Cantaloupe Classification Using Deep Learning

Basel El-Habil, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering & Information Technology,
Al-Azhar University - Gaza, Palestine

Abstract cantaloupe and honeydew melons are part of the muskmelon family, which originated in the Middle East. When picking either cantaloupe or honeydew melons to eat, you should choose a firm fruit that is heavy for its size, with no obvious signs of bruising. They can be stored at room temperature until you cut them, after which they should be kept in the refrigerator in an airtight container for up to five days. You should always wash and scrub the rind of your melon before you cut it to remove any dirt or bacteria on the outside. In this paper, cantaloupe classification approach is presented with a dataset that contains approximately 1,312 of Cantaloupes and honeydews. Convolutional Neural Network (CNN) algorithms, a deep learning technique extensively applied to image recognition was used, for this task. The results found that CNN-driven classification applications when used in farming automation have the latent to enhance crop harvest and improve output and productivity when designed properly. The trained model achieved an accuracy of 99.74% on a held-out test set, demonstrating the feasibility of this approach.

Keywords: Cantaloupe, Classification, Deep Learning

1. INTRODUCTION

We compared the nutritional contents of cantaloupe versus honeydew below using 2020 USDA and NIH data[2] or a quick recap of significant nutrients and differences in cantaloupe and honeydew: Both cantaloupe and honeydew are high in Vitamin C and potassium. Cantaloupe is an excellent source of Vitamin A.

Detailed nutritional comparison of cantaloupe and honeydew is analyzed Fig.3. You can also visualize the nutritional comparison for a custom portion or serving size and see how the nutrition compares.

Fig.1 Cantaloupe
In this paper, we present that a Deep Convolutional Neural Network (CNN) classified cantaloupe two types. In computer vision, CNNs have been known to be powerful visual models that yield hierarchies of features enabling accurate segmentation. This was considered a task only we humans could do. A researcher introduced a model capable of classifying characters with a but CNN learned features by itself [6].

Deep learning is a subset of machine learning which in turn a subset of artificial intelligence. Artificial intelligence is a technique that enables a machine to mimic human behavior. Machine learning is a technique to achieve AI through algorithms trained with data and finally deep learning is a type of machine learning inspired by the structure of the human brain in terms of deep learning. This structure is called an artificial neural network let's understand deep learning better and how it's different from machine learning, we create a machine that could differentiate between cantaloupe and Honeydew if done using machine learning we'd have to tell the Machine the features based on which the two can be differentiated these features could be the size and the type of stem on them, with deep learning on the other hand the features are picked out by the neural network without human intervention of course that kind of independence comes at the cost of having a much higher volume of data to train our machine [15-25].

One of the most common AI techniques used for processing Big Data is Machine Learning, a self-adaptive algorithm that gets gradually better analysis and patterns with experience or with new added data. If a digital payments company wanted to detect the occurrence of or potential for fraud in its system, it could use machine learning tools for this purpose. The computational algorithm built into a computer model will process all transactions happening on the digital platform, find patterns in the data set and point out any anomaly detected by the pattern [26-40].
Let's go into the working of neural networks here we have three students each of them write down the digit nine on a piece of paper notably they don't all write it identically the human brain can easily recognize the digits but what if a computer had to recognize them that's where deep learning comes in here's a neural network trained to identify each image of 128 times 128 pixels now that amounts to a total of 16,384 pixels. Neurons the core entity of a neural network is where the information processing takes place each of the 16,384 pixels is fed to a neuron in the first layer of our neural network this forms the input layer on the other end we have the output layer with each neuron representing a digit with the hidden layers. Existing between them the information is trans from one layer to another over connecting channels each of these has a value attached to it and hence is called a weighted Channel. All neurons have a unique number associated with it called bias. This bias is added to the weighted sum of inputs reaching the neuron which is then applied to a function known as the activation function. The result of the activation function determines if the neuron gets activated every activated neuron passes on information to the following layers. Continues up till the second last layer the one neuron activated in the output layer, corresponds to the input, the weights and bias are continuously adjusted to produce a well-trained network. Where is deep learning applied in customer support when most people converse with customer support agents the conversation seems so real they don't even realize that it's actually a bot on the other side in medical care neural networks clustering cantaloupe and analyze. It too faces some limitations the first as we discussed earlier is data while deep learning is the most efficient way to deal with unstructured data a neural network requires a massive volume of data to Train. Assume we always have access to the necessary amount of data processing, while this is within the capability of every Colaboratory - Google Research and that brings us to our second a computational power training and neural network requires graphical processing units which have thousands of course as compared to CPUs and GPUs and finally we come to training time deep neural networks take littl time to train. The time increases with the amount of data and number of layers in the network so here a short quiz for you arrange the following statements in order to describe the workings of a neural network a the bias is added be the weighted sum of the inputs is calculated see specific neuron once it is activated the activation function leave your, and to hurry some of the popular deep learning frameworks include tensorflow high torch Keras deep learning cognitive toolkit considering the future. Predictions for deep learning and AI we seem to have only scratched the surface in fact technology is working on a device for the blind that uses deep learning with computer vision to describe the world to the users replicating the human mind at the entirety may be not just an episode of science fiction or too long the future is indeed full of surprises [41-60].

