COARSE3D: Class-Prototypes for Contrastive Learning in Weakly-Supervised 3D Point Cloud Segmentation

Code & Video

BMVC 2022

Rong Li[1], Anh-Quan Cao[2], Raoul de Charette[2]

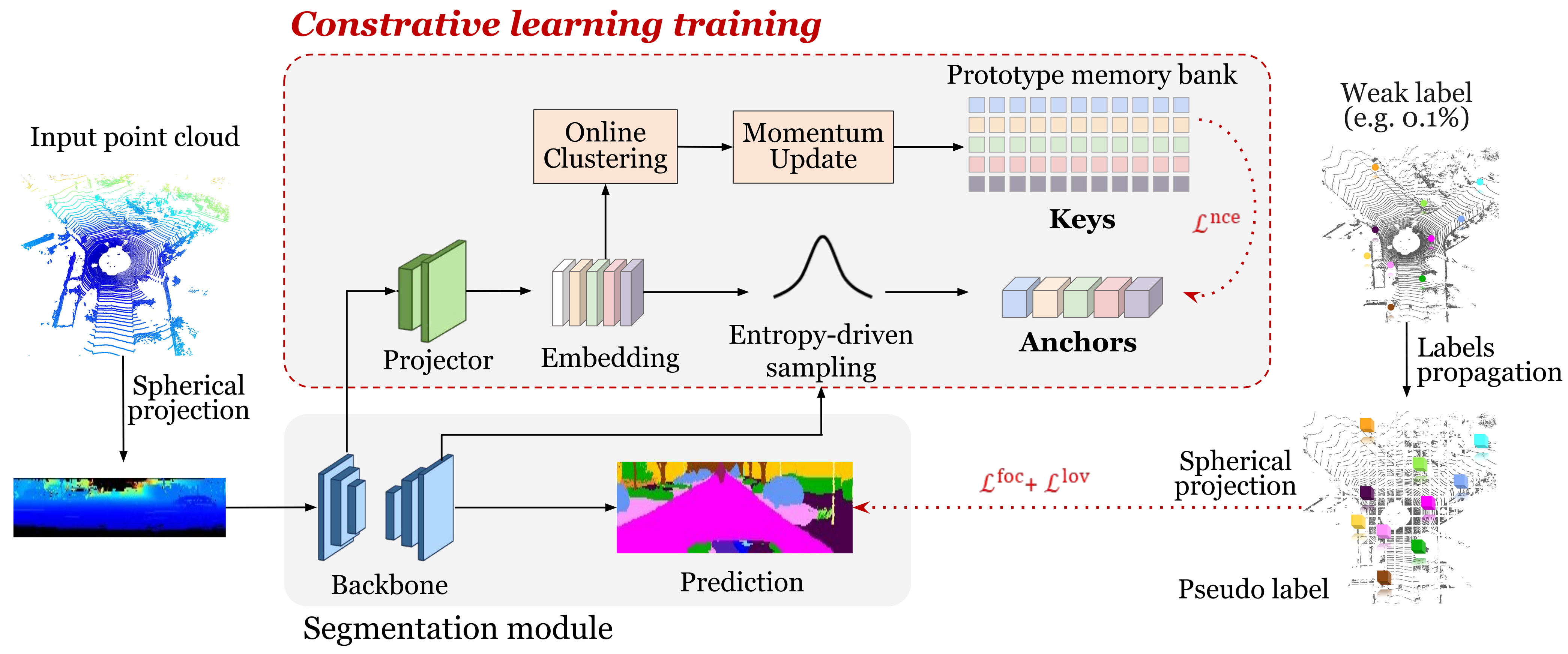
Motivation

Most semantic segmentation rely on **costly point-wise human annotation**. Instead, we explore a weakly supervised learning strategy, training with as few as **0.001% of the labels**.

Contributions

- An **architecture-agnostic framework** for weakly-supervised 3D semantic segmentation.
- A **prototype memory bank** that captures per-class dataset information with an **entropy-driven sampling** technique to sample more confident pixels as anchors.
- Results on **3 baseline architectures** and **3 datasets** demonstrate the effectiveness.

Overview



- **Contrastive learning pipeline:** architecture-agnostic plug-in in training.
- **Prototype memory bank:** cluster pixel-wise features into compact prototypes as keys at each iteration.
- **Entropy-driven anchors sampling:** select the most appropriate pixels as anchors among the abundant pseudo-labels predictions.

