

DEEP LEARNING FOR AUTOMATIC BUILDING DAMAGE ASSESSMENT: APPLICATION IN POST-DISASTER SCENARIOS USING UAV DATA

A. Calantropio^{1,*}, F. Chiabrando¹, M. Codastefano², E. Bourke²

¹ Laboratory of Geomatics for Cultural Heritage (LabG4CH), Department of Architecture and Design (DAD), Polytechnic University of Turin, Viale Pier Andrea Mattioli, 39, 10125 Torino (TO), Italy – alessio.calantropio@polito.it

² Unmanned Aircraft Systems Team (UAS), IT Preparedness and Emergency Response Branch (TECF), Technology Division (TEC), World Food Programme (WFP), Via Cesare Giulio Viola, 68, 00148 Roma, Italy

Commission I, ICWG I/II

KEY WORDS: Emergency response, Building damage assessment, deep learning, UAV photogrammetry, automatic segmentation, automatic classification.

ABSTRACT:

During the last few years, the technical and scientific advances in the Geomatics research field have led to the validation of new mapping and surveying strategies, without neglecting already consolidated practices. The use of remote sensing data for damage assessment in post-disaster scenarios underlined, in several contexts and situations, the importance of the Geomatics applied techniques for disaster management operations, and nowadays their reliability and suitability in environmental emergencies is globally recognized. In this paper, the authors present their experiences in the framework of the 2016 earthquake in Central Italy and the 2019 Cyclone Iдай in Mozambique. Thanks to the use of image-based survey techniques as the main acquisition methods (UAV photogrammetry), damage assessment analysis has been carried out to assess and map the damages that occurred in Pescara del Tronto village, using DEEP (Digital Engine for Emergency Photo-analysis) a deep learning tool for automatic building footprint segmentation and building damage classification, functional to the rapid production of cartography to be used in emergency response operations. The performed analyses have been presented, and the strengths and weaknesses of the employed methods and techniques have been outlined. In conclusion and based on the authors' experience, some operational suggestions and best practices are provided and future research perspectives within the same research topic are introduced.

1. INTRODUCTION

1.1 Geomatics tools for damage assessment

Disasters that hit communities and places strongly affect our history and culture. Nevertheless, disaster events can be monitored, sometimes predicted and their effect mitigated using the right tools.

Among different strategies and techniques applied in this specific context, Geomatics had a prominent role in documenting and archiving spatial 3D information, using image-based and range-based metric survey systems.

The easiness of deploying UAV platform and performing aerial photogrammetric survey made aerial photogrammetry the best rapid mapping method for performing emergency cartography, supporting the intervention in the field of firefighters and first responders (Calantropio et al., 2018)

Different kinds of damage classification grades have been introduced over time. While Copernicus EMS uses five damage classes (Grünthal, 1998), the BAR Methodology uses four classes ("critical visible damage"; "significant visible damage"; "minimal visible damage" and "no visible damage"); the UNOSAT classification uses a binary approach ("damaged"; "not damaged"). An exhaustive analysis of the different damage scales used for building damage assessment by the main satellite-based emergency mapping service has been discussed (Cotrufo et al., 2018), stating that different damage classes and detailed interpretation guidelines with operational examples are crucial for assuring the suitable analysis of the analyzed data.

Damages can be categorized based on the observation of the building's morphological features and their spectral features (distribution of brightness, the regularity of the texture, etc.) (Li & Tang, 2018).

A method is to use both pre-event and post-event data, which unfortunately are not always available - like pre-event and post-event TerraSAR-X (TSX) radar images, VHR remote sensing imagery from satellite (including DigitalGlobe's WorldView satellites), UAV and ground-level images.

Building footprints are usually the easiest kind of data to retrieve and are obtainable from different sources, such as pre-event vector maps or cadastral maps. These sources are unfortunately not always up to date; for this reason, when building vector footprints are not available or outdated, it is possible to exploit other kinds of pre-event data, such as VHR satellite image or aerial image that (if relatively recent) allows a manual interpretation of the building's boundaries (Li & Tang, 2018). Some experience (Cotrufo et al., 2018) underlined that it is impossible to directly apply the damage classification scales proposed (Grünthal, 1998) for addressing slight structural damages using remotely sensed images. Even if satellite images can achieve sub-meter resolution, there are still issues in identifying partially damaged buildings; for this reason, data integration from UAV-derived information remains crucial (Li & Tang, 2018). Pre-earthquake condition imagery data can be retrieved from online platforms that use different sources for providing RGB orthophotos (derived from UAV or aerial or satellite photogrammetry). Unfortunately, online data are limited in time and space (i.e., only certain regions are mapped).

