ViewNet: A Novel Projection-Based Backbone with View Pooling for Few-shot Point Cloud Classification

Jiajing Chen, Minmin Yang, Senem Velipasalar
{jchen152, myang47, svelipas}@syr.edu

Electrical Engineering and Computer Science Department
Syracuse University

THU-AM-112

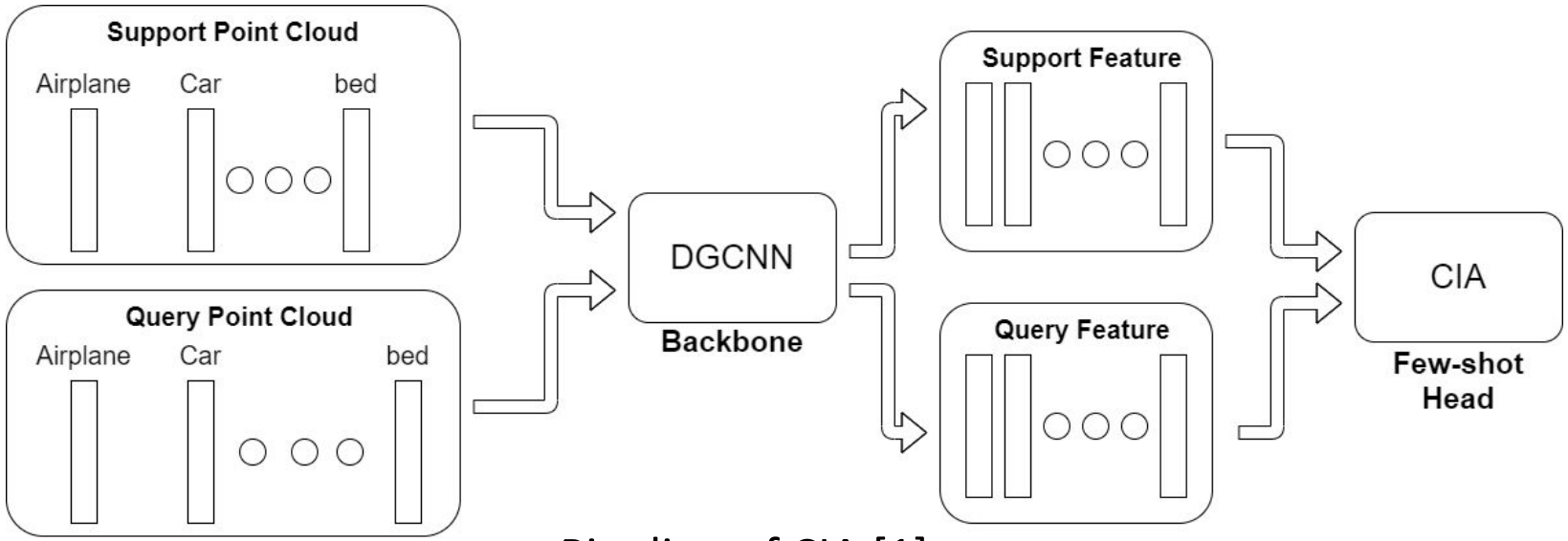
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Summary

- We propose a robust and effective 2D projection-based backbone, referred to as ViewNet, for few-shot point cloud classification
- We first provide an analysis of commonly used 3D point-based backbones in terms of point utilization, and argue that they are not well-suited for the few-shot learning task, especially with real-word point clouds obtained via scanning
- After visualizing projected depth images of point clouds, we observe that some projections are robust to missing points and deformations
- Motivated by above, we present ViewNet, which uses six different depth images that are projections of a point cloud, and employs our proposed View Pooling to generate more descriptive and distinguishing features
- ViewNet achieves SOTA performance on ScanObjectNN, ModelNet40-C and ModelNet40 datasets, and outperforms four different baselines on few-shot point cloud classification
- Ablation studies show that the ViewNet backbone can generalize and be employed together with different few-shot prediction heads, providing better performance than point-based backbones

State-of-the-art (SOTA)

- Cross-Instance Adaptation (CIA) [1] is a SOTA network, which uses a novel few-shot head with attention mechanism, referred to as CIA, for few-shot point cloud classification.
- In [1], different point cloud analysis networks [2-6] are compared as backbones, and it was shown that the best performance is achieved when DGCNN [6] is used as the backbone.
- In this work, we first show that point-based methods, such as DGCNN [6], are not the most suitable backbones for few-shot point cloud classification task, and then propose ViewNet.

