Exploratory Inference Learning for Scribble Supervised Semantic Segmentation

Chunawei Zhou, Zhen Cui*, Chunyan Xu, Cao Han, Jian Yang*

PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education,
Jiangsu Key Lab of Image and Video Understanding for Social Security,
School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.
{cwzhou, zhen.cui, cyx, hancao_work, csjyang}@njust.edu.cn

Abstract

Scribble supervised semantic segmentation has achieved great advances in pseudo label exploitation, yet suffers insufficient label exploration for the mass of unannotated regions. In this work, we propose a novel exploratory inference learning (EIL) framework, which facilitates efficient probing on unlabeled pixels and promotes selecting confident candidates for boosting the evolved segmentation. The exploration of unannotated regions is formulated as an iterative decision-making process, where a policy searcher learns to infer in the unknown space and the reward to the exploratory policy is based on a contrastive measurement of candidates. In particular, we devise the contrastive reward with the intra-class attraction and the inter-class repulsion in the feature space w.r.t the pseudo labels. The unlabeled exploration and the labeled exploitation are jointly balanced to improve the segmentation, and framed in a close-looping end-to-end network. Comprehensive evaluations on the benchmark datasets (PASCAL VOC 2012 and PASCAL Context) demonstrate the superiority of our proposed EIL when compared with other state-of-the-art methods for the scribble-supervised semantic segmentation problem.

Introduction

Semantic segmentation is a fundamental task that serves many other tasks like multi-task learning (Cui et al. 2022; Zhou et al. 2020), intelligent video analysis (Xu et al. 2021a; Zhou et al. 2021; Lv et al. 2021) etc. But it requires great laboring costs to obtain precise annotations in particular massive accurate boundary contours to train a good segmenter. To relieve the annotation burdens, the scribble-supervised semantic segmentation is supported by some flexible lines to train a segmenter. The scribble-supervised annotations can essentially promote segmenter learning in contrast to the fully unsupervised mode. Even so, the challenge of using a few scribble signals is to achieve as much precision as possible, up to the cap of the fully supervised case.

Toward the aim, several previous works have been developed to attempt to extremely exploit sparse semantic annotations. Wang et al. (Wang et al. 2019) introduced a pre-trained edge detection network to estimate outstretched boundaries of the scribbles to refine the segmentation result. In order to relieve the inconsistency of segmentation maps and the uncertainty of prediction results, Pan et al. (Pan et al. 2021) proposed the uncertainty reduction on neural representation to produce confident results, and a self-supervision way on the neural eigen-space for consistent outputs. To better utilize the topological structures between different semantic regions, Tang et al. (Tang et al. 2018a,b) employed the classic dense CRF (Adams, Baek, and Davis 2010), and graph cuts to constrain the optimization for more stable segmentation. The most current methods, including the recent works (Lin et al. 2016; Xu et al. 2021b), referred to generating the constant pseudo labels of some unknown regions by extending from initial scribbles, and then fine-tuned the segmentation model on the estimated constant labels as well as the ground-truth scribbles. Although these methods above have made great advances, they still suffer from insufficient exploration on unannotated regions while paying more attention to the exploitation of label/pseudo-label samples.

To address the above issue, in this work, we propose a novel exploratory inference learning (EIL) framework to perform label propagation through exploratory inference policies for boosting the evolved segmentation. Specifically, the exploratory label inference learning is formulated as a sequential decision-making problem by designing a policy searcher for label exploration and a policy assessor for inference evaluation. The label exploration of the policy takes account of three aspects of attributes: pixel-level features, semantic probabilities, and historical states, which are framed in a unified encoder-decoder fully convolutional network to be learned. Hereby, the next estimated state of pseudo labels is mainly decided by the exploratory policy, with scribble structures as the auxiliary constraint. To assess the reliability of label inference, the rewards (or returns) of the searched inference policies are derived from the variations of states during the segmentation evolution process. In particular, we devise a contrastive critic with the intra-class attraction and the inter-class repulsion in the feature space w.r.t the pseudo labels. The segmenter update and exploratory inference optimization are finally integrated into a unified framework to be optimized alternately. Resorting to the powerful exploration on unlabeled regions as well as the exploitation of labels/pseudo-labels, hereby, the starting scribble-annotated seeds could be progressively diffused to those distant unknown regions. Experimental results demonstrate that our

*Zhen Cui and Jian Yang are the corresponding authors

proposed exploratory inference learning method could effectively mitigate the deficiency of label exploration of the existing methods and achieve state-of-the-art performances on the public benchmarks.

To summarize, the contributions of this work are three-fold: i) We propose an exploratory inference learning (EIL) framework to model scribble-supervised segmentation evolution as a sequential decision-making problem for better exploration on the mass of unlabeled regions. ii) We develop an effective critic rule to assess and optimize exploring policies, and further integrate it with the segmenter learning to balance the exploration of unknown labels and the exploitation of supervised signals during the progressive segmentation evolution. iii) Comprehensive evaluations demonstrate the effectiveness of the proposed EIL and achieve new state-of-the-art performances on the used public datasets for the scribble-supervised semantic segmentation task.

