CS282A Final Project: Implementation of the PointNet Architecture

Arm Wonghirundacha (3038533201), Joohwan Seo (3037346750), Seunghoon Paik (3037246747)

Abstract

In this final project, we designed homework for CS182 students to understand and implement the PointNet architecture. The homework is designed in the following way: (i) The students will review and explore some basic mathematical concepts, including permutation invariance and matrix transformation, which take fundamental roles in the design of the PointNet architecture; (ii) the students will implement the PointNet architecture by filling out selectively chosen blank codes in the prepared Jupyter notebook file, including T-Net (Transformation Network) and the transformer for data augmentation, which are the key parts in the implementation of the PointNet architecture. Carrying out the coding task, the students are able to have more comprehensive understanding of the PointNet architecture by: visualizing the point cloud data; comparing different sampling methods for the point cloud data; checking the role of T-Net in the model; training PointNet models for two different tasks, classification and part segmentation; and visualizing the training loss and the test accuracy for different tasks and sampling methods. For the sake of the wall clock time of model training, the preprocessed data by us is given to the students. This leads to about 20x times faster training time compared to using the raw data, so that the students can do multiple trial-and-error runs in a short period of time, and find better hyperparameters for the models.

Contents

1 Part I. Review comments

1.1 Regarding Comments from Reviewer 1

1. The content of the homework assignment looks correct. Code solutions are provided. (Excellent work, no actions needed.)
2. The write-up contains a page of scaffolding to motivate the PointNet architecture. This gives students the relevant information to understand applications. It also summarizes why certain properties of the architecture are important as a basis for the problems. (Excellent work, no actions needed.)
3. The assignment is well written. (Excellent work, no actions needed.)
4. The commentary on the HW seems to be missing. The assignment itself has a one page preamble on the context of the problem, but is missing an explicit commentary section. (Medium improvement needed)
5. The assignment is a straightforward set of problems on the basics of PointNet. The coding portion of the assignment is well-designed and gives students good hands-on experience with the architecture. (Excellent work, no actions needed.)

1.2 Regarding Comments from Reviewer 2

1. The content matches the paper and engage with important concepts like data invariances, weight sharing and geometric transformation. The main focus of the submission is the implementation of pointNet that I found overall pretty clear. I would comment more the jupiter recalling some theory since there is not much of it. The coding part is ok. I would have complete the exercise in less than 1.30h. No LATEX source is provided. Overall good job. (Small improvement needed)
2. No sanity checks are provided. It could have been pretty easy to do it. Good code, very easy to read. The hw is self contained. (Small improvement needed)
3. You should provide directly the link to a colab. Since I had to clone the repo and upload the jupyter on colab to make it run smoothly. I liked the code. Minor typos (line16 network.py 'make' not 'ake') Typo subsection 1.2 in latex, The T-net is (a) neural net... And for clarity I would improve the paragraph that start with "we take the transformation matrix multiplied.....using a transformation matrix." (Small improvement needed)
4. The commentary is great good job. Enjoyed a lot reading it. (Excellent work, no actions needed.)
5. Simplifying the paper. Creating solid visualizations to help students understand what's going. (Excellent work, no actions needed.)

1.3 Regarding Comments from Reviewer 3

1. Coding questions engage with selected key concepts in the paper and are easy to follow and pedagogically useful: Yes.

   Implementation is doable for someone with CS 182 level mathematical maturity (not too difficult, but not completely trivial either): I think the written question were definitely pretty straightforward, but maybe more demo level than homework level (although at some level I wish the cs282a course staff had sometimes done more of this). Similarly the coding question would be pretty fast for most cs182 students. I think at least asking students to interpret results for the coding comparison would be a good addition. Anyways I hate to be the person who's saying a homework is a bit too easy (since I've spent the entire semester feeling that the actual homeworks are too long and some of the coding takes me too much time to debug), but I think this homework could probably use a couple of extra questions to engage students more before it becomes a real homework. But definitely a great start for a real homework set!

   Any exposition/mathematical background about the concepts as provided in the problem description is correct: Yes, I believe so.

   Code solutions are provided and fully correct: Yes.

   If there are analytical questions, the full LaTeX and PDF export of the assignment (and solutions if separate) are provided. LaTeX files compile correctly: PDF is included. There is no LaTeX file, but I'm not sure from this wording it that was required.

   The problems and solutions are completely correct and engage with the material in a pedagogically useful way: Yes.

   Doing the full problem (coding and written parts) would take an average CS 182 student 1.5-2hr to complete: I think 1.5 hours is a reasonable time estimate. (Small improvement needed)

2. Project provides necessary scaffolding for a student to engage with the material: Yes

   Any code that students are expected to implement is explained clearly through the use of text cells in Jupyter and/or code comments: Yes, student coding is minimal.

   If the project group uses any external packages, these are provided, or there are simple instructions on how to download them from before starting the assignment: Yes.

   If possible, autograding tests/sanity checks to make sure a student's implementation is correct are provided: I don't think so, but also student coding is minimal. Gradescope autograder files are not needed—simple sanity checks that compare actual values against expected values are enough: Yes (Excellent work, no actions needed.)

3. HW assignment and commentary are easy to read and follow: Yes.

   Any mathematical notation used is understandable, and any non-standard notation is clearly explained: Yes.

   The assignment and commentary are free of spelling/grammar errors: Relatively few spelling/grammar errors, but there are a handful. Probably worth one read through. (Excellent work, no actions needed.)

4. Project has a 2-3 page commentary on the HW that explains the key concepts in the paper and how the assignment engages with them: Very possible I just missed this, but I don't see a commentary document.

