

ComputerVisionTopics

Version 2017 by MVPLAB

Topics by CV Society

1. Low Level Vision

- Filtering / Grouping / Enhancement...

2. High Level Vision

- Detection / Recognition / Understanding...

3. Graphics

- Color / Style / Editing / Rendering...

Contents

1. Restoration/Generation
2. Enhancement
3. Segmentation
4. 3D Vision
5. Recognition/Detection/Retrieval
6. Tracking/Motion/Action
7. Understanding
8. Vision + Language
9. Graphics
10. Machine Learning

Today, We Introduce
63 Papers

Restoration/Generation

Denoising, Deblurring, SuperResolution
Completion, Generation

Image Denoising

Non-local Color Image Denoising with Convolutional Neural Networks **CVPR 2017**



(a)



(b)

Figure 1. Image denoising with the proposed deep non-local CNN model. (a) Noisy image corrupted with additive Gaussian noise ($\sigma = 25$) ; PSNR = 20.16 dB. (b) Denoised image using the 5-stage feed-forward network described in Sec. 3.3 ; PSNR = 29.53 dB.

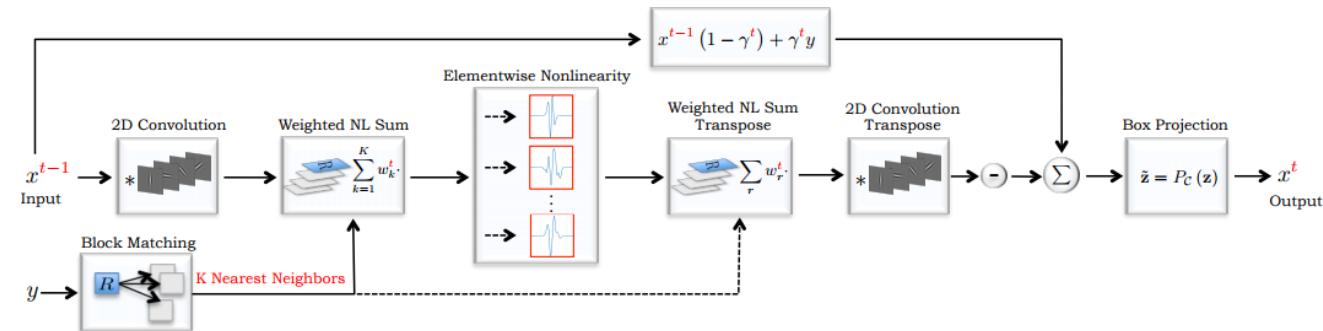


Image Deblurring

Self-paced Kernel Estimation for Robust Blind Image Deblurring **ICCV 2017**
Deep Video Deblurring for Hand-held Cameras **CVPR 2017**



(a) Blurry image



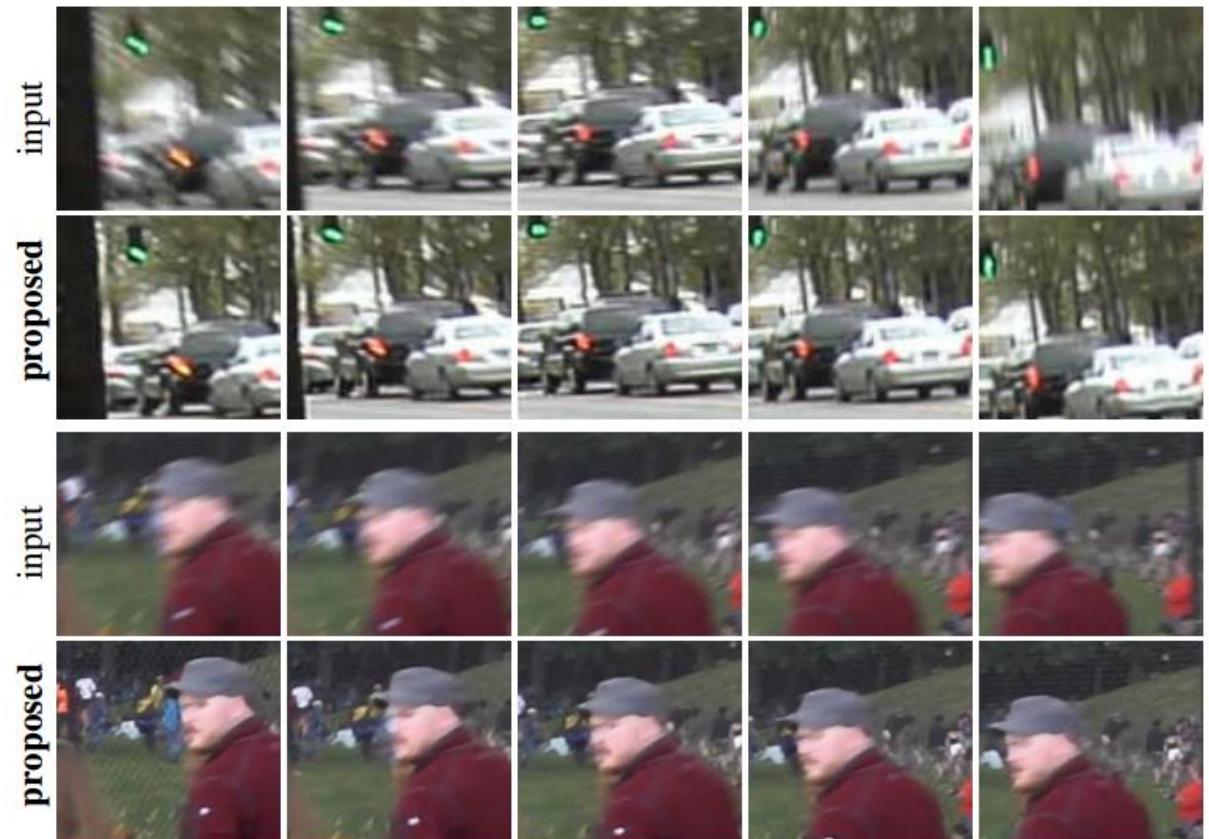
(b) Xu and Jia [37]



(c) Pan *et al.* [26]

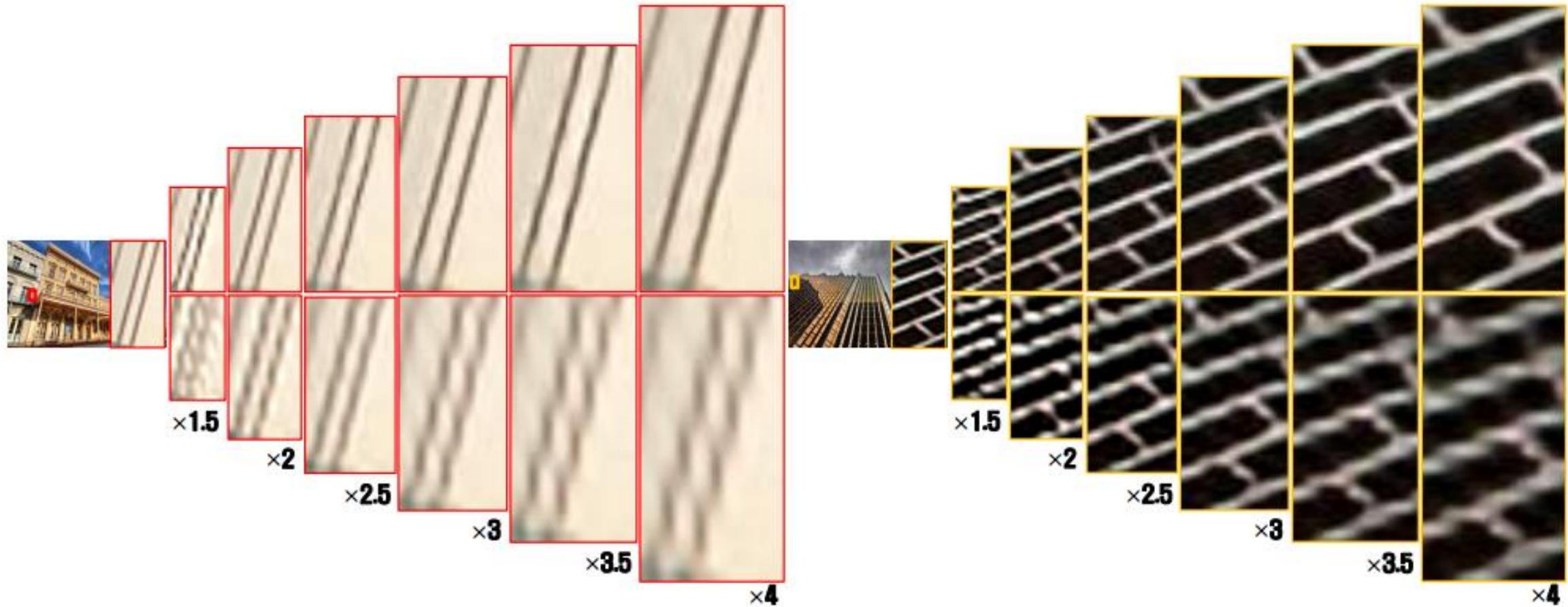


(d) Ours



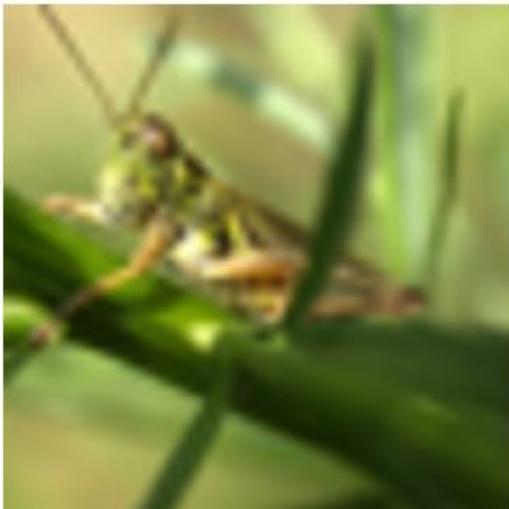
Super Resolution

Accurate Image Super-Resolution Using Very Deep Convolutional Networks CVPR 2016

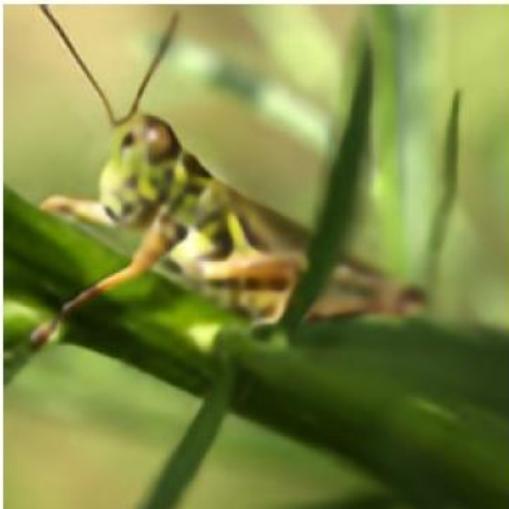


Super Resolution

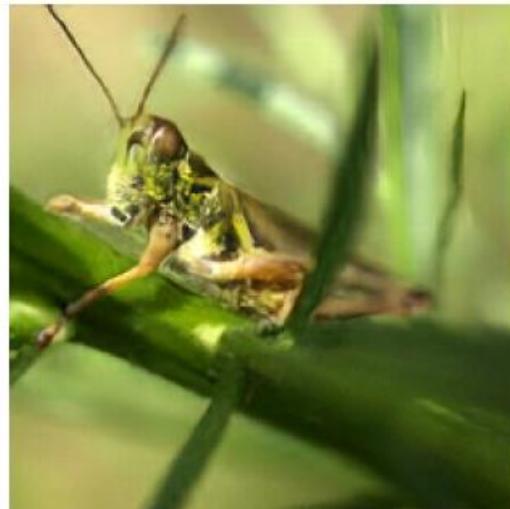
EnhanceNet:Single Image Super-Resolution through Automated Texture Synthesis **ICCV 2017**



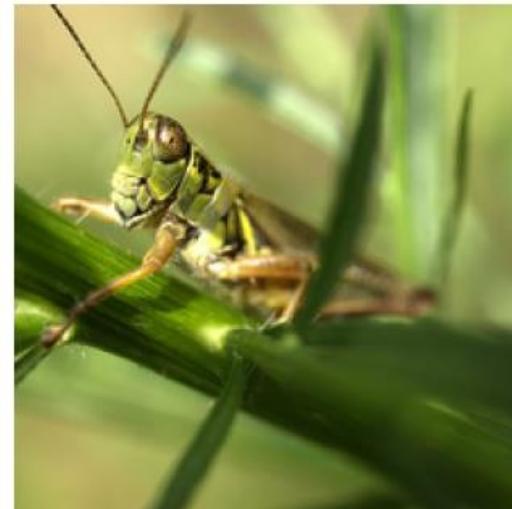
Bicubic



ENet-E



ENet-PAT



Ground Truth

Image Completion

Generative Face Completion CVPR 2017



Image Completion

Pixel RNN Arxiv 2016

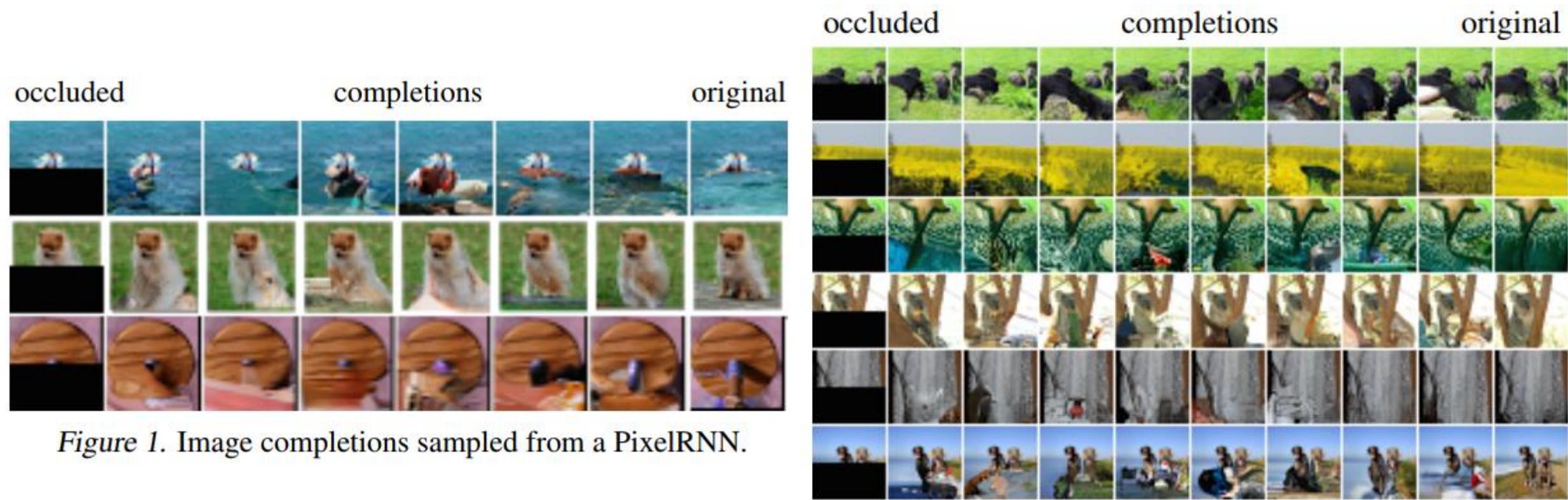


Figure 1. Image completions sampled from a PixelRNN.

Image Generation

Progressive Growing of GANs for Improved Quality, Stability, and Variation [Arxiv 2017](#)



1024x1024



Gulrajani et al. (2017) (128×128)

Our (256×256)

Enhancement

Texture Removal, Depth Enhancement/SuperResolution
Contrast Enhancement, Quality Assessment

Texture Removal

Robust Guided Image Filtering Using Nonconvex Potentials **IEEE TPAMI 2017**



(a) WLS [23].



(b) WLS [23].



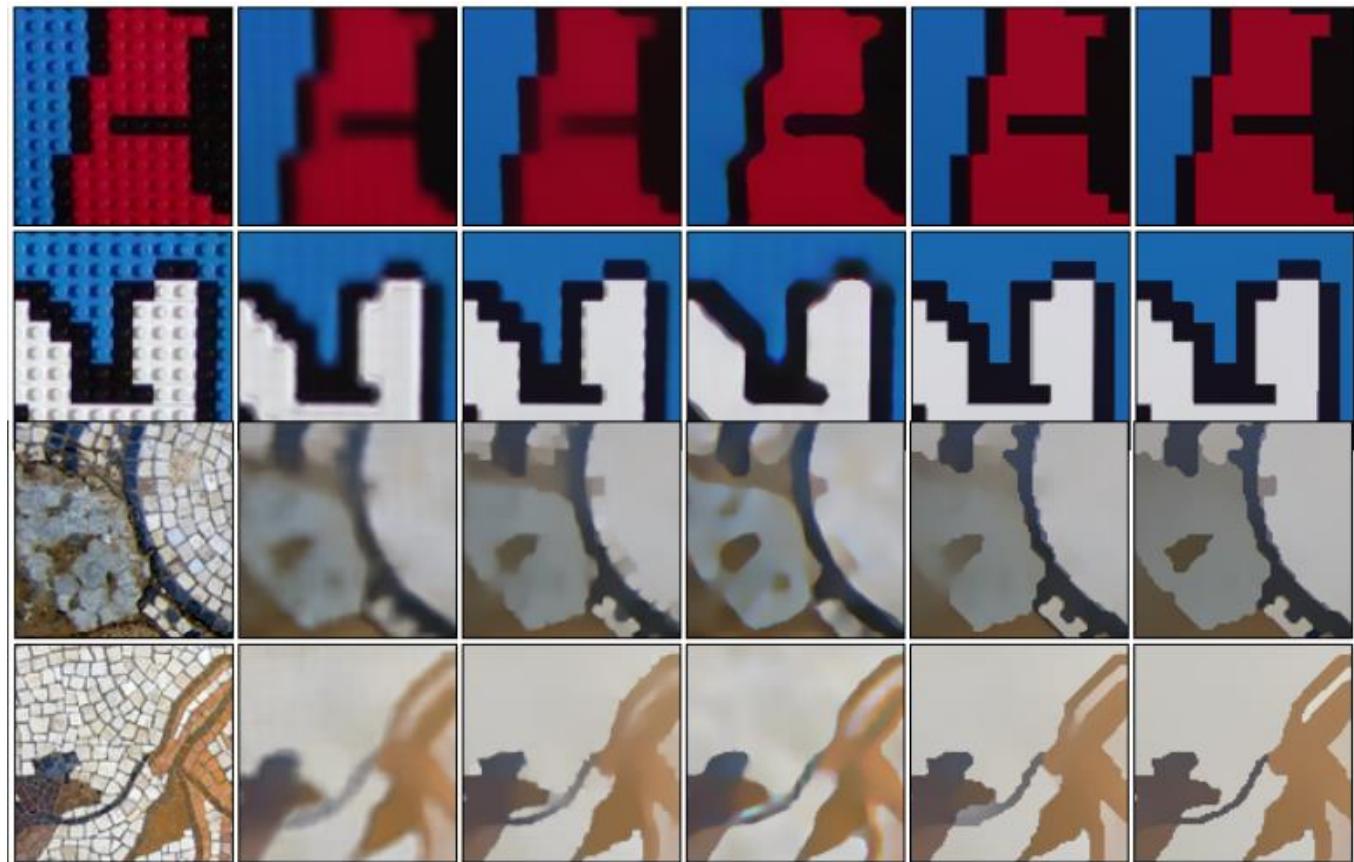
(c) RGF [16].



(d) SD filter ($u^0 = u_{l_2}$).