Related Works

Scribble-supervised Segmentation: Scribble-supervised segmentation has attracted increasing attentions recently for its promising effectiveness and fewer label requirements, and many methods have been developed to train segmenter under the limited scribbles. Some methods introduced extra regularization terms in addition to the original scribble annotations to constrain the segmenter learning. For example, NormalCut (Tang et al. 2018a) and KernelCut (Tang et al. 2018b) built topological regularization based on the conventional graph cuts methods to produce more stable segmentation. URNE (Pan et al. 2021) introduced the self-supervised consistency term on the neural-eigen space to generate more consistent segmentation. The above methods mostly focused on exploiting the extremely scarce scribble supervisions but ignored efficient label exploration on the unlabeled regions. In contrast, the proposed EIL learned exploratory label inference policy to effectively diffuse the original scribble seeds to unknown regions, and make a good balance on the unlabeled exploration and labeled exploitation. Instead of resorting to auxiliary constraints, the pseudo-label methods extended the original scribbles to the unknown regions to acquire more label annotations. For example, ScribbleSup (Lin et al. 2016) estimated pseudo labels through optimizing a CRF (Krähenbühl and Koltun 2011) model based on scribbles and network predictions. RAWKS (Vernaza and Chandraker 2017) built a random-walk process to obtain the pseudo labels by propagating the scribbles to unlabeled regions according to the hitting probability-derived transmission matrix. PSI (Xu et al. 2021b) proposed the multi-granularity context aggregation model and learned to generate the constant pseudo labels with a network module in a deterministic manner. Despite the same spirit of inferring pseudo labels, our EIL greatly differs from them as our EIL explores label inference policies while other methods develop constant label extension.

Reinforcement Learning For Vision Tasks Reinforcement learning solves tasks in a trial-and-error manner, hence it is highly suitable for sequential determination tasks like tracking. For example, ADNet (Yun et al. 2017) transferred the tracking problem into a sequential bounding box manipulation task, and searched for better tracking actions within a discrete action space. ACT (Chen et al. 2018a) searched for continuous tracking actions and addressed the problem with the DDPG (Lillicrap et al. 2015) framework in a differentiable manner. In addition to the sequential tracking problem, some segmentation tasks were also transformed into sequential decision-making problems and solved by reinforcement learning. For example, the determination of the optimal sequential user clicking locations was addressed with a deep reinforcement algorithm in SeedNet (Song, Myeong, and Lee 2018) for the interactive image segmentation problem. Yang et al. (Yang et al. 2018) detected the most informative region sequences for human matting to relieve human burdens using deep reinforcement networks. Casanova et al. (Casanova et al. 2020) learned to sequentially select the most informative regions for active annotation with deep reinforcement learning in the active segmentation problem. In this work, we first transform the successive unknown label exploration into a sequential decision-making problem, and develop an exploratory inference learning framework to sequentially search optimal inference policies.

The Proposed Method

Overview Given the training set $\mathcal{X}=\{(X_i, Y_i)|i=1, \cdots, N\}$ where X_i is the i-th image, Y_i represents the corresponding scribble labels and N is the total size of the training set. The goal of the scribble-supervised semantic segmentation is to learn a robust segmenter on the condition of the limited scribble supervision signals and a vast amount of unannotated area. An overview of the proposed exploratory inference learning (EIL) framework has been depicted in Fig. 1. The whole framework is composed of two parts: the segmenter update and the exploratory inference learning, and the latter includes the policy searcher φ and the policy assessor ψ.

At the time t, the segmenter takes the image X_i as input and produces both the intermediate features F_i^t and the segmentation probabilities P_i^t. We combine the deep features F_i^t, the probability map P_i^t, along with the current label M_i^t as the segmentation state s_i^t of the i-th sample at time t. According to the current state $s_i^t = (F_i^t, P_i^t, M_i^t)$, the policy searcher φ can sample a segmentation inference policy a_i^t in a continuous action space, which essentially empowers the label exploration. The explored inference policy a_i^t is then adopted to extend or revise the label signal M_i^t for inferring the next pseudo-label map M_i^{t+1}, which is then used as supervised signals to update the segmenter. Accordingly, we could obtain the new state at the next time $t+1$, i.e., $s_i^{t+1} = (F_i^{t+1}, P_i^{t+1}, M_i^{t+1})$, where F_i^{t+1} and P_i^{t+1} come from the results of the newly updated segmenter. To evaluate the effect of policies, we design a policy assessor ψ to evaluate the evolved segmentation results and the sampling actions with a concrete reward score, denoted $R(s_i^t, a_i^t, s_i^{t+1}) = r_i^t \in \mathbb{R}$. In the stage of exploratory inference learning, the policy searcher, as well as the policy assessor are jointly optimized with the developed policy reward to enhance the exploration capacity, toward the aim of gain maximization. The segmenter update and exploratory inference learning are encapsulated into a unified architecture, in which the exploratory operators φ, ψ, and the segmenter

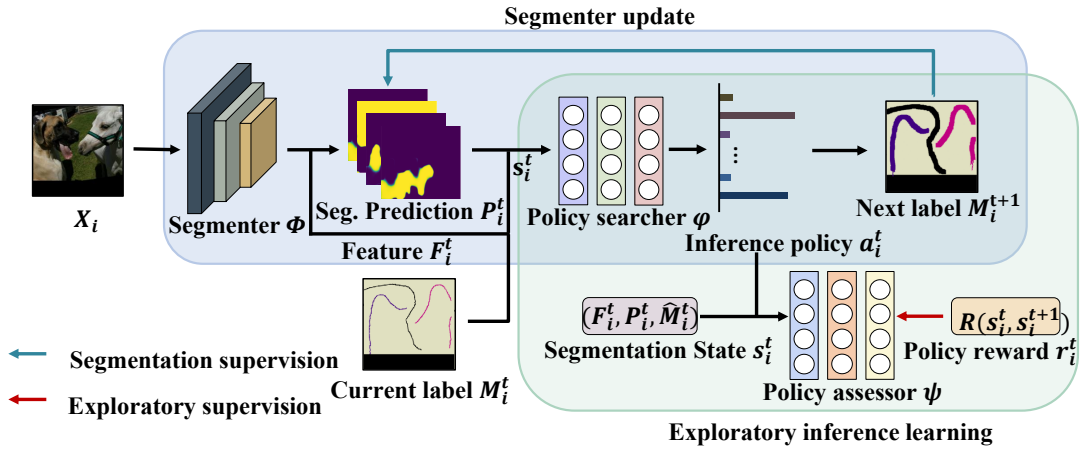


Figure 1: An overview of the proposed EIL framework. We introduce two exploratory operators (policy searcher φ and policy assessor ψ) to realize the exploratory label inference learning, and they are learned with the developed policy reward r_i^t. The explored inference policy a_i^t is then utilized to infer a pseudo label M_i^{t+1} to guide the segmenter update. The exploratory inference learning and segmenter update are enclosed in a close-looping to balance the unlabeled exploration and labeled exploitation. For more details, please refer to the manuscript.

