Remote Sensing Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/trsl20

Semantic Segmentation of High Resolution Satellite Imagery using Generative Adversarial Networks with Progressive Growing

Edward Collier, Supratik Mukhopadhyay, Kate Duffy, Sangram Ganguly, Geri Madanguit, Subodh Kalia, Gayaka Shreekant, Ramakrishna Nemani, Andrew Michaelis, Shuang Li & Auroop Ganguly

To cite this article: Edward Collier, Supratik Mukhopadhyay, Kate Duffy, Sangram Ganguly, Geri Madanguit, Subodh Kalia, Gayaka Shreekant, Ramakrishna Nemani, Andrew Michaelis, Shuang Li & Auroop Ganguly (2021) Semantic Segmentation of High Resolution Satellite Imagery using Generative Adversarial Networks with Progressive Growing, Remote Sensing Letters, 12:5, 439-448, DOI: 10.1080/2150704X.2021.1895444

To link to this article: https://doi.org/10.1080/2150704X.2021.1895444

REMOTE SENSING LETTERS
2021, VOL. 12, NO. 5, 439–448
https://doi.org/10.1080/2150704X.2021.1895444

Taylor & Francis
Taylor & Francis Group

Semantic Segmentation of High Resolution Satellite Imagery using Generative Adversarial Networks with Progressive Growing

Edward Collier[a], Supratik Mukhopadhyay[a], Kate Duffy[b], Sangram Ganguly[c], Geri Madanguit[d], Subodh Kalia[d], Gayaka Shreekant[d], Ramakrishna Nemani[e], Andrew Michaelis[e], Shuang Li[e] and Auroop Ganguly[b]

[a]Computer Science Department, Louisiana State University, Baton Rouge, USA; [b]Northeastern University, Boston, USA; [c]NASA Ames Research Center/BAERI; [d]BAER Institute; [e]NASA Ames Research Center

ABSTRACT

With increase in urbanization and Earth Sciences research into urban areas, the need to quickly and accurately segment urban rooftop maps has never been greater. Current machine learning techniques struggle to produce high accuracy maps in dense urban zones where there is high image noise and foot print overlap. In this paper, we evaluate a training methodology for pixel-wise segmentation for high-resolution satellite imagery using progressive growing of generative adversarial networks as a solution. We apply our model to segmenting building rooftops and compare these results to conventional methods for rooftop segmentation. We evaluate our approach using the SpaceNet version 2 and xView datasets. Our experiments show that for SpaceNet, progressive Generative Adversarial Network (GAN) training achieved a test accuracy of 93% compared to 89% for traditional GAN training and 87% for U-Net architecture, while for xView, we achieved 71% accuracy using progressive GAN training compared to 69% through traditional GAN training and 65% using U-Net.

ARTICLE HISTORY
Received 15 July 2020
Accepted 17 February 2021

1. Introduction

Due to the massive, and increasing, amount of satellite data available, a significant effort has been devoted to developing machine learning methods for satellite image processing. Among the higher level products sought, rooftop detection has received particular attention due to the diverse insights available from rooftop products. Rooftop detection is used to track urban growth, estimate population, assess damage from natural disasters and classify land use, among other applications.

Training rooftop segmentation models presents challenges, like the similar appearance of rooftops to other objects such as cars. Rooftops also have dissimilar appearances from city to city. Building shape, building material, and surrounding land cover vary widely from scene to scene and present challenges for transfer of models between cities. As such,

CONTACT Supratik Mukhopadhyay supratik@csc.lsu.edu Louisiana State University, Baton Rouge, USA

no generalizable model yet exists that can accurately detect roofs in the full population of satellite images.

Remote sensing provides one of the fastest, lowest cost methods to gather information about damaged areas in post-disaster damage assessment. Automated generation of high-resolution rooftop products creates a running inventory of assets, which can be leveraged to track damages.

