na, W., et al., 2020. Implementation of generative adversarial network-CLS combined with bidirectional long short-term memory for lithium-ion battery state prediction. Journal of Energy Storage. 31, 101489.
DOI: https://doi.org/10.1016/j.est.2020.101489

[8] Zhang, H., Na, W., Kim, J. (editors), 2018. State-of-charge estimation of the lithium-ion battery using neural network based on an improved thevenin circuit model. 2018 IEEE Transportation Electrification Conference and Expo (ITEC); 2018 Jun 13-15; Long Beach, CA, USA. New York: IEEE. p. 342-346.
DOI: https://doi.org/10.1109/ITEC.2018.8450162

[9] Zhang, H., Cheng, S., El Amm, C., et al., 2023. Efficient pooling operator for 3D morphable models. IEEE Transactions on Visualization and Computer Graphics. 1-9.
DOI: https://doi.org/10.1109/TVCG.2023.3255820

[10] Han, D., Wang, S., Jiang, C., et al., 2015. Trends in biomedical informatics: Automated topic analysis of JAMIA articles. Journal of the American Medical Informatics Association. 22(6), 1153-1163.
DOI: https://doi.org/10.1093/jamia/ocv157

[11] Han, D., Mulyana, B., Stankovic, V., et al., 2023. A survey on deep reinforcement learning algorithms for robotic manipulation. Sensors. 23(7), 3762.
DOI: https://doi.org/10.3390/s23073762

[12] Sze, V., Chen, Y.H., Yang, T.J., et al., 2017. Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE. 105(12), 2295-2329.
DOI: https://doi.org/10.1109/JPROC.2017.2761740

[13] Wang, W., Fu, Y., Pan, Z., et al., 2020. Real-time driving scene semantic segmentation. IEEE Access. 8, 36776-36788.
DOI: https://doi.org/10.1109/ACCESS.2020.2975640

[14] Papadeas, I., Tsochatzidis, L., Amanatiadis, A., et al., 2021. Real-time semantic image segmentation with deep learning for autonomous driving: A survey. Applied Sciences. 11(19), 8802.
DOI: https://doi.org/10.3390/app11198802

[15] Mahe, H., Marraud, D., Comport, A.I. (editors), 2019. Real-time rgb-d semantic keyframe slam based on image segmentation learning from industrial cad models. 2019 19th International Conference on Advanced Robotics (ICAR); 2019 Dec 2-6; Belo Horizonte, Brazil. New York: IEEE. p. 147-154.
DOI: https://doi.org/10.1109/ICAR46387.2019.8981549

[16] Bruce, J., Balch, T., Veloso, M. (editors), 2000. Fast and inexpensive color image segmentation for interactive robots. Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000) (Cat. No. 00CH37113); 2000 Oct 31-Nov 5; Takamatsu, Japan. New York: IEEE. p. 2061-2066.
DOI: https://doi.org/10.1109/IROS.2000.895274

[17] Du, X., Nie, Y., Wang, F., et al., 2022. AL-Net: Asymmetric lightweight network for medical image segmentation. Frontiers in Signal Processing. 2, 842925.
DOI: https://doi.org/10.3389/frsip.2022.842925

[18] Lou, A., Guan, S., Loew, M., 2023. Cfpnet-m: A light-weight encoder-decoder based network for multimodal biomedical image real-time segmentation. Computers in Biology and Medicine. 154, 106579.
DOI: https://doi.org/10.1016/j.compbiomed.2023.106579

[19] Cordts, M., Omran, M., Ramos, S., et al., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding [Internet]. Available from: https://openaccess.thecvf.com/content_cvpr_2016/papers/Cordts_The_Cityscapes_Dataset_CVPR_2016_paper.pdf

[20] Gal, Y., Islam, R., Ghahramani, Z., 2017. Deep Bayesian Active Learning with Image Data [Internet]. Available from: https://proceedings.mlr.press/v70/gal17a/gal17a.pdf

[21] Chu, H.M., Lin, H.T. (editors), 2016. Can active learning experience be transferred? 2016 IEEE 16th International Conference on Data Mining (ICDM); 2016 Dec 12-15; Barcelona, Spain. New York: IEEE. p. 841-846.
DOI: https://doi.org/10.1109/ICDM.2016.0100

[22] Hsu, W.N., Lin, H.T., 2015. Active learning by learning. Proceedings of the AAAI Conference on Artificial Intelligence. 29(1).
DOI: https://doi.org/10.1609/aaai.v29i1.9597

