Enhancing forest cover analysis through super-resolution of Sentinel-2 multispectral images

S.V. Illarionova [1], D.G. Shadrin [1], A.V. Kedrov [2]
[1] Skolkovo Institute of Science and Technology, Bolshoy Bulvar 42, bldg. 1, Skolkovo, Moscow, 143026, Russia;
[2] Space technologies and services center, Ltd, str. Lev Lavrov, 14, Perm, 614038, Russia

Abstract

Machine learning (ML) algorithms, combined with satellite observations, offer significant advantages in environmental studies, particularly in vegetation cover analysis. The varying spectral resolution and number of spectral bands of remote sensing imagery allow for different tasks to be addressed with different levels of detail and accuracy. A current limitation in advanced Geographic Information System (GIS) development is the availability and accessibility of data. High-resolution data with a wide spectral range are often expensive, while open-access data typically force researchers to choose between high spatial and temporal resolution or large number of spectral bands. In this study, we investigate this issue through a case study of forest type classification. We employed and trained a single-image super-resolution model based on the Residual Channel Attention Network (RCAN) to upscale Sentinel-2 multispectral images from 10 to 5 meters. We then compared image segmentation results from the original Sentinel-2 data, the upscaled data, and WorldView-3 images. In addition to experiments with spatial resolution, we explored the effect of number of spectral bands on segmentation quality. The results confirm our hypothesis that artificially upscaled data provide more information than low-resolution data, both for narrow and wider spectral ranges, with the increase in spatial resolution proving more significant than the increase in number of spectral bands.

Keywords: remote sensing, computer vision, super-resolution, deep learning.
Citation: Illarionova SV, Shadrin DG, Kedrov AV. Enhancing forest cover analysis through super-resolution of Sentinel-2 multispectral images. Computer Optics 2025; 49(6): 994-1001. DOI: 10.18287/2412-6179-CO-1626.

Introduction

Satellite data serve as a valuable resource for environmental analysis over vast territories, particularly for monitoring changes over extended periods. This data are widely used in forestry analysis to estimate key characteristics such as species composition, height, growing stock volume, and others [1, 2]. Although there is dense coverage of territories by remote sensing data, the quality of this data–both in terms of spatial resolution and number of spectral bands–remains a major issue for the development of efficient geo-information systems (GIS). While multispectral observations provide significant information about vegetation properties, high spatial resolution data with a wide spectral range are often expensive and may be unavailable for certain studies. Consequently, many studies focus on freely available medium or low-resolution multispectral data, such as Sentinel-2, Landsat-8, and MODIS, often sacrificing spatial resolution [3]. On the other hand, even RGB bands can be effectively employed for vegetation cover analysis through deep learning algorithms due to their rich texture information for pattern recognition [4]. The gap between high and medium spatial resolution, along with the spectral range differences between commonly used RGB and more informative multispectral data, has become a hot topic for research exploration. Artificial intelligence (AI) algorithms are expected to bridge this gap by providing additional spectral information [5] and artificially adjusting the spatial resolution of remotely sensed data [6]. Therefore, super-resolution (SR) methods are among the promising tools for enhancing the availability and reliability of remote sensing analyses based on less expensive satellite data.

In the field of computer vision, SR methods aim to improve image resolution using various techniques. These methods typically rely on high-resolution (HR) which are used In supervised artificial intelligence algorithms, high-resolution (HR) images serve as a reference to validate the generated SR images.

The Super-Resolution Convolutional Neural Network (SRCNN) stands as one of the pioneering CNN-based networks designed for super-resolution in the broader domain of computer vision [7]. Over time, several enhancements to SRCNN have been proposed [8, 9, 10, 11].

Recent advancements in the SR task are closely tied to Generative Adversarial Networks (GANs) [12]. In the case of the Super-Resolution Generative Adversarial Network (SRGAN), a dual-model approach is employed, with one model responsible for generating images and the other for distinguishing between original and generated images [13]. The generative model is optimized using a combination of content and adversarial loss functions [14]. ESRGAN, an architectural improvement upon SRGAN for realistic image SR, was proposed in [15]. The authors introduced a residual-in-residual dense block, along with adversarial and perceptual loss functions. Subsequent enhancements involve training ESRGAN with purely synthetic data utilizing high-order degradation modeling, which aligns closely with real-world degradation scenarios [16].

Another popular model in this domain is the Residual Channel Attention Network (RCAN) [17]. RCAN adopts a deep residual network architecture, incorporating channel attention mechanisms to selectively highlight crucial features in the image. The attention mechanism of the RCAN model learns to assign weights to features across various channels, enabling the generation of high-quality, sharp, and realistic images.

