Apparel-based deep learning system design for apparel style recommendation

Congying Guan[1], Shengfeng Qin[1] and Yang Long[2]

[1]School of Design, Northumbria University, Newcastle upon Tyne, UK

[2]Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, UK

Congying.guan@gmail.com

Abstract

Purpose

The big challenge in apparel recommendation system research is not the exploration of machine learning technologies in fashion, but to really understand clothes, fashion and people and know what to learn. This paper aims to explore an advanced apparel style learning and recommendation system that can recognise deep design-associated features of clothes and learn the connotative meanings conveyed by these features relating to style and the body so that it can make recommendations as a skilled human expert.

Design/methodology/approach

This study first proposes a new clothes style training data. Secondly, it designs three intelligent apparel learning models based on newly proposed training data including ATTRIBUTE, MEANING and the raw image data, and compares the models' performances in order to identify the best learning model. For deep learning, two models are introduced to train the prediction model, one is a Convolutional Neural Network joint with the baseline classifier Support Vector Machine and the other is with a newly proposed classifier Later Kernel Fusion.

Findings

The results show that the most accurate model (with average prediction rate of 88.1%) is the third model that is designed with two steps, one is to predict apparel ATTRIBUTEs through the apparel images, and the other is to further predict apparel MEANINGs based on predicted ATTRIBUTEs. The results indicate that (1) adding the proposed ATTRIBUTE data that captures the deep features of clothes design does improve the model performances (e.g. from 73.5%, Model B to 86%, Model C), and (2) the new concept of apparel recommendation based on style meanings is technically applicable.

Originality/value

The apparel data and the design of three training models are originally introduced in this study. The proposed methodology can evaluate the pros and cons of different clothes feature extraction approaches through either images or design attributes and balance different machine learning technologies between the latest CNN and traditional SVM.

Keywords

Clothes and Fashion, Smart shopping, Apparel recommendation system, Machine Learning, Deep Learning

1. Introduction

New technology especially the Artificial Intelligence stands to reshape the way of people shopping. In fashion retailing, in order to improve the customer experience, a new concept of smart shopping appears at both in-store and internet-based online-shopping markets. *Echo Look* (Amazon, 2017, Hartmans, 2017) is a camera-based personal style assistant device developed by Amazon in 2017 currently applicable on the USA market. It introduces the *Style Check* which offers a second opinion on which outfit looks best based on fit, colour, styling, and current trends using advanced machine learning algorithms and advice from fashion specialists. Fashion AI (Sun, 2017) is a professional style assistant system developed by a research team of Alibaba, China in 2017. It could recognise clothing, the tastes of designers and generate outfit matches. In the UK market, *Dressipi (Landia, 2017, Dressipi, 2018)*, a one-to-one style personalisation and recommendation start-up developed by a technology innovation company in London, has been used by large retailers such as John Lewis and Topshop. The exploration of AI-based fashion recommendation or shopping assistant system becomes a new innovative research area that attracts research interests from the academic researchers, technology developers and leading retailers.

Artificial Intelligence, Machine Learning, Deep Learning and Computer Vision are the main technologies to support the development of intelligent apparel recommendation systems and smart shopping devices in fashion field. There are two major issues in existing apparel recommendation system studies. Firstly, although huge volumes of clothing data were used to train machine learning model, however, the data itself may not accurately represent the features of clothes, which could restrict the performance of a training model. Such data being applied in developing clothes recommendation systems lack the depth of theory and knowledge relating to clothes and fashion. The other issue is regarding the practicability and usefulness of apparel-oriented machine learning models. In fact, an accurate apparel learning model may not be a good apparel recommendation model without knowing what clothes really do for people and how to meet their needs from the physical body to the mental concepts.

To address these issues, this paper aims to develop a knowledge-based apparel deep learning model which integrates the profound knowledge of fashion and clothing theories as well as practical experiences of clothes and style experts into our apparel-based deep learning system design. This paper designs three models to train and test the newly proposed apparel data which are composed of three types of datasets, an Image dataset, an ATTRIBUTE dataset and a MEANING dataset. The training methods involve with the conventional machine learning algorithms as well as the latest deep learning model such as the Convolutional Neural Networks. Based on the model performances, this study finds the pros and cons of each proposed datasets and algorithms and evaluates their capabilities of applying to a knowledge-based apparel style recommendation system.

The remaining of this paper is organized as follows: section two reviews other works relating to knowledge-based apparel recommendation system developments, especially the applied machine learning methods. Section three introduces three new models according to the characters of proposed apparel data. The results of model performances and comparisons are demonstrated in section four. Section five evaluates the feasibility of using designed model in apparel style recommendation system. The following two sections discuss the results and findings from this study and draw the conclusion.

2. Related work

Knowledge-based apparel recommendation system is an advanced recommendation system that learns clothes knowledge rather than tracking customer data. In research, based on a comprehensive review conducted by the authors (Anonymous, 2016) and latest updates, it is found that some recent studies have started to focusing on the integration of clothing and fashion knowledge, such as the fashion DNA, styling and body dressing in latest studies (Landia, 2017, Bracher et al., 2016, Perkinian and Vikkraman, 2015, Vaccaro et al., 2016, Vuruskan et al., 2015, de Barros Costa et al., 2017). The technologies required by such a system are from two aspects, namely, automatic apparel feature extractions and personalised clothes style recommendation.

