Fast American Sign Language Image Recognition Using CNNs with Fine-tuning

Qi Cui[1], Zhili Zhou[1,2], Chengsheng Yuan[1], Xingming Sun[1], Q. M. Jonathan Wu[2]

[1] School of Computer and Software and Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, China

[2] Department of Electrical and Computer Engineering, University of Windsor, Canada

cuiqiloveslife@gmail.com, {zhou_zhili, ycs_nuist, sunnudt}@163.com, jwu@uwindsor.ca

Abstract

Sign language images such as American Sign Language (ASL) images carry effective information for communication based on gestures. The significance of sign language recognition is that it can help hearing impaired people to communicate with others successfully. To facilitate the ASL images recognition for the hearing impaired people, a proper image classification is demanded urgently. As the computation of traditional image classification methods is usually extensive, it's a bad choice for real-time tasks. Due to the extraordinary performance of Convolutional Neural Networks (CNNs) on various tasks of images classification, a method is proposed to use CNNs to train classification models for ASL images recognition with fine-tuning strategy. First, a structure of CNNs is optimized to make it more suitable for ASL classification task. Secondly, the models are trained on the structure by using test images. We compare the test results of proposed approach with those of the state-of-the-arts at the end with the aim to illustrate the effectiveness of the trained CNNs models. The experimental results demonstrate that the proposed method can achieve superior recognition results for ASL images.

Keywords: Image recognition, Convolutional neural networks, Fine-tuning strategy

1 Introduction

Image recognition is an important kind of pattern recognition task. There are many advanced methods employed in image processing field based on improved conventional image features [1-2]. However, most of these methods are not suitable for real-time recognition task owing to the large computational efforts and high impact from environments [3-4]. For large-scale image recognition, many methods based on Convolutional Neural Networks (CNNs) [5] have been proved. In recent years, many image recognition tasks showed the extraordinary performances such as VGG [8], GoogleNet [9] and ResNet [10]. They made great contributions to ImageNet [12] Large-Scale Visual Recognition Challenge (ILSVRC) with approaches based on DCNNs. With the rapid development of DCNNs, many advances have been made in object detection based on deep learning, such as Faster R-CNN [7] and SSD [16]. At the same time, these methods have been successfully applied in various research fields. Meng et al. [30] used the target detection method based on deep learning to achieve information hiding in local regions. As an important member in the field of machine learning, extreme learning such a feedforward neural network can effectively solve the classification problem [6]. Compared with traditional methods [13-14] in the field of image features exaction, the approaches based on CNNs could handle the task better when facing the special cases. In the era of processing big data, intelligent retrieval based on deep leaning has been widely used in various services. Li and Li [32] proposed the image retrieval method using DCNN via efficient hashing code and addictive latent semantic layer. Xia et al. [24, 31] proposed intelligent privacy security retrieval approaches. The researches of DCNNs in the field of image forensics proves its flexibility in feature selection. Cui et al. [11] used DCNNs to extract high-frequency features of images to classify natural image and computer generated images. In this paper, an improved CNNs architecture will be introduced to extract features of images which can be applied for American Sign Language (ASL) image classification. ASL is a kind of gesture language and a method of spelling which contains 24 static gestures and 2 motion gestures. The fine-tuning strategies based on CNNs are used to train a CNNs model fast and efficiently, which is proved efficiently to train a model of CNNs by a pattern of resuming training on ASL image set.

The contributions of this paper include: (1) A CNNs architecture of training models is optimized with fine-tuning strategy on ASL image set; (2) We analyze the

*Corresponding Author: Xingming Sun; E-mail: sunnudt@163.com
DOI: 10.3966/160792642018121907023

mismatch problems about networks for training between the CaffeNet [15] and ASL image data set. We solve the problem by adjusting the network.

2 Related work

In the fine-tuning strategy operation, a CNNs model from a large-scale image set is required as the initial weights of the network at first. Next, this CNNs model is trained for extracting image features by convolutional calculation with different convolutional kernels and then be activated by specific methods. The activated values usually will be put in next intermediate layers. Finally, by the calculation of fully-connected layers, the multi-dimension output vector can be regarded as image features. The features that are exacted by CNNs have been proved to be efficient in other image tasks such as image retrieval and image copy detection tasks.

