BUILDING DAMAGE ASSESSMENT WITH DEEP LEARNING

S. May[1], A. Dupuis[1], A. Lagrange[2], F. De Vieilleville[2], C. Fernandez-Martin[2]*

[1] CNES, Toulouse, France, (stephane.may, aurore.dupuis)@cnes.fr
[2] Agenium Space, Toulouse, France (adrien.lagrange, francois.devieilleville, christine.fernandez-martin)@agenium.com

ICWG III/IVa

KEY WORDS: Deep learning, building detection, damage assessment, natural disaster, semantic segmentation, siamese network

ABSTRACT:

Global warming modifies the climate balance. Warming parameters are observed by many Earth Observation satellite systems, and the huge amount of data modifies the way to process them. This paper presents a few studies relative to damage detection on buildings, occurred during natural disasters. Recent advances in deep learning techniques are used for the building detection such as EfficientNet networks. Additional networks as Siamese models are used to evaluate the damage level with pre- and post-event images. Different techniques to merge detection masks are described and compared to a multiclass segmentation network. Results are presented and performances of the different solutions are compared.

1. INTRODUCTION

This paper presents a few studies relative to damage detection on buildings, occurred during natural disasters. Recent advances in deep learning techniques are used for the building detection, and additional networks are used to evaluate the damage level with before / after event images. Several neural networks architectures have been tested, and their evaluation is presented. Such techniques lead to interesting results for building damage assessment.

2. BACKGROUND

2.1 Natural disasters problem

Global warming modifies the climate balance. Warming parameters are observed by many Earth Observation satellite systems. Warming consequences are also monitored: among them, everyone can note the increase of natural disasters number. The population on Earth has also increased in the last century, with a major urbanization. As a consequence, the number of vulnerable areas has increased. In order to answer to the rising number of natural disaster events, the International Charter "Space and Major Disasters" (Charter) has been created. Many space agencies provide an important source of satellite imagery to respond to major natural and man-made disasters worldwide. The monitoring of major disasters and their impacts has become an issue.

2.2 Data volume problem

The number of events increased. The number of Earth Observation satellite (in particular VHR images) also led to an augmentation of the amount of data to be downloaded and processed. Building damage assessment is important to organize rescue operations. For now, damaged buildings are analyzed by photo-interpreters, which remains complex and time-consuming. For this reason, automatic solutions to detect and classify damaged buildings are welcomed. Recent advances in object recognition with deep learning algorithms offer new opportunities to do so.

2.3 How to define a damage ?

In order to offer consistent multi-scale building damage assessment, the first step is to define a proper standard nomenclature for damages. Several studies such as HAZUS, FEMA's Damage Assessment Operations Manual, the Kelman scale, and the EMS-98 propose a common ground on which it is possible to build a damage scale of four damage classes: no-damage, minor damage, major damage and destroyed. This nomenclature has been used to label the xView2 database.

2.4 Building segmentation

Building detection may be solved by semantic segmentation. Several architectures of convolutional networks may be used. (Maggiori et al., 2017) compares several architectures and shows that using Unet architecture with both encoder, decoder and skip-connections is the best way to obtain precise semantic segmentation.

Recent works based on this approach, try to refine detection results, focusing on building outlines. Some approaches as (Marmanis et al., 2017) combine images and digital surface models (DSM), and extract first building outlines and then introduces it into the segmentation network. (Wu et al., 2018) uses multi objects loss function, evaluating both segmentation and building outline maps, provided by two distinct networks. (Bischke et al., 2019) proposes an approach using SegNet with a VGG16 encoder where the masks are replaced by distance mask generated by computing for each pixel the distance to the outline. An interesting idea is also to add constraints on intermediate decoder layouts to make them look like the final segmentation at lower resolution, which permits to optimize multi scales descriptors. A similar multi-scale approach is also used in (Ji et al., 2018), and this work introduces Siamese networks working at their own different resolutions.

Enhancement of building segmentation is possible with modifications of skip-connections. (Li et al., 2018) proposes to add a compression step into skip-connections in order to reduce network weights. And (Yang et al., 2018) proposes to add an attention module into skip-connections to focus network attention on

---

* Corresponding author

the most informative parts of the image and to add bottleneck layout into the decoder.

