SpaceNet MVOI: a Multi-View Overhead Imagery Dataset Supplementary Material

Nicholas Weir[1], David Lindenbaum[2], Alexei Bastidas[3], Adam Van Etten[1], Sean McPherson[3], Jacob Shermeyer[1], Varun Kumar[3], and Hanlin Tang[3]

[1]In-Q-Tel CosmiQ Works, [nweir, avanetten, jshermeyer]@iqt.org
[2]Accenture Federal Services, david.lindenbaum@accenturefederal.com
[3]Intel AI Lab, [alexei.a.bastidas, sean.mcpherson, varun.v.kumar, hanlin.tang]@intel.com

A. Dataset

A.1. Imagery details

The images from our dataset were obtained from DigitalGlobe, with 27 different viewing angles collected over the same geographical region of Atlanta, GA. Each viewing angle is characterized as both an off-nadir angle and a target azimuth. We binned each angle into one of three categories (Nadir, Off-Nadir, and Very Off-Nadir) based on the angle (see Table 2). Collects were also separated into South- or North-facing based on the target azimuth angle.

The imagery dataset comprises Panchromatic, Multi-Spectral, and Pan-Sharpened Red-Green-Blue-near IR (RGB-NIR) images The ground resolution of image varied depending on the viewing angle and the type of image (Panchromatic, Multi-spectral, Pan-sharpened). See Table 1 for more details. All experiments in this study were performed using the Pan-Sharpened RGB-NIR image (with the NIR band removed, except for the U-Net model).

The imagery was uploaded into the spacenet-dataset AWS S3 bucket, which is publicly readable with no cost to download. Download instructions can be found at www.spacenet.ai/off-nadir-building-detection/.

Image	Resolution at 7.8°	Resolution at 54°
Panchromatic	0.46m/px	1.67m/px
Multi-spectral	1.8m/px	7.0m/px
Pan-sharpened	0.46m/px	1.67m/px

Table 1: Resolution across different image types for two nadir angles.

A.2. Dataset breakdown

The imagery described above was split into three folds: 50% in a training set, 25% in a validation set, and 25% in a final test set. 900×900-pixel geographic tiles were randomly placed in one of the three categories, with all of the look angles for a given geography assigned to the same subset to avoid geographic leakage. The full training set and building footprint labels as well as the validation set imagery were open sourced, and the validation set labels and final test imagery and labels were withheld as scoring sets for public coding challenges.

B. Model Training

B.1. TernausNet

The TernausNet model was trained without pre-trained weights roughly as described previously [5], with modifications. Firstly, only the Pan-sharpened RGB channels were used for training, and were re-scaled to 8-bit. 90° rotations, X and Y flips, imagery zooming of up to 25%, and linear brightness adjustments of up to 50% were applied randomly to training images. After augmentations, a 512×512 crop was randomly selected from within each 900×900 training chip, with one crop used per chip per training epoch. Secondly, as described in the Models section of the main text, a combination loss function was used with a weight parameter $\alpha = 0.8$. Secondly, a variant of Adam incorporating Nesterov momentum [1] with default parameters was used as the optimizer. The model was trained for 25-40 epochs, and learning rate was decreased 5-fold when validation loss failed to improve for 5 epochs. Model training was halted when validation loss failed to improve for 10 epochs.

Catalog ID	Pan-sharpened Resolution	Look Angle	Target Azimuth Angle	Angle Bin	Look Direction
1030010003D22F00	0.48	7.8	118.4	Nadir	South
10300100023BC100	0.49	8.3	78.4	Nadir	North
1030010003993E00	0.49	10.5	148.6	Nadir	South
1030010003CAF100	0.48	10.6	57.6	Nadir	North
1030010002B7D800	0.49	13.9	162	Nadir	South
10300100039AB000	0.49	14.8	43	Nadir	North
1030010002649200	0.52	16.9	168.7	Nadir	South
1030010003C92000	0.52	19.3	35.1	Nadir	North
1030010003127500	0.54	21.3	174.7	Nadir	South
103001000352C200	0.54	23.5	30.7	Nadir	North
103001000307D800	0.57	25.4	178.4	Nadir	South
1030010003472200	0.58	27.4	27.7	Off-Nadir	North
1030010003315300	0.61	29.1	181	Off-Nadir	South
10300100036D5200	0.62	31	25.5	Off-Nadir	North
103001000392F600	0.65	32.5	182.8	Off-Nadir	South
1030010003697400	0.68	34	23.8	Off-Nadir	North
1030010003895500	0.74	37	22.6	Off-Nadir	North
1030010003832800	0.8	39.6	21.5	Off-Nadir	North
10300100035D1B00	0.87	42	20.7	Very Off-Nadir	North
1030010003CCD700	0.95	44.2	20	Very Off-Nadir	North
1030010003713C00	1.03	46.1	19.5	Very Off-Nadir	North
10300100033C5200	1.13	47.8	19	Very Off-Nadir	North
1030010003492700	1.23	49.3	18.5	Very Off-Nadir	North
10300100039E6200	1.36	50.9	18	Very Off-Nadir	North
1030010003BDDC00	1.48	52.2	17.7	Very Off-Nadir	North
1030010003193D00	1.63	53.4	17.4	Very Off-Nadir	North
1030010003CD4300	1.67	54	17.4	Very Off-Nadir	North

