SATELLITE IMAGE CLASSIFICATION OF BUILDING DAMAGES USING AIRBORNE AND SATELLITE IMAGE SAMPLES IN A DEEP LEARNING APPROACH

D.Duarte [a*], F.Nex [a], N. Kerle [a], G. Vosselman[a]
[a] 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, WGII/4

KEY WORDS: multi-resolution, dilated convolutions, residual connections, multi-scale, multi-platform, machine learning, UAV, earthquake

ABSTRACT:
The localization and detailed assessment of damaged buildings after a disastrous event is of utmost importance to guide response operations, recovery tasks or for insurance purposes. Several remote sensing platforms and sensors are currently used for the manual detection of building damages. However, there is an overall interest in the use of automated methods to perform this task, regardless of the used platform. Owing to its synoptic coverage and predictable availability, satellite imagery is currently used as input for the identification of building damages by the International Charter, as well as the Copernicus Emergency Management Service for the production of damage grading and reference maps. Recently proposed methods to perform image classification of building damages rely on convolutional neural networks (CNN). These are usually trained with only satellite image samples in a binary classification problem, however the number of samples derived from these images is often limited, affecting the quality of the classification results. The use of up/down-sampling image samples during the training of a CNN, has demonstrated to improve several image recognition tasks in remote sensing. However, it is currently unclear if this multi resolution information can also be captured from images with different spatial resolutions like satellite and airborne imagery (from both manned and unmanned platforms). In this paper, a CNN framework using residual connections and dilated convolutions is used considering both manned and unmanned aerial image samples to perform the satellite image classification of building damages. Three network configurations, trained with multi-resolution image samples are compared against two benchmark networks where only satellite image samples are used. Combining feature maps generated from airborne and satellite image samples, and refining these using only the satellite image samples, improved nearly 4% the overall satellite image classification of building damages.

1. INTRODUCTION AND RELATED WORK

Building damage maps have been recurrently used in the response and recovery phase of the disaster management cycle. Damaged buildings may be a proxy for victim localization (Dell'Acqua and Gamba, 2012) and their identification can also aid to plan and delineate recovery activities (Eguchi et al., 2009). Remote sensing has been extensively used to perform the damage assessment of a given region affected by a disastrous event (Dell'Acqua and Gamba, 2012; Dong and Shan, 2013; Gerke and Kerle, 2011; Vetrivel et al., 2017). The platforms used in remote sensing usually have a wide coverage, fast deployment and high temporal frequency while the collected data allow to automate building damage assessment procedures (Ural et al., 2011).

A wide variety of remote sensing sensors mounted on different platforms have been used to map building damages (Armesto-González et al., 2010; Dell'Acqua and Polli, 2011; Gokon et al., 2015; Khoshelham et al., 2013; Marin et al., 2015; Vetrivel et al., 2017). However, there has been a growing interest regarding the use of images (Curtis and Fagan, 2013; Fernandez Galarreta et al., 2015; Vetrivel et al., 2015, 2016a, 2017).

In this regard, synoptic satellite imagery can be readily available and provide the first overview over a region struck by a disastrous event such as an earthquake (Dell'Acqua and Gamba, 2012). The International Charter (IC) (Bessis et al., 2004) and the Emergency Management Service (EMS) (Copernicus programme, European Commission), are two institutions which use such imagery to provide geoinformation to regions affected by disasters. The IC and EMS mostly rely on the manual interpretation of satellite images to identify damaged buildings, despite the amount of proposed automated methods. However, scene characteristics, cloud cover, limited resolution and viewpoint, limited time by map producers to develop new operational methods; hinder the automation of these procedures (Kerle, 2010; Vetrivel et al., 2016a).

Other platforms coupled with cameras have also been used to map damages (Sui et al., 2014; Vetrivel et al., 2016b). Manned and unmanned aerial vehicles (UAV) enable the acquisition of images at a higher-resolution and can also perform oblique flights, introducing another level of damage information regarding the façades (Tu et al., 2017). In this regard, the Joint Research Center (JRC, European Commission) awarded a contract in 2015 to a consortium of private companies for the provision of aerial imagery after a disastrous event within a European context ("CGR supplies aerial survey to JRC for emergency," n.d.). UAV images have become a normal source of information for many rescue teams in the recent earthquakes in Nepal (2015) and Italy (2016). These trends have pushed many researchers (Duarte et al., 2017; Sui et al., 2014; Vetrivel et al., 2017) to develop damage detection algorithms exploiting these high-resolution images.

The use of overlapping images may allow the generation of 3D point clouds through dense image matching. The set of geometrical information extracted from point clouds can be used alongside the images for the detection of building damages (Fernandez Galarreta et al., 2015; Vetrivel et al., 2017). Their added value can be marginal if single epoch data are considered (Duarte et al., 2017; Vetrivel et al., 2017). Furthermore, the generation of 3D point clouds is still very time consuming, hindering their use in early response tasks. The quality of these 3D data is directly related with the resolution of the input images, which limits the use of the 3D generated from satellite imagery.

