ZRG: A Dataset for Multimodal 3D Residential Rooftop Understanding

Isaac Corley[1][2] Jonathan Lwowski[2] Peyman Najafirad[1]
[1]University of Texas at San Antonio [2]Zeitview
isaac.corley@my.utsa.edu, jonathan.lwowski@zeitview.com, peyman.najafirad@utsa.edu

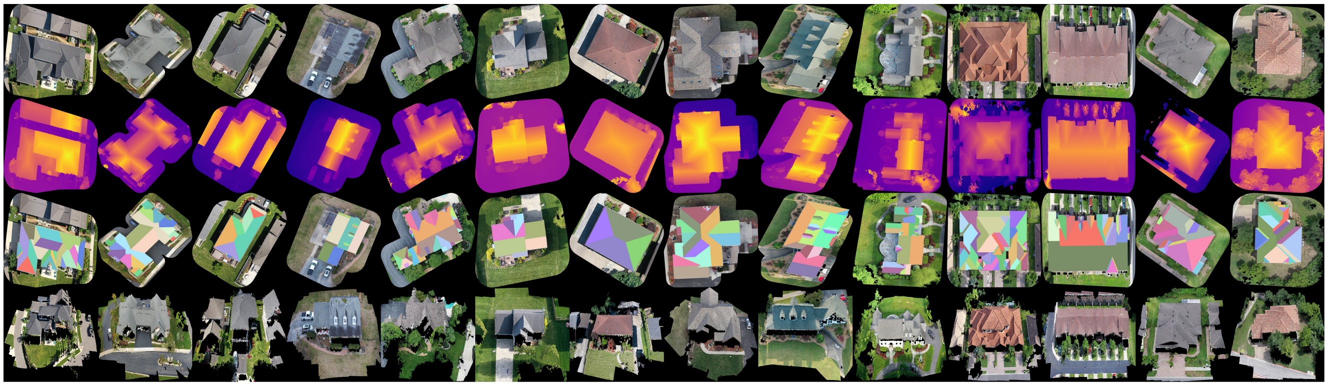


Figure 1. **Sample properties from our proposed ZRG dataset**. The dataset contains (top to bottom) high resolution RGB orthomosaics, digital surface models (DSM), 3D rooftop geometry wireframes, and 3D colored point clouds acquired from roof inspections of over 20k residential properties across the U.S.

Abstract

A crucial part of any home is the roof over our heads to protect us from the elements. In this paper we present the Zeitview Rooftop Geometry (ZRG) dataset for residential rooftop understanding. ZRG is a large-scale residential rooftop dataset of over 20k properties collected through roof inspections from across the U.S. and contains multiple modalities including high resolution aerial orthomosaics, digital surface models (DSM), colored point clouds, and 3D roof wireframe annotations. We provide an in-depth analysis and perform several experimental baselines including roof outline extraction, monocular height estimation, and planar roof structure extraction, to illustrate a few of the numerous potential applications unlocked by this dataset.[1]

1. Introduction

The roof is a vital component of a home and serves to protect owners from various environmental elements, such as rain, snow, and wind, while also contributing to the overall aesthetics and energy efficiency of the structure. A thorough understanding of a roof's condition is critical for homeowners as it allows them to make informed decisions about repair, replacement, or maintenance and the cost thereof. However, traditional methods of roof inspection are often time-consuming, labor-intensive, and subject to human error. Rapid developments in cost-effectiveness of unmanned aerial vehicles (UAV) have created safer and more efficient alternatives to manual roof inspections which capture high-resolution images of residential roofs from various angles and perspectives. As a result, there is an increasing need to augment and/or automate the roof analysis process by combining the latest advancements in machine learning and computer vision learning with large-scale residential rooftop datasets.

Given a sufficiently large residential rooftop dataset with multiple modalities, what are some of the potential applications which can be automated to further benefit society?

[1] Subsets of the dataset will be released for academic benchmarking purposes only in the future here https://github.com/isaaccorley/zrg

Dataset	Task	3D	Synthetic	Samples	Size (px)	Resolution (cm)	Modality
VWB [35]	R	×	×	2,001	256	30	RGB
Enschede [52]	R	×	×	2,000	512	8	RGB
RID [25]	S	×	×	3,648	512	10	RGB
MIPD [39]	R	✓	✓	2,539	-	-	RGB,Mesh
BuildingWF [30]	R	✓	✓	3,600	-	-	RGB,Mesh
ZRG (Ours)	S/R/H	✓	×	22,334	4,096+	<1	RGB,DSM,PC

Table 1. **Comparison of the ZRG dataset to related datasets** containing residential buildings and rooftop geometry related annotations. (S = segmentation, R = rooftop structure extraction, H = height estimation, PC = point cloud)

The primary objective of this research is to answer this question by introducing a novel large-scale rooftop dataset, which we name the Zeitview Rooftop Geometry (ZRG) dataset, for the analysis and understanding of residential rooftops.

