Convolutional Neural Network Based Hand Gesture Recognition in Sophisticated Background for Humanoid Robot Control

Ali Yildiz
Department of Mechatronics Engineering,
Bursa Technical University, Turkey
aliyildiz2588@gmail.com

Nurettin Gökhan Adar
Department of Mechatronics Engineering,
Bursa Technical University, Turkey
gokhan.adar@btu.edu.tr

Ahmet Mert
Department of Mechatronics Engineering,
Bursa Technical University, Turkey
ahmet.mert@btu.edu.tr

Abstract: *Hand gesture recognition is a preferred way for human-robot interactions. Conventional approaches are generally based on image processing and recognition of hand poses with simple backgrounds. In this paper, we propose deep learning models, and humanoid robot integration for offline and online (real-time) recognition and control using hand gestures. One thousand and two hundred of hand images belonging to four participants are collected to construct the hand gesture database. Five class (forward, backward, right, left and stop) images in six sophisticated backgrounds with different illumination levels are obtained for four participants, and then one participant's images are kept as testing data. A lightweight Convolutional Neural Network (CNN), and transfer learning techniques using VGG16, and Mobilenetv2 are performed on this database to evaluate user independent performance of the hand gesture system. After offline training, real-time implementation is designed using a mobile phone (Wi-Fi and camera), Wi-Fi router, computer with embedded deep learning algorithms, and NAO humanoid robot. Streamed video by the mobile phone is processed and recognized using the proposed deep algorithm in the computer, and then command is transferred to robot via TCP/IP protocol. Thus, the NAO humanoid robot control using hand gesture in RGB and HSV color spaces is evaluated in sophisticated background, and the implementation of the system is presented. In our simulations, 95% and 100% accuracy rates are yielded for the lightweight CNN, and transfer learning, respectively.*

Keywords: *Hand gesture recognition, human-robot interaction, robot control, convolutional neural network, deep learning.*

Received October 26, 2021; accepted January 5, 2023
https://doi.org/10.34028/iajit/20/3/9

1. Introduction

Robots' usage, which are an important part of production in Industry 4.0, are increasing continuously. Robots are expected to fulfill the task given in workplace with or without human. The fact that robots are with people in their workplace causes the human-robot interaction. Robots are desired to be used more flexibly and comfortably. Therefore, robots are expected to be compliant and collaborative.

Humanoid robots constitute the top level of robots which have many types. Humanoid robots, which are like human skeletal structure, can move with their arms and to hold and put the objects with their arms. The programming of these robots should be done by trained people, and it is time consuming. Studies are carried out to reduce the programming process of robots and to increase human-robot interaction. The camera is one of the devices that provide this interaction. The data obtained from the camera is used to control of the robot using some methods such as image processing, Convolutional Neural Network (CNN), Fuzzy logic etc., Object recognition and classification are the main studies on this subject [8].

In literature, studies on industrial robots to increase human-robot interaction continue. Static gesture or hand posture recognition is based on shape and orientation of the hand without trajectory planning [18] in the fields of Human-Robot Interaction (HRI) [15] and medical or virtual reality procedures [10, 12]. Most of the studies existing in the literature are mainly focused on image and video processing methods with machine learning algorithms [1, 3, 20, 24]. In this pattern recognition task, Gabor filters were applied to extract features of the five angles of the hand postures [9]. Thus, performance evaluation of the filter was investigated under varying illumination. In [4], the bag of the Scale-Invariant Feature Transform (SIFT) based features with multi-class Support Vector Machine (SVM) were combined for real-time gesture recognition with accuracy rate up to 96.23%. Complex background problem of the hand image was investigated in [22]. Shape and color descriptors in HSI, and YCbCr spaces were classified using SVM for the data set including 2750 hand postures and 2000 background images belonging to 40 subjects with various hand sizes and ethnicities. This task yielded accuracy level of 94.32%.

