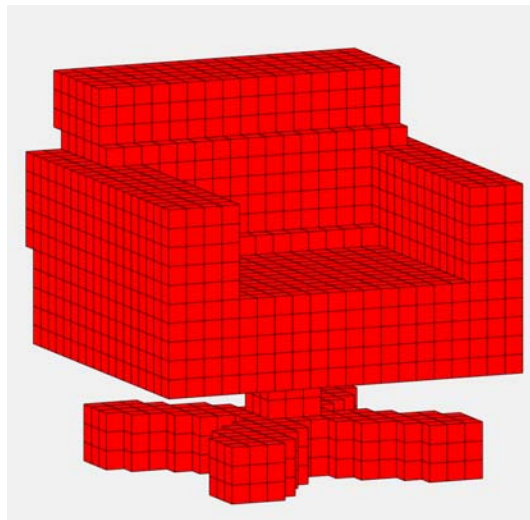
3D Deep Learning approaches
Volumetric Convnets



Evangelos Kalogerakis

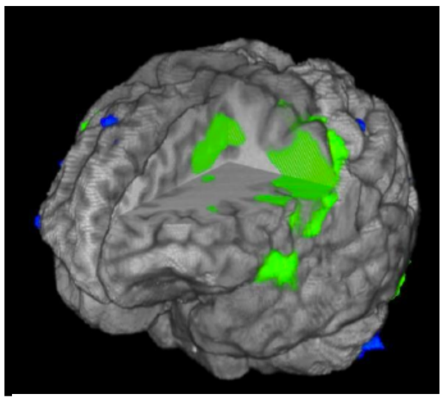
UMASS AMHERST

3D Deep Learning approaches

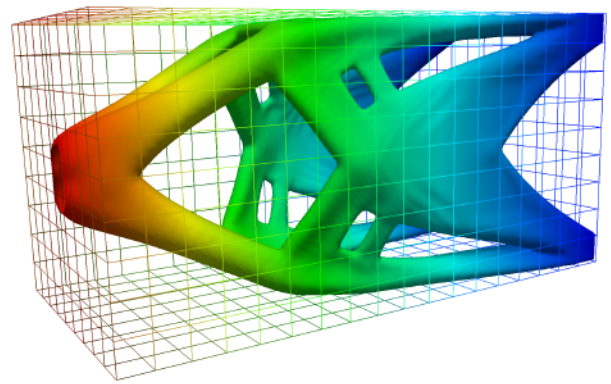
- The Multi-View approach
- **The Voxel approach**
- The Point approach
- The Graph approach

Motivation

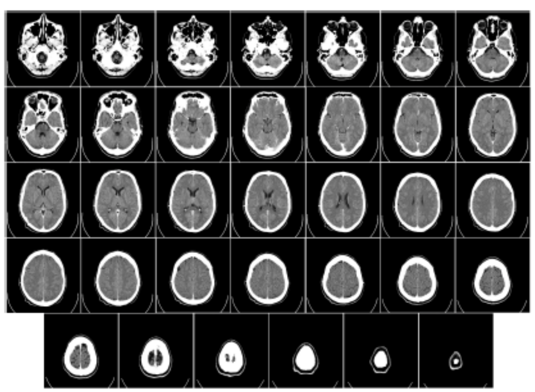
Some types of 3D data are truly volumetric (not “empty” inside)



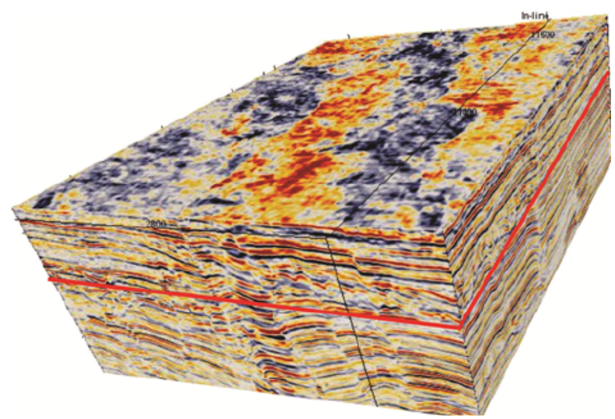
fMRI



Physical properties of 3D objects



CT

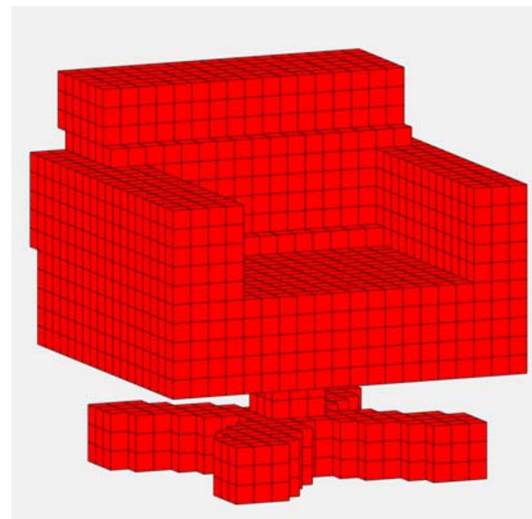


Geological data

Voxelization

Convert shape to 3D regular grid

3D polygon mesh



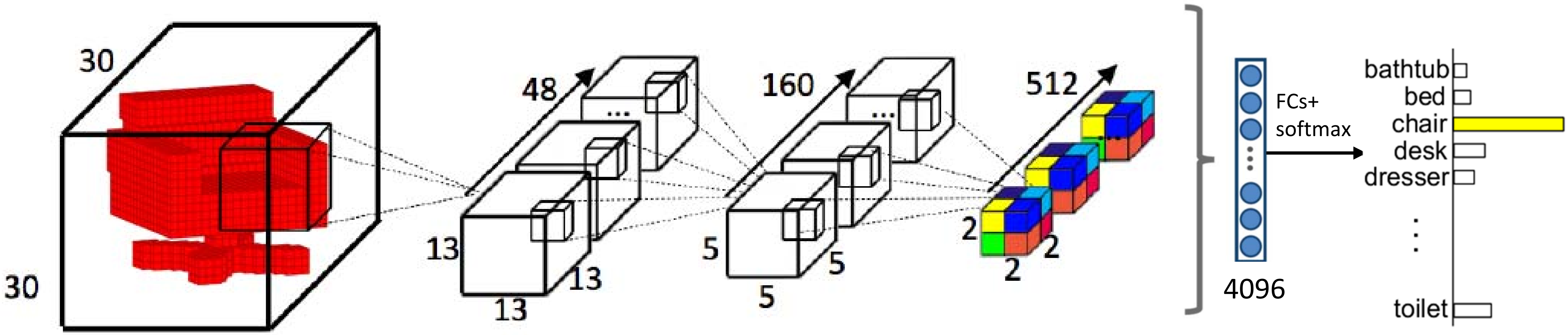
Voxels

3D Deep Learning approaches

- The Multi-View approach
- **The Voxel approach**
  - ***Dense Volumetric Nets***
  - Octree Nets
  - Sparse Tensor Nets
- The Point approach
- The Graph approach

Volumetric Network

Volumetric networks use convolution over 3D spatial input (=> **4D** feature maps)



Flatten 4D map

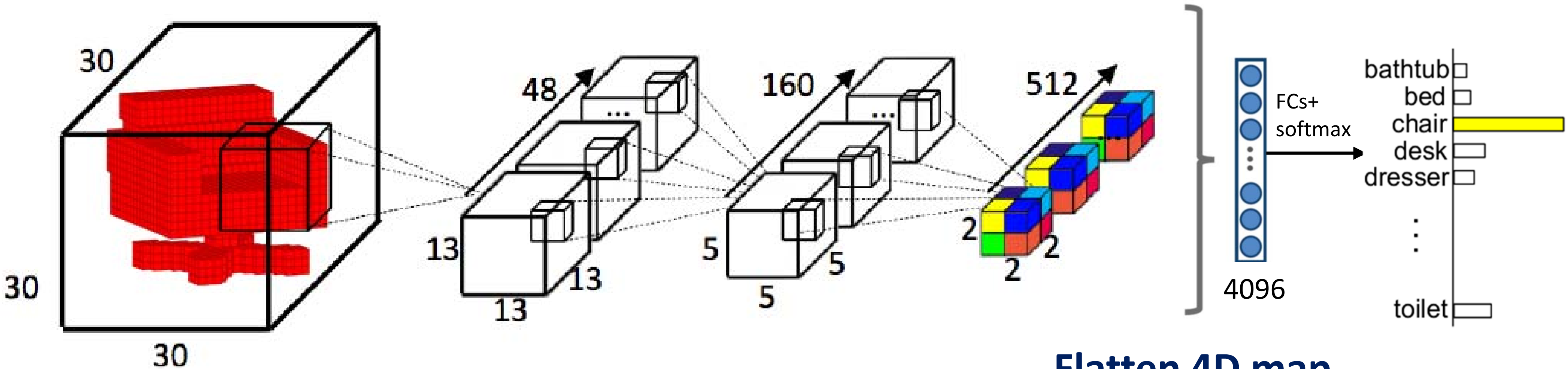
$$O(x,y,z,q) = \sum_{k=-n}^{k=n} \sum_{l=-n}^{l=n} \sum_{m=-n}^{n} \sum_{channel\ c} w_q(k,l,m,c) I(x+k, y+l, z+m, c)$$

3D ShapeNets: A Deep Representation for Volumetric Shapes, Wu et al. 2015

Volumetric Network

Volumetric networks use convolution over 3D spatial input (=> **4D** feature maps)

Computationally & memory expensive! Requires low-res input!



Flatten 4D map

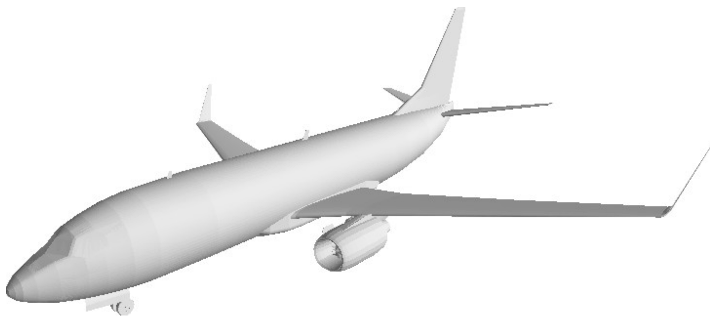
$$O(x,y,z,q)=\sum_{k=-n}^{k=n}\sum_{l=-n}^{l=n}\sum_{m=-n}^{n}\sum_{channel\ c} w_q(k,l,m,c)I(x+k,y+l,z+m,c)$$

