

Visual BMI estimation from face images using a label distribution based method

Min Jiang^a, Guodong Guo^{a,*}, Guowang Mu^b

^a Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA

^b School of Science, Hebei University of Technology, Tianjin 300401, China

ARTICLE INFO

Communicated by Stefanos Zafeiriou

Keywords:

Visual BMI estimation

Label distribution based method

Two-stage learning framework

ABSTRACT

Body mass index (BMI) analysis from face images is an interesting and challenging topic in machine learning and computer vision. Recent research shows that facial adiposity is associated with BMI prediction. In this work, we investigate the problem of visual BMI estimation from face images by a two-stage learning framework. BMI-related facial features are learned from the first stage. Then a label distribution based BMI estimator is learned by an optimization procedure that is implemented by projecting the features and assigned labels to a new domain which maximizing the correlation between them. Two label assignment strategies are analyzed for modeling the single BMI value as a discrete probability distribution over a range of BMIs. Extensive experiments are conducted on FIW-BMI, Morph II and VIP_attribute datasets. The experimental results show that the two-stage learning framework improves the performance step by step. More importantly, the proposed BMI estimator efficiently reduces the error. It outperforms regression based methods, two label distribution methods and two deep learning methods in most cases.

1. Introduction

Recent research shows that facial adiposity is associated with perceived health and is important for body mass index (BMI) prediction (Coetzee et al., 2009; Wolffechehel et al., 2015). As a body fat indicator, BMI is widely used in health monitoring and health research. There are close connections between BMI and some diseases, such as cancers, unstable angina and type 2 diabetes and cardiovascular disease (CVD), etc. (Arnold et al., 2016; Wolk et al., 2003; Meigs et al., 2006). Generally, BMI is measured in person with special devices. Therefore, automatically estimating BMI from face images is a great benefit to health monitoring and researchers who are interested in studying obesity in large populations.

BMI estimation from face images is a challenging problem. First, different from other human visual tasks, such as face recognition (Guo et al., 2016; Hjelms and Low, 2001; Yu et al., 2019), motion capture (Moeslund and Granum, 2001; Moeslund et al., 2006) which have sufficient data for training and testing, it is difficult to collect a database covering images with all BMI values. Second, the distribution of BMIs on the database is uneven. According to the BMI values, there are mainly four BMI categories: underweight ($BMI \leq 18.5$), normal ($18.5 < BMI \leq 25$), overweight ($25 < BMI \leq 30$), obese ($BMI > 30$). Very few BMIs distribute on underweight and severe obese categories. It is hard to ensure each category have enough associated images. Currently, the number of public databases for visual BMI study is limited. This

work uses three databases: Morph II (Ricanek and Tesafaye, 2006), FIW-BMI (Jiang et al., 2019) and VIP_attribute (Dantcheva et al., 2018). According to the above two reasons, we can see it is difficult to estimate BMI with limited training data. Finally, the BMI label is ambiguous. e.g., one person looks like with BMI around 25 which means that some neighbor values (24.5 or 25.5) can also be used to describe this person; and some people may look lower than their real BMI, while others may look higher than their real BMI. Fig. 1 shows some face images from Morph II and FIW-BMI datasets with corresponding BMI values. We can see some samples with the same gender and adjacent BMI values but have different facial appearances.

Single label estimation assumes one image has one label. Regression based methods directly predict the label from the images which ignores the ambiguous label that may existing in images. To describe the ambiguity associated with the labels of images, a label distribution scheme is proposed by Geng et al. (2013) to describe such ambiguity. Later on, other distribution learning based approaches have been proposed for age estimation and other tasks. These methods utilized label correlation or entropy model to solve the problem. Geng et al. (2013) proposed two label distribution based algorithms named IIS-LLD and CPNN. Comparing with other single label methods, their methods showed good performances. A multivariate label distribution (MLD) based method was also proposed by Geng and Xia (2014) for further improving the performance on head pose estimation. In addition, Xing et al. (2016)

* Corresponding author.

E-mail addresses: minjiang.aca@gmail.com (M. Jiang), guodong.guo@mail.wvu.edu (G. Guo), muguow@163.com (G. Mu).

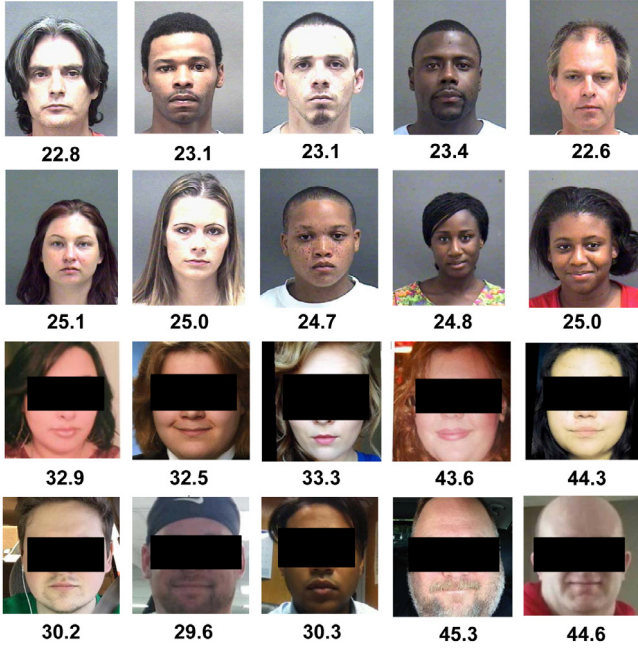


Fig. 1. Samples from Morph II and FIW-BMI datasets with corresponding BMI values. Because the images from FIW-BMI dataset are collected from social networks, here we used a black box to thinly hide the identity of each subject.

used Logistic Boosting Regression (LogitBoost) to learn a general label distribution model family which can avoid the potential influence of the specific model.

Some work explored regression based methods for visual BMI estimation. Wen and Guo (2013) proposed a computational method for automatically predicting BMI from 2D face images. This is the first work on visual BMI estimation from face images. Kocabey et al. (2017) employed the pre-trained VGG-Face model (Parkhi et al., 2015) to extract facial representation for BMI estimation. Then a support vector regression model is learned to map the facial representation to predicted BMIs. The above two works treated BMI prediction as a regression problem. Recently, convolutional neural networks (CNN) have shown promising performance in many applications (Krizhevsky et al., 2012; Schroff et al., 2015; Cao et al., 2017; Wang et al., 2017). A method using CNN for BMI estimation is proposed by Dantcheva et al. (2018), where estimating height, weight, and BMI from single-shot face images by a regression method based on the 50-layers ResNet-architecture. Recently, Guo and Jiang (2020) gave a brief overview of related methods for visual BMI estimation, and presented various potentials of the developed techniques towards practical applications.

