

Swear Words Censor in Video using Machine Learning and Computer Vision

C.J. Richard Patrick ¹

¹C.J. Richard Patrick, TechLead, CodinkTech Pvt Ltd

Abstract - Swear Words in video can be censored automatically. The swear words detection can help us in live feeds so we don't need to delay the video and censor manually. The data is applied using machine learning method CNN and computer vision to detect swear word and censor it in the video

Key Words: Machine Learning, computer vision, Swear Words Censor, Swear words detection, Artificial intelligence

1. INTRODUCTION

Based on the Machine learning methods to detect the Swear Words in video feed or live feed. One of the systems that can detect a Words in a real-time video is Swear Words recognition. Swear Words are classified according to their subject matter. One of the goals of achieving the project's objectives is to design a framework for Swear Words recognition. One of the most challenging and critical problems in computer vision is understanding hand movements. This system can be also included in the other Swear Words methods. The system uses the words in all media(text, sound, video) to know the Swear Words for censoring.

1.1 Why we need the ML model:

Every time when we need to detect a swear words, we don't need search for a word or professional swear words analyst. So the machine learning model is created to detect the swear words, based on the media. The new system can be more accurate in analysis swear words using multiple inputs. Making the ML model to connect in the direct system to increase the working ability and also the increase accuracy of the model. The usage of swear words to the model Programming and Computer vision make the system more user friendly also the computer based embedded into any edge devices.

2. Data collection and Data Preparation:

To create a good machine learning model we need data. One of the most difficult stages in any machine learning project is data preparation. The explanation for this is that each dataset is unique and tailored to the project. Nonetheless, there are enough similarities between

predictive modelling projects that we can describe a loose sequence of steps and subtasks that you can most likely complete. To collect data for the computer vision to process, the data from the dictionary and the word detecting method are used. To clean the data the video is made into frames to process, the video processing method are in the part of Computer Vision. The gesture is collected in the JPG format to process the data. The data are made in the motion base and gesture are to be processed and collected to enhance the accuracy of ML model and also Exploring data from a hand motion recognition database, It consists of 20,000 hand movements that have been labelled, this data can be used to improve the system.

The data collected on three different categories:

Hand gesture data is used to capture the hand motion and analysis it for the gesture in the system for censor. Body language data is used to analysis the body language of a human. These data are cleaned and processed as-per the machine learning model requirement as they are shown above. These data also collected in the live feed also when the system is on the detection of swear word.

3. Computer Vision:

Computer vision is a branch of computer science that aims to replicate aspects of the human vision system to allowing computers to recognise and process objects in images and videos in the same way that humans do. So one way to teach a computer to understand visual data is to feed it images, thousands, if not millions, of labelled images, and then subject them to various software techniques, or algorithms, that enable the computer to search for patterns in all the elements that relate to those labels. Objects do not necessarily have a rectangular form. Polygonal segmentations, based on this definition, are a method of data annotation in which complex polygons are used instead of rectangles to describe the shape and position of an object with greater precision. Semantic segmentation is a pixel-by-pixel annotation that assigns a class to each pixel in the image. By forming dots around the image, key-point and landmark annotation is used to detect small objects and shape variations. Face features, facial expressions, emotions, human body parts, and poses can all be detected using this type of annotation. Annotating an image with text, annotation software, or both to display the data features you want your model to understand on its own is known as image annotation. You're applying metadata to a dataset when you annotate an image.

For this system the computer vision has been used in various model to process data and inputs in various stages of process. The Gesture Analysing System and swear word Recognition are getting their inputs through the computer vision. The computer vision allows the video feed or an image to undergo the process of analysing the pixel in the image to information that can be made into the data for the machine learning models. In this system the computer vision works on two different processes, they are analysis of data and processing of data. Once we get the video feed or live feed the feed are made into the frames and these frames are feed inputs for the machine learning model. As the process data which is precollected from the resource are gathered and feed as input for the computer vision to analysis the visual part of the video or an image.

4. Machine learning model:

The Machine learning model is used to detect lies, this model have two input, the input are gathered from two different system. The system are The Gesture Analysing System and swear word Recognition System

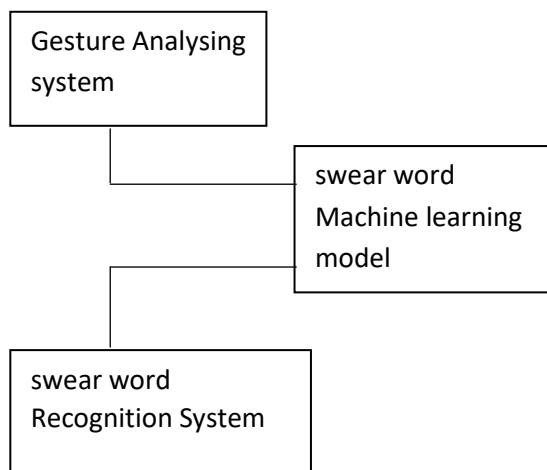


Fig -1: Machine learning model

As shown in the Fig -1 the machine learning model has two inputs these input are gathered from other system which is embedded along with the machine learning model. The lie detection model processes the input data along with the information the machine learning model has been trained.

