A simple and highly accurate object detection

Vishal Singh Bisht[1], Yash Vardhan Malpani[2], Adyaksh Sharma[3], Shivank Bansal[4], Nishi Sharma[5]

[1,2,3,4] B.Tech Student, CSE-Dept, ADGITM, Delhi, India
[5]Assistant Professor, CSE-Dept, ADGITM, Delhi, India

Abstract- **This paper presents the study of a robust and highly accurate object detection model for classification of food items. Our model uses ResNet50 with the optimization of stochastic gradient descent which uses less computational power and is space efficient. The model implements transfer learning to achieve higher accuracy even with a small dataset. In order to achieve accurate results we randomly flipped and resized the images from its base resolution to match the real world scenario.**

Keywords— Object Detection, ResNet, ResNet50, Food Classification

I. Introduction

This project focuses on multiclass object detection in single domain. The framework provides with classification of multiple food classes. Towards this end, an object detection system for classifying food into different classes is achieved. Its ability to detect and classify is very accurate. It operates on 300 x 300 pixel images from which the food items are detected. This project can further be used to classify real world objects into weapons which can then be proven very useful for security purpose.

There are two key concepts to our project for object detection. We will briefly present each of these concepts below and then explain them in detail afterwards.

The first concept is ResNet50 is a CNN which can be trained to classify multiple objects. It is a pre trained network which can be retrained on a new classification task on following simple steps. We use ResNet to implement transfer learning which trains highly precise model in less training data.

The second major concept is the highly efficient Stochastic Gradient Descent. It is a simple optimization technique which updates the model in response to output of the loss function. Stochastic Gradient Descent has been applied to large scale classification based machine learning projects, given that data is sparse.

Another contribution is a function which randomly resizes and/or flips the image in dataset to match the real world condition to improve the accuracy even with small dataset. So the data is being edited to improve the accuracy score in the testing phase.

An extremely accurate object detector will have lots of practical applications. This project can be trained to classify almost any object with little tuning if given the right dataset. In addition, the project can be easily operated over cloud without any critical system requirements. This model is trained to work even with small dataset which increases efficiency with small project which do not have large dataset in hand.

II. Features

A. ResNet50:

ResNet stands for residual network. ResNet uses shortcut connection also known as identity connection which counters the vanishing gradient during back propagation so it doesn't lead to a tiny gradient which can make the learning of initial layers intractable. Another advantage of shortcut connection is that we can train deeper network faster than possible previously [2]. There also exists ResNet101 and ResNet152 which consists of 101 and 152 layers respectively. We implement transfer learning with the help of ResNet in order to train accurate model even with a small dataset. Transfer learning uses pre-trained model of a large dataset and apply its learned feature maps for classification of a different but related dataset [8]. ResNet at its core makes use of batch normalization which increases the network performance by standardizing the inputs. Batch normalization also decreases the internal covariate shifts [4] and enables the use of larger learning rates [5]. ResNet implements deep learning using 2D convolutional neural network. It is very effective when used for classification and image recognition purposes [1, 3].

The weight update equation of neural network is as below:

$$\sum_i w_i x_i + b$$

Where,
w_i = weight of layer
x_i = input of layer
b = bias

B. Stochastic Gradient Descent

We used Stochastic Gradient Descent for optimization of the loss function. It is computationally fast as it replaces the costly operations [5] and processes one sample at a time. Since samples are being processed one by one, space is utilized much more efficiently and hence it is easier to fit into memory. Stochastic Gradient takes frequent steps towards minima of loss function which in turn helps it to hop out of local minimums of the loss function. Stochastic gradient descent generalizes the model well with fewer training errors [7]. For less amount of training time, it only leads to small generalization error.

III. WORKING

The core working is based on ResNet50 library. The 50 in ResNet indicates the number of layers as being 50. A pre-trained version of network that has been trained on large number of images from database of ImageNet has been loaded. A pre trained model was loaded from ImageNet database. This is known as transfer learning in which we use a pre trained model on a related dataset. By using this technique we achieved very high accuracy even with a small dataset. Rather than just using Convolutional Neural Network, ResNet is preferred because it is an ensemble of various mathematical techniques implemented in Deep Learning such as Convolutional2D layer, MaxPooling Layer. ResNet also uses identity connections which allow our network to have more number of layers by countering the vanishing gradient problem. This is further reinforced with add-ons like the Flatten, Dense and Dropout Layers. It can predict multiple classes of food item.

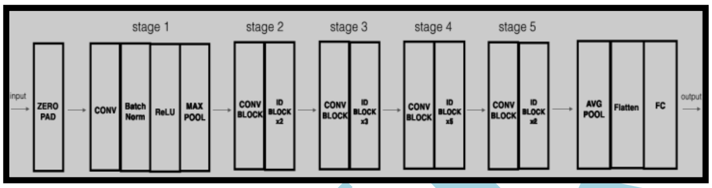

Fig. 1: Layers and processing in ResNet

As we can see from the image above, the very first layer in ResNet is 2D convolutional neural network layer. Before being passed onto the first layer, the dataset is preprocessed by undergoing zero padding. Zero padding helps preserve features on the edge of matrix. The output of the convolutional neural network is then put through batch normalization which helps in standardization of the inputs prior to activation function. Finally the max pooling layer reduces the dimensions of feature map which then decreases the number of learning parameters and thus reducing computational power.

