2017 International Conference on Applied Mechanics and Mechanical Automation (AMMA 2017)
ISBN: 978-1-60595-471-4

3D Object Classification Based on Multi Convolutional Neural Networks

Mei-qi LU, Wei LI* and Ya-guang NING

State Key Laboratory of Virtual Reality Technology and System,
Beihang University, 100191, China

*Corresponding author

Keywords: 3D object classification, Convolutional neural network, Image classification.

Abstract. With the development of 3D sensing technology in recent years, making it possible to obtain high quality color image as well as its depth information, the combination of both can improve the accuracy of image classification. Traditional methods are mostly based on well-designed features separately for color image or depth image, which does not consider relation-features between color and depth information. A 3D object classification approach based on multi convolutional neural networks is presented in this paper. Our approach consists of three parts: (1) Two resizing methods for keeping image aspect ratio were proposed, which can keep the aspect ratio information of the original image, besides, data was augmented by applying both methods. (2) Different from traditional methods, CNN (Convolutional Neural Network) was used to extract features from both color image and depth image. (3) Consider on the relation-features about color and depth, we designed a multi CNN which focuses on relation-features and trained it successful, that is, we could extract relation-features using this network. Comparing with current approaches, our experiment results show better performance on 3D object classification.

Introduction

3D object classification is one of the hardest problems in the field of computer vision. It's the key point for making robots useful in outdoor environment. 3D sensing technology has got enough development in recent years, which could provide high quality depth images. The depth image provides useful extra information for classification task, since the depth information with color invariance and lighting invariance will provide geometrical cues. Most recent methods for 3D object classification are based on hand-designed features, such as Lai K[1] uses Spin Image feature and SIFT feature, Blum M[2] uses CKD feature. Researchers in the field of computer vision have designed many kinds of features about color, shape and geometry. Also there are some methods based on CNN could extract feature automatically, but some of them such as Socher R[3] based on the unsupervised method have to learn the convolution filters in advance.

In this paper, we introduce a method based on multi CNN for 3D object classification task that can learn features from the raw RGB-D image directly. Thus, we don't need to perform segmentation or design features by hand. We use the supervised method to learn color feature, depth feature and relation-feature. All the features bring improvements for 3D object classification. Details of the related works are discussed in the following sections.

Algorithm

Data Resizing and Augmentation

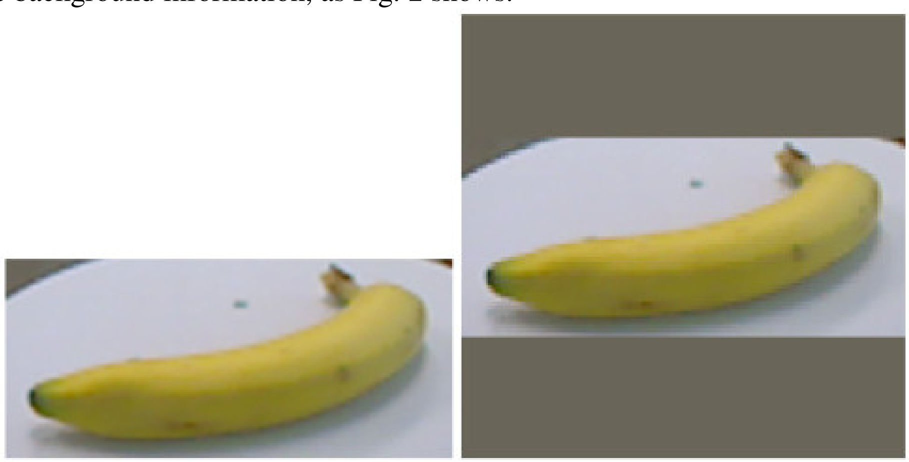
A fixed size image is required as input image in CNN. So we must resize the source image to the size required. Traditional image normalization method based on stretch has some disadvantages, the most striking shortcoming is the destruction of the aspect ratio, as to the aspect ratio of raw image is far different from the target images, the destruction will affect the performance of classification. So we provide two approaches to avoid this problem.

For the image with a large difference between length and width, we get multiple copies from the source image and combin them to one image that has similar length and width, after that we resize the combined image to the target size which CNN needed. This method reduces the influence on aspect ratio when resizing image to the target size, as Fig. 1 shows.

(left) original image (right) target image

Figure 1. Aspect ratio keeping image resize method (one).

Another approach is based on depth information, the values of pixels in the depth image present the distance between object and camera. As we know, the point located farthest from the camera must belong to the background, so we can get the color information about background throught the corresponing point in the RGB image and fill the image to a square one with background information. We try to get 30 points farthest from the camera, then calculate the average to represent the background information, as Fig. 2 shows.



(left) original image (right) target image

Figure 2. Aspect ratio keeping image resize method (two).

We use the two methods mentioned above to generate the target image, so we not only keep the aspect ratio, but also achieved the benefits from data augmentation.

Color Feature Extracting

Alex[4] has trained a CNN in the ImageNet Dataset[5], this model is used to classify 1000 categories. In order to use the features learned from ImageNet Dataset, Ali Sharif Rasavian has proven that using Alex's model as a feature extractor and SVM as a classifier is effective. In this paper, in order to obtain more effective features, we also finetune the parameters on our data.