Several neural networks architectures have been selected in this study, and trained for building segmentation purpose.

2.5 Damage classification

Damage evaluation may be considered as a classification task. The first idea is to use directly outputs of two building segmentation (images masks) and to define remaining, new or destroyed buildings between two satellite acquisitions. Other approaches are possible. There are a few references on damage classification, in particular on optical images. (Fujita et al., 2017) defines a CNN (VGG or Alexnet) that classifies building patches as destroyed or not, using several architectures as a unique network or Siamese networks, and integrating both pre and post images. (Ji et al., 2018b) proposes the same destroyed/not destroyed classification with SqueezeNet architecture and (Cao et al., 2018) with VGG16 architecture. (Gupta, 2019) introduces multi-classes damage classification, using first a Unet type network to identify buildings (segmentation) and then an architecture combining both a Resnet50 and a shallow network in parallel to classify damages.

3. PROPOSED METHOD

Our method is based on deep learning algorithms. Several satellite images with ground truth are used for the training. Two datasets are considered.

3.1 Used databases

3.1.1 xView2 dataset: The xView2 dataset is available freely since the end of 2019. This dataset, described in (Gupta, 2019), offers a wide variety of situations with 6 types of disasters : earthquake, flood, hurricane, fire, tsunami, volcano. These natural events occured in eleven countries. It includes both pre- and post-disaster Very High Resolution images (Ground Sampling Distance between 0.3 and 3 meters) for each disaster. Most importantly, this dataset provides a realistic situation with images taken from various satellites (GeoEye1, WorldView 2 and 3), sometimes old pre-image (maximum gap of 3 years) and a wide range of nadir angles (from 5 deg to 36 deg). This dataset appears perfectly suited to test the processing chains considered in this work. Classification ground truth contains 4 labels : no damage, minor damage, major damage, destroyed buildings. Figure 1, figure 2 and figure 3 are examples extracted from xView2 dataset.

Figure 1. Pre-disaster xView2 images

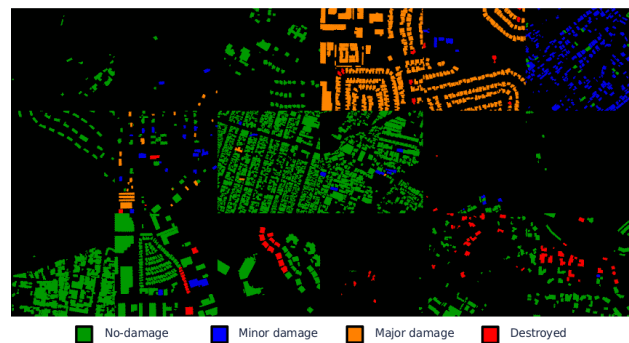

Figure 2. Post-disaster xView2 images

Figure 3. Ground truth for xView2 images

3.1.2 Haiti data: Thanks to the International Charter Space and Major disasters, many satellites images are provided by space agencies as CNES when natural events occur. Moreover, damages maps are produced by experts. For the evaluation, we considered the Matthieu hurricane that creates many damages on Haiti. Pleiades images are available on Haiti, before and after Matthieu huricane passage in october 2016. The images have been pansharpened (Panchromatic + Multispectral). They are used as validation data.

Ground truth associated to these images is a vector file, retrieved from Copernicus EMS R&R 050, containing manual labelling on building prints and damages classification between 3 classes: no damage, damaged, destroyed. Distinction between minor and major damages is not available, but this dataset is a great application case for our study.

Figure 4 presents pre and post Haiti images examples.

Figure 4. Haiti, pre and post Matthieur Huricane (oct 16), pansharpened Pleiades images

3.2 Building segmentation with deep learning

Among the different architectures, 4 neural networks are chosen for the damage detection and recognition. Evaluations are done for this networks.

