DESIGN AND IMPLEMENTATION OF COMPUTER VISION USING DEEP LEARNING FOR INSPECTION OF ASSEMBLY SYSTEMS

[1]Donthula Rahul, [2]Dr. M. Srinivasa Rao
[1]PG scholar, Department of Mechanical Engineering, JNTUH, Hyderabad.
[2]Professor and Special Officer, Department of Mechanical Engineering, JNTUH, Hyderabad

ABSTRACT

In manufacturing and process industries, to increase the production rate and reduce the defective parts, there's an increased Automation in fabrication and assembly lines. At every station, in the assembly line there might be an issue with Safety of operator while working, Manual Inspection of product for defects or good one and reject the defective parts to the Rejection bin. As a Part of this thesis, the computer vision system for Inspection of parts in production systems has been designed. The Production rate of the designed system has been calculated and compared with the existing system. The accuracies of the different pre-trained deep learning model have been analyzed. Automated Inspection has been performed in assembly of MG996R servo motor and gear rack using deep learning approach of Convolutional Neural Networks (CNN). Deep learning used in solving the problem of choosing the type of features as the weights in a neural network adapt themselves to understand features and maximize the values in the nodes that would indirectly help us in attaining our objective of selecting a good feature. This was effectively used for vision-based applications as it reduces the complexity in designing and selecting the features that are best suitable for a specific task.

Keywords: Computer vision; CNN; Deep learning for Inspection; classification.

I. INTRODUCTION

Deep learning is an Artificial Intelligence (AI) field and machine learning subset that try to imitate neurons function of the neural networks present in human brain through artificial neural networks. Artificial Neural Networks comprising many layers, drive deep learning. Deep Neural Networks (DNNs) are such types of networks where each layer can perform complex operations.

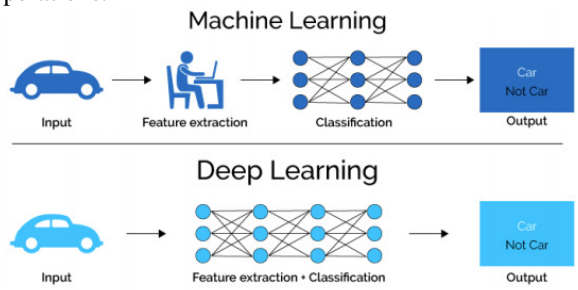


Figure 1. Neurons as layer

Neural Networks are layers of nodes, like neurons in human brain. Nodes within individual layers are connected to nodes of adjacent layers. The network is said to be deeper based on depth of its layers. A single neuron in the human brain receives thousands of signals from other neurons. Similarly, In ANN nodes in intermediate layers receives inputs from many other inputs, signals travel between nodes and assign corresponding weights. A heavier weighted node will exert more effect on the next layer of nodes. The final layer compiles the weighted inputs to produce an output.

The Deep learning concept comes under machine learning subset that had great tendency and flexibility to represent the available data as a nested hierarchy of concepts by learning, with each concept defined in relation to simpler ones, and more abstract representations computed in less abstracted terms.

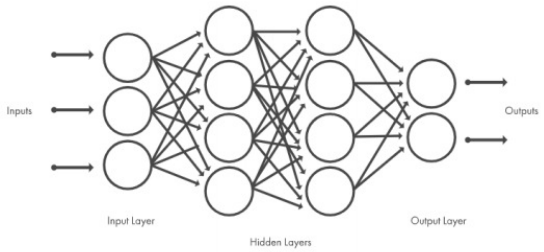


Figure 2. Multilayer Neural network

II. LITERATURE SURVEY

ToubaMalekzadeh et al. [1] Proposed algorithm is able to detect almost all the defects of the aircraft fuselage and reduce the workload of manual inspection significantly. They have reached an accuracy of 96%.

Yundong Li et al. [2] In this paper, they proposed a Fisher criterion-based SDA framework with the objective of learning more discriminative features, which can efficiently inspect defects for both periodic pattern and jacquard pattern fabrics have insufficient negative images data.

