Water Segmentation with Deep Learning Models for Flood Detection and Monitoring

Mirko Zaffaroni

LINKS Foundation

University of Turin, Computer Science Department

mirko.zaffaroni@linksfoundation.com

Claudio Rossi

LINKS Foundation

claudio.rossi@linksfoundation.com

ABSTRACT

Flooding is a natural hazard that causes a lot of deaths every year and the number of flood events is increasing worldwide because of climate change effects. Detecting and monitoring floods is of paramount importance in order to reduce their impacts both in terms of affected people and economic losses. Automated image analysis techniques capable to extract the amount of water from a picture can be used to create novel services aimed to detect floods from fixed surveillance cameras, drones, crowdsourced in-field observations, as well as to extract meaningful data from social media streams. In this work we compare the accuracy and the prediction performances of recent Deep Learning algorithms for the pixel-wise water segmentation task. Moreover, we release a new dataset that enhances well-know benchmark datasets used for multi-class segmentation with specific flood-related images taken from drones, in-field observations and social media.

Keywords

Deep Learning, Water Segmentation, Data Validation.

INTRODUCTION

A flood occurs when there is an overflow of water that inundates a portion of land that is normally dry, and it can happen in several ways. For example, a flood can be caused by an excess of rainwater in saturated ground, overflow of water bodies such as rivers and lakes, rapid snow or ice melting, storm surge or tsunami. Such events are exacerbated by climate change effects, which include more intense precipitation events, higher temperature variations. Also, bad water management can cause floods, e.g., an excess of discharge from dams can cause floods in rivers downstream, negligent bank maintenance can result in their failure during water discharge peaks, a poor sewage system will not be able to handle intense rainfall events in urban areas. Most floods are induced by extreme rainfall and therefore they can be predicted up to a certain extent. Usually, floods take several days to develop, giving residents time to get ready and follow evacuation plans. Less often, floods can develop in just a few hours, i.e., the so-called flash floods, which are extremely dangerous and more difficult to predict. During these phenomena, a quiet river can rapidly turn into a flood, bringing debris and rubble along with water in its downstream path. It is difficult to quantify how human-activities affect extreme weather events. However, it is increasingly clear that climate change has influenced several variables related to flood events (Seneviratne et al. 2012). Over the last century, the steady increase in temperatures has changed the hurricanes travelling mechanics, making them slower and consequently letting them cause more intense rainfalls. At the same time, the melting of permanent ice zones and glaciers is contributing to the worldwide sea-level rise, creating an increasing threat to coastal areas and cities like Venice (Italy), which was recently hit by a flood event of historic proportions. Annually, floods cause more than $40 billion in damage worldwide (OECD 2016). From 2007 to 2016, 5553 people died because of floods, while in 2017 alone the death toll reached 3331 (R. and P. 2018). The aforementioned figures demonstrate the increasing severeness of floods, pointing to the need for novel approaches and tools aimed to reduce floods impact worldwide. The monitoring for water flows is key to implement effective early warnings, while the analysis of in-field data can contribute to early event detection and to the real-time understanding of flood impacts. To achieve

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.

66

these goals, the automated analysis of images and videos through algorithm based on Artificial Intelligence can be of great importance, especially when applied to heterogeneous data coming from multiple sources, including fixed surveillance cameras installed near river beds or shores, geolocated images taken from drones or other aerial vehicles (Restas 2015), posts extracted from social medias (Acerbo and Rossi 2017), in-field pictures generated by ad-hoc crowdsourced mobile applications (Nguyen et al. 2017).

Therefore, we study the accuracy and the prediction performances of Deep Learning models for the detection of water in images. Specifically, we target the pixel-wise semantic segmentation of images to separate water elements from everything else, which we consider as background. This approach is not only able to understand if a given image contains a water element, but also to quantify the number of water pixels and delineate the shape of water elements. These capabilities enable multiple use cases, which can span from river monitoring through a fixed camera, flood mapping from drones, data validation in case of crowdsourced data collection activities and social media analysis. Note that image water segmentation can be effectively used within early warning systems, which can be triggered to send alerts when a certain threshold of water is reached, as well as in the emergency response phase to monitor the flood evolution. However, existing machine learning models trained for the semantic segmentation task may perform poorly in detecting water in a heterogeneous scenario, where the images to be analyzed could be produced at different resolutions, angles, lighting conditions, altitude and distance from the observed phenomena.