Artificial Neural Network (ANN) based systems were also successfully performed for hand gesture recognition in the literature. Fingertip positions and Gabor transform features were applied as an input to ANN (specifically local linear mapping) in order to obtain 3-D model [19]. In [13], sensor signals from 5DT data glove were classified using ANN and genetic algorithm. Recently, deep learning algorithms have been implemented to many kinds of engineering problems, especially to recognition tasks. Due to eliminating pre-processing stage, Deep Neural Network (DNN) based hand gesture recognition were studied by researchers [21]. CNNs have a superior success rate on image classification. Alexnet [11], GoogleNet [27], and MobileNetv2 [25] are the well-known topologies of the CNNs. In [7], a proposed CNN architecture (with 3 convolutional layer and max pooling layers followed 3-fully connected layers in depth) were used to control of flight of Unmanned Aerial Vehicle (UAV). Nine hand gesture were classified, and accuracy up to 97% was reported. The combination of the CNN with SVM was performed on hand images from Kinect camera [14], increased the recognition rate up to 98.52%. Weakly-supervised fusing shallow/deep image properties were combined with SVM and CNN to assign each hand gesture to a particular type. This study covers regional features and deep features together to yield 87.87% of accuracy level for 10-class task. 3D CNN architectures were also studied using the images from Kinect camera. The RGB images with Distance (RGB-D) was processed using 3D CNN, and the accuracy rates were yielded higher than 90.00 % for each gesture [26].

Human-robot interaction has been tried to be achieved by using human hand and skeleton in real time. 3D skeleton extraction was performed with Kinect V2 camera while a hand data set was created with the 10 volunteers and a hand gesture was implemented with Inception V3 CNN. 98.9% success was achieved with trained CNN which was applied and tested in the robotic manipulator (BAZAR robotic system) [16]. In another study using Bazaar robotic system, human skeleton extraction and hand gesture detection using a CNN was also implemented [17]. Many different studies such as kinematics and dynamics analysis, image processing, speech recognition etc. are carried out with the NAO humanoid robot. The NAO humanoid robot can be used to assist elderly people for their daily needs, and young learners [12]. The tele-manipulation of the NAO humanoid robot was carried out with Leap Motion Sensor which is a device to recognize and track hands, fingers and finger-like tools [29]. Industrial robot with six degrees of freedom was controlled by using hand position and orientation information obtained from the Leap Motion Sensor [28]. NAO humanoid robot’s pick and place task was performed with just two hand gestures which are open or close hand poses using SURF algorithm [6]. In addition to recognition tasks using hand images, bio-signal based gesture recognition studies are studied. Surface Electromyography (sEMG) signals were mapped into 2D space using wavelet transform and then classified into 12, and 6 positions [2, 5].

Firstly, in this paper, we collected 1200 hand images of 5 class gestures from 4 participants with 6 complex backgrounds. With the help of this database, a CNN based hand gesture recognition is evaluated. Transfer learning approach using the models of the VGG16, and MobileNetv2 are adopted to our task to compare classification. Unknown participant's (testing) hand postures on 6 complex backgrounds are applied to these models to check robustness of the proposed system. After training phase, real-time classification of the hand gesture using a mobile phone's camera to control NAO humanoid robot.

2. Hand Gesture Database

Image acquisition is the first step of the study. For sophisticated background benchmarking, the hand gestures have been taken from various conditions. We collected 1200 images of different four participants to train and test CNN models for performance evaluation for hand gesture recognition.

5-types of hand postures namely, forward, backward, right, left and stop actions were taken by camera from different angles (30 degree spacing) on 6-complex background (wall, furniture, flower, poster, corridor and garden) with different illumination levels (lights are on/off, sunny or not). The scheme of the image collection is shown in Figure 1.

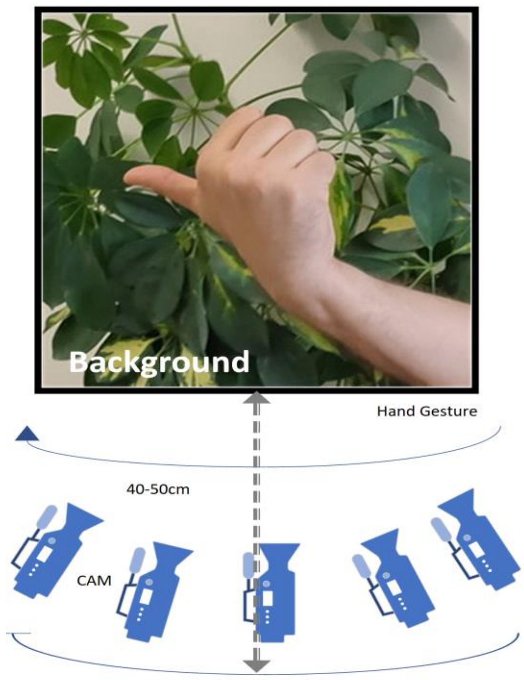

Figure 1. Hand gesture database collection.

Thus, 300 hand images of a participant were acquired. In other words, a participant's 60 photos in each posture class were taken in 6 different background and illumination levels. The detailed information of the database is demonstrated as thumbnail in Figure 2. In addition, three participants' 900 images are used to train CNN models, and the resting 300 images belonging the other participant are retained for testing purpose.