The achieved results regarding the use of airborne and UAV images are promising and their use is drastically increasing in recent years. However, satellite images are still the first and most common source for damage assessment. For this reason, a more reliable method to automate the detection from these images would be needed.

The most recent approaches to perform satellite image classification of building damages use CNN (Vetrivel et al.,

2016a). The used networks are very similar to the ones used in the computer vision domain (Krizhevsky et al. 2017). Satellite image samples are used for the training of the network, in a binary classification scheme (i.e. damaged and not damaged areas). However, the number of samples from satellite images is relatively small, while a wide variety of images acquired with airborne platforms, both manned and unmanned, are available too. These data are currently used to train a network which classifies images with the same resolution (Vetrivel et al., 2017). In computer vision and remote sensing, the use of multi-resolution data has demonstrated to improve the overall image classification and segmentation (Fu et al., 2017; Hamaguchi et al., 2017; Lin et al., 2016; Liu et al., 2016). The multi-resolution training is usually performed artificially (Fu et al., 2017; Hu et al., 2015; Li et al., 2015; Shen et al., 2015; Tang and Mohamed, 2012), up/down sampling the images at several scales. However, a multi-resolution approach using image data from different platforms and sensors has not been tested yet.

The aim of this paper is to assess if the combined use of different resolution images improves the image classification of building damages from satellite images using CNN (Figure 1).

Figure 1 Examples of damaged and undamaged regions in a) UAV (Pescara del Tronto,Italy, 2016), b) satellite (WorldView 3, Amatrice, Italy, 2016) and c) manned aerial vehicles (St Felice, Italy, 2012) imagery.

The main idea is that the native multi-resolution information of remote sensing imagery (i.e. satellite and airborne) can be captured by a CNN, improving the satellite image classification. Several CNNs configurations have been tested to assess how the image samples from different resolutions can influence the performance of the classification of building damages. Two recent developments in the computer vision domain are used: residual connections and dilated convolutions. More details regarding the developed approach are described in Section 2. This is then followed by an experiments section (3) which details the datasets (Section 3.1) used to test the approach, presents the experiments (Section 3.2) and the achieved results (Section 3.3). The discussion and the conclusions are finally given in Section 4 and Section 5 respectively.

2. METHODOLOGY

Five different CNN architectures are defined. Two are used as benchmark and the remaining three are used to test the multi-resolution approach. Regarding the benchmark networks, the first is trained from scratch and the other one is fined-tuned on the generic satellite image samples provided by Cheng et al. (2017). The three multi-resolution test networks have been conceived to analyze the best way to combine and exploit features from each image resolution level.

All the networks take advantage of residual connections and dilated convolutions. This section explains these two central components of the networks while the two basic modules of the networks are then described in Section 2.1. The networks architectures used in the tests are finally presented in Section 3.

Residual connections: The depth of CNN have shown an increase in their capabilities to retrieve relevant information from images (Telgarsky, 2016). The usual hierarchical stacking of convolutional layers allows the network to learn from lower level features to higher levels of abstraction. Nonetheless, a given layer *l* may need feature information not only from the layer *l-1* but also from other previous layers (*l-2*, etc.). Residual connections (He et al., 2016) enable this process, by feeding a given layer to the previous one, as in the classical hierarchical approach, summed with a given output of earlier layers (Figure 2). In this way, every level of a given residual network effectivilly contributes to the final recognition task. Figure 2 shows a scheme of a residual connection and its interactions within a network. In this approach, features are extracted from remote sensing imagery at different spatial resolutions, where the relevance and complexity of a given feature may vary between the considered resolution levels. Thus, it is mandatory to capture and retain all of these levels of feature complexity through the use of residual connections.

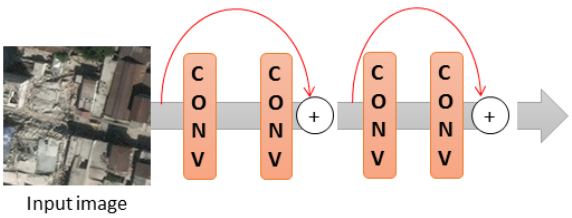

Figure 2 Simple scheme of possible residual connections within a CNN. The grey arrow shows a classical approach, while the red arrows show the new added (residual) connections.

Dilated convolutions: Another central aspect of a network capable of capturing multi-resolution information is its ability to capture spatial context. Recently, Yu and Koltun (2016) proposed the use of dilated convolutions (Figure 3) in CNN. These dilated convolutions consist of convolutions applied to a given input image with a kernel having defined gaps (Figure 3). The receptive field of the network is bigger, capturing more contextual information (Hamaguchi et al., 2017). These dilated convolutions allow the integration of knowledge of the wider context (Hamaguchi et al., 2017) and at the same time depict finer details (Yu and Koltun, 2016). This is especially relevant for a multi-resolution approach since several sizes of patterns at different resolutions may contribute to the classification task.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-89-2018 | © Authors 2018. CC BY 4.0 License.