Different from the above work, this work addresses the visual BMI estimation problem by a label distribution based method. The label distribution scheme provides a feasible method to learning the estimator with limited training data. In addition, the label distribution scheme well defines the increase and decrease of BMI as a “continuous” process. More specifically, a two-stage learning framework is shown in Fig. 2. First, the BMI-related facial representation is learned by fine-tuning the pre-trained deep face model. This step is expected to obtain sufficient visual BMI characteristics and reinforce the learning process using the limited number of BMI data. More importantly, the label distribution method models the single BMI value as a discrete probability distribution over the whole ranges of BMIs. Given the extracted facial features from the first stage, a BMI estimator is trained by an optimization procedure that is applied to the features and the assigned distribution labels. The main contributions of this work include:

1. A two-stage learning framework is presented to address the visual BMI estimation problem from face images.

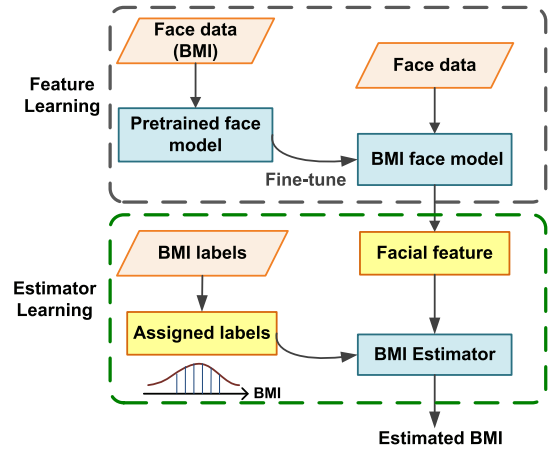


Fig. 2. The pipeline of two-stage learning framework. It consists of BMI-related feature learning and BMI estimator learning.

2. A label distribution based learning method that regards a BMI value (label) as a discrete probability distribution is proposed to learn the BMI estimator.
3. Two label assignment strategies are analyzed to model BMI labels. The output can either be a discrete probability distribution or a single value.

The remaining of the paper is organized as follows. The details of the proposed method for BMI estimation are presented in Section 2. Section 3 describes the three databases used in this work: Morph II, FIW-BMI and VIP_attribute. In Section 4, first, we describe the evaluation metrics and experimental setting; and then we provide the detailed experimental results and discussion. Finally, the conclusion is summarized in Section 5.

2. Method

Fig. 2 depicts the two-stage learning framework, which consists of BMI-related facial features learning and BMI estimator learning. The BMI-related face model is learned based on a pre-trained deep face model. Then two different strategies are analyzed for modeling the BMI labels with probability distributions. And a projection optimization is achieved by maximizing the correlation between the facial features and the assigned labels. Finally, the BMI estimator is learned from the projected features and assigned labels. Below the detailed procedure and derivation are presented.

2.1. Deep model for BMI-related facial feature

(1) *Face model*: The face structure is represented by a pre-trained face model. We utilize the feature extracted from a publicly released Centerloss face model (Wen et al., 2016). This network improves the discriminative power of the learned features by using a Centerloss function to minimize the intra-class variations while maximizing the inter-class variations by the softmax function. This model performs impressively in face recognition tasks which achieved face verification accuracy of 98.28% on Labeled Faces in the Wild Dataset, 94.9% on YouTube Faces Database. The fully connected layer 5 (fc5) of Center loss model C is used to extract facial features.

(2) *Fine-tuning*: This step is implemented by adapting from the general facial structure model to the BMI-related face model. Our goal is to estimate BMI values from face images. We tune the pre-trained Centerloss face model to the BMI face model before extracting features. FIW-BMI dataset is used to fine-tune the deep model. This step aims to learn sufficient BMI-related facial features. We replace the last 512-dimension fully connected layer (fc) with 1-dimension fc, and use

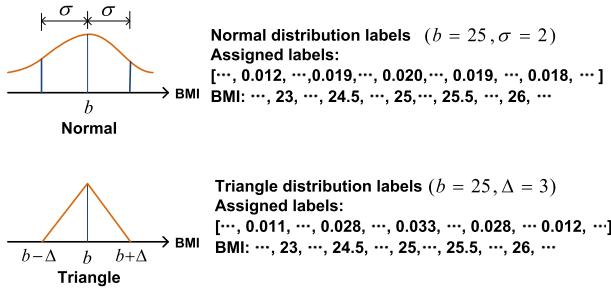


Fig. 3. Two strategies for modeling the single BMI value.

the Euclidean loss function to fine-tune the network. Euclidean loss function E computes the sum of squares of differences between two inputs, which can be written as:

$$E = \frac{1}{2N} \sum_{i=1}^N \|\hat{y}_i - y_i\|^2, \quad (1)$$

where N is the number of samples, $\hat{y}_i$ is the output from the network and y_i is the true BMI value. After the above steps, the fine-tuned face model is expected to have the capability to capture more BMI-related facial structures.

2.2. Modeling BMI values with label distribution

BMI value is labeled by a real number. BMI estimation from face images is different from other traditional regression tasks because there is ambiguous information among BMI labels. Based on this observation, we define a single BMI value as a discrete probability distribution over the range of BMIs. There are two advantages of the label distribution scheme. First, given one face image, its corresponding label distribution consists of a set of probabilities. Each probability represents the confidence that the corresponding label describes the image. The largest probability is corresponding to the true BMI label (value). With this scheme, one image not only contributes to the learning of its true BMI label but also provides auxiliary information to learn its adjacent BMI labels. Second, the label distribution scheme well defines the increase and decrease of BMI as a process. Similar definition is proposed by Geng et al. (2013) for age estimation. Different from human age, there is no exact range for BMIs. According to the BMI distribution of the two databases, we assume the BMI range is from 13 to 60 in this work. Two strategies are investigated for modeling the single BMI value, namely Gaussian distribution, and triangle distribution, which are shown in Fig. 3.

Specifically, Given an image labeled with the BMI value b , the BMI label is transformed to a discrete probability distribution $\mathbf{p} = [p_1, p_2, \dots, p_k]^T \in \mathbb{R}^k$ over the whole range of BMIs which follows a Gaussian distribution centered at b :

$$p_i = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z_i - b)^2}{2\sigma^2}\right), \quad 13 \leq z \leq 60, \quad (2)$$

where σ is the standard deviation of the Gaussian distribution. And $\mathbf{z} = [z_1, z_2, \dots, z_k]^T \in \mathbb{R}^k$ is a set of discrete values from 13 to 60 with an interval of 0.1.

For the triangle distribution, the neighbor BMIs are considered with a length of Δ on each side of the BMI value. The probability $p(z_i)$ is computed by:

$$p_i = \begin{cases} \frac{\Delta - b + z_i}{\Delta^2}, & \text{if } b - \Delta \leq z_i \leq b \\ \frac{\Delta + b - z_i}{\Delta^2}, & \text{if } b \leq z_i \leq b + \Delta \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

A normalization process is applied to the assigned labels, which defined as: $y_i = p_i / \sum_i p_i$. This leads to a discrete range of BMIs with different levels of “probabilities”: $\sum_i y_i = 1$.

