environmental sciences proceedings

MDPI

Article

Buildings' Classification Using Very High-Resolution Satellite Imagery

Ali J. Ghandour [1]*, Mohammad Dimassi [2], Hasan Nasrallah [1], Mohammad Zaraket [3], Jamal Haydar [3] and Abed Ellatif Samhat [2]

[1] National Center for Remote Sensing - CNRS, Lebanon;
[2] Lebanese University;
[3] Islamic University of Lebanon;
* Corresponding author: aghandour@cnrs.edu.lb

Abstract: Buildings' classification using satellite images is gaining importance for several applications, such as damage assessment, resource allocation, and population estimation. This work focuses on two specific classification tasks: building-type classification (residential or non-residential) and building-damage assessment (damaged or not-damaged). These two tasks have recently become of great interest, especially considering current global conflicts and natural disasters, such as the Ukrainian conflict, the 2023 Turkey–Syria earthquake, and the recent tragic flood in Libya's Derna. Existing building-type classification approaches combine optical satellite images with LIDAR data or street view images. We propose here to rely solely on RGB satellite images and follow a 2-stage deep learning-based approach, where buildings' footprints are extracted using a semantic segmentation model, followed by the classification of the segmented images. To the best of our knowledge, this is the first Building-Type Classification (BTC) method solely based on RGB aerial imagery. Optical imagery is cost-effective and widely available and can be tasked and acquired quickly following a military conflict or natural disaster. Supervised deep learning algorithms rely on high-quality labeled datasets. Four Building-Damage Assessment (BDA) datasets are considered, analyzed, and compared to choose the appropriate one for our research. For the scope of this work, we define a damaged building as a partially or wholly demolished building that might result from armed conflicts, earthquakes, tornadoes, and hurricanes. On the other hand, due to the need for an appropriate residential/non-residential building-type classification dataset, we introduce a new high-resolution satellite imagery dataset called the Beirut Buildings Type Classification (BBTC) dataset. We conduct experiments to select the best hyper-parameters and model architecture and propose an extra transfer learning stage that outperforms the classical method. Finally, we validate the performance of the proposed schemes, where BTC achieved a 94.8% accuracy using RexNet as a backbone, and BDA scored 97.3% accuracy using Focal loss and Adam optimizer. With the extra transfer learning stage, BDA improved in the last percentages of the accuracy and F1-score to reach 98.96% and 99.4%, respectively.

Keywords: Building-Type Classification; Building-Damage Assessment; Hyper-parameter tuning

Citation: Ghandour A. J.; Dimassi M.; Nasrallah H.; Zaraket M.; Haidar J.; Samhat A. Buildings' Classification Using Very High-Resolution Satellite Imagery. *Environ. Sci. Proc.* **2023**, *1*, 0. https://doi.org/

Published:

1. Introduction

Buildings classification is vital for many applications, such as buildings damage assessment (BDA) and buildings type classification (BTC). Urban areas are constantly struck by man-made and/or natural disasters, such as wars, tornadoes, and earthquakes, resulting in large-scale buildings and urban infrastructure destruction. In the early reconstruction phase, damage assessment is performed manually using crucial information about the area, amount, rate, and type of damage. In addition, buildings type classification (*e.g.*, residential/non-residential) pave the way for many real-world applications such as population estimation and resource allocation.

For these purposes, remote sensing techniques can play an important role, mainly due to their wide availability at relatively low cost, wide field of view, and fast response capacities. Using deep learning to build classification can speed up the process by reducing human intervention and saving considerable time and cost.

Unlike other work that proposes to tackle such a task by adopting a model for semantic segmentation with many classes [1,2], here we follow a 2-stage approach where we disentangle the semantic segmentation from the classification. In a nutshell, a semantic segmentation model takes an input RGB image and predicts buildings' masks (*i.e.*, stage 1). In the second stage, buildings are cropped from the original image and fed to a classification model to predict the class of each building. Also, we propose using only RGB satellite images, which is more efficient than using other additional modalities. Optical images are cost-efficient and widely available, and can be tasked and acquired quickly after a military conflict or natural disaster.