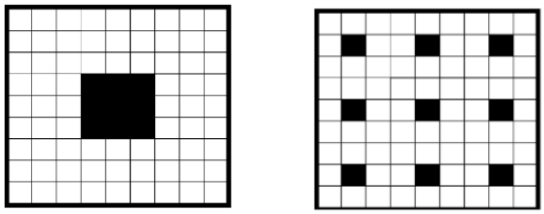

Figure 3 a) 3x3 kernel with dilation 1, b) 3x3 kernel with dilation 3

2.1 Basic convolutional set and modules definition:

The architecture of the CNN is composed by two main modules: 1) context module, followed by 2) resolution specific module (Figure 5). This structure was inspired by the works of Hamaguchi et al. (2017), Yu et al. (2017) and He et al. (2016). The general idea is that both context and resolution specific information is needed (Hamaguchi et al., 2017), hence the use of the two distinct modules.
Both modules are built stacking basic convolutional sets. These are composed of a convolution, batch normalization and ReLU (CBR, see Figure 4 a)) (He et al., 2016; Ioffe and Szegedy, 2015; Yu et al., 2017). Two basic convolutional sets bridged by a residual connection form a main CBR block, as shown in Figure 4 b). In each CBR, different number of filters and dilation values can be adopted. Both the context and resolution specific modules are composed of a sequence of CBRs with different numbers of filters and dilation rates as indicated in Figure 5.

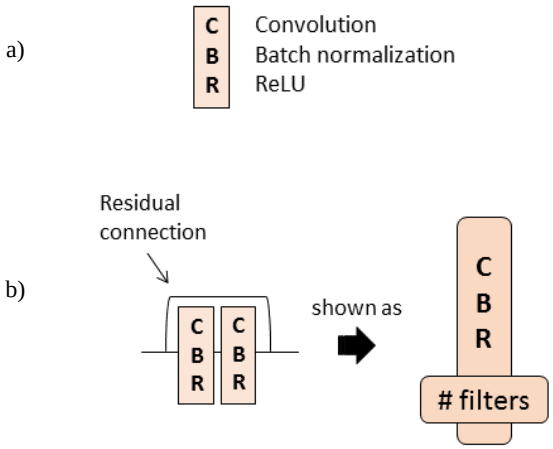

Figure 4 Basic convolutional set (a). Basic group of convolutions used to build the context and (b) resolution specific modules indicating the number of filters used

The context module (Figure 5 a)) is composed of several stacked CBRs with increasing dilation and increasing number of filters, with the objective of gradually capturing larger feature representations (Hamaguchi et al., 2017; Yu et al., 2017). The increasing number of filters over a CNN follows the state of the art approaches (He et al., 2016; Simonyan and Zisserman, 2015), more filters for higher level feature representation. The initial feature map is reduced from 224x224 (input) to 28x28px using a stride of 2, instead of 1 in the first three sets of CBRs. The use of larger stride has shown better performances than the max pooling operations, mainly because of the use of dilated convolutions (Yu et al., 2017). The kernel size of all the convolutions is 3x3 (Springenberg et al., 2015).

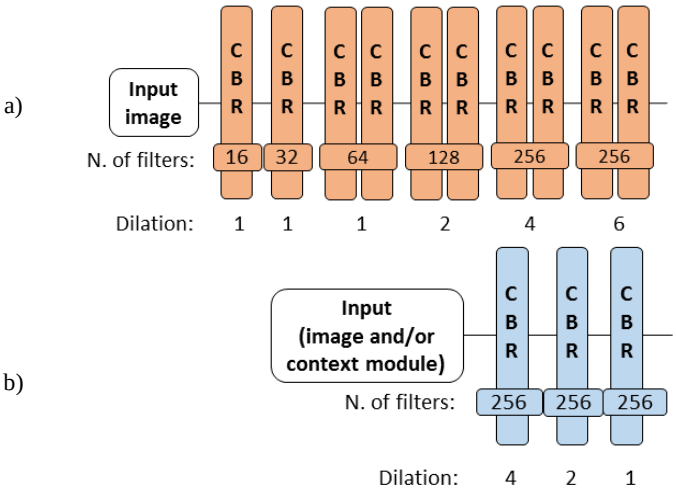

Figure 5 a) Context module, b) resolution specific module. Resolution specific module does not contain residual connections.

The increase in the dilation factor can create artificacts on the resulting feature maps, due to the gaps generated by the dilated kernel (Hamaguchi et al., 2017; Yu et al., 2017). To attenuate this drawback, the dilation increase in the context module is compensated in the resolution specific module with a gradual reduction of the dilation value and the removal of residual connections from the basic CBR blocks (Yu et al., 2017). This also allows to re-capture the more local features (Hamaguchi et al. 2017), which might be lost due to the increasing dilations in the context module.
For the classification part of the network, global average pooling followed by a convolution which maps the feature map size to the number of classes, is applied. Since this is a binary classification problem, a sigmoid function is used as activation.

