Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion

Shi Qiu [1,2], Saeed Anwar [1,2], Nick Barnes [1]
[1] The Australian National University; [2] DATA61-CSIRO, Australia

Point Cloud Semantic Segmentation

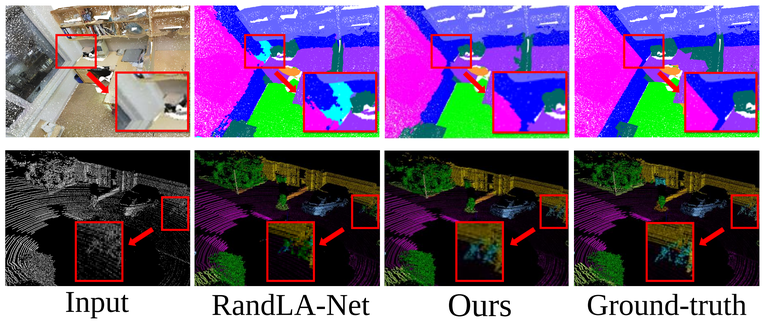
Background:

◇ Point clouds are collected by scanners, but usually *scattered, irregular, unordered, and unevenly distributed* in 3D space.

◇ It is very challenging for machine's perception on large scenes made of millions or even billions points.

◇ We focus on the semantic segmentation task to help AI-driven machines better recognize the complex surroundings in the real world.

Examples:



Current approaches:

◇ Projection-based

◇ Discretization-based

◇ Point-based

Limitations:

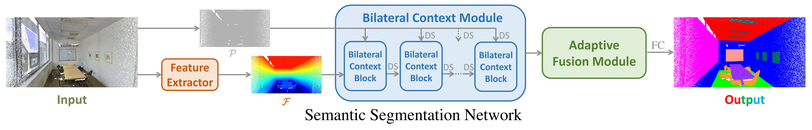
◇ Ambiguity in close points

◇ Redundant features

◇ Inadequate representations

Network Architecture

BAAF-Net: consisting of three different modules for point cloud semantic segmentation.



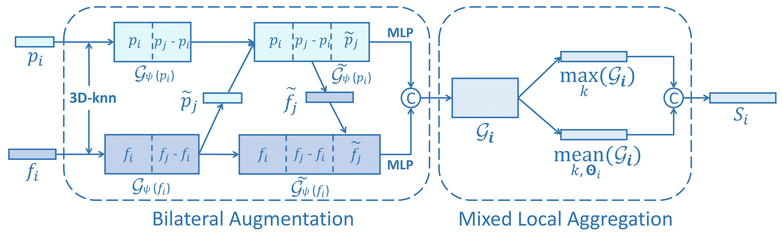
◇ **Feature Extractor**: to capture the preliminary semantic context $\mathcal{F}$ from the input data via a single-layer MLP.

◇ **Bilateral Context Module**: to deploy cascaded Bilateral Context Blocks augmenting the local context of multiple point cloud resolutions, using both semantic $\mathcal{F}$ and geometric context $\mathcal{P}$.

◇ **Adaptive Fusion Module**: to upsample the Bilateral Context Blocks' outputs, and adaptively fuses them as an output feature map. Finally, we predict semantic labels for all points via fully-connected layers.

Bilateral Context Block

Bilateral Context Block: to augment the local context of each point by involving the offsets that are mutually learned from the geometric $p_i \in \mathbb{R}^3$ and semantic $f_i \in \mathbb{R}^d$ inputs, then precisely aggregate the augmented local context to represent the point feature.



Bilateral Context Block

◇ **Bilateral Augmentation:** to learn position-related offsets from semantic features to enhance *local geometric* context and feature-related offsets from geometric features to enhance *local semantic* context.

◇ **Augmentation Loss:** to regulate the learning process of the bilateral offsets. In practice, we encourage the *geometric center* of the shifted neighbors to approach the local centroid in 3D space by minimizing the ℓ_2 distance.

◇ **Mixed Local Aggregation:** to directly collect the maximum feature from the neighbors for an overview, and obtain more details by learning the high-dimensional barycenter over the neighborhood.

Adaptive Fusion Module

Adaptive Fusion Module: to upsample the multi-resolution outputs of the Bilateral Context Module, then adaptively fuse them as a comprehensive feature map for the whole point cloud scene.

