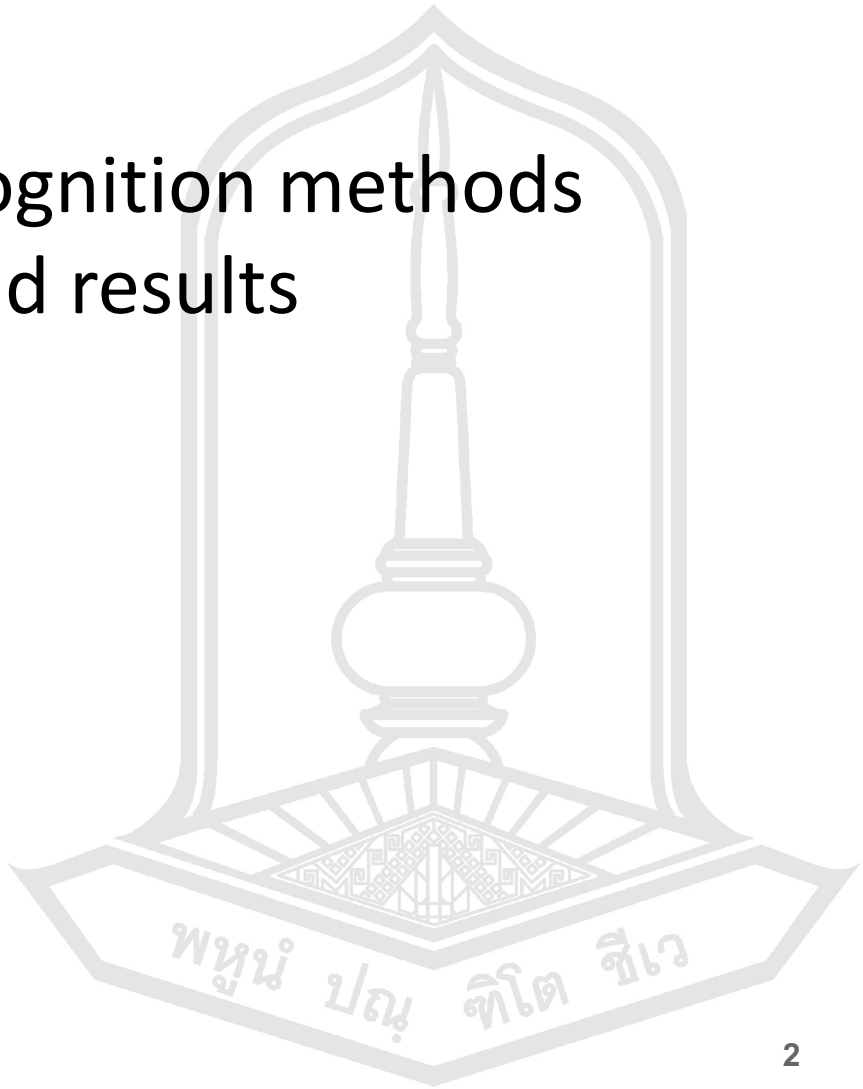
RECOGNIZING PORNOGRAPHIC IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORKS

Olarik Surinta and Thananchai Khamket
Faculty of Informatics
Mahasarakham University

Present at the 4th International Conference on Digital Arts, Media and Technology (ICDAMT2019)
January 30 – February 2, 2019

Outline:

- Introduction
- Pornographic image recognition methods
- Experimental settings and results
- Conclusion



Introduction:

- In **pornographic image recognition**, image processing and machine learning techniques are proposed to use.
- Due to **the image processing techniques**,
  - the **human skin** is extracted from the whole image.
  - The RGB is converted into **HSV and YCbCr color spaces** to extract the skin color.
  - The whole image region is calculated and decided as the pornographic image when the ratio is more than **the threshold value**.

Introduction:

- For **the machine learning technique**,
  - First, the color image is converted into **HSV, YCbCr color space** to **extract skin area**.
  - Then, **extracted the feature** from the skin area.
  - Finally, the **machine learning technique** such as SVM and MLP are used to create a model and classify.

