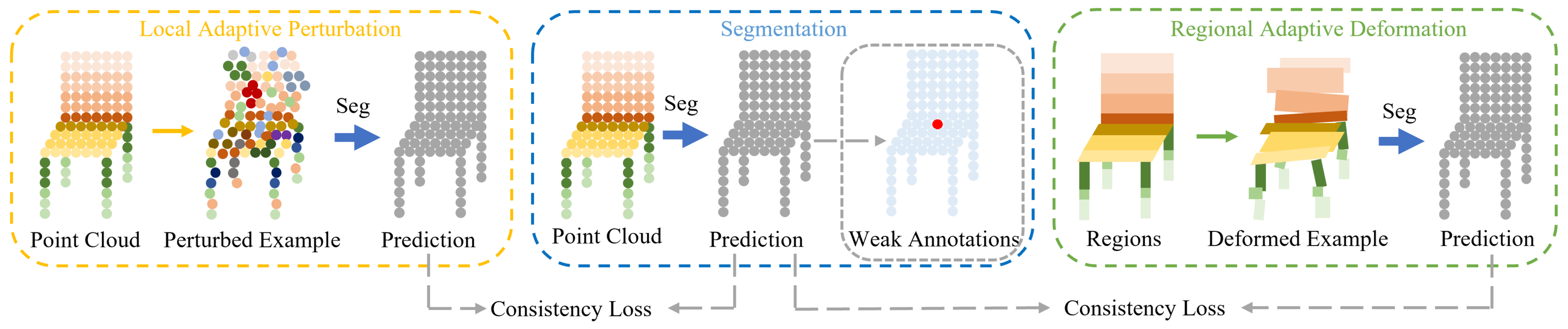
Dual Adaptive Transformations for Weakly Supervised Point Cloud Segmentation

Zhonghua Wu [1,2], Yicheng Wu [3], Guosheng Lin [1,2], Jianfei Cai [2,3], and Chen Qian [4]
[1]S-lab, NTU [2]School of Computer Science and Engineering, NTU
[3]Dept of Data Science and AI, Monash University [4]SenseTime Research

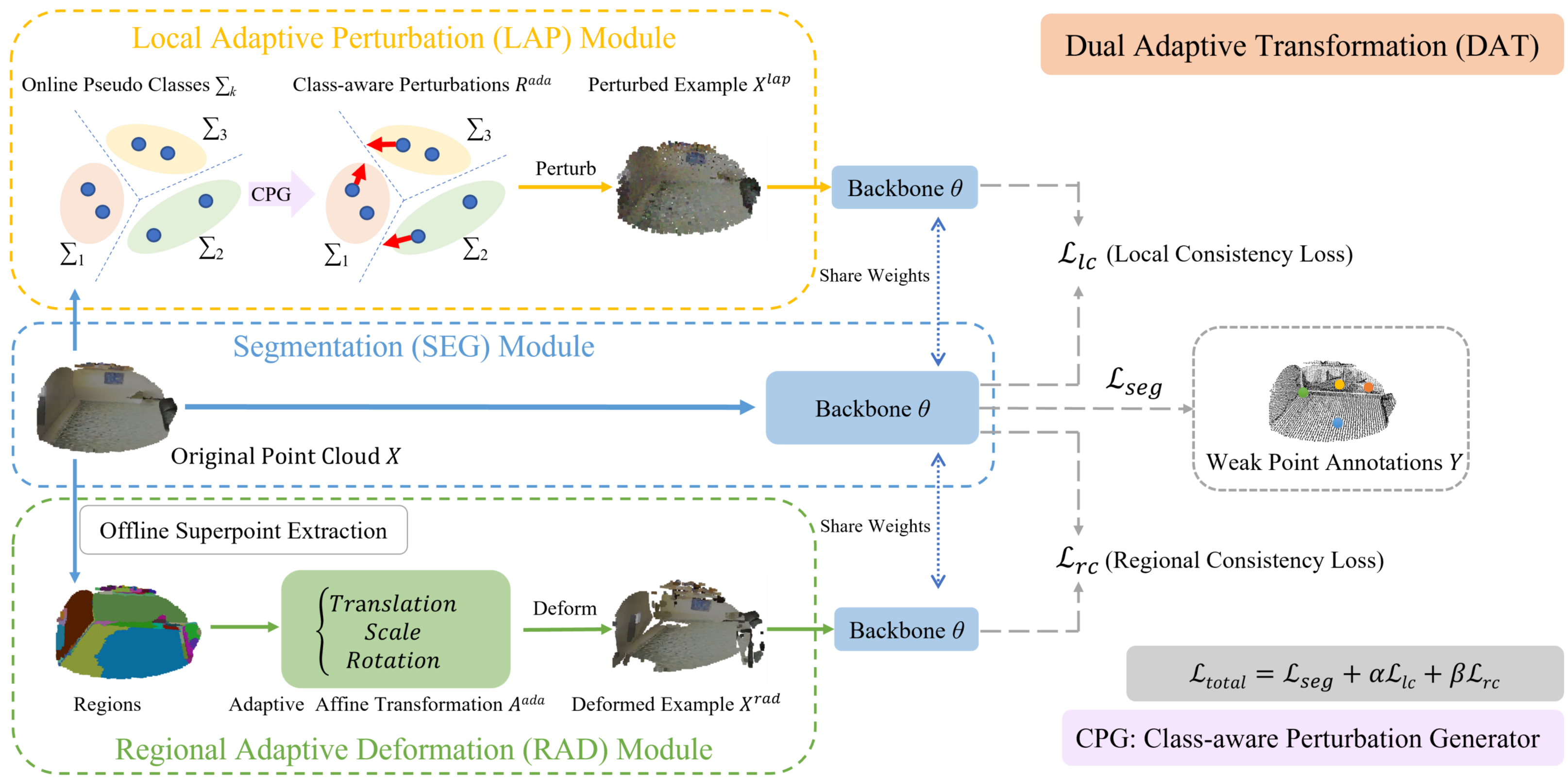
ECCV TEL AVIV 2022 — October 23-27, 2022, Tel Aviv

