Journal of Engineering Technology and Applied Physics

Convolutional Neural Network-based Transfer Learning and Classification of Visual Contents for Film Censorship

Hor Sui Lyn, Sarina Mansor*, Nouar AlDahoul and Hezerul Abdul Karim
Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia.
*sarina.mansor@mmu.edu.my
https://doi.org/10.33093/jetap.2020.2.2.5

***Abstract* - Content filtering is gaining popularity due to easy exposure of explicit visual contents to the public. Excessive exposure of inappropriate visual contents can cause devastating effects such as the growth of improper mindset and rise of societal issues such as free sex, child abandonment and rape cases. At present, most of the broadcasting media sites are hiring censorship editors to label graphic contents manually. Nevertheless, the efficiency is limited by factors such as the attention span of humans and the training required for the editors. This paper proposes to study the effect of usage of Convolutional Neural Network (CNN) as feature extractor coupled with Support Vector Machine (SVM) as classifier in an automated pornographic detection system. Three CNN architectures: MobileNet, Visual Geometry Group-19 (VGG-19) and Residual Network-50 Version 2 (ResNet50_V2), and two classifiers: CNN and SVM were utilized to explore the combination that produce the best result. Frames of films fed as input into the CNN were classified into two groups: porn or non-porn. The best accuracy was 92.80 % obtained using fine-tuned ResNet50_V2 as feature extractor and SVM as classifier. Transfer learning and SVM have improved the CNN model by approximately 10 %.**

Keywords—Convolutional Neural Network (CNN), Deep Learning, Pornography, Support Vector Machine (SVM)

I. INTRODUCTION

Tight regulations are exercised in Malaysia regarding censorship of visual contents that involve pornography and nudity on screen. Media groups have to censor sensitive visual contents from being broadcasted to prevent violation of the law and penalties that follow such violation. The focus of this paper is on pornographic related graphics that appear in films. In this work, the definition of pornography is "any sexually explicit material with the aim of sexual arousal or fantasy" [1].

With the booming consumption of Internet, the controversial contents (e.g.: violence, profanity, drug abuse, nudity) can be freely accessible and exposed to the public with or without the viewers' intentions. Possible consequences of this phenomenon involve rises in cases of sexual addiction and sexual assault, as reported in the findings of a review on impact of internet pornography on teenagers [2]. Therefore, it is critical for the society to avoid the arise of such unfavorable circumstance from the bud of the problem.

Preventive measures have been taken by moderators of channels that can expose sexual explicit visual contents to the public. However, the traditional way of having the censorship editors to look through all films to be broadcasted in order to segregate those that contain sensitive contents is tedious, slow and inefficient. A solution to this problem is to automate the pornographic detection process, allowing computers to take over humans' role in this task.

Most researchers utilised deep learning techniques such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Support Vector Machine (SVM) in their works for recognition of adult

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License.
Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

contents in images or videos. In this work, we explore the combination of transfer learning using CNN and SVM.

The organisations of the subsequent parts of this paper are as follows. Section II discusses techniques used in other works to recognise explicit visual contents. Section III briefly describes the techniques to be implemented and Section IV presents the results obtained from the experiment. This is followed by analysis of the results in Section V. The conclusion and future works are included in Section VI.

II. LITERATURE REVIEW

In recent years, deep learning is one of the popular approaches for pornographic detection. Most of the state-of-the-art solutions utilise CNN to perform transfer learning. Generally, transfer learning allows learning of a new task (target task) based on previously learnt tasks (source task) through knowledge transfer [3]. In 2015, Moustafa [4] designed a classifier using a fusion of AlexNet [5] and GoogLeNet [6] to perform categorisation of pornographic images and videos. In the following year, Nian *et al.* [7] brought up a system which used deep CNN as a pornographic image detector. Performance of the system was fine-tuned by analysing the relationship between the validation result and training set distribution. In 2017, Ou *et al.* [8] proposed a deep multicontext network to recognise adult contents in images and key frames of videos by using fine-to-coarse strategy. Different deep learning models were utilised in the local (deep faster R-CNN-based models) and global (deep CNN-based models) modules. Consequently, minor errors of local context were used to rectify the judgement of the global context.

