Analysis of Machine Vision Using Deep Learning

Abigail Keziah J[1],Aksharasree K[2],Deepika K[3],Hemalatha B[4], Gokulavasan B[5]
[1, 2, 3, 4] Dept of Electronics and Communication Engineering
[5]Professor, Dept of Electronics and Communication Engineering
[1, 2, 3, 4, 5] Sri Eshwar College Of Engineering,Coimbatore,India.

Abstract- *One of the important characteristics of fruits is its appearance. Appearance not only influences their value, the preferences and also the choice of the patron, but also their internal quality to a particular extent. Colour, texture, size, shape, similarly the visual flaws are generally examined to assess the skin quality of object. Manually controlling external internal control of object is time consuming and labour intensive. Thus for automatic external internal control of objects, computer vision systems are widely utilized by intensive beat decades. The employment of machine and computer vision technology within the field of external quality inspection of object has been published supported studies carried on spatial image and / or spectral image processing and analysis. A close overview of the method of fruit classification and grading has been presented during this paper. Detail examination of every step is completed. Some extraction methods the common features of objects like colour, size, shape and texture. Machine learning algorithms like Convolutional Neural Networks (CNN) also is discussed. Process, advantages, disadvantages, challenges occurring in object-classification and grading is discussed during this paper, which might give direction to researchers.*

Keywords- Deep Learning, Machine Vision, Object classification, Convolutional Neural Network, Object detection.

I. INTRODUCTION

Deep learning could also be machine learning in computing (AI) that has networks capable of learning supervised from data that's structured... Deep learning techniques are as a strong strategy for learning feature directly from data which have led to remarkable breakthrough within the sphere of object detection. Deep learning algorithms, particularly convolutional networks, have rapidly become a way of choice for detecting and specifying objects. Object detection, one among the foremost challenging problems in computer vision which helps to locate object from an oversized number of stored categories in real images. We show that this accuracy is reached by increasing the depth of the projet.We use deep learning for image classification, object detection and segmentation related tasks the success of deep learning is based on new model regularization techniques, improved nonlinearities design, and current hardware capabilities, among others. In particular, for Machine Vision tasks, the success of deep learning is based on convolutional neural networks (CNN) which have become the standard neural network variant to process images. This concept lies at the thought of the various deep learning algorithms: models (networks) composed of the various layers that transform file (e.g. images) to outputs while learning greater features. because the cornerstone of image understanding and computer vision, object detection forms the thought for solving complex or high level vision tasks like segmentation, scene understanding, object tracking, image captioning, event detection, and activity recognition.

Object detection supports a decent range of applications; including robot vision, consumer electronics, and security, autonomous driving, and human computer interaction, content based image retrieval, intelligent video surveillance, and augmented reality. CNN contain many layers that transform their input with convolution filters of a tiny low extent. During this paper, object is being detected and specified by comparing it with several images using CNN. Variety of the machine learning libraries like Tensor Flow, PyTorch etc. is employed. With help of Alex Net-CNN- image classification is been done and thus the item is detected.

In this part we review several alternate attempts to use neural networks and deep learning for fruit recognition.

The way for identifying and counting fruits from images in cluttered greenhouses is presented. The targeted plants are peppers with fruits of complex shapes and ranging colors the identical because the plant canopy. The aim of the appliance is to locate and count green and red pepper fruits on large, dense pepper plants growing in a very greenhouse.

The process and validation data entered during this paper consists of 28000 images of fruits. The used method to locate and count the peppers is two-step: within the initiative, the fruits are located during one image and through a second step multiple views are merged to raise the detection rate of the fruits. The approach to look out the pepper fruits during one image is based on a mixture of (1) finding points of interest, (2) applying a fancy high-dimensional feature descriptor of a patch around the point of interest and (3) employing a so-called bag-of-words for classifying the patch. Paper presents a very unique approach for detecting fruits from images using deep neural networks. For this issue the authors produce a Faster Region-based convolutional network. The aim is to make a neural network which will be utilized by specialized robots which can produce fruits. The network is trained with RGB and NIR images. The mixture of the RGB and NIR models is finished in 2 separate methods: early and late fusion. Early fusion hinted the input layer has 4 channels: 3 for the RGB image and one for the NIR image. Late fusion uses 2 independently trained models that are combined by obtaining predictions from both models and averaging the results. The result is a multi-modal network which obtains much better performance than the current networks.

