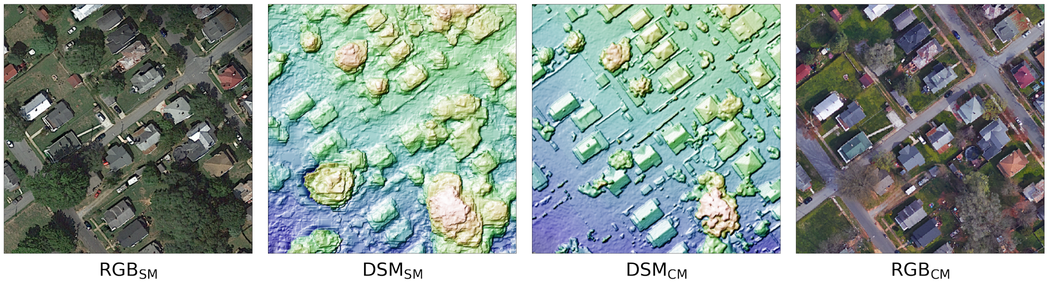
Estimating Residential Solar Potential using Aerial Data

Ross Goroshin, Alex Wilson, Andrew Lamb, Betty Peng, Brandon Ewonus, Cornelius Ratsch, Jordan Raisher, Marisa Leung, Max Burq, Thomas Colthurst, William Rucklidge, Carl Elkin
Google Inc
{goroshin, alexwilson, andrewlamb, bettypeng, bewonus, cratsch
jraisher, marisaleung, mburq, thomaswc, wjr, celkin}@google.com

Abstract

Project Sunroof estimates the solar potential of residential buildings using high quality aerial data. That is, it estimates the potential solar energy (and associated financial savings) that can be captured by buildings if solar panels were to be installed on their roofs. Unfortunately its coverage is limited by the lack of high resolution digital surface map (DSM) data. We present a deep learning approach that bridges this gap by enhancing widely available low-resolution data, thereby dramatically increasing the coverage of Sunroof. We also present some ongoing efforts to potentially improve accuracy even further by replacing certain algorithmic components of Sunroof's processing pipeline with deep learning.

1 Introduction


$\mathrm{RGB}_{\mathrm{SM}}$

$\mathrm{DSM}_{\mathrm{SM}}$

$\mathrm{DSM}_{\mathrm{CM}}$

$\mathrm{RGB}_{\mathrm{CM}}$

Figure 1: Low quality, sub-meter, imagery and its corresponding digital surface map are shown on the left ($\mathrm{RGB}_{\mathrm{SM}}$ and $\mathrm{DSM}_{\mathrm{SM}}$). High quality, centimeter-scale, of the same area collected at a different time, are shown on the right side ($\mathrm{RGB}_{\mathrm{CM}}$ and $\mathrm{DSM}_{\mathrm{CM}}$). The DSMs are rendered in 3D using the “hillshade” technique to better visualize geometric details.

Sunroof enables potential customers and policymakers to make informed decisions regarding solar energy by providing accurate solar energy estimates for individual buildings. Since its release, we estimate that Sunroof has been used in over 1 million residential solar projects. Installers report that the use of Sunroof substantially increases the likelihood of solar adoption by their customers.

In order to accurately estimate a building's solar energy potential it is necessary to first create a detailed 3D model of it and its surrounding area. Even tiny features like chimneys, air vents or AC units can impact the viable install area for solar panels. Of particular importance is the precise geometry of the roof, as the expected angle of incidence of sunlight has a major impact on annual generated energy. Sunroof mainly uses a digital surface model (DSM) and a digital terrain model (DTM) to simulate solar potential, which are currently obtained from aerial imagery (15). Both give elevations as functions of location on a grid, with the DTM giving the elevation of the terrain, and the DSM providing the elevation inclusive of all objects, such as trees and buildings. The DSM can be computed from overlapping images using standard stereo vision algorithms.

