2021 Asian Conference on Innovation in Technology (ASIANCON)
Pune, India. Aug 28-29, 2021

A Schematic Review on Applications of Deep Learning and Computer Vision

Prof.Rahul Pitale
Department of Computer Engineering
Pimpri Chinchwad College of Engineering,
Pune, India
rahulpitale3@gmail.com

Harshvardhan Kale
Department of Computer Engineering
Pimpri Chinchwad College of Engineering
Pune, India
harshvardhankale34@gmail.com

Sakshi Kshirsagar
Department of Computer Engineering
Pimpri Chinchwad College of Engineering,
Pune, India
sakshikshirsagar@gmail.com

Harshal Rajput
Department of Computer Engineering
Pimpri Chinchwad College of Engineering,
Pune, India
harshalrajput36@gmail.com

***Abstract*— Artificial intelligence(AI) and Deep learning (DL) become supreme problem-solving strategies in many areas of research and industrial applications.[1].Deep learning is one of the most efficient, accurate and cost efficient algorithms. Applications of Deep learning via computer vision in Bio medical and healthcare, security, education and latest trends in technologies are stated here. This paper focuses on giving an overview of work done by many authors on Deep learning and computer vision applications and algorithms. At last common findings from the papers are reviewed.**

Keywords—Deep Learning , Computer Vision , Biomedical security, Education and Latest Trends .

I. INTRODUCTION

Artificial intelligence is a branch of computer science which deals with the study of building smart computers or machines which are capable of working like human brains. It was introduced in the 1960s. Subpart of Artificial intelligence is machine learning which is used for data analytics and analytical model building . Deep learning is a subpart of machine learning . The concept of deep learning was introduced by mathematician Alexey Ivakhnenko in the 1960s. The concept of Deep Learning arose from getting inspired by the behavior of the human brain learning large amounts of data . As machine learning is based on artificial neural networks , neural networks were made as mimic of the human brain . The architectures of deep learning such as graph neural networks,deep neural networks, recurrent neural networks, etc are vastly applied to fields like computer vision, speech recognition, medical image analysis , image processing , etc.

Computer vision was parallely developed with Artificial intelligence in the 1960s. Deep learning is applied to computer vision techniques to help computers see what humans can see and analyse. In other words computer vision helps computers to understand and analyse digital images such as photographs or videos. Computer vision has helped worldwide in various fields by its wonderful techniques such as character recognition or face recognition, medical imaging, machine inspection, 3D model building, motion capture , surveillance , fingerprint recognition , biometrics and many more . These applications are applied on various domains like healthcare and biomedical , security , education , sports , etc .The work presented in this paper will give an effective overview of application of computer vision techniques and algorithms on multiple domains.

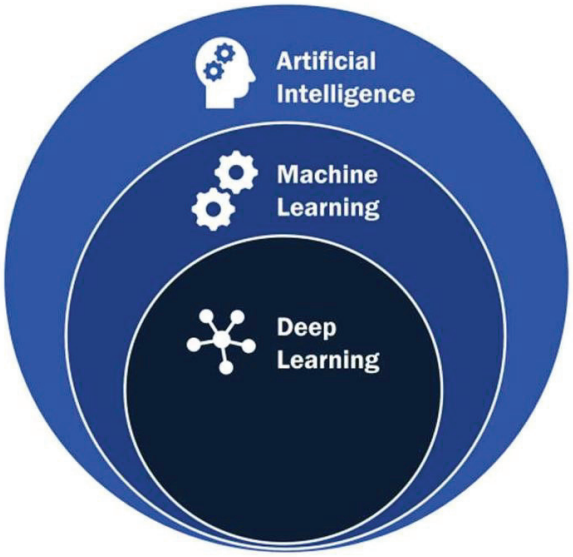

Fig. 1. Representation of Artificial Intelligence , Machine Learning and Deep Learning.

II. LITERATURE SURVEY

The work present in this paper stated about applications of Deep learning and Computer vision and the algorithms used by them . Here four algorithms namely Convolutional neural networks(CNNs), Generative Adversarial Networks (Gans), Support vector machine (SVM) and Haar Cascade Algorithms are discussed.The working of these four algorithms is presented below.

