Research on Methods of Micro-target Detection in Various Application Scenarios

Bosheng Fang [1], Siyuan Jiao [2], Chengming Wang[3, *] and Qianyu Yin [4]

[1] School of Microelectronics, Hefei University of Technology, Hefei, China

[2] School of Computing, Wuhan University, Hubei, China

[3] College of Electrical Engineering and New Energy, China Three Gorges University ,Yichang, China

[4] School of Instrument Science and Opto-electronics Engineering, Hefei University of Technology, Hefei, China

*Corresponding author: 202290090123@ctgu.edu.cn

Abstract. This paper provides a comprehensive introduction and comparison of the research on small target detection algorithms in different application scenarios and systematically explores the application of deep learning algorithms for small target detection in remote sensing imagery, visible imagery, and infrared imagery. In remote sensing images, the Deconvolution R-CNN algorithm is used to improve the accuracy of small target detection by adding a deconvolution layer. In the field of visible images, a Gaussian Mixture Model (GMM)-based approach is introduced to achieve fast detection of small targets with SIFT features and a GMM target detector. In infrared images, a comprehensive algorithm adapted to low signal-to-noise ratio and complex background is proposed to automatically detect tiny targets through background suppression and threshold selection optimization. In this paper, we compare and analyze the performance of these methods in different scenarios, and propose directions for future research, including expanding the training dataset, applying super-resolution methods to improve the image resolution, and enhancing the universality of the algorithms. This paper provides a comprehensive overview of research in the field of small target detection, as well as an outlook on future research directions, which provides a useful reference for the further development of the field.

Keywords: Small target detection; deep learning algorithm; remote sensing imagery.

1. Introduction

Target detection is one of the main research directions and fundamental tasks in computer vision, which has been widely used in intelligent security, intelligent transportation, intelligent medical care, satellite remote sensing, national defense military, and aerospace. Aerial image dataset contains images taken from different heights, angles, and locations. it is a kind of dataset characterized by scale diversity, high background complexity, and small and dense targets. Related scholars have done much research in recent years, aiming at these characteristics of aerial image datasets. Chinese researchers in the field of computer science for the application of target detection have explored the perfect, but with the development of society and scientific research continues to improve, large target detection is not enough to complete the research of the current national project, so the research from the target detection over to the detection of tiny targets makes us the most important task. The paper will also introduce the embodiment of different tiny target detection application scenarios. With the development of deep learning, the performance of target detection has been greatly improved, but effective detection of tiny targets in images has always been a challenging problem in the field of computer vision. Because the small targets themselves are small in size and contain very few pixels (the size of the tiny targets defined in the AI-TOD dataset [1] is less than 16×16 pixels) and even less target feature information can be extracted after convolution, pooling, and downsampling operations in neural networks. Pooling and downsampling operations can extract even less information about the target features [2].

With the continuous development of science and technology and the continuous improvement of economic strength, as of December 31, 2022, there are 6,718 active satellites in orbit around the world, of which more than 1,100 satellites are used for Earth observation missions, which makes the amount of accessible satellite data grow exponentially, and also provides richer data resources for scientific research. However, due to many factors such as time difference, sensor difference, and space difference, how to obtain effective information from remote sensing images has become a key issue for the development of target detection in the field of remote sensing.

This paper will introduce the corresponding algorithms for tiny target detection in three application scenarios, namely remote sensing images, visible light images, and infrared images, and discuss them in comparison.

2. Small Target Detection Methods for Remote Sensing Images

Remote sensing images are usually derived from the acquisition of ground-based information by high-altitude equipment, in which the size of the captured target is usually characterized by small pixel ratios and high background interference noise. In the target detection of remote sensing images, a two-stage target detection method is commonly used, and Faster R-CNN is a two-stage target detection method with high detection accuracy and speed. Faster R-CNN is developed based on a deep learning neural network (R-CNN), which integrates the extraction of target candidate frames into the deep learning network on top of this algorithm, and is a true end-to-end target detection algorithm [4]. Also realizing quasi-real-time target detection, Faster R-CNN has been widely used in image and video. However, during the last layer of convolution of the algorithm, the feature mapping of the model is very small, and thus the feature mapping of the last layer of convolution of the original image is just as small, resulting in the inability to detect tiny objects with high accuracy. Additive Deconvolutional Neural Network (Deconv R-CNN) was born.

