Perceptual Visual Interactive Learning

Shenglan Liu, Member, IEEE, Xiang Liu, Yang Liu, Lin Feng, Hong Qiao, Senior Member, IEEE Jian Zhou, Yang Wang

Abstract—Supervised learning methods are widely used in machine learning. However, the lack of labels in existing data limits the application of these technologies. Visual interactive learning (VIL) compared with computers can avoid semantic gap, and solve the labeling problem of small label quantity (SLQ) samples in a groundbreaking way. In order to fully understand the importance of VIL to the interaction process, we re-summarize the interactive learning related algorithms (e.g. clustering, classification, retrieval etc.) from the perspective of VIL. Note that, perception and cognition are two main visual processes of VIL. On this basis, we propose a perceptual visual interactive learning (PVIL) framework, which adopts gestalt principle to design interaction strategy and multi-dimensionality reduction (MDR) to optimize the process of visualization. The advantage of PVIL framework is that it combines computer’s sensitivity of detailed features and human’s overall understanding of global tasks. Experimental results validate that the framework is superior to traditional computer labeling methods (such as label propagation) in both accuracy and efficiency, which achieves significant classification results on dense distribution and sparse classes dataset.

Index Terms—Semantic Gap, Visual Interactive Learning, Gestalt Principle, Multi-Dimensionality Reduction

1 INTRODUCTION

Human vision and perception play an important role in data preprocessing, which has been illustrated in many previous works [1] [2] [3]. Visual interactive learning methods, transmitting human’s perceptual information of data to computer, are different from traditional machine learning methods [4]. Such as the semantic gap of image retrieval [5], the semantic segmentation of images [6] and so on. In image retrieval, relevant feedback labels related (or unrelated) images of human cognitive according to the original ranking list of query, which constructs a new ranking model to rerank the images to avoid the image semantic gap as much as possible [7]. In image segmentation, Li et al. utilize human visual perception and cognition of images to achieve semi-supervised segmentation through interaction between humans and images [8]. Besides, human-computer interaction related studies have improved the effect of image processing through eye movement [9] and region of interest (ROI) [10]. In general, the key point of visual interactive learning is extracting important information (e.g. position, contour) from data or images using visualization. Perception and cognition are two main visual processes of humans receiving information. Depending on existing technology (e.g. image segmentation, image retrieval), we re-summarize visual interactive learning from the aspects of perception and perception with cognition. Visual interactive learning is defined as a method of machine learning which interacts data using human’s perception and cognition.

It is illustrated that visual interactive labeling is a key issue in visual interactive learning, which is always used in machine learning methods such as classification [4], retrieval [11], and clustering [12]. In recent years, many related machine learning methods have been proposed. However, the SLQ problems limit the development of these technologies, especially in the classification method where a large number of labels are needed, such as deep learning methods [13] [14]. Therefore, labeling is an important issue that is difficult to solve in SLQ machine learning. The traditional classification framework for SLQ is generally classified by support vector machine (SVM) [15] after expert labels or active learning [16] [17]. However, this approach is expensive in labeling and calculating, so we hope to develop an efficient interactive learning method to solve the SLQ problem.

Classification is one of the major applications of data labeling, which can generally be fallen into non-metric models and metric-related models. Non-metric models can be regarded as perceptually unrelated models, which are difficult to understand with human perception, i.e. iterative dichotomiser 3 (ID3) [18], classification and regression tree (CART) [19] and other decision tree methods. The other is metric-related classification approaches, which are easy to visualize and conform to the human perception system. One typical classification model is k-nearest-neighbor (KNN) classifier [20]. The classical KNN model is based on the $L_2$-norm metric. To improve the performance of KNN, some extended versions (e.g. $L_1$-norm based KNN, extended nearest neighbor (ENN) and cosine metric based KNN [21] [22]) are proposed according to KNN framework. However, lazy learning methods (including KNN and the extended versions) strong relate to the distribution and the number of training samples (KNN-based methods are always difficult to apply to large-scale data sets and are quite time consuming.).

On this basis, many model-based classification methods such as SVM [15], random vector functional link (RVFL) [23] and other discriminant models are proposed, which only needs to learn a linear or nonlinear function to achieve fast classification. Since then, with the rapid development of digitalization, neural network classifiers under big data, such as the deep learning (DL) methods, have been widely used and have achieved outstanding results on ImageNet [24] [25] [26]. However, labeling is the
bottleneck of the DL methods in big data. Therefore, data labeling has become an important issue that needs to be resolved in the past two years. Traditional data labeling is a machine-to-machine learning method, such as label propagation method [27], utilizing the potential manifold structure and nonlinear metrics of the data to pass label to the unlabeled samples, can be considered as a classification method for SLQ samples.

In recent years, data labeling has achieved some remarkable results, especially in visualization labeling field. The categories of labels include video, text, time series, and scatter [28]. For example, Hoferlin used a specific interactive classifier for video visualization analysis [29]. Heimer et al. reduced the workload of obtaining labels in text search through interactive training [30]. Sarkar and Bernard et al. obtained relevant labels for semi-supervised interaction models for industrial data [31] and human motion time series data [32], respectively. Sedlmair et al. proposed an interactive model for measuring scatter-ability, correlation, and outliers [33]. In addition, the types of labeling include classification tasks, relevant feedback, correlation coefficients, etc. For example, relevant feedback in the field of image retrieval can effectively improve the accuracy of retrieval [34] [35]. Bernard et al. proposed a method that characterizing the patient’s health index by actively learning the doctor’s feedback [36]. To enhance the performance of learning tasks, human-to-machine approaches are proposed by combining interactive learning with other machine learning methods. For example, interactive labels are used to reflect similarities in metric learning, and to interpret the interrelationships between instances from user interaction features [37]. Bernard et al. pointed out that interactive labeling is superior to active learning given the condition the class distributions are separated well in dimension reduction [38].

The existing interaction methods are mainly focus on expanding the amount of labels available, improving learning result and reducing the complexity of manual labeling [38] [39], which are instructive and excellent on designing interaction strategies based on existing technical characteristics. However, most of these works are lack of considering human’s visual processing advantages, and limit the further application of visual interaction. At present, there are few interactive learning methods that combine computer process and human perception. This paper points out two important issues that interactive learning should notice: (1) human perception of data exists in a manifold, which may have significant impact on visualization and interaction (e.g. overlapping, bending manifold). (2) human has a strong visual perception over data (especially large amount of data).

