A Review on Color Recognition using Deep Learning and Different Image Segmentation Methods

Chirag Mahesh Sahasrabudhe

B. Tech Student, Dept. of Computer Science and Technology, MIT World Peace University, Pune, India.

Abstract – *With the publication of backpropagation algorithm paper by Geoffrey Hinton, deep learning has got the boost. In this paper, we talk about a deep learning model can be used to recognize various colors and impact of different segmentation methods on the color recognition.*

Key Words: Color recognition, CNN, Deep learning, Otsu's method, ReLU activation function, Watershed segmentation transfer learning.

2. TERMINOLOGIES USED.

2.1 Deep Learning.

Deep learning is a subset of machine learning which is further a subset of artificial intelligence. When fed with huge amount of raw data, deep learning can discover patterns in the given data. Further, the multi-layers of deep learning also known as neural networks, can recognize similar patterns and hence, segregate them into different classes. One advantage of deep learning over traditional machine learning algorithms is that, we can't give raw data (such as .csv) directly to the machine learning algorithm. Before giving input, we have to do pre-processing of the data. But we can give raw input to deep learning directly. Some of the examples of deep learning algorithms are convolutional neural network, recurrent neural network, generative adversarial networks etc. [1][2].

2.2 Convolutional neural networks.

Similar to traditional neural networks, [3] CNN has three different types of layers namely, input, hidden and output layer. The difference here is that, the input given to CNN is an image or pixel matrix.

The most important part of CNN is the kernel which is a 2D matrix of $N \times N$ size. In this matrix, each point has its own weight. The kernel size generally taken is of 2 X 2 matrix size.

Another characteristic which makes CNN technique to achieve higher accuracy results is large local receptive fields. [4] The receptive field size increases as the network becomes deeper and complex or a pooling layer is added to the network. A CNN works on a large receptive field (for example *48 X 48*) as on other hand traditional one's work on small receptive field such as *16 X 16*.

2.3 Watershed Image segmentation method.

As the name suggests, watershed segmentation is somewhat similar to geographical water shedding [5]. In this technique, the image is seen as a topographic landscape with ridges and valleys. The elevation points or values are the brightness of each pixel.

2.4 Otsu's method for Image segmentation.

In Otsu's threshold method [6], we iterate through all the possible values of the pixels and calculate a measure of spread for every pixel. The pixel which is in the foreground can be distinguished from the pixel in background by assigning a class level. Black label can be used for background pixels and whereas white for foreground features. Generally, grayscale histogram is passed to algorithms.

2.5 Adaptive Boosting.

Adaboost (short for Adaptive boosting) algorithm was first discussed by Schapire and Freund, in 1997[7]. It works on the concept of Majority voting. It is an ensemble type of learning.

The [8] common way to use adaptive boosting technique is with a decision tree. An adaboost with a decision tree is also known as a conventional adaboost. A tree with just one node and two leaves is known as a stump. Stump only works on one variable hence; they are also known as weak learners. The errors made by the first stump influence the output of the second stump. Stumps vary in their sizes. So, in final classification voting, some stumps get more influence (say) than the others. In adaboost every sample of the dataset is assigned with a sample weight. This sample weight indicates the importance of that sample to be correctly classified.

Sample weight is calculated as follows:

$$Weight = \frac{1}{Total\ number\ of\ samples\ present\ in\ the\ dataset.}$$

The stump having lowest Gini index or Gini impurity is taken as the first stump for classification purposes.

2.6 Rectified Linear Unit (ReLU) Activation function.

This activation function preserves the properties of linear models because it is a linear function thus, making it easy to optimize [9]. ReLU function, performs on the threshold value. If the input element is less than the threshold value, it is categorized to 0 (zero), and if greater than threshold value, it is categorized to threshold value. The function also eliminates the vanishing gradient problem as every input element less than threshold value is forced to become zero. But here overfitting of data is greater as compared to SoftMax function.

2.7 SoftMax Activation function.

SoftMax function [9] is used to compute the probability distribution. The output of the function is in the range of [0,1] where all the probabilities sum up to unity.

The SoftMax function is used for multivariate classification purposes.

