Automated Nudity Recognition using Very Deep Residual Learning Network

Rasoul Banaeeyan, Hezerul Abdul Karim, Haris Lye, Mohamad Faizal Ahmad Fauzi, Sarina Mansor, John See

Abstract*: The exponentially growing number of pornographic material has brought many challenges to the modern daily life, particularly where children and minors have unlimited access to the internet. In Malaysia, all local and foreign films should obtain the suitability approval before distribution or public viewing, and this process of screening visual contents of all the TV channels imposes a huge censorship cost to the service providers such as Unifi TV. To leverage this issue, this paper proposes to use an emerging model of Deep Learning (DL) techniques called Residual Learning Convolutional Neural Networks (ResNet), in order to automate the process of nudity detection in visual contents. The pre-trained ResNet model, with hundred and one layers, was utilized to perform transfer learning and solve a new binary classification problem of nudity versus non-nudity. The performance of the proposed model is evaluated based on a newly created dataset comprising more than 4k samples of nudity and non-nudity images. After conducting experiments on the nudity dataset, the deep learning method succeeded to achieve the best performance of 70.42% in term of F-score, 84.04% in term of accuracy, and 93.72% in term of AUC .

Index Terms*: convolutional neural network, deep learning, nudity recognition, residual learning block.

I. INTRODUCTION

Filtering inappropriate visual contents from the different sources (internet TV, web pages, etc.) is a primary concern in different environments such as schools, homes or workplaces. In Malaysia, according to Ministry of Home Affairs – Security Collective Responsibility, all the TV channel providers are expected to obtain suitability approval before granting access to their subscribers or public users. One part of suitability assessment involves nudity identification which most of the time imposes a huge censorship cost to the service providers by means of recruiting large number of manpower constantly working over the months.

The main purpose of this research is to automate this tedious and laborious task of nudity detection by proposing to exploit the power of deep learning (DL) techniques as in [1]. More particularly, it is proposed to employ a certain DL model named Convolutional Neural Networks (CNN) [2] which have recently achieved the best performances in all visual recognition tasks (classification, segmentation, detection, localization, etc.) including pornography and adult content recognition as reported in [3]–[6].

Although there are several attempts in the literature in order to address the problem of nudity detection, there is variation in the definition of the word "Nudity" in our work with those previous scholarly works [3], [7]–[11]. For instance, a women wearing a bikini is considered a regular content in USA or Europe, but it is defined as adult content in Malaysia (and even other countries like Indonesia, or Brunei). In our paper, we adhere to the definition of the "Nudity" as determined by the Ministry of Home Affairs, Malaysia. We also created a new dataset of nudity images which best reflect the nudity samples targeted in this research.

The paper is organized as follows. The next section (II) briefly overviews the recent similar works in the domains of sensitive/adult/pornography detection in the visual contents, it is followed by Section III which presents the design framework of the proposed nudity detection as well as the details of the architectural design of the CNN models employed in this study. Section IV details the experimental setup and procedures followed in our research to facilitate the reproducibility of the results. In Section V, results of the different experiments are presented and discussed; this is followed by Section VI which concludes the paper and states some possible future directions.

II. II. RELATED WORK

Conventionally, it was common to employ traditional feature descriptors such as LBP (local binary patterns) [12], SIFT (scale invariant feature transform) [13], or HOG (histogram of oriented gradients) [14] to obtain local or global image descriptors and later use them to differentiate among images with sensitive content or normal contents.

For instance, the work in [11] attempted to detect images containing nude/pornography scene by proposing to use a new variation of SIFT called Hue-SIFT to extract global image features, along with a Bag of Feature (BoF) model to obtain a global representation the image. Later the authors showed the same recognition rate as reported by those skin-based approaches is possible without detection of skin or shapes in the nudity scenes.

In another paper [15], authors approached the pornography detection

Revised Manuscript Received on August 18, 2019.

Rasoul Banaeeyan, Hezerul Abdul Karim, Haris Lye, Mohamad Faizal Ahmad Fauzi, Sarina Mansor, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia.

John See, Faculty of Computing Informatics, Multimedia University, Cyberjaya, Malaysia.

