Object-Based Augmentation for Building Semantic Segmentation: Ventura and Santa Rosa Case Study

Svetlana Illarionova Sergey Nesteruk Dmitrii Shadrin Vladimir Ignatiev
Mariia Pukalchik Ivan Oseledets
Skolkovo Institute of Science and Technology, Moscow, Russia
{S.Illarionova,Sergei.Nesteruk,dmitry.shadrin,V.Ignatiev,M.Pukalchik,i.oseledets}@skoltech.ru

Abstract

Today deep convolutional neural networks (CNNs) push the limits for most computer vision problems, define trends, and set state-of-the-art results. In remote sensing tasks such as object detection and semantic segmentation, CNNs reach the SotA performance. However, for precise performance, CNNs require much high-quality training data. Rare objects and the variability of environmental conditions strongly affect prediction stability and accuracy. To overcome these data restrictions, it is common to consider various approaches including data augmentation techniques. This study focuses on the development and testing of object-based augmentation. The practical usefulness of the developed augmentation technique is shown in the remote sensing domain, being one of the most demanded in effective augmentation techniques. We propose a novel pipeline for georeferenced image augmentation that enables a significant increase in the number of training samples. The presented pipeline is called object-based augmentation (OBA) and exploits objects' segmentation masks to produce new realistic training scenes using target objects and various label-free backgrounds. We test the approach on the buildings segmentation dataset with different CNN architectures (U-Net, FPN, HRNet) and show that the proposed method benefits for all the tested models. We also show that further augmentation strategy optimization can improve the results. The proposed method leads to the meaningful improvement of U-Net model predictions from 0.78 *to* 0.83 *F1-score.*

1. Introduction

Machine learning models depend drastically on the data quality and its amount. In many cases, using more data allows the model to reveal hidden patterns deeper and achieve better prediction accuracy [42]. However, gathering of a high-quality labeled dataset is a time-consuming and expensive process [35]. Moreover, it is not always possible to obtain additional data: in many tasks, unique or rare objects are considered [31] or access to the objects is restricted [17]. In other tasks, we should gather data rapidly [40]. The following tasks are among such challenges: operational damage assessment in emergency situations [33], medical image classification [28]. There are different approaches to address dataset limitations: pseudo labeling, special architectures development, transfer learning [2, 3, 32, 50]. Another standard method to address this issue is image augmentation. Augmentation means applying transformations (such as flip, rotate, scale, change brightness and contrast) to the original images to increase useful samples that allow training more robust algorithms [4].

In this study, we focus on augmentation techniques for the remote sensing domain. The lack of labeled data for particular remote sensing tasks makes it crucial to generate more training samples artificially and prevent overfitting [53]. Data augmentation is especially important to enhance the efficiency of deep learning applications in remote sensing [27]. This work aims to propose an object-based augmentation (OBA) pipeline for the semantic segmentation task that works with high-resolution georeferenced satellite images. Naming our augmentation methodology object-based, we imply that this technique targets separate objects instead of whole images. The idea behind the approach is to crop objects from original images using their masks and pasting them to a new background. This method is studied in the general domain [11, 52, 55], but we are the first to study its effectiveness in remote sensing applications. For this purpose, we adopt the method to work with geospatial data formats and experiment with case-specific features (such as objects' shadows and large study area size). In our approach, every object and background can be augmented independently to increase the variability of training images; shadows for pasted objects also can be added artificially. We show that our approach is superior to the classic image-based methods in the remote sensing domain despite its simplicity. The pipeline is tested in a building segmentation task using U-Net [39], Feature Pyramid Network (FPN) [25], and HRNet model [43] to reveal a relationship between convolutional neural network (CNN) architecture and augmentation benefit.

The main contributions of this paper are:

- We propose a novel for remote sensing domain simple and efficient augmentation scheme called OBA that improves CNN model generalization for satellite images.
- We test the proposed method on the building segmentation task and show that our approach outperforms common augmentation approaches.
- We show that OBA parameters can be efficiently optimized for better performance.

The OBA code is available:
https://github.com/LanaLana/satellite_object_augmentation.

2. Related Works

We can split all image augmentations into two groups according to the target. Image-based augmentations transform the entire image. On the contrast, object-based augmentation technique targets every object in the image independently [10, 30, 51]. It makes augmentations more flexible and provides a better way to handle sparse objects which is particularly useful for remote sensing problems. However, this novel approach has not been studied yet in the remote sensing domain.

In [55], for the object detection task, authors perform image transformations individually within and outside bounding boxes. They also change bounding box position regarding the background. In [52] authors clip area with the target object and replace it with the same class object from another image. However, the bounding box's background is still from the source image of the new object. It makes generated image less realistic and can affect further classification. In [10], for semantic segmentation, authors use objects' masks to create new images with pasted objects.

Another notable augmentation approach is based on generative adversarial neural networks (GANs) [34]. It generates completely new training examples that can benefit the final model performance [9]. However, this approach requires training an auxiliary model that produces training samples for the main model. In this work, we focus only on augmentation approaches that require neither major changes in the training loop nor much computational overhead.