3D ShapeNets: A Deep Representation for Volumetric Shapes, Wu et al. 2015

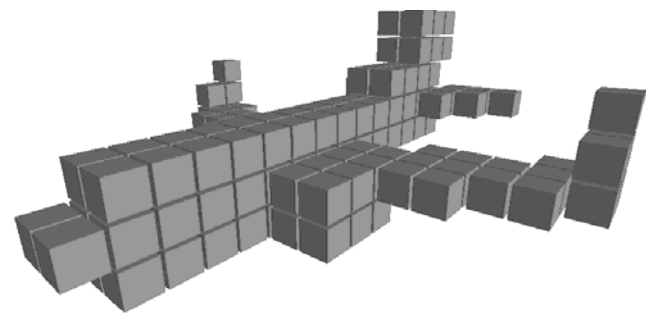
Sparsity of 3D data

Rendered Mesh

Sparsity of 3D data



Rendered Mesh

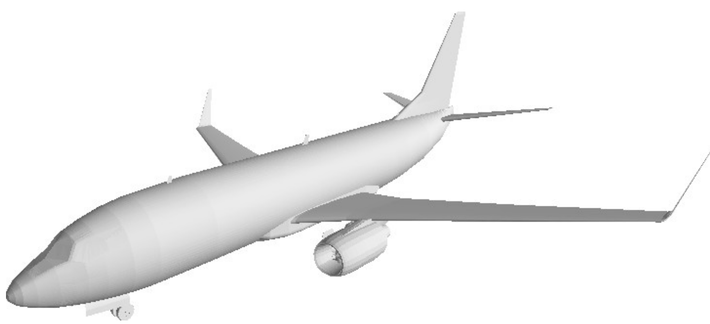


Voxelized 16^3
Occupancy 4.19%

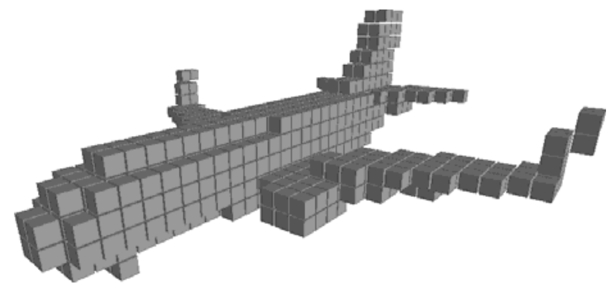
[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Sparsity of 3D data



Rendered Mesh

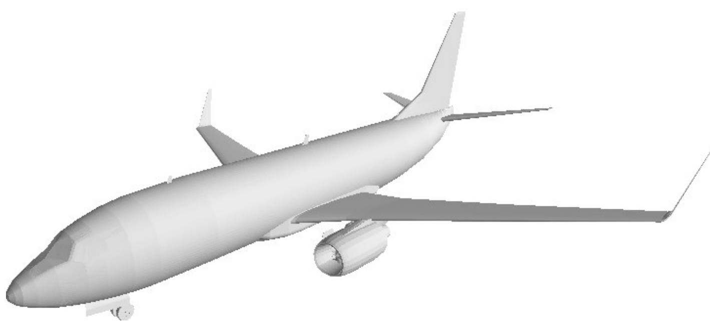


Voxelized 32^3
Occupancy 2.11%

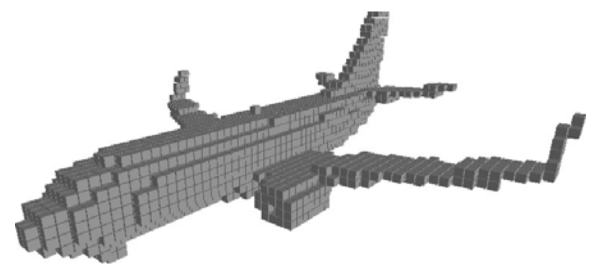
[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Sparsity of 3D data



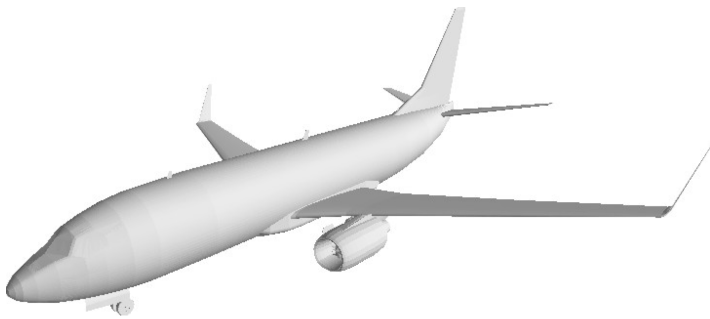
Rendered Mesh



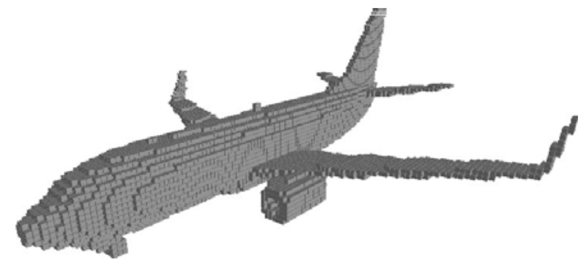
Voxelized 64^3
Occupancy 1.06%

[Slides from Riegler et al. OctNet]
OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Sparsity of 3D data



Rendered Mesh



Voxelized 128^3
Occupancy 0.56%

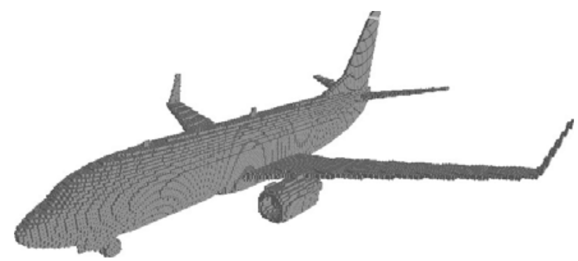
[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Sparsity of 3D data

Running convolution on so much empty space is wasteful!

Rendered Mesh



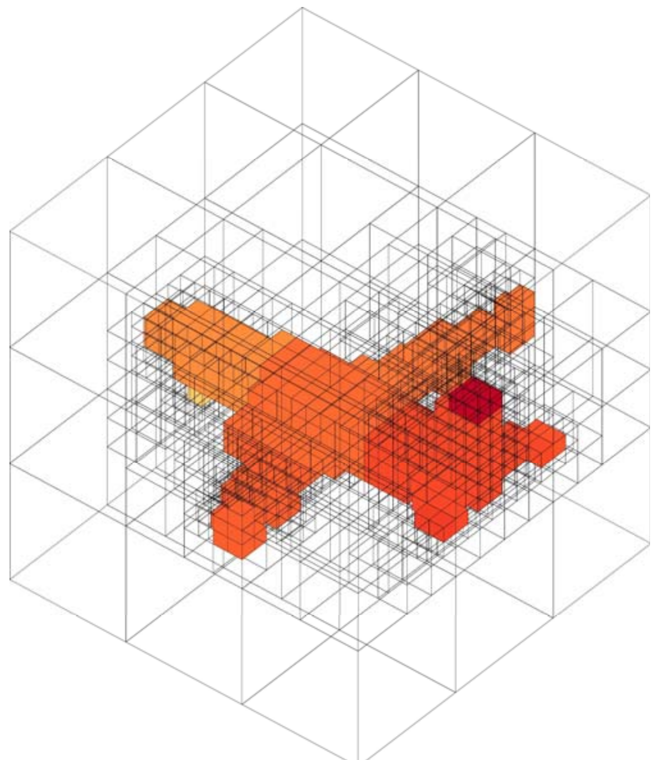
Voxelized 256^3
Occupancy 0.31%

[Slides from Riegler et al. OctNet]
OctNet: Learning Deep 3D Representations at High Resolutions, 2017

3D Deep Learning approaches

- The Multi-View approach
- The Point approach
- **The Voxel approach**
  - Dense Volumetric Nets
  - ***Octree Nets***
  - Sparse Tensor Nets
- The Graph approach

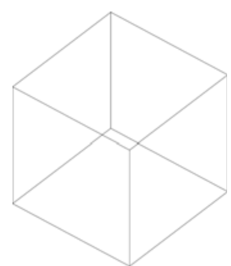
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

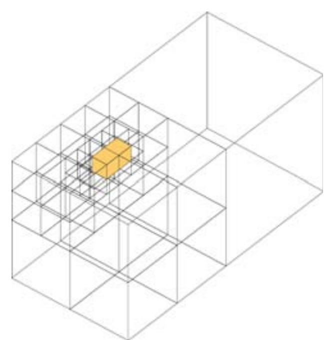
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

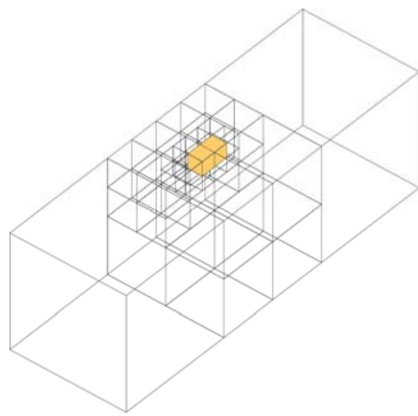
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

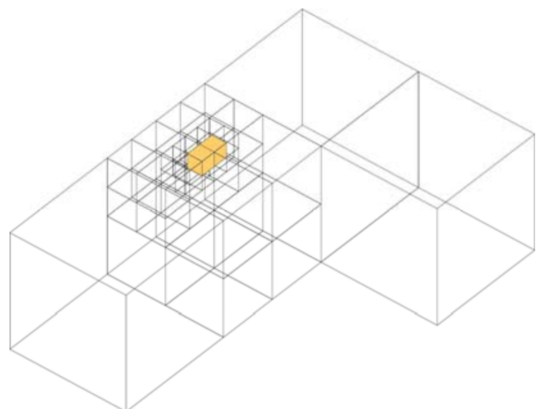
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

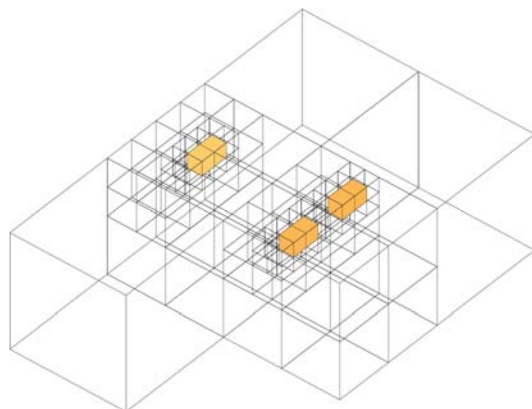
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

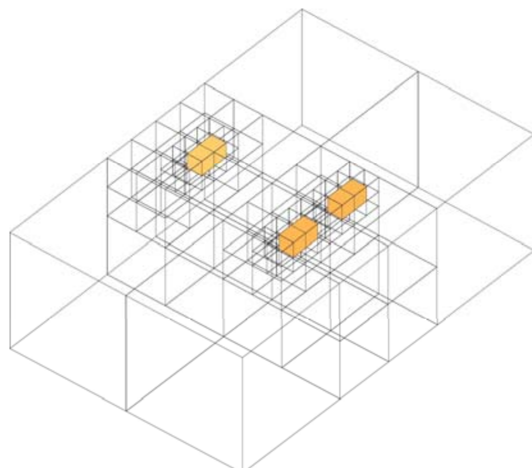
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

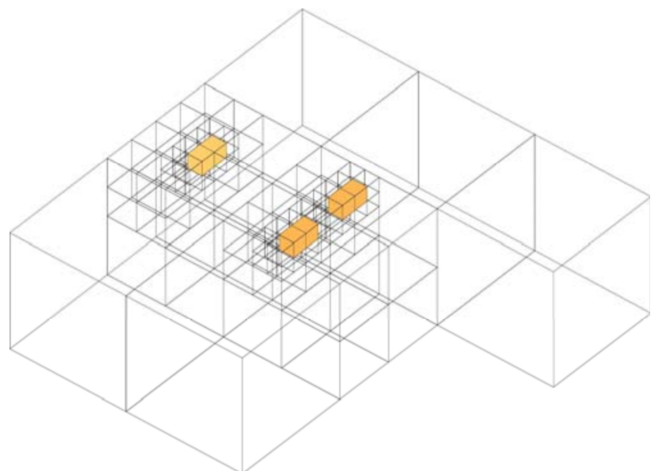
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

