DATA MINING LAB

NEURAL INFORMATION PROCESSING SYSTEMS

Sample-Efficient Multi-Round Generative Data Augmentation for Long-Tail Instance Segmentation

Byunghyun Kim, Minyoung Bae, Jae-Gil Lee

KAIST Korea Advanced Institute of Science and Technology

Introduction

- The **discovery of diffusion models** enabled **low-cost, high-quality** image generation.
  - Open-sourced models such as Stable Diffusion and DALL·E are easily accessible.

- Q. Would it be possible to use these models to **generate data for AI training**?

Stable Diffusion 3

Trees photographed under the Milky Way, the moon and twilight shine on the Valley. The full moon appears high in the sky and the twilight glow can still be seen.

DALL-E 3

High-quality images generated by diffusion-based models.

Introduction

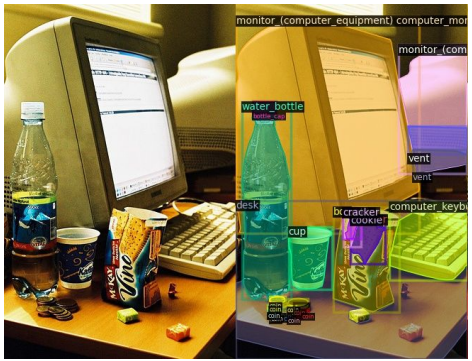
- **Object detection and instance segmentation tasks**
  - Definition: Predict the **location and class** of objects in a given image (bounding box / pixel-wise)
  - Characteristics:

(1) ***Annotation cost is high***

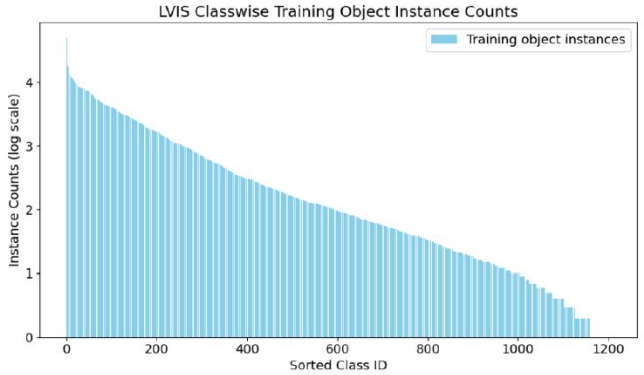
High cost is required for drawing boxes or boundary / pixel-wise annotations

(2) **Class imbalance exists**

In natural images, certain objects appear more frequently, showing a long-tail distribution



Accurate annotation is required for various objects, resulting in high cost



Class-wise object distribution of LVIS [1] dataset

Previous works

(a) **Layout-based** generative augmentation:

Train the diffusion model to generate images containing objects given a conditioned layout.

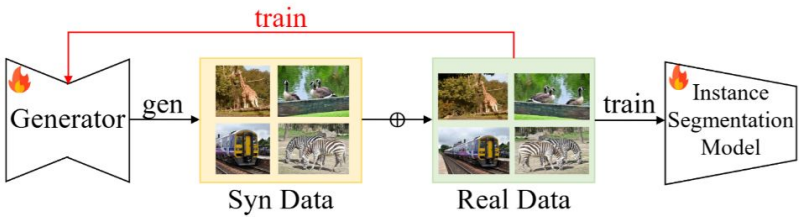
☹ High training cost and complexity for the generator.

(b) **Training-free** generative augmentation:

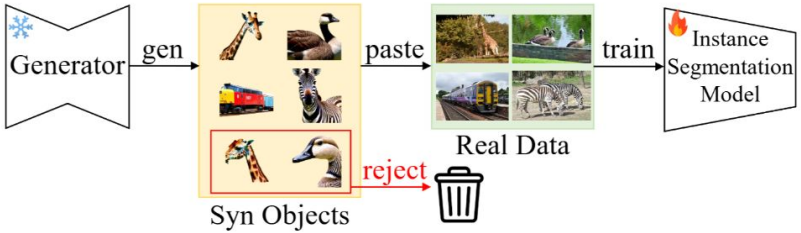
Pre-generate a large pool of image objects and augment the training data by pasting them.

☺ No training cost for the generator.

☹ Redundancy of preemptively constructed objects.



(a) Layout-based generative augmentation (existing).



(b) Training-free generative augmentation (existing).

