FASHION STYLE RECOGNITION USING COMPONENT-DEPENDENT CONVOLUTIONAL NEURAL NETWORKS

WA.PC.4

Takahisa Yamamoto*† and Atsushi Nakazawa†

*Fujitsu Laboratories Ltd., Japan †Kyoto University, Japan

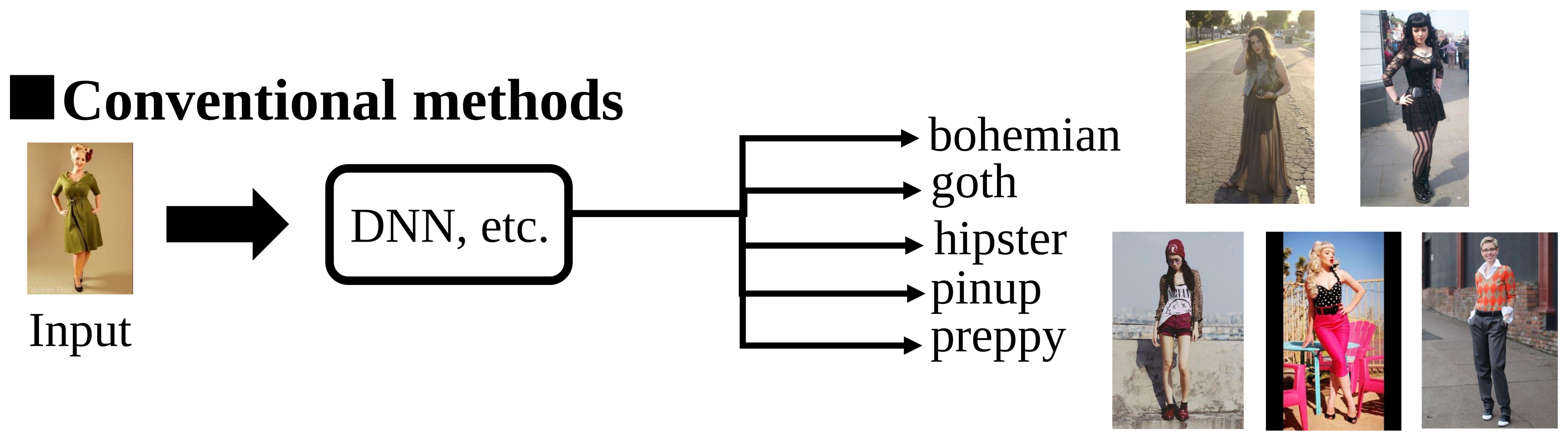
Introduction

■Motivation

The fashion style recognition is important technology in online marketing applications such as recommendations in accordance with customer preferences. Several algorithms have been proposed, but their accuracy is still unsatisfactory.

Our goal is to improve classification accuracy for fashion styles.

■Conventional methods



In contrast to the existing fine-grained recognition methods [1,2], our **component-dependent convolutional neural networks (CD-CNNs)** firstly segments an input image using semantic segmentation methods, then apply the part-specific CNNs.