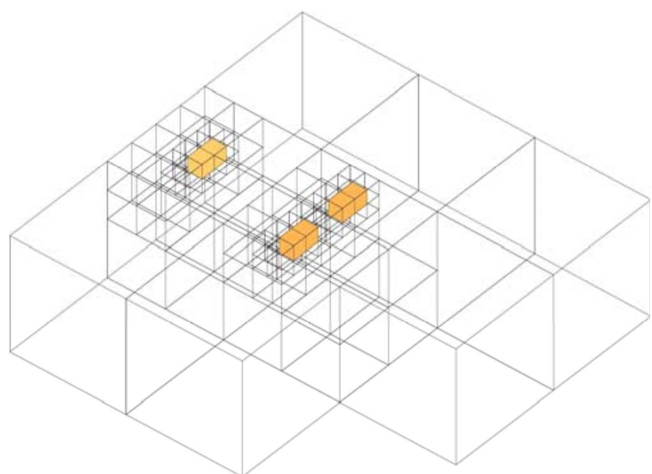
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

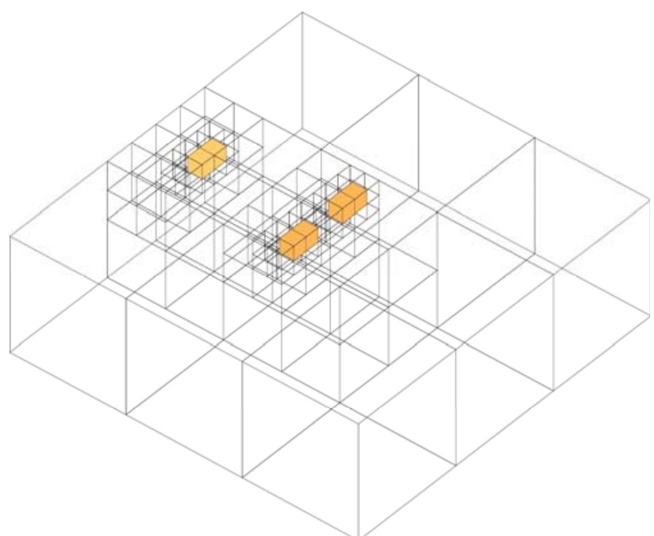
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

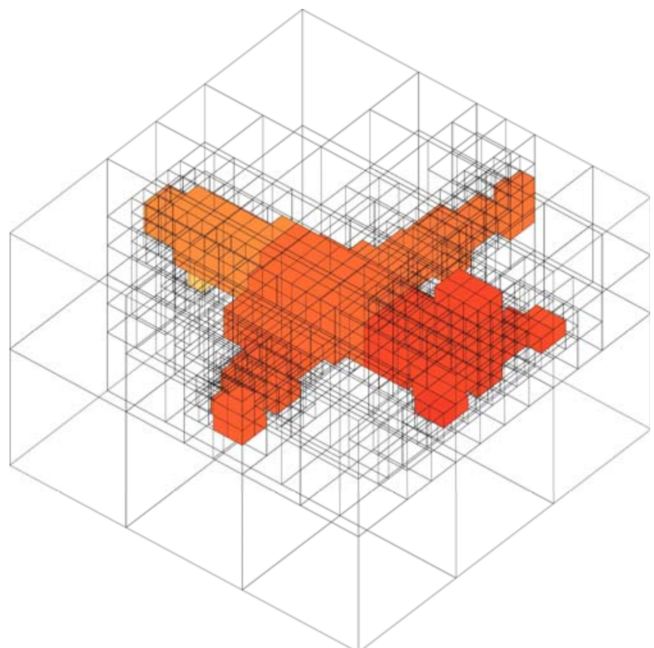
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

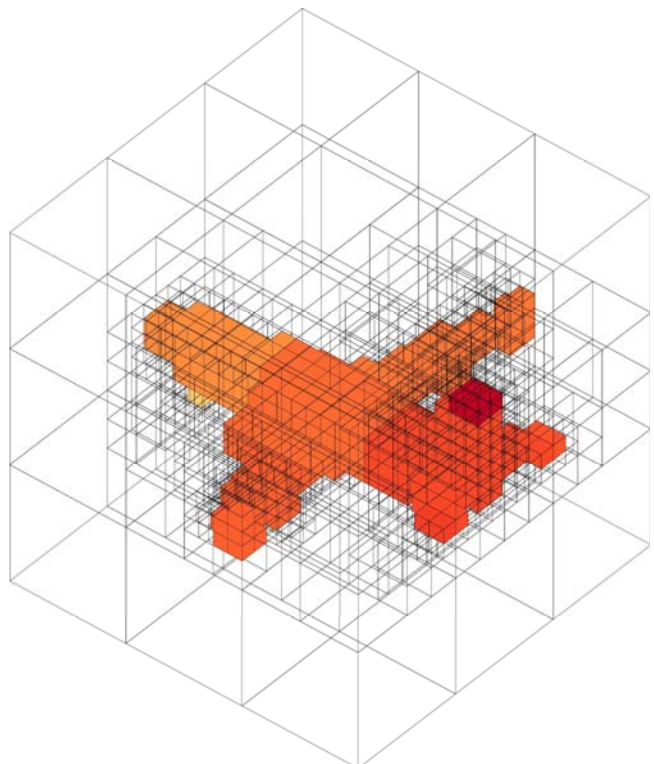
Octrees



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees



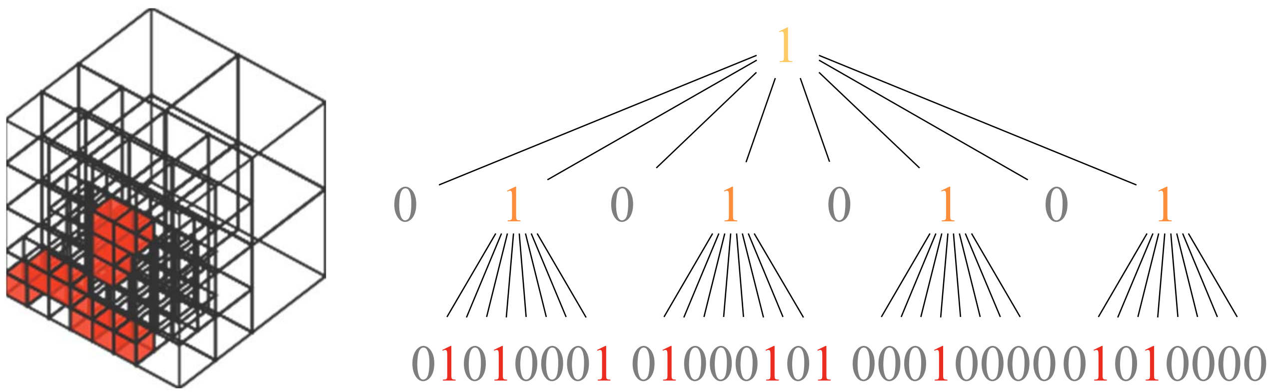
[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: representation

Octrees are efficiently encoded as bit-strings

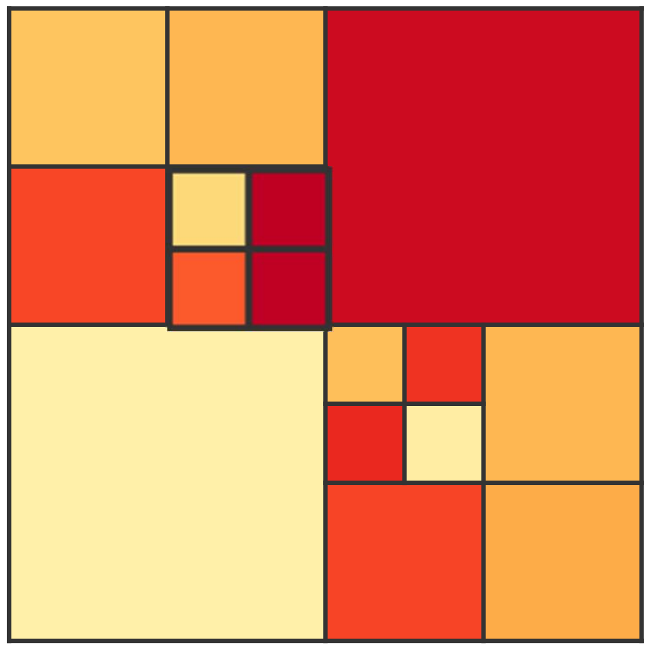
1010101010000000001010001000000000100010100000000000100000
0000000001010000



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

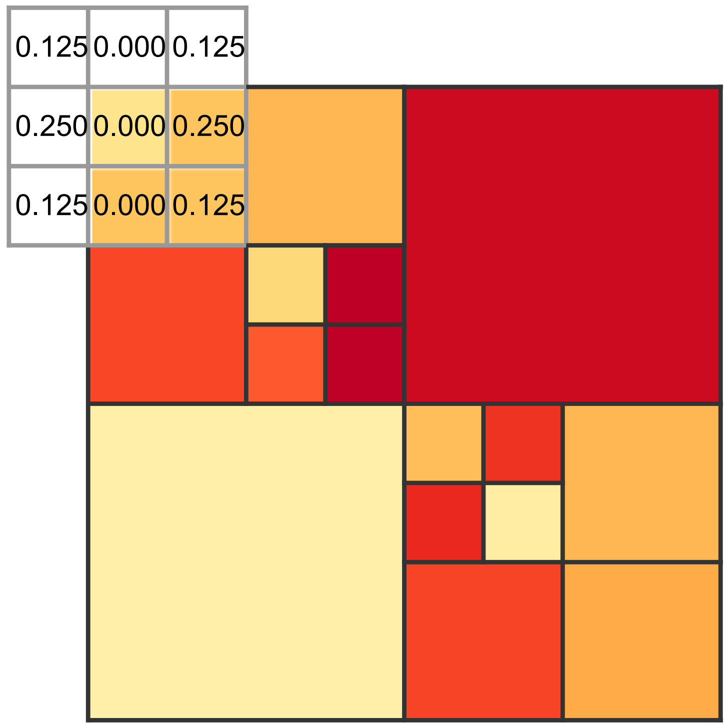
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

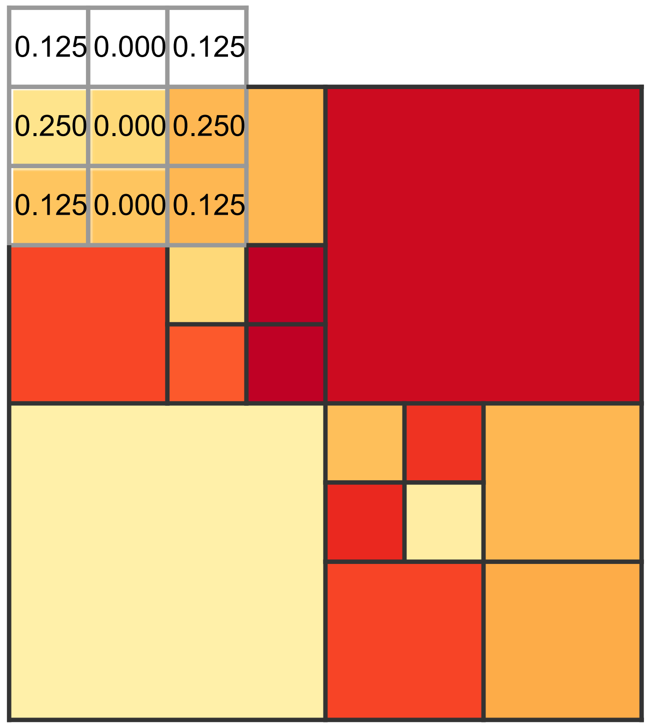
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

