Performance analysis of a U-Net landslide detection model

ITAHISA GONZÁLEZ ÁLVAREZ1, KATHRYN LEEING1, ALESSANDRO NOVELLINO1, SOPHIE TAYLOR1
1 British Geological Survey

INTRODUCTION

We use satellite images with labelled landslide maps from known events to train a machine learning algorithm to automatically identify areas where landslides have taken place. These maps are then used to evaluate the results. One of the main training sets is a dataset of aerial images that is publicly available. These images have been shown to be of particular interest when working with small training datasets, especially when convolutional data augmentation techniques are not used.

Here, we examine differences in the performance obtained with the ability of this algorithm to correctly predict the presence of landslides in satellite images of our algorithm as a result of varying inputs.

U-Net

U-Net is a type of deep convolutional neural network (CNN) with skip connections (Ronneberger et al., 2015). They are of particular interest due to their short training times and their ability to retain small details in larger training classes. Data augmentation techniques can further help the algorithm from the training and validation phases. The network consists of a contracting path with the usual pyramidal structure (Szegedy et al., 2015) and an expansion path. The number of hidden channels doubles or quadruples as we move up vertically in the network, and the number of channels in each expansion path is equal to that of one vertical expansion path. A result, each level of the network, which is not mentioned, is made up of convolutional layers followed by additional operations, which in this case are dropout, batch normalization, and activation functions. The design of the U-Net scheme for generation of segmentation networks can be found in the literature (Chen et al., 2015).

DATA

We collected satellite images from two locations in Fig. 2, in which landslides were triggered by monsoon rainfall and flood. Over 1000 tiles were randomly sampled from these images. The polygons were then used to create手工truth masks to evaluate training of the algorithm. (Figures 3).

Our input data consists of RGB and satellite land images, as well as additional input images (Fig. 4). These images were then fully augmented to create new inputs, which were completely and almost pair-wise balanced, to reduce the extensive evaluation process, and to prevent the algorithm from learning from neighbouring pixels.

RESULTS

Input selection

Validation of the performance of our U-Net model by combining different types and numbers of layers in our input. This diagram illustrates how we built 16 different runs, each with a different layer and input combination. The number of channels is denoted by the size of the blue box, and the number of channels of each layer is denoted by the text above the box. The arrows denote the different operations.

Algorithm performance evaluation

We quantify the accuracy of the predictions by comparing some of the most well-known metrics for machine learning performance: F1 Score (F1 Score = recall + precision / (recall + precision)), Kappa Coefficient (Kappa Coefficient = (O - E) / (1 - E)), Recall, Precision, and Accuracy (Accuracy = True Positives / (True Positives + False Negatives)).

Predictions

PREDICTIONS

FUTURE WORK

If further improvements are needed:

- Increasing the size of the training dataset
- Increasing the size of the test dataset
- Incorporating more steps in the workflow
- Adding more features to the model
- Enhancing the algorithm to improve prediction accuracy in complex scenarios

www.bgs.ac.uk