3.2.1 Unet: The semantic segmentation is done with a Unet architecture (Ronneberger, 2015), which is based on a multi-resolution segmentation network, with an encoder part and a decoder part. The Unet network has proven its strength in a high number of image processing configurations with deep learning. The encoder extracts features (convolutional layers) while reducing the spatial resolution. The decoder sequentially increases the resolution and, merges information from the previous layer and from the skip-connection coming from the encoder layer at the same resolution. Several Unet architecture with different encoders have been compared.

3.2.2 MobileNet V2: Another neural network architecture is considered to compare with previous Unet network. (Sandler, 2019) introduces MobileNet architecture : this architecture's particularity is the use of depthwise convolution, which allows to reduce computation cost with equal performance. A depthwise convolution consists in applying a convolution per canal and then applying classic 2D convolution of size $1x1$ (pointwise convolution instead of usual $3x3$ spatial convolution). When $H.W.C.C.n.n$ operations are needed for a $H.W.C$ block 's convolution, only $H.W.C.n.n + H.W.C$ operations are needed with depthwise convolution. (Howard, 2019) introduces MobileNetV2 architecture. This architecture uses a new block type, called mobile inverted bottleneck, in replacement of Convolution+ReLu block. Mobile inverted bottleneck consists in applying : firstly a $1x1$ convolution allowing features space expansion, secondly an activation such as ReLu, thirdly a depthwise convolution (one convolution per feature), then another ReLu activation, and finally a 1x1 convolution in order to reduce features number. In addition, a skip connection can link input and output blocks (same size), to reduce loss due to ReLu activation.

3.2.3 SE-ResNet: Among the interesting solutions, we consider the SE-ResNet network. (He, 2016) introduces ResNet architecture which aims at modeling convolution filter nearby identity in order to transfer information between layers without having small loss gradient during learning phase. To do so, it uses skip-connections allowing information to bypass convolutions.(Hu, 2019) introduces Squeeze-Excite blocks which aims to balance weights between channels, and to increase receptive field size. (Xie, 2017) improves ResNet network, introducing Squeeze-Excite blocks, to define SE-ResNet architecture. We can notice that Squeeze-Excite blocks is likely an average on the features which could lead to smooth effects. As it introduces a dependence to a large part of the image, if this model is used to infer the image with a tile processing, depending on the margins, discontinuities near tiles borders are visible.

3.2.4 EfficientNet: (Tan, 2019) proposes to integrate mobile inverted bottleneck, Squeeze-Excite and depthwise convolution into an architecture as light as possible. It also presents a method allowing to adapt the architecture hyperparameters in order to point the best compromise between performance and weight. This adaptation consists in modifying both resolution, depth and features numbers. This high performance network is considered for this study.

3.3 Damages classification with deep learning

For this second way, damages are classified with the neural network. Three approaches have been compared : direct use of the building segmentation associated with an image classifier, Siamese network to achieve segmentation with 5 classes (4 damages classes + background) and several binary networks

that generates 4 semantic segmentation, and are then merged afterwards to build the 5 classes map.

3.3.1 Decision based on building semantic segmentation: After the building semantic segmentation step, building mask are computed. For each building detected, tiles centered on this object are extracted from pre- and post-event images. Then, a classifier is used to decide the type of damages. This approach presents a major drawback : we need a perfect building segmentation which distinguish buildings individually. UNet-EfficientNet architectures, which got the best performance for building segmentation, have been chosen. Their performance is good enough for building detection, but not sufficiently precise to separate each building individually. So the ground truth of building is used to extract separate tiles. Then the classification network is trained.

3.3.2 Siamese segmentation network: It consists in implementing two parallel encoders : one to process to pre-disaster image, and one to process post-disaster image. These two encoders are identical and share the same weights. The two features outputs are then merged (by concatenation) and used as input in the decoder part. All parallel outputs of the skip-connections are also concatenated and used as input of the decoder.
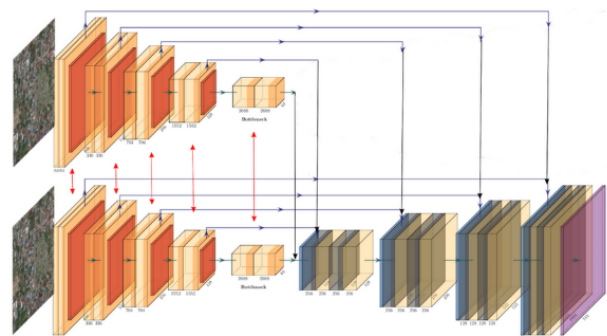
Figure 5. Architecture of Siamese network