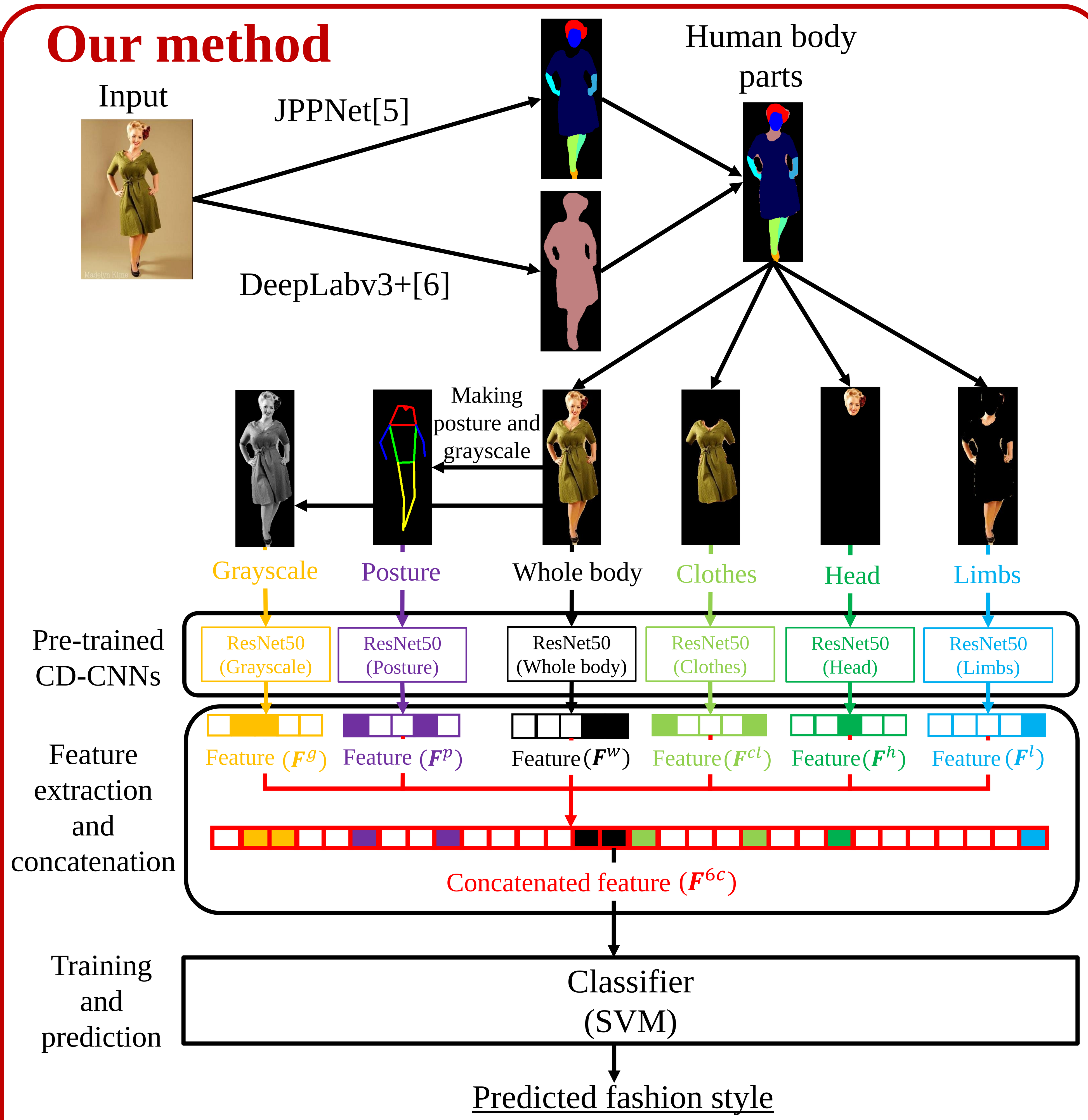
Related works

Fig.1 : FashionStyle14 dataset example images

Fig.2 : HipsterWars dataset example images

- Takagi et al. created FashionStyle14 dataset (14 class, Fig.1) and showed that ResNet50 was the best CNN architecture [3].
- Kiapour et al. created HipsterWars dataset (5 class, Fig.2) and developed the handcrafted classification algorithm [4].

Our method



・Extracting human area in a pixel unit by using JPPNet [5] and DeepLabv3+ [6].

・Creating 4 components images from the result of extracting human area.

・Assuming that pose and grayscale are important for fashion style classification, therefore, creating the images.

・Extracting feature vectors using a pre-trained ResNet50 for each component.

・Concatenating extracted feature vectors.

・5-fold cross validation to find the best SVM parameters.

Experimental results

■Table of best component combination for each used component number.

Number of components	Component image	dataset	naming
1	Whole body	Both	1-component
2	Whole body, Head	Both	2-component
3	Whole body, Clothes, Head	Both	3-component
4	Whole body, Clothes, Head, Grayscale	Both	4-component
5	Whole body, Clothes, Head, Posture, Grayscale	HipsterWars	5EL
5	Whole body, Clothes, Head, Limbs, Grayscale	FashionStyle14	5EP

■Experimental results for HipsterWars dataset

Method	Preprocessing	Feature	Accuracy
Kiapour et al. [4]	None	Hand craft	70.6
StyleNet [7]	None	CNN base	75.9
Nakajima et al. [8]	SSD and PSPNet	Pre-trained ResNet50	80.9
1-component	DeepLabv3+ and JPPNet	Pre-trained ResNet50	83.0
2-component	DeepLabv3+ and JPPNet	CD-CNNs	84.2
3-component	DeepLabv3+ and JPPNet	CD-CNNs	84.3
4-component	**DeepLabv3+ and JPPNet**	**CD-CNNs**	**85.3**
5EL	DeepLabv3+ and JPPNet	CD-CNNs	85.1
5EP	DeepLabv3+ and JPPNet	CD-CNNs	85.0
6-component	DeepLabv3+ and JPPNet	CD-CNNs	84.5

■Experimental results for FashionStyle14 dataset

Method	Preprocessing	Feature	Accuracy
Takagi et al. [3]	None	ResNet50 as End-to-End	72.0
1-component	DeepLabv3+ and JPPNet	Pre-trained ResNet50	75.5
2-component	DeepLabv3+ and JPPNet	CD-CNNs	75.9
3-component	DeepLabv3+ and JPPNet	CD-CNNs	76.9
4-component	DeepLabv3+ and JPPNet	CD-CNNs	77.3
5EL	DeepLabv3+ and JPPNet	CD-CNNs	77.4
5EP	DeepLabv3+ and JPPNet	CD-CNNs	77.6
6-component	**DeepLabv3+ and JPPNet**	**CD-CNNs**	**77.7**

[1]D. Lin et al., “Deep LAC: Deep Localization, Alignment and Classification for Fine-Grained Recognition”, CVPR, 2015.
[2]H. Zhang et al., “SPDA-CNN: Unifying semantic part detection and abstraction for fine-grained recognition”, CVPR, 2016.
[3]M. Takagi et al., “What Makes a Style: Experimental Analysis of Fashion Prediction”, ICCVW, 2017.
[4]M. Kiapour et al., “Hipster Wars : Discovering Elements of Fashion”, ECCV, 2014.
[5]L. Liang et al., “Joint Body Parsing & Pose Estimation Network and a New Benchmark”, PAMI, 2019.
[6]L.C. Chen et al., “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Separation, ECCV, 2018.
[7]E. Simo-Serra et al., “Fashion Style in 128 Floats: Joint Ranking and Classification Using Weak Data for Feature extraction”, CVPR, 2016.
[8]T. Nakajima et al., “Accuracy improvement of fashion style classification by appropriate training data and estimation of human regions”, IEICE technical report, 2018.