The training of a machine learning model is based on the algorithm and the data used to train the model. The information for the model is processed from the two systems and the output data are serial cases and categories, they are independent variables to achieve the case that can be embedded in the system. The algorithms are chosen based for two systems: Gesture analysis and swear word recognition. The combination system has specific limitations that can be solved in the algorithm using input noise clearing. Enhance the pixel based

on the requirements of the study of the lies and facial and gesture made by the subject.

4.1 Gesture Analysing System:

As shown in the Fig -2 the feed is act as input to the Gesture Analysing System. The system processes the input through the computer vision. Using the computer vision and the data the system analyses the feed. The information has been processed in the computer vision will make the base for the system to detect the gesture.

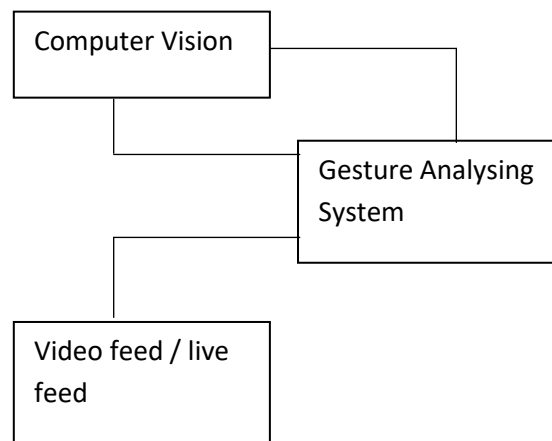


Fig -2: ML Gesture Analysing System

Our library has a vast number of common hand movements for all possible rotating, orientation, scaling, and deformation combinations. Aside from query input and database matching, one important function we want to accomplish is retrieving 3D motion parameters from the query image. Since query inputs lack pose detail, the best approach is to associate them. Due to the lack of pose information in query inputs, the best approach is to compare the query's motion parameters with the best obtained match from the database. Attaching the motion sensor to the user's hand and synchronising the picture frames with the calculated parameters is one way to calculate the corresponding motion parameters of the hand gesture. Another choice is to employ computer vision strategies to approximate the parameters directly from the database. Where the palm point and wrist point are found, an arrow may be drawn from the palm point to the wrist line's middle point at the bottom of the side. As the gesture detects.

4.2 Facial Emotion Recognition System:

As shown in the Fig -3 the feed is act as input to the swear word Recognition System. The system processes the input through the computer vision. Using the computer vision and the data the system analyses the feed. The information has been processed in the computer vision will make the base for the system to detect the swear word.

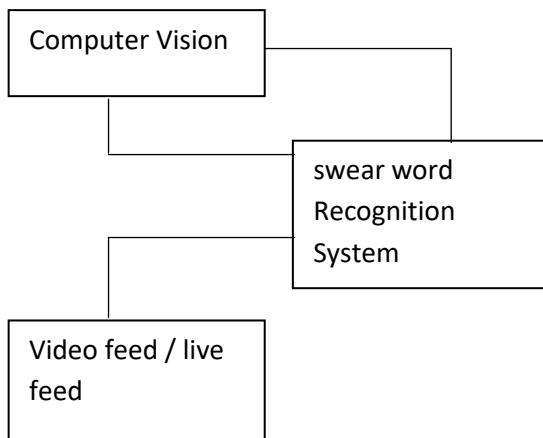


Fig -3: ML Facial Emotion Recognition System

swear word, regardless of national boundaries, ethnicity, or gender, is one of the most strong, normal, and fundamental signals for human beings to communicate their emotional states and intentions. With the advancement in artificial intelligence and deep learning, new methods for dealing with emotion knowledge in swear word representations have been proposed, which has increased recognition performance. swear word Recognition activities can be classified into two categories: static images and dynamic sequences, which take into account the dynamic interaction between constantly shifting images. The model of human emotions we embrace and interact with has a big impact on how Machine Learning models them. Based on the emotion the system detect the information that are process and feed as input into the system. The data and the learning are serve as the inducer for the enhancer for the system.

5. CONCLUSIONS

The model for swear word detection helps in various field of studies and also other two system which may helps analysis of swear word in the field of live news. swear word detection is one of the major challenges that is being faced by the child safety. The machine learning model can be detect the swear word and the process can be delivered in the more improved data based system and computer vision for analysis of swear word. This is a combination system that can be use in various system and also adaptive to any system of computation and low level devices.

REFERENCES

- [1] Tahani Bouchrika, Mourad Zaied, Olfa Jemai and Chokri Ben Amar, "Ordering computers by hand gestures recognition based on wavelet networks", Communications Computing and Control Applications, 2012.
- [2] Karen Simonyan and Andrew Zisserman, "Very deep convolutional network for large scale image

recognition", ICLR (International Conference on Learning Representations), 2015.

- [3] S. J. Pan and Q. Yang, "A survey on transfer learning", IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.
- [4] Assari, M.A.; Rahmati, M. Driver drowsiness detection using face expression recognition. In Proceedings of the IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia, 16–18 November 2011.
- [5] Wu, Y.; Ji, Q. Facial landmark detection: A literature survey. Int. J. Comput. Vis. 2019.