RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc
University of Bath
T.Nguyen-Phuoc@bath.ac.uk

Chuan Li
Lambda Labs
c@lambdalabs.com

Stephen Balaban
Lambda Labs
s@lambdalabs.com

Yong-Liang Yang
University of Bath
Y.Yang@cs.bath.ac.uk

ADVANTAGES

- A novel CNN architecture that enables both rendering and inverse rendering.
- Generalizes well to objects of unseen category and more complex scene geometry.
- Capable of producing textured images from textured voxel grids, where the input textures can be RGB colors or deep features computed from semantic inputs.
- Easy to integrate into other modules for applications, such as texturing or image-based reconstruction.

CURRENT APPROACHES

METHOD

- Rigid-body transformation (world coordinate system to camera coordinate system) followed by trilinear resampling.
- 3D convolutions compute shading color for each pixel.
- Projection unit:
  \[ I_{i,j,k} = f \left( \sum_{d} u_{k,d} \cdot V_{i,j,d} + b_{k} \right) \]
- Pixel-wise loss function:
  \[ L_{\text{recon}} = \frac{1}{n} \sum_{i} ||y_{i} - y'_{i}||^2 \]

- Pixel-wise loss function:
  \[ L_{\text{recon}} = \frac{1}{n} \sum_{i} ||y_{i} - y'_{i}||^2 \]

- Focus on losses and training regimes
- Make few assumptions about the 3D world and the image formation process
- Rotation in latent space using a CNN is hard! [1, 2]
- Do not generalise well to different object categories
- Current differentiable renderers are limited to a single fixed shader. [3, 4]

- Using adversarial loss instead of MSE or BCE
- Using more efficient voxel representation (octree)
- Considering other 3D data types (mesh, point clouds, etc.)
- Learning multiple shaders with one network

INVERSE RENDERING RESULTS

MAP estimation:
\[
\minimize_{z', \theta, \phi, \eta} \| I - f(\theta(g(z'), \theta, \eta)) \|^2
\]

where \( I \) is the observed image and \( f \) is our pre-trained 3D auto-encoder, \( \theta \) and \( \eta \) are the pose and lighting parameters, and \( \Phi \) is the texture vector.

\[
\minimize_{z'} \alpha \| I - f(g(z'), \theta) \|^2 + \beta (z - \mu)^T \Sigma^{-1} (z - \mu)
\]

where \( \mu \) and \( \Sigma \) are the mean and covariance of \( z' \) estimated from the training set respectively.

REFERENCES


COMPARISON

DISCUSSION

ACKNOWLEDGEMENT

This work was supported in part by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665992, the UK’s EPSRC Centre for Doctoral Training in Digital Entertainment (CDE), EP/L016540/1, and CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1. We also received GPU support from Lambda Labs.

Code available on GITHUB