Augmentation is extensively used in various areas. In [36], they proposed augmentation for medical images aimed to classify skin lesions. In [46], augmentation was implemented for underwater photo classification. Another sphere of study that processes images distinguished from regular camera photos is remote sensing [16].

The most frequently used augmentation approach in remote sensing is also color and geometrical transformations [18, 21, 49, 54]. In [23], rescaling, slicing, and rotation augmentations were applied in order to increase the quantity and diversity of training samples in building semantic segmentation task. In [38], authors implemented ”random rotation” augmentation method for small objects detection. The effect of three geometrical transformations (flip, translation, and rotation) on DL model performance was assessed in [49]. In [41], authors discussed advances of augmentation leveraging in the landcover classification problem with limited training samples. Another task and augmentation approach is described in [47]. The authors used 3D ship models to insert them into the background obtained from high-resolution satellite images. Another augmentation approach with 3D models leveraging for aircraft detection was described in [48]. The main limitation of listed works is related to 3D models' unavailability for most remote sensing problems. The above overview clearly states the importance of the augmentation techniques in current computer vision remote sensing research and high capabilities for making the trained models more generalized and precise. Thus, improving the augmentation techniques is crucial for developing accurate solutions for practical computer vision tasks. One can see that the application of generic geometrical and color transformations provides a very limited increase in models' performance due to the small variability of those transformations. A more promising approach is to treat every object of interest separately. It allows varying both the objects' augmentations and its surrounding. The currently used algorithms, such as ones based on 3D modeling, verify the superiority of object-based augmentations. However, these techniques are poorly scalable to the new types of objects due to extremely time-consuming manual 3D modeling. To overcome this issue, in this paper, we propose an automated object-based augmentation method.

3. Methodology

3.1. Object-based augmentation

This section describes the object-based augmentation methodology for the semantic segmentation problem.

Object-based augmentation requires images containing objects with masks and background images. Each object has its ID and shape coordinates extracted from a geojson file. Based on the information about object location, one can crop it from the original image, and paste to a new background. There were two types of background areas: from the initial dataset and new unlabeled images that aim to add diversity into data. An object and a background were chosen randomly. According to the object's coordinates, a crop with a predefined size containing the object was clipped

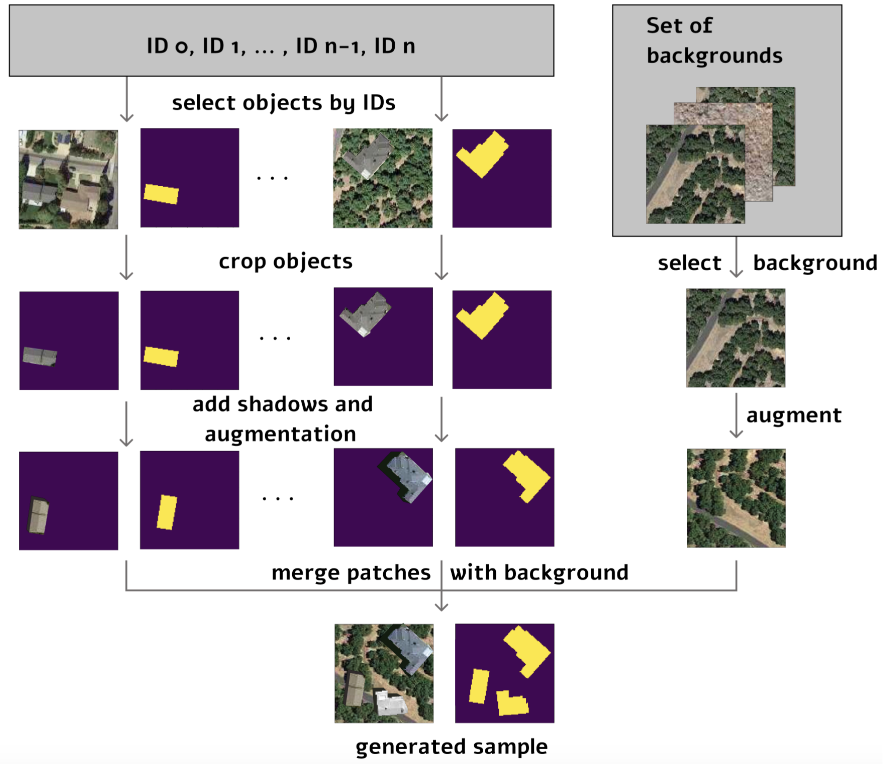

Figure 1. Object-based augmentation (OBA) scheme. For each generated sample, we choose objects from the set of IDs, crop objects according to its footprint, add shadows, conduct geometrical and color transformations, and then merge these cut objects with a new background.

