1. Najwan WAISI[1], 2. Nawal ABDULLAH[2], 3. Mohammed GHAZAL[3]

Northern Technical University (1,2,3) Iraq

ORCID: 1. 0000-0003-4259-1294, 2. 0000-0003-1217-4959, 3. 0000-0003-3159-4983

doi:10.15199/48.2021.09.18

The Automatic Detection of Underage Troopers from Live-Videos Based on Deep Learning

***Abstract**. Military service is undoubtedly among the most profound forms of service to the nation. With military service young people can develop qualities of discipline within them, but nobody should be forced to serve, and especially young children. A real-time Child Troopers detection surveillance system is built to overcome these bad acts, based on Convolutional Neural Networks (CNNs). This method is focused on the automatic face, age, and weapon detection. The proposed detection and identification system consist of many steps of process: starting with, a pre-trained deep learning model based on SSD-ResNet network to perform face detection operation. Then, an age estimation using VGG-Face model is performed, finally, a weapon detection based on MobileNetV2-SSD pretrained model. The results of these steps are combined to look for children under 18 years old with guns in the images. These models have been selected because of there fast and accurate in infering to integrate network for detecting and identifying children with weapons in images. The experimental result using global datasets of various images for faces and weapons showed that the use of this method enhances the accuracy level of detection.*

***Streszczenie.** Dzięki służbie wojskowej młodzi ludzie mogą rozwinąć w sobie cechy dyscypliny, ale nikt nie powinien być zmuszany do służby, a zwłaszcza małe dzieci. Zaproponowano jest system nadzoru wykrywający w czasie rzeczywistym Child Troopers, oparty na Convolutional Neural Networks (CNN). Ta metoda skupia się na automatycznym wykrywaniu twarzy, wieku i broni. Proponowany system detekcji i identyfikacji składa się z wielu etapów procesu: zaczynając od wstępnie wytrenowanego modelu głębokiego uczenia opartego na sieci SSD-ResNet do wykonywania operacji wykrywania twarzy. Następnie przeprowadzana jest estymacja wieku za pomocą modelu VGG-Face, a na koniec detekcja broni w oparciu o wstępnie wytrenowany model MobileNetV2-SSD. Wyniki tych kroków są łączone w celu wyszukania na zdjęciach dzieci poniżej 18 roku życia z bronią. Modele te zostały wybrane ze względu na szybkie i dokładne wnioskowanie do integracji sieci do wykrywania i identyfikacji dzieci z bronią na obrazach. Wyniki eksperymentalne wykorzystujące globalne zbiory danych różnych obrazów twarzy i broni wykazały, że zastosowanie tej metody zwiększa poziom dokładności wykrywania. (**Automatyczne wykrywanie nieletnich żołnierzy z wideo na żywo w oparciu o metodę Deep Learning**)*

Keywords: Deep Learning, Convolutional Neural Networks (CNN), Child Troopers, VGG-Face, SSD-ResNet-101, MobileNetV2-SSD.
Słowa kluczowe: Deep Learning, wykrywanie I identyfikacja twarzy, wykrywanie nieletnich.

Introduction

Recently, A tremendous interest in deep learning has become one of the main topics in the field of machine learning and artificial intelligence [1]. The goal of this paper is Child Troopers detection from a live video stream because in various countries of the world proliferation of individual weapons in the hands of civilians in general and in children under the age of eighteen in particular, has been increased. Community police in our country (Iraq), tried to impose their control by gathering places by deploying security men among the crowds. This method is ineffective and dangerous it caused clashes by weapons. In this work, we suggest installing surveillance cameras in crowds' places for observation purposes, to detect Child Troopers.

The human face carries important information about identity, gender, age, ethnicity, and emotion [2]. which attracted the interest of researchers in the domain of computer vision and deep learning to developing efficient algorithms to detect and identify faces in addition to age estimation [3,4]. The system is based on deep learning, which is provided by images from surveillance camera streams, whenever, a child under 18 years old with a gun has detected the system sends an alert message to the supervisor, which will help to impose control and speedup crime detection before occurring.

The proposed system is a part of computer vision field, a common doubt which most of researchers have is what is the difference between image classification, object detection, and image segmentation [5,6]. Object detection performs object categories recognition and finds the location of each object in the image then surrounds it by a boundary box [7,8]. In the proposed system we focused on the use of the main efforts for object detection based on deep learning technology. The preferred algorithm for object detection must have a robust understanding of semantic cues besides the spatial information about the image [9].

