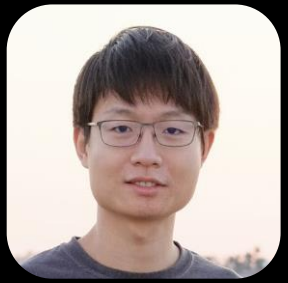
Sparse High-Dimensional and Content-Adaptive Convolutions

Hang Su
UMass Amherst

Varun Jampani
NVIDIA



Hang Su
UMass Amherst

Varun Jampani
NVIDIA

Deqing Sun
Google

Orazio Gallo
NVIDIA

Jan Kautz
NVIDIA

Ming-Hsuan Yang
UC Merced

Erik Learned-Miller
UMass Amherst

Subhransu Maji
UMass Amherst

Evangelos Kalogerakis
UMass Amherst

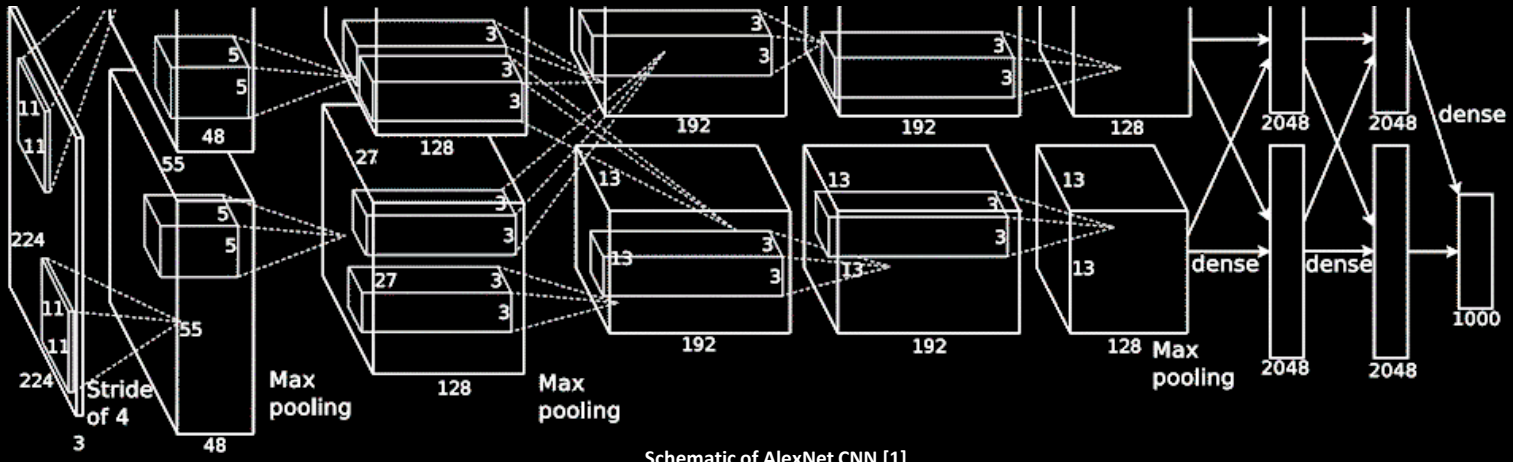
Raghudeep Gadde
Amazon

Martin Kiefel
Amazon

Peter Gehler
Amazon

Spatial convolution

- *Simplest*, *fastest* and *most used* way of propagating information
- Basic building block of most CNNs



Schematic of AlexNet CNN [1]

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *NIPS,* 2012.

Spatial convolution

- Weighted sum over local neighborhoods

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{p}_i - \mathbf{p}_j]\mathbf{v}_j + \mathbf{b}$$

output feature — filter weights — position offset — input feature

$*$

$=$

W, e.g. Gaussian filter

$$e^{-\frac{1}{2\sigma^2}\|\mathbf{p}_i - \mathbf{p}_j\|^2}$$

After Gaussian filtering (smooths image)

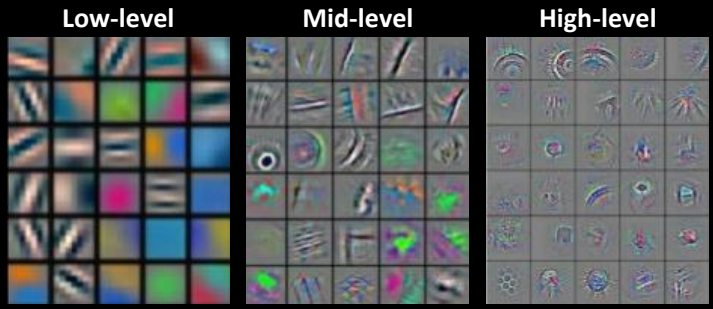
CNNs: w/ learned filters

Spatial convolution

- Weighted sum over local neighborhoods

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{p}_i - \mathbf{p}_j]\mathbf{v}_j + \mathbf{b}$$

- Simple formulation → efficient parallelization
- Spatial sharing → capturing invariance
- Hierarchical features learning

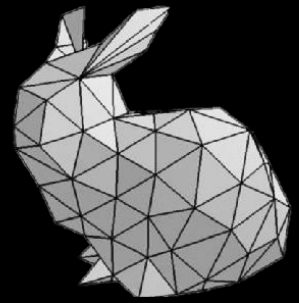


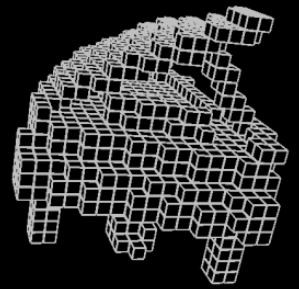
[Zeiler and Fergus, ECCV '13]

Limitations

- **Requires structured inputs**
  - Spatial convolution operates on *regularly structured grids*
  - Data in 3D or higher-dimensional spaces
    - Lack of grid structure
    - Curse of dimensionality

images, text, etc.





Limitations

- **Requires structured inputs**

- **Content-agnostic filters**
  - Loss gradients are shared across pixel locations
  - Once trained, same filter banks are used across different images

Limitations

- **Requires structured inputs**
- **Content-agnostic filters**

Goal: Extend spatial convolution to overcome the limitations while retaining the favorable properties

Bilateral filter

Isn't bilateral filter defined like this??

$$\mathbf{v}_i' = \sum_{j \in \mathcal{N}} f(\|I_i - I_j\|) g(\|\mathbf{p}_i - \mathbf{p}_j\|) \mathbf{v}_j$$

- Generalization of spatial filtering to arbitrary features [1, 2]

$$\mathbf{v}_i' = \sum_{j \in \mathcal{N}} W(\mathbf{p}_i - \mathbf{p}_j) \mathbf{v}_j \quad \longrightarrow \quad \mathbf{v}_i' = \sum_{j \in \mathcal{N}} W(\mathbf{f}_i - \mathbf{f}_j) \mathbf{v}_j$$

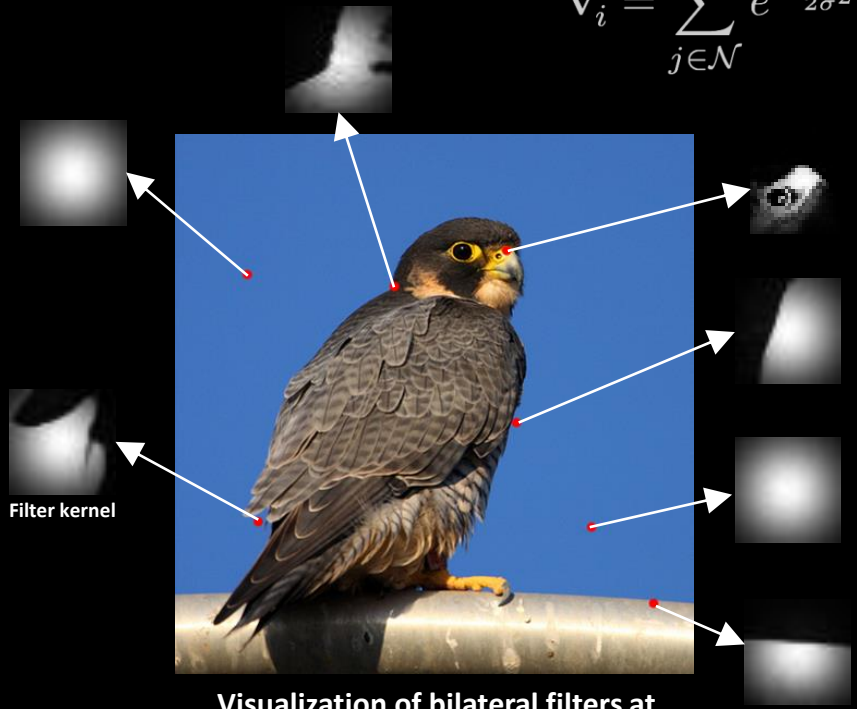
- Features for $\mathbf{f}$ are position and color: $\mathbf{f} = (x, y, r, g, b)$

- Kernel function is Gaussian: $W(\cdot) = e^{-\frac{1}{2}\|\cdot\|^2}$

1. Aurich, V., & Weule, J. Non-linear Gaussian filters performing edge preserving diffusion. In *Mustererkennung,* 1995.
2. Tomasi, C., & Manduchi, R. Bilateral filtering for gray and color images. In *ICCV,* 1998.

Bilateral filter depends on the image content

$$\mathbf{v}_i' = \sum_{j \in \mathcal{N}} e^{-\frac{1}{2\sigma^2}||\mathbf{f}_i - \mathbf{f}_j||^2} \mathbf{v}_j$$



Visualization of bilateral filters at different input locations

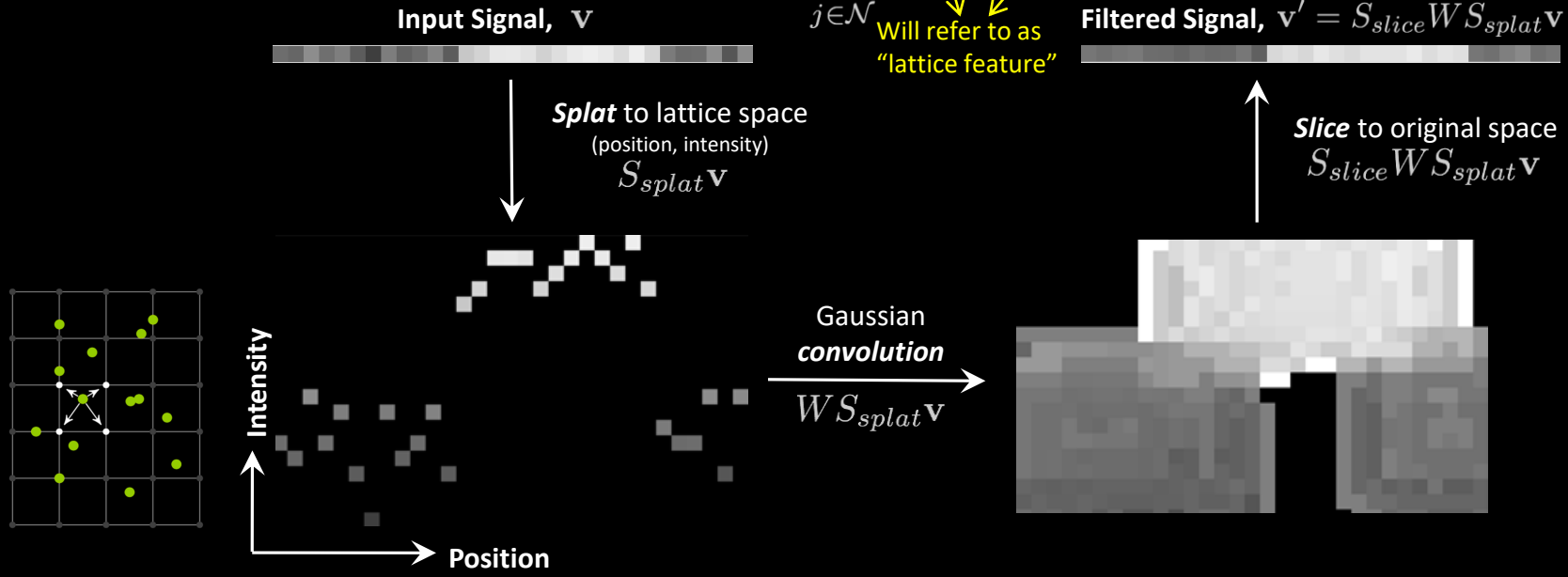
After bilateral filtering (preserves edges)

High-dimensional Gaussian filter

- Bilateral filter as convolution in high-dimensional feature space [1]

$$\mathbf{v}'_i = \sum_{j \in \mathcal{N}} W(\mathbf{f}_i - \mathbf{f}_j)\mathbf{v}_j$$

