

# CiNuSeg: Class Incremental Nuclei Segmentation via Anchor-driven Consistency Learning with Dual Region Regularization

Xuexin Wu<sup>1</sup>, Zhenhui Ding<sup>1</sup>, Huisi Wu<sup>1\*</sup>, Jing Qin<sup>2</sup>

<sup>1</sup>College of Computer Science and Software Engineering, Shenzhen University

<sup>2</sup>Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University  
2400101079@mails.szu.edu.cn, hswu@szu.edu.cn

## Abstract

Recent advances in deep learning have led to significant improvements in nuclei segmentation from histological images, particularly when labels of all classes are available simultaneously during training. However, in clinical practice, real-world scenarios require a model to perform well in an incremental learning setting, where we anticipate the model to achieve satisfactory performance on previously unseen data while effectively mitigating catastrophic forgetting of old classes. Most previous methods alleviate forgetting by distilling old class knowledge through prototypes; however, they fail to adequately capture fine-grained details to address the challenge of high class similarity, which is particularly severe in histological images. To overcome these limitations, we propose a novel incremental learning method for nuclei segmentation (we call it *CiNuSeg*), which is composed of two key innovative modules. First, we propose a new **Anchor-driven Consistency Learning (ACL)** module to construct multi-level class anchors within each sample to effectively capture fine structural and textural details of nuclei, thereby significantly mitigating forgetting. Second, we develop a **Dual Region Regularization (DRR)** module to suppress new class representations within old class regions while enhancing new class representations within new class regions, strengthening the model’s ability to discriminate between different nuclei types and improving inter-class separability. We further introduce an **Adaptive Temperature Tuning (ATT)** strategy to dynamically balance model stability and plasticity. Extensive experiments conducted on benchmarking MoNuSAC and CoNSeP pathological datasets demonstrate the effectiveness of our method, consistently achieving better performance than SOTAs in different settings.

**Code** — <https://github.com/xxwu1/CiNuSeg>

## Introduction

Nuclei segmentation is a fundamental task in histological image analysis. Precise delineation of nuclei in stained tissue sections facilitates cell-level analysis. This is essential for identifying cell types and assessing their functional states, which are critical factors in disease diagnosis, particularly cancer (Kadaskar and Patil 2023). Although recent studies (Chen, Ding, and Tao 2020; Li et al. 2021; Jiang et al.

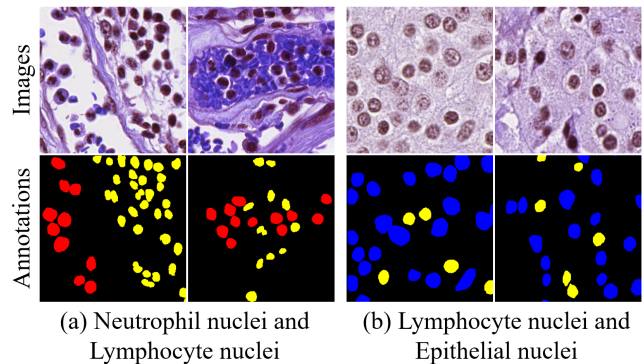


Figure 1: Comparison of nuclear morphologies in H&E-stained images: neutrophil nuclei (red), lymphocyte nuclei (yellow), and epithelial nuclei (blue). Morphological similarities and frequent co-occurrence make visual distinction challenging within a single routine diagnostic image.

2023; Oh and Jeong 2023) have achieved promising results on unchanged datasets, the practical deployment of these segmentation models in clinical settings usually necessitates the capability of performing incremental learning (Verma et al. 2021). In such scenarios, the distribution shift between newly introduced data and previous data may cause substantial adaptation of model parameters toward the new data. This adaptation results in a remarkable decline in performance on previous data, known as *catastrophic forgetting* (McClelland, McNaughton, and O’Reilly 1995).

A common strategy to mitigate forgetting in incremental learning is the replay mechanism (Niu et al. 2024; Kim et al. 2024; Liu et al. 2023), which involves storing and replaying previous data during the acquisition of new knowledge. Nevertheless, this approach requires the careful selection of representative samples and the allocation of considerable storage resources to retain historical data. When storage capacity is limited or retention of prior data is restricted, those methods become impractical. This limitation is particularly pronounced in medical image analysis, where storing previous data is highly challenging due to privacy regulations.

To meet this challenge, some studies explored architectural modification schemes (Zhang and Gao 2025) or regularization strategies (Baek et al. 2022; Park et al. 2025; Cer-

\*Corresponding Author

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

melli et al. 2020; Cheraghian et al. 2021). However, tissue sections in histological images often contain densely packed nuclei that exhibit high similarity in both color and morphological characteristics (see Fig. 1). The high similarity intensifies the risk of catastrophic forgetting in incremental learning, as it impairs the model’s ability to effectively differentiate old data distribution and new data distribution. To solve this, prototype-based methods (Wu et al. 2023) have been proposed to enhance the model’s discrimination capability by preserving prototypes of old classes. However, these methods still have two major limitations. First, most of these methods simply construct the prototypes from the features of the final layer, making them fail to capture crucial fine-grained structural details and local textures preserved in earlier layers, and thus limiting the model’s ability to retain knowledge of small nuclei. Second, prototypes are usually constructed using all samples from each class in the dataset, which obscures the subtle differences among individual samples within the same class, thereby limiting the model’s ability to detect representation drift based on previously learned knowledge.

To address these long-standing challenges, we propose CiNuSeg, a novel framework for class incremental nuclei segmentation, which is composed of two innovative components. First, we propose an anchor-driven consistency learning (ACL) module to facilitate holistic knowledge distillation from old classes. Instead of solely harnessing the features of the final layer, our ACL module constructs old class anchors at each layer of both old and new models. Then, by aligning intermediate features of the two models via these anchors, our ACL is able to prevent representation drift and further alleviate forgetting. Second, we propose a dual region regularization (DRR) module to improve discrimination between old and new classes. Compared with commonly used binary cross-entropy (BCE) loss, which predicts each class independently and thus may assign high confidence to multiple classes for a pixel, parameter updates from new class learning can shift old class representations toward high confidence but incorrect classes. The proposed DRR explicitly separates image regions by suppressing new class predictions within areas of old classes and enhancing them within areas of new classes, and hence improves the discrimination capability of the new model. Meanwhile, DRR also incorporates negative samples to reinforce new class learning, thereby further improving fine-grained discrimination and model plasticity. Moreover, we propose an adaptive temperature tuning (ATT) strategy to dynamically adjust the temperature scaling to balance the learning of new classes and the retention of old ones across different inputs, promoting uniform segmentation performance across all classes. We extensively validated our method on two benchmarking datasets for nuclei segmentation, MoNuSAC (Verma et al. 2021) and CoNSEP (Graham et al. 2019); our model consistently achieves better performance than SOTAs. Our contributions can be summarized as follows:

- We propose a novel method for class incremental nuclei segmentation in pathological images, which can operate in a plug-and-play manner without requiring access to prior annotated training data; we call it *CiNuSeg*.

