Real-Time Obscene Scene Nudity Detection and Blurring in a Video Clip

by

Jesima Rizwana
17201099
Md. Fahim Hasan
17201129
Motakabbir Hossain
17301078
Kaniz Ferdous Binte Zahangir
19101441
Humaira Mir Prothoma
19101622

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
September 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Jesima Rizwana
17201099

Md. Fahim Hasan
17201129

Motakabbir Hossain
17301078

Kaniz Ferdous Binte Zahangir
19101441

Humaira Mir Prothoma
19101622

Approval

The thesis/project titled "Real-Time Obscene Scene Nudity Detection and Blurring in a Video Clip" submitted by

1. Jesima Rizwana (ID: 17201099)
2. Md. Fahim Hasan (ID: 17201129)
3. Motakabbir Hossain (ID: 17301078)
4. Kaniz Ferdous Binte Zahangir (ID: 19101441)
5. Humaira Mir Prothoma (ID: 19101622)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on September 22, 2022.

Examining Committee:

Supervisor:
(Member)

Jia Uddin, Ph.D.
Assistant Professor
AI and Big data Department
Woosong University, Daejeon, South Korea
Associate Professor(On Leave)
Department of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh

Co Supervisor:
(Member)

Tanzim

Mr. Md. Tanzim Reza
Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

____

Md. Golam Rabiul Alam, Ph.D.
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

____

Sadia Hamid Kazi, Ph.D.
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Videos are widely consumed by people of all ages as a form of entertainment, information and education. However, not all videos are made for everyone. Many videos contain obscenities such as nudity, violence, blood, and gore which should not be watched by children or people who feel repulsed by these obscenities. Obscene content can negatively affect a child's mindset, and it can even traumatize people with weak mental constitutions. The real problem begins when these obscene videos are publicly available on the Internet, and anyone can watch them easily by downloading or streaming them online without getting any kind of warning. Moreover, people can even encounter these obscenities on live video streams or video calls. In our research, we have worked to detect and blur nude and obscene sexual content from videos in real-time. In that respect, this paper proposes a Neural Network-based approach. We have detected whether sexually explicit content is present in a video or not and blurred only the detected contents from the video frames. To detect nude and obscene contents, we have used different object detection algorithms such as Faster R-CNN, YOLOv5 and YOLOv6. These three respectively gave us mean average precision values of 0.382, 0.663 and 0.508 at 0.5 IOU threshold. Although with an mAP value less than YOLOv5, we chose YOLOv6 as it has proved to be the most optimal for our solution in regards of both accuracy and speed. And to blur, we have tried a total of five methods provided by two image processing libraries, OpenCV and PIL. Among those, we have selected the averaging method of OpenCV since it has best suited our needs. Additionally, we have attempted to reduce the rate of false positives so that any decent content does not get incorrectly labelled as obscene. This detection and blurring of obscene contents will contribute to ensuring safety in internet browsing for everyone.

Keywords: Video; Obscenity; Neural Network; Nudity Detection; YOLOv5; Faster R-CNN; OpenCV; PIL; Image Processing; Blur

Dedication

The dedication of our thesis work is respectively to our loving families, supportive sibling(s) and our beloved institution.

Acknowledgement

Firstly, all praise to the Almighty for whom our thesis have been completed without any major interruption.

Secondly, to our respective supervisor, Dr. Jia Uddin, for giving his precious guidance, motivation and time. We are lucky to have him as the friendliest and most helpful advisor throughout the thesis. Also, here is to our respective co-supervisor, Mr. Md. Tanzim Reza Sir, for his relentless effort, guidance, help and kind support in our work. We will always be grateful to these two amazing people and our best well-wishers.

Thirdly, we are very happy to express our appreciation and gratitude to the Department of Computer Science and Engineering, BRAC University and our educators for assisting us with all the fundamental help.

And finally, to our parents, without their throughout support it may not be possible. With their kind support and prayers, we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thoughts Behind The Detection and Blurring Work

Due to technological advancement, internet-enabled devices such as smartphones, laptops, tablets, smart televisions, etc., have become more affordable than they ever were. To demonstrate the current affordability and availability of these devices, more than 80 percent of people worldwide now own a smartphone, where it was only 49 percent back in 2016 [28]. These devices empower people by letting them access the vast collection of data available on the Internet. People access the Internet every day for various reasons such as communication, education, and entertainment. However, among all other things, people spend a significant amount of time every day on the Internet and other platforms to watch video content. According to Statista, in 2020, an adult person in the United States spent a staggering 312 minutes per day watching digital video on devices [32].

As a versatile media format, the usage of videos is diverse. Be it for the sake of education, work, gathering knowledge, or entertainment, videos are used everywhere. People of all ages consume videos through digital devices. Nevertheless, in the entertainment media nowadays, many videos created to entertain people contain sensitive and obscene content such as nudity and sex. These contents, in recent days, have penetrated the mainstream media deeply. Many mainstream movies and web series contain these, which are inappropriate to watch for children and people with strict moral values. Moreover, most sexual video contents objectify women, which can adversely affect the minds of our younger generation. According to Ward (2016), watching nudity and sexual content can result in "higher levels of body dissatisfaction, greater self-objectification, greater support of sexist beliefs and adversarial sexual beliefs, and greater tolerance of sexual violence toward women" (p.1) [12]. Viewers of such media can also experience changes in their sexual beliefs [13] and preferences, for example, they might get influenced to desire a particular body shape or practice sexual behaviours shown in these videos [4].

1.2 Aims and Objectives

Videos containing sexual content and nudity are more widespread on the Internet than anywhere else, and people can watch them very easily by downloading them or streaming them online. It is harmful for children to get easy access to such content all over the internet. Live streaming and online video calls are not invulnerable either. So, to protect the children from encountering these potentially harmful contents and ensure safety in internet browsing for children and adults alike, it is imperative that these contents are filtered and flagged. Furthermore, if this can be done in real-time, it will help ensure protection against these obscene sexual contents even in live video streams and video calls. That way, no one will ever have to stumble upon obscene contents without any prior warning. So, our target is to develop an algorithm that can detect sexually explicit content from videos in real-time, keeping the best possible accuracy. Also, we have tried to approach another model to conceal the nude parts in the video with black or red filled boxes. We will be proceeding with this in our following segments of research.

Which will be the most effective algorithm for real-time obscene nudity detection?

To implement this in detecting obscene scenes in a video, we have to train our neural network based on algorithms like CNN, Faster-CNN, YOLO, etc. and some related datasets. From all of these algorithms, according to [19], the YOLO algorithm is faster than other algorithms in terms of real-time object detection. The paper also mentions that while the YOLO algorithm does have some localization errors, it gives fewer false positives in the background. That is why our research will show the effective implementation of the YOLO algorithm in the pursuit of detecting real-time obscene scenes.

To solve the problem and get into the solution, we have to detect the issues that have been raised. Our foremost goal is to detect unwanted nude scenes in any video clip. To implement that, we have to use the You Only Look Once (YOLO) algorithm to detect objects within the videos. As a proposed model, it predicts the cluster centres and the model is trained for these boxes using either single instance of the picture. On the contrary, a Convolutional Neural Network (CNN) helps classify the images we find in our clips. Furthermore, through Support Vector Machine (SVM), we can distinguish between sexually explicit content and decent content.

The objectives for our research are mentioned below:

1. We will use YOLO since our primary goal is to work with real-time object detection.

2. We will be understanding CNN profoundly and work along, as we need to create classification while we are done with making or gathering datasets.

3. Our utmost goal is to detect unwanted nudity in the clips. So, we shall proceed to try some models in YOLO (YOLOv3, YOLOv4, YOLOv5 and YOLOv6).

4. Previously, we have proposed a model of YOLOv3 previously, we have studied YOLOv5, which was launched by Ultralytics in June 2020 and is now the most advanced object identification algorithm available.

5. Also, we have planned to blur the nude parts in the clips.

6. In our thesis, we would like to proceed with detecting the obscene scene at first by using three CNN models (e.g. YOLOv5, Detectron2 - for Faster R-CNN, YOLOv4). After that, we will compare the results and accuracy level.

1.3 Overview and Motivation

Obscenity in video clips, movies or documentaries is a common issue nowadays that is hard enough to be taken care of from the Internet. Especially some of these inappropriate images also take hold in people's minds, which eventually turn into many ungraded events or deeds caused by the people in real life. Sometimes, we might find some documentaries necessary and resourceful for the children and overall for all aged people. But in the middle of the video or documentary, some indecent pictures or scenes may appear, which is inappropriate in front of the children. Children nowadays are too addicted to gadgets, and eventually, one of their favourite pastimes is to watch TV or Movies. Sometimes, it is difficult to take down the unusual and unwanted scenes before it comes up in the absence of an immediate solution. Getting motivated by the journals which we have recently gone through, an excellent idea has come to our head to work on the matter [10].

Children and the young generation being attached to social media (such as video-sharing platforms like YouTube, TikTok, etc.) also impact them [18]. Sometimes, in various kinds of videos over the Internet, being uncensored and pirated might also be a significant problem. Nowadays, it has become almost a big challenge to find out the obscene contents and get those removed from the sites because there are vast amounts of these available all over the Internet. Even much of the content that is supposed to be labelled as adult content is not done. We can pluck the keyframes from the clips by using segmentation [18]. This will help us to detect the discrepancies and fix those sequentially. The foremost step to prevent this is to detect the related objects in the particular clips. We can classify the contents and detect whether that highlights nudity or not. If so, then we can use SVM (Support Vector Machine) [24] to allocate the collected characteristics into two groups (e.g., nude content and decent content). A region of interest (ROI) is a segment of a picture that we can filter or adjust. As a preprocessing stage of the process, skin-coloured area segmentation is conducted [24]. People are generally prone to violent scenes. However, there is a direct connection between watching violent scenes and mental trauma. According to [5], those who have watched the violent experience of 9/11 on television or other media were just as affected as those who watched it firsthand. It became indistinguishable from determining actual traumatization with the mimicked one. Due to watching this violence, the events were immediately traumatic to the public mass. While it is the right of every citizen to know and watch what is happening around the world, not everyone is capable of handling everything.

Some people cannot tolerate the blood or gory scenes presented on live television. I am bringing violent scenes into consideration because this way, it is easier to explain. The way violent scenes affect our minds is undeniable. That means videos have a huge impact on our minds and behaviour in certain situations. So, if we consider a kid being exposed to nude videos, we will realize that the effect can be devastating for that child. For example, if a child watches a naked video, he might think of women as objects rather than human beings just like him. Not only will the child be unable to learn the true meaning of relationships, but also, from an early stage, the child's mind will be scarred forever. According to [29], the crucial age for developing a child's brain is between 2 to 7 years. That will be our focus; that is what our research aims to protect. These are the above reasons why we cannot deny that videos, whether consciously or subconsciously, affect our behaviour and choices in life. We are hoping to mitigate these problems through our research. Our research will mainly focus on detecting the obscene scene in real-time. For this, object detection is essential. Object recognition consists of classification, localization, detection, and segmentation. What classification does is that it predicts the class of one image, that is, to classify what is inside the image, while localization predicts the location of an object. After that comes detection, which is the combination of these two, and it is a computer vision technique that works to identify and locate objects within an image or video. Specifically, object detection draws bounding boxes around the object of a given scene. Lastly, segmentation is a type of labelling where each pixel is labelled with a given concept. We can detect an object and gather feedback based on that detection by combining all of these. And then, we will use our algorithm and labelled data to conceal the obscene portions after detecting the figures/data with the help of YOLOv5, Detectron2 or YOLOv4.

