DAMAGE DETECTION ON BUILDING FAÇADES USING MULTI-TEMPORAL AERIAL OBLIQUE IMAGERY

D.Duarte [1]*, F. Nex [1], N. Kerle [1], G. Vosselman [1]

[1] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
(d.duarte, f.nex, n.kerle, george.vosselman)@utwente.nl

Commission II, WG II/4

KEY WORDS: Convolutional Neural Networks, Change Detection, Manned-platforms, Remote Sensing, Deep Learning

ABSTRACT:

Over the past decades, a special interest has been given to remote-sensing imagery to automate the detection of damaged buildings. Given the large areas it may cover and the possibility of automation of the damage detection process, when comparing with lengthy and costly ground observations. Currently, most image-based damage detection approaches rely on Convolutional Neural Networks (CNN). These are used to determine if a given image patch shows damage or not in a binary classification approach. However, such approaches are often trained using image samples containing only debris and rubble piles. Since such approaches often aim at detecting partial or totally collapsed buildings from remote-sensing imagery. Hence, such approaches might not be applicable when the aim is to detect façade damages. This is due to the fact that façade damages also include spalling, cracks and other small signs of damage. Only a few studies focus their damage analysis on the façade and a multi-temporal approach is still missing. In this paper, a multi-temporal approach specifically designed for the image classification of façade damages is presented. To this end, three multi-temporal approaches are compared with two mono-temporal approaches. Regarding the multi-temporal approaches the objective is to understand the optimal fusion between the two imagery epochs within a CNN. The results show that the multi-temporal approaches outperform the mono-temporal ones by up to 22% in accuracy.

1. INTRODUCTION

Remote sensing has been continuously used for automatic building damage assessment, since the used platforms can cover large areas and attenuate the costly and lengthy ground observations. While several sensors coupled with distinct platforms have been used (Armesto-González et al., 2010; Dell'Acqua and Polli, 2011; Gokon et al., 2015), there has been a special interest in the use of images (Duarte et al., 2018a; Tu et al., 2017; Vetrivel et al., 2017).

Several approaches have been used to detect damages from images. These usually rely on a set of features that are later used as input for a supervised classifier. While hand crafted features have been used (Vetrivel et al., 2016b), convolutional neural networks (CNN) features have recently been found to be preferable (Duarte et al., 2018a; Vetrivel et al., 2017).

Hence, current image based damage detection approaches often rely on convolutional neural networks to classify a given image into damaged and non-damaged regions (Duarte et al., 2018a; Vetrivel et al., 2017). These approaches aim at detecting damage evidences such as rubble piles and debris from satellite and aerial (manned and unmanned) imagery. While the image classification of debris and rubble piles may achieve higher accuracy, it might overlook smaller signs of damage (e.g. spalling and cracks) which are usually present on the façades. These often differ in image characteristics when compared to those of rubble piles or debris (see Figure 1 – figure with low levels of damage and with totally destroyed areas for comparison). Furthermore, when these approaches are used for façade damage detection, there is a large number of false positives (Duarte et al., 2017), i.e. many images patches depicting intact buildings which are classified as damaged. This indicates the limitation of such models trained with image samples depicting rubble piles and debris for the façade damage assessment.

In the case multi-view aerial imagery is captured with enough overlap, the computation of 3D point clouds through dense image matching is possible. These 3D models may then be used to detect geometrical deformations of the buildings (Sui et al., 2014), while the images may be used to detect rubble piles and/or debris as well as smaller signs of damage such as spalling and cracks (Fernandez Galarreta et al., 2015). The façade planes are often tilted respective to the image plane, which increases the noise present in the 3D point cloud (Rupnik et al., 2014). These dense image matching problems combined with the usual decimetre resolution of manned multi-view aerial image surveys drastically decrease the chances of identifying cracks and small signs of spalling from the point cloud.

In spite of the growing amount of literature regarding the image classification of building damages, little attention has been given to the detection of damages on the facades. Most of the studies focus on the identification of damage evidences such as debris and/or rubble piles, which may leave out damage evidences present in façades (see Figure 1), such as cracks and spalling.

Specifically focusing on the facades, a few approaches can be found in the literature. For example, considering UAV imagery, and relying both on the image and 3D features, Fernandez Galarreta et al. (2015) determined several types of damage. Among them, cracks and spalling from façades. Gerke and Kerle (2011) used multi-view aerial imagery and derived a 3D point cloud to extract features and identify damaged buildings, and at

* Corresponding author

the same time classified the damage of a given building into three classes which were based on the European Macroseismic Scale (EMS-98). More recently, Tu et al. (2017) identified damaged façades using local symmetry features and the Gini Index extracted from aerial oblique images. The authors assumed symmetric façades and considered the deviations from that symmetry as damaged façade proxies.

