



OPEN

Spatial point patterns generation on remote sensing data using convolutional neural networks with further statistical analysis

Rostyslav Kosarevych^{1✉}, Oleksiy Lutsyk¹, Bohdan Rusyn¹, Olga Alokhina¹, Taras Maksymyuk^{2✉} & Juraj Gazda³

Continuous technological growth and the corresponding environmental implications are triggering the enhancement of advanced environmental monitoring solutions, such as remote sensing. In this paper, we propose a new method for the spatial point patterns generation by classifying remote sensing images using convolutional neural network. To increase the accuracy, the training samples are extended by the suggested data augmentation scheme based on the similarities of images within the same part of the landscape for a limited observation time. The image patches are classified in accordance with the labels of previously classified images of the manually prepared training and test samples. This approach has improved the accuracy of image classification by 7% compared to current best practices of data augmentation. A set of image patch centers of a particular class is considered as a random point configuration, while the class labels are used as marks for every point. A marked point pattern is regarded as a combination of several subpoint patterns with the same qualitative marks. We analyze the bivariate point pattern to identify the relationships between points of different types using the features of a marked random point pattern.

Nowadays, we observe a tremendous increase of the number of remote sensing devices and platforms such as low-orbit satellites, unmanned aerial vehicles, etc. With all these platforms we observe an exponential growth of the corresponding remote sensing data¹. As result, so far we are observing a huge interest from the machine learning community to the analysis of remote sensing data, including object classification, pattern recognition, etc.¹⁻³. Remote sensing (RS) image analysis allows us to solve a wide range of applied environmental tasks, such as monitoring of natural resources, landscapes, ecology, etc. It also provides opportunities for the analysis of the Earth's surface in local and global scale, and make corresponding conclusions on ecosystems and their underlying biotic and abiotic controls⁴⁻⁶.

In addition to conventional 2D image cameras, there are also a number of alternative sensing solutions such as lidars or terrestrial laser scanning, which produce a scatter plot of points to reflect the footprint of ecological processes. These footprints can be detected through spatial point-pattern analysis (SPPA).

For the first time, the possibility of describing the ecological processes by the point patterns was indicated in^{7,8}. Since that, research over the last two decades has undoubtedly proven that spatial point patterns and processes play an important role in understanding of plant and animal communities⁹⁻¹³. Being widely used in plant ecology, SPPA is frequently adopted to describe biotic interactions and interpret pattern process relationships¹⁴. Inhomogeneous summary statistics can quantify the impact of heterogeneity, while mark correlation functions can include trait and phylogenetic information in the analysis of multivariate spatial patterns. Furthermore, more refined point process models can be used to realistically characterize the structure of a wide range of patterns. The main objective of SPPA in ecology is to extract the information from corresponding patterns¹⁵. Identifying the type of interaction between the objects of environmental research will provide deeper understanding of the implicit processes that link them to build predictive dependencies of their behavior. To obtain the point pattern of investigated objects in some region a time-consuming investigation are necessary. Sometimes it is possible only with remotely received data. Therefore, it is important to develop new approaches for the analysis of remote

¹Department of Remote Sensing Information Technologies, Karpenko Physico-Mechanical Institute, NAS of Ukraine, Lviv, Ukraine. ²Department of Telecommunications, Lviv Polytechnic National University, Lviv, Ukraine. ³Technical University of Kosice, Kosice, Slovakia. ✉email: kosar2311@gmail.com; taras.maksymyuk@gmail.com

sensing data, which allow to us to improve a process of identification of the existing dependencies among the elements of ecosystems.

The main contributions of this article are the following:

1. We propose a new method for the spatial point patterns generation by classifying remote sensing images using convolutional neural network.
2. We propose a new data augmentation scheme that uses a similarity between the images within the same part of the landscape for a limited observation time to improve the classification performance of convolutional neural network.
3. We analyze the bivariate point pattern to identify the relationships between different types of points using the features of a marked random point pattern.

The remainder of the paper is organized as follows. In “[Background and related work](#)” we present a background and related work on the environmental remote sensing and data analysis. In “[Spatial point patterns generation by remote sensing data classification](#)” we describe the details on our research workflow and main contributions of the paper. In “[Simulation results and performance analysis](#)” we present the experimental simulation results. In “[Discussion and future research](#)” we discuss the main outcomes and potential use cases for the proposed SPPA approach. Finally, we conclude the paper in “[Conclusions](#)”.

Background and related work

Recent trends in remote sensing. Environmental studies can be differentiated by two groups. The first one is the analysis of previously generated point patterns of the studied objects. The second is the qualitative selection of the desired objects in the image. The methods of the first group aim for a more accurate description of the environmental processes occurring in the target investigated area, while the methods of the second group aim for more accurate detection of objects. Both groups require a labor-intensive data collection to obtain corresponding spatial point patterns and accurately determine the target features of each object. Thus, comprehensive approaches for environmental monitoring are getting momentum so far^{16–19}. In¹⁶, authors describe main reasons why RS has become an important source of data, and present the different types of sensors and platforms that have been used to display the diversity of forest variables.