Chapter 2

Related Work

In the past decades, the appearance of nudity has increased excessively in the media industry. Children and young people are exposed to explicit content in greater numbers with a wider range of content [25]. It is proven that both children and adults can have an immense impact on their minds as well as their sexual lives due to exposure to sexual content. According to the research conducted by Dr Jennings Bryant [9], 66% or more boys and 40% of girls were found to desire to engage in some of the excessive habits represented in the newspapers. Keeping the danger of children's future in mind, many researchers have worked on how this could be solved as much as possible. Many blocking systems are used to block such content, and it is also helpful to have reasonable parental control. Such content may not make anyone violent or sexually twisted, but it undoubtedly affects their behaviour and responses. It normalizes objectifying women and also crumbles our morals. Many kinds of research have been done to detect obscenity from an image or video. This paragraph attempts to better analyze similar previous research on Nudity Detection Systems in pictures and videos, particularly in video clips. We evaluate the different methodologies being used to achieve the stated outcome and demonstrate the requirements needed to identify images and videos due to sensor diversity, limited computational capacity, and the massive adoption of smart devices, which makes finding nudity detection techniques more difficult.

2.1 Related Works On Detecting Obscene Objects

The paper from [18] introduced ACORDE, which is a one-of-a-kind computational intelligence design for detecting inappropriate content in videos. It merges convolutional neural networks and LSTM in sequence-to-sequence connections. Testing mainly on available online NPDI datasets indicates that ACORDE improves earlier region algorithms throughout this purpose, and it eliminates false-positive results by 51 percent and wrong mistakes by a fourth. Nowadays, NPDI is the largest online sexually explicit data source, with approximately 78 minutes of content from 801 clips, half of which contain adult content. These data were retrieved from different online platforms. Moreover, the section of the non-adult category is subsequently segmented into 202 clips that are possible to identify and 201 videos that are difficult to categorize. Additionally, ACORDE is essentially the first methodology to use a particular framework for evaluating explicit content in media in literary works.

The article from [24] investigated the weaknesses of emerging CNN methodologies by concentrating on the spatial perception of CNN on the assumed nude geographic areas within the pixels in order to minimize the FNR (False Negative Result). Comparatively tiny pornographic substance was overshadowed by CNN's current approaches in the appearance of different perspectives. In the sexually explicit material and pornographic statistical approaches, the "You Only Look Once" (YOLO) methodology was applied to identify individuals as slight zones of interest (ROIs), that were later classified employing SVM and CNN. It showed that the object detector "You Only Look Once" (YOLO) outperformed the CNN-ONLY approach. Many evaluations were performed with the mentioned dataset to draw comparisons of the contribution of different CNN algorithms. In addition, a resection survey was carried out to prove the effect of including YOLO before CNN. In terms of accuracy, YOLO-CNN surpassed CNN-ONLY by a margin of 85.6% to 89.5 percent. Furthermore, when tested on the data set and model, the motion-based colour filter achieves a 93% accuracy rate. It takes 879 nanoseconds to generate three sequential screenshots for every end-user.

The report from [15] proposes a system to detect inappropriate scenes, including nudity, drugs, gore, etc., present in video streams. It was done in three different steps, and the first one was converting the videos into several frames. After that, three different algorithms were used to detect the ill-suited scenes, and lastly, the percentage of inappropriate scenes was calculated and shown as the given result. The detection of objects and scenes was done with the help of CNN's Object Detection algorithm. The nudepy library from python helped with the nudity detection, which works based on detecting skin-coloured pixels and helps identify nudity based on pixel count and its region. Moreover, with the help of the model, detection and classification of any video as pornography is possible if it exceeds the base scale mentioned in the model. It has been shown that the devices used to create the framework gave an estimated 90 percent accuracy in determining both the nudity and violent content of the video for nudity detection.

Another study from [1] proposes a model for detecting nudity based on skin colour. The steps in determining nudity through the skin are detecting skin-coloured pixels in an image, detecting skin zones relying on identified skin pixels, examining skin zones for sexualisation or non-nudity signs, and categorize the image as practically naked or not Based on these assertions, skin pixel intensities must be distinguished from non-skin pixels. The skin screens that have been recognised are looked at to see whose are associated to constant zones. There is no significant difference in the results of the skin filtration on testing and training images, according to analytical outcomes. On the testing dataset, the skin filtration has a detection capability of 96.7% and a fake-positive rate of 9%. There are many people with different skin colours, so this is where we need colour spaces. According to [1], color space is a configuration of a coordinate frame and space time inside one mechanism for which color is represented by just one point. There are various colour spaces used for digital images, and the most widely used one is RGB colour space, where every colour appears with the combination of its primary colour components. However, the RGB colour space only considers the primary colour and its combinations, so it is not reliable enough to detect skin colour. Therefore, to determine our natural human skin colour, we must consider many factors, such as luminance. We cannot get a clear picture of luminance with the help of RGB colour space only. This is where contrapuntal color enters the picture, that we may obtain by removing light output or through some kind of conversion. Consolidated YCbCr, RGB and other modifications are popularly used within skin color experiments. Again, if we want to go for a more realistic approach, then not only luminance, but we also have to consider hue, saturation, etc. Hue defines the dominant colour, for example, the distinction between red and yellow. Saturation in colour means how saturated a particular colour is in an area. It is the colourfulness for the area of percentage to its lightness. To deal with this, we use HSV colour space. Likewise, if we want to go for a more realistic approach, we can add more colour spaces.

The paper from [7] introduces a new method called SafeVChat, using a motion-based skin detection system that has better precision and significantly higher recall. It resolves the problem of flasher detection that can occur in online video chat systems and can be problematic for underaged users. The proposed model detects nudity with the help of four identification and calculation phases. At first, it calculates the target region with the help of motion in continuous screenshots. Then, some discriminative features were specified through further analysis using the Chatroulette dataset. Generally, behaving inappropriately subscribers on web video conversations normally keep their faces hidden to avoid providing their individuality. The regular chatters show their faces most of the time, but some do not feel comfortable doing so. Given this, the method incorporates a new skin sensor module that recognizes non-face skin in the target location using different skin color schemes. Following detection, it computes the percentage of non-face area of the skin to identify area and uses the percentage to predict misbehavior.

The study in [8] proposes a new method for obscene scene recognition. The algorithm works by classifying the video files independently using three features. The first functionality is based on just one keyframe data, the third time on 3D temporal and spatial amount, while the final on motion and duration qualities. To split the clip into clip events, a new keyframe harvesting methodology was being used. It works by incorporating the complexity of two sequences with the structure's statistical features. The proposed algorithm works in three stages: preprocessing, feature extraction, and classification. In the preprocessing stage, the detection of the keyframe is performed. The previously mentioned characteristics are gathered in the second stage, followed by the classification process. Comparing the results with the proposed algorithm and the existing methods, it can be seen that the recognition rate has increased by more than 9.34 percent.

We know that people have different skin tones. Based on that fact, it is crucial to detect nudity by separating the skin types, as that helps more accurately to detect obscenity. From the following journal, we have denoted a model for skin colour to detect adult scenes in an image [3]. Though we will work with videos, we have to know how we would be able to split the videos into images and then start working with that (because we cannot apply segmentation until we split the videos into images with every possible sequence). The algorithm they have used in their paper is about dividing the images into blocks as it helps in detecting part by part.

These blocks help detect the naked blocks in the images. The matching chroma distribution can be determined by using a neural network, which can be utilized to detect the skin area [3]. The roughness characteristic is also used to filter out non-skin items, allowing the skin region to be recognized more precisely. There is also a fact to be mentioned on how we can work with the lighting condition. The one-class-one-net neural network approach has been applied in the following journal, which tells us how the object's chroma distribution can be found over the Internet [3]. So, overall we see that this entire plan can help us reaching one of the desired goals for the research: to differentiate between the obscene and decent scenes so that we can easily detect them.

2.2 Related Works On Blurring Obscene Objects

The research paper in [20] proposes a model that can detect and blur both at the same time. The model detects and blurs based on a predefined database class using fragmentation. In addition, Inappropriate scenes such as nudity, drugs, bloody scenes, weapons and so on are detected and blurred through any streaming videos. To detect explicit content, segmented images will be evaluated in a dataset to determine whether or not nudity is present. Moreover, The video would be categorized as pornography if it has been associated and outperforms a certain level. To determine whether an image is explicit or implicit, each frame must be examined. Furthermore, the requirement set for the photo to be assumed obscene for the model for 79 percent. As a result, any photos that exceed the achieved limit are deemed explicit and blurred. Likewise, the derived statistics of all other items or unnecessary sceneries would be matched to a dataset to determine the proportion of inappropriate scenes in a clip. If there is any nudity down that route, the viewer will be able to notice it in advance, and the model will urgently blur out certain parts of the clip for explicit content. The model is 93 percent reliable in producing the desired result.

The proposed model in [16] is a picture-to-picture interpretation method which is predicated on training data that indirectly retrieves and covers sexual content in responsive areas in images while maintaining the image's semantics. As a consequence, It converts a photo x through the responsive target axis X (naked women) to a photo y through the non sensitive target axis Y (women in swimsuits), with the vulnerable spots immediately hidden behind swimsuits. To find the regions, the model used complex simulations. Moreover, The method would not enable matched training dataset and returns impressively accurate outcomes, clearing the way for the unique activity of effortless nudity restrictions to be solved. However, This also plans to investigate the impact of various structural options and damage functional areas on the raw pixels, as well as encode the method in a search engine implementation to prevent the public from unlocking inappropriate content. Therefore, The outcomes are graphically remarkable, demonstrating that effortless nudity prohibition is conceivable with minimal content gathering and analysis exertion.

According to Wehrmann et al. in the proposed model in [18], ACORDE (Adult Content Recognition with Deep Neural Networks) denotes a semantic segmentation strategy that employs a convolutional architecture similar to an autoencoder as well as a Long Short-Term Memory network (LSTM). Moreover, ACORDE facilitates for extracting of input images from NPDI sequences in order to generate live stream conceptual identifiers that are used by LSTM to analyze the clip. Moreover, In the final stage, the entire stream appears to be working by removing absolute necessity, which is generally fine and retraining CNN. Furthermore, ACORDE has taken the lead in adult clip identification in NPDI.

Following the described model in [2], It results that the used method can detect approximately half of the nude-content images in a limited testing dataset, with approximately 11% of the secured photos inconsistently labelled as nude-content or, at a consisting of separate types to detect 91% of nude-content photos with a 34 percent probability of error. Unfortunately, Because protecting photos outcompete those with nude content, there are plenty of negative warnings which imply that sufficient effort is necessary. Moreover, The structure is now used for photo safe-searching and has been integrated into Google's nude-content masking architecture. Head computer vision can evaluate demographic factors of people in a photo, which could turn out to be useful tools, even if only for prediction. More detailed interpretations of the skin tone particle structure could be useful without significantly increasing computation complexity. Furthermore, Improved pattern initiatives that go beyond simple segmentation could support skin area recognition. Finally, a photo comparison corresponding to a nude-content photo data store could be advantageous. Because each strategy appears to have its own set of benefits and drawbacks, any photo structure can be used to supplement the results of a skin concealer. Therefore, A low-cost experiment set that can be used by both researchers and commercial entities would help to accelerate the progress.

This paper from [23], The Adversarial Promotional Porn Images (APPIs) are obscenely used for the subsurface branding for multiple hidden motives such as fraudulent online advertising, spoofing, and some other immoral purposes. Previously, image spam was being used, but recent time, the use of considering the potential sexual assault to draw the spectator has rapidly increased, and several juggles are being employed to hide individuals from automatic checkers. In this document, a novel approach for large scale revelation of such images, named Melena (Malicious Explicit Content Analyzer), is formed for acknowledging such adversarial images and the underground business behind them, which merely tends to focus on the zone of an image where sexualization is least ambiguous and visible to the intended audience. This method assisted in the discovery of 4,500 APPIs from 4,042,680 images from the most well-known social advertising sites, as well as in bringing to touch the different approaches used to avoid the popular overt content detection, as well as the rationalization behind the productive use of such methodologies. This method reduced dramatically the portions of the image at which noise implantation is most likely to happen. This method was accurate 91% of the time and could recall the images 84% of the time.