A. Convolutional neural network (CNNs):

In the concept of deep learning, CNN is a most efficient class of deep neural networks, mostly applied to analyze images. It uses a special technique called convolution. In mathematical terms, it is an operation on two functions that gives rise to a third function and determines how the shape of one is modified by the other. CNNs consist of 3 layers: convolutional, pooling and the fully connected layer. The convolutional layer contains filters that perform operations on the image matrix and give an output as a feature map which is fed to the pooling layer. This layer captures essential features and the dimensions of the feature map are reduced. The output from this layer is fed into the next layer, which is responsible for identifying the images.

2021 Asian Conference on Innovation in Technology (ASIANCON) | 978-1-7281-8402-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ASIANCON51346.2021.9544941

An image can be represented as: dim(image) = (nH, nW, nC) where:

nH: size of height nW: size of width

nC: number of channels

The filter can be defined as follows: dim(filter) = (f , f , nC)

where f: odd dimension of the filter

For a given filter and image:

$$\text{conv}(I, K)_{x,y} = \sum_{i=1}^{n=H} \sum_{j=1}^{n=W} \sum_{k=1}^{n=C} K_{i,j,k} I_{x+i-1,y+j-1,k}$$

CNNs have very high accuracy in image recognition problems and they automatically detect the important features without any intervention.

However, they do not encode the position and orientation of an object and they require lots of training data.

B. Generative Adversarial Networks (GANs):

GANs are deep learning algorithms which create a batch of sample data that is trained to distinguish between real and fake images .GANs architecture consists of two components: a generator,which is used in creating batches of sample data , and a discriminator, which is used in differentiating in real and fake images.The GANs have shown vast growth over a period of time. Quality enhancement of an image , generation of crafted images, reconstruction of an image and image registration and segmentation are the objectives of the GANs in the applications.

GANs help in generating 3D objects, cartoon and realistic images and characters and photographs of human faces.

Working Of GAN:

Generator : The generator takes a random vector which is of fixed lengths as input and generates a sample or a fake image .

Discriminator : Discriminator is a classification model . The output of the generator model i.e real or fake image is given input to the discriminator model. It is trained in such a way that it can differentiate the fake data and the real data of the generator.

C. Support Vector Machines :

Support Vector Machine (SVM) is a supervised machine learning algorithm that is widely used in segregation problems. SVMs are based on finding a hyperlink that severely divides data into two categories. Support vectors are data points near the hyperplane, data set points that, when removed, cause the hyperplane to change positions. Hence they are a very important part of the algorithm. A hyperplane can be thought of as an entity which divides the dataset into equal parts..

So, whenever new test data is introduced, the side of the hyperplane the data resides on determines the category that it belongs to . The distance between the hyperline and the nearest data point from each set is known as the margin. A hyperplane with a large margin should be selected between the hyperplane and any point within the training set. To separate a database, it is necessary to move from 2d data view to 3d view. Imagine that different coloured balls are spread out on a sheet and the sheet is lifted in the air. Midair, we use another sheet to divide these balls into their respective categories. This is a simplified version of what is called kernelling.

D. Haar Cascade Algorithm:

The concept of Haar cascade removed features from images using a kind of 'filter', similar to the concept of convolutional kernel.

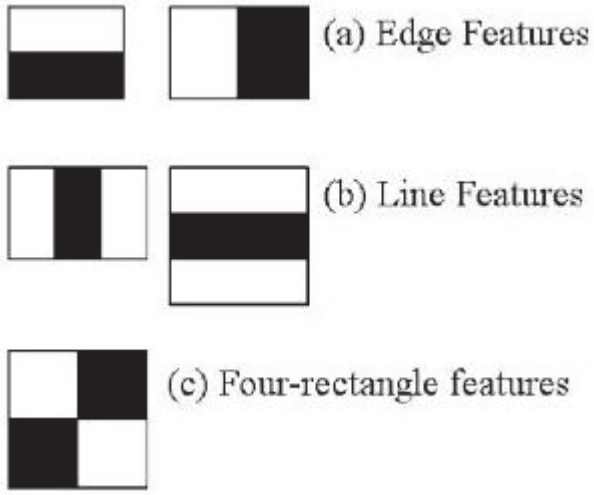

Fig. 2. Haar Cascade Features

The algorithm transfers these filters to an image, exploringone component (or window) at a time. After that, in each window, all the pixel powers, white and black parts are summarized. Finally, the value obtained by deleting those two parts is the value of the element extracted. Unwanted features are discarded through a classifier and the features that make it to the last stage of the classifier are usually accepted as a face

III. APPLICATIONS OF DEEP LEARNING VIA COMPUTER VISIONIN BIOMEDICAL DOMAIN

The paper stated here reviews eight papers which are brief about various deep learning computer vision techniques applied to healthcare and biomedical fields. The overview of these papers will help us understand various applications of various different algorithms used in bio medical and health care via computer aided softwares or techniques .

