Original Article

Classification of Actors in an Animated Video using a Novel Yolo Framework in Comparison with SVM Algorithm

Srihari V[1], Magesh kumar S[2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India, Pincode: 602105.
[2]Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India, Pincode: 602105.

Abstract

Aim: The major goal of this study is Classification of actors in an animated video using the novel YOLO (You Only Look Once) framework in comparison with the SVM (Support Vector Machine) algorithm. **Materials and Methods:** Sample groups that are considered in the project can be classified into two, one for YOLO and other for SVM, which are tested using 0.80 for G-power to determine the sample size and for t-test analysis**.Results and Discussion:-** The analysis of results show that the You Only Look Once has a high accuracy (87.45%) in comparison with the Support Vector Machine (84.74%).The statistical significance difference (two-tailed) is 0.001 ($p<0.05$).**Conclusion:** Novel YOLO framework seems to be better in classification of actors in an animated video over the SVM algorithm.

Keywords: Deep Learning, Face detection, Image Classification, Novel You Only Look Once (YOLO) framework, Supervised Learning, Support vector machine.

DOI: 10.47750/pnr.2022.13.S04.187

INTRODUCTION

Face recognition and identity systems have emerged as one of the most full-size disciplines of pc imaginative and prescient in recent years. As a preliminary level, the man or woman popularity machine includes face detection. Face detection era may be utilized in a range of industries, such as safety, biometrics, law enforcement, amusement, and personal safety, to permit actual-time monitoring and surveillance of persons(Li and Jain 2005)(Yangon, Gui, and Hu 2010). A person's Screen Time is the whole time of the film that contains the individual. It is crucial in evaluating the actor's salary and analyzing the gender bias in films(E.k., Varnima, and Ramachandran 2020).we employed You Only Look Once to apply picture classification concepts. The construction of the deep learning model to recognise and characterize articles based on recently cared for photographs is tied to grouping Both supervised and unsupervised learning approaches can be used to classify images.Supervised learning creates an array with values from 0 to 1(Brownlee 2019). These values are obtained by normalizing the pixel value of the image in the range 0-255. For accurate image classification, the model needs to understand the background, camera angle, focus, and lighting of the image. Camera light. For example, sharpness needs to be extracted by the model. Color, texture, shape histogram, number of pixels in the image created (Nixon and Aguado 2019).

There are around 8 IEEE papers and 25 google scholar papers have been published over the past 5 years. The most cited article is "Calculating screen time of Characters in a video using You Only Look Once algorithm" and "Face detection in real time video using you only look once algorithm". In the existing research the major problem is identification of the videos and its characters with good accuracy and also the total calculating time accuracy is not good. Our project's primary goal is to Calculate Screen Time of an Actor in Video Using You only look once algorithm over Support Vector Machine.

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al.

2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). Image categorization is used for many different things, including image processing, picture improvement, image restoration, image understanding, image recognition, and object recognition preparation. Face identification and image categorization are critical in many aspects of daily life. It is employed in many areas, including science and technology (Sonka, Hlavac, and Boyle 2012). Supervised Learning technology may be used to provide real-time monitoring and tracking of individuals in a range of businesses, including security, biometrics, law enforcement, escapism, and physical safety. Image classification can be used to identify anomalous border activity in military surveillance and automatically assess the lives of soldiers(Le and Li, n.d.). It can be used to understand different parts of the globe where human research is difficult. There are many ways to use image classification to calculate an actor's screen time, but this study used YOLO to get very accurate results(Zuo, Wang, and Fu 2016) with convolutional neural network to find the actor's screen time (Cheng 2020)(Lu, Zhang, and Xie 2020)**.**