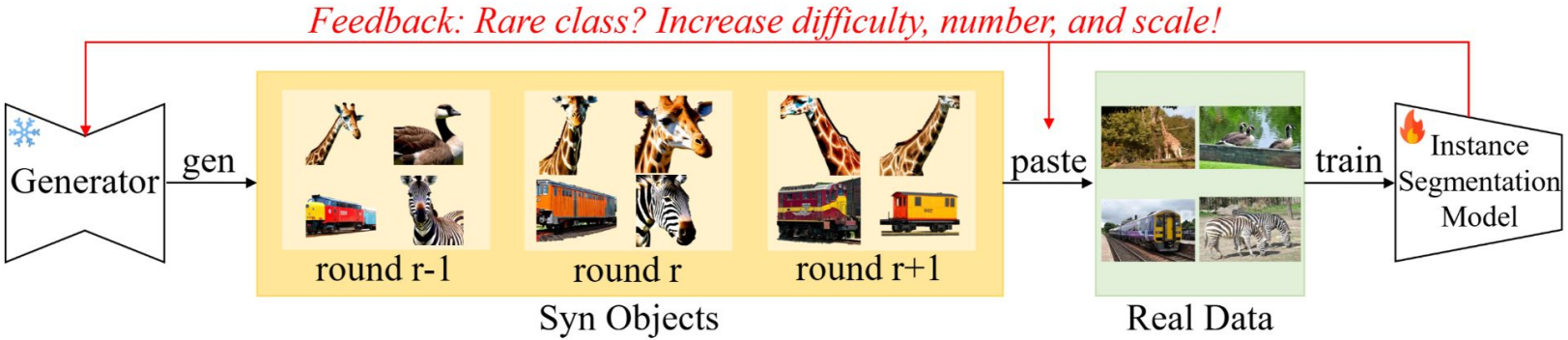
Our work

(c) **Multi-round collaborative augmentation (Ours)**:

Use feedback from the instance segmentation model (**ISM**) to the generator and pasting across multiple rounds.

- No training cost for the generator.
- Generate and paste objects with feedback from the ISM.

→ **Sample-efficient** and **higher-quality** generative data augmentation!



(c) MRCA: multi-round collaborative augmentation (ours).

Methodology

- **M**ulti-**R**ound **C**ollaborative **A**ugmentation (**MRCA**)
  - **Collaborative augmentation** between the ISM and the generator
    - ISM is trained by objects from the generator (**Gen**, **Map-Paste**, and **Training**).
    - The generator receives feedback from the ISM (**Feedback**).
  - **Multi-Round**: Provide **stable feedback** from the ISM to the generator and pasting

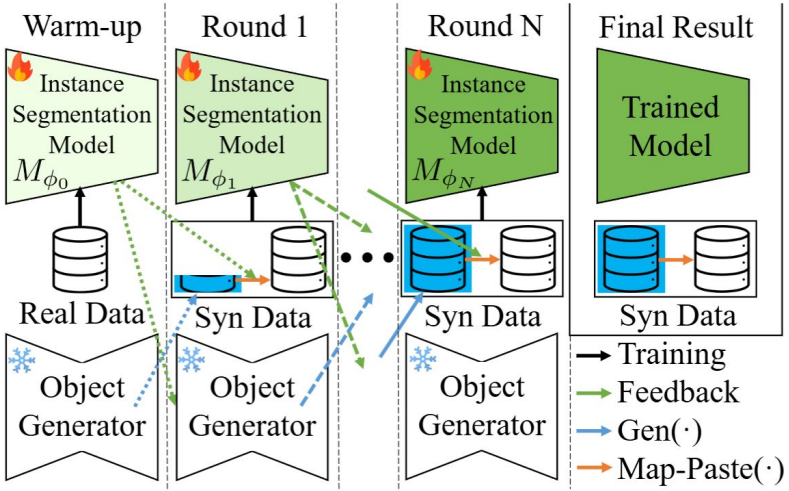
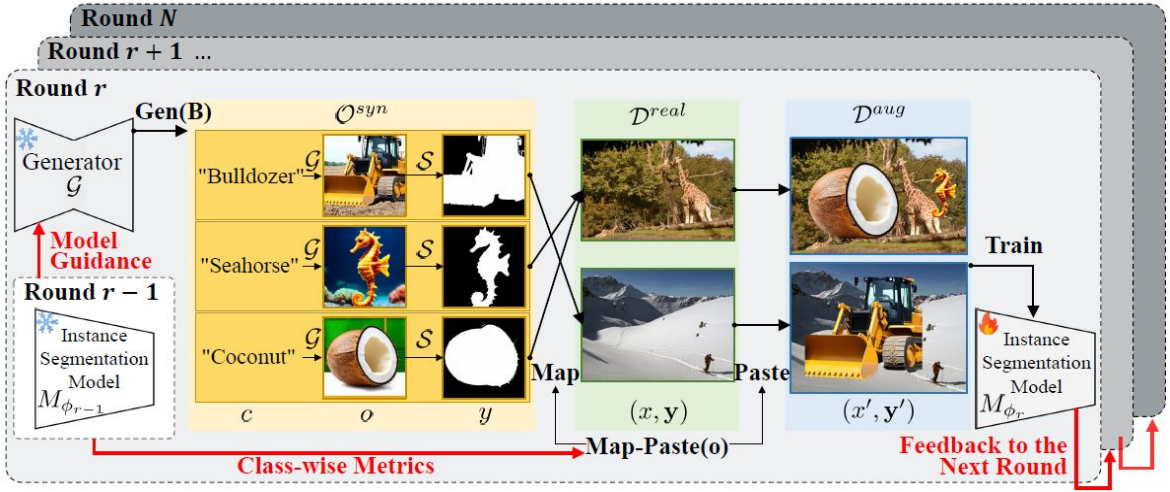