The rapid development of RS technologies supports forest management and conservation in those parts of the world where environmental issues are most critical. For example, in¹⁷ a methodology for continuous monitoring and estimation of land cover and land change areas have been tested. This methodology is based on the recent advancements in the field of environmental RS, using algorithms for time series analysis and estimation protocols. The research¹⁸ is devoted to modelling and predicting the species richness. By diversifying multiple sensors, which examine relationships over different temporal and spatial scales, authors have focused on the normalized difference vegetation index (NDVI) from passive sensors. NDVI has been associated with net primary productivity and has been hypothesized to quantify species richness and diversity based on the species energy theory^{19–26}. In particular, in¹⁹ authors review techniques that are currently used for monitoring, including close-range RS, airborne and satellite-based approaches. The implementation of optical, RADAR and LiDAR RS-techniques to assess spectral traits/spectral trait variations (ST/STV) is described in detail. They found that ST/STV can be used to record indicators of ecosystem state based on RS. Therefore, the ST/STV approach provides a framework to develop a standardized monitoring concept for ecosystem condition indicators using RS techniques that is applicable to future monitoring programs. In addition, in¹⁹ authors analyze two different approaches: empirical modelling and physical modelling for the estimation of ST and trait variations of ecosystem condition from RS data. Authors have investigated that empirical models based on statistical relationship between one or more biotic traits and one or more spectral variables that are widely used for the quantitative modelling of biophysical vegetation parameters on a global scale.

Analyzing the above-mentioned related research, we can conclude that currently there is a lack of approaches that would directly link the formation and modeling of biotic processes in ecosystem based on random point patterns generated by remote sensing data. It is worth noting that both spatial point patterns and remote sensing images are able to work in different scales. It is achieved by the identification of events and territories, as well as by the different image resolutions. Therefore, we use this similarity in our methodology of environmental research.

Classification of remote sensing data. Nowadays, it is relatively easy to obtain an enormous amount of remote sensing data from open access or commercial databases^{27,28}. However, the analysis of remote sensing images is time consuming and may depend on human subjective assessment. Recently, various automated techniques of remote sensing image analysis have been used to facilitate these tasks^{29–39}. Both classical methods of image processing^{29–35} and more modern methods based on convolutional neural networks (CNNs)^{36–39} provide a satisfactory result. Nevertheless, the CNN based algorithms often achieve higher performance in most image analysis tasks. In particular, most remote sensing tasks involve selection of quantitative characteristics of the studied objects^{32,33}, which includes image segmentation and classification. CNN also offers very fast processing in the inference stage, assuming that the model is well trained initially. Nevertheless, the training is needed also for other known classification approaches, including support vector machine, random forest, k-nearest neighbors, etc.^{34,35}.

Despite the progress in classification techniques, the training phase and the dataset quality are still key factors in achieving high classification accuracy of CNN. Its qualitative construction requires considerable effort, especially for remote sensing tasks. First, due to the specifics of their generation, most of images are distorted by the atmosphere. Another factor that has a significant impact is the image resolution. Such aspects, significantly

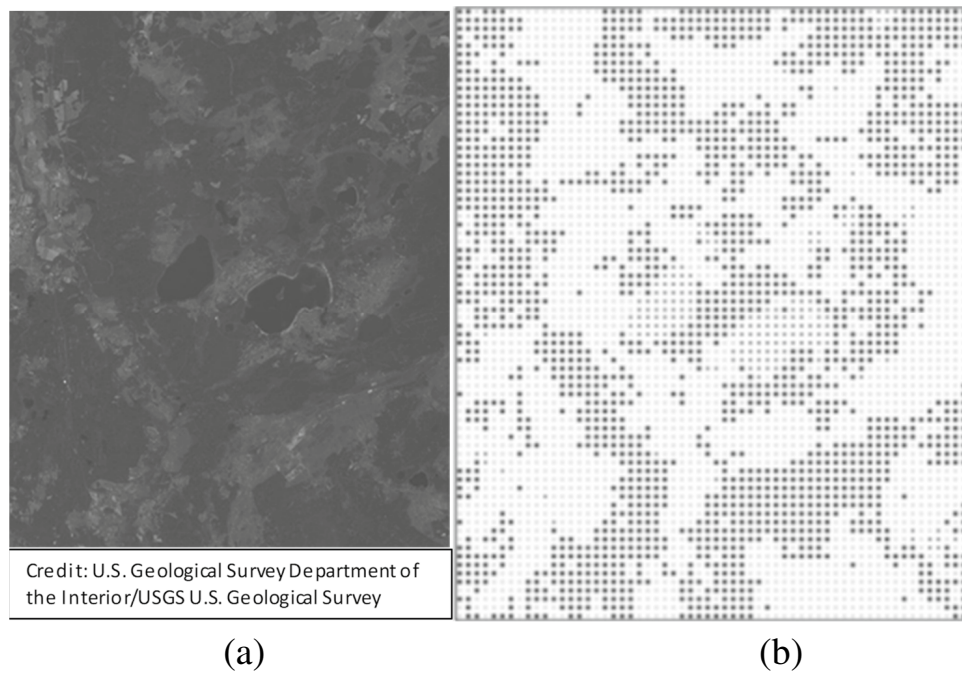


Figure 1. Remote sensing image (a)²⁸, and its point-pattern (b).

limit the scope of problems that can be considered and the choice of methods for solving them. High-resolution images (~ 1 m per pixel) of the desired object are required for local scale tasks, while lower-resolution images are sufficient for solving global scale tasks. In both cases, such images are quite large, and result in considerable computational complexity of their processing. High-quality high-resolution images are rarely available, so in most cases, images with a resolution of approximately 10 m per pixel are used for monitoring tasks. This aspect significantly limits the ability to classify required objects that might be represented by only a few pixels and makes it difficult to prepare training samples of different image classes due to their similarity. It is well known, that the accuracy of qualitative classification of images is directly proportional to the size of the training dataset. The well-known ImageNet³⁵ image dataset, which is frequently used as a benchmark for modern classification tools, contains more than one million labeled images.

