Exploiting ConvNet Diversity for Flooding Identification

Keiller Nogueira, Samuel G. Fadel, Ícaro C. Dourado, Rafael de O. Werneck, Javier A. V. Muñoz, Otávio A. B. Penatti, Rodrigo T. Calumby, Lin Tzy Li, Jefersson A. dos Santos, and Ricardo da S. Torres

***Abstract*— Flooding is the world's most costly type of natural disaster in terms of both economic losses and human causalities. A first and essential procedure toward flood monitoring is based on identifying the area most vulnerable to flooding, which gives authorities relevant regions to focus. In this letter, we propose several methods to perform flooding identification in high-resolution remote sensing images using deep learning. Specifically, some proposed techniques are based upon unique networks, such as dilated and deconvolutional ones, whereas others were conceived to exploit diversity of distinct networks in order to extract the maximum performance of each classifier. The evaluation of the proposed methods was conducted in a high-resolution remote sensing data set. Results show that the proposed algorithms outperformed the state-of-the-art baselines, providing improvements ranging from 1% to 4% in terms of the Jaccard Index.**

***Index Terms*— Flooding identification, inundation, MediaEval, natural disaster, remote sensing, satellites.**

I. Introduction

NATURAL disaster monitoring is a fundamental task to create prevention strategies as well as to help authorities act in the control of damages, coordinate rescues, and assist victims. Among natural hazards, flooding is possibly the most extensive and devastating one, destroying buildings and threatening human lives [1]. All these consequences make such events to be considered as the world's most costly type of natural disaster in terms of both economic losses and human causalities [2].

Manuscript received November 3, 2017; revised April 3, 2018 and May 21, 2018; accepted June 5, 2018. Date of publication June 27, 2018; date of current version August 27, 2018. This work was supported in part by São Paulo Research Foundation under Grant 2013/50169-1, Grant 2013/50155-0, Grant 2014/50715-9, Grant 2014/12236-1, Grant 2015/24494-8, and Grant 2016/18429-1, in part by Minas Gerais State Agency for Research and Development under Grant APQ-00449-17, in part by National Council for Scientific and Technological Development under Grant 312167/2015-6, and in part by Coordination for the Improvement of Higher Level Personnel under Grant 88881.145912/2017-01. *(Corresponding author: Jefersson A. dos Santos.)*

K. Nogueira and J. A. dos Santos are with the Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil (e-mail: keiller.nogueira@dcc.ufmg.br; jefersson@dcc.ufmg.br).

S. G. Fadel, Í. C. Dourado, R. de O. Werneck, J. A. V. Muñoz, and R. da S. Torres are with the Institute of Computing, University of Campinas, Campinas 13083-970, Brazil.

O. A. B. Penatti is with the Samsung Research and Development Institute Brazil, Campinas 13097-160, Brazil.

L. T. Li is with the Institute of Computing, University of Campinas, Campinas 13083-970, Brazil, and also with the Samsung Research and Development Institute Brazil, Campinas 13097-160, Brazil.

R. T. Calumby is with the Department of Exact Sciences, State University of Feira de Santana, Feira de Santana 44036-900, Brazil.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2018.2845549

Although extremely important, floods are difficult to monitor, because they are highly dependent on several local conditions, such as precipitation, slope of terrain, drainage network, protective structures, and land cover [3]. A first and essential step toward such monitoring is based on identifying areas most vulnerable to flooding, helping authorities to focus on such regions while monitoring inundations.

Remotely sensed data may play a crucial role in the identification of such areas since it allows the capture of whole inundated regions, allowing a better understanding of what and how they are being flooded. Because of the importance of such task, a subtask (called flood-detection in satellite images) of the 2017 multimedia satellite task [4], which was part of the traditional MediaEval benchmark, was proposed to leverage the development of methods for identifying flooding areas in high-resolution remote sensing images. In this letter, we present our proposed methods, which won the aforementioned task, to automatically identify flooding areas in high-resolution remote sensing images using deep learning.

