ZOOMING INTO UNCERTAINTIES: TOWARDS FUSING MULTI ZOOM LEVEL IMAGERY FOR URBAN LAND USE SEGMENTATION

Eike Jens Hoffmann[1], *Mohsin Ali*[2] *and Xiao Xiang Zhu*[1,2]

[1] Data Science in Earth Observation, Technical University of Munich (TUM), Munich, Germany
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

ABSTRACT

Urban land use prediction is an ill-posed problem from a remote sensing perspective. Some areas are easy to predict with aerial images, e.g. residential areas or industrial areas, whereas it is nearly impossible to predict land use in dense urban centers with highly mixed land use.

In this study, we use a fully convolutional, Bayesian neural network for urban land use segmentation that yields predictions and pixel-wise uncertainty values side-by-side. By adding aleatoric uncertainty to the output of our model, we can assess how much the model benefits from the provided data. We train our network using a dataset from four metropolitan areas in the U.S. on two different zoom levels. Our results show that adding aleatoric uncertainty can improve the IoU scores if a sufficient amount of informative data is provided.

Index Terms— Urban Land Use, Semantic Segmentation Model, Uncertainty, Multi-Zoom Level

1. INTRODUCTION

Mapping urban land use is a crucial part for understanding dynamics within a city, e.g. where do citizens live, where do they work, etc. An in-depth knowledge into this geo-spatial information helps to estimate the number of inhabitants as well as to create mobility models.

In developed countries, municipalities have detailed records about land use in their area on a parcel level. However, in emerging countries urban development proceeds at high pace, making this information out-dated in short time, if ever recorded.

Remote sensing can help to close this gap due to a growing number of earth observation satellites providing high-resolution imagery on a spatial as well as on a temporal scale. But assessing urban land use from an aerial perspective is an ill-posed task. Determining the function of an area or a building can be impossible due to unsolvable ambiguities: roofs in city centers can cover buildings of any function. Nevertheless, there are areas that can be clearly predicted using remote sensing data. Figure 1 illustrates where remote sensing data shows clear patterns for urban land use (in blue) and where there are high uncertainties (in red).

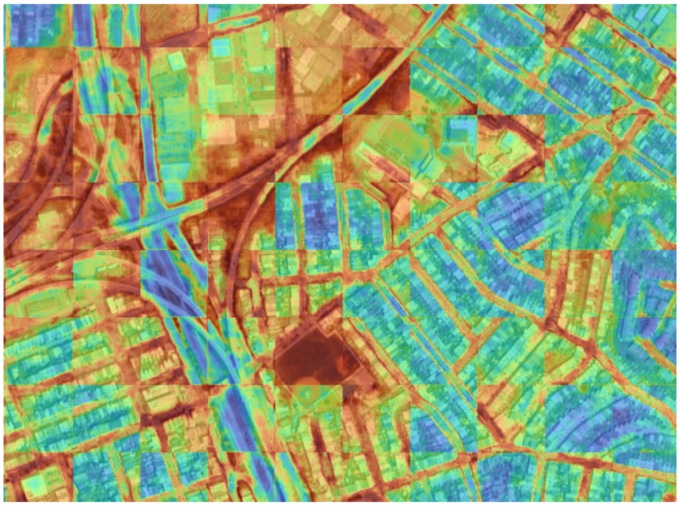

Fig. 1: Uncertainties in land use prediction in San Francisco: Blue areas are predicted with high confidence while red ones have high uncertainties

As a real world problem, areas with an obvious land use can be right next to ambiguous areas with smooth transitions between them. Therefore, we treat this problem as an semantic segmentation task on images instead of a classification task allowing different land use classes in one image. To tackle the problem, we propose a Bayesian deep learning architecture for land use segmentation that yields predictions and corresponding, pixel-wise aleatoric uncertainty scores. This uncertainty "captures noise inherent in the observations and cannot be reduced even if more data were to be collected" [1]. By corrupting the network's output with random noise it makes the network more robust. We show how this models performs on different zoom levels and discuss how aleatoric uncertainty can help to improve the segmentation results.