2.3. Learning with assigned labels

The BMI estimator is learned by a label optimization procedure based on the correlation. The optimization procedure is implemented by projecting the features and assigned labels to a new domain which maximizes the correlation between them. Then the BMI estimator is learned from the projected features and labels as a least square problem.

Given N training samples, denote $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^d]^T \in \mathbb{R}^d$, $\mathbf{y}_i = [y_i^1, y_i^2, \dots, y_i^k]^T \in \mathbb{R}^k$ as the feature vector and assigned distribution label of i th training sample, respectively. Here d is the dimension of the feature vector $\mathbf{x}_i$, and k is the length of the assigned label $\mathbf{y}_i$. We assume that both $\mathbf{x}_i$ and $\mathbf{y}_i$ are centered, i.e., $\sum_{i=1}^N \mathbf{x}_i = 0$ and $\sum_{i=1}^N \mathbf{y}_i = 0$. Denote $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{k \times N}$, the main idea of canonical correlation analysis (CCA) (Thompson, 2005) is to project the two sets of variables into latent variables (a new domain), such that the correlation ρ between them is maximized, which can be written as:

$$\rho = \max_{\mathbf{w}_X, \mathbf{w}_Y} \frac{\mathbf{w}_X^T C_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T C_{XX} \mathbf{w}_X \mathbf{w}_Y^T C_{YY} \mathbf{w}_Y}}, \quad (4)$$

here $\mathbf{w}_X$ and $\mathbf{w}_Y$ are projection vectors. Observe that the solution of Eq. (4) is invariant to re-scaling $\mathbf{w}_X$ or $\mathbf{w}_Y$ either together or independently:

$$\frac{\alpha \mathbf{w}_X^T C_{XY} \mathbf{w}_Y}{\sqrt{\alpha^2 \mathbf{w}_X^T C_{XX} \mathbf{w}_X \mathbf{w}_Y^T C_{YY} \mathbf{w}_Y}} = \frac{\mathbf{w}_X^T C_{XY} \mathbf{w}_Y}{\sqrt{\mathbf{w}_X^T C_{XX} \mathbf{w}_X \mathbf{w}_Y^T C_{YY} \mathbf{w}_Y}}. \quad (5)$$

The solution of Eq. (4) is only related to the direction of the two projection vectors $\mathbf{w}_X$ and $\mathbf{w}_Y$. To obtain a unique solution, the constraints are added. Thereby, the CCA problem is equivalent to maximizing the following problem:

$$\begin{aligned} \max_{\mathbf{w}_X, \mathbf{w}_Y} & \mathbf{w}_X^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_Y, \\ \text{s.t.} & \mathbf{w}_X^T \mathbf{X} \mathbf{X}^T \mathbf{w}_X = 1, \mathbf{w}_Y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_Y = 1. \end{aligned} \quad (6)$$

Using the Lagrangian multiplier:

$$\begin{aligned} L(\lambda, \mathbf{w}_X, \mathbf{w}_Y) = & \mathbf{w}_X^T C_{XY} \mathbf{w}_Y - \frac{\lambda_X}{2} (\mathbf{w}_X^T C_{XX} \mathbf{w}_X - 1) \\ & - \frac{\lambda_Y}{2} (\mathbf{w}_Y^T C_{YY} \mathbf{w}_Y - 1). \end{aligned}$$

Taking derivatives of $\mathbf{w}_X$ and $\mathbf{w}_Y$, respectively:

$$\frac{\partial L}{\partial \mathbf{w}_X} = C_{XY} \mathbf{w}_Y - \lambda_X C_{XX} \mathbf{w}_X = \mathbf{0}, \quad (7)$$

$$\frac{\partial L}{\partial \mathbf{w}_Y} = C_{YX} \mathbf{w}_X - \lambda_Y C_{YY} \mathbf{w}_Y = \mathbf{0}. \quad (8)$$

Then subtracting $\mathbf{w}_Y^T$ multiplies Eq. (8) from $\mathbf{w}_X^T$ multiplies Eq. (7):

$$\begin{aligned} \mathbf{0} = & \mathbf{w}_X^T (C_{XY} \mathbf{w}_Y - \lambda_X C_{XX} \mathbf{w}_X) - \mathbf{w}_Y^T (C_{YX} \mathbf{w}_X - \lambda_Y C_{YY} \mathbf{w}_Y) \\ = & \lambda_Y \mathbf{w}_Y^T C_{YY} \mathbf{w}_Y - \lambda_X \mathbf{w}_X^T C_{XX} \mathbf{w}_X. \end{aligned}$$

Taking into account the constraints $\mathbf{w}_X^T \mathbf{X} \mathbf{X}^T \mathbf{w}_X = 1$ and $\mathbf{w}_Y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_Y = 1$, we can obtain that $\lambda_Y - \lambda_X = 0$. Let $\lambda = \lambda_Y = \lambda_X$ and assuming C_{YY} is invertible, we can obtain:

$$\mathbf{w}_Y = \frac{C_{YY}^{-1} C_{YX} \mathbf{w}_X}{\lambda}. \quad (9)$$

substituting Eq. (9) into Eq. (7):

$$C_{XY} C_{YY}^{-1} C_{YX} \mathbf{w}_X = \lambda^2 C_{XX} \mathbf{w}_X. \quad (10)$$

Now Eq. (10) is a generalized eigenvalue problem of the form $Ax = \lambda Bx$. $\mathbf{w}_X$ can be obtained via solving the following generalized eigenvalue problem. To avoid the singularity problem of $\mathbf{Y} \mathbf{Y}^T$ and $\mathbf{X} \mathbf{X}^T$, we adopt regularized CCA to get $\mathbf{w}_X$ by the following form:

$$C_{XY} (C_{YY} + \eta_Y \mathbf{I})^{-1} C_{YX} \mathbf{w}_X = \lambda^2 (C_{XX} + \eta_X \mathbf{I}) \mathbf{w}_X. \quad (11)$$

Table 1

Characteristics of FIW-BMI dataset. Mean and standard deviations pertained to BMI for male and female groups.

	Male	Female
#Subjects	3192	1689
#Images	5197	2733
Mean	30.7	31.2
Std	7	6.9

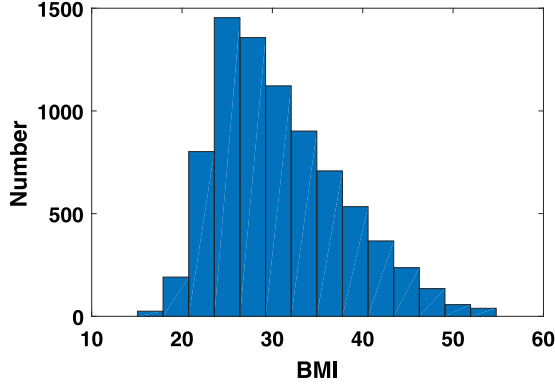


Fig. 4. Distribution of BMI values on BMI-analysis face database. The BMI values span a wide range with most of the values distribute between 20 to 50.