   The key concepts that are engaged with in the HW assignment are explained briefly in the commentary (it is fine if it is also explained in the assignment itself): Once again couldn't find this, but could just be me.... I am slightly sleep deprived here. (Medium improvement needed)

5. Going above and beyond what was in the original paper. Could include:

   Simplifying the paper: they do this a bit in their descriptions.

   Applying the results in a new way: No.

   Conducting additional systematic experimentation and explaining the results: conducted systematic experimentation, but didn't really explain the results.

   Creating solid visualizations to help students understand what's going on and helping them interpret what they see: I liked the visualizations up front and I think that counts for this category, but there wasn't as much interpreting what the students sees... although maybe that could just be a question for students in the homework.

   Writing creative analytical problems that engage with the material at a deeper level well beyond what was presented in the paper: No

   Creating synthetic datasets that illustrate certain ideas, etc. : No (Excellent work, no actions needed.)

1.4 Regarding Comments from Reviewer 4

1. The assignment offers a solid introduction to the PointNet architecture and its applications in 3D point cloud data analysis. It does a commendable job of explaining the concept of translational invariance and its significance within the architecture. However, the assignment falls short in providing a thorough explanation of the need for Transformation Networks (T-nets) and the underlying concepts involved. Additionally, both the written and coding components of the assignment could benefit from a stronger focus on understanding the design choices behind PointNet, rather than solely concentrating on its properties. (Small improvement needed)

2. The coding portion of the assignment is well-organized and presents a coherent flow, allowing students to effectively grasp the various uses of PointNet and the implications of adjusting specific variables within the model. This structured approach facilitates a deeper understanding of the PointNet architecture and its inner workings, while also providing opportunities for students to experiment with different configurations and observe their effects on the model's performance. (Excellent work, no actions needed.)

3. The assignment exhibits a reasonable degree of clarity and readability, but it occasionally suffers from a disjointed presentation of concepts, making it difficult to follow the logical progression of ideas. While some concepts are well-explained, others are not adequately

expounded upon, leading to an inconsistent learning experience. Furthermore, the assignment presents an odd mix of high-level and low-level discussions, which can make it challenging for students to maintain a consistent level of understanding throughout the material. (Small improvement needed)

4. I don't know if I'm mixing things up, but I don't see the commentary in the submission. If the attached pdf is the commentary, it was unclear because there are homework questions in the same document. (Medium improvement needed)

5. The assignment excels in certain aspects, such as its focus on translational invariance and the various factors of the Transformation Network (T-net), providing students with a solid foundation in these areas. However, there is room for improvement in terms of incorporating more visualizations related to the model and architecture decisions, which would help students better comprehend the design choices behind PointNet. On the positive side, the inclusion of point cloud visualizations effectively enhances the learning experience. Additionally, the novel implementation of the dataset loader is noteworthy, as it significantly reduces training time and allows students to experiment more efficiently. (Small improvement needed)

2 Part II. Point-to-point response to the review comments

2.1 Regarding Comments from Reviewer 1

1. The content of the homework assignment looks correct. Code solutions are provided. (Excellent work, no actions needed.)

   Thanks for the comment.

2. The write-up contains a page of scaffolding to motivate the PointNet architecture. This gives students the relevant information to understand applications. It also summarizes why certain properties of the architecture are important as a basis for the problems. (Excellent work, no actions needed.)

   Thanks for the kind comments.

3. The assignment is well written. (Excellent work, no actions needed.)

   Thanks for the kind comments.

4. The commentary on the HW seems to be missing. The assignment itself has a one page preamble on the context of the problem, but is missing an explicit commentary section. (Medium improvement needed)

   We included instructive guidelines for the task and explanations of their importance throughout the written and coding assignments, but we were not aware of the requirement for the specific commentary files about the homework. Thanks for pointing out this issue, and we included the commentary file in our final submission.

5. The assignment is a straightforward set of problems on the basics of PointNet. The coding portion of the assignment is well-designed and gives students good hands-on experience with the architecture. (Excellent work, no actions needed.)

   Thanks for the kind comments.

2.2 Regarding Comments from Reviewer 2

1. The content matches the paper and engage with important concepts like data invariances, weight sharing and geometric transformation. The main focus of the submission is the implementation of pointNet that I found overall pretty clear. I would comment more the jupiter recalling some theory since there is not much of it. The coding part is ok. I would have complete the exercise in less than 1.30h. No LATEX source is provided. Overall good job. (Small improvement needed)

   We include the LaTeX file in the final submission. Regarding the homework time, we expect 1.5 to 2 hours are needed for average CS182 students. This is possible due to 20x faster training which uses our preprocessed data. Without preprocessed data, one epoch of the training takes 40 mins.

2. No sanity checks are provided. It could have been pretty easy to do it. Good code, very easy to read. The hw is self contained. (Small improvement needed)

   We include the checkpoints where the students can see their current implementation is correct or not.

3. You should provide directly the link to a colab. Since I had to clone the repo and upload the jupyter on colab to make it run smoothly. I liked the code. Minor typos (line16 network.py 'make' not 'ake') Typo subsection 1.2 in latex, The T-net is (a) neural net... And for clarity I would improve the paragraph that start with "we take the transformation matrix multiplied.....using a transformation matrix." (Small improvement needed)

   In the original submission, we included the link to Jupyter notebook file, as well as the link to GitHub repository. We changed the typos and revised the paragraph.

