IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 3, 2019, accepted February 5, 2020, date of publication February 11, 2020, date of current version February 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2973190

Revisit Point Cloud Recognition From the Viewpoint of Data Mixture

ZHENZHONG KUANG, XIN ZHANG, AND ZHIQIANG GUO
Key Laboratory of Complex Systems Modeling and Simulation, School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding author: Zhenzhong Kuang (zzkuang@hdu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806063, Grant 61902101, Grant 61772161, Grant 61622205, and Grant 61602136, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR15F020002.

ABSTRACT In recent years, point cloud based data analysis has attracted lots of attentions of researchers from many different fields because of its simplicity and effectiveness. As a fundamental research, point cloud representation and recognition plays an important role. Although existing works achieved some good performances, they still cannot take full use of the information hidden in training data. This paper revisits the problem of point cloud representation and recognition from the viewpoint of data augmentation without incorporating additional data or more annotations. Different from existing works on 1D or 2D data, our proposed approach deals with a more complicated problem in three dimensional space for point cloud representation and recognition by mixing various training data from different object categories, which could help the classifier to better optimize the data-driven parameters. To validate the performance of our proposed approach, the popular used ModelNet40 dataset is employed as the standard benchmark. By carrying out comprehensive experiments under many different conditions, the experimental results show that our mixture method works positively towards improving the recognition performance of point cloud.

INDEX TERMS Point cloud, deep learning, point cloud mixture, feature mixture.

I. INTRODUCTION

With the development of three dimensional scanner devices, the amount of 3D models has increased a lot. Because 3D models contain much more (intuitive) information than text and image, they have attracted increasing attention of researchers and their application becomes widespread in various fields, such as robot, games and medicine. In recent years, the importance of point cloud has increased a lot because: (1) it is simple by using a collection of 3D points in Euclidean space without considering point relations; and (2) it can avoid the combinatorial irregularities and complexities of building meshes [1]–[3]. However, it is a challenging problem by directly working with such representation because there are limited information available [4], [5].

In literature, there are two ways for 3D shape representation according to the methods for modeling [3], [5]–[12]. One is the traditional hand-craft methods based on pre-defined models, while the other is the recent deep learning (DL) algorithm which is based on a data-driven modeling process. In the early years, most of the researchers pay attention to the model-based global and local shape descriptors, where the global descriptors only deals with the holistic shape description while the local keypoint descriptors could perform partial shape description and the bags-of-word (BOW) model is usually used with local descriptors for vector quantization [13], [14]. However, one significant defect of traditional methods is that they usually lack of flexibility for model selection and parameter determination (e.g. the BOW model is limited to the roughly pre-defined dictionary), which may lead to the low effectiveness for shape representation [13]–[15]. In contrast, the recent deep learning technique is more suitable for data-driven model learning by providing labeled training data. One typical example is the convolutional neural network (CNN) based deep network which has significantly boosted the performance of object recognition [16]–[20]. As a result, new 3D deep methods were proposed for object analysis and recognition (e.g. classification and segmentation), which can be put into two classes [5]: (a) new deep CNNs for end-to-end 3D data learning, such

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Wang.

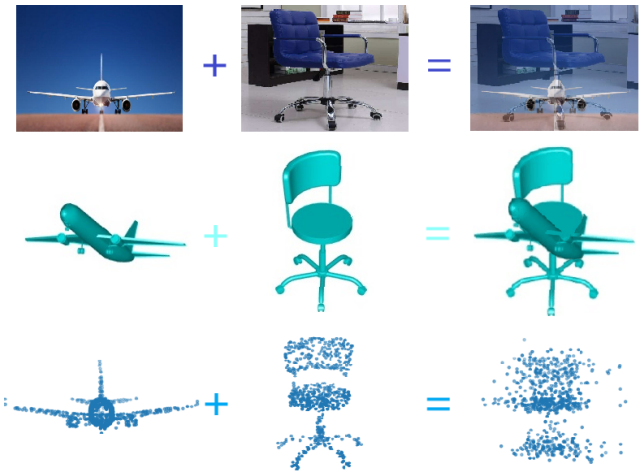


FIGURE 1. Demonstration of between class learning for data mixture. Top line: the mixture of two images. Middle line: the mixture of two 3D surface shapes. Bottom line: the mixture of two point cloud shapes.

as 3D volumetric CNN [12], [16], 3D graph CNN [8], [21] and PointNet [3]; (b) data transformation (e.g. multi-view images or feature vectors) based DL methods, such as the multiview CNN (MVCNN) [12], heat diffusion LSTM (HD-LSTM) [22] and deep geodesic moments (DeepGM) [18]. With the help of DL, the recent works have witnessed significant performance gain on shape recognition. But, there is still room to improve. For example, when building the 3D volumetric CNNs [3], [16], much complicated deep networks are used and the resolution of the constructed voxels is usually low for shape representation.

Data augmentation is widely applied for deep network training to improve the effectiveness of the resulting model, but traditional methods usually create new training samples from the original data [3], [23], [24], such as random cropping, image clipping, adding noise, data jittering and rotation. Recently, some pioneering works show that data augmentation by mixture of training examples can improve the learning performances [23], [25]–[27]. Although this kind of mixed data may not be realistic from the viewpoint of human perception, the computer algorithms can understand the data in some latent forms (e.g. waveform [26]) and can grasp more discriminative information for representational feature learning, such as the top line of Figure 1. The advantages of this approach lies in three points: (1) higher level of data variation; (2) better constrains on data distributions (e.g. gaussian); and (3) less dependency of CNN networks on large numbers of labeled training data. However, existing works only focus on dealing with the recognition problem of sound or image, but it is unclear: (1) if the data mixture method also works for 3D shape recognition; and (2) how to perform data mixture in 3D space.