Table 2: DigitalGlobe Catalog IDs and the resolution of each image based upon off-nadir angle and target azimuth angle.

B.2. U-Net

The original U-Net [7] architecture was trained for 30 epochs with Pan-Sharpened RGB+NIR 16-bit imagery, on a binary segmentation mask with a combination loss as described in the main text with $\alpha = 0.5$. Dropout and batch normalization were used at each layer, with dropout with $p = 0.33$. The same augmentation pipeline was used as with TernausNet. An Adam Optimizer [6] was used with learning rate of 0.0001 was used for training.

Type	NADIR	OFF - NADIR	VOFF - NADIR
Industrial	0.51	-0.13	-0.28
Sparse Res	0.57	-0.19	-0.37
Dense Res	0.66	-0.21	-0.41
Urban	0.64	-0.13	-0.30

Table 3: F_1 score for the model trained on all angles and evaluated evaluated on the nadir bins (NADIR), then the relative decrease in F_1 for the off-nadir and very off-nadir bins.

B.3. YOLT

The You Only Look Twice (YOLT) model was trained as described previously [2]. Bounding box training targets were generated by converting polygon building footprints into the minimal un-oriented bounding box that enclosed each polygon.

B.4. Mask R-CNN

The Mask R-CNN model with the ResNet50-C4 backbone was trained as described previously [4] using the same augmentation pipeline as TernausNet. Bounding boxes were created as described above for YOLT.

C. Geography-specific performance

C.1. Distinct geographies within SpaceNet MVOI

We asked how well the TernausNet model trained on SpaceNet MVOI performed both within and outside of the dataset. First, we broke down the test dataset into the four bins represented in main text Figure 1: Industrial, Sparse Residential, Dense Residential, and Urban, and scored models within those bins (Table 3). We observed slightly worse performance in Industrials areas than elsewhere at nadir, but markedly stronger drops in performance in residential areas as look angle increased.

C.2. Generalization to unseen geographies

We also explored how models trained on SpaceNet MVOI performed on building footprint extraction from imagery from other geographies, in this case, the Las Vegas imagery from SpaceNet [3]. After normalizing the Las Vegas (LV) imagery for consistent pixel intensities and channel order with SpaceNet MVOI, we predicted building footprints in LV imagery and scored prediction quality as described in Metrics. We also re-trained TernausNet on the LV imagery and examined building footprint extraction quality on the SpaceNet MVOI test set. Strikingly, neither model was able to identify building footprints in the unseen geographies, highlighting that adding novel looks angles does not necessarily enable generalization to new geographic areas.

		Test Set	
		MVOI 7°	SN LV
Training Set	MVOI ALL	0.68	0.01
	SN LV	0.00	0.62

Table 4: **Cross-dataset F_1.** Models trained on MVOI or SpaceNet Las Vegas [3] were inferenced on held out imagery from one of those two geographies, and building footprint quality was assessed as described in Metrics.

References

[1] Timothy Dozat. Incorporating nesterov momentum into adam. In *The 2016 International Conference on Learning Representations Workshops*, 2016.

[2] Adam Van Etten. You only look twice: Rapid multi-scale object detection in satellite imagery. *CoRR*, abs/1805.09512, 2018.

[3] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. SpaceNet: A Remote Sensing Dataset and Challenge Series. *CoRR*, abs/1807.01232, 2018.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *The 2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[5] Vladimir Iglovikov and Alexey Shvets. Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation. *CoRR*, abs/1801.05746, 2018.

[6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net - Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, 9351(Chapter 28):234–241, 2015.