Figure 1. Top: example of two completely collapsed buildings. Below: examples of extracted facades. Left, partially collapsed building. Right and centre, two damaged facades with spalling and other damages while the facade is still standing

Only one contribution used pre- and post-event multi-view aerial imagery in a multi-temporal approach to detect damaged façades. Vetrivel et al. (2016a) tested the potential of multi-temporal aerial imagery, using a simple correlation coefficient to determine the similarity between two rectified façade image patches. However, no statistical measures on the quality of the approach were reported.

In this work we assess the impact of considering multi-temporal aerial oblique imagery for the detection of damages along the façades. To this end we use three different approaches that rely on convolutional neural networks. These are then compared with two mono-temporal approaches that can be found in recent literature. Moreover, and due to the small amount of data, we take advantage of the high image overlap of aerial multi-view images to generate the needed input for a multi-temporal façade damage detection approach.

In the following section the used dataset is presented. Section 3 introduces the methodology for both the facade image patch extraction from the aerial imagery and the mono/multi-temporal approaches used in the paper. Section 4 presents the results while section 5 and 6, respectively present the discussion and conclusion.

2. DATA

Two airborne oblique acquisitions are considered in this paper. These datasets were captured in August 2008 and in May 2009, depicting the pre- and post-event of the April 2009 earthquake which occurred in central Italy. The images were captured over the city of L'Aquila and a nearby village (Tempera).

The dataset was captured with the Pictometry system which contains small format DSLR cameras, 4 obliques (one for each cardinal direction) and 1 nadir. These were acquired at a flying height of approximately 1000m, with an average ground sampling distance of 0.14 m on the oblique views. The flight was performed considering a forward overlap between 60-70% and a side overlap between 35-45%, which allowed to derive a 3D point cloud.

2.1 Limitations

This dataset contains partial and total collapsed buildings but mostly depicts damage on the façades (e.g. spalling, cracks, etc, see Figure 1). Often, it presents areas with occluded façades due to urban design, where narrow streets are common and hence not visible in the oblique images (see Figure 2). These two issues only allowed to extract 88 damaged façades.

Figure 2. Oblique view over L'Aquila, narrow streets and high buildings do not allow to visualize some of the facades

3. METHODOLOGY

In this section the multi-temporal façade damage detection approach is presented. This approach assumes as input the façade image patches of both pre- and post-event. Hence, the procedure to extract the façades image patch from the images is explained in the first sub-section. Sub-section 3.2 presents the three multi-temporal façade damage detection approaches using CNN. These muti-temporal approaches were compared with two mono-temporal approaches, presented also in sub-section 3.2.

Figure 3 shows an overview of the steps, indicating the section of each of the steps.

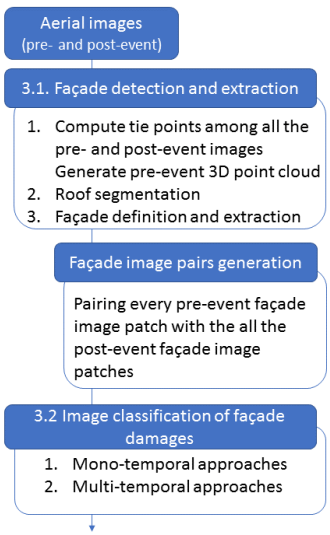

Figure 3. Overview of the steps of the facade damage detection using multi-temporal oblique images

`

3.1 Extraction of the façade image patches from both the pre- and post-event images

The extraction of the façades image patches from the multi-temporal oblique imagery was accomplished by using the pre-event 3D point cloud. This point cloud was generated considering only pre-event images. However, to have the datasets registered, the tie-points were computed from both epochs imagery, which forced them to share the same local coordinate system. To define and extract the façades, a method similar to the one defined by Duarte et al. (2017) was followed.

This approach uses the tie-point point cloud derived from highly overlapping unmanned aerial vehicles' (UAV) images, as input for the posterior point cloud plane-based segmentation. In this case, dense point clouds generated by image matching were used for this task instead of the tie-point clouds used in the original approach.

The point clouds were used to determine the building roof locations through a plane-based segmentation followed by connected component analysis. With the roofs defined, a minimum bounding rectangle was fitted to each roof segment. In this way, four façades per building were located and could be extracted from the images using the projection matrices. More details can be consulted in the paper (Duarte et al., 2017).

With this approach, buildings with round shaped roofs or with a different geometry other than a square may be impacted by assumption of having 4 façades per building. However, from the performed tests, such assumption showed a small impact in the final quality of the façade detection process (see Figure 4).

Figure 4. Example of a segmented round roof. In spite of the four façades per building assumption, the façade area is captured.