Furthermore, Ying *et al.* [9] developed a system capable of recognising images with explicit contents in the year 2018. Features produced by the CNN fine-tuned model was visualised such that the feature visualization analysis was utilised to improve the system's performance. Agastya *et al.* [10] also applied transfer learning on CNN. However, they performed 5-fold cross validation on their dataset in order to gauge the performance of their proposed method.

In addition to CNN, Wehrmann *et al.* [11] included the usage of RNN in their adult content detection system to allow sequence learning. AlDahoul *et al.* [12] chose to utilise a fast deep learning model called Local Receptive Field-Extreme Learning Machine (LRF-ELM) to allow automatic censorship to be performed in 2019.

In the context of classifier, Zhao *et al.* [13] implemented a system in 2010 to detect adult images using several SVM classifiers. Specifically, an SVM classifier was used to detect features of each class, where selected features of each pornographic category were distinct.

III. MATERIALS AND METHODS

The overall system architecture of the proposed system is illustrated in Fig. 1.

Key frames from input video files were extracted and used as inputs to the CNN. Each of the frames was pre-processed such that each RGB pixel value was normalised to a value in the range of 0 to 1. The resulting image was fed into a feature extractor, which is essentially a CNN model. Subsequently, the classifier would place a label for each frame, indicating whether that particular frame consisted of inappropriate visual content.

We proposed to use different classifiers (CNN and SVM) in order to study the effect of usage of SVM as classifier as opposed to the conventional CNN-only model. A visualisation of the general structure of a CNN is displayed in Fig. 2, where the classifier in the image can be either CNN or SVM. Three different CNN architectures, namely MobileNet [14], Visual Geometry Group-19 (VGG-19) [15] and Residual Network-50 Version 2 (ResNet50-V2) [16] were used in this work.

After all key frames from the same video file had gone through these processes and obtained their respective frame labels of either 0 ("Non Porn") or 1 ("Porn"), classification of the whole video file was performed by checking the ratio of "Porn" frames in the video against a threshold value to obtain the video's label. If the former is greater than or equal to the latter, then the video would be labelled as "Porn", otherwise it would be assigned the "Non Porn" label.

The specification of the platform selected to run the proposed system are as follow:

- AMD Ryzen 5 2600 Six-Core Processor @ 3.40 GHz
- Windows 10 Pro
- 16.0GB DDR4 at 2666 MHz
- NVIDIA GeForce GTX 1660

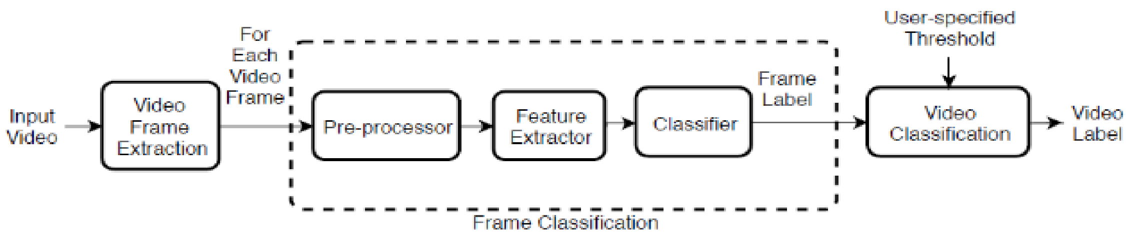

Fig. 1. System architecture of proposed system.

