arXiv:2111.11397v2 [cs.CV] 16 Dec 2021

Solar Potential Assessment using Buildings Segmentation from Aerial Images

Hasan Nasrallah, Abed Ellatif Samhat [*] Ghaleb Faour [†] Yilei Shi [‡] Ali J. Ghandour [§]

Abstract

Estimating the solar potential of buildings' rooftops at a large scale is a fundamental step for every country to utilize its solar power efficiently. However, such estimation becomes time-consuming and costly if done through on-site measurements. This paper uses deep learning-based multi-class instance segmentation to extract buildings' footprints from satellite images. Hence, we introduce Lebanon's first complete and comprehensive buildings' footprints map. Furthermore, we propose a photovoltaic panels placement algorithm to estimate the solar potential of every rooftop, which results in Lebanon's first buildings' solar potential map too. Finally, we report average and total solar potential per district and localize regions corresponding to the highest solar potential yield.

1. Introduction

Building footprints extraction from aerial imagery is essential for many urban applications, including geographical databases, land use, and change analysis. Fully automated extraction and recognition of buildings' footprints geometries can help estimate the solar potential of every rooftop. Estimating rooftops' solar potential provides more insight into how a given country can efficiently utilize renewable resources for solar power generation.

Solar power harvesting paves the way for ensuring a greener future and a better economic status. The first step to accomplish this task is acquiring rooftop geometries that we tackle via satellite imagery analysis and object segmentation.

Segmentation of urban aerial imagery is currently undergoing significant attention in the research community and notable development efforts in the industry. Remote sensing images are usually complex and characterized by significant intra-class variations, and often low inter-class variations [1].

Deep learning significantly reduces the time and cost required for aerial imagery segmentation due to its capabilities in automatically extracting features and patterns present in large scenes. This work uses deep convolutional neural networks to extract rooftop geometries and obtain Lebanon's first complete and comprehensive urban map.

We then use a solar panel placement algorithm that considers the area and morphology of every rooftop to estimate the number of panels that can fit on top. Automated solar potential estimation is made feasible given the corresponding photovoltaic (PV) power generated per unit of the installed panels.

The contribution of this paper is four-folds: *(i)* leverage the first complete Lebanese urban map via buildings' instance segmentation from satellite images, *(ii)* estimate the solar potential of every rooftop by taking into consideration its morphology, *(iii)* calculate the average and total solar potential of every Lebanese district per year. *(iv)* highlight regions with the highest solar potential yield that can be utilized in a better way.

The rest of the paper is organized as follows: Section 2 shows the literature review related to the use of deep learning models for building footprints detection from aerial images. We present in Section 4 the methodology adopted in this work and the adopted deep learning model design. Study area is described in Section 3 and the building segmentation results are presented in Section **??**. The solar potential application is presented in Section 5. Finally, Section 6 concludes the paper.

2. Related Work

Semantic Segmentation is the process of classifying every pixel present in an image into one or more predefined classes. Semantic segmentation networks [5, 19, 25, 36] are widely used nowadays to capture various features present in satellite images like ships [23] and buildings [26]. The authors in [25] propose an encoder-decoder-like architecture, where skip connections are utilized to decode feature

[*]H. Nasrallah and A. Samhat are with the Lebanese University.

[†]A.J. Ghandour and G. Faour are with the National Center for Remote Sensing, at the National Council for Scientific Research (CNRS), Beirut, Lebanon.

[‡]Y. Shi is with the Technical University of Munich.

[§]Corresponding Author: Ali J. Ghandour, aghandour@cnrs.edu.lb

maps at different spatial resolutions. In DeepLabV3+ [5], the authors use Atrous Spatial Pyramid Pooling to capture multi-scale information at different fields of view. SANet authors [39] propose squeeze and attention modules as a mechanism to re-weight the feature maps at a global and local aspect resulting in a boosted semantic segmentation accuracy.

Instance segmentation is a much more challenging task in the computer vision field compared to semantic segmentation. Instance segmentation labels objects of the same class as distinct individual instances using complex networks composed of two stages usually. Region Proposal networks, Bounding Box Regressors, in addition to classifiers such as Fast and Faster-Region-based CNN (RCNN) [12, 24], Cascade-RCNN [3], and Efficient-Det [30] are examples of such networks proposed in the literature.

