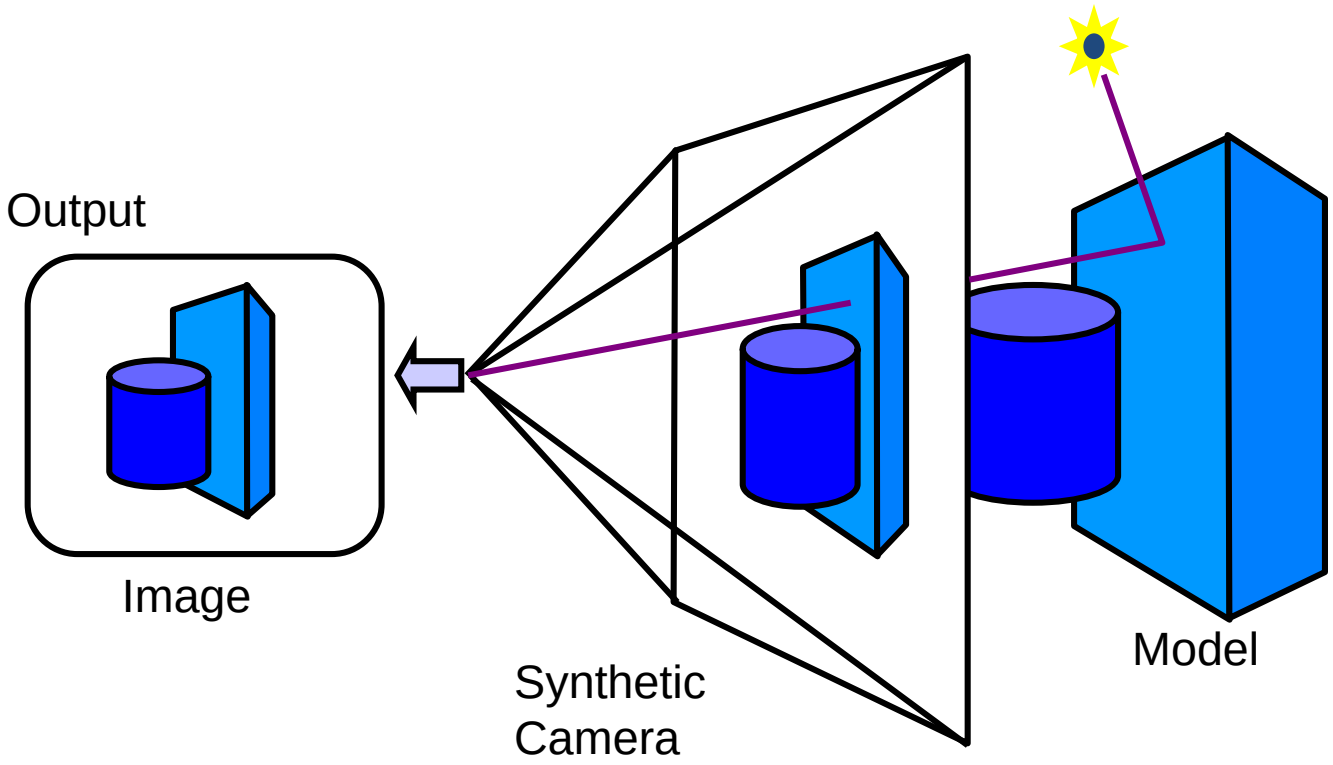
Machine Learning in Computer Vision and Graphics

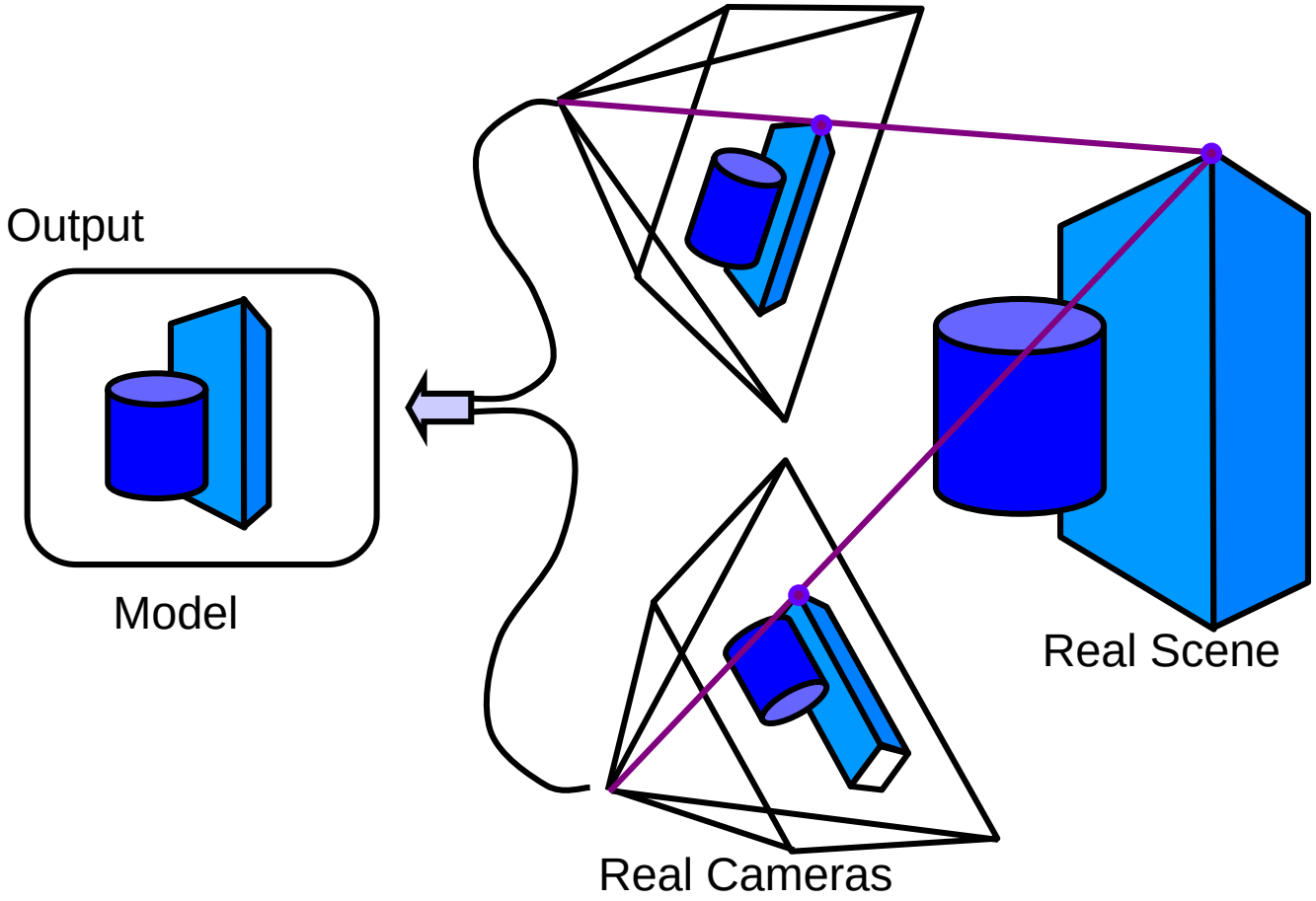
6.036 Introduction to Machine Learning

Computer Graphics – Synthesis



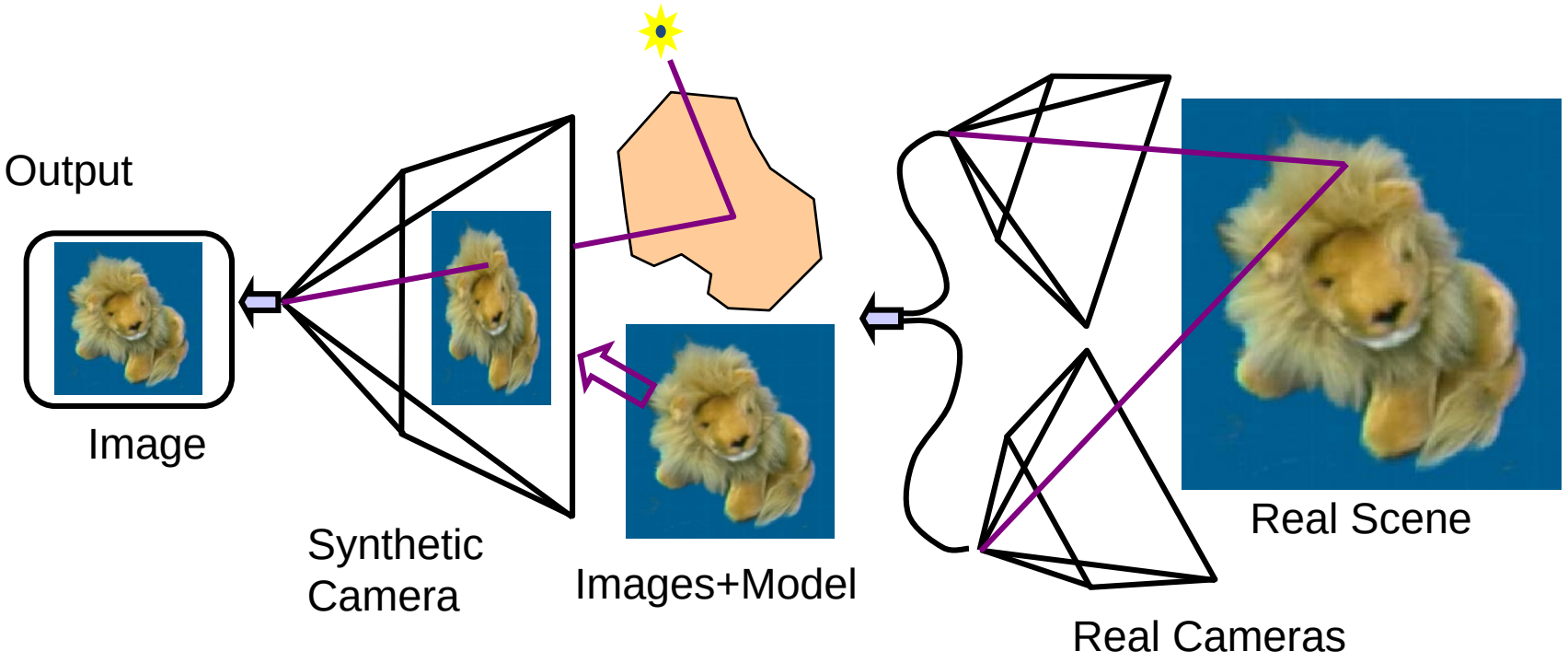
Source: R. Szeliski

Computer Vision - Analysis



Source: R. Szeliski

Often Together



Source: R. Szeliski

Machine learning is a core component of many vision and graphics algorithms

Let's start with something familiar...

Project 3

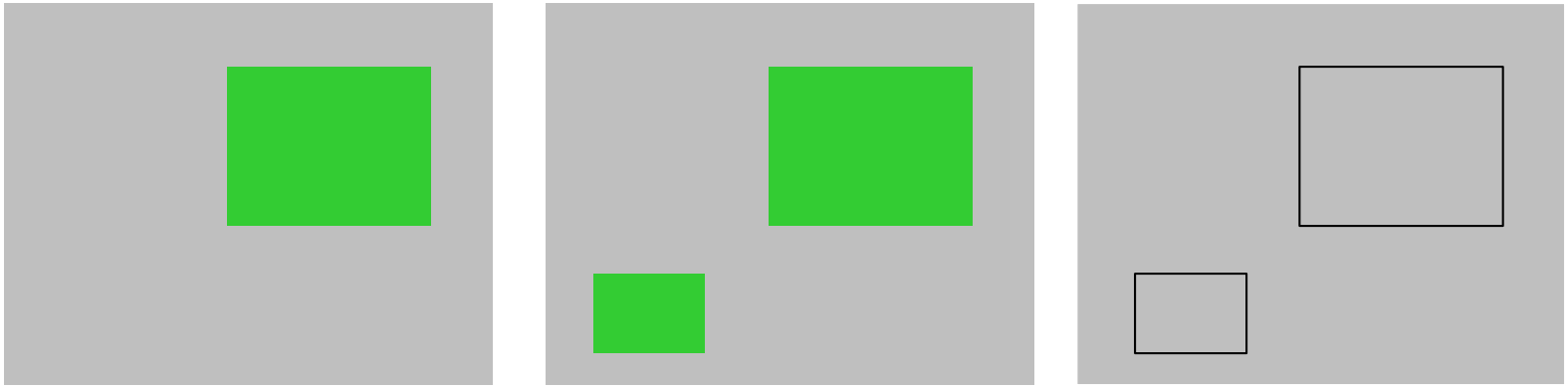
Image Segmentation

- Split an image into non-overlapping regions
- Group together similar-looking pixels

"superpixels"

Issues

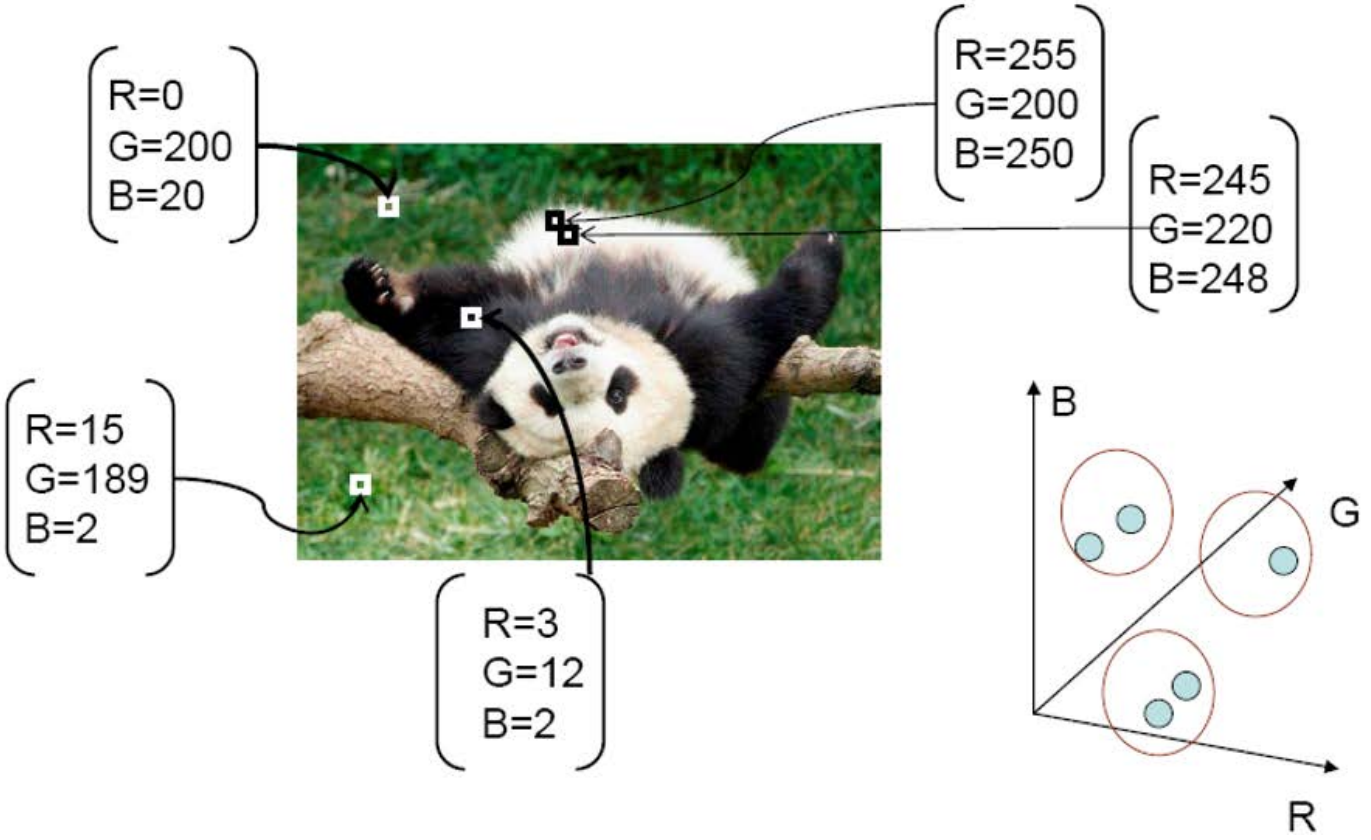
- How do we decide that two pixels are likely to belong to the same region?



- How many regions are there?

Segmentation: Clustering

- *Cluster similar pixels together*



Source: K. Grauman

A Simple Segmentation Algorithm

- Each pixel is described by a vector
  - e.g., z = [r, g, b]
- Run a clustering algorithm (e.g. k-means) using some distance between pixels:

$$D(\text{pixel}_i, \text{pixel}_j) = || z_i - z_j ||^2$$

K-Means Clustering

- Given *k*, the *k-means* algorithm consists of four steps:
  - Select initial centroids at random.
  - Assign each object to the cluster with the nearest centroid.
  - Compute each centroid as the mean of the objects assigned to it.
  - Repeat previous 2 steps until no change.

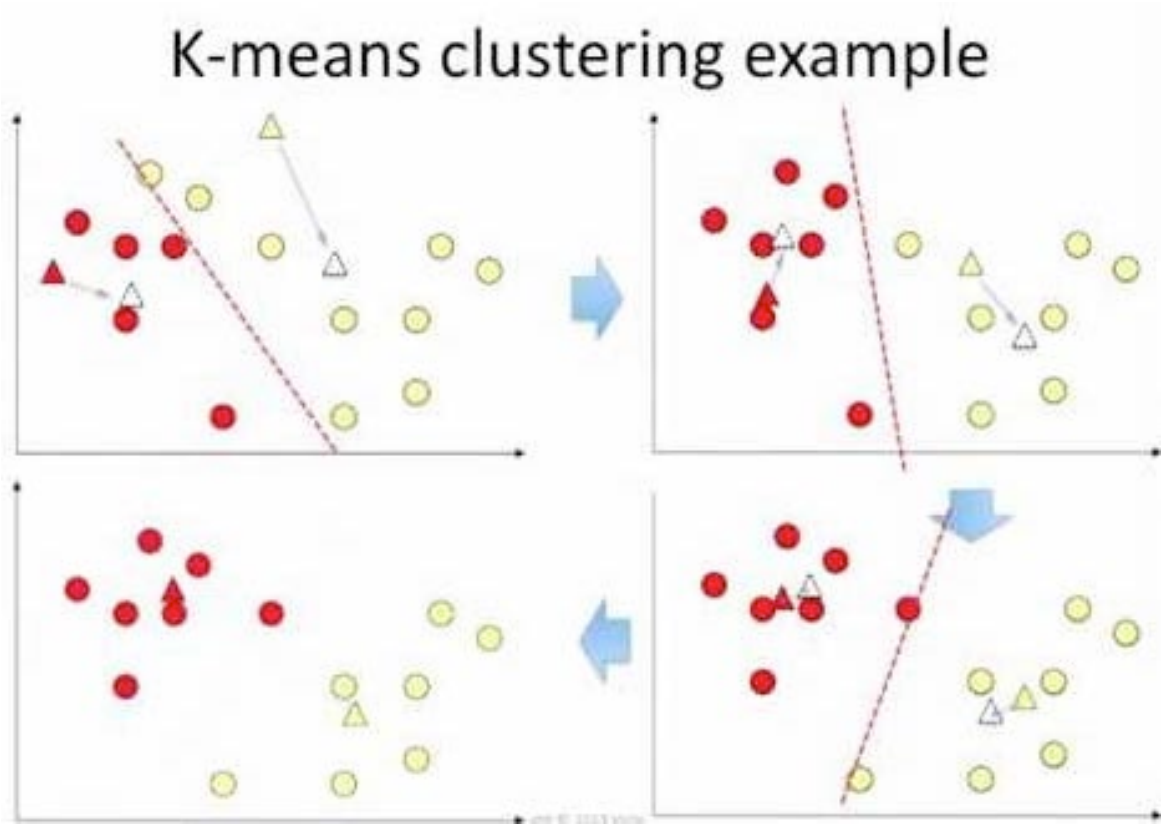
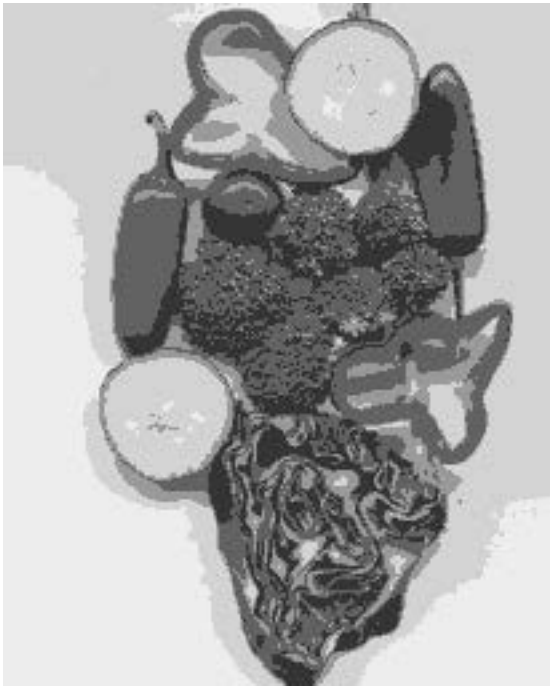


Image Segmentation: K-Means

- K-means (k=5) clustering based on intensity (middle) or color (right) is essentially vector quantization of the image attributes

Image

Intensity-based clusters



Color-based clusters

each pixel is replaced with the mean value of its cluster

K-means using color alone (k=11 clusters) Showing 4 of the segments, (not necessarily connected) Some are good, some meaningless

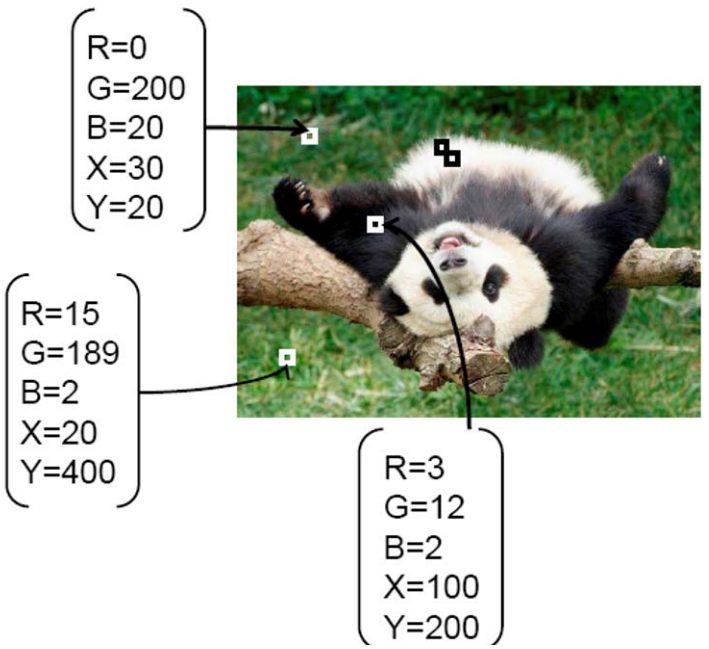
Including Spatial Relationships

Augment data to be clustered with spatial coordinates.

$$z = \begin{pmatrix} r \\ g \\ b \\ x \\ y \end{pmatrix}$$

color coordinates (or r,g,b)

spatial coordinates

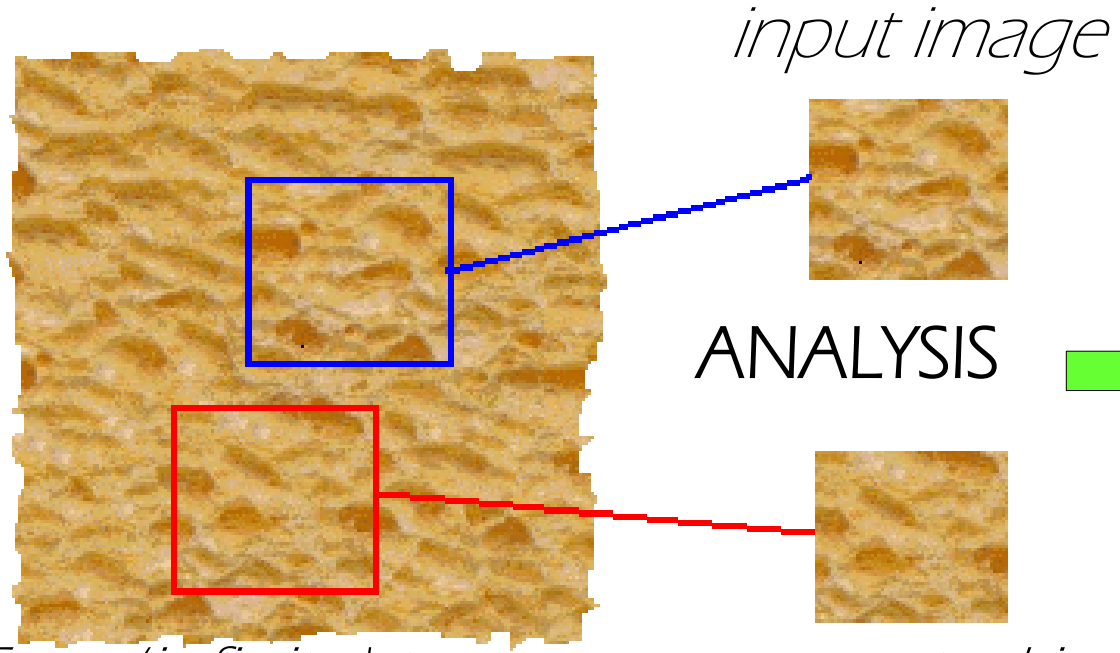


K-Means based on (r,g,b,x,y)

K-means using color and position, 20 segments

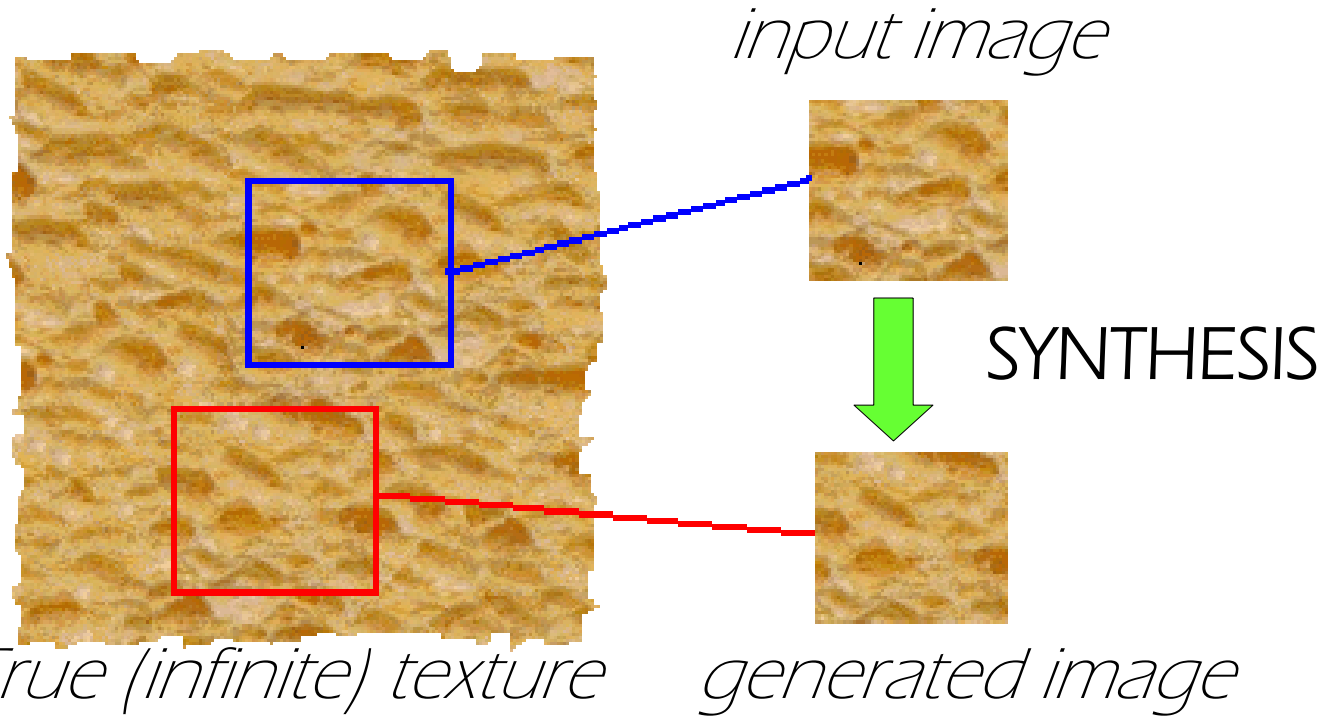
Let's look at other classic problems

Texture Analysis



Compare textures and decide if they're made of the same "stuff".

