Medical Image Analysis 105 (2025) 103728

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Rethinking data imbalance in class incremental surgical instrument segmentation

Shifang Zhao [a,1], Long Bai [a,b,*,1], Kun Yuan [b,c,d], Feng Li [b], Jieming Yu [a], Wenzhen Dong [a], Guankun Wang [a], Mobarakol Islam [e], Nicolas Padoy [c,d], Nassir Navab [b], Hongliang Ren [a,*]

[a] *Department of Electronic Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong*
[b] *Chair for Computer Aided Medical Procedures, Technical University of Munich, Munich, Germany*
[c] *University of Strasbourg, CNRS, INSERM, ICube, Strasbourg, France*
[d] *IHU Strasbourg, Strasbourg, France*
[e] *UCL Hawkes Institute, University College London, London, United Kingdom*

ARTICLE INFO

Keywords:
Continual learning
Robotic surgery
Data imbalance
Instrument segmentation
Data synthesis

ABSTRACT

In surgical instrument segmentation, the increasing variety of instruments over time poses a significant challenge for existing neural networks, as they are unable to effectively learn such incremental tasks and suffer from catastrophic forgetting. When learning new data, the model experiences a sharp performance drop on previously learned data. Although several continual learning methods have been proposed for incremental understanding tasks in surgical scenarios, the issue of data imbalance often leads to a strong bias in the segmentation head, resulting in poor performance. Data imbalance can occur in two forms: (i) class imbalance between new and old data, and (ii) class imbalance within the same time point of data. Such imbalances often cause the dominant classes to take over the training process of continual semantic segmentation (CSS). To address this issue, we propose **SurgCSS**, a novel plug-and-play CSS framework for surgical instrument segmentation under data imbalance. Specifically, we generate realistic surgical backgrounds through inpainting and blend instrument foregrounds with the generated backgrounds in a class-aware manner to balance the data distribution in various scenarios. We further propose the Class Desensitization Loss by employing contrastive learning to correct edge biases caused by data imbalance. Moreover, we dynamically fuse the weight parameters of the old and new models to achieve a better trade-off between the biased and unbiased model weights. To investigate the data imbalance problem in surgical scenarios, we construct a new benchmark for surgical instrument CSS by integrating four public datasets: EndoVis 2017, EndoVis 2018, CholecSeg8k, and SAR-RAPR50. Extensive experiments demonstrate the effectiveness of the proposed framework, achieving significant performance improvement against existing baselines. Our method demonstrates excellent potential for clinical applications. The code is publicly available at github.com/Zzsf11/SurgCSS.

1. Introduction

Image segmentation of surgical instruments is crucial for advancing fully autonomous robotic surgery, as it provides critical information to understand dynamic and complex surgical scenes (Chen et al., 2024; Nwoye et al., 2023; Maier-Hein et al., 2022; Cheng et al., 2025; Yu et al., 2024b; Ranem et al., 2024). Recently, deep learning-based methods have shown promising results in solving increasingly challenging tasks (Garcia-Peraza-Herrera et al., 2021; Colleoni et al., 2022; Lou et al., 2023; Liu et al., 2021; Alabi et al., 2025; Yu et al., 2024a). However, collecting and aggregating a large-scale dataset for surgical scenarios is difficult in practice due to high storage costs, licensing, and privacy concerns, and the frequent updating of instruments (Zia et al., 2023; Razzak et al., 2018; Wang et al., 2023b; Bai et al., 2023). Most approaches suffer from *catastrophic forgetting*, where the performance has a significant degradation when the past data are not available. To address this challenge, Continual Learning (CL) methods aim to enable models to learn new classes or tasks by training on new data

[*] Corresponding authors.

E-mail addresses: shifang.zhao@bjtu.edu.cn (S. Zhao), b.long@link.cuhk.edu.hk (L. Bai), kun.yuan@tum.de (K. Yuan), feng.li@tum.de (F. Li), jmyu@link.cuhk.edu.hk (J. Yu), dongwz@link.cuhk.edu.hk (W. Dong), gkwang@link.cuhk.edu.hk (G. Wang), mobarakol.islam@ucl.ac.uk (M. Islam), npadoy@unistra.fr (N. Padoy), nassir.navab@tum.de (N. Navab), hlren@ee.cuhk.edu.hk (H. Ren).

[1] Co-first authors.

https://doi.org/10.1016/j.media.2025.103728
Received 19 February 2025; Received in revised form 23 June 2025; Accepted 9 July 2025
Available online 22 July 2025

while alleviating the catastrophic forgetting problem (Chaudhry et al., 2018; Zhao et al., 2024; Li and Hoiem, 2017; Douillard et al., 2020; Li et al., 2025; Ayromlou et al., 2024; Wang et al., 2023c). Nevertheless, previous methods suffer significant performance degradation in robotic surgical scenarios due to biased incremental learning updates, as illustrated in Fig. 1, which shows a pronounced bias in the segmentation head weights for new and old classes.

Why is continual segmentation more challenging in robotic surgical scenarios? We believe this stems from the combination of three factors: (i) the data distribution for instrument classes is highly imbalanced due to variations in surgical workflow, with different phase durations and erratic occurrences; (ii) the imbalance is exacerbated as the new class has a high probability of occurrence and is highly concentrated, whereas the old class has a low probability of occurrence and is more dispersed; and (iii) the visual similarity of instruments, with minimal differences near class boundaries, further complicates segmentation. Even worse, the unavailability of exemplar samples due to licensing and privacy constraints makes the task even more challenging. Additionally, attributes specific to surgical scenes lead to increased difficulty. Unlike the relatively stable illumination in general settings, surgical illumination is highly directional and often positioned to highlight specific areas. The anatomical complexity also leads to various illumination degradation scenes. The diversity of surgical instruments and the similar visual characteristics also introduce further complexity. Besides, in surgical scenes, camera viewpoints are often highly constrained, leading to noise and reduced image quality, which further increases the difficulty of applying segmentation models.

To address the class imbalance challenges, classical techniques have been widely explored in machine learning to mitigate imbalanced distributions. Recent studies (He et al., 2021) have analyzed the relationship between class incremental learning and class imbalanced learning, demonstrating that techniques from class imbalanced learning are similar to class incremental learning methods. Wu et al. (2019), Liu et al. (2022) and Zhao et al. (2020) have attributed the performance degradation caused by sample imbalance to biases in the final fully connected layer. These biases are mitigated by incorporating a bias correction layer trained using a balanced validation set (Wu et al., 2019) or normalizing the output weights (Zhao et al., 2020). While these methods provide some improvement, they fall short in the context of continual learning for surgical instrument segmentation, where data imbalance is further exacerbated by privacy restrictions and the inherent complexity of surgical workflows. Current continual semantic segmentation approaches still lack effective strategies to improve performance under imbalanced distribution scenarios, which are typical in surgical environments.

In medical image analysis, data scarcity is often addressed by generating synthetic data, enabling knowledge sharing without exposing patient-level information (Colleoni and Stoyanov, 2021; Colleoni et al., 2022). Similarly, continual learning utilizes generative models to replay synthesized data (Shin et al., 2017; Wu et al., 2018). However, updating generative models incrementally introduces challenges, including high resource consumption and reduced system stability. In this case, Xu et al. (2024) blended and augmented the foreground to the synthetic background to avoid generating model updates. These methods provide the flexibility to create a class-balanced dataset using any preferred instrument, while also automatically generating the instrument annotations. However, they simply crop patches without instruments from the images and use these patches to train a generative model. In this case, (i) the generated backgrounds are often distorted and may contain unforeseen artifacts and inconsistencies, (ii) the model can only learn foreground information based on segmentation annotations from such data, but due to the distorted backgrounds, it is challenging for the model to extract effective discriminative information between the foreground and the background, and (iii) this approach limits the model to generated background patches, neglecting the effective utilization of the scarce real samples available. Therefore, although this method is straightforward, there is still significant room for improvement.

In this work, we propose **SurgCSS**, a novel plug-and-play approach specifically designed to mitigate the degradation caused by class imbalance in the continual segmentation of robotic surgical scenes. We identify a significant bias in the segmentation head, analyze its root causes, and propose a two-stage pipeline to resolve this issue. Specifically, we decouple the instrument foreground from the surgical background and construct a small class-balanced set using a class-aware blending technique, which is crucial in overcoming data scarcity in surgical scenarios. Additionally, we design the Class Desensitization Loss (CDL) function to reduce sensitivity to class boundary uncertainty, thereby minimizing the impact of visual similarity between classes. After training the model on the small set, we apply a dynamic weight fusion (DWF) strategy to merge the representation capabilities of the trained and untrained models effectively. Our approach effectively addresses the class imbalance between new and old data, as well as the class imbalance within the same time point. We summarize our contributions as follows:

- We propose a novel plug-and-play framework to address performance degradation in class incremental surgical instrument segmentation, specifically targeting imbalanced distributions under data-scarce surgical scenarios.
- We introduce a class-aware blending with an inpainting mechanism to decouple the instrument foreground from the high-fidelity surgical background, effectively tackling the data scarcity issue. The CDL is further proposed to minimize confusion between visually similar classes and correct segmentation boundary deviations.
- We develop a DWF strategy that combines the strengths of both incremental and rebalance models, ensuring detailed class representations are preserved while boosting overall segmentation accuracy.
- By integrating four existing public datasets, we establish a new large-scale benchmark for class incremental segmentation of surgical instrument segmentation. Through extensive comparisons and ablation studies on this benchmark, we demonstrate the potential of our method for clinical applications.