Authors in [37] use Mask-RCNN [13] to segment buildings' instances and then apply a polygon border regularization process to regularize noisy buildings boundaries. In [2], the authors perform instance segmentation using Fully Convolutional Neural Networks composed of a segmentation network followed by a Direction Net (DN) to predict the transform distance of every point and finally a Watershed Transform Network (WTN) to predict the energy levels applied in the watershed transform. BlendMask [4] blends together instance-level score maps and attention masks cropped using bounding box proposals extracted from the FCOS [32] detector. Finally, authors in [33] conceptually divide an image into an SxS grid, where each grid is responsible for segmenting the instance whose center falls within this grid.

Building segmentation. Several techniques tackle the problem of building segmentation from aerial imagery. In [26], the authors use gated graph convolutional neural networks to output a Truncated Signed Distance Map (TSDM), which is then converted into a semantic segmentation mask of buildings. In [21], the authors propose two plug-and-play modules to generate spatial augmented, and channel augmented features for semantic segmentation from satellite images. In [18], the authors use augmentations like slicing, rescaling, and rotations, in addition to GIS data, to improve buildings' footprint extraction. Authors in [31] use Siamese networks to both segment and classify buildings present in pre- and post- disaster images. However, a more direct approach is presented in [15] where the authors only use a semantic segmentation network with an additional output mask designating spacing between nearby buildings to separate building instances.

Rooftop Solar Potential Estimation is currently drawing the attention of geospatial deep learning researchers due to its effectiveness in accurately predicting the potential usage of rooftops for solar power generation. Significant advances in this field have been made using statistical models, computer vision, numerical analysis, and geographic information systems. In [9], the authors use collected data about the solar radiation and the rooftop areas in Beirut city to estimate solar potential, without taking into consideration each rooftop morphology. In [11], authors use high precision photographic sensors mounted on a UAV to scan and create a digital surface model (DSM) for a single building. Further statistical analysis of the output DSM model, including shading analysis, solar irradiance estimation, and panel placement, was conducted to estimate a single roof's solar potential. In [8], the authors use satellite imagery to divide and segment building rooftops into sections using convolutional neural networks. They predict each section's pitch, azimuth, and shading mask and then use a greedy algorithm to place solar panels on rooftops to estimate their solar potential. A similar approach was adopted in [38] to estimate the solar potential at a city scale in China.

To the best of the authors' knowledge, there is no previous work to estimate the rooftop solar potential at a country scale for Lebanon. The presented work is thus the first attempt to produce a Lebanese urban and solar potential map. We used a deep learning model to segment rooftops from high-resolution satellite imagery (50cm/pixel). Then, we devise a greedy algorithm to simulate photovoltaic (PV) panels' placement based on each rooftop morphology.

3. Study Area

We aim in this research to build an efficient and accurate building segmentation model from Lebanese satellite images.

We first chose a 35372x28874 GeoTIFF image with RGB channels taken by the GeoEye-1 satellite sensor of 50 cm/pixel resolution covering the Tyre district in South Lebanon. We cropped the chosen area of interest into 1024x1024 chips, resulting in 338 tiles for manual annotation using the VGG-Image-Annotator tool. Finally, we further cropped each tile into non-overlapping 512x512 sized images while preserving each image's relative labeled polygons indices.

The final training dataset includes 1,352 tiles of 512x512 dimensions with 10,000 buildings' objects. We refer to this dataset as Tyre Urban Map (TUM) dataset. Figure 1 shows the number of images present in the TUM dataset for each given range of buildings' count. The average is 28 buildings per image.

As for the test set, we selected 30 Areas Of Interest (**AOI**) from different Lebanese regions, including Beirut, Saida, Jounieh, Jbeil, and Tripoli. Chosen AOI's encompass dense urban regions (Beirut), structured urban regions (Saida), and some rural areas. This diversity would help assess the generalizability of our model over the whole Lebanese geographical area.

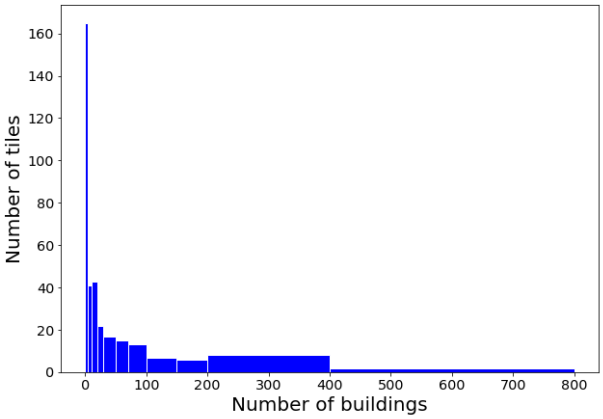

Figure 1. Number of 1024px Tiles per buildings count.

