DeepVGI: Deep Learning with Volunteered Geographic Information

Jiaoyan Chen
Heidelberg University, Germany
j.chen@uni-heidelberg.de

Alexander Zipf
Heidelberg University, Germany
zipf@uni-heidelberg.de

ABSTRACT

Recently, deep learning has been widely studied to recognize ground objects with satellite imageries. However, finding ground truths especially for developing and rural areas is quite hard and manually labeling a large set of training data is costly. In this work, we propose an ongoing research named DeepVGI which aims at deeply learning from satellite imageries with the supervision of Volunteered Geographic Information (VGI). VGI data from OpenStreetMap (OSM) and a crowdsourcing mobile application named MapSwipe which allows volunteers to label images with buildings or roads for humanitarian aids are utilized. Meanwhile, an active learning framework with deep neural networks is developed by incorporating both VGI data with more complete supervision knowledge. Our experiments show that DeepVGI can achieve high building detection performance for humanitarian mapping in rural African areas.

Keywords

Deep Learning; OpenStreetMap; VGI

1. INTRODUCTION

Recently, machine learning especially deep learning algorithms like Convolutional Neural Networks (CNNs) are widely applied in satellite image classification for building detection, scene understanding, etc. However, such learning methods usually rely on a large set of labeled samples for supervision and semi-supervision. Finding training samples or manually labeling images for a specific domain like humanitarian mapping is costly.

With the development of World Wide Web and crowdsourcing, Volunteered Geographic Information (VGI) which is a special case of the larger Web phenomenon known as user-generated content starts to harvest big geographic data provided voluntarily by individuals [1]. VGI platforms like OpenStreetMap (OSM) provide a way for free labels for satellite image classification. Mnih et al. [2] proposed to extract vector data from OSM for supervised learning with deep networks. The study defined missing error (i.e., cases when an object that appears in an satellite image does not appear in the map) and registration error (i.e., cases when the location of an object in the map is inaccurate) of the map labels and developed two loss functions for training to reduce the negative effect of such noise.

In humanitarian mapping where prediction targets are often located in rural or undeveloped areas (e.g., cottage in Africa) and not labeled on OSM (cf. (3) in Figure 1), the missing error becomes so large that only a part of the domain's supervision knowledge is covered by OSM. In this work, we propose an ongoing study named DeepVGI which aims at deeply learning from VGI and satellite images for humanitarian mapping. It incorporates free satellite image labels from both OSM and MapSwipe (cf. (1) in Figure 1) with an active deep learning framework based on deep neural networks and a cost sensitive active sampling strategy which enriches the supervision knowledge for the domain of building detection.

Figure 1: (1) Interface of MapSwipe application where volunteers will click the images that contain buildings or roads for humanitarian aids and (2)(3) two level-18 satellite images used in MapSwipe

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054250

2. PROBLEM AND FRAMEWORK

In MapSwipe, the humanitarian mapping task is to judge whether a size-fixed satellite image (256*256) has target objects (e.g., building). With DeepVGI, this task is implemented by first sliding a window (e.g., 32*32) over an image and then classifying the image tiles generated by the sliding window. Therefore, the technical problem of DeepVGI includes (1) predicting the label y of each small tile $\boldsymbol{x}$ and (2) deciding the label l of a MapSwipe image $\boldsymbol{s}$.

The workflow of DeepVGI which includes four steps is shown in Figure 2. The first step (S1) trains a multilayer artificial neural network (ANN-S1) with (1) positive tiles $(\boldsymbol{x}, y_+)$ whose centers are determined by the geographic locations of the buildings on OSM and (2) negative tiles $(\boldsymbol{x}, y_-)$

that are extracted from empty MapSwipe images. The second step (S2) includes an active sampling strategy which (1) slides a window over each training MapSwipe image (2) predicts the label of each tile using ANN-S1 with a probability threshold α and determines whether the image has any buildings, (3) divides the training images into MapSwipe-consistent ($\boldsymbol{s}^C$) and MapSwipe-inconsistent ($\boldsymbol{s}^I$) by comparing the predicted label with the MapSwipe volunteers' label. The third step (S3) manually labels a limited number of positive tiles ($\boldsymbol{x}^I, y_+$) randomly selected from the above MapSwipe-inconsistent images with type II error. The fouth step (S4) integrates the new training tiles in S3 with the old training tiles in S1 and re-trains a multilayer artificial neural network (ANN-S4).

In the above workflow of DeepVGI, we call the procedure of resampling as VGI-based active learning strategy, where the MapSwiple volunteers' labels are used to guide the selection of candidate images for manual labeling. The positive supervision knowledge from OSM are compared against those from MapSwipe. Those training MapSwipe images whose supervision knowledge are consistent with OSM are denoted as $\boldsymbol{s}^C$, while those that are inconsistent are denoted as $\boldsymbol{s}^I$ where new positive tiles are sampled to bridge the supervision knowledge gap between OSM labels and the whole humanitarian mapping domain.