Introduction:

- Rattanee and Chiracharit (2016) Nudity detection based on face color and body morphology

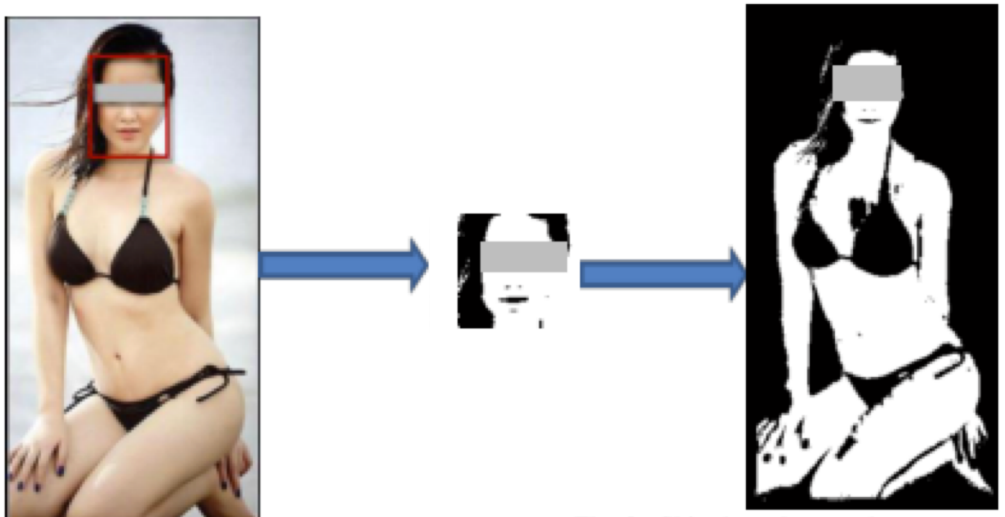


Fig. 2. Face detection result

Fig. 3. Skin detection result

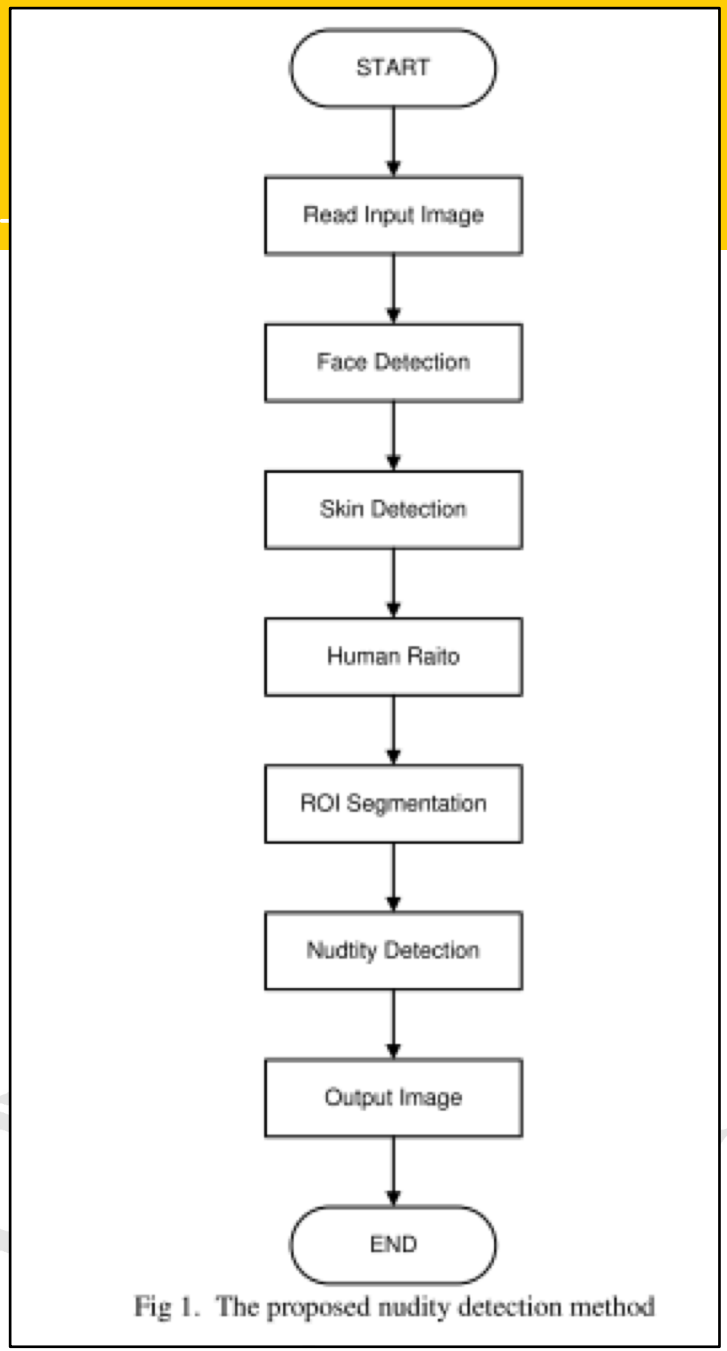


Fig 1. The proposed nudity detection method

Introduction:

- Wijaya, et al. (2015) Pornographic image recognition based on skin probability and Eiganporn of skin ROIs images

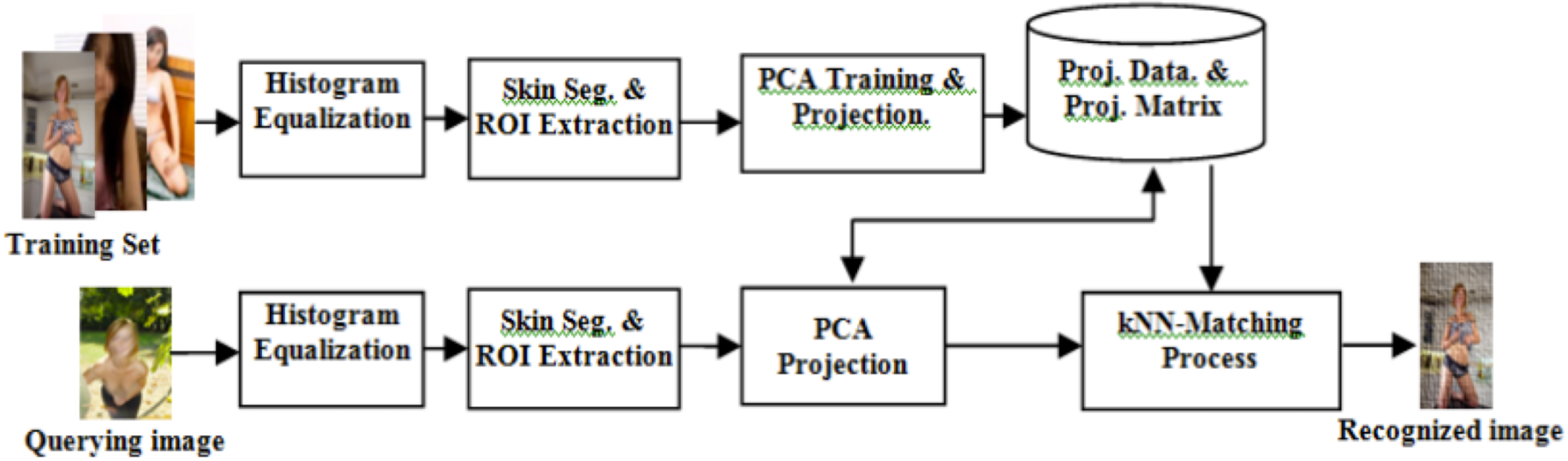


Figure 1. Pornographic image recognition diagram block

Introduction:

- Wijaya, et al. (2015) Phonographic image recognition using fusion of scale invariant descriptor

