IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.6, June 2022 E-ISSN: 2347-2693

Introduction to Computer Vision: An End-to-End Guide for Beginners

Kirti Sharma[1*], Chhaya Gupta[2]

[1,2]Vivekananda School of Information Technology, Vivekananda Institute of Professional Studies-Technical Campus, Delhi, India

[*]*Corresponding Author: kirtisharmaa.11@gmail.com, Tel.: +91-9899193941*

DOI: https://doi.org/10.26438/ijcse/v10i6.3236 | Available online at: www.ijcseonline.org

Received: 21/May/2022, Accepted: 09/Jun/2022, Published: 30/Jun/2022

Abstract— Images are very easy to understand and remembered by humans. The human brain can understand if the image belongs to a dog or a cat by just having a look at it. Computer vision is one of the subjects in artificial intelligence that made it possible for a machine to visualize objects like a human brain. Although, the machine can visualize objects like human brain but still the path is so long to reach the accuracy of a human brain. The amount of data we collect today, which is subsequently utilised to train and improve computer vision, is one of the driving drivers behind its rise. Computer vision is the science by which various objects can be detected in fraction of time with the help of neural networks. Early computer vision investigations began in the 1950s, and by the 1970s, it was being used commercially to discern between typed and handwritten text. Today, computer vision applications have evolved tremendously. In this paper, a general introduction to computer vision is provided with an understanding of all the concepts of computer vision. A basic neural network is implemented from scratch for the same. The basic neural network developed achieves an accuracy of 89.13% when trained on Fashion MNIST dataset.

Keywords— Computer Vision, Image Processing, bounding Boxes, Object Detection, Image Segmentation

I. INTRODUCTION

The term “computer vision”, itself is self-explanatory, it is concerned with extracting information about an image by analysing that image. All the images are made up of pixels and every pixel has a RGB value. The computer visualises these pixels as numbers. Every image has a label that helps identify an image like, dog, cat, apple, mango and so on. When a person teaches a baby about a dog, the dog's picture is shown to the baby and the baby tries to remember that this image belongs to a dog.

Similarly, computer is taught with images of different objects and it memorises that the image belongs to which object. This mechanism is repeated several times to make the computer recognise the images with certain confidence or accuracy. The whole concept is known as Computer Vision.

In today’s world, computer vision has an enormous number of applications in different areas such as remote sensing, radiology, document processing, microscopy, robot guidance and many more. Computer vision has dual goals, as it provides computational models and autonomous systems for human visual system.

Larry Roberts is father of computer vision, who arrived at the likelihood of extraction of 3D geometrical data from 2D perspective[1]. But researchers realised that it was required to analyse real world images. David Marr proposed a framework using a bottom-up approach to understand any image/scene [2]. Furthermore, the framework proposed was able to process 2D images. Finally, techniques were evolved to analyse 3D objects in a given scene.

Computer vision has been merged with other related fields like machine learning [3], image processing [4], [5], photogrammetry [6], object poses determination [7], computer graphics [8] and others. Many researchers feel that computer vision is a difficult task as it is always compared with a human visual system which is exceptionally strong for many tasks. For example, a human can identify face under all conditions like illumination, viewpoint, expressions, occlusions etc. Humans are so sharp in recognising any human in a photograph that has been taken so many years ago, and there is no limit on how many things, objects, and faces a human brain can store in future recognition. The open challenges are endless in this field, and researchers worldwide are working on the task to overcome some of these limitations.

In this paper, the authors provided details about basic concepts of computer vision and proposed a simple convolutional neural network-based framework for computer vision. Computer vision has a bigger scope and is used in all the aspects of routine life, but without the basic concepts of state-of-art, no individual can enhance the field with their innovations. The study helps in understanding the basic necessities of computer vision that one must know. The paper can act as a beginner guide for an amateur in the state-of-art.

II. RELATED WORK

There has been a lot of study done worldwide in the computer vision field. In this section, the authors are trying to provide a brief outline of the work in state-of-art field.