4. The commentary is great good job. Enjoyed a lot reading it. (Excellent work, no actions needed.)

   We included instructive guidelines for the task and explanations of their importance throughout the written and coding assignments, but we were not aware of the requirement for the specific commentary files about the homework. Thanks for pointing out this issue, and we included the commentary file in our final submission. (We are glad that the reviewer enjoyed it.)

5. Simplifying the paper. Creating solid visualizations to help students understand what's going. (Excellent work, no actions needed.)

   Thanks for the kind comments.

2.3 Regarding Comments from Reviewer 3

1. Coding questions engage with selected key concepts in the paper and are easy to follow and pedagogically useful: Yes.

   Thanks for the comment.

   Implementation is doable for someone with CS 182 level mathematical maturity (not too difficult, but not completely trivial either): I think the written question were definitely pretty straightforward, but maybe more demo level than homework level (although at some level I wish the cs282a course staff had sometimes done more of this). Similarly the coding question would be pretty fast for most cs182 students. I think at least asking students to interpret results for the coding comparison would be a good addition. Anyways I hate to be the person who's saying a homework is a bit too easy (since I've spent the entire semester feeling that the actual homeworks are too long and some of the coding takes me too much time to debug), but I think this homework could probably use a couple of extra questions to engage students more before it becomes a real homework. But definitely a great start for a real homework set!

   The homework is designed to be done in 1.5-2 hours for average CS182 students. Also we preprocessed the data to make the training time 20x faster, and this leads to shorter estimated time for the homework. However, we added an additional question in the final submission.

Any exposition/mathematical background about the concepts as provided in the problem description is correct: Yes, I believe so.

Thanks for the comment.

Code solutions are provided and fully correct: Yes.

Thanks for the comment.

If there are analytical questions, the full LaTeX and PDF export of the assignment (and solutions if separate) are provided. LaTeX files compile correctly: PDF is included. There is no LaTeX file, but I'm not sure from this wording it that was required.

We include the original $\LaTeX$ in zip file in the final submission.

The problems and solutions are completely correct and engage with the material in a pedagogically useful way: Yes.

Thanks for the comment.

Doing the full problem (coding and written parts) would take an average CS 182 student 1.5-2hr to complete: I think 1.5 hours is a reasonable time estimate. (Small improvement needed)

Thanks for the comment. We commented about the estimated homework time on the previous comment of the reviewer.

2. Project provides necessary scaffolding for a student to engage with the material: Yes

   Any code that students are expected to implement is explained clearly through the use of text cells in Jupyter and/or code comments: Yes, student coding is minimal.

   If the project group uses any external packages, these are provided, or there are simple instructions on how to download them from before starting the assignment: Yes.

   If possible, autograding tests/sanity checks to make sure a student's implementation is correct are provided: I don't think so, but also student coding is minimal. Gradescope autograder files are not needed—simple sanity checks that compare actual values against expected values are enough: Yes (Excellent work, no actions needed.)

   We appreciate the comment. We added the sanity checks in the final version of the jupyer notebook file in the submission.

3. HW assignment and commentary are easy to read and follow: Yes.

   Any mathematical notation used is understandable, and any non-standard notation is clearly explained: Yes.

   The assignment and commentary are free of spelling/grammar errors: Relatively few spelling/grammar errors, but there are a handful. Probably worth one read through. (Excellent work, no actions needed.)

   Thanks for the comment. We also checked the submission thoroughly and did our best to catch any grammatical error.

4. Project has a 2-3 page commentary on the HW that explains the key concepts in the paper and how the assignment engages with them: Very possible I just missed this, but I don't see a commentary document.

The key concepts that are engaged with in the HW assignment are explained briefly in the commentary (it is fine if it is also explained in the assignment itself): Once again couldn't find this, but could just be me.... I am slightly sleep deprived here. (Medium improvement needed)

We included instructive guidelines for the task and explanations of their importance throughout the written and coding assignments, but we were not aware of the requirement for the specific commentary files about the homework. Thanks for saying this issue, and we included the commentary file in our final submission.

5. Going above and beyond what was in the original paper. Could include:

   Simplifying the paper: they do this a bit in their descriptions.

   Applying the results in a new way: No.

   Conducting additional systematic experimentation and explaining the results: conducted systematic experimentation, but didn't really explain the results.

   Creating solid visualizations to help students understand what's going on and helping them interpret what they see: I liked the visualizations up front and I think that counts for this category, but there wasn't as much interpreting what the students sees... although maybe that could just be a question for students in the homework.

   Writing creative analytical problems that engage with the material at a deeper level well beyond what was presented in the paper: No

   Creating synthetic datasets that illustrate certain ideas, etc. : No (Excellent work, no actions needed.)

   Thanks for the comment.

2.4 Regarding Comments from Reviewer 4

1. The assignment offers a solid introduction to the PointNet architecture and its applications in 3D point cloud data analysis. It does a commendable job of explaining the concept of translational invariance and its significance within the architecture. However, the assignment falls short in providing a thorough explanation of the need for Transformation Networks (T-nets) and the underlying concepts involved. Additionally, both the written and coding components of the assignment could benefit from a stronger focus on understanding the design choices behind PointNet, rather than solely concentrating on its properties. (Small improvement needed)

   The concept of the T-Net and its role are explained and observed in the written assignment. Also, the effect of T-Net in the model can be checked by the final visualization of the training loss and test accuracy.

2. The coding portion of the assignment is well-organized and presents a coherent flow, allowing students to effectively grasp the various uses of PointNet and the implications of adjusting specific variables within the model. This structured approach facilitates a deeper understanding of the PointNet architecture and its inner workings, while also providing opportunities for

students to experiment with different configurations and observe their effects on the model's performance. (Excellent work, no actions needed.)