* Corresponding author

1.2 WFP experience in Mozambique after cyclone Idai: Artificial intelligence for geospatial data analysis in emergency contexts

In recent years, AI (artificial intelligence) techniques, such as ML (machine learning) and, more specifically, DL (deep learning), have evolved, supporting human activities in different fields. Apart from examples related to disaster management and damage assessment based on pre-event and post-event data (Nia & Mori, 2017), there are wide-ranging examples of Deep Learning techniques such as convolutional networks for biomedical image segmentation (Ronneberger et al., 2015) employed for medical purposes such as cancer detection (Hu et al., 2018; Shen, 2017; Shen et al., 2017)¹.

2012 was a crucial year for Deep learning, after discussing the ImageNet contest results (Krizhevsky et al., 2017) and the demonstrated capability to recognize objects comparing millions of images without human-codified instruction. When big data trained-sets are available, deep learning constitutes a good alternative compared to classical algorithms such as *Support Vector Machine* (Cortes & Vapnik, 1995) or *Random Decision Forests* (Ho, 1995), which usually better perform on structured data, such as tables and images.

After an emergency strikes, people affected are in the most vulnerable situation and taking the right decision at the right time is fundamental to minimize the impact on their lives. The WFP (World Food Programme) has seen that with an appropriate UAV prepositioning workflow (developed during the SEARCH project)², it is possible to collect high-value data in a small amount of time. The most significant limitation of this approach, amplified by the offline environment, is that it takes days to manually analyze the information in a structured and reliable way.

WFP, therefore, developed its open-source solutions tailored to the emergency context where object recognition can enrich and expedite humanitarian decision-making: DEEP (Digital Engine for Emergency Photo-analysis)³.

After Idai hit Mozambique in June 2019, DEEP was tested in Maputo, Mozambique, with partners from the INGC (*Instituto Nacional de Gestão de Calamidades* - National Institute of Disaster Management - Mozambique) and students from Eduardo Mondlane University. Ten participants annotated UAV imagery during a ten-day training workshop, later fed into the DEEP algorithm, thereby grasping machine learning basics (Codastefano, 2019). DEEP automated the analysis and processing of high-resolution images – a process that can significantly speed up humanitarian response⁴. Its ease of use and modular design based on traditional computing (instead of cloud computing) is a perfect solution for emergency settings where power is limited and internet connectivity patchy. The software, designed by TECEF, can be installed on any laptop to run a model that can automatically extract objects from drone imagery, classify them as damaged or intact, and plot these on a map (Figure 1).

As such, models need to be trained and able to use drone images at various resolutions, depending on resource availability and field conditions. With the correct data, DEEP can also help with diverse tasks such as identifying standing water pools following a cyclone to help monitor potential cholera and malaria outbreaks.

In 2019, Cyclones Idai and Kenneth ravaged large swaths of southern and northern Mozambique, destroying homes, livelihoods and displacing nearly two million people. Working closely with the government and partners on the field, it was the first time WFP deployed drones during an emergency response. A few months later, INGC and WFP drones' combined effort provided a wealth of valuable data and insight, including maps and aerial imagery. Using this data is a unique opportunity to develop machine learning capacity in a real-world situation.



Figure 1. On the left: Orthophoto of Beira in 2019 after the cyclone Idai. On the right: WFP's DEEP software shows damaged houses in red and undamaged ones in purple — patchwork drone imagery turned into a detailed map.
Image: WFP.

DEEP is wholly based on Open-Source dependencies and specifically (but not only) on QGIS⁵ and GDAL⁶. During its development, the QGIS guidelines have been followed to create a widely compatible and standardized solution that can provide a high degree of deployability comparable to similar commercial solutions.

Because a neural network is not capable to efficiently work on a large image (such as a UAV-generated orthoimage), the primary requirement and the preliminary step of the workflow is to subdivide it into tiles of 1024x1024 pixels each; DEEP exploits the VRT format to keep the georeferentiation of the generated tiles, simplifying the ML task. The other important characteristic of DEEP is that it is possible to access different libraries for managing the VRT (thanks to GDAL), recompacting files, using a filter (SIEVE) that allows ignoring segmented geometries with an area lower than a specified number of square pixels and, in the end, transforming the generated raster of the segmentation into a geojson file using the option "poligonize" (initially bugged, this option has been corrected in the framework of this research).

The original idea of the workflow on which DEEP is based is inspired by the Zanzibar Aerial Mapping Project⁷ adapted with a custom implementation to extract the building segmented by the segmentation model (building/not building) to reclassify them according to the damage level (not damaged/damaged/empty). UNET⁸ (Ronneberger et al., 2015), based on KERAS⁹, was successfully tested on the Zanzibar dataset, and therefore implemented in DEEP. To tackle the emergency generated by

¹ <https://github.com/lishen/end2end-all-conv>

² <https://medium.com/world-food-programme-insight/this-workshop-came-along-at-the-right-time-14ed453915ef>

³ <https://www.wfp.org/stories/joining-dots-how-ai-and-drones-are-transforming-emergencies>

⁴ <https://medium.com/world-food-programme-insight/humanitarian-face-of-technology-e74b822b58cc>

⁵ <https://www.qgis.org/>

⁶ <https://gdal.org/>

⁷ <https://github.com/daveluo/zanzibar-aerial-mapping>

⁸ <https://github.com/zhixuhao/UNET>

⁹ <https://keras.io/api/>

Idai, the model was re-trained using images acquired on the field after the disaster.