Materials and Methods

This study was conducted at SIMATS' Saveetha School of Engineering's Image Processing Lab. Essentially, two types of classifiers are employed: You Only Look Once and Support Vector Machine, which are used for Classification of Actors in an Animated Video. Group 1 is the You Only Look Once with a sample size of 24, and Group 2 is the Support Vector Machine with a sample size of 24, and these are compared for higher accuracy and precision score values to determine the best method. The minimum power analysis for G-power is 0.8, with a threshold of 0.05 percent and a maximum allowed error of 0.5.
Testing setup for this proposed system used a visual studio code. It is a software which is used for Classification of Actors in an Animated Video with the Novel YOLO Framework model and SVM. Hardware configuration for this proposed system is an Amd Ryzen 5 5th gen processor and requires 4GB RAM and 256GB SSD. The configuration of the system is Windows 10 operating system and python programming language 3.8.3.
The sample size has been computed, and it has been determined that 24 samples/group, for a total of 48 samples, with a standard deviation of YOLO = .63324 and SVM = 1.87293, which are tested using 0.80 for G-power to determine the sample size and for t-test analysis

Testing procedure for Classification of Actors in an Animated Video with Novel YOLO Framework
Step 1: **Preprocessing**
The preprocessing step is used to adjust your image to the model's specifications.Images with values ranging from 0 to 1 are used in some models.
Step 2: Feature Extraction
To extract the features from a video, feature extraction is employed.
Step 3: Training
After the feature extraction step, the training process is done. Training step involves Screen Dataset as input to the classifier to generate epochs.
Step 4: Test Classifier
In the test classifier is possible to import different and more datasets to test the accuracy of the classifier.
Step 5: Testing
Testing process is used to detect the Actors in an Animated Video.If the actor is present or not. In this step the classifier shows the result of how much time the actor is present.

You Only Look Once
The You Only Look Once algorithm is called YOLO in short.This algorithm can detect as well as recognise the object from the images which is an end to end neural network which makes the class probabilities and bounding boxes-predictions at once.

Algorithm:-Novel YOLO Framework
Input:-d:dataset,frames:video,X and y:variable

```
Get video_Capture(frames)                          //divide video into frames
        for frames i  to n
            n1 = plt.imread(" + i)
            X.append(n1)                           //Storing each image in X
         X=np.array(X)                             //converting list into array
         y = data.Class
/* Initialize the X array from each (224, 224) members of Δ(r, c)
for i <- 0 to 245 do
     for j <- 0 to 245 do
          (startR,startC) <-(i*224,j*224);
```

```
        (endR,endC)<-(startR+32,startC+32);
        if Any of Δ ((startR, .., endR), (startC, .., endC), (0, .., 2)) == 1 then
           X(i,j) ← '1';
        elif
           X(i,j) ← '2';
        else
           X(i,j) ← '0';
         end
     end
end
X_train,X_test,y_train,y_test<-split features set and labels into train subset and test subset
history = model.fit(X_train, y_train, epochs=50)
V<-YOLO(X_train,y_train)
score_accuracy<-evaluate(i,y_test,V)
return score_accuracy.
```

Support Vector Machine

Novel Support Vector Machine is a supervised learning machine learning method that is commonly used in classification and regression issues. This Novel Support Vector Machine technique is employed for classification in this case. This algorithm will effectively work for classification problems by dividing the groups accordingly.

Algorithm:-Support vector machine

```
Input:-d:dataset,frames:video,X and y and svc and Z:variable,
 Get video_Capture(frames)                                  //divide video into frames
     for frames i  to n
         n1 = plt.imread(" + i)
         X.append(n1)                                          //Storing each image in X
      X=np.array(X)                                          //converting list into array
      y = data.Class
  svc = svm.SVC(kernel='linear',C=1).fit(X, y)               //we create an instance of svm
   x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
   y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
   h = (x_max / x_min)/100
   X_train, X_test = np.meshgrid(np.arange(x_min, x_max, h),np.arange(y_min, y_max, h))
   Z = svc.predict(np.c[X_train.ravel(), X_test.ravel()])
   Z = Z.reshape(xx.shape)
   X_train,X_test,y_train,y_test,Z<-split features set and labels into train subset and test subset
   history = model.fit(X_train, y_train, epochs=50)
   V<-MTCNN(X_train,y_train)
         score_accuracy<-evaluate(i,y_test,V)
   return score_accuracy.
```

The dataset started to start with separated into two parts: schooling and validation units. The set of rules is then examined at the schooling and validation statistics sets. The training and validation sets are modified five times depending on the dimensions of the statistics collection. Table 1 compares the accuracy of YOLO and SVM across 5 iterations. The various parameters for the analysis are can be calculated as follows:

Equation (1) - Accuracy : It counts the number of cases successfully categorized.

$$\text{Accuracy} = \frac{True\ Positive+True\ Negative}{True\ Positive+True\ Negative+False\ Positive+False\ Negative} \quad (1)$$

The training and test sets are changed five times depending on the size of the testing dataset. Table 2 shows the time taken of characters present in the video for 5 iterations. The various parameters for the analysis can be calculated as follows:

Equation (2) - Screen Time : It shows the number of characters that were correctly presented.

$$\text{Screen Time} = \frac{Number\ of\ images\ predicted}{No\ of\ frames\ per\ sec} \quad (2)$$

Statistical Analysis
SPSS Statistics is a suite of statistical software developed by IBM .The independent sample T-test calculation for analyzing equal variable,standard error,and leven's test are evaluated.Attributes like 0 for Tom,1 for Jerry,and 2 for others are the independent variables and Image_ID,Detection and class are dependent variables. Independent sample T-test has been carried out for evaluating the accuracy

Results

On 5 Tom and Jerry videos, the proposed methodology was tested. Table 1 displays the classification-count method findings for 5 videos. The use of the You Only Look Once Algorithm with the Support Vector Machine. The usage of Novel YOLO Framework Algorithm to deploy has shown to be successful. Accuracy in calculating the amount of time spent in front of the screen was assured. characters. The screen time of characters can be used in a variety of ways, including assessing actor pay and analysing the Discrimination against Tom and Jerry based on their time. It is inferred that the mean time for T-test is far lesser than the comparison time. Moreover, the mean accuracy value of the base YOLO is around 87.45 which seems to be superior to the SVM around 84.74. The group statistics results are presented in Table 3 with mean and standard deviation. In Table 4, it was observed that the Levens test for equality of variance and its significance for YOLO is 16.968 and .001 respectively and standard error difference and confidence interval are lower than SVM. Mean accuracy and mean loss graph is depicted in Fig. 1. YOLO seems to appear better than SVM as depicted in Fig 2.

Discussion

Throughout this study, it was discovered that the Novel YOLO Framework technique appears to be more accurate than the usual way of the SVM algorithm for determining screen time.
There are similar papers on Calculating screen time and classify the actor present in the video using deep learning algorithms. In this research paper the Traditional image processing methods are difficult to achieve in the near future due to environmental complexity such as sunlight reflections, and large differences in short-range and long-range target scales(Rajinikanth et al. 2021). In this research paper proposes a YOLOv2 algorithm that uses position and video recording to identify objects in an image. This task's main goal is to detect objects in an image. That is live detection using a webcam or video recorder. Because the GPU version is so quick, you may utilise the anchor box to complete the function precisely (Raskar and Shah 2021). The goal of this research article is to extract particular vehicle-type information from images or videos that feature cars. To address existing vehicle recognition issues such as absence of vehicle type detection, poor recognition accuracy, and sluggish speed, new vehicle recognition has been designed. A multi-layer feature fusion approach was used to increase the network's feature extraction capabilities, removing the recurrent convolutional layer in elevated networks(Zhang et al. 2018).In this research paper to recognize an object from the specified image. The proposed method for recognizing objects relates to feature vector reduction by kernel principal component analysis and recognition using a support vector machine classifier. Also in this article, the feature extraction method extracts features from the global descriptor of the image. The image feature extraction process extracts global features and forms them as feature vectors. Feature vectors are generated for the entire training image and dimensionality reduction is performed using KPCA(Yu, Chang, and Tsai 2021).
According to the data, the YOLO algorithm outperforms SVM in terms of outcomes and performance. As a result of the study's findings, both experimental and statistical analysis produce clarity in performance but it has some limitations. The limitation of the study constraints are scaled when face sizes are different and classifiers are run on each image. In such scenarios, the accuracy is not particularly high and can fail in unexpected ways. Another drawback of the method presented is that it requires the use of a GPU graphics card because the calculations are very complicated. The proposed strategy has only been tested on small films. Therefore, longer videos should be used in the future to evaluate the scalability and overall performance of the proposed methodology.