In response to the above issues, This paper focuses on the method of visual interactive labeling based on human perception in SLQ datasets and proposes multi-feature selection and multi-dimension reduction approach, which makes full use of the discriminative features of data manifolds, wakens the influence of coupling on visualization, and achieves better labeling results. Meanwhile, this paper introduces Gestalt principle to understand human perception. Gestalt principles are derived from the process of human understanding of the world and have guiding significance for the design of interaction strategies. Based on the understanding perception and data manifold, we propose a perceptual visual interactive learning framework. The framework considers SLQ samples in semi-supervised learning, which significantly improve the efficiency and accuracy for classification. The key contributions of this paper to interactive learning can be summarized as follows.

(1) This paper re-summarizes visual interactive learning (VIL) from the aspects of human visual perception and cognition, and makes a division of VIL, which has important significance for comprehensively understanding the visual-based interactive technology.

(2) This paper points out the reason for the overlapping visual data, and proposes a user-oriented visualization method for multi-feature selection and multi-dimension reduction.

(3) Compared with existing labeling methods (e.g. label propagation), human have a strong visual perception ability over little amount of labels or large amount of data. We point out and analysis the above point of view, and illustrate the PVIL framework proposed in this paper superior to the traditional labeling method in correctness and time consumption, and has a better classification effect on real-world dataset.

(4) PVIL framework adopts the Gestalt principle to explain several criteria for human perception, by which the design process of the interaction strategy is guided. It provides a theoretical basis for visual interactive learning in human perception.

The rest of this paper is organized as follows. Section 2 introduces the computer labeling and human perception. Section 3 summarizes visual interactive learning and introduces our PVIL framework. The experimental results and analysis of this paper are listed in section 4, which gives a comparison between the PVIL framework and computer labeling in both accuracy and efficiency aspects, and the semi-supervised classification experiment is also conducted. The last section is conclusion.

2 RELATED WORK

In VIL community, computer labeling cooperates with human perception. As a typical method of computer labeling, label propagation [27] is briefly introduced in section 2.1. Human perception principles for classification labeling are based on the Gestalt principles [40] [41], which can be found in section 2.2.

2.1 Label Propagation (LP)

Expert labeling, as a common solution to obtain labels, is complicated and time-intensive. By contrast, the key idea of semi-supervised learning is to employ a large amount of unlabeled data and partially labeled data to structure a classifier that achieves relatively high accuracy with less human participants [15]. As a typical semi-supervised learning method, LP [27] algorithm generate labels with the help of the graph weight matrix.

Given a n samples dataset \( X = [x_1, \cdots, x_n] \in \mathbb{R}^{D \times n} \), where \( D \) indicates the dimensions of one sample. LP algorithm develops a fully connected graph by the labeled set \( D \), corresponding labeled samples \( X_l \in \mathbb{R}^{D \times l} \) (Eq. (1)) and the unlabeled set \( X_u \in \mathbb{R}^{D \times u} \) (Eq. (2)), respectively.

\[
X_l = [x_{l_1}, \cdots, x_l]
\]

\[
D_l = [y_{l_1}, \cdots, y_l]
\]

\[
X_u = [x_{l+1}, \cdots, x_n]
\]

where \( x_i \in \mathbb{R}^D, i = 1, 2, \cdots, n \), is a representation of the instance, \( y_j \in \{c_1, c_2, \cdots, c_p\}, j = 1, 2, \cdots, l \) is the class label of the \( j \)-th sample (\( p \) indicates the number of the classes), and \( n = l + u \) (always \( l << u \)).
The weight matrix $W = \{W_{ij}\} \in \mathbb{R}^{n \times n}$ describing weight of each pair of different samples $x_m$ and $x_j$ can be denoted by Eq. (3) as follows.

$$w_{i,j} = \begin{cases} \exp(-\gamma||x_m - x_j||), & x_m, x_j \in N_k(x_i) \\ 0, & \text{else} \end{cases} \quad (3)$$

where $\gamma$ is a hyper-parameter relating to the data distribution. Then, we compute the symmetrically normalize matrix of $W$, which denotes as $S = D^{-1/2}WD^{-1/2}$, where $D$ is a diagonal matrix, and $D_{ii} = \sum_{j=1}^{n} W_{ij}$. In Eq. (3), $|| \cdot ||$ can be $L1/L2$ or other norm.

We denote $Q = \{q_1, \cdots, q_n\} \in \mathbb{R}^n$ as the original label vector which satisfies $q_i = y_i$, $i = 1, 2, \cdots, l$, $q_j = 0$, $j = l + 1, l + 2, \cdots, n$. $f^t$ indicates the $t$-th iteration of label vector, and $f^*$ is the final label result with labeled $q_j$, $j = l+1, l+2, \cdots, l+u$. The iteration equation of LP can be expressed by Eq. (4) as follows.

$$f^{t+1} = \lambda S f^t + (1 - \lambda)Q \quad (4)$$

where $\lambda$ is a balance weight parameter of $Sf^t$ and $Q$. We can get $f^*$ by Eq. (4) while $||f^{t+1} - f^*|| < \varepsilon$, where $\varepsilon > 0$ is a small real number to stop the iteration. LP, as a metric-based approach, shows the advantage of computational accuracy in computer labeling. However, in real word applications, most of computer labeling methods are not ideal, which can be attributed to the high complexity of time and space. Furthermore, the process of propagation is easily misled by neighbors.

### 2.2 Gestalt Principle

Gestalt psychology arises from the research of perception (as famously noted by Koffka [42]), the process of perception could be understood as a series of Gestalt principles, which have a close relationship with human visual models, and then can be very widely used to assist the design of interactive models [27] [40] [43] [44]. In this study, the theory of visual labeling involves three principles, which are proximity, closed and continuity respectively.

![Fig. 1: Gestalt principle of human perception](image)

**2.2.1 Proximity Principle**

The proximity principle indicates that elements with proximate positions are tend to be classified into one category. Assume that there are three independent elements, which are $A_1$, $A_2$ and $A_3$ respectively. Based on this principle, the symbolized form can be written as Eq. (5).

$$[A_1, A_2, A_3] \rightarrow [A_1, A_2, A_3] \quad (5)$$

As shown in Fig. 1 (a1) and Fig. 1 (a2), $A_1$ and $A_2$ could be classified into one group when $A_3$ does not exist. Similarly when $A_3$ is added, $A_2$ and $A_3$ tend to be put together. The different distance between elements results in different classification situation in Fig. 1 (a1) and Fig. 1 (a2). It is worth noting that the distance measure plays a key role in evaluation indicator for various classification tasks. The common unsupervised classifier KNN, for example, is focuses on the neighbor distance [27]. Also, the loss function of the linear classifier’s gradient descent is based on the distance between the sample and the hyperplane [43].

