MAX-PLANCK-GESELLSCHAFT

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Learning Inference Models for Computer Vision

Varun Jampani

PhD Thesis Presentation, May 2nd 2017

MPI for Intelligent Systems and University of Tübingen, Germany

Advisor: Prof. Peter Gehler

Thesis Committee:
Prof. Peter Gehler, Prof. Hendrik Lensch, Prof. Felix Wichmann and Prof. Matthias Bethge

Three Components of a Vision System

1. Model

Model Definition of Input-Output relationship

2. Learning

Model Learning of parameters/structure

3. Inference

Model Inference for a given observation

Thesis Objective

This thesis proposes techniques for better inference in a range of computer vision models.

3. Inference

Model Inference for a given observation

Two Classes of Vision Models

Generative Models

- Models image formation process (likelihood and/or prior): $P(\mathbf{x}|\mathbf{y}, \theta)P(\mathbf{y})$
- Prior knowledge and quantified uncertainty
- Examples: Graphics systems, probabilistic graphical models etc.

Discriminative Models

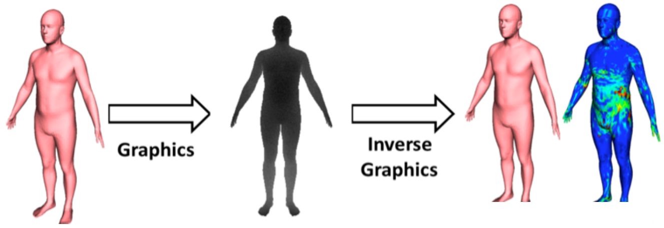
- Directly models posterior distribution: $P(\mathbf{y}|\mathbf{x}, \theta)$
- Fast and learned from data
- Examples: Random forests, CNNs etc.

Part 1: Inference in Generative Computer Vision Models

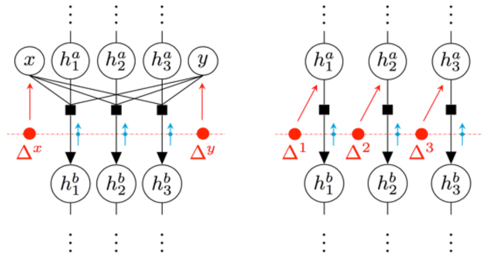
Better Inference in Generative Vision Models

Inference in *inverse graphics* and *probabilistic graphical models.*

Informed Sampler [CVIU'15]
- Learning sampling



Consensus Message Passing [AISTATS'15]
- Learning message passing



[CVIU'15]. **Jampani, V.**, Nowozin, S., Loper, M., & Gehler, P. V. The informed sampler: A discriminative approach to Bayesian inference in generative computer vision models. *CVIU*, 2015.
[AISTATS'15]. **Jampani, V.**, Eslami, S. M., Tarlow, D., Kohli, P., & Winn, J. Consensus Message Passing for Layered Graphical Models. In *AISTATS,* 2015.

Informed Sampler

for inverting graphics engines

Vision as Inverse Graphics

Vision can be tackled as inverting graphic engines

Modern graphic engines produce stunning level of realism

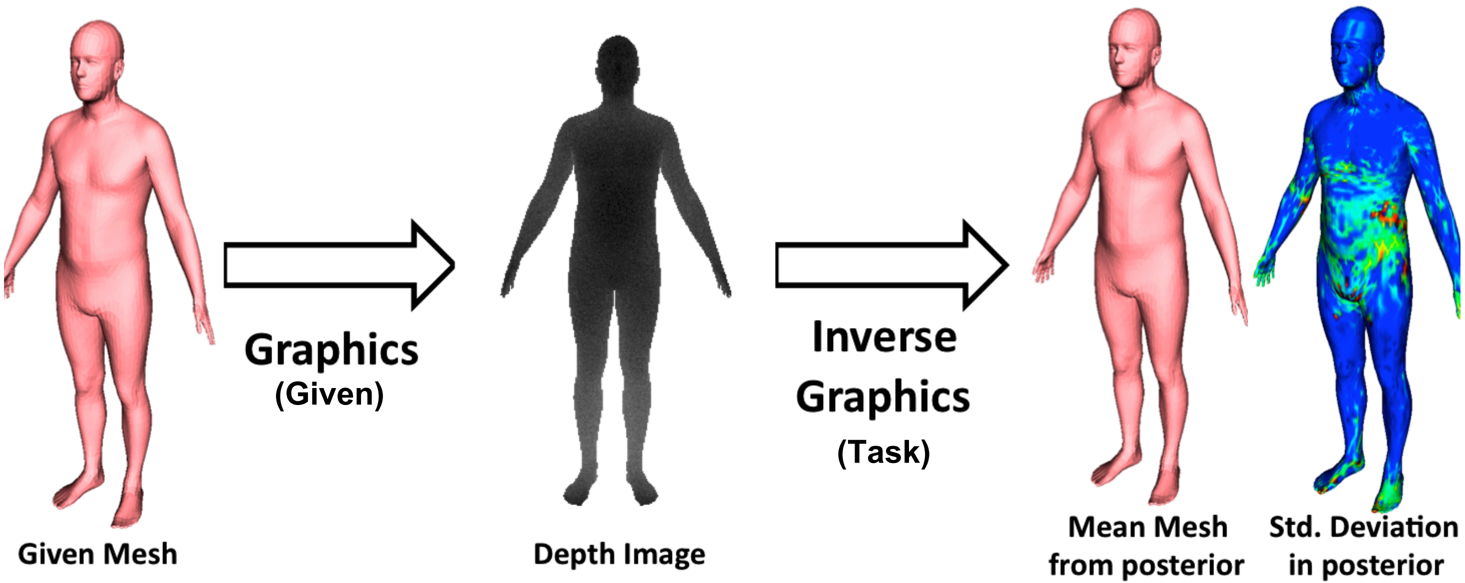
Scene Geometry
Lighting
Albedo
Camera parameters etc.

Graphics **Vision**

Image

Sample Renderings from Modern Graphics Engines

An Example Inverse Graphics Problem



Goal: Inverse graphics with *making use of* graphics renderer

Inference in Inverse Graphics Models

Markov Chain Monte Carlo (MCMC) sampling to estimate the posterior distribution

1. Sample parameters ⟶ 2. Render ⟶ 3. Compare and accept/reject parameters ⟶ Repeat 1-3

General MCMC Algorithm [1]:

At each step in Markov chain

1. Propose next sample using a *proposal distribution*

$$\bar{\mathbf{y}} \sim T(\cdot|\mathbf{y}_t)$$

2. Accept or reject it based on Metropolis Hastings (MH) acceptance rule

$$\mathbf{y}_{t+1} = \begin{cases} \bar{\mathbf{y}}, & \mathrm{rand}(0,1) < \min\left(1, \frac{\pi(\bar{\mathbf{y}})T(\bar{\mathbf{y}}\to\mathbf{y}_t)}{\pi(\mathbf{y}_t)T(\mathbf{y}_t\to\bar{\mathbf{y}})}\right), \\ \mathbf{y}_t, & \text{otherwise.} \end{cases}$$

Target posterior distribution

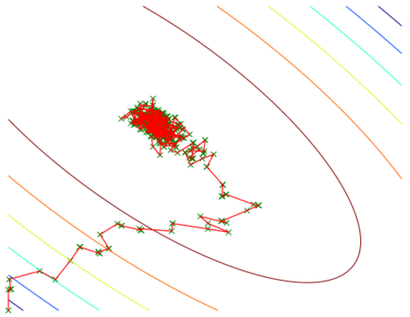


Illustration of MCMC Sampling for a 2D Gaussian distribution

MCMC sampling is *generic* but is often *too slow* and *fail to converge*.

1. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., Equation of state calculations by fast computing machines. *The journal of chemical physics,* 1953.

Informed Sampler [CVIU'15]

Mixture Sampling:

$$T_\alpha(\cdot|\mathbf{x},\mathbf{y}_t) = \alpha\, T_L(\cdot|\mathbf{y}_t) + (1-\alpha)\, T_G(\cdot|\mathbf{x}).$$

Learn global proposal distribution T_G using discriminative approaches.

We experimented with random forest and clustering techniques.

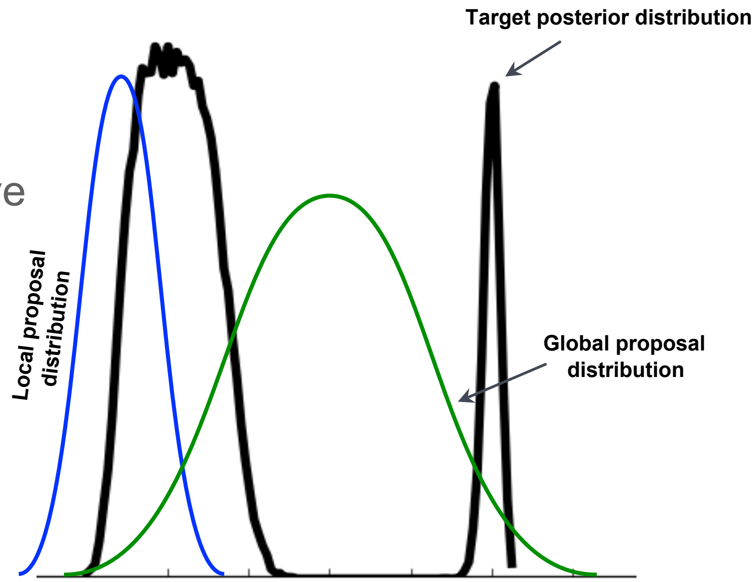


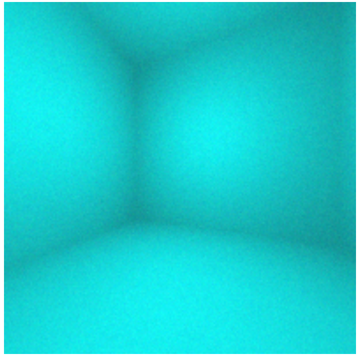
Illustration of sampling with a mixture of local (blue) and global (green) proposal distributions

[CVIU'15]. **Jampani, V.**, Nowozin, S., Loper, M., & Gehler, P. V. The informed sampler: A discriminative approach to Bayesian inference in generative computer vision models. *CVIU*, 2015.

