International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

VAZAM (Video Classifier)

Parth Vanarase[1], Dhiren Boricha[2], Suman Sasmal[3], Pravin Patil[4]

[1-3] *Student, Information Technology Engineering, Padmabhushan VasantDada Patil Pratishthan's College of Engineering, Maharashtra, India*

[4]*Professor ,Information Technology Engineering, Padmabhushan VasantDada Patil Pratishthan's College of Engineering, Maharashtra, India*

Abstract - *Video classification has found usefulness in world applications like detecting copyright issues, finding the origin sources, etc. The feature of having the ability to seek out the initial video has become crucial as we've seen a surge in video content which are uploaded not only to social media but also in thousands of various video platforms. during this paper we present a unique approach to try to video classification. Our method is capable of detecting videos that have suffered from distortions (such as change in illumination, rotation) or perhaps screened content, i.e. content that was recorded employing a smart phone during a movie house and also to urge the knowledge of the video clip. We combine several methods as thanks to measure similarity between videos and neural networks for object detection. Our results demonstrate the efficiency of mixing these methods.*

1. INTRODUCTION

Images and videos became ubiquitous on the net, which has encouraged the event of algorithms that may analyze their semantic content for various applications, including search and summarization. It has also become harder and harder to check authentication of those uploaded videos, thus protecting the copyright rights of the creator. Moreover, keeping track of the videos has proven much harder than anticipated. These videos are from different platforms, use different codec, some are for entertainment purposes some are for educational purposes. As people are obsessed over consuming more and more of the similar content that they need watched. There is no direct thanks to find the video from a clip it should be gif or 1 min montage. For Art we will use Microsoft Bing's reverse image search but nothing is obtainable for video. Our service proposes a technique that may reverse video search i.e., it'll present the user with the initial video (from which the clip been taken out). As this service progresses it will become more efficient and efficient, thus we can use the service for other purpose like detecting infringement of copyright, there are other applications planned for this service as we advance -it are often from user or company's specific need. We also are about to release a SDK for anyone who want to feature our service to their app/website.

2. METHODOLOGY

Firstly, we will be extracting the frames of the video. We extract a frame every 0.2 seconds and using this frame, we make a prediction using the Resnet-50 model. Considering we are using the transfer learning technique; we won't be extracting the final classification of the model. Instead, we are extracting the results of the last pooling layer. Until now, we had a feature map of one frame. Nevertheless, we would like to relinquish our system a way of the sequence. To do so, we aren't considering single frames to create our final prediction. We take a gaggle of frames so as to classify not the frame but a segment of the video. At the end, what we see on the screen could be a classification of the video in real-time.

3. OBJECTIVE

Aim is to classify video based on the action, face, objects represented. For the classification we will be using different types of models, like Resnet, etc. Last few years was marked with unprecedented increase of social media – like tik tok, Instagram, Instagram Reels, Facebook, Taka Tak ,etc, people spend more time on social media than any other apps in there phone. People are accustomed to the video being as short as 15sec. They watch funny clips, video performed by other people and news, yes news! This has led to influencing people with targeted fake news. There are many clips of series and movies which may or may not be credited. So here our service which not only will help the creators by giving them their deserved credit but also helping the user to find original video of that out of context clip.

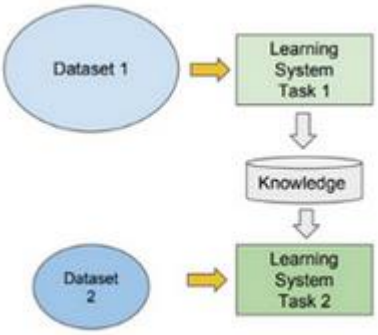
4. PROPOSED SYSTEM

Instead of just pixel comparison we will classify the whole frame, thus getting to know what that that frame contains either or all of the following trees, birds, animal, human, etc. To get more accuracy we will further detect faces, action of human, lighting condition, etc. Let's take a real-life example to know how we will handle certain situation sometimes social media or news uses clips taken out of context, which gives off wrong idea. Our service will first dissect them then remove recurring frames later it will be

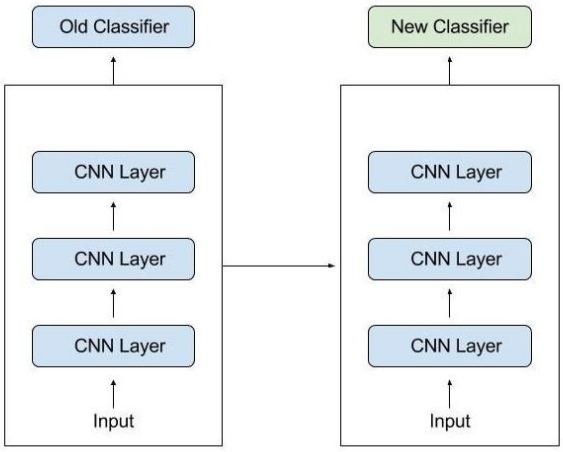
given to algorithm which will process that clip and search in the database to output the original video.

4.1 Transfer Learning

Transfer learning is not a new concept and it is usually specific to Deep Learning. There's an obvious difference between the standard approach of building and training machine learning models, and employing a methodology following transfer learning principles.



