Acupoint Detection Based on Deep Convolutional Neural Network

Lingyao Sun[1,2], Shiying Sun[1], Yuanbo Fu[3], Xiaoguang Zhao[1]

1. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
E-mails: sunlingyao2018@ia.ac.cn

2. University of Chinese Academy of Sciences school of Artificial Intelligence, Beijing 100049, China

3. The Department of Acupuncture and Moxibustion, Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing Key Laboratory of Acupuncture Neuromodulation, Beijing 100010, China

Abstract: As an important component of Traditional Chinese Medicine (TCM), science of acupoint therapy has achieved significant results in clinical practice, but recognizing and positioning acupoints is heavily depends on the skills of practitioners. In recent years, researchers have proposed a few methods of automatic acupoints detection and positioning, but most of the methods are still based on manual designed features. In this paper, we propose an acupoints detection method based on deep convolutional neural network, and an evaluation method is proposed for acupoint detection. What's more, we build an acupoint detection dataset. Experiments are performed and a promising result is achieved.

Key Words: convolutional neural network, deep learning, acupoint detection, convolutional pose machine

1 Introduction

As an important part of Traditional Chinese Medicine, acupuncture science has received more and more attention worldwide. Due to the regional differences between Eastern medicine and Western medicine, some researchers of Traditional Chinese Medicine theory have tried to verify the rigor and feasibility of acupoint therapy science through modern medical methods. Some clinical studies have shown promising results for the efficacy of acupoint therapy, for example, reducing both postoperative and chemotherapy nausea and vomiting in adults, as well as postoperative dental pain [1]. Furthermore, acupoint therapy has also shown strong efficacy in treating psychological disorders [2].

As one of the core therapeutic modalities in many Asian countries, acupoint therapy is considered as an effective treatment with low risks. In China, there is a huge demand for acupuncture treatment, but the number of acupuncturists who truly have qualified acupuncture skills is limited. According to our survey, an acupuncturist can only treat about 80 patients a day at the most. What's more, in traditional science of acupoint therapy, locating the acupoint accurately is often highly dependent on the skills of acupuncture practitioners. And for beginners of Chinese Medicine, it will take a lot of time on professional training to learn how to recognize and find the location of acupoints. Therefore, a machine or a robot that can perform the acupoint therapy automatically will reduce the workload of practitioners and benefit to people's daily health care.

Nowadays, there have been some studies on acupoint therapy robots. Wang *et al.* [3] developed a novel acupuncture robot, and designed some needling modes at different angel, depth and strength of needle penetration to imitating the acupuncturist's techniques. Dai *et al.* [4] designed a moxibustion robot based on the medical human back spine model. It is combined with the constant temperature moxibustion principle. Also, a manipulator trajectory has been established through the analysis of the position of each joint and kinematics of the robot. Gao *et al.* [5] developed a Chinese Medical massage robot. An architecture combining series structure and parallel structure is applied to the robot system, and a massage mechanical hand that can accomplish various massage manipulations is developed.

Automatically acupoint location is the basis and the first process of acupoint therapy robot. In recent years, researchers began to apply the current technologies to explore methods for automatic recognition and location of acupoint. Lin *et al.* [6] showed that the electrical resistance of the acupoint is lower than the surrounding skin, and propose a method of locating acupoints by using this electrical resistance specificity. Zhang *et al.* [7] proposed an acupoint positioning and tracking method based on template matching. At first, they place red circular marks at acupoints on the back of the human body, and use the fast template matching algorithm to obtain the center coordinates of each red mark. Then the world coordinates of acupoints are calculated by camera calibration technology. At last, the artificial marks on the human neck are tracked, and motion vector which would be added on acupoint coordinate is calculated to achieve the tracking of the acupoints. Chang *et al*. [8] proposed a method based on the traditional pattern recognition algorithm to obtain facial feature points and calculate the facial acupoints coordinates by the theory of traditional Chinese medicine bone length measurement. Firstly, ASM algorithm is used to locate facial feature points. Then the distance between the center point of eyebrows and midpoint of hairline is computed and considered as 3'Cun'. Finally, according to facial feature points and body inches, the corresponding facial acupoints are found.

*This work was Supported by the National Natural Science Foundation of China (61673378, 61421004).

