International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.7, July 2021 E-ISSN: 2347-2693

Food Image Classification Using Deep Learning Techniques

Yash Baid[1*], Avinash Dhole[2]

[1,2]Department of Computer Science & Engineering, Raipur Institute of Engineering and Technology, Raipur, India

Corresponding Author: yashbaid4500@gmail.com

DOI: https://doi.org/10.26438/ijcse/v9i7.1115 | Available online at: www.ijcseonline.org

Received: 22/Jun/2021, Accepted: 20/Jul/2021, Published: 31/Jul/2021

***Abstract*—**The recognition of image is one of the most important fields in the image processing and computer vision. Image recognition has many branches but the food image classification is very unique. In today's world people are very conscious about their health. Many people around the world use some dietary assessment system for planning of their diet. In dietary assessment system people make the use of food image classification to classify the food from the image. The classification of food images is a very difficult task as the dataset of food images is highly non-linear. In this paper, we proposed a method that can classify food images. We used pre trained models for the food image classification. The pre trained models is based on the convolutional neural network. In neural networks the CNNs is highly effective at the task of image classification and other computer vision problem. We classified a food image dataset i.e. food11 and obtained an accuracy of 96.75% in our experiment.

***Keywords*—** Deep Learning, CNN, Computer Vision, Image processing.

I. INTRODUCTION

In the today's world obesity and other health problem has been increasing day by day. In the paper [2], it is specify that the obesity has been doubled since 1980 in more than 70 countries. Obesity can cause many types of chronic diseases such as heart, diabetes, arthritis, etc and also decrease the immunity of the human system. Peoples have also shown their interest on reducing weight by calculating the calorie values of their food intake. Nutritional value of food should get more importance to prevent from diseases. To regularize the food habits of people they can make the use of the dietary management. Dietary management will help people by telling the information about the food they are eating. To get the calorie information, a system needs to detect the food from image and then analyze the dietary. For this paper the researchers is mainly concerned with the detection of food from the images.

Image processing and computer vision techniques are now the base of many domains. Food recognition from images is one of the domains of image processing. The recognition of food is a very difficult task as there is large similarity between food classes. Image processing also required more computation power than many text base data classification. The food recognition model should be very efficient, so people can run on a less expensive device too.

In the next section of the paper we will elaborate some literature, in section III we will discuss about how we have proposed the method, in section IV we will discuss about the results we get from the experiment, at last we will conclude our study.

II. LITERATURE SURVEY

The classification of food from images is a very difficult task because same class of food may have difference in images.

The author proposed a system of food image recognition using convolutional neural network. They use food-11 dataset for the training and testing of neural network. They used inception V3 model that is per-trained with the ImageNet. They get an accuracy of 92.86% in the experiment [1].

The authors discussed about the health effects of overweight and obesity in 195 countries that study over 25 years from 1980 to 2015. The authors have analyzed data from 68.5 million persons. In 2015, the obesity count on adult is 603.7 million & 107.7 million. Since the start of analysis in 1980, the obesity has doubled in more than 70 countries and has continuously increased in most other countries. 4.0 million Deaths globally has caused by High BMI. More than two thirds of deaths related to high BMI were due to cardiovascular [2].

The authors implemented a k-nearest neighbour algorithm and vocabulary tree algorithm to classify 42 food categories with 1453 images. To calculate the distance they choose L1-norm for SCD, EFD and GFD features and the L2-norm for the DCD features they get a Top 4 accuracy of 84.2% [3].

The authors implement a method that classifies the food images using SIFT and LBP (Local Binary Pattern) features in SVM classifier. SIFT feature is used to detect and describe local feature in images and LBP is simple to compute, it's a type of visual descriptor. The dataset used in this is PFI (Pittsburgh fast-food image) [4].

The authors proposed a method for the classification of food images using the sphere shaped support vector machine. They used the FCM (Fuzzy C-Means) algorithm for the segmentation of food images, FCM is very similar to k-means clustering algorithm, and sphere shaped SVM (Support Vector Machine) is used for the classification of segmented food idem. The proposed method automatically identifies the food items and also calculates their calorie value after identification. They apply this method on a dataset of Foodlog that contain total of 6512 images. In this experiment they got an accuracy of 95% [5].