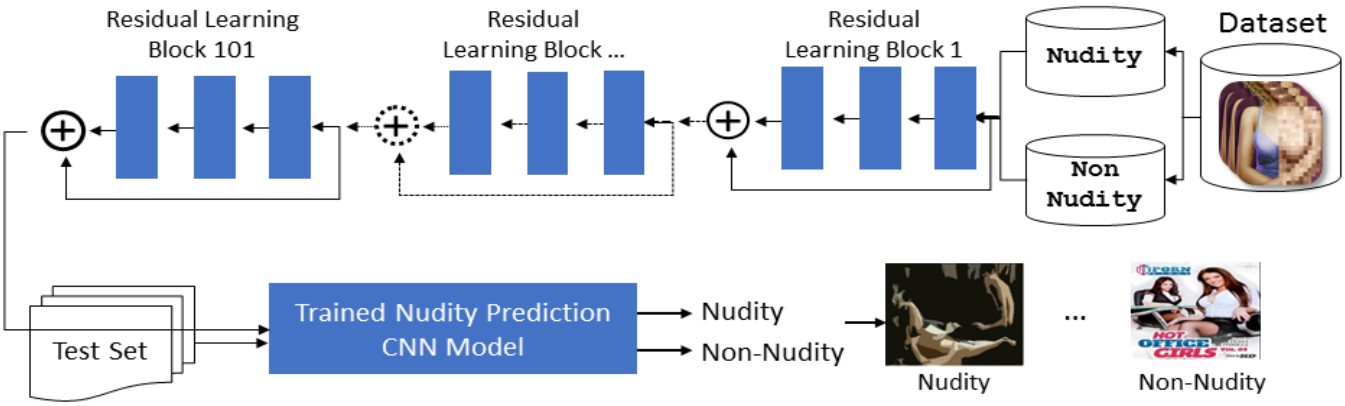

Figure 1 Presentation of the overall system design for the nudity and non-nudity detection. Sample images are first categorized into two sets of nudity and non-nudity and are used as input to the series of residual learning blocks for the purpose of abstract representation extraction. Later these image descriptors are used as input to a softmax classifier to train a binary nudity predictor model

problem by proposing to use high-level semantic features. They optimized the BoF model in order to close the gap between low-level scene features and those of the high-level by fusing the contextual information in the visual words and spatial features of the pornographic images. Later in [16], it was suggested to use the regions of interest (RoI) to address the issue of inaccurate pornography detection assuming that this task is similar to object detection and the human visual system uses visual attention model to tackle such tasks. Their proposed framework included four stage of (1) skin region detection, (2) visual saliency map construction, (3) detection of pornographic regions based on threshold segmentation, and (4) extraction of features such as color, texture, intensity, and skin. The research in [17] was the first attempt towards utilizing the mid-level image descriptors for solving the pornography detection task in videos. More particularly, the authors proposed to use a novel video frame descriptor which uses local binary patterns together with BossaNova, a powerful mid-level image representation as described in [18]. BossaNova again was used in [19] for the same task. The BoF model was used in [20] where authors presented to use a multi-instance modeling scheme based on spatial pyramid partitions (SPP) to transfer the target problem (pornography detection) into MIL problem.

Temporal Robust Features (TRoF) were for the first time employed in the task of pornography detection videos by the work in [7]. It was proposed to aggregate the TRoF features into mid-level features by using FV (Fisher vector), a new variation of BoVW (bag of visual words) model.

Since the emergence of Convolutional Neural Networks, starting with revolutionary work presented in [2], all previous performances of hand-crafted feature descriptors have been significantly enhanced, and these low-level descriptors are not in the focus of the researchers and practitioners anymore. The survey in [21], overviews the works in the domain of adult content recognition which mostly utilize hand-crafted features. The more interested reader may find the survey in [22] very informative where the authors comparatively reviewed the different local feature extraction algorithms in the domain of online pornography detection.

Recently, after the emergence of Deep Learning techniques, all aspects of visual recognition (detection, localization, classification, etc.) are significantly enhanced in all domains such as biometric [23], bioinformatics [24], etc.

The field of pornography content recognition has not been containing pornography contents.

Another research in [10], proposed to use a probability model based on uncertain inferencing along with an ensemble of CNN-based classifiers for the task of adult content recognition in still images. Finally, authors in [3] proposed to use a pre-trained CNN model called ResNet-50 [26] in order to detect the sensitive pornography contents in images.