### 2.2.2 Closed Principle

The closed principle suggests that elements with integrated boundary are tend to classified into one category. Suppose that the independent elements can be represented by $A_1$, $A_2$, $\ldots$, $A_n$, where $A = \{A_1, A_2, \ldots, A_{n-1}\}$. Its symbolization can be expressed as Eq. (6).

$$A_1, A_2, \ldots, A_{n-1}, A_n \rightarrow [A_1, A_2, \ldots, A_{n-1}, [A_n] \quad (6)$$

As illustrated in Fig. 1 (b), the closest element to the $A_{n-1}$ is $A_n$, while $A_{n-1}$ and the rest of the elements ($A_1$, $\cdots$, $A_{n-2}$) is classified into a same set, of course, $A_n$ is grouped independently. Apart from $A_n$, the remaining elements can be gathered into a meaningful set $A$. That is, set $A$ is consist with the priori combination hypothesis of a certain shape. Moreover, in human version, set $A$ has a closed tendency. In the process of perceiving the world, human vision tends to combine scattered individuals, in other words, to understand isolated elements from a macroscopic perspective [46]. In visual analysis, the significant different between human and computer is that, the perception of humans towards real world can be viewed as a collection of elements. As far as our knowledge is concerned, the advantage of human perception derived from its macroscopically discriminant ability, by sharp contrast, computer merely has ability to calculate relationships between certain isolated elements.

### 2.2.3 Good Continuity Principle

Good continuity principle suggests that the elements with shared lines, curves or planes are tend to be seen as an integral. In certain cases, an integrated element may be divided into several separated parts. However, part of continuous information (e.g. curvature continuity information, dimensional continuity information, etc.) of the origin element is retained in the separated parts. Good continuity principle indicates that human vision tend to recruit the integrated element by means of these continuous information. Assuming there are several discrete elements, $A_1$, $A_2$, $A_3$, and $A_4$ respectively, and $A$ is a visual-integrated element. The symbolized form of good continuity principle can be denoted as Eq. (7).

$$[A_1 + A_2 + A_3] = [A] \quad (7)$$

As clearly revealed in Fig. 1 (c), $A_1$ and $A_2$ are broken in $A$, and the retained continuous information mainly includes curvature continuity information. Human vision tends to make judgments that treating a manifold with continuous curvature as an integral [46]. The connection of $A_4$ and $A_1$ simply meet the first-order continuity in $M$, while the connection of $A_2$ and $A_1$ not only confirm to the first-order continuity but also meet the second-order continuity. Therefore, $A_1$, $A_2$, and $A_3$ are assigned into a visual-integrated element $A$ ($A_2$ and $A_3$ are assigned in a similar way with $A_1$ and $A_2$). The continuous information can not accurately...
perceived by computer in the process of calculation, which can be
deemed as a disadvantage of computer in comparison with human

3 Visual Interactive Learning

The key aspect of visual interactive learning is extracting useful
information from data for machine learning by using visualization
method. In section 3.1, Visual interactive learning are re-
summarized by perception and cognition of visual processes.
Section 3.2 introduces our multi-dimensionality reduction method
which can solve data overlap problem. The perceptual visual
interactive learning framework is introduced in the last subsection.

3.1 Visual Interactive Learning (VIL)

The aim of VIL is to label samples of dataset from both computer’s
computing ability and human’s visual understanding ability. The
three Gestalt principles, described in subsection 2.2, illustrate the
superiority of human visual perception which motivates us to
survey VIL on a new point of view. We re-summarize the VIL
according to distinction and relationship between computer and
visual perception of human in Fig. 2. According to the process of
VIL, A, B, C and D are introduced respectively.

A mainly includes VIL related technologies (clustering, di-

dimensionality reduction and classification) and applications (image
segmentation and image retrieval). In general, VIL interacts with
human from the aspect of vision. Of course, brain-signal-based
BCI can also be deemed as a VIL method. B is denoted as a
information extraction method. For the purpose of visualization,
B maps samples to a certain visual space. B could be seen as
concrete realization of A, which includes eye movement, ROI and
relevance feedback. Depend on the method of B, C is indicated
as a data interaction approach with human. After data is mapped
to an image in visual space, human could decide how to interact
with the image after visualization through operability (perception)
and requirement (cognition). As an interaction approach, C mainly
includes unsupervised perception, supervised perception, cogni-
tion, and combination of perception and cognition. D is expressed
as a Labeling approach, which can also be seen as an interactive
result of C. The information of perception and cognition from
human are transmitted to computer by means of labels. With
the help of the labels, the model computed by computer is in
line with human needs. D is made up of soft labeling, soft-hard
labeling, and hard labeling in perception and cognition degrees,
respectively. Unsupervised perception and supervised perception
could be explained by Gestalt principles (as is illustrated in
subsection 2.2), which are fallen into soft labeling. Cognition-driven
labeling in C is categorized into hard labeling, which is frequently
adopted in applications of goal orientations. Furthermore, soft-
hard labeling lies between soft and hard labeling, which utilizes
both perception and cognition in labeling tasks. The interaction-
label idea, embedded in D, is widely used in various technologies
and applications. The subsequent work of this paper is focus on
soft labeling.

3.2 Perceptual Visual Interactive Learning (PVIL)

An important purpose and application of interactive learning is
classification. This paper proposes a perceptual visual interaction
framework for classification problems (Fig. 2). PVIL, a perception
based VIL method, is to solve the problem that high-dimensional
space cannot be perceived and labeled by human vision in SLQ.
The framework achieved high accuracy in the case of SLQ
classification. PVIL is introduced as follows.