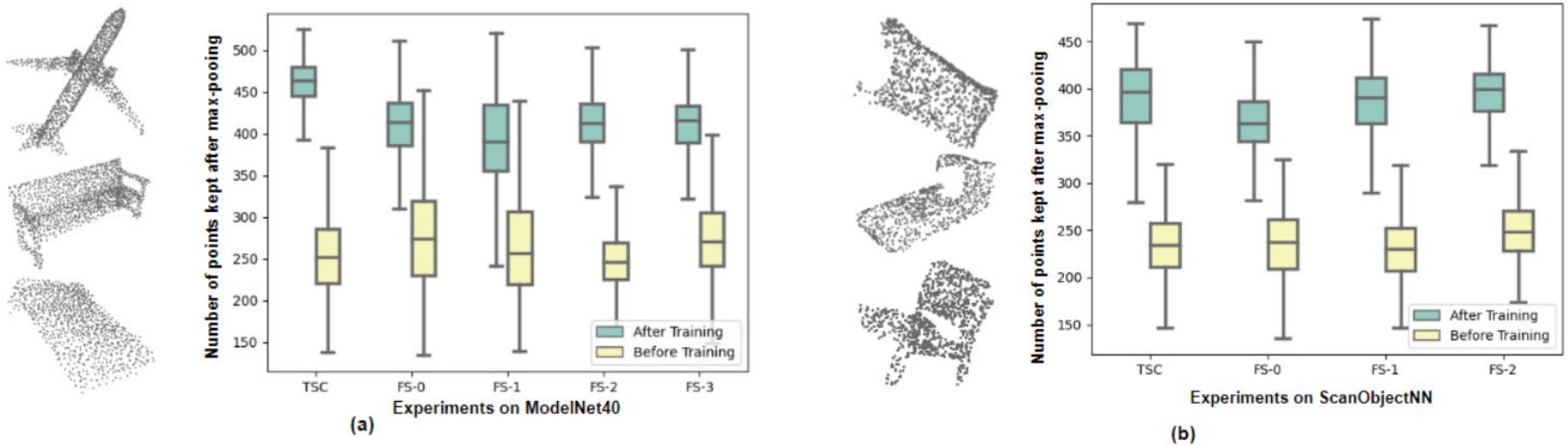


Pipeline of CIA [1]

Motivation - I

Point Utilization Analysis

- In our previous work [7], we had shown that a large number of points are discarded during max pooling.
- In this study, we analyze the number of points utilized by DGCNN, for both traditional supervised classification and few-shot classification, on ModelNet40 [8] and ScanObjectNN [9] datasets. For few-shot classification, we use ProtoNet [10] to perform the experiments.



TSC refers to traditional supervised classification, and FS-n represents few-shot point cloud classification experiment at fold n. The number of input points is 1024 for all the experiments.