Let $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q]$ denotes the matrix of top q eigenvectors of the generalized eigenvalue problem. Here $\mathbf{W}$ is the projection vector which is used to project $\mathbf{X}$ into a new domain, such that the correction ρ is maximized. For each original feature vector $\mathbf{x} \in \mathbb{R}^d$, we obtain the new representation $\mathbf{x}^{CCA} = \mathbf{W}^T \mathbf{x}$.

After obtaining the new representations of all N training samples $\mathbf{x}_i^{CCA} = \mathbf{W}^T \mathbf{x}_i$, we can obtain the BMI distribution by solve the following least square problem:

$$\min_B \sum_{i=1}^N \|\mathbf{x}_i^T \mathbf{W} \mathbf{B} - \mathbf{y}_i^T\|, \quad (12)$$

where $\mathbf{B} \in \mathbb{R}^{q \times k}$ is a coefficient matrix, which can be shown that the solution to Eq. (12) is:

$$\mathbf{B} = (\mathbf{X}^T \mathbf{W})^+ \mathbf{Y}^T, \quad (13)$$

where $(\mathbf{X}^T \mathbf{W})^+$ denotes the pseudo-inverse of $\mathbf{X}^T \mathbf{W}$. Given a test sample (feature) $\mathbf{x}_t$, the corresponding estimated assigned label distribution can be obtained by:

$$\hat{\mathbf{y}} = \mathbf{B}^T \mathbf{W}^T \mathbf{x}_t, \quad (14)$$

here $\hat{\mathbf{y}} = [\hat{y}^1, \hat{y}^2, \dots, \hat{y}^k]$ is a vector denotes the predicted probabilities distribution. $\hat{y}^i$ is a factor in vector $\hat{\mathbf{y}}$ which denotes the predicted probability belongs to the BMI label z_i . Then the estimated BMI value $\hat{b}$ is computed by:

$$\hat{b} = \sum_{i=1}^k \hat{y}^i z_i. \quad (15)$$

$\mathbf{z} = [z_1, z_2, \dots, z_k]$ is a set of discrete BMI values from 13 to 60 with the interval of 0.1.

3. Dataset

We conduct extensive experiments on three datasets. FIW-BMI dataset (Jiang et al., 2019) is used to fine-tune the deep face model. Morph II dataset (Ricanek and Tesafaye, 2006) and VIP_attribute dataset (Dantcheva et al., 2018) are utilized to evaluate the effectiveness of the proposed method.

Table 2

Characteristics of selected data from Morph II. Mean and standard deviations pertained to BMI for four gender and ethnicity groups.

	Black male	Black female	White male	White female
#Subjects	6497	1096	1565	535
#Images	19 290	2824	4862	2057
Mean	25.0	25.2	24.6	22.8
Std	4.4	6.0	4.0	5.5

Table 3

Splitting selected Morph II by gender and ethnicity.

	Training set		Test set	
	#Subject	#Images	#Subject	#Images
Black male	4568	13 574	1929	5716
Black female	873	2218	223	606
White male	1245	3856	320	1006
White female	428	1615	107	442

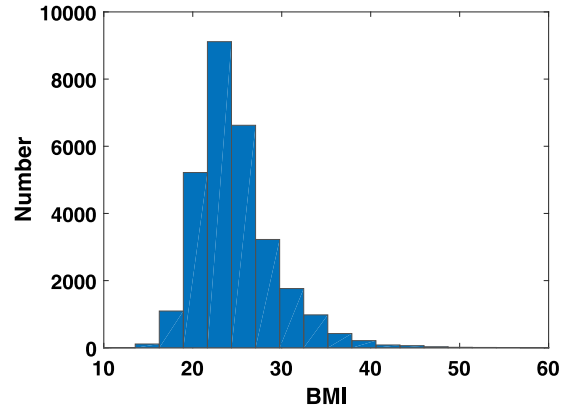


Fig. 5. Distribution of BMI values on Morph II. The BMI values mainly distribute between 15 to 35.

FIW-BMI dataset: It contains 7930 images from 4881 individuals, along with the corresponding gender, height and weight information. Among these individuals, there are 3192 males and 1689 females. Each individual has 1 to 4 images. Details about the dataset are described in Table 1. It is separated into two groups by gender. Most images in this dataset are collected from a social network—Reddit posts.¹ Because this is a social network displaying people's progress of weight loss, weight gain, or essentially any type of body transformation, the BMI values of these images distribute in a very wide range from 15 to 60 as shown in Fig. 4. Comparing with the distribution of BMI values on the Morph II database (as shown in Fig. 5), this dataset has a much wider range of BMI distribution. Thereby, we use it to fine-tune the deep face model.

Morph II dataset: It contains 55,608 passport-style frontal face images along with age, gender and ethnicity information. Moreover, there are 40,330 images have height and weight information. Considering the uneven distribution of the ethnicity in the database, only 29 033 images kept. The images are separated into four groups by gender and ethnicity. Details about the selected data are described in Table 2. Fig. 5 shows the BMI distribution of the dataset. The BMI values of Morph II mainly distribute in the range of 15 to 35. Among these, 893 are underweight, 16,582 are normal, 8237 are overweight and 3321 are obese. In this work, this dataset is used for evaluating the methods. It is split into training and test sets as shown in Table 3. The same individual does not exist in both the training and test sets. Most images from the same individual have different BMI values.

VIP_attribute dataset: It contains 1026 images from 1026 celebrities (mainly actors, singers and athletes) collected from the WWW. Among

¹ Website: <http://www.reddit.com/r/progresspics>.

Table 4

Characteristics of VIP_attribute dataset. Mean and standard deviations pertained to BMI for male and female groups.

	Male	Female
#Images	513	513
Mean	25.2	20.9
Std	3.6	3.7

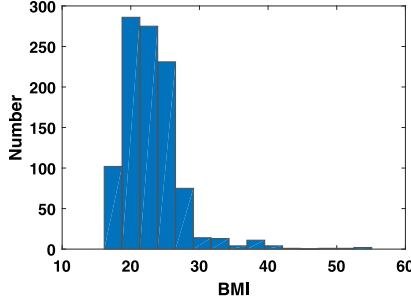


Fig. 6. Distribution of BMI values on VIP_attribute dataset. The BMI values mainly distribute between 18 to 30.

them, there are 513 males and 513 females. Table 4 describes the characteristic of the dataset. Fig. 6 shows the BMI distribution of the database. It is shown that BMI values mainly distribute between 18 to 30. Comparing with the BMI distributions of the previous two databases, this database has a much narrow and even distribution of BMI values. This dataset is used for evaluating the performance of the methods. It is split into training and test sets. There are 400 male images and 400 female images in the training set, and 113 male images and 113 females in the test set.