In the domain of remote sensing of environment, the challenge of enhancing spatial resolution of satellite data is particularly critical in scenarios where acquiring higher-resolution images is impractical or cost-prohibitive. This issue is notably relevant in remote sensing applications, such as the utilization of Sentinel-2 satellite imagery, which offers free access with a high revisit frequency but is constrained by a 10-meter ground sampling distance. SR algorithms can play a pivotal role in increasing the resolution of such images particularly in cases where multispectral data are obtained instead of RGB images commonly used in the general computer vision domain. Addressing this challenge, Michel et al. introduced the SEN2VENμS [18] dataset, that comprises cloud-free surface reflectance patches from multispectral Sentinel-2 data at both 10-meter and 20-meter resolutions. Importantly, it pairs these patches with spatially registered surface reflectance patches at a high 5 meter resolution, concurrently acquired by the VENμS satellite on the same day. This dataset spans 29 distinct locations on Earth, encompassing a total of 132,955 patches, each measuring 256×256 pixels at 5-meter resolution. SEN2VENμS serves as a valuable resource for training and benchmarking SR algorithms, enabling the enhancement of the spatial resolution of eight of the Sentinel-2 spectral bands to 5 meters, ultimately advancing the capabilities of remote sensing applications.

In this study, we set a hypothesis that artificially enhancing the spatial resolution of medium-resolution multispectral satellite imagery will improve the extraction of forest characteristics using a deep neural network. To validate this hypothesis, we developed a super-resolution (SR) model for Sentinel-2 data, evaluating two training scenarios for 10m and 20m spectral bands. In the second stage of our study, we collected a dataset based on forestry assessments conducted in Central Russia. We considered two forest species: deciduous and coniferous. A convolutional neural network was trained to recognize forest types under different scenarios using 4 and 8 spectral bands, with the original resolution of 10m and an upscaled resolution of 5m. To achieve more comprehensive results regarding the effects of spectral and spatial resolution on forest segmentation tasks, we also utilized WorldView-3 imagery, which has a spatial resolution of less than 2m, to analyze model performance when high-resolution data are available.

1. Methods and data

1.1. Problem definition

The goal of the conducted research is to explore the feasibility of super-resolution technique implementation for forest cover classification. We focus on freely available Sentinel-2 multispectral data as a valuable data source for environmental analysis with medium spatial resolution of 10m per pixel. The experiments involve two main parts:

- super-resolution model development for multispectral data;
- forest classification based on images of different spatial resolution.

For Sentinel-2 image upscaling, we selected 8 spectral bands with spatial resolution of 10 m and 20 m, that are commonly used in ecological and forestry tasks. There are the following Central wavelengths (nm) for each selected Sentinel-2 band: Blue – 490, Green – 560, Red – 665, Red edge (RE1) – 705, Red edge (RE2) – 740, Red edge (RE3) – 783, Near infrared (NIR) – 842, Narrow near infrared (NNIR) – 865. In present experiments, we set the target resolution to 5 m. Two approaches are considered for model development. The first approach assumes that two separate models are developed for upscaling 4 bands with original resolution of 10 m and separately of 20 m. Another approach relies on the idea that during model development the loss function is shared between two models and the training process is performing simultaneously to bring all 8 bands to the resolution of 5 m. The implementation details are presented in the following subsections.

To compare the upscaled images and the original high-resolution data, we also considered observations captured by the WorldView-3 satellite with the spatial resolution up to 1.24 meter. There are the following Central wavelengths (nm) for each selected WorldView band: Coastal blue – 425, Blue – 480, Green – 545, Yellow – 605, Red – 660, Red-edge – 725, Near infrared 1 – 833, Near infrared 2 – 950. The number of spectral bands always plays a significant role in forest cover recognition, therefore, we compared two scenarios where 4 or 8 bands are available for semantic segmentation model training. The corresponding spectral bands were selected both for Sentinel-2 and WorldView-3 observations. The first set of bands involve Red, Green, Blue, and Wide Near-Infrared spectral bands, while for the second scenario, we considered the wavelengths corresponding to Red, Blue, Green, Wide Near-Infrared, three Red Edge bands, and Narrow Near-Infrared. As a result, we conducted the following set of experiments to explore different spatial resolution and number of spectral bands both for Sentinel-2 and WorldView-3 data:

1) Sentinel-2 (S2), 10 m, 8 bands (B2 - B8a)
2) Sentinel-2 (S2), 10 m, 4 bands (RGB+NIR)
3) Sentinel-2 (S2), upscaled to 5 m, 8 bands (B2 - B8a), super-resolution algorithm with simultaneous model training for factor 2 and factor 4
4) Sentinel-2 (S2), upscaled to 5 m, 4 bands (RGB+NIR), super-resolution algorithm with simultaneous model training for factor 2 and factor 4
5) Sentinel-2 (S2), upscaled to 5 m, 8 bands (B2 - B8a), super-resolution algorithm with separate model training for factor 2 and factor 4 (RGB+NIR)

6) Sentinel-2 (S2), upscaled to 5 m, 4 bands (RGB+NIR), super-resolution algorithm with separate model training for factor 2 and factor 4
7) WorldView-3, 5 m, 8 bands (all bands)
8) WorldView-3, 5 m, 4 bands (RGB+NIR)
9) WorldView-3, 2 m, 8 bands (all bands)
10) WorldView-3, 2 m, 4 bands (RGB+NIR)

The same numbering of the Experiments is preserved in the Results and Discussion section.