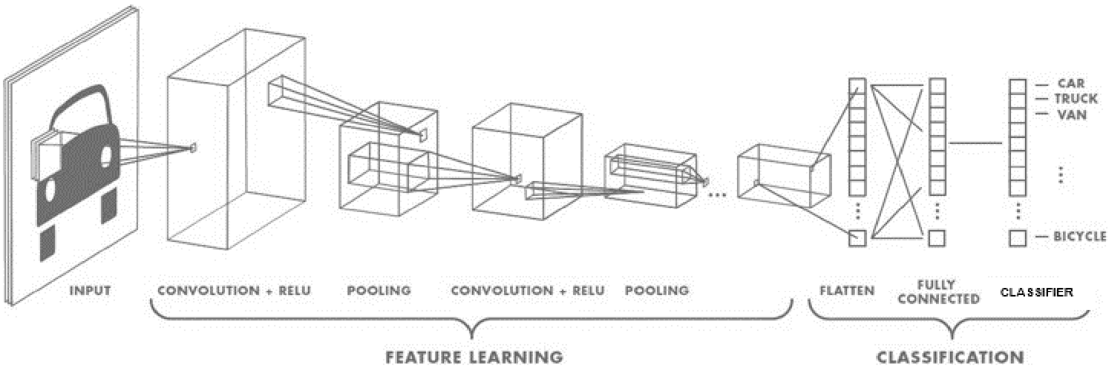

Fig. 2. Structure of CNN [17].

A total of 804 videos and 727 images obtained from the Internet with various settings [(anime and real world), angles, background, human skin colour and zoom levels], were used for the training and validation of the proposed system. Labels of 0 ("Non Porn") or 1 ("Porn") were assigned to each of the images and video frames belonging to the training dataset before the training process took place [18]. The amount of images and video files for each category is summarised in Table I.

Table I. Amount of data for each category.

	Porn	Non Porn	Total
Video	402	402	804
Video frames	5,519	14,258	19,777
Images retrieved from Internet	727	-	727
Total images	6,246	14,258	20,504

The data were split into two groups (training and validation) with similar distribution of probability. Around 80 % of the data of each class (11,872 images for "Non Porn" and 4,531 images for "Porn") were allocated for training purpose and the rest (2,386 images for "Non Porn" and 988 images for "Porn") were reserved for validation use. The splitting of data was done manually such that images from the same video were used for either training or validation but not both.

The ways of implementing CNN in the proposed system, summarised in Table II, will be explained in the following subsections. The training of CNN models was performed from Method 1 to Method 3 in ascending order.

Table II. Summary of implemented methods.

Method	Feature Extractor	Classifier
1	Traditional CNN-only	Softmax
2	Fine-tuned CNN-only	Softmax
3	Fine-tuned CNN (from Method 2)	SVM

For methods that used Softmax as classifier, two fully-connected layers with ReLU activation function and kernel regularisation function, and another fully-connected layer with softmax activation function were added to the top of the imported network to act as the classifier. Dropout layers were also added to simplify the network. Figure 3 shows a flowchart for the design of Methods 1 and 2 classifier.

Most of the CNN hyperparameters (learning rate, number of epochs, batch size et cetera) were fixed, leaving only certain types of values that could be manipulated, such as number of neuron available in the added fully-connected layers in the classifier layers and the number of layer to be fine-tuned when transfer learning was applied. RMSprop optimiser was employed in CNN of all the methods. A low learning rate of 0.0001 was used for all the CNN training so that the degree of change could be limited. Batch sizes, or number of samples per batch, of 32 and 16 were used on the training and validation dataset respectively to save memory space. Dropout rates were fixed at 0.5 or 50 %, which means 50 % of the inputs of the particular layer were excluded from each update cycle. Only L1 regulariser was used as the kernel regulariser such that 0.005 times of each weight coefficient value was added to the total network loss [19]. Rectified Linear Unit (ReLU) was the activation function used in all layers except the output layer, which utilised softmax function to allow the classification to be done. Furthermore, data augmentation was utilised to reduce overfitting and allow better generalisation by adding more data artificially [3].