In the proposed approach, a deep convolutional neural network based on SSD-ResNet-101 has been used for face detection [10], VGG-Face for age estimation, and MobileNetV2-SSD for weapon detection architectures [11,12], as shown in Figure 1 a pre-trained model for child troopers detection task illustrated.

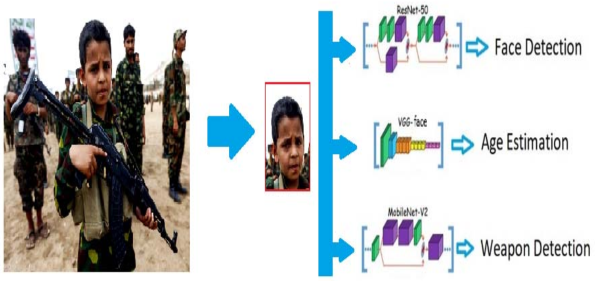

Fig. 1. The proposed system for detection a Child Troopers under 18 years.

Moreover, the classification of faces and age estimation is based on the Adience dataset [13]. Which is a collection of face photos used for age and gender recognition studies.The dataset which contains 26.580 photos with approximately 2.284 subjects and 8 labels for different age groups (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-). Moreover, the weapon datasets for both the sliding window and region proposals approaches have been selected for weapon detection [14].

Many related studies have been published on each subsystem of our project separately, which reinforces the novelty of our work. In the domain of human face detection, the studies divided into two class: the traditional method such as face texture detection [15] which fails to give good results for Face Liveness Detection under unconstrained environment, and the deep learning method which success in spoof face detection in live video using local receptive fields (LRF)-ELM and CNN as proposed in [16]. our approach is based on a pre-trained SSD-ResNet-101 model which enhances the efficiency of face detection under challenging illumination conditions. The earlier studies for age estimation from face images used the size and face proportions for detection [17], these methods are restricted on the estimation of young ages because of the nature of the human head. Later, age pattern subspace (AGES) [18], and morphable model [19] are proposed. All these methods are traditional, our proposed method is based on the VGG-Face model to perform age estimation the network trained on a very large dataset of face images which enhances the efficiency of the estimation system. A study of child troopers identification based on age and military fatigue detection has been proposed in [20] as an application of computer vision, this study did not include weapon detection, which represents an important factor to detect the child troopers. The studies of weapon detection systems involved two methods: the first is based on metal detection using X-ray or millimetric wave images [21], which is unable to detect non-metallic weapons. while the second is based on deep learning as we have proposed in our handguns detection from live video based on MobileNetV3-SSDlite pre-trained deep learning model [22], which enhance the accuracy level in identifying the real-time weapon detection. The Child Troopers detection system needs a fast and accurate combination of face detection, age estimation, and weapon detection in one system, as we have proposed in the model below.

In section 2 of this work, we described the proposed model, involving the pipeline of building the system model. The experimental model training has been proposed in Section 3, and section 4 presents the results that evaluate the performance of the adopted method. Finally, the cconclusions are reviewed in Section 5.

The proposed model

Our proposed method for the Child Troopers detection system follows a pipeline structure, consists of many steps of detection then combines them to get the required aim of this project. The system trained using a collection of face images for face detection and age estimation operations, while a handgun dataset has been used to detect weapons in the images. The following subsections describe the building structure for the proposed system.

1. Face Detection

SSD-ResNet model is selected in this paper to detect faces in various environments [10]. The ResNet [23,24] basic structure is illustrated in Figure 2.

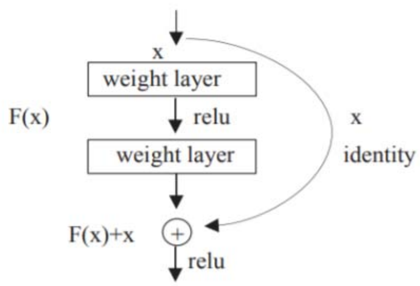

Fig. 2. ResNet: a building block [10].

The network structured in two layers, represented by the expression:

(1) $$F = W_{2\alpha}(W_1 x)$$

where α represents the non-linear function ReLU. Then obtained the output Y, through a shortcut and second ReLU.

(2) $$Y = F(x, \{w_i\}) + x$$

where x represents the input of the network, then H(x) is the learns feature, hoping that it can learn the residual.

(3) $$F(x) = H(x) - X$$

where F(x)+x represents the original learning feature.