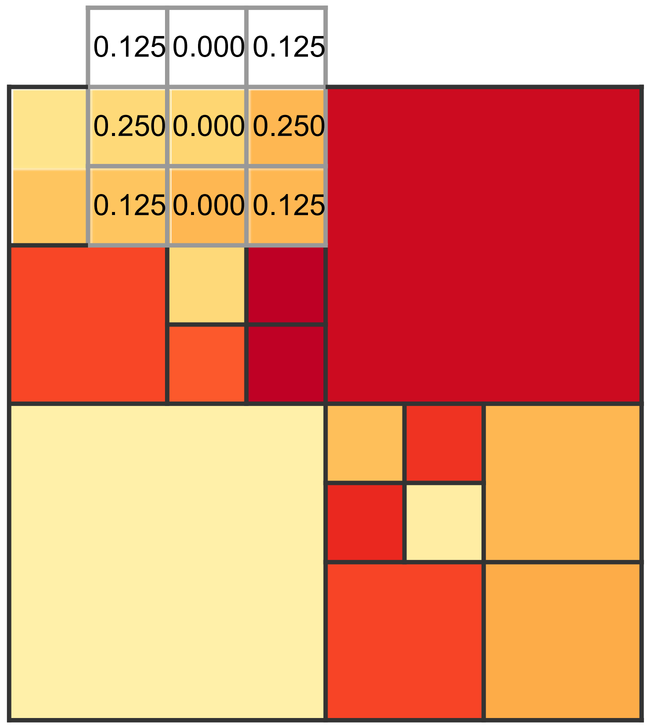
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

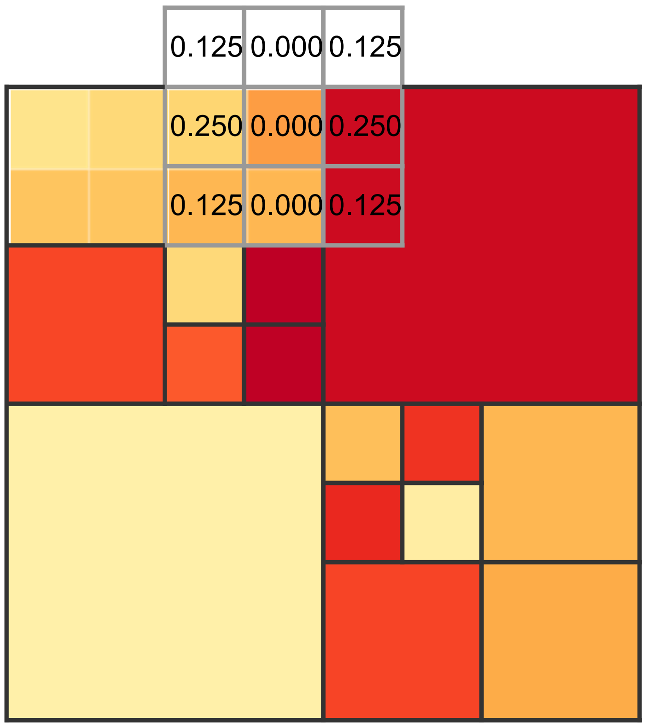
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

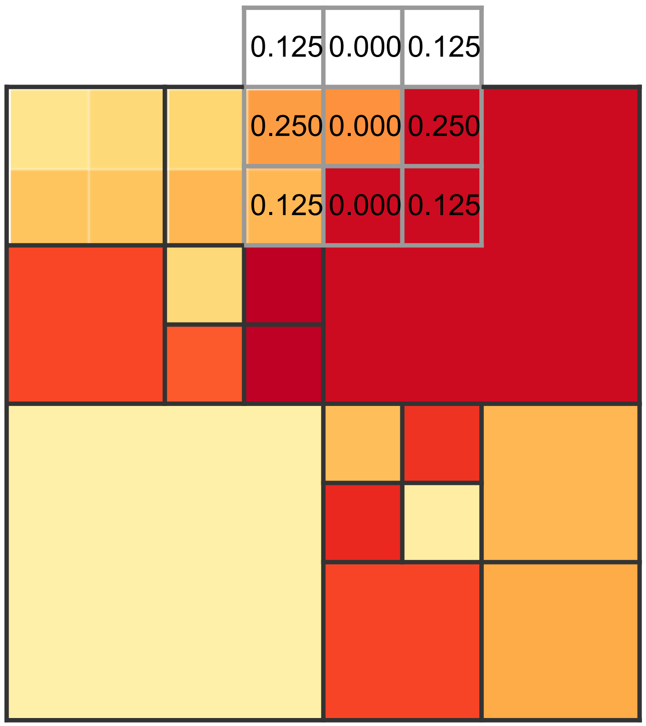
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

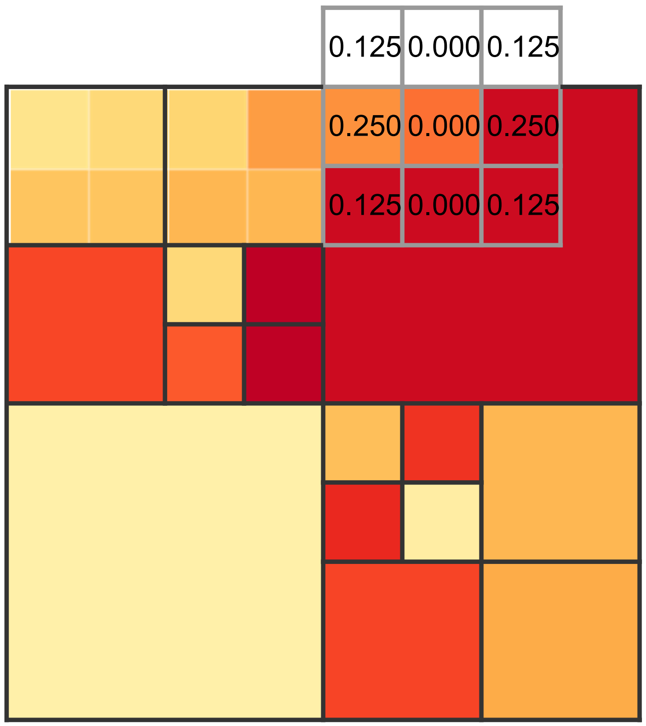
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

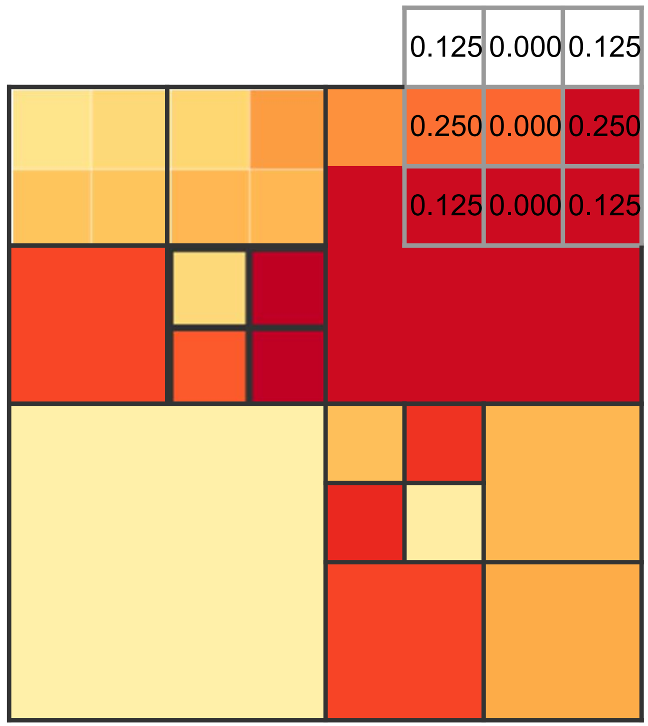
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

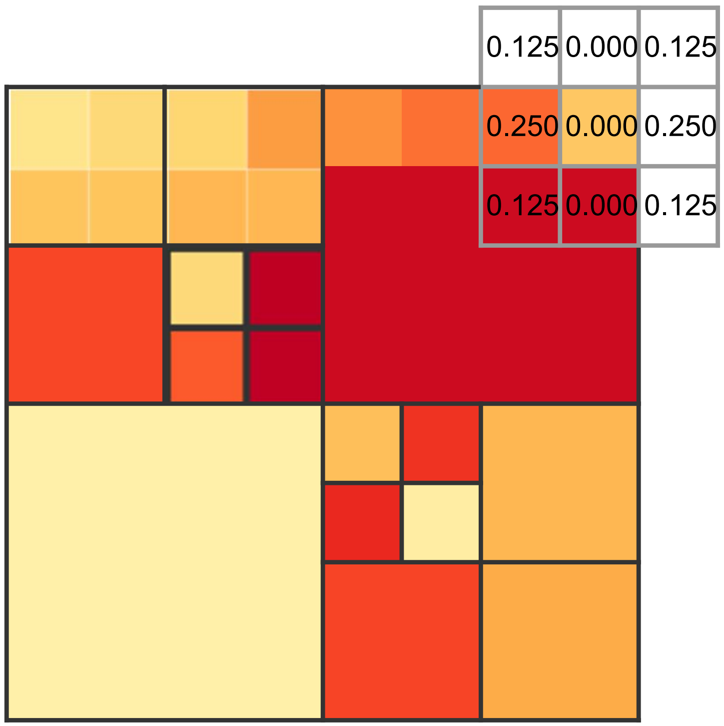
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

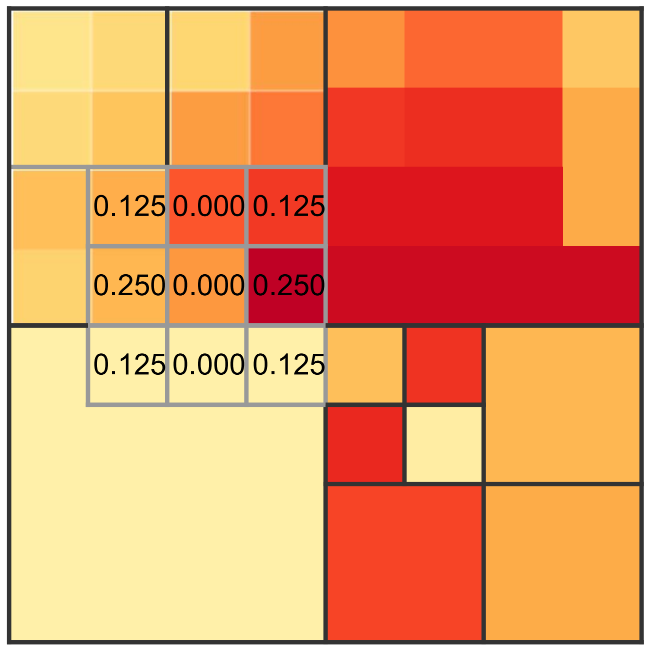
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

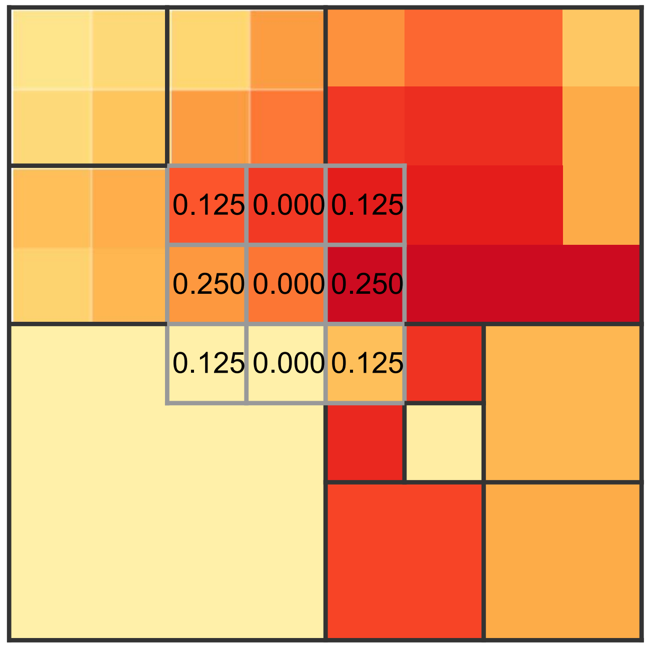
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