According to the paper from [22], proposes that, in this advanced modern environment, images and video clips are being uploaded to the internet at an alarming rate. Unfortunately, some of the photos downloaded, whether intentionally or unintentionally, include their identity document (ID), which malicious hackers can use to steal one's identity. Entering a financial institution in the victims' names, attempting to access the victim's accounting statements, forging a license or other forms of identification, and so on. As a result, the investigators addressed the task of preventing the existing challenges by detecting and censoring an entity that could also contain relevant personal sensitive data using TensorFlow and OpenCV. In this analysis, the super-fast R-CNN image recognition design is utilized to identify IDs in photos and is matched to certain other machine learning algorithms (SSD, R-FCN) by evaluating one's mean Average Precision (mAP) of TensorFlow. The mAP of 71.5 percent was achieved by the super-fast R-CNN to ResNet101 approach. For applying the solution, the scholars blurred the entity with OpenCV in Python by using bounding box coordinates generated by the image recognition method. This study will be beneficial in identifying each and every conceivably responsive information privacy outflow and it will be capable of helping individuals who are unfamiliar with the risks of submitting their proof of identity to the website.

Analyzing the usage of various techniques of image blurring and determining the effectiveness against facial detection algorithms were the sole purpose of this paper [21]. ITo put it in another way the objective is to find out which technique is the most effective in reducing the possibility of facial detection. With a view to revealing the threats imposed upon privacy anticipated from blurring images. Box blurring method, Gaussian blurring method, and a different privacy-based pixilation method were the three blurring algorithms which were designed and implemented for blurring. After considerably large tests being done on each of these algorithms, inclusive of effects of blurring on color and grayscale images, images with and without faces, and the effectiveness in hiding faces of each method, the conclusion was the privacy blurring method is more effective than mitigating facial detection.

Numerous image files have blurred zones engendered by rotation or gaussian blur. Fully automated recognition and categorization of blanked input images is critical for a variety of video production analysis assignments. This article from [6] describes an easy-to-use method for recognizing and classifying blurred regions in images. Blurred photo zones should be first identified in the suggested methodology by investigating parameter knowledge to every pixel location. The blur sorts (for example, kinematic blur or gaussian blur) are therefore decided and use an alpha-connected restriction that does not require photo-unblurring or low contrast operating system assessment. Moreover, comprehensive investigations were performed on a consecutive series of 300 blurred visual features and 300 blur-free photo zones retrieved from 150 digital files. The investigational findings show that the suggested methodology appropriately works by detecting and characterizing the two main types of photo blurs. The suggested method is applicable to a wide range of audio and visual modeling approaches, including edge detection, detail evaluation, and pattern recognition. This method yields an accuracy rate of 73.6%.

According to this journal [26], Olympians in energetic, collaborative sporting events must define, scoop up, and relate to various information to a small amount of time for responding appropriately. As a result, experiential abilities are an important predictor of aristocratic athletic effectiveness. Athletics researchers have evaluated methods to analyze, and prepare spatial reasoning, among them involves the application of blurred stimulation. In this section, it characterized the two main techniques for generating blur (Gaussian and dioptric) and thereafter discussed recent evidence in a game context. Ultimately, the research has demonstrated that the use of blur can improve beginner attendees' abilities and understanding of various sports duties, particularly when blur is implemented to observational and experimental conditions. Whereas the approximate and specialist extent contestants are moderately not affected by the appearance of blur, it remains to be seen whether it has beneficial adverse effects on knowledge gaining. There's a main segment, it examined the possible analyze of using worlds to enlarge through the present study by having great situations, as well as a number of the technical inaccuracies that limit the effectiveness of blur.

This paper from [17], shows the method of detection, tracking and blurring a face in a video frame using Viola-Jones detection algorithm for detection and KLT algorithm for tracking the attributes in the video frames. The filter for blurring is posed after the identification or detection of the face is done as well as identification of the feature points for tracking. A competition to real time detection,Viola Jones algorithm was used to detect features in a video sequence. Haar basis functions(simple rectangular features) are used in this method and is applied to determine the facial features. In order to get feature tracking Kanade Lucas Tomasi algorithm was used which is the most popular one. This algorithm helps to detect feature points which are scattered and contain enough texture for the required points in a good standard. A continuous human face tracking is possible in a video frame with the help of this algorithm when the brightness of the image is constant . After detection and tracking blurring is applied by filters using circular averaging filters.

In this project [31], an android-based software application is developed on automated face detection and blurring algorithms. This technology's one of the main applications is to protect the privacy for both the safety and legal concerns of the subjects present in the video. The successful development of this tool enables the application of behavioral observation based on video at the same time protecting the identity of the subjects who have participated. This technology includes automated human face detection and automated facial area blurring in a real-word video. False alarm and missed ratio of face detection is also included. This technology runs on a smartphone and performs the video capture and face blurring at real time. This enhanced facial detection algorithm is better than the baseline implementation and also out-performs the current implementation of the YouTube face blurring tool. A selective blurring feature is also added with a view to allowing more freedom of selecting particular facial images to be blurred at the same time leaving remaining facial images unchanged.

According to the preceding discussion, most researchers develop nudity detection models using YOLO and CNN, which are common approaches that use skin colour, motion, etc. Moreover, it would suggest the difficulties that nudity detection presents in models. For example, it necessitates important massive computational control or disk usage. Besides this, the tools are varied.

Chapter 3

Proposed Method

3.1 Neural Network Architecture

The Neural Network architecture is consist of individual units, called neurons, an imitation of the biological behaviour of the brain. The main job of a neural network is to alter input into a meaningful output. It is mainly composed of an input layer, an output layer, and one or more hidden layers. The input layer retrieves the data necessary to train our concept from an independent factor. The output layer takes input from the hidden layer and comes to an ultimate prediction based on the model's training and provides the final result. The hidden layers within these two layers are the intermediate layers, which perform all the computations and bring out the features from the given data (Figure 3.1).

3.1.1 Types of Architectures

A new version of the YOLO model was introduced in June 2020, which is more user friendly and less work than the YOLOv3(Figure 3.2). And that is YOLOv5. YOLOv5 gives a similar result as YOLOv3 but with 75 percent fewer operations. This is why our decision to work with YOLOv3 has changed to YOLOv5. Figure 3.1, 3.2, 3.3 and 3.4 show some examples of the types of architectures of the models we are working on.

