ON THE FUSION STRATEGIES OF SENTINEL-1 AND SENTINEL-2 DATA FOR LOCAL CLIMATE ZONE CLASSIFICATION

Jakob Gawlikowski[1], *Michael Schmitt*[2], *Anna Kruspe*[1], *Xiao Xiang Zhu*[2,3]

[1]Institute of Data Science, German Aerospace Center (DLR), Jena, Germany
[2]Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany
[3]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany

ABSTRACT

Local Climate Zone (LCZ) classification is the most commonly used scheme to analyze how local urban morphology affects the climate of local areas. Classification methods are often based on remote sensing data or on a fusion of several data sources. In this study, the effects of different fusion strategies of optical and synthetic aperture radar (SAR) data on the accuracy of LCZ classifications are investigated. The data processing is implemented with a convolutional neural network (CNN), where until a fusion layer, separate data sources are processed separately on branches. Strategies of splitting the data into branches and the effects of different fusion stages are compared, together with approaches based on sums of independent classifiers. For our setting, the stage of fusion does not seem to have a big influence on the accuracy. The results of this study contribute to a better understanding of cooperative usage of multispectral and SAR data.

Index Terms— Local Climate Zone Classification, Data Fusion, Fusion Network

1. INTRODUCTION

The Local Climate Zone (LCZ) classification scheme was designed to describe the physical nature of cities independent from cultural and regional differences in the descriptions [1]. It distinguishes regions based on the type of land cover (e.g. urban, vegetation, water) and their structures (e.g. height and density). The scheme consists of 17 classes, which can be seen in Figure 1. Due to its globally valid definition, the LCZ classification can be used to generate a meaningful global map, which is an important contribution in a variety of study areas, including climate change and urban development. To achieve this, the classification process needs to be automated. The challenge here lies in creating models that deliver reliable classification results even within completely unknown regions. The introduction of convolutional neural networks (CNN) led to rapidly increasing accuracies in image classification tasks. Today, CNNs and net structures based on CNNs are widely used within the remote sensing community [2] and have delivered promising results on Local Climate Zone classification tasks [3, 4, 5, 6]. Due to different data sources, different data sets, and different strategies of choosing training and validation data, it is hard to directly compare the works with each other. However, all results show that approaches based on convolutional neural networks are very promising. Former works differ not only by the used classification methods and net structures, but also by the considered data sources. While some authors work with multispectral images from Sentinel-2 or Landsat 8 [4, 5], others evaluated the usage of SAR data for this task [7]. In order to benefit from the advantages of different data sources, several data sources can be used in combination within a *data fusion* approach. Data fusion not only combines the benefits from different sources, but also mitigates their disadvantages [8]. With its independence from weather conditions and atmospheric distortions, SAR data can deliver useful additional data and complement optical data sources. Fusing optical and SAR data is therefore a promising approach in order to increase the accuracy and stability of local climate zone classification methods. In [6] and [7], CNNs are used to extract and fuse features from Polarimetric SAR (PolSAR) data and

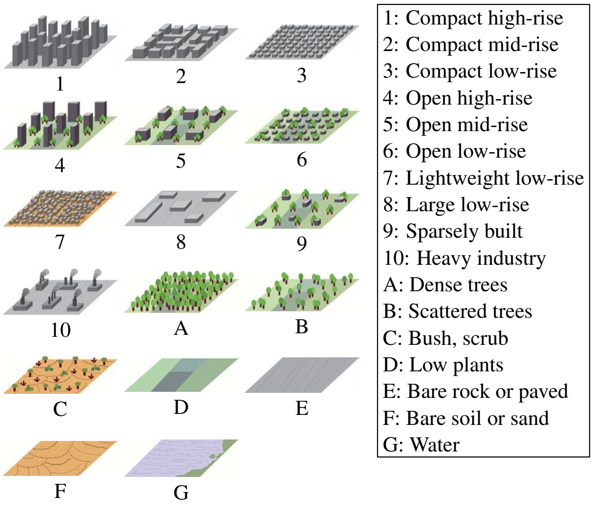

Fig. 1: The 17 local climate zones as defined in [1].

