SciencePG
Science Publishing Group

Research Article

An Active Learning Semantic Segmentation Model Based on an Improved Double Deep Q-Network

Yan Yu[*] (iD)

North China University of Technology, Brunel London School, Beijing, China

Abstract

Image semantic segmentation is essential in fields such as computer vision, autonomous driving, and human-computer interaction due to its ability to accurately identify and classify each pixel in an image. However, this task is fraught with challenges, including the difficulty of obtaining detailed pixel labels and the problem of class imbalance in segmentation datasets. These challenges can hinder the effectiveness and efficiency of segmentation models. To address these issues, we propose an active learning semantic segmentation model named CG_D3QN, which is designed and implemented based on an enhanced Double Deep Q-Network (D3QN). The proposed CG_D3QN model incorporates a hybrid network structure that combines a dueling network architecture with Gated Recurrent Units (GRUs). This novel approach improves policy evaluation accuracy and computational efficiency by mitigating a Q-value overestimation and making better use of historical state information. Our experiments, conducted on the CamVid and Cityscapes datasets, reveal that the CG_D3QN model significantly reduces the number of required sample annotations by 65.0% compared to traditional methods. Additionally, it enhances the mean Intersection over Union (IoU) for underrepresented categories by approximately 1% to 3%. These results highlight the model's effectiveness in lowering annotation costs, addressing class imbalance, and its versatility across different segmentation networks.

Keywords

Deep Reinforcement Learning, Active Learning, Image Semantic Segmentation, Dueling Network, Gated Recurrent Unit (GRU)

1. Introduction

Semantic Segmentation, as a pixel-level image semantic label classification task, aims to classify each pixel in an image to achieve accurate segmentation of semantic regions in the image [1]. Semantic segmentation plays a crucial role in popular AI fields such as autonomous driving [2], medical imaging, and augmented reality. However, current research faces two major limitations. First, compared to other image classification tasks, semantic segmentation datasets require a large number of high-quality pixel labels. This is particularly true in fields with high annotation thresholds, such as medical imaging and defense, where annotators need a high level of expertise. Additionally, manual annotation is prone to errors, and the annotation process is time-consuming and labor-intensive. Second, the results of semantic segmentation may be affected by data imbalance in the segmentation dataset. For instance, in medical imaging [3, 4], there is often a significant bias in the age and gender distribution of data samples, which can lead to the model's performance being skewed

toward the more prevalent classes. Given the challenges of large-scale data and high annotation costs faced by many semantic segmentation studies, active learning has garnered attention as a method to reduce the model's dependency on data [5].

Active Learning (AL) [6], also known as query learning or optimal experimental design, is centered on the idea of adaptively selecting the most informative samples for annotation and training to reduce annotation costs without compromising model performance. Active learning methods can be broadly categorized into two types: traditional hand-crafted heuristic methods [7, 8], which select strategies generally tailored to specific research goals or datasets, designed by experts based on their knowledge or approximate theoretical criteria, and data-driven methods [9, 10], which are built upon prior active learning experiences and trained using labeled data to develop active learning strategies. Since the process of active learning can be simulated as a sequential decision-making process—learning to make a series of decisions through interaction with the environment—Reinforcement Learning (RL) [11] offers the possibility of training active learning query strategies.

Currently, compared to image classification [12], research on active learning for semantic segmentation is relatively scarce. Traditional active learning methods for semantic segmentation mainly rely on hand-crafted heuristic approaches, with the most basic active learning algorithm being the random sampling strategy (Random), which selects samples from the unlabeled pool randomly for annotation. Cai et al. [13] proposed a cost-sensitive acquisition function based on labeled image regions; however, in practical applications, this information is not static, which limits its applicability. Mackowiak et al. [14] introduced an active learning algorithm for handling large sample segmentation datasets, which is region-based and does not consider the cost of image labeling. Gal et al. [15] proposed a decision uncertainty-based active learning method (BALD) that uses Bayesian Convolutional Neural Networks for active learning. Although the aforementioned methods have made some progress in addressing semantic segmentation, they are tailored to specific datasets, which limits the generalization and robustness of the models.

With the development of the deep learning field, deep neural networks have been introduced into the field of reinforcement learning, giving rise to Deep Reinforcement Learning (DRL). Current active learning methods using reinforcement learning typically adopt a strategy of annotating one sample at a time [16, 17] until the sample budget is reached. However, when dealing with large-scale semantic segmentation datasets, re-training the segmentation network and recalculating the corresponding rewards after each annotation results in low efficiency. Sener et al. [18] proposed an active learning algorithm based on core-set selection, which incrementally selects a batch of representative samples, improving annotation efficiency. Dhiman et al. [19]

combined DRL, active learning, and recurrent neural networks (RNN) to propose an automatic annotation model for streaming applications, enhancing retrieval accuracy and performance. Chan et al. [20] reduced the impact on the segmentation network by weighting posterior and prior class probabilities. Casanova et al. [21] proposed a reinforcement learning-based active learning method (Rails), a general approach to discovering active learning strategies from data, but it still faces the challenge of label class imbalance during the active learning process.

To address the aforementioned issues, this paper proposes a data-driven active learning semantic segmentation method, which selects and requests labels for the most relevant regions from an unlabeled image set, enabling the training of a high-performance segmentation network with only a small number of annotated pixel label samples. The main contributions are as follows:

a) Proposed Model: We introduce an active learning semantic segmentation model based on an improved Double Deep Q-Network, transforming the pool-based active learning process into a Markov decision process. The model selects critical image regions rather than entire images, improving information extraction.

b) Q-Value Overestimation and Imbalance: To address Q-value overestimation and class imbalance issues, we incorporate a Dueling Double Deep Q-Network (Dueling DDQN) and a hybrid CNN-GRU network structure, enhancing the model's robustness and performance.

c) Performance Evaluation: Evaluations on CamVid and Cityscapes datasets demonstrate that our model requests more annotations for less frequent classes, improving efficiency and addressing class imbalance. The model also outperforms original semantic segmentation methods when combined with the latest segmentation networks.