Although there are a number of previous works [5], [6] that performed flooding detection using remote sensing data in conjunction with elevation maps, augmenting the amount and variety of information available for an effective prediction, this letter is one of the first specifically focused on identifying flooding areas on only high-resolution imagery using deep learning-based approaches. Furthermore, some works [7], [8] performing surface water segmentation may be suitable for flooding detection, given the high similarity between these tasks. However, the focus of this letter is exclusively on *flooding area identification* from satellite images took during and shortly after a flood event.

As introduced, the proposed methods to perform flooding identification employ a resurgent area called deep learning [9]. Methods from this area, commonly represented as multilayered neural networks, are able to learn both the features and the classifier in a unified manner, adjusting themselves to better represent the characteristics of the data and their labels. A specific deep learning method, called convolutional (neural) networks (ConvNets) [9], is the most popular for learning visual features in computer vision applications, including remote sensing [10], [11]. This type of network relies on the natural stationary property of an image, i.e., the statistics of one part of the image are assumed to be the same as those of any other part. Furthermore, ConvNets can be considered as an inherently multiscale approach since they usually obtain different levels of abstraction for the data, ranging from local low-level information in the initial layers (e.g., corners and

1545-598X © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: Universidade Estadual de Campinas. Downloaded on October 10,2025 at 14:54:28 UTC from IEEE Xplore. Restrictions apply.

edges), to more semantic descriptors, mid-level information (e.g., object parts) in intermediate layers and high-level information (e.g., whole objects) in the final layers.

Supported by these advantages, we introduce several approaches to identify flooding areas of remote sensing images exploiting ConvNets. First, four networks have been proposed based on distinct properties, including: 1) dilated convolutions [12], which, unlike standard ConvNets, process the input without downsampling it and 2) deconvolution layers, employed similar to the SegNet [13], in which a coarse feature map is upsampled outputting a dense map with the same resolution of the original image. Then, each network is analyzed individually and in combination (with an ablation study), allowing a better understanding of their diversity, which resulted in the proposal of a combination method that aims to exploit the complementarity of these networks. In summary, the main contributions of this letter are the introduction of distinct network architectures, the analysis of their diversity, and the proposal of combination methods that try to exploit the complementary views of different networks. Results of the proposed methods represent the state of the art, in terms of the Jaccard Index, in a remote-sensing-based flooding detection task. These results made us the winner of the Flood-Detection in Satellite Images, a subtask of 2017 Multimedia Satellite Task [4].

II. Proposed Methods

This section introduces the network architectures (Section II-A) as well as the combination method (Section II-B) proposed to perform flooding identification.

A. Network Architectures

All ConvNets conceived specifically for flooding identification are illustrated in Fig. 1. Some use dilated convolutions, while others are based on deconvolutional networks. Independently of the architecture, rectified linear unit [14] was used as processing unit for all neurons.

Specifically, two architectures, presented in Fig. 1(a) and (b), are based on the concept of dilated convolutions [12]. In these layers, the convolution filter is expanded by dilation rate. Given this rate, the weights are placed far away at the given intervals and the kernel size increases by allowing gaps (or "holes") inside their filters. Therefore, networks composed of these layers allow the receptive field to expand but preserving the resolution, i.e., *without downsampling the input data*. This procedure represents a great advantage in terms of processing as well as in terms of learning since internal feature maps do not lose resolution (and maybe useful information).

The two remaining networks, presented in Fig. 1(c) and (d), are based on deconvolutional networks [13]. This type of network has two modules: the first receives input images, learns the visual features by using standard convolution and pooling layers, and outputs a coarse feature map; whereas the second receives this map as input, learns to upsample these features by using several deconvolution layers, and outputs a dense prediction map with the same resolution of the original image. Both modules work together without distinction and can be trained end to end by using the standard feedforward and backpropagation algorithms.

