Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **3**, May **2018**

Computational Resource Constrained Deep Learning Based Target Recognition from Visible Optical Images

Jia Zhai[1,2], Hao-guang Zhao[3], Qiang Ji[3] and Xiao-dan Xie[4]

[1]Communication University of China, Beijing, 100024
[2]Science and Technology on Electromagnetic Scattering Laboratory, Beijing, 100854
[3]AVIC Shenyang Aircraft Design and Research Institute, Shenyang, 110035
[4]Science and Technology on Optical Radiation Laboratory, Beijing, 100854
zhaijia_bj@163.com; Laserradar@126.com; heujq@sina.com; 13810812099@139.com

Received September, 2017; revised December, 2017

Abstract. *The current deep machine learning systems often adopt high performance computer or computing platforms to enhance accuracy of optical object image recognition, but this is not possible for Airplane or other mobile platform based object recognition systems. In order to meet urgent needs of Airplane platform based target recognition systems, this paper proposes a computational resource constrained deep learning based optical target image recognition method, and presents an architecture of collaborative learning algorithm of high performance computers and embedded systems, and then a physical simulation system is built up to verify the performance. The experimental results show that high accuracy and efficiency of the proposed Airplane-based target recognition approach can be achieved. The proposed approach can be used for the Airplane, satellites-based and other mobile platforms-based target recognition systems.*
Keywords: Visible optical image; Target recognition; Deep learning; Computational Resource Constrained

1. **Introduction.** The resolution and quality of visible optical images get higher with the development of optical sensors. Meanwhile, the application needs of object recognition based on the visible optical images become more and more urgent. Besides, the recognition process is applicable to different mobile platforms, such as drones, manned aircrafts, airships and so on, and this enables object recognition to play an increasingly important role in military and civilian fields. For example, in the modern complex battlefield environment in which targets, background and response from sensors are fused together, the obtaining, transmission, handling and application of information exist throughout the whole time of war, and thus it is crucial to obtain information effectively, accurately and timely, especially for the decision-making of war. With the rapid development of modern electronic information technology, material technology, aviation technology and many other technologies, airplane-based platform has been equipped with multiple sensors for real-time monitoring of the battlefield dynamic information, and henceforth, the processing and analysis of the original data become an important factor in the final result of the war.

At the present stage, for the object recognition on airplane platforms, the usual approach is adopting the template matching method to identify the targets with the assistance of man. There are three problems in this traditional method:

- The recognition accuracy rate depends on the accuracy of feature selection and feature modeling. Good recognition results can only be achieved from high robustness features by extracting the characteristics of goods. However features are often affected by illumination variation, background change and different shooting angles, and this leads to heavy dependence on work experience.
- The method is not effective for targets without some clear characteristics, or multiple targets which are needed to identify from massive data, and still relies on human interpretation.
- The method needs a large number of target feature templates to have a good accuracy which will affect the processing speed.

In order to achieve fast and efficient autonomous recognition on aircrafts, some ideas need to be borrowed from the fields of big data, machine learning and artificial intelligence. Deep learning[1] makes it possible to real-time recognize target and mine data based on data model without features, and it is a way of modeling patterns which is a statistical probability model based on large amount of data in the field of machine learning. In addition, after automatic modeling of various patterns (such as sound, images, etc.), it can identify these patterns with no need of manual extraction of features.

Object recognition [1][2] is one of the most remarkable fields where deep learning can be applied to. As deep learning technology becomes more and more prominent in accuracy, robustness and other aspects, its network structure is also constantly improving. Based on CNN?Convolutional Neural Network?, various network structures suitable for different needs arises, such as R-CNN (Region Based CNN)[3], Fast R-CNN (Fast Region Based CNN)[4] and Faster R-CNN (Faster Region Based CNN) [5]. In 2015, He et al.[6] proposed a 152 layers residual network (Residential Network, ResNet) to recognize ImageNet data set, and the recognition error rate was reduced to below 3.57%, which was beyond the human eye recognition ability.