A Deconvolutional Neural Network is an improved neural network architecture based on the Faster R-CNN neural network. The network adds an anti-convolutional layer to the Faster R-CNN neural network, on top of which the anti-convolutional layer can obtain an up-sampled feature map in the image. Then Regional Neural Network (RPN) is used to generate a set of suggestions from the feature mapping of the up-sampled feature maps, and finally, Faster R-CNN is used to regress, classify, and remove repetitiveness from the suggested image targets to get the final detection results [5]. As an example of detecting tiny objects (airplanes) in remote sensing images, the network architecture of Deconv R-CNN is shown in Fig 1.

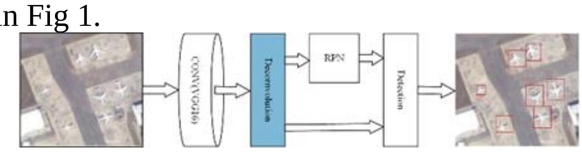

Fig.1 Deconv R-CNN network architecture used for aircraft detection

The detection steps of the Deconvolutional Neural Network are as follows.Step 1 is deconvolution. Deconvolution is a process used to reverse the effect of convolution on the data, and it is also the first and most important process performed by inverse convolutional neural networks in image processing. Deconvolution can expand the size of the image and increase the resolution of the image so that the image can be up sampled to obtain a clearer feature map. To realize this process, it is necessary to calculate each value of the original feature map with the convolution kernel, and then through a series of operations such as summing and increasing padding to get the output feature map, since each pixel feature of the image is represented by a numerical weight in the algorithm[6]. The formula for the deconvolution is as follows.

$$Y[i: i+h, j: j+w] += X[i, j] - K \tag{1}$$

Step 2 is to generate proposals using the Regional Proposal Network (RPN).To be able to match the feature elements in the inverse convolutionally processed image, it is necessary to construct a model of a region proposal network (RPN) architecture that takes as input an image of a specified size. This model is used in the feature mapping after the last convolutional layer. In the RPN network architecture, each pixel of the feature map generates a region candidate box that is fed to the Classification Layer (CLS) and the Regression Layer (REG), and after regression, RPN combines the results obtained from these two neural network layers to generate a set of proposals. The RPN architecture is shown in Fig 2.

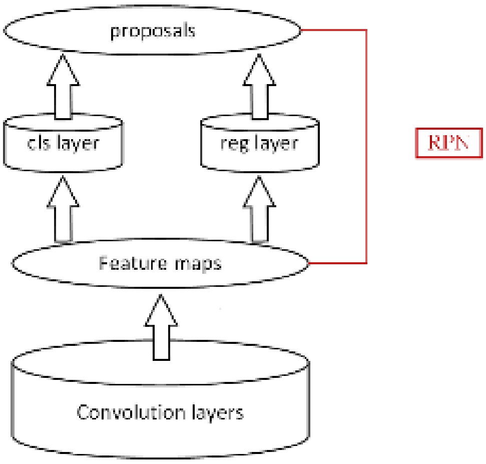

Fig. 2 The RPN architecture

After the RPN neural network generates the proposals, the Faster R-CNN network can be used to provide two output layers for each proposal, the label score and each candidate box output bounding box regression, based on which the location of each proposal and label can then be determined. The object detection network architecture is shown in Fig 3.

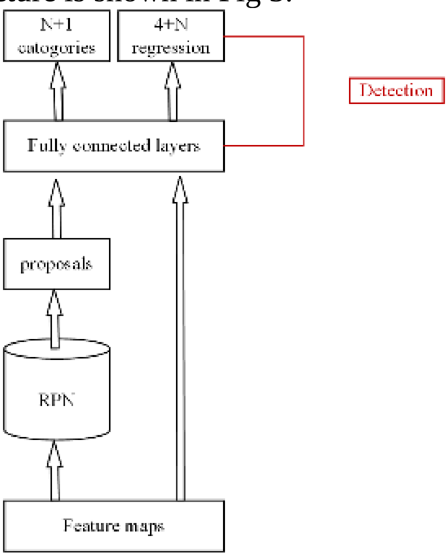

Fig.3 Small Object Detection Network Architecture

On top of the above steps, the RPN can then be trained. Firstly, a binary class label is assigned to each proposal, which in turn leads to the multitasking loss function with the following formula.