The contribution of this paper is two-fold. First, we introduce a new dataset that increases the water segmentation performances when images from different sources are considered, Second, we study the accuracy and the prediction performances of the most recent and advanced Deep Learning models for the semantic water segmentation tasks with and without the use of our dataset.

In the following sections, we first review the related works and then introduce our new dataset. Next, we describe the methodology we applied in our analysis, briefly explaining the selected Deep Learning algorithms. Then, we present the outcomes of the evaluation, comparing the performances of the selected algorithms when using previous datasets as well as when using ours. Finally, we outline the main conclusion of our study.

RELATED WORKS

Social media platforms offer the possibility of easily obtaining a lot of data about natural disasters- However, information extracted from open systems such as social media platforms would require validation before to be trusted and operationally used in emergency management contexts. For this reason, previous works focused on the use of Artificial Intelligence to analyze and classify this kind of data (Imran, Castillo, Diaz, et al. 2015; Imran, Castillo, Lucas, et al. 2014; Rossi et al. 2018). However, the majority of such works mainly use the text of social media posts to implement a binary classifier that retains only informative content. Another kind of previous works can be found in the computer vision domain (Lai et al. 2007; Filonenko et al. 2015), where the use of machine learning algorithm with handcrafted features is coupled with image processing techniques such as background subtraction, morphological operators and colour probability, light intensity, texture or colour clustering. The shortcomings of such approaches are mainly linked to the use of hand-crafted features, which usually tend to not generalize well and to poorly perform when applied to data coming from different data sources or context. Moreover, a comparison with the majority of such works is often not possible because they were evaluated in very specific scenarios and with non-publicly available datasets.

Our goal is different, we want to study the performances of recent Deep learning models for the pixel-wise water segmentation task in order to allow the automatic analysis of images coming from multiple sources, namely from social media, surveillance cameras placed near the water beds, drones, and from in-field operators. Moreover, we do not rely on handcrafted features, but we apply a data-driven semantic segmentation approach with automatic feature selection. Semantic segmentation is a common task in image processing and analysis, and it consists in assigning to each pixel a label, thus obtaining a set of regions in output. Image segmentation can be used to separate foreground from the background or to cluster regions of pixels based on common properties, such as colour or shape. Therefore, we compare the performances of recent Deep Learning models used for multi-class semantic segmentation (Ronneberger et al. 2015; Badrinarayanan et al. 2015; Zhao et al. 2016; L. Chen et al. 2016; Wu et al. 2019; Takikawa et al. 2019) trained on well-known benchmark datasets, showing how their performances can be increased through the use of a dedicated dataset, specifically designed to fit the water segmentation task. We include in our performance analysis, also the Tiramisu model, which resulted in the best algorithm for the water segmentation task when trained a specific dataset containing riverine flood images (Lopez-Fuentes, Rossi, and Skinnemoen 2017). Overall, we compare the classification accuracy and the prediction performances of four top-notch state of the art Deep Learning algorithms using different combinations. We train such algorithms on a novel dataset that we create extending a collection of previous benchmark multi-label datasets with a set of new

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   67

Table 1. Water Segmentation Open Collection (WSOC) key metrics.

Metric	Value
Min size of image	[147, 150] px
Max size of image	[2448, 3264] px
Mean size of image	[612, 465] px
Min percentage of water	10%
Max percentage of water	95%
Mean percentage of water	35%
Number of images	120061
Number of images with water presence	11900

labelled images covering the above-mentioned data sources. Our new dataset, as well as the tested algorithms, are presented in the next sections.

THE WATER SEGMENTATION OPEN COLLECTION DATASET

In this section we introduce the Water Segmentation Open Collection (WSOC), a new dataset aimed to effectively train deep Learning water segmentation algorithms. WSOC is composed by a collection derived from pre-existent publicly available datasets for image segmentation, which have been binarized in order to be used for the water segmentation task and enhanced with an additional dataset specifically created to increase the performances of the water delineation task. The public collection is composed by well-known benchmark multi-class segmentation datasets, namely COCO (Lin et al. 2014), the Semantic Drone Dataset (Mostege et al. 2019), the Microsoft Research in Cambridge v2 (MSRC v2) (Criminisi 2005), as well as by other open datasets containing water-related classes (e.g., sea, shore, river, lake, etc..), namely the Video Label Propagation (A. Y. C. Chen and J. 2010) and the River Dataset (Lopez-Fuentes and Rossi 2017). We extend this aggregated dataset with additional 490 images, which we gather from Twitter and online news using "floods" as keyword for the search query. We label the new images using human annotators with the aid of a segmentation tool as well as by using crowd-sourcing online annotation services. For each image, we select three human annotators: the first one to realize the initial annotation, the second one to refines it, and the third one to validate it. In case the final validation fails, the labelling is discarded and re-assigned. Overall, the WSOC dataset is composed of 120061 images and their labelled ground truth, which consists of binary masks where zero is assigned to the background pixels and one to the water pixels. We openly publish WSOC at the following link: https://zenodo.org/record/3642406, so that it could be further extended and used in future works.