This work is jointly supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-StG-714087], Acronym: *So2Sat*), Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de), and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg.

Our paper is structured as follows: We start with a brief discussion about urban land use prediction in remote sensing and introduce our model as well as the datasets for training and evaluation. In section 5, results, we show how this model performs in our study areas and discuss its features. Finally, we summarize our findings and give an outlook on further developments.

2. RELATED WORK

Generating a high-quality dataset for land use classification is challenging because image patches must show homogeneous areas without label noise from different classes in their vicinity. The UC Merced dataset [2] is one of the most popular benchmark datasets for land use classification since its labels are manually curated and of very high quality. Deep learning models achived big success on this dataset gaining accuracies of over 95% [3, 4].

Models trained on classification datasets with homogeneous samples achieve high accuracies when applied to image patches of the same data distribution, but lack generalization if patches show parts of multiple classes. Therefore, recent approaches aiming at high generalization decompose satellite images into structurally similar regions and predict multiple samples from each of them [5]. Alternatively, object-based classification algorithms can be used [6].

Instead of treating the problem as a classification task, segmentation models predict a class for each pixel of an image patch individually and are able to process heterogeneous areas. Segmentation models can work on data from different decades to predict agricultural land use [7] or include multiple remote sensing products from SAR and optical data [8].

3. METHOD

Our model is based on a combination of techniques that showed success in a Kaggle segmentation challenge on aerial images[1]. At the core it is a U-Net style architecture [9] with four down and four up convolution blocks with batch normalization and ELU activation functions [10] for feature extraction. These blocks are pair-wise connected with skip connections for propagating spatial information.

For calculating uncertainties, we train the network with a modified loss function that captures aleatoric uncertainty representing the inherent noise in the data, e.g. inconsistencies in ground truth or sensor noise [11]. For image segmentation the network outputs are randomly corrupted N times and integrated by Monte Carlo sampling. For N samples the loss criteria becomes:

$$\mathbb{E}(y') = -\sum_{i,j} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} y_{i,j,k} \cdot log(y'_{n,i,j,k}) \tag{1}$$

where i, j are pixel indices, k is the number of classes, and y'_n are corrupted samples of network output y'. These corrupted samples are created by making the network output extra learned parameter σ. This parameter scales the randomly generated Gaussian noise samples, which are added to network output logits and passed through softmax function. During prediction, same corruption and sampling steps are performed and entropy of softmax outputs is taken as the uncertainty measure.

We train this model using Adam [12] with a learning rate of 10e-4 and a batch size of 16 in combination with early stopping monitoring the validation loss. Our model is trained a maximum of 25 epochs.

4. DATASET

Our dataset consists of Google Maps tiles and cadastral data from four metropolitan areas in the United States: Los Angeles (LA), New York City (NYC), San Francisco (SF), and Washington D.C (WDC). Their municipalities provide land use data on a parcel level with geo-spatial coordinates in a free and open way. In each area a different land use classification scheme is applied for describing land use with different levels of granularity. Therefore, we homogenized the cadastral labels to a three class segmentation scheme: *Residential*, *non-residential*, and *background*. The first two labels are used for built-up areas focusing on buildings, whereas the last one covers roads, parks, green spaces, and other open areas.

After homogenization of our ground truth we obtained the corresponding aerial image tiles for the areas covered from Google Maps at two zoom levels: 16 and 18. If a tile covers only *background*, we omitted it, otherwise we downloaded a 256 x 256 image from Google Maps.

To minimize spatial correlation and avoid spatial overfitting, we split the data from each area using a median data split strategy in latitude and longitude. This divides each area into four parts: North-east, north-west, south-east, and south-west, so that each part covers approximately the same area. The northern parts are used for training, the part in the south-west is for validation, and the south-east is for testing.

5. RESULTS

We evaluate our model on two different zoom levels for each class and study area individually. The results are shown in Table 1 and Table 2 for zoom levels 16 and 18, respectively. To assess the results of a model with aleatoric uncertainty estimation, we compare it with a baseline model of the same architecture but a default cross-entropy loss called baseline. Zoom level 16 has a lower ground sampling distance and shows more spatial context of an area while at zoom level 18 parcels and building instances become distinguishable. Land use classes spanning across large areas can benefit from the

[1] https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection

larger spatial context, while others benefit from a higher level of detail.

