JOURNAL OF NEW RESULTS IN SCIENCE (JNRS)
ISSN: 1304-7981
http://dergipark.gov.tr/jnrs
Research Article

Volume : 7
Issue: 3
Year: 2018
Pages: 9-16

Received: 31.01.2018
Accepted: 14.09.2018

Editors-in-Chief: ***Ebubekir ALTUNTAS***
Area Editor: ***Turgut ÖZSEVEN/Bilge GÖZENER***

Classification of Different Age Groups of People by Using Deep Learning

Özkan İNİK[a*], Bülent TURAN[b]

[a,b]*Department of Computer Engineering, Gaziosmanpaşa University, Tokat/ Turkey.*
**Corresponding author: ozkan.inik@gop.edu.tr*

ABSTRACT: The Purpose of this study is to classify human images of different age groups with VggNet which is one of the Deep Learning (DL) models. Artificial intelligence, machine learning and computer vision have been carried out in recent years at very advanced level. Undoubtedly, it is a great contribution of DL in the rapid progress of these studies. Although DL foundational is based on past history, it has become popular in the imageNet competition held in 2012. This is because the top-5 error rate of 26.1% for visual object description has fallen to 15.3% for the first time with a sharp decline that year with DL. The Convolution Neural Network (CNN) is basis of DL models. It is basically composed of 4 layers. These are Convolution Layer, ReLu Layer, Pooling Layer and Full Connected Layer. DL models are designed using different numbers of these layers. In this study, people are divided into 12 classes according to age groups. These classes are man, woman, man face, woman face, old man, old woman, old man face, old woman face, boy, girl, boy face, girl face respectively. A new data set was created for people in 12 different age categories. For Each class 150 and totally 1800 images were collected. 90% of these images were used for training and the remaining 10% were used for testing. VggNet was trained with this data set. As a result of the study, it was seen that people in different age groups were estimated with 78.5% accuracy with VggNet model. DL models need to be trained with large data required. But it has been seen that training success has achieved a certain value with little data.

Keywords – *Classification of people, Deep Learning, VggNet, CNN*

1. Introduction

DL is a subdivision of artificial intelligence and became popular in 2012 for the first time. The 2012 ImageNet contest was won by AlexNet (Krizhevsky et al., 2012) , a DL model, which has been a huge success in object classification. The ImageNet competition has been gained with continuous DL models since 2012. Due to this success of DL, the use of deep networks has become widespread in areas such as voice recognition (Amodei et al., 2016; Bahdanau et al., 2016; Graves et al., 2013; Hinton et al., 2012), natural language processing (Hermann et al., 2015; Jozefowicz et al., 2016; Lample et al., 2016; Luong et al., 2015), robotics (Lenz et al., 2015; Levine et al., 2016) and object identification (Long et al., 2015; Redmon et al., 2016; Ren et al., 2015). DL performs the learning process on raw data (Bengio et al., 2013). The basic structure of DL explores knowledge by creating representations in different layers (LeCun et al., 2015). Deep nets do not have a pre-processing step such as cropping as in traditional methods for identifying objects on an image (Krizhevsky et al., 2012). In the literature, more than one study has been done about human identification and perception of movements. Some of these are; Alexander Toshev and Christian Szegedy used DL in the prediction of the human pose (Toshev and Szegedy, 2014). DL used for pedestrian detection (Tian et al., 2015; Zeng et al., 2013). It has redefined people from different characteristics through DL (Ahmed et al., 2015). Face identification, which is composed of 10,000 classes, was performed by DL (Sun et al., 2014). The deep learning-based methods have been developed for human redefinition (Yi et al., 2014).

The structure of this work is as follows: Section II describes VggNet and Data Set. In Section III, experimental work with VggNet model is presented. Finally, Section IV describes the Discussion and Conclusion.

2. Material and Methods

A. VggNet

VggNet is a DL model designed by Karen Simonyan and Andrew Zisserman (Simonyan and Zisserman, 2014). The model has been the winner ImageNet competition with a %7.3 top-5 error rate at the 2014. Its basic layer structure is similar to the AlexNet(Krizhevsky et al., 2012) model. VggNet model is important because of it is deeper than previous models. Thus, VggNet has reinforced the idea that DL networks should have a deep layer of network for the study of the hierarchical representation of visual data.