2. Methods

2.1. Problem Definition

"Given k unlabeled samples, they are placed into an unlabeled sample pool U_t. The active learning semantic segmentation method selects sample regions from U for annotation while simultaneously learning a query network that serves as a discriminative method for selecting regions to annotate. The annotated samples are then placed into the labeled sample pool L_t, and the semantic segmentation model is trained using the samples in L_t, iterating until the annotation budget B is reached. To reduce the impact of annotationbudget and class imbalance, the sample selection strategy is crucial.

To address this, this paper proposes a CG_D3QN active learning semantic segmentation model, which transforms the active learning semantic segmentation problem into a Mar-

kov Decision Process (MDP), represented by the tuple (S, A, R, S_{t+1}), defined as follows:

a) State Set S: Represents a set of state values. For each state $s_t \in S$, the agent selectswhich sample regions to label from U_t by performing an action $a_t \in A$.

b) Action Set A: Represents the action $a_t \in \{a_1, \dots, a_n\}$, composed of n sub-actions, each representing labeling a region on a sample, determined based on the semantic segmentation network, the labeled sample pool L_t, and the unlabeled sample pool U_t.

c) Reward Set R: Represents the reward value $r_{t+1} \in R$ obtained after each active learning iteration, calculated based on the difference in performance of the segmentation network on DR between the current and previous rounds. Here, DR is a separate subset of datasamples used to evaluate the performance of the segmentation network."

d) State Set S_{t+1}: Represents the state value at the next time step.

The model adopts a pool-based active learning process framework as the overall architecture, with Feature Pyramid Networks (FPN) [24] used as the semantic segmentation network. The process framework is shown in Figure 1. The query network is modeled as the reinforcement learning agent, while the other components are modeled as the reinforcement learning environment. The state subset DS includes data samples from all classes, representing a representative subset of the entire dataset. During training, the agent obtains state and action representations from the environment and trains the query network using a reinforcement learning model and samples from the experience buffer. The query network selects an action a_t and adds the annotated region to the labeled sample pool. These mantic segmentation network FPN is updated and the reward is calculated, with iterative training continuing until the annotation budget is reached.
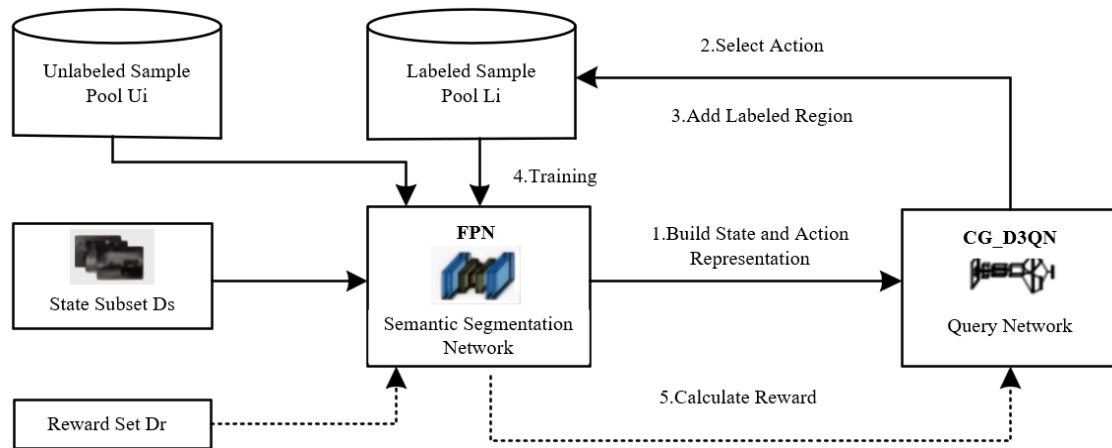

Figure 1. *Active learning semantic segmentation workflow framework.*

2.2. Construct State Representation and Action Representation for Semantic Segmentation

"Since semantic segmentation is a pixel-level semantic label classification task, to avoid consuming a large amount of memory, the state representation for reinforcement learning is constructed using a state subset DS. The samples in DS are divided into multiple patches and feature vectors are calculated for all patches. During the construction of the state representation, first, the information entropy at each pixel position within the image sample regions of the state subset is calculated. Three pooling operations maximum, minimum and average are then applied to the entropy values to downsample them, generating the first set of feature vectors. Next, the segmentation network is used to predict the number of pixels for each class, and these predicted values are normalized to form the second set of feature vectors. Finally, the

two sets of feature vectors for each sample region are concatenated to encode the state s_t.