III. METHOD

The overall system design of the proposed nudity detection is depicted in Fig.1.

After construction of the new nudity dataset, dataset images are categorized into two groups of nudity and non-nudity. Later these classes of images are used as input to the series of residual learning blocks designed to obtain the abstract and high-level presentation of the images. The system uses these image representations as input into a binary classifier model for the training purpose. Finally, the testing images are classified based on the probability values outputted by the trained CNN model.

Deep learning [1] models are comprised of multiple computational layers which enable the learning of abstract representation of input data in and end-to-end and iterative manner. Owing to the remarkable power of Convolutional Neural Networks (CNN) – a branch of DL techniques – state-of-the-art performances are recently obtained in all tasks of visual recognition such as classification, detection, recognition, or segmentation. CNNs [2] are designed to automatically identify and detect the patterns in the visual contents as well as speech (or any type of 2D, 3D, and 4D data), and have outperformed the human accuracy and speed in many recognition tasks. Some exemplary CNN models are VGG-Net [27], Google-Net [28], Inception [29], and ResNet [26].

In this research, we propose to use a variation of ResNet comprising hundred and one layers of learning blocks which is recently achieved the best performance in other visual recognition tasks such as object recognition in [30]. The

design of the learning blocks is detailed in sub-section C. The key difference between our employed version of ResNet and the one used in [3] is the presence of additional sixty-seven layers which makes the model more resistance to the problem of vanishing gradients while facilitate extraction of more abstract features through large number of layers.

A. Construction of the Nudity Dataset

We used a third-party software to automatically search and download the samples of nudity and non-nudity from Google image website. To this purpose, two search terms were used as "Nude Body" and "People" to respectively collected instances of the full body nudity and non-nudity. In total, a collection of 4,380 samples were obtained. After filtering the dataset, out of all, 1470 instances represent the class of nudity and the remaining 2910 instances represent the class of non-nudity.

Sample images of the nudity dataset are presented in Fig.2.

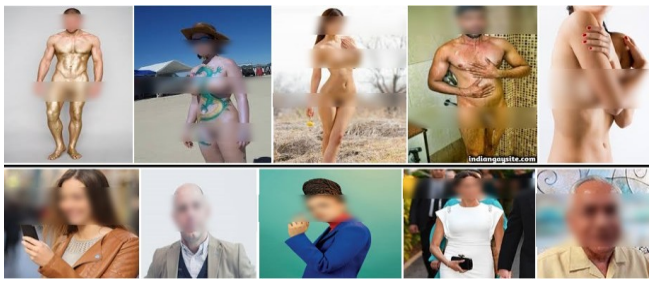

Figure 2 Presentation of the sample nudity and non-nudity images. The top row presents the positive (nudity) instances while the bottom row presents negative instances (non-nudity).

B. Layers in Deep Learning

A typical CNN model is comprised of different computational layers some with learnable parameters, and some solely mathematical operations with different purposes. In the following, these layers and their functions are explained.

- Convolutional Layer: performs a set of randomly initialized 2D convolutions in order to increase the depth of the input image or the previous layer.
- Batch Normalization: speeds up the training process by normalizing channels of the input image or previous layer.
- Rectified Linear Unit (ReLU): performs thresholding on all channels of the input image or previous layer to add non-linearity.
- Max Pooling Layer: applies down-sizing on the input image or previous layer and generates an output of the decreased width and height but same depth.
- Fully Connected Layer (FC): multiplies all the values of the input image or previous layer by its weights.
- Softmax Layer: assigns the values in an FC layer to their corresponding probability values.

C. The architecture of Residual Networks

The ResNet model [26] employs the residual learning block (Fig.3) aiming for solving the issue of vanishing gradients and therefore decreases the error rate in the case of considerably deeper CNN models (a large number of computational layers). Residual Learning blocks were designed to enhance the prediction by reference to the input layers rather than unreferenced functions as depicted in Fig.3. In our experiment, we took advantage of the one-hundred-one-layer architectural design of the ResNet.