So far, the methods of acupoint positioning are mainly based on artificial design features or geometric calculation and no supervised training algorithms is applied. Due to the huge success in classification and object detection tasks, deep learning has become a research hotspot nowadays and is gradually applied to various fields. So we came up with an idea of applying deep convolutional neural network to adaptively extract features of acupoints so as to predict the position of acupoints end-to-end.

One of the applications of deep learning is human key point detection. It is an important basis of human pose estimation and human behavior prediction. Since the applicability of deep neural networks for human pose estimation has been examined [9], more and more researchers consider to apply deep learning algorithms in human key point detection. Wei *et al.* [10] proposed a multi-stage cascaded network named Convolutional Pose Machines (CPM). With this method, convolution layers are used to transfer texture information and spatial information at the same time. It has achieved an excellent detection results in human pose estimation. Newell *et al.* [11] introduced a "stacked hourglass" convolutional network architecture for the task of human pose estimation, and the ability to extract features at different scales is improved. Yang *et al.* [12] designed a Pyramid Residual Module (PRMs) to enhance the invariance in scales of deep convolutional neural network. Xiao *et al.* [13] proposed to incorporate convolutional neural networks with a multi-context attention mechanism into an end-to-end framework for human key point detection.

The distribution of acupoints in the human body is highly dependent on the key points of the human skeleton. Therefore, we apply the human key point detection approach into acupoints detection, and propose to use Convolutional Pose Machine (CPM) for acupoint detection because of the good performance in leaning local and spatial features. A dataset for acupoint detection is built by collecting images of human body. What's more, an evaluation index for acupoint detection is also proposed and the experiments are conducted.

The organization of this paper is as follows. Section 2 depicts the detail of the proposed method which consists of the base network of feature extractor, the network structure and the loss function. Section 3 shows the experiment and gives the analysis of experimental results. Finally, Section 4 concludes this paper.

2 Proposed method

This section introduces the framework of the proposed method, which includes the feature extractor, the multi-stage network structure and the loss function. The whole network architecture is shown in Fig 1. The original images of human body and the labeled images are input into the network. During training, the predicted heatmap is generated by the network. The loss between predicted heatmap and ground truth heatmap is calculated based on the loss function for back-propagating. When the model converges, test images can be input into the network to get predicted heatmap. At last, the point with largest response value on the heatmap is found to get the coordinate as the location of the acupoint.

