*International Journal of Latest Trends in Engineering and Technology*
*Vol.(14)Issue(2), pp.017-022*
*DOI: http://dx.doi.org/10.21172/1.142.03*
*e-ISSN:2278-621X*

# SURVEY AND HIGH LEVEL DESIGN OF ACTIVITY MONITORING FOR ICU PATIENTS

Nikitaa Magi[1], B. G. Prasad[2], Priyanka Jigalur[3]

**Abstract -ICU patients are immobile and their medical conditions are very sensitive, any minute response shown by them have to be immediately reported to the hospital authority so that the patients are treated as quickly as possible. ICU patients require constant monitoring in order to detect and identify any movement exhibited by them. One of the conventional methods of monitoring ICU patients is to have assigned observer who observes the patients all the time. This conventional method for monitoring ICU patients has several limitations such as observer must be available by the side of the patients always or observer might overlook some minute actions performed by patient. This paper presents a survey on smart and automatic vision based patient monitoring system using deep learning methods and image processing. Deep leaning provides several methods for activity monitoring, detection and identification such as 2D convolution networks, 3D convolution networks, etc. This paper discusses classification, challenges, application and methods for activity detection and also presents high level system design for activity monitoring system.**
**Keywords –Deep learning, Neural Networks, Patient Monitoring, Image processing, Activity analysis.**

## 1. INTRODUCTION

The Patients in critical condition need intense monitoring and care. ICU patients who are in coma state are static, they do not show any movement in that state. Patients may show minute activity such as hand or leg movement, and when they show such activity it has to be notified to hospital authority as early as possible and the hospital authorities have to treat the patient.The conventional method of monitoring ICU patients requires dedicated person to look after the patient, the person who is monitoring the patient should be proactive and available by the side of the patient all the time. Another issue is one person can only be able to monitor one patient at a time. Nowadays shortage of Intensivist and Critical Care nurses to look after the ICU patients is the major problem faced by the hospitals so, this research aims to present survey on the system which overcomes the issues of conventional methods. To overcome such problems automatic vision based patient monitoring systems are needed.For reducing the workload of the person and to automate some of the humans handled task we need the machine based interface which will take the decisions automatically weather the patient has performed some activity or not. Such system must to able to monitor the patients in ICU and assist the person or doctor available at the physical location in case of emergency. Automatic monitoring systems can provide round the clock monitoring, they are cost efficient and reliable. The proposed survey on the system in this paper is about object detection to identify the type of persons available in the ICU room along with activity analysis that is used to automatically detect several unusual activities done by the patient. As soon as any unusual activity is detected the system notifies the same to hospital authority on the type of activity and persons available in the ICU Room.

Deep learning methods are becoming more and more powerful in classification, recognition, identification, location and analysis of spatial and temporal information. There are many types of neural network available for doing different type of tasks. 2D neural networks has the ability to learn spatial data, they are used for tasks such as image classification, recognition, object detection, etc. 2D neural networks do not have capability to capture temporal data, they cannot be used for real time applications. 3D convolution neural networks has the ability to learn spatial data as well as temporal data, they can be used for tasks such as activity analysis. 3D convolution networks are slow in learning as well as inferencing. Since, Activity monitoring for ICU patient is hard real time system, inferencing time and accuracy cannot be compromised. There is a need to develop a new approach to obtain fast and accurate results for our application of activity monitoring. Collaboration of Video Processing and Deep Learning network opens up the door for many real-time applications. Real-time object detection from video frame is not an easy task to do, especially when the objects are moving and a background of frames is also changing. Video processing is the essential part of our proposed system. As we monitor the patients in real time, video processing is required. The ICU room has the camera mounted on the top, this camera captures information in the form of real time vides and sends data to the activity monitoring system. We faced various challenges regarding the processing of videos and data

[1] Department of Computer Science and Engineering, B. M. S College of Engineering, Basavanagudi, Bangalore, Karnataka, India
[2] Department of Computer Science and Engineering, B. M. S College of Engineering, Basavanagudi, Bangalore, Karnataka, India
[3] Department of Computer Science and Engineering, K.L.E Technological University, Hubballi, Karnataka, India

collection for experimental evaluation of system e.g. too many moving objects are available in particular scene or frame which makes patient motion detection task tough, various action done by the patients differs from patient to patient which is challenging task at time of detection.

