Skateboarding State Detection

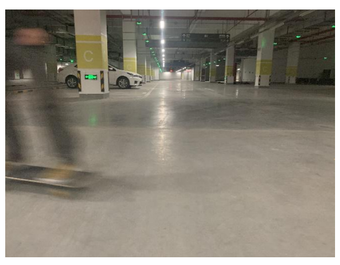
Shiwen Han, Jinze Wang, Yunfei Zhang, Duo Xu, Jiahe Song

ShanghaiTech University, Computer Vision I Project

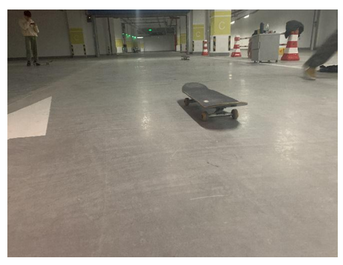
INTRODUCTION

Object detection and image classification are widely popular today. Use these two methods, we can realize skateboarding state detection. The details are as follows: given a photo of skateboarding, we can tell its state (static/sliding/overhead).

overhead



sliding



static

DATASETS

We use the toy dataset provided by TA. Moreover, we take some photos by ourselves with same background. Since the dataset is not enough for us to train and skateboard dataset is not available on Internet, we use FCN-ResNet model, whose dataset is PASCAL VOC dataset. It can detect the skateboards as birds or planes, which has no influence on our result as long as we can detect its position.

METHODS

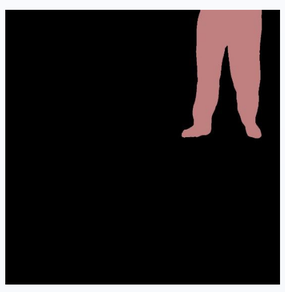
BOVW + SVM

First use K-means clustering, cluster hundreds of sifts into K-cluster to build a BOVW model. Then use this model to find the eigenvector of each image in training sets and testing set. Use SVM to train the eigenvector of training sets, get an SVM classifier. Then we can input eigenvector of testing sets to the SVM classifier to get the classification result of each testing image.

SEGMENT + BOVW + SVM

Combination of object detection and K-means:

First apply the object detection method to the images, so that we can remove some distracting factors like the background. Then apply method(1) to these images and predict the results.



CNN

We construct a CNN network with 2 layers of convolution and 3 layers fully connected layers and use ReLU in between. Inputs are image with RGB channels. We also resize image into 256x256 square and normalized them with common parameters.

SEGMENT + CNN

We combine image segmentation and CNN to classify this dataset.

First we segment both training and validation dataset and save them as RGB image array. Then put the array into same CNN network and get the results.

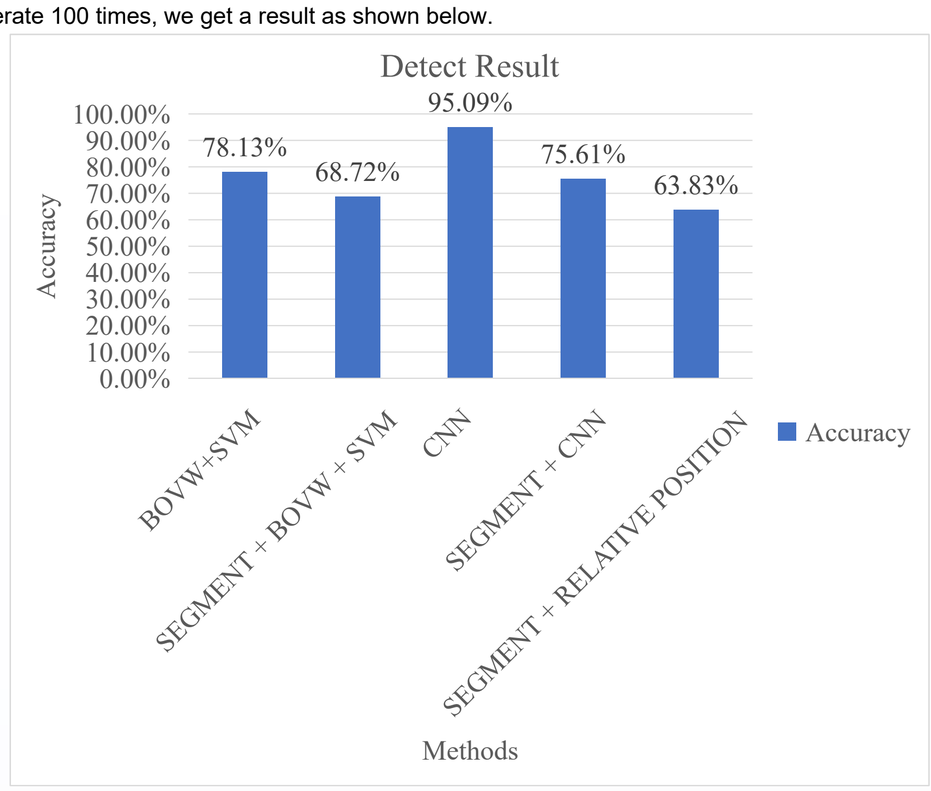
SEGMENT + RELATIVE POSITION

Given the preprocessed images obtained from segmentation, we iterate every pixels in this image and get the maximum and minimum x and y coordinates of the person and the skateboard respectively.

If (person_y_min - skateboard_y_max) is large enough, which means there is some distance between them, then we regard this image as 'overhead'. Otherwise, it is 'sliding'. Moreover, if one of (person_x_min-skateboard_x_max) or (skateboard_x_min- person_x_max) is large, which means the person stands beside the skateboard instead of standing on it, we classify it as 'static'.

RESULTS

After iterate 100 times, we get a result as shown below.



CONCLUSION

In this project, we mainly tried three tasks:

1. **Run pre-trained segment model FCN-ResNet101 on our dataset.**
2. **Made a simple CNN network for classification.**
3. **Use BOVW to extract feature into SVM classifier for classification.**

From the results, we can see simple CNN works quite well in this toy dataset, but segment before CNN works bad. We think CNN and BOVW extract many feature about background, since in this example, different class seems have different background and object is small in our image.

Low accuracy in classifier using segment results have these reasons:

1. **Many image is blurry in target object.**
2. **The pre-trained model is trained on a dataset without skateboard.**
3. **We did not extract skeleton feature about segmented object, simply put them into CNN will loss even more information.**

IMPROVEMENTS

- **The size of the given dataset is limited, so our model may be underfitted. We may take more photos to extend the dataset and thus get a better result.**
- **We use FCN-ResNet model based on PASCAL VOC dataset. However, this dataset doesn`t contain skateboards. With more time, we could label the images of skateboards and use these images to train our model.**

REFERENCE

Lowe, David G. "Object recognition from local scale-invariant features." Proceedings of the seventh IEEE international conference on computer vision. Vol. 2. Ieee, 1999.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

CODE LINK

For complete code, please go to our github repository.