In a term of a stack, when the residual is 0, only identity mapping is carried out on the stack, at least the network performance will not decline. The residual enables the stack to learn new features based on the input feature, to obtain better performance. Its expressed by formula as follows:

(4) $$Y = F_{(x\{wi\})} + w_s x$$

The ResNet-101 is adopted to improve the network accuracy, In this work The Single Shot Detector (SSD) has been employed as the basic structure of the CNN model and replace the inside VGG16 with ResNet-101 model. Figure 3 shows the last 3 layers use original SSD layers to get the feature map for face detection. We have employed the SSD-ResNet-101 model for face detection to enclosing the faces in images by a boundary box.

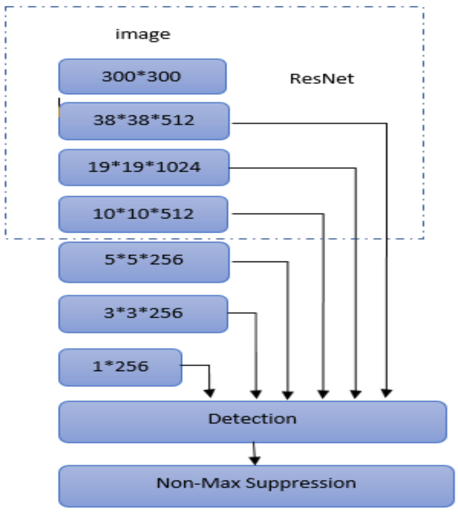

Fig. 3. Face Detection model based on SSD-ResNet-101 [10].

2. Age Estimation

The VGG-Face model is applied as a facial feature extractor for age estimation purposes from the face images [25]. The model consists of eleven layers, 8 convolutional layers, and 3 fully connected layers. in addition to, a rectification layer after each convolutional layer, followed by a max pool layer operate at the end of each convolutional block [26], the architecture of the VGG-Face model is shown in Figure 4.

In this work, the VGG-Face model is trained with one of the newest datasets which are designed for age and gender recognition. The Adience benchmark dataset is unfiltered face collection images. It attempts to capture all the variations in appearance, noise, pose, lighting, and more.

3. Weapon Detection

The handgun detection is a subsystem of the Child Troopers detection system, it's used SSD based on the MobileNetV2 model [12,27]. MobileNetV2 represents an extended version of the MobileNet model, where the Inverted Residuals with Linear Bottlenecks layer module has been added in this version.

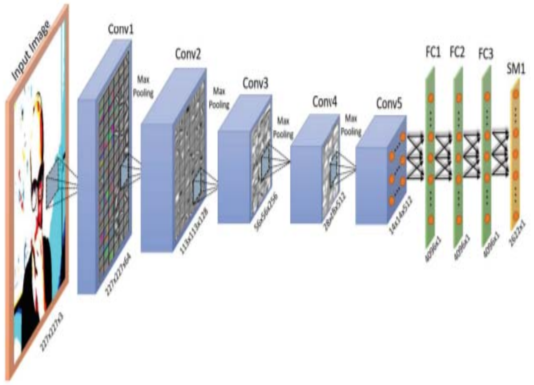

Fig. 4. The VGG-Face architecture.

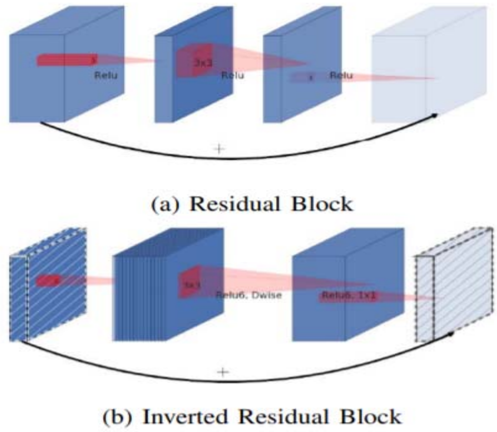

Fig. 5. Inverted Residual Block [27].

Table 1. The Adience Dataset

	LABELS IN YEARS								TOTAL
GENDER	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	
FEMALE	682	1234	1360	919	2589	1056	433	427	9411
MALE	745	928	934	734	2308	1294	392	442	8192

Inverted residuals inherit the concept of adding the first and the last layer of a convolution block with a skip connection from residual networks [28]. Moreover, it rearranged the squeeze-expand convolution operations opposite to that of a residual block. Both residual convolution block is illustrated in Figure 5, as described in the original residual networks implementation, and inverted residual convolution block. SSD model is used to produce a bounding boxes for the existence object and confidence scores in a single forward pass.