3.3.3 Binary siamese segmentation networks: Is it more efficient to train one multiclass network or to train several binary networks ? In order to answer this question, several binary Siamese networks "all vs one" (one per class) have been implemented with EfficientNet-B0 as encoder. Then, a post treatment retrieves all results and defines the final multi-classes result, as defined in the following schema:

Neural networks are trained following classical backpropagation procedure. A Stochastic Gradient Descent is used. Tiles with 224x224 pixels are sent to the network. Batch size is equal to 32. Each class is represented into the batch with an approximate ratio: 1 tile with background and no building, 2 tiles with no-damage, 9 with minor damages, 9 with major damages, and 9 tiles with destroyed buildings. The global loss includes a balanced cross-entropy and dice loss. 200 epochs are computed. The dropout rate is equal to 0.01. Data augmentations are done on the input images as vertical and horizontal flips, light zoom into the range $[0.9; 1.1]$, some rotations (random value into $[0; 360°]$), and modifications of brightness into the range $[0.8; 1.2]$.
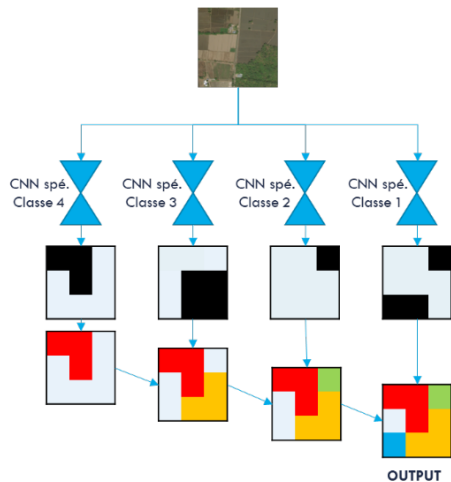
Figure 6. Architecture with fusion of 4 specialized binary networks

4. RESULTS

In this part, results are given for the building semantic segmentation and damages detection.

4.1 Building segmentation

This study has been focused on U-Net type architecture for building segmentation part. Evaluations have been run on the xView2 test database. As usual, test database has not been used during learning phase. But the test database is related to the same disasters and it is extracted from the same original dataset as the learning database.

Model	Parameters number (millions)	Precision	Recall	Score F1
MobileNet-V2	5.1	0.702	**0.887**	0.784
SE-ResNet-101	62.3	0.78	0.822	0.801
EfficientNet-B0	8.6	0.771	0.859	0.813
EfficientNet-B5	59	**0.805**	0.866	**0.834**

Table 1. Synthesis of building segmentation results

Among the different architectures, Unet-EfficientNet-B5 rises the best performance, with F1-score up to 83%. Unet-EfficientNet networks performance is confirmed on this use case. We can notice that Unet-EfficientNet-B0 got better perfomance than Unet-SE-Resnet-101 while it got only 8.6M parameters.

The largest studied network EfficientNet-B5 gives the best results, but the computation time for the training is larger, higher occupied memory into the Graphical Processing Unit. Also the inference time and computation cost is larger with such network compared to EfficientNet-B0 and MobileNet-V2.

4.2 Damages classification

The results presented here are computed for all pixels of the images, whereas in the xView2 challenge, results are computed at pixel level on buildings area only.

Model	No damage	Minor damage	Major damage	De-stroyed
EffNet-B0	**0.725**	**0.548**	**0.587**	**0.729**
EffNet-B2	0.717	0.535	0.577	0.713

Table 2. Synthesis of damage classification results (F1 score)

4.2.1 Classification using building segmentation: The first damage detection network classifies each couple of tiles centered of buildings. Results of classification networks are given in table 2.