Neha Sharma et al. [3] have performed the analysis of CIFAR-100 and CIFAR-10 datasets by each pretrained networks of AlexNet, GoogleNet, ResNet-50. Two different methods were used to calculate the results. The first method only considers 10 most probable classes and the second registers the position of the correct class in the full probability range. It is to be noted that the complex frames often create confusion

for the network to detect and recognize the scene. The results suggested that trained networks with transfer learning performs better than existing ones. It is concluded that, more the number of layers, more will be the training and therefore, higher the rate of accuracy in prediction will be achieved. The hardware requirements may not allow the network to be trained on normal desktop

Prof. NilimaBargal et al. [4] In this project report, they have tried to create a setup that will decrease human effort and succeeded to an extent by using the low-cost automation system (LAC) to avoid risk, improve accuracy, increase speed of production and reduce the cycle time. This setup couldn't perform sorting for other physical consideration.

Madhav Patil et al. [5] If comparing to the current inspection systems, the proposed Universal Quality Inspection and Control system implemented and described along this paper can be adapted and used at any stage of the production and manufacturing processes for different types of shapes and size of product inspection with minimum effort. Malfunctioning of equipment might give wrong Feedback to PLC in turn It would Provide Wrong Results.

Denni Kurniawan et al. [6] Results from the tests showed that the system is capable to perform inspection tests at efficiency of 78 bottles per minute. The system assists in enhancing manual visual inspections work to automated visual inspection. Can be used only for one type of inspection.

III. METHODOLOGY

In this project, the image data is captured manually which include various defects of the part and in different lightening.

Procedure

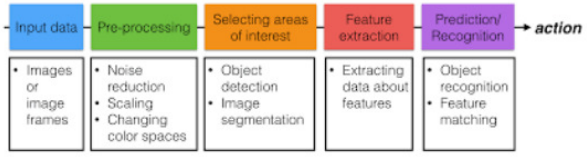


Figure 3. Computer vision pipeline

In this project, when the image is taken, it goes for pre-processing where the size of the image is altered as per the requirement and this process is known as scaling. In few images, the noise that is present in the images can be reduced. The noise in an image adds more unreliability of understanding a feature for a model. Even though a human can detect an object even with a certain amount of noise, but a trained model cannot identify the features due to the presence of noise. The extraction of features is the most important task performed by any model as it helps in identifying and localizing an object of interest and as an image is made up of a set of three matrices of the same size comprising of the values of the red, blue and green values in the image. The red, green and blue colour model is the most basic additive colour model in which colours on adding one above another provide a unique colour which never repeats in this entire three colour boundary. Each value of a colour ranges from zero to two hundred and fifty-five. The zero represents that the value of the colour is dim and as the value gets higher and reaches two hundred and fifty-five the colour becomes bright.

Figure 4. Computer Vision system

All LCD displays can be seen or thought of as an arrangement of arrays of millions of little red, green and blue lights and these lights seem to have their own dimmers. The conversion of the red, green and blue colour model into other colour models or colour spaces. Subtractive models can also be employed and different colour spaces like the HSV, CIELCh and HSL and CMYK can also be used. The HSV and HSL are from the primitives or the red, green and blue but a transformation of that Cartesian system along with the colorimetry and its components to its relative color space. The HSV stands for hue, saturation and value while HSL stands for hue, saturation and lamination. The other variants of this type are HSI-hue, saturation and intensity, HSB-hue, saturation and brightness, HSD-hue, saturation and darkness.

The project is implemented by taking the input image from the camera in the form of a matrix with numbers ranging from one to two hundred and fifty-five. The image as a matrix of numbers goes as an input into the detection model which is made up of a number of layers such as convolution and maxpooling. A convolution layer is a layer which has multiple filters passing on the image in a sliding window format. In a sliding window, the filter of a small square matrix passes on the entire image skipping few positions horizontally and vertically in a specific order. Filters are square matrices which have fixed position values and these when slide on an image go through elemental operations and give out a single output that gets stacked into an image with reduced dimensions. These filters are not a fixed set of filters as they are learned in the process of backpropagation. In the maxpooling layer the image matrix is reduced to a dimension of a scale of the image size. This layer takes the maximum of a small set of square matrices in an image and adds it into the next

