Water level prediction from social media images using deep learning

P. Chaudhary, S. D'Aronco, J.P. Leitao, K. Schindler, J.D. Wegner

Floods are among the most frequent and catastrophic natural disasters and affect millions of people worldwide. It is important to create accurate flood maps to plan flood mitigation and rescue operations. We introduce a computer vision system that estimates water depth from social media images in order to build flood maps in near real time. Our approach is motivated by the observation that a main bottleneck for image-based water level estimation is training data: it is difficult and requires a lot of effort to annotate images with the correct depth. We demonstrate how to effectively learn a predictor of water level from a small set of annotated water levels and a larger set of weaker annotations that only indicate in which of two images the water level is higher, and are much easier to obtain.

Überschwemmungen gehören zu den häufigsten und katastrophalsten Naturkatastrophen und betreffen weltweit Millionen von Menschen. Um Hochwasserschutzmassnahmen und Rettungsaktionen zu planen, ist die Verfügbarkeit genauer Hochwasserkarten von grosser Bedeutung. Wir stellen ein Computer-Vision System vor, das es ermöglicht, den Wasserstand während eines Hochwassers anhand von Bildern aus sozialen Medien zu schätzen, um somit Hochwasserkarten nahezu in Echtzeit zu generieren. Unser Ansatz ist durch die Beobachtung motiviert, dass die Anzahl Trainingsdaten den Hauptengpass bei der bildbasierten Schätzung des Wasserstands darstellt: Es ist schwierig und erfordert viel Aufwand, den richtigen Wasserstand für Bilder zu bestimmen. Im Gegenzug ist es für den Menschen viel einfacher zu bestimmen, in welchem Bild eines Bildpaares der Wasserstand höher ist. Wir zeigen, wie ein Modell für die exakte Wasserstandschätzung aus Bildern gelernt werden kann, indem wir eine kleine Anzahl von exakt annotierten Bildern mit einer grösseren Anzahl von Bildern mit einer paarweisen – schwächeren – Annotation ergänzen.

Les inondations appartiennent aux catastrophes naturelles les plus fréquentes et désastreuses et touchent des millions de personnes au monde. Afin de planifier des mesures de protection contre les crues et d'organiser des actions de sauvetage la disponibilité de cartes de crues précises est d'importance primordiale. Nous présentons un système de vision sur ordinateur permettant d'évaluer les niveaux d'eau à l'aide d'images diffusées par des médias sociaux afin de créer des cartes de crues quasiment en temps réel. Notre approche est motivée par l'observation que le nombre des données d'entraînement constitue le principal obstacle lors de l'évaluation du niveau d'eau sur la base d'images. Cela est difficile et demande beaucoup d'effort pour déterminer le vrai niveau d'eau pour des images. En revanche il est bien plus simple pour l'oeil humain de voir dans quelle paire d'images le niveau d'eau est plus élevé. Nous expliquons un modèle qui montre comment l'on peut apprendre à évaluer avec exactitude le niveau d'eau sur la base d'images en complétant un petit nombre d'images annotées exactement avec un plus grand nombre d'images par deux annotées plus faiblement.

1. Introduction

The frequency of weather-related disasters is increasing rapidly: During the period of 1995–2015, floods have accounted for 47 % of all weather-related disasters and have affected over 2 billion people [1]. The number of floods has also soared up to an average of 171 floods per year between 2005–2014, compared to 127 floods per year during 1995–2004 [1]. To mitigate the damage caused by such flood events and for effective disaster response and emergency plans, the rapid analysis of data collected from the affected area is essential. There are various sources from where observations can be gathered: stream gauge data, remote sensing data and field data collection. The field data collection approach consists of sending people to the affected areas to survey and document data after the flood event. However, implementing this approach in real-time is expensive, labour intensive and difficult to obtain. Data collected from stream gauges provide accurate, near real-time information of water height for the monitored locations, but gauges are sparsely distributed leading to extremely sparse observations. Due to these dispersed locations the information provided is often not sufficient to map the flooded area.