Experiments

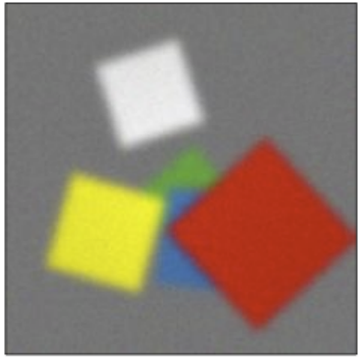
Graphics

Room Renderer



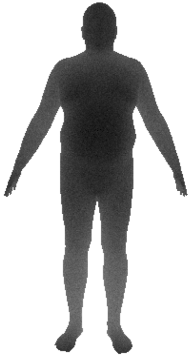
Camera Localization

Tile Renderer



Occlusion Reasoning

Body Depth Renderer

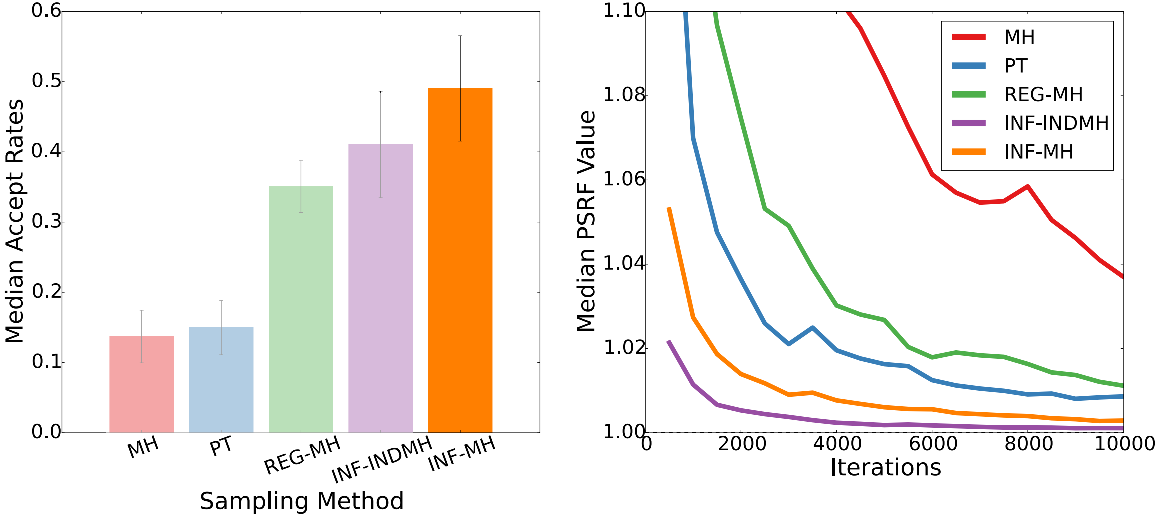


3D Shape Estimation

Inverse Graphics

Observations

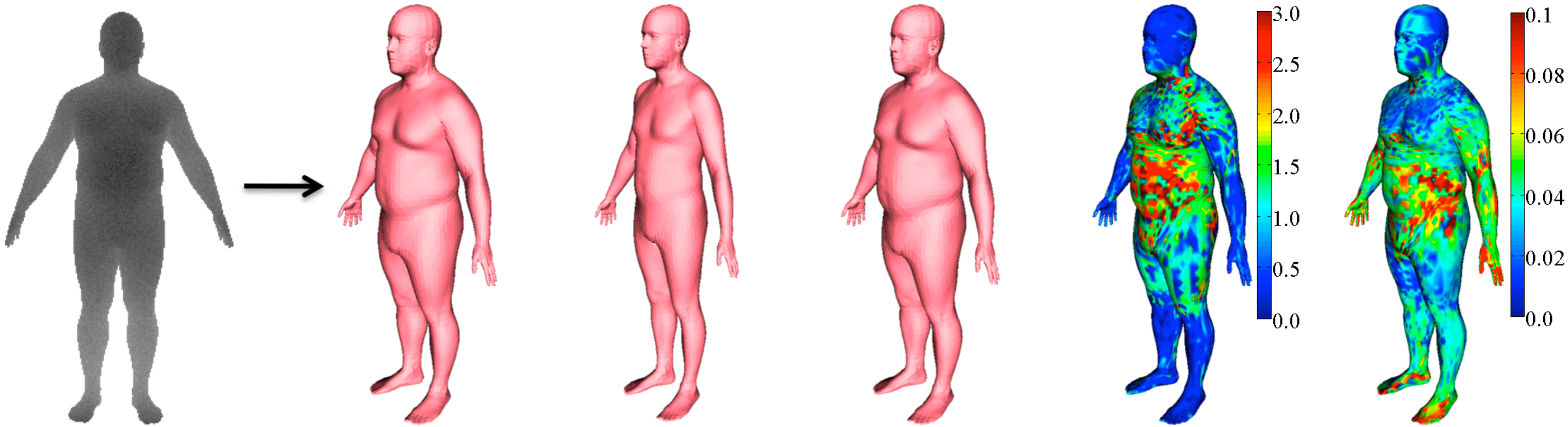
Higher acceptance rates and **faster convergence** for informed sampler in comparison to wide range of baseline samplers.



Results of experiments on Camera Localization.

MH: Metropolis Hastings**, PT:** Parallel Tempering**, REG-MH:** Regenerative-MH**, INF-INDMH:** *Informed* Global Sampling**, INF-MH:** *Informed* Sampler

Quantified Uncertainties with Inverse Graphics



Given Image | Ground Truth | Mean Body Shape | Estimated Shape (Informed Sampler) | Error | Std. Deviation in Posterior

An example result of 3D shape estimation using *Informed Sampler*

Remarks: Informed Sampler [CVIU’15]

Most existing sampling methods *fail* when dealing with complex multi-modal distributions

Informed Sampler consistently improved over existing samplers

- Global proposals helps in **quick mixing**
- Local proposals helps in **higher acceptance rate**

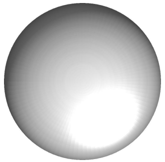
Inverse graphics can

- yield accurate solutions even with incomplete evidence
- give quantified uncertainties

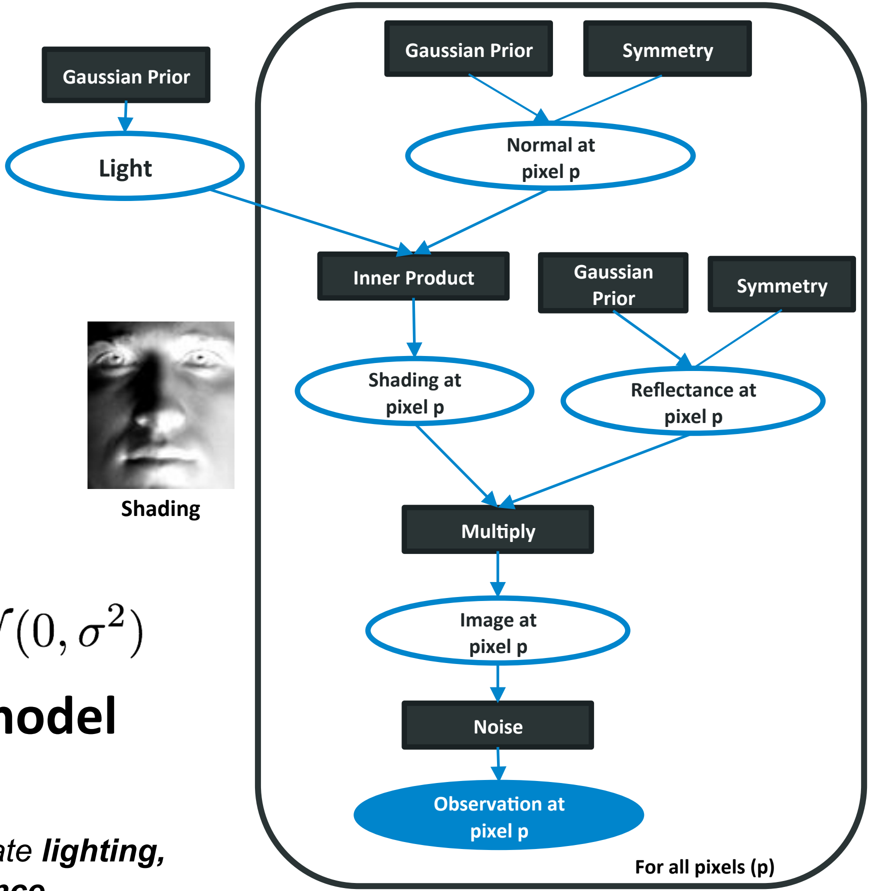
[CVIU’15]. **Jampani, V.**, Nowozin, S., Loper, M., & Gehler, P. V. The informed sampler: A discriminative approach to Bayesian inference in generative computer vision models. *CVIU*, 2015.