Chhaya Gupta et al. [9] proposed a two phase model for detecting face-masks in the pandemic situation of COVID-19. The authors used CNN as the base for the model proposed. The research determines whether or not a person is wearing a mask in real time. A red bounding box surrounds the human face if he/she is not wearing a mask. Andre Esteva et al. [10] provided a deep learning enabled computer vision review for medical imaging. The authors provided a review in different fields of medical imaging like ophthalmology, radiology, pathology, and dermatology.

Shuyuan Xu et al. [11] provided a critical review in the field of architecture and engineering with the help of computer vision. Construction sites are a tedious task to be managed and monitor due to clutter and disorder, and hence authors have done extensive research in the state-of-art methods and their applications. Chuan-Zhi Dong [12] gives a broad review of computer vision concepts, methodologies, and real-world applications in the field of healthcare. The authors reviewed 2D computer vision structural health monitoring applications.

Brian H.W.Guo et al. [13] adopted a 5-step review approach and presented a critical review in state-of-art methods in safety science and management in construction. To identify computer vision applications in the construction industry, the authors suggested a three-level computer vision development framework. Junyi Chai et al. [14] gave a critical assessment of recent advances and applications in computer vision. In four cases, they found eight techniques: recognition, visual tracking, semantic segmentation, and image restoration.

Umair Iqbal et al. [15] presented a review on different computer vision techniques which are useful in flood management. The authors provided a systematic review and proposed a need-based evaluation of all the applications. Innocent Nyalala et al. [16] provided a systematic study of using computer vision to evaluate the weight and volume of poultry and goods.

Chen Chen et al. [17] focuses on computer vision methods for behaviour recognition in pigs and cattle. The authors worked on recognising their behaviours which might be helpful for their health and welfare like aggression, drinking habits, mounting, lameness, tail biting, eating habits, and nursing. Guoming Li et al. [18] provide a thorough evaluation of CNN-based computer vision systems' applications in animal husbandry was conducted, with the research divided into five categories: picture classification, object detection, semantic/instance segmentation, pose estimation, and tracking.

Baid et al. [24] have utilised many different sorts of neural network models to demonstrate which neural network delivers the best accuracy and efficiency in the recognition of food photos. They have worked with a food image dataset (food-11), which includes 16643 photos. Baid et.al [25] in their other paper proposed a method for classifying food photos in this research. For the food image classification, they have employed pre-trained models. The convolutional neural network is used in the pre-trained models. CNNs are extremely good in image classification and other computer vision problems in neural networks. In their experiment, researchers categorised a food image dataset called food11 with an accuracy of 96.75 percent.

Zhangnan Wu et al. [19] presented a review on weed detection methods with the help of computer vision. The review has been elaborated on methods based on deep learning and standard image processing methods Ling Yang et al. [20] presents an evaluation of computer vision methods for fish detection and their behaviour analysis. The authors also discussed 2D and 3D image recognition systems. The study also reviews motion-based, appearance-based, and deep learning-based computer vision models for intelligent aquaculture. Tanzila Saba [21] presents a comparative study on skin cancer diagnosis methods with handcrafted and non-handcrafted features using computer vision.

Anuja Bhargava et al. [22] presents a detailed review of different pre-processing, segmentation, feature extraction and classification techniques for quality of vegetables and fruits on the basis of colour, shape, size, texture, and defects. In addition, the authors have provided a comparison of different algorithms. Claudio Michael Louis et al. [23] presented a comprehensive review of computer vision applications in the realm of in vitro fertilisation for improvement in evaluating the development of embryo.

III. COMPUTER VISION

Computer vision helps in computers and smartphones in visualising the world. The phenomenon of re-creating a human eye started in the early 50's and there's been a lot of study done till now. The human eye is a complex structure and is equally complex in understanding [1]. Therefore, it is the most tedious task to make a machine visualise things the same way a human eye does. Computer vision is mainly categorised into four tasks:

Image Classification: Image classification helps in identifying an image belongs to which class whether it's a dog, cat, lion, human face etc.
Object Detection: Object detection helps detect more than one object in a scene/image. It helps in not only classifying objects but also in localising objects too.
Semantic Segmentation: Semantic segmentation helps in identifying objects into different classes such as animals, humans, bikes, cars etc.