The input image size on VggNet is 224x224x3. After the input layer, the convolution layer comes and it connected to the input layer with 3x3 receptive fields. In other words, 3x3 filters are designed in the convolution layer. Filters are routed on the image by 1x1 step (stride 1). The loss of information has been eliminated from the previous layer to the next layer with the number of stride being 1. In the model, after some convolution layers, the Pooling layer comes. The filter size in the Pooling layers is 2x2 and the stride is 2. A stack of convolution and pooling layer is followed by 3 fully connected layers. The output size of the first two fully connected layers is 4096. The output value of the last layer is originally 1000. This value is equal to the number of classes. In this study, this layer output size was updated to 12. Because the number of classes we do classify is 12. After all hidden layers, Rectification Linear Unit (ReLu) layer is used. ReLu makes the negative values to zero and leaves the other values in the same.

The VggNet model is given in Figure 1. It has 6 different architecture of VggNet in Figure 2. The configuration D (VggNet-16) produced best result. Therefore this model is used in classifications.

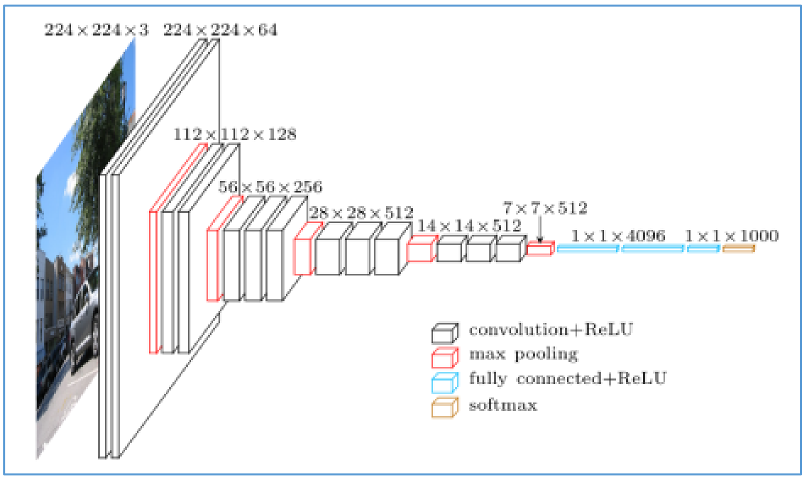
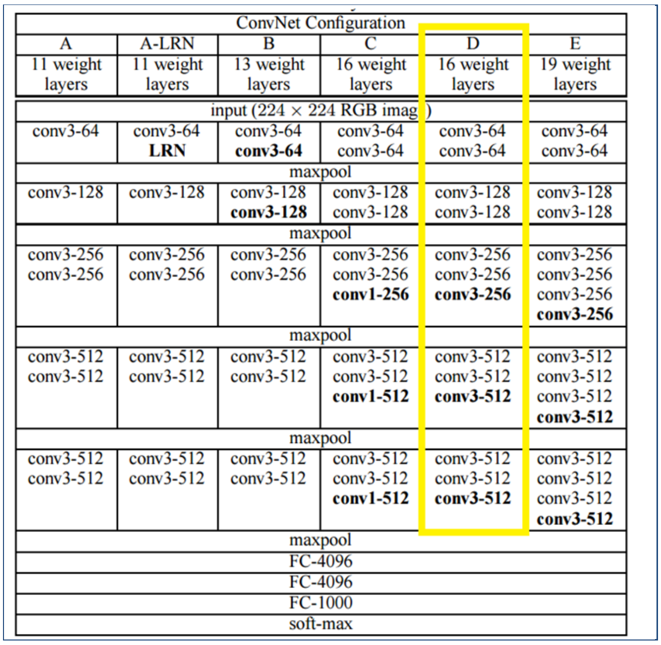


Figure 1. VggNet model architecture(Heuritech, 2018).



ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2. Different architecture of VggNet model (Deshpande, 2018).

B. Data Set

A new data set was created for this study. The data set is designed to be composed of 12 classes according to estimated human gender and age. These classes are; boy, girl, man, women, old man, old women, boy face, girl face, man face, women face, old man face and old women face. Data set consist of 1800 image and each class has 150 images. 90% of these image used for training and remaining images used for testing. The classes of the prepared dataset are given in Figure 3.