The researchers proposed a novel method for the classification of food images using Random Forests (RF). Random Forests allow us to mine for parts simultaneously for all classes and to share knowledge among them. For the improvement in efficiency of mining and classification, the researchers are only considering the patches that are aligned with the image super pixels. The authors also created a novel and challenging dataset for the method, the dataset consist 101 food categories that have 101,100 images in it. For the labelling of food items the authors uses either the nutrition experts or Amazon Mechanical Turk. The researchers got 50.76% accuracy using the RFDC approach [6].

The researchers proposed a system for the improvement of accuracy in the dietary assessment. It can be done by analyzing the food images captured using the Smartphone. The main technique innovation of the researchers is to use of deep learning-based food image recognition algorithm. It uses the convolutional neural network based food image recognition algorithm to address the problem. The proposed system is test on the UEC-256. They obtained a top-1 accuracy of 54.7% and 81.5% accuracy of top-5. The convolutional neural networks are more scalable for large datasets, that's the reason CNNs is used more in food image classification [7].

The researchers proposed a system with deep convolutional neural network (DCNN) for the recognition of food from the food photo/image. In this system the authors uses a dataset that is already fine tuned and pre-trained for training. ImageNet dataset are used because it contains 1000 food-related categories. The researchers get an accuracy of 78.77% top-1 [8].

The researchers used the pretrained model GoogLeNet for the classification of Thai fast food images in TFF food dataset. The researchers are able to obtain an accuracy of 88.33 for 11 classes [9].

The authors proposed a model and compared the proposed model with several convolutional neural network models with food-11 dataset. They got 70.12% with their proposed approach, 80.51% with Caffenet and 82.07% accuracy with Alexnet [10].

The authors proposed a method that is used to increase the network in ways that aim at utilizing the added computation as efficiently as possible by suitably factorized convolution and aggressive regularization. The authors benchmark the system with the ILSVRC 2012 classification challenge validation set [11].

The authors introduced a new database for the image recognition. The author named the database as ImageNet, this database built upon the backbone of the WordNet structure. The authors also include the analysis of the new dataset (ImageNet) in its current state: 12 sub-trees with 5247 synsets and 3.2 million images in total, the author also says that the ImageNet is much larger in scale and diversity and more accurate than the current image dataset. The authors also make the use of the Amazon Mechanical Turk for the data collection scheme. The authors also show the usefulness of ImageNet with a simple object recognition program [12].

The authors proposed a method that uses the Multi-column Deep Neural Network, in this approach small receptive fields of Convolutional winner take all neurons yield large network depth, resulting in roughly as many sparsely connected neural layers as found in mammals between retina and visual cortex, only winner neurons are trained. Using this method several deep neural columns become experts on inputs pre-processed in different ways: predictions are averaged. It is also very efficient. It also uses graphics cards which allow the fast training of models. [13].

The authors develop a mobile application that analysis and tell the calories based on the food image captured using mobile device. The author creates a database know as food dairy for storing the food image. The food dairy database will be filled by the data entered by the user in the mobile app; here every user will have a separate food dairy database [14].

The authors create a web based application for the identification of food images from any image. The application can extract food images from other images, it also analyze the food balance, and visualize the log. The authors make the use of Food pyramid to estimate the balance of food. It is very simple and makes logging very feasible. The authors use flickr for the extraction of food images. The authors also implemented an interface for correction of some value, as system not always give the correct estimation [15].

The authors develop a mobile application that will be used for the accounting of daily food and nutrient intake. In creating this the authors use the image analysis tools for the identification and quantification of food that is consumed at a meal, in this application we capture the image two times first before consuming the food and second after consuming the food to estimate the amount and type of food consumed. Using two images help in determine the near to exact amount of food. The authors also convert the RGB image to YCbCr color space. The authors also make the use of the support vector machine for the identification of food items using statistical pattern recognition technique [16].

