FLOODNET CLASSIFICATION - POST FLOOD SCENE UNDERSTANDING

Arjun Arora

Technische Universität Braunschweig

ABSTRACT

Visual scene understanding is an important task in Disaster Management response. Computer vision is being deployed more and more to help automate this task. Image datasets are used from past floods to understand and plan out the response for future calamities. Unmanned Aerial Vehicles(UAV) can effortlessly access these difficult situations and gather cheap and high-quality data. Floodnet is one such dataset gathered after Hurricane Harvey using UAVs. It contains post-flood scene images from both flooded and non-flooded regions. We can use this labeled data to perform multiple tasks like classification, segmentation and visual question answering. In this paper we will analyse the classification task and train it using transfer learning for efficient and quick classifications using Floodnet dataset.

of areas as flooded or non-flooded which is the most high-level task required in disaster response. Using this model we can dispatch help in required areas quickly. In this pa

Index Terms— Floodnet, classification, Resnet, Transfer learning, Semantic segmentation.

1. INTRODUCTION

The frequency and severity of natural disasters has increased manifolds in recent years due to climate change. Our response to these calamities has to adapt as well for timely response and saving lives. For a post flood response, timely and accurately understanding the scene is the most important task. Computer vision is the field of applying machine learning tasks on visual data like images and videos. Tasks like classification and segmentation are fitting in situations like this, as they can help make quick and better decisions.

Applying computer vision algorithms has some challenges though. Like we need a large amount of annotated data, which in real-world is almost always not available in sufficient amount. Labelling data is an expensive human task. Another challenge is, the available datasets cater not to post disaster understanding but other specific tasks for which they were made. So the dataset is quite context

Figure 1. Floodnet dataset overview.

specific as well. Also, there is generally a class imbalance in the data available generally. Like in our task, we have classifications between Flooded and Non-flooded data, and when we will check the labeled data we will have only a few of the images from the Flooded sample as seen in the real-world.

So coupled with these challenges, the scope of computer vision was limited in disaster management. Floodnet dataset aims at solving some of these problems. The most important factor being that it is designed specifically for post disaster understanding from a real world disaster(Hurricane Harvey). So the data represents practical approach. Also the data is annotated for several tasks using masks. Classification wise it is annotated in our two classes. For segmentation, the training samples are annotated pixel wise into 10 classes. We also have data available for VQA tasks to perform after segmentation to better understand a situation using questions.

Our approach to this paper aims at practically achieving classification on a local system using approaches like transfer learning. We discuss about the tasks and related works in the next section. Then we will discuss our pipeline for classification and what makes it efficient in 3. In 4 we describe our experimental setup of the classification task. Finally in 5 we summarise the results including conclusion and future works.

2. RELATED WORKS

Image classification: Classification is a fundamental task that comprehends an entire image to as a whole. The goal is to classify the image to one of the given specific label. It can either be Single-label classification or Multi-label classification. In single-label classification, one annotation is present for each image. Therefore, the model outputs a single value or prediction for each image. We can use a softmax or sigmoid as activation function. On the other hand, in multi-label classification, an image can contain more than one label. This problem is more complex than single-label classification because the model needs to train and predict probabilities of all the labels in each image. This is generally used in identifying diseases(more than one) from X-rays.

In deep learning. Several architectures and methods are have been developed for classification using extensive datasets like ImageNet. We will use ResNet in our implementation.

ImageNet is a dataset of over 15 million labeled high resolution images with around 22000 categories. In a plain network with fully connected layers, we have the problem of vanishing/exploding gradients. During back-propagation, partial derivatives of the error function with respect to the current weight are multiplied n of these small/large numbers to compute gradients. When n of these small numbers are multiplied, they tend to become zero. When they are large, the gradients explode. So training a deep plain network is problematic.

ResNet introduces skip/shortcut connections to add the input to the output after a few weight iterations. So what the weight learn is a kind of residual mapping. So even if we have a vanishing gradient, we still have the input available to transfer it back to original. ResNet has a VGG-19, some more conv layers and 34-layer residual network with addition of skip connections.

