DiffAugmentation: Data Augmentation with DDPM for Object Recognition

02501 Advanced Deep Learning in Computer Vision

Dimitrios Papadopoulos
Associate Professor, DTU Compute

10 October, 2023

Keywords/lectures: Diffusion models, DDPM, text2image, Data augmentation, Limited labels

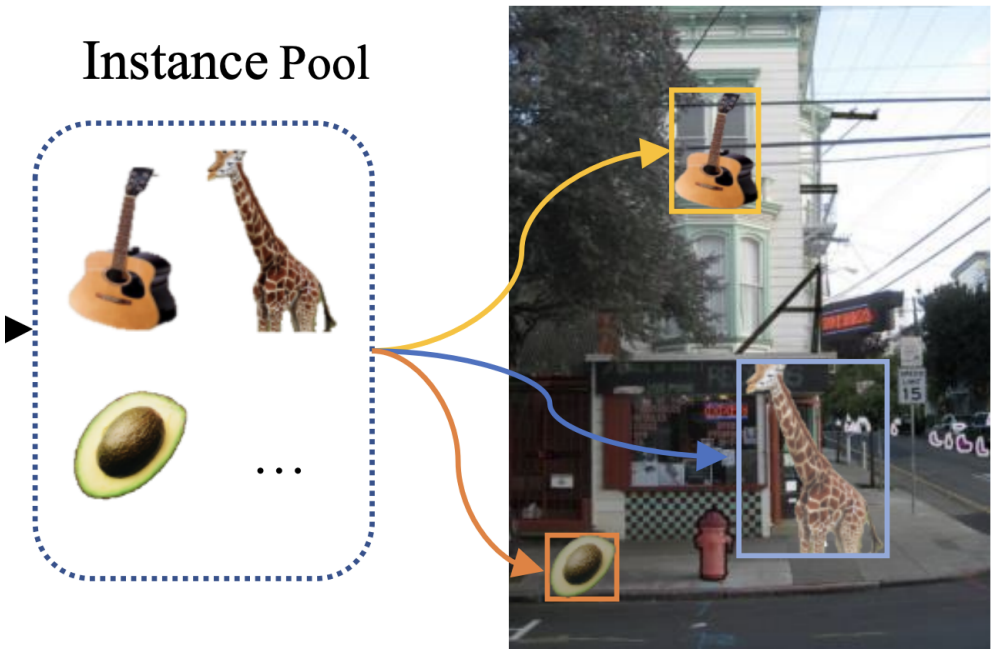

Figure 1: **Data Augmentation for Instance Segmentation.**

1 Project description

Instance segmentation [4] is a fundamental task in computer vision with very broad applications. In order to get plausible performance for one specific category, most existing methods rely on a large number of images annotated for this category, which is not only expensive but also time-consuming.

As one simple yet effective data augmentation strategy, Copy-Paste [3, 2, 1] has been extensively studied to improve data efficiency. By randomly pasting object instances onto background images, it can generate a combinatorial number of training data for free and boost the instance segmentation

model performance, especially for rare categories. Intuitively, if we can utilize more diverse object instances in Copy-Paste, more performance gain can be achieved.

The goal of this project is to combine the idea of Copy-Paste with the advances of generative models (i.e., diffusion models) [7, 5, 8] and examine the impact on the task of instance segmentation.

2 Data

In this project, you can use the COCO dataset [6], as standard for the instance segmentation task. Since COCO has 120K images, you can use either a subset of COCO (e.g., minicoco) or a smaller dataset (e.g., PASCAL VOC).

3 Tasks

In this project, you could work on the following tasks:
Task 1: Reproduce a simple Copy-Paste augmentation algorithm [3]. Evaluate the results of this augmentation algorithm on a standard instance segmentation problem. You can also compare the results with several other baselines, such as standard augmentation, RandAugment, or AutoAugment.
Task 2: Generate new isntances. Use Stable diffusion or ControlNet to generate new instances for the target object categories.
Task 3: Copy-Paste and Stable Diffusion. Combine the generated instances with the Copy-Paste strategy and augment your training set with more object instances.
Task 4: ...

References

[1] Nikita Dvornik, Julien Mairal, and Cordelia Schmid. Modeling visual context is key to augmenting object detection datasets. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 364–380, 2018.

[2] Debidatta Dwibedi, Ishan Misra, and Martial Hebert. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1301–1310, 2017.

[3] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. 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*, pages 2918–2928, 2021.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[8] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.