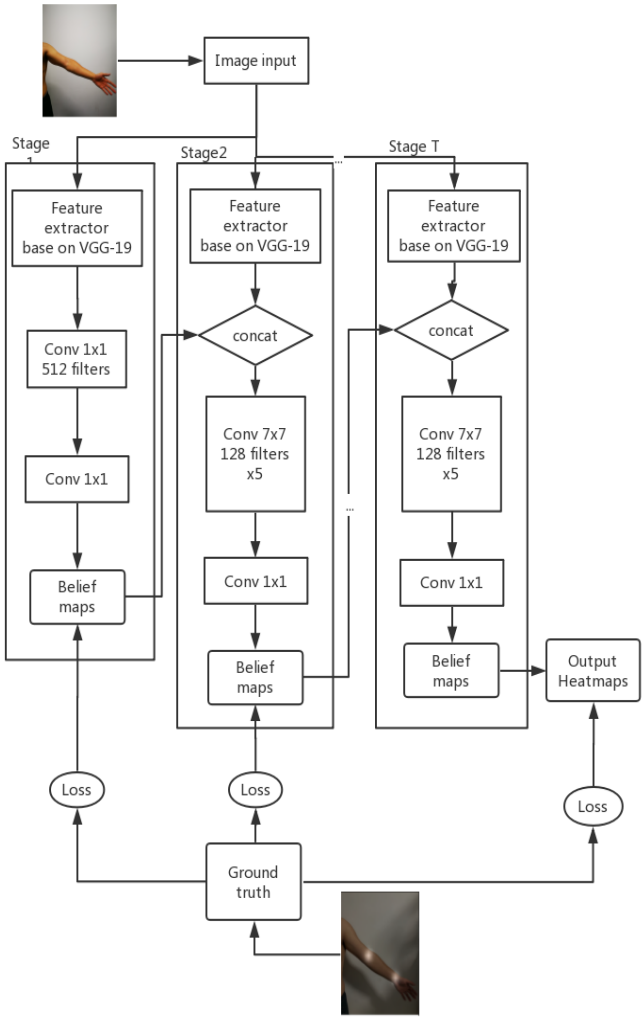

Figure 1: Network Structure

2.1 Feature extractor network

The network's backbone in this paper is based on VGG-19[14] and it is used as a feature extractor. VGG is a CNN structure developed from AlexNet[15], and it is inherited from the basic architecture of AlexNet. But the difference is that VGG replaces the larger convolution kernel in AlexNet with multiple smaller kernels (3×3). The author believes that increasing the number of the convolutional layers can deepen the depth of the network, and get better performance. At the same time, a smaller convolutional kernel can reduce the network's parameters. For example, one 5×5 convolutional kernel has 25 parameters, while two 3×3 convolutional kernels have only 18 parameters, and the size of the receptive field obtained by stacking two 3×3 convolution layers is equivalent to a layer of 5×5 convolution. In addition, VGG increases the number of channels, and the maximum number of channels is 512. With the increasing number of channels, more features can be extracted. In this paper, the first five convolutional layers of VGG-19 are used as feature extractors. The specific network structure is shown in Table. 1.

Table 1: Feature extractor

Type	Filters	Size
Convolutional	64	3×3
Convolutional	64	3×3
MaxPooling[2,2]		
Convolutional	128	3×3
Convolutional	128	3×3
MaxPooling[2,2]		
Convolutional	256	3×3
Convolutional	256	3×3
Convolutional	256	3×3
Convolutional	256	3×3
MaxPooling[2,2]		
Convolutional	512	3×3
Convolutional	512	3×3
Convolutional	512	3×3
Convolutional	512	3×3
Convolutional	512	3×3
Convolutional	512	3×3
Convolutional	128	3×3

2.2 Multi-stage network structure

CPM uses a multi-stage cascaded network structure. In the first stage, the input images are passed into the feature extractor, and the outputs is a belief map of the image after a 1×1 convolutional layer. Belief map reflects the spatial response of acupoints. Assuming that there are p acupoints in total, the belief map has p+1 layers. Each layer represents a acupoint heatmap, the left one is an empty response, which represents the background of the image. The confidence of a heatmap's corresponding acupoint is reflected by each pixel in the heatmap. The larger the pixel value is, the greater the probability of the acupoint locating at this position is.

When stage$\geqq$2, the input of the network contains two contents: 1) Feature map, extracted by the original image. It represents the image texture information. 2) Belief map, generated by the previous stage, represents all spatial information for acupoints. The network outputs a belief map to the next stage, and the output of the last stage is used as the result. The network structure is shown in Fig. 1.

2.3 Loss Function

There are many methods for constructing loss function. DeepPose [9] use a method of direct regression coordinates and define the loss function as the Euclidean distance between the predicted acupoint and the ground truth. This method can directly obtain the coordinates of the acupoint. However, the method of directly regression to the coordinates only provide one supervision information during training: the offset between the acupoint and ground truth, which makes it difficult for the network to converge. The method of constructing a heatmap is to perform regression at each pixel and calculate the probability of whether each pixel is a acupoint, so it can provide more supervision information. In addition, pixel-level prediction can improve the positioning accuracy.

In this paper, we choose to construct a heatmap to calculate the loss function. By placing a Gaussian response at the acupoint of the image, a ground truth's heatmap is built. The loss is obtained by calculating the L2 norm between the predicted heatmap and the true heatmap. Assuming that k represent the corresponding acupoint, $b(k)$ represents the predicted believe map and $b(k)^*$ represents the true value believe map, the loss function can be expressed as:

$$loss = \sum_{k} \left\| b(k) - b(k)^* \right\|_2^2 \qquad (1)$$

The loss is calculated at the end of each stage, and finally the losses of every stage is added to calculate the gradient and back-propagate. The ground truth's heatmap is shown in the Fig.2.