The goal of this paper is to recognize 3D shapes in the format of point clouds by addressing two issues: (1) if existing data mixture methods (e.g. between class learning [26]) work for point cloud without any apparent structure information (e.g. the bottom line of Figure 1); and (2) how to mix inter-class point cloud data to obtain reliable network input for discriminative deep feature learning. In the experimental part, we validate our method on the ModelNet dataset and the results have revealed the superior performances of our approach on point cloud recognition. Then, we summarize the main contributions of the paper as follows:

- We first explore if the traditional between class learning method can improve the performance of point cloud recognition by directly mixing pairwise inter-class shapes.
- We develop a random point sampling based mixture method for data augmentation and perform ablation study by imposing different mixing settings.
- Beyond the raw point cloud, we further investigate the possibility of mixing the internal features of the deep networks for more discriminative feature learning;
- We perform experimental tests on ModelNet dataset to validate the effectiveness of the studied approaches for improving the performance of point cloud recognition.

II. RELATED WORK

Retrieving similar 3D objects from the same category as with the query has become an increasing important issue, where shape representation and retrieval is one of the most popular way for addressing this problem. The main difficulty of shape representation lies in the large inter-class similarity and the large intra-class variance, which makes it difficult for designing effective algorithms to grasp discriminative information [6], [28]. On the one hand, although some objects have the same function, like sitting, they have quite different shapes (e.g. bench vs. armchair). On the other, although some objects belong to different categories, they share similar shapes (e.g. microwave oven vs. square box) [12], [16]. Thus, how to support effective shape representation is quite challenging for machine learning by developing effective algorithms to capture object properties that could distinguish shapes from different categories. Over the past decades, many different trials were proposed to deal with the problem. At first, the traditional hand-craft features attracted more attention of researchers for 3D shape representation. But, recently, the deep learning techniques have further boosted the performance of 3D shape recognition in an end-to-end fashion so that researchers do not need to manually design models, and the shape representation problem becomes to designing an effective network and preparing sufficient number of raw training data. In this paper, we mainly focus on 3D point cloud which is a fundamental form of 3D shapes and which is also much easier to capture compared with the other data forms (e.g. 3D mesh) [3], [16]. Next, we present some of the most related works on point cloud recognition and some related works on data augmentation by using data mixture.

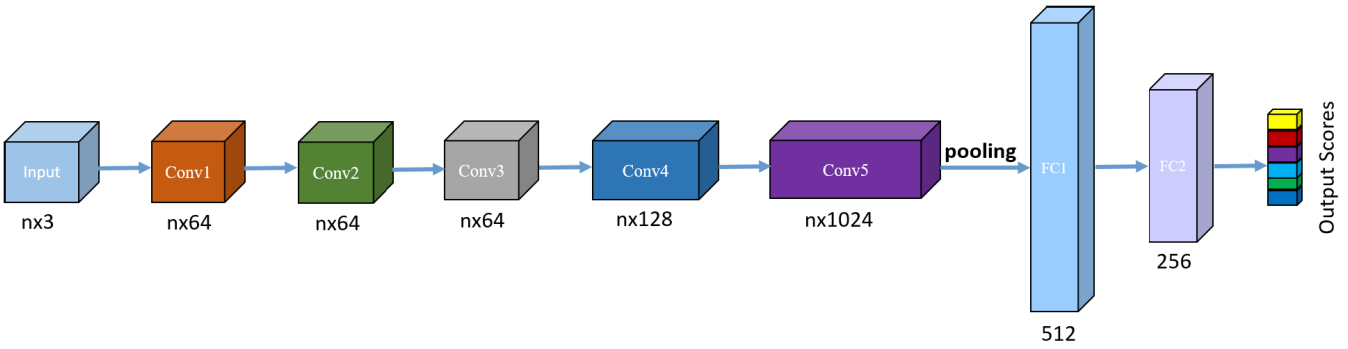


FIGURE 2. The basic configuration of PointNet [3] network which takes *n*-sized 3D point cloud as input, going through five convolutional (Conv) layers, one MAX pooling layer, two fully connected (FC) layers and one FC-based classifier layer.

A. TRADITIONAL POINT CLOUD DESCRIPTORS

As stated in previous works [1], [2], [4], each point cloud object only consists of a set of points, but it lack of structure information, which would make it more difficult for feature learning. In the early years, the traditional hand-craft methods are popular for point cloud recognition and there were limited number of works focusing on this topic. In [6], Funkhouser et al. presented a typical shape distributions method to describe shapes using the statistical histogram of pairwise point distances and it has already been proved to be a robust method for shape description [5], [9]. In [1], the diffusion distance [29] was employed for 3D point cloud recognition by borrowing the idea of heat diffusion for shape description. In [2], Limberger et al. organized a track to retrieve the point cloud toys and many different methods were proposed and compared to find a good solution for point cloud [8]. Although point cloud representation and recognition received some progress by using traditional methods, most of them cannot provide flexible models for obtaining discriminative features and are limited to empirical design for feature extraction.

B. DEEP LEARNING ON 3D SAPES

The development of deep learning has promoted the performance of 3D shape recognition by creating lots of new algorithms [5], [11], [19], [30]. In [16], a deep 3D volumetric approach was proposed by denoting a 3D object as a probability distribution of 3D voxel grid. In [30], the authors proposed a multi-view CNN approach for 3D shape classification and the result was superior to that of 3D volumetric approach. In [18], a DeepGM approach was proposed by studying geodesic moments (GM) and stacked sparse auto-encoder. In [22], a HD-LSTM method was proposed by integrating the heat distribution and LSTM for shape description. Although prior works have some achievements, they suffer from different drawbacks and most of the up-to-date deep learning algorithms focus on 3D meshes instead of point clouds [5]. Recently, Qi *et al.* [3] designed a point cloud based CNN network that receives an unordered set of 3D positions and labels, but it neglected the latent spatial relations between pairwise points. Similarly, a dynamic graph CNN approach was presented in [21] by incorporating the k-nearest neighbor elements to discover the spatial relations for learning features and classifiers.

