

TOWARDS AUTOMATED SATELLITE IMAGE SEGMENTATION AND CLASSIFICATION FOR ASSESSING DISASTER DAMAGE USING DATA-SPECIFIC FEATURES WITH INCREMENTAL LEARNING

A. Vetrivel*, N. Kerle, M. Gerke, F. Nex, G. Vosselman

University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), P.O. Box 217, 7500 AE, Enschede, The Netherlands

{a.vetrivel, n.kerle, m.gerke, f.nex, george.vosselman}@utwente.nl

KEY WORDS: Online learning, CNN features, building damage, satellite imagery, streaming training samples

ABSTRACT:

Automated damage assessment based on satellite imagery is crucial for initiating fast response actions. Several methods based on supervised learning approaches have been reported as effective for automated mapping of damages using remote sensing images. However, adopting these methods for practical use is still challenging, as they typically demand large amounts of training samples to build a supervised classifier, which are usually not readily available. With the advancement in technologies local and detailed damage assessment for individual buildings is being made available, for example through analysis of images captured by unmanned aerial vehicles, monitoring systems installed in buildings, and through crowdsourcing. Often such assessments are being done in parallel, with results becoming available progressively. In this paper, an online classification strategy is adopted where a classifier is built incrementally using the streaming damage labels from various sources as training samples, i.e. without retraining it from the scratch when new samples stream in. The Passive-Aggressive online classifier is used for the classification process. Apart from the classifier, the choice of image features plays a crucial role in the performance of the classification. The features extracted using recently reported deep learning approaches such as Convolutional Neural Networks (CNN), which learns features directly from images, have been reported to be more effective than conventional handcrafted features such as gray level co-occurrence matrix and Gabor wavelets. Thus in this study, the potential of CNN features is explored for online classification of satellite image to detect structural damage, and is compared against handcrafted features. The feature extraction and classification process is carried out at an object level, where the objects are obtained by over-segmentation of the satellite image. The proposed online framework for damage classification achieves a maximum overall accuracy of about 73%, which is comparable to that of batch classifier accuracy (74%) obtained for the same training and testing samples, however at a significantly lesser time and memory requirements. Moreover, the CNN features always significantly outperform handcrafted features.

1. INTRODUCTION & RELATED WORKS

Very high resolution satellite images, which are usually made available within a few hours after a disaster event such as an earthquake, serve as an ideal data source for rapid damage assessment over large areas for fast response actions (Kerle and Hoffman, 2013). Though many automated methods have been proposed for damage mapping using satellite imagery, in practice operational damage mapping continues to be based on manual interpretation of satellite images, which is time- and labour-intensive. There are many reasons for automated methods not yet being adopted for automatic processing, including the limited spatial resolution of satellite imagery compared to increasingly available aerial alternatives, and the complexity of the scene. However, the primary reason is that most of the automated methods are based on supervised learning approaches that require a large volume of training samples to build accurate models which are often not readily available (Dong and Shan, 2013). Also adopting a pre-trained supervised model that was previously developed for a different geographic area or image data type typically has limited transferability. In such cases at least a small number of training samples reflecting the study area is required to fine-tune the pre-trained model. Either for constructing a new model or for calibrating the existing model, a significant number of study-area-specific training samples is required. However, training samples representing the distribution of damage characteristics of the study area, are usually not readily available. Moreover, manual construction of a new training dataset with a

large number of samples is not feasible at the time of emergency. With the advancement in technologies, local and detailed damage assessments for individual buildings are made available from various sources. For example, it is becoming increasingly common to assess damages locally after an earthquake. UAV images substantially exceed satellite images in terms of spatial resolution and multi-perspective coverage of individual buildings (Vetrivel et al., 2015). Therefore, automated damage estimation based on UAV images can be more accurate and reliable. Furthermore, damage information for specific locations is made available on-line from many other sources such as buildings equipped with automated monitoring systems (Sdongos et al., 2014) and crowdsourcing (Ghosh et al., 2011; Sdongos et al., 2014). Local damage estimates from such sources could be used to construct the study-area-specific training data for building a required supervised classifier. However, two major challenges need to be addressed when constructing a supervised classifier with such training data:

- 1) **Handling streaming training samples:** In the considered scenario damage assessment results from the aforementioned sources arrive at different points in time, also depending on whether data processing takes place locally or remotely. Therefore, a classifier is needed that can be trained dynamically, i.e., when new samples arrive the classifier should learn without retraining from scratch, and reclassify the image if required. This kind of learning is referred to as incremental or on-line learning (i.e. learning without having access to all the samples at once) (Crammer

* Corresponding author.

et al., 2006). Conventional batch-learning methods such as Random Forests or Support Vector Machines (SVM) are not suitable for this kind of learning (Narayanan et al., 2016). Many on-line learning algorithms have been developed, and have been shown to perform similarly to batch-learning methods (Hoi et al., 2014; Wang et al., 2014).

- 2) **Data-specific feature for damage classification:** Even when a large number of training samples is available, it is critical to choose the features and a representation strategy that is suitable for the specific data, study area and application. For example, in earlier work (cf. Vetrivel et al. (2016)) we examined various image features for identifying damage using images from various geographic locations. It was observed that specific features are performing well for specific study areas. Hence, selection of appropriate features specific to the study area is crucial for improved assessment. Recent research revealed that supervised feature learning methods such as Convolutional Neural Networks (CNN) could learn the data-specific features and their representation directly from the image pixel values (Bengio et al., 2013). These features are found to be far superior to conventional handcrafted features, which are described in the later section (Bengio et al., 2013).

Generally, the regions corresponding to heavy damage are determined through the identification of damage patterns corresponding to rubble piles, debris and spalling in an image (Kerle and Hoffman, 2013). The recognition process of those damage patterns can be performed by analysing features extracted either at pixel or region level (Dong and Shan, 2013; Kaya et al., 2010; Miura et al., 2013). However, the pixel level analysis is not meaningful for very high spatial resolution images, particularly in the context of damage assessment, as the evidences are identified based on the characteristics of their radiometric distribution pattern, which can be captured more precisely at a region- or object-level. Therefore, super-pixels or segments derived from object-based image analysis approaches are considered as the primary entity for performing feature extraction and classification. Moreover, this kind of segment-based (super-pixels) approach has been demonstrated as a more efficient approach in several studies compared to pixel-based methods, particularly in applications dealing with very high resolution images (Blaschke, 2010; Blaschke et al., 2014).

Another challenging task is how to compile information from various sources to label each super-pixel as damaged or undamaged. However, this is not the focus of this paper. Instead we center on how to build an incremental classifier if the labels for super-pixels are made available from streaming data sources. Hence, in this study the labels for super-pixels are manually annotated and they are synthetically framed as a stream of training labels obtained from various sources to carry out the online-classification.

To the best of our knowledge online classification with CNN features has not yet been tested for remote sensing applications, particularly for damage assessment. Thus the objective is to develop a framework to use online-learning and CNN together to build an incremental classifier with data-specific features for automated satellite image-based damage classification from streaming samples.

2. METHODOLOGY

The proposed methodology comprises three pipeline processes. As a first step, the satellite image is split into super-pixels using

an over-segmentation approach. In the second step, feature extraction is carried for the super-pixels, and as a step 3 an online classifier is constructed based on the extracted features, by considering them as streaming training samples.

Step 1: Over-segmentation of image

Super pixel construction is a mandatory pre-processing step in many image processing applications. For that a range of methods has been reported (Achanta et al., 2012; Aksoy and Akcay, 2005; Salem et al., 2013). Among them Simple Linear Iterative Clustering (SLIC) (cf. Achanta et al. (2012)) is widely used and reported to be effective for obtaining objects in uniform size, which is suitable to create rectangular image patches, the input format required by common CNN implementation for extracting features.

