Learning models for semantic classification of insufficient plantar pressure images

Article

Published Version

Learning Models for Semantic Classification of Insufficient Plantar Pressure Images

Yao Wu[1], Qun Wu[2, 3, 4]*, Nilanjan Dey[5], R. Simon Sherratt[6]

[1] Department of Industrial Design, Wenzhou Business College, Wenzhou (PR China)
[2] Institute of Universal Design, Zhejiang Sci-Tech University, Hanzhou (PR China)
[3] Collaborative Innovation Center of Culture, Creative Design and Manufacturing Industry of China Academy of Art, Hangzhou (PR China)
[4] Zhejiang Provincial Key Laboratory of Integration of Healthy Smart Kitchen System, Hangzhou (PR China)
[5] Department of Information Technology, Techno India College of Technology, West Bengal (India)
[6] Department of Biomedical Engineering, the University of Reading, Berkshire (United Kingdom)

uniR
LA UNIVERSIDAD
EN INTERNET

ABSTRACT

Establishing a reliable and stable model to predict a target by using insufficient labeled samples is feasible and effective, particularly, for a sensor-generated data-set. This paper has been inspired with insufficient data-set learning algorithms, such as metric-based, prototype networks and meta-learning, and therefore we propose an insufficient data-set transfer model learning method. Firstly, two basic models for transfer learning are introduced. A classification system and calculation criteria are then subsequently introduced. Secondly, a data-set of plantar pressure for comfort shoe design is acquired and preprocessed through foot scan system; and by using a pre-trained convolution neural network employing AlexNet and convolution neural network (CNN)-based transfer modeling, the classification accuracy of the plantar pressure images is over 93.5%. Finally, the proposed method has been compared to the current classifiers VGG, ResNet, AlexNet and pre-trained CNN. Also, our work is compared with known-scaling and shifting (SS) and unknown-plain slot (PS) partition methods on the public test databases: SUN, CUB, AWA1, AWA2, and aPY with indices of precision (tr, ts, H) and time (training and evaluation). The proposed method for the plantar pressure classification task shows high performance in most indices when comparing with other methods. The transfer learning-based method can be applied to other insufficient data-sets of sensor imaging fields.

KEYWORDS

I. INTRODUCTION

DATA in our social society, scientific research, and other fields, are generating at an unprecedented rate and being widely collected and stored. To further exploring the information or knowledge hiding in the datasets, intelligent technologies-based data processing has been developed and becomes the consensus of the current academic and industrial circles. However, the lack of labeled data has become increasingly prominent with the accumulation of many new data-sets. While not every field will spend as much manual labeling to produce some data like ImageNet [1], supervised learning can solve many important problems. Transfer Learning (TL) plays a very important role for processing the unlabeled datasets. In traditional classification learning models, there are two basic assumptions to guarantee the accuracy and reliability of the classification model; the first is that how the training samples for learning to the new test samples satisfying the same distribution independently; the second is that there must be enough available training samples to learn a good classification model [2]–[3].

In practice however, these two conditions are often unsatisfactory. Over time, the previously available labeled sample data may become unavailable resulting in semantic and distribution barriers with new testing samples. Reliable labeled sample data are often scarce and hard to be obtained. This resulted in a very important problem in Machine Learning (ML), how to use insufficient labeled samples or source domain data to predict target domains with different data distribution of reliable models. Recently, TL has attracted extensive attention and research [4]. Transfer learning uses existing knowledge to solve problems in different but related fields; it aims to solve the learning problem in the target region by transferring the existing knowledge, thus relaxing the two basic assumptions in traditional machine learning, while only a small amount of labeled sample data is available in the target region. The more factors shared by two different fields then the easier it will be. Otherwise, it will be more difficult, and even "negative transfer" will occur, which has side effects in practice [5].

Scholars have carried out extensive research on TL recently; and many of them have focused on different technologies to study TL algorithms. TL arises from the mining technology of incomplete or insufficient data-sets; and as one of the Deep Learning (DL) technologies, it has been used in various fields to solve problems. The target accuracy of more than 90% was easily achieved by

previous studies [6]– [7]. However, DL is a data-hungry technology that requires many annotated samples to work with. In practice, there are many problems due to the lack of labeled data, and the cost of obtaining labeled data is also very large, such as in the field of medical treatment and information security. Therefore, this inevitably leads to the well-known small sample problem, with new classes in the training process that have never been seen before. It can only use a few labeled samples of each class without changing the trained model [8]. Taking image classification data as an example, the traditional method is to obtain a model based on the training set, and then annotate the test set automatically. The small sample problem is that many classified data-sets can be processed using a few labeled data-sets. When the amount of tagged data-set is relatively small, these rare categories need to be generalized without additional training. Few-shot, one-shot [9]– [11], and Zero-Shot Learning (ZSL) models are widely used to address this issue. TL is highly related to the few-shot and one-shot learning models [12]. Abderrahmane et al. [13] introduced a zero-shot haptic recognition algorithm for robots in interacting with their environment. Ji et al. [14]-[15] introduced a Manifold-regularized Cross Modal Embedding (MCME) approach for ZSL. Combining ontology and reinforcement learning for Zero-Shot Classification (ZSC) [16]–[17], Kernelized Linear Discriminant Analysis (KLDA), Central-Loss based Network (CLN) and Kernelized Ridge Regression (KRR) for ZSL [18] has also been developed subsequently. Yu et al. [19] introduced a regularized cross-modality ranking (ReCMR) to capture the semantic information from heterogeneous sources by using intra/intermodals.

In ZSL, intuitive decision-making is more reliable based on costs and benefits [20]. Some remarkable classification results deployed plenty of new class by using fast prototype-based learning algorithms [21]-[24]. From a TL aspect, the distribution of labeled source domain data is generally different from that of unlabeled target domain data, so the labeled sample data is not necessarily useful [25]. Therefore, the selection of the training samples which are most beneficial to target domain classification is of utmost importance. The TL algorithms can be divided into those applied to single-source domains and those applied to multiple source domains. TL has been widely applied in text classification and clustering, emotional classification, image classification and collaborative filtering. Soudani and Barhoumi [26] extracted features from the convolutional part to improve the segmentation performance using two pre-trained architectures (VGG16 and ResNet50).

Plantar pressure test and gait analysis is an internationally advanced technology based on the principle of biomechanics to detect the structure of the lower limbs of the human body, to assess and predict future foot diseases, and to provide scientific rehabilitation methods. The reaction force of the ground can be divided into static and dynamic foot pressure, which respectively represent the ground reaction force that the human body receives when standing statically and dynamically walking and running and jumping. Most of the foot diseases are related to normal pressure changes in the soles of the feet, and the two usually affect each other. Therefore, understanding the distribution of normal foot pressure is not only an effective tool for the diagnosis and evaluation of foot diseases but also is an important guider in restoring the normal or near-normal distribution of the foot pressure and restoring the function of the foot.

