A. Appendix

A.1. PointNeXt Backbone Experiments

PointNeXt [37] is a concurrent work which proposes a lightweight backbone based on PointNet++ and in particularly it gives promising results on the ScanObjectNN benchmark. In order to demonstrate the effectiveness of our ULIP on this most recent backbone, we pre-train PointNeXt using ULIP, and use the pre-trained weights to finetune on the ScanObjectNN dataset.

As shown in Table 7, ULIP significantly improves PointNeXt in both Overall Accuracy and Class-mean Accuracy.

Model	Overall Acc	Class-mean Acc
PointNeXt* [37]	87.4	85.8
PointNeXt + ULIP	**89.2** (↑ 1.8)	**88.0** (↑ 2.2)
PointNeXt †*	87.5	85.9
PointNeXt †+ ULIP	**89.7** (↑ 2.2)	**88.6** (↑ 2.7)

Table 7. 3D classification results on ScanObjectNN for PointNeXt. †indicates a model uses 2K sampled points and all others use 1K sampled points. * indicates it's reproduced result.

A.2. Details of Evaluation Sets in Zero Shot Classification

When evaluating zeroshot classification, we notice that there are some common classes between our pre-train dataset, ShapeNet55, and ModelNet40. Evaluations on these common classes might introduce an unfair comparison of zeroshot performance. Therefore, we introduced three different validation sets for evaluating our models and our baselines on ModelNet40.

All Set: Includes all the categories in ModelNet40 as shown in Table 8.

airplane	bathtub	bed	bench	bookshelf
bottle	bowl	car	chair	cone
cup	curtain	desk	door	dresser
flower_pot	glass_box	guitar	keyboard	lamp
laptop	mantel	monitor	night_stand	person
piano	plant	radio	range_hood	sink
sofa	stairs	stool	table	tent
toilet	tv_stand	vase	wardrobe	xbox

Table 8. ModelNet40 All Set.

Medium Set: We remove categories whose exact category names exist in our pre-training dataset. The resulting categories in this set is shown in Table 9.

cone	cup	curtain	door	dresser
glass_box	mantel	monitor	night_stand	person
plant	radio	range_hood	sink	stairs
stool	tent	toilet	tv_stand	vase
wardrobe	xbox			

Table 9. ModelNet40 Medium Set.

Hard Set: We remove both extract category names and their synonyms in our pre-training dataset. The final *Hard Set* is shown in Table 10

cone	curtain	door	dresser	glass_box
mantel	night_stand	person	plant	radio
range_hood	sink	stairs	tent	toilet
tv_stand	xbox			

Table 10. ModelNet40 Hard Set.

A.3. Indoor 3D Detection Experiments

In order to show our potential on 3D scene applications, we conduct experiments on ScanNet-v2 dataset and benchmark 3D detection performance based on one of SOTA 3D detection frameworks, Group-Free-3D [27]. In our setting, we use the Group-Free-3D basic model and observe significant improvements as shown in Table 11.

	mAP@0.5	mAP@0.5 Averaged
Group-Free-3D	48.9	48.4
Group-Free-3D + ULIP	50.2 (↑ 1.3)	49.6 (↑ 1.2)

Table 11. Experiments on indoor 3D Detection. We use Group-Free-3D basic model as our detection framework, and we follow the same metric computation as in [27].