CNN-powered body and face detection for intelligent people counting in Covid-19 restricted places

Michał Wieczorek[1]

[1]Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland

Abstract

In the reality of 2021's Covid-19 pandemic there are a lot of government's restrictions made to reduce the virus spread speed in the society. One of the examples of such restrictions are the people per square meter limit in public places and shopping malls. Because manual counting of each person in such places is not possible due to limited time and money resources, this limit restriction is widely abused making the pandemic outbreak more dangerous. To face this problem author has presented a novel face and body detection model for the CCTV (Closed Circuit Television) monitoring systems, that automatically counts the amount of people in the monitored area by the use of Deep Learning.

Keywords

Convolutional Neural Network, Detection, Recognition, Image processing, Deep Learning

1. Introduction

Object detection is important field of image processing. We can conclude many important operations related to detection and recognition of various objects. Among them very important are body detection and face recognition. We can find many applications of these ideas and a variety of models developed for particular purposes.

Model presented in Li [1] used for body pose detection a complex model based on SVM (Support Vector Machine) co-working with PSO (Particle Swarm Optimization). Results have shown that heuristic model was able to detect locations which were further classified by SVM classifier. The idea presented in Chen, Wang, Li, and Hong [2] was based on repetitive assembling of information from object detection. In Wang, Chen, Zheng, and Li [3] facial rotation model was based on marking changes compared by proposed classifier. Network construction based on clustering was proposed in Zhao, Luo, Quan, Liu, and Wang [4]. The idea has introduced cluster-wise processing for body detection. A model proposed in our previous research Woźniak, Wieczorek, Siłka, and Połap [5] was developed for sensoric data. The idea was oriented on evaluation of numerical information concerning body position read from sensors located on human body. We have developed a model of Recurrent Neural Network. The research presented in Yun, Park, and Cho [6] is oriented on pose rotation detection, where the main detection part was implemented by using self-supervised learning procedure. Supervised learning was also presented for eye tracking techniques Huang, Chen, Zhou, and Xu [7]. In Nadeem, Jalal,

IVUS'21: 26th International Conference Information Society and University Studies, April 23, 2021, Kaunas, Lithuania
michal_wieczorek@hotmail.com (M. Wieczorek)
0000-0002-5319-3366 (M. Wieczorek)

and Kim [8] was proposed a model of neural network evaluating actions of body to recognize the state in which human was. The model proposed in Winnicka, Kęsik, Połap, Woźniak, and Marszałek [9] was oriented on Convolutional Neural Networks working as a part of intelligent home infrastructure, in which actions of humans were evaluated from images. In Barra, Barra, Bisogni, De Marsico, and Nappi [10] was proposed web-shaped model, where pose detection was based on sampling comparisons. Important part of each training process is data augmentation. From this part we can get better data for training. When initial images are not well fitted for the model we can perform modification to improve the set. In Abayomi-Alli, Damaševičius, Wieczorek, and Woźniak [11] was proposed a model of based on principal resampling. The images were analyzed and for the key features repetionions were proposed. In Woźniak and Połap [12] was proposed a composition of neural networks with soft sets. The model was giving classes of detected objects from image, while comparisons and decisions were based on soft set classifier. The model proposed in Bin, Chen, Wei, Chen, Gao, and Sang [13] was developed by using Graph Convolutional Neural Networks. This complex structure was able to recognize variants of human body structure.

In this paper a model of a Lightweight Convolutional Neural Network is proposed. The model is focused on fast detection of human upper-body and faces to roughly count the amount of people in the desired area.

2. Neural Network Architecture

The final network model was developed with an intention to be as light-weight as possible, maintaining the high accuracy. In order to achieve this the number of layers, as well as, the synapse count was kept at a bare minimum. The final model can be seen in Fig. 1.

As presented, the developed CNN (Convolutional Neural Network) contains 3 main convolutional segments followed by max polling. In the first, and third segment the kernel size is set to 5x5. In the middle segment however the filter sizes are varying and are set to 7x7, 5x5 and 3x3 in order to extract as many features as possible. All convolutional layers are followed by the ReLU (Rectified Linear Unit) activation function.