Instance Segmentation: Instance segmentation is a complex version of semantic segmentation. Instance segmentation helps in classifying an object within a class into different objects of that class, like, if an object belongs to animal class, the instance segmentation helps in classifying object as dog, cat, lion etc. In many cases, instance segmentation is not used because it is an

expensive approach. If the problem is simple, then its always better to keep things simple and fast. If the problem is to detect a ball in an image, then it is not preferred to use instance segmentation.

To make a machine visualise things just like a human eye, the following steps have to be followed:
Step 1: The colours are represented in terms of hexadecimal numbers. Hexadecimal number is the only way to make a machine understand about colours of an image.
Step 2: Image segmentation is done to distinguish foreground and background colours of an image. Colour gradient techniques are used to find edges of different objects.
Step 3: Images are churned up for more information regarding other features such as corners. Corners are the features known as building blocks and help gather more information contained in an image.
Step 4: To identify the image accurately, texture in the image is determined. Differentiating texture between two objects makes it easier for a machine to identify an object.
Step 5: Once all the steps have been implemented, the machine makes a nearly-right prediction about an object's class and then matches the image to those present in the dataset.
The complete procedure for computer vision is represented in figure 1.

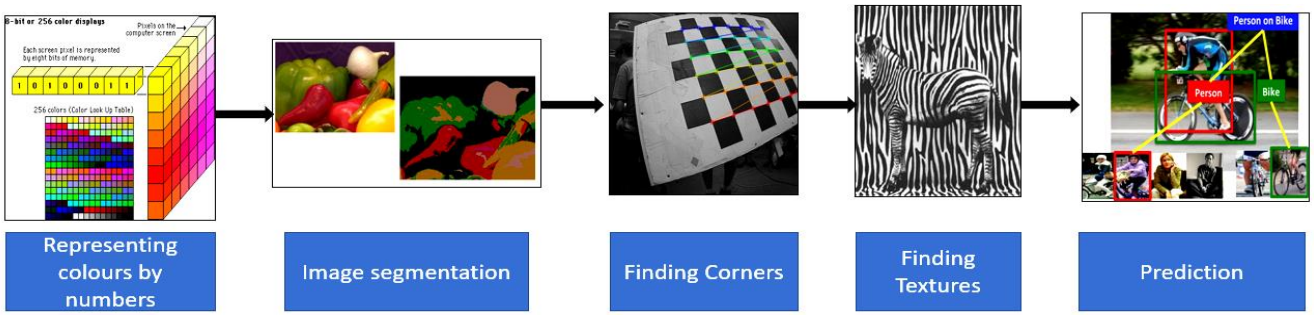

Figure 1. Basic steps followed in Computer Vision

IV. METHODOLOGY USED

Computer vision has a simple rule of gathering the images, transforming the images in the form of NCHW (Number of batches, Channels, Height, Width), preparing the model and finally classifying the images into appropriate classes. In this study, fashion MNSIT dataset is being used which is freely available on Kaggle repository. The dataset consists of a training set of 60000 images of different apparel and a testing set of 10000 images of different apparel. Each image is a 28x28 grayscale image and are categorised into ten classes. In this research, authors have created a new neural network from scratch with a base of LeNet-5 model. The structure of LeNet-5 model is depicted in figure 2.