CONCLUSION
In this research the YOLO algorithm seems to appear with a better accuracy percentage (87.45%) than SVM (84.74%) improving facial recognition and calculating the screen time of an actor.

DECLARATION
Conflict of Interests
No conflict of interest in this manuscript.

Author Contribution
Author SH was involved in data collection, data analysis, and manuscript writing. Author SMK was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgments
The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for offering the infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. VBLP Tech Solution, Hyderabad.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Brownlee, Jason. 2019. *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*. Machine Learning Mastery.
2. Cheng, Richeng. 2020. "A Survey: Comparison between Convolutional Neural Network and YOLO in Image Identification." *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1453/1/012139.
3. E.k., Varnima, E. K. Varnima, and C. Ramachandran. 2020. "Real-Time Gender Identification from Face Images Using You Only Look Once (yolo)." *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. https://doi.org/10.1109/icoei48184.2020.9142989.
4. Le, Hung-Son, and Haibo Li. n.d. "Can the Surveillance System Run Pose Variant Face Recognition in Real Time?" *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. https://doi.org/10.1109/vspets.2005.1570917.
5. Li, Stan Z., and Anil K. Jain. 2005. *Handbook of Face Recognition*. Springer Science & Business Media.
6. Lu, Yonghui, Langwen Zhang, and Wei Xie. 2020. "YOLO-Compact: An Efficient YOLO Network for Single Category Real-Time Object Detection." *2020 Chinese Control And Decision Conference (CCDC)*. https://doi.org/10.1109/ccdc49329.2020.9164580.
7. Nixon, Mark, and Alberto Aguado. 2019. *Feature Extraction and Image Processing for Computer Vision*. Academic Press.
8. Rajinikanth, Venkatesan, E. Priya, Hong Lin, and Fuhua Lin. 2021. "Medical Image Processing with Hybrid Image Processing Method." *Hybrid Image Processing Methods for Medical Image Examination*. https://doi.org/10.1201/9781003082224-5.
9. Raskar, Punam Sunil, and Sanjeevani Kiran Shah. 2021. "REAL TIME OBJECT-BASED VIDEO FORGERY DETECTION USING YOLO (V2)." *Forensic Science International*. https://doi.org/10.1016/j.forsciint.2021.110979.
10. Sonka, Milan, Vaclav Hlavac, and Roger Boyle. 2014. *Image Processing, Analysis and Machine Vision*. Springer.
11. Yangon, Qiufen, Weihua Gui, and Huosheng Hu. 2010. "A Novel Real-Time Face Tracking Algorithm for Detection of Driver Fatigue." *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*. https://doi.org/10.1109/iitsi.2010.180.
12. Yu, Shun-Hsin, Jen-Shuo Chang, and Chia-Hung Dylan Tsai. 2021. "Grasp to See-Object Classification Using Flexion Glove with Support Vector Machine." *Sensors* 21 (4). https://doi.org/10.3390/s21041461.
13. Zhang, Zhongbao, Hongyuan Wang, Ji Zhang, and Wei Yang. 2018. "A Vehicle Real-Time Detection Algorithm Based on YOLOv2 Framework." *Real-Time Image and Video Processing 2018*. https://doi.org/10.1117/12.2309844.
14. Zuo, Lingxuan, Hanli Wang, and Jie Fu. 2016. "Screen Content Image Quality Assessment via Convolutional Neural Network." *2016 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/icip.2016.7532725

TABLES AND FIGURES

Table 1. Accuracy achieved during evaluation of Screen time of an actor using test and mapping dataset with YOLO algorithm and Comparison of SVM algorithm for different iterations.

ITERATIONS	ACCURACY(%)	
	YOLO	SVM
1	88.43	86.87
2	88.41	86.54
3	87.67	86.65
4	87.54	86.98

5	87.43	84.34
6	87.32	84.54
7	87.28	83.53
8	87.12	83.23
9	86.76	82.53
10	86.43	82.23

Table 2. Screen time of Tom and Jerry calculated using the Classification count method.
Sample data sets with 283 images depicting the screen time of various videos. It consists of attributes like video no, Duration of videos, Screen time of tom, Screen time of Jerry.

