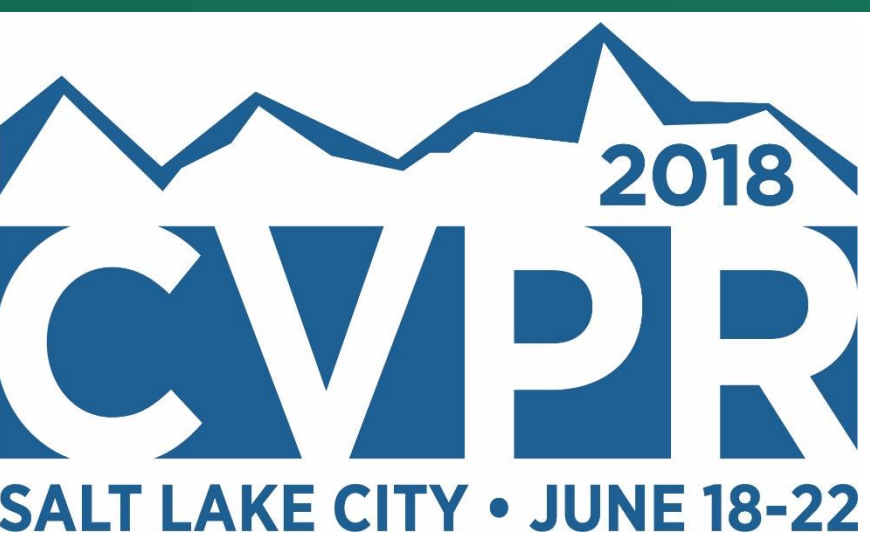# Pointwise Convolutional Neural Networks
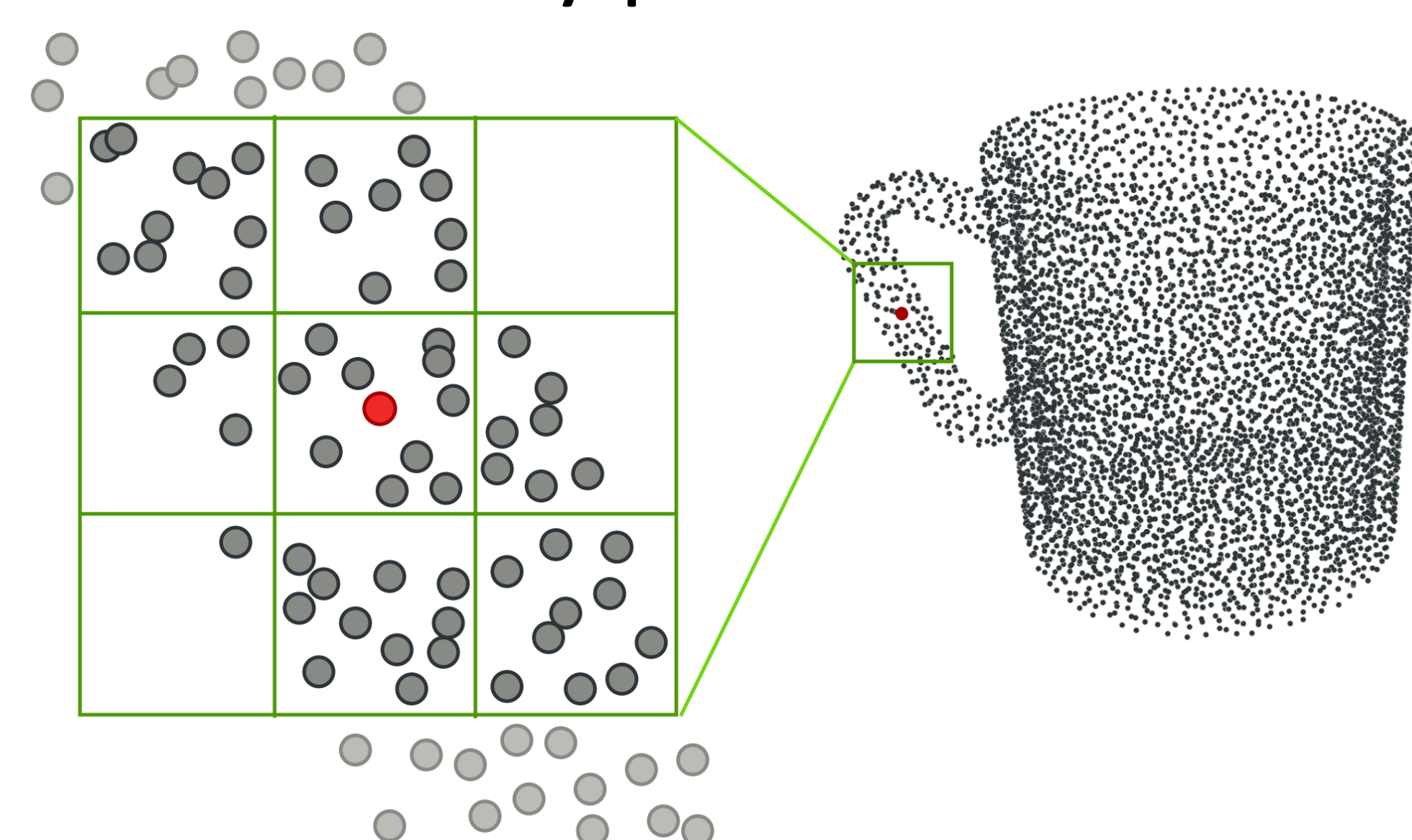
Binh-Son Hua