Due to the high overlap of aerial multi-view surveys, a given facade might be visible from several images. Analogously, each façade should be visible in both epochs. Image pairs were, therefore, created associating each pre-event façade image patch to all the post-event image patches of the same facade. Table 1 presents the number of façades and corresponding image pairs used in this study (~25 image pairs per façade). Given the low number of damaged façades; non-damaged façades had to be discarded so that the amount of image pairs was the same for both classes. In this way it was possible to carry out the current study, otherwise there would only be around 180 image samples. Figure 5 depicts various façade image patches extracted from different images but from the same epoch.

	Image pairs	Façades
Damaged	2274	88
Not damaged	2272	90
Total	4546	178

Table 1. Number of façades and image pairs considered in this study.

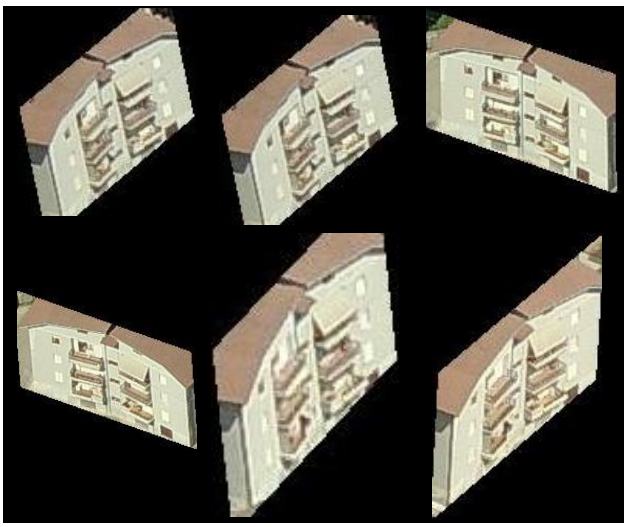

Figure 5. Examples of the same façade and same epoch, from different images.

3.2 Mono- and multi-temporal façade damage detection using convolutional neural networks.

Five different CNN approaches were tested using the image pairs (or post-event images in the mono-temporal case) to assess the value of multi-temporal image data in the image classification of façade damages. In the following sub-section, the two mono-temporal experiments are described, while sub-section 3.2.2 explains the multi-temporal experiments.

Since the paper published by Krizhevsky et al. (2012), the use of convolutional neural networks has become an established machine learning technique for, among others, image-based tasks. Convolutional neural networks are built by hierarchically stacking of convolutions that enable a network to learn from lower level features to higher levels of abstraction. Several improvements have been proposed by the computer vision community and were also successfully used in image-based remote sensing applications. The depth of the proposed networks has increased since then. However, deeper networks are usually harder to train, given the high number of parameters. Residual connections, where the input of a given layer is the summation of previous layers, have been then proposed (He et al., 2016). This allowed to have deeper networks while maintaining the number of parameters low. Moreover, to consider feature information not only from the previous layer but also from preceding layers, allows to consider more feature information that otherwise could be lost in backpropagation (Yu et al., 2017).

Another major development of convolutional neural networks is the use of dilated convolutions (Yu and Koltun, 2016). These

`

convolutions are applied to a given input using a kernel with pre-defined gaps. For example, a convolution with a kernel of 3x3px and dilation 2, has a receptive field of 7x7px, while maintaining a low number of parameters. In this way, more context is captured without the need of aggressive down-sampling of the feature maps throughout the networks.

In the proposed work, the dilated convolutions aimed at capturing the context of the fine details such as the damage along the façades (see Figure 1). In this case a maximum dilation of 4 was adopted, with a receptive field of 11x11px. The smallest feature map size of the network was 28x28px.

In Figure 6, the base network (*stream* in the figure) used in the approach is shown (this is the *stream* used in Figure 7). This was built with sets of convolutions, batch normalizations and *relu* (blue rectangles in Figure 6) (Ioffe and Szegedy, 2015). These may or may not have a residual connection indicated by a + in the figure. While these connections are used during the increasing dilation, they are eliminated in the last set of convolutions where the dilation is decreasing. These latter connections were removed in order to capture the more local features that might have been lost due to the aggressive dilation increase (Yu et al., 2017). The down-scale of the feature maps is performed with striding 2 instead of 1, red rectangles in Figure 6.

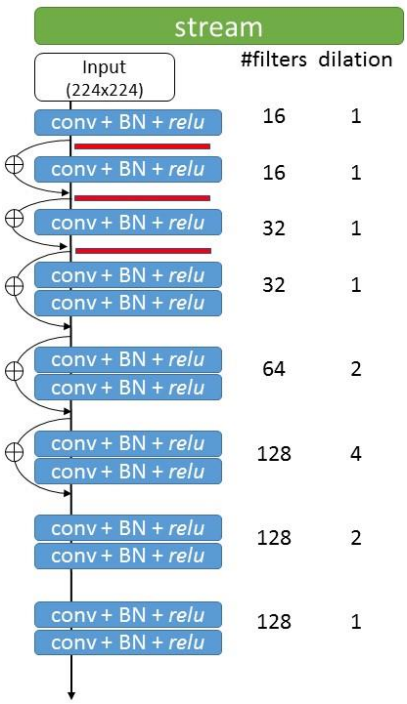

Figure 6. Network used for the mono- and multi-temporal approaches. Residual connections are indicated with a + sign. The red rectangles means that striding of 2 instead of 1 was used. *conv* stands for convolution + batch normalization (BN) + *relu*.