The pressure distribution of the human foot can directly reflect the pressure value of the various parts of the foot when standing or running and can also indirectly reflect the structure and function of the foot and the control of the knee, hip, spine and even the entire body posture. Testing and analyzing the plantar pressure can obtain the pressure information of the normal and abnormal soles of the human body in various postures. Through the dynamic collection of the pressure data of the sole, the biomechanical properties of the plantar pressure can

be determined for the early prediction, diagnosis and treatment of deformed foot, various types of foot diseases and diabetic foot ulcers; quantitative assessment of the degree of joint disease and postoperative efficacy in orthopedics and orthopedic surgery; Walking training and designing intelligent prosthetics and other fields. Therefore, a comprehensive understanding of the changes in plantar pressure facilitates the understanding of the biomechanics and function of normal feet, clinical medical diagnosis, disease degree determination, postoperative efficacy evaluation, rehabilitation research and the design of various types of orthopedic shoes and sports shoes, which are of great significance.

Deschamps et al. [27] proposed a high-resolution pixel-level analysis and 12 of the standardized foot-pressure barometers in the original experimental data-set (including 97 diabetic patients and 33 non-diabetic patients). The new cohort of diabetic patients determines the classification recognition rate and its sensitivity and specificity to assess the classification effect. Ramirez-Bautista et al. [28] used plantar pressure data to detect disease progression and the use of different electronic measurement systems for corresponding analysis suggests a hybrid algorithm for classifying plantar images. Sommerset et al. [29] combined other physiological signals to study the effects of plantar pressure on compressible atherosclerosis. Xia et al. [30] analyzed the background plantar pressure image (PPI) of high temporal and spatial resolution and experimental results perform higher effectiveness in terms of mean square error (MSE), exclusive OR (XOR) and mutual information (MI). Plantar pressure imaging is different from the general image. Pressure images are formed by discrete values of pressure. Colors are used to reflect different pressure and pressure values. These pressure values will be applied to the image intelligent classification method in this paper. Therefore, we screened the current popular deep learning method to process the image, thus transforming the data set of discrete pressure values into image processing problems. The research motivation of our work is to find an effective transfer learning method applied for such insufficient datasets, including lack of labelled datasets. The finding of the research is that we used the images forms of the sensor generated dataset for training and classification by using an improved convolutional neural network, and the significance of the network models performs high efficiency by comparing with VGG, ResNet, AlexNet and pre-trained CNN, etc., in the classification indices.

The structure of the rest of the paper is: Section 2 introduces the basic models for plantar pressure image classification, and improved models are described in Section 3. Section 4 deploys the results and discussion, while Section 5 includes concluding remarks and future works.

II. Methods

A. Finetune-Based Modeling

DL requires a large amount of data and computing resources and takes a lot of time to train the model, but it is difficult to meet these requirements in practice. The use of TL can effectively reduce the amount of data, calculations, and calculation time, and can be customized in the new business needs of the scenario. Transfer learning is not an algorithm but a machine learning idea, and its application to deep learning is fine-tune. By modifying the structure of the pre-trained network model (such as modifying the number of sample category outputs), selectively load the weight of the pre-trained network model (usually loading all the previous layers except the last fully connected layer, also called the bottleneck layer), and then use it retraining the model on the dataset is the basic step of fine-tuning. Fine-tuning can quickly train a model with relatively small amount of data achieving good results.
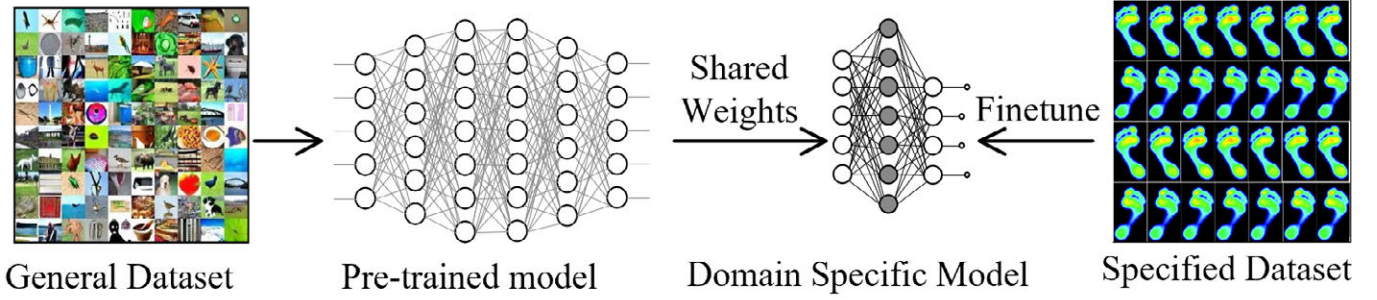
Fig. 1. Pre-trained convolutional neural network for shared weights finetune specified data-set using general image database- ImageNet.

Finetune-based models have been widely used in the field of DL. By obtaining a certain amount of tagging data, a basic network can be established for fine-tuning. This basic network is acquired through large-scale data sets with rich tags, such as ImageNet, or e-commerce data, known as universal data domains. Then, training is performed on specific data domains. While the parameters of the basic network will be fixed in the process, and the domain-specific network parameters will be trained, including to set the fixed layer and learning rate, etc. This model method can be relatively fast and does not significantly depend on the amount of data. Fig. 1 shows the framework of a finetune-based model for the plantar pressure image specified data-set.

B. Siamese Neural Networks

Siamese Neural Networks (SNN) measure how similar the two inputs are. The twin neural network has two inputs (Input1 and Input2), and the two input feeds into two neural networks (network1 and network2). These two neural networks map the inputs to the new space respectively, forming the input in the new space. Trough the calculation of Loss, the similarity of the two inputs is evaluated. This method restricts the input structure and automatically discovers features that can be generalized from new samples. Supervised metric learning based on twin networks is trained, and then one/few-shot learning is performed by reusing the features extracted from that network [31] – [34]. It is a two-way neural network. It combines samples of different classes into pairs for the network training. At the top level, it calculates loss through a distance cross-entropy. In forecasting, taking 5 way-5 shot as an example, five samples are randomly selected from five classes, and the data with 25 mini-batch is input into the network. Finally, 25 values are obtained. The category with the highest score is taken as the forecasting result, as shown in Fig. 2.

The network structure is shown in Fig. 3, it is an eight-layer deep convolution twin network. The graph shows only one of the calculations. After the 4096-dimensional full-connection layers of the network, the component-wise L1 distance calculation is performed to generate a 4096-dimensional eigenvector, a probability of 0 to 1 is obtained by sigmoidal activation as the result of whether two input samples are similar.
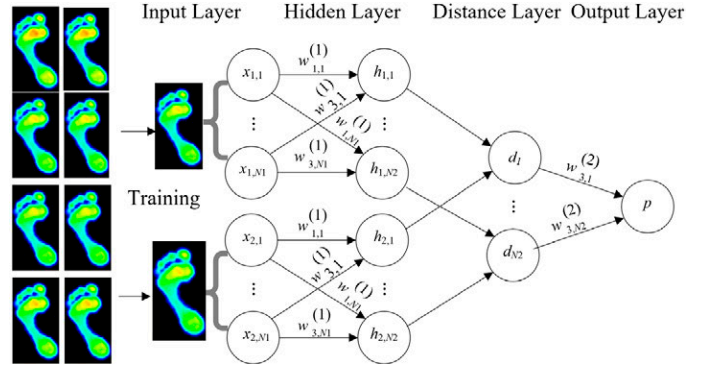

Fig. 2. Typical Siamese Neural Networks for plantar pressure image data-set training using distance layer and shared weights.