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum L_{cls}(p_i, p_i^*) + \lambda\frac{1}{N_{reg}}\sum p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

Immediately after that, the detection results are obtained by combining the gradient descent method with continuous iteration and constant setting of updated weights. The results of different magnification scales and a comparison of algorithms are shown in the following table 1.

Table1. Comparison of results of different algorithms

	mAP	Ship	Plane
Deconv-scale×2	52.0	61.2	42.8
Deconv-scale× 4	54.0	61.8	46.2
Deconv-scale× 8	55.6	62.4	48.7
Faster R-CNN	42.5	50.6	34.3
New method	55.6	62.4	48.7

In addition to the two-stage tiny target detection algorithm based on convolutional neural networks, the improved YOLOv3 algorithm is also capable of accurately detecting tiny targets in images. This approach uses an EIOU Loss function to replace the original target localization loss term and optimizes the regression of the prediction framework, which enables the reduction of the loss of target information and to detection of tiny targets more accurately [7]. In the whole algorithm, to accomplish the task of tiny target detection, the residual block reconstruction operation is first carried out using RS-Darknet53, the structure of this network is shown in Fig. 4

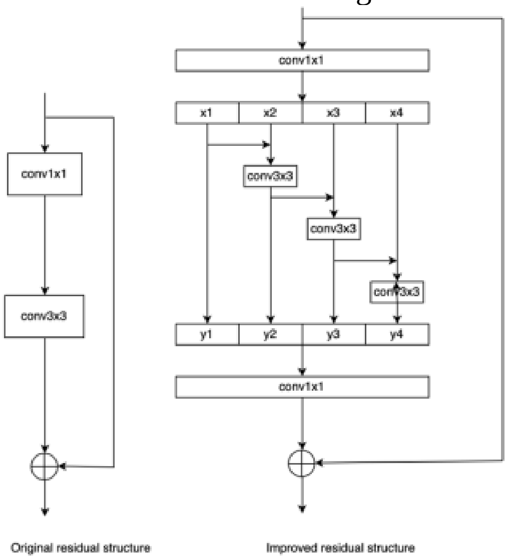

Fig.4 Structure of RS-Darknet53

RS-Darknet53 segments image targets of different sizes by increasing or decreasing the number of residual units after downsampling and by constructing residual blocks, where the design of the residual block size is based on the following equation.

$$y_i = \begin{cases} x_i, i = 1 \\ Conv_{3\times3}(x_i + y_{i-1}), 1 < i < 5 \end{cases} \# \quad (3)$$

After the image segmentation is completed, to be able to realize the target detection based on retaining the spatial details of the image, it is necessary to construct the shallow feature enhancement module (SFEM) to localize the tiny targets in the image. On this basis, the use of void convolution and the attention machine based on the development of the CSMA to capture the information of the small targets in the image and to be able to make the network pay more attention to the key information features of the shallow feature map, to realize the detection of the tiny targets. For the error analysis of the results obtained after training, the EIOU Loss loss function is used. Unlike the original method of calculating the mean square error MSE, the EIOU Loss establishes a correlation between the regression of the bounding box and the overlapping area, aspect ratio, and centroid distance of the two boxes, which can improve the precise target localization by calculating the width and height of the prediction box and the truth box, and thus more accurately train the network is trained quickly. The EIOU Loss loss function is formulated as follows.

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(\hbar, \hbar^{gt})}{c_\hbar^2} \quad (4)$$

The researchers used images of road scenes at night for tiny target detection and the detection results are shown Fig 5[7].

Fig.5 Small target detection for nighttime road scenes [7]

For this image, the researchers used different algorithms to compare the same image for the experiment and the results are shown in the table 2.