The pooling layers were set to reduce the image by a factor of 2.

3. System Model

In presented detection model two neural networks were used for the final prediction. The first one was detecting 2 abstract classes:

- body,
- no-body.

And the second one:

- face,
- no-face.

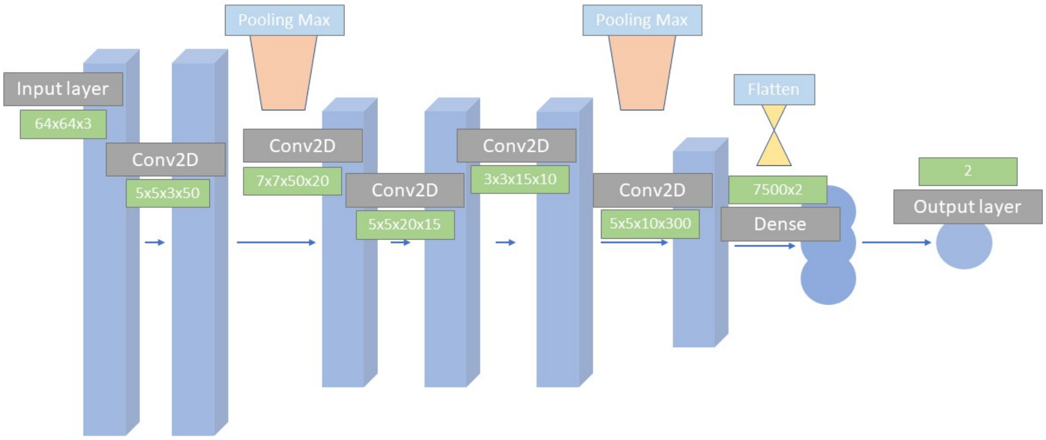


Figure 1: Lightweight CNN Architecture

This approach, despite the fact of being slower, have the advantage over the classical one-network models in terms of the final customization. Because these two main classes are separated we can fine-tune the detection system after the training and set different detection thresholds for bodies and faces which can lead to much more accurate people count predictions.

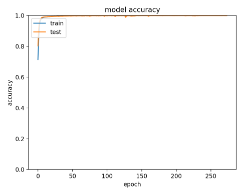
4. Model's performance

Presented neural network model has achieved a high accuracy of over 99.94% for both the face and the body detection. Final metrics are shown in Tab. 2 and the training plots can be found in Fig. 2. The final confusion matrices are in Fig. 3 and Fig. 4.
Used metrics are:

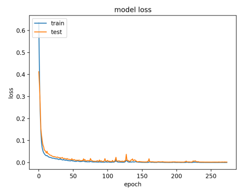
- Accuracy,
- Precision,
- Recall,
- F1,
- Specificity,
- FDR (False Discovery Rate),
- FPR (False Positive Rate),
- FNR (False Negative Rate),
- FOR (False Omission Rate),
- NPV (Negative Predictive Value).

Table 1
Most common machine learning metrics for the final model

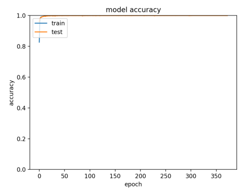
Type	Accuracy	Precision	Recall	F1	Specificity	FDR	FPR	FNR	FOR	NPV
Face	99.95%	99.95%	99.94%	2.0	99.94%	99.97%	0.06%	0.04%	0.09%	99.91%
Body	99.94%	99.94%	99.92%	2.0	99.96%	99.98%	0.04%	0.07%	0.13%	99.87%



