Motivation

- Goal: 3D reconstruction from sparse viewpoint
- Assumptions of previous works:
  - Dense Viewpoints
  - Lambertian and non-uniform albedo
  - Shape Prior
  - Single-view [Kar et al., Aubry et al.]
  - Multi-view [Bao et al.]
  - General lighting condition
  - Dense viewpoints
- Contributions
  - Reconstruction free from assumptions above
  - Unifying single- and multi-view reconstruction
  - 3D convolutional LSTM
  - Attention for view specific update

Recurrent Neural Network

- Memory

- Long Short-Term Memory

- Gated Recurrent Unit

Recurrent Reconstruction Neural Network

- Encoder

- 3D-Convolutional LSTM

- Decoder

3D-Convolutional LSTM

- Attention for viewpoint specific update:
  - LSTM cell assigned for each 3D region

- 3D Convolution: Regularization

3D Convolutional LSTM & GRU

$$f_t = \sigma(W^{f}\mathbf{x}_t + U^{f}\mathbf{h}_{t-1} + b_t)$$

$$h_t = f_t \odot \mathbf{h}_{t-1} + (1 - f_t) \odot \text{tanh}(W^{h}\mathbf{x}_t + U^{h}\mathbf{h}_{t-1} + b_h)$$

Deep Residual GRU/LSTM Network

- 3D Convolution: Regularization

GRU Input Gate Analysis

Multi-view Evaluations

- Cross entropy loss

- Voxel IoU

- Pre-category IoU

Recurrent Stereo vs. Ours


- Multi-view: ShapeNet [1] and real images

Multi-view Stereo vs. Ours

- ShapeNet vs. Ours

- PASCAL 3D+ vs. Ours

References


Acknowledgement

Christoher B. Choy, Danfei Xu*, JunYoung Gwak*, Kevin Chen, Silvio Savarese
Stanford University

* Indicates equal contribution