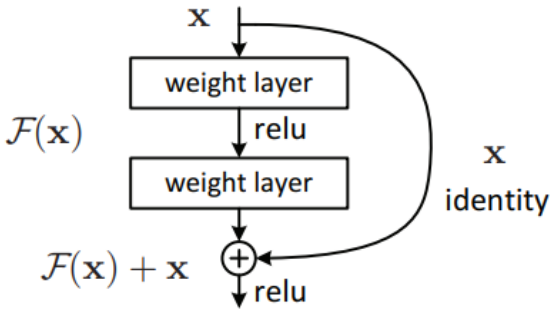
In this method, we can leverage the knowledge which includes features, weights from a previously trained model and use this knowledge for training newer models. So we basically attempt to exploit what has been learned in one task to strengthen generalization in another. This also solves the problem of having less data for the newer model.



In transfer Learning, we seek to transfer as maximum amount of knowledge from the previous trained model to the new model. This method has many advantages such as saving training time, better performance of neural networks and not needing plenty of information. As we are using transfer learning, the training time is reduced as it can take days or even weeks to train a deep neural network from scratch! We have used the ResNet-50 Model as our pre-trained model.

4.2 ResNet-50

ResNet-50 is a widely used deep learning architecture which is 50 layers deep. The model has over 23 million trainable parameters and is trained on a million images of 1000 categories from the ImageNet database. Using a pre-trained model may be a highly effective approach, compared if you would like to create it from scratch, where you would like to gather great amounts of knowledge and train it yourself. There are many pre-trained model which includes AlexNet, GoogleNet, Inception Network,etc. We are using ResNet-50 in this project since it has fewer error rates on recognition task and it also has excellent generalization performance.



ResNet introduces skip connections in the network and permits the gradient to go with the flow without getting elevated with weight matrices several times. Specifically speaking, the ResNet-50 model has 5 stages with a residual block. This residual block has 3 layers with both 1*1 and 3*3 convolutions. The concept of the traditional neural networks is that every layer feed into the following layer. But in a network with residual blocks, every layer feed into the following layer and immediately into the layers about 2–three hops away, called identity connections. This makes residual blocks quite simple.

4.3 Implementation

If we take a close look at the video classification problem, we come to know that it is not so different from an image classification problem. In the image classification task, we take images as inputs, use feature extractors such as convolutional neural networks to extract features from those images. After extracting features, we classify them based on these extracted features. We know videos are just a series of images. So we first extract frames from the given video and then follow the same steps as we do for an image classification task. So video classification is just and extra step to image classification problem. It's the simplest way to deal with video classification problem. In our model overfitting was a common issue here. Overfitting usually exist when you achieve a good fit of your model on the training data but it does not generalize on the unseen

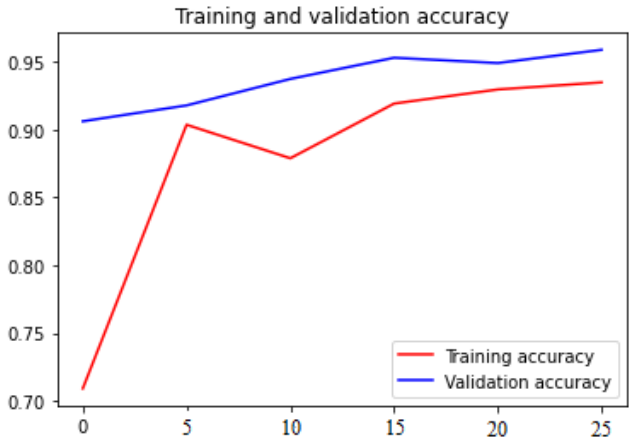
data. To prevent overfitting, we have added Dropout layer. Dropout is nothing but ignoring some of the neurons during the training phase which is chosen randomly.

Let's summarize all the steps that we did to build the classification model:

- First, we extracted frames from the videos to get the images for dataset.
- We use training set to train the model and validation set to evaluate that trained model.
- As video is series of image frames, we extract frames from all the videos in the training and the validation set.
- We then Pre-process these frames and train a model in the training set. After training we evaluate the trained model in the validation set.
- Evaluation the model in validation set is one of the most crucial part as it helps to know where out model stands in front of unseen data.
- We can train the model again and again by tweaking the parameters of we are not satisfied with the validation set results.
- Once we are satisfied, we can use the final trained model to classify videos.

5. RESULTS

After experimenting with different network architectures and tuning hyper parameters, the best result that we could achieve was 95% accuracy.



We can also see that the classes are constantly changing, as well as the respective accuracy for that class. These values constantly update every second until the video is over. We can use our service, video classifier, in the following way by attaching a camera to a gimbal and analyzing the video feed in real-time. Furthermore, we can use our service by training it with the particular dataset to detect different kinds of activities.

6. CONCLUSION

Our system is successfully able to classify videos in real-time situations. With different datasets and appropriate data, we can be able to detect various kinds of activities in the video.

7. REFERENCES

1. Video Classification using Machine Learning Shaunak Deshpande1, Ankur Kumar2, Abhishek Vastrad3, Prof. Pankaj Kunekar.

2. Large-scale Video Classification with Convolutional Neural Networks.

3. Video Activity Recognition: State-of-the-Art Itsaso Rodríguez-Moreno 1,∗ , José María Martínez-Otzeta 1, Basilio Sierra 1, Igor Rodriguez 1 and Ekaitz Jauregi 2.

4. Action Recognition in Video Sequences using Deep Bi-Directional LSTM with CNN Features.

5. 3D ResNets for 3D object classification Anastasia Ioannidou, Elisavet Chatzilari, Spiros Nikolopoulos, and Ioannis Kompatsiaris Information Technologies Institute, Centre for Research and Technology Hellas, Thermi 57001, Greece.