1.2. Satellite data and inventory field-based measurements

Due to the significant volume and numerous spectral channels in both input and reference images, which enable the effective utilization of spectral information from high-resolution imagery for forest-related tasks, the SEN2VENμS dataset is employed for super-resolution purposes. This dataset allows us to develop a model to upscale Sentinel-2 data using VENμS data as a high-resolution reference. VENμS is a New Micro-Satellite launched by French and Israeli space agencies. There are the following Central wavelengths (nm) for each selected Venus band: B3 – 490, B4 – 555, B7 – 672, B8 – 702, B9 – 742, B10 – 782, B11 – 865. B1, B2, B5, B6, B12 were excluded from the further study. The dataset was divided into training, validation, and test sets in proportions of 75 %, 20 %, and 5 %, respectively. As a preprocessing step, the values of all pixels were divided by 10^5, and then truncated to fall within the range [0, 0.5]. The total coverage of the SEN2VENμS dataset represents 29 sites across the world.

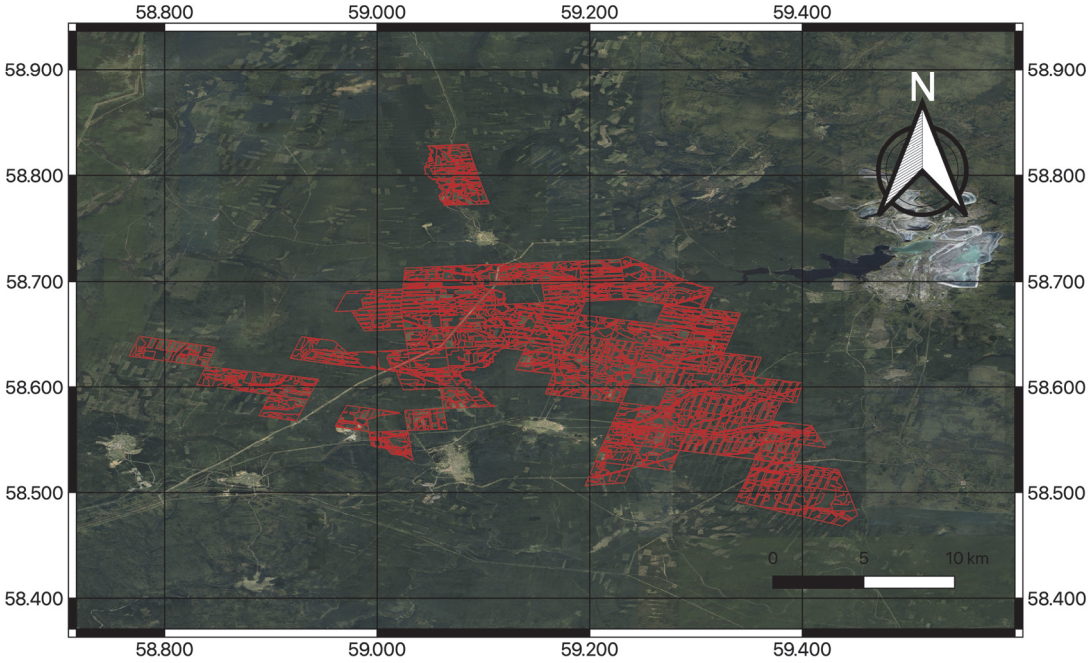

Fig. 1. Study area in the Gornozavodsky District in Perm Krai. The red polygons are for forestry individual stands

For the segmentation part, we prepared forest inventory data along with satellite images from Sentinel-2 and WorldView-3 for 2017. The study area is located in Gornozavodsky District in Perm Krai, Russia (Fig. 1). The forest inventory was conducted in 2017 and covered an area of more than 346 km². The data were collected in the format of individual stands with the averaged properties for each stand. Two classes, deciduous and coniferous, were selected for further analysis and converted to raster files, aligning the satellite data with the appropriate spatial resolution for each experiment. To reduce the effect of cloud coverage, cloudless composite imagery from Sentinel-2 and WorldView-3 was prepared for the summer period of 2017. Sentinel-2 data were preprocessed in the same way as they were prepared for the super-resolution part, while WorldView-3 images were simply normalized. The dataset is divided into training, validation, and test sets based on forestry quarters. The quarters do not overlap and are selected randomly for each subset in the defined proportions. The target label distribution is presented in Table 1. To exclude any areas with logging from consideration, we employed the developed model for forest mask estimation, as presented in [19]. Therefore «background» pixels were not used for segmentation model development and evaluation.

1.3. Experimental setup

1.3.1. Super-resolution

We consider the attention-based model RCAN for the SR task that has already shown remarkable results for image resolution adjustment both in general and remote sensing domains.

Tab. 1. Distribution of the target classes in km²

	Deciduous	**Coniferous**
Train area	67	81
Validation area	10	12
Test area	18	16

SEN2VENμS dataset uses 8 bands of Sentinel-2 images, some of which have resolution 10 m (B2, B3, B4,

B8), while others have resolution 20 m (B5, B6, B7, B8a). Thus, we train three different models with RCAN architecture - one for each input resolution (RCAN 2× and RCAN 4×) and one for all of the bands (RCAN FUSE). Therefore, three models were trained:

1) RCAN ×2 for B2, B3, B4, B8 bands;
2) RCAN ×4 for B5, B6, B7, B8a bands;
3) RCAN FUSE for all bands simultaneously.