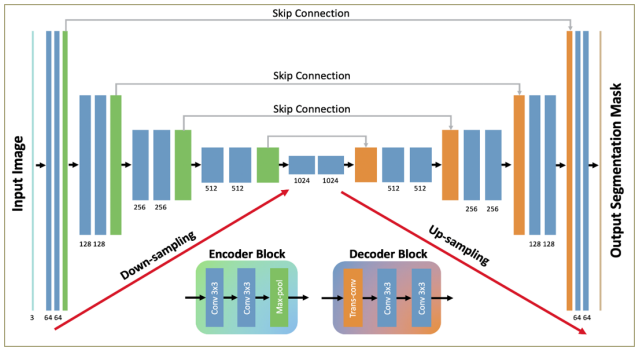

Figure 2. Encoder-Decoder UNet Architecture with Skip Connections

4. Buildings' Footprints Map

4.1. Model Architecture

UNet [25] is an end-to-end fully convolutional neural network that serves the purpose of semantic segmentation of input images over multiple classes. It consists of a contracting path [Encoder] and an expansive path [Decoder] with skip connections added between every two symmetrical blocks of same size in the encoder and the decoder. The Encoder is comprised of repeated down-sampling convolutional and pooling layers followed by an activation function at every stage (i.e. RELU) to output feature maps of higher semantics and lower resolution (half) from the previous stage. Furthermore, the Decoder is made up of consecutive up-sampling blocks consisting of 3x3 convolution and an activation function followed by an up-sampling layer that doubles the resolution of the input maps. Each decoder block takes as input the output of the previous decoder block concatenated to the feature maps of same size from the parallel encoder block [skip connection]. Skip connections allow information to flow directly from the low level to high-level feature maps without alternations that even further improve localization accuracy and speed up convergence [15]. The selection of the activation function at the last decoder block depends on whether the output classes are dependent/independent and overlapping/non-overlapping, which will explain in later sections our choice of using the Sigmoid activation over the Softmax activation in our approach [22].

The overall is a U-shaped encoder-decoder architecture that takes an image as input, in our case 'RGB image', of size HxW and outputs a multi-channel map of same size where every channel designates a pixel-wise probability map for a certain class as shown in Figure 2. Typically, we choose a probability threshold (0.5 by default) to classify each pixel to a given class across all channels spatially.

The Encoder part of the architecture can be chosen from powerful and deep convolutional neural networks like Residual Nets [14, 34], Inception Nets [28], dual-path nets [6], or the newly introduced Efficient-Nets [29]. Efficient-Net-B0 is the baseline of a multi-objective neural architecture search that optimizes both accuracy and memory usage. Efficient-Nets-B1 to Efficient-Nets-B7 can be obtained using a compound scaling method. In our experiments, Efficient-Net-B3 was found to perform better in terms of accuracy and variance than other members of the Efficient-Net family members and other encoders like ResNet34, ResNeXt50, InceptionV4, InceptionResNetV2 and DPN92 as shown in Table 1, where distance is defined as the Fscore difference on the training and validation datasets.

Backbone	F-score(%)	Distance(%)
ResNet34	82.8	4.33
ResNeXt50	83.7	4.24
Inception-ResNetV2	84.0	3.55
InceptionV4	84.1	4.95
DPN92	83.8	3.74
EfficientNetB2	83	2.93
EfficientNetB3	**84.3**	**2.88**
EfficientNetB4	83.8	2.97

Table 1. F-score and Distance percentages on the validation set using different backbones.

As for the loss function, we chose a normalized weighted sum of losses across all output channels. Each single channel loss is a combination of Dice loss and Binary Cross-Entropy loss [16] in order to leverage the benefits of both.

4.2. Training Pipeline

Our best model is trained over 100 epochs using Adam Optimizer [17] and the One-Cycle learning rate policy [27] starting with an initial learning rate = $\frac{0.0001}{20}$ and increases for 40 epochs in a cosine annealing manner till it reaches a maximum of 0.0001 and then decreases for the rest 60 epoch in the same annealing fashion to a much lower learning rate = $\frac{\frac{0.0001}{20}}{1000}$. When learning rate is high, the the One-Cycle learning rate policy works as a regularization method and keep network from over-fitting.

