Towards *DeepSentinel*

An extensible corpus of labelled Sentinel-1 and -2 imagery and a proposed general purpose sensor-fusion semantic embedding model.

Lucas Kruitwagen*
Smith School of Enterprise and the Environment, School of Geography and the Environment
University of Oxford
Oxford, UK

Abstract

Earth observation offers new insight into anthropogenic changes to nature, and how these changes are effecting (and are effected by) the built environment and the real economy. With the global availability of medium-resolution (10-30m) synthetic aperature radar (SAR) Sentinel-1 and multispectral Sentinel-2 imagery, machine learning can be employed to offer these insights at scale, unbiased to company- and country-level reporting. In this proposal, we document the development of an extensible corpus of labelled and unlabelled Sentinel-1 and Sentinel-2 imagery for the purposes of sensor fusion research. We make a large corpus and supporting code publicly available. We propose our own experiment design for the development of *DeepSentinel*, a general-purpose semantic embedding model. Our aspiration is to provide pretrained models for transfer learning applications, significantly accelerating the impact of machine learning-enhanced earth observation on climate change mitigation.

1 Earth observation for climate change mitigation

Satellite-based earth observation plays a central role in measuring climate change impacts and risks [5]. Medium-resolution satellites (10-30m spatial resolution), despite being initially designed for environmental monitoring, are being increasingly used in applications focusing on the interface between the environment and the real economy for the purposes of financial risk measurement: estimating carbon dioxide emissions[11], methane emissions[32], and localising large fixed-capital assets[18]. For the purposes of assessing financial risk due to climate change, these satellites provide the globally exhaustive and unbiased view required by financial decision makers. Deploying analysis at this global scale can only be accomplished with the use of machine learning.

These climate change risk applications are impaired by two perennial challenges with satellite-based earth observation: the presence of atmospheric interference, and a shortage of training labels. Atmospheric interference is not equally distributed around our planet. Excessive cloud cover makes surface retrievals using multispectral imagery very challenging in certain geographies, negating its otherwise ‘exhaustive‘ coverage. Many cloudy geographies are in the global south, precisely where populations are the most vulnerable to climate change and where conventional reported data is the most sparse. These same geographies are where financial institutions in the global north concentrate their risk exposure to generate outsize returns, and are where civil society groups must be most vigilant to identify and respond to neocolonial practises by these same institutions.

The shortage of training labels for machine learning with earth observation data is well documented. Land use and land cover data is available with moderate spatial and temporal resolution in Europe (e.g. the Copernicus Corine Land Cover[27]) and the US (e.g. the USDA Cropland Data Layer [28]),

*Contact: lucas.kruitwagen@smithschool.ox.ac.uk; Bio: https://lkruitwagen.github.io/

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

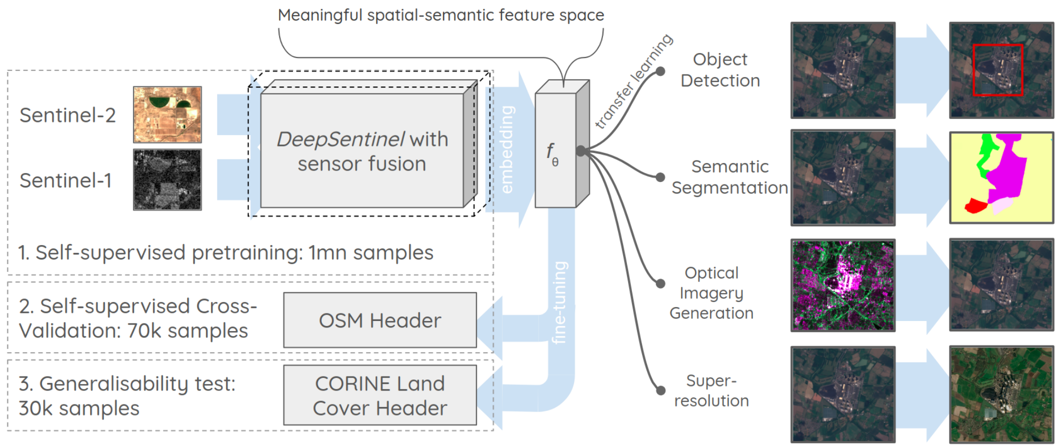

Figure 1: *DeepSentinel* summary, showing pretraining and fine-tuning curriculum and the variety of use case applications.