Experimental model training

The system training experiments have been achieved by using a workstation equipped with an Intel® Core™ i9-9880H CPU @ 4.8GHZ, NVIDIA® Quadro® Graphics Processing Unit (GPU) of 384 CUDA cores and 16GB memory RAM, with Ubuntu 20.04.1 LTS operating system. The input images are rescaled to 300*300 pixels to fit the system requirements. The stochastic gradient descent optimization technique is used by setting the mini-batches size to 256 with a momentum value of 0.9. This method is used to find the connected layer's parameters that minimize the predication of the softmax, log, and loss functions of the Child Troopers detection system. To evaluate the performance and efficiency of the proposed system, many experiments have been performed based on tuning parameters like epochs, learning rate. A list of age labels and the number of images per label for both genders' male / female are listed in Table 1 of the Adience dataset which is used for age estimation.

Fig. 6. Samples of Weapon Dataset.

Two datasets for weapons detection systems have been used to evaluate the efficiency in this work: The sliding window approach dataset, which includes 102 classes and around 9261 images, and the region proposals approach dataset, which includes 608 images of handguns. Some of the handguns images from both datasets are shown in Figure 6.

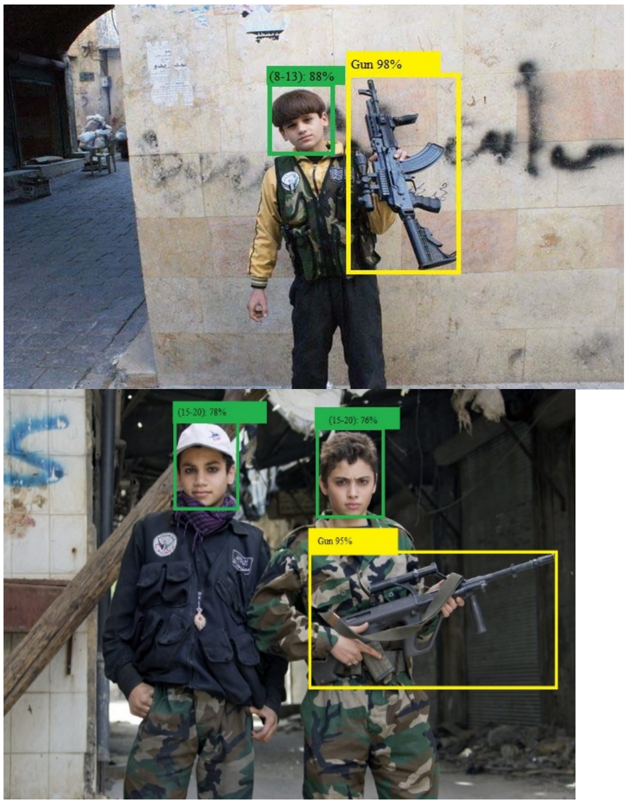

Fig 7. Some of the test result images of the Child Troopers detection system.

Results and discussion

To Detect and identify the Child Troopers in live-video three-stage detection subsystems combined in a pipeline structure. First, for extract the human face area in the image, the SSD-ResNet-101 network was fine-tuned with 70k iterations, thus the trained network reaches 91.5%

training accuracy rate which represents a very promising result. The experimental results for age estimation based on the VGG-Face model showed that 0-2 years old have the highest accuracy 92.20% because infants images contain special features enable the system to recognize this age group easily, while the 15-20 years old ages classified with lower accuracy 87.30%. The final detection subsystem is for weapon detection based on the MobileNetV2-SSD model achieved a 95% training accuracy rate. The results of these steps are combined and used to determine if a child with a weapon in images detected as shown in Figure 7. The suggested method is suitable for the Child Troopers detection system as it delivers a perfect correspondence between prediction speed and accuracy. The speed achieved by the experiments is ~25–28 msec. The experiments have been performed in real-time detection in live videos and images with a minimum level of false detection less than 10%.

CONCLUSION

The proposed Child Troopers detection system is an applicable system that helps to detect children under 18 years old with weapons in live video. Our work shows that the use of deep learning networks (SSD-ResNet-101, VGG-Face, and MobileNetV2-SSD) has big potential to perform the desired tasks, besides the use of the appropriate datasets for training the system. Moreover, the system can be used in various applications, such as security systems, and monitor social media videos.

***Authors**: Najwan Zuhair Waisi obtained her M.Sc. degree from Computer Science Department, Mosul University, Mosul, Iraq in 2014, E-mail: najwan.tuhafi@ntu.edu.iq; Nawal Younis Abdullah obtained her M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2014, E-mail: nawal_younis@ntu.edu.iq; Mohammed Talal Ghazal obtained his M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2016, E-mail: mohammed.ghazal@ntu.edu.iq;;*

REFERENCES

[1] J. Chmielińska and J. Jakubowski, "Application of convolutional neural network to the problem of detecting select ed symptoms of driver fatigue," Przegląd Elektrotechniczny, vol. 93, no. 10, pp. 6-10, 2017.

