ELG7186
Computer Vision Fall 2022

2.

Architecture Selection & Literature Review

Group 8

Sports products Multi-Class Classification

1. Problem Statement

Nowadays a computer is a critical life need like food and water and we cannot imagine our life without the computer because our all knowledge, activities, works, memories, history, and lifestyles are managed and stored by the computer, in the work area it facilitates and manage many tasks very fast and in an organized way. One of the main problems that computer solve is classify products, in previous people used papers to classify products then they used some basic computers software to manage it but in case of using images as products description we need something different to use computer to solve this problem, our thoughts turns quickly to computer vision or machine learning techniques to help us to work with image classification. So, in our project we will work on deep learning project to help us in Sports products multi-class classification problem.

2. Literature Review

[1] Robust Sports Image Classification Using InceptionV3 and Neural Network

2.1.1 Summary:

Computer vision is assuming a fundamental part in the field of image or video distinguishing proof and examination. This paper presents a robust framework for the classification of various categories of sports using Neural Networks and InceptionV3. Sports images are divided into six classes, namely, rugby, basketball, tennis, badminton, cricket, and volleyball. Furthermore, the future scope of this paper is wide open. Our framework has successfully achieved an average accuracy of 96.64% over six categories, which demonstrates the effectiveness of the framework and can be used for the detection and classification of various sports activities. Comparisons have been made with other classifiers like Random Forest, K-Nearest Neighbors, and Support Vector Machine (SVM). The framework has achieved an average accuracy of 96.64%. Neural Networks have achieved the best accuracy as compared to other classifiers due to their ability to handle large datasets efficiently. These accuracies can also be increased as there is scope for the rectification of images, as some images are blurred and dark, which can be preprocessed.

2.1.2 Methodology:

There are two fundamental processes in image classification. Calculation and categorization of feature descriptors several sports videos from each category are first acquired from YouTube and then frames are taken from them. This data collection is sent into an image embedded for feature descriptor value extraction. Inception V3 is an image recognition model that extracts information using convolutional neural networks. In our case, 70% of our data is stored as training data, while the remaining 30% is kept as testing data. A data sampler is used to collect data samples for training and testing. Finally, a confusion matrix was created to evaluate the effectiveness of our training.

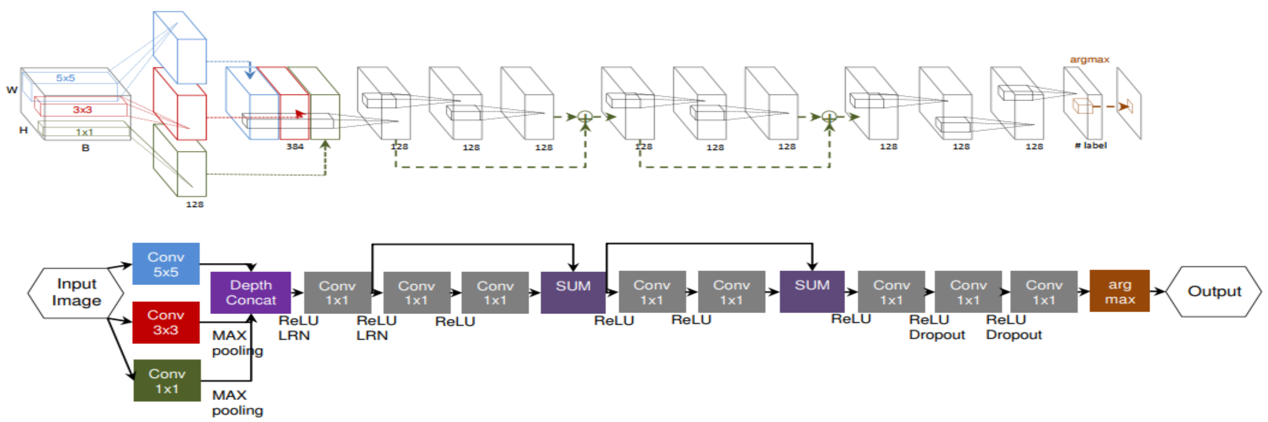
[2] Going Deeper with Contextual CNN for Hyper-spectral Image Classification

2.2.1 Summary

The authors used a deep convolution neural network (DCNN) unlike existing simple CNN, it can optimal explore local conceptual intersections by the help of the relationship of the neighboring pixels, the proposed architecture tested on three benchmarks datasets: the Indian Pines dataset (145X145 image size), the Salinas dataset (512X217 image size) and the university of Pavia dataset (610X340 image size).