- We develop the ACL module to comprehensively align class representations between the old and new models across multiple feature scales, and the DRR module to enhance the discrimination between old and new classes. We further develop the ATT strategy for to balance the learning of new classes and the retention of old ones.
- Extensive experiments conducted on the MoNuSAC and CoNSEP pathological datasets demonstrate the effectiveness of our CiNuSeg, outperforming SOTAs by a considerable margin.

## Related Work

### Multi-class Nuclei Segmentation

Nuclei segmentation aims to identify and delineate individual nuclei in microscopic images (e.g., H&E-stained histopathology). Early approaches based on hand-crafted features and classical image processing (Yang, Li, and Zhou 2006; Ali and Madabhushi 2012; Liao et al. 2016) often failed in complex scenarios such as overlapping or occluded nuclei. Recent advances in deep learning have enabled convolutional models (Graham et al. 2019; Alemi Koohbanani et al. 2020; He et al. 2021) to achieve high segmentation accuracy under full supervision, where all labels are available during training. However, more challenging scenarios, such as the potential for nuclei morphology to evolve over time, necessitate a model capable of continuous adaptation.

### Class Incremental Learning

Conventional supervised learning utilizes all labeled data simultaneously, whereas class incremental learning (CIL) operates under sequential label availability: at each step, the model learns only from current classes, lacking previous class labels. While CIL enables scalable continual learning, it suffers from catastrophic forgetting. Existing methods address this through three main strategies: Replay-based methods store and reuse old class samples, features (Borsos, Mutny, and Krause 2020; Castro et al. 2018; Wu et al. 2019; Kemker and Kanan 2018) or generate pseudo-samples of old classes (Wu et al. 2018; Ostapenko et al. 2019) for rehearsal during new class training; Architecture-based methods dynamically expand the model, assigning distinct components to different classes (Zhang and Gao 2025; Hu et al. 2023; Zhang et al. 2022a); Regularization-based methods constrain model limit parameter drift, either at the feature (Dhar et al. 2019; Douillard et al. 2020; Kang, Park, and Han 2022) or logit level (Dong et al. 2023; Li and Hoiem 2018).

### Class Incremental Semantic Segmentation

Class incremental semantic segmentation (CISS) adapts incremental learning to semantic segmentation. MiB (Cermelli et al. 2020) pioneered solutions to the background shift, and PLOP (Douillard et al. 2021) used pseudo-labels and local pooling to preserve spatial context. CoNuSeg (Wu et al. 2023) mitigated forgetting by enforcing prototype consistency and enhancing inter-class separation. EWF (Xiao et al. 2023) directly fused weights from old and new models,

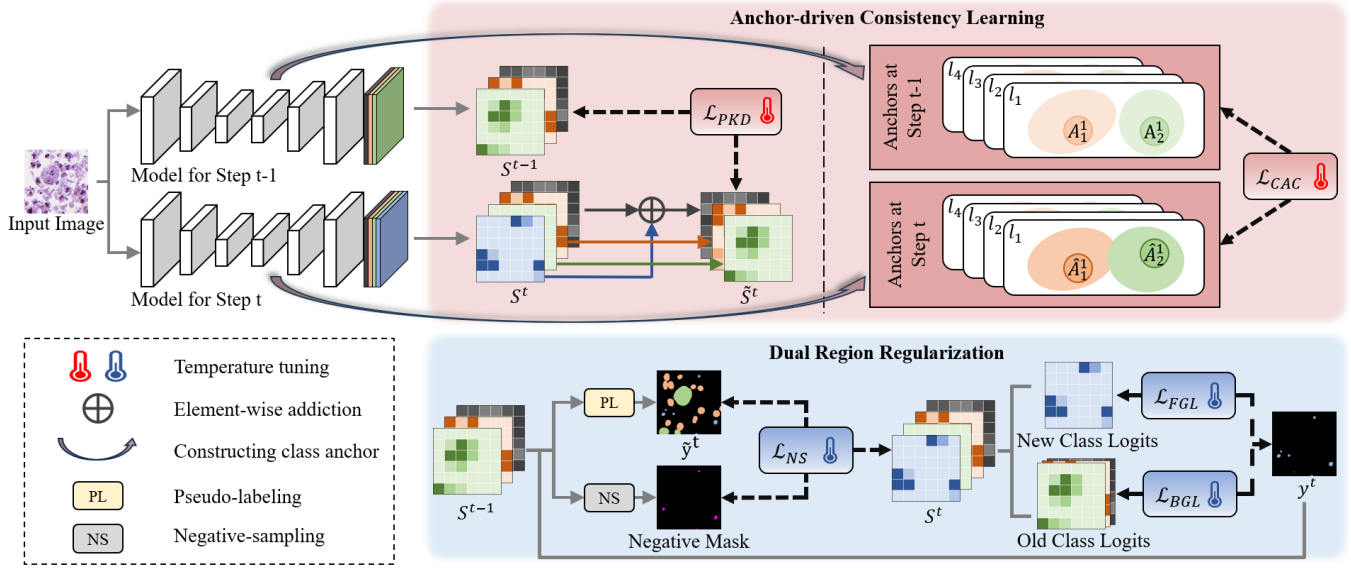


Figure 2: Overview of our CiNuSeg. This framework consists of three modules: ACL, DRR, and ATT. ACL preserves old class knowledge by constructing multi-level anchors; DRR partitions new classes and old classes regions to guide class-specific learning; and ATT adaptively balances new class acquisition and old class retention to achieve optimal performance.

while NeST (Xie et al. 2025) pre-tuned classifiers via transformation matrices. More recent methods such as SSULM (Cha et al. 2021), MicroSeg (Zhang et al. 2022b), and INS (Wang, Wu, and Qin 2024) improve plasticity by anticipating future classes during training. BARM (Zhang and Gao 2025) further introduced a dynamic background adaptation mechanism by modeling residuals.