Our GPU card is a Titan-Xp video card with 12GB of virtual RAM allowing us to train our model on a batch size of 10 512x512 pixel satellite images, however since this is a low batch size and the images are relevantly small, the model will see data of low diversity at each iteration , and thus we needed to utilize a higher batch size given our resources, so we used mixed precision training [20] where some operations like convolutions are performed over 16-bit instead of 32 saving up half the memory needed for these computations. And hence, we were able to increase our batch size to 16 images.

During training, we apply randomly train time augmentations on images, where each image undergoes a series of augmentations. Furthermore, we use CutMix [35] data augmentation to further enhance our train time augmentations and train a robust model.

Finally, it is worthy to note that we also used inference-time augmentations and we employed post-processing techniques to extract instance masks from semantic segmentation. The details of those steps are beyond the scope of this manuscript.

5. Solar Potential Assessment

In this section, we will discuss how we utilized the generated buildings footprints in order to calculate the solar potential per year of rooftops over all Lebanon. In order to calculate the solar potential of each rooftop, we first have to place solar panels within the designated area of the extracted building footprints . After placing the solar panels, we will then have information about how many solar panels can fit in each roof. Finally we use the following formula to calculate the solar potential of every rooftop:

$$SolarPotential(MWh/year) = N_{panels} * P_{nominal} * PV_{out}$$

Where N_{panels} is the number of panels placed on the roof, $P_{nominal}$ is the maximum power of the solar panel module in (KWp), and PV_{out} is the specific photovoltaic power in (MWh/KWp)

5.1. Building Footprint Regularization

A simple but yet mandatory step before solar panel placement, is the regularization of building footprint polygons. This procedure not only results in more eye-pleasing polygonal shapes, but also removes irregularities and noise at the footprint boundaries. Thus it assures that less solar panels will be ruled out of being placed at the boundaries due to pixel shifts caused by these shape irregularities. We use ArcGIS Pro Software to do so with the help of the **arcpy.ddd.RegularizeBuildingFootprint** module, and we choose the **RIGHT ANGLES AND DIAGONALS** method to output regularized building footprints with 90 and 45 degrees corners as shown in Figure 3.

Figure 3. Building Footprint Regularization

5.2. PV Panels Placement

The first step in calculating the solar potential of rooftops, is to find how many solar panels can fit in each roof. But since we lack information like the relative slope and azimuth angle of each rooftop, we assume that the roofs are flat and at an orthogonal angle with the satellite sensor. Other Factors like shading are also not taken into consideration due to insufficient knowledge of the building heights. We use a standard commercial type of solar panel modules of dimensions (1x1.98) meters and a nominal power $P_{nominal}$ = 0.4 KWp. Simply dividing the area of the rooftop by the area of the module does not take the rooftop's morphology into account and thus it is not accurate enough. We use an algorithm to place the panels similar the one proposed here [7] as follows:

For each building footprint polygon, we find its minimum bounding rectangle. Then we designate the rectangle's longest axis as the main axis of the rooftop. The solar panels are placed in a greedy way next to each other in the rectangle along the main axis. Finally we remove the solar panels that extend out of the original rooftop polygon geometry. The whole process is shown in Figure4.

5.3. Solar Potential Map

PV_{out} data were obtained from the "Global Solar Atlas 2.0", a free, web-based application that utilizes SolarGIS data. We obtain a 240x360 pixel GeoTIF image/map of resolution (0.0083x0.0083), where an average yearly PV_{out} is