C. Matching Networks

Without changing the network model, matching networks can generate labels for unknown categories. Its main innovation lies in the process of modeling and training. For the innovation of the modeling process, a matching network based on memory and attention is proposed, which makes it possible to learn quickly. For the innovation of the training process, this work is based on a principle of traditional ML, that is training and testing are to be carried out under the same conditions [35]-[36]. It is proposed that the network should constantly look at the insufficient samples of each type during training, which will be consistent with the testing process. Specifically, it attempts to obtain a mapping from support set S (composed of K samples and tags) to classifier $\hat{Y}$, which is a network of $P(\hat{Y}|\hat{X},S)$. It gives the label $\hat{Y}$ for each unknown test sample $\hat{X}$ based on the current S, and the label maximizes P. This model is described as follows,

$$\arg \max_y \ P(Y|\hat{X},S) \tag{1}$$

$$\hat{Y} = \sum_{i=1}^{k} a(\hat{X},x_i)y_i \tag{2}$$

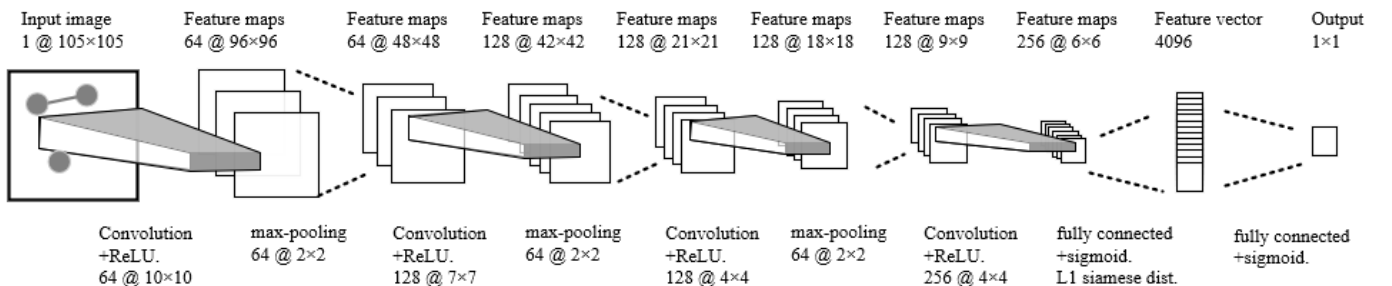where, a is an attention (i.e. the distribution of classes on S) being a

Fig. 3. Eight deep convolutional Siamese neural network using feature maps and max-pooling.

linear combination of class attentions on S; for x_i which is farthest from $\hat{X}$, its attentions under a certain measure is 0, and its value is the weight fusion of the corresponding labels corresponding to x_i which is similar to $\hat{X}$. The attentions mentioned above are embedding the training sample x_i and the test sample $\hat{X}$ separately; and then are inputed by software-max transforms, such as:

$$a\left(\hat{X}, x_i\right) = \frac{e^{c(f(\hat{x}),g(x_i))}}{\sum\limits_{j=1}^{k} e^{c\left(f(\hat{x}),g(x_j)\right)}} \tag{3}$$

where, c is the cosine distance. Two of the embedding models are shared, such as CNN. a is related to metric learning. For the sample x to be classified, it needs to be aligned with those labeled y, and other misalignments. Furthermore, the support set sample embedding model g can continue to be optimized, and the support set sample should be used to modify the embedding model f of the test sample. This can be solved in the following two aspects: i) embedding based on two-way LSTM learning training set, so that embedding of each training sample is the function of other training samples; ii) embedding of test samples based on attention-LSTM, so that embedding of each test sample is the function of embedding of training set. This work calls Full-Conditional Embedding (FCE). Although the above classification is done on the whole sample of support set, the embedding used for cosine distance calculation is independent of each other. Therefore, this work changes the embedding of support set samples to $g(x_i, S)$, which is useful when x_j is very close to x_i. This model uses Bidirectional LSTM (BiLSTM), forward LSTM and the backward LSTM are combined into a BiLSTM. By doing so, we can effectively use past features (through forward state) and future features (through backward state) within a specified time frame. We use back-propagation through time (BPTT) to train a bidirectional LSTM network. Over time, the forward and backward transfers on the expanded network are similar to the forward and backward transfers in conventional networks, except that we need to expand the hidden state for all time steps. We also need special processing at the beginning and end of the data points. Supposed that S is a random sequence, and then codes each x_i:

$$g(x_i, S) = \vec{h}_i + \overleftarrow{h}_i + g'(x_i) \tag{4}$$

$$(h_i, c_i) = LSTM\left(g'(x_i), \vec{h}_{i-1}, \vec{c}_{i-1}\right) \tag{5}$$

$$\left(\overleftarrow{h}_i, \overleftarrow{c}_i\right) = LSTM\left(g'(x_i), \overleftarrow{h}_{i-1}, \overleftarrow{c}_{i-1}\right) \tag{6}$$

where, $g'(x_i)$ is originally only dependent on its own embedding, x_i is for information exchange by BiLSTM. On the optimization of f, support set samples can be used to modify the embedding model of the test samples. This can be solved by a fixed-step LSTM and attention model of support set as:

$$f(\hat{x}, S) = atLSTM\left(f'(x), g(S), K\right) \tag{7}$$

$$\hat{h}_k, c_k = LSTM\left(f'(\hat{h}), [h_{k-1}, r_{k-1}], c_{k-1}\right) \tag{8}$$

$$h_k = \hat{h}_k + f'(\hat{x}) \tag{9}$$

$$r_{k-1} = \sum_{i=1}^{|S|} a\left(h_{k-1}, g\left(x_i\right)\right) g\left(x_i\right) \tag{10}$$

$$a\left(h_{k-1}, g(x_i)\right) = soft\max(h_{k-1}^T \, g(x_i)) \tag{11}$$

where, $f'(x)$ is only dependent on the characteristics of the test

sample itself; as the input of LSTM (unchanged at each step), K is the step of LSTM, $g(S)$ is the embedding of support set. As a result, the model ignores some samples in support set S. The embedding functions f and g optimize the feature space to improve the accuracy.

D. Insufficient Data-set Transfer Learning Modeling

For the convenience of introducing our zero-sample problem, here we first briefly introduce the recognition problem which is usually referred to in computer vision. We can summarize a simple pattern. First, we choose a suitable learning machine: convolutional neural network (CNN), prepare enough training data and corresponding labels for the learning machine, and use training data to train the learning machine. The main feature of the Convolutional neural network (CNN) is the use of convolutional layers. This is actually a simulation of the human visual neuron. A single neuron can only respond to certain specific image features, such as horizontal or vertical edges. It is very simple in itself, but these simple neurons form a layer. After the number of layers is enough, the system can obtain sufficient features. CNN is the most popular neural network model for image classification problems. An important idea behind CNN is the local understanding of the image. Its parameters will greatly reduce the time required for learning and reduce the amount of data required to train the model. The CNN has enough weights to see small patches of the image, rather than a fully connected network of weights from each pixel.

In practical use, the test data is input into the trained learner, and the learner can output the corresponding label of the test data. It is worth mentioning here that in this general pattern, the learner can only predict the existing categories in the training set when testing.