2.2.2 Methodology

The authors proposed an architecture contains of fully convolutional neural network as show in the figure, the model contains of 9 layers and its deeper than other CNNs that related work used to work with the mentioned dataset and 9 layers are not very large to avoid over fitting, they also used data augmentation to increase the data samples. The performance of this model is better than the baseline model by 2.58% in Indian Pines dataset, 2.47% in the Salinas dataset and 2.86% for the Indian dataset.



[3] Deep Image Classification of a Wild Data Set for Olympic Sports

2.3.1 Summary and Methodology:

"Deep Image Classification of a Wild Data Set for Olympic Sports" (2016) is a paper that shows how to increase the success of image classification by using a deep learning technique. It used "GoogleLetNet", which is the topology for the Convolutional Neural Networks (CNN), to train 26 of the Olympic sports and 5362 images. It can reduce the number of parameters in the network and replace the fully connected layers with sparse ones, so the decision of classification was spread through this network. The authors collected these images from the web using "Google's Custom Search Engine tool." They split these images into three parts (training, validation, and testing) as 3763, 802, and 797. They also used data augmentation to increase the dataset that became 22522 images. To solve the imbalanced dataset, they used oversampling. So, they categorized some sports easily.

2.3.2 Experiments and Results

In this study, "GoogleLetNet" was applied to 256 x 256-pixel photos using 30 training epochs, 8 images per batch size, and a 1-epoch validation interval. Furthermore, it utilized "the Adaptive Gradient (AdaGrad)" and the learning rate (0.001) with a "Sigmoid Decay strategy" and 50% step and gamma (0.1). As can be seen in table 1, the outcomes varied depending on whether they trained the model with or without fine-tuning and with both the augmented set and balanced set.

	Raw	Raw(fine-tune)	Augmented	Balanced
Acc.(val.)	0.6262	0.8873	0.8997	0.9158
Acc.(test.)	0.6161	0.8733	0.8683	0.8745
Top5 Acc.(val.)	0.9047	0.9801	0.9826	0.9851
Top5 Acc.(test.)	0.8921	0.9749	0.9686	0.9787

Fig: Table 1 (Accuracy for the Olympic data set)

2.3.3 Strengths and Weaknesses

The obtained data set was used by the authors to train their model to obtain the unique accuracy for each class. After using the augmented, it did better in 7 classes while doing worse in 12 others. After using the balanced, it did better in 12 classes while doing worse in 12 others. Therefore, after using data augmentation, the best basketball forecast was 0.4963. The best handball forecast, using balanced data, was 0.999. Due to a conflict between two classes with similar probability, it therefore projected both very good results in some circumstances and bad results in others.

3. Proposed solution:

We will try different CNN architectures with hyperparameter tuning, for the proposed architectures we have:

a. The AlexNet CNN architecture, was developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, this architecture won the 2012 ImageNet ILSVRC challenge with a 17% top-5 error rate while the second-best achieved 26% top-5 error rate.
b. VGG (Visual Geometry Group) invented the VGG-16 architecture, which has 13 convolutional and 3 fully-connected layers, carrying with them the Relu activation function from AlexNet.
c. ResNet architecture it is developed by Kaiming He et al, this architecture won the ILSVC 2015 challenge with a top-5 error rate under 3.6%, using an extremely deep CNN composed of 152 layers.

4. References

[1] Joshi, K., Tripathi, V., Bose, C., & Bhardwaj, C. (2020). Robust sports image classification using InceptionV3 and neural networks. Procedia Computer Science, 167, 2374-2381.

[2] Lee, H., & Kwon, H. (2017). Going deeper with contextual CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, *26*(10), 4843-4855.

[3] Deep Image Classification of a Wild Data Set for Olympic Sports, Daniel Ferreira Moreira, C. Vasconcelos, A. Paes, Luis Velho less, Published 2016 link: https://www.addlabs.uff.br/workpedia2016/documentos/Artigo_08.pdf