Ryad Zemouri,et. al, [2]conducted a deep survey regarding deep learning applications in biomedical fields of - Omics, Bio and medical imaging , Brain and body machine interfaces Most effective Deep Learning Model used – Convolutional Neural Network (CNN) and Generative Adversarial Networks (GANs).] Zengqi Yue , et. al, [3] stated development and classification of samples of blood plasma to detect ovarian cancer.. Laser-induced breakdown spectroscopy (LIBS) - a machine learning data treatment process was used here to test the fingerprint of human blood plasma using back-propagation neural network (BPNN) with accuracy of 71.4% to 86.5% .Junfeng Gao , et. al , [4] focuses on 4 major domains namely medical image analysis , computer vision for predictive analysis and therapy , fundamental algorithms for medical images. Rapid prototyping and 3-d modeling techniques such as CT and MRI have shown great results in diagnosis and treatment of

diseases. Deep learning algorithms used in this paper are convolutional neural networks(CNN)

,stacked autoencoders(SAN) and deep neural networks (DNN) . Esteva, A , et. al [5]stated a survey focusing on the applications of computer vision techniques on static medical imagery ,cardology , radiology, pathology, ophthalmology, and dermatology are growing fields in which high structured images are required. According to the authors , in the fields of ophthalmology CNNs can accurately diagnose or predict thediabetic macular edema,, glaucoma, age-related macular degeneration, manifest visual field loss, childhood blindness, and others.Anwaar Ulhaq , et al ,[6] focuses on Computer Vision is very helping in fighting against COVID-19 , contributing to treatment and management,disease diagnosis, prevention, control,, prognosis . A nBayesian Convolutional Neural Networks (BCCN) model was used to train X-ray images of lungs with COVID-19 cases and gained 88.39% accuracy. Zongwei Zhou, el al,[7] , stated a method called AIFT(active, incremental fine-tuning) which reduces the annotation time and cost while applying CNNs in biomedical imaging . After testing it on 3 biomedical image analysis ,it boosts the performance of CNN. It was applied for classification of lung diseases , automatically retrieving missing information from magnetic resonance imaging.] Shivangi Jain,et al, [8],refers to computer-assisted viewing techniques such as image imaging of the lesion, separation of the skin lesion in the skin region, removal of lesion blob features and feature separation. They are used with computer-assisted software to detect Melanoma Skin Cancer more accurately using image processing tools. Howard Lee,et al, [9],did a deep survey regarding computer vision techniques of image processing for lung , prostate , breast and skin cancer detection . According to the paper , for the detection of breast cancer , image segmentation technique of mammography is used for diagnosis.For prostate cancer detection, SVM with pixel intensity is used for TRUS and for lung cancer Supervised Learning Bayesian network and support vector machine algorithm for X-ray CT.

A. Common Findings :

From all the above papers it can be concluded , the most common and effective deep learning techniques under computer vision are image processing , image segmentation, texture , feature extraction and classification, which have worked wonders in the field of biomedical and health care. For these techniques, CNN is the most accurate algorithm used followed by Support vector Machine , DNNs , etc . These computer vision techniques proved to be very helpful for the clinicians for the diagnosis and treatments of various diseases.

IV. APPLICATIONS OF DEEP LEARNING VIA COMPUTER VISIONIN SECURITY DOMAIN

Computer Vision is used widely in the security domain.Wherever there are CCTV cameras, computer vision comes into the picture. Capturing data through these cameras is the easy part, but identifying and analyzing the images proves to be difficult for the naked eye. Computer Vision systems attempt to describe the contents of a scene in great detail.