Sunroof relies on high quality, centimeter-level, aerial imagery in order to resolve details that are necessary for producing high accuracy solar estimates. However, even aerial imagery has varying degrees of quality. High flying aircraft equipped with a single camera can cover a large area at lower resolution, while lower flying aircraft equipped with calibrated multi-camera rigs are able to capture and register images at very high resolution. Other factors such as image registration quality and number of cameras used for stereoscopy, determine the signal to noise ratio of elevation data and therefore influence the effective resolution of the data. Figure 1 shows example DSMs computed from low and high quality image inputs, which we will refer to as sub-meter (DSM_{SM}) and centimeter (DSM_{CM}) scale, respectively. Sunroof relies on DSM_{CM} data to compute solar potential estimates. Unfortunately, high quality aerial imagery is much more limited in its coverage and update frequency. In Section 3 we will discuss how this limitation was overcome and enabled Sunroof to use widely available lower quality data, thus expanding its coverage, and thereby potentially increase the rate of solar adoption in new areas.

2 The Sunroof Algorithm

Sunroof estimates the solar power potential for the buildings in a given area by performing five major processing steps outlined in Algorithm 1. Steps (1-4) are involved in computing the viable solar potential of individual buildings. *The resulting energy predicted by steps (1-3) was physically validated by the National Renewable Energy Laboratory (NREL) (8). Furthermore, the entire pipeline has been validated by several major solar install companies which have queried millions of addresses via our API.* Details of each step are described in Appendix A.

Algorithm 1 Sunroof Algorithm

```
$Inputs:$ RGB, DSM, building footprints and probabilities
(1) footprints $\leftarrow$ SegmentBuildings(RGB, DSM, footprints, probabilities)
for each building $\in$ footprints do
    (2) Segment roof DSM into planes and remove obstacles
    (3) Efficiently compute solar flux with fast ray-tracing
    (4) Compute panel layout & power produced by each panel
    (5) Financial calculations
end for
$Outputs:$ Potential solar power and cost savings
```

3 Enabling Sunroof to Process Low Quality Data

Our low quality data coverage is about $\sim 10\times$ the area of the high quality data coverage, and is updated much more frequently. Therefore we seek a way to apply the Sunroof algorithm to the low quality data. As outlined in the previous section, Sunroof mainly relies on the DSM to compute the solar potential. Unfortunately, Sunroof does not generate accurate solar estimates from low quality DSMs.

In order to overcome this limitation we train a deep network to enhance the low quality DSM, by fusing information from multiple input modalities. If the enhanced DSM is a sufficiently accurate estimate of the high resolution DSM, it can be input to the unmodified Sunroof algorithm to estimate the solar potential. After presenting this approach and its results, we discuss a potential future improvement to the Sunroof algorithm itself, which replaces the graph-cut based roof segmentation (Subsection A.2) with a segmentation derived from an additional output head of our model.

3.1 Architecture

The inputs to our enhancement model are: 1-a visible spectrum RGB image of the area, 2-a low quality DSM and DTM, 3-building probabilities (see Appendix A.1). The outputs of our multi-head model are: 1-an enhanced DSM, 2-enhanced footprints, and 3-planar roof segments. All inputs, outputs, and targets have the same shape and spatial extent (512×512 covering the same area) but inputs are generally lower quality (i.e. have lower underlying resolution and are more noisy than the

