

DETECTION OF TERRAIN STRUCTURES IN AIRBORNE LASER SCANNING DATA USING DEEP LEARNING

B. Kazimi*, F. Thiemann, M. Sester

Leibniz University Hannover, Institute of Cartography and Geoinformatics
{kazimi,thiemann,sester}@ikg.uni-hannover.de

KEY WORDS: Deep Learning, Airborne Laser Scanning, Digital Elevation Models, Multi-Modal Data Fusion

ABSTRACT:

Automated recognition of terrain structures is a major research problem in many application areas. These structures can be investigated in raster products such as Digital Elevation Models (DEMs) generated from Airborne Laser Scanning (ALS) data. Following the success of deep learning and computer vision techniques on color images, researchers have focused on the application of such techniques in their respective fields. One example is detection of structures in DEM data. DEM data can be used to train deep learning models, but recently, Du et al. (2019) proposed a multi-modal deep learning approach (hereafter referred to as MM) proving that combination of geomorphological information help improve the performance of deep learning models. They reported that combining DEM, slope, and RGB-shaded relief gives the best result among other combinations consisting of curvature, flow accumulation, topographic wetness index, and grey-shaded relief. In this work, we approve and build on top of this approach. First, we use MM and show that combinations of other information such as sky view factors, (simple) local relief models, openness, and local dominance improve model performance even further. Secondly, based on the recently proposed HR-Net (Sun et al., 2019), we build a tinier, Multi-Modal High Resolution network called MM-HR, that outperforms MM. MM-HR learns with fewer parameters (4 millions), and gives an accuracy of 84.2 percent on ZISM50m data compared to 79.2 percent accuracy by MM which learns with more parameters (11 millions). On the dataset of archaeological mining structures from Harz, the top accuracy by MM-HR is 91.7 percent compared to 90.2 by MM.

1. INTRODUCTION

Deep Learning (DL) techniques have gained a lot of popularity in many research fields. They are used to learn abstract representations of their inputs. The representations are exploited for solving a problem, or reaching a decision. The inputs, type of representation, and the decision could vary depending on the task. For example, deep learning models learn compressed, low dimensional vector representations (*features*) of an *input* image in order to produce a class label (*decision*) for it. DL models have achieved outstanding results in image classification (Voulodimos et al., 2018), object detection (Zhao et al., 2019), speech recognition (Nassif et al., 2019), medical imaging analysis (Kumar and Bindu, 2019) and neural machine translation (Stahlberg, 2019), among others. Researchers in remote sensing community use DL methods to extract useful information from hyperspectral images, LiDAR, and Radar data, to name a few (Zhu et al., 2017). In this work, we focus on applications of DL techniques for pattern recognition in Airborne LiDAR or ALS data. ALS refers to measuring the travel time for a light pulse emitted from a flying laser beam and reflected back from the ground (Shan and Toth, 2008; Farid et al., 2008). The recorded measurements in ALS point cloud data are not uniform. They are dense for certain locations, and sparse for others, essentially making them unstructured. Many DL models are adapted well to structured data such as images or videos. Therefore, it is advantageous to create regular raster grids such as DEMs from the ALS point clouds which could be fed to DL models for training (Guiotte et al., 2020). Values represented by DEM cells however show either absolute distance from the terrain to the acquisition device or relative elevations based on a reference surface, and in cases where the shape of objects and

structures are relevant regardless of how high or low of a terrain they are located at, only elevations relative to neighboring cells matter. Therefore, while training a DL model with DEM data, it is necessary to apply some preprocessing techniques such as local normalization or scaling (Torres et al., 2019). There exist multiple raster data visualizations created from DEMs that help scientists visually inspect interesting patterns and structures. Each visualization is produced by calculating the values of grid cells relative to the elevations in the neighboring cells only. Examples include slope, RGB-shaded relief, Curvature, Flow Accumulation (FA), Topographic Wetness Index (TWI), and Grey-Shaded Relief (GSR). While these visualizations are created to make features visually perceptible by humans, studies show their positive impact in recognition of patterns with DL techniques as well. Du et al. (2019) used a multi-modal approach (MM) for landform recognition and reported the most effective combination to be DEM, slope and RGB-shaded relief. The goal of this research is to study the effects of other such visualizations, either individually or combined with others, on the performance of DL models for detecting patterns.

Our contributions are two-fold. First, we use the MM approach and show that combination of other raster products such as Simple Local Relief Model (SLRM), Sky View Factor (SVF), Local Dominance (LD), Positive Openness (POS), and Negative Openness (NEG) helps deep learning models to perform even better, and detect objects with higher accuracy. Secondly, we build a multi-modal high resolution network (referred to as MM-HR) based on High Resolution Net (Sun et al., 2019) that has fewer parameters compared to MM, and gives a higher prediction accuracy.