Step 2: Feature extraction

Many kinds of feature extraction techniques, such as statistical, filtering and morphological operations, have been reported for image processing applications (Zhang and Tan, 2002). Among them filtering is recognized as the most effective approach (Arivazhagan et al., 2006; Tian, 2013). For example, many popular features such as Gabor-, Sobel-, Gaussian- and wavelet-features are based on filtering techniques. Such filtering-based features have proved to be effective for many image processing applications, particularly for image classification (Arivazhagan et al., 2006; Tian, 2013). These filters are designed based on standard mathematical functions and they are referred to as hand-crafted features. However, these features are not especially designed for specific data types or applications. It is challenging to choose appropriate features, i.e. designing filters with appropriate weights that give the best image representation for a specific application (Vetrivel et al., 2016). CNN is one of the deep learning approaches where the filters' weights are learned directly from the images chosen for a specific application, instead of using a mathematical function (Krizhevsky et al., 2012). It has been reported that filters that are directly learned from images outperform conventional handcrafted features (Antipov et al., 2015; Chen et al., 2015; Krizhevsky et al., 2012).

In this study, both hand-crafted- and CNN features are examined and compared for the damage classification capacity.

- a) **Hand-crafted features:** Two kinds of hand-crafted texture features that have been widely reported as effective features for damage classification are considered. They are 1) features based on gray-level co-occurrence matrix (GLCM) and 2) Gabor wavelet features. Details about the extraction of GLCM- and Gabor wavelet- features can be found in Preethi and Sornagopal (2014) and Arivazhagan et al. (2006), respectively.
- b) **Deep learning features based on CNN:** CNN can be used for classification in three different scenarios:
 - 1) **Training from scratch:** Designing and training of new CNN requires a large amount of training data to avoid overfitting. This method can be adopted if large number of training sample is available.
 - 2) **Tuning a pre-trained model:** Another common approach is to adopt a pre-trained CNN model that is trained for a related domain (e.g., general image classification) using a large amount of training samples: the network weights of a pre-trained model are fine-tuned using a domain specific

training samples. This approach also requires relatively many training samples.

- 3) **Pre-trained model without tuning:** If only a small amount of training sample is available for the designated application (here damage classification), then a common approach is to extract the features using the pre-trained model and use them to perform classification based on any supervised classifier such as SVM. The activations of one of the fully connected layers in CNN architecture are considered as the CNN features for a given image patch (super-pixels). For more details about the layers in CNN architecture refer to Zeiler and Fergus (2014) and Krizhevsky et al. (2012).

The third approach is most suitable for our application where it is usually common to obtain only small amount of training samples from the aforementioned sources i.e., few hundreds to thousands of samples depending on the nature of the disaster event. Moreover, this approach has been reported to be effective for image classification in various domains, including remote sensing applications such as land cover classification using very high resolution images (Hu et al., 2015). Hence, this approach is adopted in this study where the features for super-pixels are obtained using the pre-trained model and using these features an independent supervised classifier is built for performing the final classification.

Step 3: online classification

Many online-learning algorithms have been proposed and among them the widely used Passive-Aggressive (PA) algorithm is adopted here for building the online classifier (Crammer et al., 2006). The classifier is built incrementally by providing one sample at a time, where it predicts the label and confidence rate of the provided unseen sample. Further, the classifiers gets updated for each wrong and low confidence prediction. The details of the PA algorithm can be found in Crammer et al. (2006).

The overall work flow is depicted in Figure 1.

