UDC 621.391 https://doi.org/10.15407/vidbir2025.53.071

OBJECT LOCALIZATION IN IMAGES BASED ON CLUSTERING OF ATTENTION ZONES

R. Y. Kosarevych [1, 2], O. A. Lutsyk [1], B. P. Rusyn [1, 2], D. V. Ivchenko [1]

[1] Karpenko Physico-Mechanical Institute of the National Academy of Sciences of Ukraine, Lviv;

[2] Lviv Polytechnic National University, Lviv

E-mail: r.y.kosarevych@gmail.com

A novel methodology for the localization of objects within images is proposed. Contrary to the well-known deep learning approach, which involves the use of samples with matched boxes around similar objects to find their exact position on an image, a method based on object properties as the location of pixels with similar characteristics is developed. The notion of cluster point patterns to detect single parts of an object is used. The concept of an entire object as a composite of proximate, amalgamated clusters is proposed.

Keywords: *object detection, point pattern, classification.*

ЛОКАЛІЗАЦІЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ ШЛЯХОМ КЛАСТЕРИЗАЦІЇ ЗОН УВАГИ

Р. Я. Косаревич [1, 2], О. А. Луцик [1], Б. П. Русин [1, 2], Д. В. Івченко [1]

[1] Фізико-механічний інститут ім. Г. В. Карпенка Національної академії наук України, Львів;

[2] Національний університет "Львівська політехніка"

Запропоновано нову методологію локалізації об'єктів на зображеннях. На відміну від добре відомого підходу глибокого навчання, який передбачає використання зразків зі обмежувальними рамками навколо подібних об'єктів для визначення їхнього точного положення на зображенні, розроблено метод на властивостях об'єкта як місця розташування пікселів з подібними характеристиками. Використано поняття кластерів точкових образів для виявлення його окремих частин. Концептуалізовано весь об'єкт як сукупність близьких, об'єднаних кластерів. Запропоновано підхід, який використовує притаманну об'єктам властивість, зумовлену їхньою внутрішньою структурою. Для будь-якого об'єкта його вродженим компонентом є забарвлення. Це забарвлення є постійним в області, визначеній межами об'єкта, або є комбінацією кількох забарвлень. Отже, якщо ідентифікувати всі такі фрагменти постійного кольору, можна цим обмежити область на зображенні, яку займає об'єкт.

Припущення ґрунтуються на гіпотезі, що будь-який скінченний та обмежений фізичний об'єкт на зображенні є множиною точок, які мають подібні характеристики та розташовані у вигляді кластерних груп. Це не обов'язково мають бути однозв'язні множини. Водночас фон об'єкта складається з множин точок, які регулярно та/або випадково розташовані у площині зображення.

Сучасні підходи дають змогу досить добре вирішувати задачу локалізації об'єктів на зображенні, але мають суттєве обмеження, яке виникає через необхідність навчання локалізації на наборі зразків, подібних до об'єктів, які зустрічаються в поточному завданні. Якщо під час навчання таких зразків не було, система не зможе їх розрізнити. Саме тому методи, які не потребують етапу навчання, а використовують характеристики, властиві об'єкту для локалізації, можуть мати переваги в конкретних завданнях.

Ключові слова: *локалізація об'єктів, точковий шаблон, класифікація.*

Introduction. In computer vision, object localization is the process of identifying regions of interest in an image or video which contain a specific object. This task is a key component of more sophisticated systems aimed at automatic recognition and analysis of visual data. Typically, the result of localization is a bounding box which

precisely delineates the location of the object. This box provides information about the spatial location of the object, which is critical for many practical applications.

To fully understand object localization, it is necessary to clearly distinguish it from other key tasks of computer vision, such as segmentation, classification, or detection. These tasks are often interrelated, but adress different questions about the content of the image: segmentation is the most detailed task, dividing the image into segments, assigning each pixel a label of a specific class, thus highlighting the exact contours of the objects, rather than just bounding boxes. Localization, in general, involves not only classifying the object, but also determining its exact position in the image using a bounding box. This approach usually assumes that there is only one object in the image that needs to be found. The classification task consists in assigning a single class label from a defined set to an image. Object detection is a combination of classification and localization, which detects and outlines with bounding boxes several objects of different classes in a single image. This is one of the most common computer vision tasks, which is fundamental to video surveillance systems and autonomous vehicles.