In apparel feature extractions, the earliest idea was using simple image features such as the edges and boundaries to train apparel classification model (Cheng and Liu, 2008). Then, the methods of Scale Invariant Feature Transform (SIFT), Histograms of Oriented Gradient (HOG), Local Binary Pattern (LBP), Color histogram and body detector have been used for apparel image feature extractions (Chen et al., 2012, Kalantidis et al., 2013, Hidayati et al., 2012). Liu et al. proposed an occasion-oriented clothing recommendation system that successfully applied HOG, LBP and colour histogram etc. for apparel-based attributes detection and classifications on dressing occasions (Liu et al., 2012a).

The methods applied in system modelling are dominated by the rapid growth of machine learning technology, from the earlier Expert System, Genetic Algorithm and Neural Networks to the latest Deep Learning algorithms. Currently, the Convolution Neural Network with powerful feature recognition function becomes the dominant approach in image recognition, including the apparel (Hara et al., 2016, Perlin and Lopes, 2015, Liu et al., 2016, Li et al., 2017). Google developed a Style Detection feature in Google Cloud Vision API based on deep learning of CNNs (Ward, 2016). Zalando researchers (Bracher et al., 2016) proposed a fashion DNA-based recommendation model based on a deep neural network architecture to learn customer style. Chen et al. applied

CNN model for developing image-based product recommendation involving product category recognition and similarity calculation (Chen et al., 2017).

The deep learning model CNN has enormous potential in developing knowledge-based apparel recommendation systems because it could not only extract low-level features that are similar to SIFT and HOG etc. but also train feature detector to recognise highly sensitive clothes features to adapt specific classification tasks. That is why CNNs can achieve higher accuracy levels in image feature extraction. This study intends to construct a novel ConvNets model that is able to extract high-dimensional deep attributes of apparel such as visual design from apparel images (signs) and make intricate classifications according to learning outputs (meanings).

3. New apparel-based machine learning system design

3.1 New apparel data design

The development of apparel datasets is based on the specific theories identified in clothes and fashion, which are the clothes communication and semiotics theories describing the relationship between clothes and people. Signs and Meanings derived from the semiotics theory are two main components to interpret the substitute of clothes and the meanings of clothes that are communicative for people. The structure of apparel datasets is connected with the clothes communication networks. Our previous paper has detailed the apparel data development methods based on these theories (Anonymous, forthcoming 2018). As shown in figure 1, the raw apparel image dataset represents the original visual signs of clothes. The ATTRIBUTE dataset represents the denotation meanings of clothes that indicate apparel lines, shapes, colours, patterns and other design features. The MEANING labelling dataset represents the connotation meanings of clothes that are generated by the mental concept of people.
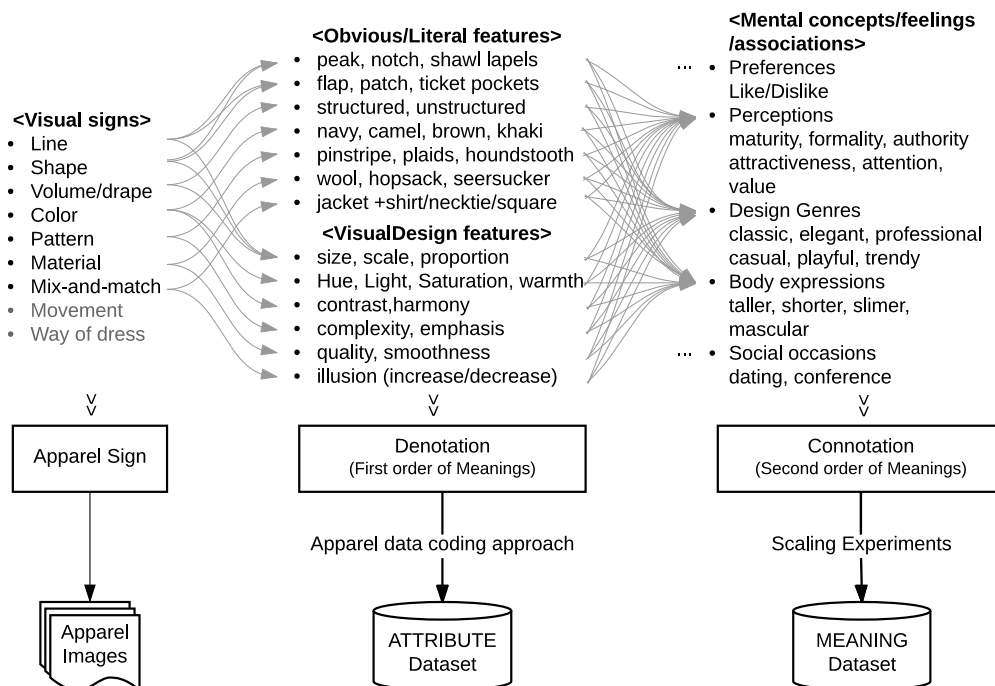

Figure 1 The development of three apparel datasets (Anonymous, forthcoming 2018)

Image-based data

The image-based dataset is composed of raw images of 200 mix-and-matched clothing sets selected from online shops. It covers four menswear categories, suit jackets/blazers, shirts, neckties and pocket squares. This part of data is applied to image-based apparel deep learning model design in next section.