## Method

### Problem Formulation

Different from ordinary image segmentation, in CISS, a model is trained sequentially over  $T$  steps. At each step  $t = \{1, \dots, T\}$ , the model receives a dataset  $D^t$  and learns a set of new classes  $C^t$ , distinct from all previously learned classes ( $C^i \cap C^j = \emptyset$  for  $i \neq j$ ), with  $C^0$  representing the background class. The dataset  $D^t = \{(x_i^t, y_i^t)\}_{i=1}^N$  consists of images  $x_i^t \in \mathbb{R}^{H \times W \times 3}$  and their corresponding ground-truth labels  $y_i^t \in \mathbb{R}^{H \times W}$ . Crucially, during training at step  $t$ , only pixels belonging to the current classes  $C^t$  retain their original labels. Pixels belonging to previously seen classes  $C^{1:t-1}$ , future classes  $C^{t+1:T}$ , or unlabeled regions are set to the background class  $C^0$ . The model’s objective is to effectively learn the current classes  $C^t$  while mitigating forgetting of past classes  $C^{1:t-1}$ , using only the labels provided for  $C^t$ . An overview of our CiNuSeg is illustrated in Fig. 2.

### Anchor-driven Consistency Learning

In incremental learning, new class learning can shift representations of previous classes in the new model, leading to catastrophic forgetting. As previously mentioned, static old class features in the last layer cannot effectively preserve the knowledge of old classes within the new model. To address this, we propose an anchor-driven consistency learning

(ACL) module to achieve knowledge distillation by generating class anchors. These anchors maintain the new model’s awareness of old class representations during incremental learning, effectively reducing forgetting. The anchors are constructed as follows.

We first employ the previous model with frozen parameters to generate pseudo-labels by leveraging the ground truth ( $y^t$ ) of the current image:

$$\tilde{y}^t = \begin{cases} y^t & \text{if } y^t \in C^t \\ c & \text{if } y^t \notin C^t \text{ and } \max_{c \in C^{1:t-1}} S_c^{t-1} > \tau, \\ c_0 & \text{otherwise} \end{cases} \quad (1)$$

where  $y^t$  is the available ground truth (valid for  $C^t$  and  $C^0$ ), and  $S_c^{t-1}$  is the old model’s prediction score for class  $c$ .

Subsequently, we construct an anchor  $A_c^l$  for class  $c$  at each encoder layer  $l = \{1, 2, 3, 4\}$  using pseudo-labels  $\tilde{y}$  (as masks) and intermediate features  $f^l$ :

$$A_c^l = \frac{\sum_{i=1}^{H^l \times W^l} f_i^l \cdot \mathbf{1}(\tilde{y}_i = c)}{\sum_{i=1}^{H^l \times W^l} \mathbf{1}(\tilde{y}_i = c)}, \quad (2)$$

here,  $\mathbf{1}(\cdot)$  is the indicator function, and pseudo-labels are resized to match the spatial dimensions of  $f^l$  (see Fig. 3). After constructing class-specific anchors, we refrain from using weighted averaging for fusion in order to preserve the distinctiveness of each anchor. Instead, we directly apply cosine similarity to enforce consistency between the anchors generated by the current model and those from the previous model. The Class Anchor Consolidation (CAC) loss is:

$$\mathcal{L}_{CAC} = \frac{1}{L} \sum_{l=1}^L \sum_{c \in C^{1:t-1}} \left[ 1 - \cos \left( A_c^l, \hat{A}_c^l \right) \right], \quad (3)$$

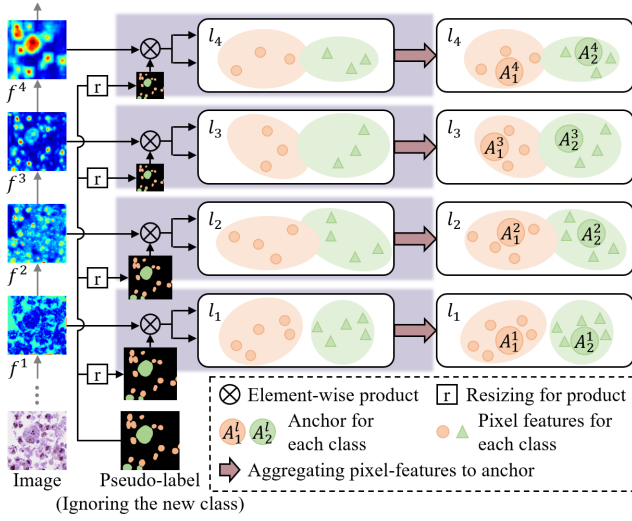


Figure 3: The ACL module constructs anchors by first generating masks for each old class from pseudo-labels. These masks are then applied to the features from each layer via element-wise multiplication to isolate class-specific representations. These isolated features in each layer are aggregated to form the corresponding class anchors.

where  $A_c^l$  and  $\hat{A}_c^l$  denote anchors from the old and new models in layer  $l$ , respectively.  $\cos(\cdot)$  is the cosine similarity. In contrast to conventional coarse prototypes, our method establishes multi-level old class anchors to achieve more accurate representations. This hierarchical design enables the capture of detailed cellular structures and ensures discriminative feature representations within each training batch.

To further mitigate semantic drift in the classifier, we incorporate conventional output-level distillation. Specifically, this is implemented by minimizing the Prediction Knowledge Distillation (PKD) loss:

$$\tilde{S}_c^t = \begin{cases} S_c^t & \text{if } c \neq c_0 \\ S_{c_0}^t + \sum_{k \in C^t} S_k^t & \text{if } c = c_0 \end{cases} \quad (4)$$

$$\mathcal{L}_{\text{PKD}} = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} \sum_{c \in C^{0:t-1}} S_{c,i}^{t-1} \log \tilde{S}_{c,i}^t \quad (5)$$

Our final ACL loss combines both objectives:

$$\mathcal{L}_{\text{ACL}} = \mathcal{L}_{\text{CAC}} + \mathcal{L}_{\text{PKD}} \quad (6)$$

### Dual Region Regularization

To mitigate catastrophic forgetting during new class learning, prior methods apply BCE loss to each class. However, without reliable labels for old classes, relying solely on BCE loss fails to establish a clear distinction between new and old classes. Therefore, we propose a dual region regularization (DRR) module, which explicitly separates regions by suppressing new class predictions in old regions and enhancing them in new ones, improving model discrimination among classes. DRR incorporates the following losses.