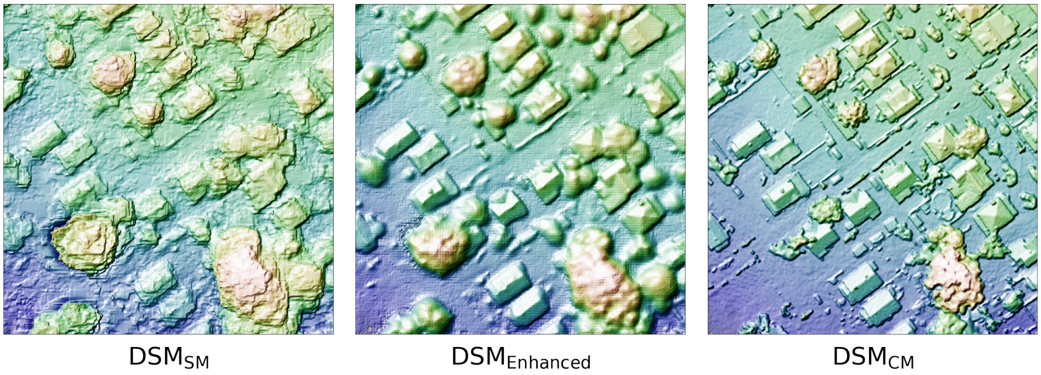


Figure 2: DSM enhancement by our model of the same area as Figure 1. DSM_{SM} is input to the model while the DSM_{CM} is used as ground truth.

corresponding high quality DSM). We use the corresponding high quality data (i.e. DSM, footprint, and roof segments) as the target for the corresponding output head. High and low quality data are often collected several years apart (“temporal skew”), therefore high quality targets should be regarded as imperfect/misaligned ground truth. For example, in Figure 1 the tree at lower left of the DSM_{SM} is missing in the DSM_{CM}. We use a UNet (9)-like architecture with a ResNet-50 encoder (5) backbone pretrained on ImageNet classification (10). Instead of using transposed-convolution, our architecture uses pixel-shuffle layers inspired by super-resolution architectures (13). The full architecture is depicted in Figure 7.

3.2 Enhanced DSM

Inferring absolute elevation (or similarly depth) from monocular images is an ill-posed problem (2). Therefore, we train our network to infer relative elevation using a low resolution digital terrain map (DTM) as a reference (6). This is achieved by inputting DSM_{SM} - DTM_{SM} to our model, and then adding the DTM_{SM} to the output before computing the error with the DSM_{CM} as a target. Our loss consists of per-pixel modified L_1 regression and surface normal terms, specifically:

$$L_1 = |h_o - h_t - \lambda * \sum_x \sum_y [h_o - h_t] \,|$$

$$L_{sn} = 1 - \frac{n_o \cdot n_t}{||n_o||\;||n_t||}$$

The second term in the L_1 loss partially discounts the average deviation (over all spatial locations) between the output and target elevation maps. We have found that setting $\lambda = \frac{1}{2}$ yields the best performance in metrics discussed in Section 4. The terms n_o and n_t are the normal vectors computed at each point of the output and target DSMs, respectively. To compute the outward pointing normal vector we use the following expression:

$$n = [-g_x, g_y, 1]$$

Where g_x and g_y denote the x and y components of the gradient of the DSM, respectively. These are computed using finite differences and local smoothing. Surface normal loss minimization has the effect of enhancing high frequency details of the output DSM, as well directly enforcing that the normal vectors of roof segments derived from the output DSM are accurate. Similar approaches have been used in (6) and (7).

3.3 Additional Outputs

Our network also outputs refined building probability and semantic roof segmentation maps. Building probabilities corresponding to low and high quality images are obtained automatically using another network specifically trained to perform building segmentation (similar to (14)). Our network is trained to enhance building probabilities corresponding to low quality data to match those corresponding to high quality data using a binary cross-entropy loss. Roof plane segmentation is achieved

Data	MAPE	MAPE@5kW	Temporal Separation
USA	$29.14\% \pm 10.50$	$4.62\% \pm 1.58$	$<$ 1 year
EU	$31.00\% \pm 9.36$	$6.43\% \pm 1.89$	$<$ 1 year
Temporal	$20.10\% \pm 12.71$	$4.43\% \pm 2.58$	1-5 years

Table 1: MAPE and MAPE@5kW errors in the EU and USA. The Temporal row shows the error due to temporal skew.

using a single shot affinity mask approach with graph partition post-processing to obtain instance segments similar to (3). Briefly, the N-dimensional output at each location predicts whether its N neighboring pixels belongs to the same instance label. Instance segments are obtained by applying a graph partitioning algorithm, to a graph whose edge weights are computed from the affinity maps output by the network (see Figure 6). Ground truth roof instance segments are obtained by applying Sunroof's graph-cut algorithm to the corresponding high quality DSMs described in Appendix A.2. Results presented in the next section were obtained by feeding enhanced DSMs and building probabilities to the Sunroof algorithm. Note that despite being trained to output the roof segmentation, the results presented in the next section used Sunroof's original segmentation approach described in Appendix A.2.

