Convolutional Neural Networks and Training Strategies for Skin Detection

Yoonsik Kim, Insung Hwang, and Nam Ik Cho

terryoo@ispl.snu.ac.kr

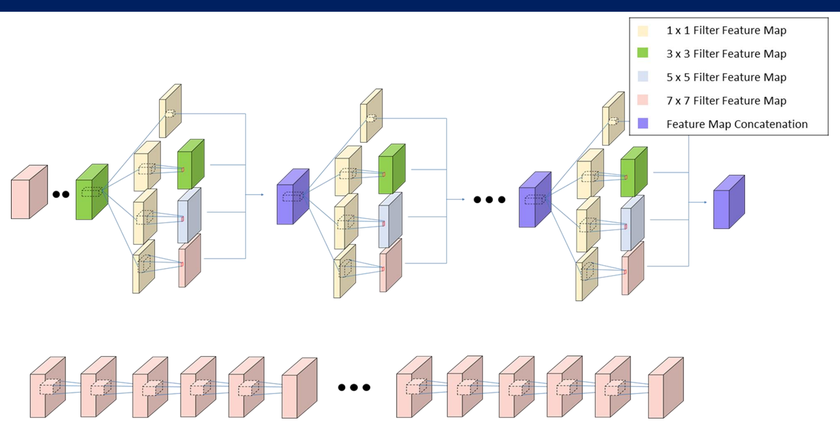
Dept. of Electrical and Computer Engineering, Seoul National University

Abstract This paper presents two convolutional neural networks (CNN) and their training strategies for skin detection. The first CNN is VGG based network and the other is network in network (NiN) architecture which is the modification of Inception Block. These CNNs are trained by patch based and image based methods which focus local features and overall shape respectively. Proposed CNNs yield better performance than the conventional methods and the other deep learning based method.

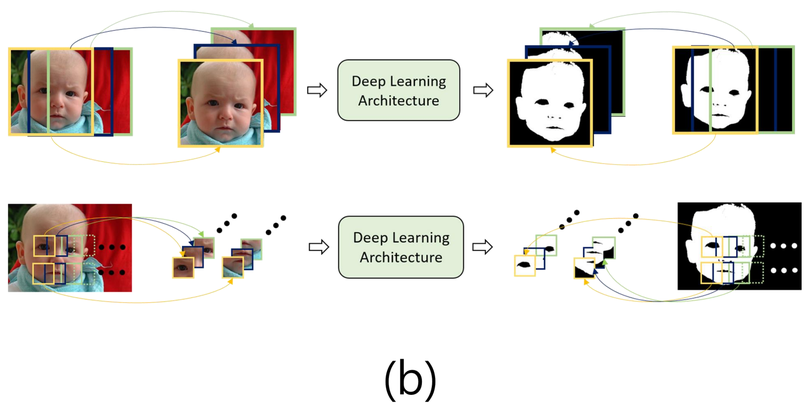
MOTIVATION

- ✓ Skin detection is to find skin pixels, which is one of the important pre-processing steps in many image processing and computer vision.
- ✓ Specifically, skin detection is used for image enhancement, face and human detection, gesture analysis, pornographic contents filtering, surveillance systems.
- ✓ Skin detection is considered a challenging problem due to diverse variations such as change of illumination, skin color variations of races and makeup, and skin-like backgrounds.

ALGORITHM



(a)



(b)

Figure 1. Framework of proposed method, (a) Proposed CNN architectures, (b) Proposed training strategies

- ✓ NiN architecture is composed of modified inception module[1] to adjust skin detection which is composed of 20 convolutional layer.
- ✓ The other architecture is VGG based network which consists of 3 x 3 kernels stacking 20 convolutional layer.
- ✓ Training strategies are also proposed, namely patch based method and image based method.
- ✓ In summary, four CNN approaches are proposed.

[1] Szegedy, Christian, et al. "Going deeper with convolutions", CVPR 2015

EXPERIMENTAL RESULTS

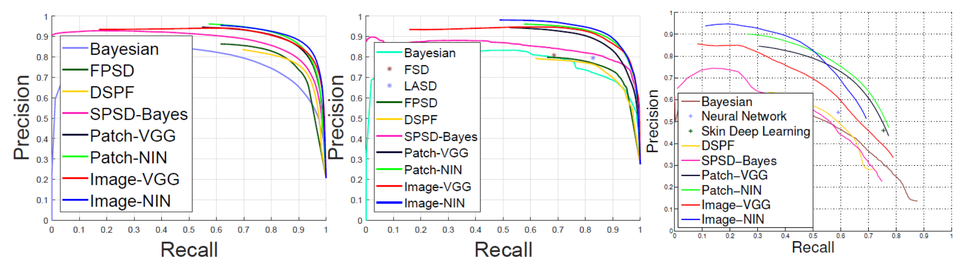


Figure 2. Comparison of ROC curves on ECU, Pratheepan and VT-AAST.