Summarizing all abovementioned issues, we can conclude that CNN allows to perform initial training on known dataset or use the previously trained weights (transfer learning) with further fine-tuning on the target data. A significant limitation is that features of the target images must match the imported features, which is not always the case. Therefore, the generation of CNN training data for the classification task and their further analysis remain an unsolved urgent problem in environmental studies.

Spatial point patterns analysis in environmental studies. Tasks that require a statistical description of the sequence of events that occur at certain points in space often arise in many fields of technology and economics. Additionally, knowledge of the characteristics of the spatial distribution patterns of natural resources is crucial for developing an understanding in many fields of ecology. In the simplest one-dimensional case, the sequence of random events occurring in time can be characterized by random moments of their appearance. Such a sequence of events is often called a random flow³⁶ or a random point process (RPP). For a two-dimensional case, the presence of related events is even more important, provided that these events occurred within a short period of time. The analysis of RPP can establish arrangement patterns of random events and hence reveal their influence on each other and its type: mutual or asymmetric, or none of these. Establishing the properties of the process allows us to gain more information about the object itself.

Historically, object analysis using RPP was first applied in disciplines where the object could have been located on the map: astronomy, seismology, economics, sociology, materials science, etc.^{37,38}. Therefore, over the last two decades, SPPA has become increasingly popular in ecological and other research³⁹. The division of the input image into points and their subsequent classification allows the generation of various points, which correspond to different classes. By substituting each image fragment with a point at its center, we can produce a configuration displaying the location of fragments of the same class, as shown in Fig. 1²⁸.

Recently, ecologists have increasingly used such datasets to quantify the characteristics of observed spatial patterns. Most of the research aimed at deriving the hypotheses on underlying processes or testing hypotheses derived from ecological processes or theory. One of the main characteristics of the observed point configuration is the type of point location. The cluster, regular or random configuration types are distinguished according to the nature of the interaction between them. For cluster point configuration, the presence of interaction is assumed, causing the attraction of elements to each other, while for regular, the repulsion from elements to each

Training sample size	Accuracy, %, augmentation/no augmentation	Training time, s, augmentation/no augmentation
11,792	80/83	30/105

Table 1. Comparison of classification accuracy with and without data augmentation.

other. Finally, the random configuration assumes no interaction between its elements. The homogeneous Poisson point process is a fundamental model for RPP, and it also serves as a basis for more complex models. This process reflects the concept of "absolute" randomness, which is expressed by the absence of a definite internal structure of the set of points that correspond to this point pattern. A number of tests have been developed to test the hypothesis of complete spatial randomness. The most commonly used is the Clark-Evans test⁴⁰, which is based on examining the normal distribution of the standardized average distance to the nearest neighbor. Another way to analyze RPP is to calculate its second-order characteristics, including the K-function, g-function, J-function and others^{36,41,42}.

Wider opportunities are provided in the case of the simultaneous analysis of a mutual arrangement of two or more species. Examining biological objects, for instance, makes it possible to establish interspecific relations. Conventionally, point configurations are formed by manually plotting the locations of events or objects of study. This is a time-consuming process requiring knowledge of the exact distances between the objects of study, which are often distributed over large areas. Therefore, aerial photography images are the preferable source for the generation of spatial point patterns. Although they are detailed enough to reflect the location of the studied objects, they are suitable only for the analysis of local areas and relevant processes. On the other hand, in our opinion, images obtained by satellites are well suited for the study of more global processes. Therefore, in this paper, we propose a framework that uses satellite remote sensing data transformed to point patterns with the aim of describing biotic interactions on a global scale.

Spatial point patterns generation by remote sensing data classification

Remote sensing data augmentation based on the spatiotemporal similarity between images. Currently, the CNN based models are considered as the most effective in image classification, which outperform other methods in training, inference accuracy and generalization^{43–46}. To achieve a minimal training error CNN parameter must well align with training data. In the vast majority of practical tasks, the size of the training sample is insufficient, so that it is necessary to increase it. The traditional way to increase the training and test sets of images is the data augmentation^{47–49}.

Typically, the methods of data augmentation apply the various transformations to image data: rotation, shift, scaling, changing the color palette and their combinations^{47–49}. In general, the use of data augmentation can increase the accuracy of classification by 5–10%. However, existing data augmentation solutions the gain in accuracy is observed only in case of the significant difference between images. This is not applicable for remote sensing image fragments due to their proximity. In this case, augmentation produces copies of image fragments of the original sample and does not affect the classification accuracy (Table 1).