Data Augmentation is widely used in the process of learning deep networks to deal with limited labeled training data and avoid overfitting by enlarging the training set. In [24], the AlexNet augmented the deep data by random sub-image cropping, horizontal flipping and changing the intensity of each channel, which could improve the robustness of the algorithm towards translation, reflection and illumination. In [3], the data sampling and jittering operations were employed to enhance the robustness of the learned model. One recent relevant work is to augment the training data by mixing pairwise data in a linear way [23], [25]–[27]. Tokozume *et al.* [27] first presented a between class (BC) learning method for sound recognition by mixing a pair of sounds with a random ratio as the input of deep network, and then they extended this approach to deal with image classification by regarding images as wave signals and proposed an improved version BC+ [26]. On this basis, there appeared some follow up works that achieved comparable results by discussing some related mixture strategies [23], [25]. The main advantage of BC learning is to constrain the feature distributions, which is difficult for standard learning methods to reach and which enables related methods to boost the performance of features. In this paper, we extend this approach for learning 3D point cloud features to achieve better shape recognition performance. Next, we present our approach and experimental results in different sections.

III. PRELIMINARY ON BETWEEN CLASS LEARNING

Between class (BC) learning was first studied for improving the performance of sound recognition and then was successfully extended to images. Let (x_1, l_1) and (x_2, l_2) be two (data, label) pairs, the main idea of BC learning is to blend one (data, label) pair by using

$$x = rx_1 + (1 - r)x_2 \tag{1}$$

and $l = rl_1 + (1 - r)l_2$, where l_1 and l_2 are one-hot category label vectors, and $r \in (0, 1))$ is the mixture ratio which is

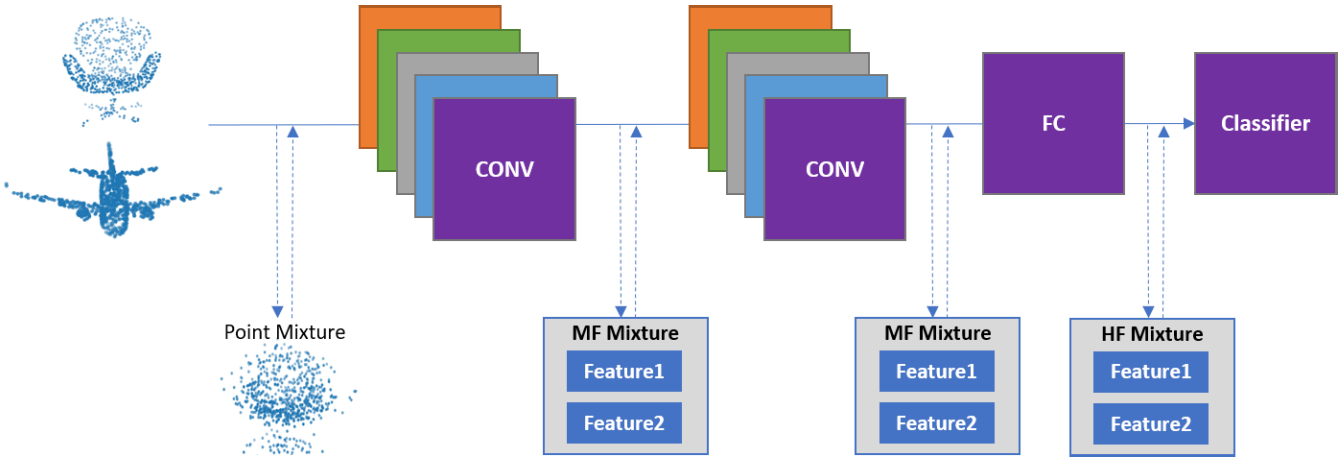


FIGURE 3. A brief flowchart of our approach for 3D point cloud recognition by mixing data at different layers of the employed deep neural network: point mixture by using the raw point coordinates, middle-level feature (MF) mixture at the intermediate layers, and high-level feature (HF) mixture at the last few linear layers.

usually determined by random. It has been shown in [27] and [26] that BC learning can regularize the learned feature distributions to be more compact.

BC plus (BC+) learning interprets any data (e.g. sounds&images) as waveform data. For any piece of sound, it has absolute center 0 and the distance from the center denotes the sound energy. But, for an image, it does not have absolute center and there is no concept of energy, which makes it hard to directly apply data mixture. To deal with this issue, each image is divided into two parts: static component and wave component. Before performing image mixture, the static component is first removed and then the mixture of pairwise data is formulated as

$$x = \frac{p(x_1 - \mu_1) + (1-p)(x_2 - \mu_2)}{\sqrt{p^1 + (1-p)^2}}, \tag{2}$$

$$p = \frac{1}{1 + \frac{\sigma_1}{\sigma_2} * \frac{1-r}{r}}, \tag{3}$$

where μ_1 and σ_1 denotes the mean value and variance of image x_1, respectively. The authors in [26] have verified the effectiveness of this approach for image classification.

IV. DEEP MIXTURE SHAPE REPRESENTATION

For the general 3D shapes, they are supposed to have no background and the point coordinates are used to describe each of the shape. Thus, each shape follows a certain point distribution that would vary for shapes from different categories. This property just follows the basic idea of BC learning by regarding each image as some waveform data. In this section, we wonder if data mixture still work for unstructured 3D point cloud data because blending point cloud data becomes much harder for humans to understand without any structure information. For example, the mixed shape using two 3D surface shapes in the middle line of Figure 1 is much easier to understand for humans compared with the point cloud mixture results in the last line. Let $S = \{X, y\}$ be the point cloud shape, $X = \{x_1, x_2, x_3, \cdots, x_n\}$ be the point set and $y \in C = \{c_1, c_2, c_3, \cdots, c_M\}$ be the atom category labels, the goal of this paper is to learn shape feature $f(S)$ and classifier $\phi(S)$ by optimizing the parameter set W of a well designed deep network $\mathbb{N}(S|W)$.

