

DEEP BAYESIAN ACTIVE LEARNING IN HIGH-RESOLUTION SATELLITE IMAGES FOR CHANGE DETECTION IN URBAN AND SUBURBAN AREAS

¹Lemonia Ragia and ^{2*}Antigoni Panagiotopoulou

^{1,2}Information Management Systems Institute, Athena Research Center, Artemidos 6, Marousi 151 25 Greece

¹leragia@athenarc.gr

^{2*}antigoni27aug@gmail.com

KEY WORDS: Bayesian CNN, Deep Active Learning, BALD, Dropout, QuickBird, WorldView, Change Detection.

ABSTRACT:

In this work the problem of change detection in high-resolution (HR) satellite images is addressed. The active learning (AL) algorithm Bayesian active learning disagreement (BALD) is applied on WorldView images of urban and suburban areas in the island of Crete, Greece. Comparisons with results from random sampling (RS) on AL are carried out. Several cases of selecting different amounts of images in the training set of a convolutional neural network (CNN) are experimented. The results show that the validation accuracy of classification as changed or unchanged of the BALD algorithm is superior to that of the RS algorithm. Indeed, the BALD algorithm achieves zero test error against the test errors 34.6% and 38.5% of the RS algorithm. Actually, as the amount of training images increases, the accuracy also increases. Interesting experiments could be executed in the future utilizing estimators from robust statistics inside the AL acquisition function framework. Up to now in the literature no other work has appeared to present deep AL on WorldView images for change detection.

1. INTRODUCTION

A significant challenge in remote sensing (RMSS) applications is obtaining labelled data as changed or unchanged (Ruzicka et al., 2020). Intercontinentally, maps over huge areas have to be kept up-to-date by being renewed with gradual renovations. In fact, data which have been acquired across aerial or satellite surveys serve for recognizing the potential changes which have to be introduced into the map. Therefore, change detection through automatic image analysis is a problem needed to be addressed in RMSS.

Machine learning (ML) and in specific active learning (AL) can face the above-mentioned challenge of change detection in RMSS. In AL frameworks a system has the potential to learn from small amounts of data and choose unattended what data it would prefer to be labelled by the user. Cost and time can be saved when training a ML system via AL, due to the reduced amount of required labelling.

In the literature, several works utilizing AL across RMSS have been presented. Gaussian processes (GP) and Dirichlet processes serve for creating a probabilistic framework in Sun et al. (2015), Wu, Prasad (2016), with which the model uncertainty is estimated and the data points to be labelled are selected. More specifically, in Sun et al. (2015) three new AL heuristics that rely on the posterior probability output of the GP classifiers are introduced for the selection of the most uncertain candidate data samples from the unlabeled pool. Also, repeated training of the GP classifiers is avoided by means of an incremental model updating scheme. The proposed AL approach can be used in conjunction with other classifiers, which deal with the multiclass

classification problem. In Wu, Prasad (2016) the AL problem is addressed in the context of domain shift and RMSS, where source and target domains, along with an optimal transport method, are utilized for data labelling and efficient AL. An AL framework, which is robust in classification with least manual labelling attempt, is proposed. Simultaneously, undisclosed classes are brought to light. Local information density serves for the query strategy, where clustering based on Dirichlet process mixture model gives the local density. The proposed methodology is advantageous for numerous applications where it is often impossible that the initial training library includes all classes or accounts for multimodal distributions caused by class variability. Also, in Haut et al. (2018) a new AL-driven framework for classification is presented, where both the spectral and the spatial contextual information of the hyperspectral data is utilized. Bayesian convolutional neural networks (CNN) are incorporated in the proposed framework. Robust hyperspectral image classification with very short training sets is achieved, due to avoiding the dimensionality curse and the overfitting problem. RMSS datasets serve for evaluating various deep AL techniques. Monte Carlo (CM) dropout is demonstrated as capable of effective model uncertainty estimation.

Furthermore, AL is proposed in Hamrouni et al. (2021) to be utilized for adapting a classifier trained on a source image to spatially distinct target images with as small as possible labelling load. Poplar plantations are classified among other tree types in an operational framework. A local model is adjusted into a global model being appropriate for a national scale mapping. Fitting samples from the unexplored areas are queried as well as novel classes get uncovered. Experiments are carried out on Sentinel-2 time series and poplar plantations are identified at a local scale with an average F-score ranging from 89.5% to 99.3%. In Feng et al. (2019) an AL method for training a LiDAR 3D object detector with as few as possible labeled data is presented. In order to decrease the search space of objects and accelerate the

* Author to whom correspondence should be addressed

learning process, the detector leverages 2D region proposals having been created from the RGB images. According to the experimental results, the proposed method works under various uncertainty estimations and query functions while up to 60% of the labelling efforts can be saved without sacrificing network performance.