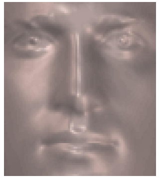
Consensus Message Passing

for layered graphical models



Light

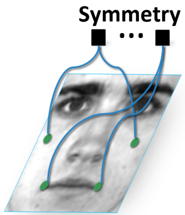




Normal Map

Shading

Reflectance



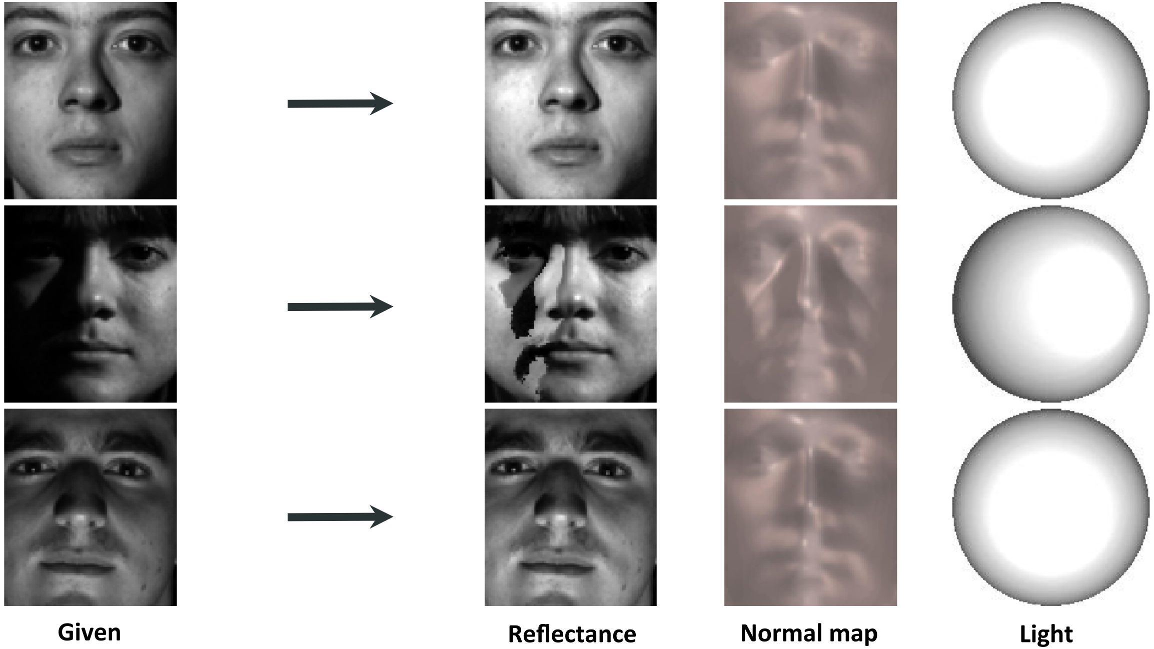
Observation

$$I = R \times (\vec{N}.\vec{L}) + \mathcal{N}(0, \sigma^2)$$

Face generative model

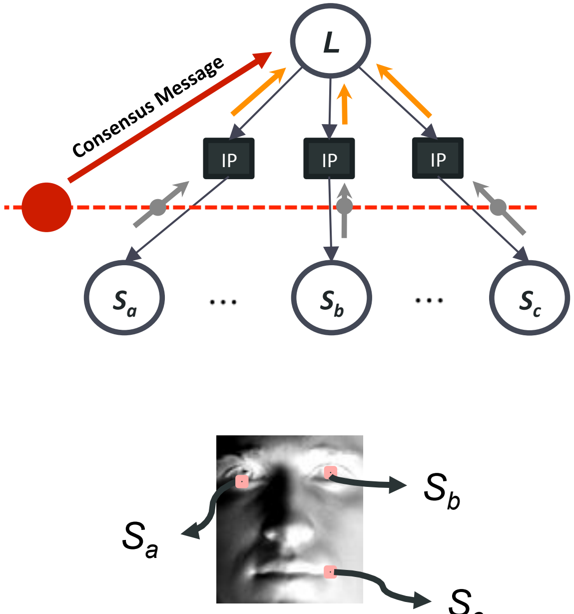
Task:
*Given image observation, estimate **lighting, normals, shading** and **reflectance**.*

Results with Variational Message Passing [1]

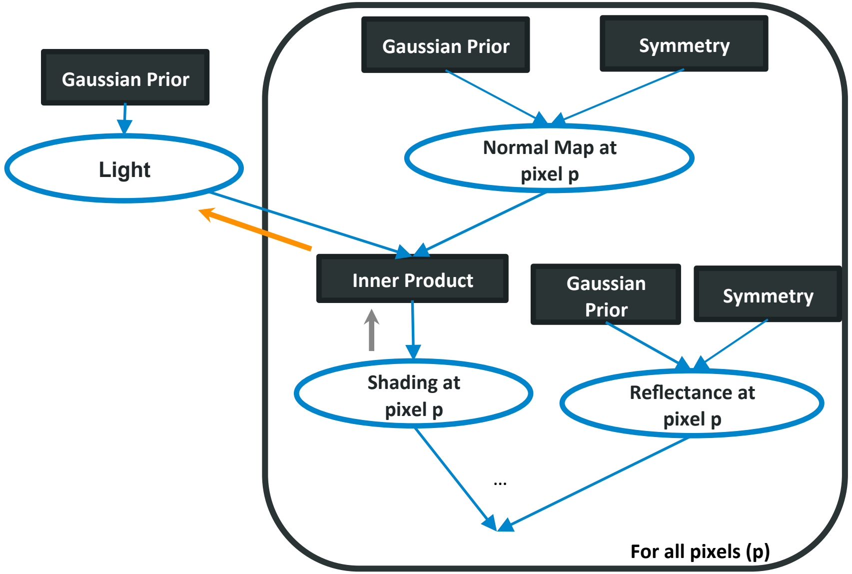


1. Winn, J. and Bishop, C.M., Variational message passing. *Journal of Machine Learning Research,* 2005.

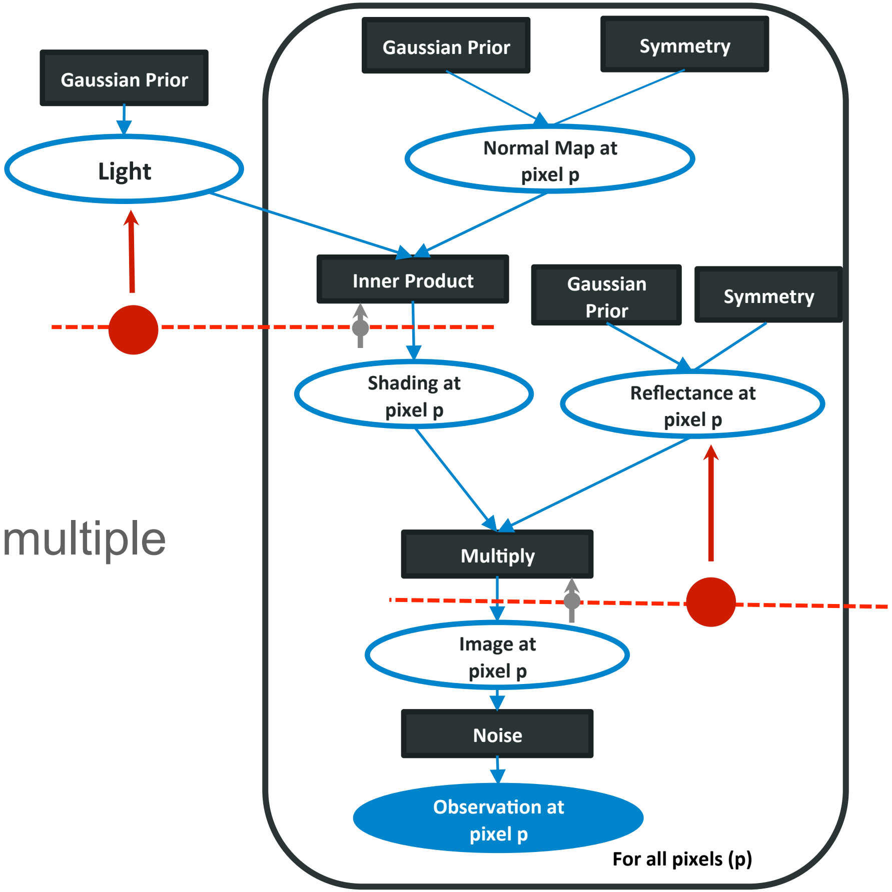
Consensus Message Passing [AISTATS'15]



Example Shading



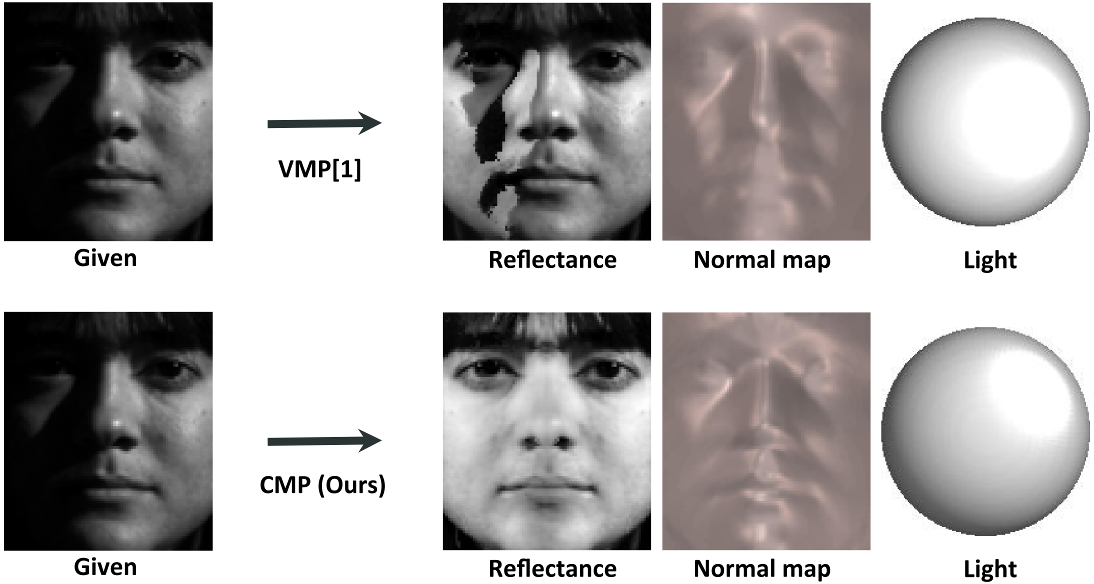
[AISTATS'15]. **Jampani, V.**, Eslami, S. M., Tarlow, D., Kohli, P., & Winn, J. Consensus Message Passing for Layered Graphical Models. In *AISTATS*, 2015.