This paper focuses on the second classification stage and conducts experiments to find the best hyper-parameters and model architecture for the underlying task. Moreover, we propose an extra transfer learning stage that extends the classical fine-tuning approach by an additional stage that makes the model's layers more consistent and specific to the downstream task. Finally, we propose a new dataset for BTC. For space limitations, we showed hyper-parameters optimization results for BDA while we focused on the architecture design for BTC.

The rest of this paper is organized as follows: Section 2 reviews some of the related works. Section 3 presents a brief background and details the methodology adopted in this work. In Sections 4 and 5, we validate the proposed approach using Building Damage Assessment and Building Damage Assessment applications, respectively, and discuss the proposed dataset. Finally, Section 6 concludes this manuscript.

2. Related Work

2.1. Buildings Type Classification

In [3], the authors use an ensemble of machine learning models to classify buildings as sprayable and non-sprayable based on characteristics of the buildings such as size, shape and proximity to neighboring features. Similarly, in [4], authors use classical machine learning approaches to predict if the building is residential or not based on several input variables. Work presented in [5] propose to classify buildings type using geospatial data (*e.g.*, point-of-interest (POI) data, building footprints, land use polygons, and roads) based on NLP and ratio-based techniques. In [6], an iterative clustering method is introduced to classify buildings based on spatiotemporal data (*e.g.*, population density and people interaction). The random forest classifier is used in [7] to classify the footprint of buildings from different data sources (e.g., topographic raster maps, cadastral databases, or digital landscape models). The authors in [8] use object-Based Image Analysis (OBIA) and machine learning methods to extract and classify buildings from Airborne Laser Scanner (ALS). Gaussian finite mixture model is proposed in [9] to classify buildings based on several metrics extracted from high-resolution satellite images. Finally, [10] combines street view with satellite images to classify buildings using CNN models.

To the best of our knowledge, no method in the literature is solely based on RGB aerial images to classify buildings types.

2.2. Buildings Damage Assessment

Authors in [11] assess the impact of the combined use of different resolution satellite images on improving classification accuracy of damaged buildings using CNN.In [12], the proposed method detects Flooded/Damaged buildings from satellite imagery of an area affected by a hurricane. The study in [13] proposes an algorithm for building damage detection from post-event aerial imagery, using a data expansion Single Shot multi-box Detector (SSD) algorithm for a small data set of Hurricane Sandy. In [14], the authors use

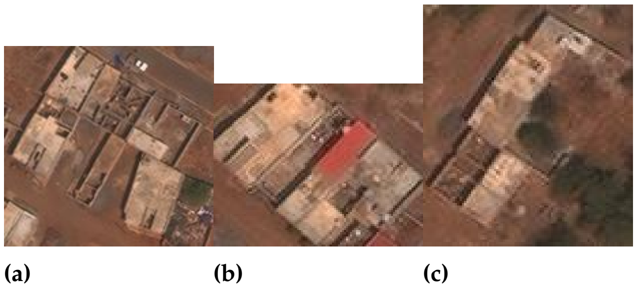


Figure 1. Sample damaged buildings scenes from the xView dataset.

CNN on a small set of candidates damaged buildings to reduce the required processing time. Deep learning is also used in [15] to improve the detection of rooftop hail damage.

3. Methodology

In this section, we will focus on buildings classification where the model takes an RGB image of a building and outputs the class of the building (damaged/non-damaged; residential/non-residential).

Multiple loss functions exist and can be used during training. In this work, we focus on two of the widely used losses in order to compare their performance and choose the appropriate one for the task: (*i*) Binary Cross-Entropy loss and (*ii*) Focal loss [16].