Illustration of multi-round collaborative augmentation

Methodology

- **Overall Pipeline**
  - Gen: objects obtained by generator & dichotomous segmentation
  - Map: map objects to real data images
  - Paste: paste the generated objects to real data images to create augmented data
  - **Train ISM**, and **give feedback to the next round!**



Detailed Illustration of multi-round collaborative augmentation

Methodology

- **Two-fold feedback** of MRCA
  - Feedback on **Object Generation**
    - Class-Wise Budget Optimization
    - Feedback-Guided Object Generation
  - Feedback on **Map-Paste**
    - Quota-Balanced Unique Mapping
    - Accuracy-Based Object Resizing

Methodology

- Feedback on Object Generation
  - **Class-Wise Budget Optimization**

    Each round r, generate objects by allocating more budget $B_r(c)$ to classes c with:
    - low accuracy ($A_{r,c}$),
    - low number of instances in the training data ($S_{r,c}$),
    - high diversity (measured by difference in classifier weights $C_{r,c}$).

$$r \leq 2 : B_r(c) = \frac{B}{c_{max}}, \quad r > 2 : B_r(c) = \alpha_r(1 - A_{r-2,c}) \cdot \|C_{r-2,c} - C_{r-3,c}\| \cdot \frac{1}{S_{r-2,c}}$$

Equation for deciding the number of objects ($B_r(c)$) to create for a class $c \in \{1, \cdots, c_{max}\}$ at the round r

  - Feedback-Guided Object Generation

Methodology

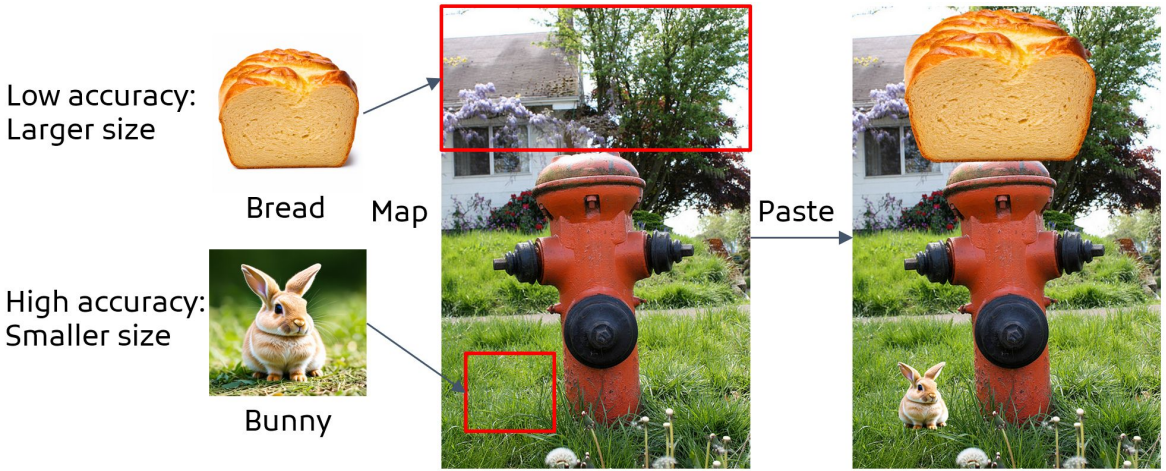
- Feedback on Object Generation
  - Class-Wise Budget Optimization
  - **Feedback-Guided Object Generation**
    - Provide feedback on diffusion sampling with the entropy criterion function [2]

$$\nabla_x \log p_{\theta,\gamma,\omega}(x|y) = \nabla_x \log p_\theta(x) + \gamma \nabla_x \log p(y|x) + \omega \nabla_x \mathcal{C}(x, y, M_\phi)$$

Equation for diffusion sampling process conditioned by the criterion function respect to the ISM

Methodology

- Feedback on **Map-Paste**
  - **Quota-Balanced Unique Mapping**
    - Randomly mapping objects may hinder our class-wise budget optimization.
    - Therefore, remove the drawn objects from the object pool and refill when empty.
  - ***Accuracy-Based Object Resizing***
    - Provide richer visual cues for underrepresented classes



Experimental results

- Quantitative analysis
  - Results in instance segmentation and object detection tasks for LVIS [1] data.
  - Despite using **only 6% of the generated objects** compared to previous works X-Paste [3], BSGAL [4], and DiverGen [5], MRCA outperforms in both tasks.