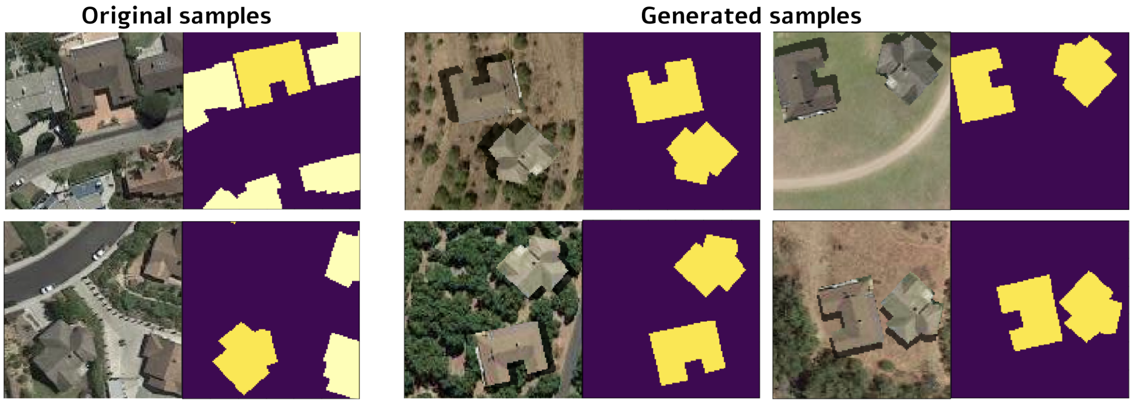

Figure 2. Examples of augmented samples reconstructing various environmental conditions. Objects and backgrounds are from different images and have different color and geometrical transformations. Shadows are added artificially for the generated samples.

for RGB channels and masks (Figure 1). The background crop has the same size. With some set probability, the object's crop and the background's crop were augmented separately or together using base color and geometrical transformations from Albumentations package [4]. This package is popular both for semantic segmentation task in general and remote sensing domains. The considered in our study transformations are described in Table 1. Since most of the works in remote sensing do not specify albumentations parameters for images augmentation [21, 54] we also set default parameters.

The object extension was then merged with a new background by placing it in a random position strictly within the image crop. Objects number within each image crop was chosen randomly in a predefined range. Overlapping between objects was prohibited.

To make generated samples more realistic, we add shadows using objects' footprints (Figure 2). The mask of the

Transformation	Description
RandomRotate90	Randomly rotate the input by 90 degrees zero or more times
Flip	Flip the input either horizontally, vertically or both
Additive GaussianNoise	Add Gaussian noise to the input image
HueSaturationValue	Randomly change hue, saturation and value of the input image
CLAHE	Apply Contrast Limited Adaptive Histogram Equalization to the input image
OpticalDistortion	Apply Barrel Distortion [12] to the image
RandomContrast	Randomly change contrast of the input image
RandomBrightness	Randomly change brightness of the input image
IAAEmboss	Emboss the input image and overlays the result with the original image
MotionBlur	Apply motion blur to the input image using a random-sized kernel

Table 1. Base color and geometrical transformations from ablumentations.

shadowed area is blended with initial background pixels with different intensities.

The difference between the general and remote sensing domains often relates to image size in a dataset. The average image resolution in ImageNet dataset is $469 * 387$ pixels, while in many remote sensing datasets image is significantly larger. Images in DOTA dataset have size about $4000 * 4000$ pixels and may contain large-size images with only a handful of small instances [44]. Image size for the remote sensing domain often depends on the study area scale. A single satellite image can cover an entire city or a large county. Moreover, target objects in remote sensing tasks usually have dramatically lower density (as in the beforementioned DOTA dataset) in comparison to general domain images. It is necessary to split an initial image into crops that a CNN model can accept for training. Therefore, sampling strategy is crucial for the remote sensing domain as simple image partition into tiles is unproductive for large study areas [45]. Our framework supports an efficient sampling strategy that uses objects coordinates to crop training patches within large georeferenced images (Figure 1).

The entire new sample generating process is conducted during model training. It aimed to ensure greater diversity without memory restrictions related to additional sample storage. Therefore, all functions for object-based augmentation were implemented into the data-loader and generator. New generated samples are also alternated with original samples.

In summary, OBA includes the following options:

- Shadows addition (length and intensity may vary);
- Objects number per crop selection (default: up to 3 extra objects);
- Selection of base color and geometrical transformations probability (default: 50%);
- Background images selection (default: 60%);
- Selection of original and generated samples mixing probability (default: 60%).

We compared this augmentation approach with the following alternatives: *Baseline* is training a CNN model using just base color and geometrical augmentations (from Albumentations framework); *Baseline_no_augm* is training a CNN model without any augmentations; *OBA_no_augm* is applying OBA cropping and pasting without generic augmentations; *OBA_no_shadow* is applying OBA cropping and pasting without adding generated shadows to the objects; *OBA_no_background* is applying OBA cropping and pasting using a background from the same image only. Note that all the tested OBA approaches except *OBA_no_background* use crops from other images to form a background. The summary of the experiments is reflected in Table 5.

3.2. Optimization

The task of optimal augmentation policy choice is a significant part of algorithm adjustment. Many works are devoted to this topic [6, 8, 24]. It is often defined as a combinatorial optimization problem of optimal transformations search within some available set.