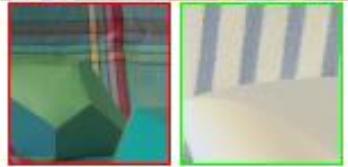


(e) SD filter ($u^0 = u_{l_1}$).

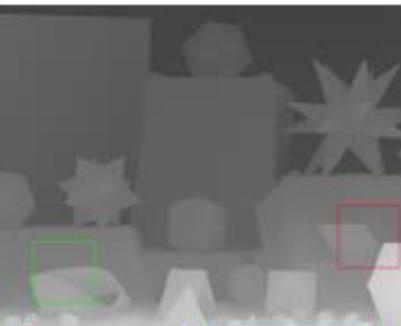


Depth Enhancement

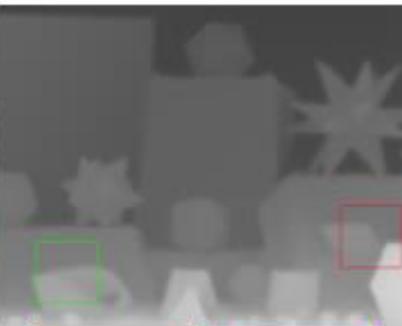
Learning Dynamic Guidance for Depth Image Enhancement **CVPR 2017**



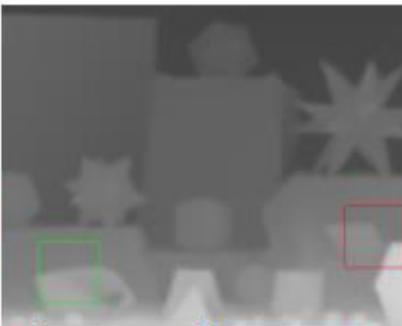
(a) Color Image



(b) Ground Truth



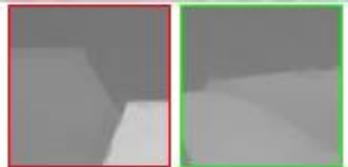
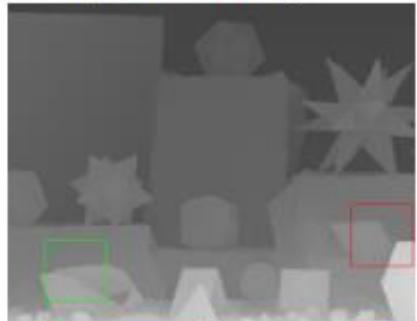
(c) GF [13]



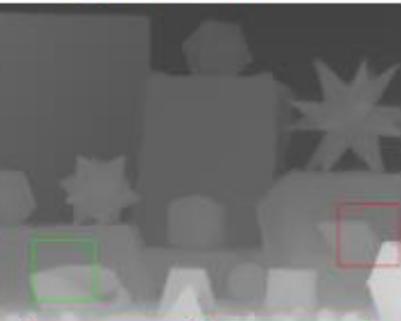
(d) MRF [4]



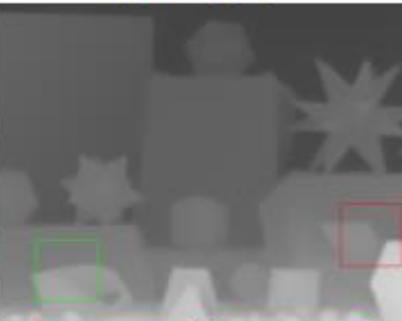
(e) Yang et al. [37]



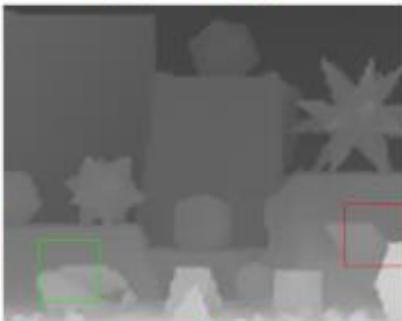
(f) Park et al. [28]



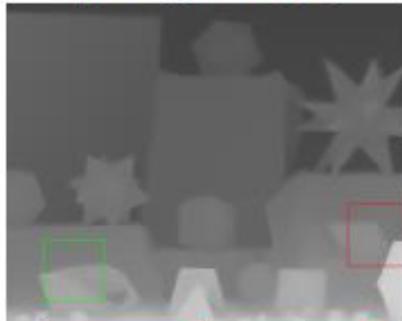
(g) TGV [6]



(h) SDF [12]



(i) DJF [22]



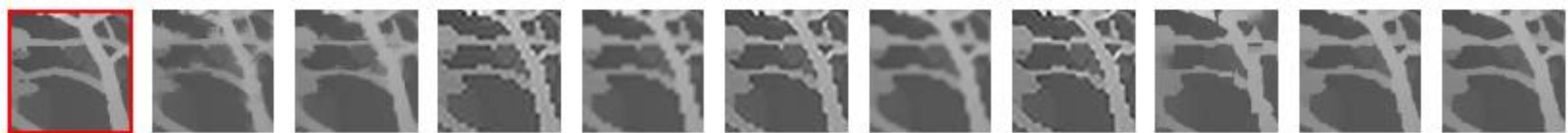
(j) Ours

Depth Super Resolution

Edge-Guided Single Depth Image Super Resolution **IEEE TIP 2016**



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Contrast Enhancement

Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution **IEEE TIP 2013**



(a)

(b)

(c)

(d)

(e)



(f)

(g)

(h)

(i)

(j)

Quality Assessment

Convolutional Neural Networks for No-Reference Image Quality Assessment **CVPR 2014**

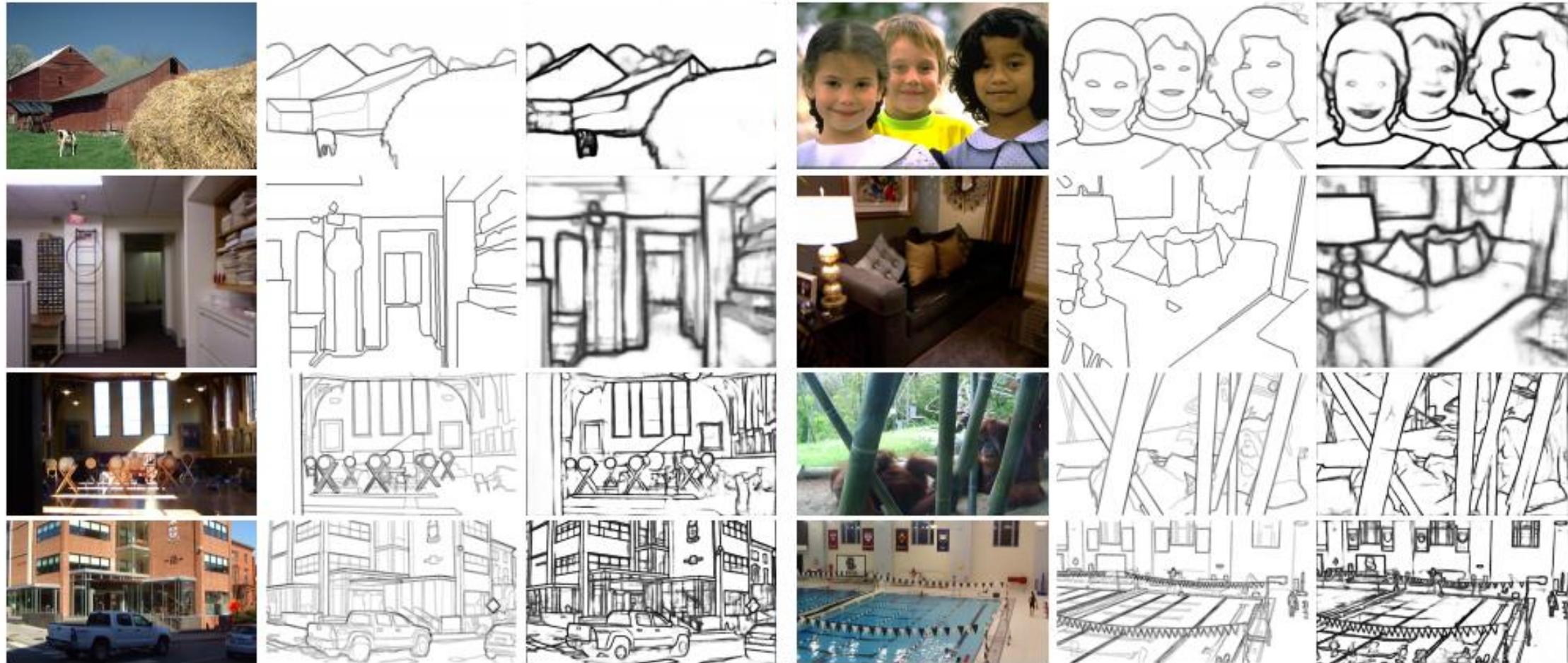


Segmentation

Edge Detection, Superpixels
Semantic Instance Segmentation, Video Segmentation

Edge Detection

Richer Convolutional Features For Edge Detection **CVPR 2017**

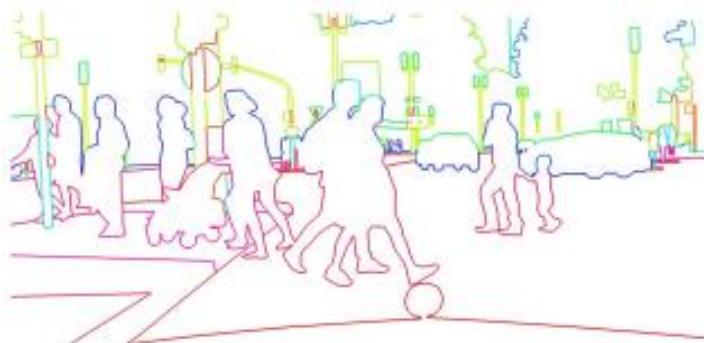


Semantic Edge Detection

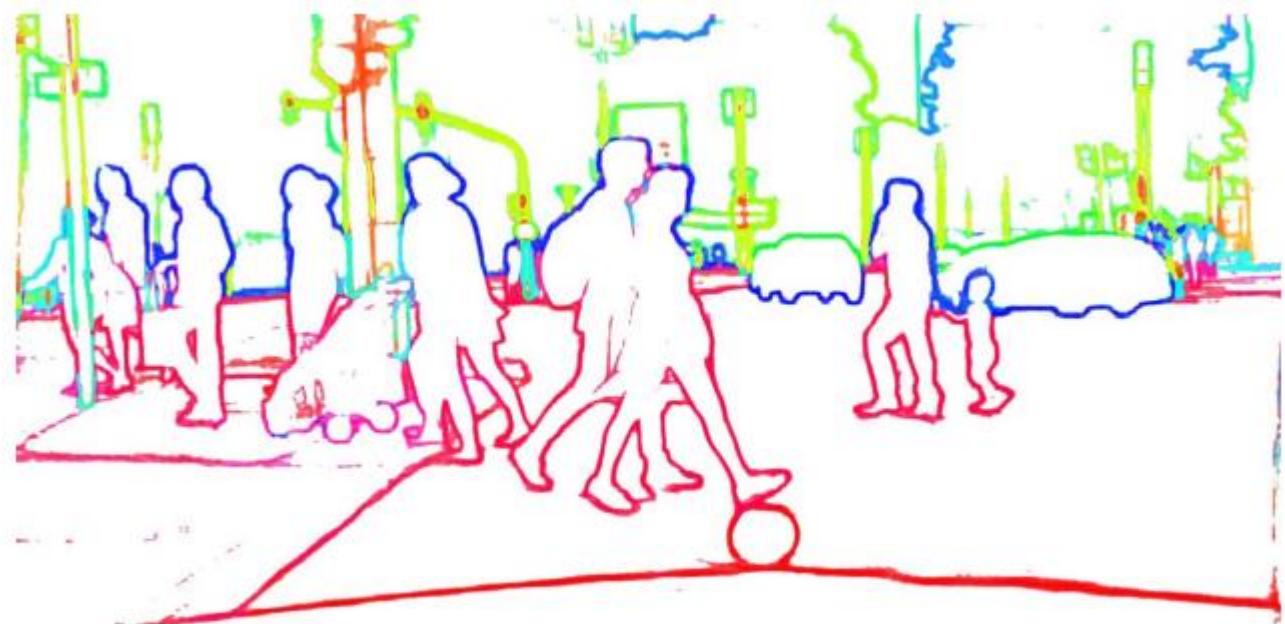
CASENet: Deep Category-Aware Semantic Edge Detection **CVPR 2017**



(a) Input image



(b) Ground truth



(c) CASENet output

building+pole	road+sidewalk	road	sidewalk+building	building+traffic sign	building+car	road+car
building	building+vegetation	road+pole	building+sky	pole+car	building+person	pole+vegetation

Superpixels

Real-Time Coarse-to-fine Topologically Preserving Segmentation **CVPR 2015**

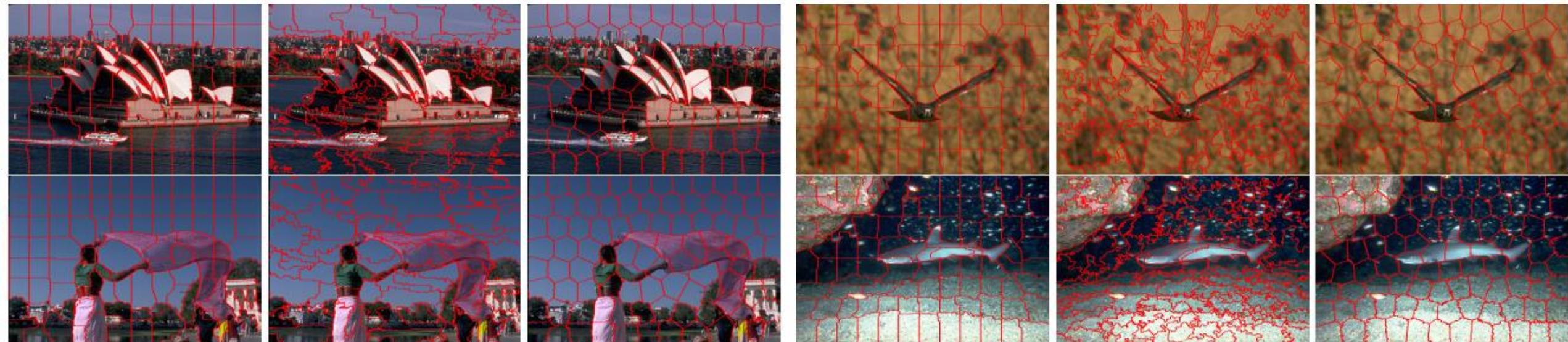
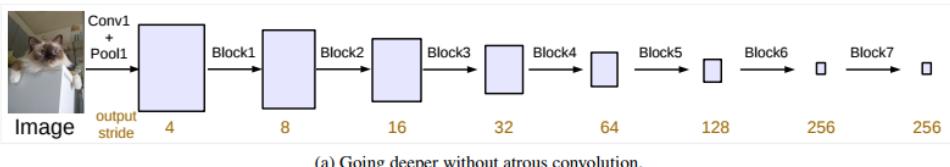


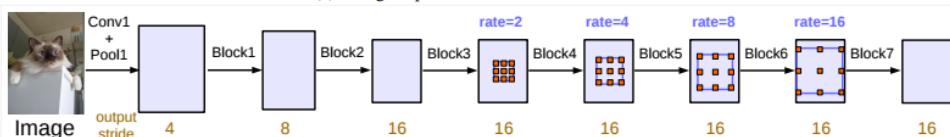
Figure 6. **BSD**: Qualitative results. Left to Right: Ours, SEEDS [7], SLIC [1]. Our approach better snaps to the image boundaries.

Semantic Segmentation

Rethinking Atrous Convolution for Semantic Image Segmentation (DeepLab-3)
Arxiv 2017

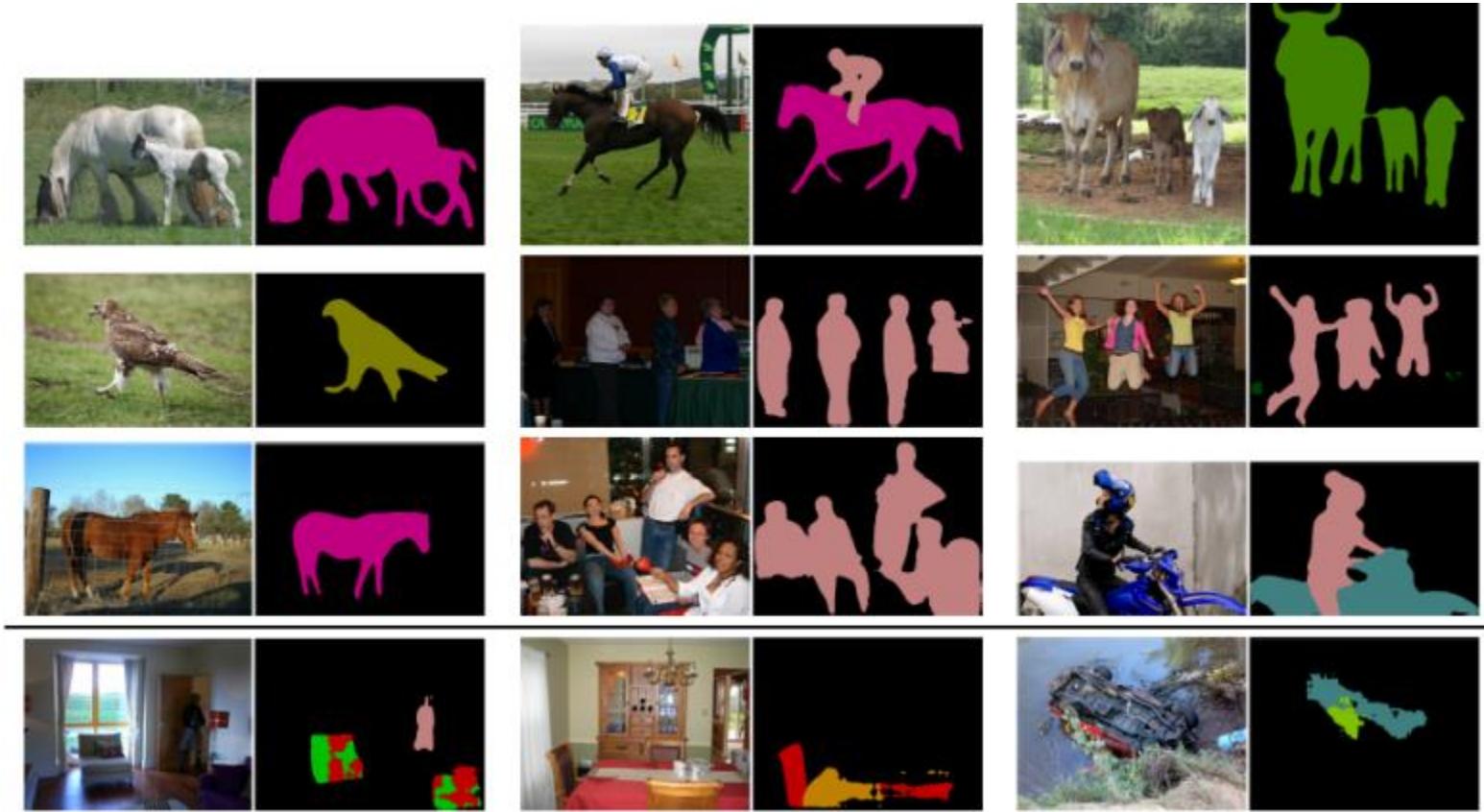


(a) Going deeper without atrous convolution.



