

# Comparison of Multi-Object Recognition Models

Vitalius Parubochyi  
Department of System Design  
Ivan Franko National University  
Lviv, Ukraine  
vitalius.parubochyi@lnu.edu.ua

Roman Shuvar  
Department of System Design  
Ivan Franko National University  
Lviv, Ukraine  
shuwar@electronics.lnu.edu.ua

**Abstract**—The multi-object recognition on the single image is one of the most difficult issues of the object recognition. However, this particular model of the object recognition is best suited to the requirements of a large number of application systems. The paper analyzes the main aspects of the multi-object recognition, examines the problems of the multi-object recognition and the ways of their solution. Comparison of the basic approaches to the multi-object recognition is realized, and the possibility of adapting these approaches to different types of recognition systems is analyzed.

**Index Terms**—multi-object recognition, object search, object recognition, segmentation, exhaustive search

## I. INTRODUCTION

Multi-object recognition is one of the most important, but the most complex areas of the object recognition. At the same time, the main problem in solving this issue isn't the object recognition process, but its search, scaling and localization in relation to other objects in the image. The solution to this problem providing in two main areas: the image segmentation and exhaustive object search.

For image segmentation usually considers that the characteristics of target objects are unknown and the partition of the image on objects is carried out due to the color and texture features of the image. This approach to the object search in the image has found application in a large number of recent researches [1-4].

However, the color and texture features of the image aren't a universal way for object search, since they take into account only the physical properties of objects and don't analyze the context information of the image. In particular, for the recognition of complex objects or living creatures, the most important is the context-sensitive information. Therefore, along with methods based on the segmentation, the exhaustive search methods of objects have been acquired the significant development [5-8]. The main advantage of search methods is the possibility of a more flexible selection of approaches for analyzing image fragments. However, such methods require a much larger amount of the computation, since it is necessary to analyze the entire image multiple times.

Multi-object recognition has the number of other problems, which need to be considered for choosing a model for searching objects in the image.

In this paper, the problem of the multi-object recognition is analyzed, as well as the most common and effective object

search models for the multi-object recognition are considered. In addition, the possibility of adapting these models to different types of systems is considered.

## II. MULTI-OBJECT RECOGNITION PROBLEMS

Although the complexity of the multi-object recognition on the single image is primarily due to the search and localization of objects on the image, along with these problems, there are several other reasons that greatly complicate the practical use of the multi-object recognition systems. Let's consider the main ones:

### 1) Computational complexity

The solution of an issue of the multi-object recognition requires the multiple analysis of image fragments of different sizes in order to identify objects on them, as well as identify and classify found objects. However, in most cases, the analysis and recognition of individual fragments of an image can be done independently of each other, which allows using the parallel computing for implementation of the recognition process.

### 2) Presence of objects information

In many cases, object search is vastly due to the context of the data, which requires the additional information that could be used to set the properties of the object. On the other hand, the presence of the additional information makes the system a more complex and less flexible, which makes it much more difficult to adapt to other types of problems.

### 3) Object scalability

The size of objects can be very different, so window size for selecting image fragment also must changing to provide the correct size for recognition the full object or its part. And the number of these fragments can be really large.

### 4) The irregular shape of objects

Most objects have not only different sizes but also incorrect shape. So the search for such objects cannot be performed by analyzing the same type of fragments, which greatly increases the number of possible variants for analysis.

### 5) Overlap of objects

Objects can overlap on one another, which can also complicate the process of their recognition. At the same time, it should also take into account that fragments of other objects can hugely affect the recognition of other objects. Therefore

the correct representation of the complex shape of objects has a significant role. Another aspect of this problem may be the impossibility of fully object recognition and the need to transition to a partial recognition of complex objects.

#### 6) Search for complex objects

The color and texture heterogeneity of the object may be due to its complexity. In this case, the object can look like as the overlap of several other objects, which can hugely complicate its search and recognition.