To increase the training sample size, we take advantage of the unique property of remote sensing images, namely, the periodicity of receiving such images. It is especially applicable for satellite imagery but can also generally be implemented for aerial imagery. Having the period of rotation of satellites around the Earth and the period of rotation of the Earth, we can collect a series of images of the target area of interest during a specified period of time. These will be similar images, each differently affected due to circumstances of remote sensing, and primarily by meteorological factors. Accurate distinguishing of the same area of the surface in these images is an effective kind of transformation that must be applied to restore the original image. The particular nature of the transformation does not matter, but the fact that images are different is essential. By selecting one of the images as the source and marking its fragments according to the set of preset classes, we can assign the same classes to the corresponding fragments of other images. In this way, an extended set of marked fragments is produced that can be used for both the training and test sets.

Dataset preparation. Training and test image samples were taken from the images of the Shatsk National Nature Park (SNNP), which were obtained by the Landsat 8 satellite in the visible spectral range²⁸. The data set was collected during 2017–2020. Images obtained during June–September were selected for the research, as they fully depict the vegetation of the investigated region. The images display a section of SNNP with an area of 450 km². The size of the original image is 3420 × 4380 pixels with a precision of approximately 10 pixels per meter. Based on the selected image, training, test and validation samples were generated with image patches of 10 × 10 pixels. The number of fragments in each category is 11,792, 1664, and 482, respectively. To extend these samples, it was proposed to supplement them with corresponding fragments present in a number of other images, since the generated dataset reflects the same area of the Earth's surface. For that purpose, we assigned class labels to such additional fragments according to the initial training sample and then conducted CNN training. As result, we have extended a dataset by classifying the fragments of supplementary images to the 47,213, 10,443 and 2182 fragments, respectively. Overall, by supplementing a number of images, we have further produced a following data samples: training—186,027; test—45,019; validation—22,349.

CNN training for the classification of remote sensing images. To classify the remote sensing images, we have chosen a powerful CNN configuration (EffNet)⁵⁰ and the simple CNN configuration (LeNet)⁵¹.

Data set capacity	Accuracy (top-1) %		Accuracy (top-5) %	
	EffNet	LeNet	EffNet	LeNet
11,792/1664/482	96/80/68	80/76/65	100/98/79	100/99/75
47,213/10,443/2182	96/86/77	86/88/67	100/99/90	99/99/89
186,027/45,019/22,349	97/90/89.7	87/89/88.9	100/99/95	99/99/93

Table 2. Classification accuracy for different CNN architectures.

The training has been conducted by the following parameters: optimizer—Adam; batch size—256; number of epochs—30; number of classes—6; learning rate—not defined explicitly. As shown in Table 2, an unexpected similarity has been observed in the classification results of different CNNs. Specifically, for the above-mentioned EffNet and LeNet, the corresponding results differ by no more than 2%. This can be explained by the domination of large-scale features in the investigated images. Therefore, for the current remote sensing images, the more complex CNN architectures, which are available nowadays, will not provide a sufficient advantage, especially considering the inevitable increase of computational complexity. Moreover, regardless of the CNN architecture, the proposed approach of data augmentation has improved the accuracy of classification by more than 10% (Table 2).

Simulation results and performance analysis

To analyze several configurations of points, the properties and characteristics of marked point processes are considered, which differ from the ordinary ones by the inclusion of an additional parameter—a marker that describes both qualitative and quantitative characteristic. To study these properties, we use the two-dimensional analogy of ordinary point processes characteristics, such as K , g , J —functions and others^{36,42}. Let's define the corresponding expressions for these functions. The expression for the K -function in the case of a single point configuration is given by:

$$\lambda K(r) = E_o(N(b(o, r)/\{o\})),$$

where λ is the number of points per unit area, and $K(r)$ denotes the average number of points in a circle with radius r whose center is located at a typical point that is excluded from consideration. In the case of multiple point configurations, this function looks similar:

$$\lambda_i K_{ij}(r) = E_{oi}(N_j(b(o, r)/\{o\})),$$

where λ_i is the number of points of type i per unit area, and $K_{ij}(r)$ is the average number of points of type j in a circle of radius r , the center of which is located at a typical point of type i . When considering all pairs of points separated by distance r , mark connection functions $p_{lm}(r)$ yield the conditional probability of the first point being of type l and the second of type m ^{10,41}. It gives a measure of dependency between two such points of the process a distance r apart. If the marks attached to the points are independent and equally distributed, then $p_{lm}(r) = p_l p_m$, where p_l denotes the probability of the point being of type l . Otherwise, $p_{lm}(r) > p_l p_m$ suggests a positive association between the two types, while $p_{lm}(r) < p_l p_m$ indicates a negative association. The Mark correlation function $k_{mm}(r)$ is a second-order summary statistic adapted to quantitatively marked patterns. $k_{mm}(r)$ visits all pairs of points separated by distance r , estimates the mean of the product of their marks, and divides this by the corresponding quantity taken over all pairs of points. It can be used to detect correlations in the case of mutual stimulation or inhibition of individuals⁴². If there is no interaction between points, then the function is equal to 1 for all distances. If $k_{mm}(r) > 1$, points located at distance r indicate a mutual stimulation or positive correlation. An opposite situation, $k_{mm}(r) < 1$, indicates mutual inhibition between points⁴². An example of applying the proposed approach to the analysis of remote sensing images is shown in Fig. 2. It displays the relationships between the components of a random point pattern that contains 4 constituents. These dependencies were built based on the assumption of the mutual presence of elements of one component in the vicinity of a typical point of another component. The number of elements of one type in the vicinity of another can serve as a measure of the compatibility of these types. In particular, the dependencies between point configurations for pairs of classes were constructed. Such dependencies may show, for example, trends in the development of the region's vegetation, expansion or contraction of areas of habitat of individual species.