Despite its lower number of parameters (5.3M related to 9.2M), the EfficientNet-B0 network gives better results than the EfficientNet-B2. In order to evaluate the performance of the classification network, evaluation is done considering a perfect building detection : consequently, the ground truth of building is used to create each tiles. The buildings ground truth contains well separated buildings map contrary to the output map of the MobileNet or EfficientNet semantic segmentation with fused buildings. Global performance results should also take into account previous building segmentation model errors.

4.2.2 Siamese network using pre and post images: UNet-EfficientNet have been used as encoder in the siamese network. It leads to the results (F1 score) in table 3.

Model	No damage	Minor damage	Major damage	De-stroyed
EffNet-B0	**0.790**	**0.412**	**0.635**	**0.702**
EffNet-B5	0.556	0.097	0.222	0.609

Table 3. Siamese networks results

UNet-EfficientNet-B0 results are the best and satisfying, except for minor damages class. UNet-EfficientNet-B5 evaluation suffers from machine limitations that conduct to a batch size limited to 2. This explains why the EfficientNet-B5 results are not satisfying. These results are computed for all pixel of the building semantic segmentation, and so integrate building detection errors.
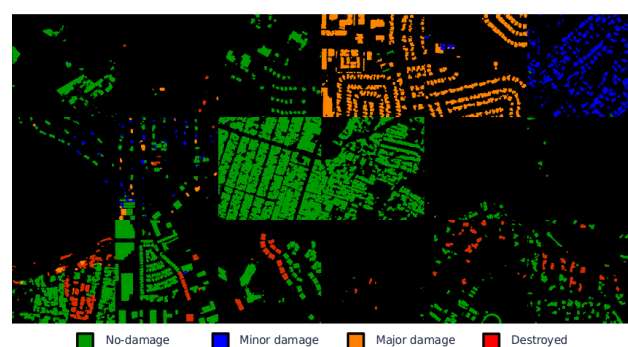

Figure 7. Results with siamese network

4.2.3 Binary siamese networks: Several fusion methods have been tested, and compared to multiclasse. It leads to the results given in table 5.

For each cascade network, only pixels that have not be classified previously are classified by this network.

These results show firstly the difficulty to distinguish minor an major damages. Secondly, it is interesting to note that using

id	Description
1	EfficientNet-B0 (multiclass)
2	Binary DNN: no damage → destroyed → major → minor
3	Binary DNN: no damage → destroyed → major + minor + major and minor separation afterwards
4	EfficientNet-B0 (multiclass) major and minor damages classes are fused

Table 4. Description of fusion policies

Model	No damage	Minor damage	Major damage	De-stroyed
1	0.790	0.412	0.635	0.702
2	0.805	0.328	0.538	0.635
3	0.805	0.603		0.691
5	0.790	0.624		0.702

Table 5. Binary siamese networks results

several binary networks and merge them, does not increase performance compared to a multiclass network. The difficulty of the multiclass network training is ensuring that all classes are well represented by samples into each batch or into each epoch. Moreover, classes weights shall be taken into account into the cross-entropy loss.

4.3 Validation

Building detection validation tests on Haiti data have been realized with the 5-class segmentation model UNet-EfficientNet-B5. Figure 8, figure 9 and figure 10 present an example.

Figure 8. Haiti validation image

The output of building detection has a good quality and may help image analyst to achieve rapid mapping operations.

The evaluation of building damages is also interesting. Some area with major damages in red contrasts with the other building are. It gives a preliminary idea of where the hurricane created more damages, with an automatic processing chain.

5. CONCLUSION

In this work, several architectures of neural networks have been trained to evaluate damages between pre-disaster and post-disaster images. The class "minor damage" appears to be the more complicated to learn. So an approach with specialized networks is a potential solution to alleviate this problem. Some improvements have been studied, as merging the results of damage
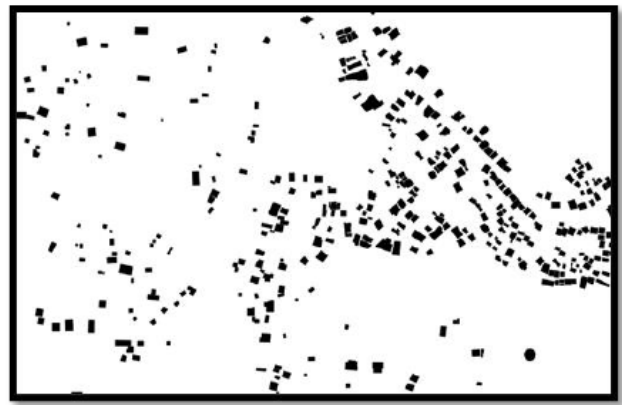

