

CISA: Context Substitution for Image Semantics Augmentation

Subjects: **Computer Science, Artificial Intelligence**

Contributor: Sergey Nesteruk , Ilya Zherebtsov , Svetlana Illarionova , Dmitrii Shadrin , Andrey Somov , Sergey V. Bezzateev , Tatiana Yelina , Vladimir Denisenko , Ivan Oseledets

Context is essential for generic neural networks. In Computer Vision background defines context. Therefore, smart background substitution can improve the generalization capability of a trained model.

image augmentation

computer vision

data collection

image retrieval

1. Introduction

Deep learning and computer vision (CV) algorithms have recently shown their capabilities in addressing various challenging industrial and scientific problems [1]. Successful application of machine learning and computer vision algorithms for solving complex tasks is impossible without relying on comprehensive and high-quality training and testing data [2][3]. CV algorithms for solving classification, object detection, and semantic and instance segmentation require a huge variety of input data to ensure robust work of the trained models [4][5][6]. There are two major ways to enlarge a training dataset. The first one is obvious and implies physical collection of the dataset samples in various conditions to ensure high diversity of the training data. There is a set of huge datasets that have been collected for solving computer vision problems. These datasets are commonly used as the benchmark [7][8][9][10]. One of the specifics of these datasets is that they are general-domain sets. Unfortunately, general-domain-labeled data can be almost useless for solving specific industrial problems. One of the feasible applications of such well-known datasets is that they can serve as a good basis for pre-training of neural networks (transfer learning) [11][12]. Using these pre-trained neural networks, it is possible to fine-tune them and adapt them to address specific problems. However, in some cases, even for fine-tuning, a comprehensive dataset is in high demand. Some events are rare, and it is possible to collect only a few data samples [13][14][15]. Thus, a second approach for enhancing the characteristics of the dataset can help. This approach is based on artificial manipulations with the initial dataset [16][17]. One of the well-developed techniques is data augmentation, where original images are transformed according to special rules [18]. Usually, the goal of image augmentation is to make the training dataset more diverse. However, augmentation can be used to deliberately shift the data distribution. If the distribution of the original training dataset differs from the distribution of the test set, it is important to equalize them as much as possible.

The agricultural domain is part of the industrial and research areas for which the development of artificial methods for improvement of training datasets is vital [19][20][21]. This demand appears due to the high complexity and variability of the investigated system (plant) that has to be characterized by computer vision algorithms [22]. The difficulty of the agricultural domain makes it a good candidate for testing augmentation algorithms.

There are many different plant species, and plants grow slowly. Thus, collecting and labeling huge datasets for each specific plant growing in each specific stage is a complex task [23]. Overall, it is difficult to collect datasets [24], especially for plants, and it is expensive to annotate them [25].

2. Image Augmentation

Computer vision models require many training data. Therefore, it becomes challenging to obtain a good model with limited datasets. Namely, a small-capacity model might not capture complex patterns, while a big capacity model tends to overfit if small datasets are used [26]. Slight changes in test data connected with surrounding and environmental conditions might also lead to a decrease in model performance [27].

To overcome this issue, researchers use various image augmentation techniques. Data augmentation aims to add diversity to the training set and to complicate the task for a model [28]. Among these plant image augmentation approaches, it can be distinguished: basic computer vision augmentations, learned augmentation, graphical modeling, augmentation policy learning, collaging, and compositions of the ones above.

Basic computer vision augmentations are the default methods preventing overfitting in most computer vision tasks. They include image cropping, scaling, flipping, rotating, and adding noise [29]. There are also advanced augmentation methods, connected with distortion techniques and coordinate system changes [30]. Since these operations are quite generic, most popular ML frameworks support them. However, although helpful, these methods demonstrate limited use, as they bring insufficient diversity to the training data for few-shot learning cases.

Learned augmentation stands for generating training samples with an ML model. For this purpose, conditional generative adversarial networks (cGANs) and variational autoencoders (VAEs) are frequently used. In the agricultural domain, there are examples of applying GANs to *Arabidopsis* plant images for the leaf counting task [31] [32]. The main drawback of this approach is that generating an image with a neural network is quite resource-intensive. Another disadvantage is the overall pipeline complexity: the errors of a model that generates training samples are accumulated with the errors of a model that solves the target task.

Learned augmentation policy is a series of techniques used to find combinations of basic augmentations that maximize model generalization. This implies hard binding of the learned policy to the ML model, the dataset, and the task. Although it is shown to provide systematic generalization improvement on object detection [33] and classification [34], its universal character as well as the ability to be performed along with multi-task learning are not supported with solid evidence.

Collaging presupposes cropping an object from an input image with the help of a manually annotated mask and pasting it to a new background with basic augmentations of each object [19]. In [35], a scene generation technique using object mask was successfully implemented for an instance detection task. It boosted model performance significantly compared with the use of only original images. The study on image augmentation for instance

segmentation using a copy–paste technique with object mask was extended in [36]. The importance of scene context for image augmentation is explored in [37][38].