The architectural design of the pre-trained ResNet 101 used in our study is detailed as follows:

- Convolution [(1×1 (64) + 3×3 (64) + 1×1 (256)] ×3
- Convolution [(1×1 (128) + 3×3 (128) + 1×1 (512)] ×4
- Convolution [(1×1 (256) + 3×3 (256) + 1×1 (1024)] ×23
- Convolution [(1×1 (512) + 3×3 (512) + 1×1 (2048)] ×3
- Fully Connected Layer
- Softmax Layer

The models consists of 3 replications of a block of 3-layer convolutional operations, followed by 4 replication of 3-layer convolutional operations, followed by 23 replications of another 3-layer block of convolutions, and finally 3 replications of the last 3-layer convolution operations. In total there are 99 layers ending with one fully connected and one softmax layer (101 layers).

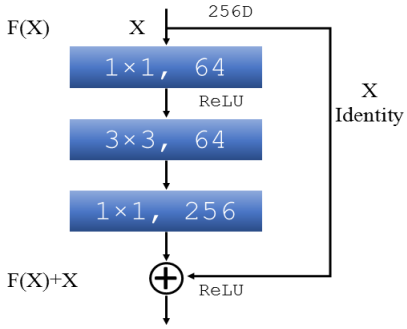

Figure 3 Design of the Residual Learning block used in the architecture of ResNet 101.

D. Training Hyper Parameters

For the training, the optimizer was set to SGDM (stochastic gradient descent momentum), the learning rate was set to 0.001 for early convergence and avoiding the problem of vanishing gradients, the training set was shuffled one before starting the new epoch, the values for the momentum and weight decay were both set to 0.9. The loss function in the classification layer was set to Cross Entropy

IV. EXPERIMENTS

This section details the experimental procedure, the choice of the datasets, performance metrics as well as the system specifications.

A. Training, Validation, and Testing Partitions

To train the proposed CNN model, the dataset was randomly divided into three parts for training, validating, and testing the performance, each partition containing 60%

(2,628), 10% (438), and 30% (1,314) of the images in the collection.

B. Performance Metric

Three different evaluation metrics were used to assess the performance of the proposed binary classification models (nudity vs. non-nudity) as *Accuracy*, *AUC*, *F-score* formulated in (1), (2), (3), and (4).

In addition, a ROC performance curve was also generated to provide more insight into the behavior of the classifier with respect to different prediction threshold values (ranging from 0 to 1). AUC is the area under the ROC curve indicating the performance of a binary classifier.

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

$$F\text{-}score = 2\times(Precision\times Recall)/(Precision+Recall) \quad (2)$$

$$Precision = TP/(TP+FP) \quad (3)$$

$$Recall = TP/(TP+FN) \quad (4)$$

Where TP is the number of true positive samples (correctly classified as nudity), TN is the number of true negatives (correctly classified as non-nudity), FP is the number of false positives (incorrectly classified as nudity), and FN is the number of false negatives (incorrectly classified as non-nudity).

C. Software and Hardware

All the implementations were carried out using Matlab R2018b (academic version), Image Processing Toolbox, Computer Vision Toolbox, and Statistics and Deep Learning Toolbox. The experiments were conducted on a desktop computer with Windows 7 (64-bit), 16 GB RAM, Intel Core i7-4790 CPU @ 3.60 GHz, and an NVIDIA GPU with 4 GB internal memory (GTX 749).

D. Method of Comparison

The results of our research are compared with a very recent and similar work presented in [3]. The authors of this work employed an earlier version of the ResNet called ResNet 34 comprising thirty-four blocks of residual learning. Therefore, we have replicated their model on our dataset for the purpose of comparison.

V. RESULTS AND DISCUSSION

This section presents the experimental results on the newly created nudity dataset; it is followed by a discussion on results. The ROC curve of the proposed model is presented by Fig.4. As it can be observed, the ResNet 101 achieved a much better prediction performance of 93.72 in term of AUC while its performance is, at all possible false positive rate, better than the other method.

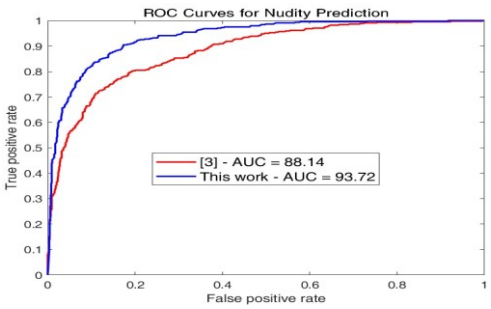

Figure 4 Performance curves (ROC) of the proposed nudity prediction model and the previous work presented in [3].