However, generation of high-resolution images presents challenges for traditionally trained deep neural networks. The existence of false positives or multiple foot prints blending together can give in accurate assessments. This is particularly a problem in condensed and noisy urban areas, where traditional neural network methods struggle. In this paper, we aim to solve the short comings of neural networks applied to rooftop segmentation by adapting the progressive training of a generative adversarial network (GAN) Karras et al. (2017) to segmentation. We introduce progressive training to the decoder of and encoder-decoder generator, while allowing the full encoder to learn the best latent encoding for the map. We evaluate the efficacy of rooftop segmentation using multi-spectral satellite images and show how using progressive training can limit the number of false positives and product blending while still producing accurate segmentations. This is, to the best of our knowledge, the first results of progressive training for semantic segmentation. A preliminary version of this paper was published in DMESS 2018, a satellite workshop of ICDM 2018. The GAN Goodfellow et al. (2014) consists of a generator and a discriminator, which are linked through an adversarial training algorithm. The generator learns to generate mappings from the input to the target and the discriminator learns to evaluate them. Feedback from the discriminator enables the generator to produce highly realistic outputs. We employ U-Net architecture, a convolutional neural network consisting of an encoder-decoder, as the generator. We apply progressive growing of the generator and the discriminator. Progressive growing is a transfer learning process wherein increasingly deep networks are trained to learn increasingly complex features. Accuracy of rooftop classification is assessed and results are compared with those of a traditionally trained generative model and with those of non-generative U-Net. Our progressively trained GAN approach beats both traditional GAN and non-generative U-Net in accuracy, by four percent and eight percent respectively on the Spacenet spa (2018) dataset, and by 2% and 6% respectively on the xView Lam et al. (2018) dataset.

2. Related work

Significant accomplishments have been made in computer vision, resulting in increasingly effective state-of-the-art methods for image processing Karki et al. (2017); Basu et al. (2016). Early efforts in automatic rooftop segmentation used methods like edge detection, corner detection, and segmentation into homogeneous regions via k-means clustering or support Vector Machines (SVM) to identify candidate rooftops in Joshi et al. (2014). Discriminative features used to evaluate candidate rooftops include building shadows, geometry, and spectral characteristics Ren et al. (2009); Jin and Davis (2005). Several approaches have used LiDAR alone or in addition to multi-spectral images Wang et al. (2011); Bittner and Korner (2018) Newer-generation machine learning techniques Basu et al. (2017) have also been applied in satellite image classification Basu et al. (2015b); Liu

et al. (2020) and in rooftop segmentation specifically Basu et al. (2015a); Chen et al. (2018). Convolutional neural networks (CNNs) have greatly improved the state-of-the-art in semantic segmentation tasks wherein each pixel in an image is associated with a class label Long, Shelhamer, and Darrell (2015). High-resolution rooftop detection presents a dense prediction problem in which proper pixel-wise labelling is paramount to a produce a product with well-defined rooftops. In Khalel and El-Saban (2018), stacked U-Nets were used that enhanced the results of the previous U-Net. This study found that stacking of just two CNNs outperforms the state-of-the-art method. Introduced in 2015, U-Nets utilize skip connections and an encoder-decoder structure to learn a latent translation from input to output Ronneberger, Fischer, and Brox (2015). CNN performance is sometimes hampered by blurry results, which satisfy the loss function by reducing the Euclidean distance between predictions and the target Pathak et al. (2016). Generative adversarial networks (GAN) address this pitfall by simultaneously training a discriminator network to differentiate between real and generated images Goodfellow et al. (2014). The original classic GAN algorithm Goodfellow et al. (2014) is further improved upon by progressively grown GANs Karras et al. (2017). In working with high-resolution images, GANs run into issues with real and generated images being too easy to discriminate. Progressively grown GANs address this challenge by utilizing transfer learning in deep neural networks Karras et al. (2017).