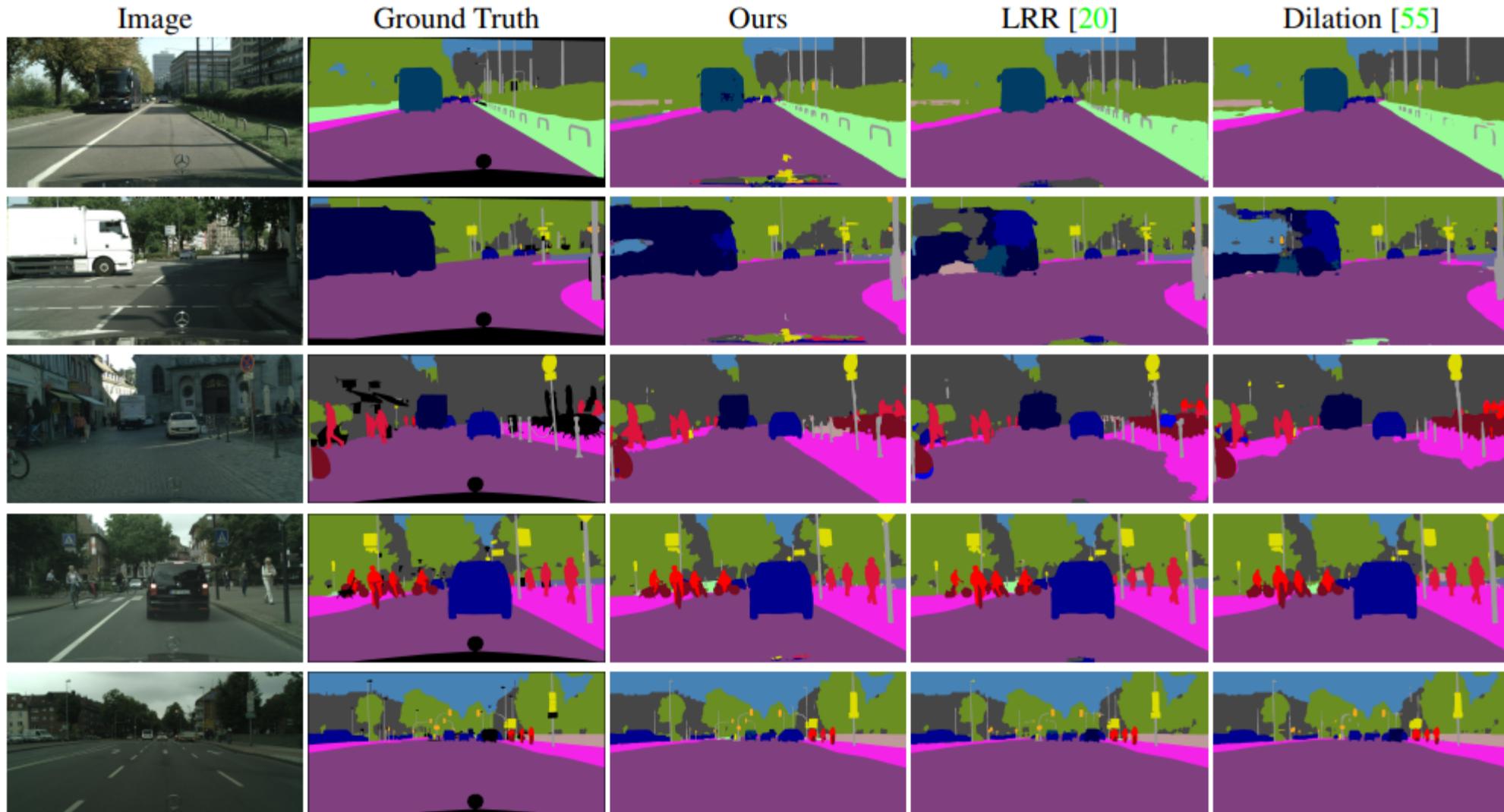
(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output_stride = 16$.

Figure 3. Cascaded modules without and with atrous convolution.



Semantic Segmentation

Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes **CVPR 2017**



Void	Road	Sidewalk	Building	Wall	Fence	Pole	Traffic Light	Traffic Sign	Vegetation
Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	Motorcycle	Bicycle

Instance Segmentation

Mask R-CNN ICCV 2017

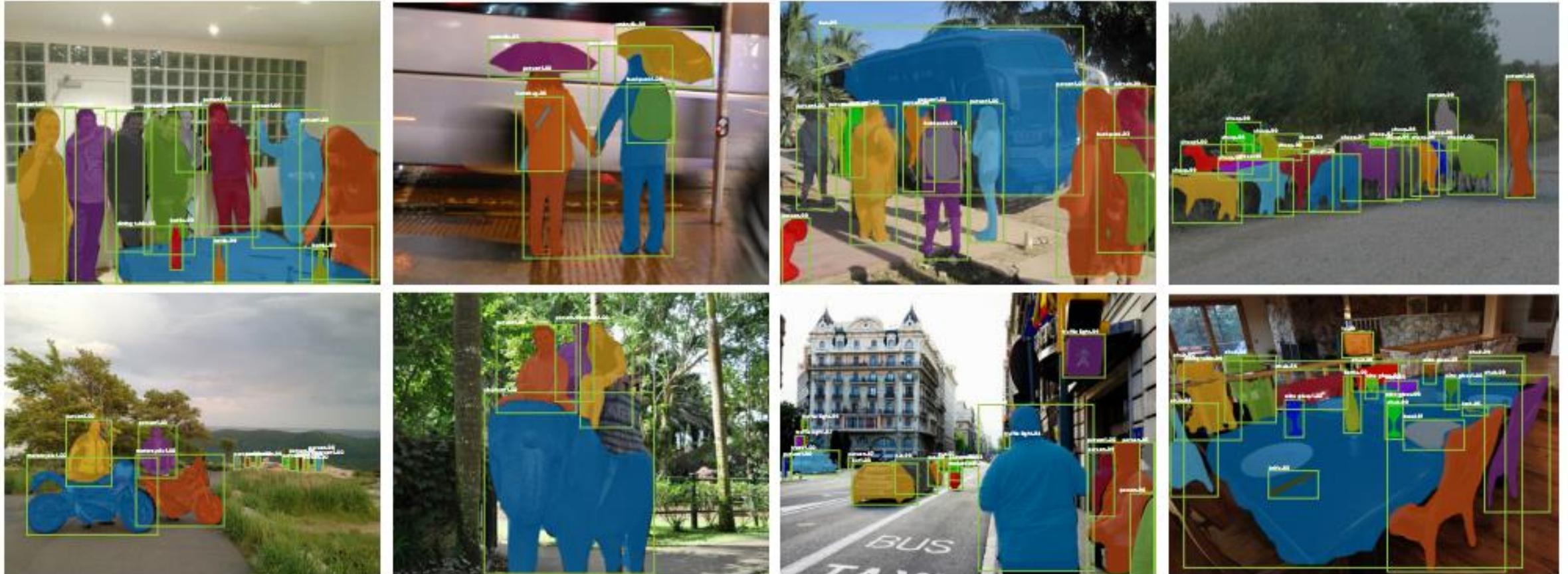
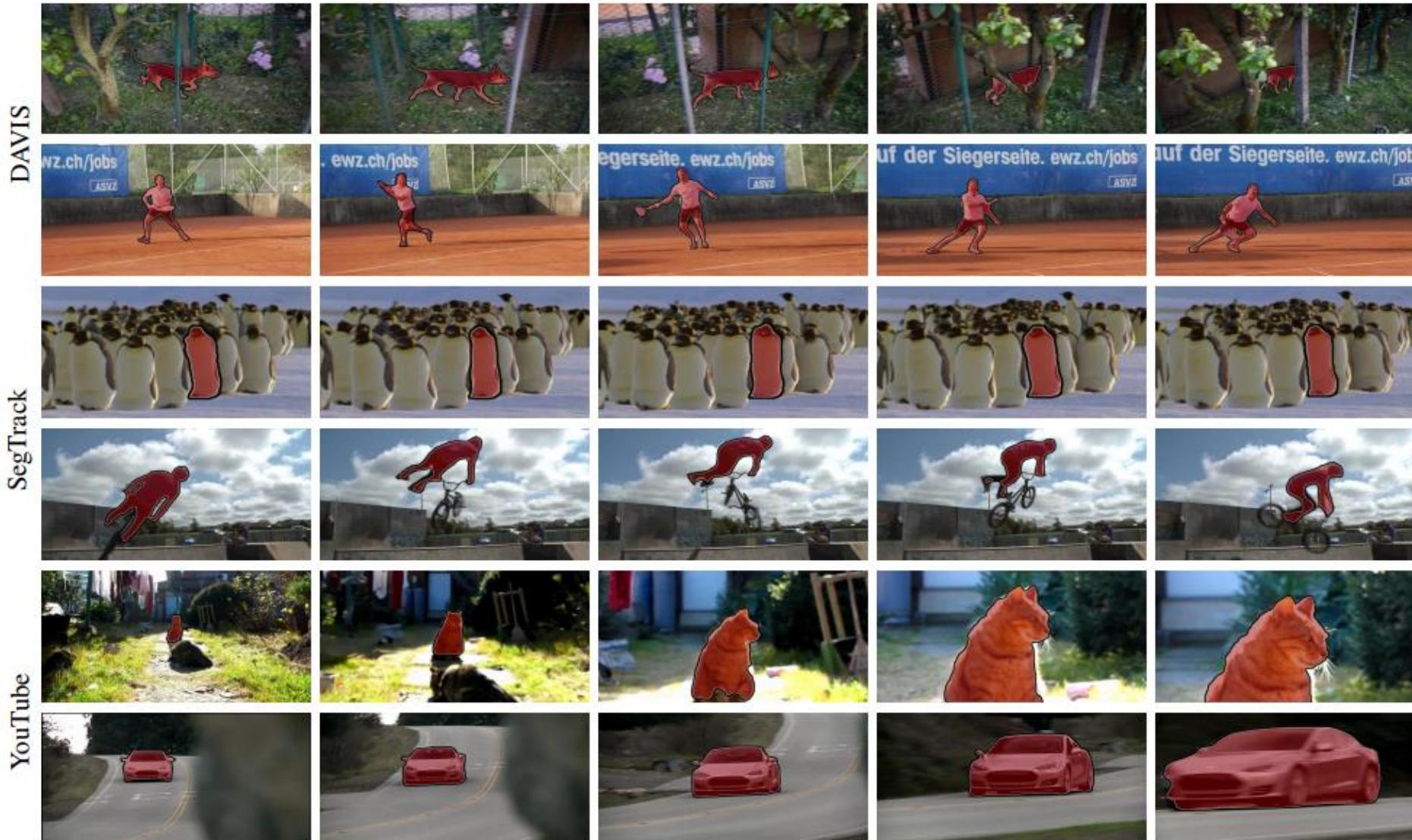


Figure 2. **Mask R-CNN** results on the COCO test set. These results are based on ResNet-101 [19], achieving a *mask AP* of 35.7 and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.

Video Segmentation

Learning Video Object Segmentation from Static Images **CVPR 2017**



1st frame, GT segment

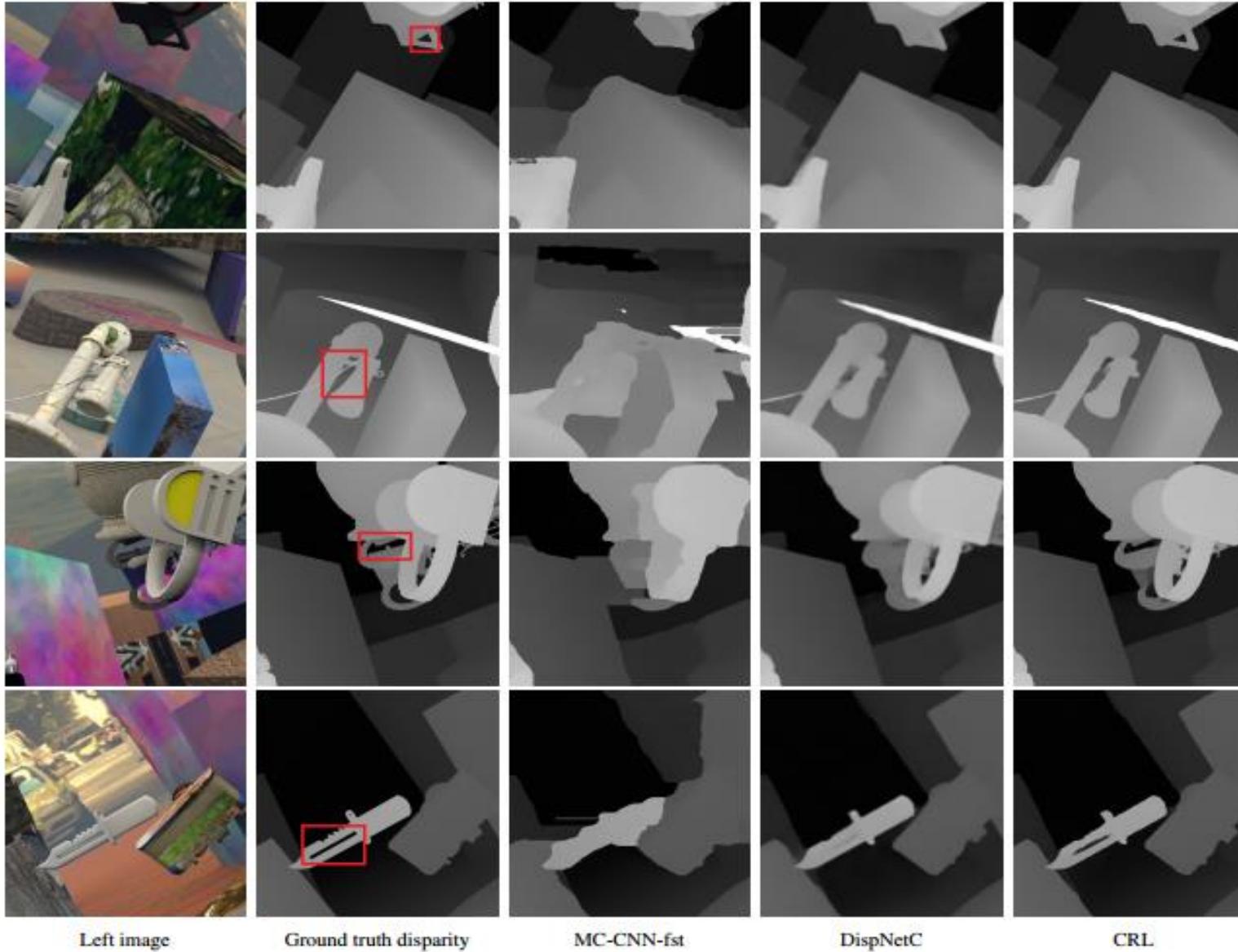
Results with MaskTrack, the frames are chosen equally distant based on the video sequence length

3D Vision

Stereo Matching, 2D to 3D
Intrinsic Image Decomposition, Visual SLAM

Stereo Matching

Cascade Residual Learning: A Two-stage Convolutional Neural Network for Stereo Matching [Arxiv 2017](#)



2D to 3D

KillingFusion: Non-rigid 3D Reconstruction without Correspondences **CVPR 2017**

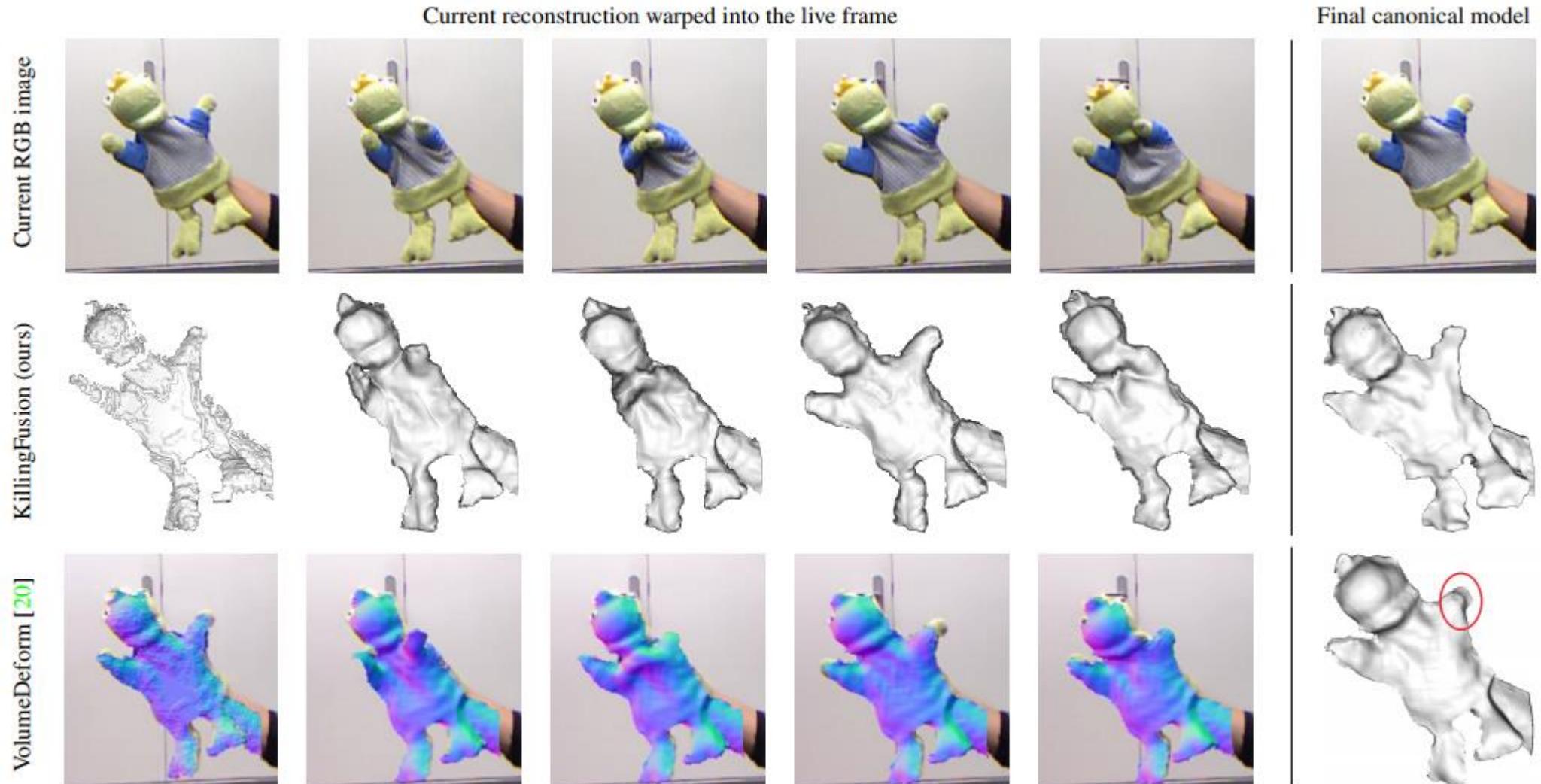
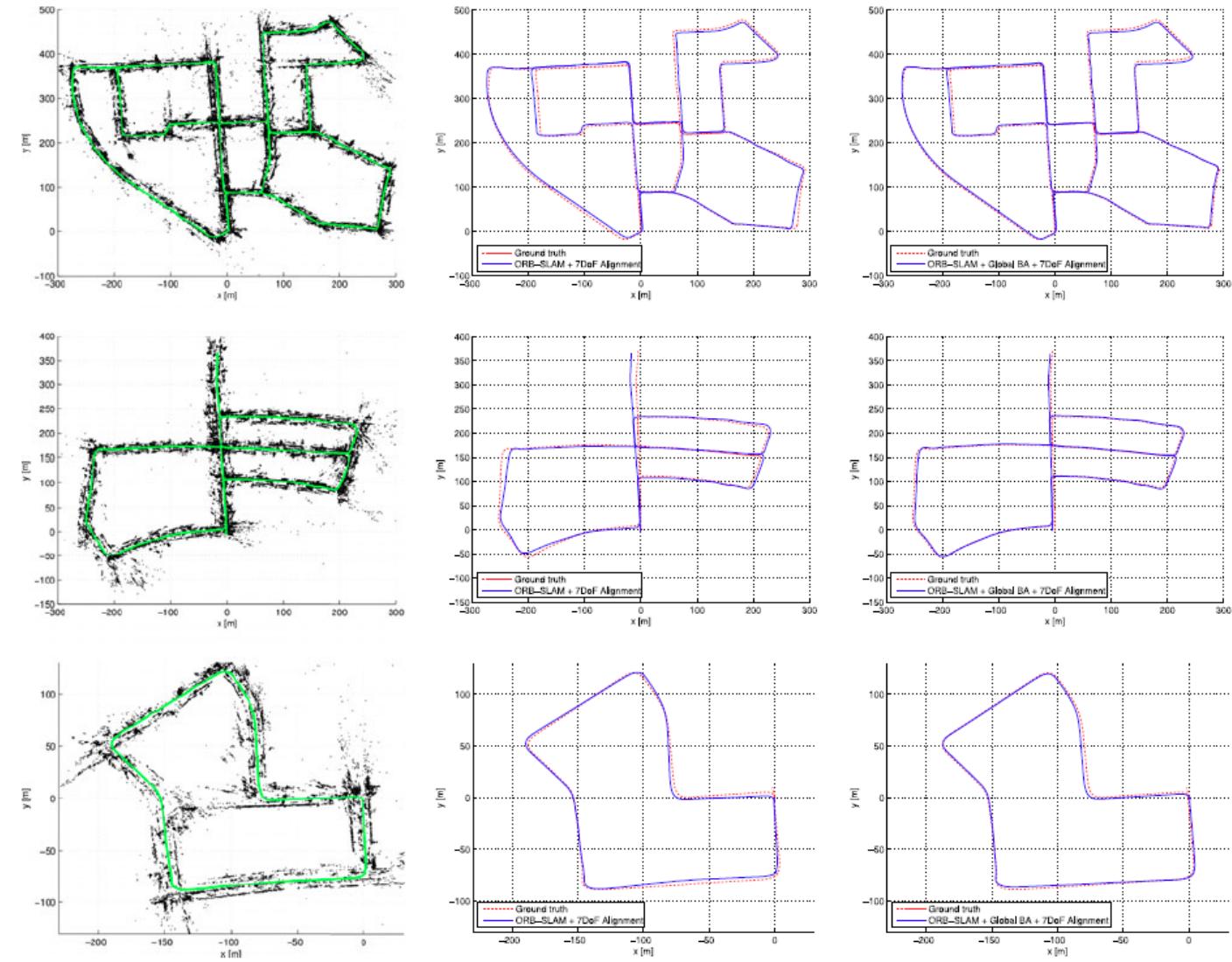