In China, deep learning is only applied to image and voice recognition in the commercial area, while in the United States and some other western countries, artificial intelligence has been combined with aircraft (especially unmanned aircraft) technology, which has started a new round of military revolution. On December 14, 2015, the United States Secretary of Defense clearly put forward that the autonomous learning systems, which were closely related to the intelligent unmanned aircraft technology, were one of the five key technologies supporting the U.S. military technology in its top position in the world. For reconnaissance and attack, the US DARPA has made great efforts to apply the latest technology of artificial intelligence to the rapid and accurate object recognition of complex platforms such as unmanned aerial vehicles (UAVs).

In summary, in order to realize the transition from traditional aircraft to the autonomous intelligent ones, and to solve the traditional airplane object recognition problems, a research on intelligent sensing technology based on deep learning for airplane platform needs to be carried out to achieve fast and efficient autonomous on-board recognition. The deep learning recognition method in the battlefield has prominent advantages as well as challenges. For one thing, the battlefield environment is complex and changeable, in which occlusion and interference are normal, and therefore the general deep learning object recognition method should be optimized and adapted to realize high speed real time autonomous recognition in order to meet the real-time needs in the high dynamic and strong jamming battlefield environment. For another, since intelligent awareness for airplane platform relies on computing resources, application research on intelligent sensing device needs to be carried out to solve the problem caused by the limitation of airplane

hardware capacity, such as localization, miniaturization, low power consumption and poor environmental adaptability, and to make use of these devices for actual combat.

This paper intends to apply deep learning technology to intelligent airplane sensing and awareness area. It proposes a deep learning method of target recognition based on visible optical image using resource-limited device, and a deep learning algorithm framework based on the collaboration of high performance computer and embedded systems. And the method is equipped with a physical simulation system to verify the performance. The experimental results show that this method can effectively improve the accuracy and speed of airplane object recognition and it can be applied to the recognition of objects carried on LEO satellite as well as other mobile platforms besides airplane.

2. **Overall structure.**

In order to solve the traditional Airplane object recognition problem and to realize fast and efficient implementation of autonomous recognition, it is necessary to combine the characteristics of feature generation technology based on photoelectric mechanism and intelligent awareness of deep learning. Power consumption of the airplane carried platform also needs to be minimized with careful design, and this can be realized with an on-board device for aircraft. Another work is the big data service platform at the backend. The overall architecture is shown in figure 1.

In general, the big data service application platform consists of hardware layer, data layer, algorithm layer and application layer. And the layers are organically integrated and support each other. The storage server hardware layer provides HDFS storage management services for the data layer. The training server provides computing resources for algorithms training. The recognition server provides services for the application layer by realizing the recognition and the collection of the recognition rate statistics function.

The data layer, with the measured data and feature generation data in it, provides the algorithm layer with a large number of training samples. The algorithm layer automatically mines key target features and the output of the model will be used by the recognition layer.

The statistics of accuracy rate is calculated in the application layer, and network model will be improved after using these statistics to form a more accurate network model. And the new images and videos input can also be used as effective supplementary for the data layer.

The front end intelligent object recognition device uses the Storm based distributed architecture to integrate ARM type CPU and multi-GPU. It can load deep learning network models to achieve object recognition from video stream and develops a small-size intelligent object recognition device with low power consumption.

3. **Computational resource constrained object recognition algorithm.**

3.1. **Work flow of object recognition based on deep learning.**

The architecture of deep-learning-based target recognition is shown in Figure 1, where the algorithm layer and the front-end smart object recognition device are the focuses.