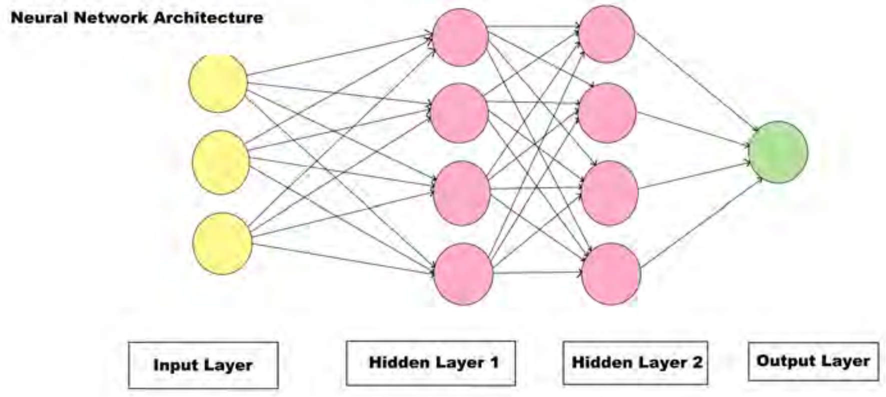


Figure 3.1: Neural Network Architecture

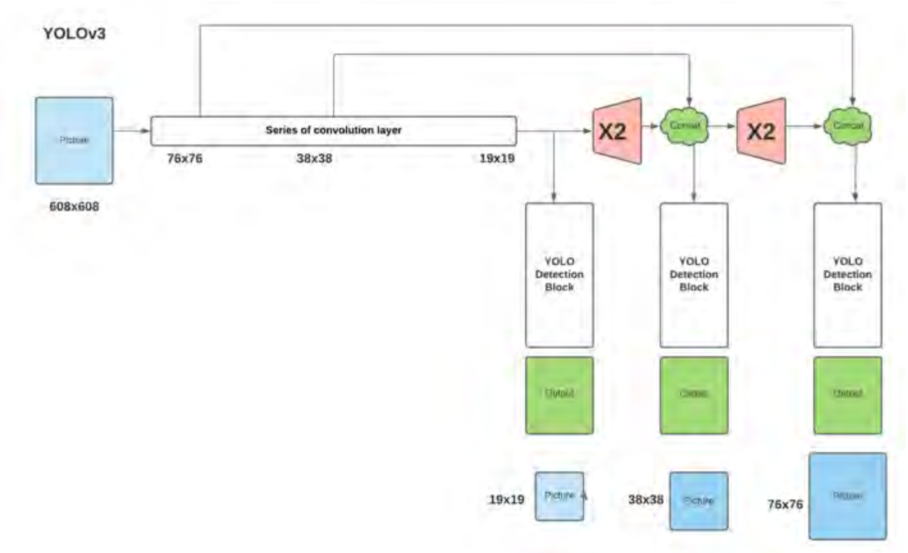


Figure 3.2: Architecture of YOLOv3 Algorithm

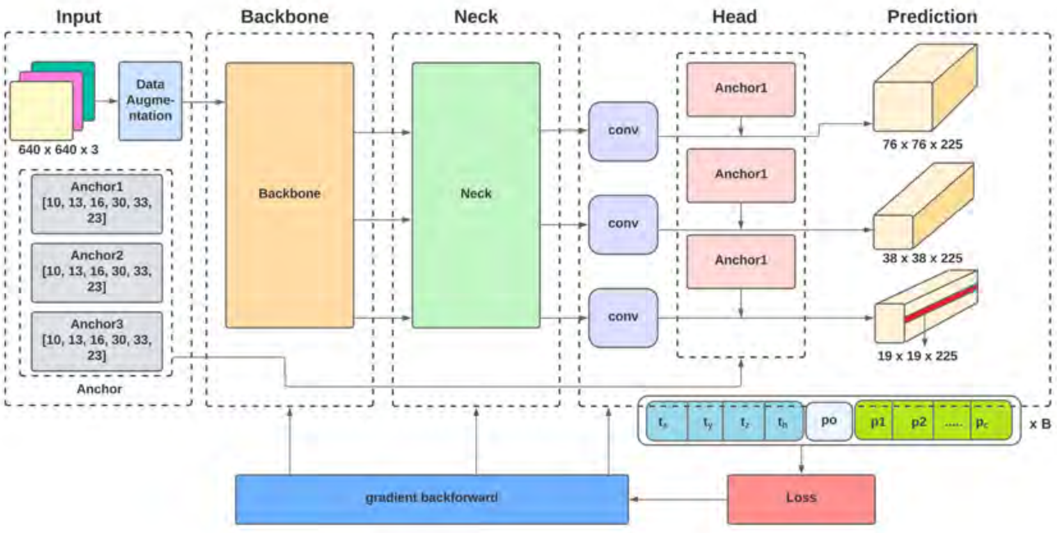


Figure 3.3: Architecture of YOLOv5 Algorithm

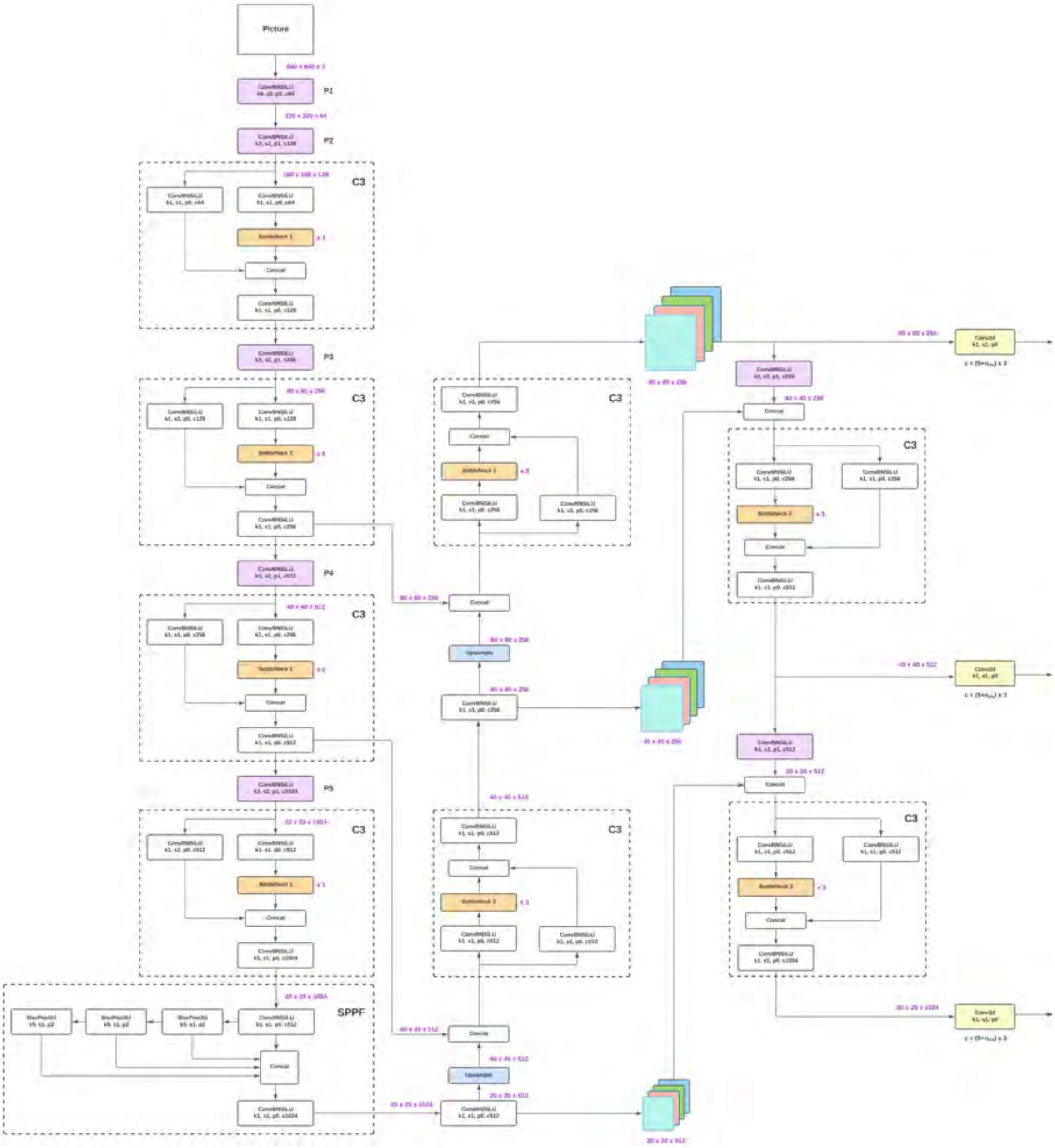


Figure 3.4: Architecture of YOLOv5 Algorithm (detailed)

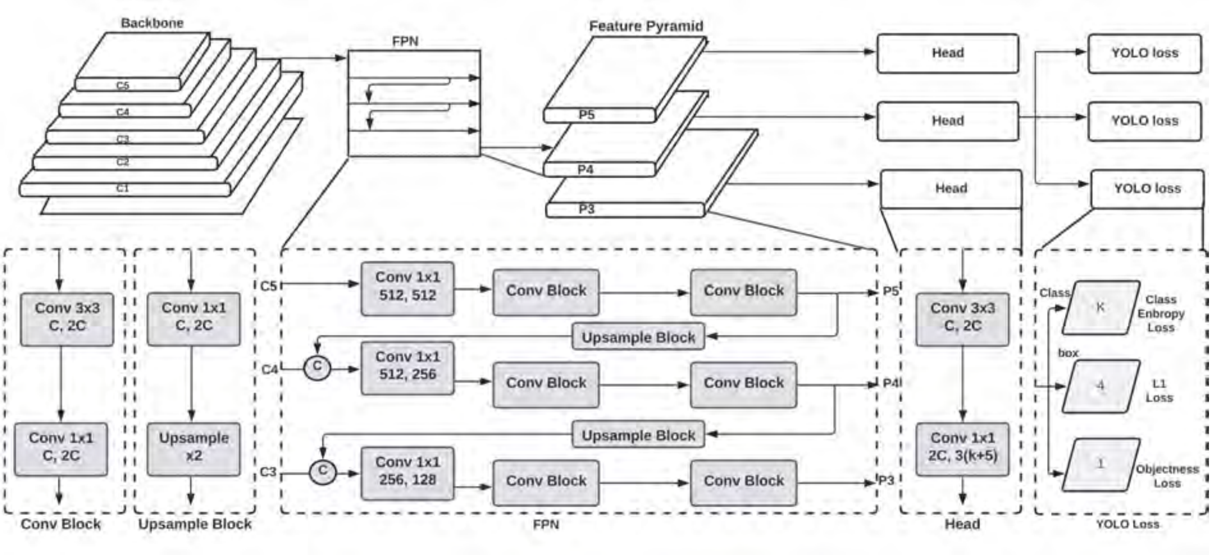


Figure 3.5: Architecture of YOLOv6(Detailed)

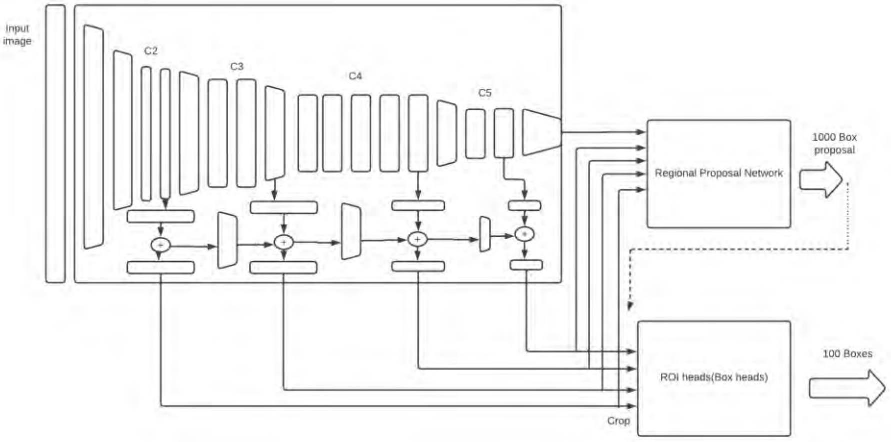


Figure 3.6: Meta architecture of Base RCNN FPN

Detectron2(Faster R-CNN with FPN)

Detectron2 is Facebook AI Research's next-generation library that provides state-of-the-art detection and segmentation algorithms. It is the successor of Detectron and maskrcnn-benchmark. It supports a number of computer vision research projects and production applications on Facebook. We tried to work with the following algorithm but eventually this slower than YOLO(v3,v4,v5,v6) [30].

Faster R-CNN

Faster R-CNN is formed using two modules. The first module is a deep, fully convolutional network that proposes regions, and the second module is the Fast R-CNN detector that uses the proposed regions. Instead of using the slow selective search algorithm, which is used in both R-CNN and Fast R-CNN, a fast neural network is used to predict the region proposals in Faster R-CNN. This makes Faster R-CNN fast enough to detect objects even in real-time. The entire system works as a single unified network.

YOLOv5 Architecture

YOLOv5 You only look once, or YOLO is composed of a single neural network and the best way to detect objects in real-time [24]. As humans look into something once and can easily detect what they are looking at. YOLO works in a similar way; it is a fast and accurate algorithm to detect objects in real-time. YOLO takes processed data as inputs, and through training, it is possible to detect objects, body parts etc., in real-time. It detects a particular thing through a bounding box where the probability of the trained object being present is shown. YOLOv5 is used in our case of nudity detection as in the case of accuracy and precision, YOLOv5 gives the most suitable solution compared to YOLOv4 and YOLOv3 [14]. The following figure shows the architecture of YOLOv5.

YOLOv6 Architecture

Outperforming YOLOv5 in detection accuracy and inference speed, a new improved version of YOLO architecture is introduced named YOLOv6. The diverse requirements for speed and accuracy was kept in consideration while making this version of YOLO. This version's structure is different from the previous versions as well. The EFFICIENTrep backbone is used in YOLOv6 which can make use of hardware computing power such as GPU. The neck in this version is called Rep-PAN Neck which is more accurate and faster. The performance improvement comes from the decoupled head meaning a layer is present between the network and the final head.

3.1.2 Proposed Model for Nudity Detection and Blurring

The initially proposed nudity detection model is intended to identify naked figures in videos. Therefore, the model should establish a process using YOLO, CNN, and other techniques to accomplish this. Using a specific scenario of the existing diagram, the model would collect data from the video clip and comprehensively review the data, generating assumptions of nodes and preparing paths of input for each video clip as packages. In this contradicting situation, a convolutional neural network (CNN) makes it possible to categorize the pictures in our frames, whereas Support Vector Machine (SVM) separates nudity and non-nudity. The Figures(3.7 and 3.8) given below illustrate a strong summary of the model design.