The authors proposed a DCNN that is used to achieve the new height for classification and detection in the ILSVRC14. The new incarnation that is submitted into ILSVRC14 is GoogLeNet, a 22 layers deep network [17].

The authors proposed a method that combines the residual connection network with inception architecture. The authors find that the training with residual connections increase the acceleration of training of inception networks [18].

The authors proposed methods that help in ease the training of networks. The authors reformulate the layers. The proposed methods of residual network is easy to optimize and can gain accuracy from considerably increased depth [19].

III. METHODOLOGY

We used convolutional neural network in our approach to classify food images. There are many types of different neural network but we choose convolutional neural network because it has been proven very efficient in image classification. The CNN learns the filters that are present in traditional algorithms were hard-engineered. We used inception V3 [11] model pre-trained with ImageNet [12] dataset, GoogLeNet[17] model pre-trained with ImageNet [12] dataset, ResNet101[19] & InceptionResNet[18]. Our task for classifying food images consisting of four processes:

- Selection of food dataset.
- Pre-processing of image.
- Training dataset using deep learning algorithm.
- Food image classification.

A. Selection of food dataset:

We are using the Food-11 dataset for our research. Food-11 dataset was created by the authors [8]. In the dataset, we get a total of 16643 images which is grouped into 11 food categories. The categories of food are Bread, Dairy products, Egg, Fried food, Meat, Pasta, Rice, Seafood, Soup, & Vegetables. We create a sample dataset of 3060 images from the Food-11 dataset. We divide the sample dataset into two parts: training set with 2295 images and validation set with 765 images.

B. Pre-processing of image:

We change the size of our sample dataset images to 299 x 299 x 3 to fit in InceptionResNet & Inception V3. For GoogLeNet & ResNet101, we change the size of our sample dataset to 224 x 224 x 3. The decrease in size helps in the increase of processing time.

C. Deep Learning Models:

In our research, we are utilizing the inception V3[11] , GoogLeNet[17], ResNet101[19] & InceptionResNet[18] models. The architecture of each model is given in figure.

The layers in the inception V3, GoogLeNet, ResNet101 & InceptionResNet are:

A. **Convolution Layer**: The convolution layer begins with the input size of 299 x 299 x 3 for the inception v3 and InceptionResNet model & for GoogLeNet and ResNet101 the input size is 224 x 224 x 3. The feature maps are created by convolving input images.
B. **Average Pooling Layer:** It reduces the variance and complexity of the data. It divides the input into rectangle pooling regions and computes the average of each matrix to downsample the features.
C. **Max Pooling Layer:** It is a sample-based discretization process. It is perform by applying max filter to non-overlapping sub regions of the input matrices. It extracts the very important features like vertical edges and horizontal edges.
D. **Concat Layer:** It is used to concatenate the multiple input blobs to single output blob.
E. **Fully Connected Layer:** It represents the feature vector for the input. It holds the information that is vital to the input.
F. **Dropout Layer:** It drops the elements randomly from a layer in the neural network. It is used to improve over-fit on neural network.
G. **Softmax Layer:** This layer assigns the decimal probabilities to each class in multi-class recognition problem. It must add up to 1.0. This additional constraint makes training quickly.

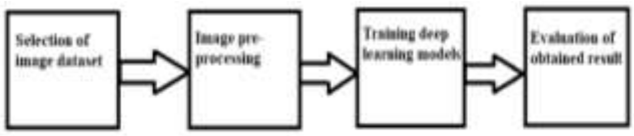

Figure 1:- Food image classification Methodology

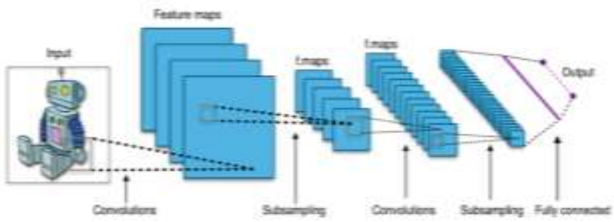

Figure 2:- Convolutional Neural Network Architecture

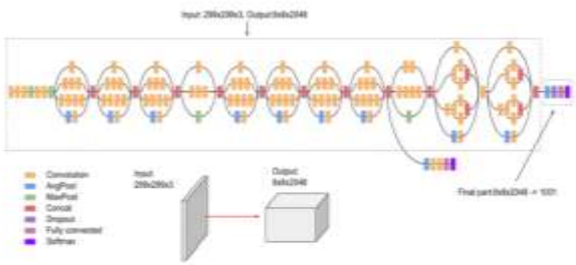

Figure 3:- Architecture of Inception V3[11]

