- MATHEMATICAL THEORY OF IMAGES AND SIGNALS REPRESENTING, PROCESSING, ANALYSIS, RECOGNITION, AND UNDERSTANDING
- Published: 18 March 2022

Object Detection in Video Surveillance Based on Multiscale Frame Representation and Block Processing by a Convolutional Neural Network

- Rykhard Bohush,
- Guangdi Ma,
- Yang Weichen &
- Sergey Ablameyko

Pattern Recognition and Image Analysis **volume 32**, pages1–10 (2022)Cite this article

Abstract

A method for detecting objects in high-resolution images is proposed that is based on representing an image as a set of its copies of decreasing scale, splitting it into blocks with overlap at each level of the image pyramid except for the top one, detecting objects in the blocks, and analyzing objects at the boundaries of adjacent blocks to merge them. The number of pyramid layers is determined by the size of the image and the input layer of the convolutional neural network (CNN). At all levels except for the top one, a block splitting is performed, and the use of overlap allows one to improve the correct classification of objects, which are divided into fragments and located in adjacent blocks. The decision to merge such fragments is made based on the analysis of the metric of intersection over union and membership in the same class. The proposed approach is evaluated for 4K and 8K images. To carry out experiments, a database is prepared with objects of two classes, person and vehicle, marked in such images. Networks of the You Only Look Once (YOLO) family of the third and fourth versions are used as CNNs. A quantitative

assessment of the detection efficiency of objects is performed using the mAP metric for various combinations of parameters such as the degree of threshold confidence of the CNN and the percentage of intersection of blocks in the hierarchical representation of images. The results of the investigations are presented.

REFERENCES

1. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, arXiv:2004.10934 [cs.CV].
2. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **38**, 142–158 (2016). https://doi.org/10.1109/TPAMI.2015.2437384

Article Google Scholar

3. R. Girshick, "Fast R-CNN," in *IEEE Int. Conf. on Computer Vision (ICCV), Santiago, 2015* (IEEE, 2015), pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169
4. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. (Pearson Education, New York, 2018).

Google Scholar

5. K. Goulekas, *Visual Effects in a Digital World: A Comprehensive Glossary of over 7000 Visual Effects Terms*, (Morgan Kaufmann, San Diego, Calif., 2001).

Google Scholar

6. He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition, in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las*

Vegas, NV, USA, 2016 (IEEE, 2016), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
7. Jiang, X., Hadid, A., Pang, Y., Granger, E., and Feng X., *Deep Learning in Object Detection and Recognition*, (Springer, Singapore, 2019). https://doi.org/10.1007/978-981-10-5152-4
8. P. Korshunov and T. Ebrahimi, "UHD video dataset for evaluation of privacy," in *Sixth Int. Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 2014* (IEEE, 2014), pp. 232–237. https://doi.org/10.1109/QoMEX.2014.6982324
9. LabelImg is a graphical image annotation tool and label object bounding boxes in images. https://github.com/tzutalin/labelImg. Cited December 20, 2020.
10. London, United Kingdom. Virtual travel. 360 video in 8K. https://www.youtube.com/watch?v=KGerjHMa90s. Cited March 16, 2021.
11. New York City 8K - VR 360 Drive. https://www.youtube.com/watch?v=2Lq86MKesG4. Cited March 12, 2021.
12. J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016* (IEEE, 2016), pp. 779–788. https://doi.org/10.1109/CVPR.2016.91
13. J. Redmon and F. Farhadi, "YOLOv3: An incremental," 2018. arXiv:1804.02767 [cs.CV]
14. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell. **36**, 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031

Article **Google Scholar**

15. V. Ruzicka and F. Franchetti, “Fast and accurate object detection in high resolution 4K and 8K video using GPUs,” in *IEEE High Performance Extreme Computing Conf. (HPEC), Waltham, Mass., 2018* (IEEE, 2018), pp. 1–7. https://doi.org/10.1109/HPEC.2018.8547574
16. C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, “Inception-v4, inception-ResNet, and the impact of residual connections on learning,” in *Proc. of the Thirty-First AAAI Conf. on Artificial Intelligence, San Francisco, 2017* (AAAI Press, 2017), pp. 4278–4284.
17. F. Ö. Ünel, B. O. Özkalayci, and C. Çiğla, “The power of tiling for small object detection,” in *IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, Calif., 2019* (IEEE, 2019), pp. 582–591. https://doi.org/10.1109/CVPRW.2019.00084
18. D. Vorobjov, I. Zakharova, R. Bohush, and S. Ablameyko, “An effective object detection algorithm for high resolution video by using convolutional neural network,” in *Advances in Neural Networks–ISNN2018*, Ed. by T. Huang, J. Lv, C. Sun, and A. Tuzikov, Lecture Notes in Computer Science, vol. 10878 (Springer, Cham, 2018), pp. 503–510. https://doi.org/10.1007/978-3-319-92537-0_58

