# Impression Estimation of Suit Patterns Based on Style Features Using Multi-scale CNN

Eiki Tsumura[1], Kesnke Tobitani[2][0000−0002−3898−8435],
Miyuki Toga[1][0000−0002−2385−4389], and Noriko Nagata[1][0000−0002−2037−1947]

[1] Kwansei Gakuin University, 2–1 Gakuen, Sanda-shi, Hyogo 669–1337, Japan
{E.T-tsumu17,toga.m,nagata}@kwansei.ac.jp
[2] University of Nagasaki, 1–1–1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki
851–2195 ,Japan
tobitani@sun.ac.jp

**Abstract.** In the field of fashion design, impressions evoked by the texture of materials, such as “flashy” and “cool” (affective texture), attract attention. The affective texture is considered necessary in evaluating and judging the quality of an object and personal preferences. In the fashion domain, there is a high demand for personalization of designs when diversifying user needs. One example is a custom-made suit service. However, there is a problem that it is labor-intensive to find a suit that matches image from many patterns, and colors. It is, therefore, necessary to understand affective texture. Many studies evaluate aesthetics using product style, suggesting that style is highly associated with affective texture. Multi-scale CNN has attracted attention for image recognition and is more accurate than single-scale CNN. However, no model that combines style and multi-scale CNN has been developed in previous studies.
In this study, we propose a method for affective texture (visual impressions) evoked by suit patterns, corresponding to different scales of the pattern. (1) Suit patterns are collected in different resolution images, and impression evaluation experiments quantify (2) affective texture. (3) Model the relationship between the affective texture and physical characteristics (style features) of the pattern images using a multi-scale CNN. Then, the correlation coefficient between the impression values of the test data and the estimated impression values of the models. The results showed that multi-scale CNN has better accuracy than single-scale CNN, confirming the effectiveness of this method.

**Keywords:** Fashion · Suit · Style · Multi-scale CNN · Impression Estimation Models.

## 1 Introduction

In the field of fashion design, impressions (including aesthetics) evoked by the texture of materials, such as “flashy” and “cool” (affective texture), attract attention. The affective texture is considered necessary in evaluating and judging the quality of an object and personal preferences. In the fashion domain, the impression that clothes make on people plays a vital role in expressing a person's impression [1]. The pattern and color give different impressions to people. Therefore, one must dress appropriately for different occasions when wearing a suit. Therefore, technologies are required to quantify, index, and model affective texture [2].

With the improvement of information technology, users can quickly obtain various of product information through the Internet and SNS. User needs and preferences have become increasingly diverse in recent years, and demand for customization and personalization of products and designs has grown. Therefore, it is essential to accurately grasp users' affective values, such as their preferences and satisfaction [3].

The tailor-made suit service is an example of the customization and personalization of design in the fashion field. However, finding a suit that matches one's taste and image from many available materials, patterns, and colors is time-consuming and labor-intensive. The perception of a pattern's motif may significantly affect the pattern's overall impression (a unit that constitutes the subject of the patterns) [4]. The entire motif is observed at a low resolution if the motif is large. If the motif is small, the details are observed at high resolution (motifs can be combined to form a concept motif at a higher level). In this way, people form an impression of the pattern while appropriately changing the resolution according to the size of the motif.

In this study, we propose a method to construct impression estimation models for suit patterns based on affective texture by modeling the pattern's visual impression and physical characteristics. In particular, we aim to extend the model to impression estimation using multi-resolution (multi-scale) images to accommodate differences in pattern scale.

## 2 Previous Research

Using machine learning techniques, numerous studies have been conducted to model the impressions evoked by products and their physical characteristics. These include technologies that use a product's color, texture, and shape, as physical characteristics [5] and retrieval technologies based on impressions of images [6].

There has been much research on textures [7–9]. On the deep learning framework, Gatys et al. proposed a style transformation algorithm that content features and style features extracted from VGG19 [10] which is a CNN used for general object recognition [11]. Content features are feature maps output from middle layer of VGG19 style features is grammaticalized versions of content features. Gatys et al. have proposed that content features hold a lot of shape information necessary for general object recognition. In contrast, style features are texture features that hold many detailed appearance information, such as color and pattern, in an image.

Many studies have related to image style and kansei (sensibility) [12, 13]. These studies have built models that use style to evaluate the aesthetics of photographs [12] and to estimate the affective texture of clothing patterns [13]. The results suggest that style information in images is a feature with a high affinity to affective texture. However, since the emotions and impressions evoked by products and textures vary depending on the object's size, an estimation model that can respond to differences in the object's scale is desirable.

Multi-scale CNN, a deep learning model that considers differences in image scale, has attracted attention in image recognition. Wetteland et al. proposed a multi-scale CNN with pre-trained VGG16 concatenated in parallel [14]. It is a classification network that takes individual input images of different resolutions, and concatenates the features of each resolution image. Global Average Pooling (GAP) and Dropout layer are added to prevent over-learning. The results showed that multi-scale CNN improved classification accuracy over single-scale CNN.

In this study, we construct impression estimation style features with high affinity to affective texture and a multi-scale CNN framework to deal with differences in the scale of suit patterns (motifs). No model that combines style and multi-scale has been developed in previous studies.