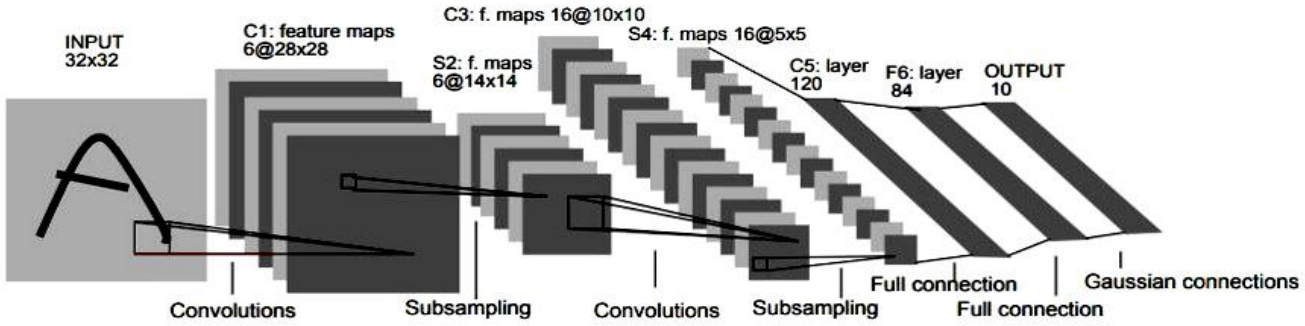

Figure 2. Architecture of LeNet-5 model

LeNet act as the base for creating the new neural network in this study. According to the architecture of this model, the new neural network has been created with two convolutional layers, that are useful for extracting features from the image. Two Maxpooling layers have been added in the neural network, that are responsible for reducing a number of parameters to be trained. Max pooling layers are added after every convolutional layer. A flatten layer is added to flatten the data into one dimensional data. This one-dimensional data is passed as an input to the fully connected layer or the dense layer. In this new network, there are three fully connected layers. Finally, a loss function is determined to check whether the neural network is performing well or not. In this study, Cross-entropy loss function paired with softmax is used to squeeze the error values between o and 1. The summary of the model created is shown in figure 3.


```
<bound method Block.summary of Sequential(
  (0): Conv2D(1 -> 6, kernel_size=(5, 5), stride=(1, 1), Activation(relu))
  (1): MaxPool2D(size=(2, 2), stride=(2, 2), padding=(0, 0), ceil_mode=False, global_pool=False, pool_type=max, layout=NCHW)
  (2): Conv2D(6 -> 16, kernel_size=(3, 3), stride=(1, 1), Activation(relu))
  (3): MaxPool2D(size=(2, 2), stride=(2, 2), padding=(0, 0), ceil_mode=False, global_pool=False, pool_type=max, layout=NCHW)
  (4): Flatten
  (5): Dense(400 -> 120, Activation(relu))
  (6): Dense(120 -> 84, Activation(relu))
  (7): Dense(84 -> 10, linear)
)>
```

Figure 3. The summary of the newly created neural network

For evaluating the neural network, accuracy metric has been used and the new neural network achieved an accuracy of 89.68%. the model has achieved the accuracy after ten epochs. The gathered data was splitted in the ratio of .2.

V. COMPARATIVE RESULTS

Python is used to implement the underlying work. The LeNet function, which is available in TensorFlow's Keras package, is used to extract the features. The experiment is split into two sections: training and testing. The characteristics are taken from the training examples and fed into the fully connected layer, which is utilised as the classifier, during the training phase. The classifier uses these extracted features, which are stored in LeNet's knowledge base, to produce predictions. The classifier's predictions are confirmed by the test photos.

The proposed LeNet based architecture has got the 89.13% accuracy on test images (Table 1). Table 2 depicts the accuracy analysis. Figure 4 shows few of the accurately classified images. All the images are predicted correctly.

Table 1 Confusion matrix for LeNet based architecture

Class	Label	0	1	2	3	4	5	6	7	8	9
T-Shirt/top	**0**	**912**	0	9	3	2	0	72	0	3	0
Trouser	**1**	0	**996**	0	1	0	0	1	0	1	0
Pullover	**2**	24	1	**910**	5	26	0	35	0	0	0
Dress	**3**	18	6	7	**916**	17	0	34	0	1	0
Coat	**4**	1	0	20	9	**945**	0	23	0	0	0
Sandal	**5**	0	0	0	0	0	**893**	0	5	1	1
Shirt	**6**	91	0	35	8	74	0	**789**	0	3	0
Sneaker	**7**	0	0	0	0	0	6	0	**980**	0	9
Bag	**8**	1	1	0	2	0	1	0	0	**993**	0
Ankle Boot	**9**	0	0	1	0	0	9	0	38	0	**952**