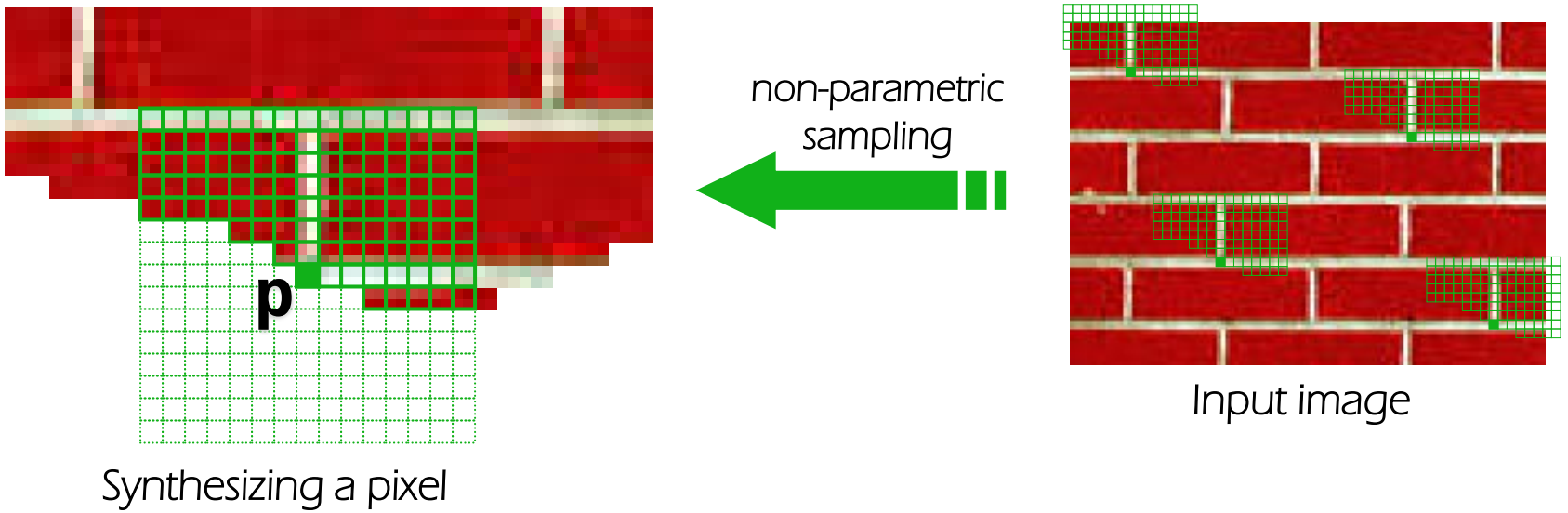
Texture Synthesis



Given a finite sample of some texture, the goal is to synthesize other samples from that same texture

- The sample needs to be "large enough"

Textures Synthesis

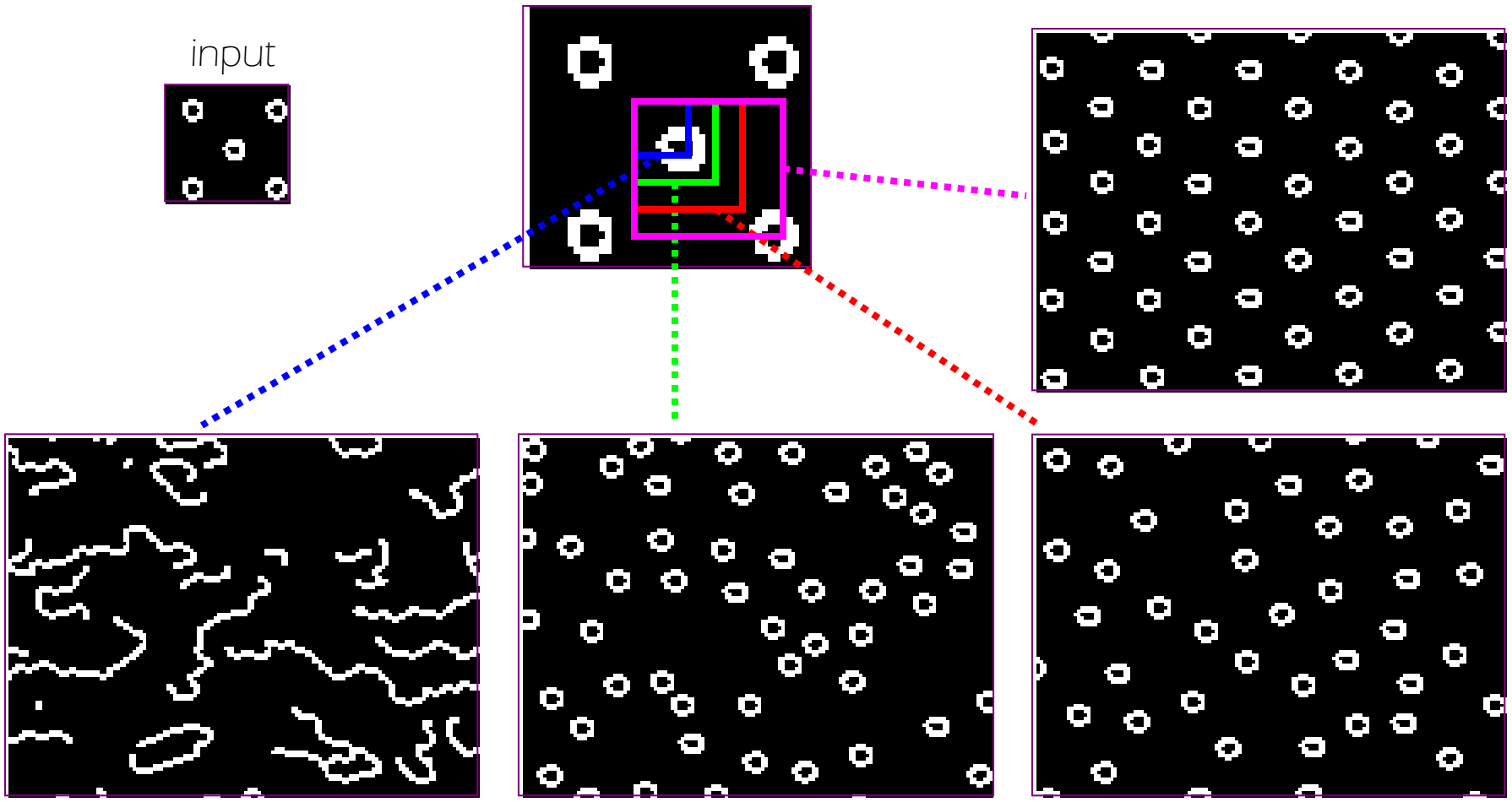


Assuming Markov property, compute P(p|N(p))

- Building explicit probability tables infeasible
- Instead, we *search the input image* for all similar neighbourhoods — that's our pdf for p
- To sample from this pdf, just pick one match at random