## 3 Proposed Method

In this study, we propose a method for estimating affective texture (visual impressions) that the suit patterns evoke in response to differences in pattern scale. Fig. 1 presents an overview of the proposed method. (1) First, a scanner captured the suit's fabric which is collected as pattern images at various resolutions. (2) Next, we conducted an impression evaluation experiment on the pattern image of the suit to quantify the affective texture. (3) We then used the style feature as the suit patterns ' physical characteristics. We modeled relationship between the quantified affective texture and the physical characteristics using a multi-scale CNN. Finally, we conducted impression estimation on the test data based on the constructed models to confirm the proposed method's effectiveness.

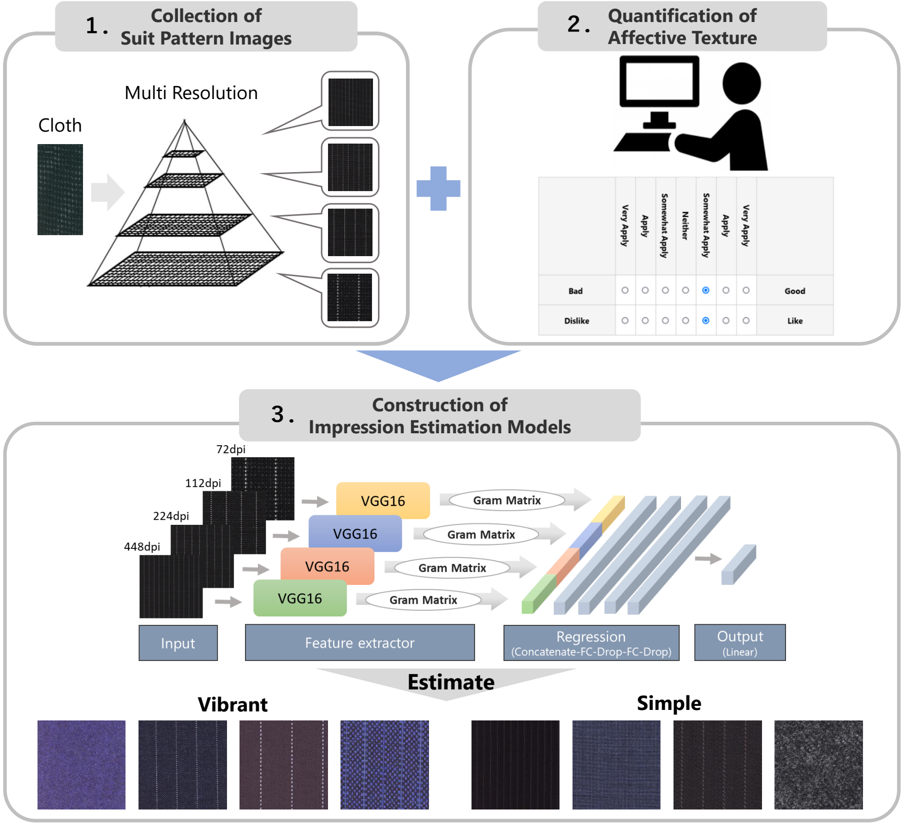


**Fig. 1.** Overview of our proposed method.

## 4 Quantification of Affective Texture

### 4.1 Collection of image data

The suit fabrics were photographed with high dynamic range (HDR) images at different resolutions to collect images representing different scales. We collect sample images at different resolutions for 613 different suit cloths based on the image pyramid of Tada et al. [15]. The resolution was set to 72 dpi, 112 dpi, 224 dpi, and 448 dpi, and a total of 2,452 pattern images were collected. Without loss of generality, capturing with HDR images preserves tone and light regions, resulting in a more transparent representation of woven patterns such as stripes and checks. However, HDR images cannot be viewed on a standard display, so they must be converted to low dynamic range (LDR) images, which can be viewed by compressing their dynamic range. In this study, we used “Photographic Tone Reproduction” by Reinhard et al. [16], a representative tone mapping method, and converted the images to LDR images. We normalized all image sizes to 224 × 224pixel. Examples of suit patterns collected from the process are shown in Fig. 2. A scanner (Epson GT -X830) and software SilverFast 8 were used to perform the analysis. The suit fabrics used for the photography dealt with some from the 2016-2019 SS and AW.