Different from 2D images, point cloud has the concept of centroid by averaging the coordinates of each shape. Thus, we do not need to remove the static component as with [26] did, and before apply data augmentation, we first normalize each shape by putting it into a unit ball centered at the origin (0, 0, 0). Given a random pair of inter-class point cloud shapes S_1 and S_2, it is very important to perform data mixture by addressing three critical issues: (1) whether data mixture is required for all the training data; (2) what kind of data (e.g. original point cloud or deep learned features extracted from some layer of the learned deep network $\mathbb{N}(S|W)$) to be used for data mixture; (3) how to mix two point cloud shapes (e.g. average of point coordinates or mixing of data points). As shown in Figure 3, we validate various ways of mixing inter-class data to improve the performance of deep learned features and classifiers. Next, we present the details of the adopted approach.

A. RANDOM MIXTURE

In BC learning, all the input samples are mixed in a pairwise manner. But, it may only produce limited performance gain without explicitly training the deep network $\mathbb{N}(S|W)$ using the atom category information. According to Figure 1, it is easy to recognize the original shape based on images and 3D meshes, but the case becomes difficult for 3D point cloud, which is hard to recognize the original shapes after mixing. To address this issue, we employ a random mixture approach

$$\hat{S} = \begin{cases} \theta(S_1, S_2), & if\ \sigma \geq th \\ S, & otherwise \end{cases} \tag{4}$$

by feeding both the mixed data $\theta(S_1, S_2)$ and the original data S according a random variable $\sigma \in [0, 1]$ to the deep network $\mathbb{N}(\hat{S}|W)$, where S_1 and S_2 are two randomly sampled inter-class point cloud shapes, $\theta(\cdot, \cdot)$ is the adopted mixture function and $th(= 0.5)$ is a threshold. On top of this modeling, the deep network can percept both the original data and the mixed data to regularize the feature distributions.

B. POINT-BASED DATA MIXTURE

Although 3D point cloud data has only one more dimension compared with 2D image, the data has confronted significant difference in complexity because 3D point cloud concentrates more on unordered points and 2D images can be seen as a projection along some viewpoint. In this part, we discuss how to mix the 3D data points of pairwise point cloud shapes.

1) WEIGHTED MIXTURE OF POINTS (WMP)

One simplest way to perform data augmentation is to follow Equation (4) for mixing two 3D point clouds by using the weighted sum of corresponding point coordinates. Similar to BC and BC+, we denote our method on point cloud as WMP and WMP+.

2) POINT SAMPLING BASED MIXTURE (PSM)

Instead of fusing corresponding points, we mix a pair of point clouds by sampling different number of points from each point cloud with a ratio $r \in [0, 1]$. First, we sample $s_1 = r \times n$ points from 3D point cloud X_1 and $s_2 = (1 - r) \times n$ points from 3D point cloud X_2 by using a sampling function $\gamma(\cdot)$. Then, the sampled s_1 and s_2 points are concatenated to obtain a new mixed training point cloud data

$$X_m = \gamma(X_1|s_1) \cup \gamma(X_2|s_2) \tag{5}$$

To obtain good mixture results, we propose to try three different methods to realize the point sampling function $\gamma(\cdot)$. The first method is to leverage a continuous PSM (**cPSM**) strategy, i.e. sampling s_1 continuous points from X_1 and s_2 continuous points from X_2, respectively. The second method is to leverage a random PSM (**rPSM**) strategy, i.e. sampling s_1 random points from X_1 and s_2 random points from X_2. The third method to leverage a crop-replace PSM (**crPSM**) strategy, i.e. cropping s_2 continuous points from X_1 and replacing them by using s_2 continuous points randomly picked up from X_2. Then, the labels of the mixed shape X_m is calculated by $l_m = rl_1 + (1 - r)l_2$.

3) WEIGHTED PSM (WPSM)

To integrate the advantages of point sampling and WMP, we propose a two step deep mixture method. First, we follow PSM method to generate a mixed point cloud X_m for X_1 and X_2. Then, we follow WMP to further mix X_m and X_1 to obtain a weighted mixture result X_w. In this way, we can further increase the variance of data mixture and keep the basic shape information. By adopting different sampling methods of PSM and mixture functions of WMP, we can obtain different WPSM results: **rWPSM** for using rPSM and WMP, **cWPSM** for using cPSM and WMP, **crWPSM** for using crPSM and WMP. When using WMP+, we can similarly obtain **rWPSM**+, **cWPSM**+ and **crWPSM**+.

C. FEATURE-BASED DATA MIXTURE

In this section, we study the feature-based data mixture. As shown in the Figure 3, we divide the mixing methods into three categories from the viewpoint of features: (1) low-level mixture of the original input point cloud data (see Section IV-B); (2) middle-level mixture of features by using the output of the middle layers of the adopted deep network (or the intermediate point features); and (3) high-level mixture of features by using the output of the last two layers of the adopted deep network (i.e. the global point cloud features). In Section IV-B, we have already discussed (1). Thus, we only need to discuss issue (2) and issue (3) next.

We rely on a pretrained deep network for mixing pairwise features. For network training, we only train the remaining network layers after the mixed feature layer. For example, if we mix the output features of Conv2 in Figure 2, then we only train all the layers after Conv2 by taking the mixed data as input. In all, we summarize the learning and testing process of feature-based data mixture as follows.

- **Step1.** Train a general deep network $\mathbb{N}(\hat{S}|W)$ by using the point cloud data.
- **Step2.** Extract the output of each network layer as features to be used for data mixture.
- **Step3.** Choose one intermediate layer Z for mixture of features and mix pairwise features by using the methods discussed in Section IV-B.
- **Step4.** Train the remaining deep network layers after Z.
- **Step5.** Test the resulting deep network by taking the non-mixed data as input for classification and feature extraction.