Consensus message predictors in multiple model layers

Random forest regression

A Visual Result



1. Winn, J. and Bishop, C.M., Variational message passing. *Journal of Machine Learning Research,* 2005.

Visual results – Reflectance
For different subjects

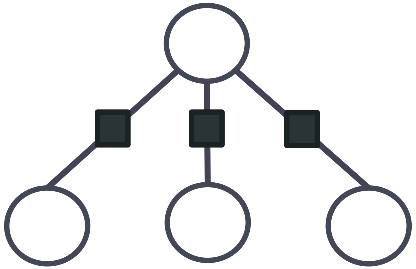
1. Winn, J. and Bishop, C.M., Variational message passing. *Journal of Machine Learning Research,* 2005.
2. Biswas, S., Aggarwal, G. and Chellappa, R., Robust estimation of albedo for illumination-invariant matching and shape recovery. *PAMI,* 2009.

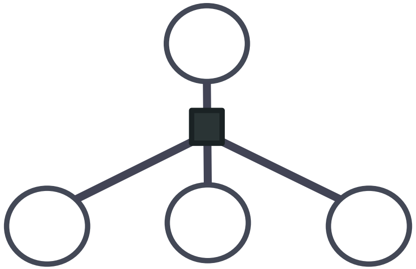
Remarks: Consensus Message Passing [AISTATS'15]

CMP helps message passing converge to better solutions.

Demonstrated using 3 different layered graphical models.

Can be seen as a good trade-off between factored and joint models.





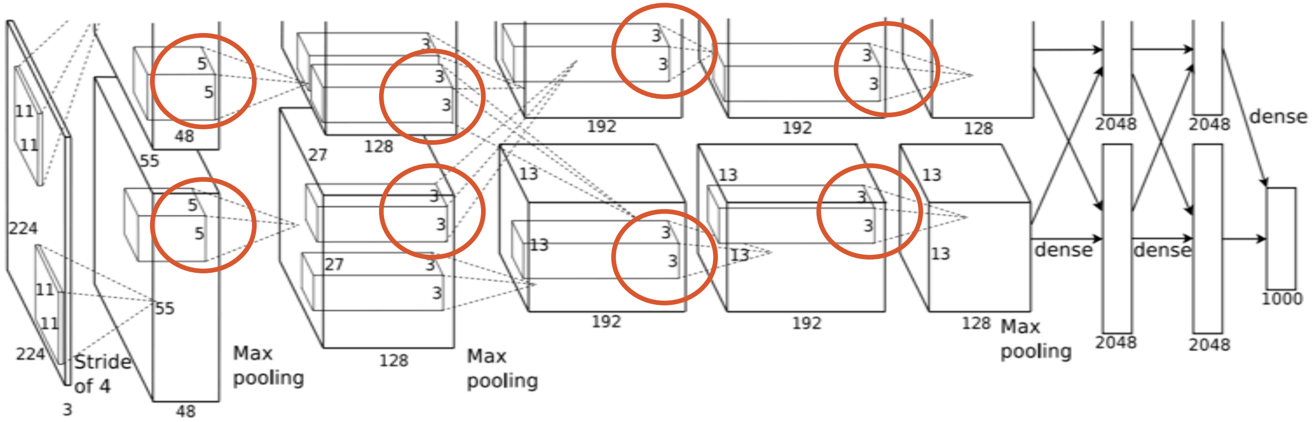
[AISTATS'15]. **Jampani, V.**, Eslami, S. M., Tarlow, D., Kohli, P., & Winn, J. Consensus Message Passing for Layered Graphical Models. In *AISTATS,* 2015.

Part 2: Inference in Discriminative Computer Vision Models

Convolutional Neural Networks (CNN)

Spatial convolutions form the building block of most CNN architectures

Perhaps the *simplest*, *fastest* and *most used* way of propagating information across pixels

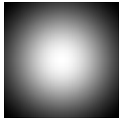


Schematic of AlexNet CNN [1]

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

Spatial Filter

$*$



Gauss filter

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

$=$

After Gaussian Filtering (smoothes image)

Filter values depends on position offset with respect to the center

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

Output filter Position Input

Goal: *To equip neural networks with richer class of filters*

Bilateral Filter

Generalization of spatial convolution to arbitrary features [1, 2]

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

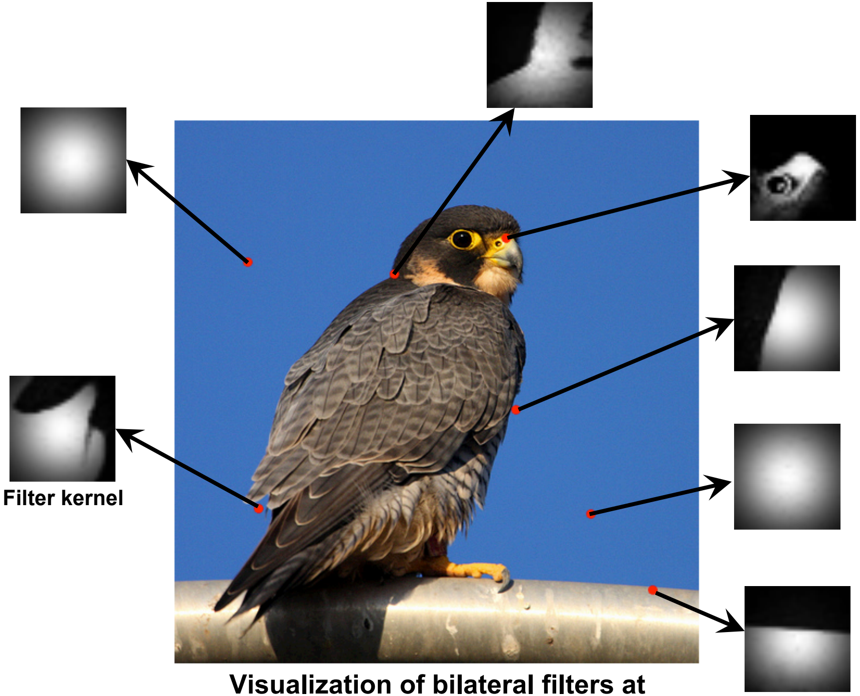
Spatial Filtering **Bilateral Filtering**

Position and color features : $\mathbf{f} = (x, y, r, g, b)^\top$

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

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



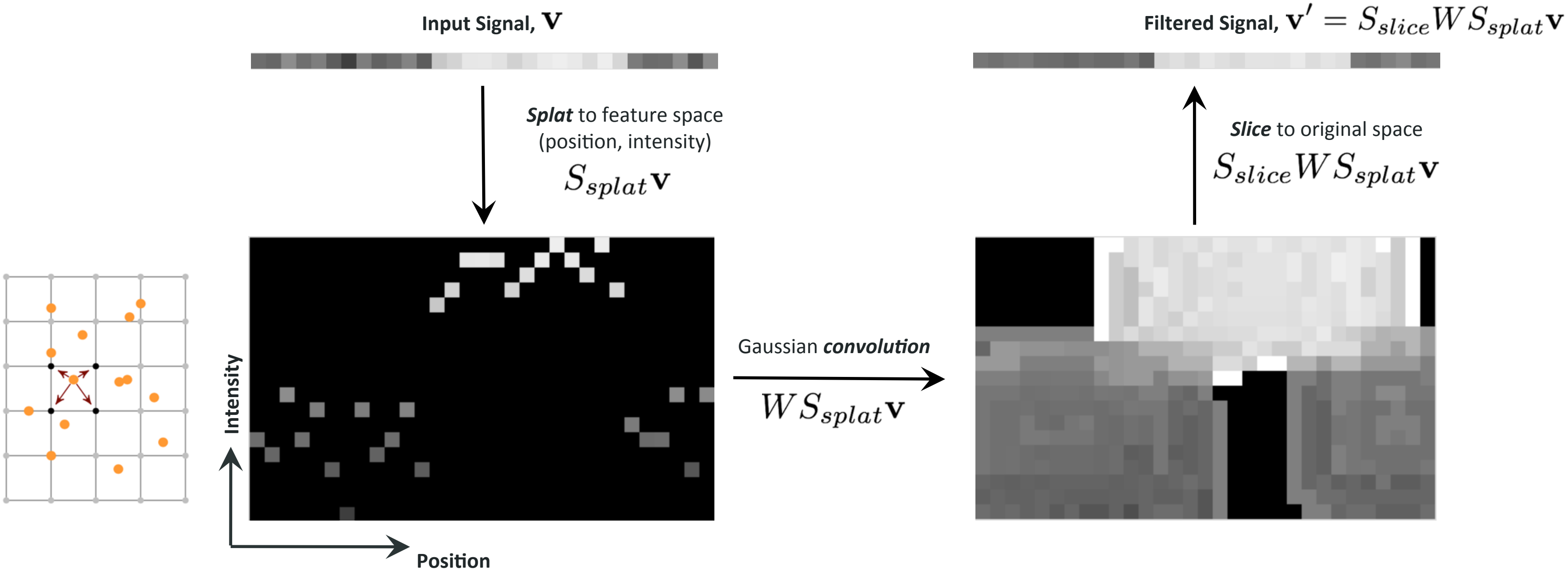
Visualization of bilateral filters at different input locations

After bilateral filtering (preserves edges)

Computationally expensive!

