Table of contents

KEYNOTES

Reverse Engineering the Human Visual System
Jack L. Gallant
most related to the following paper: http://arxiv.org/pdf/1407.5104v1.pdf

- for each brain voxel learn feature weights (on feature vector extracted from pixel features) to predict voxel response
- can learn weights on different layers of CNNs to determine which layers weighted most for which voxels
- correspond layers of CNNs to brain areas

What's Wrong with Deep Learning?
Yann LeCun
talk slides available here: https://drive.google.com/file/d/0BxKBnD5y2M8NVHRiVXBnOVpiYUk/edit

- resurgence of methodology attempted previously
- What's wrong?:
- 1) There is some theory missing!
- 2) Memory problem
- 3) Unsupervised learning

CNNs for different tasks

Hypercolumns for Object Segmentation and Fine-Grained Localization [full paper] [ext. abstract]
Bharath Hariharan, Pablo Arbeláez, Ross Girshick, Jitendra Malik

- outputs of all units at a particular location stacked into a vector = hypercolumn
- captures the visual content's representation at multiple scales
- hypercolumn per pixel passed through linear classifier -> solve prediction task: is this part of object, part, or keypoint?
- use linearity of classifiers to interpolate classification weights instead of feature representations themselves
- applications: segmentation and part labeling

Going Deeper With Convolutions [full paper] [ext. abstract]
Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich

- inception module replacing traditional convolutional layers
- GoogleNet submission beats R-CNN
- ensemble of 6 nets improves prediction

Predicting Eye Fixations Using Convolutional Neural Networks [full paper] [ext. abstract]
Nian Liu, Junwei Han, Dingwen Zhang, Shifeng Wen, Tianming Liu

- CNNs trained on fixated and non-fixated image regions (patches)
- multi-scale CNNs to predict saliency values at different locations in an image to obtain saliency maps

Is Object Localization for Free? - Weakly-Supervised Learning With Convolutional Neural Networks [full paper] [ext. abstract]
Maxime Oquab, Léon Bottou, Ivan Laptev, Josef Sivic

- object localization by searching for highest-scoring window

Designing Deep Networks for Surface Normal Estimation [full paper] [ext. abstract]
Xiaolong Wang, David Fouhey, Abhinav Gupta

Finding Action Tubes [full paper] [ext. abstract]
Georgia Gkioxari, Jitendra Malik

- classifying actions based on spatial and motion cues
- linking regions across frames based on action predictions

Cross-Scene Crowd Counting via Deep Convolutional Neural Networks [full paper] [ext. abstract]
Cong Zhang, Hongsheng Li, Xiaogang Wang, Xiaokang Yang

The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification[full paper] [ext. abstract]
Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, Zheng Zhang

- mine object parts and use them to help with classification

End-to-End Integration of a Convolution Network, Deformable Parts Model and Non-Maximum Suppression [full paper] [ext. abstract]
Li Wan, David Eigen, Rob Fergus

Sketch-Based 3D Shape Retrieval Using Convolutional Neural Networks [full paper] [ext. abstract]
Fang Wang, Le Kang, Yi Li

MatchNet: Unifying Feature and Metric Learning for Patch-Based Matching [full paper] [ext. abstract]
Xufeng Han, Thomas Leung, Yangqing Jia, Rahul Sukthankar, Alexander C. Berg

Deep Filter Banks for Texture Recognition and Segmentation [full paper] [ext. abstract]
Mircea Cimpoi, Subhransu Maji, Andrea Vedaldi

understanding/visualizing CNNs

Understanding Image Representations by Measuring Their Equivariance and Equivalence [full paper] [ext. abstract]
Karel Lenc, Andrea Vedaldi

- investigation of representational qualities of different features - i.e. CNN filters
- equivariance: how representation changes based on transformations of input (predictable changes in CNN indicate geometry is represented)
- invariance: and how it builds with depth
- equivalence: same representation when trained differently (and in some cases for different tasks)

Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images [full paper] [ext. abstract]
Anh Nguyen, Jason Yosinski, Jeff Clune

- fooling images: evolutionary algorithms or gradient descent to find noise/pattern images that have high-confidence CNN predictions but are unrecognizable to humans

Understanding Deep Image Representations by Inverting Them [full paper] [ext. abstract]
Aravindh Mahendran, Andrea Vedaldi

- where does invariance come about?