layer as input, thus providing the reduced image. The process of learning is of three types like supervised, unsupervised and reinforcement learning. In the supervised learning the input data is given along with the annotation data or the label data. The label data tells what the output of the input data is. The unsupervised learning is one the most important form of learning in which the learning is done without providing any input annotations or label data to the inputs. The model extracts feature such as color, shape, structure, etc., and separates the data or learns to identify the differences between the inputs. The model learns on its own to separate the data. In reinforcement learning the model is partially trained through supervised data and then trained using unsupervised data to generalize the model to all types of situations. The model learns by the process of backpropagation. In the backpropagation, all the filters with the values ranging between zero and one are changed depending upon the error. The error is identified when the image is forward propagated and the real output with the expected output is calculated as the error. Depending upon the error obtained from the forward propagation, the change in weights is obtained by applying the delta rule with respect to the weights. The image passes through this learned model and then generates detections. The detections consist of the scores, bounding boxes, classes, and the number of detections. The scores consist of the accuracy that the model feels that an object is exactly a 13 object whereas the bounding boxes tell where the object with that score is in the image and the classes tells what object it is. The detector provides the output of the object if an object exists in the image along with the location of the object on the image. Then preprocessing is done on the image to extract the object from the image and then the object image is fed into a classifier which consists of convolution layers and maxpooling layers. The output of the classifier is a SoftMax output that puts out the class of the object. The predicted class of the object signals the PLC which controls the Pneumatic cylinder actuation.

IV. EXPERIMENTAL SETUP

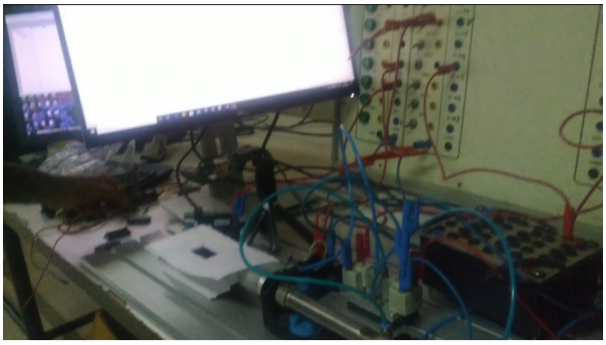


Figure 6. Experimental setup

This project requires a hardware which is computationally fast and can operate huge levels of computations without any failure. The commercially available products that are suitable for this application are raspberry pi 4b and Jetson nano. The raspberry pi 4 has Cortex A72 CPU that processes at the speed of 1.5GHz whereas the Jetson nano with Cortex A57 CPU processes at the speed of 1.42GHz. The Jetson nano of the same variant as that of the raspberry pi costs as twice as that of the raspberry pi 4b. The Raspberry pi 4b is desirable as it has high speed, capability of maximizing through output of mathematical models and low cost.

In the modern world, with the implementation and introduction of new neural network architectures by AI practitioners and researchers which increase the accuracy and provide better performance. With the existence of a wide variety of AI models and frameworks to perform deep learning or neural networks. So, the Raspberry pi 4b is highly desirable for AI applications.

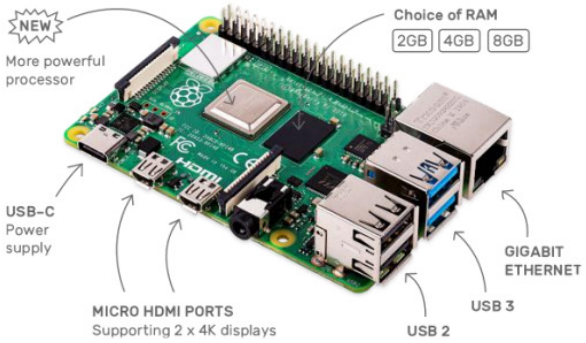


Figure 7. Raspberry pi 4b

A PLC (Programmable Logical Controller) is a ruggedized computer used for industrial automation. These controllers can automate a specific process, machine function, or even an entire production line. For any typical industry PLC used is with following specifications: RAM: 512 kb, Digital I/O Module: 20, Analog I/O Module: 20, Power supply: 24 V DCSatisfying above specifications SIMATIC S7-400 designed by Siemens is available.

Figure 8. SIMATIC S7-400 by siemens

Selection of Camera is based on the Megapixel count, Frame Rate, Auto focus, size. For any typical Deep

learning Tasks Camera used is with following specifications with available GPU s.Image Resolution: 640 X 480 p is required.Satisfying above specifications HP webcam 1300 is available.

