Journal of Research Proceedings

Under the delegate of "Journal of Research Proceedings," we anchor a bimonthly electronic journal enclosing the diverse realms of the educational research field. JRP is providing a platform for the researchers, academicians, professionals, practitioners, and students to impart and share knowledge in the form of high quality empirical and theoretical research papers, case studies, literature reviews, and book reviews.

JRP Publications
www.i-jrp.com
journalrp.editor@gmail.com
9353189468

Image Segmentation and Object Recognition

Aishwarya S Kumar[1], GuruKiran B L[2], M Saqlaen Ahmed[3], Kiran B V[4], Prof. Geetha C Megharaj[5]
[1,2,3,4] CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India
[5] Associate Professor, CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India

ABSTRACT

The work focuses on object recognition to perform a set of closely related tasks in the field of computer vision, involving object detection and identification. Image classification involves an array of tasks like identifying the class of an object in an image. An object that is identified can also be precisely localized and a bounding box is drawn around it. Object detection combines both localization and identification and classifies one or more objects in image. For Object recognition, a technique called 'Mask Region-based Convolution Neural Networks' or Mask R-CNN is used. It is an extremely efficient approach for object localization and recognition tasks. It is an extension over Faster RCNN method by adding a parallel process for the prediction of a highly accurate object segmentation mask that conforms to the bounds of each detected object in the Region of Interest (RoI) by performing pixel by pixel analysis and classification while still staying highly performant. By using the above methods, individual objects can be identified more precisely than with precise localization. It is also easy to generalize the algorithm to other tasks like estimating the pose of a human in the image or applying a color filter to the image selectively on any object

Keyword: Image Segmentation, Object Recognition, Computer Vision

1. INTRODUCTION

Image processing is a technique used to perform a wide range of computations on an image, in order to enhance the image or extract some useful information out of it [1-3].Digital image processing enables a wide range of algorithms to be applied onto an input image and perform tasks like refining the image of its noise, improve brightness, contrast, saturation and other characteristics of the image. It also enables advanced operations like object detection and localization, human pose estimation, etc. [4-7].

Image processing analyses images the way they are built - in two dimensions of x and y axes, however the algorithms used in processing these images are modelled as multidimensional systems. The image data goes through a series of processes namely pre-processing, enhancement, display and information extraction. It attempts to automate tasks of the human visual system. The process of partitioning an image with the aim to simplify, and change the representation of the image into something which is easier for analysis into multiple segments is known as image segmentation [8]. Image segmentation also involves the process of assigning a label to every pixel in an image. Pixels with the same label share a certain set of characteristics and are said to belong to the same class. Processing an entire image is not a great idea as there will be regions in an image with no useful information and this would cause an additional burden on the processor. Thus, the image is divided into specific segments - called the Regions of Interest - which may potentially contain important features for analysis. This reduces load on the processor while also directing focus of the analysis where needed [9-11].

The technique of building bounding boxes in an image by considering each class in the image is called as object detection and localization.

The shape of an object is not described in object detection. While bounding boxes only localize the objects in an image, a need for marking the exact border of an image arises, for use cases like computational photography or obstacle avoidance. Thus, in image segmentation, a pixel-wise mask is created for each and every object present in an image. It also enables granular understanding of the objects in an image [12-15].

Advancements in these realms have been hugely driven by powerful baseline algorithms utilizing Convolutional Neural Network frameworks for object detection and segmentation. These methods are conceptually intuitive and offer flexibility and robustness, together with fast training and inference time [16].

2. RELATED WORK

A detailed explanation of the papers to explore the existing research, the approaches and the drawbacks identified is given in the following section.

The work discussed in this paper majorly draws inspiration from the paper titled Mask RCNN by Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick presented in the Facebook AI Research, 2018 by implementing the Mask RCNN algorithm using simpler frameworks. The research paper titled ‘Fast R-CNN proposes’ a Fast Region-based Convolutional Network. It improves training and testing times and accuracy in very innovative ways. The paper ‘Towards real-time object detection’ introduces utilization of a Region Proposal Network that works in full synchrony with the network used for object detection. It predicts object extents and prediction scores with a single pass with high-quality region proposals. The ImageNet framework is very effective in generating region proposals and is one of the best contenders when choosing an FCN to identify ROIs and identifying the pixel set [17].

