Automated Detection of Visual Contents for Film Censorship

Student Name : Hor Sui Lyn

Student ID : 1161300122

Majoring : CE

Supervisor : Dr. Sarina binti Mansor

Moderator : Dr. Hezerul Bin Abdul Karim

INTRODUCTION

- Unconstrained and freely accessible through the Internet
- Adolescents and children may take it as normality or curious to try
- Distort healthy sexual development
- Censorship prevents exposure

PROBLEM STATEMENT

- Manual detection:
  - Limited human attention span
  - Training to know the standards

⇒**Automation of pornographic scene detection**

Challenge:

Wide range of **colours** and **textures** of exposed skin

PROJECT OBJECTIVE

- To obtain and study pornographic data set from existing data set
- To study various image processing techniques to detect explicit scene from video
- To evaluate various CNN features and their encoding strategies

PROJECT SCOPE

- Detection of pornographic visual contents
- Pornography: “any sexually explicit material with the aim of sexual arousal or fantasy” [1]
- Deep learning via CNN
  - Specialises in automatic pattern and object detection
  - Automated detection without human intervention is achievable
- Frames extracted from videos = image

[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. Available: 10.1089/cyber.2010.0477.

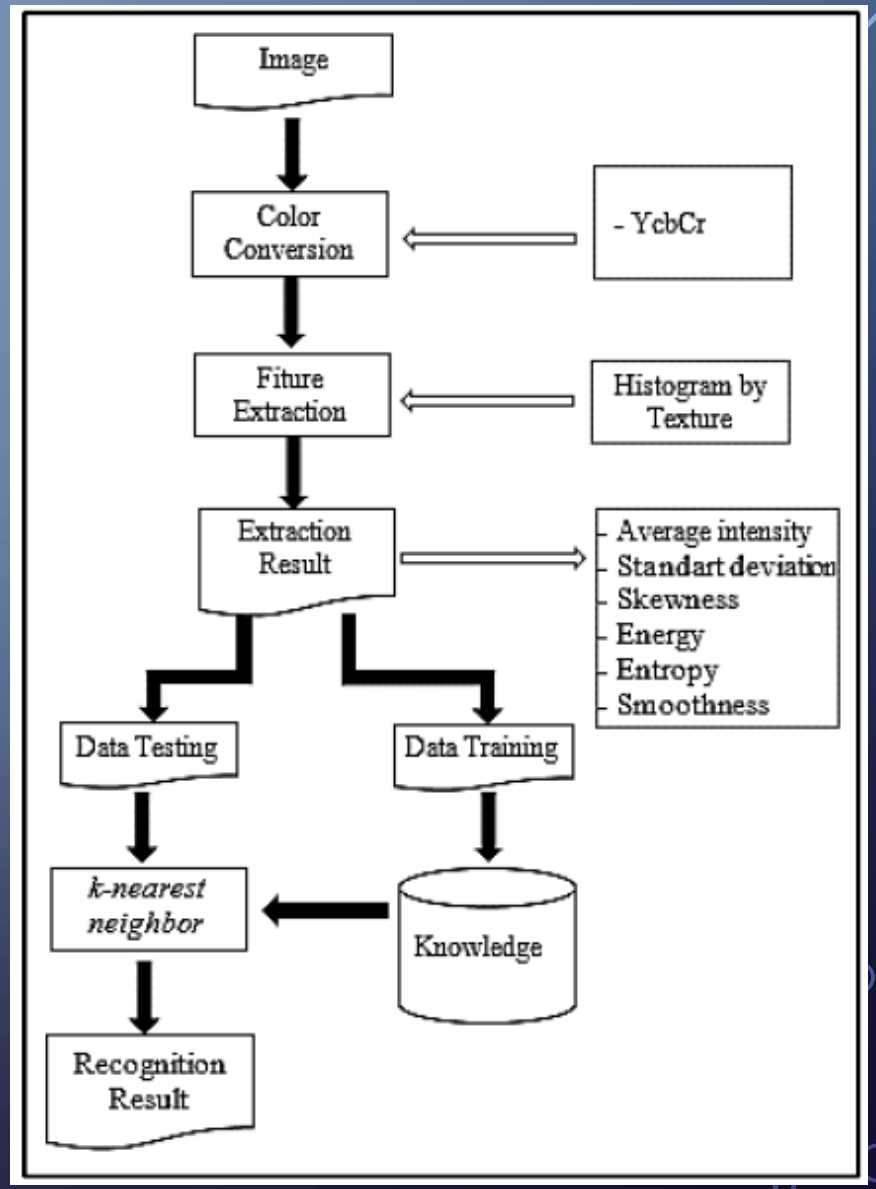
LITERATURE REVIEW

- Digital Image Processing Techniques: 2
- SVM Classifiers: 1
- Deep Learning: 7

DIGITAL IMAGE PROCESSING TECHNIQUES (1)

Implementation of K-NN Based on Histogram at Image Recognition for Pornography Detection [2]

- Training: 2 images per category
- Accuracy: 90% (on 60 Google images)
- Limited to cases where distribution of exposed skin area is similar to training data

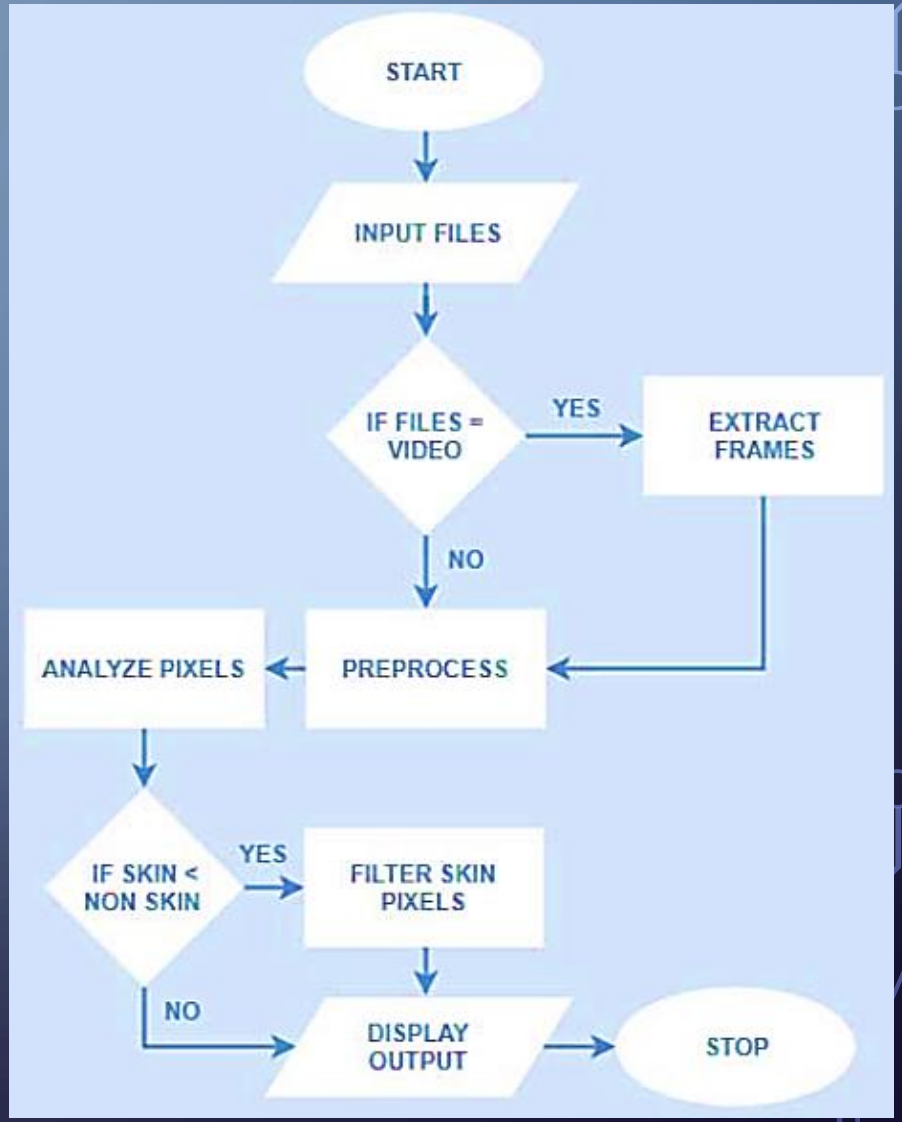


[2] S. Nuraisha, F. Pratama, A. Budianita and M. Soeleman, "Implementation of K-NN based on histogram at image recognition for pornography detection", in *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, 2017.

DIGITAL IMAGE PROCESSING TECHNIQUES (2)

A Pornographic Image and Video Filtering Application Using Optimized Nudity Recognition and Detection Algorithm [3]

- Pixel-by-pixel segmentation
- Rule-based classification
- Various lighting condition
- Accuracy: 80.23%
- Assumption: pornographic contents include nudity
- Tedious to specify the conditional expressions

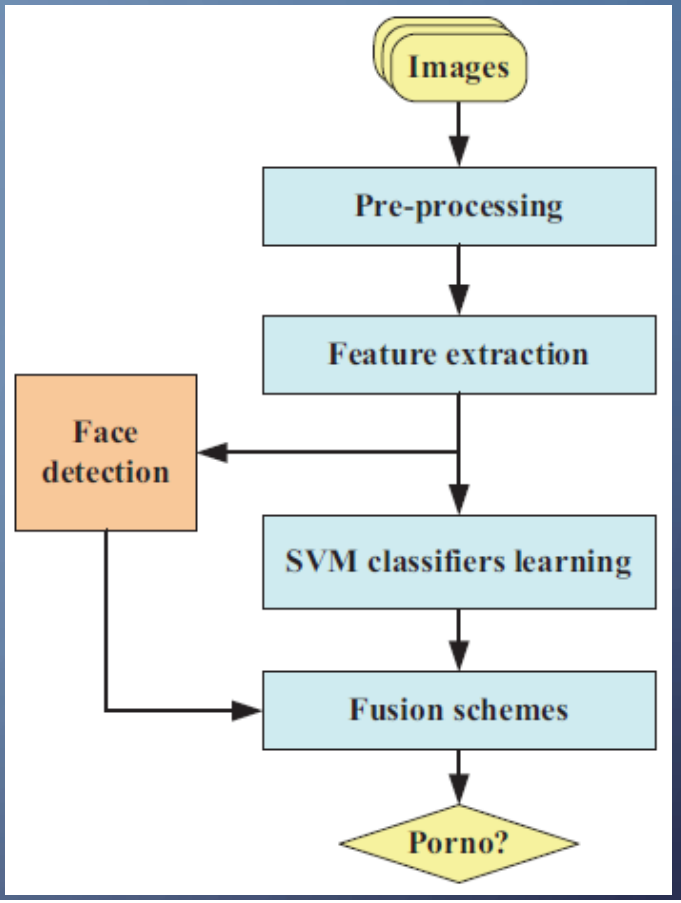


[3] M. Garcia, T. Revano, B. Habal, J. Contreras and J. Enriquez, "A pornographic image and video filtering application using optimized nudity recognition and detection algorithm", in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Baguio City, 2018, pp. 1-5.

