COMP309-CNN image Classification Project

Jiawei Luo, *Victoria University of Wellington*School of Engineering and Computer Science

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—Deep Learning, Machine Learning, CNN

# Introduction

In the field of modern artificial intelligence, convolutional neural networks(CNN) have been widely studied and used in various industries. Among them, image recognition is the most important application of CNN. In this paper, The objective of this paper is to solve an Image Recognition Problem for cherries, strawberries and tomatoes. A baseline model with one convolutional layer was constructed initially to basically understand the process of neural networks and obtain the basic performance matric of the model. After that, a CNN with multiple layers of convolutional layers and pooling layers is developed based on the baseline. The model that be constructed by CNN should correctly classify the label of a given image between cherry, strawberry and tomato.

# EDA Discussion and Problem investigation

## Initial data analysis

The initial dataset contains 3 Classes, which is three kinds of Fruit: Cherry, Strawberry and Tomato. The original dataset was split to training set and test set, 4500 instances and 1500 instances are in the training set and test set respectively. For training set that is given, Each folder of 3 classes includes 1500 .JPG images, and the Size of each image is 300\*300 which is preprocessed in terms of size of image. Therefore, the training set is well-balanced as the number of each class is the same.

The images are represented at a RGB model. Generally, RGB has 256 levels of Luminance, represented by numbers from 0,1,2.. Until 255. Although the highest number is 255,0 is also one of the values, so there are 256 levels. All colors can be created by mixing the red, green, and blue lights in different proportions. A set of red, green, and blue is the smallest display unit. Any color on the screen can be recorded and represented by a set of RGB values. So an image in training set is basically made up of 300 x 300 RGB units, each pixel having a set of values representing the corresponding color Luminance.

In training data, each folder contains some noise pictures. There are six kinds of noise images in figure 1. In the first cherry image, compare to a normal cherry, the small bubbles around the cherry could be seen as noise when training the CNN model. In RGB color image, this kind of noise can be seen as Salt and Pepper noise which an change the number of color combination based on the noise radio. Similarly, the cherry in the fifth image has different color as a general one, the number of RGB color would be significantly degrade the accuracy when classifying this kind of images.

The second image only contains a part of a strawberry, which may not be fully classified as the shape of the strawberry is hard to catch by the trained model. Also, the third image does not include a cherry at all, the word ‘CHERRY’ is not be trained as a cherry classification. The fourth and sixth image shows the occlusion of an image, which only a small portion of the target object could be visible and the target object is very small and invisible. In reality, the cherry and strawberry could be smaller than a tomato, which makes the process of classification harder.



On the other hand, some images have background with similar RGB color or sharp should be considered as noise images as they provide many extra information in the image. When zoomed to 300x300, some of the images became distinctly slanted or faded. However, in real life it would be unrealistic to expect pictures of these three fruits to contain no noise. The quality of image dataset is quite random, and it provides significant variation and randomness to the dataset, preprocessing and model-training process. However, reasonable noise should be include in the training dataset as it can introduce the capability of recognize noisy images to the test set. But the training performance would be degraded if many noise images are involved.

It is noticeable that, in training set, a few image is grayscale. The gray scaled image only have one dimension when the color model is ‘grayscale’. It can also be used as RGB model, but the ‘color’ feature will not be the same as other RGB images.

## Data Preprocessing

The all images in the dataset were previously preprocessed to ‘JPG’ images with 300\*300 image width. To further process the image dataset, some of the method in the keras ImageDataGenerator could be useful.

Some methods of Data Augmentation is the generation of training data from existing data sets. it is the processing of an image that already exists in the training Dataset and the processing of that image to create a version of many changes to the same image. This not only provides more image training, but also helps expose our classifier to a wider range of both and colors, making our classifier more robust.

**featurewise\_center/std\_normalization:** it is used to make the image have a mean pixel value of 0 and a standard deviation of 1, but I've tried it several times and it's not very intuitive.

**zca\_whitening:** The function of Zca Whitening is to reduce the dimension of Images by PCA, to reduce the redundant information and to preserve the most important features. When my image resize to 256 \* 256, the code reported a memory error that was too large during the calculation.