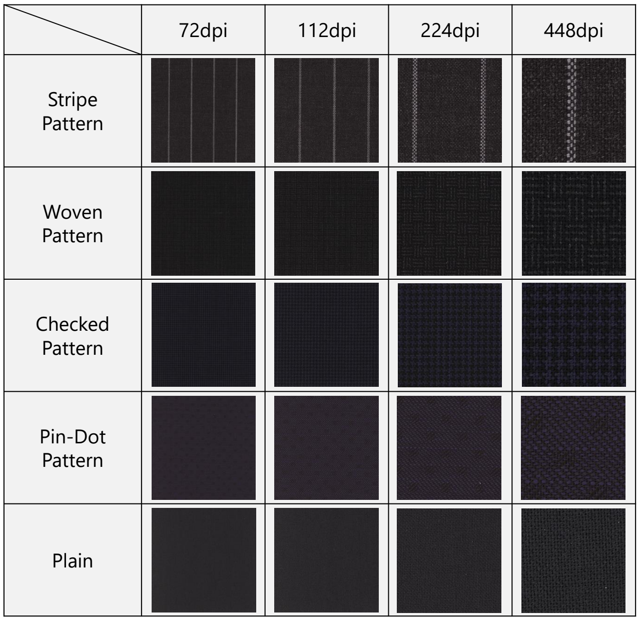


**Fig. 2.** Pattern images of suits with different resolutions.

### 4.2 Impression evaluation experiment

To quantify the affective texture evoked by suit patterns, we conducted an impression evaluation experiment using crowd-sourcing on a total of 2,452 pattern images of 613 types and four resolutions. We used Lancers as a crowd-sourcing service. In kansei (sensibility including aesthetics) research, models of kansei are often expressed in a hierarchical structure [17] consisting of three layers: emotion, impression, and physical element（Fig. 3）. By quantifying the impression layer, we assume that we can clarify the basis (causal relationship) for the formation of emotional values in the correspondence between ”people” (emotions) and ”objects” (physical characteristics). In the experiment, to select evaluation words to be used in the experiment, more than 60 words were listed through discussions with the suit experts, and we asked the suit experts to evaluate their impressions based on these evaluation words. We conducted factor analysis on the evaluation data obtained, and we adopted words with high factor loadings as suitable words for evaluating suit patterns. Table 1 lists the selected evaluation terms. We selected 44 words, 4 for the emotional layer and 40 for the impression layer, corresponding to the hierarchical structure of kansei (Fig. 3). The experiment participants were 3,080 of gender and age and were asked to rate a single stimulus for 20 people.

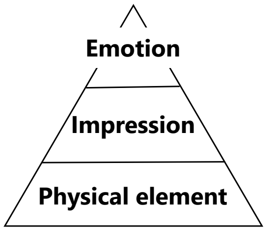


**Fig. 3.** Hierarchical structure of kansei (sensibility).

**Table 1.** Forty-four evaluation words were used in the impression evaluation experiment.

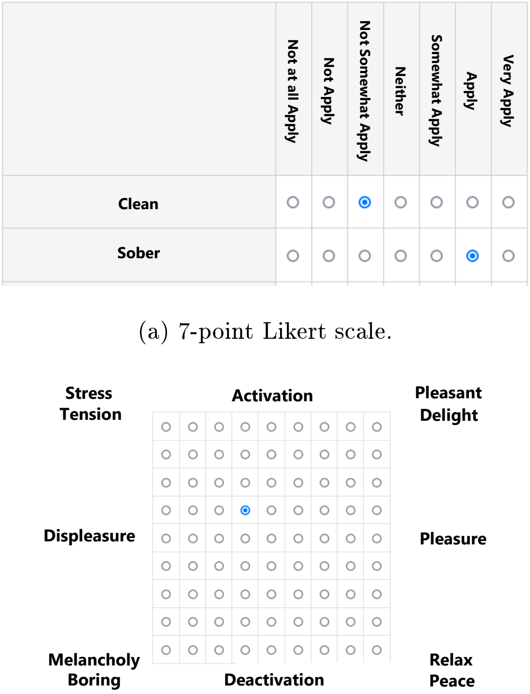
| | | | |
|---|---|---|---|
| crisp | calm | clear | formal |
| plain | unique | vibrant | suitable for everyone |
| sharp | fashionable | simple | sober |
| well-tailored | adult | standard | youthful |
| refreshingly cool | dressy | massive | trendy |
| elegant | powerful | modern | sexy |
| classical | kind | neat | smart |
| supple | prominent | squeezed | stylish |
| gorgeous | good impression | clean | serious |
| light feeling | cool | humility | warm |
| deactivation - activation | displeasure - pleasure | dislike-like | bad-good |

The experimental procedure was conducted using a Likert scale and the Affect-Grid method of Russell et al. [17]. Participants were asked to observe the pattern images of the suits displayed on the screen. They were asked to respond to the degree to which each evaluation term applied to them using a seven-point scale consisting of “Not at all Apply,” “Not Apply,” “Not Somewhat Apply,” “Neither,” “Somewhat Apply,” “Apply,” and “Very Apply. The scores for each rating scale were divided into 1-point increments, with 1 point representing “Not at all Apply” and 7 points representing “Very Apply. Two words, “displeasure - pleasure” and “deactivation - activation” was evaluated using the Affect-Grid method. In this study, we used an affect grid of two dimensions, with the horizontal axis indicating “displeasure - pleasure” and the vertical axis indicating “deactivation - activation.” We asked respondents to respond on a nine-point scale from 1 to 9 for each dimension. The experimental screen is shown in Fig. 4.

The above process was used to obtain the evaluation data of 44 evaluation words for 20 participants for each stimulus.

After that, the obtained evaluation data was cleaned for respondents who were dishonest or only completed the survey for a short amount of time. When assessing the stimuli, dummy images are prepared, and those who have not performed the evaluations instructed beforehand on dummy images are considered dishonest respondents. Among the dishonest respondents, the peak (mode) response time was calculated, and those whose response time was shorter than the peak were defined as the short time of the respondents. Due to the cleaning process, 90% of the 3,080 total respondents were valid.

Finally, a discrete probability distribution with the scored evaluation values as the random variable was used as the impression distribution of the suit pattern images. In this case, the impression distribution was normalized by the number of raters for each stimulus because the number differed between stimuli based on cleaning the evaluation data. The obtained impression distribution is converted to expected values and used as the teacher data for the impression estimation models to be constructed later.



| | Not at all Apply | Not Apply | Not Somewhat Apply | Neither | Somewhat Apply | Apply | Very Apply |
|---|---|---|---|---|---|---|---|
| Clean | ○ | ○ | ● | ○ | ○ | ○ | ○ |
| Sober | ○ | ○ | ○ | ○ | ○ | ● | ○ |

(a) 7-point Likert scale.