To choose the best augmentation strategy, we set experiments with the optimizer from the Optuna software framework [1]. It uses a multivariant Tree-structured Parzen Estimator. Optuna helps to search hyperparameters efficiently and shows significant improvements for various machine learning and deep learning tasks [14, 19]. The optimizer supports earlier pruning to reject weak parameters initialization. As a pruner, we used Optuna's implementation of MedianPruner. Loss function value after each epoch was evaluated to choose new parameters' values in the searching space. For object-based augmentation, the following parameters were considered:

- Number of generated objects within one crop;
- Probability of the base color transformations;
- Probability of object-based augmentation;

	Baseline_no_augm			*Baseline*			*OBA*		
Training set size	1/3	2/3	1	1/3	2/3	1	1/3	2/3	1
F1-score	0.415	0.43	0.45	0.751	0.785	0.788	0.787	0.81	0.829

Table 2. Augmentation approaches comparison for different training set size using U-Net with Resnet34 encoder (F1-score for the test set).

	Baseline_no_augm			*Baseline*			*OBA*		
Model	Resnet18	Resnet34	Resnet50	Resnet18	Resnet34	Resnet50	Resnet18	Resnet34	Resnet50
FPN	0.325	0.367	0.186	0.741	0.762	0.784	0.802	0.813	**0.826**
U-Net	0.435	0.45	0.34	0.766	0.788	0.766	0.807	**0.829**	0.824
	Resnet101			Resnet101			Resnet101		
HRNet	0.23			0.741			**0.812**		

Table 3. Augmentation approaches comparison for different CNN models (FPN, U-Net, HRNet) with different encoders (Resnet18, Resnet34, Resnet50, Resnet101). F1-score for the test set.

- Probability of extra background usage.

For this study, we set 12 epochs and the same validation samples representation without any modifications to obtain the most equivalent criteria as possible for earlier pruning.

4. Experiments

4.1. Dataset

	Train	Validation	Test
Objects number	955	226	282
Area in hectars	390	100	93
Extra background area in hectars	2000	500	500

Table 4. Dataset description.

	Base augm.	Shadow	Extra background
Baseline_no_augm	✗	✗	✗
Baseline	✓	✗	✗
OBA_no_augm	✗	✓	✓
OBA_no_shadow	✓	✗	✓
OBA_no_background	✓	✓	✗
OBA	✓	✓	✓

Table 5. Experiments with different augmentation setups.

We evaluated the developed augmentation pipeline in the remote sensing semantic segmentation problem, namely the buildings segmentation task. It is an important problem for remote sensing, and it was considered in different studies [23, 37]. Lack of labelled training data makes it suitable for the OBA approach evaluation.

For building segmentation, we used the dataset described in [33]. This dataset was collected for damage assessment in the emergency and included images before and after wildfires in California in 2017. However, we leveraged just data before the event. It covers Ventura and Santa Rosa counties (the total area is about 580 hectares). Very high-resolution RGB images for this region were available through Digitalglobe within their Open Data Program. We used 955 buildings for training and 226 for validation from Ventura and 282 buildings from Santa-Rosa for the test (see Table 4). Objects' masks are presented both in raster TIFF format and vector shapes. Image that was used for training is shown in Figure 3.

We selected high-resolution extra background without target objects from Maxar serves [29] (image id is *lnu-lightning-complex-fire*, April 15, 2020, California). We cut test, validation and train images with the total area of about 3000 ha. It includes various land-cover types such as: lawns, individual trees, roads, and forested areas.

4.2. Effect of the train dataset size

To assess the effect of the dataset size on the final model score we considered the following subset of samples for training dataset:

- The entire training dataset;
- 2/3 of the entire training dataset;
- 1/3 of the entire training dataset.

For each experiment, we fixed the same validation set that was not reduced further. For the reduced training dataset, we ran a model on different subsets. There were 2 and 3 subsets for each of the mentioned dataset sizes. The final results for each training subset size were defined as an average. We conducted these experiments for three different training modes: without augmentation (*Baseline_no_augm*), with base color and geometrical transformations (*Baseline*), with object-based augmentation (*OBA*).

To evaluate how original and generated samples affect the final score the following experiment was conducted:

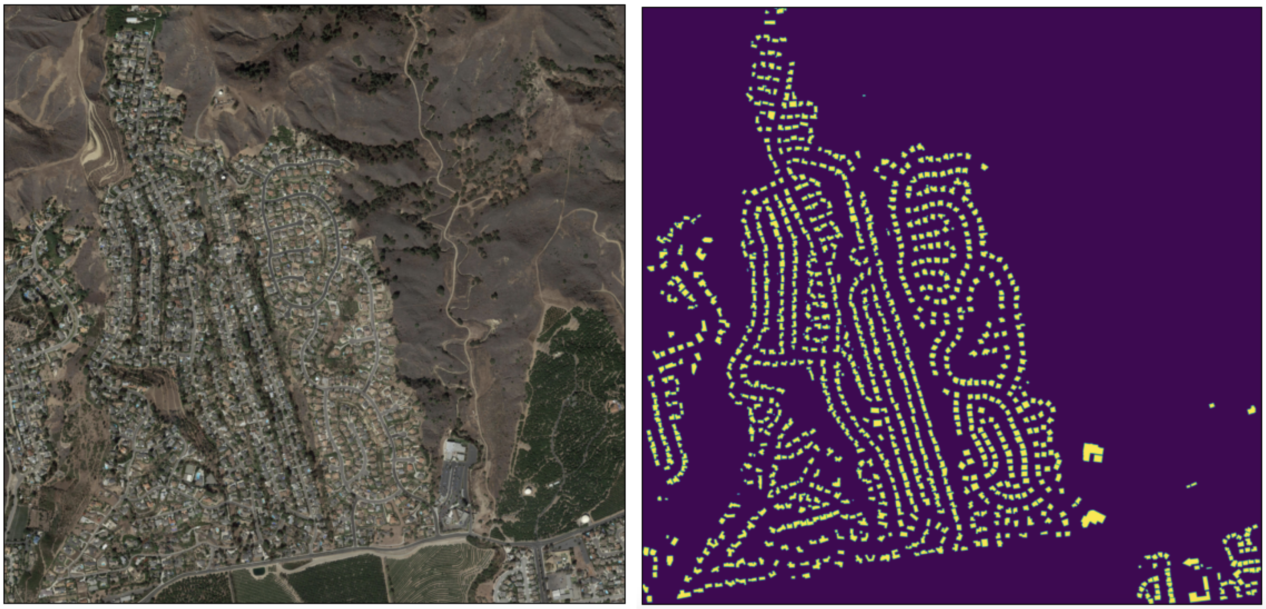

Figure 3. Train area and mask. The image size is $4418 * 4573$ pixels

1. Pretrain model using just generated samples for predefined fixed number of epochs: 5, 10, 15, or 20;
2. Continue training using just original samples for predefined fixed number of epochs: 2, 4, or 8.

Therefore, we aimed to obtain 12 models that utilize for training different proportions of generated and original samples. Such methodological experiments allow us to obtain results that provide important information on the best possible training strategy in order to achieve the highest score. Also, results are useful for performing further analysis of the sensitivity of models performance and training procedure to the developed augmentation technique which in turn allow using the most beneficial aspects of the proposed augmentation algorithm.

	Fine-tuning epochs		
Pretrain epochs	2	4	8
5	0.742	0.774	0.727
10	0.698	**0.795**	0.739
15	0.708	0.736	0.747
20	0.72	0.747	0.763

Table 6. F1-score results with augmentation pretraining, and fine-tuning on original data (U-Net with ResNet-34 encoder).

4.3. Neural Networks Models And Training Details

To evaluate the object-based augmentation approach on different fully convolutional neural networks architectures, we considered FPN [25] and U-Net [39] with three encoders' sizes: ResNet-18, ResNet-34, ResNet-50 [15]. Both U-Net and FPN are popular CNN architectures for semantic segmentation tasks in remote sensing domain [20, 22]. We also train a contemporary high-resolution HR-Net model [43] with the ResNet101 backbone. All models used weights pre-trained on "ImageNet" classification dataset [7].

The training of all the neural network models was performed at a PC with GTX-1080Ti GPUs. For each model, the following training parameters were set. An RMSprop optimizer with a learning rate of 0.001, which was reduced with the patience of 3. There were 50 (except the experiment with augmentation strategies) epochs with 100 steps per epoch and 30 steps for validation. Early stopping was chosen with the patience of 4, then the best model according to validation score was considered. The batch size was specified to be 30 with a crop size of $128 * 128$ pixels. Such a crop size is a typical choice in remote sensing tasks with CNN models [20, 26]. The batch size was chosen according to GPU memory limitations. As a loss function, binary cross entropy (Equation 1) was used.

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N} y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i}), \quad (1)$$

where N is the number of target mask pixels, y is the target mask, $\hat{y}$ is the model prediction.

4.4. Evaluation

The model outputs were binary masks of target objects, which were evaluated against the ground truth with pixelwise F1-score (Equation 2). F1-score is robust for unsymmetrical datasets, and it is a commonly used score for semantic segmentation tasks [5], in particular in the remote sensing domain [20].

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \qquad (2)$$

where TP is True Positive (number of correctly classified pixels of the given class), FP is False Positive (number of pixels classified as the given class while in fact being of other class, and FN is False Negative (number of pixels of the given class, missed by the method).

For each experiment, we run a CNN model three times with different random seeds and average the results.

5. Results and discussion

Standard augmentation	Augmentation	F1-score
No	*Baseline_no_augm*	0.45
	OBA_no_augm	0.66 (+21%)
Yes	*Baseline*	0.788
	OBA_no_shadow	0.811 (+2.3%)
	OBA_no_background	0.81 (+2.2%)
	OBA	0.829 (+4.1%)
	OBA + optimization	0.835 (+4.7%)

Table 7. Experiments with different augmentation setups (F1-score for the test set, U-Net with ResNet-34 encoder).

5.1. Object-based augmentation

We compared different augmentation approaches and presented results in Table 2. Model predictions for the test region are presented in Figure 4. *OBA* allows us to improve the F1-score for the entire dataset size from 0.788 to 0.829 compared with the base color and geometrical transformations (*Baseline*). As experiments clearly indicate, the model trained without any data augmentation (*Baseline_no_augm*) performed significantly poorly (0.45 F1-score).