Additionally, we calculated the average performance for upscaled super-resolution images created by RCAN ×2 and RCAN ×4. We denoted this experiment RCAN ×2 + RCAN ×4. This assessment was conducted to compare independent models training for super-resolution and model training in a single pipeline (RCAN FUSE).

RCAN 2× and RCAN 4× are purely identical except for the up-sample factor. Both networks are trained for 50 epochs using a mean square error (MSE) loss function and Adam as optimizer.

For the last model, we concatenate two RCAN models - the first one takes input images with resolution 20 m and up-sample them to 10 m. After that up-sampled images concatenates with other bands with original 10 m resolution and concatenated tensor uses as input the second RCAN model, which up-samples all of the bands to 5 m. Here, we also train a network for 50 epochs using MSE, but now we calculate loss function twice - one time for each sub-network and its own input and output, and then sum up values to be able to train the whole network end-to-end.

1.3.2. Forest cover segmentation

To investigate the importance of spatial and spectral data quality in the forest cover classification task, we conduct experiments with the MA-Net [20] neural network architecture, which have proven its capability in remote sensing tasks and, more generally, in segmentation tasks [21].

MA-Net, being an improvement of UNet, has exhibited remarkable performance in previous remote sensing tasks by explicitly modeling multi-scale features. This capability allows it to effectively account for the varying scales present in remote sensing images, capturing everything from large geographical features to smaller, finer structures. The MA-Net's focus on multi-scale feature integration aligns well with the heterogeneous nature of remote sensing data.

We train several models with MA-Net architecture for all combinations of input bands, resolutions and data sources, including super-resolution Sentinel-2 images gained with RCAN 2×, RCAN 4× and RCAN FUSE models. Each model is trained using Dice-Focal Loss, which is essentially a weighted combination of dice [22] and focal [23] losses, for 150 epochs. Geometrical augmentations such as rotation and flipping were employed to enlarge the dataset size while preserving the spectral properties.

1.4. Evaluation metrics

1.4.1. Evaluation metrics for super-resolution

To evaluate the developed SR models, we compute the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), both commonly used metrics for image adjustment tasks. It measures both the quality and diversity of the generated images by comparing their feature representations to those of real images. We average the metrics for all patches to achieve the ultimate value. The metrics are computed according to the following equations:

$$PSNR\left(I_1, I_2\right) = 20log_{10}\left(\frac{MAX\left(I_1\right)}{MSE\left(I_1, I_2\right)}\right),$$

where MAX represents the maximum possible pixel value of the image (i.e., 255 for an 8-bit grayscale image), and MSE represents the mean squared error between the original image and the compressed or distorted image.

$$SSIM\left(I_1, I_2\right) = \frac{\left(2\mu_1\mu_2 + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_1^2 + \mu_2^2 + c_1\right)\left(\sigma_1^2 + \sigma_2^2 + c_2\right)},$$

where μ_1 and μ_2 are the pixel sample means, σ_x and σ_y are variances, σ_{xy} is covariance, c_1 and c_2 are variables to stabilize the division with weak denominators.

1.4.2. Evaluation metrics for semantic segmentation

To assess the quality of the forest cover segmentation task, we considered commonly used metrics such as Precision, Recall, F1-score, and Intersection over Union (IoU). They are computed using the following formulas:

$$Recall = \frac{TruePositive}{TruePositive + TrueNegative},$$

$$Precision = \frac{TruePositive}{TruePositive + FalseNegative},$$

$$F1 - score = \frac{2Precision * Recall}{Precision + Recall},$$

$$IoU = \frac{AreaOfOverlap}{AreaOfUnion},$$

where *True Positive*, *False Negative*, and *True Negative* represent number of correct classified pixels of positive class, number of pixels that were wrongly classified as a target class, number of pixels that were correctly ascribed to the negative class, respectively; *Area Of Overlap* and *Area Of Union* represent overlapping between predicted and ground truth masks. The results are presented as averaged for two classes.

2. Results and discussion

In this study, we investigated the importance of spatial resolution and number of spectral bands in the task of forest cover classification. To improve artificially the quality of satellite images, we employed a super-resolution technique based on RCAN architecture. The study involved multispectral Sentinel-2 with simulation enhancement of several spectral bands. Then, we also compared pattern recognition capability based on various input images using semantic segmentation MA-Net architecture.