Since patients in ICU can be accompanied by other members such as doctor, nurse or family member there is need to detect target object (patient). Very popular deep learning methods for object or person detection from the scene are Convolutional Neural Network. At present different type of architectures are available for Convolutional Neural Network (CNN). Object detection neural network models such as RCNN, Fast RCNN, Faster RCNN, SDD, YOLO, etc. can be used to identify, and locate target object (patient) in ICU room from real time video data. Once the target object is being identified there is need to detect if the target object has performed some activity. Simple classification neural network can be repurposed to perform task of activity detection. Image classification model has to be trained to classify image into two class, normal or abnormal. No activity is said to have being detected as long as all frames from real time data are classified as normal. Once the frame from real time video is classified as abnormal, activity is said to have being detected.

*1.1 Types of human activity analysis*

Different types of activity analysis are classified based on various factors as illustrated below in Figure. 1.[1]

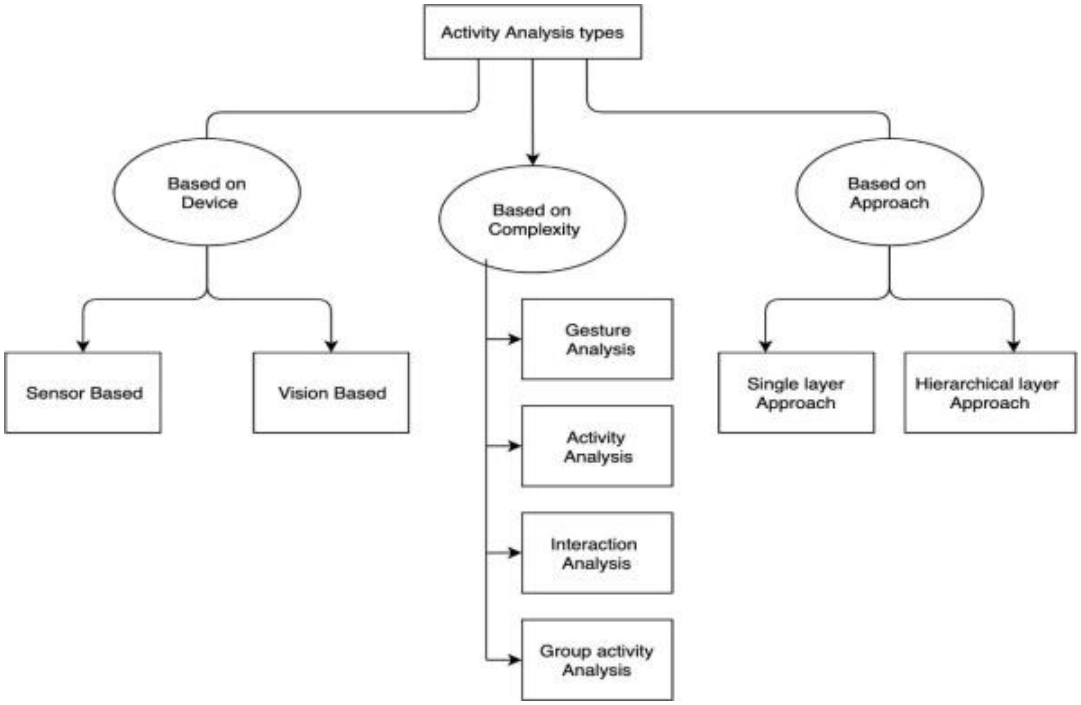


Figure 1. Classification of different types of activity analysis based on devices, complexity and approach

Based on Device - Sensor based activity analysis is one of the activity analysis methods where network of different physical sensors are used to sense different types of activities performed by target object [2]. Sensors generate different output voltages corresponding to different activities, this voltage is further analyzed to identify type of activity performed. Vision based activity analysis uses devices that capture the event in the form of videos or images, later this image or video is processed to analyze activity.