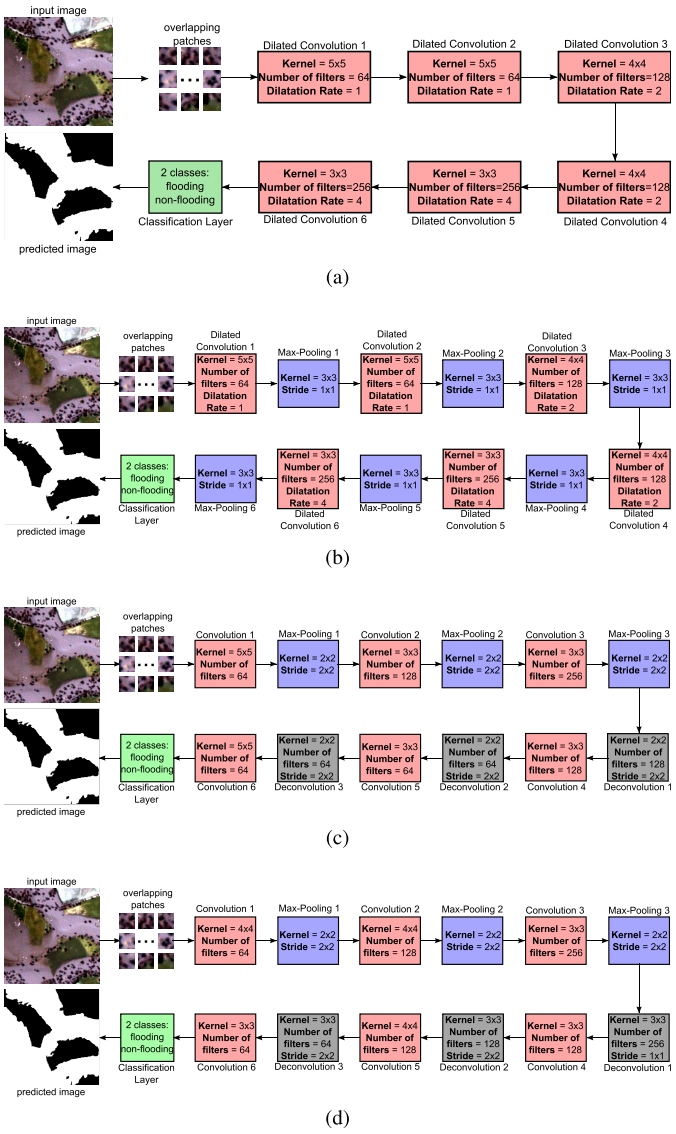

Fig. 1. ConvNet architectures proposed in this letter. (a) Dilated ConvNet #1. (b) Dilated ConvNet #2. (c) Deconvolution ConvNet #1. (d) Deconvolution ConvNet #2.

B. Combination

We also proposed another strategy to solve the flooding detection task, which aims to exploit the diversity of distinct ConvNets. The main premise is that the previous presented ConvNets learn and produce distinct outcomes (dense prediction maps). Those differences should make ConvNets complementary to each other. Therefore, a clever combination of such outcomes should improve the final prediction map if compared with the ConvNets individual results. We propose a combination method using support vector machines (SVMs).

The proposed method is divided into the following three main steps.

1) *Extraction:* In this phase, an image is processed by all proposed networks, which produce distinct outcomes (i.e., different probability or prediction maps). All these maps (that have the same resolution of the original input image) are then concatenated creating a feature vector that, in fact, represents the input image.

Authorized licensed use limited to: Universidade Estadual de Campinas. Downloaded on October 10,2025 at 14:54:28 UTC from IEEE Xplore. Restrictions apply.