4. Experiments

We evaluate the proposed methods on two datasets. The performance metrics and experimental settings are introduced in this section. Then the experimental results and analysis are presented in detail.

4.1. Performance metrics

Two measure metrics are utilized for evaluating the performance of BMI estimators. The first one is mean absolute error (MAE) which is motivated by that used in age estimation, e.g., Guo et al. (2009). It is defined as the average of the absolute error between the estimated BMIs and the ground truth BMIs: $MAE = \frac{1}{N} \sum_{j=1}^N |\hat{b}_j - b_j|$, here b_j is the ground truth BMI for j th image, $\hat{b}_j$ is the estimated BMI, N is the number of test images. Another metric is the accuracy of predicted categories. According to the estimated BMI value, one can predict the image belong to which category (underweight, normal, overweight or obese). The accuracy of the predicted category is the proportion of the total number of predictions that are correct.

Both the two metrics have their advantages and limitations. For example, given an image with the true BMI value is 24, the estimated value is 19. Though the absolute error is 5, the predicted category (normal) is correct. However, if the true BMI of an image is 30 and the estimated value is 30.5, though the absolute error is 0.5, the predicted category (obese) is incorrect. Thereby, we combine them together to evaluate the performance of each method.

4.2. Experimental settings

(1) *Data preprocessing*: Given the images, we first applied face detection and landmark localization using the Openface toolkit (Amos et al., 2016). Then the images are aligned by the eye locations and cropped with the size of 96×112 . In addition, two geometric facial

BMI models, namely PIGF and PF, are utilized to further evaluate the effectiveness of the proposed method. For these two geometric models, the face images are cropped with the size of 256×256 .

(2) *Implementation details for BMI-related feature learning*: The fine-tuning of Centerloss network is implemented by the Caffe framework (Jia et al., 2014). We fine-tune the network parameters using face images with BMI labels from FIW-BMI dataset. In this fine-tuning step, we used mini-batch stochastic gradient descent (SGD) with momentum settings. The mini-batch size is set to 64 and momentum is set to 0.9. We initialize the learning rate to 0.00001. The learning rate decreases in polynomial decay with a power of 0.1. The training procedure stops after 20 000 iterations. A feature vector of 512 dimensions is extracted from layer fc5 of the Centerloss. We follow the official usage (Wen et al., 2016) which extracts the features for each image and its horizontally flipped one, and concatenates them as the representation with the size of 1024×1 .

(3) *Implementation details for evaluating different BMI estimators*: After extracting the facial features, five estimators are learned: Support Vector Regression (SVR) (Drucker et al., 1997), PCA-SVR, Gaussian Process Regression (GPR) (Williams and Rasmussen, 1996), Canonical Correlation Analysis (CCA) (Thompson, 2005), and Partial Least Square analysis (PLS) (Abdi, 2003). Considering the dimension of the deep facial features, principal component analysis (PCA) (Wold et al., 1987) is applied to the features before training the SVR, which denoted as PCA-SVR. The PCA percentage of explained variance for different SVR is various, but all selected based on the best performance. In our implementation, SVR is trained with RBF kernel, and GPR is trained with the rational quadratic kernel. The parameters for each SVR and GPR lead to the best result are utilized.

(4) *Implementation details for label distribution based estimator*: For the label distribution based method, the first step is to convert the BMI labels to distribution labels as described in Section 2.2. Particularly, the corresponding BMI range is from 13 to 60 with an interval of 0.1. The assigned labels are calculated according to the true BMI value. With the label distribution based method, we train the BMI estimators which are named LD-CCA and LD-PLS, respectively.

4.3. Experimental results

4.3.1. Evaluation of BMI estimation based on feature learning

First, we conduct the evaluation based on features extracted from the pre-trained Centerloss face model and fine-tuned BMI-related face model, respectively. In order to explore the performance based on different BMI estimators using the deep features, we conduct experiments using five different estimators: SVR, PCA-SVR, GPR, PLS and CCA. Facial features extracted from the deep network are fed into the estimators for training. Table 5 presents the experimental results. All the results are obtained based on separated training and testing on each of the gender and ethnicity groups. Face model means the feature directly extracted from the pre-trained Centerloss model. Fine-tuned means the feature extracted from the fine-tuned facial BMI model. As described before, FIW-BMI dataset is used to fine-tune the deep model. From the results, one can see that the performance (MAEs and the accuracy of predicted category) based on fine-tuned model are all better than the face model, which shows that fine-tuning the deep model using facial BMI data derives more robust representations for BMI estimation. No matter which estimator is used, the errors are all reduced. This demonstrates that the fine-tuned facial BMI model is more capable of capturing BMI-related facial features.

Since all the results presented in Table 5 are obtained from the estimators trained by original BMI labels (the single BMI label for each face image), they will be compared with the experimental results from label distribution based methods in the following section.

Table 5
BMI estimation results using five estimation methods on Morph II dataset.

Feature	Method	Black male		Black female		White male		White female	
		MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)
Face model	SVR	2.47	75.7	3.67	66.8	2.48	73.8	2.89	74.4
	PCA-SVR	2.50	75.7	3.73	66.8	2.50	73.8	2.86	73.5
	GPR	2.54	75.8	3.73	66.3	2.52	75.4	3.09	75.1
	PLS	2.52	76.5	4.76	54.1	2.82	70.4	4.64	51.8
	CCA	2.51	76.3	4.42	56.1	2.78	70.9	4.05	60.7
Fine-tuned	SVR	2.45	76.7	3.40	68.6	2.37	76.2	2.78	75.8
	PCA-SVR	2.57	75.7	3.56	67.8	2.34	78.1	3.47	69.0
	GPR	2.53	76.8	3.62	68.3	2.38	76.7	2.79	78.1
	PLS	2.47	76.5	4.54	56.2	2.71	73.0	4.27	58.4
	CCA	2.46	76.6	4.36	55.7	2.67	74.2	3.84	62.9

Table 6
BMI estimation results using label distribution based method on Morph II dataset.

Feature	Method	Black male		Black female		White male		White female	
		MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)
Face model	LD-PLS	2.49	77.0	3.49	66.0	2.48	74.7	2.96	71.5
	LD-CCA	2.42	76.5	3.50	67.1	2.38	77.1	2.86	75.9
Fine-tuned	LD-PLS	2.41	76.6	3.69	59.4	2.54	74.3	2.98	72.0
	LD-CCA	2.35	77.0	3.40	67.3	2.25	75.6	2.72	73.8

Table 7
BMI estimation results using the label distribution based method on Morph II dataset by geometric features.