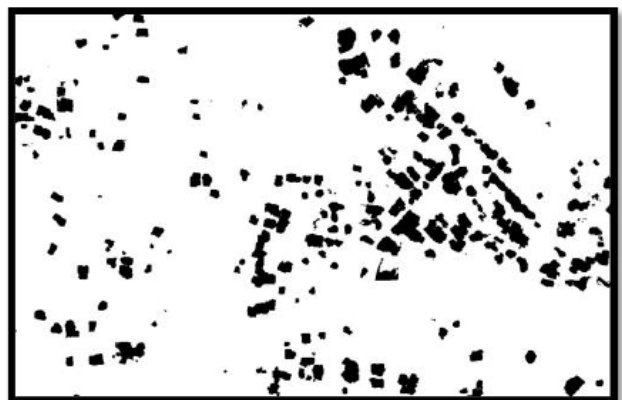Figure 9. Haiti building detection ground truth
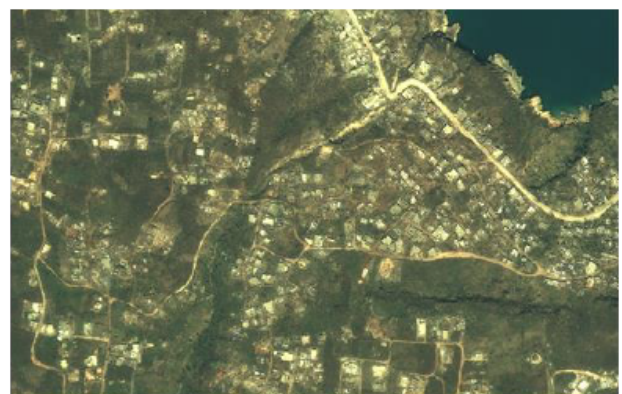

Figure 10. Haiti building detection

Figure 11. Haiti validation image 2

segmentation with a preliminary building segmentation (made only with the pre-disaster image). Several proposed solutions give interesting results for photo-interpreters.

REFERENCES

[Bischke et al., 2019] Bischke, Benjamin, Patrick Helber, Joachim Folz, Damian Borth, et Andreas Dengel, 2019. Multi-task learning for segmentation of building footprints with deep neural networks. *IEEE International Conference on Image Processing (ICIP)*, 14801484.
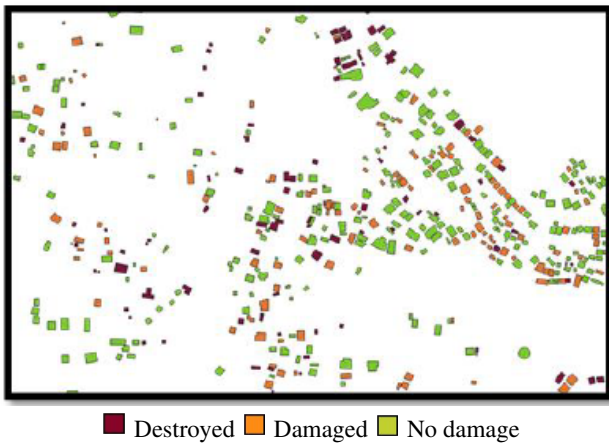
[Cao et al., 2018] Cao, Quoc Dung, et Youngjun Choe, 2018.