In March 2020, the Open Cities AI Challenge: Segmenting Buildings for Disaster Resilience¹⁰ took place, and after studying the result of the challenge, it turned out that the 3rd classified solution, provided by Michael Busta¹¹ and based on (Lin et al., 2016), had suitable characteristics for being included (after practical implementations in pre-processing and post-processing) in DEEP, therefore substituting its original embedded ML segmentation model.

The damage classification model considers a minimum tile that includes the previously segmented buildings and classifies them according to the damage. The level of damage is assigned as an attribute of the previously segmented shape of the building footprint.

Because image classification (a whole image is classified and a single label is given as an output) is generally more straightforward than image segmentation (each pixel of the image is classified and a label is assigned to each pixel) (Blaschke, 2003), through negative mining, it is possible to create an "empty" class to detect (using the classification model) the segmented part of the images that were erroneously classified as buildings, to purge them from the final classification output. In this way, the "empty" class allows an optimization of the segmentation output, exploiting the fact that the classification model performs better in this task compared to the segmentation one.

DEEP works well with UAV-generated orthoimages at the current optimization and development (from 6 to 9 cm/pix GSD). For this reason, the use of satellite images, which are usually lower in resolutions, does not produce good results.

1.3 Politecnico di Torino experience in 2016 Central Italy earthquake

The 2016 earthquake experience in central Italy has been a validation field for many survey and mapping strategies proposed during recent years by the Geomatics research community. Exploiting the last decade's technical and technological advances for emergency response purposes is one of the more crucial topics, intending to obtain and efficiently organize high-scale reliable geospatial data for the early warning, impact, and recovery phases. In this sense, Geomatics applied techniques in DM (disaster management) are primarily devoted to enhancing search and rescue, analysis and assessment, damage monitoring, and emergency management activities.

BDA (building damage assessment) operations involve experimental sensors and approaches that work together or are combined with more traditional and consolidated practices. Remote sensing data in BDA is traditionally used through visual interpretation of operators; this is a time-consuming task. The new requests from the FR (first responders) are nowadays oriented towards quasi-real-time data processing and fast production of emergency cartography on different scales.

Based on previous similar experiences (Calantropio et al., 2018), UAV flights over damage sites capturing high-resolution video footage and photos for obtaining 3D models, and orthoimages using SfM techniques are nowadays consolidated methodologies for obtaining rapid and reliable information of the areas affected by a disaster.

In this scenario, different tests have been carried out during the post-earthquake documentation activities conducted by the Politecnico di Torino Geomatics group in cooperation with the

SAPR team of the Italian Fire Fighters and the GEER association¹² (Geotechnical Extreme Events Reconnaissance association).

Politecnico di Torino is an Italian technical university that has contributed to this direction, applying knowledge and resources to the study of calamities and disasters and their effect on the natural and built environment. Especially after the seismic swarms that hit Italy in August 2016, a task force made of volunteers among professors and researchers was created to document and study each phase of the event and their effects and guide the necessary interventions. Several UAVs and sensors were employed to define standards and guidelines for data acquisition and processing according to the different platforms, payloads, imaging sensors, and required accuracy of the final output.

Summarising the obtained results and according to the achieved final products, it is possible to define some essential items that need to be considered to get suitable outputs for the BDA operations. Concerning the image-acquisition phase, it is recognized that, in addition to the nadir acquisitions, the use of oblique images (Aicardi et al., 2016; Duarte et al., 2017; Ezequiel et al., 2014) is crucial and fundamental for improving the 3D reconstruction part (especially in urban areas) and it is also essential during the BBA (bundle block adjustment) since the use of those images allow to increase the rigidity of the photogrammetric block and, consequently, allowing the possibility of decreasing the number of the used GCPs (ground control points) when platforms without RTK/PPK are employed (Nesbit & Hugenholtz, 2019; Ostrowski & Bakula, 2016).

Another important aspect is related to the GCPs acquisition and the accuracy evaluation. If, on the one hand, it is crucial to obtain data very quickly during an emergency, on the other hand, it is essential to verify the final accuracy of the achieved output. To follow these objectives, using a minimum number (4-5) of well-measured points on the field is an essential requirement in case of acquisition performed using platforms without RTK/PPK capability. Following this practice could be a plus for the platforms with accurate onboard GNSS since the use of the surveyed data is necessary for evaluating the final accuracy (using them as checkpoints) or to improve the quality of the camera calibration (Gabrlík et al., 2018).

Finally, concerning the data-processing and apart from the possibility of having in a short time reliable and high-resolution 3D models useful for BDA activities, the possibility of combining multi-temporal acquisitions in a typical project (Aicardi et al., 2016) allows the monitoring of the affected area across time accurately.