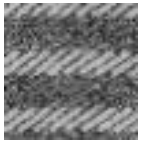
Source: Efros

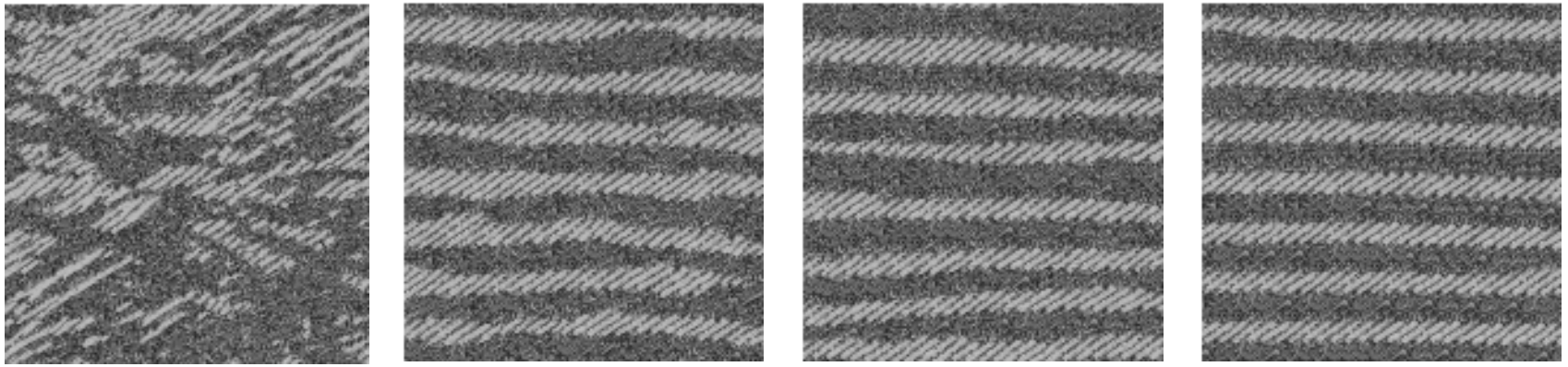
Neighborhood Window

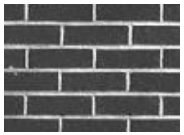


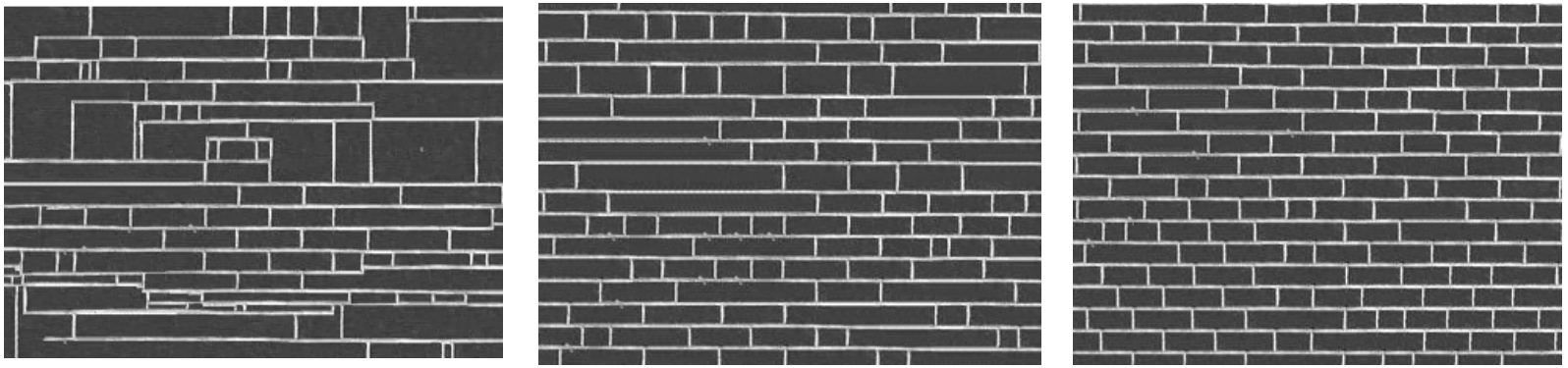
Source: Efros

Varying Window Size







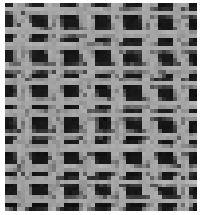


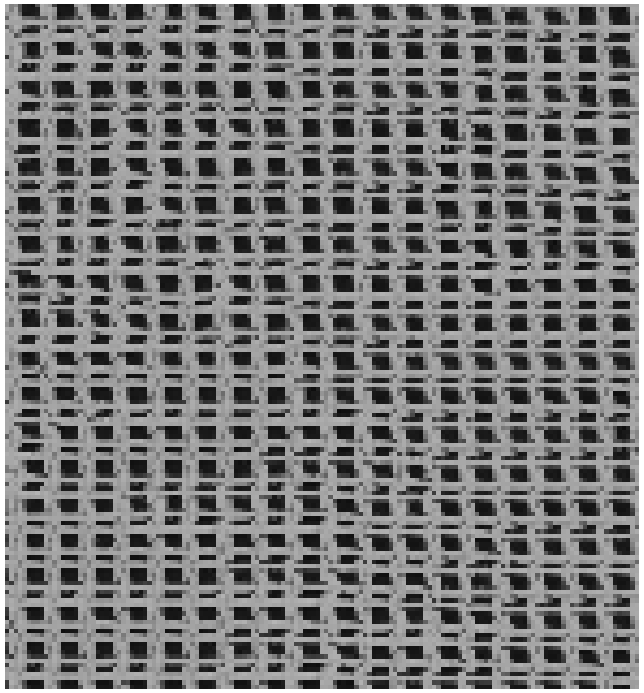
Increasing window size

Source: Efros

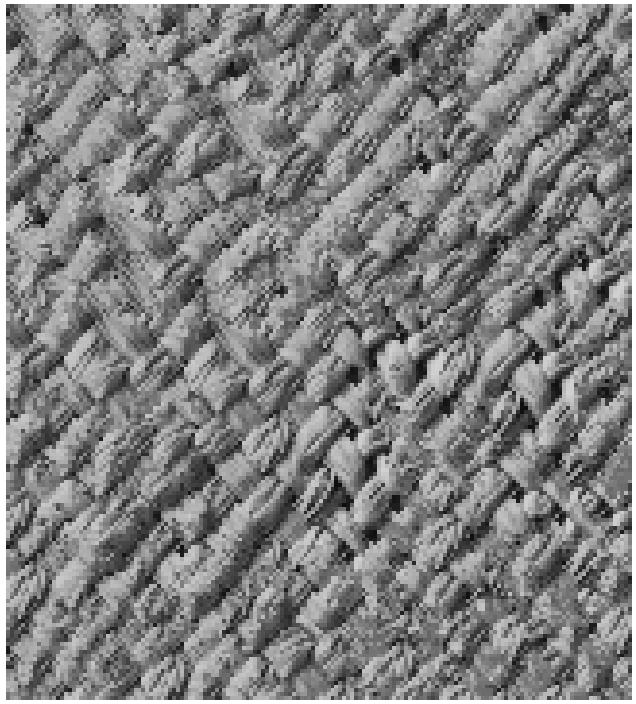
Synthesis Results

French canvas





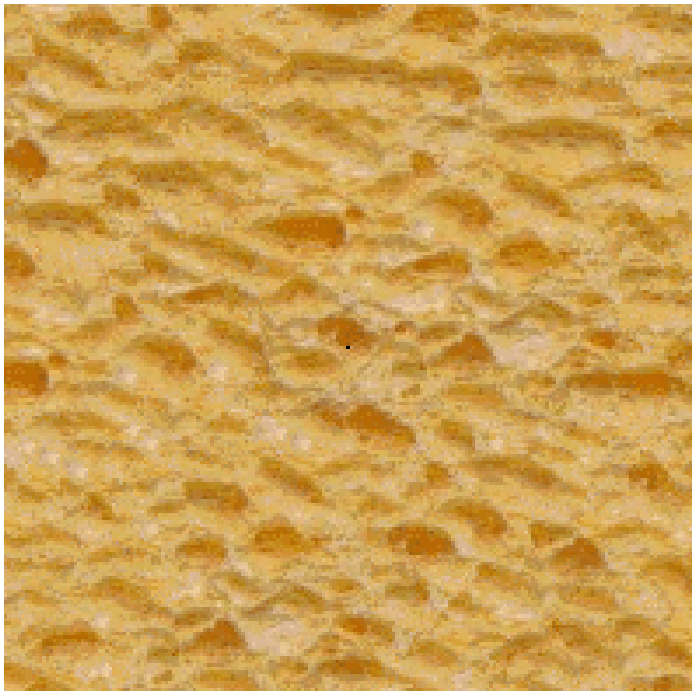
rafia weave



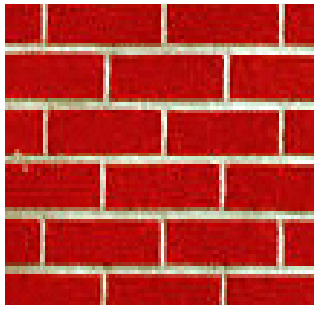
Source: Efros

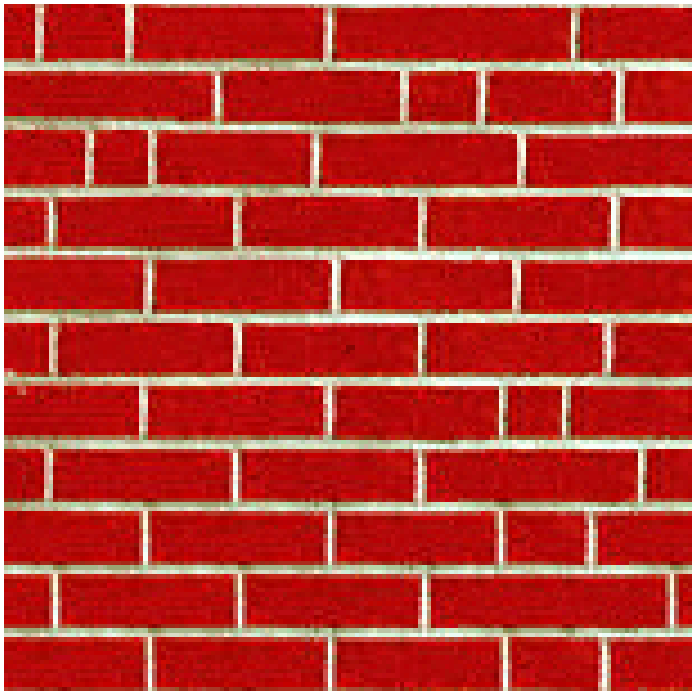
More Results

white bread



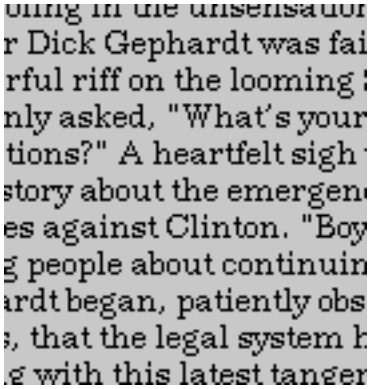
brick wall

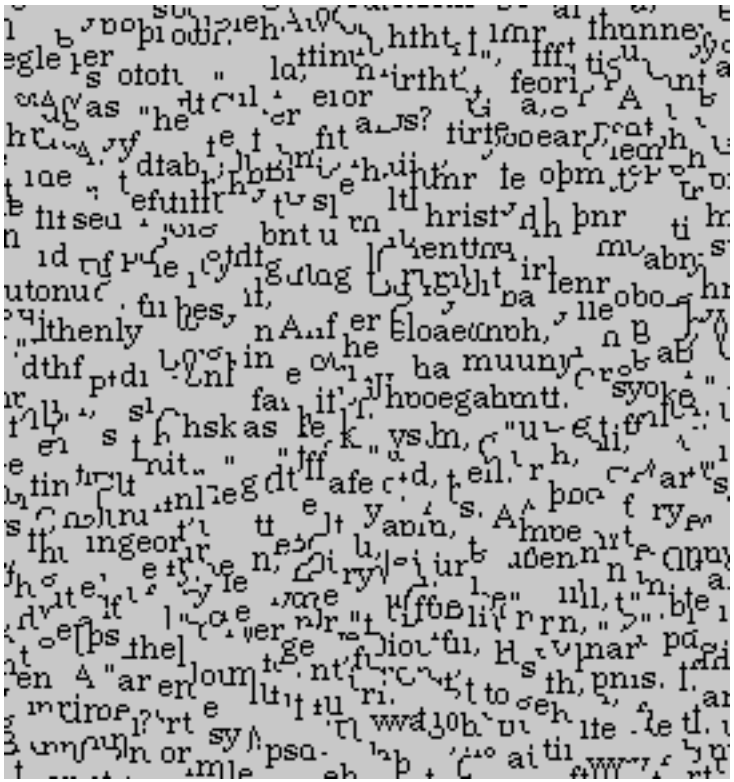


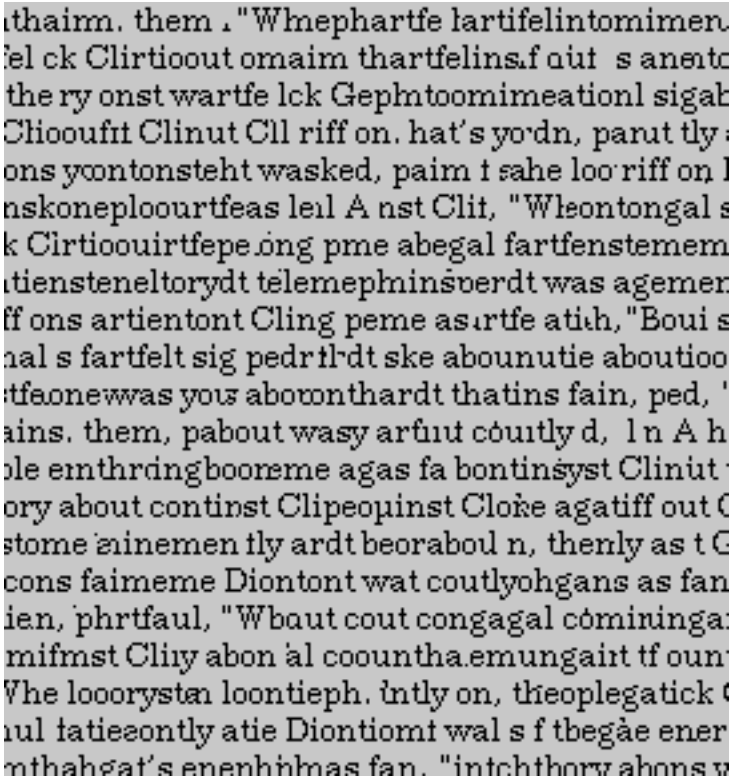


Source: Efros

Homage to Shannon

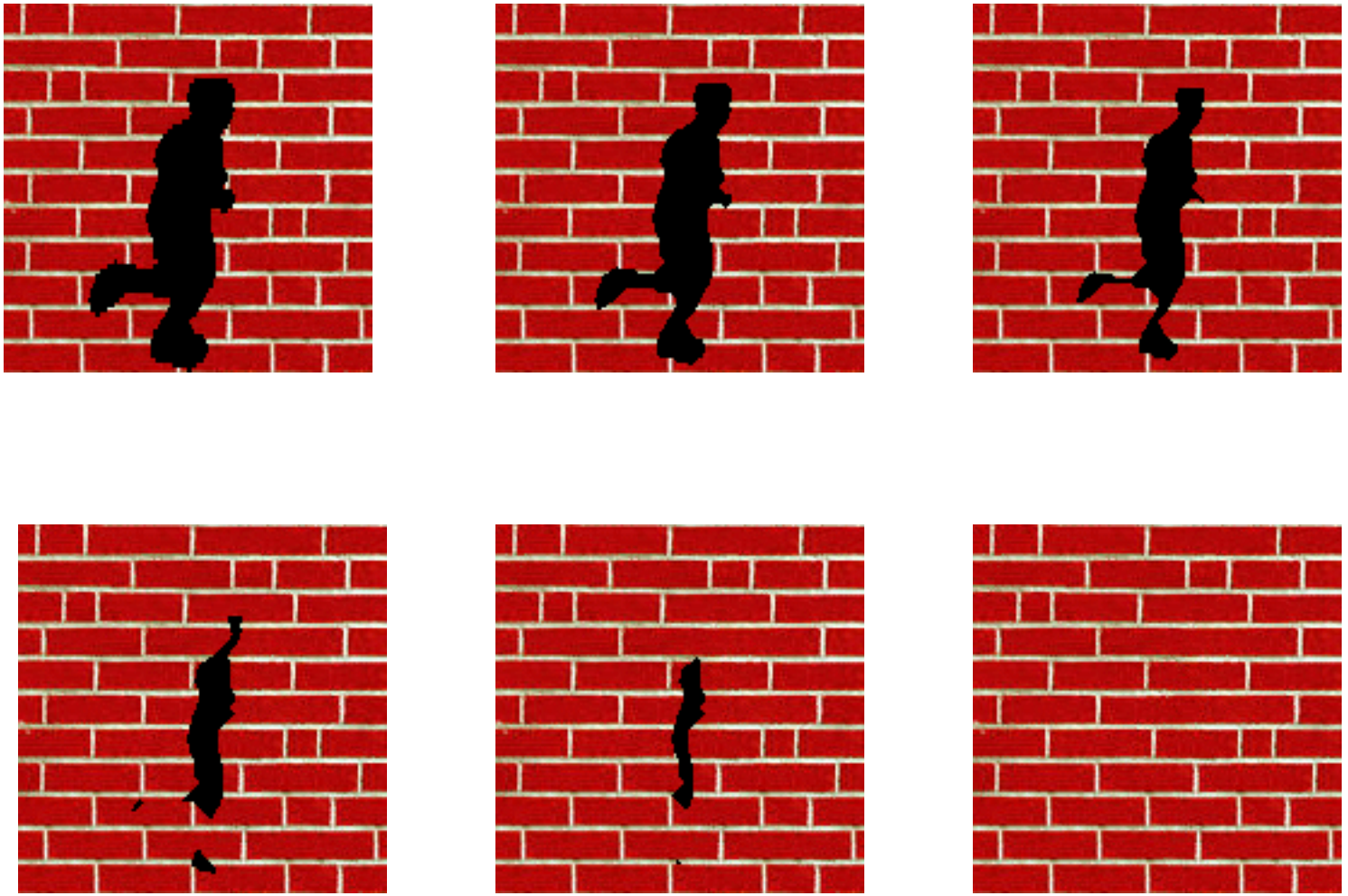






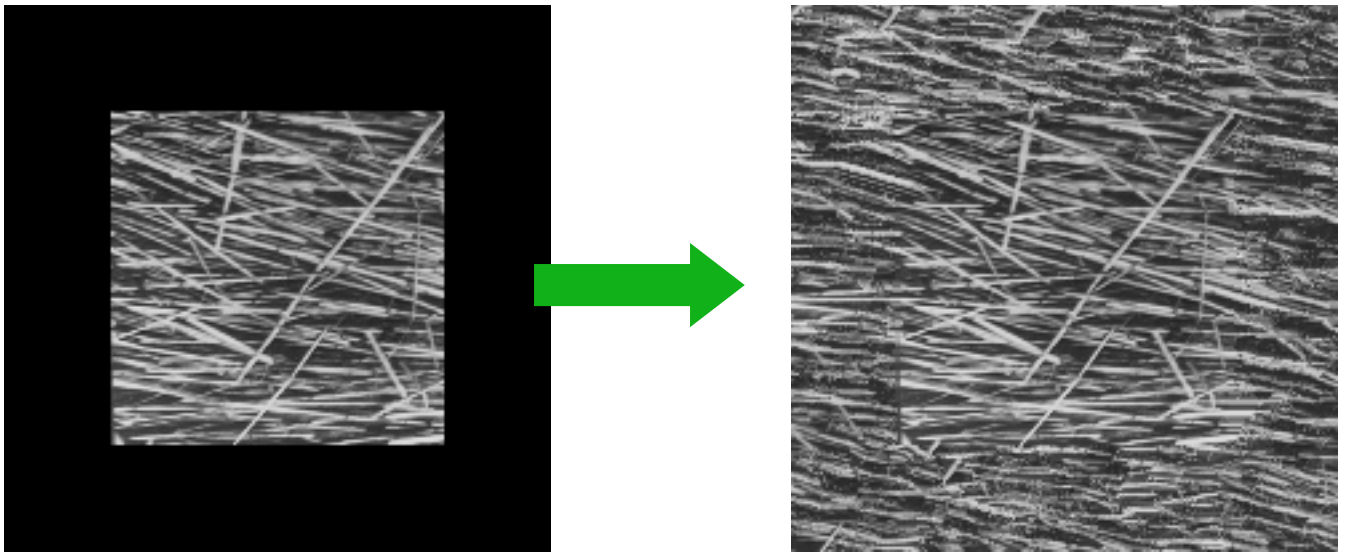
Source: Efros

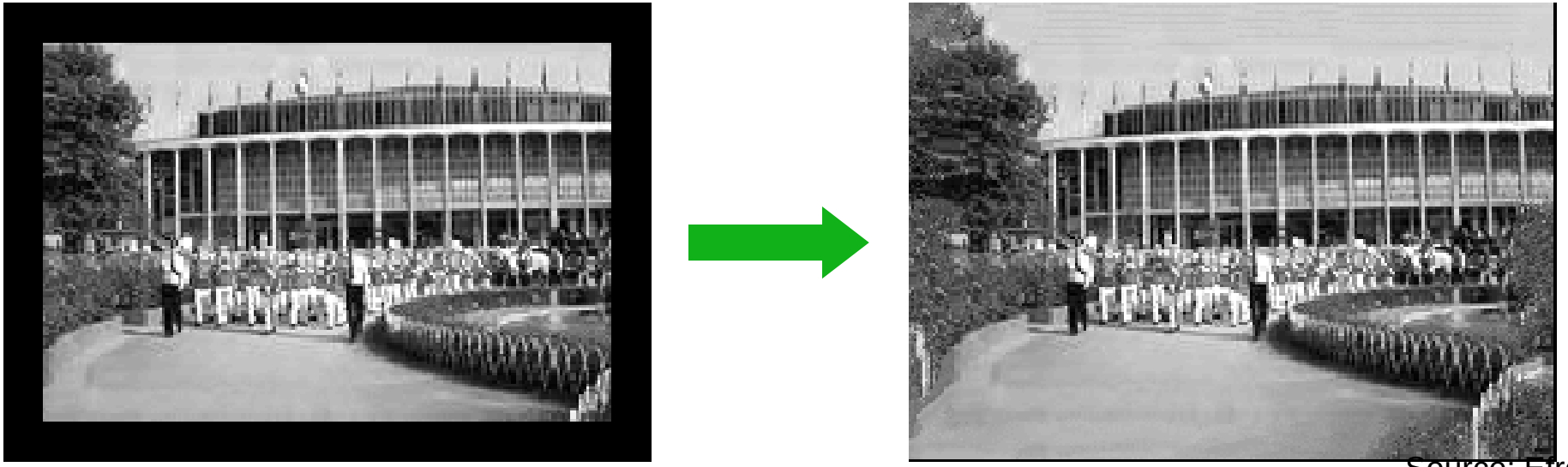
Hole Filling



Source: Efros

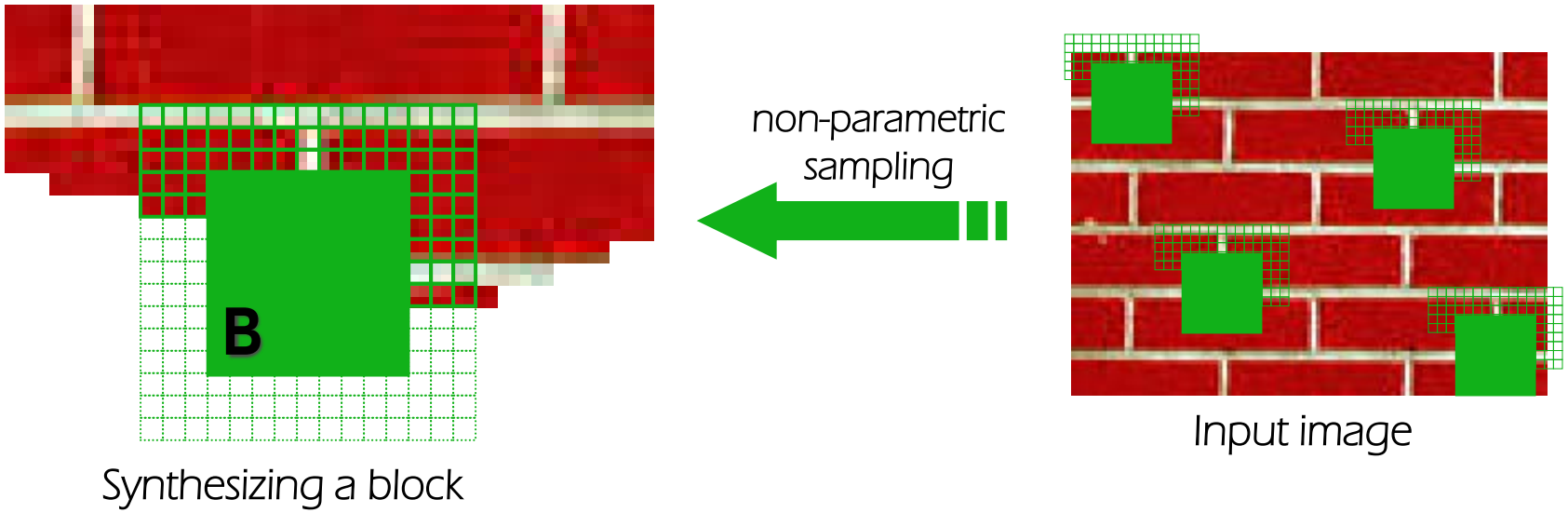
Extrapolation





Source: Efros

Image Quilting

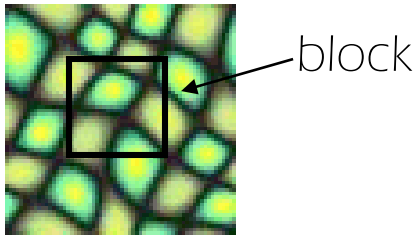


- Observation: neighbor pixels are highly correlated

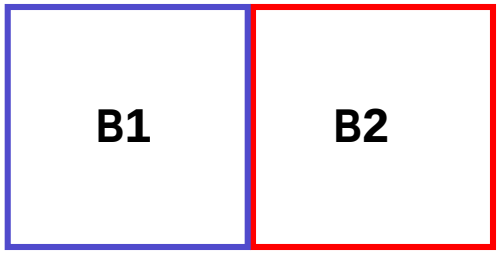
Idea: unit of synthesis = block

- Exactly the same but now we want P(B|N(B))
- Much faster: synthesize all pixels in a block at once
- Not the same as multi-scale!

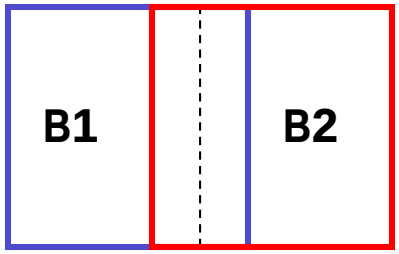
Source: Efros



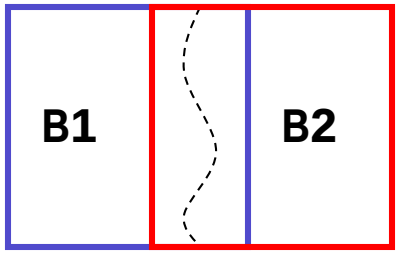
Input texture



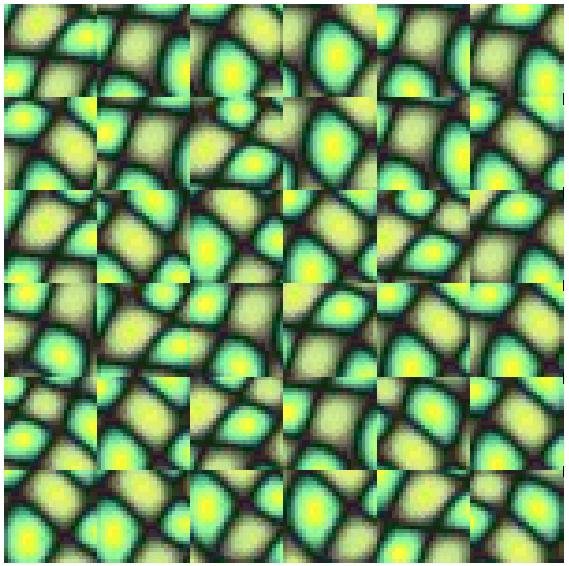
Random placement of blocks

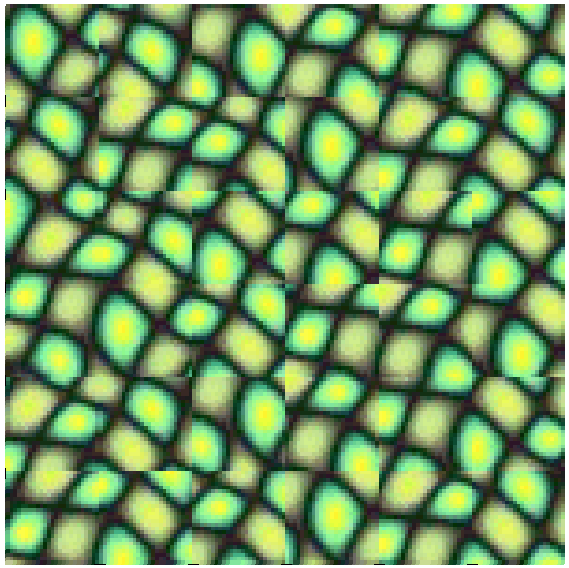


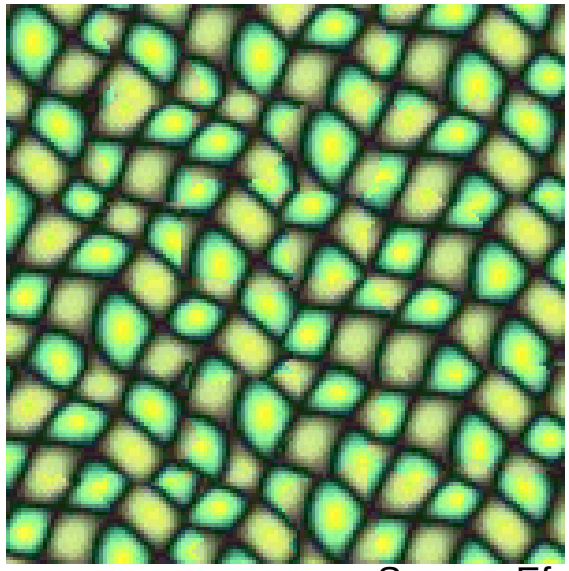
Neighboring blocks constrained by overlap



Minimal error boundary cut

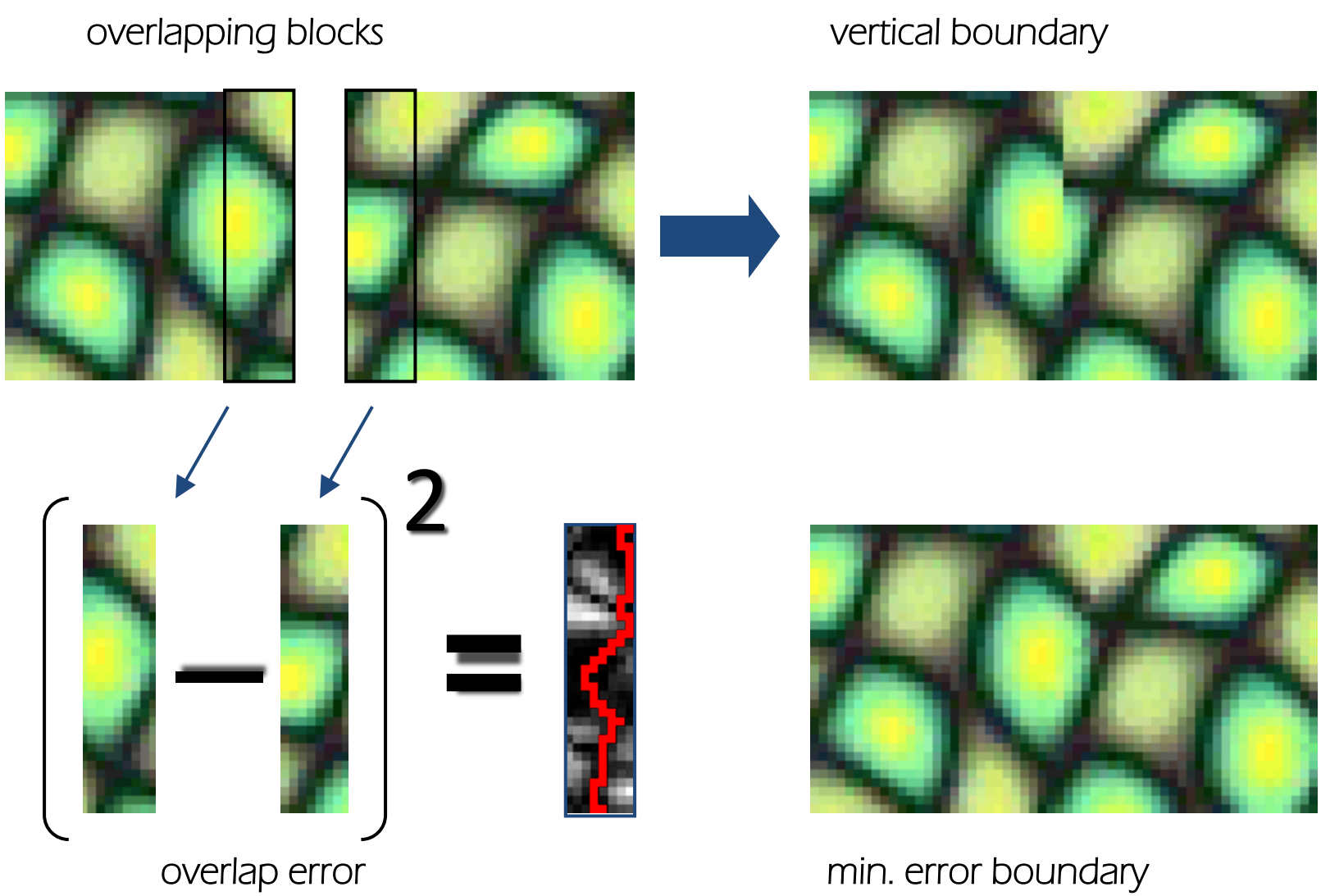






Source: Efros

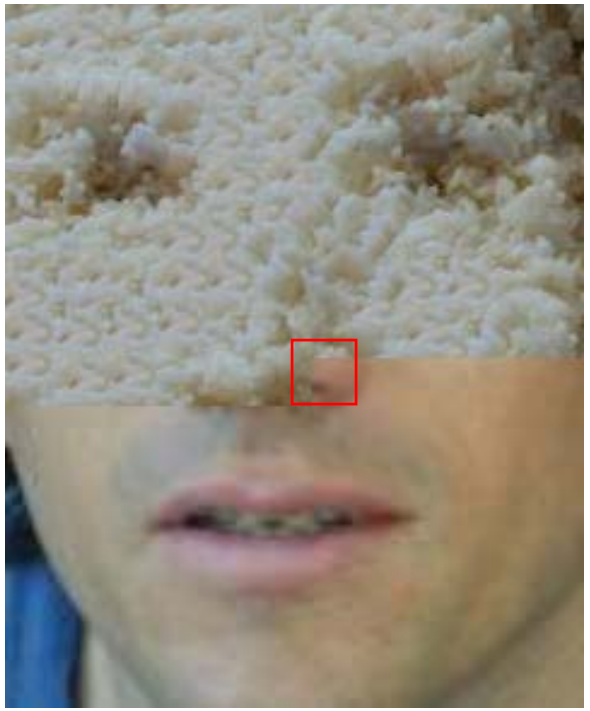
Minimal Error Boundary



Source: Efros

Texture Transfer

- Take the texture from one object and "paint" it onto another object
  - This requires separating texture and shape
  - That's HARD, but we can cheat
  - Assume we can capture shape by boundary and rough shading



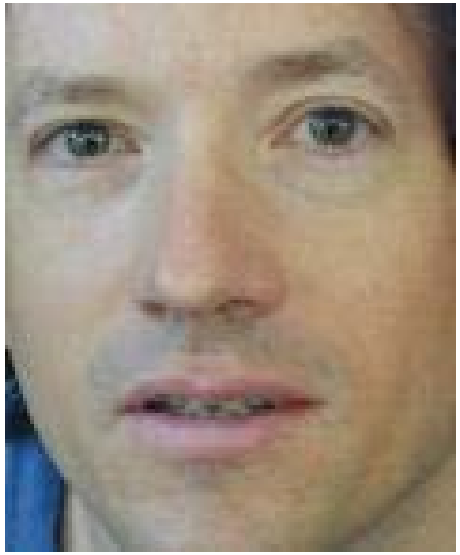
Then, just add another constraint when sampling: similarity to underlying image at that spot