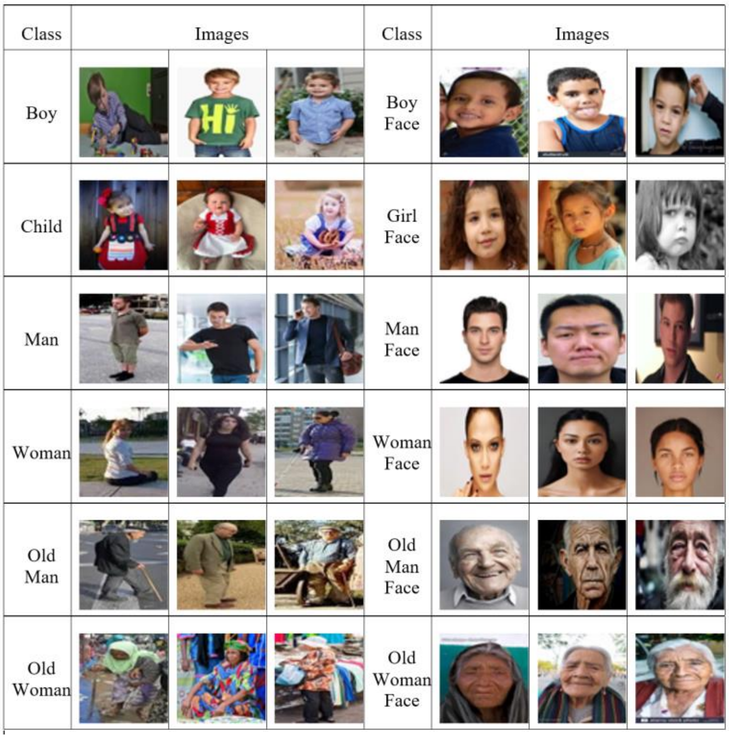


Figure 3. The classes of dataset

3. Experimental Studies

In the experimental work, Intel Core i7 7700HQ 2.8GHz processor, 16GB Ram and GeForce GTX1050 graphics card were used. Applications have been made on Matlab R2017a 64bt (win64).

Precision, Accuracy, Recall and F-Measure are usually used for model performance in classification. Calculation of these values is carried out with the Equation (1-4).

		PREDICTION	
		Positive	Negative
ACTUAL	Positive	TP	FN
	Negative	FP	TN

TP: True Positive

FN: False Negative

FP : False Positive

TN: True Negative

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precission + Recall} \tag{4}$$

The confusion matrix for the classification is given in Table 1. Looking at Table 1, the classes at the top are model predicted, while the classes at the left are actual classes. In Table 2, explanations of the classes in the confusion matrix are given. In Table 1, the predicted of the Boy who is indicated with 1, it is estimated as 60% Boy, 10% Man, 20% Man Face. When we look at the classification in this test images, 100% of the gender was estimated as male.

Table 1. Confusion matrix for classification people withVggNet

		PREDICT											
	Num.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
ACTUAL	**1**	0,6	0	0,1	0,1	0	0	0	0	0,2	0	0	0
	2	0	0,9	0	0	0	0	0	0	0	0,1	0	0
	3	0	0	0,67	0,13	0	0	0	0	0,07	0	0,13	0
	4	0	0	0	1	0	0	0	0	0	0	0	0
	5	0	0	0	0	0,87	0	0	0,13	0	0	0	0
	6	0	0	0	0	0	1	0	0	0	0	0	0
	7	0	0	0	0,08	0	0	0,92	0	0	0	0	0
	8	0	0,08	0	0	0,15	0	0	0,77	0	0	0	0
	9	0,28	0	0	0	0	0	0,07	0	0,57	0	0,07	0
	10	0	0	0	0	0	0,1	0	0,1	0	0,7	0	0,1
	11	0	0	0,12	0,06	0	0	0,06	0	0,12	0	0,65	0
	12	0	0	0	0	0	0	0	0	0	0,2	0	0,9

The classification success of the model is 78.5% accuracy. Other classification performance values are given in Table 3**.**

Table 2. Classes in the confusion matrix

Num.	**Class**	**Num.**	**Class**	**Num.**	**Class**	**Num.**	**Class**
1	Boy	**4**	Woman	**7**	Boy Face	**10**	Woman Face
2	Girl	**5**	Old Man	**8**	Girl Face	**11**	Old Man Face
3	Man	**6**	Old Woman	**9**	Man Face	12	Old Woman Face