The network was empirically tested. For example, the number of filters had to be kept to a maximum of 128. Otherwise, the network would easily overfit the training data. In this case it was found to be better to have a deeper network but with a smaller number of filters. Increasing the dilation also did not affect the result.

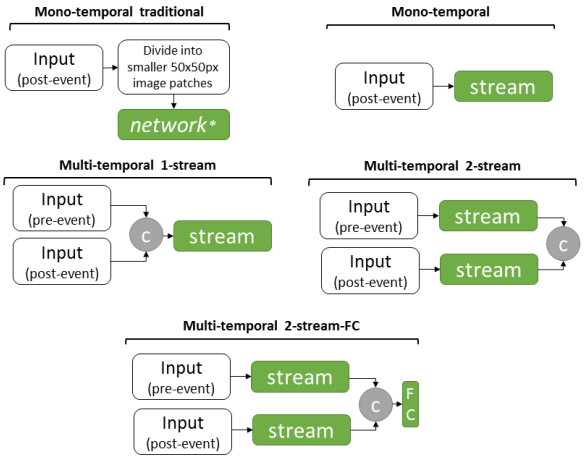

Figure 7. Mono- and multi-temporal approaches considered in this study. * refers to the network used in Duarte et al. (2018a).

3.2.1 Mono-temporal approaches

Two mono-temporal approaches were tested, depicted in Figure 7. The mono-temporal traditional (MN-trd) approach directly used a network trained with aerial image patches containing debris and rubble piles, as in Duarte et al. (2018a). It was trained using aerial image patches of several geographical locations, including datasets similar to the L'Aquila one (e.g. Amatrice, Italian city with similar urban design of L'Aquila). Instead of considering the whole façade as in the other mono-temporal approach, it divides the post-event façade image patches into smaller 50x50px image patches which are then fed to the network as described in Duarte et al. (2018a). In the case a given façade image patch contains at least 1 smaller image patch classified as damage the whole façade is considered damaged.

The other mono-temporal approach (MN-fac) used as input only the post-event façade image patches defined in section 2, divided in two classes, damaged and not damaged. These were then fed to the stream defined in Figure 6.

3.2.2 Multi-temporal approaches

As indicated in Figure 7, three multi-temporal approaches were tested. These aimed at understanding how the different streams of data (pre- and post-event) could be merged together for an optimal image classification of façade damages. The merging of the different epochs/modalities of data within a CNN has been the focus of many recent research in change detection (Daudt et al., 2018; Wang et al., 2018), but also in merging multi-modal data (Audebert et al., 2018; Xu et al., 2017). Two distinct approaches are tested: early and late fusion of the epoch-specific streams. For example, multi-temporal 1-stream (MT-1str) concatenated the pre- and post-event images in the image channels direction which was subsequently fed to the network. On the other hand, multi-temporal 2-stream (MT-2str) considered a different set of convolutions for each epoch and then concatenated these two streams. Multi-temporal 2-stream-FC (MT-2str-FC) is similar to the 2-stream; however, it has a set of two fully connected layers after the concatenation of the two streams. These fully connected layers were intended to merge the different streams of features before the classification layer, instead of considering the concatenation directly (Vo and Hays, 2016).

In addition, dot product and Euclidean distance were also tested instead of the concatenation; however, these performed poorly.

`

Given the small amount of data several fine-tuning approaches were tested: a) *Resnet* (He et al., 2016) with ImageNet weights, b) using built *vs* non-built weights (Duarte et al., 2018b), and c) using a network used for the image classification of debris and rubble piles from aerial images (Duarte et al., 2018a). However, these approaches did not perform as well as training from scratch.

4. EXPERIMENTS AND RESULTS

All the networks were trained with learning rate of 0.1 and weight decay of 10^-4. Only one loss function (binary cross-entropy) was used in each experiment. The classification part of the network was performed using a sigmoid activation over the last layer of a given network. The experiments were performed with early stopping, e.g. when the validation data loss stopped improving.

Data augmentation was performed to further address the issue of low number of samples. However, given the fact that we are using a network with batch normalization, light data augmentation was used (Ioffe and Szegedy, 2015). The skew and zoom were limited to small intervals (+/- 5 deg.) in order to make the damages still visible on the façade image patch. Apart from the zoom and skew, horizontal flips were also used. The same data augmentation parameters are applied to both inputs (pre- and post-event images).