Deformable Part Models are Convolutional Neural Networks [full paper] [ext. abstract]
Ross Girshick, Forrest Iandola, Trevor Darrell, Jitendra Malik

- can construct a CNN equivalent to a DPM

new datasets

SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite [full paper] [ext. abstract]
Shuran Song, Samuel P. Lichtenberg, Jianxiong Xiao

Material Recognition in the Wild With the Materials in Context Database [full paper] [ext. abstract]
Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala

- efficient way of gathering labels for materials
- material classification with GoogleNet performance

Deeply Learned Attributes for Crowded Scene Understanding [full paper] [ext. abstract]
Jing Shao, Kai Kang, Chen Change Loy, Xiaogang Wang

ActivityNet: A Large-Scale Video Benchmark for Human Activity Understanding [full paper] [ext. abstract]
Fabian Caba Heilbron, Victor Escorcia, Bernard Ghanem, Juan Carlos Niebles

- most varied activity dataset
- taxonomy of activities

SALICON: Saliency in Context [full paper] [ext. abstract]
Ming Jiang, Shengsheng Huang, Juanyong Duan, Qi Zhao

- mouse maps to gather saliency locations
- why? rank object categories, suggest new categories, suggest objects for scene description, etc.

crowdsourcing labels

Building a Bird Recognition App and Large Scale Dataset With Citizen Scientists: The Fine Print in Fine-Grained Dataset Collection [full paper] [ext. abstract]
Grant Van Horn, Steve Branson, Ryan Farrell, Scott Haber, Jessie Barry, Panos Ipeirotis, Pietro Perona, Serge Belongie

- comparison between citizen scientist and AMT populations

Best of Both Worlds: Human-Machine Collaboration for Object Annotation [full paper] [ext. abstract]
Olga Russakovsky, Li-Jia Li, Li Fei-Fei
x

- Markov Decision Process to determine best question to ask to maximize understanding of a scene (labels) while minimizing user labeling effort

Image Segmentation in Twenty Questions [full paper] [ext. abstract]
Christian Rupprecht, Loïc Peter, Nassir Navab

Efficient Label Collection for Unlabeled Image Datasets [full paper] [ext. abstract]
Maggie Wigness, Bruce A. Draper, J. Ross Beveridge

- learn bottom-up (feature-based) clustering of images (create a tree)
- query people for labels to describe images sampled at a particular node in the tree, propagate label

generative approaches

Picture: A Probabilistic Programming Language for Scene Perception [full paper] [ext. abstract]
Tejas D. Kulkarni, Pushmeet Kohli, Joshua B. Tenenbaum, Vikash Mansinghka

- Monte Carlo estimation for inference in high-dimensional space to construct proposals to explain data
- use proposals to infer parameters of data: viewpoint, pose, shape

24/7 Place Recognition by View Synthesis [full paper] [ext. abstract]
Akihiko Torii, Relja Arandjelović, Josef Sivic, Masatoshi Okutomi, Tomas Pajdla

- synthesize novel views of place to do place recognition (and potentially label transfer tasks

Learning to Generate Chairs With Convolutional Neural Networks[full paper] [ext. abstract]
Alexey Dosovitskiy, Jost Tobias Springenberg, Thomas Brox

- generative network to produce new views of object, interpolate between styles, etc.

unsupervised (semi-supervised) approaches

Unsupervised Object Discovery and Localization in the Wild: Part-Based Matching With Bottom-Up Region Proposals [full paper] [ext. abstract]
Minsu Cho, Suha Kwak, Cordelia Schmid, Jean Ponce

FlowWeb: Joint Image Set Alignment by Weaving Consistent, Pixel-Wise Correspondences [full paper] [ext. abstract]
Tinghui Zhou, Yong Jae Lee, Stella X. Yu, Alyosha A. Efros

Fine-Grained Recognition Without Part Annotations [full paper] [ext. abstract]
Jonathan Krause, Hailin Jin, Jianchao Yang, Li Fei-Fei

DEEP-CARVING: Discovering Visual Attributes by Carving Deep Neural Nets [full paper] [ext. abstract]
Sukrit Shankar, Vikas K. Garg, Roberto Cipolla

language and vision

From Captions to Visual Concepts and Back [full paper] [ext. abstract]
Hao Fang, Saurabh Gupta, Forrest Iandola, Rupesh K. Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C. Platt, C. Lawrence Zitnick, Geoffrey Zweig

Show and Tell: A Neural Image Caption Generator [full paper] [ext. abstract]
Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan

Deep Visual-Semantic Alignments for Generating Image Descriptions [full paper] [ext. abstract]
Andrej Karpathy, Li Fei-Fei

- relationship between text snippets and labels
- bidirectional RNN

Long-Term Recurrent Convolutional Networks for Visual Recognition and Description [full paper] [ext. abstract]
Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, Trevor Darrell

Mind's Eye: A Recurrent Visual Representation for Image Caption Generation [full paper] [ext. abstract]
Xinlei Chen, C. Lawrence Zitnick