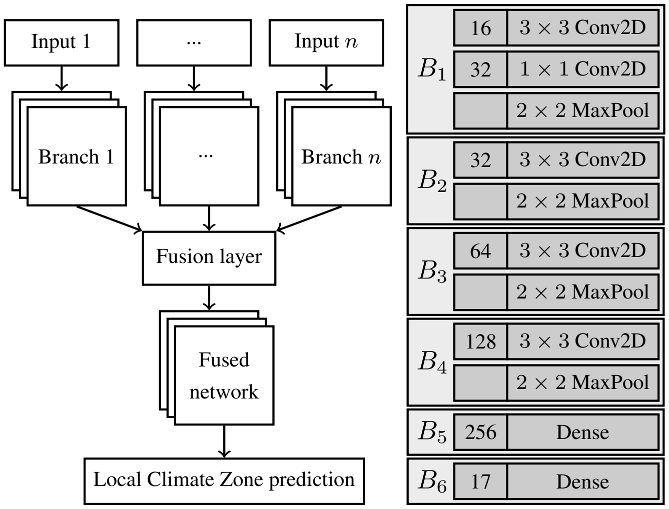

Fig. 2: The considered network structures for the data fusion of n input sets (left) and the basic network structure split into 6 blocks B_1 to B_6 (right).

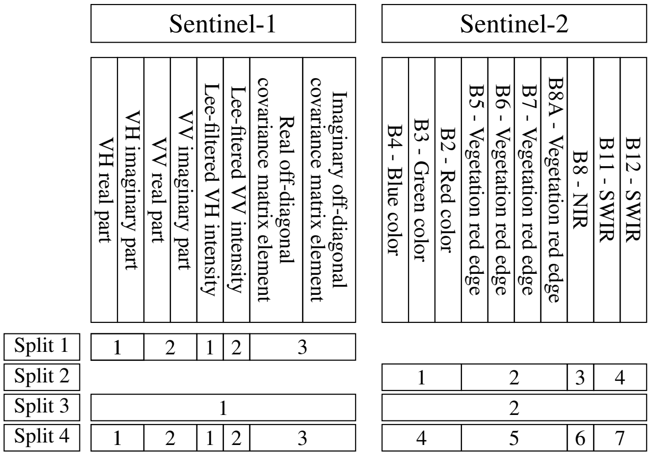

Fig. 3: The four considered possibilities of splitting the data into disjoint subsets as input for the fusion network. The numbers in the blocks represent the split set to which these channels belong to.

optical images in order to classify LCZs and urban scenes. In both studies, it can be seen that fusing optical data with SAR data increases the accuracy in classification tasks compared to working on optical or SAR data alone. Nevertheless, an optimal way of fusing the two very different data sources is not obvious and both approaches fuse the data after a fixed number of layers.

In this study, we investigate the influence of the stage of fusion for optical Sentinel-2 data and Sentinel-1 SAR data and figure out how different stages affect the accuracy of classification tasks on known and on unknown regions. We use relatively simple CNNs with different stages of fusion to combine the feature extraction and the data fusion processes. Furthermore, different types of input splittings are compared.

2. CONVOLUTIONAL FUSION NETWORK

Data fusion is widely used in a variety of research areas. One crucial aspect is the stage within the data processing where the fusion takes place. Combining the data in the beginning of the information extraction pipeline is called *early* or *feature fusion*. Fusing different pieces of information that have been extracted from different data sources before is called *late* or *decision fusion*. In practice, there are many other stages between early and late fusion. In [9], the authors show that the optimal stage of fusion also depends on the task and the data sources. In the following, a convolutional network structure for comparing different stages of fusion is presented.