II. LITERATURE SURVEY

In this section we analyse several preceding attempts to use neural networks and deep learning for fruits recognition. A ability for recognizing and counting fruits from images in cluttered greenhouses is presented. The targeted plants are peppers with fruits of complex shapes and ranging colors the same as the plant canopy. The goal of the applying is to detect and count green and red pepper fruits on large, dense pepper plants growing in a very greenhouse.

2.1Convolutional Layers

A convolutional layer consists of groups of neurons that frame kernels. The kernels have little size but they always have the identical depth because the input. The neurons from a kernel are connected to little region of the input, called the receptive field, because it's highly inefficient to link all neurons to all or any previous outputs within the case of inputs of high dimensions like images. As an example, a 100 x 100 image has 10000 pixels and if the primary layer has 100 neurons, it'd lead to 1000000 parameters. Rather than each neuron having weights for the total dimension of the input, a neuron holds weights for the dimension of the kernel input.

2.2Deep Belief Network

Yet another model that's a part of the deep learning algorithms is that the deep be- life network. A deep belief network could be a probabilistic model composed by multiple layers of hidden units. The usages of a deep belief network are the identical because the other presented networks but may also be wont to pre-train a deep neural network so as to boost the initial values of the weights. This process is very important because it can improve the standard of the network and might reduce training times. Deep belief networks may be combined with convolutional ones so as to get convolutional deep belief networks which exploit the benefits offered by both styles of architectures.

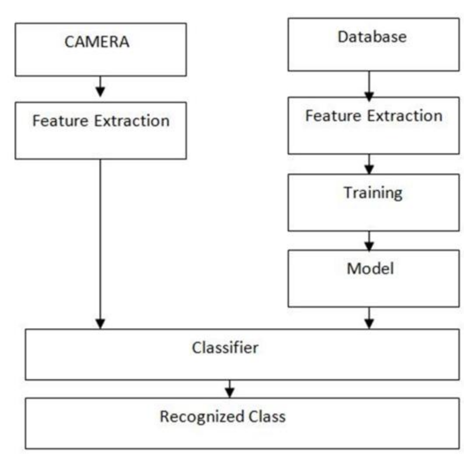
III. EXISTING METHOD

The training and recognition data employed in this paper consists of 28000 images of over 1000 plants and their fruits. The used method to locate and count the peppers is two-step: within the beginning, the fruits are located in an exceedingly single image and in an exceedingly second step multiple views are combined to extend the detection rate of the fruits.The approach to search out the pepper fruits during a single image relies on a mix of (1) finding points of interest, (2) applying a posh high-dimensional feature descriptor of a patch round the point of interest and (3) employing a so-called bag-of-words for classifying the patch.

IV. METHODOLOGY

The input sheet of the network contains 30,000 neurons as computer file, representing the quality RGB image of size 100×100 pixel. The primary hidden sheet is that the convolutional layer 1 which has 64 filters with a kernel of size 3×3 pixels and Rectified Linear Units (ReLU) as activation function. The second convolutional sheet is that the convolutional layer 2 where 64 filters with the kernel size of 3×3 pixels and ReLU were employed as on the primary convolutional layer. The convolutional sheet is employed for the feature extraction from input data.

BLOCK DIAGRAM



V. IMAGE SEGMENTATION

In image processing, image segmentation are often defined as a "process of partitioning a digital image into multiple segments" (sets of pixels, also observed as super pixels). The aim of image segmentation is to clearly change the representation of a image, which is more significant and easier to investigate.The aim of image segmentation is to classify or change the representation of an image, which is more similar and easy to analyse. Image segmentation methods are divided on the concept of two things of discontinuity and similarity. Methods based on discontinuities called boundary-based methods, and methods based on similarity called region-based methods. The output of the segmentation is either a restriction of the thing from the background or the area itself. In the tone image segmentation, distinct tone spaces such as RGB, HSI and Celia are used, with the image segmentation.

5.1 Feature Extraction

Feature extraction is a low-level image processing application. For an image, the attribute is the "interest" part. In the specimen identification literature, the name feature is mostly used to appoint a descriptor. Repeatability is the desirable property of a feature device after image segmentation, the next level is to take out art features useful in narrate fruits. Various attribute can be taken out from the image: hue, shape, size, texture. There are some community feature detector and visual descriptor, which are used for thing recognition and classification. Some of them are Speeded Up Robust Features (SURF), Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP). All these attributes, characteristic detectors and sight descriptors are explained in next section.