Following the strategy mentioned above, the experiences carried out during the 2016 Earthquakes were deployed mainly using COTS (commercial off-the-shelf) multi-rotor and fixed-wing platforms. In September, October, November 2016, and February 2017, the acquisitions were performed over different villages with multi-sensors documentation of buildings and urban areas heavily and repeatedly damaged by the numerous seismic events. Several villages have been surveyed, such as Pescia, Pescara del Tronto, Cittareale, Accumoli, Norcia, Castelluccio, Amatrice, Campi di Norcia, etc. Multi-scale documentation has been achieved to allow the different researchers involved in the emergency activities to enrich the knowledge of the areas affected by the events.

In the next chapter of this paper, the first results of the use of DEEP are described and analyzed, starting from the achieved orthophoto to the performed building damage assessment.

¹⁰ <https://github.com/ntarn/open-cities-ai-challenge>

¹¹ <https://github.com/drivendataorg/open-cities-ai-challenge/tree/master/3rd%20Place>

¹² <http://www.geerassociation.org>

2. APPLICATION OF DEEP ON THE 2016 CENTRAL ITALY EARTHQUAKE DATASET

According to the experiences carried out by the WFP on the development of DEEP and the expertise of the Geomatics group of the Politecnico di Torino, in 2020, a joint research project focused on testing the possibilities offered by the proposed Deep Learning approach on the data collected during the 2016 post-earthquake surveys. The first tests related to the Village of Pescara del Tronto (Figure 2), seriously affected by the different shocks, are reported in detail in the following sections. These tests' primary objective is to develop an entirely new automatic methodology for damage assessment, not affected by a subjective evaluation that could be found in the Copernicus EMS classification.

2.1 Segmentation model

The first test concerned the segmentation model of DEEP that can be used (in addition to the existing cartography) to automatically detect the building footprint, which will be later classified by the classification module.

After the preliminary test, the building segmentation model was performed as expected on the Pescara del Tronto orthophoto (Figure 3). Therefore, there was no need to perform any additional adjustments to the model. The preliminary evaluation of the results showed that a minor amount of non-building items (roads, parking lots, etc.) was segmented (recognized as buildings). To mitigate the abovementioned issue, a "sieve" filter has been applied before generating the final building footprint vector, excluding all the items with an area lower than 5 m².

It is essential, however, to state that the segmentation model is valid only when the building footprint is still clearly recognizable; for this reason, in case of destroyed buildings, in a situation in which even a trained operator would have difficulties in recognizing the original building footprint, it is necessary to employ an existing layer of cartography that can provide the building footprint. Therefore, the segmentation model will be used to detect new buildings or additions built after the ones represented in the existing cartography to provide a comprehensive and updated cartographic base for the following BDA.

This step's output is a GeoJSON format file containing each shape of the buildings with a unique ID and a closed geometry.

2.2 Classification model

The second step was using the image classification model, previously trained with Mozambique and Zanzibar open data, for running a preliminary test using the Keras implementation of the VGG16 classification algorithm (Simonyan & Zisserman, 2015) on the orthophoto of Pescara del Tronto (Figure 3), generated by using UAV images gathered on August 2016 (after the first seismic swarm).

The results are shown in the following image (Figure 4). As it is possible to observe, the results are inconsistent and do not resemble the actual damage classification provided by Copernicus, sometimes providing random outputs.

The reported results were expected because the typology of debris produced by Mozambique's catastrophic event are substantially different from those in central Italy. It is essential to clarify that the challenge of applying this system and method to different cities and countries is related to the fact that various material and construction techniques generate different debris patterns; therefore, the features recognized by the DEEP classification model are not comparable at this stage.



Figure 2. Orthophoto of Pescara del Tronto (Italy) after the 2016 seismic event.



Figure 3. Automatic building segmentation performed by Deep on the Pescara del Tronto 2016 orthophoto. The recognized building footprints are represented in blue color.



Figure 4. The image shows the damage classification performed by DEEP on the Pescara del Tronto 2016 orthophoto, using the Mozambique classification model. The not-affected buildings are marked in white, while the damaged buildings are marked in red colors.

With the aim of improving the results obtained in this test, the classification layers of the CNN (Convolutional neural network) of the image classification model have been re-trained using data samples obtained from the UAV surveys performed in three other villages affected by the same seismic swarm in Central Italy (Figure 5). Thus, the additional samples were selected from the orthophotos of the cities of Amatrice, Accumoli, and Norcia (August 2016). Moreover, the preliminary test showed the adopted solution was still not optimal, as there were again some areas of the orthophoto segmented as building. This problem outlined the need for more advanced training with three classes: "not damaged", "damaged", and "empty". This additional class "empty" can be later used to refine the erroneously segmented items, eliminating them before the BDA operation.