2.1. Convolutional data fusion network

We use a very simple and relatively shallow net structure based on five convolutional layers, one dense layer, and a softmax output layer. Additionally, we consider a fusion layer, which fuses the data from different input branches into one single branch. The fusion layer can be placed at any position in the structure, leading to a network where, up to the fusion layer, each input has its own branch as described by the structure. Within this work, the fusion layer represents a simple concatenation of the channels of the single branches. The basic structure and a visualization of the resulting fusion networks are given in Figure 2.

2.2. Processing the input data

We use polarimetric SAR data from Sentinel-1, the same data processed by a Lee filter, and optical images from Sentinel-2. The SAR channels contain the real and imaginary parts of the signals for VV and VH polarization, the Lee-filtered intensities for VV and VH polarization, and the real and imaginary off-diagonal elements of the Lee-filtered covariance matrix. From Sentinel-2, we consider the bands B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. The bands B1, B9, and B10 are not considered since they represent mostly atmospheric properties, which contain no information for the classification of local climate zones. The bands B2, B3, B4, B8A, B12 have a Ground Sampling Distance (GSD) of 10 m. For the purpose of data consistency, the bands with 20 m GSD (B5, B6, B7, B8, B11) are upsampled to 10 m. The used optical bands can be grouped into RGB (Sentinel-2, B4, B3 and B2), vegetation red edge (Sentinel-2, B5, B6, B7, B8A), near infrared (NIR) (Sentinel-2, B8), and short wavelength infrared (SWIR) (Sentinel-2, B11, B12) information.

All in all, we have 18 channels of data and we consider four different ways of feeding the data into our network. The first and the second approach only use SAR and optical data, respectively. The third approach separates the input data into