To realize our idea, the classical PointNet [3] is adopted as the base network $\mathbb{N}(\hat{S}|W)$ to verify our idea. According to Figure 2, we perform feature mixture only using the output of the convolutional layer (Conv) and the fully connected layer (FC): Conv1, Conv2, Conv3, Conv4, Conv5, FC1 and FC2, where data mixture at FC2 layer belongs to category (3), and data mixture at the other layers belong the category (2).

V. EXPERIMENT

In this section, we evaluate the performance of our studied method on the popular ModelNet dataset (i.e. the point cloud version) which contains 127, 915 3D shapes from 622 object categories, where two subsets ModelNet10 and ModelNet40 are popular used for algorithm evaluation. ModelNet10 contains 4, 899 3D shapes belonging to 10 categories, where 3, 991 shapes are used for training and 908 shapes are used for testing. ModelNet40 contains 12, 311 3D shapes belonging to 40 categories, where 9, 843 shapes are used for training and 2, 468 shapes are used for testing. More details about the datasets can be found in [3], [16], [30]. The network

TABLE 1. The ablation study results on random mixture.

Method	th	Dataset	Accuracy
PointNet(baseline)	–	ModelNet40	89.08
WMP	0.0	ModelNet40	88.59
WMP	0.5	ModelNet40	**89.52**
WMP+	0.0	ModelNet40	88.87
WMP+	0.5	ModelNet40	89.44

TABLE 2. Results of different 3D point cloud mixture strategies.

Method	Dataset	PointNet(vanilla)	PointNet
Baseline	ModelNet40	87.45	89.08
WMP	ModelNet40	88.39	89.52
WMP+	ModelNet40	87.98	89.44
cPSM	ModelNet40	88.35	89.61
rPSM	ModelNet40	88.10	89.61
crPSM	ModelNet40	87.90	89.40
cWPSM	ModelNet40	87.86	89.69
cWPSM+	ModelNet40	88.02	89.65
rWPSM	ModelNet40	87.78	89.20
rWPSM+	ModelNet40	**88.43**	89.32
crWPSM	ModelNet40	88.10	89.44
crWPSM+	ModelNet40	88.18	**89.77**

classification accuracy rate in percentage (%) is used as the basic evaluation measure.

Implementation Details: The classical PointNet [3] is adopted as the baseline method which has two implementation versions PointNet (vanilla) and PointNet, where PointNet (vanilla) does not use the transformation net compared with PointNet, The experimental platform is based on GTX 1080Ti GPU, Intel Xeon CPU 3.5GHZ HP Z440 with tensorflow framework under Ubuntu 16.04 System. Each of the adopted point cloud data is evenly sampled from the original shape with 1024 points and has been normalized to a unit sphere. Similar to PointNet, each point cloud is augmented by performing random rotation along the up-axis and adding zero mean gaussian noise with 0.02 standard deviation for data jittering.

A. ABLATION STUDY ON RANDOM MIXTURE

We carry out ablation study to verify whether random mixture works after applying data mixture, where PointNet is used as the baseline.

Our classification accuracy rates are listed in Table 1. Without using random mixture, the performance of WMP ($th = 0.0$) has decreased the baseline (i.e. 88.59% vs. 89.08%), while WMP ($th = 0.5$) has improved the baseline. The reason may lie in that, without using the original atom category for classification, WMP ($th = 0.0$) network may not know how to fit the mixed data without some reference. With the help of random mixture, PointNet can grasp both the atom category and the transition category (i.e. the mixed category) between pairwise point clouds. Thus, the random mixture operation is more likely to produce positive gains on classification, which is adopted for all follow up settings.

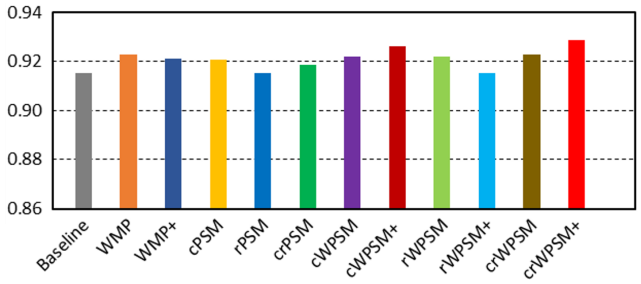


FIGURE 4. The classification performance on ModelNet10 dataset.

TABLE 3. The ModelNet10 classification results of feature-based data mixture at different layers of the adopted deep network.

Layers	PointNet(vanilla): 91.51				
	WMP	WMP+	cPSM	rPSM	crPSM
Conv1	93.30	**93.41**	92.29	92.18	92.52
Conv2	92.74	**93.08**	92.63	92.18	92.52
Conv3	92.74	93.08	91.96	**93.41**	92.07
Conv4	92.52	**92.85**	92.63	92.41	92.63
Conv5	92.22	92.00	92.44	92.66	**93.11**
FC1	92.55	92.44	92.11	92.44	**92.66**
FC2	91.77	91.66	91.77	92.00	**92.22**

B. RESULTS OF POINT-BASED DATA MIXTURE

In Table 2, we present the classification accuracy rates by using different mixture strategies on 3D point cloud.

According to the results, we can have several observations: (1) all the tested data mixture methods can improve the baseline performance; (2) WMP is superior to WMP+; (3) as for the low-level point cloud based data mixture, cPSM and rPSM are superior to crPSM; (4) as for the middle-level feature based data mixture, cWPSM+ performs better than cWPSM, which is the same for rWPSM+ and crWPSM+; (5) the performance improvement for all compared methods on PointNet(vanilla) is larger than on PointNet, which may be contributed to the employment of transformation operation and which reveals another advantage of the transformation layer of PointNet. The different WPSM+ results reveal that point-based data mixture outperform traditional coordinate based data mixture method (e.g. crWPSM+ vs. WMP+ on ModelNet40). The results also suggest that WPSM can further improve the performance of point-based data mixture.