The algorithm layer uses deep learning algorithm to realize the whole process of object recognition, which is divided into two parts, offline training part and online real-time recognition part. The overall process of target recognition is shown in Fig.2. First, it constructs the training sample image database in the offline training. The training data, including the measured data and the generated feature data set, are the input of the deep learning network. Then, the trained deep learning model with better recognition accuracy can be achieved by adjusting parameters of the deep learning network. In the online real-time recognition part, the CPU and multi-GPU embedded recognition system

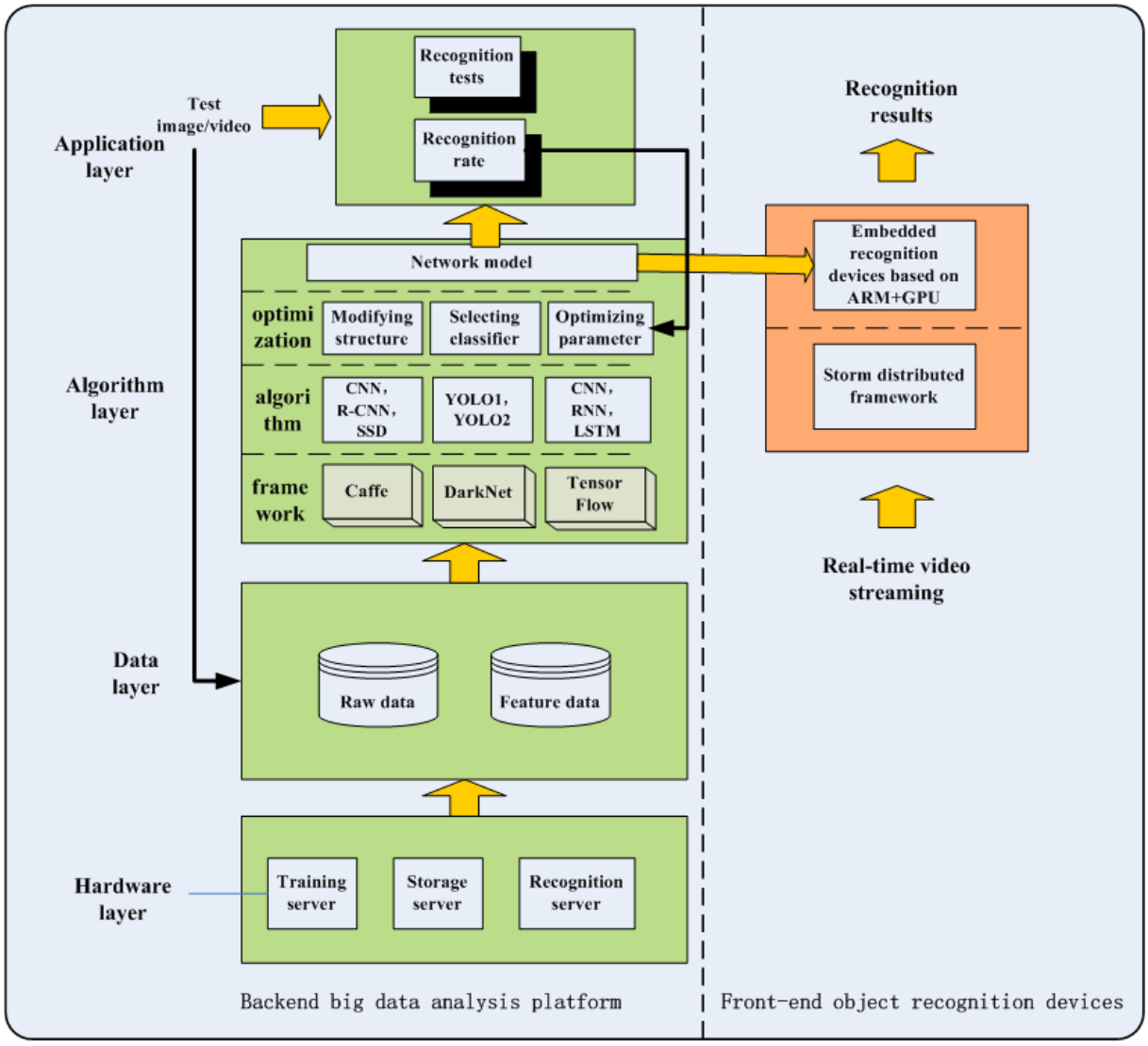

FIGURE 1. Architecture of deep-learning-based target recognition using high performance computing platform and embedded computing platform

is the key part. The ARM type CPU is used here to reduce power consumption. The preprocessing video capture is transferred to the trained deep learning model in the front-end intelligent object recognition device to realize real-time target recognition of the video. Every GPU uses the deep network model to detect objects in parallel and CPU merges results together according to time sequence. Then the final recognition results can be displayed and restored.