In Table 1 we report the key WSOC metrics, including the amount of water present in the images and the variation in terms of image size. The biggest images are the drone aerial images, while the smallest ones are the pictures shared in social media.

METHODOLOGY

We decided to test WSOC with four different Deep Learning algorithms for image segmentation, testing each of them with different backbones, namely: VGG16 (Simonyan and Zisserman 2014), ResNet50 (He et al. 2015) and MobilNet (Howard et al. 2017). A backbone is the initial part of a deep learning model for image segmentation and it is usually used for feature extraction. Having a good feature extractor is crucial, because it has been demonstrated that high-quality features enable better performance on many machine learning tasks (Kornblith et al. 2018). We selected VGG16, ResNet50 and MobilNet because they are among the most recent and performing models for image classification that allows being re-trained and specialized for semantic segmentation task by substituting the last classification layers with other Deep Learning models. Therefore, we use such backbones to exploit the features they learned on another dataset. VGG16 is a classic CNN, one of the first Deep learning networks that outperformed classic machine learning algorithms on image classification. Instead, ResNet50 is a more recent model that uses a new kind of connection between subsequent layers called *residual block*, which have been proven to increase the performance on very deep models by allowing better information transfer through layers. MobileNet is a simplified CNN that has been designed to be used in embedded systems at a cost of a slight performance loss. The backbone is usually pretrained on well-known benchmark dataset such as ImageNet (Russakovsky et al. 2014), COCO (Lin et al. 2014) or PascalVOC (Everingham et al. n.d.). We used all the aforementioned backbones pre-trained on ImageNet, because it contains 1000 classes and therefore it is able to create very rich and different features. For this reason, models pre-trained on ImageNet are considered good feature extractors. We evaluate each backbone in combination

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   68
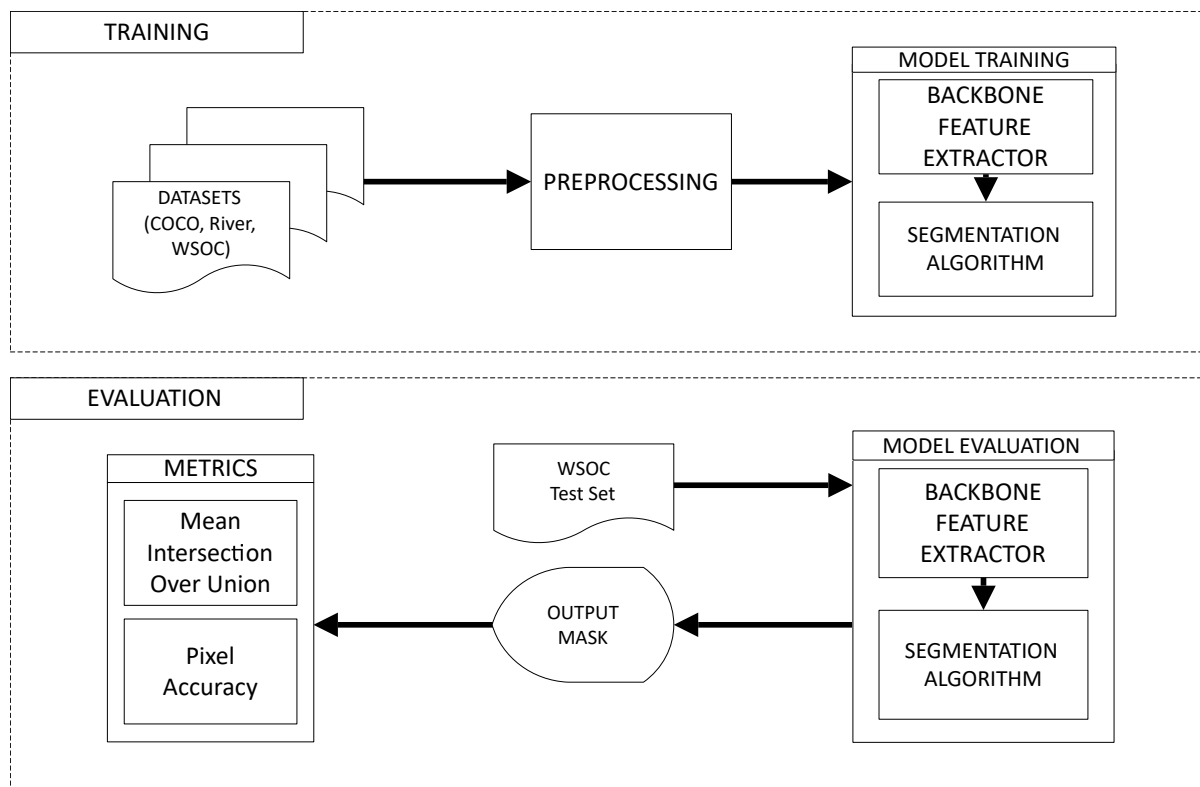
Figure 1. A graphical example of our proposed methodology: images contained inside the dataset are preprocessed and then used to train the model, which is composed by a backbone network which is responsible for the features extraction and a segmentation algorithm which uses the features to create a segmentation mask. Then the trained model is evaluated on Mean Intersection over Union (MIoU) and Pixel Accuracy (PA). Those metrics are used to score the model performances.