In the active learning semantic segmentation process, action representation involves labeling the unlabeled regions pixel by pixel. However, each action request requires calculating the features for every region in the unlabeled samples, which incurs a high computational cost. To address this issue, during the construction of the action representation, at each time step t, n unlabeled regions are uniformly sampled from the unlabeled sample pool to form a region pool n_p^t, which approximately represents the entire set of unlabeled samples. Then, a candidate region x_t is selected from the region pool n_p^t, and the normalized count of predicted pixels for each class is calculated. Subsequently, the KL divergence between the class distributions predicted by the segmentation network for the labeled and unlabeled regions is computed, forming two sets of feature vectors. Finally, these vectors are concatenated with the state representation to form the action repre-

sentation n_a^t. The state and action representations for se-
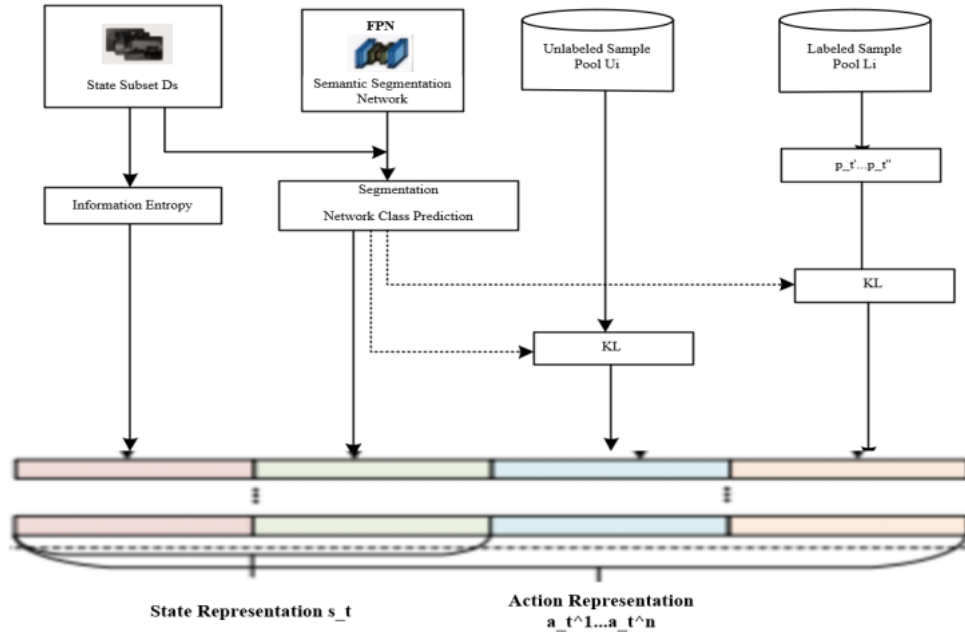
mantic segmentation are illustrated in Figure 2.

Figure 2. State representation and action representation.

2.3. Web Framework

2.3.1. Dual Deep Q Network

Double Deep Q Network (DDQN) is an improved algorithm of the Deep Q Network (DQN). The primary difference between DDON and DON lies in its use of double Q-learning to find the optimal policy, By decoupling the action selection and value evaluation processes for the target Q-value, DDQN aims to eliminate overestimation bias. The DDON algorithm uses deep convolutional neural networks to approximate the state-action value function:

$$(s_t, a_t) = Q(s_t, a_t; \theta) \tag{1}$$

where θ represents the parameters of the main network. The network takes the state sample s_t, and the action a_t, under that state as input and then outputs the corresponding Q-value. During training, the action that yields the maximum Q-value produced by the main network is selected, and this action is input into the target network to evaluate the state-action value function:

$$y_t = r_{\{t+1\}} +$$

$$\gamma \cdot Q(s_{\{t+1\}}, \{argmax\}_{\{a_{\{t+1\}}\}} Q(s_{\{t+1\}}, a_{\{t+1\}}; \theta); \theta^-) \tag{2}$$

Where θ^- represents the parameters of the target network, r_{t+1} is the immediate reward value, and γ is the discount factor. The goal of training is to minimize the error between the target value and the predicted value, commonly known as the Temporal Difference (TD) error. The loss function of the main network is defined as:

$$L(\theta) = E\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2\right] \tag{3}$$

where the parameters of the main network and the target network are updated asynchronously. This approach effectively decouples the sample data from the network training.

2.3.2. Dueling Network

Dueling Deep Q Network (Dueling DON) introduces the dueling network structure to both the main network and the target network, The structure of the main network is shown in Figure 3.
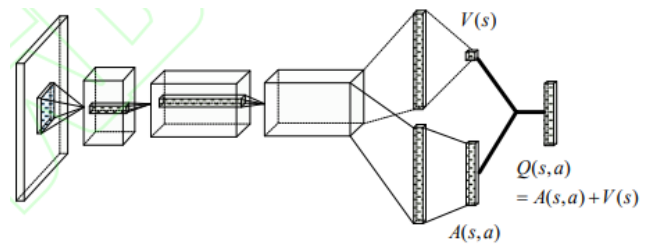

Figure 3. Dueling network structure.

In the dueling network structure, the Q-value function Q (s,

a) is explicitly decomposed into two parts: one part is the value function V(*s*) under the state *s*; the other part is the advantage function A (*s, a*) of taking action a under the state *s*.

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \beta) + \left(A(s, a; \theta, \alpha) - \frac{1}{|A|} \sum_{a'} A(s, a'; \theta, \alpha) \right) \qquad (4)$$

where:

θ represents the shared network parameters,

β represents the network parameters for the state value function,

α represents the network parameters for the action advantage function,

$V(s; \theta, \beta)$ is the value function for state s_t, indicating whether the current state is favorable for obtaining future cumulative rewards,

$A(s, a; \theta, \alpha)$ is the advantage function for action a, indicating how beneficial each possible action is for the current state,

a' represents possible actions,

$\frac{1}{|A|} \sum_{a'} A(s, a'; \theta, \alpha)$ is the average value of the advantage function across actions.

By combining these two evaluation values and calculating the advantage of each action, the dueling network can better understand the differences between state values and various actions, thereby estimating the Q-value more effectively.