Selective search is a viable option for object recognition when the issue of generating all possible ROIs becomes increasingly exhaustive. It uses image of an image for sampling while also attempting to capture all possible ROIs exhaustively. It generates a set of ROIs, reinforced by the data in the image. In other works of research, it is found that very deep CNNs increase detection accuracy in image recognition [18-19].

Feed-forward networks with a top-down refinement approach provide great analysis of pixel level data as well as ROI-level information with over 15% increase in accuracy in comparison to conventional approaches when a bottom-up/top-down architecture is used.

Instance-aware semantic image segmentation tasks have been recently implemented using Fully Convolutional Networks. An image is fed to the FCN, where it passes through a series of convolutional layers, and generates likelihood score maps for all classes on a pixel-by-pixel level [20].

There are always Speed/accuracy trade-offs while performing object detection tasks using modern convolutional networks. An architecture has to be decided for the CNN that achieves the right balance of speed, memory and accuracy for the use case scenario in question and the hardware available. The Faster R-CNN framework, RFCN and SSD systems are viewed as “meta-architectures”. At a high level, they consist of one CNN containing feature extractors. These CNNs are trained with the objective of instance classification. A sliding window approach is used to analyze each available ROI that has been generated earlier and make the best prediction out of overlapping pixel regions in the collection of the ROIs. Some samples of work in the paper Simultaneous detection and segmentation attempts to detect all instances of a particular class once one instance is found. It has shown promising results with a 7 point boost over the baselines on SDS, a 5 point boost over semantic segmentation, and performance in object detection [21-26].

Multitask Network Cascades are explored for performing instance-aware semantic segmentation, in the paper titled ‘Fully Convolutional instance-aware semantic segmentation’. The model consists of three networks, one each for differentiating instances, estimating masks, and categorizing objects [27-31]. The Multitask networks are designed to share their convolutional features for achieving the segmentation tasks with significantly lesser load on the processing unit [32-37].

3. METHODOLOGY

The computational system is laid out in two steps by design. Figure 1 explains the flow of image-pixel data through the structure of the Mask RCNN Algorithm.

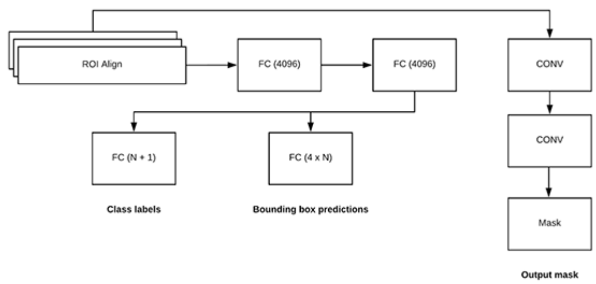

Figure 1: Mask RCNN Data flow Diagram

First stage: A CNN called the region proposal network (RPN) is used to scan the whole Feature Pyramid network. This process generates a set of Regions of Interest and creates anchors for each ROT. An anchor is a bounding box around the extent of a proposed ROI [37-42]. This stage classifies a ROI only as either an object or a part of the background.

Second stage: An algorithm called ROIAlign figures out the relevant ROIs from the ROIpool without the help of anchors and feeds these ROIs to two parallel branches [35]. One of these branches runs the image through multiple Convolution Neural Networks and generates a pixel-level or pixel-group level segmentation mask around the object while the other branch runs it through multiple fully convolution networks reinforced by pre-trained weights to identify the object within the ROI [43-48].

4. IMPLEMENTATION AND RESULTS

The system improves on the pre-existing approach of Faster Region Based Convolution Neural Network, by adding a mask prediction branch towards the end of the bounding box prediction flow as seen in Figure 2. The major Convolution Neural Network (CNN) utilized is the ResNet-50, a CNN that's 50 layers deep and is very commonly used for processing feature rich images.

A Feature Pyramid Network is constructed that utilizes lateral connections within the network from the input to extract features of an object within anRoI. If the ground-truth of anRoI evaluates to a probability of over 0.5, then it is picked for bbox generation, otherwise it is discarded.