The CNNs' parameters are tuned for small-scale data set S. The model required is pre-trained on a large-scale data set L. In training process of S, the model trained by the network will be modified to make it suitable for S by decreasing the value of loss function using BP algorithm [19]. Besides, there is no need to keep the data similarity at a high level for set S and set L. Actually, many approaches based on this theory have been proposed. Yanai and Kawano [17] applied fine-tuning DCNNs to food image classification. In the field of biology and medicine, researchers are applying convolution neural networks and transfer learning to research subjects, and prove their effectiveness. Yuan et al. [22] proposed a method for fingerprint liveness detection based on CNNs. Gurusamy and Subramaniam [18] used a machine learning approach for MRI Brain Tumor Classification. Shin et al. [20] used DCNNs for computer-aided detection.

CNNs based methods are proved effective in classification tasks. This paper focuses on the application of ASL for classification based on CNNs. As Figure 1 shows, an image set is collected by ASL which contains 12000 gesture images is divided into 24 English alphabet letters without class J and Z as their dynamic expression. The images are divided into two groups including training group and test group.

There are some classification approaches for ASL images. For traditional methods, the calculation effort is high that this method can not match real-time task well.

In this paper, ASL images are trained with a CNN based approach with fine-tuning. There are some classes with high similarity that can verify the serviceability of the proposed networks, such as the class of S and the class of T.

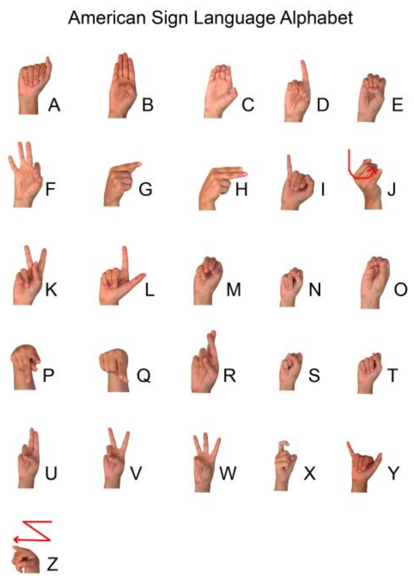


Figure 1. Some examples of ASL image set. The class of S and the class of T have a certain degree of similarity since the two classes have similar outlines

3 Methods

Fine-Tuning strategy is an approach of transfer machine learning [21]. Comparing with the learning approach in scratch manner, Fine-tuning approach that initializes the weights of the CNNs is closer to the point of the best matching one during the gradient decent algorithm execution. In addition, the model is usually trained from a large-scale data set. Accordingly, the weights of transfer learning process will access to the specific region to match the model that requires rapidly training. This approach will solve the problem that a small-scale data set which is prone to over-fitting. Since the scales of our own image set are not large enough for the way of training by scratch, the Fine-tune way is chosen to transfer learning on CNNs. All the experiments are performed by the deep learning framework Caffe [15]. An improved CNNs architecture performs well for fast images classification with fine-tuning strategy.

The gradient conjugate optimization method has been proved to be effective and expansible in solving practical problems [29]. However, in Caffe, the optimization algorithm uses a variety of gradient descent functions, such as Momentum and Adagrad. In Caffe framework, the word depth represents the performed convolutional networks with multiple input channels. For an input $W \times H$ image of depth D at each input location, using a $K \times K$ patch, which could be considered as a $K \times K \times D$ vector. When applying M filters to it, the calculation of convolutional process could be illustrated as Algorithm 1:

Algorithm 1. Convolutional Process of Caffe

for w *in* $1...W$
  for h *in* $1...H$
   for x *in* $1...K$
    for y *in* $1...K$
     for m *in* $1...M$
      for d *in* $1...D$
       $output(w,h,m) += $
       $input(w+x,h+y,d)\times filter(m,x,y,d)$
      end
     end
    end
   end
  end
end

Supposing the softmax function is defined as the following formula:

$$S_1(z) = \left(\frac{e^{z_i}}{\sum_{j=1}^{m} e^{z_j}}\right), i = l,...,m \quad \textbf{(1)}$$

Since calculation the value of each, the predict likelihood of each class could be got.

