Generating Large Scale Image Datasets from 3D CAD Models

Baochen Sun, Xingchao Peng, Kate Saenko
Computer Science Department
University of Massachusetts Lowell
Lowell, Massachusetts, US
{bsun, xpeng, saenko}@cs.uml.edu

Abstract

Datasets power computer vison research and drive breakthroughs. Larger and larger datasets are needed to better utilize the exponentially increasing computing power. However, datasets generation is both time consuming and expensive as human beings are required for image labelling. Human labelling cannot scale well. How can we generate larger image datasets easier and faster? In this paper, we provide a new approach for large scale datasets generation. We generate images from 3D object models directly. The large volume of freely available 3D CAD models and mature computer graphics techniques make generating large scale image datasets from 3D models very efficient. As little human effort involved in this process, it can scale very well. Rather than releasing a static dataset, we will also provide a software library for dataset generation so that the computer vision community can easily extend or modify the datasets accordingly.

1. Introduction

Datasets play an important role in computer vision research. From PASCAL VOC [3] to ImageNet [1], many breakthroughs were powered by them, i.e. Deformable Part Models [4], Deep Convolutional Neural Network [5], etc. Thanks to Moore's law, the computing power is increasing exponentially. As the models used in computer vison research are getting bigger and bigger (i.e. AlexNet [5] has 60 million parameters), larger and larger datasets are needed. However, datasets collection still requires a human labelling phase. This is both time cousuming and expensive and can not scale well. ImageNet is one of the largest datasets in the computer vision community, containing 14 million images (only 1 million images contain bounding box information). How can we go from 14 million to 100 million or even billions of images?

The other drawbacks of the current datasets are that they are static and users can not add new categories or make changes. For example, how can a user add a new category (i.e. books) to PASCAL VOC dataset? What if a user want to test an algorithm with a different lighting condition in the Office dataset [8]?

In this abstract, we provide a novel approach of generating large scale image datasets from 3D CAD models directly. More and more freely available 3D object models are generated; a simple search for "table" returns 45,443 results (queried on 04/29/2015) from Google 3D Warehouse. The computer graphics community spent decades working on generating photorealistic images and made many breakthroughs. These techniques have been successfully used in a lot of Hollywood movies, where it is hard to tell whether a scene is real or synthetic as illustrated in Figure 1. Thanks to these large volume 3D models and computer graphics techniques, we can easily generate a large amount of images in minutes or hours rather than months or years. Since the only human effort involved is 3D model selection and parameter settings (i.e. lighting, pose, texture, etc.) and a lot of image metadata (i.e. image label, object outline, etc.) can be generated automatically, it can scale very well. Rather than releasing a static dataset, we will also provide the code for datasets generation. This allows users to add new categories with same image statistics or generate new datasets with different parameters.

Virtual data has already been used in many applications [9] [7]. [9] showed that modern domain adaptation techniques can provide a way to bridge the gap between synthetic and real training data. In this abstract, we want to take further steps and provide large scale benchmarks for many object categories (up to thousands) and novel applications. This is an ongoing project and we plan to provide a website with download information before the workshop.

2. Synthetic Dataset Generation

As illustrated in Figure 2, our dataset generation process contains three stages: 3D model selection, parameter setting, and rendering and labelling.

3D Model Selection As mentioned in section 1, a large

Figure 1. Modern computer graphics techniques can be used to generate extremely photorealistic images. For example, here is a comparison of a real police car and a synthetically rendered one. It is very difficult to tell the real from the synthetic one. Ground truth: the one on the left side is real and the one on the right side is synthetic. (Best viewed in color)