(b) Affect-Grid.

**Fig. 4.** Experiment screen.

# 5 Modeling the Relationships Between Affective Texture and Physical Characteristics

## 5.1 Construction of Multi-scale CNN

In this study, we construct a multi-scale CNN based on the deep-learning model of Wetteland et al. [13] . As a detail of the model to be constructed, the input images are images of four resolutions: 72dpi, 112dpi, 224dpi, and 448dpi. We use the pre-trained VGG16 as a feature extractor. VGG16 is characterized by its emphasis on texture information [19, 20]. We use it as the basis for this model. Multi-scale CNN places four VGG16s in parallel to generate multiple feature vectors. These feature vectors are concatenated before being input to the FC layer. The FC layer has the same size as the original VGG16: 4,096 neurons.

Style features in Gatys et al.'s style transfer [11] are used as physical characteristics to represent the pattern image of the suit. In this case, style features are a Gram matrix of feature maps from the middle layer of CNN and have been extracted from the second pool layer of VGG16. However, the dataset is small and multi-scale CNN has many parameters, making them prone to overfitting during training. Therefore, a Dropout layer was added after each FC layer to prevent overfitting. Both Dropout rates were set to 0.5. Since VGG16 is a model that solves a classification problem, ia regression problem replaces it when estimating visual impressions. Therefore, we focused on the activation functions in each layer of VGG16, replacing the ReLU function in the middle layer with the hyperbolic tangent function, and the SoftMax function in the output layer with the identity function. The VGG16 architecture handled by multi-scale CNN is shown in Fig. 5. The architecture of all four VGG16s is the same as in Fig. 5.

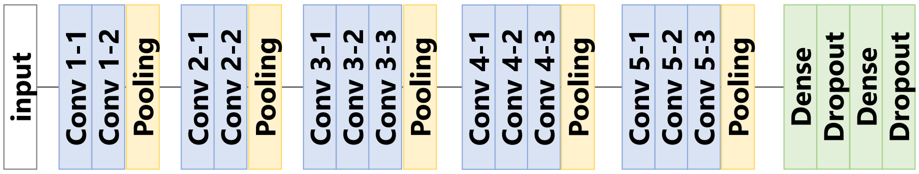


**Fig. 5.** VGG16 architecture.

## 5.2 Training

To estimate affective texture of suit patterns, we used the framework described in Section5.1. In doing so, we solved a regression problem in which the expected value of the impression distribution (impression value) created in Section 4.2 was the objective variable, and the style features extracted from the suit pattern images were the explanatory variables. Multi-scale learning and evaluation were performed using impression values of 72 dpi, 112 dpi, 224 dpi, and 448 dpi images as ground truth for a dataset of 2,452 images. Cross-validation was used to train and evaluate the models, and 10-fold cross-validation was used to divide the dataset into Train: Validation: Test = 9: 1: 1 .

The Adam optimizer was used during training. We used Mean Squared Error (MSE) for the loss and evaluation functions. The learning rate was set to 0.001, the batch size to 64, and the number of epochs to 100. To avoid overfitting, we used Early Stopping, which terminates training when no accuracy improvement is expected. Learning is stopped when there has been no improvement in accuracy over the past ten epochs.

## 6 Results and Discussions

### 6.1 Accuracy comparison between Single-scale CNN and Multi-scale CNN

To validate the effectiveness of the proposed method, we compared the estimation accuracy of the model by 10-fold cross-validation between single-scale CNN and multi-scale CNN. We use the correlation coefficient between the impression value for the test data and the impression value estimated by the models as a measure of estimation accuracy. Table 2 shows the correlation coefficients for the impression values of 72 dpi, 112 dpi, 224 dpi, and 448 dpi images as ground truth for each evaluation term. Table 2 shows the correlation results for the single-scale CNN in the left column and the multi-scale CNN in the right column. As a result, moderate or higher positive correlations were found for most evaluation terms. This confirms that the suit pattern's style features have a high affinity with people's affective texture. There were 37 evaluation terms an exceptionally high correlation (0.4 or higher), with an average coefficient of 0.57. Among them, "deactivation - activation," "vibrant," "calm," and "sober" showed high correlations at all resolutions. We believe that these evaluation words tend to have high correlations because they are low-order impressions associated with the color and characteristics of the pattern. The average correlation coefficients for all evaluation words were more accurate with the multi-scale CNN than with the single-scale CNN, confirming the effectiveness of the multi-scale CNN.

### 6.2 Human Visual angle and Resolution

We discuss visual processing and resolution when humans look at the pattern of a suit. The visual angle is the angle the object projected onto the eye makes and is the angle at which the object is viewed. In this study we compare the relationship between the range of resolution at which the human looks the suit patterns and the resolution in the constructed impression estimation models. The viewing angle is determined by the size of the visual object and the viewing distance when observing it and is calculated from Eq. (1).

$$v = 360/\pi \times \arctan(s/2d) \tag{1}$$

Where $v$ is the viewing angle, $s$ is the object's size, and $d$ is the viewing distance. We then calculated the resolution at which a person sees the suit pattern from

**Table 2.** Correlation coefficients for each evaluation term.

(Left column: single-scale CNN, Right column: multi-scale CNN.)