3. Image Synthesis

Graphical modeling is another popular method in plant phenomics. It involves creating a 3D model of the object of interest and rendering it. The advantage of this process is that it permits the generation of large datasets [39] with precise annotations, as the labels of each pixel are known. However, this technique is highly resource-intensive; moreover, the results obtained using the existing solutions [40][41] seem artificial. More realistic synthesis is very time-consuming. This approach is suitable when there are not many variations of the modeled object. If there are many different object types, it can be easier to collect and annotate new images.

4. Neural Image Generation and Image Retrieval

To gain new training images for CV tasks, one can implement GAN-based or diffusion-based models. Currently, they allow for the creation of rather realistic images and meet the demands of different domains, such as agricultural [42], manufacturing processes [43], remote sensing [44], or medical [45]. Such models can be considered as a part of an image recognition pipeline. Moreover, recent results in Natural Language Processing (NLP) offer opportunities to extend image generation applications via textual description. For instance, an image can be generated based on a proposed prompt, namely, a phrase or a word. Such synthetic images help to extend the initial dataset. The same target image can be described by a broad variety of words and phrases that lead to diverse visual results. Another way to obtain additional training images is a data retrieval approach. It supposes to search for existing images from the Internet or some database according to a user's prompt. For instance, the CLIP model can be used to compute embedding of a text and to find images that match it better based on distance in a special embedding space [46].

References

1. Kwon, O.; Sim, J.M. Effects of data set features on the performances of classification algorithms. *Expert Syst. Appl.* 2013, 40, 1847–1857.
2. Sbai, O.; Couprie, C.; Aubry, M. Impact of base dataset design on few-shot image classification. In Proceedings of the European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2020; pp. 597–613.
3. Zendel, O.; Murschitz, M.; Humenberger, M.; Herzner, W. How good is my test data? Introducing safety analysis for computer vision. *Int. J. Comput. Vis.* 2017, 125, 95–109.
4. Barbedo, J.G.A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 2018, 153, 46–53.

5. Zheng, S.; Song, Y.; Leung, T.; Goodfellow, I. Improving the robustness of deep neural networks via stability training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4480–4488.
6. Hendrycks, D.; Basart, S.; Mu, N.; Kadavath, S.; Wang, F.; Dorundo, E.; Desai, R.; Zhu, T.; Parajuli, S.; Guo, M.; et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. *arXiv* 2020, arXiv:2006.16241.
7. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 20–25 June 2009; pp. 248–255.
8. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
9. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3974–3983.
10. Caba Heilbron, F.; Escorcia, V.; Ghanem, B.; Carlos Niebles, J. Activitynet: A large-scale video benchmark for human activity understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 961–970.
11. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In Proceedings of the International Conference on Artificial Neural Networks; Springer: Berlin/Heidelberg, Germany, 2018; pp. 270–279.
12. Lemikhova, L.; Nesteruk, S.; Somov, A. Transfer Learning for Few-Shot Plants Recognition: Antarctic Station Greenhouse Use-Case. In Proceedings of the 2022 IEEE 31st International Symposium on Industrial Electronics (ISIE), Anchorage, AL, USA, 1–3 June 2022; pp. 715–720.
13. Vannucci, M.; Colla, V. Classification of unbalanced datasets and detection of rare events in industry: Issues and solutions. In Proceedings of the International Conference on Engineering Applications of Neural Networks; Springer: Berlin/Heidelberg, Germany, 2016; pp. 337–351.
14. Nesteruk, S.; Shadrin, D.; Pukalchik, M.; Somov, A.; Zeidler, C.; Zabel, P.; Schubert, D. Image compression and plants classification using machine learning in controlled-environment agriculture: Antarctic station use case. *IEEE Sensors J.* 2021, 21, 17564–17572.
15. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv. (CSUR)* 2020, 53, 1–34.
16. Illarionova, S.; Nesteruk, S.; Shadrin, D.; Ignatiev, V.; Pukalchik, M.; Oseledets, I. Object-based augmentation for building semantic segmentation: Ventura and santa rosa case study. In

Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1659–1668.