Then the calculation of loss function could be summarized as the following formula:

$$L(y,z) = -\log\left(\frac{e^{z_y}}{\sum_{j=1}^{m} e^{z_j}}\right) \quad \textbf{(2)}$$

For y is the real class of the input *x,* Function z outputs the result of forward calculation on the vectors of filters.

3.1 Training a Model on CIFAR-10 for Fine-tuning

Considering the similarity of image sizes and categories' numbers between ASL and CIFAR-10 image data set, we are inspired by the solution based on CNNs of the CIAFR-10 image recognition. CIFAR-10 data set consists of 60000 $32\times 32\times 3$ color images which are grouped into 10 classes, divided into 50,000 training images and 10,000 test images. Figure 2 shows the examples of CIFAR-10 data set. A model pre-trained on CIFAR-10 data set since fine-tuning approach requires a model with large-scale training images. However, the scale of CIFAR-10 images set is not big enough to lead the model trained to an over-fit result. In this case, the improved CNNs for training models on CIFAR-10 image set which is composed by 2 convolutional layers and 3 fully connected layers as showed in Figure 3. Both Convolutional layers are followed by a rectified linear units (relu) layer, a pooling layer. Each all-connected layer followed by a dropout layer except the last output layer. It is expected that the dropout strategy would reduce over-fitted situation. In this network, all the parameters are initialized with random Gaussian distributions. In the input layer, mirror parameters have been set to increase samples, the batch size is set to 50 images. Base learning rate is set to 0.001, decreased by a factor of 10 through each 3 epochs until the learning rate decreases to 0.0001. Momentum is set to 0.9 and weight decay is set to 0.0005. After training process of CIFAR-10 data set, a model is gained as the initial weight of our own 10-class image data set.

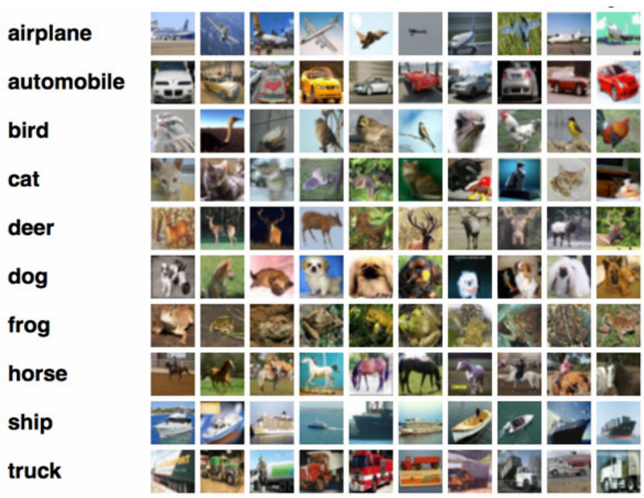


Figure 2. Some examples of the 10 classes image set on CIFAR-10 [17]

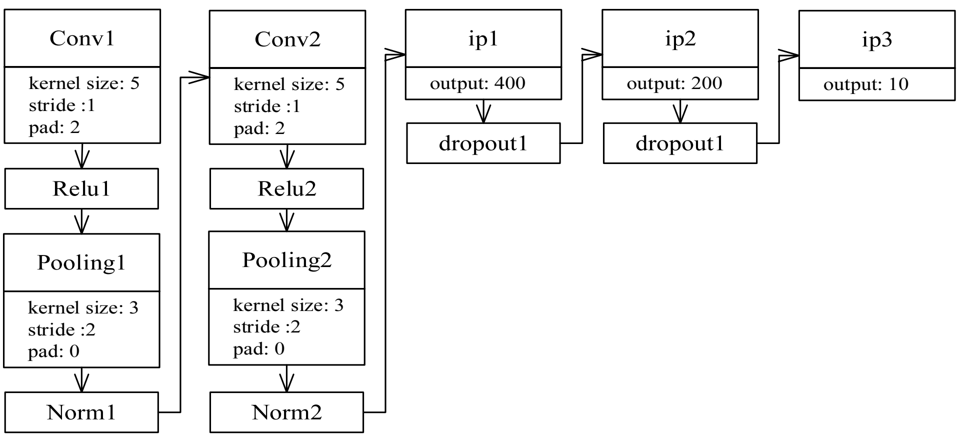


Figure 3. The network architecture of CNNs for training models on CIFAR-10 data set and the small-scale data set