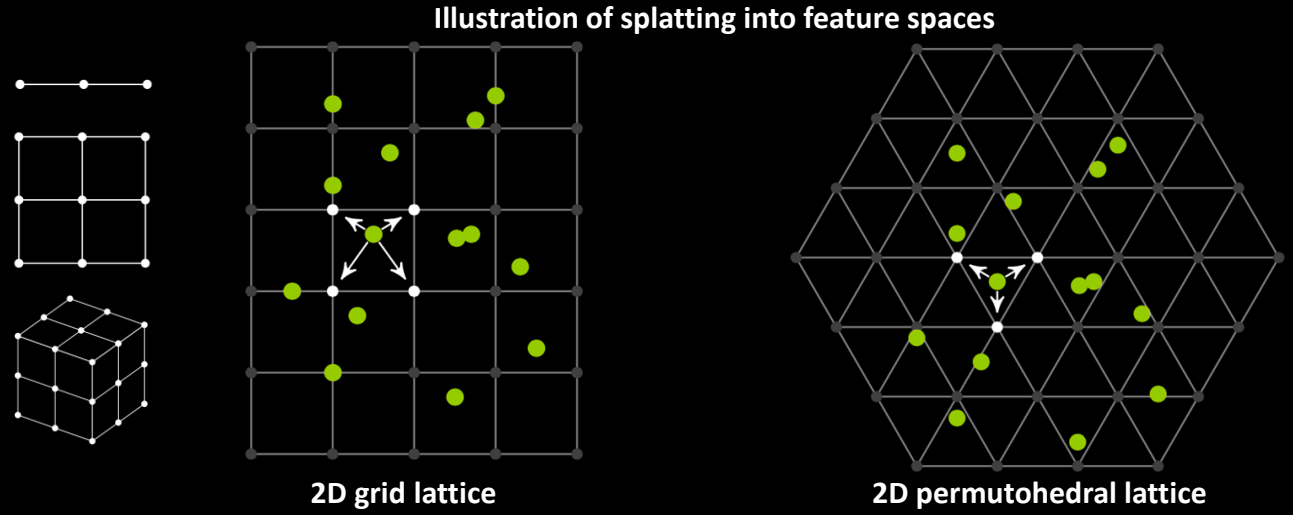
Will refer to as "lattice feature"



1. Paris, S., & Durand, F. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV,* 2006.

N-d grid vs. Permutohedral lattice

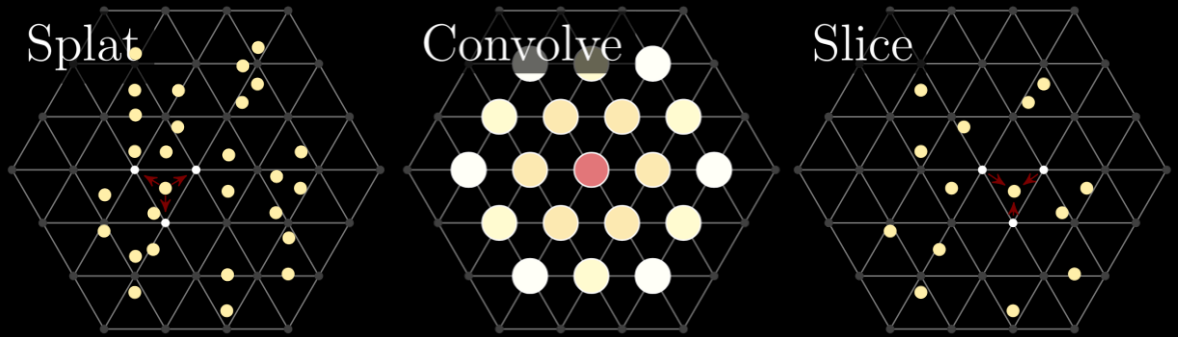
- Grid cell corner points increase exponentially with dimensions (2^d)
- Permutohedral lattice[1] cells have fewer corner points (linear in dimensions - d+1)



1. Adams, A., Baek, J., & Davis, M. A. Fast High-Dimensional Filtering Using the Permutohedral Lattice. In *Computer Graphics Forum*, 2010.

Bilateral filtering with high-dimensional filtering

$$\mathbf{v}' = \boldsymbol{S}_{slice}\boldsymbol{W}\boldsymbol{S}_{splat}\mathbf{v}$$

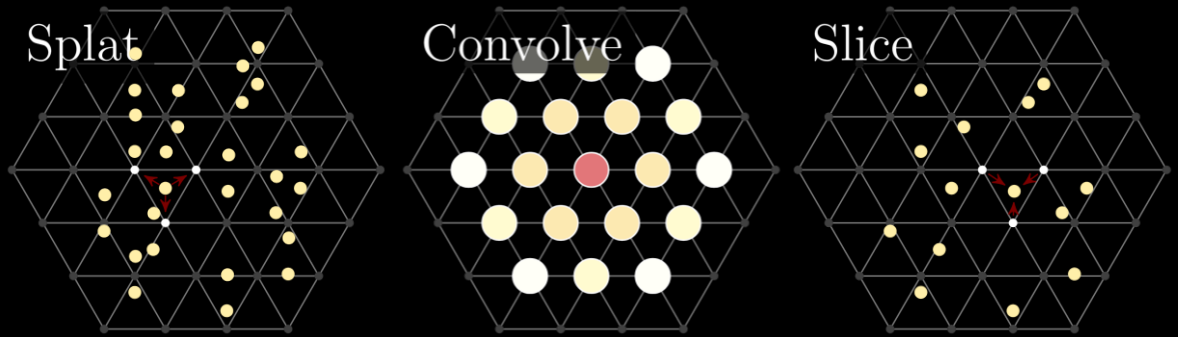


With Gauss filter

Learning the filters [CVPR'16, ICLRw'15]

- Parameterize the filter instead of using Gaussian

$$\mathbf{v}' = \boldsymbol{S}_{slice}\boldsymbol{W}\boldsymbol{S}_{splat}\mathbf{v}$$

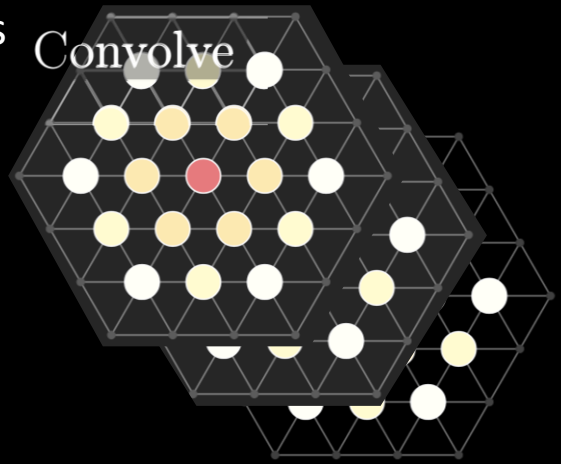


Learn filter weights via back-propagation

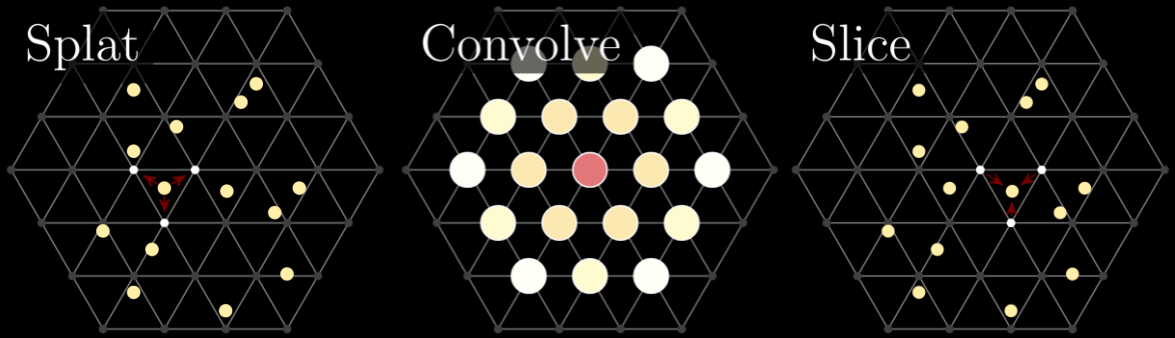
[CVPR'16] Jampani, V., Kiefel, M., & Gehler, P. V. Learning Sparse High Dimensional Filters: Image Filtering, Dense CRFs and Bilateral Neural Networks. *In CVPR,* 2016.
[ICLRw'15] Kiefel, M., Jampani, V., & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

Learning the filters [CVPR'16, ICLRw'15]

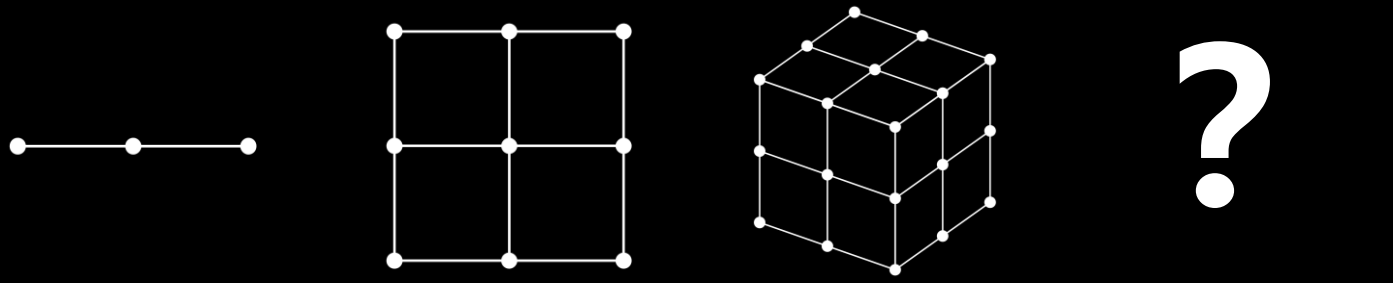
- Learning bilateral filters allows the use of stacked filters
- Randomly initialize and learn end-to-end
- **BCL**: Bilateral Convolution Layer
- **BNN**: Bilateral Neural Network



High-dimensional filter bank



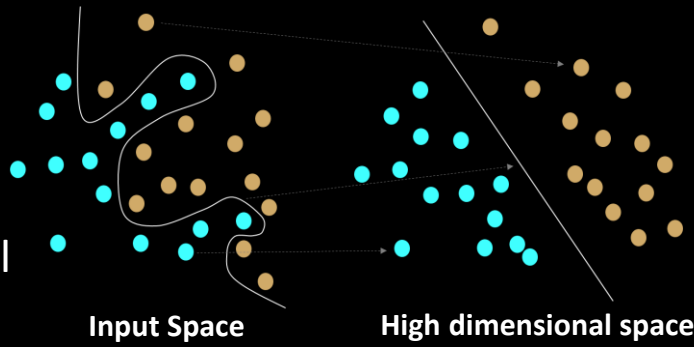
[CVPR'16] Jampani, V., Kiefel, M., & Gehler, P. V. Learning Sparse High Dimensional Filters: Image Filtering, Dense CRFs and Bilateral Neural Networks. *In CVPR,* 2016.
[ICLRw'15] Kiefel, M., Jampani, V., & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

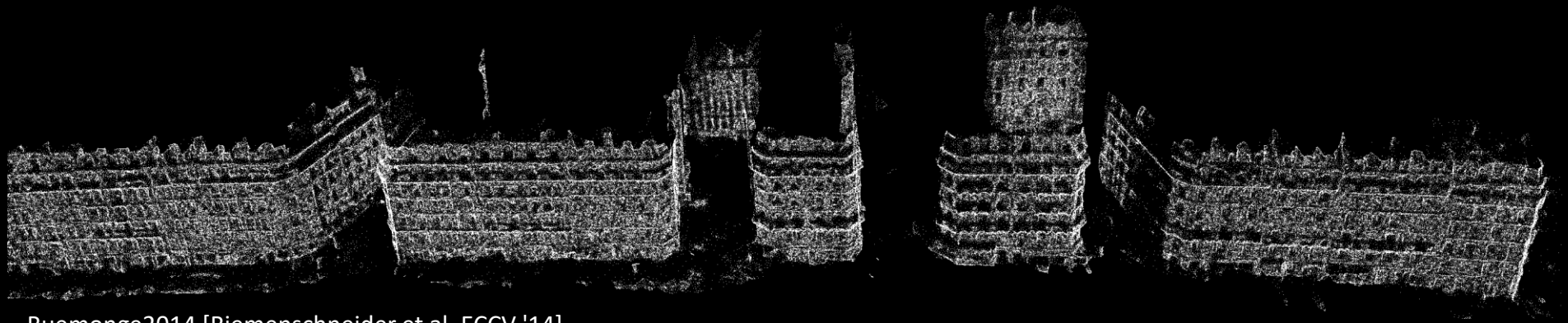


Equip neural networks with richer class of
Sparse High-Dimensional Filters

Why sparse and high-dimensional networks?

