Supplementary Material for Data Augmentation Approaches for Satellite Imagery

Laurel M. Hopkins[1], **Weng-Keen Wong**[1], **Hannah Kerner**[2], **Fuxin Li**[1], **Rebecca A. Hutchinson**[1,3]

[1]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331
[2]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281
[3]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR 97331
{hopkilau,wongwe}@oregonstate.edu, hkerner@asu.edu, {lif,rah}@oregonstate.edu

Appendix

Implementation Details

Sat-CutMix Similar to the original CutMix algorithm, Sat-CutMix generates new ($\tilde{x}$, $\tilde{y}$) samples in an online manner (i.e., during training). The new samples are then used to train the model. The original (x, y) instances are only used for generating ($\tilde{x}$, $\tilde{y}$) pairs and are not used for model training. Implementing Sat-CutMix is straightforward; this module can be called after retrieving a mini-batch and before feeding the mini-batch to the model.

Sat-Trivial For random erasing, we randomly erase anywhere from $\{0 \ldots 9\}$ small pixel groups in the image. We perform the same procedure for random saturation but rather than setting the pixels to black to remove the data, we set the pixels to white to mimic oversaturation. We set the pixels to black/white across all bands within an image to produce the highest level of augmentation. We apply a random amount of Gaussian noise with zero mean and random standard deviation sampled from Uniform(0, 0.04). We performed a parameter sweep over a random set of tasks to determine the ranges for these augmentation methods (Fig. A3, Fig. A4).

Color & Geometric Whenever we used color or geometric augmentations, we used the same magnitude ranges as in the automated methods (as defined by the PyTorch implementation). The only exception was translate which we set to a maximum of 10% of the image height and width. In both the color and geometric augmentation sets, we randomly sample one transformation from the set of possible transformations and apply it with 100% probability, with the exception of flip and rotate which are applied with 50% probability. For the transformations which have magnitudes, we randomly sample a magnitude from the range based on the linear scale described in the main text. Tab. A1 details the ranges of each augmentation.

Transformation	Magnitude range
identity	-
rotate	-
flip	-
translate	0 - 0.10
shear	0.3 & 0.99
auto contrast	-
brightness	0.01 - 1.99
color	0.01 - 1.99
contrast	0.01 - 1.99
equalize	-
posterize	2 - 8
sharpness	0.01 - 1.99
solarize	0 - 255

Table A1: List of transformations, associated ranges.

Tasks

Classification Tasks We selected three classification tasks: UC Merced Land Use (Yang and Newsam 2010), Brazilian Coffee Scenes (Penatti, Nogueira, and Dos Santos 2015), and EuroSAT (Helber et al. 2019). The UC Merced Land Use Dataset consists of aerial images from 21 different land use classes. The classes span categories such as beach, parking lot, buildings, forest, and overpass. The images were collected from 20 cities across the United States and were manually annotated. Each class contains 100 images (Yang and Newsam 2010).

The Brazilian Coffee Scenes dataset is comprised of SPOT satellite images collected in 2005 over four counties in Brazil. Images were labeled by agricultural experts and labeled coffee if more than 85% of the pixels contained coffee and non-coffee if less than 10% of the pixels contained coffee. The dataset consists of 2876 images with an equal split of coffee and non-coffee (Penatti, Nogueira, and Dos Santos 2015).

EuroSAT is made up of 27,000 Sentinel-2 images spanning ten land use and land cover classes collected over 34 countries. The images were visually verified and images with incorrect or unobservable labels were removed. The dataset consists of a set of 13-band multispectral images and a set of 3-band RGB images (Helber et al. 2019). We used the multispectral images in our analysis.

Regression Tasks We selected three regression tasks from Rolf et al. (2021) with increasing complexity: percent forest cover, nighttime light intensity, and elevation. Forest cover is directly observable from satellite imagery and, therefore, should be the most straightforward to predict. Nighttime light intensity itself is not observable from daytime satellite images, however, proxies for nighttime light intensity (e.g., dense urban areas vs. open land) are observable. Elevation on the other hand, is much more difficult to estimate solely from a satellite image. Images for all three tasks were collected across the contiguous United States and were collected based on the sampling schemes of Rolf et al. (2021).

Data preprocessing

To help address spatial autocorrelation, we used the blockCV R package (Valavi et al. 2019) to split datasets containing geographic location information into spatial blocks (label_preprocessing.R). We then used the spatial blocks to assign data points to train, test, and validation sets. For the tasks without location information, we randomly split the data into train, test, and validation sets.