To validate the performance of MM, and MM-HR, both models are trained and tested under the same settings on the ZISM50m

* Corresponding author

dataset provided by Du et al. (2019). We calculate SVF, SLRM, NEG, POS, and LD from the original dataset¹ and make them available online². We also run experiments on our own dataset collected from Harz, Lower Saxony with the same raster data variations and similar settings.

The rest of the paper is designed as follows. Section 2 lists related work. Section 3 describes the contributions of this research, followed by the experiments in Section 4. Results of the experiments and discussions are included in Section 5, and finally Section 6 concludes this study and points out future research in this direction.

2. RELATED WORK

Deep learning techniques, especially Convolutional Neural Networks (CNNs), became popular after AlexNet (Krizhevsky et al., 2012) won the ImageNet classification challenge (Russakovsky et al., 2015). AlexNet consists of a series of convolutional and max pooling layers, followed by two dense layers, and a final classification layer. The final layer uses a softmax function producing class probabilities for a given image. There have been many improvements in image classification using CNNs after AlexNet. Simonyan and Zisserman (2014) introduced VGGNet which uses smaller convolutional kernels and strides, and more layers leading to faster training time and better generalization. He et al. (2016) introduced residual learning in the ResNet model which facilitates adding more layers and non-linearities for learning complex mappings between the input and output while not harming simpler mappings, if any. Szegedy et al. (2014) proposed inception modules which rather than using one fixed-size kernel, use multiple kernels with different sizes to account for the variability in the size of important features in an image. DenseNet is proposed by Huang et al. (2017) which uses the output of each layer as input to every subsequent layer, contrary to previous approaches that only feed the output of one layer only to the following layer. This leads to higher prediction accuracy with reduced number of parameters. Other examples include Xception (Chollet, 2016), MobileNets (Howard et al., 2017; Sandler et al., 2018), and EfficientNet (Tan, Le, 2019). Most of the previous methods perform down-sampling, i.e., max pooling or striding, to reduce the resolution of the features and increase efficiency. As a result a lot of information is lost during the process. Recently High Resolution Network (HRNet) (Wang et al., 2019; Sun et al., 2019) was proposed that maintains high resolution, and produces rich semantic representations of the input through multiple parallel high and low resolution convolution and consistent exchange of information among them. This high resolution representation has proved to be superior for many tasks in computer vision such as semantic and instance segmentation, and object detection. If classification is intended, only the final outputs of multiple parallel convolutional branches are downsampled to be given to a classifier.

In addition to their success in natural images, deep learning techniques have been widely used in pattern recognition from ALS raster data. Many researchers use DEM as input data. For example Marmanis et al. (2015) use DEM to train a deep classification model for above-ground objects in urban area, Torres et al. (2018) apply deep learning to identify mountain summits in DEM data, and Politz et al. (2018) and Kazimi et al.

(2019a,b) explore detection of archaeological objects with deep learning. Patterns in DEM, especially smaller changes, are not easily visually comprehensible by the human eye. Other raster derivatives such as SLRM, SVF, POS, NEG, and LD help make the DEM patterns understandable. Kokalj and Hesse (2017) give detailed explanations and examples of previously mentioned raster derivatives created from DEM, but we give a short description of the ones we use in this research in Section 2.1 below:

2.1 ALS raster data derivatives

- **Simple Local Relief Model (SLRM):** SLRM is calculated by smoothing the original DEM, extracting points that are similar in both, the original and the smoothed DEM, and finally subtracting it from the original DEM. It is used to remove the major features in a DEM and highlight small structures such as those in archaeological mining (Gallwey et al., 2019).
- **Local Dominance (LD):** LD is the steepness angle of a point relative to its surrounding surface, calculated for a specified radius around the point. LD is suitable for protruded features such as barrows, and also deep features such as hollow ways.
- **Sky View Factor (SVF):** SVF for a point is calculated relative to its surrounding points within a specified radius to show what portion of the sky is visible. SVF is well suited for archaeological structures such as mining sink-holes (Kokalj, Hesse, 2017).
- **Openness:** Openness is calculated using the mean zenith or mean nadir angles for a certain point relative to its surrounding points within a defined radius. Mean zenith angle defines Positive Openness (POS), and mean nadir angle defines Negative Openness (NEG) (Doneus, 2013). While POS highlights protruded features such as rims in bomb craters, and ridges between hollow ways, NEG highlights deep features such as the actual hollow ways.

Since the aforementioned raster derivatives can help identify structures, they should be helpful in automatic learning tasks as well. Gallwey et al. (2019) used raster products such as SLRM, POS and NEG separately to train deep learning models for detection of mining pits. Other researchers used SLRM for their deep learning tasks (Trier et al., 2019; Verschoof-van et al., 2019). Recently, Du et al. (2019) conducted experiments and proved that the addition of other rasters in a multi-modal fashion improve deep learning models' performances. They report combining DEM with RGB-shaded relief, and slope lead to a higher accuracy in classification using a multi-modal learning approach explained in detail in Section 2.2.