5.2 Fruits-360 dataset

In this section we explain how the info set was generated and what it holds. The images were acquired by filming the fruits while they're go round by motor and so extracting frames. Fruits were sowed within the shaft of a coffee speed motor (3 rpm) and a brief film of 20 seconds was noted. Behind the fruits we kept a white sheet of paper as background

To grasp the complication of background-removal steps we've got depicted in Figure1a fruit with its original background and after the background was removed and also the fruit was ascend right down to 100 x 100 pixels. The obtained dataset has 82110 images of fruits and vegetables spread across 120 tags. Each art contains one fruit or vegetable. Separately, the dataset has another 103 pictures of multiple fruits. The info set is obtainable on Github and Kaggle.

Figure 1. Left-side: original image. Notice the background and the motor shaft. Right-side: the fruit after the background removal and after it was scaled down to 100x100 pixels.

5.3 Tensorflow Library

For the motive of executing, instructing and testing the network explained during this paper we used the Tensor Flow library. This is often an open source structure for machine learning created by Google for numerical action using data flow graphs. Nodes within the graph represent mathematical functions, while the graph edges constitute the multidimensional data arrays called tensors.

Tensor Flow offers some efficient attributes such as: it allows computation mapping to multiple machines, unlike most other similar frameworks; it's in-built uphold for mechanized gradient computation; it can half results sub graphs of the full graph and it can add constraints to devices, like placing nodes on devices of a specific type, ensure that two or more objects are placed within the similar space etc.TensorFlow is utilized in several researches, just like the Inception Image Classification Model. This project introduced a state of the art network for classification and detection within the Image Net Large-Scale Visual Recognition Challenge 2014. During this research the usage of the computing assets is improved by modifying the network width and depth while keeping the computational budget constant.

A convolutional layer is defined like this:

```
conv2d (
        input,
        filter,
        strides,
        padding,
        use_cudnn_on_gpu=True,
        data_format='NHWC',
        dilations=[1 , 1, 1,1],
        name=None
```

10)

Performs the max pooling operation on the input. The k size and strides parameters are often tuples or lists of tuples of 4 elements. K size represents the scale of the window for every dimension of the input tensor and strides represent the stride of the window for every dimension of the input tensor. The padding parameter is often ‘’VALID’‘or ‘’SAME’‘

5.4The structure of the neural network employed in experiments

For this research we nearly new a convolutional neural network. As early explained the sort of network, it makes use of convolutional sheets, pooling layers, ReLU layers, fully connected layers and loss layers.

In a typical CNN design, each convolutional sheet is followed by a Rectified Linear Unit (ReLU) layer, then a Pooling layer then one or more convolutional sheet and at last one or more completely connected layer.

Note again that an attribute that sets apart the CNN from a regular neural network is taking into report the shape of the images while processing them. A regular neural network changes the input in a one dimensional array which makes the trained classifier less careful to positional changes.

Table1. Structure of neural network in this paper

Layer type	Dimensions	Output
Convolutional	5 x 5 x 4	16
Max pooling	2 x 2 — Stride: 2	-
Convolutional	5 x 5 x 16	32
Max pooling	2 x 2 — Stride: 2	-
Convolutional	5 x 5 x 32	64
Max pooling	2 x 2 — Stride: 2	-
Convolutional	5 x 5 x 64	128
Max pooling	2 x 2 — Stride: 2	-
Fully connected	5 x 5 x 128	1024
Fully connected	1024	256
Softmax	256	60

A visual depiction of the neural network used is given

The first layer (Convolution #1) is a convolutional sheet which bid 16 5 x 5 filters. On this sheet we apply max pooling with a filter of structure 2 x 2 with stride 2 which makes the pooled regions do not overlap (Max-Pool #1). This also decreases the width and height to 50 pixels each.

The second convolutional (Convolution #2) sheet applies 32 5 x 5 filters which results 32 activation maps. We apply on this sheet the same kind of max pooling (Max-Pool #2) as on the first sheet, structure 2 x 2 and stride 2.

The third convolutional (Convolution #3) sheet applies 64 5 x 5 filters. Following is another max pool sheet (Max-Pool #3) of shape 2 x 2 and stride 2.

The fourth convolutional (Convolution #4) sheet applies 128 5 x 4 filters. Following a final max pool sheet (Max-Pool #4).

5.6 Numerical Experiments

For the experiments we used the 82110 images divide in 2 parts: training set which consists of 61488 images of fruits and testing set which is created of 20622 images. The opposite 103 images with multiple fruits weren't employed in the training and testing of the network. The data was tied up into a record file (specific to Tensor Flow). This can be a computer file that contains protocol buffers with a feature map. During this map it's possible to store information like the image height, width, depth and even the raw image.