Another type of classifier implemented was SVM [20] which tries to find the optimal decision boundary or hyperplane between points belonging to two different classes in order to solve a classification problem [19]. The architecture of SVM is illustrated in Fig. 4.

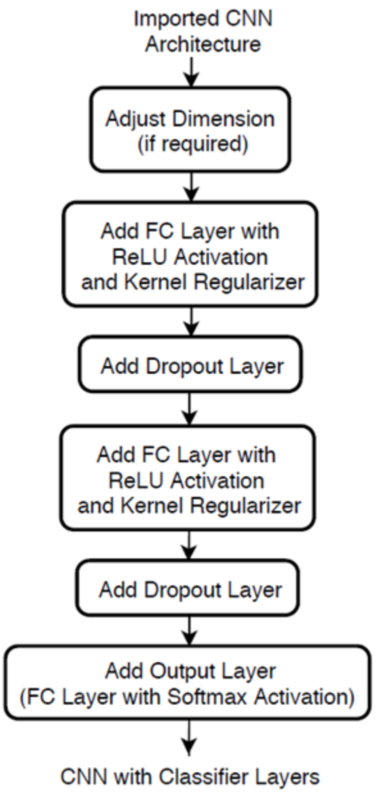

Fig. 3. Flowchart of design of Methods 1 and 2 classifier.

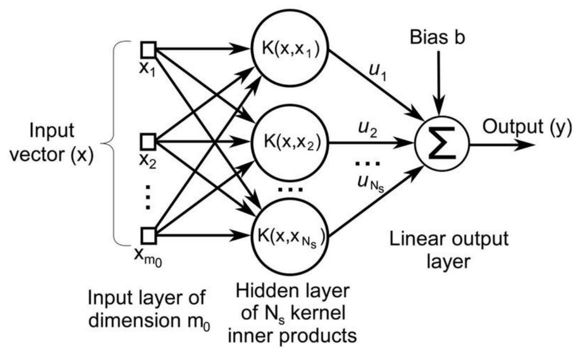

Fig. 4. Architecture of SVM [21].

A. Method 1: Traditional CNN-only

This method played the role of a control experiment, setting the baseline for CNN performance in the detection of pornographic visual contents.

A CNN architecture (up to the convolutional layers) acting as the feature extractor was imported with random initial weights. Then, the procedures in Fig. 3 were carried out to form the classifier layers. All the convolutional layers in the selected CNN (base model) were trained to adjust the connection weights according to the targeted dataset. Three combinations of numbers of neuron were experimented: (64, 8), (96, 24) and (128, 36), where the numbers enclosed in the parentheses denotes the number of neurons present in the first and second added fully-connected layer of the classifier respectively.

B. Method 2: Fine-tuned CNN-only

The effect of transfer learning was examined through the implementation of this method. Similar to Method 1, a CNN architecture (up to the convolutional base) acting as the feature extractor was imported. However, unlike Method 1, it was imported with initial weights from ImageNet classification task so that transfer learning was performed. Again, procedures in Fig. 3 were performed to enable the pre-trained CNN to act as the classifier aside from feature extractor. For each of the CNN architectures, the combination of number of neuron in the classifier which gave the best result (from Method 1) was used. This time, only the "number of layer with trainable weights" parameter was manipulated.

C. Method 3: Fine-tuned CNN + SVM Classifier

This method was implemented to study the effect of using a fine-tuned CNN as feature extractor coupled with SVM classifier to perform pornographic image classification. For each of the CNN architectures, the fine-tuned CNN with the best outcome obtained from Method 2 was used for feature projection while the classifier was an SVM.

IV. RESULTS

For all the CNN training involved, only 20 epochs were run to reveal the initial trends, sufficient to show whether transfer learning allowed CNN to perform better in the starting stage of training so as to reduce the training time required. Besides that, since performance of CNN was bound to saturate at a certain point, inference was made from the steepness or trend of the curves.