The study of labeled point configurations using the characteristics of the 2nd order allows us to draw certain conclusions about a particular species as well as interspecific connections. Obviously, we deal with traditional characteristics when performing the analysis of point configuration of one/single type; hence, we can focus exclusively on labeled point configurations. A number of point configurations were selected for the SNPP study area. Analysis of these configurations can explain the dynamics of the ecosystem of this region. Any significant works that might affect the environment are prohibited in the territory of SNPP; hence, the formed point images reflect the natural processes that take place there. This is especially true for afforestation. A number of conclusions can be drawn from a series of graphs showing the relationship between forests and other objects, as presented in Fig. 2.

Let us consider these images individually (Fig. 3). The first image (Fig. 3a) shows the relationship between marked configurations describing areas of forest and fields. The level of dependency is the probability of the presence of elements of both at a certain distance. Based on the plot, we can conclude that even at distances up to 2 km (mark 200 on the graph), the probability of their joint occurrence is quite low (0.2). Another plot

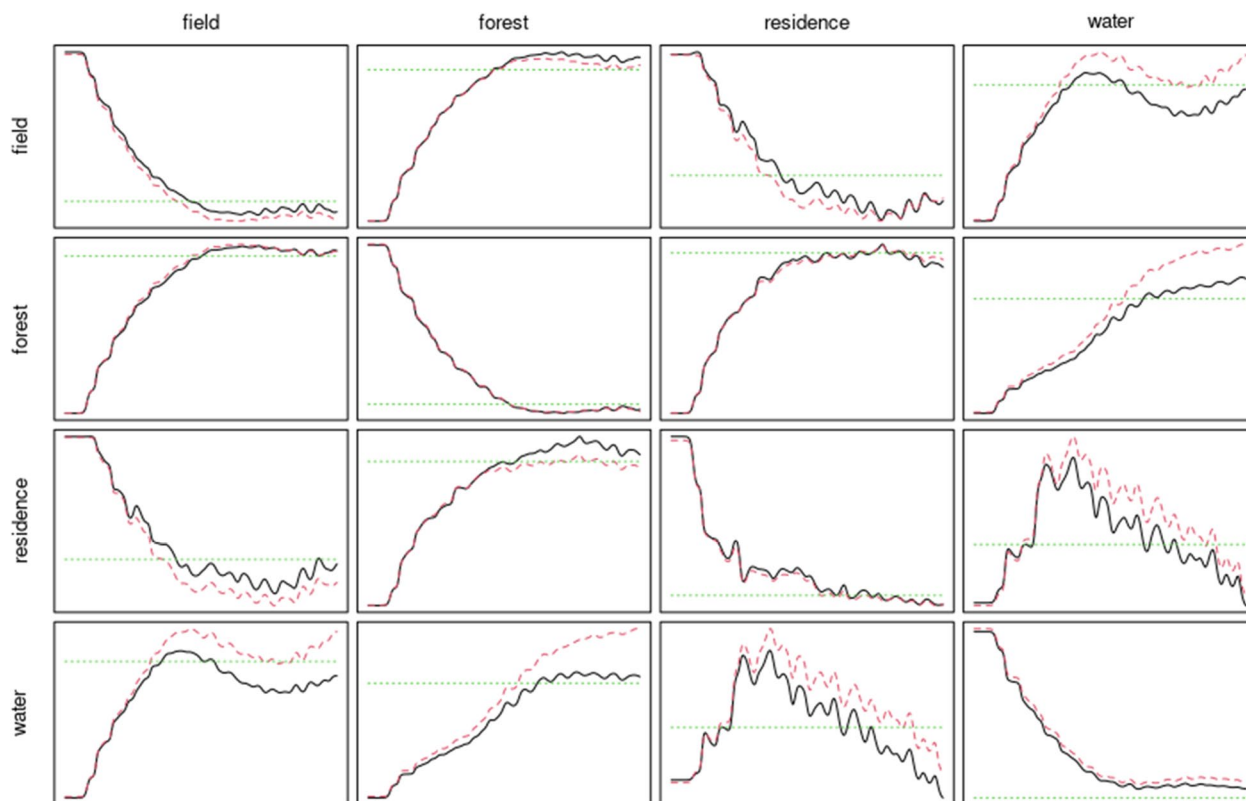


Figure 2. An array of mark connection functions for point patterns from Fig. 1b.

(Fig. 3b) indicates a low (0.01) probability of finding buildings in the forest, which is as expected for restricted areas. Similarly, comparing dependency plots of water-field and water-forest pairs (Fig. 3c, d), we can conclude that water resources are mostly surrounded by forests. The last plot (Fig. 3e) indicates the probability of finding buildings near the water.

Discussion and future research

Statistical analysis of remote sensing data for a particular region is both time consuming and resource-intensive task. Usually it requires to estimate the volumes of available species and their key characteristics and then evaluate their growth reliability. Today, it is possible to collect relevant data with the help of advanced technologies of remote sensing, satellite and aerial imaging. Nowadays, remote sensing of the Earth surface has multidisciplinary practical value, where the ecology of the landscape and associated statistical methods are becoming increasingly important⁴¹.