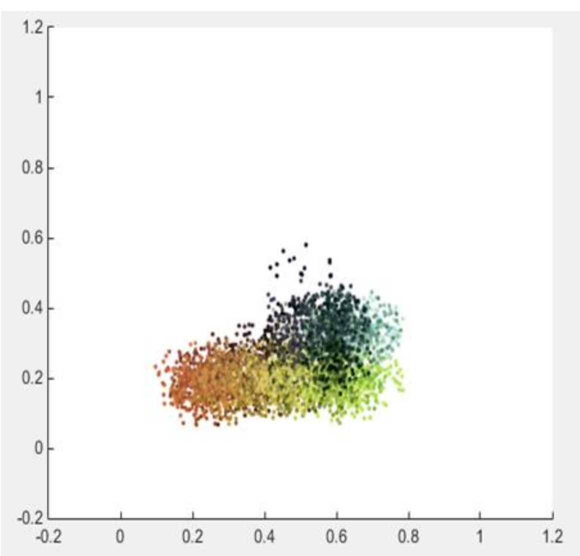
Using these files we are able to create queues so as to feed the information to the neural network. By calling the strategy shuffle batch we offer randomized input to the network. The way we used this technique was on a condition that it example tensors for images and labels and it returned tensors of shape batch size x image dimensions and batch size x labels. This helps greatly lower the prospect of using the identical batch multiple times for training, which successively improves the standard of the network. We ran multiple scenarios during which the neural network was trained using different levels of knowledge augmentation and preprocessing.

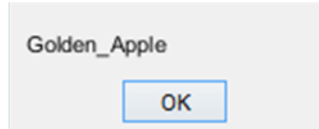
Table 2. Results of training neural network on 360 datasheet

Scenario	Accuracy on training set	Accuracy on test set
Gray scale	99.82%	92.65 %
RGB	99.82%	94.43 %
HSV	99.80%	94.40 %
HSV + Gray scale	99.78%	94.74 %
HSV + Gray scale + hue/saturation change + flips	99.58%	95.23 %

It is important to note that training the gray scale image only yielded the simplest results on the toy but very weak results on the test set. We explored this cause and that we have discovered that lots of images containing apples are wrongly divided on the test set. So as to further investigate the problem we ran a round of coaching and testing on just the apple classes of images. The outputs were same, with high correctness on the train data, but low correctness on the test data. We attribute this to over fitting, because the gray scale images lose too many features, the network doesn't learn properly the way to classify the photographs. In order to see the simplest network configuration for classifying the photographs in our dataset, we took multiple configurations, used the toy to coach them then calculated their accuracy on the test and training set.

VI. PROPOSED OUTPUT





VII. CONCLUSION AND FUTURE WORK

We described a replacement and sophisticated database of images with fruits. Also we made some numerical experiments by using Tensor Flow library so as to classify the pictures per their content. From our point of view one amongst the most objectives for the longer term is to enhance the accuracy of the neural network. This requires further testing with the formation of the network. Various tweaks and changes to any layers similarly because the introduction of latest layers can provide completely different results.

Another choice is to replace all sheets with convolutional layers. This has been shown to provide some upgrade over the networks that have fully connected sheets in their form. A result of replacing all sheets with convolutional ones is that there will be a rise in the number of framework for the network. Another chance is to replace the rectified linear parts with exponential linear parts. According to paper, this decreases computational risk and add notably better generalization performance than rectified linear parts on networks with more than 5 sheets. We would like to figure out these practices and also to try to find new layout that provide interesting solution. In the near future we plan to create a mobile application which takes images of fruits and tag them accordingly. Another unbiased is to expand the dataset to include more fruits. This is a time consuming process since we want to include object that were not used in most others related papers

REFERENCES

[1] Bargoti,S.,andUnderwood,J.Deepfruitdetectioninorchards. In*2017IEEEInternationalConferenceonRoboticsandAutomation(ICRA)*(May2017),pp.3626–3633.

[2] BAutomatic fruit classification using deep learning for industrial applicationsMS Hossain, M Al-Hammadi… - IEEE Transactions on …, 2018 - ieeexplore.ieee.org

[3] Chan,T.F.,andVese,L.A.Activecontourswithoutedges.*IEEE TransactionsonImageProcessing10*,2(Feb2001),266–277.

[4] Counting apples and oranges with deep learning: A data-driven approachSW Chen, SS Shivakumar, S Dcunha… - IEEE Robotics and …, 2017 - ieeexplore.ieee.org

[5] Performance Analysis of Gradient Descent Methods for Classification of Oranges using Deep Neural Network

P Pathak, H Gangwar, AS Jalal – 2020 7th International …, 2020 –ieeexplore.ieee.org