I. Aydin and N. A. Othma[10] used IoT in combination with computer vision technique “Haar Cascade Classifier” to detect faces outside a home and send a notification to the homeowner when a face was indeed detected. When motion was detected via a PIR sensor, the PI camera would take pictures and the facial features would be detected and sent to the homeowner.

Sujith B [11] talks about the techniques applied to detect and consequently avoid crimes in an ATM.The core idea behind this was that more than one person is involved in ATM thefts. The proposed method uses Multiple Object Detection (MOD), where the presence of an object is detected through multiple frames using background subtraction, statistical methods, temporal differencing and optical flow.

In [12], the authors try to detect objects like a knife, gun and blood (which are usually part of a crime scene) in an image. The model used consists of Rectified Linear Unit (ReLU), convolutional layer, fully connected layer and dropout function of CNN.CNNs are implemented through Tensorflow. An accuracy of 90.2% was achieved for the dataset used.Similar techniques were used in [13] and an average accuracy of 73.33% was obtained and time taken to detect objects per image was 0.2s

X-ray examination systems have been used to protect areas where access control is important. Security checkpoints have been set up at gates leading to many public places to detect unauthorized items, such as explosives. In general, human operators handle these tasks as the automatic recognition of luggage testing is still very far away. Computer vision assisted X-ray testing can be used to help these operators. [14] attempts to contribute to object recognition in X-ray examination by examining various computer imaging techniques suggested over the years. An accuracy of 95% was obtained using sparse KNNs.

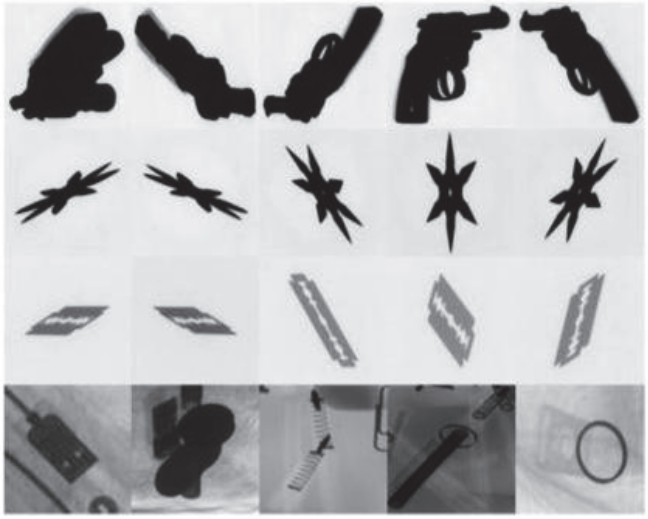

Fig. 3. Some training X-ray images used in [14]

[15] attempts to detect road crimes, where the criminals, after performing the crime, escape quickly in their vehicles. Machine learning techniques such as KNNs, SVMs are used to detect and match faces in the criminal database and an accuracy of 85% was obtained.

Lastly, Jong Sun Kim [16] along with his peers attempt to build an unmanned anti-theft system using a network camera. Two cameras, one inside and one outside the restricted area, are used. Once the outside camera detects motion, To track the attacker, a tracking algorithm using background and histogram completion is used. Scale

Invariant Feature Transform (SIFT) algorithm has been used to determine whether a criminal is stealing valuables or not.

A. Common Findings

Summarizing the content of the above discussed papers, one can find that face detection algorithms such as Haar Cascade Classifier and classification algorithms such as SVM are used extensively in security applications of computer vision. The applications generally have to do with first detecting the criminal, and then identifying the criminal by comparing with a database.

V. APPLICATIONS OF DEEP LEARNING VIA COMPUTER VISIONIN EDUCATION

Online Education is one of the best learning platforms through which we can learn tremendous things from any part of the world. While taking any Education or course or any degree we need to give the exams, as giving Exams online is a big challenge for everyone. In [17] they had proposed a Multimedia Analytics System which performs the task of Automatic Exam online proctoring[1]. In this the system hardware includes mostly Webcam and Microphone to monitor visual and acoustic background of the candidate who is giving that exam. This complete system of proctoring includes mostly 6 basic components which continuously analyses the key behaviour of thatcandidate.