Finally, these are applied to CNN models in RGB and HSV color spaces to evaluate performance on the user independent recognition scenario with complicated backgrounds.

A lightweight CNN, VGG16, and MobileNetV2 structures have modified to 5-class output with pre-trained coefficients, and the images are resized to CNN inputs as 224×224×3 in RGB and HSV for training and testing.

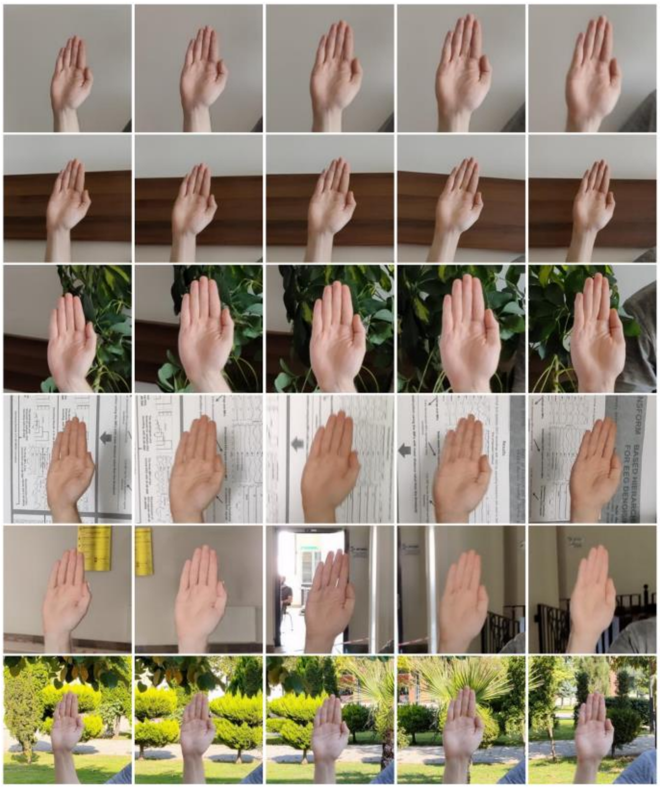

Figure 2. The thumbnails of the collected stop gesture with six backgrounds (belonging to a participant when lights are on.

3. Convolutional Neural Network and Transfer Learning

A CNN structure consists of convolutional filters, pooling layers, and activation blocks, which are connected to Neural Network (NN) layers for classification tasks. The learnable filter approach in convolutional layers is the main advantage. These can be also called learnable feature extraction method added to NN layers. In feature extraction, 2D convolution processes with fixed size learnable kernels are applied to the input images, and then max or average pooling algorithm is for lower dimensional space. Kernel size is preferred from 3×3 to 11×11 up to around a few hundred filters. These layers are also connected using an activation function preferably rectified linear unit (ReLu). It is a linear activation function firstly used in AlexNet CNN model, outputting the input directly when it is positive, otherwise, zero.

In AlexNet topology, input image size is 227×2×3 and there are 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. In each convolutional layer, 96, 256, 384, 383, and 256 kernels are existed with stride, and ReLU. VGG consists of 16 and 19 layers called VGG-16, and VGG-19 with input image size of 224×224×3. In VGG-16 architecture, 12 convolutional layers with from 64 to 512 3×3 kernels while VGG-19 has 14 convolutional layers. MobileNetV2 has been designed to be effective and fast computing in mobile devices without losing classification capability. Inverted residual connections are developed between the bottleneck layers. It contains convolutional layer with 32 filters and residual bottleneck layers. 60, 140, and 3.4 million parameters will be trained for AlexNet, VGG, and MobileNet V2, respectively. Depending on GPU performance, it requires weeks to train the network. For this reason, transfer learning approach is developed using the trained coefficients of the network (learnables in convolutional blocks and FCS).

4. Proposed Recognition Models and Robot Integration

4.1. Proposed Deep Learning Models

After collecting 1200 images of four participants' hang gestures, 300 images @ 224×224×3 in RGB and HSV format of a participant are retained as testing and the others for training. A lightweight CNN model, and transfer learning approach on VGG16, and Mobilenetv2 are deployed to 5-class hand gesture recognition. In other words, 5-class gesture in 6 complex backgrounds with different illumination levels are classified using these deep learning models. The overall processes studied in this paper are shown in Figure 3.