Φ could be alternately learned for mutual promotion.

Exploratory Inference Learning The exploratory inference learning aims to realize the adaptive exploration of the label inference policy in a differentiable manner, which is different from the most previous methods (Lin et al. 2016; Vernaza and Chandraker 2017; Xu et al. 2021b) focusing on exploitation, not exploration. To this end, we propose to formulate the segmentation policy exploration as sequential decision-making, where the exploration of unknown regions and exploitation of the known/explored labels are balanced to learn a robust segmenter. To enable efficient segmentation inference exploration, a policy searcher φ is introduced to learn to sample actions in a continuous space according to the specific segmentation states. The segmentation policy a_i^t is defined as a two-dimensional vector, i.e., $a_i^t = (\tau_o, \tau_b)$, where τ_o and τ_b separately represent the thresholds for the object and background regions. The one-step segmentation inference is then performed by extending the current label map M_i^t to the unlabeled regions by selecting those pixels whose segmentation probabilities are believed confident (i.e., beyond the explored thresholds):

$$m_i^{x,t+1} = \begin{cases} c, & \text{if} \quad p_{i,c}^{x,t} \geq \tau_c, \\ \sigma, & \text{otherwise}, \end{cases} \tag{1}$$

$$\tau_c = \begin{cases} \tau_o, & \text{if} \quad c \geq 1, \\ \tau_b, & \text{if} \quad c = 0, \end{cases} \tag{2}$$

where $m_i^{x,t+1}$ is the class value of the next label M_i^{t+1} at location x, and $p_{i,c}^{x,t}$ denotes the c-th class segmentation probability of P_i^t at location x. σ is a constant value, and it means that the pixel remains unlabeled. The one-step label inference is constrained in a region whose pixel distances w.r.t the currently labeled regions is at most a radius ρ. To realize effective exploratory inference learning, we specifically introduce two exploratory operators including a policy searcher φ and a policy assessor ψ to jointly perform the policy action sampling and label inference evaluation.

Segmentation Policy Assessor ψ: The segmentation policy assessor ψ is a neural network, which is consisted of several residual blocks to produce the confidence criterion of the currently explored segmentation policy a_i^t. It takes the segmentation state s_i^t as well as the segmentation policy a_i^t as input to output a scalar as the segmentation policy criterion. A positive policy criterion value means that the currently explored policy is reliable, and it could produce positive segmenter update guidance, and vice versa. The training of the segmentation policy assessor ψ is realized by closing the gap of two criterion values which are obtained separately from the segmentation states s_i^t and s_i^{t+1} as well as their corresponding segmentation policies. On one hand, we can directly obtain the segmentation policy criterion value for the current time step by forwarding the current segmentation state s_i^t and inference policy a_i^t into the policy assessor ψ:

$$q_i^t = f_\psi(s_i^t, a_i^t), \tag{3}$$

where q_i^t is the segmentation policy criterion value directly obtained by the policy assessor, and $f_\psi(\cdot)$ denotes the forward process of the policy assessor ψ.

On the other hand, the score value for the current time could also be achieved by jointly considering the reward value r_i^t for the current time and the criterion score for the next time step. We first need to obtain the segmentation policy a_i^{t+1} and the criterion value q_i^{t+1} for the next time. Due to the exploratory nature, it requires obtaining robust estimates of the next segmentation policy as well as the criterion value to stabilize the training of the exploratory operators. To this end, we introduce two target exploratory operators $\{\bar{\varphi}, \bar{\psi}\}$ by taking their momentum means to obtain the robust estimates $\bar{a}_i^{t+1}$ and $\bar{q}_i^{t+1}$ for the next time. The next segmentation state s_i^{t+1} is forwarded into the target policy searcher

$\bar{\varphi}$ to obtain the robust estimate of the next segmentation policy, i.e., $\bar{a}_i^{t+1} = g_{\bar{\varphi}}(s_i^{t+1})$, and $g_\varphi(\cdot)$ indicates the forward process of the policy searcher φ. Afterward, the target next segmentation policy $\bar{a}_i^{t+1}$, along with the next segmentation state s_i^{t+1} are input into the target policy assessor $\bar{\psi}$. A robust estimation of the policy score value $\bar{q}_i^{t+1}$ of the next time is afterward attained by inputting $\bar{a}_i^{t+1}$ and s_i^{t+1} into $\bar{\psi}$, i.e., $\bar{q}_i^{t+1} = f_{\bar{\psi}}(s_i^{t+1}, \bar{a}_i^{t+1})$. We can now obtain a robust estimate $\bar{q}_i^t$ of the current policy criterion value by combing r_i^t and $\bar{q}_i^{t+1}$ following:

$$\bar{q}_i^t = r_i^t + \gamma \cdot \bar{q}_i^{t+1}, \tag{4}$$

where γ is a discount ratio, and it is set to 0.9 in this work.