PVIL is divided into three levels: the input layers, the model
layers, and the output layer. The input layers include data ini-

tialization and preprocessing, which is detailed by equation (8)
and (9) in subsection 2.1. Labeled data and unlabeled data are
initialized by Eq. (8) as follows.

\[ H_1 = \begin{bmatrix} X_m \\ X_n \end{bmatrix} label = \begin{bmatrix} 0 \\ Y_n \end{bmatrix} \]  (8)

And preprocessing is denoted as Eq. (9).

\[ H_2 = Pro(H_1) \]  (9)
where $\text{Pro} = \{\text{Normalize, Regulation, . . .}\}$ In general, $H_1, H_2 \subseteq X_{(m+n)+h}$.

The samples here include scatter, image, music, video, etc. In this paper, we adopt scatter and image as examples (part1, part2 is scatter, part3 is image) and the pre-processing part performs appropriate scale transformation for different sample types. For example, the image is normalized [47] to a standard normal distribution. The model layers contain feature selection, dimensionality reduction, visualization, and human-computer interaction. Feature selection/extraction (Eq. (10)) refers to the representation of original data [48], such as the scale invariant feature transform (SIFT) feature [49], the histogram of oriented gradient (HOG) feature [50], and the deep convolution feature [51] in image understanding.

$$H_3 = FS(H_2) \quad (10)$$

where $H_2 \subseteq X_{s+h}$ then $H_3 \subseteq X_{s+t}$. As for dimensionality reduction (Eq. (11)), it refers to reducing the data dimensions on the premise of maintaining sample discriminative information as much as possible in classification issue.

$$H_4 = Dr(H_3) \quad (11)$$

where $H_4 \subseteq X_{s+t}$. Dimensionality reduction methods can fall into many categories, such as linear methods principal component analysis (PCA) [52], maximal similarity embedding (MSE) [22], and nonlinear ones isometric mapping (ISOMAP) [53]. Locally Linear Embedding (LLE) [54], local tangent space alignment (LTSA) [55], t-distributed stochastic neighbor embedding (TSNE) [56], symmetric positive definite (SPD) manifold dimension reduction [57], path based Isomap [58] and so on. In our proposed PVIL, visualization (Eq. (12)) aims to reducing high-dimensional data to dimensions that humans can perceive, typically two-dimensional or three-dimensional.

$$H_5 = Vis(H_4) \quad (12)$$

Generally speaking, visualization can be treated as a special application of dimensionality reduction, that is $Vis = DR_2, H_5 \subseteq X_{s+2}$. In addition, the human-machine interface is used to control the loop of the above process (Fig. 3), which interact with user directly. It judges whether multiple dimension reduction and feature selection are required based on user input. The algorithm processing steps of the human-computer interface are listed in Algorithm 1.

Algorithm 1 Human-Computer Interface

Input:

1. $H_5 \subseteq X_{s+2}$

Output:

$$LabelOrd[LabelIndex_{k+1}, Label_{k+1}]$$

1. $res = LabelIndex_{k+1} \cap X_n$
2. if $Label[\text{res}].\text{argmax}() < Label[\text{res}].\text{count} \ast \eta$
   2. $H_2 = \text{Pro}(H_1[\text{res},:])$
   2. JUMP TO $H2$
3. else
   3. $LabelOrd[LabelIndex_{k+1}, 1] = Label[\text{res}].\text{argmax}()$
4. end if
5. if not $Smit$ then
   5. JUMP TO $H5$
6. end if
7. END

where $H_5$ is the result of visualization. We illustrate that $LabelIndex_{k+1} \subseteq \{1, 2, 3, . . . , k\}$, $Label_{k+1} \subseteq \{c_1, c_2, . . . , c_k\}$, $Smit \in \{0, 1\}$, where $Smit$ indicates whether the user completed the experiment, $\eta \in [0.8, 0.9]$ is a hyper-parameter. At last, the output layer saves class information calculated in the previous layer.

3.3 MDR Visualization using PVIL

Dimensionality reduction which preserves the important relationship between samples is a key approach of implementing PVIL for completing interactive tasks with human perception. However, visualization utilizing dimensionality reduction may lead to some incorrect results. For example, data overlapping makes human perception inefficient in low-dimensional space (Fig. 4). In the case of (a) in Fig. 4: Data Distribution: data overlaps itself in original space (see class A1 and class A2 in Fig. 4). Any feature is difficult to separate it. For case (b) in Fig. 4: Feature Selection: the origin dataset is separable in high-dimensional space, or original data (e.g. text or image separable). We need to represent the dataset by a feature extraction method. However, it is difficult to separate overlapping samples by using the inappropriate feature for human interaction (e.g. class B1 and class B2 in ellipse B of Fig. 4). (c) Dimension Reduction: If the samples are separable in the first two stage (a) and (b) of Fig. 4, inappropriate dimensionality reduction method may also leads to data overlapping (see class C1 and class C2 in ellipse C of Fig. 4) [59].

3.3.1 Compatibility problem of dimension reduction visualization and data distribution

In this subsection, we will illustrate why dimension reduction and data distribution combining lead an inappropriate projection subspace in PVIL (see the case of Fig. 4). We point out that model structure of dimension reduction method should fit the data distribution. Here, we use PCA as a dimension reduction method,

1. Here, feature extraction means computing a feature by domain knowledge. For example, feature extraction of an image uses perceptual uniform descriptor (PUD) feature [21].
Feature Selection

Data Distribution

-60
-40
-20
0
1
50
PVIL reduction and data distribution/feature selection is important for compatibility between dimension can conclude the problem that of LDA and MMC [59] (see Fig. 5 (b) and (c)). Therefore, we 4.3 or reference [59]). This problem also happens in the scene similar conditions (see Fig. 5, and details can be refer to subsection 4.3). This is also the key idea of the perceptual visual interactive framework proposed in this paper.

Regarding dimensions, the methods are similar to features. But dimensionality reduction also involves the imbalance of data distribution. Multiple dimensionality reduction only selects overlapping area(s) of visualization graph (e.g. Fig. 4 or Fig. 5), so that the real subspace (e.g. Fig. 5 (d)) which can achieve high performance in classification is not affected by the distribution of data. The detail of Compatibility problem of dimension reduction visualization and data distribution can be referred section 3.3.1. In essence, the above method achieves the goal that linearly separable data in high-dimensional space is still holding in low-dimensional one. This is also the key idea of the perceptual visual interactive framework proposed in this paper.