Thanks for the kind comments.

3. The assignment exhibits a reasonable degree of clarity and readability, but it occasionally suffers from a disjointed presentation of concepts, making it difficult to follow the logical progression of ideas. While some concepts are well-explained, others are not adequately expounded upon, leading to an inconsistent learning experience. Furthermore, the assignment presents an odd mix of high-level and low-level discussions, which can make it challenging for students to maintain a consistent level of understanding throughout the material. (Small improvement needed)

   We added more figures on the written assignment and the Jupyter notebook file refer those figures for better understanding of the task.

4. I don't know if I'm mixing things up, but I don't see the commentary in the submission. If the attached pdf is the commentary, it was unclear because there are homework questions in the same document. (Medium improvement needed)

   We included instructive guidelines for the task and explanations of their importance throughout the written and coding assignments, but we were not aware of the requirement for the specific commentary files about the homework. Thanks for saying this issue, and we included the commentary file in our final submission.

5. The assignment excels in certain aspects, such as its focus on translational invariance and the various factors of the Transformation Network (T-net), providing students with a solid foundation in these areas. However, there is room for improvement in terms of incorporating more visualizations related to the model and architecture decisions, which would help students better comprehend the design choices behind PointNet. On the positive side, the inclusion of point cloud visualizations effectively enhances the learning experience. Additionally, the novel implementation of the dataset loader is noteworthy, as it significantly reduces training time and allows students to experiment more efficiently. (Small improvement needed)

   We added more figures on the written assignment and the Jupyter notebook file refer those figures for better understanding of the task.

3 Part III. Final submission: see next page

To present the contents of the final submission in a tidy format, we present the final version of the homework on next few pages.

It is the written assignment part of the homework. The link for Jupyter notebook file (and GitHub repository as well) for the coding assignment is included in the written assignment.

CS 282A Deep Neural Networks

Homework : Implementation of the PointNet Architecture

Homework Designer: Arm Wonghirundacha, Joohwan Seo, Seunghoon Paik

1 Understanding PointNet Architecture

We introduce PointNet[2], a novel neural network architecture that directly processes point cloud data for tasks such as image classification, image segmentation and semantic segmentation, where we see examples of these tasks in Figure 1.

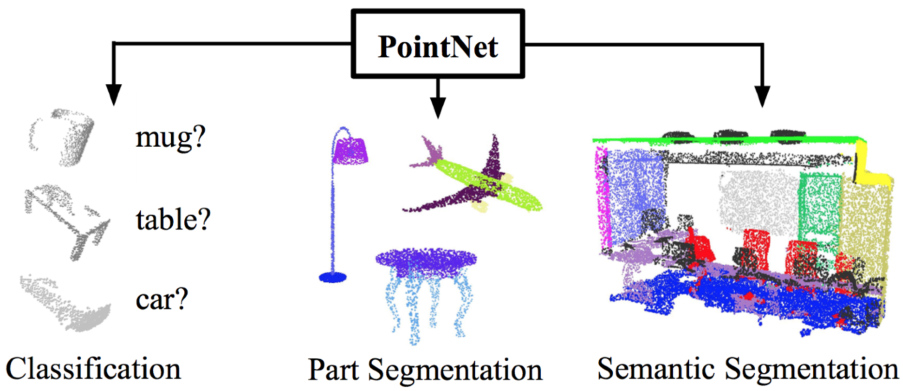


Figure 1: Tasks for PointNet [2]

In order to understand the motivation behind the architecture, we need to understand why point clouds are a desirable three-dimensional representation. Firstly, three-dimensional sensors such as LiDAR (Light Detection and Ranging) and Depth sensors record data that are close in similarity to point cloud data. It is also a simple and unified structure that avoids the combinatorial irregularities and complexities of meshes, making them easier to learn from. Point clouds also can represent a wide range of geometric structures, including irregular shapes and surfaces, which may be difficult to capture with other data types like voxel grids or images. Lastly, point clouds are inherently unordered sets of points, which makes them permutation invariant and thus more flexible for processing with neural networks.

With this in mind, we can see why learning from point clouds is beneficial. The primary challenge to learning from point clouds is that the model needs to be invariant to permutations. This is due to the unordered nature of the data. We also must ensure the model has invariance under geometric transformations, meaning point cloud rotations should not alter classification results.

These two main challenges lead us to the architecture design choices in PointNet: using symmetric functions (max pooling) for permutation invariance and learning a spatial transformer network (T-Net) for invariance under geometric transformations. These choices can be seen in Figure 2.

