Predicting Landsat Reflectance with Deep Generative Fusion

Shahine Bouabid[*]   **Maxim Chernetskiy**   **Maxime Rischard**   **Jevgenij Gamper**

Cervest Ltd.
London, UK

Abstract

Public satellite missions are commonly bound to a trade-off between spatial and temporal resolution as no single sensor provides fine-grained acquisitions with frequent coverage. This hinders their potential to assist vegetation monitoring or humanitarian actions, which require detecting rapid and detailed terrestrial surface changes. In this work, we probe the potential of deep generative models to produce high-resolution optical imagery by fusing products with different spatial and temporal characteristics. We introduce a dataset of co-registered Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat surface reflectance time series and demonstrate the ability of our generative model to blend coarse daily reflectance information into low-paced finer acquisitions. We benchmark our proposed model against state-of-the-art reflectance fusion algorithms.

1   Introduction

Amid climate-induced stress on land resource management, detecting land cover changes associated with vegetation dynamics or natural disaster in a timely manner can critically benefit agriculture, humanitarian response, and Earth science. Orbiting remote sensing devices stand as prime assets in this context, continuously providing multi-spectral terrestrial surface imaging [29]. They assist in decision-making with valuable large-scale perspectives that enable assessment of vegetation indices [22, 18], forest wildfire spread [28], and flooding risks [27, 2].

Unfortunately, a trade-off between spatial and temporal resolution in this type of imaging is often observed. For example, sensors with a large scanning swath are able to cover wide regions at once, enabling high revisit frequency but also resulting in poor spatial resolution. The Moderate Resolution Imaging Spetroradiometer (MODIS) sensors [1, 20] offer a daily revisit cycle that is well suited to tracking rapid surface changes but only capture ground resolution cells ranging from 250-500m. Conversely, sensors with greater spatial resolution provide more precise views but take longer to revisit. For instance, Landsat missions [16, 25] provide imagery at 30m resolution, which is sufficient to discern individual crops. However, these missions suffer from 16-day revisit cycles and frequently encounter issues of cloud occluding. These technical constraints undermine the availability of free remote sensing imagery with sufficient coverage to meet the needs of precision agriculture or geohumanitarian actions.

Surface reflectance fusion is the task of combining imagery products with different characteristics to synthesize a mixed reflectance product with sufficient spatiotemporal resolution for land-cover monitoring. The complementarity of low-paced fine-resolution Landsat acquisitions and coarse daily MODIS updates makes them natural candidates for surface reflectance fusion. Statistical approaches have been successfully introduced [10, 35, 33] that leverage consistency between Landsat and MODIS in reflectance [19], but these techniques rely heavily on the temporal density of the available satellite

---

[*]Work completed as part of the Cervest Residency program. Correspondence to `jev@cervest.earth`

acquisitions. Alternatively, deep learning has shown promise when applied to specific remote sensing tasks [8, 12, 31, 36, 4, 26]. Notably, deep generative models have demonstrated their capacity to fuse multiple frames with low spatial resolution into a higher resolution one [6, 4].

In this work, we propose to address surface reflectance fusion with deep generative models applied to Landsat and MODIS images. Our contribution is threefold: **(i)** we propose a deep generative framework to estimate Landsat-like reflectance; **(ii)** we introduce a dataset of paired Landsat and MODIS reflectance time series; **(iii)** we conduct a quantitative and qualitative evaluation against state-of-the-art reflectance fusion algorithms.

2 Background

The introduced dataset contains time series of co-registered Landsat and MODIS reflectance images at various sites. Let $\mathcal{D} = \{\mathbf{l}_{t_i}, \mathbf{m}_{t_i}\}_{i=1}^N$ denote a time series with $\mathbf{l}_{t_i}, \mathbf{m}_{t_i} \in \mathbb{R}^{B \times W \times H}$ representing the Landsat and MODIS fields, respectively, at date t_i. The image dimensions B, W, H are respectively the number of spectral bands, image width, and image height. Note that the MODIS frames have been georeferenced and resampled to Landsat resolution. Furthermore, Landsat and MODIS spectral bands typically cover the same domains such that they can be considered consistent in reflectance [19].