Source: Efros

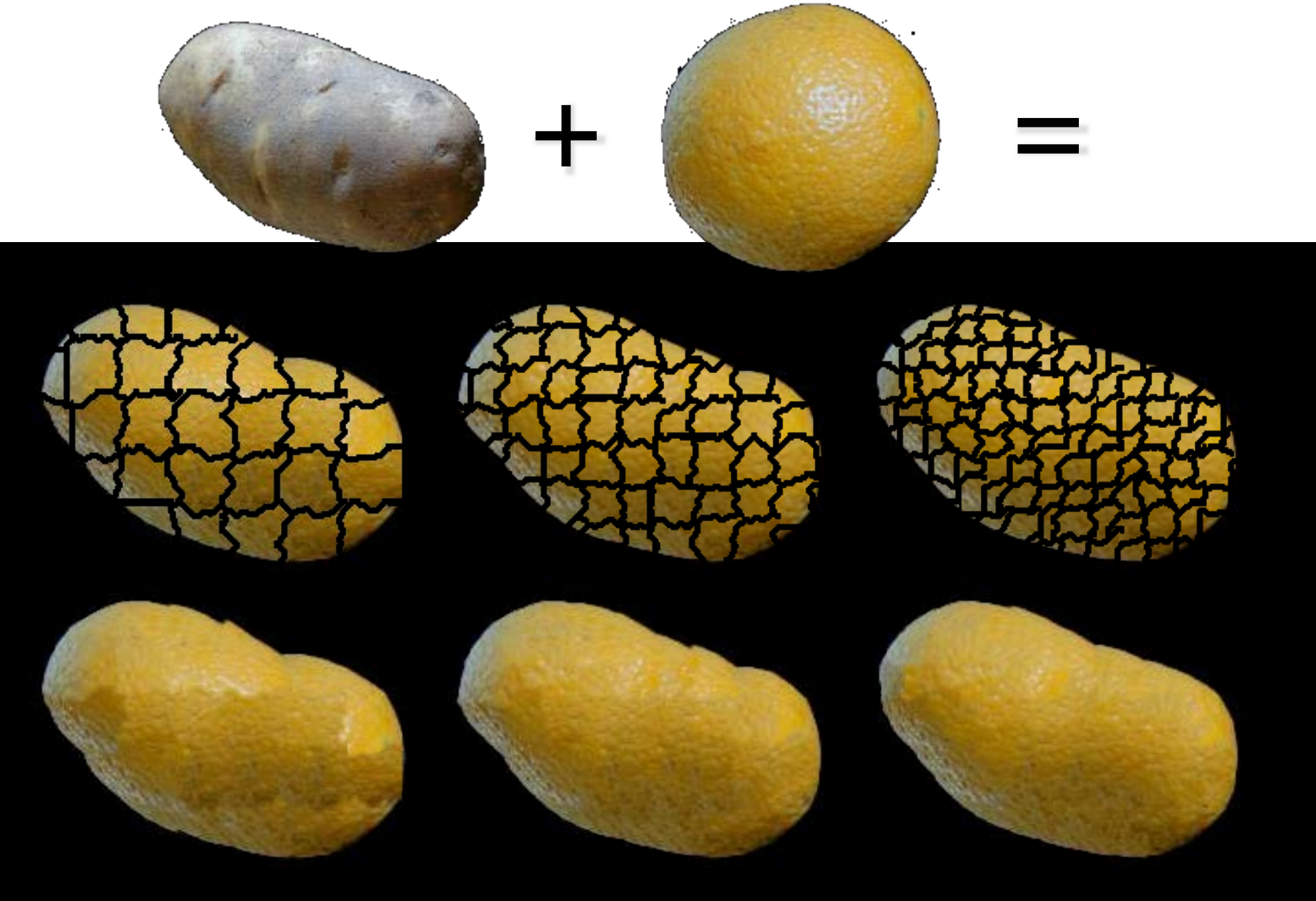
parmesan

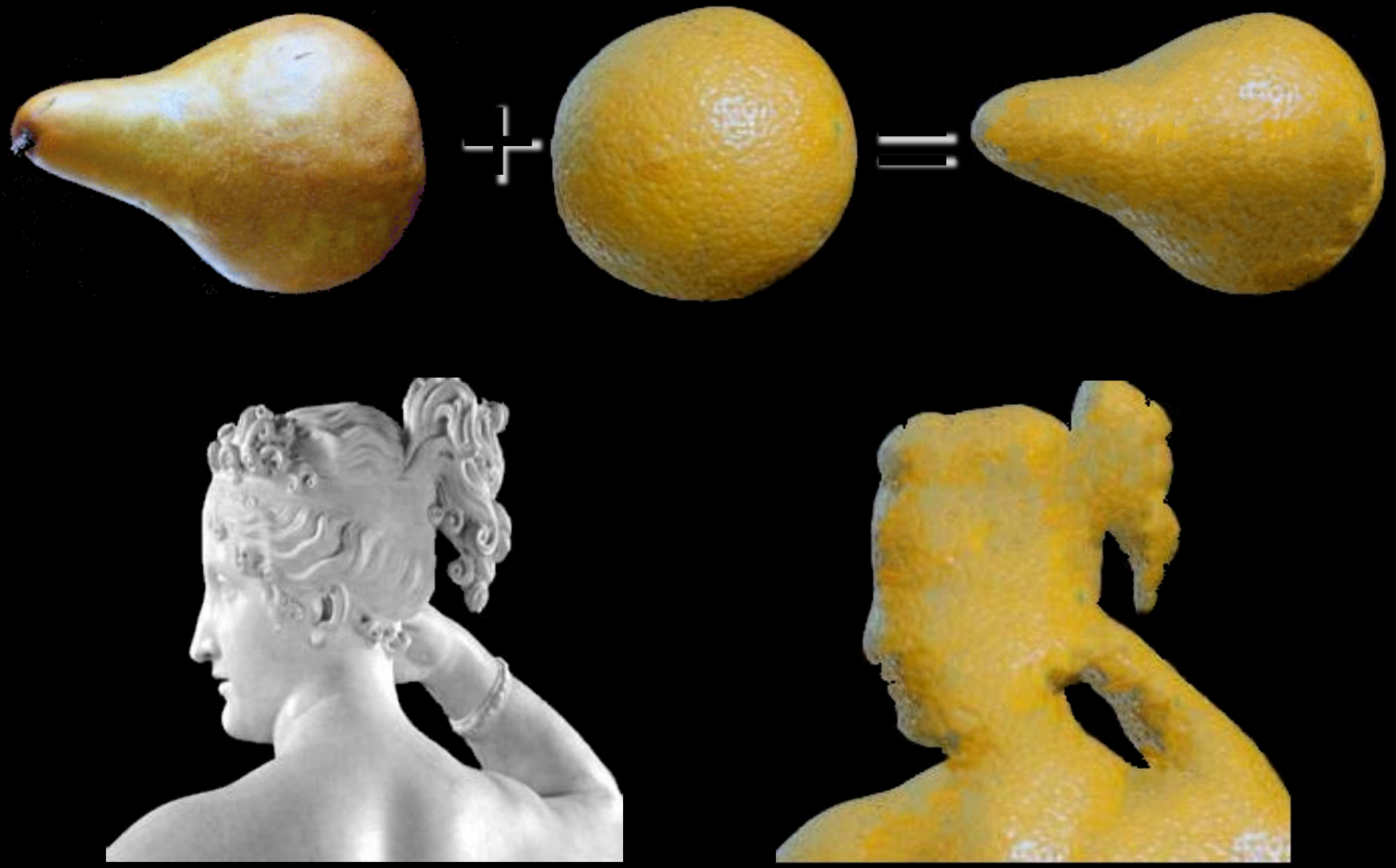


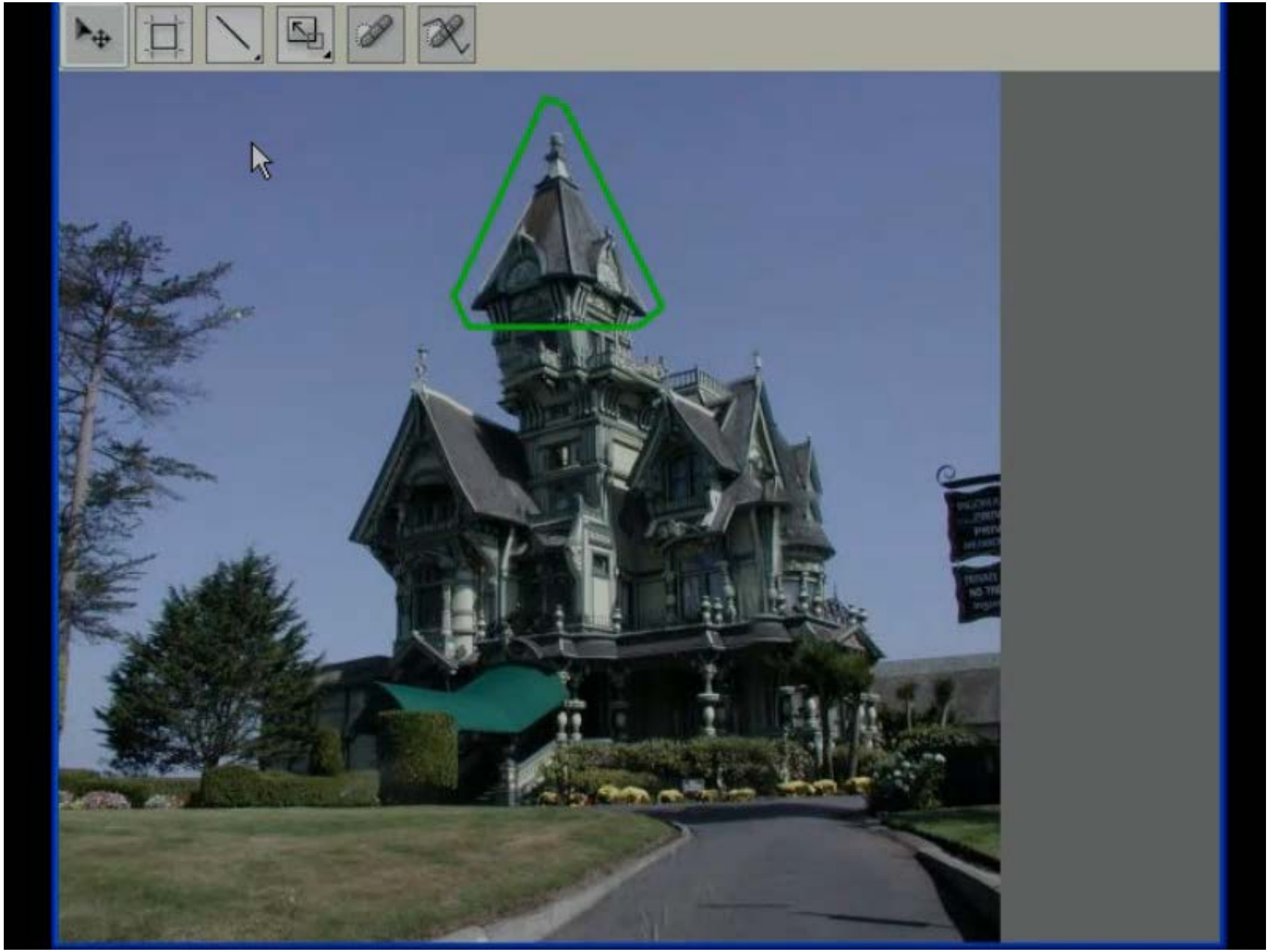
rice



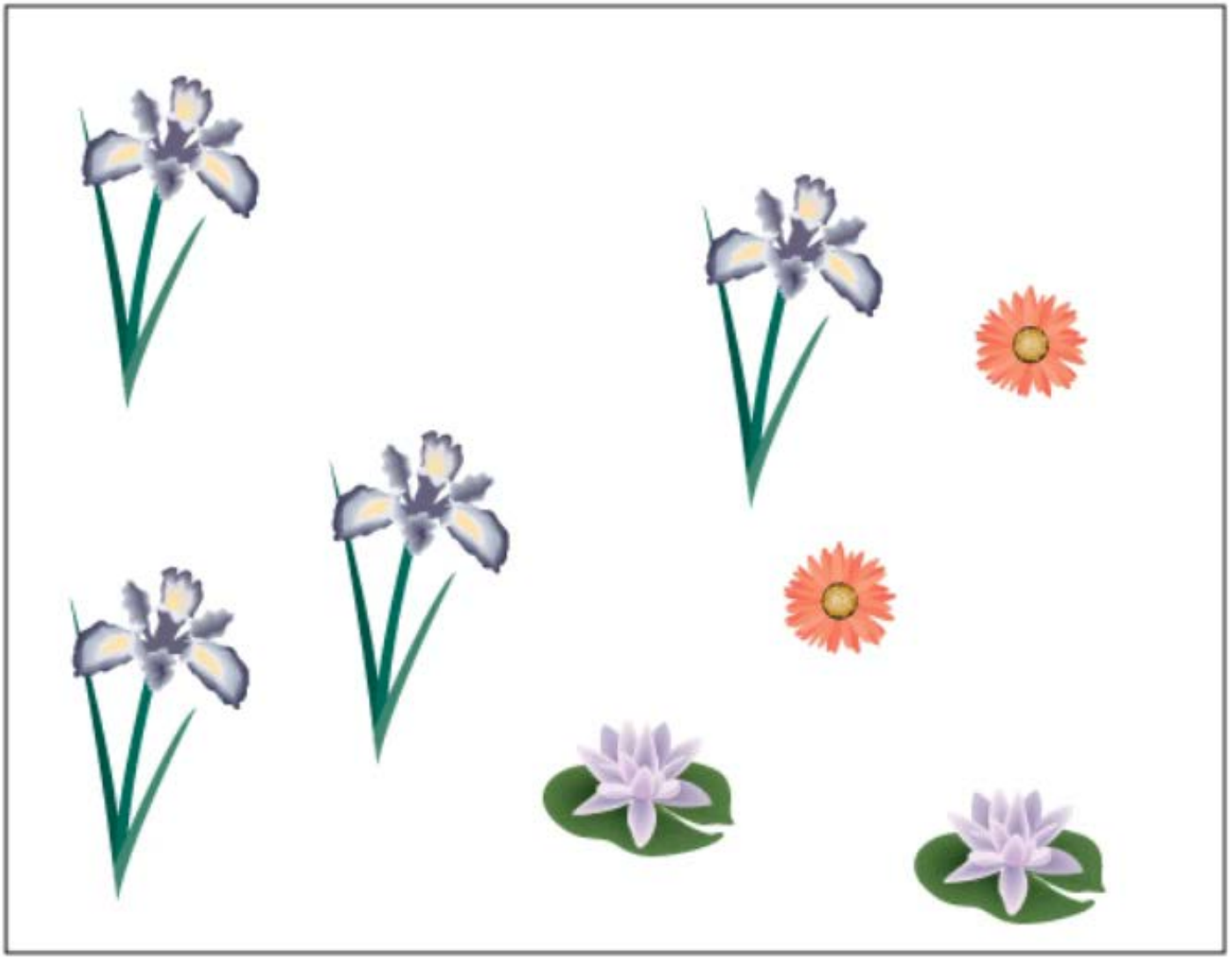
Source: Efros



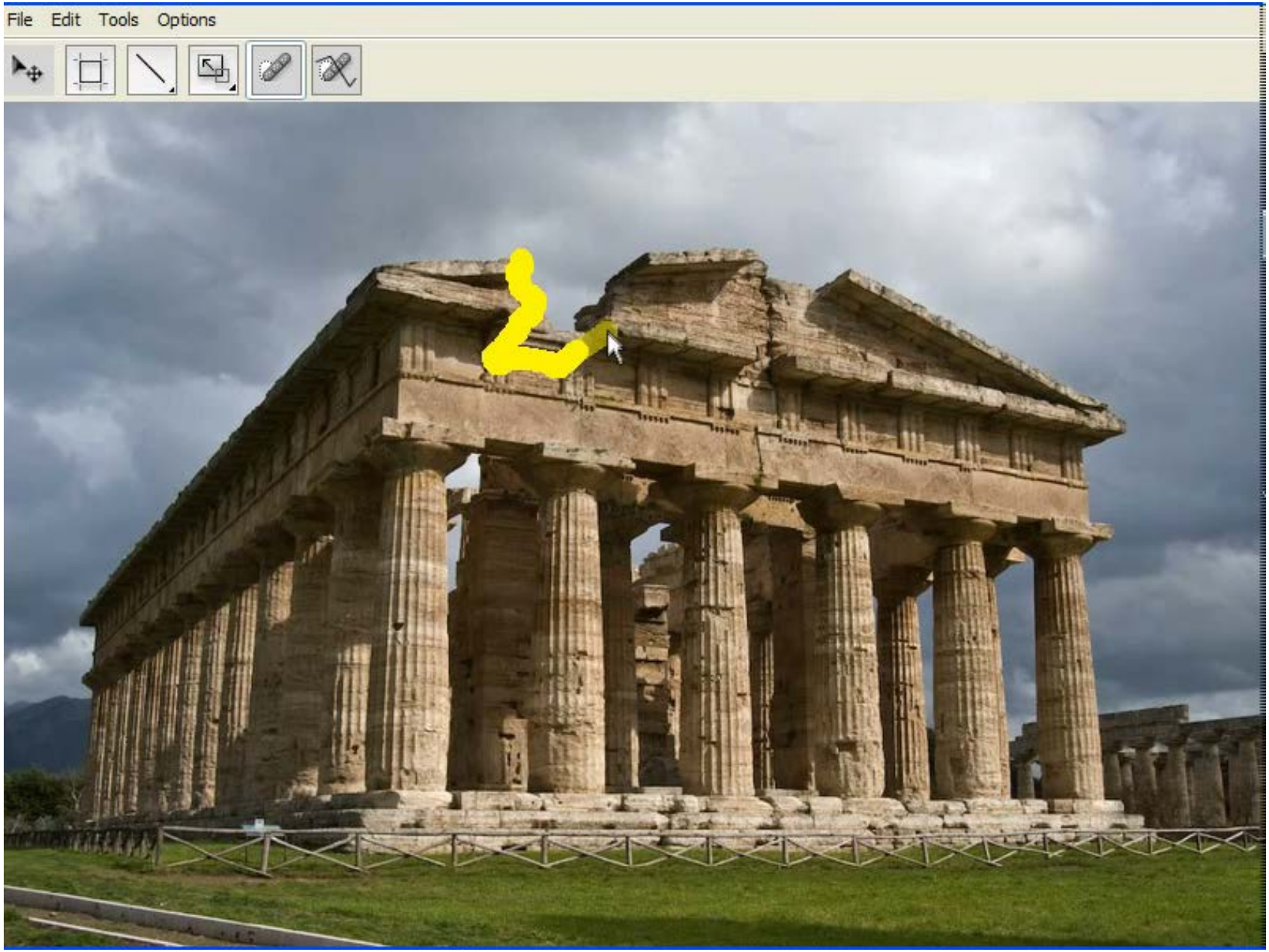




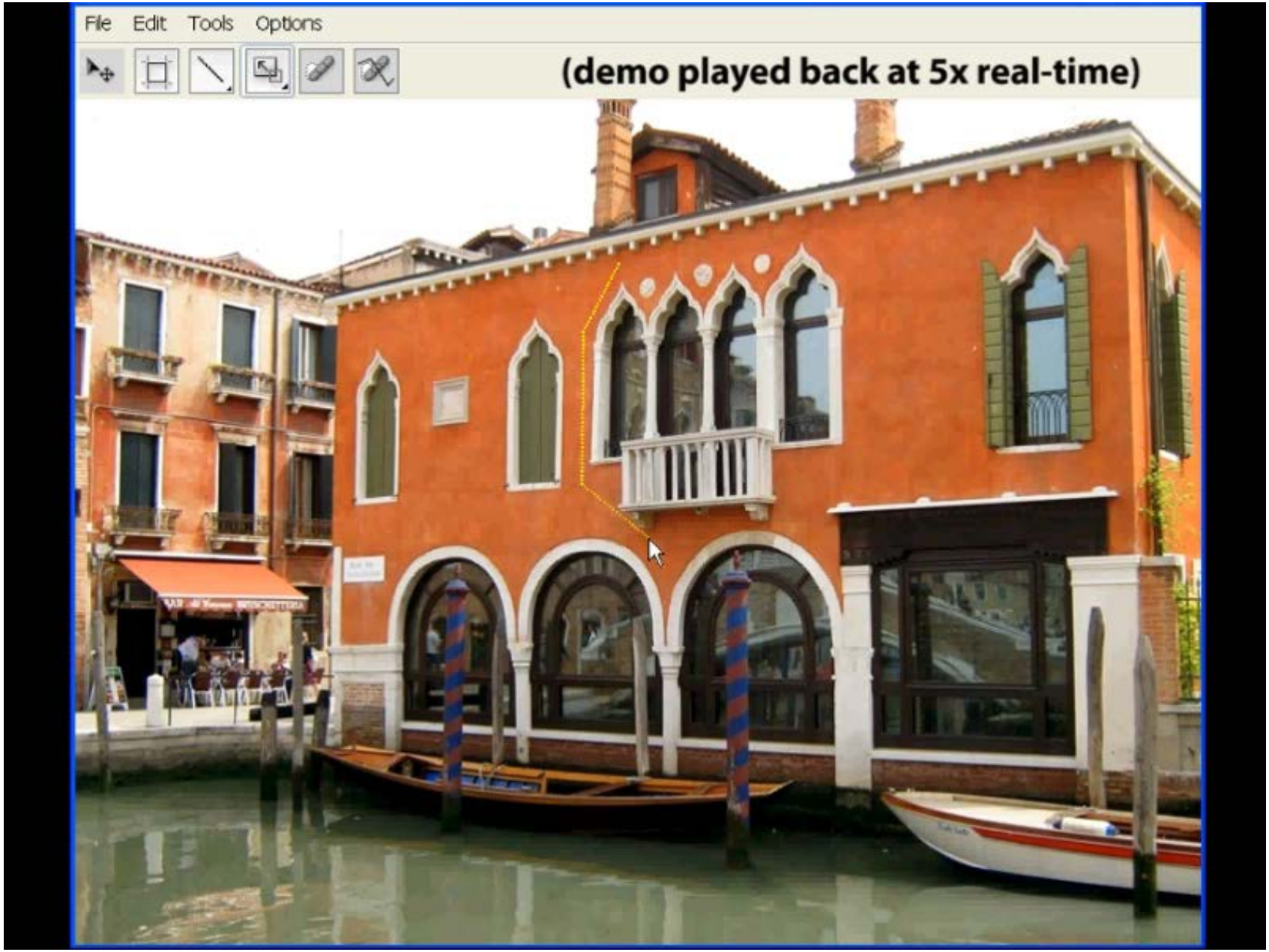
PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing
ACM Transactions on Graphics (Proc. SIGGRAPH), August 2009



PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing
ACM Transactions on Graphics (Proc. SIGGRAPH), August 2009

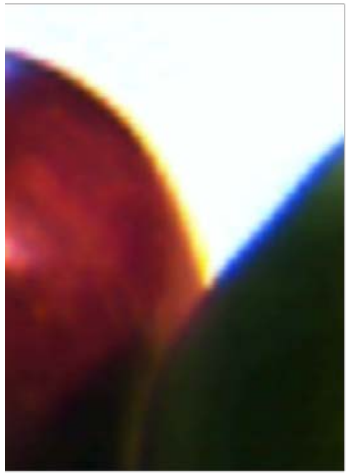


PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing
ACM Transactions on Graphics (Proc. SIGGRAPH), August 2009



PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing
ACM Transactions on Graphics (Proc. SIGGRAPH), August 2009

Image Analogies



?

Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies

Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies

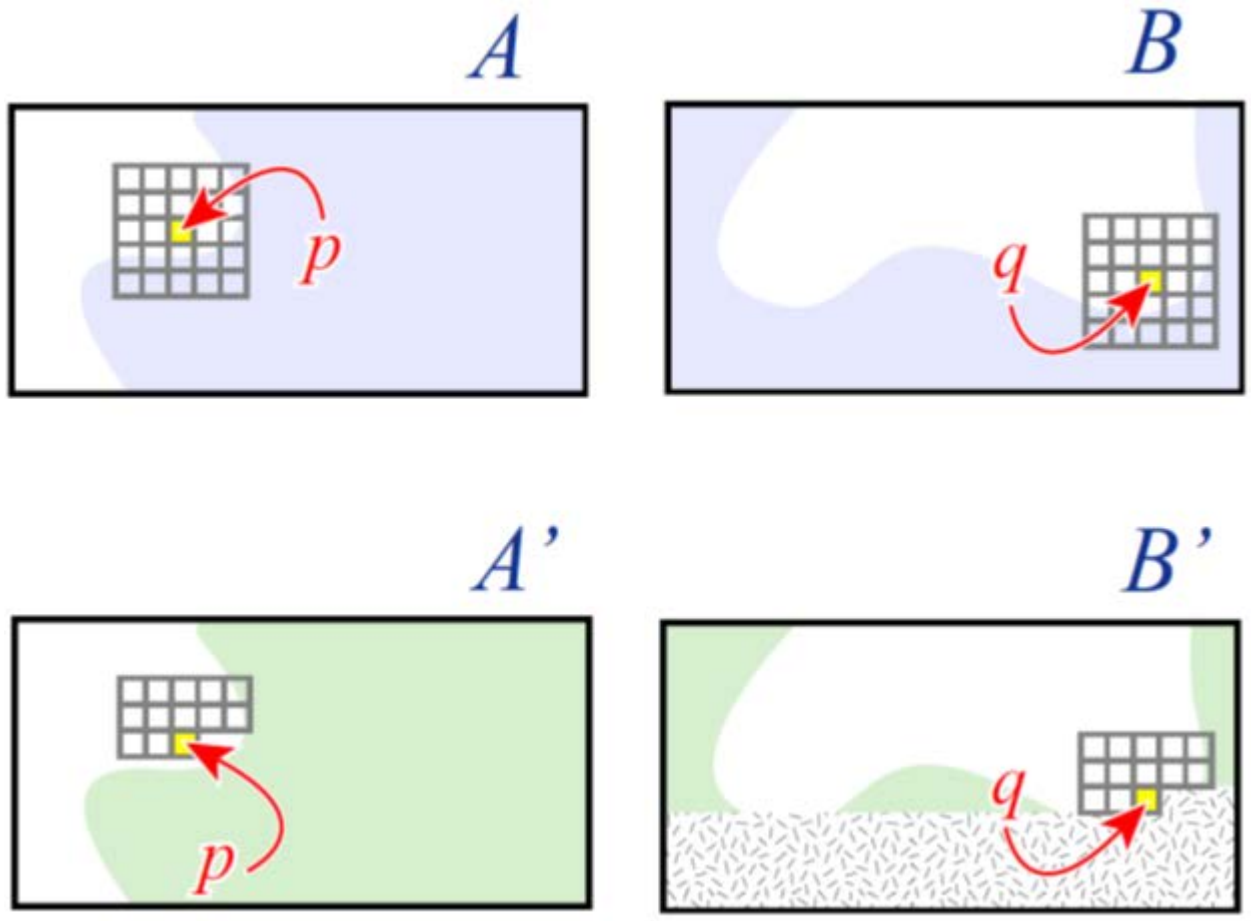


Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies

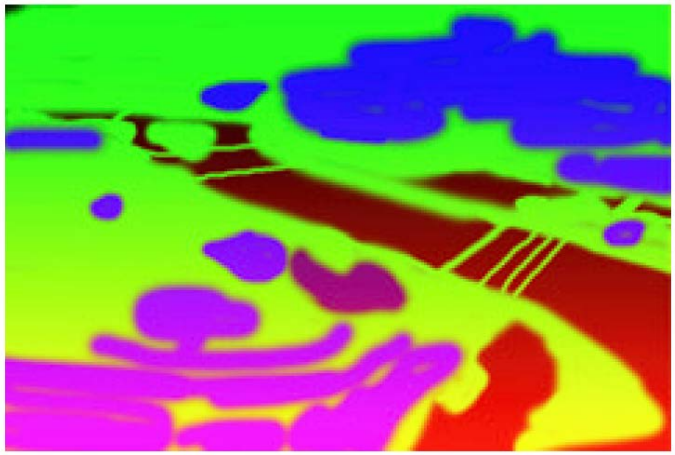
?

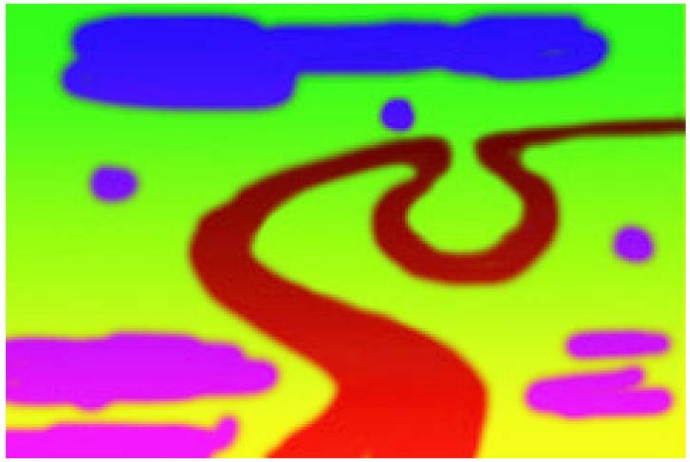
Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies

Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies

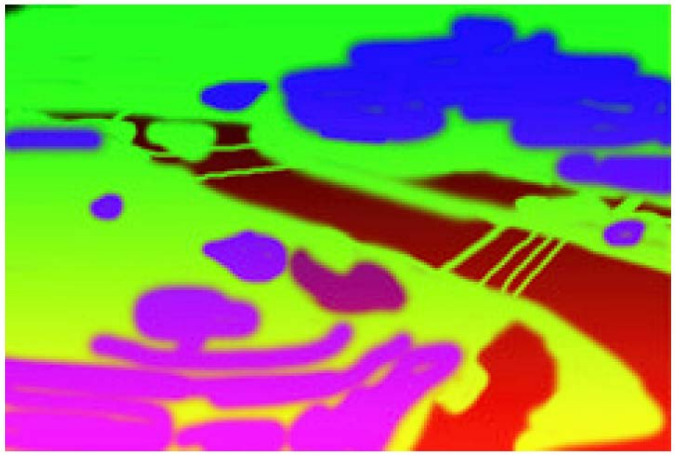




?

Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

Image Analogies



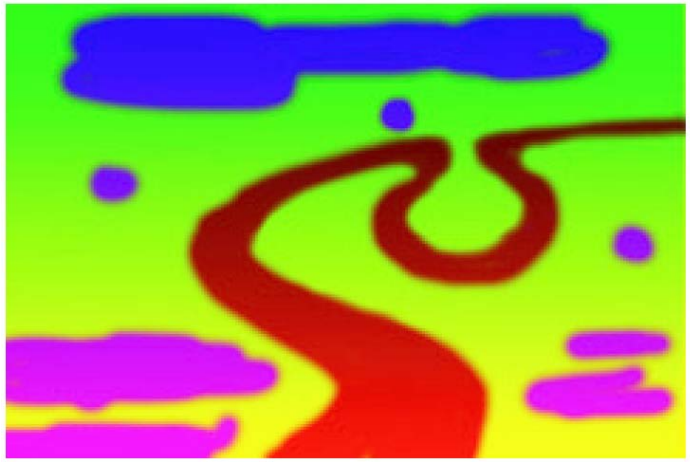
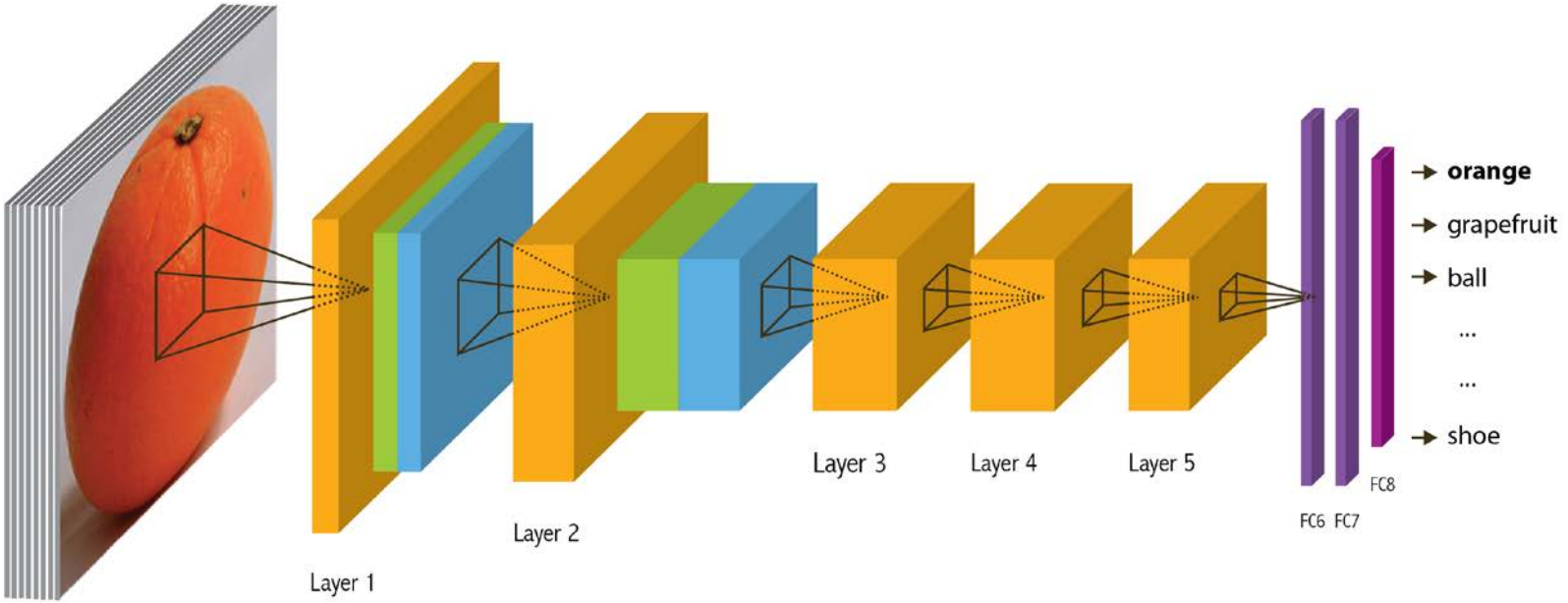


Image Analogies, A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, D. Salesin. *SIGGRAPH 2001*

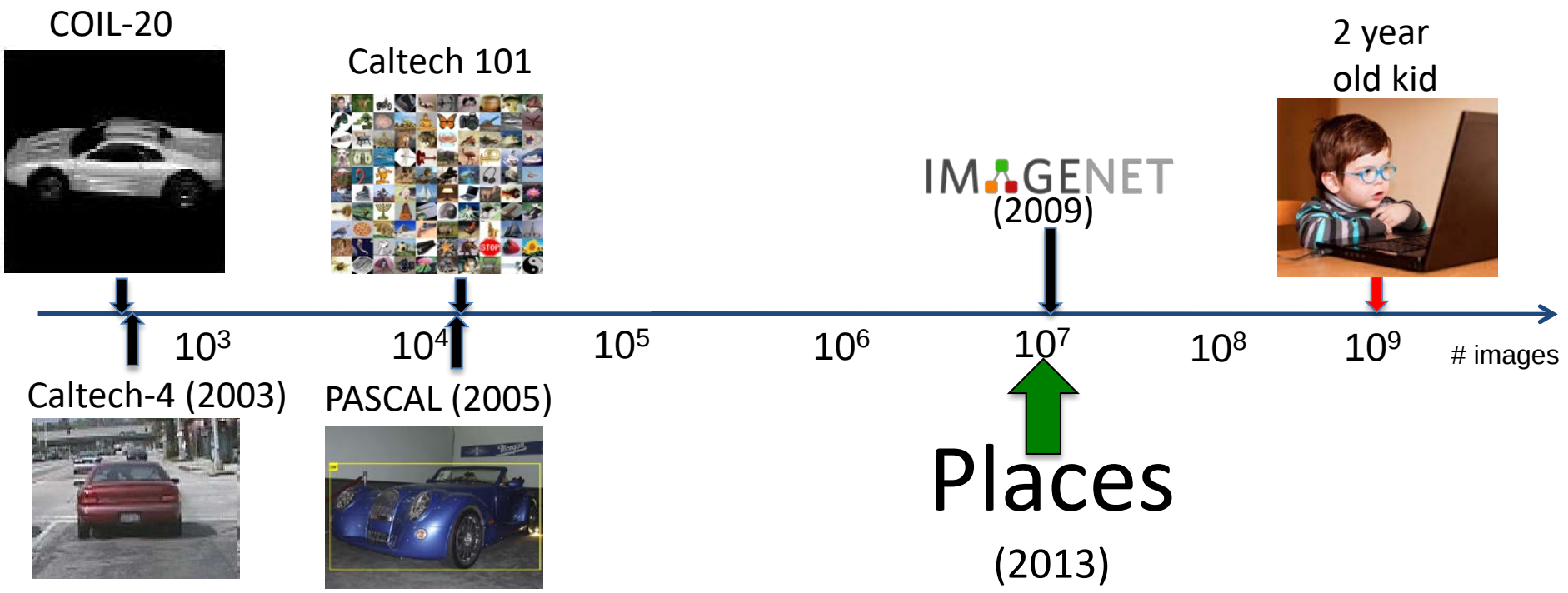
ML in Vision & Graphics Today

- Main methods: Deep Neural Networks
  - Geoffrey Hinton, Yann LeCun
  - Used both for analysis and synthesis



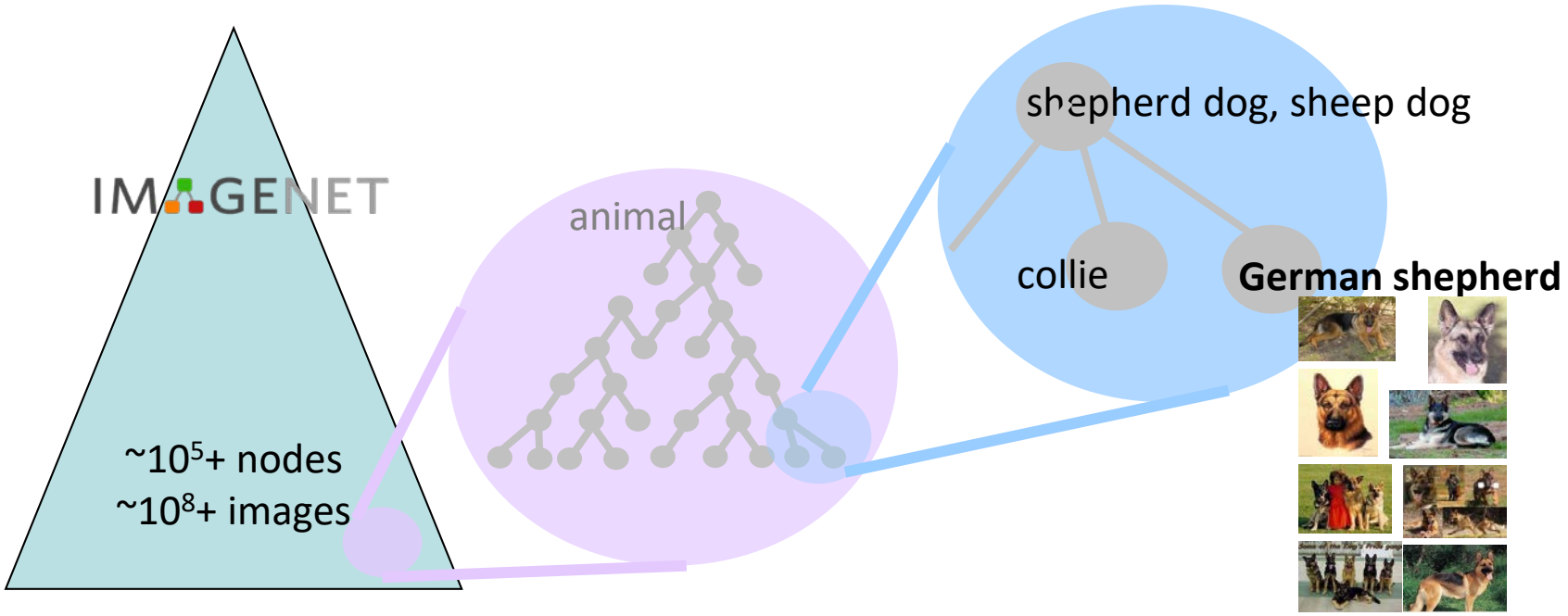
Input image

Advances in Dataset Collection



IMAGENET

- An ontology of images based on WordNet
- ImageNet currently has
  - 13,000+ categories of visual concepts
  - 10 million human-cleaned images (~700im/categ)
  - 1/3+ is released online @ www.image-net.org



Deng, Dong, Socher, Li & Fei-Fei, CVPR 2009

Application to ImageNet

[Deng et al. CVPR 2009]

- ~14 million labeled images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk

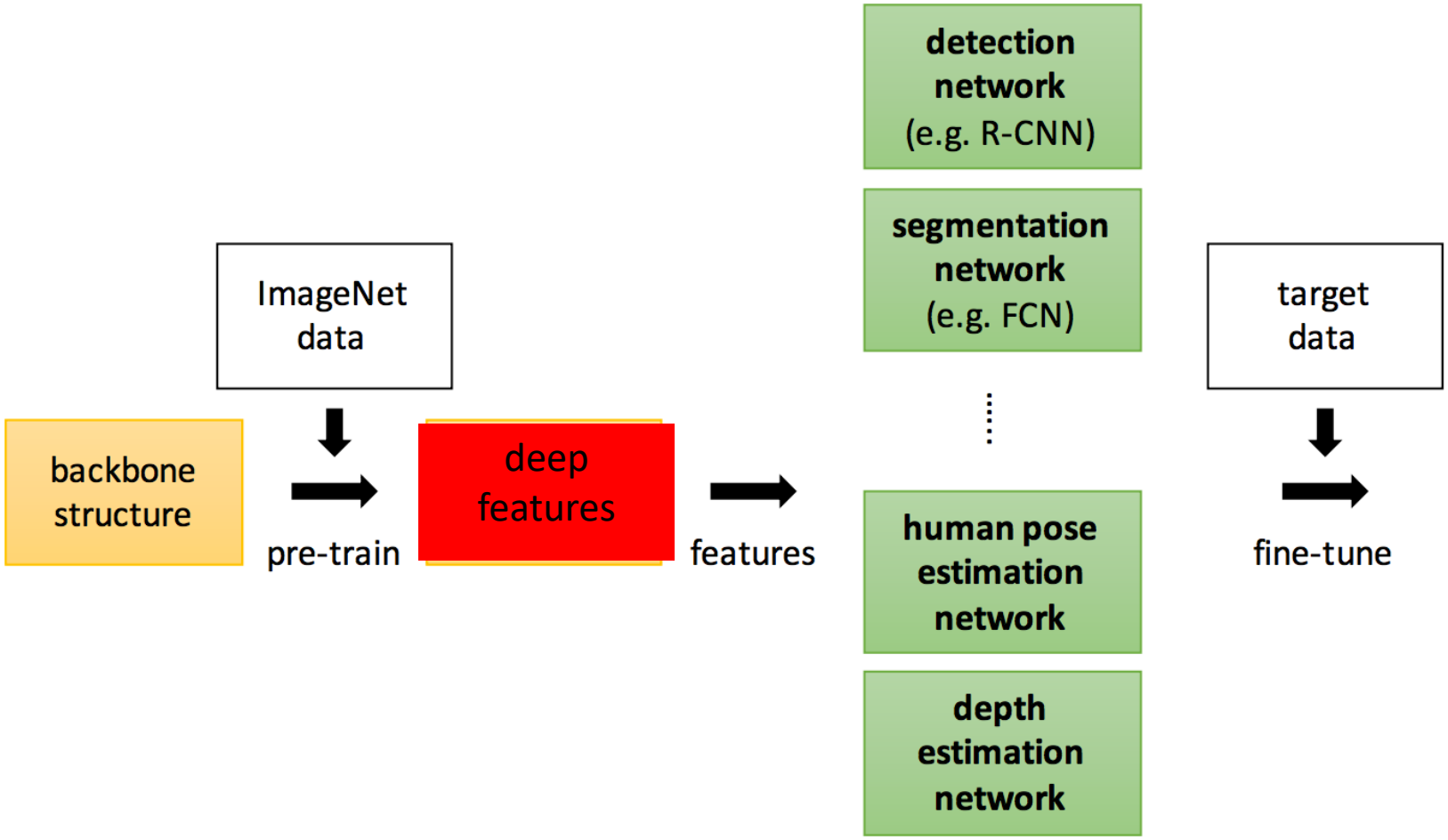
ImageNet Classification with Deep Convolutional Neural Networks [NIPS 2012]

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

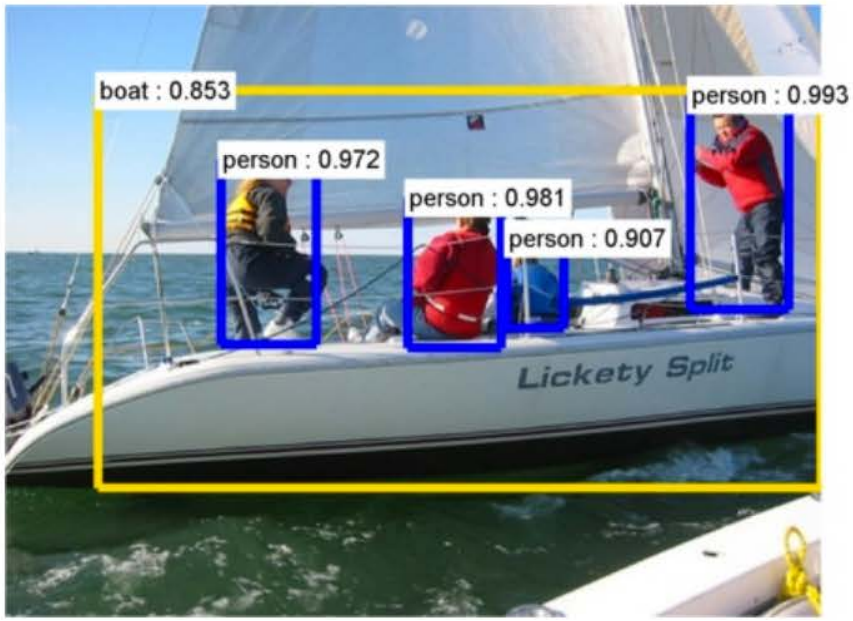
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Solving Different Tasks



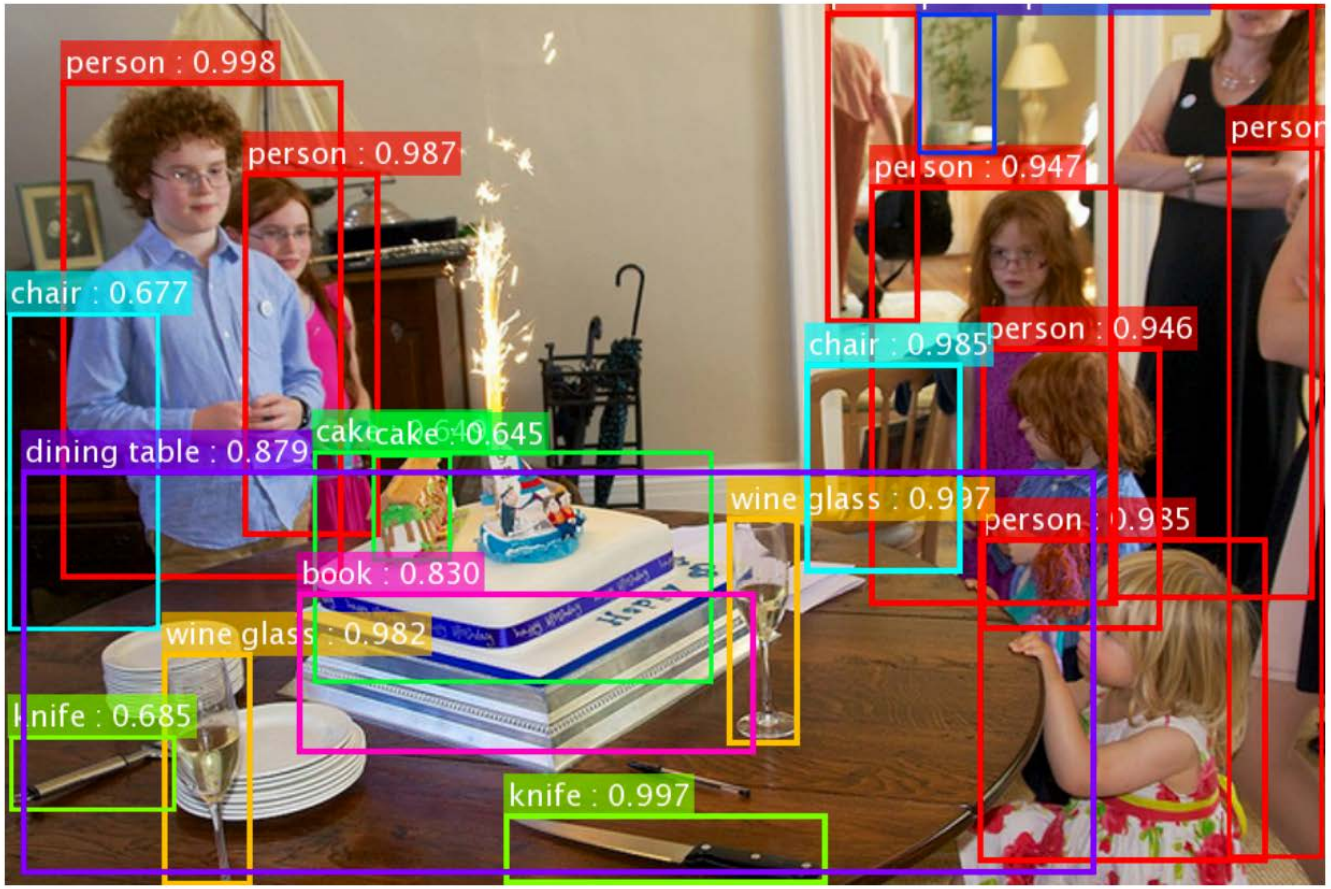
Slide Credit: http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

Object Detection: What and Where



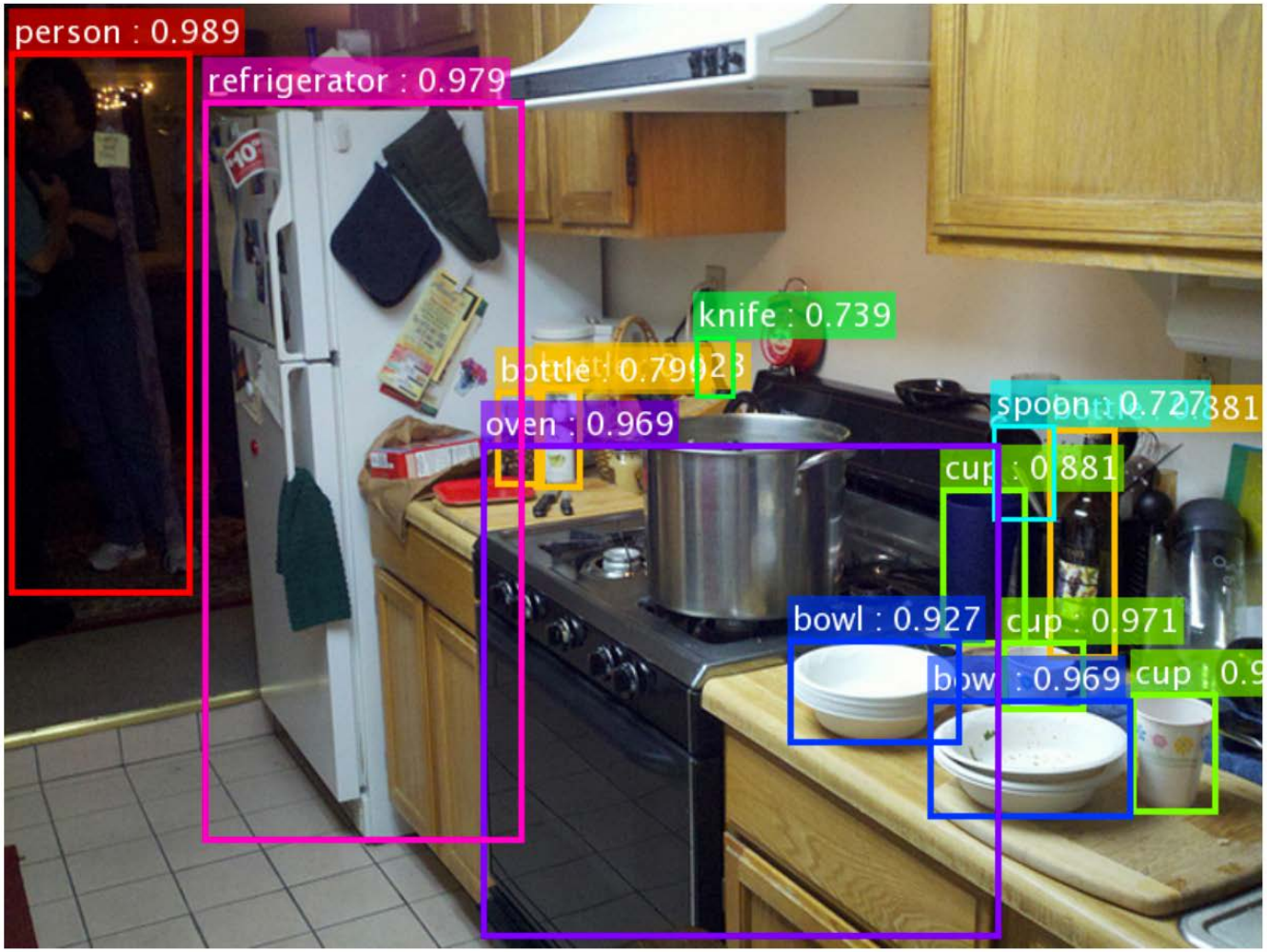
Slide Credit: http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

Object Detection Results



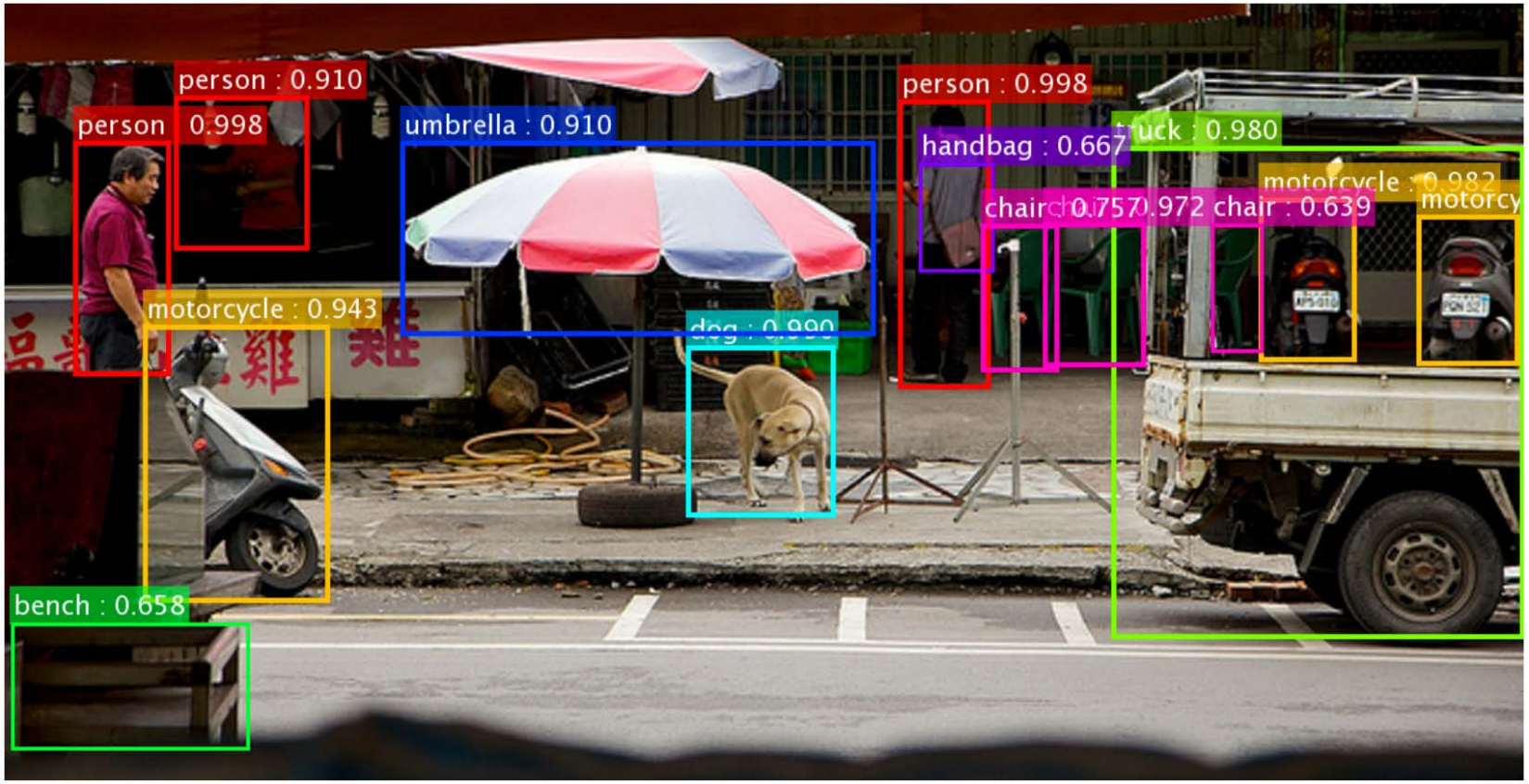
Slide Credit: http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