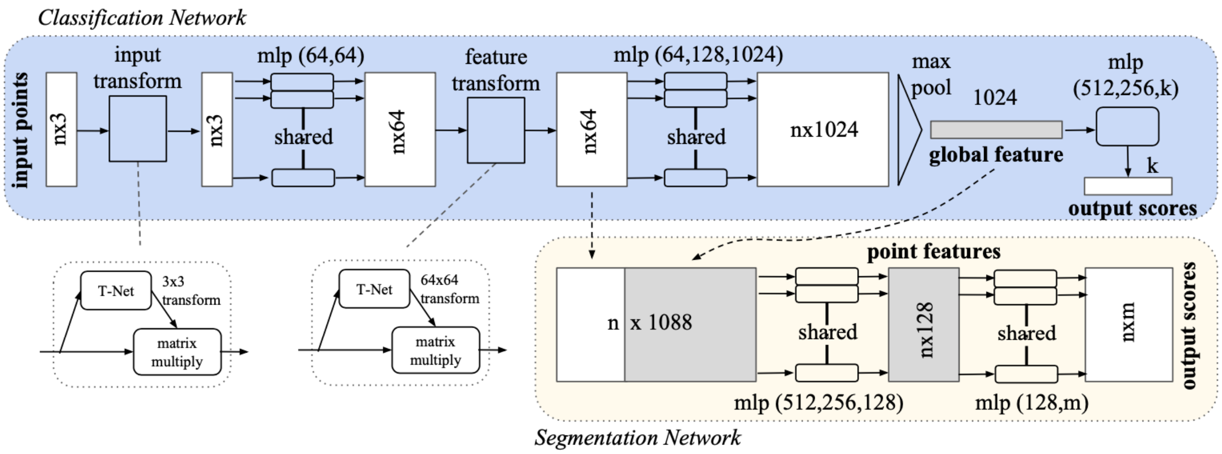


Figure 2: PointNet architecture [2]

1.1 Permutation Invariance

A critical concept of the PointNet is permutation invariance: the results of the PointNet should not change as the sequence of the points changes. Consider, for example, a set of 3 points, $p_i = [x_i, y_i, z_i]^T$, for $i = 1, 2, 3$. Then, the outputs of the PointNet function $f(\cdot)$ should satisfy:

$$f(p_1, p_2, p_3) = f(p_1, p_3, p_2) = \cdots = f(p_3, p_2, p_1)$$

Question 1.
Select all permutation invariant functions and note the reason why briefly. You may present a counterexample if the function is not permutation invariant. Note that without loss of generality, $x_i \in \mathbb{R}^n$, $f_j : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$

1. $f_1(x_1, x_2) = \frac{1}{2}(x_1 + x_2)$
2. $f_2(x_1, x_2) = \max(x_1, x_2)$
3. $f_3(x_1, x_2) = w_1 x_1 + w_2 x_2$, $w_1, w_2 \in \mathbb{R}$, $w_1 \neq w_2$
4. $f_4(x_1, x_2) = (a^T x_1) x_2$, $a \in \mathbb{R}^n$
5. $f_5(x_1, x_2) = w_1 x_1 + w_1 x_2$, $w_1 \in \mathbb{R}$

(Sol) 3 and 4 are not permutation invariant. For 3, since the weight is not shared, the output is not the same. For 4, consider the case that x_1 is orthogonal to a, but x_2 is not orthogonal to a.

1.1.1 Weight sharing for permutation invariance

As you can see in the previous part, weight-sharing is needed to guarantee the permutation invariance of the PointNet. There is an easy way to implement the weight-shared linear layer in PyTorch-`torch.nn.Conv1d`, and we will implement the code using this function.

Question 2.
Please refer to `https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html` for the usage of `Conv1d` and the notations. Suppose that our $C_{in} = 3$, the Euclidean coordinates, $C_{out} = 64$, $L = 1$, $N = n_{batch}$.

1. Briefly describe how `Conv1d` work as the weight-shared fully connected linear layer.

   (Sol) `Conv1d` with the kernel size 1 and stride value 1 aggregates the channel information and outputs the a scalar. Since there are 64 kernels, the output size will be $N \times 64$, where the input size is $N \times 3$. Therefore, if we make the size comparison, we can also understand that `Conv1d` is mapping 3 to 64 for each batch - which works as a linear layer. In this sense, `Conv1d` provides a single-line implementation for the weight-shared linear layer.

2. For `Conv1d` to work as the weight-shared fully connected linear layer, what should be the value of stride and kernel size?

   (Sol) Filter size and the stride should be 1.

1.2 Invariance under geometric transformations

The T-Net is a neural network architecture used in PointNet to perform the alignment and transformation of input points and features. The T-Net consists of a small multi-layer perceptron (MLP) followed by a matrix multiplication layer that outputs a 3×3 or 64×64 transformation matrix. This matrix is then applied to the input points or features to perform alignment and transformation. The transformation matrix is desirable since it helps re-orientate the 3D object and makes the model more robust. An example of this is, if a chair is upside down, it is still a chair, and we want the model to not be confused by an upside down, or an upright chair, so we learn a transformation matrix to re-orientate the chair into a good position to be learned. The T-Net is used twice in PointNet: once for input point alignment/transformation (T1) and once for feature alignment/transformation (T2).

Figure 3 shows us the effect of a learned affine transformation matrix on an input image.

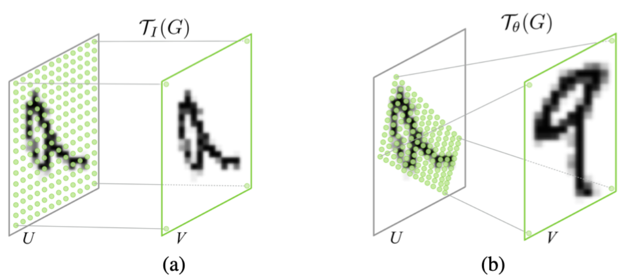


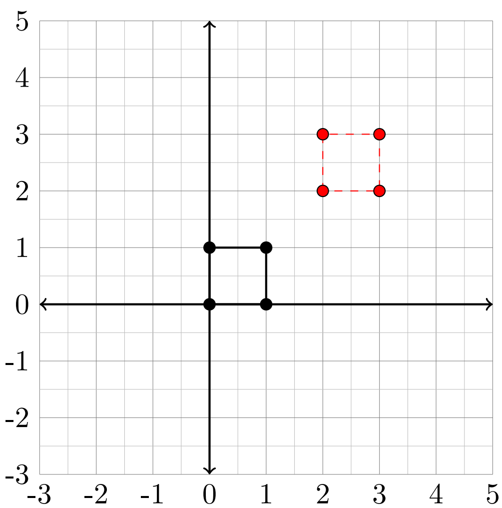
Figure 3: Applying an affine transformation [1]

In order to better understand what the T-Net is learning, we will walk through a short example of transformation matrices in two dimensions.