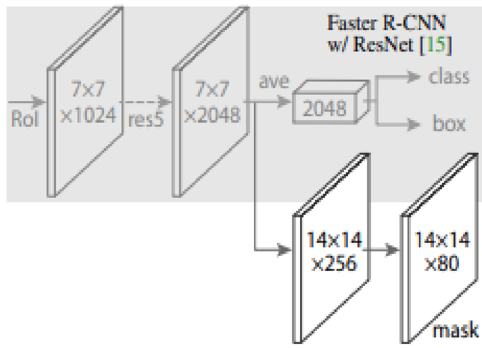

Figure 2: Network Head Implementation of the algorithm

For the branch that generates masks, a smoothening factor has been implemented to vary the load on the processor by tweaking the edge accuracy of the mask as they're found to vary inversely.

Following are the individual modules that are in the prescribed architecture:

Backbone: A CNN is implemented to extract features from the input image starting from the lower level detail like the edges and extents of an object progressing towards more precise detail towards the later layers in the CNN. Three backbone structures are commonly implemented: ResNet101, ResNet50 and MobilNetV1.

Feature Pyramid Network: This part of the architecture allows the detection of small details in the image. It creates a pyramid-like representation of the feature network that has been generated earlier and reinforces the detection of intricate details in a detected object.

Region Proposal Network: The RPN is a neural network that is not as deep as the conventional CNNs. It conducts a progressive scan over the image with a sliding-window approach and enlists all potential Regions of Interest and assigns anchors to each of them. These ROIs overlap and are of varying sizes and aspect ratios to get the most out of the object recognition task that relies on these anchors.

The results achieved with the implementation of this work were satisfactory with accuracy of class prediction averaging slightly less than accuracy of bounding box generation and mask generation.