2.1. Super-resolution

The numerical results for image super-resolution are presented in Tab. 2. The results for ×2 upsampling are higher than for the ×4 factor both in terms of SSIM and PSNR metrics as the less upscaling factor the easier the problem for the AI algorithm. It is important to note that the model trained for factor ×2 works just with the Sentinel-2 bands with original resolution of 10 m (B2, B3, B4, B8), while another model with factor ×4 upscales the spectral bands with the 20 m spatial resolution (B5, B6, B7, B8a). The fused model proceeds all 8 bands simultaneously by sharing the loss functions for upscale factors of 2 and 4. Fig. 2 – 4 represent the achieved results for multispectral image super-resolution. It can be observed that the edges and the general structure of the surface become more detailed and accurate for all considered bands. Moreover, we compared the approach where RCAN ×2 and RCAN ×4 are used for individual bands to obtain multispectral images with 5 m per pixel resolution. This approach denoted as RCAN ×2 + RCAN ×4 demonstrates the PSNR and SSIM equal to 33.9 and 0.915 respectively, while the RCAN FUSE approach leads to PSNR and SSIM of 32.5 and 0.91. There is no statistically significant difference between the super-resolution of bands with the same upscaling factor. However, for the factor of ×2, the differences in PSNR and SSIM are smaller compared to the model with a factor of 4×. The variance of PSNR values for upscaling with a factor of x2 is 0.06, while for a factor of ×4, it is 0.09. When we trained the RCAN FUSE model, the same behavior was observed for bands with an original resolution of 10 m and for bands with a resolution of 20 m.

Tab. 2. Results of Super-resolution models on test subset. RCAN ×2 + RCAN ×4 represents super-resolution setup where two models are trained independently

Model or setup	PSNR	SSIM
RCAN ×2	35.9	0.95
RCAN ×4	31.9	0.88
RCAN ×2 + RCAN ×4	33.9	0.915
RCAN FUSE	32.5	0.91

2.2. Forest cover segmentation

Following image super-resolution, we conducted experiments with different spatial resolution and number of spectral bands to assess the model's capability of meaningful feature extraction in different scenarios. The numerical results are presented in Tab. 3. The average F1-score for the original Sentinel-2 images with spatial resolution of 10m per pixel equals to 0.64 for all bands, and 0.62 for RGB and NIR bands. The visual assessment was conducted for some test samples presented in Fig. 5 – 8. Then, we compared several datasets with upscaled Sentinel-2 observations with the spatial resolution of 5 m per pixel. The same as for the original resolution, we compared the usage of 8 and 4 spectral bands and their significance for the target classes recognition. The RCAN FUSE approach slightly outperforms the approach with separate RCAN 2× and RCAN 4× models for 8 bands for F1-score metrics and for 4 bands for IoU metrics. It is important to note, that the model trained on the WorldView-3 data downscaled to 5 meters shows the same F1-score as the one trained on upscaled Sentinel-2 data. Both of them significantly outperforms the original Sentinel-2 data with the 10-meters spatial resolution that support the initial hypothesis.

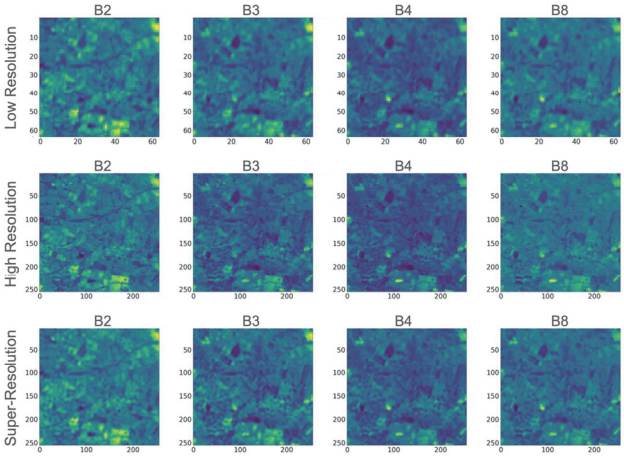

Fig. 2. Prediction of RCAN ×2 model for multispectral image super-resolution for spectral bands B2, B3, B4, B8. The colormap format for individual band representation is «viridis»

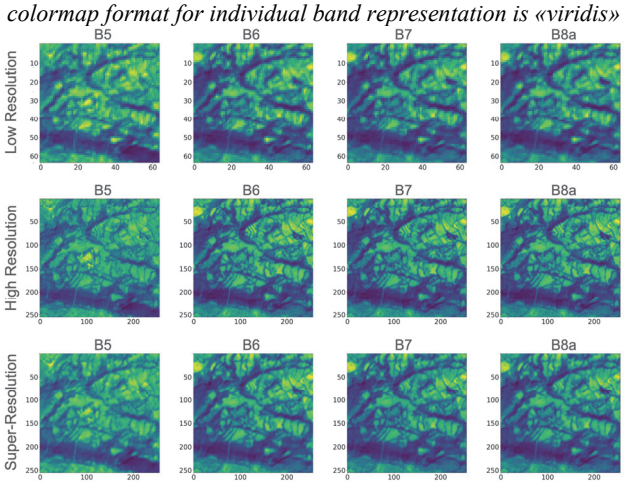

Fig. 3. Prediction of RCAN ×4 model for multispectral image super-resolution for spectral bands B5, B6, B7, B8a. The colormap format for individual band representation is «viridis»

We also considered confusion matrix to assess the segmentation performance (Fig. 9 – 10). The matrices are computed for the experiments with the highest performance based on average score for coniferous and deciduous classes. The Experiment 3 shows the highest performance among experiments with Sentinel-2 data. This experiment involves data brought to 5 m using the RCAN FUSE model. For the WorldView-3 the best performance was achieved in the Experiment 9, where all bands were used and the spatial resolution equal to 2 m. The same percentage of the correctly segmented regions for these two models are observed (67 %). On the other hand, the model in the Experiment 9 significantly outperformed the model in the Experiment 3 in coniferous region segmentation (69 % and 81 %, respectively).