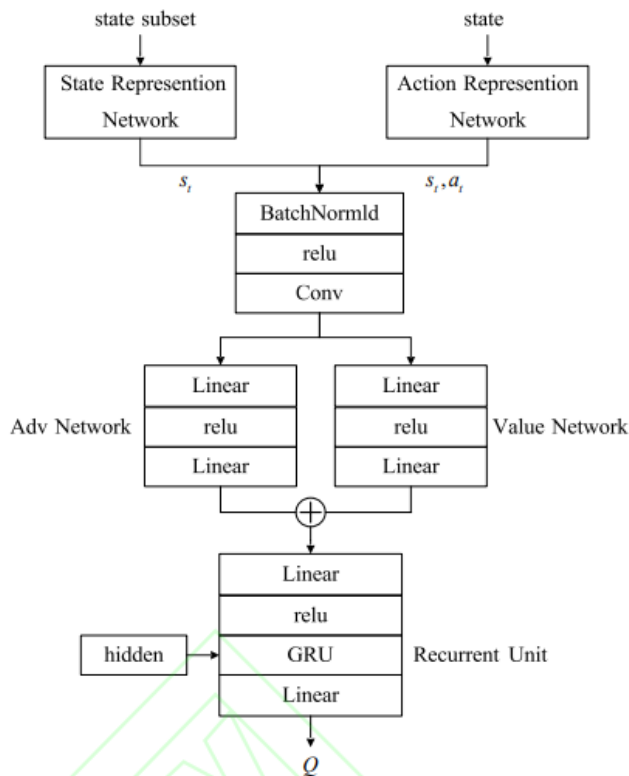### 2.3.3. CG_D3QN Structure

Figure 4. CG_D3QN network framework.

Dueling DON improves the accuracy of action value function predictions by decoupling the value function in this way. The final expression for the output Q-value function is:

In order to enable the query network model to better understand the differences between state values and different actions in semantic segmentation, improve the model's learning efficiency, and alleviate the overestimation problem in Deep Q Networks, the DDQN is combined with the dueling network structure to form the Dueling Double Deep Q Network (D3QN). Additionally, since the state information comes from local regions of image samples, indicating that the environment is a Partially Observable Markov Decision Process (POMDP), the Q-value is not only related to the current state and action but also to historical state information. Therefore, a hybrid network model combining CNN and GRU (CG) is introduced into the D3QN, forming the CG D3QN model. The CG D3QN model uses the CG network to fit the Q function and optimizes the entire network structure through the D3QN network, achieving a high-performance segmentation network with a small amount of labeled data. The framework of the CG D3QN network model is shown in Figure 4.

The design approach of the CG_D3QN network model is as follows:

First, the state and action information are combined and feature extraction is performed. The KL distance distribution features of the action representation, calculated by the Bias network, are used as coefficients to weight the state-action values, obtaining a more accurate action-state value. Then, the obtained action-state value undergoes both value evaluation and advantage evaluation. Finally, the CG network is used to encode the historical information of the state, and this information is recorded in the hidden layer. This allows the model to fully learn previous state information during Q-value evaluation, thereby improving the decision-making performance of the model.

3. Evaluation

The Materials and Methods section should provide comprehensive details to enable other researchers to replicate the study and further expand upon the published results. If you have multiple methods, consider using subsections with appropriate headings to enhance clarity and organization.

3.1. Dataset and Metrics

To verify the feasibility of the CG_D3QN model, the CamVid and Cityscapes datasets were selected to evaluate the semantic segmentation performance. The CamVid dataset

was collected using a car-mounted camera and contains 370 training images, 104 validation images, and 234 test images, with a resolution of 360×480. It provides pixel-level labels for 11 categories, covering classes such as roads, buildings, cars, and pedestrians. The Cityscapes dataset is a large-scale dataset of urban street scenes, containing 3475 high-quality images with a resolution of 2048×1024, of which 2975 are used for training and 500 for validation, covering 19 categories in total. The experimental dataset was divided into four subsets, with detailed information provided in Table 1.

Table 1. *Algorithm hyperparameter setting.*

Dataset	CamVid	Cityscapes
the state subset D_S	10	10
The training subset D_T	100	150
The evaluation subset D_V	260	2615
The reward subset D_R	104	200

The training subset D_T, is used to train the query network under a fixed budget B for labeled pixel regions. Both the state subset D_S and the training subset D_T, are obtained through uniform sampling from the training set. The reward subset D_R is derived from the remaining data obtained through uniform sampling from the validation set or the training set. The evaluation subset D_V consists of a large number of training data samples retained after sampling.

The experiment uses Mean Intersection over Union (MIoU) as the performance evaluation metric for the segmentation network. MIoU is calculated as the arithmetic mean of the IoUs for all categories, providing a comprehensive evaluation of the pixel overlap across the entire dataset. The calculation formula is as follows:

$$MIoU = \frac{1}{n+1} \sum_{i=1}^{n} \frac{p_{ii}}{\sum_{j=0}^{n} p_{ij} + \sum_{j=0}^{n} p_{ji} - p_{ii}} \quad (5)$$

In this context, p_{ii} represents the number of correctly classified pixels; p_{ij} represents the number of pixels belonging to class *i* but predicted as class *j*; and *n* represents the total number of classes.

3.2. Experimental Environment and Parameter Settings

The programming language used for the experiment is Python 3.8, and the framework is PyTorch 1.11. The hardware environment for the experiment includes an NVIDIA GeForce RTX 3090 SUPER graphics card, an i9 13900 processor, 32GB of video memory, and the Windows 11 operating system.

To improve the efficiency of the experiment, the network parameters are updated in batches from the experience replay buffer. The hyperparameters of the reinforcement learning model are shown in Table 2.

