Original Residual Learning Block
(Inefficient Implementation)

(N, D_{in})
1×1 Conv
$N \times D_{in}$
$(N \times D_o)$
Side 1×1 Conv
PG-Net Operator
$N \times D_o$
$(N \times D_o)$
1×1 Conv
$(N \times D_o)$
$\oplus$
$(N \times D_o)$

Residual Learning Block 1 (RLB1 when $D_{in} \neq D_{out}$)

(N, D_{in})
1×1 Conv
$N \times D_{in}$
$(N \times \frac{D_o}{M})$
Side 1×1 Conv
PG-Net Operator
$(N \times \frac{D_o}{M})$
$N \times D_o$
1×1 Conv
$N \times D_o$
$\oplus$
$N \times D_o$

Residual Learning Block 2 (RLB2 when $D_{in} = D_{out}$)

(N, D_{in})
1×1 Conv
$(N \times \frac{D_{in}}{M})$
PG-Net Operator
$(N \times \frac{D_{in}}{M})$
1×1 Conv
$N \times D_{in}$
$\oplus$
$N \times D_{in}$

Point Clouds Encoder

$N \times C_{in}$
1×1 Conv
$N \times 8$
RLB1
RLB2
$\frac{N}{5} \times 32$
IDS
RLB1
RLB2
$\frac{N}{25} \times 64$
IDS
RLB1
RLB2
$\frac{N}{125} \times 128$
IDS
RLB1
RLB2
$\frac{N}{625} \times 256$
GSS
RLB1
$M_{in}: \frac{N}{625} \times 256$
Global Feature Extraction
$\frac{N}{625} \times 256$
1×1 Conv
$\frac{N}{625} \times 256$

Max Pooling
1×256
1×1 Conv
$1 \times N_{cls}$

Cls. Predictions

Contextual Loss

car   parking
bicycle   sidewalk
motorcycle   truck
other vehicles buildings
other ground fence
traffic sign vegetation
bicyclist   trunk
road   terrain
pole   person

Scale 5   Scale 4   Scale 3   Scale 2   Scale 1

Multi-Layer Fusion

1×1 Conv   US   Loss
$N \times N_{cls}$

Point Clouds Decoder

US
1×1 Conv
$\frac{N}{125} \times 128$
US
1×1 Conv
$\frac{N}{25} \times 64$
US
1×1 Conv
$\frac{N}{5} \times 32$
US
1×1 Conv
$N \times 8$

$N \times N_{cls}$
$N \times 5N_{cls}$
1×1 Conv
Loss
$N \times N_{cls}$

Seg. Predictions