**Rescale:** The function of a resale is to multiply each pixel of the image by a scaling factor, which is performed before all other transformation operations. In some models, directly entering the pixel value of the original image may cause the activation function to fail. Therefore, setting the scaling factor to 1 / 255 and scaling the Pixel Value Between 0 and 1 is beneficial to the convergence of the model and ensures the normal operation of the neural network. After an image has been rescaled, the images stored locally are indistinguishable to the eye. If we print the image values directly in memory, they will all be floating point values of 0 to 1.

On the other hand, some images have a high pixel range and some images have a low pixel range. Because these images share the same model, they have the same weight and learning speed. Without scaling, images with a High Pixel range have a greater impact on the weight update and cause more loss, so these images will have a greater weight in the training, but this is not the case. However, the loss of the low-pixel range image is small, and the influence of the weight update is small. The total loss is reduced by scaling the range of the image, regardless of the range of the original image.

**width\_shift\_range/height\_shift\_range:** These two parameters are the horizontal position shift and the Upper and lower position shift of the picture respectively. The parameters can be floating-point numbers between [0,1] or greater than 1. It is set to 0.2 in my program. When images are moved, areas beyond the original image range are typically filled in according to the fill parameters. This is significant for the training data which has target objects in the corner or the edge of the image.

**rotation\_range:** it is used to specify the image rotation angle range, its parameters can specify an integer, but is not fixed at this angle for rotation, but in the [0, designated angle] range for random angle rotation. In my program, i set it to 20 as the training data have already provided many different angle of target objects.

**shear\_range:** The effect of Shear is to leave the x (or y) coordinates of all points unchanged, while the corresponding y (or x) coordinates are shifted proportionally, and the magnitude of the shift is proportional to the vertical distance from the point to the x (or Y) axis. After this operation, the shape of the image is usually changed.

**zoom\_range:** This allows the image to be zoomed randomly within the range of a set number and improve the performance on the scaled objects in the images.

Considering the color of three target objects: cherry, tomato and strawberry, They are similar in terms of color when standard images are provided. To reduce the dimensionality of trained model, Grayscale in the function ‘flow\_from\_directory’ were attempted. The training speed of gray-scaled images was significantly improved as the dimension of color model was decreased from 3(RGB) to 1(Grayscale). However, it did not work properly as the shapes of three objects are similar too, The accuracy was degraded and the loss was increased. So i decided to use RGB model to introduce more features and parameters to the model and improve the performance.

## Test/Validation set

The original training data has 4500 instances, I add other 1368 images from google images with same preprocessing(modify size to 300\*300, change to ‘JPG’ images and use the same format on image names). There are 2 ways for using the dataset. Firstly, i only using the provided 4500 images as training data, and using the other 1368 images as test and validation dataset only, treat them as unseen instances and not used for training. Then, i add the two dataset together as a entire dataset, than split 20% (a validation factor of 0.2) as validation set.

The final step of preprocessing is that i deleted the unreasonable images in dataset(both given training set and added dataset). Those data only contain incorrect information such as the third image in figure 1, it only have a ‘CHERRY’ in English, but not a real cherry.

# Methodology

To properly train and construct the model to do the classification of three-label dataset, A baseline network, a CNN and a transfer learning method ‘VGG16’ is used.

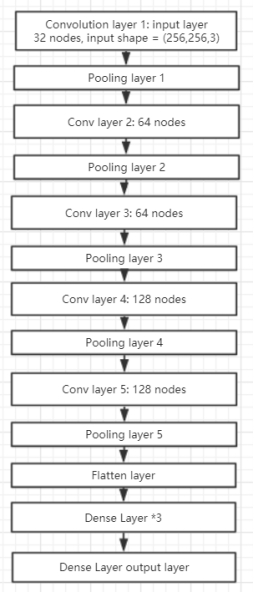
Machine image recognition is not to recognize a complex image completely at once, but to segment a complete image into many small parts and extract the features of each part, Then combine these small parts of the features together.

## 3.1.Baseline network

A one-layer CNN with a Convolution layer and 3 Dense layer is used as my baseline network. I tried to use Multi-layer perceptron, but the MLP did not work very well in my program, the accuracy can only converge about 33% with huge loss.

## CNN

The CNN that i constructed include 5 convolution layers, 5 pooling layers and Dense layers. The BatchNormalization was used in some layers. Also, dropout was used to prevent the overfitting in the CNN.



## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**