Based on Complexity- Gesture recognition involves analysis of simple motion of human body parts such as waving of hands, closing and opening of palm, blinking of eye, etc. Activity analysis – Analyze activity performed by single actor such as jumping, walking and running. Interaction Analysis – Analyze activity between two actors such as hand shake, greeting each other, waving at each other, etc. Group activity analysis – Analyze actions performed by group of actors.

Based on Approach [3] – Single layer approach where sub activities are identified independently, it does not involve combining of sub activity to be identified as whole activity. Hierarchical layer approach – this method identifies sub activity from input and then combines all of them to produce final outcome. Playing football involves running, kicking the ball, screaming, etc. When single layer approach is used all these sub activities are identified individually as running or kicking ball but are not identified as whole activity of playing football, where as in case of hierarchical layer approach all these sub activities are identified and are combined together to be recognized as one whole activity of playing football.

*1.2 Applications of Activity Analysis*

Human activity analysis finds its applications in various fields like

Behavioral Bio-metrics-In traditional bio-metrics, physical properties of a person, such as finger-print or iris scan is used to uniquely identify a person. Whereas Behavioral Bio-metrics involves methods to analyze distinctive motion history of people over time in order to uniquely identify them [4].

Security and Surveillance- Analyzing human activity through surveillance videos in real time to detect any unusual and abnormal activity such as road accidents, fire accidents, people fighting, etc. and immediately alert the concerned authority in order to provide safety and security [5].

Animation and synthesis- Involves understanding human appearance, movements, responses to stimuli, behavior, etc., and incorporate these human features into gaming and animation in order to make it look more realistic

Healthcare Systems –Involves methods to continuously evaluate patient's behaviors in order to identify abnormality or disorders, these applications reduces work load on medical staff, enables faster detection of disorders and are cost efficient. Healthcare systems can not only be used in detecting disorders but can also be used in monitoring patients who are in coma. One of the applications of Activity analysis in health care is early diagnose of autism.

*1.3 Challenges of vision based activity analysis*

There are several challenges of vision based activity analysis.

Camouflage – Segregation of target object from rest of the background is difficult because of the resemblance of appearance between target object and the background

Human activity are highly versatile and largely complex

Moving background creates as illusion of moving target object

Shadow effect - Target object's shadow creates impression of presence of other objects, causing false activity identification.

Factors to be considered from image or video perspective like angle of lighting, illumination, data format, data quality, noisy images, jitter, etc.

Appearance of human - Human can be of different age group, genderbody shape and height