however data for the rest of the planet are sparse (e.g. OpenStreetMaps [23]). These datasets include only broad categories of land use and land cover, and are not fit for the purpose of localising specific categories of industrial infrastructure. Financial institutions just beginning to reckon with geospatial data, leaving a large gap in existence and availability of spatially-localised asset-level data.[4]

We propose a general-purpose sensor fusion semantic embedding model to overcome these dual challenges of atmospheric interference and label availability. *DeepSentinel* will use sensor fusion of SAR Sentinel-1 data and multispectral Sentinel-2 data to provide latent space embeddings of surface conditions even in excessively cloudy conditions. Self-supervised pretraining followed by fine tuning on land-use and land-cover data will create a general purpose embedding model suitable for a wide range of downstream applications using transfer learning, see Figure 1. This proposal describes the current state-of-the-art in sensor fusion, our progress towards a data corpus for training purposes, and our proposed experiment design for delivering *DeepSentinel*.

2 Sensor fusion

SAR imagery carries complementary information to multispectral imagery and so is useful beyond its penetration of atmospheric conditions. Sentinel-1 SAR C-band backscatter is sensitive to moisture and surface types, and so has found applications in both natural environments (e.g. with classification of forests[26] and crop types[22, 29, 21]), and the built human environment (e.g. road classification [35]). These properties are complimentary to the multispectral data provided by Sentinel-2, which has sensors designed to detect aerosols, water vapour, chlorophyll, and visual spectra.

Sensor fusion describes the complementary combination of data from multiple sources to improve inference quality beyond what would be possible from either sensor individually. Fusion of Sentinel-1 and Sentinel-2 imagery has been studied for cloud removal[7, 12, 2, 19], synthetic imagery generation [14, 3, 9], and land cover classification [31, 8, 30].

With the exception of [19], most of these studies obtain a limited corpus for a specific area of interest. With *DeepSentinel*, we propose a general-purpose encoder than can be used for any land-surface area-of-interest on the planet, and fine-tuned for any of the applications above. Relative to [19], we propose a training dataset more than 6x larger, and made ‘production-ready’, i.e. using best-available atmosphere-corrected data, see Table 1. Our goal is to unlock increasingly niche applications where limited training data is available, allowing a proliferation of earth observation use cases with impact analogous to the release of pre-trained conventional imagery convolutional neural networks of the mid-2010s (e.g. ResNet, VGG-16/19, etc).

3 DeepSentinel - Data preparation

We prepare two novel datasets for the development of *DeepSentinel*. The first dataset is prepared without labels for the purpose of self-supervised pretraining. We obtain random sample patches

Table 1: Sensor Fusion Datasets

Dataset (Study)	Components*	Pixels	Resolution	Geography
DeepSentinel labelled (ours, proposed)	S1 (VV+VH) + S2 (L2A - all bands) + CCLC + OSM	6,554 Mpx	10m	EU27+GB
SEN12MS-CR (Meraner et al. 2020)	S1 (VV+VH) + S2 (L1C - all bands)	10,323 Mpx	10m	Global
DeepSentinel unlabelled (ours, proposed)	S1 (VV+VH) + S2 (L2A - all bands)	65,536 Mpx	10m	Global

* S1: Sentinel-1; S2: Sentinel-2; CCLC: Copernicus CORINE Land Cover; OSM: OpenStreetMaps; L1C: Level 1-C; L2A: Level 2-A

from the planet's land surface area. The square patches are sampled at a pixel resolution of 10m with 250-pixel side length. Patches are obtained for both Sentinel-1 and -2, only where the image acquisition dates are within 3 days of each other. For Sentinel-2, we sample all 13 multispectral bands. For Sentinel-1, we sample VV and VH polarisation bands of interferometric wide swath (IW) retrievals, the retrieval mode used over land. For maximum reproduceability and impact, we obtain samples for all patches using both Google Earth Engine[10] and Descartes Labs computation platform[1]. Sentinel-2 acquisitions from Google Earth Engine are provided at 'Level 2A' surface reflectance level, from Descartes Labs they are obtained at 'surface-level' using Descartes Labs proprietary atmosphere correction algorithm. Sentinel-1 acquisitions from Google Earth Engine have been thermal noise corrected, radiometrically calibrated, and terrain corrected. Sentinel-1 acquisitions from Descartes Labs have been similarly terrain corrected.