Motivation - I

Point Utilization Analysis Cont'd

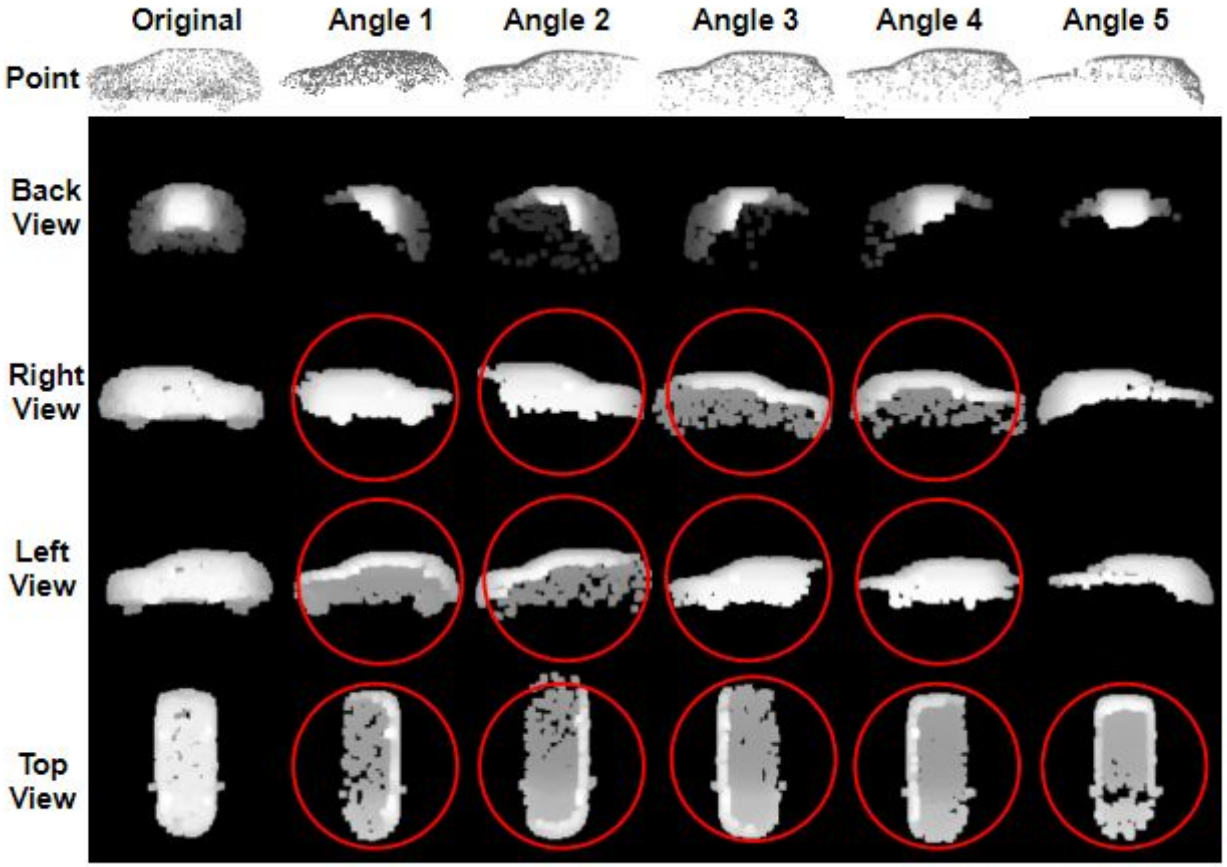
Dataset	Experiment Name	MED of no. of kept pnts	Accuracy
ModelNet40	TSC	252→464	92.51%
ModelNet40	FS-0	274→414	89.97%
ModelNet40	FS-1	257→390	83.46%
ModelNet40	FS-2	246→413	74.08%
ModelNet40	FS-3	271→416	76.13%
ScanObjectNN	TSC	234→397	83.10%
ScanObjectNN	FS-0	237→363	50.58%
ScanObjectNN	FS-1	230→391	62.17%
ScanObjectNN	FS-2	248→400	62.59%

TSC is the traditional supervised point cloud classification, and FS-n is few-shot point cloud classification at fold n. a → b shows the median value of the number of utilized points before and after training, respectively.

- With the number of points utilized for few-shot classification being less than that for TSC, it is hard to expect DGCNN to extract the best set of features to describe 3D objects.
- Compared to ModelNet40 dataset, DGCNN uses less points on ScanObjectNN, and provides lower accuracy.
- Thus, it can be inferred that missing points and deformed shapes can negatively affect max-pooling, causing to pick up inadequate points to represent a 3D shape.

Motivation - II

Point Projection Analysis



- The first row: a point cloud sampled from a CAD model (referred to as Original), and point clouds with simulated missing points seen from five different angles.
- Rows 2 - 5: projection images on different planes
- Although the missing portion of the point cloud can be different, some projection images can provide robustness to varying occlusions