The University of Tokyo

Minh-Khoi Tran   Sai-Kit Yeung

Singapore University of Technology and Design

Pointwise Convolution

> Convolution at every point of the cloud

> **On-the-fly** uniform grid for nearest neighbour search
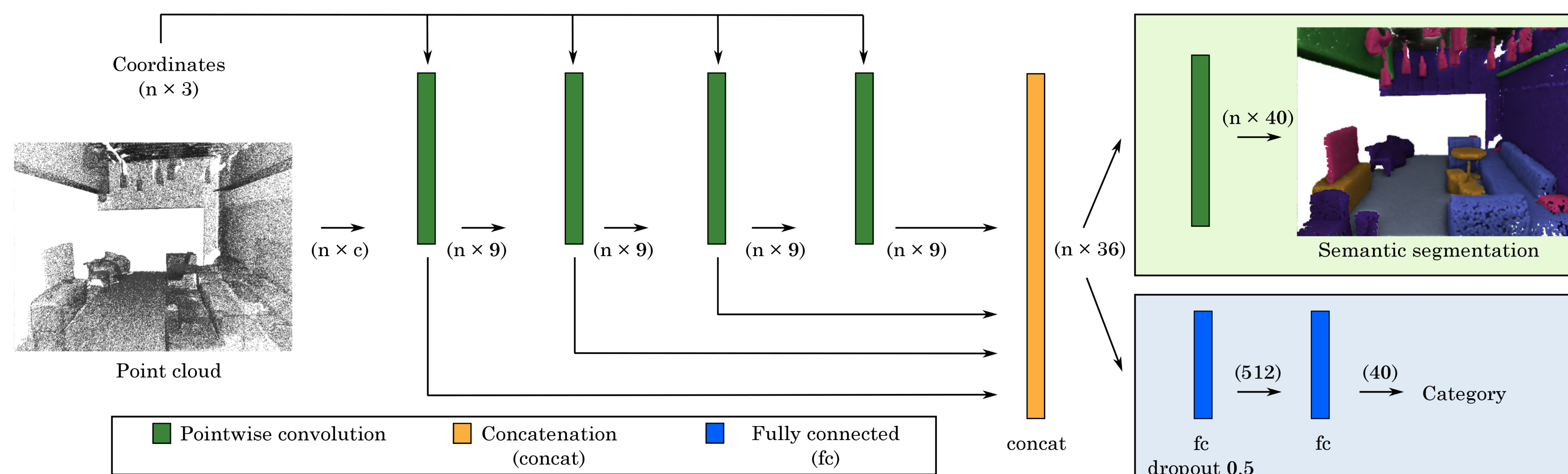> Forward convolution