SVM CLASSIFIERS

Combining Multiple SVM Classifiers for Adult Image Recognition [4]

- Training: 5-fold cross-validation, supervised learning
- TP: 87.68%, FP: 14.17% (on 50k Internet images)
- Different features for different categories
- Face detection algorithm may eliminate pornographic images

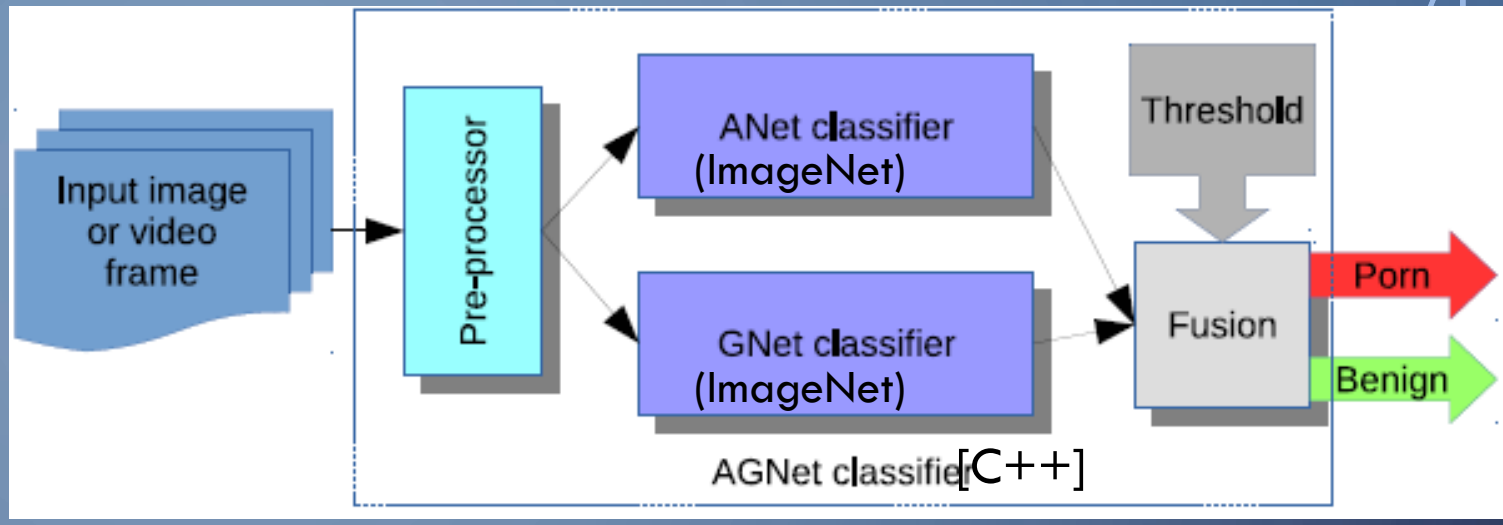


Erotic category	Features
Hardcore	HSV correlogram, Gabor, SIFT, skin, RGB moment
Nude	HSV correlogram, skin, LBP, RGB moment
Blowjob	HSV correlogram, Gabor, LBP, skin, RGB moment
Breast	HSV correlogram, Gabor, LBP, EDH, skin, RGB moment

[4] 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, 2010, pp. 149-153.

DEEP LEARNING (1)

Applying Deep Learning to Classify Pornographic Images and Videos [5]

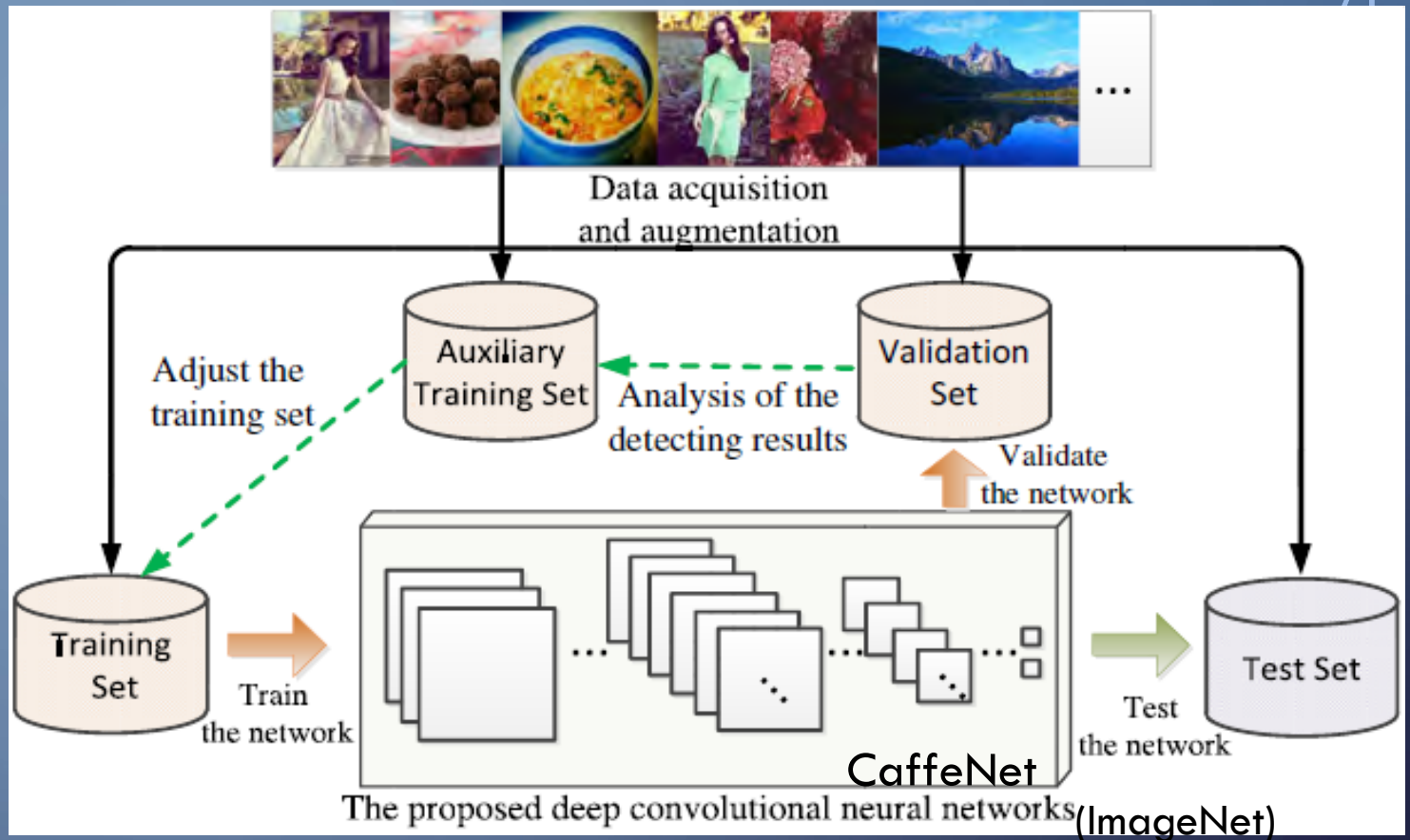


- ANet and GNet: Transfer learning, only last layer re-trained
- For frame: Compare threshold with averaged scores (AGNet) or max score (AGbNet)
- For video: Majority voting on key frames
- Accuracy: 93.8% (AGNet, on NPDI dataset); 94.1% (AGbNet, on NPDI dataset)
- Assumption: ANet and GNet produce different classification errors
- Higher computational power (if run in parallel)

[5] 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.

DEEP LEARNING (2)

Pornographic Image Detection Utilizing Deep Convolutional Deep Networks [6]



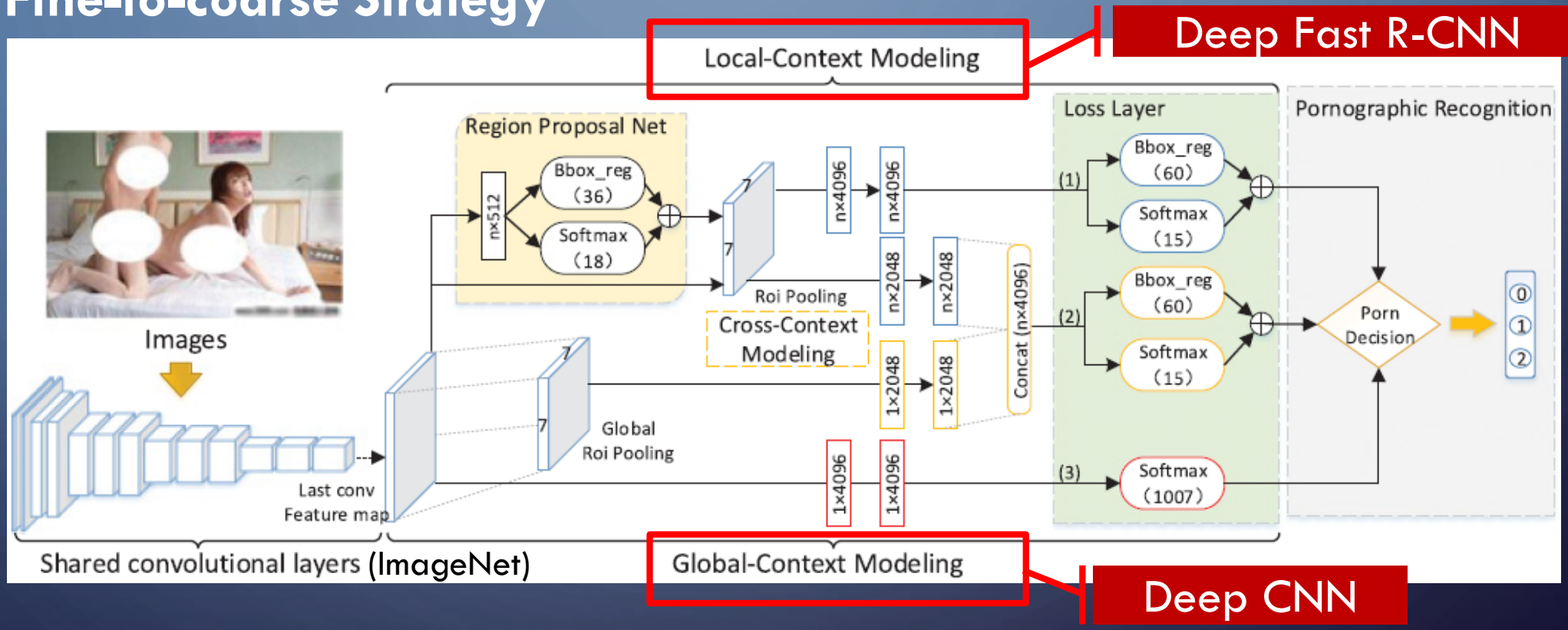
- Transfer learning, parameters trainable
- Data augmentation
- Relationship between validation results and training set distribution used to fine-tune network, repeat until test accuracy converges
- Fixed-point algorithm: to speed up computation time (-0.86% accuracy)
- Accuracy: 98.6% (on test data)

[6] 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. Available: 10.1016/j.Neucom.2015.09.135.

DEEP LEARNING (3)

Adult Image and Video Recognition by a Deep Multicontext Network and Fine-to-coarse Strategy [7]

DMCNet [CAFFE]



[7] 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. Available: 10.1145/3057733.

