Solution ML3D endterm all

Machine Learning for 3D Geometry (Technische Universität München)

Department of Informatics, Technical University of Munich

Exam on Machine Learning for 3D Geometry

Prof. Dr. Angela Dai

February 9, 2022 Winter 2021 90 Minutes

General Information:

- You have **90 Minutes** to solve the exam, which contains a total of 31 questions.
  You can achieve a maximum of 100 points.
- No additional resources are allowed.
- Do not write with red or green colors nor use pencils.
- Only submit your personalized blackened exam sheet that you downloaded from TUMExam.
  DO NOT submit this exam sheet with questions.

Multiple Choice Questions:

For the multiple choice questions, any number of answers can be correct:
You get points individually per box,

- for a correct answer that is checked, and
- for a wrong answer that is not checked.

There are no negative points.

Part I: Multiple Choice

1. (2 points) Check all that apply: For a triangle mesh which is a manifold, which of the following is true?

    ☐ The intersection of any two triangles must be non-empty.

    √ **Every edge must have at most two adjacent triangles.**

    ☐ The intersection of any two triangles must have a common edge.

    √ **Some edges may have only a single adjacent triangle.**

2. (2 points) Check all that apply: Which of the following is true for 3d representations?

    ☐ Point clouds can be converted to an implicit representation using Marching Cubes.

    √ **Point clouds represent surfaces more efficiently than dense grids.**

    ☐ Given two implicit surfaces A and B, their intersection is always defined as min(A, B).

    ☐ One major disadvantage of dense grids is that neighbor operations are expensive.

3. (2 points) Check all that apply: Multi-view learning can be used in combination with:

    √ **Dense volumetric neural networks.**

    √ **Point cloud neural networks.**

    √ **Mesh neural networks.**

    √ **Implicit surface neural networks.**

4. (2 points) Check all that apply: In a polygon mesh,

    ☐ Edges may intersect each other.

    √ **Every edge belongs to at least one polygon.**

    ☐ All vertices must have the same degree.

    √ **All polygons are closed and simple.**

5. (2 points) Check all that apply: When estimating the transform between two rigid shapes,

    ☐ The transform contains rotation and translation for a total of 4 degrees of freedom.

    √ **One usually aims for more correspondences than degrees of freedom.**

    ☐ The geometry of both shapes must be identical.

    ☐ A shape descriptor is always used to establish point correspondences.

6. (2 points) Check all that apply: A good distance measure for shape alignment

    √ **Is often measured by aggregating correspondences between the shapes.**

    ☐ Is an ℓ_2 distance.

    ☐ Must be symmetric and follow the triangle inequality.

    √ **Supports partial matches.**

7. (2 points) Check all that apply: The following deep network approaches can be used to learn semantic segmentation directly on a 3D mesh:
   - ☐ 3D sparse convolutions.
   - √ **Geodesic convolutions.**
   - √ **Message-passing along mesh edges and/or faces.**
   - √ **Convolutions on mesh edges as a basis.**

8. (2 points) Check all that apply: DeepSDF, which learns an implicit reconstruction of 3D shapes,
   - ☐ Takes only a 3D (x, y, z) point location as input.
   - ☐ Directly outputs a 3D mesh reconstruction of a shape.
   - ☐ Requires only a single forward pass through the network to generate a shape.
   - √ **Is not bound to any explicit surface resolution.**

9. (2 points) Check all that apply: Learning deformations of a template mesh:
   - √ **Can be used on CAD models to real scene observations.**
   - √ **Can be done by predicting vertex offsets.**
   - ☐ Can allow for representation of arbitrary topologies.
   - ☐ Cannot result in self-intersections.

10. (2 points) Check all that apply: sparse 3D convolutions
    - √ **Are more memory-efficient for training than dense 3D convolutions for the same voxel resolution.**
    - ☐ Store fewer weights per convolution than dense 3D convolutions for the same kernel size.
    - ☐ Can operate directly on raw point cloud data.
    - ☐ Cannot be used for generative or reconstruction tasks.

Part II: 3D Surface Representations

1. (2 points) Explain the main steps in converting an oriented point cloud into a surface mesh representation without the use of a deep network.

   + points to implicit with poisson
   + implicit to mesh with marching cubes

2. (2 points) Compare and contrast the use of dense volumetric data with point cloud input for a deep network. Name one advantage and one disadvantage for each data modality.