- Project data to higher dimensions
  - E.g. Bilateral filter
- Data can be inherently sparse and high-dimensional
  - E.g. 3D data, BRDF material properties, video





Ruemonge2014 [Riemenschneider et al. ECCV '14]

Advantages of BCL

Separate lattice and point features

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{p}_i - \mathbf{p}_j]\mathbf{v}_j \longrightarrow \mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{f}_i - \mathbf{f}_j]\mathbf{v}_j$$

"where" lattice feature

point feature "what"

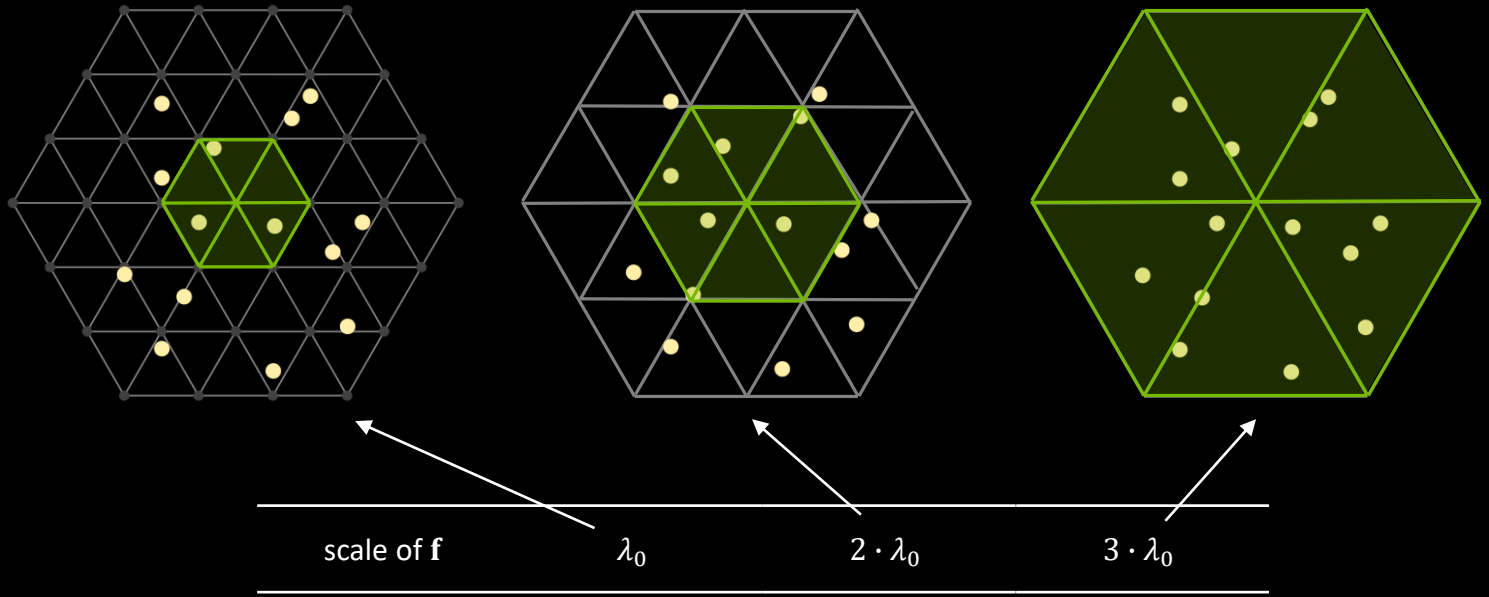
spatial convolution

sparse high-dim. filtering

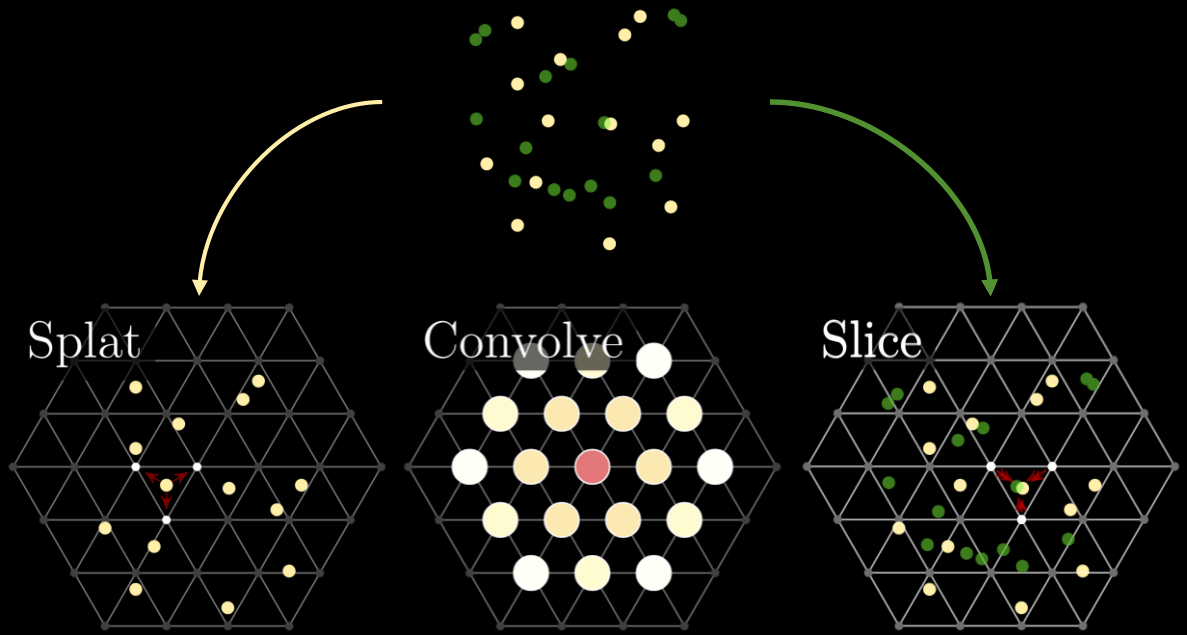
Image	point feature $\mathbf{v}$	(r,g,b)	(r,g,b)	$\mathbf{v}(\dots)$	$\mathbf{v}(\dots)$
	lattice feature $\mathbf{f}$	(x,y)	(x,y,r,g,b)	(x,y)	$(x,y,depth)$
3D Point cloud	point feature $\mathbf{v}$	1	(x,y,z)	(r,g,b)	$\mathbf{v}(\dots)$
	lattice feature $\mathbf{f}$	(x,y,z)	(x,y,z)	(x,y,z,n_x,n_y,n_z)	(x,y,z)

Controlling receptive fields

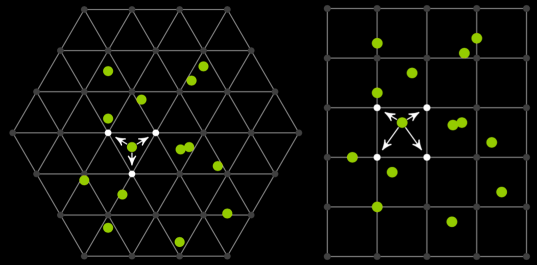
$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{f}_i - \mathbf{f}_j]\mathbf{v}_j$$

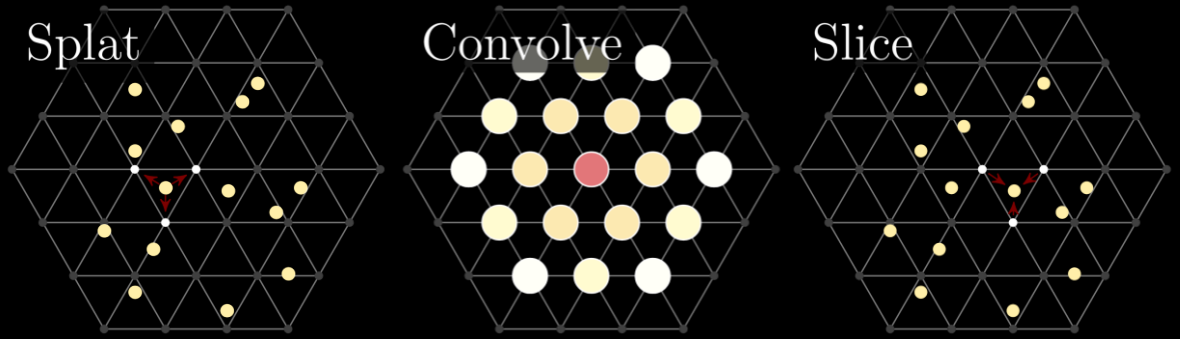


Input and output can be at different points

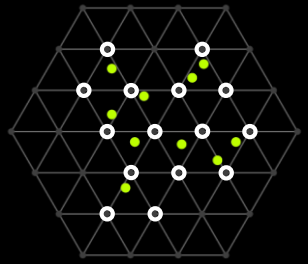


Efficient computation





- Permutohedral lattice[1]
- Hash table



[1] Adams *et al.* Fast high-dimensional filtering using the permutohedral lattice. Computer Graphics Forum '10

Applications in 2D, video and 3D vision

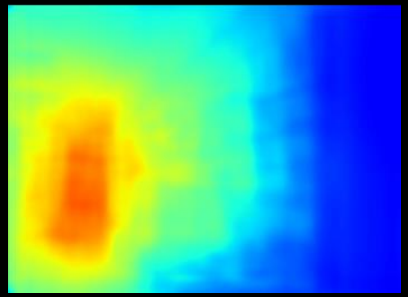
- BCL & BNNs have wide range of applications

- Examples:
  1. Joint image filtering
  2. Video information propagation
  3. 3D point cloud segmentation

Joint image filtering

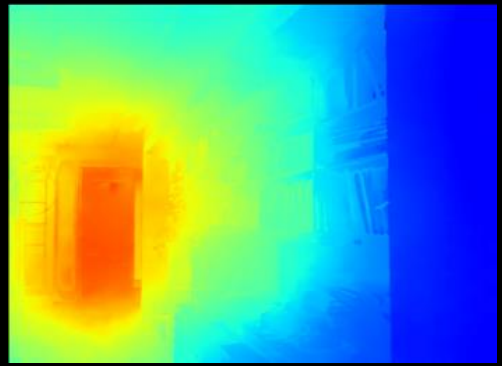
Task: Enhance an input with the help of a given guidance image

Example: Depth upsampling



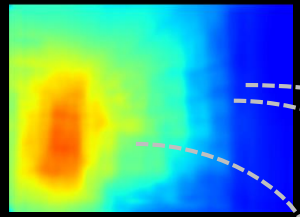
Low-resolution input

Guidance RGB image

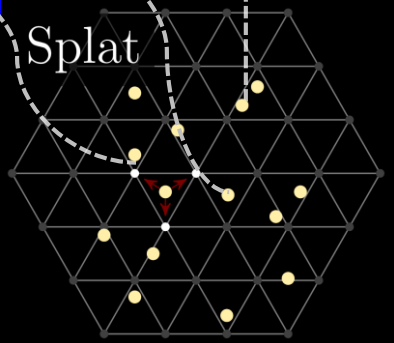


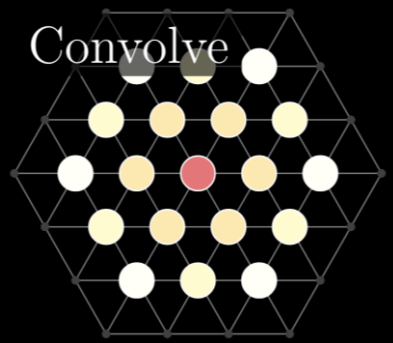
After joint image filtering

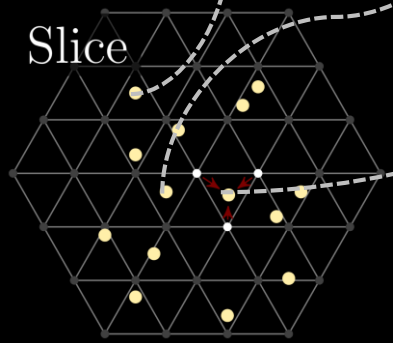
Joint image filtering with BCL



Low-resolution input
(Input features)

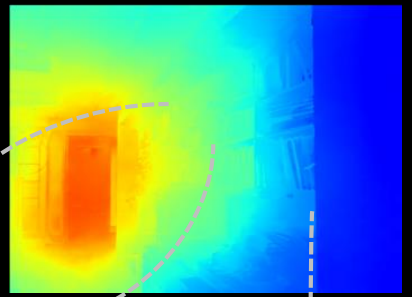






Bilateral Convolution Layer (BCL)

High-resolution output



Guidance image
(Lattice features)

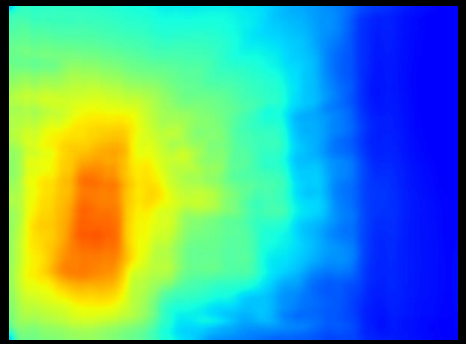
Depth upsampling results

CNN depth estimation results[1] are often low resolution

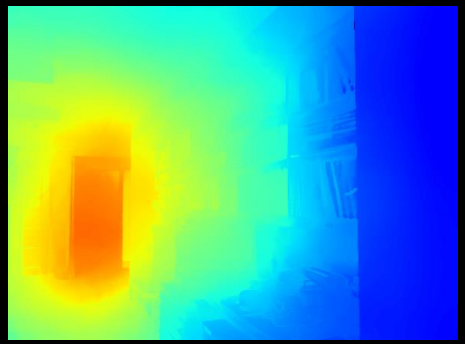
Image (guidance)