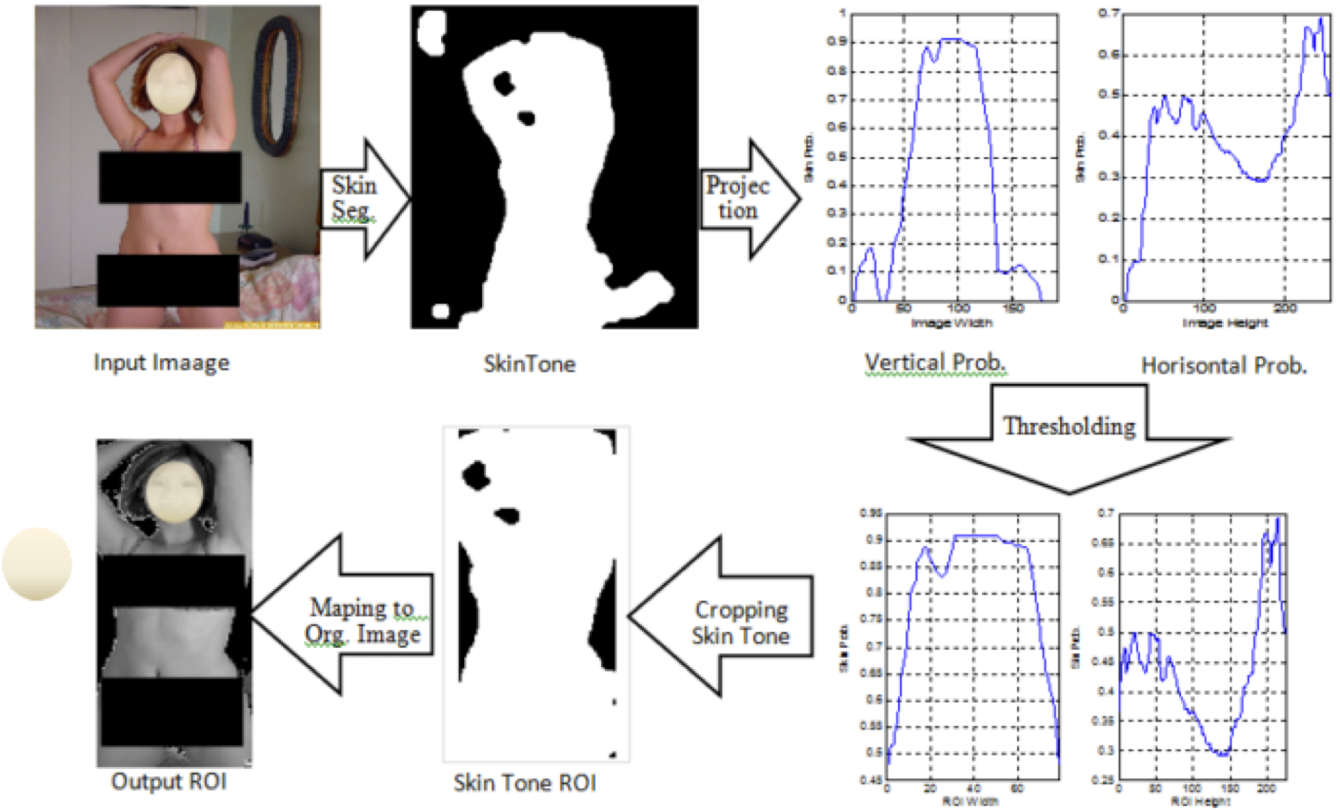


Figure 3. The ROI image extraction

Introduction:

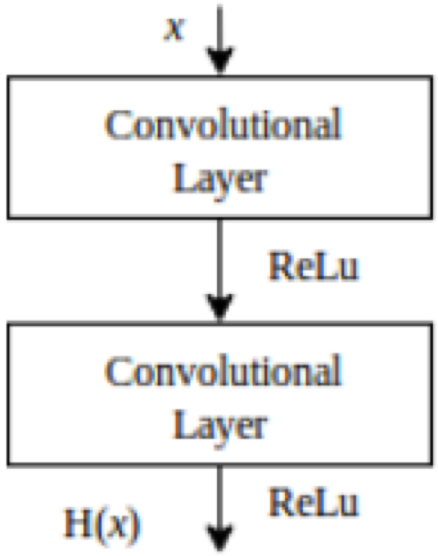
- We evaluate the performance of **16 different techniques** on a **TI-UNRAM pornographic image dataset**.
- The use of existing deep **CNN architectures** (ResNet, GoogLeNet, and AlexNet) and a **BOW method** are presented.
- This paper is combining **three well-known local descriptor methods**, called *LBP, HOG, and SIFT* and **three machine learning technique** (*SVM, MLP, and KNN*).