Feature	Method	Black male		Black female		White male		White male	
		MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)
PIGF	SVR	2.66	72.7	3.73	65.8	2.71	70.8	2.96	70.7
	GPR	2.72	74.2	3.74	66.5	2.74	72.1	2.99	70.7
	PLS	2.76	71.9	3.81	67.1	2.74	71.2	3.15	74.5
	CCA	2.77	71.8	3.77	67.1	2.74	71.3	3.14	74.5
	LD-CCA	2.62	72.4	3.56	66.6	2.62	71.6	2.96	73.4
	LD-PLS	2.64	72.5	3.61	67.7	2.70	71.1	2.87	72.0
PF	SVR	2.63	73.6	3.65	68.3	2.68	70.8	3.12	72.5
	GPR	2.68	75.0	3.79	68.4	2.79	71.0	3.09	73.8
	PLS	2.73	73.1	3.69	66.5	2.73	70.9	3.37	72.0
	CCA	2.71	73.4	3.68	66.8	2.71	70.9	3.38	72.0
	LD-CCA	2.57	73.7	3.52	67.2	2.58	69.5	2.91	76.9
	LD-PLS	2.61	73.5	3.50	65.9	2.64	69.4	3.02	75.2

4.3.2. Evaluation of label distribution based methods

Now we conduct the experiment using our proposed estimator for BMI estimation. The experiment uses the features extracted from Centerloss face model and fine-tuned facial BMI model. The results of applying label distribution based methods to the facial feature are presented in Table 6. The estimated BMI values are obtained based on separated training and testing on each of the gender and ethnicity groups. Note that the method is based on the Gaussian distribution model as mentioned in Section 2.2. The comparison of the performance between the two label assignment strategies will be analyzed in Section 4.3.3.

Comparing with the results given in Table 5, LD-CCA and LD-PLS outperform the previous five estimators-SVR, PCA-SVR, GPR, PLS and CCA. More specifically, with the features extracted from the face model, MAEs of CCA and PLS are 2.51 and 2.52, respectively; while MAEs of LD-CCA and LD-PLS are 2.42 and 2.49, respectively. With the features extracted from the fine-tuned model, MAEs of CCA and PLS are 2.46 and 2.47, respectively; while MAEs of LD-CCA and LD-PLS are 2.35 and 2.42, respectively. This result shows the advantages of the proposed estimator when utilizing the label distribution schemes. Since the number of databases for visual BMI study is very limited, it is challenging to learn a good BMI estimator with limited training data. The label distribution scheme provides a feasible method to learn estimator with limited training data. More specifically, given a face image labeled with BMI value 20, its corresponding label distribution consists of a set of probabilities. Each probability represents the confidence that the corresponding label describes the image. The largest probability is

corresponding to BMI label 20. With this scheme, one image not only contributes to the learning of its BMI label but also provides auxiliary information to the learning of its adjacent BMI labels. In addition, the label distribution scheme well defines the increase and decrease of BMIs as a process.

To further evaluate the performance, Fig. 7 shows the overall MAE of the estimated BMI values in each step of the two-stage learning method. The overall MAE is calculated from the four test sets. From this figure, one can see that the MAE is reduced step by step using the proposed method on deeply learned representations. The MAE of applying CCA to Centerloss face feature is 2.75. After the feature learning stage, it drops from 2.75 to 2.71. Then by applying LD-CCA to deeply learned BMI features, the MAE significantly drops to 2.42. This further demonstrates the effectiveness of the proposed two-stage learning method.

4.3.3. Performance of different label assignment strategies

As mentioned in Section 2.2, there are two strategies for modeling BMI values with label distributions: Gaussian distribution and triangle distribution. We compare the performance of the two label assignment strategies. Three facial features are used in this experiment, they are Centerloss, PIGF and PF. Note that Centerloss features are extracted from the fine-tuned BMI face model. Two label distribution methods, LD-CCA and LD-PLS, are implemented for the comparison. The results are shown in Fig. 8. One can see that the Gaussian distribution performs better than the triangle distribution in the three cases. This result indicates that the Gaussian distribution is more suitable for defining

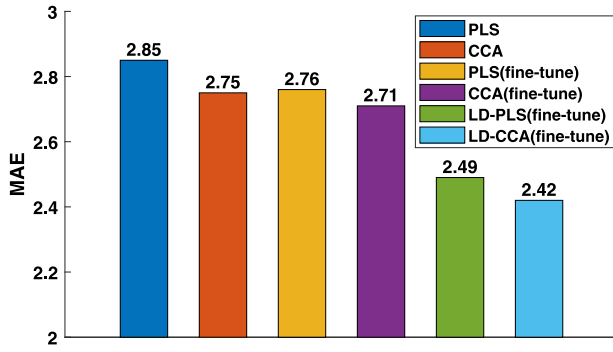


Fig. 7. Comparison of overall MAE on Morph II dataset in each step of our proposed method.

BMI labels than the triangle distribution. Because changes in facial appearance are non-rigid, the BMI-related changes in facial appearance are non-linear. The correlation between facial appearance and BMI is related to ages and genders (Pham et al., 2011). This means with different age or gender the BMI-related changes in facial appearance are different. However, the triangle distribution describes the BMI-related changes in facial appearance as a linear process. Thereby, the Gaussian distribution is more suitable for defining BMI labels than the triangle distribution. In addition, it can be observed that in most cases LD-CCA method shows better performance than LD-PLS by both two label assignment strategies.

To further analyze the two label assignment strategies, we evaluate their performances with different parameters. This experiment uses the features extracted from the fine-tuned BMI face model. Fig. 9 shows the experimental results. It can be seen that the best performance is achieved by setting σ to 4, Δ to 3 for LD-CCA, and setting σ to 5, Δ to 2 for LD-PLS.

4.3.4. Evaluate label distribution based estimators on geometric features

To further evaluate the effectiveness of the proposed label distribution based estimator, we apply it to two geometric features: psychology inspired geometric feature (PIGF) (Wen and Guo, 2013) and pointer feature (PF) (Jiang et al., 2019). PIGF consists of seven facial metrics: cheek-to-jaw-width ratio (CJWR), face width-to-height ratio (WHR) and face perimeter to area ratio (PAR), metrics-eye size (ES), lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH) and mean of eyebrow height (MEH). PF is the geometric facial representation that can well define the facial shape. It consists of coordinates of 68 facial landmarks which are extracted by the Openface toolkit (Amos et al., 2016). The coordinates of detected landmarks are simply concatenated as: $[x_1, y_1, \dots, x_n, y_n, \dots, x_{68}, y_{68}]^T$. The dimension of the feature is 136. The experimental results are given in Table 7. It can be seen that LD-CCA and LD-PLS perform better than the other

four methods in most cases. This demonstrates that the proposed BMI estimator performs well on geometric features.