Consider a coupled pair of Landsat and MODIS images at date t, $\{\mathbf{l}_t, \mathbf{m}_t\}$, and a corresponding pair at some future prediction date t_p, $\{\mathbf{l}_{t_p}, \mathbf{m}_{t_p}\}$. The surface reflectance fusion problem may then be framed as predicting $\mathbf{l}_{t_p}$ given $\{\mathbf{l}_t, \mathbf{m}_t\}$ and $\mathbf{m}_{t_p}$. Formally, this entails that homogeneous pixels in spatial position $[i, j]$ should satisfy $\mathbf{l}_{t_p}[:, i, j] = \mathbf{m}_{t_p}[:, i, j] + \mathbf{l}_t[:, i, j] - \mathbf{m}_t[:, i, j]$.

Adaptative Reflectance Fusion Models

Gao, et al. [10] propose a spatial and temporal adaptative reflectance fusion model (STARFM) based on the above relationship to estimate daily Landsat-like surface reflectance. By extending the interpolation range to a surrounding spatial, temporal and spectral window around pixel $[i, j]$, they better account for heterogeneous pixels and changes between acquisition dates.

STARFM has since been successfully applied to estimate daily Landsat-like reflectance and assist vegetation monitoring [7, 30, 9, 13]. Improved version such as Enhanced STARFM (ESTARFM) [35] and Unmixing STARFM (USTARFM) [33] further focus on the treatment of endmembers from heterogeneous coarse pixels and allowed to better cope with fine grained landscapes and cloud contamination.

Deep Generative Imagery in Remote Sensing

Neural networks have displayed compelling performance at generative tasks in remote sensing such as cloud removal from optical imagery [12], SAR-to-Optical image translation [8, 31, 36] or spatial resolution enhancement [6, 17, 4]. In particular, work on multi-frame super-resolution has demonstrated the capacity of deep neural networks to combine information from multiple low resolution images to produce finer grained outputs [6, 4]. Rudner, et al. [26] have also pointed out how fusing satellite products with complementary spatial, temporal and spectral information can benefit tasks such as flooded building segmentation.

Generative adversarial networks (GANs) [11] have demonstrated impressive capacity for natural image generation [5], image translation tasks [14] and single-image super-resolution [3]. These successes highlight the ability of GANs to impute learnt details from images with poor resolution. The conditional GANs (cGANs) paradigm has received particular interest in remote sensing as it allows generated samples to be conditioned on specific inputs. For example, conditioning on radar images has shown promising performance at generating realistic cloud-free optical remote sensing imagery [12, 31, 8].

3 Method

In the following, we address the task of predicting a Landsat-like image $\mathbf{l}_{t_i}$ at date t_i given MODIS image at the same date $\mathbf{m}_{t_i}$ and last known Landsat image $\mathbf{l}_{t_{i-1}}$. Although we use Landsat and

MODIS reflectance, we bring to the reader's attention that this fusing rationale remains compatible with other remote sensing products.

Supervised cGAN for Landsat Reflectance Prediction

Conditional GANs are a class of generative models where a generator G learns a mapping from random noise $\mathbf{z}$ and a conditioning input $\mathbf{c}$, to an output $\mathbf{y}$. That is, $\mathbf{y} = G(\mathbf{z}|\mathbf{c})$. The generator is trained adversarially against a discriminator D, which in turn learns to estimate the likelihood $D(\mathbf{y}|\mathbf{c})$ of sample $\mathbf{y}$ being real or generated by G.

Suppose we want to estimate the Landsat surface reflectance $\mathbf{l}_{t_i}$ at date t_i. Doing so would require ground-level structural information of the site, which could be obtained from the last known Landsat image $\mathbf{l}_{t_{i-1}}$. However, information about actual reflectance at the current date t_i could be inferred from the corresponding coarse resolution MODIS image $\mathbf{m}_{t_i}$. Let $\mathbf{c}_{t_i}$ denote the concatenation of these two fields along spectral bands, $\mathbf{c}_{t_i} = \mathrm{Concat}\left(\mathbf{l}_{t_{i-1}}, \mathbf{m}_{t_i}\right)$.