A preliminary version of this paper appeared in Collier et al. (2018). The present version extends that in Collier et al. (2018) by first introducing a modified algorithm for progressive training that includes smooth fading. This method doubles the number of training cycles by introducing a residual connection over the new layer to the output layer. This step aims to increase the accuracy by preserving the features learned in the previous layer when the new layer is added. Additionally, we further evaluate the methodology on the xView dataset. xView introduces noisy masks with large areas of background noise labelled with positive pixels which helps further evaluate the progressively trained model's ability to minimize false positives and blending.

3. Data preparation

Our experiments are run on the SpaceNet version 2 dataset spa (2018), and the xView dataset Lam et al. (2018). These datasets contain high resolution commercial satellite images along with the masks of building and road footprints, as depicted in Figure 1. The following experiments are run strictly on rooftop segmentation for both SpaceNet and xView. Both datasets, that we have used, are limited to the greater Las Vegas area. We leave other datasets and class segmentations for future work and evaluation. The xView dataset does not contain segmentation masks, but ROI's (region of interests) contained in a geojson file. To circumvent this, we translate the ROI's for each image into an image mask. The positive segments of the mask have a 1 to 1 translation to the area of the image contained with an ROI. The resulting masks are different from those in the SpaceNet data set because they do not snap to the building foot prints. In xView, all the ROIs are rectangles aligned with the image axis. The results can leave artefacts and background in the segmentation that are not actually part of a building footprint. Each mask is paired with the original image and then split into train and test data to complete the dataset. Figure 2 gives the flow of our data preparation for converting a ground truth geojson to a binary mask.

Figure 1. Example labelled images from the Las Vegas SpaceNet dataset.

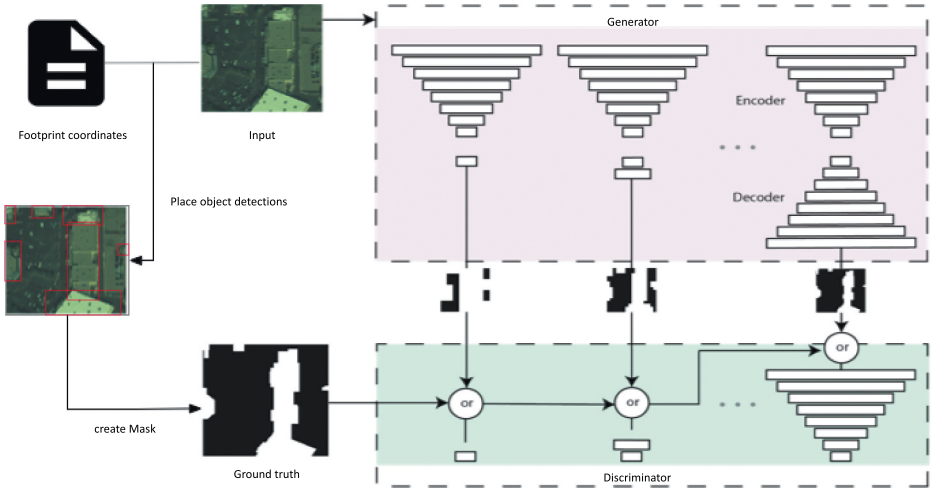

Figure 2. Schematic flow diagram of the data preparation and proposed model that contains U-Net architecture as the generator.

There is a class imbalance in the xView dataset. Building footprints makeup only a portion of the objects contained in xView. If not handled, then the majority of the train and test datasets would contain masks with no positive examples, leading to poor performance. To get around this, we created our own masks using the ROI's for classes that we identified as being buildings with rooftops. The shortfall of this dataset is that the building footprints are not exact outlines of the buildings, but just segment out their area. This provides a different challenge compared to the SpaceNet dataset.

4. Proposed method

Our training algorithm incorporates two primary components: adversarial training and progressive growing. Our method is unique to previous works in progressive growing due to the architecture of the generator and the discriminator Karras et al. (2017). In previous works the generator and discriminator mirror one another; in our model, the generator

instead has an encoder-decoder structure. Our proposed model's architecture and progressive growth are presented in Figure 2.