During CNN training, the role of the optimizer is to update the weight parameters to minimize the loss function. Multiple optimizers are suggested in the literature, such as Momentum, Regular Gradient Descent (RGD), Stochastic Gradient Descent (SGD), Adam, and RectifierAdam for classification tasks [17]. Although non-adaptive optimizers, such as SGD, help obtain better minima and generalization properties. Adam optimizer is widely used, as it leads to faster convergence due to its adaptive learning rate. However, Adam suffers from significant variance at the beginning of training. Rectifier-Adam (also known as RectAdam) [18] was proposed to improve the convergence of Adam. In a nutshell, no optimizer works best for all the applications; thus, we propose a comparison of these optimizers in this work.

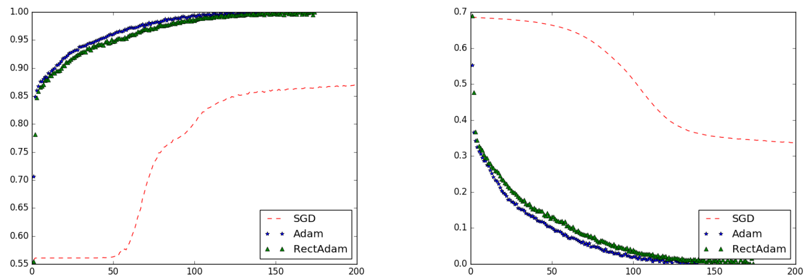
3.1. Transfer Learning

Transfer learning is considered an innovative approach in CNN training to achieve high accuracy with minimal time and effort, relying on the existing pre-trained model. The idea is to use a model trained on a large dataset and transfer its knowledge to the application at hand. The classical transfer learning approach is as-follows: (*i*) a model is trained from scratch on a large dataset (*i.e.*, pre-training on ImageNet [19]), then (*ii*) the last layers of the model, are retrained for the downstream application. Here, we propose to extend this approach by adding a third and final stage: (*iii*) we freeze the last layers of the network and re-train initial layers (previously frozen) in order to ensure the consistency between those layers and the last ones, and also to adjust all the weights according to the specific target application.

4. Buildings Damage Assessment

4.1. Dataset

We considered four different datasets as potential candidates for the BDA application: ABCD Dataset [20], Flooding Dataset [12], NOAA Dataset [21] and xView Dataset [22]. We used several parameters for the datasets benchmark where "damage type" was the critical metric. "Damage type" differs based on the disaster event: war, earthquakes, and wind (hurricanes and tornadoes) usually result in completely or partially demolished and ruined buildings. Flooding produces water in areas where it is not normally expected,



(a) **(b)**

Figure 2. (a) and (b) show the accuracy and loss, respectively, in function of time for BDA: the model is trained with SGD, Adam and RectAdam optimizers. The model trained with SGD struggles to converge. Adam optimizer leads to faster convergence.

Dataset	Number of Objects	Resolution (in pixels)	Format	Generalizable
ABCD dataset [20]	8,500	128x128	.tif	No
Flooding dataset [12]	10,000	128x128	.jpeg	No
NOAA dataset [21]	500,000	9351x9351	.tif	No
xView dataset [22]	1,000,000	3000x3000	.tif	Yes

Table 1. Datasets survey and comparison.

while a volcano erupts lava. For the scope of this work, we define a damaged building as a partially or completely demolished building. This definition covers the most common type of damage resulting from armed conflict, earthquakes, tornadoes, and hurricanes.

Table 1 shows a comparison between these four datasets. One can conclude that xView dataset is the best candidate for our research. xView dataset is well labeled and large enough. Figure 1 shows sample damaged scenes from the xView dataset.

4.2. Experimental Results

In this section, we present the results for the different scenarios discussed in Section 3 based on the xView dataset. For each scenario, we rely on a CNN network made from three layers of Convolutions + MaxPooling followed by the Flatten, Dense, and finally Softmax layer. In our work, we used the Sigmoid activation function for each convolution layer and the ReLU activation function for the dense layer. We carried out an evaluation study of the test data. The convergence condition is to reach a high accuracy value (99%) and low loss threshold (≤ 0.001) in trained data.