   0.5 for each +/-.
   volumetric: -cubic memory/compute, +regular grid structure/easy neighbor queries, +easy pooling
   points: -unstructured/no easy neighbors, +efficient surface representation

3. (1 point) Name one advantage of incorporating multi-view information into a neural network which previously operated only on surface geometry information.

   high resolution color information, efficient (pre-)training from 2d datasets

4. (1 point) Name one reason to use a signed distance field representation over an occupancy grid representation to characterize a surface.

   superresolution, information in empty space, inside/outside or visible/unknown information

5. (2 points) For two implicit surfaces f and g defined with positive values outside the surface and negative values inside the surface, explain how to compute the boolean union $f \cup g$ and

the boolean subtraction $f - g$.

0.5: union min(f,g)
1.5: subtraction max(f, -g)

Part III: Geometric Registration

1. (5 points) Given perfect correspondences $\{x_i\}, \{y_i\} \in \mathbb{R}^3$, state the optimization objective (as a formula) and describe a non-iterative approach to estimate rotation $\mathbf{R}$ and translation $\mathbf{t}$.

- $\min_{R,t} \sum_{i=1}^{N} ||Rx_i + t - y_i||_2^2$
- Remove translation by mean-centering
- Compute SVD: UDV^T
- Define S as I if det(U)det(V) = 1 else diag(1, ..., 1, -1)
- R = USV^T

2. (2 points) You are given two rigid shapes A and B without any correspondences between them. Describe a method (non deep learning based) that can be used to find the rigid transformation between them.

- ICP; Iterate over the following points:
- Find corresponding points P_A and P_B based on proximity
- Find optimal transform R, t
- Apply optimized R, t

3. (2 points) Name and explain (in one sentence each) two possible approaches that can be taken for a global registration of shapes in arbitrary positions.

Any two of:
- Exhaustive Search (sample space of possible initial alignments, and use best)
- Normalization (center all shapes at origin and ise PCA to find principal directions)
- Random Sampling (iterate picking random pair of points on both shapes, estimate, guess and verify)
- Invariance (calculate matching descriptors at invariant feature points) - Deep Network (with good explanation)

Part IV: 3D Object Classification

1. (4 points) PointNet is a popular network architecture that takes in shapes as point clouds and predicts their class while being agnostic to the input point ordering. How does it achieve this?

- MLP with shared weights as encoder (1)
- Input and feature transform networks (1)
- Global feature aggregation (.5) and MLP for output score prediction (.5)

- Symmetric function (max pooling) for feature aggregation (1)

2. (2 points) Imagine you have version of PointNet, pretrained for classification on the ShapeNet database. You now want to use this to compute a global shape descriptor for previously unseen shapes during test time. How would you modify the network architecture and after which layer do you extract the global descriptor features?

- Remove decoder / use only encoder (1)
- Take global features after max pooling operation (1)

3. (2 points) Name one follow-up work that improves classification performance by also capturing local structures of shapes. How does it do so and what are the essential operations newly introduced in this architecture?

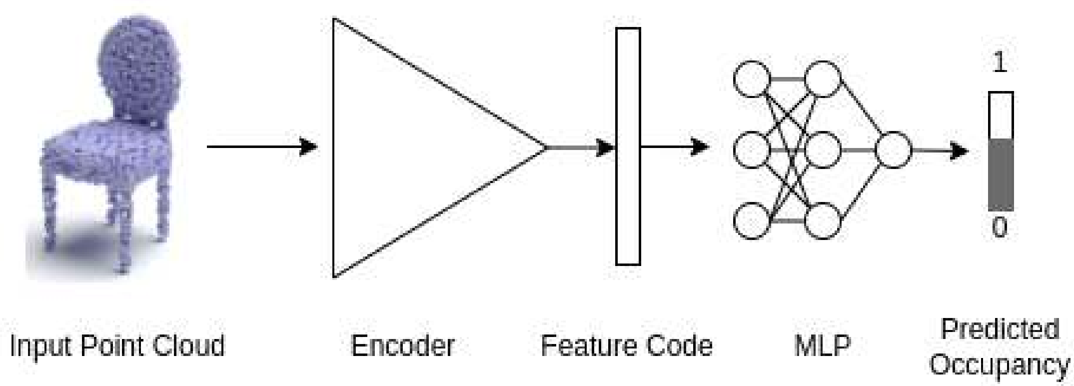
- PointNet++ (.5)
- Hierarchical point set feature learning (.5)
- Sampling & Grouping (1) / Set abstraction layers (1)

Part V: Shape and Scene Segmentation

1. (1 point) You are provided with 5000 shapes from the ShapeNet dataset, all belonging to the chair class. The shapes come only as geometry, i.e. without any texture or semantic information. Your task is to perform part segmentation for each of these shapes. Name a method you can use in this scenario for this task.

BAE Net unsupervised part segmentation, unsupervised co-segmentation,

2. (2 points) The following figure shows a surface reconstruction method inspired from Occupancy Networks. How would you modify its architecture and training supervision such that it can be used for part segmentation of objects?