For EuroSAT, the only multispectral dataset in our analysis, we converted the 16-bit Sentinel-2 images to 8-bit images by clipping pixel values to 2750 and then scaling them to be between 1-255 as recommended by Helber et al. (2019).

Model Training & Parameter Tuning

For each task, we tuned the learning rate, batch size, and weight decay. We did not perform an exhaustive search over the hyperparameters, but we performed enough tuning to find suitable values for the more sensitive hyperparameters (learning rate and batch size, in our case) as is recommended by Lacoste et al. (2024). We evaluated learning rates from 1e-6 to 1e-3, batch sizes from 16 to 128, and weight decay from 0.001 to 0.1. We fine-tuned the models for a maximum of 3k epochs and performed early stopping based on the validation set. Tab. A2 details the final hyperparameter values.

The models were trained on a cluster containing a mix of GPUs. The GPUs utilized for our analysis included Nvidia Tesla V100s with 32GB VRAM and 1.5 TB RAM, Nvidia Quadro RTX 8000 with 44 GB VRAM and 768 GB RAM, Nvidia GeForce GTX 1080 Ti GPUs with 11 GB VRAM and 256 GB RAM, Nvidia GeForce GTX 980 Ti GPUs with 6 GB VRAM and 128 GB RAM. The cluster has a Centos Linux 7.9 operating system and CUDA 12.2. Our virtual environment ran Python 3.8, Torch 2.1.1. and Torchvision 0.16.1. The rest of the package versions can be found in our requirements.txt file.

For our proposed method which required parameter tuning (Sat-CutMix), we performed a parameter sweep on a random subset of tasks. We ensured that there was at least one regression task and one classification task. The random subset includes forest cover, Brazilian Coffee Scenes and UC Merced Land Use. We ran five models per task and parameter value and evaluated the performance on the validation set.

Task	Learning rate	Batch size	Weight decay
Brazilian Coffee Scenes	1e-5	64	0.1
UC Merced Land Use	1e-6	50	0.01
EuroSAT	1e-6	16	0.01
Elevation	1e-6	100	0.01
Forest cover	1e-5	32	0.01
Nighttime lights	1e-5	64	0.01

Table A2: Hyperparameters for each task.

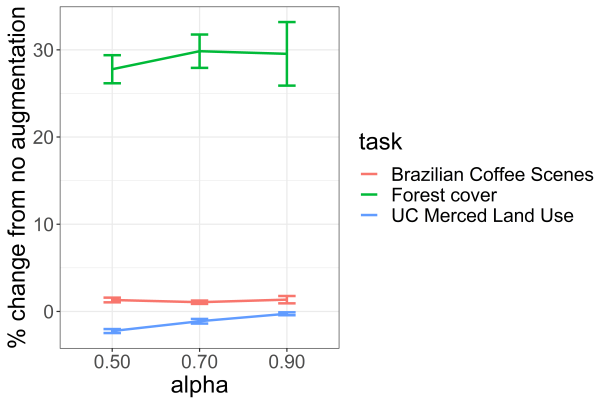

Figure A1: Parameter sweep of α (the lower bound for how much of the base image to keep) for Sat-CutMix with $\gamma = 3$. $\alpha = 0.9$ achieves the best performance across tasks.

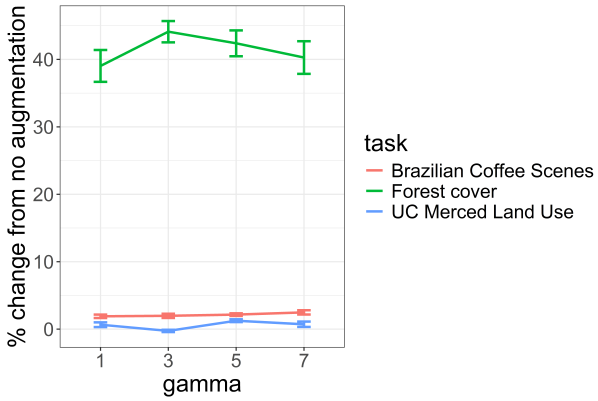

Figure A2: Parameter sweep of γ (number of pairs created) for Sat-CutMix with $\alpha = 0.9$. $\gamma = 3$ achieves the best performance across tasks.