3.2 Fine-tuning CaffeNet

The CaffeNet architecture includes over 60 million parameters which is trained from scratch way. In the pre-trained model, weights have already been initialized weights in each convolutional layer. Usually, the units' number of final output layer is re-set which is equal to the number of classes of the image set. Different from learning from scratch approach in which all the parameters of the network's are initialized randomly, this strategy is efficient where the network's weights of each layer are initialized from a closer position to an optimal value through gradient descent process. In this case, this approach will reduce the iterations of training process of the optimal value. Meanwhile, with the parameters initialized from a large-scale data set, this approach will avoid the likelihood of over-fitting [23]. The network architecture of CaffeNet's off-the-self network is modified to the structure shown in Figure 4. The fc-8 (fully connected) layer's output number is replaced with 24, the same as the number of our own image set and left the parameters of other layers unchanged. Through decreasing the ratio of weight-decay, we prevent the network over-fitting.

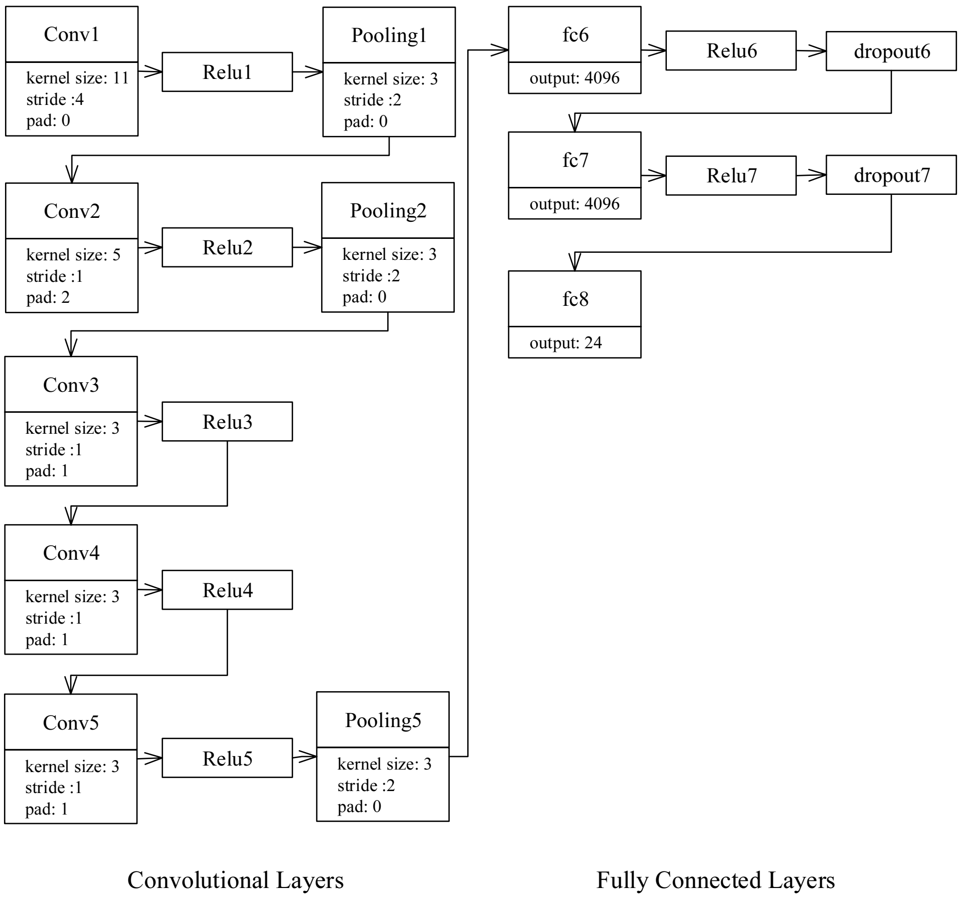


Figure 4. The improved Architecture of CaffeNet The output number of the last fully connected layer is modified to 24. The left part are 5 convolutional layers with regular layers and pooling layers. The right part are 3 fully connected layers with regular layers and drop out layers

4 Experiments

By verifying the scalability and effectiveness of the fine-tuning strategy in the first experiment, we apply this strategy to the gesture classification experiment and verify that the fine-tuning strategy is suitable for it by analyzing and comparing the experimental results.