RMSE results on NYU depth dataset [2] (8x upsampling)

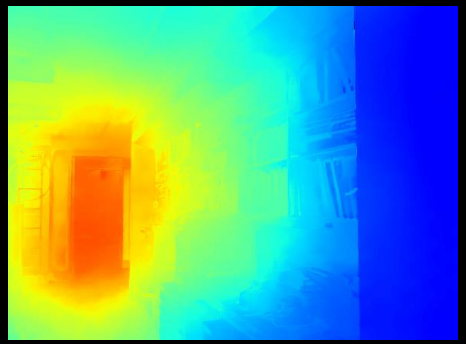
	Bicubic	Gauss Bilateral	Learnt Bilateral
Test set	0.753	0.752	**0.748**



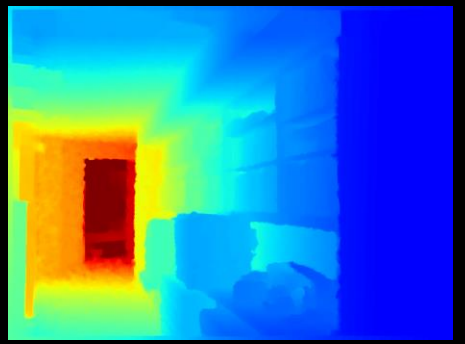
CNN depth (bicubic interpolated)



Gauss bilateral upsampling



Learnt bilateral upsampling



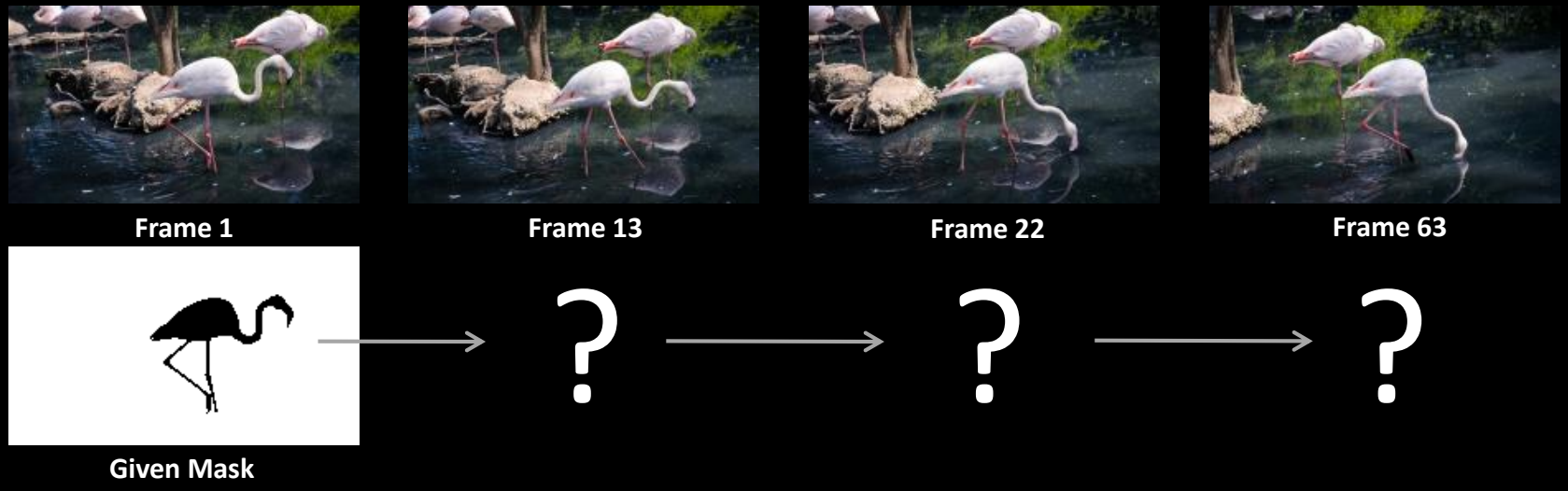
Ground Truth depth

1. Eigen, D., Puhrsch, C., & Fergus, R. Depth map prediction from a single image using a multi-scale deep network. NIPS '14
2. Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from RGBD images. ECCV '12

Video propagation

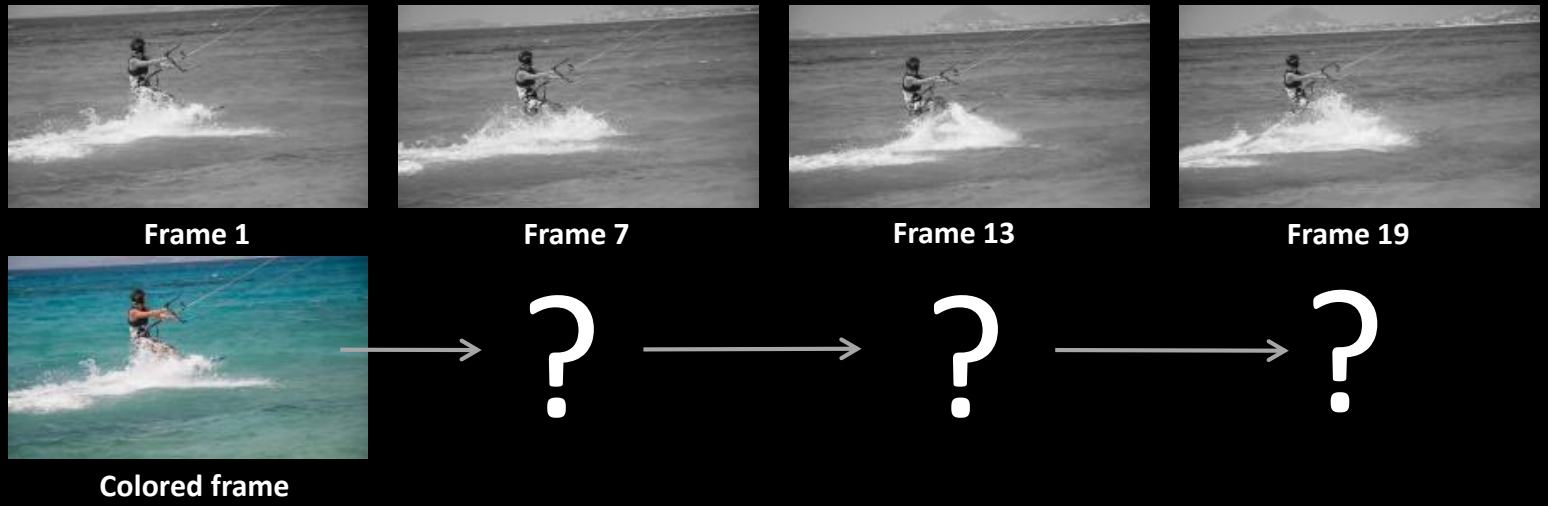
How can we propagate information across video frames?

Example: Video object segmentation



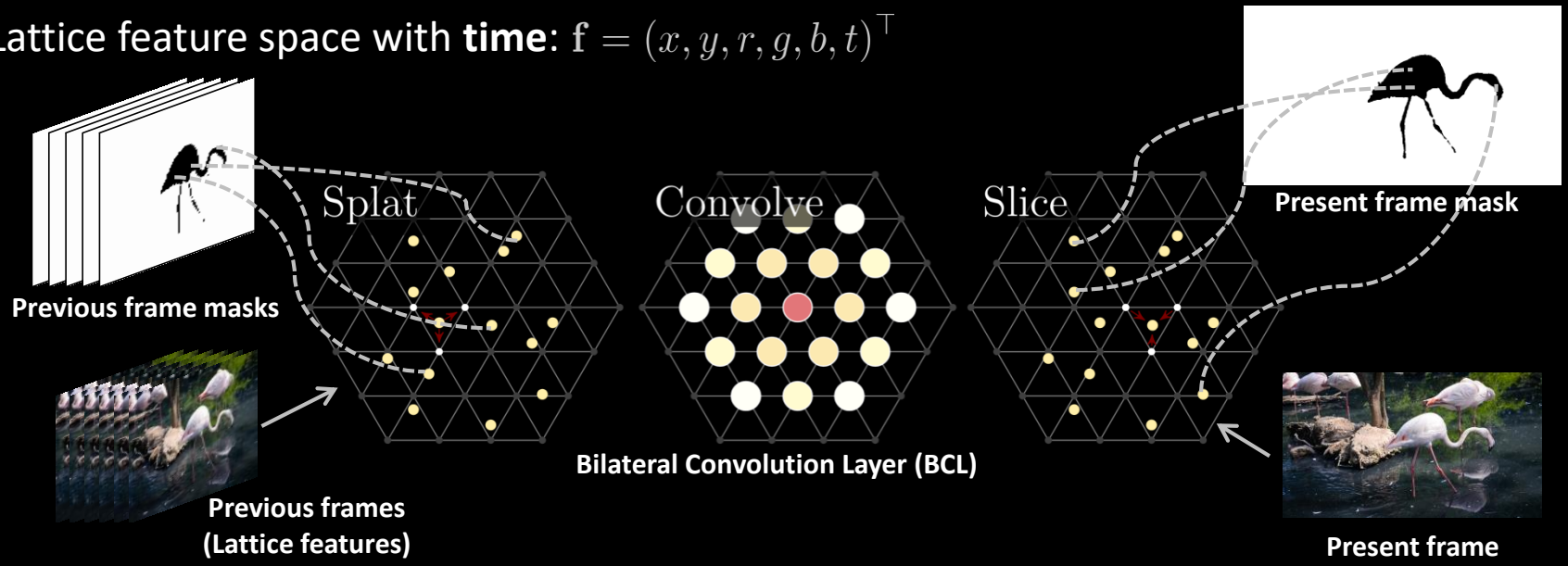
Video propagation

Another example: Color propagation



Video propagation with BCL

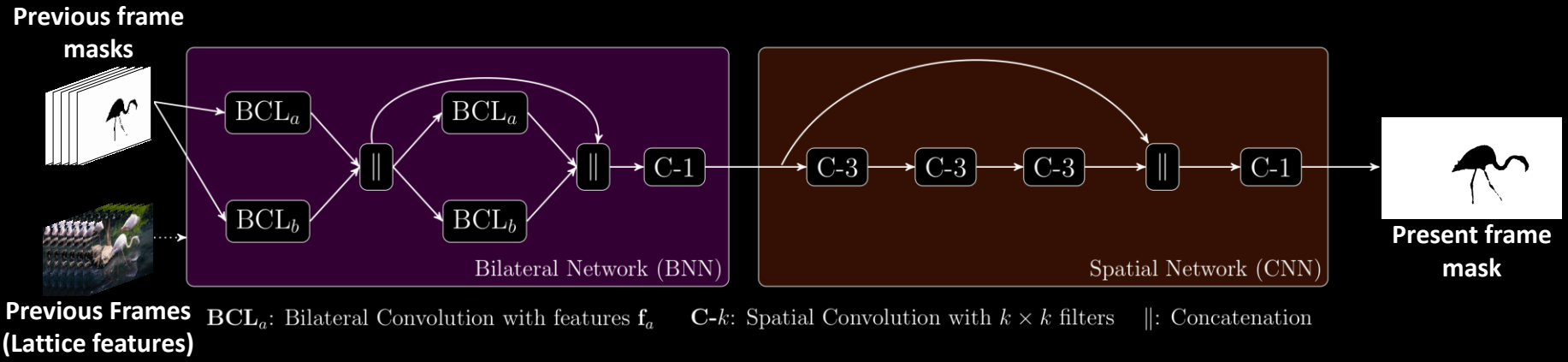
- Splat *previous f*rame results, convolve and slice at the *present* frame.
- Lattice feature space with **time**: $\mathrm{f} = (x, y, r, g, b, t)^\top$



Video Propagation Network (VPN) [CVPR'17]

Two components in VPN

- Bilateral Network (BNN): For propagation from previous frames to present one
- Spatial Network (CNN): For refining features and prediction for the present frame



[CVPR'17] Jampani, V., Gadde, R. and Gehler, P.V., Video Propagation Networks. In *CVPR,* 2017.