In this work our **contributions** can be described as following:

- *Zeitview Rooftop Geometry (ZRG) Dataset* - We present a novel large-scale, high quality, high resolution, multimodal dataset for residential rooftop understanding and analysis. The dataset consists of high resolution RGB orthomosaics, digital surface models (DSM), 3D rooftop geometry wireframes, and 3D colored point clouds acquired from roof inspections of over 20k residential properties across the U.S. We perform an in-depth analysis to highlight the value of our dataset.

- *Baseline Experiments* - We provide baseline experiments for several common rooftop understanding tasks to display a few of the potential applications of our proposed dataset, including roof outline extraction, monocular height estimation, and planar roof structure extraction.

2. Background

While there is significant prior research into the mapping of building structure [19, 45], building height estimation [28, 34], and building change [7, 8, 10, 44] from remotely sensed imagery, the understanding of rooftop geometry and structure in particular is typically neglected from these works. Furthermore, there are few works and little focus specifically on residential buildings as opposed to commercial or industrial buildings. To further highlight the importance of automated residential rooftop understanding we detail the following notable applications:

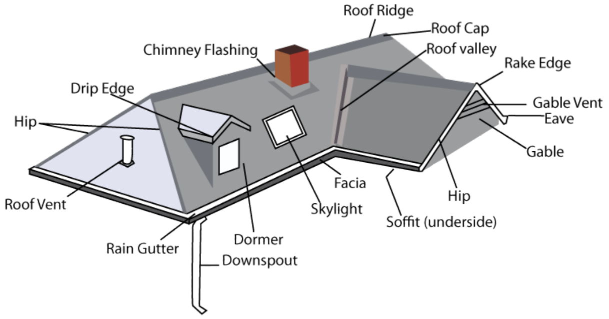


Figure 2. **A visual example of the anatomy of a residential rooftop** [41]. Each rooftop is annotated with 3D wireframe polygons for each roof face. Individual wireframe edges are also labeled with 18 edge categories.

Roof Damage Inspection and Detection Automatic identification and categorization of various types of roof damage from aerial imagery [16], such as missing or damaged shingles, cracks, or leaks, enables rapid and accurate damage assessment, significantly decreasing the cost of inspections and ultimately reducing the risk of further damage to the property.

Residential Solar Rooftop Potential Planar rooftop surface and structure extraction can assist in determining the feasibility of the installation of solar panels [2, 4, 20, 25, 46], by taking into account factors such as roof orientation, shading, and estimating the available surface area. Understanding the 3D geometry of roof structures allows for the optimal placement and configuration of solar panels to maximize energy production and reduce unnecessary costs to the homeowners.

3D Modeling and Digital Twins The automatic translation of man-made structures from aerial imagery to 3D models enables the ability to simulate a realistic virtual environment for various applications such as urban planning [22] and humanitarian assistance and disaster response (HADR) [16].

3. Related Work

Residential rooftop understanding datasets are scarce. The available datasets are typically generated from either

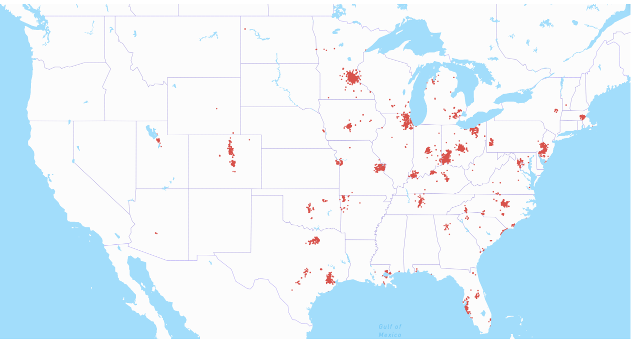
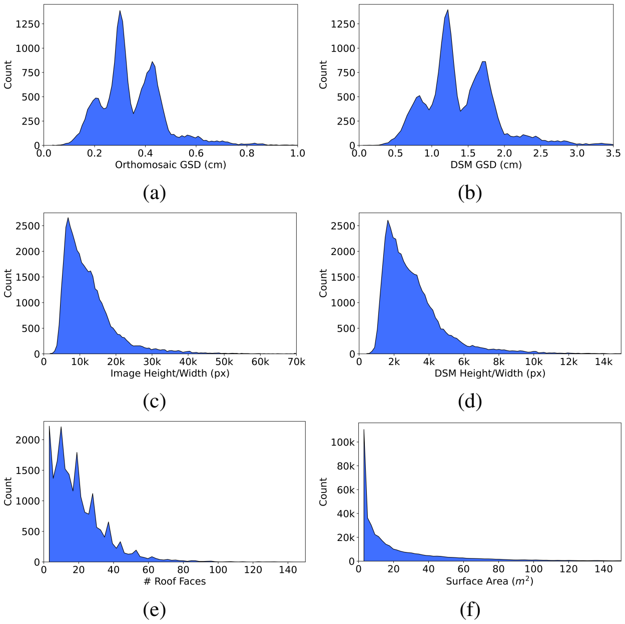