We trained two models using cross-entropy loss and focal loss until they converged. Both models converge to 99%, but focal loss converges faster during training (127 vs. 138 epochs). Also, focal loss scores slightly higher accuracy over the testset (97.56% vs 97.33%), and thus it will be adopted as part of the following scenario.

No single optimizer choice can be made for all different CNN downstream applications. We compared three optimizers: (*i*) SGD, (*ii*) Adam and (*iii*) RectAdam. Figure 2a and b show the accuracy metric and the loss metric, respectively, as a function of time (number of epochs) for SGD, Adam, and RectAdam optimizers. The results reveal that SGD struggled to converge. In terms of accuracy, Adam and RectAdam score a better result (97. 3% and 97. 2%, respectively) than SGD (85.1%). Regarding the number of epochs needed to converge, Adam is much faster than RectAdam (137 vs. 174 epochs). We will adopt Adam optimizer for the damaged building classification application in the following scenarios.

	Epochs	Accuracy (%)
TL-VGG16	**23**	96.12
E-TL-VGG16 (ours)	69 (23 + 46)	**98.96**
Baseline	127	97.33

Table 2. Comparison of different trained paradigms for BDA; ours (E-TL-VGG16) is the best in terms of accuracy and faster to train than the baseline.

Transfer Learning Results

We considered the following three scenarios:

- Scenario 1: network trained from scratch without using any pre-trained weights. We will refer to this model as Baseline model.
- Scenario 2: network trained based on transfer learning approach using VGG16 pre-trained weights on ImageNet dataset [23]. We freeze all layers and train only the last layers of the network. We will refer to this model as the TransferLearning-VGG16 (TL-VGG16) model.
- Scenario 3: extra transfer leaning stage (discussed in Section 3.1) where an additional final step is added. We will refer to this model as the Extra-TransferLearning-VGG16 (E-TL-VGG16) model.

Table 2 clearly shows the difference in the convergence time, where TL-VGG16 iterates only for 23 epochs to converge, while the Baseline model needed 127 epochs. However, this does not result in higher accuracy where TL-VGG16 achieved 96.12% accuracy over the test data, which is lower than the 97.33% scored by the Baseline model.

The rationale behind the E-TL-VGG16 model is to improve the accuracy of the transfer learning approach while maintaining good enough convergence speed. At epoch 23, when we un-freeze the VGG16 layers and de-freeze the last layers, accuracy metric sharply decreases, and loss sharply increases. After this glitch in performance, the network spends 69 epochs to re-converge to previously attained thresholds. E-TL-VGG16 performance tabulated in Table 2 reveals an important improvement in the last percentages of accuracy to reach 98.96%. Finally, we also compute the F1-score achieved by E-TL-VGG16, which turns out to be 99.4%.

5. Buildings Type Classification

To the best of our knowledge, there is no high-resolution satellite image data set to classify buildings into residential / non-residential. We introduce a new "Beirut Buildings Type Classification" (BBTC) dataset to help the community develop this field further. BBTC dataset consists of 17,033 RGB images, where each image contains a single building and is annotated as residential (15,330 buildings) or non-residential buildings (1,703 buildings, which represents 10% of the annotated objects). The tiles are extracted from 50-cm optical images that cover the city of Beirut. Some examples can be seen in Figure 3.