Ground Truth

BVS-Result

VPN-Result(Ours)


Ground Truth
Levin et al.
VPN(Ours)

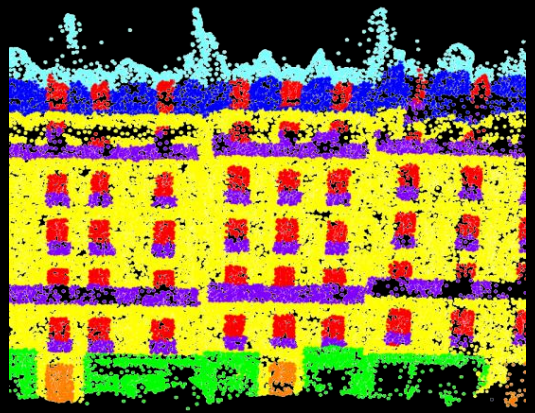
3D point cloud segmentation

Goal: End-to-end trainable networks for analyzing point cloud data

Challenges:

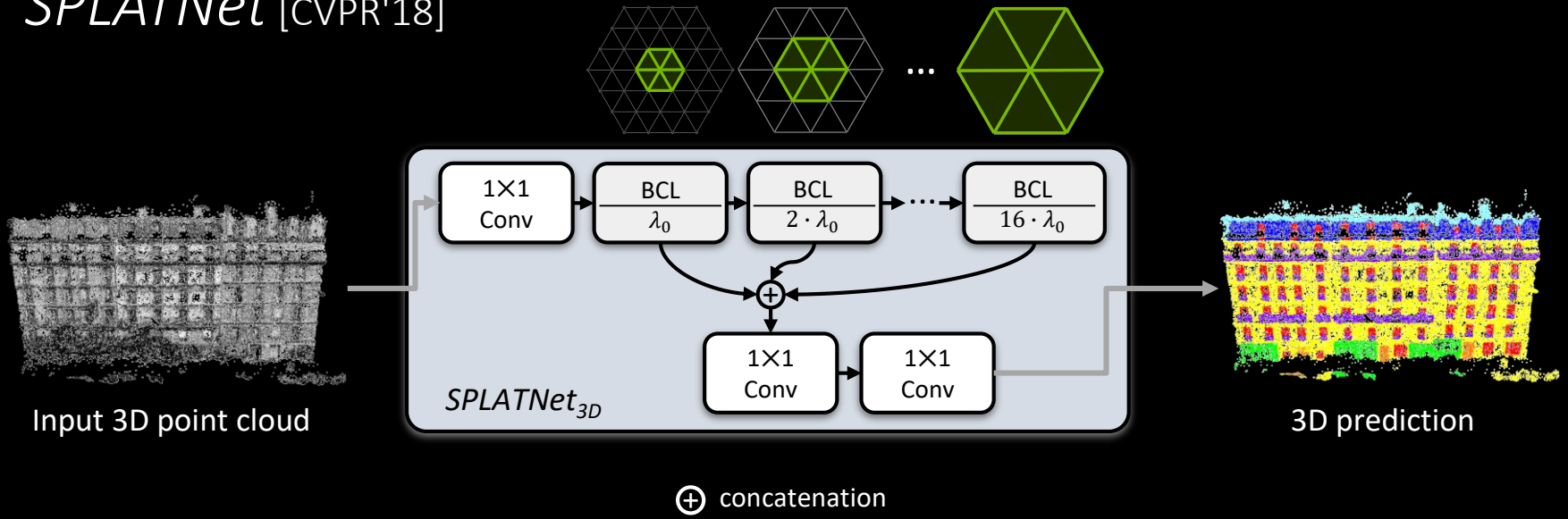
- Point clouds are *unordered* and *sparse*
- No readily defined *connections* between points

Input point cloud



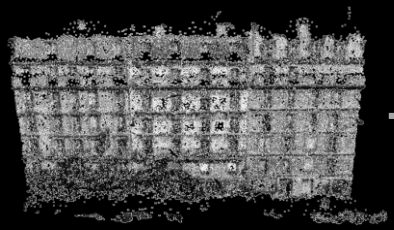
Semantic Segmentation

SPLATNet [CVPR'18]

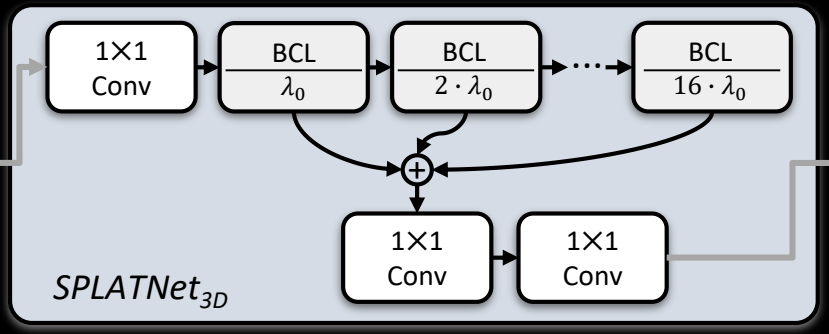


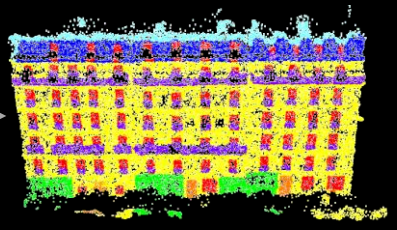
[CVPR'18] Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.H. and Kautz, J., SPLATNet: Sparse Lattice Networks for Point Cloud Processing. In *CVPR,* 2018.

Joint 2D-3D processing



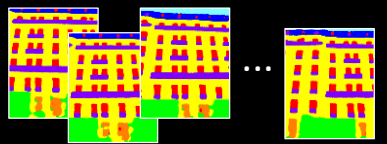
Input 3D point cloud





3D prediction

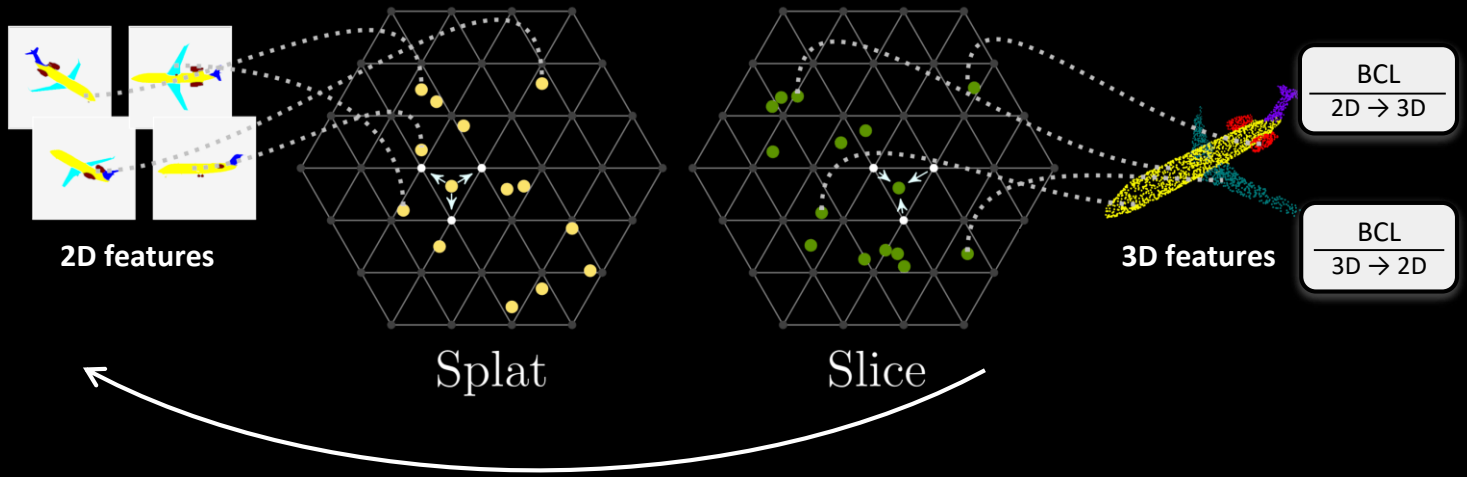
Input images



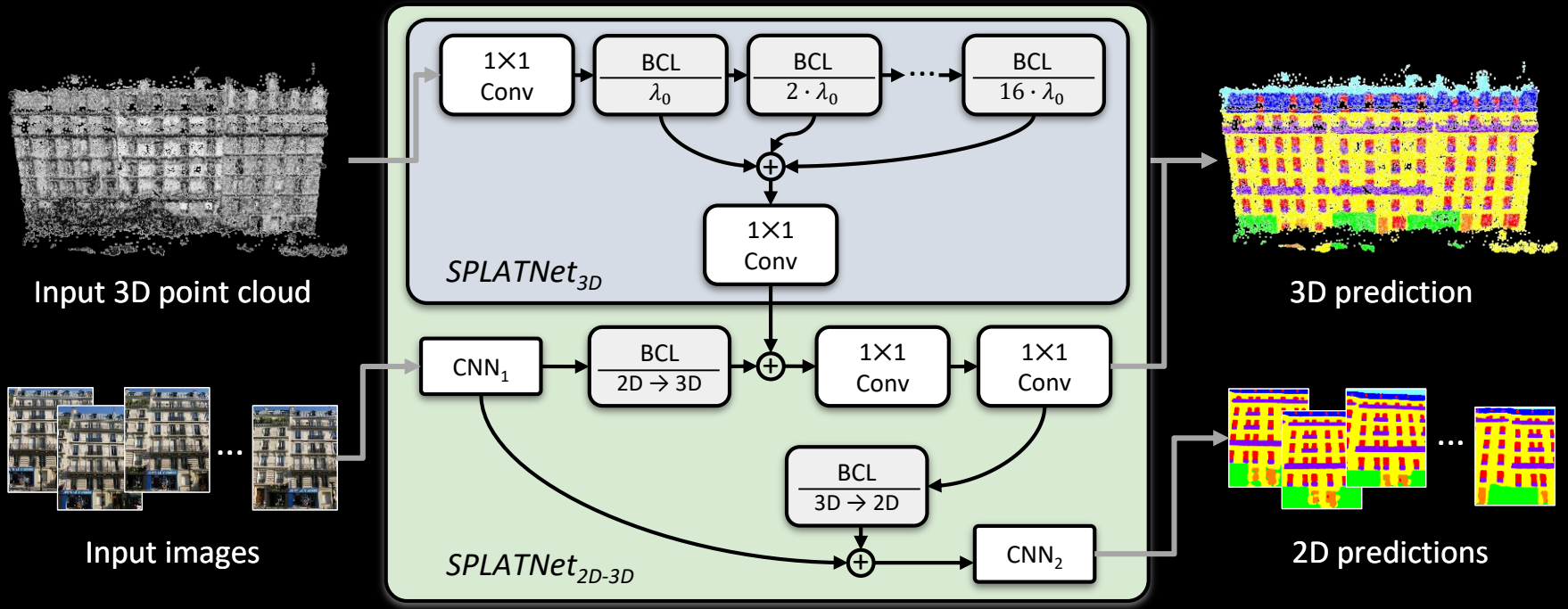
2D predictions

Smooth projection between 2D and 3D

- Each 2D pixel p: $\mathbf{v}(p)$, (x, y, z)
  → Input feature: $\mathbf{v}(p)$
  → Lattice feature: (x, y, z)



SPLATNet$_{2D\text{-}3D}$ [CVPR'18]



[CVPR'18] Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.H. and Kautz, J., SPLATNet: Sparse Lattice Networks for Point Cloud Processing. In *CVPR,* 2018.

Facade segmentation (Ruemonge2014 [1])

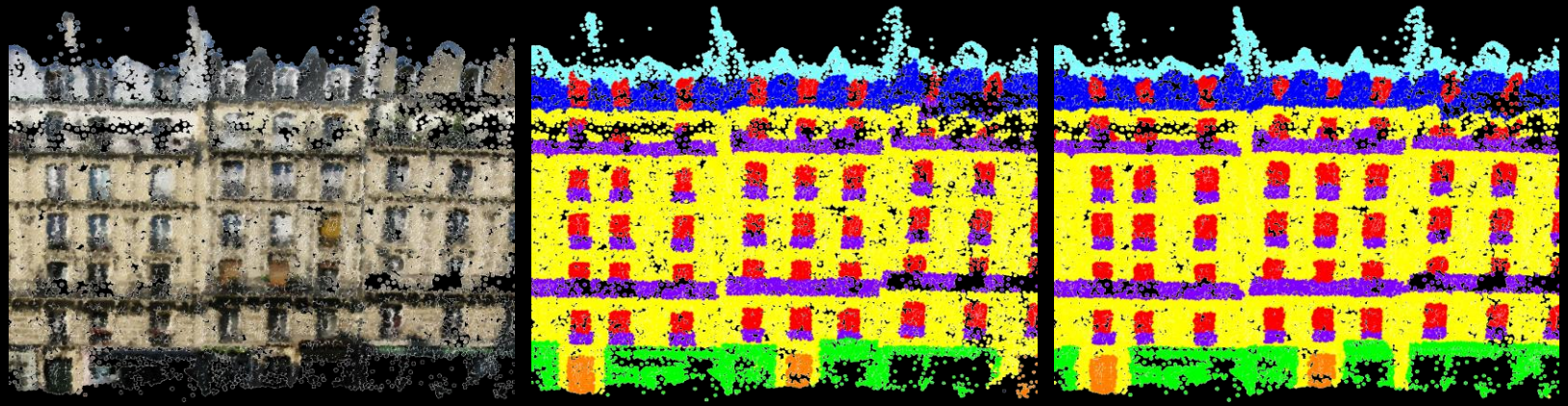
with only 3D data

	IoU	runtime (min)
OctNet [2]	59.2	-
Autocontext$_{3D}$ [3]	54.4	16
SPLATNet$_{3D}$	**65.4**	**0.06**

with both 2D & 3D data

	IoU	runtime (min)
Autocontext$_{2D-3D}$ [3]	62.9	87
SPLATNet$_{2D-3D}$	**69.8**	**1.2**

