Neurocomputing 174 (2016) 988–998

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An efficient and effective convolutional auto-encoder extreme learning machine network for 3d feature learning

Yueqing Wang [a,b,*], Zhige Xie [b], Kai Xu [b], Yong Dou [a,b], Yuanwu Lei [b]

[a] National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha, China
[b] College of Computer, National University of Defense Technology, Changsha, China

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form
5 October 2015
Accepted 11 October 2015
Communicated by G.-B. Huang
Available online 19 October 2015

Keywords:
Convolutional
Extreme learning machine
Auto-encoder
Feature learning

ABSTRACT

3D shape features play a crucial role in graphics applications, such as 3D shape matching, recognition, and retrieval. Various 3D shape descriptors have been developed over the last two decades; however, existing descriptors are handcrafted features that are labor-intensively designed and cannot extract discriminative information for a large set of data. In this paper, we propose a rapid 3D feature learning method, namely, a convolutional auto-encoder extreme learning machine (CAE-ELM) that combines the advantages of the convolutional neuron network, auto-encoder, and extreme learning machine (ELM). This method performs better and faster than other methods. In addition, we define a novel architecture based on CAE-ELM. The architecture accepts two types of 3D shape representation, namely, voxel data and signed distance field data (SDF), as inputs to extract the global and local features of 3D shapes. Voxel data describe structural information, whereas SDF data contain details on 3D shapes. Moreover, the proposed CAE-ELM can be used in practical graphics applications, such as 3D shape completion. Experiments show that the features extracted by CAE-ELM are superior to existing hand-crafted features and other deep learning methods or ELM models. Moreover, the classification accuracy of the proposed architecture is superior to that of other methods on ModelNet10 (91.4%) and ModelNet40 (84.35%). The training process also runs faster than existing deep learning methods by approximately two orders of magnitude.

1. Introduction

3D shape feature extraction is a vital issue covered in the high-level understanding of 3D shapes. Extensive efforts have been exerted to solve this important problem with the aid of recent advances on deep learning techniques. Existing feature extraction approaches based on deep learning can be broadly categorized as semi-automatic and fully-automatic methods.

In semi-automatic methods such as [1,2], researchers first extract several popular hand-crafted features from input 3D shapes and then utilize deep learning methods to combine these features further. This category of methods relies strongly on the adopted human-designed features. The extraction of these features consumes much time; hence, these methods cannot handle large-scale 3D datasets.

Numerous fully automatic deep learning methods have been proposed recently, such as convolutional deep belief network (CDBN) [3], auto-encoder (AE) [4], deep Boltzmann machines [5], convolutional neuron network (CNN) [6], and stacked local convolutional AE [7] approaches. These techniques are utilized to learn 3D features given the feature learning capability of these methods. In addition, these methods were first proposed for 2D image classification tasks.

3D shapes with reasonable resolutions have the same dimensions as high-resolution images. Thus, training deep networks on large-scale 3D datasets is time consuming. Furthermore, mastering this category of feature learning methods consumes time because of the black-box property of the deep learning method. Most of these deep learning methods convert 3D shapes into 2D representations for input [7–9]; thus, much of the 3D geometry information of 3D shapes is lost. Several works [3,10] attempt to apply 3D cubes, such as the volumetric representations of 3D shapes, as inputs. However, the training processes of these works are time consuming because of the additional dimension of input data. Therefore, the input resolution of these methods is limited.

To overcome the shortcomings of the existing methods, we propose a novel 3D shape feature extraction method called convolutional AE extreme learning machine (CAE-ELM) in this paper. This approach combines the advantages of CNN, AE, and extreme learning machine (ELM). AE is a typical unsupervised learning

* Corresponding author.
E-mail address: yqwang2013@163.com (Y. Wang).

http://dx.doi.org/10.1016/j.neucom.2015.10.035
0925-2312/

algorithm that can extract good features without supervised labels. However, the AE network is fully connected; thus, additional parameters must be learned. CNN restricts the connections between the hidden layer and the input layer through locally connected networks. Nevertheless, this network is an extensive computational method that is used with 3D shape datasets because of its convolutional operation. To reduce computational complexity, ELM [11] is often considered for its high efficiency and effectiveness.

Additionally, different input representations exert varied effects. For example, voxel data describe the structural information of 3D shapes because these data are expressed only as 0 and 1, which indicate that the voxel is outside and inside the mesh surface, respectively. Signed distance field (SDF) data are represented as a grid sampling of the minimum distance to the surface of an object that is represented as a polygonal model. The convention of applying negative and positive values within and outside the object, respectively, is frequently applied; thus, additional 3D shape details can be derived. To extract the global and local features collectively, we define a novel architecture that accepts both voxel and SDF as inputs. By combining these two types of data, our architecture can classify 3D shapes effectively.