The second dataset is prepared with labels for the purposes of fine-tuning and cross-validation. Land use and land cover labels are obtained from the 2018 Copernicus CORINE Land Cover inventory, rasterised at 10m, as well as OSM Point, Line, and Polygon datasets. For the second dataset, samples are only obtained for European Union (plus the United Kingdom) where Copernicus CORINE Land Cover data are available and OSM data is of substantially higher quality. The two datasets are made publicly available via Google Cloud Storage and Microsoft Azure Storage. Two demonstration datasets have been prepared for the NeurIPS Climate Change AI workshop: a 10,000 sample dataset without labels, and a 1,000 sample dataset with CORINE and OSM land use and land cover labels.[2] The code for sampling the earth observation, land cover, and OSM data is available via Github. [3] We propose to scale up our two novel datasets to 1,000,000 and 100,000 samples respectively, which would make them the two of the largest datasets of their kind, see Table 1.

4 DeepSentinel - Proposed experimentation

With DeepSentinel, we seek to produce semantically-meaningful feature embeddings from either or both Sentinel-1 and -2 imagery. In this proposal, we seek feedback on our experiment design from the NeurIPS 2020 Climate Change AI workshop. Our success with *DeepSentinel* will depend on two main design decisions: our choice of embedding architecture, and our training curriculum.

For our embedding architecture, we will seek an optimal architecture learning from the experiences of the computer vision research community. Embedding architectures for experimentation include ResNet[13], UNet[25], and DenseNet[15] variants, HRNet[34], and context-based attention encoders [33]. For our training curriculum, we propose to use our large unlabelled data corpus in a self-supervised learning implementation. Many self-supervised learning algorithms are available that may be appropriate for *DeepSentinel*. We will experiment with a range of methods: contrastive learning[6], anchor-neighbour-distant triplet loss[16], auto-encoders [17], and adversarial network variants (e.g. cGAN[20], ALEA[24]). The challenge with self-supervision methods is the development of an evaluation criteria suitable for our target use case. This is why we have created our second labelled dataset. Using 70% of this dataset, we will fine-tune the encoder using labelled data, and use fine-tuned performance to cross-validate and optimise the performance of our self-supervised pretraining method. We will experiment with patch multi-class classification, object detection, and semantic segmentation tasks on OpenStreetMap labels using conventional weighted log loss and intersection-over-union as

[2]Datasets are accessible at `https://console.cloud.google.com/storage/browser/deepsentinel`

[3]*DeepSentinel* code is available at `https://github.com/Lkruitwagen/deepsentinel.git` and the OSM server at `https://github.com/Lkruitwagen/deepsentinel-osm.git`

evaluation criteria. Finally, with the remaining 30% of the data, we will test the generalisability of the feature embeddings on new land use and land cover classes previously unknown to the model.

Acknowledgments and Disclosure of Funding

I am immensely grateful for the ongoing support of Descartes Labs who have provided platform access for this project as part of their Impact Science Program. This work is also being prepared with the generous support of Microsoft AI for Earth and Google Cloud Platform research credits. Without the support of these organisations this work would not be possible; they have my immense gratitude.

References

[1] C. M. Beneke, S. Skillman, M. S. Warren, T. Kelton, S. P. Brumby, R. Chartrand, and M. Mathis. A Platform for Scalable Satellite and Geospatial Data Analysis. In *AGU Fall Meeting Abstracts*, volume 2017, pages IN32C–04, Dec. 2017.

[2] J. Bermudez, P. Happ, D. Oliveira, and R. Feitosa. Sar to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4(1), 2018.

[3] J. D. Bermudez, P. N. Happ, R. Q. Feitosa, and D. A. Oliveira. Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1220–1224, 2019.

[4] B. Burks. Space, The Next Frontier: Spatial Finance And Environmental Sustainability. `https://www.spglobal.com/ratings/en/research/articles/200122-space-the-next-frontier-spatial-finance-and-environmental-sustainability-11317146`.

[5] B. Caldecott, L. Kruitwagen, M. McCarten, X. Zhou, D. Lunsford, O. Marchand, P. Hadjikyriakou, V. Bickel, T. Sachs, and N. Bohn. Climate risk analysis from space: remote sensing, machine learning, and the future of measuring climate-related risk. 2018.

[6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

[7] R. Eckardt, C. Berger, C. Thiel, and C. Schmullius. Removal of optically thick clouds from multi-spectral satellite images using multi-frequency sar data. *Remote Sensing*, 5(6):2973–3006, 2013.