Ideally, the directly computed policy criterion value q_i^t and its robust estimate $\bar{q}_i^t$ should be equal. We hence build a loss function L_{assessor} to evaluate their difference, and the gradient descent algorithm is then adopted to update the policy assessor ψ as follows:

$$\psi \leftarrow \psi - \alpha \cdot \frac{\partial L_{\text{assessor}}(\bar{q}_i^t, q_i^t)}{\partial \psi}, \tag{5}$$

where α is the learning rate, and L_{assessor} is the MSE loss.

Segmentation Policy Searcher φ The segmentation policy searcher φ consists of several residual blocks and it aims to explore effective label inference policies to benefit robust segmenter training. It produces continuous values $\{\tau_o, \tau_g\}$ to perform the label inference indicated in Eq. 1 2. Considering the characteristics of the probability thresholds, we append a sigmoid layer to the policy searcher outputs to regularize the policy actions into the range $[0.0, 1.0]$. Note the hard thresholding operation in Eq. 1, it is unable to directly back-propagate the gradients through the policy inferred segmentation map M_i^{t+1}. Instead, we resort to the trained policy assessor ψ with the wish that the currently explored segmentation policy a_i^t could result in the maximum criterion gain to promote the following segmenter update as much as possible. We hence employ the negative policy criterion value of the current time as the optimization objective L_{searcher} to train the policy searcher φ, and the corresponding gradient descent process is:

$$\varphi \leftarrow \varphi - \beta \cdot \frac{\partial L_{\text{searcher}}(\varphi, s_i^t)}{\partial \varphi}, \tag{6}$$

$$L_{\text{searcher}}(\varphi, s_i^t) = -f_\psi(s_i^t, a_i^t) = -f_\psi(s_i^t, g_\varphi(s_i^t)), \tag{7}$$

where β is the learning rate.

Segmentation Policy Reward The segmentation policy reward r_i^t is a scalar value that is computed from variations of the segmentation states. It is a critical component to realize the segmentation policy learning since the successful training of the exploratory operators largely depend on a favorable policy reward as is shown in Eq. 4 5. The core is to obtain an effective critic for the reliability of the label maps so that the inferred label M_i^{t+1} is more reliable than the current-time label M_i^t to train a robust segmenter. To develop an effective label reliability critic, we take inspiration from the conventional clustering objectives and obtain a critic value by jointly considering the intra-class closeness and inter-class repulsion. Our wish is that the updated segmenter could produce better feature representations where the same-class features distribute more tightly while those different-class features depart farther so that maximum gain could be attained.

Assume the feature of pixel x for the input image X_i is denoted as $f_{i,x}^t$. We next take the current-time label map M_i^t as an example to elaborate its reliability critic value V_i^t. In order to achieve the reliability critic value for the whole label map, we first need to obtain the reliability critic of each class which begins by computing the class feature centers. Take the k-th class as an instance, we can compute the class mean feature $\mu_i^{k,t}$ by taking the average of all the pixel features which are annotated as class k following:

$$\mu_i^{k,t} = \frac{\sum_x f_{i,x}^t \cdot \mathbb{I}[m_{i,x}^t = k]}{\sum_x \mathbb{I}[m_{i,x}^t = k]}, \tag{8}$$

where $m_{i,x}^t$ is the class value of M_i^t at location x. $\mathbb{I}[\cdot]$ is the indicator function whose value equals 1 if the condition in the bracket is true, otherwise, it equals 0. We then derive the reliability critic value of labeling the pixel x as class k by considering its distribution in the whole feature space. Ideally, it should distribute as close to the k-th class center while as far from other centers, therefore its label critic value $v_{i,x}^t$ is devised in a contrastive manner following:

$$v_{i,x}^t = \frac{\exp\{-\delta \cdot d(f_{i,x}^t, \mu_i^k)\}}{\sum_{c=0}^{C} \exp\{-\delta \cdot d(f_{i,x}^t, \mu_i^c)\}}, \tag{9}$$

where δ is a hyper-parameter to enlarge its separation from the class centers, and we set $\delta = 10.0$ in this work. The operator $d(\cdot)$ computes the distance of two feature vectors and we use the '1-cos' distance here as $d(f_1, f_2) = 1 - \cos(f_1, f_2)$. Note the pixel label critic value $v_{i,x}^t$ will reach its extreme if $d(f_{i,x}^t, \mu_i^k) = 0$ and $d(f_{i,x}^t, \mu_i^c)|_{c \neq k} = 2$. Hence the derived contrastive pixel label critic value $v_{i,x}^t$ could simultaneously reflect the intra-class attraction and the inter-class repulsion in the feature space w.r.t the current label M_i^t.

The class-level reliability critic value $V_i^{k,t}$ is then obtained by averaging all the pixel reliability critic values with the same class annotation k:

$$V_i^{k,t} = \frac{\sum_x v_{i,x}^t \cdot \mathbb{I}[m_{i,x}^t = k]}{\sum_x \mathbb{I}[m_{i,x}^t = k]}. \tag{10}$$

Afterward, the label reliability critic value V_i^t for mask M_i^t is reached as the mean value of the class-level reliability critic values of all the classes:

$$V_i^t = \frac{\sum_{c=0}^{C} V_i^{c,t}}{\sum_{c=0}^{C} \mathbb{I}[c]}. \tag{11}$$

The label reliability critic value V_i^{t+1} of the next time could be obtained in the same manner by using M_i^{t+1} and F_i^{t+1}. The inferred label M_i^{t+1} and F_i^{t+1} should bring better feature distribution if it is favorable for segmenter update. The segmentation policy reward r_i^t of the explored policy a_i^t could now be obtained by using the variation of the label reliability critic values as:

$$r_i^t = V_i^{t+1} - V_i^t. \tag{12}$$

It could be observed that a positive policy reward brings tighter intra-class distribution and more inter-class separation, hence a robust segmenter could be expected to optimize.