2.2 Multi-Modal network

The MM network by Du et al. (2019) takes n input types and extract features for each of them in parallel. It then fuses the extracted features together and uses them to classify the given inputs. The parallel feature extractors for each input type works better than using a single feature extractor where all input types are fed together. This is because rather than extracting general features for the input types as a whole, extracting distinct features individually for all input types help the following layers discriminate objects and structures better. The overall architecture of the MM model is shown in Figures 1, 2, and 3.

¹ <http://www.adv-ci.com/download/geomorphology/>

² <https://seafle.cloud.uni-hannover.de/d/17ce9a0f343e415aaff1/>

The current paper builds on top of the MM approach, exploring better raster alternatives, and a better multi-modal network for this purpose, details of which are given in Section 3.1.

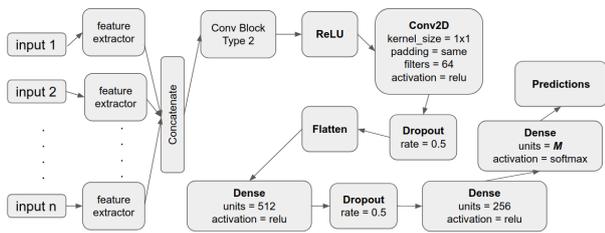


Figure 1. Architecture of the MM model. M and n denote the number of classes and input raster types, respectively. The *feature extractor* module is illustrated in Figure 2. *Conv Block Type 2* is shown in Figure 3

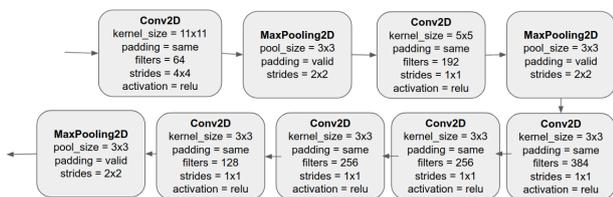


Figure 2. Feature extractor for the MM model shown in Figure 1.

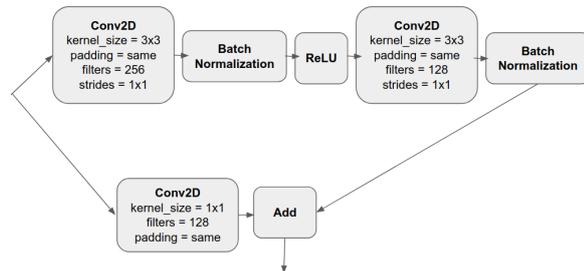


Figure 3. Convolutional Block Type 2 for the MM model shown in Figure 1.

3. METHOD

The main contribution of this research is proposing a multi-modal high resolution network referred to as MM-HR. It follows the idea of MM, but extends it to a simpler version of the recently proposed HRNet model proposed by Sun et al. (2019). Additionally, we propose and confirm the use and efficiency of other raster derivatives such as SLRM, SVF, LD, POS, and NEG for multi-modal learning tasks. Details of the proposed MM-HR model are given in Section 3.1.

3.1 Multi-Modal High Resolution network

The MM-HR model we propose follows the approach used in MM for feature extraction. The layers following feature fusion have the same structure as that of HRNet (Sun et al., 2019), but with fewer number of blocks and layers, and fewer parameters. MM-HR, similar to HRNet, does not downsample features in the layers after concatenation of the parallel feature extractor outputs. It rather maintains the high resolution throughout the

network till the end, before feeding it to a classification layer. Not downsampling leads to reduced loss of information, and better prediction. The MM-HR architecture is illustrated in Figures 4, 5, 6, and 7.

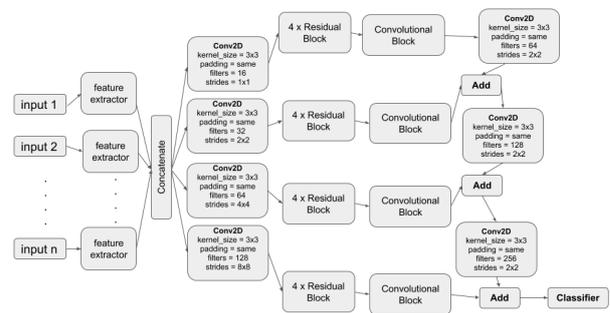


Figure 4. Architecture of the MM-HR model. M and n denote the number of classes and input raster types, respectively. The *feature extractor* module is illustrated in Figure 5. *Convolutional Block* and *Residual Block* are shown in Figures 6 and 7

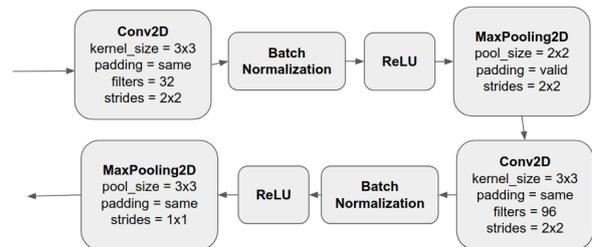


Figure 5. Feature extractor for the MM-HR model shown in Figure 4.