## 2. LITERATURE SURVEY

Our primary task to implement activity monitoring system for ICU patients is to identify and detect target person i.e., patient in our case. One of the prominent and distinguishing character to identify human is to identify human face. The very first successful work in the area of computer vision was the development of an algorithm that could detect human face [6]. This algorithm was developed by Viola P. and Jones M., hence the name Viola Jones algorithm. Some of the prominent features of human face such as eyes, nose, cheeks, etc. were mapped to be identical to Haar filters. 24*24 sized windows was used to look at the input image to identify haar features. There were 160,000+ haar filter calculation for each window. Adaboost was used to eliminate some of the dispensable features, this algorithm used 7000 feature calculations for each window. This algorithm was further optimized dividing features called Cascade [7]. Even after using Adaboost, Integral Image and cascading Viola Jones algorithm was not efficient for real time applications of detecting faces as it was computationally expensive and slow. Viola Jones algorithm failed to detect faces at different angles. The next efficient technique for human detection was Histogram of Oriented gradients [8] implemented for pedestrians detection. The primary aim of HOG is to compare darkness of a given pixel with its neighboring pixels and draw a gradient in the angle of darker pixel. Every pixel was compared with all its neighboring pixels and convert input image to the gradients. Gradient image was finally compare with hard coded gradient to detect pedestrian. Deformable parts model [9] viewed humans as set of discrete features separated by respected distance, but the basic approach remains same as that of HOG. [10] Presents Histogram of Oriented gradients for detecting human face. Disadvantages of HOG method is that it is slow for real time object detection and requires hard coded features for human detection. The above mentioned algorithms, Viola Jones algorithm and Histogram of Oriented gradients used some hard coded features for object detection and these methods were not able to learn by themselves given set of supervised data. Development of neural network opened the door for many real time applications because neural networks had the ability of self-learning when trained on set of supervised data without need of any hard coded features. LeNet was developed in 1996 [11] and was used to identify hand written digits. Neural networks achieved breakthrough moment in 2012 due to availability of large computational power coming from advanced processors, GPUs and also availability of large dataset for training and testing

ILSVRC held vision based competition every year to award best algorithm for image classification and detection. Super Vision (AlexNet)[12]won ILSVRC in 2012, AlexNet presented convolutional implementation of neural networks. AlexNet had 8 layers, first 5 convolutional layers and 3 fully connected layers, final layer was 1000 way softmax indicating probability of 1000 classes. 17% was the Top5 error rate when tested on Image Net dataset. Detailed explanation of CNN is give in [13]. 2013 ILSVRC was won by ZFNet [14], it was implemented as improvisation on top of AlexNet. Output of every layer in ZFNet was analyzed using deconvolution neural network. Two issues were found in first two layers of AlexNet, first issues were resolved by reducing first layers size to 7*7 and second issue was fixed by changing first layer stride size of convolution to 2. 14.8% was the Top5 error rate when tested onILSVRC dataset. In 2014, ILSVRC winner was GoogleNet [15] for object classification and detection, with 22 layers. GoogleNet used parameters that were 1/12th of the parameters used in AlexNet.

Lin M. proposed an idea called Network-in-Network in which 1*1 convolution layers are added to a network in pursuance of increasing the depth of the network[16]. Google Net incorporated batch normalization, inception model and image distorting to improve performance. Google Net had Top5 validation error rate of 6.7% on ILSVRC dataset.

VGGNet[17]was selected as next best algorithm after GoogleNet by ILSVRC in 2014. VGGNet involved preprocessing set by eliminating mean RGB value of input image. VGGNet had 16 convolution layers and attempted to expand depth of the convolution neural network by using tiny convolution kernals of size 3*3 or 1*1. Hidden layers was implemented with ReLU. Local response normalization was used to implement one layer of VGGNet. VGG Net training was computationally expensive in order to handle 138 million parameters. This network was not only tested on ILSVRC dataset but also was testes on PASCAL, VOC and Caltech datasets. Top5 validation error rate was 7.3% for VGGNet on ILSVRC dataset. ResNet (Residual Network) [18] won ILSVRC for image classification and detection in 2015, it was 152 layered network. Even though Res Net had 152 layers it was computationally inexpensive as it skipped some of the connections between layers based on residual capacity. ResNet accuracy was more than that of human with Top5 error rate of 3.57 on ILSVRC dataset. In fig. 2. X axis shows ILSVRC winning network and the year of winning, Y axis shows Top5 error rate for classifying images.