Table 3. Performance values of the VggNet model for classification of people

Num.	Class	Precision	Recall	F- Measure
1	Boy	0,682	0,600	0,638
2	Girl	0,918	0,900	0,909
3	Man	0,753	0,670	0,709
4	Woman	0,723	1	0,839
5	Old Man	0,853	0,870	0,861
6	Old Woman	0,909	1	0,952
7	Boy Face	0,876	0,920	0,897
8	Girl Face	0,770	0,770	0,770
9	Man Face	0,594	0,576	0,585
10	Woman Face	0,700	0,700	0,700
11	Old Man Face	0,765	0,643	0,699
12	Old Woman Face	0,900	0,818	0,857
	Mean Average	**0,786917**	**0,788917**	**0,7847858**

Deep learning models detect the properties of objects from the filters in the convolution layers. Thus deep nets get the best filters to discover image properties during the training phase. In the study, the filters obtained in the first convolution layer of VggNet model are given in Figure 4. In Figure 1, it is seen that the VggNet model has a filter size of 3x3 and a filter number of 64 in the first convolution layer. Each of these filters is applied to the input image and output images are generated. These images are called activation map or feature map. The structure of filters in the first convolution layer of a trained network can tell us about network training. If the filters in this layer are meaningful, then the network training process is meaningful. It can be seen from Figure 4 that the filters a meaningful structure. The activation map generated by applying an image of the filters in the first convolution layer in the test phase is shown in Figure 5. Looking at the activation map in Figure 5, it is seen that the effect of each filter is different. So each filter tries to discover a different feature in the image.