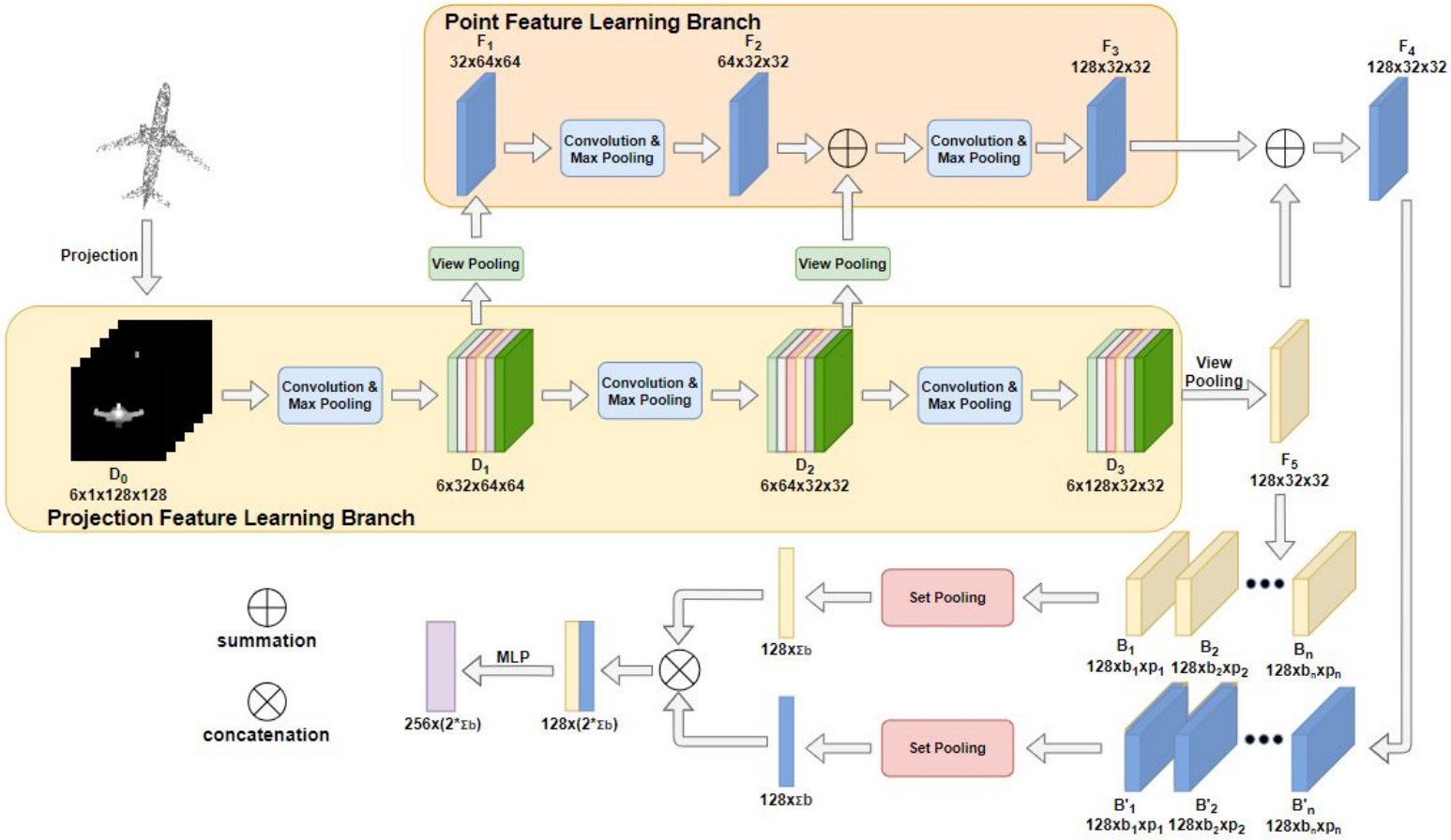
Motivation - III

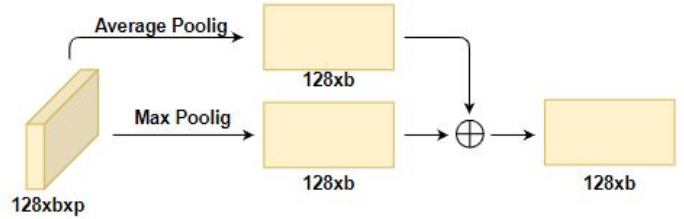
- Although projection-based approach [12] has been used together with ResNet for supervised point cloud classification, a traditional image classification backbone, such as ResNet, is not the most suitable for few-shot classification.
- Traditional CNN-based backbones are composed of convolution layers and process all depth images separately, without a module for extracting distinguishing features among all views' feature maps.

	Model	fold 0	fold 1	fold 2	fold 3	Mean
5-way 1-shot	DGCNN+ProtoNet	**85.42%**	**79.46%**	**70.06%**	70.73%	**76.42%**
	ResNet+ProtoNet	83.29%	79.35%	64.44%	**74.42%**	75.38%%
5-way 5-shot	DGCNN+ProtoNet	**93.99%**	**88.65%**	**84.76%**	85.56%	**88.24%**
	ResNet+ProtoNet	92.61%	87.39%	80.91%	**86.96%**	86.97%%

Comparison of ProtoNet's performance on ModelNet40, with DGCNN and ResNet as backbones, for 5-way 1-shot and 5-way 5-shot classification.