Figure 3. **Residential property locations** of the ZRG dataset. The dataset consists of a diverse set of samples from various population settings (urban vs. rural), property type (single-family (SFH) vs. multi-family homes (MFH)), and regions across the U.S. (primarily northeast, south, midwest)

multiple view reconstruction techniques [15] or from LiDAR [30] scans which typically contain only point clouds without corresponding high resolution orthomosaics. Additionally, LiDAR comes with a high data acquisition cost, is generally low resolution, and can contain significant noise. On the other hand, aerial rooftop datasets typically do not contain any 3D modalities such as DSM or point clouds or are too low resolution to perform accurate planar rooftop structure extraction. While there are some datasets [30, 39] which contain full 3D meshes of residential buildings and rooftops, they are generated using synthetic height and textures which may lead to undesirable performance in a real world roof analysis setting. The following datasets are the most closely related works in the area of residential rooftop geometry and scene understanding

Vectorizing World Buildings The VWB dataset [35] is a modification of the SpaceNet 1 challenge dataset [47] which consists of 30cm spatial resolution Maxar WorldView-3 satellite RGB imagery. Patches of size 256 × 256 were manually cropped from the larger images around individual building instances. 2D planar graphs of building roof structures are annotated for 2,001 buildings from the cities of Atlanta, Paris, and Las Vegas. This dataset only contains RGB imagery and 2D wireframe annotations, but no 3D information.

Enschede The Enschede dataset [52] consists of inner and outer roofline vectors of buildings in 8cm spatial resolution aerial orthomosaics taken over the Enschede, Netherlands area. The vector annotations were extracted from the BAG dataset [37] and overlayed onto the georeferenced imagery instead of performing manual annotation which can lead to inaccurate labels. The dataset contains 3,648 512 × 512 image patches cropped around individual buildings. This dataset only contains RGB imagery, 2D wireframe annotations, but no 3D information.



(a) (b) (c) (d) (e) (f)

Figure 4. Distribution plots of ground sampling distance (GSD) (cm/px) for (a) orthomosaic and (b) DSM, image sizes (height & width) (px) for (c) orthomosaic and (d) DSM, (e) roof faces per property, and (f) surface area per roof face (m^2).

Roof Information Dataset (RID) The RID [25] is a dataset of image patches centered around rooftops containing solar panel installations and is intended for photovoltaic potential analysis. This dataset contains some rooftop features within the semantic segmentation categories such as roof dormers and chimneys. However, due to the focus on solar panel detection and low spatial resolution, this dataset becomes insufficient for accurate roof structure extraction and wireframe generation. This dataset only contains RGB imagery, 2D wireframe annotations, but no 3D information.

BuildingWF The BuildingWF dataset [30] is composed of 3,600 polygon meshes of residential buildings with *synthetic* textures along with ground truth 3D wireframes of the *synthetic* meshes.

Mesh-Image Paired Dataset The Mesh-Image paired dataset [39] consists of 2,539 samples of roof geometries extracted from 2D images of residential buildings and then converted to 3D meshes using *synthetic* height information.

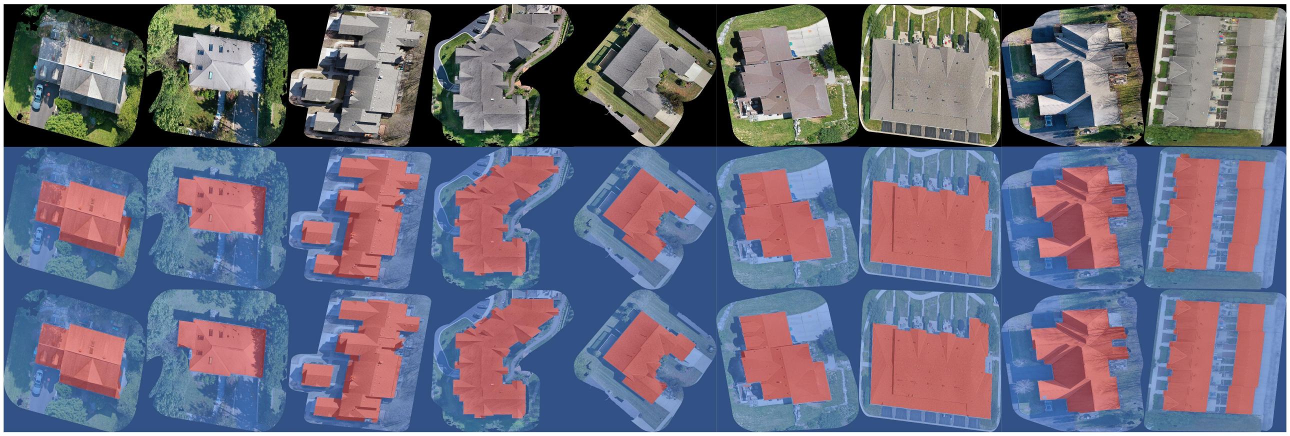


Figure 5. **Roof Outline Extraction** samples and predictions from the ZRG-Test subset using the DeepLabV3-ResNet50 model trained on the ZRG-10k subset. From top to bottom: orthomosaic, ground truth, predictions. (red=**roof**, blue=**background**).