[23] Yoo, D., Kweon, I.S., 2019. Learning Loss for Active Learning [Internet]. Available from: https://openaccess.thecvf.com/content_CVPR_2019/papers/Yoo_Learning_Loss_for_Active_Learning_CVPR_2019_paper.pdf

[24] Lookman, T., Balachandran, P.V., Xue, D., et al., 2019. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials. 5(1), 21.
DOI: https://doi.org/10.1038/s41524-019-0153-8

[25] Fasel, U., Kutz, J.N., Brunton, B.W., et al., 2022. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. Proceedings of the Royal Society A. 478(2260), 20210904.
DOI: https://doi.org/10.1098/rspa.2021.0904

[26] Hu, Z., Bai, X., Zhang, R., et al., 2022. Lidal: Inter-frame uncertainty based active learning for 3d lidar semantic segmentation. European

Conference on Computer Vision. 13687, 248-265.
DOI: https://doi.org/10.1007/978-3-031-19812-0_15

[27] Lenczner, G., Chan-Hon-Tong, A., Le Saux, B., et al., 2022. Dial: Deep interactive and active learning for semantic segmentation in remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 15, 3376-3389.
DOI: https://doi.org/10.1109/JSTARS.2022.3166551

[28] Xie, B., Yuan, L., Li, S., et al., 2022. Towards Fewer Annotations: Active Learning via Region Impurity and Prediction Uncertainty for Domain Adaptive Semantic Segmentation [Internet]. Available from: https://openaccess.thecvf.com/content/CVPR2022/papers/Xie_Towards_Fewer_Annotations_Active_Learning_via_Region_Impurity_and_Prediction_CVPR_2022_paper.pdf

[29] Dhiman, G., Kumar, A.V., Nirmalan, R., et al., 2023. Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. Multimedia Tools and Applications. 82(4), 5343-5367.
DOI: https://doi.org/10.1007/s11042-022-12178-7

[30] Zhou, W., Li, J., Zhang, Q., 2022. Joint communication and action learning in multi-target tracking of UAV swarms with deep reinforcement learning. Drones. 6(11), 339.
DOI: https://doi.org/10.3390/drones6110339

[31] Hu, M., Zhang, J., Matkovic, L., et al., 2023. Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions. Journal of Applied Clinical Medical Physics. 24(2), e13898.
DOI: https://doi.org/10.1002/acm2.13898

[32] Casanova, A., Pinheiro, P.O., Rostamzadeh, N., et al., 2020. Reinforced active learning for image segmentation. arXiv preprint arXiv:2002.06583.
DOI: https://doi.org/10.48550/arXiv.2002.06583

[33] Wang, Z., Schaul, T., Hessel, M., et al., 2016. Dueling Network Architectures for Deep Reinforcement Learning [Internet]. Available from: https://proceedings.mlr.press/v48/wangf16.pdf

[34] Schaul, T., Quan, J., Antonoglou, I., et al., 2015. Prioritized experience replay. arXiv preprint arXiv:1511.05952.
DOI: https://doi.org/10.48550/arXiv.1511.05952

[35] Fortunato, M., Azar, M.G., Piot, B., et al., 2017. Noisy networks for exploration. arXiv preprint arXiv:1706.10295.
DOI: https://doi.org/10.48550/arXiv.1706.10295

[36] Wang, C., Ross, K., 2019. Boosting soft actor-critic: Emphasizing recent experience without forgetting the past. arXiv preprint arXiv:1906.04009.
DOI: https://doi.org/10.48550/arXiv.1906.04009

[37] Lillicrap, T.P., Hunt, J.J., Pritzel, A., et al., 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.
DOI: https://doi.org/10.48550/arXiv.1509.02971

[38] Tokic, M., Palm, G., 2011. Value-difference based exploration: Adaptive control between epsilon-greedy and softmax. Annual conference on artificial intelligence. Springer: Berlin. pp. 335-346.
DOI: https://doi.org/10.1007/978-3-642-24455-1_33

[39] Brostow, G.J., Shotton, J., Fauqueur, J., et al. (editors), 2008. Segmentation and recognition using structure from motion point clouds. Computer Vision-ECCV 2008: 10th European Conference on Computer Vision; 2008 Oct 12-18; Marseille, France. Berlin: Springer. p. 44-57.
DOI: https://doi.org/10.1007/978-3-540-88682-2_5