2.8 Transfer learning property of deep learning.

Transfer learning is the method used to transfer the knowledge of the source domain (a trained CNN) to the target CNN (next CNN) efficiently. In transfer learning, the training and test datasets need not to be identically distributed. Datasets can be imbalanced.[10]

2.9 Residual Network Architecture.

Error rate of training and testing dataset increases as number of layers are increased. This can be verified from Fig. 1 of [11] where 56 layers were added to the network as compared to 20 layers. This analysis was performed on the infamous CIFAR-10 dataset. The solution to the problem is Residual Blocks or Residual Network architecture (ResNet). A slightly different version of ResNet is known as Skip connections where the network skips training from a few layers and directly connects to the output. This architecture also uses parametric gates which decides how much information should be passed through a connection.

3. LITERATURE REVIEW

In 2020, Feng Jiang *et al* proposed a new method for detection of rice diseases by combining deep learning and SVM model. [12]. Mean shift method was employed as image segmentation method, because the area extracted by this technique is large. Choosing the appropriate segmentation method can reduce the time complexity of the model. A total of 8,900 images were taken as the dataset. Three new features of the crop leaf other than traditional methods such as shape features, area, roundness etc. were considered. Eigenvalues from these shape features were calculated and further passed to the SVM model. Three color spaces were taken such as RGB, HSI and YCbCr color space. In CNN, six different layers were modelled. Out of which three were convolutional, two subsampling layers and the remaining one as feature layer which gives sigmoid activation function as the output. At every layer, feature maps were generated of certain pixel resolution which were then passed onto the next layer. The sigmoid activation function along with three shape features and total of nine color spaces were given as input to the SVM model. The initial weights between layers were taken randomly at first and then adjusted accordingly using a backpropagation algorithm. SVM was used because penalty parameter (C) in SVM can remove some redundant features. Grid search algorithm was used to compare the accuracy results with different combinations of penalty parameter C and kernel function g.

After using 10-cross fold validation, with C = 1 and g = 50 has highest accuracy compared to others. To fairly evaluate the model, ROC curve was chosen. Accuracy of 96.8% was seen with deep learning and SVM.

The authors [13] have proposed a deep neural network (DNN) model with ResNet to categorize the radioactive wastes properly. The model works on six types of labelled data, four of which are radioactive wastes such as vinyl, rubber, cotton and paper, and other two such as no objects in the image and no objects and no work tables. As we increase the number of layers, in all deep learning networks, training and testing error rate also increases with it. This phenomenon is known as Vanishing/ Exploding gradient [14]. This happens because the gradient related to deep learning network suddenly becomes zero or too large. A dataset was created through a video, captured at a sorting worktable. Images were further refined to a certain resolution which was suitable for training. For proper extraction, the camera was placed latterly over the worktable.

Phases of DNN:

a. In the first phase of DNN training, pre-processing of images is done. For example, resizing the image, cropping, color jittering and reducing the resolution of all images to 512 x 512 resolution.

b. In the second phase, to avoid the creation of an imbalanced dataset equal number of samples were extracted.

c. They have used mini-batch training to solve the over-fitting problem which is caused when the algorithm captures noise present in the dataset. The batch size was 128 each.

d. In fourth phase, they have further divided the batches into 8 subdivisions to maximize the benefits. Sending a fixed size of batches to the GPU reduces the hardware load.

e. After passing the data through a dropout layer, they have passed it to the ResNet50.

The processing time of the overall model was 0.268 seconds. The accuracy of the ResNet model increases as the number of layers increases and this was verified through experimental analysis. The overall accuracy on all six labels was 99.67%.

Research Gap:

The study was only carried out on a single type of waste. Many different types of wastes get overlapped in nature and they are not segregated. However, the authors conclude that this type of categorization is under consideration.