2. Related work

2.1. Class-incremental continual learning

Class-Incremental Continual Learning aims for a model to learn distinguishing features for new classes while preventing catastrophic forgetting of previously learned classes. This requires balancing the need for learning flexibility with maintaining memory stability (Wang et al., 2024). In Continual Semantic Segmentation, a specific challenge is background shift, which happens when pixels labeled as background correspond to classes the model has already seen or will see later. Early methods like MiB (Cermelli et al., 2020) addressed this by calibrating the loss for background pixels using prior model predictions. PLOP (Douillard et al., 2021) used a multi-scale pooling approach for distillation. SSUL (Cha et al., 2021) used auxiliary data to improve distillation and prevent forgetting. Other approaches focused on preventing feature changes; RCIL (Zhang et al., 2022) and EWF (Xiao et al., 2023) worked to reduce feature plasticity problems by combining parameters from the new and old models. CIT (Ge et al., 2024) converts the output of existing semantic segmentation models into a form that does not depend on the specific class. IPSeg (Yu et al., 2025) uses image posterior probabilities to align the optimization process across different stages and reduce negative effects from optimizing each stage separately. MoDA (Yang et al., 2024) helps the SAM encoder create features that are clearly separated for different task types, providing

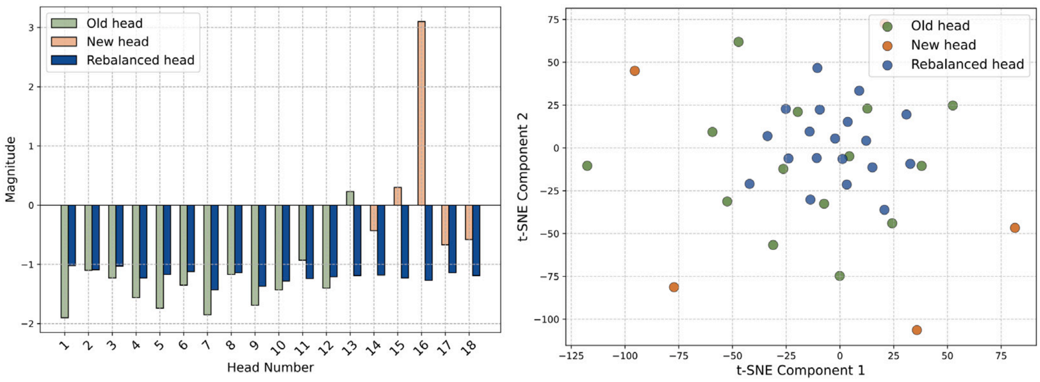


Fig. 1. Experimental results on a surgical dataset confirm the presence of bias in the final layer, as observed through weight visualization of the segmentation head in task "13-1". Heads 1–13 correspond to the old class, while heads 14–18 represent incremental classes, one by one (Best viewed in color). We visualize the bias within each class-specific output channel of the final convolutional layer using two methods: a summation average of activations (left) and a t-SNE embedding of the feature space (right).

accurate task-specific information for continual learning. EIR (Yin et al., 2025) reduces background changes in new images by combining stored examples with the new images. CoMBO (Fang et al., 2025) adds an extra part to handle learning new classes while preserving the original information used for distillation. These recent methods represent the current advances in general continual segmentation scenarios.

Continual segmentation is also being explored in medical image analysis. Xu et al. (2024) uses synthetic data with class-aware temperature normalization to enhance model rigidity while preserving privacy in robotic surgery. Zhang et al. (2023) propose replacing the conventional output layer with a set of lightweight, class-specific heads. Ji et al. (2023) demonstrates that continually training, followed by freezing the encoder, combined with incrementally added decoders, can effectively extract sufficiently representative image features for new classes, allowing them to be validly segmented. Elskhawy et al. (2020) introduces an approach that disentangles the feature space into task-specific and task-invariant features. Zhu et al. (2024) employs lightweight low-rank adaptation (LoRA) to efficiently extend pre-trained segmentation models for segmenting new organs. Sadegheih et al. (2025) introduces a mixture-of-experts mechanism and dual knowledge distillation for brain lesion segmentation. Anand et al. (2024) proposes the construction of a common label space to facilitate incremental learning across different datasets. Guo et al. (2025) introduces segmentation models that can handle a comprehensive set of fine-grained whole-body anatomies from diverse, partially labeled datasets in CT segmentation.

However, previous continual segmentation methods that use distillation may struggle with visual similarities between different instruments or anatomical structures, particularly near class boundaries. Methods focused on parameter fusion or regularization may not inherently address the severe class imbalance. Additionally, while data limitations represent a significant challenge in medical continual learning tasks, previous research in medical image analysis has primarily focused on modifying model architectures to improve performance, often neglecting the critical issue of data scarcity and imbalance. In this paper, we further explore the specific challenges and solutions related to surgical instrument semantic segmentation, a domain marked by imbalanced data, privacy concerns, and the complexity of surgical workflows.

2.2. Learning with class imbalance

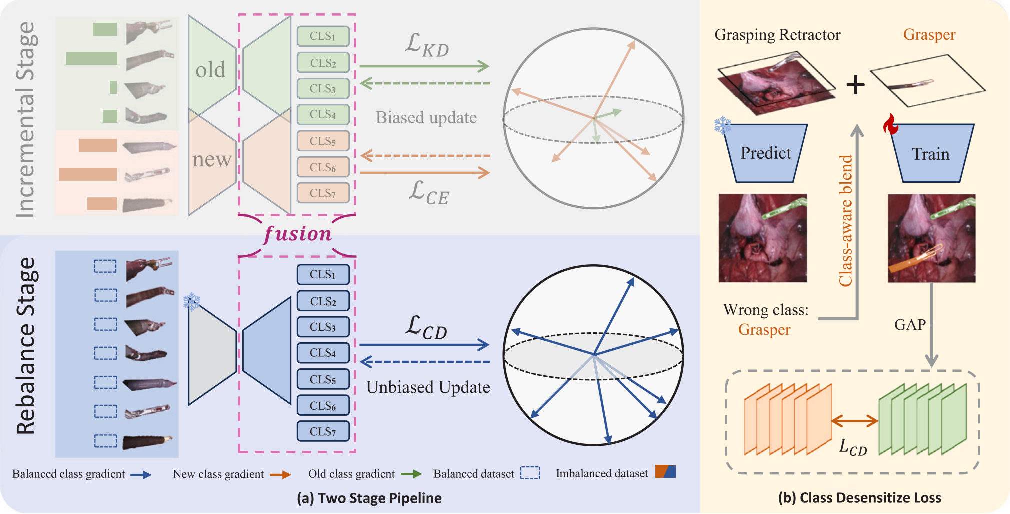
Class imbalance in deep learning is a significant challenge arising from the disproportionate representation of classes in training datasets, which often leads to biased models favoring the majority class. To address this issue, various techniques have been proposed, broadly categorized into data-level, algorithm-level, and hybrid methods. *Data-level methods* (Hensman and Masko, 2015; Lee et al., 2016) focus on modifying the dataset to reduce class imbalance by adjusting the training data distribution, either through duplicating samples from the minority class or removing samples from the majority class. In contrast, *algorithm-level approaches* (Lin, 2017; Wang et al., 2016) involve modifying the learning process or the loss function to account for class imbalance. Recently, researchers have increasingly focused on addressing class imbalance in continual learning. For instance, He et al. (2021) examined the relationship between class imbalance and performance degradation in continual learning. Wu et al. (2019) proposed a bias correction method that targets the fully connected layer, addressing the bias caused by class imbalance and effectively narrowing the performance gap between old and new classes. Furthermore, Liu et al. (2022) tackled class imbalance in long-tailed datasets by introducing a learnable weight-scaling layer to mitigate bias. This approach is particularly relevant to continual learning scenarios and aligns closely with our task of semantic segmentation for surgical instruments.

2.3. Image synthesis

Surgical image scarcity, driven by privacy concerns and workflow complexities, challenges deep learning application development. To address this, synthetic datasets blend open-source surgical instrument images with real or simulated backgrounds, enabling targeted data generation (Wang et al., 2022; Garcia-Peraza-Herrera et al., 2021). However, these methods face limitations such as manual data collection costs and limited source availability. Generative models, particularly GANs, have improved data augmentation by synthesizing surgical images (Zaffino et al., 2021; Shin et al., 2018; Hamghalam et al., 2020; Lee et al., 2019; Rivoir et al., 2021), and unpaired image-to-image translation has enhanced image quality (Pfeiffer et al., 2019). Recently, diffusion models (DMs) have demonstrated superior performance, employing semantic maps for high-quality surgical image synthesis (Zhou et al., 2024) and latent consistency-distilled approaches for unpaired translation (Venkatesh et al., 2024). Advanced methods now generate high-resolution surgical videos from text prompts alone (Cho et al., 2024). Compare to other data augmentation methods like copy-paste (Ghiasi et al., 2021) and cutpaste (Li et al., 2021), image synthesis methods offer the capability to generate entirely new visual contexts and object appearances, create variations that extend beyond basic geometric or color transformations, and potentially produce more realistic and diverse data that better reflects the complexity of surgical environments. Our inpainting-based approach further addresses data limitations by generating realistic, class-balanced datasets.

3. Methodology

The proposed two-stage method addresses a continual learning scenario relevant to real-world clinical applications, where models must



(a) Two Stage Pipeline

(b) Class Desensitize Loss

Fig. 2. Framework of our two-stage class-incremental continual segmentation method. Following the incremental training stage, similar to previous approaches, we introduce (a) an additional rebalance stage using a synthetic balanced dataset. During this stage, (b) a Class Desensitization Loss is applied to reduce class bias introduced in the incremental stage. A Dynamic Weight Fusion, shown in the purple dotted box, is performed between the biased model and the unbiased model to integrate knowledge derived from more diverse real-world data.

adapt incrementally to new classes after deployment. The training process is divided into two stages, as illustrated in Fig. 2. In the incremental stage, similar to existing methods, the process begins with a model that has already been trained on prior classes at step $t-1$. The goal is to incrementally update the model to recognize a new class at step t while preserving its performance on previously learned classes. We introduce an additional rebalance stage to address the class imbalance problem caused during incremental training. In this stage, the model from the incremental stage undergoes a post-training process specifically designed to further enhance performance. Unlike conventional approaches, the rebalance stage balances weight updates while keeping the backbone parameters frozen. This is achieved using a synthetic dataset for fine-tuning, ensuring that core feature extraction remains unaltered. By introducing the rebalance stage, our method effectively addresses challenges in class-incremental learning, significantly improving robustness and stability in robotic surgical settings.

3.1. Problem formulation: Incremental stage

Let us define the dataset at each step as $\mathcal{D}_t = \{x_i, y_i\}$, where x_i represents the data and y_i is the associated label. The label space for the training task T is given as $C_t \cup \{c_b\}$, where C_t contains all classes involved in the task, and c_b corresponds to the background. Typically, classification loss is handled using cross-entropy loss. Due to the disjoint nature of class sets across tasks, i.e., $C_i \cap C_j = \emptyset$, the objectives differ significantly between tasks, often causing catastrophic forgetting. To mitigate this, knowledge distillation is a widely adopted method. In continual learning, distillation is generally categorized into two types: feature-based distillation and logits-based distillation, which can be expressed as:

$$\mathcal{L}_{FD} = \frac{1}{|\mathcal{D}_t|} \sum_{(x_i,y_i)\sim\mathcal{D}_t} \left\|\Psi_{old}(x_i) - \Psi_{new}(x_i)\right\|^2 \tag{1}$$

$$\mathcal{L}_{LD} = \frac{1}{|\mathcal{D}_t|} \sum_{(x_i,y_i)\sim\mathcal{D}_t} KL\left(\Phi_{old}(\Psi_{old}(x_i)), \Phi_{new}(\Psi_{new}(x_i))\right) \tag{2}$$

Ψ_{old}/Φ_{old} and Ψ_{new}/Φ_{new} denote the feature extractor/classifier from the previous old model and incremental new model, respectively, and D is the corresponding dataset.