3. EXPERIMENTS

3.1 Dataset and training samples

There are two subsets of data: a) a multi-resolution dataset formed by three sets of images corresponding to satellite and airborne images (manned and UAV platforms) and b) a set of generic satellite image samples, which is used in one of the benchmark approaches.
Regarding the multi-resolution data, three sets of images, one set for each level of resolution, are considered: satellite, manned and unmanned aerial vehicles (Table 1). Most of the datasets depict real earthquake-induced building damages; however, there are also images from controlled demolitions.
The satellite images cover five different geographical locations in Italy, Ecuador and Haiti (Table 1). The satellite imagery was collected with WorldView 3 (Amatrice, Pescara del Tronto and Portoviejo) and GeoEye 1 (L’Aquila, Port-au-Prince). These data are pansharpened and have a variable resolution between 0.4 and 0.6m.
The airborne imagery consists of nadir and oblique imagery with a ground sampling distance (GSD) of 12-18 cm for the manned vehicles and of 2-10 cm for the UAV. The differences in image content at a given level of resolution (different illumination settings, view angles, sensors characteristics, morphology of buildings and urban landscape) are further increased by the multi-resolution aspect.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-89-2018 | © Authors 2018. CC BY 4.0 License.

Location	N. of samples		Month/Year of event
	Damaged	Not damaged	
	Satellite samples		
Aquila	115	118	April 2009
Port-au-Prince	732	701	January 2010
Portoviejo	147	163	April 2016
Amatrice	165	180	August 2016
Pesc. Tronto	93	94	August 2016
Total	**1252**	**1256**	
	Airborne (manned vehicles) samples		
Aquila,	336	385	April 2009
St Felice	587	593	May 2012
Amatrice	320	362	August 2016
Tempera	259	260	April 2009
Bidonville	229	229	January 2010
Port-au-Prince	749	712	January 2010
Onna	387	365	April 2009
Total	**2867**	**2906**	
	Airborne (UAV) samples		
Aquila	113	131	April 2009
Wesel	90	94	++
Portoviejo	216	208	April 2016
Pesc. Tronto	218	264	August 2016
Katmandu	309	288	April 2015
Taiwan	187	611	February 2016
Gronau	457	501	++
Mirabello	502	453	May 2012
Lyon	312	310	++
Total	**2704**	**2860**	

Table 1 Overview of the location and quantity of satellite and airborne samples. The ++ locations indicate controlled demolitions of buildings.

The samples are extracted for each resolution from the set of images indicated before. First, damaged and undamaged image regions are manually delineated, see Figure 6. Every cell that contains more than 60% of its area covered by one of the classes is cropped and used as an image sample for that same class. The grid size varies according to the resolution: satellite 80x80px, airborne (manned vehicles) 100x100px and airborne (UAV) 160x160px. The variable size of the image samples is set in order to keep in count the different resolution and the extension of the area captured in each patch. Due to the scarcity of satellite image samples (Table 1), to consider a smaller patch in this level of resolution, allowed to extract a higher number of samples.
The number of samples is approximately the same for the damaged and undamaged classes. However, the number of samples is not balanced among the 3 levels of resolution. The number of satellite image samples is two-fold lower when compared to the other two levels of resolution.

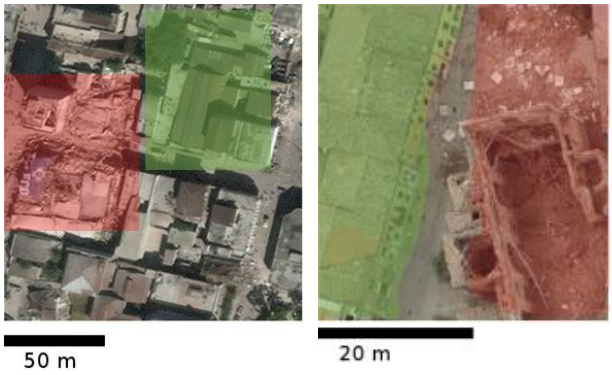

Figure 6 Examples of damaged (red) and non-damaged (green) areas digitized in satellite (GeoEye 1, Port-au-Prince, Haiti, 2010), left. Airborne (manned platform) (St Felice, Italy, 2012) imagery, right.

The generic satellite images samples are taken from a freely available benchmark dataset: NWPU-RESISC45 (Cheng et al., 2017). This benchmark dataset contains 45 classes with 700 satellite image samples per class. From these, fourteen classes were selected and divided into two broader classes, built and non-built (Table 2). Instead of considering the total 31500 samples, only fourteen classes are considered (9800) to reduce the computational cost of the approach

Built	Non-built
Airport	Beach
Commercial area	Circular farmland
Dense residential	Desert
Freeway	Forest
Industrial area	Mountain
Medium residential	Rectangular farm
Sparse residential	Terrace

Table 2 Fourteen classes of the benchmark dataset (NWPU-RESISC45) divided in built and non-built classes. Each class contains 700 samples, totalling 9800 image samples.