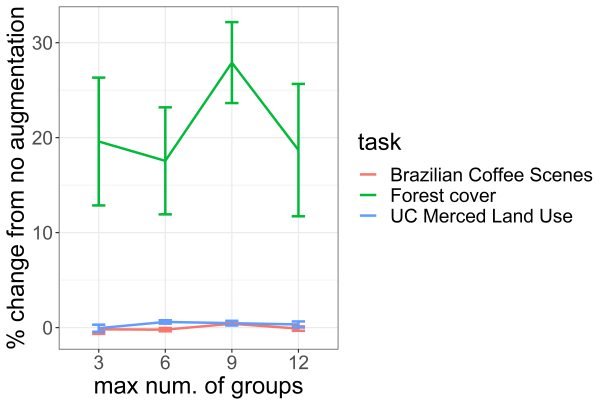

Figure A3: Parameter sweep over the maximum number of pixel groups for the random erase and random saturate augmentations. Setting the maximum number of groups to 9 gives good performance across all tasks.

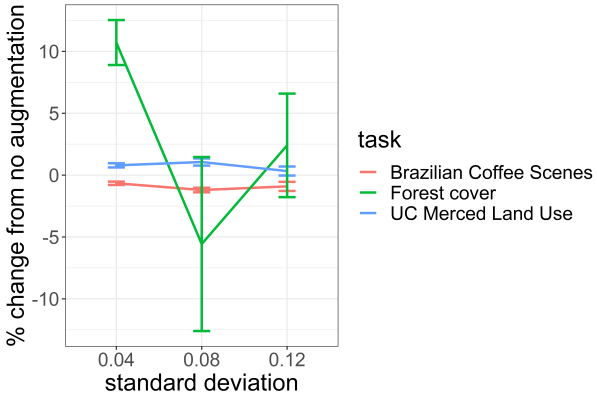

Figure A4: Parameter sweep over different standard deviations for the Gaussian noise. Setting the standard deviation to 0.04 gives the best performance across all tasks.

Additional Results

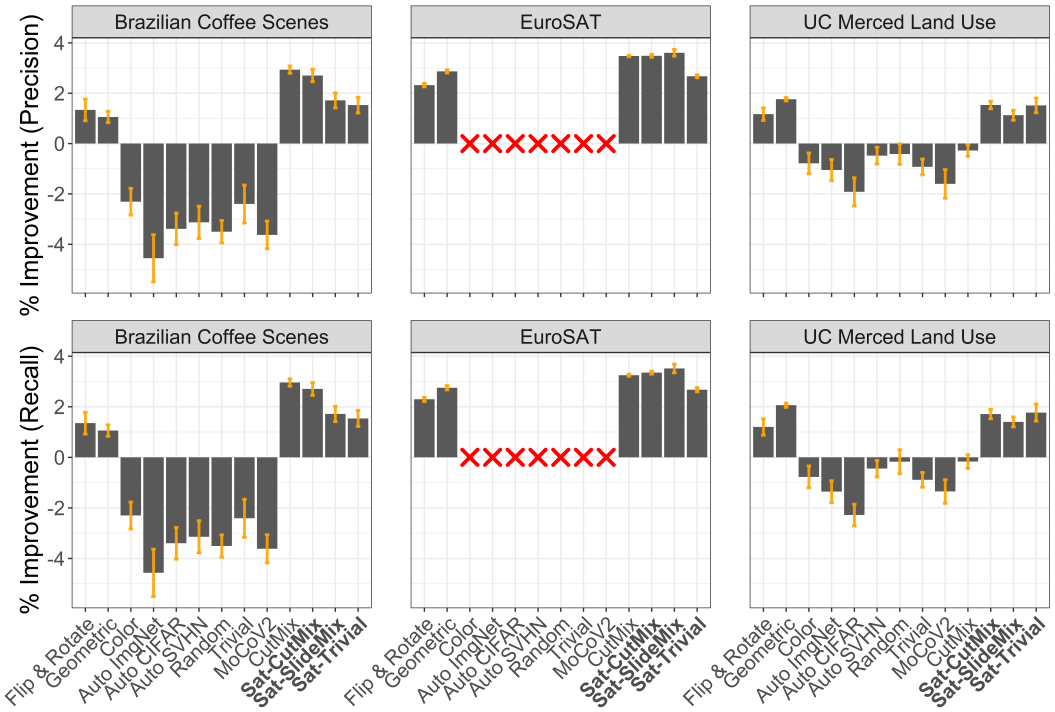

Figure A5: Percent improvement over no augmentation for classification tasks.