Table I. depicts the confusion matrices for the prediction results of both the proposed method as well as the one used for comparison [3].

Table I Confusion matrix corresponding to the results of the proposed nudity prediction model (on the top), and the work in [3] (on the bottom).

Total: 1,314		**Predicted**	
		Nude	Not Nude
Actual	Nude	320	121
	Not Nude	105	768

Total: 1,314		**Predicted**	
		Nude	Not Nude
Actual	Nude	250	191
	Not Nude	19	854

From the ROC curves, it is quite clear that the proposed ResNet with 101 layers succeeded to improve the AUC by 5.58 units indicating the distinctive prediction power of this model over the previous one with 50 layers.

The confusion matrices also confirm the superior performance of the ResNet 101 in term of accuracy where it is in enhanced from 82.80% to 84.04 (by 1.24). However, with respect to the F-score, it can be seen that the previous ResNet 34 [3] still claims the higher value at 73.90 while the proposed one achieved a value of 70.42. This difference in values can be explained by the fact that the previous method obtained better precision as opposed to the proposed one (precision=72.56), while the new model was achieved the best recall value (recall=92.93).

The detailed results of our work are presented in Table II. with respect to different metrics along with those of the work in [3]. Overall, out proposed model achieved three out of the five best performances with respect to Recall, Accuracy, and AUC respectively at 92.93%, 84.04%, and 93.72%. More importantly, one can see a 5.58% improvements in term of AUC which considered a more appropriate indicator of performance in the task of binary classification.

Table II Presentation of the results in this work with regard to different evaluation metrics and compared to [3].

Metric	Method in [3]	This Work
Precision	**72.56**%	56.69%
Recall	75.29%	**92.93**%
F-score	**73.90**%	70.42%
Accuracy	82.80%	**84.04**%
AUC	88.14%	**93.72**%

VI. CONCLUSION AND FUTURE WORKS

In this research, we proposed to use a recent successful architecture of convolutional neural networks called ResNet for the task of binary classification of nudity contents (nudity vs. non-nudity). We constructed a new dataset of nudity images comprising more than 4k instances of nudity and non-nudity conforming to the specific definition of nudity according to Ministry of Home Affairs, Malaysia which is referred to during visual content censorship of TV channels.

After conducting experiments, an enhancement in the performance of the nudity classifier was obtained by the ResNet 101 CNN model and state-of-the-art results were achieved in terms of Recall, Accuracy and AUC.

In the future work, we intend to extend our dataset to include more aspects of nudity instances (partial nudity, complicated nudity, etc.) as well as other classes of sensitive contents such as different types of pornography acts. In addition, future attempts may benefit from CUDA-enabled implementation in order to enable deployment of the deep learning models on embedded platforms for high-speed utilization of nudity detection in high-resolution (SD/HD/FHD) video frames.

ACKNOWLEDGMENT

This research was fully funded by TM R&D, Malaysia.