In order to solve the problem of section 3.3.1, this paper proposes a “multiple feature selection and multiple dimensionality reduc-

\[ W_{opt}^{PCA} = \arg \max_W tr(W S_t W^T) \] (13)

where \( W \in \mathbb{R}^{D \times d} \), \( W^TW = I \), and \( S_t = \hat{X}^T \hat{X} \) denotes covariance matrix of \( X \), where \( \hat{X} = 1/\sqrt{n-1} [x_1 - \mu, \ldots, x_n - \mu] \), \( \mu = 1/n \sum_{i=1}^n x_i \). \( W_{opt}^{PCA} \) can be got by the eigenvectors corresponding the first \( d \) largest eigenvalues of \( S_t \). PCA aims to project the original (or after feature selection) dataset to a low-dimensional subspace and preserves the largest variance of samples, which makes data overlapping by the incompatibility between model and data distribution in four classes dataset or similar conditions (see Fig. 5 and details can be refer to subsection 4.3 or reference [59]). This problem also happens in the scene of LDA and MMC [59] (see Fig. 5 (b) and (c)). Therefore, we can conclude the problem that compatibility between dimension reduction and data distribution/feature selection is important for PVIL.

Data Distribution | Feature Selection | Dimension Reduction

(a) (b) (c)

Fig. 4: Three cases of data overlap in low-dimensional space

Fig. 5: The four classes dataset

the optimization problem of PCA can be written by Eq. (13) as follows.

4 Experimental Results and Analysis

To validate the effectiveness and efficiency of VIL and the proposed PVIL framework, experiments are conducted on synthetic datasets and real-world dataset. Note that, user participation are required in the the process of VIL. Details of the experimental settings are introduced in section 4.1. To evaluate the influences of PVIL on labeling, experiment is conducted on a synthetic dataset and label propagation (LP) is adopted as the comparing method (Details in section 4.2). Moreover, dimensional reduction experiment is also conducted to fully reveal the effectiveness of MDR (Details in section 4.3). The proposed PVIL aims at combining computer’s sensitivity of detailed features and human’s overall understanding of a task. Hence, classification experiment is conducted to evaluate the effectiveness of PVIL (Details in section 4.4).

4.1 Experimental Introduction

In our experiment, 20 participants (12 males, 8 females) are recruited whose ages are ranging from 18 to 23. Each participant has a normal vision ability and receives college education over one year or more. In our experiment, participants are required to have no idea of the process of VIL. Even some of the participants are interested in interactive learning, no one has expertise in the field, or have seen the detailed distribution of the dataset. An experimental tool is designed and arranged on a high-performance server in our experiment to realize user-machine interface. Each

2. We only use multi-dimensionality reduction visualization because multiple feature selection utilizes the similar approach and get the similar results in the rest of this paper.
participant connects to the tool through a web page. The use of the experimental tool can be divided into 4 steps, as shown in Fig. 7.

To ensure the completeness and objectivity of the experiments, experimental process is detailed introduced to the participants. In particular, the best labeling results will be rewarded. During the experiment, each user is required to record the labeling strategy used during the labeling process, such as how to handle critical data, choose aggressive or conservative ways to complete the data labeling task. The strategies are recorded by the participants on an issued form. The entire experiment time is approximately 90 minutes. We mainly select the following two indicators to judge the differences between different labeling methods: (1) Unlabeled rate is defined as the percentage of the number of unlabeled data ($DATA_{unlabeled}$) to the number of total data ($DATA$), which is used to show the proportion of the labels. It is denoted as $R_{unlabeled} = DATA_{unlabeled}/DATA$.

(2) The data accuracy is measured by the F1-measurement, which is commonly used in classification experiments. Given a set of data, according to the classification criteria, we count number $N_{TP}$ of True Positives, False Positives ($N_{FP}$) and False Negatives ($N_{FN}$). We compute the following performance measure: recall ($R = N_{TP}/(N_{TP} + N_{FN})$), precision ($P = N_{TP}/(N_{TP} + N_{FP})$) and $F_1$ score ($Acc_{F1} = 2PR/(P + R)$).

### 4.2 Labeling Performance Comparison between LP and PVIL on Synthetic Dataset

We conduct an experiment on a two moons synthetic dataset as is shown in Fig. 9 (a) to assess the labeling performance of LP and PVIL. The origin dataset is consisted of 200 samples with two classes [60]. We also increase the amount of samples up to $N(10^5)$, where $N(\cdot)$ indicates samples’ amount.

3. “x” shape is only used for illustrating the good continuity of human perception.
Fig. 9: The comparison of LP and PVIL in accuracy and time consuming. (a) The accuracy of LP and PVIL at different unlabeled quantities. (b) The time consuming of LP and PVIL with different data levels.

direction similar to S1, S2 (as shown in Fig. 8 (a)). Worse still, since the choice of improper propagation direction, wrong labeling results would be calculated. It can be seen in Fig. 9 (a), when $R_{unlabeled}$ is above 95%, the Acc$_{f1}$ is under 60% in certain processes of LP. The similar situation of Fig. 8 (a) is particularly evident in the data manifold of Fig. 8 (b). As the distance of two manifolds is relative small in the directions of S3 and S4 (as shown in Fig. 8 (b)), it is difficult to select a proper subspace in the process of LP algorithm.

In contrast, PVIL has a distinct advantage over LP, which derives mainly from the human visual perception ability. During the process of similarity calculation, closed or warped data can be inferred as uniform manifold by PVIL. Euclidean distance, thus, will be put on the back burner in the decision process. As is shown in Fig. 8 (a), improper propagation directions (S1 and S2) are ignored by PVIL. Similarly, proper propagation directions of S3 and S4 in Fig. 8 (b) are chosen, respectively. As illustrated in Fig. 9 (a), the PVIL algorithm achieves the Acc$_{f1}$ of 99% under $R_{unlabeled}$ around 99%. Hence, the PVIL could outperform LP on dense distribution and sparse classes dataset.

TABLE 1: Time cost comparison between LP and PVIL

| $N (10^k)$ | LP        | PVIL     |
|------------|-----------|----------|
| $10^2$     | $10^{-3}s$| < $10^1s$|
| $10^3$     | $10^{-3}s$| $< 10^1s$|
| $10^4$     | $10^1s$   | $< 10^1s$|
| $10^5$     | $10^3s$   | $< 10^1s$|
| $10^6$     | None      | $< 10^1s$|

PVIL realizes labeling by utilizing human perception, which is quite time-efficient regardless of the amount of samples. Hence, PVIL will be more time-saving than LP under a large-size dataset. The time complexity of LP is $O(n^2)$, while the time complexity of PVIL is $O(1)$

5. The task is difficult for supervised dimensionality reduction methods such as linear discriminant analysis (LDA) due to the fact that the unlabeled rate $R_{unlabeled}$ of the dataset (95.5%) is too high.