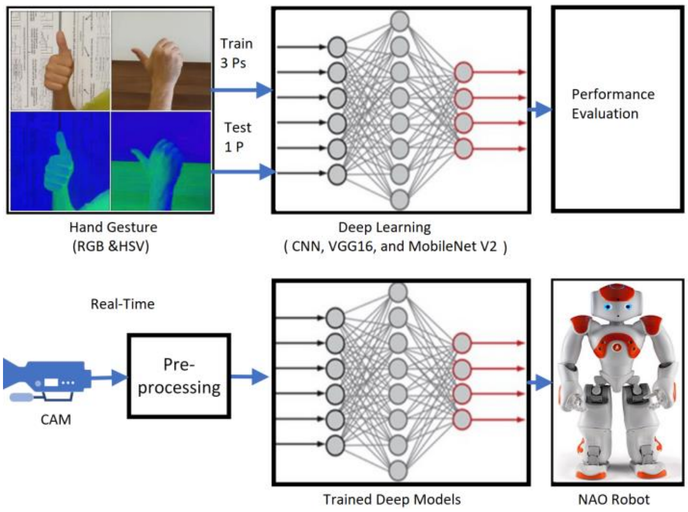

Figure 3. The proposed deep learning-based recognition and control (P denotes participant).

The hand gesture images in the collected database are resized to 224×224×3 to make compatible the input to the CNN models. The lightweight CNN is trained using training part of the dataset in RBG, and HSV, separately. The VGG-16, and MobileNetV2 are applied to transfer learning process by freezing the layers except the last output in Fully-connected (Fc) layer. Thus, advanced CNN structures are trained and adopted to new hand gesture classes. After training process of the three models, their performances are evaluated by testing with

300 hand gestures in terms of accuracy. Eventually, the most successful topology is trained and carried out. The second stage is the real-time implementation of the robot control using a camera, image pre-processing, and trained model so that the humanoid robot can move according to the hand pose taken by the camera. The three deep model are trained through transfer learning approach, but the lightweight CNN model given in Figure 4 is constructed and trained in this study.

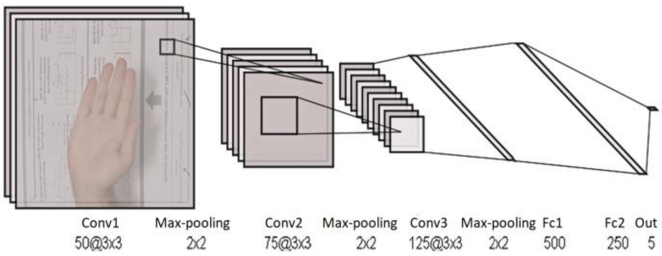

Figure 4. The lightweight CNN model in this paper.

In this CNN model, 3 convolutional layers (with 50@3×3,75@3×3, 125@3×3) with 3 max-pooling blocks (2×2) for feature extraction are followed by 3 Fc layers (with 500, 250, and 5 neurons). After each stage, ReLu is used. The dropout (50%) process is also existed among the FCS to avoid over-fitting. The popular CNN models VGG16, and Mobilenetv2 have superior recognition performance on image-net challenge with 1000 classes [23], reducing error rate up to 3.1%. Due to high computational cost and long-term training process up to a few weeks, it is preferred to transfer learnable parameters and weights to specialized classification task. Generally, the output of the CNN is revised according to new number of the classes. Thus, the weights between the last Fc and output are trained while other learnables are kept frozen to make faster learning. The demonstration of the transfer learning concept is shown in Figure 5.

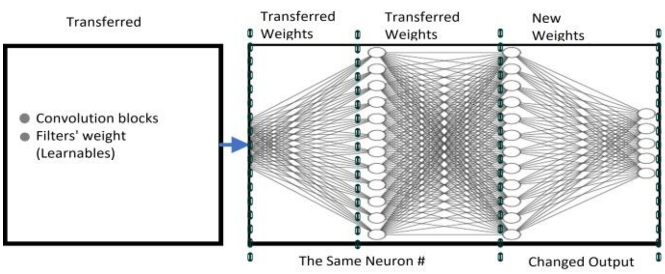

Figure 5. Transfer learning principle.

For hand gesture recognition, we adopt VGG16, and MobileNetV2 to 5-class output, and then these are trained using images in RGB and HSV color spaces to compare accuracy rates.