Among various available approaches for remote sensing data analysis the CNN-based solutions are the most promising ones in terms of accuracy. A significant part of research on CNN and other deep learning models is devoted to the problem of acquisition of the training, test and validation data. Therefore, the proposed approach for spatial point patterns generation opens a new perspective area of investigation that can significantly increase the potential of the remote sensing image analysis. The use of remote sensing data allows to analyze problems of different scales, but in most cases, it comes down to extracting the target information from the image. Interpretation of such information is a simple statistical analysis, such as calculating the number, area or perimeter, and other parameters. This does not reveal the full potential of such data. For the purpose of deeper analysis, the transformation of such data into spatial point images is proposed. Their apparatus can establish both simple quantitative characteristics such as intensity, and more complex, which reflect the relationships between different parts of the data. The results of our research give an example of establishing such relationships between different classes of objects on the example of SHNP.

Typically, the point patterns are formed based on the exact location of each element in the image. However, in our case, we artificially create an element of the point image as the center of the image fragment. Proposed approach is different from the fully stochastic solutions, which are known in the literature, such as⁵² where the point images are generated based of forest fires and the element of the point image is chosen as the center of the fire area. Among other examples: the original dataset is a pattern of small blobs, and the points are the blob centres; the original dataset is a collection of line segments, and the points are the endpoints, crossing points, midpoints etc.; the original dataset is a space-filling tessellation of biological cells, and the points are the centres of the cells. All cases prove that point process methodology can be more powerful or more flexible than the existing methodology for the raw image data. Nevertheless, the origin of the point pattern may lead to artefacts which must be considered in the analysis⁵³. In our research, we have shown a possibility of the application of spatial point patterns generation for the analysis of real data based on the example of SNNP satellite images.

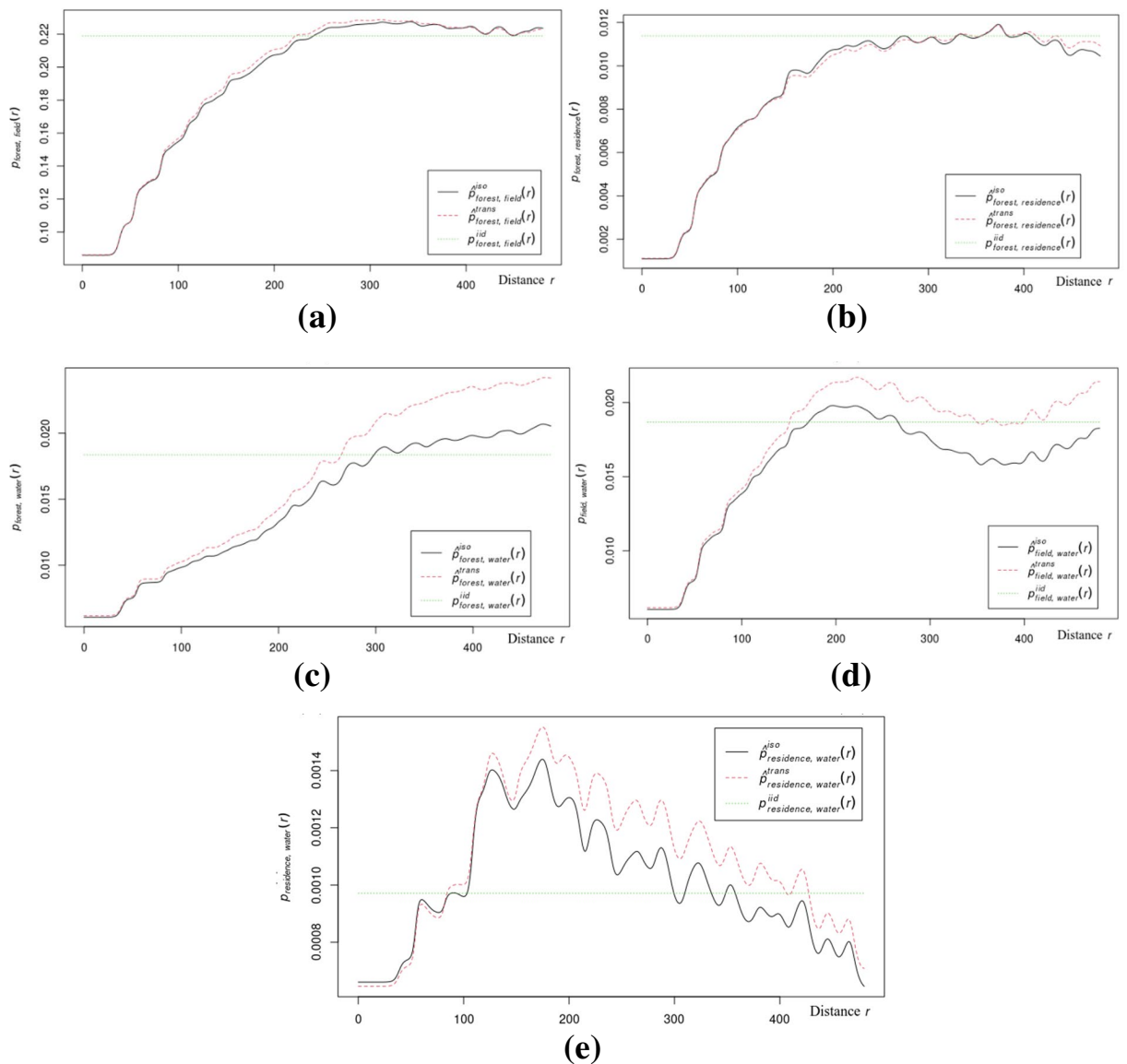


Figure 3. Plots of mark connected functions for bivariate marked point pattern. (a) “forest-water”; (b) “forest-residence”; (c) “water-forest”; (d) “water-field”; (e) “water-residence”.