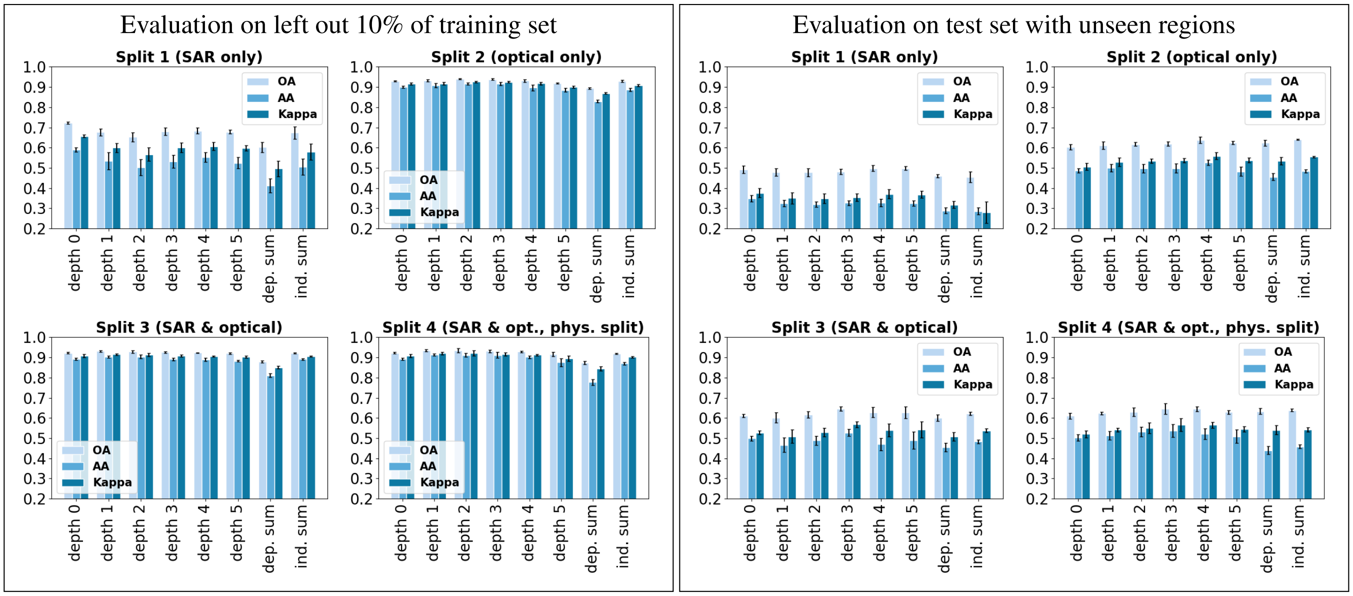

Fig. 4: Mean values and standard deviations of overall accuracy (OA), average accuracy (AA) and Kappa (K) based on 5 test runs with evaluation on left out 10% of the training set (left) and on the separate test set from the LCZ42 data set (right).

optical and SAR data. The fourth approach splits the data in the same way as in [6] by the physical concepts of the channels. The case of using all data without a split is covered by configuration with fusion stage 0. A visualization of the four considered splitting approaches is given in Figure 3.

3. EXPERIMENTS AND RESULTS

To test the network structure presented in Section 2.1, we use the LCZ42 dataset [10], provided by the Chair of Signal Processing in Earth Observation of the Technical University of Munich (TUM). The dataset can be downloaded from the TUM library: https://mediatum.ub.tum.de/1483140. The data set contains co-registered patches of size 32 $\times$ 32 pixels with dual-Pol data from Sentinel-1 and multispectral images from Sentinel-2. The data is split into a training set of 352366 patches and a validation and test set containing 24188 and 24119 patches, respectively, sampled from regions different to the regions of the training set.

We use the *Adam* optimizer with a learning rate of $\alpha = 10^{-3}$. In order to counteract the very unbalanced data set, we weight the loss for a sample of class i by $w_i = \frac{n}{n_i}$ where n is the total number of samples in the training set and n_i is the number of samples in the training set that belong to class i. After the first dense layer, dropout is applied with a rate of $d = 0.4$. The models are trained for 100 epochs with early stopping if the best reached overall accuracy does not change for 25 epochs. We consider two settings to train, validate and test the networks. For the first setting, the training set of the LCZ42 data set is split into random fractions of 80%, 10%, and 10% for training, validation, and testing. For the other setting, the whole training set is used for training and the validation and test sets are used for validation and testing. In the following, these settings are referenced as *80/10/10* and *train/val/test*. We test the four different splitting strategies described in 2.2 together with a fusion at stage $s \in \{0, 1, 2, 3, 4, 5, 6\}$, where s describes the number of blocks from the basic structure used for the single branches. The case $s = 0$ is equal to the case of no data splitting, the case $s = 6$ is equal to the fusion of decisions from the single branches, for which we simply sum up the corresponding softmax results. For $s = 6$, we consider two different procedures: the dependent and the independent sum. While for the dependent sum fusion all branches are summed already during the training process, the independent sum fusion sums up the results from independently trained networks, each representing one branch. Each experiment is run five times on different seed values and the results are averaged over these runs.

The test results for the two settings are shown in Figure 4. For the 80/10/10-setting and fusing optical and SAR data, a fusion between stages 0 and 5 delivers very good results with a mean of the average accuracy from 88.9% up to 91.2%. For the cases where optical or SAR data are used alone, the optimal fusion depth is at stage 3 and stage 0, respectively. While applying the dependent sum of the single branches delivers results with less accuracy for all considered types of input data splitting, the independent sum fusion performs differently, depending on the input data splitting and the considered accuracy measure. For the evaluation on the separate test set, accuracies are significantly lower, with an optimal average accuracy of 53.73% for seven input branches of optical and SAR data fused at stage 3.