1. Basic Modeling

The concept of zero-shot learning and the initial solution, called attribute migration between classes, are given. We need to create a table to record the values of attributes of each category. We can map these pictures to the classification attribute table by looking up the table. These attribute descriptions can be summarized in advance because it is difficult to collect many pictures of unknown categories, but it is feasible to summarize the corresponding attributes of each category only. The core idea of attribute transfer between classes is that although objects have different categories, they have the same attributes, extract the attributes corresponding to each category and use several learners to learn. In the process of testing, the attributes of test data are predicted, and then the predicted attributes are combined with corresponding categories to realize the category prediction of test data. A simple model of zero sample learning can be summarized as shown in Fig. 4. The image space and label space are the initial image space and label space respectively. In zero-sample learning, images are usually mapped into feature space by some methods, which is called feature embedding. The same label is mapped into a label embedding, learning the linear or non-linear relationship between feature embedding and label embedding. Predictive transformation during testing replaces previous direct learning from images space to label space.
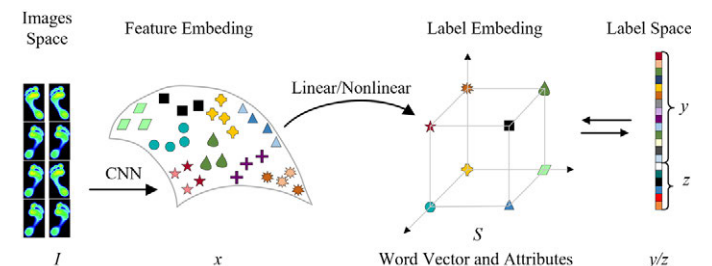

Fig. 4. Zero-shot learning using word vector and attributes.

2. Discriminative Learning of Latent Feature

Compared with the inter-class attribute migration method, it mainly adds two ideas: designs a special network to find the main part of the target in the picture; and designs the attributes using manual annotation, and adds label embed to search potential attributes from the original picture. The model diagram is shown in Fig. 5.
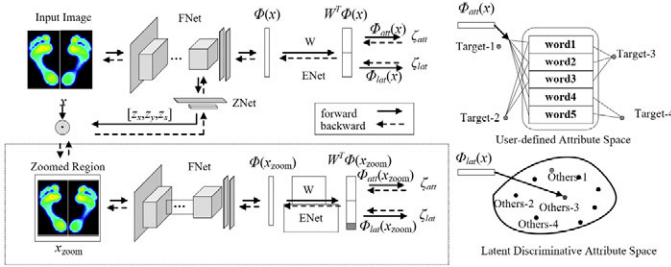

Fig. 5. Discriminative learning of latent feature-based modeling.

3. Improved Framework for the Proposed Model

Universal Zero Sample Learning (ZSL) requires accurate prediction of categories that have been seen (WS) and have not been seen (WU) (ZSL only requires prediction of categories that have not been seen). The main idea of this method is to train a WS first by using training data and training categories [37]–[39]. Then WS trains a graph convolution network using word vectors as input, and uses the graph convolution network to input the word vectors of test categories to get the WU. WS and WU together form a learner W to predict all categories, as shown in Fig. 6.
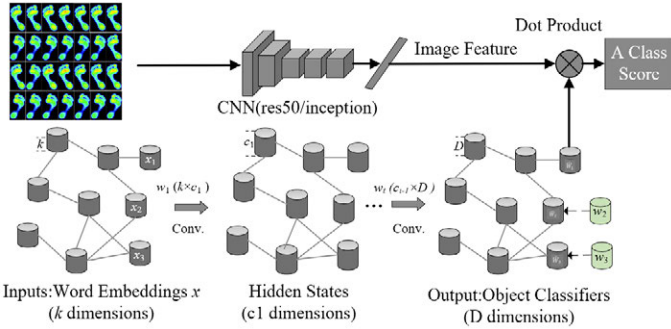

Fig. 6. Semantic (word embeddings) classifiers for plantar pressure image

III. SEMANTIC CLASSIFICATION OF PLANTAR PRESSURE IMAGE RESULTS

A. Semantic Classification of Plantar Pressure Image Results

Training a good CNN model for image classification requires not only computational resources but also a long time, especially when the model is complex and the amount of data is large. Standard desktop computers typically need to be trained for a few days if they are immobile. In order to train our plantar pressure image classifier quickly, we can use the model parameters that others have trained, and train our model on this basis. This belongs to transfer learning. Images from the basic data-set obtained are shown in Fig. 7.
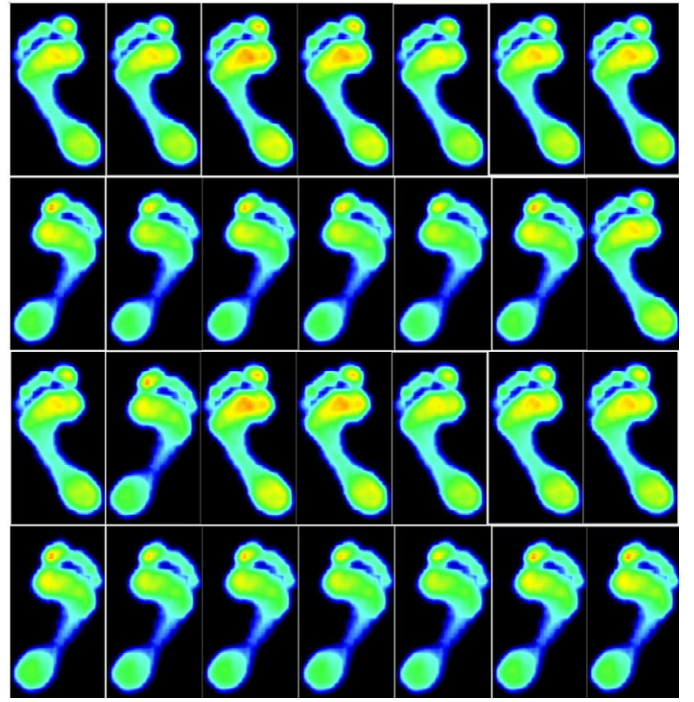

Fig. 7. Basic data-set acquired from the pressure sensors.

B. Pre-trained CNN using AlexNet for Classifiers

The data are divided into a training data set and a validation data set. 70% of the images were used for training, 15% for testing and 15% for verification. In the AlexNet system, SplitEachLabel divides image data stored into two new data storage areas, i.e. [imdsTrain, imdsValidation] = splitEachLabel (imds, 0.7,'randomized'); this small data set contains 55 training images and 20 validation images. AlexNet was trained on more than a million images, which can be grouped into 1,000 object categories. Therefore, the model has learned a rich feature representation based on many images. Let Net=alexnet and by using "analyzeNetwork", the network architecture and detailed information about the network layer can be presented visually and interactively. The steps are as follows:

Step 1: The first layer (image input layer) requires an input image of 277×277×3 in size, where 3 is the number of color channels.

Step 2: Replace the final layer. The last three layers of the pre-training network net are configured for 1000 classes. These three layers must be fine-tuned for the new classification problem. All layers except the last three layers are extracted from the pre-training network. By replacing the last three layers with a full connection layer, soft Max layer and classification output layer, the layer is migrated to the new classification task. -The options for a new full connection layer are specified based on the new data. To make learning faster in the new layer than in the migrating layer, the full connection layer Weight Learn Rate Factor (WLRF) and Bias Learn Rate Factor (BLRF) values are increased.