Methods	ECU dataset				Pratheepan dataset				VT-AAST dataset			
	Accuracy	*Precision*	*Recall*	*F-measure*	*Accuracy*	*Precision*	*Recall*	*F-measure*	*Accuracy*	*Precision*	*Recall*	*F-measure*
Bayesian [4]	0.8910	0.7292	0.8220	0.7728	0.8237	0.6881	0.8972	0.7788	0.8798	0.4857	0.5740	0.5262
Neural Network [24]	-	-	-	-	-	-	-	-	0.8856	0.5426	0.5907	0.5656
Skin Deep Learning [16]	-	-	-	-	-	-	-	-	0.8881	0.4605	**0.7538**	0.5717
FPSD [22]	0.9106	0.7948	0.8534	0.8231	0.8419	0.7387	0.8991	0.8070	0.8918	0.5333	0.5583	0.5455
SPSD [13]	0.9306	0.8085	0.8805	0.8430	0.8782	0.7659	**0.9328**	0.8412	0.8871	0.5136	0.5477	0.5301
Patch-VGG	0.9479	0.8577	0.8913	0.8742	0.9299	0.8563	0.8750	0.8655	0.9243	0.6850	0.6455	0.6647
Patch-NiN	0.9492	0.8696	0.8923	0.8808	0.9334	0.8802	0.8972	0.8886	**0.9272**	0.7001	0.6539	**0.6762**
Image-VGG	0.9486	0.8499	0.9037	0.8760	0.9313	0.8577	0.9069	0.8816	0.9103	0.6225	0.5798	0.6004
Image-NiN	**0.9562**	**0.8720**	**0.9122**	**0.8917**	**0.9484**	**0.9003**	0.8912	**0.8957**	0.9249	**0.7183**	0.5826	0.6434

Table 1. Evaluation on ECU, Pratheepan and VT-AAST datasets at peak F-measure.

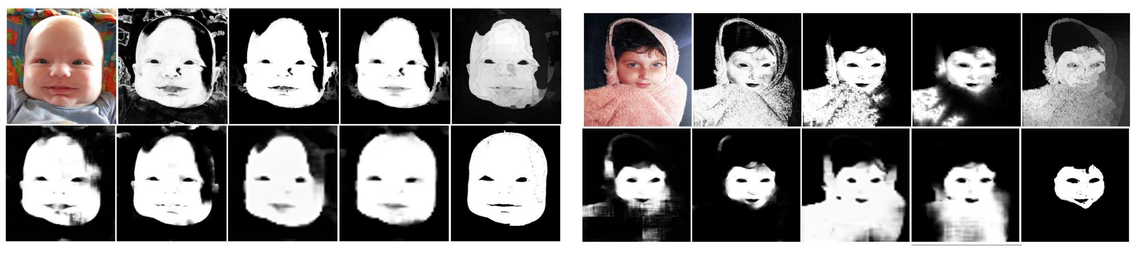


Figure 3. Visual comparison with other methods

- ✓ Proposed methods achieve higher quantitative performance than other conventional methods including graph and deep learning methods.
- ✓ Proposed NiN architecture with image based training method yields robust results on illumination variation input by learning a human shape information.
- ✓ Proposed patch based strategy yields more precise skin-maps by learning a skin texture information.

CONCLUSION

- ✓ We have proposed two deep neural network architectures: VGG and NiN based ones, and we have also proposed two training schemes: patch-based and whole-image-based methods.
- ✓ Proposed CNNs outperform the conventional methods that are based on graph representation, and auto-encoder based deep learning method.
- ✓ The NiN architecture generally works better than the VGG network for the skin detection, though the NiN needs less parameters than the VGG.
- ✓ The whole-image-based training finds the human shape features better so that it is robust to illumination and color variations
- ✓ The patch-based method finds skin texture very well so that it can reject the skin-colored background when it has different texture from the skin.