Table 2 Accuracy Analysis of fashion MNIST database with various methods

Methodology	Accuracy (%)
HOG + SVM	86.53
ANN + SVM	80.72
LeNet	89.13

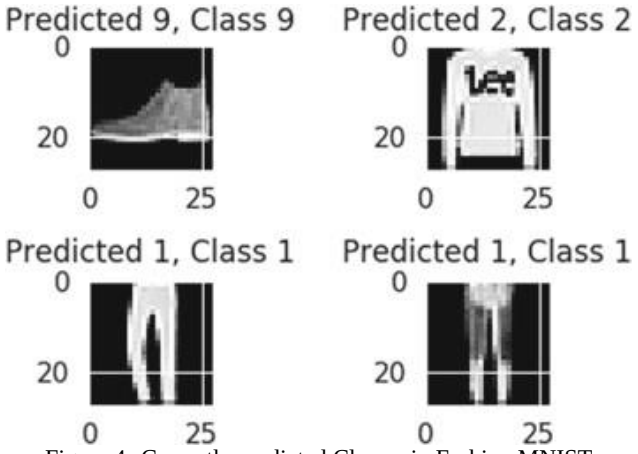

Figure 4. Correctly predicted Classes in Fashion MNIST

VI. CONCLUSION AND FUTURE SCOPE

The primary goal of this study is to provide guidance to newcomers to computer vision on the correct path of how to begin and where to begin. This paper taught about the basics of computer vision, what it is, how anyone can start working in this field and finally how to create a new neural network on the basis of a pre-trained neural network. For this, Fashion MNIST dataset has been used that consists of 60000 images of apparels for training purpose and 10000 images for testing purpose and the dataset is available freely on Kaggle repository. The dataset was analysed with a new neural network which was created on the basis of LeNet-5 model. The experimental results make it clear that the new neural network created achieves an accuracy of 89.68%; in the future, the neural network can be further enhanced according to the new requirements.

REFERENCES

[1] L. G. Shapiro, "Computer Vision : the Last Fifty Years," **pp. 1–8, 2015.**

[2] D. Marr, Part I: Introduction and philosophical preliminaries. 2010.

[3] A. A. Khan, A. A. Laghari, and S. A. Awan, "Machine Learning in Computer Vision : A Review," pp. 1–11.

[4] A. Bhargava and A. Bansal, "Novel coronavirus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review," Multimed. Tools Appl., **vol. 80, no. 13, pp. 19931–19946, 2021**, doi: 10.1007/s11042-021-10714-5.

[5] R. Sohail et al., "A review on machine vision and image processing techniques for weed detection in agricultural crops," Pakistan J. Agric. Sci., **vol. 58, no. 1, pp. 187–204, 2021**, doi: 10.21162/PAKJAS/21.305.