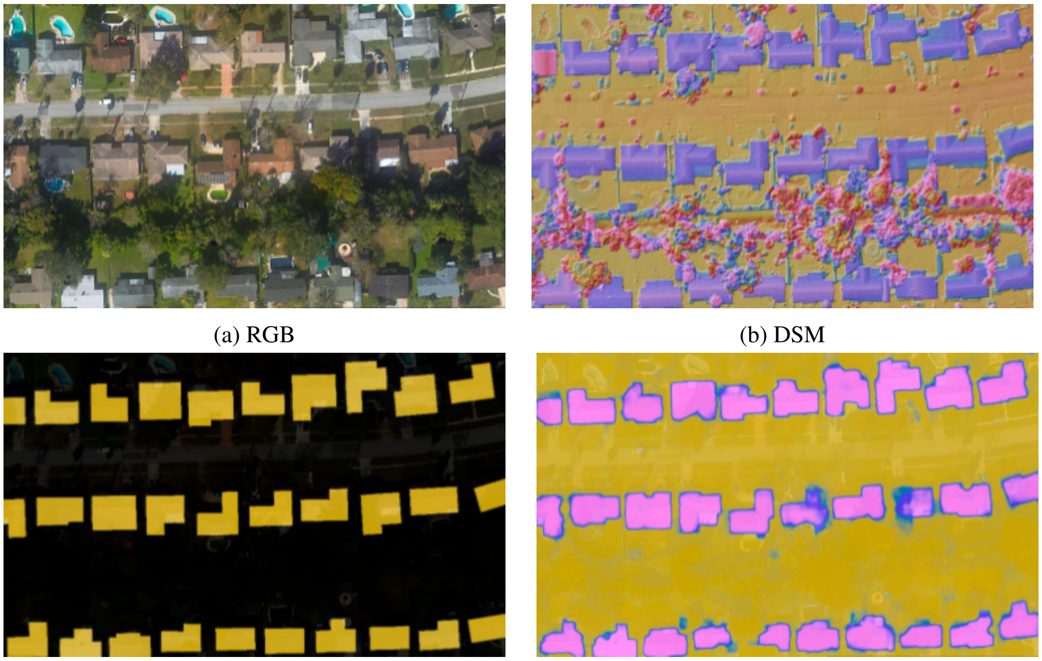
4 Performance and Future Work

We use two, physically grounded, performance measures to evaluate the performance of the refinement model (see Table 1). The first measures the percentage mean absolute power error (MAPE), that is the error between the total power, over the course of one year, predicted by Sunroof corresponding to the refined data and high quality data over the same area. Total power implies tiling the entire viable roof area in solar panels (see Figure 4d), something typical consumers rarely do. Therefore we introduce a second error measure, MAPE@5kW which measures the error corresponding to a much smaller, but more typical 5kW array ($\approx$ 10 panels). These panels are optimally positioned by Sunroof in the most productive areas of the roof predicted using the solar flux (Figure 4c). Thus the MAPE effectively measures the error for the entire roof, while the MAPE@5kW only measures the error corresponding to the most productive portions of the roof. Our enhancement model was trained on data collected over cities in the Southeastern USA, excluding any cities in Florida, which we reserved for validation. Our dataset is limited to cities where high/low quality pairs are available. The results presented in Table 1 are over test cities which are not present in the training or validation sets. We also report performance on Western European cities. To minimize temporal skew between inputs and ground truth (e.g. new construction, seasonal foliage, etc.), we selected high/low quality dataset pairs that were collected at most one year apart. Finally, in order to estimate the effect of temporal skew we evaluated the error between *high quality assets* collected between 1-5 years apart. This evaluation does not involve our enhancement model but simply compares Sunroof's predictions corresponding to temporally separated high quality datasets. This evaluation leads to a substantially lower MAPE value, which confirms that despite some temporal skew, the enhanced data still lags real high quality data in performance by about 10%. In ongoing work, we hope to achieve even better performance by replacing it with our model's output. We have found that the segmentation is extremely sensitive to the enhanced DSM and is thus a major source of MAPE error.