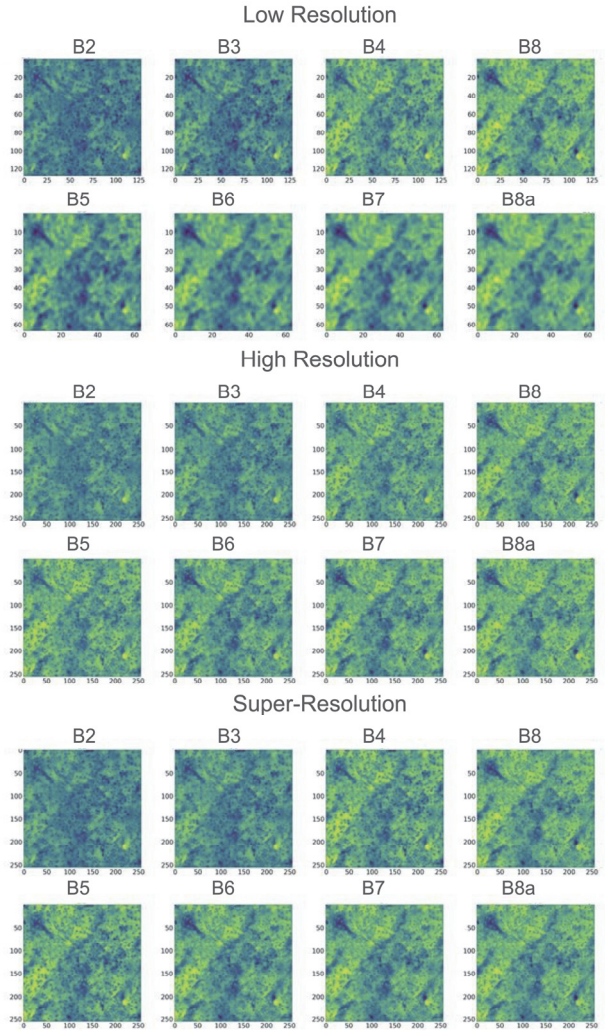

Fig. 4. Prediction of RCAN FUSE model for multispectral image super-resolution for all 8 bands. The colormap format for individual band representation is «viridis»

Tab. 3. Segmentation metrics on test subset for different spatial resolution and number of spectral bands (is for super-resolution using RCAN FUSE, ** is for super-resolution using RCAN 2× + RCAN 4×). # denotes the experiment's number*

#	Satellite	Resolution	Bands	F1-score	IoU
1	S2	10 m	B2 - B8a	0.64	0.46
2	S2	10 m	RGB + NIR	0.62	0.45
3	S2	5 m*	B2 - B8a	0.68	0.51
4	S2	5 m*	RGB + NIR	0.67	0.5
5	S2	5 m**	B2 - B8a	0.67	0.51
6	S2	5 m**	RGB + NIR	0.67	0.49
7	WV3	5 m	all bands	0.68	0.52
8	WV3	5 m	RGB + NIR	0.68	0.52
9	WV3	2 m	all bands	0.74	0.58
10	WV3	2 m	RGB + NIR	0.72	0.54

There are the following avenues for future study. In recent research, we focus mainly on hypothesis validation using a single super-resolution algorithm, RCAN. This algorithm can be considered as a baseline for further exploration as it has proven its capability in various computer vision tasks in different modalities with comprehensively high performance. In the future study, more advanced algorithms can be employed such as diffusion models. The same for the semantic segmentation task where comparison between different DL algorithms can be conducted.

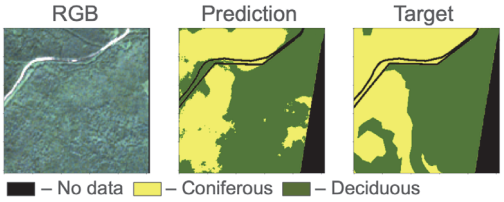

Fig. 5. Prediction for Sentinel-2 image with 10 m resolution and B2-B8a bands

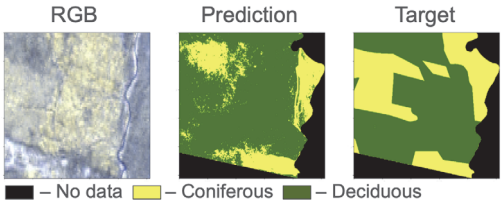

Fig. 6. Prediction for Sentinel-2 image with 5 m resolution (upscale with RCAN FUSE) and B2-B8a bands

