3D INDOOR SCENE LONG TAIL SEGMENTATION

DLCV FINAL PROJECT TEAM - PeiyuanWu

R11942152 WANG, YU
R10921A16 LI, YING-SHUO
R10942198 LIN, CHUNG-WEI
R10942147 HUANG, CHUN-KAI
R11946012 WANG, YI-FANG

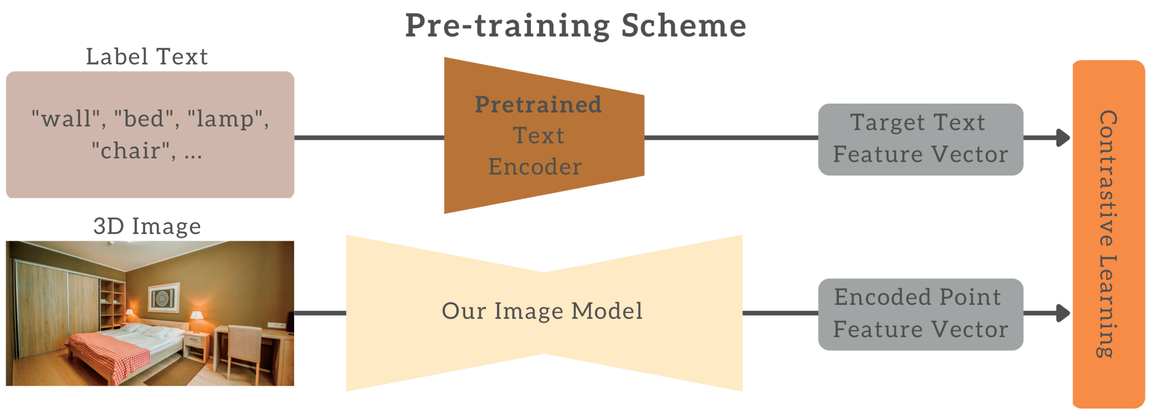
MAIN CHALLENGE

ScanNet200 has much more classes and class imbalance issues. Since the performance of previous work has poor performance on tail categories of ScanNet200, we follow the work of the Language-Grounded 3D Semantic Segmentation model to deal with the class imbalance issue.

DATA PREPROCESS

Language-Grounded Contrastive Learning

The previous works in 3D semantic segmentation introduce augmentation alternatives for 3D pre-training. These methods will lead to low precision on tail categories. In contrast to this work, our model introduces CLIP, a powerful zero-shot classification language model, to learn a more robust feature representation space with the contrastive learning method.



Class-Balanced Loss

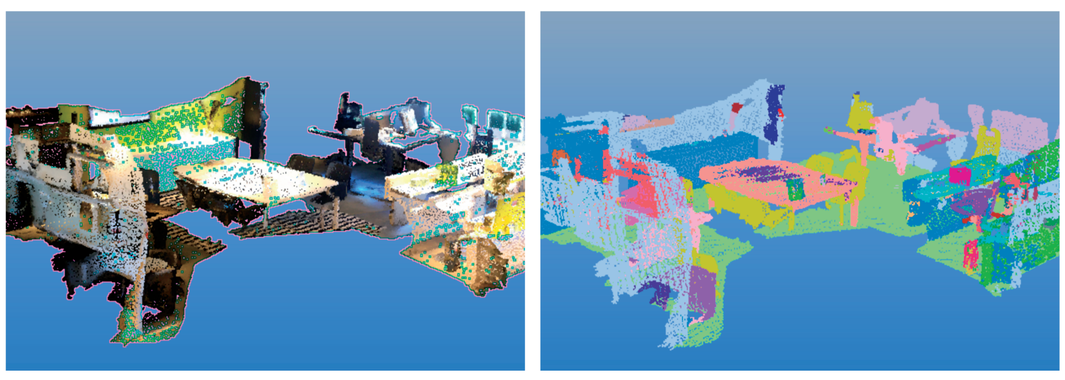
The focal loss has shown that it can better deal with the less frequency or high misclassification example compared to the cross-entropy criterion. Recently, focal loss has also improved the performance of class-balanced 3D semantic segmentation tasks. We adapt the original focal loss by multiplying with a weight based on the number of voxels of the train set:

$$FL\left(p_t\right) = -\alpha_i \left(1-p_t\right)^{\gamma} \log\left(p_t\right), \ \alpha_t = \frac{\log(n_i)}{\sum_{j=1}^{N_{class}} \log(n_j)}$$