The training and validation split were performed at an image pair level, hence 70% of the image pairs were used for training and 30% for validation. The split also took into account that no façade is present in both training and validation, as several image pairs exist per façade. Since the split was made considering the image pairs, there were 19 damaged façades and 8 not damaged façades used for validation.

The two tables below (Table 2 and Table 3) show the results on validation data: recall precision and accuracy. This is presented both at an image/image-pair level and on a façade level in Table 2. Table 3 presents the TP, TN, FP, FN at a façade level. An error analyses only at an image pair level would not be enough to describe the behaviour of the tested approaches. Hence, the error is also determined at a façade level, which considered a majority filtering on the predictions of the image pairs (or image in the mono-temporal approaches) related to a façade.

Equations 1, 2 and 3 formalize the used accuracy metrics.

$$accuracy = \frac{TP + TP}{\#\ validation\ samples} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

	Image/image-pair level			Façade level		
	Rec.	Prec.	Acc.	Rec.	Prec.	Acc.
MN-trd	0.67	0.50	0.51	0.66	0.63	0.52
MN-fac.	0.52	0.26	0.52	0.83	0.26	0.44
MT-1str	0.76	0.39	0.63	0.37	1.00	0.56
MT-2str	0.75	0.50	0.67	0.92	0.58	0.66
MT-2str-FC	0.73	0.78	0.75	0.92	0.58	0.66

Table 2. Precision, recall and accuracy (0-1) for the defined models. This is given at an image or image pair level and on a facade level.

Overall, from Table 2 the best performing approach was MT-2stream-FC. Regarding only the mono-temporal approaches, the use of a traditional approach was found to be preferable. In MN-fac, the results were worse than randomly assigning a label to the image pairs, i.e. the network was not able to learn from the given input.

While at an image pair level MT-2stream-FC had better results, it did not affect the final result at a façade level when compared with MT-2stream.

From Table 3, the MT-2stream and MT-2strean-FC had the same result with both the higher count of true instances and lower count of false instances.

	Façade level			
	TP	TN	FP	FN
MN-trd	12	2	7	6
MN-fac.	5	7	14	1
MT-1str	7	8	0	12
MT-2str	11	7	8	1
MT-2str-FC	11	7	8	1

Table 3. True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) when evaluating the models at a facade level. 27 façades are considered in validation (19 damaged and 8 not damaged)

Figure 8 shows the results on several façades regarding the multi-temporal approach. The left columns present the pre and post event façade image patches and on the right the activation of the post event façade when fed to the MT-2stream-FC. The coloured bar on the left indicates the prediction for a given image pair, D for damaged and ND for not damaged. The green and red colour stand for a correct or wrong prediction, respectively. The activations were extracted from the activation layer after the last set of convolutions of each of the approaches. The rows *a*, *c* and *f* indicate image pairs that were labelled as damaged, and the remaining as not damaged. Only in the last example the network failed to classify the image pair as damaged. The activations indicate the focus of the network on a given post-event façade image patch. As can be observed, signs of spalling were detected by the multi-temporal approach (row *a* and *b* of Figure 8), while the detection of cracks (row *e* of Figure 8) was unsuccessful.

Figure 9 shows some results of the traditional approach, which for example does not detect the spalling (*a* and *e*), whereas it was detected by the MT-2stream-FC (see Figure 8 a). In spite of detecting the façade patch relative to the cracks (*d*), the traditional approach detected a cluttered part of that façade as damaged. This figure also shows that since this network is trained on image samples containing debris and rubble piles, partial collapses were correctly classified as damaged. In façade *b*) a not damaged area was classified as damaged due to the cluttered scene.

5. DISCUSSION

Overall, the proposed framework using multi-temporal imagery outperformed the mono-temporal approaches when performing the image classification of façade damages. However, the poor results reported in this study reflect the difficulty in the detection of damage along the façades from manned aerial oblique imagery. Especially when these damage evidences are smaller signs of damage such as cracks and small portions of spalling.