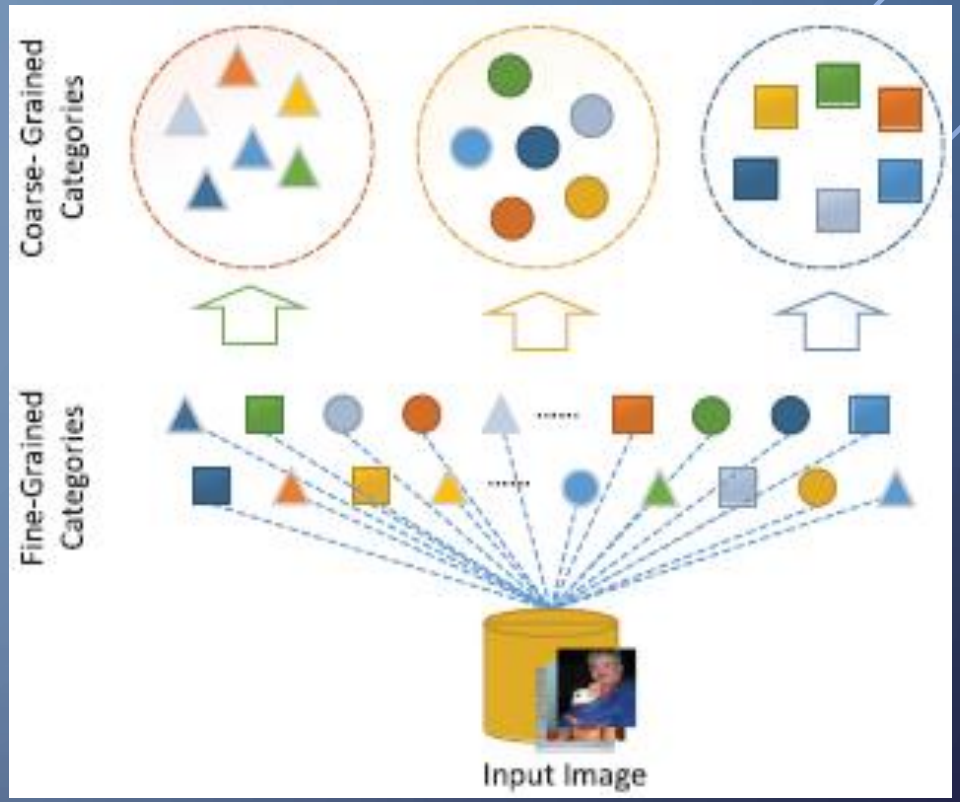
DEEP LEARNING (3)

Classes: S00, S01, S02

- Transfer learning, network fine-tuned
- Fine-to-coarse strategy: accelerate detection process
- Accuracy: 99.1% (on L2 dataset), 97.8% (on L3 dataset)

Classes: S00, S01+S02

- Higher complexity leads to longer development time

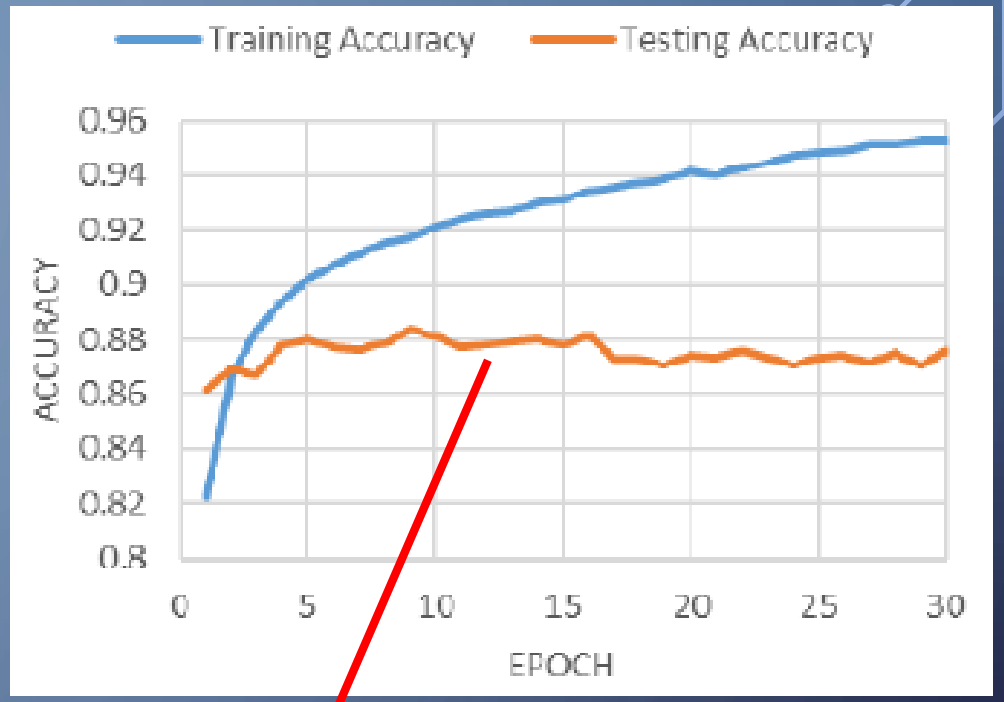


Dataset:
S00: normal
S01: adult
S03: unsuitable for children

DEEP LEARNING (4)

Convolutional Neural Network for Pornographic Images Classification [8]

- VGG-16 [Python, Keras framework]
- Transfer learning
- 5-fold cross-validation
- Accuracy: 95.4% (training on NPDI dataset), 93.8% (testing on NPDI dataset)

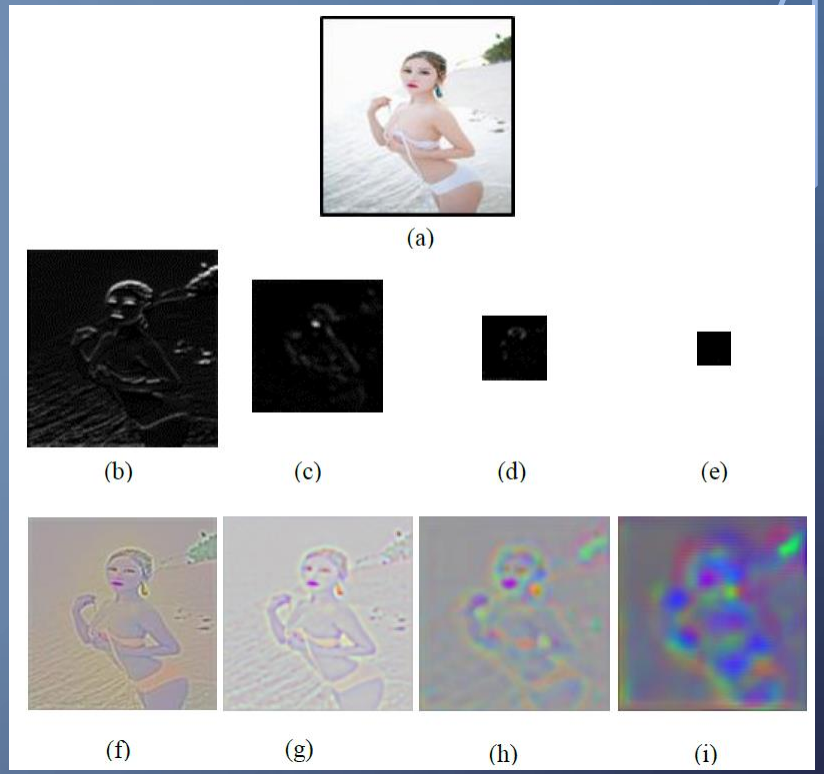


- Difficulty in classification in current fold is decreased significantly compared to other folds

[8] 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.

DEEP LEARNING (5)

A Deep Network for Pornographic Image Recognition Based on Feature Visualization Analysis [9]



- VGG-16 and CaffeNet
- Transfer learning (exclude last 3 FC)
- Features extracted visualised to determine data augmentation method
  - Display forward output (direct)
  - De-convolute feature, then output as image
- Max accuracy: 94.7% (increase of 2.6% from using fine-tuned VGG-16)
- Requires human intervention to decide data augmentation method to improve results

[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, 2018, pp. 212-216.

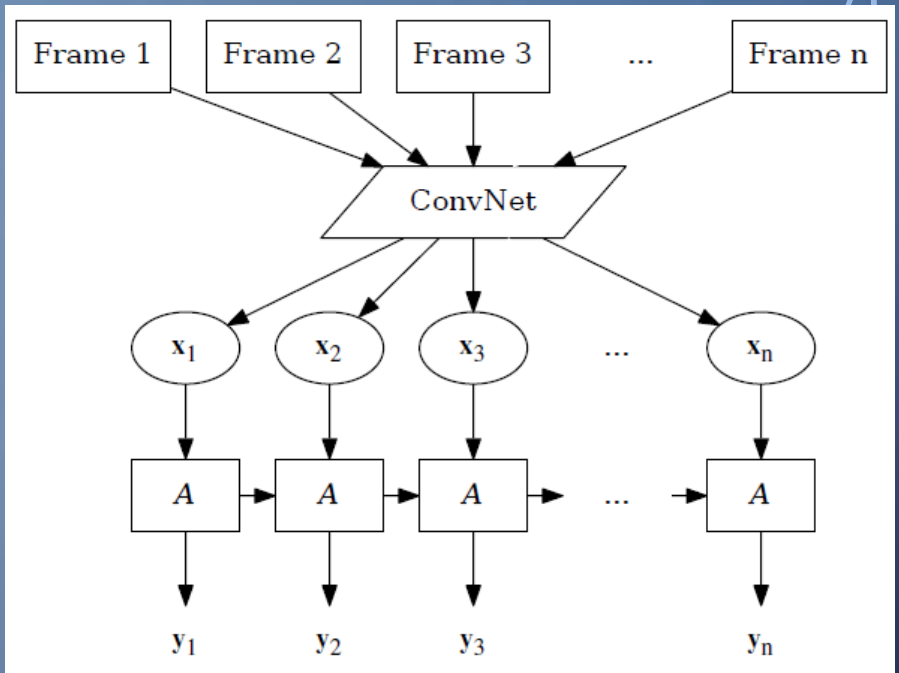
DEEP LEARNING (6)

Adult Content Detection in Videos with Convolutional and Recurrent Neural Networks [10]

- transfer learning
- training not required

- Video frames extracted
- CNN + multiple cropping for feature extraction
- LSTM (RNN) for sequence learning
- Video classification: majority voting
- Max accuracy: 95.6% by ACORDE-101 (on NPDI dataset)
- Useful in real-time applications

- training required

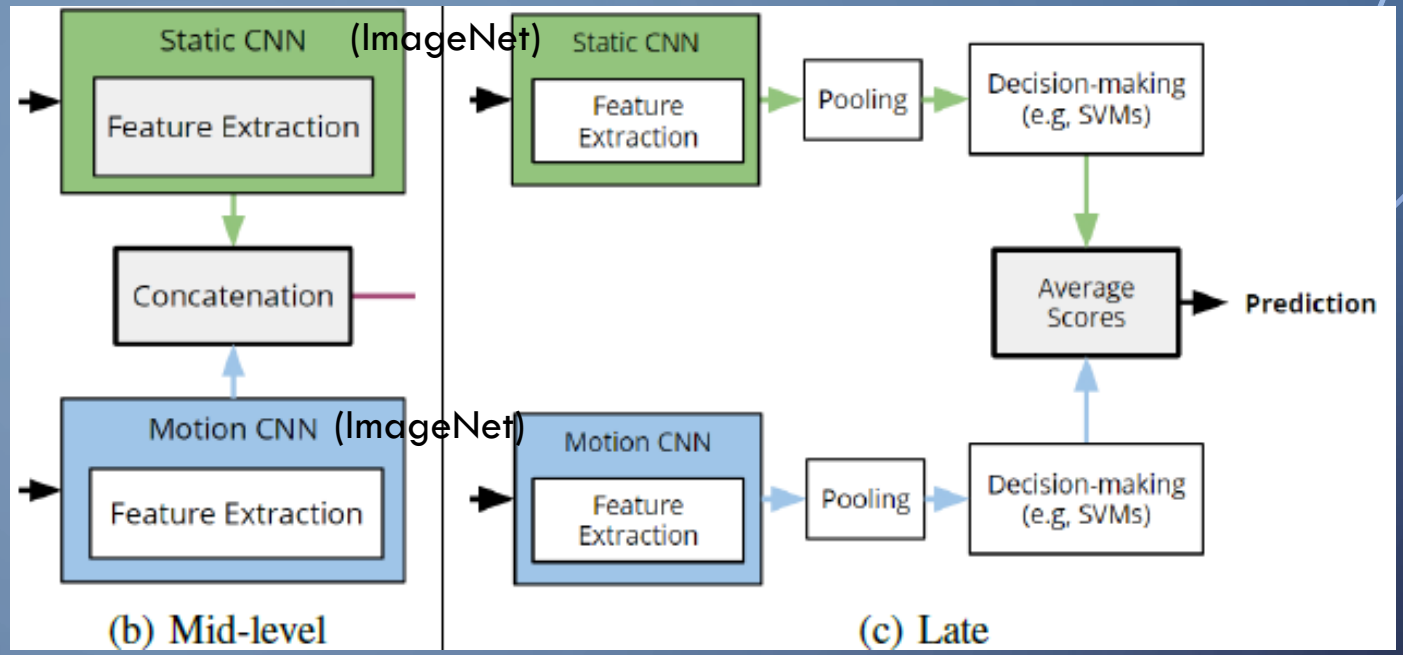


ACORDE

[10] 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. Available: 10.1016/j.Neucom.2017.07.012.