Figure 3. Comparison under **topological changes**. Our level-set-based KillingFusion fully evolves into the correct geometric shape between frames, while VolumeDeform [20] does so only partially (3rd and 5th live frames), which is reflected as artifacts in the final reconstruction.

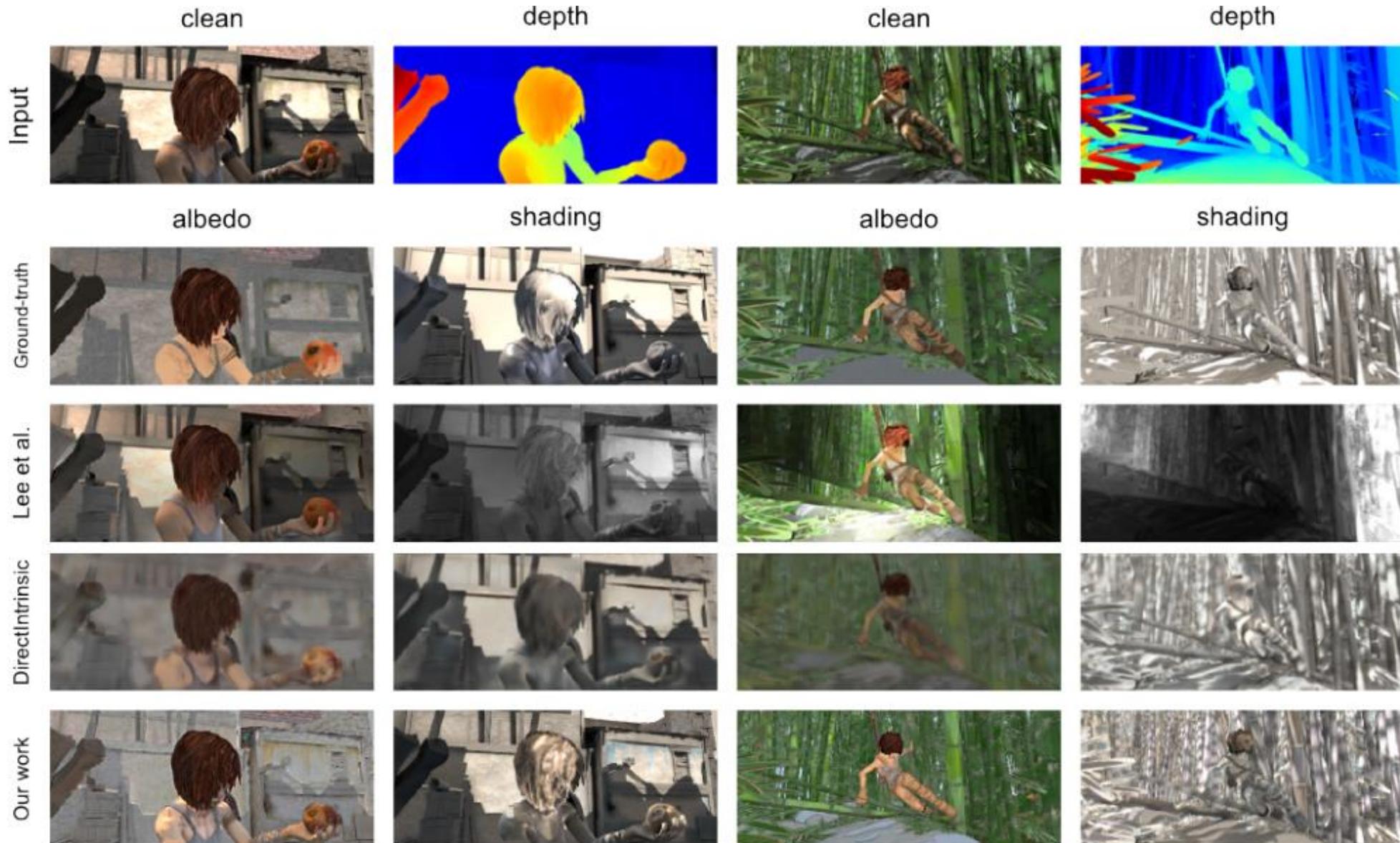
Visual SLAM

ORB-SLAM IROS 2015
NID-SLAM CVPR 2017



Intrinsic Image Decomposition

DARN: a Deep Adversarial Residual Network for Intrinsic Image Decomposition [Arxiv 2016](#)



Recognition/Detection/Retrieval

Object/Face Detection, Fine Grained Recognition

Pose Estimation, Keypoint/Landmark Detection

Retrieval

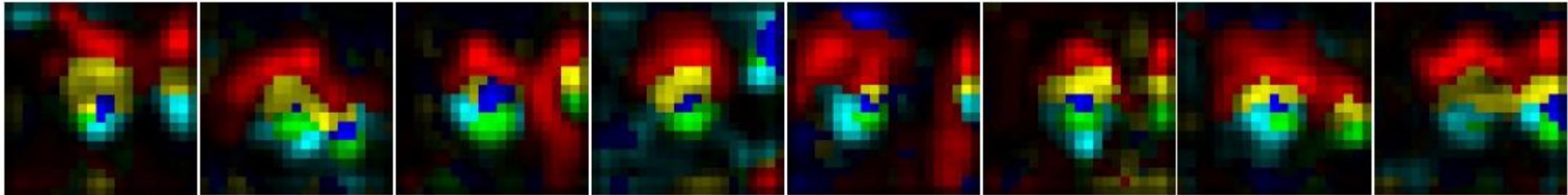
Object Detection

YOLO 9000: Better, Faster, Stronger **CVPR 2017**

<https://www.youtube.com/watch?v=VOC3huqHrss>

Face Detection

Faceness-Net: Face Detection through Deep Facial Part Responses [Arxiv 2017](#)



(a)

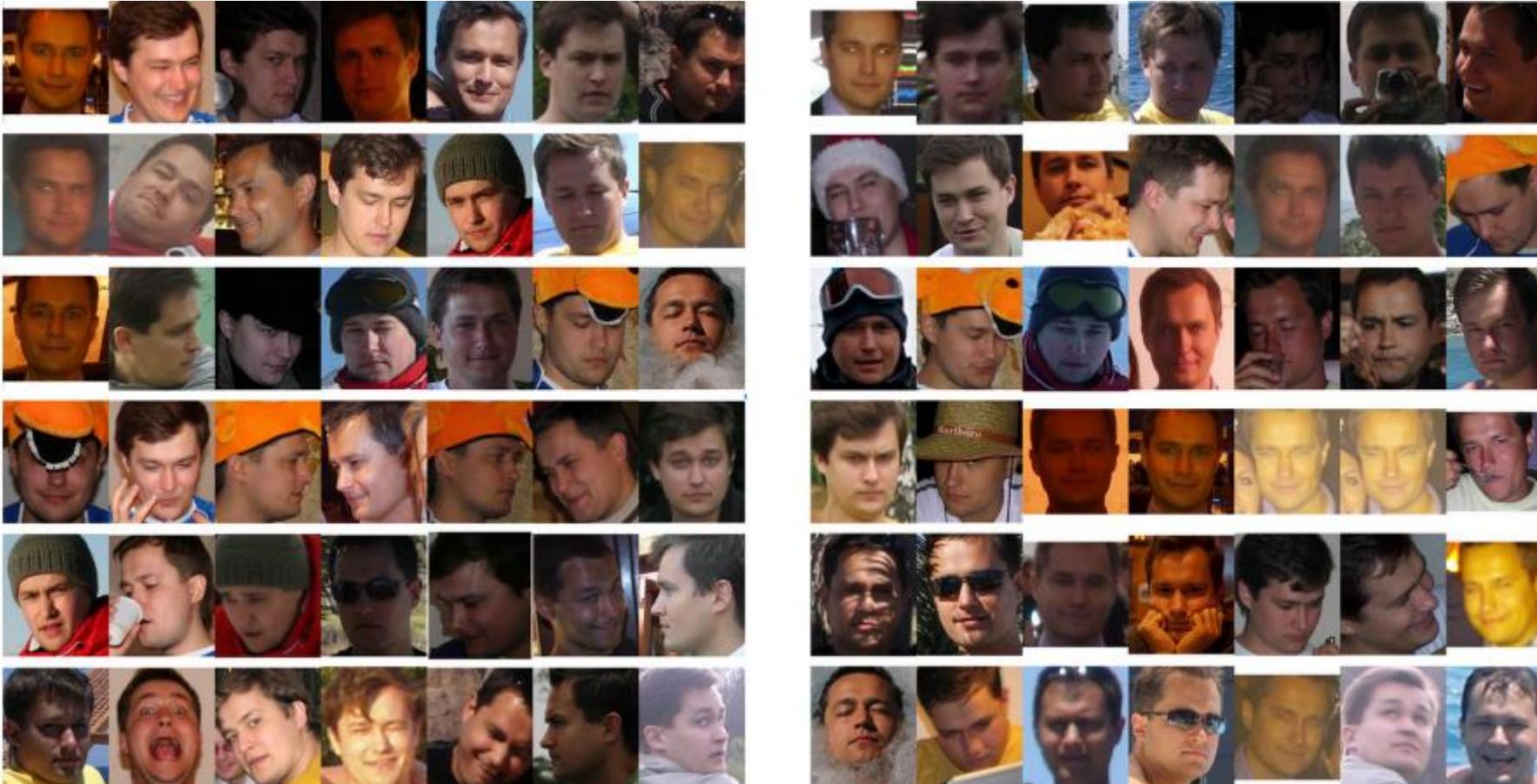


(b)

■ Hair ■ Eye ■ Nose ■ Mouth ■ Beard

Face Recognition

FaceNet: A Unified Embedding for Face Recognition and Clustering **CVPR 2015**



Fine-Grained Recognition

Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition
CVPR 2017

CUB-200-2011



Stanford Dogs



Stanford Cars



Pose Estimation

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields **CVPR 2017**

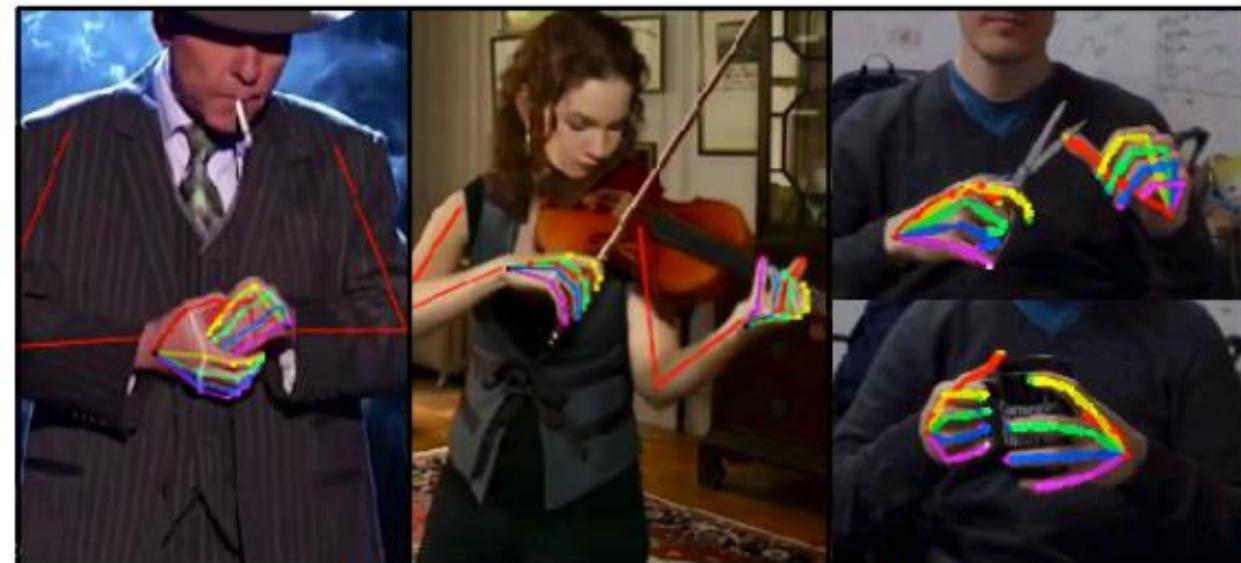
<https://www.youtube.com/watch?v=pW6nZXeWIGM>



Hand Keypoints Detection

Hand Keypoint Detection in Single Images using Multiview Bootstrapping **CVPR 2017**

<https://www.youtube.com/watch?v=q4xbmEQp3VE>



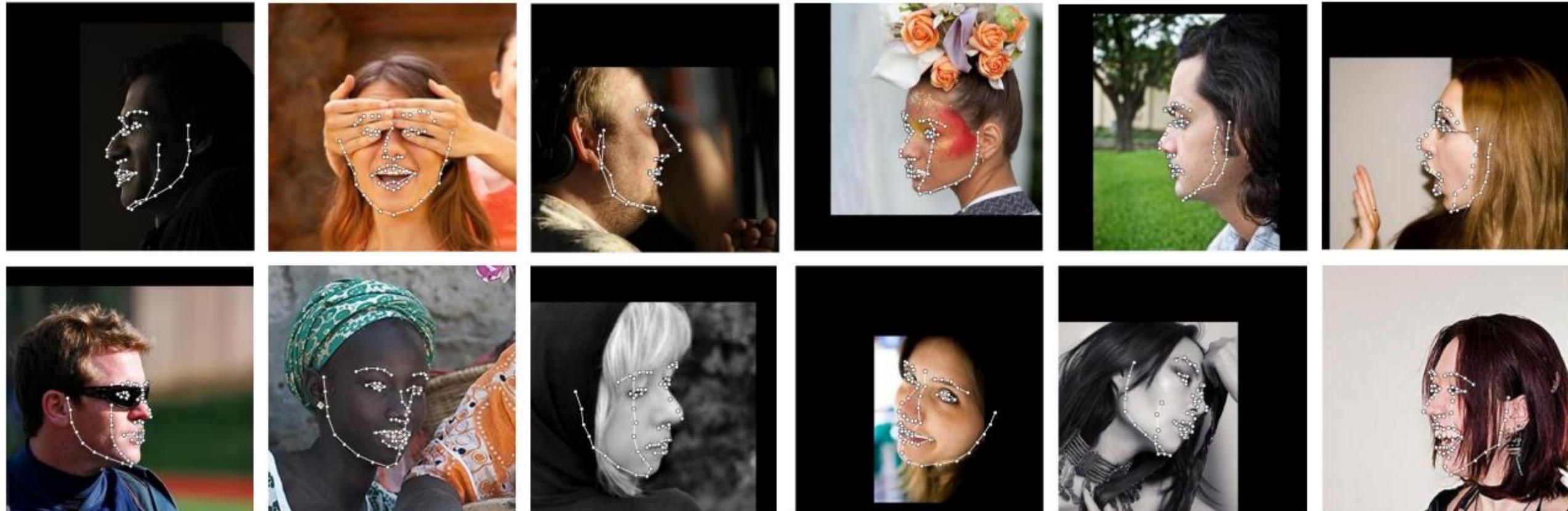
(a) Realtime 2D Hand Detection on YouTube and Webcam Videos



(b) 3D Hand Motion Capture by Triangulating Multiple 2D Detections

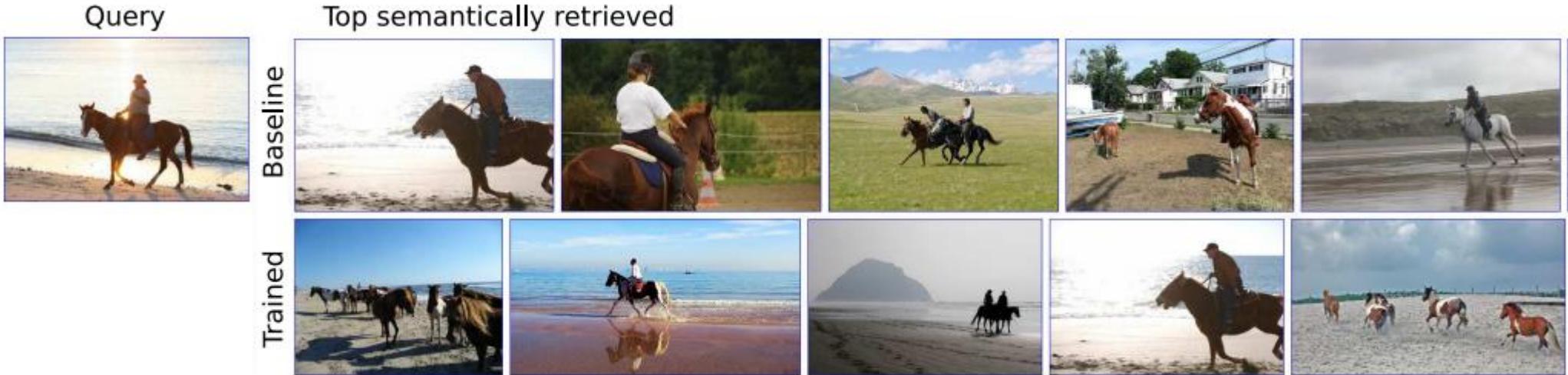
Facial Landmark Detection

Binarized Convolutional Landmark Localizers for Human Pose Estimation and Face Alignment with Limited Resources
ICCV 2017