4. The ZRG Dataset

In this paper, we present the Zeitview Rooftop Geometry (ZRG) dataset: a large-scale high resolution multimodal dataset with a focus on residential building rooftops for damage assessment, planar roof structure extraction, and 3D reconstruction.

4.1. Comparison to Related Datasets

Datasets containing residential rooftop buildings either only contain imagery but not realistic height information in the form of DSM or point clouds, or contain point clouds only but no imagery or geometric roof information. Related datasets are also generally too low resolution to support high quality and accurate understanding of roof structures. Table 1 presents comparisons between the ZRG dataset and closely related roof structure datasets detailed in Section 3.

4.2. Data Acquisition

The data collection process utilized several commercially available DJI drones outfitted with high-resolution cameras to acquire imagery for performing residential roof inspections and analysis. The drones were programmed to navigate in a systematic lawn mower pattern, maintaining an altitude of 10-15 feet above the highest point of the roof. Additionally, oblique images were acquired for performing multi-view 3D reconstruction to infer the geometric structure of the rooftops. As illustrated in Figure 3, data for residential properties were collected from various regions across the U.S. and include single-family homes (SFH) and multi-family homes (MFH) or apartment complexes. There is a natural concentration in clients seeking roof inspections in the central and eastern regions of the U.S. This is due to a higher concentration of hail storms occurring in these states [43].

4.3. Post-Processing

Upon completion of the data acquisition phase, the captured images are stitched together and georeferenced to generate an orthomosaic. Further, Digital Surface Model (DSM) and colored point clouds were generated using 3D multi-view reconstruction techniques. Note that the techniques utilized to generate DSMs with lesser GSDs with a factor of $3 - 3.5$, in this case the DSM height and widths are also less than their corresponding orthomosaic. Distribution plot of the GSD and height and widths of the orthomosaics and DSMs are illustrated in Figure 4. The end result is a large-scale dataset of sub-centimeter resolution RGB orthomosaics, DSMs, and colored point clouds of a total of 22,334 properties.

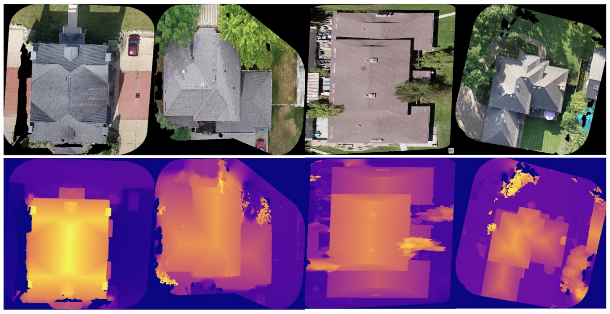


Figure 6. Examples of sample images and DSM pairs containing noise as a result of invalid pixels, overhanging vegetation, and shadows.

4.4. Wireframe Annotation

Our labeling team consists of residential properties inspection domain experts. We utilize a custom annotation tool for generating 3D wireframe annotations into geojson

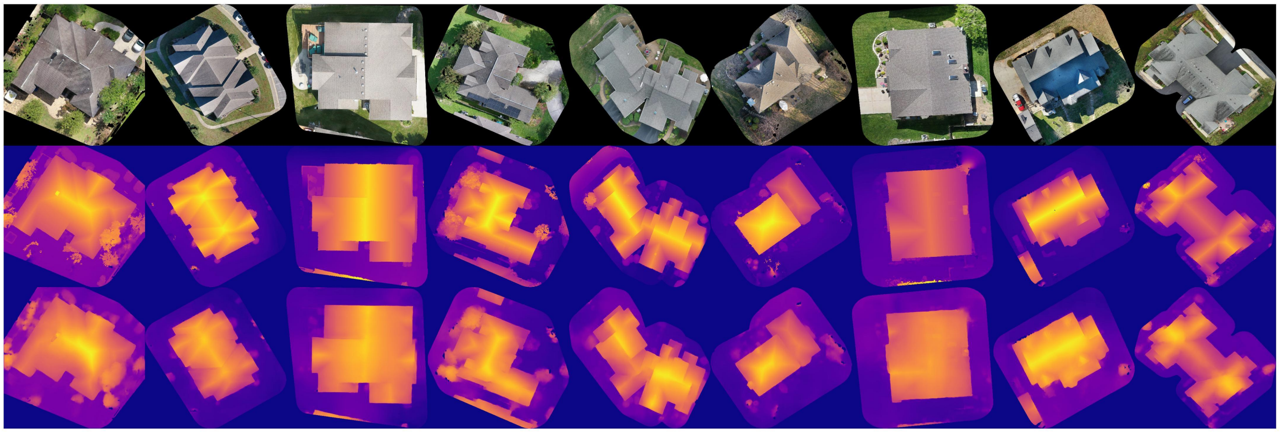


Figure 7. **Monocular Height Estimation** samples and predictions from the ZRG-Test subset using the DeepLabV3-ResNet50 model trained on the ZRG-10k subset. From top to bottom: orthomosaic, ground truth DSM, height predictions. The DSM and predictions are color mapped such that *brighter* indicates *greater height*.

format. Each roof face, or distinct plane on a roof, is annotated with a 3D polygon and separate line geometries for each individual edge. Additional labels such as surface area of the polygon as well as 18 distinct edge type labels are recorded in the file metadata. The 18 edge classes include common roof edge categories such as flashing, ridge, drip edge, hip, and valley. Examples of these categories are provided in Figure 2. A total of 425,660 total faces were annotated with a median of 16 faces per property. Distribution plots of the roof faces per property and surface area for each roof face are provided in Figure 4.

4.5. Limitations and Challenges

Due to the focus on the rooftop during data acquisition, there are some edge cases that arise where image acquisition stage fails to capture the entire scene surrounding the residential building. This can result in invalid pixels being present in the final stitched orthomosaic. We elect to not clean these properties from the dataset as performing inspection and analysis with models robust enough to learn even with the presence of this noise is an important and necessary task. Additionally, natural challenges such as overhanging vegetation and shadows add additional complexity for learning rooftop structure as can be seen in Figure 6.

4.6. Dataset Subsets

Due to the large scale of the ZRG dataset, we split the data into several subsets to make machine learning experimentation simpler and reproducible. First, 1k properties are sampled from the total dataset which we use as a holdout test set called **ZRG-Test**. Then, from the remaining 21k we sample 10k, 1k, and 100 properties which we coin the **ZRG-10k**, **ZRG-1k**, and **ZRG-100** subsets, respectively. Since certain major cities and regions contain more samples, we perform weighted sampling by number of properties per state such that each subset will be more geospatially diverse.

5. Experiments

We conduct several experiments to provide simple baselines using canonical architectures for the tasks of roof outline extraction, monocular height estimation, and planar roof structure extraction.

5.1. Common Training Details

We perform all experiments on a NVIDIA DGX server with 1x NVIDIA A100 with 40GB memory. During training we use the following augmentations: horizontal and vertical flip, random rotation, random perspective, gaussian blur, color jitter, and scale jitter [14] primarily to train models to generalize across variations in data acquisition altitude, seasonal changes, and daily changes resulting in shadows. We use the AdamW [29] optimizer with a learning rate of $\alpha = 3e^{-4}$ throughout. We train each model using automated mixed precision (fp16) for 150 epochs with a batch size of 8 and mixed precision. Additionally, we normalize each orthomosaic using ImageNet statistics.

5.2. Roof Outline Extraction

We pose the task of roof outline extraction as a binary segmentation problem where we seek to segment rooftop pixels from the background. These predictions are important for providing additional information to downstream tasks to assist in the focus on the rooftop. For this experiment we train several canonical segmentation models including U-Net [40], U-Net++ [53], PSPNet [51], and DeepLabV3+ [9], with ResNet [18] encoder backbones.

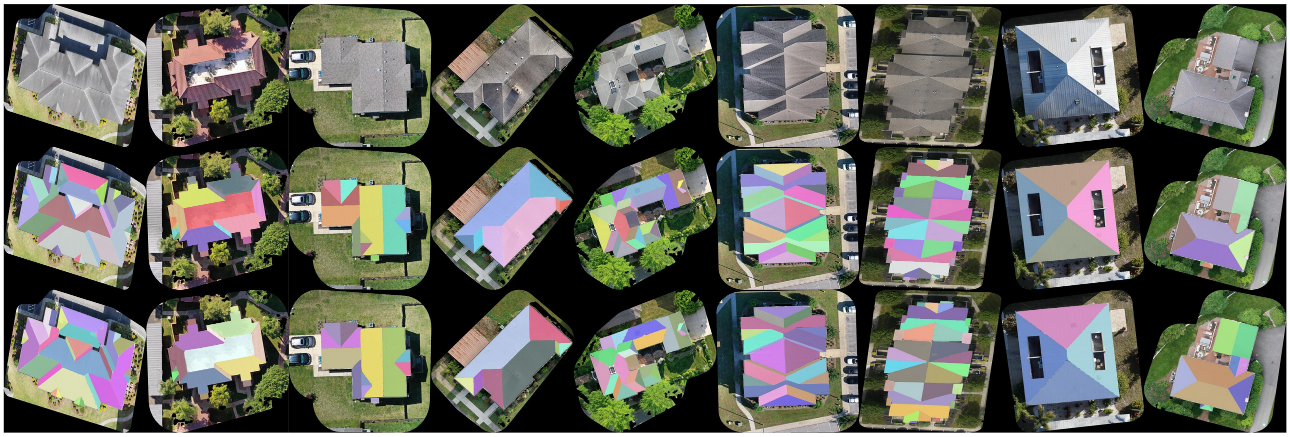


Figure 8. **Planar Roof Structure Extraction** instance segmentation samples and predictions from the ZRG-Test subset using the best performing MaskRCNN model trained on the ZRG-10k subset. From top to bottom: orthomosaic, ground truth, predictions.