ATTRIBUTE and MEANING data

The ATTRIBUTE and MEANING datasets represent visual design features such as lines, shapes and patterns extracted from images and the meanings conveyed by these visual design features. The ATTRIBUTE data captures the denotative meanings of apparel itself which nominally denote the literal meanings of clothes, such as lapel shape is denoted as peak, notch or shawl. An extended concept of the denotative meaning was proposed to introduce the deep design features, such as the proportion of lapel shapes and the contrast of colour schemes. A coding system is introduced which identifies a total number of 45 attributes in 220 sub-categories to describe jackets, shirts, neckties and pocket squares. The MEANING data represents the connotation meanings outside of apparel which denote the feelings, thoughts, beliefs and desires of clothes based on mental concepts. For instance, a peak lapel with a wider shape and higher position of gorge can create a muscular chest look. Eleven semantic words relating to perceptual feelings (to express style genres) and body meanings (to indicate body shapes) are used to label apparel samples.

An example of transferring apparel image to ATTRIBUTE data:
ATTRIBUTE (Image A) =[2,2,2,2,1,3,5,4,3,1,3,4,1,2,1,3,2,3,1,2,3,1,1,1,1,1,1,2,1,3,2,3,3,3,2,4,1,1,1,
4,3,1,1,4,1]

Table 1 Generation of ATTRIBUTE codes through a sample image

Codes	Representations of codes		
2	Peak	Jacket lapel	Attribute_1
2	Patch	Jacket pocket	Attribute_2
2	Double breasted	Jacket breast	Attribute_3
2	Two-row	Jacket button(s)	Attribute_4
1	Slim	Jacket fit	Attribute_5
3	Figured pattern	Jacket fabric	Attribute_6
5	Navy	Jacket colour	Attribute_7
4	Plaids	Jacket pattern	Attribute_8
3	Wide	Lapel width	Attribute_9
1	High	Gorge position	Attribute_10
3	Low	First button position	Attribute_11
4	Wide long	Jacket proportion	Attribute_12
1	Structured	Jacket drape	Attribute_13
2	Waistline	Jacket silhouette	Attribute_14
1	Blue	Jacket Hue	Attribute_15
3	Dark	Jacket Light	Attribute_16
2	Dull	Jacket Saturation	Attribute_17
3	Large	Jacket pattern size	Attribute_18
1	Low	Jacket pattern contrast	Attribute_19
2	Matte	Jacket material surface	Attribute_20
3	Spread	Shirt collar	Attribute_21
1	Blue	Shirt Hue	Attribute_22
1	Light	Shirt Light	Attribute_23
1	Intense	Shirt Saturation	Attribute_24
1	Vertical stripes	Shirt pattern	Attribute_25
1	Small	Shirt pattern size	Attribute_26
1	Low	Shirt pattern contrast	Attribute_27
2	Regular	Tie note	Attribute_28
1	Blue	Tie Hue	Attribute_29
3	Dark	Tie Light	Attribute_30
2	Dull	Tie Saturation	Attribute_31
3	Slant stripes	Tie pattern	Attribute_32
3	Large	Tie pattern size	Attribute_33
3	Large	Tie pattern contrast	Attribute_34
2	Unfinished	Square shape	Attribute_35
4	Yellow	Square Hue	Attribute_36
1	Light	Square Light	Attribute_37
1	Intense	Square Saturation	Attribute_38
1	Printed pattern	Square pattern	Attribute_39
4	j+s+t+p	Outfit-pairing	Attribute_40
3	Saturation	Outfit colors contrast	Attribute_41

1		Monochromatic	Outfit colors harmony	Attribute_42
1		Increase-body	Outfit shapes illusions	Attribute_43
4		4-type patterns	Outfit emphasis	Attribute_44
1		Jacket style	Outfit complexity	Attribute_45

An example of transferring apparel image to MEANING data:
MEANING (Image A) =[0,0,0,0,0,1,0,0,0,0,1]

Table 2 Generation of MEANING codes through a sample image

Codes	Representations of codes		
0	Negative	Casual style	Meaning_1
0	Negative	Elegant style	Meaning _2
0	Negative	Playful style	Meaning _3
0	Negative	Professional style	Meaning _4
0	Negative	Trendy style	Meaning _5
1	Positive	Classic style	Meaning _6
0	Negative	Circle body	Meaning _7
0	Negative	Invert triangle body	Meaning _8
0	Negative	Triangle body	Meaning _9
0	Negative	Square body	Meaning _10
1	Positive	Slim body	Meaning _11