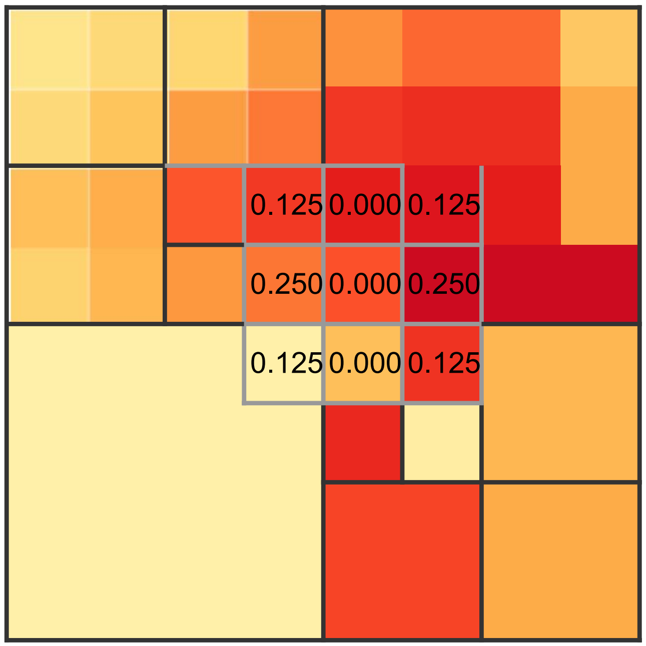
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

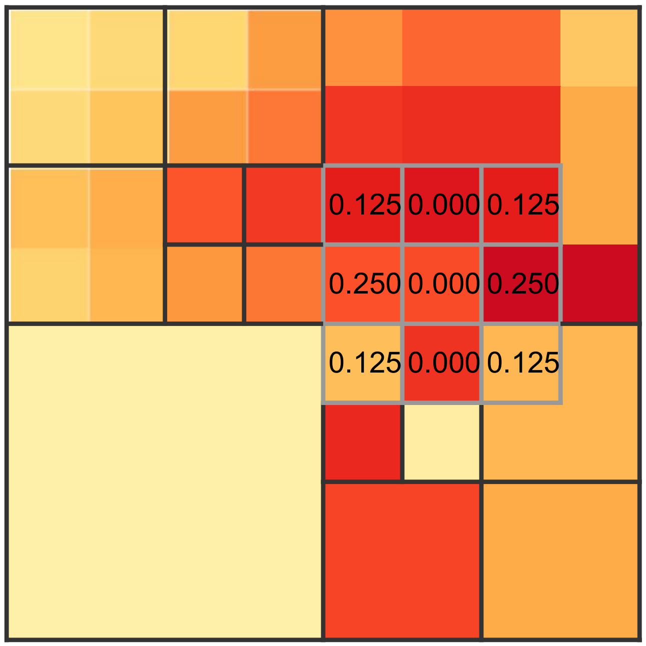
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

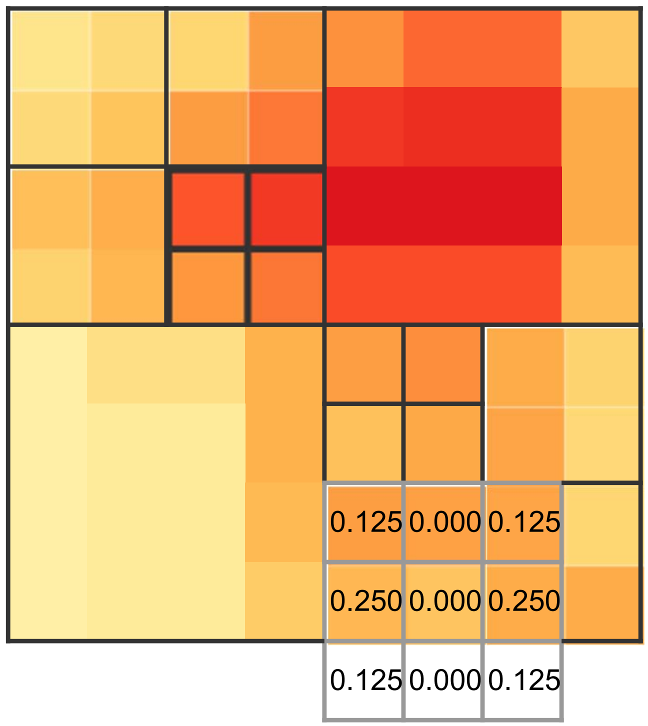
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

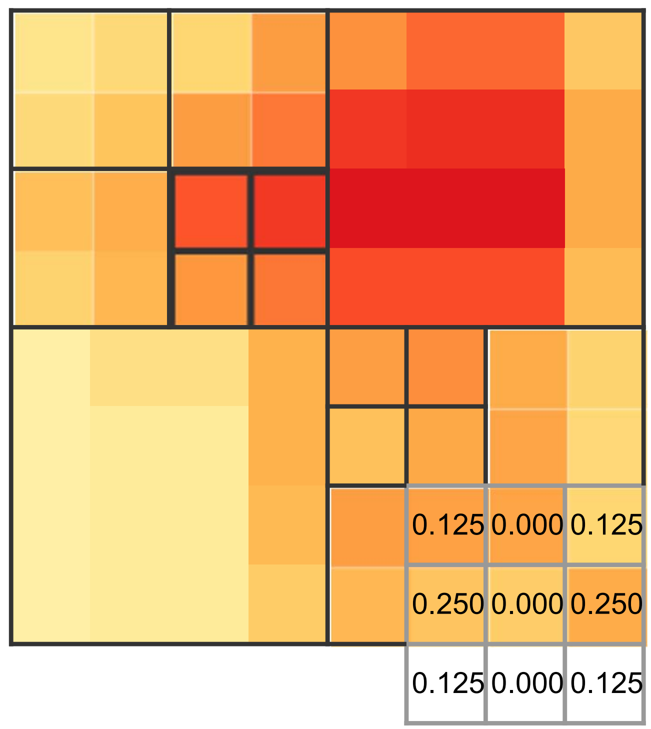
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

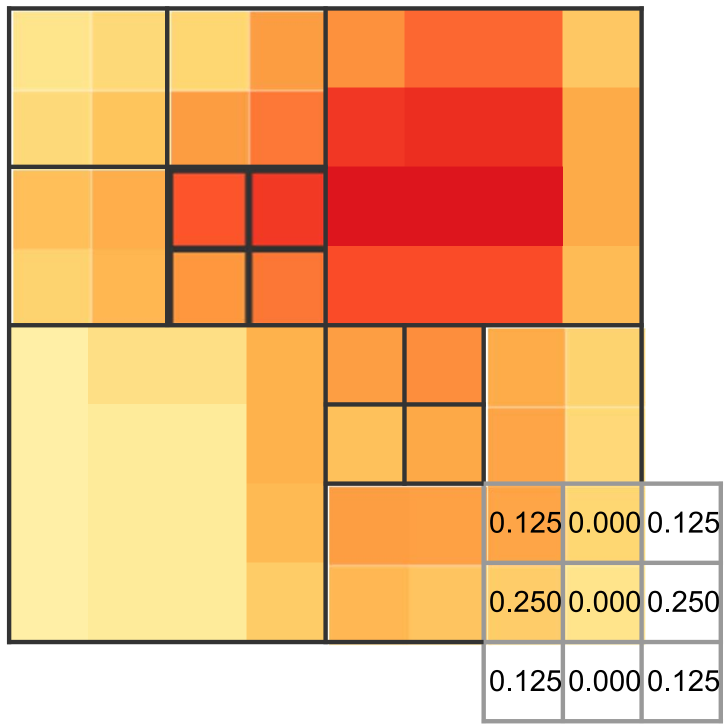
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

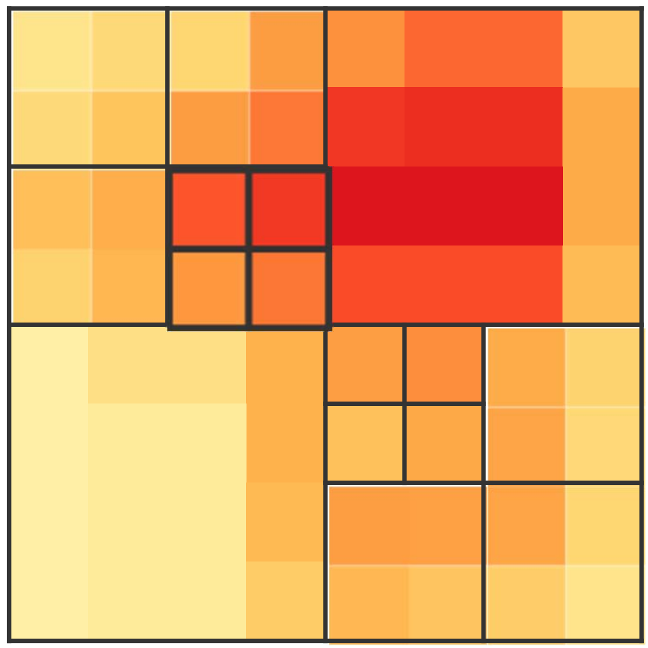
Octrees: convolution (2D example)



[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: convolution (2D example)

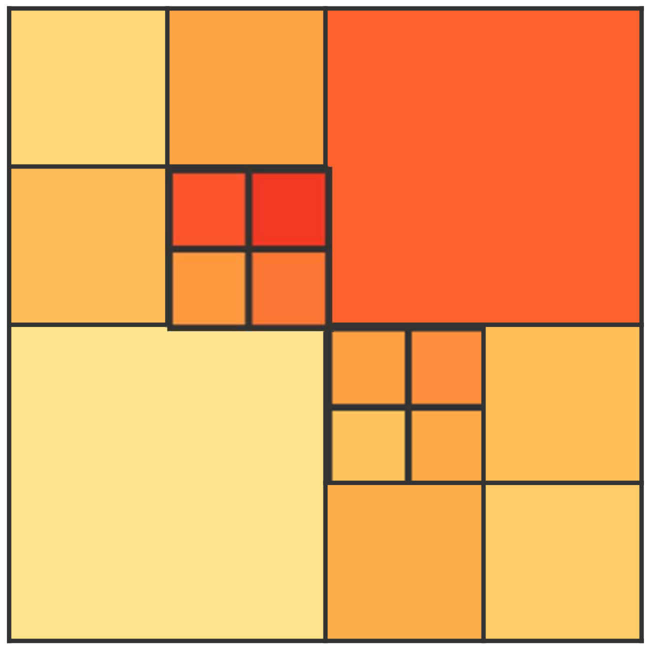


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: convolution (2D example)

Pool responses within each cell (e.g., mean or max-pooling)

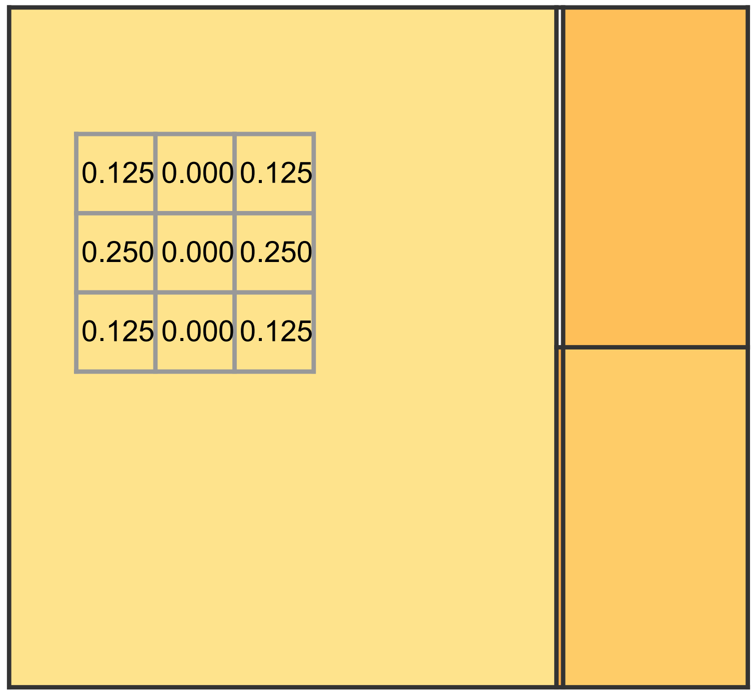


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: convolution (2D example)