In the present work the AL algorithm called Bayesian active learning disagreement (BALD) (Gal et al., 2017) is applied on WorldView images of urban and suburban areas, demonstrating the application of change detection. The results that are obtained are compared against those from AL using random sampling (RS). The study area is Georgioupoli, in the island of Crete, Greece. A CNN model is trained on the WorldView images by various cases of selecting different amounts of images in the training set. Experimentation demonstrates that the testing accuracy of classification as changed or unchanged of the BALD algorithm is superior to that of the RS algorithm. Actually, the validation accuracy increases as the number of training images also increases. The novelty of the present work lies on applying deep AL on HR satellite images, i.e. the WorldView images, to perform change detection.

This work is organized into five sections. In Section 2 the BALD acquisition function into CNNs is presented. AL for change detection utilizing the WorldView images is given in detail in Section 3. Discussion is carried out in Section 4 while the conclusions are drawn in Section 5.

2. ACQUISITION FUNCTION: BAYESIAN ACTIVE LEARNING DISAGREEMENT

In AL a model gets trained on an initial dataset of small amount and an acquisition function determines which data points to ask to be labelled by an external oracle. There is a pool of unlabeled data points, lying outside of the training set, from which the acquisition function selects one or more points. Then, the selected data points are labelled by an oracle, they are added to the training set and a novel model gets trained on the re-equipped training dataset. The above-mentioned process is repeated while the data is progressively enlarged (Gal et al. 2017).

For AL to be performed, a certain model should exhibit learnability under small amounts of data and also act for its uncertainty over unknown data. Practically, Bayesian deep learning is combined with AL to cope with high dimensional data. In this work AL is demonstrated on such image data and the employed model presents the potential to represent prediction uncertainty on the data. The Bayesian equivalent of CNNs (Gal, Ghahramani, 2016) is applied on image data. These Bayesian CNNs present prior probability distributions that are set over a group of model parameters:

$$w = \{W_1, \dots, W_L\}; w \sim p(w) \quad (1)$$

where $p(w)$ may be a standard Gaussian prior. For classification, a likelihood model:

$$p(y = c|x, w) = \text{softmax}(f^w(x)) \quad (2)$$

is also defined, where $f^w(x)$ is the model output.

Practical approximate inference in multiplex deep models, such as the Bayesian CNNs, can be performed by utilizing stochastic regularization techniques like the dropout technique (Panagiotopoulou, Anastassopoulos, 2009), (Hinton et al., 2012), (Srivastava et al., 2014), (Gal, Ghahramani, 2016).

Specifically, in Bayesian CNNs inference is reached through training a model with dropout every weight layer and by executing dropout during testing in addition to sample from the approximate posterior. These are stochastic forward passes and have been named MC dropout. Bayesian nets prove efficient with small amounts of data. In addition, these nets hold uncertainty information which can be exploited in conjunction with existing acquisition functions. In fact, acquisition functions for the classification task are of interest in the present work.

Different AL algorithms or strategies make use of different metric or acquisition function for the selection of data to be labeled. The goal is to minimize the number of data annotation that is queried from an oracle during training. Therefore, an acquisition function $A_{cq}(x, M)$ is a function of x , through which the AL algorithm decides where to query next:

$$x^* = \underset{x \in D_p}{\text{argmax}} A_{cq}(x, M) \quad (3)$$

where M denotes the model and D_p stands for the pool of data. Queries from informative areas and not queries from noise should be made. So, the selected data for labeling should decrease model uncertainty.

The BALD acquisition function relies on Bayesian uncertainty (Houlsby et al., 2011) and can be utilized in view of the classification task. According to BALD principle, the pool data points which are anticipated to maximize the information gained about the model parameters, that is the common information between predictions and model posterior, should be chosen. The model presents on average uncertainty regarding the data points that maximize the BALD acquisition function. However, some model parameters give disagreeing predictions of high certainty, which brings points of high variance in the softmax layer input. Thereafter, the highest probability assigned to a different class would be presented by each stochastic forward pass across the model.

3. ACTIVE LEARNING FOR CHANGE DETECTION IN URBAN AND SUBURBAN AREAS

3.1 Image data

The study area is Georgioupoli in the island of Crete, Greece (Ragia, Krassakis, 2019). The experimental data consist of two different scenes near the shore. WorldView images of spatial resolution 30cm/pixel and QuickBird images of 60cm/pixel are utilized. In fact, there is a time difference of thirteen years among the satellite images.