Segmenter Update The ultimate goal of the exploratory inference learning is to infer favorable pseudo-label maps, according to which a robust segmenter could be trained. Hence, after training the exploratory operators, we infer and select confident pseudo labels to guide the segmenter update. When training the segmenter, a teacher segmenter $\bar{\Phi}$ is first obtained by taking the moving average of the segmenters in previous epochs. The teacher segmenter is to produce a robust segmentation state, and it is fixed to attain the robust feature $\bar{F}_i^t$ and segmentation probability $\bar{P}_i^t$, and the robust segmentation state is then achieved as $\bar{s}_i^t = (\bar{F}_i^t, \bar{P}_i^t, M_i^t)$. A favorable inference policy is then adopted to perform one-step label inference to obtain the next label map M_i^{t+1} following:

$$M_i^{t+1} = \begin{cases} \mathbf{H}(M_i^t, g_\varphi(\bar{s}_i^t)), & \text{if} \quad g_\psi(\bar{s}_i^t, g_\varphi(\bar{s}_i^t)) > 0, \\ M_i^t, & \text{otherwise}, \end{cases} \tag{13}$$

where $\mathbf{H}(\cdot)$ indicates the inference process in Eq. 1. Considering the effective learning of the exploratory operators, we could expect a favorable segmenter update from the inferred pseudo label M_i^{t+1}. The inferred label M_i^{t+1} is then utilized to guide the parameter update of the segmenter Φ as:

$$\Phi \leftarrow \Phi - \lambda \cdot \frac{\partial L_{\text{seg}}(h_\Phi(X_i), M_i^{t+1})}{\partial \Phi}, \tag{14}$$

where $h_\Phi(\cdot)$ represents the probability prediction process of the segmenter Φ, $L_{\text{seg}}(\cdot, \cdot)$ is the segmentation loss function which is set as the cross entropy loss here, and λ is the learning rate. The segmenter is first pre-trained for 100 epochs with solely the original scribbles, and the label exploration and selection is then performed every 20 epochs with a radius of 21 to obtain the pseudo labels for segmenter update and following policy exploration.

Experiment

Datasets: The PASCAL VOC 2012 dataset contains 20 foreground classes in total as well as a background class. The original segmentation training subset is composed of 1464 images which is then extended by (Hariharan et al. 2011) to a full set that includes a total of 10582 images. We utilize the full set to train the main framework, and the ablation studies are conducted on both the subset and the full set. The validation dataset is composed of 1449 fully annotated image samples. The experiments are also conducted on the PASCAL Context dataset (Mottaghi et al. 2014), which has 59 semantic classes as well as a background class. The PASCAL Context dataset is composed of 4998 training images along with 5105 validation images. All the scribbles for the model training are from (Lin et al. 2016).

Implementation Details: The DeepLabV3+ (Chen et al. 2018b) which is equipped with the CPP module (Xu et al. 2021b) is utilized as the segmenter Φ, and ResNet34 (He et al. 2016) is adopted as the backbone for both the policy and critic networks. The first convolutional layers and the last fully connected layers of the two exploratory operators, as well as the CPP module (Xu et al. 2021b) of the segmenter, are initialized with the 'Kaiming_init' method (He et al. 2015), and the remaining layers are initialized from the ImageNet (Deng et al. 2009) pre-trained networks. The SGD optimizer with momentum and weight decay being 0.9 and 5e-4 is adopted to train the segmenter Φ, and its learning rate is initially set 1e-4 and then slowly decayed with a 'poly' schedule. We utilize two SGD optimizers whose momentum and weight decay equal 0.9 and 0.02 to train the exploratory operators, and the learning rates are initialized 1e-3 and decayed with a 'cos' schedule. Random augmentations including scaling ($[0.5, 2.0]$), flipping ($p = 0.5$), rotation ($[-10, 10]$) and cropping (512×512) are adopted in the training stages. The multi-scale, as well as the flipping strategies, are adopted during the testing phase, but no CRF post-processing is utilized as in current works (Wang et al. 2019; Xu et al. 2021b). All the experiments are conducted with the PyTorch framework (Paszke et al. 2019) in an RTX Titan GPU. Our code will be available at our site[1]. Following the previous literature (Lin et al. 2016), we adopt the mean Intersection-over-Union (mIoU) score as the evaluation metric to compare different methods.

Method	Sup.	Backbone	mIoU
DeepLab (Chen et al. 2017)	F	ResNet101	76.8
TreeFCN (Song et al. 2019)	F	ResNet101	80.9
FickleNet (Lee et al. 2019)	I	ResNet101	61.2
OAA (Jiang et al. 2019)	I	VGG16	63.1
IAL (Wang et al. 2020)	I	VGG16	62.0
ICD (Fan et al. 2020)	I	VGG16	64.0
BoxSup (Dai, He, and Sun 2015)	B	VGG16	62.0
WSSL (Papandreou et al. 2015)	B	VGG16	60.6
SDI (Khoreva et al. 2017)	B	VGG16	65.7
ScribbleSup* (Lin et al. 2016)	S	VGG16	63.1
RAWKS (Vernaza and Chandraker 2017)	S	ResNet101	59.5
NormalCut (Tang et al. 2018a)	S	ResNet101	72.8
KernelCut (Tang et al. 2018b)	S	ResNet101	73.0
BPG (Wang et al. 2019)	S	ResNet101	73.2
SPML (Ke, Hwang, and Yu 2021)	S	ResNet101	74.2
PSI (Xu et al. 2021b)	S	ResNet101	74.9
URNE (Pan et al. 2021)	S	ResNet101	76.1
A2GNN (Zhang et al. 2021)	S	TreeFCN	76.2
EIL (Ours)	S	ResNet101	**77.9**

Table 1: Comparison with state-of-the-art methods on the PASCAL VOC 2012 validation set. 'F','I','B' and 'S' separately mean the full supervision, the image-level tags, the boxes and the scribbles. The symbol '*' means CRF post-processing.