For efficiency, convolutions inside the cell need to be done once

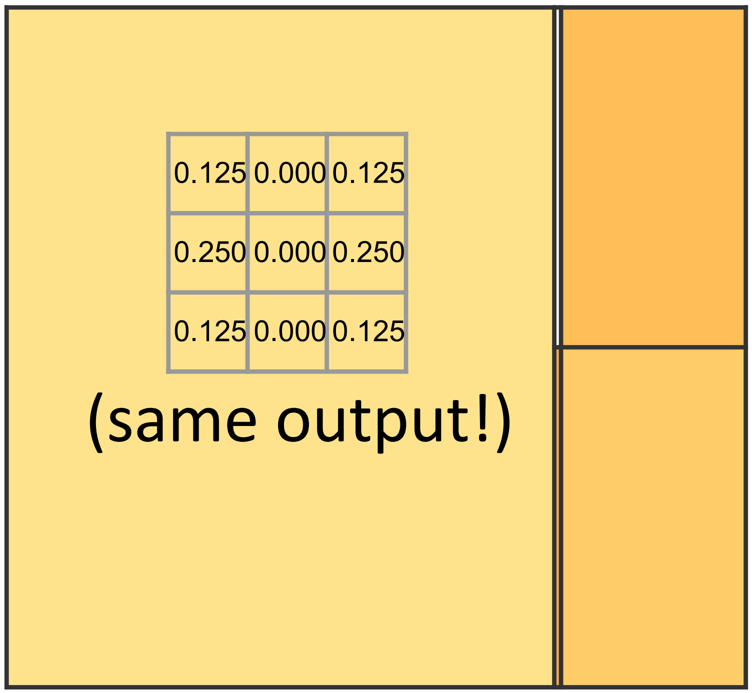


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: convolution (2D example)

For efficiency, convolutions inside the cell need to be done once

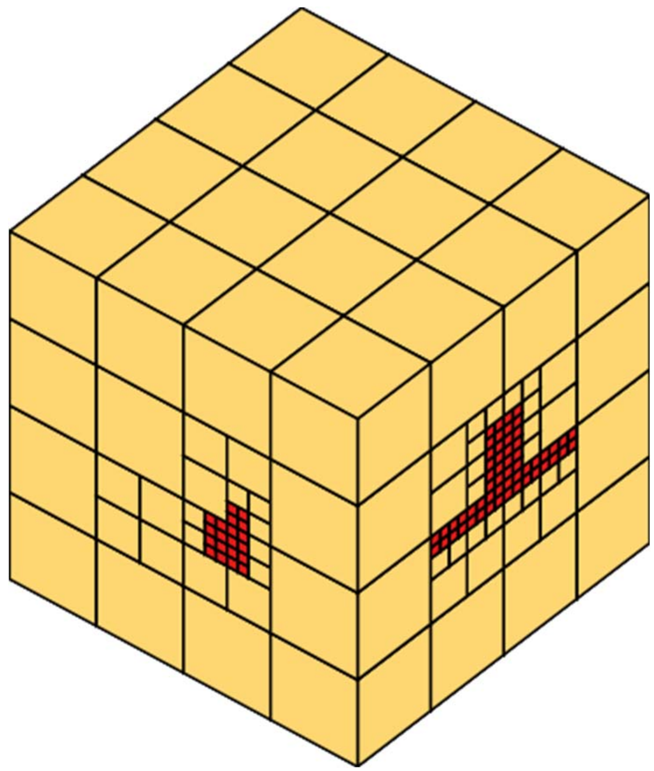


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: pooling

Before pooling

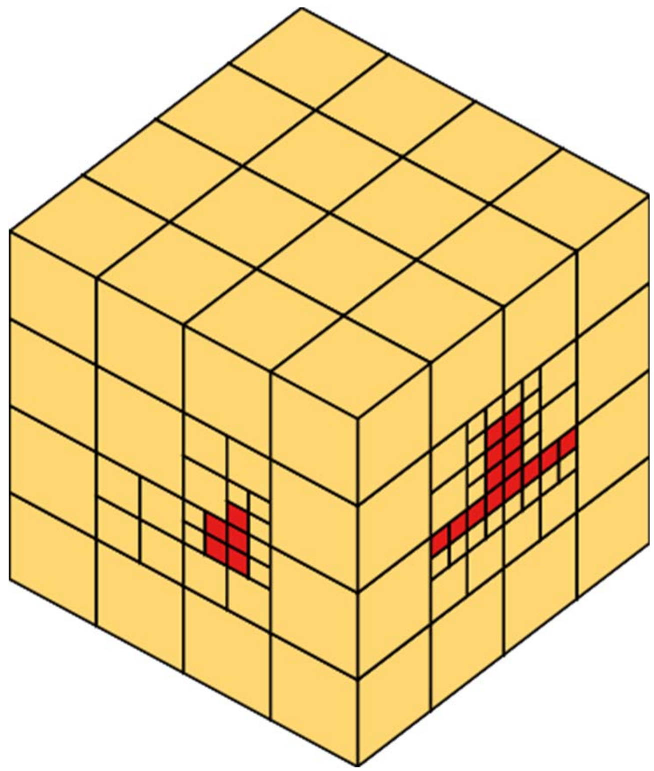


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Octrees: pooling

After pooling

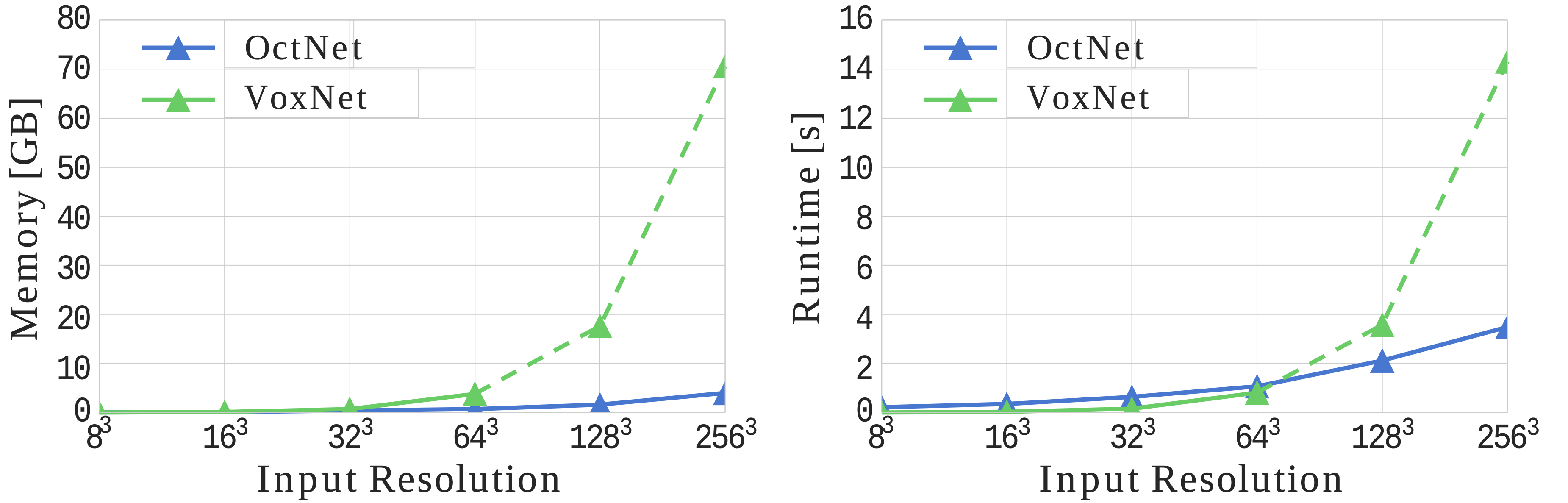


[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Experiments

Memory consumption and runtime of classification network versus dense convolutions (VoxNet)



Network evaluation with batch size 32

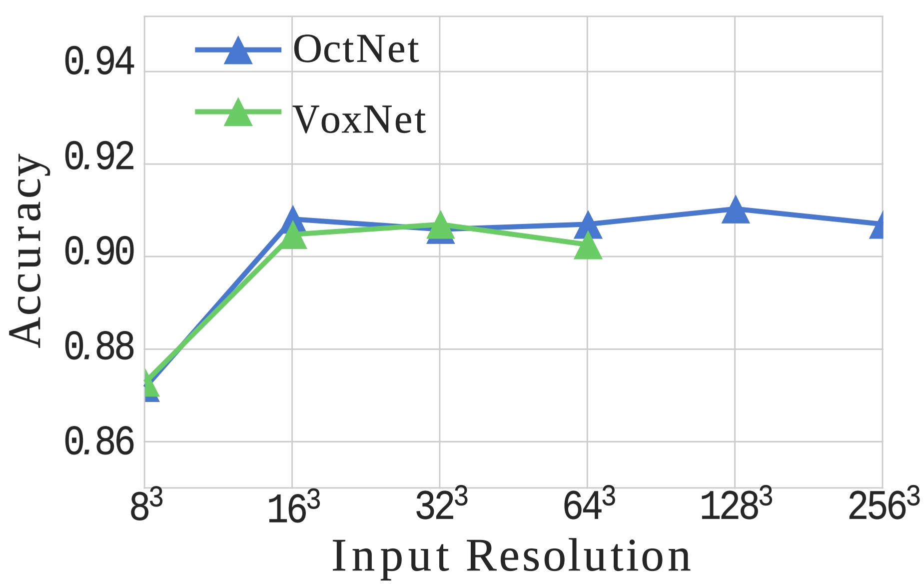
Voxelized ModelNet10 meshes

[Slides from Riegler et al. OctNet]

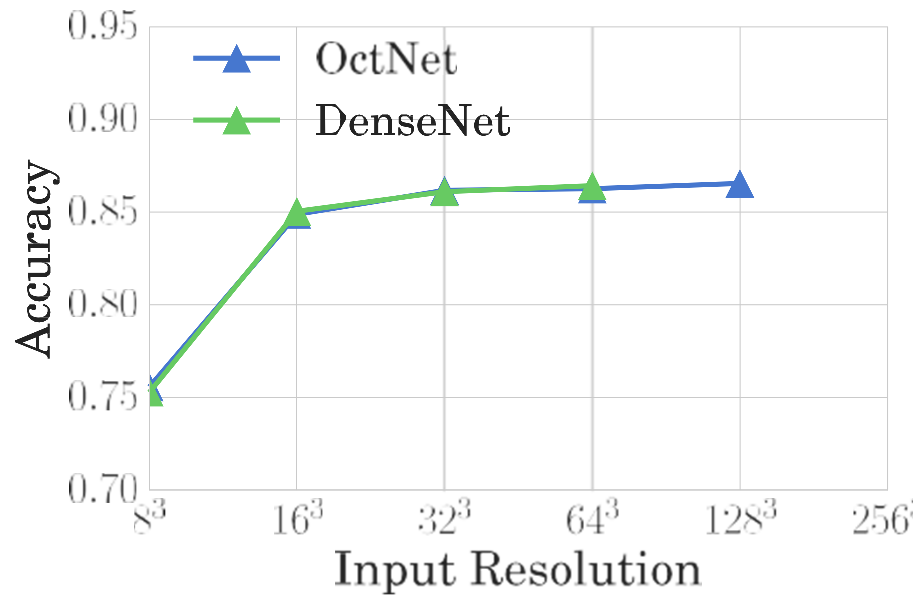
OctNet: Learning Deep 3D Representations at High Resolutions, 2017

