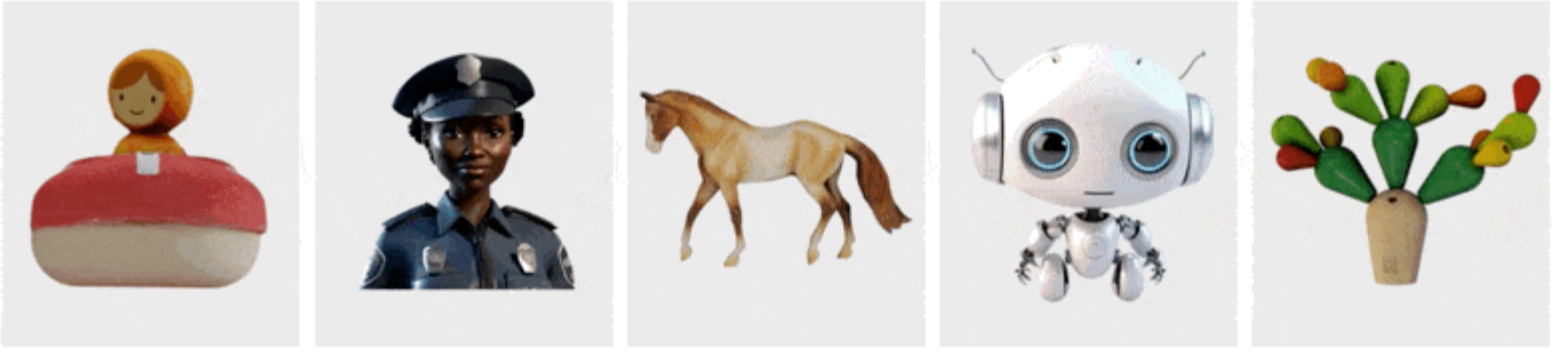
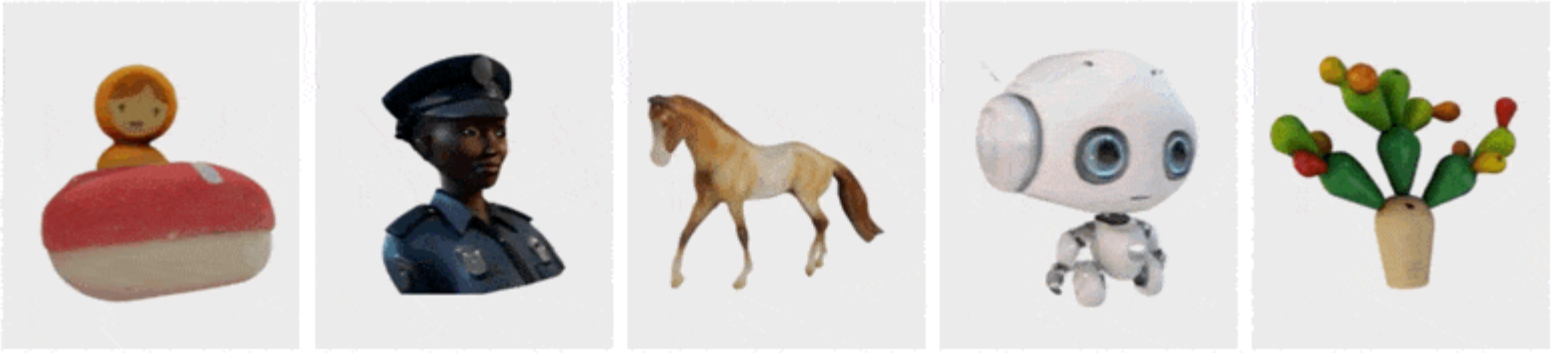
3D OBJECT RECONSTRUCTION MODELS

Input images



TripoSR outputs



GOAL

Generate accurate 3D model from just a single image
→ streamline process to be just as easy as taking videos and pictures

CHALLENGES

- How to detect depth of object
- How can model find out how the back of the given object looks like
- Whats the scale of the given object

BACKGROUND

SHAPE FROM SHADING

- Most basic technique, dates back to 1989
- Use shading information to infer 3D structure
- Unrealistically assumes surface are Lambertian (reflects light uniformly in all directions)