Comparison with State-of-the-Art Methods

PASCAL VOC 2012: We first compare our proposed EIL with other state-of-the-art methods on the PASCAL VOC 2012 validation dataset (Everingham et al. 2010). The detailed results have been listed in Table 1. When compared

[1]https://vgg-ai.cn/resources/

Method	Sup.	mIoU(%)
PSPNet (Zhao et al. 2017)	F	47.8
DANet (Fu et al. 2019)	F	52.6
OCR (Yuan, Chen, and Wang 2020)	F	56.2
BoxSup (Dai, He, and Sun 2015)	Semi	40.5
ScribbleSup* (Lin et al. 2016)	S	36.1
RAWKS (Vernaza and Chandraker 2017)	S	36.0
DeepLabV3+ (Xu et al. 2021b)	S	37.1
PSI (Xu et al. 2021b)	S	43.1
CPP(fP)	S	41.7
CPP(qP)	S	42.0
CPP(mqP)	S	42.2
EIL (Ours)	S	**43.8**

Table 2: Comparison with state-of-the-art methods on the PASCAL Context validation dataset following a pure weakly supervised setting. 'F' and 'S' separately mean the full supervisions and the scribbles. The symbol '*' means the CRF post-processing.

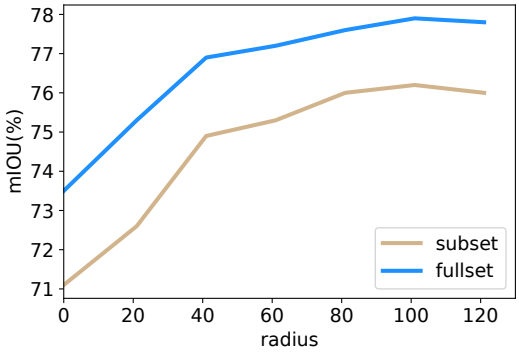


Figure 2: The segmentation accuracy changing curves with growing extension radius on the Pascal VOC 2012 validation set (Everingham et al. 2010).

Policy	None	Rand/PS	Rand+PA	PS+PA
subset	71.1	50.3	72.2	76.2
fullset	73.5	51.7	74.8	77.9

Table 3: The segmentation results on the PASCAL VOC 2012 validation set of four segmenter variants trained with different label inference policies . 'PS' and 'PA' separately refer to policy searcher and policy assessor.

with other scribble-supervised algorithms, our proposed EIL achieves the best performance of 77.9%, and it surpasses other state-of-the-art scribble-supervised methods which verifies the effectiveness of the proposed unlabeled exploration method. Our EIL achieves better performance than those methods which heavily rely on the extreme exploitation of the sparse scribble such as BPG (Wang et al. 2019), KernelCut (Tang et al. 2018b), etc. It shows that our proposed EIL achieves a good balance between unlabeled exploration and labeled exploitation. It is worth noting that our EIL could greatly bridge the performance gap between the weakly-supervised methods and the fully supervised ones, and it even outperforms some fully supervised models such as DeepLab (Chen et al. 2017). It further validates the effectiveness of exploring favorable label supervision for robust segmenter training.

PASCAL Context: We further conduct comparison experiments on the PASCAL Context (Hariharan et al. 2011) dataset which is with more complex scenarios as well as more semantic classes. The detailed results have been listed in Table 2. It could be observed that our proposed EIL achieves notable performance improvement over the current methods. We have also trained the segmenter with some constant pseudo-label generation schemes including the fixed value method (fP), the p-quantile-based model (qP) and the moving p-quantile-based one (mqP). 'qP' utilized the p-quantile of the corresponding class probabilities as the threshold, while 'mqP' adopted a moving average of the p-quantiles over consecutive epochs. We here adopt $p = 0.9$ which achieves the best performance. When compared with the constant pseudo label generation methods, our proposed EIL achieves superior performance to these methods which mainly focus on label exploitation. It shows that our proposed unlabeled exploration could promote the labeled exploitation and our EIL could result in a good balance between the unlabeled exploration and labeled exploitation to train a robust segmenter. Considering the complexity of the PASCAL Context dataset, the superior performances of the proposed EIL further validate the effectiveness of the proposed exploratory label inference policy learning method.

Ablation Studies

In this part, we conduct some ablation studies to verify the adopted parts, and models are separately trained in the training subset ('ss') and full set ('fs') which contains 1464 and 10582 image samples in total for the training of the networks.

Exploratory Operators: We conduct experiments to study the effectiveness of the exploratory networks including the policy searcher (PS in short) and policy assessor (PA in short). The results are listed in Table 3. A baseline that applies no label inference (None) is first trained and achieves 71.1% and 73.5% on the two sets. When the policy searcher is solely introduced to implement policy exploration, it fails to train and is actually equal to adopting random inference policies (Rand/PS). The performances drop greatly by about 20% since the random policy brings many harmful label extensions. We then generate random policies and use the policy assessor to select those confident policies for label inference, and it could bring performance increments of 1.1% and 1.3% over the baseline. It verifies the effectiveness of the policy assessor in evaluating the policy reliability and selecting confident pseudo labels. When both the policy searcher and policy assessor are jointly trained, the performances further promote by 4.0% and 3.1%. It demonstrates that a well-trained policy assessor could provide favorable optimization guidance for the policy searcher, and optimal segmentation inference policies could then be effectively sampled by the

policy searcher.