Object Detection Results



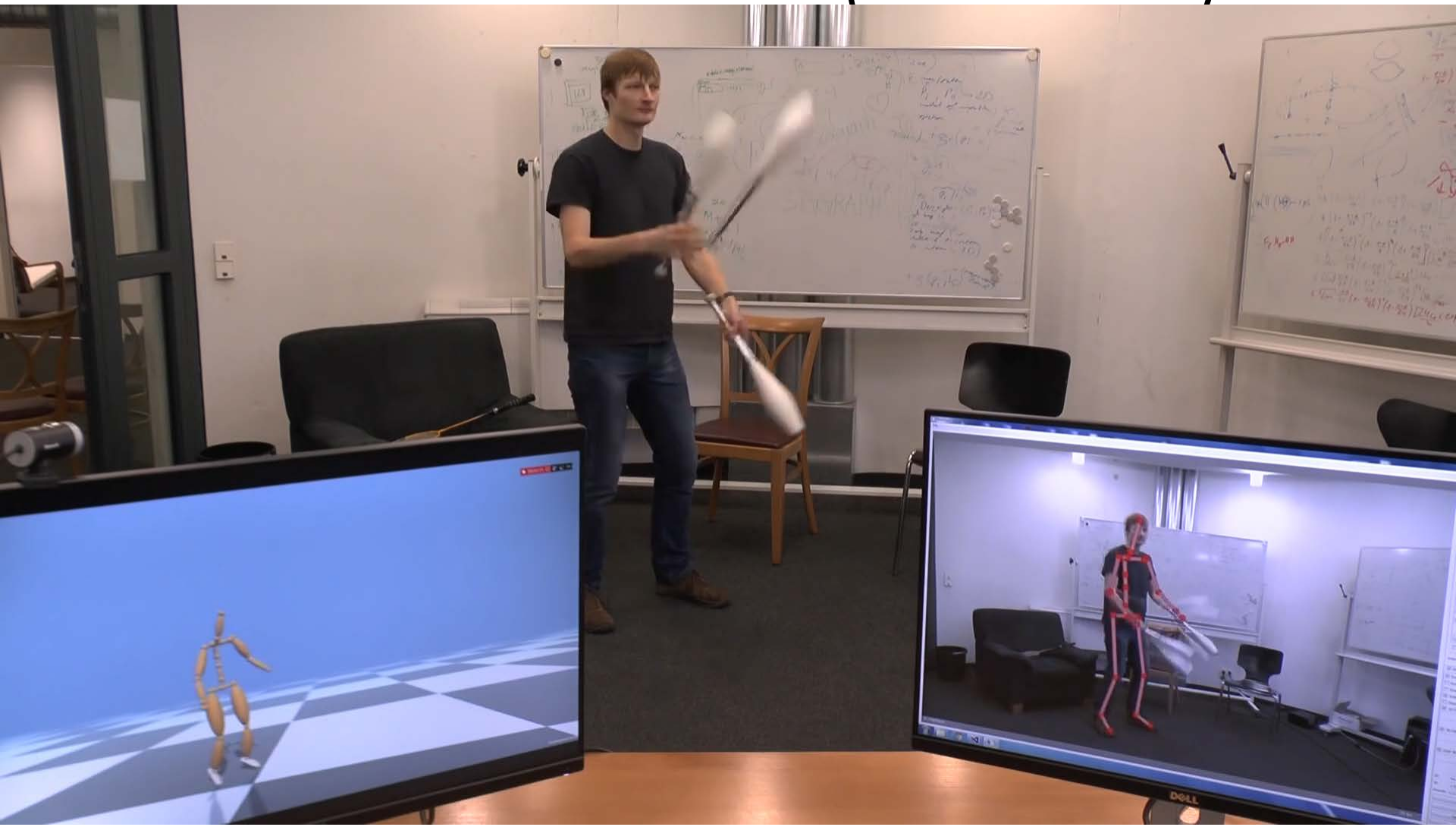
Slide Credit: http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

Object Detection Results



Slide Credit: http://kaiminghe.com/icml16tutorial/icml2016_tutorial_deep_residual_networks_kaiminghe.pdf

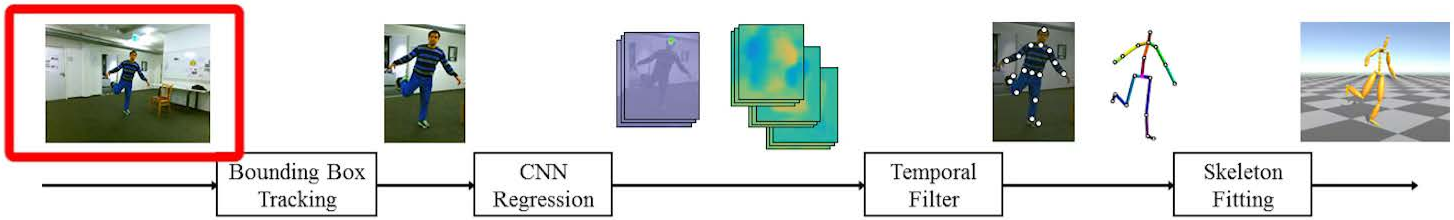
Pose Estimation (2D & 3D)



VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera – SIGGRAPH 2017

https://www.youtube.com/watch?v=W1ZNFfftx2E

Pose Estimation (2D & 3D)

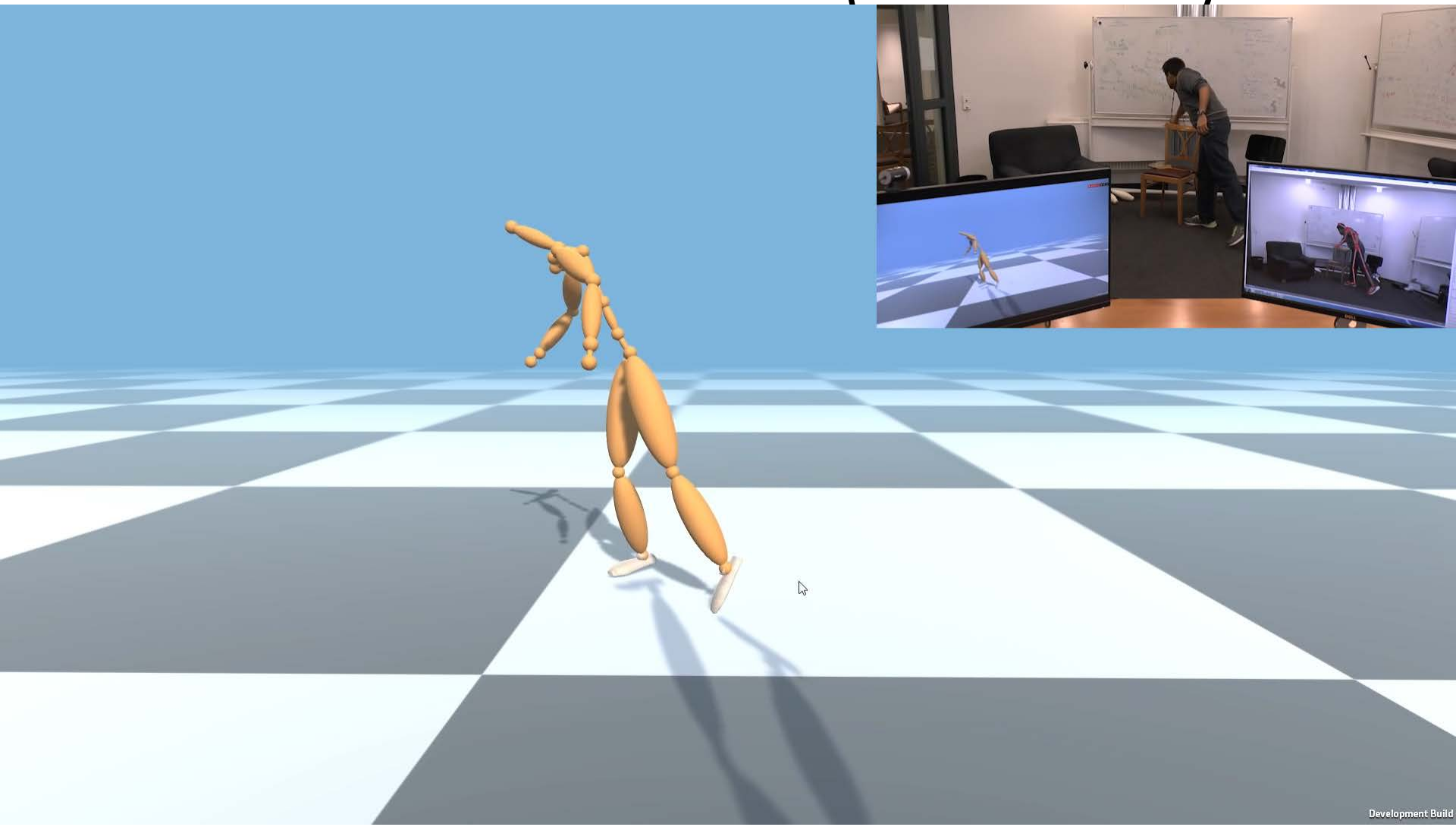


Full-frame Input from Single RGB Camera

VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera – SIGGRAPH 2017

https://www.youtube.com/watch?v=W1ZNFfftx2E

Pose Estimation (2D & 3D)



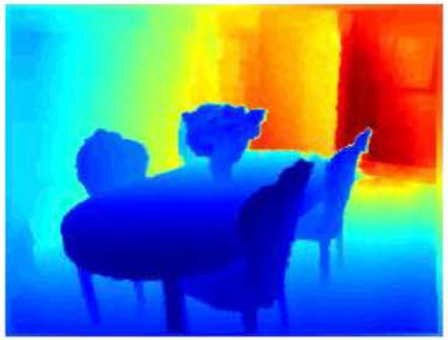
VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera – SIGGRAPH 2017

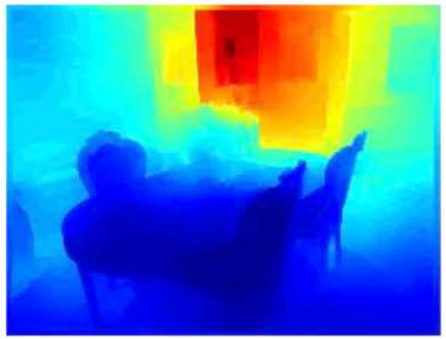
https://www.youtube.com/watch?v=W1ZNFfftx2E

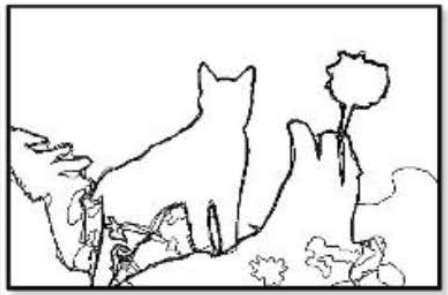
Segmentation, Depth, & More

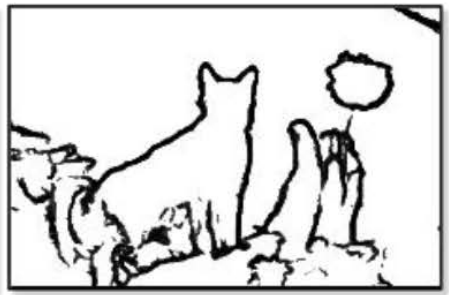
semantic segmentation

monocular depth estimation (Liu et al. 2015)





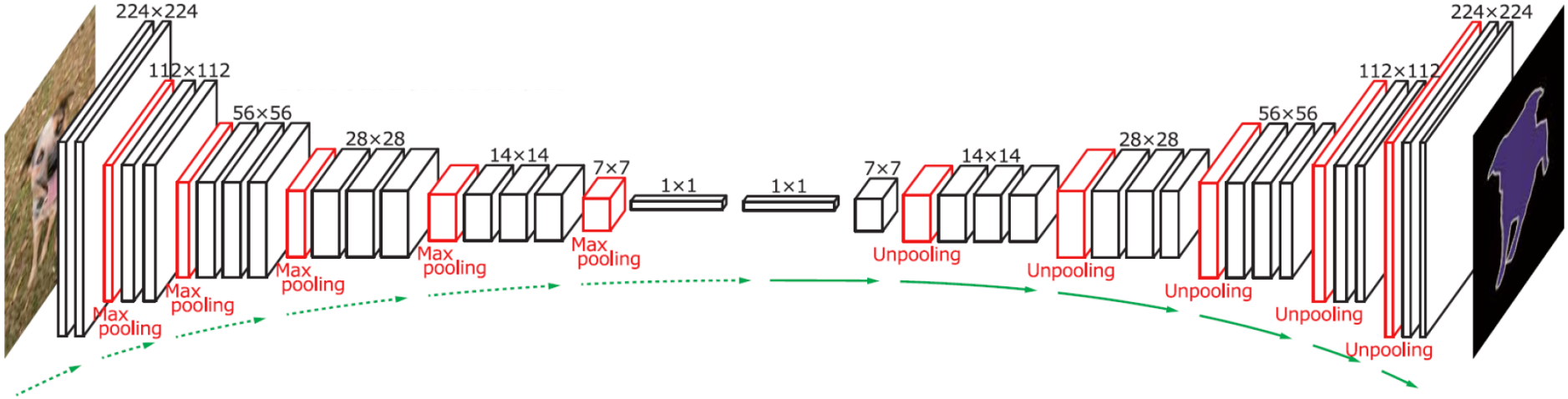




boundary prediction (Xie & Tu 2015)

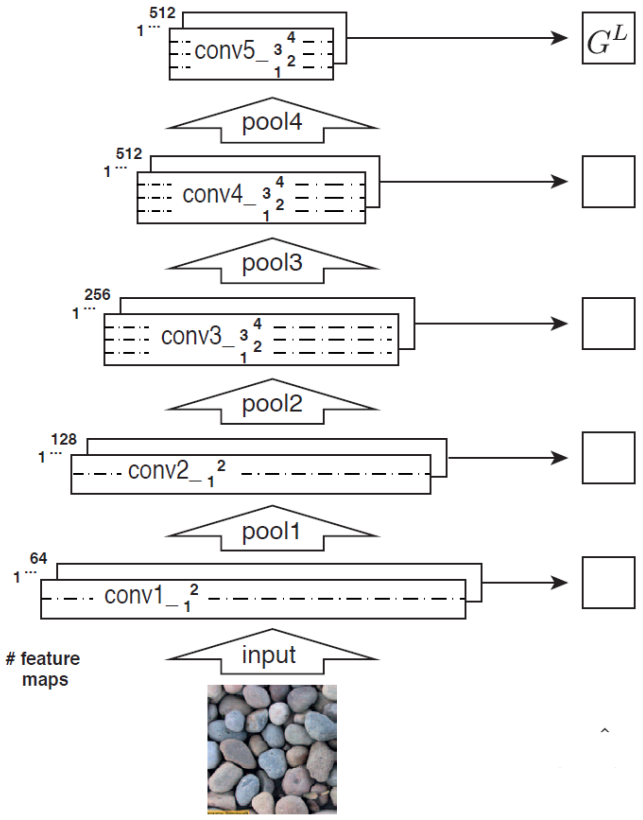
Slide Credit: https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Typical Architecture



H. Noh et.al. Learning Deconvolution Network for Semantic Segmentation, ICCV 2015

Texture Synthesis using CNNs

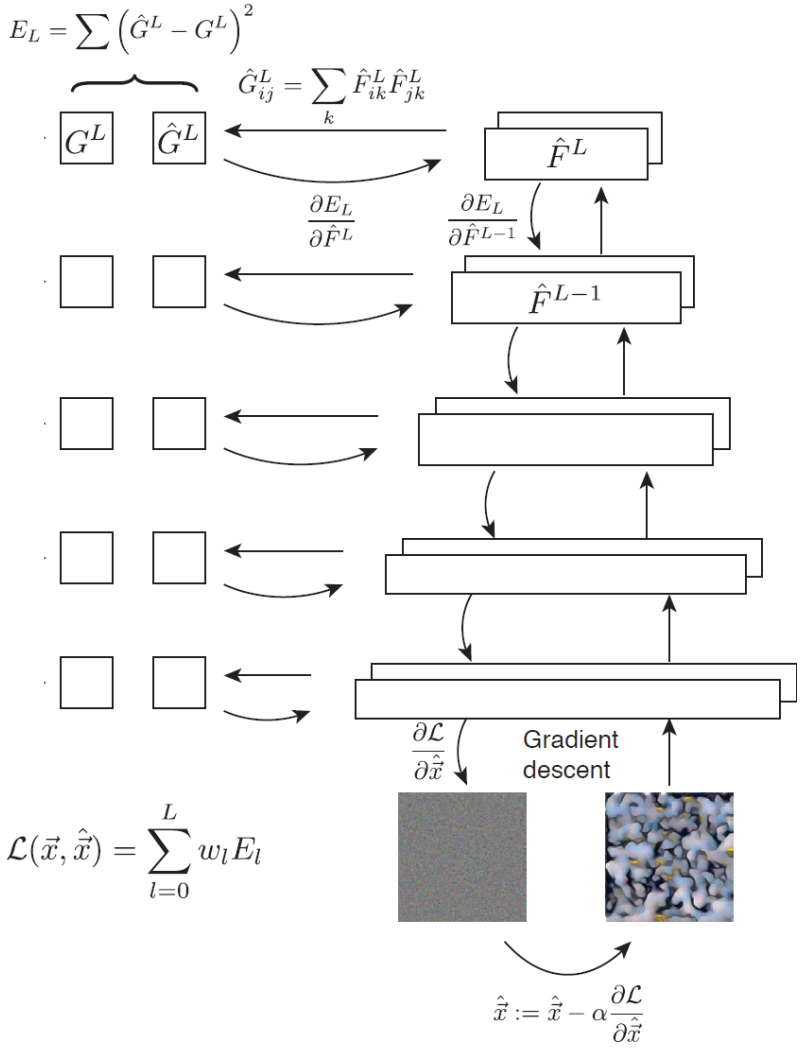


Analysis

Gatys et al., Texture Synthesis Using Convolutional Neural Networks, NIPS 2015

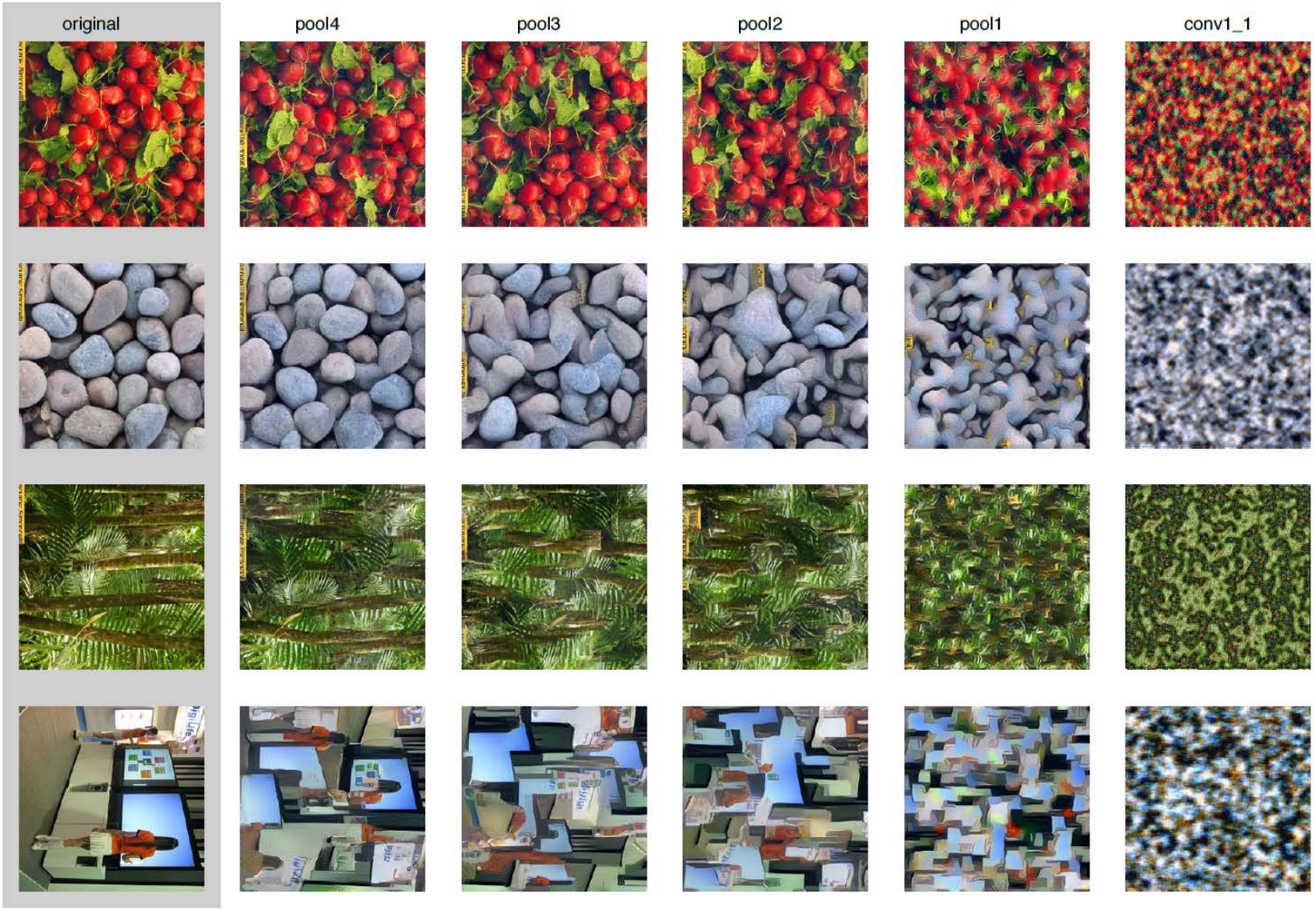
Texture Synthesis using CNNs

Synthesis



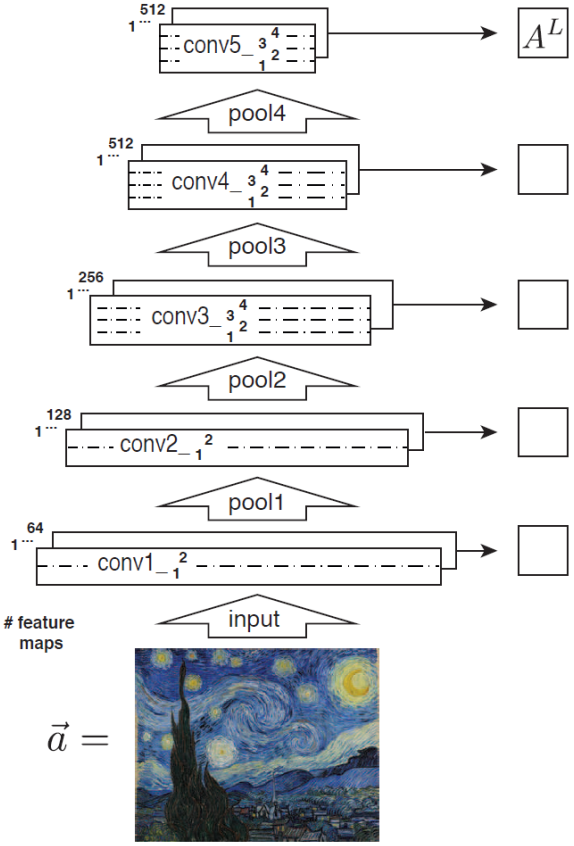
Gatys et al., Texture Synthesis Using Convolutional Neural Networks, NIPS 2015

Texture Synthesis using CNNs



Gatys et al., Texture Synthesis Using Convolutional Neural Networks, NIPS 2015

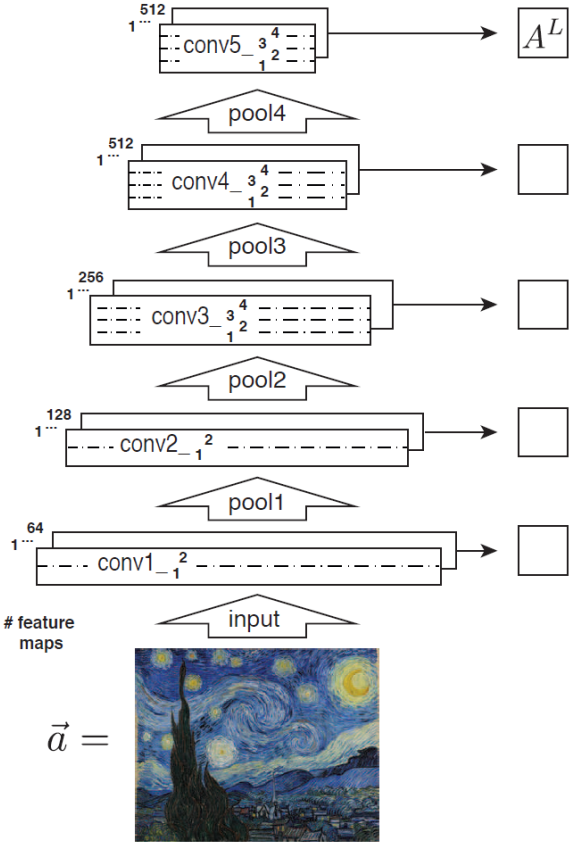
Image Analogies/Style Transfer



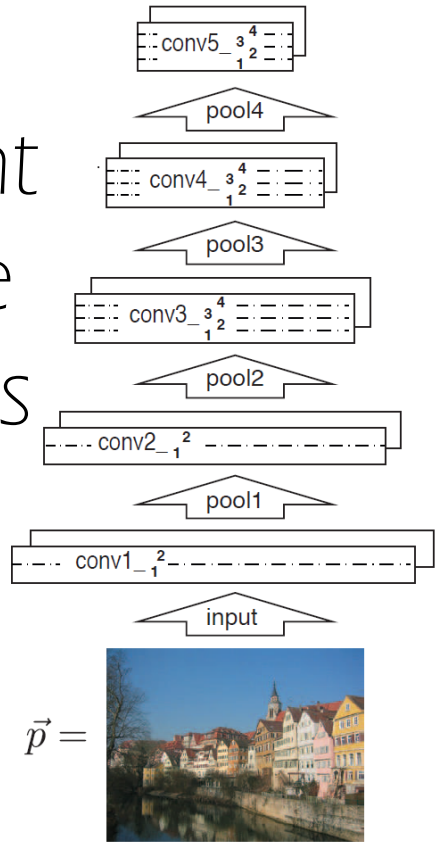
Style Source Analysis

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image Analogies/Style Transfer