4. For convenience, the time complexity of dimensionality reduction is ignored in this experiment.

4.3 Labeling Performance of MDR

To review the performance of MDR with PVIL, we compare it with other unsupervised dimensional reduction methods (PCA, ISOMAP and T-SNE). A four-class synthetic dataset containing 4800 instances (as is shown in Fig. 10) is involved in this experiment. The covariance matrix and the means of the four classes dataset are listed in Table. 2 [59]. And the scatters of the dataset are showed in Fig. 5 (d).

Fig. 10: The dimension reduction (PCA, ISOMAP, T-SNE and Twice PCA) results of four-classes dataset. The dataset cannot be separated by one-time global methods (PCA and ISOMAP).

TABLE 2: The parameter settings of four classes dataset

| Mean       | Covariance Matrix |
|------------|-------------------|
|            |                   |
| 0.3        | 0.04              | 0.06             |
| 0.04       | 0.2               | 0.05             |
| 0.06       | 0.05              | 0.2              |

TABLE 3: Comparison of accuracy and time complexity of PCA, ISOMAP, T-SNE and MDR(Twice PCA)

|                | Acc$_{f1}$ | Time(s)  |
|----------------|------------|----------|
| PCA            | 0.583      | 0.004    |
| ISOMAP         | 0.812      | 6.204    |
| T-SNE          | 0.999      | 55.536   |
| MDR(Twice PCA) | 0.994      | 0.010    |

4. For convenience, the time complexity of dimensionality reduction is ignored in this experiment.
As detailed in subsection 3.3, overlap may exist in the dimensionality reduction result of some datasets. The primary reason for the overlap is the existence of outliers. These outliers mislead the process of dimensionality reduction by affecting the choice of subspace. It can be seen in Table 3 that the Acc$_{f1}$ of the PCA and ISOMAP are 58% and 81%, respectively. The result indicates that the data in our four-classes dataset can not be separated by one-time global methods (PCA and ISOMAP), and the dimensionality reduction process is influenced by outliers (as shown in Fig. 10).

The MDR method could capture semantic information better than that of one-time global methods. Concretely speaking, the distribution of outliers can be located by the first order dimensionality reduction with the help of VIL. Furthermore, the existence of the outliers can be ignored during the process of the second order dimensionality reduction. The advantage of MDR mainly derives from the localization of outliers. As is shown in Table 3, the Acc$_{f1}$ of MDR method is around 99%. MDR achieves comparable performance with T-SNE, which is much better than the other one-time global methods (PCA and ISOMAP).

The time complexities of these dimensionality reduction algorithms are discussed as follows. Note that, the time complexity of MDR(Twice PCA) is near PCA, which can be regarded as PCA algorithm. To realize dimensionality reduction for n samples and p features, the time complexity of PCA is $O(p^2n + p^3)$. The optimal time complexity of ISOMAP algorithm is $O(n^3)$ [33]. As for T-SNE, the solution must be approximated through gradient descend iterations, which is quite time consuming. In our experiment, the time consumption of the three methods is $\text{Time}_{\text{PCA}} < \text{Time}_{\text{ISOMAP}} < \text{Time}_{\text{T-SNE}}$. In general, only the PCA algorithm can meet the time requirement of interactive learning in practice.

4.4 Performance of PVIL on Real-World Dataset

In this experiment a real-world handwritten dataset (MNIST) is adopted to evaluate the effectiveness of PVIL. The MNIST dataset is made up of 60,000 training instances and 10,000 test instances. Each instance is a 0 to 9 handwritten digit image, which is represented by a $28 \times 28$ matrix. In our experiment, the matrix is transformed to a 784-dimensional vector to represent the original grayscale image.

4.4.1 Labeling Performance Comparison between LP and PVIL on Real-World Dataset

To evaluate the labeling performance of LP and PVIL, the unlabeled rate is set as $R_{\text{unlabeled}}$ of 99.5% in our experiment. In the experiment, the labeling results of 20 participants using PVIL are recorded. For fair comparison, the LP algorithm is executed 20 times with random initialization, the result of which in each time is also recorded.

|                | $R_{\text{unlabeled}}$ | Acc$_{f1}$ (worst) | Acc$_{f1}$ (average) | Time(s) |
|----------------|------------------------|--------------------|----------------------|--------|
| LP             | 99.5%                  | 0.0995             | 0.0999               | 1776.1 |
| PVIL           | 99.5%                  | 0.9504             | 0.9654               | 494.7  |

The semantic gap is one of the hard-to-solve problems in computer vision. It has been explained in subsection 2.2 that the perception ability of computers is weaker than that of humans for closed or warped data in visualization. The advantage of human perception is particularly evident in the labeling tasks of real-world dataset. As is shown in Table 3 the labeling accuracy of LP for 99.5% $R_{\text{unlabeled}}$ data is about 10%, which is far from the request of labeling tasks. While PVIL method can better tolerate overlapping and bending information through interaction, which achieves an accuracy of 96.54% on average for the 99.5% $R_{\text{unlabeled}}$ data.

In terms of computational efficiency, the measurement of relationships between samples is isolated with the LP algorithm. To obtain the weight map, the measurement of the similarity between every two samples is calculated by a kernel function, which results in high computation cost. As for the MNIST dataset, LP takes an average of 1776.1s in time consumption, which is unacceptable in most real-world applications. Note that, the human perception for a large number of samples is holistic. The process of human perception can skip the specific neighborhood similarity calculations, and greatly reduce the time of constructing a weight map. However, it takes some time for user to interactive with computers. In our experiments, 20 participants spent an average of 494.1s to complete a group of labeling. It is obvious that PVIL is superior to LP in efficiency, which ensures the real-world applications of PVIL.