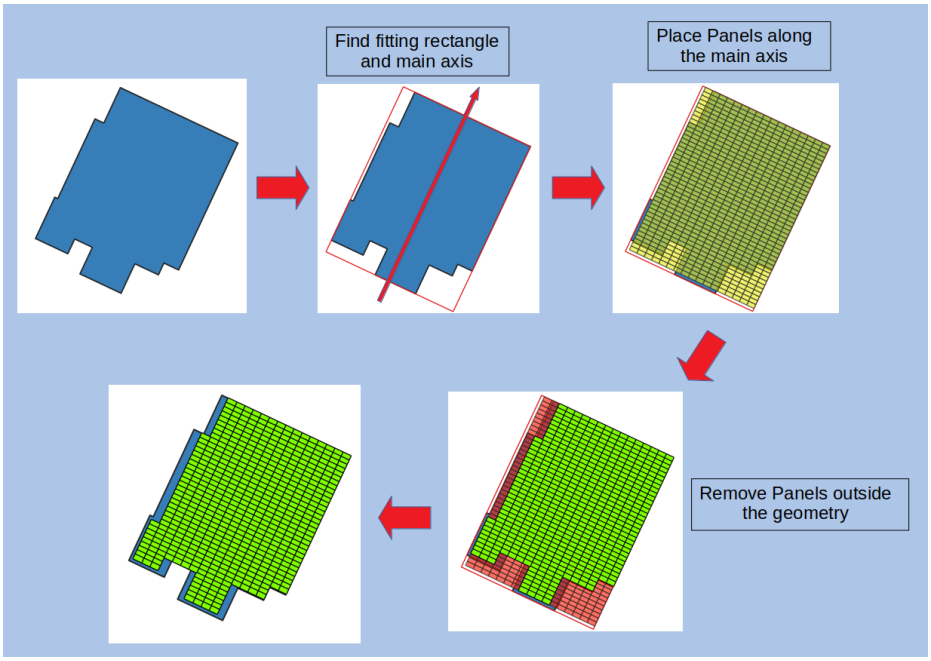

Figure 4. Solar Panels Placement Algorithm

calculated over $NxNm^2$ tiles. For every rooftop, we locate the corresponding tile it lies within in the PV_{out} map to obtain its relative PV_{out} value in (MWh/KWp). Thus we base our calculation on more accurate and precise PV_{out} map than other studies [10], where they use a fixed value over the entire city. The PV_{out} heatmap is shown in Figure 5.

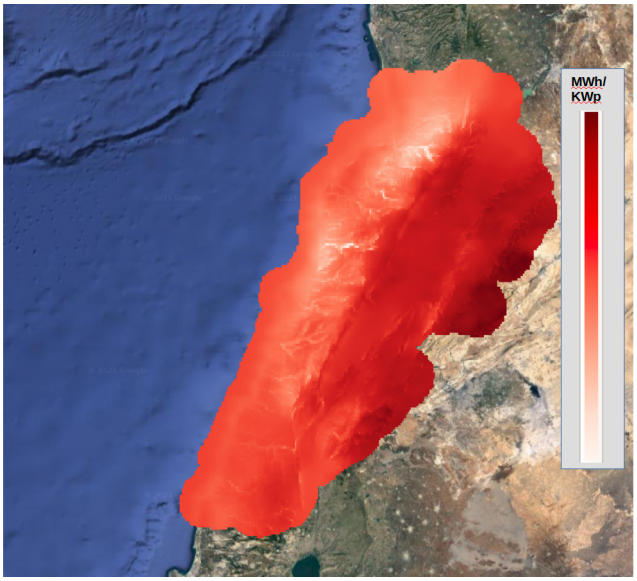

Figure 5. PV_{out} heatmap

Finally, we proceed on calculating the solar potential of every building rooftop based on the equation stated in the introduction of this section. For every rooftop we count the number of placed solar panels and locate the specific tile it belongs to in the PV_{out} map to obtain its relative PV_{out} value. Since we use a greedy algorithm for the solar panels placement, this results in over-placement of these modules on the roof. In reality, less solar panels could be placed on the rooftop due to some obstacles or the inclined slope of the surface. So we further multiply the number of solar panels with ratios **[0.1, 0.25 ,0.5, 0.75]** for comparison and to obtain more realistic results, because in reality less than 75% of these greedily placed panels could be really placed on the roof. Figure 6 shows how every rooftop is assigned a specific solar potential value.

For every district, we calculate the total solar potential (GWh) and the average solar potential produced by the rooftops(MWh) of its buildings. Figure 7 shows these produced maps. From map(a), we can show that Beirut and Tripoli Districts maintain the highest average solar potential per rooftop (92 and 70 $(MWh/year)$) respectively if 50% of solar panels were placed. This reflects the fact that the rooftops in these district have higher relative areas that other districts. However as shown in Figure 9, these areas contribute to the most glowing regions indicating a high average solar potential per $4Km^2$ which is due to the high building density present. If effects like shadowing were taken into consideration, we expect a lower average solar