Moreover, we analyze the classification results on ModelNet10 in Figure 4. According to the barchart results, it is easy to see that crWPSM+ has achieved the best performance. To intuitively demonstrate the performance of crWPSM+, we also plot the 2D t-SNE embedding [31] results of the extracted features (i.e. FC2 layer in Figure 2) in Figure 5. By comparing crWPSM+ with baseline (i.e. PointNet), the inter-cluster overlappings of crWPSM+ have reduced a lot, such as the pink cluster vs. the black cluster, which reveals that, after data mixture, the features of crWPSM+ becomes more separated.

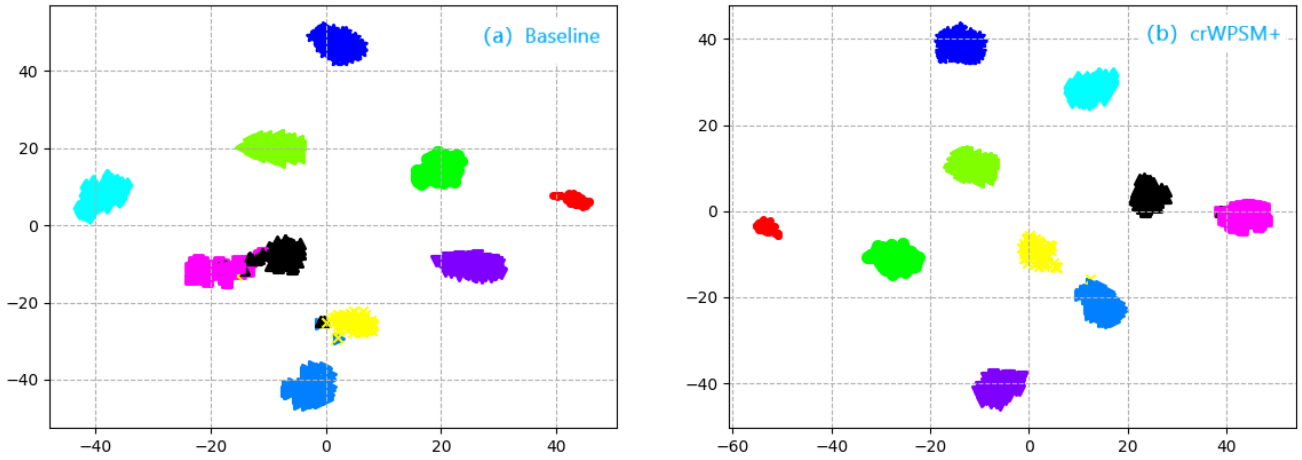


FIGURE 5. Comparison of 2D t-SNE embedding [31] results before and after data mixture for PointNet on ModelNet10: (a) baseline and (b) crWPSM+.

TABLE 4. The ModelNet40 classification results of feature-based data mixture at different layers of the adopted deep network.

Layers	PointNet(vanilla):87.45					PointNet: 89.08				
	WMP	WMP+	cPSM	rPSM	crPSM	WMP	WMP+	cPSM	rPSM	crPSM
Conv1	88.87	**89.32**	88.79	89.04	88.23	89.52	**89.81**	89.12	89.28	89.36
Conv2	**89.20**	88.83	88.59	88.40	88.27	89.28	89.04	**89.69**	89.16	89.20
Conv3	88.96	88.92	88.79	**89.20**	88.27	**89.85**	89.81	89.44	89.16	88.92
Conv4	89.16	**89.32**	88.92	89.04	88.55	89.16	89.40	89.69	89.48	**89.73**
Conv5	89.28	**89.32**	89.00	89.12	88.75	**89.80**	89.40	89.73	89.28	89.48
FC1	**89.24**	89.04	88.83	88.63	88.83	89.40	89.04	88.59	89.06	**89.44**
FC2	87.70	**87.78**	87.21	87.21	87.29	88.71	88.75	88.92	88.71	**89.04**

C. RESULTS OF FEATURE-BASED DATA MIXTURE

In this section, we discuss the performance of feature-based data mixture. We test data mixture at intermedia layer (i.e. middle-level feature or high-level feature) Z of the deep network shown in Figure 2 from Conv1 layer to FC2 layer based on the pretrained PointNet(vanilla) or PointNet.

We first present our initial results on ModelNet10 in Table 3 based on the pretrained PointNet (vanilla) network. According to each line of the table, we find that: (1) WMP+ performs better than WMP in most cases; (2) the overall performances of WMP and WMP+ outperform cPSM, rPSM and crPSM; (3) crPSM has achieved the best performance on Conv5, FC1 and FC2 layers; and (4) all the data mixture results have improved the baseline accuracy rate (i.e. 91.51%).

In Table 4, we present our results on ModelNet40 dataset by using both PointNet(vanilla) and PointNet as baselines, respectively. According to each line of this table, we observe that: (1) with PointNet(vanilla), WMP+ outperforms WMP in most cases; (2) with PointNet(vanilla), the overall performances of WMP and WMP+ outperform cPSM, rPSM and crPSM; (3) with PointNet, WMP+, WMP, cPSM, rPSM and crPSM have comparable performances; (4) all the results can improve the corresponding baselines in most cases from Conv1 to Conv5 layers, but it is not the case for FC1 and FC2 layers, especially for PointNet; (5) the performance lift is higher for PointNet(vanilla) than that of PointNet; and (6) the results in Table 4 is in consistent with Table 3.

The above results reveals that we can improve the performance of PointNet(vanilla) and PointNet by mixing data at convolutional layers (e.g. Conv1, Conv2, Conv3, Conv4, Conv5 as middle-level features), but it may become harder for the high-level features (e.g. FC1 and FC2). The reason may lie in that, at the final FC layers (or the high-level global features), there does not exist enough nonlinear space for representative feature learning. Although the best results of the feature-based data mixture (i.e. WMP: 89.85%) is slightly superior to the best point-based data mixture method (i.e. crWPSM+: 89.77%) on ModelNet40, the feature-based method requires to pretrain a basic network in ahead of using the mixed data for training, which is more complex than the point-based mixture method.