We designe a model as shows in Fig. 3. In the training phrase, we set the front four layers's learning rate to 0, the followed three layers's learning rate to 1 and the last layers's learning rate to 5. Considered the last layer is initialized with random weights, a big learning rate should be set for accerlerating convergence. Through this way, we could inherit the low level features from Alex's model and finetune the high level features to adapt our task.

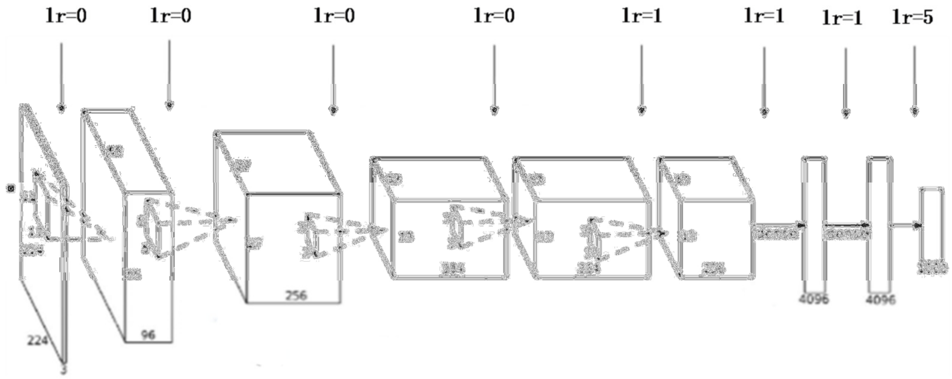


Figure 3. Convolutional neural network for color feature.

Depth Feature Extracting

There are two problems on depth feature. On the one hand, the depth image is so different with color image that we couldn't inherit features from Alex's model directly. On the other hand, the amount of the dataset is too small to train a complex CNN. To solve the above problems, we design a small CNN to learn depth feature, as Fig. 4 shows.

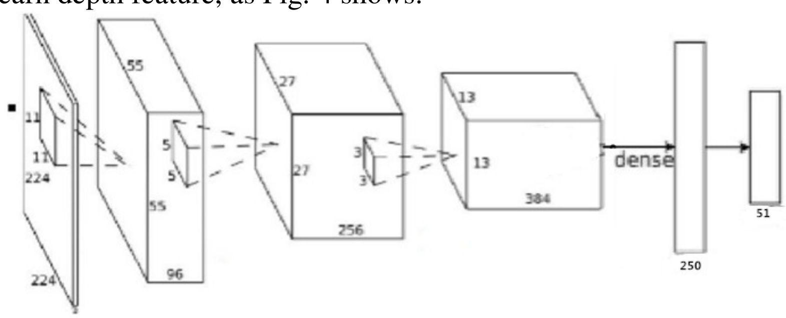


Figure 4. Convolutional neural network for depth feature.

Relation Feature Extracting

For classification task, adding an effective feature would be able to improve the accuracy. As we know, there are consistent one-to-one match between color image and depth image on every pixels. We try to learn the relation-features between color image and depth image, and design a model to find relation-features based on color and depth features. Similarity with the extraction of color features, the designed model inherit the color and depth features learned before, as Fig. 5 shows. In this training phrase, we don't adjust the parameters on the color and depth feature extraction model, color and depth features have been concated as the input to the last layer which is responsible for relation-feature learning.

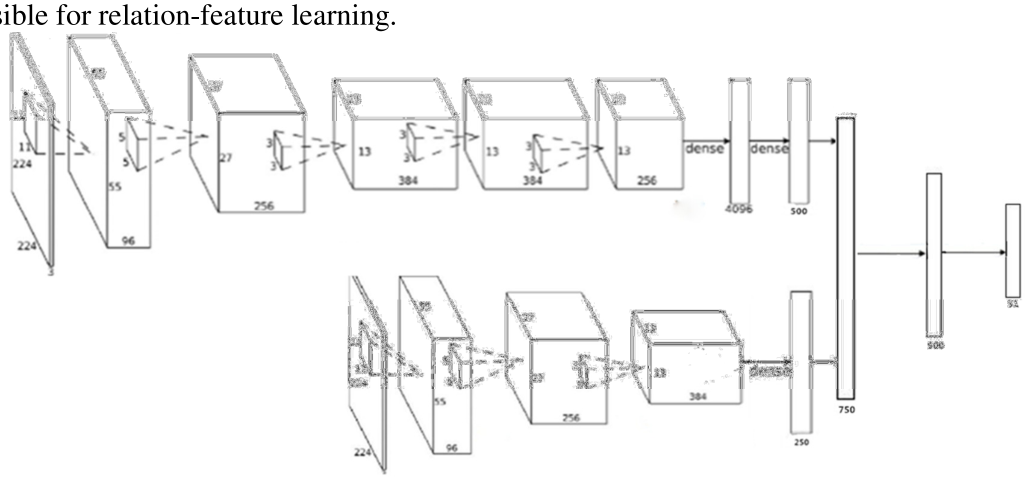


Figure 5. Convolutional neural network for relation feature.