4.3.5. Comparing with label distribution based methods

We compare our methods with two label distribution learning methods, namely LDL-IIS and LDL-CPNN (Geng et al., 2013) on Morph II dataset. The two methods are first proposed for facial age estimation. One assumption made in the IIS-LLD is the derivation of conditional probability $p(y|x)$ as the maximum entropy model (Berger et al., 1996). A strategy similar to improved iterative scaling (IIS) (Della Pietra et al., 1997) is used to optimize the cost function. Different from LDL-IIS, LDL-CPNN uses a three-layer network to approximate $p(y|x)$ to replace the above assumption. The results of the comparison are presented in Table 8. One deep feature and two geometric features are utilized for this experiment. The deep features are extracted from the fine-tuned BMI face model. From Table 8, one can see that among the two compared methods, LDL-IIS performs better than LDL-CPNN. Though LDL-IIS achieves the best performance on the black female set, our methods outperform LDL-IIS and LDL-CPNN in most other cases.

4.3.6. Comparing with deep learning based methods

We further compare the proposed method with two deep learning methods on VIP_attribute dataset. One is a regression based BMI estimation method proposed by Dantcheva et al. (2018). It is based on a 50-layers ResNet architecture which replaces the last 1000-dimension fc with the 1-dimension fc. And L1 loss is used to fine-tune the deep network. The weights of the pre-trained ResNet-50 (He et al., 2016) are used to initialize the network. Another is an ordinal regression based method for BMI category classification, namely native noisy binary search (NNBS) (Polania et al., 2019). Noisy binary search algorithms based pairwise comparisons are utilized to exploit the ordinal relationship among BMI categories. Table 9 shows the comparisons of the BMI estimation. For a fair comparison, the four methods (LDL-IIS, LDL-CPNN, LD-PLS and LD-CCA) utilize the deep features extracted by the Centerloss model (Wen et al., 2016) without the BMI feature learning stage. Because NNBS is a BMI category classification method, it outputs a predicted BMI category for each image rather than an estimated BMI value. Here we report its accuracy of predicted categories. For BMI value estimation (MAE), the proposed LD-CCA method outperforms the two label distribution based methods. And it performs slightly better than the ResNet based method. For BMI category prediction (accuracy), the NNBS method achieves the best accuracy (74.3%) on the male set, and ResNet based method achieves the best accuracy (85.8%) on the female set. All of the above experimental results demonstrate the effectiveness of the proposed method.

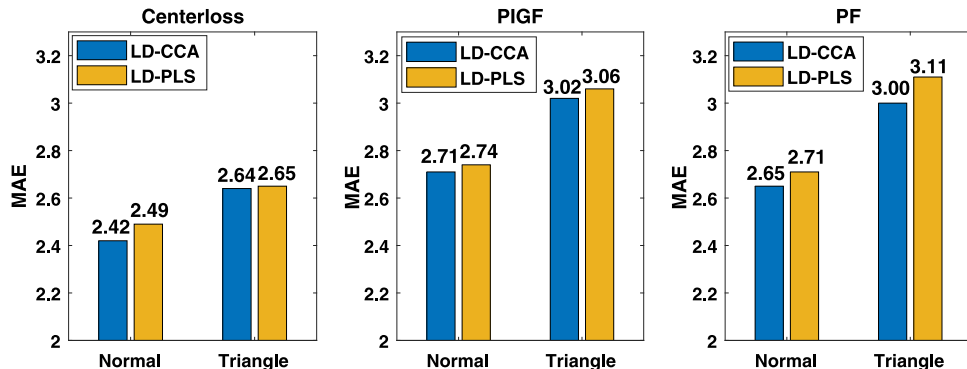


Fig. 8. Comparison of BMI estimation results on Morph II dataset using the two label assignment strategies.

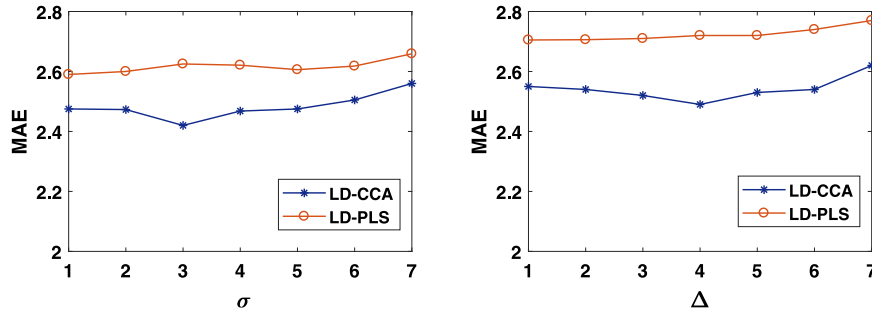


Fig. 9. The performance of the two label assignment strategies with different parameters.

Table 8

Comparison of BMI estimation between the proposed methods and two label distribution based methods on Morph II dataset.

Feature	Method	Black male		Black female		White male		White female	
		MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)
Fine-tuned	LDL-IIS	2.48	75.6	3.33	69.0	2.39	76.5	2.89	73.4
	LDL-CPNN	7.60	30.4	4.20	60.1	3.73	59.8	4.66	54.0
	LD-PLS (ours)	2.41	76.6	3.69	59.4	2.54	74.3	2.98	72.0
	LD-CCA (ours)	2.35	77.0	3.42	67.3	2.25	75.6	2.72	73.8
PIGF	LDL-IIS	2.99	85.6	4.10	74.4	3.08	77.2	3.34	66.8
	LDL-CPNN	2.98	80.0	4.70	50.9	3.19	78.4	3.26	66.4
	LD-PLS (ours)	2.64	72.5	3.61	67.7	2.70	71.1	2.87	72.0
	LD-CCA (ours)	2.62	72.4	3.56	66.6	2.62	71.6	2.69	73.4
PF	LDL-IIS	3.40	60.7	4.50	72.4	3.45	59.5	3.75	64.1
	LDL-CPNN	3.70	54.7	5.24	24.5	4.47	30.5	6.45	13.4
	LD-PLS (ours)	2.61	73.5	3.50	65.9	2.64	69.4	3.02	75.2
	LD-CCA (ours)	2.57	73.7	3.52	67.2	2.58	69.5	2.91	76.9

Table 9

Comparison of BMI estimation between the proposed methods and other four methods on VIP_attribute dataset.