[2] Riegler *et al.* CVPR '17 [3] Gadde *et al.* PAMI '17



Input point cloud | Ground-truth | Ours (*SPLATNET$_{2D-3D}$*)

[1] Riemenschneider et al. ECCV '14

2D predictions

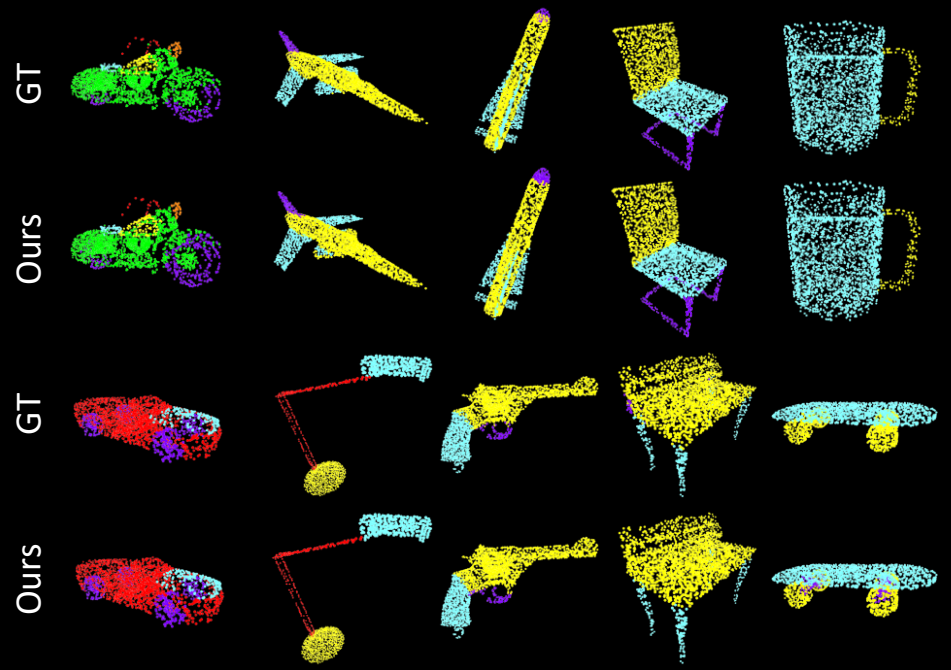
	IoU	runtime (min)
Autocontext$_{2D}$ [1]	60.5	117
Autocontext$_{2D\text{-}3D}$ [1]	62.7	146
2D CNN [2] only	69.3	0.84
SPLATNet$_{2D\text{-}3D}$	70.6	4.34

[1] Gadde *et al.* PAMI '17
[2] Chen *et al.* ICLR '15

Input image | Ground-truth | Our prediction

3D object part labeling (ShapeNet [1])

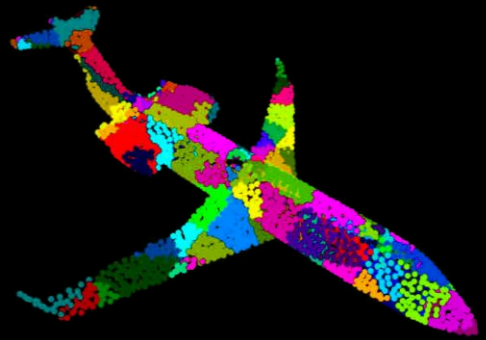
	class avg. IoU	instance avg. IoU
Yi *et al.* [1]	79.0	81.4
3DCNN [2]	74.9	79.4
Kd-network [3]	77.4	82.3
PointNet [2]	80.4	83.7
PointNet++ [4]	81.9	85.1
SyncSpecCNN [5]	82.0	84.7
SPLATNet$_{3D}$	**82.0**	**84.6**
SPLATNet$_{2D-3D}$	**83.7**	**85.4**



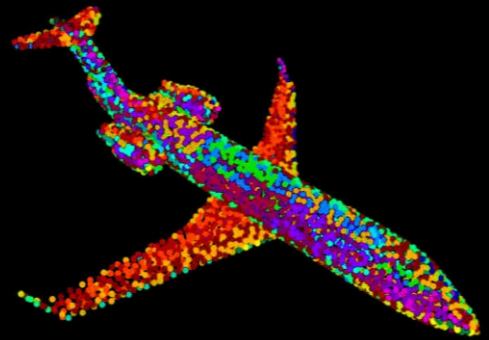
[1] Yi *et al.* SIGGGRAPH Asia '16
[2] Qi *et al.* CVPR '17
[3] Klokov and Lempitsky ICCV '17
[4] Qi *et al.* NIPS '17
[5] Yi *et al.* CVPR '17

Filtering in other lattice spaces

- Adding additional (x, y, z, n_x, n_y, n_z) filtering → +0.2 IoU



(x, y, z)



(n_x, n_y, n_z)

Remarks

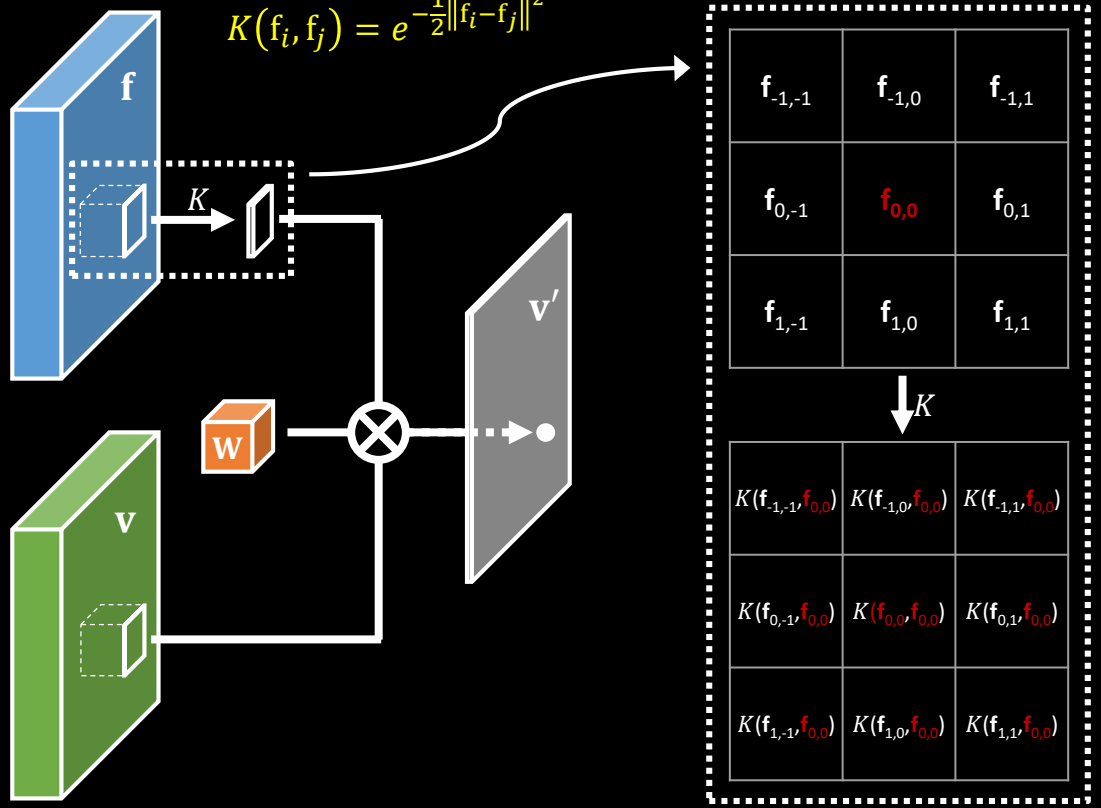
$$\mathbf{v}'_i = \sum_{j \in \mathcal{N}} W(\mathbf{p}_i - \mathbf{p}_j)\mathbf{v}_j \longrightarrow \mathbf{v}'_i = \sum_{j \in \mathcal{N}} W(\mathbf{f}_i - \mathbf{f}_j)\mathbf{v}_j$$

- General sparse high-dimensional neural networks
- Advantages
  - Efficient computation with permutohedral lattice and hash tables
  - Flexible specification of lattice features and receptive fields
  - Input and output can be at different points
  - Wide range of applications in 2D, video and 3D vision
- Disadvantages
  - Slower than standard convolution when filtering 2D data
  - Manual specification of lattice features

Pixel-Adaptive Convolution (PAC)

For our experiments, we use

$$K(\mathbf{f}_i, \mathbf{f}_j) = e^{-\frac{1}{2}\|\mathbf{f}_i - \mathbf{f}_j\|^2}$$



- $\mathbf{v}$ input features
- $\mathbf{v}'$ output features
- p (x, y) coordinates
- $\mathbf{W}$ filter weights
- $\mathbf{f}$ adapting features
- K adapting kernel

[CVPR'19] Su, H., Jampani, V., Sun, D., Gallo, O., Learned-Miller, E. and Kautz, J., Pixel-Adaptive Convolutional Neural Networks. In *CVPR*, 2019.

Pixel-Adaptive Convolution (PAC)

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} K(\mathbf{f}_i, \mathbf{f}_j)\mathbf{W}[p_i - p_j]\mathbf{v}_j + \mathbf{b}$$

Pixel-adaptive convolution

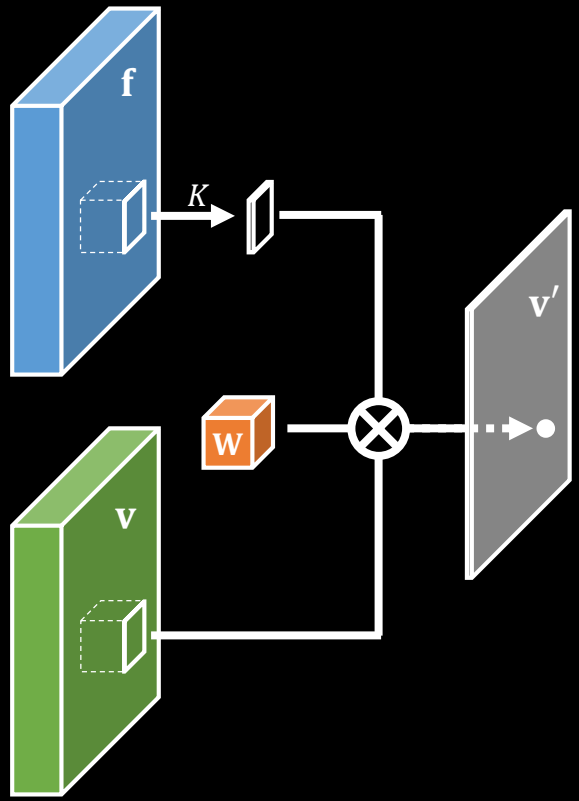
$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[p_i - p_j]\mathbf{v}_j + \mathbf{b}$$

Standard convolution

factorization

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{f}_i - \mathbf{f}_j]\mathbf{v}_j + \mathbf{b}$$

High-dimensional convolution



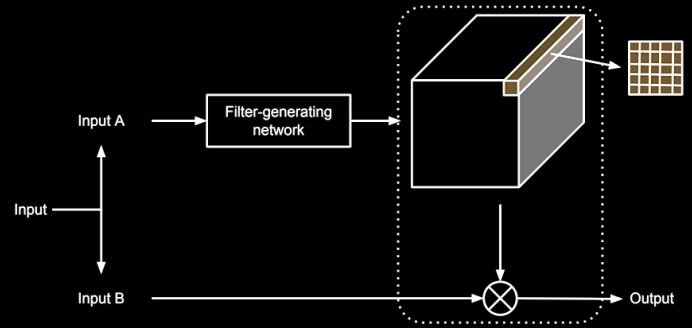
PAC is content-adaptive

Effective filters are different across images and pixel locations

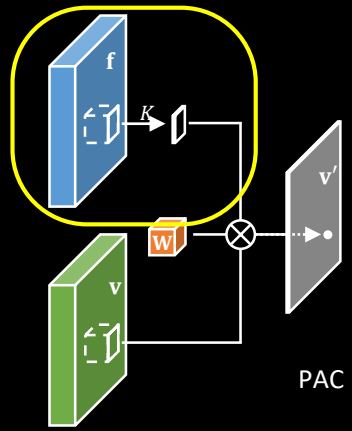
Content-adaptive filters

- Kernel-prediction networks [1,2,3]