4.1 Fine-tuning CaffeNet

CIFAR-10 data set contains 60,000 RGB images which is grouped into 10 categories, including 50,000 training images and 10,000 test images. As all the images are small in the size of 32×32 pixels, the networks is employed in Figure 3 by decreasing learning rate by 10 times every 3 epochs during the training process. After 60,000 iterations, the model is trained. Finally, we chose the model with accuracy of 67.06% as pre-trained model to proceed fine-tuning process on our small-scale images set.

In the following transfer training process, the architecture which used in the previous process are remained. The network which consists of two dropout layers would reduce the problem of samples over-fitting. Figure 5 shows the test accuracy when removing the dropout layers. After 6,000 iterations, the result is still at a lower standard. It validates the importance of using dropout layers. As Figure 6 shows, when adding the dropout architectures, the top-1 and top-5 classification accuracy is respectively 73% and 82%. Obviously, it shows that the accuracy of training process is increased rapidly and then becomes smoothly and holds wave nearby a specific value. As the fine-tuning strategy mentioned in the previous section, the fast ascending of accuracy curve shows that this strategy can make a rapidly gradient descent process.

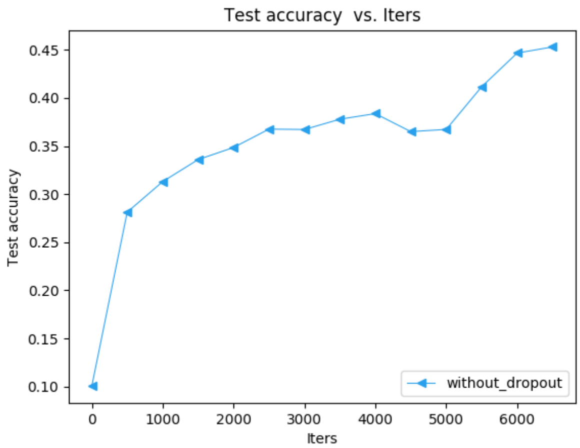


Figure 5. Traces of testing accuracy during training process when removing the dropout layers

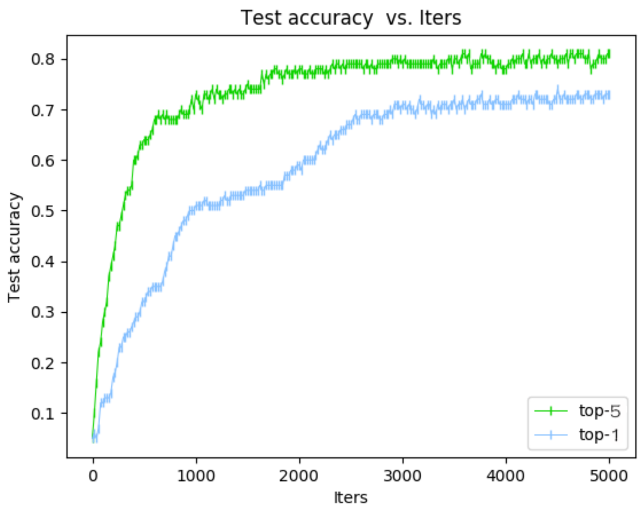


Figure 6. Traces of testing accuracy during the training process with fine-tuning Cifar-10 model. Top-1accuracy was 73%, Top-5 accuracy was 82%

4.2 Experiments on CaffeNet with fine-tuning

In this experiment, the output number of fc-8 is modified to 24 since employed ASL set has 24 classes. Other parameters of the network are unchanged. After the process of training, 24 classification accuracy of 24 alphabets have been gotten, the final accuracy is showed in the Figure 7. The specific values are shown in the Table 1. The highest accuracy value is 1 of class F, G and Y. The Minimum accuracy value is 0.62 of class Y. In view of the small-scale images set is prone to over-fit, implementing another experiment by decreasing the value of weight-decay from 0.005 to 0.0005. By decreasing the punishment term, the over-fitting phenomenon is improved. A Comparison of the training results about test accuracy with S. Ameen's ConvNet [25] is shown in the Figure 8.