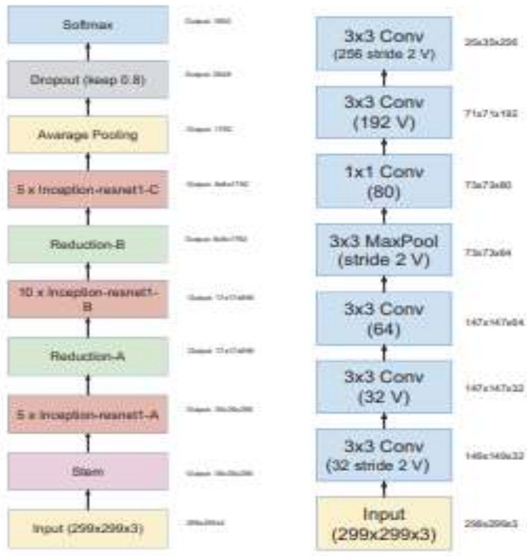

Figure 4:- InceptionResNet Architecture [18]

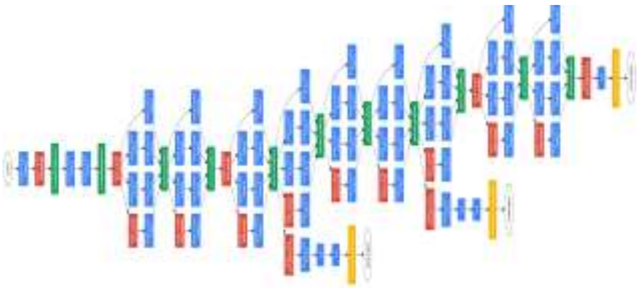
Figure 5:- GoogLeNet Architecture [17]

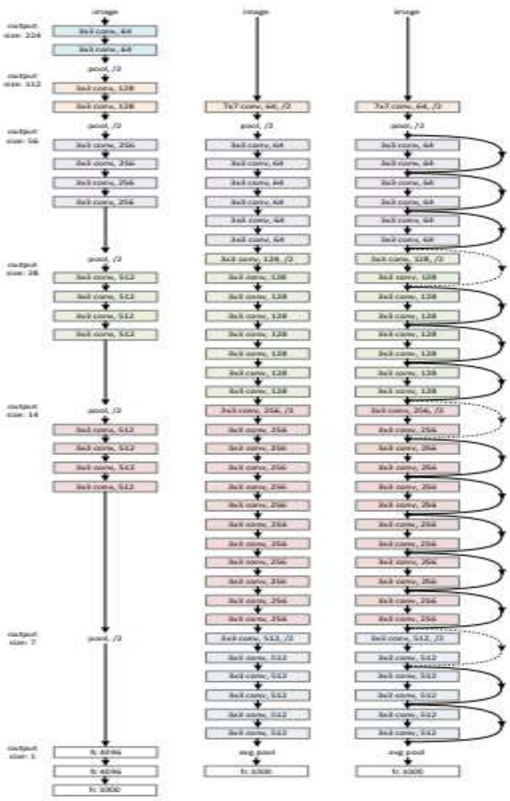
Figure 6:- ResNet101 Architecture [19]

D. Classification:
We trained our sample dataset with SGDM optimizer with initial learning rate of 0.01.

IV. RESULT ANALYSIS

In this section, we will describe the results found in our experiments.

A. Evaluation of model:-
We created our sample dataset from the food-11 dataset. Our sample dataset is consists of two parts: training, validation. We use training dataset for the training of models and validation dataset for evaluation.