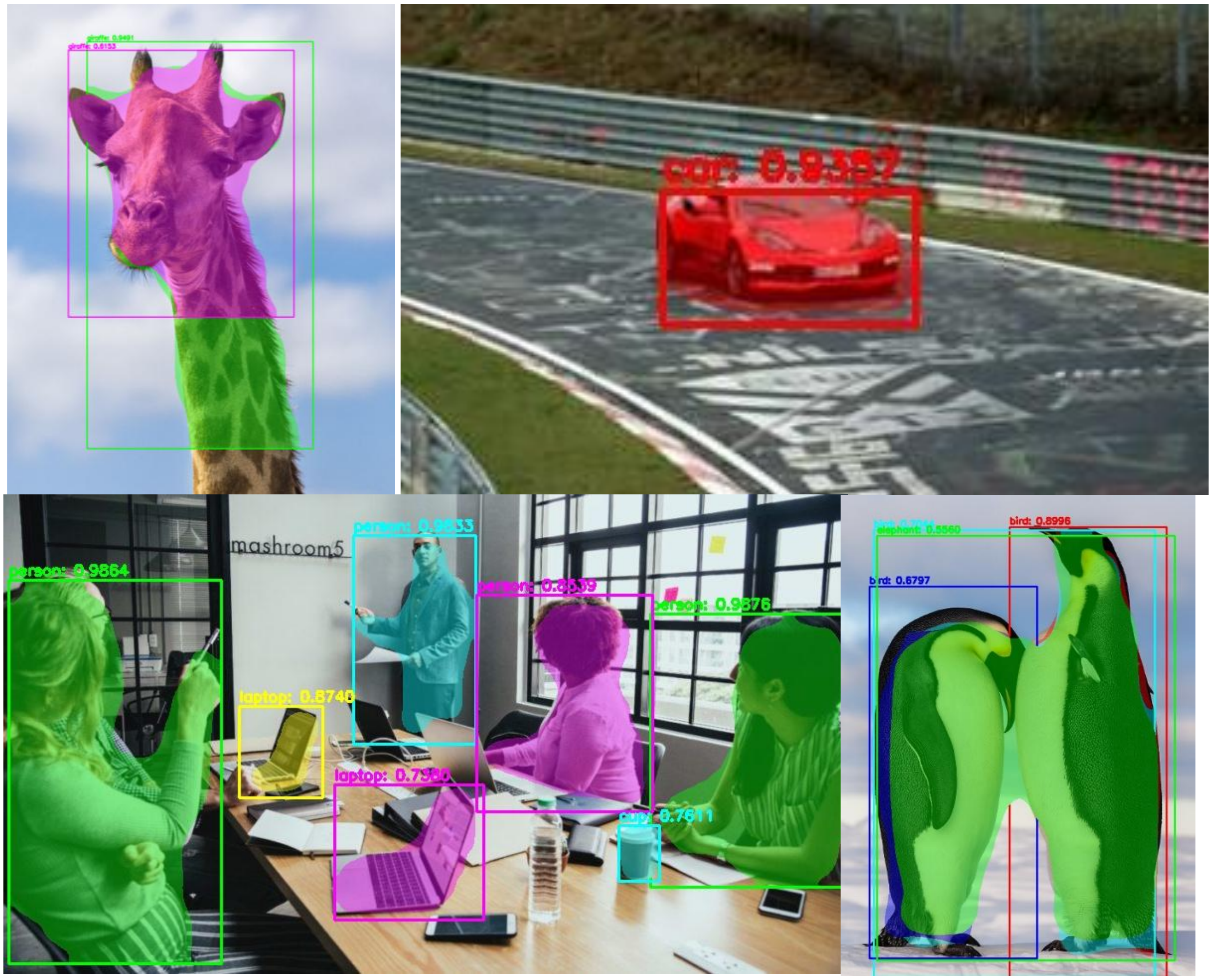

Figure 3: Compilation of results

As seen in the Compilation in Figure 3, the algorithm predicts bounding boxes around each instance of recognized objectin each RoI. Further on, it is found to draw a segmentation mask around the rough extents of each of the identified object instances within their RoI. It is also found that varying the smoothening factor during of mask generation inversely affects the mask edge accuracy. A class label is mentioned on the top left of the bounding box along with its confidence factor ranging between 0 and 1.

Some inferences are found to be erroneous as with the sole nature of prediction algorithms but these have been found to be in less than 5% of the testing instances.

5. ACKNOWLEDGEMENT

We would like to thank Dr. Shantharam Nayak and Prof. Geetha C Megharaj for their valuable suggestion, expert advice and moral support in the process of preparing this paper.

6. REFERENCES

[1] Mask R-CNN. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, FAIR, 2018.

[2] R. Girshick. Fast R-CNN, ICCV, 2015

[3] Hu, Liwen, Ngoc-Tu Nguyen, Wenjin Tao, Ming C. Leu, Xiaoqing Frank Liu, Md Rakib Shahriar, and SM Nahian Al Sunny. "Modeling of cloud-based digital twins for smart manufacturing with MT connect." Procedia manufacturing 26 (2018): 1193-1203.

[4] Dwivedi, R., Dey, S., Chakraborty, C. and Tiwari, S., 2021. Grape disease detection network based on multi-task learning and attention features. IEEE Sensors Journal.

[5] Nguyen, Ngoc-Tu, and Bing-Hong Liu. "The mobile sensor deployment problem and the target coverage problem in mobile wireless sensor networks are NP-hard." IEEE Systems Journal 13, no. 2 (2018): 1312-1315.

[6] Babu, N.V., 2014. Design and performance analysis of clustering mechanisms for wireless sensor networks.

[7] Hemalatha, K. L., S. M. Ashitha, and S. R. Meghana. "Design and implementation of modified FCM in the detection of brain tumor." Int. J. Adv. Sci. Res. Eng 3, no. 4 (2017): 2850-2858.

[8] Bhuvaneswary, N., S. Prabu, S. Karthikeyan, R. Kathirvel, and T. Saraswathi. "Low Power Reversible Parallel and Serial Binary Adder/Subtractor." Further Advances in Internet of Things in Biomedical and Cyber Physical Systems (2021): 151.

[9] Puttamadappa, C., and B. D. Parameshachari. "Demand side management of small scale loads in a smart grid using glow-worm swarm optimization technique." Microprocessors and Microsystems 71 (2019): 102886.

[10] Hemalatha, K. L., SUNILKUMAR MANVI, and HN SURESH. "ADAPTIVE WEIGHTED-COVARIANCE REGULARIZED KERNEL FUZZY C MEANS ALGORITHM FOR MEDICAL IMAGE SEGMENTATION." Journal of Theoretical & Applied Information Technology 95, no. 14 (2017).