`

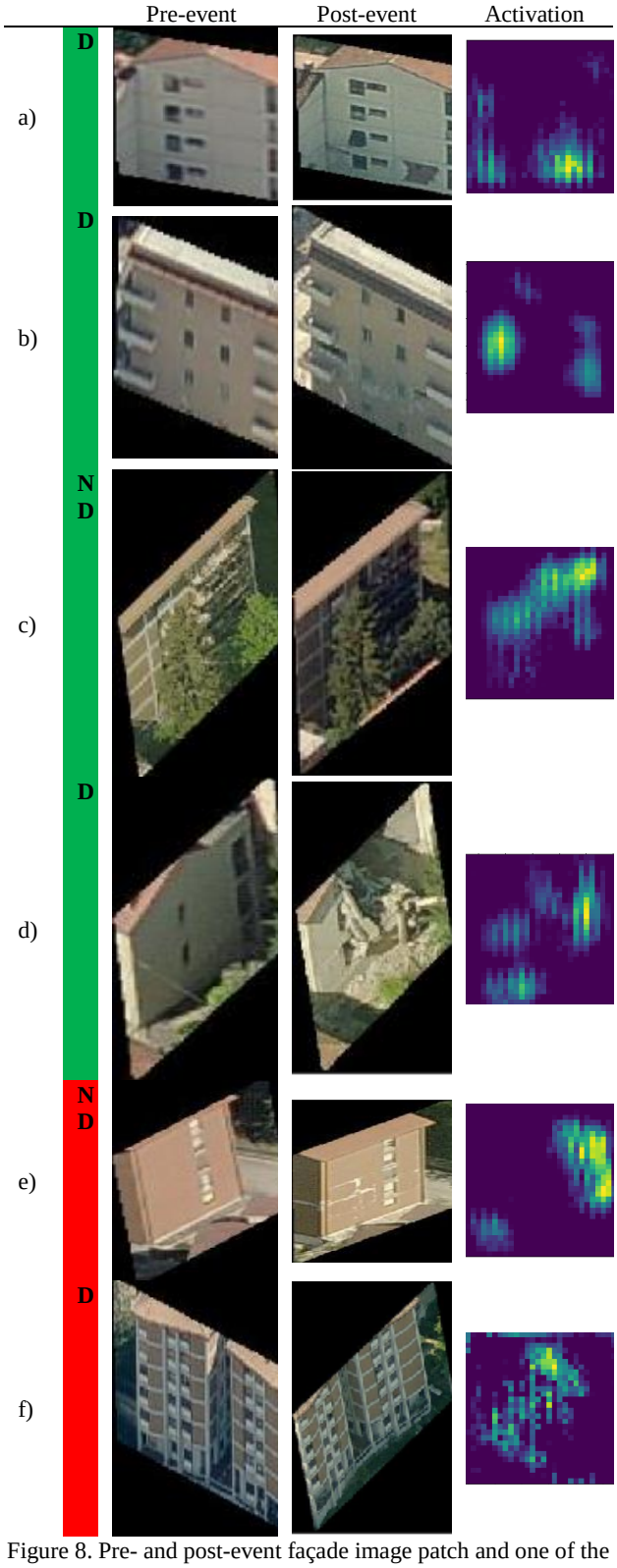

Figure 8. Pre- and post-event façade image patch and one of the activations coming from the activation layer of the last convolutional set of the MT-2str-c. On the left column, red means the prediction was correct and green means the prediction was incorrect. D stands for damaged and ND for not-damaged

The traditional mono-temporal approach, which used a network trained with image samples containing mostly debris and rubble piles resulted in a high rate of false positives and negatives. This can be seen in the accuracy metrics, and also in the examples of predictions shown in the results. This type of behaviour was previously reported in studies that used such type of network for façade damage detection (Duarte et al., 2017). The other mono-temporal approach that considered as training the façades was not able to outperform the traditional mono-temporal approach.

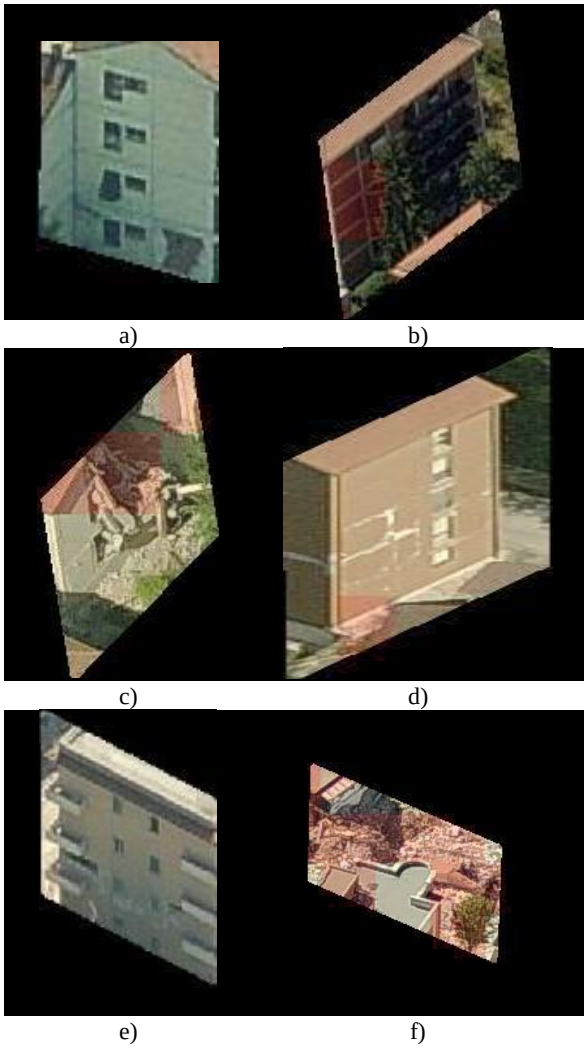

Figure 9. Several examples of classification with the mono-temporal traditional approach. Red squares indicate damaged regions within each façade patch.