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}_i[p_i - p_j]\mathbf{v}_j + \mathbf{b} \quad \leftarrow \quad \mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[p_i - p_j]\mathbf{v}_j + \mathbf{b} \quad \rightarrow \quad \mathbf{v}_i' = \sum_{j \in \Omega(i)} K(\mathbf{f}_i, \mathbf{f}_j)\mathbf{W}[p_i - p_j]\mathbf{v}_j + \mathbf{b}$$



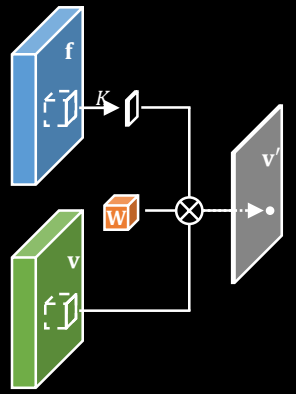
Kernel-prediction networks (fig. from [1])



PAC

[1] Brabandere et al. Dynamic filter networks. NIPS '16
[2] Bako et al. Kernel-Predicting Convolutional Networks for Denoising Monte Carlo Renderings. TOG '17
[3] Wu et al. Dynamic Sampling Convolutional Neural Networks. ECCV '18

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} K(\mathbf{f}_i, \mathbf{f}_j)\mathbf{W}[p_i - p_j]\mathbf{v}_j + \mathbf{b}$$



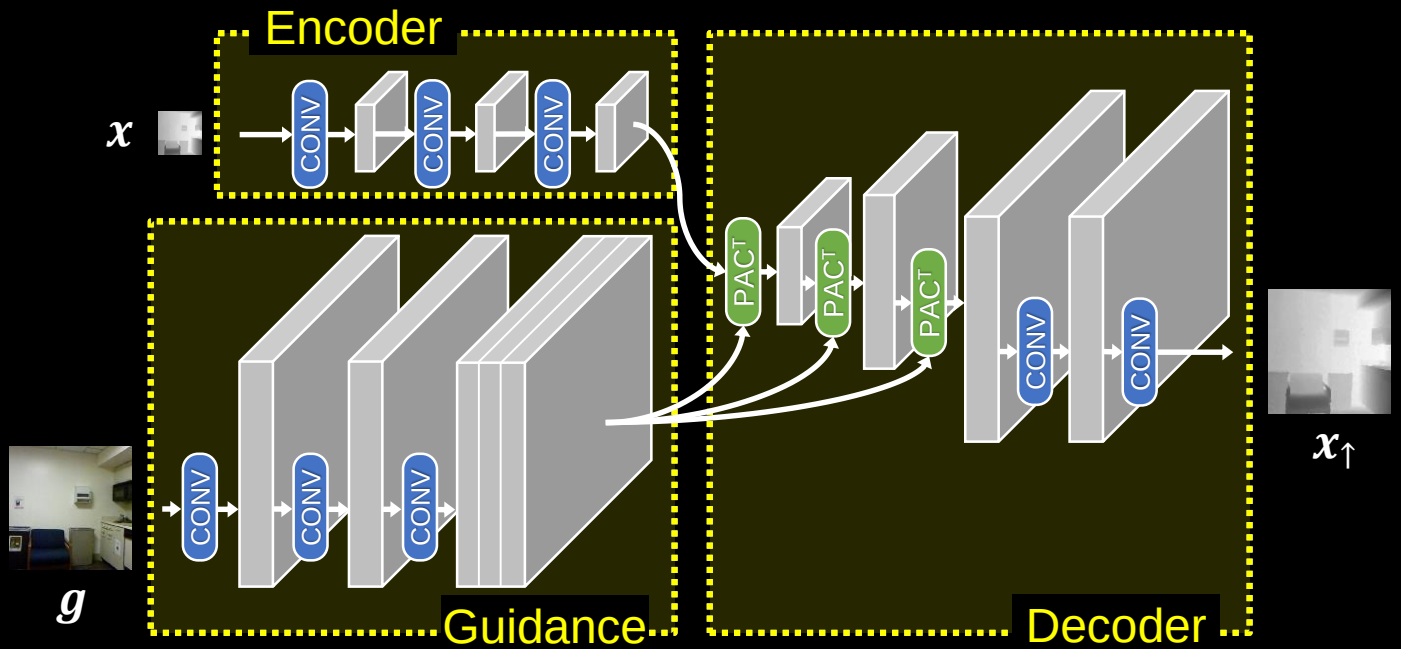
- Spatial convolution $K(\mathbf{f}_i, \mathbf{f}_j) = 1$
- Bilateral filtering $\mathbf{f} = (r, g, b), K(\mathbf{f}_i, \mathbf{f}_j) = \exp(-\frac{1}{2\alpha_1}\|\mathbf{f}_i - \mathbf{f}_j\|^2)$

  $\mathbf{W}[p_i - p_j] = \exp(-\frac{1}{2\alpha_2}\|p_i - p_j\|^2)$
- Average pooling $K(\mathbf{f}_i, \mathbf{f}_j) = 1, \mathbf{W}[p_i - p_j] = \frac{1}{F^2}$
- Detail-preserving pooling [1] $K(\mathbf{f}_i, \mathbf{f}_j) = \alpha + (\|\mathbf{f}_i - \mathbf{f}_j\|^2 + \epsilon^2)^\lambda$

PAC generalizes many existing filtering techniques

[1] F. Saeedan, et al. Detail Preserving Pooling in Deep Networks. CVPR '18

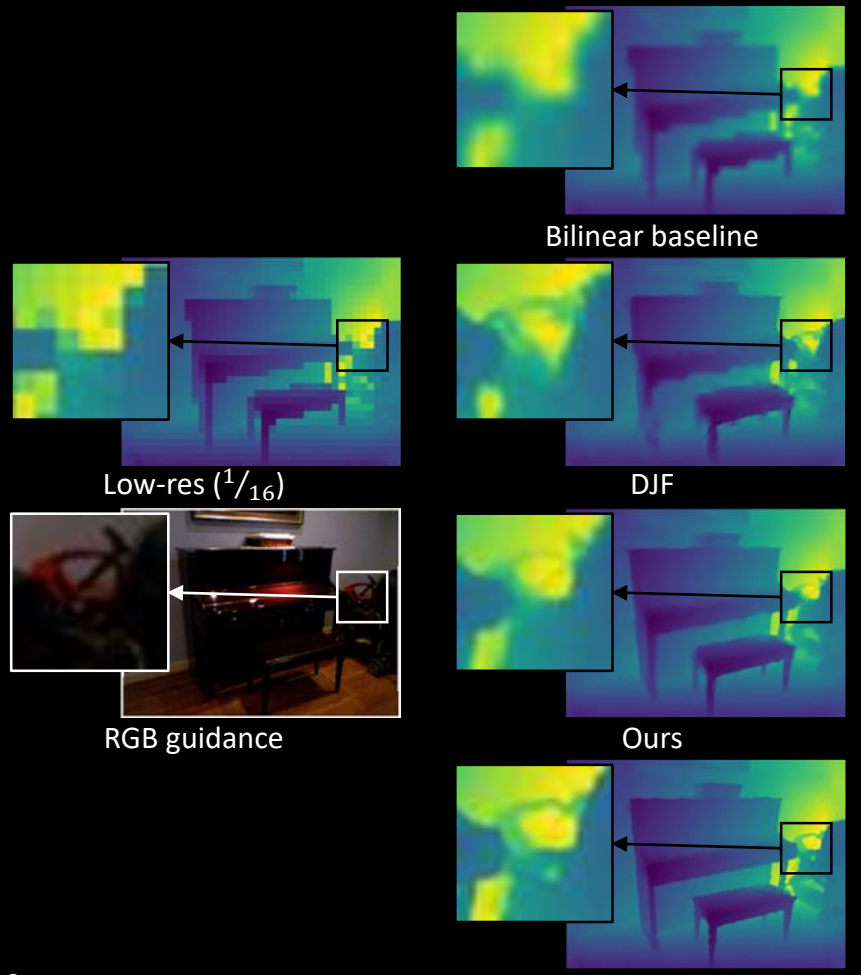
Joint upsampling network with PAC



Joint depth upsampling

Method	4×	8×	16×
Bicubic	8.16	14.22	22.32
MRF	7.84	13.98	22.20
GF [1]	7.32	13.62	22.03
JBU [2]	4.07	8.29	13.35
DMSG [3]	3.78	6.37	11.16
FBS [4]	4.29	8.94	14.59
DJF [5]	3.54	6.20	10.21
DJF+ [6]	3.38	5.86	10.11
Ours	**2.39**	**4.59**	**8.09**

RMSE on NYU-V2 depth maps

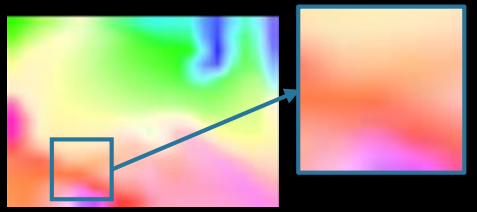


[1] He *et al.* PAMI '13
[2] Kopf *et al.* ToG '07
[3] Hui *et al.* ECCV '16
[4] Barron and Poole. ECCV '16
[5] Li *et al.* ECCV '16
[6] Li *et al.* PAMI '18

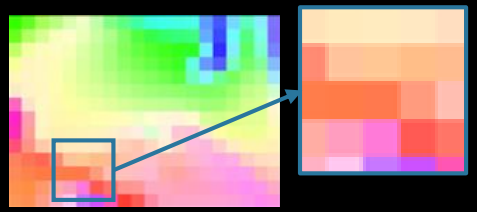
Joint optical flow upsampling

Method	4×	8×	16×
Bicubic	0.465	0.901	1.628
DJF	0.176	0.438	1.043
Ours	**0.105**	**0.256**	**0.592**

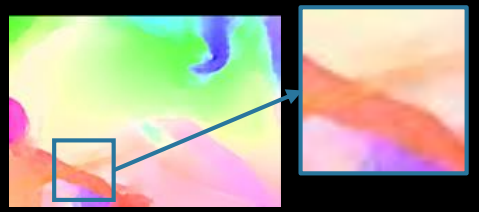
End-Point-Error (EPE) on Sintel



Bilinear baseline

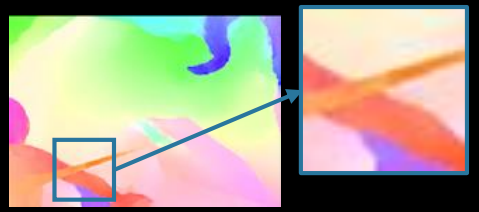


Low-res ($^1/_{16}$)

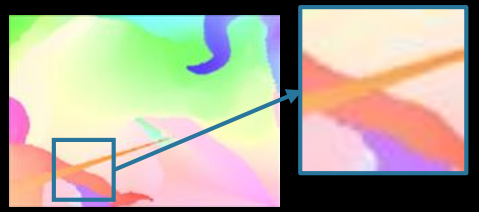


DJF

RGB guidance

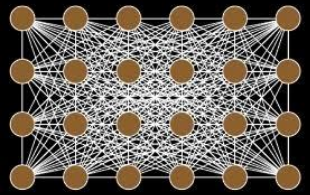


Ours



GT

PAC for CRF inference



- Fully Connected CRF[1] : connection between every pixel pairs

$$p(\mathbf{l}|I) \propto \exp\left\{-\sum_i \psi_u(l_i|I) - \sum_{i<j} \psi_p(l_i, l_j|I)\right\}$$

unary potential

pairwise potential

- Applications: segmentation[1], optical flow[2], intrinsic images[3], etc.
- Mean-field inference with sparse high-dim filtering[1]

$$Q_i^{(t+1)}(l) \leftarrow \frac{1}{Z_i} \exp\left\{-\psi_u(l) - \sum_{l' \in \mathcal{L}} \mu(l, l') \sum_{j \neq i} K(\mathbf{f}_i, \mathbf{f}_j) Q_j^{(t)}(l')\right\}$$

sparse high-dim filtering

[1] Krähenbühl, P., & Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. NIPS '11
[2] Sun, D., Wulff, J., Sudderth, E. B., Pfister, H., & Black, M. J. A fully-connected layered model of foreground and background flow. CVPR '13
[3] Bell, S., Bala, K., & Snavely, N. Intrinsic images in the wild. TOG '14

PAC for CRF inference

- Fully Connected CRF

$$\psi_p(l_i, l_j|I) = \mu(l_i, l_j) K(\mathbf{f}_i, \mathbf{f}_j) \qquad K(\mathbf{f}_i, \mathbf{f}_j) = w_1 \exp\left\{-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\theta_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\theta_\beta^2}\right\} + w_2 \exp\left\{-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\theta_\gamma^2}\right\}$$

label compatibility

- Generalized pairwise potential

$$\psi_p(l_i, l_j|I) = W(\mathbf{p}_j - \mathbf{p}_i, l_i, l_j) K(\mathbf{f}_i, \mathbf{f}_j) \qquad K(\mathbf{f}_i, \mathbf{f}_j) = \exp\left\{-\frac{1}{2}\|\mathbf{f}_i - \mathbf{f}_j\|^2\right\}$$