Extra background usage improves prediction quality compared with models that use only initial background areas both for the original test set (F1-score from 0.81 to 0.829). Additional backgrounds make a model more universal for new regions. It is promising in cases where we want to switch between different environmental conditions without extra labeled datasets.

Even without extra background images, remote sensing task specificity frequently offers an opportunity to add more diversity in training samples. Target objects are often too small compared with the entire satellite image that is leveraged for a particular task. Moreover, target objects can be distributed not regularly which creates large areas free of them. We show that even these areas can be successfully used to create new various training samples (see *OBA_no_background* in Table 7).

We studied artificial shadows importance in the proposed approach. As it is shown in Table 7, shadows allow us to improve the model performance from 0.811 to 0.829 (F1-score). Therefore, a shadow is an essential descriptor for objects observed remotely from satellites. It distinguishes OBA for remote sensing tasks from the copy-paste approach [10] applied in the general computer vision domain.

We tested the proposed approach with different neural networks architectures. The results of the experiments are shown in Table 3. Models with different capacities perform better on different tasks; however, for both U-Net and FPN architectures with different encoders sizes, and the HRNet model the object-based approach outperforms the base augmentation strategy. Object-based augmentation clearly improves generalization in our experiments. One can also see from the results that models with high capacity tend to overfit on small training data when augmentation is not sufficient. It reaffirms the hypothesis that it is essential to research data preprocessing, but not only model architectures.

In Table 6 we study the effect of different proportions of original and generated samples. In this experiment, we show the ability to use the augmented set for model pretraining, and then to tune it further on the original set. The intention is to apply such a pipeline is to start training with a bigger augmented set to learn general patterns and to continue training with a smaller set that is closer to the distribution of the test set to learn more precise patterns. However, this approach faces the ”catastrophic forgetting” problem [13] that means that after a long fine-tuning on a new set, a model forgets the patterns trained on an old set. Another problem is that model can overfit either on augmented or on original set. To solve this, we try several combinations of pretrain and fine-tune epochs. The results show that pretraining in the object-based augmentation mode (without original sample usage) for 10 epochs and further training in the base augmentation mode for 4 epochs for our task leads to the best result for the considered experiment. This experiment indicates that separate training on the original and generated data during different epochs is not an optimal choice in this task. A more efficient approach is to set a probability to add augmented and original samples into each batch during training.

The advantage of this method is that it does not require much computational overhead. It needs just one model training on the generated dataset and tuning the model from several checkpoints on the original dataset. However, as Table 3 shows, the strategy of mixing generated and original images during the training process leads to better results than separating image sources.

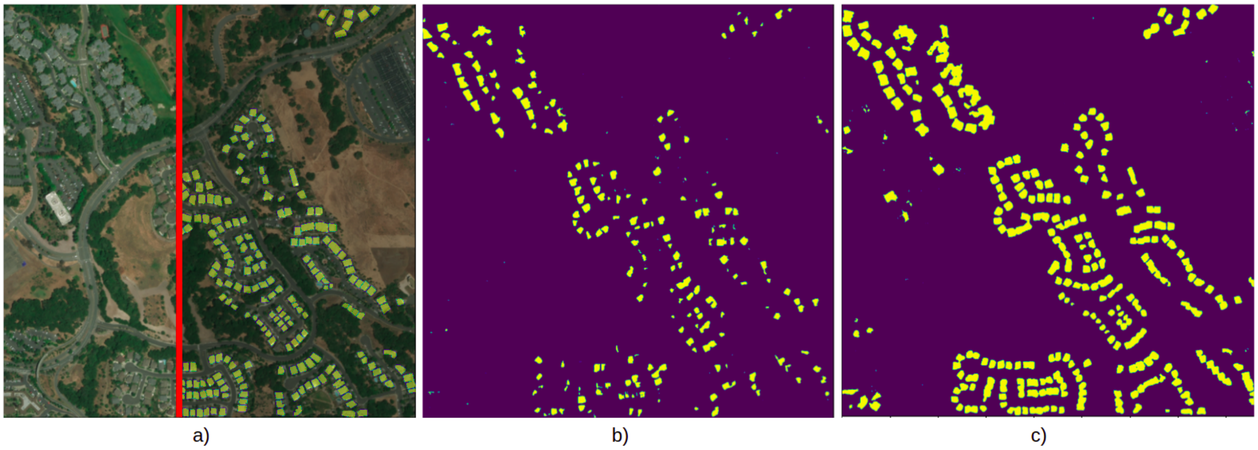

Figure 4. Sample results on the test set of buildings dataset (U-Net with ResNet-34): a) input RGB with ground truth on the right of the red line; b) prediction without augmentation; c) prediction with object-based augmentation.

As we evaluated the OBA approach for remote sensing tasks with man-made objects, one of the future study directions is to implement the described method to wider classes, in particular, vegetation objects, such as agricultural crops or individual trees.

