Peer Reviewed Paper openaccess Paper Presented at SWIIms 2022, September 2022, Rome, Italy

Comparison of 2D and 3D semantic segmentation in urban areas using fused hyperspectral and lidar data

Agnieszka Kuras,[a] Anna Jenul,[a] Maximilian Brell[b] and Ingunn Burud[a]

[a]Faculty of Science and Technology, Norwegian University of Life Sciences, PB 5003, 1430 Aas, Norway

[b]Helmholtz Center Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Contact:
Agnieszka Kuras: agnieszka.kuras@nmbu.no
https://orcid.org/0000-0002-0541-5614
Anna Jenul: anna.jenul@nmbu.no
https://orcid.org/0000-0002-6919-3483
Maximilian Brell: maximilian.brell@gfz-potsdam.de
https://orcid.org/0000-0002-3759-7483
Ingunn Burud: ingunn.burud@nmbu.no
https://orcid.org/0000-0003-0637-4073

Multisensor data fusion has become a hot topic in the remote sensing research community. This is thanks to significant technological advances and the ability to extract information that would have been challenging with a single sensor. However, sensory enhancement requires advanced analysis that enables deep learning. A framework is designed to effectively fuse hyperspectral and lidar data for semantic segmentation in the urban environment. Our work proposes a method of reducing dimensions by exploring the most representative features from hyperspectral and lidar data and using them for supervised semantic segmentation. In addition, we chose to compare segmentation models based on 2D and 3D convolutional operations with two different model architectures, such as U-Net and ResU-Net. All algorithms have been tested with three loss functions: standard Categorical Cross-Entropy, Focal Loss and a combination of Focal Loss and Jaccard Distance—Focal–Jaccard Loss. Experimental results demonstrated that the 3D segmentation of U-Net and ResU-Net with Focal and Focal–Jaccard Loss functions had significantly improved performance compared to the standard Categorical Cross-Entropy models. The results show a high accuracy score and reflect reality by preserving the complex geometry of the objects.

Keywords: semantic segmentation, 3D segmentation, urban environment, hyperspectral imaging, lidar, hyperspectral lidar fusion, data fusion, deep learning, multisensor fusion, remote sensing

Introduction

An urban environment is a complex mosaic of diverse materials and surfaces constantly undergoing natural and anthropogenic processes resulting in rapid urbanisation.[1,2] This introduces environmental and ecological challenges, such as urban flooding, poor air quality, urban heat island issues and microclimate changes.[3–6] Due to these challenges and the complex and heterogeneous nature of urban areas, a proper advanced urban analysis

Correspondence
A. Kuras: agnieszka.kuras@nmbu.no

Received: 28 October 2022
Accepted: 30 October 2022
Publication: 7 November 2022
doi: 10.1255/jsi.2022.a11
ISSN: 2040-4565

Citation
A. Kuras, A. Jenul, M. Brell and I. Burud, "Comparison of 2D and 3D semantic segmentation in urban areas using fused hyperspectral and lidar data", *J. Spectral Imaging* **11,** a11 (2022).
https://doi.org/10.1255/jsi.2022.a11

is a prerequisite to getting information about urban land surfaces.

Over the last decades, passive and active remote sensing has been widely used in diverse analyses of urban environments, including land cover/material abundance mapping,[7,8] vegetation monitoring,[9–12] urban and suburban use analysis,[13] and three-dimensional urban geometry investigation.[14–16] Such advanced analyses are often achievable considering spectral, spatial and structural properties.[7,17] One sensor is often insufficient to obtain all that information.[18] This can be overcome with multisensory applications such as the fusion of hyperspectral (HS) data with lidar (light detection and ranging) scanning data,[19–22] synthetic aperture radar (SAR)[23] or thermal sensors.[24]

Hyperspectral imaging (HSI) provides detailed spectral information about materials, classifying complex urban structures and effectively distinguishing pixel-based soil, water, vegetation and impervious surfaces.[25] However, the classification is only considered at the near-surface without any elevation information. A land cover classification based on HS data in the spectral domain has been proposed in many studies, applying shallow machine learning methods.[26,27] However, a pure spectral analysis is often insufficient in object-based classification since it does not consider the spatial context—producing salt–pepper noise on the resulting image.[28] An object is a mixture of materials and structures differently sensitive to various atmospheric factors and illumination. This results in distorted results, bias and incomplete analysis.

Lidar scanning in the analysis of urban areas focuses primarily on geometric and textural information of objects and surfaces.[29] Lidar can improve the urban land cover analysis, especially of objects of the same material but differing in geometry, height and structure, such as high and low vegetation.[19,30] Since lidar, as an active sensor, is less sensitive to atmospheric influence and illumination, which challenge HS-based semantic segmentation, it can recompensate with a physical shadow correction for HS data analysis.[31,32]

Therefore, recent technological advances make multisensory fusion possible to exploit the potential of HS and lidar properties for urban analysis (HL-Fusion).[33–38] Such an HL-Fusion can be conducted on different levels focusing on the fusion of various data, products and application purposes based on physical and empirical approaches.[39] An HL-Fusion involves a physical understanding of sensors and awareness of the increased complexity of the analysis process. Therefore, no general process chain for an HL-Fusion exists yet, although deep learning networks have become promising for different classification and segmentation purposes in recent years. However, classifications based on deep learning models, such as convolutional neural networks (CNN),[40,41] deep residual U-Net (deep ResU-Net)[42] or deep residual net (deep Resnet)[33] are often carried out for HS or lidar data separately.

In urban land cover classification based on HS and lidar data, it is crucial to develop an algorithm that deals effectively with high-dimensional data, considers spectral, spatial and geometrical information simultaneously, and is stable and transferable to other areas with similar problem setups. Since HS data are high-dimensional (3D hyperspectral data cubes), they contain inherent spatial (x, y) and spectral information (λ). One of the main characteristics of an HS image is that it exhibits a strong correlation between the adjacent bands in the spectrum. Therefore, segmentation models that consider spatial features (2D convolutional operations) and the spectral dimension (3D convolutional operations) are believed to improve the performance since they allow for the finding of more patterns (Figure 1).

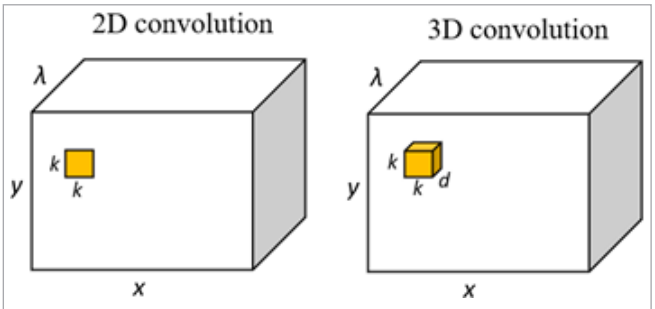

Figure 1. Comparison of 2D and 3D convolutional operations.

The 2D convolutional operation applied to hyperspectral and lidar data is powerful but has some limitations, especially in the spectral domain, since a 2D kernel slides only in the x and y directions. On the other hand, the 3D convolutional operation preserves the spectral information of the input HS data by moving in all three directions: x, y, λ. Furthermore, 3D convolutional operations are designed to exploit spatial continuity further and suppress noisy prediction.[43]

While many different 3D segmentation networks have been proposed in recent studies[43–46] for HL-Fusion, especially in urban environments, little research about 3D convolution-based networks for fused HS and lidar data exists.[47] Inspired by HL-Fusion and deep learning, we propose a method to analyse fused HS and lidar data for urban land cover classification. We compare two model architectures—U-Net and deep ResU-Net (residual deep

U-Net) at 2D and 3D convolutional operation levels applying different loss functions for multiclass segmentation. The two key contributions of this study are:

1) We suggest the design of an HL-Fusion on the feature extraction level for urban land cover classification.
2) We present the first study in HL-Fusion using 3D convolutional operations in U-Net and deep ResU-Net for urban land cover classification.

Theoretical background

Segmentation models

U-Net

The U-Net architecture, introduced in biomedical image segmentation by Ronneberger *et al.*,[48] is based on convolutional neural network layers. The architecture copies low-level features to the corresponding high-level features (Figure 2).

Due to the contracting path propagating between low-level features and high-level semantics in a facilitated way, U-Net takes context into account and hence helps to localise objects precisely.

ResU-Net

A residual neural network is built of stacked residual units.[49] The intuition behind the residual network is to introduce skip connections that prevent accuracy degradation in very deep networks with multiple hidden layers (Figure 3).[50]

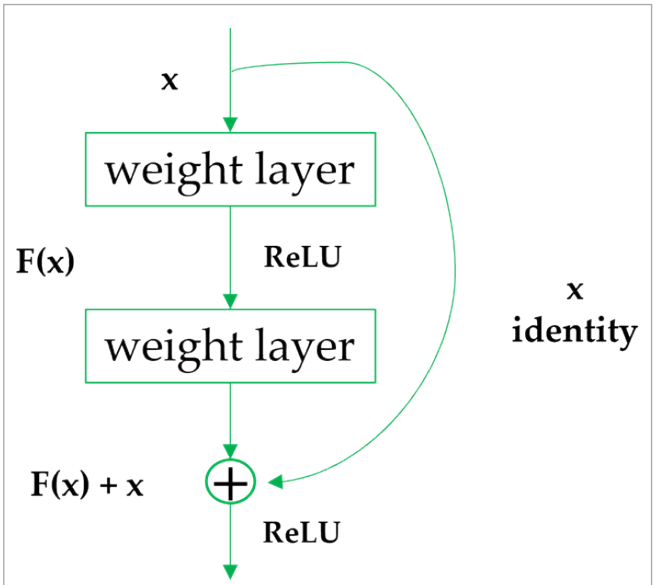

Figure 3. Deep residual network unit.

Skip connections learn identity functions that help follow higher layers perform at least as well as the previous lower layers. Therefore, information loss is reduced and the risk of vanishing gradients during training decreases. The residual unit facilitates the training and skip connections, making the segmentation stable and transferring the information from one layer to another without any information loss. This allows a calculation with fewer parameters.[42] The 3D ResU-Net consists of three parts: encoding, bridge and decoding (Figure 4). Bridge connects both encoding and decoding paths. The encoding part transforms the

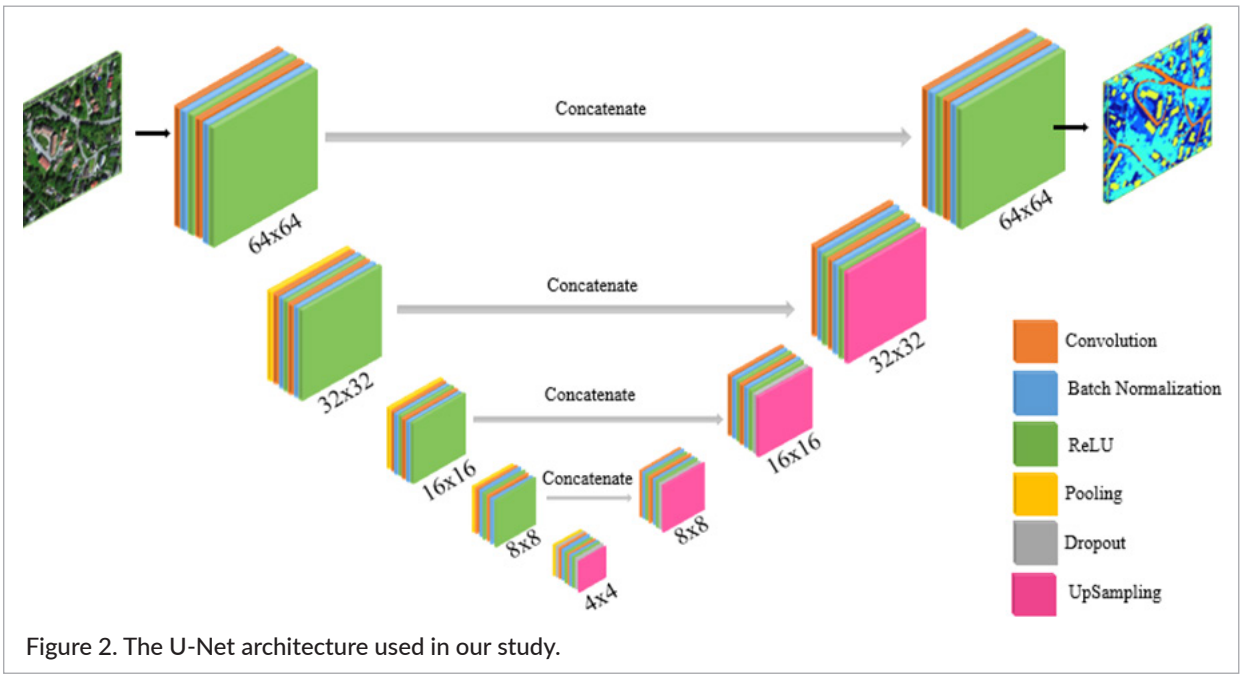

Figure 2. The U-Net architecture used in our study.

input into a compact representation. The decoding part puts the compact representations into a pixel-wise categorisation. All three parts include residual units built of a batch normalisation layer, a ReLU activation layer and a 3D convolutional layer. Both encoding and decoding paths do not contain a pooling operation that downsamples the output to maintain the spectral and spatial dependency in high-dimensional HS and lidar data.

Losses

A loss function must be specified to optimise the weights of the segmentation algorithms. The selection of the loss function is crucial and can significantly impact model optimisation. In Reference 51, Ma compares and categorises different loss functions; Duque *et al.*[52] divide them into two main groups: statistical-based and geometric-based loss functions. While statistical-based loss functions try to minimise the dissimilarity between distributions (derived from Kullback–Leibler divergence), geometric-based functions focus on overlapping regions between prediction and ground truth. Our work compares two statistical-based [Categorical Cross-Entropy (CE) Loss and Focal Loss] and one geometric-based representative (Jaccard Loss). Furthermore, to get the best out of both, we define the Focal–Jaccard Loss, combining Focal and Jaccard losses in an additive way.

Only the (ground truth) positive class contributes to the loss function as we use one-hot encoded target for training the network. In our notation, p_t is the probability of assigning a pixel t to its true class resulting from a SoftMax activation. The SoftMax function normalises a multidimensional network output to a probability distribution over the model classes. The desired result is that the probability for the true class is significantly higher than for the other classes.

Widely used in classification problems, the CE Loss function is defined as

$$CE(p_t) = -\log(p_t). \tag{1}$$

Hence, the higher p_t is, the better the prediction and the lower the CE Loss is. Although CE is quite common and often the first choice when selecting loss functions, it shows weakness when applied to a model with imbalanced classes.[53] Suppose the dataset is heavily imbalanced, and the number of easy-to-classify samples is much higher than the number of difficult-to-classify samples. In that case, the CE Loss will learn more representations from easy-to-classify samples instead of representations from hard-to-classify samples.

Furthermore, the loss for pixels that are easy to classify (usually the background) is relatively high compared to hard-to-classify samples (such as houses or streets), as shown in Figure 5. As an example, we compare the CE Loss for $p_t = 0.6$ and $p_t = 0.3$. While for $p_t = 0.6$, the correct pixel classification is almost certain, and the CE Loss is still around 0.5 compared to $p_t = 0.3$, where a correct

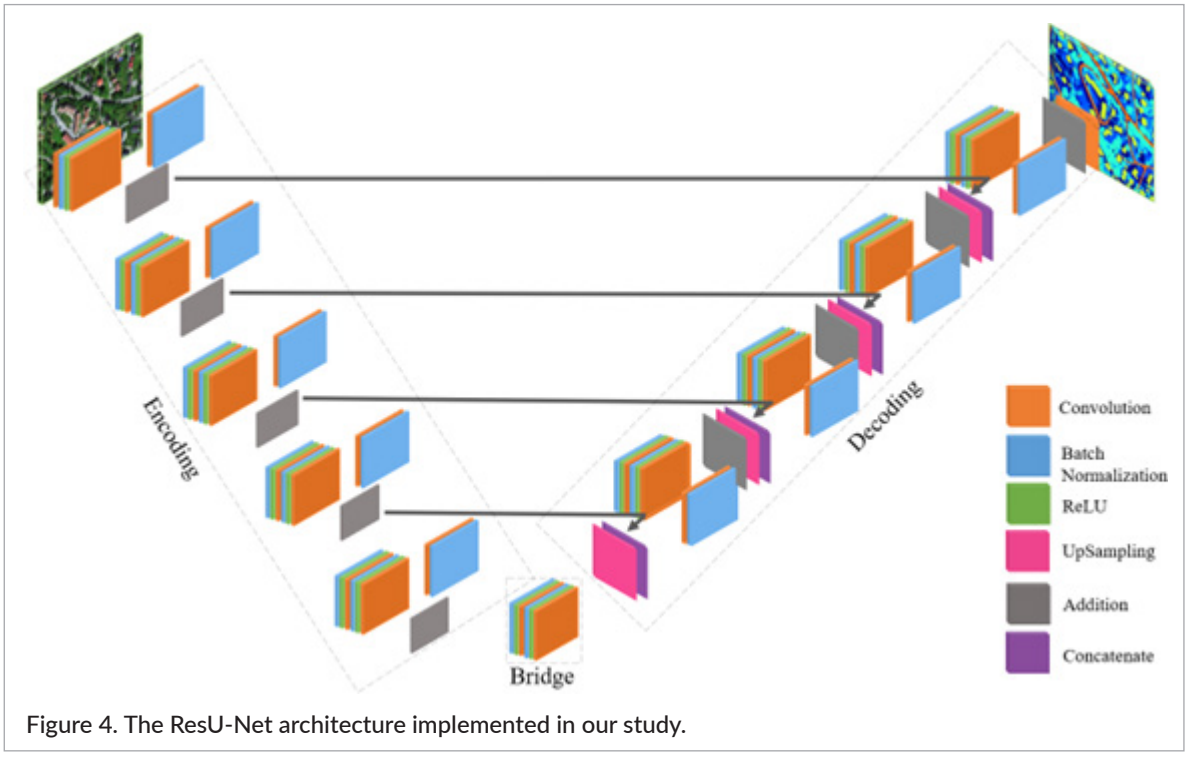

Figure 4. The ResU-Net architecture implemented in our study.

classification is unlikely, but where the loss between 1 and 1.5 is not much higher than 0.5. For this reason, Tsung-Yi Lin *et al.*[53] invented Focal Loss, which extends CE Loss by a multiplicative weight term α (assigning a higher weight to under-represented classes) and a modulating factor $(1-p_t)^\gamma$, where $\gamma > 0$ is a shape parameter. Focal loss is defined as

$$FL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t). \tag{2}$$

If $\gamma = 0$ and $\alpha_t = 1$, the Focal Loss is equal to CE Loss. Both γ and α_t are hyperparameters that control the learning process. α_t helps weight classes and balance the importance of positive and negative labelled pixels.[53] The higher γ gets, the more Focal Loss focuses on difficult pixels, see Figure 5 where it does not penalise a lot when $p_t \gtrsim 0.6$, i.e. when the certainty in correct prediction increases. Nevertheless, if γ is too high, already weak probabilities would get a very low loss—a good hyperparameter selection is, therefore, essential. Different strategies can be found in the original paper.[53] In our work, we set $\gamma = 2$ and $\alpha_t = 0.25 \forall t$.

The Jaccard Loss relies on the Jaccard Index,[52] measuring the Intersection over the Union of two sets. For our one-hot encoded setup, we compute the Jaccard Loss for a single sample t as

$$1 - \frac{p_t + \varepsilon}{1 + \varepsilon} \tag{3}$$

which is a decreasing linear function with ε representing a small constant. In our case, the function's slope changes slightly, but in the original definition, it intends to prevent division by zero. By default, we set epsilon, ε, to 1.

The fourth loss in our study is the additive combination of Focal Loss and Jaccard Loss which we refer to Focal–Jaccard Loss (FJL). The loss is given as

$$FJL(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t) + 1 - \frac{p_t + \varepsilon}{1 + \varepsilon} \tag{4}$$

In Figure 5, we observe that FJL behaves like Focal Loss for small p_t, whereas it converges towards Jaccard Loss for increasing p_t.

Dimension reduction: spectral unmixing

Airborne-based optical data are characterised by high spectral resolution but correspondingly lower spatial resolution.[54] A single scene pixel usually contains various materials creating a mixed pixel.[55] In order to separate urban materials regardless of pixel size and mixture complexity, spectral unmixing is used to define the amount of a given material in a pixel (abundance map). Various scientific fields have applied spectral unmixing in HS data analysis for decades.[56–60]

Here, abundance maps for each endmember are generated with unsupervised endmember extraction and spectral unmixing, simplifying the analysis and effectively reducing dimensions. Hence, we avoid a curse of dimensionality,[61] which occurs easily when working with limited training data. In such scenarios, the deep learning model has to deal with many features to achieve reliable and accurate results.[19]

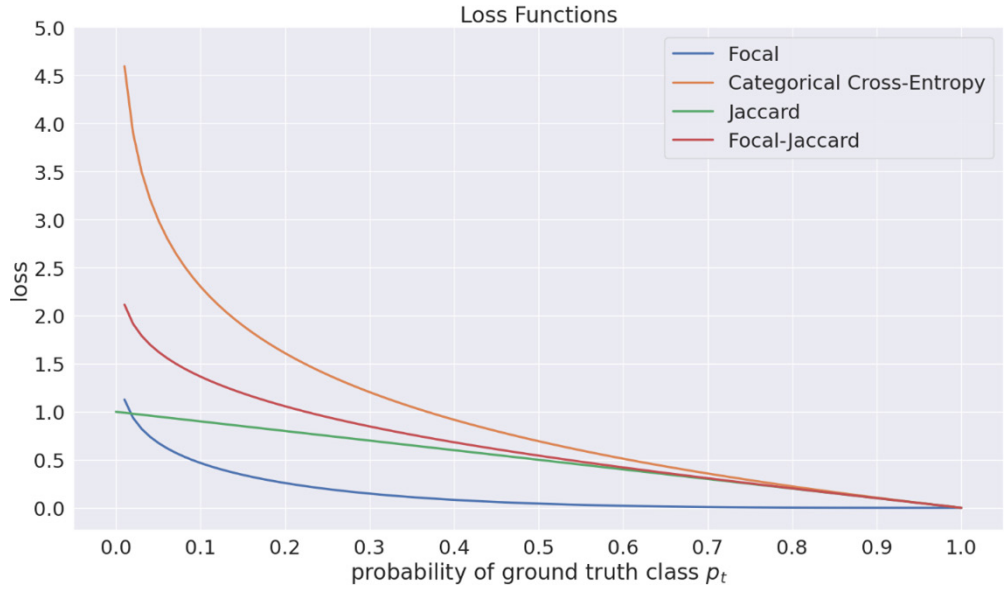

Figure 5. Comparison of the four loss functions used in this study. The limit towards 0 for CE Loss, Focal Loss and Focal–Jaccard Loss is infinite (due to the logarithm).

N-FINDR

The N-FINDR algorithm for endmember extraction was introduced by Winter.[62] This iterative approach aims to find endmembers corresponding to the purest spectra in the scene. N-FINDR assumes that the largest volume formed by an *n*-simplex with vertices can only be achieved by those purest pixels. The algorithm finds the final endmembers by randomly selecting the pixel set as initial endmembers, further iteratively investigating all spectra in the scene and searching for the largest volume of the simplex. The exact position of the defined endmember is then used to set the original spectrum before the dimension reduction as the final endmember signature. Each pixel in the scene is described as a linear mixture of the defined endmembers using non-negativity-constrained least squares fitting.[63]

Dataset

Terratec AS Company collected HS and lidar data in June 2021 over Baerum municipality in Norway (Figure 6).

The dataset represents a suburban environment with typical urban settlements, infrastructure and diverse vegetation. The airborne-based data were acquired under cloud-free conditions, including HS images and lidar-based 2.5-dimensional rasters. The HS data consist of images from HySpex sensors: VNIR-1800 (0.4–1.0 µm) and SWIR-384 (1–2.5 µm) with 0.3 m and 0.7 m spatial resolution, respectively. The lidar data were acquired using a Riegl VQ-1560i laser scanner, with five pulses per m^2 and intensity at 1.064 µm relevant for urban environments. The hyperspectral data were georeferenced and orthorectified by the Terratec AS using the PARGE software (Parametric Geocoding and Orthorectification for Airborne Optical Skanner Data).[64] The program corrects the geometry for each HS image pixel using a digital elevation model of the region, GPS position and attitude (ReSe Applications). The geocoded radiance data were converted to reflectance, adjusting illumination levels using ATCOR-4 (Atmospheric and Topographic Correction for airborne imagery). Absorption features associated with H_2O and OH close to bands at 1.4 µm

Figure 6. Study area representing a suburban environment in Høvik, near the capital Oslo in Norway.

and 1.9 µm were excluded from further analyses. The images from all flight lines with residual 176 channels for VNIR and 262 channels for SWIR were mosaicked separately using ENVI software (Environment for Visualizing Images). The noisy bands and outliers were set to bad bands and excluded from analyses resulting in 398 hyperspectral bands. The mosaic layers of VNIR and SWIR were stacked by applying the Savitzky–Golay filter[65] to all spectra. Our study area is located in Høvik with a coordinate extent of 588060, 6641500; 588878, 6641735 WGS 84/UTM zone 32N (Figure 6). The dataset contains the Joint Geospatial Database (FKB) that includes polygons of buildings, roads and other urban surfaces from 2010 to 2019. We carefully controlled FKB-polygons and adapted some modifications for labelling data. The residual classes (high and low vegetation) were extracted automatically.

Proposed method

Semi-automatic label preparation

In our analysis, we used five main classes extracted from the study area: low vegetation, high vegetation, buildings, roads and railway. Our segmentation is limited to these classes to facilitate the analysis and verify and deepen the understanding and optimisation of the method. In the semi-automatic label extraction, we identified high and low vegetation using the advantage of the high amount of relevant HS and lidar features. First, we masked out classes assigned to buildings to exclude potential errors. We applied the Normalised Difference Vegetation Index to HS data,[66] calculating the average for the red $\overline{x}_{RED}$ (660–670 nm) and near infrared band $\overline{x}_{NIR}$ (810–835 nm), respectively:

$$NDVI = \frac{\overline{x}_{NIR} - \overline{x}_{RED}}{\overline{x}_{NIR} + \overline{x}_{RED}} \quad (5)$$

We concatenated the vegetation mask with raster-based lidar features such as normalised Digital Surface Model, surface roughness and multiple returns. Significant are multiple returns, which distinguish between low vegetation and high vegetation. The assumption is that high vegetation is represented by more returns than low vegetation, usually located at ground points or impervious surfaces through which the laser does not penetrate multiple times. The rest of the classes of interest, such as roofs, railways and roads, were visually validated and extracted from the FKB database. The ground truth map is displayed in Figure 7.

Endmember extraction and abundance maps

Our study separately implemented state-of-art iterative endmember extraction (EA) algorithm N-FINDR[62] for HS and lidar data. We applied the non-negativity-constrained least squares algorithm for abundance map generation. The preprocessed reflectance normalised image in the HS analysis was used to retrieve 27 endmembers. For lidar EA, we built a lidar feature space where the five most relevant raster-based features have been extracted, including slope, the intensity from the first return, multiple returns, normalised Digital Surface Model (nDSM) and point density. All the features have been then normalised separately before EA. The initially extracted endmembers for both HS and lidar data were used to generate abundance maps for each endmember, retrieving 32 abundance maps.

Semantic segmentation

In this study, we applied two semantic segmentation model architectures: a U-Net and a deep residual U-Net that takes full advantage of U-Net architecture and a residual neural network. Both architectures were tested on two- and three-dimensional convolutions. All four model types (2D U-Net, 2D ResU-Net, 3D U-Net and 3D ResU-Net) performed with different loss functions

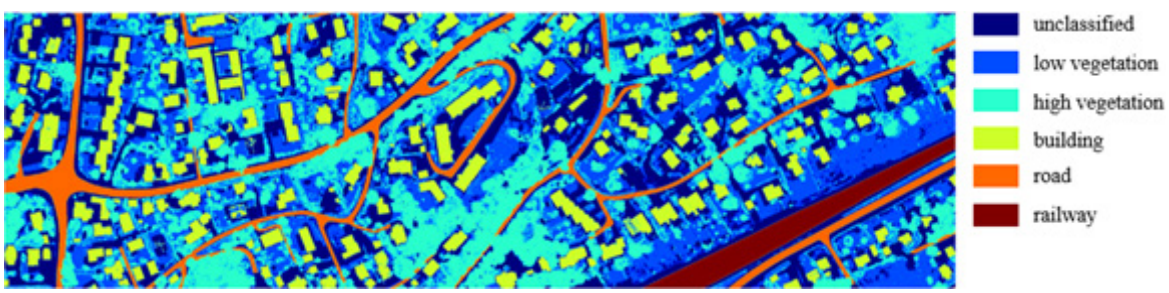

Figure 7. Ground truth map of the study area with five classes.

described above, resulting in 12 segmentation model combinations (Figure 8).

Implementation details

All 12 segmentation models were implemented in Python using Tensorflow, including GPU functionalities.[67] In all models, the maximum number of epochs was set to 500; however, we applied an early stopping function to save time, stop training when the model stopped improving and avoid model overfitting. For both U-Nets and ResU-Nets, we applied the Adam optimiser[68] with the learning rate of 1e–03, epsilon of 1e–07, $\beta_1 = 0.99$ and $\beta_2 = 0.999$. We implemented the Matthews Correlation Coefficient (MCC) to evaluate the algorithm's performance. MCC deals with unbalanced classes calculating the accuracy for each class separately.[69] Compared to F1-score, it takes all true positives, true negatives, false positives and false negatives into account, giving a more reliable performance result. MCC is bounded between –1 and 1, where –1 means that all predictions are incorrect, 1 indicates that all predictions are correct and a score of 0 denotes random predictions. Hence, the higher MCC gets, the better the results are. The input data dimensions for both 2D and 3D convolution-based models can be found in Table 1, splitting the study area into 80% training and 20% testing. The split between training and validation is carefully adapted so that all defined classes are involved equally to ensure that training data contain all known classes (Figure 9).

The input data for the 2D model consist of N number of patches with a size of 64×64 pixels and 32 abundance maps from both HS and lidar data. The ground truth for the 2D model contains five classes with the same N number and size of patches as the input. For the 3D model, the input data have to be expanded by one additional dimension compared to the 2D model. In the ground truth for the 3D model, we must stack the ground truth values 32 times for each abundance map, creating a matrix with the dimension $N \times 64 \times 64 \times 32 \times 5$ (Table 1).

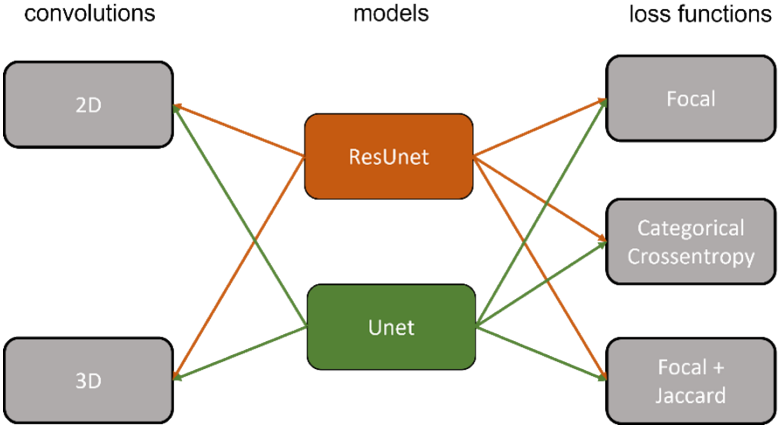

Figure 8. Twelve semantic segmentation combinations were implemented in the study.

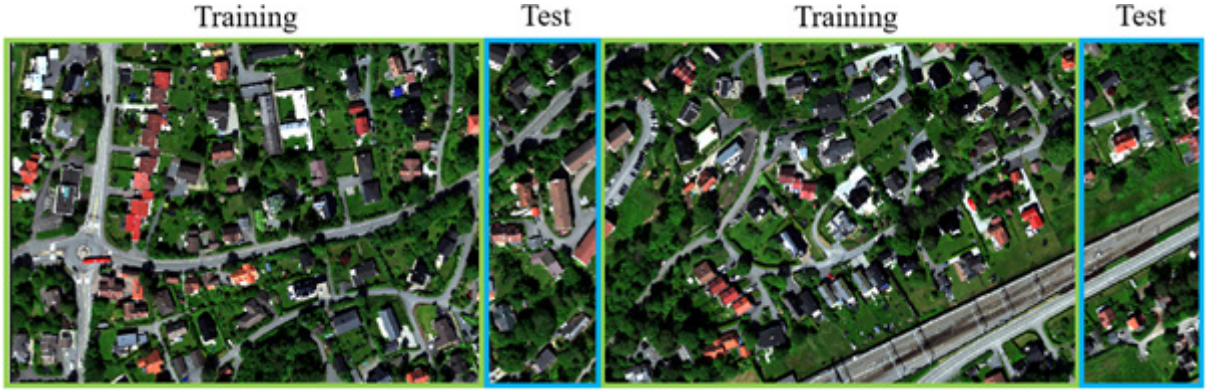

Figure 9. Training and test data separation. Green and blue outlines represent training and test sets, respectively.

Table 1. The input data dimensions for both 2D and 3D convolution-based models.

	2D model	3D model
Input	$N \times 64 \times 64 \times 32$	$N \times 64 \times 64 \times 32 \times 1$
Ground truth	$N \times 64 \times 64 \times 5$	$N \times 64 \times 64 \times 32 \times 5$

Experimental results and discussion

Four segmentation methods, including 2D U-Net, 2D ResU-Net, 3D U-Net and 3D ResU-Net, with three different loss functions: CE, Focal Loss and combined Focal–Jaccard Loss, were compared and evaluated. Table 2 presents the experimental results of all segmentation models reporting the MCC metric for all five classes and the overall accuracy of the test set. In this section, the results are presented and discussed in detail.

In order to test and compare 2D and 3D convolutional operation-based segmentation models, we configured U-Net and ResU-Net models for 2D and 3D convolutions and kept the parameters consistent by comparing loss functions such as CE, Focal Loss and Focal–Jaccard Loss. All segmentation maps from the whole scene are shown in Figure 10.

Model evaluation

Based on the results presented in Table 2, it can be noted that the best results for individual classes (shaded fields in grey) were obtained mainly by models based on 3D convolutional operations. Comparing the number of classes best identified in the model, the 3D ResU-Net with combined Focal–Jaccard Loss outperforms, recognising 98% of high vegetation and 99% of buildings and railways.

Both 2D and 3D segmentation algorithms within the ResU-Net model architecture have shown that using either Focal Loss or combined Focal–Jaccard Loss significantly outperforms the overall accuracy score (based on MCC) of CE.

Looking at the overall accuracy of each model, we can derive the two best models, such as 2D U-Net with CE and 3D ResU-Net, using Focal Loss. Looking at the total segmentation result for each class, these two models achieved more than 80% for each class. In the case of the simplest of the proposed models (2D U-Net), CE performed almost as well as the best of the models (3D ResU-Net), obtaining slightly lower accuracy than 3D ResU-Net. The 2D U-Net, despite the simplicity, is more stable and not sensitive to loss functions.

Single class accuracy

Starting the interpretation of the accuracy of a single class, we want to point out that the most challenging class

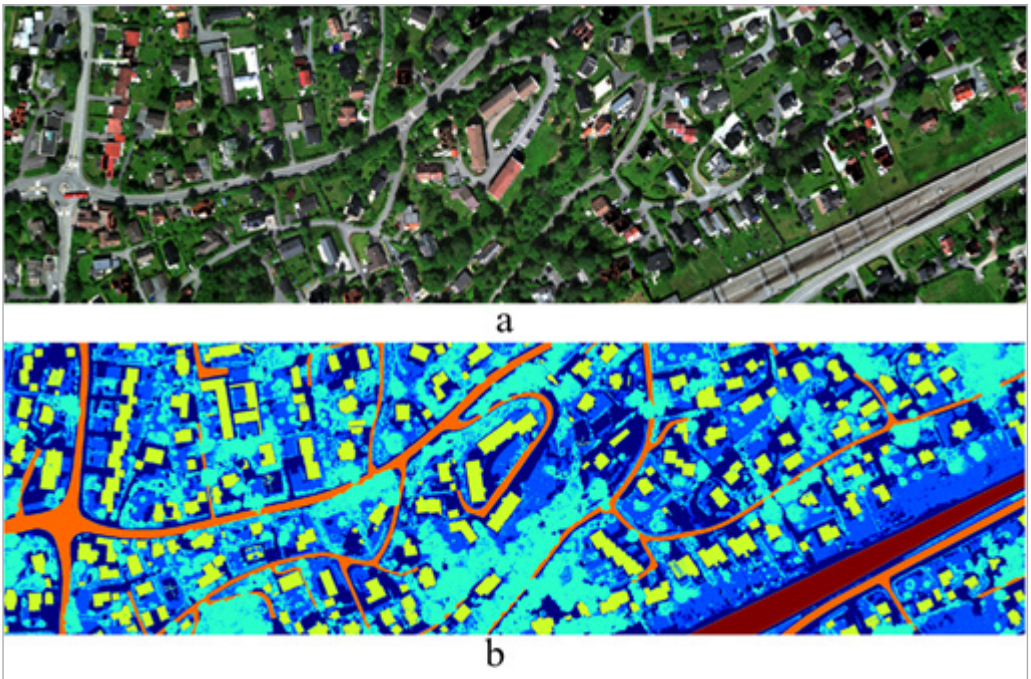

Figure 10. Qualitative comparison of 2D (c–h) and 3D (I–n) convolution-based model in urban land cover classification. a) HS image in RGB colour representation, b) ground truth with five classes such as low vegetation, high vegetation, building, road and railway.

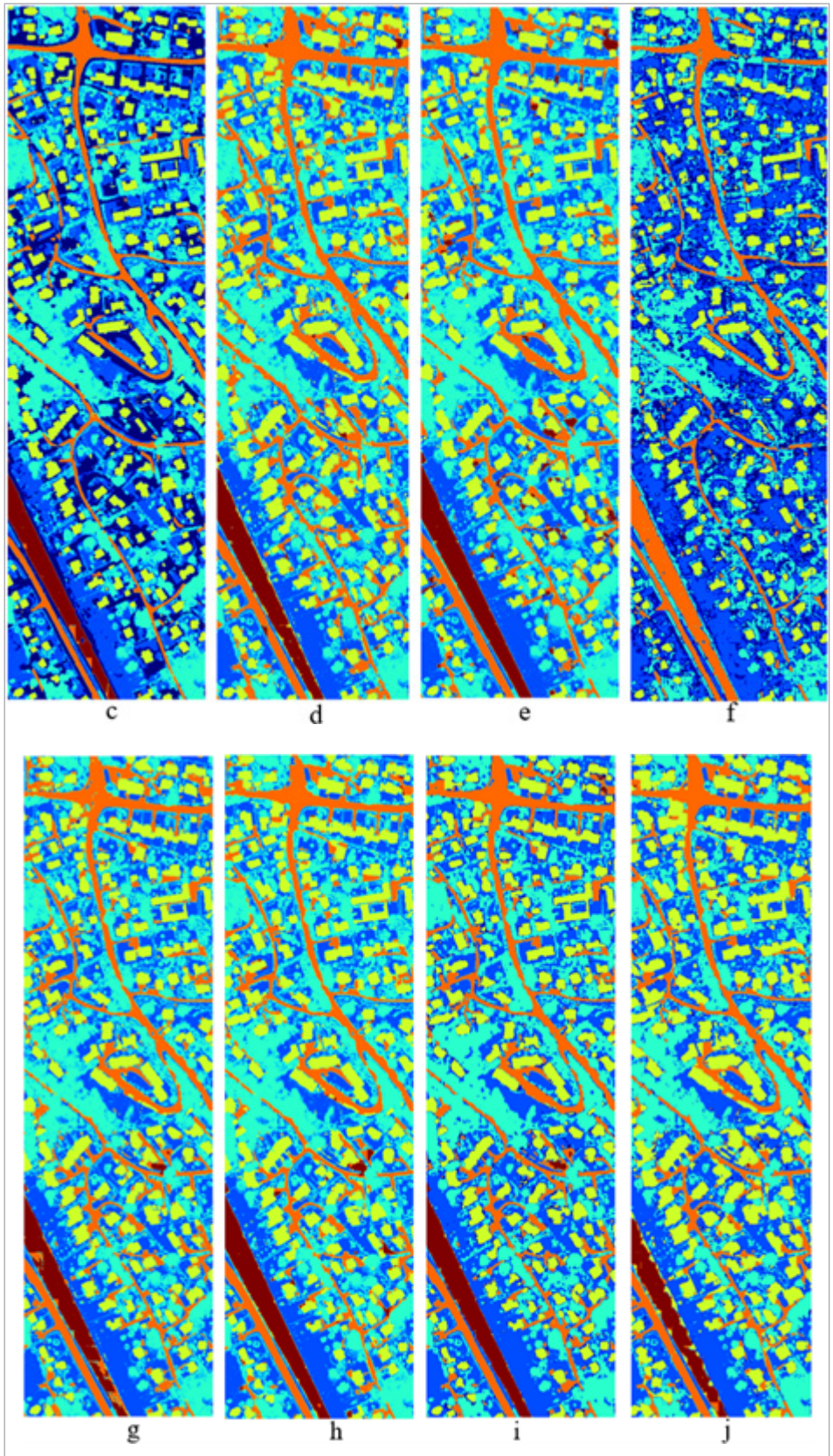

Figure 10. Qualitative comparison of 2D (c–h) and 3D (I–n) convolution-based model in urban land cover classification. c) 2D U-Net CE, d) 2D U-Net Focal Loss, e) 2D U-Net Focal–Jaccard loss, f) 2D ResU-Net CE, g) 2D ResU-Net Focal Loss, h) 2D ResU-Net Focal–Jaccard Loss, i) 3D U-Net CE, j) 3D U-Net Focal Loss.

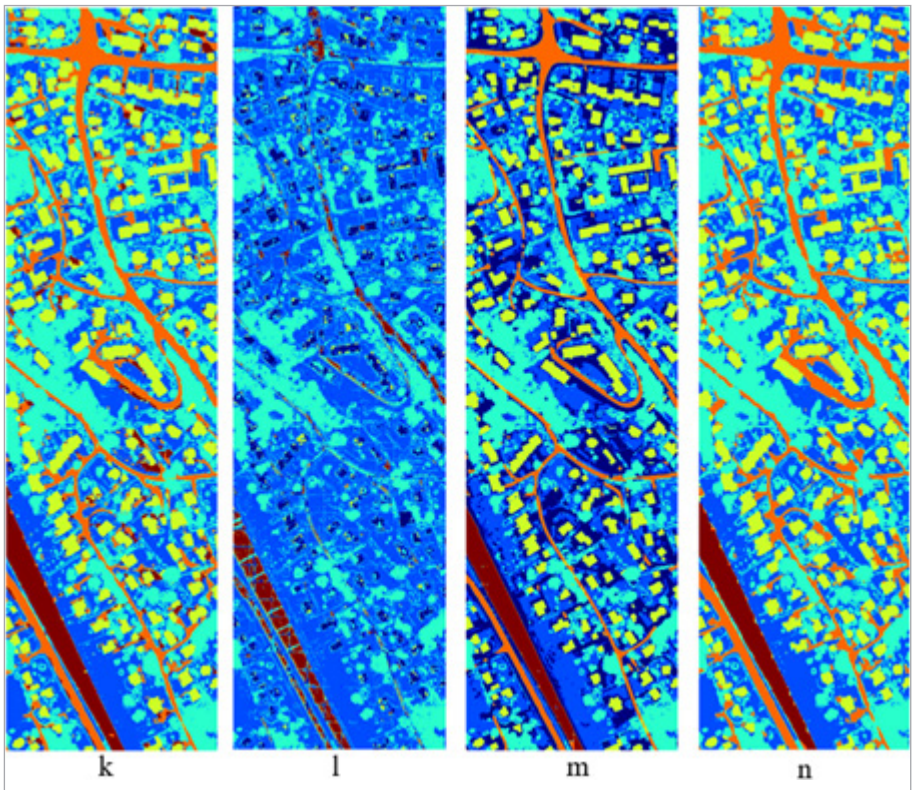

Figure 10. Qualitative comparison of 2D (c–h) and 3D (I–n) convolution-based model in urban land cover classification. k) 3D U-Net Focal–Jaccard Loss, l) 3D ResU-Net CE, m) 3D ResU-Net Focal Loss, n) 3D ResU-Net Focal–Jaccard Loss.

Table 2. Segmentation accuracies obtained by MCC. The grey shaded boxes indicate the highest accuracy score comparing all models for the five classes. Fields shaded in green represent the best accuracy result obtained by comparing loss functions within one model.

	2D U-Net			2D ResU-Net			3D U-Net			3D ResU-Net		
	CE	Focal	Focal–Jaccard	CE	Focal	Focal–Jaccard	CE	Focal	Focal–Jaccard	CE	Focal	Focal–Jaccard
Low vegetation	0.82	0.59	0.64	0.71	0.63	0.62	0.61	0.56	0.63	0.54	0.94	0.61
High vegetation	0.95	0.89	0.91	0.59	0.94	**0.98**	0.88	0.82	0.91	0.82	0.86	**0.98**
Building	0.94	**0.99**	**0.99**	0.89	0.97	**0.99**	0.97	**0.99**	0.98	0.16	0.95	**0.99**
Road	0.90	0.97	0.96	0.83	0.95	0.94	0.96	0.95	**0.98**	0.21	0.92	0.97
Railway	0.95	**0.99**	**0.99**	0.05	0.88	0.95	**0.99**	0.93	**0.99**	0.10	0.98	**0.99**
Overall accuracy (MCC)	**0.84**	0.80	0.82	0.62	0.80	**0.81**	0.78	0.73	**0.80**	0.31	**0.88**	0.84

to identify was the low vegetation class, reaching the accuracy of most models from 54 % to 71 %. However, in the 2D U-Net with CE (Figure 10c) and 3D ResU-Net with Focal Loss (Figure 10m), the low vegetation class was correctly classified, reaching 82 % and 94 %, respectively. Any supervised classification requires ground truth preparation. In our case, dealing with an airborne-based dataset, our ground truths are extracted directly from the image (see above), where low and high vegetation are extracted from the laser features. This way speeds up the labelling process but can cause, in particular, sparse low vegetation pixel labelling as "not present".

All other defined classes belonging to impervious objects were classified correctly, even in 2D convolution-based models. We assume that separating these three classes, such as buildings, roads and railways, was

supported by a preprocessing method for HS and lidar data and a proper selection of lidar features based mainly on the review article from Kuras *et al.*[19]

Loss comparison

Considering the choice of loss functions in each model, Focal and combination of Focal–Jaccard Loss outperform because Focal Loss converges to a lower loss faster than CE, focusing more on classes that are difficult to predict (Figure 5). Measurements on Jaccard distance take more time than Focal Loss and need to be close to the truth before the loss converges. These losses fill each other out and take advantage of both functions: handling extremely imbalanced data and recognising objects.

2D and 3D models

Several important aspects can be noted by visually comparing 2D and 3D models (Figure 10). First, sometimes in 2D convolution-based segmentation models, despite their comparable accuracy score to 3D models, so-called edge effects appear which are the ends of the 64 × 64 patches into which the whole scene has been divided before feeding into the segmentation algorithm (Figure 11).

This problem is well-known in 2D patch-based image segmentation processes and has been reported in several studies.[70,71] In order to avoid additional post-processing steps and the potential generation of other artefacts to the final segmentation map, we compared 3D segmentation models in which the problem of edge effects has been mitigated. The 3D ResU-Net with CE (Figure 10l) is an exception, where the parameters were probably chosen incorrectly. However, since we compare the same 2D and 3D model architectures, we necessarily wanted to present all the results.

Another aspect that needs to be discussed is geometric accuracy. In Figure 10 and Figure 11, we can see (especially in the building class) that 2D and 3D models do not perform well with objects with irregular shapes. Often, we can see inaccurate blurred contours and, in some places, complete disruption of the geometry. However, in the 2D U-Net with CE and the 3D ResU-Net with Focal Loss, the geometric of identified objects/surfaces are adequate or close to reality and ground truth.

It must be noted that an important issue appeared in the attempt to classify the "unknown" class. Only in the case of the 2D U-Net with CE and the 3D ResU-Net with Focal Loss undefined pixels (no ground truth) were correctly identified as "unknown", while the rest of the segmentation models assigned a class to each pixel, creating many false positives. Identification of unknown pixels is crucial, especially in a heterogeneous complex urban environment where our selected scene consists of more than the five defined classes. This "unknown" class has been assigned to those pixels where no defined class has reached 50 % in the prediction.

Furthermore, the proper input preparation and the reduction of dimensions through endmember extraction allowed us to skip statistical transformation methods, such as Principal Component Analysis (PCA). Due to its purely statistical nature, PCA loses spectral information since the spectral positions are not considered in the calculations.[72] Hence the extracted (spectral) endmembers can be used for further analysis or to generate local spectral libraries.

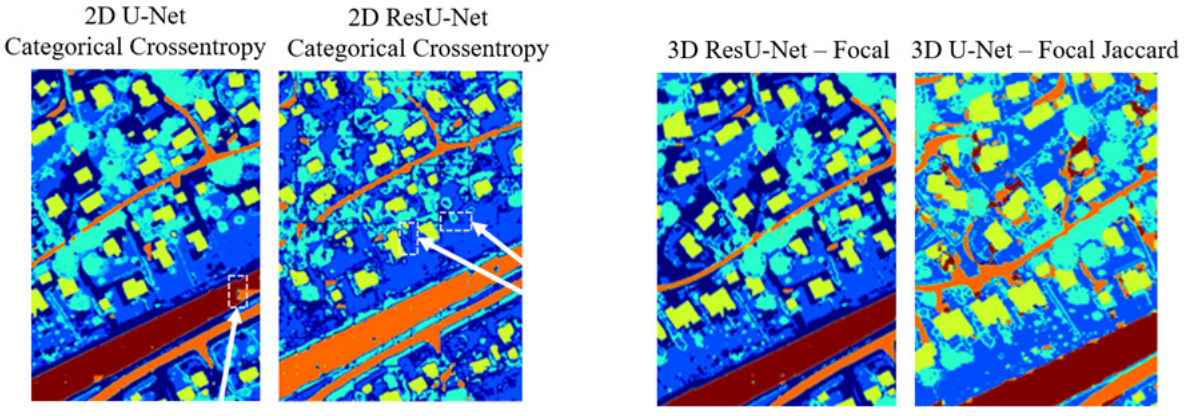

Figure 11. 2D and 3D model segmentation map clippings. In both 2D models, we can see edge effects that have been compensated for in 3D models. White arrows and dashed rectangles point to the exact positions of edge effects.

The last but not least aspect relates to time consumption comparing 2D and 3D segmentation models. Due to the requirements of advanced calculations in 3D convolutional operations, 3D models took about 20 times longer than 2D models. For this reason, with a simple segmentation of high-dimensional HS and lidar data, simple 2D models seem to be sufficient. However, we are convinced that analysis based on much more advanced segmentation, intraclass segmentation, as well as identification of materials and their properties in an urban environment will require the choice of 3D convolutional operations despite the time consumption.

Conclusion and future perspectives

This paper explores diverse deep learning models with different parameters and convolutional dimensions for urban land cover classification using fused HS and lidar data. In particular, four deep learning models have been compared, 2D-U-Net (Figure 10c–e), 2D-ResU-Net (Figure 10f–h), 3D-U-Net (Figure 10i–k) and 3D-ResU-Net (Figure 10l–n). All models have been trained with three different loss functions: state-of-art CE, Focal Loss, and a combination of Focal Loss and Jaccard Loss (Focal–Jaccard Loss).

As a whole, we can conclude that segmentation of both U-Net and ResU-Net performed very well, especially implementing Focal and Focal–Jaccard Loss functions. However, an important factor to consider is the choice of the model architecture and the proper selection of parameters. Despite the out-performance of the 3D U-Net and Res-U-Net models, the training often requires more training data than analysing shallow machine learning or deep 2D convolution-based algorithms. Alternatively, to save time, reduce computational resources and limit the need for a large amount of training data, we will explore the potential of hybrid 2D–3D models and carry out analyses based on ensemble learning where several models are combined. Thus, it is possible to reduce the number of parameters in the deep learning model and, at the same time, use the information from the 3D convolution-based algorithm.

Another important factor to consider when working with image segmentation on HS and lidar data is the correct labelling of the ground truth. Due to the fuzzy class assignment of some pixels responding from semi-automatic label preparation, the network may try to learn incorrect patterns. Further, a single pixel can contain more than one class, but the model assumes that the pixel contains only a single class, which also hinders the learning process. Hence, in future research, we will focus on unsupervised segmentation to avoid other errors in creating ground truths and develop a method to identify unknown classes, such as testing open set recognition for deep learning algorithms, e.g. for change detection based on HL-Fusion data.

Funding

This work is part of the project “FKB maskinlæring” funded by RFF “Oslo og Akershus Regionale forskningsfond”.

References

1. X.-P. Song, J.O. Sexton, C. Huang, S. Channan and J.R. Townshend, “Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover”, *Remote Sens. Environ.* **175,** 1–13 (2016). https://doi.org/10.1016/j.rse.2015.12.027
2. J.A. Hepinstall, M. Alberti and J.M. Marzluff, “Predicting land cover change and avian community responses in rapidly urbanizing environments”, *Landsc. Ecol.* **23,** 1257–1276 (2008). https://doi.org/10.1007/s10980-008-9296-6
3. Y.B. Chen, H.L. Zhou, H. Zhang, G.M. Du and J.H. Zhou, “Urban flood risk warning under rapid urbanization”, *Environ. Res.* **139,** 3–10 (2015). https://doi.org/10.1016/j.envres.2015.02.028
4. F. Chen, H. Kusaka, R. Bornstein, J. Ching, C.S.B. Grimmond, S. Grossman-Clarke, T. Loridan, K.W. Manning, A. Martilli, S. Miao, D. Sailor, F.P. Salamanca, H. Taha, M. Tewari, X. Wang, A.A. Wyszogrodzki and C. Zhang, “The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems”, *Int. J. Climatol.* **31(2),** 273–288 (2011). https://doi.org/10.1002/joc.2158
5. J.H. Lee and K.B. Woong, “Characterization of urban stormwater runoff”, *Water Res.* **34(6),** 1773–1780 (2000). https://doi.org/10.1016/S0043-1354(99)00325-5
6. Y. Qian, W. Zhou, W. Yu and S.T.A. Pickett, “Quantifying spatiotemporal pattern of urban

greenspace: new insights from high resolution data", *Landsc. Ecol.* **30,** 1165–1173 (2015). https://doi.org/10.1007/s10980-015-0195-3

7. M. Herold, M.E. Gardner and D.A. Roberts, "Spectral resolution requirements for mapping urban areas", *IEEE Trans. Geosci. Remote Sens.* **41(9),** 1907–1919 (2003). https://doi.org/10.1109/TGRS.2003.815238
8. S. Roessner, K. Segl, U. Heiden and H. Kaufmann, "Automated differentiation of urban surfaces based on airborne hyperspectral imagery", *IEEE Trans. Geosci. Remote Sens.* **39(7),** 1525–1532 (2001). https://doi.org/10.1109/36.934082
9. Y.-C. Lin, C. Lin, M.-D. Tsai and C.-L. Lin, "Object-based analysis of LiDAR geometric features for vegetation detection in shaded area", *XXIII ISPRS Congress*. Prague, Czech Republic (2016). https://doi.org/10.5194/isprsarchives-XLI-B1-43-2016
10. M. Rutzinger, B. Höfle and N. Pfeifer, "Detection of high urban vegetation with airborne laser scanning data", *Forestsast*. Montpellier, France (2007).
11. M.K. Ridd, "Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities", *Int. J. Remote Sens.* **16,** 2165–2185 (1995). https://doi.org/10.1080/01431169508954549
12. M. Alonzo, B. Bookhagen, J.P. McFadden, A. Sun and D.A. Roberts, "Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry", *Remote Sens. Environ.* **162,** 141–153 (2015). https://doi.org/10.1016/j.rse.2015.02.025
13. C. Jürgens, "Urban and suburban growth assessment with remote sensing", *OICC 7th International Seminar on GIS Applications in Planning and Sustainable Development*. 13–15 February, Cairo, Egypt (2001).
14. M. Batty and P. Longley, *Fractal Cities: A Geometry of Form and Function*. Academic Press, London and San Diego (1994).
15. M.A. Stretton, W. Morrison, R.J. Hogan and S. Grimmond, "Evaluation of the SPARTACUS-urban radiation model for vertically resolved shortwave radiation in urban areas", *Bound.-Lay. Meteorol.* **184,** 301–331 (2022). https://doi.org/10.1007/s10546-022-00706-9
16. V.S. Saravanarajan, R.C. Chen and L.S. Chen, "Geometric feature learning network for detecting the objects in urban streets", *SSRN*, preprint (2022). https://doi.org/10.2139/ssrn.4218483
17. E. Ben-Dor, N. Levin and H. Saaroni, "A spectral based recognition of the urban environment using the visible and near-infrared spectral region (0.4–1.1 µm). A case study over Tel-Aviv, Israel", *Int. J. Remote Sens.* **22(11),** 2193–2218 (2001). https://doi.org/10.1080/01431160117759
18. P. Gamba, F. Dell'Acqua and B.V. Dasarathy, "Urban remote sensing using multiple data sets: Past, present, and future", *Inf. Fusion* **6,** 319–326 (2005). https://doi.org/10.1016/j.inffus.2005.02.007
19. A. Kuras, M. Brell, J. Rizzi and I. Burud, "Hyperspectral and lidar data applied to the urban land cover machine learning and neural-network-based classification: a review", *Remote Sens.* **13(17),** 3393 (2021). https://doi.org/10.3390/rs13173393
20. R. Senchuri, A. Kuras and I. Burud, "Machine learning methods for road edge detection on fused airborne hyperspectral and lidar data", *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2021*. pp. 1–5 (2021). https://doi.org/10.1109/WHISPERS52202.2021.9484007
21. F. Priem and F. Canters, "Synergistic use of LiDAR and APEX hyperspectral data for high-resolution urban land cover mapping", *Remote Sens.* **8(10),** 787 (2016). https://doi.org/10.3390/rs8100787
22. M. Alonzo, B. Bookhagen and D.A. Roberts, "Urban tree species mapping using hyperspectral and lidar data fusion", *Remote Sens. Environ.* **148,** 70–83 (2014). https://doi.org/10.1016/j.rse.2014.03.018
23. D. Borghys, M. Shimoni and C. Perneel, "Change detection in urban scenes by fusion of SAR and hyperspectral data", *Proc. SPIE 6749, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology VII* 67490R (2007). https://doi.org/10.1117/12.738767
24. K. Segl, S. Roessner, U. Heiden and H. Kaufmann, "Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data", *ISPRS J. Photogramm. Remote Sens.* **58(1–2),** 99–112 (2003). https://doi.org/10.1016/S0924-2716(03)00020-0
25. J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches", *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **5(2),** 354–379 (2012). https://doi.org/10.1109/JSTARS.2012.2194696

26. G. Hegde, J.M. Ahamed, R. Hebbar and U. Raj, "Urban land cover classification using hyperspectral data", *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XL-8,** 751–754 (2014). https://doi.org/10.5194/isprsarchives-XL-8-751-2014
27. S. Amini, S. Homayouni and A. Safari, "Semi-supervised classification of hyperspectral image using random forest algorithm", *2014 IEEE Geoscience and Remote Sensing Symposium*. pp. 2866–2869 (2014). https://doi.org/10.1109/IGARSS.2014.6947074
28. X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery", *IEEE Trans. Geosci. Remote Sens.* **51(1),** 257–272 (2013). https://doi.org/10.1109/TGRS.2012.2202912
29. A. Serna and B. Marcotegui, "Urban accessibility diagnosis from mobile laser scanning data", *ISPRS J. Photogramm. Remote Sens.* **84,** 23–32 (2013). https://doi.org/10.1016/j.isprsjprs.2013.07.001
30. M. Dalponte, L. Bruzzone and D. Gianelle, "Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas", *IEEE Trans. Geosci. Remote Sens.* **46(5),** 1416–1427 (2008). https://doi.org/10.1109/TGRS.2008.916480
31. M. Brell, K. Segl, L. Guanter and B. Bookhagen, "Hyperspectral and Lidar intensity data fusion: a framework for the rigorous correction of illumination, anisotropic effects, and cross calibration", *IEEE Trans. Geosci. Remote Sens.* **55(5),** 2799–2810 (2017). https://doi.org/10.1109/TGRS.2017.2654516
32. M. Brell, C. Rogass, K. Segl, B. Bookhagen and L. Guanter, "Improving sensor fusion: a parametric method for the geometric coalignment of airborne hyperspectral and LiDAR data", *IEEE Trans. Geosci. Remote Sens.* **54(6),** 3460–3474 (2016). https://doi.org/10.1109/TGRS.2016.2518930
33. C. Ge, Q. Du, W. Sun, K. Wang, J. Li and Y. Li, "Deep residual network-based fusion framework for hyperspectral and LiDAR data", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14,** 2458–2472 (2021). https://doi.org/10.1109/JSTARS.2021.3054392
34. D. Hong, L. Gao, R. Hang, B. Zhang and J. Chanussot, "Deep encoder-decoder networks for classification of hyperspectral and LiDAR data", *IEEE Geosci. Remote Sens. Lett.* **99,** 1–5 (2020). https://doi.org/10.1109/LGRS.2020.3017414
35. R. Hang, Y. Li, P. Ghamisi, D. Hong, G. Xia and Q. Liu, "Classification of hyperspectral and LiDAR data using coupled CNNs", *IEEE Trans. Geosci. Remote Sens.* **68(7),** 4939–4950 (2020). https://doi.org/10.1109/TGRS.2020.2969024
36. S.Y. Sadjadi and S. Parsian, "Combining hyperspectral and LiDAR data for building extraction using machine learning technique", *Int. J. Comput.* **2,** 88–93 (2017).
37. Q. Feng, D. Zhu, J. Yang and B. Li, "Multisource hyperspectral and LiDAR data fusion for urban land-use mapping based on a modified two-branch convolutional neural network", *ISPRS Int. J. Geoinf.* **8(1),** 28 (2019). https://doi.org/10.3390/ijgi8010028
38. H. Li, P. Ghamisi, U. Soergel and X.X. Zhu, "Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks", *Remote Sens.* **10(10),** 1649 (2018). https://doi.org/10.3390/rs10101649
39. H. Torabzadeh, F. Morsdorf and M.E. Schaepman, "Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems", *ISPRS J. Photogramm. Remote Sens.* **97,** 25–35 (2014). https://doi.org/10.1016/j.isprsjprs.2014.08.001
40. J. Yin, C. Qi, Q. Chen and J. Qu, "Spatial-spectral network for hyperspectral image classification: a 3-D CNN and Bi-LSTM framework", *Remote Sens.* **13(12),** 2353 (2021). https://doi.org/10.3390/rs13122353
41. H. Gong, Q. Li, C. Li, H. Dai, Z. He, W. Wang, H. Li, F. Han, A. Tuniyazi and T. Mu, "Multiscale information fusion for hyperspectral image classification based on hybrid 2D-3D CNN", *Remote Sens.* **13(12),** 2268 (2021). https://doi.org/10.3390/rs13122268
42. Z. Zhang, Q. Liu and Y. Wang, "Road extraction by deep residual U-net", *IEEE Geosci. Remote Sens. Lett.* **15(5),** 749–753 (2018). https://doi.org/10.1109/LGRS.2018.2802944
43. C. Shi and C.M. Pun, "Superpixel-based 3D deep neural networks for hyperspectral image classification", *Pattern Recognit.* **74,** 600–616 (2018). https://doi.org/10.1016/j.patcog.2017.09.007
44. C. Yu, R. Han, M. Song, C. Liu and C.I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13,** 2485–2501 (2020). https://doi.org/10.1109/JSTARS.2020.2983224

45. M. Rao, P. Tang and Z. Zhang, "A developed Siamese CNN with 3D adaptive spatial-spectral pyramid pooling for hyperspectral image classification", *Remote Sens.* **12(12),** 1964 (2020). https://doi.org/10.3390/rs12121964
46. Q. Li, Q. Wang and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution", *Remote Sens.* **12(10),** 1660 (2020). https://doi.org/10.3390/rs12101660
47. M.K.K. Singh, S. Mohan and B. Kumar, "Fusion of hyperspectral and LiDAR data using sparse stacked autoencoder for land cover classification with 3D-2D convolutional neural network", *J. Appl. Remote Sens.* **16(3),** 034523 (2022). https://doi.org/10.1117/1.JRS.16.034523
48. O. Ronneberger, P. Fischer and T. Brox, "U-Net: convolutional networks for biomedical image segmentation", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Ed by N. Navab, J. Hornegger, W. Wells and A. Frangi. Lecture Notes in Computer Science. Springer, Vol. 9351, pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
49. K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
50. K. He and J. Sun, "Convolutional neural networks at constrained time cost", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 5353–5360 (2015). https://doi.org/10.1109/CVPR.2015.7299173
51. J. Ma, "Segmentation loss odyssey", *arxiv:2005* 13449v1 (2020). https://doi.org/10.48550/arXiv.2005.13449
52. D. Duque-Arias, S. Valesco-Forero, J.E. Deschaud, F. Goulette, A. Serna, E. Decenciere and B. Marcotegui, "On power Jaccard losses for semantic segmentation", *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Vol. 5, pp. 561–568 (2021). https://doi.org/10.5220/0010304005610568
53. T.Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, "Focal loss for dense object detection", *2017 IEEE International Conference on Computer Vision (ICCV)*. pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324
54. M. Hasanlou and S.T. Seydi, "Hyperspectral change detection: an experimental comparative study", *Int. J. Remote Sens.* **39(20),** 7029–7083 (2018). https://doi.org/10.1080/01431161.2018.1466079
55. A. Ertürk, M.K. Güllü, D. Çeşmeci, D. Gerçek and S. Ertürk, "Spatial resolution enhancement of hyperspectral images using unmixing and binary particle swarm optimization", *IEEE Geosci. Remote Sens. Lett.* **11(12),** 2100–2104 (2014). https://doi.org/10.1109/LGRS.2014.2320135
56. A. Ertürk, M.-D. Iordache and A. Plaza, "Sparse unmixing with dictionary pruning for hyperspectral change detection", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **20(1),** 321–330 (2017). https://doi.org/10.1109/JSTARS.2016.2606514
57. S. Liu, L. Bruzzone, F. Bovolo and D. Peijun, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images", *IEEE Trans. Geosci. Remote Sens.* **54(5),** 2733–2748 (2016). https://doi.org/10.1109/TGRS.2015.2505183
58. Y. Fang, Y. Wang, L. Xu, R. Zhuo, A. Wong and D.A. Clausi, "BCUN: Bayesian fully convolutional neural network for hyperspectral spectral unmixing", *IEEE Trans. Geosci. Remote Sens.* **60,** 5523714 (2022). https://doi.org/10.1109/TGRS.2022.3151004
59. C.V.S.S.M. Kumar, S.S. Jha, R.R. Nidamanuri and B.K. Dadhwal, "Benchmark studies on pixel-level spectral unmixing of multi-resolution hyperspectral imagery", *Int. J. Remote Sens.* **43(4),** 1451–1484 (2022). https://doi.org/10.1080/01431161.2022.2040755
60. A. Gomez-Sanchez, M. Marro,M. Marsal, S. Zacchetti, R.R. de Oliveira, P. Loza-Alvarez and A. de Juan, "Linear unmixing protocol for hyperspectral image fusion analysis applied to a case study of vegetal tissues", *Sci. Rep.* **11,** 18665 (2021). https://doi.org/10.1038/s41598-021-98000-0
61. G.F. Hughes, "On the mean accuracy of statistical pattern recognizers", *IEEE Trans. Inf. Theory* **14,** 55–63 (1968). https://doi.org/10.1109/TIT.1968.1054102
62. M.E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data", *Proc. SPIE 3753, Imaging Spectrometry V* **3753,** 266–275 (1999). https://doi.org/10.1117/12.366289
63. R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm", *J. Chemometr.* **11(5),** 393–401 (1997). https://doi.org/10.1002/

(SICI)1099-128X(199709/10)11:5<393::AID-CEM483>3.0.CO;2-L

64. D. Schläpfer and R. Richter, "Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: Parametric orthorectification", *Int. J. Remote Sens.* **23(13),** 2609–2630 (2002). https://doi.org/10.1080/01431160110115825
65. C. Ruffin and R.L. King, "The analysis of hyperspectral data using Savitzky-Golay filtering-Theoretical basis (part 1)", *IEEE 1999 International Geoscience and Remote Sensing Symposium IGARSS'99* **2,** 756–758 (1999). https://doi.org/10.1109/IGARSS.1999.774430
66. P.S. Thenkabail, R.B. Smith and E. De Pauw, "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics", *Remote Sens. Environ.* **71(2),** 158–182 (2000). https://doi.org/10.1016/S0034-4257(99)00067-X
67. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, "Tensorflow: large-scale machine learning on heterogeneous distributed systems", arXiv:160304467 (2016). https://doi.org/10.48550/arXiv.1603.04467
68. E. Adam, O. Mutanga and D. Rugege, "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review", *Wetl. Ecol. Manag.* **18,** 281–296 (2010). https://doi.org/10.1007/s11273-009-9169-z
69. B.W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme", *Biochim. Biophys. Acta-Protein Structure* **405(2),** 442–451 (1975). https://doi.org/10.1016/0005-2795(75)90109-9
70. N. Pielawski and C. Wählby, "Introducing Hann windows for reducing edge-effects in patch-based image segmentation", *PLoS One* **15(3),** e0229839 (2020). https://doi.org/10.1371/journal.pone.0229839
71. N.G. Kurale and M. Vaidya, "Retinal hemorrhage detection using splat segmentation of retinal fundus images", *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pp. 364–375 (2017). https://doi.org/10.1109/ICCUBEA.2017.8463939
72. H.A. Khan, M.M. Khan, K. Khurshid and J. Chanussot, "Saliency based visualization of hyperspectral images", *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1096–1099 (2015). https://doi.org/10.1109/IGARSS.2015.7325961