Modern object localization and detection are mostly based on deep learning architectures, particularly convolutional neural networks (CNNs). CNNs efficiently detect local and global features in an image, which helps in the accurate identification of objects and their boundaries. Instead of using traditional, heuristic methods such as SIFT [1, 2] or SURF [3], which require manual feature extraction, deep learning has made it possible to automate this process, significantly increasing accuracy and speed. Neural networks learn to recognize patterns and regularities in input data, which allows us to detect objects even under different background and lighting conditions. This is achieved by using several types of layers: convolutional layers for feature extraction, pooling layers to reduce the spatial dimension of the data, and fully connected layers for classification.

The main architectures for object localization and detection are currently two-stage and one-stage models such as R-CNN based [4–6] and YOLO [7].

Two-stage architectures, such as Faster R-CNN [8], are among the most well-known in the field of object detection. They work according to a two-stage principle. In the first stage, the model analyzes the image using a convolutional neural network to create a feature map. Then, based on this map, a special network of regional proposals (Region Proposal Network, RPN) generates a set of potential regions where the objects are likely to be located. In the second stage, these proposed regions pass to another part of the network, where they are classified to determine which object is in each frame, and their bounding boxes are refined for maximum accuracy. The advantage of Faster R-CNN is high accuracy, especially in detecting small and partially overlapping objects. The speed and efficiency of the algorithm are achieved by using a shared feature map between the RPN and another CNN, making it one of the best for object detection, especially for applications where accuracy is a priority.

In contrast to two-stage models, single-stage models such as YOLO perform object detection in a single pass through the neural network. The principle of their operation is that they divide the image into a grid and for each cell of this grid predict the bounding boxes and object class probabilities. The main advantage of YOLO models is their extreme speed, which makes them ideal for applications which require real-time analysis, such as autonomous vehicles. Newer versions continue to optimize the architecture, introducing blocks for efficient feature extraction and attention mechanisms for better focusing on important regions of the image. This allows us to significantly improve accuracy, especially when detecting small objects, without losing their main advantage – speed.

Despite the undeniable advantages of the above-mentioned methods, they have one important drawback – the training stage. As is known [9], this stage is the most

expensive, in terms of resources and time, in preparing the training sample. In addition to dividing the volume of this sample into classes, which should be thousands, or even tens of thousands of samples, it is still necessary to specify the limiting boxes for objects for each image. Also it is difficult to imagine a universal training sample for the entire variety of object classes. Therefore, there is still a need to develop methods for locating objects on images that use paradigms other than deep learning.

In this paper, an approach is proposed which uses an inherent property of objects due to their internal structure. For any complexity of an object, its innate component is its coloring. This coloring is constant in the area defined by the boundaries of the object, or it is a combination of several coloring in this area. So if we identify all such fragments of constant color for an object, we can thereby limit the area in the image which the object occupies.

Methods and materials. Any events or objects which can be represented on the surface by multiple points, such as planting sites, bird or insect breeding grounds, earthquake epicenters, diseases, industrial, cultural, or scientific locations will reproduce random point configurations. The object itself can be considered as a configuration of points with different properties. Suitable model for such kind of object model is random point pattern (RPP) [10, 11].