Step 3 Training the network. The network requires the size of the input image to be 277×277×3, but the images stored in image data have different sizes. The size of training image can be automatically adjusted by using enhanced image data storage. The training image in both horizontal and vertical directions is up to 30 pixels. Data enhancement helps to prevent network over-fitting and memorizes the details of training images. Specify training options. For migration learning, the shallow features of the pre-training network (migration layer weight) need to be retained. This combination of learning rate settings only accelerates the learning speed in the new layer, but slows

down the learning speed in other layers. When implementing transfer learning, the number of training rounds required is relatively small. A round of training is a complete training cycle for the whole training data set. Small batch sizes and validation data are specified. The software verifies the network every Validation Frequency iteration in the training process.

Step 4: Training the network consisting of migration layer and new layer.

Step 5: Calculate the classification accuracy for the verification sets.

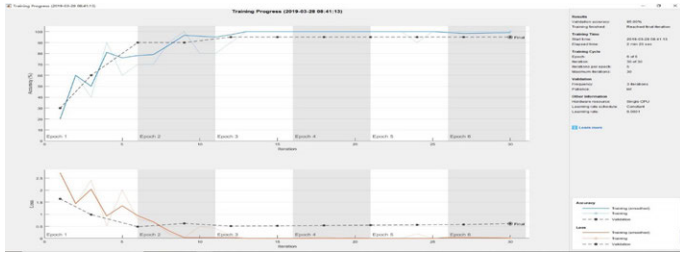Fig. 8 and Fig. 9 present the accuracy and loss of 25 layers of the CNN in AlexNet.

Fig. 8. Accuracy as a function of iteration in 25 layers of the CNN in AlexNete pressure sensors.
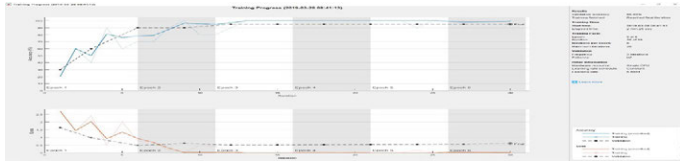

Fig. 9. Loss as a function of iteration in 25 layers of the CNN in AlexNet.

C. CNN based Transfer Modeling

The CNN model has two parts, the convolution layer in front and the full connection layer in the back. The function of the convolution layer is to extract image features while the function of the full connection layer is to classify features. Our method is to modify the full connection layer and retain the convolution layer on the inception-v3 network model. The parameters of convolution layer are trained by others. The parameters of the full connection layer require to initialize and use our own data to train and learn, as shown in Fig. 10.

- Convolution
- AvgPool
- MaxPool
- Cancat
- Dropout
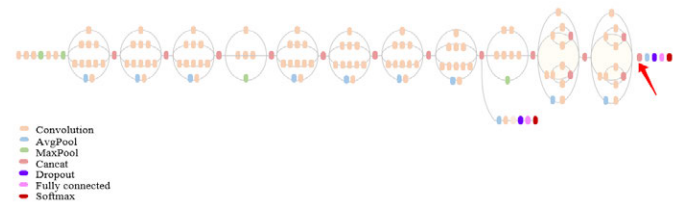- Fully connected
- Softmax

Fig. 10. Loss as a function of iteration in 25 layers of the CNN in AlexNet.

The front part of the red arrow in the inception-v3 model in Fig. 10 is the convolution layer, and the back part is the full connection layer. We need to modify the full connection layer and change the final output of the model to 5. Because the TensorFlow framework is used here, we need to get the tensor BOTTLENECK_TENSOR_NAME (output value of the last convolution activation function, number 2048) and the tensor JPEG_DATA_TENSOR_NAME of the initial input data of the model. The purpose of acquiring these two tensors is to obtain the image trained data through JPEG_DATA_TENSOR_NAME tensor input model and BOTTLENECK_TENSOR_NAME tensor after the convolution layer. By inception-v3 model including pre-

trained parameters, the 2 tensors above mentioned are acquired. For the classification issue, the full connected layers need to be adjusted and here only one layer is added; the input data is BOTTLENECK_TENSOR_NAME; finally, the cross-entropy loss function is defined. Because the training data set is relatively small, all pictures are input into the model through JPEG_DATA_TENSOR_NAME tensor, and then the values of BOTTLENECK_TENSOR_NAME tensor are obtained and saved to the hard disk. During model training, the value of the saved BOTTLENECK_TENSOR_NAME tensor is read from the hard disk as the input data of the full connection layer, the value of tensor of BOTTLENECK_TENSOR_NAME is acquired by the inputting images. The final accuracy of the test was found to be 93.5% over 4180 steps.

D. Discussion

As a result, the pre-trained CNN to assess the required level achieves performance improvements in plantar pressure image data-set, and the classification was also incrementally fine-tuned. Detection and characterization of plantar pressure image data-set by using deep learning with indices of Training-Testing Data Splitting (TTDS), Class Type (CT), Area Under the Curve (AUC), Average Precision Score (APS), recall, precision, f1 score (an index used to measure the accuracy of binary classification model in statistics -it considers both the accuracy and recall of the classification model), N, UN, Ave are compared and summarized in Table I [37],[40]-[42]. Table II shows the details of the public test database of SUN, CUB, AWA1, WAW2, and aPY [37]. Scaling and shifting (SS) is the known partitioning method while plain slot (PS) is the unknown partitioning method, tr is precision on known classification, ts is precision on unknown classification. The training and evaluation time for SS and PS are shown in Table III. The performance of the proposed method for public data-set of SUN CUB AWA1, AWA2, aPY [25] are shown in Table IV. By using inception-v3, the proposed model was also compared by using a public test database which H is a comprehensive index of "ts" and "tr", as shown in Table V. Therefore, the proposed CNNTM has better performance on primary databases on known classification.

IV. Conclusion

Large-scale methods are currently used to process most of the labeled data, but existing problems include lack of knowledge and many training samples. In practice, it is difficult to accumulate data for most categories and large-scale methods are not fully applicable. Therefore, learning many data-sets for a certain category is not feasible. While in each situation, a new category can be learned quickly with only a small number of samples. There are two main solutions under consideration at present: the first is that an object has not been seen before which is named as zero sample learning; and the second is to learn knowledge from existing tasks and apply it to future model training, which can be considered as a problem of Transfer Learning (TL). The method of the transfer-based learning refers to the use of these auxiliary data sets for transfer learning when there are other data sources. The idea is to learn a generalized representation from the data which can be directly used in the target data and small sample category learning process. Sample-based transfer learning is an attempt to give new weight to each sample in the source data so that it can better serve the new learning task. Samples that are more like the target data are selected from the source data to participate in the training, while samples that are not like the target data are excluded. The insufficient data-set learning methods are introduced, and an improved framework of transfer learning model is proposed and compared by other typical methods. The plantar pressure image data-set semantical classification issue performs high effectiveness with indices that were defined by current researchers.