Table 2. *Hyperparameter in CG D3QN model.*

Hyperparameters	CamVid	Cityscapes
Region_size	80*90	128*128
Al_algorithm	/	/
Rl_episodes	100	100
Rl_buffer	600	1000
lr	0.001	0.0001
gamma	0.998	0.998
Train_batch_size	32	16
Val_batch_size	4	1
patience	10	10
Num_each_iter	24	256
R1_pool	10	10

3.3. Active Learning Comparison Experiment

To validate the final performance of the model, experiments were conducted using the CamVid and Cityscapes datasets. The FPN network, pretrained on the GTAV dataset, was used as the backbone network for the model. First, the entire training set was set to 1 epoch, and five independent experiments were conducted using different random seeds. The training process consisted of 100 episodes, ultimately producing the query network for active learning.
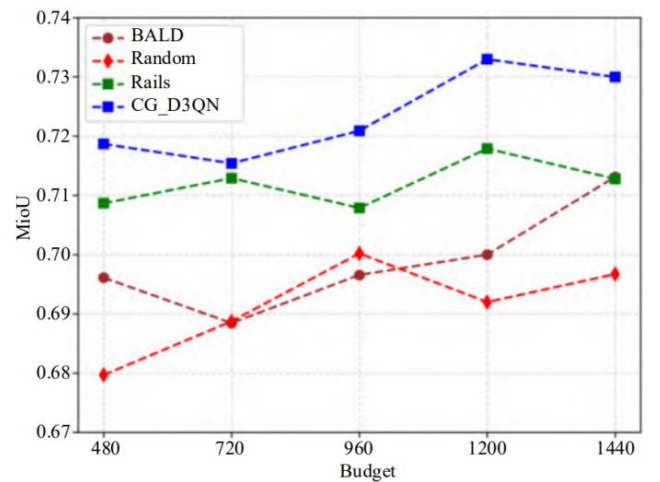

Figure 5. *Comparison of experimental results for various algorithms on the Camvid dataset.*

This study compares the proposed active learning semantic segmentation model based on CG_D3QN with three other active learning methods: Rails, BALD, and Random. The comparison was performed under different pixel region budgets. The effectiveness of the model training was evaluated by comprehensively analyzing the MIoU during the validation phase. The experimental results of the four active learning methods are shown in Figures 5 and 6, where the x-axis represents the number of labeled pixel regions used for training, and the y-axis represents the MIoU level.

By observing the experimental results on the small-scale CamVid dataset shown in Figure 7, it is evident that the traditional random sampling method (Random) and the maximum uncertainty method (BALD) both perform poorly under different budgets. This suggests that training with newly acquired labels does not provide additional information. Moreover, the CG_D3QN method shows a 1% to 5% improvement compared to other models, indicating that a larger label budget contributes to enhanced model performance. The experimental results demonstrate that the region selection strategy of CG_D3QN can help the segmentation model avoid local optima and improve the overall performance of the semantic segmentation model. Due to the small sample size of the CamVid dataset, all results exhibit considerable variance, leading to further validation on the large-scale Cityscapes dataset.
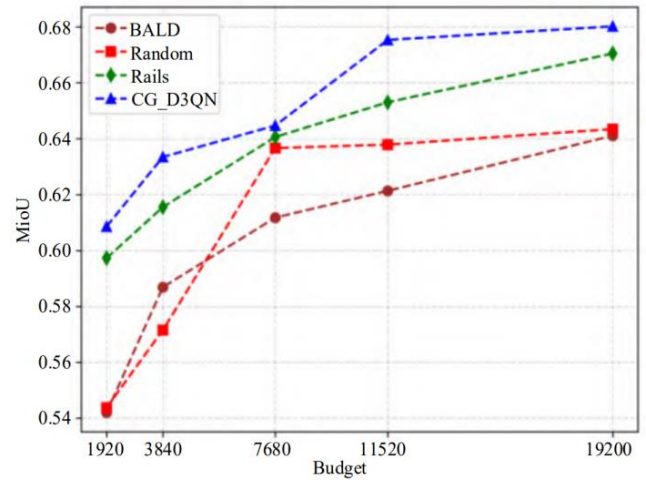

***Figure 6**. Comparison of experimental results for various algorithms on the Cityscapes dataset.*
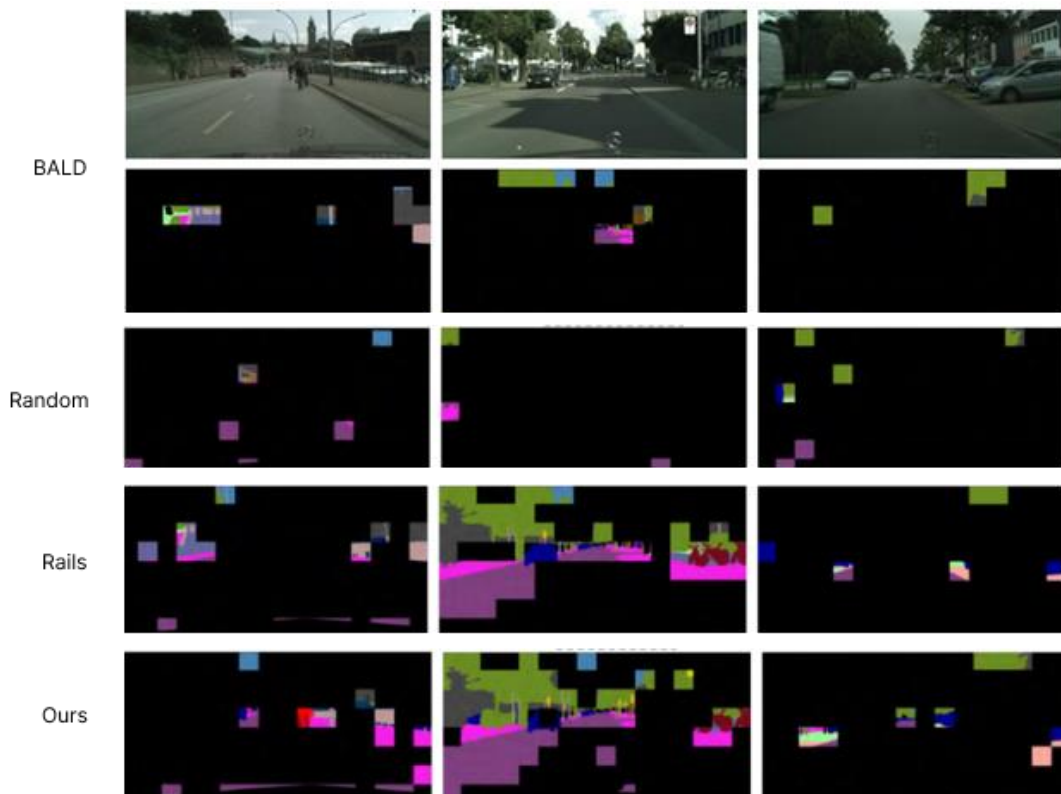

