Wide-Area Land Cover Mapping With Sentinel-1 Imagery Using Deep Learning Semantic Segmentation Models

Sanja Šćepanović ⓘ, Oleg Antropov ⓘ, *Member, IEEE*, Pekka Laurila, Yrjo Rauste, Vladimir Ignatenko, and Jaan Praks ⓘ, *Member, IEEE*

Abstract—Land cover (LC) mapping is essential for monitoring the environment and understanding the effects of human activities on it. Recent studies demonstrated successful applications of specific deep learning models to small-scale LC mapping tasks (e.g., wetland mapping). However, it is not readily clear which of the existing state-of-the-art models for natural images are the best candidates to be taken for the particular remote sensing task and data. In this article, we answer that question for mapping the fundamental LC classes using the satellite imaging radar data. We took ESA Sentinel-1 C-band SAR images acquired during the whole summer season of 2018 in Finland, which are representative of the land cover in the country. CORINE LC map was used as a reference, and the models were trained to distinguish between the five major CORINE-based classes. We selected seven among the state-of-the-art semantic segmentation models so that they cover a diverse set of approaches: U-Net, DeepLabV3+, PSPNet, BiSeNet, SegNet, FC-DenseNet, and FRRN-B, and further fine-tuned them. Upon evaluation and benchmarking, all the models demonstrated solid performance with overall accuracy between 87.9% and 93.1%, with good to a very good agreement (Kappa statistic between 0.75 and 0.86). The two best models were fully convolutional DenseNets (FC-DenseNet) and SegNet (encoder-decoder-skip), with the latter having a much shorter inference time. Overall, our results indicate that the semantic segmentation models are suitable for efficient wide-area mapping using satellite SAR imagery and provide baseline accuracy against which the newly proposed models should be evaluated.

Index Terms—C-band, CORINE, deep learning, image classification, land cover (LC) mapping, semantic segmentation, sentinel-1 data, synthetic aperture radar (SAR).

I. INTRODUCTION

MAPPING of land cover (LC) and its change has a critical role in the characterization of the current state of the

environment. The changes in LC can be due either to human activities as well as caused by climate changes on a regional scale. The LC, on the other hand, affects climate through water and energy exchange with the atmosphere and by changing carbon balance. Because of this, LC belongs to the essential climate variables [1]. Hence, timely assessment of LC and its change is one of the most important applications in satellite remote sensing. Thematic maps are needed annually for various purposes in medium resolution (circa 250 m) with less than 15% measurement uncertainty, and in high resolution (10–30 m) with less than 5% uncertainty.

CORINE LC (CLC) is a notable example of a consistent Pan-European LC mapping initiative [2], [3] coordinated by the European Environment Agency (EEA).[1] CORINE stands for *coordination of information on the environment*. It is an ongoing long-term effort providing harmonized land cover data in Europe with updates approximately every four years. The CORINE maps are an important source of LC information suitable for operational purposes also for various customer groups in Europe. It has altogether 44 classes, though many of them are not strictly ecological classes but rather land use classes. On the continental scale, CORINE provides a harmonized map with a 25-ha minimum mapping unit (MMU) for areal phenomena, and a minimum width of 100 m for linear phenomena [4]. National LC maps in the CORINE framework can exhibit smaller mapping units. In Finland, the latest revision of the CORINE LC map at the time of this study was the year 2018 version produced by the Finnish Environment Institute. The map has an MMU of 20×20 m and was produced by a combined automated and manual interpretation of the high-resolution satellite optical data followed by the data integration with existing basic map layers [5].

The state-of-the-art approaches used for LC mapping mainly rely on satellite optical imagery. The key role is played by the Landsat imagery often augmented by the MODIS or SPOT-5 imagery [6]–[8]. Other sources of information employed for LC mapping include digital elevation models (DEM) and very high-resolution imagery [9]. When it comes to large-scale and multitemporal LC mapping, a more recent optical imagery

---

[1][Online]. Available: https://land.copernicus.eu/pan-european/corine-land-cover

source is Copernicus Sentinel-2. With a revisit of five days, it has become another key data source [10].

International programs, such as the European Space Agency's (ESA's) Copernicus [11] behind the Sentinel satellites are taking significant efforts to make Earth observation (EO) data freely available for commercial and noncommercial purposes. The Copernicus programme is a multibillion investment by the EU and ESA aiming to provide essential services based on accurate and timely data from satellites. Its main goals are to improve the ways of managing the environment, to help mitigate the effects of climate change, and enable the creation of new applications and services, such as environmental monitoring and urban development.

The provision of free satellite data for mapping in the framework of such programs also enables the application of methods that could not be used earlier because they require vast and representative datasets for training, for example, *deep learning*. In recent years, deep learning has brought about several breakthroughs in pattern recognition and computer vision [12]–[14]. The success of the deep learning models can be attributed to both their *deep multilayer structure* creating nonlinear functions and, hence, allowing extraction of hierarchical sets of features from the data, and to their *end-to-end training scheme* allowing for simultaneous learning of the features from the raw input and predicting the task at hand. In this way, the heuristic feature design is removed. This is advantageous compared to the traditional machine learning methods [e.g., support vector machine (SVM) and random forest (RF)], which require a multistage feature engineering procedure. In deep learning, such a procedure is replaced with a simple end-to-end deep learning workflow. One of the key requirements for the successful application of deep learning methods is a large amount of data available from which the model can automatically learn the representative features for the prediction task [15]. The availability of open satellite imagery, such as from Copernicus, offers just that.

The LC mapping systems based solely on optical imagery suffer from issues with cloud cover and weather conditions, especially in the tropical areas, and with a lack of illumination in the polar regions. Among the free satellite data offered by the Copernicus programme are *synthetic aperture radar (SAR)* images from the Sentinel-1 satellites. SAR is an active radar imaging technique that does not require illumination and is not hampered by cloud cover due to penetration of microwave radiation through clouds. The utilization of SAR imagery, hence, would allow mapping such challenging regions and increasing the mapping frequency in the orchestrated efforts like CORINE. One of the significant issues previously was the absence of timely and consistent high-resolution wide-area SAR coverage. With the advent of Copernicus Sentinel-1 satellites, operational use of imaging radar data becomes feasible for consistent wide-area mapping. The first Copernicus Sentinel-1 mission was launched in April 2014. First, Sentinel-1 A alone was capable of providing C-band SAR data in up to four imaging modes with a revisit time of 12 days. Once Sentinel-1B was launched in 2016 the revisit time has reduced to six days [11].

We studied wide-area SAR-based LC mapping by methodologically combining the two discussed recent advances: The improved methods for large-scale image processing using deep learning and the availability of SAR imagery from the Sentinel-1 satellites.

A. Land Cover Mapping With SAR Imagery

While using optical satellite data is still a mainstream in LC and LC change mapping [5], [16]–[19], SAR data has been getting more attention as more suitable sensors appear. To date, several studies have investigated the suitability of SAR for LC mapping, focusing primarily on L-band, C-band, and X-band polarimetric [20], [21], multitemporal, and multifrequency SAR [22] [23], as well as, on the combined use of SAR and optical data [24]–[28].

Independently of the imagery used, the majority of LC mapping methods so far are based on traditional supervised classification techniques [29]. Widely used classifiers are support vector machines (SVM), decision trees, random forests (RF), and maximum likelihood classifiers (MLC) [7], [9], [22], [29]. However, extracting a large number of features needed for classification, i.e., the *feature engineering process*, is time intensive, and requires lots of expert work in developing and fine-tuning classification approaches. This limits the applications of the traditional supervised classification methods on a large scale.

Backscattered microwave radiation is composed of multiple fundamental scattering mechanisms determined by the vegetation water content, surface roughness, soil moisture, horizontal and vertical structure of the scatterers, as well as imaging geometry during the datatake. Accordingly, a considerable number of classes can be differentiated in SAR images [20], [30]. However, majority of SAR classification algorithms use fixed SAR observables (e.g., polarimetric features) to infer specific LC classes, despite the large temporal, seasonal, and environmental variability between different geographical sites. This leads to a lack of generalization capability and a need to use extensive and representative reference data and SAR data. The latter means the need to account for not only all variation of SAR signatures for a specific class but also the need to consider seasonal effects, as changes in moisture of soil and vegetation, as well as frozen state of land [31] that strongly affect SAR backscatter. On the other hand, when using multitemporal approaches, such seasonal variation can be used as an effective discriminator among different LC classes.

When exclusively using SAR data for LC mapping, reported accuracy often turns out to be relatively low for operational LC mapping and change monitoring. Methodologically, reported solutions utilized supervised approaches, linking SAR observables and class labels to pixels, superpixels, or objects in parametric or nonparametric manner [19], [20], [30], [32]–[37].

However, tackling relatively large number of classes was considered only in several studies, often with relatively low reported accuracies. For instance, in [38] it was found that P-band PolSAR imagery was unsatisfactory for mapping more than five classes with the iterated conditional mode contextual classifier applied to several polarimetric parameters. They achieved a Kappa value of 76.8% when mapping four classes. Classification performance of the L-band ALOS PALSAR and C-band RADARSAT-2 images was compared in the moist

tropics [39]. L-band provided 72.2% classification accuracy for a coarse LC classification system and C-band only 54.7%. In a similar study in Lao PDR, ALOS PALSAR data were found to be mostly useful as a back-up option to optical ALOS AVNIR data [19]. Multitemporal Radarsat-1 data with HH polarization and ENVISAT ASAR data with VV polarization (both C-band) were studied for classification of five LC classes in Korea with moderate accuracy [40]. Waske *et al.* [22] applied boosted decision tree and RF to multitemporal C-band SAR data reaching accuracy up to 84%. Several studies [41], [20] investigated specifically SAR suitability for the boreal zone, with reported accuracy up to 83% depending on the classification technique (maximum likelihood, probabilistic neural networks, etc.) when five super-classes (based on CORINE data) were used.

The potential of Sentinel-1 imagery for CORINE-type thematic mapping was assessed in a study that used Sentinel-1 A data for mapping class composition in Thuringia [30]. Long-time series of Sentinel-1 SAR data are considered especially suitable for crop type mapping [42]–[45], with increased number of studies attempting LC mapping in general [46], [47].

Moreover, as Sentinel-1 data are presently the only free source of SAR data routinely available for wide-area mapping at no cost for users, it seems to be the best candidate data for development and testing of improved classification approaches. Previous studies indicate a necessity for developing and testing new methodological approaches that can be effectively applied to a large-scale and deal with the variability of SAR observables concerning ecological LC classes. We suggest adopting state-of-the-art deep learning approaches for this purpose.

B. Deep Learning in Remote Sensing

The advances in the deep learning techniques for computer vision, in particular, convolutional neural networks (CNNs) [12], [48], have led to the application of deep learning in several domains that rely on computer vision. Examples are self-driving cars, image search engines, medical diagnostics, and augmented reality. Deep learning approaches are becoming extensively applied in the remote sensing domain, as well.

Zhu *et al.* [49] provided a discussion on the specificities of remote sensing imagery (compared to ordinary RGB images) that result in specific deep learning challenges in this area. For example, remote sensing data are georeferenced, often multimodal, with particular imaging geometries, there are interpretation difficulties, and the ground-truth or labeled data needed for deep learning is still often lacking. Additionally, most of the state-of-the-art CNNs are developed for three-channel input images (i.e., RGB) and so certain adaptations are needed to apply them on the remote sensing data [50].

Nevertheless, several research papers tackling remote sensing imagery with deep learning techniques were published in recent years. Zhang *et al.* [51] review the field and find applications to image preprocessing [52], target recognition [53], classification [54], [55], and semantic feature extraction and scene understanding [56]. The deep learning approaches are found to outperform standard methods applied up to several years ago, i.e., SVMs and RFs [57].

When it comes to deep learning for *LC* or *land use mapping*, applications have been limited to optical satellite [50], [54], [58] or aerial [59] imagery, and hyperspectral imagery [55], [58] owing to the similarity of these images to ordinary RGB images studied in computer vision [50].

When it comes to SAR images, Zhang *et al.* [51] found that there is already a significant success in applying deep learning techniques for object detection and scene understanding. However, for classification on SAR data, applications are scarce and advances are yet to be achieved [51]. Published research includes deep learning for crop types mapping using combined optical and SAR imagery [57], as well as the use of SAR images exclusively [60]. However, those methods applied deep learning only to some part of the task at hand and not in an end-to-end fashion. Wang *et al.* [54], for instance, just used deep neural networks for merging oversegmented elements, which are produced using traditional segmentation approaches. Similarly, Tuia *et al.* [55] applied deep learning to extract hierarchical features, which they further fed into a multiclass logistic classifier. Duan *et al.* [60] used first unsupervised deep learning and then continued with a couple of supervised labeling tasks. Chen *et al.* [58] applied a deep learning technique (stacked autoencoders) to discover the features, but then they still used traditional machine learning (SVM, logistic regression) for the image segmentation. Unlike those methods, we applied the deep learning in an *end-to-end fashion*, i.e., from supervised feature extraction to the land class prediction. This makes our approach more flexible, robust, and adaptable to the SAR data from new regions, as well as more efficient.

When it comes to the *end-to-end* approaches for SAR classification, there are several studies where the focus was on a small area and a specific LC mapping task. For instance, Mohammadimanesh *et al.* [61] used fully polarimetric SAR (PolSAR) imagery from RADARSAT-2 to classify wetland complexes, for which they have developed a specifically tailored semantic segmentation model. However, the authors have tackled a small test area (around 10×10 km) and have not explored how their model generalizes to other types of areas. Similarly, Wang *et al.* [62] adapted existing CNN models into a fixed-feature-size CNN that they have evaluated on a *small-scale* RADARSAT-2 or AIRSAR (i.e., airborne SAR data). In both cases, they have used more advanced fully polarimetric SAR imagery at better resolution as opposed to Sentinel-1, which means the imagery with more input information to the deep learning models. Importantly, it is only Sentinel-1 that offers open operational data with up to every six days repeat. Because of this, the discussed approaches developed and tested specifically for PolSAR imagery at a higher resolution cannot be considered applicable for a wide-area mapping, yet. Similarly, Ahishali *et al.* [63] applied end-to-end approaches to SAR data. They have also worked with single polarized COSMO-SkyMed imagery. However, all the imagery they considered was X-band SAR contrary to the C-band imagery we use here and again only on a small scale. The authors proposed a compact CNN model that they found had outperformed some of the off-the-shelf CNN methods, such as Xception and Inception-ResNet-v2. It is important to note that compared to those, the off-the-shelf models that we consider

here are more sophisticated semantic segmentation models, some of which employ Xception or ResNet but only as a module in their feature extraction parts.

In summary, the capabilities of the deep learning approaches for the classification have been investigated to a lesser extent for SAR imagery than for optical imagery. The attempts to use SAR data for LC classification were relatively limited in scope, area, or the number of used SAR scenes. Particularly, wide-area LC mapping was never addressed. The reasons for this include comparatively poor availability of SAR data compared to optical (greatly changed since the advent of Sentinel-1), complex scattering mechanisms leading to ambiguous SAR signatures for different classes (which makes SAR image segmentation more difficult than the optical image segmentation [64]), as well as the speckle noise caused by the coherent nature of the SAR imaging process.

C. Study Goals

Present study addresses the identified research gap of a lack of wide-area land cover mapping using SAR data. We achieve this by training, fine-tuning, and evaluating a set of suitable state-of-the-art deep learning models from the class of semantic segmentation models, and demonstrating their *suitability for land cover mapping*. Moreover, our work is the first to examine and demonstrate the suitability of deep learning for LC mapping from SAR images on a *large-scale, i.e., across the whole country*.

Specifically, we applied the semantic segmentation models on the SAR images taken over Finland. We focused on the images of Finland because there is the LC mask of a suitable resolution that can be used for training labels (i.e., CORINE). The training is performed with the seven selected models (Seg-Net [65], PSPNet [66], BiSeNet [67], DeepLabV3+ [68], [69], U-Net [70], [71], FRRN-B [72], and FC-DenseNet [73]), which have encoder modules *pretrained* on the large RGB image corpus ImageNet 2012.[2] Those models are freely available.[3] In other words, we reused semantic segmentation architectures developed for natural images with pretrained weights on RGB images and we fine-tuned them on the SAR images. Our results (with over 90% overall accuracy) demonstrate the effectiveness of the deep learning methods for the LC mapping with SAR data.

In addition to having the *high-resolution CORINE map that can serve as a ground-truth (labels)* for training the deep learning models, another reason that we selected Finland is that it is a *northern country with frequent cloud cover*, which means that using optical imagery for wide-area mapping is often not feasible. Hence, demonstrating the usability of radar imagery for LC mapping is particularly useful here.

Even though Finland is a relatively small country, there is still considerable heterogeneity present in terms of LC types and how they appear in the SAR images. Namely, *SAR backscattering is sensitive to several factors that likely differ between countries or between distant areas within a country*. Examples of such factors are moisture levels, terrain variation and soil roughness,

predominant forest biome and tree species proportions, types of shorter vegetation and crops in agricultural areas, and specific types of built environments. We did not contain our study to a particular area of Finland where the SAR signatures might be consistent, but we obtained the images across a wide area. Hence, demonstrating the suitability of our methods in this setting hints at their potential generalizability. Namely, the semantic segmentation models can be fine-tuned and adapted to work on data from other regions or countries with different SAR signatures.

On the other hand, we took into account that the same areas *will appear somewhat different on the SAR images across different seasons*. Scattering characteristics of many LC classes change considerably between the summer and winter months, and sometimes even within weeks during seasonal changes [20], [74]. These include snow cover and melting, freeze/thaw of soils, ice on rivers and lakes, crops growing cycle, and leaf-on and leaf-off conditions in deciduous trees. Because of this, in the present study, we focused only on the scenes acquired during the summer season. However, we did allow our training dataset to contain several images of the same area, taken during different times during the summer season. This way not only spatial but also the temporal variation of SAR signatures is introduced.

Our contributions can be summarized as follows.
C1: We thoroughly *benchmarked* seven selected state-of-the-art semantic segmentation models covering a diverse set of approaches for LC mapping using Sentinel-1 SAR imagery. We provided insights on the best models in terms of both *accuracy* and *efficiency*.
C2: Our results demonstrated the power of deep learning models along with SAR imagery for accurate *wide-area land cover mapping* in the *cloud obscured boreal zone and polar regions* These results can serve as *baselines* when developing new, specialized approaches to SAR imagery.

II. DEEP LEARNING TERMINOLOGY

As with other *representation learning* models, the power of deep learning models comes from their ability to *learn rich features* (representations) from the dataset automatically [15]. The automatically learned features are usually better suited for the classifier or other task at hand than hand-engineered features. Moreover, thanks to a *large number of layers* employed, it has been proven that the deep learning networks can discover *hierarchical representations* so that the higher-level representations are expressed in terms of the lower level, simpler ones. For example, in the case of images, the low-level representations that can be discovered are edges and using them, the mid-level ones can be expressed, such as corners and shapes, and this helps to express the high-level representations, such as object elements and their identities [15].

The deep learning models in computer vision can be grouped according to their main task in three categories. In Table I, we provide a description for those categories. However, the deep learning terminology for those tasks does not always correspond well to the terminology used in the remote sensing community. Relevant to our task, several remote sensing studies use the term

---

[2][Online]. Available: http://image-net.org/challenges/LSVRC/2012
[3][Online]. Available: https://github.com/tensorflow/models/tree/master/research/slim#pre-trained-models

TABLE I
SELECTED TERMINOLOGY IN THE DEEP LEARNING AND REMOTE SENSING
COMMUNITIES

Deep learning	Remote sensing	Task description
Classification [13]	Image Annotation, Scene Understanding, Scene Classification	Assigning a *whole image* to a class based on what is (mainly) represented in it, for example a ship, oil tank, sea or land.
Object Detection, Localization, Recognition [15]	Automatic Target Recognition	Detecting (and localizing) presence of particular *objects* in an image. These algorithms can detect several objects in the given image. For instance ship detection in SAR images.
Semantic Segmentation [76]	Image Classification, Clustering	Assigning a class to each *pixel* in an image based on which image object or region it belongs to. These algorithms not only detect and localize objects in the image, but also output their exact areas and boundaries.

Fig. 1. Architecture of fully convolutional neural networks (FCNs), image source: IEEE [75].

classification in the context of LC mapping, inherently meaning pixel- or region-based classification, which in the deep learning terminology corresponds to *semantic segmentation*. In Table I we list the corresponding terminology that we encountered being used for each task in both, the deep learning and remote sensing communities. This is helpful to disambiguate when talking about different and recognize when talking about the same tasks in the two domains. In the present study, the focus is on LC mapping. Hence, we tackle *semantic segmentation* in the deep learning terminology and *image classification, i.e., pixelwise classification*, in the remote sensing terminology. Another terminology issue that often arises is about the dataset types used. The dataset that is held out from the training set and used to give an estimate of the model's performance during the training phase is referred to as a *development dataset* or *validation dataset* in the deep learning context. From a remote sensing viewpoint, both training and development/validation datasets belong to training phase data. Further, the term *validation data* in remote sensing context is typically reserved for the datasets used during the final evaluation (accuracy assessment) on completely independent data not involved in the training phase, i.e., what is called a *test dataset* in deep learning. Hence, to avoid any confusion, we will avoid using the *validation* term in the text, calling respective datasets as training, development, and test (accuracy assessment) data.

CNNs [12], [13] are the deep learning models that have transformed the computer vision field. Initially, CNNs were defined to tackle the *image classification* (deep learning terminology) task. Their structure is inspired by the visual perception of mammals [76]. CNNs are named after one of the most important operations, which is particular to them compared to other neural networks, i.e., *convolutions*. Mathematically, convolution is a combination of two other functions. A convolution is applied to the image by sliding a *filter (kernel)* of a given size $k \times k$ which is usually small compared to the original image size. Different purpose filters are designed; for example, a filter can serve as a vertical edge detector. Application of such a convolution operation on an image results in a feature map.
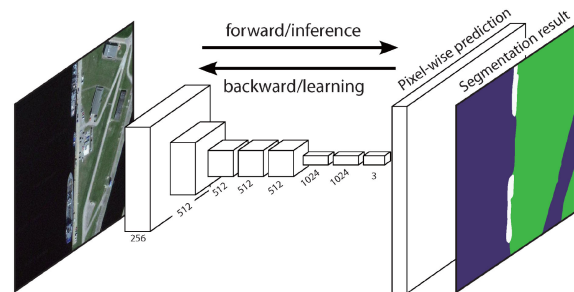
Another common operation that is usually applied after a convolution is *pooling*. Pooling reduces the size of the feature map while providing robustness to the extracted features. Common CNNs end with a *fully connected layer* which is used for final predictions, commonly for image classification. By employing a large number of convolutional layers (depth), CNNs can extract gradually more complex and abstract features. The first CNN model to demonstrate its impressive effectiveness in image classification (of hand digits) was LeNet [12]. Several years later, Krizhevsky *et al.* [13] developed AlexNet, the deep CNN to dramatically push the limits of classification accuracy on the famous ImageNet computer vision challenge [77]. Since then, a variety of CNN-based models have been proposed. Some notable examples are: VGG network [14], ResNet [78], DenseNet [79], and Inception V3 [80]. The effectiveness of CNNs has been also proven in various real-world applications [81], [82].

Once CNNs have proven their effectiveness to classify images, Long *et al.* [75] were the first to discover how they can augment a given CNN model to make it suitable for the *semantic segmentation* task—they proposed the *fully convolutional neural network (FCN)* framework. This generic architecture can be used to adapt any CNN network used for classification into a segmentation model. Namely, the authors have shown that by replacing the last, fully connected layer, with an appropriate convolutions layer, so that they will upsample and restore the resolution of the input at the output layer, CNNs can be transformed to classify each individual pixel (instead of the whole image). The basic idea is as follows. The *encoder* is used to learn the feature maps, and is usually based on a pretrained deep CNN for classification, such as ResNet, VGG, or Inception. The *decoder* part serves to upsample the discriminative features that the encoder has learned from the coarse-level feature map to the fine, pixel level. Long *et al.* [75] have shown that this upsampling (backward) computation can be efficiently performed using backward convolutions (deconvolutions). Moreover, this means that the specific CNN models, such as those mentioned above, can all be incorporated in the FCN framework for segmentation, giving rise to FCN-AlexNet [75], FCN-ResNet [78], FCN-VGG16 [75], FCN-DenseNet [73], etc. We present a diagram of the generic FCN architecture in Fig. 1.

III. MATERIALS AND METHODS

Here, we first describe the study site, SAR, and reference data. This is followed by an in-depth description of the deep
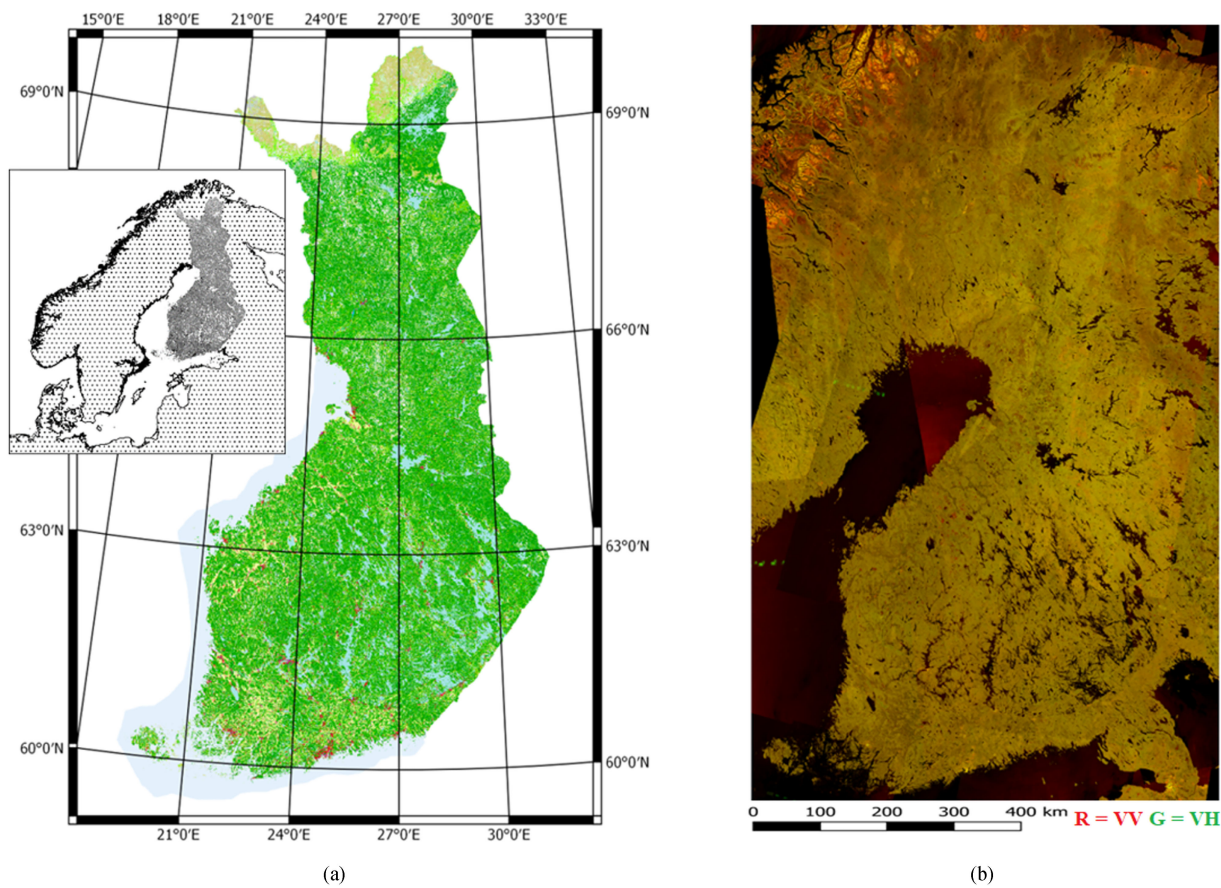
Fig. 2. Study area in Finland. (a) Reference CORINE land cover data. (b) Example of compiled Sentinel-1 SAR mosaic that includes the whole country.

learning terminology and the models used in the study. We finish with the description of the experimental setup and the evaluation metrics.

A. Study Site

Our study site is covering the territory of Finland located to the south of $66.0°$ latitude, which is effectively the whole country without Finnish Lapland. The study area is shown in Fig. 2. Southern Finland is primarily covered by boreal forests with lakes, marshes, open bogs, agricultural areas, and urban settlements. We have omitted Lapland due to considerably different LC composition and topography compared to the rest of the country. The terrain height variation within the study area is moderate and mostly within the $100 - 300$-m range.

B. SAR Data

Presently, Sentinel-1 is a C-band SAR dual-satellite system with two satellites orbiting $180°$ apart [11], launched in 2014 and 2016, respectively. The operational acquisition modes are stripmap (SM), interferometric wide-swath (IW), extra wide swath (EW), and wave mode (WV). The IW-mode is the default mode over land, providing a 250-km wide swath composed of three subswaths, with a single look image at 5 by 20-m spatial resolution. It uses the so-called terrain observation with progressive scan (TOPS) SAR mode.

SAR data acquired by Sentinel-1 satellites in IW mode is used in our study. Specifically, we used only Sentinel-1 A imagery acquired during the summer of 2018.

Original scenes were downloaded as Level-1 ground range detected (GRD) products. They represent focused SAR data that has been detected, multilooked and projected to ground range using an Earth ellipsoid. The images were orthorectified using the Technical Research Centre of Finland (VTT) in-house software employing the local digital terrain model (with 20-m resolution) available from the National Land Survey of Finland. The pixel spacing of orthorectified scenes was set to 20 m. Orthorectification included terrain flattening to obtain the backscatter signal in gamma-nought format [83]. The scenes were further reprojected to the ERTS89/ETRS-TM35FIN projection (EPSG:3067) and resampled to a final pixel size of 20 m.

The Sentinel-1 images were mosaiced into seven homogeneous SAR mosaics covering the whole territory of Finland. Each mosaic was compiled from approximately 90 Sentinel-1 IW scenes (both ascending and descending paths), and it takes about 12 days to collect enough imagery to have the whole country covered. Altogether seven SAR mosaics were produced during summer 2018. These SAR mosaics are further used for sampling the training, development, and testing images that are input to deep learning models as described in detail in Section III-F. The geographical coverage of each SAR mosaic is shown in Fig. 2.
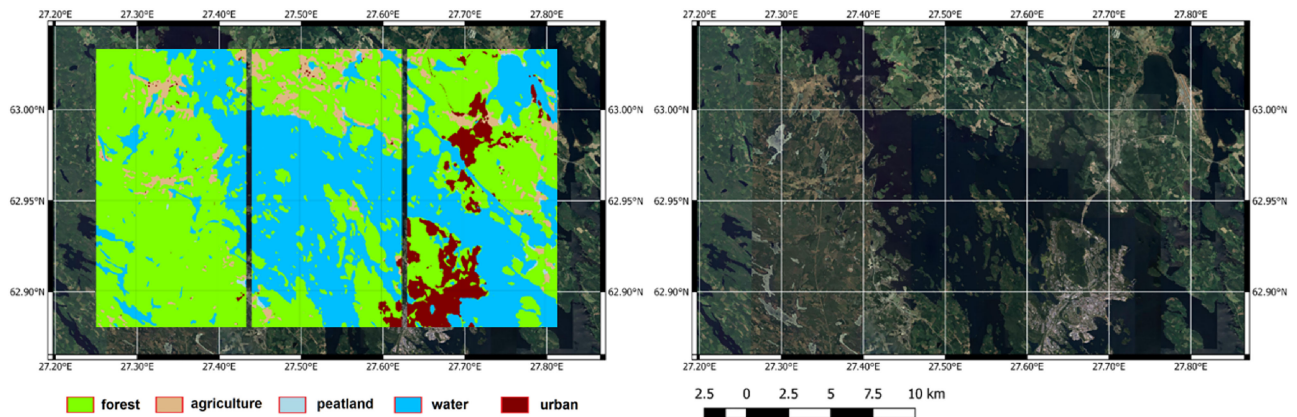
Fig. 3. Zoomed in area fragment with reference CORINE-based map shown on top (left) and along with the Google Earth layer (right), image source: Google Earth.

TABLE II
DESCRIPTION OF CORINE-BASED LAND COVER CLASSES AND THEIR MAP COLOR CODES

LC class	RGB	description
urban	128,0,0	urban fabric, industrial and commercial units, constructions sites, dump sites
agriculture	222,184,135	agricultural and agro-forestry areas, fruit trees and berry plantations, pastures
forested areas	127,255,0	broad-leaved, coniferous and mixed forest, transitional woodland/shrub
peatland	173,216,230	peatland, bogs, inland marshes and salt marshes
water bodies	0,191,255	rivers, lakes, sea

C. Reference Data

In Finland, the Finnish Environment Institute (SYKE) was responsible for the production of the CORINE maps. While for most of the EU territory, the CORINE mask of $100 \times 100 - m$ spatial resolution is available, the national institutions might choose to create more precise maps, and SYKE, in particular, had produced a $20 \times 20 - m$ spatial resolution mask for Finland (Fig. 3), with the first one in 2000. Since then, the updates have been produced regularly, with the latest one CLC2018 that well corresponds to the acquisition timing of our SAR data. There are 48 different land use classes in the map that can be hierarchically grouped into four CLC Levels. In detail, there are 30 classes on CLC Level-3, 15 classes on CLC Level-2, and 5 top CLC Level-1 classes. According to the information provided by SYKE for CLC2012, the accuracy of the CLC Level-3 was 61%, of the CLC Level-2, 83%, and of the CLC Level-1, it was 93%. In this study, we use its updated and revised version, CLC2018, having good results on both internal and external quality control. The selected classes and their corresponding color codes used for our segmentation results are shown in Table II. Our superclasses generally correspond to CLC Level-1 classes, with minor corrections for "artificial surfaces" class that is not fully included in the urban class, but some elements are distributed to other classes; most notably green urban areas were included in the forest class in our study as those are essentially parks and mixed boreal forestland enclosed within urban-designated areas.

Until the most recent CORINE production round, EEA member countries adopted national approaches for the production of CORINE. EEA technical guidelines include manual digitalization of LC change based on visual interpretation of optical satellite imagery. In Finland, the European CLC was not applicable for the majority of national users due to the large minimal mapping unit (MMU). Thus national version was produced with a somewhat modified nomenclature of classes [84]. The national high-resolution CLC2018 data are in raster format of 20 m, with corresponding MMU. In the provision of the 2018 update of CLC, obtaining optical imagery over Scandinavia and Britain was particularly challenging because of the frequent clouds, thus *calling for the use of radar imagery* to meet user requirements on accuracy and coverage [30]. CORINE map itself is normally built from high-resolution satellite images acquired primarily during the summer and, to a smaller extent, during the spring months [2].

D. Semantic Segmentation Models

We selected following seven state-of-the-art [85] semantic segmentation models to test for our LC mapping task: SegNet [65], PSPNet [66], BiSeNet [67], DeepLabV3+ [68], [69], U-Net [70], [71], FRRN-B [72], and FC-DenseNet [73]. The models were selected to cover a wide set of approaches to semantic segmentation. In the following, we describe its specific architecture for each of these DL models. We will use the following common abbreviations: `conv` for convolution operation, `concat` for concatenation, `max pool` for max pooling operation, `BN` for batch normalisation, and `ReLU` for the rectified linear unit activation function.

1) Bilateral Segmentation Network (BiSeNet): The BiSeNet model is designed to decouple the functions of encoding additional spatial information and enlarging the receptive field. These functions are fundamental to achieving good segmentation performance. As can be seen in Fig. 4(a), there are two main components to this model: Spatial path (SP) and context path (CP). Spatial path serves to encode rich spatial information. Context path serves to provide sufficient receptive field and uses global average pooling and pretrained Xception [86] or ResNet [78] as the backbone. The goal of the creators was not only to obtain superior performance but to achieve a balance
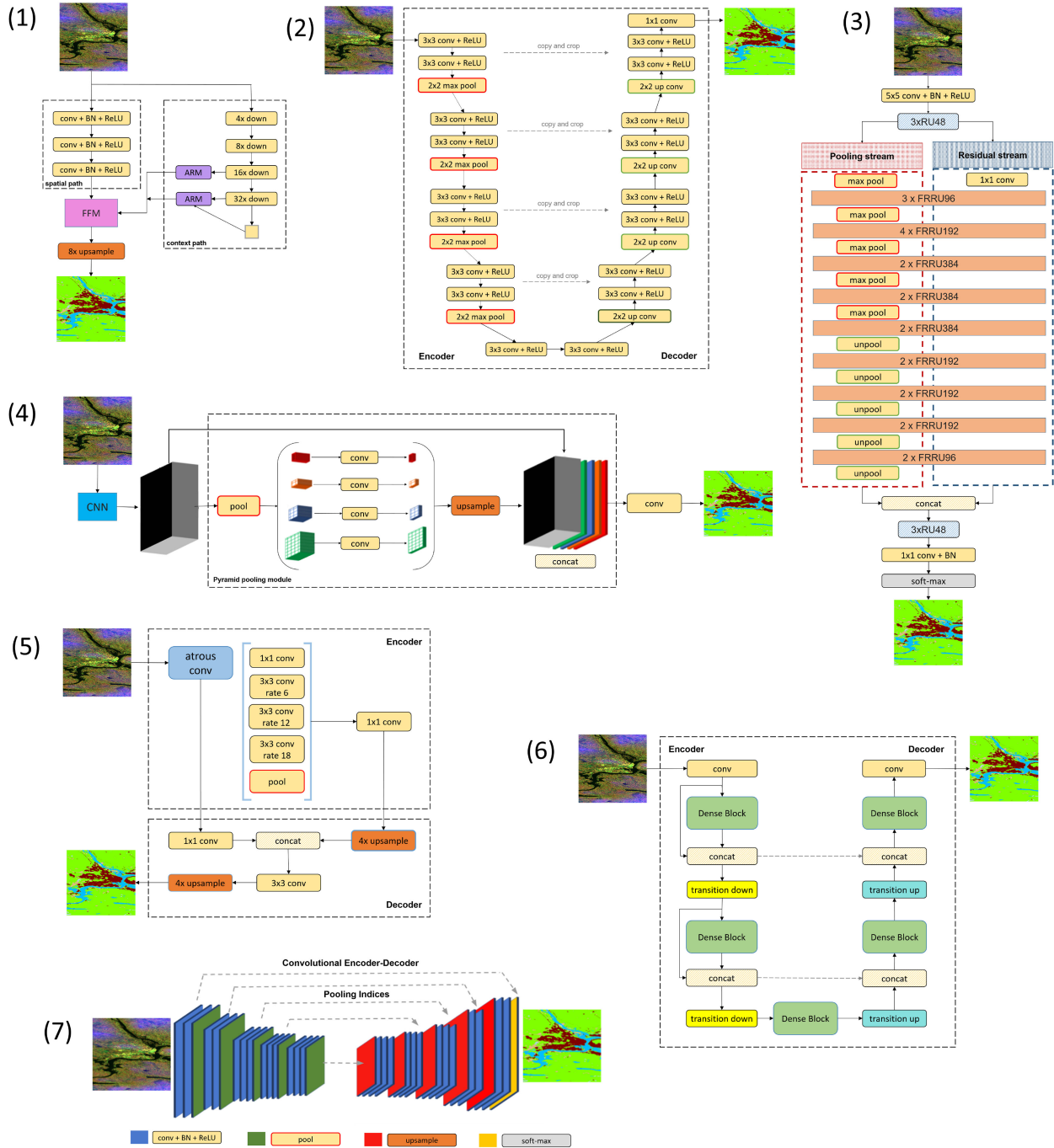
Fig. 4. Architectures of (a) BiSeNet [67], (b) U-Net [70], (c) FRRN-B [72], (d) PSPNet [66], (e) DeepLabV3+ [68], (f) FC-DenseNet [73], and (g) SegNet-based encoder-decoder with skip connections [65]. For BiSeNet, ARM stands for the attention refinement module and FFM for the feature fusion module. For FRRN-B, RU_n and FRRU_n stand for residual units and full-resolution residual units with n-channel convolutions, respectively. FRRUs simultaneously operate on the two streams.

between speed and performance. Hence, BiSeNet is a relatively fast semantic segmentation model.

2) Mobile U-Net: Mobile U-Net is based on the U-Net [70] semantic segmentation architecture shown in Fig. 4(b). In designing U-Net, a fully convolutional approach was generally employed with the following modification. Their upsampling part of the architecture has no fully convolutional layer but is nearly symmetrical to the feature extraction part due to the use of similar feature maps. This results in a *u-shaped* architecture [see Fig. 4(b)], and hence the name of the model.

While originally developed for biomedical images, the U-net architecture has proven successful for image segmentation in other domains, as well. Here, we somewhat modify the U-Net architecture, according to MobileNets [71] framework, to improve its efficiency. In particular, the MobileNets framework uses depthwise separable convolutions, a form which factorizes standard convolutions (e.g., 3×3) into a depthwise convolution (applied separately to each input band) and a pointwise (1×1) convolution to combine the outputs of depthwise convolution.

3) Full-Resolution Residual Networks (FRRN-B): As we have seen, most of the semantic segmentation architectures are based on some form of FCN, and so they utilize existing classification networks, such as ResNet or VGG16 as encoders. We also discussed the main reason for such approaches, which is to take advantage of the learned weights from those architectures pretrained for the classification task. Nevertheless, one disadvantage of the FCN approach is that the resulting network outputs of the encoder part (particularly, after the pooling operations) are at a lower resolution, which deteriorates the localization performance of the overall segmentation model. Pohlen *et al.* [72] proposed to tackle this by having two parallel network streams processing the input image: A pooling and a residual stream [Fig. 4(c)]. As the name says, the pooling stream performs successive pooling and then unpooling operations, and it serves to obtain good recognition of the objects and classes. The residual stream computes residuals at the full image resolution, which enables that low-level features, i.e., object pixel-level locations, are propagated to the network output. The name of the model comes from its building blocks, i.e., full-resolution residual units. Each such unit simultaneously operates on the pooling and the residual stream. In the original paper [72], the authors propose two alternative architectures FRRN-A and FRRN-B, and they show that FRRN-B achieves superior performance on the Cityscapes benchmark dataset. Hence, we employ the FRRN-B architecture.

4) Pyramid Scene Parsing Network (PSPNet): Zhao *et al.* [66] proposed the pyramid scene parsing as a solution to the challenge of making the local predictions based on the local context only, and not considering the global image scene. In remote sensing, an example for this challenge happening could be when a model wrongly predicts the water with waves present in it as the dry vegetation class, because they appear similar and the model did not consider that these pixels are being part of a larger water surface, i.e., it missed the global context. In similarity to the other FCN-based approaches, PSPNet uses a pretrained classification architecture to extract the feature map, in this case, ResNet. The main module of this network is the pyramid pooling, which is enclosed by a square in Fig. 4(d). As can be seen in the figure, this module fuses features at four scales: From the coarse (red) to the fine (green). Hence, the output of each level in the pyramid pooling module contains the feature map of a different resolution. In the end, the different features are stacked together yielding the final pyramid pooling global feature for predictions.

5) DeepLab-V3+: DeepLab-V3+ [68] is an improved version of DeepLab-V3 [87], while the latter is an improved version the original DeepLab [69] model. This segmentation model does not follow the FCN framework like the previously discussed models. The main features that distinguish the DeepLab model from FCNs are the *atrous convolutions* for upsampling and the application of probabilistic machine learning models, concretely, *conditional random fields (CRFs)* for a finer localization accuracy in the final fully connected layer. Atrous convolutions, in particular, allow enlarging the context from which the next layer feature maps are learned while preserving the number of parameters (and, thus, the same efficiency). Using a chain of atrous convolutions allows computing the final output layer of a CNN at an arbitrarily high resolution (removing the need for the upsampling part as used in FCNs). In the follow-up work, proposing DeepLab-V3, Chen *et al.* [87] changed the approach to atrous convolutions to gradually double the atrous rates, and showed that with an adapted version, their new algorithm outperforms the previous one, even without including the fully connected CRF layer. Finally, in their newest adaption to the model, called DeepLab-V3+, Chen *et al.* [68] turned to a similar approach to the FCNs, i.e., they add a decoder module to the architecture [see Fig. 4(e)]. That is, they employ the features extracted by the DeepLab-V3 module in the encoder part, and add the decoder module consisting of 1×1 and 3×3 convolutions.

6) Fully Convolutional DenseNets (FC-DenseNet): This semantic segmentation algorithm is built using DenseNet CNN [79] as a basis for the encoder, followed by applying the FCN approach [73]. The specificity of the DenseNet architecture is the presence of blocks, where each layer is connected to all other layers in a feed-forward manner. Fig. 4(f) shows the architecture of FC-DenseNet, where the blocks are represented by the *dense block* units. According to [79], such architecture scales well to hundreds of layers without any optimization issues, while yielding excellent results in classification tasks. In order to efficiently upsample the DenseNet feature maps, Jegou *et al.* [73] substituted the upsampling convolutions of FCNs by *dense blocks* and *transitions up*. The *transition up* modules consist of transposed convolutions, which are then concatenated with the outputs from the input skip connection [the dashed lines in Fig. 4(f)].

7) SegNet (Encoder-Decoder-Skip): Similarly to BiSeNet, SegNet is also designed with computational performance in mind, this time, particularly during inference. Because of this, the network has a significantly smaller number of trainable parameters compared to most of the other architectures. The encoder in SegNet is based on VGG16: It consists of its first 13 convolutional layers, while the fully connected layers are omitted. Hence, the novelty of this network lies in its decoder part, as follows. The decoder consists of one decoder layer for each encoder layer and so it also has 13 layers. Each decoder layer utilizes max-pooling indices memorized from its corresponding encoder feature map. The authors have shown that this enhances boundary delineation between classes. Finally, the decoder output is sent to a multiclass soft-max function yielding classification for each pixel [see Fig. 4(g)].

E. Training Approach

To improve performance, we can pretrain semantic segmentation models (their encoder modules) on a set of images of another type (such as natural images). When the model pretrained on natural images is further trained with the limited set of SAR images, the knowledge is effectively transferred from the natural to the SAR task [88]. To accomplish such transfer, we employed models whose encoders were pretrained for the ImageNet classification task and fine-tuned them on our SAR dataset (described next).

F. Experimental Setup

In this section, we first describe how we prepared the SAR images for training with the deep learning models that are originally designed for natural images, and then we provide the details of our models' implementation and the hardware setup used.

1) SAR Data Preprocessing: Sentinel-1 imagery comes in two polarization channels (VH and VV), each of them being particularly informative about certain types of land cover. Hence, using their combination is expected to yield better LC mapping results than using any of them independently.

The backscatter amplitude for both polarizations (VH and VV) represented the first two channels for our dataset. As the third channel, after some preliminary tests, we decided to include a VH-to-VV amplitude ratio (also known as cross-pol ratio). This dataset was called *RGB SAR*. In addition, for the deep learning models, each band should be normalized so that the distribution of the pixel values would resemble a Gaussian distribution centered at zero. This is done to yield a faster convergence during the training. Hence, each channel was normalized by channel-specific calibration factor using percentile contrast stretching, with no more than 1% of pixel values clipped.

The naming of the dataset comes from the process used to create the images. Namely, VH-pol data of a Sentinel-1 image is assigned to R and VV-pol to G channel. For the third, B channel, the cross-pol ratio of the Sentinel-1 data is used. Given that the semantic segmentation models expect RGB pixel values in the range (0,255), we scaled the normalized channel values for both datasets to this range.

2) Train/Development and Test (Accuracy Assessment) Dataset: The original images needed to be split into $512 \times 512 - $ px partial images (further in the text called *imagelets*) used for model training and testing. Thus, each imagelet represented an area of roughly 10×10 km^2. The first reason for this preprocessing has to do with the squared shape: Some of the selected models required square-shaped images. Some other of the models were flexible with the image shape and size but we wanted to make the setups for all the models the same so that their results are comparable. The second reason for this preprocessing has to do with computational capacity: With our hardware setup (described below), this was the largest image size that we could work with.

Given the geography of Finland, for representative training data, it is useful to include imagelets from the whole country (including the large cities) aside from the Finnish Lapland, where the land classes are distinctly different. On the other hand, some noticeable differences are found also in the gradient from east to west of the country. Hence, to achieve a representative training dataset, we selected all imagelets between the longitudes of 25° and 29° for the accuracy assessment (so-called "unobserved data" for model testing), and all the other imagelets we used for the model training (that is training/validation in the computer vision terminology). In this way, we prevented the situation in which two images of the same area, but acquired at different times, were used one for training and the other one for testing. In other words, we kept our training/development and test sets completely independent from each other.

The areas for training/development and model testing are shown in Fig. 5. From each of the seven SAR mosaics, 1000 imagelets were generated using random sampling, while controlling for no spatial overlap between the imagelets. Among those 1000 imagelets, 400 were sampled from the testing area and set aside for the accuracy assessment, while the remaining 600 were sampled from the training/development area. The procedure resulted in 4200 images in the training and development set and 2800 images in the test (accuracy assessment) set. Finally, we used 60% from the training/development set for training and the rest for the development of the deep learning models.

3) Data Augmentation: Further, we have employed the *data augmentation* technique. The main idea behind the data augmentation is to enable improved learning by reusing original images with slight transformations such as rotation, flipping, adding Gaussian noise, or slightly changing the brightness. This provides additional information to the model, and the dataset size is effectively increased. Moreover, an additional benefit of the data augmentation is in helping the model to learn some invariant data properties for which no examples are present in the original dataset. Given the sensitivity of the SAR backscatter, we did not want to augment the images in terms of color, brightness, or by adding noise. However, we could safely employ rotations and flipping. For rotations, we only used the 90° increments, giving three possible rotated versions of an image. For image flipping, we applied horizontal and vertical flipping, or both at the same time, giving another three possible versions of the original image.[4] Notice that our images are square, so the transformations did not change the image dimensions. Finally, we applied the *online augmentation*, as opposed to the offline version. In the online process, each augmented image is seen only once, and so this process yields a network that generalizes better.

4) Implementation: To apply the described semantic segmentation models, we adapted the open-source semantic segmentation suite. We used Python with TensorFlow [89] backend.

5) Hardware and Training Setup: We trained and tested separately each of the deep learning models on a single GPU (NVIDIA GeForce GTX 1080) on a machine with 32 GB of RAM.

For all the models, we used the Adam optimization method [90] with the learning rate of 0.0001, and with the exponential decay rate for the first moment estimates of 0.9, and for the second-moment estimates of 0.999. We applied the early stopping criterion so that, for each model, the training would automatically stop after there was no improvement in the development (validation) loss for 10 epochs. Such early stopping criteria resulted in different models being trained for a different number of epochs: From 69 (for DeepLabV3+) up to 126 (for SegNet) epochs. In each case, the checkpoint for the latest model with the best result before stopping was saved. Then we used that model for prediction on the test set and we report those results.

The general processing flowchart is shown in Fig. 6.

---

[4]Vertical flip operation switches between top-left and bottom-left image origin (reflection along the central horizontal axis), and horizontal flip switches between top-left and top-right image origin (reflection along the central vertical axis)
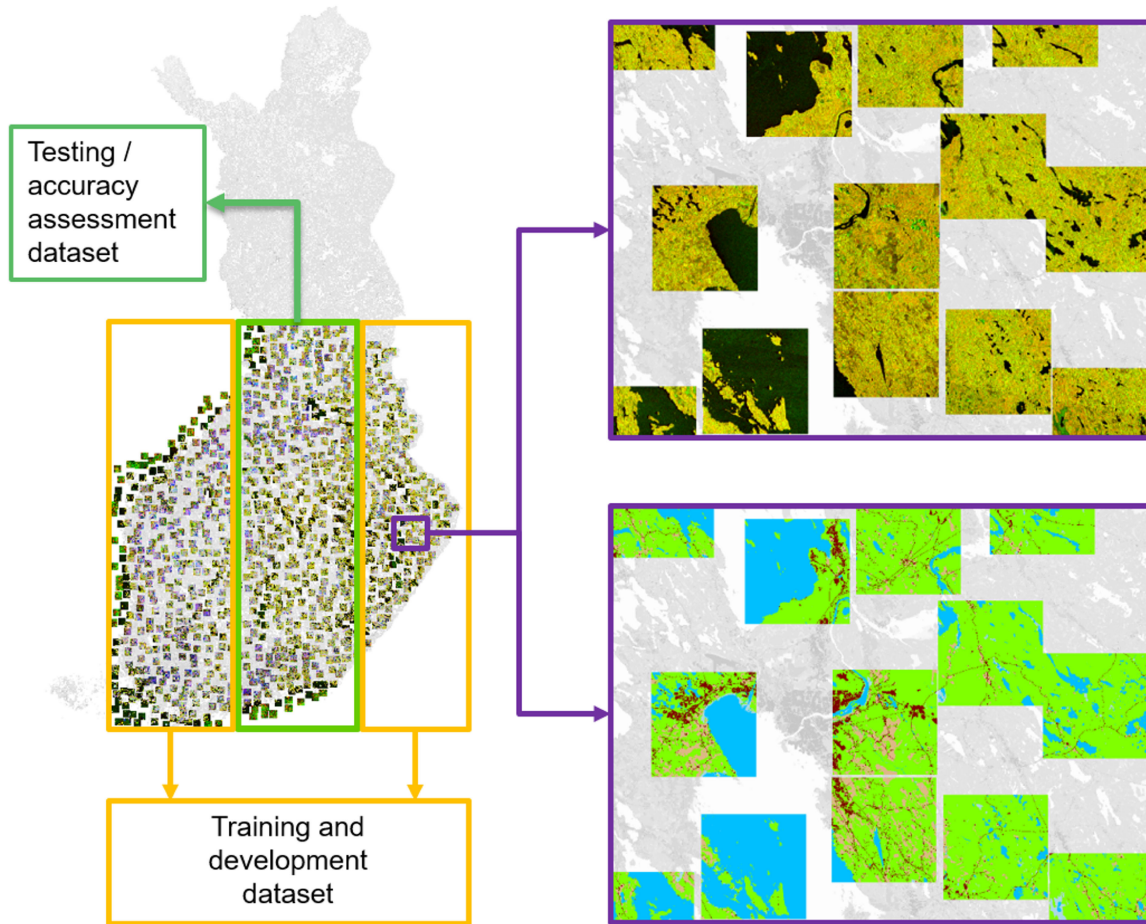
Fig. 5.   Sampling of SAR and land cover imagelets and division into training and development and testing datasets.
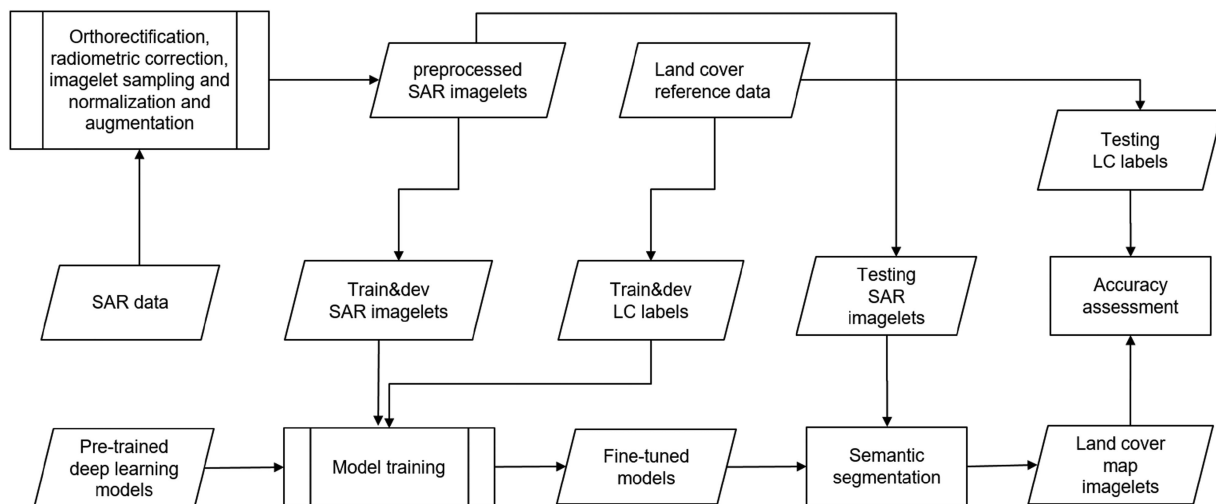

Fig. 6.   General processing flowchart for RGB SAR dataset.

G. Evaluation Metrics

In the review on the metrics used in LC classification, Costa *et al.* [91] have found a lack of consistency, complicating the intercomparison of different studies. To avoid such issues and ensure that our results are easily comparable with the literature, we thoroughly evaluated our models. For each model and class, we report the following measures of accuracy: Precision, also known as producer's accuracy (PA), recall, also known as user's

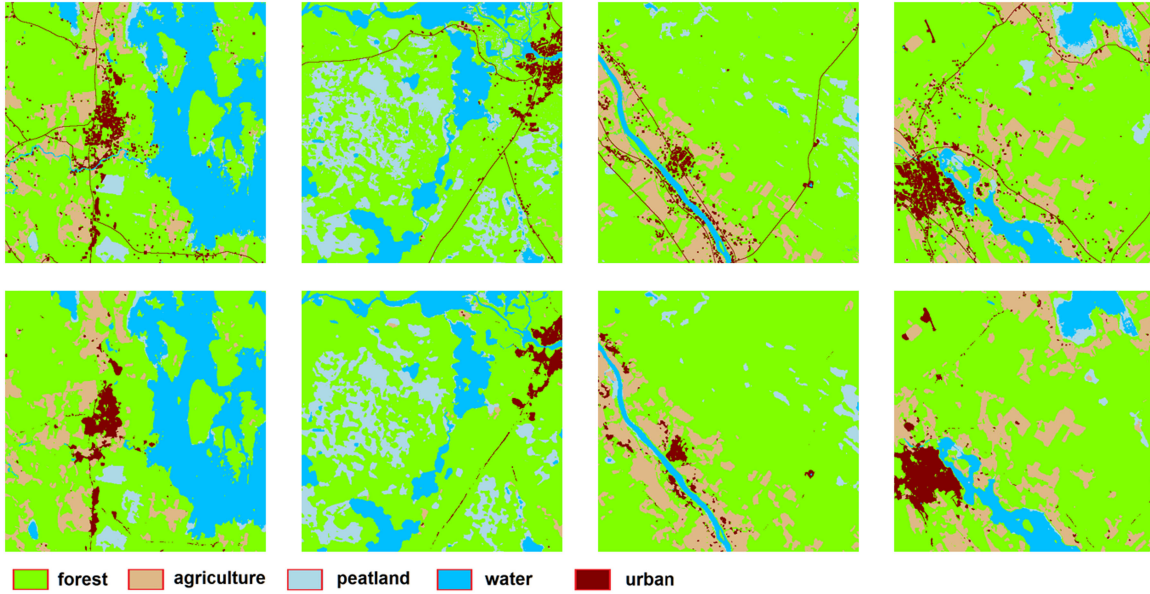■ forest ■ agriculture ■ peatland ■ water ■ urban

Fig. 7. Illustration of the FC-DenseNet model performance: Selection of classification results, i.e., direct output of the network, without any post-processing (bottom row) versus reference CORINE based LC (upper row).

accuracy (UA), and overall accuracy and Kappa coefficient. The formulas are as follows.

For each segmentation class (LC type) c, we calculate *precision (producer's accuracy)*:

$$P_c = \frac{Tp_c}{Tp_c + Fp_c}$$

and *recall (user's accuracy)*

$$R_c = \frac{Tp_c}{Tp_c + Fn_c}$$

where Tp_c represents true positive, Fp_c false positive, and Fn_c false negative pixels for the class c.

When it comes to accuracy [92], we calculate *per class accuracy*[5]

$$\mathrm{Acc}_c = \frac{C_{ii}}{G_i}$$

and *overall pixel accuracy*:

$$Acc_{OP} = \frac{\Sigma_{i=1}^{L} C_{ii}}{\Sigma_{i=1}^{L} G_i},$$

where C_{ij} is the number of pixels having a ground-truth label i and being classified/predicted as j, G_i is the total number of pixels labelled with i, and L is the number of classes. All these metrics can take values from 0 to 1.

Finally, we also use a Kappa statistic (Cohen's measure of agreement), indicating how the classification results compare to the values assigned by chance [93]. Kappa statistics can take values from 0 to 1. Starting from a k by k confusion matrix with

[5]Effectively, per class accuracy is defined as the recall obtained on each class.

TABLE III
PROPERTIES OF THE EXAMINED SEMANTIC SEGMENTATION ARCHITECTURES

Architecture	Base model	Parameters
BiSeNet	ResNet101	24.75M
SegNet	VGG16	34.97M
Mobile U-Net	Not applicable	8.87M
DeepLabV3+	ResNet101	47.96M
FRRN-B	ResNet101	24.75M
PSPNet	ResNet101	56M
FC-DenseNet	ResNet101	9.27M

elements f_{ij}, the following calculations are done:

$$P_o = \frac{1}{N} \sum_{j=1}^{k} f_{jj} \tag{1}$$

$$r_i = \sum_{j=1}^{k} f_{ij} \forall i, \text{ and } c_j = \sum_{i=1}^{k} f_{ij} \forall j \tag{2}$$

$$P_e = \frac{1}{N^2} \sum_{i=1}^{k} r_i c_i \tag{3}$$

where P_o the observed proportional agreement (effectively the overall accuracy), r_i and c_j are the row and column totals for classes i and j, and P_e is the expected proportion of agreement. The final measure of agreement is given by such statistic [93]

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \tag{4}$$

Depending on the value of Kappa, the observed agreement is considered as either poor (0.0 to 0.2), fair (0.2 to 0.4), moderate (0.4 to 0.6), good (0.6 to 0.8), or very good (0.8 to 1.0).
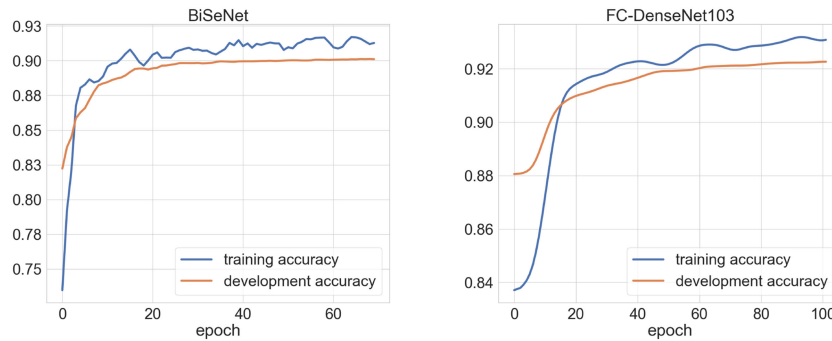
Fig. 8. Accuracy curves during training and development for the fastest (BiSeNet) and the slowest (FC-DenseNet) model. The early-stopping criteria with 10 epochs of no improvement for development loss was applied.

TABLE IV
SUMMARY OF THE CLASSIFICATION PERFORMANCE AND EFFICIENCY OF DEEP LEARNING MODELS

LC classes	Test scale (km^2)	Accuracy (UA, PA %)						
		BiSeNet	DeepLabV3+	SegNet	FRRN-B	U-Net	PSPNet	FC-DenseNet
Urban fabric	7906	59, 29	67, 29	70, 33	68, 33	65, 31	65, 31	69, 34
Agricultural areas	22659	63, 69	76, 70	79, 79	80, 77	69, 77	69, 77	80, 79
Forested areas	199327	91, 96	92, 97	94, 97	94, 97	94, 96	94, 96	94, 97
Peatland, bogs and marshes	10686	77, 48	83, 42	81, 61	77, 65	77, 51	77, 51	78, 65
Water bodies	53022	96, 86	97, 97	98, 98	98, 98	97, 97	97, 97	98, 98
Overall Accuracy (%)		88.48	91.37	**92.78**	92.69	92.25	91.37	**92.78**
Kappa		0.758	0.818	**0.851**	0.849	0.839	0.823	**0.851**
Average inference time (s / image)		0.043	**0.031**	0.073	0.143	0.085	0.053	0.196

[5]UA: User's accuracy, PA: Producer's accuracy.
The results with the best Kappa score and the smallest inference time are highlighted in bold.

TABLE V
CONFUSION MATRIX FOR CLASSIFICATION WITH FC-DENSENET MODEL

LC classes	Sentinel-1 class					*Total*	*PA*
	urban	agriculture	forest	peatland	water		
Urban fabric	6674628	2567434	10281910	127574	114168	19765714	33.8
Agricultural areas	406663	44490302	11169843	514909	66019	56647736	78.5
Forested areas	2485460	7729542	483087561	3583613	1432249	498318425	96.9
Peatland, bogs and marshes	47057	844064	7010103	17231477	1583214	26715915	64.5
Water bodies	84660	221139	2159293	594050	129496268	132555410	97.7
Total	9698464	55852481	513708710	22051623	132691918	73400320	
UA	68.8	79.7	94.0	78.1	97.6		92.8

[5]UA: User's accuracy, PA: Producer's accuracy.

IV. RESULTS AND DISCUSSION

Using the experimental setup described in the previous section, we evaluated the seven selected semantic segmentation models: SegNet [65], PSPNet [66], BiSeNet [67], DeepLabV3+ [68], [69], U-Net [70], [71], FRRN-B [72], and FC-DenseNet [73]. The overall classification performance statistics for all studied models is gathered in Table IV. Fig. 7 shows maps produced for several imagelets with the best performing model, FC-DenseNet. Fig. 8 illustrates training and development progress for the fastest (BiSeNet) and the slowest (FC-DenseNet) models. Obtained results are compared to prior work and classification performance for different LC classes is discussed further.

A. Classification Performance

All the models performed relatively well on both datasets achieving an overall accuracy above 87% for each model. Four models performed particularly well, achieving an accuracy score above 92% on both datasets; those are FRRN-B, U-Net, SegNet, and FC-DenseNet. The two latter models were also somewhat better than others in terms of Kappa statistics, and, along with FRRN-B, were the best models also concerning classwise user's and producer's accuracy.

The advantage for SegNet is that its training and inference times were 2.5 times better compared to the FC-DenseNet of similar accuracy. BiSeNet and DeepLabV3 were performing somewhat worse than other five models particularly in terms

of agreement (Kappa was 0.75–0.82) but also overall accuracy was lower, most strongly for BiSeNet. Overall accuracy and classwise accuracies obtained on completely independent test dataset were still remarkably high compared to other reported results in the literature when more conventional statistical or traditional machine learning approaches were used with C-band SAR data [30], [94]. Further in-depth comparison can be found in Section IV-C.

Before further analysis, let us recall that CORINE is not exclusively a LC map, but rather a LC and land use map, thus some specific classes can differ from ecological classes observed by Sentinel-1. Also, the aggregation to CLC Level-1 is sometimes not strictly "ecological" nor it complies with physics surface scattering considerations. For example, airports, major industrial areas, and road networks often appear similar to fields, the presence of trees and green vegetation near summer cottages can cause them to exhibit signatures close to the forest rather than urban, sometimes forest on the rocky terrain can be misclassified as urban instead due to presence of very bright targets and strong disruptive features, while confusion between peatland and agricultural and grassland areas is also common. Finally, the accuracy of the CORINE data is only somewhat higher than 90%.

As for the results across the different land classes, all the models performed particularly well in recognizing the water bodies and forested areas, while the urban fabric represented the most challenging class for all the models. The urban class was particularly challenging for the following main reasons. First, this is still essentially a land-use class, with continuous urban fabric (easy to recognize by radar) representing only a moderate fraction of the whole class. It also changes the most, as new houses, roads, and urban areas are built. Second, the CORINE map itself does not have perfect accuracy, neither aggregation rules are perfect. As a matter of fact, in a majority of studies where SAR-based classification was done versus CLC or similar data, a poor or modest overall agreement was observed for urban land use areas [20], [41], [74], [95], while the user's accuracy was strongly higher than producer's [21]. The latter is exactly due to radar being able to sense sharp boundaries and bright targets very well whereas such bright targets often do not dominate the whole urban land-use class. Importantly, relatively good performance was obtained in mapping agricultural and wetland areas, particularly well differentiating between them while this is often problematic with other remote sensing instruments.

We mentioned the issues of SAR backscattering sensitivity to several ground factors so that the same classes might appear differently on the images between countries or between distant areas within a country. An interesting indication of our study, however, is that the deep learning models might be able to deal with this issue. Namely, we used the models pretrained on ImageNet and fine tuned-them with a relatively small number of Sentinel-1 images. The models learned to recognize varying types of backscattering signal across the country of Finland. This indicates that with a similar type of fine-tuning, present models could be relatively easily adapted to the other areas and countries, with different SAR backscattering patterns. Such robustness and adaptability of the deep learning models come

from their automatic learning of feature representation, without the need for a human expert predefining those features.

B. Computational Performance

The training times with our hardware configuration took from 1 up to 2.5 days for the different models. This could be significantly improved by training each model using a multi-GPU system instead of a single-GPU in our experiments.

In terms of the inference time, we also saw the differences in the performance. In Table IV, we present the average inference time per $512 \times 512 - px$ imagelet that we worked with. The results show that there is a tradeoff between classification and computational performance: The best models in terms of classification results (i.e., FC-DenseNet and FRRN-B) take several times longer inference time compared to the rest. A positive exception in this regard is the SegNet model, which achieved the best classification results together with FC-DenseNet but with 2.5 times better inference time. Depending on the application, this might or might not be of particular importance.

C. Comparison to Similar Work

Obtained results compare favorably to previous similar studies on LC classification with SAR data [20], [28], [30], [41], [74], [95]. Depending on the level of classes aggregation (4–5 major classes or more), using mostly statistical or classical machine learning approaches reported classification accuracies were as high as 80–87% to as low as 30% when only SAR imagery were used.

Two recent studies that employed neural networks to SAR imagery classification (albeit in combination with satellite optical data) for LC mapping were [28] and [57], with reported classification accuracies of up to 97.5% and 94.6%, respectively.

The best models in our experiments achieved the overall accuracy of 93%. However, our results are obtained using solely the SAR imagery. In contrast, SAR imagery (PALSAR) alone yielded the overall accuracy of 78.1% in [28]. The types of classes they studied are also different compared to ours (crops versus vegetation versus LC types) and our study is performed on a larger area. Importantly, the previous studies have applied different types of models (regular NNs versus CNN versus semantic segmentation). In particular, the CNN models work on the 7×7 resolution windows, while we have applied more advanced semantic segmentation models, which work on the level of a pixel. Keeping in mind findings from [28] that the addition of optical images on top of SAR improved the results by over 10%, we expect that our models would perform comparably well or outperform these previous works if applied to a combined SAR and optical imagery.

In terms of the deep learning setup, the most similar to ours are the studies [50] and [61]. However, RapidEye optical imagery at 5-m spatial resolution was used in [50], and the test site was considerably smaller. Study [61], similar to our research, relied exclusively on SAR imagery, however, fully polarimetric images, and acquired by RADARSAT-2 at considerably better resolution. They have developed an FCN-type of a semantic

segmentation model *"specifically designed for the classification of wetland complexes using PolSAR imagery."* Using this model to classify eight wetland map classes, they achieved an overall accuracy of 93%. However, because their model is designed specifically for wetland complexes, it is not clear if such a model would generalize to other types of areas. Compared to our study, they have focused on a considerably smaller area (nearly the size of a single imagelet we used), and on a very specific task (wetland types mapping). Thus, it is not readily clear how general their approach is and how it compares to our presented approach.

D. Outlook and Future Work

There are several lines for potential improvement based on the results of this study, as well as future work directions.

First, using even a larger set of Sentinel-1 images can be recommended since for the supervised deep learning models large amounts of data are crucial. Here, we processed only 7 K imagelets altogether, but deep learning algorithms become efficient typically only once they are trained with hundreds of thousands or millions of images.

Second, if SAR images and reference data of a higher resolution are used, we expect better classification performance, too, as smaller details could be potentially captured. An interesting extension of the study could be to use a digital elevation model as additional input along with SAR data to deep learning models, as this could potentially improve accuracies for several land classes. In our opinion, such models should, however, be not global topographical DEMs but rather high-resolution lidar digital surface models, if improvements in mapping, for example, urban class are sought.

The reference maps should represent reality as accurately as possible. The models in our experiments were certainly limited by the CORINE's limited accuracy.

Third, in this study, we have tested the effectiveness of off-the-shelf deep learning models for LC mapping from SAR data. While the results show their effectiveness, it is also likely that the novel types of models, specifically developed for the radar data (such as [61]), will yield even better results. Based on our results, we suggest DenseNet- and SegNet-based models as a starting point. In particular, one could develop the deep learning models to handle directly the SLC data which preserve the phase information.

Focusing on a single season is both an advantage and a limitation. Importantly, we have avoided confusion between SAR signatures varying seasonally for several LC classes. However, multitemporal dynamics itself can be potentially used as an additional useful class-discriminating parameter. Incorporating seasonal dynamics of each LC pixel (as a time series) is left for future work, perhaps with the additional need to incorporate recurrent neural networks into the approach.

As discussed in Section IV-A, it could be suitable to use more detailed (specific) LC classes, as an aggregation of smaller LC classes into CORINE super-classes is not exactly ecological, leading to mixing several distinct SAR signatures in one class and thus causing additional confusion for the classifier. Later, classified specific classes can be aggregated into larger classes, potentially showing improved performance [19] .

Finally, we have used only SAR images for the presented large-scale LC mapping. If one were to combine other types of remote sensing images, in particular the optical images, we expect that the results would significantly improve. This is true for those areas where such imagery can be collected due to cloud coverage, while in operational scenario it would potentially require the use of at least two models (with and without optical satellite imagery). It is also important to access the added value of SAR imagery with deep learning models when optical satellite images are available, as well as possible data fusion and decision fusion scenarios before a decision on the mapping approach is done [19].

V. CONCLUSION

Our study demonstrated the potential for applying state-of-the-art semantic segmentation models to SAR image classification with high accuracy. Several models were benchmarked in a countrywide classification experiment using Sentinel-1 IW-mode SAR data, reaching nearly 93% overall classification accuracy with the best performing models (SegNet and FC-DenseNet). This indicates strong potential for using pretrained CNNs for further fine-tuning and seems particularly suitable when the number of training images is limited (to thousand or tens of thousands instead of millions). In addition to suggesting the best candidate semantic segmentation models for LC mapping with SAR data (that is, the SegNet or DenseNet-based models), our study offers baseline results against which the newly proposed models should be evaluated. Several possible improvements for future work were identified, including the necessity for testing multitemporal approaches, data fusion, and employing DEM models, very high-resolution SAR imagery, as well as developing models specifically for SAR.

V. DATA AVAILABILITY

The implementation scripts with documentation are available on GitHub,[6] the original Sentinel-1 images can be downloaded from SciHub,[7] and the processed train/development and test data are published on Zenodo[8] and IEEE DataPort.

REFERENCES

[1] S. Bojinski, M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, "The concept of essential climate variables in support of climate research, applications, and policy," *Bull. Amer. Meteorological Soc.*, vol. 95, no. 9, pp. 1431–1443, 2014.

[6][Online]. Available: https://github.com/sanja7s/DL_SemSAR_with_docs
[7][Online]. Available: https://scihub.copernicus.eu
[8][Online]. Available: https://zenodo.org

[2] G. Büttner, J. Feranec, G. Jaffrain, L. Mari, G. Maucha, and T. Soukup, "The CORINE land cover 2000 project," *EARSeL eProceedings*, vol. 3, no. 3, pp. 331–346, 2004.

[3] M. Bossard, J. Feranec, and J. Otahel, *CORINE Land Cover Technical Guide: Addendum 2000*. Copenhagen, Germany: Eur. Environ. Agency. 2000.

[4] G. Büttner, "CORINE land cover and land cover change products," in *Land Use and Land Cover Mapping in Europe*. Berlin, Germany: Springer, 2014, pp. 55–74.

[5] M. Törmä, T. Markkanen, S. Hatunen, P. Härmä, O.-P. Mattila, and A. Arslan, "Assessment of land-cover data for land-surface modelling in regional climate studies," *Boreal Environ. Res.*, vol. 20, no. 2, pp. 243–260, 2015.

[6] J. Chen et al., "Global land cover mapping at 30 m resolution: A POK-based operational approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 103, pp. 7–27, 2015.

[7] C. A. d. Almeida, A. C. Coutinho, J. C. D. M. Esquerdo, M. Adami, A. Venturieri, C. G. Diniz, N. Dessay, L. Durieux, and A. R. Gomes, "High spatial resolution land use and land cover mapping of the brazilian legal amazon in 2008 using landsat-5/TM and modis data," *Acta Amazonica*, vol. 46, no. 3, pp. 291–302, 2016.

[8] C. Homer et al., "Completion of the 2011 national land cover database for the conterminous United States-representing a decade of land cover change information," *Photogrammetric Eng. Remote Sens.*, vol. 81, no. 5, pp. 345–354, 2015.

[9] Y. Zhao et al., "Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data," *Remote Sens. Environ.*, vol. 183, pp. 170–185, 2016.

[10] P. Griffiths, C. Nendel, and P. Hostert, "Intra-annual reflectance composites from Sentinel-2 and landsat for national-scale crop and land cover mapping," *Remote Sens. Environ.*, vol. 220, pp. 135–151, 2019.

[11] R. Torres et al., "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, 2012.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[16] W. Cohen and S. Goward, "Landsat's role in ecological applications of remote sensing," *BioScience*, vol. 54, no. 6, pp. 535–545, 2004.

[17] S. Goetz et al., "Mapping and monitoring carbon stocks with satellite observations: A comparison of methods," *Carbon Balance Manage.*, vol. 4, 2009, Art. no. 2.

[18] C. Atzberger, "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs," *Remote Sens.*, vol. 5, no. 2, pp. 949–981, 2013.

[19] T. Hame et al., "Improved mapping of tropical forests with optical and SAR imagery, Part I: Forest cover and accuracy assessment using multi-resolution data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 1, pp. 74–91, Feb. 2013.

[20] O. Antropov, Y. Rauste, H. Astola, J. Praks, T. Hame, and M. Hallikainen, "Land cover and soil type mapping from spaceborne PolSAR data at l-band with probabilistic neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5256–5270, Sep. 2014.

[21] O. Antropov, Y. Rauste, and T. Hame, "Volume scattering modeling in PolSAR decompositions: Study of ALOS PALSAR data over boreal forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3838–3848, Oct. 2011.

[22] B. Waske and M. Braun, "Classifier ensembles for land cover mapping using multitemporal SAR imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 5, pp. 450–457, 2009.

[23] L. Bruzzone, M. Marconcini, U. Wegmüller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.

[24] T. Ullmann et al., "Land cover characterization and classification of arctic tundra environments by means of polarized synthetic aperture X- and C-band radar (PoLSAR) and landsat 8 multispectral imagery - Richards Island, Canada," *Remote Sens.*, vol. 6, no. 9, pp. 8565–8593, 2014.

[25] N. Clerici, C. A. V. Calderón, and J. M. Posada, "Fusion of Sentinel-1a and Sentinel-2a data for land cover mapping: A case study in the lower magdalena region, Colombia," *J. Maps*, vol. 13, no. 2, pp. 718–726, 2017.

[26] C. Castañeda and D. Ducrot, "Land cover mapping of wetland areas in an agricultural landscape using SAR and landsat imagery," *J. Environ. Manage.*, vol. 90, no. 7, pp. 2270–2277, 2009.

[27] Y. Ban, H. Hu, and I. M. Rangel, "Fusion of quickbird MS and radarsat SAR data for urban land-cover mapping: Object-based and knowledge-based approach," *Int. J. Remote Sens.*, vol. 31, no. 6, pp. 1391–1410, 2010.

[28] G. V. Laurin et al., "Optical and SAR sensor synergies for forest and land cover mapping in a tropical site in West Africa," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 21, pp. 7–16, 2013.

[29] R. Khatami, G. Mountrakis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sens. Environ.*, vol. 177, pp. 89–100, 2016.

[30] H. Balzter, B. Cole, C. Thiel, and C. Schmullius, "Mapping CORINE land cover from Sentinel-1 A SAR and SRTM digital elevation model data using random forests," *Remote Sens.*, vol. 7, no. 11, pp. 14876–14898, 2015.

[31] S.-E. Park, "Variations of microwave scattering properties by seasonal freeze/thaw transition in the permafrost active layer observed by ALOS PALSAR polarimetric data," *Remote Sens.*, vol. 7, no. 12, pp. 17135–17148, 2015.

[32] J. D. T. De Alban, G. M. Connette, P. Oswald, and E. L. Webb, "Combined landsat and l-band SAR data improves land cover classification and change detection in dynamic tropical landscapes," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 306.

[33] N. Longepe, P. Rakwatin, O. Isoguchi, M. Shimada, Y. Uryu, and K. Yulianto, "Assessment of ALOS PALSAR 50 m orthorectified FBD data for regional land cover classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2135–2150, Jun. 2011.

[34] T. Esch, A. Schenk, T. Ullmann, M. Thiel, A. Roth, and S. Dech, "Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1911–1925, Jun. 2011.

[35] J. W. Cable, J. M. Kovacs, J. Shang, and X. Jiao, "Multi-temporal polarimetric RADARSAT-2 for land cover monitoring in northeastern Ontario, Canada," *Remote Sens.*, vol. 6, no. 3, pp. 2372–2392, 2014.

[36] X. Niu and Y. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 1–26, 2013.

[37] T. L. Evans, M. Costa, K. Telmer, and T. S. F. Silva, "Using ALOS/PALSAR and RADARSAt-2 to map land cover and seasonal inundation in the Brazilian Pantanal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 4, pp. 560–575, Dec. 2010.

[38] C. da Costa Freitas et al., "Land use and land cover mapping in the Brazilian Amazon using polarimetric airborne p-band SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2956–2970, Oct. 2008.

[39] G. Li, D. Lu, E. Moran, L. Dutra, and M. Batistella, "A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region," *ISPRS J. Photogrammetry Remote Sens.*, vol. 70, pp. 26–38, 2012.

[40] N. Park and K. Chi, "Integration of multitemporal/polarization C-band SAR data sets for land-cover classification," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4667–4688, 2008.

[41] A. Lonnqvist, Y. Rauste, M. Molinier, and T. Hame, "Polarimetric SAR data in land cover mapping in Boreal zone," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3652–3662, Oct. 2010.

[42] E. Tomppo, O. Antropov, and J. Praks, "Cropland classification using Sentinel-1 time series: Methodological performance and prediction uncertainty assessment," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2480.

[43] D. B. Nguyen, A. Gruber, and W. Wagner, "Mapping rice extent and cropping scheme in the Mekong delta using Sentinel-1a data," *Remote Sens. Lett.*, vol. 7, no. 12, pp. 1209–1218, 2016.

[44] A. Veloso et al., "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sens. Environ.*, vol. 199, pp. 415–426, 2017.

[45] G. Satalino, A. Balenzano, F. Mattia, and M. W. Davidson, "C-band SAR data for mapping crops dominated by surface or volume scattering," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 384–388, Feb. 2014.

[46] A. W. Jacob et al., "Sentinel-1 InSAR coherence for land cover mapping: A comparison of multiple feature-based classifiers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 535–552, 2020.

[47] S. Ge, O. Antropov, W. Su, H. Gu, and J. Praks, "Deep recurrent neural networks for land-cover classification using Sentinel-1 InSAR time series," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 473–476.

[48] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1995.

[49] X. X. Zhu *et al.*, "Deep learning in remote sensing: A review," 2017, *arXiv:1710.03959*.

[50] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1119.

[51] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[52] J. Zhang, P. Zhong, Y. Chen, and S. Li, "$L\_1/2$ regularized deconvolution network for the representation and restoration of optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2617–2627, May 2014.

[53] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[54] J. Wang *et al.*, "Deep hierarchical representation and segmentation of high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4320–4323.

[55] D. Tuia, R. Flamary, and N. Courty, "Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 272–285, 2015.

[56] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, Mar. 2016.

[57] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[58] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[59] G. Wu *et al.*, "Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks," *Remote Sens.*, vol. 10, no. 3, 2018, Art. no. 407.

[60] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, 2017.

[61] F. Mohammadimanesh, B. Salehi, M. Mahdianpari, E. Gill, and M. Molinier, "A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem," *ISPRS J. Photogrammetry Remote Sens.*, vol. 151, pp. 223–236, 2019.

[62] L. Wang, X. Xu, H. Dong, R. Gui, and F. Pu, "Multi-pixel simultaneous classification of POLSAR image using convolutional neural networks," *Sensors*, vol. 18, no. 3, 2018, Art. no. 769.

[63] M. Ahishali, S. Kiranyaz, T. Ince, and M. Gabbouj, "Dual and single polarized SAR image classification using compact convolutional neural networks," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1340.

[64] Z. Li, Z. Yang, and H. Xiong, "Homogeneous region segmentation for SAR images based on two steps segmentation algorithm," in *Proc. Int. Conf. Comput., Commun., Syst.*, 2015, pp. 196–200.

[65] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[66] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 2881–2890.

[67] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiseNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 325–341.

[68] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018, *arXiv:1802.02611*.

[69] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[71] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[72] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4151–4160.

[73] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2017, pp. 1175–1183.

[74] O. Antropov, Y. Rauste, A. Lonnqvist, and T. Hame, "PolSAR mosaic normalization for improved land-cover mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1074–1078, Nov. 2012.

[75] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 3431–3440.

[76] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the Cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[77] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[78] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.

[79] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4700–4708.

[80] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, June 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[81] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[82] T. N. Sainath, A.-R. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 8614–8618.

[83] D. Small, L. Zuberbühler, A. Schubert, and E. Meier, "Terrain-flattened gamma nought Radarsat-2 backscatter," *Can. J. Remote Sens.*, vol. 37, no. 5, pp. 493–499, 2012.

[84] P. Härmä, R. Teiniranta, M. Törmä, R. Repo, E. Järvenpää, and M. Kallio, "The production of finnish CORINE land cover 2000 classification," in *Proc. 20th Int. Arch. Photogrammetry Remote Sens. Congr.*, 2004, pp. 1330–1335.

[85] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*.

[86] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 1251–1258.

[87] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[88] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. Int. Conf. Mach. Learn. Workshop Unsupervised Transfer Learn.*, 2012, pp. 17–36.

[89] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th Symp. Operating Syst. Des. Implementation*, 2016, pp. 265–283.

[90] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., May 2015.

[91] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," *Remote Sens. Environ.*, vol. 205, pp. 338–351, 2018.

[92] G. Csurka, D. Larlus, F. Perronnin, and F. Meylan, "What is a good evaluation measure for semantic segmentation?," in *Proc. Brit. Mach. Vision Conf.*, 2013, pp. 32.1–32.11.

[93] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychological Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[94] C. Thiel, O. Cartus, R. Eckardt, N. Richter, C. Thiel, and C. Schmullius, "Analysis of multi-temporal land observation at C-band," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, pp. III-318–III-321.

[95] P. Lumsdon, S. R. Cloude, and G. Wright, "Polarimetric classification of land cover for Glen Affric radar project," *IEE Proc.—Radar, Sonar Navigation*, vol. 152, no. 6, pp. 404–412, Dec. 2005.
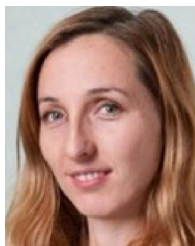
Sanja Šćepanović received the double M.Sc. degree in computer science from Aalto University, Espoo, Finland and the University of Tartu, Tartu, Estonia, in 2010, the Ph.D. degree in data science from Aalto University, Espoo, Finland, in 2018, and the Doctoral degree in innovation and entrepreneurship education from the EIT Digital Academy in 2019. She completed Space Studies Program from the International Space University, Strasbourg, France, in 2012.

She is currently, a Research Scientist with the Social Dynamics group, Nokia Bell Labs, Cambridge, U.K. Earlier, she was with ICEYE, Espoo, Finland, where the world's first microsatellite SAR instrument and SAR smallsat constellation were developed. Her research interests include health and human behaviour through data coming from various sociotechnical systems, such as social media traces and remote sensing.
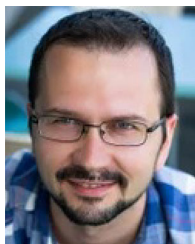
Oleg Antropov (Member, IEEE) was born in Almaty, Kazakhstan, in 1984. He received the M.S. degree in computer science and the candidate of sciences degree in radio physics from Dnepropetrovsk National University, Ukraine, in 2006 and 2010, respectively, and the Ph.D. degree in space technology from Aalto University, Espoo, Finland, in 2014.

Since 2010, he has been with the Remote Sensing Group, VTT Technical Research Centre of Finland, Espoo, Finland, as a Research Scientist and Senior Scientist. He was also a Postdoctoral Associate with Aalto University, during 2014–2018, and a Research Consultant with ICEYE, Espoo, Finland, in 2018. His research interests include the application of multiparameteric SAR data for various environmental applications, most centrally forestry, and crops.

Pekka Laurila is the Chief Strategy Officer (CSO) and co-founder of ICEYE, Espoo, Finland. As CSO, he has been instrumental in establishing and directing the company's strategy, and previously as CFO, raising the initial $20 M funds for ICEYE. He joined forces with ICEYE Co-founder Rafal Modrzewski to create SAR microsatellites to solve various large-scale issues the world faces, including those related to the ice monitoring for maritime safety, agricultural financing with objective data, and a wide range of further Earth observation analytics. He has deep domain expertise in geographic information systems. Prior to co-founding ICEYE, he played an instrumental role in Finland's Aalto University nanosatellite Program Aalto-1 .

Mr. Laurila was the recipient of the 2018 Forbes 30 under 30 Technology award based on the world-first achievements of ICEYE.

Yrjö Rauste was born in Espoo, Finland, in 1956. He received the M.S. degree in surveying and mapping, the Licentiate of Technology and the doctor of science (Tech.) degrees in photogrammetry and remote sensing from the Helsinki University of Technology, Espoo, in 1983, 1989, and 2006, respectively.

Since 1979, except for visits to other research centers and his military service, in 1983, he has been with the VTT Technical Research Centre of Finland, Espoo, Finland, where he is currently a Senior Research Scientist with the Remote Sensing Group. From 1986 to 1987, he was a Visiting Scientist with the Institute for Image Processing and Computer Graphics, Graz Research Center, Graz, Austria. From 1997 to 1999, he was a Postdoctoral Fellow with the Joint Research Centre of the European Commission, Ispra, Italy. His research interests include application of synthetic aperture radar image analysis and processing (particularly in forestry and land cover applications).

Dr. Rauste served as the Secretary of the Finnish Society of Photogrammetry and Remote Sensing, from 1994 to 1996.
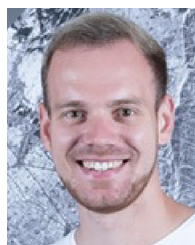
Vladimir Ignatenko received M.S. degree (with honors) in telecommunications from Saint Petersburg Electrotechnical Uiversity "LETI," Saint-Petersburg, Russia, in 2013. Since 2013, has been working toward the doctoral degree in signal processing with Aalto University, Espoo, Finland.

He has been a Research Engineer with ICEYE Oy, Espoo, Finland, since 2013. Since 2018, he has been SAR Imaging Tech Lead at ICEYE Oy. His research interests include SAR focusing algorithms and imaging modes, coordinating team of SAR experts on various projects on SAR data processing, quality assessment, SAR payload calibration, development of new imaging modes for ICEYE SAR constellation.

Jaan Praks (Member, IEEE) received the M.Sc. degree in physics from the University of Tartu, Tartu, Estonia, in 1996, and the Ph.D. degree in space technology from Aalto University, Espoo, Finland, 2012.

He is an Assistant Professor (tenure track) with Aalto University, Espoo, Finland, and a Leader of Aalto Space Technology and Microwave Remote Sensing research group. During his career, he has participated in numerous remote sensing projects and measurement campaigns, instrument developments, and led the development of Aalto-1, Aalto-2, Aalto-3, and Foresail nanosatellites. He is currently leading the Satellite Platforms team in the Finnish Centre of Excellence in Research of Sustainable Space. To date, his research projects have lead to several New Space and Earth Observation start-up companies in Finland. His research interests include microwave remote sensing, especially SAR Interferometry and SAR polarimetry, and nanosatellite technology.