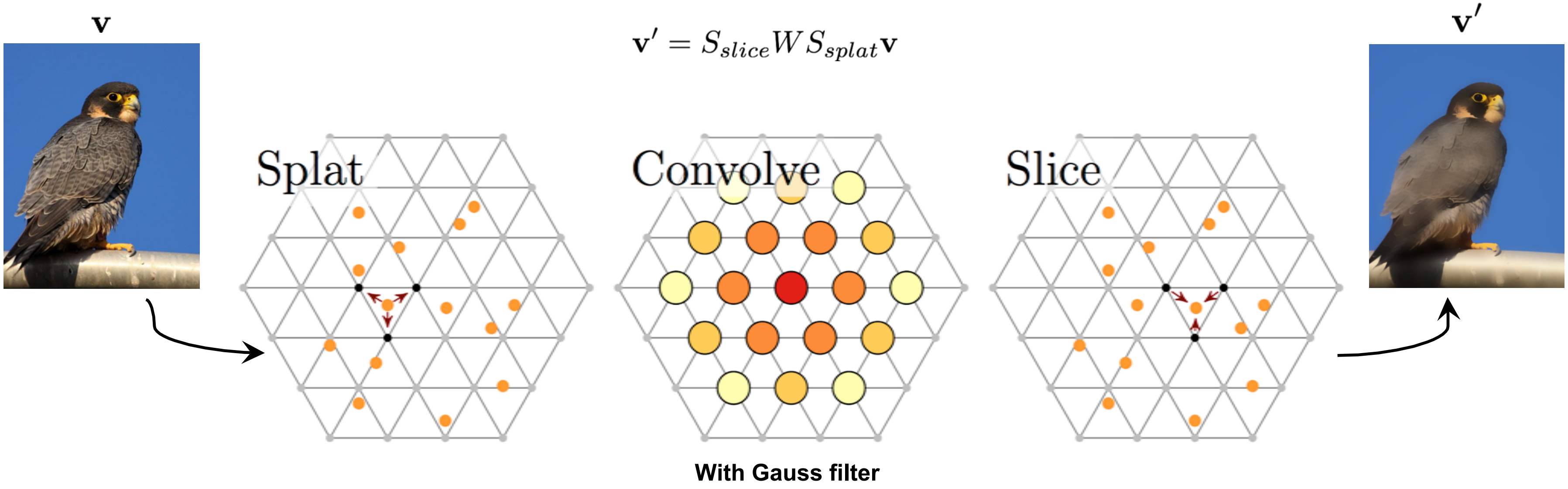
High dimensional linear approximation

Convolution in high-dimensional feature space [1]



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

Using the permutohedral lattice [1]



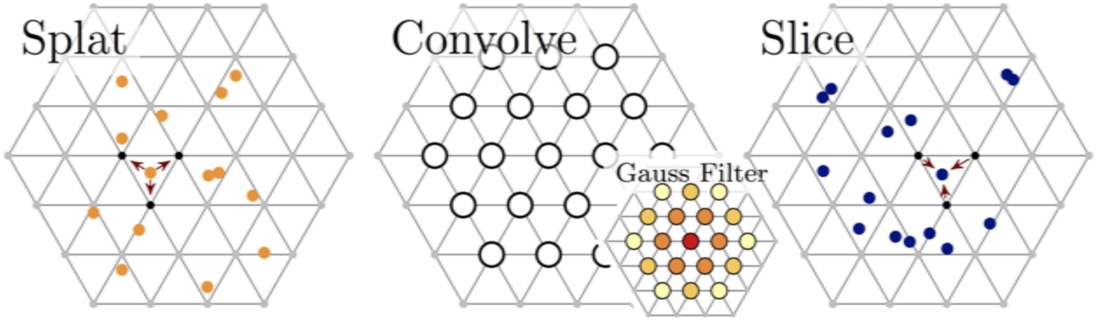
All the existing works use *hand-designed* filter weights (mostly, Gaussian)!

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

Learning Bilateral Filters [CVPR'16, ICLR Workshop'15]

Parameterize the filter instead of using Gaussian

$$\mathbf{v}' = S_{slice} W S_{splat} \mathbf{v}$$



Back-propagate through linear operations of *splat*, *convolve* and *slice*

Can be easily integrated into any existing deep learning architectures.

[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.
[ICLR Workshop'15]. Kiefel, M., **Jampani, V.**, & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

Contributions [CVPR’16, ICLR Workshop’15]

1. Learn **problem specific bilateral filters**

2. Generalize CNNs to **Bilateral Neural Networks** (BNN)

3. **Generalize Dense CRFs** (fully-connected Conditional Random Fields)

[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.
[ICLR Workshop’15]. Kiefel, M., **Jampani, V.**, & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

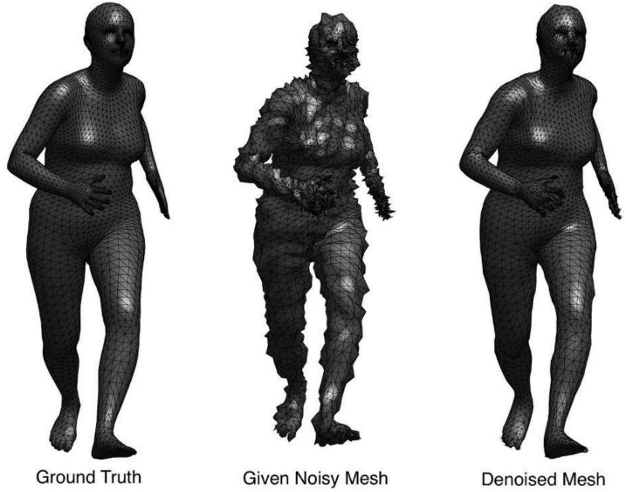
Learning Problem Specific Bilateral Filters

Applications of bilateral filter

Many vision, graphics and image processing applications [1]

We studied the use of filter learning [CVPR'16] in following applications:

- Joint bilateral upsampling [2]
- Image denoising [3]
- 3D mesh denoising [4,5]



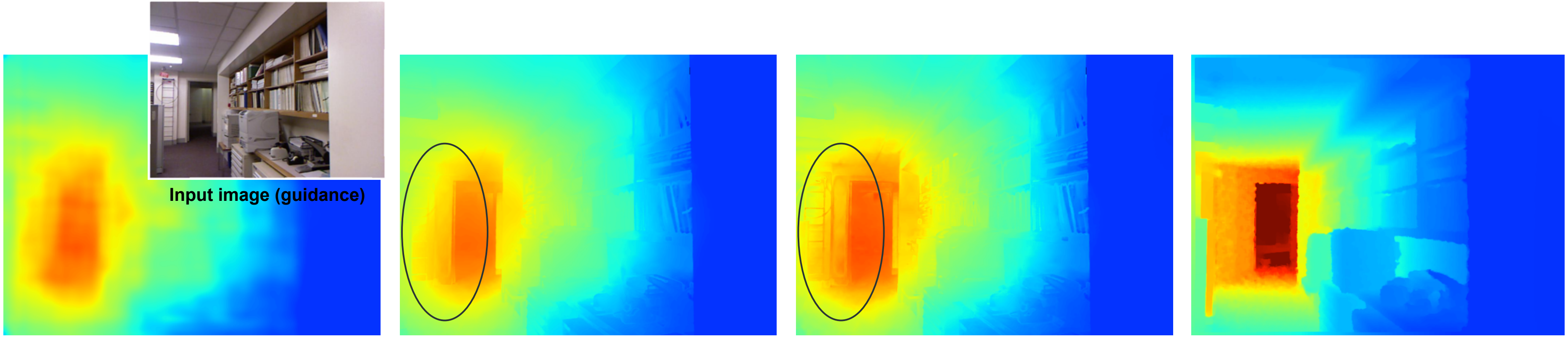
3D mesh denoising using *learned* bilateral filtering, with normals as features

[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.
1. Paris, S., Kornprobst, P., Tumblin, J., & Durand, F. Bilateral filtering: Theory and applications. *Now Publishers Inc.*, 2009.
2. Kopf, J., Cohen, M. F., Lischinski, D., & Uyttendaele, M. Joint bilateral upsampling. *ACM Transactions on Graphics (TOG)*, 2007.
3. Tomasi, C., & Manduchi, R. Bilateral filtering for gray and color images. In *ICCV,* 1998.
4. Fleishman, S., Drori, I., & Cohen-Or, D. Bilateral mesh denoising. In *ACM Transactions on Graphics (TOG),* 2003.
5. Jones, T. R., Durand, F., & Desbrun, M. Non-iterative, feature-preserving mesh smoothing. In *ACM Transactions on Graphics (TOG)* 2003.

Observations

Consistent improvement of learned bilateral filter w.r.t. Gaussian filter.