with recent Deep segmentation models designed for semantic segmentation, namely: SegNet (Badrinarayanan et al. 2015), PSPNet (Zhao et al. 2016), UNet (Ronneberger et al. 2015) and FCN32 (Long et al. 2015). SegNet is a fully convolutional autoencoder (Baldi 2012), where each encoding layer is connected with his decoding layer with which pooling information are shared. Thanks to this particular architecture, the model is able to share higher-resolution information between layers, leading to better segmentation accuracy. The UNet was developed for biomedical image segmentation and its architecture was created in order to work with few training images without sacrificing segmentation accuracy. Its name is due to his U-shaped architecture where the network uses connections between layers to propagate context information to higher resolution layers. PSPNet uses a pyramid pooling module to transfer information from the higher layer to the lower ones. The pyramidal structure aims at passing additional contextual information between layers. FCN32 uses a deep fully convolutional network that combines semantic information from coarse layer with outputs of fine layer to produce accurate and detailed segmentations. Our goal is to test and evaluate all the aforementioned segmentation networks with the different backbone and compare them with Tiramisu, which has been previously trained and evaluated for the water segmentation task using only the River Dataset. We evaluate each model combination on three parameters: accuracy, prediction speed and memory usage (size and VRAM).

EVALUATION METRICS AND CONFIGURATIONS

We divided the WSOC dataset between training and test set that we use to train the models and evaluate their performances, respectively. The training set is composed by all WSOC images including at least one water pixel (11900 images), excluding the test ones, which we select as the 10% (139 images) of a WSOC subset composed by the following categories: the River, the Semantic Drone Dataset and the newly added images. We select as test set this subset because such images are the most representatives for the flood monitoring scenarios we are considering. We use only images containing at least one water pixel because this configuration leads to the best accuracy, although the test set contains images that have no water element. The models have been trained using Tversky loss function (Salehi et al. 2017) with $\alpha = 0.2$ and $\beta = 0.8$. Note that a high value for β leads to higher sensitivity (recall) and to lower specificity (Glaros and Kline 1988), which is the desired condition for our task

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   69

Table 2. Model accuracy comparison in terms of average and standard deviation of Mean Intersection over Union (MIoU) and Pixel Accuracy (PA) with different training dataset (COCO, River, and WSOC) and with WSOC test set.