One of the main characteristics of RPP is their appearance: clustered, regular or random, which indicates the nature of the arrangement of the elements and accordingly the type of interaction between them (Fig. 1). For cluster RPP, it is assumed that there is an interaction that leads to the attraction of elements to each other, for a regular one, there is an interaction that leads to the repulsion of elements from each other, and for random RPP, there is no interaction between the elements. This feature allows it to be used in the analysis of images, since the image objects, namely the textures representing them, can be matched to one or another type of RPP or a their combination [11]. The homogeneous Poisson point process is a fundamental model for RPP which typify the concept of complete spatial randomness (CSR) of events; it also serves as a basis for building more complex models. Perhaps even more importantly, the Poisson process can be used as a reference model to distinguish point patterns which are random or tend to form clusters (aggregations) or, conversely, to repel, that is, a regular arrangement at which the distance between points cannot be less than some threshold. A large number of tests of the CSR hypothesis have been developed. Most CSR tests are constructed as follows. A summary characteristic is estimated for the data and compared with the relevant general summary characteristic for a Poisson process. If there is a significant difference between both characteristics, the Poisson null hypothesis is rejected. The tests may be based on either numerical summary characteristics, i.e., a single value, or functional summary characteristics, i.e., a function of the distance between the points. The Clarke Evans test [12] is the most frequently used for the testing of CSR. It is based on similarity with a normal distribution of the nearest neighbor distance standardized mean value of the RPP elements. Test of the CSR can classify the point pattern as cluster, regular or random. (Fig. 1) To obtain a point pattern from an image different ways can be used. To do this it is necessary to build associations between an image fragment and a point. This can be done with different image processing techniques. For example, point patterns can be obtained by splitting the image in a binary template each of which is a set of pixels of certain intensity from the image range. Another approach is to set a threshold to get a binary image and next thin it to get the points of bifurcation. A bit complicated mean is described in [13], according to which an image is evenly divided into patches and local maxima of patch intensities histograms are determined. If one can choose a certain intensity from a maxima set and mark every patch with this intensity, a patch centers will be displayed as point pattern.

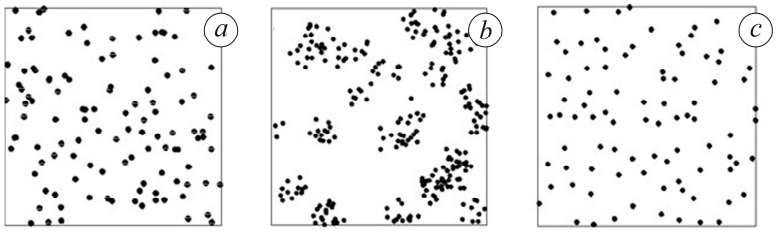

Fig. 1. Types of RPP: *a* – random; *b* – clustered; *c* – regular.

Approach and Results. Our assumption is based on the hypothesis that any finite and bounded physical object represented in the image is a set of points which have similar color characteristics and are arranged in the form of cluster groups. These do not necessarily have to be simply connected sets. At the same time, the background of the object consists of sets of points which are regularly and/or randomly located in the image plane. Let us describe the meta-algorithm for the proposed approach.