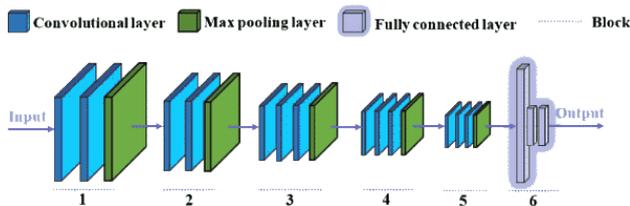


Figure 5. The image shows the neural network blocks that have been re-trained in two steps: at first, the classification layers (block 6); then the classification layer plus the last block of the feature extraction part (block 6 + block 5).

2.3 Training and validation: the dataset

The training and validation dataset comprises 440 damaged buildings, 440 not affected buildings, and 440 Negative examples (roads, crop fields, etc.), taken from the orthophotos of Accumoli, Norcia, and Amatrice after the seismic swarm of August 2016. UAV orthophoto used in the training and validation dataset refers to flights performed between the first and the second seismic event. So, they are in line and thus comparable with the Copernicus EMSR activations 177¹³ and 190¹⁴. However, even if the initial data was already labelled and divided into five damage classes, according to EMS98 (Grünthal, 1998), the different classes were highly unbalanced. Moreover, Copernicus's damage classification was performed on aerial data (lower resolution than the UAV orthophotos used in this research). The classification was subject to human errors, as it was an operator-dependent classification. For the reasons mentioned above, the data have been divided into 2 classes of damage, using a binary classification of "not damaged" (class 0) and damaged (class 1). As previously mentioned, a class "empty" containing negative examples have been introduced.

The dataset was therefore divided as follows:

- 80% Training set.
- 20% Validation set.

And contained:

- 440 damaged buildings.
- 440 not affected buildings.
- 440 negative examples (roads, crop fields, etc.).

The data augmentation parameters used for the training phase are reported in the appendix at the end of the article. With a batch size of 8, there were 660 steps per epoch.

2.4 Training and validation: the results

Step A: At first, only a re-training of the classification layer has been performed. Since the model was already pre-trained on the Mozambique data, there was no need for a long training process. It was interrupted after epoch 4 – validation accuracy improved from 0.91 to 0.95. The details of this step are reported in Table 1 and Table 2.

		Actual Class		
		Undamaged	Damaged	Empty
Predicted Class	Undamaged	78	10	0
	Damaged	2	85	1
	Empty	0	1	87

Table 1. The confusion matrix (264 images belonging to 3 classes) shows the validation results obtained after the first training.

Class	Precision	Recall	f1-score	support
Undamaged	0.97	0.89	0.93	88
Damaged	0.89	0.97	0.92	88
Empty	0.99	0.99	0.99	88
Accuracy			0.95	264
Macro avg. ¹	0.95	0.95	0.95	264
Weighted avg. ²	0.95	0.95	0.95	264

¹ averaging the unweighted mean per label.
² averaging the support-weighted mean per label.

Table 2. Main classification metrics after the first training (epoch 4 – validation accuracy improved from 0.91 to 0.95 – validation loss: 0.14). The table shows a summary of the precision, recall, and F1 score for each class.

Step B: In this step, a fine-tuning of the whole neural network has been performed, re-training not only the last layer as done in the first step but also the last feature extraction block. The process was interrupted after epoch 2 – validation accuracy improved from 0.89 to 0.97. The details of this step are reported in Table 3 and Table 4.

		Actual Class		
		Undamaged	Damaged	Empty
Predicted Class	Undamaged	83	5	0
	Damaged	3	85	0
	Empty	1	0	87

Table 3. The confusion matrix (264 images belonging to 3 classes) shows the validation results obtained after the second training.

¹³ <https://emergency.copernicus.eu/mapping/list-of-components/EMSR177>

¹⁴ <https://emergency.copernicus.eu/mapping/list-of-components/EMSR190>

Class	Precision	Recall	f1-score	support
Undamaged	0.95	0.94	0.95	88
Damaged	0.94	0.97	0.96	88
Empty	1.00	0.99	0.99	88
Accuracy			0.97	264
Macro avg. ¹	0.97	0.97	0.97	264
Weighted avg. ²	0.97	0.97	0.97	264

¹ averaging the unweighted mean per label.
² averaging the support-weighted mean per label.

Table 4. Main classification metrics after the second training (epoch 2 – validation accuracy improved from 0.89 to 0.97 – validation loss: 0.10). The table shows a summary of the precision, recall, and F1 score for each class.

2.5 Binary shades for multi classes classification

After the training and the validation phases, the classification model has been tested again on the Pescara del Tronto orthophoto. It is essential to clarify that the models have never seen Pescara del Tronto's data during the training and validation phases.

After the segmentation, the buildings detected by DEEP were subtracted from the buildings provided in the existing cartography, generating a layer containing only buildings not present in the current cartography. The 3-class classification model ("not damaged"; "damaged"; and "empty") has been performed only on this last layer, as it might contain wrongly segmented items (erroneously segmented as buildings). The items classified as "empty" were therefore discarded. The other buildings extracted from the existing cartography were classified using the 2-class classification model ("not damaged" and "damaged") because it was sure that they contained only buildings.