4.2. Humanoid Robot (Nao) Integration

NAO is a medium-sized humanoid robot and has been developing by the Aldebaran Robotics since 2004. Nao Academic Edition which can be programmable is used in universities and laboratories for research and educational purposes. It has totally 25 DOF and its height and weight are 57 cm and 4.5 kg respectively. It is equipped with sensors such as CMOS video camera, gyrometer, accelerometer, infrared sensor, ultrasonic sensor, loudspeaker, microphone etc. It has some features which are dancing, speaking, recognizing face and object etc. Today, researches continue to study on subjects such as being assistant, being autonomous, and developing a walking algorithm.

The NAO humanoid robot is integrated to a mobile phone camera, Wi-Fi router, and computer in our study. After trained deep learning models are deployed a computer, hand images taken by telephone transferred to assign hand gesture to correct command. TCP/IP operated NAO robot is then got the command via Wi-Fi link and makes the command action with our embedded code in the robot. The scheme of the real-time NAO humanoid robot control is given in Figure 6.

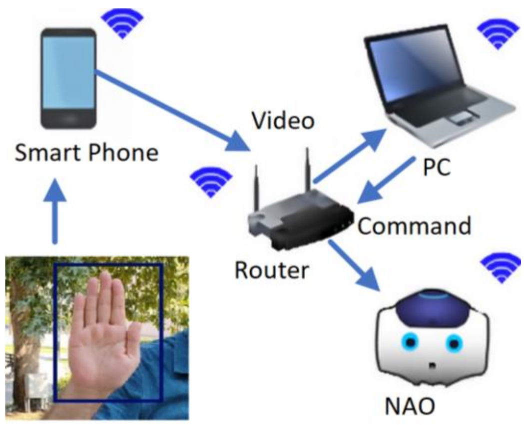

Figure 6. The scheme of the NAO robot integration.

After described CNN models running computer integrated to NAO based system, we have investigated offline hand gesture recognition performance of the deep learning models, and its real-time integration to a NAO robot control.

5. Experimental Results and Discussion

300 of 1200 hand gesture images in the dataset are remained as testing data for the deep learning models so that one participant out validation performance is evaluated. 5-hand gesture classification with 6 sophisticated backgrounds under different illumination levels is aimed using a lightweight CNN, and transfer learning models including VGG16, and Mobilenetv2. First, 224×224×3 images are extracted and converted to HSV space, and the effect of the color spacing on gesture recognition is investigated in gray scale, HSV, and RGB. Stochastic Gradient Descent with Momentum (SGDM) learning method with learning rate=0.001 and momentum=0.9 is used, and the gray scale, HSV, and RGB training graphs of the lightweight CNN are shown in Figure 7.

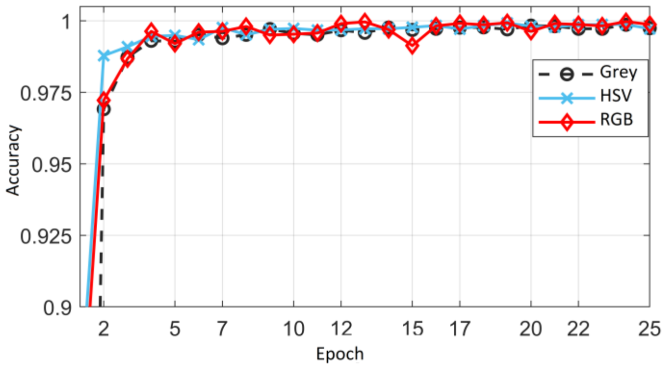

Figure 7. The training processes of the lightweight CNN in gray scale, HSV and RGB spaces.

The CNN model yields min-max 51.00-99.86% with the mean value of 97.55%, 86.48-99.92% with the mean value of 99.12%, and 84.72-99.97% with mean value of 98.97%, respectively. After training, these models are tested using 300 images of other participant, and the confusion matrices of them are given in Figure 8. The classification using RGB images of the hand gestures are more successful when compared to gray scale and HSV as given in Table 1.

Table 1. The results of the lightweight CNN.

Color Space	Training	Testing
Gray scale	97.55% (51.00-99.86%)	94.40%
HSV	99.12% (86.48-99.92%)	87.40%
RGB	98.97% (84.72-99.97%)	95.00%

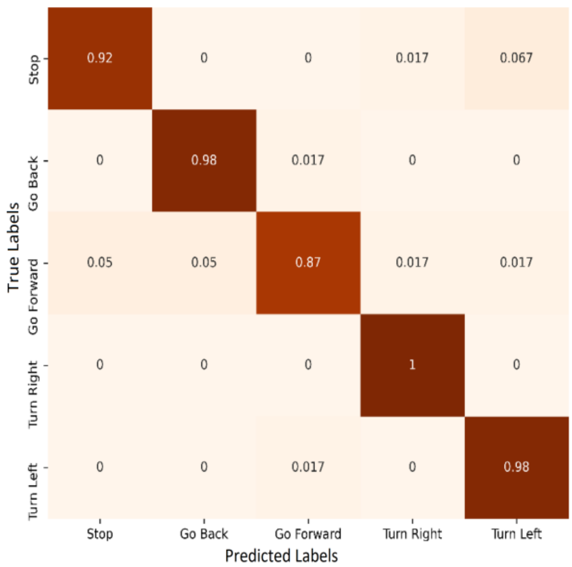

Figure 8. The confusion matrix of the lightweight CNN.