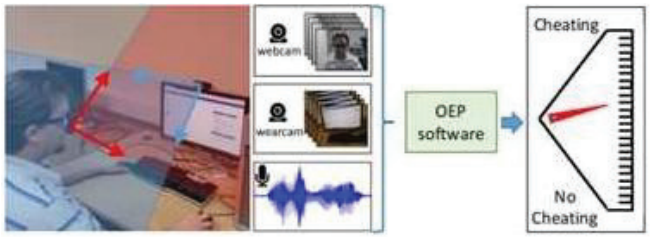

Fig. 4. Representation of OEP and its components

These six basic features include - user authentication, text acquisition, voice recognition, active window acquisition, eye measurement and phone recognition. By looking at all of these things together and using a temporary temporary window, they had designed high-quality features to distinguish whether the candidate was cheating at any time during the test or not. reads and analyzes the result of whether the candidate is cheating or not. The method of self-examination of handwritten Essays for baptismal candidates is time-consuming and tedious. Therefore in view of this, [18] The automation of the handwriting work is done. The main purpose of this is to give marks and scores to the student compared to human scorers and by combining these two AI technologies called Optical Handwriting Recognition and Automated Essay score. In this case the first AI technology of handwriting recognition recognizes student handwriting and the second AI technology of Automated Essay Scoring performs the task of awarding marks according to various topics. OHR and AES are used for this monitoring and evaluation analysis.

Nowadays we make extensive use of digital technology to read and deliver modern content, connect teachers and students around the world, and we can at any time, any learning is increasing, but keeping students focused on this technology-based learning is a major challenge. [19] existing measles engagement in technology mediation learning. They have identified the strengths and limitations and constraints of existing objects and explored the various possible ways to improve the measurement of student involvement. at that moment, it reaches the level of the perfect student experience at school. They used this to analyze skills and show details to improve student engagement. Articles are coded for this category of research content and no one participates in the study, the type of subject matter involved, the location of the study material and the technology used. The use and level of engagement were measured using Skinner and Pitzer framework[2012].

Proper identification of students' learning environments can provide a great deal of information for teachers to analyze and identify problematic students in real time. In [20] To improve this accuracy of learning engagement detection,they first collected two aspects of student behavior data when initiating facial data (using Local Gray Code Patterns for facial recognition) and mouse communication. They also developed a new algorithm for obtaining learning engagement based on collected data (student behavior), from cameras and a mouse connected to a Pc or Laptop in this online learning environment. The main methods of extracting the objects taken from this [20] were to convert the sequence of images into a video containing the students' facial expressions and snapshots and then to the facial features using machine learning techniques. The traditional Local Binary Pattern (LBP) algorithm has been used to analyze this facial expressions. They had also used Complete Local Binary Pattern(CLBP) in some cases as this is more comprehensive than LBP.

A. Common Findings :

From all the above papers we had seen common algorithms which mostly work on face data and voice recognition which are further used in exams to examine and moderate the students. Online Education is our future so for complete analysis of candidates through online mode helps the teacher to get complete detailed analysis of candidates which can further lead to improvement of the candidate. LBP is mostly used to study face recognition patterns.

VI. CURRENT TRENDS

In today's era latest technologies are mostly based on computer vision in different aspects. Image Preprocessing and Video Interpolation are mostly used in different applications and various new technologies are based on this.In [21] they proposed motion measurement and compensation-driven network-driven video frame integration network. A new compatible layer was created to integrate both flow particles and composites to separate the directing frames into it. In this technology they had added virtual frames in between the existing two frames to increase the fps of that video. In such a way a 30 fps video can be converted to 60 and also in some cases 120fps by using this MEMC Technology. This technology is also used in gaming field to enhance the gaming experience

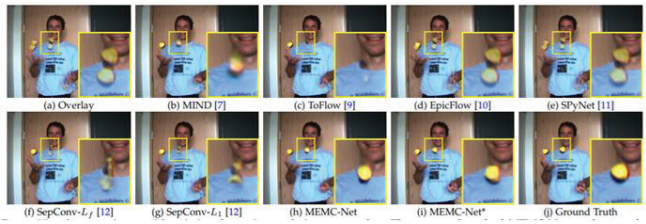

Fig. 5. Representation of main image and virtualframes added in it

Movement vectors and compensation filters are rated by CNN. They also suggest a layer of flexible warping, meaning visual frames are based on optical filters and compensation filters to make new pixels and frames. Finding lost colors in underwater images was a major challenge. In [22] they had proposed an original color retrieval method with their revised model, using RGBD images. They have suggested a visual detection measurement using very dark pixels and their known width. More than 1100 images from two different bodies of water are used in comparison with those of the basic images in space. Sea-thru is the first algorithm to use an underwater model as it has the great advantage of having a comprehensive map.

