Technische Universität München

Advances In Uncertainty-Guided Local Climate Zone Classification

Christoph Koller [1,2], Katharina Hechinger [3], Muhammad Shahzad [1], Göran Kauermann [3], Xiao Xiang Zhu [1]

[1]: Data Science in Earth Observation (TUM) , [2]: Remote Sensing Technology Institute (DLR) , [3]: Department of Statistics (LMU)

Uncertainty Quantification

In the machine or deep learning context, we generally distinguish between two types of uncertainties:

- **Aleatoric** or data uncertainty inherent in the data
- **Epistemic** or model uncertainty, which reflects the model's confidence in its prediction

While the aleatoric uncertainty cannot be reduced, because it belongs to the data by nature, epistemic uncertainty can be reduced by finding more suitable machine learning algorithms or by modeling the data in a more clever way.

On a very broad level, we can separate two major streams of uncertainty quantification methods:

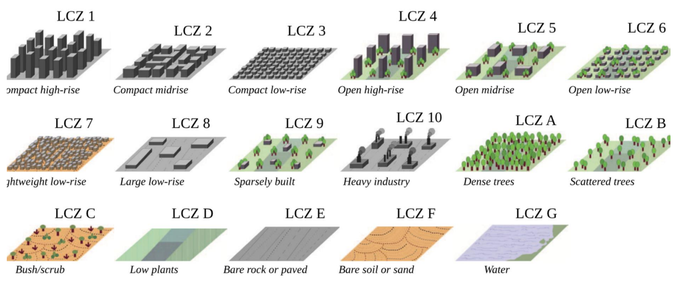
- **Probabilistic** methods including ensembles of networks or machine learners, dropout mechanisms, and Bayesian Neural Networks
- **Deterministic** approaches such as *Prior Neural Networks* (Malinin, Gales 2018), distance-preserving hidden mappings or spectral-normalized networks

Why modeling uncertainties? Ideally, **a high predictive uncertainty** indicates **lack of knowledge or confidence** in the prediction. This can hint at an *out-of-distribution* or an expressive new data point. The latter can be a promising candidate which could be added to the initial data set and help the model to generalize better to unseen data.

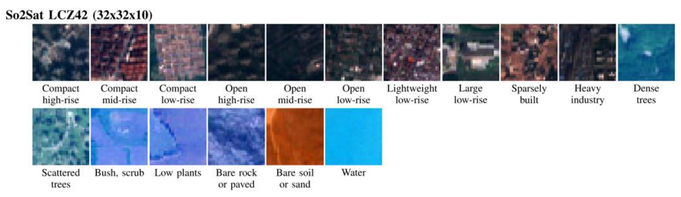
Local Climate Zones (LCZs)

The 17-class scheme of Local Climate Zones (LCZs) as defined by (Stewart, Oke 2011) consists of 10 urban and 7 vegetation zones and describes urban conglomerates. Early works using LCZ maps have focussed on:

- Urban Heat Island (UHI) detection
- Urban planning
- Climatological applications



LCZ Classification Scheme (Samsonov and Trigub, 2017)



LCZ example images (Gawlikowski et al., 2022)

Human Label Uncertainty

For the data set *So2Sat LCZ42* (Zhu et al. 2020), we study a subset which was created to measure **label confidence** and has J=10 label votes for each image.

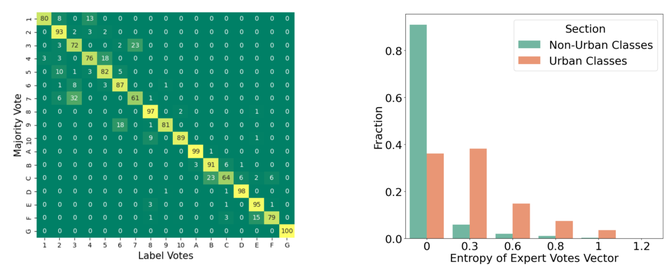
From the individual votes $V_1, \dots, V_J$ belonging to classes $k = 1, \dots, K$, we can formulate **vote counts**

$$Y_k^{(i)} = \sum_j 1_{\{V_j^{(i)}=k\}}$$

These are combined to a label with distributional form:

$$y_{distr}^{(i)} = \boldsymbol{Y}^{(i)}/M$$

The studied subset covers 10 European cities. Below, we look at the inherent **human label uncertainty:**