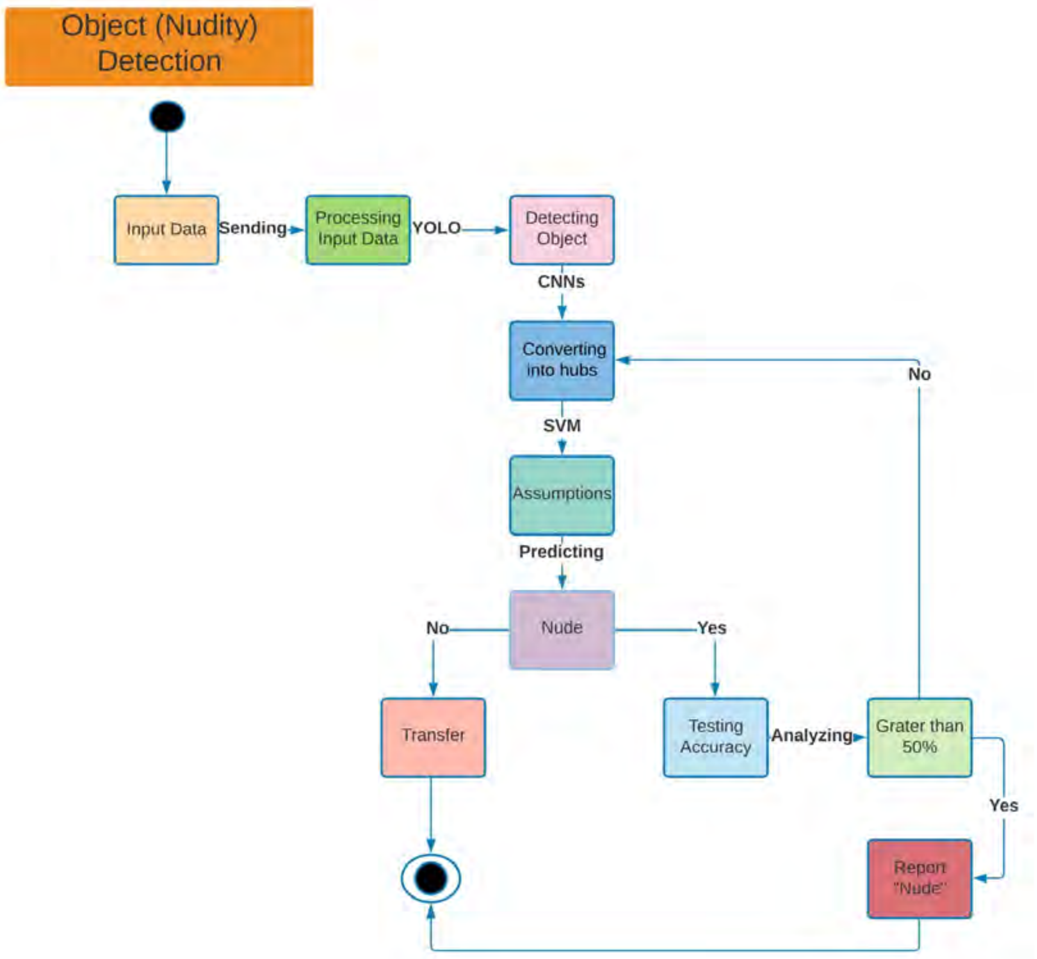


Figure 3.7: Object (Nudity) Detection - Flow Chart

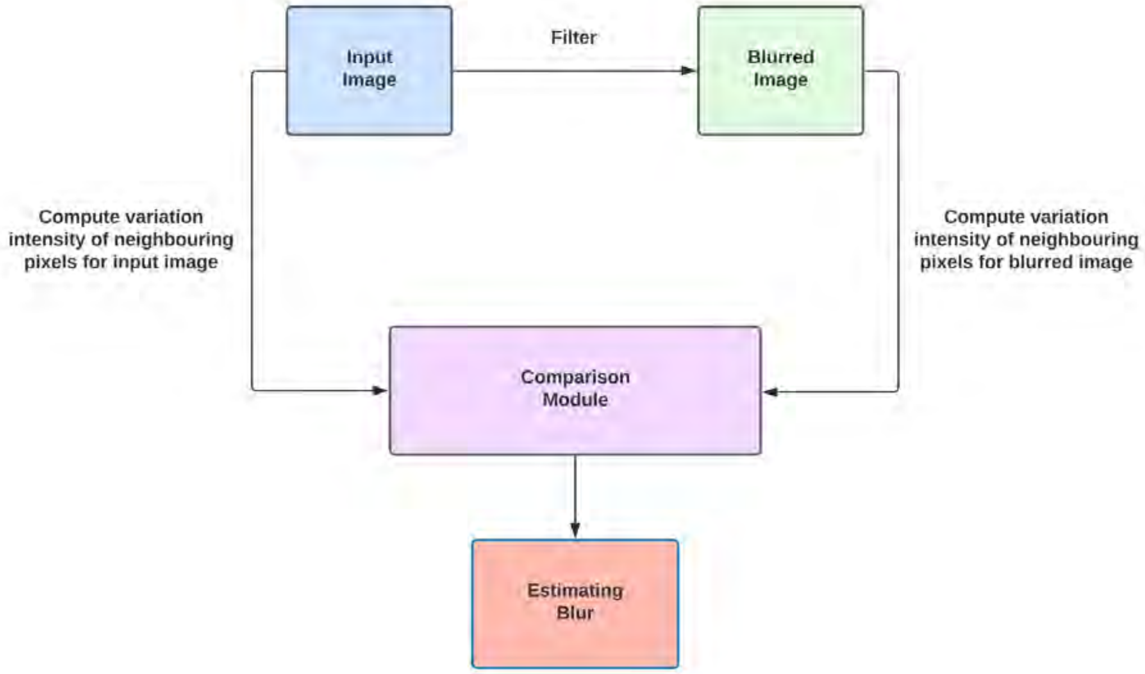

Input Image
Filter
Blurred Image
Compute variation intensity of neighbouring pixels for input image
Compute variation intensity of neighbouring pixels for blurred image
Comparison Module
Estimating Blur

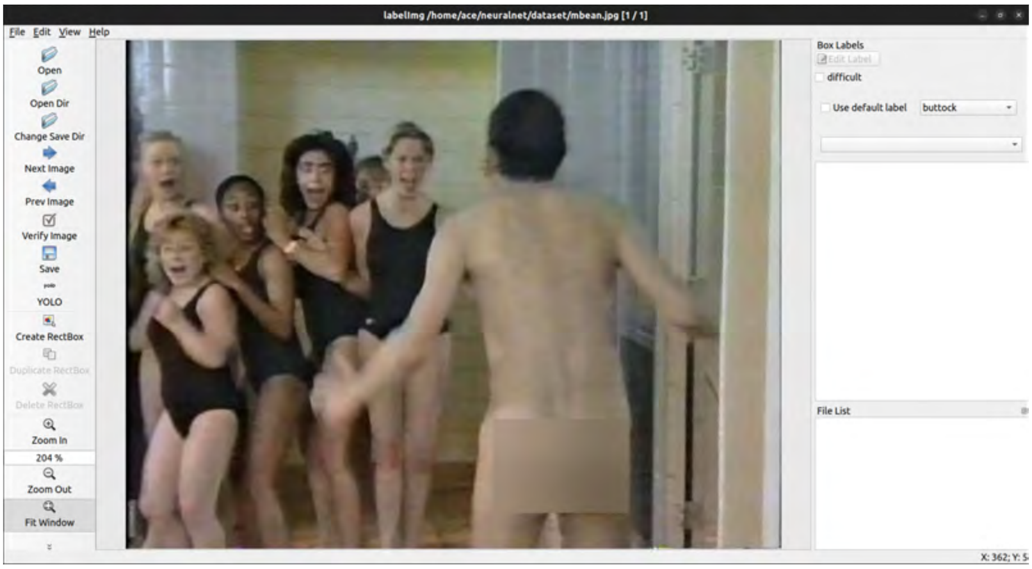
Figure 3.8: Object (Nudity) Blurring - Flow Chart

Chapter 4

Training and Results

4.1 Datasets, Labelling and Training

For training our object detection models, first of all we made our datasets accordingly with the help of taking images from various movies, series, YouTube clips, videos and some other google images. Like this way, we could collect almost 3500 images and then we labelled these accordingly (Figure 4.1). The images were resized to a shape of 640 x 640 pixels and instead of stretching the images, the aspect ratio was preserved by white padding where necessary. The images were then split among train, validation and test categories with a ratio of 70%, 20% and 10% respectively before passing them as inputs in the object detection models.



□ Use default label buttock

Figure 4.1: Labelling image

After letting the datasets to be trained, we get this bar chart of the quantity of bounded objects (Figure 4.3).

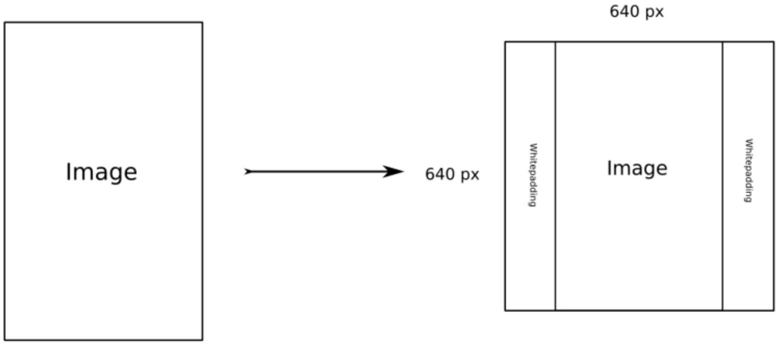


Figure 4.2: Resizing Images For Labelling

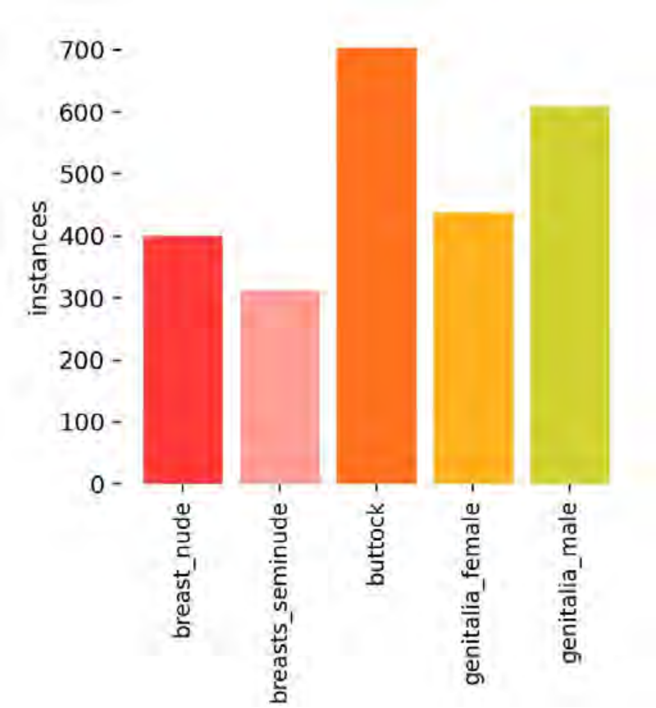


Figure 4.3: After Giving The Datasets To Train The Codes

4.2 Intial Model Testing

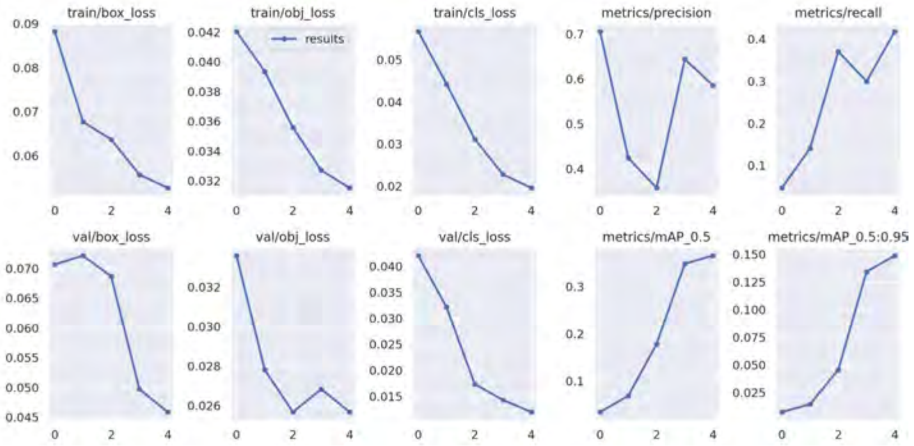


Figure 4.4: Initial Data Train Loss and Precision

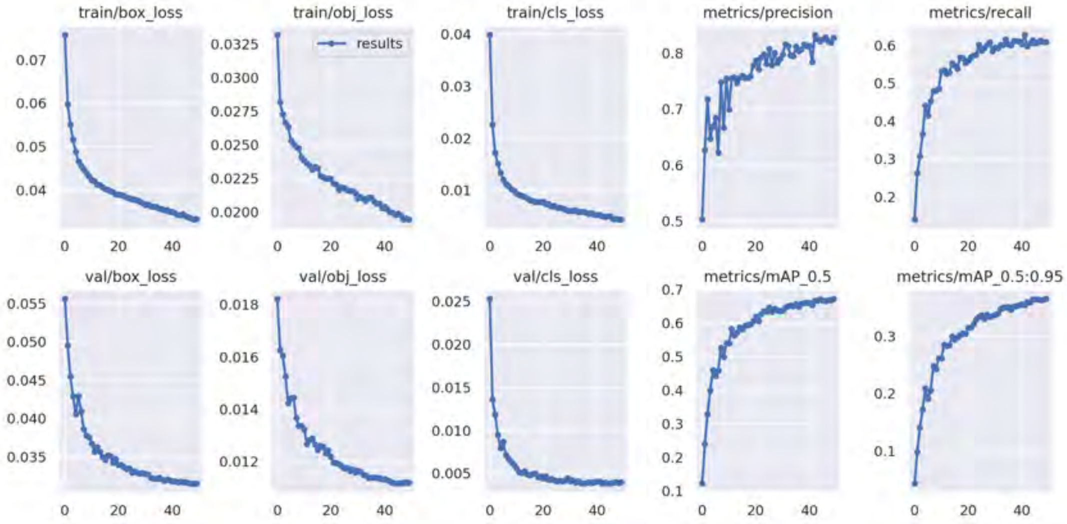


Figure 4.5: Initial Data Train Loss and Precision

And then we trained the datasets with more epochs using YOLOv5 again.