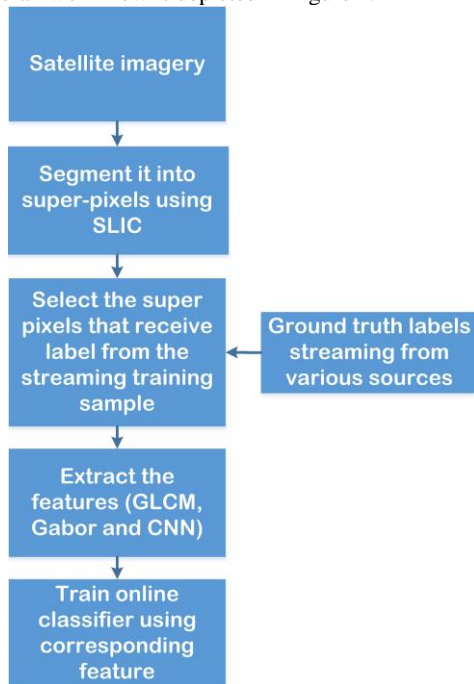


Figure 1. Overall workflow

3. EXPERIMENTS

3.1 Data used:

A subset of a Geosy satellite image with 50 cm nominal ground resolution of Port-au-Prince captured after the 2010 Haiti earthquake was considered for this experiment. The damaged and undamaged regions in the image were manually delineated using a polygon and annotated as damaged and undamaged, respectively, to generate the training samples for building the classifier. These polygons are considered as the streaming ground truth information for analyzing the proposed online classifier.

3.2 Experimental steps and implementation details:

Step 1: The super-pixels were generated for the selected image subset using SLIC method. As stated earlier, we need super-pixels with more uniform size and shape. To achieve this, the parameter 'm' in SLIC that controls the compactness of a super-pixel was empirically determined as 40.

Step 2: The manually delineated polygons annotated with damage labels were overlaid on the super-pixels of the image. The super-pixels having at least 50% of their area overlapping with a polygon were assigned with corresponding polygon's label, and considered for training and testing the classifier. In total, 2553 super-pixels were labelled, of which 70% of the samples were considered for building the classifier, while the remaining 30% samples were used for testing.

Step 3: Three kinds of features (GLCM, Gabor and CNN features) as described in the methodology section were extracted for the selected super-pixels. The CNN features were extracted using a pre-trained CNN model 'imagenet-caffe-alex' developed by Krizhevsky et al. (2012). This model demands the input of image patch with size 227x227x3. Hence the super-pixels were converted into rectangular patches and then scaled to above said dimension. In general the images are scaled to larger size using an interpolation technique that significantly degrades the quality of the image. This might have an impact on the quality of the CNN features. Hence, in this study the image patches based on super-pixels were resized to aforementioned dimensions using two approaches: 1) images resized based on interpolation techniques and 2) images resized based on zero padding. CNN features from image patches obtained based on these two approaches are compared as well.

In total four different features – GLCM, Gabor, CNN_pad (image resized by zero padding) and CNN_nopad (image resized by interpolation technique) are analyzed independently by fitting the PA online classifier.

3.3 Results

The results of the online classification are shown in Figure 2, which depicts the cumulative error rate for each feature. Also this figure implicitly shows the number of times the classifier gets updated. The overall accuracy estimated based on their error rates for GLCM, Gabor, CNN_pad and CNN_nopad are 66.5%, 70.0%, 71.0% and 74.5%, respectively.

It is evident that the online learning algorithm does not need to be re-trained from scratch every time a sample streams in, leading to significantly superior time and memory efficiency. However, this might adversely affect the accuracy of the model. Hence, we intend to determine whether the accuracy diminishes in the online learning setting, and if so, how significant would that be. To this end, we compare the online classifier with canonical batch classifier SVM, using all aforementioned features. The SVM classifier was trained using the same 70% of the samples.

Subsequently, both SVM and PA classifiers were evaluated using the remaining 30% of samples and the accuracies are reported in Table 1.

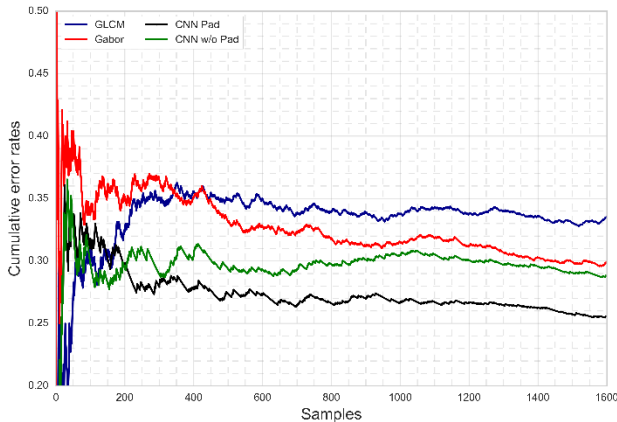


Figure 2. The cumulative error plot of PA online classifiers when associated with GLCM, Gabor and CNN features

Classifier	Overall accuracy in %			
	GLCM	Gabor	CNN_nopad	CNN_pad
PA	57.3	59.5	68.1	73.2
SVM	63.4	63.7	70.9	74.7

Table 1. Overall accuracy of online- and batch- classifiers when associated with different features.