TABLE I. Classifiers of VGG, ResNet, AlexNet and Pre-trained CNN for Plantar Pressure Image Data-set

Classifier	Training-Testing Data Splitting	Class Type	Precision	Recall	F1 Score	Accuracy	AUC	APS
		N	0.910	0.912	0.920			
	90-10%	UN	0.910	0.914	0.920	92.21	93.12	90.50
		Avg/Total	0.910	0.908	0.920			
		N	0.915	0.914	0.919			
VGG16+LR	80-20%	UN	0.915	0.918	0.918	91.12	92.10	90.72
		Avg/Total	0.915	0.917	0.918			
		N	0.915	0.912	0.907			
	70-30%	M	0.912	0.909	0.907	91.34	91.21	93.12
		Avg/Total	0.905	0.908	0.911			
		N	0.910	0.909	0.911			
	90-10%	UN	0.905	0.909	0.911	90.52	92.50	92.64
		Avg/Total	0.910	0.909	0.912			
		N	0.875	0.912	0.901			
VGG19+LR	80-20%	UN	0.872	0.909	0.902	91.28	91.34	91.12
		Avg/Total	0.875	0.904	0.912			
		N	0.870	0.880	0.899			
	70-30%	UN	0.912	0.889	0.892	92.10	86.12	90.45
		Avg/Total	0.910	0.878	0.878			
		N	0.835	0.812	0.832			
	90-10%	UN	0.890	0.831	0.823	80.32	86.22	91.25
		Avg/Total	0.850	0.835	0.825			
		N	0.835	0.848	0.845			
ResNet50+LR	80-20%	UN	0.869	0.879	0.889	80.15	83.24	86.25
		Avg/Total	0.889	0.877	0.887			
		N	0.825	0.808	0.832			
	70-30%	UN	0.824	0.811	0.821	80.25	81.25	86.75
		Avg/Total	0.829	0.812	0.821			
		N	0.908	0.870	0.899			
	90-10%	UN	0.908	0.914	0.883	92.32	93.07	92.65
		Avg/Total	0.909	0.921	0.883			
		N	0.912	0.919	0.876			
Pre CNN25+AlexNet	80-20%	UN	0.908	0.909	0.912	92.66	90.07	92.12
		Avg/Total	0.909	0.908	0.912			
		N	0.914	0.911	0.921			
	70-30%	UN	0.889	0.900	0.919	92.65	93.01	93.11
		Avg/Total	0.917	0.911	0.919			
		N	0.925	0.938	0.943			
	90-10%	UN	0.929	0.940	0.942	91.75	92.66	94.21
		Avg/Total	0.930	0.935	0.940			
		N	0.927	0.937	0.922			
inception-v3-CNN+TM (+)	80-20%	UN	0.921	0.931	0.921	92.75	91.25	94.36
		Avg/Total	0.919	0.923	0.921			
		N	0.915	0.912	0.922			
	70-30%	UN	0.912	0.913	0.910	93.50	94.22	95.55
		Avg/Total	0.907	0.907	0.910			

+ proposed method, Normal(N) and Unnormal (UN)

TABLE II. Number of Classes (Y(ts): test classes, Y(tr): training classes)

Data-set	Size	Detail	Attributes	Y	Y(tr)	Y(ts)
SUN	M	F	102	717	580+65	72
CUB	M	F	312	200	100+50	50
AWA1	M	C	85	50	27+13	10
AWA2	M	C	85	50	27+13	10
aPY	S	C	64	32	15+5	12

TABLE III. Number of Images, Training Time and Evaluation Time of TR and TS

| | | At Training Time | | | | At Evaluation Time | | | |
| | | SS | | PS | | SS | | PS | |
Data-set	Total	Y(tr)	Y(ts)	Y(tr)	Y(ts)	Y(tr)	Y(ts)	Y(tr)	Y(ts)
SUN [43]	14340	12900	0	10320	0	0	1440	2580	1440
CUB [44]	11788	8855	0	7057	0	0	2933	1764	2967
AWA1 [45]	30475	24295	0	19832	0	0	6180	4958	5685
AWA2 [45]	37322	30337	0	23527	0	0	6985	5882	7913
aPY [46]	15339	12695	0	5932	0	0	2644	1483	7924

TABLE IV. Classification Precision on Public Data-Set Comparing with Current Methods

| | SUN | | CUB | | AWA1 | | AWA2 | | aPY | |
Methods	SS	PS	SS	PS	SS	PS	SS	PS	SS	PS
DAP [45]	38.9	39.9	37.5	40.0	57.1	44.1	58.7	46.1	35.2	33.8
IAP [45]	17.4	19.4	27.1	24.0	48.1	35.9	46.9	35.9	22.4	36.6
CONSE [47]	44.2	38.8	36.7	34.3	63.6	45.6	67.9	44.5	25.9	36.6
CMT [48]	41.9	39.9	37.3	34.6	58.9	39.5	66.3	37.9	26.9	28.0
SSE [49]	54.5	51.5	43.7	43.9	68.8	60.1	67.5	61.0	31.1	34.0
LATEM [50]	56.9	55.3	49.4	49.3	74.8	55.1	68.7	55.8	34.5	35.2
ALE [50]	59.1	58.1	53.2	54.9	78.6	59.9	80.3	62.5	30.9	39.7
DEVISE [47]	57.5	56.5	53.2	52.0	72.9	54.2	68.6	59.7	35.4	39.8
SJE [51]	57.1	53.7	55.3	53.9	76.7	65.6	69.5	61.9	32.0	32.9
ESZSL [52]	57.3	54.5	55.1	53.9	74.7	58.2	75.6	58.6	34.4	38.3
SYNC [53]	59.1	56.3	54.1	55.6	72.2	54.0	71.2	46.6	39.7	23.9
SAE [54]	42.4	40.3	33.4	33.3	80.6	53.0	80.7	54.1	8.3	8.3
GFZSL [55]	62.9	60.6	53.0	49.3	80.5	68.3	79.3	63.8	51.3	38.4
CNNTM(+)	**63.9**	59.6	**55.12**	49.9	79.2	**70.7**	**80.5**	**67.2**	**52.4**	**39.9**

TABLE V. Comparing Analysis on Inception-v3-CNN+TM (+) Noted as CNNTM (+)

Methods	SUN			CUB			AWA1			AWA2			aPY		
-	ts	tr	H	ts	tr	H	ts	tr	H	ts	tr	H	ts	tr	H
DAP	4.2	25.1	7.2	1.7	67.9	3.3	0.0	88.7	0.0	0.0	84.7	0.0	4.8	78.3	9.0
IAP	1.0	37.8	1.8	0.2	72.8	0.4	2.1	78.2	4.1	0.9	87.6	1.8	5.7	65.6	10.4
CONSE	6.8	39.9	11.6	1.6	72.2	3.1	0.4	88.6	0.7	0.5	90.6	1.0	0.0	91.2	0.0
CMT	8.1	21.8	11.8	7.2	49.8	12.6	0.9	87.6	1.8	0.5	90.0	1.0	1.4	85.2	2.8
CMT2	8.6	28.0	13.3	4.7	60.1	8.7	8.4	86.9	15.3	8.7	89.0	15.9	10.9	74.2	19.0
SSE	2.1	36.4	4.0	8.5	46.9	14.4	7.0	80.5	12.9	8.1	82.5	14.8	0.2	78.9	0.4
LATEM	14.7	28.8	19.5	15.2	57.3	24.0	7.3	71.7	13.3	11.5	77.3	20.0	0.1	73.0	0.2
ALE	11.8	33.1	26.3	23.7	62.8	34.4	16.8	76.1	27.5	14.0	81.8	23.9	4.6	73.7	8.7
DEVISE	16.9	27.4	20.9	23.8	53.0	32.8	13.4	68.7	22.4	17.1	74.7	27.8	4.9	76.9	9.2
SJE	14.7	30.5	19.8	23.5	59.2	33.6	11.3	74.6	19.6	8.0	73.9	14.4	3.7	55.7	6.9
ESZSL	11.0	27.9	15.8	12.6	63.8	21.0	6.6	75.6	12.1	5.9	77.8	11.0	2.4	70.1	4.6
SYNC	7.9	41.3	13.4	11.5	70.9	19.8	8.9	87.3	16.2	10.0	90.5	18.0	7.4	66.3	13.3
SAE	8.8	18.0	11.8	7.8	54.0	13.6	1.8	77.1	3.5	1.1	82.2	2.2	0.4	80.9	0.9
GFZSL	0.0	39.6	0.0	0.0	45.7	0.0	1.8	80.3	3.5	2.5	80.1	4.8	0.0	83.3	0.0
CNNTM(+)	21.7	42.9	32.1	25.5	58.7	66.1	12.5	81.2	30.1	12.8	82.1	27.3	14.1	83.2	20.1