Area	Class Method	Background	Non-residentiall	Residential
LA	Baseline	0.307	0.358	0.523
	Aleatoric	0.299	0.298	0.519
NYC	Baseline	0.462	0.098	0.347
	Aleatoric	0.368	0.069	0.314
SF	Baseline	0.388	0.297	0.376
	Aleatoric	0.385	0.287	0.357
WDC	Baseline	0.593	0.029	0.357
	Aleatoric	0.531	0.036	0.356

Table 1: Intersection over Union results at zoom level 16 per study area, method, and class

Area	Class Method	Background	Non-residential	Residential
LA	Baseline	0.488	0.510	0.688
	Aleatoric	0.507	0.535	0.706
NYC	Baseline	0.287	0.427	0.480
	Aleatoric	0.362	0.473	0.531
SF	Baseline	0.298	0.398	0.489
	Aleatoric	0.274	0.410	0.540
WDC	Baseline	0.156	0.295	0.420
	Aleatoric	0.157	0.338	0.445

Table 2: Intersection over Union results at zoom level 18 per study area, method, and class

At zoom level 16, adding aleatoric uncertainty does not yield any improvement compared to a baseline model: In most cases, the IoU values decrease. Moreover, the *background* class shows the highest IoU values compared to *non-residential* and *residential* with *non-residential* performing the worst. Taking a closer look at the study areas, LA shows the highest values on average, whereas NYC and WDC are lower due to their IoU values for the *non-residential* class.

Increasing the zoom level to 18, the findings change: Adding aleatoric uncertainty increases the IoU values in almost every aspect. Each class yields an improvement of up to 10% with one exception: the *background* class in SF. On a class level, the *residential* class shows the highest IoU values, with *non-residential* being second and the background class on the third place. Again, LA shows the highest values on average, but in this case NYC scores second ahead of SF.

Comparing the results on the two zoom levels, built-up areas benefit from the increased resolution. Especially the *non-residential* class shows high gains in NYC and WDC. As a drawback, the *background class* yields lower IoU scores with more details in the imagery. Furthermore, adding aleatoric uncertainty increases the values only at zoom level 18.

To dive deeper into the differences between the two zoom levels we analyze the distribution of uncertainty values in LA for each class in Figure 2 and Figure 3 for zoom levels 16 and 18, respectively.

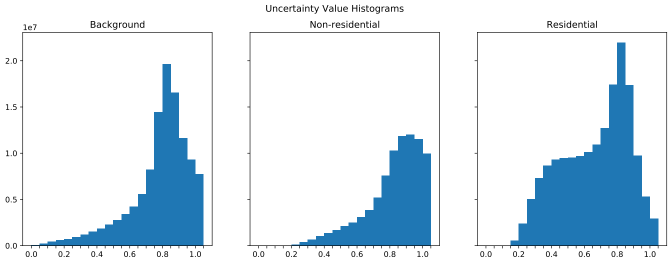

Fig. 2: Uncertainty value histograms at zoom level 16 in Los Angeles for each class

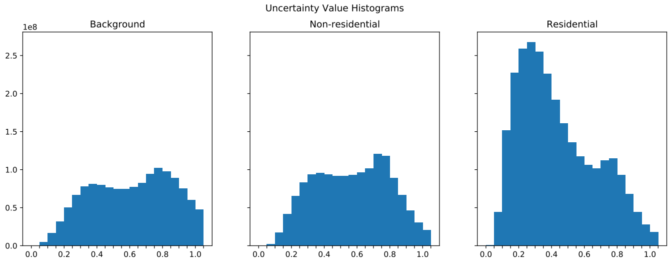

Fig. 3: Uncertainty value histograms at zoom level 18 in Los Angeles for each class

At zoom level 16, all histograms are skewed towards higher values (for each class) indicating that the model is mostly uncertain about its predictions (Figure 2). Increasing the zoom level to 18 leads to a shift, especially for the *residential* class. In this case, the distribution of uncertainty values is skewed towards lower values meaning that the model is mostly very confident about its predictions for *residential* areas (Figure 3).