We utilize the model implementations from the Segmentation Models PyTorch library [21] with ImageNet pretrained weight initialization [11] in the encoders. We train each model using a joint loss function consisting of the sum of the cross entropy loss and the Intersection-over-Union (IoU) loss. Evaluation is performed on the ZRG-Test holdout subset of the ZRG dataset. Random samples from the ZRG-Test set are displayed in Figure 5 and semantic segmentation metrics for each model are provided in Table 4.

5.3. Monocular Height Estimation

Estimating a surface model represented by the height of each pixel in a monocular (single) overhead view of a building is a difficult yet important problem for automated mapping purposes as explored in [13, 23, 26, 28, 34, 36, 48]. We pose this task as a dense regression problem similar to depth estimation. We utilize the same experimental setup as detailed in Section 5.2. We min-max normalize each DSM to their relative heights using the sample statistics after filtering invalid pixels values. We train each model using a masked L1 loss function to not penalize predictions on invalid pixels. We use and modify the the segmentation model architectures described in Section 5.2 by fixing the output layer to contain a single channel with a ReLU activation [12] for continuous outputs, similarly to the architectures used in [38]. Random samples and predictions from the ZRG-Test set are displayed in Figure 7 and regression metrics for each model are provided in Table 3.

5.4. Planar Roof Structure Extraction

For the task of planar roof structure extraction, we experiment with instance segmentation architectures for segmenting each individual face of the roof from a single overhead view. We experiment by fine-tuning the torchvision [31] implementation of the MaskRCNN architecture [17] which is pretrained on the MS-COCO dataset [27]. Random samples from the ZRG-Test set are displayed in Figure 8. In Table 2 We report mean Average Precision (mAP) at different IoU thresholds and at different scales, (medium and large roof faces based on area). For simplicity, we pose the single-view planar roof structure extraction task as an instance segmentation problem. However, we note that it is also common to utilize corner and junction detection methods [52] in combination with graph neural network (GNN) based architectures [5] as an alternative to solve this problem.

Subset	**mAP**	**mAP$_{50}$**	**mAP$_{75}$**	**mAP$_{M}$**	**mAP$_{L}$**
ZRG-100	40.4	45.3	35.5	37.1	66.3
ZRG-1k	67.9	91.4	44.5	66.4	92.4
ZRG-10k	**72.1**	**97.0**	**47.1**	**90.0**	**96.1**

Table 2. **Planar Roof Structure Extraction** instance segmentation mean Average Precision (mAP) results of a MaskRCNN with a ResNet50-FPN backbone pretrained on COCO and trained on each ZRG subset and evaluated on the ZRG-Test subset. Best results marked in **bold**.

6. Discussion

6.1. Effects of Dataset Size

To explore the necessity of creating a large scale dataset, we repeat the experiments in Section 5.4 for planar roof structure extraction on the ZRG-100, ZRG-1k, and ZRG-10k subsets to evaluate how dataset size affects performance. We train each method for 1k iterations and record the mean Average Precision (mAP) at different thresholds and for different size roof faces (medium (M) and large (L)). As seen in Table 2, the increasing size of the dataset re-

Model	Backbone	MAE	MSE	RMSE
PSPNet	ResNet18	0.0815	0.0167	0.1292
U-Net	ResNet18	0.0738	0.0144	0.1198
U-Net++	ResNet18	0.0727	0.0138	0.1174
DeepLabV3+	ResNet18	**0.0696**	**0.0131**	**0.1143**
PSPNet	ResNet50	0.0793	0.0162	0.1274
U-Net	ResNet50	0.0723	0.0133	0.1155
U-Net++	ResNet50	0.0699	0.0127	0.1129
DeepLabV3+	ResNet50	**0.0679**	**0.0123**	**0.1107**

Table 3. **Monocular Height Estimation results** of various models trained on the ZRG-10k subset and evaluated on the ZRG-Test subset. We report mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) dense regression metrics. Metrics are computed on the relative height values after min-max normalization. Best results are marked in **bold**.