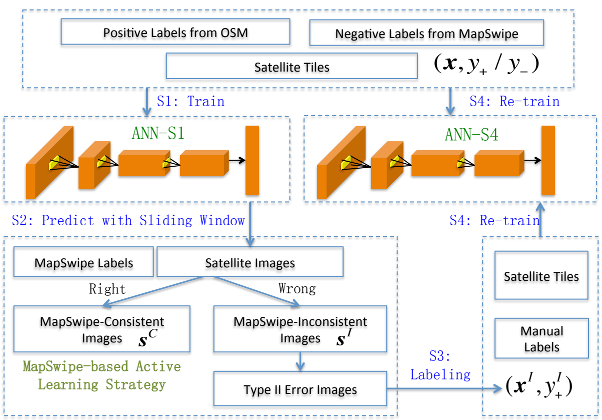
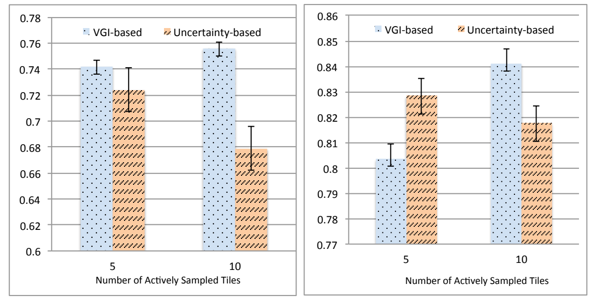


Figure 2: The Workflow of DeepVGI

3. EVALUATION AND DISCUSSION

In the evaluation of this poster, 1910 MapSwipe remote sensing images from Southern Malawi Districts in Africa are used. They are randomly divided into 1590 training images and 320 testing images. Two classic CNNs named LeNet and AlexNet as well as a multilayer perception are tested. We (1) analyze the performance of the VGI-based active learning strategy in comparison with a typical uncertainty-based strategy which selects the most uncertain samples for labeling, and (2) compare the overall testing performance of DeepVGI with Deep-OSM [1] and MapSwipe volunteers. All the used codes, data and a tutorial are open[2].

Figure 3 presents the average testing performance of the VGI-based active learning strategy and the uncertainty-based active learning strategy, where multiple batches of active samples (with size of 5 or 10) are randomly selected for testing. We can find that the VGI-based strategy achieves the best performance with 10 actively sampled tiles, and outperforms the uncertainty-based strategy in most cases.



(a) F1 Score (b) Accuracy

Figure 3: Testing performance using VGI-based active learning strategy and uncertainty-based active learning strategy

Table 1 shows the overall testing performance of DeepVGI, Deep-OSM and MapSwipe, where we can find DeepVGI's F1 score and accuracy are (1) significantly (p-value $\ll 0.05$) larger than Deep-OSM (2) but still smaller than the MapSwipe volunteers. The first result indicates that adding actively sampled positive tiles do brings additional supervision knowledge and increases the generalization performance of the ANN model. The second result means that DeepVGI is still weaker than the volunteers of MapSwipe which each image is voted by three volunteers. This is may be caused by two factors: (1) the noise (e.g., big rocks on bare land) and (2) the small size of images used in this evaluation.

	Precision	Recall	F1 Score	Accuracy
DeepVGI	0.775	0.737	0.756	0.841
Deep-OSM	0.632	0.875	0.734	0.788
MapSwipe	0.738	0.938	0.826	0.868

Table 1: Testing performance of DeepVGI (10 actively sampled tiles), DeepOSM and MapSwipe (volunteers)

To further improve the accuracy, we will on one hand adopt much more training data for higher generalization performance with a scalable implementation on the Spark cluster. On the other hand, refined OSM supervision knowledge (e.g. building contour) as well as other VGI will be utilized with a more carefully designed object function. By the way, we will also implement the deep learning algorithm in [2] to extend the evaluation.

4. ACKNOWLEDGMENTS

This work has been supported by the Klaus Tschira Foundation (KTS) Heidelberg. The authors thank Benjamin Herfort and Melanie Eckle from MapSwipe and Missing Maps.

5. REFERENCES

[1] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[2] V. Mnih and G. E. Hinton. Learning to label aerial images from noisy data. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 567–574, 2012.

[1]Deep-OSM only uses OSM labels for supervision deep learning (i.e., ANN-S1 in Figure 2), which follows the basic idea of https://github.com/trailbehind/DeepOSM and https://github.com/geometalab/OSMDeepOD

[2]https://gitlab.com/giscience/DeepVGI