Classifier Model

We achieve the color features, depth features and relation features through above work. At last, we combine all the three features for 3D object classification task. The model designed for classification is shown in Fig. 6. In this phrase, we also train the last layer in the model whose input is a union of color feature, depth feature and relation-feature and output is the class of the input image.

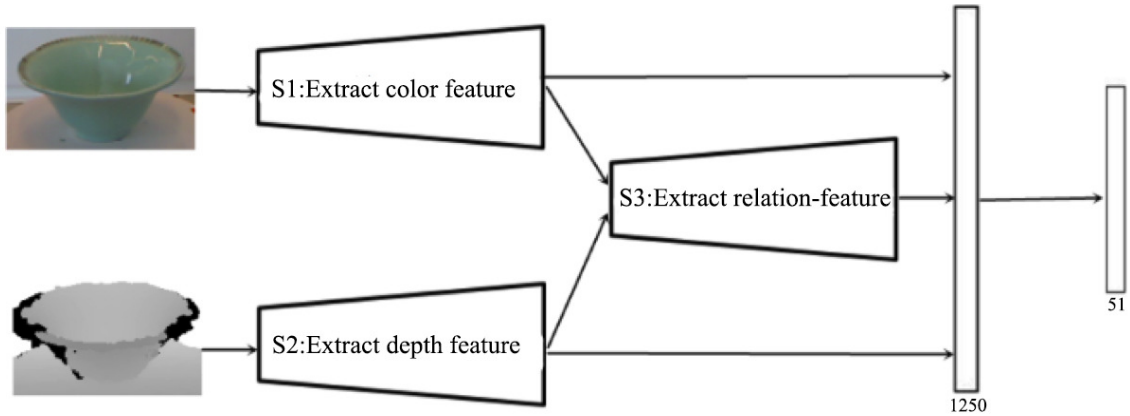


Figure 6. Convolutional neural network for 3D object classification.

Experiment

The dataset we used is the RGB-D Dataset built by Lai K, Bo L and Ren X[1], which contains 300 objects organized into 51 categories. The results of testing are shown in Table 1.

Table 1. Top1 accuracy compare with other popular algorithm.

Method	**Depth(%)**	**RGB(%)**	**Both(%)**
Linear SVM[1]	53.1 ± 1.7	74.3 ± 3.3	81.9 ± 2.8
Kernel SVM[1]	64.7 ± 2.2	74.5 ± 3.1	83.9 ± 3.5
Random Forest[1]	66.8 ± 2.5	74.7 ± 3.6	79.6 ± 4.0
SVM[6]	78.8 ± 2.7	77.7 ± 1.9	86.2 ± 2.1
CKM[2]	--	--	86.4 ± 2.3
SP+HMP[7]	81.2 ± 2.3	82.4 ± 3.1	87.5 ± 2.9
CNN-RNN[3]	78.9 ± 3.8	80.8 ± 4.2	86.8 ± 3.3
SSL[8]	77.7 ± 1.4	81.8 ± 1.9	87.2 ± 1.1
Subset-SAE-RNNs[9]	81.8 ± 2.6	82.8 ± 3.4	88.5 ± 3.1
Ours	**77.5 ± 1.3**	**84.2 ± 2.1**	**90.8 ± 1.8**

We can draw three conclusions from the experiment results. (1) The model is effective when classifier only use the color feature or depth feature, which proved the feature we learned from color image or depth image was available. (2) there is a higher accuracy rate for 3D object classification when all the features are used rather than only used color feature or depth feature, which proves the relation-feature between color and depth is effective. (3) Our method obtains a higher accuracy rate than other methods.

Conclusion

In this paper we propose an 3D object classification approach, an algorithm that could keep image aspect ratio when resizing, as well as a multi CNN for 3D object classification task. Compared with other popular methods, ours shows better performance. If more 3D object dataset are provided for our model, a better feature extraction model would be trained, especially for depth feature, and the accuracy would improve correspondingly.

References

[1] Lai K, Bo L, Ren X, et al. A large-scale hierarchical multi-view rgb-d object dataset[C]//Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011: 1817-1824.

[2] Blum M, Springenberg J T, Wülfing J, et al. A learned feature descriptor for object recognition in rgb-d data[C]//Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012: 1298-1303.

[3] Socher R, Huval B, Bath B, et al. Convolutional-recursive deep learning for 3d object classification[C]//Advances in Neural Information Processing Systems. 2012: 656-664.

[4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[5] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248-255.

[6] Bo L, Ren X, Fox D. Depth kernel descriptors for object recognition[C]//Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. IEEE, 2011: 821-826.

[7] Bo L, Ren X, Fox D. Unsupervised feature learning for RGB-D based object recognition[C]//Experimental Robotics. Springer International Publishing, 2013: 387-402.

[8] Cheng Y, Zhao X, Huang K, et al. Semi-supervised learning for rgb-d object recognition[C]//Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 2377-2382.

[9] Bai J, Wu Y, Zhang J, et al. Subset based deep learning for RGB-D object recognition[J]. Neurocomputing, 2015, 165: 280-292.