| evaluation word | 72dpi | | 112dpi | | 224dpi | | 448dpi | |
|---|---|---|---|---|---|---|---|---|
| | single-scale | multi-scale | single-scale | multi-scale | single-scale | multi-scale | single-scale | multi-scale |
| deactivation - activation | 0.79 | 0.76 | 0.71 | 0.69 | 0.63 | 0.77 | 0.70 | 0.68 |
| displeasure – pleasure | 0.30 | 0.40 | 0.19 | 0.25 | 0.39 | 0.37 | 0.30 | 0.43 |
| dislike-like | 0.59 | 0.63 | 0.42 | 0.44 | 0.34 | 0.41 | 0.41 | 0.47 |
| bad-good | 0.72 | 0.68 | 0.38 | 0.24 | 0.43 | 0.47 | 0.35 | 0.26 |
| crisp | 0.64 | 0.57 | 0.35 | 0.46 | 0.36 | 0.47 | 0.18 | 0.32 |
| clear | 0.62 | 0.70 | 0.30 | 0.49 | 0.57 | 0.61 | 0.31 | 0.43 |
| plain | 0.43 | 0.68 | 0.62 | 0.64 | 0.49 | 0.59 | 0.46 | 0.50 |
| vibrant | 0.76 | 0.80 | 0.66 | 0.74 | 0.68 | 0.71 | 0.77 | 0.66 |
| sharp | 0.54 | 0.52 | 0.29 | 0.53 | 0.25 | 0.41 | 0.15 | 0.27 |
| simple | 0.68 | 0.74 | 0.66 | 0.73 | 0.55 | 0.67 | 0.64 | 0.61 |
| well-tailored | 0.16 | 0.45 | 0.07 | 0.20 | 0.18 | 0.07 | 0.26 | 0.22 |
| standard | 0.65 | 0.78 | 0.60 | 0.68 | 0.63 | 0.66 | 0.52 | 0.43 |
| refreshingly cool | 0.51 | 0.63 | 0.49 | 0.63 | 0.36 | 0.53 | 0.48 | 0.55 |
| warm | 0.42 | 0.47 | 0.27 | 0.35 | 0.38 | 0.47 | 0.38 | 0.47 |
| Massive | 0.73 | 0.70 | 0.64 | 0.67 | 0.63 | 0.64 | 0.22 | 0.36 |
| elegant | 0.36 | 0.39 | 0.34 | 0.39 | 0.20 | 0.21 | 0.23 | 0.31 |
| modern | 0.16 | 0.23 | 0.22 | 0.30 | 0.27 | 0.23 | 0.29 | 0.13 |
| classical | 0.72 | 0.68 | 0.44 | 0.58 | 0.29 | 0.49 | 0.29 | 0.30 |
| neat | 0.78 | 0.75 | 0.67 | 0.70 | 0.63 | 0.69 | 0.58 | 0.60 |
| supple | 0.16 | 0.32 | 0.13 | -0.06 | 0.38 | 0.29 | 0.39 | 0.40 |
| squeezed | 0.57 | 0.68 | 0.51 | 0.54 | 0.44 | 0.57 | 0.35 | 0.37 |
| gorgeous | 0.62 | 0.72 | 0.44 | 0.50 | 0.25 | 0.46 | 0.55 | 0.49 |
| clean | 0.40 | 0.55 | 0.37 | 0.45 | 0.34 | 0.40 | 0.30 | 0.45 |
| light feeling | 0.67 | 0.69 | 0.60 | 0.64 | 0.35 | 0.53 | 0.40 | 0.27 |
| calm | 0.78 | 0.82 | 0.67 | 0.76 | 0.71 | 0.76 | 0.71 | 0.70 |
| formal | 0.75 | 0.80 | 0.70 | 0.74 | 0.63 | 0.68 | 0.53 | 0.44 |
| unique | 0.67 | 0.75 | 0.68 | 0.70 | 0.59 | 0.67 | 0.75 | 0.69 |
| suitable for everyone | 0.74 | 0.79 | 0.70 | 0.71 | 0.64 | 0.69 | 0.64 | 0.57 |
| fashionable | 0.51 | 0.50 | 0.41 | 0.42 | 0.33 | 0.45 | 0.46 | 0.43 |
| sober | 0.77 | 0.79 | 0.70 | 0.75 | 0.57 | 0.75 | 0.68 | 0.75 |
| adult | 0.67 | 0.67 | 0.59 | 0.65 | 0.45 | 0.68 | 0.27 | 0.54 |
| youthful | 0.59 | 0.58 | 0.43 | 0.52 | 0.32 | 0.53 | 0.27 | 0.45 |
| dressy | 0.54 | 0.61 | 0.49 | 0.51 | 0.39 | 0.38 | 0.48 | 0.42 |
| trendy | 0.29 | 0.45 | 0.49 | 0.48 | 0.22 | 0.39 | 0.10 | 0.31 |
| powerful | 0.68 | 0.70 | 0.55 | 0.60 | 0.49 | 0.58 | 0.15 | 0.50 |
| sexy | 0.61 | 0.59 | 0.39 | 0.53 | 0.22 | 0.46 | 0.04 | 0.36 |
| kind | 0.40 | 0.52 | 0.45 | 0.35 | 0.26 | 0.22 | 0.34 | 0.34 |
| smart | 0.39 | 0.58 | 0.31 | 0.50 | 0.50 | 0.55 | 0.30 | 0.29 |
| prominent | 0.69 | 0.63 | 0.38 | 0.48 | 0.48 | 0.56 | 0.24 | 0.34 |
| stylish | 0.49 | 0.52 | 0.21 | 0.33 | 0.48 | 0.39 | -0.03 | 0.25 |
| good impression | 0.41 | 0.67 | 0.46 | 0.43 | 0.46 | 0.38 | 0.37 | 0.39 |
| serious | 0.74 | 0.77 | 0.66 | 0.74 | 0.63 | 0.68 | 0.67 | 0.66 |
| cool | 0.31 | 0.61 | 0.28 | 0.37 | 0.28 | 0.51 | 0.43 | 0.39 |
| humility | 0.73 | 0.79 | 0.76 | 0.78 | 0.64 | 0.74 | 0.69 | 0.67 |
| average | 0.56 | 0.63 | 0.47 | 0.53 | 0.44 | 0.52 | 0.40 | 0.44 |