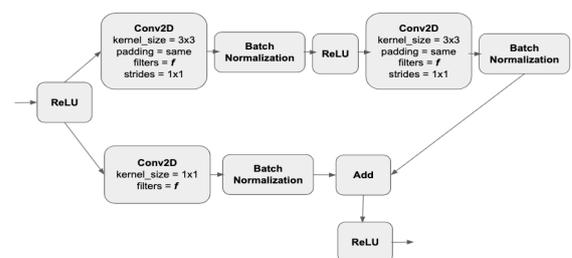


Figure 6. Convolutional Block for the MM-HR model shown in Figure 4. f depends on the number of input filters to the block.

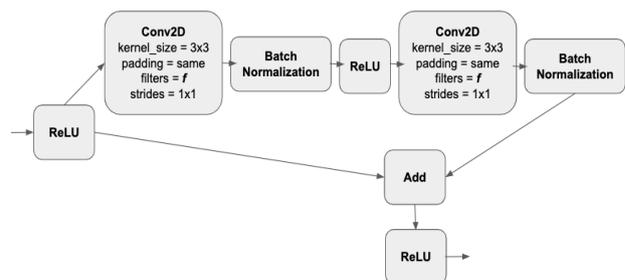


Figure 7. Residual Block for the MM-HR model shown in Figure 4. f depends on the number of input filters to the block.

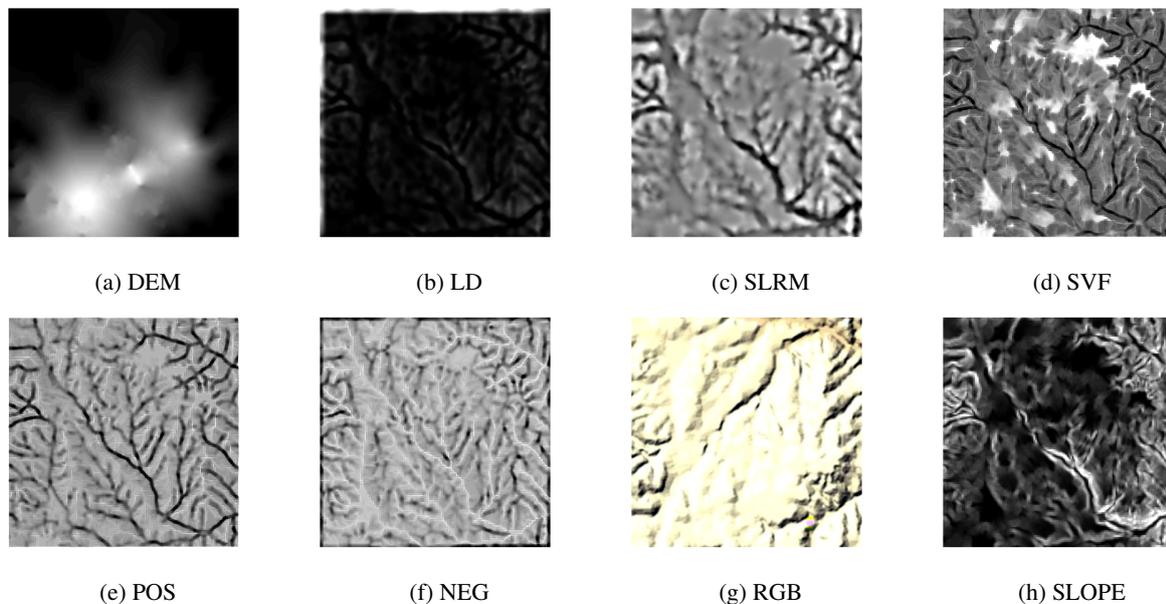


Figure 8. Examples of different raster derivatives for a fluvial landform.