Experiments

Accuracy is not affected



ModelNet10



ModelNet40

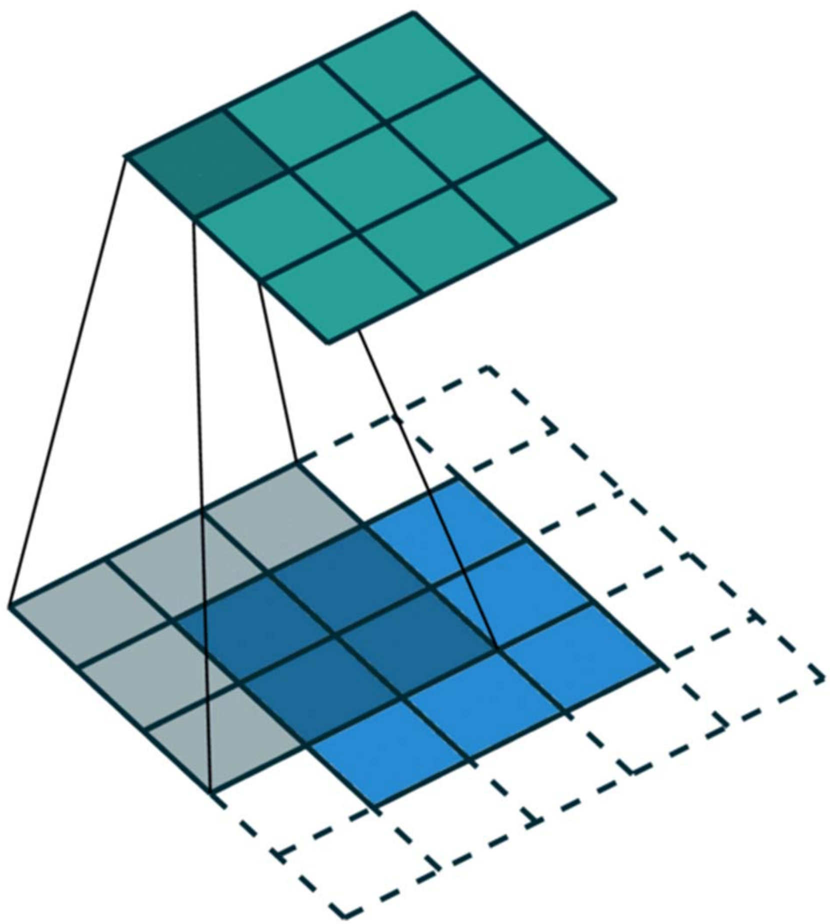
[Slides from Riegler et al. OctNet]

OctNet: Learning Deep 3D Representations at High Resolutions, 2017

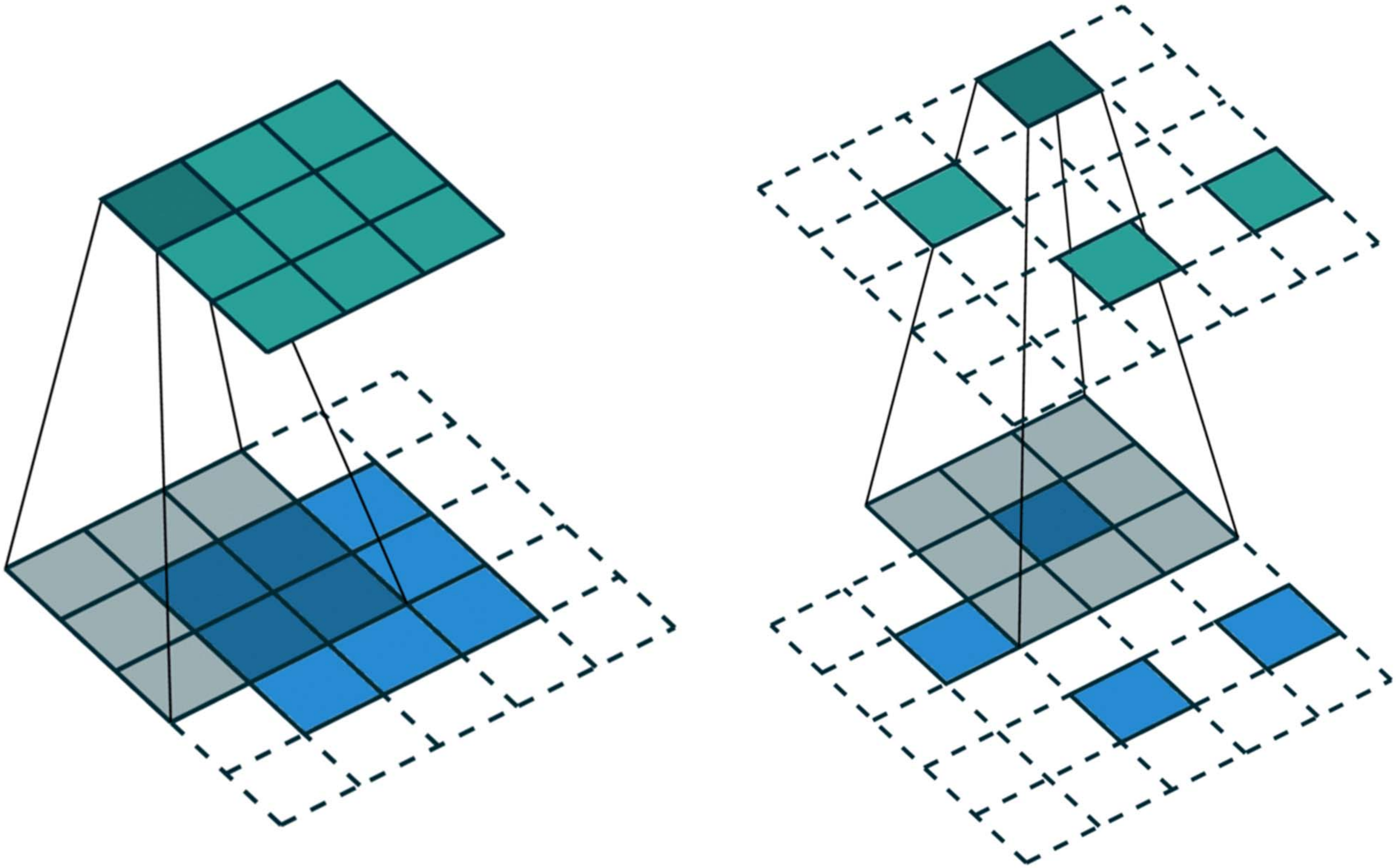
3D Deep Learning approaches

- The Multi-View approach
- **The Voxel approach**
  - *Dense Volumetric Nets*
  - Octree Nets
  - **Sparse Tensor Nets**
- The Point approach
- The Graph approach

Dense Tensor Convolution vs Sparse Tensor Convolution



Dense Tensor Convolution vs Sparse Tensor Convolution

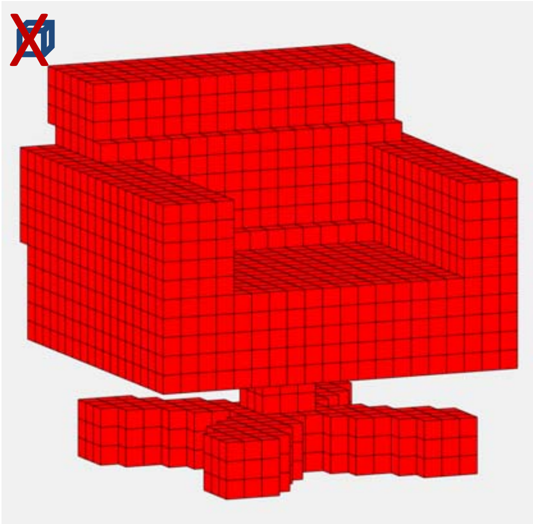


https://nvidia.github.io/MinkowskiEngine/sparse_tensor_network.html

Dense Tensor Convolution vs Sparse Tensor Convolution

Main differences:

a) Do not perform convolution in empty space

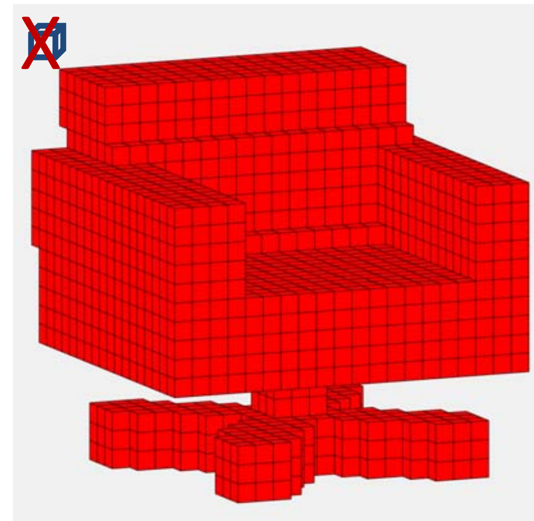


Dense Tensor Convolution vs Sparse Tensor Convolution

Main differences:

a) Do not perform convolution in empty space

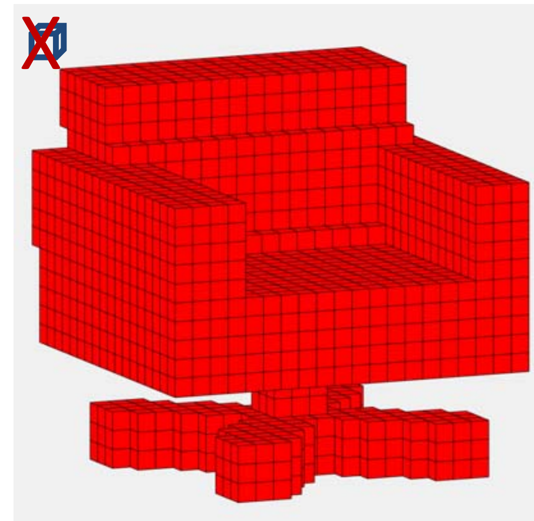
b) Do not store any features in empty space



Dense Tensor Convolution vs Sparse Tensor Convolution

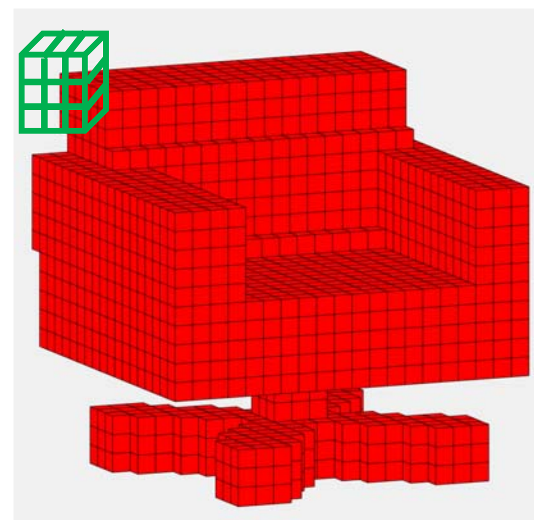
Main differences:

a) Do not perform convolution in empty space



b) Do not store any features in empty space

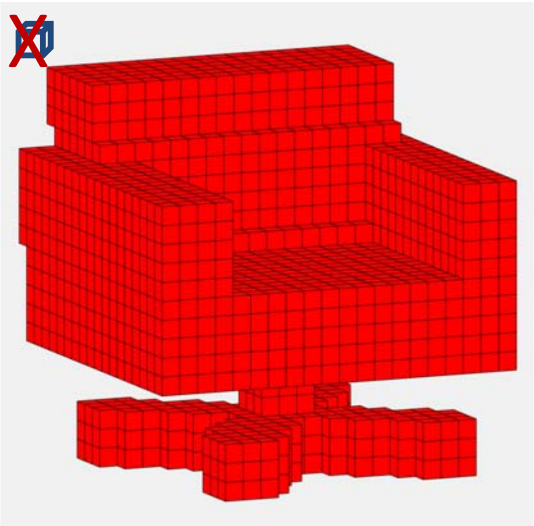
c) While performing convolution, skip multiplying filter weights with empty space