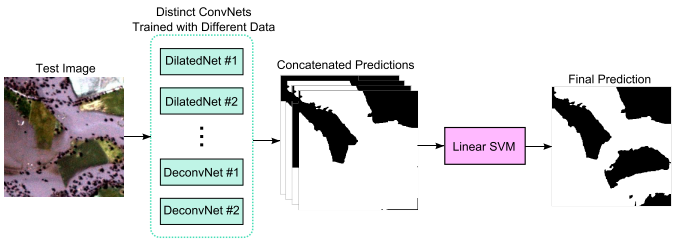

Fig. 2. Pipeline for the prediction phase of the combination approach.

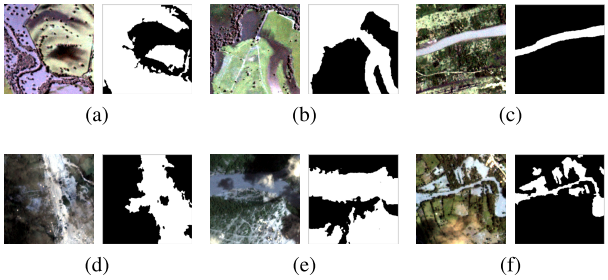

Fig. 3. Examples of patches for all locations of the training set with its respective ground truth, in which white regions refer to flooding areas and black ones correspond to background. (a) Location 1. (b) Location 2. (c) Location 3. (d) Location 4. (e) Location 5. (f) Location 6.

2) *Learning:* In this step, the SVM receives the aforementioned feature vector as well as the ground-truth flooding map for all training data. Then, it independently processes each pixel of these images, learning which and when each classifier is better.
3) *Prediction:* This final step receives feature vectors of testing images and, using the trained SVM, outputs the improved prediction map for each test image. This final step is illustrated in Fig. 2.

III. Experiments

In this section, we present the experimental setup, including the data set, evaluation metric, protocol, and baselines.

A. Data Set

The data set consists of satellite image patches collected from eight different flooding events June 1, 2016 to May 1, 2017. Each image patch is composed of four bands (red, green, blue, and near infrared bands) and has resolution of 320×320 pixels with a ground-sample distance of 3.7 m and an orthorectified pixel size of 3 m [4]. Each pixel is classified into two classes: background and flooded area. While the former class includes everything that is not related to water, the latter, mainly supported by the similarity, includes water in general and does not distinguish between flood and normal surface water (such as river).

The training set is composed of 462 image patches unevenly extracted from *six* locations. Among these images, 92 (20%) were employed as internal validation set to evaluate the proposed algorithms, whereas the remaining 370 images were used to train the proposed methods. Some examples of image patches for each of these locations are presented in Fig. 3. Two test sets were released in this data set: the *Same Locations* test set contains 216 unseen patches unevenly extracted from the same region presented in the training set, whereas the *New Locations* test set contains 58 unseen patches extracted from a region not present in the training set. Fig. 4 presents some examples of these test sets. Until the submission of this letter, ground-truth information of the test sets was not released by the organization of the competition.

B. Experimental Evaluation

In order to assess the performance of generated segmentation masks for flooded areas in the satellite image patches, the intersection-over-union (IoU) metric (also known as the Jaccard Index), was used. The metric measures the accuracy for the pixelwise classification and is defined as $\text{IoU} = (\text{TP}/(\text{TP} + \text{FP} + \text{FN}))$, where TP, FP, and FN are the numbers of true positive, false positive, and false negative pixels, respectively, determined over the whole test set.

C. Experimental Protocol

First, we trained all presented ConvNets using 50% overlapping patches of size 25×25 extracted from all training images. In the prediction phase, we extracted overlapping patches with the same resolution from the testing images and averaged the probabilities outputted by the network. Among all networks, the best one is reported as *ConvNet* 25×25.

Another proposed method relied on training the aforementioned ConvNets using larger overlapping patches, with 50×50 pixels, also extracted from all training images. The motivation behind this strategy is based on the entire context that could be extracted from the input patches to improve the learning process. The prediction phase is similar to the previous strategy. Considering this configuration, the best network is referred to, in Section IV, as *ConvNet* 50×50.