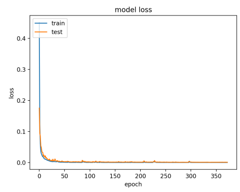
(a) Accuracy Face



(b) Loss Face

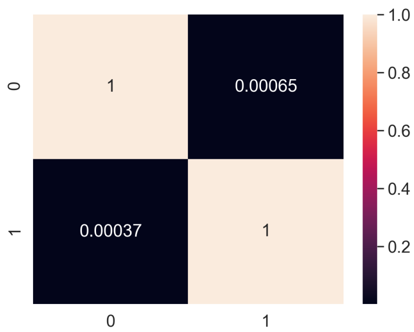


(c) Accuracy Body

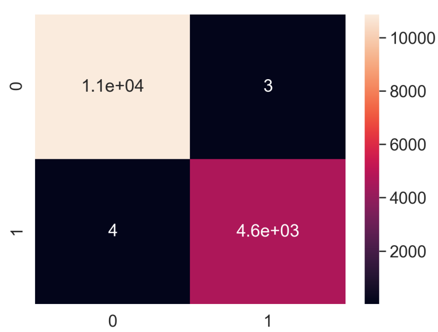


(d) Loss Body

Figure 2: Training metrics over epochs

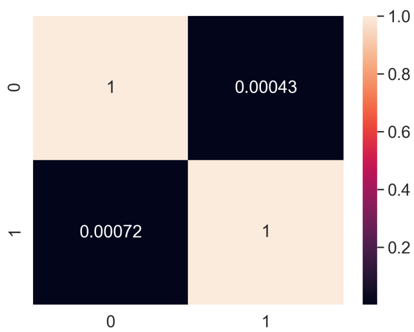


(a) Confusion Matrix Normalized

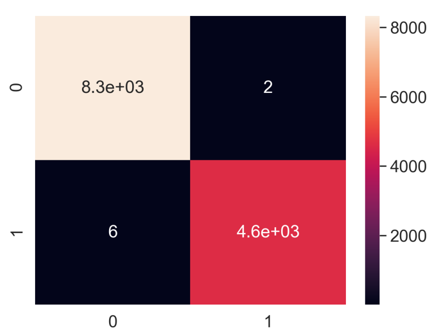


(b) Confusion Matrix

Figure 3: Confusion Matrix Face



(a) Confusion Matrix Normalized



(b) Confusion Matrix

Figure 4: Confusion Matrix Body

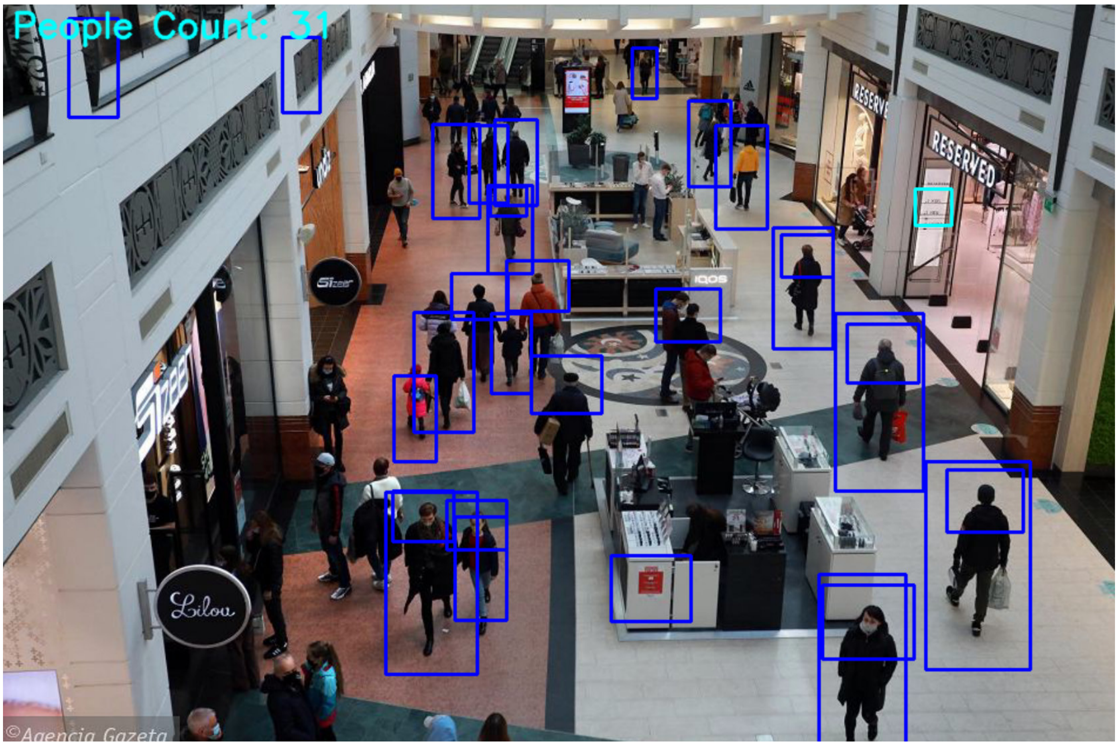


Figure 5: Example1

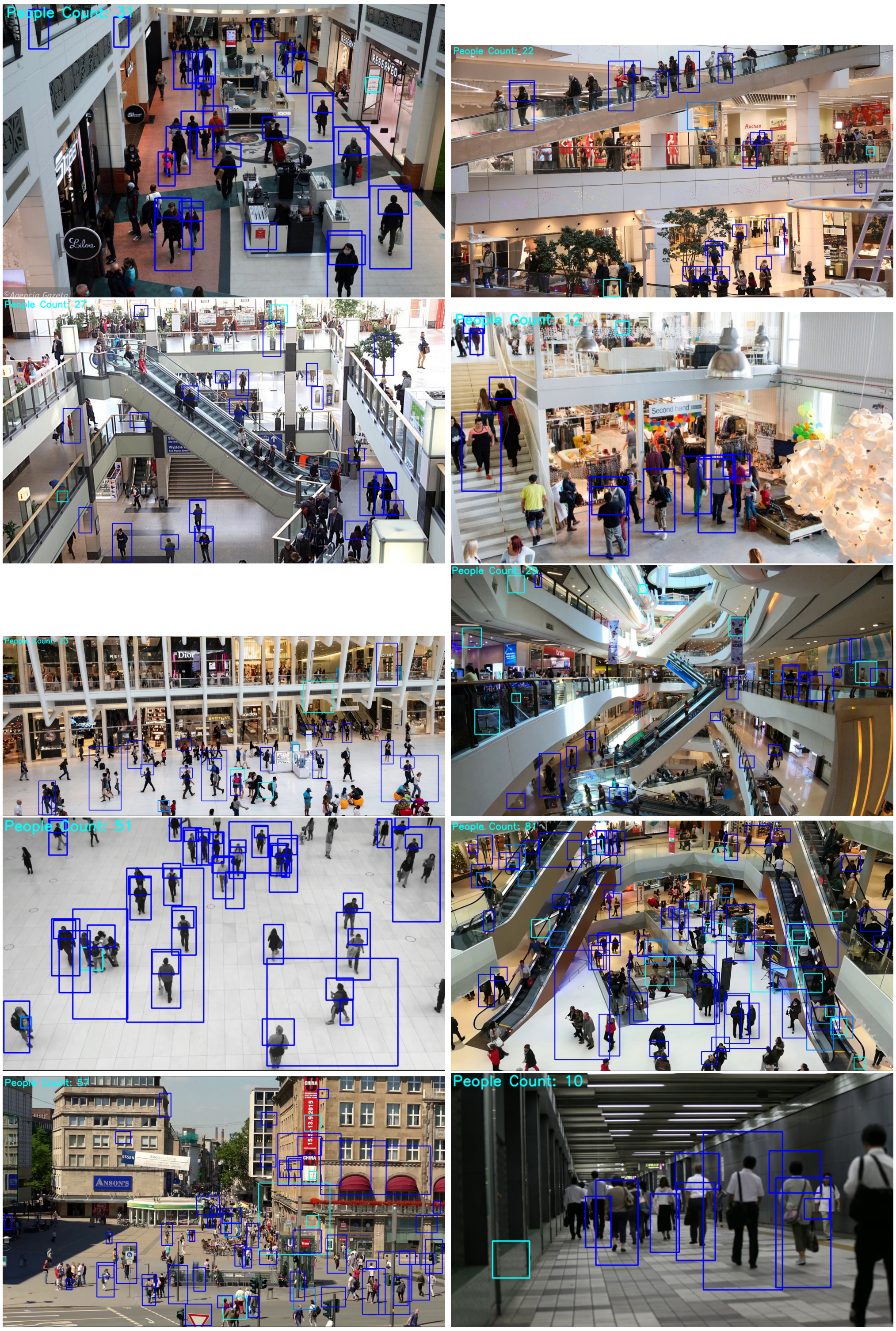


Figure 6: Examples of system in action

5. Results

The results of the final system are shown in Fig. 6. In all examples the system was set with the same, default parameters to present the accuracy "out of the box". Because of the dual-CNN model there is a possibility to better fit the parameters to the specific camera and view making the predictions even more accurate.

In Fig. 5 there is one of the examples of the system in action. As we can see in most cases the CNN correctly detects human silhouettes on the image and draws a bounding box around them. In some cases, however, especially on the dark background the system has a problem to correctly detect the body and the certainty is too low to recognize this batch as a human. In some other regions of the image (especially on the left top) there are some inverted examples, when batches of the image are falsely recognized as the body.