***Figure 7**. 9Visualization results on the Cityscapes dataset.*

Figure 6 presents the performance on the Cityscapes dataset under different budget levels. With a pixel budget of 3840, the CG_D3QN method achieved an MIoU level of 63.3%, while the baseline algorithm Ralis required an additional 65% of labeled pixels to reach the same performance. The experimental results further indicate that the CG_D3QN method can reliably and effectively select the image pixel regions that need to be labeled when handling large-scale

semantic segmentation datasets.

Table 3 provides detailed records of the MIoU results, along with standard deviations, for the 19 classes in the Cityscapes dataset under a pixel region budget of 19,200 for the four active learning methods. The bolded numbers indicate the maximum MIoU values. The experiments demonstrate that across different classes, the CG_D3QN method maintains a relatively high MIoU level compared to other active learning methods. Additionally, for classes with smaller sample sizes, such as Person, Motorcycle, and Bicycle, CG_D3QN also maintains a high MIoU level, confirming the effectiveness of this method in addressing the class imbal-ance problem in image datasets.

To visually demonstrate the advantages of CG_D3QN, this section presents a visual analysis of the selected pixel regions in specific images under consistent budget conditions. The specific results are shown in Figure 7. Compared to traditional active learning methods such as BALD, Random, and reinforcement learning-based active learning method Rails, it can be observed that CG_D3QN includes more informative labels in the selected annotation regions. Furthermore, CG_D3QN focuses more on selecting underrepresented regions, thereby further enhancing the overall performance of the model.

Table 3. Miou results for all image categories when the budget is set to 19200.

Method	Road	SideWalk	Building	Wall	Fence	Pole	Traffic Light
Bald	96.32±0.03	74.74±0.15	89.77±0.06	42.28±0.17	46.91±0.20	49.44±0.17	52.51±0.28
Random	93.94±0.06	65.13±0.22	88.28±0.11	37.70±0.47	44.81±0.43	45.70±0.23	48.86±0.39
Ralis	95.74±0.06	73.13±0.25	89.17±0.10	43.61±0.30	48.01±0.28	47.33±0.17	50.05±0.29
CG_D3QN	96.99±0.03	77.55±0.14	90.85±0.06	45.58±0.12	50.03±0.14	52.18±0.13	56.53±0.23

	Traffic Sign	Vegetation	Terrain	Sky	Person	Rider	Car
Bald	59.56±0.22	89.31±0.05	59.08±0.12	92.64±0.05	73.01±0.10	32.46±0.34	91.52±0.06
Random	55.47±0.39	87.92±0.10	54.58±0.29	91.73±0.17	69.70±0.17	28.98±0.51	88.82±0.12
Ralis	57.98±0.26	88.63±0.08	57.26±0.17	90.18±0.18	92.96±0.17	33.41±0.52	91.11±0.12
CG_D3QN	64.22±0.19	89.84±0.05	59.60±0.07	93.45±0.04	74.96±0.08	41.54±0.03	92.76±0.05

	Truck	Bus	Train	Motorcycle	Bicycle
Bald	30.29±0.40	27.13±0.29	38.40±0.51	37.29±0.39	61.08±0.21
Random	21.29±0.66	23.66±0.69	37.55±0.89	25.99±0.67	57.38±0.42
Ralis	36.98±0.73	35.43±0.61	54.26±0.77	34.24±0.39	61.30±0.30
CG_D3QN	38.43±0.29	35.94±0.22	54.19±0.33	44.32±0.27	64.97±0.18

3.4. Ablation Experiment

To validate the impact of the key techniques used in the CG_D3QN model on its performance, two sets of ablation models were designed to separately test the effects of the Dueling network module and the combined Convolutional Neural Network module (CG) on the overall network performance. The comparison algorithms include the original DDQN model, the Dueling DDQN model, and the CG_D3QN model. These three models were independently trained for 100 episodes under the same experimental parameter settings, and the experimental results are shown in Figure 8.