Contrastive learning

Pixel-prototype-based contrastive loss with InfoNCE

$$\mathcal{L}^{\text{Pix2Proto}} = \frac{1}{N_{\text{a}}} \sum_{a_i \in \mathcal{A}} -\log \frac{\sum_{p_j^+ \in \mathcal{P}^+} \exp\left(a_i \cdot p_j^+ / \tau\right)}{\sum_{p_j^+ \in \mathcal{P}^+} \exp\left(a_i \cdot p_j^+ / \tau\right) + \sum_{p_j^- \in \mathcal{P}^-} \exp\left(a_i \cdot p_j^- / \tau\right)}$$

with $\mathcal{A}$ the anchors, $\mathcal{P}^+/\mathcal{P}^-$ the positive/negative keys (prototypes)

Entropy-driven sampling

Using information theory we sample relevant pseudo-labels predictions to serve as anchors. The sampling writes:

$$\rho(x_i) = \frac{\exp -H(x_i)^2}{\sum_{x_j \in \mathcal{X}} \exp -H(x_j)^2}$$

with $H(x_i)$ the Shannon entropy

Prototype memory bank

Our prototypes memory bank avoid greedy pixels storage. We apply **online prototype clustering,** to compute pixel-prototypes mapping framed as an optimal transport problem using Sinhorn algorithm. A **momentum update** ($\sigma = 0.999$) is applied to prototypes.

j^{th} prototype $\{P_k\}_j$ of class k is updated as:

$$\{P_k\}_j = \sigma \{P_k\}_j + (1-\sigma) \frac{1}{\sum_{i=1}^{N_k} [\![m(x_i)=j]\!]} \sum_{i=1}^{N_k} x_i [\![m(x_i)=j]\!]$$

$m(x_i)$: prototype mapping of pixel x_i

$\{P_k\}_j$: j-th prototype of class k

Experiments

We evaluate COARSE3D on 3 datasets and 3 projection-based backbones.

Unless mentioned otherwise, we use SalsaNext [2] backbone.

SemanticKITTI (hidden set)

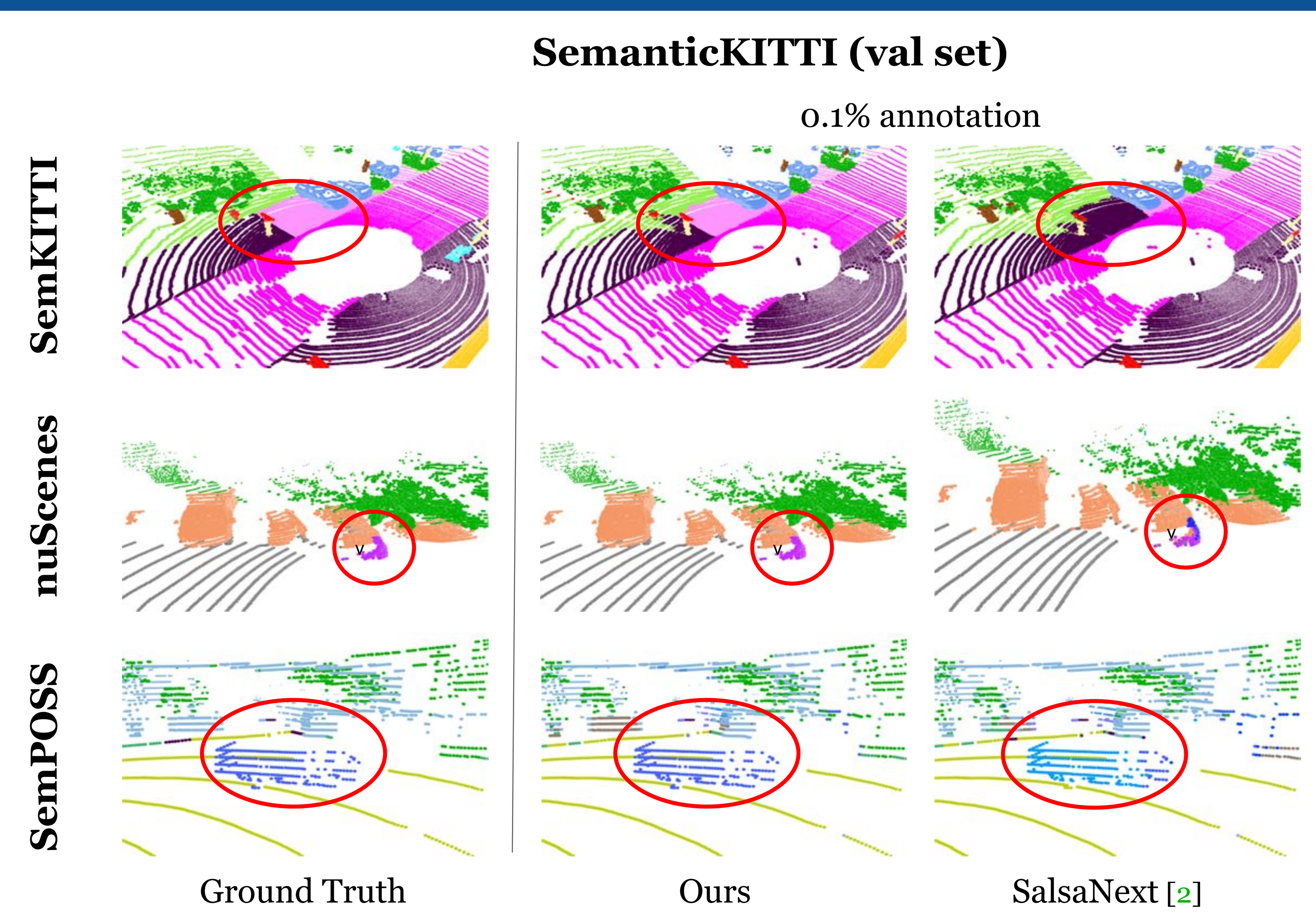
Anno. (%)	Method	Proj	mIoU (%)
100	(AF)²S3Net [5]	×	69.7
	SquSegV3 [4]	√	55.9
	SalsaNext [2]		59.5
0.1	SQN [1]	×	50.8
	SalsaNext [2]	√	50.1
	Ours		**55.7**
0.01	SQN [1]	×	39.1
	SalsaNext [2]	√	42.6
	Ours		**46.2**