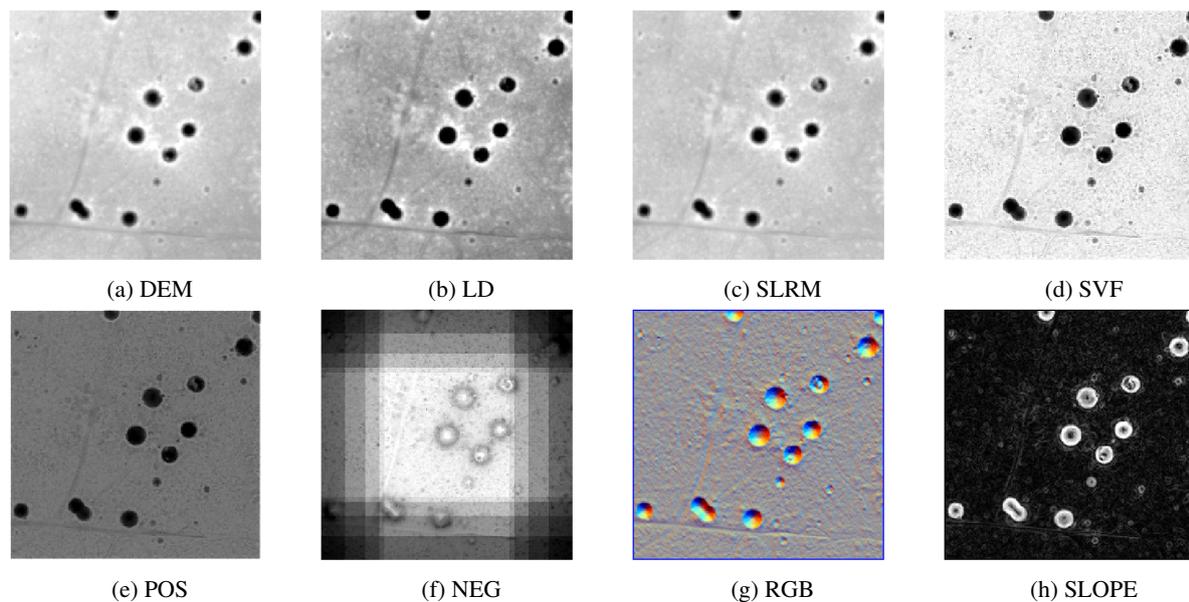


Figure 9. Examples of different raster derivatives for bomb craters from Harz.

Experiments are conducted training and evaluating both models under the same settings on both, ZISM50m and Harz dataset, details and results of which are given in Sections 4 and 5.

4. EXPERIMENTS

4.1 Data

For this research, we use two datasets. The first one is ZISM50m dataset provided by Du et al. (2019). It contains examples of six typical landform categories including aeolian, arid, loess, karst, fluvial, and periglacial from central China. There are 8400 examples (1400 for each category) in this dataset which are divided into 80, 10, and 10 percent splits for training, validation, and test set, respectively. The data has a resolution of 50 meters

per pixel. Each example is around 600x600 pixels, which represents a region of 900 km^2 . For memory restrictions and to speed up the experiments, we resize each example to 224x224 pixels each during training.

The second dataset is DEM acquired from the Harz Region in Lower Saxony which includes examples of archaeological mining structures such as bomb craters, charcoal kilns, barrows, and mining sinkholes. The DEM has a resolution of 0.5 meters per pixel. There are 1107 bomb craters, 1042 charcoal kilns, 1293 barrows and 2666 mining sinkholes. Each example has a size of 256x256 pixels. This dataset is divided into 80, 10, and 10 percent for training, validation, and testing, respectively.

With the available DEMs from both datasets, we calculate other raster products such as SVF, SLRM, LD, POS, Slope, RGB-