The softmax cross-entropy loss is used as the classification loss, which is computed as follows:

$$\mathcal{L}_{CE} = \frac{1}{|\mathcal{D}_t|} \sum_{(x_i,y_i)\sim\mathcal{D}_t} -y_i \log\left[\Phi_{new}(x_i)\right] \tag{3}$$

$\Phi_{new}(x_i)$ is the output probability of new classifier.

This approach introduces a significant bias toward new classes due to the imbalance between the abundant samples of new classes and the limited exemplars from old classes. During training, the current model starts with the weights of the previous model, with an additional output channel added to handle segmentation for the new class. These class-specific parameters are highly prone to overfitting on the current class. By examining the weights of the segmentation head for both new and old classes, as illustrated in Fig. 1, it is evident that the bias in class predictions results from uneven weight updates. Our work aims to address and correct the bias in the segmentation head.

3.2. Proposed method: Rebalance stage

Building on our finding that the segmentation head is heavily biased, we propose a simple and effective bias correction method. Our approach introduces additional training stages, as illustrated in Fig. 2. The basic idea is to alleviate the bias by learning a balanced distribution of both old and new classes (shown in Fig. 2). First, we train the feature extractor and segmentation head using the baseline method. In the rebalance stage, we freeze the feature extractor and fine-tune the segmentation head on a synthetic, class-balanced set with a class-desensitizing loss, using a few epochs to address the imbalance between new and old classes through balanced updates. After fine-tuning, we apply a DWF between the original and fine-tuned weights to integrate prior knowledge, enhancing the model's generalization ability. This section details the generation of the synthetic class-balanced set, the class-desensitizing loss, and the DWF method.

3.2.1. Inpaint and blending

A balanced dataset is essential for rebalancing and removing bias. However, in incremental surgical segmentation, acquiring such a dataset is challenging. We pioneer the integration of inpainting and blending techniques into the continual learning setting. Specifically, we utilize an inpainting model to generate background images. The foreground instrument is cropped separately and then overlaid onto the generated background images to address data limitations. We carefully control the sample class distribution to avoid class imbalance, effectively addressing issues of data scarcity and reducing the need for costly, time-intensive labeling.

In surgical settings, backgrounds are highly valuable, but due to privacy concerns and operational constraints, obtaining and storing a

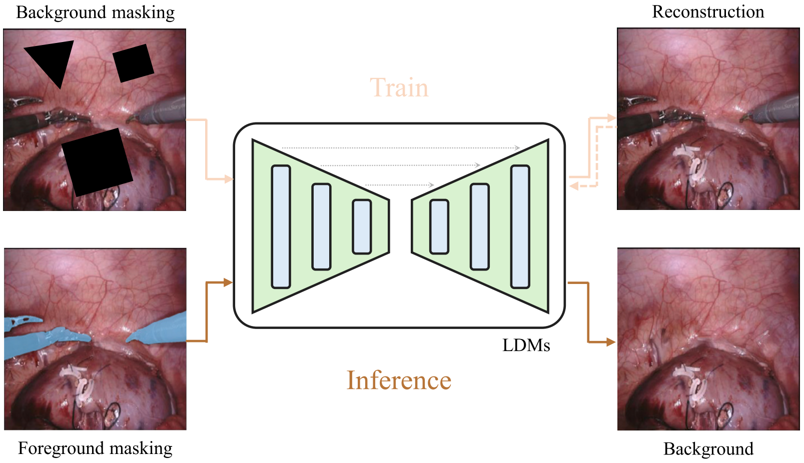


Fig. 3. Construction of a rebalance dataset through inpainting: During training, the model performs a reconstruction task on background-masked regions. At inference time, it is used to remove existing foreground instruments to isolate the background.

large number of background images is not feasible. Available samples are often coupled with specific foreground categories, resulting in an unbalanced dataset. Although recent efforts have used generative models for data synthesis, these models often struggle with high-quality generation due to the detailed nature of surgical backgrounds and the similarity of many surgical instruments in the foreground (Xu et al., 2024). By using inpainting, we can reuse existing backgrounds and flexibly create a balanced dataset. Compared to direct generative models, inpainting provides better detail preservation as it retains more of the original information.

Given the remarkable performance demonstrated by diffusion models recently, we chose to use latent diffusion models (LDM) to generate the background images (Rombach et al., 2022). However, robotic surgical images differ significantly from natural images, so directly applying existing models is not ideal due to domain gaps. To address this, we fine-tune the pre-trained LDM on our surgical dataset. Specifically, we first fine-tune the pre-trained LDM autoencoder, and then train the U-Net component. During training, we mask out the surgical background regions in the images, allowing the model to learn the background inpainting process. Subsequently, during inference, the inpainting model is used to remove the foreground surgical instruments from the given data samples, generating clean background images. The inpainting data samples are illustrated in Fig. 3. The dataset is denoted as $\mathcal{D}_r = \{\mathcal{B}, \mathcal{F}\}$, where $\mathcal{B}$ represents the background obtained through inpainting, and $\mathcal{F} = \{f^i\}$ denotes the foreground instrument, with each f^i corresponding to different class i.

In the surgical setting, various environmental disturbances can lead to performance deviations between training and test datasets. Insufficient model robustness under these conditions may result in prediction errors, potentially compromising procedural safety (Garcia-Peraza-Herrera et al., 2021; Wang et al., 2023a). To address this, we enhance both backgrounds and foregrounds through augmentations such as rotation, scaling, and flipping. In this procedure, we allow up to a specified number of surgical instruments to appear simultaneously in a single background image. We then blend the augmented foreground and background images to generate composite images along with corresponding multi-class masks.

3.2.2. Class desensitization loss

After obtaining clean background images, our next challenge is to use these backgrounds for model training in a way that allows for balanced updates, reducing bias, and improving performance. Additionally, it is also a challenge to avoid introducing a gap between synthetic and real data. As demonstrated in Fig. 1, the average summation and t-SNE have shown that the segmentation heads have a strong bias toward certain classes.

Motivated by contrastive learning (Michieli and Zanuttigh, 2021; Wu et al., 2023), we design the Class Desensitization Loss (CDL) as our objective function, employing contrastive learning with class-aware blending to mitigate the bias introduced in incremental stages. Contrastive learning supports representation learning for both old and new classes by enhancing intra-class compactness and inter-class separation. By enabling the model to learn more robust and discriminative feature representations, CDL also addresses biases that might arise from the synthetic-real domain shift by acting as a corrective mechanism.

We use our existing backgrounds and instrument foregrounds to perform class-aware blending. First, we randomly select an instrument foreground and blend it into the background according to the blending method described above. Then, we use the existing model to predict the label of this foreground. As previously discussed, biased models tend to misclassify certain classes as other similar classes, which may lead to errors in category prediction when applied to the initial blended image. Based on the predicted label, we select a foreground of the predicted class and blend it again. The resulting samples are used for training and loss computation, as shown in Fig. 2(b). The overall algorithm of CDL is shown in Algorithm 1.

Algorithm 1 Class Desensitization Loss

```
Input: Model $\mathcal{F}$; rebalanced dataset $\mathcal{D}_{\text{rebalance}} = \{\mathcal{B}, \mathcal{F}\}$; blending number $N$
Output: Class Desensitization Loss $\mathcal{L}_{\text{CD}}$
Randomly select a foreground object $f_0^i$ and a background $b_0$ from dataset $\mathcal{D}$
Blend $f_0^i$ with $b_0$ to create a blended image $I_0$ and assign it label $Y_0$
Use model $\mathcal{F}$ to predict the label $\hat{Y}_0$ for the foreground in $I_0$
if $\hat{Y}_0$ is the background class then
    Set the new blending class $C \leftarrow i$
else if $\hat{Y}_0 \neq Y_0$ then
    Set the new blending class $C \leftarrow c$, where $c$ is the incorrectly predicted class
end if
for $j = 1$ to $N$ do
    Randomly select a foreground object $f_j^C$
    Blend $f_j^C$ into $I_{j-1}$ to create a blended image $I_j$ with label $Y_j$
end for
Compute the loss $\mathcal{L}_{\text{CD}}(\mathcal{F}(\text{I}_j), Y_j)$
```

The samples contain different classes that can confuse the model, which can be addressed by the CDL using contrastive learning, defined as:

$$\mathcal{L}_{CD} = \frac{1}{2}\left(y \cdot \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 + (1-y) \cdot \max(0, m - \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2)\right) \tag{4}$$

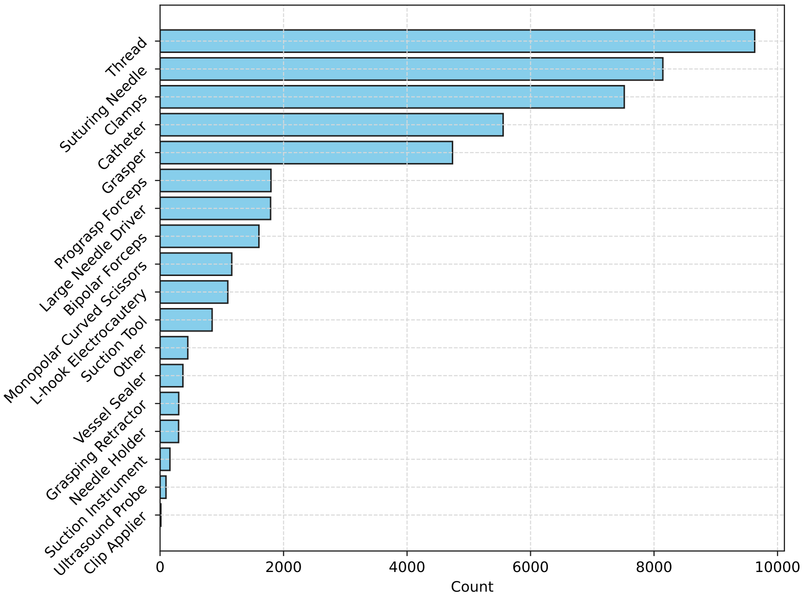


Fig. 4. Illustration of the class distribution in the collected dataset, ordered from the most frequent to the least frequent class.

$$\mathcal{L}_{CE} = -\sum_{c=1}^{C} Y \log(\hat{Y}) \tag{5}$$

$$\mathcal{L}_{BCE} = -[Y \log(\hat{Y}) + (1 - Y)\log(1 - \hat{Y})] \tag{6}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ represent the features of the confused classes, computed via global averaging. The variable y serves as the label to indicate the positive or negative class, Y and $\hat{Y}$ are the predicted probabilities and ground truth, and m represents the margin for the contrastive difference. This loss operates alongside the conventional pixel-wise cross-entropy loss, L_{CE}, and the binary cross-entropy loss, L_{BCE}, offering a complementary approach to improve segmentation performance. The combination is intentional: while pixel-wise cross-entropy loss focuses on predicting the correct label for each sample, CDL targets effective data representation by considering the relationships between different samples.