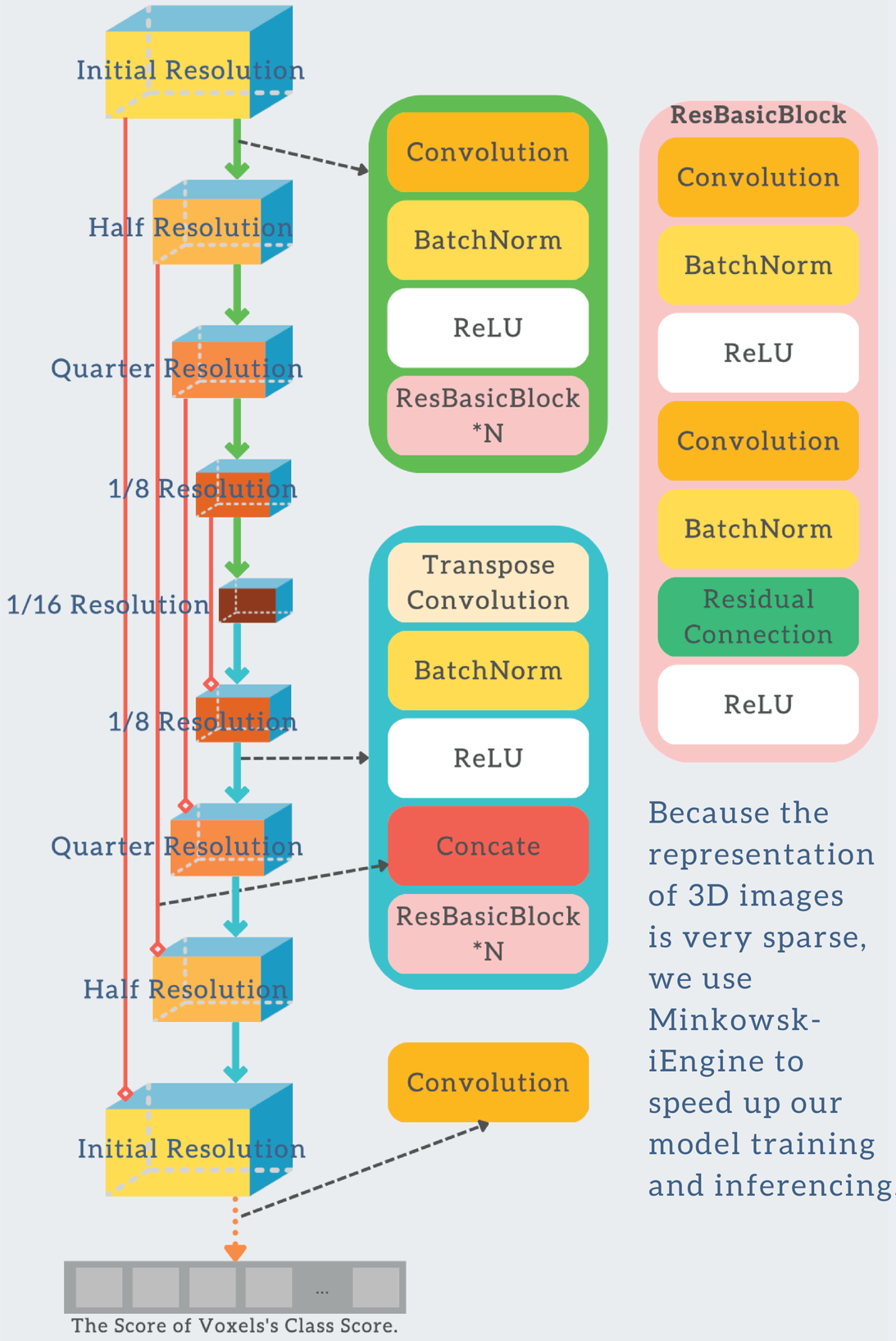
where n_i is the number of voxels of the i-th category in the train set.

Re-Balancing Instance Augmentation

Since rare classes and small objects do not have enough voxels, they will be overfitted on the specific surrounding context in the training set. We introduce an augmentation by placing instances of infrequently seen categories in scenes to break the dependencies of the surrounding context.



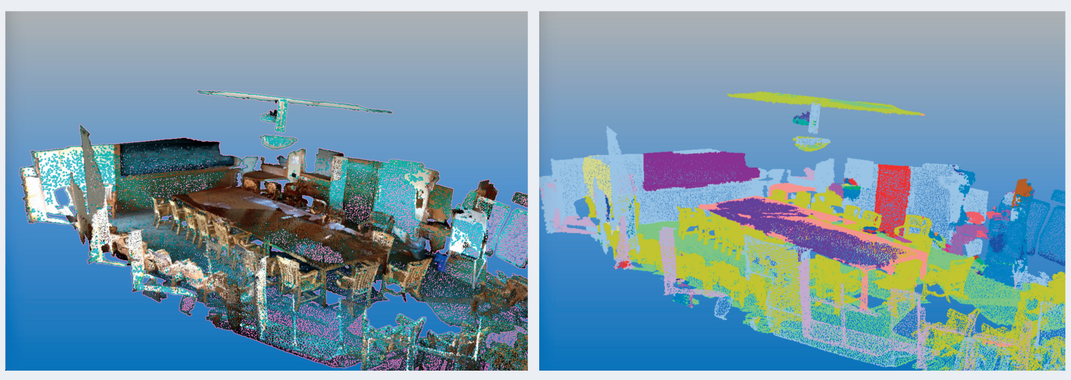
MODEL ARCHITECTURE



Because the representation of 3D images is very sparse, we use Minkowsk-iEngine to speed up our model training and inferencing.

EXPERIMENT

Pretraining	Loss	Augmentation	Test mIoU
No Pretrained	CrossEntropy	w/o	7.708
Pretrained	CrossEntropy	w/o	10.473
Pretrained	CrossEntropy	w/	
Pretrained	Class-balanced Focal Loss	w/o	
Pretrained	Class-balanced Focal Loss	w/	19.760



Reference: https://github.com/RozDavid/LanguageGroundedSemseg.git