**Foreground Guidance Loss (FGL).** This loss directly utilizes the true new class labels to reinforce learning in the

corresponding image regions. Specifically, it minimizes the binary cross-entropy between the model’s predictions ( $S_{c,i}^t$ ) and the pixel-level labels associated with the new class:

$$N = \sum_{i=1}^{H \times W} \mathbb{1}(y_i^t \in C^t), \quad (7)$$

$$\mathcal{L}_{\text{FGL}} = -\frac{1}{N} \sum_{i=1}^{H \times W} \sum_{c \in C^t} \text{BCE}(y_{c,i}^t, S_{c,i}^t) \cdot \mathbb{1}(y_i^t = c), \quad (8)$$

where  $\text{BCE}(\cdot)$  is the binary cross-entropy loss:

$$\text{BCE}(y_1, y_2) = y_1 \log(y_2) + (1 - y_1) \log(1 - y_2). \quad (9)$$

**Background Suppression Loss (BGL).** To explicitly reduce interference and improve discrimination, this loss suppresses predictions for old classes ( $C^{0:t-1}$ ) within regions identified as belonging to new classes ( $y_i^t \in C^t$ ). It penalizes the standard  $L_2$  normalization of the old class scores in these regions, explicitly minimizing their magnitude to enhance segmentation clarity:

$$\mathcal{L}_{\text{BGL}} = \frac{1}{N} \sum_{i=1}^{H \times W} \sum_{c \in C^{0:t-1}} \|S_{c,i}^t\|_2^2 \cdot \mathbb{1}(y_i^t \in C^t). \quad (10)$$

### Weighted BCE Loss with Negative Sample Focus.

While true labels replace pseudo-labels for new regions during training, we observe that the old model often misclassifies some true new class pixels as old classes. These regions represent critical hard negative examples for learning the new classes. To focus the model’s attention on these challenging cases, we construct a negative sample mask ( $O_i^t$ ) and incorporate it with a weighting factor ( $\omega$ ) into the binary cross-entropy loss for all classes ( $C^{0:t}$ ) and pixels:

$$O_i^t = \begin{cases} 1 & \text{if } y_i^t \in C^t \text{ and } \max_{c \in C^{1:t-1}} S_{c,i}^{t-1} > \tau \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

$$\mathcal{L}_{\text{NS}} = -\frac{1}{I} \sum_{i=1}^I \sum_{c \in C^{0:t}} (1 + \omega O_i^t) \text{BCE}(\tilde{y}_{c,i}^t, S_{c,i}^t), \quad (12)$$

where  $I = H \times W$  and  $\mathcal{L}_{\text{NS}}$  represent the BCE loss with negative sample focus. Pixels identified as negative samples ( $O_i^t = 1$ ) receive a higher weight, resulting in larger gradients during backpropagation and promoting the correct classification of these previously misclassified areas.

The DRR losses are calculated by directly summing up the three losses mentioned above:

$$\mathcal{L}_{\text{DRR}} = \mathcal{L}_{\text{FGL}} + \mathcal{L}_{\text{BGL}} + \mathcal{L}_{\text{NS}}. \quad (13)$$

### Adaptive Temperature Tuning

Even though the ACL and DRR modules can respectively preserve old classes and learn new classes for the model. In incremental learning, intra-image class imbalance poses a significant challenge: when training images contain predominantly new class pixels with sparse old class representations, models develop a bias toward new classes. This occurs because scarce old class pixels weaken anti-forgetting constraints, even when regularization techniques are applied.

Method	1-1 (4 tasks)			2-1 (3 tasks)			2-2 (2 tasks)			3-1 (2 tasks)		
	Old	New	Mean	Old	New	Mean	Old	New	Mean	Old	New	Mean
MiB (Cermelli et al. 2020)	53.45	62.47	60.22	57.91	71.20	64.55	57.11	70.45	63.78	61.99	75.39	65.34
PLOP (Douillard et al. 2021)	63.42	63.94	63.81	66.52	69.96	68.24	63.06	72.24	67.65	67.69	74.79	69.46
CoNuSeg (Wu et al. 2023)	65.90	67.59	67.17	64.02	71.79	67.90	63.13	70.99	67.06	69.46	73.02	70.35
EFW (Xiao et al. 2023)	68.77	63.23	64.62	63.14	69.90	66.52	65.77	69.42	67.59	70.76	63.56	68.96
IDEC (Zhao, Yuan, and Shi 2023)	63.87	68.20	67.12	62.37	74.66	68.51	65.77	74.40	70.08	62.27	76.55	69.59
NeST (Xie et al. 2025)	62.54	68.99	67.38	61.68	73.99	67.83	65.06	73.77	69.41	68.90	76.70	70.85
BARM (Zhang and Gao 2025)	69.82	68.56	68.88	63.45	75.28	69.36	64.65	76.58	70.62	70.26	76.40	71.80
INS (Wang, Wu, and Qin 2024)	68.52	69.48	69.24	63.90	75.60	69.75	66.12	76.25	71.19	70.32	77.10	72.02
Ours	<b>70.28</b>	<b>71.13</b>	<b>70.92</b>	<b>64.22</b>	<b>76.58</b>	<b>70.40</b>	<b>68.76</b>	<b>77.10</b>	<b>72.93</b>	<b>71.33</b>	<b>78.18</b>	<b>73.04</b>
Offline	73.09	75.12	74.61	71.21	78.01	74.61	71.21	78.01	74.61	73.08	79.20	74.61

Table 1: Quantitative comparison of CISS methods on the MoNuSAC dataset using mDice (%). Results are reported under four incremental learning protocols (1-1, 2-1, 2-2 and 3-1), showing the performance on old classes, new classes, and their mean. Offline upper-bound results are provided in gray for reference. Best incremental results are highlighted in **bold**.