Proposed ViewNet



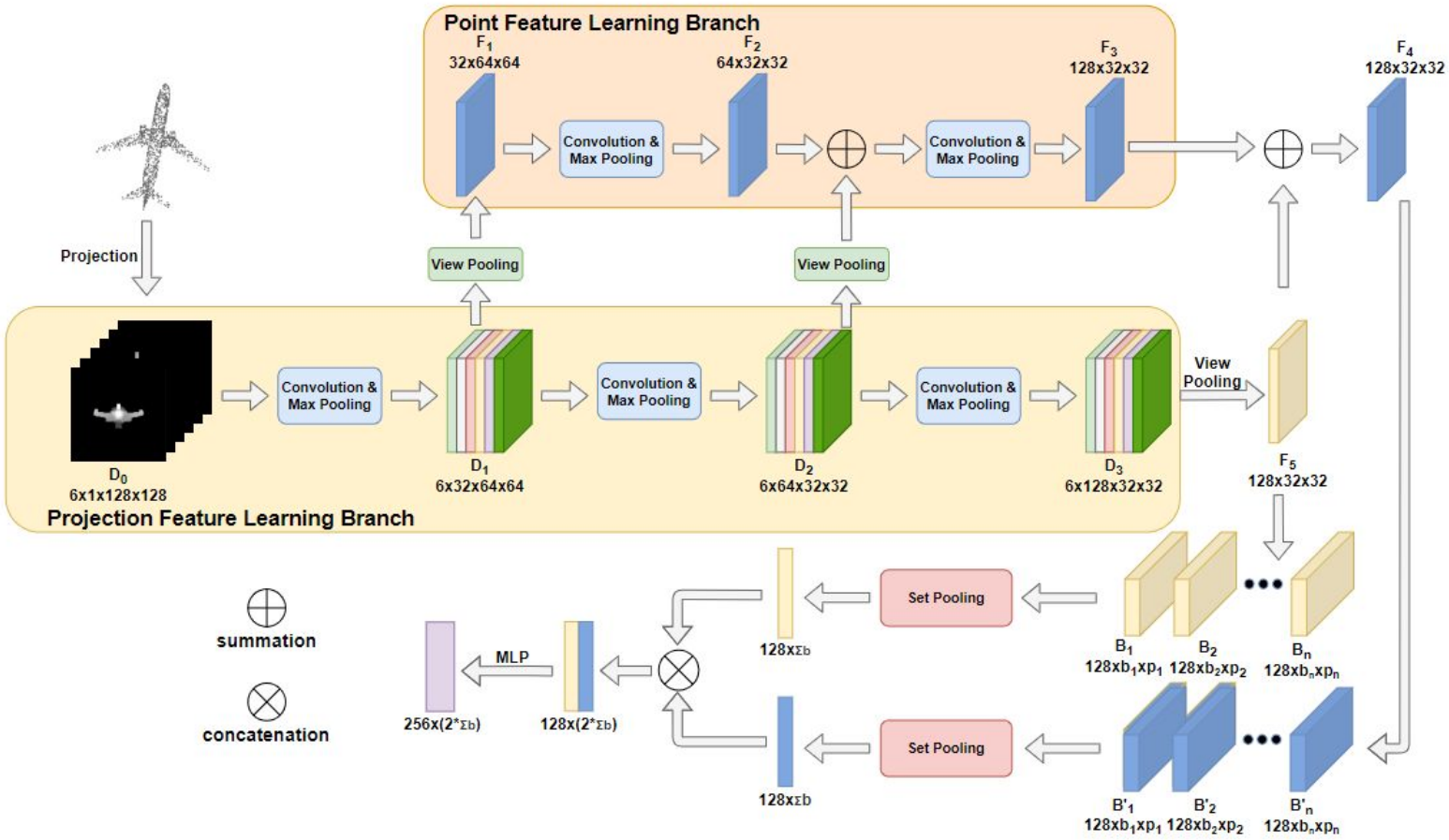


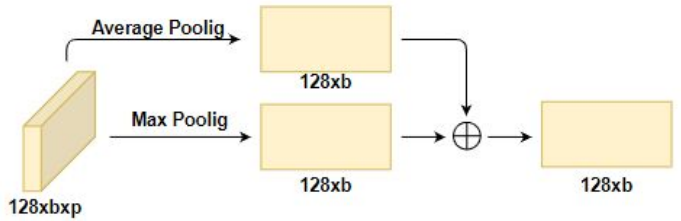
Set Pooling. b is the number of bins, p is the number of pixels in each bin. The pooling is performed along the pixel dimension

- In Projection Feature Learning branch, convolution and max pooling are used to process each depth image independently, and obtain intermediate feature maps $\{Di|i \in \{1, 2, 3\}\}$ for our proposed View Pooling.