An example depth upsampling result:



CNN depth [1] (bicubic interpolated) | **Gauss bilateral upsampling** | **Learnt bilateral upsampling (Ours)** | **Ground Truth depth**

1. Eigen, D., Puhrsch, C., & Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In *NIPS,* 2014 (pp. 2366-2374).
2. Silberman, N., Hoiem, D., Kohli, P., & Fergus, R. (2012). Indoor segmentation and support inference from RGBD images. In *ECCV, 2012* (pp. 746-760).

Bilateral Neural Networks

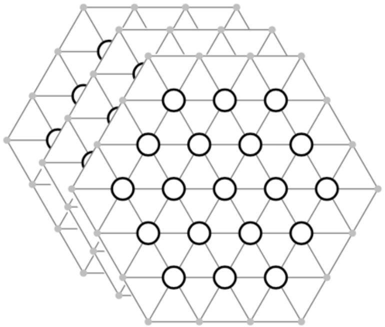
Bilateral neural networks [CVPR’16, ICLR Workshop’15]

Learning bilateral filters allows the use of stacked filters

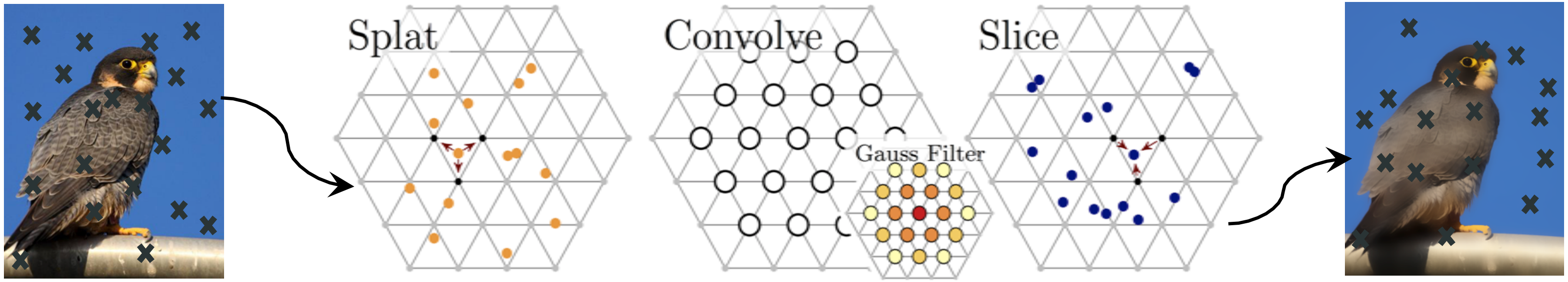
BNN: Bilateral Neural Network

Advantages of BNNs over CNNs:

- Image adaptive Filtering
- Filtering unordered set of points (e.g., sparse 3D points)
- Input and Output points can be different



Bilateral 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.
[ICLR Workshop’15]. Kiefel, M., **Jampani, V.**, & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

Application: Video Propagation

How can we propagate information across video frames?

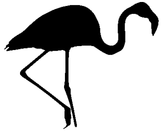
E.g.: Video Object Segmentation


Frame 1


Frame 13


Frame 22


Frame 63

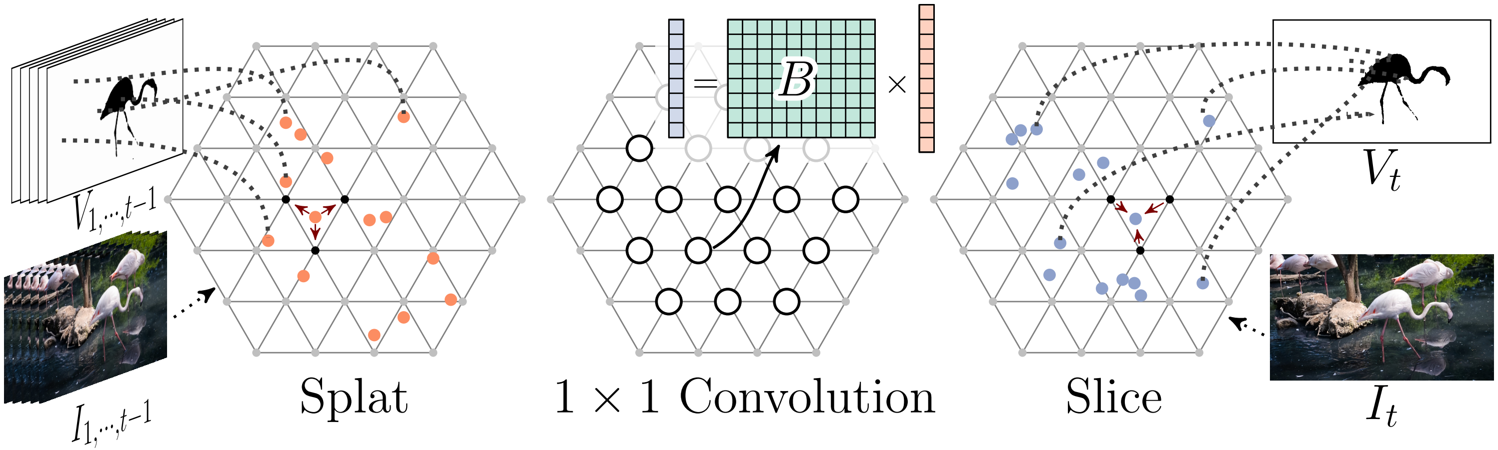

Given Mask

→ ? → ? → ?

Video Propagation with Bilateral Filters [CVPR'17]

Splat *previous* frame results, convolve and slice at the *present* frame.

Bilateral feature space with **time**: $\mathbf{f} = (x, y, r, g, b, t)^\top$

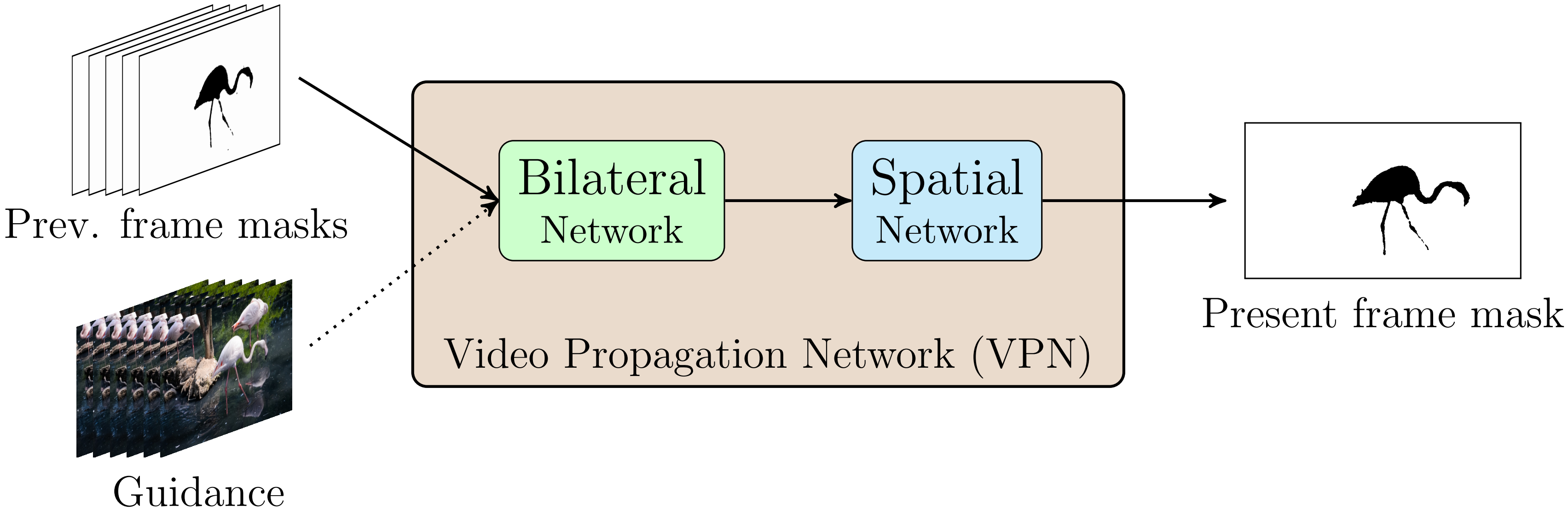


Bilateral Convolution Layer (BCL)