V. NEURAL NETWORK

Neural networks are defined as a way of doing machine learning, in which the machine learns to analyze data that is supplied for training and performs a task. Neural networks were developed based on how the human brain functions. In the human brain, there are thousands and millions of neurons that are interconnected in a densely manner.

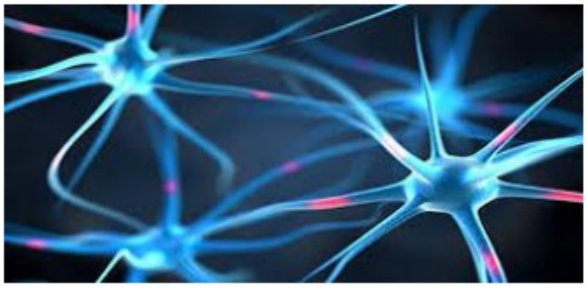


Figure 9. Neural connections

In the neural networks, there is an arrangement of layers of artificial neurons or nodes that allow data or information to travel in one direction known as feed-forward. Every individual node can be connected to the node in any layer after it and also to the node any layer before it from which it might receive data or transmit data to. All the links or the interconnections are assigned some values that are known as weights. Whenever the network is active, the data received by the node is always a different number as it multiplies the connections with the incoming data at the previous layer and adds up every connection data received at a particular node in the present layer. Every neuron or node has a threshold that determines whether the node has to be active or not. The summation of the inputs and the connections from the previous layer to the node in the present layer is passed through an activation function which gives out a value that is compared with the threshold and then decision is made to fire the neuron or not.

Convolutional Neural Networks

Convolutional Neural Networks are one of the most common networks that are used not only in vision applications but also in natural language processing. They are regularized versions of fully connected networks. Fully connected networks are networks in which every neuron in a layer is connected to every neuron in the next layer. This dense connectivity can lead to over fitting of the data. Convolutional neural networks were developed from the biological processes in which connectivity pattern between neuron resembles animal vision or visual cortex. The CNNs use very less pre-processing when compared to other image classification algorithms. For the normal algorithms, every filter has to be hand engineered but the CNNs reduce the complexity of designing each filter as it learns the filters.

VI. RESULTS

A Deep Learning model to detect the quality of the product has been built which detects whether the product is defective or not. The model has been built on two different stages that help to detect and classify the image. The comparison of the model has been made with the most commonly used pre trained model.