Hand gesture recognition using HSV color space is more successful than RGB at training phase. However, classification accuracy in RGB space (95.00%) is higher than HSV (87.40%).

It can be summarized that HSV learning is overfitted. After these experiments, we only apply RGB gesture images to the other models namely VGG-16, and MobileNetV2. These are more complicated model consists of 13 Convolutional Layers (CLs) and 3 Fully Connected layers (FCs) with drop-out, relu, and max-pooling for VGG16 (totally in 16 layers deep), inverted residual blocks and bottlenecks for Mobilenetv2. These are all trained for 1000-class image recognition, so their output and the learnable weights between output and the last FC. This requires shorter training period. We changed the output to 5-classes, and the last weights are trained. The confusion matrices of the transfer learning approaches are given in Figure 9, and the results are summarized in Table 2.

Captured hang image using a mobile phone camera is sent to PC with deep learning models using Wi-Fi, and recognized command is sent transferred to NAO robot using TCP/IP protocol. Wireless communication between the Nao robot and the computer is provided by the Socket Library in Python. TCP server is created with this library. IP address and port number is assigned to the server and then server communicate with port. At the same time, the client command in NAO humanoid robot is running. Then prediction of hand gesture information which is into the text file send to control the NAO humanoid robot. A box diagram written in Choregraphe runs on the NAO operating system given in Figure 10.

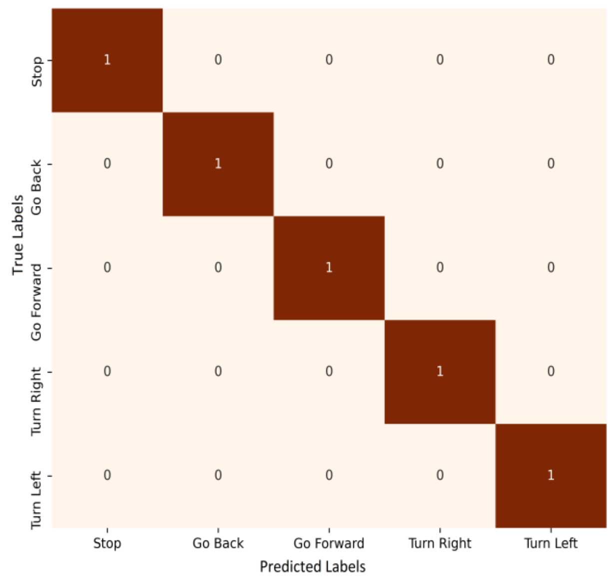

Figure 9. The confusion matrices of the transfer learning using VGG16, and MobileNetV2 (100% accuracy for both).

Table 2. The lightweight CNN and transfer learning results.

Model	Testing
Lightweight CNN	95.00%
VGG & MobileNet V2	100.00%

In this diagram, the NAO humanoid robot receives a command from the server in the client box. It runs its output according to the commands received. For example, if the forward command is received, the client box activates the go forward box, and NAO humanoid robot walks forward a certain distance. It continues in a loop. If the CNN is closed, a command is sent from the server to stop, and it stops in stand-up position as shown in Figure 11.

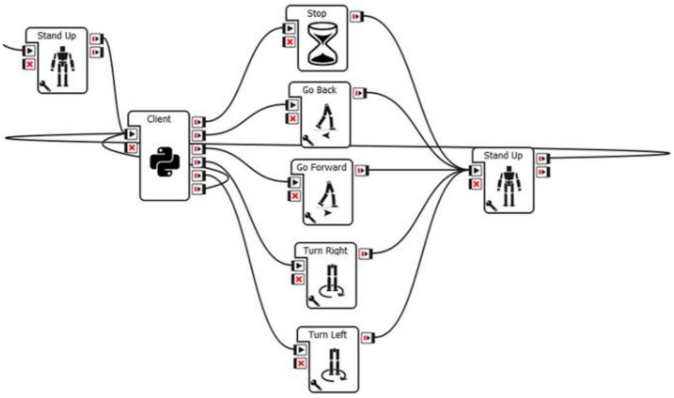

Figure 10. The box diagram written in Choregraphe runs on the NAO humanoid robot.