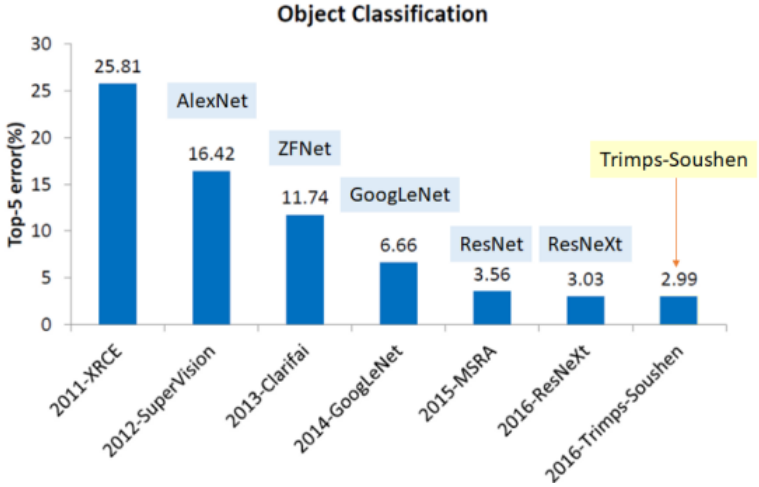


Figure 2. Comparison of ILSVRC winning networks and their Top5 error rates.

Real time applications of image analysis using neural networks are not as simple as that of image classification. For real time application first, the objects in the image has to be classified. Second, different objects has to be located using bounding boxes, this task is referred as object detection. Object detection cannot be implemented by using simple CNN because output of object detection is the bounding boxes around each object detected in the image and the size of these bounding boxes are not fixed.

Deformable parts model implementation for object detection was shown by Felzenszwalb P., and others in [19]. New technique of hand coded SVM is used at the end of object detection algorithm called as latent SVM, which is reconstructed using MI-SVM w.r.t latent variables. For non-positive examples, latent SVM is semiconvex and for positive examples, latent SVM is fully convex. This method could correctly detect 10 object classes with average confidence score of 0.6 on PASCAL VOC 2006 dataset and could correctly detect 20 object classes with average confidence score of 0.3 on PASCAL VOC 2007 dataset. Object detection using deformable parts model was not based on neural network,Szegedy C. with others [20]Showed that simple convolution neural networks designed for the task of classifying images can be modified to support object detection. AlexNet was used as base neural network for implementing object detection. Binary mask was produced by replacing last softmax layer of AlexNet with regression layer. Binary mask was used as object representor. The object binary box is then refined and resized by continuing process of generating sub images. Object mask provides the final output. This method was very slow.

One of the methods for object detection was R-CNN (Region-CNN) [21]which predicted possible areas that might contain some object in it, as a part of preprocessing stage. Initially the input image was preprocessed using an operation called segmentation. Segmentation process creates 2000 region proposal on the input image. Segmentation operation processes input image across different sized window and attempts to assemble pixels with similar color, texture and intensity. Image with proposed region is then passed to pre trained AlexNet for extracting features. Once the features are extracted they are then passed to SVM for image classification, R-CNN final output contains object class and bounding box for each object detected around the object. R-CNN is not suitable for real time applications as it is computationally expensive. R-CNN produced mAP (mean Average Precision) of 53.7% on PASCAL VOC 2010 and mAP of 31.4% on ILSVRC 2013 dataset. The drawbacks of [21]were addressed in Fast R-CNN [22]. In this method input image was still preprocessed using segmentation but used convolution implementation of sliding window to classify all proposed regions at once. Region proposals were reshaped into specific size using region of interest pooling layer and were fed to fully connected CNN all at once. Finally softmax layer was used to detect class. Fast R-CNN was superior to R-CNN with respect to speed as it used ConvNet to generate region proposal

instead of using segmentation. Fast R-CNN produced mAP of 66% on PASCAL VOC 2012. [21][22]used image separate preprocessing stage to generate region of proposal, this preprocessing stage is slow in both the cases. Faster R-CNN [23] was designed in order to eliminate preprocessing stage. A dedicated convolution neural network was designed to generate region proposal instead of segmentation process, this approach was faster compare to other two methods. Faster R-CNN was most satisfactory methods Real-time object detection. It produced mAP of 67% on PASCAL VOC 2007 and mAP of 41.5% on COCO dataset.