Book **Google Scholar**

19. Walk in Shinjuku, Tokyo, Japan @8K 360° VR / Sep 2020. https://www.youtube.com/watch?v=YYQufxYrBiU. Cited March 14, 2021.
20. J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, “Sun database: Large-scale scene recognition from abbey to zoo,” in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, San*

Francisco, 2010 (IEEE, 2010), pp. 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970
21. YOLO4.weights. https://drive.google.com/u/0/open?id=1cewMfusmPjYWbrnuJRuKhPMwRe_b9PaT. Cited April 5, 2021.
22. L. Yongxi and T. Javidi, "Efficient object detection for high resolution images," in *53rd Ann. Allerton Conf. on Communication, Control, and Computing (Allerton), Monticello, Ill., 2015* (IEEE, 2015), pp. 1091–1098. https://doi.org/10.1109/ALLERTON.2015.7447130

Download references

Author information

Affiliations

1. **Polotsk State University, 211440, Novopolotsk, Republic of Belarus**
   Rykhard Bohush
2. **EarthView Image Inc., 313200, Huzhou, China**
   Guangdi Ma & Yang Weichen
3. **Belarusian State University, 220030, Minsk, Republic of Belarus**
   Sergey Ablameyko
4. **United Institute for Informatics Problems, National Academy of Sciences of Belarus, 220012, Minsk, Republic of Belarus**
   Sergey Ablameyko

Corresponding authors

Correspondence to Rykhard Bohush, Guangdi Ma, Yang Weichen or Sergey Ablameyko.

Ethics declarations

COMPLIANCE WITH ETHICAL STANDARDS

This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the *PRIA* Editorial Board decides not to accept it for publication.

Conflict of Interests

The process of writing and the content of the article do not give grounds for raising the issue of a conflict of interest.

Additional information

Rykhard Bohush. Graduated from the Polotsk State University in 1997. In 2002 he received his PhD in the field of Information Processing at the Institute of Engineering Cybernetics, the National Academy of Sciences of Belarus. Head of Computer Systems and Networks Department of Polotsk State University. His scientific interests include image and video processing, object representation and recognition, intelligent systems and machine learning.

Guangdi Ma. Born in 1985. Graduated from the Chinese Academy of Surveying and Mapping in 2011. Chief Engineer of EarthView Image Inc. Scientific interests: image analysis, photogrammetry, point cloud, and oblique photography aided real 3D reconstruction.

Yang Weichen. Born in 1979. Graduated from the Jilin University, China in 2001. General manager of EarthView Image Inc. Scientific interests: image analysis, photogrammetry, geographical information systems. Pioneered the business service mode of remote sensing target recognition to assist refined social governance in China.

Sergey Ablameyko. Born in 1956, DipMath in 1978, PhD in 1984, DSc in 1990, Prof. in 1992. Professor of Belarusian State University. Scientific interests: image analysis, pattern recognition, digital geometry, knowledge-based systems,

geographical information systems, medical imaging. He is in Editorial Board of *Pattern Recognition and Image Analysis*, *Nonlinear Phenomena in Complex Systems*, and many other international and national journals. He is Fellow of IAPR, Fellow of AAIA, Academician of National Academy of Sciences of Belarus, Academician of the European Academy, and many other academies. He was a First Vice-President of International Association for Pattern Recognition IAPR (2006–2008), Chairman of Belarusian Association for Image Analysis and Recognition.

Cite this article

Bohush, R., Ma, G., Weichen, Y. *et al.* Object Detection in Video Surveillance Based on Multiscale Frame Representation and Block Processing by a Convolutional Neural Network. *Pattern Recognit. Image Anal.* **32,** 1–10 (2022). https://doi.org/10.1134/S1054661822010035

- Received28 October 2021
- Revised28 October 2021
- Accepted28 October 2021
- Published18 March 2022
- Issue DateMarch 2022
- DOI https://doi.org/10.1134/S1054661822010035

Keywords:

high-resolution images

small-size objects

pyramidal representation of images

accuracy of object detection