

Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images

Nanyang Wang¹, Yinda Zhang², Zhuwen Li³,
Yanwei Fu⁴, Wei Liu⁵, Yu-Gang Jiang^{1†}

¹Shanghai Key Lab of Intelligent Information Processing,
School of Computer Science, Fudan University

²Princeton University ³Intel Labs ⁴School of Data Science, Fudan University ⁵Tencent AI Lab
nywang16@fudan.edu.cn yindaz@cs.princeton.edu lizhuwen@gmail.com
yanweifu@fudan.edu.cn wl2223@columbia.edu ygj@fudan.edu.cn

Abstract. We propose an end-to-end deep learning architecture that produces a 3D shape in triangular mesh from a single color image. Limited by the nature of deep neural network, previous methods usually represent a 3D shape in volume or point cloud, and it is non-trivial to convert them to the more ready-to-use mesh model. Unlike the existing methods, our network represents 3D mesh in a graph-based convolutional neural network and produces correct geometry by progressively deforming an ellipsoid, leveraging perceptual features extracted from the input image. We adopt a coarse-to-fine strategy to make the whole deformation procedure stable, and define various of mesh related losses to capture properties of different levels to guarantee visually appealing and physically accurate 3D geometry. Extensive experiments show that our method not only qualitatively produces mesh model with better details, but also achieves higher 3D shape estimation accuracy compared to the state-of-the-art.

Keywords: 3D shape generation · Graph convolutional neural network · Mesh reconstruction · Coarse-to-fine · End-to-end framework

1 Introduction

Inferring 3D shape from a single perspective is a fundamental human vision functionality but is extremely challenging for computer vision. Recently, great success has been achieved for 3d shape generation from a single color image using deep learning techniques [6, 9]. Taking advantage of convolutional layers on regular grids or multi-layer perception, the estimated 3D shape, as the output of the neural network, is represented as either a volume [6] or point cloud [9]. However, both representations lose important surface details, and is non-trivial to reconstruct a surface model (Fig. 1), i.e. a mesh, which is more desirable for many real applications since it is lightweight, capable of modelling shape details, easy to deform for animation, to name a few.

In this paper, we push along the direction of single image reconstruction, and propose an algorithm to extract a 3D triangular mesh from a single color image. Rather

indicates equal contributions.

[†] indicates corresponding author.