To mitigate this bias, we propose adaptive Temperature Tuning (ATT), which dynamically scales loss intensities based on each image’s pixel composition. When new class pixels dominate an image, the loss temperature is increased for old classes to reinforce learning, and decreased for new classes to reduce bias. When the class distribution is more balanced, default temperatures are preserved. The dominance ratio  $r$  is computed as:

$$r = \frac{\sum_{i=1}^I \mathbb{1}(\tilde{y}_i^t \in C^t)}{\sum_{i=1}^I \mathbb{1}(\tilde{y}_i^t \in C^{1:t-1}) + \sum_{i=1}^I \mathbb{1}(\tilde{y}_i^t \in C^t)}, \quad (14)$$

where  $I = H \times W$ . The temperature parameters for the new class learning loss ( $t_n$ ) and the old class retention loss ( $t_o$ ) are then adjusted based on  $r$ :

$$t_n = \begin{cases} \alpha_n & \text{if } r > \sigma \\ 1 & \text{otherwise} \end{cases}; \quad (15)$$

$$t_o = \begin{cases} \alpha_o & \text{if } r > \sigma \\ 1 & \text{otherwise} \end{cases}, \quad (16)$$

where  $\sigma$  is a threshold indicating significant dominance of new classes,  $\alpha_n$  and  $\alpha_o$  are scalar hyper-parameters controlling the degree of adjustment (typically  $\alpha_n < 1$ ,  $\alpha_o > 1$ ).

### Loss Function

Finally, the old class retention loss ( $\mathcal{L}_{ACL}$ ) and new class learning loss ( $\mathcal{L}_{DRR}$ ) are combined through the ATT strategy, yielding the following final loss function:

$$\mathcal{L} = t_o \mathcal{L}_{ACL} + t_n \mathcal{L}_{DRR}. \quad (17)$$

## Experiments

### Experimental Details

**Datasets.** We conduct experiments on two public datasets: MoNuSAC and CoNSeP. The MoNuSAC dataset, introduced in the 2020 Multi-organ Nuclei Segmentation and Classification Challenge, contains H&E stained tissue images with annotations for four cell types: epithelial cells,

lymphocytes, macrophages, and neutrophils. The CoNSeP dataset, provided by Hover-net (Graham et al. 2019), contains 41 H&E stained images with seven original classes. We re-group these into three final segmentation classes: Epithelial (healthy/dysplastic/malignant), Spindle-shaped (fibroblasts/muscle/endothelial), and Other.

**Experimental Protocols.** For a general CISS task, two basic configurations exist: overlapped and disjoint. In the disjoint configuration, pixels in  $D^t$  may belong to both old and new classes ( $C^{1:t-1} \cup C^t$ ). Conversely, in the overlapped configuration, pixels in  $D^t$  encompass old classes, new classes, and potentially future classes ( $C^{1:t-1} \cup C^t \cup C^{t+1:T}$ ), reflecting a more realistic scenario. Our work employs the overlapped configuration. Furthermore, incremental learning settings vary. We denote them as X-Y, where X is the number of classes learned initially, and Y is the number of new classes added per subsequent step. Since nuclei in medical images are small and occupy relatively small areas, we adopt the mean Dice Similarity Coefficient (mDice) as our evaluation metric.

**Implementation Details.** We employ a ResNet-101 backbone pre-trained on ImageNet as the feature encoder. Hyper-parameters are configured as follows: learning rate = 0.01 (initial step) and 0.001 (incremental steps), epochs = 100, batch size = 12. For the loss function, we set  $\alpha_o = 10$ ,  $\alpha_n = 0.5$ ,  $\tau = 0.7$ ,  $\omega = 0.1$ , and  $\sigma = 0.7$ . All training stages use SGD with momentum (0.9) and weight decay (0.01). Experiments were conducted on NVIDIA RTX 3090 GPUs.

### Comparison with State-of-the-art Methods

To validate the effectiveness of our method, we compared it with several CISS methods. As shown in Tab. 1, different methods exhibit varying results under identical experimental conditions. Methods like IDEC (Zhao, Yuan, and Shi 2023) achieves promising performance in learning new class but suffer from significant forgetting of old classes due to limited constraints on knowledge retention, reducing overall effectiveness. In contrast, approaches such as BARM (Zhang and Gao 2025) maintain old class knowledge effectively but

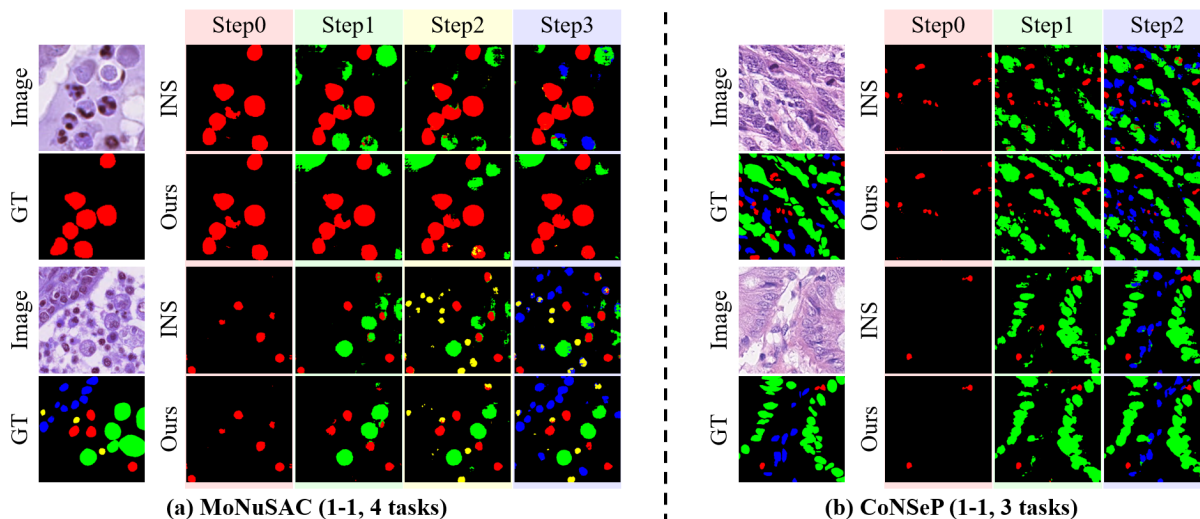


Figure 4: Visual comparison of segmentation results between our CiNuSeg and INS on the MoNuSAC and CoNSeP datasets: (a) Results on the MoNuSAC dataset; (b) Results on the CoNSeP dataset.

