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**

- DNN, etc.
- bohemian
- goth
- hipster
- pinup
- preppy

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

- 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

Input: JPPNet[5]

DeepLabv3+[6]

Human body parts

- Grayscale
- Pose
- Whole body
- Clothes
- Head
- Limbs

Feature extraction and concatenation

Pre-trained CD-CNNs

- ResNet50 (Grayscale)
- ResNet50 (Pose)
- ResNet50 (Whole body)
- ResNet50 (Clothes)
- ResNet50 (Head)
- ResNet50 (Limbs)

Training and prediction

Classifier (SVM)

Predicted fashion style

- Extracting human area in a pixel unit by using JPPNet [5] and DeepLabv3+ [6].
- Creating 4 component 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.**

<table>
<thead>
<tr>
<th>Number of components</th>
<th>Component image</th>
<th>Dataset example images</th>
<th>Component name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Whole body</td>
<td>HipsterWars SEL</td>
<td>83.0</td>
</tr>
<tr>
<td>2</td>
<td>Whole body, Head</td>
<td>HipsterWars SEL</td>
<td>84.2</td>
</tr>
<tr>
<td>3</td>
<td>Whole body, Clothes, Head</td>
<td>HipsterWars SEL</td>
<td>84.3</td>
</tr>
<tr>
<td>4</td>
<td>Whole body, Clothes, Head, Grayscale</td>
<td>HipsterWars SEL</td>
<td>85.3</td>
</tr>
<tr>
<td>5</td>
<td>Whole body, Clothes, Head, Pose, Grayscale</td>
<td>HipsterWars SEL</td>
<td>85.1</td>
</tr>
<tr>
<td>6</td>
<td>Whole body, Clothes, Head, Limbs, Grayscale</td>
<td>HipsterWars SEL</td>
<td>85.0</td>
</tr>
</tbody>
</table>

**Experimental results for HipsterWars dataset**

<table>
<thead>
<tr>
<th>Method</th>
<th>Preprocessing</th>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-trained ResNet50</td>
<td>None</td>
<td>ResNet50 as End-to-End</td>
<td>72.0</td>
</tr>
</tbody>
</table>

**Experimental results for FashionStyle14 dataset**

<table>
<thead>
<tr>
<th>Method</th>
<th>Preprocessing</th>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-trained ResNet50</td>
<td>None</td>
<td>ResNet50 as End-to-End</td>
<td>72.0</td>
</tr>
</tbody>
</table>

Acknowledgment

This work was supported in part by JST, CREST and in part by JSPS KAKENHI Grant Number 16H06195.