4.4.2 Classification Performance Comparison between PVIL and Ground-truth Labels

![Fig. 11: Classification Performance of ground-truth labels (a) and PVIL labels (b). Blue points are True Positives and red ones are False Negatives.](image)

To further evaluate the performance of the labeling results of PVIL in real-world application, a classification experiment is conducted. Several mainstream classifiers, including logistic regression (LR) [62], gradient boosted decision trees (GBDT) [63], random forest (RF) [64] and convolutional neural network (CNN) [65], are adopted in our experiment. The classifiers are trained by the labels obtained by PVIL algorithm and ground-truth labels respectively.

The process of PVIL in acquiring labels can combine computer’s sensitivity of detailed features and human’s overall understanding ability of a task. Hence, the labeling results of PVIL could reveal the actual labeling distribution better. The experiment results (Table 3) show that PVIL labels achieve significant classification results on real-world dataset. As is listed in Table 4 under the premise of 96.5% label accuracy, the highest classification accuracy of PVIL could reach 97.5% (CNN), which is approaching 6. LeNet is chosen as the convolutional network model rather than deeper one such as VGG16 [25] or ResNet [26] for the reason that the the task complexity is relatively low.

7. It is difficult to continue the classification experiment by LP labels since the Acc$_{f1}$ of the LP labels in subsection 4.4.1 is extremely low (<10%).
the performance with ground-truth labels. The advantage of PVIL mainly lies in the perception of the backbone manifold, which can effectively ensure the high labeling accuracy.

As for the shortcomings, the error of PVIL mainly comes from the lack of accurate measurement ability. Isolated samples, located in the boundary area of a manifold, is difficult to distinguish even for humans. As is shown in Fig. 11, most errors of the PVIL labels are distributed in the boundary area. However, It does not appear a significant loss of classification accuracy in our experiment for the reason that the isolated samples are relatively sparse.

5 CONCLUSION

The main purpose of this paper is to re-summarize the related technologies and applications of interactive learning from the perspective of VII. We thus propose a framework called PVIL that structures an interaction strategy based on Gestalt principle. PVIL takes advantage of the idea of multi-dimensionality reduction to further boost the performance of visualization. The experiments validate that the PVIL framework significantly outperforms the labeling propagation for most visualization tasks. Furthermore, PVIL achieves promising classification results on real-world datasets.

In the future, we would attempt to develop a perceptual-cognitive framework with diversification for the aim of extending current PVIL to more learning scenes such as weak supervised learning and transfer learning.

ACKNOWLEDGEMENTS

This study was funded by National Natural Science Foundation of People’s Republic of China (61672130, 61602082, 91648205), the National Key Scientific Instrument and Equipment Development Project (No. 61627808), the Development of Science and Technology of Guangdong Province Special Fund Project Grants (No. 2016B090910001). All the authors declare that they have no conflict of interest.

REFERENCES

[1] M. C. F. D. Oliveira and H. Levkowitz, “From visual data exploration to visual data mining: A survey,” IEEE Transactions on Visualization & Computer Graphics, vol. 9, no. 3, pp. 378–394, 2003.
[2] D. A. Keim, Information Visualization and Visual Data Mining. IEEE Educational Activities Department, 2002.
[3] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, “Challenges in visual data analysis,” in Tenth International Conference on Information Visualization, 2006, pp. 9–16.
[4] B. Demir, C. Persello, and L. Bruzzone, “Batch-mode active-learning methods for the interactive classification of remote sensing images,” IEEE Transactions on Geoscience & Remote Sensing, vol. 49, no. 3, pp. 1014–1031, 2011.
[5] J. S. Hare, P. H. Lewis, P. G. B. Enser, and C. J. Sandom, “Mind the gap: another look at the problem of the semantic gap in image retrieval,” in Multimedia Content Analysis, Management, and Retrieval, 2006.
[6] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 39, no. 4, pp. 640–651, 2014.
[7] G. Salton and C. Buckley, Improving Retrieval Performance by Relevance Feedback. Cornell University, 1988.
[8] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum, “Lazy snapping,” ACM Transactions on Graphics, vol. 23, no. 3, pp. 303–308, 2004.
[9] D. F. Fisher, R. A. Monty, and J. W. Senders, “Eye movements: cognition and visual perception,” American Journal of Psychology, vol. 95, no. 3, p. 515, 2017.
[10] M. Brett, J. Anton, R. Valabregue, and J. Poline, “Region of interest analysis using an spm toolbox,” Neuroimage, vol. 16, no. 2, pp. 210–217, 2002.
[11] A. Mountzidou, T. Mirzissidis, E. Apostolidis, F. Markatopoulou, A. Ioannidou, I. Gialampoukidis, K. Averinakis, S. Vrochidis, V. Mezaris, and I. Kompatsiaris, “Verge: A multimodal interactive search engine for video browsing and retrieval,” in Int. Conf. on Multimedia Modeling, 2016.
[12] P. Awasthi, M. F. Balcan, and K. Voevodski, “Local algorithms for interactive clustering,” in International Conference on Machine Learning, 2014, pp. 550–558.
[13] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 35, no. 8, pp. 1915–1929, 2013.
[14] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning.” Nature, vol. 521, no. 7553, p. 436, 2015.
[15] O. Chapelle, P. Haffner, and V. N. Vapnik, “Support vector machines for histogram-based image classification.” IEEE Trans Neural Netw, vol. 10, no. 5, pp. 1055–1064, 1999.
[16] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, “Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features,” Scientific Data, vol. 4, no. 170117, p. 170117, 2017.
[17] K. Brinker, On Active Learning in Multi-label Classification. Springer Berlin Heidelberg, 2006.
[18] M. Umanol, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, and J. Kinoshita, “Fuzzy decision trees by fuzzy if3 algorithm and its application to diagnosis systems,” in IEEE International Fuzzy Systems Conference, 1994, pp. 2113–2118 vol.3.
[19] Y. Sahin and E. Duman, “Detecting credit card fraud by decision trees and support vector machines,” Lecture Notes in Engineering & Computer Science, vol. 2188, no. 1, 2011.
[20] H. Zhang, A. C. Berg, M. Maire, and J. Malik, “Svm-knn: Discriminative nearest neighbor classification for visual category recognition,” Proc. IEEE Conf. Computer Vision & Pattern Recognition, vol. 2, pp. 2126–2136, 2006.
[21] S. Liu, J. Wu, L. Feng, H. Qiao, Y. Liu, W. Luo, and W. Wang, “Perceptual uniform descriptor and ranking on manifold for image retrieval,” Information Sciences, vol. 424, 2017.
[22] L. Feng, S. L. Liu, Z. Y. Wu, and B. Jin, “Maximal similarity embedding,” Neurocomputing, vol. 99, no. 1, pp. 423–438, 2013.
[23] D. Husmeier, Random Vector Functional Link (RVFL) Networks. Springer London, 1999.
[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems, 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.
[25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Computer Science, 2014.
[26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[27] X. Zhu, “Semi-supervised learning with graphs,” in Advances in Neural Information Processing Systems, 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 23–32.
[28] F. Heimerl, S. Koch, H. Bosch, and T. Ertl, “Visual classifier training for text document retrieval,” IEEE Transactions on Visualization & Computer Graphics, vol. 18, no. 12, pp. 2839–2848, 2012.
[31] A. Sarkar, M. Spott, A. F. Blackwell, and M. Jamnik, “Visual discovery and model-driven explanation of time series patterns,” in Visual Languages and Human-Centric Computing, 2016, pp. 78–86.