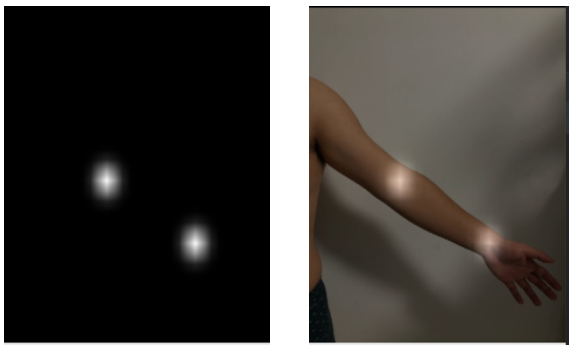

Figure 2: Ground truth's heatmap

The belief map of the previous stage and the image features extracted from the current stage are combined by this network, refines the final belief map by stacking multiple stages. Then intermediate supervision is performed on each stage to calculate the loss to ensure that gradient can be transmitted back effectively, which prevents the phenomenon of gradient vanishing.

3 Experiments

3.1 Experiment data Set

There isn't a dataset for acupoint detection, so we build a dataset by capturing the images of human body. The dataset consists of 719 images, and each image contains only one upper limb of a person. Because the data set is relatively small by now, the acupoint near the joints have more obvious features and can be detected easier. We select two acupoints near the joint of elbow and wrist, named Quze and Daling. The Quze point is in the center of the horizontal line of elbow, the side edge of the biceps brachii, and the Daling point is located at the midpoint of the horizontal line of the wrist, as is shown in Fig. 3.

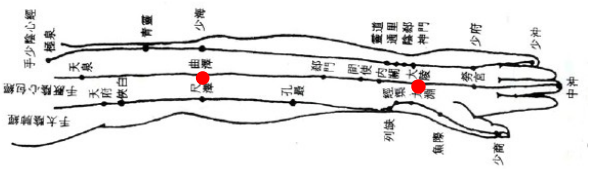

Figure 3: Position of acupoints on upper limb

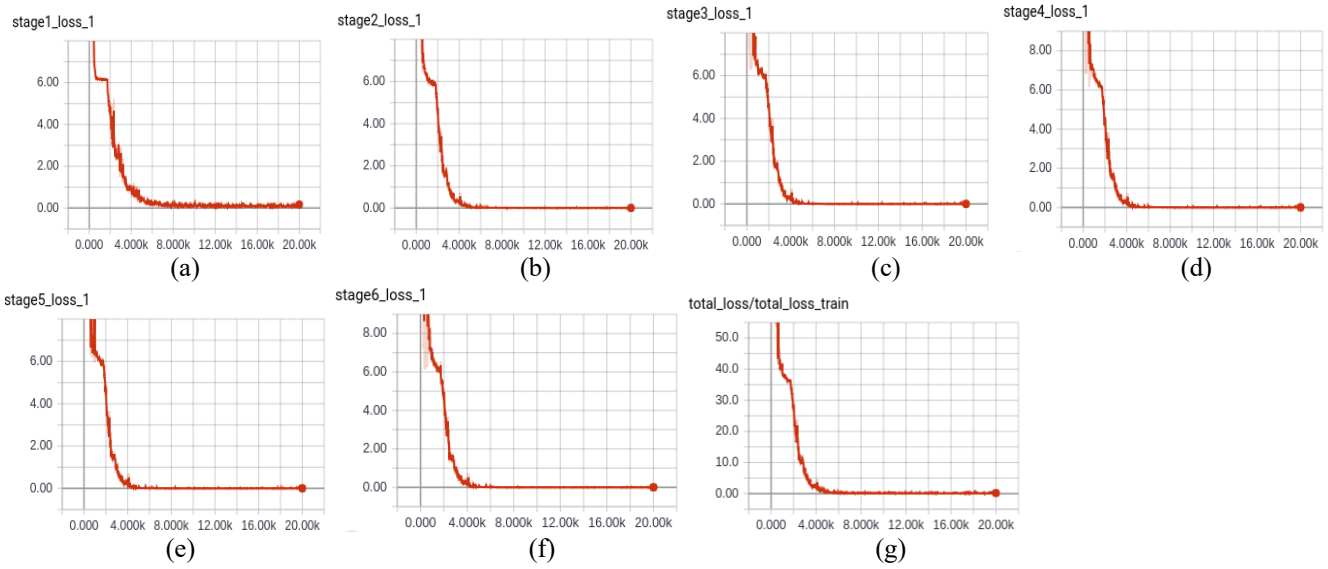

Figure 5: Loss curves, (a)-(f) are loss curves of stage1-6, (g) is the total loss.