Table 1. Results of average loss and average accuracy for ASL fine-tuning CaffeNet

Class	Average Loss	Average Accuracy
A	0.645	0.84
B	0.187	0.96
C	0.205	0.94
D	0.481	0.82
E	0.271	0.94
F	0.003	1.00
G	0.045	1.00
H	0.115	0.94
I	0.306	0.88
K	0.113	0.96
L	0.129	0.96
M	0.999	0.76
N	0.976	0.76
O	1.036	0.76
P	0.991	0.62
Q	0.158	0.98
R	0.481	0.86
S	0.155	0.96
T	0.353	0.90
U	1.056	0.82
V	0.312	0.92
W	0.288	0.88
X	0.417	0.88
Y	0.227	1.00

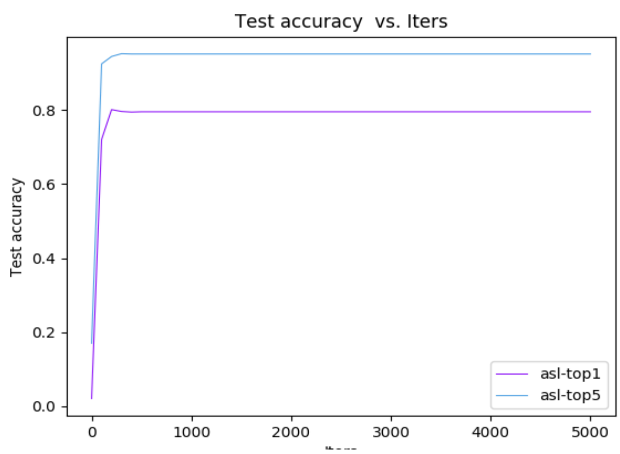


Figure 7. Traces of testing accuracy during the training process with fine-tuning CaffeNet

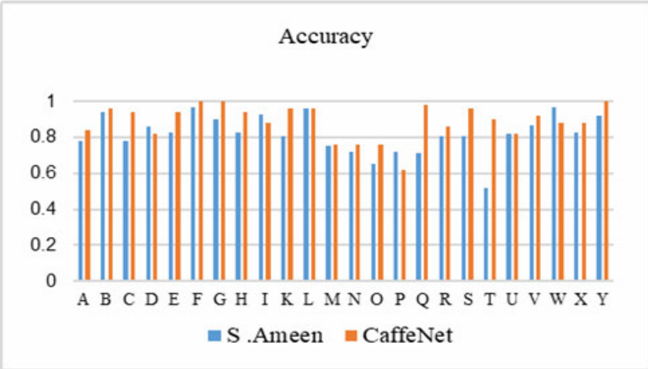


Figure 8. Test accuracy contrast of S. Ameen's ConvNet and our fine-tuning CaffeNet

The average accuracy results over 5-fold are showed in Table 2.

Table 2. 5-fold validation results of average accuracy

Model	Average Accuracy
Trained CNNs from CIFAR-10	0.79
Fine-tuned CaffeNet	0.89

5 Conclusion

In this paper, an approach of CNNs with fine-tuning is employed for ASL image recognition by transfer learning. At the same time, the high performance of this strategy is witnessed. The several experiments in section 4.1 show the importance of using dropout architecture when the training images are in small-scale. After 6,000 iterations of training, the model trained by the strategy of fine-tuning has a better accuracy than the one removed dropout architecture. In the experiments in section 4.2, through using the pre-trained model that has been initialed by training with a large-scale of images, the test accuracy of the networks could be promoted to 0.95, which is far higher than the result in the previous section. These results indicate that this strategy will solve the problem about the small-scale images that are prone to over-fit. In the single classify experiments, the best accuracy results come to 1 of class F, G and Y. Of course, the differences among some classes are not clear enough, such as class M and N. Furthermore, the outlines of the two classes are similar. Therefore, the output of loss function is hard to decrease to a low value at 0.76, and it leads to the test results of the classes performing badly. Meanwhile, it achieves the classification task well by using this fine-tuning strategy.

In the future work, we will continue to explore the approaches that could help the accuracy of fine-tuning training result to a higher standard than state-of-the-art methods [26-28]. Moreover, we will try to decrease the value of the loss function in training process. Last but not the least, we will improve the proposed architecture to make it more suitable for the other classification tasks, such as image copy detection.