Based on the experimental results, the following conclusions can be drawn: The DDQN model exhibits significant performance fluctuations under different pixel budget conditions, and it fails to achieve notable improvements under high budget conditions, indicating that this method cannot effectively utilize new label information for decision-making. In

contrast, the Dueling DDQN model, which introduces the Dueling network module, achieves a high MIoU level even under low labeling budget conditions, and its performance gradually improves as the budget increases. This demonstrates that the Dueling network structure can understand action advantages and state values, effectively addressing the Q-network's value overestimation problem, thereby formulating more effective image region selection strategies. The CG_D3QN model, which incorporates the combined Convolutional Neural Network (CG) module on top of the Dueling DDQN structure, achieves significant performance improvements under high budget conditions. This suggests that the recurrent network structure can effectively leverage historical state information in reinforcement learning, enabling the model to learn more valuable information and further enhancing model performance with a large amount of state information.
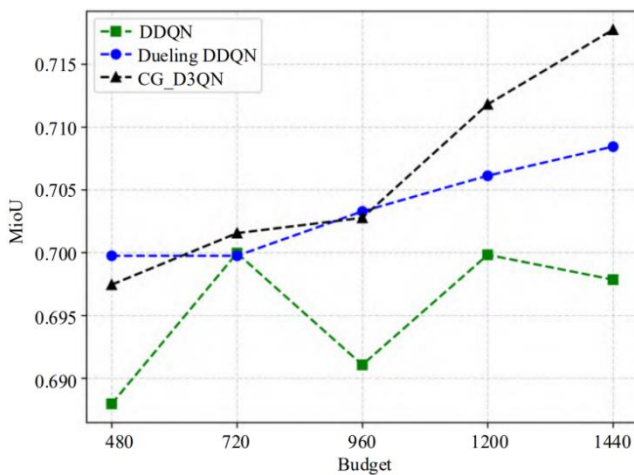

Figure 8. Ablation experiment results on the Camvid dataset.

3.5. Segmentation Model Comparison Experiment

To verify that the CG_D3QN model can still improve the performance of segmentation models when using different image semantic segmentation algorithms, the following experiment was designed: DDRNet [25] and BiSeNet [26], both pretrained on the ImageNet dataset, were used as the segmentation networks within the active learning framework. A comparison experiment was conducted on the CamVid dataset against the original semantic segmentation models. The image region budget was set to 480, and the performance of the algorithms was evaluated after training for 10 epochs. The original semantic segmentation networks used a random strategy to select image regions for segmentation training, while the comparison method used the CG_D3QN model to select regions for training. When transferring the new semantic segmentation algorithms to the CG_D3QN model, the same hyperparameters were used for both experiments. The evaluation metrics included accuracy and MIoU, with the experimental

results on the validation set shown in Table 4.

Table 4. Model performance under different basic segmentation algorithms.

Mthod	Accuracy	MIoU
DDRNet	75.99	34.76
CG_D3NQ+DDRNet	75.71	35.92
BiSeNet	77.46	34.08
CG_D3NQ+ BiSeNet	82.41	38.61

According to the results in Table 4, after applying different semantic segmentation networks with the CG_D3QN model, the MIoU level of the segmentation method using the CG_D3QN model improved under the same budget conditions. This demonstrates that the proposed method can still enhance model performance when using different semantic segmentation networks, validating the applicability of the model. Additionally, this section includes a visual analysis of the active learning region selection strategy. The results in Figure 9 show that on the CamVid dataset, by introducing the CG_D3QN active learning model, the amount of label information in the sample regions selected by both DDRNet and BiSeNet significantly increased, further validating the superiority of this model.
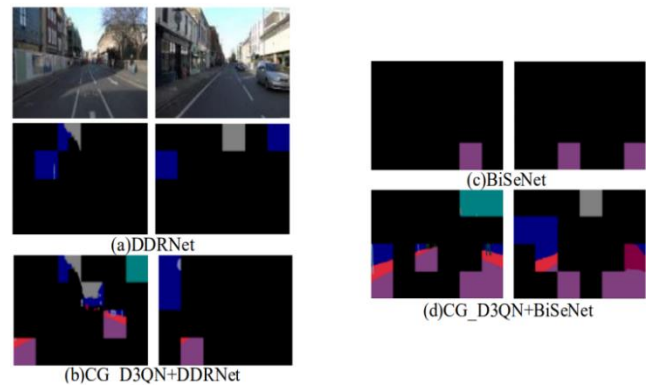

Figure 9. Visualization results on the Camvid dataset.

4. Conclusion

This paper proposes a region-based, data-driven active learning semantic segmentation model, CG_D3QN. The model aims to address the challenges in semantic segmentation tasks, such as the difficulty and large quantity of annotation work, as well as the class imbalance in image samples. The CG_D3QN model builds on the Double Deep Q-Network structure by incorporating a dueling network module and a combined convolutional neural network module to learn the acquisition function during

the active learning process, achieving high-performance semantic segmentation with a smaller amount of labeled samples. Experimental results show that this model demonstrates significant advantages in MIoU across multiple datasets and also improves performance for classes with fewer samples in large-scale datasets, validating the model's effectiveness. Moreover, after applying different semantic segmentation networks, the performance of the segmentation networks was further enhanced, thus validating the algorithm's applicability. In future research, the reinforcement learning algorithm could be further optimized by constructing new state and action representations for region selection, thereby improving the feature representation ability of the reinforcement model and further reducing the amount of annotations required.

Author Contributions

Yan Yu is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Csurka G, Volpi R, Chidlovskii B. Semantic image segmentation: Two decades of research [J]. Foundations and Trends® in Computer Graphics and Vision, 2022, 14 (1-2): 1-162. http://dx.doi.org/10.1561/0600000095

[2] Liao Wensen, Xu Cheng, Liu Hongzhe, et al. Real-time semantic segmentation method for road scenes based on multi-branch networks [J]. Computer Applications Research, 2023, 40 (8): 2526-2530.

[3] Shu Xiu, Yang Yunyun, Xie Ruicheng, et al. Als: Active Learning-Based Image Segmentation Model for Skin Lesion [J/OL]. SSRN Electronic Journal, 2022. (2022-06-21) [2024-02-05]. http://dx.doi.org/10.2139/ssrn.4141765

[4] Zhang Meng, Han Bing, Wang Zhe, et al. Thyroid Cancer Pathological Image Classification Method Based on Deep Active Learning [J]. Journal of Nanjing University: Natural Sciences, 2021. 57 (1): 21-28.