- Advantages:
  - Can learn more general compatibility transformation
  - Use learned features in kernel function

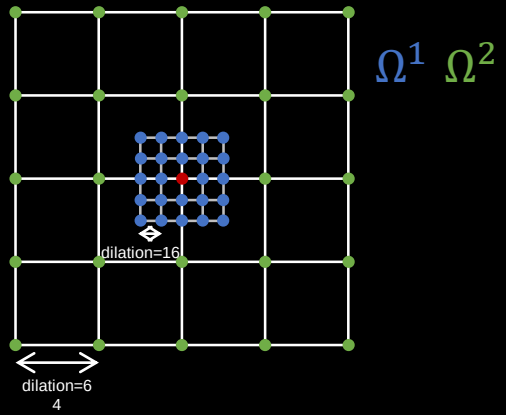
PAC for CRF inference

- The new update rule

$$Q_i^{(t+1)}(l) \leftarrow \frac{1}{Z_i} \exp\left\{-\psi_u(l) - \sum_{l' \in \mathcal{L}} \sum_{j \in \Omega(i)} W(\mathbf{p}_j - \mathbf{p}_i, l, l') K(\mathbf{f}_i, \mathbf{f}_j) Q_j^{(t)}(l')\right\}$$

PAC

- Increasing the reach of local neighborhoods
  - PAC considers only pairs within certain distance
  - Use dilated convolution
  - Combine multiple dilation factors

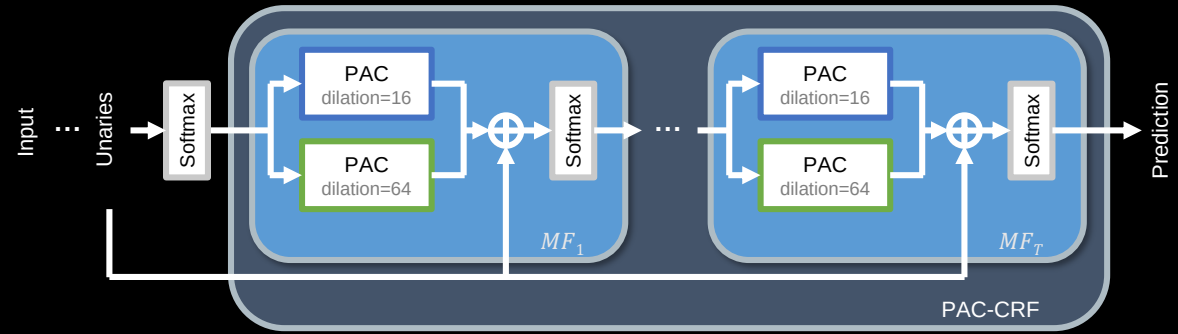


PAC for CRF inference

- Easy integration with DNNs

$$Q_i^{(t+1)}(l) \leftarrow \frac{1}{Z_i} \exp\left\{-\psi_u(l) - \sum_{l' \in \mathcal{L}} \sum_{j \in \Omega(i)} W(\mathbf{p}_j - \mathbf{p}_i, l, l') K(\mathbf{f}_i, \mathbf{f}_j) Q_j^{(t)}(l')\right\}$$

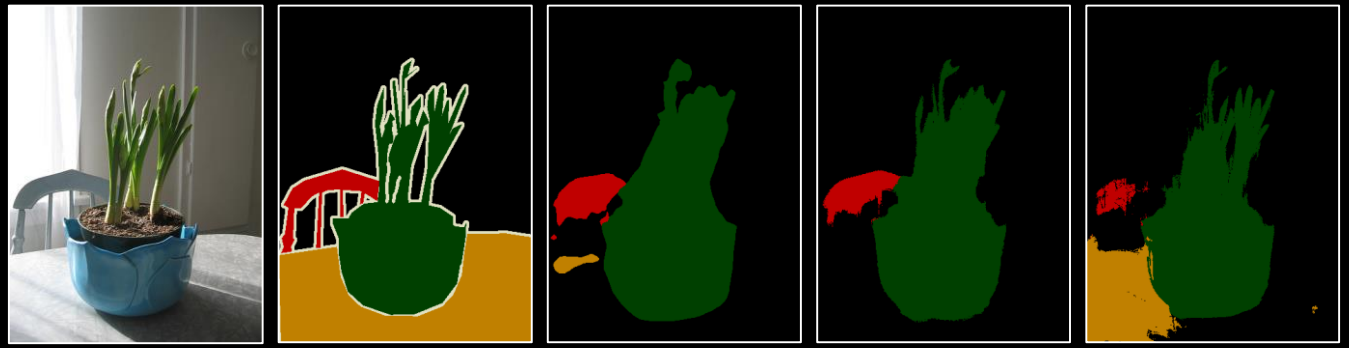
PAC



PAC for CRF inference

Method	mIoU	Runtime
FCN only	67.20	39ms
Full-CRF [1]	+2.45	+629ms
Conv-CRF [2]	+1.57	+38ms
PAC-CRF	**+2.62**	+78ms

Evaluated on Pascal VOC test images

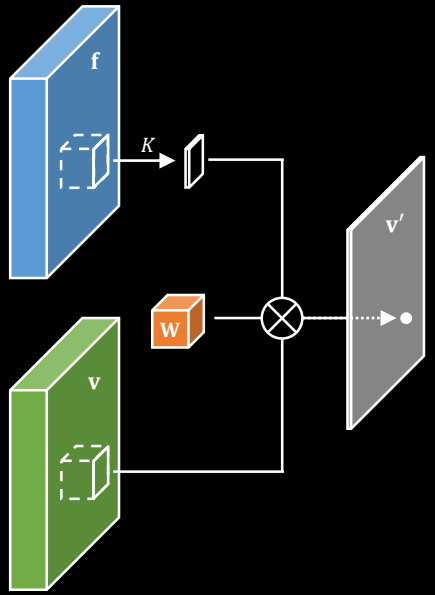


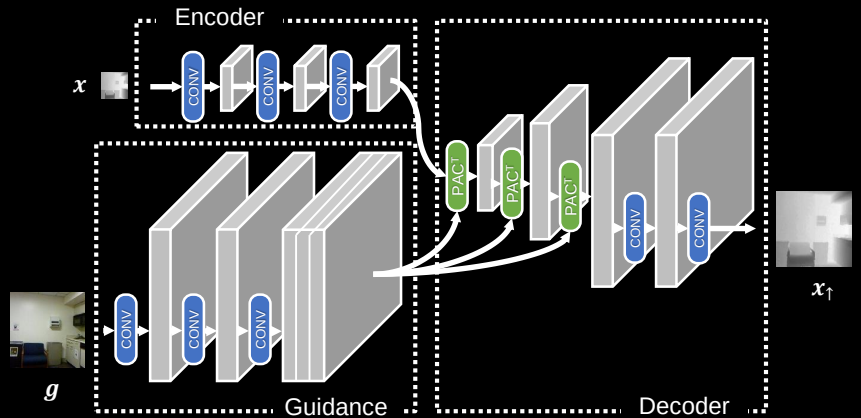
Input GT Unary Full-CRF [1] Ours

[1] Krähenbühl, P., & Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. NIPS '11
[2] Teichmann, M.T. and Cipolla, R., Convolutional CRFs for Semantic Segmentation. arXiv '18

Layer “hot-swapping” with PAC

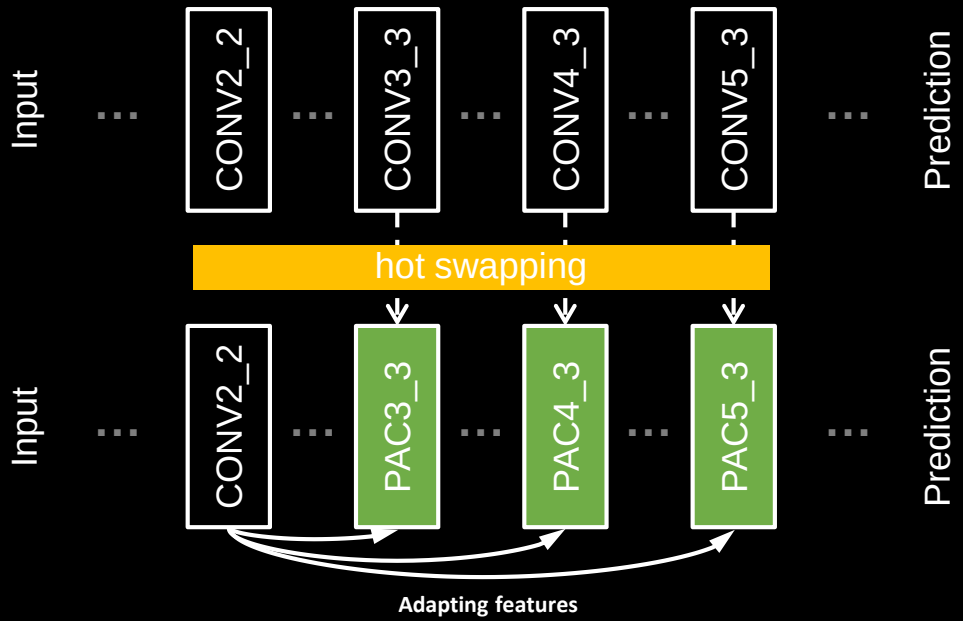
- Two ways to get adapting features: (1) dedicated branch





Layer “hot-swapping” with PAC

- Two ways to get adapting features: (1) dedicated branch (2) re-using features

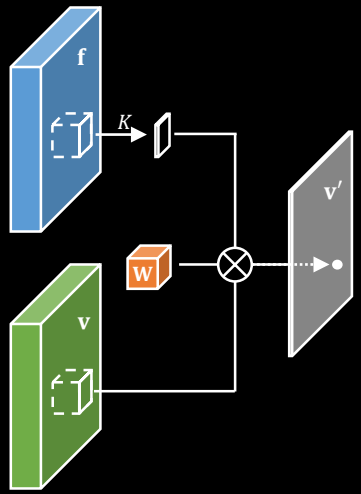


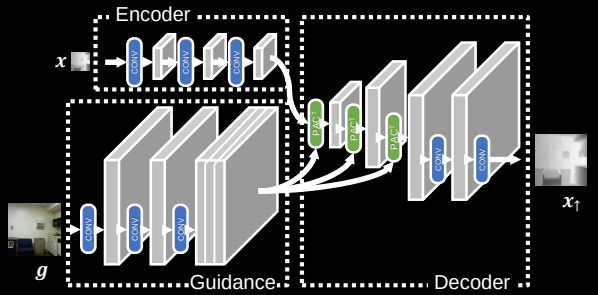
Method	mIoU	Runtime
FCN	67.20	39ms
PAC-FCN	69.18	41ms
PAC-FCN + PAC-CRF	**71.34**	118ms

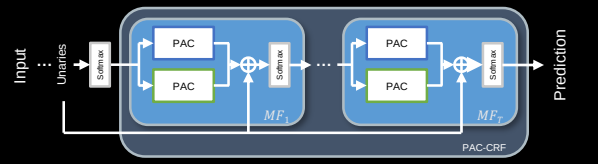
Evaluated on Pascal VOC test images

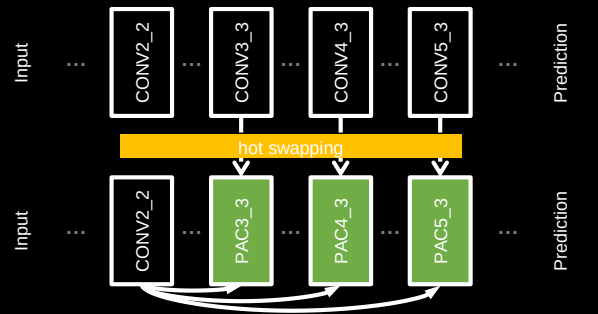
PAC: Remarks

- Pixel-Adaptive Convolution:
  - Content-adaptive
  - Generalizes several existing filtering techniques
- Three use case examples:
  - Joint upsampling networks
  - Efficient CRF inference
  - Network layer hot-swapping









Summary and Future outlook

Sparse High Dimensional Convolutions

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} \mathbf{W}[\mathbf{f}_i - \mathbf{f}_j]\mathbf{v}_j$$

- Can filter sparse and high-dimensional data.
- Helps in long-range information propagation.
- Manual specification of feature spaces.

- Future outlook: Learn lattice features via backpropagation.

Pixel Adaptive Convolutions

$$\mathbf{v}_i' = \sum_{j \in \Omega(i)} K(\mathbf{f}_i, \mathbf{f}_j)\mathbf{W}[\mathbf{p}_i - \mathbf{p}_j]\mathbf{v}_j$$

- Learns guidance features.
- Fast and can hot-swap standard convolutions.
- Works only on dense 2D data.

- Future outlook: Other applications and filtering sparse and high-dimensional data.

Thank you

Comments and suggestions are most welcome!

Hang Su
https://suhangpro.github.io/

Varun Jampani
https://varunjampani.github.io/