We approach this problem using the cGANs framework and propose to train a generator to predict Landsat-like surface reflectance, $\hat{\mathbf{l}}_{t_i} = G(\mathbf{z}|\mathbf{c}_{t_i})$. We augment the objective with a supervised component to constrain G to output images close to $\mathbf{l}_{t_i}$. Specifically, we incorporate an L_1 penalty to capture low-frequency image components, while inducing less blurring than an L_2 norm. Also, a structural similarity index measure (SSIM) [32] criterion fosters the generation of higher frequency components expressed through local pixels dependencies.

Subsequently, the generative reflectance fusion problem is written as the minimax two-player game

$$\min_{G} \max_{D} \mathcal{L}_{\mathrm{cGAN}}(G, D) + \alpha \mathcal{L}_{L_1}(G) + \beta \mathcal{L}_{\mathrm{SSIM}}(G), \tag{1}$$

where the loss terms are given by $\mathcal{L}_{\mathrm{cGAN}}(G, D) = \mathbb{E}\left[\log D(\mathbf{l}_{t_i}|\mathbf{c}_{t_i})\right] + \mathbb{E}\left[\log\left(1 - D(G(\mathbf{z}|\mathbf{c}_{t_i})|\mathbf{c}_{t_i}))\right)\right]$, $\mathcal{L}_{L_1}(G) = \mathbb{E}\left[\|\mathbf{l}_{t_i} - G(\mathbf{z}|\mathbf{c}_{t_i})\|_1\right]$ and $\mathcal{L}_{\mathrm{SSIM}}(G) = \mathbb{E}\left[1 - \mathrm{SSIM}(\mathbf{l}_{t_i}, G(\mathbf{z}|\mathbf{c}_{t_i}))\right]$. Expectations are taken over all possible $\{\mathbf{l}_{t_i}, \mathbf{c}_{t_i}\}$ pairs and $\alpha, \beta > 0$ are supervision weight hyperparameters.

Dataset

The study area is in the department of Marne, within the Grand Est region of France. It is mostly constituted of crops, forest and urban areas. We acquire Landsat-8 30m reflectance imagery for 14 dates spanning from 2013 to 2020, and MODIS 500m surface reflectance product (MCD43) for the same dates. Images are reprojected to the same coordinate system and MODIS frames are bilinearly resampled to the Landsat resolution and bounds. We limit the spectral domain to red, near-infrared, blue and green bands. Contaminated image regions are discarded using quality assessment maps.

We extract 256×256 non-overlapping registered Landsat and MODIS patches at each date, resulting in 548 distinct patches and a total of 5671 Landsat-MODIS pairs when accounting for multiple time steps. The dataset is split into 383, 82 and 83 patches locations for training, validation and testing.

4 Experiments

Experimental setup

Inspired by the success of `pix2pix` [14] based approaches in remote sensing [12, 31, 8], we use U-Net architecture [24] for the generator and a PatchGAN discriminator [14]. U-Net skip connections directly pass the spatial structure from the last know Landsat frame across the network. This allows the intermediate layers to only learn variations from this baseline. We provide stochasticity in the training process using dropout layers. The PatchGAN discriminator jointly classifies local regions of the image, unlike image-level discriminators that process the entire input as a whole. This localized analysis fosters generation of high-frequency components to mimic realistic image textures. All code is made available [2].

Quantitative evaluation of generated samples relies on several full-reference image quality metrics: peak-signal-to-noise ratio (PSNR), SSIM [32], and spectral angle mapper (SAM) [34] scores.