	COCO				River				WSOC			
	MIoU [%]		PA [%]		MIoU [%]		PA [%]		MIoU [%]		PA [%]	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Tiramisu	0.35	0.20	0.65	0.22	0.33	0.19	0.63	0.22	0.38	0.17	0.73	0.24
SegNet-M	**0.80**	**0.10**	**0.90**	**0.05**	0.64	0.13	**0.75**	**0.17**	**0.85**	**0.07**	0.92	0.05
UNet-M	0.76	0.12	0.86	0.08	0.64	0.13	0.75	0.20	0.83	0.09	**0.92**	**0.03**
FCN32-M	0.62	0.15	0.73	0.17	**0.67**	**0.13**	0.74	0.17	0.77	0.11	0.86	0.08
PSPNet-M	0.75	0.12	0.88	0.08	0.62	0.13	0.74	0.19	0.81	0.08	0.90	0.07
SegNet-R	**<u>0.82</u>**	**<u>0.10</u>**	0.90	0.05	**<u>0.74</u>**	**<u>0.12</u>**	**<u>0.83</u>**	**<u>0.13</u>**	0.85	0.08	**<u>0.94</u>**	**<u>0.01</u>**
UNet-R	0.81	0.10	<u>0.92</u>	<u>0.01</u>	0.55	0.14	0.69	0.23	**<u>0.88</u>**	**<u>0.06</u>**	0.94	0.03
FCN32-R	0.79	0.12	0.75	0.12	0.59	0.13	0.74	0.19	0.79	0.10	0.87	0.08
PSPNet-R	0.81	0.10	0.90	0.04	0.62	0.12	0.74	0.22	0.83	0.08	0.90	0.07
SegNet-V	**0.71**	**0.16**	0.87	0.05	**0.69**	**0.13**	**0.79**	**0.15**	0.82	0.09	0.90	0.05
UNet-V	0.49	0.27	0.68	0.13	0.30	0.03	0.59	0.31	0.50	0.26	0.67	0.15
FCN32-V	0.67	0.19	0.90	0.03	0.58	0.16	0.70	0.19	0.76	0.13	0.91	0.02
PSPNet-V	0.61	0.23	**0.91**	**0.02**	0.12	0.10	0.61	0.38	**0.82**	**0.09**	**0.92**	**0.02**
	X: Best score model-wise				<u>X</u>: Best score metric-wise							

Table 3. Models performance comparison in terms of VRAM, size and prediction speed.

	ResNet50			MobileNet			VGG16		
	Size [Mb]	VRAM [Mb]	Prediction time [ms]	Size [Mb]	VRAM [Mb]	Prediction	Size [Mb]	VRAM [Mb]	Prediction time [ms]
SegNet	**57**	**2461**	27	**21**	**2665**	**9**	**44**	**1193**	22
UNet	62	2563	29	24	2674	10	47	1792	21
FCN32	1732	4188	250	867	3255	115	516	3334	56
PSPNet	114	2557	**24**	37	4352	11	65	1824	**16**

because we want to predict as much water pixels as possible while accepting lower performance on the background ones. As optimizer was used *Adadelta* with the following parameters: $lr = 0.001$ and $\rho = 0.95$ (Zeiler 2012). In order to avoid over-fitting we stop the training as soon as the loss computed on the validation set ceases to decrease for 5 epochs. We evaluate the results of each algorithm on the test set with the two most commonly used metrics for segmentation tasks (Everingham et al. 2010; Thoma 2016), namely the Mean Intersection over Union (MIoU):

$$MIoU = \frac{\frac{1}{C} \sum_i^N n_{ii}}{t_i + \sum_j^N (n_{ij} - n_{ii})}$$

and the Pixel-wise Accuracy (PA):

$$PA = \frac{\sum_i^N n_{ii}}{\sum_i^N t_i}$$

where n_{ij} corresponds to the number of pixels from class i which have been wrongly classified as belonging to class j, n_{ii} represents the pixels from class i which have been correctly classified, C is the total number of classes (C=2 in our case, because we are only trying to predict water pixels in the scene) and t_i is the total number of pixels belonging to class i. For each of these metrics we calculated the mean value and the standard deviation. We perform both the training and the testing on a desktop workstation with an Intel(R) Core(TM) i9-7940X CPU, 128 GB RAM and 4 NVIDIA GeForce GTX 1080 Ti.