The *Location ConvNets* strategy is based on the idea of creating specialized ConvNets for each flooding event. Since the data set has six distinct flooding event locations, we propose to train a specific Dilated ConvNet #1 (using patches of 25×25) for each location. The prediction is similar to the other proposed protocols, except for the fact that, in this case, each ConvNet was used in its respective location. For the *New Locations* test set, we combined the outcomes extracted from each ConvNet (trained specifically for each location) using a linear SVM, as proposed in Section II-B.

The *Fusion-SVM* strategy expands above-mentioned idea, i.e., an SVM is used to create prediction maps for *both test sets* and not only for the *New Locations* one. Based on the premise that distinct ConvNets (trained using different input data) produce distinct (and possibly complementary) outcomes, we propose to combine the predictions extracted from above ConvNets using a linear SVM, as presented in Section II-B. In this way, the SVM should learn when and how these networks complement each other in order to improve the final performance.

Another strategy relies on exploiting the diversity of distinct ConvNets by combining outcomes of previous methods using a majority voting scheme, referred as *Fusion-MV*.

All proposed methods were created using TensorFlow, and the code has been made publicly available at https://github.com/keillernogueira/FDSI/. During training, all the aforementioned protocols used the same hyperparameters, which were defined based on convergence analyses.

Authorized licensed use limited to: Universidade Estadual de Campinas. Downloaded on October 10,2025 at 14:54:28 UTC from IEEE Xplore. Restrictions apply.

TABLE I

ABLATION STUDY OVER THE PROPOSED MODELS

Type	Network(s)	Events Considered	Validation	
			25×25 patches	50×50 patches
Single	Dilated ConvNet #1	1,2,3,4,5,6	85.60	84.98
	Dilated ConvNet #1	1	62.86	
	Dilated ConvNet #1	2	66.52	
	Dilated ConvNet #1	3	51.56	
	Dilated ConvNet #1	4	35.75	
	Dilated ConvNet #1	5	37.16	
	Dilated ConvNet #1	6	35.46	
	Dilated ConvNet #2	1,2,3,4,5,6	84.21	83.21
	Deconvolution ConvNet #1	1,2,3,4,5,6	83.96	83.04
	Deconvolution ConvNet #2	1,2,3,4,5,6	83.62	82.03
Combination	Dilated ConvNet #1 + #2	1,2,3,4,5,6	83.96	84.69
	Deconvolution ConvNet #1 + #2	1,2,3,4,5,6	84.66	84.63
	Dilated ConvNet #1 + Deconvolution ConvNet #1	1,2,3,4,5,6	86.87	87.32
	Dilated ConvNet #1 + Deconvolution ConvNet #2	1,2,3,4,5,6	86.88	86.54
	Dilated ConvNet #2 + Deconvolution ConvNet #1	1,2,3,4,5,6	87.03	87.54
	Dilated ConvNet #2 + Deconvolution ConvNet #2	1,2,3,4,5,6	88.93	89.10
	Dilated ConvNet #1 + #2 + Deconvolution ConvNet #1	1,2,3,4,5,6	88.99	88.91
	Dilated ConvNet #1 + #2 + Deconvolution ConvNet #2	1,2,3,4,5,6	88.01	89.52
	Dilated ConvNet #1 + Deconvolution ConvNet #1 + #2	1,2,3,4,5,6	87.57	88.95
	Dilated ConvNet #2 + Deconvolution ConvNet #1 + #2	1,2,3,4,5,6	87.01	88.65
	Dilated ConvNet #1 + #2 + Deconvolution ConvNet #1 + #2	1,2,3,4,5,6	88.91	88.62
	ALL (submission)	1,2,3,4,5,6	88.71	

Specifically, learning rate, weight decay, momentum, and number of epochs are 0.01, 0.0005, 0.9, and 20, respectively. After every five epochs, the learning rate is reduced 10 times to allow a finer adjustment during the training process.

