ANATOMIZATION OF HUMAN ACTIVITY RECOGNITION WITH RESNET, OPENCV AND DEEP LEARNING

Vinitha.V[1], Velantina.V[2]

[1] *Student, Department of Master of Computer Application, AMC Engineering College Bangalore, Karnataka, India*
[2] *Student, Department of Computer Science and Engineering (M.Tech), C.B.I.T Kolar, Karnataka, India*

---***---

Abstract - *Human activity recognition involves predicting the movement of a person based on sensor data and traditionally involves deep domain expertise and methods from signal processing to correctly engineer features from the raw data in order to fit a machine learning model. The practical application of human activity recognition consists of automatically classifying/categorizing a dataset of videos on disk, Training and monitoring a task to perform correctly, verifying the task is done. Recently, deep learning methods such as convolution neural networks and recurrent neural networks have shown capable and even achieved results by automatically learning features from the raw sensor data.*

Key Words: **Deep Learning, OpenCV, ResNet , Kinetics dataset.**

1. INTRODUCTION

The Kinetics dataset, the dataset used to train the human activity recognition model. The description of the DeepMind Kinetics human action video dataset contains 400 human action classes, with at least 400 video clips for each action. Each clip lasts around 10s and is taken from a different YouTube video. The actions are human focused and cover a broad range of classes including human object interactions such as playing instruments, as well as human-human interactions such as shaking hands. The statistics of the dataset how it was collected, and give some baseline performance figures for neural network architectures trained and tested for human action classification on this dataset.

1.1 3D ResNet for Human Activity Recognition

In this architecture shows how existing state-of-the-art 2D architectures (such as ResNet, ResNeXt, DenseNet, etc.) can be extended to video classification via 3D kernels. These architectures have been successfully applied to image classification.

- The large-scale ImageNet dataset allowed such models to be trained to such high accuracy.
- The Kinetics dataset is also sufficiently large.

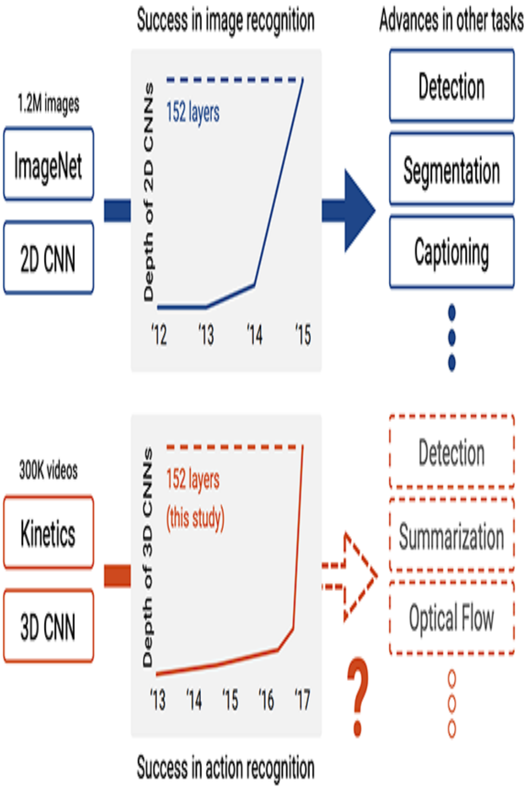

Fig -1: Deep neural network advances on image classification with Image Net have also led to success in deep learning activity recognition.

By modifying both the input volume shape and the kernel shape, the accuracy obtained as follows:

- 78.4% accuracy on the Kinetics test set
- 94.5% accuracy on the UCF-101 test set
- 70.2% accuracy on the HMDB-51 test set

In order to determine whether current video datasets have sufficient data for training very deep convolution neural networks (CNNs) with spatio-temporal three-dimensional (3D) kernels. Recently, the performance levels of 3D CNNs in the field of action recognition have improved significantly. However, to date, conventional research has only explored relatively shallow 3D architectures. We examine the architectures of various 3D CNNs from relatively shallow to very deep ones on current video datasets. Based on the results of those experiments, the following conclusions could be obtained:

(i) ResNet-18 training resulted in significant over fitting for UCF-101, HMDB-51, and Activity Net but not for Kinetics.

(ii) The Kinetics dataset has sufficient data for training of deep 3D CNNs, and enables training of up to 152 Res Nets layers, interestingly similar to 2D ResNets on ImageNet. ResNeXt-101 achieved 78.4% average accuracy on the Kinetics test set.

(iii) Kinetics pertrained simple 3D architectures outperforms complex 2D architectures, and the pretrained ResNeXt-101 achieved 94.5% and 70.2% on UCF-101 and HMDB-51, respectively. The use of 2D CNNs trained on ImageNet has produced significant progress in various tasks in image. We believe that using deep 3D CNNs together with Kinetics will retrace the successful history of 2D CNNs and ImageNet, and stimulate advances in computer vision for videos.

1.2 DEEP LEARNING

Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features. Deep learning algorithms seek to exploit the unknown structure in the input distribution in order to discover good representations, often at multiple levels, with higher-level learned features defined in terms of lower-level features. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to AI deep learning. Deep learning excels on problem domains where the inputs (and even output) are analog. Meaning, they are not a few quantities in a tabular format but instead are **images of pixel data, documents of text data or files of audio data.** Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.

1.3 OpenCV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library, there are many startups such as Applied Minds, Video Surf, and Zeitera, that make extensive use of OpenCV. OpenCV's deployed uses span the range from stitching street view images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects at Willow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a template interface that works seamlessly with STL containers.

Fig – 2: Human activity recognition

2. CONCLUSION

As the technology are blooming with emerging trends the model architecture had been modified to utilize 3D kernels rather than the standard 2D filters, enabling the model to include a temporal component for activity recognition. Finally, we implemented human activity recognition using OpenCV's dnn module Based on the results, the human activity recognition model is performing efficiently.

REFERENCES

[1] Neha Gaba, Neelam Barak and Shipra Aggarwal, "Motion Detection, Tracking and Classification for Automated Video Surveillance", IEEE 1 st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), pp. 1-5, 2016.

[2] L. Weixin, M. Vijay, and V. Nuno, "Anomaly detection and localization in crowded scenes", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 1, pp. 1975-1981,2014.

[3] G. Gayathri, S. Giriprasad, "Anomaly Detection for Intelligent Video Surveillance: A Survey", pp. 48-50, 2015.

[4] Qiang Li and Weihai Li, "Novel Framework For Anomaly Detection in Video Surveillance Using Multi-Feature Extraction", 9 th International Symposium on Computational Intelligence and Design (ISCID), Vol. 1, pp. 455-459, 2016.

[5] Gaoya Wang, Huiyuan Fu and Yingxin Liu, "Real Time Abnormal Croud Behavior Detection Based On Adjacent Flow Location Estimation", 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 476-479, 2016.

[6] R. Poppe, "A survey on vision-based human action recognition," Image and Vision Computing, vol. 28, pp. 976–990, 2010.

[7] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," Computer Vision and Image Understanding, vol. 104, pp. 90–126, 2006.

[8] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," Computer Vision and Image Understanding, vol. 117, pp. 633–659, 2013.