4.1. Network architecture

Many out of the box segmentation models use the U-Net architecture because of its ability to learn a latent translation between the input and target sets. Additionally, mirrored layers in U-Net contain skip connections that allow structural information to be preserved when decoding from the learned latent encoding. This architecture has become a common generator structure in many domains of GANs. It is for these reasons along with its popularity that we have chosen to use the U-Net architecture in our framework as well.

4.2. GAN training

In the most basic form of a GAN, the generator learns a mapping of $\mathbf{z} \rightarrow \mathbf{y}$, where $\mathbf{z}$ is some random latent vector that is translated onto the feature space defined by the task $\mathbf{y}$. If a GAN is being used to translate one image to another, then the task of the generator is to learn a mapping $\mathbf{x} \rightarrow \mathbf{y}$ from input set $\mathbf{x}$ to target $\mathbf{y}$. This is done by mapping $\mathbf{x}$ to a latent encoding $\mathbf{z}$, $\mathbf{x} \rightarrow \mathbf{z}$, which can be decoded to $\mathbf{y}$, $\mathbf{z} \rightarrow \mathbf{y}$. In our case we seek to learn a mapping between a high-resolution satellite image and the rooftop segment of that image. GANs learn these mappings between inputs and targets via a min/max game, $\min_{\Gamma}\max_{\Delta}(\mathrm{L}(\Gamma, \Delta))$, played between the generator Γ, with inputs $\mathbf{x}$ and $\mathbf{z}$ expressed as $\Gamma(\mathbf{x}, \mathbf{z})$, and the discriminator RLES_A_1899393, with inputs $\mathbf{x}$ and $\Gamma(\mathbf{x}, \mathbf{z})$ denoted by $\Delta(\mathbf{x}, \Gamma(\mathbf{x}, \mathbf{z}))$, with loss $\mathrm{L}(\Gamma, \Delta)$. We express the standard GAN's objective function as Goodfellow et al. (2014):

$$\min_{\Gamma}\max_{\Delta}\mathrm{L}(\Gamma, \Delta) = \mathrm{E}_{\mathbf{y}}[\log_{10}\Delta(\mathbf{y})] + \mathrm{E}_{\mathbf{x},\mathbf{z}}[\log_{10}(1 - \Delta(\mathbf{x}, \Gamma(\mathbf{x}, \mathbf{z})))] \tag{1}$$

In the case of segmentation, we desire the outputs of our generator to be as near as practicable to the ground truth mask. To do this we add the L_1 distance to the objective:

$$\mathrm{L}_1(\Gamma) = \mathrm{E}_{\mathbf{x},\mathbf{y},\mathbf{z}}[|\mathbf{y} - \Gamma(\mathbf{x}, \mathbf{z})|] \tag{2}$$

This imposes a second objective for generator's output: to mirror the ground truth by forcing the generator to minimize the absolute distance between its output and the ground truth mask. Absolute error (L_1 distance) is used rather than mean squared error (L_2 distance) to discourage blurring.

4.3. Progressive growing

In a progressive growing algorithm, layers are added to the generator and discriminator as training moves forward. As layers are added to the networks, generated images increase in spatial resolution. While all layers remain trainable throughout the training period, progressive growing allows Generator Γ and Discriminator Δ to learn increasingly fine scaled features on increasingly high-resolution images. Learning step by step presents a series of simpler tasks to the model. Progressive training is consequently more stable and more efficient than traditional training. Layers in progressive growing are not added to the network for training one after another. Instead, layers are added using a technique called smooth fading. In smooth fading, new higher resolution layers are added in two steps. First