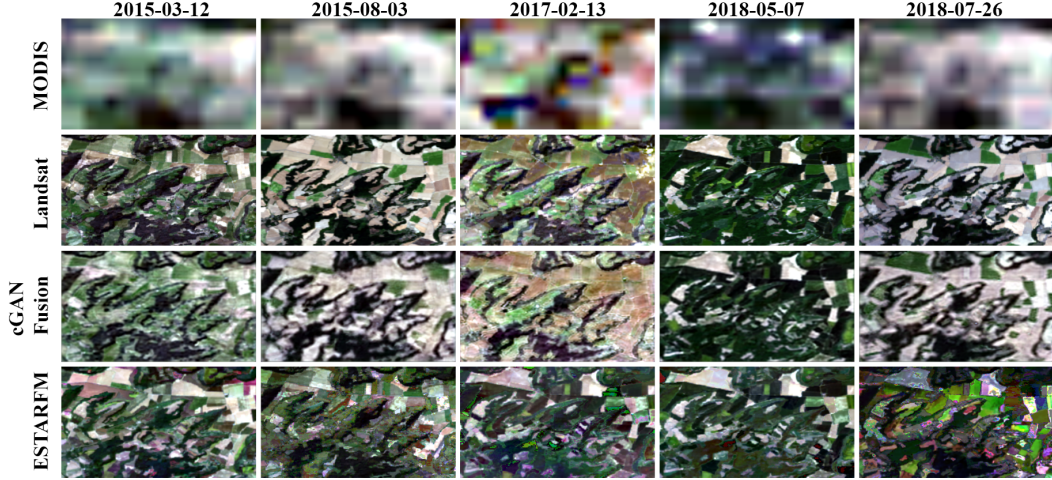[2] `https://github.com/Cervest/ds-generative-reflectance-fusion`

Figure 1: Example of Landsat-like surface reflectance prediction time series, each column is a different time step; "cGAN Fusion" is infered from MODIS at same date and previous Landsat; "ESTARFM" is infered from MODIS at same date and Landsat-MODIS pairs at previous and next dates

Results

Method	PSNR				SSIM				SAM (10^{-2})
Band	NIR	R	G	B	NIR	R	G	B	
Bilinear Upsampling	20.0	19.0	21.0	21.1	0.568	0.550	0.633	0.639	3.87
ESTARFM [35]	19.6	20.2	21.8	22.3	0.555	0.640	0.688	0.696	4.88
cGAN Fusion + L_1	22.1	21.8	23.7	23.8	0.675	0.697	0.747	0.747	2.75
cGAN Fusion + L_1 + SSIM	**22.3**	**22.0**	**23.9**	**24.0**	**0.694**	**0.714**	**0.761**	**0.760**	**2.70**

Table 1: Image quality scores on testing set; cGAN models scores are averaged over 3 independently trained models

We compare our method against ESTARFM implementation [23], which requires at least two pairs of Landsat-MODIS images at distinct dates and MODIS reflectance at prediction date. Table 1 highlights the substantial improvement in image quality metrics on an independent testing set for which the SSIM objective has not been optimized.

Figure 4 provides a qualitative comparison of generated Landsat-like surface reflectance from the cGANs approach and the ESTARFM method. We see that the cGANs approach demonstrates flexibility at capturing and blending MODIS reflectance into Landsat images, sensibly respecting the shape of ground-level instances. ESTARFM, being more conservative with regards to image sharpness, produces quite realistic looking samples. However, it struggles to recover the correct spectral values when images used as inputs are too temporally distant from the prediction date.

Furthermore, we observe that the SSIM loss term improves the stability of the adversarial training process and prevents GANs-induced up-sampling artifacts such as checkerboard patterns [21]. It encourages the discriminator to instill realistic contrast levels and image structures (e.g., fields boundaries and forested/urban areas), which have finer details. However, we notice a subsequent forcing in luminance that pushes generated crops toward lighter tones, while an L_1-only supervision renders more faithful colors.

5 Conclusion

Using cGANs, we can develop implicit generative models capable of producing visually promising results at surface reflectance fusion. We demonstrate their capacity to faithfully capture the broad features from coarse reflectance images and fuse it into detailed images. This can help circumvent limited access to public imagery with sufficient spatial and temporal resolution to assist precision agriculture and humanitarian response.