[8] S. Ferrant, A. Selles, M. Le Page, P.-A. Herrault, C. Pelletier, A. Al-Bitar, S. Mermoz, S. Gascoin, A. Bouvet, M. Saqalli, et al. Detection of irrigated crops from sentinel-1 and sentinel-2 data to estimate seasonal groundwater use in south india. *Remote Sensing*, 9(11):1119, 2017.

[9] M. Fuentes Reyes, S. Auer, N. Merkle, C. Henry, and M. Schmitt. Sar-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits. *Remote Sensing*, 11(17):2067, 2019.

[10] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017.

[11] M. Gray, L. Watson, S. Ljungwaldh, and E. Morris. Nowhere to hide: Using satellite imagery to estimate the utilisation of fossil fuel power plants. 2018.

[12] C. Grohnfeldt, M. Schmitt, and X. 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*, pages 1726–1729. IEEE, 2018.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[14] W. He and N. Yokoya. Multi-temporal sentinel-1 and-2 data fusion for optical image simulation. *ISPRS International Journal of Geo-Information*, 7(10):389, 2018.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks, 2018.

[16] N. Jean, S. Wang, A. Samar, G. Azzari, D. Lobell, and S. Ermon. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3967–3974, 2019.

[17] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2014.

[18] L. Kruitwagen, K. Stoy, J. Friedrich, S. Skillman, and C. Hepburn. A global census of solar facilities using deep learning and remote sensing. In *NeurIPS Climate Change AI Workshop 2019*. NeurIPS Climate Change AI, 2019.

[19] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333–346, 2020.

[20] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[21] E. Ndikumana, D. Ho Tong Minh, N. Baghdadi, D. Courault, and L. Hossard. Deep recurrent neural network for agricultural classification using multitemporal sar sentinel-1 for camargue, france. *Remote Sensing*, 10(8):1217, 2018.

[22] A. Nelson, T. Setiyono, A. B. Rala, E. D. Quicho, J. V. Raviz, P. J. Abonete, A. A. Maunahan, C. A. Garcia, H. Z. M. Bhatti, L. S. Villano, et al. Towards an operational sar-based rice monitoring system in asia: Examples from 13 demonstration sites across asia in the riice project. *Remote Sensing*, 6(11):10773–10812, 2014.

[23] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org`, 2020.

[24] S. Pidhorskyi, D. Adjeroh, and G. Doretto. Adversarial latent autoencoders, 2020.

[25] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[26] M. Rüetschi, M. E. Schaepman, and D. Small. Using multitemporal sentinel-1 c-band backscatter to monitor phenology and classify deciduous and coniferous forests in northern switzerland. *Remote Sensing*, 10(1):55, 2018.

[27] C. L. M. Service. Corine land cover, 2020.

[28] N. A. S. Service. Cropland data layer, 2020.

[29] T. Setiyono, F. Holecz, N. Khan, M. Barbieri, E. Quicho, F. Collivignarelli, A. Maunahan, L. Gatti, and G. Romuga. Synthetic aperture radar (sar)-based paddy rice monitoring system: Development and application in key rice producing areas in tropical asia. In *IOP Conference Series: Earth and Environmental Science*, volume 54, page 012015. IOP Publishing, 2017.

[30] N. Torbick, X. Huang, B. Ziniti, D. Johnson, J. Masek, and M. Reba. Fusion of moderate resolution earth observations for operational crop type mapping. *Remote Sensing*, 10(7):1058, 2018.

[31] K. Van Tricht, A. Gobin, S. Gilliams, and I. Piccard. Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: a case study for belgium. *Remote Sensing*, 10(10):1642, 2018.

[32] D. J. Varon, D. J. Jacob, D. Jervis, and J. McKeever. Quantifying time-averaged methane emissions from individual coal mine vents with ghgsat-d satellite observations. *Environmental Science & Technology*, 54(16):10246–10253, 2020.

[33] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual attention network for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2017.

[34] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao. Deep high-resolution representation learning for visual recognition, 2020.

[35] Q. Zhang, Q. Kong, C. Zhang, S. You, H. Wei, R. Sun, and L. Li. A new road extraction method using sentinel-1 sar images based on the deep fully convolutional neural network. *European Journal of Remote Sensing*, 52(1):572–582, 2019.