The general equation for calculation of accuracy is shown in Eq. (1). In all the implemented methods, confusion matrices were generated to aid user in understanding the classification errors made. The interpretation of confusion matrix is shown in Table III. As this algorithm is designed to detect pornographic images, detection of adult content is perceived as positive result. On the contrary, negative outcome means no inappropriate visual content is detected.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

, where TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

Table III. Interpretation of confusion matrix.

True Negative (TN): "Non Porn" image labelled correctly	**False Positive (FP)**: "Non Porn" image labelled as "Porn"
False Negative (FN): "Porn" image labelled as "Non Porn"	**True Positive (TP)**: "Porn" image labelled correctly

Precision is computed using Eqs. (2) and (3) whereas the computation for recall is shown in Eqs. (4) and (5) [22]. F1 score, which is calculated using Eq. (6) [23], is the weighted average of both the precision and recall. When F1 score approaches 1, the model is perceived as having a good performance, and the opposite if it approaches 0.

$$\text{Precision (Porn)} = TP / (TP + FP) \quad (2)$$

$$\text{Precision (Non Porn)} = TN / (TN + FN) \quad (3)$$

$$\text{Recall (Porn)} = TP / (TP + FN) \quad (4)$$

$$\text{Recall (Non Porn)} = TN / (TN + FP) \quad (5)$$

$$F1 = 2*[(\text{Precision}*\text{Recall}) / (\text{Precision}+\text{Recall})] \quad (6)$$

, where Precision = Percentage of relevant results irrespective whether they are classified correctly,
Recall = Percentage of relevant results that are classified correctly

The following subsections present the experimental results for Methods 1–3, where the highest validation accuracy obtained by each CNN model is bolded. The confusion matrices were generated using the validation dataset, consisting of 3,374 images in total.

A. Method 1: Traditional CNN-only

Table IV shows the results obtained from implementation of Method 1. Consequently, Table V presents the precisions, recalls and F1 scores of combinations that generated the highest validation accuracy for each of the CNN models using Method 1.

Table IV. Results of Method 1 implementation.

CNN Model	Number of Neuron	Number of Images (Total = 3,374)				Validation Accuracy (%)
		TP	*TN*	*FP*	*FN*	
Mobile Net	(64, 8)	2,103	812	283	176	86.40
	(96, 24)	2,135	798	251	190	86.93
	(128, 36)	2,190	726	196	262	86.43
VGG-19	(64, 8)	2,192	664	194	324	84.65
	(96, 24)	1,950	821	436	167	82.13
	(128, 36)	2,225	696	161	382	83.91
ResNet 50_V2	(64, 8)	2,156	708	230	280	84.88
	(96, 24)	2,199	721	187	267	86.54
	(128, 36)	2,237	689	149	299	86.72

Table V. Performance metrics of combinations that produced the highest validation accuracies using Method 1.

CNN Model	Number of Neuron	Precision (%)	Recall (%)	F1 Score (%)
MobileNet	(96, 24)	87	87	87
VGG-19	(64, 8)	84	85	84
ResNet50_V2	(128, 36)	86	87	86

B. Method 2: Fine-tuned CNN-only

Table VI shows the results obtained from implementation of Method 2. For this method, the best combination of number of neuron obtained from Method 1, which were (96, 24) for MobileNet, (64, 8) for VGG-19 and (128, 36) for ResNet50_V2, was noted and used to train each of the CNN architectures. This time, the parameter that was tuned was the number of layer towards the end (top) of the convolutional layers for each of the CNN architectures. Similar to Table V, Table VII presents the three performance metrics of the combinations that generated the highest validation accuracy for each of the CNN models using Method 2.

Table VI. Results of Method 2 implementation.