[6] R. Qin and A. Gruen, "The role of machine intelligence in photogrammetric 3D modeling–an overview and perspectives," Int. J. Digit. Earth, **vol. 14, no. 1, pp. 15–31, 2021**, doi: 10.1080/17538947.2020.1805037.
[7] Z. Fan, Y. Zhu, Y. He, Q. Sun, H. Liu, and J. He, "Deep Learning on Monocular Object Pose Detection and Tracking: A Comprehensive Overview," ACM Comput. Surv., **vol. 1, no. 1, 2021**, doi: 10.1145/3524496.
[8] F. Okura, "3D modeling and reconstruction of plants and trees: A cross-cutting review across computer graphics, vision, and plant phenotyping," Breed. Sci., **vol. 72, no. 1, pp. 31–47, 2022**, doi: 10.1270/jsbbs.21074.
[9] C. Gupta and N. S. Gill, "Coronamask: A face mask detector for real-time data," Int. J. Adv. Trends Comput. Sci. Eng., **vol. 9, no. 4, pp. 5624–5630, 2020,** doi: 10.30534/ijatcse/2020/212942020.
[10] A. Esteva et al., "Deep learning-enabled medical computer vision," npj Digit. Med., **vol. 4, no. 1, pp. 1–9, 2021**, doi: 10.1038/s41746-020-00376-2.
[11] S. Xu, J. Wang, W. Shou, T. Ngo, A. M. Sadick, and X. Wang, "Computer Vision Techniques in Construction: A Critical Review," Arch. Comput. Methods Eng., **vol. 28, no. 5, pp. 3383–3397, 2021**, doi: 10.1007/s11831-020-09504-3.
[12] C. Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," Struct. Heal. Monit., **vol. 20, no. 2, pp. 692–743, 2021**, doi: 10.1177/1475921720935585.
[13] B. H. W. Guo, Y. Zou, Y. Fang, Y. M. Goh, and P. X. W. Zou, "Computer vision technologies for safety science and management in construction: A critical review and future research directions," Saf. Sci., **vol. 135, no. January, p. 105-130, 2021**, doi: 10.1016/j.ssci.2020.105130.
[14] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," Mach. Learn. with Appl., **vol. 6, no.** August**, pp. 100-134, 2021**, doi: 10.1016/j.mlwa.2021.100134.
[15] U. Iqbal, P. Perez, W. Li, and J. Barthelemy, "How computer vision can facilitate flood management: A systematic review," Int. J. Disaster Risk Reduct., **vol. 53, no. January, p. 102030, 2021**, doi: 10.1016/j.ijdrr.2020.102030.
[16] I. Nyalala, C. Okinda, C. Kunjie, T. Korohou, L. Nyalala, and Q. Chao, "Weight and volume estimation of poultry and products based on computer vision systems: a review," Poult. Sci., **vol. 100, no. 5, p. 101072, 2021**, doi: 10.1016/j.psj.2021.101072.
[17] C. Chen, W. Zhu, and T. Norton, "Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning," Comput. Electron. Agric., **vol. 187, no.** January**, p. 106255, 2021**, doi: 10.1016/j.compag.2021.106255.
[18] G. Li et al., "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review," Sensors, **vol. 21, no. 4, pp. 1–42, 2021**, doi: 10.3390/s21041492.
[19] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," Sensors, **vol. 21, no. 11, pp. 1–23, 2021**, doi: 10.3390/s21113647.
[20] L. Yang et al., Computer Vision Models in Intelligent Aquaculture with Emphasis on Fish Detection and Behavior Analysis: A Review, vol. 28, no. 4. Springer Netherlands, 2021.
[21] T. Saba, "Computer vision for microscopic skin cancer diagnosis using handcrafted and non-handcrafted features," Microsc. Res. Tech., **vol. 84, no. 6, pp. 1272–1283, 2021**, doi: 10.1002/jemt.23686.
[22] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," J. King Saud Univ. - Comput. Inf. Sci., **vol. 33, no. 3, pp. 243–257, 2021**, doi: 10.1016/j.jksuci.2018.06.002.
[23] C. M. Louis, A. Erwin, N. Handayani, A. A. Polim, A. Boediono, and I. Sini, "Review of computer vision application in in vitro fertilization: the application of deep learning-based computer vision technology in the world of IVF," J. Assist. Reprod. Genet., **vol. 38, no. 7, pp. 1627–1639, 2021**, doi: 10.1007/s10815-021-02123-2.
[24] Y. Baid and A. Dhole, "Food Image Classification Using Machine Learning Techniques: A Review," *Int. J. Comput. Sci. Eng.*, **vol. 9, no. 7, pp. 11–15, 2021**, doi: 10.26438/ijcse/v9i7.1115.
[25] Y. Baid and A. Dhole, "Food Image Classification Using Deep Learning Techniques," *Int. J. Comput. Sci. Eng.*, **vol. 9, no. 7**, **pp. 11–15, 2021**, doi: 10.26438/ijcse/v9i7.1115.

AUTHORS PROFILE

Kirti Sharma is designated as Assistant Professor in Vivekananda Institute of Professional Studies-Technical Campus, Delhi. She has completed her MCA from Maharshi Dayanand University, Rohtak in 2006. She completed her graduation from University of Delhi in 2003 in the stream of computer applications. She is a research scholar pursuing PhD in Computer Science and Applications from Lovely Professional University, Punjab.

Chhaya Gupta is designated as Assistant Professor in Vivekananda Institute of Professional Studies-Technical Campus, Delhi. She has completed her MCA from Guru Gobind Singh Indraprastha University, Delhi in 2012. She completed her graduation from Delhi University, Delhi in 2009 in the stream of computer Science. She is a Gold Medallist from GGSIPU, Delhi. She qualified her UGC NET exam in January, 2019. She is a research scholar pursuing PhD in Computer Science and Applications from Maharshi Dayanand University, Rohtak.