D. DISCUSSION

According to the experimental results, one can see that data mixture works positively for 3D point cloud data, because that we can improve the recognition performance by mixing pairwise point-based data and pairwise feature-based data. The point-based data mixture is easier to realize compared with the feature-based data mixture. Thus, we suggest to use the point-based data mixture for 3D point cloud. We also find that it would be easier to improve the performance of

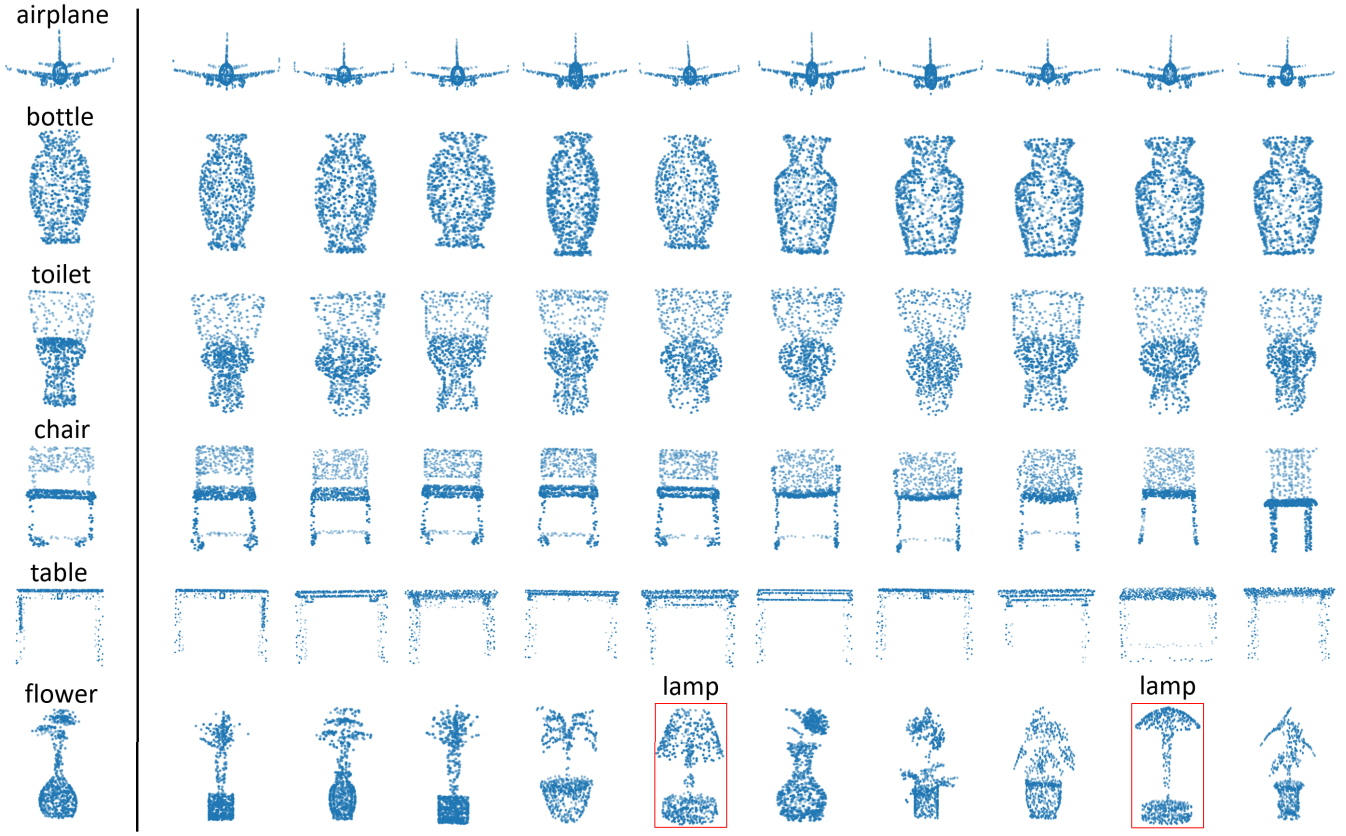


FIGURE 6. Representative examples of 3D point cloud retrieval on ModelNet40 dataset. The first column shows the query input point cloud shapes, and the corresponding shapes on the right of each line show the top10 retrieved results. The shapes in red boxes are error results.

simple networks by comparing PointNet(vanilla) and PointNet, which means that the extra networks (e.g. transformer network) adopted in PointNet may have the function of data regularization and data mixture may produce limited effects.

Further, we extract PointNet feature (FC2 layer) depending on point-based data mixture to perform 3D point cloud retrieval on ModelNet40. In Figure 6, we plot some representative retrieval results and it is easy to see that our results are very promising by automatically picking up intra-class shapes from the dataset. Although there are some error results (see the last line of Figure 6), one can observe that the shapes of the error results are quite similar to the query input, which can be contributed to the inter-class similarity of point cloud without structured information.

VI. CONCLUSION

This paper has revisited 3D point cloud recognition from the viewpoint of data mixture. In order to verify whether data mixture works for 3D point cloud, we have designed two distinct ways of performing data mixture for training deep network. We first present point-based methods for data mixture and then discuss the feature-based data mixture methods. Our experimental results show that both kinds of mixture methods can help to improve the performance of 3D point cloud recognition, where point-based data mixture is more practical without additional operations. In conclusion, we have verified that data mixture-based method could improve the performance of shape recognition, but the performance gain may be restricted to the deep models or the mixture methods. In the future work, we would study more effective data mixture methods to improve the discriminative ability of shape features and classifiers for point cloud recognition.