### III. MULTI-OBJECT DETECTION MODELS

In the most cases, multi-object recognition models strongly depended on methods of searching for potential objects in the image and their recognition. However, this approach can be nonoptimal, as the object search method becomes a bottleneck, requiring a significant amount of computing operations and time, and generally slows down image recognition considerably. The one way to optimize this process is to integrate the search process directly into the recognition process. Although this approach is much more effective in terms of computational cost and effectiveness, it is less flexible. The main advantage of these methods is a possibility to adjust the object search using results of their recognition. However, since it is developed directly for a specific model of object recognition system, it is more difficult to adapt for use in other systems.

#### A. Image Sequential Block Analysis

The most simple approach to implementing object search. The main idea of the method is to split the image into blocks, each of which described by the position (usually the coordinates of the upper left corner), as well as the linear dimensions - height and width. Depending on the task, for this method defining the resizing range, as well as the displacement step of the blocks.

Despite the simplicity of this approach, its main disadvantage is a significant number of blocks that need to be analyzed. Optimization of the method is possible by choosing the range of block sizes and the block displacement step, but the correct choice of these parameters requires the availability of information about the nature of the images which are processing.

#### B. Selective Search

Selective search proposed by J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W. M. Smeulders [9] is a combination of segmentation and a full search of objects. The basis of the method is the image segmentation using the algorithm proposed by P. F. Felzenszwalb and D. P. Huttenlocher [3], after which segmented regions are estimating based on the color, texture and spatial characteristics. Then regions with the closest features are uniting using computed estimation.

The usage of the textural and spatial features of the image, as well as the quick method of selecting regions in the image, led to the widespread practical application of the selective search. For example, the method was used in the neural

network for multi-object recognition, called R-CNN [10], and its modifications, in particular, Fast R-CNN [11].

#### C. Objectness

The Objectness method proposed by Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari [12]. The basis of the method is the calculation for each region in the image five signals: Multi-scale Saliency (MS), Color Contrast (CC), Edge Density (ED), Superpixels Straddling (SS) and the Location and Size (LS). Based on these signals, the Bayesian classifier is built, which learns to evaluate the regions of the image.

The main difficulty of using Objectness method is the need to train a classifier on a large database of images.

#### D. Category Independent Object Proposals

Category Independent Object Proposals [13, 14] is very close to the selective search [9]. At the first stage, the method uses the hierarchical segmentation of the image, after which each region described and estimated. The main difference from selective search is the use of a learning model to choose a better region.

#### E. Multiscale Combinatorial Grouping

The Multiscale Combinatorial Grouping [15, 16] method based on the use of image segmentation of different linear sizes to produce multiple sets of segments and their combinations. The process is carried out in several stages. In the first step, the image is scaled to get multiple images with different linear sizes. After that, the all images are segmented. In the second stage, the resulting segments scaled to the size of the original image, after that segments combined based on their spatial position. This approach allows solving the problem of excessive segmentation of a high-resolution image and can also solve the problem of complex object search.

#### F. MultiBox Detection

The main disadvantage of neural networks R-CNN [10] and Fast R-CNN [11] constructed using selective search is the speed of the objects search in the image. To solve this problem were proposed new types of neural networks - YOLO [17, 18] and SSD [19]. The main difference between these neural networks was that the object search method directly added to the neural network, and instead of the detected object region, the complete image used at the input of the network.

The objects search implemented in SSD [19] is to apply the object search method at each level of the network. In this case, for each level of the network, the image is divided into a set of fragments each of which corresponds to several windows. Each window is adjusted and evaluated. Based on the estimation, the network associates the object with a particular class of objects. The main advantage of this approach is the automatic solution of the image scaling problem since at each level the network analyzed objects of another size. However, this method requires information about the identity of the object to a particular class, which limits the number of available classes, and the addition of other classes requires retraining of the entire system.

## IV. COMPARISON OF MULTI-OBJECT RECOGNITION MODELS

Although the list of methods used for objects search is not complete and there are other methods for finding objects [20, 21], however, discussed methods allow to determine the basic models of the multi-object recognition:

1) *Sequential Block Analysis*

The simplest example of this model is the image sequential block analysis.