Thus, the training objective is to optimize the following combined loss:

$$\mathcal{L} = \alpha\mathcal{L}_{CD} + \beta(\mathcal{L}_{CE} + \mathcal{L}_{BCE}) \tag{7}$$

where α, β represent the respective weights for each loss component.

3.2.3. *Dynamic weight fusion*

The primary concept of our Dynamic Weight Fusion (DWF) is to enhance the model's generalization capability while maintaining balanced representations. Although the incremental model may exhibit a significant bias toward certain classes, it often encodes more detailed knowledge about these classes. This is because the training dataset during the incremental stage is typically more diverse than the one used in the rebalance stage. To effectively retain general knowledge while preserving an unbiased representation, we propose the DWF strategy to further improve performance. Specifically, after the rebalance stage, we have two models: one trained during the incremental stage (θ_{ori}) and another after the rebalancing process (θ_{re}). We introduce an additional parameter λ, which serves as a weight fusion factor, enabling us to combine the parameters of these two models in a specified ratio. This ratio acts as a balancing factor. The operation for dynamic weight fusion can thus be expressed as:

$$\theta_{\text{fused}} = \lambda \cdot \theta_{\text{ori}} + (1 - \lambda) \cdot \theta_{\text{re}} \tag{8}$$

This formulation ensures that the model benefits from both the detailed knowledge retained in θ_{ori} and the balanced representation provided by θ_{re}, allowing for improved performance while mitigating biases. While λ serves as a control ratio for parameter magnitudes, we also introduce a constraint on the selection of fusion parameters to mitigate the bias present in θ_{ori}. To address this, we define α, a dynamic threshold parameter, based on the most variable parameters after rebalancing. These parameters are likely to represent general knowledge learned from a large dataset, which is at risk of being lost during rebalancing. The threshold α is computed as follows:

$$\alpha = \min(\text{top-}k(|\theta_{\text{ori}} - \theta_{\text{re}}|)), \tag{9}$$

where top-$k(|\theta_{\text{ori}} - \theta_{\text{re}}|)$ identifies the k-largest differences between θ_{ori} and θ_{re}, capturing the parameters most affected by rebalancing. The overall strategy can be expressed as:

$$\theta_{\text{fused}} = \begin{cases} \theta_{\text{ori}}, & \text{if } |\theta_{\text{ori}} - \theta_{\text{re}}| < \alpha, \\ (1-\lambda)\theta_{\text{ori}} + \lambda\theta_{\text{re}}, & \text{if } |\theta_{\text{ori}} - \theta_{\text{re}}| \geq \alpha, \end{cases} \tag{10}$$

indicating that weight fusion is applied selectively to parameters likely to be significantly influenced by class imbalance. This targeted fusion ensures the preservation of general knowledge while addressing the issues introduced during the incremental training stage.

4. Experiments

4.1. *Implementation*

4.1.1. *Datasets*

We collected a new dataset to assess our method. The motivation behind collecting the new dataset stems from the limitations identified in previous datasets, which include a restricted number of instrument classes and the challenge of class imbalance in real-world surgical workflows. In surgical environments, some instruments are used more frequently than others, leading to highly imbalanced data distributions. Addressing this imbalance is crucial for evaluating the robustness of Continual Learning (CL) models under realistic conditions, where the model must adapt to new classes without access to past data (a scenario prone to catastrophic forgetting).

Our dataset is merged from four public datasets: *EndoVis 2017 Dataset* (Allan et al., 2019), *EndoVis 2018 Dataset* (Allan et al., 2020), *CholecSeg8k* (Hong et al., 2020), and *SAR-RAPR50* (Psychogyios et al., 2023). The *EndoVis 2017 and 2018* datasets primarily focus on the segmentation of robotic instruments in minimally invasive surgeries,

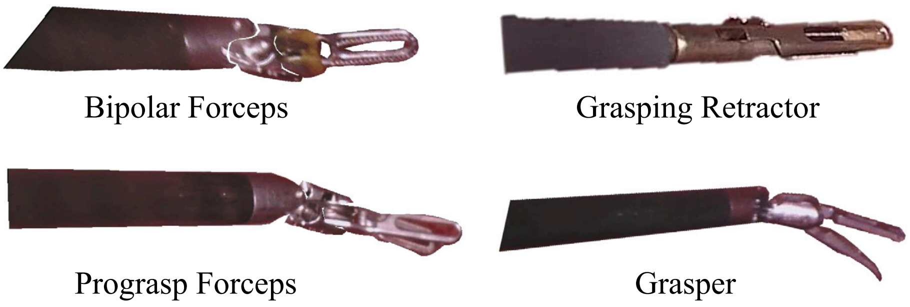


Fig. 5. Demonstration of visually similar samples across different classes in our dataset. We present several surgical instruments with similar appearances, including Bipolar Forceps, Prograsp Forceps, Grasping Retractor, and Grasper.

Table 1
Performance comparison across different continual learning tasks (13-1, 13-5, 9-3, and 17-1, "x-y" denotes that x is the number of old classes and y is the number of incremental classes introduced in each step). Results are reported for methods without continual learning (WO CL), including the old model at time point $t = 0$ and naive fine-tuning (FT), and those with continual learning (W CL). The metrics presented are for base, new, and all classes, with the offline training serving as an upper bound.

Method	13-1 (5 steps)			13-5 (1 step)			9-3 (3 steps)			17-1 (1 step)		
	Base	New	All	Base	New	All	Base	New	All	Base	New	All
Old Model (t = 0)	61.29	0.00	45.16	61.29	0.00	45.16	60.62	0.00	31.90	56.93	0.00	53.93
FT	0.00	17.67	8.53	0.00	32.39	13.71	0.00	10.99	9.52	0.00	5.49	8.77
EWC (Kirkpatrick et al., 2017)	0.00	10.98	5.30	0.00	**52.60**	22.25	0.00	7.65	8.75	0.00	5.47	8.73
LwF (Li and Hoiem, 2017)	1.48	21.82	11.30	14.04	14.23	20.11	3.56	27.42	18.35	23.14	4.64	23.31
LwF-MC (Rebuffi et al., 2017)	0.32	17.48	9.05	10.07	13.49	14.35	5.63	19.31	16.02	10.62	15.18	15.98
ILT (Michieli and Zanuttigh, 2019)	0.92	21.70	11.23	16.10	23.03	23.16	10.77	20.04	19.01	35.51	7.13	35.77
MiB (Cermelli et al., 2020)	7.98	**25.21**	17.11	13.56	19.41	19.52	30.79	7.86	24.98	35.04	7.45	36.79
RCIL (Zhang et al., 2022)	50.08	4.86	40.64	26.22	48.53	42.66	0.38	22.95	22.64	52.32	8.41	50.59
CAT-SD (Xu et al., 2024)	44.70	6.29	37.32	49.50	11.24	41.45	4.95	5.55	9.67	55.31	9.00	52.58
MBS (Park et al., 2024)	30.85	7.64	24.40	46.93	40.54	45.15	35.92	18.16	34.12	46.43	49.41	46.59
NeST (Xie et al., 2024)	29.61	24.08	28.08	39.59	30.05	37.89	18.56	8.26	13.41	45.64	9.44	43.63
PLOP (Douillard et al., 2021)	12.09	19.69	18.49	44.14	23.38	41.47	4.56	18.37	14.90	39.11	8.32	41.06
PLOP+Ours	52.50	18.74	45.99	55.02	23.37	**48.93**	34.63	11.94	28.33	57.19	**19.25**	54.43
Improvement	*+40.41*	*−0.95*	*+27.50*	*+10.88*	*−0.01*	*+7.46*	*+30.07*	*−6.43*	*+13.43*	*+18.08*	*+10.93*	*+13.37*
EWF (Xiao et al., 2023)	54.82	5.12	43.98	50.90	26.83	43.22	35.96	0.00	23.77	57.15	0.05	53.93
EWF+Ours	**62.37**	8.70	**50.13**	**60.83**	20.18	45.62	**63.96**	18.80	**46.68**	**59.01**	8.88	**55.33**
Improvement	*+7.55*	*+3.58*	*+6.15*	*+9.93*	*−6.65*	*+2.40*	*+28.00*	*+18.80*	*+22.91*	*+1.86*	*+8.83*	*+1.40*
Offline	58.00	60.43	60.77	58.00	60.43	60.77	58.00	60.43	60.77	58.00	60.43	60.77

offering detailed pixel-wise annotations for da Vinci robotic tools, including articulating parts such as shafts, jaws, and wrists, as well as surgical devices like ultrasound probes and suturing instruments. *CholecSeg8k* provides segmentation data for key instruments used in laparoscopic cholecystectomy, including graspers and electrocautery tools, alongside complex anatomical backgrounds. *SAR-RARP50* contains segmentation annotations for robotic instruments in the context of radical prostatectomy surgeries. Our newly collected dataset expands the number of instrument classes compared to previous datasets, as shown in Fig. 4. The integrated dataset contains 18 instrument classes, with a highly imbalanced distribution. It includes 66 videos with a total of 26,638 frames, of which 19,938 are in the training set and 6705 are in the test set. The resolution ranges from 1920 to 800 pixels, originating from RGB endoscopic videos. Each frame was annotated instance by instance, with separate annotated masks provided for each object. Additionally, the dataset includes several visually similar classes, resembling real robotic surgical scenes. Fig. 5 presents examples of categories with similar appearances.

4.1.2. Experiments setting

In our surgical instrument segmentation scenario, we assume that the training data from the time point(t = 0) is no longer accessible during the subsequent incremental learning stages. We define four tasks: 13-1 (5 steps), 13-5 (1 step), 9-3 (3 steps), and 17-1 (1 step), all set under the overlapped protocol. "x-y" denotes that x is the number of old classes and y is the number of incremental classes introduced in each step. At each step, the model has access to the new dataset, but previous datasets are not available for further training. Similar to the overlapped protocol described earlier, the training set at each stage may include images containing pixels belonging to classes that are yet to be learned in future steps, with these pixels labeled as background. This setting allows us to evaluate the effectiveness of the incremental learning approach when handling new instrument classes while dealing with the limitations imposed by inaccessible prior data.