the new layer is added to the network, but treated as a residual block with a skip connection. For the encoder in the generator, the upsampled encoding is passed through an RGB output layer and merged with the RGB output of the new high-resolution layer to produce a faded output that is fed to the discriminator. In the discriminator, the faded output from the generator feeds into both the new higher resolution layer and directly to the following lower resolution layer with a downsampling and skip connection. Progressive learning takes advantage of a deep neural networks' ability to learn features from generic to specific, or low to high resolution. At each progressive step, the weights learned for all the layers in the last step are transferred to identical layers in the next step. This transfer leaves only one untrained layer at each step. By progressively adding layers, the network learns the features at each resolution independently, easing the learning task of each progressive network. We employ this technique to produce masks that mirror the input high resolution in sharpness. Traditionally, progressive GANs are employed for generative tasks. We, however, seek to apply it to translation, specifically segmentation. By using an encoder-decoder structure in the generator, we rely on the encoder to map the high-resolution input to a latent vector which is translated by the decoder. Like in traditional progressively growing GANs, the decoder is progressively trained. Because we desire the decoder to decode from a latent vector containing all the information contained in the high spatial resolution of our input, the encoder is not progressively grown. The encoder instead maintains its full structure throughout the progressive training. The discriminator grows in sequence with the decoder. This trains each successive layer to discriminate specific resolutions.

5. Experimental evaluation

In this section, we compare the results of our progressive GAN model to that from a standard U-Net model and a traditionally-trained GAN model Goodfellow et al. (2014) that is not progressively trained. We choose to use the U-Net model with residual connections as it is a traditional model that has been well researched and adapted to segmentation many times Long, Shelhamer, and Darrell (2015); Khalel and El-Saban (2018). The U-Net is also built identically to the generator used in the progressive GAN, allowing us to further isolate the effects of progressive training. Similarly, we use a standard GAN built identically to our progressive GAN to discern the difference between standard training and progressive training. We compare the results both visually and numerically by taking the per-pixel error of the masks.

5.1. Implementation details

For our experiment, the U-Net model has an encoder-decoder architecture. The encoder is built of eight hidden layers with 64, 128, 256, 512, 512, 512, 512 and 512 hidden units per layer. The decoder is built of eight hidden layers that mirror the encoder. This U-Net is used as the generator of the GAN. The GAN discriminator is built with the same architecture as the decoder, and grows in conjunction with it. Batch normalization with momentum = 0.9 and dropout with probability = 0.5 are employed during training to discourage overfitting.

5.2. Results

From Figure 3, we can see that model inferences of rooftop location for SpaceNet generally match the size and shape of ground truth. While border regions of rooftops leave room for improvement, we show progressive growing improves the definition of individual buildings compared with its counterparts. In the two non-progressive methods, we can see that the building segments tend to blend together more than in progressive growing. Additionally, we can see that the standard GAN Goodfellow et al. (2014) and non-GAN approaches suffer from false positives where the progressive growing is able to minimize this. The xView dataset poses a more difficult problem for the progressive model as we can see in Figure 4. Because xView is not naturally made for segmentation, the

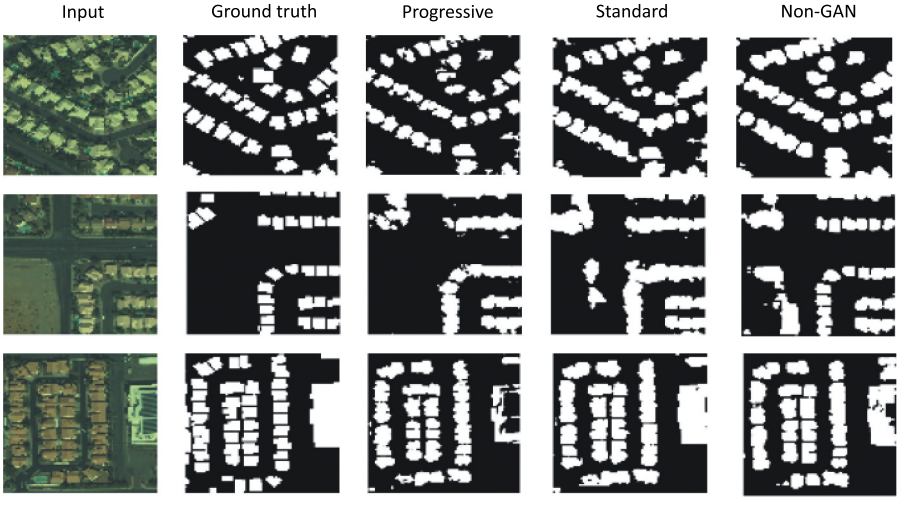

Figure 3. Results of applying the three tested methods to sample images along with the input and the ground truth mask.