- View Pooling extracts more descriptive and distinguishing features from different combinations of views, which are then fed into the Point Feature Learning Branch for further processing

Proposed ViewNet

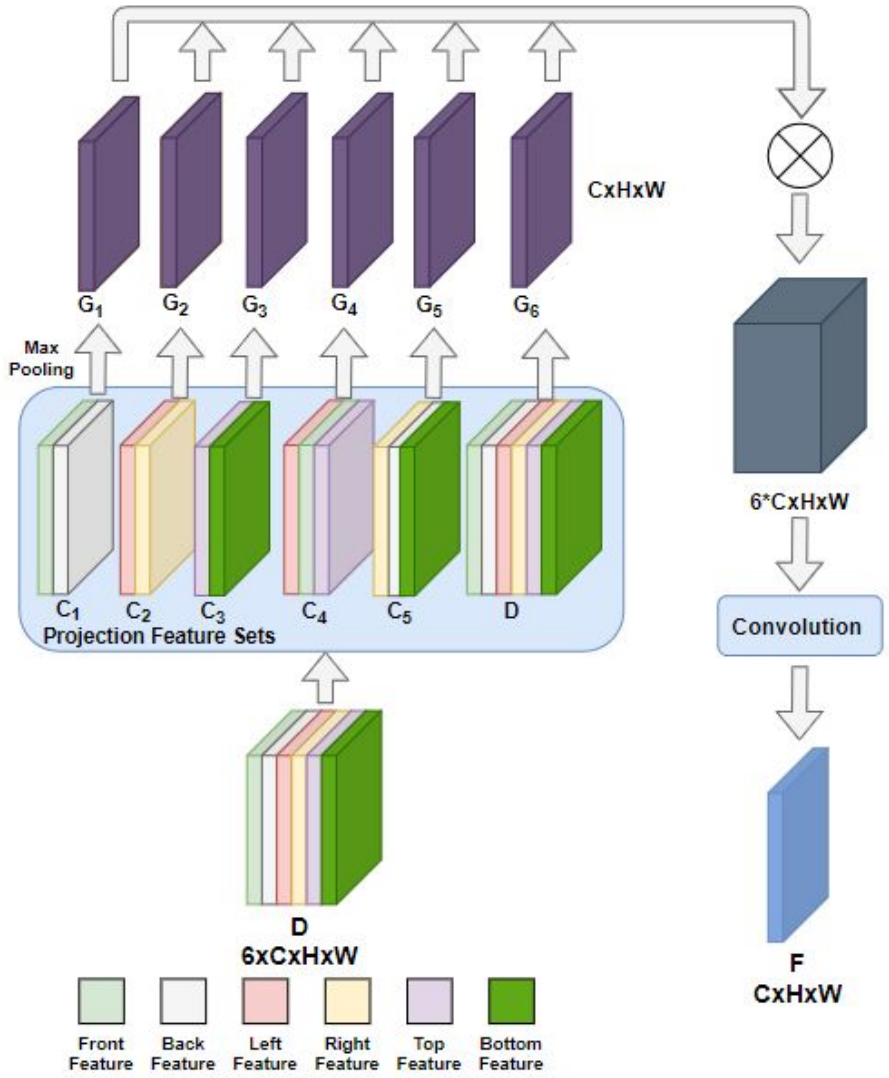




Set Pooling. b is the number of bins, p is the number of pixels in each bin. The pooling is performed along the pixel dimension

- The Point Feature Learning Branch learns the features describing the point cloud's shape from the feature maps of all six projections
- To learn discriminative features in different receptive fields, the pixels in feature maps F4 and F5 are divided into n-many bins
- Each bin's feature is fed into few-shot head, and the bin's loss is obtained. Final loss is computed by summing up n losses

View Pooling



- Five new feature maps (Ci) are obtained from six view features in D by using them in different combinations
- Three of the five feature maps are obtained by taking all pairs of opposite projection feature maps into account, i.e. {(left, right), (front, back), (top, bottom)}. If a small part of points is missing, it is likely that at least one projection image in this combination of opposite projections is not affected greatly
- Remaining two of the five feature maps come from triplet combinations {(left, front, top), (right, back, bottom)}, since these three-view drawings are able to depict an object's 3D shape