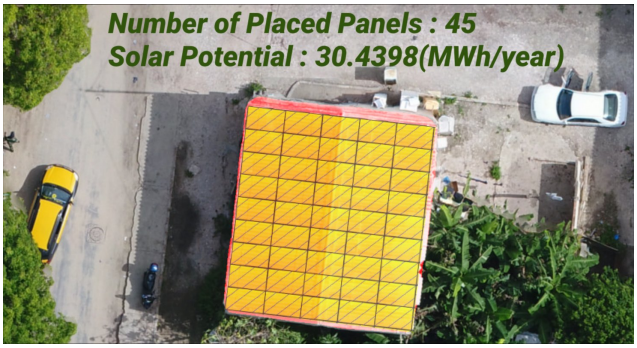

Figure 6. Solar Potential Value for a given rooftop after solar panel placment.

potential in these high density districts. Map(b) shows that districts like Baalbeck,Zahle and Akkar have the highest total solar potentials (2885,1979 and 1775 $(GWh/year)$) despite holding a low average solar potential value per building. Now these mentioned districts contain vast unpopulated areas which makes them capable of producing even more solar energy. We also infer from districts having both low average and total solar potential like Hermel, Hasbaya and Jezzine , that their areas are not utilized properly and we can take advantage of their void spaces for solar energy production.

We also show the number of buildings for each solar potential range if 50% of solar panels were placed in Figure 8. We deduce that one can expect that his rooftop can produce a solar potential up-to $100(MWh/year)$ if he places 50%. Rooftops with higher solar potential are a rare case, and mainly values above $350(MWh/year)$ are due to segmentation errors in high dense areas like slums. The distribution even shows that rooftops with lower solar potentials $(1up-to10(MWh/year)$ correspond to more than $120K$ detected buildings of supposedly small roof areas considering that PV_{out} is nearly uniform in lebanon as shown in Figure 5.

Figure 9 visualizes a heatmap of the total solar potential in $(MWh/year)$ for every $4Km^2$ tile. Rather than inspecting the orange/yellowish rare parts of the map, we should take into consideration the nearly uniformly blue huge part in north Lebanon. This part of low average solar potential also holds a darker red color in the PV_{out} map (Figure 5 than most of the other regions which indicates a higher Pv_{out} value. The fact that this area belongs to the western Lebanese mountain range(needs discussion to continue)....

We also show the Table 2 the total solar potential (GWh/year) for every district given a percentage of solar panels that can be placed on the rooftops the greedy algorithm. In reality it is not feasible to place 100% of the proposed solar panels.

6. Conclusion

In this paper we presented how we used a 2-class segmentation model to extract building footprints from satellite imagery for Lebanon. We showed how we used our generated complete urban map to estimate the solar potential of rooftops over all Lebanon via solar panel placement.

References

[1] Rasha Alshehhi, Prashanth Reddy Marpu, Wei Lee Woon, and Mauro Dalla Mura. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:139 – 149, 2017. 1

[2] M. Bai and R. Urtasun. Deep watershed transform for instance segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2858–2866, 2017. 2

[3] Zhaowei Cai and N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6154–6162, 2018. 2

[4] Hao Chen, Kunyang Sun, Zhi Tian, Chunhua Shen, Yongming Huang, and Youliang Yan. Blendmask: Top-down meets bottom-up for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[5] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv e-prints*, page arXiv:1706.05587, June 2017. 1, 2

[6] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and Jiashi Feng. Dual path networks. In *NIPS*, 2017. 3

[7] Daniel de Barros Soares, François ANDRIEUX, Bastien HELL, Julien LENHARDT, JORDI BADOSA, Sylvain GAVOILLE, Stéphane GAIFFAS, and Emmanuel BACRY. Predicting the solar potential of rooftops using image segmentation and structured data. In *NeurIPS 2020 Workshop on Tackling Climate Change with Machine Learning*, 2020. 4

[8] Daniel de Barros Soares, François Andrieux, Bastien Hell, Julien Lenhardt, Jordi Badosa, Sylvain Gavoille, Stéphane Gaïffas, and Emmanuel Bacry. Predicting the solar potential of rooftops using image segmentation and structured data. *CoRR*, abs/2106.15268, 2021. 2

[9] Hossein Eslami, Sara Najem, Dana Abi Ghanem, and Ali Ahmad. The potential of urban distributed solar energy in transition economies: The case of beirut city. *J. Environ. Manage.*, 285(112121):112121, May 2021. 2

[10] Hossein Eslami, Sara Najem, Dana Abi Ghanem, and Ali Ahmad. The potential of urban distributed solar energy in transition economies: The case of beirut city. *Journal of Environmental Management*, 285:112121, 2021. 5