RESULTS

In Table 2 we report the results in terms of prediction accuracy, where we highlight the best model both metric-wise and model-wise. Such results are obtained by scoring all the models on the same test set while varying the training set. Each model is scored in terms of Pixel-wise Accuracy and Mean Intersection over Union. Model-wise, SegNet

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   70
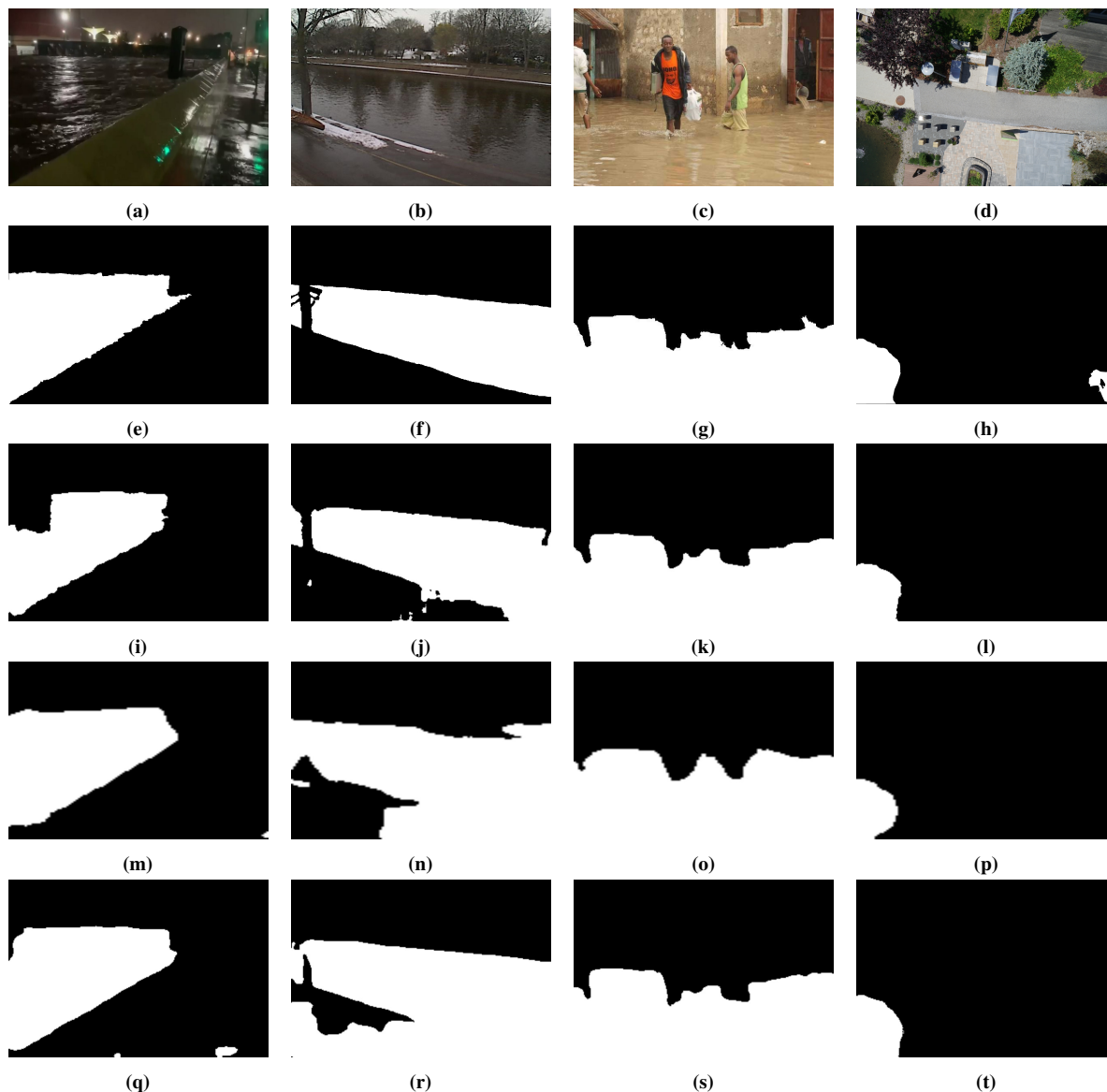
Figure 2. Results of the best performing algorithms on the best sample of the validation set, where (a-d) corresponds to the original images, (e-h) are the ground truths, (i-l) SegNet with ResNet50 as backbone, (m-p) SegNet with MobileNet as backbone, (q-t) UNet with ResNet50 as backbone.

has the higher number of best scoring entries, although sometimes it was slightly worse than the UNet. Metric-wise, the models with ResNet50 as backbone result consistently as the best performing. This increase in performance is due to the residual layers of ResNet50 which, as reported in the original paper, allow a better backpropagation of the gradient during training, thus increasing the accuracy of the model. Moreover, both model-wise and metric-wise, the models trained with WSOC achieve increased accuracy against the training done with COCO and the River Dataset. Moreover, the Tiramisu model, which was our reference benchmark, achieved by far the worst performances.