$$x_i^\ell = \sum_k w_k \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} x_j^{\ell-1},$$
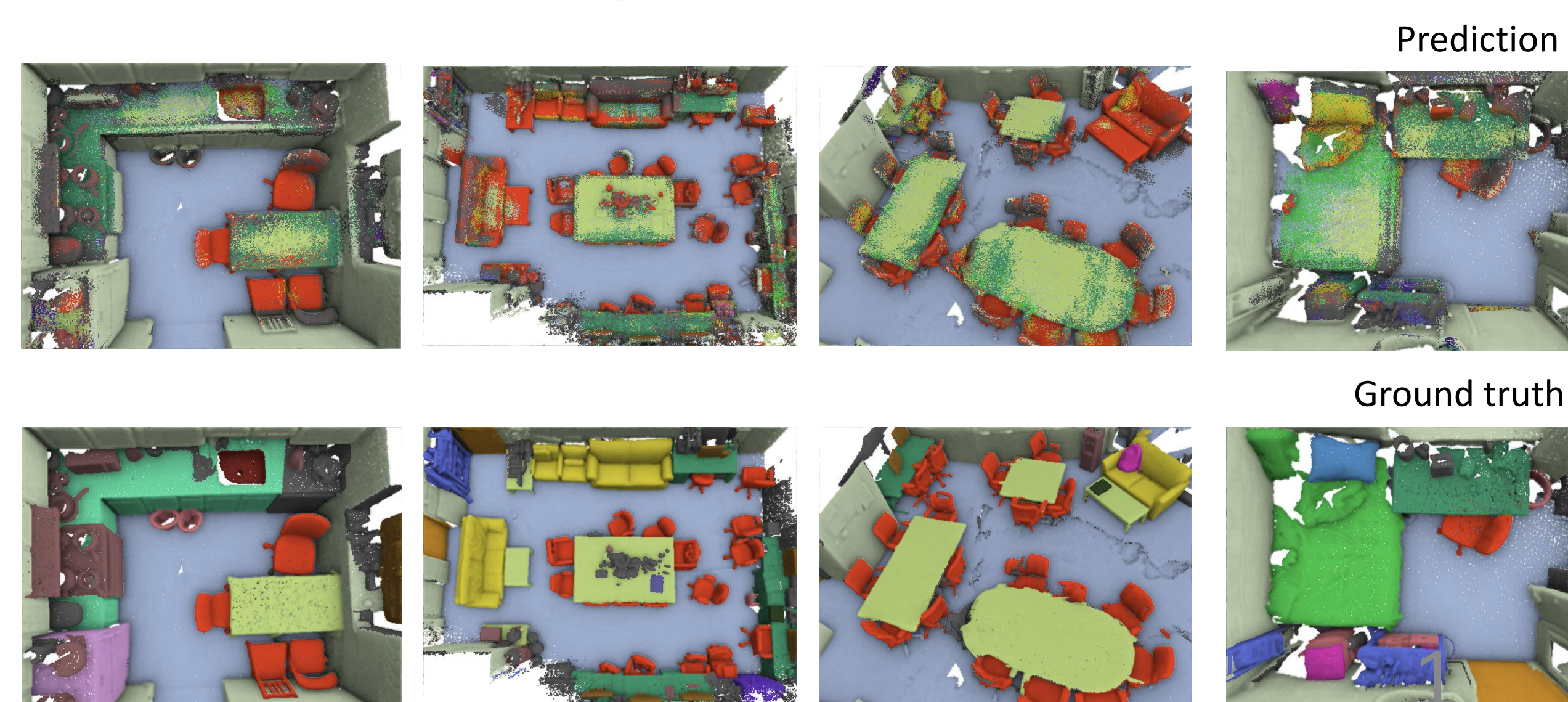
> Backward propagation

$$\frac{\partial L}{\partial x_j^{\ell-1}} = \sum_{i \in \Omega_j} \frac{\partial L}{\partial x_i^\ell} \frac{\partial x_i^\ell}{\partial x_j^{\ell-1}} \qquad \frac{\partial x_i^\ell}{\partial x_j^{\ell-1}} = \sum_k w_k \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} 1$$

$$\frac{\partial L}{\partial w_k} = \sum_i \frac{\partial L}{\partial x_i^\ell} \frac{\partial x_i^\ell}{\partial w_k} \qquad \frac{\partial x_i^\ell}{\partial w_k} = \frac{1}{|\Omega_i(k)|} \sum_{p_j \in \Omega_i(k)} x_j^{\ell-1}$$

> À-trous convolution
> Self-normalizing activation function (SeLU)
> CUDA and multi-GPU implementation

> Source code and data

pointwise.scenenn.net

Coordinates (n × 3)

Point cloud

(n × c) → (n × 9) → (n × 9) → (n × 9) → (n × 9) → (n × 36)

| ■ Pointwise convolution | ■ Concatenation (concat) | ■ Fully connected (fc) |

concat

(n × 40) →

Semantic segmentation

(512) → (40) → Category

fc   fc   dropout 0.5

Semantic Segmentation: SceneNN

Prediction

Ground truth

Class	Ours	PointNet	Class	Ours	PointNet
wall	0.868	**0.897**	table	**0.412**	0.235
floor	0.864	**0.891**	counter	**0.144**	0.052
cabinet	**0.214**	0.090	desk	**0.362**	0.310
bed	**0.513**	0.457	pillow	**0.175**	0.067
chair	**0.639**	0.596	tv	**0.178**	0.114
sofa	**0.298**	0.167	box	0.141	**0.163**

Future Works

> Adapt neural network design from 2D to 3D with pointwise convolution.
> Global feature learning.
> Applications: denoising, upsampling, colorization.

Object Recognition

Base	Concat.	À-trous	SELU	Dropout	Accuracy
✓					78.6
✓	✓				78.0
✓		✓			75.0
✓	✓	✓			82.5
✓			✓		81.7
✓	✓		✓		81.9
✓	✓		✓	✓	85.2
✓	✓	✓	✓	✓	86.1

Convergence

(a) Scene segmentation   (b) Object recognition