The unprecedented global spread of low-cost sensors, especially in smartphones, together with the rise of the internet and social media, opens the possibility of community-based mapping initiatives. Recognition is increasing for the utility of social media when it comes to capturing real-time information during and immediately after a flood, using «citizens-as-sensors» [1]. Gathering this real-time information might be useful to improve rescue operations in episodes of flash floods or similar, where satellites do not always offer a viable source of information. In earlier work [2] we have presented a model to predict flood height from imag-

[1] For more information on citizen science projects, please refer to this link: https://citizenscience.ch/de/

es gathered from social media platforms in a fully automated way using object instance segmentation and predicted water level whenever an instance of some specific object was detected. Although the trained model performs rather well, the effort required to build a large, pixel-accurate annotated dataset for instance segmentation of flood images is considerable. To tackle this problem, we propose in this paper a deep learning approach where we define the flood estimation as a per-image *regression* problem and combine it with a *ranking loss* to further reduce the labelling load. We propose to avoid the tedious, and hardly scalable, procedure of pixel-accurate object instance labelling per image by *(i)* directly regressing one representative water level value per image and, more importantly, *(ii)* exploiting relative ranking of the water levels in pairs of images, which is much easier to annotate. Estimating an absolute water level from individual images is hard for humans. We thus pursue the strategy also used in our previous work [2] and look for partially submerged objects of roughly known size as «scale bars». We consider 11 water levels, from *level0,* which means no water, to *level10,* which represents a human body of average height completely submerged in water.

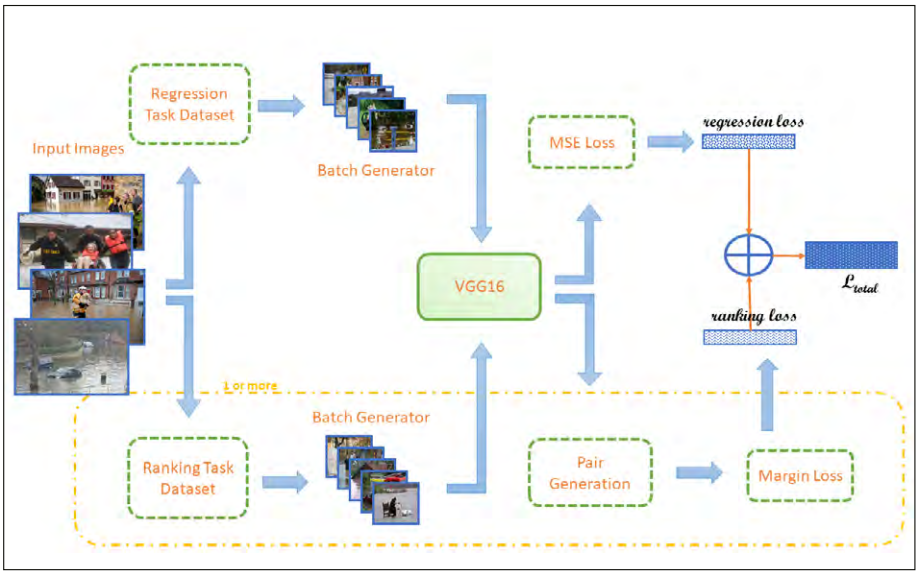

Fig. 1: The architecture of our multi-task learning with ranking loss method. MSE is mean squared error. L_{total} is the total loss function for the model. The phrase «1 or more» in the figure means that for more than 1 image pair we use more than one batch generator (for more details see [8]).

2. Methodology

Moving from pixel-accurate object delineation as in [2] to annotating only a single water depth per image comes at a price. While the regression task might, in principle, be easier than detailed object detection and segmentation, the ground-truth information it provides to a machine learning system is much weaker (or less quantitatively) because we no longer tell the system to turn its attention to certain types of objects that reoccur with similar height in an image. Furthermore, even in the presence of known objects it is often hard for a human operator to determine the water depth of individual images with an absolute value. On the contrary, it is a much simpler task to rank images via pairwise comparisons. People can, with no or little training, quickly decide which of two images shows a higher water level. In this way it becomes feasible to outsource the labelling effort to large groups of untrained annotators, for instance through an online/internet tool. Using ranking as a complementary task can be seen as a variant of *weak supervision,* or alternatively the ranking information can be interpreted as a *regulariser* for the otherwise data-limited regression task. The idea is that a large volume of weaker ranking labels should be able to largely compensate for the small amount of strong water depth labels, and lead to better regression performance. Here, by the term strong and weak, we mean the information quantity of ground-truth we provide to the model during training. As providing the water depth label for an image is giving more information for the model to learn rather than telling which image has just relative higher water depth label in an image pair, is more valuable but also more difficult to acquire.