Suppose our input is given, and the input is a vertex (or vertices) in the PointNet. Let's say the transformation matrix is M. Take the matrix multiplication of M to the left side of the input, then we get the resulting transformed vertex point. Numerical examples for this is visualized in the figure below. In the 3 by 3 matrix below, the first two elements represent the (x, y) coordinate, and the third element represents the 'transformation value'. Check the given examples of translation and shearing of the unit square using a transformation matrix.

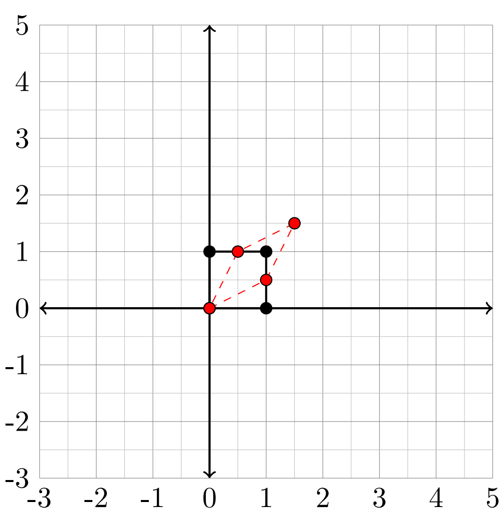
These examples only contain simple affine transformations, however spatial transformers networks like T-Net can learn many different transformations such as plane projective transformations, piece-wise affine, or thin plate spline [1].

Matrix Translation



- $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$

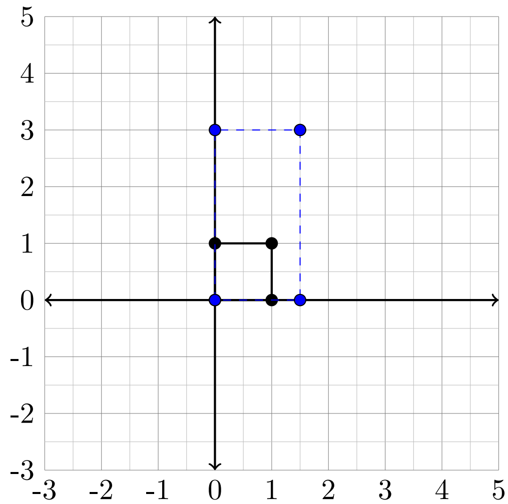
Matrix Shearing



- $\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1.5 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$

Question 3a. Calculate the transformed vertices and plot them on the graph.

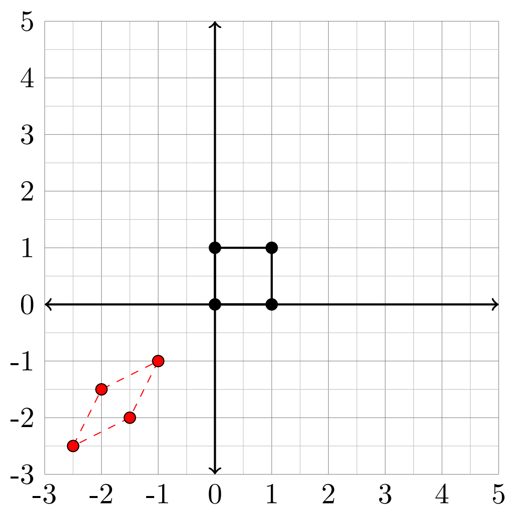
Matrix Scaling



- $\begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 3 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$

Question 3b. Calculate the transformation matrix to get the resulting vertices.

Matrix shear and translate



- $\begin{pmatrix} 1 & 0.5 & -2.5 \\ 0.5 & 1 & -2.5 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2.5 \\ -2.5 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & -2.5 \\ 0.5 & 1 & -2.5 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & -2.5 \\ 0.5 & 1 & -2.5 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1.5 \\ -2 \\ 1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0.5 & -2.5 \\ 0.5 & 1 & -2.5 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -1.5 \\ 1 \end{pmatrix}$

1.3 Robustness to Data Corruption

In the PointNet paper, the authors demonstrate that the model is very robust to data corruption. In Figure 4 we can see that the model still has high accuracy even with less data points. You can observe that even with half the data points missing, the model's accuracy drops by only 1-4%, which shows that the model is incredibly robust to data corruption!

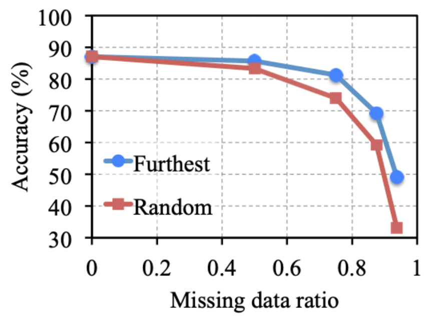


Figure 4: Missing data ratio vs. Accuracy [2]

Question 4. What architecture design choice makes the model so robust to data corruption? Why does this feature help with robustness? *Hint:* What did we do to obtain the 1024 global features?

(Sol) The max pooling function directly leads to the model robustness! This is because the max pooling only takes the critical points of the point cloud and uses those as the global features which are learned by the MLP. Losing out on points that are not critical points won't affect the model much at all.