In 2019,[15] Mazen, F.M.A. *et al.* used Tamura's texture feature to classify ripeness of the bananas. Brown spots formed on bananas were used for classification purposes. Identification of the correct maturity stage in fresh bananas can reduce farmers' work. The database consisted of four class labels such as green, yellowish green, mid-ripen and overripe bananas. HSV color space was used over RGB color space as HSV models describe the given color similarly to how a human perceives the color. The need of this proposed model was, as the classification of bananas is done by humans, time to segrate them and misclassification increases. First Image Acquisition was done. The dataset consisted of 300 images out of which 30% was used for testing purposes. After which Image pre-processing was done. A guided filter for edge preserving and smoothening of banana images was used. In image segmentation Otsu's method was implemented. Further, ripening factor of every banana was calculated as:

$$Ripening\ factor\ = \frac{Total\ area\ of\ Brown\ spots}{Total\ area\ of\ Banana}$$

To calculate the area of brown spots, a mask was used to detect the brown spots on the bananas. Tamuras' texture features were used to aid model in surface, shape and class determination of fruits. Tamura's texture features are based on psychological studies of how a human perceives. Further Levenberg-Marquardt backpropagation algorithm was used as the backpropagation algorithm. The input layer of ANN consisted of four neurons. The hidden layer comprised of 10 neurons and sigmoid activation function was implemented. The models' accuracy was calculated with the help of confusion matrix. Accuracy of 100% was achieved for class 1 and class 4 and of 97.75% for class 2(yellowish-green) and class 3 (mid-ripen) when tamuras' texture features was used with ANN. The model took 18 secs to classify the 89 images correctly.

4. CONCLUSION

Hereby, we conclude that we can use different segmentation methods for color recognition depending upon the task. Many models/prototypes cannot be implemented in real life as they are huge in size and many people don't know how to use them. Like using it in Malls on counters. Color detection/recognition mainly depends upon the images you show to the model. It has been seen that background changes in a particular image can affect its illumination overall. Many types of fruits and vegetables are subject to significant variations depending upon how they ripe. Even though, this model or technique is emerging, it can be used to solve many problems like proper pricing the vegetables and fruits that we purchase. Controlling the no-driver cars (Google or Tesla).

5. REFERENCES

1. https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac.

2. LeCun, Yann & Bengio, Y. & Hinton, Geoffrey. (2015). Deep Learning. Nature. 521. 436-44. DOI: 10.1038/nature14539.

3. X. Lei, H. Pan and X. Huang, "A dilated CNN model for image classification," in IEEE Access, vol. 7, pp. 124087-124095, 2019, doi: 10.1109/ACCESS.2019.2927169. [Accessed on 28.06.2021]

4. Wu, Jianxin. "Introduction to Convolutional Neural Networks." (2017)

5. Bernhard Preim, Charl Botha, Chapter 4 - Image Analysis for Medical Visualization,

   Visual Computing for Medicine (Second Edition), Morgan Kaufmann, 2014. https://doi.org/10.1016/B978-0-12-415873-3.00004-3

6. http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html

7. Yoav Freund, Robert E Schapire," A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", Journal of Computer and System Sciences, Volume 55, Issue 1, 1997.

8. Schapire, R.E. and Freund, Y. (2013), "Boosting: Foundations and Algorithms", Vol. 42 No. 1, pp. 164-166.

9. Chigozie Nwankpa, *et al.*, "Activation Functions: Comparison of trends in practice and research for deep learning" (2018). https://arxiv.org/abs/1811.03378v1.

10. Chuanqi Tan, *et al.* "A survey on deep transfer learning" The 27th International Conference on Artificial Neural Networks (ICANN 2018). https://arxiv.org/abs/1808.01974v1

11. K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778. [Accessed on 24.06.2021]

12. Feng Jiang *et al*." Image recognition of four rice leaf diseases based on deep learning and support vector machine", Computers and Electronics in Agriculture, Volume 179, 2020. https://doi.org/10.1016/j.compag.2020.105824

13. Jeong-Guk Kim *et al.*, "A study on object recognition using deep learning for optimizing categorization of radioactive waste", Progress in Nuclear Energy, Volume 130, 2020. https://doi.org/10.1016/j.pnucene.2020.103528

14. https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/

15. Mazen, F.M.A., Nashat, A.A. Ripeness Classification of Bananas Using an Artificial Neural Network. *Arab J Sci Eng* **44,** 6901–6910 (2019). https://doi.org/10.1007/s13369-018-03695-5