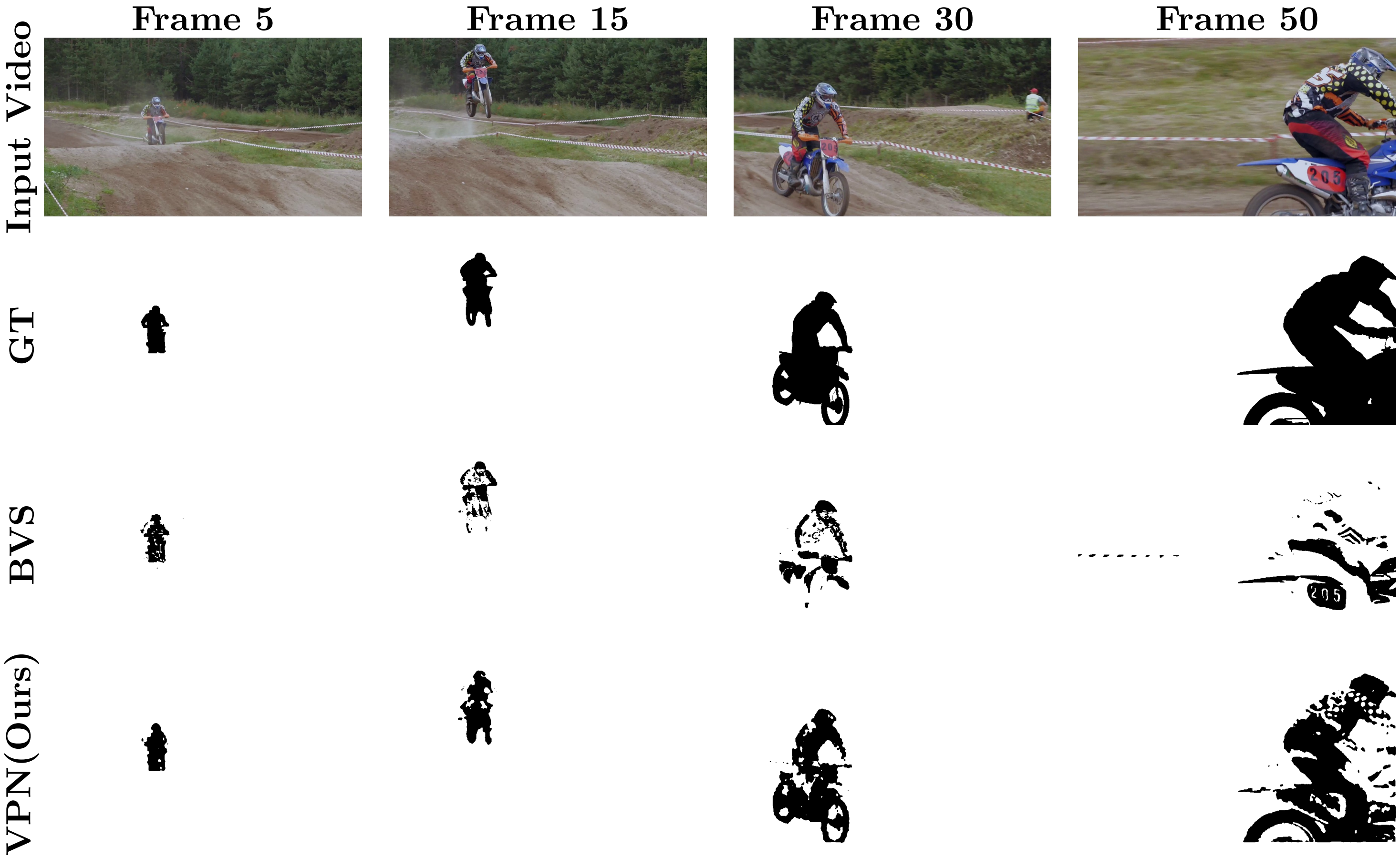
[CVPR'17]. **Jampani, V.**, Gadde, R. and Gehler, P.V., Video Propagation Networks. In *CVPR,* 2017.

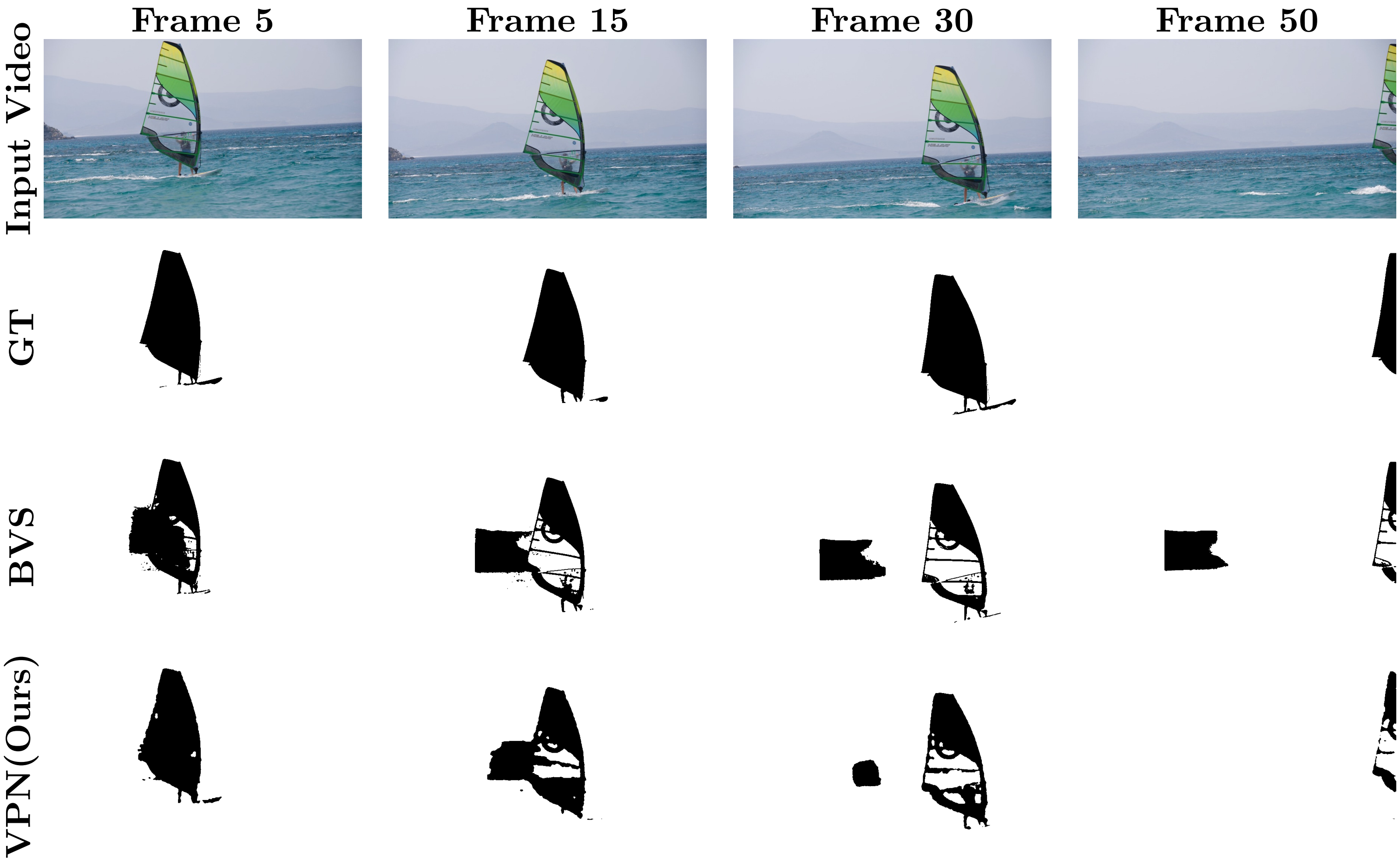
Overview of Video Propagation Networks [CVPR'17]

End-to-end trained *fast* neural network with *video-adaptive* receptive fields



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


Frame 5
Frame 15
Frame 30
Frame 50
Input Video
GT
BVS
VPN(Ours)


Frame 5
Frame 15
Frame 30
Frame 50
Input Video
GT
BVS
VPN(Ours)

Generalize Dense CRFs

Dense CRFs (review)

Every pixel is connected to every other pixel via pairwise terms [1]

$$p(x|v) \propto \exp\left(-\sum_i \underbrace{\psi_u(x_i)}_{\text{Unary}} - \sum_{i>j} \underbrace{\psi_p(x_i, x_j)}_{\text{Pairwise}}\right)$$

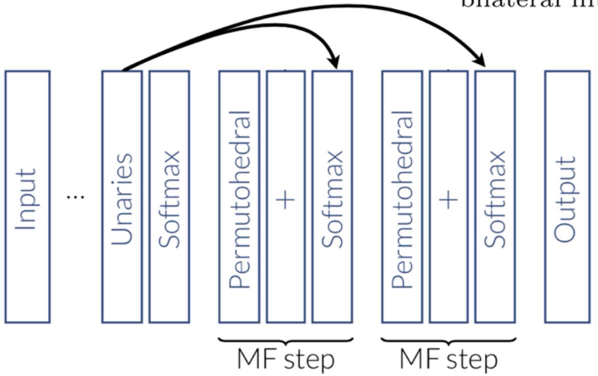
Many applications in vision: segmentation [1], optical flow [2], intrinsic images [3] etc.

1. Krähenbühl, P., & Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *NIPS,* 2011.
2. Sun, D., Wulff, J., Sudderth, E. B., Pfister, H., & Black, M. J. A fully-connected layered model of foreground and background flow. In *CVPR,* 2013.
3. Bell, S., Bala, K., & Snavely, N. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 2014.

Learning Pairwise Potentials

Mean-field updates can be computed using bilateral filter [1]:

$$\underbrace{q_i^{t+1}(x_i)}_{\text{beliefs}} = \frac{1}{Z_i}\exp\{-\underbrace{\psi_u(x_i)}_{\text{unary}} - \sum_{x_j \in \mathcal{L}} \underbrace{\sum_{j \neq i} \overbrace{\psi_p(x_i, x_j)}^{\text{pairwise}}\, q_j^t(x_j)}_{\text{bilateral filtering}}\}$$



Learning bilateral filters → *Learning* pairwise potentials (learned CRF)

1. Krähenbühl, P., & Koltun, V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *NIPS,* 2011 (pp. 109-117).

Segmentation results with learned CRF [CVPR'16]

Experiments on **Semantic segmentation** and **Material segmentation**

Average IoU Scores for Semantic Segmentation [2]

Deeplab [1] (72.1 / 67.0)	**1-step MF**	**2-step MF**	***loose* 2-step MF**
Gauss CRF	+2.5	+3.4	+3.4 / +3.0
Learned CRF (Ours)	+3.0	**+3.7**	**+3.9 / +3.4**

Average pixel accuracy for Material Segmentation [3]

AlexNet (67.2)	**1-step MF**	**2-step MF**	***loose* 2-step MF**
Gauss CRF	+7.9	+9.7	+9.7
Learned CRF (Ours)	+9.5	**+11.9**	**+11.9**

Input | **Ground Truth** | **Deeplab CNN result** | **+ 2-stepMF Gauss CRF** | **+ 2-step MF *learned* CRF (Ours)**

[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.
1. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *ICLR,* 2015.
2. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. The Pascal visual object classes (VOC) challenge. IJCV, 2010, *88*(2), 303-338.
3. Bell, S., Upchurch, P., Snavely, N. and Bala, K., Material recognition in the wild with the materials in context database. In *CVPR,* 2015.