DEEP LEARNING (7)

Video Pornography Detection through Deep Learning Techniques and Motion Information [11]



- Static feature extraction and all classifications: GoogLeNet (CNN)
- Motion (temporal) information extraction:
  - Optical flow displacement field (magnitude and direction)
  - MPEG motion vector (Movement of macroblock represented using motion vector)
- Accuracy: 96.4% with Late Fusion (on Pornography-2k dataset), 97.9% with Mid-level or Late Fusion using optical flow method (on Pornography-800 dataset)

transfer learning

[11] M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S. Goldenstein, A. Rocha, "video pornography detection through deep learning techniques and motion information", neurocomputing, vol. 230, pp. 279-293, 2017. Available: 10.1016/j.Neucom.2016.12.017.

SUMMARY OF LITERATURE REVIEW

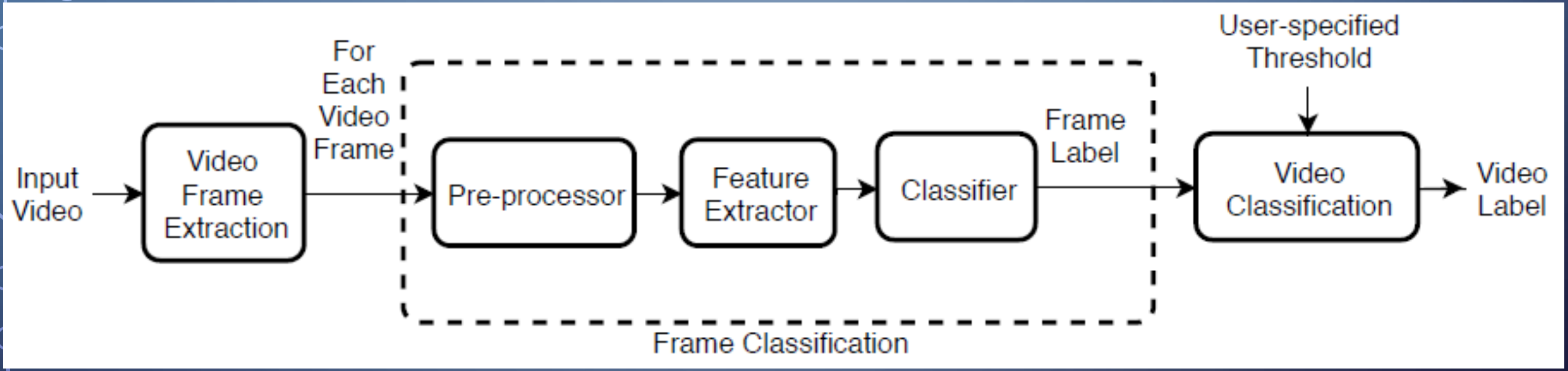
- Digital image processing techniques and SVM classifiers:
  - Rely on assumptions that affect accuracy
  - Human intervention required
- Deep learning (CNN):
  - Learn without human intervention
  - Feature to be extracted decided automatically
  - Dataset has to cover a wide range in an unbiased way
- K-fold cross-validation not suitable

=> manual distribution of training and validation dataset

Experiment on:

- CNN with and without transfer learning to extract features
- Different classifiers (CNN and SVM)

SYSTEM ARCHITECTURE



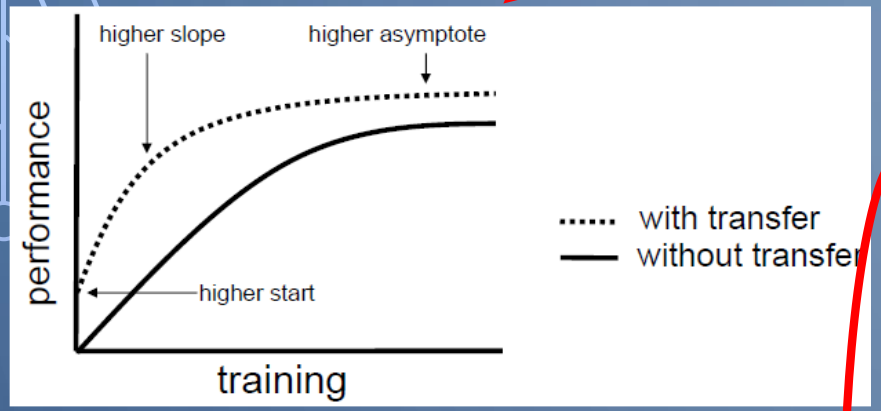
DATA DISTRIBUTION & CNN ARCHITECTURES

Dataset	"Non Porn" Images	"Porn" Images	Total
Training	11,872	4,531	16,403
Validation	2,386	988	3,374
Total	14,258	5519	19,777

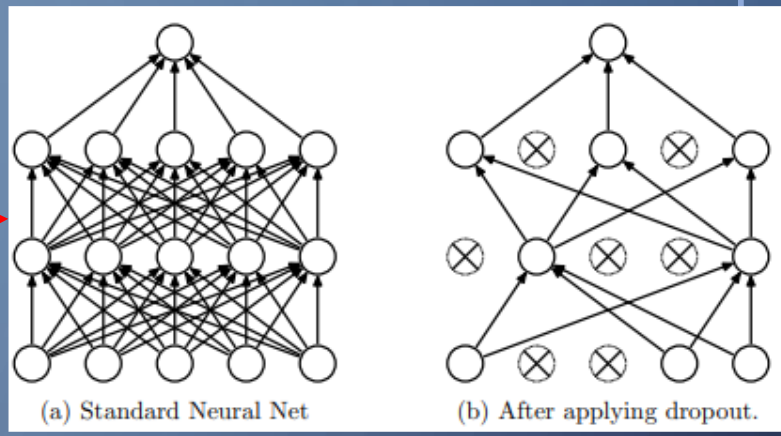
804 videos (402 "Non Porn" and 402 "Porn") and 727 images (all "Porn") from Internet

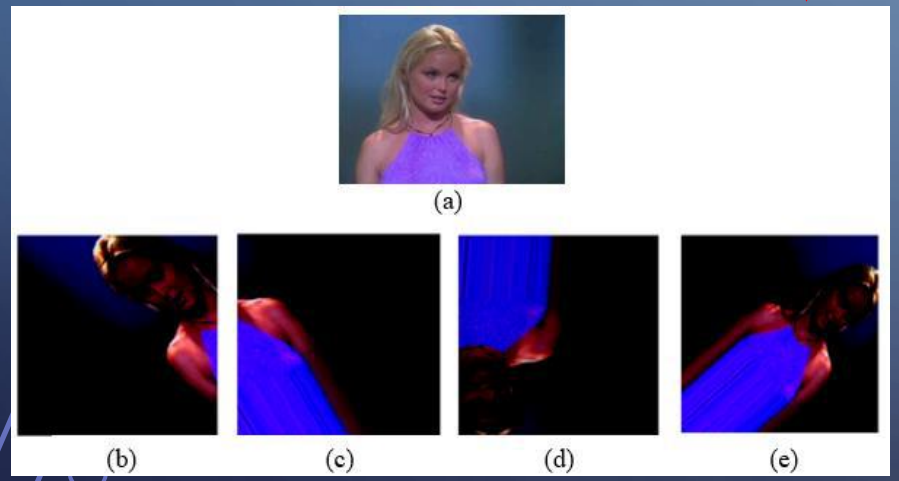
CNN Architecture	Number of Layer
MobileNet	30
VGG-19	19
ResNet50_V2	50

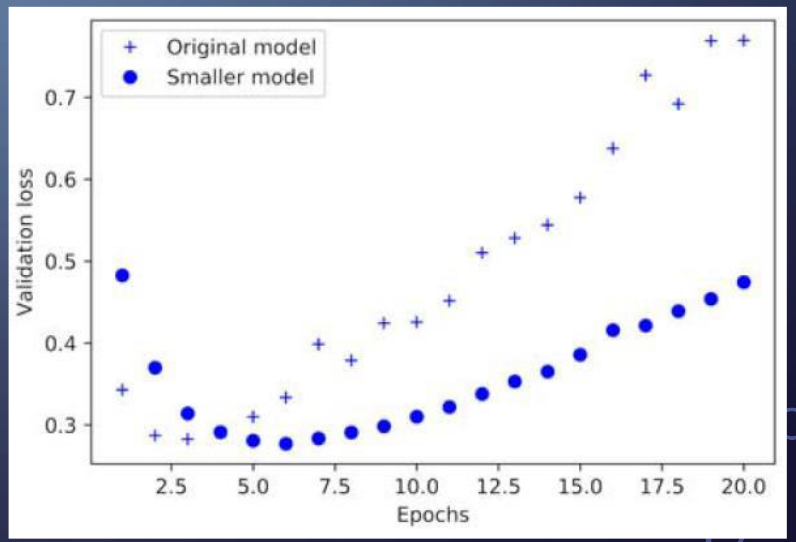
CNN IMPLEMENTATION TECHNIQUES



- Transfer learning (ImageNet)
- Data augmentation
- Dropout layer
- Reduction of network capacity
- L1 kernel regularisation