[11] Jose Eduardo Fuentes, Francisco David Moya, and Oscar Danilo Montoya. Method for estimating solar energy potential based on photogrammetry from unmanned aerial vehicles. *Electronics*, 9(12), 2020. 2

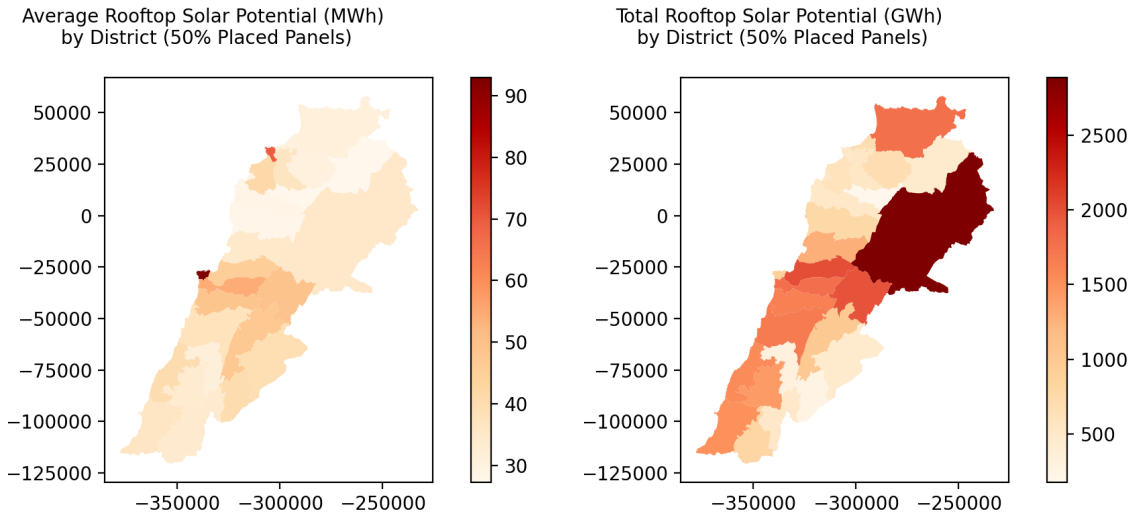

Figure 7. Average and Total Solar Potential Maps By District

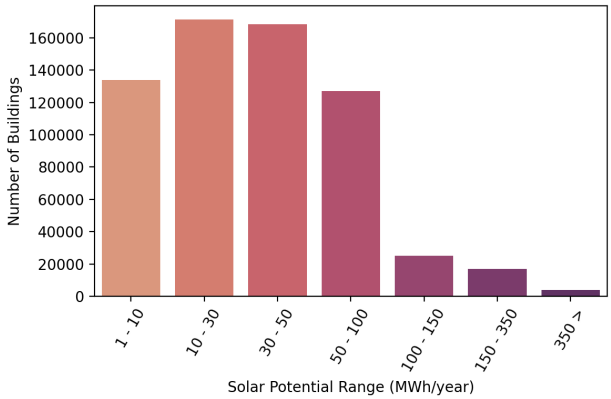

Figure 8. Building Count for Every Solar Potential Range.

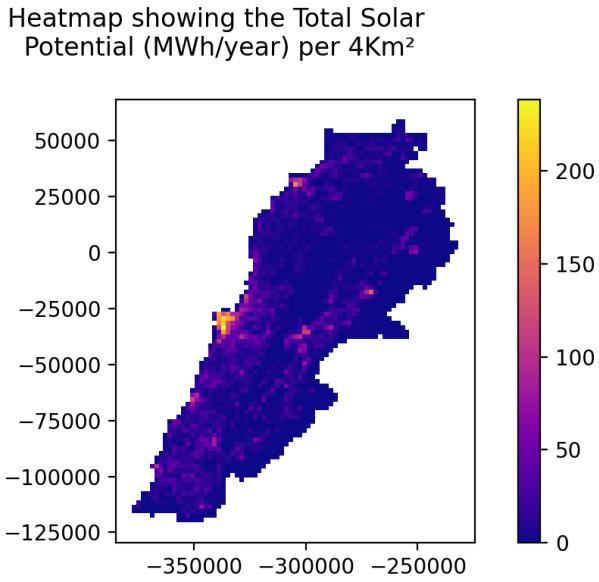

Figure 9. Heatmap showing the Total Solar Potential (MWh/year) per $4Km^2$ when placing 50% of solar panels on the rooftops