REFERENCES

[1] M. Mahmoudi and G. Sapiro, "Three-dimensional point cloud recognition via distributions of geometric distances," *Graph. Models*, vol. 71, no. 1, pp. 22–31, Jan. 2009.

[2] F. A. Limberger, R. C. Wilson, and M. Aono, N. Audebert, A. Boulch, B. Bustos, A. Giachetti, A. Godil, B. L. Saux, B. Li, and Y. Lu, "Point-cloud shape retrieval of non-rigid toys: SHREC'17 track," in *Proc. Eurographics Workshop 3D Object Retr.*, 2017, pp. 75–84.

[3] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[4] J. Liang, R. Lai, T. Wai Wong, and H. Zhao, "Geometric understanding of point clouds using Laplace-Beltrami operator," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 214–221.

[5] Z. Kuang, J. Yu, S. Zhu, Z. Li, and J. Fan, "Effective 3-D shape retrieval by integrating traditional descriptors and pointwise convolution," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3164–3177, Dec. 2019.

[6] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 807–832, 2002.

[7] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. Symp. Geometry Process.*, 2003, pp. 156–164.

[8] D. Boscaini, J. Masci, S. Melzi, M. M. Bronstein, U. Castellani, and P. Vandergheynst, "Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks," *Comput. Graph. Forum*, vol. 34, no. 5, pp. 13–23, Aug. 2015.

[9] Z. Kuang, Z. Li, Q. Lv, T. Weiwei, and Y. Liu, "Modal function transformation for isometric 3D shape representation," *Comput. Graph.*, vol. 46, pp. 209–220, Feb. 2015.

[10] Z. Lian, J. Zhang, and S. Choi and, "SHREC'15 track: Non-rigid 3D shape retrieval," in *Proc. Eurographics Workshop 3D Object Retr.*, 2015, pp. 101–108.

[11] J. Xie, G. Dai, F. Zhu, E. K. Wong, and Y. Fang, "DeepShape: Deep-learned shape descriptor for 3d shape retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1335–1345, Jul. 2017.

[12] C. R. Qi, H. Su, M. Niebner, A. Dai, M. Yan, and L. J. Guibas, "Volumetric and multi-view CNNs for object classification on 3D data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5648–5656.

[13] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape Google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graph.*, vol. 30, no. 1, pp. 1–20, Jan. 2011.

[14] H. Laga, T. Schreck, and A. Ferreira and, "Bag of words and local spectral descriptor for 3D partial shape retrieval," in *Proc. Eurographics Workshop 3D Object Retr.*, 2011, pp. 41–48.

[15] Z. Kuang, Z. Li, X. Jiang, Y. Liu, and H. Li, "Retrieval of non-rigid 3D shapes from multiple aspects," *Comput.-Aided Des.*, vol. 58, pp. 13–23, Jan. 2015.

[16] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1912–1920.

[17] Z. Kuang, J. Yu, Z. Li, B. Zhang, and J. Fan, "Integrating multi-level deep learning and concept ontology for large-scale visual recognition," *Pattern Recognit.*, vol. 78, pp. 198–214, Jun. 2018.

[18] L. Luciano and A. Ben Hamza, "Deep learning with geodesic moments for 3D shape classification," *Pattern Recognit. Lett.*, vol. 105, pp. 182–190, Apr. 2018.

[19] Z. Kuang, J. Yu, J. Fan, and M. Tan, "Deep point convolutional approach for 3D model retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[20] Z. Wang, Z. Kuang, Z. Guo, S. Zhu, and M. Tan, "Isometric shape representation by integrating shape function maps and deep learning," *IEEE Access*, vol. 7, pp. 158503–158513, 2019.

[21] Y. Wang, Y. Sun, and Z. Liu, "Dynamic graph CNN for learning on point clouds," 2018, *arXiv:1801.07829*. [Online]. Available: https://arxiv.org/abs/1801.07829

[22] F. Zhu, J. Xie, and Y. Fang, "Heat Diffusion long-short term memory learning for 3D shape analysis," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 305–321.

[23] H. Inoue, "Data augmentation by pairing samples for images classification," 2018, *arXiv:1801.02929*. [Online]. Available: http://arxiv.org/abs/1801.02929

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[25] C. Summers and M. J. Dinneen, "Improved mixed-example data augmentation," 2018, *arXiv:1805.11272*. [Online]. Available: http://arxiv.org/abs/1805.11272

[26] Y. Tokozume, Y. Ushiku, and T. Harada, "Between-class learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 5486–5494.

[27] Y. Tokozume, Y. Ushiku, and T. Harada, "Learning from between-class examples for deep sound recognition," in *Proc. ICLR*, 2018, pp. 1–13.

[28] Z. Kuang, J. Yu, and S. Zhu, "Deformable point cloud recognition using intrinsic function and deep learning," in *Proc. Pacific-Rim Conf. Multimedia*, 2018, pp. 89–101.

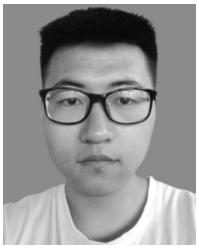
[29] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.

[30] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 945–953.

[31] L. Van Der Maaten and G. Hinton, "Visualizing non-metric similarities in multiple maps," *Mach. Learn.*, vol. 87, no. 1, pp. 33–55, Apr. 2012.

ZHENZHONG KUANG received the Ph.D. degree from the China University of Petroleum, Qingdao, China, in 2017. From 2015 to 2017, he was with The University of North Carolina at Charlotte, Charlotte, USA. He is currently with the School of Computer Science and Technology, Hangzhou Dianzi University. His research interests include visual recognition, privacy protection, multimedia analysis, and machine learning.

XIN ZHANG is currently pursuing the master's degree with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. His research interest is visual object analysis using deep learning methods.



ZHIQIANG GUO is currently pursuing the master's degree with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. His research interest is visual object analysis using deep learning methods.