D. Baselines

The baselines evaluated in this letter were, in fact, the approaches proposed for the Flood-Detection in Satellite Images subtask of the 2017 Multimedia Satellite Task. An overview of the such methods (which includes the state-of-the-art methods, such as Generative Adversarial, Deconvolutional, and Fully ConvNets) is presented in Table II.

IV. RESULTS AND DISCUSSION

In this section, we present the ablation study performed to analyze the diversity of the ConvNets. The final results reported by the competition organizers are also presented.

A. Ablation Study

An ablation study, presented in Table I, was performed using our validation set in order to evaluate the ConvNets and their diversity. In this letter, features from all networks (trained using patches of size 25×25 and 50×50) were extracted and used (individually or in combination) as input for a linear SVM, replicating the proposed *Fusion-SVM* approach.

Considering the ConvNet individually, all networks (trained with all events) achieved very similar results. Among them, Dilated ConvNet #1 yielded slightly better results independent of the patch size. Therefore, these networks were the ones submitted to the competition and reported in the remaining of this letter as ConvNet 25×25 and ConvNet 50×50.

Considering the combination, it is possible to see that networks based on same paradigm (first two combinations) do not produce higher performances, suggesting that these networks learn very similar patterns and, therefore, are not complementary. However, the combination of distinct networks (dilated with deconvolutional ones, for instance) generates some considerable improvement in the final result for the validation set. These results may be due to the distinct pattern extraction process performed by these ConvNets. On one hand, dilated networks learn all feasible information without downsampling the input image, which allows them to capture more details mainly in small objects [12]. On the other hand, deconvolutional networks downsample the input and then learn the final prediction based on a coarse feature map, making them able to give more attention to specific patterns of the image presented in the coarse feature map [13]. In addition, the combination of distinct networks trained using the same patch size (all lines in the combination part of Table I, except for the last one) generates slightly different results with trending to networks trained with 50×50 patches. This may be justified by the fact that these networks learn distinct features given the different input context and, therefore, provide complementary views about the image content. Also, according to the last row, one may notice that the combination of all trained networks leads to an improvement in the final result. This increase is comparable to other improvements generated by different network combinations. Hence, although we believe that some combinations could produce better results, these would be marginal improvements that do not justify an extensive and costly evaluation of combination scenarios. Thus, when concerning the combination strategies, our final submission to the competition was using all previous networks.

TABLE II

IOU (%) RESULTS OF THE PROPOSED METHOD AND BASELINES FOR BOTH TEST SETS. HIGHER VALUES OF IOU INDICATES BETTER PERFORMANCE.

	Methods	Overview	Test Set	
			Same Locations	New Locations
Baselines	WISC [15]	NDVI plus SVM-RBF	80	83
		K-Means to cluster and classify	81	77
	CERTH-ITI [16]	Mahalanobis dist. with stratified cov.	75	56
	BMC [17]	ResNet-152 and random forest	37	40
	UTAOS [18]	Gen. Adv. Net. with 0.78 threshold	82	73
		Gen. Adv. Net. with 0.94 threshold	80	70
		Gen. Adv. Net. with 0.50 threshold	83	74
		Gen. Adv. Net. with 0.35 threshold	83	74
		Gen. Adv. Net. with 0.12 threshold	81	73
	DFKI [19]	VGG13-FCN with RGB data	73	69
		VGG13-FCN with RGB and NIR data	84	70
		VGG13 adapted to be a DeconvNet	84	74
Proposed	Dilated 25×25	Dilated ConvNet #1 (25×25 patches)	87	82
	Dilated 50×50	Dilated ConvNet #1 (50×50 patches)	86	80
	Location ConvNets	Dilated ConvNet #1 trained per location	87	**84**
	Fusion-SVM	SVM over concatenated predictions	**88**	**84**
	Fusion-MV	MV over concatenated predictions	78	49

B. Final Results

All results for the test sets are presented in Table II. These are the official results released by the Mediaeval since no ground truth for the test set was released yet. For both test sets, the best solution was obtained by combining the probabilities of all trained ConvNets using a linear SVM. This technique yielded the state-of-the-art results in both test set, outperforming all baselines by, at least, 4% in the *Same Locations* test set and 1% in the *New Locations* test set (in terms of the Jaccard Index). Some samples of this results are presented in Fig. 4.