3.2 Data training model design

The main contribution of this paper is the design of the apparel data training method. The purpose of model design is to 1) compare different training data between the knowledge-encoded apparel ATTRIBUTE dataset and the raw apparel image dataset, 2) compare different machine learning algorithms in apparel MEANINGs predictions and 3) eventually evaluate the feasibility of the designed model as an intelligent apparel recommendation model. According to the newly constructed datasets, three models with machine learning and deep learning combined methods are designed for data training and evaluation. As shown in figure 2, Model A is a machine learning model using ATTRIBUTE features to predict apparel MEANINGs. Model B is a deep learning model using image features to predict apparel MEANINGs. Model C is a two-step deep learning model using image-attribute collaborated features to predict apparel ATTRIBUTEs and MEANINGs.
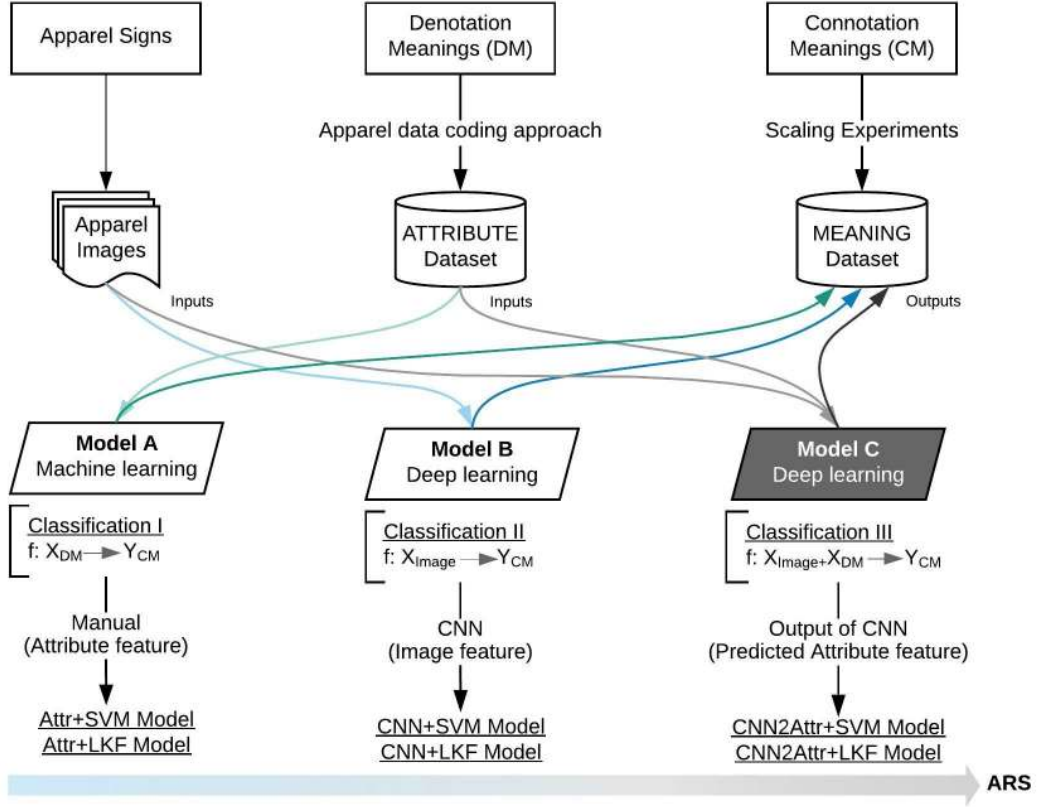
Figure 2 Methods design of three models based on new proposed apparel data

Model A

The design of model A aims to examine how accurate is the proposed ATTRIBUTE dataset as training data to predict apparel MEANINGs. SVM classifier is applied to predict apparel meanings by the training inputs of apparel attribute dataset. This method is denoted as Attr+SVM. SVM is a discriminative classifier. Apart from the linear classification, it could handle high dimensional feature spaces in a non-linear classification task. For individual cases where the number of attribute dimensions is greater than the number of samples, it is still effective by applying kernel functions to avoid over-fitting. However, if the training size is extremely small, the performance of conventional classifiers, e.g. SVM may suffer degradation.

Hence, this study proposes a new algorithm Later Kernel Fusion (LKF) that can effectively handle the small data. Instead of learning the class boundaries, LKF is a distribution based classification algorithm which focuses on maximising the likelihood function between the test samples and training classes, which can be measured by the Kullback-Leibler divergence:

$$\min_i KL\big(p(\hat{x})\|p(\hat{x}|y^c = i)\big) = p(\hat{x})\log\frac{p(\hat{x})}{p(\hat{x}|y^c = i)}, i \in \mathbb{B} \quad (1)$$

where $\hat{x}$ is a test sample. In other words, for each style c, we regroup the images according to whether they have the style. The prediction is made by minimizing the KL divergence to either the group of 'have'(1) or 'not have'(0) the style. Since $\hat{x}$ is a known constant that does not affect the *min* operator, we can reduce the prior $p(\hat{x})\log p(\hat{x})$. The KL divergence problem is reduced to a Maximum Likelihood classifier:

$$\min_i - p(\hat{x}) p(\hat{x}|y^c = i) = \max_i p(\hat{x}|y^c = i), i \in \mathbb{B} \quad (2)$$

In this study, due to the unbalanced distribution of samples in each labelled class, the weak classes with few samples may be sacrificed entirely for the denser ones. For instance, since the number of apparel samples in the

class of 'playful' is much larger than that of 'trendy', samples with 'trendy' may be sacrificed for 'playful'. In order to mitigate the unbalanced prior, Probability Density Estimation (PDE) (Raykar, 2002) is introduced to measure any arbitrary distribution from a set of training samples. First adapted PDE approach is Maximum Likelihood estimation with a kernel function K:

$$p(\hat{x}|y^c = i) = \frac{1}{|i|}\sum_{j=1}^{|i|} K\left(\hat{x} - x_j^i\right) \quad (3)$$

where $|i|$ is the cardinality, such as number of positive or negative samples; K is the Parzen Kernel Function which is non-negative and integrates to 1. In order to compare with SVM, same kernel RBF is adopted. However, due to the long-tail property, not all of the samples are effective for the PDE. Therefore, improved performance can be achieved by using the K-nearest neighbours of a test sample in the positive samples for PDE and compare that of negative samples. This method is denoted as Attr+LKF.