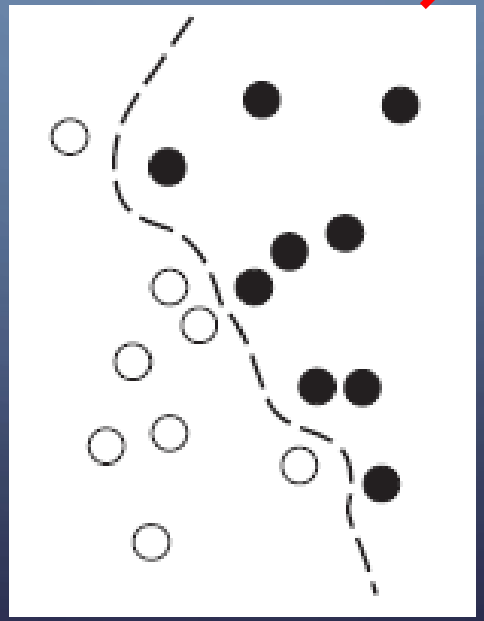


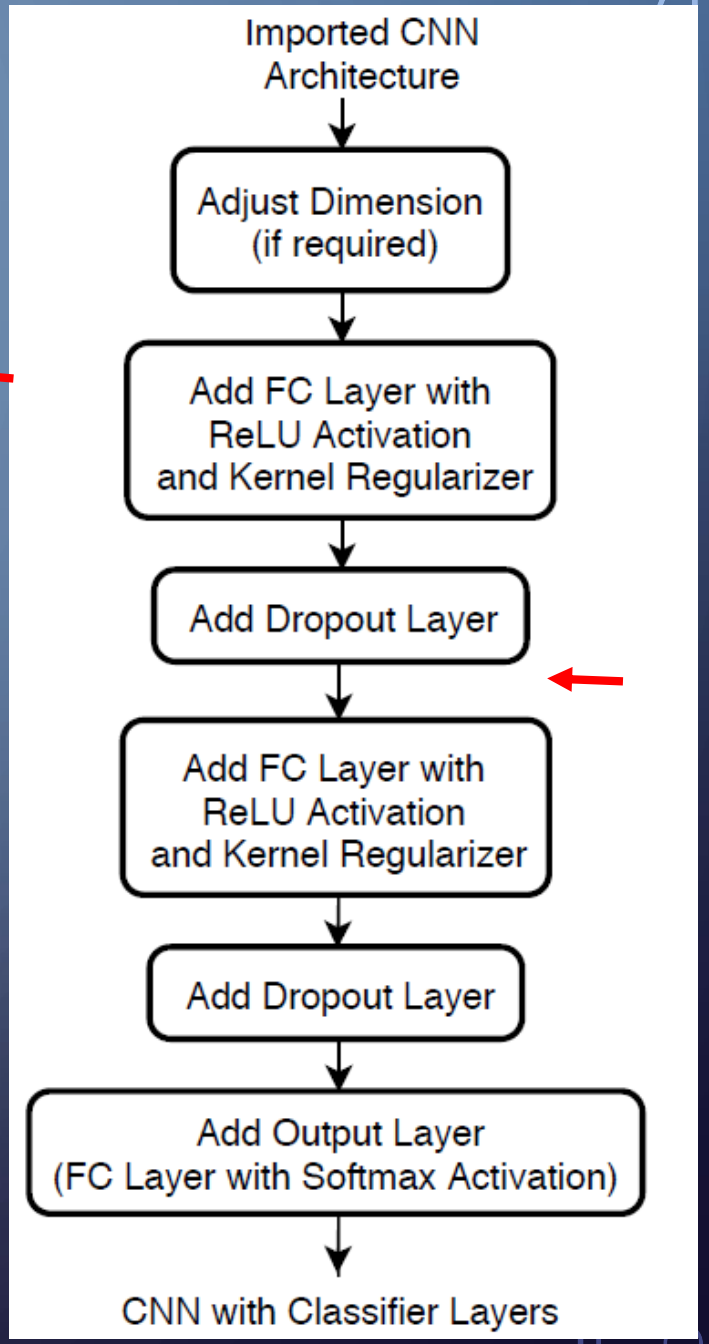


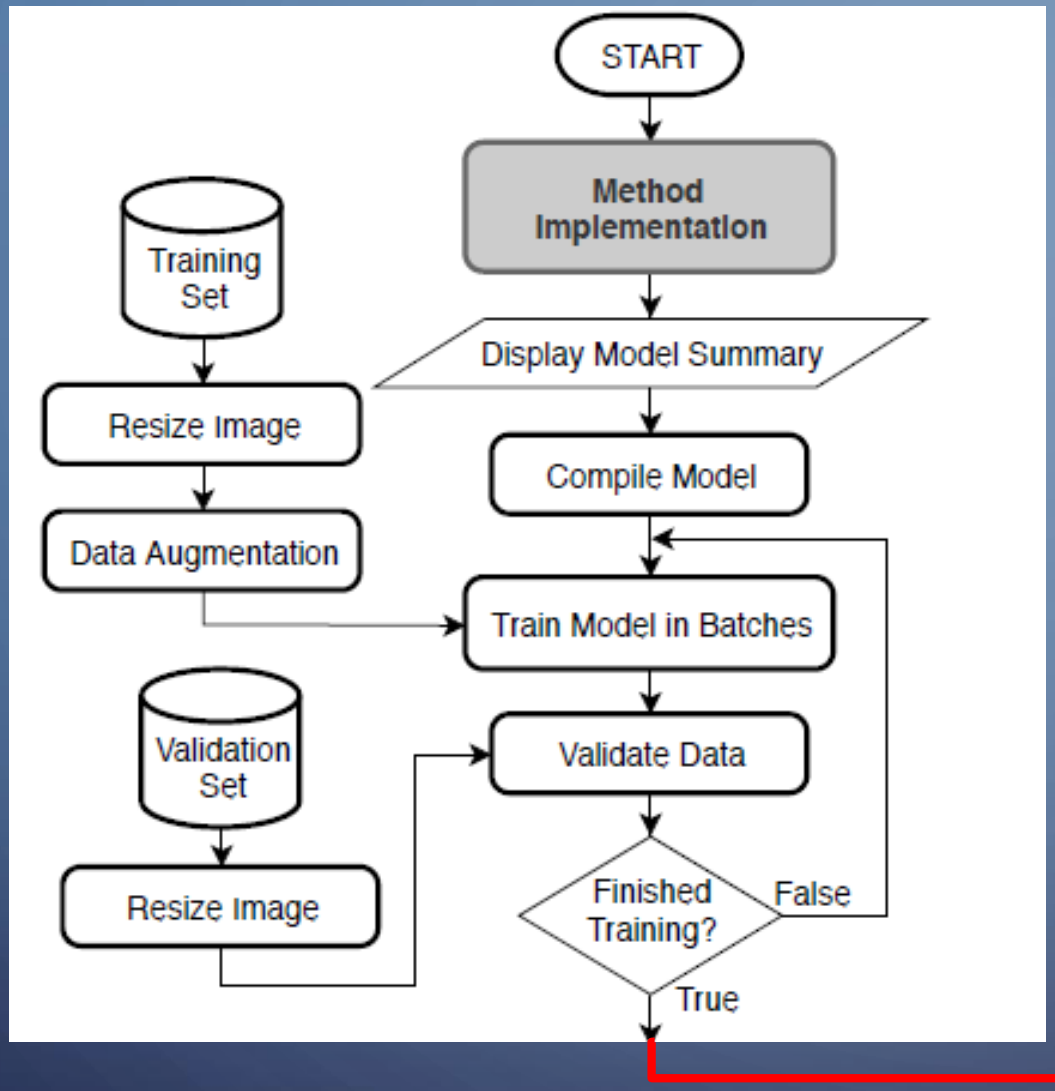
METHODS OF CNN IMPLEMENTATION

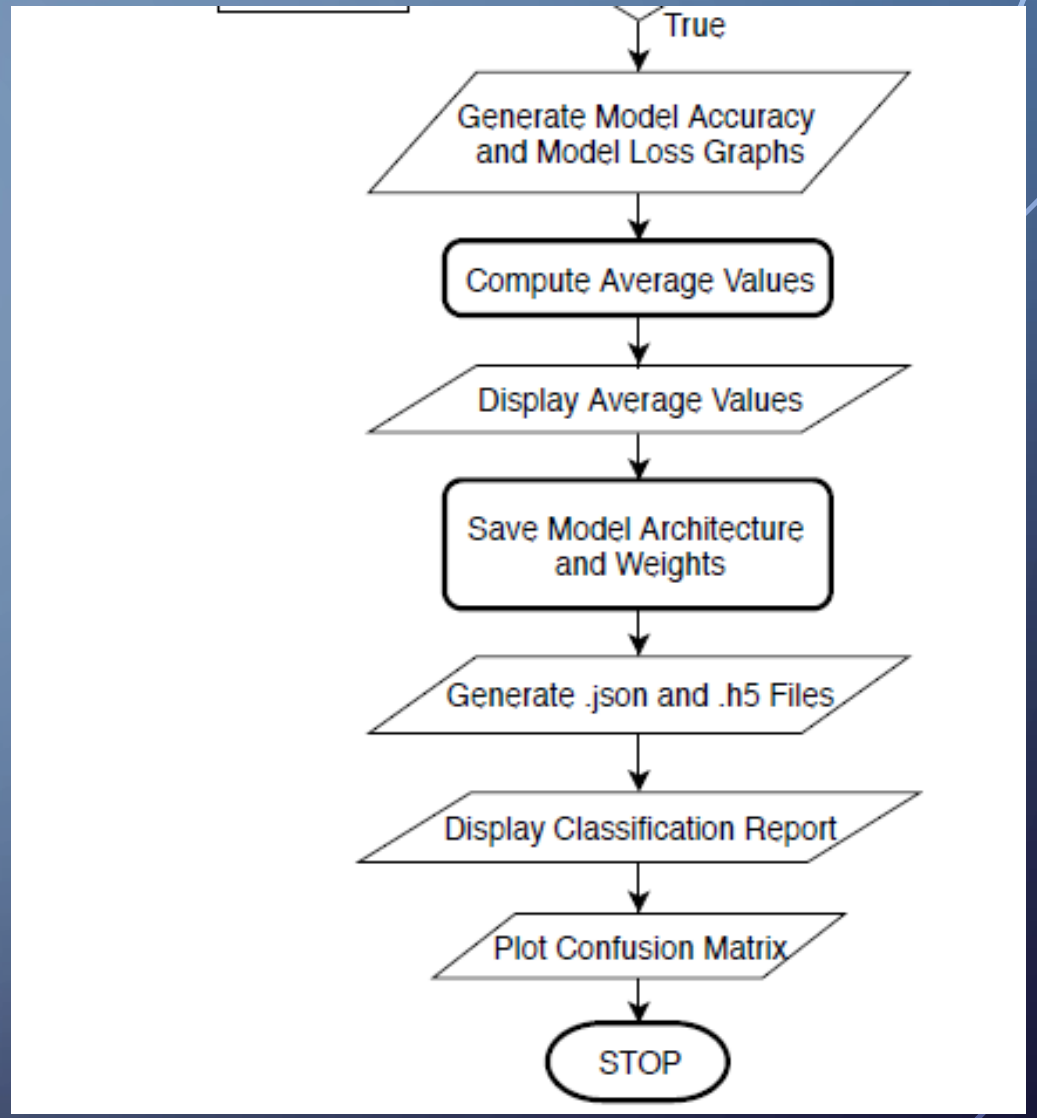
Method	Feature Extractor	Classifier
1	Traditional CNN	CNN
2	Fine-tuned CNN	CNN
3	Traditional CNN (from Method 1)	SVM
4	Fine-tuned CNN (from Method 2)	SVM

- RMSprop optimizer
- Categorical cross entropy loss
- Learning rate = 0.0001
- Batch size (train) = 32
- Batch size (validation) = 16
- Number of epoch = 20









Flowchart of Method 1 and Method 2

RESULTS (METHOD 1)

CNN Model	Number of Neuron	Number of Trainable Parameter	Elapsed Time (hh:mm:ss)	Validation Accuracy (%)
MobileNet	(64, 8)	3,273,114	02:00:24	86.40
	(96, 24)	3,307,754	02:01:14	86.93
	(128, 36)	3,342,894	02:02:53	86.43
VGG-19	(64, 8)	21,630,618	02:01:22	84.65
	(96, 24)	22,435,306	02:02:08	82.13
	(128, 36)	23,240,494	02:03:07	83.91
ResNet50_V2	(64, 8)	29,942,490	02:06:42	84.88
	(96, 24)	33,155,626	02:07:36	86.54
	(128, 36)	36,369,262	02:08:02	86.72

RESULTS (METHOD 2)

CNN Model and Number of Neuron	Number of Trainable Layer	Number of Trainable Parameter	Elapsed Time (hh:mm:ss)	Validation Accuracy (%)
MobileNet (96, 24)	10	1,689,002	01:10:02	88.77
	20	1,963,434	01:10:32	90.78
	30	2,496,426	01:11:36	91.82
VGG-19 (64, 8)	5	11,045,466	01:10:43	91.79
	8	15,765,082	01:11:36	88.53
	10	19,305,050	01:12:26	87.20
ResNet50_V2 (128, 36)	20	20,729,582	01:13:54	92.68
	30	27,293,422	01:14:12	91.35
	50	29,202,158	01:15:07	89.72

RESULTS (METHOD 3 & METHOD 4)

CNN Model	Number of Neuron	Number of Feature		Elapsed Time (mm:ss)	Validation Accuracy (%)
		Train	Validation		
MobileNet	(96, 24)	1,574,688	323,904	01:15	87.05
VGG-19	(64, 8)	1,049,792	215,936	02:14	84.91
ResNet50_V2	(128, 36)	2,099,584	431,872	02:21	86.84

CNN Model and Number of Neuron	Number of Trainable Layer	Number of Feature		Elapsed Time (mm:ss)	Validation Accuracy (%)
		Train	Validation		
MobileNet (96, 24)	10	1,574,688	323,904	01:08	92.77
VGG-19 (64, 8)	5	1,049,792	215,936	02:11	90.57
ResNet50_V2 (128, 36)	50	2,099,584	431,872	02:15	92.80

DISCUSSION OF FINDINGS (1)

- Number of neuron & number of trainable layer – number of trainable parameter
- SVM classifier: Number of neuron in layer used as input to classifier – number of extracted feature
- Number of trainable parameter – elapsed time
- Number of trainable parameter is less with transfer learning (Method 1 > Method 2)
- No particular pattern for hyperparameter (number of neuron, number of trainable layer) manipulation