the viewing distance and the minimum resolution of the human eye. Suppose a person's visual acuity is 1.0. In that case, the minimum resolution of the human eye is approximately 1/60 degree in terms of visual angle, and the resolution at which a human can look is calculated from Eq. (2).

$$dpi = \frac{2.54}{2 \times d \times \tan(v_1/2)} \tag{2}$$

Where $d$ is the viewing distance and $v_1$ is the viewing angle (1/60 degree). In this study, we use Eq. (1) and Eq. (2) to compute the range of resolution over which a person sees a suit pattern. One obtains the finest resolution is when looking at the fabric up close. At this time, assuming the normally quoted viewing distance of 57cm, the resolution was about 153dpi, calculated from Eq. (2). Next, one obtains the coarsest resolution when looking at the entire suit jacket. Here, the effective viewing angle for a person is assumed to be 30 degrees, which is generally defined. The visual distance for observing the suit pattern was obtained from Eq. (1), used the actual size of the suit as 75 cm (average value). As a result, the viewing distance for the suit was 140 cm. Calculations using Eq. (2) based on the calculated viewing distance indicate a resolution of approximately 62dpi.

Substituting this into the constructed impression estimation models, we can expect the highest estimation accuracy when the impression value of the 72dpi or 112dpi images is used as ground truth. In this study, the actual size of the suit fabric was small, making it difficult to capture images at a resolution of 72dpi or lower. Table 2 shows the estimation accuracy at each resolution. The average correlation coefficient for all evaluation words is highest when 72dpi is used as ground truth, and accuracy tends to decrease as resolution increases. These results suggest that the relationship between the resolution at which one look at the suit patterns and the optimal resolution in the impression estimation models is close. This confirms that the relationship between the range of resolution at which people actually see the pattern of the suit and the resolution in the impression estimation model is a close result. Looking at the evaluation terms that differ in accuracy between 72dpi and 448dpi, "massive" and "light feeling" are listed. Both impressions expressed "weight-lightness," suggesting a more macroscopic observation of the pattern in these evaluation terms. Furthermore, it was suggested that the affective texture may change depending on the pattern's scale.

### 6.3 Confirmation of estimation results

Visually confirm the relationship between the estimation results by the constructed impression estimation models and the pattern image of the suit. We show the top and bottom two suit pattern images with the highest estimation accuracy (Fig. 6) for each resolution. The evaluation words for the estimation results are four impression words, "simple," "unique," "vibrant," and "calm," and two emotion words, "deactivation - activation" and "dislike-like. The models with the highest accuracy in cross-validation were used for estimation.

In the “simple” category, the top patterns primarily were solid black or gray and pinstripe patterns (thin stripes), while the bottom patterns were often loudly colored stripes and staggered plaid patterns. In the “unique” category, the top patterns were often loudly colored stripes and staggered plaid, while the bottom patterns were standard patterns such as plain black or gray and stripes. In the “vibrant” category, the top patterns were often brightly colored solid colors and stripes with blue as the base color, while the lower patterns were often plain colors and stripes of subdued colors. In the “calm” category, the top patterns for “calmness” were plain black or gray and shadow stripe patterns, while the bottom patterns were brightly colored stripes and staggered lattices. In the “deactivation - activation” category, most activation patterns are brightly colored stripes or staggered plaid, while deactivation patterns are calmly colored, such as plain gray or black. In the “dislike-like” category, standard patterns dominated responses, such as pinstripes and solid colors for the “like” response. In contrast, patterns with gaudy colors and geometric patterns were selected for the “dislike” response. These results confirm that the estimation results are close to the human image.

### 6.4 Relationship between resolution and pattern motifs

Based on the constructed impression estimation model, we discuss the relationship between resolution and pattern motifs. Using the actual impression values and the model-estimated impression values, we calculate the correlation coefficient for each pattern, and visually confirm the pattern with the highest correlation. We do so at each resolution. Fig. 7 shows some of the patterns with high correlation coefficients (estimation accuracy) at each resolution.

From Fig. 7, patterns with large motifs, such as alternate stripes and solid colors, are more common at 72 dpi. Alternate stripes are stripes with alternating lines of different colors and are believed to have a larger motif than a typical stripe pattern. At 112 dpi, there are medium-patterned stripes with medium-sized motifs, such as chalk stripes. On the other hand, patterns with large plain-like motifs are also mixed. Similarly, 224dpi has many medium-sized stripes with medium-sized motifs, such as pinstripes. At 448dpi, patterns with small motifs, such as staggered plaid and dot patterns, were observed. These results suggests that people change their resolution (viewing angle) according to the motif’s size and that the affective texture obtained from the pattern may depend on the motif’s size.