CNN Model	Number of Trainable Layer	Number of Images (Total = 3,374)				Validation Accuracy (%)
		TP	*TN*	*FP*	*FN*	
Mobile Net	10	2,357	638	29	350	88.77
	20	2,333	730	53	258	90.78
	30	2,343	755	43	233	91.82
VGG-19	5	2,280	817	106	171	91.79
	8	2,108	879	278	109	88.53
	10	2,070	872	316	116	87.20
ResNet 50_V2	20	2,316	811	70	177	92.68
	30	2,344	738	42	250	91.35
	50	2,354	673	32	315	89.72

Table VII. Performance metrics of combinations that produced the highest validation accuracies using Method 2.

CNN Model	Number of Trainable Layer	Precision (%)	Recall (%)	F1 Score (%)
MobileNet (96, 24)	30	92	92	92
VGG-19 (64, 8)	5	92	92	92
ResNet50_V2 (128, 36)	20	93	93	93

C. Method 3: Fine-tuned CNN + SVM Classifier

Table VIII shows the results obtained from implementation of Method 3, where Val Accuracy (last column) refers to validation accuracy. In this method, SVM classifiers replaced the role of CNN as classifiers in Method 2. The trained or fine-tuned CNNs that produced the best result for each of the CNN architectures using Method 2 was loaded to perform feature extraction. Due to memory limitation, instead of the output of convolutional base, the output of the first added dropout layer was fed as input to the SVM to allow classification to be performed. The precisions, recalls and F1 scores of combinations that generated the highest validation accuracy for each of the CNN models using Method 3 are recorded in Table IX.

Table VIII. Results of Method 3 implementation.

CNN Model	Number of Trainable Layer	Number of Images (Total = 3,374)				Validation Accuracy (%)
		TP	*TN*	*FP*	*FN*	
Mobile Net (96, 24)	30	2,313	817	73	171	92.77
VGG-19 (64, 8)	5	2,314	742	72	246	90.57
ResNet 50_V2 (128, 36)	20	2,294	837	92	151	92.80

Table IX. Performance metrics of combinations that produced the highest validation accuracies using Method 3

CNN Model	Number of Trainable Layer	Precision (%)	Recall (%)	F1 Score (%)
MobileNet (96, 24)	30	93	93	93
VGG-19 (64, 8)	5	91	91	90
ResNet50_ V2 (128, 36)	20	93	93	93

V. DISCUSSIONS

A. Comparison Between Methods

Table X, Table XI and Table XII compare the highest validation accuracies achieved by each of the methods explained in Section III for the different CNN architectures. The highest validation accuracy obtained by each CNN model is bolded.

A significant outcome that can be extracted from the aforementioned tables is validation accuracies obtained using Method 1 are the lowest regardless of the CNN model employed. For MobileNet and ResNet50_V2, validation accuracies achieved using Method 3 are the highest. On the other hand, for VGG-19, highest validation accuracy is obtained using Method 2. Interestingly, differences between accuracies obtained using Method 2 and Method 3 is in the range of 1.5 %.

Table X. Comparison among usages of MobileNet with (96, 24) neurons.

Method	Validation Accuracy (%)	Difference Between Current Validation Accuracy and Baseline Validation Accuracy (%)
1 (Baseline)	86.93	-
2	91.82	4.89
3	92.77	5.84

Table XI. Comparison among usages of VGG-19 with (64, 8) neurons.

Method	Validation Accuracy (%)	Difference Between Current Validation Accuracy and Baseline Validation Accuracy (%)
1 (Baseline)	84.65	-
2	91.79	7.14
3	90.57	5.92

Table XII. Comparison among usages of ResNetV2_50 with (128, 36) neurons.

Method	Validation Accuracy (%)	Difference Between Current Validation Accuracy and Baseline Validation Accuracy (%)
1 (Baseline)	86.72	-
2	92.68	5.96
3	92.80	6.08

B. Hyperparameters

An interesting finding is observed from the data in Table IV. No particular pattern (increasing or decreasing trend) for the number of neuron in classifier can be detected in order to obtain a better network performance. Similarly, data in Table V show that there is no specific way (increase or decrease) to manipulate the number of trainable layer in the pre-trained CNN architectures in order to improve the results. This implies that the optimal CNN solution cannot be deduced by performing simple experiments as there are still limitless possibilities for combinations of all the CNN hyperparameters. On the contrary, hyperparameters of CNN models should be tuned in a systematic and effective manner in order to allow improvement of network performance. This is an important issue for future research.