Replace head to output semantic classes.
Train with semantic labels: Should describe how the data is sampled & supervision is performed

3. (1 point) One challenge associated with moving from part segmentation of objects to semantic segmentation of scenes is the issue of the arbitrary scale of scenes (e.g., a room vs a building). Describe one approach to handle this scale variety in semantic segmentation of scenes.

Train on crops evaluate on scenes, must be fully-conv; sliding window

4. (1 point) Describe the conclusion that RevealNet provides for the relation between completion and segmentation for instance segmentation.

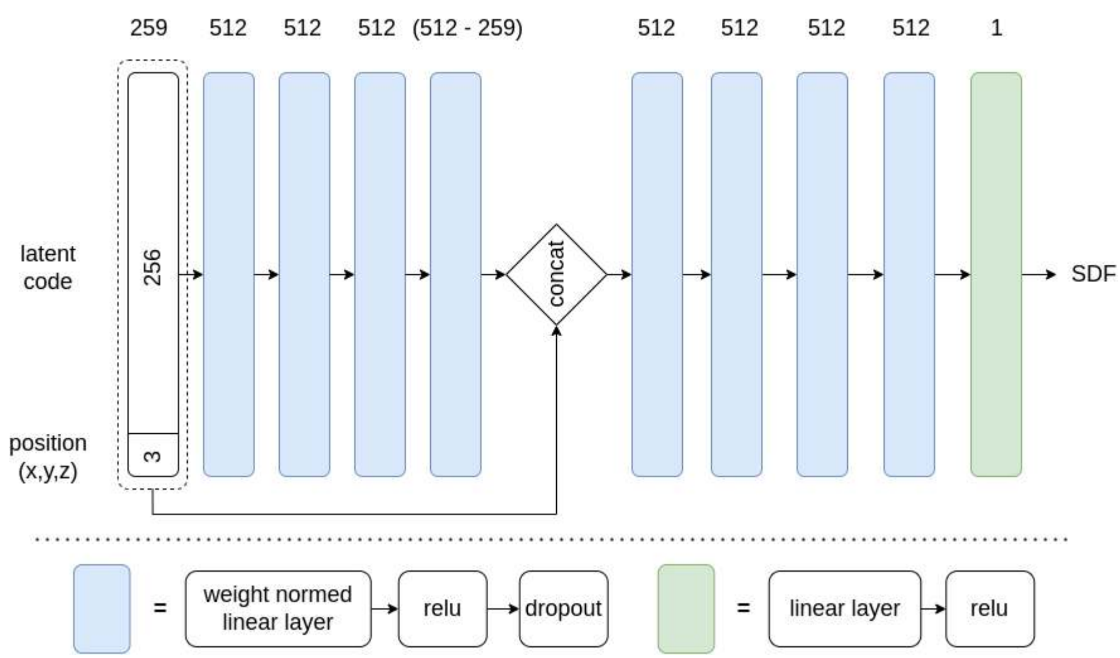
Completion helps segmentation

Part VI: Generative Models for 3D Reconstruction

1. (2 points) In the lecture, you were introduced to DeepSDF and Occupancy Networks, two very similar methods for shape reconstruction and representation. Describe two differences between these methods.

   - Representation (occupancy vs sdf)
   - Encoder vs Autodecoding

2. (2 points) Your colleague came up with an implicit model for shape representation inspired from DeepSDF, but their model, shown below, cannot even overfit to one shape. Point out a fundamental issue with the model and explain why it is a problem.



   Relu in the output
   Problem because: No negative outputs

3. (3 points) Explain the idea behind Convolutional Occupancy Networks. For what kinds of reconstruction problems do you expect Convolutional Occupancy Networks to significantly

outperform methods like Occupancy Networks and DeepSDF?

- (2) Local implicit method (convolutions to get local features followed by implicit decoder)
- (1) Scene reconstruction

4. (3 points) Supervised scene completion methods are typically trained with synthetic data, since real world scanned data is incomplete due to physical limitations in the scanning process, and thus cannot be used for supervision. Name and explain a method which overcomes this issue and learns scene completion from real world scans.

- SGNN [1]
- Describe SGNNs self supervision [2]

Part VII: Project Questions

Please keep your answers short (do not exceed the given space below the questions!). Also try not to spend more than 15 minutes on this part! We would advise you to work on this part in the last 15 minutes of the exam.

1. (20 points) Describe your project task; i.e., what is your problem statement and how did you plan to approach it? For instance describe the theoretical foundation behind the approach (i.e., what you promised in your proposal).

   - Individual project description has to make sense

2. (20 points) Give a high-level overview of the technical solution of your problem.

   - Individual project description has to make sense