B. Obtained Results:-
The obtained results by running different models with our sample dataset are given in below Table 1.

Table 1. Test Result in Different models for same dataset

Model	Accuracy
InceptionResNet	94.75%
GoogLeNet	86.0%
ResNet101	87.75%
Inception V3	96.75%

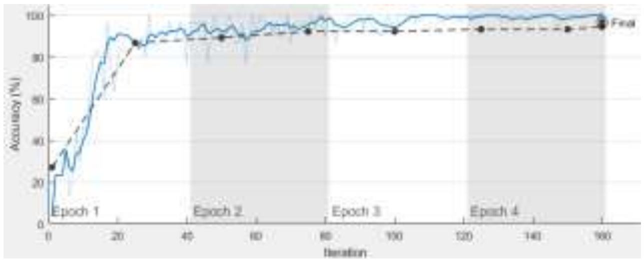

Figure 7:- Plot of Model accuracy on training and validation on Inception V3.

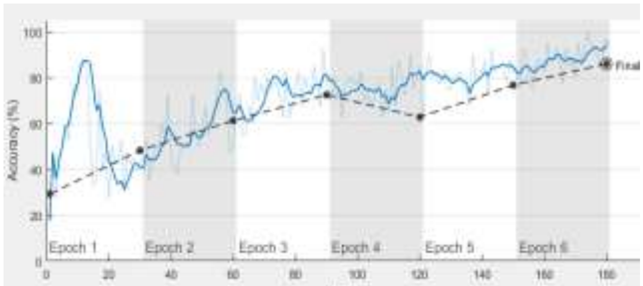

Figure 8:- Plot of Model accuracy on training and validation on GoogLeNet.

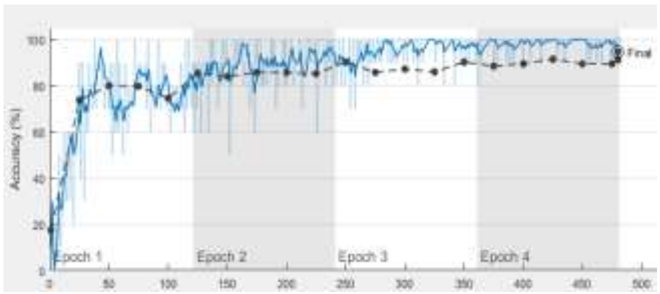

Figure 9:- Plot of Model accuracy on training and validation on InceptionResNet.

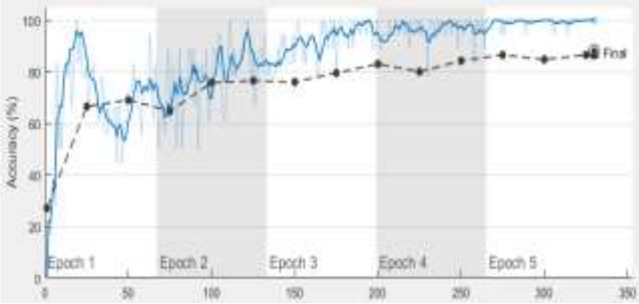

Figure 10:- Plot of Model accuracy on training and validation on ResNet101

From the above figures we can see that the graph of Inception V3 is the least fluctuating graph that helps us in getting the best result among all other models. In our experiment, we are using transfer learning technique. The transfer learning technique make the use of previous learned knowledge in new training dataset to classify images that is why our proposed approach has a better accuracy. In our proposed approach which uses transfer learning technique with inception V3 CNN has an accuracy of 96.75%, the next to Inception V3 is the InceptionResNet with an accuracy of 94.75% but requires much more time and powerful hardware. The last two is ResNet101 and GoogLeNet with an accuracy of 87.75% and 86.0% respectively but they both are faster than Inception V3 and InceptionResNet.