Retrieval

Leveraging captions to learn a global visual representation for semantic retrieval **CVPR 2017**
Deep Sketch Hashing: Fast Free-hand Sketch-Based Image Retrieval **CVPR 2017**



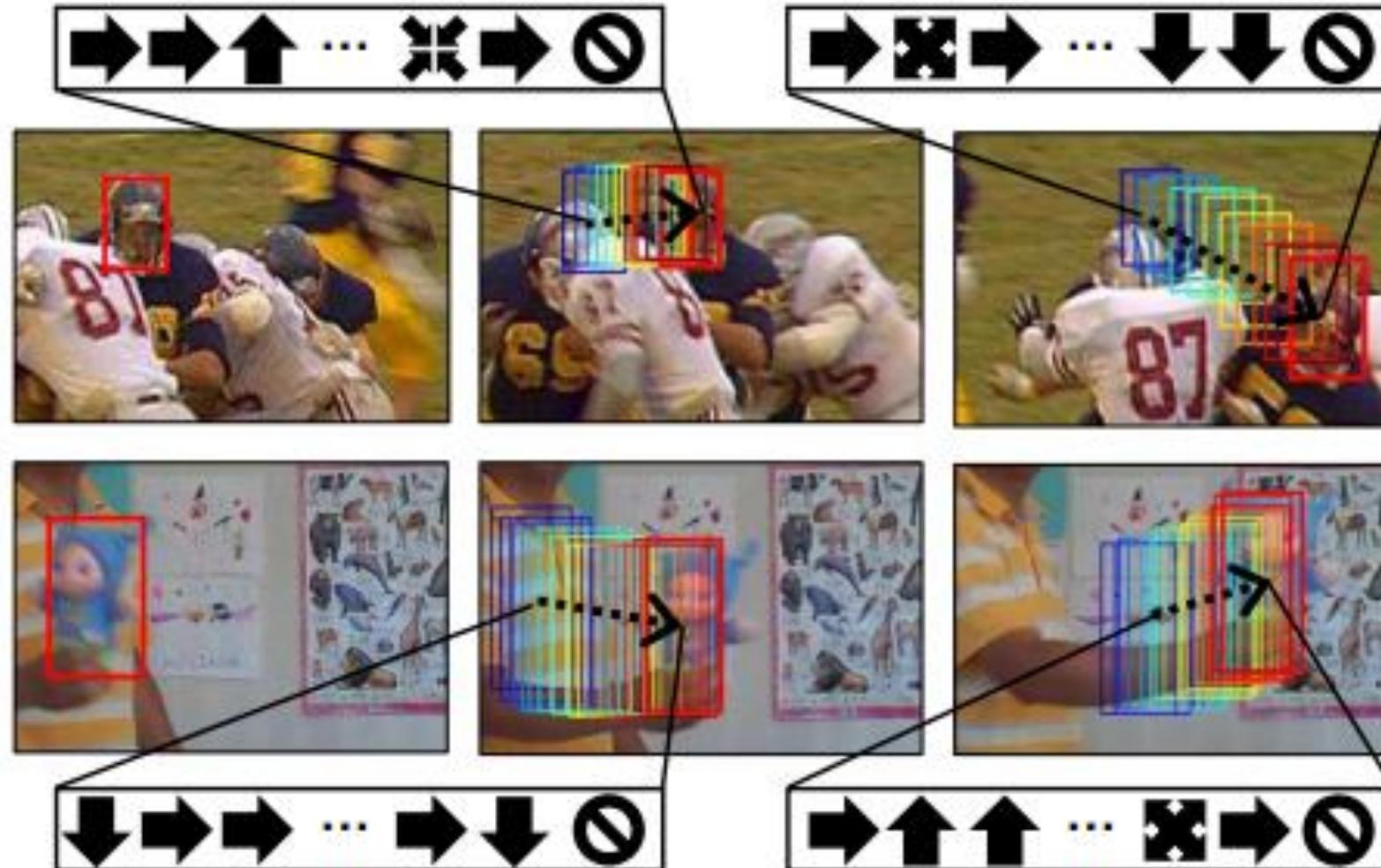
Free-hand sketch	Natural image (top)/Sketch-token(bottom)			Free-hand sketch	Natural image (top)/Sketch-token(bottom)		
	 	 	 		 	 	 
	 	 	 		 	 	 
	 	 	 		 	 	 

Tracking/Motion/Action

Tracking, Optical Flow, Prediction
Gesture Recognition, Action Recognition

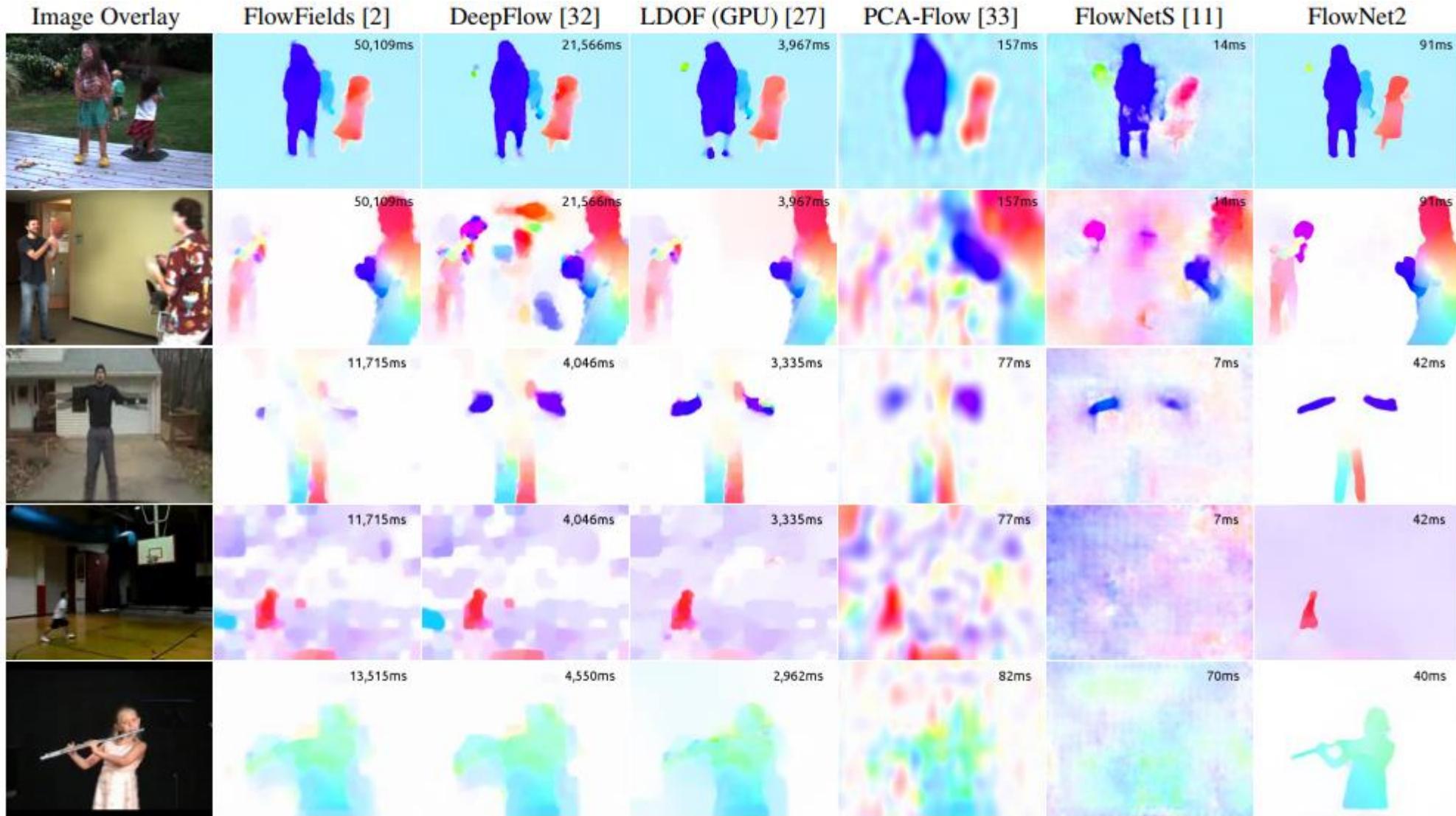
Tracking

Action–Decision Networks for Visual Tracking With Deep Reinforcement Learning **CVPR 2017**



Optical Flow

FlowNet 2.0 CVPR 2017



Optical Flow

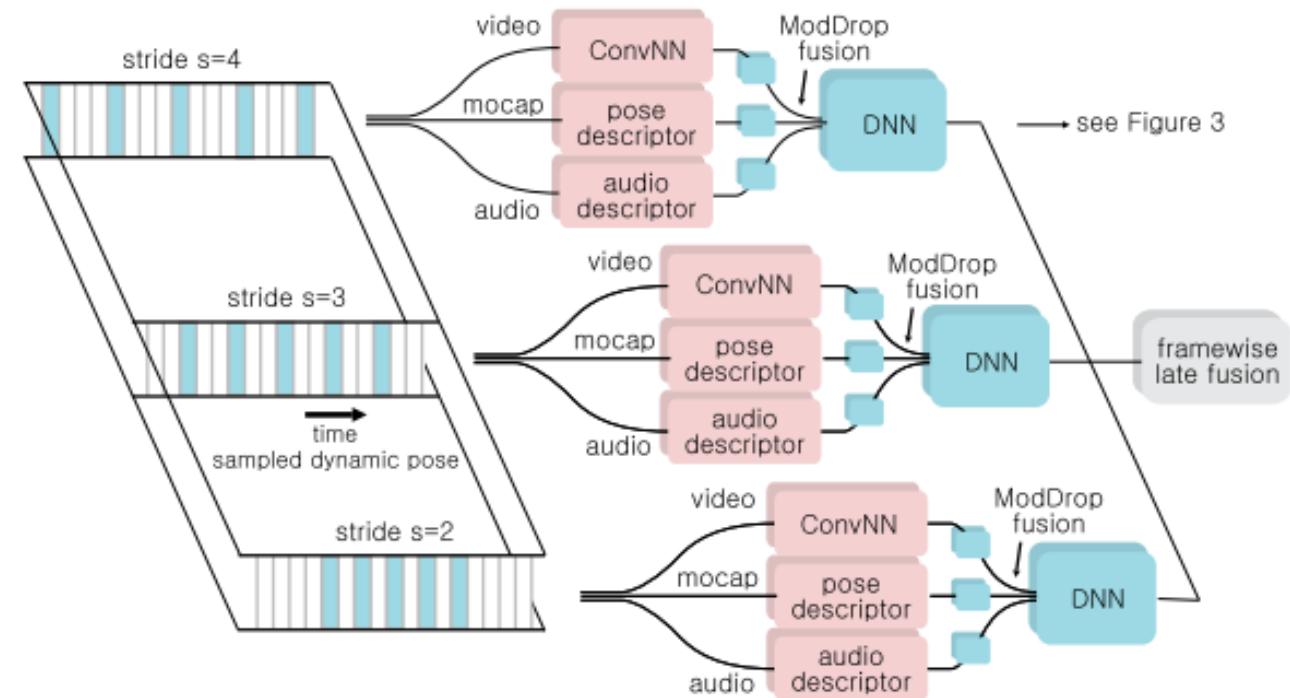
MirrorFlow: Exploiting Symmetries in Joint Optical Flow and Occlusion Estimation **ICCV 2017**



Figure 1. Results of our symmetric optical flow approach given two consecutive images from the KITTI benchmark [12]. Our method jointly predicts forward and backward optical flow (*overlaid on top row, from left to right, color coding of the benchmark*) as well as corresponding occlusion maps for each view (*overlaid on bottom row*).

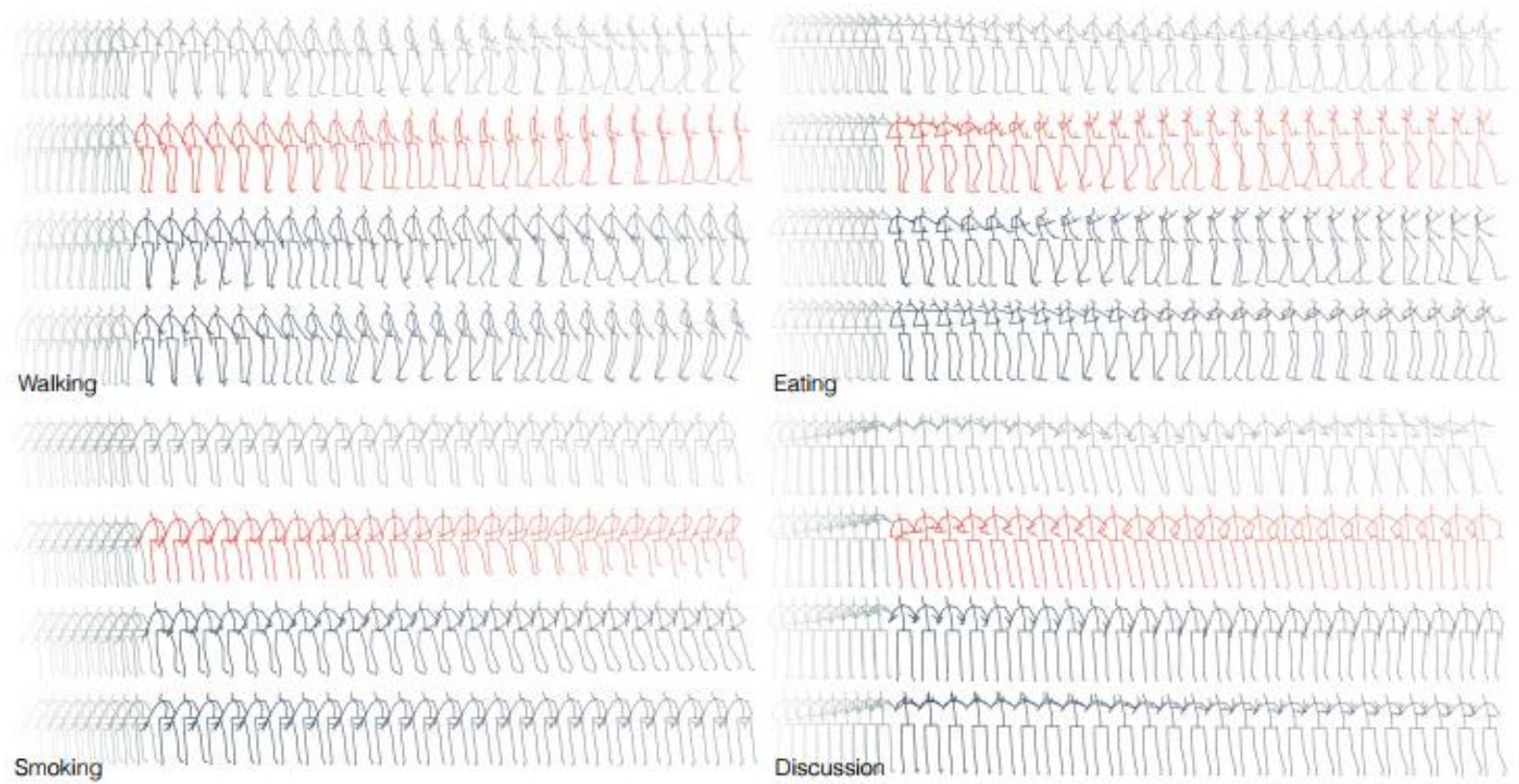
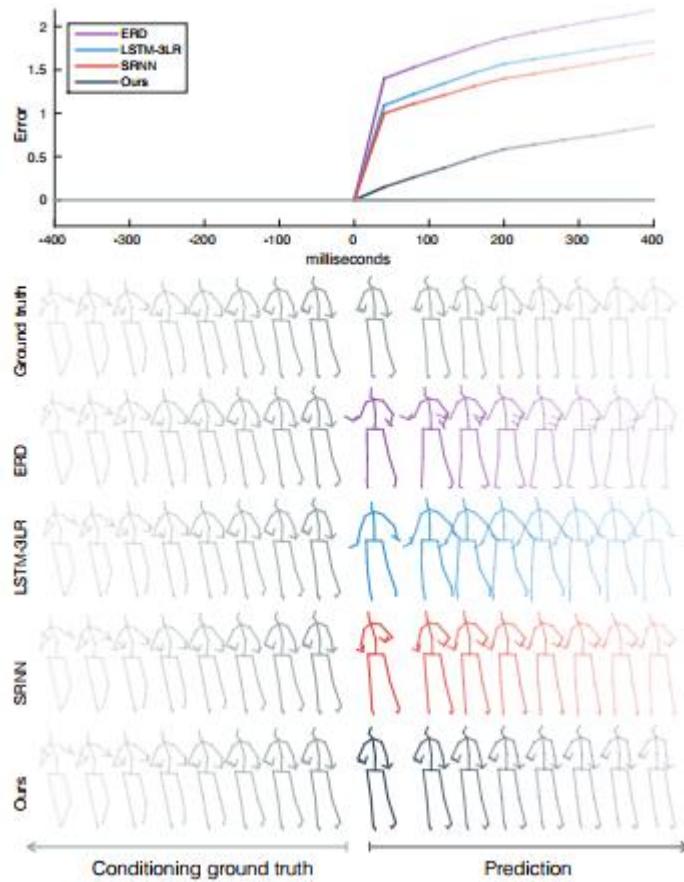
Gesture Recognition