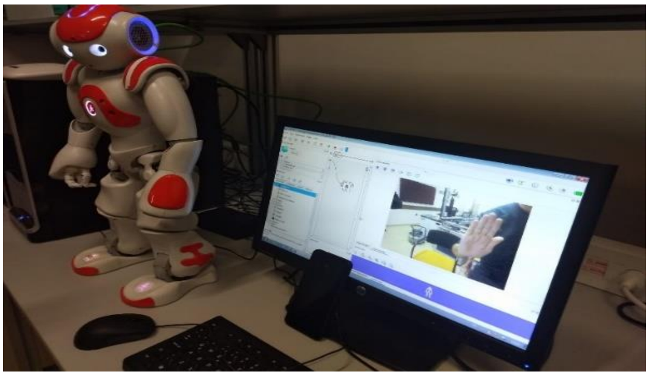

Figure 11. The NAO and computer integration.

6. Conclusions

In this paper, we proposed NAO humanoid robot control using the classification of the hand gesture images. 1200 hand images of the four participants are taken to collect 5-class with 6 sophisticated backgrounds under varying illumination levels, and then deep learning models have investigated to obtain high accuracy rate. A lightweight CNN with 3 convolution layers has correct rate of 95.00%, while transfer learning approaches using VGG16 and Mobilenetv2 has the perfect recognition rate of 100%. These deep learning models are successfully validated for recognizing the hand gestures (turn left and right, go forward and backward, and stop) in sophisticated background under different light levels according to one-participant out scheme (i.e, testing performance without using the participants' images in training). After training processes, NAO humanoid robot is integrated to real-time control implementing a mobile phone camera and a computer with running deep learning algorithms and TCP/IP communication. These offline and real-time analysis and implementations in this paper show that mentioned device configuration with deep learning can be successfully used for robot control, and human-robot interaction.

References

[1] Ahmad N. and Asif H., "Perceptual Quality Assessment of Digital Images Using Deep Features," *Computing and Informatics*, vol. 39, no. 3, pp. 385-409, 2020.

[2] Chen L., Fu J., Wu Y., Li H., and Zheng B., "Hand Gesture Recognition Using Compact CNN Via Surface Electromyography Signals," *Sensors*, vol. 20, no. 3, pp. 672, 2020.

[3] D'Orazio T., Marani R., Reno V., and Cicirelli G., "Recent Trends in Gesture Recognition: How Depth Data Has Improved Classical Approaches," *Image and Vision Computing*, vol. 52, pp. 56-72, 2016.

[4] Dardas N. and Georganas N., "Real-Time Hand Gesture Detection and Recognition Using Bag-Of-Features and Support Vector Machine Techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592-3607, 2011.

[5] Erozen A., "A New CNN Approach for Hand Gesture Classification Using sEMG Data," *Journal of Innovative Science and Engineering*, vol. 4, no. 1, pp. 44-55, 2020.

[6] Hafiane S., Salih Y., and Malik A., "3D Hand Recognition for Telerobotics," *in Proceedings of the IEEE Symposium on Computers and Informatics*, Langkawi, pp. 132-137, 2013.

[7] Hu B. and Wang J., "Deep Learning-Based Hand Gesture Recognition and UAV Flight Controls," *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 17-29, 2020.

[8] Hu Y., Sirlantzis K., Howells G., Ragot N., and Rodr´ıguez P., "An Online Background Subtraction Algorithm Deployed on A Nao Humanoid Robot-Based Monitoring System," *Robotics and Autonomous Systems*, vol. 85, pp. 37-47, 2016.

[9] Huang D., Hu W., and Chang S., "Gabor Filter-Based Hand-Pose Angle Estimation for Hand Gesture Recognition under Varying Illumination," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6031-6042, 2011.

[10] Jacob M., Wachs J., and Packer R., "Hand-Gesture-Based Sterile Interface for The Operating Room Using Contextual Cues for The Navigation of Radiological Images," *Journal of the American Medical Informatics Association*. vol. 20, pp. 183-186, 2013.

[11] Krizhevsky A., Sutskever I., and Hinton G., "Imagenet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[13] Levinson L., Gvirsman O., Gorodesky I., Perez A., Gonen E., and Gordon G., "Learning in Summer Camp with Social Robots: A Morphological Study," *International Journal of Social Robotics*, vol. 13, no. 2, pp. 999-1012, 2021.