Model	Backbone	OA	F1	mIoU
PSPNet	ResNet18	92.37	90.37	76.99
U-Net	ResNet18	96.89	96.28	92.93
U-Net++	ResNet18	97.59	97.15	95.91
DeepLabV3+	ResNet18	**97.73**	**97.32**	**96.65**
PSPNet	ResNet50	96.48	95.80	92.52
U-Net	ResNet50	97.52	97.06	95.78
U-Net++	ResNet50	97.31	96.82	95.12
DeepLabV3+	ResNet50	**97.81**	**97.42**	**96.83**

Table 4. **Roof Outline Extraction results** of models trained on the ZRG-10k subset and evaluated on the ZRG-Test subset. We report overall accuracy (OA), average F1 score, and mean Intersection-over-Union (mIoU) semantic segmentation metrics. Best results are marked in **bold**.

sults in a significant increase in performance across all metrics with a particularly large increase in AP_M performance in for medium sized roof faces. This illustrates the need for large-scale and high quality labeled datasets to provide greater performance gains rather than iterating with various model architectures.

6.2. Results

With regards to the roof outline extraction and monocular view height estimation tasks, it is clear that the DeepLabV3+ model with a ResNet50 backbone outperforms other models in all cases. Predictions in Figures 5 and 7 visually show that the model is able to properly segment the roof from the background and accurately estimate pixelwise height of the buildings.

For planar roof structure extraction, we can see that there is an increasing relationship between the performance and the size of the training set, particularly with a large increase of 23.6 in mAP for medium sized roof faces when increasing the dataset size from 1k to 10k samples. Additionally, Figure 8 shows that the model is able to extrapolate the segmentation of roof face structures even in the presence of occlusion of areas of the face due to overhanging vegetation.

6.3. Future Work

While each task may not appear to be as useful independently, the bigger picture of combining predicted roof outlines, height estimations, and segmented roof faces to generated 3D reconstructions and wireframes of rooftops allows for deeper analysis and insights into the condition and surface area breakdown of each roof. However, we leave this combination of model outputs or joint learning for future work.

We plan to perform further annotation of the dataset to include labels for classification of roof types [1, 6, 33], e.g. gable, complex, pyramidal, as well as labels for classification of building type, particularly single vs. multi-family homes, similar to the work described in [3]. Both label categories can be used to further benefit fine-grained analysis of rooftops.

While we do not explicitly utilize the multi-view imagery used to generate the DSM and point cloud, we acknowledge that there are numerous recent reconstruction methods related to Neural Radiance Fields (NeRF) [32, 49, 50] which can more accurately generate 3D models and novel views of each property.

Due to the regional diversity of the properties, our dataset can result in subpopulation distribution shifts described in [24, 42]. Further analysis of the generalization and geographic bias across regions of the U.S. and distance from urban metropolitan areas is outside the scope of this paper and we leave for future work.

7. Conclusion

In this paper, we presented ZRG, a 3D residential rooftop understanding dataset, which we have shown through thorough analysis and several baseline experiments what is possible with a large-scale dataset with multiple modalities. We hope that our work advances and generates novel ideas for additional applications of residential rooftop structure extraction and understanding and inspires the research community to develop additional residential rooftop datasets.

References

[1] Fatemeh Alidoost and Hossein Arefi. A cnn-based approach for automatic building detection and recognition of roof types using a single aerial image. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86:235–248, 2018. 7

[2] Dan Assouline, Nahid Mohajeri, and Jean-Louis Scartezzini. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Solar Energy*, 141:278–296, 2017. 2

[3] Abhilash Bandam, Eedris Busari, Chloi Syranidou, Jochen Linssen, and Detlef Stolten. Classification of building types in germany: A data-driven modeling approach. *Data*, 7(4):45, 2022. 7

[4] Katalin Bódis, Ioannis Kougias, Arnulf Jäger-Waldau, Nigel Taylor, and Sándor Szabó. A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the european union. *Renewable and Sustainable Energy Reviews*, 114:109309, 2019. 2

[5] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. 6

[6] M Buyukdemircioglu, R Can, and S Kocaman. Deep learning based roof type classification using very high resolution aerial imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:55–60, 2021. 7

[7] Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021. 2

[8] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662, 2020. 2

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 5

[10] Isaac Corley and Peyman Najafirad. Supervising remote sensing change detection models with 3d surface semantics. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3753–3757. IEEE, 2022. 2

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[12] Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological cybernetics*, 20(3-4):121–136, 1975. 6

[13] Pedram Ghamisi and Naoto Yokoya. Img2dsm: Height simulation from single imagery using conditional generative adversarial net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):794–798, 2018. 6

[14] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2918–2928, 2021. 5

[15] Jianwei Guo, Yanchao Liu, Xin Song, Haoyu Liu, Xiaopeng Zhang, and Zhanglin Cheng. Line-based 3d building abstraction and polygonal surface reconstruction from images. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 3

[16] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, and Matthew Gaston. xbd: A dataset for assessing building damage from satellite imagery. *arXiv preprint arXiv:1911.09296*, 2019. 2

[17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 6

[18] 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. 5

[19] Mehdi P Heris, Nathan Leon Foks, Kenneth J Bagstad, Austin Troy, and Zachary H Ancona. A rasterized building footprint dataset for the united states. *Scientific data*, 7(1):207, 2020. 2