The overall architecture of the front-end smart object recognition device is shown in Figure 3, which consists of an ARM CPU and multiple Jetson TX1 GPUs. The storm stream processing framework is deployed on the hardware structure to improve the efficiency of the distributed computing. The video data can be collected and processed without relying on a remote server, and all functions can be completed locally. The platform can guarantee low power consumption and real-time object detection, and it also has good extensibility.

To summarize, the intelligent target awareness algorithm based on deep learning, and miniaturization low power consumption design of airplane platform equipment for the front-end intelligent object recognition is our two key technologies.

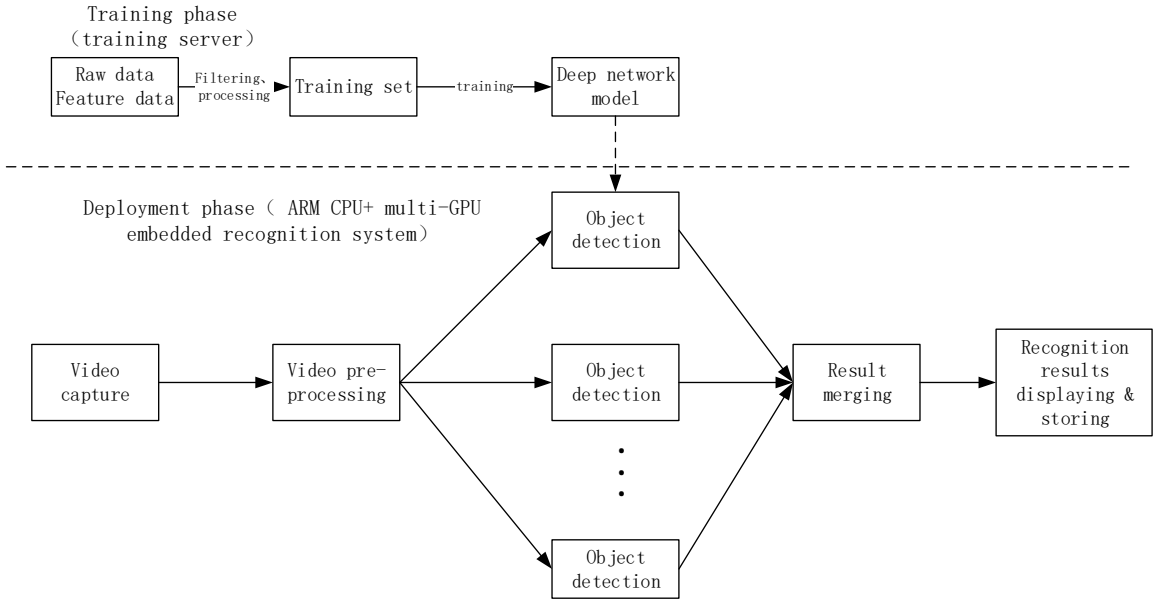

FIGURE 2. The overall process of target recognition

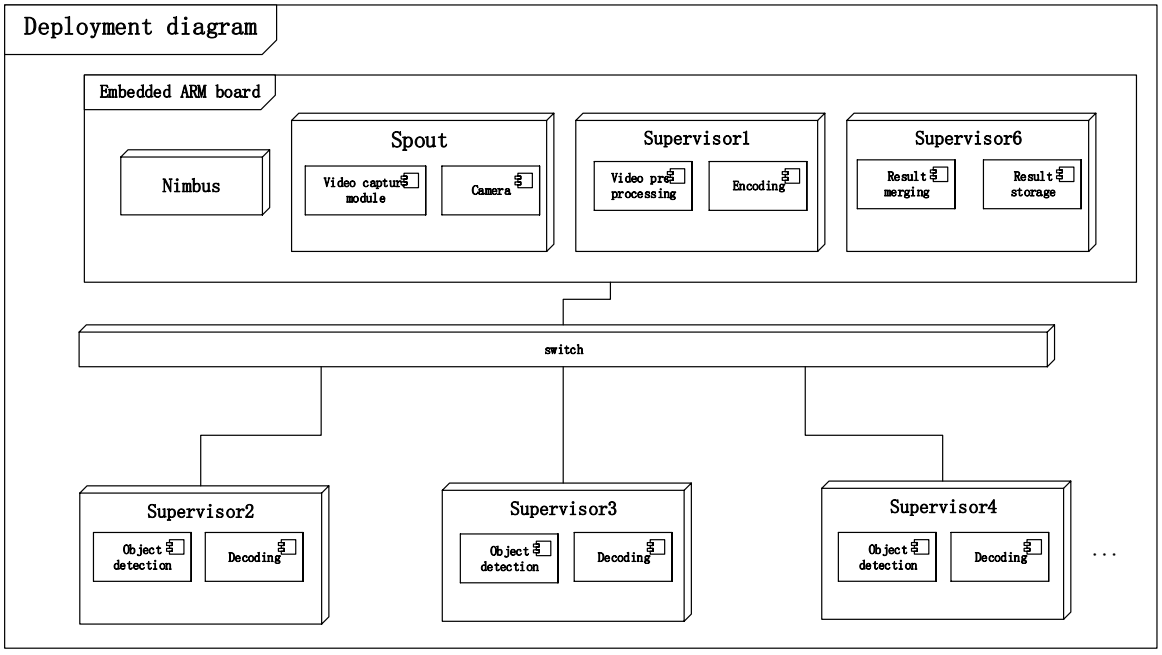

FIGURE 3. The intelligent target recognition front-end application architecture

3.2. **Object recognition algorithm based on SSD network.** Identifying video objects in each frame which often contains multiple targets to be detected is needed for airplane autonomous recognition. Therefore, the SSD (Single Shot Multi-Box Detector) method is adopted to improve the recognition capability of multiple objects detection while guaranteeing the performance. SSD produces a series of fixed-sized bounding boxes

and the possibility of containing object instances in each box based on a forward propagation convolutional neural network. SSD obtains the rectangle region in the image, its type and the type score corresponding to that region. The use of features of images in various scales of each location to conduct regression, which can ensure both speed and accuracy, is its superiority.

Firstly, the target image database is constructed, and then the training data set including target detection features are built by using the target image database. Secondly, the training data set is input to the SSD network for training, and then the SSD network parameters are continuously adjusted so that it can have better recognition accuracy. While the ratios of length and width of SSD final candidate frames are variable, the usual ratios of width and length of the targets are roughly the same. Therefore, this paper aims to improve the recognition frame and optimize the network by making its proportion in the range of prior knowledge so as to be more suitable for the detection of specific target.

The SSD training process consists of four steps, which is divided into two stages, as shown in figure 4:

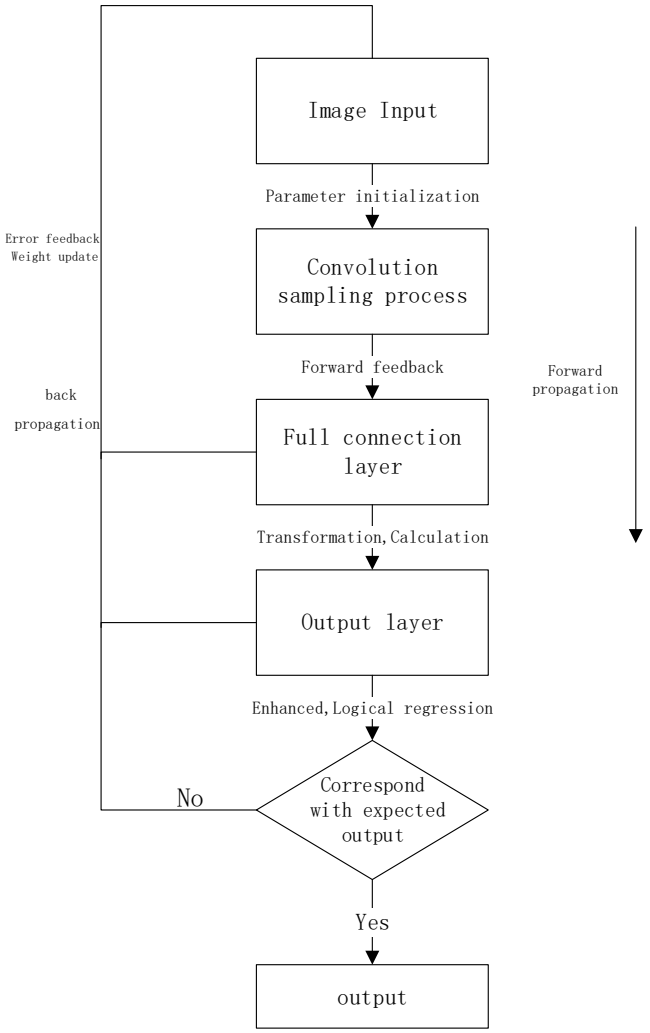

FIGURE 4. The training process of SSD network model

The first stage is the forward propagation process. It consists of two main steps: Step 1, take a sample from the sample set and type it into the network; Step 2, calculate the

corresponding actual output. At this stage, the input information is transformed layer by layer to the final output layer. The calculation process is actually performed by the network input and weight matrix multiplication in each layer, plusing some deviation, and finally getting the output.

The second stage is the reverse propagation process, which also consists of two main steps, i.e., calculating the difference between the actual output and the expected output and adjusting the weight matrix according to the method of minimizing error.

The training process of network includes forward propagation and back propagation. Forward propagation is mainly used for feature extraction and classification calculation. The back propagation is the inverse feedback of error and the updating calculation of weights. After the image input, neurons of all layers are initialized. Convolution and sampling are used to extract and map image features, where multiple convolution and sampling procedures can be employed. The multi-layers extraction process can extract useful information from the image. After the feature extraction is completed, the extracted features are fed forward to the full connection layer which contains multiple hidden layers. Then the data is transformed and calculated by the hidden layers. Finally, the result is fed back to the output layer and will be output if the test results are consistent with the expected results.

If the test results and expected results do not meet the need, the weights and biases will be reversely propagated to the training network. They will be transferred from the output layer backward to the fully connected layer and convolution sampling layer until every layer gets their gradient. Then, the weights are updated to start a new round of training until the optimal neural network is obtained.

3.3. **A computational resource constrained deep learning object recognition system.** In order to realize object recognition by using deep learning for airplane, an online processing model based on parallel computing with embedded GPUs is adopted to establish an application system with low power consumption. High performance deep neural network with large scale is achieved by high power consumption cost, but the computing power of mobile devices is limited by weight and battery capacity. Although mobile cloud computing helps to transfer a portion of the computation to the back-end cloud server, the bandwidth, latency and availability will be severely tested when addressing high traffic and real-time streaming data. Therefore, it is necessary to design an embedded distributed platform for efficient calculation of deep learning networks.

Deep learning hardware accelerators require data level and work flow parallelism, multithreading and high memory bandwidth. A GPU can execute more instructions in each instruction cycle. Therefore, GPU is more suitable for large scale matrix convolution operation in deep learning than CPU. NVIDIA dominates the current deep learning market with its massively parallel GPU and dedicated GPU programming framework, CUDA. As is shown in Figure 6, NVIDIA Jetson TX1 is an embedded GPU designed for deep learning. It is based on the NVIDIA Maxwell^{TM} architecture which has 256 CUDA cores, which provide more than 1 trillion performance floating-point operations, and 64 ARM CPU. The board integrates CPU and GPU functions into one unit and provides high floating point throughput with extremely low power consumption (10W).