A Sunroof Details

A.1 Building Footprint Refinement

An initial set of footprints and probabilities are input to the Sunroof algorithm. Both of these are output by other, separately trained, models. Building footprints are rough polygons, but often separate individual residential addresses although they may appear connected. Building probabilities take on values close to 1.0 where the corresponding pixel is likely to belong to a building. A graph is created whose edge weights are computed by fusing information from the footprints/ probabilities, and DSM (height discontinuities captured in the DSM are often good cues to detect the presence



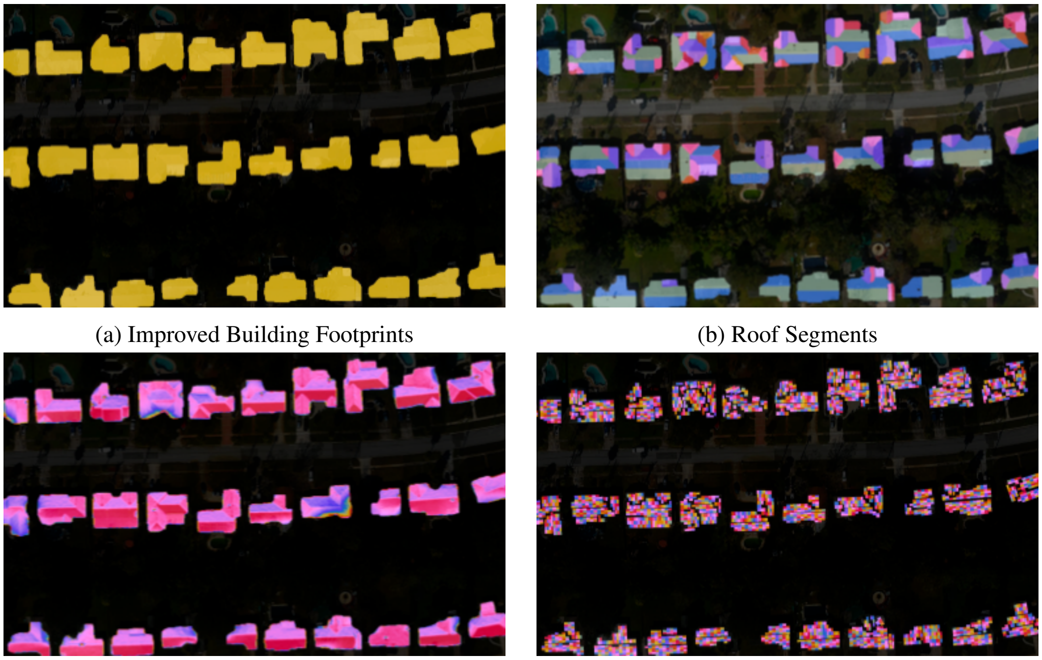
(a) RGB

(b) DSM

(c) Building Footprints

(d) Building Probabilities

Figure 3: Illustrations of the inputs to Sunroof.



(a) Improved Building Footprints

(b) Roof Segments

(c) Solar Flux

(d) Solar Panel Layout

Figure 4: Illustrations of the outputs of the processing steps of Sunroof.

of buildings). Footprint refinement is performed by running a graph-cut algorithm on this graph. The refined footprints often remove “tree overhangs” or buildings that are entirely occluded by trees, thanks to information from the building probabilities, while preserving separation between residences using information from the footprints.