■ Destroyed ■ Damaged ■ No damage

Figure 12. Haiti damages classification ground truth

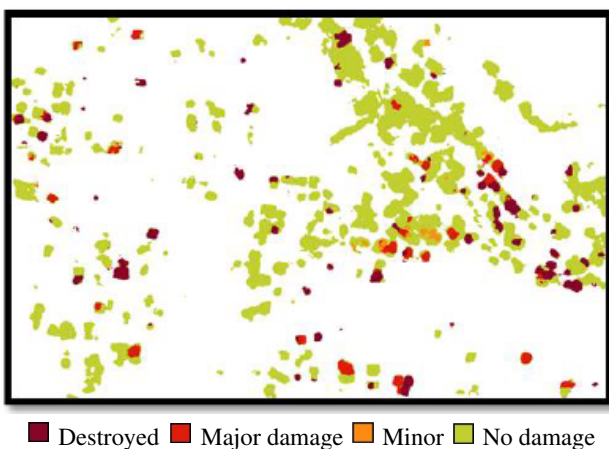■ Destroyed ■ Major damage ■ Minor ■ No damage

Figure 13. Haiti damages classification

Deep learning based damage detection on post-hurricane satellite imagery. arXiv:1807.01688 [cs.CV].

[Fujita et al., 2017] Fujita, Aito, Ken Sakurada, Tomoyuki Imaizumi, Riho Ito, Shuhei Hikosaka, et Ryosuke Nakamura, 2017. Damage detection from aerial images via convolutional neural networks, *Fifteenth IAPR International Conference on Machine Vision Applications* (MVA), 58.

[Gupta, 2019] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, Matthew Gaston, 2019. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. arXiv:1911.09296 [cs.CV].

[He, 2016] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, et Jian Sun. Deep residual learning for image recognition, 2016. *IEEE conference on computer vision and pattern recognition*, 77077.

[Howard, 2019] Howard, Andrew, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, et al. Searching for MobileNetV3 . arXiv:1905.02244 [cs.CV].

[Hu, 2019] Hu, Jie, Li Shen, Samuel Albanie, Gang Sun, et Enhua Wu, 2019. Squeeze-and-Excitation Networks. arXiv:1709.01507 [cs.CV].

[Ji et al., 2018] Shunping Ji, Shiqing Wei, et Meng Lu, 2018. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing* 57, no 1 : 574586.

[Ji et al., 2018b] Ji, Min, Lanfa Liu, et Manfred Buchroithner, 2018. Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 Haiti earthquake. *Remote Sensing* 10, no 11: 1689.

[Li et al., 2018] Lin Li, Jian Liang, Min Weng, et Haihong Zhu, 2018. A multiple-feature reuse network to extract buildings from remote sensing imagery. *Remote Sensing* 10, no 9: 1350.

[Maggiori et al., 2017] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, 2017. High-resolution semantic labeling with convolutional neural networks, *IEEE TGRS*, vol.55/12.

[Marmanis et al., 2017] Dimitrios Marmanis, Konrad Schindler, Jan Dirk Wegner, Silvano Galliani, Mihai Datcu, et Uwe Stilla, 2017. Classification With an Edge: Improving Semantic Image Segmentation with Boundary Detection. arXiv:1612.01337 [cs.CV].

[Ronneberger, 2015] Ronneberger, Fischer, Brox, 2015. U-Net : Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV].

[Sandler, 2019] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, et Liang-Chieh Chen, 2019. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381 [cs.CV].

[Smith, 2017] Smith, Leslie N, 2017. Cyclical Learning Rates for Training Neural Networks. arXiv:1506.01186 [cs.CV].

[Tan, 2019] Tan, Mingxing, et Quoc V. Le., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946 [cs.LG].

[Wu et al., 2018] Guangming Wu, Zhiling Guo, Xiaodan Shi, Qi Chen, Yongwei Xu, Ryosuke Shibasaki, et Xiaowei Shao, 2019. A boundary regulated network for accurate roof segmentation and outline extraction. *Remote Sensing* 10, no 8: 1195.

[Xie, 2017] Xie, Saining, Ross Girshick, Piotr Dollr, Zhuowen Tu, et Kaiming He, 2017. Aggregated residual transformations for deep neural networks. *IEEE conference on computer vision and pattern recognition*. arXiv:1611.05431 [cs.CV].

[Yang et al., 2018] Yang, Hui, Penghai Wu, Xuedong Yao, Yanlan Wu, Biao Wang, et Yongyang Xu, 2018. Building extraction in very high resolution imagery by denseattention networks. *Remote Sensing* 10, no 11: 1768.