[2] J. Chmielińska and J. Jakubowski, "Face recognition based on deep learning techniques and image fusion," Przegląd Elektrotechniczny, vol. 95, no. 11, pp. 150-154, 2019.

[3] A. Othmani, A. R. Taleb, H. Abdelkawy, and A. Hadid, "Age estimation from faces using deep learning: A comparative analysis," Computer Vision and Image Understanding, p. 102961, 2020.

[4] M. T. Ghazal and K. Abdullah, "Face recognition based on curvelets, invariant moments features and SVM," Telkomnika, vol. 18, no. 2, pp. 733-739, 2020.

[5] W. Qiong, L. Zhang, L. Yan, and K. Kpalma, "Overview of deep-learning based methods for salient object detection in videos," *Pattern Recognition,* p. 107340, 2020.

[6] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232, 2019.

[7] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," ACM Computing Surveys (CSUR), vol. 52, no. 4, pp. 1-35, 2019.

[8] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer Vision and Image Understanding,* vol. 189, p. 102805, 2019.

[9] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing,* 2020.

[10] X. Lu, X. Kang, S. Nishide, and F. Ren, "Object detection based on SSD-ResNet," in 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), 2019: IEEE, pp. 89-92.

[11] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, "Deep convolutional neural network for age estimation based on VGG-face model," arXiv preprint arXiv:1709.01664, 2017.

[12] N. S. Sanjay and A. Ahmadinia, "MobileNet-Tiny: A Deep Neural Network-Based Real-Time Object Detection for Rasberry Pi," in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019: IEEE, pp. 647-652.

[13] K. Zhang et al., "Fine-grained age estimation in the wild with attention LSTM networks," IEEE Transactions on Circuits and Systems for Video Technology, 2019.

[14] Y. Elmir, S. A. Laouar, and L. Hamdaoui, "Deep Learning for Automatic Detection of Handguns in Video Sequences," in JERI, 2019.

[15] S. L. Fernandes and G. J. Bala, "Developing a novel technique for face liveness detection," Phys. Procedia, vol. 78, pp. 241-247, 2016.

[16] Y. Akbulut, A. Şengür, Ü. Budak, and S. Ekici, "Deep learning based face liveness detection in videos," in 2017 international artificial intelligence and data processing symposium (IDAP), 2017: IEEE, pp.1-4.

[17] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, vol. 1: IEEE, pp. 387-394.

[18] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 29, no. 12, pp. 2234-2240, 2007.

[19] K. Scherbaum, M. Sunkel, H. P. Seidel, and V. Blanz, "Prediction of Individual NonLinear Aging Trajectories of Faces," in Computer Graphics Forum, 2007, pp. 285-294.

[20] https://www.pyimagesearch.com/2020/05/11/an-ethical-application-of-computer-vision-and-deep-learning-identifying-child-soldiers-through-automatic-age-and-military-fatigue-detection.

[21] Flitton G, Breckon TP, Megherbi N. " A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery". Pattern Recognition. 2013;46(9):2420-36.

[22] M. T. Ghazal, N. Waisi and N. Abdullah, " The Detection of Handguns from Live-Video in Real-Time Based on Deep Learning," Telkomnika, vol. 18, no. 6, pp. 3030-3036, 2020.

[23] C. Zhenzhou and D. Pengcheng, "Face recognition based on improved residual neural network," in 2019 Chinese Control And Decision Conference (CCDC), 2019: IEEE, pp. 4626-4629.

[24] H. Ling, J. Wu, L. Wu, J. Huang, J. Chen, and P. Li, "Self residual attention network for deep face recognition," IEEE Access, vol. 7, pp. 55159-55168, 2019.

[25] F. Dornaika, S. E. Bekhouche, and I. Arganda-Carreras, "Robust regression with deep CNNs for facial age estimation: An empirical study," *Expert Systems with Applications,* vol. 141, p. 112942, 2020.

[26] M. K. Islam and S. U. Habiba, "Human Age Estimation and Gender Classification Using Deep Convolutional Neural Network," in *International Conference on Cyber Security and Computer Science*, 2020: Springer, pp. 503-514.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510-4520.

[28] L. Hu and Q. Ge, "Automatic facial expression recognition based on MobileNetV2 in Real-time," in Journal of Physics: Conference Series, 2020, vol. 1549, no. 2: IOP Publishing, p. 022136.