Table 2. Comparison target category results of different algorithms

Method	Backbone	Target Category						mAP@0.5
		Pedestrian	Bicycle	Car	Van	Truck	Motor	
Faster R-CNN	ResNet-101	20.9	7.3	51.0	29.7	19.5	21.2	21.8
CDNet	ResNeXt-101	35.6	13.8	55.8	42.1	38.2	29.3	34.2
MSA-YOLO	CSPDarknet	33.4	11.2	76.8	41.6	41.4	31.0	34.7
DMNet	ResNet-50	28.5	15.9	56.8	37.9	30.1	29.2	30.3
CenterNet	Hourglass-104	22.6	14.6	59.7	24.0	21.3	23.7	26.2
DBAI-Net	ResNeXt-101	36.7	14.7	47.4	38.0	41.4	16.6	28.0
YOLOv3-LITE	Darknet-53	34.5	7.9	70.8	31.3	21.9	32.7	28.5
New Method	Darknet-53	54.5	19.4	82	46.7	38.3	49.6	36.7

3. Small Target Detection Methods for Visible Non-remote Sensing Images

In a typical visible light image, colors are made up of three colors, commonly referred to as RGB. Algorithms for detecting relevant tiny targets in remotely sensed RGB images have been described above, and some more advanced and stable algorithms are available for non-remotely sensed images as well.

3.1. Gaussian Mixture Model Based Small Target Detection Method

Among RGB images, image matching can be done by finding SIFT (Scale Invariant Feature Transform) features, but the computational cost of extracting SIFT features in the overall image is very high and the number of template images to be compared is very high, making the method very time consuming. To solve this problem, the researchers constructed a small target detection method based on Gaussian Mixture Model (GMM). The GMM hybrid model is able to use the SIFT feature descriptors as its feature vectors while estimating the initial configuration parameters of the model to generate the binary image, and then using maximum likelihood estimation for parameter estimation with iterative updating of the configuration parameters to find the SIFT features more quickly [8]. The training flow of the GMM is shown in Fig. 6.

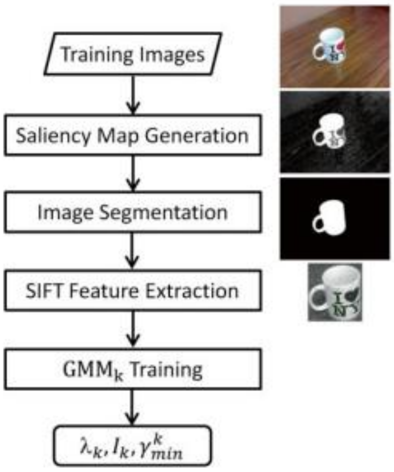

Fig 6. Training flow of GMM [8]

In the whole detection process, the researcher extracts several image chunks from the RGB-D camera image, extracts SIFT features from them, and then sends these extracted features to the Gaussian Mixture Model (GMM) target detector for target recognition, and ultimately, the systematic training of the GMM target detector is used to more accurately detect tiny targets in the image.

3.2. Tiny Target Detection Methods for Fuzzy Visible Non-remote Sensing Images

Highly blurred images are characterized by the difficulty of obtaining and identifying small targets in the image. In order to realize the acquisition of small targets in blurred images, the experimenter used a convolutional neural network to realize the detection of small targets [9]. Throughout the detection process, the researcher first suppresses the background of the blurred image, thus facilitating the input of the processed image into the convolutional neural network. Then it goes through the convolutional layer and pooling layer of the convolutional neural network to get the output results to achieve the detection.

