Multi-objective Identification of UAV Based on Deep Residual Network

Wang JiaQi\textsuperscript{1,1}, Dai JiYang\textsuperscript{1,2}, Zhai JinYou\textsuperscript{1,3} and Ying Jin\textsuperscript{1,4}

\textsuperscript{1}School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China

1196006635@qq.com\textsuperscript{1}, djiyang@163.com\textsuperscript{2}, 710644534@qq.com\textsuperscript{3} and azx519@foxmail.com\textsuperscript{4}

Abstract. Because of the classical RPN (Region Proposal Net) exists the defect of large computation and high time complexity when extracting targets candidate region, a search mode called CRPN (Cascade Region Proposal Network) mode was proposed to ameliorate it in this paper. In order to suppress the degradation phenomenon in deep convolutional neural network training, the residual learning theory was introduced, a novel Mu-ResNet (multi-strapdown deep residual network) was proposed. Combined the Mu-ResNet with CPRN, a network model for multi-target identification of UAV was designed and tested. Compared with the network model that combines ResNet with RPN, the identification accuracy was increased nearly 2 percentage points.

1. Introduction
The convolutional neural network is a branch of the artificial neural network. It has been successfully applied in computer vision, pattern recognition and other fields. The advantages of the convolutional neural network are obvious. Firstly, it processes input datas through convolution operations and down sampling, which and lower the computational complexity. Secondly, its network weights are shared with each other, which can effectively reduce unnecessary weight connections and lower model complexity. Thirdly, the network structure has high degree of scale invariance, rotation invariance and other affine invariance, so the learned features have topological correspondence and robustness.

Imagenet Large Scale Visual Recognition Challenge (ILSVRC) competitions at recent years, the top three participants are both used neural networks algorithms, such as literature [1]. This phenomenon revealed the strong vitality of CNN and its generalization capabilities. Deep convolutional neural networks [2,3] are CNN with deeper level networks, have led a series of breakthroughs for image classification [4,5]. Deep CNN naturally integrate low/mid/advanced functions and categories at end-to-end pattern [6], and the ‘levels’ of functions can be enriched by stacking more layers (depth). Recent evidences [7,8] reveals that network depth is crucial importance, and the most advanced results of image classification [9,10,11] on the challenging ImageNet dataset [12] all exploited very ‘deep’ [7] models.

However, with the deepening of the number of network layers, data amount and computation amount are also increasing dramatically which caused the notorious problem of vanishing/exploding gradients\[13\]. Not only hamper convergence from the beginning but also exposed degradation problem. That means, with the network depth increasing result accuracy gets saturated and then degrades rapidly [14].
2. Deep Residual Network Construction

The residual learning module is a part of the deep convolutional neural network and it can effectively solve the problem of accuracy degradation. Residual learning does not map the image features successively to the next layer, but directly maps the features to the residual layer[14]. Training deep neural networks is very difficult, however, using the deep residual module to build the network can greatly reduce the amount of training data and achieve deeper network training. The two-layers and three-layers residual learning module learning mechanism was shown in figure 1 and figure 2 respectively.

In this paper, a Strapdown ResNet with multiple shortcut channels was designed, which combines the advantages of the two-layers and three-layers residual modules. The Strapdown ResNet uses a five-layer stack, its structure and connect model was shown in figure 3. By setting multiple shortcuts to ease the training difficulty and makes it is easier to converge in deep network training. And using dimensionality reduction and convolution to reduced the time complexity of training.

Aiming at the multi-class target images with arbitrary size, UAV target identification system was combined with CNN and residual learning model. The five layers residual learning module was used to build Multi-Strapdown ResNet with different depths.

3. Target Identification Algorithm

RBG et al. proposed the Regional Proposal Network (RPN) [15] and gives three kinds of area \{128^2, 256^2, 512^2\} and three scales \{1:1, 1:2, 2:1\} which constitute total 9 types of anchors to extract detection areas, as shown in figure 4. Anchor mechanism gives the size of a standard sliding window, then different size windows can obtained according to different multiples and length-to-width ratios.