As expected, all the multi-temporal approaches outperformed the mono-temporal ones. Specifically, the multi-temporal approach MT-2stream-FC had the best results when considering the evaluation on an image pair level. However, at a façade level the results were the same as MT-2stream.

The results also indicate that it is better to have a set of epoch specific convolutions that are merged at a later stage of the network (see MT-1stream *vs* MT-2stream and MT-2stream-FC). Ahmed et al. (2015) reported that for person re-identification it would be better to have a specific stream for each branch, being merged at a later stage. However, Daudt et al. (2018) and Vo and Hays (2016) reported that it would be preferable to have the images concatenated and then have a single set of convolutions, rather than having epoch specific convolutions for each epoch.

Given the literature and the reported results, it seems that the merging of the streams is application dependent. Hence, for the

`

image classification of façade damages, intra-epoch features play a role in the final damage score.

Furthermore, while the concatenation of the feature maps was the best approach to merge the different streams, this may be further improved as reported in Ahmed et al. (2015). The authors successfully used a cross-input neighbourhood differences which improved the result.

The results shown for the MT-2stream-FC (Figure 8) indicate that in spite of localizing correctly several instances of spalling, it struggled to detect cracks.

Another issue is the occlusions due to urban design. In this study many façades were not visible in the aerial oblique images. Hence, such areas should be detected so a further UAV flight can be planned and the occluded façades surveyed (Nex et al., in review).

In this study a low amount of data was used, with only 88 damaged façades being considered. The overlap on such aerial imagery datasets was taken advantage of by creating multiple image pairs per façade. However, overall the amount of data is still low, which likely impacted the results. This is especially visible in the mono-temporal approach where the network was not able to learn.

6. CONCLUSIONS AND FUTURE WORK

This paper assessed the use of multi-temporal aerial oblique imagery for the image classification of façade damages. Three different multi-temporal approaches using convolutional neural networks were tested and compared against two mono-temporal approaches.

The mono-temporal approaches all performed worse. However, it was found that it is preferable to use a traditional approach trained with image samples depicting rubble piles and debris, rather than use a mono-temporal approach trained only with the set of façades. This may be due to the lack of damaged façades, where the model is not able to generalize the damage appearance given the low amount of data, while the traditional approach can at least identify the partial and totally collapsed façades.

Regarding the multi-temporal approaches, not all the methods performed in the same way. It was found that it is preferable to have an epoch-specific set of convolutions, instead of considering a single stream network where both epochs' inputs are concatenated and then fed to a network. Hence, epoch-specific feature information is valuable for the image classification of façade damages.
A transversal issue to this study were the occlusions due to urban design. In such cases the manned-aerial platform could not observe the façade, where the use of a more directed UAV flight could aid in surveying such occluded façades.

A central aspect of the contribution was to take advantage of the redundant information in multi-view aerial imagery. While the number of damaged façades was low (88), a mean of approximately 25 image pairs per façade is considered due to the redundancy present in such datasets. Only in this way it was possible to test the presented approach.

Despite the better results using multi-temporal imagery, the accuracy is still only 66%. Hence, more research is needed to improve the detection of façade damages. One of the issues might be the low number of damaged façades; non-damaged ones were discarded with the objective of having a balanced set of damaged and non-damaged image pairs for training. In future work this unused set of data should be taken advantage of, for example, with oversampling where new image samples are generated from the initial set of images (Buda et al., 2018). Given the small signs of damage, sometimes only affecting a few pixels, a segmentation and/or localization approach could improve the façade damage detection rate.

ACKNOWLEDGEMENTS

This work was funded by INACHUS (Technological and Methodological Solutions for Integrated Wide Area Situation Awareness and Survivor Localisation to Support Search and Rescue Teams), a FP7 project with grant number: 607522.

