Use of Deep Learning to Create 3D BMW-10 Car Models from 2D Images

Lenin Kamma

(Class DSC680)

GitHub Portfolio URL: https://databinary.github.io/

Domain

Image transformation - Cars

Data

The data that is used for this project is extracted from the Stanford cars dataset. The cars dataset contains more than 500 images of BMW class 10 cars[1]. Each image contains one car taken at different locations at different angles.

Data Description

The data is placed in 11 different files and each file contains around 50 images.

URL

https://ai.stanford.edu/~jkrause/cars/car_dataset.html

References

The following references provide helpful information about 2D to 3D image conversion.

[1] Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.

In extensive experiments on existing and newly proposed datasets, this paper shows 3D object representations outperform their state-of-the-art 2D counterparts for fine-grained categorization and demonstrate their efficacy for estimating 3D geometry from images via ultra-wide baseline matching and 3D reconstruction.

[2]https://www.researchgate.net/publication/342199083_3D_Reconstruction_of_2D_Images_using_Deep_Learning_on_the_Nvidia_Jetson_Nano

This model is trained on the Shape-net dataset in which the training set includes 2D images with their respective 3D structures

[3] https://github.com/lkhphuc/pytorch-3d-point-cloud-generation

A Pytorch implementation of the paper: Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction

[4] https://github.com/chenhsuanlin/3D-point-cloud-generation

Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction

[5] https://arxiv.org/abs/1706.07036

This paper proposes a novel 3D generative modeling framework to efficiently generate object shapes in the form of dense point clouds.

[6] https://medium.com/vitalify-asia/create-3d-model-from-a-single-2d-image-in-pytorch-917aca00bb07

In this post author explores a recent attempt of extending deep learning to the Single image 3D reconstruction task, one of the most important and profound challenge in the field of 3D computer graphics.

[7] **http://ai2-website.s3.amazonaws.com/publications/Deep3D-2.pdf**

In this paper the authors proposed a fully automatic 2D-to-3D conversion algorithm based on deep convolutional neural networks.

[8] **https://newsroom.ucla.edu/releases/2d-images-converted-3d-deep-learning**

In this study published in Nature Methods, the scientists reported that their framework, called "Deep-Z," could take 2D images from one type of microscope and virtually create 3D images of the sample as if they were obtained by another.

[9] **https://arxiv.org/pdf/1906.06543.pdf**

This article provides a comprehensive survey of the recent developments in this field of image based 3D reconstruction. This paper focuses on the works which use deep learning techniques to estimate the 3D shape of generic objects either from a single or multiple RGB images.

[10] **https://venturebeat.com/2020/03/05/microsofts-ai-generates-3d-objects-from-2d-images/**

New preprint paper published by a team hailing from Microsoft Research detail a framework that they claim is the first "scalable" training technique for 3D models from 2D data. The paper says it can consistently learn to generate better shapes than existing models when trained with exclusively 2D images.

[11] **https://www.mdpi.com/2072-4292/12/11/1729/htm**

While deep learning on 3D point clouds has shown good performance on several tasks, including classification, parts, and semantic segmentation, there are still areas which require more attention. This paper contains a review of the recent state-of-the-art deep learning techniques, mainly focusing on raw point cloud data.

Research and Analysis

A lot of research is being done to convert 2D images into 3D using modern deep learning methods. Conversion into 3D images requires deep neural networks to recognize the depth in the images. As part of this project, I will use the convolutional learning technique (CNN) to carry out this process. First, Keras model will be used to convert each image into an RGB value. Each RGB model is then fed to encoder and decoder methods to create structures. Point cloud fusion or Voxel grids can be used to create 3D images from the RGB values. The following questions will be answered while developing the model.

- What kind of images can be converted?
- How to classify the images?
- Does the size of the image make any difference in conversion?
- What information is needed to convert 2D images into 3D images?
- How does the algorithm create depth using 2D images?
- What is the best model to create RGB values for each image?
- What deep learning methods can be used for conversion?
- What kind of loss is expected in the images?
- How to convert occlusions or cluttered images?
- What are the computational advantages of using deep learning for conversion?

Methods & Machine learning models.

Convolution Neural Networks (CNN)

Keras – Deep learning API running on top of Tensor Flow

Voxel Grids – Converts spatial pixels into volume grid pixels. Convolutional Neural Networks can be directly applied but it has high tradeoffs between details and resources.

Geometric forms – This method focuses on compact representation but it cannot directly apply CNN[6]

Point Cloud Fusion – Set of points defined in a 3D space. Point clouds have become one of the most significant data formats for 3D representation and are gaining increased popularity[11]

Challenges

- The effectiveness of different image conversion techniques like Voxel grids, Triangular mesh and Point Cloud
- Measuring the performance of each model
- Conversion of each 2D image into RGB values using Keras library

Conclusion

The goal of this project is to take 2D images from the BMW class-10 image dataset and convert them into 3D images using deep learning techniques. Although creating 3D images give opportunities to visualize realistic spatial arrangements, there are many challenges in recognition, classification, segmentation, detection, and conversion of 2D into 3D images.