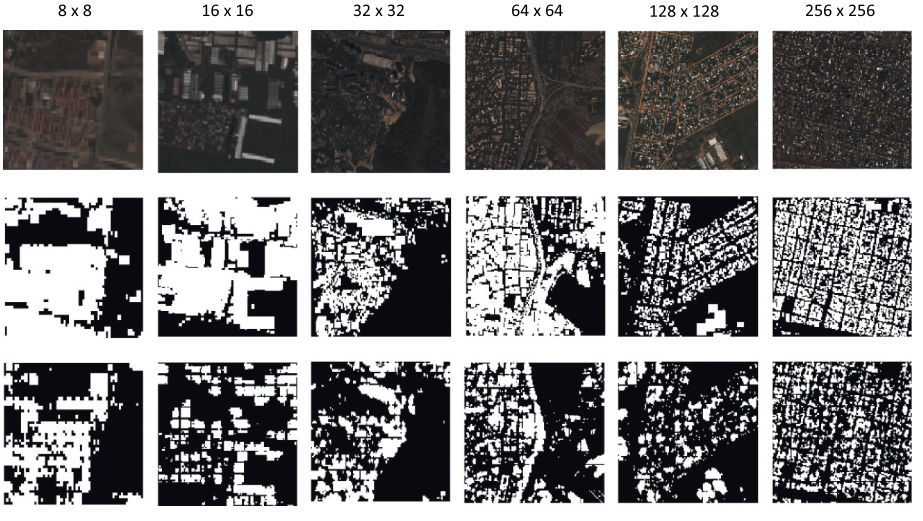

Figure 4. Progressive GAN results by resolution for xView images of different resolutions up to 256 × 256.

footprints in the masks blend together and background features commonly get caught in the footprint. This can cause the accuracy to fall off, with progressive GAN showing high rates of false negatives.

The progressive model shows the ability to limit false positives compared to the other methods. The cause of this is due to the specificity of features in the later layers. This results in the progressively trained model preferring to not label pixels over committing false positives. Figure 5 presents progressive GAN output demonstrating this phenomenon. As a whole, the progressively trained GAN produces building footprints that snap to the original nicely while also minimizing the amount of false-positive pixels compared to the standard methods. This is more conspicuous with the xView dataset. The mixing of background and building features confuses the progressive model, which tends heavily towards not guessing, as it has difficulty separating foreground from background, due to the training dataset limitations.

We present our accuracy scores as the per-pixel error between the ground truth mask and the masks produced by our models. The per-pixel provides a good view of how well the produced masks fit to the high-resolution buildings. From Table 1 we can see that the progressively trained GAN outperforms its counterparts in this metric. We also present both the training and testing accuracy of our models to verify that none have over-fit to the dataset.

In Figure 6, we present graphs for the accuracy of each method during training on SpaceNet. We can see that for the progressively trained GAN that each progressive step builds on top of the previous. The decreased loss and quicker convergence at each step show that there is good transfer of knowledge between the previous and successive steps. Another interesting result is the closeness of the higher resolution layers. This suggests that there exists an image resolution, in our case 256 $\times$ 256 pixels, for which all following image resolutions cannot be used to learn increasingly fine features.