You only look once (YOLO) [24]was best suitable for real time object detections, it looked at the image just once, hence the name You Only Look Once. Instead of reusing object classification as base for object detection, YOLO used regression approach. Input image is divided into N*N cells, each of the cell is responsible to predict objects. Each cell is also responsible to predict M different regions representing objects. Each cell outputs confidence score for detecting some object and M bounding box whose value is 5 tuples i.e., x, y, height, width and class probability. Each cell is also associated with twenty conditional class probability (20 object class). Class confidence score is product of bounding box confidence score and conditional class probability. Some threshold value T is defined such that if the class confidence score is more than defined T then that particular object is detected and located within that bounding box. YOLO was trained on PASCAL VOC dataset, it was restricted only to classify objects belonging to 20 class. YOLO showed result of object detection at the rate of 45 frames per second on PASCAL VOC dataset using TITAN X graphics processing unit. YOLO was enhance in YOLOv2 which was much faster and was able to detect objects belonging to more than 20 classes [25]. YOLOv2 showed result of object detection at the rate of 67 frames per second on PASCAL VOC, COCO dataset. YOLOv2 used Batch normalization, multiple shapes of anchor boxes and dimension clustering. YOLO9000 was developed on top of YOLOv2 and was trained on top 9000 classes form Image Net along with COCO dataset. YOLOv3 [26] was introduced to improvise on YOLO and YOLOv2. YOLOv3 used multiple class labels. YOLOv3 uses logical regression for each bounding box. YOLOv3 mean average precision was same as that of SSD [27]but it was three times faster than SSD. YOLOv3 showed result of object detection at the rate of 22 frames per millisecond with 28.2 mAP.

## 3. SYSTEM DESIGN

Figure 3 describes the complete System Design to implement activity monitoring for ICU Patients

Image frame are extracted from ICU room real time video, this image is passed through person detection trained model.

Person detection trained model returns bounding box around each person present in the input image. Number of bounding boxes are counted in order to get the count of number of people in the frame.

When object count is more than one i.e., when patient is accompanied by family members or hospital staff, the system continues Person detection trained model for upcoming incoming frame. When object count is one i.e., patient is alone, system crops the image and passes only the bounding box around the person to the next step.

Open CV libraries are used to detect patients hand from the cropped patient image This patient image is further cropped and only patients hand image is sent to next stage of the system.

Hand classification model inputs cropped hand image of the patient, classifies it as normal or abnormal. If the hand is classified as abnormal then activity is said to be detected, if hand is classified as normal then entire process repeats for next frame from real time video.

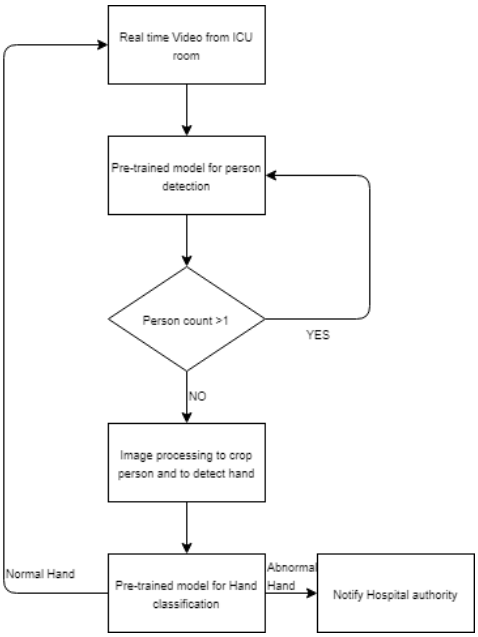


Figure 3. Complete System Design to implement activity monitoring for ICU Patients