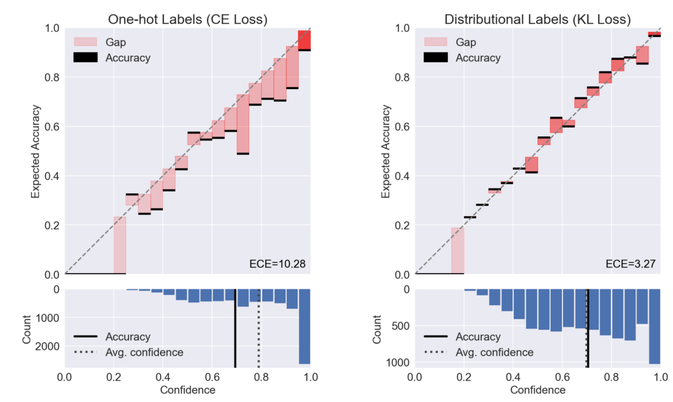


(Left) Confusion matrix of individual label votes; (Right) Entropies of the resulting voting distributions.

Eventually, we implement these new labels into a supervised deep learning architecture by using the following Kullback-Leibler (KL-) divergence loss:

$$-\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K} y_{distr,k}^{(i)} \cdot \log \frac{y_{distr,k}^{(i)}}{p_\theta(y^{(i)}=k|x^{(i)})}$$

We show the results of the comparison between the regular model (*one-hot* labels based on majority vote, cross-entropy) loss and the distributional approach outlined above:



(a) One-Hot Encoding (b) Label Distribution Encoding

Reliability Diagrams of the Modeling Approaches with and without including Human Uncertainty, visualized using (Hollemans 2020)

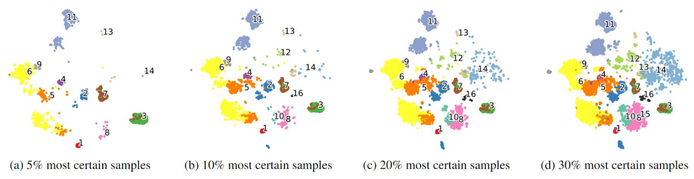
	CE One-hot ↓	CE Distr. ↓	ECE ↓	MCE ↓	SCE ↓
One-hot	1.12 ± 0.05	1.38 ± 0.07	9.79 ± 3.18	23.14 ± 3.97	$\mathbf{1.03 \pm 0.50}$
+ LS	1.05 ± 0.01	1.23 ± 0.03	7.33 ± 2.62	19.80 ± 4.82	1.11 ± 0.25
+ TS	1.00 ± 0.13	1.17 ± 0.07	4.15 ± 2.37	15.88 ± 10.60	1.44 ± 0.22
+ LS & TS	1.02 ± 0.03	1.18 ± 0.02	$\mathbf{3.21 \pm 0.96}$	$\mathbf{11.30 \pm 4.26}$	1.32 ± 0.03
+ MC-Drop	1.12 ± 0.05	1.37 ± 0.06	9.57 ± 3.11	23.10 ± 4.22	1.04 ± 0.49
+ LS & MC-Drop	1.05 ± 0.01	1.23 ± 0.03	7.11 ± 2.41	33.22 ± 26.60	1.12 ± 0.24
Distr.	1.06 ± 0.07	1.21 ± 0.07	5.80 ± 1.07	15.57 ± 4.22	1.21 ± 0.20
+ LS	0.98 ± 0.03	1.08 ± 0.02	8.31 ± 2.46	17.32 ± 4.84	1.73 ± 0.06
+ TS	0.96 ± 0.09	1.07 ± 0.07	5.89 ± 2.50	15.37 ± 3.94	1.72 ± 0.15
+ LS & TS	$\mathbf{0.95 \pm 0.04}$	$\mathbf{1.05 \pm 0.04}$	4.21 ± 1.38	15.35 ± 1.95	1.58 ± 0.10

Table 1: Cross-Entropies between predicted softmax probabilities and labels on the test set as well as calibration errors, averaged over five runs. CE = Cross entropy, LS = Label smoothing, TS = Temperature Scaling, MC-Drop = Monte Carlo Dropout. Binning was performed using 20 equally-sized bins.

For more information, check out our preprint on arxiv (Koller, Kauermann, Zhu 2022) !