## 7 Conclusion

In this study, we proposed a method for estimating affective texture that the suit patterns evoke in response to differences in pattern scale and conducted the following procedure. (1) collect pattern images of suits at different resolutions, and (2) quantify the affective texture of 44 evaluation words through an impression evaluation experiment. (3) Style features extracted from an image are used as physical characteristics, and the relationship between affective texture and

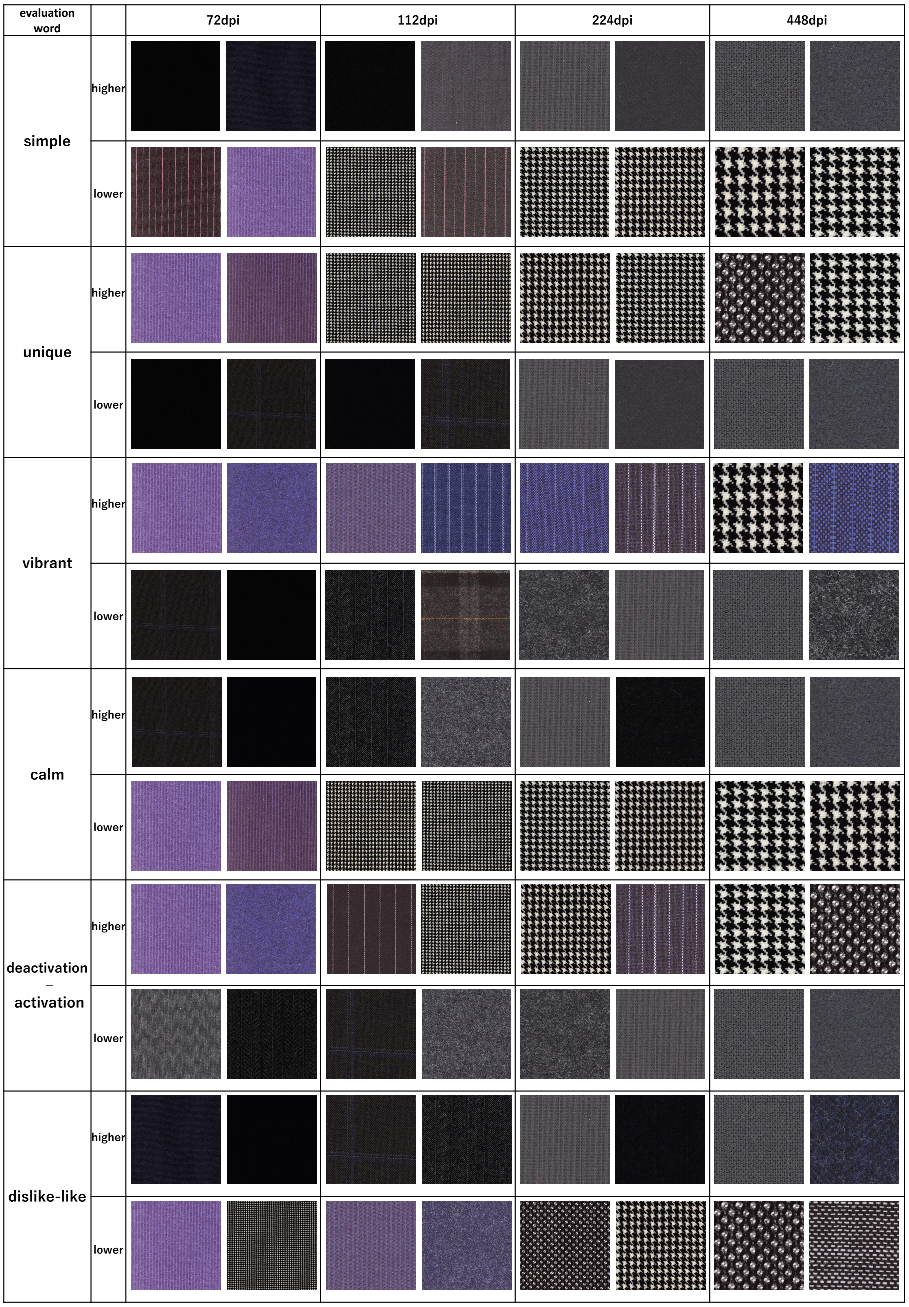


**Fig. 6.** Estimation results from impression estimation models.

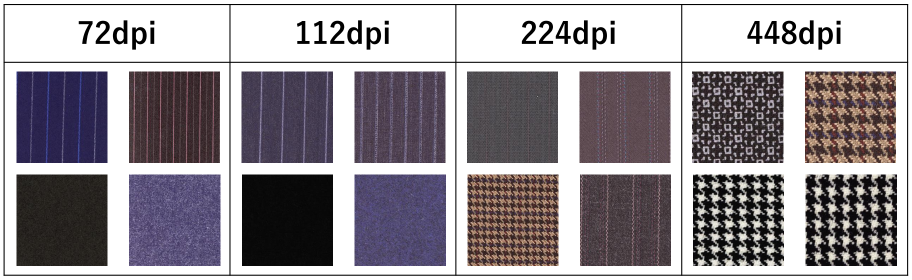


**Fig. 7.** Suit patterns with high estimation accuracy at each resolution

physical characteristics is modeled using a multi-scale CNN. A 10-fraction cross-validation was performed, and the correlation coefficients between the impression values of the test data and those estimated by the models were calculated. Results showed moderate or higher positive correlations for most evaluation terms. We found that multi-scale CNN improves the accuracy compared to single-scale CNN, confirming the effectiveness of this method. The resolution at which the suit patterns are viewed from the human visual angle was calculated and compared with the results of the constructed impression estimation models. They confirmed that the estimation accuracy was highest for images at 72dpi, close to the human resolution.