Pornographic image recognition methods:

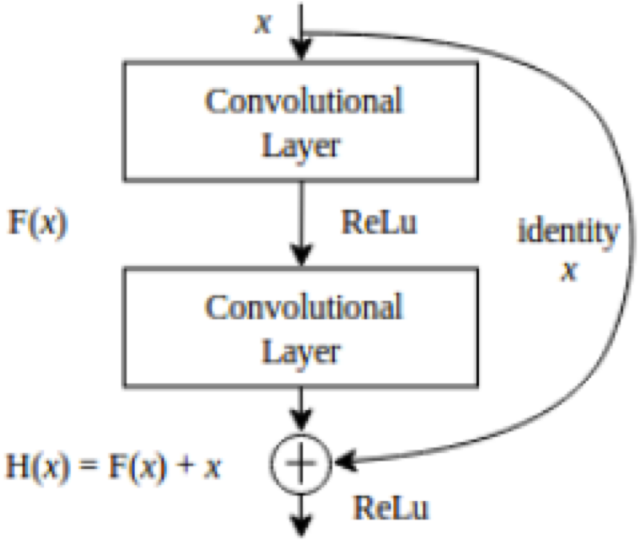
- **Deep Residual Networks (ResNet)**
  - ResNet architecture has **very deep network** and shown good performance in many image recognition.
  - *He et al.* proposed the deep ResNet architecture with a depth of 18, 34, 50, 101, and 152 layers.
  - The **ResNet-152** is deeper 22 and 7 times than AlexNet and GoogLeNet, respectively.

Pornographic image recognition methods:

- The novel architecture called ***shortcut connections***, is proposed.
- The shortcut directly uses the input of the previous layer to the next output.