The results for experiments with different dataset sizes are present in Figure 2. Object-based augmentation allows avoiding the drastic drop in prediction quality when dataset size is reduced. For buildings segmentation with object-based augmentation, dataset size decreasing to one-third leads to F1-score decreasing from 0.829 to 0.787, while with the base augmentation it decreases from 0.788 to 0.751. That makes object-based augmentation suitable for few-shot learning, especially when high-capacity models are used.

For the optimization task, we tested U-Net with ResNet-34 encoder using the entire dataset. Optuna package was leveraged to find better values for augmentation parameters, namely, extra objects number ($[0, 1, 2, 3]$), the probability to use additional background ($0 - 1$), object-based augmentation probability ($0 - 1$), and color augmentation probability ($0 - 1$). We run 20 trials; for each trial, parameter values varied. For the optimization process, Optuna utilized loss function values on the validation set after each epoch. As the pruner method, we used MedianPruner.

Augmentation strategy search for U-Net model increases the final performance from 0.829 to 0.835 (*OBA* and *OBA + optimization* in Table 7). The found optimal parameters are as follows: *extra objects* $= 3$; *background prob* $= 0.53$; *object-based augmentation probability* $= 0.787$; *color augmentation probability* $= 0.35$.

As it is shown, the optimizer allows us to set up better augmentation parameters according to the particular task specificity.

6. Conclusion

This study proposes an advanced object-based augmentation approach that outperforms standard color and geometrical image transformations in building semantic segmentation task. The presented method combines target objects from georeferenced satellite images with new backgrounds to produce more diverse realistic training samples. Our framework extends object-based augmentation methods to work with the remote sensing domain. It considers the satellite imagery data format and adds features that improve the accuracy of our case study. We also explicate the importance of augmentation hyperparameters tuning and describe a practical way to find optimal object-based augmentation parameters. Our results show promising potential for real-life remote sensing tasks making CNN models more robust for new environmental conditions even if the labeled dataset size is highly limited.

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[2] Bjorn Barz and Joachim Denzler. Deep learning on small datasets without pre-training using cosine loss. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1371–1380, 2020.

[3] Joseph Bullock, Carolina Cuesta-Lázaro, and Arnau Quera-Bofarull. Xnet: A convolutional neural network (cnn) implementation for medical x-ray image segmentation suitable for small datasets. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10953, page 109531Z. International Society for Optics and Photonics, 2019.

[4] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.

[5] Gabriela Csurka, Diane Larlus, Florent Perronnin, and France Meylan. What is a good evaluation measure for semantic segmentation?. In *BMVC*, volume 27, pages 10–5244, 2013.

[6] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[8] Alhussein Fawzi, Horst Samulowitz, Deepak Turaga, and Pascal Frossard. Adaptive data augmentation for image classification. In *2016 IEEE international conference on image processing (ICIP)*, pages 3688–3692. Ieee, 2016.

[9] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 289–293. IEEE, 2018.

[10] 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. *arXiv preprint arXiv:2012.07177*, 2020.

[11] 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.

[12] KT Gribbon, CT Johnston, and Donald G Bailey. A real-time fpga implementation of a barrel distortion correction algorithm with bilinear interpolation. In *Image and Vision Computing New Zealand*, pages 408–413, 2003.

[13] Steven Gutstein and Ethan Stump. Reduction of catastrophic forgetting with transfer learning and ternary output codes. In *2015 International Joint Conference on Neural Networks (IJCNN)*, volume 1, pages 1–8, 2015.

[14] Jad Haddad, Olivier Lézoray, and Philippe Hamel. 3d-cnn for facial emotion recognition in videos. In *International Symposium on Visual Computing*, pages 298–309. Springer, 2020.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Guoquan Huang, Zining Wan, Xinggao Liu, Junpeng Hui, Ze Wang, and Zeyin Zhang. Ship detection based on squeeze excitation skip-connection path networks for optical remote sensing images. *Neurocomputing*, 332:215–223, 2019.

[17] Hai Huang, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, and Ai-Yun Zang. Faster r-cnn for marine organisms detection and recognition using data augmentation. *Neurocomputing*, 337:372–384, 2019.

[18] Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Neural-based hierarchical approach for detailed dominant forest species classification by multispectral satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1810–1820, 2020.

[19] Naoko Kato, Hiroki Masumoto, Mao Tanabe, Chikako Sakai, Kazuno Negishi, Hidemasa Torii, Hitoshi Tabuchi, and Kazuo Tsubota. Predicting keratoconus progression and need for corneal crosslinking using deep learning. *Journal of clinical medicine*, 10(4):844, 2021.

[20] Teja Kattenborn, Jens Leitloff, Felix Schiefer, and Stefan Hinz. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:24–49, 2021.

[21] Atakan Körez, Necaattin Barışçı, Aydın Çetin, and Uçman Ergün. Weighted ensemble object detection with optimized coefficients for remote sensing images. *ISPRS International Journal of Geo-Information*, 9(6):370, 2020.

[22] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020.

[23] Weijia Li, Conghui He, Jiarui Fang, Juepeng Zheng, Haohuan Fu, and Le Yu. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. *Remote Sensing*, 11(4), 2019.

