Learning Visual Representations

Amir R. Zamir

CS231A: Spring 2017
“Solving a problem simply means representing it so as to make the solution transparent.”

— Herbert Simon, Sciences of the Artificial
“Representation?”
My heart beats as if the world is drooping, you may not feel the love but do its a heart breaking moment of your life, enjoy the times that we have, it might not sound good but one thing it rhymes it might not be romantic but i think it is great, the best rhyme i've ever heard.
Macbeth was guilty.
Macbeth was guilty.

Representation

\[\{81, 20, 84, 64, 58, 39, 17, 54, 72, 15\}\]

Mathematical Model
(e.g., classifier)

\[f(x)\]

“Transcript”
~12 lbs

~8 lbs
~12 lbs

~8 lbs
~12 lbs

~8 lbs

Representation

Mathematical Model (Classifier)

Weight \( (w) \)

\( w > 11 \)

Type A

Type B
Represent these cats for a cat detector!
Represent these cats for a cat detector! (II)
Represent these cats for a cat detector! (II)
Represent these cats for a cat detector! (III)
Represent these cats for a cat detector! (IV)
Color Histograms

Deformable Part based Models (DPM)

Models based Shapes

Histogram of Gradients (HOG)

Felzenszwalb et al. 2010.
Dalal and Triggs, 2005.
Not always as easy (Happy vs Sad)
Not always as easy (Sad)
Learning Representations
Two approaches to representation learning

Supervised

Representation constrained on task(s).

Unsupervised

Representation constrained on reconstruction.

LeCun et al. 1998.
Hinton et al. 2006.
Two approaches to representation learning

**Supervised**

Representation constrained on task(s).

**Unsupervised**

Representation constrained on reconstruction.
Two approaches to representation learning

**Supervised**

Representation constrained on task(s).

**Unsupervised**

Representation constrained on reconstruction.

LeCun et al. 1998.
Hinton et al. 2006.
Pros and Cons

**Supervised**

Representation constrained on task(s).

**Unsupervised**

Representation constrained on reconstruction.

---

LeCun et al. 1998.
Hinton et al. 2006.
Neural Networks

Convolutional X

input layer

hidden layer 1

hidden layer 2

output layer

Neural Net

A Neuron
Understanding Representations

Understanding Representations

Inverting a representation

[ 81 20 84 64 58 39 17 54 72 15]
Understanding Representations

Dosovitskiy and Brox, 2015.

Inverting a representation
Representations in NLP, Brain, Speech, etc.

Word2Vec (NLP)

FMRI Scan (brain)

Mikolov et al. 2013
Macbeth was guilty.
Now that we’re done with building some background...
Visual Polymath

a.k.a generic/comprehensive/broad representation/perception
The main lesson of thirty-five years of AI research is that **the hard problems are easy and the easy problems are hard**. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived...
The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old that we take for granted – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived...

As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.

–Steven Pinker
An Exciting Time!

Fully Supervised Learning
Fully Supervised Learning

Isolation ~ “Idiot Savant”

Task Interplay
Task Interplay
Colorization $\rightarrow$ Object Detection

Zhang et al., Colorful Image Colorization, 2016.
Task Interplay
• Proposition:
  • (Instead of providing supervision over the desired tasks)
  • Provide supervision over a set of selected foundational tasks \( \Rightarrow \)
    generalization to novel tasks and abstraction capabilities.

• But how to pick the foundational tasks?
  • Biology
  • Inspirations from developmental stages of visual skills in brain
Generic 3D Representation Learning

Generic 3D Representation via Pose Estimation and Matching.
Amir Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, Silvio Savarese.
ECCV 2016.
Generic 3D Representation Learning

<table>
<thead>
<tr>
<th>Images</th>
<th>Input</th>
<th>Multi-Task ConvNet</th>
<th>Supervised Foundational Tasks</th>
<th>Unsupervised Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Patch 1" /></td>
<td><img src="image2.png" alt="Patch 2" /></td>
<td><img src="convnet.png" alt="Multi-Task ConvNet" /></td>
<td><img src="pose.png" alt="Relative Camera Pose" /></td>
<td><img src="tasks.png" alt="Unsupervised Tasks" /></td>
</tr>
</tbody>
</table>

Generic 3D Representation via Pose Estimation and Matching.

Amir Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, Silvio Savarese.