This problems are mostly due to the fact that the model was made to be more general and to work in most common scenarios. To improve the classification and to better fit the specific camera, the fine-tuning of detection parameters, as well as, some post-training would be necessary.

Table 2
Comparison of accuracy with other detection models

Detection Model	Accuracy
This work	99.94%
Hsu, Abdel-Mottaleb, and Jain [14]	99.12%
Wu, Yin, Wang, and Xu [15]	98.12%
Cuimei, Zhiliang, Nan, and Jianhua [16]	98.01%
Chi, Zhang, Xing, Lei, Li, and Zou [17]	96.4%
Zhang, Chi, Lei, and Li [18]	96.2%
Rowley, Baluja, and Kanade [19]	90.3%

6. Discussion

Because of the global Covid-19 pandemic many governments prepared restrictions to the amount of people per square meter in public places such as shops and shopping malls. This approach, except of allowing shops to be run at all, created a huge logistical problem of counting the amount of people inside the desired area. To follow this restrictions every shop owner had to create some way of dealing with this problem. In smaller, regional shops the problem is almost non visible due to the fact that there are only allowed 2-3 people for entire building and the shop owner can easily count the people manually. In bigger shops however, where limits are much bigger, the problem starts to occur because the shop personnel has to serve the clients and have no time and possibility to count every new guest. Some owners deal with it by

hiring additional personnel just for this purpose, however it is very expensive and especially in times of global pandemic, when the income is much smaller, this method is far from optimal. Even worse situation is in much bigger places such as shopping malls and other big, public places. Because there are many entrances and exits the manual counting even with additional personnel would be near impossible from the financial and logistical point of view, so in most places government restrictions are violated.
Presented system was made to address these problems with a minimal financial footprint, allowing shop owners to efficiently control the number of people in the monitored area and because of that respect the official restrictions and actively fight with the pandemic outbreak.

7. Conclusion

As we can see presented Lightweight-CNN solution allows the user to quickly count people located in a store, in a shopping mall or in any open public space with small resource, money and power consumption. Because of the high effectiveness of face and body detection the presented system counts the amount of people with high accuracy and due to the light-weighted architecture it does it in real-time even on high resolution cameras. What's more the small architecture allows the system to run even on smaller and less powerful devices such as laptops or, for example, Raspberry Pie. Because of that it reduces the need of more expensive and much more power hungry PCs making it more environment friendly.