[24] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. *arXiv preprint arXiv:1905.00397*, 2019.

[25] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[26] Yunling Liu, Chaojun Cen, Yingpu Che, Rui Ke, Yan Ma, and Yuntao Ma. Detection of maize tassels from uav rgb imagery with faster r-cnn. *Remote Sensing*, 12(2):338, 2020.

[27] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019.

[28] Axel H Masquelin, Nicholas Cheney, C Matthew Kinsey, and Jason HT Bates. Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning. *Histochemistry and Cell Biology*, pages 1–9, 2021.

[29] MAXAR. California and colorado fires. `https://www.maxar.com/open-data/california-colorado-fires`, 2017. Accessed: 2021-01-10.

[30] Sergey Nesteruk, Dmitrii Shadrin, and Mariia Pukalchik. Image augmentation for multitask few-shot learning: Agricultural domain use-case, 2021.

[31] S. Nesteruk, D. Shadrin, M. Pukalchik, A. Somov, C. Zeidler, P. Zabel, and D. Schubert. Image compression and

plants classification using machine learning in controlled-environment agriculture: Antarctic station use case. *IEEE Sensors Journal*, 2021.

[32] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449, 2015.

[33] German Novikov, Alexey Trekin, Georgy Potapov, Vladimir Ignatiev, and Evgeny Burnaev. Satellite imagery analysis for operational damage assessment in emergency situations. In *International Conference on Business Information Systems*, pages 347–358. Springer, 2018.

[34] Vladislav Ostankovich, Rauf Yagfarov, Maksim Rassabin, and Salimzhan Gafurov. Application of cyclegan-based augmentation for autonomous driving at night. In *2020 International Conference Nonlinearity, Information and Robotics (NIR)*, pages 1–5. IEEE, 2020.

[35] N. Paton. Automating data preparation: Can we? should we? must we? *In Proceedings of the 21st International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data*, 2019.

[36] Fábio Perez, Cristina Vasconcelos, Sandra Avila, and Eduardo Valle. Data augmentation for skin lesion analysis. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pages 303–311. Springer, 2018.

[37] Geesara Prathap and Ilya Afanasyev. Deep learning approach for building detection in satellite multispectral imagery. In *2018 International Conference on Intelligent Systems (IS)*, pages 461–465, 2018.

[38] Yun Ren, Changren Zhu, and Shunping Xiao. Small object detection in optical remote sensing images via modified faster r-cnn. *Applied Sciences*, 8(5):813, 2018.

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[40] D. Shadrin, A. Menshchikov, A. Somov, G. Bornemann, J. Hauslage, and M. Fedorov. Enabling precision agriculture through embedded sensing with artificial intelligence. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4103–4113, 2020.

[41] Radamanthys Stivaktakis, Grigorios Tsagkatakis, and Panagiotis Tsakalides. Deep learning for multilabel land cover scene categorization using data augmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1031–1035, 2019.

[42] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[43] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[44] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018.

[45] Guang Xu, Xuan Zhu, and Nigel Tapper. Using convolutional neural networks incorporating hierarchical active learning for target-searching in large-scale remote sensing images. *International Journal of Remote Sensing*, 41(11):4057–4079, 2020.

[46] Yifeng Xu, Yang Zhang, Huigang Wang, and Xing Liu. Underwater image classification using deep convolutional neural networks and data augmentation. In *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–5. IEEE, 2017.

[47] Yiming Yan, Zhichao Tan, and Nan Su. A data augmentation strategy based on simulated samples for ship detection in rgb remote sensing images. *ISPRS International Journal of Geo-Information*, 8(6):276, 2019.

[48] Yiming Yan, Yumo Zhang, and Nan Su. A novel data augmentation method for detection of specific aircraft in remote sensing rgb images. *IEEE Access*, 7:56051–56061, 2019.

[49] Xingrui Yu, Xiaomin Wu, Chunbo Luo, and Peng Ren. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience & Remote Sensing*, 54(5):741–758, 2017.

[50] Guanwen Zhang, Jien Kato, Yu Wang, and Kenji Mase. How to initialize the cnn for small datasets: Extracting discriminative filters from pre-trained model. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 479–483. IEEE, 2015.

[51] Lingzhi Zhang, Tarmily Wen, Jie Min, Jiancong Wang, David Han, and Jianbo Shi. Learning object placement by inpainting for compositional data augmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 566–581. Springer, 2020.

[52] Yingwei Zhou. Slot based image augmentation system for object detection. *arXiv preprint arXiv:1907.12900*, 2019.

[53] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[54] Z Zong, C Chen, X Mi, W Sun, Y Song, J Li, Z Dong, R Huang, and B Yang. A deep learning approach for urban underground objects detection from vehicle-borne ground penetrating radar data in real-time. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2019.

[55] Barret Zoph, Ekin D Cubuk, Golnaz Ghiasi, Tsung-Yi Lin, Jonathon Shlens, and Quoc V Le. Learning data augmentation strategies for object detection. In *European Conference on Computer Vision*, pages 566–583. Springer, 2020.