The proposed CAE-ELM can also be used in practical graphics applications, such as in 3D shape completion. Optical acquisition devices often generate incomplete 3D shape data because of occlusion and unfavorable surface reflectance properties. These incomplete 3D shapes are challenging to repair; to fix incomplete data, we compare the features of broken and complete shapes before the CAE-ELM classifier as well as obtain the broken locations and values. Although the completion results are imperfect, CAE-ELM serves as a new approach to solve this problem.

The contributions of our approach are summarized as follows:

(1) *CAE-ELM*: We propose a new ELM-based designed network that performs well and learns quickly. To the best of our knowledge, our proposed model is the first to combine the advantages of CNN, AE, and ELM to learn the features of 3D shapes. This method has been used in practical graphics applications. We provide the source code[1] so that researchers can master it in a short time.
(2) *Increased classification accuracy*: The classification accuracy of the designed architecture is higher than that of other methods [10,12,13,9] on ModelNet10 (91.41%) and ModelNet40 (84.35%).
(3) *3D shape completion*: CAE-ELM can repair a broken 3D shape by using the features before the classifier.
(4) *Rapid 3D shape feature extraction*: Our method runs faster than existing deep learning methods by approximately two orders of magnitude, thus facilitating large-scale 3D shape analysis.

The experiment results show that the features learned by CAE-ELM significantly outperform hand-crafted features and other deep learning methods in terms of 3D shape classification. CAE-ELM can also repair the broken locations of 3D shapes with learned features for 3D shape completion. Furthermore, our method is efficient and easy-to-implement; thus, it is practical for real 3D applications.

2. Related work

2.1. 3D shape descriptors

3D shape descriptors play a crucial role in graphics applications such as 3D shape matching, recognition, and retrieval [14–17].

A variety of 3D shape descriptors have been developed during the last two decades [18,13,19,15]. Existing 3D descriptors are hand-crafted features which are labor-intensively designed and are unable to extract discriminative information from the data. Instead, we learn shape features from 3D shapes using automatically feature learning method.

2.2. 3D feature learning via deep learning

Researchers have successfully built deep models, such as convolutional neural network (CNN) [20], deep autoencoder networks [21], deep belief nets (DBN) [22] and extreme learning machine (ELM) [23] and etc., to automatically extract features with the superior discriminatory power for 2D image and shape representation in computer vision and machine learning [24]. A few very recent works attempt to learn 3D shape features via deep learning methods.

Zhang et al. [25] use ELM to determine an optimal fabrication in 3D printing considering a perceptual model. Wu et al. [3] use voxelization of 3D meshes as network's input and adopt 3D deep belief nets (DBN) [22] as their networks. Their work obtains good results on a subset of Princeton ModelNet [3]. However, their method is timing consuming and discards the pooling operations in CDBN, which makes their network fail to handle shape rotation invariance. As a result, they have to manually align all the input meshes into the same direction, in order to avoid uncertainty in extracting their features. Zhu et al. [4] use autoencoder to learn a 3D shape feature based on the depth images. However, they treat 2.5D depth images as traditional 2D images and can only get the global feature, which make their method have to combine with hand-crafted 2D image features (SIFT) to finish the 3D shape classification task. Xie et al. [9] propose Multi-View Deep Extreme Learning Machine (MVD-ELM) which adopts the multi-view depth image representation can achieve fast and quality projective feature learning for 3D shapes. However, as mentioned before, using 2.5D depth images as the input of network will lose 3D geometry and structure information of 3D shapes, and further influences the classification accuracy. Our method can handle large scale of 3D shapes with large rotation and geometry invariance through using voxel and SDF representations of 3D shapes. Moreover, due to the efficiency of ELM, our method runs faster than existing deep learning methods by approximately two orders of magnitude.

2.3. Extreme learning machines

Extreme learning machines (ELM) was proposed for generalized single-hidden layer feedforward neural network (SLFNs) [11,26,27] where the hidden layer need not be neuron alike. Unlike other neural networks with back propagation (BP) [28], the hidden nodes in ELM are randomly generated, as long as the activation functions of the neurons are nonlinear piecewise continuous. The weights between the hidden layer and the output layer have analytical solution and can be calculated using a formula. There are two phases in training process of ELM: feature mapping and output weights solving.

ELM feature mapping: Given input data $\mathbf{x} \in \mathbb{R}^D$, the output function of ELM for generalized SLFNs is

$$f(\mathbf{x}) = \sum_{i=1}^{L} \beta_i h_i(x) = \mathbf{h}(\mathbf{x})\beta, \tag{1}$$

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \cdots, h_L(\mathbf{x})]$ is the output vector of the hidden layer and $\beta = [\beta_1, \cdots, \beta_L]^T$ denotes the output weights between the hidden layer (L nodes) and the output layer (m nodes). The procedure of getting $\mathbf{h}$ is called ELM feature mapping which maps the input data from $\mathbb{R}^D$ to the feature space $\mathbb{R}^L$. In real applications, $\mathbf{h}$

[1] https://github.com/yqwang2006/CAE-ELM