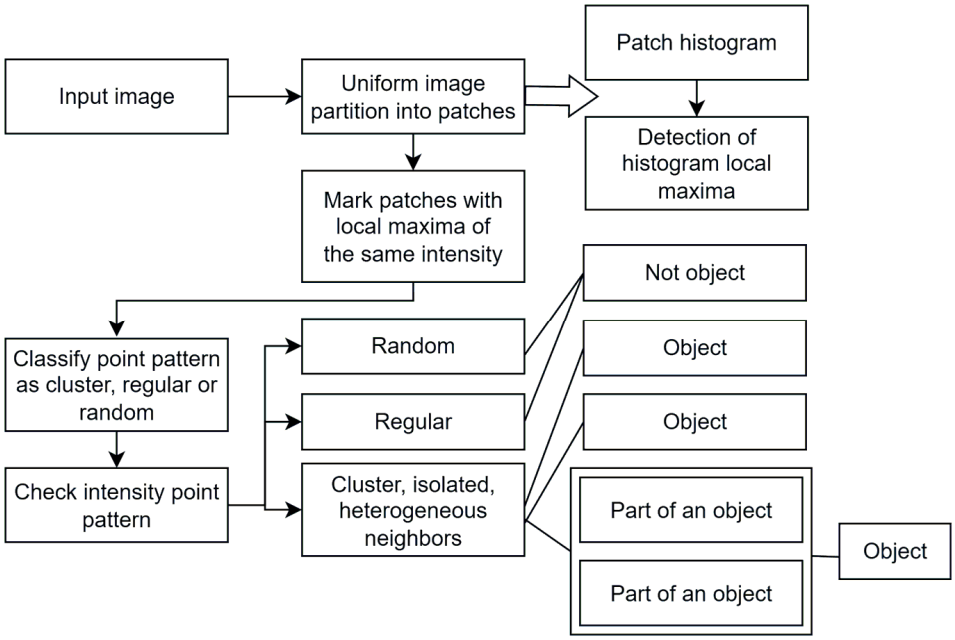

Fig. 2. A flowchart of the proposed approach.

1. At the beginning, we need to select the image brightnesses that form cluster formations. For this, it is advisable to average the brightnesses in the image, since even in the seemingly locally uniform color neighborhood of the image there is a non-zero brightness dispersion. Use of a median filter will reduce it.

2. Separately for each of the brightnesses available in the dynamic range of the image, we form a random point configuration using the approach described earlier [13].

3. We determine the type of point configuration for each brightness.

4. We select separate groups of points within the cluster formations and set their centers.

5. We combine the centers of groups of points of different brightnesses based on their proximity.

6. We determine the boundaries of the object as a union of sets of image points which belong to the corresponding cluster configurations.

We have summarized the algorithm in the form of a flowchart in Fig. 2.

Fig. 3 shows examples of the proposed approach. For comparison, a method based on the YOLO approach was used.

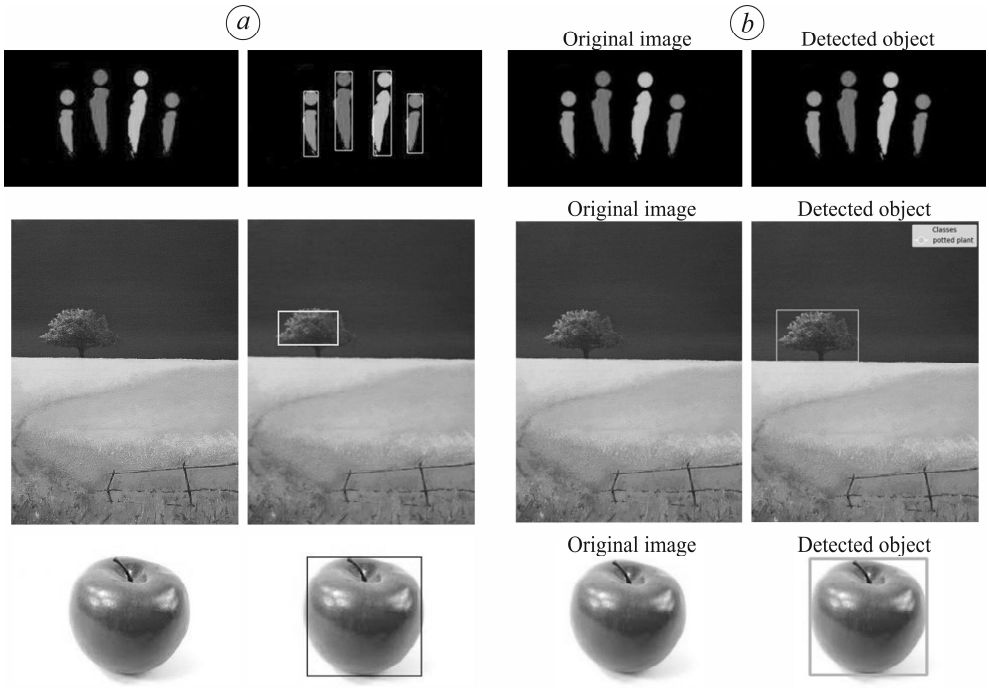

Fig. 3. Examples of object detection by the proposed approach and YOLO model:
a – proposed approach; *b* – YOLO model.