IV. RESULT

A. Training Dataset:

Our training dataset consists of 750 image samples of each class each of them being 300 x 300 as base resolution. Further the images in dataset are resized and flipped to match the real world scenario and to achieve maximum accuracy. Initially our model shows a training accuracy of 46.8% which goes way high to 90%. Model takes approximately 20 minutes to train to achieve the maximum accuracy.

Fig. 2: Sample images of training dataset

B. Test Dataset:

Our validation dataset or test dataset consists of 250 image samples of each class. Each of the image samples are of base resolution of 300 x 300. Initially our model shows a testing accuracy of 74.6%. Later on the accuracy increases to 92%.

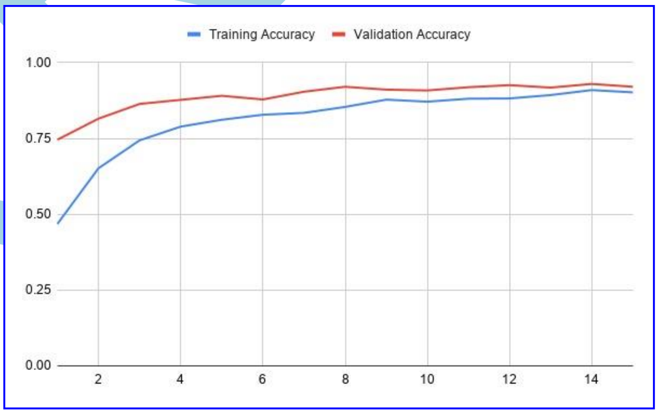

Fig. 3: Graph depicting training accuracy and validation accuracy against epoch

C. Real World Test Cases:

Apart from validation dataset, some image samples were also taken in our study. The model performed up to the expectations and identified the food items correctly. Below are some examples.

Fig. 4: Input image samples other than the dataset

V. CONCLUSION

Our study depicted a robust and accurate object detection model using ResNet and Stochastic Gradient Descent. Our model is very efficient as it can be trained from a small dataset and requires low computational power. It correctly classifies multiple food items in a given image sample in very less amount of time. Previous studies conducted by Gianluigi Ciocca et al., 2018 [8] has shown relevant results in similar domain with a lower accuracy than our model with ResNet50. In future our model can be utilized as base model for classification of objects into the weapons in surveillance.

REFERENCES

[1] Kelek MO, Calik N, Yildirim T. Painter classification over the novel art painting data set via the latest deep neural networks. Procedia Comput Sci. 2019 Jan 1;154:369-76.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[3] Ren S, He K, Girshick R, Zhang X, Sun J. Object detection networks on convolutional feature maps. IEEE transactions on pattern analysis and machine intelligence. 2016 Aug 17;39(7):1476-81.

[4] Santurkar S, Tsipras D, Ilyas A, Madry A. How does batch normalization help optimization?. Advances in neural information processing systems. 2018;31:2483-93.

[5] Chahar, P., Dalal, S.: Deadlock resolution techniques: an overview. Int. J. Sci. Res. Publ. 3(7), 1–5 (2013).

[6] Bjorck N, Gomes CP, Selman B, Weinberger KQ. Understanding batch normalization. InAdvances in Neural Information Processing Systems 2018 (pp. 7694-7705).

[7] Bottou L. Stochastic gradient learning in neural networks. Proceedings of Neuro-Nımes. 1991 Nov;91(8):12.

[8] Hardt M, Recht B, Singer Y. Train faster, generalize better: Stability of stochastic gradient descent. InInternational Conference on Machine Learning 2016 Jun 11 (pp. 1225-1234). PMLR.

[9] Torrey L, Shavlik J. Transfer learning. InHandbook of research on machine learning applications and trends: algorithms, methods, and techniques 2010 (pp. 242-264). IGI global.

[10] D. N. Le, B. Seth and S. Dalal, "A hybrid approach of secret sharing with fragmentation and encryption in cloud environment for securing outsourced medical database: a revolutionary approach," Journal of Cyber Security and Mobility, vol. 7, no. 4, pp. 379–408, 2018.

[11] [38] B. Seth, S. Dalal, V. Jaglan, D. N. Le, S. Mohan and et al., "Integrating encryption techniques for secure data storage in the cloud," Transactions on Emerging Telecommunications Technologies, vol. e4108, 2020.

[12] Ciocca G, Napoletano P, Schettini R. CNN-based features for retrieval and classification of food images. Computer Vision and Image Understanding. 2018 Nov 1;176:70-7.