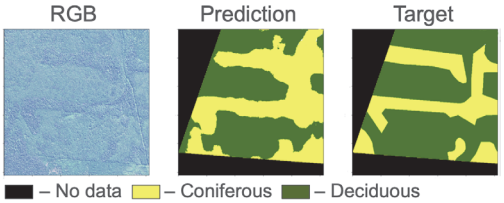

Fig. 7. Prediction for WorldView-3 image with 5m resolution and all bands

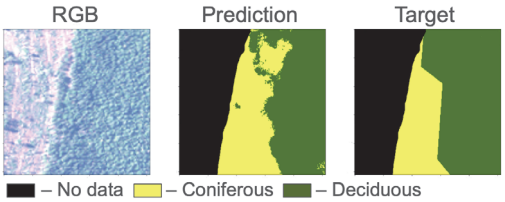

Fig. 8. Prediction for WorldView-3 image with 2 m resolution and all bands

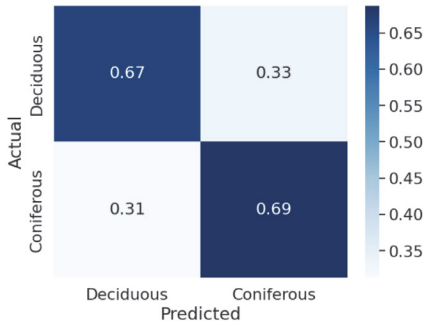

Fig. 9. Confusion matrix for Experiment #3: Sentinel-2 with all bands upscaled to 5 m. Values are normalized and represent the proportion of correct and misclassified areas

It is important to notice, that for the conducted experiments, a small forestry dataset was used and enlarging the dataset size may lead to higher segmentation quality. However, such dataset size represents a typical situation for environmental studies where field-based measurements for

reference data might be extremely limited. It is usually crucial for more specific vegetation characteristics for which multispectral data are required. Also, in this study, we used forestry inventory data instead of manual annotation of high-resolution satellite or aerial photogrammetry. The reason is that it is a common data source for environmental studies that corresponds to the national forest management instructions [24].

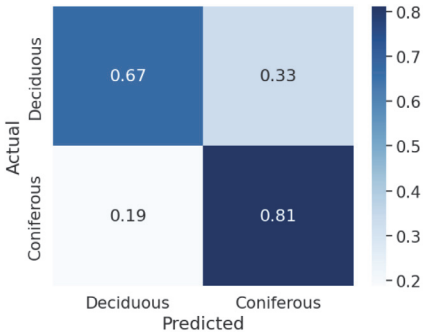

Fig. 10. Confusion matrix for Experiment #9: WorldView-3 image with 2 m resolution and all bands. Values are normalized and represent the proportion of correct and misclassified areas

3. Conclusion

Multispectral satellite data serve as a valuable source of information regarding land cover characteristics. However, researchers often face the challenge of choosing between high spatial and temporal resolution and number of spectral bands. While high-resolution satellite data with frequent revisit times and wide spectral ranges can be prohibitively expensive, integrating AI algorithms into the data processing pipeline can help mitigate the need for costly imagery while minimizing the loss of useful information. In this study, we explored the potential of adjusting the spatial resolution of multispectral images for forest cover analysis. We collected a unique dataset representing coniferous and deciduous forest species in Central Russia. The super-resolution algorithm RCAN was trained to upscale the resolution of Sentinel-2 multispectral images from 10 to 5 meters. We then compared the results of forest cover segmentation using multispectral WorldView-3 images at an high resolution of 2 meters per pixel, downscaled to 5 meters, alongside Sentinel-2 data at its original 10-meter resolution and artificially upscaled to 5 meters. The study involved experiments with eight spectral bands, including a narrow range utilizing only RGB and NIR bands. Our findings demonstrate that artificially upscaled data are more effective for remote sensing image segmentation than original lower-resolution observations. However, we did not observe significant improvements when all eight bands were upscaled compared to using only four upscaled bands. Overall, the proposed approach shows promise in addressing the challenges posed by the high costs of acquiring high-resolution multispectral satellite images.

Acknowledgements

The work was funded by the Russian Science Foundation under project No. 23-71-01122.

References

[1] Lei T, et al. Review of remote sensing-based methods for forest aboveground biomass estimation: Progress, challenges, and prospects. Forests 2023; 14(6): 1086. DOI: 10.3390/f14061086.

[2] Illarionova S, Smolina A, Shadrin D. Primary forest characteristics estimation through remote sensing data and machine learning: Sakhalin case study. E3S Web of Conf 2024; 542: 04003. DOI: 10.1051/e3sconf/202454204003.

[3] Illarionova S, et al. A survey of computer vision techniques for forest characterization and carbon monitoring tasks. Remote Sens 2022; 14(22): 5861. DOI: 10.3390/rs14225861.

[4] Illarionova S, et al. Estimation of the canopy height model from multispectral satellite imagery with convolutional neural networks. IEEE Access 2022; 10: 34116-34132. DOI: 10.1109/ACCESS.2022.3161568.

[5] Liu L, et al. Hyperspectral remote sensing imagery generation from RGB images based on joint discrimination. IEEE J Sel Top Appl Earth Obs Remote Sens 2021; 14: 7624-7636. DOI: 10.1109/JSTARS.2021.3099242.