ModDrop: Adaptive Multi-Modal Gesture Recognition **PAMI 2016**



Action Recognition

On Human Motion Prediction using RNNs Arxiv 2017



Understanding

Semantic Matching, Future Prediction, Video Modeling

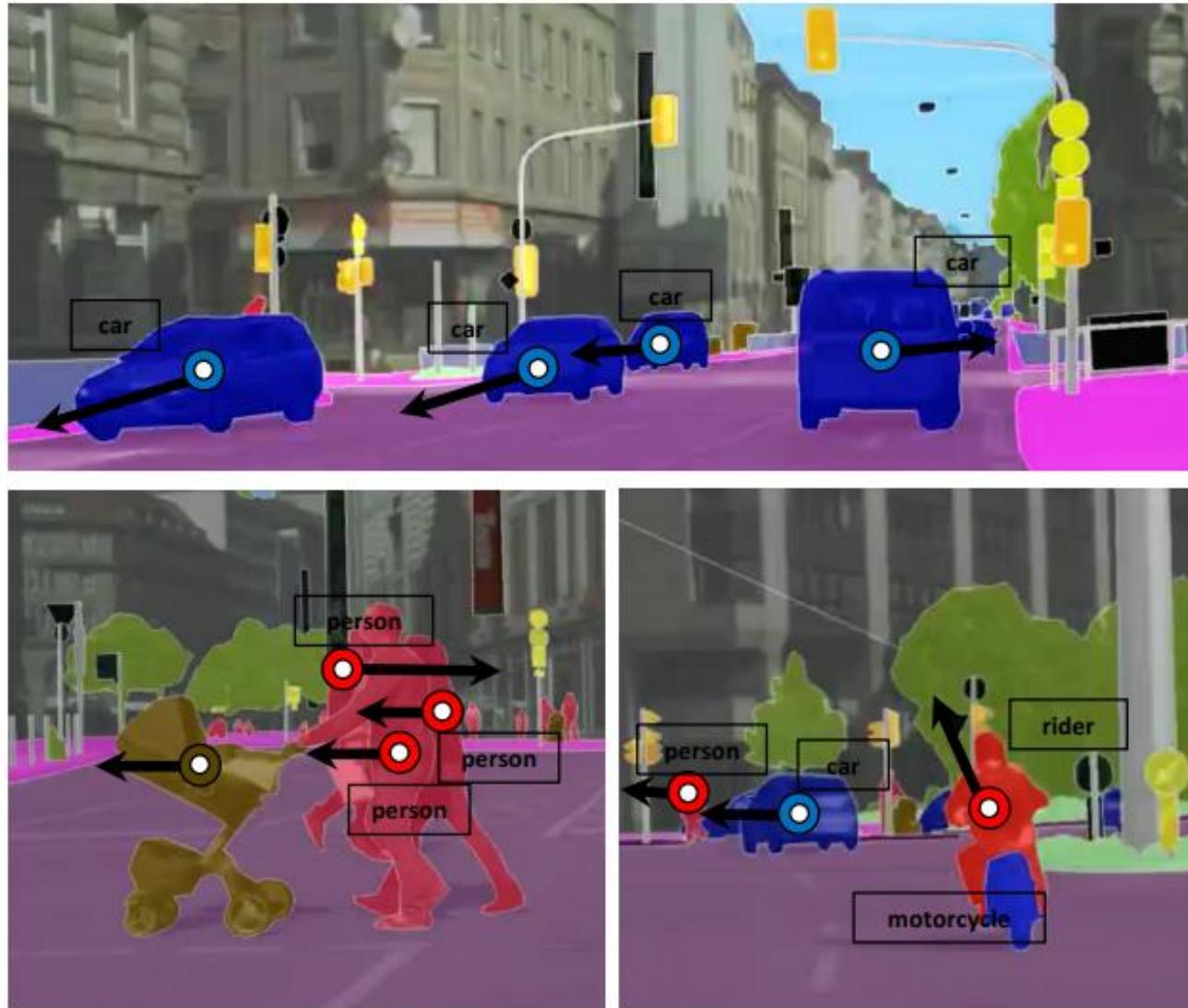
Semantic Flow

Proposal Flow: Semantic Correspondences from Object Proposals **PAMI 2017**



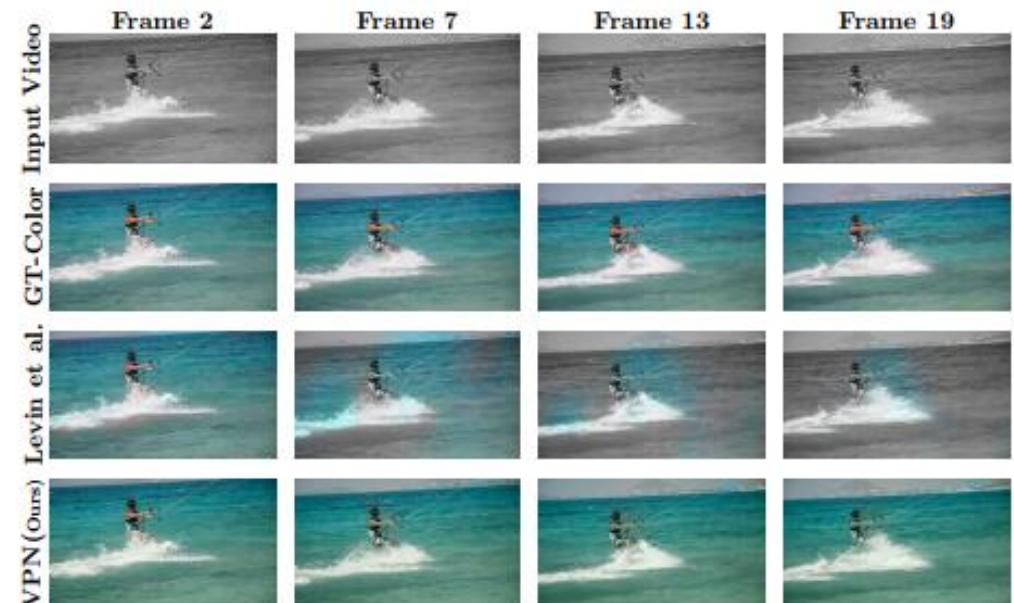
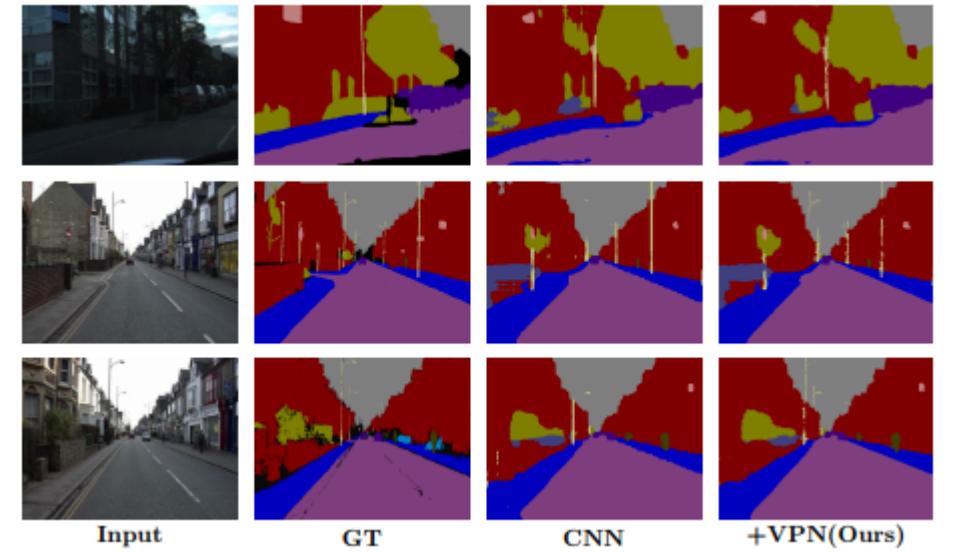
Future Prediction

Predicting Deeper into the Future of Semantic Segmentation **ICCV 2017**



Video Modeling

Video Propagation Networks CVPR 2017

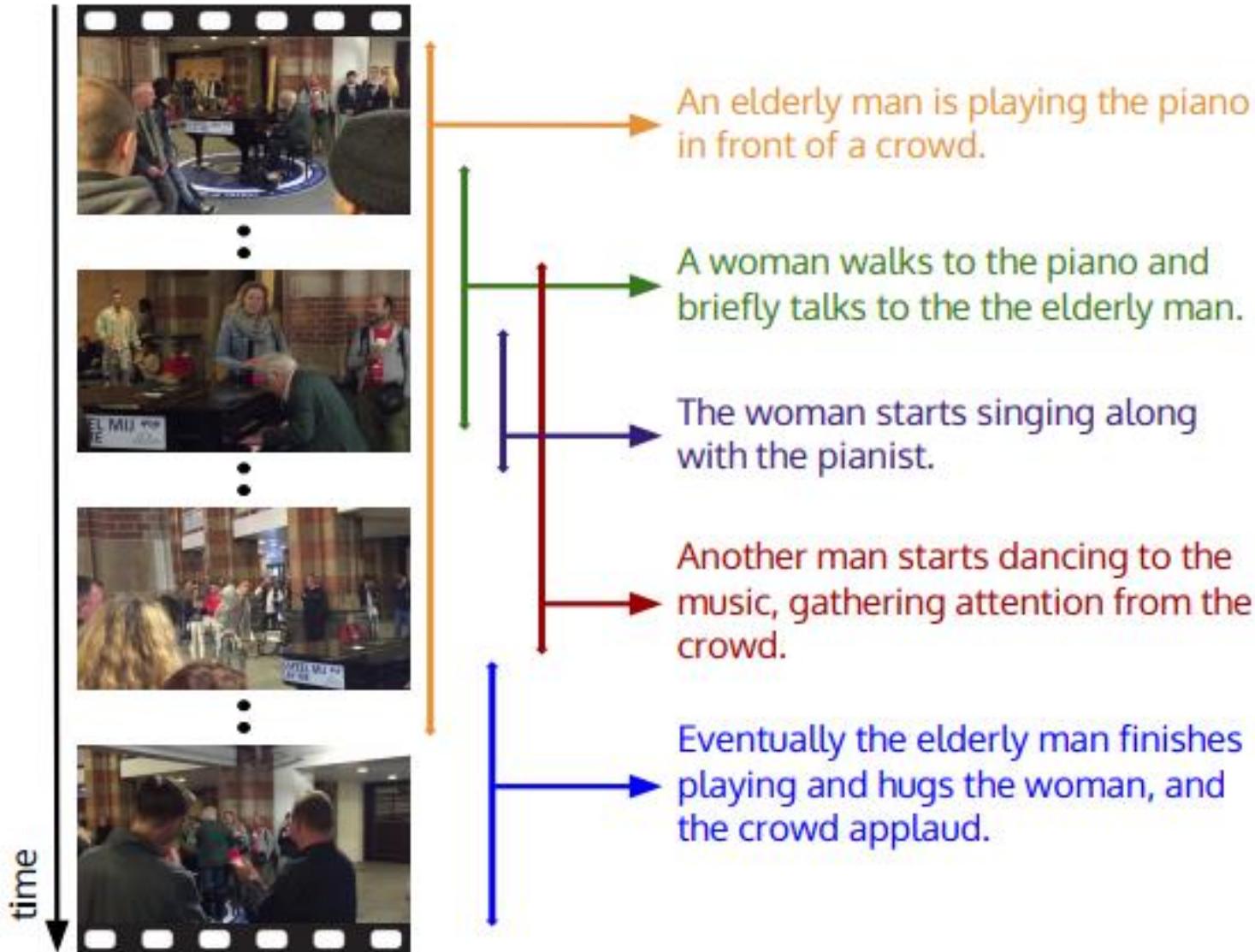


Vision + Language

Captioning, Q&A, Reasoning, Text to Image

Captioning

Dense-Captioning Events in Videos ICCV 2017



Q&A / Reasoning

Inferring and Executing Programs for Visual Reasoning **ICCV 2017**



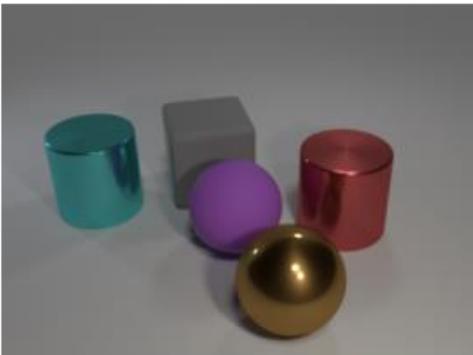
How many chairs are at the table?



Is there a pedestrian in my lane?



Is the person with the blue hat touching the bike in the back?



Is there a matte cube that has the same size as the red metal object?

Q: What shape is the...

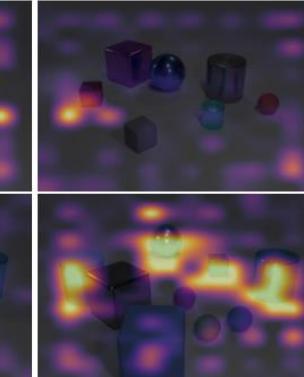
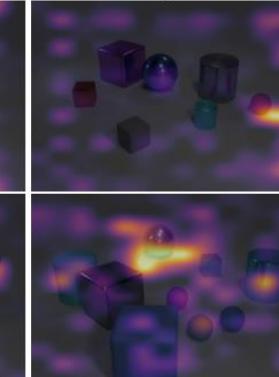
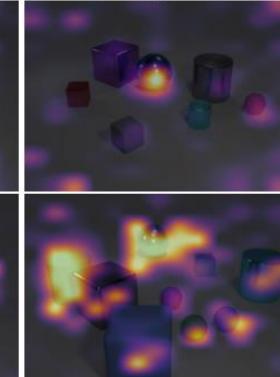
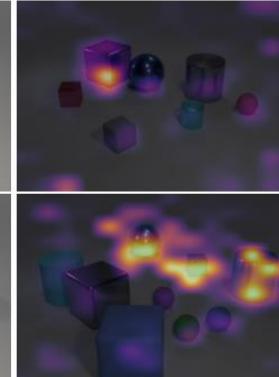
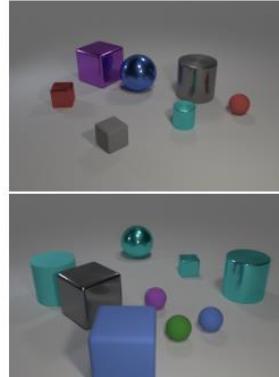
...purple thing?

...blue thing?

...red thing right of the blue thing?

...red thing left of the blue thing?

A: cube



Q: How many cyan things are...

...right of the gray cube? ...left of the small cube? ...right of the gray cube and left of the small cube? ...right of the gray cube or left of the small cube?

A: 3

A: 2

A: 1

A: 4

Text to Image

StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks [Arxiv 2017](#)

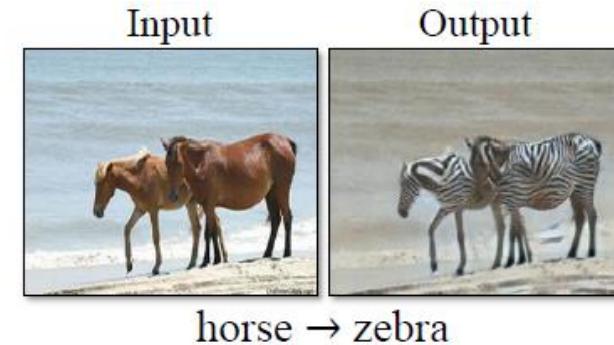
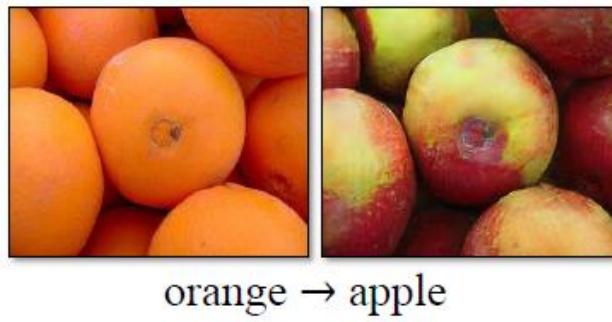
Text description	A couple of men riding horses on top of a green field	A train coming to a stop on the tracks out side	A big airplane flying in the big blue sky	A group of boats on a body of water	Two public transit buses parted in a lot	The white kitchen features very contemporary cabinet arrangements	The man is standing in the water holding his surfboard
Stage-II images							
Text description	A big building with a parking lot in front of it	There is a lot of electrical sitting on the table	A couple of computer screens sitting on a desk	Three zeebras standing in a grassy field walking	A herd of cows standing on a grass covered field	A group of people standing around and posing for a picture	People who are dressed for skiing standing in the snow
Stage-II images							

Graphics

Image Translation, Style Transfer, Attribute Transfer, Matting

Image Translation

CycleGAN ICCV 2017
DiscoGAN ICML 2017



(b) Handbag images (input) & Generated shoe images (output)



(c) Shoe images (input) & Generated handbag images (output)

Style Transfer

Deep Photo-Style Transfer **CVPR 2017**

Real-Time Neural Style Transfer for Videos **CVPR 2017**



(a) Input image

(b) Reference style image

(c) Reinhard et al. [12]

(d) Pitié et al. [11]

(e) Our result

Attribute Transfer

Neural Color Transfer Between Images Arxiv 2017



Source

Reference

Reinhard et al. [2001]

Pitié et al. [2005]

Ours



A portrait photograph of a man with dark hair and a mustache, wearing a black shirt. He is looking directly at the camera with a neutral expression.



A close-up photograph of a woman's eyes. She has light-colored eyes and is wearing a subtle, smoky purple eyeshadow. Her gaze is directed slightly to the right of the camera.

Source

Reference

Shih et al. [2014]

Ours

Source

Reference

Tong et al. [2007]

Ours

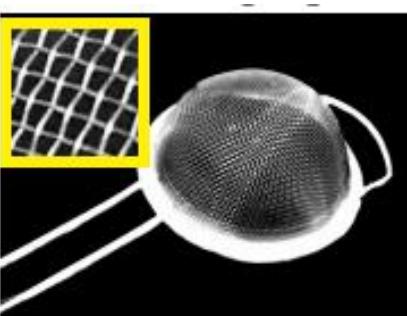
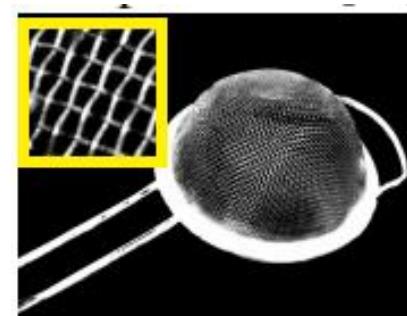
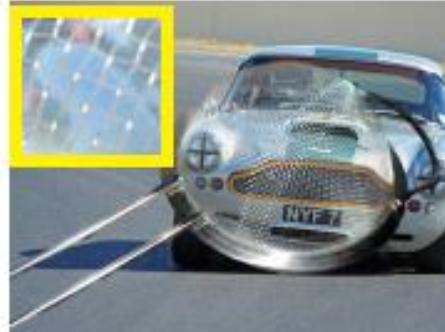
Matting

Deep Image Matting CVPR 2017



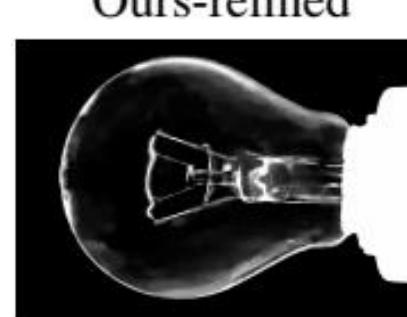
Ours-refined

GT



Ours-refined

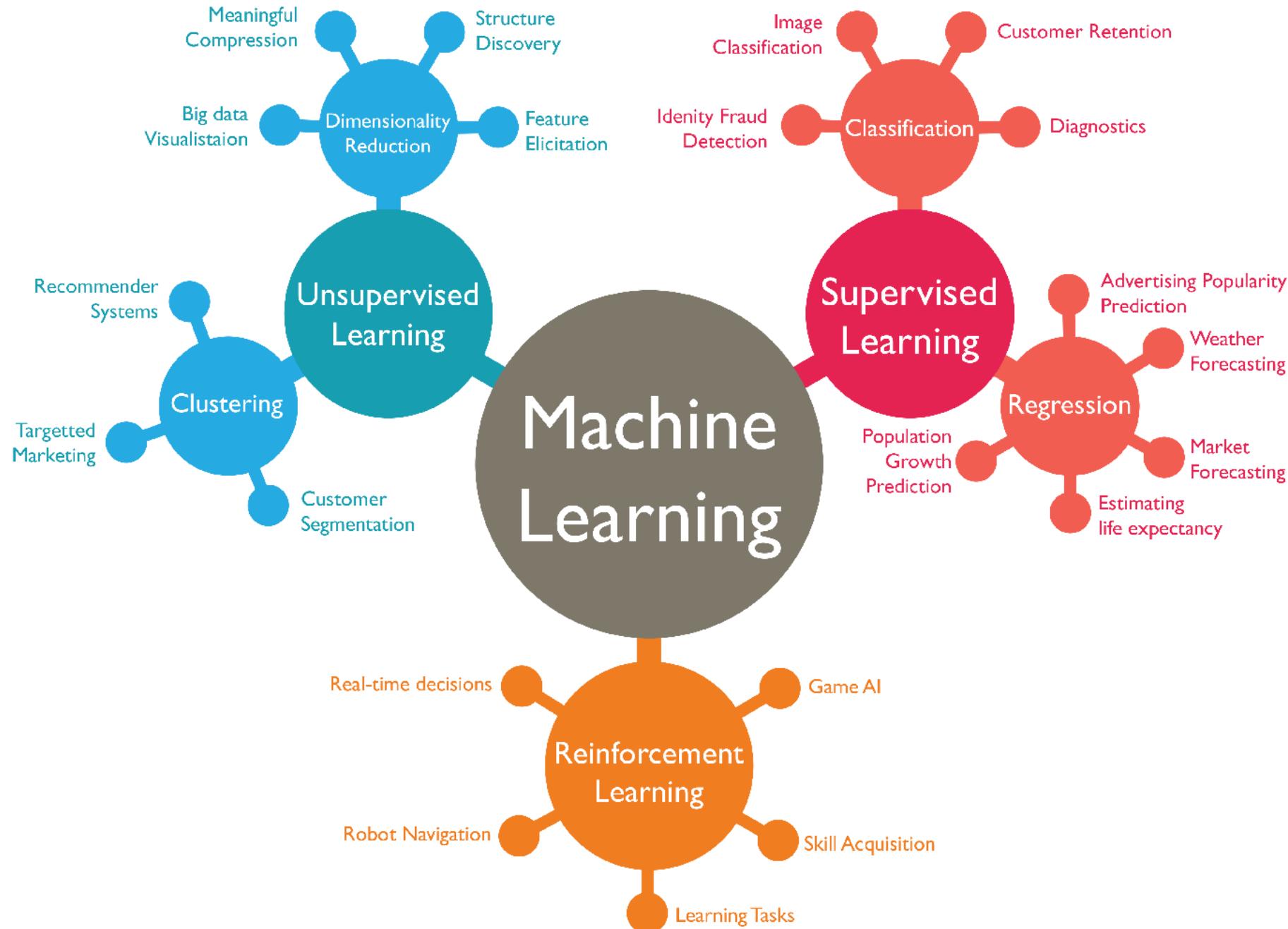
GT



Ours-refined

GT

Machine Learning



Supervised Learning

Regression

- From Training Data, Predict Real-Value
- Linear Regression / ARMA / SVR / GP

Classification

- From Training Data, Predict Discrete-Value
- Naïve Bayes / Logistic Regression
- SVM / Random Forests / Neural Networks