[40] Angluin, D., 1988. Queries and concept learning. Machine Learning. 2, 319-342.
DOI: https://doi.org/10.1023/A:1022821128753

[41] King, R.D., Whelan, K.E., Jones, F.M., et al., 2004. Functional genomic hypothesis generation and experimentation by a robot scientist. Nature. 427(6971), 247-252.
DOI: https://doi.org/10.1038/nature02236

[42] Dagan, I., Engelson, S.P. (editors), 1995. Committee-based sampling for training probabilis-

tic classifiers. Machine Learning Proceedings 1995; 1995 Jul 9-12; Tahoe, California. p. 150-157.
DOI: https://doi.org/10.1016/B978-1-55860-377-6.50027-X

[43] Krishnamurthy, V., 2002. Algorithms for optimal scheduling and management of hidden Markov model sensors. IEEE Transactions on Signal Processing. 50(6), 1382-1397.
DOI: https://doi.org/10.1109/TSP.2002.1003062

[44] Lewis, D.D., 1995. A Sequential Algorithm for Training Text Classifiers: Corrigendum and Additional Data [Internet]. Available from: https://dl.acm.org/doi/pdf/10.1145/219587.219592

[45] Ren, P., Xiao, Y., Chang, X., et al., 2021. A survey of deep active learning. ACM Computing Surveys (CSUR). 54(9), 1-40.
DOI: https://doi.org/10.1145/3472291

[46] Shui, C., Zhou, F., Gagné, C., et al., 2020. Deep Active Learning: Unified and Principled Method for Query and Training [Internet]. Available from: https://proceedings.mlr.press/v108/shui20a/shui20a.pdf

[47] Settles, B., 2012. Active learning, volume 6 of synthesis lectures on artificial intelligence and machine learning. Morgan & Claypool.

[48] Settles, B., 2011. From Theories to Queries: Active Learning in Practice [Internet]. Available from: https://proceedings.mlr.press/v16/settles11a/settles11a.pdf

[49] Hernández-Lobato, J.M., Adams, R., 2015. Probabilistic Backpropagation for Scalable Learning of Bayesian Neural Networks [Internet]. Available from: http://proceedings.mlr.press/v37/hernandez-lobatoc15.pdf

[50] Osugi, T., Kim, D., Scott, S. (editors), 2005. Balancing exploration and exploitation: A new algorithm for active machine learning. Fifth IEEE International Conference on Data Mining (ICDM'05); 2005 Nov 27-30; Houston, TX, USA. New York: IEEE.
DOI: https://doi.org/10.1109/ICDM.2005.33

[51] Long, C., Hua, G., 2015. Multi-Class Multi-Annotator Active Learning with Robust Gaussian Process for Visual Recognition [Internet]. Available from: https://openaccess.thecvf.com/content_iccv_2015/papers/Long_Multi-Class_Multi-Annotator_Active_ICCV_2015_paper.pdf

[52] Gong, J., Fan, Z., Ke, Q., et al., 2022. Meta Agent Teaming Active Learning for Pose Estimation [Internet]. Available from: https://openaccess.thecvf.com/content/CVPR2022/papers/Gong_Meta_Agent_Teaming_Active_Learning_for_Pose_Estimation_CVPR_2022_paper.pdf

[53] Sadigh, D., Dragan, A.D., Sastry, S., et al., 2017. Active preference-based learning of reward functions. UC Berkeley: Berkeley.
DOI: https://doi.org/10.15607/rss.2017.xiii.053

[54] Kunapuli, G., Odom, P., Shavlik, J.W., et al. (editors), 2013. Guiding autonomous agents to better behaviors through human advice. 2013 IEEE 13th International Conference on Data Mining; 2013 Dec 7-10; Dallas, TX, USA. New York: IEEE. p. 409-418.
DOI: https://doi.org/10.1109/ICDM.2013.79

[55] Ezzeddine, A., Mourad, N., Araabi, B.N., et al., 2018. Combination of learning from non-optimal demonstrations and feedbacks using inverse reinforcement learning and Bayesian policy improvement. Expert Systems with Applications. 112, 331-341.
DOI: https://doi.org/10.1016/j.eswa.2018.06.035

[56] Liu, M., Buntine, W., Haffari, G. (editors), 2018. Learning how to actively learn: A deep imitation learning approach. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2018 Jul 15-20; Melbourne, Australia. p. 1874-1883.
DOI: https://doi.org/10.18653/v1/P18-1174

[57] Pang, K., Dong, M., Wu, Y., et al., 2018. Meta-learning transferable active learning policies by deep reinforcement learning. arXiv preprint arXiv:1806.04798.
DOI: https://doi.org/10.48550/arXiv.1806.04798

[58] Contardo, G., Denoyer, L., Artières, T., 2017.

