

A Multi-Scale Image Detection Method for Power Line Inspection

Nan Zhang, Fu Zhao and Jinglin Guo*

Beijing Guodiantong Network Technology Co., Ltd. Chuangye Middle Road, Haidian District, Beijing 100085, China

*Corresponding author

Abstract—A multi-scale image detection method for power line inspection is provided by combining deep learning with traditional image algorithm. The method of Gauss blur and differential pyramid is used to enhance the data set. Experiments show that the method can effectively improve the efficiency of line inspection. Using FPGA to realize the Gauss blur and difference pyramid, using GPU to realize the Yolo V3, and using heterogeneous computing mode, a set of multi-scale image detection equipment is designed.

Keywords—power line inspection, image processing, deep learning, data augmentation

I. INTRODUCTION

As a means of delivery, the helicopter carries detection personnel, infrared thermal imager, digital camera and other equipment to detect and maintain the transmission line to replace the traditional manual way. The problem of traditional manual detection can be classified as visible light detection. In the process of line inspection, the speed of helicopter is not low. How to locate the interest point of image quickly is an engineering challenge. Furthermore, the large scale change of detection content brings difficulty to image detection. For example, the integrity of tower base is a large-scale problem, and the problem of conductor broken strand is also a small-scale image problem. How to accurately identify these problems in line patrol is also a challenge. Therefore, to some extent, there is a trade off between accurate recognition and fast recognition.

Aiming at the first problem, this paper adopts the heterogeneous computing mode based on PCI-E to speed up the image computing speed. Aiming at the second problem, this paper uses the method of enhancing image data set to improve the accuracy of image recognition.

II. FRAMEWORK DESIGN

Image processing is a fast developing subject. The image processing involves video frame capture, feature recognition and so on. With the development of deep learning technology, CNN (convolution neural network), RNN (recurrent neural network) and DNN (deep neural network) have brought the recognition of image features to the new dimension [1,2,3].

Image processing involves three kinds of equipment: 1) airborne or hand held devices; 2) field device processing (disaster relief command vehicle); 3) data center equipment processing. Among them, airborne or hand held devices

process video or static images; field device processing usually includes image initial processing, recognition based on training models and real-time application needs (for example, Panoramic Map of disaster scene). The data center equipment can cover all kinds of image application requirements.

In the implementation of airborne equipment, the heterogeneous computing mode is adopted, the main FPGA accelerates the traditional image algorithm, and the h.265 decoding card is used to improve the speed of video coding and decoding, and GPU is used to realize the image recognition based on deep learning. Furthermore, in the airborne equipment of the project, the tiny Yolo framework is adopted[4].

A. PCI-E Communication Module

PCI-E is a high-speed serial point-to-point dual channel high bandwidth transmission. Its main advantage is high data transmission rate. The highest 16X 2 version now achieves 10GB/s. In fact, in general applications, high speed Ethernet technology can be used in data communication, and the advantage of using PCI-E is that the real time is guaranteed.

B. FPGA Module

FPGA is more efficient than CPU even GPU, mainly due to its architecture without instruction and no memory sharing. There are two functions of memory in the von Neumann structure. One is the save state, and the two is the inter cell communication. Because memory is shared, we need to do access arbitration. In order to make use of the locality of access, every executive cell has a private cache, which means we need to maintain consistency between components. For the needs of saving state, registers and on chip memory (BRAM) in FPGA belong to their own control logic, without unnecessary arbitration and caching. For communication needs, the connection between each logical unit of FPGA and surrounding logic units has been determined during reprogramming, and no communication is required through shared memory. In this work, is the selection of Xilinx Virtex-6 (Xilinx).

C. Performance Optimization

In order to improve the performance, in the case of engineering application precision, this paper made three changes during the implementation process.

The decode image is completed by using the H.265 decoding card.

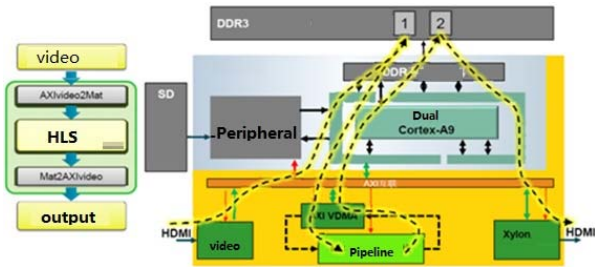


FIGURE 1. IMAGE DECODING PROCESS

A drop sampling for an image is to get a point every few lines and columns to form a new image. The scale factor of 2 down sampling: an image is for every row to take a bit. Because FPGA is pipelined operation. After original image input, there is no register preservation. If we want to achieve the down sampling, we need very precise timing control, which increases the overall complexity of control. Therefore, this paper stripped the process from the classical SIFT algorithm and provided the data with the H.265 decoding card. The synthesis control is implemented at the back end of the algorithm process.