The next layer takes the second layer’s information and includes raw data like geographic location and makes the machine’s pattern even improved. This goes on across all levels of the neuron network. Practical application of Deep Learning is fraud detection system. Using the fraud detection system mentioned above with machine learning, we can create a deep learning example. If the machine learning system created a model with parameters built around the amount of dollars a user sends or receives, the deep learning method can start building on the results offered by machine learning. Each layer of its neural network builds on its previous layer with added data like retailer, sender, user, social media event, credit score, IP address and a host of other features that may take years to connect together if processed by a human being. Deep learning algorithms are trained to not just create patterns from all transactions, but to also know when a pattern is signaling the need for a fraudulent investigation. The final layer relays a signal to an analyst who may freeze the user’s account until all pending investigations are finalized [8]. Deep learning is used across all industries for a number of different tasks. Commercial apps that use image recognition, open source platforms with consumer recommendation apps and medical research tools that explore the possibility of reusing drugs for new ailments are a few of the examples of deep learning incorporation.

2. RELATED WORK

The Authors in [9] used deep learning to detect five tomato leaves diseases. They achieved a high accuracy in detecting the tomato disease. The authors in [10] provided a dataset that is aimed at ground-based weed or specie spotting and also suggested a benchmark measure to researchers to enable easy comparison of classification results. The authors in [11] demonstrated the effectiveness of a convolutional neural network to learn unsupervised feature representations for 44 different plant species with high accuracy.

In the course of exploring the right architecture for our model, we consider the work of [11] in classifying leaves using the VGGNet16 architectures. The authors in [12] implemented a 26-layer deep learning model consisting of 8 residual blocks in their classification of 10,000 images of 100 ornamental plant species achieving classification rates of up to 91.78%. The authors in [13] addressed the problem of CNN-based semantic segmentation of crop fields separating sugar beet plants, weeds, and background solely based on RGB data by proposing a deep encoder-decoder CNN for semantic segmentation that is fed with a 14-channel image storing vegetation indexes and other information that in the past has been used to solve crop-weed classification.

3. METHODOLOGY

In this section we describe the proposed solution as selected pre-trained VGGNet16 architectures convolutional network (ConvNet) architecture with some tuning and discuss associated design choices and implementation aspects.

3.1 Dataset
The dataset used, was provided by the 1,312 images, balanced 656 for each class (Cantaloupe type) See Fig. 4 for the samples.

![Fig.4: Images of the samples](image)

3.2 Architecture

VGGNet is a well-documented and globally used architecture for convolutional neural networks [11]. This ConvNet became very widespread by accomplishing excellent performance on the ImageNet [12] dataset. It comes in several variations of which the two best-performing (with 16 and 20 weight layers) have been made publicly available. In this work, the VGG16 architecture was selected, since it has been shown to generalize well to other datasets. The input layer of the network expects a 128x128 pixel RGB image. The input image is passed through five convolutional blocks. Small convolutional filters with a receptive field of 3_3 are used. Each convolutional block includes a 2D convolution layer operation (the number of filters changes between blocks). All hidden layers are equipped with a ReLU (Rectified Linear Unit) as the activation function layer (nonlinearity operation) and include spatial pooling through use of a max-pooling layer. The network is concluded with a classifier block consisting of three Fully Connected (FC) layers.