4.1.3. Implementation details

We used the DeepLabv3+ architecture (Chen et al., 2018) with the ResNet-101 backbone (He et al., 2016) for segmentation tasks. We compared our proposed method against the following baselines and state-of-the-art solutions: EWC (Kirkpatrick et al., 2017), LwF (Li and Hoiem, 2017), LwF-MC (Rebuffi et al., 2017), ILT (Michieli and Zanuttigh, 2019), MiB (Cermelli et al., 2020), RCIL (Zhang et al., 2022), CAT-SD (Xu et al., 2024), MBS (Park et al., 2024), NeST (Xie et al., 2024), PLOP (Douillard et al., 2021), and EWF (Xiao et al., 2023). Consistent with prior approaches, in-place activated batch normalization was set in the backbone. All input images were resized to 224 × 224 for uniformity. Using the SGD optimizer, we set the initial learn rate as 0.06 and 0.006 for the continual learning steps, and the batch size is set to 12. For the incremental stage, we set the training to 20 epochs per step. Taking the 13-1 (5 steps) setting as an example, PLOP requires approximately 6 h for training, and other baseline methods also require similar training time. The rebalance stage is set to 1 step with 3 epochs, which takes around 0.5 h. The training of the rebalance stage requires around 9 GB of VRAM. All experiments are conducted on two NVIDIA RTX 3090 GPUs.

Table 2
Performance comparison across different datasets for the 13-1 continual learning task. Results are reported for methods without continual learning (WO CL), including the old model at time point $t = 0$ and naive fine-tuning (FT), and those with continual learning (W CL). The metrics include performance on base, new, and all classes, with offline training serving as the upper bound.

Method	CholecSeg8k			EndoVis 2017			EndoVis 2018			SAR-RAPR50		
	Base	New	All	Base	New	All	Base	New	All	Base	New	All
Old Model (t = 0)	62.53	0.00	44.66	83.82	0.00	55.88	73.33	0.00	61.10	61.69	0.00	52.97
FT	0.00	0.00	31.60	0.00	0.00	11.00	0.00	0.00	12.02	0.00	75.05	24.49
EWC (Kirkpatrick et al., 2017)	0.00	0.00	41.26	0.00	0.00	16.63	0.21	0.97	12.91	0.00	**74.82**	23.45
LwF (Li and Hoiem, 2017)	0.00	**78.39**	58.30	0.00	0.44	11.78	0.00	0.05	12.81	0.00	66.23	22.69
LwF-MC (Rebuffi et al., 2017)	0.00	18.40	37.86	0.00	0.00	11.03	0.00	0.00	12.14	0.00	19.56	16.55
ILT (Michieli and Zanuttigh, 2019)	0.00	59.66	51.94	0.00	0.00	11.26	0.00	0.00	12.24	0.00	36.75	18.70
MiB (Cermelli et al., 2020)	47.41	15.80	31.60	22.00	2.75	11.00	14.49	6.21	12.42	16.15	20.93	20.93
RCIL (Zhang et al., 2022)	66.63	31.19	65.31	44.07	21.47	41.91	38.50	22.34	43.85	51.03	0.00	49.75
CAT-SD (Xu et al., 2024)	42.22	53.63	64.44	33.74	3.26	30.36	22.55	2.13	30.35	37.06	33.73	43.82
MBS (Park et al., 2024)	27.20	34.53	41.52	35.54	10.83	37.12	18.93	10.16	23.12	27.96	18.32	23.21
NeST (Xie et al., 2024)	31.23	29.53	40.95	28.79	13.37	29.62	18.02	12.61	25.49	28.28	23.77	26.42
PLOP (Douillard et al., 2021)	0.00	68.52	55.65	14.59	6.18	21.80	13.16	6.90	24.15	0.00	43.28	18.61
PLOP+Ours	59.70	45.69	67.86	46.28	67.98	60.84	**63.68**	25.55	62.91	11.55	12.66	23.82
Improvement	*+59.70*	*−22.83*	*+12.21*	*+31.69*	*+61.80*	*+39.04*	*+50.52*	*+18.65*	*+38.76*	*+11.55*	*−30.62*	*+5.21*
EWF (Xiao et al., 2023)	63.46	67.46	76.46	**57.65**	31.25	52.73	51.67	41.99	56.82	49.53	0.01	48.99
EWF+Ours	**67.12**	70.04	**78.57**	50.31	48.05	**55.36**	58.90	**49.07**	**63.00**	**51.68**	0.02	**50.84**
Improvement	*+3.66*	*+2.58*	*+2.11*	*−7.34*	*+16.80*	*+2.63*	*+7.23*	*+7.08*	*+6.18*	*+2.15*	*+0.01*	*+1.85*
Offline	64.68	74.97	79.59	64.47	65.33	59.10	50.11	56.55	49.56	77.91	67.15	58.76

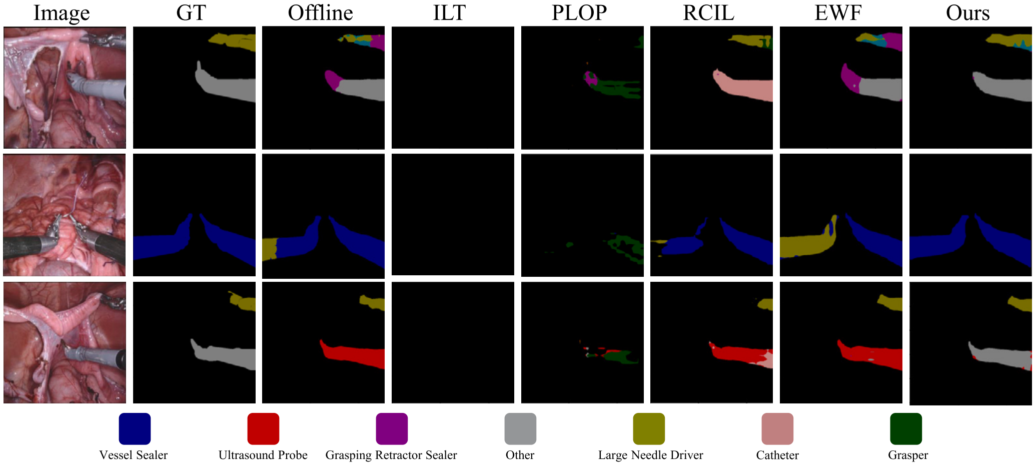


Fig. 6. Visual comparison with different state-of-the-art methods (Task 13-1 on our surgical dataset).

The model testing and weight initialization followed the same approach as outlined earlier. Specifically, the best-performing model from each step was selected based on the highest mean Intersection over Union (IoU) achieved on the validation set. For each continual learning step, the model was initialized with the weights of the previously trained model, ensuring that knowledge from earlier steps was retained. During the continual learning step, the old model acted as a teacher, providing logits for distillation loss calculation, while the model was optimized using both cross-entropy loss and distillation losses. This approach enabled us to learn new classes incrementally while maintaining performance in previously learned classes.

4.2. Experimental results

We conduct extensive experiments and evaluations to validate our proposed approach and investigate the effectiveness of privacy-preserving continual learning in robotic surgery. Our method is compared against several state-of-the-art continual learning approaches, which can be categorized into four groups. One category includes approaches based on penalty computation, such as EWC (Kirkpatrick et al., 2017), which employ strategies to compute the importance of parameters for old classes, safeguarding those critical parameters to prevent forgetting. The second category consists of methods based on knowledge distillation, originally designed for continual image classification, including LwF (Li and Hoiem, 2017) and LwF-MC (Rebuffi et al., 2017). These methods have demonstrated impressive performance when applied to semantic segmentation tasks. The third category focuses on approaches specifically designed for continual semantic segmentation, such as ILT (Michieli and Zanuttigh, 2019), which combines both feature- and output-level knowledge distillation, and MiB (Cermelli et al., 2020), which relies solely on output-level knowledge distillation to address the background shift problem. Recently, some weight fusion methods have been proposed to improve knowledge transfer and enhance model robustness, including RCIL (Zhang et al., 2022) and EWF (Xiao et al., 2023). Additionally, we record the results of the fine-tuning approach (FT) as the lower bound for segmentation performance, since FT updates the model with new classes without any additional constraints, leading to the most severe catastrophic forgetting. The upper bound is represented by the results from the offline training approach (Offline), where images and annotations for all classes are available during training. To ensure a fair comparison, we implement all competitors using the same base segmentation network, as well as identical computing environments and common hyperparameter settings.

In Tables 1 and 2, we present a statistical comparison of our method with existing approaches across tasks 13-1, 13-5, 9-3, and 17-1 on our

Table 3
Ablation study on the Class Desensitization Loss and dynamic weight fusion in our rebalance stage.

Setting	$\mathcal{L}_{CD}$	WF	13-1	13-5	9-3	17-1
Incremental stage	–	–	43.98	43.22	23.77	53.93
Rebalance stage	✗	✗	45.88	45.29	41.38	50.09
	✓	✗	49.83	43.06	45.75	54.64
	✓	✓	**50.13**	**45.62**	**46.68**	**55.33**

surgical dataset, as well as task 13-5 on the four-component dataset. Our observations indicate that, regardless of the method employed, old class forgetting remains a challenge when learning new classes. Furthermore, learning new classes can be negatively impacted when methods impose excessive constraints. While weight fusion methods maintain base class performance, they limit the plasticity required for learning new classes. By addressing the class bias between base and new classes, our method not only preserves the knowledge of base classes but also enhances their performance. For instance, on 13-1 setting, our algorithm achieves performance gains of 27.5% and 6.15% mIoU for PLOP and EWF, respectively, by effectively mitigating bias. This mitigation further improves the performance of the base class by 40.41% and 7.55%, respectively, while the slight decrease in the performance of the new class may be attributed to the reduction in predictable pixels. The differences in improvements between the two baseline methods are primarily because PLOP (Douillard et al., 2021) relies on feature distillation to limit representation capacity, while EWF (Xiao et al., 2023) employs a more robust weight fusion technique. The distinction in their principles leads to differences in plasticity, which results in varying degrees of performance improvement. Notably, as continual learning progresses, the model updates increasingly favor new classes, resulting in greater bias. This trend further highlights the effectiveness of our method in addressing such challenges. Fig. 1 provides evidence for the bias mitigation capabilities of our proposed framework. The increased uniformity in the distribution of the summation average and t-SNE embeddings demonstrates a significant reduction in class bias.

These findings underscore the robustness and effectiveness of our approach in mitigating bias, preserving base class knowledge, and improving overall performance across diverse tasks. Fig. 6 shows the visual comparison with the offline and partial competitors. We observe that our method consistently provides better predictions for Vessel Sealer, Other, Ultrasound Probe, Grasping Retractor Sealer, Large Needle Driver, Catheter, and Grasper, each represented by different colors. Additionally, our predictions are comparable to those of the best-performing offline method, as evidenced by fewer instances of over-predicting and under-predicting, as well as reduced confusion among different instrument classes.