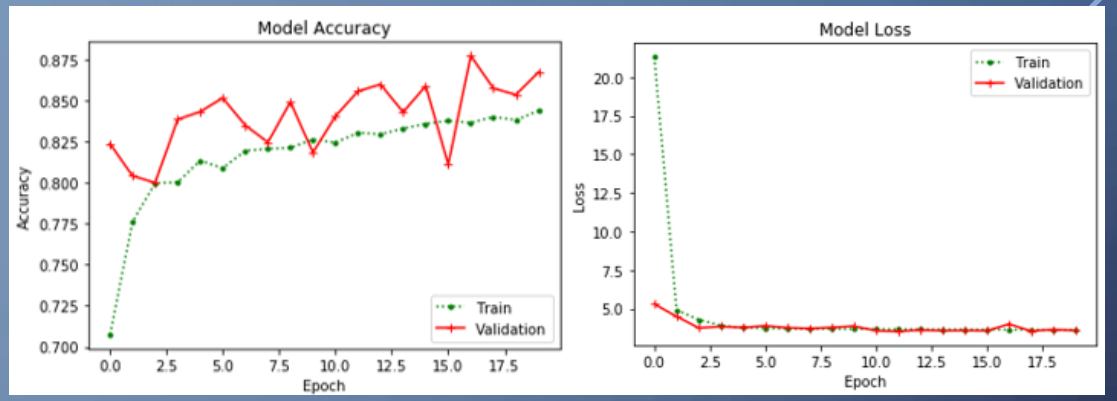
$$\text{Number of parameter} = \text{output size} \times (\text{input size} + 1)$$

where output size = number of neuron specified in the densely-connected layer,

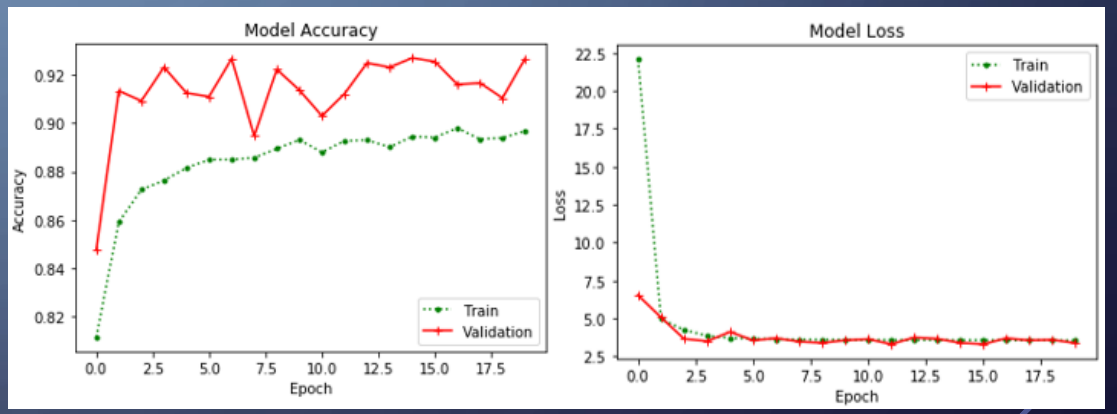
input size = output size of the previous layer

DISCUSSION OF FINDINGS (2)

- Transfer learning
- Effectiveness affected by relationship between source task and target task
- Fluctuation of validation accuracies
- SVM > CNN as classifier



ResNet50_V2, (128, 36) Model Accuracy and Loss Graphs [Method 1]



ResNet50_V2 (20 Layers) Model Accuracy and Loss Graphs [Method 2]

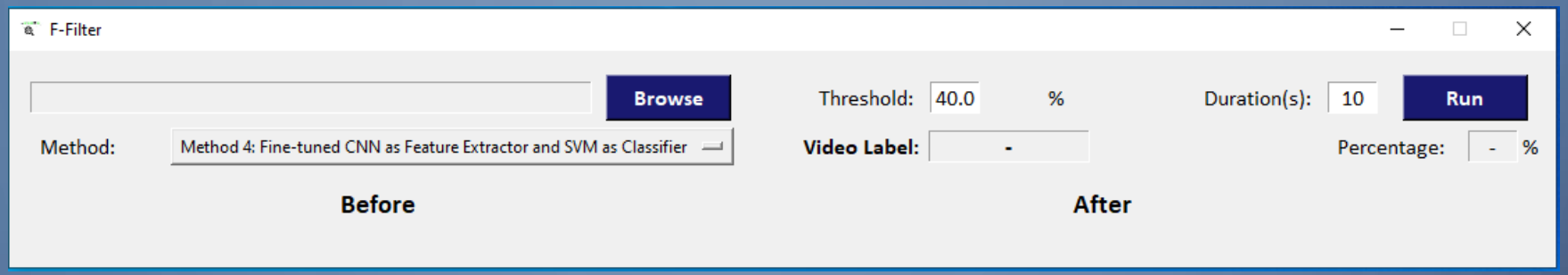
DISCUSSION OF FINDINGS (3)

Method	CNN Model
1	MobileNet with (96, 24) neurons
2	ResNet50_V2 with (128, 36) neurons and last 20 layers in convolutional base trainable
3	MobileNet with (96, 24) neurons
4	ResNet50_V2 with (128, 36) neurons and last 20 layers in convolutional base trainable

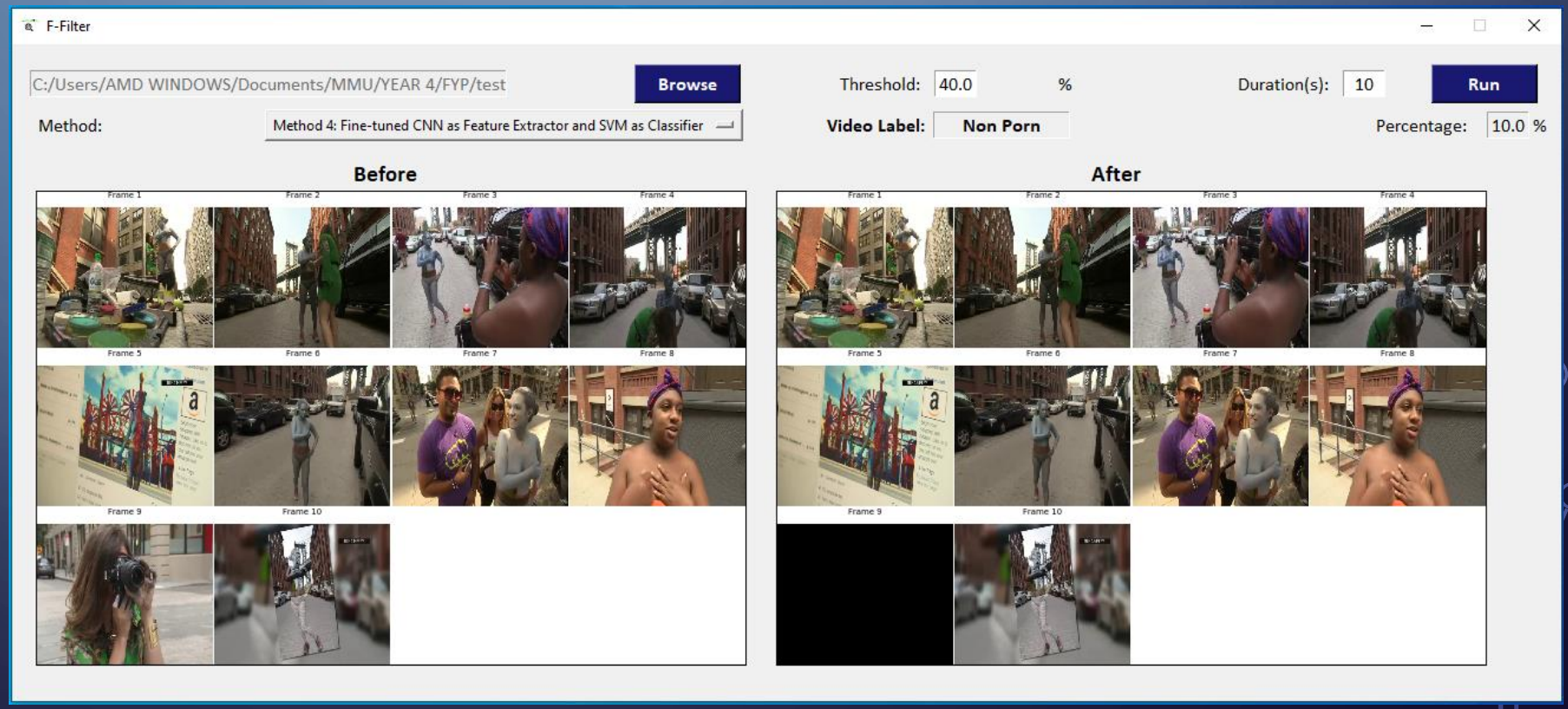
Automated Detection Achieved!

IMPLEMENTED GUI (F-FILTER)