To further improve the quality of the final output, and to make it comparable with other damage classification standards (such as EMS-98), instead of assigning to each building a value "not damaged" or "damaged", it was considered the classification likelihood score (from 0 to 1) of belonging to one of the two classes. This value has been assumed as a function of the level of damage for each of the analyzed buildings.

The whole workflow is summarized in Figure 6, and the classification output is reported in Figure 7.

All the steps detailed in this chapter were undertaken to refine the results initially obtained in Figure 4. The output is a heatmap of the damages that exploits the doubts of the classification model, using the probability of a building to be "not damaged" or "damaged" and assigning a score from 0 to 1 (from white to red). However, it is essential to check how the model performed compared to the ground truth data (obtained by Copernicus EMS).

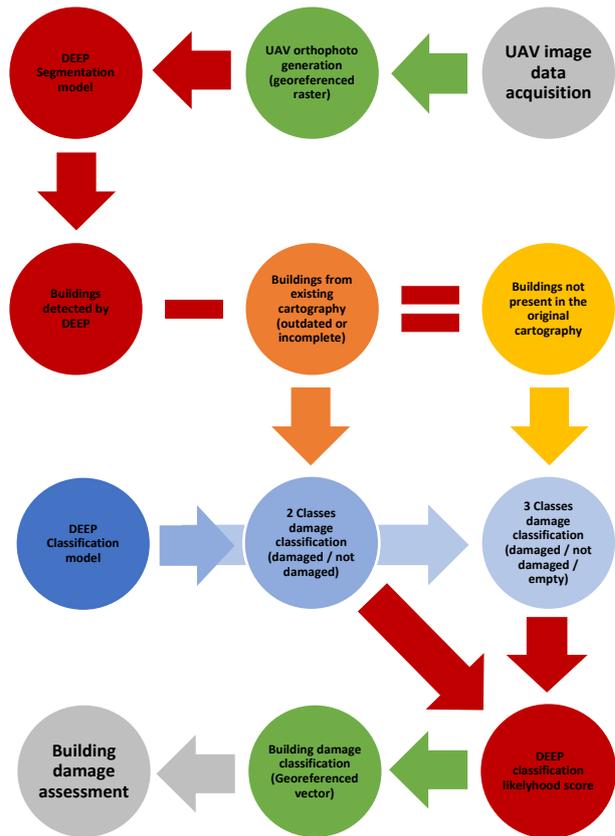


Figure 6. DEEP operative workflow. The diagram shows all the steps performed, from image acquisition to damage assessment.

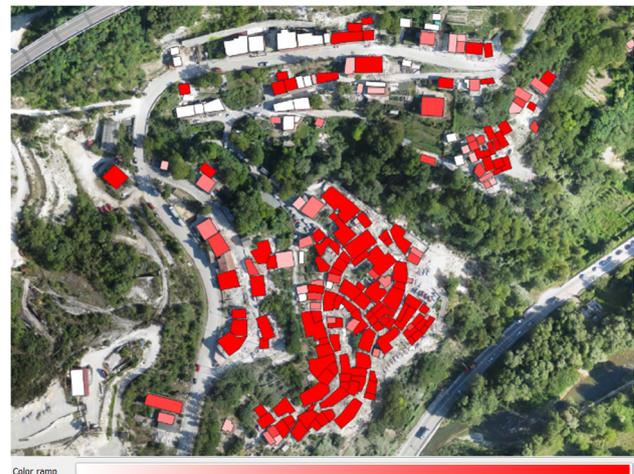


Figure 7. Damage classification performed by DEEP on the Pescara del Tronto 2016 orthophoto, using the new classification model based on the central Italy earthquake dataset. The color scale shows the classification likelihood score of belonging to the "not damaged" class (0 - in white) or to the "damaged" class (1 - in red).

2.6 Copernicus verification and comparison

It is not easy to define at the current moment the interval of the DEEP's scores that correspond to each of the classes in compliance with the EMS-98 classification adopted by Copernicus. For this reason, a confusion matrix of the actual and predicted level of damages cannot be provided. To give an idea for quantifying the method's performance in this specific case, it is possible to provide a qualitative comparison between the likelihood score assigned by DEEP (0-1 white-red color ramp) and the damage level set by Copernicus (Figure 8).

At this stage, the possibility of assigning, for each class, an interval of the score is possible. It will require further tests involving a higher number of ground-truth data that can be used for a suitable calibration of the score conversion.

It will also be necessary to re-create a ground-truth dataset, as Copernicus's classification is not only subject to bias because of its operator-related nature but also performed on aerial images with a lower GSD that makes it difficult to assign the right level of damage unequivocally.