Use the variable length Gauss template to replace the fixed template in the classic software.

In general, the SIFT realization uses Gauss's fuzzy semigroup properties and transforms the two-dimensional Gauss matrix into a one-dimensional matrix. But in FPGA implementation, due to the error effect caused by the fixed template, for the sake of concurrent consideration, the general way is to enlarge the template size, for example, the length of one dimension template is 9.

The so-called Gauss semigroup property is that in the classical SIFT algorithm, every layer of the Gauss operation depends on the result of the last Gauss operation, which can reduce the computation cost. However, in FPGA implementation, this method restricts the capability of parallel computing. By using the variable length template, Gauss Pyramid can be obtained by multiplying the original and variable length templates without using this property. The corresponding template lengths for each layer were 1, 3, 5, 7, 9, 11, 13. The input image and the corresponding template are multiplied to get the data of each layer directly, and the calculation dependence is avoided. The length of the template is 1 in fact representing that the layer does not do Gauss blur. The template value of each layer is fixed in advance and is placed in FPGA in advance. Through simulation test, it is found that the variable length template does not affect the accuracy, but it can improve the parallel ability at the same time.

The polynomial function is used to approximate the exponential function.

It takes more time to call the exponent operation, so the polynomial method is used to approach.

For example, in the weight calculation:

$$\text{Obin} = \text{grad_ori} * \text{bins_per_rad};$$

Weighting

$$W = \exp(-(- (c_rot * c_rot + r_rot * r_rot) / \exp_d);$$

Adjust to

$$W = r * r / (r * r + c_rot * c_rot + r_rot * r_rot);$$

This adjustment affects only about 3% of the feature points. But the difficulty of the project is reduced and the speed of the algorithm is improved.

Application scene

The airborne or hand-held devices output high-definition video streams through the H.265 coding module.

The scene of the disaster: the video stream is processed by the field equipment, the decoding module completes the frame extraction at high speed, and the FPGA completes the image splicing and the feature recognition task.

Data center: the frame extraction is completed at high speed and parallel by the decoding module, which can generate 256 1024*768 resolution images at the same time, and the sampling frequency is 30 frames per second. The image is passed through the PCI-E bus to the GPU based deep learning framework to complete the training or recognition. The PCI-E bus solves the problem of large throughput.

III. DATA AUGMENTATION

In general, deep neural network needs a lot of training data to get better results. In the case of limited data, data augmentation can be used to increase the diversity of training samples, improve the robustness of the model and avoid over fitting. In computer vision, typical data augmentation methods include flip, rotate, scale, random crop or pad, etc[5].

In this paper, the method of using Gaussian blur and difference pyramid to generate image is used to enhance the data, and good results are achieved. This work uses part of the engineering implementation of SIFT algorithm and has a high degree of reuse.



FIGURE II. GAUSSIAN BLUR AND DOG($\Delta=1.5$; $\Delta=2$)

In the absence of other methods, Gaussian blur and differential pyramid can effectively improve the image recognition accuracy. For large-scale targets, with Tiny Yolo, the recognition accuracy increases by about 10 percentage points (without data augmentation, the recognition accuracy is about 76%).



IV. SUMMARY

In this work, the recognition effect of large-scale interest points is better than that of small-scale interest points. In addition, how other data augmentation algorithms fit with Gaussian blur is also our work direction. In the next work, we plan to implement Yolo V3 on FPGA.

REFERENCES

- [1] Hinton, G. E., Osindero, S. and Teh, Y., A fast learning algorithm for deep belief nets. *Neural Computation* 18:1527-1554, 2006
- [2] Yoshua Bengio, Pascal Lamblin, Dan Popovici and Hugo Larochelle, Greedy Layer-Wise Training of Deep Networks, in J. Platt et al. (Eds), *Advances in Neural Information Processing Systems 19 (NIPS 2006)*, pp. 153-160, MIT Press, 2007
- [3] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra and Yann LeCun Efficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), *Advances in Neural Information Processing Systems (NIPS 2006)*, MIT Press, 2007
- [4] YOLO9000: Better, Faster, Stronger, *Computer Vision and Pattern Recognition (cs.CV)*, arXiv:1612.08242 [cs.CV]
- [5] Cui, Xiaodong, Goel, Vaibhava, Kingsbury, Brian, Data augmentation for deep convolutional neural network acoustic modeling- *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015