3.2 Experiments

Using the modules defined before in section 2.2, five different networks are derived from the architectures shown in Figure 7. The first two networks are used as benchmarks for the other tests involving the multi-resolution architecture. In the first benchmark network (Figure 7 a)), the satellite training samples are fed into a network composed of the context module and the resolution specific module. The second benchmark uses the same architecture as defined in Figure 7 c) (mresB). It feeds the generic satellite image samples (Table 2) into the context module, while the resolution specific is only fed with the satellite samples. Due to the low number of damage domain satellite image samples (2508) when compared to the other levels of resolution (around 5700), training a network from scratch may not be optimal (Tajbakhsh et al., 2016). For this reason, the second benchmark (henceforth referred as benchmark_ft), fine tunes the learned features from generic satellite samples, with damage domain specific satellite image samples.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-2-89-2018

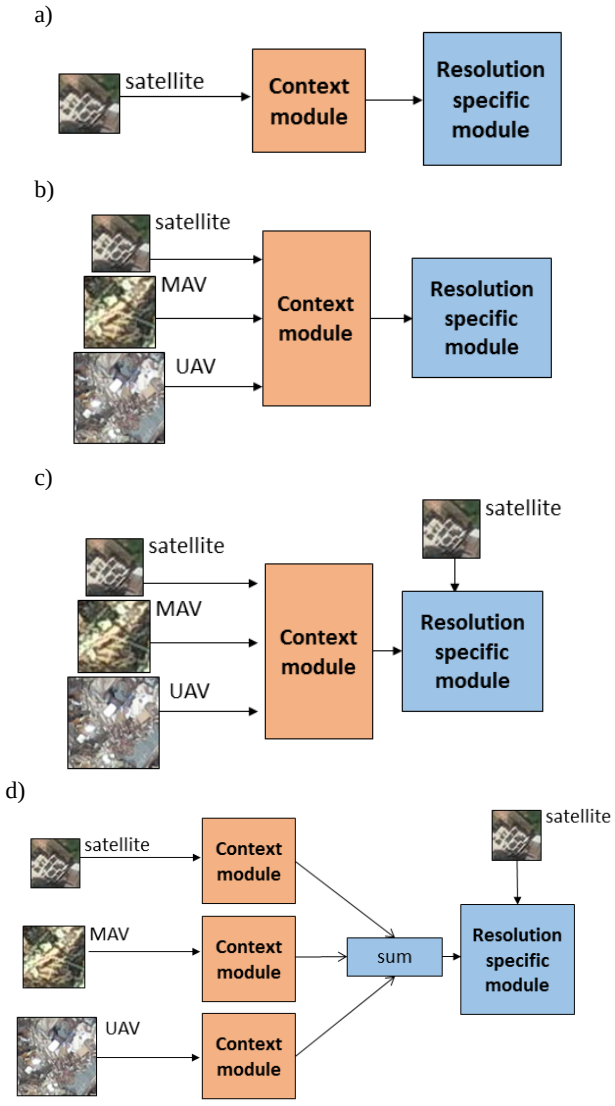

Figure 7 Tested network configurations: a) benchmark, b) multi-resolution A (mresA), c) multi-resolution B (mresB) and d) multi-resolution C (mresC). Details on the text.

The other three networks combine both the context and the resolution specific modules. The overall aim of these tests is to understand if sharing features between resolutions (Figure 7 b) and c)) captures more relevant information than merging the output of each separate context module (Figure 7 d)). A more detailed explanation of these three networks is given below:

mresA: feeds the training data of all resolutions to the context followed by the resolution specific module. In this way the extracted features of both modules are shared between resolutions, Figure 7 b).

mresB: all the training data of all resolutions are fed into the context module. However, the resolution specific module is only fed with the satellite samples. In this case the context module serves as base model with its weights that are tuned in the resolution specific module, Figure 7 c).

mresC: each data resolution is given to a different context module. The output of these modules is subsequently summed. These summed feature maps are used to initialize the resolution specific module that considers only satellite image samples, Figure 7 d).

The stockastic gradient descent (Wilson et al., 2017), with momentum of 0.9 and with a decreasing learning rate, is used in the optimization. The initial learning rate is of 10^{-2}, decreasing by a factor of 10 every 30 epochs (total of 120), with a weight decay of 10^{-2}. This is set for the benchmark and mresA networks. For the other two networks, the context and resolution specific modules are executed separately. In these cases, the context module is performed with the same learning rate parameters of the benchmark and mresA. However, the resolution specific learning rates differ. The mresB (and benchmark_ft) resolution specific module has the learning rate initially set at 10^{-3}, decreasing by a factor of 10 every 30 epochs, with a weight a decay of 10^{-6}. In the case of the mresC the learning rate is set initially to 10^{-4}, with the same decreasing rate and weight decay as mresB. These parameters are obtained empirically. In the benchmark and mresA the networks are learning from scratch, hence the aggressive learning rate. While in the benchmark_ft,mresB and mresC, the resolution specific module intends to take advantage of the weights obtained by the context module, hence the lower learning rate parameters. In this way, the multi-resolution context information is refined for the specific case of the satellite image classification of building damages.