A.2 Roof Segmentation and Obstacle Removal

Next, the roof pixels in the DSM are fitted to a small number of planes using a RANSAC (1) algorithm. This is essential as solar panels can only be laid out on flat roof segments. After the RANSAC procedure, the points assigned to each segment are refined using a graph cut (4) approach – if a point could reasonably be assigned to multiple planes, the graph cut will prefer to assign it to the same plane as its neighbors. Specifically, the graph cut algorithm attempts to minimize a cost function consisting of: (i) the projection distance from a point to its assigned plane, (ii) a second cost that minimizes the number of planes with similar normal vectors. This is intended to make it harder for two very similar planes to partition a flat area of the roof. The cost can be expressed as:

$$\sum_{p} \sum_{\mathbf{P}} \left[d(p, \mathbf{P}_p) + m \sum_{q \in \mathbf{N}(\mathrm{p})} 1 + max\left(0, \hat{n}_{\mathbf{P}_p} \cdot \hat{n}_{\mathbf{P}_q}\right) \right] \tag{1}$$

Where $d(p, \mathbf{P_p})$ denotes the projection distance when point p is assigned to plane $\mathbf{P}$, $\mathbf{N}(p)$ denote the set of points neighboring point p, and $\hat{n}_{\mathbf{P}_p}$ denotes the normal vector to plane $\mathbf{P}$. After running the graph cut, each roof plane is refit based on its new points, and the graph cut plus refitting procedure is repeated several times. Finally, the roof segments are filtered by size to remove tiny segments. This step also removes roof obstacles such as air-vents and air-conditioners.

A.3 Solar Flux

The solar flux calculation estimates the incident solar energy on a building over the course of a year (irradiance). Factors that affect solar flux include: the latitude, pitch angle of the roof segments, and surroundings which may occlude sunlight and cast shadows on the roof (usually trees or other buildings). The flux calculation is parallelized by partitioning the data into tiles. These tiles overlap, with each tile having a core plus margins. The margins are used so that nearby obstructions are taken into account when calculations are performed on the core. This means that the effects of distant occluders outside the margin area, such as distant mountain ranges, will not be factored into the calculation. The main flux calculation is performed using a method similar to (16), and its computational complexity is linear in the number of pixels, with a constant that depends on latitude, compared to $O(n^3)$, for direct ray tracing. The irradiance is summarized as two quantities: Direct Normal Irradiance (DNI), which is sunlight received directly from the sun, and Diffuse Horizontal Irradiance (DHI) which is indirect sunlight received by a flat plane. Both are measured in units of *Watts*$/m^2$. DNI and DHI are obtained from publicly available datasets, such as the National Solar Radiation Database (NSRDB) (12). Finally, both of these are further attenuated using a correction factor derived from the air temperature and wind speed using the model from (11). This reflects the decrease in silicon solar panel efficiency in elevated temperatures.

A.4 Optimal Panel Layout and Power Prediction

In order to get an upper bound on the solar potential of buildings, the solar panel placement algorithm tiles a roof with as many panels as the viable roof area can support. Viable roof areas include roof segments that are not overly steep and are free of obstacles. For sloped roof segments, solar panels are laid out in the 2D coordinates defined by the “eaves” and “up-slope” vectors (see Figure 5). If the unit normal is $\hat{n} = [n_x, n_y, n_z]$ then the eaves and up-slope vectors are defined as:

$$\hat{e} = \left[\frac{n_y}{\sqrt{n_x^2 + n_y^2}}, -\frac{n_x}{\sqrt{n_x^2 + n_y^2}}, 0 \right]$$
$$\hat{u} = \hat{e} \times \hat{n}$$

For a horizontal roof ($n_x = n_y = 0$), the eaves and up-slope vectors are chosen arbitrarily.

Panels are tiled over these flat roof segments in a way that maximize both energy production and compactness of the layout – roughly minimize the area of the rectangle that bounds all panels in a given roof segment. The flux allows simple estimation of the expected power generated by each panel, this facilitates finding optimal configurations for smaller solar installations. For example, Sunroof can be used to find the optimal panel layout of a, more typical, 5kW array consisting of 10-20 panels.