Initially we trained our datasets to detect objects with the help of models like RCNN and YOLO(v3,v4,v5). At first we took less epochs to test the model and ran the codes using YOLOv5 algorithm (Figure 4.4 and Figure 4.5). You only look once, or YOLO is composed of a single neural network and the best way to detect objects in real-time [11]. As humans look into something once and can easily detect what they are looking at. YOLO works in a similar way; it is a fast and accurate algorithm to detect objects in real-time. YOLO takes processed data as inputs, and through training, it is possible to detect objects, body parts etc., in real-time. It detects a particular thing through a bounding box where the probability of the trained object being present is shown. YOLOv5 is used in our case of nudity detection as in the case of accuracy and precision, YOLOv5 gives the most suitable solution compared to YOLOv4 and YOLOv3 [27].

4.3 Implementations and Results

4.3.1 Initial Results

Since it already has been established that YOLOv5 is faster than YOLOv4, the results of Detectron 2 (Faster R-CNN with Feature Pyramid Network) and YOLOv5 are considered [27]. After running both models for 50 epochs, the final result of mean average precision or, in short, mAP was calculated. Mean Average Precision (mAP) is a widely used unit of measurement in object detectors, instance and semantic segmentation. It takes into account both true positives (TP) and false positives (FP), and thus, it incorporates the trade-off between precision and recall. This attribute makes mAP applicable for most use cases.
The mAP is calculated based on the following formula in (Figure 4.6). After training both of the models, the final mAP values calculated at IOU threshold 0.5 were respectively 0.3823 (Detectron2) and 0.4798 (YOLOv5), which shows a clear difference between the accuracy of the two models.

$$mAP = \frac{1}{|Classes|} \sum_{c \epsilon Classes} \frac{|Truepositive_c|}{|Falsepositive_c| + |Truepositive_c|}$$

Figure 4.6: mAP Calculated based on the following formula

If any obscene scene was detected then the immediate thing to do was to blur it. However, while trying to blur obscene scenes in real time we ran into one issue and that is the blurring was not fast enough. In this case the GPU was the NVIDIA GeForce 1050ti (Mobile). In OpenCV, at first Gaussian blurring method was used. In Gaussian blur, the pixels closest to the center are given more weight than the farthest ones. Here, channel by channel averaging is done and then the average channel value is used as the pixel in the filtered image. However, this blurring method was not fast enough. After Gaussian blur, OpenCV median blur method was used. In this blurring method, the central pixel gets replaced by the median of all pixels in a kernel. Kernel will always be an odd matrix and the total elements of a kernel will be N*N. Among these N*N elements what median blur does is to choose the (N*N+1)/2^{th} elements and replaces the central pixel with the new chosen one. After the Gaussian filter the median filter is quite robust on noise. This time the blurring method worked fast however, still another blurring method was used to get an even faster result. The last one that was used, was the average blur method. Implications of these findings an image with an adjusted box filtration produces this blurring. It starts by averaging all of the pixel value in a firmware and then supplants it with the core feature.

4.3.2 Results of YOLOv5 and YOLOv6 Algorithm

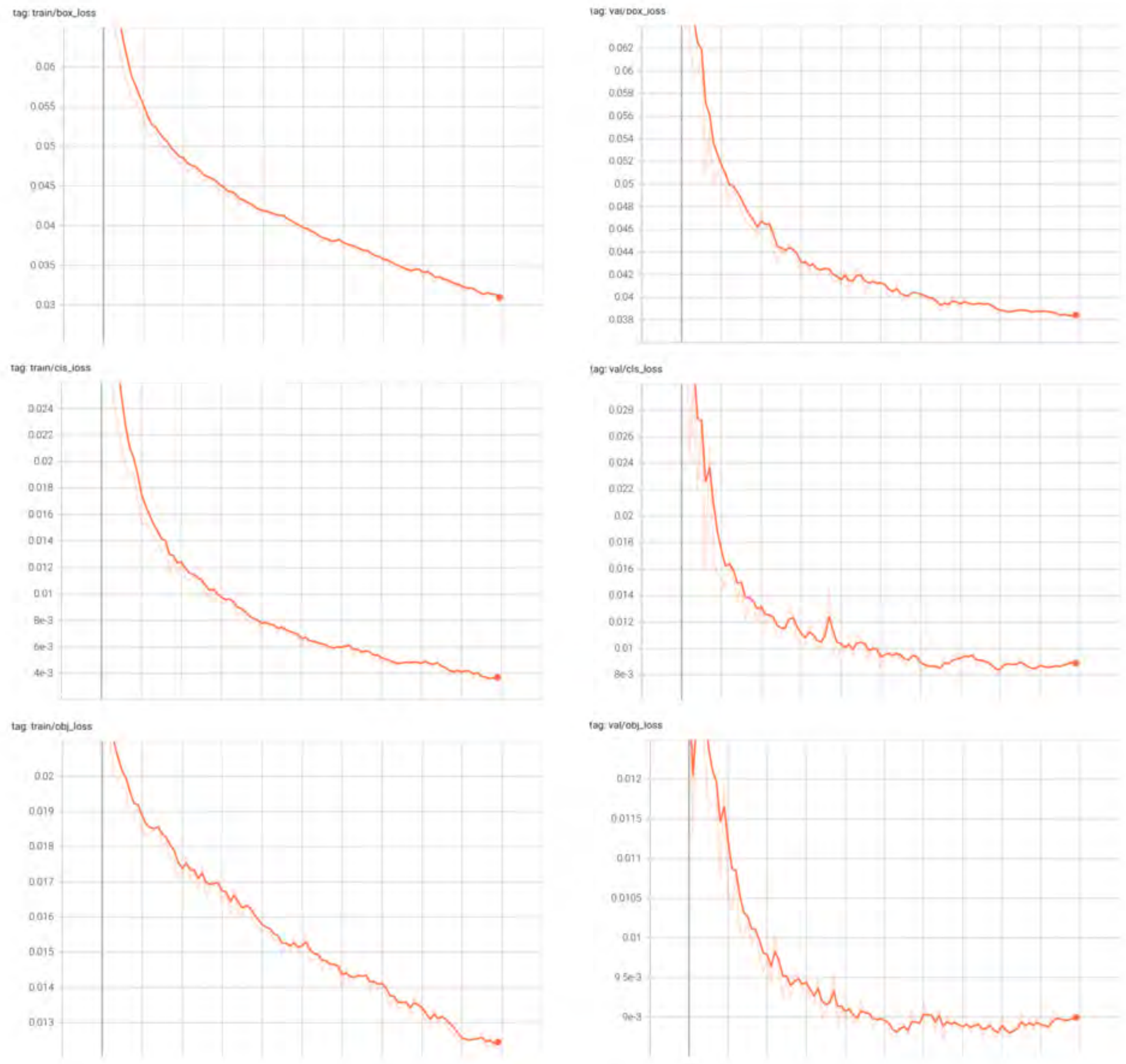


Figure 4.7: YOLOv5 Train Loss and Validation Loss

How well the algorithm can detect the centre is determined by box loss. It also determines how well it can predict bounding box cover in an object (Figure 4.7). Probability of an object existing in a region of existence is determined by object loss. Lastly, the capability of predicting classes of a given object is determined by class loss. These are the box loss, object loss and class loss that we got from our algorithms.

The mean average precision, also known as mAP, compares the given true bounding box(bbox) with the detected bounding box and returns a score. Which means if the score is higher then the accuracy level, it will be higher as well. The mAP scores, mAP@0.5 that we got from our YOLOv5 model is 0.6626 and again mAP@0.50:0.95 score is 0.338. In case of YOLOv6, the mAP@0.5 is 0.5083 and the mAP@0.50:0.95 is 0.2497 (Figure 4.8 and Figure 4.14). On the other hand, in Faster R-CNN (De-

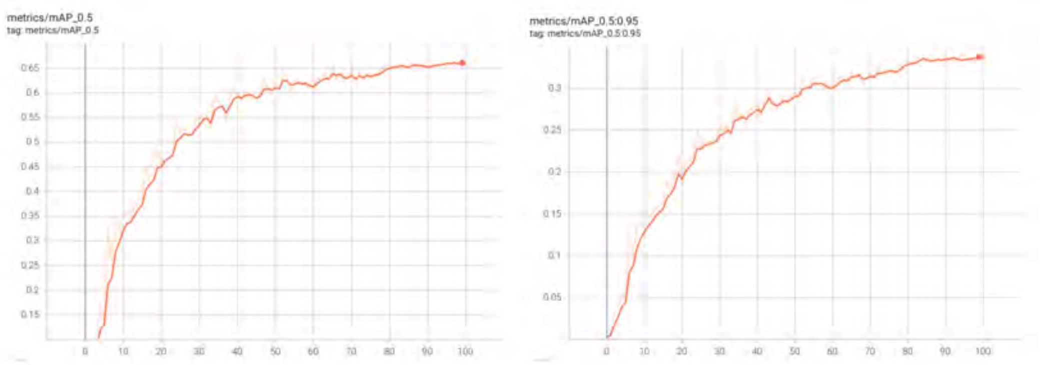


Figure 4.8: YOLOv5 mAP

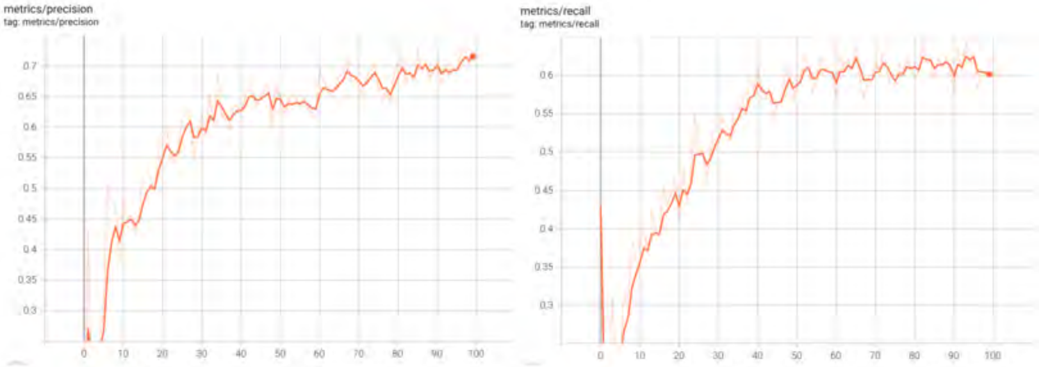


Figure 4.9: YOLOv5 Precision and Recall

tectron2), the mAP score was considerably low, which is why we were prompted not to move any further with it.

This PR curve gives us an understanding of high recall and high precision (Figure 4.9). This also relies on the low false positive and negative rates (Figure 4.10).