2 Implementation of the PointNet Code

For the coding part, we will implement the classification and part segmentation models for the PointNet architecture. Follow the instructions in this Jupyter notebook file.
For the reviewers who hope to see the solution homework file, here is the Jupyter notebook solution. If the link is not working, please follow the instruction below:

1. Go to `https://github.com/Joohwan-Seo/CS282A_final_project`.
2. Download `pointnet`$_s$`ol.ipynb` and run the Jupyter notebook file.
3. The detailed description is provided in the Jupyter notebook file!

You need to implement PointNet architectures in `networks.py` and run the model in the Jupyter notebook. Figures 5, 6, and 7 in the next page may be useful to complete the tasks. The above Jupyter file will let you know when each figure is relevant to certain tasks.

Question 5. In the final part of the Jupyter notebook file, you plot the training loss and the validation accuracy. Attach that figure in the written part of the homework, together with the test accuracy.

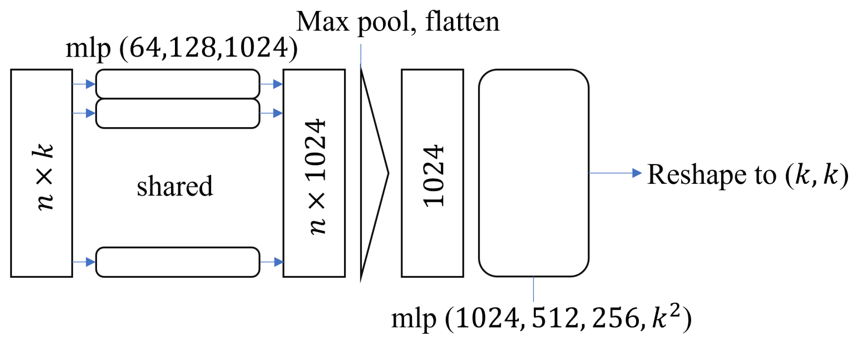


Figure 5: Architecture of T-Net used in this assignment (Drawn by ourselves; contents from [2]).

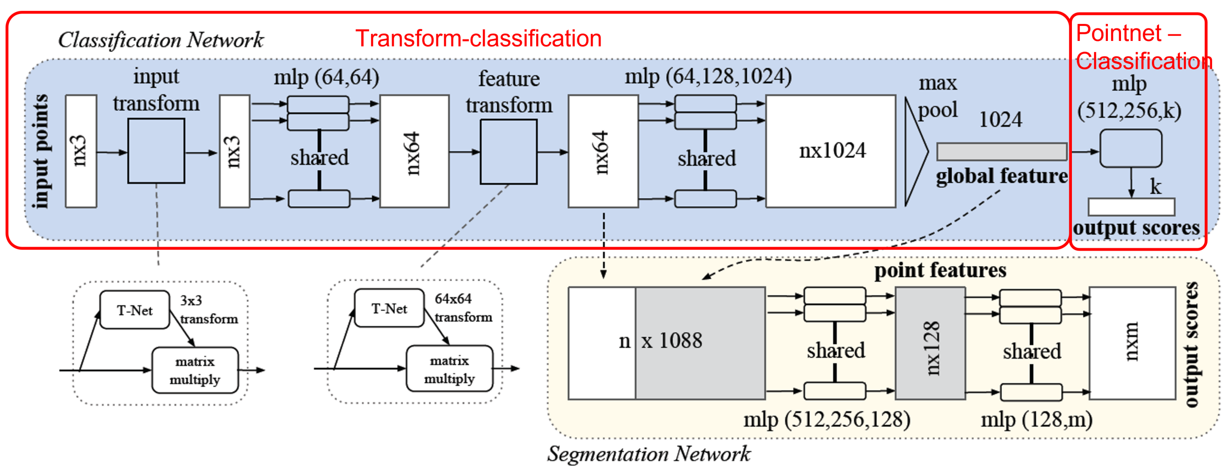


Figure 6: Architecture of classification network part used in this assignment [2].

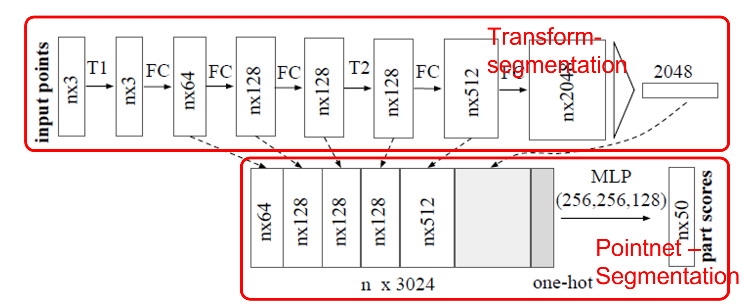


Figure 7: Architecture of segmentation network part used in this assignment [2].

3 Commentary on the homework

3.1 Motivation

The goal of the PointNet architecture identified in the paper is to create a novel architecture that can learn three-dimensional structures directly from their point cloud representations. Given that the data we want to learn from is a point cloud, which can also be called a point set, the unique characteristics of these sets become the primary driver for the design of the PointNet architecture. Two characteristics of the point sets that present challenges are that they are unordered point sets, so if you take a point cloud and permute it, it will still represent the same object, and that they are geometrically transformation invariant, meaning if we take a point cloud and flip it or rotate it, it will still represent the same object, just in a different orientation. Since these motivated the major architectural decisions, we focus our homework on these two aspects as well, with the hopes that they clearly reflect the design choices made by the authors of the paper.