In future research, there are still open areas for SPPA general adoption. In particular, this is due to some skepticism in using static point images to describe and analyze a dynamic ecosystem. Moreover, there is still an incomplete understanding of the correspondence between random processes that simulate the location of random point configurations and underlying ecological processes in the study region. Despite the obvious successes of both the theory of point processes and its numerous applications, including environmental monitoring, there are a number of problems limiting its wider application. First, apart from their complexity, there is only a small number of ready-to-use zero-models of RPPs to hand, as well as tests for their verification on real data. This means that each application requires its own mathematical apparatus for formalization.

Nevertheless, the advantage of using SPPA for the real data analysis is not only its ability to describe existing objects but also the ability to reflect existing connections between them. The latter, as already mentioned, is the main task of environmental monitoring. While the traditional and widely used SPPA apparatus for the analysis of real objects accounts only for their location, it is also important to account for their individual characteristics by applying models of marked random processes and determining their parameters.

Finally, there is still a wide area of research related to the usage of advanced CNN and other deep learning applications in combination with spatial point patterns, which may result in significant boost of the remote sensing data analysis.

Conclusions

The tasks of environmental monitoring require improvements to existing methods as well as the development of new methods for the analysis of remote sensing data. The most common source of such information is remote sensing images of the Earth's surface. The essential step in the analysis of such images is their classification. Most effectively, at present, this problem is solved by using convolutional neural networks. However, for such an approach to be successful, large training and test data sets are needed. To enhance such data, it is proposed to supplement them with additional augmented data, which is similar to the original data. Proposed data augmentation scheme outperforms the current best practices by at least 7% in classification accuracy. In addition, the methodology for the remote sensing images analysis is proposed based on a combination of image fragment classification by CNN and determining the relationship between the components of the marked random point field generated based on these fragments. The proposed approach allows to reflect the relationships between structural elements of the investigated areas of the ecosystem.

Data availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available because they are a fragment of other publicly available data.

Received: 21 April 2022; Accepted: 16 August 2022

Published online: 22 August 2022

References

- Appel, M., Lahn, F., Buytaert, W. & Pebesma, E. Open and scalable analytics of large earth observation datasets: From scenes to multidimensional arrays using SCIDB and GDAL. *ISPRS J. Photogramm. Remote Sens.* **138**, 47–56 (2018).
- Audebert, N., Saux, B. L. & Lefvre, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS J. Photogramm. Remote Sens.* **140**, 20–32 (2018).
- Ball J. E., Anderson D. T., & Chan C. S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* <https://doi.org/10.1117/1.JRS.11.042609> (2017).
- Proceedings of the Royal Society B: Biological Sciences*. Vol. 282. 20141657 (2015).
- Velázquez, E., Paine, C. T., May, F. & Wiegand, T. Linking trait similarity to interspecific spatial associations in a moist tropical forest. *J. Veg. Sci.* **26**, 1068–1079 (2015).
- Ben-Said, M. Spatial point-pattern analysis as a powerful tool in identifying pattern-process relationships in plant ecology: an updated review. *Ecol. Process.* **10**, 1–23 (2021).
- Watt, A. S. Pattern and process in the plant community. *J. Ecol.* **35**, 1–22 (1947).
- Pielou, E.C. *Mathematical Ecology; Number 574.50151 P613 1977*. (Wiley, 1977).
- Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
- Brown, C., Law, R., Illian, J. B. & Burslem, D. F. Linking ecological processes with spatial and non-spatial patterns in plant communities. *J. Ecol.* **99**, 1402–1414 (2011).
- Detto, M. & Muller-Landau, H. C. Fitting ecological process models to spatial patterns using scalewise variances and moment equations. *Am. Nat.* **181**, E68–E82 (2013).
- May, F., Huth, A., & Wiegand, T. Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings. Biological sciences* **282**(1802), 20141657. <https://doi.org/10.1098/rspb.2014.1657> (2015).
- Kerr, J. T. & Ostrovsky, M. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* **18**, 299–305 (2003).
- Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P. & Saatchi, S. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **32**, 203–221 (2008).
- He, J., Zhang, L., Wang, Q. & Li, Z. Using diffusion geometric coordinates for hyperspectral imagery representation. *IEEE Geosci. Remote Sens. Lett.* **6**(4), 767–771 (2009).
- Lechner, A.M., Foody, G.M., & Boyd, D.S. Applications in remote sensing to forest ecology and management. *One Earth* **2.5**, 405–412 (2020).
- Arévalo, P., Olofsson, P. & Woodcock, C. E. Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series: A test methodology for REDD+ reporting. *Remote Sens. Environ.* **238**, 111051 (2020).
- Gillespie, T.W. *et al.* Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* **32.2**, 203–221 (2008).
- Lausch, A., Erasmi, S., King, D. J., Magdon, P. & Heurich, M. Understanding forest health with remote sensing-part II—A review of approaches and data models. *Remote Sens.* **9**(2), 129 (2017).
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., *et al.* Monitoring vegetation systems in the Great Plains with ERTS. in *NASA Special Publication*. Vol. 351. 309 (1974).
- Chen, J. M. & Black, T. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* **15**, 421–429 (1992).
- Zha, Y., Gao, J. & Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **24**, 583–594 (2003).
- Zhao, S. *et al.* Remote detection of bare soil moisture using a surface-temperature-based soil evaporation transfer coefficient. *Int. J. Appl. Earth Obs. Geoinf.* **12**, 351–358 (2010).
- Gao, B. C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**, 257–266 (1996).
- Wan, Z. & Dozier, J. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.* **34**, 892–905 (1996).
- Xu, H., Wang, Y., Guan, H., Shi, T. & Hu, X. Detecting ecological changes with a remote sensing based ecological index (RSEI) produced time series and change vector analysis. *Remote Sensing* **11**, 2345 (2019).
- List of Top 10 Sources of Free Remote Sensing Data* (2017).
- USGS Earth Explorer: Download Free Landsat Imagery* (2021).
- Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote. Sens.* **65**, 2–16 (2010).
- Li, M., Zang, S., Zhang, B., Li, S. & Wu, C. A review of remote sensing image classification techniques: The role of spatio-contextual information. *Eur. J. Remote Sens.* **47**, 389–411 (2014).
- Gómez-Chova, L., Tuia, D., Moser, G. & Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **103**, 1560–1584 (2015).
- Alajlan, N., Bazi, Y., Melgani, F. & Yager, R. R. Fusion of supervised and unsupervised learning for improved classification of hyperspectral images. *Inf. Sci.* **217**, 39–55 (2012).