[20] Taehoon Hong, Minhyun Lee, Choongwan Koo, Kwangbok Jeong, and Jimin Kim. Development of a method for estimating the rooftop solar photovoltaic (pv) potential by analyzing the available rooftop area using hillshade analysis. *Applied Energy*, 194:320–332, 2017. 2

[21] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019. 6

[22] Umit Isikdag and Sisi Zlatanova. Interactive modelling of buildings in google earth: A 3d tool for urban planning. In *Developments in 3D Geo-Information Sciences*, pages 52–70. Springer, 2009. 2

[23] Savvas Karatsiolis, Andreas Kamilaris, and Ian Cole. Img2ndsm: Height estimation from single airborne rgb images with deep learning. *Remote Sensing*, 13(12):2417, 2021. 6

[24] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pages 5637–5664. PMLR, 2021. 7

[25] Sebastian Krapf, Lukas Bogenrieder, Fabian Netzler, Georg Balke, and Markus Lienkamp. Rid—roof information dataset for computer vision-based photovoltaic potential assessment. *Remote Sensing*, 14(10):2299, 2022. 2, 3

[26] Xiang Li, Mingyang Wang, and Yi Fang. Height estimation from single aerial images using a deep ordinal regression network. *IEEE Geoscience and Remote Sensing Letters*, 2020. 6

[27] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 6

[28] 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. 2, 6

[29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[30] Yicheng Luo, Jing Ren, Xuefei Zhe, Di Kang, Yajing Xu, Peter Wonka, and Linchao Bao. Learning to construct 3d building wireframes from 3d line clouds. *arXiv preprint arXiv:2208.11948*, 2022. 2, 3

[31] TorchVision maintainers and contributors. Torchvision: Pytorch's computer vision library. `https://github.com/pytorch/vision`, 2016. 6

[32] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 7

[33] Nahid Mohajeri, Dan Assouline, Berenice Guiboud, Andreas Bill, Agust Gudmundsson, and Jean-Louis Scartezzini. A city-scale roof shape classification using machine learning for solar energy applications. *Renewable Energy*, 121:81–93, 2018. 7

[34] Lichao Mou and Xiao Xiang Zhu. Im2height: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network. *arXiv preprint arXiv:1802.10249*, 2018. 2, 6

[35] Nelson Nauata and Yasutaka Furukawa. Vectorizing world buildings: Planar graph reconstruction by primitive detection and relationship inference. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 711–726. Springer, 2020. 2, 3

[36] Emmanouil Panagiotou, Georgios Chochlakis, Lazaros Grammatikopoulos, and Eleni Charou. Generating elevation surface from a single rgb remotely sensed image using deep learning. *Remote Sensing*, 12(12):2002, 2020. 6

[37] Publieke Dienstverlening Op de Kaart (PDOK). Dataset: Basisregistratie adressen en gebouwen (bag). 3

[38] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12179–12188, 2021. 6

[39] Jing Ren, Biao Zhang, Bojian Wu, Jianqiang Huang, Lubin Fan, Maks Ovsjanikov, and Peter Wonka. Intuitive and efficient roof modeling for reconstruction and synthesis. *arXiv preprint arXiv:2109.07683*, 2021. 2, 3

[40] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 5

[41] Premiere Roofing, Mar 2023. 2

[42] Shiori Sagawa, Pang Wei Koh, Tony Lee, Irena Gao, Sang Michael Xie, Kendrick Shen, Ananya Kumar, Weihua Hu, Michihiro Yasunaga, Henrik Marklund, et al. Extending the wilds benchmark for unsupervised adaptation. *arXiv preprint arXiv:2112.05090*, 2021. 7

[43] Joseph T Schaefer, Jason J Levit, Steven J Weiss, and Daniel W McCarthy. The frequency of large hail over the contiguous united states. In *Preprints, 14th Conf. on Applied Climatology, Seattle, WA, Amer. Meteor. Soc*, volume 3, 2004. 4

[44] Li Shen, Yao Lu, Hao Chen, Hao Wei, Donghai Xie, Jiabao Yue, Rui Chen, Shouye Lv, and Bitao Jiang. S2looking: A satellite side-looking dataset for building change detection. *Remote Sensing*, 13(24):5094, 2021. 2

[45] 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. 2

[46] Kiti Suomalainen, Vincent Wang, and Basil Sharp. Rooftop solar potential based on lidar data: Bottom-up assessment at neighbourhood level. *Renewable Energy*, 111:463–475, 2017. 2

[47] Adam Van Etten, Dave Lindenbaum, and Todd M Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018. 3

[48] Siyuan Xing, Qiulei Dong, and Zhanyi Hu. Sce-net: Self-and cross-enhancement network for single-view height estimation and semantic segmentation. *Remote Sensing*, 14(9):2252, 2022. 6

[49] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 7

[50] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 7

[51] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017. 5

[52] Wufan Zhao, Claudio Persello, and Alfred Stein. Extracting planar roof structures from very high resolution images using graph neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:34–45, 2022. 2, 3, 6

[53] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018. 5