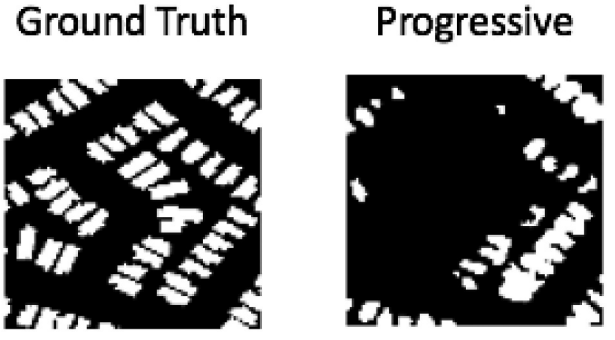

Figure 5. Results of applying progressively trained model showing how the progressively trained model tends to leave space blank rather than classify possible false positives.

Table 1. Summary of testing and training performance for U-Net, GAN and Progressive GAN.

	U-Net	GAN	Progressive GAN
SpaceNet Training accuracy	0.87	0.91	0.94
SpaceNet Testing accuracy	0.85	0.89	0.93
xView Training accuracy	0.69	0.70	0.73
xView Testing accuracy	0.65	0.69	0.71

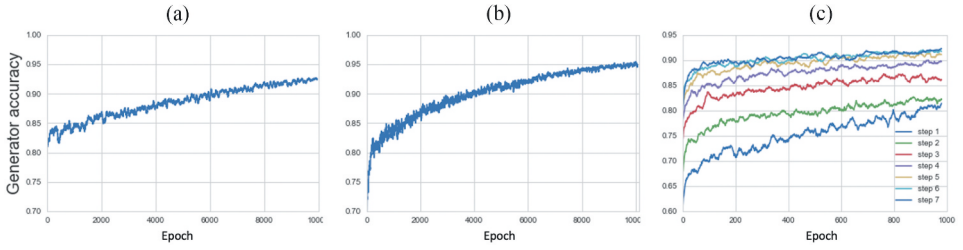

Figure 6. Generator loss and accuracy over training Epochs for U-Net (a), GAN (b) and Progressive GAN (c).

6. Conclusion

This paper presents a novel approach to semantic segmentation for high-resolution satellite imagery that draws upon recently developed machine learning techniques. We use progressive training for semantic segmentation of rooftop products to create tighter fitting segmentation masks with less false positives that previous approaches. Our method was tested on both the SpaceNet and xView datasets. Each posed their own unique set of challenges for the model, such as the background noise being labelled as positive pixels in the xView ground truth. The experiments show that using progressive training is indeed able to reduce the number of false positives and product blending, even in the noisy xView data. The trade-off to this approach is that there is the potential for more false negatives than the traditional U-Net approaches tested.

Acknowledgments

This project was supported by the NASA Earth Exchange (NEX) and NASA CMS grant #NNH16ZDA001N-CMS.

Funding

This work was supported by the NASA Earth Exchange (NEX) and NASA CMS grant [NNH16ZDA001N-CMS.].

References

Basu, S., S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani. 2015a. "Deepsat: A Learning Framework for Satellite Imagery." *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems* 37 (10):1-37. ACM.

Basu, S., S. Ganguly, R. R. Nemani, S. Mukhopadhyay, G. Zhang, C. Milesi, A. R. Michaelis et al. 2015b. "A Semiautomated Probabilistic Framework for Tree-Cover Delineation from 1-m NAIP Imagery Using A High-Performance Computing Architecture." *IEEE Transactions on Geoscience and Remote Sensing* 53 (10): 5690–5708. doi:10.1109/TGRS.2015.2428197.

Basu, S., M. Karki, S. Ganguly, R. DiBiano, S. Mukhopadhyay, S. Gayaka, R. Kannan, and R. R. Nemani. 2017. "Learning Sparse Feature Representations Using Probabilistic Quadtrees and Deep Belief Nets." *Neural Processing Letters* 45 (3): 855–867. doi:10.1007/s11063-016-9556-4.