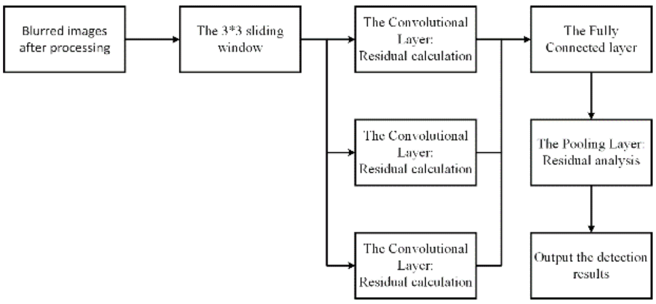

Fig 7. Process of tiny target detection by convolutional neural network

4. Tiny Target Detection in Infrared Images

An infrared image is formed by a thermal infrared scanner that receives and records the thermal radiation emitted by a target [10]. Compared to visible light images, infrared images are characterized by poor resolution, low contrast, low signal-to-noise ratio, and blurred visual effects. Generally, people use simple long-distance detection methods to detect tiny targets, but there is still the problem that the signal-to-noise ratio (SNR) gradually decreases leading to a decrease in the ability to distinguish objects for recognition, so a more complex algorithm is needed to solve the problem. Researchers have proposed an algorithm that can operate at high detection rates and low signal-to-noise ratios to automatically detect tiny objects in three-dimensional infrared images, as shown in fig 8[11]. In this algorithm, the first task is to enable a generalized thresholding algorithm using a background suppression method to preprocess the image to improve the differentiation of low signal-to-noise objects from the background. In this process, an open top-line transform is used to suppress the interference of the image environment around the measured target. Immediately after that, the researcher used an averaging filter of size 21 × 21 as a pass filter to make the horizon and object size in the image more salient compared to the fixed-line transform. After obtaining a clear image with the background suppressed by the calculation, it is possible to detect the target using the thresholding method. The researcher used the ptile method to calculate the general threshold level and then update the threshold value weighted according to the coefficients and reflect the strong and weak flares in the image by the updated threshold value [11]. The formula for updating the threshold and judging the strong and weak flares is as follows.

$$th_1 = \alpha * th, \qquad th_2 = \beta th \tag{5}$$

$$\begin{aligned} I_1(x,y) &= 1, S(x,y) \geq th_1 \\ I_1(x,y) &= 0, S(x,y) \leq th_1 \\ I_2(x,y) &= 1, S(x,y) \geq th_2 \\ I_2(x,y) &= 0, S(x,y) \leq th_2 \end{aligned} \tag{6}$$

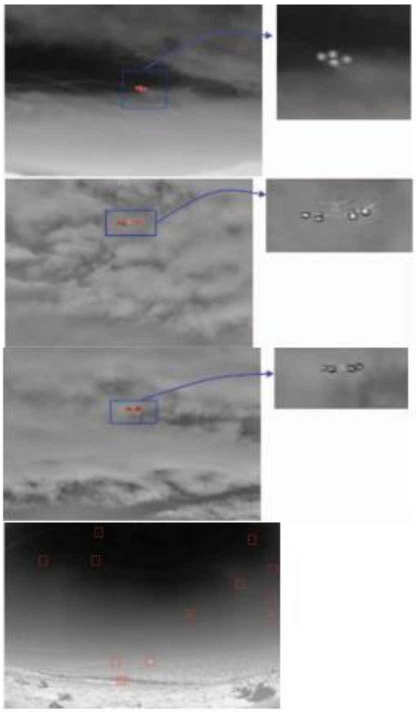

Fig 8. Test results of the model [11]

5. Results and Discussion

In the traditional methods of micro-target detection in infrared images, the target signal is enhanced and the background noise is suppressed by calculating the contrast and saliency graphs of the images. However, this approach faces some challenges, such as the real-time and robustness of the algorithm needs to be further enhanced, especially in dynamic environments. In addition, complex backgrounds can interfere with the effect of background suppression, which requires more efficient background modeling techniques. To solve these problems, researchers have developed some novel methods. For example, the universal threshold method combined with explicit top-hat transformation and background suppression can effectively improve detection performance. At the same time, the adaptability of the algorithm to different scenes can be enhanced by using disc structure elements and adaptive filters.

In the field of tiny object detection in remote sensing images, researchers propose some optimized SSD algorithms, which enhance feature extraction and multi-scale fusion by combining DenseNet and FPN networks. In addition, the image is analyzed by using a graph neural network (GNN), the graph structure with the node as a detection unit is constructed, and the edge graph is created for training. However, this approach still has room for improvement, such as the need for more efficient network structures to handle small target detection in high-density and complex backgrounds. At the same time, it is also crucial to improve the accuracy of detection and recall rate under the condition of low signal-to-noise ratio. To address these challenges, DenseNet and FPN are introduced to enhance feature extraction capabilities, while GNN is applied to process graph structure data, providing a new solution strategy for complex scenarios.