Contributions:

- ➢ Advocate the consistency constraint under various perturbations to effectively regularize unlabeled 3D points.
- ➢ Propose a novel DAT (Dual Adaptive Transformations) model, via an adversarial strategy at both point-level and region-level, enforcing the local and structural smoothness constraints.



Overview:



- ➢ Overall pipeline of our proposed Dual Adaptive Transformation (DAT) model.
- ➢ Segmentation (SEG) module: adopts KPConv backbone.
- ➢ Local Adaptive Perturbation (LAP) module: generate class-aware perturbed examples on each point.
- ➢ Regional Adaptive Deformation (RAD) module: generates structural deformed data on each region.
- ➢ During testing, we only employ SEG module.

Local Adaptive Perturbation Module:

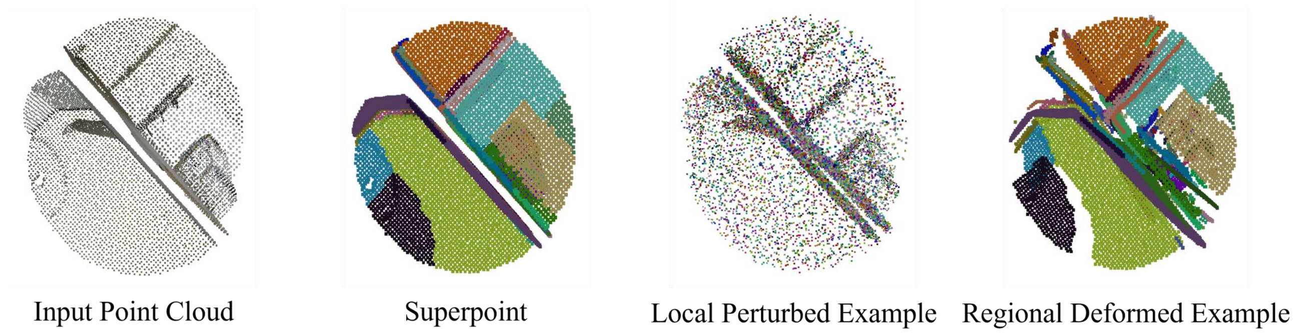
- ➢ Encourage our model to generate consistent outputs between each input point x and its perturbed version $x + r_{ada}$:
  - $LDS(x;\,\theta) = D\left[p(\hat{y}|x;\,\theta), p\left(\hat{y}|x + r^{ada};\theta\right)\right]$
  - $g = \nabla_R \, D[p(\hat{y}|x,\theta), p(\hat{y}|x + r,\theta)]|_{r=\xi d}$
  - $r^{ada} = \epsilon g/\|g\|_2$

Class-aware Perturbation Generator.

- ➢ d_f: sampling from the up-to-date class-aware multivariate Gaussian distribution.
- ➢ d_c: sampling from the iid Gaussian distribution.
- ➢ The LDS loss becomes:
  - $LDS(x;\,\theta) = D\left[p(\hat{y}|c,f;\,\theta), p\left(\hat{y}|c + \xi_c d_c, f + \xi_f d_f;\,\theta\right)\right]$
  - $g_c = \nabla_{\xi_c d_c} LDS(x,\theta)$, $g_f = \nabla_{\xi_f d_f} LDS(x,\theta)$
  - $r_c^{ada} = \epsilon_c g_c/\|g_c\|_2$, $r_f^{ada} = \epsilon_f g_f/\left\|g_f\right\|_2$

Regional Adaptive Deformation Module:

- ➢ For each superpoint S_i, generate the initial affine transformation matrices $A_{i,j}$.
- ➢ Deform each superpoint as:
  - $S_i^{int} = S_i \cdot \prod_{j=1}^{K_a} \xi_a A_{i,j}$
- ➢ The LDS loss becomes:
  - $LDS(X;\,\theta) = D\left[p(\hat{y}|x;\,\theta), p\left(\hat{y}|x^{int};\,\theta\right)\right]$
  - $g_{A_{i,j}} = \nabla_{\xi_A A_{i,j}} LDS(x;\,\theta)$
  - $A_{i,j}^{ada} = \epsilon_A g_{A_{i,j}}/\left\|g_{A_{i,j}}\right\|_2$
- ➢ The deformed superpoints are computed as:
  - $S_i^{ada} = S_i * \prod_{j=1}^{K_a} A_{i,j}^{ada}$



Experimental Results:

Table 1. Comparison of our DAT with several existing methods on the S3DIS Area-5 set. Note that, we report the performance as final results based on the KPConv [38] backbone.

Method	Supervision (%)	mIoU (%)
PointNet [32]	100%	41.1
PointCNN [21]	100%	57.3
Xu et al. [53]	0.2%	44.5
Xu et al. [53]	10%	48.0
GPFN [39]	16.7% 2D	50.8
GPFN [39]	100% 2D	52.5
1T1C [24]	0.02% (OTOC)	50.1
1T1C [24]	0.06% (OTTC)	55.3
Our DAT	0.02% (OTOC)	**56.5**
Our DAT	0.06% (OTTC)	**58.5**
Our Upper Bound	100%	65.4

Table 2. Comparison of our DAT with its variant methods with the KPConv framework. Note that, all experiments are conducted under the OTOC setting on the S3DIS dataset

Method	Random Noises		LAP	RAD	mIoU (%)
	Features	Coordinates			
Our Baseline					50.1
Ours w/ Noise	✓				49.1
Ours w/ Noise		✓			52.9
Ours w/ Noise	✓	✓			52.6
Ours w/ PAP			✓		53.9
Ours w/ RAD				✓	54.8
Our DAT			✓	✓	**56.5**

Table 6. Comparison of our DAT model with several existing methods on the ScanNet-v2 test set. " Our DAT†" denotes that our DAT is built upon the 1T1C [24] model.

Method	Supervision	mIoU (%)
Pointnet++ [33]	100%	33.9
PointCNN [21]	100%	45.8
MinkowskiNet [5]	100%	73.6
Virtual MVFusion [15]	100%+2D	74.6
MPRM [44]	scene-level	24.4
MPRM [44]	subcloud-level	41.1
MPRM+CRF [44]	subcloud-level	43.2
CSC_LA_SEM [11]	20 points	53.1
Viewpoint_BN_LA_AIR [25]	20 points	54.8
PointContrast_LA_SEM [52]	20 points	55.0
1T1C [24]	20 points	59.4
Our Baseline	20 points	51.6
Our DAT	20 points	55.2
Our DAT†	20 points	**62.3**
Our Upper Bound	100%	68.4

Qualitative Results:

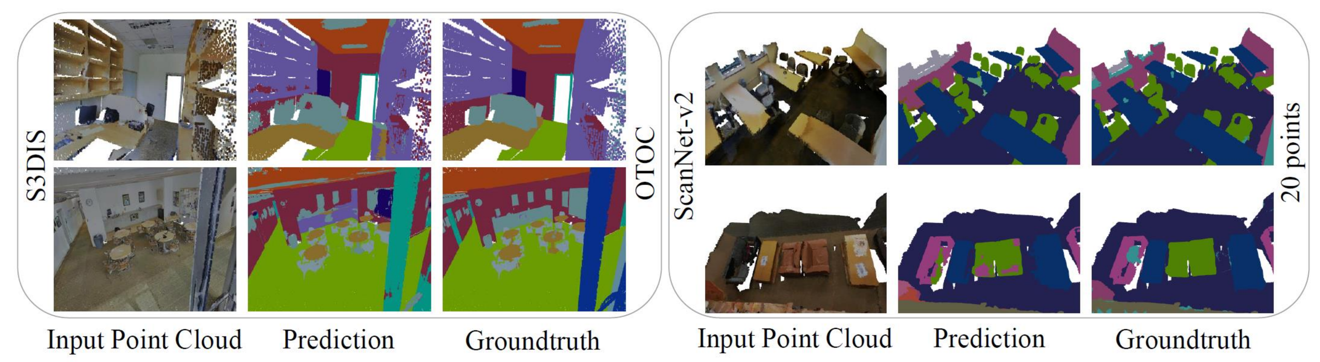


Fig. 5. Two results of our DAT on the S3DIS (first two rows, under the "OTOC" setting) and ScanNet-v2 datasets (last two rows, under the "20 points" setting).