REFERENCES

Ahmed, E., Jones, M., Marks, T.K., 2015. An improved deep learning architecture for person re-identification, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 3908–3916. https://doi.org/10.1109/CVPR.2015.7299016

Armesto-González, J., Riveiro-Rodríguez, B., González-Aguilera, D., Rivas-Brea, M.T., 2010. Terrestrial laser scanning intensity data applied to damage detection for historical buildings. Journal of Archaeological Science 37, 3037–3047. https://doi.org/10.1016/j.jas.2010.06.031

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing 140, 20–32. https://doi.org/10.1016/j.isprsjprs.2017.11.011

Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Daudt, R.C., Le Saux, B., Boulch, A., Gousseau, Y., 2018. Urban change detection for multispectral earth observation using convolutional neural networks. Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Valencia, pp. 2115–2118. https://doi.org/10.1109/IGARSS.2018.8518015

Dell'Acqua, F., Polli, D.A., 2011. Post-event only VHR radar satellite data for automated damage assessment. Photogrammetric Engineering & Remote Sensing 77, 1037–1043. https://doi.org/10.14358/PERS.77.10.1037

Duarte, D., Nex, F., Kerle, N., Vosselman, G., 2018a. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. Remote Sensing 10, 1636. https://doi.org/10.3390/rs10101636

Duarte, D., Nex, F., Kerle, N., Vosselman, G., 2018b. Satellite image classification of building damages using airborne and satellite image samples in a deep learning approach, in: ISPRS

`

Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 89–96. https://doi.org/10.5194/isprs-annals-IV-2-89-2018

Duarte, D., Nex, F., Kerle, N., Vosselman, G., 2017. Towards a more efficient detection of earthquake induced facade damages using oblique UAV imagery, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 93–100. https://doi.org/10.5194/isprs-archives-XLII-2-W6-93-2017

Fernandez Galarreta, J., Kerle, N., Gerke, M., 2015. UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. Natural Hazards and Earth System Science 15, 1087–1101. https://doi.org/10.5194/nhess-15-1087-2015

Gerke, M., Kerle, N., 2011. Automatic structural seismic damage assessment with airborne oblique Pictometry© imagery. Photogrammetric Engineering & Remote Sensing 77, 885–898. https://doi.org/10.14358/PERS.77.9.885

Gokon, H., Post, J., Stein, E., Martinis, S., Twele, A., Muck, M., Geiss, C., Koshimura, S., Matsuoka, M., 2015. A method for detecting buildings destroyed by the 2011 Tohoku earthquake and tsunami using multitemporal TerraSAR-X data. IEEE Geoscience and Remote Sensing Letters 12, 1277–1281. https://doi.org/10.1109/LGRS.2015.2392792

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: CVPR. https://doi.org/10.1109/CVPR.2016.90

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift, in: 34th International Conference on Machine Learning. Sydney, Australia.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems. pp. 1907–1105.

Nex, F., Duarte, D., Kerle, N., Steenbeek, A., in review. Towards real-time building damage mapping with low-cost UAV solutions.

Rupnik, E., Nex, F., Remondino, F., 2014. Oblique multi-camera systems orientation and dense matching issues. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-3/W1, 107–114. https://doi.org/10.5194/isprsarchives-XL-3-W1-107-2014

Sui, H., Tu, J., Song, Z., Chen, G., Li, Q., 2014. A novel 3D building damage detection method using multiple overlapping UAV images. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL–7, 173–179. https://doi.org/10.5194/isprsarchives-XL-7-173-2014

Tu, J., Sui, H., Feng, W., Sun, K., Xu, C., Han, Q., 2017. Detecting building façade damage from oblique aerial images using local symmetry feature and the Gini Index. Remote Sensing Letters 8, 676–685. https://doi.org/10.1080/2150704X.2017.1312027

Vetrivel, A., Duarte, D., Nex, F., Gerke, M., Kerle, N., Vosselman, G., 2016a. Potential of multi-temporal oblique airborne imagery for structural damage assessment. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III–3, 355–362. https://doi.org/10.5194/isprsannals-III-3-355-2016

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2017. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS Journal of Photogrammetry and Remote Sensing. https://doi.org/10.1016/j.isprsjprs.2017.03.001

Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2016b. Identification of structurally damaged areas in airborne oblique images using a Visual-Bag-of-Words approach. Remote Sensing 8, 231. https://doi.org/10.3390/rs8030231

Vo, N.N., Hays, J., 2016. Localizing and orienting street views using overhead imagery, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Amsterdam, pp. 494–509. https://doi.org/10.1007/978-3-319-46448-0_30

Wang, Q., Zhang, X., Chen, G., Dai, F., Gong, Y., Zhu, K., 2018. Change detection based on Faster R-CNN for high-resolution remote sensing images. Remote Sensing Letters 9, 923–932. https://doi.org/10.1080/2150704X.2018.1492172

Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., 2017. Multisource remote sensing data classification based on convolutional neural network. IEEE Transactions on Geoscience and Remote Sensing 1–13. https://doi.org/10.1109/TGRS.2017.2756851

Yu, F., Koltun, V., 2016. Multi-scale context aggregation by dilated convolutions, in: ICLR.

Yu, F., Koltun, V., Funkhouser, T., 2017. Dilated residual networks, in: CVPR.

`