Dense CRFs are good but time taking.

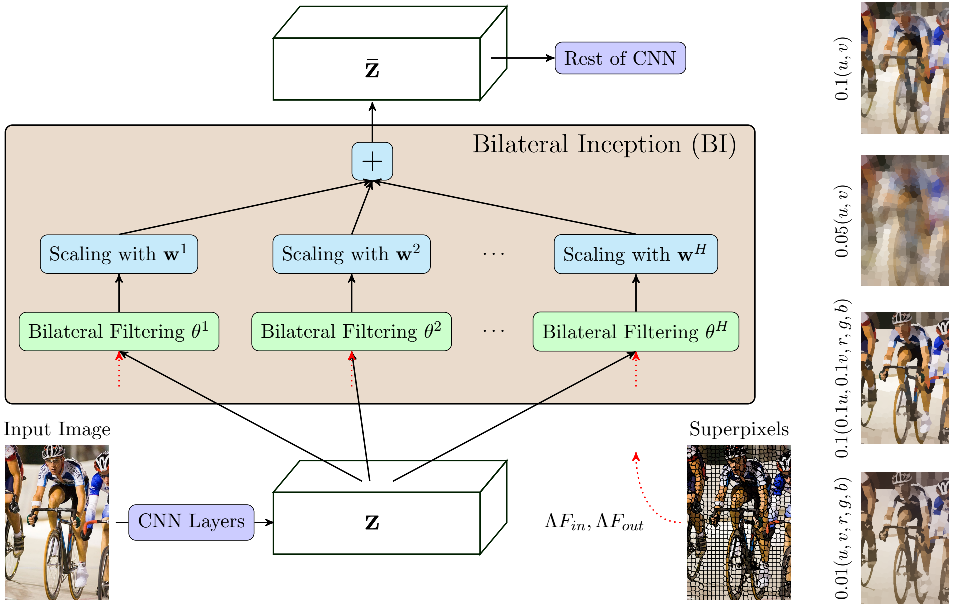
*Can we do **dense information propagation** inside CNN itself?*

Bilateral Inception Module [ECCV'16]

Bilateral filtering intermediate CNN representations with different feature scales

Long-range information propagation between CNN units

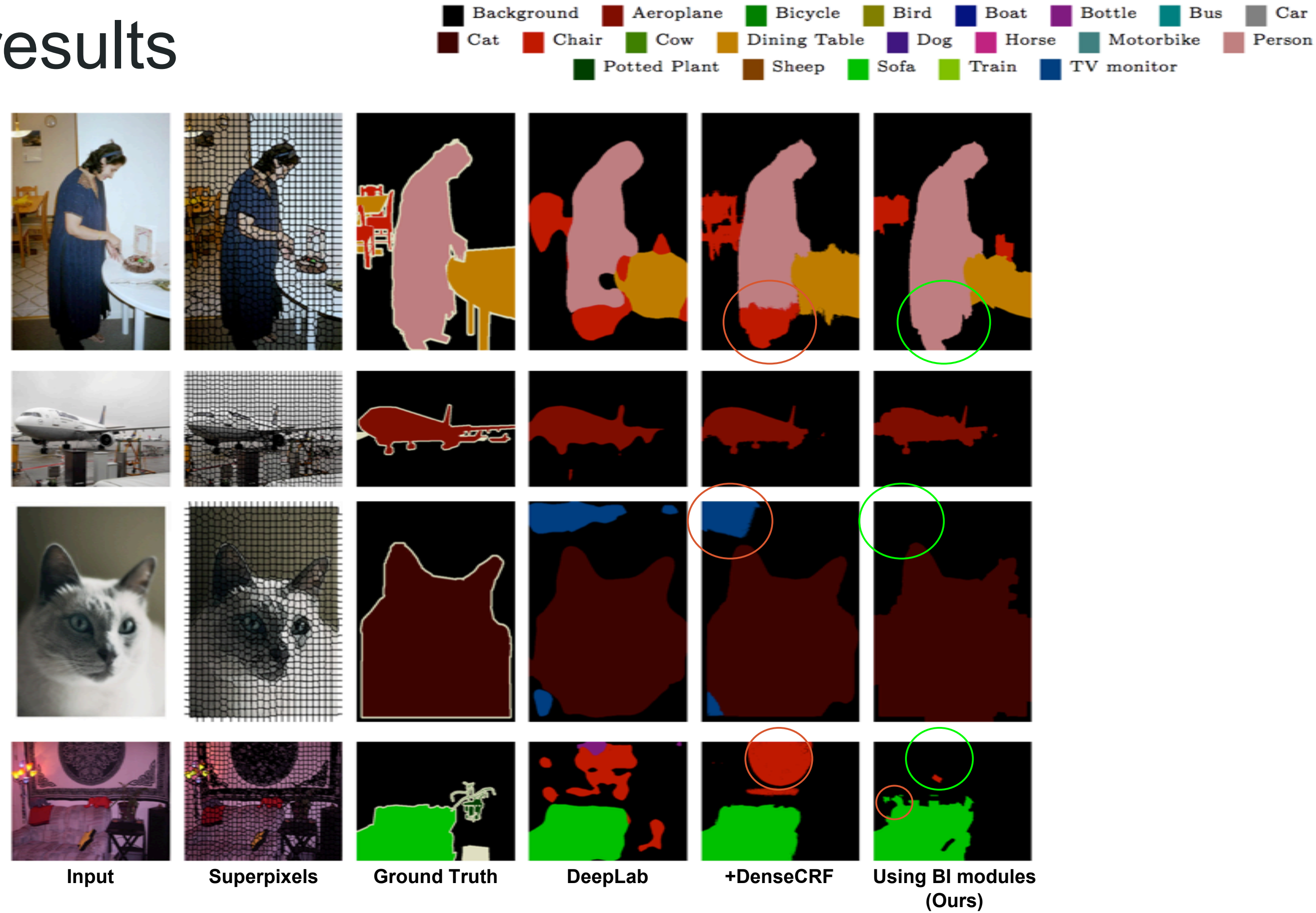
Faster and *better* compared to DenseCRF



Schematic of bilateral inception module

[ECCV'16]. Gadde, R., **Jampani, V.**, Kiefel, M., Kappler, D. and Gehler, P.V., Superpixel convolutional networks using bilateral inceptions. In *ECCV*, 2016.

Visual results



Remarks

Bilateral filters provide simple yet rich framework for information propagation

Learning bilateral filters with several contributions [ICLR Workshops’15, CVPR’16, CVPR’17]

➔ **Problem specific filters** can be learned

➔ Edge-aware CNNs: **Bilateral neural networks**

➔ Generalize the dense CRF to **non-Gaussian pairwise potentials**

Dense information propagation inside CNNs (**bilateral inception modules**) [ECCV’16]

Wide range of applications in vision and graphics

[ICLR Workshop’15]. Kiefel, M., **Jampani, V.**, & Gehler, P. V. Permutohedral Lattice CNNs. *In ICLR Workshop,* 2015.

[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.

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

[ECCV’16]. Gadde, R., **Jampani, V.**, Kiefel, M., Kappler, D. and Gehler, P.V., Superpixel convolutional networks using bilateral inceptions. In *ECCV,* 2016 *(pp. 597-613).*

Conclusion and Future Work

Generative and *discriminative* approaches for vision

Informed sampler and **Consensus message passing** for inference in generative models

Learning bilateral filters help in bringing prior knowledge *into* CNNs

Future outlook:

- How to bring more prior knowledge into CNNs while maintaining fast runtime?
- Bridging the gap between generative and discriminative models

Publications

Thesis related publications:

- **Jampani, V.**, Nowozin, S., Loper, M., & Gehler, P. V. The informed sampler: A discriminative approach to Bayesian inference in generative computer vision models. *CVIU*, 2015.
- **Jampani, V.**, Eslami, S. M., Tarlow, D., Kohli, P., & Winn, J. Consensus Message Passing for Layered Graphical Models. In *AISTATS*, 2015.
- Kiefel, M., **Jampani, V.**, & Gehler, P. V. Permutohedral Lattice CNNs. In *ICLR Workshop*, 2015.
- **Jampani, V.**, Kiefel, M., & Gehler, P. V. Learning Sparse High Dimensional Filters: Image Filtering, Dense CRFs and Bilateral Neural Networks. In *CVPR*, 2016.
- Gadde, R., **Jampani, V.**, Kiefel, M., Kappler, D. and Gehler, P.V., Superpixel convolutional networks using bilateral inceptions. In *ECCV*, 2016.
- **Jampani, V.**, Gadde, R. and Gehler, P.V., Video Propagation Networks. In *CVPR*, 2017.

Others:

- **Jampani, V.**, Gadde, R. and Gehler, P.V., Efficient Facade Segmentation Using Auto-Context. In *WACV*, 2015.
- Sevilla-Lara, L., Sun, D., **Jampani, V.** and Black, M.J., Optical flow with semantic segmentation and localized layers. In *CVPR*, 2016.
- Gadde,R., **Jampani, V.**, Marlet, R. and Gehler, P.V., Efficient 2D and 3D Facade Segmentation using Auto-Context. *PAMI*, 2017.