Both models were trained using early stopping to avoid overfitting and none of them was trained until the maximum number of epochs. So the data available to the model was sufficient for some training, but still there are open questions to the model. There are ambiguities that the model is not able to resolve based on the patterns it has seen so far.

Training the model on the same area with tiles at zoom levels 16 and 18 leads to a multiplication of the data by factor 16 in the number of samples. Nevertheless, we see that the *background* class can be predicted well at zoom level 16 and moreover, yields lower IoU scores at higher resolutions. At zoom level 16 large scale structures become more visible while classes that require a higher resolution are blurred.

Figure 4 shows an example for the output of our model

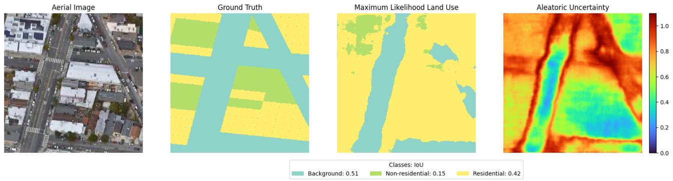

Fig. 4: Example for segmentation of urban land use at zoom level 18 with uncertainty

in a dense urban area with parcels of different land use next to each other. The plot on the very right depicts the aleatoric uncertainty for each pixel with blue as low uncertainty and red as high uncertainty.

The large road is segmented well as *background*, while smaller streets are partially missed and labeled as *residential*. However, the uncertainty plot reveals that these narrow street areas are predicted with low confidence. The *residential* area in the lower right corner with small green spaces in the backyard is correctly predicted at very low uncertainty. We see that *non-residential* areas are often wrongly classified, but with high uncertainty, indicating that the decision was very close.

6. CONCLUSION AND OUTLOOK

In this study, we propose a Bayesian deep neural network to address an ill-posed task in remote sensing. We add aleatoric uncertainty to a segmentation model with a special loss function and use this combination for urban land use segmentation. By evaluating this model on four study areas in the U.S., we show its behaviour on different data and compare it to a baseline model without uncertainty estimation. Adding aleatoric uncertainty to a model makes it more robust towards noise given enough data and provides further details into what has been learned.

In our future work, we plan to integrate this knowledge into models that use multi-scale data in an end-to-end fashion leveraging large-scale overviews and focused views in a combined manner. Uncertainty values can help to identify where higher-resolution data is necessary or helpful to resolve ambiguities. This paves the way to a smart approach of fusing multi-scale imagery.

7. REFERENCES

[1] Alex Kendall and Yarin Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[2] Yi Yang and Shawn Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.

[3] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *arXiv preprint arXiv:1508.00092*, 2015.

[4] Francois PS Luus, Brian P Salmon, Frans Van den Bergh, and Bodhaswar Tikanath Jugpershad Maharaj, "Multiview deep learning for land-use classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 12, pp. 2448–2452, 2015.

[5] Bo Huang, Bei Zhao, and Yimeng Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sensing of Environment*, vol. 214, pp. 73–86, 2018.

[6] Ce Zhang, Isabel Sargent, Xin Pan, Huapeng Li, Andy Gardiner, Jonathon Hare, and Peter M Atkinson, "An object-based convolutional neural network (ocnn) for urban land use classification," *Remote sensing of environment*, vol. 216, pp. 57–70, 2018.

[7] Antoine Richard, Assia Benbihi, Cédric Pradalier, V Perez, Rosalinde van Couwenberghe, and P Durand, "Automated segmentation of land use from overhead imagery," in *International Conference on Precision Agriculture*, 2018.

[8] Mohamed Barakat A Gibril, Mohammed Oludare Idrees, Helmi Zulhaidi Mohd Shafri, and Kouame Yao, "Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data," *Journal of Applied Remote Sensing*, vol. 12, no. 1, pp. 016036, 2018.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[11] Quoc V Le, Alex J Smola, and Stéphane Canu, "Heteroscedastic gaussian process regression," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 489–496.

[12] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.