MULTI-VIEW STEREO

” [. . .] the structure from motion theorem which states that the structure of 4 non-coplanar points is recoverable from 3 orthographic projections.” (Ullman and Brenner 1997)

MULTI-VIEW STEREO

- Already covered in computer vision lecture
- Combination of Corner Harris, Difference of Gaussians, ...
- Great relevance in AR/VR to map environment

CONVOLUTIONAL NEURAL NETWORKS (CNN)

- Type of deep learning models especially good for image data
- Designed to adaptively learn spatial hierarchies of features from the data (image)
- Proved to work quite a bit better than previous state-of-the-art

GENERATIVE ADVERSARIAL NETWORKS (GAN)

- Deep learning model consisting of two networks
  - Generator: Generated new data samples
  - Discriminator: Tries to distinguish between real and generated data
- Both models work in competition → Generator tries to fool discriminator

MODELS

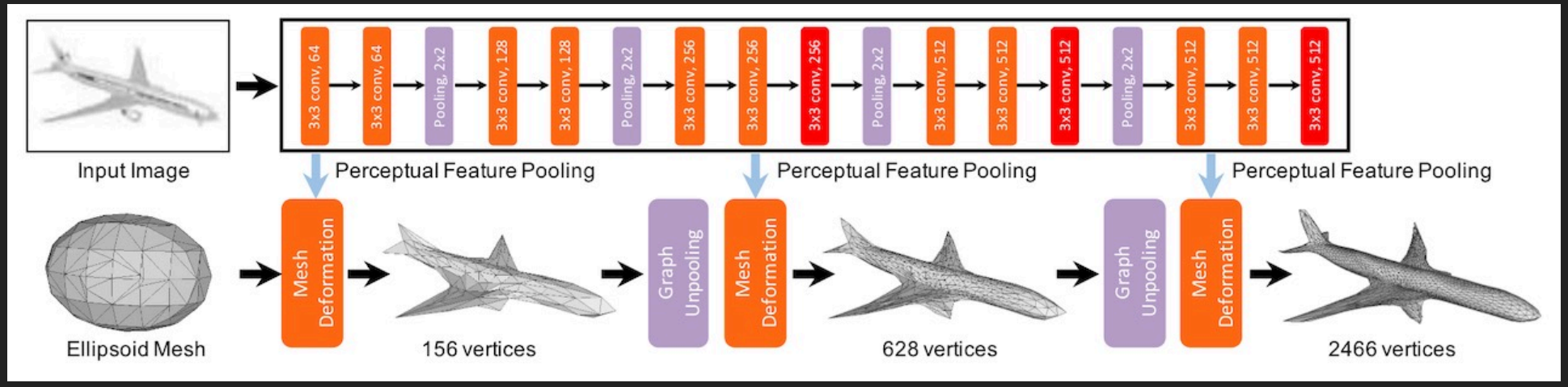
PIXEL2MESH

- Built with two components
  - Image feature network (CNN): extracts perceptual features
  - Cascaded mesh deformation network

PIXEL2MESH

1. Input image is analyzed by image feature network and features are extracted
2. Cascaded mesh deformation network initializes with ellipsoid mesh
3. Features extracted from image are taken to refine shape of the mesh
4. Mesh model is refined iteratively in 3 steps (each step increases mesh resolution)

PIXEL2MESH


Input Image
3x3 conv, 64
3x3 conv, 64
Pooling, 2x2
3x3 conv, 128
3x3 conv, 128
Pooling, 2x2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
Pooling, 2x2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Pooling, 2x2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Perceptual Feature Pooling
Perceptual Feature Pooling
Perceptual Feature Pooling
Ellipsoid Mesh
Mesh Deformation
156 vertices
Graph Unpooling
Mesh Deformation
628 vertices
Graph Unpooling
Mesh Deformation
2466 vertices

PIXEL2MESH++

- Extension of original Pixel2Mesh
- Improves performance of output by using Generative Adversarial Network (GAN)

APPLICATIONS

Field is still relatively new → no mainstream applications

DEVELOPMENT AND ENTERTAINMENT

- Most prominent application
- Could revolutionize asset creation process
- Especially beneficial for small studios or indie developers

Models generated by One-2-3-45++

TODO

- Neural Implicit Functions (NeRF - Neural Radiance Fields)
- Analyze more models to have a better overview
- Refine section: Applications of 3D Mesh Reconstruction