For the *Same Locations* test set, training a network for each location (Location ConvNets) or training a ConvNet with all available data (ConvNet 25×25) achieved the same result. This may indicate that the proposed architecture can, in fact, extract and interpret all feasible information from the whole data, which is a great advantage given that it reduces the number of networks to train and, consequently, the processing time. This conclusion does not hold for the *New Locations* test set. In this set, training a specific network for each location

Authorized licensed use limited to: Universidade Estadual de Campinas. Downloaded on October 10,2025 at 14:54:28 UTC from IEEE Xplore. Restrictions apply.

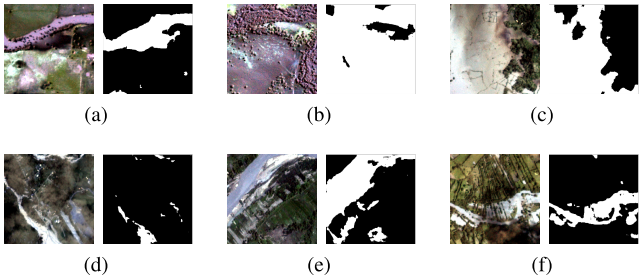

Fig. 4. Examples of some test images and the obtained results achieved by using SVM with the aggregated probabilities. White areas refer to flooded regions and black areas correspond to background. (a) Location 1. (b) Location 2. (c) Location 3. (d) Location 4. (e) Location 5. (f) Location 6.

(Location ConvNets) achieved higher performance (aside the Fusion-SVM strategy) when compared to unique networks trained with the whole training set (such as ConvNet 25×25 and ConvNet 50×50). This indicates that specific location network can learn details that may not be useful for classification of a known image but that is important for unseen data, which is the case.

Another relevant outcome is that increasing the size of the input patch (ConvNet 50×50) decreases the final result, a conclusion that holds for both data sets. We believe that this is because of the amount of training patches generated in each case. More specifically, a large amount of data may be used for training with smaller patch sizes, while large patches means less data to train. This corroborates with the fact that deep learning really needs a large amount of labeled data to train [9].

Finally, for both sets, the worst result was obtained using the majority voting scheme (Fusion-MV). It indicates that Majority Voting is not so robust to aggregate information from multiple networks when they disagree in the classification. This fact can be overcome by using a machine learning technique to capture the opinions of the ConvNets.

In addition, Fig. 4(a), which (clearly) contains a flooding river, suggests that the proposed method may be also suitable for water segmentation. Our hypothesis, left for future validation, is that the proposed methods could be able to distinguish between such classes if the data set allows a specific training for flood and normal water surface.

V. Conclusion

In this letter, we proposed and analyzed the diversity of four distinct deep networks, based on dilated convolutions [12] and deconvolution layers [13], to perform detection of flooding areas in remote sensing images. Specifically, each network was analyzed individually and in combination (with an ablation study), allowing a better understanding of their diversity, which resulted in the proposal of a combination method that aims to exploit complementary views of different networks.

Experimental results showed that the methods are effective and robust. We achieved the state-of-the-art performance, in terms of the Jaccard Index, in a specific data set proposed for the Flood-Detection in Satellite Images subtask of the 2017 Multimedia Satellite Task. The proposed methods outperformed all baselines, winning that subtask challenge. Such results show that our proposed approaches are effective and robust to identify flooding areas (independent if it is for a recurrent or atypical event). This identification process performed by our proposed algorithms may help authorities keep focus on most vulnerable regions while monitoring forecast inundations, which may aid in coordinate rescues and help victims. As a future work, we intend to analyze the proposed method using more data sets to evaluate if it will be able to distinguish between water surfaces.