4.3. Ablation study

In this part, we demonstrate and analyze the effectiveness of the Class Desensitization Loss and its blending method. We also provide an analysis of our Dynamic Weight Fusion Strategy. We use 13-1 (5 steps) for the ablation experiments. Table 3 presents the ablation study on the impact of each proposed module, including the rebalance stage, Class Desensitization Loss, and Dynamic Weight Fusion. With the inclusion of the rebalance stage, performance improves by 1.90%, 2.07%, and 17.61% in the 13-1, 13-5, and 9-3 tasks, respectively. This demonstrates that a simple bias adjustment step can significantly enhance the performance of continual segmentation models. However, in the 17-1 task, a slight decrease in performance is observed, primarily due to the fewer continual learning steps, where incremental classes have not introduced significant class bias. This effect is mitigated by the Class Desensitization Loss, which results in a 4.55% improvement in the 17-1 task by clarifying the class boundary. We also perform a quality analysis to ensure that the bias has been mitigated, as demonstrated by the t-SNE feature in Fig. 7, accompanied by qualitative cluster metrics in Table 4. The DB Index, CH Index, and WCSS have significantly improved, showing a more uniform distribution, which serves as evidence of the efficiency. Additionally, the Dynamic Weight Fusion strategy proves to be a versatile method for combining previous models, enhancing the generalization ability on real samples. Ultimately, our method strikes a balance between enabling the learning of new classes and maintaining the ability to distinguish base classes.

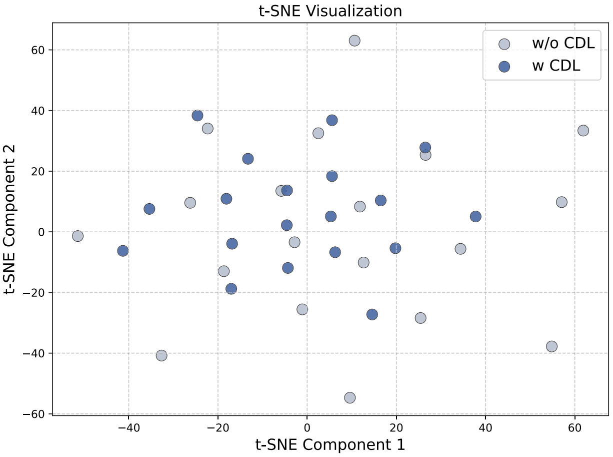


Fig. 7. Ablation study on the class desensitization loss via the t-SNE visualization on the feature space.

Table 4
Ablation study on the Class Desensitization Loss using cluster metrics on the t-SNE embedding.

$\mathcal{L}_{CD}$	DB Index ↓	CH Index ↑	WCSS ↓
✗	3.56	1.13	31445.56
✓	**2.13**	**3.02**	**11 366.84**

Table 5
Comparisons of our ablation study on different blending methods and sample numbers.

Method	Number	Base classes	New classes	All classes
Random	100	44.20	2.64	36.02
	300	48.39	5.43	39.66
	500	50.83	7.69	41.93
	1000	54.08	7.33	44.05
Class-aware	100	46.65	7.02	38.78
	300	49.38	4.50	40.09
	500	58.77	6.21	46.97
	1000	**62.02**	**8.49**	**49.83**

4.3.1. Class-aware blending

In Table 5, we demonstrate the performance of different blending methods and sample numbers. We compare random blending with class-aware blending, verifying the effectiveness of our algorithm in alleviating model bias caused by class imbalance through the use of synthetic class samples. The results show that smaller sample sizes (e.g., 100 samples) can negatively impact model performance, particularly for new classes. However, as the number of samples increases, the performance improves significantly, with the class-aware method outperforming random blending. The class-aware blending method leads to a better balance between base and new classes, resulting in enhanced overall performance.

Table 6 illustrates that directly cropping the background from the sample would reduce the effectiveness of the Class Desensitization Loss enhancement. Moreover, in Table 7, we explore the synthesis quality

Table 6
Comparisons of different background synthetic methods. Crop refers to directly cropping the background from samples.

Method	Base classes	New classes	All classes
Crop	56.10	6.47	45.21
Generate (DDPM, Ho et al., 2020)	44.02	5.79	36.49
Inpaint (LDM, Rombach et al., 2022)	**62.37**	**8.70**	**50.13**

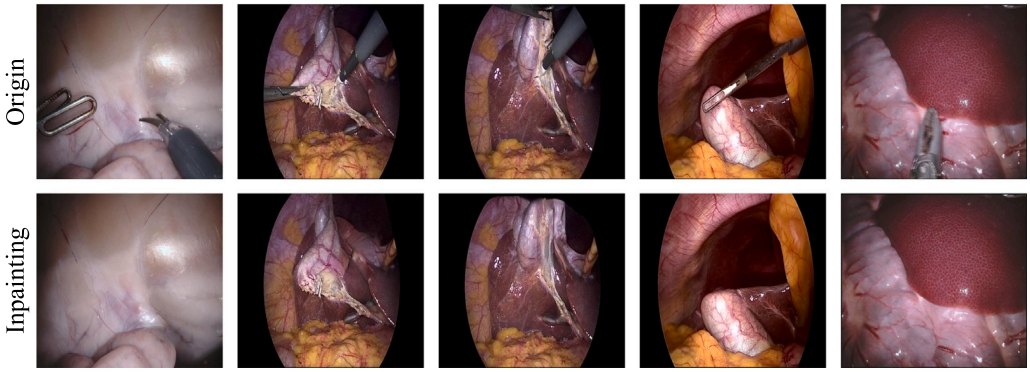


Fig. 8. Visualization samples of our synthetic background inpainting.

Table 7
Synthesis quality comparison of different generation methods.

Method	FID ↓	IS ↑	PSNR ↑
DDPM (Ho et al., 2020)	266.77	0.99	11.25
StyleGAN-XL (Sauer et al., 2022)	138.46	1.09	12.16
FLUX.1-Fill-dev (Labs, 2024)	158.60	**1.86**	11.06
LDM (**Ours**, Rombach et al., 2022)	**48.26**	1.30	**12.40**

Table 8
Comparisons of our ablation study on the weight of the Class Desensitization Loss (Task 13-1 on our surgical dataset).

α	β	Base classes	New classes	All classes
1.0	0.0	56.54	7.76	45.92
0.0	1.0	61.33	**9.90**	49.65
1.0	**1.0**	**61.98**	8.74	**49.84**
1.0	2.0	59.98	5.79	47.74
2.0	1.0	61.65	7.82	49.36
2.0	2.0	57.22	4.64	45.48

of four generative models (DDPM, Ho et al., 2020, StyleGAN-XL, Sauer et al., 2022, FLUX.1-Fill-dev, Labs, 2024, and LDM, Rombach et al., 2022) applied to our inpainting task and finally select LDM (Rombach et al., 2022) as our inpainting model. LDM exhibits strong performance compared to other methods, as other methods mostly rely on cropped patches instead of using the latent space, thus losing completeness. Fig. 8 presents visualization examples of our inpainting solution, and Fig. 9 provides the qualitative comparison of different synthetic methods. Additionally, the performance of the generation method significantly degrades due to its low quality. In contrast, our inpainting method demonstrates a substantial performance improvement. Besides, Table 8 shows the performance comparison under different loss weight configurations, specifically varying the values of α and β. When $\alpha = 1.0$ and $\beta = 1.0$, the model performs best on base classes and overall classes, leading to an improved performance of 3.96%.

In summary, our class-aware blending strategy combines foreground and background through a simple overlay. While this approach may introduce certain data inconsistencies and biases due to the lack of effective modeling between foreground and background, it has already proven effective in improving the performance of downstream segmentation tasks. Our work does not focus on developing generative models specifically for surgical applications. Instead, the generative model can serve as a simple and effective solution to address the downstream continual segmentation task in the surgical domain. Furthermore, our generation-based approach can also enhance privacy protection and improve data scalability (Wang et al., 2022, 2023a).

4.3.2. Fusion strategy

In Table 9, we demonstrate the performance of different fusion strategies and their corresponding parameter selections. The Replace method involves directly selecting the top-k parameter values from the previous model to rebalance the current model. However, this approach results in a performance decline, as it can reintroduce biased parameters into the model. For example, when top-k = 1%, the performance drops to 45.88% for all classes, with a notable decrease in base class accuracy (55.98%). As k increases (e.g., top-k = 50%), the performance continues to degrade further (43.80% overall), illustrating the detrimental effect of bringing biased parameters back into the model. In contrast, our Weighted Fusion strategy balances the knowledge from the previous model while maintaining the model's state unbiased. This method shows a clear improvement over the Replace strategy. Specifically, when using top-k = 10% and $\lambda = 0.01$, the model achieves the highest overall performance at 50.13%, with a 0.3% improvement over the baseline (49.83%). The weighted fusion approach also allows for better retention of both base and new class performance, as evidenced by the slight improvements in both categories. Notably, limiting the number of fused parameters (e.g., top-k = 10%) helps prevent performance degradation, keeping the model less prone to bias. These results emphasize the effectiveness of the weighted fusion strategy in enhancing model performance, offering a robust solution for continual learning scenarios where knowledge from previous models must be incorporated without introducing bias.

5. Discussion and conclusion

In this work, we propose a novel two-stage pipeline to address the critical challenge of class imbalance in continual segmentation for robotic surgical scenes, a challenge exacerbated by privacy constraints and the complexity of surgical workflows. Our SurgCSS framework includes the class-aware blending with inpainting to generate a class-balanced dataset by separating instrument foregrounds from surgical

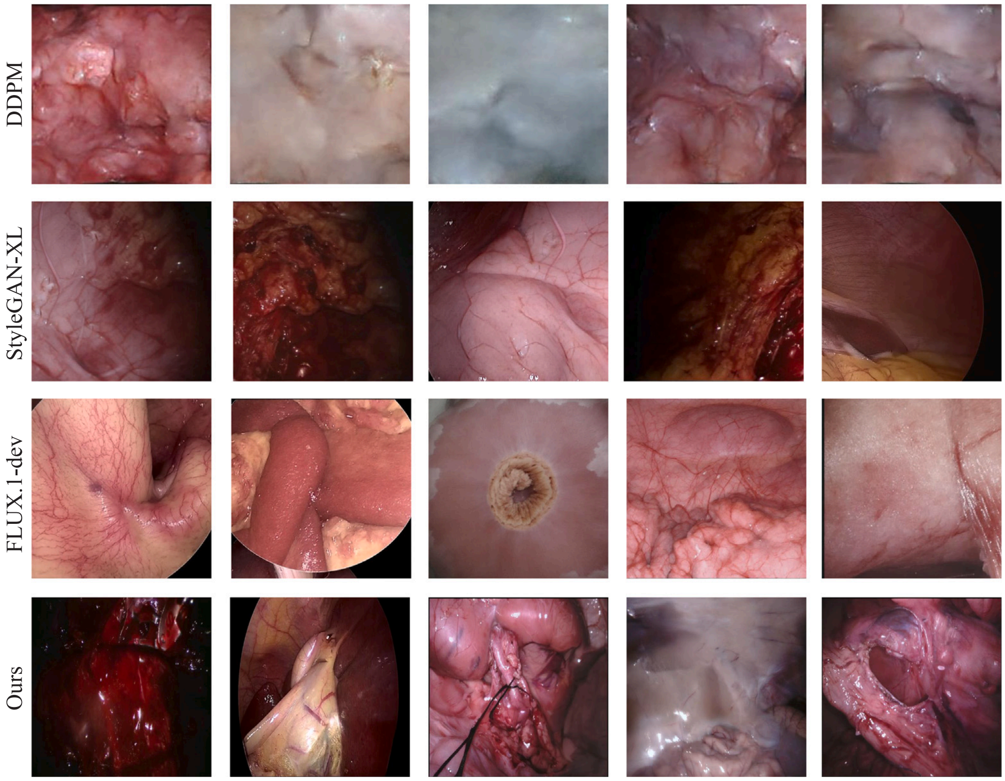


Fig. 9. Qualitative comparison of different generation methods.