The Evolved Inference Process: We inspect the label evolution process learned by EIL, where the extension radii are separately 0, 21, 41, 61, 81, 101, and 121. The corresponding evolution curves with growing extension radii are plotted in Fig. 2. On both two sets, the performances first increase when the radius grows from 0 to 101, but start to drop at radius=121. This is because the explored inference rule could continually mine useful label information but the accumulated errors will harm the segmenter optimization in the last few stages. Three instances are plotted in Fig. 3 with the inferred label maps, and the extension radii are separately 21, 61, and 101. It could be seen that in most cases, the explored label inference policies could continually provide useful information. But the proposed framework struggles in some complicated scenes, especially with the thin-structure objects, and it requires further investigations.

Labeled Exploitation v.s. Unlabeled Exploration: We further conduct experiments to compare the performances of our EIL which balances the labeled exploitation and unlabeled exploration with the constant pseudo-label generation methods which focus on the labeled exploitation. We additionally design two variants of EIL where the exploratory operators separately utilize the ResNet18 and ResNet34 as the backbone. The comparison bars on the two sets are displayed in Fig. 4(a). It could be observed that under the both two settings, the proposed EIL with either ResNet18 or ResNet34 could achieve much better performances than the constant policy methods. It sufficiently validates that our proposed unlabeled exploration could lead to better labeled exploitation, verifying the effectiveness of the proposed exploratory label inference policy learning.

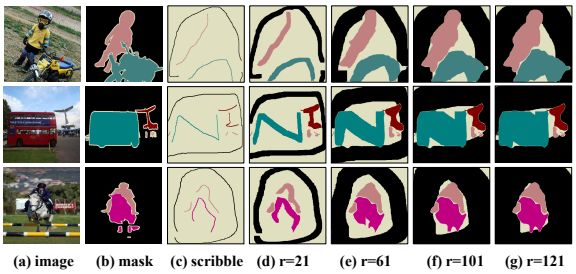
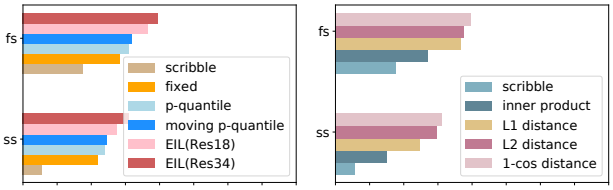


Figure 3: Three representative training instances, the brown color means unlabeled.

Policy Reward: To validate the robustness of the used pseudo label reliability criterion, we design several variants where the distance functions d separately take the form of inner product, L1 distance, L2 distance, and '1-cos' distance. The comparison results are displayed in Fig. 4(b). It could be observed that the 'L1', 'L2' and '1-cos' variants make little difference, especially under the full set setting. This indicates that our proposed pseudo-label reliability criterion is robust to the distance function, which further validates the effectiveness of our proposed pseudo-label evaluation criterion.

Policy Space: The inference policy explored by the policy searcher plays a critical role in training a robust segmenter.



(a) (b)

Figure 4: (a) The comparison bars for different label inference policies on the PASCAL VOC 2012 validation set. (b) The comparison bars for different reward function variants on the PASCAL VOC 2012 validation set.

Inference rule	binary mask	AOD=1	AOD=2	AOD=C
subset	—	75.0	76.2	75.3
fullset	—	76.3	77.9	76.8

Table 4: The segmentation results of four inference rules on the PASCAL VOC 2012 validation set. The symbol '—' means that the algorithm fails to converge. 'AOD' represents the action output dimension.

We hence inspect different policy action space choices in this experiment. The policy actions under consideration are separately the binary masks, the one-dimensional (AOD=1), two-dimensional (AOD=2), and C-dimensional (AOD=C) thresholds where C is the total number of the classes. For the binary masks, the policy is represented as a binary mask, and the adopted pixels are denoted 1. The results are listed in Table 4. It shows that the 'binary mask' variant fails to converge, while 'AOD=2' achieves the best performance among the remaining three candidates since it provides an appropriate space size for the label inference policy exploration.

Conclusion

In this paper, to resolve the insufficient label exploration for the mass of unannotated regions, we propose a novel exploratory inference learning (EIL) framework to facilitate favorable unlabeled pixel probes, and boost the segmentation evolution by selecting those confident inference policy candidates. We formulate the unknown region exploration as a sequential decision-making problem and introduce two exploratory operators (a policy searcher and a policy assessor) to perform effective inference policy learning which is then adopted for the segmenter update. The policy reward to optimize the exploratory operators is devised upon the contrastive reliability criterion which measures the intra-class closeness and inter-class repulsion in the feature space w.r.t the label maps. We encapsulate the inference policy learning and segmenter update into a unified close-looping framework to jointly balance the unlabeled exploration and labeled exploitation to train a robust segmenter. Comparative evaluations as well as ablation studies are conducted to evaluate the effectiveness of our proposed EIL framework. In the future, we plan to apply the framework to other weakly supervised or cross-domain tasks which require unknown exploration.

Acknowledgments

The authors would like to thank all reviewers for their instructive comments. This work was supported by the National Science Fund of China under Grant Nos. 61972204 and 62072244, the fundamental research funds for the central universities under Grant 30919011232.

References

Adams, A.; Baek, J.; and Davis, M. A. 2010. Fast high-dimensional filtering using the permutohedral lattice. In *Computer Graphics Forum*, 753–762.

Casanova, A.; Pinheiro, P. O.; Rostamzadeh, N.; and Pal, C. J. 2020. Reinforced active learning for image segmentation. In *International Conference on Learning Representations*.

Chen, B.; Wang, D.; Li, P.; Wang, S.; and Lu, H. 2018a. Real-time'actor-critic'tracking. In *Proceedings of the European Conference on Computer Vision*, 318–334.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 834–848.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018b. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision*, 801–818.