ECCV 2016.
Learn it from the world!
<table>
<thead>
<tr>
<th>Supervised Tasks</th>
<th>Unsupervised Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Camera pose estimation</td>
<td>• Surface Normal</td>
</tr>
<tr>
<td>• Matching (wide-baseline)</td>
<td>• 3D Object Pose</td>
</tr>
<tr>
<td></td>
<td>• 3D Scene Layout</td>
</tr>
<tr>
<td>State-of-the-art Human-level</td>
<td>• Visual Abstraction</td>
</tr>
<tr>
<td></td>
<td>State-of-the-art unsupervised</td>
</tr>
</tbody>
</table>
MIT Places
3D Object Pose - ImageNet

Airship (n02692877)
3D Object Pose - ImageNet

Chest (n03014705)
Demo

http://3drepresentation.stanford.edu/

Query Image

Generic 3D Representation

ImageNet (AlexNet)
3D Object Pose Estimation - Abstraction

- Wang & Gupta, 2015
- Agrawal et al., 2015
- Russakovsky et al., 2015
- Ozuysal et al., 2009.
# 3D Object Pose Estimation - Abstraction

<table>
<thead>
<tr>
<th>Query</th>
<th>Ours</th>
<th>Wang &amp; Gupta</th>
<th>Agrawal et al.</th>
<th>AleNet - Pool5</th>
<th>AleNet - FC7</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Query 1" /></td>
<td><img src="image2" alt="1st NN" /></td>
<td><img src="image3" alt="1st NN" /></td>
<td><img src="image4" alt="1st NN" /></td>
<td><img src="image5" alt="1st NN" /></td>
<td><img src="image6" alt="1st NN" /></td>
</tr>
<tr>
<td><img src="image7" alt="2nd NN" /></td>
<td><img src="image8" alt="2nd NN" /></td>
<td><img src="image9" alt="2nd NN" /></td>
<td><img src="image10" alt="2nd NN" /></td>
<td><img src="image11" alt="2nd NN" /></td>
<td><img src="image12" alt="2nd NN" /></td>
</tr>
<tr>
<td><img src="image13" alt="3rd NN" /></td>
<td><img src="image14" alt="3rd NN" /></td>
<td><img src="image15" alt="3rd NN" /></td>
<td><img src="image16" alt="3rd NN" /></td>
<td><img src="image17" alt="3rd NN" /></td>
<td><img src="image18" alt="3rd NN" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Query</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
<th>1st NN</th>
<th>2nd NN</th>
<th>3rd NN</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image19" alt="Query 2" /></td>
<td><img src="image20" alt="1st NN" /></td>
<td><img src="image21" alt="2nd NN" /></td>
<td><img src="image22" alt="3rd NN" /></td>
<td><img src="image23" alt="1st NN" /></td>
<td><img src="image24" alt="2nd NN" /></td>
<td><img src="image25" alt="3rd NN" /></td>
<td><img src="image26" alt="1st NN" /></td>
<td><img src="image27" alt="2nd NN" /></td>
<td><img src="image28" alt="3rd NN" /></td>
<td><img src="image29" alt="1st NN" /></td>
<td><img src="image30" alt="2nd NN" /></td>
<td><img src="image31" alt="3rd NN" /></td>
<td><img src="image32" alt="1st NN" /></td>
<td><img src="image33" alt="2nd NN" /></td>
<td><img src="image34" alt="3rd NN" /></td>
<td><img src="image35" alt="1st NN" /></td>
<td><img src="image36" alt="2nd NN" /></td>
<td><img src="image37" alt="3rd NN" /></td>
</tr>
<tr>
<td><img src="image38" alt="Query 3" /></td>
<td><img src="image39" alt="1st NN" /></td>
<td><img src="image40" alt="2nd NN" /></td>
<td><img src="image41" alt="3rd NN" /></td>
<td><img src="image42" alt="1st NN" /></td>
<td><img src="image43" alt="2nd NN" /></td>
<td><img src="image44" alt="3rd NN" /></td>
<td><img src="image45" alt="1st NN" /></td>
<td><img src="image46" alt="2nd NN" /></td>
<td><img src="image47" alt="3rd NN" /></td>
<td><img src="image48" alt="1st NN" /></td>
<td><img src="image49" alt="2nd NN" /></td>
<td><img src="image50" alt="3rd NN" /></td>
<td><img src="image51" alt="1st NN" /></td>
<td><img src="image52" alt="2nd NN" /></td>
<td><img src="image53" alt="3rd NN" /></td>
<td><img src="image54" alt="1st NN" /></td>
<td><img src="image55" alt="2nd NN" /></td>
<td><img src="image56" alt="3rd NN" /></td>
</tr>
</tbody>
</table>