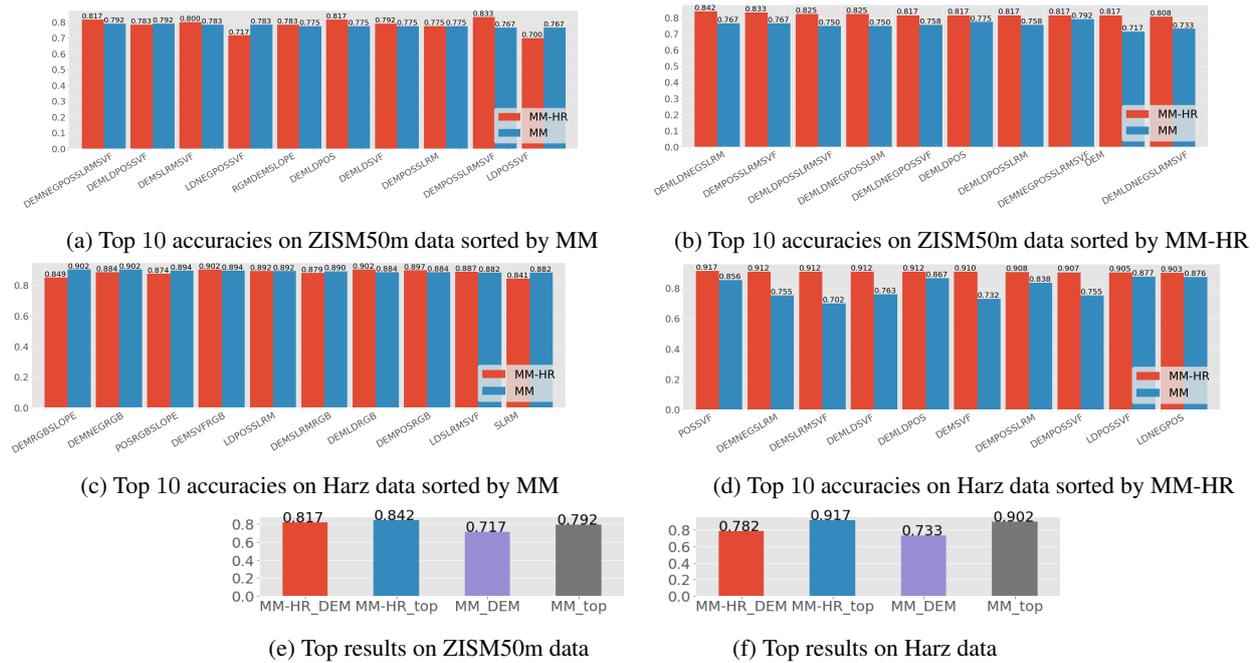


Figure 10. Top 10 accuracies among different combinations for both models.

shaded relief, and NEG using the Relief Visualization Toolbox (Kokalj, Somrak, 2019). An example of fluvial landform from ZISM50m dataset is shown in Figure 8, and an example of bomb craters from Harz is illustrated in Figure 9 .

4.2 Multi-Modal learning

To observe the effects of different raster products on detection of structures using deep learning, we use different combinations of 6 raster products: SVF, SLRM, LD, POS, NEG, and the original DEM. The 6 raster products can form 63 different combinations without repetition. They include single products and combinations of two, three, four, five, and finally six raster derivatives. Additionally, we use the best combination: RGB-shaded relief, slope and DEM , reported by Du et al. (2019). Thus, we run both models 64 times each.

For the Harz data, we use SVF, SLRM, LD, POS, NEG, DEM, RGB-shaded relief, and slope. Due to time constraints, we only try combinations of 1, 2, and 3 raster derivatives. Consequently, we run both models 92 times each on the Harz data. The results are detailed in Section 5.

4.3 Model training

Both models discussed in Section 3 are trained, and evaluated using the exact same settings, and on both datasets . The implementations are in Python using the Keras (Chollet et al., 2015) library. Models are trained to minimize sparse categorical cross entropy, aiming for the maximum categorical accuracy. Optimization function is stochastic gradient descent with the learning rate of 0.00001 and momentum of 0.9. Training is done using two GPUs for 100 epochs and batch size of 50 on ZISM50m data. The number of training epochs and batch size for Harz dataset are 50 and 32, respectively. Parameters leading to maximum accuracy on validation data during training are saved to disk, and used for evaluation on the test data.

5. RESULTS AND DISCUSSION

In this section, we first evaluate and compare the performances of both methods: MM and MM-HR on the two datasets. Additionally, we compare both methods in terms of computation and memory efficiency.

5.1 Detection performance

To compare the results of combinations of different raster derivatives, and the performance of both models, we first run the MM model on each (of the 64) combinations and calculate its accuracy on the test data. We sort the accuracies in descending order and show the corresponding accuracy achieved for each combination using the MM model on both datasets in Figures 10a and 10c. We only show the results for top 10 for visibility.

It is observed in Figure 10a that the highest accuracy by MM is 79.2% achieved with the combination of 5 raster types, namely *DEM*, *NEG*, *POS*, *SLRM*, and *SVF*. It is followed by combination of fewer features with a slightly lower accuracy, 78.3%. Combination of *RGB-shaded relief*, *DEM*, and *SLOPE* which was reported to have the highest accuracy in the study by Du et al. (2019) gives the 5th best accuracy, 77.5%.

It is also observed that almost for all the combinations, the accuracy produced by MM-HR is higher than that of MM. The top accuracy by MM-HR (91.7%) on Harz dataset is achieved using POS and SVF which is also higher than that of MM (90.2%) using DEM, RGB and slope. In Figures 10b and 10d, we illustrate the top 10 accuracies by MM-HR in a sorted manner. As illustrated in Figures 10e and 10f, MM-HR can output predictions with a higher accuracy using solely the *DEM* on both datasets. For further comparison, the confusion matrices by the best performing combination with both models on both datasets are shown in Figure 11. Numbers 0 to 5 indicate classes: aeolian, arid, loess, karst, fluvial, and periglacial, respectively for ZISM50m. For Harz, numbers 0 to 3 represent bomb craters, charcoal kilns, mining sinkholes and barrows, respectively. In

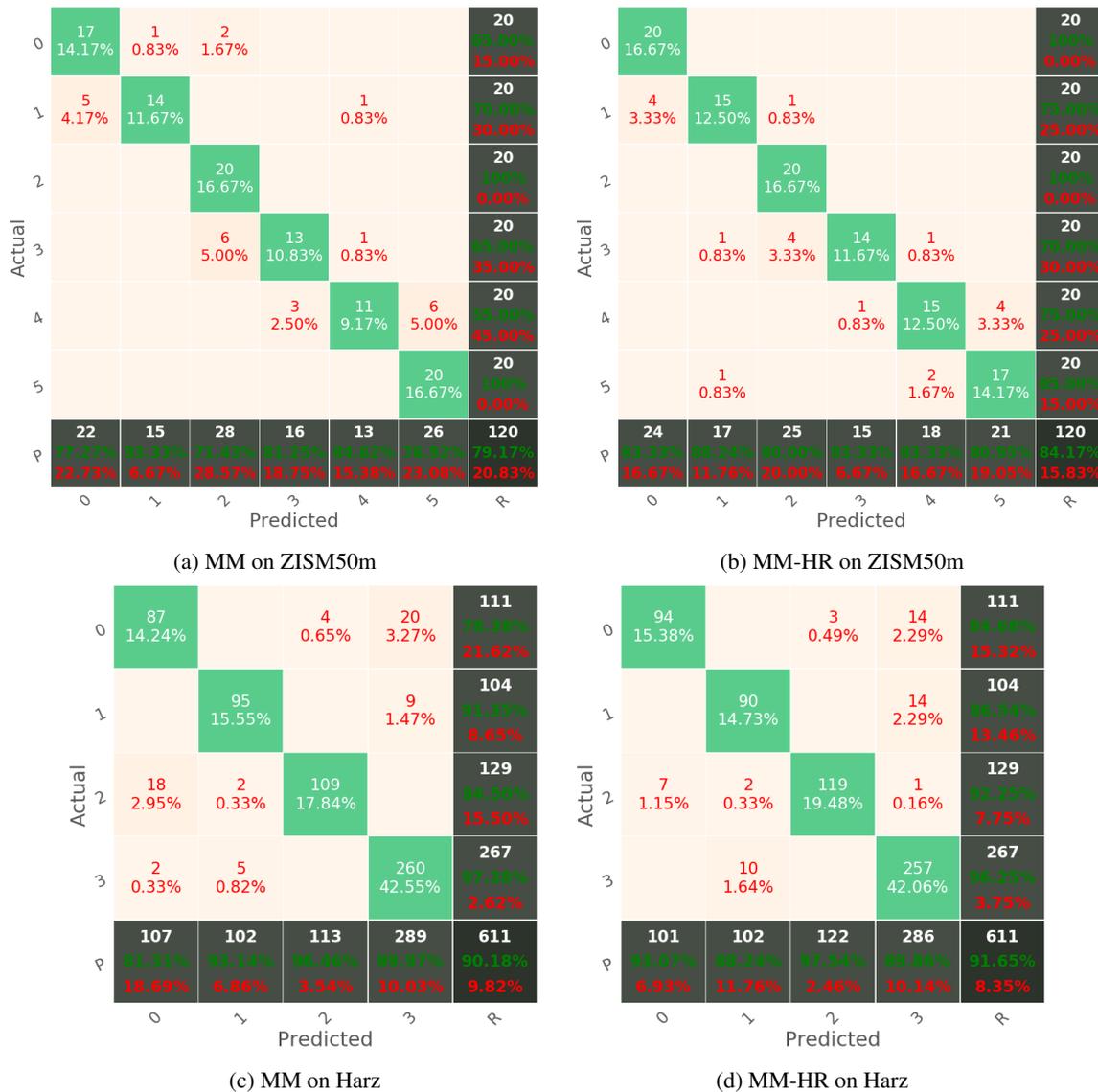


Figure 11. Confusion matrices.

the cells excluding the final row and column, the number at the top shows the either the number of correctly classified example (indicated by white) or that of incorrectly classified examples (indicated by red). The number at the bottom shows the percentage of its contribution to the total accuracy shown in the bottom right cell. The last row shows precision (P) and the last column shows recall (R) for each category.

5.2 Efficiency

In addition to giving a higher prediction accuracy as detailed in Section 5.1, MM-HR model has fewer parameters compared to that of MM model, making it almost 3 times smaller in size. On the other hand, since the feature extractor in MM downsamples features rapidly, and into very small resolutions compared to the gradual downsampling and high resolution features in MM-HR, MM model requires less memory than that of MM-HR and runs faster. Details of model parameters for both models are given in Table 1.

Model	Trainable	Non-trainable	Total
MM	10,814,598	768	10,815,366
MM-HR	4,021,446	7,968	4,029,414

Table 1. Parameters for both models. The parameters are for 3 parallel feature extractor modules for both models with input channels of 3, 1, and 1.

6. CONCLUSION AND OUTLOOK

In this research, we conducted experiments to see the effect of including other raster derivatives LRM, SVF, LD, POS, and NEG in addition to the commonly used DEM, in detection of structures using deep learning. We based our experiments on the recently proposed MM approach by (Du et al., 2019), and showed that the best accuracy (79.2%) is achieved by using combinations of DEM, NEG, POS, SLRM and SVF on ZISM50m data. The best accuracy by MM on Harz data is 90.2%. Additionally, we build a multimodal high resolution network, MM-HR, that is based on the High Resolution Network (Sun et al., 2019) and learns with fewer parameters (4 millions) than MM (11 millions). It produces a higher accuracy (84.2%) than that

of MM (79.2%) on ZISM50m data. The MM-HR accuracy (91.7%) is also higher than that of MM (90.2%) on Harz data. Classification methods are good when the desired task is to determine the existence of an object in a given location. However, if accurate location and boundaries of objects are expected, then we need to create models that perform pixel level predictions and give bounding box coordinates. Therefore, future research in this direction includes extension of the MM-HR approach for semantic and instance segmentation.

ACKNOWLEDGEMENTS

The project is funded by the Ministry of Science in Lower Saxony. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

REFERENCES

- Chollet, F., 2016. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*.
- Chollet, F. et al., 2015. keras.
- Doneus, M., 2013. Openness as Visualization Technique for Interpretative Mapping of Airborne Lidar Derived Digital Terrain Models. *Remote Sensing*, 5(12), 6427–6442. <https://www.mdpi.com/2072-4292/5/12/6427>.
- Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 63 - 75.
- Farid, A., Goodrich, D., Bryant, R., Sorooshian, S., 2008. Using airborne lidar to predict Leaf Area Index in cottonwood trees and refine riparian water-use estimates. *Journal of Arid Environments*, 72(1), 1 - 15.
- Gallwey, J., Eyre, M., Tonkins, M., Coggan, J., 2019. Bringing Lunar LiDAR Back Down to Earth: Mapping Our Industrial Heritage through Deep Transfer Learning. *Remote Sensing*, 11(17). <https://www.mdpi.com/2072-4292/11/17/1994>.
- Guiotte, F., Pham, M., Dambreville, R., Corpetti, T., Lefvre, S., 2020. Semantic Segmentation of d Points Clouds: Rasterization Beyond Digital Elevation Models. *IEEE Geoscience and Remote Sensing Letters*, 1-4.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.
- Kazimi, B., Thiemann, F., Sester, M., 2019a. Object instance segmentation in digital terrain models. M. Vento, G. Percannella (eds), *Computer Analysis of Images and Patterns*, Springer International Publishing, Cham, 488–495.
- Kazimi, B., Thiemann, F., Sester, M., 2019b. Semantic Segmentation of Manmade Landscape Structures in Digital Terrain Models. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7, 87-94.
- Kokalj, Ž., Hesse, R., 2017. *Airborne laser scanning raster data visualization: A Guide to Good Practice*.
- Kokalj, Ž., Somrak, M., 2019. Why not a single image? Combining visualizations to facilitate fieldwork and on-screen mapping. *Remote Sensing*, 11(7), 747.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097–1105.
- Kumar, E., Bindu, C., 2019. *Medical Image Analysis Using Deep Learning: A Systematic Literature Review*. 81–97.
- Marmanis, D., Adam, F., Datcu, M., Esch, T., Stilla, U., 2015. Deep neural networks for above-ground detection in very high spatial resolution digital elevation models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3), 103.
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., Shaalan, K., 2019. Speech recognition using deep neural networks: A systematic review. *IEEE Access*, 7, 19143–19165.
- Politz, F., Kazimi, B., Sester, M., 2018. Classification of laser scanning data using deep learning. *38th Scientific Technical Annual Meeting of the German Society for Photogrammetry, Remote Sensing and Geoinformation*, 27.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3), 211-252.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.
- Shan, J., Toth, C. K., 2008. Topographic laser ranging and scanning: principles and processing.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Stahlberg, F., 2019. Neural Machine Translation: A Review. *arXiv preprint arXiv:1912.02047*.
- Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., Mu, Y., Wang, X., Liu, W., Wang, J., 2019. High-Resolution Representations for Labeling Pixels and Regions. *CoRR*, abs/1904.04514. <http://arxiv.org/abs/1904.04514>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. *CoRR*, abs/1409.4842. <http://arxiv.org/abs/1409.4842>.
- Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 6105–6114.

Torres, R. N., Fraternali, P., Milani, F., Frajberg, D., 2018. A deep learning model for identifying mountain summits in digital elevation model data. *2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, IEEE, 212–217.

Torres, R. N., Fraternali, P., Milani, F., Frajberg, D., 2019. Mountain summit detection with Deep Learning: evaluation and comparison with heuristic methods. *Applied Geomatics*, 1–22.

Trier, . D., Cowley, D. C., Waldeland, A. U., 2019. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2), 165-175. <https://onlinelibrary.wiley.com/doi/abs/10.1002/arp.1731>.

Verschoof-van, d. V. W., Lambers, K. et al., 2019. Learning to Look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. *Journal of Computer Applications*, 2, 10.

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X. et al., 2019. Deep high-resolution representation learning for visual recognition. *arXiv preprint arXiv:1908.07919*.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., Wu, X., 2019. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212–3232.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.