[6] Wang P, Bayram B, Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth Sci Rev 2022; 232: 104110. DOI: 10.1016/j.earscirev.2022.104110.

[7] Dong C, et al. Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Machine Intell 2015; 38(2): 295-307. DOI: 10.1109/TPAMI.2015.2439281.

[8] Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 1646-1654. DOI: 10.1109/CVPR.2016.182.

[9] Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 1637-1645. DOI: 10.1109/CVPR.2016.181.

[10] Shi W, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 1874-1883. DOI: 10.1109/CVPR.2016.207.

[11] Liu J, et al. Residual feature aggregation network for image super-resolution. 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2020: 2356-2365. DOI: 10.1109/CVPR42600.2020.00243.

[12] Chen H, et al. Real-world single image super-resolution: A brief review. Inf Fusion 2022; 79: 124-145. DOI: 10.1016/j.inffus.2021.09.005.

[13] Ledig C, et al. Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2017: 105-114. DOI: 10.1109/CVPR.2017.19.

[14] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In Book: Leibe B, Matas J, Sebe N, Welling M, eds. Computer Vision – ECCV 2016. 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II. Cham, Switzerland: Springer International Publishing AG; 2016: 694-711. DOI: 10.1007/978-3-319-46475-6_43.

[15] Wang X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks. In Book: Leal-Taixé L, Roth S, eds. Computer Vision – ECCV 2018 Workshops. Munich, Germany, September 8-14, 2018, Proceedings, Part V. Cham, Switzerland: Springer International Publishing AG; 2018: 63-79. DOI: 10.1007/978-3-030-11021-5_5.

[16] Wang X, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. 2021 IEEE/CVF Int Conf on Computer Vision Workshops (ICCVW) 2021: 1905-1914. DOI: 10.1109/ICCVW54120.2021.00217.
[17] Zhang Y, et al. Image super-resolution using very deep residual channel attention networks. In Book: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. Computer Vision – ECCV 2018. 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII. Cham, Switzerland: Springer International Publishing AG; 2018: 294-310. DOI: 10.1007/978-3-030-01234-2_18.
[18] Michel J, et al. SEN2VENμS, a dataset for the training of Sentinel-2 super-resolution algorithms. Data 2022; 7(7): 96. DOI: 10.3390/data7070096.
[19] Illarionova S, et al. Augmentation-based methodology for enhancement of trees map detalization on a large scale. Remote Sens 2022; 14(9): 2281. DOI: 10.3390/rs14092281.
[20] Fan T, et al. MA-Net: A multi-scale attention network for liver and tumor segmentation. IEEE Access 2020; 8: 179656-179665. DOI: 10.1109/ACCESS.2020.3025372.
[21] Sharma S. Semantic segmentation for urban-scene images. arXiv Preprint. 2021. Source: <https://arxiv.org/abs/2110.13813>. DOI: 10.48550/arXiv.2110.13813.
[22] Milletari F, Navab N, Ahmadi S-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth Int Conf on 3D Vision (3DV) 2016: 565-571. DOI: 10.1109/3DV.2016.79.
[23] Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. 2017 IEEE Int Conf on Computer Vision (ICCV) 2017: 2999-3007. DOI: DOI: 10.1109/ICCV.2017.324.
[24] Schepaschenko D, et al. Russian forest sequesters substantially more carbon than previously reported. Sci Rep 2021; 11: 12825. DOI: 10.1038/s41598-021-92152-9.

Authors' information

Svetlana Vladimirovna Illarionova. Svetlana received the bachelor's and master's degrees in computer science from Lomonosov Moscow State University, Moscow, Russia, in 2017 and 2019, respectively. She received the PhD degree in data science in 2023. Currently, she is a head of Research Group of remote sensing data processing at Skolkovo Institute of Science and Technology (Skoltech), Russia. Her scientific interests include computer vision, remote sensing of environment, and machine learning. E-mail: *s.illarionova@skoltech.ru*

Dmitrii Germanovich Shadrin is PhD in data science. He is a head of department at Skolkovo Institute of Science and Technology (Skoltech), Russia. His research interests are computer vision and precision agriculture. E-mail: *d.shadrin@skoltech.ru*

Alexander Vladimirovich Kedrov, born in 1986. Graduated from the Perm State Agricultural Academy named after D.N. Pryanishnikov (Perm) in 2009, specializing in engineering of garden and park construction. Since 2009 and up to the present time, he has been working at the Perm Agrarian-Technological University named after D.N. Pryanishnikov (Perm) in the Department of Forestry and Landscape Architecture as a senior lecturer. Since 2015, he has also been working at the LLC "Center for Space Technologies and Services" as deputy director, responsible for projects related to forest inventory technologies based on laser scanning data. Alexander's scientific interests include laser scanning, forest inventory, remote methods, and GIS technologies. E-mail: *kedalex@gmail.com*

Code of State Categories Scientific and Technical Information (in Russian – GRNTI)): 28.23.15
Received October 22, 2024. The final version – March 13, 2025.