[11] Subramani, Prabu, Ganesh Babu Rajendran, Jewel Sengupta, Rocío Pérez de Prado, and Parameshachari Bidare Divakarachari. "A block bi-diagonalization-based pre-coding for indoor multiple-input-multiple-output-visible light communication system." Energies 13, no. 13 (2020): 3466.

[12] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, NIPS, 2015.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and Semantic segmentation, CVPR, 2014.

[14] Vu, D.L., Nguyen, T.K., Nguyen, T.V., Nguyen, T.N., Massacci, F. and Phung, P.H., 2020. HIT4Mal: Hybrid image transformation for malware classification. Transactions on Emerging Telecommunications Technologies, 31(11), p.e3789.

[15] Lingappa, H., Suresh, H. and Manvi, S., 2018. Medical image segmentation based on extreme learning machine algorithm in kernel fuzzy c-means using artificial bee colony method. Int. J. Intell. Eng. Syst, 11, pp.128-136

[16] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, Selective search for object recognition, IJCV, 2013.

[17] Ngo, T.D., Bui, T.T., Pham, T.M., Thai, H.T., Nguyen, G.L. and Nguyen, T.N., 2021. Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration. Journal of Real-Time Image Processing, pp.1-14.

[18] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al., Speed/accuracy trade-offs for modern convolutional object detectors, CVPR, 2017

[19] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik. Simultaneous detection and segmentation, ECCV. 2014

[20] Babu, R.G., Maheswari, K.U., Zarro, C., Parameshachari, B.D. and Ullo, S.L., 2020. Land-Use and Land-Cover Classification Using a Human Group-Based Particle Swarm Optimization Algorithm with an LSTM Classifier on Hybrid Pre-Processing Remote-Sensing Images. Remote Sensing, 12(24), p.4135.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton, ImageNet classification with deep convolutional neural networks, NIPS, 2012

[22] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, Fully convolutional instance-aware semantic segmentation, CVPR, 2014

[23] Subramani, Prabu, Fadi Al-Turjman, Rajagopal Kumar, Anusha Kannan, and Anand Loganthan. "Improving medical communication process using recurrent networks and wearable antenna s11 variation with harmonic suppressions." Personal and Ubiquitous Computing (2021): 1-13.

[24] Parameshachari, B. D., Rashmi P. Kiran, P. Rashmi, M. C. Supriya, Rajashekarappa, and H. T. Panduranga. "Controlled partial image encryption based on LSIC and chaotic map." In ICCSP, pp. 60-63. 2019.

[25]

[26] C. Szegedy, S. Ioffe, and V. Vanhoucke, Inception-v4, inception-resnet and the impact of residual connections on learning, ICLR Workshop, 2016.

[27] Shahriar, Md Rakib, SM Nahian Al Sunny, Xiaoqing Liu, Ming C. Leu, Liwen Hu, and Ngoc-Tu Nguyen. "MTComm based virtualization and integration of physical machine operations with digital-twins in cyber-physical manufacturing cloud." In 2018 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), pp. 46-51. IEEE, 2018.

[28] Prabu, S., Balamurugan, V. and Vengatesan, K., 2019. Design of cognitive image filters for suppression of noise level in medical images. Measurement, 141, pp.296-301.

[29] Parameshachari, B. D., H. T. Panduranga, and S. K. Naveenkumar. "Partial encryption of medical images by dual DNA addition using DNA encoding." In 2017 international conference on recent innovations in signal processing and embedded systems (RISE), pp. 310-314. IEEE, 2017.

[30] Muhammad, L.J., Algehyne, E.A., Usman, S.S., Ahmad, A., Chakraborty, C. and Mohammed, I.A., 2021. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. SN computer science, 2(1), pp.1-13.

[31] Nguyen, Ngoc-Tu, Bing-Hong Liu, Shao-I. Chu, and Hao-Zhe Weng. "Challenges, designs, and performances of a distributed algorithm for minimum-latency of data-aggregation in multi-channel WSNs." IEEE Transactions on Network and Service Management 16, no. 1 (2018): 192-205.