BBTC dataset was created and labeled using the following routine: (*i*) buildings polygons are extracted from OpenStreetMap as a shapefile then (*ii*) annotation is done using QGIS software where the shapefile is overlayed on two basemaps for visual confirmation (Bing Virtual Earth and Google Satellite Hybrid) and each building is labeled as either residential or non-residential. Non-residential objects include universities, schools, government facilities, mosques, churches, commercial centers, and industrial facilities. (*iii*) Then, 50-cm resolution tile for each building is cropped (slightly larger than the corresponding polygon to account for any inaccuracy) and saved. Duplicate buildings were ignored, reaching a total of 17,033 objects. BBTC dataset can be accessed through the following GCP bucket: https://storage.googleapis.com/bbtc/bbtc_dataset.tar.gz

Figure 3. Sample of buildings images from the proposed BBTC dataset.

Model	Avg. time (per epoch)	Val. Acc	Epoch	Test Acc
VGG16	3m20s	0.953	95	0.935
ResNet 50	5m30s	0.947	44	0.933
Res2Net	9m30s	0.960	**40**	0.943
EfficientNet	**3m13s**	0.960	49	0.941
RexNet	3m15s	**0.962**	99	**0.948**

Table 3. Ablation study for BTC on the proposed dataset; We compare different architectures. RexNet outperforms all other models in terms of accuracy.

5.1. Experimental Results

We compared 5 different backbones including state-of-the-art ones; VGG16, ResNet50 [24], Res2Net [25], EfficientNet-B0 [26] and RexNet [27]. Table 3 shows that RexNet gives the best results in terms of accuracy in the test set (94.8%), while average time per epoch (3minutes 15 seconds) is almost equal to EfficientNet. RexNet is a newly proposed architecture where the authors claim to provide high accuracy while achieving a low time complexity comparable to EfficientNet.

Finally, we adopt the winning model (*i.e.*, RexNet) and train it using different optimizers. From Table 4, one can notice that the SGD optimizer was able to converge, in contrast to the previous performance witnessed in Table **??**. SGD achieved the best accuracy on the test set (94.8%) compared to Adam and RectAdam, although it is the slowest with convergence time equal to 99 epochs.

Optimizer	Avg. time (per epoch)	Val. Acc	Epoch	Test Acc
SGD	3m15s	**0.962**	99	**0.948**
Adam	3m15s	0.957	90	0.945
RectAdam	3m15s	0.958	**68**	0.947

Table 4. Ablation study for BTC on the proposed dataset; RexNet is trained with several optimizers. SGD leads to the best accuracy, while RectAdam leads to faster convergence.

6. Conclusions

In this paper, we focus on buildings classification using only RGB satellite images. We conducted experiments to find the best hyperparameters, backbone, and we proposed an extra transfer learning stage. Due to the lack of an appropriate residential/non-residential buildings classification dataset, we proposed the Beirut Buildings Type Classification (BBTC) dataset. We validated the proposed approach on two applications: Buildings Damage Assessment and Buildings Type Classification over the xView and BBTC datasets.

Funding: This research received no external funding.