Acknowledgments

This work is supported by the National Key R&D Program of China under grant 2018YFB1003205; by the National Natural Science Foundation of China under grant U1836208, U1536206, U1836110, 61602253, 61672294; by the Jiangsu Basic Research Programs-Natural Science Foundation under grant numbers BK20181407; by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund; by the Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET) fund, China.

References

[1] C. Yuan, Z. Xia, X. Sun, Coverless Image Steganography Based on SIFT and BOF, *Journal of Internet Technology*, Vol. 18, No. 2, pp. 435-442, March, 2017.

[2] C. Yuan, X. Sun, R. Lv, Fingerprint Liveness Detection Based on Multi-Scale LPQ and PCA, *China Communications*, Vol. 13, No. 7, pp. 60-65, September, 2016.

[3] Z. Fu, K. Ren, J. Shu, X. Sun, F. Huang, Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement, *IEEE Transactions on Parallel & Distributed Systems*, Vol. 27, No. 9, pp. 2546-2559, December, 2016.

[4] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, K. Ren, A Privacy-Preserving and Copy-Deterrence Content-Based Image Retrieval Scheme in Cloud Computing, *IEEE Transactions on Information Forensics & Security*, Vol. 11, No. 11, pp. 2594-2608, July, 2016.

[5] Y. Lecun, P. Haffner, L. Bottou, Y. Bengio, Object Recognition with Gradient-Based Learning, in: D. A. Forsyth, J. L. Mundy, V. di Gesú, R. Cipolla (Eds.), *Shape, Contour and Grouping in Computer Vision*, Vol. 1681, Springer, Heidelberg, 1998, pp. 319-345.

[6] S. Dutta, R. Murthy, D. Kim, P. Samui, Prediction of Compressive Strength of Self-Compacting Concrete Using Intelligent Computational Modeling, *CMC: Computers, Materials & Continua*, Vol. 53, No. 2, pp. 157-174, October, 2017.

[7] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Trans Pattern Anal Mach Intel*, Vol. l39, No. 6, pp. 1137-1149, June, 2015.

[8] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *ARXIV*, Vol. 14, No. 09, pp. 1556, September, 2014.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision*

and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[11] Q. Cui, S. McIntosh, H. Y. Sun, Identifying Materials of Photographic Images and Photorealistic Computer Generated Graphics Based on Deep CNNs, *CMC: Computers, Materials & Continua*, Vol. 55, No. 2, pp. 229-241, May, 2018.

[12] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, ImageNet: A Large-scale Hierarchical Image Database, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, 2009, pp. 248-255.

[13] Z. Zhou, Y. Wang, Q. M. J. Wu, C. N. Yang, X. Sun, Effective and Efficient Global Context Verification for Image Copy Detection, *IEEE Transactions on Information Forensics and Security* Vol. 12, No. 1, pp. 48-63, August, 2017.

[14] I. Amerini, L.Ballan, R.Caldelli, A. D. Bimbo, G. Serra, A Sift-based Forensic Method for Copy–move Attack Detection and Transformation Recovery, *Information Forensics & Security IEEE Transactions*, Vol. 6, No. 3, pp. 1099-1110, March, 2011.

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrel, Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding, *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando, FL, 2013, pp. 675-678.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, SSD: Single Shot MultiBox Detector, *European Conference on Computer Vision*, Cham, Switzerland, 2016, pp. 21-37.

[17] K. Yanai, Y. Kawano, Food image recognition using deep convolutional network with pre-training and fine-tuning, *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Turin, Italy, 2015, pp. 1-6.

[18] R. Gurusamy, D. V. Subramaniam, A Machine Learning Approach for MRI Brain Tumor Classification, *CMC: Computers, Materials & Continua*, Vol. 53, No. 2, pp. 91-108, October, 2017.

[19] L. K. Saul, S. T. Roweis, Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds, *Journal of Machine Learning Research*, Vol. 4, No. 2, pp. 119-155, June, 2003.

[20] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Deep Convolutional Neural Networks for Computer-aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1285-1298, January, 2016.

[21] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, 2014, pp. 1717-1724.