Feature Space Uncertainty

Neural networks learn *implicit* **high-dimensional representations** of the input data. A simple uncertainty measure was introduced in our recent work (Koller, Shahzad, Zhu 2022), where we investigate the distance of points in the feature space to their respective **class focal points**. Visual and quantitative results are displayed below:

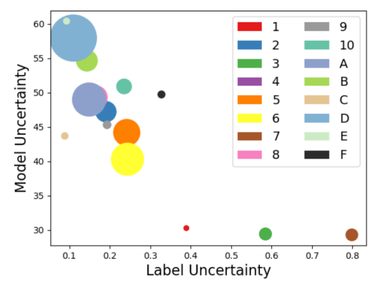


(a) 5% most certain samples (b) 10% most certain samples (c) 20% most certain samples (d) 30% most certain samples

Fig. 1: TSNE visualization of the 2048d deep features learned by a ResNet50. The model was pretrained on ImageNet and finetuned on LCZ42 with distributional labels. The distances of the data points to their respective geometric class medians were sorted and only the subset samples are shown in the TSNE plots.

	OA	AA	κ	CE
R / UT / UTS (5 %)	**66.2** / 60.1 / 58.5	**31.3** / 20.9 / 23.6	**56.7** / 46.1 / 47.9	**1.39** / 1.61 / 1.44
R / UT / UTS (10 %)	62.9 / **63.5** / 61.2	**27.7** / 24.0 / 27.0	**52.9** / 51.6 / 51.8	1.37 / 1.47 / **1.33**
R / UT / UTS (15 %)	**65.8** / 64.9 / 61.4	**31.1** / 24.4 / 26.8	**56.3** / 53.9 / 51.7	**1.25** / 1.35 / 1.31
R / UT / UTS (20 %)	**67.8** / 66.0 / 65.5	**32.4** / 26.1 / 31.7	**58.9** / 55.7 / 56.9	**1.20** / 1.31 / 1.26
R / UT / UTS (30 %)	68.0 / 65.4 / **68.0**	34.3 / 27.3 / **34.8**	59.2 / 55.0 / **59.7**	1.20 / 1.21 / **1.17**

Table 1: Performance metrics on hold-out test set (R = random subsets, UT = uncertainty threshold & UTS = stratified uncertainty threshold, both based on most uncertain samples). The metrics are overall accuracy (OA), average accuracy (AA), Kappa score (κ), and cross-entropy (CE), shown as averages over 3 independent runs. Further performance metrics showed similar trends. The results validate the hypothesis that the most "*uncertain*" samples extracted by the proposed uncertainty-guided representation learning approach induce more diversity during training, and hence makes the trained model more generic.



Conclusion & Outlook

- We have observed strong **generalization** and **calibration** performance when including human uncertainty
- Uncertainty quantification in the **feature space** is very promising, good results already for simple approach with single-class uncertainties

Outlook:

- More suitable loss functions and distance measures
- Link model uncertainty with human label uncertainty
- Use Bayesian reasoning to learn parameters of continuous distributions

References

Gawlikowski, J., Saha, S., Kruspe, A., & Zhu, X. X. (2022). An advanced dirichlet prior network for out-of-distribution detection in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *60*.

Hollemans, M., 2020. Reliability diagrams. https://github.com/hollance/reliability-diagrams.

Koller, C., Kauermann, G., & Zhu, X. X. (2022). Going Beyond One-Hot Encoding in Classification: Can Human Uncertainty Improve Model Performance?. *arXiv preprint arXiv:2205.15265*.

Koller, C., Shahrad, M., & Zhu, X. X. (2022). Uncertainty-Guided Representation Learning in Local Climate Zone Classification. In *IEEE International Geoscience and Remote Sensing Symposium* 2022.

Malinin, A., & Gales, M. (2018). Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, *31*.

Samsonov, T., & Trigub, K. (2017). Towards computation of urban local climate zones (LCZ) from openstreetmap data. In *Proceedings of the 14th International Conference on GeoComputation*.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, *93*(12).

Zhu, X. X., Hu, J., Qiu, C., Shi, Y., Kang, J., Mou, L., ... & Wang, Y. (2020). So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [Software and Data Sets]. *IEEE Geoscience and Remote Sensing Magazine*, *8*(3).

Technische Universität München
TUM School of Engineering and Design
Department of Aerospace and Geodesy

This present contribution is supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS".