Deep Learning for Regular Change Detection in Ukrainian Forest Ecosystem With Sentinel-2

Kostiantyn Isaienkov, Mykhailo Yushchuk, Vladyslav Khramtsov, and Oleg Seliverstov

***Abstract*—The logging is the leading cause for the reduction in the forest area in the world. At the same time, the number of forest clearcuts continues to grow. However, despite the massive scale, such incidents are difficult to track in time. As a result, huge areas of forests are gradually being cut down. Therefore, there is a need for regular and effective monitoring of changes in forest cover. The multitemporal data sources like Copernicus Sentinel-2 allow enhancing the potential of monitoring the Earth's surface and environmental dynamics including forest plantations. In this article, we present a baseline U-Net model for deforestation detection in the forest-steppe zone. Training and evaluation are conducted on our own dataset created on Sentinel-2 imagery for the Kharkiv region of Ukraine (31 400 km^2). As a part of the research, we present several models with the ability to work with time-dependent imagery. The main contribution of this article is to provide a baseline model for the forest change detection inside Ukraine and improve it adding the ability to use several sequential images as an input of the segmentation model.**

***Index Terms*—Change detection, convolutional neural network (CNN), deep learning, deforestation, logging, LSTM, optical imagery, semantic segmentation, U-Net.**

I. Introduction

FORESTS play a crucial role in climate regulation and carbon sequestration. Particularly, forests in Ukraine, covered 15% of territory, can fetch more than $\sim 200 \times 10^6$ tons of carbon dioxide from the atmosphere and emit $\sim 180 \times 10^6$ tons of oxygen [11]. But woodlands in Ukraine similar to most forests in other countries are at the risks associated with anthropogenic pressure. Illegal deforestation is carried out with next main objectives—for forestry and the sale of wood and for new farms or construction projects. Some forest areas have the highest conservation value as habitats of rare species of animals and plants and are protected by the Berne Convention [44]. However, Ukrainian forestry in most cases is not able to follow the requirements of the Berne Convention due to the imperfection of the legislative framework.

Official statistics of illegal logging in Ukraine shows a decrease from 10 000 cases of illegal logging per year to 5000 for the last 10 years' period [43], but independent observations indicate an increase in the number of violations [58]. At the same time, according to state and civilian estimates, the amount of illegal logging is increasing and, in 2019, amounted to more than 100 000 m^3. Moreover, the largest number of cases of illegal logging and volumes of illegally harvested wood accounted for just forest-steppe regions of Ukraine and not mountain.

One of the effective mechanisms to reduce illegal deforestation is the influence of civil activists and nonprofit environmental organizations. To support their activity, systems of operational monitoring of logging are actively developed (see, e.g., [4]). Systems of this kind are aimed to obtain data by the date and area of logging actually carried out, identifying inconsistencies, and taking preventive measures for violators, and identify illegal deforestation at the initial stages and use civil control capabilities to suspend.

Today, deforestation systems are mostly based on remote sensing, which established as an effective tool for forest change detection. Most earlier works dedicated to the analysis of deforestation detection with the aid of remote sensing used imaging data such as Landsat [41] and MODerate Resolution Imaging Spectroradiometer [42]. For instance, [19] presented 30-m spatial resolution maps of global forest change using data from Landsat mission; [20] applied machine learning techniques to the Landsat and MODIS data with the aim to map the forest cover and the forest change in Congo Basin. The work [5] presented a review of different approaches applied to the Landsat data for forest change monitoring.

Although spatial and spectral resolution offered by Landsat and MODIS is well-suited for analysis of the deforestation [13], [27], [30], [60], its temporal resolution is not appropriate for high-frequency forest change detection. While the Landsat data were limited by annual and subannual frequency of observations, the Sentinel-2 mission [15] may provide the high-resolution (up to 10 m) images every 5 d. The potential of Sentinel-2 in detecting small-scale deforestation regions was shown in [35] and [38], where Sentinel-2 imagery showed better performance than Landsat in detecting forest change in Amazon.

With remote sensing, deforestation monitoring systems use a bunch of automatic methods based on the change detection techniques[12], [21], [26], e.g., time-series analysis of vegetation indices [38], image preprocessing to detect changes between a pair of images [49], machine learning classification [24], [57], etc. But, recently, the deep learning has demonstrated great potential in remote sensing due to its ability to extract features from the spectral–spatial–temporal image data

Manuscript received June 17, 2020; revised September 8, 2020 and October 17, 2020; accepted October 20, 2020. Date of publication October 27, 2020; date of current version January 6, 2021. *(Corresponding author: Vladyslav Khramtsov.)*

Kostiantyn Isaienkov, Mykhailo Yushchuk, and Vladyslav Khramtsov are with the Quantum, 61072 Kharkiv, Ukraine (e-mail: k.isaienkov@quantumobile.com; m.yushchuk@quantumobile.com; v.khramtsov@quantumobile.com).

Oleg Seliverstov is with the V. N. Karazin Kharkiv National University, SCGIS Ukraine, 61022 Kharkiv, Ukraine (e-mail: oleg.seliverstov@physgeo.com).

Digital Object Identifier 10.1109/JSTARS.2020.3034186

[40], providing state-of-the-art results in remote sensing change detection.

In [34], the authors demonstrate general principles of the convolutional neural networks (CNNs) usage for the change detection problem on satellite images. In the recent work of [46], the authors introduce the integration of U-Net [50] model with recurrent models for the urban changes detection on the Sentinel-2 data. Scores obtained in the experiments show that deep learning models can deal with the problem of change detection based on the satellite data.

Another type of architecture developed to work with Sentinel-2 imagery is present in [62]. Originally, it was used for land cover classification but can be converted into a form suitable for the change detection problem. The input for the neural network is a 3-D stack of sequential images that represent the time-dependent series from the common location.

In the work [29], the instance segmentation models, namely region-based CNN [23] and a variation of U-Net model, were used to detect urban changes on the high resolution aerial images. The models based on Siamese CNN to speed up the change detection on satellite images were presented by [10].

Despite the fact that deep learning is a quite popular tool in change detection from remote sensing, the number of studies related to the application of CNNs to the deforestation detection is relatively small. Most of these works are dedicated to the Amazonian rainforests—most cut down forest in the world [2], [3], [37]. In [13], the better performance of deep learning (SharpMask [47], U-Net, and ResUNet [61]) models with respect to classic machine learning (random forest [7] and multilayer perceptron [25]) algorithms to track the change detection of deforestation in the Amazon using Landsat data are presented. The work [45] reviewed several deep learning methods like Early Fusion CNN model and Siamese CNN model [10] for deforestation detection in Amazon forests.

In this work, we present the results of high-frequency forest change monitoring within the Ukraine region using data from the Copernicus Sentinel-2 mission. The aim of this study is to set first result in remote sensing deforestation detection in Ukraine. To do so, we constructed large dataset of deforestation within some regions of Ukraine. We solved the image segmentation task, which aims at classifying each pixel of entire image; in our case, we tried to distinguish pixels corresponding to the forest change from the background. By solving the segmentation task, we investigate two set of CNNs—time-dependent and single-image based models—to check the ability of deep learning models to utilize the temporal information in segmenting forest changes. Also, our work is the first research that compares different variants of time-dependent CNNs to detect deforestation from high-resolution and high-frequency data. Namely, we used and evaluated seven models (based on the U-Net network) that utilize the temporal information and predict the change of the deforestation regions. The proposed models could detect the forest change within extremely short timescales (up to temporal resolution of satellite data, equal to 5 d) due to the construction of the manually created training dataset.

The structure of rest of this article is as follows. The description of the dataset used for training of the models as well as the description of models are presented in Section II. In Section III, we describe the results of forest change detection based on our datasets. In Section IV, we provide the discussion of our results as well as the principal limitations of the models. The conclusions are presented in Section V.

II. Materials and Methods

As we note in Section I, there is a wide list of methods, able to detect forest changes in remote sensing images. We solve this task in a supervised learning manner—having some dataset of training observations with known ground truth deforestation regions, we build the model, which maps the remote sensing images into the deforestation images. As such a model, we used deep artificial network which is able to train on remote sensing data to predict changes with a good quality (see references in Section I for example). To train this model, we construct our own dataset of deforestation regions in particular regions of Ukraine. Further in this section, we describe the pipeline of our forest change detection methodology.

A. Dataset Labeling

There is no large-scale public research in the field of forest change detection inside Ukraine territory; therefore, it is hard to find a good dataset to set experiments and to train models. So we decided to collect our own dataset. We selected the most typical forest types from Ukrainian territory. We work with only the Kharkiv region for this purpose. The results are applicable to the entire forest and steppe zone which makes up most of the territory of Ukraine. The exception is only the Crimean and Carpathian mountains which require additional labeling due to their landscape and trees species. But it is out of the scope of this article and we do not focus on it. We selected two neighboring Sentinel-2 tiles (36UYA and 36UXA) within the period from 2016 to 2019. Each tile comprises $\sim 12\,000$ km^2 in area. They have a maximum forest area, including the most typical forests: coniferous, broad-leaved and mixed the woods. On these tiles, forest areas different by size (from small to large) are presented (see Fig. 1).

Labeling was performed using the entire sequence of Sentinel-2 [15] images from selected tiles, with cloud cover up to 20%. We used ArcGIS Pro and QGIS for this purpose [48]. The initial data sources are Sentinel-2 image services [1] and locally stored images in the combination of natural color channels. For the refinement, the additional composites and indices [like infrared bands, normalized difference vegetation index (NDVI), and normalized difference moisture index (NDMI)] were used.[1] The labeling was carried out manually by analyzing a series of three or more sequential images. Each new section of increasing a clearcut was marked as a separate polygon; the output for the masks were geoJSON files. If the clearcut fell on a tile with clouds, this image was not used to mark the clearcut. Clouds labeling was not performed since it is available in Sentinel-2 Level A products, as well as could be done automatically with

[1]Some of these channels are used for deep learning segmentation; for more details see Table I.

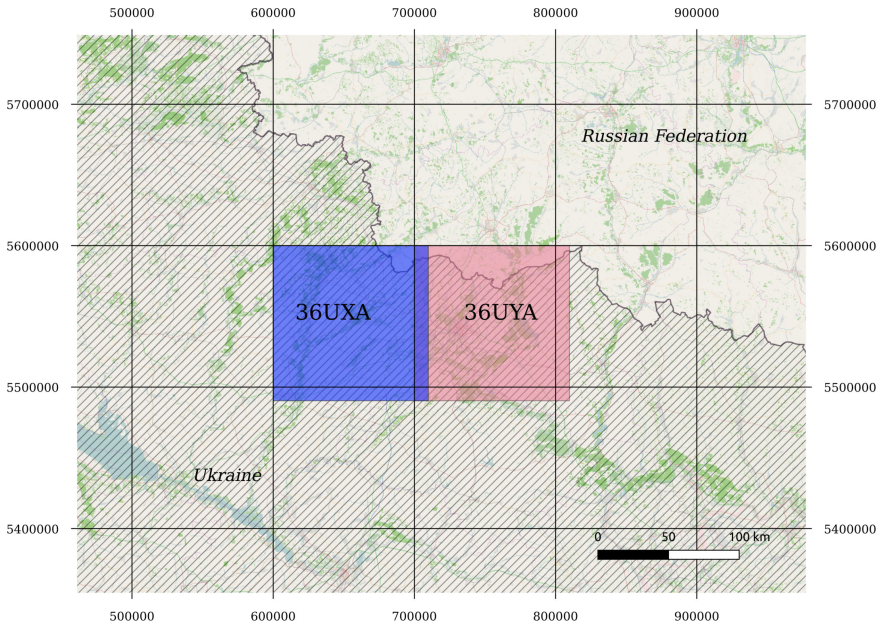

Fig. 1. Location of the study areas (tiles 36UXA, blue, and 36UYA, red) within the region of Ukraine (dashed area). Names of countries are provided within corresponding regions. The map is in the WGS-84 UTM zone 36N projection.

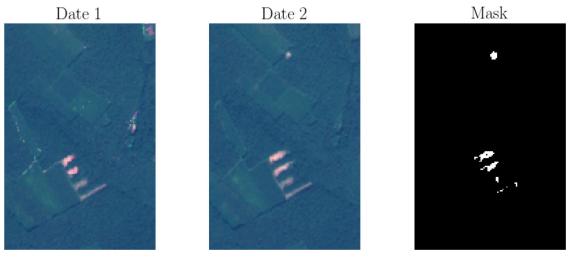

Fig. 2. Example of images from the dataset. The figure consists of two images with neighbor dates and a corresponding mask with clearcutting changes for this period.

using Sentinel-2 cloud detector [52]. The example of labeling procedure output is available in Fig. 2, where we show the two serial images of the region with clearcuttings and the corresponding masking of deforestation dynamics.

During the labeling, the main types of inconsistencies that we faced were the following:

1) missed clearcutting (main reason: small area);
2) clearcutting is marked on a similar object that is not logging (meadow or haymaking);
3) the clearcut of less than the threshold area is marked (within the precision of georeferencing of images);
4) insufficient positional accuracy of borders (less precision of georeferencing of images); and
5) the date for clearcutting was incidentally determined incorrectly.

The quality of the labeling is controlled by the following methods:

1) repeated selective verification;
2) cross-check of the labeling by different annotators; and
3) verification of the results of segmentation errors after training the model.

As a result of the labeling procedure, we obtained a dataset with a general area of ∼50 000 hectares. The total time spent on the whole labeling procedure is ∼800 h. During the labeling, we found out that the number and area of felling are evenly distributed over all seasons of the year, with a slight dominance of clearcuttings in spring and summer.

B. Data Preprocessing

In our research to probe the possibility of deep learning forest change detection, we used about 4000 polygons, with an area of ∼30 000 hectares. The reason for using a subsample of the labeled region only is the availability of high-frequency monitored clearcuts within it, while other regions are labeled rarely. Selected subsamples of the regions contain a large cumulative number of clearcuts, most of which were monitored every 5–10 d, which gives an advantage in training time-dependent models. As we describe below, we analyze two sets of models (baseline and time-dependent), and these sets use the following labeling results:

1) baseline: 2318 polygons, tile 36UYA and tile 36UXA, 2016–2019 years, 26 images in total; and
2) time-dependent: tile 36UYA (two samples of separated annotations—for spring and summer, with number of polygons equal to 278 and 123 respectively, 2019 year) and tile 36UXA (1404 polygons, 2017–2018 years), 36 images in total.

As the input images for the segmentation models we use (see Table I) the following:

1) True color images;
2) B8, B8A, B11, and B12 bands;

TABLE I
CHANNELS PRESENT IN THE DATASET

Channel	Resolution	Description
True Color Image	10 m px^{-1}	RGB image built from the B02 (Blue), B03 (Green), and B04 (Red) Bands.
B8 band	10 m px^{-1}	Visible and Near-Infrared (VNIR) wavelength ranges. B8 band is good for mapping vegetation. [32, 59]
B8A band	20 m px^{-1}	Narrow VNIR band. B8A band is good for vegetation classification. [6]
B11 band	20 m px^{-1}	Short Wave Infrared (SWIR) band. B11 band helps to classify different types of vegetation. [6, 33]
B12 band	20 m px^{-1}	SWIR band. B12 band helps to measure the soil and vegetation moisture maintenance. [16]
NDVI	10 m px^{-1}	Vegetation Index for quantifying green vegetation [51]: $\text{NDVI} = \frac{\text{B8}-\text{B4}}{\text{B8}+\text{B4}}$
NDMI	20 m px^{-1}	Vegetation Index for identification vegetation water stress problems [17]: $\text{NDMI} = \frac{\text{B8A}-\text{B11}}{\text{B8A}+\text{B11}}$

3) NDVI and NDMI.

The description of the Sentinel-2 bands is available in [56, Section 3.10].

All images are Level C products[2] and have the same resolution—10 m per pixel. The same resolution of the entire channels was provided by the resampling of low-resolution channels to the target resolution.

After downloading images, which correspond to the dates of the entire dataset polygons, we converted polygons from the vector representation into the raster one. During conversion, we made separate masks of deforestation, each of which is corresponding to the one image (in the sense of spatial and temporal information).

To prepare data for the baseline model we need to crop each Sentinel-2 image ($\sim$10 000 pixels per axis) with the corresponding clearcuts and cloud masks into 224 $\times$ 224 fragments. The chosen size of the fragments is a most common size of the input images for the deep learning models. After we remove the fragments, which are covered by clouds (coverage $> 70\%$, which was determined directly with `sentinel2-cloud-detector`) and contain no masks (except the pieces, containing field crops), it is important to save them due to the visual similarity of fields with cutting areas.

For the models with time dependency, we crop each Sentinel-2 image with the corresponding clearcuts and cloud masks into 56 $\times$ 56 pixel fragments. We decided to use smaller size of images for the time-dependent models in respect to the size of images for the baseline model due to the relatively small area of forest changes. During our experiments, we noted that such small masks of forest changes on the 224 $\times$ 224 images resulted in the very small training dataset (in order of 100 images) and forced our models with time dependency to predict empty masks. Also due to the small masks of clearcut changes, we artificially increase the size of masks to about one pixel from each border, to help the models converge faster and neatly. As in the case of the baseline, we drop fragments with clouds such that we include in the one particular sequence cloudless images only, until completing the sequence.

To ensure the preservation of time dependence, we form the input datasets as sequences of images which are separated by time at 30 d in maximum. Sequences of images are transformed using histogram matching for each particular channel to make the images in a group most similar. This transformation provides an opportunity to detect changes in clearcuts easier, ignoring seasonal changes which are important in subannual clearcut detection [18].

Due to the high cloud cover and the small number of clearings in the winter, we removed winter fragments from the dataset.

The described data preprocessing pipeline is represented as a flowchart in Fig. 3. After all manipulations, we got datasets that consist of 4700 fragments for the baseline models, 6500 samples for models accepting sequence of two images, and 1500 samples for UNet-LSTM, which require the sequence of five images as the input. For the training procedure, we split data into the train-valid-test parts, and got 3812, 559, 383/4825, 1098, and 681 fragments in training, validation, and testing splits respectively for baseline and time-dependent models. The data were split by using the spatial coordinates rather than random splitting, meaning that the region, which occur, for example, in the training sample, will not be included in test and validation datasets. Each fragment consists of 1) one image in nine channels for the baseline models and 2) the sequence of images, where images are separated by 5–30 d in time, in nine channels for the time-dependent models. The dataset was constructed in this way to train our models to detect clearcuts within extremely short time scales. This is quite a natural way to provide the certain ability for the model via training sample because the deep learning models are very sensitive to the training data points.

[2][Online]. Available: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products

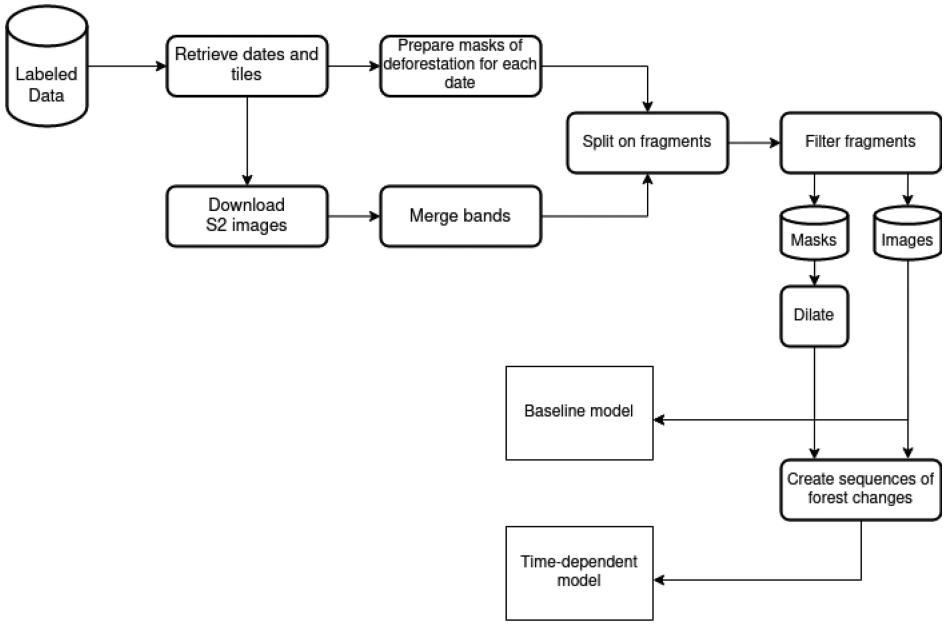

Fig. 3. Workflow of the data preparation for segmentation with baseline and time-dependent models.

Thus, we believe that our models could deal with detecting forest change regions every 5–30 d.

C. Quality Assessment

The segmentation quality metric is Dice coefficient [14]:

$$\text{Dice}(y, p) = \frac{2\sum_i y_i p_i}{\sum_i y_i^2 + p_i^2} \tag{1}$$

which represent the similarity between two binary images (the mask of clearcuts y, and the predicted regions of deforestation p in our case). The summation $\sum_i$ is occurred over all pixels of images. In the case of ideal alignment between these two binary images, we will get $\text{Dice}(y, p) = 1$; otherwise, in the worst case (if two binary images are in mismatch), the Dice score will be equal to zero.[3]

Although the Dice coefficient is a popular metric for segmentation models evaluation, it could not deeply describe the situation with the segmentation of images without clearcut changes. In the case if it is not critical to obtain exactly the full area of the clearcut but important to get the approximate location of the problematic zone, we have to implement another quality metric. This metric has to be focused on calculating the number of clearcuts, detected correctly and incorrectly.

In general, we expect the following four types of detections:

1) true positive (TP)—the clearcut instance was detected successfully;
2) true negative (TN)—there is no actual clearcut instance on the mask and the clearcut was not detected;
3) false positive (FP)—there is no actual clearcut instance on the mask but it was detected by model;
4) false negative (FN)—there is an actual clearcut instance on the mask, but it was not detected.

The next combination of the numbers of these detections defines the F1-score:

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{2}$$

The Dice score is the same as F1-score in terms of detections types, but, in the first case, we compare the predicted and ground truth values in each pixel of images, and, in the second case, we work with numbers of instance detections of different types.

D. Neural Network Models Description

To understand the possibility of forest change detection and evaluate the accuracy, we prepared the two sets of models. We include in the first set the models, which work with images in spectral–spatial dimensions, and do not accept temporal information; these models were trained to find clearcuts on each image. The second set is composed of the time-dependent models, which were trained to find the changes of deforestation regions in sequence of images. We name these sets "baseline models" and "time-dependent models," respectively.

Necessity of introducing these two sets is, in our aim, to investigate the possibility of deep learning models performing with additional temporary information to detect deforestation. Also, assuming a wide range of available deep learning models, we can find the best model within a set of baseline models and use it for the subsequent analysis, adding the dependency on time.

To analyze a set of baseline models, we implemented several experiments. We use ResNet-50 and ResNet-101 [22] networks wrapped into the U-Net and feature pyramid network (FPN,

[3]We note that with such a definition, Dice metric degenerates when both images contain zero pixels only. Practically, we implemented a slightly different equation of the Dice score, considering this degenerate case: $(2\sum_i y_i p_i + \varepsilon)/(\sum_i y_i^2 + p_i^2 + \varepsilon)$, where ε is a small additive ($\varepsilon \ll 1$). This representation results in $\text{Dice}(y, p) = 1$ if $\sum_i y_i = 0$ and $\sum_i p_i = 0$.

TABLE II
LEARNING TIME AND NUMBER OF BEST EPOCH FOR TIME-DEPENDENT MODELS

Model	Time per epoch, sec	Best epoch
‘ UNetDiff	22	93
UNetCH	28	51
UNet2D	23	133
UNet3D	44	107
SiamConc	27	118
SiamDiff	27	177
UNet-LSTM	44	31

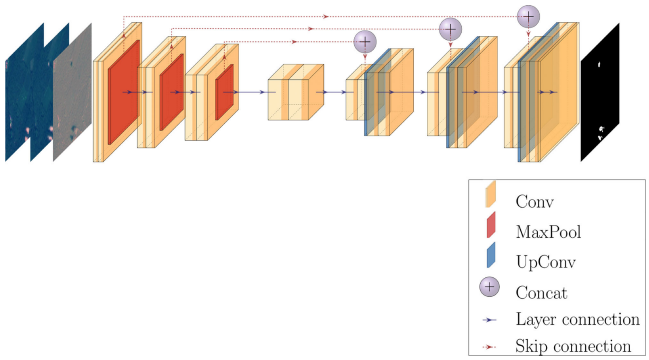

Fig. 4. Schematic representation of UNet-diff model.

[36]) architectures to obtain scores for the baseline. The chosen architectures are drawn from the one family of neural networks, which is called autoencoders. Autoencoders compact through convolutional filters and downsampling the input images into the latent space (the so-called encoder network) to learn the low-level features and then upsample these latent features to the new image with an original shape (via decoder network). For the segmentation task, autoencoder should have additional classification layer, which transforms output image of decoder into the mask, which should be restored from input image. The main advantage of the U-Net and FPN is that these architectures use connections between encoder and decoder—not only through the layer, which form the latent features, but also between corresponding layers at each level of sampling (for example, see schematic representations of the time-dependent models). These connections combine low-level and high-level features to infer the mask details more accurately and to take into account learnt latent features as well as the high-level features of image. This property is a principal reason for the picking up these architectures for our analysis.

Wrapping the chosen ResNet CNNs into architectures, we also use different initial learning rates and optimizers (see Table II in Section III for more details, [31]).

The baseline models do not take into account the time dependence in the clearcuts. In the next experiments, we use several approaches for the training of neural networks considering the time dependence. All of the networks described below are wrapped into the U-Net model. These networks use Adam optimizer (200 epochs of training in total, starting learning rate equal to 10^{-2}, and further learning rate decreasing on the factor of 0.1 at 10, 40, 80, and 150 epochs), with balanced batches (with regard to number of images with empty masks and without), size of which is equal to 64.

To increase the generalization ability of all of the models, we used data augmentation procedures, applied to the images from training sample:

1) random crop of images (70% of image size in each axis with further rescaling to the initial size), random brightness, and contrast changing;
2) elastic and grid transformations (for more details, see [9] and [54]); and
3) mask dropout (zeroing out mask/image regions corresponding to instance at mask),

which were applied with some predefined probability for each training image.

As we noted in the previous section, we use Dice score to measure the quality of our deep learning segmentation. But, in addition to the monitoring of the Dice score with the aim to analyze the quality of CNNs, we can optimize the model with respect to this score. We can use the Dice score as the loss function, but instead of maximizing it, the model should minimize the “inverse” Dice score: $1 - \text{Dice}(y, p)$. So, for the loss function for our models, we used the weighted sum of binary cross-entropy and Dice losses [39]:

$$L_{\text{BCE}}(y,p) = -\sum_i \left(y_i \log(p_i) + (1-y_i)\log(1-p_i)\right) \quad (3)$$

$$L_{\text{Dice}}(y,p) = 1 - \frac{2\sum_i y_i p_i}{\sum_i y_i^2 + p_i^2} \quad (4)$$

$$L(w_1, w_2) = w_1 L_{\text{BCE}} + w_2 L_{\text{Dice}} \quad (5)$$

where y, p are ground truth and predicted masks, respectively, and summing is held over all pixels. Hereafter, for all models, which use the weighted sum of binary cross-entropy and Dice losses, we assume $w_1 = 0.2$ and $w_2 = 0.8$.

We assume that segmentation quality of the model is the best when the loss on the validation dataset is the highest. We also note that we used a full validation dataset for the first two models from the list below, and the remaining models were validated with using images with nonempty masks only. This specific selection of validation data allowed us to obtain satisfactory evaluation results on the test sample for all of the models. The list of models is presented in the following.

1) U-Net model trained on image difference and concatenation (UNet-diff). The model is based on the ResNet-18 architecture. The depth of the encoder and decoder is equal to 3 (see Fig. 4).
   The input for the UNet-diff model is a pair of consecutive images and their difference. The output is a mask of differences in clearcuts between two images from a sequence. All input fragments are concatenated by the channel axis, resulting in the 27-channel data-in. To train the UNet-diff network, we use Adam optimizer which minimizes the weighted sum of BCE and Dice losses [see (5)]. The advantage of this network is that it uses as the input not only the sequence of images but also the

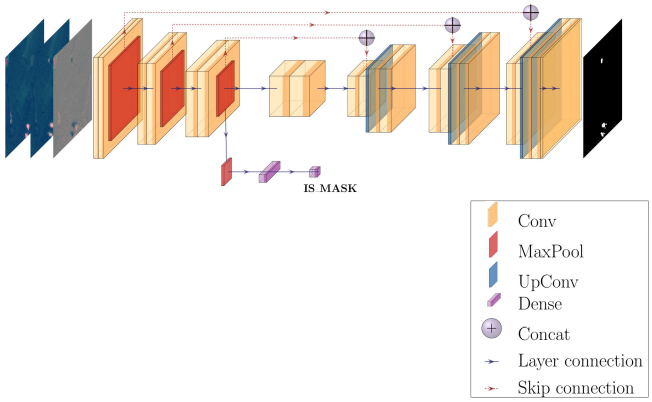

Fig. 5. Schematic representation of UNet-CH model.

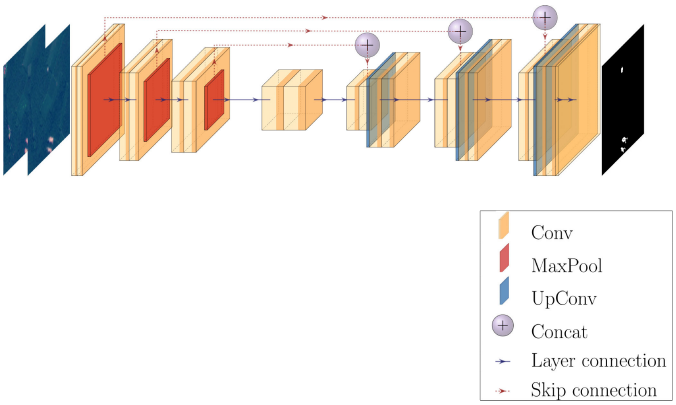

Fig. 6. Schematic representation of UNet2D model.

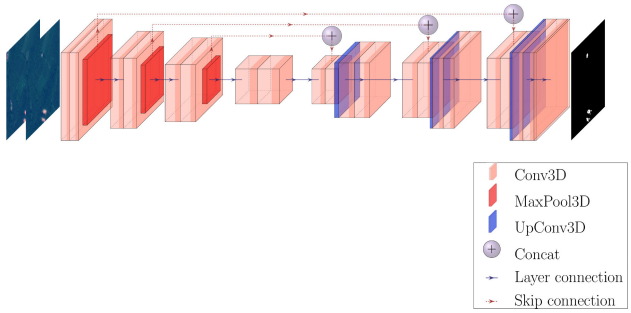

Fig. 7. Schematic representation of UNet3D model.

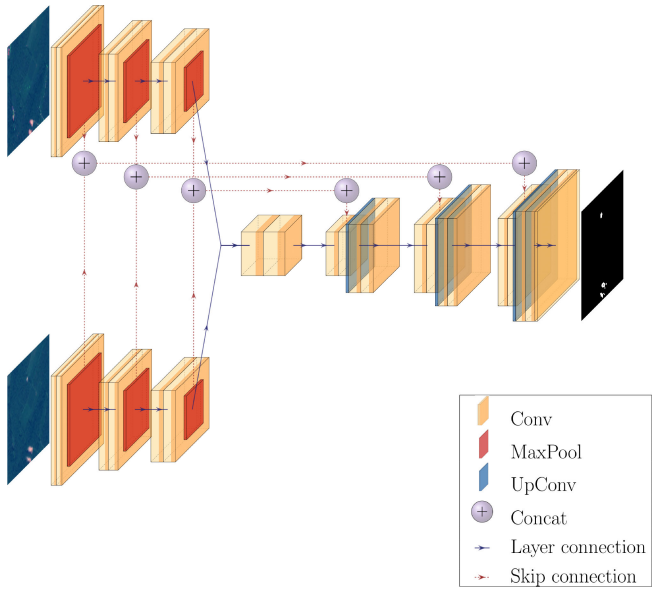

Fig. 8. Schematic representation of SiamConc model.

difference between them, which directly represents the changes in images.

2) U-Net model with classification head trained on concatenation of two images from sequence and image difference (UNet-CH). This model has the construction similar to the UNet-diff, excepting the auxiliary classification head (Fig. 5). Classification branch is placed at the top of the encoder and consists of the max-pooling layer followed by the flattening and the dense layer with one output, expressed via sigmoid activation function. Classification head predicts if an input stacked image contains forest changes or not. The network is trained with the same loss function as the UNet-diff model [see (5)]. UNet-CH requires the same input images that are presented for UNet-diff model and predicts the forest change difference between two images from the sequence. All input fragments are concatenated by the channel axis, resulting in the 27-channel images.
3) U-Net model trained on concatenation of two images by channels axis (UNet2D, [10]). UNet2D model contains three max-pooling and three upsampling layers. Also the Dropout [55] technique is adapted to this network after each convolutional layer with the dropout rate equaling 20% (Fig. 6). UNet2D is trained with a weighted sum of BCE and Dice losses [see (5)]. The input is a pair of consecutive 9-channel images, concatenated with each other by channel axis (producing 18-channel images) and the prediction is the difference of masks of two images.
4) 3-D U-Net model (UNet3D). It takes as the input pair of images, concatenated in the new depth axis. UNet3D model is the model same as UNet2D, excepting the replacement of 2-D operations (convolution, pooling, etc.) with their 3-D equivalents (with using no pooling across the new depth axis, Fig. 7). This network was trained with Adam optimizer, minimizing the Dice loss function [see (4)]. The UNet3D requires as the input two nine-channels images, which are stacked into a single $2\times9\times56\times56$ image.
5) Siamese U-Net network with concatenating the skip connections (SiamConc). It consists of two encoders and one decoder, depth of which is equal to four layers (Fig. 8). Encoders have shared weights, as in the usual Siamese network. The skip connections in SiamConc are organized such that the concatenation of the layers of encoders feeds to the decoder layers. SiamConc was trained with the Dice loss function [see (4)]. The input for each encoder is a nine-channel image from a consecutive pair. The output is the difference of masks of two images.

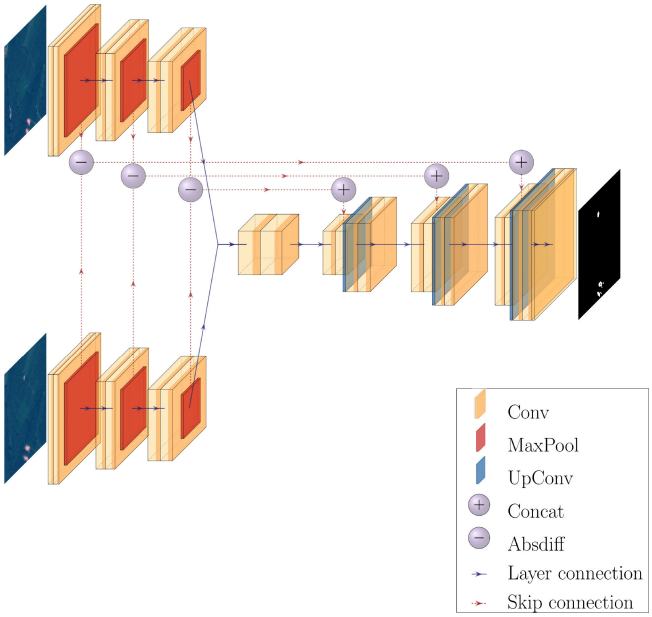

Fig. 9. Schematic representation of SiamDiff model.

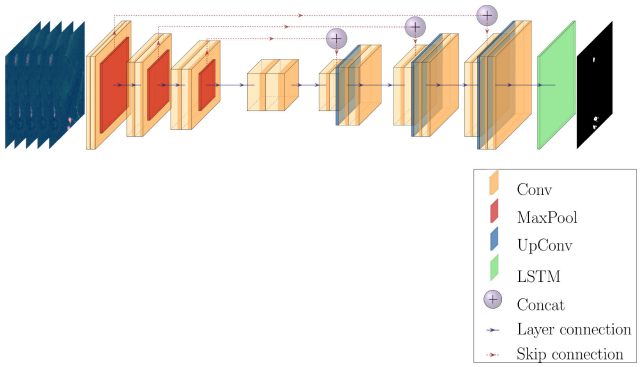

Fig. 10. Schematic representation of UNet-LSTM model.

6) Siamese U-Net network with concatenating the absolute values of difference from two encoders at skip connections (SiamDiff). This network has the same architecture as SiamConc but with the absolute differences between connections of both encoders before stacking with the decoder parts. We used the Dice loss function [see (4)] to train the SiamDiff model (Fig. 9).
7) U-Net LSTM model (UNet-LSTM). Presented as a model similar to UNet2D with addition of a convolutional LSTM unit [53] to the decoder's output (Fig. 10). The input is a sequence of 5 nine-channel images, and the output, unlike other time-dependent models, is the last mask from the sequence. Due to the higher length of input images sequence regarding other models, UNet-LSTM has the smallest training dataset. This model also was trained with a weighted sum of BCE and Dice losses [see (5)].

All these models are different in the sense of used architecture and input data. Namely, UNet2D is very similar to the UNet-diff and UNet-CH models, excepting the physical sense of the input data. While UNet2D is forced to find the forest changes itself, UNet-diff and UNet-CH models are able to use information about the difference of two sequenced images in predicting the mask. UNet-CH also has auxiliary classification branch, which, as we assumed, should help in creating more reliable latent feature space and, as a result, in generating more accurate predictions. UNet3D, in its turn, should try to find forest changes in the manner different from the UNet2D due to the specific construction of spectral–spatial–temporal convolutions. The advantage of the Siamese models is manipulating with images in different branches but with shared weights; it means that these networks should try to find forest changes comparing latent and downsampled features from these two branches. And the last model, namely UNet-LSTM, is drawn from the family of recurrent models; these models have been established as a great tool to process sequence of signals.

Neural networks were implemented with `PyTorch`.[4] To train the models, we used GPU GeForce GTX 1080Ti. The principal characteristics of learning time-dependent models (time of learning per epoch, and the number of the best epoch) are given in Table II.

We can see that the most optimal learning time was reached with UNet-diff and UNet-CH models, which show relatively low (< 30 s epoch^{-1}) learning time and fast convergence (< 100 epochs).

III. Results

In this section, we discuss the results of forest change detection algorithms in use. We summarize the results in the terms of segmentation quality metrics on the training, validation, and testing samples.

We have analyzed the segmentation results with the Dice coefficient. According to Table III, the best baseline network on the test score is U-Net model with the ResNet-50 backbone with learning rate equal to 10^{-3} and SGD as optimizer. We can see that the second best model according test score and the best model according to the validation score is also UNet-50 with learning rate equal to 10^{-3} but with Adam optimizer. From this, we can see that the UNet-50 model is the most suitable architecture for this task. The best Dice score is equal to 0.46 on the test sample and 0.52 on the validation one.

In the analysis of models with time dependency, we separate two cases of analyzed data: nonempty masks (corresponding images of which have signs of changing clearcuts) and all masks (empty and nonempty). This was done in order to analyze the segmentation quality, looking from the two sides of the problem: segmentation of actual clearcuts and ability of the model to differentiate between images with and without clearcuts.

Table IV shows that the best validation/test Dice score on nonempty masks can be achieved by the UNet-diff model, while the UNet-CH model introduces the highest Dice score on all masks. It means that by adding the classification head to the UNet-diff model, we have lost some segmentation quality of clearcuts, but, at the same time, we have won in differing images

[4][Online]. Available: https://pytorch.org/

TABLE III
SCORES FOR THE BASELINE MODEL SELECTION (SET BY MODEL TYPE, OPTIMIZER, AND LEARNING RATE). THE SCORE FOR THE TRAIN, VALIDATION, AND TEST DATASETS IS DICE.

Architecture	Optimizer	lr	train	valid	test
U-Net50	Adam	10^{-2}	0.5958	0.4973	0.4403
		10^{-3}	0.6513	**0.5214**	0.4636
		10^{-4}	0.7401	0.4464	0.424
	SGD	10^{-3}	0.6528	0.5193	**0.4651**
U-Net101	Adam	10^{-2}	0.6010	0.4410	0.4556
		10^{-3}	0.6466	0.4919	0.4275
		10^{-4}	0.7549	0.3945	0.4271
	SGD	10^{-3}	0.6663	0.4718	0.4451
FPN_50	Adam	10^{-2}	0.6279	0.4747	0.4159
		10^{-3}	0.639	0.497	0.445
		10^{-4}	0.641	0.469	0.4104
	SGD	10^{-3}	0.624	0.482	0.4447
FPN_101	Adam	10^{-2}	0.577	0.479	0.4341
		10^{-3}	0.618	0.502	0.4408
		10^{-4}	0.642	0.462	0.4179
	SGD	10^{-3}	0.563	0.48	0.4279

The Best Result on Validation and Testing is Highlighted in Bold.

TABLE IV
DICE SCORES FOR ALL MODELS WITH TIME DEPENDENCY, SEPARATELY FOR ALL MASKS AND NONEMPTY MASKS

Masks type	Model	train	validation	test
All masks	UNet-CH	0.9086	0.8085±0.0232	0.8600±0.0202
	UNet-diff	0.9125±0.0071	0.7983	0.8406
	UNet2D	0.7676	0.7252	0.7932
	UNet3D	0.7792	0.7366	0.7968
	SiamConc	0.7660	0.7354	0.7904
	SiamDiff	0.7643	0.7328	0.8193
	UNet-LSTM	0.4025	0.4508	0.4155
Non-empty masks	UNet-CH	0.6859	0.3807	0.5185
	UNet-diff	0.7171±0.0201	0.3838±0.0478	0.5534±0.0607
	UNet2D	0.4115	0.3364	0.4363
	UNet3D	0.4412	0.3611	0.4406
	SiamConc	0.3632	0.3008	0.3697
	SiamDiff	0.2732	0.1884	0.2895
	UNet-LSTM	0.4052	0.4548	0.4013

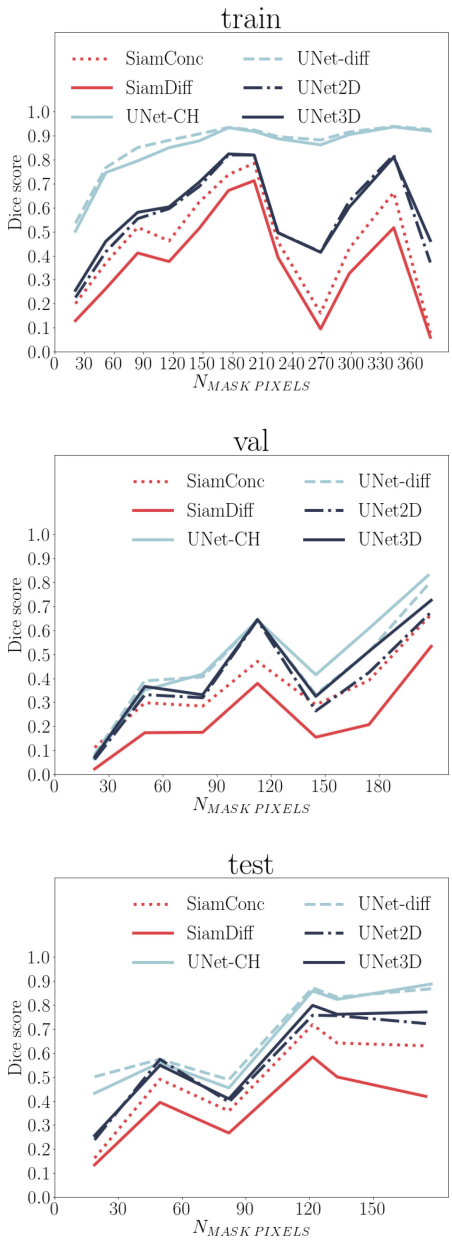

Fig. 11. Dependence of Dice score on the number of masked pixels for the training (top), validation (middle), and testing (bottom) samples for all of the models.

with and without changing clearcuts. We note that UNet-diff and UNet-CH models performed better than baseline models described in the previous section. This could be explained by the fact that these models receive as the input not only the images at each date but also their difference, which directly represent the changed regions. Other networks, which work with the sequences of images without auxiliary image differences, could not infer the regions of the forest changes as better as UNet-diff and UNet-CH and, thus, have a lower segmentation quality. Possibly, we do not have enough training data to train our models in finding the changes without difference images.

Another notable result is that UNet-LSTM estimates the clearcutting regions with lower quality, despite the fact that output of this model is the whole clearcut, not its change. Possibly, this fact could be explained by the smaller amount of the training data in use.

We also analyzed the Dice score as a function of clearcut size in pixels for the nonempty masks, and we noted that the Dice score increases with the number of masked pixels. For the best-performed models (UNet-diff and UNet-CH), the Dice score could reach values $0.4 \div 0.5$ for the smallest ($\sim$20 pixels) and $0.8 \div 0.9$ for the largest ($\sim$180 pixels) clearcuts in the testing dataset (see Fig. 11).

To measure the F1-score, we have calculated the number of detections of all types for each predicted or ground truth instance. We assume that a single detection is TP if the corresponding intersection over the union (IoU) score between predicted and ground truth instances is higher than some predefined thresholds.

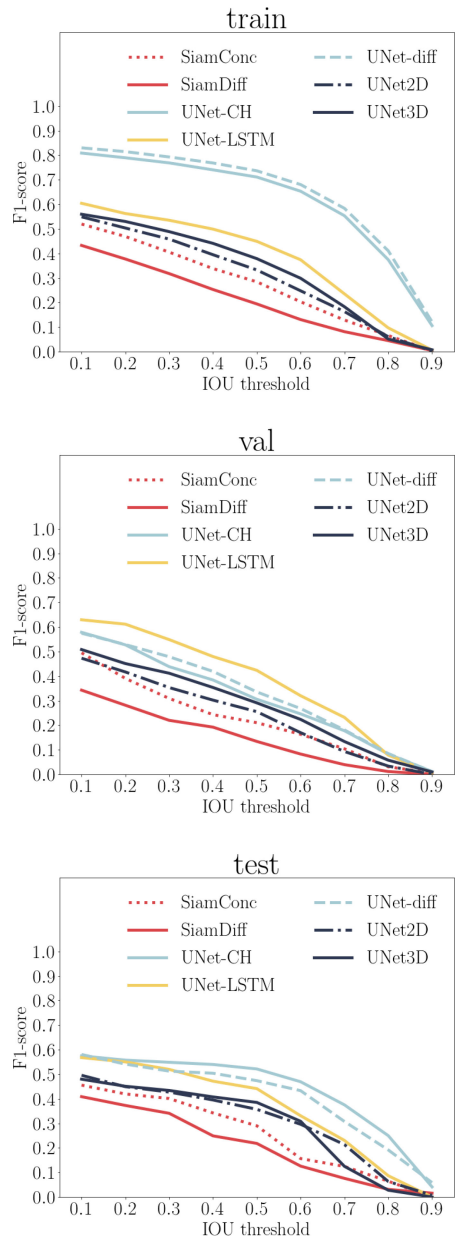

Fig. 12. Dependence of F1-score on IoU threshold (right) for the training (top), validation (middle), and testing (bottom) samples for all of the models.

We have collected such statistics for different IoU thresholds that were ranging between 0.1 and 0.9 with a step equal to 0.1. As we can see in Fig. 12, the F1-score for all of the models is strongly dependent on IoU threshold, reaching relatively high values at low IoU thresholds. This could be explained by accidental segmentation of the full clearcut on image rather than its masked changing possibly due to the artificial increase of the mask size.

Also we observe the overfitting of UNet-CH and UNet-diff models, which show a great performance on the training data but have worse results on the test dataset. But despite these facts, we conclude that UNet-CH and UNet-diff models, which show highest Dice score F1-score values, are able to provide a high-frequency monitoring of forest changes with relatively high quality. The representative comparison of models output is shown in Fig. 13 for randomly selected images. With this comparison, we can see the relative segmentation quality of the models in use and observe the good enough segmentation with UNet-diff and UNet-CH, which, unlike other ones, were caught in all logging regions.

IV. Discussion

A. Limitations of the Models

Some restrictions to our models are apparently caused by our training sample that does not fully reflect the actual out-of-sample data. Such limitations are in spatial sampling, temporal properties of the tracked changes, and seasonal effects.

As we mentioned in Section II, the data which were sampled for our investigations cover only the type of forests inherent to the forest-steppe zone; thus, the utilization of the trained models to detect the clearcuts within other forest types could be doubtful.

The next important limitation of our models is a seasonal one. We note the strong inconsistency between Dice scores of the models with different seasons: The score, in average, tends to be the lowest in spring and the highest in summer for the testing and validation samples (see Fig. 14). Also, our data include no images from the winter season that could result in insecure segmentation of winter forest changes.

Another limitation of our models was spotted during observing the clearcuts predictions on the whole tiles. We noted that models are not able to ignore some types of changes on images (e.g., ripening crop fields, shadows of the clouds, or changes caused by geographical misalignment of two tile frames) and segmentation output of the models for such cases was nonempty. Therefore, the utilization of the models trained with our dataset should be done with sophisticated postprocessing, such as filtering predictions with the maps of land cover (see, e.g., [8]), cloud shadows detections, etc.

One more limitation of trained models is the timing of forest changes. Our sampling of images was constructed in such a way that a maximum time interval between a pair of images does not exceed ∼30 d. But due to the different conditions (overcast mostly), the pair of images of the investigated forest area could be obtained within a larger period of time that was not presented for our models. And thus the quality of change segmentation over the large time intervals between images in sequence might be inconsistent with testing results obtained in this work mostly due to the different seasonal effects which we described in Section II.

B. Perspectives and Monitoring System Deployment

Importance of the deforestation detection problem could not be understated. But, at the same time, the community is limited in developing forest change monitoring systems and methods due to the lack of enough training data. Thus, we make publicly available the dataset with annotated deforestation regions, which was presented in this article. In future, we plan to increase the number of annotated forest changes within Ukrainian territory,

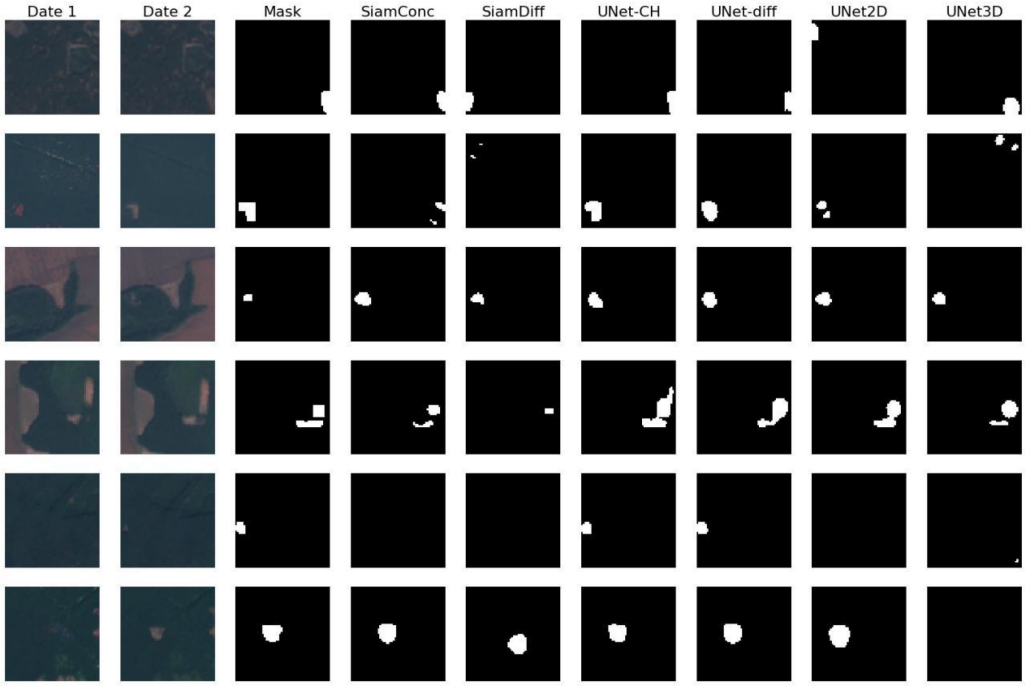

Fig. 13. Predicted masks produced for several images from the testing sample. First three columns represent time-separated images of some region and corresponding changing mask (third columns), and other columns correspond to the predictions of different models.

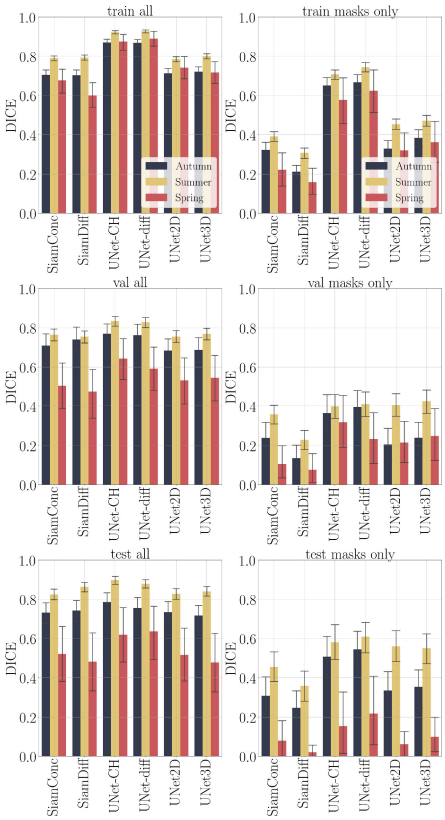

Fig. 14. Dice score seasonal distribution for the training (top), validation (middle), and testing (bottom) samples for all of the change-detection models. Each bar plot represents one season, and descriptions are available at the top plot. Error bars represent the 95% confidence interval.

covering as forest-steppe regions, as well as mountain ones, and improve the models used in our research.

Also, we deployed the open-source online monitoring service.[5] The app processes Sentinel-2 data with the highest frequency once every 5 d; the platform performs segmentation of clearcuts, converts results to the polygons, and stores the results. Users have the opportunity to check the clearcutting regions and track the forest changes during the specified period of time. The image data and deforestation regions are displayed on the OpenStreetMap,[6] currently for 13 image tiles, covered part of Ukraine.

V. Conclusion

In this article, we investigated deep learning approaches to probe the possibility of detecting clearcuts on Ukraine territory from Sentinel-2 imagery. For this purpose, we collected and developed our own dataset that contains annotations of deforestation regions for several Sentinel-2 tiles.

We conducted a set of experiments to get a baseline model according to the dataset without time dependency. To understand how time dependency can improve the process of clearcuts detection, we choose several models and some different techniques to work with images in the time axis. The results of these experiments provided the information that usage pairs of images with close dates can improve the score. We have obtained the highest Dice score and F1-score values for UNet-diff and

[5][Online] Available: https://clearcut.quantumobile.com/
[6][Online] Available: https://www.openstreetmap.org/

UNet-CH models both based on the difference of images in pairs. Also we found out that not all models with time dependency are able to provide improvements. In some cases, the reason is the small amount of data required to train the model.

Our future steps include the extension of the existing dataset by new tiles within Ukrainian territory and developing and testing new architectures of neural networks to improve accuracy of clearcut detection.

The results of our research are available in the `github` repository: https://github.com/QuantuMobileSoftware/forest_change_detection.

ACKNOWLEDGMENT

The authors would like to acknowledge the colleagues from the Quantum company for their help in this project; especially, they would like to thank Andrey Nesmyanovich and Nikita Luzan for their work, which laid the foundation for the current research. They would also like to thank the European Space Agency (ESA) for providing open access to Sentinel-2 data, Esri Conservation Program (ECP), and Group on Earth Observations (GEO) for provision of software and satellite imagery services used for the training dataset and validation dataset preparation. The schemes of neural networks were produced with a software of [28].

REFERENCES

[1] Accessed: Nov. 15, 2020. [Online]. Available: https://www.sentinel-hub.com/develop/capabilities/wms

[2] T. M. Aide *et al.*, "Deforestation and reforestation of latin america and the caribbean (2001–2010)," *Biotropica*, vol. 45, no. 2, pp. 262–271, 2013

[3] A. Amin *et al.*, "Neighborhood effects in the Brazilian Amazônia: Protected areas and deforestation," *J. Environ. Econ. Manag.*, vol. 93, pp. 272–288, 2019.

[4] G. P. Asner, D. E. Knapp, A. Balaji, and G. Paez-Acosta, "Automated mapping of tropical deforestation and forest degradation: CLASlite," *J. Appl. Remote Sens.*, vol. 3, no. 1, pp. 1–24, 2009. [Online]. Available: https://doi.org/10.1117/1.3223675

[5] A. Banskota *et al.*, "Forest monitoring using landsat time series data: A review," *Can. J. Remote Sens.*, vol. 40, no. 5, pp. 362–384, Sep. 2014. [Online]. Available: https://doi.org/10.1080/07038992.2014.987376.

[6] C. Bolyn, A. Michez, P. Gaucher, P. Lejeune, and S. Bonnet, "Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery," *Biotechnol., Agronomy, Soc. Environ. (BASE) J.*, vol. 22, no. 8, pp. 172–187, 2018.

[7] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[8] M. Buchhorn *et al.*, "Copernicus global land service: Land cover 100m: Epoch 2015: Globe, 2019," presented at the ESA Living Planet Symp., Milan, Italy, 2019. [Online]. Available: https://zenodo.org/record/3243509

[9] A. Buslaev *et al.*, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, Feb. 2020, Art. no. 125. [Online]. Available: http://dx.doi.org/10.3390/info11020125

[10] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional siamese networks for change detection," in *Proc. IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.

[11] M. V. Chernyavsky, "Forests of Ukraine and improvements of forest management," in *Restoration Forests*. Dordrecht, Netherlands: Springer Netherlands, 1997, pp. 195–204. [Online]. Available: https://doi.org/10.1007/978-94-011-5548-9_14

[12] J. B. Collins and C. E. Woodcock, "An assessment of several linear change detection techniques for mapping forest mortality using multitemporal landsat TM data," *Remote Sens. Environ.*, vol. 56, no. 1, pp. 66–77, 1996.

[13] P. de Bem, O. de C. Junior, R. F. Guimarães, and R. T. Gomes, "Change detection of deforestation in the Brazilian Amazon using landsat data and convolutional neural networks," *Remote Sens.*, vol. 12, no. 6, Mar. 2020, Art. no. 901. [Online]. Available: http://dx.doi.org/10.3390/rs12060901

[14] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[15] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[16] A. Ghulam, Z.-L. Li, Q. Qin, H. Yimit, and J. Wang, "Estimating crop water stress with ETM+ NIR and SWIR data," *Agric. Forest Meteorol.*, vol. 148, no. 11 pp. 1679–1695, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168192308001603

[17] N. R. Goodwin, N. C. Coops, M. A. Wulder, S. Gillanders, T. A. Schroeder, and T. Nelson, "Estimation of insect infestation dynamics using a temporal sequence of landsat data," *Remote Sens. Environ.*, vol. 112, no. 9, pp. 3680–3689, Sep. 2008.

[18] E. Hamunyela, J. Verbesselt, and M. Herold, "Using spatial context to improve early detection of deforestation from landsat time series," *Remote Sens. Environ.*, vol. 172, pp. 126–138, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425715301942

[19] M. C. Hansen *et al.*, "High-resolution global maps of 21st-century forest cover change," *Science*, vol. 342, no. 6160, pp. 850–853, Nov. 2013. [Online]. Available: https://doi.org/10.1126/science.1244693

[20] M. C. Hansen, D. P. Roy, E. Lindquist, B. Adusei, C. O. Justice, and A. Altstatt, "A method for integrating MODIS and landsat data for systematic monitoring of forest cover and change in the Congo basin," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2495–2513, May 2008. [Online]. Available: https://doi.org/10.1016/j.rse.2007.11.012

[21] D. J. Hayes and S. A. Sader, "Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series," *Photogramm. Eng. Remote Sens.*, vol. 67, no. 9, pp. 1067–1075, 2001.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017, *arXiv:1703.06870*.

[24] M. G. Hethcoat, D. P. Edwards, J. M. B. Carreiras, R. G. Bryant, F. M. França, and S. Quegan, "A machine learning approach to map tropical selective logging," *Remote Sens. Environ.*, vol. 221, pp. 569–582, Feb. 2019. [Online]. Available: https://doi.org/10.1016/j.rse.2018.11.044

[25] G. E Hinton, "Connectionist learning procedures," *Artif. Intell.*, vol. 40, no. 1, pp. 185–234, 1989.

[26] M. Hirschmugl *et al.*, "Methods for mapping forest disturbance and degradation from optical earth observation data: A review," *Current Forestry Rep.*, vol. 3, no. 1, pp. 32–45, Feb. 2017. [Online]. Available: https://doi.org/10.1007/s40725-017-0047-2

[27] I. W. Housman, R. A. Chastain, and M. V. Finco, "An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: Case studies in the United States," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1184. [Online]. Available: https://www.mdpi.com/2072-4292/10/8/1184

[28] H. Iqbal, *Harisiqbal88/plotneuralnet v1.0.0*, 2018. [Online]. Available: https://zenodo.org/record/2526395

[29] S. Ji, Y. Shen, M. Lu, and Y. Zhang, "Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples," *Remote Sens.*, vol. 11, no. 11,Jun. 2019, Art. no. 1343. [Online]. Available: http://dx.doi.org/10.3390/rs11111343

[30] S. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, Sep. 2017.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[32] E. B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sens. Environ.*, vol. 1, no. 3, pp. 155–159, 1970. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425770800219

[33] R. M Korobov and V. Ya Railyan, "Canonical correlation relationships among spectral and phytometric variables for twenty winter wheat fields," *Remote Sens. Environ.*, vol. 43, no. 1, pp. 1–10, Jan. 1993. [Online]. Available: https://doi.org/10.1016/0034-4257(93)90059-7

[34] M. Larabi, Q. Liu, and Y. Wang, "Convolutional neural network features based change detection in satellite images," in *Proc. 1st Int. Worksh. Pattern Recog.*, Jul. 2016, Art. no. 100110W.

[35] T. A. Lima *et al.*, "Comparing Sentinel-2 MSI and Landsat 8 OLI imagery for monitoring selective logging in the Brazilian Amazon," *Remote Sens.*, vol. 11, no. 8, Apr. 2019, Art. no. 961. [Online]. Available: https://doi.org/10.3390/rs11080961
[36] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 936–944.
[37] K. MacDicken *et al.*, "Global forest resources assessment 2015: How are the world's forests changing?" 2016.
[38] D. Masiliūnas, *Evaluating the Potential of Sentinel-2 and Landsat Image Time Series for Detecting Selective Logging in the Amazon*, Wageningen, The Netherlands: Wageningen University and Research Centre, 2017.
[39] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 2016 4th Int. Conf. 3D Vision*, 2016, pp. 565–571.
[40] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, Feb. 2018.
[41] NASA, Landsat. [Online]. Available: https://landsat.gsfc.nasa.gov/
[42] NASA, Modis. [Online]. Available: https://modis.gsfc.nasa.gov/
[43] Ministry of ecology and natural resources of Ukraine, State forest agency Public Report 26022020, Ministry of ecology and natural resources of Ukraine, Accessed: Nov. 15, 2020. [Online]. Available: https://mepr.gov.ua/files/images/news_2020/26022020/
[44] Council of Europe, "Bern convention on the conservation of european wildlife and natural habitats," 2020. [Online]. Available: https://www.coe.int/en/web/bern-convention
[45] M. Ortega, J. Castro, P. N. Happ, A. Gomes, and R. Feitosa, "Evaluation of deep learning techniques for deforestation detection in the Amazon forest," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inform. Sci.*
[46] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzalos, "Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data," Oct. 2019, *arXiv:1910.07778*.
[47] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in *Proc. Eur. Conf. Comput. Vision*, Springer, 2016, pp. 75–91.
[48] QGIS Development Team, *QGIS Geographic Information System*. Beaverton, OR, USA: Open Source Geospatial Foundation, 2009. [Online]. Available: http://qgis.osgeo.org
[49] A. Rasuly, R. Naghdifar, and M. Rasoli, "Detecting of Arasbaran forest changes applying image processing procedures and GIS techniques," *Procedia Environ. Sci.*, vol. 2, pp. 454–464, 2010. [Online]. Available: https://doi.org/10.1016/j.proenv.2010.10.050
[50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.
[51] J. W. Rouse Jr., R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERSTS," in *Proc. 3rd Earth Resour. Technol. Satellite-1 Symp.*, vol. 351, 1974, Art. no. 309.
[52] Sentinel-Hub, "Sentinel2-cloud-detector," [Online]. Available: https://github.com/sentinel-hub/sentinel2-cloud-detector
[53] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Jun. 2015, *arXiv:1506.04214*.
[54] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 958–963.
[55] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html
[56] ESA Sentinel-2 Team, "GMES Sentinel-2 mission requirements document," ESA (or ESA Sentinel-2 Team), Tech. Rep. EOP-SM/1163/MR-dr, Jan. 2007, Accessed: Nov. 15, 2020. [Online]. Available: https://earth.esa.int/pub/ESA_DOC/GMES_Sentinel2_MRD_issue_2.0_update.pdf
[57] L. Wu *et al.*, "Multi-type forest change detection using bfast and monthly landsat time series for monitoring spatiotemporal dynamics of forests in subtropical wetland," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 341. [Online]. Available: https://www.mdpi.com/2072-4292/12/2/341
[58] WWF-Ukraine, "WWF and interpol team up to address forest crime in CEE," 2019. [Online]. Available: https://wwf.panda.org/our_work/governance/?355710/WWF-and-INTERPOL-Ukraine
[59] Q. Xie *et al.*, "Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1482–1493, May 2018.
[60] A. Yismaw, A. B. Gedif, S. Addisu, and F. Zewudu, "Forest cover change detection using remote sensing and GIS in Banja district, Amhara region, Ethiopia," *Int. J. Environ. Monit. Anal.*, vol. 2, Jan. 2014, Art. no. 354.
[61] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
[62] A. Zupanc *et al.*, "Spatio-temporal deep learning: An application to land cover classification," in *Proc.Living Planet Symp.*, May 2019.

Kostiantyn Isaienkov received the M.Sc. degree in computer science from Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine, in 2018.

He is currently with the Data Science Department, Quantum, Kharkiv, Ukraine. His research interests include deep learning and computer vision algorithms in the field of applied robotics. His work consists of developing algorithms for automated machine learning including image processing models.

Mykhailo Yushchuk received the M.Sc. degree in mechanical engineering from Pryazovsiy State Technical University, Mariupol, Ukraine, in 2013.

He is currently leading the Data Science Department, Quantum, Kharkiv, Ukraine. His research interests include remote sensing image processing, image registration, computer vision, and machine learning. His work consists of computer vision applications in remote sensing.

Vladyslav Khramtsov received the B.Sc. degree in astronomy and space informatics from V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, in 2019, where he is currently working toward the M.Sc. degree in astronomy.

He is currently with the Data Science Department, Quantum, Kharkiv, Ukraine. His research interests include machine learning, image processing, and astroinformatics.

Oleg Seliverstov received the B.Sc. and M.Sc. degrees in geography and cartography from V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, in 2001 and 2002, respectively.

He was the founder and leader of Geospatial Laboratory, V. N. Karazin Kharkiv National University from 2002 to 2008. In 2008, he studied at the analytical center of the global cartographic mapping company NAVTEQ, Fargo, ND, USA. He is currently leading a Ukrainian chapter of Society for Conservation Geographical Information Systems and lecturing at the V. N. Karazin Kharkiv National University. His research interests include the application of machine learning to spatial modeling of landscapes and habitats and biodiversity conservancy.