Model B

Model B is a deep learning model using image features to predict apparel MEANINGs. The development of model B is characterised as an image classification problem which is to categorise images into one of the several predefined classes. For image-based data, the traditional machine-learning methods request manually designed feature extractor to transform the raw data into a feature vector. By contrast, deep learning as a type of representation learning approach could automatically convert raw images into multi-layered representations from the first layer with raw data to higher layers with more abstract features (LeCun et al., 2015). The Convolutional Neural Network is the leading deep learning model to tackle image classification tasks. Rawat and Wang summarised the remarkable works of CNN from a single label, multi-label to semantic classification problems (Rawat and Wang, 2017).
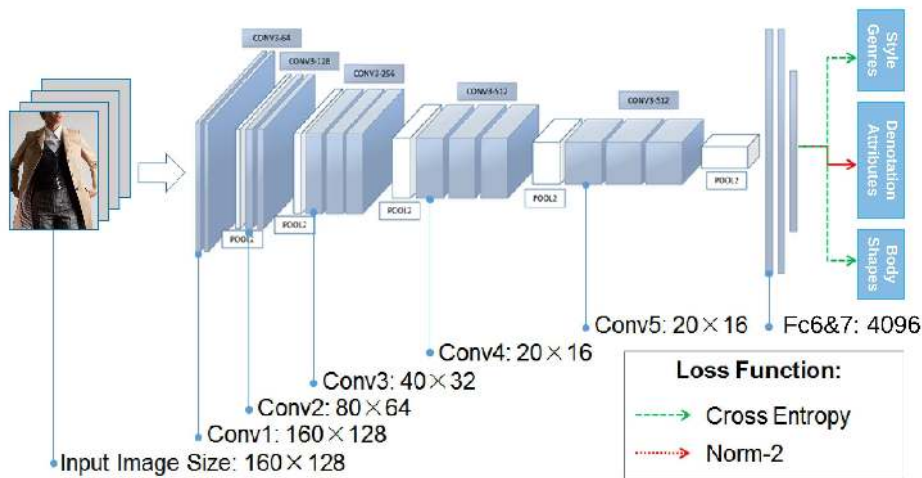

Figure 3 The structure of designed deep learning model. Conv: convolution layers; Fc: fully connection layers.

This study designed a deep CNN structure which serves as a multi-task model considering all of the three levels of predictions. Given a training set of images, this model aims to extract their visual image features X and predict the corresponding clothes MEANINGs including style genres, body shapes and ATTRIBUTEs, which are denoted as $f: X \rightarrow Y$, $f: X \rightarrow S$, and $f: X \rightarrow A$, respectively. Figure 3 demonstrates the main structure of the proposed CNN model. It is composed of five convolutional layers, five pooling layers and two fully connected layers. The design from the first convolutional layer to the last pooling layer serves for the feature learning. Each convolution level has two layers in the same size, e.g. Conv1 has Conv1_1 and Conv1_2 in the size of 160 ×128. The max-pooling layers between two levels make the size shrink to half of the original size. After Conv5, all of the patches are concatenated and mapped to two layers of fully connection, Fc6 and Fc7. The resultant feature space is in $X = [x_n] \in \mathbb{R}^{N \times 4096}$, where $1, ..., n, ...N$ indexes the training samples.

The design from fully connected layers to the final outputs serves for classifications. There are two types of objective functions. For style genre and body shape predictions, the supervision is discrete values. Therefore, a binary vector is adopted to represent these labels: $Y = [y_n] \in \mathbb{B}^{N \times 6}$, and $S = [s_n] \in \mathbb{B}^{N \times 5}$. Suppose Mutual Information loss does not differ Y and S, the objective function, Cross Entropy for Y also applies for S:

$$\mathcal{L}_1(X, Y) = \frac{1}{N \times 6} \sum_{n=1}^{N} \sum_{c=1}^{6} \left[y_n^c \log \mathrm{p}(y_n^c | x_n) - \overline{y_n^c} \log\left(1 - \mathrm{p}(y_n^c | x_n)\right) \right] \quad (4)$$

where $y_n^c = 1$ denotes that the $n - th$ image possesses the style c, and $y_n^c = 0$ otherwise.

For clothing attributes, however, it is a non-linear regression problem. For example, the clothes 'has three buttons' are highly correlated to 'has four buttons' compared to 'has ten buttons'. Its objective function is based on the Norm-2 distance:

$$\mathcal{L}_2(X, A) = \frac{1}{N} \| \Phi(x_n) - a_n \|_2^F \quad (5)$$

where $\|.\|_F$ is the Forbenius Norm, and Φ denotes the non-linear projection between X and A. the final loss is the sum of the three losses:

$$\mathcal{L} = \lambda_1 \mathcal{L}(X, Y) + \lambda_2 \mathcal{L}(X, S) + \lambda_3 \mathcal{L}(X, A) \quad (6)$$

where the three λ balance the weight of each loss.

However, since the CNN model is based on a multi-task loss, empirically, the performance of each individual task may have been sacrificed for the sake of balanced overall loss. The possible solution is to keep the structure of feature learning parts and replace the loss functions in fully connected layers by other separate classification models. Therefore, the proposed CNN model is used only for feature extractions from apparel images. Then, a conventional classifier, Support Vector Machine is used to train two separate classification tasks of style genre and body shape. This method is denoted as CNN+SVM. Alternatively, an optimised classification algorithm Later Kernel Fusion is also applied for comparisons. This method is denoted as CNN+ LKF.

Model C

The third prediction model is based on collaborated data of joining apparel images and ATTRIBUTEs. The training is divided into two stages. The task in the first stage is to train feature extractors to recognise features that are discriminative to apparel-related attributes from visual images, and then train classifiers to output apparel attributes based on extracted features. This procedure is called semantic attributes learning in visual recognition. It usually learns obvious or literal attributes of apparel such as the shape of a collar. The challenging part for the proposed model is to allow high dimensional apparel attributes, including literal and design attributes to be predicted. In fact, the earlier proposed CNN model in model B has involved with attributes learning in a multi-task feature extraction and classification model. Hence, the method in the first stage of model C is using previous proposed CNN model to predict the output of attributes, which is denoted as CNN2Attr.