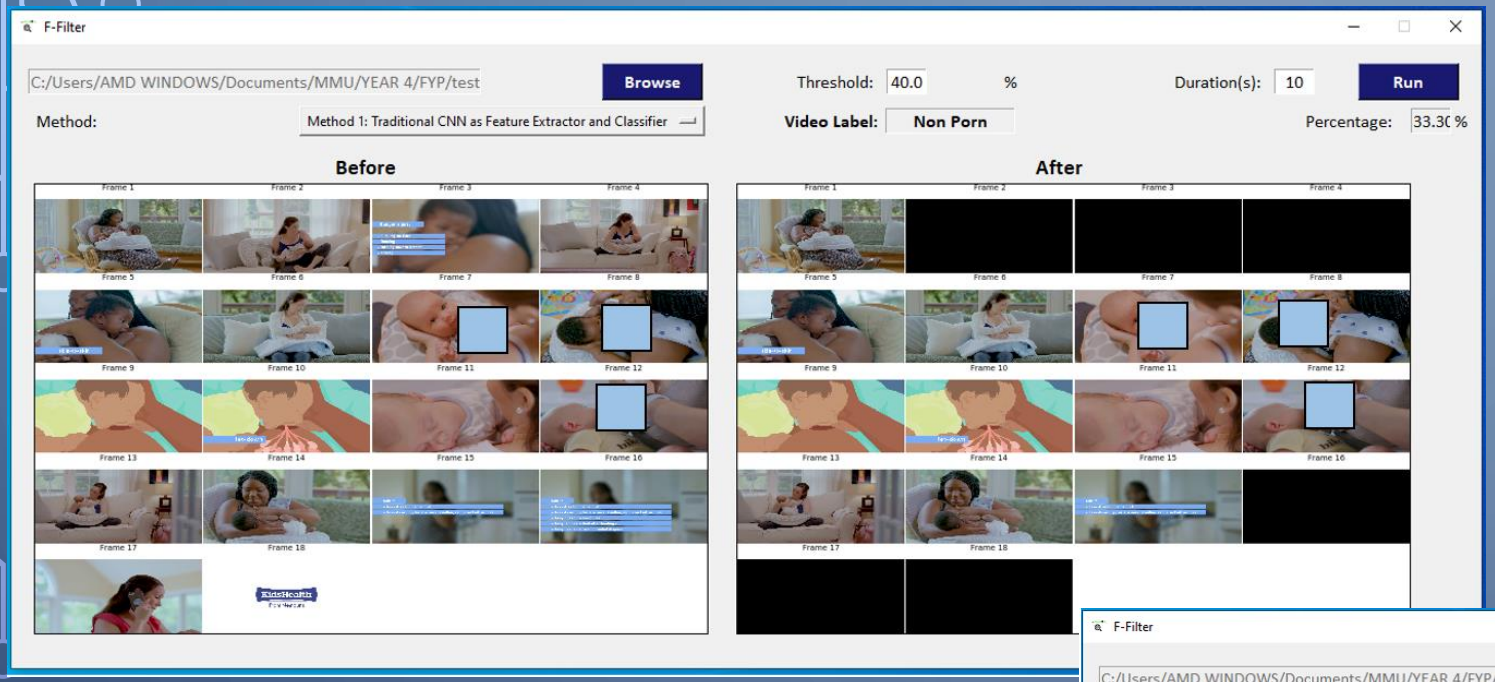
Upon Launching



After a successful run

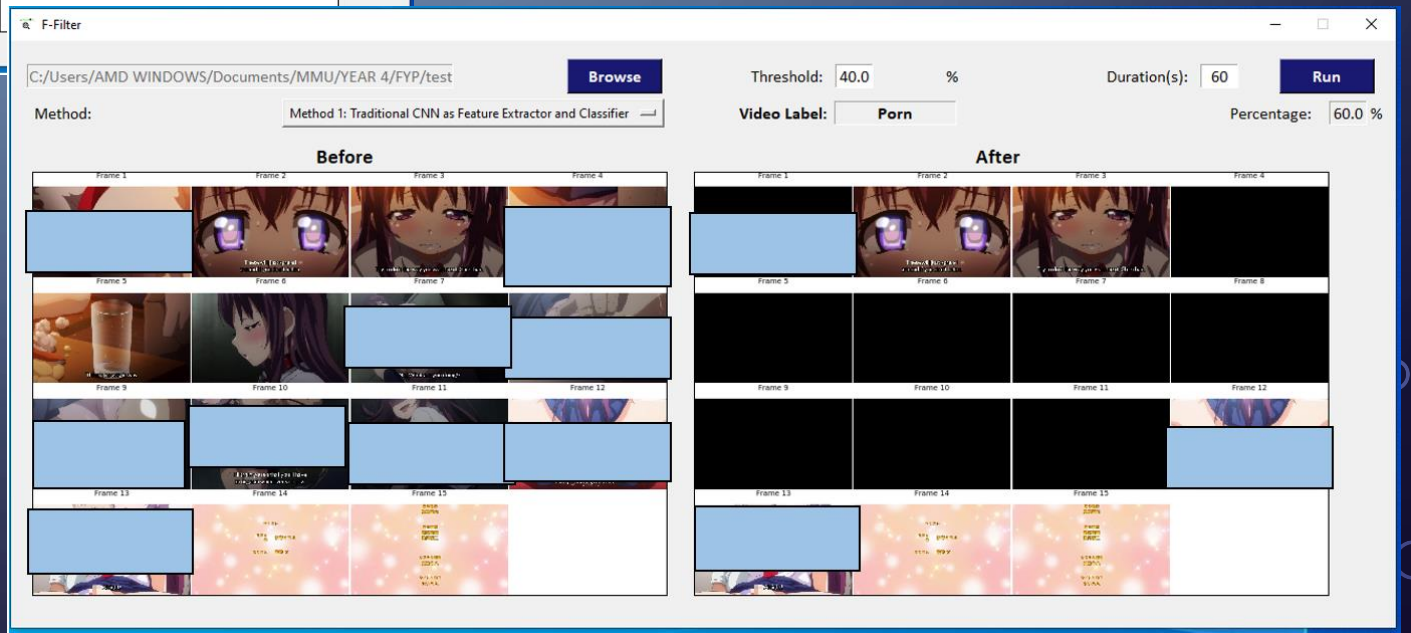


IMPLEMENTED GUI (F-FILTER) – METHOD 1

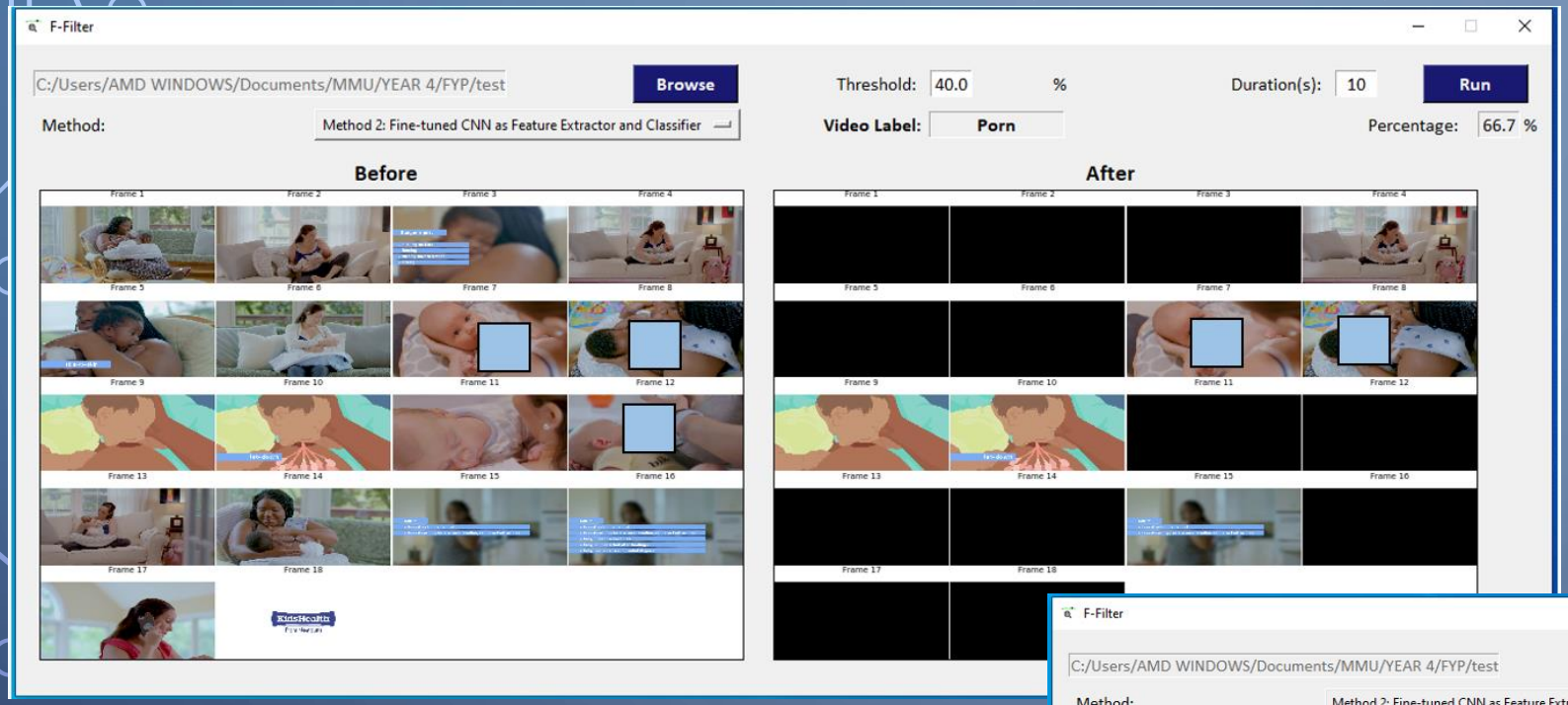


Breastfeeding video

Pornographic anime

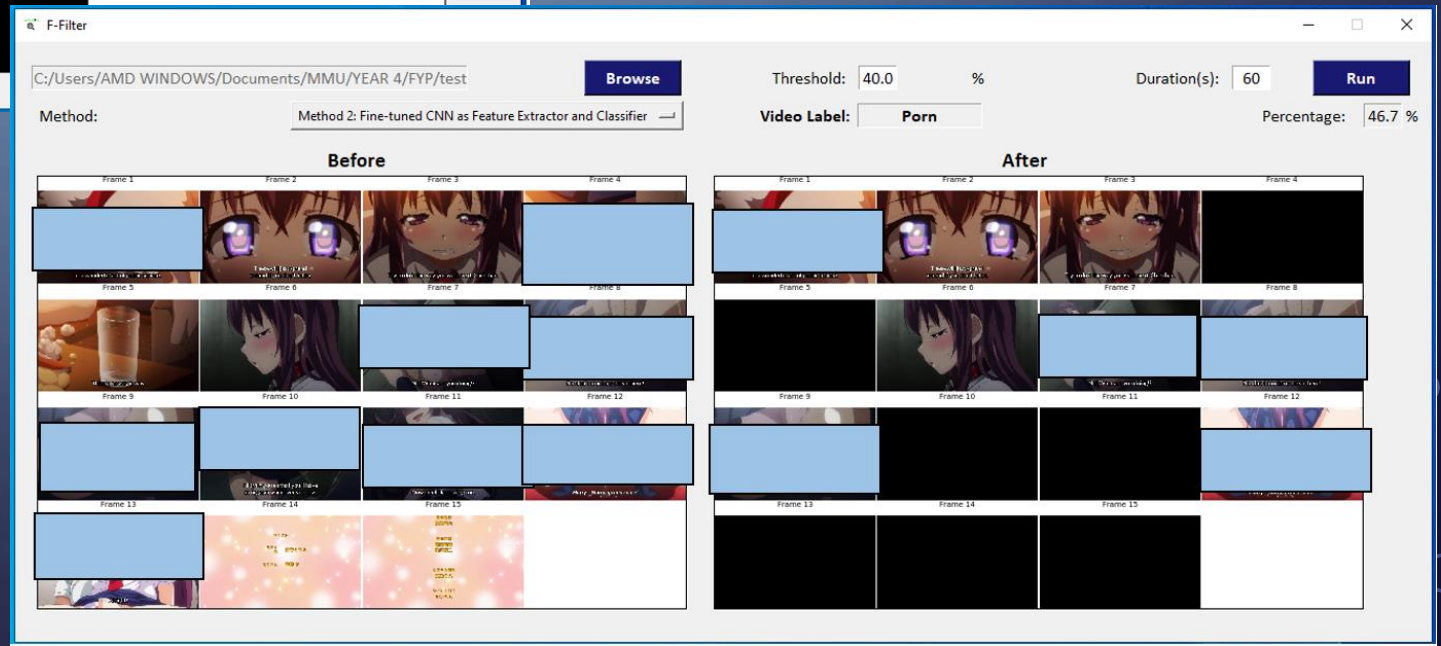


IMPLEMENTED GUI (F-FILTER) – METHOD 2

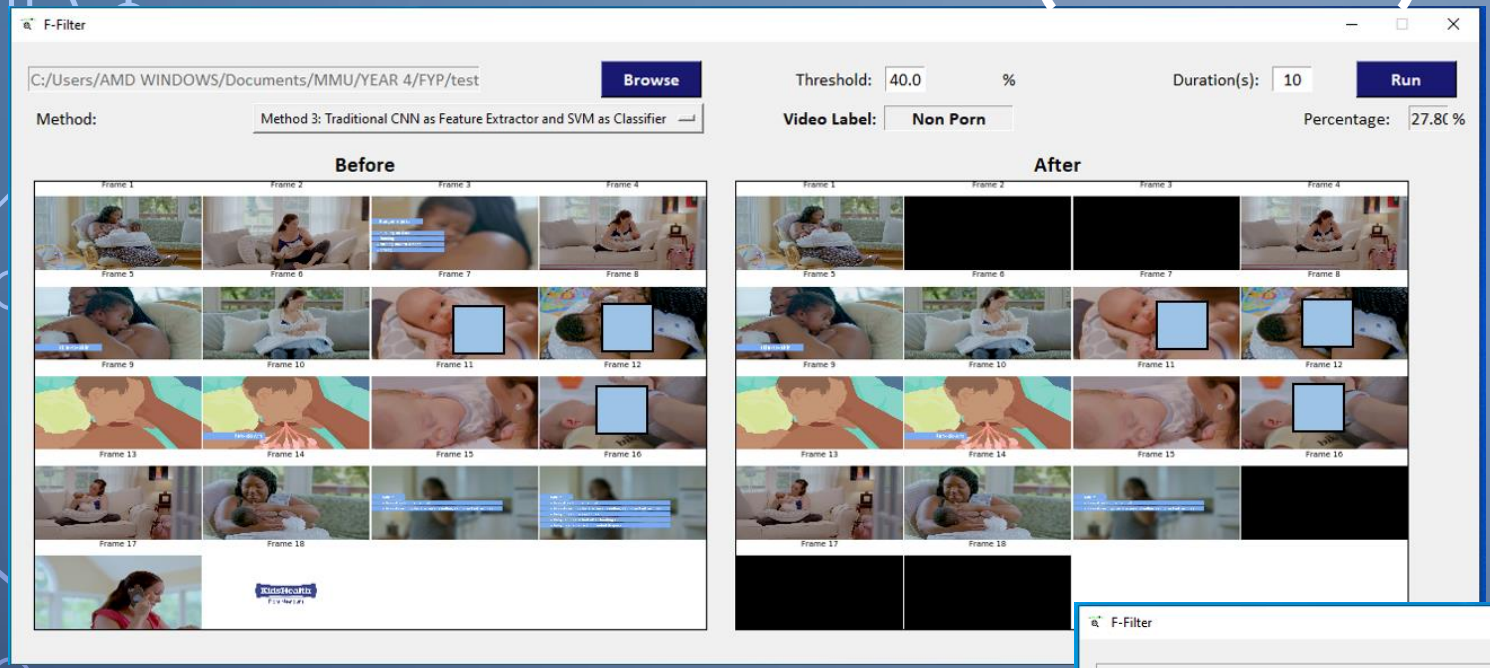


Breastfeeding video

Pornographic anime

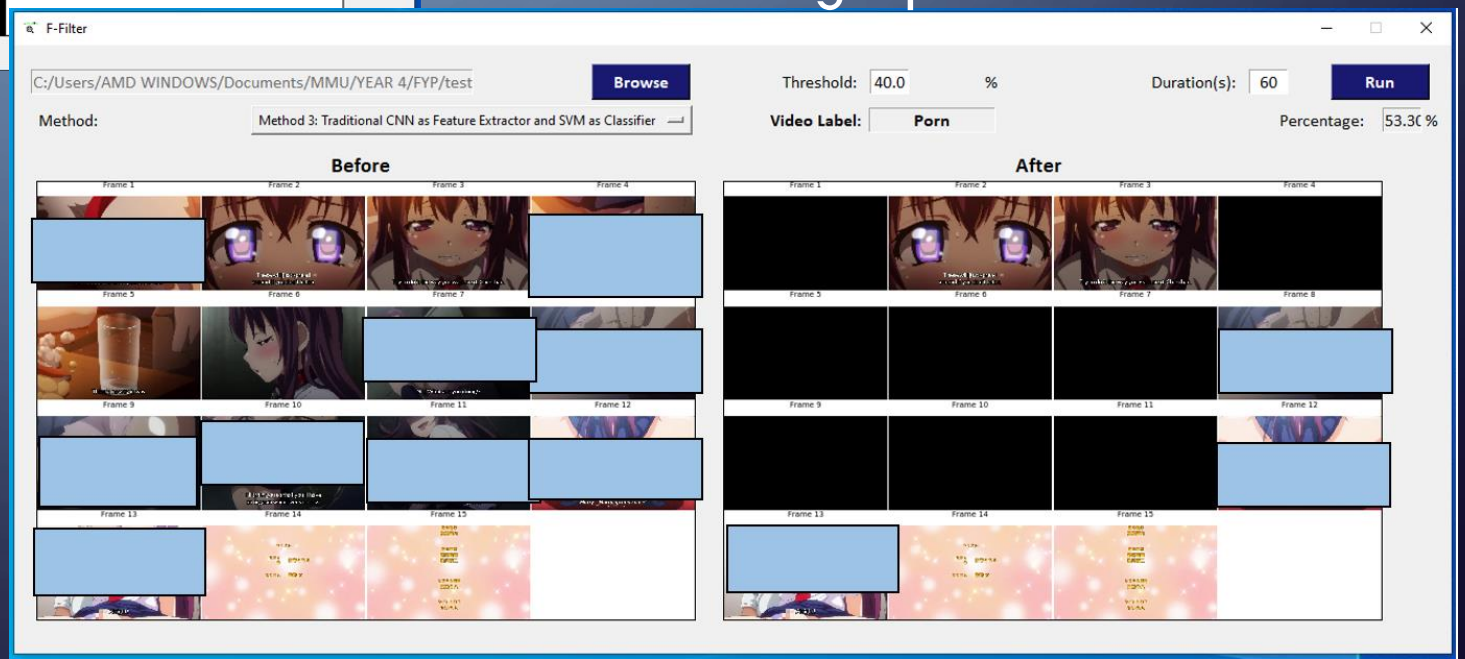


IMPLEMENTED GUI (F-FILTER) – METHOD 3

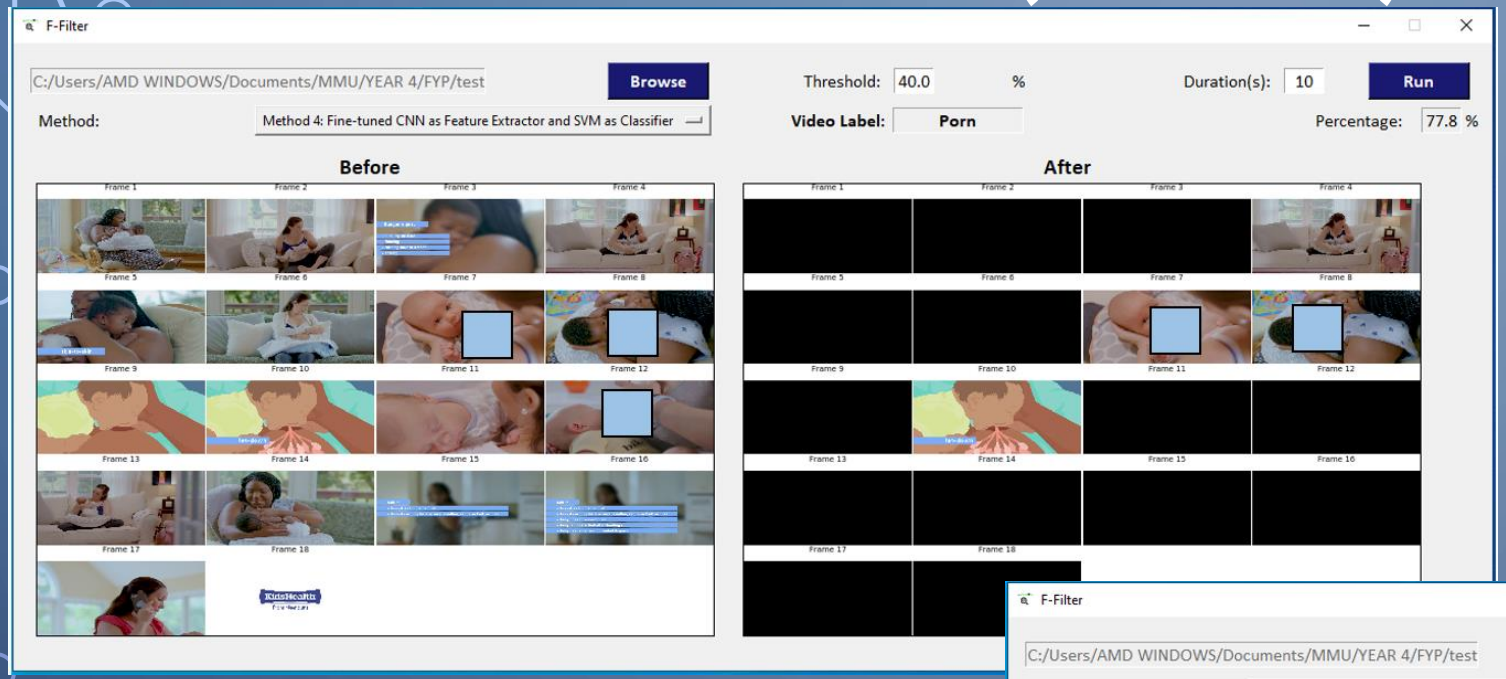


Breastfeeding video

Pornographic anime

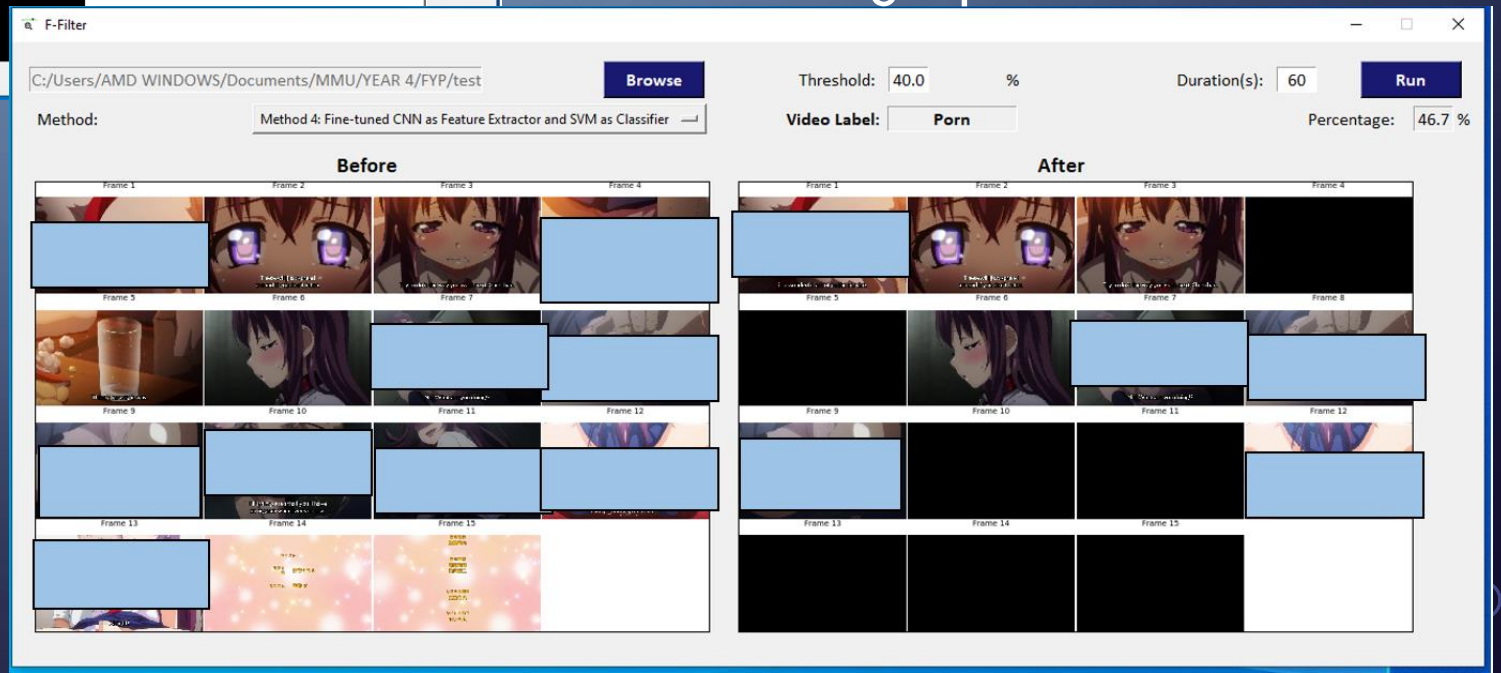


IMPLEMENTED GUI (F-FILTER) – METHOD 4



Breastfeeding video

Pornographic anime



CONCLUSION

- Best method: Method 4 [using ResNet50_V2, (128, 36), 20 layers trainable] with validation accuracy of 92.80%
- Deep learning via CNN can be used for automated detection of inappropriate visual content in films
- Transfer learning can improve CNN performance
- SVM classifier performed better than CNN classifier

FUTURE WORK

- Hyperparameter tuning or model optimization
- Trying other optimisers
- LSTM (RNN) which allows sequence learning [10]
- Adjustments to training data [6]
- Incorporate filtering process