Acknowledgments

Author would like to acknowledge contribution to this research from the Rector of the Silesian University of Technology, Gliwice, Poland, under program "Initiative of Excellence—Research University" grant no. 08/IDUB/2019/84.

References

[1] S. Li, Global face pose detection based on an improved pso-svm method, in: Proceedings of the 2020 International Conference on Aviation Safety and Information Technology, 2020, pp. 549–553.
[2] C. Chen, T. Wang, D. Li, J. Hong, Repetitive assembly action recognition based on object detection and pose estimation, Journal of Manufacturing Systems 55 (2020) 325–333.
[3] W. Wang, X. Chen, S. Zheng, H. Li, Fast head pose estimation via rotation-adaptive facial landmark detection for video edge computation, IEEE Access 8 (2020) 45023–45032.
[4] Y. Zhao, Z. Luo, C. Quan, D. Liu, G. Wang, Cluster-wise learning network for multi-person pose estimation, Pattern Recognition 98 (2020) 107074.
[5] M. Woźniak, M. Wieczorek, J. Siłka, D. Połap, Body pose prediction based on motion sensor data and recurrent neural network, IEEE Transactions on Industrial Informatics 17 (2020) 2101–2111.
[6] K. Yun, J. Park, J. Cho, Robust human pose estimation for rotation via self-supervised learning, IEEE Access 8 (2020) 32502–32517.

[7] B. Huang, R. Chen, Q. Zhou, W. Xu, Eye landmarks detection via weakly supervised learning, Pattern Recognition 98 (2020) 107076.

[8] A. Nadeem, A. Jalal, K. Kim, Human actions tracking and recognition based on body parts detection via artificial neural network, in: 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), IEEE, 2020, pp. 1–6.

[9] A. Winnicka, K. Kęsik, D. Połap, M. Woźniak, Z. Marszałek, A multi-agent gamification system for managing smart homes, Sensors 19 (2019) 1249.

[10] P. Barra, S. Barra, C. Bisogni, M. De Marsico, M. Nappi, Web-shaped model for head pose estimation: An approach for best exemplar selection, IEEE Transactions on Image Processing 29 (2020) 5457–5468.

[11] O. O. Abayomi-Alli, R. Damaševičius, M. Wieczorek, M. Woźniak, Data augmentation using principal component resampling for image recognition by deep learning, in: International Conference on Artificial Intelligence and Soft Computing, Springer, 2020, pp. 39–48.

[12] M. Woźniak, D. Połap, Soft trees with neural components as image-processing technique for archeological excavations, Personal and Ubiquitous Computing 24 (2020) 363–375.

[13] Y. Bin, Z.-M. Chen, X.-S. Wei, X. Chen, C. Gao, N. Sang, Structure-aware human pose estimation with graph convolutional networks, Pattern Recognition 106 (2020) 107410.

[14] R.-L. Hsu, M. Abdel-Mottaleb, A. K. Jain, Face detection in color images, IEEE transactions on pattern analysis and machine intelligence 24 (2002) 696–706.

[15] W. Wu, Y. Yin, X. Wang, D. Xu, Face detection with different scales based on faster r-cnn, IEEE Transactions on Cybernetics 49 (2019) 4017–4028. doi:10.1109/TCYB.2018.2859482.

[16] L. Cuimei, Q. Zhiliang, J. Nan, W. Jianhua, Human face detection algorithm via haar cascade classifier combined with three additional classifiers, in: 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), IEEE, 2017, pp. 483–487.

[17] C. Chi, S. Zhang, J. Xing, Z. Lei, S. Z. Li, X. Zou, Selective refinement network for high performance face detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 8231–8238.

[18] S. Zhang, C. Chi, Z. Lei, S. Z. Li, Refineface: Refinement neural network for high performance face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (2020) 1–1. doi:10.1109/TPAMI.2020.2997456.

[19] H. A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Transactions on pattern analysis and machine intelligence 20 (1998) 23–38.