REFERENCES

[1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[2] A. Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*, vol. 4, no. 4, pp. 253–262, 2012.
[3] A. Nurhadiyatna, S. Cahyadi, F. Damatraseta, and Y. Rianto, "Adult content classification through deep convolution neural network," *Proc. - 2017 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Comput. Sci. Eng. IC3INA 2017*, vol. 2018-Janua, pp. 106–110, 2018.
[4] X. Jin, Y. Wang, and X. Tan, "Pornographic Image Recognition via Weighted Multiple Instance Learning," *IEEE Trans. Cybern.*, vol. PP, pp. 1–9, 2018.
[5] 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.
[6] K. Zhou, L. Zhuo, Z. Geng, J. Zhang, and X. G. Li, "Convolutional neural networks based pornographic image classification," *Proc. - 2016 IEEE 2nd Int. Conf. Multimed. Big Data, BigMM 2016*, pp. 206–209, 2016.
[7] D. Moreira *et al.*, "Pornography classification: The hidden clues in video space–time," *Forensic Sci. Int.*, vol. 268, pp. 46–61, 2016.
[8] M. D. More, D. M. Souza, and R. C. Barros, "Seamless Nudity Censorship : an Image-to-Image Translation Approach based on Adversarial Training," *IEEE Int. Jt. Conf. Neural Networks*, 2018.
[9] A. P. B. Lopes, S. E. F. De Avila, A. N. A. Peixoto, R. S. Oliveira, M. D. M. Coelho, and A. D. A. Araújo, "Nude detection in video using bag-of-visual-features," *Proc. SIBGRAPI 2009 - 22nd Brazilian Symp. Comput. Graph. Image Process.*, pp. 224–231, 2009.
[10] R. Shen, F. Zou, J. Song, K. Yan, and K. Zhou, "EFUI: An ensemble framework using uncertain inference for pornographic image recognition," *Neurocomputing*, vol. 322, pp. 166–176, 2018.
[11] A. P. B. Lopes, S. E. F. De Avila, A. N. A. Peixoto, R. S. Oliveira, and A. De A. Araújo, "A bag-of-features approach based on Hue-SIFT descriptor for nude detection," *Eur. Signal Process. Conf.*, no. Eusipco, pp. 1552–1556, 2009.
[12] W. Zhou, A. Ahrary, and S. I. Kamata, "Image description with local patterns: An application to face recognition," *IEICE Trans. Inf. Syst.*, vol. E95-D, no. 5, pp. 1494–1505, 2012.
[13] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, pp. 1150–1157 vol.2, 1999.
[14] N. Dalal, B. Triggs, and D. Europe, "Histograms of Oriented Gradients for Human Detection," 2005.
[15] L. Lv, C. Zhao, H. Lv, J. Shang, Y. Yang, and J. Wang, "Pornographic images detection using high-level semantic features," *Proc. - 2011 7th Int. Conf. Nat. Comput. ICNC 2011*, vol. 2, pp. 1015–1018, 2011.
[16] J. Zhang, L. Sui, L. Zhuo, Z. Li, and Y. Yang, "An approach of bag-of-words based on visual attention model for pornographic images recognition in compressed domain," *Neurocomputing*, vol. 110, no. July 2012, pp. 145–152, 2013.
[17] C. Caetano, S. Avila, S. Guimar, and A. D. A. Ara, "Pornography Detection using BOSSANOVA Video Descripor," pp. 2–6, 2014.
[18] S. Avila, N. Thome, M. Cord, E. Valle, and A. De A. Araújo, "Pooling in image representation: The visual codeword point of view," *Comput. Vis. Image Underst.*, vol. 117, no. 5, pp. 453–465, 2013.
[19] C. Caetano, S. Avila, W. R. Schwartz, S. J. F. Guimarães, and A. de A. Araújo, "A mid-level video representation based on binary descriptors: A case study for pornography detection," *Neurocomputing*, vol. 213, pp. 102–114, 2016.
[20] D. Li, N. Li, J. Wang, and T. Zhu, "Pornographic images recognition based on spatial pyramid partition and multi-instance ensemble learning," *Knowledge-Based Syst.*, vol. 84, pp. 214–223, 2015.
[21] C. X. Ries and R. Lienhart, "A survey on visual adult image recognition," *Multimed. Tools Appl.*, vol. 69, no. 3, pp. 661–688, 2014.
[22] Z. Geng, L. Zhuo, J. Zhang, and X. Li, "A comparative study of local feature extraction algorithms for Web pornographic image recognition," *Proc. 2015 IEEE Int. Conf. Prog. Informatics Comput. PIC 2015*, pp. 87–92, 2016.
[23] R. Nejad, Elaheh Mahraban and Affendey, Lilly Suriani and Latip, Rohaya Binti and Ishak, Iskandar Bin and Banaeeyan, "Transferred Semantic Scores for Scalable Retrieval of Histopathological Breast Cancer Images," pp. 1–8, 2018.
[24] R. Banaeeyan, H. Lye, M. F. Ahmad Fauzi, H. Abdul Karim, and J. See, "Semantic facial scores and compact deep transferred descriptors for scalable face image retrieval," *Neurocomputing*, 2018.
[25] K. Fernandes, J. S. Cardoso, and B. S. Astrup, "A deep learning approach for the forensic evaluation of sexual assault," *Pattern Anal. Appl.*, vol. 21, no. 3, pp. 629–640, 2018.
[26] R. G. Crane, "Deep Residual Learning for Image Recognition 2015," no. (ed.), Oxford, U.K., Pergamon Press PLC, 1989, Section 3, pp.111-120. (ISBN 0-08-036148-X), pp. 1–9, 1989.
[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Int. Conf. Learn. Represent.*, pp. 1–14, 2015.
[28] G. Smoluk, "Going deeper with convolutions," *Mod. Plast.*, vol. 57, no. 3, pp. 62–63, 2015.
[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015.
[30] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