In Table 3 we present the performance of the model in terms of model size and prediction time for one image. Such results depict a different situation, where SegNet is clearly the lightest model in terms of size (memory usage) and it is also the fastest one when coupled with MobilNet as backbone. This combination (MobileNet plus SegNet) is the winning one both in terms of size, VRAM and speed, being three times faster compared with ResNet50 as backbone.

Overall, we can state that the ResNet50 backbone is the best one in terms of accuracy. While we do not have a unique winner model-wise, with UNet and Segnet being the best options provided that there are no particular needs in terms of memory and prediction speed. Whenever the application requires faster prediction and lighter models, able to get good results and in a shorter time, SegNet with MobileNet is the best choice. Moreover, we note that the introduction of WSOC was able to increase the performances of all models in the considered test scenario.

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   71

Figure 2 shows representative images from the test set (first row) with their relative ground truths (second row) and the predicted binary masks (third, fourth, fifth rows) of the following backbone-model combinations: SegNet-ResNet50, SegNet-MobileNet and Unet-ResNet50. In Figure 2 (a) is taken from Twitter, (b) from the River dataset, (c) from online news and (d) from the drone dataset. As it can be seen, the models that use ResNet give more fine-grained masks, whereas results with MobileNet are slightly coarser. Specifically, it can be observed that in (k) and (s) the people legs are clearly defined, the same applies in (j) and (r) where the shape of the structure of the tree is at least sketched, although it is not completely identified.

CONCLUSION AND FUTURE WORKS

In this work we presented and released a new dataset for water segmentation (WSOC), enhancing a previous collection of publicly available datasets composed by water-related labels (river, sea, waves, waterdrops, etc...). This collection of data is useful for the task of water segmentation, in particular for implementing services aimed at flood event monitoring and information validation. Moreover, we studied several state of the art Deep Learning models for semantic segmentation, combining different backbones used for image labelling with semantic segmentation models. We trained and tested all models with WSOC, comparing their performances against training done with previously available datasets. *SegNet* ad UNet, both with ResNet as backbone, performed slightly better than the other algorithms in terms of accuracy, while SegNet was the best performing in terms of prediction speed and memory usage, VRAM and size. Considering accuracy, ResNet was the best performing backbone overall, but it was also the slowest. Therefore, the model to be used for the water segmentation task should be selected according to the application requirements in terms of accuracy and prediction speed.

Future works will include the creation of an operational service aimed to monitor flood-prone areas and to support social media analysis. Moreover, we plan to further improve the dataset by evaluating the possibility of using synthetically generated data.

ACKNOWLEDGEMENTS

REFERENCES

Acerbo, F. S. and Rossi, C. (2017). "Filtering Informative Tweets During Emergencies: A Machine Learning Approach". In: *Proceedings of the First CoNEXT Workshop on ICT Tools for Emergency Networks and DisastEr Relief*. I-TENDER '17. Incheon, Republic of Korea: ACM, pp. 1–6.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015). "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *CoRR* abs/1511.00561. arXiv: 1511.00561.

Baldi, P. (July 2012). "Autoencoders, Unsupervised Learning, and Deep Architectures". In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Ed. by I. Guyon, G. Dror, V. Lemaire, G. Taylor, and D. Silver. Vol. 27. Proceedings of Machine Learning Research. Bellevue, Washington, USA: PMLR, pp. 37–49.

Chen, A. Y. C. and J., C. J. (2010). "Propagating Multi-class Pixel Labels Throughout Video Frames". In: *Proceedings of Western New York Image Processing Workshop*.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2016). "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *CoRR* abs/1606.00915. arXiv: 1606.00915.

Criminisi, A. (2005). *MSRC-v2 image database*.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (June 2010). "The Pascal Visual Object Classes (VOC) Challenge". In: *International Journal of Computer Vision* 88.2, pp. 303–338.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (n.d.). *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

Filonenko, A., Wahyono, W., Hernandez, D., Seo, D., and Jo, K.-H. (Nov. 2015). "Real-time flood detection for video surveillance". In: pp. 004082–004085.

Glaros, A. G. and Kline, R. B. (1988). "Understanding the accuracy of tests with cutting scores: the sensitivity, specificity, and predictive value model." In: *Journal of clinical psychology* 44 6, pp. 1013–23.