The whole dataset is divided as follows: 519 training images, 100 validation images and 100 test images. We labeled the acupoint manually by using 'Labelme' to get the corresponding json file, as is shown in Fig. 4.

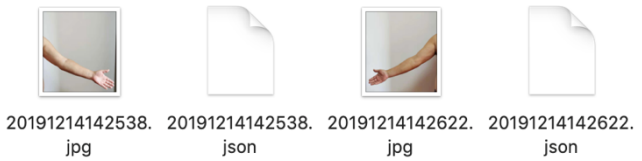

Figure 4: Data set

3.2 Training

We implemented the network with Tensorflow and the model is trained using the dataset. The platform for training is the 64-bit Ubuntu 16.04 operating system with a hardware configuration of Intel Xeon E5-2630 processor. The GPU is NVIDIA GeForce GTX 1080 Ti with memory of 11GB. The training time is about 11 hours.

We set 20,000 steps. When the training reaches about 10,000th step, the model basically converges. The various stage loss and total loss curves are shown in Fig.5. After training, we tested our model on a platform of Macbook with 3.1 GHz Intel Core i5 CPU, and the processing time of detecting acupoints is 2.36s per image on average.

3.3 Experimental results and analysis

After training, the method is tested with 100 images of upper limbs in the test set. We calculate the offset between predicted points and the labeled points. Because of the different scale of images, the Euclidean distance cannot be directly used to evaluate the accuracy. The error should to be normalized. So the definition of offset error is given by:

$$offset\ error = \frac{\left\| P_i - \hat{P}_i \right\|_2}{d} \qquad (2)$$

Where P_i is the labeled value of acupoint and $\hat{P}_i$ is the predicted value. d is the scale normalization factor, which is defined as the Euclidean distance between two acupoints in the labeled image:

$$d = \left\| P_1 - P_2 \right\|_2 \qquad (3)$$

The relationships between d, P_i and $\hat{P}_i$ are shown in Fig. 6.

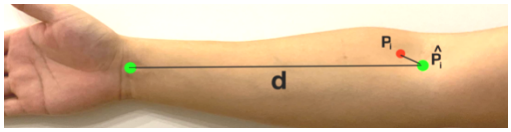

Figure 6: Relationship between P_i, $\hat{P}_i$ and d

We calculated the detection rate under different offset error. And the relationship between offset error and detection rate is shown in Fig.7.

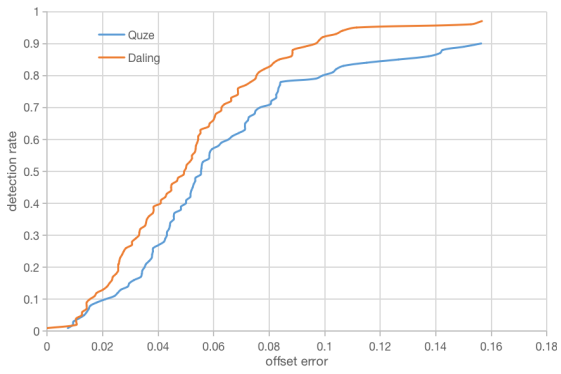

Figure 7: Detection rate curves

The length of the forearm of an adult is 25cm on average, During the process of acupressure, the radius of the finger contact surface is assumed to be 2cm. Under this condition, d is assumed as 25cm and $\left\| P_i - \hat{P}_i \right\|_2$ is assumed to be

2cm, the allowable offset error is calculated by Equation (2) , and is equal to 0.08. In other words, when the offset error is lower than 0.08, it is considered as a successful detection. Under this threshold, the detection rate of Quze point is 70% , and the detection rate of Daling point is 82%, which is a promising result. The examples of detection results are shown in Fig. 8.