References:
- Wang & Gupta, 2015
- Agrawal et al., 2015
- Russakovsky et al., 2015
- Ozuysal et al., 2009.
3D Layout Estimation - Abstraction

<table>
<thead>
<tr>
<th>Query</th>
<th>Ours</th>
<th>1st NN</th>
<th>2nd NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Query Image]</td>
<td>![Ours Image]</td>
<td>![1st NN Image]</td>
<td>![2nd NN Image]</td>
</tr>
</tbody>
</table>

Wang & Gupta. 2015
Agrawal et al. 2015
Russakovsky et al. 2015
Zhang et al. 2016.
### 3D Layout Estimation - Abstraction

<table>
<thead>
<tr>
<th>Query</th>
<th>Ours 1st NN</th>
<th>Ours 2nd NN</th>
<th>Wang &amp; Gupta 1st NN</th>
<th>Wang &amp; Gupta 2nd NN</th>
<th>Agrawal et al. 1st NN</th>
<th>Agrawal et al. 2nd NN</th>
<th>AlexNet - Pool5 1st NN</th>
<th>AlexNet - Pool5 2nd NN</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Query 1" /></td>
<td><img src="image2" alt="Ours 1st NN 1" /></td>
<td><img src="image3" alt="Ours 2nd NN 1" /></td>
<td><img src="image4" alt="Wang &amp; Gupta 1st NN 1" /></td>
<td><img src="image5" alt="Wang &amp; Gupta 2nd NN 1" /></td>
<td><img src="image6" alt="Agrawal et al. 1st NN 1" /></td>
<td><img src="image7" alt="Agrawal et al. 2nd NN 1" /></td>
<td><img src="image8" alt="AlexNet - Pool5 1st NN 1" /></td>
<td><img src="image9" alt="AlexNet - Pool5 2nd NN 1" /></td>
</tr>
<tr>
<td><img src="image10" alt="Query 2" /></td>
<td><img src="image11" alt="Ours 1st NN 2" /></td>
<td><img src="image12" alt="Ours 2nd NN 2" /></td>
<td><img src="image13" alt="Wang &amp; Gupta 1st NN 2" /></td>
<td><img src="image14" alt="Wang &amp; Gupta 2nd NN 2" /></td>
<td><img src="image15" alt="Agrawal et al. 1st NN 2" /></td>
<td><img src="image16" alt="Agrawal et al. 2nd NN 2" /></td>
<td><img src="image17" alt="AlexNet - Pool5 1st NN 2" /></td>
<td><img src="image18" alt="AlexNet - Pool5 2nd NN 2" /></td>
</tr>
<tr>
<td><img src="image19" alt="Query 3" /></td>
<td><img src="image20" alt="Ours 1st NN 3" /></td>
<td><img src="image21" alt="Ours 2nd NN 3" /></td>
<td><img src="image22" alt="Wang &amp; Gupta 1st NN 3" /></td>
<td><img src="image23" alt="Wang &amp; Gupta 2nd NN 3" /></td>
<td><img src="image24" alt="Agrawal et al. 1st NN 3" /></td>
<td><img src="image25" alt="Agrawal et al. 2nd NN 3" /></td>
<td><img src="image26" alt="AlexNet - Pool5 1st NN 3" /></td>
<td><img src="image27" alt="AlexNet - Pool5 2nd NN 3" /></td>
</tr>
</tbody>
</table>
Task Interplay
Task Interplay
Intelligence as Efficiency

• Yielding higher value for less resources.

Intelligence as Efficiency

- Yielding higher value for less resources.

Practice: Prediction on a Budget

Time

Resources

65 mph

1 second
40 meters
Practice: Prediction on a Budget

A tradeoff game

Time

Resources

✓ wheeled vehicle

✗ foliage

✓ road bike

✓ flat bar bike

× tandem bike

65 mph
Feedback networks

Amir Zamir*, Te-Lin Wu*, Lin Sun, William Shen, Bertram Shi,
Jitendra Malik, Silvio Savarese
CVPR 2017

http://feedbacknet.stanford.edu
Feedforward model.
Feedforward model.
Feedforward model.
Feedforward model.

Feedback model.

an alternative with several advantages
Feedback model unrolled.
Advantage I: Early Prediction

Feedback enables making early predictions of the output.
Advantage II: Taxonomic Prediction
Feedback predictions naturally conform to a taxonomy (even when trained without a taxonomy)

**Advantage II: Taxonomic Prediction**
Experimental Results
Qualitative results on CIFAR100 test set
Thank you!

Amir R. Zamir
zamir@cs.stanford.edu