V. CONCLUSION

The Convolutional neural network uses a variation of multilayer Perceptrons and requires minimal pre-processing. We have used Convolutional neural network to classify images. We can conclude from our experiment that convolution neural network works well in food classification; we get an accuracy of 96.75% when we use our proposed model of Inception V3. The InceptionResNet is also gives a result of 94.75% but requires more time and power than Inception V3, ResNet101 and GoogLeNet is efficient than Inception V3 but provide less accuracy. The Convolutional neural network take time for computation compare to some neural network but once it is trained is very quick to classify. In future we can implement the proposed method in a mobile or web application for the identification of food images and calculating the calories of the food images; we can also test the proposed method on different dataset.

.

REFERENCES

[1] Md Tohidul Islam, Sagidur Rahman, B.M. Nafiz Karim Siddique, Taskeed Jabid, "*Image Recognition with Deep Learning*," 2018 International Conference on Intelligent Informatics and Biomedical, Bangkok, Thailand**, pp. 106-110, 2018.**

[2] "*Health effects of overweight and obesity in 195 countries over 25 years*,", New England Journal of Medicine 377, **no. 1, 13–27, PMID: 28604169, 2017**.

[3] Y. He, C. Xu, N. Khanna, C. J. Boushey and E. J. Delp, "*Analysis of food images: Features and classification*," IEEE International Conference on Image Processing (ICIP), Paris, **pp. 2744-2748, 2014**.

[4] Z. Zong, D. T. Nguyen, P. Ogunbona and W. Li, "*On the Combination of Local Texture and Global Structure for Food Classification*," IEEE International Symposium on Multimedia, Taichung, **pp. 204- 211, 2010**.

[5] S. J. Minija and W. R. S. Emmanuel, "*Food image classification using sphere shaped — Support vector machine*," International Conference on Inventive Computing and Informatics (ICICI), Coimbatore , **pp. 109-113, 2017**.

[6] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool, *"Food-101 – mining discriminative components with random forests*," Computer Vision – ECCV (Cham) (David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, eds.), Springer International Publishing, **pp. 446–461, 2014**.

[7] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma, "*Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment*," **CoRR abs/1606.05675, 2016**.

[8] K. Yanai and Y. Kawano, "*Food image recognition using deep convolutional network with pre-training and fine-tuning*," IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, **pp. 1-6, 2015**.

[9] N. Hnoohom and S. Yuenyong, "*Thai fast food image classification using deep learning*," International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), Chiang Rai, **pp. 116-119, 2018**.

[10] G. Özsert Yiğit and B. M. Özyildirim, "*Comparison of convolutional neural network models for food image classification*," IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, **pp. 349-353, 2017**.

[11] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, *"Rethinking the inception architecture for computer vision,"* **CoRR abs/1512.00567 ,2015**.

[12] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li and Li Fei-Fei, "*ImageNet: A large-scale hierarchical image database*," IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, **pp. 248- 255**, **2009**.

[13] D. Ciregan, U. Meier and J. Schmidhuber, "*Multi-column deep neural networks for image classification*," IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, **pp. 3642-3649, 2012**.

[14] Kiyoharu Aizawa, Makoto Ogawa, "*FoodLog: Smartphone based Multimedia Food Recording Tool*" International Conference on Artificial Reality and Telexistence (ICAT), Tokyo, Japan, **pp.143-144, 2013**.

[15] Keigo Kitamura, Toshihiko Yamasaki, Kiyoharu Aizawa, "*FoodLog: Capture, Analysis and Retrieval of Personal Food Images via Web*" CEA, Beijing, China, **pp. 23-29, 2009**.

[16] Fengqing Zhu, Marc Bosch, Insoo Woo, SungYe Kim, Carol J. Boushey, David S. Ebert & Edward J. Del, "*The Use of Mobile Devices in Aiding Dietary Assessment & Evaluation*", IEEE Journal of Selected Topics In Signal Processing, **pp. 756-766, 2010**.

[17] Christian Szegedy, Wei Liu, Yangqing Jio, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going Deeper with Convolution",CVPR, **pp. 1-9, 2015.**

[18] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", AAAI Conference on Artifical Intelligence, **pp. 4278-4284, 2017.**

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition, **pp. 770-778, 2016**.