Generative Model = Probabilistic Model

- Learn $P(x, y) \rightarrow$ Learn $P(y | x)$
- Naïve Bayes, VAE, GAN

Discriminative Model = Function Fitting

- Learn $P(y | x)$ Directly
- Logistic Regression, Neural Networks

Discriminant Function

- SVM

Unsupervised Learning

Dimension Reduction

- High Dimension Data → Low Dimension Data
- Linear: PCA / LDA
- Manifold: ISOMAP / LLE / t-SNE

Clustering

- Cluster Un-labeled Data
- K-means / GMM / HMM

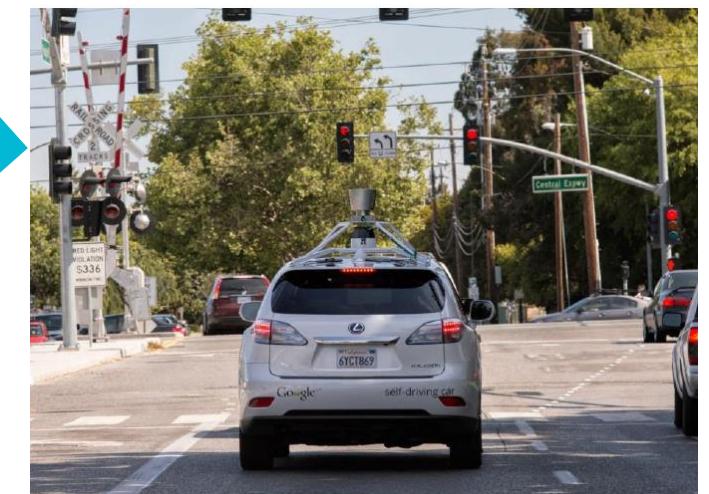
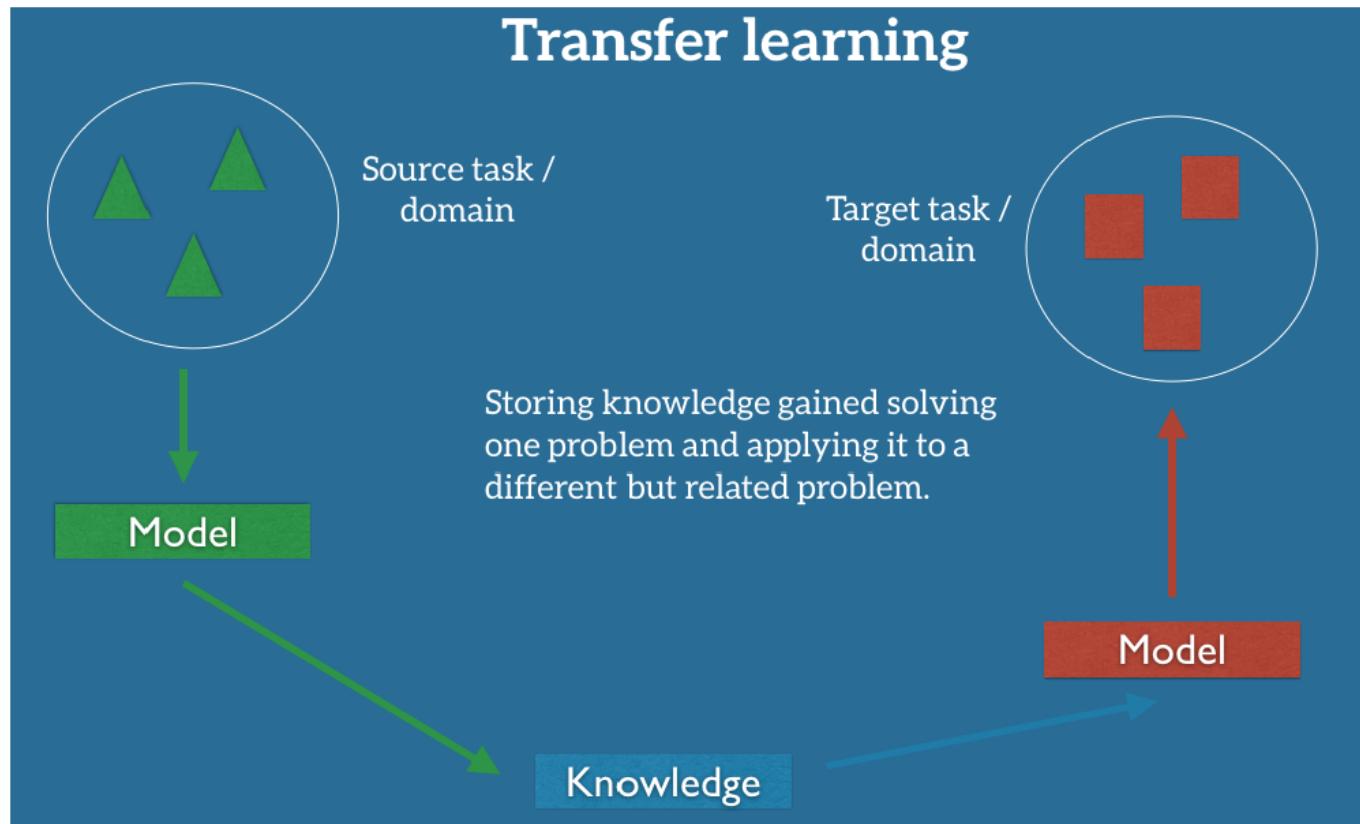
Unsupervised Learning

Feature Learning

- Learning Features from Training Data
- Also Called Pre-training
- Autoencoder / RBM
- VAE / GAN / CNN
- Semi-supervised Learning
- Weakly-supervised Learning

Transfer Learning

- Train on A, Prediction on B



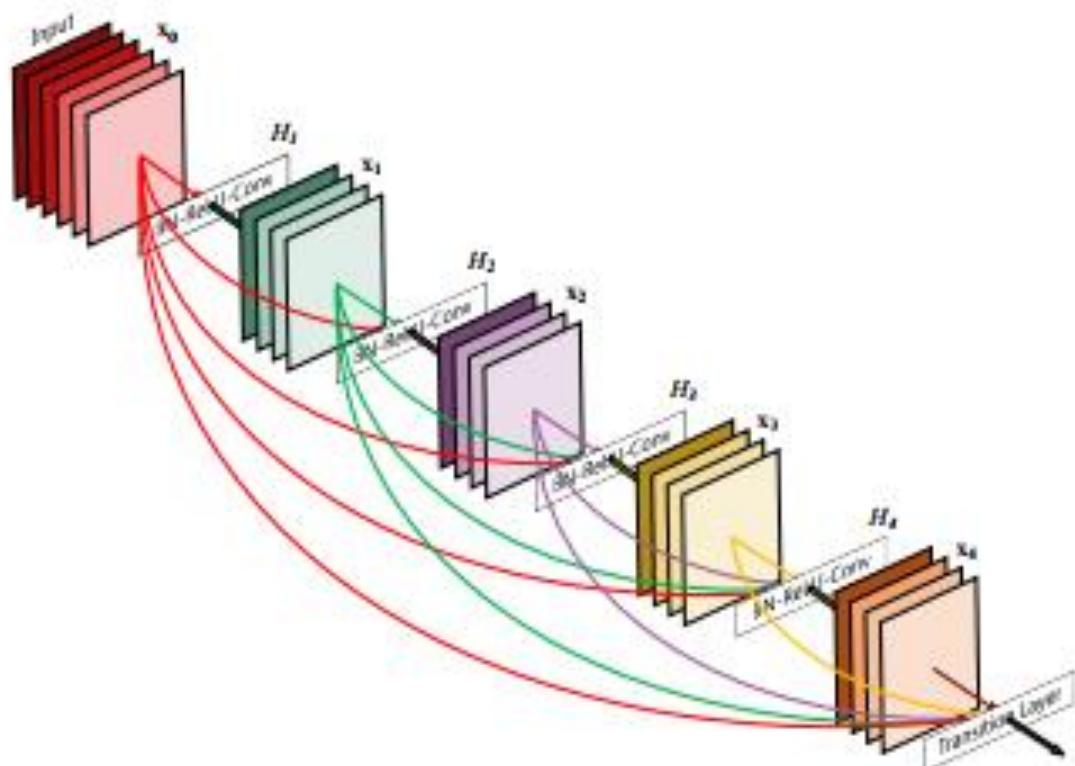
Reinforcement Learning

Learning from Actions & Rewards

- Dynamic Programming
- Deep Q Learning
- Policy Gradients
- Asynchronous Advantage Actor-Critic

CNN

ResNet CVPR 2016 DenseNet CVPR 2017

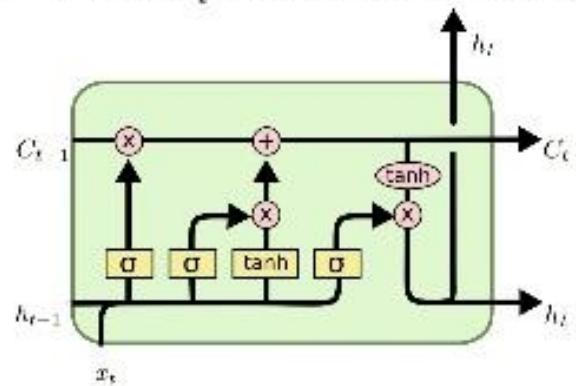


RNN

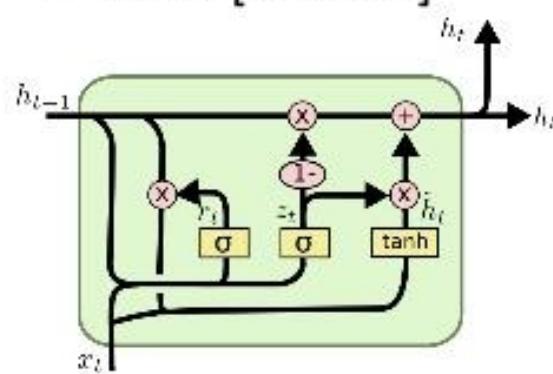
LSTM NeuralComputation 1997
GRU Arxiv 2014

LSTM and GRU

- LSTM [Hochreiter&Schmidhuber97]



- GRU [Cho+14]



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

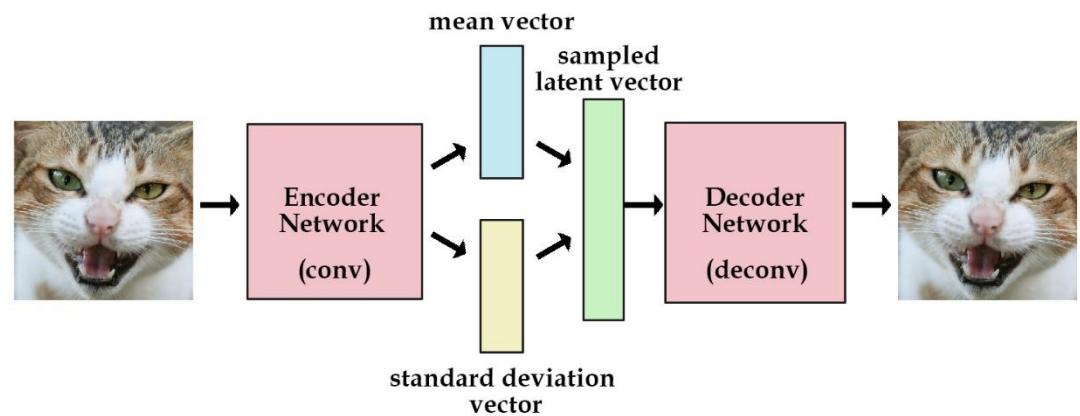
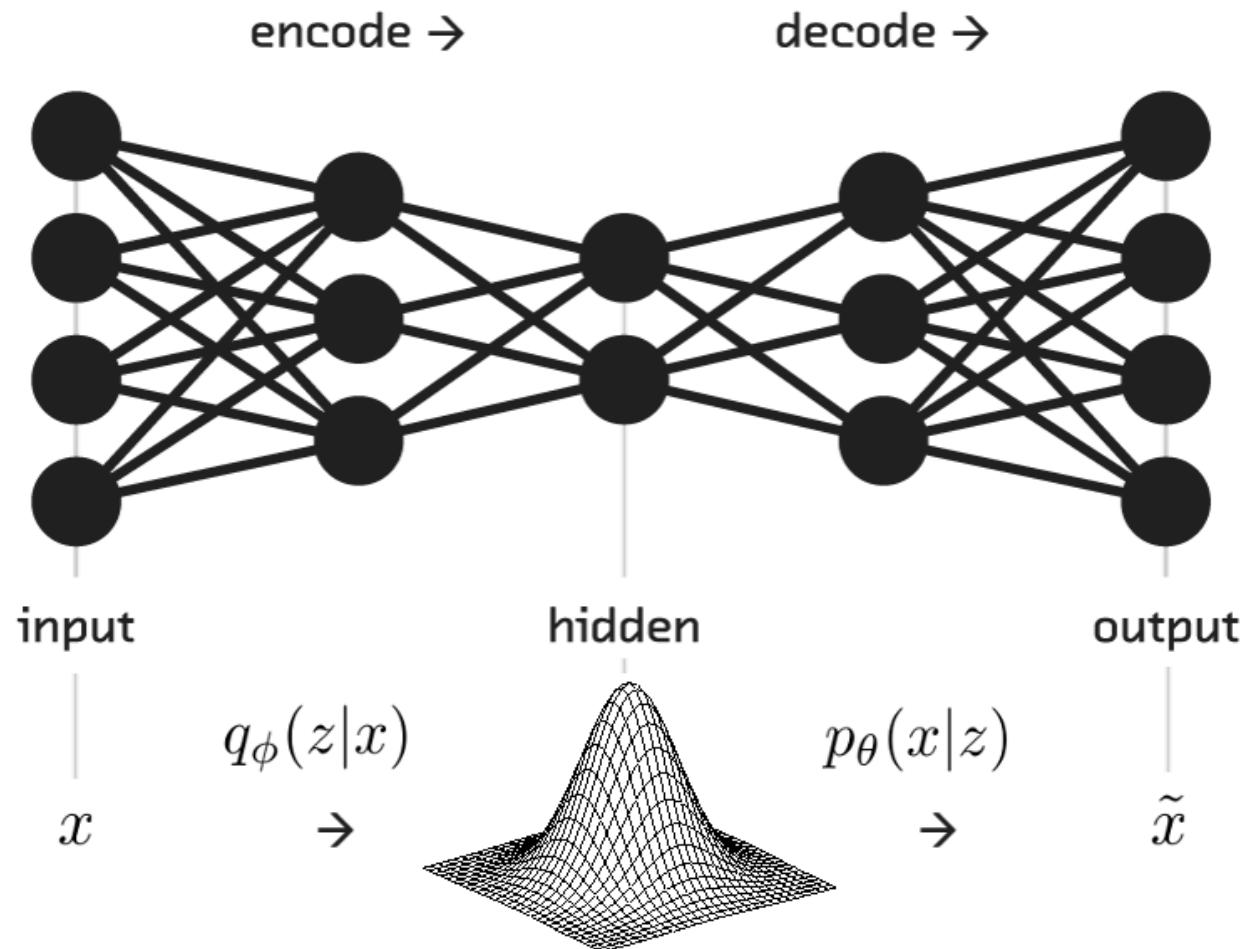
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

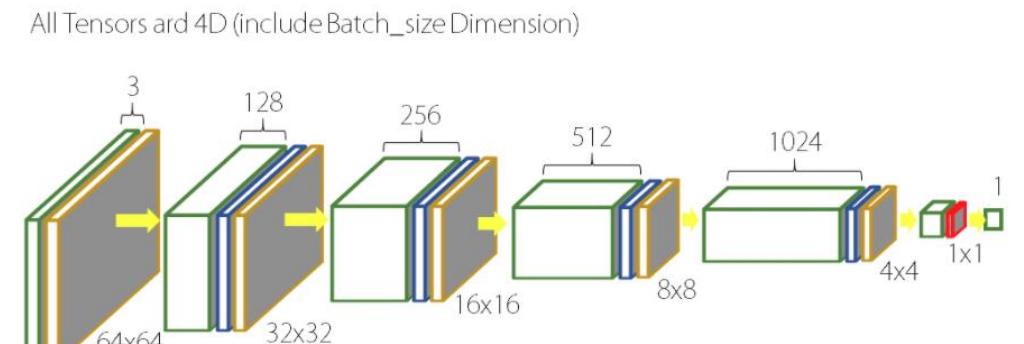
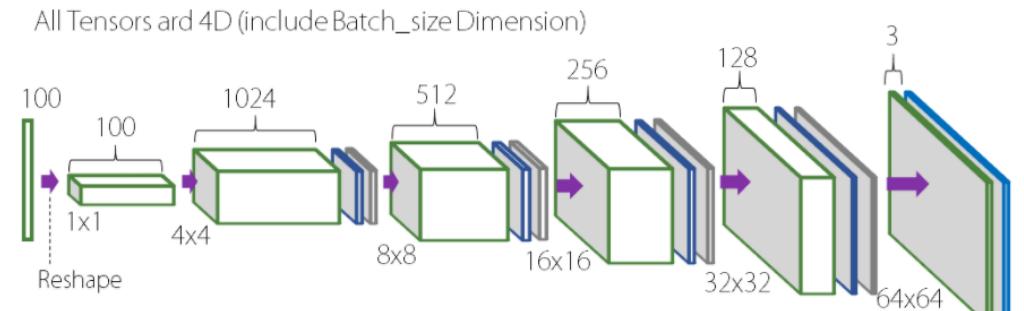
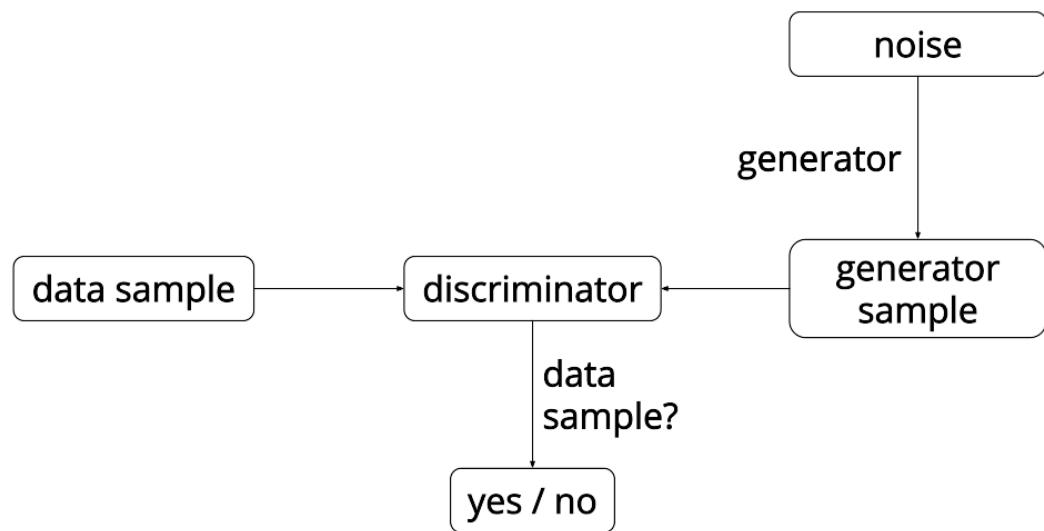
VAE