Table 1: Comparison with the state-of-the-art methods using the **ResNet-50** backbone.

Method	# Gen Objects	AP^{box}	AP^{mask}	AP^{box}_r	AP^{mask}_r	AP^{box}_c	AP^{mask}_c	AP^{box}_f	AP^{mask}_f
No Aug.	0	31.50	28.20	22.60	20.20	29.30	26.70	37.80	33.40
X-Paste	1200k	34.20	30.39	24.33	22.21	33.23	29.57	39.63	34.89
X-Paste + CLIP	1200k	34.35	30.70	25.99	24.38	32.83	29.41	39.71	34.92
BSGAL	1200k	35.40	31.56	27.95	25.43	34.14	30.56	40.07	35.37
MRCA	**72k**	**35.56**	**31.81**	**28.14**	**25.93**	**34.33**	**30.86**	**40.18**	**35.44**

Table 2: Comparison with the state-of-the-art methods using the **Swin-L** backbone.

Method	# Gen Objects	AP^{box}	AP^{mask}	AP^{box}_r	AP^{mask}_r	AP^{box}_c	AP^{mask}_c	AP^{box}_f	AP^{mask}_f
No Aug.	0	47.43	42.30	41.00	36.75	47.53	43.10	50.14	43.83
X-Paste	1200k	49.57	43.85	44.87	39.66	49.74	44.64	51.46	44.82
X-Paste + CLIP	1200k	49.80	44.51	45.28	40.62	49.33	44.96	**52.30**	**45.72**
BSGAL	1200k	50.47	44.85	47.55	42.37	50.43	45.47	51.79	45.26
DiverGen	1200k	51.24	45.48	50.07	45.85	51.33	45.83	51.64	44.96
MRCA	**72k**	**51.80**	**45.91**	**51.58**	**46.84**	**51.86**	**46.31**	51.84	45.05

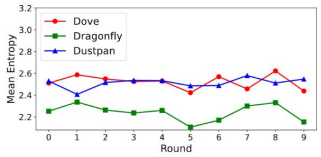
  - Extensive ablation studies can be found in our paper!

Experimental results

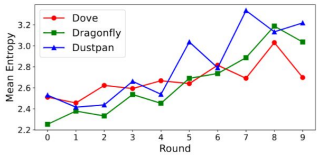
- **Round-*W*ise Object *V*isualization and Entropy**
  - The entropy respect to the feedback model remains similar across rounds
  - The entropy respect to the initial model increase over rounds

  → MRC*A* forms an **effective easy-to-hard curriculum**

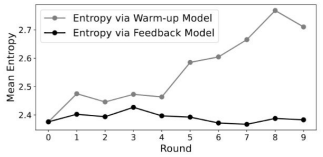
(a) Examples of generated objects for each round.



(b) Mean entropy of objects respect to the feedback model (M_{ϕ_r}).



(c) Mean entropy of objects respect to the warm-up model (M_{ϕ_0}).



(d) Class average entropy respect to the warm-up and feedback models.

Takeaways

- **Collaborative augmentation** between the target model for training and the data generator can provide room for **performance improvement**
  - Important to keep the **feedback stable** (in our case, through round-wise feedback)

- ***Augmenting training data online*** can increase **sample-efficiency**
  - Few powerful synthetic data can decrease the total cost of augmentation

References & Links

[1] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In CVPR, pages 5356–5364, 2019.
[2] Reyhane Askari-Hemmat et al. Feedback-guided data synthesis for imbalanced classification. arXiv preprint arXiv:2310.00158, 2023.
[3] Hanqing Zhao et al. X-paste: Revisiting scalable copy-paste for instance segmentation using clip and stablediffusion. In ICML, pages 42098–42109. PMLR, 2023.
[4] Muzhi Zhu et al. Generative active learning for long-tailed instance segmentation. In ICML. PMLR, 2024.
[5] Chengxiang Fan et al.. Divergen: Improving instance segmentation by learning wider data distribution with more diverse generative data. In CVPR, pages 3986–3995, 2024.

QR code to our paper