Because of the large number of search areas in the classical RPN network, the computational complexity and time complexity were high when target candidate areas were extracted. In order to solve this problem, this paper used a three-level CRPN search mode. In the first level CNN, a 32x32 convolution kernel were used, sliding window with 4 step size on the WxH size feature map to get detection window. Due to small search area scale, the first-level CRPN can filter out more than 90% regional windows. Then use non-maximum suppression to eliminate candidate areas with high coincidence rate. This method greatly reduces the computational complexity and time complexity of lower level networks. Next, the target candidate areas obtained through the first level network was adjusted to 64x64 size and input to second-level CRPN to filter out 90% candidate region windows, then the remaining windows passes through the NMS to eliminate the candidate region of high coincidence rate. Finally, the candidate region of the second level CRPN’s output was adjusted to 128x128. After the same operation, the background area and the height overlap area are filtered out.
and obtained 300 rectangle candidates in the end. The final candidate region classification layer and window regression layer are implemented with a convolutional layer of 1x1 size. The CRPN model was shown in figure 5.

4. Target Identification Experiment
In this paper, the image set for UAV target identification mainly includes 10 kinds of targets, such as aircraft, pedestrians, cars, large buses, ships and flying birds. Those are common targets when UAV execution tasks. Data sources mainly come from ImageNet, major open source databases and our own aerial images taken by drones, total 100,000 images. Among them, 70,000 images were used as the training set, and 30,000 images were used as the test set.

In order to study the effect of feature extraction network with different performance parameters on identification results, this paper constructs three kinds of network with different depth for performance
test. That are, Mu-ResNet-22, Mu-ResNet-37, Mu-ResNet-57, represent 22/37/57 layers depth residual network separately.

For conduct horizontal comparison, the above three networks are compared with VGG16, ZF and classical residual networks, and the CRPN is combined for performance comparison. Firstly, training on the target training set to obtain each network model, then testing the accuracy of target identification and detecting time consumed for each picture on the test set. The specific results are shown in table 1. From table 1, we can see that the identification accuracy of Mu-ResNet-57 is slightly higher than ResNet-56, and the detection time is less than ResNet-56. The overall performance of Mu-ResNet-22 and Mu-ResNet-37 is also better than VGG16 and ZF networks. Therefore, using the 5-layers residual learning module proposed in this paper to build the network model can improve the performance of the algorithm more effectively.

Table 2 shows the detection curvature of each type target in the UAV data set by combining VGG16, ResNet-56 and Mu-resnet-57 with RPN and CRPN respectively. The proposal parameters of RPN set 2000 can get best effect, according to [15]. The proposal parameters of CRPN in this paper takes 1000. From Table 2, we can see that Mu-ResNet is better than VGG16 and ResNet when classifying airplane, human, automobile and truck. And it’s worse than ResNet in classification of ship, bird and horse. The overall performance of CRPN were better than RPN. Through above experiments, considering detection time of each image and detection accuracy, the 57-layer Mu-ResNet combined with CRPN constructed by a 5-layer residual learning module can achieve 90.4 identification rate on the data set and detect each image only needs 0.093s, so this network model is more suitable for UAV target recognition.

Figure 6 shows the test results based on the Mu-ResNet-57 combined with CRPN. The model has good effects on the identification of targets and has good adaptability that it’s not influenced by objects size.

| Table 1. Comparison of different networks experiment results. |
|-----------------|-----------------|-----------------|
| net             | times(s)        | accuracy        |
| ZF              | 0.035           | 79.2%           |
| VGG16           | 0.116           | 86.6%           |
| ResNet-56       | 0.105           | 89.1%           |
| Mu-ResNet-22    | 0.076           | 87.8%           |
| Mu-ResNet-37    | 0.885           | 88.2%           |
| Mu-ResNet-57    | 0.093           | 90.4%           |