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.

72

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385. arXiv: 1512.03385.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". In: *CoRR* abs/1704.04861. arXiv: 1704.04861.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. (June 2015). "Processing Social Media Messages in Mass Emergency: A Survey". In: *ACM Comput. Surv.* 47.4, 67:1–67:38.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014). "AIDR: Artificial Intelligence for Disaster Response". In: *Proceedings of the 23rd International Conference on World Wide Web*. WWW '14 Companion. Seoul, Korea: ACM, pp. 159–162.

Kornblith, S., Shlens, J., and Le, Q. V. (2018). "Do Better ImageNet Models Transfer Better?" In: *CoRR* abs/1805.08974. arXiv: 1805.08974.

Lai, C. L., Yang, J. C., and Chen, y.h. (2007). "A Real Time Video Processing Based Surveillance System for Early Fire and Flood Detection". In: *2007 IEEE Instrumentation and Measurement Technology Conference IMTC 2007*, pp. 1–6.

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). "Microsoft COCO: Common Objects in Context". In: *CoRR* abs/1405.0312. arXiv: 1405.0312.

Long, J., Shelhamer, E., and Darrell, T. (June 2015). "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440.

Lopez-Fuentes, L., Rossi, C., and Skinnemoen, H. (Dec. 2017). "River segmentation for flood monitoring". In: *2017 IEEE International Conference on Big Data (Big Data)*, pp. 3746–3749.

Lopez-Fuentes, L. and Rossi, C. (Oct. 2017). *River segmentation dataset*.

Mostege, C., Maurer, M., Heran, N., Puerta, J. P., and Fraundorfer, F. (2019). *Semantic Drone Dataset*.

Nguyen, Q. N., Frisiello, A., and Rossi, C. (2017). "Co-design of a Crowdsourcing Solution for Disaster Risk Reduction". In: *Proceedings of the First CoNEXT Workshop on ICT Tools for Emergency Networks and DisastEr Relief*. I-TENDER '17. Incheon, Republic of Korea: ACM, pp. 7–12.

OECD (2016). *Financial Management of Flood Risk*, p. 136.

R., B. and P., W. (2018). *Annual Disaster Statistical Review 2017*.

Restas, A. (2015). "Drone Applications for Supporting Disaster Management". In: *World Journal of Engineering and Technology*. Scientific research, pp. 316–321.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *CoRR* abs/1505.04597. arXiv: 1505.04597.

Rossi, C., Acerbo, F., Ylinen, K., Juga, I., Nurmi, P., Bosca, A., Tarasconi, F., Cristoforetti, M., and Alikadic, A. (2018). "Early detection and information extraction for weather-induced floods using social media streams". In: *International Journal of Disaster Risk Reduction* 30, pp. 145–157.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. S., et al. (2014). "ImageNet Large Scale Visual Recognition Challenge". In: *CoRR* abs/1409.0575. arXiv: 1409.0575.

Salehi, S. S. M., Erdogmus, D., and Gholipour, A. (2017). "Tversky loss function for image segmentation using 3D fully convolutional deep networks". In: *CoRR* abs/1706.05721. arXiv: 1706.05721.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., Mcinnes, K., Rahimi, M., et al. (Apr. 2012). "Changes in climate extremes and their impacts on the natural physical environment: An overview of the IPCC SREX report". In: pp. 12566–.

Simonyan, K. and Zisserman, A. (Sept. 2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *arXiv 1409.1556*.

Takikawa, T., Acuna, D., Jampani, V., and Fidler, S. (2019). "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation". In: *CoRR* abs/1907.05740. arXiv: 1907.05740.

Thoma, M. (2016). "A Survey of Semantic Segmentation". In: *CoRR* abs/1602.06541. arXiv: 1602.06541.

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   73

Wu, H., Zhang, J., Huang, K., Liang, K., and Yu, Y. (2019). "FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation". In: *CoRR* abs/1903.11816. arXiv: 1903.11816.

Zeiler, M. D. (2012). "ADADELTA: An Adaptive Learning Rate Method". In: *CoRR* abs/1212.5701. arXiv: 1212.5701.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2016). "Pyramid Scene Parsing Network". In: *CoRR* abs/1612.01105. arXiv: 1612.01105.

WiP Paper – AI Systems for Crisis and Risks
Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020
Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, eds.   74