SemanticPOSS (val set)

Anno. (%)	Method	Proj	mIoU (%)
100	RandLANet[8]		53.5
	KPConv [9]	×	55.2
	JS3C-Net [10]		60.2
	SquSegV2[11]	√	29.8
	SalsaNext [2]		45.0
0.1	SalsaNext [2]	√	38.9
	Ours		**43.0**
0.01	SalsaNext [2]	√	27.4
	Ours		**31.1**

nuScenes (val set)

Anno. (%)	Method	Proj	mIoU (%)
100	PolarNet [6]		72.2
	Cylinder3D [7]	×	76.1
	(AF)²S3Net [5]		78.0
	RangeNet[3]	√	65.5
	SalsaNext [2]		72.2
0.1	SalsaNext [2]	√	56.5
	Ours		**58.7**
0.01	SalsaNext [2]	√	**44.5**
	Ours		42.9



Varying architectures

Methods	SemPOSS mIoU (%)	SemKITTI mIoU (%)
Rangenet-21 [3]	25.1	40.7
Ours (Rangenet-21)	**28.9 (+3.8)**	**44.5 (+3.8)**
SqueezeSegV3-21 [4]	30.4	42.5
Ours (SqueezeSegV3-21)	**36.7 (+6.3)**	**48.5 (+6.0)**
SalsaNext [2]	38.9	52.4
Ours (SalsaNext)	**43.0 (+4.1)**	**57.6 (+5.2)**

Architecture ablation

Methods	mIoU (%)
Ours	**57.57**
w/o contrast module	55.44
w/o anchor sampling	56.32
w/o prototype (5k pxl)	56.10
w/o voxel propagation	56.26
w/o Focal loss	42.41
w/o Lovasz loss	56.10

Effect of varying annotation

Anno.	mIoU (%)	
	SalsaNext [2]	Ours
0.001%	30.39	31.69
0.01%	44.00	47.13
0.1%	52.43	56.61
1%	56.16	58.30
100%	56.44	58.39

[1] SQN. Hu et al. ECCV 2022.
[2] SalsaNext. Tiago et al. ISVC 2020.
[3] Rangenet. Milioto et al. IROS 2019.
[4] SqueezeSegV3. Xu et al. ECCV 2020.
[5] (AF)2S3Net. Cheng et al. CVPR 2021.
[6] PolarNet. Zhang et al. CVPR 2020.
[7] Cylinder3D. Zhu et al. CVPR 2021.
[8] RandLANet. Hu et al. CVPR 2020.
[9] KPConv. Thomas et al. ICCV 2019.
[10] JS3C-Net. Yan et al. AAAI 2021.
[11] SqueezeSegV2. Wu et al. ICRA 2018.