Semantic segmentation: Semantic segmentation is a task that aims at labelling each pixel of an image with a corresponding class of what is being presented. It is also called dense prediction because we are predicting each pixel. We are not separating instances of the same class though, so each object of the same class will be identified as the same object by the classifier. The task of identifying the objects of the same class if called Instance Segmentation. Segmentation models are quite helpful in scenarios like: Autonomous vehicles, medical image diagnostics, scene understanding, which is our task at hand. With classification we answered a high level question that if the image of an area is flooded or not. But we also need to go deeper that that. We can do that with semantic segmentation which can tell us if the pixels in our image can be predicted as a building is flooded, a road is flooded etc by looking at parts of the objects. As we can observe, it is a much more complicated task than just classification. This requires data that is labeled pixel-wise through masks corresponding to each image. Some models used for semantic segmentation are PSPNet, DeepLab, RescueNet etc. All these models address the semantic segmentation of specific objects like rivers, buildings, and roads but not at the whole scene post-disaster.

Visual Question Answering : This task aims at correctly answering questions asked to a machine in a natural language regarding an image input. The aim is to build a model that can answer questions related to an image like a human could. This task is even more challenging than segmentation as this is a computer vision task coupled with natural language integration. Implementing VQA models in distaster response would be beneficial as responders can write simple English sentences and the models can automatically provide answers corresponding to the state of the area. Like a responder could ask if the building in the image is flooded or not and the classifier can automatically give an answer.

To find the right answers however, VQA systems need to be model questions and answers together. We can build and test VQA on Floodnet we can construct baseline models like Stacked Attention network, and MFB with co-attention network.

3. CLASSIFICATION

The dataset has 2343 images split in two tracks. Track 1 follows classification and segmentation, and track 2 is VQA. We follow data from Track 1, and specifically train labelled images(398) and not considering the unlabelled data(1047).

The image dimensions are roughly 3000 x 4000 x 3. The labels for classification tasks are Flooded and Non-flooded.

3.1 Data and preprocessing

Out of the 398 images, we have 51 Flooded samples and 347 Non-Flooded samples. As we can see there is a large class imbalance in our dataset, so we need to use a weighted sampling strategy while loading.

The other problem is when we work with images of this high resolution, while performing the actual training, the systems may run out of memory because the pixels 3000x4000 multiplied with the large number of conv layers and flattening will make millions of parameters and running on multiple epochs is a problem. So the next task in our data preparation pipeline is reshaping the data to 300 x 400, so we can actually do the training in our systems without running out of memory.

Also we will augment the labeled dataset to increase the number of samples as much as we can for better training. The image samples were randomly resized, shifted, cropped, and flipped in both axes.

3.2 Methodology

After preprocessing of data, we have a training set, a validation set and we have set aside 10% of the data for testing the model at the end. The model chosen is ResNet50 for binary classification with our added input layer and the final layer with sigmoid activation function. We chose sigmoid activation because we are working in binary classification and we just want the model to tell which class is more probable for each image without explicit probabilities.
Model compilation parameters: Adam optimiser, Binary cross-entropy and accuracy as metric. Adam is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments.
Binary cross-entropy is the loss function for our binary classification to flooded or non-flooded.
Classification accuracy is the perfect metric to measure how many of the test samples were correctly classified by our model.

3.3 Experiments

ResNet50 was chosen because its computably efficient. While training with the experimental setup described above, we started getting really well accuracies in a few epochs due to resnet's pre-trained architecture and reshaped image samples. After about 20 epochs, we get the following results:

	Accuracy	Loss
Train	95.90%	0.10
Validation	97.44%	0.15

Table 1: Training results

4. RESULTS

Our system performed really well - computationally and accurately. The model got the accuracies in table 1 for training and validation samples.
Test accuracy for the remaining 10% samples turned out to be around 80% and we got 87% F1-score.

5. DISCUSSION AND CONCLUSION

In this research we use Floodnet dataset to help understand post disaster scenes. After a natural flood, we want to prioritise and plan the response. Computer vision tasks work well in these settings and can be used to build a model that classifies an area as flooded or non-flooded.
We use the available annotated training data, split it into train, test and validation. Reshape it into smaller sizes and augment it to get maximum number of images. Then we modify and use ResNet50 model according to our dataset. ResNet50 provides us a sufficiently accurate model and we use that to predict labels of the training data and compare it to the original labels.
This model is efficient and accurate and can be practically used in real-life scenarios confidently.

We can further the research with segmentation and VQA tasks. Also we can deploy semi-supervised learning to use the large amount of unlabelled data available.

6. REFERENCES

[1] Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., & Murphy, R. (2021). FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. IEEE Access, 9, 89644-89654.

[2] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[3] Post-Disaster Segmentation Using FloodNet.

[4] Gupta, Rohit and Mubarak Shah. "RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery." 2020 25th International Conference on Pattern Recognition (ICPR) (2021): 4405-4411.

[5] He, Kaiming et al. “Deep Residual Learning for Image Recognition.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.

[6] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660.