Method	1-1 (3 tasks)			2-1 (2 tasks)		
	Old	New	Mean	Old	New	Mean
MiB (Cermelli et al. 2020)	58.37	66.17	63.57	70.48	59.70	66.88
PLOP (Douillard et al. 2021)	63.67	68.67	67.00	71.65	62.61	68.63
CoNuSeg (Wu et al. 2023)	65.16	68.95	67.69	72.62	62.66	69.30
EWf (Xiao et al. 2023)	<b>67.76</b>	69.45	68.89	72.78	63.37	69.64
IDEC (Zhao, Yuan, and Shi 2023)	65.74	69.67	68.36	71.28	64.14	68.90
NeST (Xie et al. 2025)	64.50	67.29	66.36	73.70	60.32	69.24
BARM (Zhang and Gao 2025)	65.66	69.41	68.16	71.81	63.25	68.96
INS (Wang, Wu, and Qin 2024)	65.96	70.26	68.83	<b>74.46</b>	63.05	70.65
Ours	66.07	<b>70.69</b>	<b>69.15</b>	73.98	<b>64.36</b>	<b>70.77</b>
Offline	68.59	72.55	71.23	74.46	64.77	71.23

Table 2: Quantitative comparison of CISS methods on the CoNSeP dataset using mDice (%) under two protocols (1-1 and 2-1). Best incremental results are indicated in **bold**, while offline upper-bound performance is provided in gray for reference.

hinder the learning of new class, resulting in suboptimal average performance. Our method achieves superior results on the MoNuSAC dataset across all tasks. Notably, compared to the SOTA method INS (Wang, Wu, and Qin 2024), our approach achieves substantial performance gains:

- Task 1-1: +1.76 (old), +1.65 (new), +1.68 (mean)
- Task 2-1: +0.32 (old), +0.98 (new), +0.65 (mean)
- Task 2-2: +2.64 (old), +0.85 (new), +1.74 (mean)
- Task 3-1: +1.01 (old), +1.08 (new), +1.02 (mean)

This indicates that CiNuSeg achieves the best performance. In addition, as shown in Tab. 2, we evaluate our method on the CoNSeP dataset to further validate its effectiveness.

**Qualitative Comparison.** Beyond quantitative results, Fig. 4 provides a visual comparison of segmentation outputs. Compared with INS (Wang, Wu, and Qin 2024), our method produces more accurate and complete segmentation. Notably, it exhibits significantly reduced forgetting of previously learned classes during the incremental learning pro-

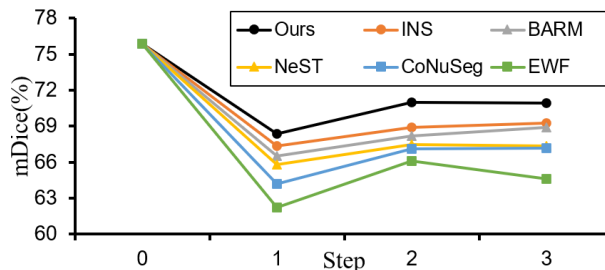


Figure 5: Evolution of mean Dice metric across incremental learning steps for CISS (Task 1-1 on the MoNuSAC dataset).

cess. After all classes have been introduced, our approach consistently outperforms INS (Wang, Wu, and Qin 2024) in terms of segmentation quality, effectively alleviating catastrophic forgetting. Furthermore, the qualitative observations are supported by clear visual evidence in Fig. 5.

$\mathcal{L}_{ACL}$	$\mathcal{L}_{DRR}$	ATT	1-1 (4 tasks)		
			old	new	mean
			60.98	69.93	67.69
✓			68.22	69.18	68.94
	✓		50.12	70.35	65.29
✓	✓		67.65	<b>71.95</b>	70.87
✓	✓	✓	<b>70.28</b>	71.13	<b>70.92</b>

Table 3: Ablation study of our method on the MoNuSAC dataset under the incremental protocol 1-1. **Bold** denotes best mDice (%) scores.

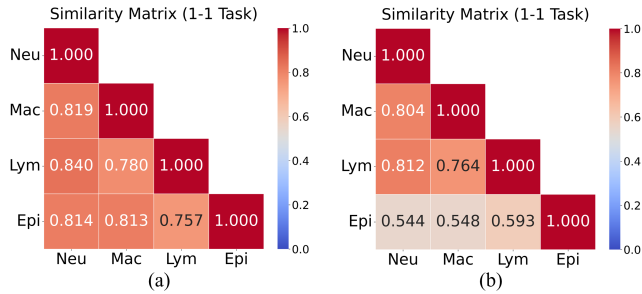


Figure 6: Similarity matrix of nuclear prototypes: (a) Raw similarity matrix of nucleus prototypes; (b) Prototype similarity separation with DRR module. (Neu: neutrophils, Mac: macrophages, lym: lymphocytes, Epi: epithelial)

## Ablation Studies

**Component Analysis.** To evaluate the effectiveness of our method, we conduct ablation studies on the MoNuSAC dataset using the 1-1 task setting. As shown in Tab. 3, the baseline relies solely on pseudo-labels combined with BCE loss to constrain the new model’s learning. Incorporating the anchor-driven consistency learning (ACL) module improves segmentation performance on old classes, demonstrating enhanced retention of previously acquired knowledge. However, this improvement comes at a slight cost to new class performance, due to the added anti-forgetting constraints.

Adding the dual region regularization (DRR) alone improves performance on new classes compared to the baseline. Yet, without explicit constraints to preserve old classes, the model exhibits forgetting of previously learned classes.

Based on the baseline with ACL and DRR, incorporating the adaptive temperature tuning (ATT) strategy further improves segmentation performance on both new and old classes, demonstrating the effectiveness of ATT in balancing knowledge retention and acquisition.

**Further Analysis.** To assess the impact of our DRR module on inter-class similarity, we calculated the similarity between prototypes of each class. As shown in Fig. 6, the introduction of the DRR module leads to a noticeable reduction in the similarity between class prototypes, indicating that the DRR enhances the separability of learned classes in the feature space. This enhanced discriminability contributes to more effective learning of new classes.