Joint Photo Stream and Blog Post Summarization and Exploration [full paper] [ext. abstract]
Gunhee Kim, Seungwhan Moon, Leonid Sigal

VisKE: Visual Knowledge Extraction and Question Answering by Visual Verification of Relation Phrases [full paper] [ext. abstract]
Fereshteh Sadeghi, Santosh K. Kumar Divvala, Ali Farhadi

- train on subject,verb and verb,object combinations
- for a new (subject,verb,object) triple check for presence of 2 enclosed relationships and their spatial relationship in image for reasoning

Image Retrieval Using Scene Graphs [full paper] [ext. abstract]
Justin Johnson, Ranjay Krishna, Michael Stark, Li-Jia Li, David Shamma, Michael Bernstein, Li Fei-Fei

Learning Semantic Relationships for Better Action Retrieval in Images [full paper] [ext. abstract]
Vignesh Ramanathan, Congcong Li, Jia Deng, Wei Han, Zhen Li, Kunlong Gu, Yang Song, Samy Bengio, Charles Rosenberg, Li Fei-Fei

evaluation

Image Specificity [full paper] [ext. abstract]
Mainak Jas, Devi Parikh

- motivation: for better scoring image captioning algorithms
- measure of how diverse a set of human-produced captions are for an image

CIDEr: Consensus-Based Image Description Evaluation [full paper] [ext. abstract]
Ramakrishna Vedantam, C. Lawrence Zitnick, Devi Parikh

- how to evaluate image captioning systems using a collection of human ground-truth captions

attentional mechanisms, saliency, and related

Social Saliency Prediction [full paper] [ext. abstract]
Hyun Soo Park, Jianbo Shi

- predict where groups of people look (center of attention in the image)
- estimate face pose of people in image

Finding Distractors In Images [full paper] [ext. abstract]
Ohad Fried, Eli Shechtman, Dan B. Goldman, Adam Finkelstein

- like saliency but for secondary objects that may be distracting from main elements
- goal: automatically discover and remove these from photographs

Learning To Look Up: Realtime Monocular Gaze Correction Using Machine Learning [full paper] [ext. abstract]
Daniil Kononenko, Victor Lempitsky

- correct your gaze automatically so that when you are looking down (e.g. Skyping), real-time correction to make it look like you are looking straight

VIP: Finding Important People in Images [full paper] [ext. abstract]
Clint Solomon Mathialagan, Andrew C. Gallagher, Dhruv Batra

Learning a Sequential Search for Landmarks **[full paper]** **[ext. abstract]**
Saurabh Singh, Derek Hoiem, David Forsyth

new problems

Building Proteins in a Day: Efficient 3D Molecular Reconstruction[full paper] [ext. abstract]
Marcus A. Brubaker, Ali Punjani, David J. Fleet

A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions [full paper] [ext. abstract]
Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, Andrew C. Gallagher

- predicting distribution of elicited emotions per image

Neuroaesthetics in Fashion: Modeling the Perception of Fashionability [full paper] [ext. abstract]
Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, Raquel Urtasun

- trend: using social networks and crowdsourcing likes/clicks as labels for data

Dataset Fingerprints: Exploring Image Collections Through Data Mining [full paper] [ext. abstract]
Konstantinos Rematas, Basura Fernando, Frank Dellaert, Tinne Tuytelaars

Learning Scene-Specific Pedestrian Detectors Without Real Data [full paper] [ext. abstract]
Hironori Hattori, Vishnu Naresh Boddeti, Kris M. Kitani, Takeo Kanade

- location-specific geometry aware pedestrian detection system
- trained on millions of synthesized pedestrian images

Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset) [full paper] [ext. abstract]
Jared Heinly, Johannes L. Schönberger, Enrique Dunn, Jan-Michael Frahm

Category-Specific Object Reconstruction From a Single Image [full paper] [ext. abstract]
Abhishek Kar, Shubham Tulsiani, João Carreira, Jitendra Malik

EgoSampling: Fast-Forward and Stereo for Egocentric Videos [full paper] [ext. abstract]
Yair Poleg, Tavi Halperin, Chetan Arora, Shmuel Peleg

- This is a significantly simpler version to create stable hyperlapse, compared to the popular siggraph paper from 2014. although perhaps less applicable to general complex situations (e.g. climbing). Simultaneous to the development of this paper, another group has developed a very similar method that will be presented at SIGGRAPH 2015.

Transport-Based Single Frame Super Resolution of Very Low Resolution Face Images [full paper] [ext. abstract]
Soheil Kolouri, Gustavo K. Rohde