B Additional Figures

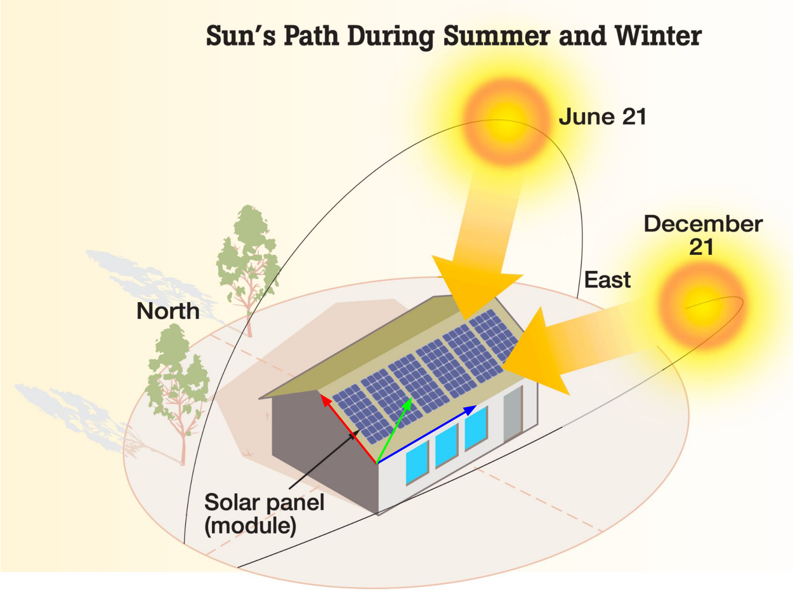
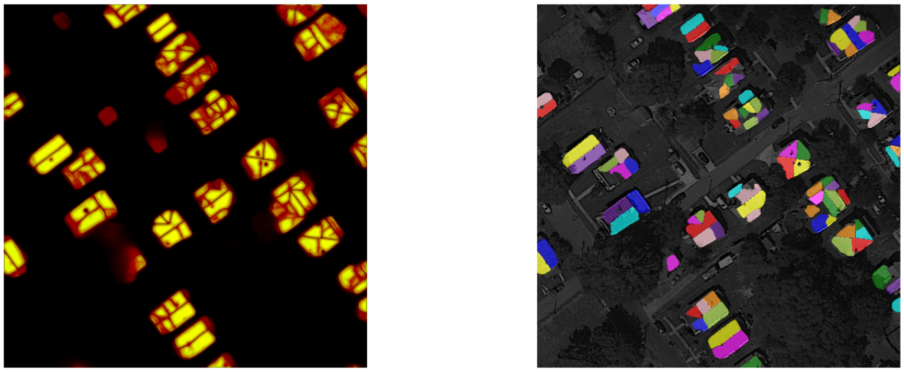


Figure 5: Depiction of irradiance and roof coordinate system used for panel layout. Red, green, and blue vectors are the up-slope, normal, and eaves vectors, respectively. Image credit: U.S. Department of Energy.



(a) One Channel from the Affinity Mask Output

(b) Instance Segments

Figure 6: Instance roof segments are obtained by post-processing the affinity mask output.