Table 9
Comparisons of our ablation study on the fusion parameter selection top-*k* and merged weight *λ* (Task 13-1 on our surgical dataset).

Method	top-*k*	*λ*	Base classes	New classes	All classes
None	–	–	62.02	8.49	49.83
Replace	1%	–	55.98	9.31	45.88
	10%	–	56.38	5.87	45.25
	30%	–	54.76	5.32	44.00
	50%	–	54.69	5.05	43.80
Weighted	10%	0.1	**62.41**	8.13	49.97
	10%	0.3	62.34	8.11	49.92
	30%	0.05	61.34	7.94	49.19
	50%	0.1	58.56	7.16	47.09
	10%	0.01	62.37	**8.70**	**50.13**

backgrounds, alongside the CDL based on contrastive learning to reduce confusion between similar classes. Additionally, a DWF strategy combines incremental and rebalance models to enhance class representation and improve overall segmentation accuracy. Experimental results demonstrate that (i) there is significant performance degradation due to class imbalance when directly applying conventional segmentation techniques; (ii) existing continual learning methods, while alleviating catastrophic forgetting, fail to address class imbalance effectively in surgical contexts; and (iii) our two-stage pipeline successfully balances performance across old and new classes, outperforming state-of-the-art continual learning methods with average 11.82% improvement in robotic surgical scenarios.

This is a fundamental work focusing on resolving the combined issues of class imbalance, catastrophic forgetting, and privacy concerns in the context of robotic surgical segmentation. The framework can be extended to other medical imaging tasks (e.g., organ segmentation or lesion detection), as these tasks often share similar challenges such as imbalanced class distributions and insufficient data. To address these challenges, our proposed inpainting synthetic data generation method and Class Desensitization Loss offer flexible, generalized solutions that reduce class bias and improve model robustness across various applications in the medical domain. However, our proposed framework still has some limitations: (i) Although the final deployed segmentation model remains identical to the previous model without introducing additional parameters, it still requires offline data generation and extra training for the generative model, which significantly increases the training cost. (ii) Any defects or artifacts in the synthetic data can negatively affect the model's performance, particularly when distinguishing between visually similar instrument classes. (iii) The generative model, trained on the background characteristics of the initial dataset, may require retraining or substantial updates to effectively generate data for significantly different surgical domains, leading to extra computational costs. Therefore, future work should rigorously validate the feasibility, robustness, and applicability of the proposed continual learning approach and generative framework in real-world clinical environments, including addressing lightweight deployment and real-time inference challenges in actual hospital settings. Additionally, our continual learning approach can be effectively applied to adapt visual foundation models, such as SAM (Kirillov et al., 2023), to various downstream medical applications, mitigating catastrophic forgetting of the originally learned computer vision patterns.

CRediT authorship contribution statement

Shifang Zhao: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Long Bai:** Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Kun Yuan:** Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Feng Li:** Writing – review & editing, Visualization, Software, Investigation, Formal analysis. **Jieming Yu:** Validation, Investigation, Data curation. **Wenzhen Dong:** Validation, Investigation, Data curation. **Guankun Wang:** Writing – review & editing, Validation, Software, Methodology, Formal analysis.

Mobarakol Islam: Writing – review & editing, Investigation, Formal analysis. **Nicolas Padoy:** Writing – review & editing, Supervision, Resources, Formal analysis, Conceptualization. **Nassir Navab:** Writing – review & editing, Supervision, Resources, Investigation, Funding acquisition. **Hongliang Ren:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of using Generative AI and AI-assisted technologies

The authors have used ChatGPT to improve the readability and language of the work. The process of using ChatGPT was done with human oversight and control. The authors have carefully reviewed and edited the results. The authors have not used ChatGPT to provide citations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Hong Kong Research Grants Council GRF 14204524, GRF 14203323, GRF 14216022, GRF 14211420, CRF C4026-21GF; Guangdong Basic & Applied Research Fund Regional Joint Fund Project 2021B1515120035 (B.02.21.00101); Shenzhen-Hong Kong-Macau Technology Research Programme (Type C) STIC Grant SGDX20210823103535014 (202108233000303).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103728.

References

Alabi, O., Vercauteren, T., Shi, M., 2025. Multitask learning in minimally invasive surgical vision: A review. Med. Image Anal. 103480.

Allan, M., Kondo, S., Bodenstedt, S., Leger, S., Kadkhodamohammadi, R., Luengo, I., Fuentes, F., Flouty, E., Mohammed, A., Pedersen, M., et al., 2020. 2018 robotic scene segmentation challenge. arXiv preprint arXiv:2001.11190.

Allan, M., Shvets, A., Kurmann, T., Zhang, Z., Duggal, R., Su, Y.-H., Rieke, N., Laina, I., Kalavakonda, N., Bodenstedt, S., et al., 2019. 2017 robotic instrument segmentation challenge. arXiv preprint arXiv:1902.06426.

Anand, D., Das, B., Dangeti, V., Jerald, A., Mullick, R., Patil, U., Sharma, P., Sudhakar, P., 2024. Label sharing incremental learning framework for independent multi-label segmentation tasks. arXiv preprint arXiv:2411.11105.

Ayromlou, S., Tsang, T., Abolmaesumi, P., Li, X., 2024. CCSI: Continual Class-Specific impression for data-free class incremental learning. Med. Image Anal. 103239.

Bai, L., Islam, M., Ren, H., 2023. Revisiting distillation for continual learning on visual question localized-answering in robotic surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 68–78.

Cermelli, F., Mancini, M., Bulo, S.R., Ricci, E., 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. pp. 9233–9242.

Cha, S., Yoo, Y., Moon, T., et al., 2021. Ssul: Semantic segmentation with unknown label for exemplar-based class-incremental learning. Adv. Neural Inf. Process. Syst. 34, 10919–10930.

Chaudhry, A., Dokania, P.K., Ajanthan, T., Torr, P.H., 2018. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 532–547.

Chen, Z., Zhang, Z., Guo, W., Luo, X., Bai, L., Wu, J., Ren, H., Liu, H., 2024. Asi-seg: Audio-driven surgical instrument segmentation with surgeon intention understanding. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 13773–13779.

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

Cheng, Z., Guo, J., Zhang, J., Qi, L., Zhou, L., Shi, Y., Gao, Y., 2025. Mamba-sea: A mamba-based framework with Global-to-local sequence augmentation for generalizable medical image segmentation. arXiv preprint arXiv:2504.17515.

Cho, J., Schmidgall, S., Zakka, C., Mathur, M., Kaur, D., Shad, R., Hiesinger, W., 2024. SurGen: Text-guided diffusion model for surgical video generation. arXiv preprint arXiv:2408.14028.

Colleoni, E., Psychogyios, D., Van Amsterdam, B., Vasconcelos, F., Stoyanov, D., 2022. SSIS-Seg: Simulation-supervised image synthesis for surgical instrument segmentation. IEEE Trans. Med. Imaging 41 (11), 3074–3086.

Colleoni, E., Stoyanov, D., 2021. Robotic instrument segmentation with image-to-image translation. IEEE Robot. Autom. Lett. 6 (2), 935–942.

Douillard, A., Chen, Y., Dapogny, A., Cord, M., 2021. Plop: Learning without forgetting for continual semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4040–4050.

Douillard, A., Cord, M., Ollion, C., Robert, T., Valle, E., 2020. Podnet: Pooled outputs distillation for small-tasks incremental learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Springer, pp. 86–102.

Elskhawy, A., Lisowska, A., Keicher, M., Henry, J., Thomson, P., Navab, N., 2020. Continual class incremental learning for ct thoracic segmentation. In: MICCAI Workshop on Domain Adaptation and Representation Transfer. Springer, pp. 106–116.

Fang, K., Zhang, A., Gao, G., Jiao, J., Liu, C.H., Wei, Y., 2025. CoMBO: Conflict mitigation via branched optimization for class incremental segmentation. arXiv preprint arXiv:2504.04156.

Garcia-Peraza-Herrera, L.C., Fidon, L., D'Ettorre, C., Stoyanov, D., Vercauteren, T., Ourselin, S., 2021. Image compositing for segmentation of surgical tools without manual annotations. IEEE Trans. Med. Imaging 40 (5), 1450–1460.

Ge, J., Zhang, B., Liu, A., Phan, M.H., Chen, Q., Shu, Y., Zhao, Y., 2024. CIT: Rethinking Class-incremental semantic segmentation with a class independent transformation. arXiv preprint arXiv:2411.02715.

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E.D., Le, Q.V., Zoph, B., 2021. Simple copy-paste is a strong data augmentation method for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2918–2928.

Guo, D., Ji, Z., Su, Y., Zheng, D., Guo, H., Wang, P., Yan, K., Wang, Y., Yu, Q., Li, Z., et al., 2025. A continual Learning-driven model for accurate and generalizable segmentation of clinically comprehensive and Fine-grained Whole-body anatomies in CT. arXiv preprint arXiv:2503.12698.

Hamghalam, M., Lei, B., Wang, T., 2020. High tissue contrast MRI synthesis using multi-stage attention-GAN for segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34, pp. 4067–4074.

He, C., Wang, R., Chen, X., 2021. A tale of two cils: The connections between class incremental learning and class imbalanced learning, and beyond. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3559–3569.

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

Hensman, P., Masko, D., 2015. The impact of imbalanced training data for convolutional neural networks. In: Degree Project in Computer Science. KTH Royal Institute of Technology.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hong, W.-Y., Kao, C.-L., Kuo, Y.-H., Wang, J.-R., Chang, W.-L., Shih, C.-S., 2020. Cholecseg8k: a semantic segmentation dataset for laparoscopic cholecystectomy based on cholec80. arXiv preprint arXiv:2012.12453.