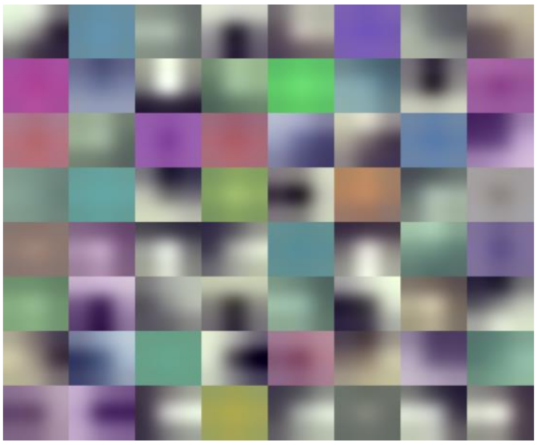


Figure 4. Filters in the first convolution layer

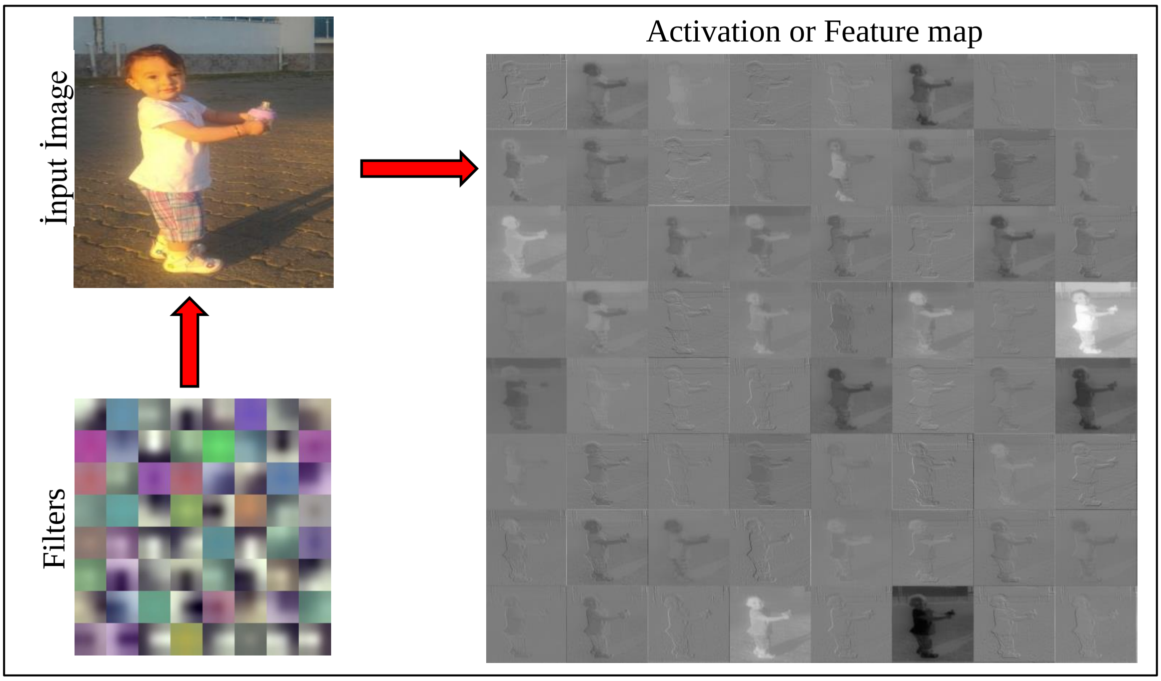


Figure 5. The activation map that occurs after applying the filters of the first convolution layer to the input image

4. Discussion and Conclusion

Automatic classification of people by intelligent systems is very important for many applications. Especially it needed for automatic biometric identification, security checks in public areas, and examination of traffic monitoring camera records. Various artificial intelligence methods have been developed for this classification. However, it is a difficult problem. Perhaps it might be easier to separate genders from each other, but it was harder to predict the age groups of different gender. For example, it is hard to distinguish between Boy and Man or Girl and Woman. In order to overcome these difficulties, this study was conducted to learn about the performance of DL model.

In this study, the classification of people determined according to age and gender criteria was done with VggNet model. It is more appropriate to use this model in the human classification problem because the success of this model in the ImageNet competition compared to other models is higher. A new data set has been created to train the pretrained VggNet model. The data set consists of 1800 images. As a result of the experimental studies, the classification of the people was done 78.5% accuracy with VggNet. Considering the necessity of training deep learning models with large data, it has been seen that accuracy of model has achieved an acceptable rate with little data. It has been understood that the data set has to be increased in order to achieve higher success rates.

5. Acknowledgment

This study was supported by Gaziosmanpaşa University under Scientific Research Projects. Project No: 2017/39 Project Name: " Derin Öğrenme İle Nesne Tanımlama".

6. References

Ahmed, E., Jones, M., Marks, T.K., 2015. An improved deep learning architecture for person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3908-3916.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., 2016. Deep speech 2: End-to-end speech recognition in english and mandarin, International Conference on Machine Learning, pp. 173-182.

Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., Bengio, Y., 2016. End-to-end attention-based large vocabulary speech recognition, Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, pp. 4945-4949.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation Learning: A Review and New Perspectives. Ieee T Pattern Anal 35, 1798-1828.

Deshpande, A., 2018. https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html.

Graves, A., Mohamed, A.-r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks, Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. IEEE, pp. 6645-6649.

Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P., 2015. Teaching machines to read and comprehend, Advances in Neural Information Processing Systems, pp. 1693-1701.

Heuritech, 2018. https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine 29, 82-97.

Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., Wu, Y., 2016. Exploring the limits of language modeling. arXiv preprint arXiv:1602.02410.

Krizhevsky, A., Sutskever, I., Hinton, G., 2012. ImageNet classification with deep convolutional neural networks. In NIPS'2012 . 23, 24, 27, 100, 200, 371, 456, 460.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C., 2016. Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436-444.

Lenz, I., Lee, H., Saxena, A., 2015. Deep learning for detecting robotic grasps. The International Journal of Robotics Research 34, 705-724.

Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., Quillen, D., 2016. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. The International Journal of Robotics Research, 0278364917710318.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440.

Luong, M.-T., Pham, H., Manning, C.D., 2015. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks, Advances in neural information processing systems, pp. 91-99.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sun, Y., Wang, X., Tang, X., 2014. Deep learning face representation from predicting 10,000 classes, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891-1898.

Tian, Y., Luo, P., Wang, X., Tang, X., 2015. Pedestrian detection aided by deep learning semantic tasks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5079-5087.

Toshev, A., Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1653-1660.

Yi, D., Lei, Z., Liao, S., Li, S.Z., 2014. Deep metric learning for person re-identification, Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, pp. 34-39.

Zeng, X., Ouyang, W., Wang, X., 2013. Multi-stage contextual deep learning for pedestrian detection, Proceedings of the IEEE International Conference on Computer Vision, pp. 121-128.