Advantages:

- Simplicity;
- Flexible configuration.

Disadvantages:

- The slowness of the method;
- The accuracy depends linearly on speed.

2) *Image Segmentation*

One of the most effective methods based on segmentation is Graph-Based Image Segmentation [3] and Multiscale Combinatorial Grouping [15, 16].

Advantages:

- Usage of the color and texture features of the image;
- Flexible configuration.

Disadvantages:

- Exclusion of contextual information;
- Excessive segmentation;
- Relative slowness of the method, which depends on the segmentation level.

3) *Hybrid models*

An example of such an approach is the selective search [9], which applies both segmentation and spatial information.

Advantages:

- Usage of the color and texture features of the image;
- Contextual information;
- Flexible configuration.

Disadvantages:

- Relative slowness of the method, which depends on the level of segmentation.

4) *Learning models*

Learning methods include, for example, Objectness [12] and Category Independent Object Proposals [13, 14].

Advantages:

- Usage of the color and texture features of the image;
- Usage of the contextual information.

Disadvantages:

- The need for training, which requires significant amounts of data and time to train model.

Separately, it is possible to determine multi-object recognition models, in which the search process directly related to the recognition process. An example of this approach is the MultiBox Detection used in the SSD [19]. The main disadvantage of this approach is the inability to use the method in other systems.

## V. CONCLUSIONS

The multi-object recognition on the single image is a complex, but the promising direction of object recognition that best suits the real conditions of the application systems. Main problems of the multi-object recognition are the optimization of computational processes that can be realized using parallel computing, as well as the choice of an optimal method for objects search. Among all considered models, methods based on the segmentation [1-4, 9, 15, 16] look very promising, since they have optimal performance and can be used in simple image recognition systems. Another advantage of these methods is the lack of need for a large number of data, that can be critical for simple image recognition systems. On the other hand, integrating the objects search method into the recognition process is a more optimal choice for the implementation of more complex systems.

## REFERENCES

- [1] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour Detection and Hierarchical Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898-916, May 2011.
- [2] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, May 2002.
- [3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," in *International Journal of Computer Vision*, vol. 59, no. 2, 2004, pp. 167-181.
- [4] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, August 2000.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, vol. 1, 2005, pp. 886-893.
- [6] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, September 2010.
- [7] H. Harzallah, F. Jurie, and C. Schmid, "Combining efficient object localization and image classification," 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009, pp. 237-244.
- [8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol.1, 2001, pp. 511-518.
- [9] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," *International Journal of Computer Vision*. vol. 104, no. 2, 2013, pp. 154-171.
- [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,"

- 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus (CVPR), OH, 2014, pp. 580-587.
- [11] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.
- [12] B. Alexe, T. Deselaers and V. Ferrari, "Measuring the Objectness of Image Windows," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2189-2202, November 2012.
- [13] I. Endres and D. Hoiem, "Category independent object proposals," in Proceedings of the 11th European conference on Computer vision: Part V (ECCV'10), Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Springer-Verlag, Berlin, Heidelberg, 2010, pp. 575-588.
- [14] I. Endres and D. Hoiem, "Category-Independent Object Proposals with Diverse Ranking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 2, pp. 222-234, February 2014.
- [15] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, "Multiscale Combinatorial Grouping," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 328-335.
- [16] J. Pont-Tuset, P. Arbeláez, J. T. Barron, F. Marques, and J. Malik, "Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 1, pp. 128-140, January 1 2017.
- [17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.
- [18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," in Computing Research Repository (CoRR), 2018.
- [19] W. Liu et al., "SSD: Single Shot MultiBox Detector," in Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham, 2016, pp. 21-37.
- [20] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 7, pp. 1312-1328, July 2012.
- [21] A. Albayrak and G. Bilgin, "Mitosis detection using convolutional neural network based features," 2016 IEEE 17th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, 2016, pp. 335-340.