Auto Encoding Variational Bayes Arxiv 2013



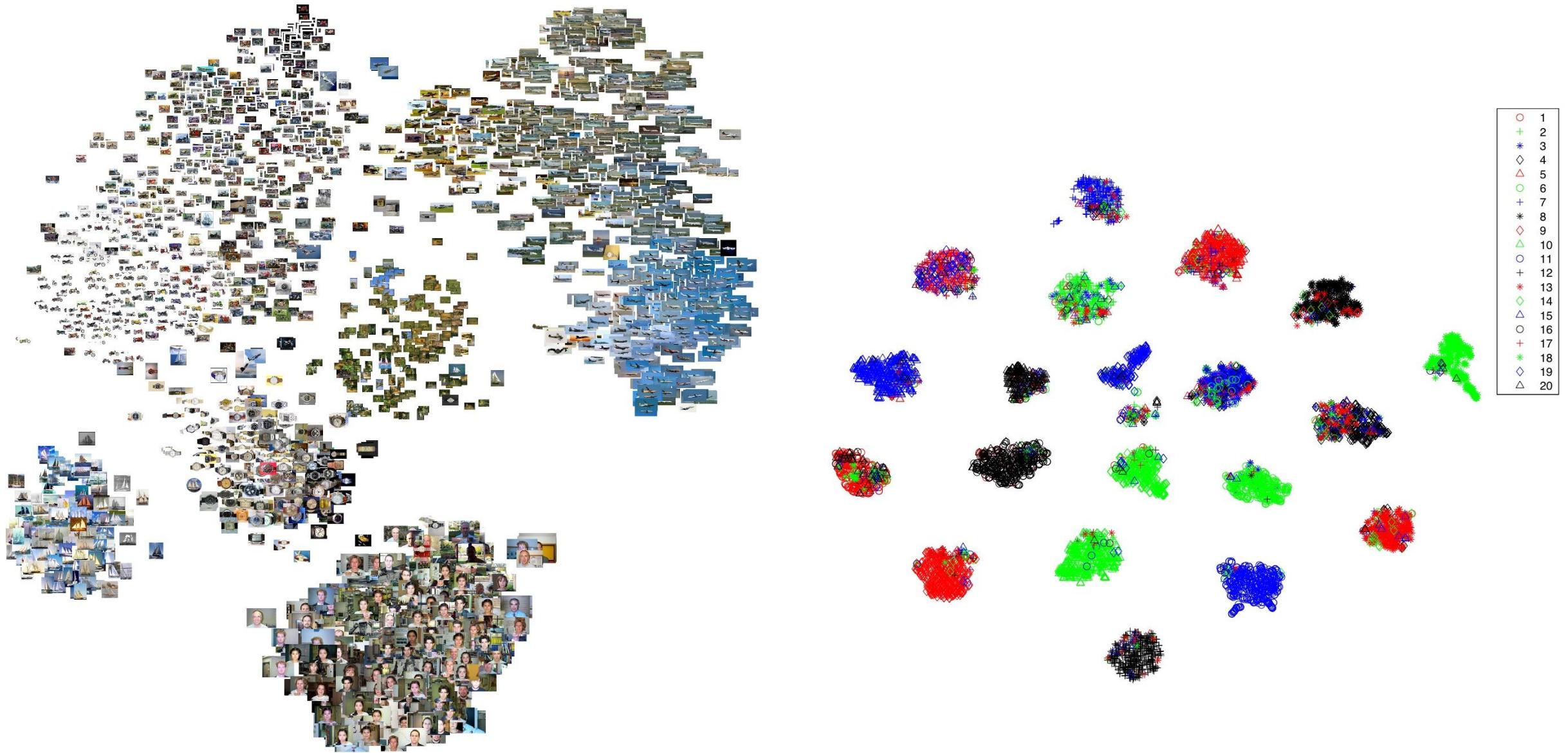
GAN

Generative Adversarial Nets NIPS 2014
DCGAN Arxiv 2015



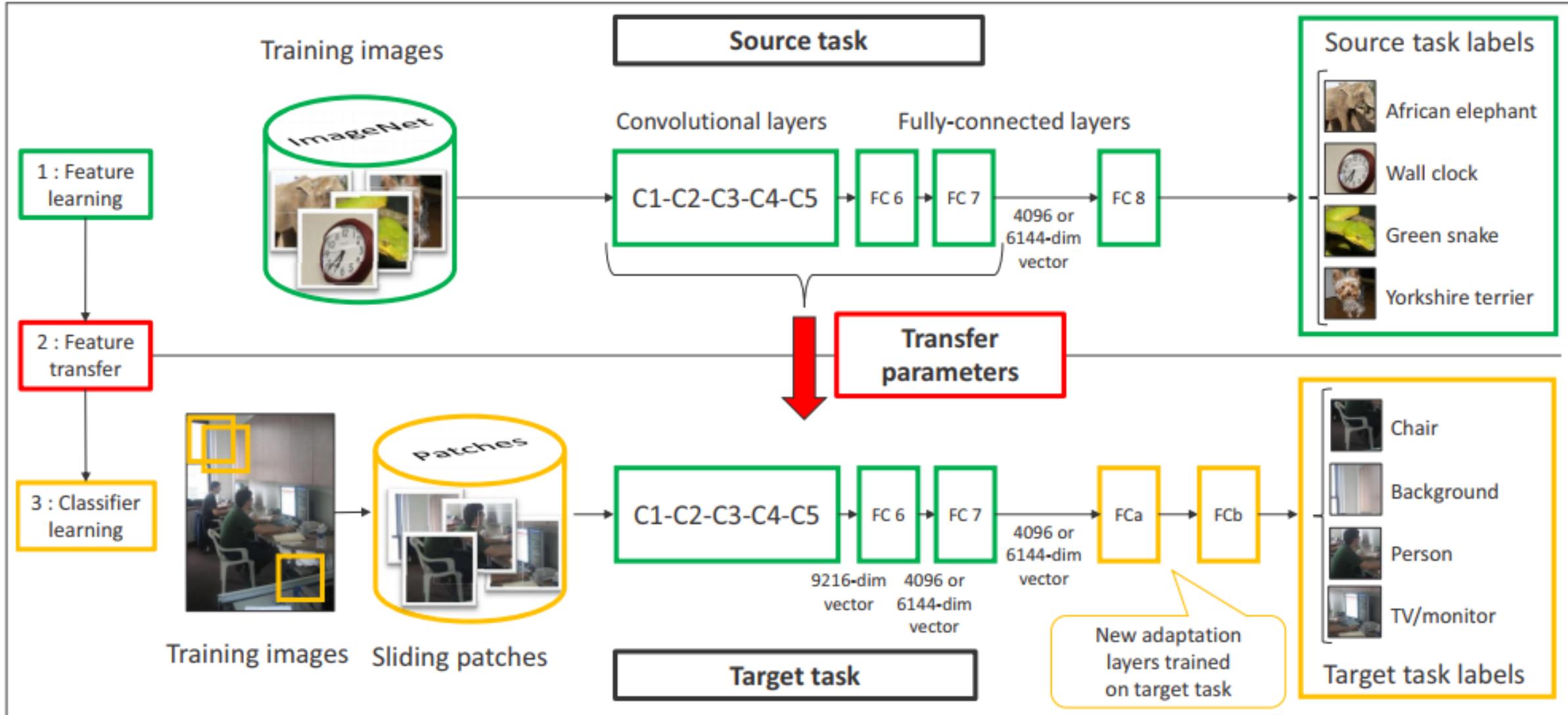
Dimension Reduction

Visualization Data using t-SNE JMLR 2008



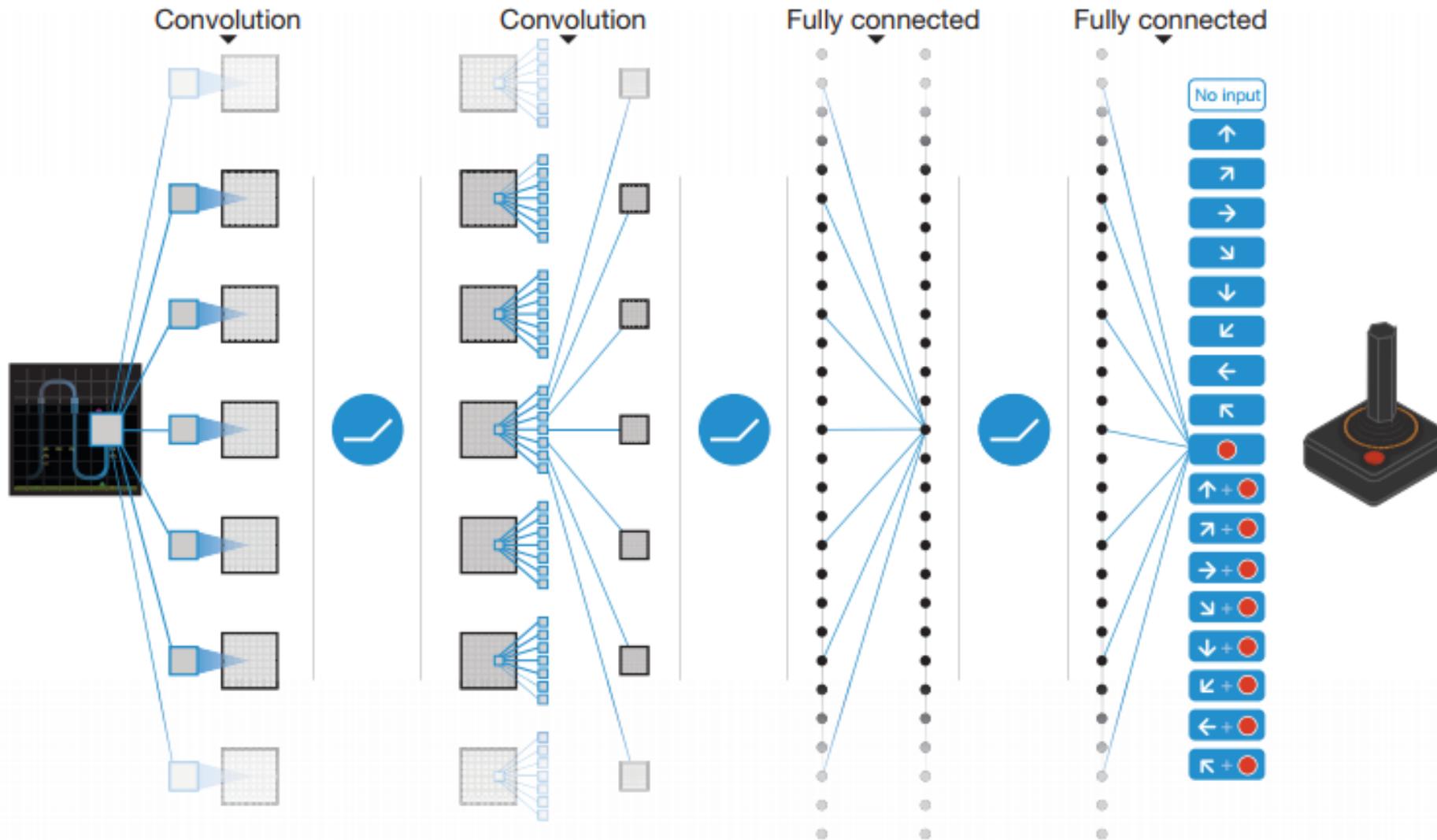
Transfer Learning

Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks CVPR 2014



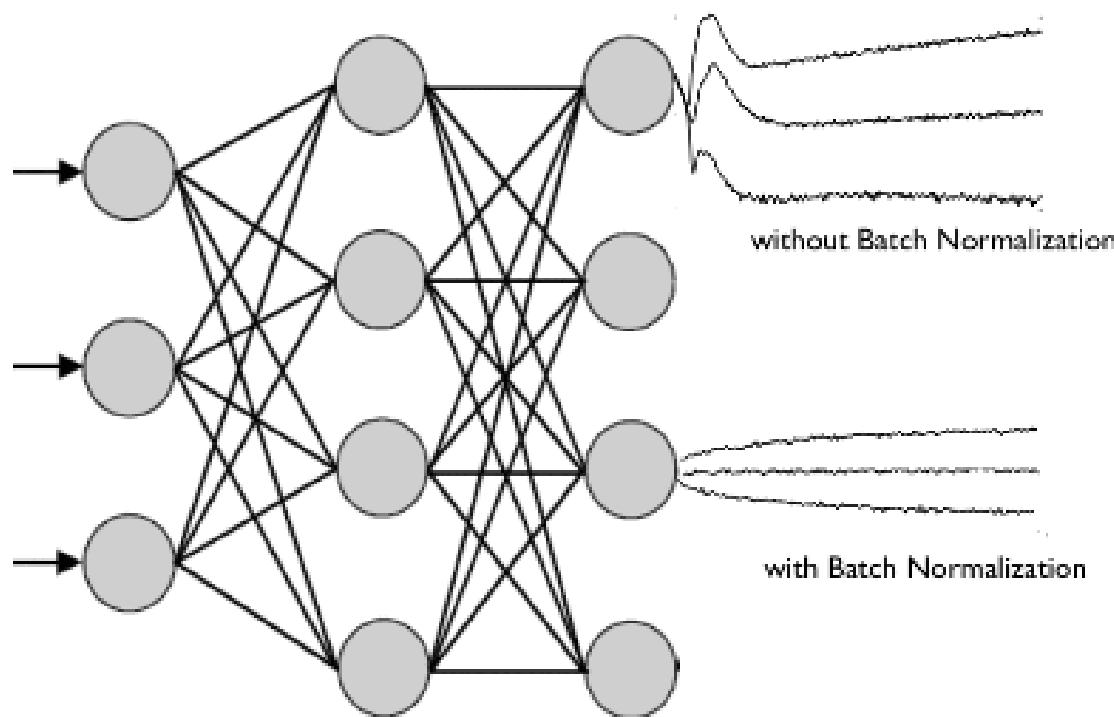
Reinforcement Learning

Human Level Control Through Deep Reinforcement Learning **NATURE 2015**



Training: Normalization

Batch Normalization JMLR 2015



Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1 \dots m\}$;

Parameters to be learned: γ, β

Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{normalize}$$

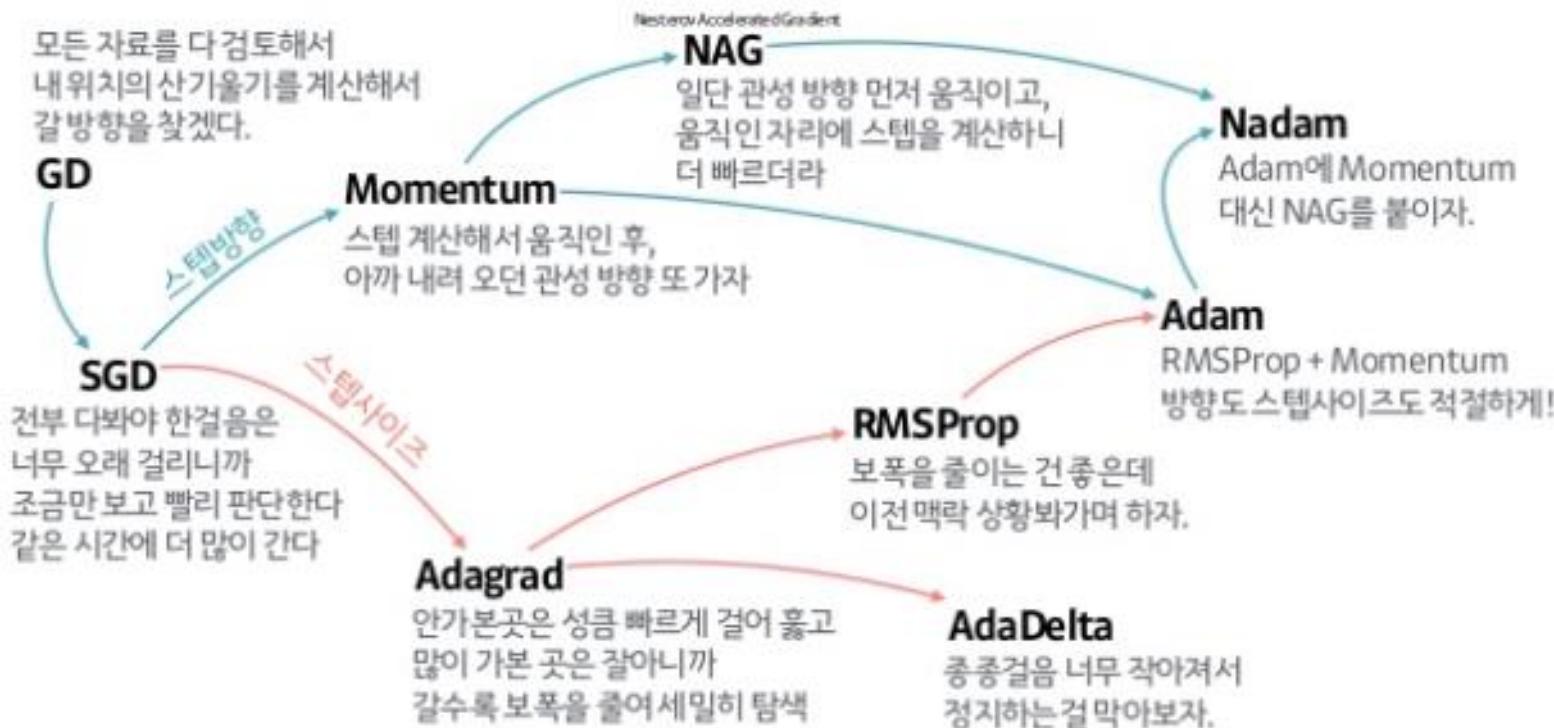
$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

Training: Optimization

Adam: A Method for Stochastic Optimization [Arxiv 2014](#)

산내려오는 작은 오솔길 잘찾기(Optimizer)의 발달 계보



Thank You
MVPLAB