Moreover, the next step of the implementation will be to use 3 level of damage ("not damaged"; "damaged"; "destroyed"; plus the "empty" class) because it is relatively easy to distinguish between "damaged" and "destroyed" buildings, but the thematic accuracy of the intermediate classes of the EMS-98 is not sufficient due to the right-from-above point of view of the orthophotos (Cotrufo et al., 2018). This additional step will, of course, require a re-training and a new fine-tuning of DEEP with opportunely classified classes.

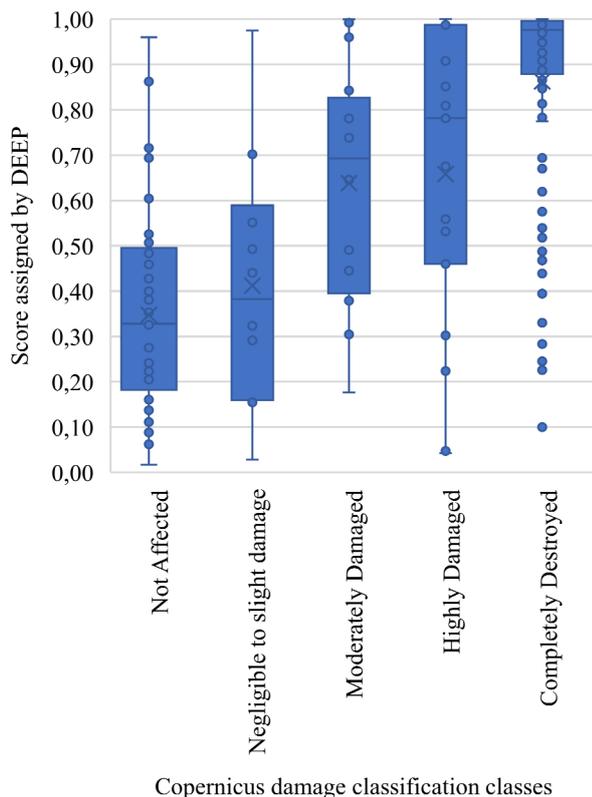


Figure 8. For each of the classified buildings, a box plot diagram shows the damage class assigned by Copernicus (on the x-axis) and the relative likelihood score given by DEEP (on the y-axis) for each of the items (buildings). The whiskers indicate the variability outside the first and last quartiles; the X symbol indicates the average value for a given class.

3. CONCLUSIONS

During the development of the tests related to this research, it has been observed that DEEP works well with orthoimages that have a GSD of $9 \text{ cm} \pm 3 \text{ cm}$; this is a good compromise in terms of the orthoimages' quality, the number of images required for its generation, flight altitude, and time of flight. For example, a DJI Phantom 4 RTK can obtain a 9 cm/pix GSD with a flight altitude of about 330 m, while the SenseFly eBeeX can get a 9 cm/pix GSD with a flight altitude of approximately 390 m.

It is important to note that the optimal resolution in real case scenarios is a function of the flight altitude (and not vice-versa) because in the early aftermath of the emergency the drone might have to perform a flight. Simultaneously, other aircraft (such as helicopters) might need to operate to assist with the ground operations and logistics.

It has also been observed that higher resolutions (lower GSD) introduce false positive detections during the segmentation phase and increase the overall processing time; lower resolutions (higher GSD) instead does not make the classification model working correctly, making it difficult or impossible to detect the level of damage accurately.

The proposed method can deliver a thematic map for a very early assessment of the situation in less than one hour after the arrival on the site (10-15 min flight time, 5 min data transfer, 20 min orthophoto generation, 10 min DEEP classification). The DEEP segmentation model is helpful if one wants to update the existing cartography if the retrieved building shapes are not updated.

The following steps will be devoted to investigating this tool's potentialities in real case applications, thanks to an ongoing collaboration with the Italian National Fire Corps - Corpo Nazionale dei Vigili del Fuoco, in two steps. A first step through creating a comprehensive dataset of orthophotos acquired in emergencies that can provide valuable material for the optimization of DEEP, to improve its performance in the future. A second step concerns the automatization of the processing part (reading the orthophoto, auto-detect the GSD, and opportunely set a value for the sieve filter, and so forth) to enhance the algorithm deployability on the field.

Principal errors are related to the fact that the training set does not contain a relatively sufficient number of terraces or other kinds of roofs and some dubious situations where even a human operator will have difficulty assigning a damage level.

Per the publication date of this paper, this is the very first research presented about the DEEP network (that has not yet – but will be soon – publicly released and will be freely accessible and usable). Future research work will be focused on implementing an instance segmentation model that will substitute the current segmentation model (due to some issues reported in the case of buildings too close to each other, recognized as a single entity). Moreover, the possibility of performing a visual interpretation of more orthoimages will provide an up-to-date and reliable ground truth that will allow performing a more comprehensive validation of the employed method.