We design a deep learning approach that combines global per-image regression and relative, pairwise ranking. The overall architecture of the proposed method is shown in Fig. 1. The backbone of our network architecture consists of a VGG16 [3] network pre-trained on the ImageNet dataset, but any standard network architecture could be used here. We replace the final layers of the network to predict water depth. Because the method does absolute water level estimation per image as well as relative ranking simultaneously, we feed two separate training sets to the model. The first part (regression task dataset) has a known absolute flood water level for each image – that is still necessary, since one cannot predict the absolute water depth value with only relative measurements. The second part (ranking task dataset) only knows the ordering relation for each pair of images. The images from the regression training set are fed to the network in conventional mini-batches. For those images, we use a standard mean squared error regression loss to train the network parameters (through back-propagation).

For the images that belong to the ranking training set, the procedure is slightly different. We first prepare a mini-batch of images and feed it to the network to obtain a water level prediction for each of them. For these images we cannot evaluate the regression loss, as we do not

have access to the ground truth values. We can, however, assemble all possible image pairs and test whether they follow the ground truth ranking. We jointly learn the regression sub-task and the ranking sub-task, by defining the total loss function as sum of regression and ranking loss and a weighting parameter to balance the contributions of the two terms. At test time, the network only receives a single image and pushes it through the regression task to obtain a water level. It takes approximately three seconds for the model to predict a water level per image. The ranking task is not used for testing as it acts as a *regulariser* for predicting water depth.

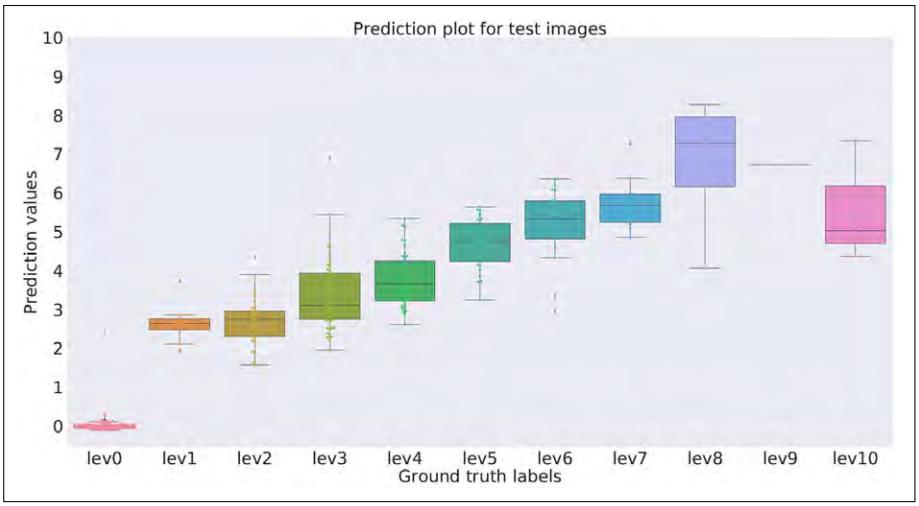

Fig. 2: Prediction plot for *Reg+Rank* experiment on test images. Each box shows the minimum predicted flood level (after excluding outliers), the 25–50–75 percentiles, and the maximum level (after excluding outliers). Note that only a single sample exists for lev9. The different colors are for visualisation purpose only, and have no other meaning.