[5] Liu Xiaoyu, Zuo Jie, Sun Pinjie. Research Progress of Machine Learning Algorithms Based on Active Learning [J]. Modern Computer, 2021 (3): 32-36.

[6] Siméoni O, Budnik M, Avrithis Y, et al. Rethinking deep active learning: Using unlabeled data at model training [C] // Proc of the 25th International conference on pattern recognition. NJ: IEEE Press, 2021: 1220-1227. https://arxiv.org/abs10.1109/ICPR48806.2021.9412716

[7] Konyushkova K, Sznitman R, Fua P. Learning Active Learning from Data [J]. Advances in neural information processing systems, 2017, 30.

[8] Ren Pengzhen, Xiao Yun, Chang Xiaojun, et al. A Survey of Deep Active Learning [J]. ACM Computing Surveys, 2022: 1-40. https://doi.org/10.1145/3472291

[9] Budd S, Robinson E C, Kainz B. A survey on active learning and humanin-the-loop deep learning for medical image analysis [J]. Medical Image Analysis, 2021: 102062. https://doi.org/10.1016/j.media.2021.102062

[10] Hu Zeyu, Bai Xuyang, Zhang Runze, et al. LiDAL: Inter-frame Uncertainty Based Active Learning for 3D LiDAR Semantic Segmentation [C] // Proc of European Conference on Computer Vision. Cham: Springer, 2022: 248-265. https://doi.org/10.1007/978-3-031-19812-0_15

[11] Wiering M A, Van Otterlo M. Reinforcement learning [J]. Adaptation, learning, and optimization, 2012, 12 (3): 729.

[12] Fan Yingying, Zhang Shanshan. Hyperspectral Remote Sensing Image Classification Method Based on Deep Active Learning [J]. Journal of Northeast Normal University: Natural Sciences Edition, 2022, 54 (4): 64-70.

[13] Cai L, Xu X, Liew J H, et al. Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs [C] // Proc of the IEEE/CVF conference on computer vision and pattern recognition. NJ: IEEE Press, 2021: 10988-10997.

[14] Mackowiak R, Lenz P, Ghori O, et al. CEREALS-Cost-Effective REgion-based Active Learning for Semantic Segmentation [J/OL]. British Machine Vision Conference, 2018. (2018-10-23) [2024-02-05]. https://doi.org/10.48550/arXiv.1810.09726

[15] Gal Y, Islam R, Ghahramani Z. Deep Bayesian Active Learning with Image Data [C] // International conference on machine learning. New York: PMLR, 2017: 1183-1192.

[16] Hu Mingzhe, Zhang Jiahan, Matkovic L, et al. Reinforcement Learning in Medical Image Analysis: Concepts, Applications, Challenges, and Future Directions [J]. Journal of Applied Clinical Medical Physics, 2023, 24 (2): e13898.

[17] Zhou Wenhong, Li Jie, Zhang Qingjie. Joint Communication and Action Learning in Multi-Target Tracking of UAV Swarms with Deep Reinforcement Learning [J]. Drones, 2022, 6 (11): 339. http://dx.doi.org/10.3390/drones6110339

[18] Sener O, Savarese S. Active Learning for Convolutional Neural Networks: A Core-Set Approach [EB/OL]. (2018-06-01) [2024-02-05]. https://arxiv.org/abs/1708.00489

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

[20] Chan R, Rottmann M, Hyuger F, et al. Application of Decision Rules for Handling Class Imbalance in Semantic Segmentation [EB/OL]. (2019-01-24) [2024-02-05]. https://arxiv.org/abs/1901.08394

[21] Casanova A, Pinheiro Pedro O, Rostamzadeh N, et al. Reinforced active learning for image segmentation [J/OL]. International Conference on Learning Representations, 2020. (2020-02-16) [2024-02-05].

[22] Wang Ziyu, Schaul T, Hessel M, et al. Dueling Network Architectures for Deep Reinforcement Learning [C] // International conference on machine learning. New York: PMLR, 2016: 1995-2003.

[23] Huang Guoyang, Li Xinyi, Zhang Bing, et al. PM2. 5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition [J]. Science of the Total Environment, 2021, 768: 144516.

[24] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // Proc of the IEEE conference on computer vision and pattern recognition. NJ: IEEE Press, 2017: 2117-2125.

[25] Pan Huihui, Hong Yuanduo, Sun Weichao, et al. Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Traffic Scenes [J]. IEEE Trans on Intelligent Transportation Systems, 2022, 24 (3): 3448-3460. https://doi.org/10.48550/arXiv.2101.06085

[26] Yu Changqian, Wang Jingbo, Peng Chao, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation [C] // Proc of the European conference on computer vision. Berlin: Springer, 2018: 325-341.

Biography

Yan Yu is an undergraduate student at Brunel London School, North China University of Technology, majoring in Data Science and Big Data Technology. Her primary research interests include graphical processing and unsupervised one-point attacks based on differential evolution. As the first author, he has published an EI-indexed conference paper that introduced an unsupervised adversarial attack method for detecting security vulnerabilities in critical network domains. Currently, Yan is working on an innovative evolutionary algorithm that leverages overlapping route information to accelerate the optimization of Vehicle Routing Problems (VRP) and Discrete Split Delivery Problems (DSDVRP). This work is under review in an EI-indexed journal. Additionally, Yan leads a municipal-level innovation and entrepreneurship project focused on applying image semantic segmentation methods to high-resolution remote sensing imagery, where he has achieved significant progress.

Research Field

Yan Yu: Graphical processing, unsupervised one-point attacks, image semantic segmentation.