The next stage is using the predicted apparel attributes in first result to train the final model. Since CNN feature is obtained by learning the multi-task loss including style genres, body shapes and apparel attributes, the predicted attributes may have been adjusted for the loss of style genres and body shapes prediction. Therefore, using the predicted attributes from the output of CNN model could further benefit the style genres and body shapes prediction, i.e. better than that of adopting original image or encoded attribute data separately. Eventually, to feed the predicted attributes to separate classifiers SVM and LKF is the completion of the third prediction model of apparel meanings, which are denoted as CNN2Attr+SVM and CNN2Attr+LKF.

4. Results Comparisons

Model A

The training of three proposed models is implemented by Python script. The training input and output data in model A are the apparel ATTRIBUTE dataset and MEANING label dataset respectively. The training results indicate the prediction accuracy of proposed two methods in model A. Table 3 shows the predictability of predefined eleven classes, including Casual, Elegant, Playful, Professional, Trendy and Classic in style genres and Circle, Invert Triangle, Triangle, Square and Slim in body shapes, as well as the average predictability on overall classes. The values indicate the percentage of samples that are correctly predicted by Attr+SVM and Attr+LKF for each class, which are between 50% (random guess) and 100% (entirely predictable).

Table 3 Prediction accuracy of Model A: SVM vs LKF

		Predictability of Classes (style genres &body shapes) (%)										
		Casual	Elegant	Playful	Professional	Trendy	Classic	Circle	Invert triangle	Triangle	Square	Slim	Mean
Model A	Attr+SVM	79.1	85.2	76.5	83.2	91.2	86.4	88.1	83.8	82.8	90.9	93.0	85.5
	Attr+LKF	83.4	85.2	79.8	86.6	92.3	86.9	89.5	85.6	84.9	91.7	93.4	87.2

The distributions of prediction accuracy on eleven classes are differentiated widely from "*Playful*" (76.5% and 79.8%) to "*Slim*" (93.0% and 93.4%). To compare the two methods, the performance of the originally designed LKF classifier is slightly better than conventional classifier SVM in model A, with 1.7% average improvement. The reason to cause low accuracy in some classes is from either the feature extraction or classification methods. Manually encoded apparel attributes may not be discriminative to the predefined classes of apparel meanings. Alternatively, the classifiers, SVM and LKF may not effective on classifications of these classes. For further investigation, next model will adopt same classifiers to test the other feature data that were automatically extracted from apparel images.

Model B

The training input and output data in model B are the apparel image dataset and MEANING label dataset respectively. To compare the training results between model A and model B (Table 4), it is found that the image-based prediction model is less effective than the ATTRIBUTE-based prediction model in all classes of MEANINGs. That is to say, the ATTRIBUTE-based feature data are more precise than the Image-based feature data so that the predictability is better. While, in terms of the training algorithms, the LKF in image-based model makes the average prediction rate increased (10.8%, from SVM to LKF) more dramatically than in attribute-based model (1.7%, from SVM to LKF). So the image-based deep learning model CNN combined with the separate LKF classifier is the more efficient method. The next model will take both the ATTRIBUTE-based data and image-based data into account with SVM and LKF algorithms applied in models A and B.

Table 4 Predictability of Model A vs Model B

		Predictability of Classes (style genres &body shapes) (%)											
		Casual	Elegant	Playful	Professional	Trendy	Classic	Circle	Invert triangle	Triangle	Square	Slim	Mean
Model A	Attr+SVM	79.1	85.2	76.5	83.2	91.2	86.4	88.1	83.8	82.8	90.9	93.0	85.5
	Attr+LKF	83.4	85.2	79.8	86.6	92.3	86.9	89.5	85.6	84.9	91.7	93.4	87.2
Model B	CNN+SVM	72.9	72.6	70.1	76.2	77.5	70.7	75.8	69.5	64.4	79.7	79.2	73.5
	CNN+LKF	82.4	83.8	77.9	84.6	91.2	85.7	84.4	82.5	79.6	86.4	89.2	84.3

Model C

Table 5 demonstrates the results of predictability of model C and compares the results with that from previous two models. By introducing CNN2Attr method to extract feature data, the performances further improved particularly for those classes received lower rate in previous two models. Specifically, with the LKF classifier, the accuracy of model C is higher than other two models in all classes. For instance, CNN2Attr+LKF achieves 87.8% recognition rate in the class of "*Triangle*", while Attr+LKF is 84.9% and CNN+LKF is only 79.6%. With the SVM classifier, some classes received remarkable improvements such as the prediction rate on "*Trendy*" from 77.5% by CNN+SVM method to 91.8% by CNN2Attr+SVM method. Therefore, the design of model C is the best solution to tackle the new apparel data and apparel recommendation system.