Experiments on ModelNet40

- ModelNet40 [8] is a commonly used point cloud dataset, which contains 12,311 CAD models of 40 man-made object categories
- We sort those classes by their ID in an ascending order, and divide 40 classes into 4 folds to perform 4-fold cross validation experiments
- Our method outperforms all baselines in terms of mean accuracy for both 1-shot and 5-shot classification

		fold 0	fold 1	fold 2	fold 3	Mean
5-way 1-shot	MetaOpt	82.87±0.72	75.77±0.83	65.31±0.92	66.97±0.93	72.73±0.85
	RelationNet	82.14±0.69	77.46±0.80	66.09±0.91	69.47±0.84	75.23±0.81
	ProtoNet	85.42±0.64	79.46±0.76	70.06±0.39	70.73±0.42	76.42±0.55
	CIA	89.97±0.63	**83.46±0.83**	74.08±0.95	76.13±0.86	80.91±0.82
	Ours	**92.57±0.52**	82.68±0.80	**75.28±0.90**	**80.95±0.75**	**82.87±0.74**
5-way 5-shot	MetaOpt	92.37±0.38	86.44±0.62	82.10±0.58	83.15±0.55	86.02±0.53
	RelationNet	91.53±0.38	85.11±0.61	79.36±0.63	83.01±0.52	84.75±0.53
	ProtoNet	93.99±0.29	88.65±0.54	84.76±0.51	85.56±0.48	88.24±0.45
	CIA	94.61±0.30	89.15±0.55	85.00±0.51	86.71±0.50	88.87±0.47
	Ours	**96.23±0.26**	**89.64±0.55**	**85.74±0.51**	**90.18±0.45**	**90.45±0.44**

Experiments on ModelNet40-C

- ModelNet40-C [11] contains the point clouds of the same 40 classes as ModelNet40, and different from ModelNet40, ModelNet40-C contains point clouds with different types of corruption to simulate real-world scenarios
- Again 4 folds are used for cross validation
- Our method outperforms all the baselines for each fold and for both 1-shot and 5-shot classification and improvement margins are larger compared to ModelNet40 dataset

		fold 0	fold 1	fold 2	fold 3	Mean
5-way 1-shot	Metaopt	78.28±0.79	75.34±0.84	58.07±0.86	66.29±0.91	69.50±0.85
	RelationNet	79.59±0.74	74.63±0.84	59.03±0.81	68.38±0.86	70.41±0.81
	ProtoNet	81.29±0.71	75.83±0.79	61.76±0.84	69.83±0.84%	72.18±0.80
	CIA	85.70±0.75	79.67±0.90	65.68±1.0	74.32±0.94	76.34±0.89
	Ours	**89.47±0.58**	**81.05±0.78**	**69.56±0.89**	**76.29±0.85**	**79.09±0.78**
5-way 5-shot	Metaopt	91.09±0.40	84.19±0.57	75.10±0.73	81.34±0.53	82.93±0.56
	RelationNet	87.12±0.46	83.55±0.54	70.18±0.78	79.01±0.58	79.97±0.59
	ProtoNet	90.97±0.39	86.21±0.50	76.99±0.65	83.19±0.51	84.34±0.51
	CIA	92.07±0.36	86.81±0.56	76.11±0.71	83.71±0.51	84.68±0.54
	Ours	**94.95±0.31**	**88.75±0.49**	**81.53±0.60**	**86.78±0.46**	**88±0.47**

Experiments on ScanObjectNN

- ScanObjectNN [9] dataset contains 15k objects from 15 categories
- Since the point clouds are scanned from real world objects, missing points due to occlusion frequently occur, posing greater challenges
- The classes are sorted in ascending order based on their ID, and then evenly divided into 3 folds for 3-fold cross validation
- Improvement margins are highest on this dataset showing that our proposed method is also effective on point clouds scanned from the real-world

		fold 0	fold 1	fold 2	Mean
5-way 1-shot	MetaOpt	41.92±0.72	61.12±0.66	53.87±0.78	52.30±0.72
	RelationNet	50.29±0.76	54.23±0.63	51.45±0.64	51.99±0.68
	ProtoNet	50.81±0.73	60.46±0.67	58.72±0.78	56.66±0.73
	CIA	50.58±0.82	62.17±0.68	62.59±0.74	58.45±0.75
	Ours	**60.90±0.76**	**66.48±0.60**	**64.10±0.77**	**63.83±0.71**
5-way 5-shot	MetaOpt	63.86±0.56	67.73±0.45	70.19±0.49	67.26±0.50
	RelationNet	58.65±0.53	66.72±0.50	65.94±0.52	63.77±0.52
	ProtoNet	68.42±0.54	70.20±0.52	68.76±0.49	69.13±0.52
	CIA	62.94±0.51	71.31±0.45	70.21±0.48	68.15±0.48
	Ours	**73.66±0.48**	**74.77±0.45**	**77.46±0.46**	**75.3±0.46**

Ablation Studies

We perform ablation studies to analyze:

- Effectiveness of bin-wise loss
- ViewNet's Generalizability as a backbone
- Effectiveness of View Pooling

		fold 0	fold 1	fold 2	Mean
5-way 1-shot	DGCNN+MetaOpt	41.92%	61.12%	53.87%	52.3%
	ViewNet+MetaOpt	**48.74% (↑6.82%)**	**61.62% (↑0.5%)**	**58.95% (↑5.08%)**	**56.44% (↑4.14%)**
	DGCNN+RelationNet	50.29%	54.23%	51.45%	51.99%
	ViewNet+MetaOpt	**55.73% (↑5.44%)**	**60.32% (↑6.09%)**	**59.10% (↑7.65%)**	**58.38% (↑6.39%)**
	DGCNN+ProtoNet	50.81%	60.46%	58.72%	56.66%
	ViewNet+ProtoNet	**56.02% (↑5.21%)**	**64.06% (↑3.6%)**	**64.05% (↑5.33%)**	**61.37% (↑4.71%)**
	DGCNN+CIA	50.58%	62.17%	62.59%	58.45%
	ViewNet+CIA	**60.81% (↑10.23%)**	**65.84% (↑3.67%)**	**64.19% (↑1.6%)**	**63.61% (↑5.16%)**
5-way 5-shot	DGCNN+MetaOpt	63.86%	67.73%	70.19%	67.26%
	ViewNet+MetaOpt	**67.97% (↑4.11%)**	**73.04% (↑5.31%)**	**75.12% (↑4.93%)**	**72.04% (↑4.78%)**
	DGCNN+RelationNet	58.65%	**66.72%**	65.94%	63.77%
	ViewNet+RelationNet	**67.49% (↑8.84%)**	66.51% (↓0.21%)	**72.01% (↑6.08%)**	**68.67%(↑4.9%)**
	DGCNN+ProtoNet	68.42%	70.2%	68.76%	69.13%
	ViewNet+ProtoNet	**75.13% (↑6.71%)**	**74.41% (↑4.21%)**	**77.07% (↑8.31%)**	**75.54% (↑6.41%)**
	DGCNN+CIA	62.94%	71.31%	70.21%	68.15%
	ViewNet+CIA	**72.69% (↑9.75%)**	**73.56% (↑2.25%)**	**75.33% (↑5.12%)**	**73.86% (↑5.71%)**

ViewNet's Generalizability as a Backbone

	Backbone Feature	fold 0	fold 1	fold 2	Mean
5-way 1-shot	O'	57.29%	64.47%	62.52%	61.43%
	O	**60.90%**	**66.48%**	**64.10%**	**63.83%**
5-way 5-shot	O'	73.28%	74.41%	75.42%	74.37%
	O	**73.66%**	**74.77%**	**77.46%**	**75.30%**

Analysis of the Bin-wise Loss

	View Pooling Type	fold 0	fold 1	fold 2	Mean
5-way 1-shot	Without C_i	**60.98%**	63.41%	63.81%	62.73%
	With C_i	60.90%	**66.48%**	**64.10%**	**63.83%**
5-way 5-shot	Without C_i	73.14%	72.76%	75.85%	73.92%
	With C_i	**73.66%**	**74.77%**	**77.46%**	**75.30%**

Analysis of View Pooling

References

[1] Ye, Chuangguan, et al. "What Makes for Effective Few-shot Point Cloud Classification?." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
[2] Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[3] Qi, Charles Ruizhongtai, et al. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." Advances in neural information processing systems 30 (2017).
[4] Liu, Yongcheng, et al. "Relation-shape convolutional neural network for point cloud analysis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
[5] Liu, Yongcheng, et al. "Densepoint: Learning densely contextual representation for efficient point cloud processing." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[6] Wang, Yue, et al. "Dynamic graph cnn for learning on point clouds." Acm Transactions On Graphics (tog) 38.5 (2019): 1-12.
[7] Chen, Jiajing, et al. "Why Discard if You Can Recycle?: A Recycling Max Pooling Module for 3D Point Cloud Analysis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[8] Wu, Zhirong, et al. "3d shapenets: A deep representation for volumetric shapes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[9] Uy, Mikaela Angelina, et al. "Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[10] Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical networks for few-shot learning." Advances in neural information processing systems 30 (2017).
[11] Sun, Jiachen, et al. "Benchmarking robustness of 3d point cloud recognition against common corruptions." arXiv preprint arXiv:2201.12296 (2022).
[12] Goyal, Ankit, et al. "Revisiting point cloud shape classification with a simple and effective baseline." International Conference on Machine Learning. PMLR, 2021.

Thanks!

Questions?

jchen152@syr.edu