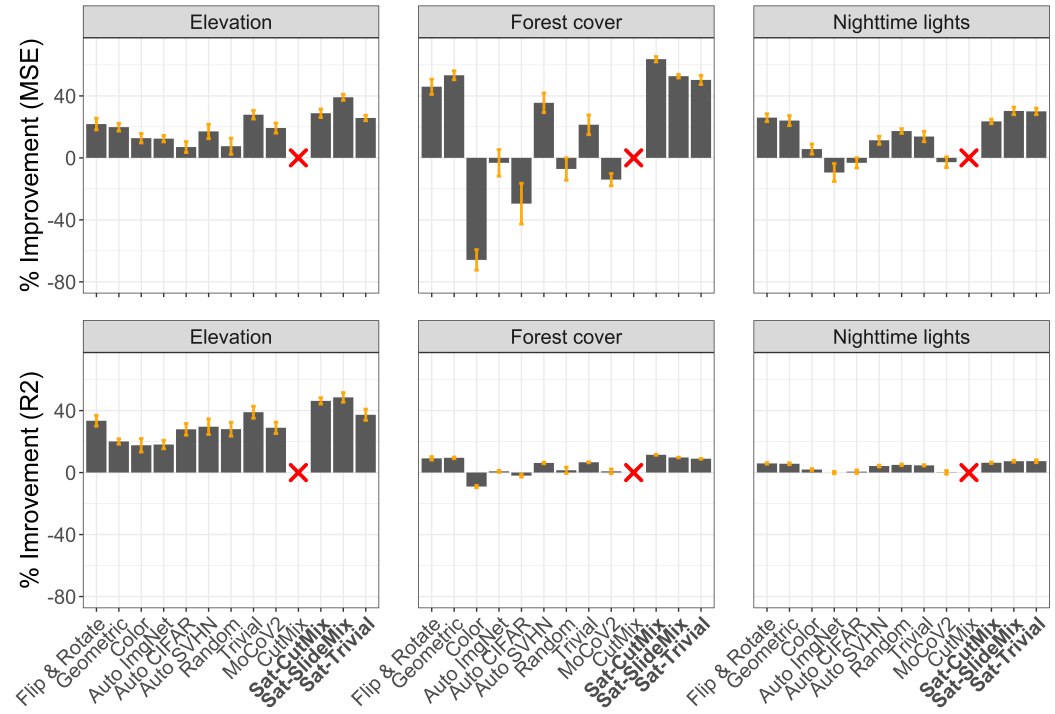

Figure A6: Percent improvement over no augmentation for regression tasks.

Task	Method	No aug.	Flip & Rotate
Brazilian Coffee Scenes	Sat-CutMix	**3.44**	1.83
Brazilian Coffee Scenes	Sat-SlideMix	**2.34**	0.73
Brazilian Coffee Scenes	Sat-Trivial	1.21	-0.40
EuroSAT	Sat-CutMix	**3.49**	**0.96**
EuroSAT	Sat-SlideMix	**3.54**	**1.02**
EuroSAT	Sat-Trivial	**2.63**	0.11
UC Merced Land Use	Sat-CutMix	**1.58**	-0.20
UC Merced Land Use	Sat-SlideMix	1.45	-0.33
UC Merced Land Use	Sat-Trivial	**1.91**	0.13
Elevation	Sat-CutMix	**19.29**	**11.04**
Elevation	Sat-SlideMix	**29.00**	**20.74**
Elevation	Sat-Trivial	**20.30**	**12.04**
Forest cover	Sat-CutMix	**45.43**	18.72
Forest cover	Sat-SlideMix	**43.19**	16.49
Forest cover	Sat-Trivial	**38.38**	11.68
Nighttime lights	Sat-CutMix	**13.55**	-0.60
Nighttime lights	Sat-SlideMix	**19.65**	5.49
Nighttime lights	Sat-Trivial	**18.82**	4.66

Table A3: Mean difference in percent improvement over no augmentation between proposed methods and no augmentation and Flip & Rotate. Tukey-Kramer tests were used to establish statistical significance. Differences found to be statistically significant (p-value < 0.05) are in bold.

References

Helber, P.; Bischke, B.; Dengel, A.; and Borth, D. 2019. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7): 2217–2226.

Lacoste, A.; Lehmann, N.; Rodriguez, P.; Sherwin, E.; Kerner, H.; Lütjens, B.; Irvin, J.; Dao, D.; Alemohammad, H.; Drouin, A.; et al. 2024. Geo-bench: Toward foundation models for earth monitoring. *Advances in Neural Information Processing Systems*, 36.

Penatti, O. A.; Nogueira, K.; and Dos Santos, J. A. 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 44–51.

Rolf, E.; Proctor, J.; Carleton, T.; Bolliger, I.; Shankar, V.; Ishihara, M.; Recht, B.; and Hsiang, S. 2021. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Commun*, 12(4392).

Valavi, R.; Elith, J.; Lahoz-Monfort, J. J.; and Guillera-Arroita, G. 2019. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2): 225–232.

Yang, Y.; and Newsam, S. 2010. Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, 270–279. New York, NY, USA: Association for Computing Machinery. ISBN 9781450304283.