6 Acknowledgments

References

[1] William L. Barnes, Thomas S. Pagano, and Vincent V. Salomonson. Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. *IEEE Transactions on Geoscience and Remote Sensing*, 1998.

[2] Paul D. Bates. Remote sensing and flood inundation modelling. *Hydrological Processes*, 2004.

[3] David Berthelot, Peyman Milanfar, and Ian Goodfellow. Creating high resolution images with a latent adversarial generator, 2020.

[4] Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, and Enrico Magli. DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images. *arXiv:1907.06490*, 2020.

[5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

[6] Michel Deudon, Alfredo Kalaitzis, Md Rifat Arefin, Israel Goytom, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. HighRes-net: Multi-frame super-resolution by recursive fusion. *arXiv:2002.06460*, 2019.

[7] Roberto Filgueiras, Everardo Mantovani, Elpídio Fernandes-Filho, Fernando Cunha, Daniel Althoff, and Santos H. Dias. Fusion of MODIS and Landsat-like images for daily high spatial resolution NDVI. *Remote Sensing*, 2020.

[8] Mario Fuentes Reyes, Stefan Auer, Nina Merkle, Corentin Henry, and Michael Schmitt. SAR-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits. *Remote Sensing*, 2019.

[9] Feng Gao, Thomas Hilker, Xiaolin Zhu, Martha Anderson, Jeffrey Masek, Peijuan Wang, and Yun Yang. Fusing Landsat and MODIS data for vegetation monitoring. *IEEE Geoscience and Remote Sensing Magazine*, 2015.

[10] Feng Gao, Jeffrey Masek, Matt Schwaller, and Forrest Hall. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 2006.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*. 2014.

[12] Claas Grohnfeldt, Michael Schmitt, and Xiaoxiang Zhu. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018.

[13] Thomas Hilker, Michael Wulder, Nicholas Coops, Nicole Seitz, Joanne White, Feng Gao, Jeffrey Masek, and Gordon Stenhouse. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, 2009.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

[15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[16] Edward J. Knight and Geir Kvaran. Landsat-8 operational land imager design. *Remote Sensing of Environment*, 2014.

[17] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, and Emmanuel Baltsavias. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.

[18] Yan Liu, Michael J. Hill, Xiaoyang Zhang, Zhuosen Wang, Andrew D. Richardson, Koen Hufkens, Gianluca Filippa, Dennis D. Baldocchi, Siyan Ma, Joseph Verfaillie, and Crystal B. Schaaf. Using data from Landsat, MODIS, VIIRS and PhenoCams to monitor the phenology of California oak/grass savanna and open grassland across spatial scales. *Agricultural and Forest Meteorology*, 2017.

[19] Jeffrey G. Masek, Eric F. Vermote, Nazmi E. Saleous, Robert Wolfe, Forrest G. Hall, Karl Fred Huemmrich, Feng Gao, Jonathan Kutler, and Teng-Kui Lim. A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, 2006.

[20] Edward Masuoka, Albert Fleig, Robert E. Wolfe, and Fred Patt. Key characteristics of MODIS data products. *IEEE Transactions on Geoscience and Remote Sensing*, 1998.

[21] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016.

[22] Lillian Kay Petersen. Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. *Remote Sensing*, 2018.

[23] G. Qingfeng and X. Peng. cuESTARFM. https://github.com/HPSCIL/cuESTARFM, 2018.

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[25] David P. Roy, Michael A. Wulder, Thomas R. Loveland, Curtis E. Woodcock, Richard G. Allen, Martha C. Anderson, Dennis L. Helder, James R. Irons, Daniel M. Johnson, Robert Kennedy, Ted A. Scambos, Crystal B. Schaaf, John R. Schott, Yongwei Sheng, Eric F. Vermote, Alan S. Belward, Robert Bindschadler, Warren B. Cohen, Feng Gao, James D. Hipple, Patrick Hostert, Justin Huntington, Christopher O. Justice, Ayse Kilic, Valeriy Kovalskyy, Zhongping Lee, Leo Lymburner, Jeffrey G. Masek, Joel McCorkel, Yanmin Shuai, Ricardo Trezza, James Vogelmann, Randolph H. Wynne, and Zhe Zhu. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 2014.