Dense Tensor Convolution vs Sparse Tensor Convolution

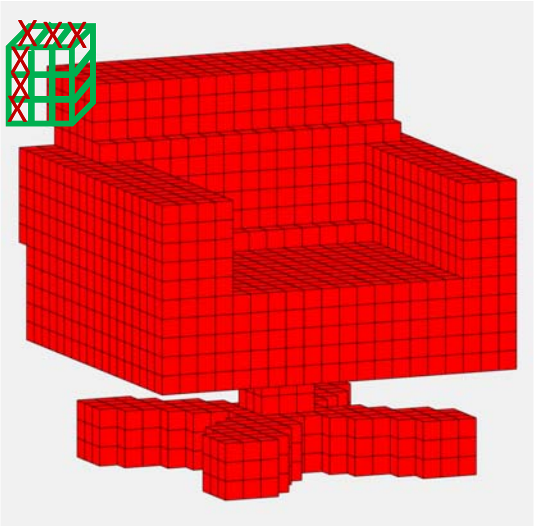
Main differences:

a) Do not perform convolution in empty space



b) Do not store any features in empty space

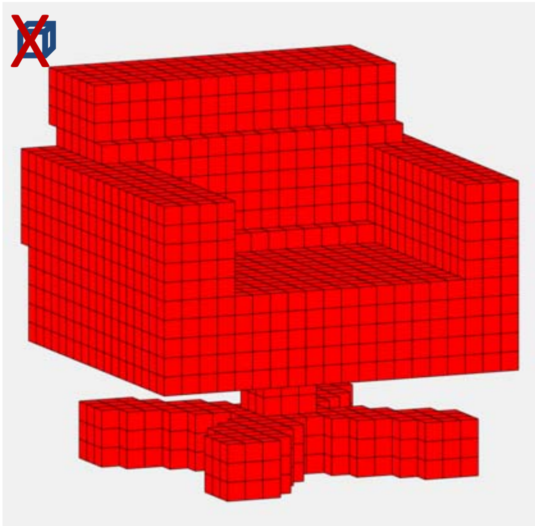
c) While performing convolution, skip multiplying filter weights with empty space



Dense Tensor Convolution vs Sparse Tensor Convolution

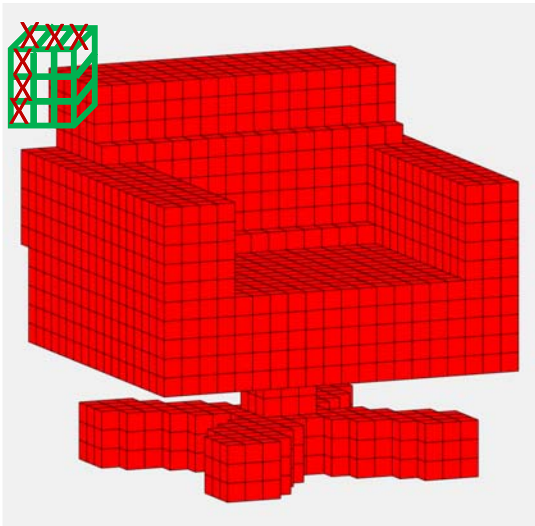
Main differences:

a) Do not perform convolution in empty space



b) Do not store any features in empty space

c) While performing convolution, skip multiplying filter weights with empty space



d) Use sparse representations to store features

e.g. COO format stores a matrix as:

(row1, col1, value1)
(row2, col2, value2)
....

Dense Tensor Convolution vs Sparse Tensor Convolution

$$O(x,y,z,q)=\sum_{k=-n}^{k=n}\sum_{l=-n}^{l=n}\sum_{m=-n}^{n}\sum_{channel\ c} w_q(k,l,m,c)I(x+k,y+l,z+m,c)$$

$$O(x,y,z,q)=\sum_{\{k,l,m\}\in Nb(x,y,z)}\sum_{channel\ c} w_q(k,l,m,c)I(x+k,y+l,z+m,c)$$

This means that you access only offsets that contain non-empty voxels in the neighborhood of voxel at (x,y,z)

Dense Tensor Convolution vs Sparse Tensor Convolution

$$O(x,y,z,q) = \sum_{k=-n}^{k=n} \sum_{l=-n}^{l=n} \sum_{m=-n}^{n} \sum_{channel\ c} w_q(k,l,m,c) I(x+k, y+l, z+m, c)$$

$$\downarrow$$

$$O(x,y,z,q) = \sum_{\{k,l,m\} \in Nb(x,y,z)} \sum_{channel\ c} w_q(k,l,m,c) I(x+k, y+l, z+m, c)$$

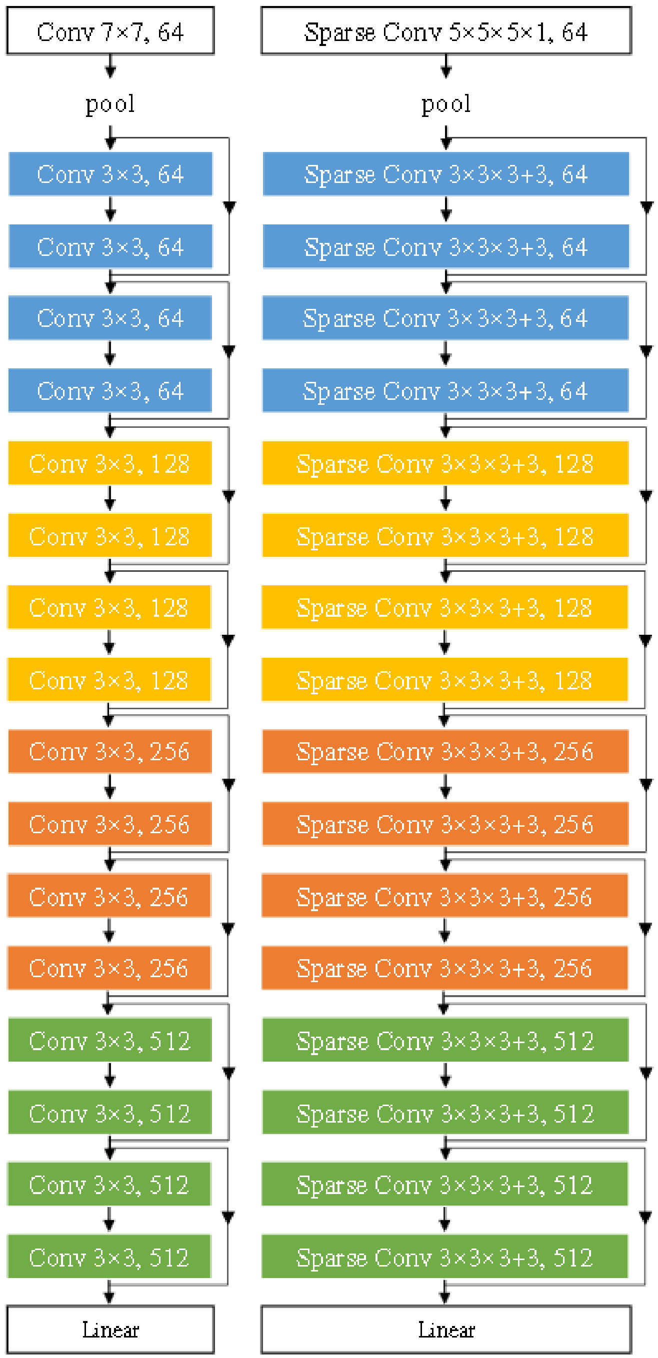
This means that you access
only offsets that
contain non-empty voxels
in the neighborhood of
voxel at (x,y,z)

See also:
4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks, CVPR19
for generalization of this convolution operation in any number of dimensions

MinkowskiNet

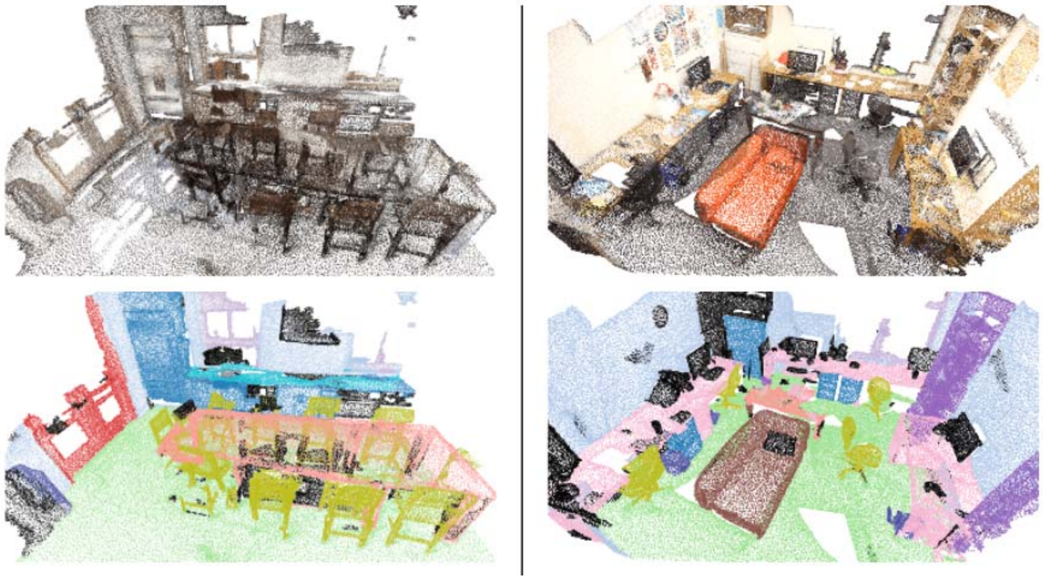
Replaces traditional convolutions (left) with sparse convolutions

Can deal with both 3D and 4D data (uses 4D sparse convolution for point cloud sequences)



MinkowskiNet

Close to state-of-the-art performance for 3D/4D scene labeling



Method	mIOU
ScanNet [5]	30.6
SSC-UNet [10]	30.8
PointNet++ [24]	33.9
ScanNet-FTSDF	38.3
SPLATNet [29]	39.3
TangetConv [30]	43.8
SurfaceConv [21]	44.2
3DMV‡ [6]	48.4
3DMV-FTSDF‡	50.1
PointNet++SW	52.3
MinkowskiNet42 (5cm)	**67.9**

ScanNet predictions

Implementation:

https://nvidia.github.io/MinkowskiEngine/

Paper:

4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks, CVPR19

Volumetric 3D Deep Learning Advantages

- **Octree and Sparse Tensor Networks offer excellent performance for shape/scene segmentation and labeling**
  (given large enough depth in octrees/high voxel resolution)

- **All 2D convolution/pooling operations and modern network architectures (residual blocks) can be adapted to 3D**

- **Well-suited for analyzing volumetric data**
  (e.g., shapes with interior structure/physical properties)

Volumetric 3D Deep Learning Disadvantages

- **Point clouds/meshes must be voxelized => artifacts**
  (a few details may be lost e.g., several points end up in the same voxel)

- **Voxel resolution (or octree depths) needs to be carefully selected**