A meta-learning approach to one-step active learning. arXiv preprint arXiv:1706.08334. DOI: https://doi.org/10.48550/arXiv.1706.08334

[59] Sener, O., Savarese, S., 2017. Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489. DOI: https://doi.org/10.48550/arXiv.1708.00489

[60] Mittal, S., Niemeijer, J., Schäfer, J.P., et al., 2023. Revisiting deep active learning for semantic segmentation. arXiv preprint arXiv:2302.04075. DOI: https://doi.org/10.48550/arXiv.2302.04075

[61] Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal. 27(3), 379-423. DOI: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[62] Houlsby, N., Huszár, F., Ghahramani, Z., et al., 2011. Bayesian active learning for classification and preference learning. arXiv preprint arXiv:1112.5745. DOI: https://doi.org/10.48550/arXiv.1112.5745

[63] Kampffmeyer, M., Salberg, A.B., Jenssen, R., 2016. Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks [Internet]. Available from: https://www.cv-foundation.org//openaccess/content_cvpr_2016_workshops/w19/papers/Kampffmeyer_Semantic_Segmentation_of_CVPR_2016_paper.pdf

[64] Jain, S.D., Grauman, K., 2016. Active Image Segmentation Propagation [Internet]. Available from: https://openaccess.thecvf.com/content_cvpr_2016/papers/Jain_Active_Image_Segmentation_CVPR_2016_paper.pdf

[65] Vezhnevets, A., Buhmann, J.M., Ferrari, V. (editors), 2012. Active learning for semantic segmentation with expected change. 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16-21; Providence, RI, USA. New York: IEEE. p. 3162-3169. DOI: https://doi.org/10.1109/CVPR.2012.6248050

[66] Konyushkova, K., Sznitman, R., Fua, P., 2015. Introducing Geometry in Active Learning for Image Segmentation [Internet]. Available from: https://openaccess.thecvf.com/content_iccv_2015/papers/Konyushkova_Introducing_Geometry_in_ICCV_2015_paper.pdf

[67] Aklilu, J., Yeung, S., 2022. ALGES: Active Learning with Gradient Embeddings for Semantic Segmentation of Laparoscopic Surgical Images [Internet]. Available from: https://proceedings.mlr.press/v182/aklilu22a

[68] Shu, X., Yang, Y., Xie, R., et al., 2022. ALS: Active learning-based image segmentation model for skin lesion. DOI: http://dx.doi.org/10.2139/ssrn.4141765

[69] Golestaneh, S.A., Kitani, K.M., 2020. Importance of self-consistency in active learning for semantic segmentation. arXiv preprint arXiv:2008.01860. DOI: https://doi.org/10.48550/arXiv.2008.01860

[70] Mackowiak, R., Lenz, P., Ghori, O., et al., 2018. Cereals-cost-effective region-based active learning for semantic segmentation. arXiv preprint arXiv:1810.09726. DOI: https://doi.org/10.48550/arXiv.1810.09726

[71] Van Hasselt, H., Guez, A., Silver, D., 2016. Deep reinforcement learning with double q-learning. Proceedings of the AAAI Conference on Artificial Intelligence. 30(1). DOI: https://doi.org/10.1609/aaai.v30i1.10295

[72] Hasselt, H., 2010. Double Q-learning. Advances in Neural Information Processing Systems. 23, 2613-2621.

[73] Babaeizadeh, M., Frosio, I., Tyree, S., et al., 2016. Reinforcement learning through asynchronous advantage actor-critic on a gpu. arXiv preprint arXiv:1611.06256. DOI: https://doi.org/10.48550/arXiv.1611.06256

[74] Richter, S.R., Vineet, V., Roth, S., et al. (editors), 2016. Playing for data: Ground truth from computer games. Computer Vision-ECCV 2016: 14th European Conference; 2016 Oct 11-14; Amsterdam, The Netherlands. Cham: Springer International Publishing. p. 102-118. DOI: https://doi.org/10.1007/978-3-319-46475-6_7