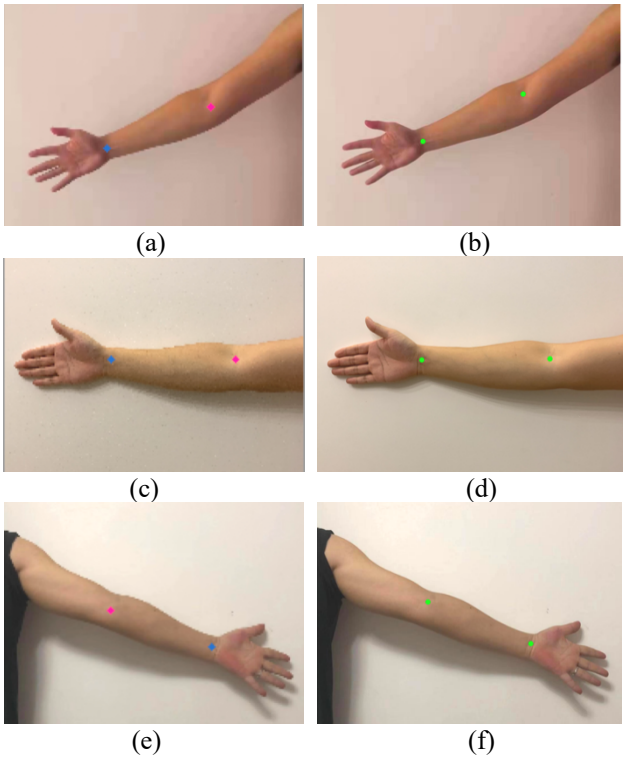

Figure 8: Examples of detection results. (a), (c), (e) show predicted points while (b), (d), (f) show labeled points.

4 Conclusion

In this paper, we use a deep learning method to detect acupoints automatically. An acupoints detection dataset is built and CPM method is trained. Then an evaluation method is proposed. The experiments are conducted on the proposed dataset and the result is promising. By now, we only labeled and detected two acupoints near the joints due to the limited number of train images. However, the experimental data is continuously being collected, the detection of more acupoints and further experimental results will be given in the future work.

References

[1] National Institutes of Health (NIH), Consensus development panel on acupuncture acupuncture—NIH consensus conference, The Journal of the American Medical Association, vol. 280, no. 17, pp. 1518 - 1524, 1998.

[2] Feinstein, David. Acupoint stimulation in treating psychological disorders: Evidence of efficacy.[J]. *Review of General Psychology, 2012*, 16(4):364-380.

[3] Yu Wang, Chen Zhao. Novel acupuncture robot[J]. *Robot Technique and Application*, 2007, 000(004):37-39.

[4] Yaonan Dai, Xubing Chen. Design and Implementation of Moxibustion Robot with Constant Temperature on Human Back Spine[J]. *Computer Engineering and Applications*, 2019, 55(09):222-228+276.

[5] Huanbin Gao, Shouyin Lu, Tao Wang, et al. Research and Development of Chinese Medical Massage Robot[J]. *Robot,* 2011(05):43-52.

[6] Dongdong Lin, Huijie Lin.A preliminary study of acupuncture location using high-frequency pulses[C]. *Progress in China's Biomedical Engineering-Proceedings of the 2007 China Biomedical Engineering Joint Academic Conference (Part 2)*. 2007.

[7] Huakai Zhang, Shouyin Lu, Guangyue Du. Acupoint Positioning and Tracking based on Template Matching[J]. *Bulletin of Science and Technology*, 2011(05):34-38.

[8] Menglong Chang, Qing Zhu. Automatic location of facial acupuncture-point based on facial feature points positioning[C]*2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*. 2017.

[9] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus*, OH, 2014, pp. 1653-1660.

[10] S. Wei, V. Ramakrishna, T. Kanade and Y. Sheikh, "Convolutional Pose Machines," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Las Vegas, NV, 2016, pp. 4724-4732

[11] Newell A. , Yang K. , Deng J. (2016) Stacked Hourglass Networks for Human Pose Estimation. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, vol 9912.

[12] W. Yang, S. Li, W. Ouyang, H. Li and X. Wang, "Learning Feature Pyramids for Human Pose Estimation," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 1290-1299

[13] X. Chu, W. Yang, W. Ouyang, C. Ma, A. L. Yuille and X. Wang, "Multi-context Attention for Human Pose Estimation," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu,* HI, 2017, pp. 5669-5678.

[14] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409-1556.

[15] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]. *Proceedings of the 25th International Conference on Neural Information Processing Systems* - Volume 1December 2012, pp. 1097 - 1105