3. Dataset and Experiments

We built a new dataset *DeepFlood* that, in total, contains 8145 ground-level images with water level annotations and extends our original dataset of [2]. From that earlier work, there are 1259 images with pixel-level object annotations. Additionally, *DeepFlood* has 5395 flood images with only a single flood depth label per image. Moreover, we add 1491 images from the Mapillary Vistas dataset [4]. These images have similar characteristics and scene content as our flood images. The images from *DeepFlood* dataset are required for the network to learn how scenes from non-flooded areas look like, as the images in the *DeepFlood* dataset are from various flood events and there are no non-flooded images. Mapillary has pixel-level instance annotations for 37 classes, we randomly pick images from the Mapillary training set that contain at least one of the objects *Person, Car, Bus, Bicycle or Building/House* that act as basis for our water-level estimation approach.

We partition *DeepFlood* into two separate sub-datasets *DF-Obj* and *DF-Img*. *DF-Obj* contains 1862 images (1259 with flooding from our previous database, 603 Mapillary images without flooding) that all have pixel-accurate object instance annotations and annotations of the water level per object. *DF-Img* contains 6283 images (5395 with flooding, 888 without flooding) that are annotated with a single water level per image, which is zero for images without flooding. The *DF-Obj* subset makes it possible to compare to our earlier, object-driven work [2], for which instance-level segmentations are required during training. We evaluate our method *Reg+Rank on the DeepFlood* dataset and compare against [2] *Classification,* and two baselines approaches *(Regression and Regression++):*

- **Regression:** A pure regression network without additional supervision with ranked pairs, equivalent to *Reg+Rank* with only the regression loss, trained on *DF-Obj.* This regression-only approach with a small training set and no pair regularisation serves as sanity check and lower performance bound.
- **Regression++:** Uses the same network and loss function as Regression, but is trained on a combination of *DF-Obj* and *DF-Img,* using absolute water levels for all training images as supervision. This corresponds to the ideal case where strong supervision by regression targets is available for the entire training dataset, and serves as an upper bound for the possible performance of *Reg+Rank*.
- **Classification:** This is the object-driven approach, where water levels are predicted via object detection and segmentation, using pixel-accurate object instance masks as supervision. Here we use a ResNet101 [5] and Feature Pyramid Network (FPN) [6] as backbone and train on the *DF-Obj* subset, for which the necessary ground truth masks are available.
- **Reg+Rank:** We evaluate our proposed multi-task ranking approach, which combines ranking loss and regression loss, as described in Sec. 2. We train the regression loss on *DF-Obj* with the absolute water level labels per image like for *Regression.* Our ranking loss is trained on the *DF-Img* data subset but, unlike *Regression++,* without using absolute water levels per image. Instead, each image inside a pair of images is only labelled to have a higher water level than the other image (same level does not influence the prediction).

We use VGG16 [3] pre-trained on ImageNet as network backbone for *Reg+Rank, Regression,* and *Regression++* experiments. The object-driven classification approach is an extension of Mask R-CNN [7], hence we use a ResNet-101-FPN backbone, as suggested by the creators of Mask R-CNN [7].

4. Results

We compare the proposed multi-task ranking approach *(Reg + Rank)* with *Regression, Regression++* and *Classification* [2]. All results are shown in Tab. 1. As expected, *Reg + Rank* outperforms *Regression* trained only on the *DF-Obj* data subset. The ≈ 22 % drop in avgRMSE (cm) is the benefit one gets from additional ranked pair supervision. More interestingly, the multi-task *(Reg + Rank)* approach performs almost on par with the upper bound *Regression++* trained with strong supervision from the entire training data. I.e., up to a small difference of ≈ 3.5 % regarding the avgRMSE(cm), the ranking information can compensate for the five times larger training set of *Regression++*. We further display the distribution of the water level predictions from our multi-task ranking approach on the test images for our *Reg + Rank* approach (Fig. 2). In general, *Reg + Rank* tends to overestimate low water levels and underestimate very high water levels. We point out that high water levels are in general underrepresented in the data, as people are less likely to capture and upload images in such extreme circumstances. E.g., for the very high water level *level9* we have only a single image in our test set. We qualitatively illustrate water level predictions of all four tested models for an example test image in Tab. 2.

Regression, 6.2
Regression++, 9.0
Classification [2], 6.7
Reg + Rank, 8.3

Tab. 2: Example of water level predictions on test image for all four approaches. On the left column, we see the test image with ground-truth water level 8 value and on the right column, we have the methods used in the experiments with their predictions.