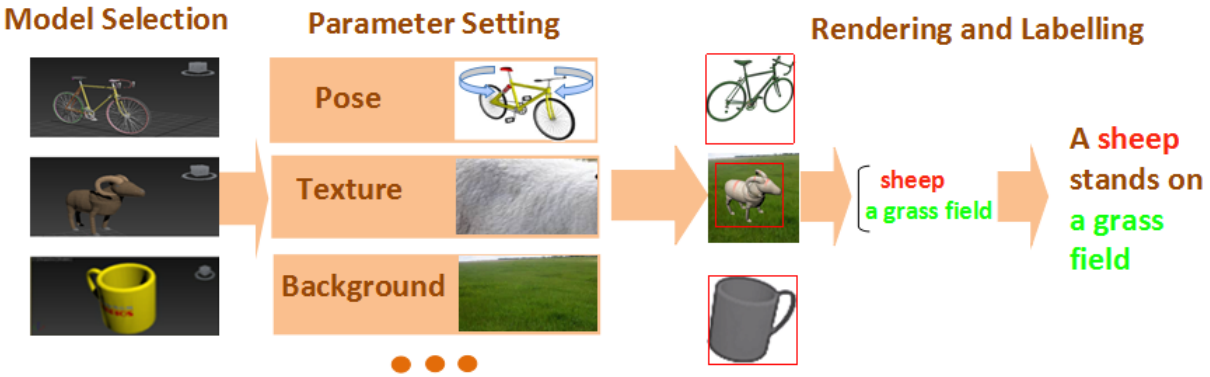

Figure 2. Overview of our dataset generation process. The first stage is 3D model selection. The second stage is parameter setting for the rendering process. The last stage is rendering and labelling. (Best viewed in color)

amount of 3D models can be downloaded from Google 3D Warehouse or other websites freely. We plan to use all the 3 million 3D models from ShapeNet first and extend it to cover all the 100,000 synsets of WordNet eventually. Before going to the next stage, we also clean the 3D models to remove noise.

Parameter Setting In order to get a more diverse and realistic dataset, we also need to set some parameters, i.e. lighting, pose, texture, background, shape deformation, etc. All these settings can be coded in a script (i.e. MAXScript of Autodesk 3ds Max) so there is no need to do it per image. For example, we could set the pose to be randomly rotated 5 to 10 degrees in each direction.

Rendering and Labelling We use Autodesk 3ds Max to render images, other softwares (i.e. Autodesk Maya) could be used as well. Image level labels, bounding box information, and other sematic labels can be generated automatically during the dataset generation process as all the information of an image are accurate and available. This process is the same as [9] and [7].

3. Applications

In this section, we give a brief overview of the possible applications of our datasets.

Image Classification [7] used virtual datasets for image classification. However, the number of categories is very small (20 for PASCAL VOC dataset and 31 for Office dataset). We believe that large scale virtual datasets can help researchers tackle more challenging problems. For example, with thousands of categories and millions of images, we could train a deep convolutional neural network (i.e. AlexNet) from scratch rather than just fine-tuning an existing one.

Object Detection With the accurate bounding box information, [9] showed that an object detector trained on a limited number (20 per category) of virtual images can achieve same performance as a classifier trained on much larger (150 to 2000 per category) ImageNet dataset.

Pose Estimation PASCAL 3D [11] used 3D models to estimate the pose of an object in a 2D image. However, it only contains 12 categories. With the help of large scale datasets containing accurate pose information, researchers can scale pose estimation to much larger exciting challenges.

Semantic Scene Labelling Since we know exactly "which objects are where", more structured semantic information can be generated and facilitate large scale semantic scene understanding and labelling.

Image to Text Recurrent neural networks have been successfully used in generating text descriptions for images or

videos [2] [10]. Our approach can generate more diverse images than those in current image captioning benchmarks such as COCO [6].

Domain Adaptation Since virtual dataset can be generated to have different visual characteristics than real image datasets, it is a natural fit for large scale domain adaptation experiments. The Office dataset, the current benchmark dataset for domain adaptation, only contains 31 categories in three domains with 8 to 90 images per category.

4. Conclusion

In this abstract, we provide a novel approach of generating large scale image datasets from 3D object models directly. As little human effort is involved in this process, it can scale very well. We will also provide a software library for dataset generation to facilitate user specific experiments.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 1

[2] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. *arXiv preprint arXiv:1411.4389*, 2014. 3

[3] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge 2012, 2012. 1

[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010. 1

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1

[6] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014*, pages 740–755. Springer, 2014. 3

[7] X. Peng, B. Sun, K. Ali, and K. Saenko. Exploring invariances in deep convolutional neural networks using synthetic images. *arXiv preprint arXiv:1412.7122*, 2014. 1, 2

[8] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010*, pages 213–226. Springer, 2010. 1

[9] B. Sun and K. Saenko. From virtual to reality: Fast adaptation of virtual object detectors to real domains. In *Britich Machine Vision Conference (BMVC)*, 2014. 1, 2

[10] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, and K. Saenko. Translating videos to natural language using deep recurrent neural networks. *arXiv preprint arXiv:1412.4729*, 2014. 3

[11] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pages 75–82. IEEE, 2014. 2