Ji, Z., Guo, D., Wang, P., Yan, K., Lu, L., Xu, M., Wang, Q., Ge, J., Gao, M., Ye, X., et al., 2023. Continual segment: Towards a single, unified and non-forgetting continual segmentation model of 143 whole-body organs in ct scans. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21140–21151.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al., 2017. Overcoming catastrophic forgetting in neural networks. Proc. Natl. Acad. Sci. 114 (13), 3521–3526.

Labs, B.F., 2024. FLUX. https://github.com/black-forest-labs/flux.

Lee, K., Choi, M.-K., Jung, H., 2019. Davincigan: Unpaired surgical instrument translation for data augmentation. In: International Conference on Medical Imaging with Deep Learning. PMLR, pp. 326–336.

Lee, H., Park, M., Kim, J., 2016. Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In: 2016 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 3713–3717.

Li, Z., Hoiem, D., 2017. Learning without forgetting. IEEE Trans. Pattern Anal. Mach. Intell. 40 (12), 2935–2947.

Li, C.-L., Sohn, K., Yoon, J., Pfister, T., 2021. Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9664–9674.

Li, W., Zhang, Y., Zhou, H., Yang, W., Xie, Z., He, Y., 2025. CLMS: Bridging domain gaps in medical imaging segmentation with source-free continual learning for robust knowledge transfer and adaptation. Med. Image Anal. 100, 103404.

Lin, T., 2017. Focal loss for dense object detection. arXiv preprint arXiv:1708.02002.

Liu, J., Guo, X., Yuan, Y., 2021. Graph-based surgical instrument adaptive segmentation via domain-common knowledge. IEEE Trans. Med. Imaging 41 (3), 715–726.

Liu, X., Hu, Y.-S., Cao, X.-S., Bagdanov, A.D., Li, K., Cheng, M.-M., 2022. Long-tailed class incremental learning. In: European Conference on Computer Vision. Springer, pp. 495–512.

Lou, A., Tawfik, K., Yao, X., Liu, Z., Noble, J., 2023. Min-max similarity: A contrastive semi-supervised deep learning network for surgical tools segmentation. IEEE Trans. Med. Imaging 42 (10), 2832–2841.

Maier-Hein, L., Eisenmann, M., Sarikaya, D., März, K., Collins, T., Malpani, A., Fallert, J., Feussner, H., Giannarou, S., Mascagni, P., et al., 2022. Surgical data science–from concepts toward clinical translation. Med. Image Anal. 76, 102306.

Michieli, U., Zanuttigh, P., 2019. Incremental learning techniques for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

Michieli, U., Zanuttigh, P., 2021. Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1114–1124.

Nwoye, C.I., Yu, T., Sharma, S., Murali, A., Alapatt, D., Vardazaryan, A., Yuan, K., Hajek, J., Reiter, W., Yamlahi, A., et al., 2023. CholecTriplet2022: Show me a tool and tell me the triplet—An endoscopic vision challenge for surgical action triplet detection. Med. Image Anal. 89, 102888.

Park, G., Moon, W., Lee, S., Kim, T.-Y., Heo, J.-P., 2024. Mitigating background shift in Class-Incremental semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 71–88.

Pfeiffer, M., Funke, I., Robu, M.R., Bodenstedt, S., Strenger, L., Engelhardt, S., Roß, T., Clarkson, M.J., Gurusamy, K., Davidson, B.R., et al., 2019. Generating large labeled data sets for laparoscopic image processing tasks using unpaired image-to-image translation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22. Springer, pp. 119–127.

Psychogyios, D., Colleoni, E., Van Amsterdam, B., Li, C.-Y., Huang, S.-Y., Li, Y., Jia, F., Zou, B., Wang, G., Liu, Y., et al., 2023. Sar-rarp50: Segmentation of surgical instrumentation and action recognition on robot-assisted radical prostatectomy challenge. arXiv preprint arXiv:2401.00496.

Ranem, A., González, C., dos Santos, D.P., Bucher, A.M., Othman, A.E., Mukhopadhyay, A., 2024. Continual atlas-based segmentation of prostate MRI. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7563–7572.

Razzak, M.I., Naz, S., Zaib, A., 2018. Deep learning for medical image processing: Overview, challenges and the future. Classif. BioApps: Autom. Decis. Mak. 323–350.

Rebuffi, S.-A., Kolesnikov, A., Sperl, G., Lampert, C.H., 2017. Icarl: Incremental classifier and representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2001–2010.

Rivoir, D., Pfeiffer, M., Docea, R., Kolbinger, F., Riediger, C., Weitz, J., Speidel, S., 2021. Long-term temporally consistent unpaired video translation from simulated surgical 3d data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3343–3353.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695.

Sadegheih, Y., Kumari, P., Merhof, D., 2025. Modality-Independent brain lesion segmentation with Privacy-aware continual learning. arXiv preprint arXiv:2503.20326.

Sauer, A., Schwarz, K., Geiger, A., 2022. Stylegan-xl: Scaling stylegan to large diverse datasets. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10.

Shin, H., Lee, J.K., Kim, J., Kim, J., 2017. Continual learning with deep generative replay. Adv. Neural Inf. Process. Syst. 30.

Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3. Springer, pp. 1–11.

Venkatesh, D.K., Rivoir, D., Pfeiffer, M., Speidel, S., 2024. Surgical-CD: Generating surgical images via unpaired image translation with latent consistency diffusion models. arXiv preprint arXiv:2408.09822.

Wang, G., Bai, L., Wu, Y., Chen, T., Ren, H., 2023b. Rethinking exemplars for continual semantic segmentation in endoscopy scenes: Entropy-based mini-batch pseudo-replay. Comput. Biol. Med. 165, 107412.

Wang, A., Islam, M., Xu, M., Ren, H., 2022. Rethinking surgical instrument segmentation: A background image can be all you need. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 355–364.

Wang, A., Islam, M., Xu, M., Ren, H., 2023a. Generalizing surgical instruments segmentation to unseen domains with One-to-Many synthesis. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4608–4614.

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J., 2016. Training deep neural networks on imbalanced data sets. In: 2016 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 4368–4374.

Wang, G., Ren, T.-A., Lai, J., Bai, L., Ren, H., 2023c. Domain adaptive sim-to-real segmentation of oropharyngeal organs. Med. Biol. Eng. Comput. 61 (10), 2745–2755.

Wang, L., Zhang, X., Su, H., Zhu, J., 2024. A comprehensive survey of continual learning: theory, method and application. IEEE Trans. Pattern Anal. Mach. Intell..

Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., Fu, Y., 2019. Large scale incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 374–382.

Wu, C., Herranz, L., Liu, X., Van De Weijer, J., Raducanu, B., et al., 2018. Memory replay gans: Learning to generate new categories without forgetting. Adv. Neural Inf. Process. Syst. 31.

Wu, H., Wang, Z., Zhao, Z., Chen, C., Qin, J., 2023. Continual nuclei segmentation via prototype-wise relation distillation and contrastive learning. IEEE Trans. Med. Imaging.

Xiao, J.-W., Zhang, C.-B., Feng, J., Liu, X., van de Weijer, J., 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. pp. 7204–7213.

Xie, Z., Lu, H., Xiao, J.-w., Wang, E., Zhang, L., Liu, X., 2024. Early preparation pays off: New classifier pre-tuning for class incremental semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 183–201.

Xu, M., Islam, M., Bai, L., Ren, H., 2024. Privacy-Preserving synthetic continual semantic segmentation for robotic surgery. IEEE Trans. Med. Imaging.

Yang, J., Wu, Y., Cen, J., Huang, W., Wang, H., Zhang, J., 2024. Continual learning for segment anything model adaptation. arXiv preprint arXiv:2412.06418.

Yin, H., Feng, T., Lyu, F., Shang, F., Liu, H., Feng, W., Wan, L., 2025. Beyond background shift: Rethinking instance replay in continual semantic segmentation. arXiv preprint arXiv:2503.22136.

Yu, J., Bai, L., Wang, G., Wang, A., Yang, X., Gao, H., Ren, H., 2024a. Adapting sam for surgical instrument tracking and segmentation in endoscopic submucosal dissection videos. arXiv preprint arXiv:2404.10640.

Yu, X., Fang, Y., Zhao, Y., Wei, Y., 2025. Ipseg: Image posterior mitigates semantic drift in Class-Incremental segmentation. arXiv preprint arXiv:2502.04870.

Yu, J., Wang, A., Dong, W., Xu, M., Islam, M., Wang, J., Bai, L., Ren, H., 2024b. Sam 2 in robotic surgery: An empirical evaluation for robustness and generalization in surgical video segmentation. arXiv preprint arXiv:2408.04593.

Zaffino, P., Marzullo, A., Moccia, S., Calimeri, F., De Momi, E., Bertucci, B., Arcuri, P.P., Spadea, M.F., 2021. An open-source covid-19 ct dataset with automatic lung tissue classification for radiomics. Bioengineering 8 (2), 26.

Zhang, Y., Li, X., Chen, H., Yuille, A.L., Liu, Y., Zhou, Z., 2023. Continual learning for abdominal multi-organ and tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 35–45.

Zhang, C.-B., Xiao, J.-W., Liu, X., Chen, Y.-C., Cheng, M.-M., 2022. Representation compensation networks for continual semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7053–7064.

Zhao, Y., Bai, L., Zhang, Z., Wu, Y., Islam, M., Ren, H., 2024. Transferring knowledge from high-quality to low-quality MRI for adult glioma diagnosis. arXiv preprint arXiv:2410.18698.

Zhao, B., Xiao, X., Gan, G., Zhang, B., Xia, S.-T., 2020. Maintaining discrimination and fairness in class incremental learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13208–13217.

Zhou, Y., Towning, R., Awad, Z., Giannarou, S., 2024. Image synthesis with Class-Aware semantic diffusion models for surgical scene segmentation. arXiv preprint arXiv:2410.23962.

Zhu, V., Ji, Z., Guo, D., Wang, P., Xia, Y., Lu, L., Ye, X., Zhu, W., Jin, D., 2024. Low-rank continual pyramid vision transformer: Incrementally segment Whole-Body organs in CT with light-weighted adaptation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 371–381.

Zia, A., Bhattacharyya, K., Liu, X., Berniker, M., Wang, Z., Nespolo, R., Kondo, S., Kasai, S., Hirasawa, K., Liu, B., et al., 2023. Surgical tool classification and localization: results and methods from the miccai 2022 surgtoolloc challenge. arXiv preprint arXiv:2305.07152.