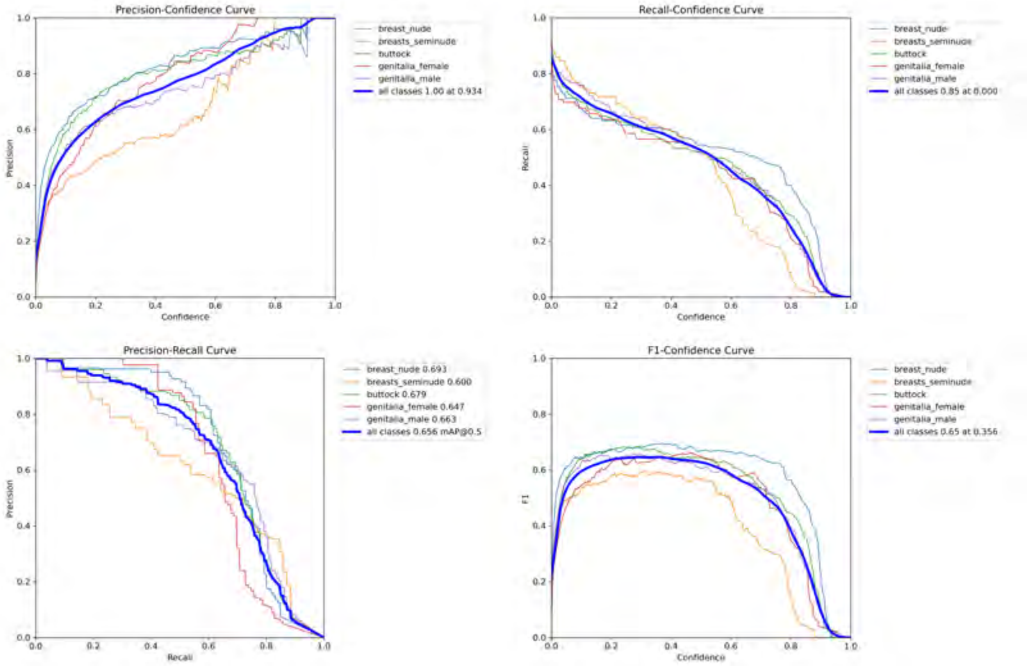


Figure 4.10: YOLOv5 P, R and PR Curve

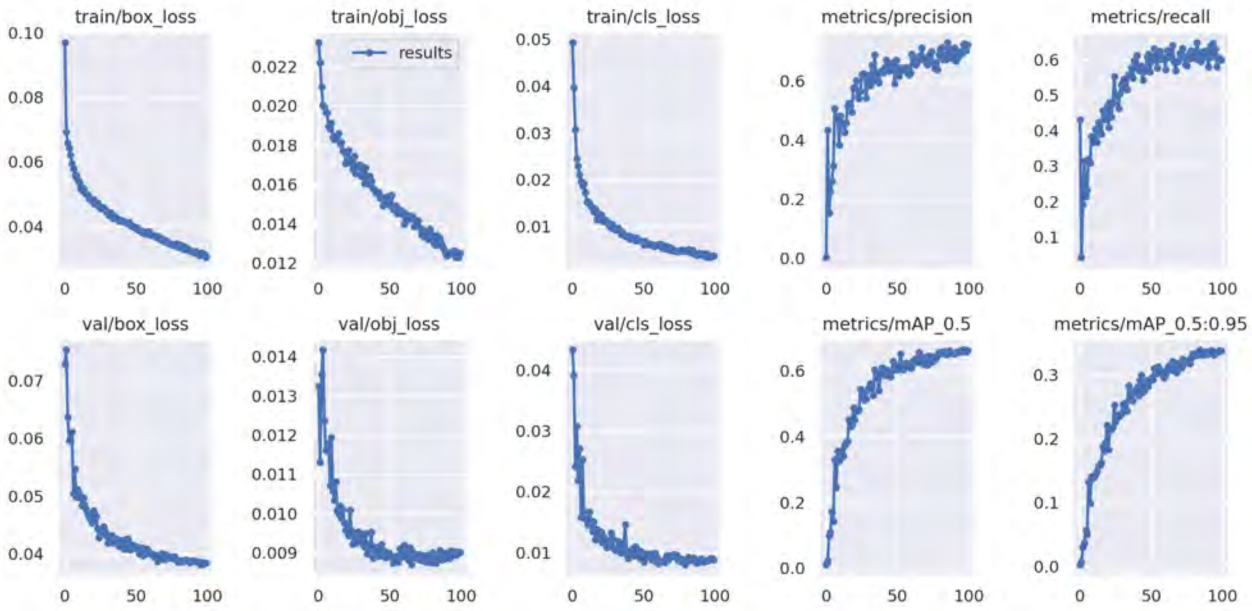


Figure 4.11: Result - YOLOv5

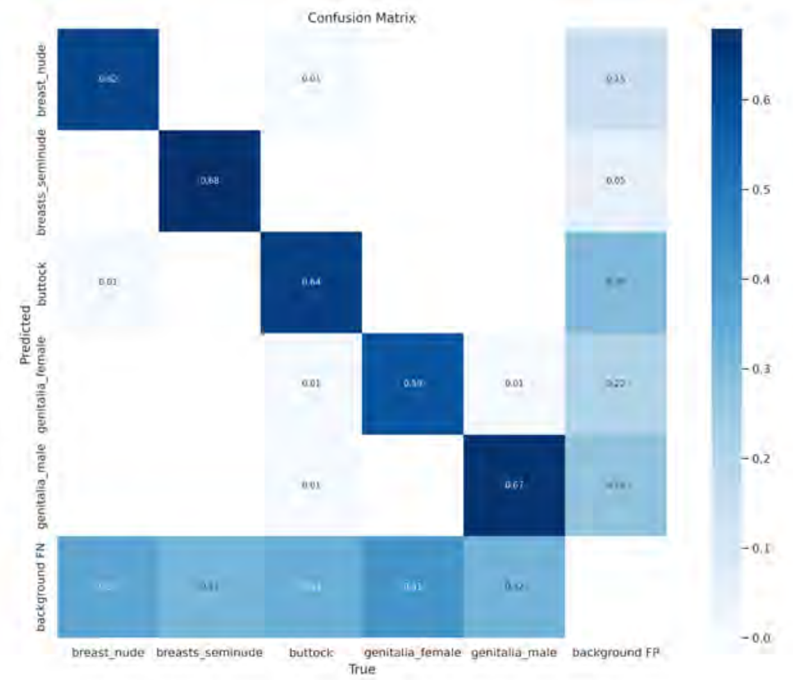


Figure 4.12: YOLOv5 Confusion Matrix

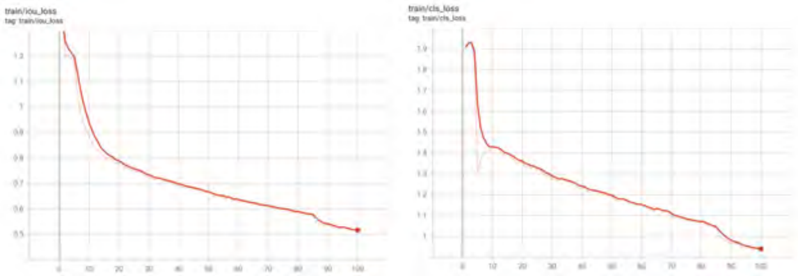


Figure 4.13: YOLOv6 - LOSS

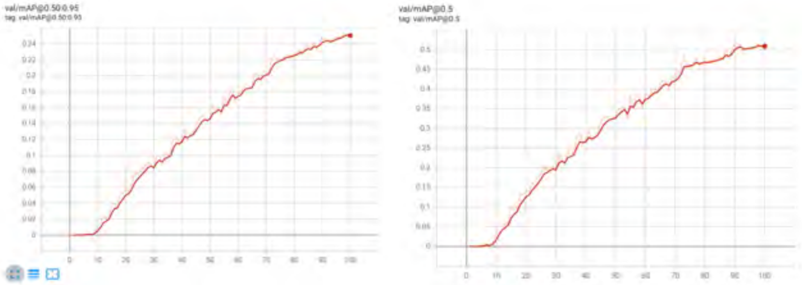


Figure 4.14: YOLOv6 - mAP

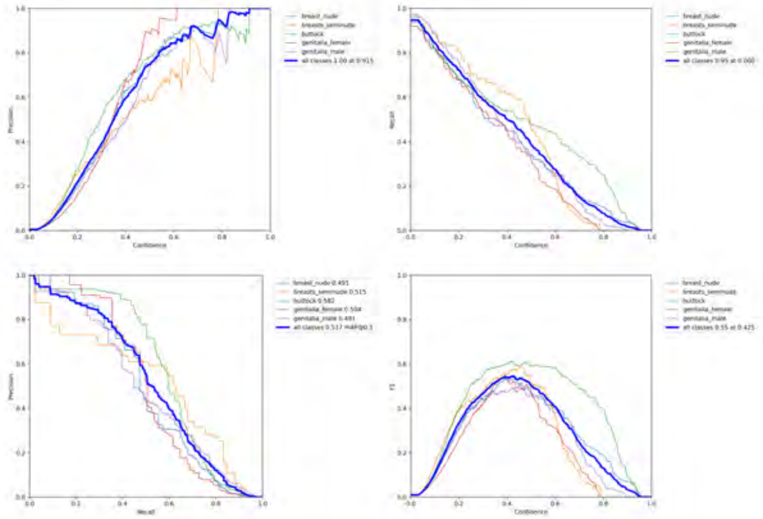


Figure 4.15: YOLOv6 - P, R, PR and F1

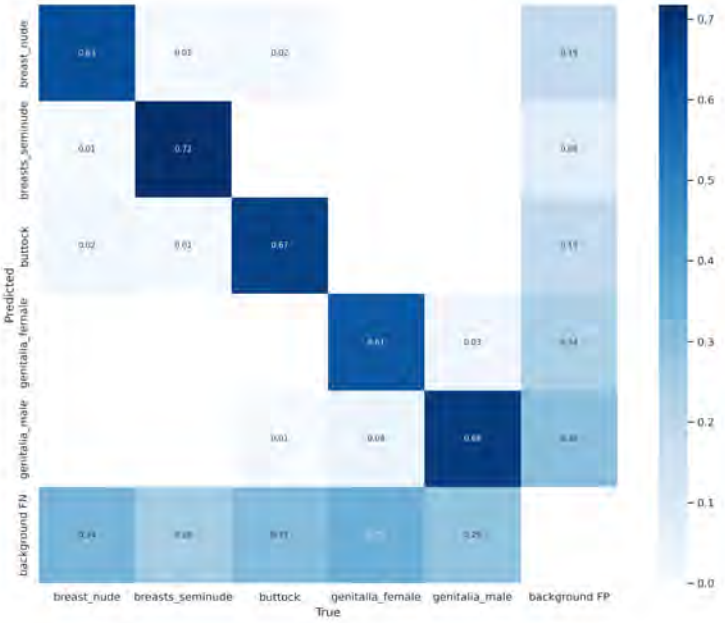


Figure 4.16: YOLOv6 - Confusion Matrix

In our training phase four possible cases may occur, these are, true positives(TP), false positives(FP), false negatives(FN) and true negatives(TN). Through these four cases it is possible to measure accuracy. Figure 4.11 shows all of trained and validation loss of YOLOv5 model training in our work.

Confusion matrix is the way to measure the errors or accuracy of predicted classes vs the actual classes. In our case the diagonal being the most accurate numbers means that the model worked well. This is the confusion matrix that we got by implementing the dataset in YOLOv5n (Figure 4.12) and YOLOv6n (Figure 4.16).