References

[1] *Floods: The Awesome Power*, NOAA/NWS, Norman, OK, USA, 2010, p. 16.

[2] (2017). *Annual Disaster Statistical Review. Centre for Research on the Epidemiology of Disasters (CRED)*. [Online]. Available: http://www.emdat.be/

[3] V. Klemas, "Remote sensing of floods and flood-prone areas: An overview," *J. Coastal Res.*, vol. 31, no. 4, pp. 1005–1013, 2014.

[4] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, A. Dengel, and D. Borth, "The multimedia satellite task at MediaEval 2017: Emergence response for flooding events," in *Proc. MediaEval Workshop*, Dublin, Ireland, 2017, pp. 1–3.

[5] L. Pulvirenti, N. Pierdicca, G. Boni, M. Fiorini, and R. Rudari, "Flood damage assessment through multitemporal COSMO-SkyMed data and hydrodynamic models: The Albania 2010 case study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2848–2855, Jul. 2014.

[6] A. D'Addabbo, A. Refice, G. Pasquariello, F. P. Lovergine, D. Capolongo, and S. Manfreda, "A Bayesian network for flood detection combining SAR imagery and ancillary data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3612–3625, Jun. 2016.

[7] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, no. 7633, pp. 418–422, 2016.

[8] F. Isikdogan, A. C. Bovik, and P. Passalacqua, "Surface water mapping by deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 4909–4918, Nov. 2017.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[10] K. Nogueira, W. O. Miranda, and J. A. Dos Santos, "Improving spatial feature representation from aerial scenes by using convolutional networks," in *Proc. 28th Conf. Graph., Patterns Images (SIBGRAPI)*, 2015, pp. 289–296.

[11] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, Jan. 2017.

[12] F. Yu and V. Koltun. (2015). "Multi-scale context aggregation by dilated convolutions." [Online]. Available: https://arxiv.org/abs/1511.07122

[13] V. Badrinarayanan, A. Kendall, and R. Cipolla. (2015). "SegNet: A deep convolutional encoder-decoder architecture for image segmentation." [Online]. Available: https://arxiv.org/abs/1511.00561

[14] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[15] N. Tkachenko, A. Zubiaga, and R. Procter, "WISC at MediaEval 2017: Multimedia satellite task," in *Proc. MediaEval Workshop*, 2017, p. 2. [Online]. Available: http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_12.pdf

[16] K. Avgerinakis *et al.*, "Visual and textual analysis of social media and satellite images for flood detection @ multimedia satellite task MediaEval 2017," in *Proc. MediaEval Workshop*, 2017, p. 2. [Online]. Available: http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_31.pdf

[17] X. Fu, Y. Bin, L. Peng, J. Zhou, Y. Yang, and H. T. Shen, "BMC@MediaEval 2017 multimedia satellite task via regression random forest," in *Proc. MediaEval Workshop*, 2017, p. 2. [Online]. Available: http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_46.pdf

[18] K. Ahmad, P. Konstantin, M. Riegler, N. Conci, and P. Holversen, "CNN and GAN based satellite and social media data fusion for disaster detection," in *Proc. MediaEval Workshop*, 2017, p. 2. [Online]. Available: http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_15.pdf

[19] B. Bischke, P. Bhardwaj, A. Gautam, P. Helber, D. Borth, and A. Dengel, "Detection of flooding events in social multimedia and satellite imagery using deep neural networks," in *Proc. MediaEval Workshop*, 2017, p. 2. [Online]. Available: http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_51.pdf

Authorized licensed use limited to: Universidade Estadual de Campinas. Downloaded on October 10,2025 at 14:54:28 UTC from IEEE Xplore. Restrictions apply.