Current TL algorithms mainly want to improve the performance of the same probability y in the feature space at different times. In the plantar pressure image data-set, the morphological changes of the anterolateral prefrontal cortex and the medial prefrontal cortex, which belong to the visceral motor network system are found to be effective by plantar pressure structural imaging studies. The improved insufficient data-set learning method proposed in this work can solve the problem of insufficient training and accomplish excellent automatic recognition. While TL is mainly applied to small and volatile data sets (such as sensor data, text categorization, image categorization, etc.) in future research it is needed to consider how to use the TL method in a wider range of data scenarios.

References

[1] J. Deng, et al, "ImageNet: A large-scale hierarchical image database," in *Proc. CVPR*, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848

[2] S. Pachori, A. Deshpande and S. Raman, "Hashing in the zero shot framework with domain adaptation," *Neurocomputing*, 275, 2137–2149. doi:10.1016/j.neucom. 2017.10.061

[3] Y. He, Y. Tian, and D. Liu, "Multi-view transfer learning with privileged learning framework," *Neurocomputing*, 335, 131–142. doi:10.1016/j.neucom.2019.01.019

[4] A. Lumini and L. Nanni, "Deep learning and transfer learning features for plankton classification," *Ecological Informatics*, 51, 33–43. doi:10.1016/j.ecoinf.2019.02.007

[5] A. Tavanaei, at al, "Deep learning in spiking neural networks," *Neural Networks*, 111, 47–63. doi:10.1016/j.neunet.2018.12.002

[6] H. Zhang, Y. Long and L. Shao, "Zero-shot hashing with orthogonal projection for image retrieval," *Pattern Recognition Letters*, 117, 201–209. doi:10.1016/j.patrec.2018.04.011

[7] A. M. Anam and M. A. Rushdi, "Classification of scaled texture patterns with transfer learning," *Expert Systems with Applications*, 120, 448–460. doi:10.1016/j.eswa.2018.11.033

[8] Y. Wu, Y. Su, and Y. Demiris, "A morphable template framework for robot learning by demonstration: Integrating one-shot and incremental learning approaches," *Robotics and Autonomous Systems*, 62, 1517–1530. doi:10.1016/j.robot.2014.05.010

[9] S. Cisse, *The Fortuitous Teacher*, pp. 119-143, 2016, Chandos Publishing.

[10] Y. Ji, Y. Yang, X. Xu, and H. T. Shen, "One-shot learning based pattern transition map for action early recognition," *Signal Processing*, 143, 364–370. doi:10.1016/j.sigpro.2017.06.001

[11] U. Mahbub, et al. "A template matching approach of one-shot-learning gesture recognition," *Pattern Recognition Letters*, 34, 1780–1788. doi:10.1016/j.patrec.2012.09.014

[12] X. Xu, S. Gong, and T. M. Hospedales, "Zero-shot crowd behavior recognition," pp. 341–369. In V. Murino, M. Cristani, S. Shah and S. Savarese (eds) *Group and Crowd Behavior for Computer Vision*. doi:10.1016/B978-0-12-809276-7.00018-7

[13] Z. Abderrahmane, G. Ganesh, A. Crosnier and A, Cherubini, "Haptic zero-shot learning: recognition of objects never touched before," *Robotics and Autonomous Systems*, 105, 11–25. doi:10.1016/j.robot.2018.03.002

[14] Z. Ji, et al., "Zero-shot learning with multi-battery factor analysis," *Signal Processing*, 138, 265–272. doi:10.1016/j.sigpro.2017.03.023

[15] Z. Ji, et al., "Manifold regularized cross-modal embedding for zero-shot learning," *Information Sciences*, 378, 48–58. doi:10.1016/j.ins.2016.10.025

[16] M. Liu, D. Zhang and S. Chen, "Attribute relation learning for zero-shot classification," *Neurocomputing*, 139, 34–46. doi:10.1016/j.neucom.2013.09.056

[17] B. Liu, et al., "Combining ontology and reinforcement learning for zero-shot classification," *Knowledge-Based Systems*, 144, 42–50. doi:10.1016/j.knosys.2017.12.022

[18] T. Long, et al., "Zero-shot learning via discriminative representation extraction," *Pattern Recognition Letters*, 109, 27–34. doi:10.1016/j.patrec.2017.09.030

[19] Y. Yu, Z. Ji, J. Guo and Y. Pang, "Zero-shot learning with regularized cross-modality ranking", *Neurocomputing*, 259, 14–20. doi:10.1016/j.neucom.2016.06.085

[20] A. Bear and D. G. Rand, "Modeling intuition's origins", *J. Applied Research in Memory and Cognition*, 5, 341–344. doi:10.1016/j.jarmac.2016.06.003

[21] S. Blaes and T. Burwick, "Few-shot learning in deep networks through global prototyping," *Neural Networks*, 94, 159–172. doi:10.1016/j.neunet.2017.07.001

[22] A. Chen, et al., "On the systematic method to enhance the epiphany Ability of individuals," *Procedia Computer Science*, 31, 740–746. doi:10.1016/j.procs.2014.05.322

[23] K. Dombroski, "Learning to be affected: Maternal connection, intuition and elimination communication," *Emotion, Space and Society*, 26, 72–79. doi:10.1016/j.emospa.2017.09.004

[24] J. Li, et al., "What is in a name?: The development of cross-cultural differences in referential intuitions," *Cognition*, 171, 108–111. doi:10.1016/j.cognition.2017.10.022

[25] X. Li, et al., "Zero-shot classification by transferring knowledge and preserving data structure," *Neurocomputing*, 238, 76–83. doi:10.1016/j.neucom.2017.01.038

[26] A. Soudani, and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," *Expert Systems with Applications*, 118, 400–410. doi:10.1016/j.eswa.2018.10.029

[27] K. Deschamps, et al., "Efficacy measures associated to a plantar pressure-based classification system in diabetic foot medicine," *Gait and Posture*, 49, 168–175. doi:10.1016/j.gaitpost.2016.07.009

[28] J. A. Ramirez-Bautista, et al., "Review on plantar data analysis for disease diagnosis," *Biocybernetics and Biomedical Engineering*, 38, 2, 342–361. doi:10.1016/j.bbe.2018.02.004

[29] J. Sommerset, et al., "Plantar acceleration time: A novel technique to evaluate arterial flow to the foot," *Annals of Vascular Surgery*, 60, 308–314. doi: 10.1016/j.avsg.2019.03.002

[30] Y. Xia, et al., "A convolutional neural network cascade for plantar pressure images registration," *Gait and Posture*, 68, 403–408. doi:10.1016/j.gaitpost.2018.12.021