AUTHORS PROFILE

Rasoul Banaeeyan completed his B.Eng (Computer Engineering) in University of Applied Sciences, Iran in 2010, and then his M.Sc. (Computer Science) and Ph.D. in University Putra, Malaysia in 2013 and Multimedia University in 2018, respectively. Formerly, he was working as a researcher at Faculty of Computer Science and Information Technology, University of Malaya in 2014. His areas of interest lie at the intersection of deep learning, image processing, and scalable computations.

Hezerul Abdul Karim obtained his B.Eng. degree in Electronics with Communications from University of Wales Swansea, UK, in 1998 and M. Eng Science degree from Multimedia University, Malaysia in 2003. He obtained his PhD degree from Center for Communication Systems Research (CCSR), University of Surrey, UK in 2008. He is an Associate Professor at Multimedia University since November 2015. He is also Deputy Dean of Student Affair and Alumni at Faculty of Engineering, Multimedia University. He has been teaching multimedia and computing engineering subjects. His research interests include telemetry, 2D/3D image/video coding and transmission, error resilience and multiple

description video coding, and deep learning in image and video. He is currently supervising and co-supervising several postgraduate students. He is Senior Member of IEEE and currently the Vice Chair of IEEE Signal Processing Society, Malaysia Section.

Haris Lye received the B.Eng (Electrical and Electronics) degree from University of Science Malaysia in 1996, and the M.S. (Information Technology) degree from Multimedia University in 2004. His research interests include deep learning, computer vision, and artificial intelligence. Currently, he is a lecturer at the Faculty of Engineering, Multimedia University, Malaysia.

Mohammad Faizal Ahmad Fauzi (MIEM, SMIEEE) received his B.Eng. in Electrical and Electronic Engineering degree from Imperial College London in 1999 and his Ph.D. in Electronics and Computer Science degree from the University of Southampton in 2004. He is currently an Associate Professor at the Faculty of Engineering, while also serving as the Deputy Director for the Collaboration and Innovation Center (CIC), at Multimedia University, Malaysia. Mohammad Faizal has published more than 80 journal and conference articles to date. His main research interests are in the area of signal and image processing, pattern recognition, computer vision and biomedical informatics. He has delivered keynote and invited speeches at several international conferences such as DPCA2016 (Kuala Lumpur, Malaysia), AMS2017 (Kota Kinabalu, Malaysia), ICCCSP2018 (Chennai, India), ICCSN2018 (Chengdu, China), and DPCA2019 (Tokyo, Japan). He is also a recipient of several awards such as the Fulbright Award, MCMC Senior Researcher Award and the TM GCEO Merit Award.

John See is currently Senior Lecturer with Multimedia University, Malaysia, and Chair of the Centre of Visual Computing (CVC) at the Faculty of Computing and Informatics. From 2018, he is also Visiting Research Fellow at Shanghai Jiao Tong University, China. He received his B.Eng., M.EngSc. and Ph.D degrees from Multimedia University, Malaysia. His current research interests span the areas of computer vision and pattern recognition in general, including emerging fields of video surveillance, affective computing particularly facial micro-expressions and computational image aesthetic and style. He has published over 70 articles in international journals and conferences. He is also currently the Associate Editor of EURASIP Journal of Image and Video Processing and IEEE Access. He is the Program Chair for the 21st IEEE MMSP, has co-chaired numerous workshops and challenges, and regularly serves in the TPC of reputable conferences (ICCV, ACCV, ACM MM, ICIP, ICME, FG). He is a Senior Member of IEEE.

Sarina Mansor is a Senior Lecturer in the Faculty of Engineering at Multimedia University (MMU) Malaysia. She is currently the Program Coordinator of B.Eng. (Hons) Electronics Majoring in Computer. She received her B.Eng (Hons) Electronic and Electrical from University of College London in 1998 and M.EngSc degree from Multimedia University in 2002. She completed her DPhil in Biomedical Engineering from University of Oxford (UK) in 2009. Her research interests are Image Processing, Medical Imaging, Computer Vision, Machine Learning and Cloud Computing.