C. Transfer Learning

Higher training and validation accuracies are obtained in Method 2 as compared to Method 1. Besides the un-optimised CNN used, it seems possible for the relationship between source task and target task to be a reason for this phenomenon. As mentioned in [3], the source task and its relationship with the target task affect the effectiveness of transfer method. In this case, detection of sexually explicit visual content is not part of the source ImageNet classification task, which may be an explanation for this observation.

D. Classification Reports and Confusion Matrices

Classification reports and confusion matrices show that most of the accuracies, precisions and recall rates of "Non Porn" images and video frames are higher than those of the "Porn" category. This inconsistency may be due to more data belonging to the "Non Porn" class than the "Porn" class as shown in Table I.

Another possible explanation is features of data belonging to the "Porn" class are not definite and clear-cut, especially with the inclusion of "Non Porn" data which have close characteristic to Porn images categorised such as sumo wrestling, breast feeding, beach wears and people undressing.

E. SVM versus CNN as Classifier

Findings show that usage of SVM as classifier yielded better accuracy values than CNN (in the range of 2 %) in all but one case. This is likely because inputs to the SVM were not outputs obtained from the convolutional base but rather from the first dropout layer added to the trained model, which forms part of the classifier. The reason for making such decision was due to the limitation of memory space.

Another possible explanation for this is the trained CNNs used were not the optimum ones as the hyperparameters that can affect the outcomes were randomly assigned. No specific strategy was used to tune them and only a small amount of experiments was conducted to test different combinations of hyperparameter values. On the contrary, SVM is a matured classifier that can be implemented easily as compared to CNN.

F. Best Combination

Based on the findings, the CNN models that produced the most desirable results for each method are shown in Table XIII. Overall, the best accuracy performance was 92.80 % achieved using the (128, 36) as the combination of number of neuron in fine-tuned ResNet50_V2 and SVM as classifier (Method 3).

Table XIII. Compilation of CNN model with best results.

Method	CNN Model
Method 1: Traditional CNN-only	MobileNet with (96, 24) neurons
Method 2: Fine-tuned CNN-only	ResNet50V2_50 with (128, 36) neurons and last 20 layers in convolutional base trainable
Method 3: Fine-tuned CNN + SVM Classifier	ResNet50V2_50 with (128, 36) neurons and last 20 layers in convolutional base trainable

VI. CONCLUSION

Deep learning applied using CNN allowed automated detection of inappropriate visual content in films by feeding a wide range of examples during the training process. Once the training is done, new unseen video's frames were applied to be filtered utilising the trained model. As demonstrated in the study, effectiveness of transfer learning varies according to the degree of similarity between the source task, where initial weights are imported from, and the target task.

Overall, the best CNN model with a validation accuracy of 92.80 % was obtained by applying Method 3 which uses SVM as classifier. The CNN model utilised was fine-tuned ResNet50_V2 CNN architecture, with (128, 36) as the combination of number of neuron and 20 layers towards the classifier set as trainable. This work strengthens the idea that transfer learning can improve CNN performance. Notwithstanding the limitations of this project, SVM classifier performed better than CNN-only classifier.

This research has thrown up many questions in need of further investigation and confirmation. As the experiments in this project involved usage of hyperparameters which were decided randomly, further research might explore the effects of hyperparameter tuning methods in optimising CNN model to improve its performance. Considerably more work will need to be done to determine the optimiser that can yield better performance than RMSprop optimiser utilised in this project. Another natural progression of this work is to apply the adjustment to training data explored in [13] to analyse the possible improvements that can be brought by this technique.