[13] Li D., Li Y., Li J., and Fu Y., "Gesture Recognition Based on BP Neural Network Improved by Chaotic Genetic Algorithm," *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 267-

276, 2018.

[14] Li G., Tang H., Sun Y., Kong J., Jiang G., Jiang D., Tao B., Xu S., and Liu H., "Hand Gesture Recognition Based on Convolution Neural Network," *Cluster Computing*, vol. 22, no. 2, pp. 2719-2729, 2019.

[15] Liu H. and Wang L., "Gesture Recognition for Human-Robot Collaboration: A Review," *International Journal of Industrial Ergonomic*s, vol. 68, pp. 355-367, 2018.

[16] Mazhar O, Navarro B, Ramdani S., Passama R., and Cherubini A., "A Real-Time Human-Robot Interaction Framework with Robust Background Invariant Hand Gesture Detection," *Robotics and Computer-Integrated Manufacturing*, vol. 60, pp. 34-48, 2019.

[17] Mazhar O., Ramdani S., Navarro B, Passama R., and Cherubini A., "Towards Real-Time Physical Human-Robot Interaction Using Skeleton Information and Hand Gestures," *in Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, Madrid, pp. 1-6, 2018.

[18] Mirza N., "Robotic Path Planning and Fuzzy Neural Networks," *The International Arab Journal of Information Technology*, vol. 17, no. 4A, pp. 615-620, 2020.

[19] Nolker C. and Ritter H., "Visual Recognition of Continuous Hand Postures," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 983-994, 2002.

[20] Phung V. and Rhee E., "A Deep Learning Approach for Classification of Cloud Image Patches on Small Datasets," *Journal of Information and Communication Convergence Engineering*, vol. 16, no. 3, pp. 173-178, 2018.

[21] Pisharady P. and Saerbeck M., "Recent Methods and Databases in Vision-Based Hand Gesture Recognition: a Review," *Computer Vision and Image Understanding*, vol. 141, pp. 152-165, 2015.

[22] Pisharady P., Vadakkepat P., and Loh A., "Attention Based Detection and Recognition of Hand Postures against Complex Backgrounds," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 403-419, 2013.

[23] Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., and FeiFei L., "Imagenet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, 2015.

[24] Sagayam K. and Hemanth D., "Hand Posture and Gesture Recognition Techniques for Virtual Reality Applications: a Survey," *Virtual Reality*, vol. 21, no. 2, pp. 91-107, 2017.

[25] Sandler M., Howard A., Zhu M., Zhmoginov A., and Chen L., "MobilenetV2: Inverted Residuals and Linear Bottlenecks," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, pp. 4510-4520, 2018.

[26] Sokhib T. and Whangbo T., "A Combined Method of Skin-and Depth-Based Hand Gesture Recognition," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 137-145, 2020.

[27] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., and Rabinovitch A., "Going Deeper with Convolutions," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 1-9, 2015.

[28] Tang G. and Webb P., "The Design and Evaluation of An Ergonomic Contactless Gesture Control System for Industrial Robots," *Journal of Robotics*, vol. 2018, no. 2, pp. 1-10, 2018.

[29] Yu N., Xu C., Wang K., Yang Z., and Liu J., "Gesture-Based Telemanipulation of a Humanoid Robot for Home Service Tasks," *in Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems*, Shenyang pp. 1923-1927, 2015.

Ali Yildiz received his B.Sc. Degree in Mechatronics Engineering from Bursa Technical University in 2021. He does research in image processing, deep learning, and robotics applications.

Nurettin Gökhan Adar received the B.Sc. degree in Mechanical Engineering in 2004 from Sakarya University. M.Sc. degree in Mechanical Engineering in 2008 from Istanbul Technical University, and the Ph.D degree in Mechanical Engineering in 2016 from Sakarya University. He has been with the Department of Mechatronics Engineering, Bursa Technical University since 2018 where he currently holds an Assistant Professor position. His research interests are modeling, dynamic system control, image processing, and robotics.

Ahmet Mert received the B.Sc. degree in Electrical and Electronics Engineering in 2006, M.Sc. degree in Mechatronics Engineering in 2009 from Marmara University, and the Ph.D degree in Biomedical Engineering in 2015 from Istanbul University. He has been with the Department of Mechatronics Engineering, Bursa Technical University since 2016 where he currently holds an Associate Professor position. His research interests are biomedical signal processing and pattern recognition.