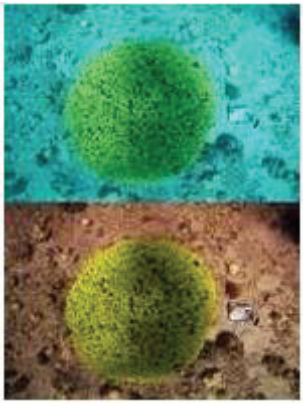

Fig. 6. Representation of original image and imagewith complete removal of water in it.

In [23] they had proposed a method which takes input as a set of images of a scene which has different lighting conditions and it produces the output as 3D representation that can be rendered from new viewpoints under different lighting conditions. They had also performed alternatives for recovering relightable 3D Scenes representations, and also various alternatives which performs well in complex lighting environments. This technique of creating 3D realistic models from 2D images allows anyone to use real world objects in game development, photography and different video creators.

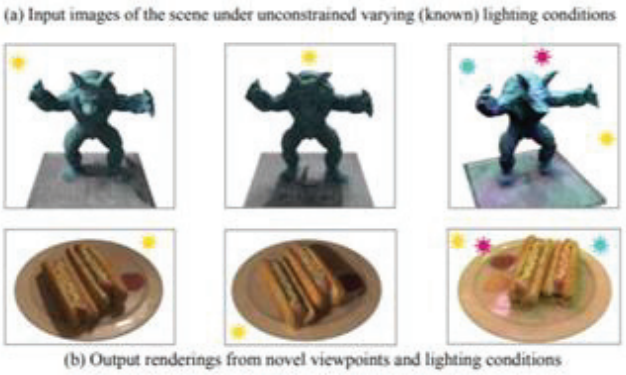

Fig. 7. Representation of original image and itsconversion to 3D image

In [24] the main aim of the author in this paper was to construct a high-resolution (HR) image from a corresponding low resolution (LR) image input. They had proposed an alternative formulation based on creating realistic original super resolved images which downscale correctly. They had used CNN Algorithm for conversion of LR image to HR image.

Common Findings:- From the above papers we had seen that in latest technology trends CNN is the most common Algorithm used in almost every field. As CNN is having high accuracy its also used in gaming technologies. The working on different frames is the main motive for future scope too as most of the things in Computer vision are based on frames.

VII. CONCLUSION

Deep Learning is the most widely unfurling field of machine learning. Computer Vision, a part of deep learning had lead great applications in this today's modern era. Computer Vision is over 50 years old. Although as a research field it has been contributing to many challenging and exciting problems, depending on successful engineering applications. Today and also in future computer vision is going to be very useful in many domains. Recently, though, a number of exciting programs have appeared where great contributions can be made. There were also tremendous researches using this DNN algorithm in energy fields for higher efficiency and accurate results. CNN is the most used Computer Vision Algorithm. It is used in various applications and also this algorithm has increasing accuracy.

REFERENCES

[1] Kersting Kristian , Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines Frontiers in Big Data 1 2018, 2018.0000610.3389

[2] Zemouri, R.; Zerhouni, N.; Racoceanu, D. "Deep Learning in the Biomedical Applications: Recent and Future Status." Appl.Sci. 2019, 9,1526.

[3] Zengqi Yue, Chen Sun, Fengye Chen, Yuqing Zhang, Weijie Xu, Sahar Shabbir, Long Zou, Weiguo Lu, Wei Wang, Zhenwei Xie, Lanyun Zhou, Yan Lu, and Jin Yu, "Machine learning-based LIBS spectrum analysis of human blood plasma allows ovarian cancer diagnosis," Biomed. Opt. Express 12, 2559-2574 (2021)

[4] .JOUR Gao, Junfeng Yang, Yong Lin, Pan Park, Dong Sun , 2018 , Computer Vision in Healthcare Applications 5157020 2018 2040-2295 https://doi.org/10.1155/2018/5157020,Journal of Healthcare Engineering , Hindawi