Table 5 Predictability of Model A vs Model B vs Model C

		Predictability of Classes (style genres &body shapes) (%)											
		Casual	Elegant	Playful	Professional	Trendy	Classic	Circle	Invert triangle	Triangle	Square	Slim	Mean
Model A	Attr+SVM	79.1	85.2	76.5	83.2	91.2	86.4	88.1	83.8	82.8	90.9	93.0	85.5
	Attr+LKF	83.4	85.2	79.8	86.6	92.3	86.9	89.5	85.6	84.9	91.7	93.4	87.2
Model B	CNN+SVM	72.9	72.6	70.1	76.2	77.5	70.7	75.8	69.5	64.4	79.7	79.2	73.5
	CNN+LKF	82.4	83.8	77.9	84.6	91.2	85.7	84.4	82.5	79.6	86.4	89.2	84.3
Model C	CNN2Attr+SVM	79.6	85.1	76.9	84.3	91.8	86.9	88.4	84.5	83.8	91.4	93.2	86.0
	CNN2Attr+LKF	84.5	85.4	80.1	88.5	92.4	87.1	90.4	86.3	87.8	92.6	93.5	**88.1**

Based on the result comparisons on three-models', it is found that firstly the proposed a series of visual design attributes of apparel are recognisable from images by the CNN2Attr method and secondly, the recognised apparel attributes are the most precise feature data to adapt the tasks of apparel meaning classifications. Regarding classification algorithms, the proposed LKF is more efficient than the baseline model SVM in particular on image-based model B, with 10.8% higher prediction rates. Overall, if a prediction rate of 80% deserves a good model, then all proposed methods in model A, B and C are effective with one exception of CNN+SVM which is below 80% in all classes as highlighted. In summary, model C is the most applicable model for prototyping apparel recommendation systems with the method of CNN2Attr+LKF which could reach the prediction accuracy of 88.1% in average.

5. Evaluation of Model C in developing an apparel style recommendation system

In this study, the development of knowledge-based apparel recommendation system requests a specific machine learning model which could understand the implicit meanings or rules behind the visual signs or characters of clothes. A well-trained model could generally learn a wide category of clothes since the features are obvious, natural and easily recognisable based on common sense such as the long dresses for red carpet and suits for conference. The learning challenge of the proposed model is to distinguish apparel samples with closer features that are even hard to achieve by general public except of the trained style professionals. For example, a men's blazer with plain, textured or printed patterns could change the feelings of style.

Our solution is to mimic human visual perception based on the theory of apparel signs, denotation and connotation meanings in apparel communication system. To understand clothes is a step-by-step process which is to recognise obvious visual features, deep features and eventually to perceive the meanings affected by these features. For example, a combination of lines (sign) is recognised to be a plaid pattern, and plaid pattern is more discriminative to classic style than other patterns. The design of model C is in accordance with this criterion. This section is to further evaluate whether this model could be used as a smart recommendation engine that matches online clothes with specific meanings or perceptual concepts that people are willing to express through the dress.

The evaluation of MEANING prediction model C shows the results of recommendations on style genres and body shapes. Suppose the user is looking for a style that feels elegant, the recommendation engine gives suggests/ not suggests on 10 random testing samples. All in all, the new model could support apparel recommendation on eleven classes of apparel-associated meanings including style genres and body shape corrections, with average 88.1% of recommendation results are in accordance with actual results given by real style professionals. In real-life applications, additional errors need to be considered. In this model, the ground truth apparel MEANINGs are labelled by individual style experts. Although this group of people has better knowledge and skill in clothes and fashion, however, the judgments may not fully social agreeable. The perceptual differentiations cause such unpredictable errors. Regarding occasional recommendations, it is suggested by stylists that the occasions are highly related to style genres, such as classic style for racing, elegant for opera, working dinner etc., playful for a stage party or going out and so on. The proposed model could train occasion-associated output data and further demonstrate the correlations between style genre and occasion recommendations.

6. Discussion and future work

As an exploratory study, the machine learning methodology and apparel data construction methods are new for apparel-oriented smart systems and visual recognition study. Proposed apparel data involve with a broad scope of data format and apparel knowledge, including three datasets: a semantic attribute dataset representing natural and visual design features, a visual image dataset from real shopping platforms and a proficient labelled dataset by clothes experts. In terms of the training model, this study challenges the dominant method of deep learning to tackle a particular visual recognition task which is the visual perceptions of internal apparel meanings, not only the natural recognisable visual appearances. For the particular data and classification tasks, the proposed methodology examined every possibility to collaborate semantic attribute and visual images as complementary learning data, as well as the advantages of conventional machine learning and deep learning algorithms on feature extractions and classifications.

This study addressed three kinds of apparel learning models. The third model that joins Convolutional Neural Network and Later Kernel Fusion methods could significantly enhance the capability of smart apparel systems regarding apparel attribute extraction and apparel recommendation. The multi-task CNN model could recognise 45 valuable semantic attributes associated with literal and visual design. And, the predictability of internal apparel meanings is higher than the general image-based deep learning model and the attribute-based machine learning model. However, the proposed system is currently limited by the learning categories on menswear, the

relatively small example size, unbalanced distribution of training samples, as well as the apparel meanings classes.

In the future, the targeted apparel could be extended to other categories, e.g. women's dresses, beyond the investigated menswear categories. The performances of apparel attribute and meaning prediction models are improvable by increasing the size of data and balancing the distribution of data. The functions of apparel recommendation system are decided by the learning outputs of apparel meaning prediction model which is the labels of meaning classes. Hence, the improvements are achievable by introducing more predefined classes of apparel meanings, e.g. dandy, understated style or body, hair colours and collecting more labelled data accordingly. Further study could examine above occasion classes as a parallel task in prediction models.

In addition, the current study reinterpreted the theory of signs, semantics and visual design as the in-depth knowledge of clothing to support the constructions of apparel datasets and learning models. However, the exploration of this study should not be limited to these early-identified theories in fashion and clothing field. The latest concepts and issues in fashion research, such as clothing and gender neutrality could be addressed in further developments of apparel learning models.