4.3.3 Blurring Results and Comparison Between YOLOv5 and YOLOv6

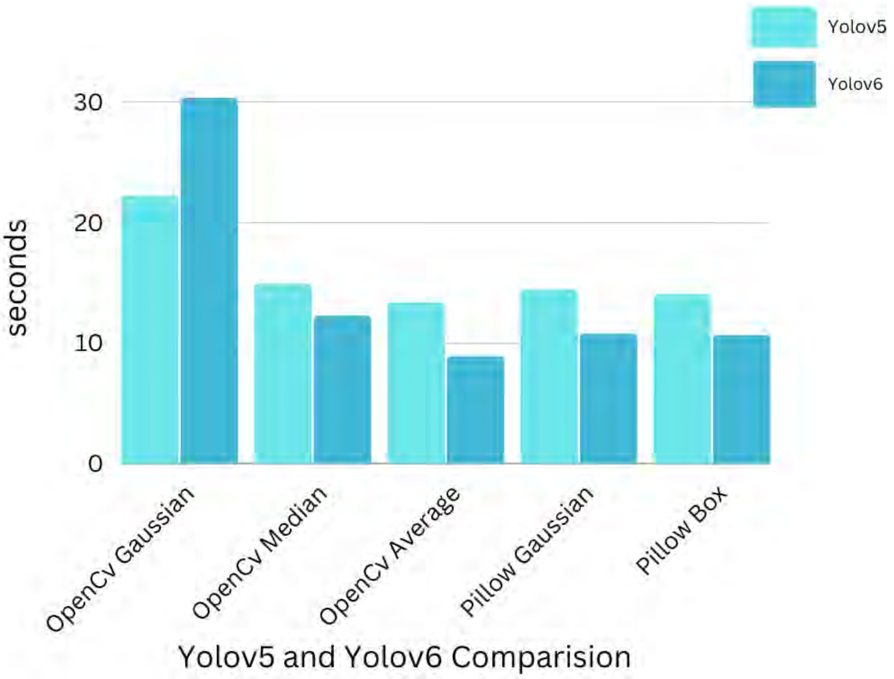


Figure 4.17: Speed Comparison Bar Chart

Combination of YOLOv6 and OpenCV average blur worked faster in comparison to both models. In the bar chart above (Figure 4.17), the lowest bar stands for the faster speed in case of blurring. Our limitation was not having access to a better GPU, due to which we were facing some timing issues in blurring speeds.

Table 4.1: Blur Timings - OpenCV/YOLOv5

OpenCV		
Median Blur	Gaussian Blur	Average Blur
14.09757137	21.69613719	12.7930088
15.43405342	21.67938781	13.20445609
15.29584622	22.4433012	13.84958696
14.82680988	23.06503415	13.21205378
14.71921921	22.1556921	13.54963064
14.87470002	22.20791049	13.32174726

Table 4.2: Blur Timings - PIL/YOLOv5

PIL	
Gaussian Blur	Box Blur
13.463434934616	14.8195593357086
14.3288419246673	13.9096252918243
14.7272689342498	13.9061765670776
15.4157016277313	14.6888945102691
14.1470916271209	12.6811256408691
14.41646781	14.00107627

Table 4.3: Blur Timings - OpenCV/YOLOv6

OpenCV		
Median Blur	Gaussian Blur	Average Blur
12.49306393	30.66199422	9.415367842
12.4664228	31.84477472	8.666805267
12.51437354	29.78105116	9.191741467
12.05743599	30.22601199	8.482309103
11.64797616	28.93356848	8.605831385
12.23585448	30.28948011	8.872411013

Table 4.4: Blur Timings - PIL/YOLOv6

PIL	
Gaussian Blur	Box Blur
10.64559317	10.69285965
10.29630065	11.28603435
10.91101742	9.702563047
11.33054686	9.906082153
10.54887247	11.5955081
10.74646611	10.63660946

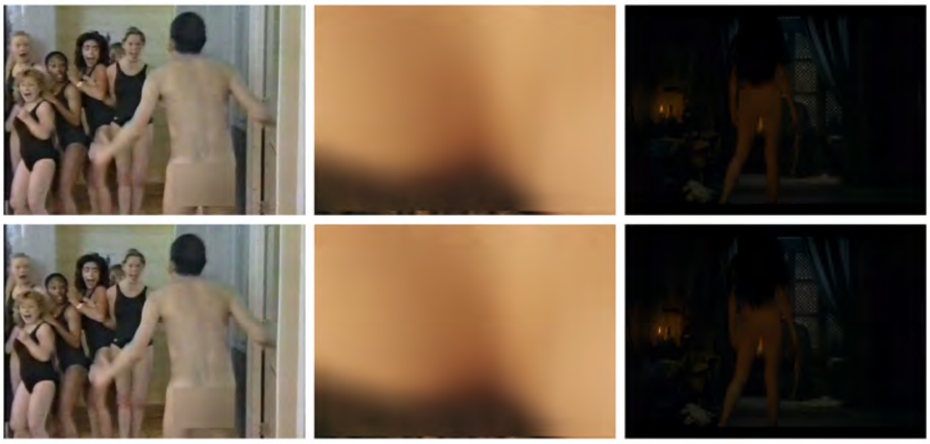


Figure 4.18: Blur Comparison Between YOLOv5 and YOLOv6 Model Training

After running each of the blurring methods 5 times with our YOLOv5n and YOLOv6n object detection models on a sample video, we recorded the timings and averaged them. The unit that stands for our blurring times in the given tables (Table 4.1, 4.2, 4.3 and 4.4) of values is ”seconds/sec”. Table 4.1 and 4.2 reflect the times for YOLOv5n and Table 4.3 and 4.4 reflect the same for YOLOv6n. Although the differences in times are just in seconds or milliseconds, even the slightest difference is important if we attempt to reach real-time. Now, from the tables, it is evident that OpenCV Average Blur works the fastest among all of the blurring methods while OpenCV Gaussian Blur is the slowest.

Chapter 5

Conclusion and Future Work

There are countless videos on the Internet, and to know whether a video is safe to watch or not, we need to determine if the video in question contains obscene content. On top of that, real-time detection and blurring will be beneficial to safeguard the users from obscenities displayed during live video streams and video calls. Therefore, we have made improvements on the existing research that has been previously done on detecting obscene content from video clips and develop an algorithm using YOLOv6 and OpenCV to achieve real-time detection and blurring of obscene content without giving up on the accuracy. With a more specific approach, this can be utilized to censor unnecessary vulgarities in mainstream media so that more people can experience them without encountering unwanted obscene content. This can make the Internet a safer environment for everyone, especially children. With a larger and more specific dataset, we plan to refine and improve this research even further. Our goal from this work is to develop our research skill to a greater extent and to put more contribution to the related field.

Bibliography

[1] R. Ap-Apid, "An algorithm for nudity detection," in *5th Philippine Computing Science Congress*, 2005, pp. 201–205.

[2] H. A. Rowley, Y. Jing, and S. Baluja, "Large scale image-based adult-content filtering," 2006.

[3] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin color model," *Pattern recognition*, vol. 40, no. 8, pp. 2261–2270, 2007.

[4] M. B. Andersen, "Ka 322 seminarhold 2 i klinisk psykologi v. karin riber," 2010.

[5] A. Meek, *Trauma and media: Theories, histories, and images.* Routledge, 2011.

[6] B. Su, S. Lu, and C. L. Tan, "Blurred image region detection and classification," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 1397–1400.

[7] X. Xing, Y.-L. Liang, H. Cheng, *et al.*, "Safevchat: Detecting obscene content and misbehaving users in online video chat services," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 685–694.

[8] A. Behrad, M. Salehpour, M. Ghaderian, M. Saiedi, and M. N. Barati, "Content-based obscene video recognition by combining 3d spatiotemporal and motion-based features," *EURASIP Journal on Image and Video Processing*, vol. 2012, no. 1, pp. 1–17, 2012.

[9] C. Ross, "Overexposed and under-prepared: The effects of early exposure to sexual content," *Psychology Today*, 2012.

[10] S. Karavarsamis, N. Ntarmos, K. Blekas, and I. Pitas, "Detecting pornographic images by localizing skin rois," *International Journal of Digital Crime and Forensics (IJDCF)*, vol. 5, no. 1, pp. 39–53, 2013.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[12] L. M. Ward, "Media and sexualization: State of empirical research, 1995–2015," *The Journal of Sex Research*, vol. 53, no. 4-5, pp. 560–577, 2016.

[13] R. L. Collins, V. C. Strasburger, J. D. Brown, E. Donnerstein, A. Lenhart, and L. M. Ward, "Sexual media and childhood well-being and health," *Pediatrics*, vol. 140, no. Supplement_2, S162–S166, 2017.

[14] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[15] K. N. Tofa, F. Ahmed, A. Shakil, *et al.*, "Inappropriate scene detection in a video stream," Ph.D. dissertation, BRAC University, 2017.

[16] M. D. More, D. M. Souza, J. Wehrmann, and R. C. Barros, "Seamless nudity censorship: An image-to-image translation approach based on adversarial training," in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2018, pp. 1–8.

[17] F. S. Al-Mukhtar, "Tracking and blurring the face in a video file," *Al-Nahrain Journal of Science*, no. 1, pp. 202–207, 2018.

[18] J. Wehrmann, G. S. Simões, R. C. Barros, and V. F. Cavalcante, "Adult content detection in videos with convolutional and recurrent neural networks," *Neurocomputing*, vol. 272, pp. 432–438, 2018.

[19] S. Geethapriya, N. Duraimurugan, and S. Chokkalingam, "Real-time object detection with yolo," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 3S, 2019.

[20] R. S. Islam, R. Siddiqui, and D. Roy, "Blurring of inappropriate scenes in a video using image processing," Ph.D. dissertation, Brac University, 2019.

[21] N. Melad, "Detecting and blurring potentially sensitive personal information containers in images using faster r-cnn object detection model with tensorflow and opencv," 2019.

[22] E.-M. Pulfer, "Different approaches to blurring digital images and their effect on facial detection," 2019.

[23] K. Yuan, D. Tang, X. Liao, *et al.*, "Stealthy porn: Understanding real-world adversarial images for illicit online promotion," in *2019 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2019, pp. 952–966.

[24] N. AlDahoul, H. Abdul Karim, M. H. Lye Abdullah, *et al.*, "Transfer detection of yolo to focus cnn's attention on nude regions for adult content detection," *Symmetry*, vol. 13, no. 1, p. 26, 2020.

[25] K. Kavitha, B. J. Sikandar, *et al.*, "Digital parenting: Issues, challenges and nursing implications," *Journal of Pediatric Surgical Nursing*, vol. 10, no. 3, pp. 100–104, 2021.

[26] A. Limballe, R. Kulpa, and S. Bennett, "Using blur for perceptual investigation and training in sport? a clear picture of the evidence and implications for future research," *Frontiers in Psychology*, vol. 12, 2021.

[27] M. Rocha, M. Claro, L. Neto, K. Aires, V. Machado, and R. Veras, "Malaria parasites detection and identification using object detectors based on deep neural networks: A wide comparative analysis," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–18, 2022.

[28] A. Turner, "How many people have smartphones worldwide (mar 2022)," *BankMyCell*, 2022.

[29] S. Awortwe, "Sabina blog no.,"

[30] *GitHub - facebookresearch/detectron2: Detectron2 is a platform for object detection, segmentation and other visual recognition tasks. — github.com*, https://github.com/facebookresearch/detectron2, [Accessed 18-Sep-2022].

[31] Z. Liu, M. Hao, and Y. Hu, "Visual anonymity: Automated human face blurring for privacy-preserving digital videos,"

[32] *U.S. video consumption by device 2023 — Statista — statista.com*, https://www.statista.com/statistics/420791/daily-video-content-consumption-usa-device/, [Accessed 18-Sep-2022].