3.2 Point sets are unordered

Since point sets are unordered, the neural network has to be permutation invariant, meaning it can still learn the point cloud representation regardless of the order of the points. This leads to permutation invariant functions. Question 1 engages with the concept of permutation invariant functions, where given multiple functions, we hope students will be able to choose and reason why each function is considered to be permutation invariant. This question is heavily inspired by previous questions in the course in which we looked at permutation invariant functions for graph neural networks, with aspects of weight sharing, addition, multiplication, and the max function. The goal is for students to fully understand what it means for a function to have permutation invariance and that permutation invariance can arise in multiple ways.

Question 2 continues on the concept of permutation invariance, linking the theory to a PyTorch function that students will see later in the coding portion of the homework. Question 1 shows that weight sharing is necessary for permutation invariance, and the `Conv.1d` function is an easy way to implement a weight-shared linear layer. This question invites students to think about how the `Conv.1d` function works in theory and how to set the parameters such that it represents a weight shared layer. The question also helps to set up why we use the `Conv.1d` function later on in the coding portion of the assignment.

3.3 Point sets are geometrically transformation invariant

Since the point set can be flipped or rotated while still representing the same object, we must ensure that our neural network architecture will be robust to such data transformations. If a chair is upside down, it is still very much a chair and should be classified as such. One way to ensure that the model is invariant to geometric transformations is to use a transformation matrix to transform the data back into a state where it can be learned. Continuing with the chair analogy, if a chair is on its side, we put it back up, and if a chair is upside down, we flip it back up before we can sit on it. This action of moving the chair back into its original orientation is the role of the transformation matrix.

However, this transformation matrix shouldn't be predetermined or treated as a hyperparameter we can tune; rather, it should be learned by the network during training. This presents the role of the T-Net, a simple deep neural network used to learn an effective transformation matrix. We want students to engage with this idea of the transformation matrix and how it can be used to make the model robust to geometric transformations. Since it is difficult to compute such matrices in three dimensions, we reduce this to a toy problem in two dimensions, where students work through some hand calculations for transforming a unit square with some transformation matrices. Question 3 addresses just that. It provides two examples for students to see how the transformation matrix works and how it affects the shape and orientation of the unit square. Question 3a has students compute the resulting shape of the unit square after a scaling transformation, while Question 3b has students work backward to find the transformation matrix responsible for the transformation from the unit square into some shape.

While there is some conceptual understanding of what the T-Net is doing in theory, we wanted to also show the role of the T-Net in the coding portion of the assignment, particularly its role in providing better validation performance of the model. To investigate the role of the T-Net, we have students train different versions of the PointNet model, one with the T-Net module and one without, and compare their performance by plotting the validation curves. Ideally, the model with the T-Net module will outperform the model that does not contain the T-Net module, highlighting the effectiveness and merits of the choice to add T-Net to the PointNet architecture.

3.4 The Data

Apart from focusing just the T-Net module, we also have a portion of the coding assignment dedicated to sampling the data, that is, to see how these point sets were generated from a three-dimensional mesh representation of an object. Since three-dimensional objects can be represented in multiple ways, synthetic datasets of 3D objects are typically done in a mesh form. For our case, we had to sample points from these mesh surfaces to construct the point clouds, which are then trained. The paper presents various methods for sampling points from the mesh surfaces of the objects, and so we present two methods for sampling the data. Random sampling from the surfaces and using farthest point sampling, where we find the centroid of the mesh surface and sample the farthest point from it. These methods directly affect the model's performance, so we wanted students to see this by having them train multiple models. Such models use the randomly sampled data as an input and the farthest point data as an input. Again we visualize the performance by comparing their validation curves.

We also have students implement a transformation function to slightly augment the point cloud data to simulate better real-world conditions where the data might not be perfectly orientated. This helps students learn how to transform and work with point cloud data.

We also have Question 4, based on the model's robustness to data corruption. This was pointed out as one of the strengths of the model in comparison to other models that learn 3D data. Given the choice of various permutation invariant functions, we present this question in the case that some students may wonder why the authors chose to use max pooling. The model robustness highlights the choice of the max pooling function since max pooling only learns the critical points of the data so that noncritical functions can be corrupt, and the model will still be able to make accurate classifications. The question helps students think critically about the author's architecture choices and the benefits of max pooling for learning point clouds.

3.5 Model layers and training

Apart from these key points of permutation invariance, geometric transformation invariance, and sampling and transforming the data, we have functions in the Jupyter Notebook and Python files for students to fill in to transfer their understanding of the PointNet architecture design into PyTorch code. Such examples include implementing the different layers of the neural network, along with the forward pass, filling in the loss function, and implementing the training function for the model. These concepts are well tested in the class since we have coded these portions for most homework. However, we still felt it was essential to understand the different layers used, including the fully connected layers, convolutional layers, and batch norm layers and their respective dimensions.

References

[1] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks, 2016.

[2] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

4 Part IV. Team members contributions

Our team works in very cooperative manner in most of the tasks in the project. All individuals have equal contributions.

- After the first few initializing stages of the project, all members work together.
- Arm completed the first version of the written assignment.
- Seunghoon completed the first building draft of the code.
- Joohwan completed data preprocessing and the first draft of the code.

5 Part V. Codes, data, and supplementary

- GitHub repository: `https://github.com/Joohwan-Seo/CS282A_final_project`.

We included the links to the GitHub repository and the Jupyter notebook file (which is also contained in the GitHub repository) in the written assignment as well to make the homework self-contained. The data is preprocessed by us, and it is included in the GitHub repository.

The LaTeX zip file is also included in the GitHub.