Algorithm 1: Adaptive Fusion Module Pipeline

input: M multi-resolution feature maps $\{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_M\}$.
output: $\mathcal{S}_{out}$ for semantic segmentation.
for $\mathcal{S}_m \in \{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_M\}$ **do**
  upsample: $\tilde{\mathcal{S}_m} \leftarrow \mathcal{S}_m$;
  summarize: $\phi_m \leftarrow \tilde{\mathcal{S}_m}$;
end for
obtain: $\forall \tilde{\mathcal{S}_m} \in \{\tilde{\mathcal{S}_1}, \tilde{\mathcal{S}_2}, ..., \tilde{\mathcal{S}_M}\}$, $\tilde{\mathcal{S}_m} \in \mathbb{R}^{N \times c}$; and $\forall \phi_m \in \{\phi_1, \phi_2, ..., \phi_M\}$, $\phi_m \in \mathbb{R}^N$.
regress: $\{\Phi_1, \Phi_2, ..., \Phi_M\} \leftarrow \{\phi_1, \phi_2, ..., \phi_M\}$, where $\Phi_m \in \mathbb{R}^N$.
return:
$\mathcal{S}_{out} = \sum_{m=1}^{M} \Phi_m \times \tilde{\mathcal{S}_m}$.

Experiments

S3DIS (6-fold):

Methods	mAcc	OA	mIoU
KPConv	79.1	-	70.6
Seg-GCN	77.1	87.8	68.5
PointASNL	79.0	88.8	68.7
RandLA-Net	82.0	88.0	70.0
BAAF-Net	**83.1**	**88.9**	**72.2**

Semantic3D (semantic-8):

Methods	OA	mIoU
PointNet++	85.7	63.1
PointConv-CE	92.3	71.0
RandLA-Net	92.4	71.8
SPG	92.9	**76.2**
BAAF-Net	**94.9**	75.4

SemanticKITTI:

Methods	PointASNL	RandLA-Net	PolarNet	MinkNet42	FusionNet	**BAAF-Net**
mIoU	46.8	53.9	54.3	54.3	**61.3**	59.9

Ablation Studies

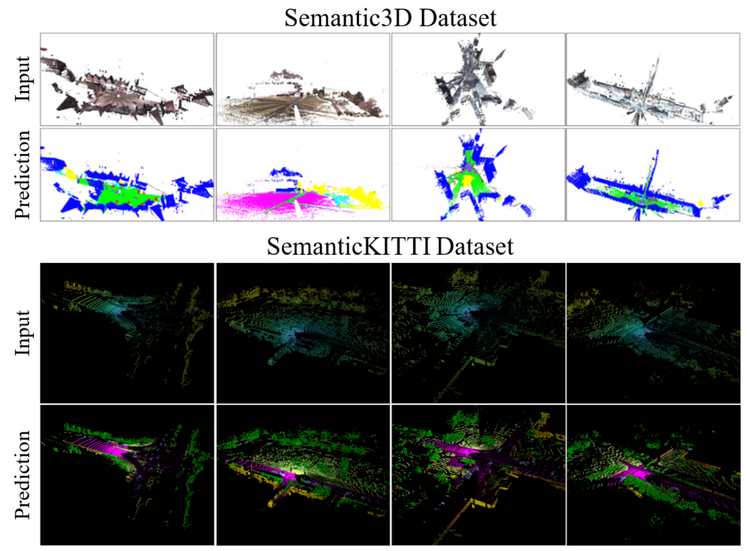
Bilateral Context Block:

bilateral offsets	augmentation loss	local aggregation	mIoU
$\tilde{f}_i \rightarrow \tilde{p}_i$	$\mathcal{L}(f_i)$	mixed	64.2
$\tilde{p}_i \rightarrow \tilde{f}_i$	$\mathcal{L}(p_i) + \mathcal{L}(f_i)$	mixed	64.3
$\tilde{p}_i \rightarrow \tilde{f}_i$	none	mixed	64.2
$\tilde{p}_i \rightarrow \tilde{f}_i$	$\mathcal{L}(p_i)$	max	64.6
$\tilde{p}_i \rightarrow \tilde{f}_i$	$\mathcal{L}(p_i)$	mean	64.8
$\tilde{p}_i \rightarrow \tilde{f}_i$	$\mathcal{L}(p_i)$	mixed	**65.4**

Adaptive Fusion Module:

upsampled feature map	fusion parameters	$\mathcal{S}_{out}$	mIoU
$\tilde{\mathcal{S}_M}$	none	$\tilde{\mathcal{S}_M}$	64.1
$\{\tilde{\mathcal{S}_m}\}$	none	$\sum \tilde{\mathcal{S}_m}$	64.7
$\{\tilde{\mathcal{S}_m}\}$	none	$\prod \tilde{\mathcal{S}_m}$	64.2
$\{\tilde{\mathcal{S}_m}\}$	none	concat($\{\tilde{\mathcal{S}_m}\}$)	65.1
$\{\tilde{\mathcal{S}_m}\}$	$\{\Psi_m\}$	$\sum \Psi_m \times \mathcal{S}_m$	65.1
$\{\tilde{\mathcal{S}_m}\}$	$\{\Phi_m\}$	$\sum \Phi_m \times \mathcal{S}_m$	**65.4**

Visualization



Conclusions

◇ Augmenting the local context bilaterally.

◇ Fusing multi-resolution features for each point adaptively.

◇ Evaluating the network on different point cloud benchmarks.

◇ Expecting different frameworks and downstream tasks in the future.