[32] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollar, Learning to refine object segments, ECCV, 2016.

[33] Rajendran, Ganesh B., Uma M. Kumarasamy, Chiara Zarro, Parameshachari B. Divakarachari, and Silvia L. Ullo. "Land-use and land-cover classification using a human group-based particle swarm

optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images." Remote Sensing 12, no. 24 (2020): 4135.

[34] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, Fully convolutional instance-aware semantic segmentation, CVPR, 2017

[35] Do, Dinh-Thuan, Tu Anh Le, Tu N. Nguyen, Xingwang Li, and Khaled M. Rabie. "Joint impacts of imperfect CSI and imperfect SIC in cognitive radio-assisted NOMA-V2X communications." IEEE Access 8 (2020): 128629-128645.

[36] Parameshachari, B. D., H. T. Panduranga, and Silvia liberata Ullo. "Analysis and computation of encryption technique to enhance security of medical images." In IOP Conference Series: Materials Science and Engineering, vol. 925, no. 1, p. 012028. IOP Publishing, 2020.

[37] J. Hosang, R. Benenson, P. Dollar, and B. Schiele, What makes for effective detection proposals? PAMI, 2015

[38] Parameshachari, B.D., Panduranga, H.T. and liberata Ullo, S., 2020, September. Analysis and Computation of Encryption Technique to Enhance Security of Medical Images. In IOP Conference Series: Materials Science and Engineering (Vol. 925, No. 1, p. 012028). IOP Publishing.

[39] Subramani, Prabu, K. Srinivas, R. Sujatha, and B. D. Parameshachari. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." Personal and Ubiquitous Computing (2021): 1-14.

[40] Y. Gong, L. Zhang, R. Liu, K. Yu and G. Srivastava, "Nonlinear MIMO for Industrial Internet of Things in Cyber–Physical Systems," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5533-5541, Aug. 2021, doi: 10.1109/TII.2020.3024631.

[41] C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv and S. Mumtaz, "Attribute-Based Encryption with Parallel Outsourced Decryption for Edge Intelligent IoV," IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 13784-13795, Nov. 2020, doi: 10.1109/TVT.2020.3027568.

[42] S. Chen, L. Zhang, Y. Tang, C. Shen, R. Kumar, K. Yu, U. Tariq, and A. K. Bashir, “Indoor Temperature Monitoring Using Wireless Sensor Networks: A SMAC Application in Smart Cities”, Sustainable Cities and Society, vol. 61, p. 102333, July 2020.

[43] Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, A. Shalaginov, “Deep Graph Neural Network-based Spammer Detection Under the Perspective of Heterogeneous Cyberspace”, Future Generation Computer Systems, https://doi.org/10.1016/j.future.2020.11.028.

[44] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava and P. Chatterjee, “Efficient and Privacy-Preserving Medical Research Support Platform Against COVID-19: A Blockchain-Based Approach”, IEEE Consumer Electronics Magazine, doi: 10.1109/MCE.2020.3035520.

[45] Z. Guo, Y. Shen, A. K. Bashir, M. Imran, N. Kumar, D. Zhang and K. Yu, “Robust Spammer Detection Using Collaborative Neural Network in Internet of Thing Applications”, IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9549-9558, 15 June15, 2021, doi: 10.1109/JIOT.2020.3003802.

[46] L. Tan, N. Shi, K. Yu, M. Aloqaily, Y. Jararweh, “A Blockchain-Empowered Access Control Framework for Smart Devices in Green Internet of Things”, ACM Transactions on Internet Technology, vol. 21, no. 3, pp. 1-20, 2021

[47] Z. Guo, A. K. Bashir, K. Yu, J. C. Lin, Y. Shen, “Graph Embedding-based Intelligent Industrial Decision for Complex Sewage Treatment Processes”, International Journal of Intelligent Systems， 2021, doi: 10.1002/int.22540.

[48] Z. Guo, K. Yu, A. Jolfaei, A. K. Bashir, A. O. Almagrabi, and N. Kumar, “A Fuzzy Detection System for Rumors through Explainable Adaptive Learning”, IEEE Transactions on Fuzzy Systems, doi: 10.1109/TFUZZ.2021.3052109.