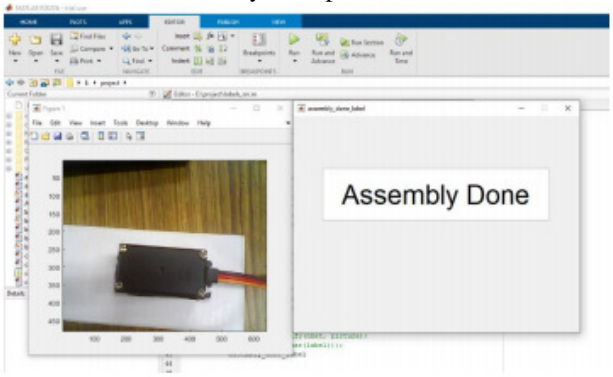


Figure 10. Label for assembly done part

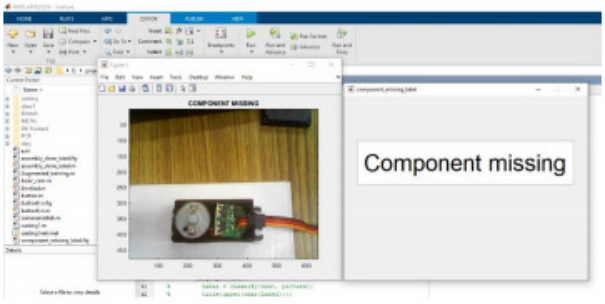


Figure 11. Label for Component missing

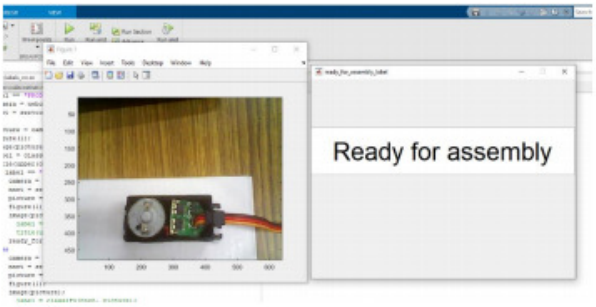


Figure 12. Label for Ready for assembly

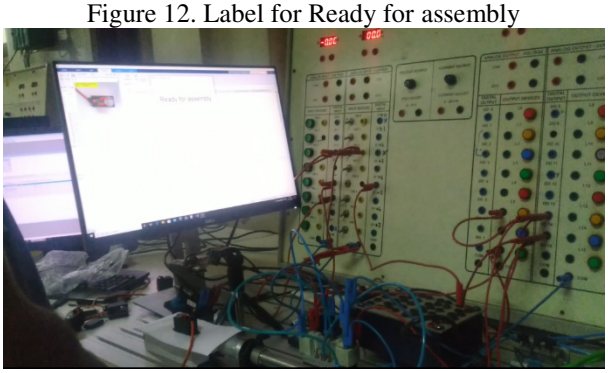


Figure 13. MG996r servo Inspection before assembly labeled " Ready for assembly"

The inspection system detects the product for different image labels assigned in deep learning models. The detection was based on different model architectures, best among them chosen for deploying in computer vision system.

Analysis of computer vision systems

The proposed Computer vision system is evaluated for accuracy.

Table 1. MG996r servo motor's test accuracies for different deep learning models.

MODEL NAME	TEST ACCURACY
GOOGLENET	87.89
RESNET50	81.82
RESNET101	81.82
RESNET18	83.33
ALEXNET	81.82
VGG16	72.73
4 CNN LAYERED NETWORK	83.33

Table 2. Gear rack's test accuracies for different deep learning models.

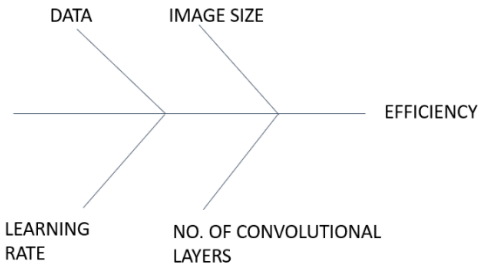
MODEL NAME	TEST ACCURACY
GOOGLENET	94.4
RESNET50	83.3
RESNET101	88.89
RESNET18	91.67
ALEXNET	88.89
VGG16	86.11
4 CNN LAYERED NETWORK	88.89

An inspection Rate is calculated for Gear Rack which is manufactured at Vasantha Tools crafts Pvt Ltd. And comparison is made with present system.

Table 3 . Inspection rate Manual vs automated.

	Manual Inspection	Automated Inspection
Inspection Rate	8 units/min	15 units/min

Factors effecting the accuracy of the model



VII. CONCLUSION

A Deep Learning Vision system is developed for inspection which achieved Inspection rate of 15 units/min that outperformed existing system with Inspection rate of 8 units/min. Accuracy analysis is performed for various pretrained models. 94.4% is achieved for Gear Rack with Google Net Architecture and 87.89 is achieved for MG996R servo motor with Google Net Architecture. Deep Learning Model with highest accuracy is deployed to Raspberry pi 4 Model B and Integrated with PLC using I/O modules. Online Inspection Analysis performed at Vasantha Tool Crafts Pvt Ltd for Gear Rack. As a future scope, a model can be developed for 3D inspection. It can be improved to show the missing component location in digital image. PCB circuit defects can be inspected using high quality Camera.

REFERENCES

[1].ToubaMalekzadeh:Aircraft Fuselage Defect Detection Using Deep Neural Networks

[2]. Yundong Li: Deformable Patterned Fabric Defect Detection with Fisher Criterion-Based Deep Learning.
[3]. Neha Sharma:An Analysis of Convolutional Neural Networks for Image Classification.
[4]. Prof. NilimaBargal:PLC Based Object Sorting Automation
[5]. Madhav Patil:Integration of Vision System, Intelligent ROBO Actuator, HMI and PLC to Design a Universal Quality Inspection or Control Machine.
[6]. Denni Kurniawan:Design and Implementation of Visual Inspection System in Automatic Bottling System Based on PLC.