During the training of every network, data augmentation is performed since this has shown to avoid overfitting and improve the overall image classification (Krizhevsky et al., 2017; Simonyan and Zisserman, 2015). The used data augmentation consists of random translations, rotations, image normalization and up/downsampling of the images. The networks were run for 120 epochs with a batch size of 8. The input size for the network is 224x224px. The image samples are zero padded to fit in this template, instead of being resized (Vetrivel et al., 2016a).

The training is performed using 70% of the samples of each resolution, while the validation uses 30% of the satellite image samples. This ratio is applied to each location separatly. The selected samples for both the training and validation remains the same for all the experiments.

3.3 Results

The achieved results of the use of the five network architectures are presented below in Table 2.

Network	Accuracy	Parameters	Training samples
benchmark	0.905	8.6M	1718
benchmark_ft	0.904	8.6M	11518
mresA	0.898	8.6M	8685
mresB	0.924	8.6M	8685
mresC	0.944	18.4M	8685

Table 3 Results of experiments

As indicated in this table, the benchmark network trained from scratch (benchmark) marginally outperforms the one which used generic satellite image samples in the context module and posteriorly fine-tuned it with the damage domain samples (benchmark).

Most of the multi-resolution approaches overcome the benchmark networks. Only mresA underperformed the two benchmark networks. The best performing network was mresC with an accuracy increase of almost 4% compared to the benchmark. This network also outperformed mresB by 2%. The network mresC is also the one with the higher number of parameters since 3 context modules were added before the resolution specific module. The number of training samples is of

1718 for the benchmark 11518 for the benchmark_ft and 8685 for the rest of the networks.
To better understand and validate the networks behaviour, a second test was conducted by feeding them with new and unused satellite image patches. These input patches were of 224x224 px (i.e. different from the sample sizes of 80x80 px). Figure 8 and Figure 9 show activations given by the last set of filters of all the multi-resolution networks and the benchmark one with the higher accuracy (benchmark, Table 2). In particular, for each network and from the set of 256 feature maps, the one with the higher average activation value is visualized.

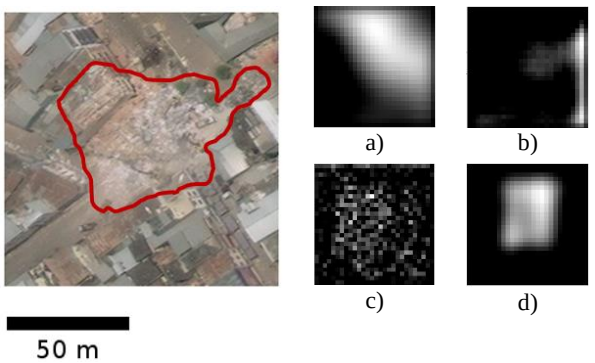

Figure 8 Satellite image sample (collected with WorldView-3, Porto Viejo, Ecuador, 2016), with damaged area manually outlined in red, fed into the network. Higher activation value of the last set of feature maps of the benchmark b), mresA c), mresB d) and mresC e) networks.

The activation from mresC (Figure 8 d)) shows a stronger agreement with the damaged area in red, when considering all the presented activations. However, smaller damaged areas are not considered as damaged. The activation from the benchmark (Figure 8 a)) also shows localization capabilities, but it is less discriminative in correspondence of non-damaged areas. Figure 8 b) presents the activation from mresA, where some difficulty to localize the damaged area from the given patch is evident. The mresB (Figure 8 c)), fails to localize the damage.

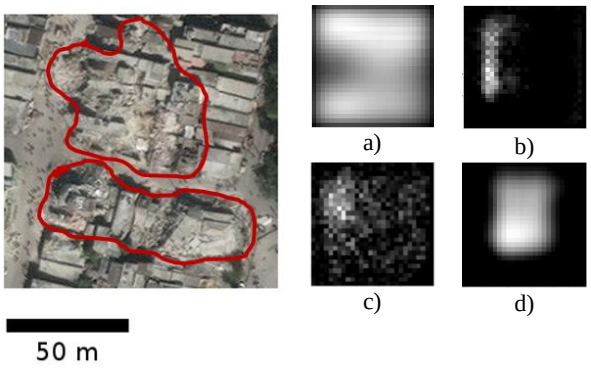

Figure 9 Satellite image sample, with the damage manually outlined in red (GeoEye 1, Port-au-Prince, Haiti, 2010) fed into the network. Higher activation value of the last set of feature maps of the benchmark a), mresA b), mresB c) and mresC d) networks

Another example is presented in Figure 9, left. In this case the mresC activation (Figure 9, d)), from the four activations, is the one that shows the better agreement with the damaged region. As in the previous case, there are smaller damage regions that are not identified in the activation. The benchmark (Figure 9, a)) activation goes across the whole image sample, including areas which are not damaged. mresA (Figure 9, b)) and mresB (Figure 9, c) only focus on the damaged area on the left upper part of the sample.
Both figures, mresA and mresB, present noisier activations than the benchmark and the mresC.