| Table 2. Results of different network models for each category of data set. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| net             | VGG16 +RPN     | VGG16 +CRPN    | ResNet-56 +RPN | ResNet-56 +CRPN | Mu-ResNet-57 +RPN | Mu-ResNet-57 +CRPN |
|                 | mAP(%)         |                 | mAP(%)         |                 | mAP(%)         |                 |
| airplane        | 92.13%         | 92.36%         | 94.21%         | 91.895          | 93.03%         | 95.07%         |
| automobile      | 90.27%         | 90.55%         | 93.20%         | 89.38%          | 90.06%         | 94.08%         |
| bird            | 89.35%         | 90.10%         | 92.05%         | 89.14%          | 88.31%         | 89.30%         |
| balloon         | 94.78%         | 96.48%         | 95.11%         | 93.75%          | 95.17%         | 96.14%         |
| cat             | 86.33%         | 85.31%         | 86.31%         | 87.86%          | 84.93%         | 90.11%         |
| dog             | 93.27%         | 93.17%         | 95.07%         | 93.15%          | 95.02%         | 96.22%         |
| human           | 90.30%         | 93.09%         | 94.30%         | 91.81%          | 93.34%         | 94.87%         |
| horse           | 87.92%         | 88.16%         | 91.82%         | 83.09%          | 86.79%         | 90.60%         |
| ship            | 90.12%         | 91.07%         | 91.74%         | 88.92%          | 90.88%         | 93.17%         |
| truck           | 82.06%         | 85.30%         | 86.26%         | 88.48%          | 90.12%         | 88.40%         |
Figure 6. Target identification effect diagram.
5. Conclusion

Aiming at the real-time and accuracy requirements of UAV target recognition, based on convolutional neural network, a variety of frameworks of deep residual network models are built for training. Through compared performance of the network, we ultimately find most suitable network for UAV target recognition. Compared with the traditional target recognition algorithm, target recognition algorithm based on depth residual network can avoid artificial error when target feature extraction and greatly improved recognition accuracy. For the situation of targets in complex background the deep residuals network’s shortcut feedback mechanism also can effectively reduce the difficulty of network training and enable deeper networks can be effectively trained and used. The target detection using the deep residual network is applied to the UAV target recognition and the comprehensive performance of the accuracy and detection time is superior to the traditional deep convolutional neural network.

However, the target data sets studied in this paper are less. In future research, we need to acquire more and more drone targets data sets, build deeper networks to get better performance and use more efficient feature extraction methods to improve targets detection accuracy while lower detection time.

References

[1] Chang, L., Duarte, M. M., Sucar, L. E., and Morales, E. F. 2012. A bayesian approach for object classification based on clusters of sift local features. Expert Systems with Applications An International Journal, 39(2), 1679-1686.

[2] Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., and Hubbard, W., et al. 2014. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4), 541-551.

[3] Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems (Vol.60, pp.1097-1105). Curran Associates Inc.

[4] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and Lecun, Y. 2013. Overfeat: integrated recognition, localization and detection using convolutional networks. Eprint Arxiv.

[5] Zeiler, M. D., & Fergus, R. 2014. Visualizing and Understanding Convolutional Networks. European Conference on Computer Vision (Vol.8689, pp.818-833). Springer, Cham.

[6] Srivastava, R. K., Greff, K., and Schmidhuber, J. 2015. Highway networks. Computer Science.

[7] Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. Computer Science.

[8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., and Anguelov, D., et al. 2014. Going deeper with convolutions. 1-9.

[9] Saxe, A. M., McClelland, J. L., and Ganguli, S. 2013. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. Computer Science.

[10] He, K., Zhang, X., Ren, S., and Sun, J. 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. 1026-1034.

[11] Ioffe, S., Szegedy, C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. 448-456.

[12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., and Ma, S., et al. 2014. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252.

[13] X. Glorot, Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In AISTATS, 2010.

[14] He, K., Zhang, X., Ren, S., and Sun, J. 2015. Deep residual learning for image recognition. 770-778.

[15] Ren, S., He, K., Girshick, R., and Sun, J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. International Conference on Neural Information Processing Systems (Vol.39, pp.91-99). MIT Press.