[5] Esteva, A., Chou, K., Yeung, S. et al. Deep learning-enabled medical computer vision. npj Digit. Med. 4, 5 (2021)

[6] Anwaar Ulhaq and Asim Khan and Douglas Gomes and Manoranjan Paul, Computer Vision For COVID-19 Control: A Survey, 2020, 2004.09420

[7] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, Jianming Liang; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7340-7351

[8] Shivangi Jain, Vandana jagtap, Nitin Pise, Computer Aided Melanoma Skin Cancer Detection Using Image Processing,

[9] Howard Lee, Yi-Ping Phoebe Chen,Image based computer aided diagnosis system for cancer detection,Expert Systems with Applications,Volume 42, Issue 12 2015,Pages 5356-5365,ISSN 0957-4174

[10] I. Aydin and N. A. Othman, "A new IoT combined face detection of people by using computer vision for security application," 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1-6, doi: 10.1109/IDAP.2017.8090171.

[11] Sujith B / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6068-6071

[12] M. Nakib, R. T. Khan, M. S. Hasan and J. Uddin, "Crime Scene Prediction by Detecting Threatening Objects Using Convolutional Neural Network," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465583.

[13] TY - BOOK, AU - Saikia, Surajit, AU - Fidalgo, Eduardo, AU - Alegre, Enrique, AU - Fernández-Robles, Laura, PY - 2017/09/15, SP - 14, EP - 24, SN - 978-3-319-68547-2, T1 - Object Detection

for Crime Scene Evidence Analysis Using Deep Learning, DO - 10.1007/978-3-319-68548-9_2

[14] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra and S. Banerjee, "Modern Computer Vision Techniques for X-Ray Testing in Baggage Inspection," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 4, pp. 682-692, April 2017, doi: 10.1109/TSMC.2016.2628381.

[15] TY - CHAP, AU - Jain, Rachna. AU - Nayyar, Anand, AU - Bachhety, Shivam, PY - 2020/01/01, SP - 503. EP - 516, SN - 978-981-32-9948-1, T1 - Factex: A Practical Approach to Crime Detection, DO - 10.1007/978-981-32-9949-8_35

[16] TY - JOUR, AU - Kim, Jong, AU - Yeom, Dong, AU - Joo, Young Hoon, AU - Park, Jin, PY - 2010/10/01, SP - 967, EP - 974, T1 - Intelligent Unmanned Anti-theft System Using Network Camera. VL-8, DO - 10.1007/s12555-010-0505-0, JO - International Journal of Control, Automation and Systems

[17] Automated Online Exam Proctoring, Yousef Atoum, Liping Chen,Alex X. Liu, Stephen D. H. Hsu, and Xiaoming Liu

[18] On the Automatic Scoring of Handwritten Essays,Sargur Srihari, Rohini Srihari, Pavithra Babu, Harish Srinivasan, Center of Excellence for Document Analysis and Recognition (CEDAR), University at Buffalo, State University of New York Amherst, New York 14228, U.S.A.

[19] Measuring student engagement in technology-mediated learning: A review, Curtis R. Henrie* , Lisa R. Halverson, Charles R.Graham, Brigham Young University, Provo, UT 84602, USA

[20] Engagement Detection via Facial Expression and Mouse Behavior Recognition Technology, Zhaoli Zhang1 , Zhenhua Li1,2, Hai Liu1,Taihe Cao1 , and Sannyuya Liu

[21] MEMC-Net: Motion Estimation and Motion Compensation Driven Neural Network for Video Interpolation and Enhancement, Wenbo Bao, Wei-Sheng Lai, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang

[22] Sea-thru: A Method For Removing Water From Underwater Images, Derya Akkaynak Tali Treibitz, University of Haifa derya.akkaynak@gmail.com, ttreibitz@univ.haifa.ac.il

[23] NeRV: Neural Reflflectance and Visibility Fields for Relighting and View Synthesis, Pratul P. Srinivasan, Google Research, Boyang Deng, Google Research, Xiuming Zhang MIT Matthew Tancik UC Berkeley Ben Mildenhall UC Berkeley Jonathan T. Barron Google Research

[24] PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models, Sachit Menon*, Alexandru Damian*, Shijia Hu, Nikhil Ravi, Cynthia Rudin, Duke University,Durham, NC