17. Illarionova, S.; Shadrin, D.; Ignatiev, V.; Shayakhmetov, S.; Trekin, A.; Oseledets, I. Augmentation-Based Methodology for Enhancement of Trees Map Detalization on a Large Scale. *Remote. Sens.* 2022, 14, 2281.
18. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* 2019, 6, 60.
19. Kuznichov, D.; Zvirin, A.; Honen, Y.; Kimmel, R. Data Augmentation for Leaf Segmentation and Counting Tasks in Rosette Plants. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.
20. Fawakherji, M.; Potena, C.; Prevedello, I.; Pretto, A.; Bloisi, D.D.; Nardi, D. Data Augmentation Using GANs for Crop/Weed Segmentation in Precision Farming. In Proceedings of the 2020 IEEE Conference on Control Technology and Applications (CCTA), Montreal, QC, Canada, 24–26 August 2020; pp. 279–284.
21. Wu, Q.; Chen, Y.; Meng, J. DCGAN Based Data Augmentation for Tomato Leaf Disease Identification. *IEEE Access* 2020.
22. Nesteruk, S.; Shadrin, D.; Kovalenko, V.; Rodriguez-Sanchez, A.; Somov, A. Plant Growth Prediction through Intelligent Embedded Sensing. In Proceedings of the IEEE 29th International Symposium on Industrial Electronics (ISIE), Delft, The Netherlands, 17–19 June 2020; Volume 2020, pp. 411–416.
23. Nesteruk, S.; Illarionova, S.; Akhtyamov, T.; Shadrin, D.; Somov, A.; Pukalchik, M.; Oseledets, I. XtremeAugment: Getting More From Your Data Through Combination of Image Collection and Image Augmentation. *IEEE Access* 2022, 10, 24010–24028.
24. Nesteruk, S.; Bezzateev, S. Location-Based Protocol for the Pairwise Authentication in the Networks without Infrastructure. In Proceedings of the 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, Finland, 15–18 May 2018; pp. 190–197.
25. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferreo, E.; Agapow, P.-M.; Zirtz, M.; Hoffman, M.M.; et al. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 2018, 15, 20170387.
26. Feng, R.; Gu, J.; Qiao, Y.; Dong, C. Suppressing Model Overfitting for Image Super-Resolution Networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.
27. Illarionova, S.; Nesteruk, S.; Shadrin, D.; Ignatiev, V.; Pukalchik, M.; Oseledets, I. MixChannel: Advanced augmentation for multispectral satellite images. *Remote. Sens.* 2021, 13, 2181.

28. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the Computer Vision—ECCV 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833.

29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 2017, **60**, 84–90.

30. Buslaev, A.; Parinov, A.; Khvedchenya, E.; Iglovikov, V.I.; Kalinin, A.A. Albumentations: Fast and flexible image augmentations. *Information* 2020, **11**, 125.

31. Zhu, Y.; Aoun, M.; Krijn, M.; Vanschoren, J.; Campus, H.T. Data Augmentation using Conditional Generative Adversarial Networks for Leaf Counting in Arabidopsis Plants. In Proceedings of the BMVC, Newcastle, UK, 3–6 September 2018; p. 324.

32. Valerio Giuffrida, M.; Scharr, H.; Tsaftaris, S.A. ARIGAN: Synthetic Arabidopsis Plants Using Generative Adversarial Network. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, Venice, Italy, 22–29 October 2017.

33. Zoph, B.; Cubuk, E.D.; Ghiasi, G.; Lin, T.; Shlens, J.; Le, Q.V. Learning Data Augmentation Strategies for Object Detection. *arXiv* 2019, arXiv:1906.11172.

34. Lemley, J.; Bazrafkan, S.; Corcoran, P. Smart Augmentation Learning an Optimal Data Augmentation Strategy. *IEEE Access* 2017, **5**, 5858–5869.

35. Dwibedi, D.; Misra, I.; Hebert, M. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1301–1310.

36. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. 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, Online, 19–25 June 2021; pp. 2918–2928.

37. Dvornik, N.; Mairal, J.; Schmid, C. On the importance of visual context for data augmentation in scene understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 2019, **43**, 2014–2028.

38. Su, Y.; Sun, R.; Lin, G.; Wu, Q. Context decoupling augmentation for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 7004–7014.

39. Flores-Fuentes, W.; Trujillo-Hernández, G.; Alba-Corpus, I.Y.; Rodríguez-Quiñonez, J.C.; Mirada-Vega, J.E.; Hernández-Balbuena, D.; Murrieta-Rico, F.N.; Sergiyenko, O. 3D spatial measurement for model reconstruction: A review. *Measurement* 2023, **207**, 112321.

40. Barth, R.; IJsselmuider, J.; Hemming, J.; Henten, E.V. Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. *Comput. Electron. Agric.* 2018, **144**,

284–296.

41. Ward, D.; Moghadam, P.; Hudson, N. Deep Leaf Segmentation Using Synthetic Data. *arXiv* 2018, [arXiv:1807.10931](https://arxiv.org/abs/1807.10931).
42. Lu, Y.; Chen, D.; Olaniyi, E.; Huang, Y. Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. *Comput. Electron. Agric.* 2022, **200**, 107208.
43. Liu, K.; Li, Y.; Yang, J.; Liu, Y.; Yao, Y. Generative principal component thermography for enhanced defect detection and analysis. *IEEE Trans. Instrum. Meas.* 2020, **69**, 8261–8269.
44. Illarionova, S.; Shadrin, D.; Trekin, A.; Ignatiev, V.; Oseledets, I. Generation of the nir spectral band for satellite images with convolutional neural networks. *Sensors* 2021, **21**, 5646.
45. Chen, Y.; Yang, X.H.; Wei, Z.; Heidari, A.A.; Zheng, N.; Li, Z.; Chen, H.; Hu, H.; Zhou, Q.; Guan, Q. Generative adversarial networks in medical image augmentation: A review. *Comput. Biol. Med.* 2022, **105382**.
46. Beaumont, R. Clip Retrieval: Easily Compute Clip Embeddings and Build a Clip Retrieval System with Them. 2020. Available online: <https://github.com/rom1504/clip-retrieval> (accessed on 27 February 2023).

Retrieved from <https://encyclopedia.pub/entry/history/show/114054>