As can be seen from the presented results the proposed approach (Fig. 3*a*) gain nearly such good results as the YOLO model (Fig. 3*b*) which is the benchmark object detection means. It should be noted that the proposed approach also allows obtaining results in the case of specific objects in the image for which there are no training examples in the database. This can be achieved by the use of a set of heuristic rules, which may require changes in individual parameters for different tasks. It should also be noted that due to the specific feature of the proposed approach, which includes image preprocessing, the dimensions of the bounding rectangle may vary slightly.

It is also necessary to note the limitations of the proposed approach with regard to partially closed objects. Since the main source of information is only visual characteristics, it is impossible to localize an object based on formally absent features. Thus, only its visible part will be considered an object.

CONCLUSIONS

Localizing objects in images is a fundamental task of computer vision, which is the basis for numerous innovative applications. Modern approaches allow us to solve the problem of detecting objects in an image quite well, but they have a significant limitation. Their success in object localization is largely based on the high accuracy of their classification abilities. Their limitation arises from the need to train localization on a set of samples similar to objects that occur in the current task. If there are no such samples during training or an error occurs during classification, the system will not be able to distinguish them. That’s why a method based on the inherent characteristics of the object, which do not need to be established based on the studies of similar objects, has been proposed. Such approaches still can have an advantage in specific detection tasks.

1. Lowe, D.G. Object Recognition from Local Scale Invariant Features. *Proceedings of the 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, September 20–27, 1999; pp 1150–1157.https://doi.org/10.1109/ICCV.1999.790410
2. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* **2004,** *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94
3. Ba, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up Robust Features. *Computer Vision – ECCV 2006. Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria, May 7–13, 2006; Leonardis, A., Bischof, H., Pinz, A. Eds.; Lecture Notes in Computer Science, Vol. 3951; Springer: Berlin, Heidelberg 2006; pp 404–417. https://doi.org/10.1007/11744023_32
4. Zitnick, C.L.; Dollar, P. Edge Boxes: Locating Object Proposals from Edges. *Computer Vision – ECCV 2014. Proceedings of the 13th European Conference on Computer Vision*, Zurich, Switzerland, September 6–12, 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. Eds.; Lecture Notes in Computer Science, Vol. 8693; pp 391–4050. https://doi.org/10.1007/978-3-319-10602-1_26
5. Girshick, R.; Donahue, J.; Darrell, T.; Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 23–28, 2014; pp 580–587. https://doi.org/10.1109/CVPR.2014.81
6. Girshick, R. Fast R-CNN. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, December 7–13, 2015; pp 1440–1448. https://doi.org/10.1109/ICCV.2015.169
7. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, June 27–30, 2016; pp 779–788. https://doi.org/10.1109/CVPR.2016.91
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2016**, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
9. Zhao, W.; Fu, H.; Luk, W.; Yu, T.; Wang, S.; Feng, B.; Yang, G. F-CNN: An FPGA-Based Framework for Training Convolutional Neural Networks. *Proceedings of the 2016 IEEE 27th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, London, United Kingdom, July 6–8, 2016, pp 107–114. https://doi.org/10.1109/ASAP.2016.7760779
10. Stoyan, D.; Penttinen, A. Recent Applications of Point Process Methods in Forestry Statistics. *Statistical Science* **2000**, *15*(1), 61–78. https://doi.org/10.1214/ss/1009212674
11. Illian, J.; Penttinen, A.; Stoyan, H.; Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns*; John Wiley & Sons, 2008. https://doi.org/10.1002/9780470725160
12. Clark, P.J.; Evans, F.C. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology* **1954**, *35*(4), 445–453. https://doi.org/10.2307/1931034
13. Kosarevych, R.; Lutsyk, O.; Rusyn, B. Detection of Pixels Corrupted by Impulse Noise Using Random Point Patterns. *The Visual Computer* **2022**, *38*(11), 3719–3730. https://doi.org/10.1007/s00371-021-02207-1

Одержано 17.07.2025

ORCID iDs

Kosarevych R.Y. https://orcid.org/0000-0001-9108-0365
Lutsyk O.A. https://orcid.org/0000-0003-1707-3532
Rusyn B.P. https://orcid.org/0000-0001-8654-2270
Ivchenko D.V. https://orcid.org/0000-0002-6715-5782