Method	Male		Female		All	
	MAE	Accuracy (%)	MAE	Accuracy (%)	MAE	Accuracy (%)
LDL-IIS	4.75	55.8	4.04	69.0	4.39	62.4
LDL-CPNN	5.37	51.3	4.02	70.0	4.69	60.6
ResNet based	2.32	70.8	2.30	85.8	2.36	76.6
NNBS (VGG)	–	74.3	–	80.5	–	77.4
LD-PLS (ours)	2.25	72.6	2.28	81.4	2.26	77.0
LD-CCA (ours)	2.19	72.6	2.27	85.0	2.23	78.8

5. Conclusion

In this work, we study the problem of BMI estimation from face images by a two-stage learning framework. More specifically, first, a BMI-related face model is fine-tuned to learn more BMI-related facial features. Then we model the BMI labels with discrete probability distributions. Finally given the BMI-related facial features from the first step and the probability distributions, a BMI estimator is learned by maximizing the correlation between them. Two different label assignment strategies are presented in this work. Extensive experiments are conducted on three datasets: FIW-BMI, Morph II and VIP_attribute. The experimental results show that the two-stage framework reduces the estimated errors step by step. The proposed label distribution based estimator shows more robustness than regression based methods and methods without label distribution schemes. We also evaluated the effectiveness of the estimator on two geometric features. Furthermore, our method outperforms the two label distribution based methods and two deep learning based methods.

CRedit authorship contribution statement

Min Jiang: Conceptualization, Methodology, Writing - original draft, Writing review & editing. **Guodong Guo:** Funding acquisition, Supervision. **Guowang Mu:** Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work is partly supported by an NSF, USA grand IIS-1450620, an NSF-CITeR, USA grant, and an WV-HEPC, USA grant. The authors would also like to thank the editors and the reviewers for their comments and suggestions that improved the paper.

References

- Abdi, H., 2003. Partial least square regression (PLS regression). Encyclopedia Res. Methods Soc. Sci. 6 (4), 792–795.
- Amos, B., Ludwiczuk, B., Satyanarayanan, M., 2016. Openface: A General-Purpose Face Recognition Library with Mobile Applications, Vol. 6. 6, CMU School of Computer Science, p. 2.
- Arnold, M., Leitzmann, M., Freisling, H., Bray, F., Romieu, I., Renehan, A., Soerjomataram, I., 2016. Obesity and cancer: an update of the global impact. Cancer Epidemiol. 41, 8–15.
- Berger, A.L., Pietra, V.J.D., Pietra, S.A.D., 1996. A maximum entropy approach to natural language processing. Comput. Linguist. 22 (1), 39–71.
- Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 7291–7299.
- Coetzee, V., Perrett, D.I., Stephen, I.D., 2009. Facial adiposity: A cue to health? Perception 38 (11), 1700–1711.
- Dantcheva, A., Bremond, F., Bilinski, P., 2018. Show me your face and I will tell you your height, weight and body mass index. In: 24th International Conference on Pattern Recognition (ICPR). IEEE, pp. 3555–3560.
- Della Pietra, S., Della Pietra, V., Lafferty, J., 1997. Inducing features of random fields. IEEE Trans. Pattern Anal. Mach. Intell. 19 (4), 380–393.
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. In: Advances in Neural Information Processing Systems. pp. 155–161.
- Geng, X., Xia, Y., 2014. Head pose estimation based on multivariate label distribution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1837–1842.

- Geng, X., Yin, C., Zhou, Z.-H., 2013. Facial age estimation by learning from label distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (10), 2401–2412.
- Guo, A.Z., Jiang, M., 2020. Artificial Intelligence Techniques as Potential Tools for Large Scale Surveillance and Interventions for Obesity, Vol. 3. Crimson Publishers.
- Guo, G., Mu, G., Fu, Y., Huang, T.S., 2009. Human age estimation using bio-inspired features. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 112–119.
- Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J., 2016. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In: *European Conference on Computer Vision*. Springer, pp. 87–102.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 770–778.
- Hjelmås, E., Low, B.K., 2001. Face detection: A survey. *Comput. Vis. Image Underst.* 83 (3), 236–274.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM International Conference on Multimedia*. ACM, pp. 675–678.
- Jiang, M., Shang, Y., Guo, G., 2019. On visual BMI analysis from facial images. *Image Vis. Comput.* 89, 183–196.
- Kocabey, E., Camurcu, M., Ofli, F., Aytar, Y., Marin, J., Torralba, A., Weber, I., 2017. Face-to-BMI: using computer vision to infer body mass index on social media. In: *11th International AAAI Conference on Web and Social Media*.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. pp. 1097–1105.
- Meigs, J.B., Wilson, P.W., Fox, C.S., Vasan, R.S., Nathan, D.M., Sullivan, L.M., D'agostino, R.B., 2006. Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease. *J. Clin. Endocrinol. Metab.* 91 (8), 2906–2912.
- Moeslund, T.B., Granum, E., 2001. A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.* 81 (3), 231–268.
- Moeslund, T.B., Hilton, A., Krüger, V., 2006. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* 104 (2–3), 90–126.
- Parkhi, O.M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition. pp. 1–12.
- Pham, D.D., Do, J.-H., Ku, B., Lee, H.J., Kim, H., Kim, J.Y., 2011. Body mass index and facial cues in sasang typology for young and elderly persons. *Evidence-Based Complement. Altern. Med.* 2011.
- Polania, L., Fung, G., Wang, D., 2019. Ordinal regression using noisy pairwise comparisons for body mass index range estimation. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 782–790.
- Ricanek, K., Tesafaye, T., 2006. Morph: A longitudinal image database of normal adult age-progression. In: *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, pp. 341–345.
- Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 815–823.
- Thompson, B., 2005. Canonical correlation analysis. In: *Encyclopedia of Statistics in Behavioral Science*. Wiley Online Library.
- Wang, Q., Zheng, Y., Yang, G., Jin, W., Chen, X., Yin, Y., 2017. Multiscale rotation-invariant convolutional neural networks for lung texture classification. *IEEE J. Biomed. Health Inf.* 22 (1), 184–195.
- Wen, L., Guo, G., 2013. A computational approach to body mass index prediction from face images. *Image Vis. Comput.* 31 (5), 392–400.
- Wen, Y., Zhang, K., Li, Z., Qiao, Y., 2016. A discriminative feature learning approach for deep face recognition. In: *European Conference on Computer Vision*. Springer, pp. 499–515.
- Williams, C.K., Rasmussen, C.E., 1996. Gaussian processes for regression. In: *Advances in Neural Information Processing Systems*. pp. 514–520.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometr. Intell. Lab. Syst.* 2 (1–3), 37–52.
- Wolffhechel, K., Hahn, A.C., Jarmer, H., Fisher, C.I., Jones, B.C., DeBruine, L.M., 2015. Testing the utility of a data-driven approach for assessing BMI from face images. *PLoS One* 10 (10), e0140347.
- Wolk, R., Berger, P., Lennon, R.J., Brilakis, E.S., Somers, V.K., 2003. Body mass index. *Circulation* 108 (18), 2206–2211.
- Xing, C., Geng, X., Xue, H., 2016. Logistic boosting regression for label distribution learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 4489–4497.
- Yu, Y.-F., Wang, Q., Jiang, M., 2019. Discriminative common feature subspace learning for age-invariant face recognition. *IET Biometrics*.