We further visualize the evolution of attention regions in

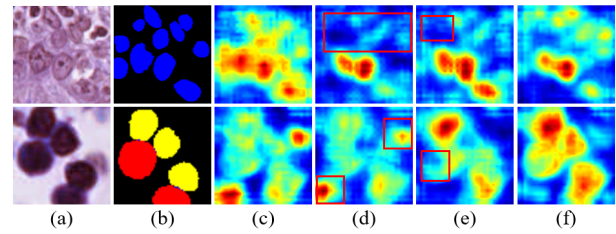


Figure 7: Components ablation on the MoNuSAC dataset 1-1 (4 tasks) : (a) Image; (b) GT; (c) Baseline; (d) w/ACL; (e) w/(ACL+DRR); (f) w/(ACL+DRR+ATT).

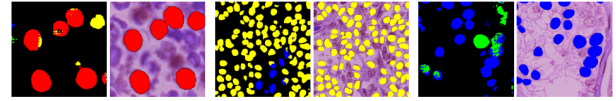


Figure 8: Failure cases. For each case, the left shows the prediction of the model, while the right displays the corresponding original image and ground truth.

the feature maps as the proposed modules are progressively integrated. As shown in Fig. 7, the baseline model produces chaotic and diffuse attention maps. With the addition of the ACL and DRR modules, the model exhibits increasingly focused attention regions. However, some areas remain either insufficiently or excessively activated, as indicated by the red borders. After integrating the ATT module, the attention regions show strong alignment with the annotations, highlighting the effectiveness of our method.

## Discussions and Limitations

While our method achieves SOTA performance on the MoNuSAC and CoNSeP datasets, it also holds promise for adaptation to other datasets that present similar challenges. However, our method has certain limitations. Specifically, due to its excessive focus on fine details, the model occasionally misclassifies non-nuclear regions, as illustrated in Fig. 8. This issue will be further explored in future work.

## Conclusion

We propose a novel CIL method for nuclei segmentation in pathological images, named CiNuSeg. First, we introduce the ACL module, which constructs the model’s prior knowledge into anchors. By anchoring representations of old classes, the model mitigates representation drift during new classes learning, thereby alleviating catastrophic forgetting and enhancing model stability. Second, we propose the DRR module, which partitions the model output into regions corresponding to new and old class and applies distinct regularization to each. This design enhances inter-class discrimination while promoting model plasticity. Third, we present the ATT strategy to dynamically balance the stability-plasticity trade-off in the incremental learning model. Our CiNuSeg achieves SOTA performance on two pathological datasets under different incremental protocols.

## Acknowledgments

This work was supported partly by National Natural Science Foundation of China (No. 62273241), Natural Science Foundation of Guangdong Province, China (No. 2024A1515011946), the Shenzhen Research Foundation for Basic Research, China (No. JCYJ20250604181940054), the grant under the Collaborative Research with World-leading Research Groups scheme of The Hong Kong Polytechnic University (project no. G-SACF) and the Shenzhen-Hong Kong-Macao Science and Technology Plan Project (Category C Project) under Shenzhen Municipal Science and Technology Innovation Commission (project no. SGDX20230821092359002).

## References

- Alemi Koohbanani, N.; Jahanifar, M.; Zamani Tajadin, N.; and Rajpoot, N. 2020. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Analysis*, 65: 101771.
- Ali, S.; and Madabhushi, A. 2012. An Integrated Region-, Boundary-, Shape-Based Active Contour for Multiple Object Overlap Resolution in Histological Imagery. *IEEE Transactions on Medical Imaging*, 31(7): 1448–1460.
- Baek, D.; Oh, Y.; Lee, S.; Lee, J.; and Ham, B. 2022. Decomposed Knowledge Distillation for Class-Incremental Semantic Segmentation. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 10380–10392. Curran Associates, Inc.
- Borsos, Z.; Mutny, M.; and Krause, A. 2020. Coresets via Bilevel Optimization for Continual Learning and Streaming. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 14879–14890. Curran Associates, Inc.
- Castro, F. M.; Marin-Jimenez, M. J.; Guil, N.; Schmid, C.; and Alahari, K. 2018. End-to-End Incremental Learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Cermelli, F.; Mancini, M.; Bulò, S. R.; Ricci, E.; and Caputo, B. 2020. Modeling the Background for Incremental Learning in Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Cha, S.; Kim, B.; Yoo, Y.; and Moon, T. 2021. SSUL: Semantic Segmentation with Unknown Label for Exemplar-based Class-Incremental Learning. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 10919–10930. Curran Associates, Inc.
- Chen, S.; Ding, C.; and Tao, D. 2020. Boundary-Assisted Region Proposal Networks for Nucleus Segmentation. In Martel, A. L.; Abolmaesumi, P.; Stoyanov, D.; Mateus, D.; Zuluaga, M. A.; Zhou, S. K.; Racoceanu, D.; and Joskowicz, L., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 279–288. Cham: Springer International Publishing. ISBN 978-3-030-59722-1.
- Cheraghian, A.; Rahman, S.; Fang, P.; Roy, S. K.; Petersson, L.; and Harandi, M. 2021. Semantic-Aware Knowledge Distillation for Few-Shot Class-Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2534–2543.
- Dhar, P.; Singh, R. V.; Peng, K.-C.; Wu, Z.; and Chellappa, R. 2019. Learning Without Memorizing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Dong, J.; Liang, W.; Cong, Y.; and Sun, G. 2023. Heterogeneous Forgetting Compensation for Class-Incremental Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 11742–11751.
- Douillard, A.; Chen, Y.; Dapogny, A.; and Cord, M. 2021. PLOP: Learning Without Forgetting for Continual Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4040–4050.
- Douillard, A.; Cord, M.; Ollion, C.; Robert, T.; and Valle, E. 2020. PODNet: Pooled Outputs Distillation for Small-Tasks Incremental Learning. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 86–102. Cham: Springer International Publishing. ISBN 978-3-030-58565-5.
- Graham, S.; Vu, Q. D.; Raza, S. E. A.; Azam, A.; Tsang, Y. W.; Kwak, J. T.; and Rajpoot, N. 2019. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58: 101563.
- He, H.; Huang, Z.; Ding, Y.; Song, G.; Wang, L.; Ren, Q.; Wei, P.; Gao, Z.; and Chen, J. 2021. CDNet: Centripetal Direction Network for Nuclear Instance Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4026–4035.
- Hu, Z.; Li, Y.; Lyu, J.; Gao, D.; and Vasconcelos, N. 2023. Dense Network Expansion for Class Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11858–11867.
- Jiang, H.; Zhang, R.; Zhou, Y.; Wang, Y.; and Chen, H. 2023. DoNet: Deep De-Overlapping Network for Cytology Instance Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 15641–15650.
- Kadaskar, M.; and Patil, N. 2023. Image analysis of nuclei histopathology using deep learning: a review of segmentation, detection, and classification. *SN Computer Science*, 4(5): 698.
- Kang, M.; Park, J.; and Han, B. 2022. Class-Incremental Learning by Knowledge Distillation With Adaptive Feature Consolidation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16071–16080.
- Kemker, R.; and Kanan, C. 2018. FearNet: Brain-Inspired Model for Incremental Learning. arXiv:1711.10563.
- Kim, J.; Cho, H.; Kim, J.; Tiruneh, Y. Y.; and Baek, S. 2024. SDDGR: Stable Diffusion-based Deep Generative Replay