Figures 1 and 2 depict the QuickBird and WorldView images, respectively, of "Scene 1". Concerning "Scene 2", the satellite images are shown in Figures 3 and 4. The four spectral bands of the images are given in Table 1. The bands 2, 3, 4 and 5 of the WorldView image correspond to the blue, green, yellow and red channels, respectively. Regarding the QuickBird image, the bands 1, 2, 3 and 4 represent the blue, green, red and near-

infrared channels, correspondingly.

All images that are shown in Figures 1-4 have the same spatial resolution (Bratsolis et al., 2018) equal to 30cm/pixel. Actually, QuickBird images have been bicubically interpolated per the factor of 2. Additionally, the QuickBird images have been spatially or geometrically co-registered to the WorldView images. Actually, the WorldView images demonstrate a more developed area than the QuickBird images, due to human intervention during the time of thirteen years.

WorldView	QuickBird
2, 3, 4, 5	1, 2, 3, 4
Blue, green, yellow, red	Blue, green, red, near-infrared

Table 1. Satellite image bands.



Figure 1. QuickBird image of “Scene 1” at spatial resolution 30cm/pixel.



Figure 2. WorldView image of “Scene 1” at spatial resolution 30cm/pixel, after thirteen years have passed from the Quickbird image of Figure 1.

The two scenes of interest, “Scene 1” and “Scene 2”, get

tiled into a total of 210 tiles and 190 tiles, respectively, of size 28x28 pixels. Then, the QuickBird images serve for labeling the WorldView images into “changed” and “unchanged”. Before the labeling into two classes is performed, the four channel satellite images are converted to grayscale images. Specifically, per-tile labels are calculated, where tiles with >10% changed pixels are considered as changed while tiles with <=10% changed pixels are regarded unchanged. Pixel intensities which differ in value more than 0.02 are considered as changed. An example is shown in Figure 5. The labeling threshold is data dependent and can be decided through trial and error. The datasets are highly unbalanced. There are only 15 and 33 tiles with changes in cases of “Scene 1” and “Scene 2”, respectively.



Figure 3. QuickBird image of “Scene 2” at spatial resolution 30cm/pixel.



Figure 4. WorldView image of “Scene 2” at spatial resolution 30cm/pixel, after thirteen years have passed from the QuickBird image of Figure 3.

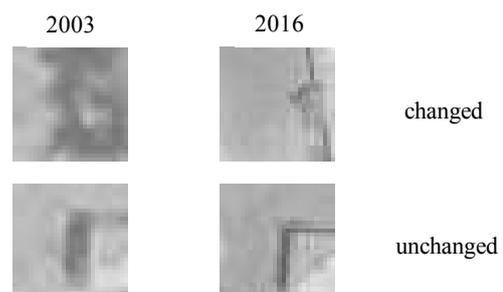


Figure 5. Example image tiles for change detection. The labelling threshold is equal to 10% of pixels differing in value more than 0.02.

3.2 Experimental Procedure and Results

In this work the BALD acquisition function with the Bayesian CNN trained on the Georgioupoli dataset is studied. In specific, the BALD function is assessed with the following model structure: convolution-relu-convolution-relu-max pooling-dropout-dense-relu-dropout-dense-softmax with 32 convolution kernels, 4x4 kernel size, 2x2 pooling, dense layer with 128 units and dropout probabilities 0.25 and 0.5. All experiments are carried out with learning rate equal to 0.005, with momentum of gradient descent of value 0.1 and during training of 10 epochs.

The model is trained on the WorldView images, separately for the two scenes, by several cases of selecting different amounts of images in the training set. In particular, for “Scene 1”, the model is trained with 42, 84, 126, 168 and 210 images, whereas testing is always performed with 38 undisclosed to the model images, Table 2.

As far as “Scene 2” is concerned, model training is carried out with the following numbers of images 76, 114, 152 and 190, Table 3. Regarding the number of testing images, it always equals 42 and the images are unknown to the model. The above mentioned image groups are formed by random split of the initial whole set of images.

In the current study using a small model the system achieves zero test error on Georgioupoli WorldView data with 84, 126, 168 and 210 labelled images of “Scene 1”, while it presents 54.6% test error when only 42 labelled images are utilized. The above results regard the BALD acquisition function. However, in the case of the RS algorithm, the larger test error equal to 38.5% is reached when 84 and 126 labelled images are used. Also, RS reaches 34.6% test error with 168 and 210 sample training images as well as 63.6% error in case of 42 labelled images, Table 2.