## 4. CONCLUSION

Intensive care for ICU patients is critical, ICU patients need continues monitoring in order to observe their movements. Conventional method of monitoring ICU patient needs observer to take care of the patient all day long and notify the hospital authority in case patient shows any activity. There are several drawbacks of this conventional method such as observer must be present 24 hours of the day, observer might overlook some of the minute movements shown by the patient. Development of deep learning has paved way for many real time applications due its self-learning capability. Deep leaning merged with image processing provides excellent framework for vision based automatic systems. This paper presents high level system design for vision based survey on different deep leaning algorithms for image classification, object detection and activity analysis. . This paper also presents high level system design for vision based automatic activity monitoring system using deep learning and image processing. Among many object detection methods, YOLO has shown fastest results. There are different versions of YOLO such as YOLO, YOLOv2 and YOLOv3. This paper presents used of 2D convolution network for image classification repurposed to act as activity analysis method. There are several advantages of vision based automatic activity monitoring system over conventional method. Vision based method is cost efficient, does not require manpower, accurate and feasible.

## 5. REFERENCES


[1] A. G. D'Sa and B. G. Prasad, "A Survey on Vision Based Activity Recognition, its Applications and Challenges," in International Conference on Advanced Computational and Communication Paradigms, sikkim, 2019.

[2] J. T. Sunny, S. M. Gellar and J. J. Kizhakkethottam, "Applications and Challenges of Human Activity Recognition using Sensors in a Smart Environment," International Journal for Innovative Research in Science & Technology, vol. 2, no. 4, 2015.

[3] G. Cheng, Y. Wan, A. N. Saudagar, K. Namuduri and B. P. Buckles, "Advances in Human Action Recognition: A Survey," arXiv, 2015.

[4] E. Hossain and G. Chetty, "Multimodal Feature Learning for Gait Biometric Based Human Identity Recognition," in Neural Information Processing, Springer, Berlin, 2013.

[5] R. K. Tripathi, A. S. Jalal and S. C. Agrawal, "Suspicious human activity recognition: a review," Artificial Intelligence Review, vol. 50, no. 2, pp. 283-339, 2018.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in IEEE Conference on Computer Vision and Pattern Recognition, Kauai, 2001.

[7] A. Sahitya, M. T. N and V. N, "A Survey on Face Recognition Technology - Viola Jones Algorithm," in IJCA Proceedings on National Conference on Recent Trends in Information Technology NCRTIT, 2016.

[8] N. Dalal, B. Triggs and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance," in Computer Vision – ECCV , Springer, Berlin, 2006.

[9] H. Azizpour and I. Laptev, "Object Detection Using Strongly-Supervised Deformable Part Models," in Computer Vision – ECCV , Springer, Berlin, 2012.

[10] o. Deniz, G. Bueno, J. Salido and F. D. L. Torre, "Face recognition using Histograms of Oriented Gradients," Pattern Recognition Letters, vol. 32, no. 12, pp. 1598-1603 , 2011.

[11] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[12] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[13] Y. LeCun, K. Kavukcuoglu and C. Farabet, "Convolutional networks and applications in vision," in Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, 2010.

[14] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," arXiv, 2013.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," arXiv, 2014.

[16] M. Lin, Q. Chen and S. Yan, "Network In Network," arXiv, 2014.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in International Conference on Learning Representations, 2015.

[18] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer vision and Pattern Recognition, 2016.

[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1627-45, 32 September 2010.

[20] C. Szegedy, A. Toshev and D. Erhan, "Deep Neural Networks for Object Detection," in Neural Information Processing Systems, 2013.

[21] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in IEEE Conference on Computer Vision and Pattern Recognitio, 2014.

[22] R. Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision, 2015.

[23] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in Neural Information Processing Systems, 2015.

[24] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once:," in IEEE International Conference on Computer Vision, 2016.

[25] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in IEEE International Conference on Computer Vision and Pattern Recognition, 2016.

[26] J. Redmon and A. Farhadi, "YOLOv3: an Incremental Improvement," in IEEE International Conference on Computer Vision and Pattern Recognition, 2018.

[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, "SSD: Single Shot MultiBox Detector," in IEEE Conference on Computer vision and Pattern Recognition, 2016.