Due to the limited computing power of TX1, a single block TX1 cannot meet the performance requirements for complex deep neural networks. Therefore, a parallel embedded framework is designed to achieve high real-time processing.

The whole framework takes Storm as the underlying infrastructure and uses Nimbus management cluster to realize cluster fault tolerance and scalability. The video capture

module captures real-time video which is decoded after adjusting the size and other operations through the video processing module in ARM CPU which can reduce the workload of GPUs. The decoded frame transits through the network to the image processing module to process each frame. The processing results are sent to the regular and unified storage module for data information integration. By using the self-designed Storm scheduling strategy, a specific bolt runs on a specific device to meet the overall needs.

4. **Experimental results and analysis.** After using the collected data to train different sizes of deep learning networks off-line, the accuracy and speed of deep learning networks can be compared. The input is HD real-time video (25frames/s), using ARM CPU as the main control unit, Jetson TX1 GPU as the processing unit. With the increasing number of Jetson TX1 GPU modules, the video processing speed is improved. The system is designed through the comprehensive measure of network scale, processing speed and integration costs. The optimized deep learning network model is deployed on the platform. Next, the platform is tested by firstly recording the standby state of the platform and then starting a Jetson TX1 object detection and recording the speed and power consumption on the platform with increasing number of Jetson TX1 one by one.

As shown in table 1, when only one Jetson TX1 is used for object detection, the speed is 11fps and the power consumption is 17W. With the number of boards increasing, the speed and power consumption also increase. When the number of Boards reaches three, the speed reaches 34fps and the power consumption is 46W. This proves that the platform can well meet real-time requirements while it maintains low power consumption and good scalability and thus can be applied to airplane platforms. From the speed power consumption ratio, one can see that with the number of TX1 increasing, the energy efficiency is on the rise while the rise rate is gradually reduced. Consequently, this paper argues that three or four boards can reach the optimal solution for speed and power consumption for the real-time video processing for airplane platforms.

Table 1. The relationship between the number of TX1 and the energy efficiency ratio of the system

TX1	0	1	2	3	4	5	6	7	8
speed(fps)	0	11	23	34	44	54	66	79	90
Power (W)	8	17	27	36	45	53	64	76	83
Power/speed ration(fps/W)	-	0.647	0.852	0.944	0.978	1.019	1.031	1.039	1.084

For comparison with the proposed platform, the same object detection algorithm is also tested to run on the GTX TITAN X GPU. To achieve the same speed (40fps), the GTX TITAN X GPU requires the power to be 90W and the energy efficiency is 0.450fps/W. One can see that the power needs to be twice as much as that is needed by the proposed platform while the energy efficiency is only half. This comparison shows that the proposed platform in this paper has high efficiency with low power consumption and thus is highly recommended to be applied to the airplane platform for real-time video processing.

5. **Conclusion.** Considering the limitation of computation resources of the embedded platform, this paper tries to resolve some key problems of optimizing the deep learning algorithm for such kind of devices by proposing an architecture for deep learning using stream processing and embedded GPUs, which effectively improves the accuracy and robustness of object recognition for airplane platforms. The miniaturization and autonomy

design, good scalability and low power design make the proposed approach suitable for airplane intelligent recognition device. The work can also be used to realize real-time detection of sea targets and self-recognition, and improve the intelligent level of future airplane combat platform.

REFERENCES

[1] Y. Lecun, Y. Bengio, and G. Hinton, Deep Learning, *Nature*, vol.521, pp.436–444, 2015.

[2] Y. Lecun, K. Kavukcuoglu, and C. Farabet, Convolutional Networks and Applications in Vision, *IEEE International Symposium on Circuits & Systems*, vol.14, pp.253–256, 2010.

[3] C. Szegedy, A. Toshev, and D. Erhan, Deep Neural Networks for object detection, *Advances in Neural Information Processing Systemse*, vol.26, pp.2553–2561, 2013.

[4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol.86, pp.2278–2324, 1998.

[5] 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*, vol.39, pp.1137–1149, 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition, *Computer Vision and Pattern Recognition*, pp.770–778, 2016.