- for Class Incremental Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 28772–28781.
- Li, X.; Yang, H.; He, J.; Jha, A.; Fogo, A. B.; Wheless, L. E.; Zhao, S.; and Huo, Y. 2021. Beds: Bagging Ensemble Deep Segmentation For Nucleus Segmentation With Testing Stage Stain Augmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 659–662.
- Li, Z.; and Hoiem, D. 2018. Learning without Forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12): 2935–2947.
- Liao, M.; qian Zhao, Y.; hua Li, X.; shan Dai, P.; wen Xu, X.; kai Zhang, J.; and ji Zou, B. 2016. Automatic segmentation for cell images based on bottleneck detection and ellipse fitting. *Neurocomputing*, 173: 615–622.
- Liu, Y.; Cong, Y.; Goswami, D.; Liu, X.; and van de Weijer, J. 2023. Augmented Box Replay: Overcoming Foreground Shift for Incremental Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 11367–11377.
- McClelland, J. L.; McNaughton, B. L.; and O’Reilly, R. C. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3): 419.
- Niu, C.; Pang, G.; Chen, L.; and Liu, B. 2024. Replay-and-Forget-Free Graph Class-Incremental Learning: A Task Profiling and Prompting Approach. arXiv:2410.10341.
- Oh, H.-J.; and Jeong, W.-K. 2023. DiffMix: Diffusion Model-Based Data Synthesis for Nuclei Segmentation and Classification in Imbalanced Pathology Image Datasets. In Greenspan, H.; Madabhushi, A.; Mousavi, P.; Salcudean, S.; Duncan, J.; Syeda-Mahmood, T.; and Taylor, R., eds., *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 337–345. Cham: Springer Nature Switzerland. ISBN 978-3-031-43898-1.
- Ostapenko, O.; Puscas, M.; Klein, T.; Jahnichen, P.; and Nabi, M. 2019. Learning to Remember: A Synaptic Plasticity Driven Framework for Continual Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Park, G.; Moon, W.; Lee, S.; Kim, T.-Y.; and Heo, J.-P. 2025. Mitigating Background Shift in Class-Incremental Semantic Segmentation. In Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; and Varol, G., eds., *Computer Vision – ECCV 2024*, 71–88. Cham: Springer Nature Switzerland. ISBN 978-3-031-72973-7.
- Verma, R.; Kumar, N.; Patil, A.; and Kurian, e. a. 2021. MoNuSAC2020: A Multi-Organ Nuclei Segmentation and Classification Challenge. *IEEE Transactions on Medical Imaging*, 40(12): 3413–3423.
- Wang, H.; Wu, H.; and Qin, J. 2024. Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11408–11417.
- Wu, C.; Herranz, L.; Liu, X.; wang, y.; van de Weijer, J.; and Raducanu, B. 2018. Memory Replay GANs: Learning to Generate New Categories without Forgetting. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.
- Wu, H.; Wang, Z.; Zhao, Z.; Chen, C.; and Qin, J. 2023. Continual Nuclei Segmentation via Prototype-Wise Relation Distillation and Contrastive Learning. *IEEE Transactions on Medical Imaging*, 42(12): 3794–3804.
- Wu, Y.; Chen, Y.; Wang, L.; Ye, Y.; Liu, Z.; Guo, Y.; and Fu, Y. 2019. Large Scale Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Xiao, J.-W.; Zhang, C.-B.; Feng, J.; Liu, X.; van de Weijer, J.; and Cheng, M.-M. 2023. Endpoints Weight Fusion for Class Incremental Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7204–7213.
- Xie, Z.; Lu, H.; Xiao, J.-w.; Wang, E.; Zhang, L.; and Liu, X. 2025. Early Preparation Pays Off: New Classifier Pre-tuning for Class Incremental Semantic Segmentation. In Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; and Varol, G., eds., *Computer Vision – ECCV 2024*, 183–201. Cham: Springer Nature Switzerland. ISBN 978-3-031-73347-5.
- Yang, X.; Li, H.; and Zhou, X. 2006. Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 53(11): 2405–2414.
- Zhang, A.; and Gao, G. 2025. Background Adaptation with Residual Modeling for Exemplar-Free Class-Incremental Semantic Segmentation. In Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; and Varol, G., eds., *Computer Vision – ECCV 2024*, 166–183. Cham: Springer Nature Switzerland. ISBN 978-3-031-72943-0.
- Zhang, C.-B.; Xiao, J.-W.; Liu, X.; Chen, Y.-C.; and Cheng, M.-M. 2022a. Representation Compensation Networks for Continual Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7053–7064.
- Zhang, Z.; Gao, G.; Fang, Z.; Jiao, J.; and Wei, Y. 2022b. Mining Unseen Classes via Regional Objectness: A Simple Baseline for Incremental Segmentation. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 24340–24353. Curran Associates, Inc.
- Zhao, D.; Yuan, B.; and Shi, Z. 2023. Inherit With Distillation and Evolve With Contrast: Exploring Class Incremental Semantic Segmentation Without Exemplar Memory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10): 11932–11947.