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Iglovikov, V.; Mushinskiy, S.; Osin, V. Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *arXiv preprint arXiv:1706.06169* **2017**.
2. Guérin, E.; Oechslin, K.; Wolf, C.; Martinez, B. Satellite Image Semantic Segmentation. *arXiv preprint arXiv:2110.05812* **2021**.
3. Sturrock, H.J.; Woolheater, K.; Bennett, A.F.; Andrade-Pacheco, R.; Midekisa, A. Predicting residential structures from open source remotely enumerated data using machine learning. *PloS one* **2018**, *13*, e0204399.
4. Lloyd, C.T.; Sturrock, H.J.W.; Leasure, D.R.; Jochem, W.C.; Lázár, A.N.; Tatem, A.J. Using GIS and Machine Learning to Classify Residential Status of Urban Buildings in Low and Middle Income Settings. *Remote Sensing* **2020**, *12*. https://doi.org/10.3390/rs12233847.
5. Chen, W.; Zhou, Y.; Wu, Q.; Chen, G.; Huang, X.; Yu, B. Urban Building Type Mapping Using Geospatial Data: A Case Study of Beijing, China. *Remote Sensing* **2020**, *12*. https://doi.org/10.3390/rs12172805.
6. Zhuo, L.; Shi, Q.; Zhang, C.; Li, Q.; Tao, H. Identifying Building Functions from the Spatiotemporal Population Density and the Interactions of People among Buildings. *ISPRS International Journal of Geo-Information* **2019**, *8*. https://doi.org/10.3390/ijgi8060247.
7. Hecht, R.; Meinel, G.; Buchroithner, M. Automatic identification of building types based on topographic databases – a comparison of different data sources. *International Journal of Cartography* **2015**, *1*, 18–31, [https://doi.org/10.1080/23729333.2015.1055644]. https://doi.org/10.1080/23729333.2015.1055644.
8. Belgiu, M.; Tomljenovic, I.; Lampoltshammer, T.J.; Blaschke, T.; Höfle, B. Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data. *Remote Sensing* **2014**, *6*, 1347–1366. https://doi.org/10.3390/rs6021347.
9. Jochem, W.C.; Leasure, D.R.; Pannell, O.; Chamberlain, H.R.; Jones, P.; Tatem, A.J. Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science* **2021**, *48*, 1161–1179, [https://doi.org/10.1177/2399808320921208]. https://doi.org/10.1177/2399808320921208.
10. Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Building instance classification using street view images. *ISPRS journal of photogrammetry and remote sensing* **2018**, *145*, 44–59.
11. Duarte, D.; Nex, F.; Kerle, N.; Vosselman, G., Satellite Image Classification Of Building Damages Using Airborne And Satellite Image Samples In A Deep Learning Approach. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2 ed.; International Society for Photogrammetry and Remote Sensing (ISPRS), 2018; Vol. IV, pp. 89–96. https://doi.org/10.5194/isprs-annals-IV-2-89-2018.
12. Cao, Q.D.; Choe, Y. Deep Learning Based Damage Detection on Post-Hurricane Satellite Imagery, 2018.
13. Li, Y.; Hu, W.; Dong, H.; Zhang, X. Building Damage Detection from Post-Event Aerial Imagery Using Single Shot Multibox Detector. *Applied Sciences* **2019**, *9*. https://doi.org/10.3390/app9061128.
14. Nex, F.; Duarte, D.; Steenbeek, A.; Kerle, N. Towards Real-Time Building Damage Mapping with Low-Cost UAV Solutions. *Remote Sensing* **2019**, *11*. https://doi.org/10.3390/rs11030287.
15. Hezaveh, M.M.; Kanan, C.; Salvaggio, C. Roof Damage Assessment using Deep Learning. In Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2017, pp. 6403–6408. https://doi.org/10.1109/AIPR.2017.8457946.
16. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.
17. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* **2016**.
18. Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the Variance of the Adaptive Learning Rate and Beyond. In Proceedings of the International Conference on Learning Representations, 2019.
19. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
20. Fujita, A.; Sakurada, K.; Imaizumi, T.; Ito, R.; Hikosaka, S.; Nakamura, R. Damage detection from aerial images via convolutional neural networks. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), 2017, pp. 5–8. https://doi.org/10.23919/MVA.2017.7986759.
21. Hurricane Harvey: Emergency Response Imagery of the Surrounding Regions. https://storms.ngs.noaa.gov/storms/harvey/index.html.
22. Lam, D.; Kuzma, R.; McGee, K.; Dooley, S.; Laielli, M.; Klaric, M.; Bulatov, Y.; McCord, B. xview: Objects in context in overhead imagery. *arXiv preprint arXiv:1802.07856* **2018**.
23. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
25. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2021**, *43*, 652–662. https://doi.org/10.1109/TPAMI.2019.2938758.
26. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning. PMLR, 2019, pp. 6105–6114.

27. Han, D.; Yun, S.; Heo, B.; Yoo, Y. Rethinking Channel Dimensions for Efficient Model Design. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 732–741.