<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output Shape</th>
<th>Param #</th>
</tr>
</thead>
<tbody>
<tr>
<td>input_1 (InputLayer)</td>
<td>(None, 128, 128, 3)</td>
<td>0</td>
</tr>
<tr>
<td>block1_conv1 (Conv2D)</td>
<td>(None, 128, 128, 64)</td>
<td>1792</td>
</tr>
<tr>
<td>block1_conv2 (Conv2D)</td>
<td>(None, 128, 128, 64)</td>
<td>36928</td>
</tr>
<tr>
<td>block1_pool (MaxPooling2D)</td>
<td>(None, 64, 64, 64)</td>
<td>0</td>
</tr>
<tr>
<td>block2_conv1 (Conv2D)</td>
<td>(None, 64, 64, 128)</td>
<td>73856</td>
</tr>
<tr>
<td>block2_conv2 (Conv2D)</td>
<td>(None, 64, 64, 128)</td>
<td>147584</td>
</tr>
<tr>
<td>block2_pool (MaxPooling2D)</td>
<td>(None, 32, 32, 128)</td>
<td>0</td>
</tr>
<tr>
<td>block3_conv1 (Conv2D)</td>
<td>(None, 32, 32, 256)</td>
<td>295168</td>
</tr>
<tr>
<td>block3_conv2 (Conv2D)</td>
<td>(None, 32, 32, 256)</td>
<td>590080</td>
</tr>
<tr>
<td>block3_conv3 (Conv2D)</td>
<td>(None, 32, 32, 256)</td>
<td>590080</td>
</tr>
</tbody>
</table>
### 3.3 Design considerations

The original VGGNet16 must be modified to suit the current solution: the final fully-connected output layer must perform 12 classes only. See Fig 5
Fig. 5: CNN Architecture
3.3.1 Preprocessing

Input images must be preprocessed by:

- Read the picture files.
- Decode the JPEG content to RGB grids of pixels.
- Convert these into floating-point tensors.
- Rescale the pixel values (between 0 and 255) to the [0, 1] interval (as you know, neural networks prefer to deal with small input values).

3.3.2 Data augmentation

Data augmentation takes the approach of generating more training data from existing training samples, by augmenting the samples via a number of random transformations that yield believable-looking images. The goal is that at training time, your model will never see the exact same picture twice. This helps expose the model to more aspects of the data and generalize better. In Keras, this can be done by configuring a number of random transformations to be performed on the images read by the ImageDataGenerator instance.

3.4 Possible solutions

The modified VGG16 ConvNet can be used in three different ways:

- Training the ConvNet from scratch;
- Using the transfer learning paradigm to leverage features from a pre-trained VGG16 on a larger dataset; and
- Keeping the transfer learning paradigm and fine-tuning the ConvNets architecture. These variants (named Method 1, Method 2, and Method 3, respectively) are described next.

3.4.1 Training from scratch

The architecture is initialized with random weights and trained for a number of epochs. After each epoch, the model learns features from data and computes weights through backpropagation. This method is unlikely to produce the most accurate results if the dataset is not significantly large. However, it still can serve as a baseline for comparison against the two other methods.

3.4.2 ConvNet as feature extractor

Due to the relatively small number of images of Cantaloupe datasets, this method initializes the model with weights from the VGG16 trained on a larger dataset (such as ImageNet [11]), a process known as transfer learning. The underlying assumption behind transfer learning is that the pre-trained model has already learned features that might be useful for the classification task at hand.

This corresponds, in practice, to using selected layer(s) of the pre-trained ConvNet as a fixed feature extractor, which can be achieved by freezing all the convolutional blocks and only training the fully connected layers with the new dataset.

3.4.3 Fine-tuning the ConvNet

Another common transfer learning technique consists of not only retraining the classifier on the top of the network with the new dataset, but also applying a fine-tuning of the network by training only the higher-level portion of the convolutional layers and continuing the backpropagation.

In this work, we propose to freeze the lower level layers of the network because they contain more generic features of the dataset. We are interested in training only the top layers of the network due to their ability to perform extraction of more specific features. In this method, the first four convolutional layers in the final architecture are initialized with weights from the ImageNet dataset. The fifth, and final, convolutional block is initialized with weights saved and loaded from the corresponding convolutional layer in Method 1. This method was adapted in our current research.
4. EXPERIMENTS AND DISCUSSIONS
We have done the experiment with Fine-tuning the ConvNet as described above. We used the original Cantaloupe dataset that consists of 1,312 images after resizing the images to 128x128 pixels. We divided the data into training (70%), validation (15%) and testing (15%). The training accuracy was 99.74% and the validation accuracy was 100%, with CPU times: user 1.07 s, sys: 15.3 ms, total: 1.09s.

5. CONCLUSION
We proposed a solution for assisting farmers to optimize cantaloupe. More specifically, we have designed and implemented a two-class classifier that takes cantaloupe images with 2 different species as input, builds a model using deep learning convolutional neural networks, and uses this model to predict the type of (previously unseen) images of cantaloupe. The proposed approach achieves promising results – most notably, validation accuracy of 99.74%.
64. Ng, Andrew; Dean, Jeff (2012). 'Building High-level Features Using Large Scale Unsupervised Learning', arXiv:1112.6209[cs.LG].