Cui, Z.; Zhou, L.; Wang, C.; Xu, C.; and Yang, J. 2022. Visual Micro-Pattern Propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Dai, J.; He, K.; and Sun, J. 2015. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, 1635–1643.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 303–338.

Fan, J.; Zhang, Z.; Song, C.; and Tan, T. 2020. Learning Integral Objects With Intra-Class Discriminator for Weakly-Supervised Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4283–4292.

Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; and Lu, H. 2019. Dual attention network for scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3146–3154.

Hariharan, B.; Arbeláez, P.; Bourdev, L.; Maji, S.; and Malik, J. 2011. Semantic contours from inverse detectors. In *Proceedings of the IEEE International Conference on Computer Vision*, 991–998.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Jiang, P.-T.; Hou, Q.; Cao, Y.; Cheng, M.-M.; Wei, Y.; and Xiong, H.-K. 2019. Integral object mining via online attention accumulation. In *Proceedings of the IEEE International Conference on Computer Vision*, 2070–2079.

Ke, T.-W.; Hwang, J.-J.; and Yu, S. X. 2021. Universal Weakly Supervised Segmentation by Pixel-to-Segment Contrastive Learning. In *International Conference on Learning Representations*.

Khoreva, A.; Benenson, R.; Hosang, J.; Hein, M.; and Schiele, B. 2017. Simple does it: Weakly supervised instance and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 876–885.

Krähenbühl, P.; and Koltun, V. 2011. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in Neural Information Processing Systems*, 109–117.

Lee, J.; Kim, E.; Lee, S.; Lee, J.; and Yoon, S. 2019. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5267–5276.

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Lin, D.; Dai, J.; Jia, J.; He, K.; and Sun, J. 2016. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3159–3167.

Lv, H.; Chen, C.; Cui, Z.; Xu, C.; Li, Y.; and Yang, J. 2021. Learning normal dynamics in videos with meta prototype network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 15425–15434.

Mottaghi, R.; Chen, X.; Liu, X.; Cho, N.-G.; Lee, S.-W.; Fidler, S.; Urtasun, R.; and Yuille, A. 2014. The role of context for object detection and semantic segmentation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 891–898.

Pan, Z.; Jiang, P.; Wang, Y.; Tu, C.; and Cohn, A. G. 2021. Scribble-Supervised Semantic Segmentation by Uncertainty Reduction on Neural Representation and Self-Supervision on Neural Eigenspace. In *Proceedings of the IEEE International Conference on Computer Vision*, 7416–7425.

Papandreou, G.; Chen, L.-C.; Murphy, K. P.; and Yuille, A. L. 2015. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, 1742–1750.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 8026–8037.

Song, G.; Myeong, H.; and Lee, K. M. 2018. Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1760–1768.

Song, L.; Li, Y.; Li, Z.; Yu, G.; Sun, H.; Sun, J.; and Zheng, N. 2019. Learnable tree filter for structure-preserving feature transform. In *Advances in Neural Information Processing Systems*.

Tang, M.; Djelouah, A.; Perazzi, F.; Boykov, Y.; and Schroers, C. 2018a. Normalized cut loss for weakly-supervised cnn segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1818–1827.

Tang, M.; Perazzi, F.; Djelouah, A.; Ben Ayed, I.; Schroers, C.; and Boykov, Y. 2018b. On regularized losses for weakly-supervised cnn segmentation. In *Proceedings of the European Conference on Computer Vision*, 507–522.

Vernaza, P.; and Chandraker, M. 2017. Learning random-walk label propagation for weakly-supervised semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7158–7166.

Wang, B.; Qi, G.; Tang, S.; Zhang, T.; Wei, Y.; Li, L.; and Zhang, Y. 2019. Boundary Perception Guidance: A Scribble-Supervised Semantic Segmentation Approach. In *International Joint Conference on Artificial Intelligence*, 3663–3669.

Wang, X.; Liu, S.; Ma, H.; and Yang, M.-H. 2020. Weakly-supervised semantic segmentation by iterative affinity learning. *International Journal of Computer Vision*, 1736–1749.

Xu, C.; Wei, L.; Cui, Z.; Zhang, T.; and Yang, J. 2021a. Meta-vos: Learning to adapt online target-specific segmentation. *IEEE Transactions on Image Processing*, 4760–4772.

Xu, J.; Zhou, C.; Cui, Z.; Xu, C.; Huang, Y.; Shen, P.; Li, S.; and Yang, J. 2021b. Scribble-Supervised Semantic Segmentation Inference. In *Proceedings of the IEEE International Conference on Computer Vision*, 15354–15363.

Yang, X.; Xu, K.; Chen, S.; He, S.; Yin, B. Y.; and Lau, R. 2018. Active matting. In *Advances in Neural Information Processing Systems*, volume 31.

Yuan, Y.; Chen, X.; and Wang, J. 2020. Object-contextual representations for semantic segmentation. In *Proceedings of the European Conference on Computer Vision*.

Yun, S.; Choi, J.; Yoo, Y.; Yun, K.; and Young Choi, J. 2017. Action-decision networks for visual tracking with deep reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2711–2720.

Zhang, B.; Xiao, J.; Jiao, J.; Wei, Y.; and Zhao, Y. 2021. Affinity attention graph neural network for weakly supervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J. 2017. Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2881–2890.

Zhou, C.; Xu, C.; Cui, Z.; Zhang, T.; and Yang, J. 2021. Self-Teaching Video Object Segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 1623–1637.

Zhou, L.; Cui, Z.; Xu, C.; Zhang, Z.; Wang, C.; Zhang, T.; and Yang, J. 2020. Pattern-structure diffusion for multi-task learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4514–4523.