Video No.	Duration of Video	Screen Time of Tom	Screen Time of Jerry
1	11 min 25 sec	276 sec	180 sec
2	9 min 10 sec	326 sec	129 sec
3	14 min 25 sec	600 sec	120 sec
4	8 min 45 sec	372 sec	147 sec
5	4 min 23 sec	135 sec	126 sec

Table 3. Consequences of institution records (imply of YOLO set of rules 87.45 seems to be more compared with SVM set of rules 84.74 and preferred blunders suggest for YOLO set of rules is .200 and SVM set of rules is .592). Descriptive SPSS employs the unbiased pattern test of Accuracy and Precision on the dataset. It defines identical variances with and without assuming a T-check rating of organizations with a reaction charge of ten in each case.

	Group	N	Mean	Std. Deviation	Std error mean
Accuracy	YOLO	10	87.45	.63324	.20025
	SVM	10	84.74	1.87293	.59227

Table 4. The dataset is subjected to the Independent Sample T- test, with the confidence interval set to 95% and the level of significance set to 0.010. (The You Only Look Once Algorithm looks to outperform the Support Vector Machine)

	Lavene's test for equality of variances	T-test for Equality of Means	95% confidence interval of the difference

		F	Sig	t	df	sig(2 tailed)	Mean diff	Std.error	Lower	Upper
Accuracy	**Equal Variances assumed**	16.968	.001	4.311	18	.000	2.69500	.62521	1.31490	4.00851
	Equal variances not assumed			4.311	11.031	.001	2.69500	.62521	1.31940	4.07060

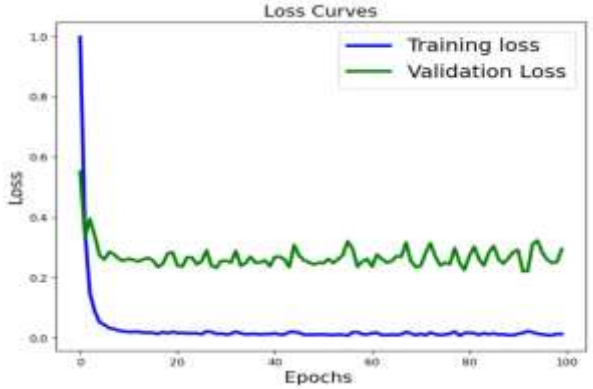

Fig. 1. Loss and Epoch Comparison in Terms of Training and Validation Loss. Here Epochs start from 0 to 100. it explores the Loss YOLO slightly better than SVM.X-axis:-Epochs,Y-axis:Loss.

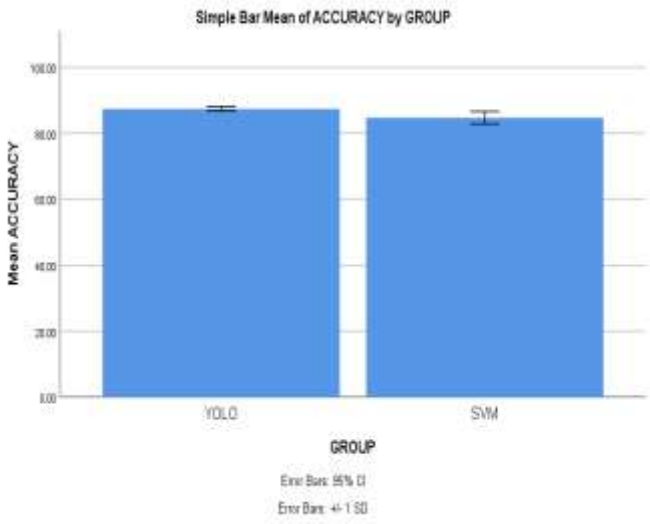

Fig. 2. In phrases of mean accuracy, the simplest look as soon as the set of rules outperformed the aid Vector gadget by means of 87.4 factors. It investigates that the mean accuracy is relatively better than that of the support Vector gadget, and the usual deviation is notably decreased. The bar graph graphically represented the usage of group identification as X-axis YOLO versus SVM, Y-axis supplying the error bars with an average accuracy of detection +/- 1 SD.