[31] C. Wang and G. Tzanetakis, "Singing style investigation by residual Siamese convolutional neural networks," in *Proc. ICASSP*, 2018, pp 116–120. doi:10.1109/ICASSP.2018.8461660

[32] J. Lu, et al., "Boosting few-shot image recognition via domain alignment prototypical networks," in Proc. ICTAI, 2018, pp. 260–264. doi: 10.1109/ICTAI.2018.00048

[33] J. Shi, et al., "Concept learning through deep reinforcement learning with memory-augmented neural networks," *Neural Networks*, 110, 47–54. doi:10.1016/j.neunet.2018.10.018

[34] T. L. Griffiths, et al., "Doing more with less: Meta-reasoning and meta-learning in humans and machines," *Current Opinion in Behavioral Sciences*, 29, 24–30. doi:10.1016/j.cobeha.2019.01.005

[35] S. Pramanik and A. Hussain, "Text normalization using memory augmented neural networks," *Speech Communication*, 109, 15–23. doi:10.1016/j.specom.2019.02.003

[36] A. E. Gutierrez-Rodríguez, et al., "Selecting meta-heuristics for solving vehicle routing problems with time windows via meta-learning," *Expert Systems with Applications*, 118, 470–481. doi:10.1016/j.eswa.2018.10.036

[37] Y, Xian, et al., "Zero-shot learning - a comprehensive evaluation of the good, the bad and the ugly," *IEEE Trans. Pattern Analysis and Machine Intelligence*, early-access. doi: 10.1109/TPAMI.2018.2857768

[38] S. J, Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowledge and Data Eng.*, 22, 1345–1359. doi: 10.1109/TKDE.2009.191

[39] B. Gong, Y. Shi, F. Sha and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in Proc. CVPR, 2012, pp. 2066–2073.

doi:10.1109/CVPR.2012.6247911

[40] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?" *ICT Express*, 4, 247–254. doi:10.1016/j.icte.2018.10.007

[41] H. Hermessi, O. Mourali and E. Zagrouba, "Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning," *Expert Systems with Applications*, 120, 116–127. doi:10.1016/j.eswa.2018.11.025

[42] P. Herent, et al., "Detection and characterization of MRI breast lesions using deep learning," *Diagnostic and Interventional Imaging*, 100, 219–225. doi:10.1016/j.diii.2019.02.008

[43] G. Patterson and J. Hays, "SUN attribute database: Discovering, annotating, and recognizing scene attributes," in *Proc. CVPR*, 2012, pp. 2751–2758. doi:10.1109/CVPR.2012.6247998

[44] P. Welinder, et al., Caltech-UCSD Birds 200. Computation & Neural Systems Technical Report CNS-TR: 1–15. https://authors.library.caltech.edu/27468

[45] C. H. Lampert, H Nickisch and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 36, 453–465. doi:10.1109/TPAMI.2013.140

[46] A. Farhadi, I. Endres, D. Hoiem and D. Forsyth, "Describing objects by their attributes," in *Proc. CVPR*, 2009, pp. 1778–1785. doi:10.1109/CVPR.2009.5206772

[47] N. M. Norouzi, et al., "Zero-shot learning by convex combination of semantic embeddings," arXiv:1802.10408.

[48] R. Socher, M. Ganjoo, C. D.Manning and A. Ng, "Zero-shot learning through cross-modal transfer," in *Proc. Int. Conf. Neural Information Processing Systems*, p.p. 935–943.

[49] Z. Zhang and V. Saligrama, "Zero-shot learning via semantic similarity embedding," in *Proc. ICCV*, 2015, pp. 4166–4174. doi: 10.1109/ICCV.2015.474

[50] Y. Xian, et al., "Latent embeddings for zero-shot classification," in *Proc. CVPR*, 2016, pp. 69–77. doi:10.1109/CVPR.2016.15

[51] Z. Akata, et al., "Evaluation of output embeddings for fine-grained image classification," in *Proc. CVPR*, 2015, pp. 2927–2936. doi:10.1109/CVPR.2015.7298911

[52] B. Romera-Paredes and P. H. S. Torr, "An embarrassingly simple approach to zero-shot learning," pp. 11–30. In: R. Feris, C. Lampert and D. Parikh (eds), *Visual Attributes. Advances in Computer Vision and Pattern Recognition*. doi:10.1007/978-3-319-50077-5_2

[53] S. Changpinyo, W. Chao, B. Gong and F. Sha, "Synthesized classifiers for zero-shot learning," in *Proc. CVPR*, 2016, pp. 5327–5336. doi:10.1109/CVPR.2016.575

[54] E. Kodirov, T. Xiang and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proc. CVPR*, 2017, pp. 4447–4456. doi:10.1109/CVPR.2017.473

[55] V. K. Verma and P. Rai, "A simple exponential family framework for zero-shot learning," in: M. Ceci, J. Hollmén, L. Todorovski, C. Vens and S. Džeroski (eds), *Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science*, 10535. doi:10.1007/978-3-319-71246-8_48

Yao Wu

Yao Wu is an Assistant Professor of Department of Industrial Design, Wenzhou Business College, China. He holds a master's degree in software engineering from Huazhong University of Science and Technology in Wuhan, China, in 2014. and a Bachelor degree of consumer product design from Shaanxi University of Science and Technology, China, in 2005. His research interests include integrated innovation design and sustainable design.

Qun Wu

Qun Wu, is an Associate Professor of Human Factor at the Institute of Universal Design, Zhejiang Sci-Tech University, China. He received his Ph.D. in College of Computer Science and Technology from Zhejiang University, China, in 2008. He holds a B.E. degree in Industrial Design from Nanchang University, China, in 2001, and a M.E. degree in Mechanical Engineering from Shaanxi University of Science and Technology, China, in 2004. His research interests include machine learning, human factor and product innovation design.
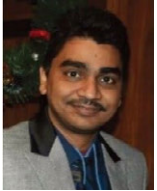
Nilanjan Dey

Nilanjan Dey, was born in Kolkata, India, in 1984. He received his B.Tech. degree in Information Technology from West Bengal University of Technology in 2005, M. Tech. in Information Technology in 2011 from the same University and Ph.D. in digital image processing in 2015 from Jadavpur University, India. In 2011, he was appointed as an Asst. Professor in the Department of Information Technology at JIS College of Engineering, Kalyani, India followed by Bengal College of Engineering College, Durgapur, India in 2014. He is now employed as an Asst. Professor in Department of Information Technology, Techno India College of Technology, India. His research topic is signal processing, machine learning, and information security. Dr. Dey is an Associate Editor of IEEE Access and is currently the Editor-in-Chief of the International Journal of Ambient Computing and Intelligence, and Series co-editor of Springer Tracts of Nature-Inspired Computing (STNIC).

R. Simon Sherratt

R. Simon Sherratt (M'97–SM'02–F'12), received the B.Eng. degree in Electronic Systems and Control Engineering from Sheffield City Polytechnic, UK in 1992, M.Sc. in Data Telecommunications in 1994 and Ph.D. in video signal processing in 1996 from the University of Salford, UK. In 1996, he has appointed as a Lecturer in Electronic Engineering at the University of Reading where he is now Professor of Biosensors. His research topic is signal processing and communications in consumer devices focusing on wearable devices and healthcare. He is a reviewer for the IEEE Sensors Journal is now the Emeritus Editor-in-Chief of the IEEE Transactions on Consumer Electronics.