7. Conclusion

This paper presents a new apparel-based deep learning system design and evaluates three new models designed for knowledge-based apparel recommendation system. The design of three models is based on new training data including ATTRIBUTEs, raw apparel images and trained ATTRIBUTEs by images. The applied training methods are widely ranged from traditional machine learning to the up-to-date deep learning. A multi-task CNN deep model, SVM and a newly proposed LKF classifier were introduced to train the three datasets. The model testing results confirmed that the proposed method of CNN2Attr+LKF in the third model could produce highly accurate predictions on apparel MENINGs, with 80.1% to 93.5% results could match the clothes experts' judgments. Applying this method in apparel style recommendation systems is technically feasible. It could achieve two applications. The ATTRIBUTE prediction model can be used for Menswear advanced search system that allows users to filter not only the natural but deep design features. The MEANING prediction model can be used as a Menswear style recommendation engine that distinguishes online clothes by specific style genres or body shapes.

8. Reference

AMAZON. 2017. *Echo Look | Hands-Free Camera and Style Assistant with Alexa—includes Style Check to get a second opinion on your outfit* [Online]. Available: https://www.amazon.com/Amazon-Echo-Look-Camera-Style-Assistant/dp/B0186JAEWK [Accessed 22 January 2018].

ANONYMOUS 2016. Details omitted for double-blind reviewing.

ANONYMOUS forthcoming 2018. Details omitted for double-blind reviewing.

BRACHER, C., HEINZ, S. & VOLLGRAF, R. 2016. Fashion DNA: Merging Content and Sales Data for Recommendation and Article Mapping. *arXiv preprint arXiv:1609.02489*.

CHEN, H., GALLAGHER, A. & GIROD, B. 2012. Describing clothing by semantic attributes. *Computer Vision–ECCV 2012*, 609-623.

CHENG, C. I. & LIU, D. S. 2008. An Intelligent clothes search system based on fashion styles *International Conference on Machine Learning and Cybernetics*. P.R.China: IEEE

DE BARROS COSTA, E., ROCHA, H. J. B., SILVA, E. T., LIMA, N. C. & CAVALCANTI, J. 2017. Understanding and Personalising Clothing Recommendation for Women. *In:* ROCHA, Á., CORREIA, A. M., ADELI, H., REIS, L. P. & COSTANZO, S. (eds.) *Recent Advances in Information Systems and Technologies: Volume 1.* Cham: Springer International Publishing.

DRESSIPI. 2018. *Personalising the Entire Customer Experience* [Online]. Available: https://dressipi.com/ [Accessed 24 January 2018].

HARA, K., JAGADEESH, V. & PIRAMUTHU, R. Fashion apparel detection: the role of deep convolutional neural network and pose-dependent priors. Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, 2016. IEEE, 1-9.

HARTMANS, A. 2017. *Amazon's new Echo device is a hands-free camera that helps you decide what to wear* [Online]. Available: http://uk.businessinsider.com/amazon-look-camera-outfit-analysis-2017-4 [Accessed 22 January 2018].

HIDAYATI, S. C., CHENG, W.-H. & HUA, K.-L. Clothing genre classification by exploiting the style elements. Proceedings of the 20th ACM international conference on Multimedia, 2012. ACM, 1137-1140.

KALANTIDIS, Y., KENNEDY, L. & LI, L.-J. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, 2013. ACM, 105-112.

LANDIA, N. Building Recommender Systems for Fashion: Industry Talk Abstract. Proceedings of the Eleventh ACM Conference on Recommender Systems, 2017. ACM, 343-343.

LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature,* 521**,** 436-444.

LI, R., FENG, F., AHMAD, I. & WANG, X. 2017. Retrieving real world clothing images via multi-weight deep convolutional neural networks. *Cluster Computing***,** 1-12.

LIU, Z., LUO, P., QIU, S., WANG, X. & TANG, X. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1096-1104.

PERKINIAN, C. & VIKKRAMAN, P. 2015. An intelligent apparel recommendation system for online shopping using style classification. *International Journal of Applied Business and Economic Research* 13**,** 671-686.

PERLIN, H. A. & LOPES, H. S. 2015. Extracting human attributes using a convolutional neural network approach. *Pattern Recognition Letters,* 68**,** 250-259.

RAWAT, W. & WANG, Z. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*.

RAYKAR, V. C. 2002. Probability density function estimation by different methods. *Enee 739Q Spring***,** 1-8.

SUN, Y. 2017. *Alibaba's AI Fashion Consultant Helps Achieve Record-Setting Sales* [Online]. Available: https://www.technologyreview.com/s/609452/alibabas-ai-fashion-consultant-helps-achieve-record-setting-sales/ [Accessed 22 January 2018].

VACCARO, K., SHIVAKUMAR, S., DING, Z., KARAHALIOS, K. & KUMAR, R. The Elements of Fashion Style. Proceedings of the 29th Annual Symposium on User Interface Software and Technology, 2016. ACM, 777-785.

VURUSKAN, A., INCE, T., BULGUN, E. & GUZELIS, C. 2015. Intelligent fashion styling using genetic search and neural classification. *International Journal of Clothing Science and Technology,* 27**,** 283-301.

WARD, M. 2016. *Introducing Style Detection for Google Cloud Vision API* [Online]. Available: https://cloud.googleblog.com/2016/04/introducing-Style-Detection-for-Google-Cloud-Vision-API.html [Accessed 27 September 2017].