3D DEEP LEARNING APPROACHES

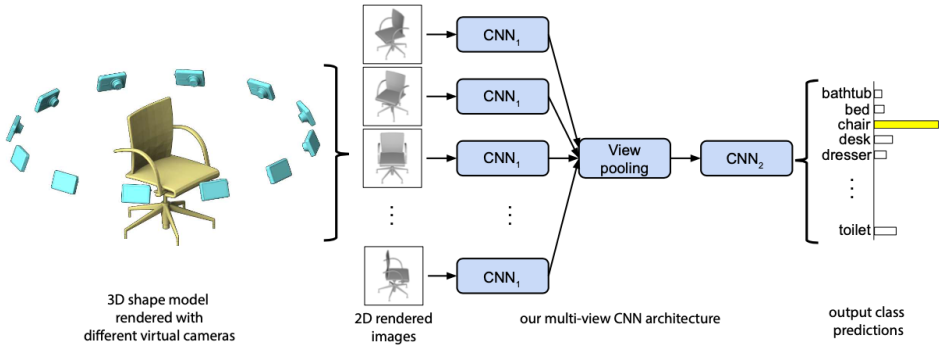
Author
Vinitra Muralikrishnan
11th March, 2024

Contents

1 Multi View Approach **3**

1 Multi View Approach

- Image based networks can process individual shape renderings
- They are basically convolution layer networks combined for reasoning across multiple rendered shape views.



- For classification
  - Get renderings of a shape from different viewpoints. Cover different camera rotations, distances, viewpoints
  - Feed all views of a shape to an FCN.
  - For every view get the predicted probability of all labels
  - Aggregate the probability across all views. The aggregator can be max or mean pooling
  - The maximum value across all labels then would be the prediction
- For segmentation
  - For each input shape, infer a set of viewpoints that covers the maximum (99.99999%) surface
  - Cover each view across multiple distances and with different camera rotations (to cover different orientations as CNN is not rotation invariant)
  - Render shape images (normal dot view vector) encoding surface normals
  - Render depth images encoding surface position relative to the camera
  - Feed each pair of images to an FCN (fully convoluted networks)
  - The views are not ordered so FCNs can share parameters
  - For each input, the FCN will output the confidance map per part label
  - The confidence maps then can be aggregated and projected to get a single surface map
  - For each surface element, final all pixels painted by it in all views. Surface confidence is max of these pixel confidences per label
- For correspondences

  - For every shape, patches of the shape are generated and such patches are fed into an FCN
  - The FCN gives shape descriptors for each of the "patches"
  - View pooling and dimensionality reduction is done to prune no of datapoints or features identified between the 2 shapes
  - Contrastive loss is applied on both shape descriptors and points of difference are identified
  - Example: Siamese Architecture

- Pros
  - Excellent performance for shape classification
  - Re use of powerful 2D image architectures
  - Use of pretrained models
  - Good resolution for shape analysis

- Cons
  - Cannot process invisible areas
  - Can be slow
  - 2d proximity is different from 3d proximity after projection
  - Heuristics for viewpoint positions/orientations need to be carefully designed. Optimal heuristics need to be identified every time