4. DISCUSSIONS, CONCLUSIONS AND FUTURE WORK

A framework was developed for automated mapping of building damage from satellite imagery, using an online classifier that dynamically learns from streaming training samples based on local damage assessments from different sources. In this study, the potential of CNN features from pre-trained model, in combination with online PA classifier was explored for automated damage classification of satellite imagery. Further, the accuracy retrieved from CNN features was compared against the standard hand-crafted features GLCM and Gabor by fitting them using the PA classifier. The results show that CNN features are performing better than both handcrafted features. Moreover, the cumulative error graph (cf. Figure 2) shows that CNN features generalize better with fewer training samples than the handcrafted features. For example, the error curve (cf. Figure 2) of CNN tends to be smoother (shows better generalization) after being trained with 1000 samples, while the error curves of handcrafted features fluctuates, indicating that it could not learn fast and generalize well compared to CNN features. Also the error graph shows that the choice of image resizing approach used to resize the super-pixel based image patch in a dimension required for CNN feature extraction have a significant impact on the performance of the final classification accuracy. For example, CNN features extracted from the image patch resized using an interpolation technique were found to be significantly inferior to the CNN features extracted from image patch resized using zero padding (cf. Figure 2).

The features evaluated in online learning settings were also evaluated using the canonical batch mode classifier SVM. The results show that the selected online classifier performs slightly inferior to the batch classifier, irrespective of the features (cf. Table 1). However, the difference in accuracy is only around 1% concerning CNN features. There is significant difference between the accuracies of online and batch classifiers when

employing the handcrafted features. This is because the online classifiers based on the handcrafted features could not learn and generalize well. The overall conclusion is that the choice of features has significant impact on the performance of the online classifier. Also it is observed that compared to handcrafted features the CNN features are more effective and consistent for online classification, as they yield similar accuracy when used in the batch learning setting (cf. Table 1).

Though CNN features were reported to be effective for many remote sensing application such as land cover classification, the maximum classification accuracy obtained for damage classification using satellite imagery was around 73% (cf. Table 1), signalling a continuing limitation of automated damage classification with satellite imagery (Kerle, 2010). This inferior accuracy might be due to many reasons such as insufficient number of training samples used for building the classifier, and complexity of the scene. However, the major reason would be the limitation of the satellite imagery where the considered image resolution is around 50 cm. With this level of spatial resolution even manual interpretation is difficult, and similar problems have previously been reported (Kerle, 2010). The available pre-event satellite imagery could be of help and is usually available. In such a case, the classifier can be trained with CNN features from both epochs, which could provide better results and we intend to explore this in the future. On the other hand, the training samples considered in this study are from single satellite imagery of a single geographic location, where the characteristics of the streaming samples are not highly variable. However, the usefulness and potential of online learning can be realized when there is high variability in the streaming samples in terms of characteristics of the study area, weather condition and spatial resolution of the images. Hence, as a future work we intend to explore the potential of the proposed framework by applying it to much better spatial resolution satellite or aerial imagery obtained for different geographic locations varying in aforementioned characteristics.

ACKNOWLEDGEMENTS

The work was funded by FP7 project: RECONASS (Reconstruction and Recovery Planning: Rapid and Continuously Updated Construction Damage and Related Needs Assessment; grant no 312718).

REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (11), pp. 2274-2282.
- Aksoy, S., Akcay, H.G., 2005. Multi-resolution segmentation and shape analysis for remote sensing image classification, *Proceedings of 2nd International Conference on Recent Advances in Space Technologies, RAST 2005*, pp. 599-604.
- Antipov, G., Berrani, S.-A., Ruchaud, N., Dugelay, J.-L., 2015. Learned vs. hand-crafted features for pedestrian gender recognition. *Proceedings of the 23rd ACM international conference on Multimedia*, Brisbane, Australia, pp. 1263-1266.
- Arivazhagan, S., Ganesan, L., Priyal, S.P., 2006. Texture classification using Gabor wavelets based rotation invariant features. *Pattern Recognition Letters*, 27 (16), pp. 1976-1982.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8), pp. 1798-1828.

- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (1), pp. 2-16.
- Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., Tiede, D., 2014. Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp. 180-191.
- Chen, M., Zhang, L., Allebach, J.P., 2015. Learning deep features for image emotion classification, *IEEE International Conference on Image Processing (ICIP)* pp. 4491-4495.
- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y., 2006. Online Passive-Aggressive algorithms. *The Journal of Machine Learning Research*, 7, pp. 551-585.
- 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, pp. 85-99.
- Ghosh, S., Huyck, C.K., Greene, M., Gill, S.P., Bevington, J., Svekla, W., DesRoches, R., Eguchi, R.T., 2011. Crowdsourcing for rapid damage assessment: The global earth observation catastrophe assessment network (GEO-CAN). *Earthquake Spectra*, 27 (S1), pp. S179-S198.
- Hoi, S.C., Wang, J., Zhao, P., 2014. Libol: A library for online learning algorithms. *The Journal of Machine Learning Research*, 15 (1), pp. 495-499.
- 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 (11), pp. 14680.
- Kaya, G.T., Ersoy, O.K., Kamasak, M.E., 2010. Spectral and spatial classification of earthquake images by support vector selection and adaptation. *International Conference of Soft Computing and Pattern Recognition (SoCPar)*, pp. 194-197.
- 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 (6), pp. 466-476.
- Kerle, N., Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural hazards and earth system sciences (NHES)*, 13 (1), pp. 97-113.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097-1105.
- Miura, H., Midorikawa, S., Kerle, N., 2013. Detection of building damage areas of the 2006 Central Java, Indonesia, earthquake through digital analysis of optical satellite images. *Earthquake Spectra*, 29 (2), pp. 453-473.
- Narayanan, A., Liu, Y., Chen, L., Liu, J., 2016. Adaptive and scalable android malware detection through online learning. *arXiv preprint arXiv:1606.07150*.
- Preethi, G., Sornagopal, V., 2014. MRI image classification using GLCM texture features, *International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pp. 1-6.
- Salem, M., Ibrahim, A.F., Ali, H.A., 2013. Automatic quick-shift method for color image segmentation, *8th International Conference on Computer Engineering & Systems (ICCES)*, pp. 245-251.
- Sdongos, E., Tsertou, A., Georgakopoulos, N., Loupos, K., Amditis, A., Joram, N., Ellinger, F., Lindner, B., Edwan, E., Ehlerding, A., Camarinopoulos, S., Kallidromitis, V., Bairaktaris, D., Bairaktaris, E., Naundrap, J., Kerle, N., Gerke, M., Vierhuß-Schloms, K., Markus, M., 2014. A novel & practical approach to structural health monitoring -The RECONASS vision: Local positioning, sensor networks, secure communications and remote sensing at the service of structural monitoring to assess construction damage and related needs, *IEEE Workshop on Environmental Energy and Structural Monitoring Systems (EESMS)* pp. 1-6.
- Tian, D.P., 2013. A review on image feature extraction and representation techniques. *International Journal of Multimedia and Ubiquitous Engineering*, 8 (4), pp. 385-396.
- Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 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, pp. 61-78.
- Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2016. Identification of Structurally Damaged Areas in Airborne Oblique Images Using a Visual-Bag-of-Words Approach. *Remote Sensing*, 8 (3), pp. 231.
- Wang, J., Zhao, P., Hoi, S.C.H., 2014. Cost-sensitive online classification. *IEEE Transactions on Knowledge and Data Engineering*, 26 (10), pp. 2425-2438.
- Zeiler, M.D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks. *European Conference on Computer Vision (ECCV)*, pp. 818-833.
- Zhang, J., Tan, T., 2002. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35 (3), pp. 735-747.