4. DISCUSSION

The presented results indicate an improvement in the satellite image classification of building damages thanks to the use of different training samples from different spatial resolutions.
Only one multi-resolution network did not improve the classification accuracy compared to the used benchmarks. Two factors could have contributed to this: 1) this network was the only multi-resolution network where the resolution specific module was not trained only considering the satellite image samples; 2) the number of satellite training samples is twofold lower if compared with the other two resolutions. This might have led the networks to discard features which might be relevant for the satellite resolution.
The other two networks, which take input samples from all the resolution levels in their context module, outperform the benchmark tests. In this regard, the sum of the feature maps coming from the context module of each of the resolutions (mresC) seems to be more beneficial than feeding all of them into the same context module (mresB). In the case the context module is shared, the network might discard satellite features, due to an unbalanced number of training samples between the different image resolutions. This is in agreement with other remote sensing studies where the up/down sampled image samples are fed into a different network (or parts of the network) and each feature map is then summed to provide a stronger classifier (Fu et al., 2017; Maggiori et al., 2017). The number of parameters is also higher in the best performing resolution; this might have a positive effect on the performance.
Considering previous works (Vetrivel et al., 2016a), there was an increase (around 15%) in the accuracy of satellite image classification of building damages, even without considering the multi-resolution aspect. This accuracy difference is, however, closely related with recent advancements in the image classification algorithms using CNN (He et al., 2016; Krizhevsky et al., 2017).
The activation maps confirm the results provided by the accuracy assessment; also in this case mresC outperform the other methods. However, the activations of this network appear to be smoother; smaller signs of damage might not be considered. In contrast, the activation maps of networks which shared the context module present a noisier activation and seem to generate artefacts as indicated in Hamaguchi et al. (2017) and Yu et al. (2017), even after decreasing the dilation value in the resolution specific module.
The learning rate was found to be critical. The used parameters were tuned empirically and a small change in the parameter values showed to have a high impact on the final result. The presented results represent the best accuracy values achieved with each network configuration.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper assessed the combined use of remote sensing imagery with different resolutions within a CNN approach, to perform the satellite image classification of building damages.
The combined use of several resolutions and their different combination in the training of the CNN, improved the accuracy of satellite image classification of building damages by nearly 4%. The addition of feature maps from the different resolutions has shown to capture more relevant information than having these shared in a single network. The activations of the best performing network, which sums the feature maps coming from the several resolutions, have shown a better agreement with

manually defined damaged regions. However, the activations also show that this network is not able to identify smaller signs of damage, which can be critical for any decision maker considering a damaged map generated by such an automated approach.

Since the shown results are only related with the overall accuracy and behaviour of the networks, more research is needed to assess in which specific conditions this multi-resolution approach improves damage mapping. The datasets used in this experiment mostly refer to the same geographical regions (Haiti and Italy) and the same disastrous events, which could be one of the reasons for the reported results.

With the expected increase in the amount of collected imagery from several different platforms (both manned and unmanned platforms), this multi-resolution aspect of CNN can be beneficial in many practical cases. The trained networks would be very useful in the damage assessment at regional level, where satellite images are currently the only used source of information. This model could be further refined adding location specific samples in an online learning approach (Vetrivel et al., 2016a). In an early post-disaster setting, this multi-resolution capability is even more meaningful, due to the different sources of imagery that might be collected. While satellite may be the first set of available data, there is a continuous capture of airborne multi-resolution data from the initial stages of the response phase.

New tests will be performed using the same number of samples for every resolution. This would allow to better understand the impact of using unbalanced number of data with different resolutions. The use of only airborne samples as training to classify damages from satellite imagery will be then considered in order to assess the transferability of learned features to different resolutions.

The successful use of multi-resolution remote sensing image samples should also be extended to other image classification problems with more classes. There is an increasing amount of multi-resolution image data available and, in that sense, a multi-resolution approach taking advantage of such large amount of data would be beneficial.

6. ACKNOWLEDGEMENTS

The 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.

The authors would like to thank the DigitalGlobe Foundation (www.digitalglobefoundation.com) for providing satellite images on Italy and Ecuador.

7. REFERENCES

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

Bessis, J.-L., Béquignon, J., Mahmood, A., 2004. The International Charter “Space and Major Disasters” initiative. Acta Astronautica 54, 183–190. https://doi.org/10.1016/S0094-5765(02)00297-7

CGR supplies aerial survey to JRC for emergency [WWW Document], n.d. . CGR spa. URL http://www.cgrspa.com/news/cgr-fornira-il-jrc-con-immagini-aeree-per-le-emergenze/ (accessed 11.9.15).

Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE 1–19. https://doi.org/10.1109/JPROC.2017.2675998

Curtis, A., Fagan, W.F., 2013. Capturing damage assessment with a spatial video: an example of a building and street-scale analysis of tornado-related mortality in Joplin, Missouri, 2011. Annals of the Association of American Geographers 103, 1522–1538. https://doi.org/10.1080/00045608.2013.784098

Dell’Acqua, F., Gamba, P., 2012. Remote sensing and earthquake damage assessment: experiences, limits, and perspectives. Proceedings of the IEEE 100, 2876–2890. https://doi.org/10.1109/JPROC.2012.2196404

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

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS Journal of Photogrammetry and Remote Sensing 84, 85–99. https://doi.org/10.1016/j.isprsjprs.2013.06.011

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

Eguchi, R.T., Huyck, C.K., Ghosh, S., Adams, B.J., McMillan, A., 2009. Utilizing new technologies in managing hazards and disasters, in: Showalter, P.S., Lu, Y. (Eds.), Geospatial Techniques in Urban Hazard and Disaster Analysis. Springer Netherlands, Dordrecht, pp. 295–323. https://doi.org/10.1007/978-90-481-2238-7_15

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

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sensing 9, 498. https://doi.org/10.3390/rs9050498

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

Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T., Hikosaka, S., 2017. Effective use of dilated convolutions for segmenting small object instances in remote sensing images arXiv:1709.00179.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. IEEE, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing 7, 14680–14707. https://doi.org/10.3390/rs71114680

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

Kerle, N., 2010. Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? International Journal of Applied Earth Observation and Geoinformation 12, 466–476. https://doi.org/10.1016/j.jag.2010.07.004

Khoshelham, K., Oude Elberink, S., Sudan Xu, 2013. Segment-Based classification of damaged building roofs in aerial laser scanning data. IEEE Geoscience and Remote Sensing Letters 10, 1258–1262. https://doi.org/10.1109/LGRS.2013.2257676

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90. https://doi.org/10.1145/3065386

Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection. IEEE, pp. 5325–5334. https://doi.org/10.1109/CVPR.2015.7299170

Lin, G., Shen, C., Hengel, A. van den, Reid, I., 2016. Efficient piecewise training of deep structured models for semantic segmentation. IEEE, pp. 3194–3203. https://doi.org/10.1109/CVPR.2016.348

Liu, W., Rabinovich, A., Culurciello, E., 2016. Parsenet: looking wider to see better, in: ICLR 2016. Presented at the ICLR 2016.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing 55, 645–657. https://doi.org/10.1109/TGRS.2016.2612821

Marin, C., Bovolo, F., Bruzzone, L., 2015. Building change detection in multitemporal very high resolution SAR images. IEEE Transactions on Geoscience and Remote Sensing 53, 2664–2682. https://doi.org/10.1109/TGRS.2014.2363548

Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J., 2015. Multi-scale convolutional neura lnetworks for lung nodule classification, in: Ourselin, S., Alexander, D.C., Westin, C.-F., Cardoso, M.J. (Eds.), Information Processing in Medical Imaging. Springer International Publishing, Cham, pp. 588–599. https://doi.org/10.1007/978-3-319-19992-4_46

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-Scale image recognition, in: ICLR 2015. pp. 1–13.

Springenberg, J., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: The all convolutional net, in: ICLR 2015.

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

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Transactions on Medical Imaging 35, 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

Tang, Y., Mohamed, A.R., 2012. Multiresolution deep belief networks, in: International Conference on Artificial Intelligence and Statistics. Presented at the International Conference on Artificial Intelligence and Statistics, Canary Islands, Spain.

Telgarsky, M., 2016. Benefits of depth in neural networks, in: 29th Annual Conference on Learning Theory. pp. 1–23.

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

Ural, S., Hussain, E., Kim, K., Fu, C.-S., Shan, J., 2011. Building Extraction and Rubble Mapping for City Port-au-Prince Post-2010 Earthquake with GeoEye-1 Imagery and Lidar Data. Photogrammetric Engineering & Remote Sensing 77, 1011–1023. https://doi.org/10.14358/PERS.77.10.1011

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

Vetrivel, A., Kerle, N., Gerke, M., Nex, F., Vosselman, G., 2016a. Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning. Presented at the GEOBIA 2016, GEOBIA 2016, Enschede, The Netherlands. https://doi.org/10.3990/2.369

Vetrivel, A., Markus Gerke, Norman Kerle, George Vosselman, 2015. Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. ISPRS Journal of Photogrammetry and Remote Sensing 105, 61–78. https://doi.org/10.1016/j.isprsjprs.2015.03.016

Wilson, C., Roelofs, R., Stern, M., Srebro, N., Recht, B., 2017. The marginal value of adaptive gradient methods in machine learning. arXiv:1705.08292.

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

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