In the detection of small objects in visible images, the YOLOv3 algorithm is used for end-to-end detection, while improving the prediction and anchor frame mechanism. In addition, the genetic algorithm and fractal feature are combined to optimize threshold selection. Still, real-time performance can suffer when dealing with a large number of prediction boxes, and the algorithm also needs to be improved to accommodate tiny targets of different scales and shapes. To solve these problems, researchers propose some innovative methods, such as introducing the BP layer and anchor box to improve prediction accuracy and combining genetic algorithms and fractal features to optimize the threshold selection process. In summary, multi-scale feature fusion, graph structure learning, adaptive and intelligent algorithms, and end-to-end learning are some of the key innovations in the field of micro-target detection, which together improve the accuracy and efficiency of detection, simplify the detection process, and improve the detection speed and accuracy.

6. Conclusion

This paper presents the current state of research on small target detection through typical algorithms in different application fields. Firstly, it introduces the classification of small target detection in the field of remote sensing images, with a focus on the process of the Deconvolutional Neural Network (Deconv R-CNN) method. Secondly, it introduces the small target detection algorithm based on the Gaussian Mixture Model (GMM) in the visible light image field. Then, it elaborates on the methods to improve the accuracy of small target detection results in the infrared image application field. Finally, it compares and discusses the algorithms in each application field.

Although the small target detection algorithm has improved in recent years, there are still many problems to be solved. At present, there are few small target detection algorithms, and there is a lack of a large amount of theoretical and algorithmic references. In most real-world scenarios, small targets are easily clustered or obscured by other objects, leading to a large loss of information, which in turn reduces the accuracy of detection results. Small targets are prone to large positioning errors due to their size. Most small targets in real-world scenarios contain less information and are difficult to extract features.

In order to improve the accuracy and efficiency of small target detection algorithms, this paper points out some future research directions. The training volume of existing small target detection algorithms is insufficient. In the future, the small target dataset can be expanded to increase the training volume of small target detection algorithms, thereby improving the accuracy and efficiency of the algorithm. Through various super-resolution methods, the resolution of small targets is further improved, thereby increasing the features of small targets and improving the accuracy of detection. Most of the existing small target detection algorithms are aimed at special scenarios, and more attention should be paid to the general adaptability of small target detection algorithms in future research.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] YU Ye, AI Hua, HE Xiaojun, et al. Attention-based feature pyramid networks for ship detection of optical remote sensing image[J].Journal of Remote Sensing, 2020, 24(02):107- 115.

[2] LIU L, OUYANG W, WANG X, et al. Deep learning for generic object detection: a survey[J]. International Journal of Computer Vision, 2020, 128: 261-318.

[3] LIU Y, SUN P, WERGELES N, et al. A survey and performance evaluation of deep learning methods for small object detection[J]. Expert Systems with Applications, 2021, 172: 114602.

[4] Du,Peng,Chen,Ming,Su,Tonghua.Deep Learning and Target Detection.Publishing House of Electronics and Industry,Beijing,pp:8-9,2020.

[5] W. Zhang, S. Wang, S. Thachan, J. Chen and Y. Qian, "Deconv R-CNN for Small Object Detection on Remote Sensing Images," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain,pp. 2483-2486,2018.

[6] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning[J]. arXiv preprint arXiv:1603.07285, 2016.

[7] Sun,C.,Li,J and Xu,G. "Improved YOLOv3 algorithm for small object detection in remote sensing images," 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, pp. 1752-1756, 2023

[8] Qian,X, Liu,X and Ye,C, "A Gaussian-mixture-model-based visual feature matching scheme for small-object detection from RGB-D data," 2017 IEEE International Conference on Real-time Computing and Robotics (RCAR), Okinawa, Japan, pp. 91-96,2017.

[9] Liu,Jinhua, Convolutional neural network based tiny target detection method for fuzzy images, Information recording material,2022

[10] Gerald,C.,Holst.Testing and Evaluation of Infrared Imaging Systems.National Defense and Industry Press,pp:25-26,2023.

[11] M. Özbay and M. C. Şahingil, "A fast and robust automatic object detection algorithm to detect small objects in infrared images," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, pp. 1-4,2017.