[26] Tim G. J. Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopačková, and Piotr Biliński. Multi3Net: segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[27] Joy Sanyal and Xi Xi Lu. Application of remote sensing in flood management with special reference to monsoon Asia: A review. *Natural Hazards*, 2004.

[28] David M. Szpakowski and Jennifer L. R. Jensen. A review of the applications of remote sensing in fire ecology. *Remote Sensing*, 2019.

[29] Tamme van der Wal, Bauke Abma, Antidio Viguria, Emmanuel Prévinaire, Pablo J. Zarco-Tejada, Patrick Serruys, Eric van Valkengoed, and Paul van der Voet. Fieldcopter: unmanned aerial systems for crop monitoring services. In *Precision agriculture '13*. 2013.

[30] Jessica Walker, Kirsten de Beurs, Randolph Wynne, and Feng Gao. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. *Remote Sensing of Environment*, 2011.

[31] Lei Wang, Xin Xu, Yue Yu, Rui Yang, Rong Gui, Zhaozhuo Xu, and Fangling Pu. SAR-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access*, 2019.

[32] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004.

[33] Dengfeng Xie, Jinshui Zhang, Xiufang Zhu, Yaozhong Pan, Hongli Liu, Zhoumiqi Yuan, and Ya Yun. An improved STARFM with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors*, 2016.

[34] Roberta H. Yuhas, Alexander F. H. Goetz, and Joe W. Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. *The Spectral Image Processing System (SIPS)*, 1993.

[35] Xiaolin Zhu, Jin Chen, Feng Gao, Xuehong Chen, and Jeffrey Masek. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 2010.

[36] Michael Zotov and Jevgenij Gamper. Conditional denoising of remote sensing imagery using cycle-consistent deep generative models. *arXiv preprint arXiv:1910.14567*, 2019.

A Experimental and Implementation details

The U-Net architecture used as generator in all experiments uses 5 down and upsampling levels, leading to $6{\times}6{\times}1024$ latent features. We use batch normalization and parameterized ReLU activations on encoding and decoding layers. Models details are provided in Table 2.

The PatchGAN model used as a discriminator, similarly downsamples input into a $15{\times}15{\times}1$ feature map on top of which a sigmoid activation is employed to classify the underlying region of each cell. Batch normalization and leaky ReLU activations are used on the downsampling layers. See Table 2 for architecture details.

The Adam optimizer [15] with safe initial learning rate of $3{\cdot}10^{-4}$ — decayed by 0.99 at each epoch — is used for training both generator and discriminator. To make up for the additional guidance provided to the generator by the supervision objectives, we backpropagate on the discriminator twice as often. A batch size of 64 is used during training. Finally, in order for the different components of the objective function to land in comparable ranges, we use supervision weights $\alpha = 0.1$ and $\beta = 100$.

U-Net
Encoder
Conv2D(in=8, out=64, kernel=4, stride=2, padding=1)
Conv2D(in=64, out=128, kernel=4, stride=2, padding=1) BatchNorm(channels=128) PReLU()
Conv2D(in=128, out=256, kernel=4, stride=2, padding=1) BatchNorm(channels=256) PReLU()
Conv2D(in=256, out=512, kernel=4, stride=2, padding=1) BatchNorm(channels=512) PReLU()
Conv2D(in=512, out=1024, kernel=4, stride=2, padding=1) BatchNorm(channels=1024) PReLU()
Conv2D(in=1024, out=1024, kernel=4, stride=1, padding=1) BatchNorm(channels=1024) PReLU()
Decoder
ConvTranspose2D(in=1024, out=1024, kernel=4, stride=1, padding=1) BatchNorm(channels=1024) Dropout(p=0.4) PReLU()
ConvTranspose2D(in=2048, out=512, kernel=4, stride=2, padding=1) BatchNorm(channels=512) Dropout(p=0.4) PReLU()
ConvTranspose2D(in=1024, out=256, kernel=4, stride=2, padding=1) BatchNorm(channels=256) PReLU()
ConvTranspose2D(in=512, out=128, kernel=4, stride=2, padding=1) BatchNorm(channels=128) PReLU()
ConvTranspose2D(in=256, out=64, kernel=4, stride=2, padding=1) BatchNorm(channels=1024) PReLU()
ConvTranspose2D(in=128, out=64, kernel=4, stride=2, padding=1)
Output Layer
ConvTranspose2D(in=64, out=4, kernel=3, stride=1, padding=1)