A fruitful research area for future work is the utilisation of other deep learning techniques like LSTM. Usage of this RNN technique allows sequence learning to be performed aside from learning of spatial data, which may be beneficial to detection problem in films. In addition, filtering of contents of videos such that output of the system is the censored version of original videos is possible with application of this method.

REFERENCES

[1] M. Short, L. Black, A. Smith, C. Wetterneck and D. Wells, "A Review of Internet Pornography Use Research: Methodology and Content from the Past 10 Years," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 1, pp. 13-23, 2012.

[2] E. Owens, R. Behun, J. Manning and R. Reid, "The Impact of Internet Pornography on Adolescents: A Review of the Research," *Sexual Addiction & Compulsivity*, vol. 19, no. 1-2, pp. 99-122, 2012.

[3] L. Torrey and J. Shavlik, "*Transfer Learning*," in Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques, E. Soria, J. Martin, R. Magdalena, M. Martinez and A. Serrano, Ed. IGI Global, pp. 242-264, 2009.

[4] M. Moustafa, "Applying Deep Learning to Classify Pornographic Images and Videos," in *7th Pacific-Rim Symposium on Image and Video Technology (PSIVT 2015)*, Auckland, 2015.

[5] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.

[7] F. Nian, T. Li, Y. Wang, M. Xu and J. Wu, "Pornographic Image Detection Utilizing Deep Convolutional Neural Networks," *Neurocomputing*, vol. 210, pp. 283-293, 2016.

[8] X. Ou, H. Ling, H. Yu, P. Li, F. Zou and S. Liu, "Adult Image and Video Recognition by A Deep Multicontext Network and Fine-To-Coarse Strategy," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 5, pp. 1-25, 2017.

[9] Z. Ying, P. Shi, D. Pan, H. Yang and M. Hou, "A Deep Network for Pornographic Image Recognition Based on Feature Visualization Analysis," in *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, pp. 212-216, 2018.

[10] I. Agastya, A. Setyanto, Kusrini and D. Handayani, "Convolutional Neural Network for Pornographic Images Classification," in *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, Subang Jaya, 2018.

[11] J. Wehrmann, G. Simões, R. Barros and V. Cavalcante, "Adult Content Detection in Videos with Convolutional and Recurrent Neural Networks," *Neurocomputing*, vol. 272, pp. 432-438, 2018.

[12] N. AlDahoul, H. A. Karim, M. H. Lye Abdullah, M. F. Ahmad Fauzi, S. Mansour and J. See, "Local Receptive Field-Extreme Learning Machine-based Adult Content Detection," in *2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, pp. 128-133, 2019.

[13] Z. Zhao and A. Cai, "Combining Multiple SVM Classifiers for Adult Image Recognition," in *2010 2nd IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, pp. 149-153, 2010.

[14] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Adreetto and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *preprint*:1704.04861, 2017.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, 2015.

[16] K. He, X. Zhang, S. Ren and J. Sun, "Identity Mappings in Deep Residual Networks," in *European Conference on Computer Vision*, pp. 630-645, 2016.

[17] "A Comprehensive Guide to Convolutional Neural Networks - the ELI5 Way," 2020. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Accessed: 18- Nov- 2020].

[18] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2010.

[19] F. Chollet, *Deep Learning with Python*, 1st ed. Manning Publications, pp. 104-110, 2017.

[20] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] R. Ruiz-Gonzalez, J. Gomez-Gil, F. J Gomez-Gil and V. Martínez-Martínez, "An SVM-Based Classifier for Estimating The State of Various Rotating Components in Agro-Industrial Machinery with A Vibration Signal Acquired from A Single Point on The Machine Chassis," *Sensors*, vol. 14, pp. 20713-20735, 2014.

[22] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in *Proceedings of The 23rd International Conference on Machine Learning*, pp. 233-240, 2006.

[23] "Scikit-learn: Machine Learning in Python - scikit-learn 0.22.1 documentation", Scikit-learn.org, 2020. [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 16- Jan- 2020].