4. DISCUSSION

The results show that the evaluation on unseen regions is significantly harder than on known regions, indicating that the models are overfitted on (the regions of) the training data and do not generalize well enough. The tests on the LCZ42 test set show that differentiating classes within urban and non-urban classes is a hard task. While dense trees (Class A), low plants (Class D), and water (Class G) have class accuracies from 70% to 99% in most cases, most others have weak accuracies from 10% to 45%. Especially the bush class is most often classified as low plants, resulting in acccuracies less than 4%. The accuracies on SAR-only input data are significantly lower than all other cases, showing that this model structure is not suited well enough for SAR data or that SAR data does not contain enough information in order to do a LCZ classification. With our model architecture and training strategy, the models suffer from overfitting, especially when evaluating on unseen regions. For the fusion of optical and SAR data the standard deviations of OA and AA lie between 0.2% and 2% for the 80/10/10-setting and between 0.6% and 4.2% for the train/val/test-setting. This makes the small differences in the results for most cases negligible.

Furthermore, the results raise the question whether splitting and fusing input data necessarily leads to increased accuracy. Not splitting the input data could lead to comparable or even better accuracy. Especially for the SAR-only experiments, not splitting the data leads to the highest accuracies. This might be due to increased overfitting caused by the additional parameters. Also, potential inference might be withheld from the model by splitting the data. In our case, the batch normalization after the first layer is a possible reason for the lower performance of the SAR-only input split.

5. CONCLUSION

In this work, we compare different split and fusion strategies of Sentinel-1 and Sentinel-2 data in order to perform a LCZ classification on the LCZ42 data set.

The results have shown that the transfer of the classes from one region to another and the handling of overconfidence of the classifiers are interesting fields for following research. Also the fact that the optical-only results are comparable to the results on optical and SAR data underline that a naive way of just putting data sources together will not be able to give a big boost. Thus, more advanced net architectures and training strategies should be considered. All in all, future work should focus on a better generalization and an optimal way of introducing the data sources, either based on more advanced net designs or (pre-) training configurations. Especially procedures that focus on complementing uncertain data of the single sources should be taken into account. Usage of data augmentations could also be considered. Due to the comparatively good results of the ensembling of independent branches and the uncertainty in the resulting models, we also want to investigate ensembling strategies.

6. REFERENCES

[1] I. D. Stewart and T. R. Oke, "Local Climate Zones for Urban Temperature Studies", *Bull. Amer. Meteor. Soc.*, vol. 93, no. 12, pp. 1879–1900, Dec. 2012.

[2] X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources", *IEEE Geosci. and Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017.

[3] C. Yoo, D. Han, J. Im, and B. Bechtel, "Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images", *ISPRS Journal of Photogramm. and Remote Sens.*, vol. 157, pp. 155–170, Nov. 2019.

[4] C. P. Qiu, L. Mou, M. Schmitt, and X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network", *ISPRS Journal of Photogramm. and Remote Sens.*, vol. 154, pp. 151–162, Aug. 2019.

[5] J. Rosentreter, R. Hagensieker, and B. Waske, "Towards large-scale mapping of local climate zones using multitemporal Sentinel 2 data and convolutional neural networks", *Remote Sensing of Environment*, vol. 237, pp. 111472, Feb. 2020.

[6] P. Feng, Y. Lin, J. Guan, Y. Dong, G. He, Z. Xia, and H. Shi, "Embranchment cnn based local climate zone classification using sar and multispectral remote sensing data", in *IGARSS 2019 - 2019 IEEE Int. Geosci. and Remote Sensing Symposium*. IEEE, 2019, pp. 6344–6347.

[7] J. Hu and X. Zhu, "Exploring Sentinel-L Data for Local Climate Zone Classification", in *IGARSS 2018 - 2018 IEEE Int. Geosci. and Remote Sensing Symposium*, July 2018, pp. 4677–4680.

[8] M. Schmitt and X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship", *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.

[9] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie, "Multilevel sensor fusion with deep learning", *IEEE sensors letters*, vol. 3, no. 1, pp. 1–4, 2018.

[10] X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Häberle, Y. Hua, R. Huang, L. Hughes, H. Li, Y. Sun, G. Zhang, S. Han, M. Schmitt, and Y. Wang, "So2Sat LCZ42: A benchmark dataset for global local climate zones classification", *arXiv preprint arXiv:1912.12171*, 2019.