[12] R. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015. 2

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397, 2020. 2

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3

[15] Vladimir Iglovikov, Selim Seferbekov, Alexander Buslaev, and Alexey Shvets. Ternausnetv2: Fully convolutional network for instance segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 2, 3

[16] S. Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7, 2020. 3

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015. 4

[18] Weijia Li, Conghui He, Jiarui Fang, Juepeng Zheng, Hao-

District \ Panels(%)	Total Solar Potential (GWh/year)				
	100%	75%	50%	25%	10%
Baalbek	5771	4328	2885	1442	577
El Metn	4029	3022	2014	1007	402
Zahlé	3958	2968	1979	989	395
Baabda	3572	2679	1786	893	357
Akkar	3550	2662	1775	887	355
Chouf	3359	2519	1679	839	335
Aley	3282	2461	1641	820	328
Saida	3095	2321	1547	773	309
Sour	3006	2255	1503	751	300
Nabatiyé	2860	2145	1430	715	286
Kesrouane	2556	1917	1278	639	255
Békaa Ouest	2019	1514	1009	504	201
Beyrouth	1814	1360	907	453	181
Bent Jbail	1637	1227	818	409	163
Jbail	1556	1167	778	389	155
Minié-Danniyé	1409	1057	704	352	140
Koura	1206	904	603	301	120
Zgharta	1079	809	539	269	107
Batroun	1046	784	523	261	104
Tripoli	1032	774	516	258	103
Marjayoun	1017	762	508	254	101
Hermel	959	719	479	239	95
Rachaiya	954	716	477	238	95
Jezzine	623	467	311	155	62
Hasbaiya	532	399	266	133	53
Bcharré	353	265	176	88	35

Table 2. Total Solar Potential($GWh/year$) per district for different percentages of placed solar panels.

huan Fu, and Le Yu. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data. *Remote Sensing*, 11(4), 2019. 2

[19] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 936–944, 2017. 1

[20] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed precision training. In *International Conference on Learning Representations*, 2018. 4

[21] Lichao Mou, Yuansheng Hua, and Xiao Xiang Zhu. A relation-augmented fully convolutional network for semantic segmentation in aerial scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[22] Chigozie Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *ArXiv*, abs/1811.03378, 2018. 3

[23] Venkatesh R and Anand Metha. Segmenting ships in satellite imagery with squeeze and excitation u-net. *CoRR*, abs/1910.12206, 2019. 1

[24] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. 2

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 1, 3

[26] Y. Shi, Q. Li, and X. X. Zhu. Building extraction by gated graph convolutional neural network with deep structured feature embedding. In *IGARSS 2020 - 2020 IEEE International*

Geoscience and Remote Sensing Symposium, pages 3509–3512, 2020. 1, 2

[27] Leslie N. Smith and Nicholay Topin. Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates. *arXiv e-prints*, page arXiv:1708.07120, Aug. 2017. 4

[28] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 4278–4284. AAAI Press, 2017. 3

[29] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019. 3

[30] M. Tan, R. Pang, and Q. V. Le. Efficientdet: Scalable and efficient object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10778–10787, 2020. 2

[31] Alexey Trekin, German Novikov, Georgy Potapov, Vladimir Ignatiev, and Evgeny Burnaev. Satellite imagery analysis for operational damage assessment in emergency situations. *CoRR*, abs/1803.00397, 2018. 2

[32] Ning Wang, Yang Gao, Hao Chen, Peng Wang, Zhi Tian, Chunhua Shen, and Yanning Zhang. Nas-fcos: Fast neural architecture search for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2

[33] Xinlong Wang, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. Solov2: Dynamic and fast instance segmentation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17721–17732. Curran Associates, Inc., 2020. 2

[34] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995, 2017. 3

[35] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6022–6031, 2019. 4

[36] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017. 1

[37] Kang Zhao, Jungwon Kang, Jaewook Jung, and Gunho Sohn. Building extraction from satellite images using mask r-cnn with building boundary regularization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 2

[38] Teng Zhong, Zhixin Zhang, Min Chen, Kai Zhang, Zixuan Zhou, Rui Zhu, Yijie Wang, Guonian Lü, and Jinyue Yan. A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy*, 298:117132, 2021. 2

[39] Zilong Zhong, Zhong Qiu Lin, Rene Bidart, Xiaodan Hu, Ibrahim Ben Daya, Zhifeng Li, Wei-Shi Zheng, Jonathan Li, and Alexander Wong. Squeeze-and-attention networks for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2