As far as “Scene 2” is concerned, the model with BALD acquisition function becomes able of 100% test accuracy when 114, 152 and 190 sample training images are utilized. Nevertheless, zero test accuracy is observed for 76 training images. Regarding the model test performance with RS algorithm, there is 9.1% accuracy with 76 labelled images, while 61.5% accuracy is reached with the greater numbers of 114 and 152 training images. When 190 training samples are utilized, the RS acquisition function achieves 65.4% test accuracy, Table 3.

The validation accuracies or test errors are calculated by averaging after 25 rounds of repetition in all cases apart from the cases of training images number equal to 76 or 42 where 10 rounds of repetition are followed. These repetition round numbers are chosen after experimentation to get favorable model performance. Figures 6 and 7 demonstrate graphical plot of test accuracy as a function of the number of training images.

Number of Training Images	Test Error %	
	BALD	RS
42	54.6	63.6
84	0	38.5
126	0	38.5
168	0	34.6
210	0	34.6

Table 2. Georgioupoli “Scene 1”: Test errors as resulting from the different acquisition functions a) Bayesian active learning disagreement and b) Random sampling, for the various amounts of training images.

Number of Training Images	Test Error %	
	BALD	RS
76	100	90.9
114	0	38.5
152	0	38.5
190	0	34.6

Table 3. Georgioupoli “Scene 2”: Test errors as resulting from the different acquisition functions a) Bayesian active learning disagreement and b) Random sampling, for the various amounts of training images.

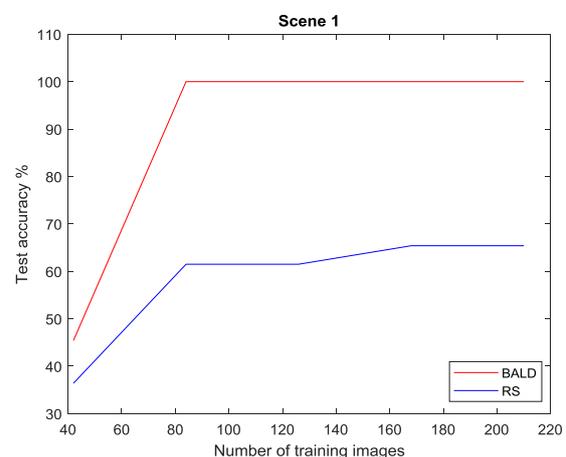


Figure 6. Test accuracy as a function of the number of training images. Results from the acquisition functions Bayesian active learning disagreement and random sampling are demonstrated for “Scene 1”.

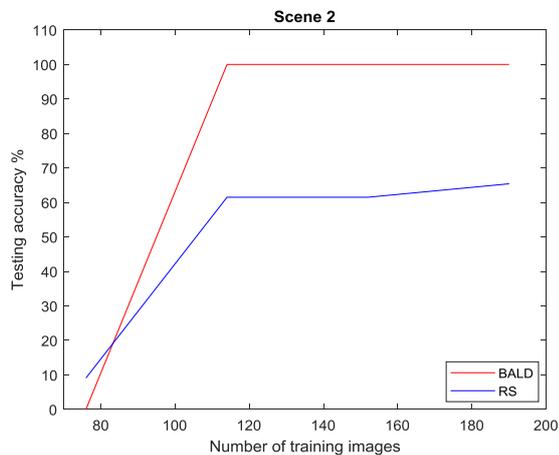


Figure 7. Test accuracy as a function of the number of training images. Results from the acquisition functions Bayesian active learning disagreement and random sampling are demonstrated for “Scene 2”.

4. DISCUSSION

During AL the object of minimization is epistemic uncertainty, that relates to the CNN parameters. In Bayesian models dropout estimation for regularization is employed. In fact, uncertainty is propagated across the model which results in great influence on the model measuring its confidence. The acquisition functions which rely on Bayesian uncertainty avoid the selection of noisy points that are close-by images for which several noisy labels of different classes are available. The particular data points present large aleatoric uncertainty rather than epistemic uncertainty. The former uncertainty cannot be reduced (Gal et al., 2017).