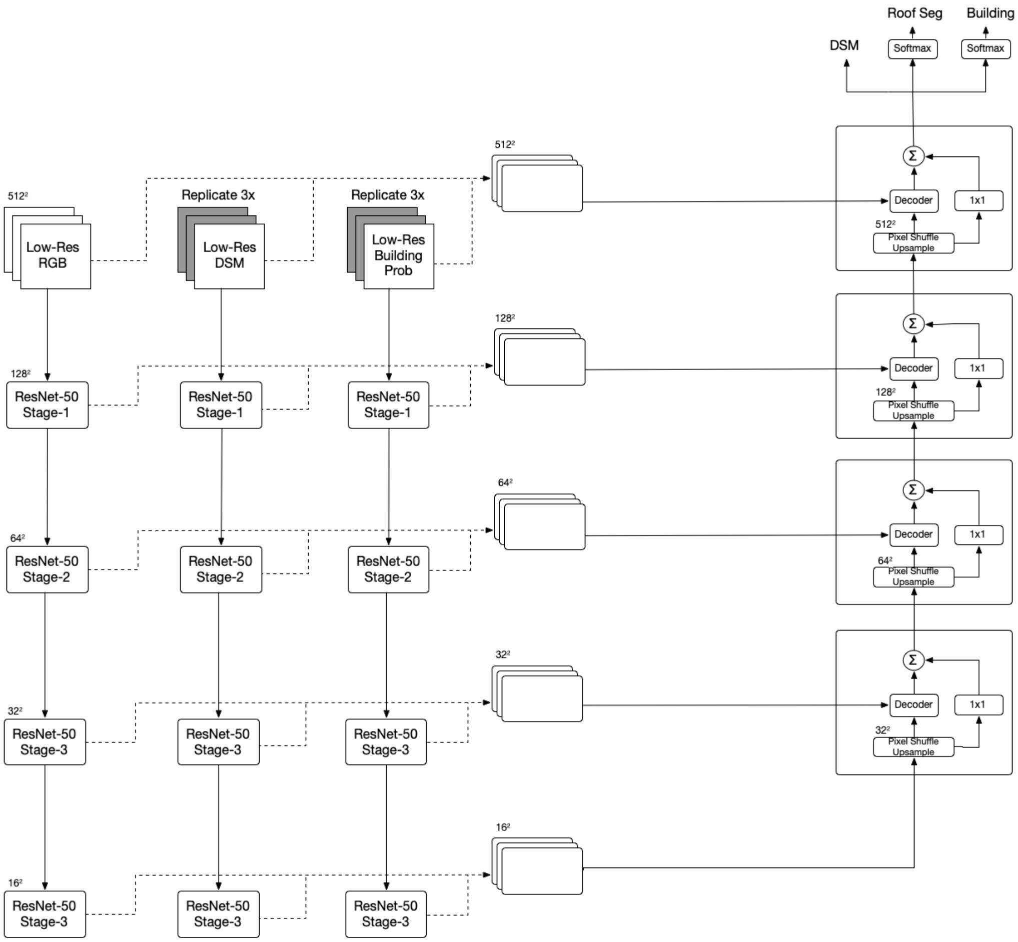


Figure 7: Sunroof UNet architecture

REFERENCES

[1] H Cantzler. Random sample consensus (ransac). *Institute for Perception, Action and Behaviour, Division of Informatics, University of Edinburgh*, 3, 1981.

[2] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014.

[3] Naiyu Gao, Yanhu Shan, Yupei Wang, Xin Zhao, Yinan Yu, Ming Yang, and Kaiqi Huang. Ssap: Single-shot instance segmentation with affinity pyramid. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 642–651, 2019.

[4] Dorothy M Greig, Bruce T Porteous, and Allan H Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):271–279, 1989.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] Lukas Liebel, Ksenia Bittner, and Marco Körner. A generalized multi-task learning approach to stereo dsm filtering in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:213–227, 2020.

[7] Chao-Jung Liu, Vladimir A Krylov, Paul Kane, Geraldine Kavanagh, and Rozenn Dahyot. Im2elevation: Building height estimation from single-view aerial imagery. *Remote Sensing*, 12(17):2719, 2020.

[8] NREL. Validating the accuracy of sighten's automated shading tool.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[10] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[11] Clemens Schwingshackl, Marcello Petitta, Jochen Ernst Wagner, Giorgio Belluardo, David Moser, Mariapina Castelli, Marc Zebisch, and Anke Tetzlaff. Wind effect on pv module temperature: Analysis of different techniques for an accurate estimation. *Energy Procedia*, 40:77–86, 2013.

[12] Manajit Sengupta, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. The national solar radiation data base (nsrdb). *Renewable and sustainable energy reviews*, 89:51–60, 2018.

[13] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[14] Wojciech Sirko, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John Quinn. Continental-scale building detection from high resolution satellite imagery. *arXiv preprint arXiv:2107.12283*, 2021.

[15] A. Sole and A. Valanzano. *Digital Terrain Modelling*, pages 175–194. Springer Netherlands, Dordrecht, 1996.

[16] Ville Timonen and Jan Westerholm. Scalable height field self-shadowing. In *Computer Graphics Forum*, volume 29, pages 723–731. Wiley Online Library, 2010.