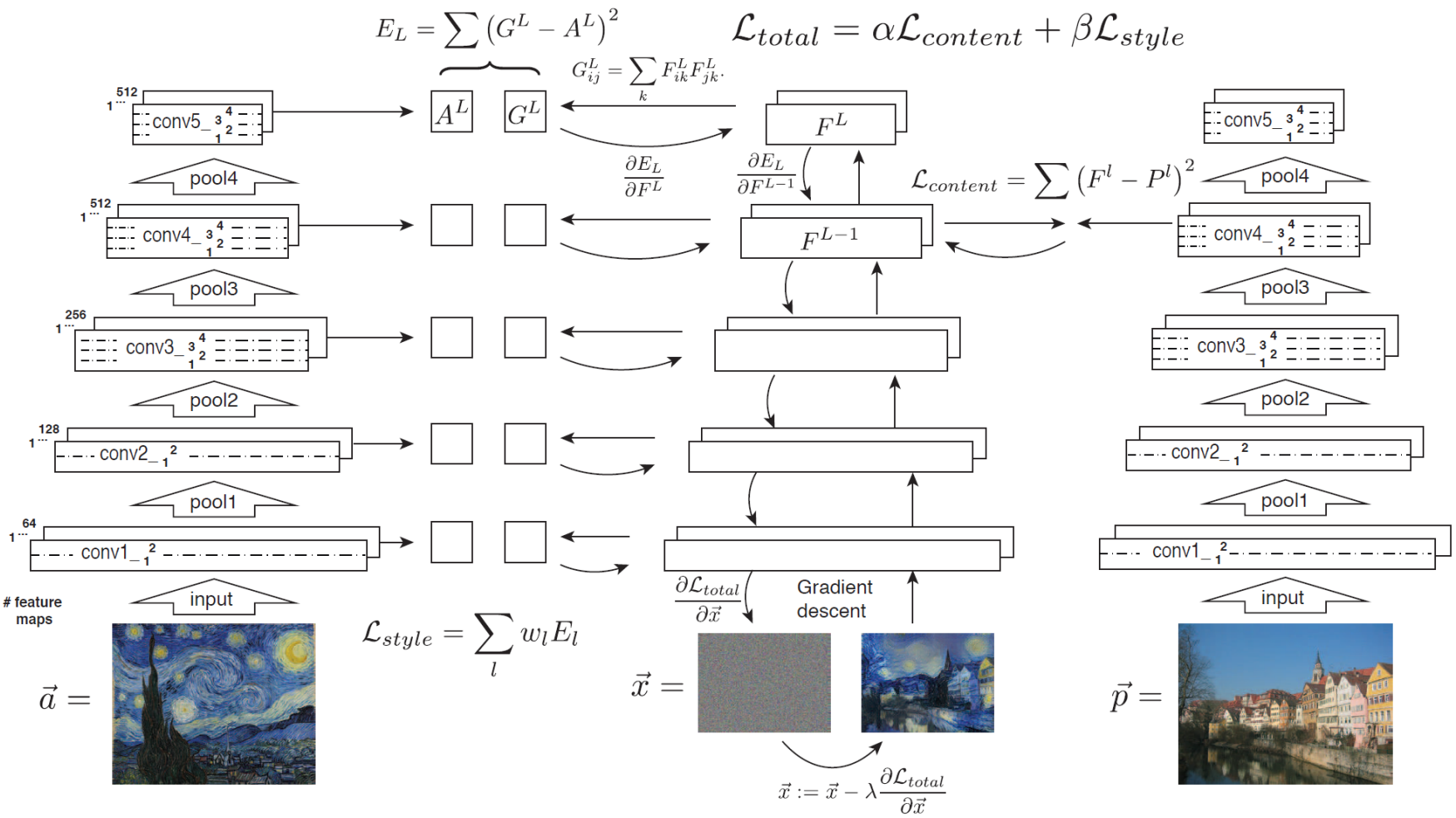


Content Source Analysis



Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image Analogies/Style Transfer



Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image Analogies/Style Transfer

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image Analogies/Style Transfer

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

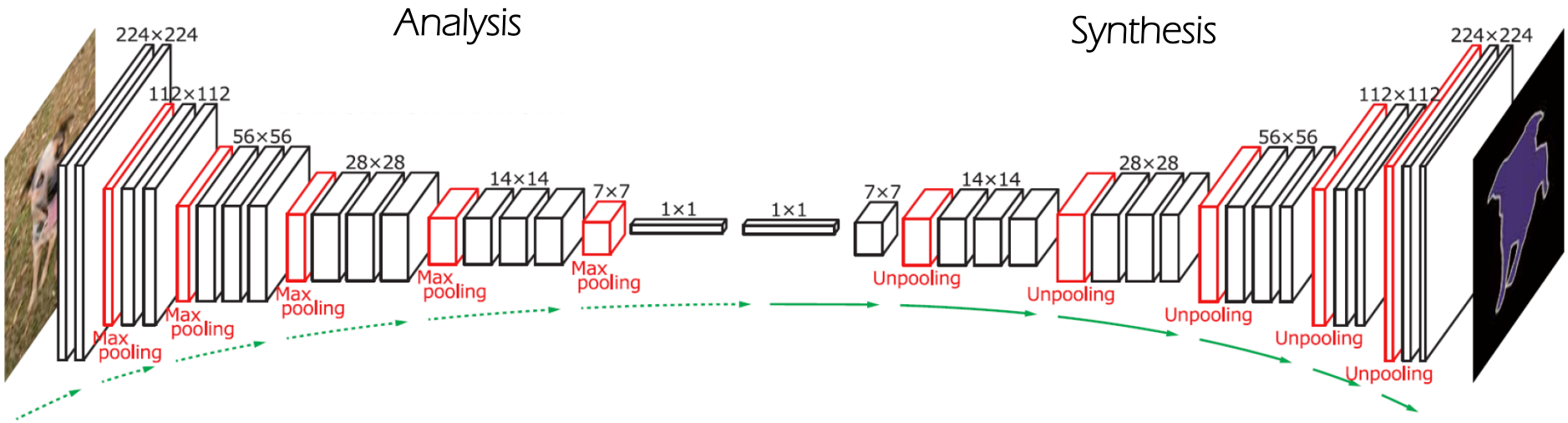
Image Analogies/Style Transfer

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

Image Analogies/Style Transfer

Gatys et al., Image Style Transfer Using Convolutional Neural Networks, CVPR 2016

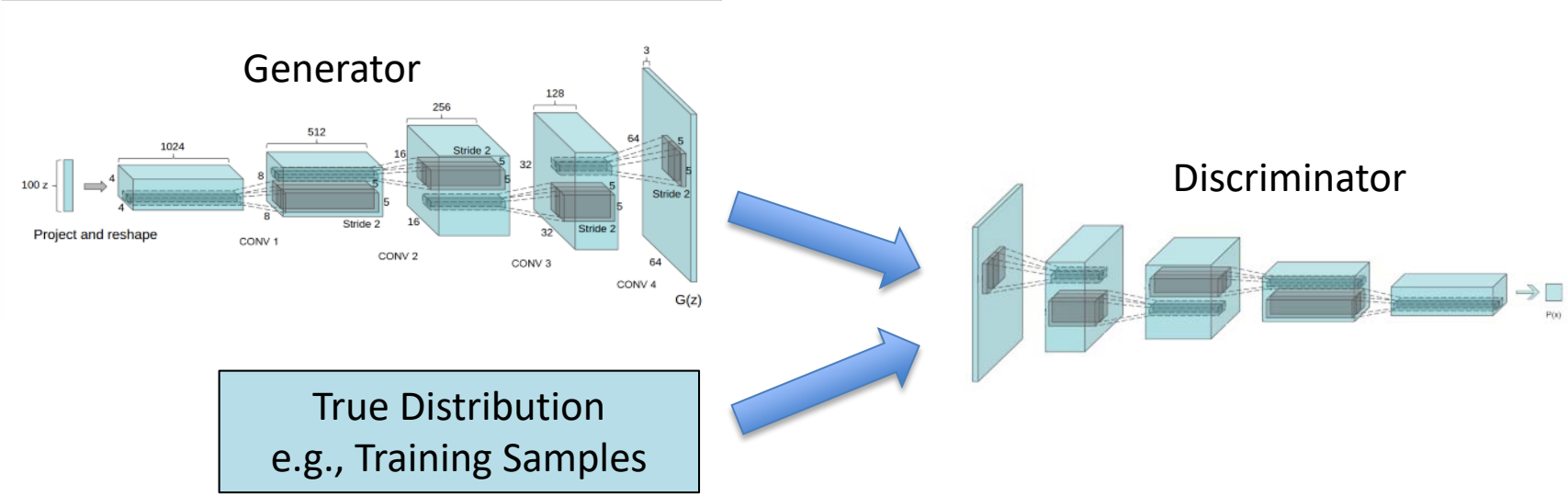
Recall: Typical Architecture



H. Noh et.al. Learning Deconvolution Network for Semantic Segmentation, ICCV 2015

Generative Adversarial Networks (GANs)

- Two neural networks competing against each other
  - Generator is taught to map to a desired distribution
  - Discriminator is taught to discriminate between data from true distribution and generator



GANs for Image Synthesis

- Automatically generated bedrooms

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Alec Radford, Luke Metz, Soumith Chintala, *ICLR 2016*

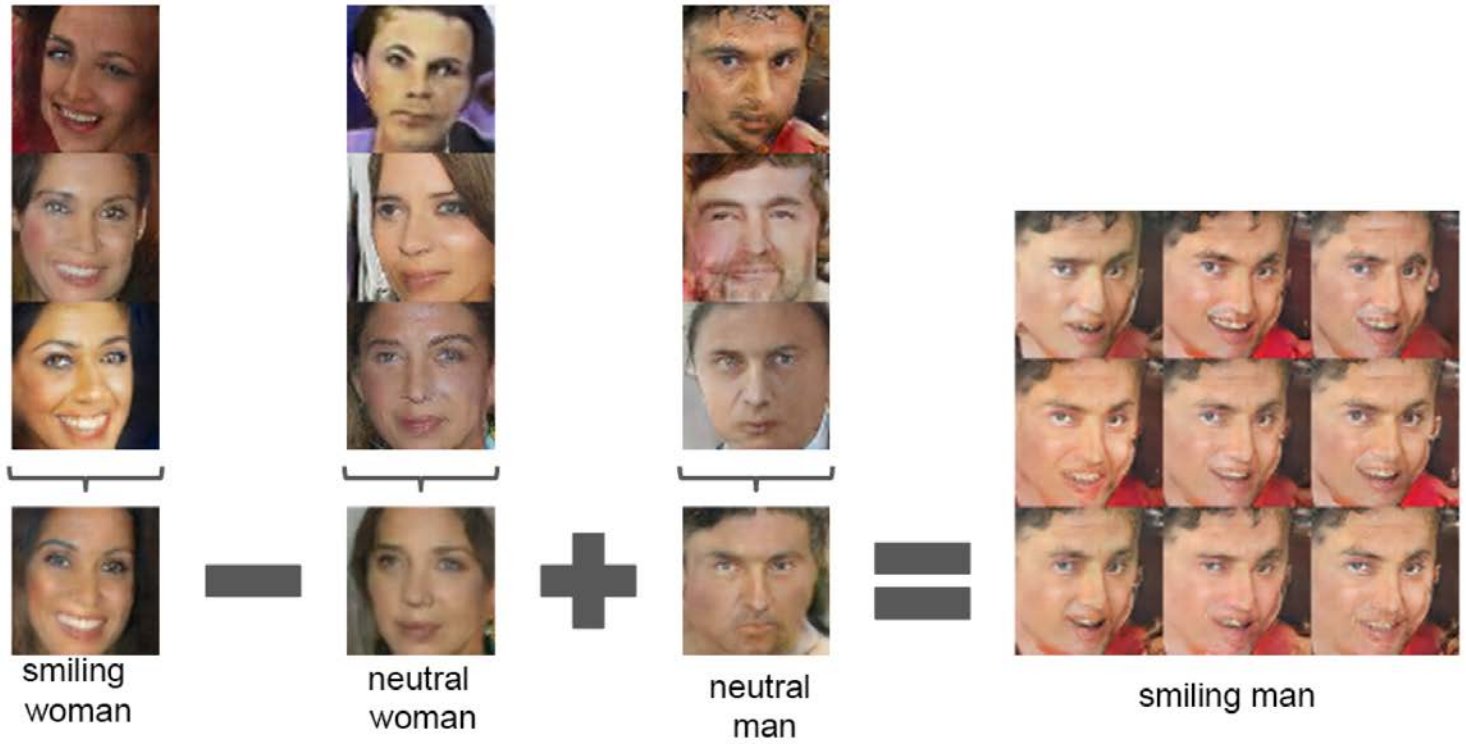
GANs for Image Synthesis

- Interpolation

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Alec Radford, Luke Metz, Soumith Chintala, *ICLR 2016*

GANs for Image Synthesis

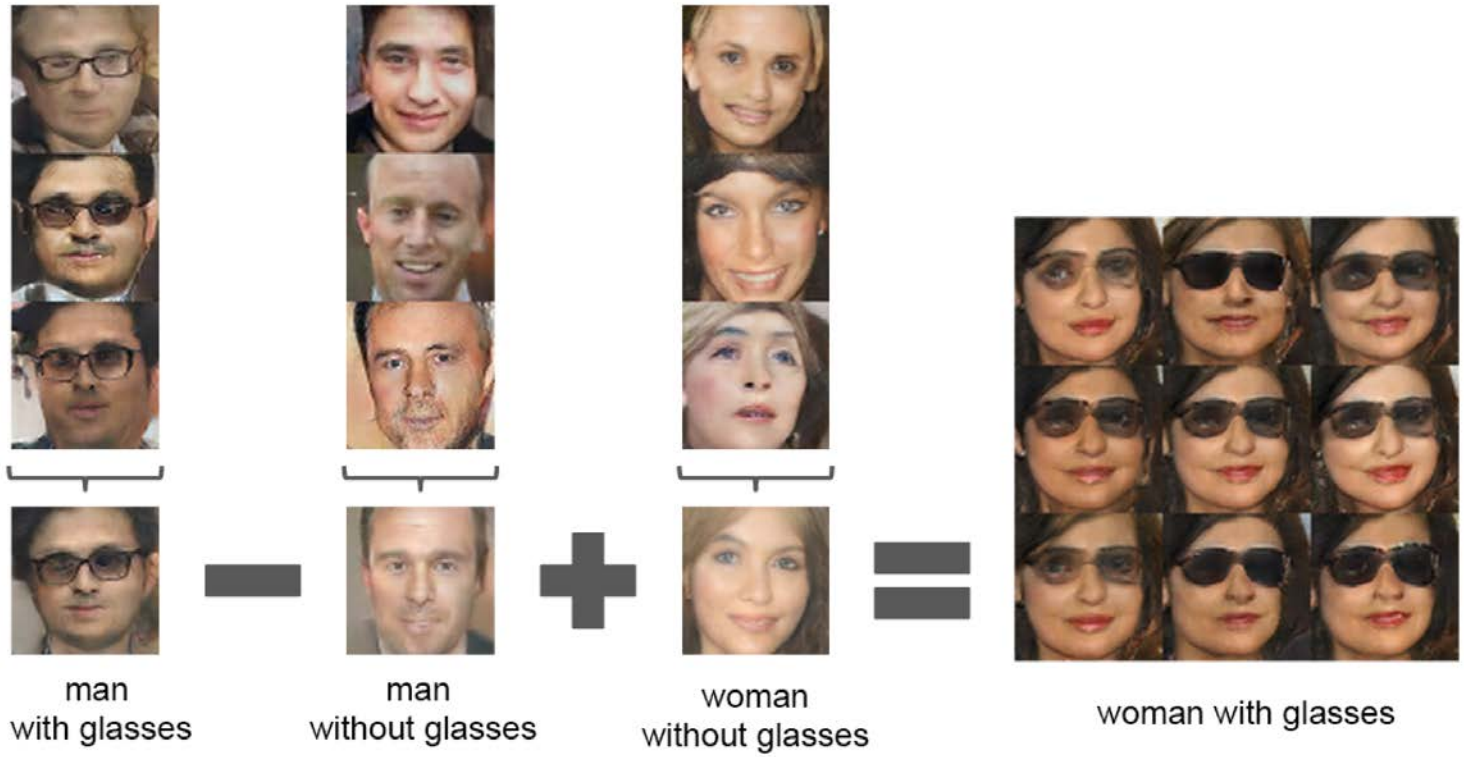
- *Image Arithmetic*



Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Alec Radford, Luke Metz, Soumith Chintala, *ICLR 2016*

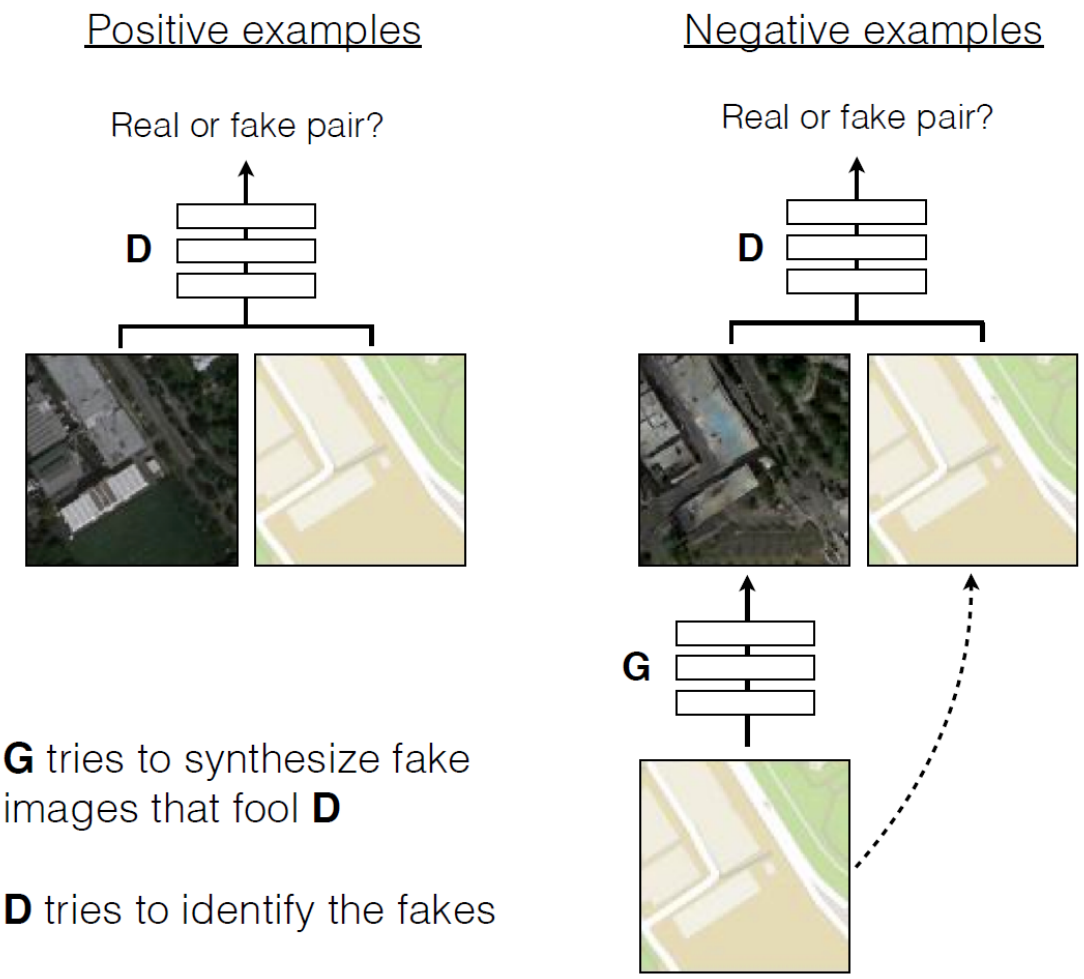
GANs for Image Synthesis

- *Image Arithmetic*



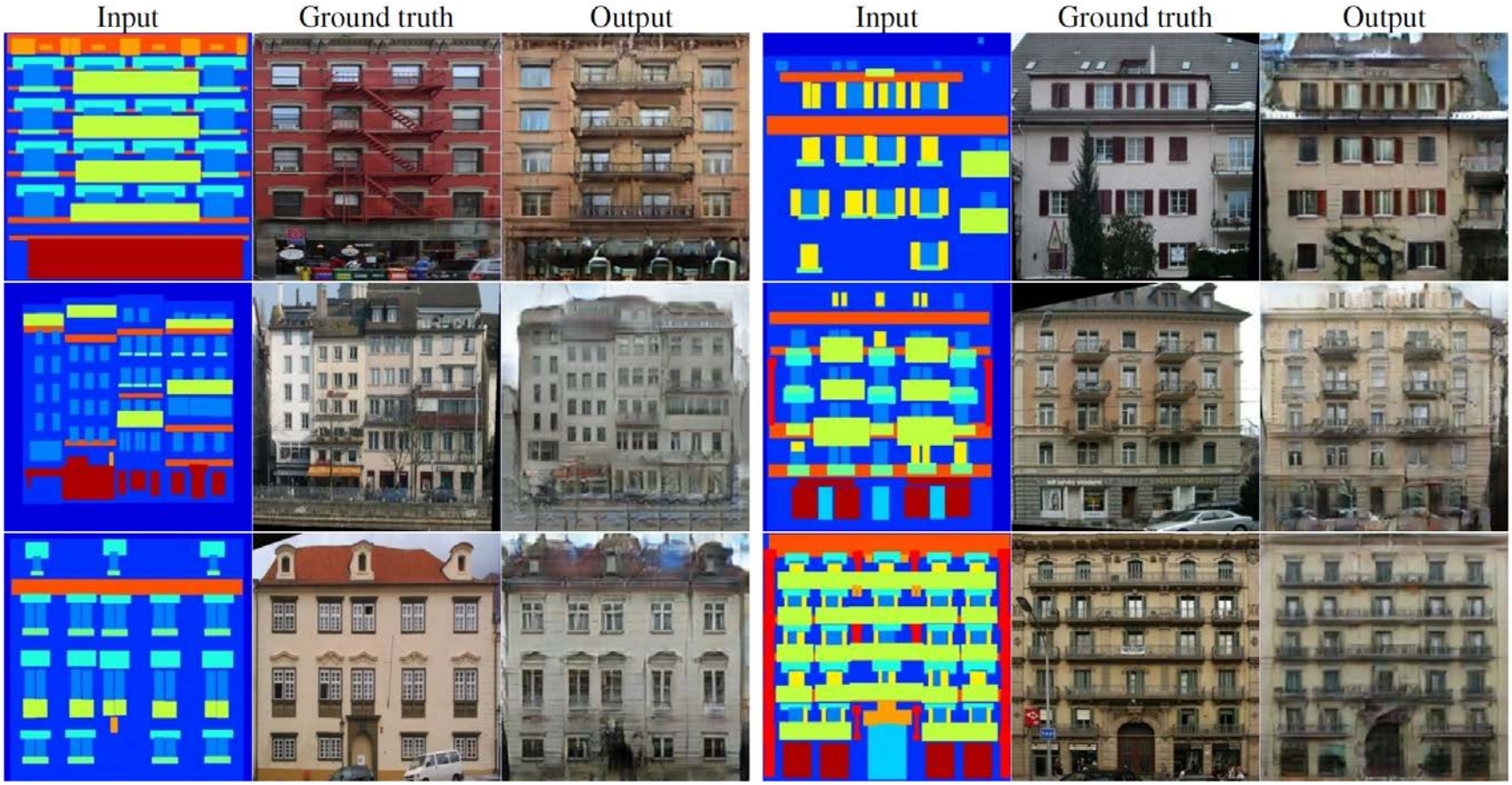
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Alec Radford, Luke Metz, Soumith Chintala, *ICLR 2016*