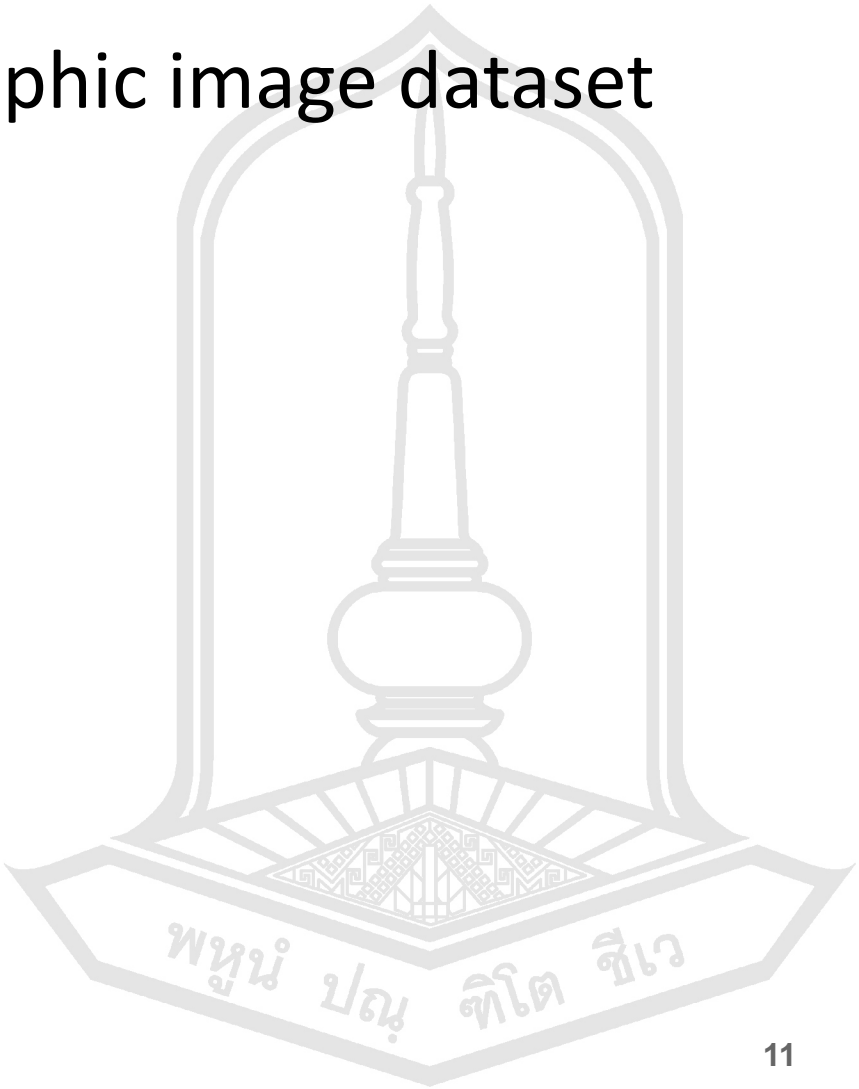
Plain network



Residual network

Experimental settings and results:

- The TI-UNRAM pornographic image dataset
- Experimental setup
- Experimental results



TI-UNRAM dataset:

- This dataset includes two classes and contains 685 pornographic, 715 non-pornographic images (**1400 images**)
- These images are collected from the Internet
- We randomly **divided 50%** of the whole dataset into training and test set

Non-pornographic images:

Complex images:

Can you guess which images are pornographic?







Experimental setup:

- We use **2-fold cross validation** according to Wijaya et al. (2015a, 2015b).
- We compute **the average and standard deviation** for evaluating the test performance of
  - deep CNN architectures
  - Local descriptors combined with machine learning techniques
  - bag of words (BOW)

Experimental results:

Recognition results using deep CNN methods

Deep CNN Methods	Layer	Test Accuracy (%)
ResNet	50	**88.00± 0.37**
GoogLeNet	22	87.20± 0.18
AlexNet	8	86.10± 0.35
LeNet	5	85.90± 0.04

Experimental results:

Recognition results using different local descriptors and machine learning techniques

Methods	Test Accuracy (%)
LBP+SVM	**87.80± 0.13**
HOG+SVM	78.00± 0.02
SIFT+SVM	78.00± 0.01
LBP+MLP	85.80± 0.30
HOG+MLP	75.87± 0.01
SIFT+MLP	74.28± 0.02
HOG+BOW	80.71± 0.34
BOW	79.00± 0.21
LBP+KNN	73.50± 0.12
HOG+KNN	70.00± 0.01
SIFT+KNN	66.43± 0.02
FD+YCbCr [7]	83.97

Conclusion:

- We have presented a **comparative study** on the TI-UNRAM pornographic image dataset including
  - local descriptors combined with machine learning techniques
  - a bag of visual words (BOW)
  - deep convolutional neural networks (CNNs)

Conclusion:

- **First**, we proposed to use the **LBP, HOG, and SIFT** as for the local descriptor methods.
- These three descriptor methods combined with 3 machine learning techniques;
  - **SVM, MLP, and KNN**
- The results show that the ***LBP+SVM*** outperforms the other combinations.
- The LBP+SVM method also gives a better result than the BOW method.

Conclusion:

- **Second**, we compared **three deep CNN architectures**
  - ResNet, GoogLeNet, and AlexNet architectures
- To make a fair comparison, in these experiments, *the transfer learning* and *the data augmentation* **are not performed**.
- The results show that **the best recognition accuracy is the ResNet**, GoogLeNet, and AlexNet, respectively.

Conclusion:

- **Finally**, the *ResNet architecture* which is *the best result in our experiment*, also slightly higher than the *LBP+SVM*.
- Future work:
  - We want to improve the result of the deep CNN by using transfer learning and data augmentation.
  - We also consider the deep learning approach that requires less memory usage and a decrease in training computing time.

ICDAMT2019:

- Thank you for your kind attention.