33. Csillik, O. Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. *Remote Sens.* **9**, 243 (2017).
34. Thanh Noi, P. & Kappas, M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* **18**, 18 (2018).
35. Jiang, S., Zhao, H., Wu, W., & Tan, Q. A novel framework for remote sensing image scene classification. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, 42 (2018).
36. Baddeley, A. *Spatial Point Process Modelling and Its Applications*. Vol. 20. (Publicacions de la Universitat Jaume I, 2004).
37. Vasudevan, K., Eckel, S., Fleischer, F., Schmidt, V. & Cook, F. Statistical analysis of spatial point patterns on deep seismic reflection data: A preliminary test. *Geophys. J. Int.* **171**, 823–840 (2007).
38. Cheng, Y. & Luo, J. Statistical analysis of metastable pitting events on carbon steel. *Br. Corros. J.* **35**, 125–130 (2000).
39. Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A. & Wiegand, T. An evaluation of the state of spatial point pattern analysis in ecology. *Ecography* **39**, 1042–1055 (2016).
40. Clark, P. J. & Evans, F. C. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* **35**, 445–453 (1954).
41. Stoyan, D., & Penttinen, A. Recent applications of point process methods in forestry statistics. *Stat. Sci.* **2000**, 61–78 (2000).
42. Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns*. Vol. 70. (Wiley, 2008).
43. Brodrick, P. G., Davies, A. B. & Asner, G. P. Uncovering ecological patterns with convolutional neural networks. *Trends Ecol. Evol.* **34**, 734–745 (2019).
44. Liu, S., Luo, H., Tu, Y., He, Z., & Li, J. Wide contextual residual network with active learning for remote sensing image classification. in *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*. 7145–7148 (IEEE, 2018).
45. Lee, H. & Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **26**, 4843–4855 (2017).
46. Cheng, G., Xie, X., Han, J., Guo, L. & Xia, G. S. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 3735–3756 (2020).
47. Lewy, D., & Mandziuk, J. An overview of mixing augmentation methods and augmentation strategies. *arXiv preprint arXiv:2107.09887* (2021).
48. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q.V. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501* (2018).
49. Naveed, H. Survey: Image mixing and deleting for data augmentation. *arXiv preprint arXiv:2106.07085* (2021).
50. Freeman, I., Roese-Koerner, L. & Kummert, A. Effnet: An efficient structure for convolutional neural networks. 25th IEEE international conference on image processing (ICIP). *IEEE* **2018**, 6–10 (2018).
51. LeCun, Y. *et al.* Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541–551 (1989).
52. Raeisi, M., Bonneau, F. & Gabriel, E. A spatio-temporal multi-scale model for Geyer saturation point process: Application to forest fire occurrences. *Spatial Stat.* **41**, 100492 (2021).
53. Baddeley, A. Analysing spatial point patterns in R. in *Workshop Notes Version*. Vol. 3 (2008).

Acknowledgements

This research was supported by the National Academy of Sciences of Ukraine, project #0117U000519, Ministry of Education and Science of Ukraine, project #0120U102193 and by the Slovak Research and Development Agency, project number APVV-18-0214.

Author contributions

The contribution of each of the authors in this research articles can be divided as follows: conceptualization—R.K. and O.A.; methodology—R.K.; software—O.L.; validation—O.A.; formal analysis—T.M. and J.G.; investigation—R.K., O.L.; data curation—O.A.; writing—original draft preparation—R.K.; writing—review and editing—B.R., T.M.; supervision—B.R.; proofreading and funding acquisition—J.G. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to R.K. or T.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022