Conditional GANs/Image Translation

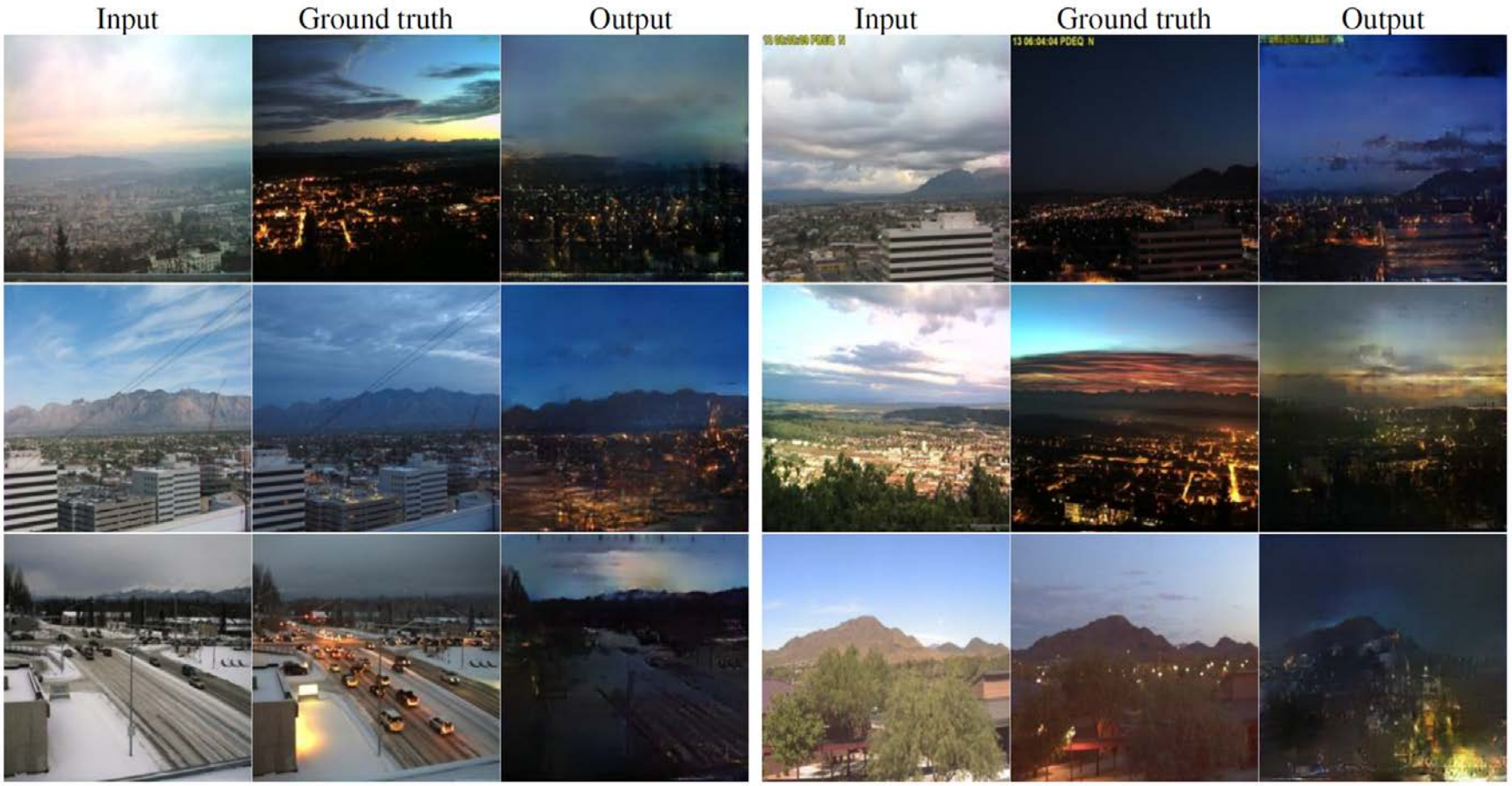


Conditional GANs/Image Translation

Conditional GANs/Image Translation



Conditional GANs/Image Translation

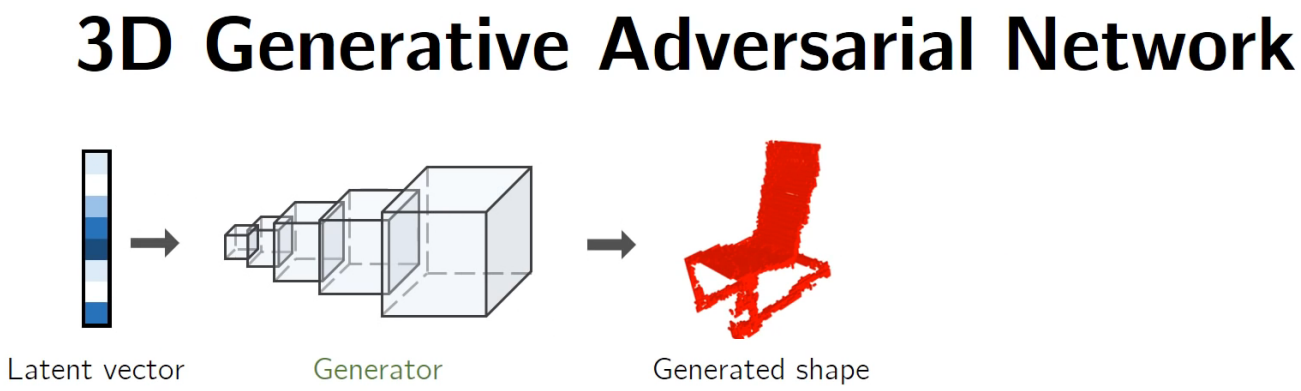


Conditional GANs/Image Translation

Conditional GANs/Image Translation

More than Images

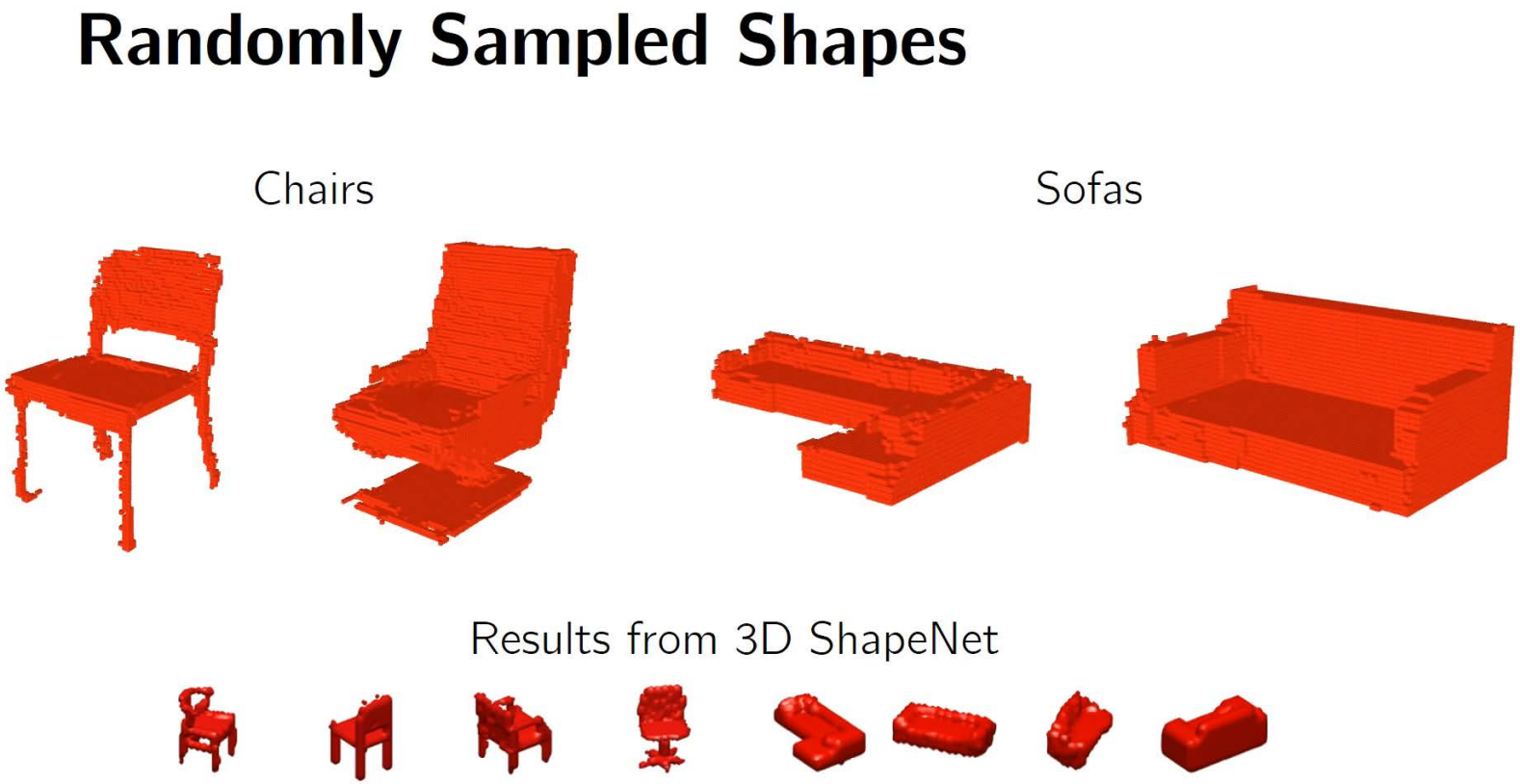
- Shape Synthesis



Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, NIPS 2016

More than Images

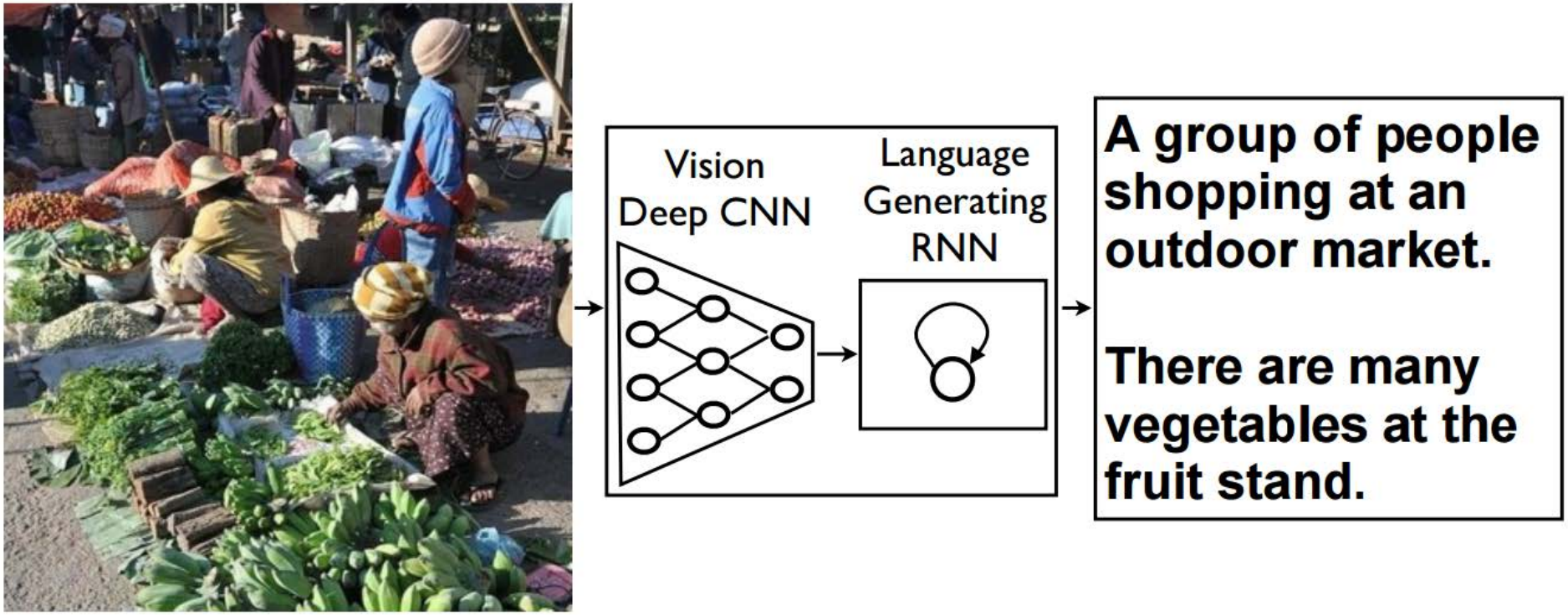
- Shape Synthesis



Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, NIPS 2016

More than Images

- From Image to Captions



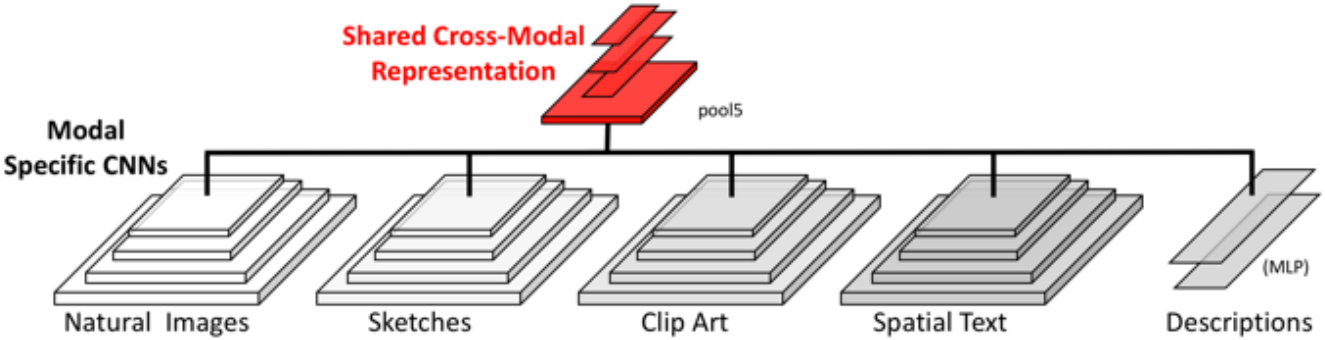
Show and Tell: A Neural Image Caption Generator, Vinyals et. al., CVPR 2015

Image Caption Generator Results

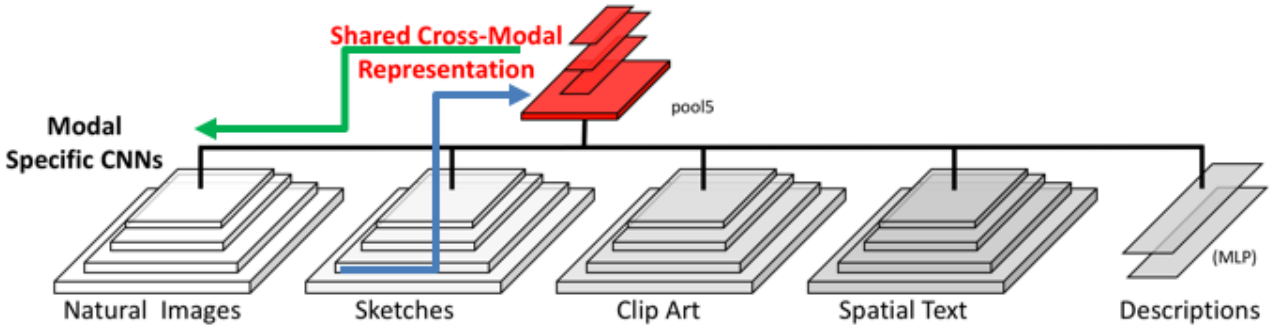
https://pdollar.wordpress.com/2015/01/21/image-captioning/

Show and Tell: A Neural Image Caption Generator, Vinyals et. al., CVPR 2015

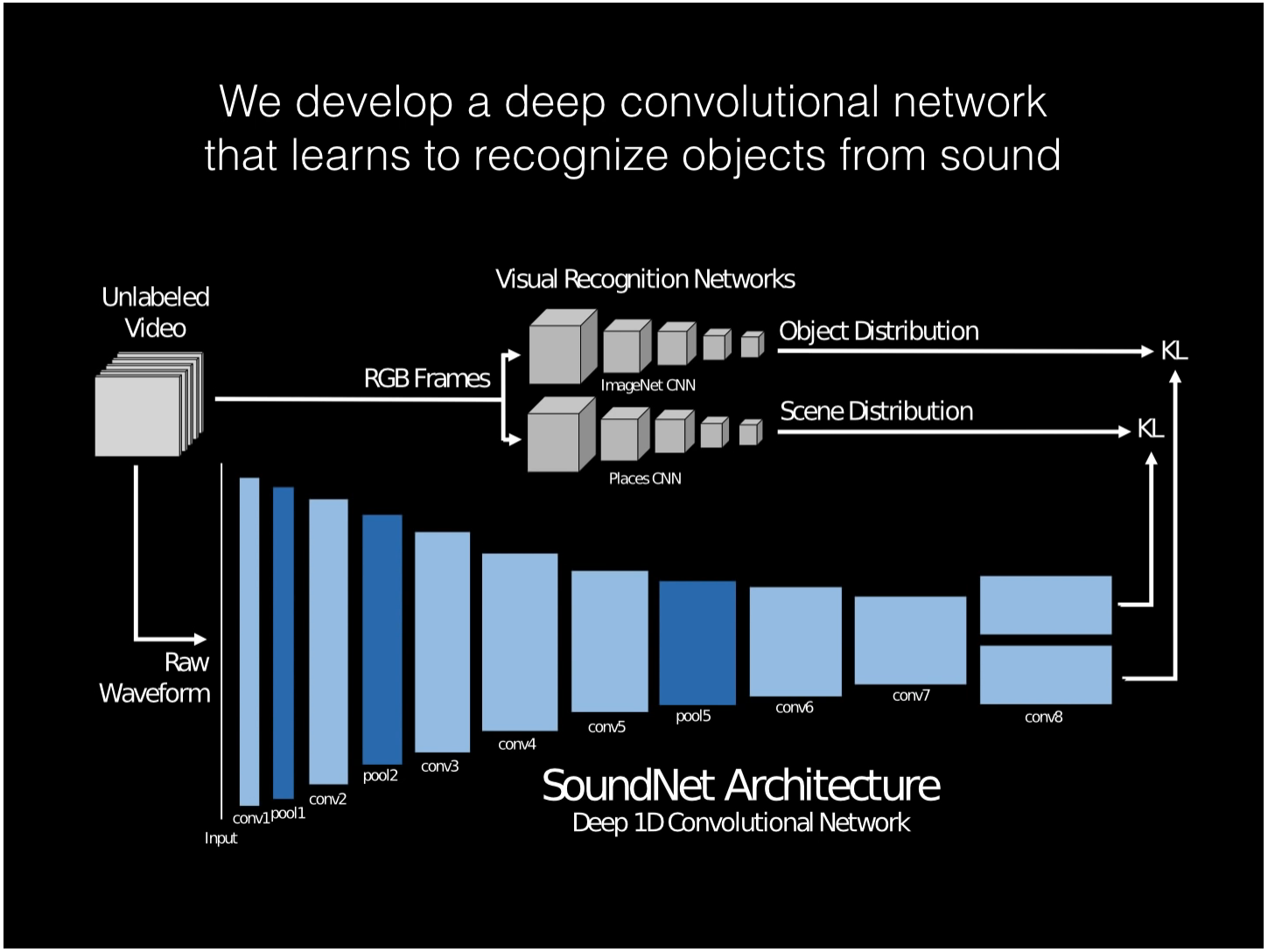
Cross Modal Networks



Cross-Modal Inversion



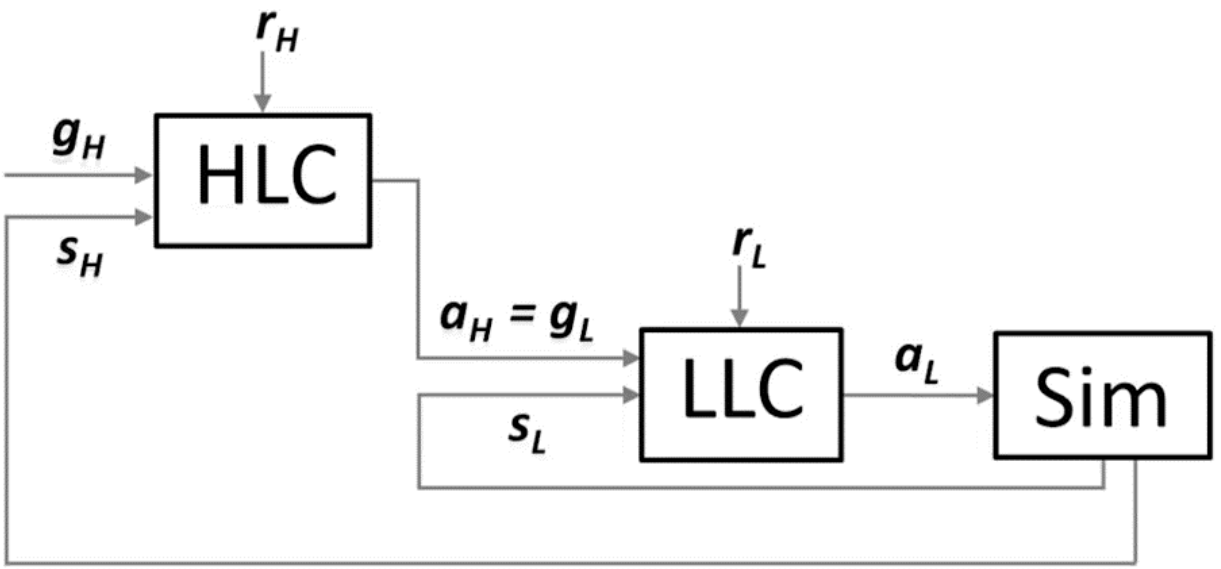
Cross-Modal Transfer: SoundNet



Y. Aytar, C. Vondrick, Antonio Torralba, SoundNet: Learning Sound Representations from Unlabeled Video, NIPS 2016
,

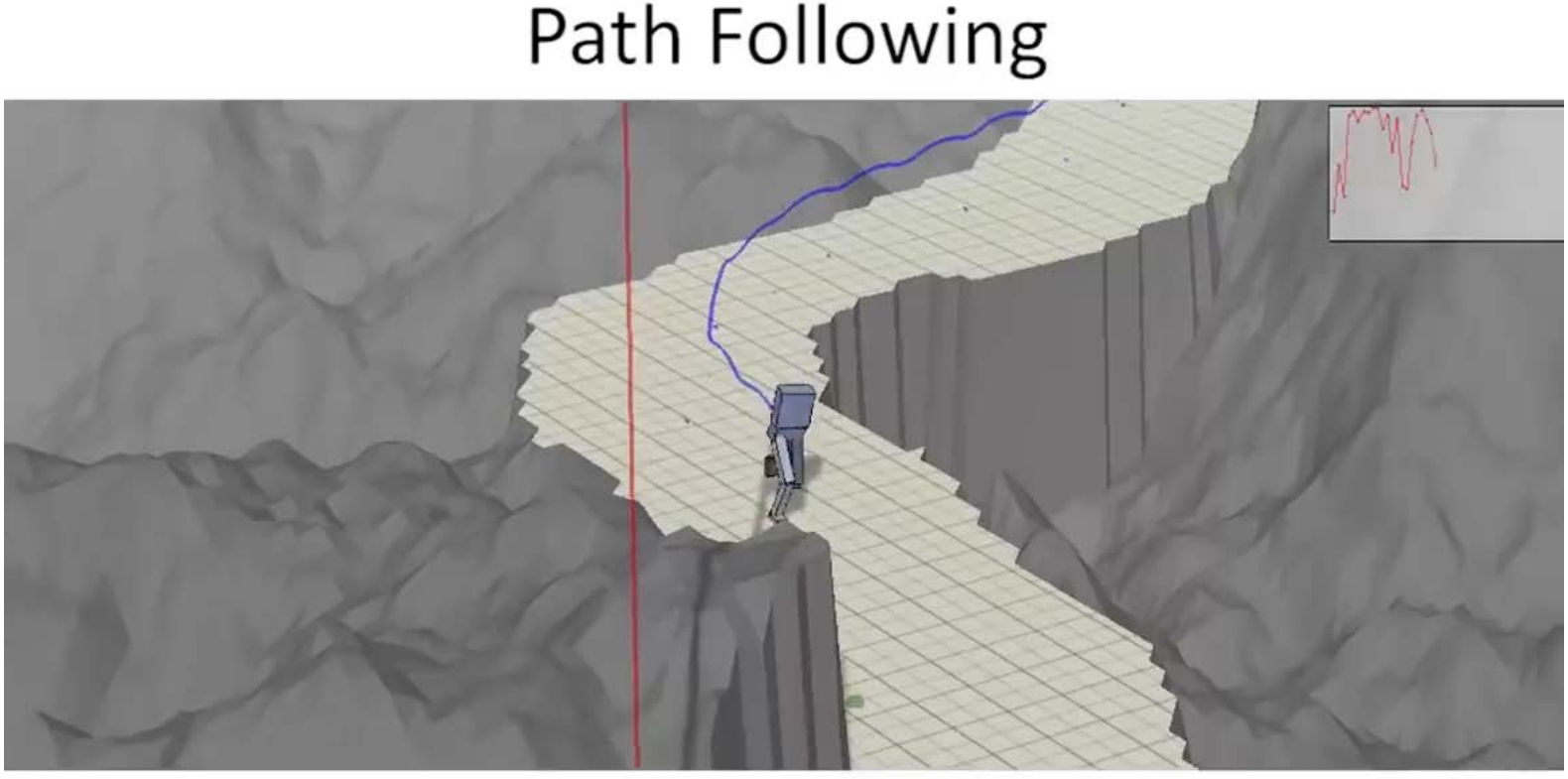
More than Images

- Character Control: Reinforcement Learning



More than Images

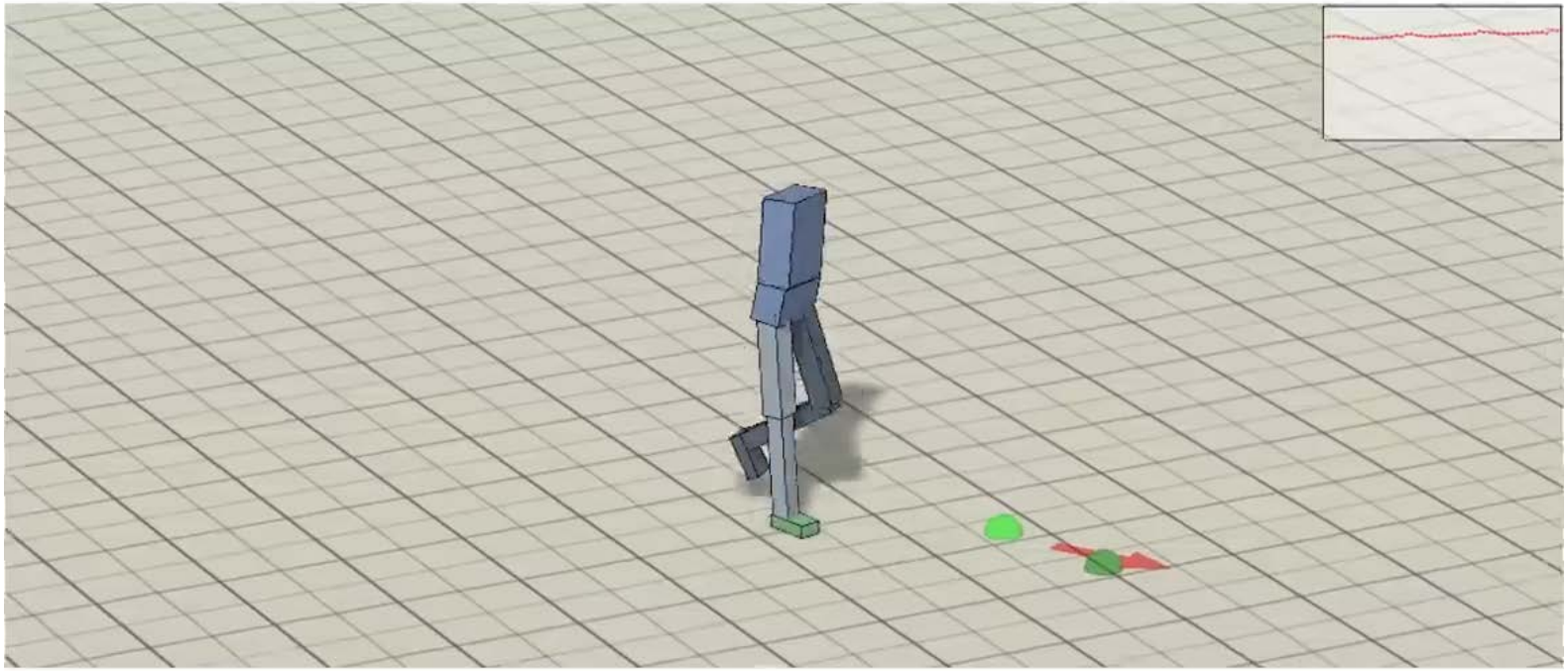
- Character Control: Reinforcement Learning



More than Images

- *Character Control: Reinforcement Learning*

LLC: Walk



The LLC is first trained to locomote while following random footstep plans.

Summary

- Vision: Analysis: Discriminative Models
- Graphics: Synthesis: Generative Models
- But many approaches combine both
- ML methods crucial for vision & graphics
- Rapidly moving and exciting field – get involved!