Basu, S., M. Karki, S. Mukhopadhyay, S. Ganguly, R. R. Nemani, R. DiBiano, and S. Gayaka. 2016. "A Theoretical Analysis of Deep Neural Networks for Texture Classification." In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada*, 24-29 July 2016, 992–999.

Bittner, K., and M. Korner. 2018. "Automatic Large-Scale 3D Building Shape Refinement Using Conditional Generative Adversarial Networks." In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June. Salt Lake City, UT.

Chen, Q., L. Wang, W. Yifan, W. Guangming, Z. Guo, and S. L. Waslander. 2018. "Aerial Imagery for Roof Segmentation: A Large-Scale Dataset Towards Automatic Mapping of Buildings." *arXiv Preprint* arXiv: 1807.09532.

Collier, E., K. Duffy, S. Ganguly, G. Madanguit, S. Kalia, S. Gayaka, R. R. Nemani, et al. 2018. "Progressively Growing Generative Adversarial Networks for High Resolution Semantic Segmentation of Satellite Images." In *2018 IEEE International Conference on Data Mining Workshops, ICDM Workshops*, *Singapore, Singapore*, 17- 20 November 2018, edited by H. Tong, Z. J. Li, F. Zhu, and Y. Jeffrey, 763–769. IEEE. doi: 10.1109/ICDMW.2018.00115.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, X. Bing, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative Adversarial Nets". *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 2672–2680. Curran Associates, Inc. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

Jin, X., and C. H. Davis. 2005. "Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information." *EURASIP Journal on Advances in Signal Processing* 2005 (14): 745309. doi:10.1155/ASP.2005.2196.

Joshi, B., H. Baluyan, A. A. Hinai, and W. L. Woon. 2014. "Automatic Rooftop Detection Using a Two-Stage Classification." In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, March, 286–291. Cambridge, United Kingdom.

Karki, M., R. DiBiano, S. Basu, and S. Mukhopadhyay. 2017. "Core Sampling Framework for Pixel Classification." In *Artificial Neural Networks and Machine Learning - ICANN 2017-26th International Conference on Artificial Neural Networks*, *Alghero, Italy*, 11-14 September 2017, *Proceedings, Part II*, 617–625.

Karras, T., T. Aila, S. Laine, and J. Lehtinen. 2017. "Progressive Growing of GANs for Improved Quality, Stability, and Variation." *CoRR* abs/1710.10196. http://arxiv.org/abs/1710.10196

Khalel, A., and M. El-Saban. 2018. "Automatic Pixelwise Object Labeling for Aerial Imagery Using Stacked U-Nets." *arXiv Preprint* arXiv: 1803.04953.

Lam, D., R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric, Y. Bulatov, and M. Brendan. 2018. "Xview: Objects in Context in Overhead Imagery." *arXiv Preprint* arXiv: 1802.07856.

Liu, Q., S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani. 2020. "DeepSat V2: Feature Augmented Convolutional Neural Nets for Satellite Image Classification." *Remote Sensing Letters* 11 (2): 156–165. doi:10.1080/2150704X.2019.1693071.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June. Boston, Massachusetts.

Pathak, D., P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. 2016. "Context Encoders: Feature Learning by Inpainting." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2536–2544. Las Vegas, Nevada.

Ren, K., H. Sun, Q. Jia, and J. Shi. 2009. "Building Recognition from Aerial Images Combining Segmentation and Shadow." *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems* 4 (November): 578–582.

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *CoRR* abs/1505.04597. http://arxiv.org/abs/1505.04597

2018. "SpaceNet on Amazon Web Services (AWS)." https://spacenetchallenge.github.io/datasets/datasetHomePage.html

Wang, X., Y. Luo, T. Jiang, H. Gong, S. Luo, and X. Zhang. 2011. "A New Classification Method for LIDAR Data Based on Unbalanced Support Vector Machine." In *2011 International Symposium on Image and Data Fusion*, August, 1–4. Tengchong, Yunnan, PR China.