Future work, we expect to improve the estimation accuracy given the small dataset in this training by increasing the training data using GAN [18].

**Acknowledgements** This work was supported by JST COI Grant Number JPMJCE1314.

## References

1. Yamamoto, M., Onisawa, T.: Interactive Fashion Design and Coordinate System Considering User's KANSEI. Transactions of Japan Society of Kansei Engineering 15(1), (2016).
2. Yamazaki, Y., Imura, M., Tobitani, K., Tani, Y., Nagata, N.: Development of measurement and simulation scheme for digitalization of tactile perception. In: Asia International Symposium on Mechatronics (AISM), pp. 981-986 (2019).
3. Tobitani, K., Shiraiwa, A., Katahira, K., Nagata, N., Nikata, K., Arakawa, K.: Modeling of "high-class feeling" on a cosmetic package design. Journal of the Japan Society of Precision Engineering 87(1), 134–139 (2021).
4. Tani, S., Matsunashi, K., Shimazaki, K.: A Study on the Configuration of Curtains (Part 5) —The Influence of Polka-dot Patterns on the Apparent Configuration of Curtains—. Journal of the Japan Research Association for Textile End-Uses 54(7), 646-655 (2013).
5. Niwa, S., Aoyama, Y., Sudo, K., Taniguchi, Y., Kato, T.: Modeling relationship between visual impression of commodities and their graphical features. In: IPSJ SIG Technical Reports 2013-HCI-152, pp. 1–4 (2013).

6. Chen, Y.W., Huang, X., Chen, D., Han, X.H.: Generic and specific impressions estimation and their application to KANSEI based clothing fabric image retrieval. J. Pattern Recognit. and Artif. Intell 32(10), 1854024 (2018).
7. Tani, Y., Nagai, T., Koida, K., Kitazaki, M., Nakauchi, S.: Experts and novices use the same factors-but differently-to evaluate pearl quality. PLOS ONE 9(1), 1–7 (2014).
8. Tobitani, K., Matsumoto,T., Tani, Y., Fujii, H., Nagata, N.: Modeling of the relation between impression and physical characteristics on representation of skin surface quality. The Journal of The Institute of Image Information and Television Engineers 71(11), 259-268 (2017).
9. Doizaki, R., Iiba, Saki., Okatani, T., Sakamoto, M.: Possibility to Use Product Image and Review Text Based on the Association between Onomatopoeia and Texture. Transactions of the Japanese Society for Artificial Intelligence : AI, 30(1), 124–137 (2015).
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: The 3rd International Conference on Learning Representations (ICLR), pp. 1–14 (2015).
11. Gatys, L. A., Ecker, A. S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414–2423 (2016).
12. Gao, F., Li, Z., Yu, Jun., Yu, Junze., Huang, Q., Tian, Q.: Style-adaptive photo aesthetic rating via convolutional neural networks and multi-task learning. Neurocomputing 395, 247-254 (2020).
13. Sunda, N., Tobitani, K., Tani, I., Tani, Y., Nagata, N., Morita, N.: Impression estimation model for clothing patterns using neural style features. In: Stephanidis, C., Antona, M. (eds.) HCI International 2020 - Posters. HCII 2020, Communications in Computer and Information Science, vol. 1226, pp. 689–697. Springer, Cham (2020).
14. Wetteland, R., Engan, K., Eftestøl, T., Kvikstad, V., Emiel A., Janssen, M.: A Multiscale Approach for Whole-Slide Image Segmentation of five Tissue Classes in Urothelial Carcinoma Slides. Technology in Cancer Research & Treatment 19 (2020).
15. Tada, M., Kato, T.: Similarity Image Retrieval System Using Step-by-Step Hierarchical Classification. IPSJ Transactions on Databases (TOD) 44(8), 37–45 (2003).
16. Reinhard, E., et al.: Photographic tone reproduction for digital images. ACM Transactions on Graphics 21(3), 267–276 (2002).
17. Miyai, S., Katahira, K., Sugimoto, M., Nagata, N., Nikata, K., Kawasaki, K.: Hierarchical structuring of the impressions of 3D shapes targeting for art and non-art university students. In: Stephanidis, C., (ed.) HCI International 2019 - Posters. HCII 2019, Communications in Computer and Information Science, vol. 1032, pp. 385–393. Springer, Cham (2019).
18. Russell, W., Mendelsohn, G.A.: Affect-Grid: A Single-Item Scale of Pleasure and Arousal. Journal of Personality and Social Psychology 57(3), 493-502 (1989).
19. Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., Brendel, W.: ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231 (2018).
20. Tuli, S., Dasgupta, I., Grant, E., Griffiths, T. L.: Are convolutional neural networks or transformers more like human vision?. arXiv preprint arXiv:2105.07197 (2021).
21. Tsumura, E., Tani, I., Tobitani, K., Nagata, N.: Textile-GAN: Generation of Texture for Woven Pattern Using Generative Adversarial Networks. In: The Institute of Electronics, Information and Communication Engineers (IEICE), pp.19-20 (2021).