Novel acquisition functions that would be prosperous in AL could be formed by employing robust estimators (Huber, 1981), (Tukey, 1983). In fact, interesting experiments of prediction for classification could be carried out with the Var-norm estimator (Panagiotopoulou, 2013) inside an acquisition function formulation in a Bayesian CNN framework. The mutual pieces of information between model predictions and model posterior could be maximized through the minimization of a Var-norm cost function that takes as argument the differences between prediction and prior expected results. Iterative forward passes through the model, where error back-propagation will have been incorporated, would lead to the selection of data points which are preferred for labelling and therefore, decrease model uncertainty.

5. CONCLUSIONS

Change detection through automatic image analysis is a problem needed to be addressed in remote sensing. Time and resources can be saved when training a system via active learning due to the reduced amount of required labelling. In the present work the BALD algorithm is applied on WorldView images of urban and suburban areas, for the application of change detection. The results are compared with those from random sampling. The study area is Georgioupoli, in the island of Crete, Greece. A CNN model is trained on the WorldView images by various cases of selecting different amounts of images in the training set. Experimentation demonstrates that the testing

accuracy of classification as changed or unchanged of the BALD algorithm is superior to that of the random sampling algorithm. Actually, the validation accuracy increases as the number of training images also increases. The testing accuracies are calculated by averaging after 25 rounds of repetition in most cases apart from certain cases of few training images where 10 rounds of repetition are followed. Up to now in the literature no other work has been presented with deep active learning on WorldView images for change detection.

In future work additional active learning experiments will be executed to assess how many images in random sampling algorithm an expert is necessary to label so that to reach the same test accuracy with the BALD algorithm. Furthermore, interesting experiments of prediction for classification could be carried out utilizing estimators from robust statistics inside the acquisition function formulation in a Bayesian CNN model.

REFERENCES

- Bratsolis, E., Panagiotopoulou, A., Stefouli, M., Charou, E., Madamopoulos, N., Perantonis, S. 2018: Comparison of optimized mathematical methods in the improvement of raster data and map display resolution of Sentinel-2 images. *IEEE ICIP Proc.*, 2521-2525.
- Feng, D., Wei, X., Rosenbaum, L., Maki, A., Dietmayer, K., 2019: Deep active learning for efficient training of a LiDAR 3D object Detector. *2019 IEEE IV*, 667-674.
- Gal, Y., Ghahramani, Z., 2016: Bayesian convolutional neural networks with Bernoulli approximate variational inference. *ICLR workshop track*.
- Gal, Y., Ghahramani, Z., 2016: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *ICML 2016*, 48, 1050-1059.
- Gal, Y., Islam R., Ghahramani, Z., 2017: Deep Bayesian active learning with image data. *ICML 2017*, 70, 1183-1192.
- Hamrouni, Y., Paillassa, E., Cheret, V., Monteil, C., Sheeren, D., 2021: From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171(1), 76-100.
- Haut, J.M., Paoletti, M.E., Plaza, J., Li, J., Plaza, A., 2018: Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11), 6440-6461.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.
- Houlsby, N., Huszar, F., Ghahramani, Z., Lengyel, M., 2011: Bayesian active learning for classification and preference learning. arXiv preprint arXiv:1112.5745.
- Huber, P.J., 1981. *Robust Statistics*. John Wiley & Sons, New

York, USA.

Panagiotopoulou, A., Anastassopoulos, V., 2009: Regularized super-resolution image reconstruction employing robust error norms. *Optical Engineering* 48(11), 117004-1 – 117004-14.

Panagiotopoulou, A., 2013. Iterative multi-frame super-resolution image reconstruction via variance-based fidelity to the data. *International Journal of Remote Sensing Applications* 3(4), 203-214.

Panagiotopoulou, A., Ragia, L., 2021: Automatic change detection for high-resolution satellite images of urban and suburban areas. *International Journal of Geological and Environmental Engineering* 15(1), 32-37.

Ragia, L., Krassakis, P., 2019: Monitoring the changes of the coastal areas using remote sensing data and geographic information systems. *SPIE Proc. of 7th Int. Conf. RSCy2019*, 111740X.

Ruzicka, V., D'Aronco, S., Wegner, J.D., Schindler, K., 2020: Deep active learning in remote sensing for data efficient change detection. *2020 MACLEAN CEUR*, 2766.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014: Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 15(1), 1929-1958.

Sun, S., Zhong, P., Xiao, H., Wang, R., 2015: Active learning with gaussian process classifier for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 53(4), 1746-1760.

Tukey, J., 1983. *Understanding Robust and Exploratory Data Analysis*. John Wiley and Sons, New York, USA.

Wu, H., Prasad, S., 2016: Dirichlet process based active learning and discovery of unknown classes for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8), 4882-4895.