5. Conclusions

We have proposed a fully automated method for water level estimation in social media images of flood events. The main idea of our approach is that it is much easier for a human annotator to decide in which of two images the water level is higher, rather than assign an absolute water level to a single image, let alone segment pixel-accurate object instance labels. We implement pairwise ranking as a form of weak supervision that regularises the training of the regression task.

The experimental comparison with a lower and upper performance bound for regression and an alternative classification scheme shows that the proposed weakly supervised method *(Reg + Rank)* is able to perform almost as well as fully supervised regression with a much larger training set *(Regression++)*. Moreover, *Reg + Rank* also outperforms *Classification* [2], although the necessary training data is, arguably, much easier to obtain for *Reg + Rank*. Weak supervision via pairwise ranking thus provides a promising alternative to costly and time-consuming, fine-grained labelling. We hope that our approach can help to overcome the label scarcity problem not only for water level prediction, but for many other regression tasks in the environmental and geo-sciences where collecting a sufficient amount of accurate labels is very laborious, and large datasets as needed for training deep learning are rare. More results and more detailed explanations of this research project can be found in [8].

Experiments	*avgRMSE [cm]*	*stdDev [cm]*	*avgRMSE [level]*	*stdDev [level]*
Regression	14.4	0.45	0.78	0.01
Regression++	10.9	0.85	0.61	0.05
Classification [2]	13.6	0.70	0.80	0.03
Reg + Rank	11.3	0.64	0.62	0.03

Tab. 1: Quantitative results of experiments on the *DeepFlood* dataset.

Literature:

[1] P. Wallemacq, R. Below, D. McClean, The human cost of weather related disasters (1995-2015). Available at https://www.unisdr.org/we/inform/publications/46796 (accessed 4 Aug. 2020).

[2] Chaudhary, P., D'Aronco, S., Moy de Vitry, M., Leitão, J. P., Wegner, J. D., 2019. Flood-water level estimation from social media images. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W5, 5–12.

[3] Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

[4] Neuhold G., Ollmann T., Bulo S. R., Kontschieder P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. The IEEE International Conference on Computer Vision (ICCV). Available at https://research.mapillary.com/img/publications/ICCV17a.pdf (accessed 4 Aug. 2020).

[5] He K., Zhang X., Ren S., Sun J., 2016. Deep residual learning for image recognition. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Available at https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf (accessed 4 Aug. 2020).

[6] Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S., 2017. Feature pyramid networks for object detection. The IEEE Conference on Computer Vision and Pattern

Recognition (CVPR). Available at https://openaccess.thecvf.com/content_cvpr_2017/papers/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.pdf (accessed 4 Aug. 2020).

[7] He K., Gkioxari G., Dollar P., Girshick R., 2017. Mask R-CNN. The IEEE International Conference on Computer Vision (ICCV). Available at https://openaccess.thecvf.com/content_ICCV_2017/papers/He_Mask_R-CNN_ICCV_2017_paper.pdf (accessed 4 Aug. 2020).

[8] Chaudhary, P., D'Aronco, S., M., Leitão, Schindler, K., J. P., Wegner, J. D., 2020. Water level prediction from social media images with a multi-task ranking approach. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 167, pp. 252-262. Available at https://arxiv.org/abs/2007.06749 (accessed 4 Aug. 2020).

Priyanka Chaudhary
Stefano D'Aronco
Konrad Schindler
Jan Dirk Wegner
EcoVision Lab, Photogrammetry and Remote Sensing group
Institute of Geodesy and Photogrammetry
ETH Zürich
CH-8093 Zürich
priyanka.chaudhary@geod.baug.ethz.ch
stefano.daronco@geod.baug.ethz.ch
schindler@ethz.ch
jan.wegner@geod.baug.ethz.ch

Joao Paulo Leitao
Department Urban Water Management
Eawag Swiss Federal Institute of Aquatic Science and Technology
Überlandstrasse 133
CH-8600 Dübendorf
joaopaulo.leitao@eawag.ch