[32] J. Bernard, E. Dobermann, A. Vgele, B. Krger, J. Kohlhammer, and D. Fellner, “Visual-interactive semi-supervised labeling of human motion capture data,” Electronic Imaging, vol. 2017, no. 1, pp. 34–45, 2017.

[33] M. Sedlmair and M. Aupetit, Data-driven Evaluation of Visual Quality Measures. The Eurographics Association & John Wiley & Sons, Ltd., 2015.

[34] J. Joseph, “Rocchio. 1971. relevance feedback in information retrieval,” The SMART retrieval system: experiments in automatic document processing, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[35] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, “Relevance feedback: a power tool for interactive content-based image retrieval,” Circuits & Systems for Video Technology IEEE Transactions on, vol. 8, no. 5, pp. 644–655, 1998.

[36] J. Bernard, D. Sessler, A. Bannach, T. May, and J. Kohlhammer, “A visual active learning system for the assessment of patient well-being in prostate cancer research,” in Proceedings of the 2015 Workshop on Visual Analytics in Healthcare. ACM, 2015, p. 1.

[37] J. Bernard, C. Ritter, D. Sessler, M. Zeppelzauer, J. Kohlhammer, and D. Fellner, “Visual-interactive similarity search for complex objects by example of soccer player analysis,” arXiv preprint arXiv:1703.03385, 2017.

[38] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair, “Comparing visual-interactive labeling with active learning: An experimental study,” IEEE Transactions on Visualization & Computer Graphics, vol. PP, no. 99, pp. 1–1, 2017.

[39] J. Bernard, M. Zeppelzauer, M. Sedlmair, and W. Aigner, “Vial: a unified process for visual interactive labeling,” Visual Computer, no. 2, pp. 1–19, 2017.

[40] D. G. Pelli, N. J. Majaj, N. Raizman, C. J. Christian, E. Kim, and M. C. Palomares, “Grouping in object recognition: The role of a gestalt law in letter identification,” Cognitive Neuropsychology, vol. 26, no. 1, pp. 36–49, 2009.

[41] L. Demany, B. Mckenzie, and E. Vurpillot, “Rhythm perception in early infancy,” Nature, vol. 266, no. 5604, pp. 718–9, 1977.

[42] K. Kolka, Principles of Gestalt psychology. HARCOURT, BRACE, 1935.

[43] D. J. Peterson and M. E. Berryhill, “The gestalt principle of similarity benefits visual working memory.” Psychonomic Bulletin & Review, vol. 20, no. 6, pp. 1282–1289, 2013.

[44] D. Todorovic, “Gestalt principles,” Scholarpedia, 2008.

[45] T. Joachims, “Making large-scale svm learning practical,” Technical Report, vol. 8, no. 3, pp. 499–526, 1998.

[46] B. A. Wandell, Foundations of Vision. Sinauer Associates, 1995.

[47] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazerbarclay, K. J. Antonellis, U. Scherf, and T. P. Speed, “Exploration, normalization, and summaries of high density oligonucleotide array probe level data.” Biostatistics, vol. 4, no. 2, pp. 249–64, 2003.

[48] I. Guyon, An introduction to variable and feature selection. JMLR.org, 2003.

[49] P. C. Ng and S. Henikoff, “Sift: Predicting amino acid changes that affect protein function.” Nucleic Acids Research, vol. 31, no. 13, pp. 3812–3814, 2003.

[50] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[51] A. Romero, C. Gatta, and G. Camps-Valls, “Unsupervised deep feature extraction for remote sensing image classification,” IEEE Transactions on Geoscience & Remote Sensing, vol. 54, no. 3, pp. 1349–1362, 2016.

[52] A. Mackiewicz and W. Ratajczak, “Principal components analysis (pca),” in Foundations of Statistical Learning, Springer, 2013, pp. 321–328.

[53] F. E. Harrell, “Ordinal logistic regression,” in Regression modeling strategies. Springer, 2015, pp. 311–325.

[54] M. Balasubramanian and E. L. Schwartz, “The isomap algorithm and topological stability,” Science, vol. 295, no. 5552, p. 7, 2002.

[55] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, 2000.

[56] Zhang, Zha, and Hongyuan, “Principal manifolds and nonlinear dimensionality reduction via tangent space alignment,” Journal of Shanghai University, vol. 8, no. 4, pp. 406–424, 2004.

[57] L. V. D. Maaten and G. Hinton, “Visualizing data using t-sne,” Journal of Machine Learning Research, vol. 9, no. 2605, pp. 2579–2605, 2008.

[58] A. Najafi, A. Joudaki, and E. Fatemizadeh, “Nonlinear dimensionality reduction via path-based isometric mapping.” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 38, no. 7, pp. 1452–1464, 2016.

[59] S. Liu, L. Feng, and H. Qiao, “Scatter balance: an angle-based supervised dimensionality reduction.” IEEE Transactions on Neural Networks & Learning Systems, vol. 26, no. 2, p. 277, 2015.

[60] D. Zhou, O. Bousquet, T. N. Lal, and J. Weston, “Learning with local and global consistency,” in International Conference on Neural Information Processing Systems, 2003, pp. 321–328.

[61] A. J. Izenman, “Linear discriminant analysis,” in Modern multivariate statistical techniques. Springer, 2013, pp. 237–280.

[62] A. Liaw, M. Wiener et al., “Classification and regression by randomforest.” R news, vol. 2, no. 3, pp. 18–22, 2002.

[63] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition.” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.