ACKNOWLEDGEMENTS

The authors would like to thank Politecnico di Torino (Prof. Foti S.) for the earthquake task force organization and for kindly providing technical and financial support for the reconnaissance personnel. The authors also thank the firefighters RPAS group (Di Lolli A., Lorusso O., Feliziani F., Fiorini M.), which has continuously accompanied us in the fieldwork, assisting with people and resources, and the Italian Department of Civil Protection; thanks to the GEER Association, with whom we have made a fruitful collaboration. Thanks to INGC for allowing the use of the Mozambique image dataset.

REFERENCES

- Aicardi, I., Nex, F., Gerke, M., & Lingua, A. M. (2016). An image-based approach for the Co-registration of multi-temporal UAV image datasets. *Remote Sensing*, 8(9), 779. <https://doi.org/10.3390/rs8090779>
- Blaschke, T. (2003). Object-based contextual image classification built on image segmentation. *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*, 113-119
- Calantropio, A., Chiabrando, F., Sammartano, G., Spanò, A., & Losè, L. T. (2018). UAV strategies validation and remote sensing data for damage assessment in post-disaster scenarios. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3W4), 121-128. <https://doi.org/10.5194/isprs-archives-XLII-3-W4-121-2018>
- Codastefano, M. (2019). *Machine Learning in Emergency Response*. https://www.itu.int/dms_pub/itu-d/oth/07/1d/D071D0000050001PDFE.pdf
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/bf00994018>
- Cotrufo, S., Sandu, C., Giulio Tonolo, F., & Boccardo, P. (2018). Building damage assessment scale tailored to remote sensing vertical imagery. *European Journal of Remote Sensing*, 51(1), 991-1005. <https://doi.org/10.1080/22797254.2018.1527662>
- Duarte, D., Nex, F., Kerle, N., & Vosselman, G. (2017). Towards a more efficient detection of earthquake induced façade damages using oblique UAV imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2W6), 93-100. <https://doi.org/10.5194/isprs-archives-XLII-2-W6-93-2017>
- Ezequiel, C. A. F., Cua, M., Libatique, N. C., Tangonan, G. L., Alampay, R., Labuguen, R. T., Favila, C. M., Honrado, J. L. E., Canos, V., Devaney, C., Loreto, A. B., Bacusmo, J., & Palma, B. (2014). UAV aerial imaging applications for post-disaster assessment, environmental management and infrastructure development. *2014 International Conference on Unmanned Aircraft Systems, ICUAS 2014 - Conference Proceedings*, 274-283. <https://doi.org/10.1109/ICUAS.2014.6842266>
- Gabrlík, P., Cour-Harbo, A. la, Kalvodova, P., Zalud, L., & Janata, P. (2018). Calibration and accuracy assessment in a direct georeferencing system for UAS photogrammetry. *International Journal of Remote Sensing*, 39(15-16), 4931-4959. <https://doi.org/10.1080/01431161.2018.1434331>
- Grünthal, G. (1998). *EMS-98 (European Macroseismic Scale)*. <https://www.gfz-potsdam.de/en/section/seismic-hazard-and-stress-field/data-products-services/ems-98-european-macroseismic-scale/>
- Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1, 278-282. <https://doi.org/10.1109/ICDAR.1995.598994>
- Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis - A survey. *Pattern Recognition*, 83, 134-149. <https://doi.org/10.1016/j.patcog.2018.05.014>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>
- Li, S., & Tang, H. (2018). Building damage extraction triggered by earthquake using the UAV imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(3), 929-936. <https://doi.org/10.5194/isprs-archives-XLII-3-929-2018>
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2016). *Feature Pyramid Networks for Object Detection*. <http://arxiv.org/abs/1612.03144>
- Nesbit, P. R., & Hugenholtz, C. H. (2019). Enhancing UAV-SfM 3D model accuracy in high-relief landscapes by incorporating oblique images. *Remote Sensing*, 11(3), 239. <https://doi.org/10.3390/rs11030239>
- Nia, K. R., & Mori, G. (2017). Building Damage Assessment Using Deep Learning and Ground-Level Image Data. *2017 14th Conference on Computer and Robot Vision (CRV)*, 95-102. <https://doi.org/10.1109/CRV.2017.54>
- Ostrowski, W., & Bakula, K. (2016). Towards efficiency of oblique images orientation. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3/W4, 91-96. <https://doi.org/10.5194/isprs-archives-xl-3-w4-91-2016>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- Shen, L. (2017). End-to-end training for whole image breast cancer diagnosis using an all convolutional design. In *arXiv*. <https://doi.org/10.1038/s41598-019-48995-4>
- Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R. B., & Sieh, W. (2017). Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography. *ArXiv*. <https://doi.org/10.1038/s41598-019-48995-4>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. <http://www.robots.ox.ac.uk/>

APPENDIX

Data augmentation parameters used during the training phase:

```
train_datagen = ImageDataGenerator(  
    rescale=1. / 255,  
    rotation_range=60,  
    horizontal_flip=True,  
    vertical_flip=True,  
    brightness_range=[0.2, 1.0],  
    zoom_range=0.2,  
    fill_mode='nearest')
```