PatchGAN
Conv2D(in=12, out=128, kernel=4, stride=2, padding=1) LeakyReLU(α=0.2)
Conv2D(in=128, out=256, kernel=4, stride=2, padding=1) BatchNorm(channels=256) LeakyReLU(α=0.2)
Conv2D(in=256, out=512, kernel=4, stride=2, padding=1) BatchNorm(channels=512) LeakyReLU(α=0.2)
Conv2D(in=512, out=512, kernel=4, stride=2, padding=1) BatchNorm(channels=512) LeakyReLU(α=0.2)
Conv2D(in=512, out=1, kernel=4, stride=1, padding=1) Sigmoid()

Table 2: U-Net Generator and PatchGAN Discriminator architectures
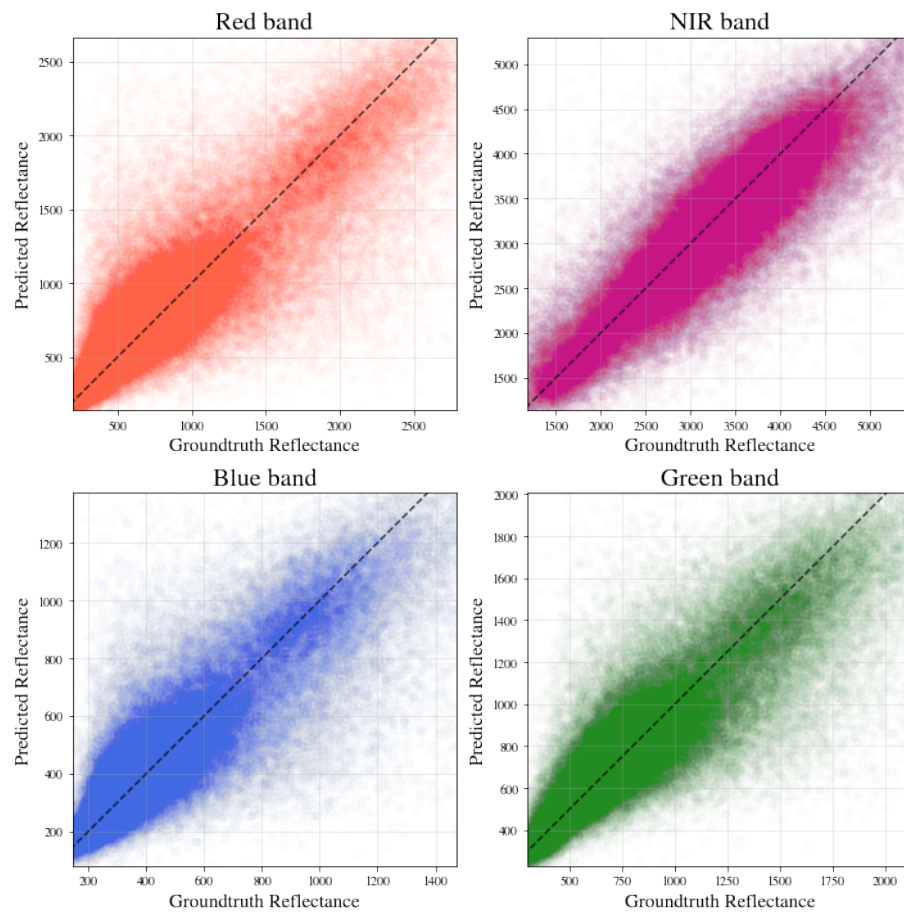
B Figures

Figure 2: Scatter plot of predicted pixel values by band against groundtruth Landsat pixel values on the testing set