[22] C. Yuan, X. Li, Q. M. J. Wu, X. Sun, Fingerprint Liveness Detection from Different Fingerprint Materials Using Convolutional Neural Network and Principal Component Analysis, *CMC: Computers, Materials & Continua*, Vol. 53, No. 3, pp. 357-371, November, 2017.

[23] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing coadaptation of feature detectors, *ARXIV*, Vol. 3, No. 4, pp. 212-223, July, 2012.

[24] Z. Xia, X. Ma, Z. Shen, X. Sun, N. N. Xiong, B. Jeon, Secure Image LBP Feature Extraction in Cloud-based Smart Campus, *IEEE ACCESS*, Vol. 6, No. 1, pp. 30392-30401, June, 2018.

[25] S. Ameen, S. Vadera, A Convolutional Neural Network to Classify American Sign Language Fingerspelling from Depth and Colour Images, *Expert Systems*, Vol. 34, No. 3, pp. e12197, June, 2017.

[26] Z. Zhou, C. Yang, B. Chen, X. Sun, Q. Liu, Q. M. J. Wu, Effective and Efficient Image Copy Detection with Resistance to Arbitrary Rotation, *Ieice Transactions on Information & Systems*, Vol. 99, No. 6, pp. 1531-1540, June, 2016.

[27] Z. Zhou, X. Sun, X. Chen, C. Chang, Z. Fu, A Novel Signature Based on the Combination of Global and Local Signatures for Image Copy Detection, *Security & Communication Networks*, Vol. 7, No. 11, pp. 1702-1711, November, 2014.

[28] Z. Zhou, H. Sun, R. Harit, X. Chen, X. Sun, Coverless Image Steganography Without Embedding, *International Conference on Cloud Computing and Security*, Nanjing, China, 2016, pp. 123-132.

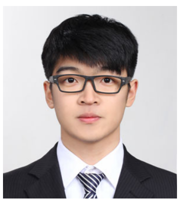
[29] H. Chen, B. Yu, H. Zhou, Z. Meng, Comparison of CS, CGM and CS-CGM for Prediction of Pipe's Inner Surface in FGMs, *CMC: Computers, Materials & Continua*, Vol. 53, No. 4, pp. 271-290, December, 2017.

[30] R. Meng, S. G. Steven, J. Wang, X. Sun, A Fusion Steganographic Algorithm Based on Faster R-CNN, *CMC: Computers, Materials & Continua*, Vol. 55, No. 1, pp. 001-016, April, 2018.

[31] Z. Xia, N. N. Xiong, A. V. Vasilakos, X. Sun, EPCBIR: An Efficient and Privacy-preserving Content-based Image Retrieval Scheme in Cloud Computing, *Information Sciences*, Vol. 387, pp. 195-204, May, 2017.

[32] J. Li, J. Li, Prompt Image Search with Deep Convolutional Neural Network via Efficient Hashing Code and Addictive Latent Semantic Layer, *Journal of Internet Technology*, Vol. 19, No. 3, pp. 949-957, May, 2018.

Biographies



Qi Cui received the B.S. degree in Software Engineering from Nanjing University of Information Science and Technology, China in 2017. At present, he is currently pursuing his PhD in Meteorological Information Technology at the same university. His research interests include deep learning, information hiding, Steganalysis and artificial neural network security.

Zhili Zhou received the M.S. and Ph.D. degrees in computer application from the School of Information Science and Engineering, Hunan University, in 2010 and 2014, respectively. His current research interests include near-duplicate image/video retrieval, image search, image/video copy detection, coverless information hiding, digital forensics, and image processing.

Chengsheng Yuan received his B.S. in Software Engineering from Nanjing University of Information Science and Technology, China in 2014. At present, he is currently pursuing his PhD in Meteorological Information Technology at the same university. His research interests include digital forensic, image processing and network security.

Xingming Sun received his Ph.D. in computing science from Fudan University, China, in 2001. He is currently a Professor at the College of Computer and Software, Nanjing University of Information Science and Technology, China. His research interests include network and information security, database security, information hiding and natural language processing.

Q. M. Jonathan Wu received his Ph.D. in electrical engineering from the University of Wales, Swansea, U.K., in 1990. His current research interests include 3-D computer vision, active video object tracking and extraction, interactive multimedia, sensor analysis and fusion, and visual sensor networks.