Big Visual Data, Deep Learning, and Open Source

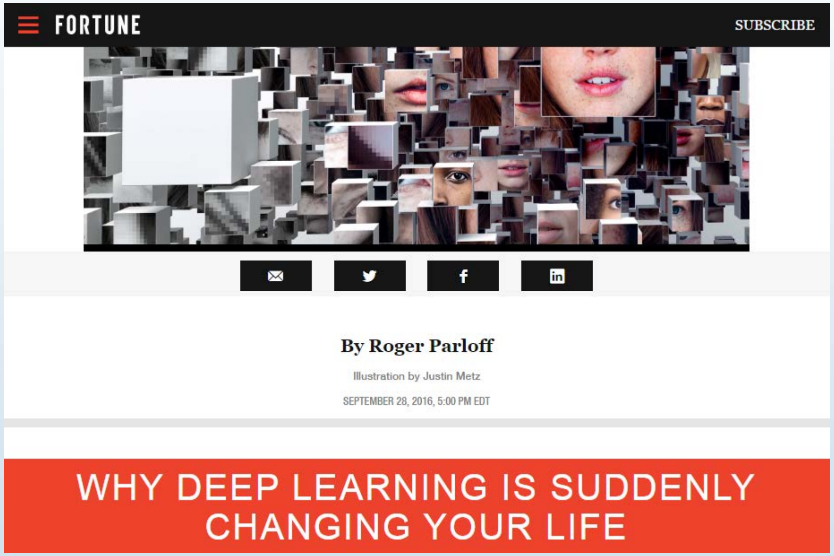
Dr. Anthony Hoogs
anthony. hoogs@kitware.com
Senior Director of Computer Vision
http://www.kitware.com/company/team/hoogs.html

Kitware, Inc.
Clifton Park, NY

The Deep Learning Revolution

“Neural nets aren’t new. What’s changed is that today computer scientists have finally harnessed both the vast computational power and the enormous storehouses of data—images, video, audio, and text files strewn across the Internet—that, it turns out, are essential to making neural nets work well.”

Fortune magazine
Sep. 28, 2016


FORTUNE
SUBSCRIBE

By Roger Parloff
Illustration by Justin Metz
SEPTEMBER 28, 2016, 5:00 PM EDT

WHY DEEP LEARNING IS SUDDENLY CHANGING YOUR LIFE

“Google had two deep-learning projects underway in 2012. Today it is pursuing more than 1,000.”

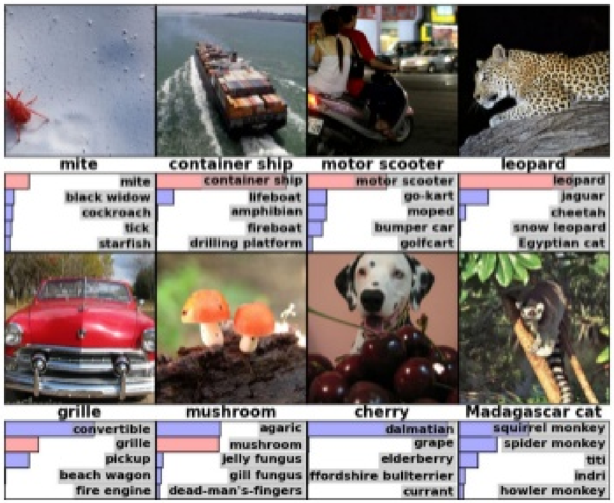
“[Google, Amazon, Microsoft, Apple] all have features that let you search or automatically organize collections of photos with no identifying tags. You can ask to be shown, say, all the ones that have dogs in them, or snow, or even something fairly abstract like hugs.”

Kitware

The (Re-)Birth of Convolutional Neural Networks

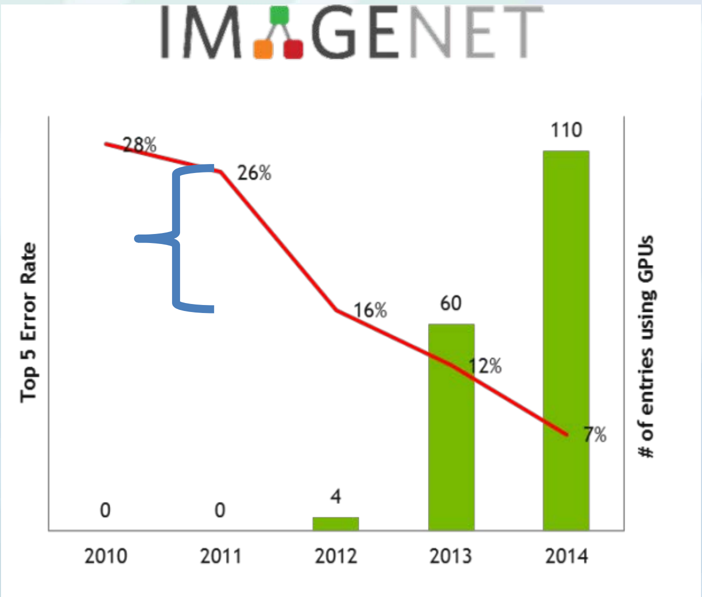
ImageNet Challenge

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



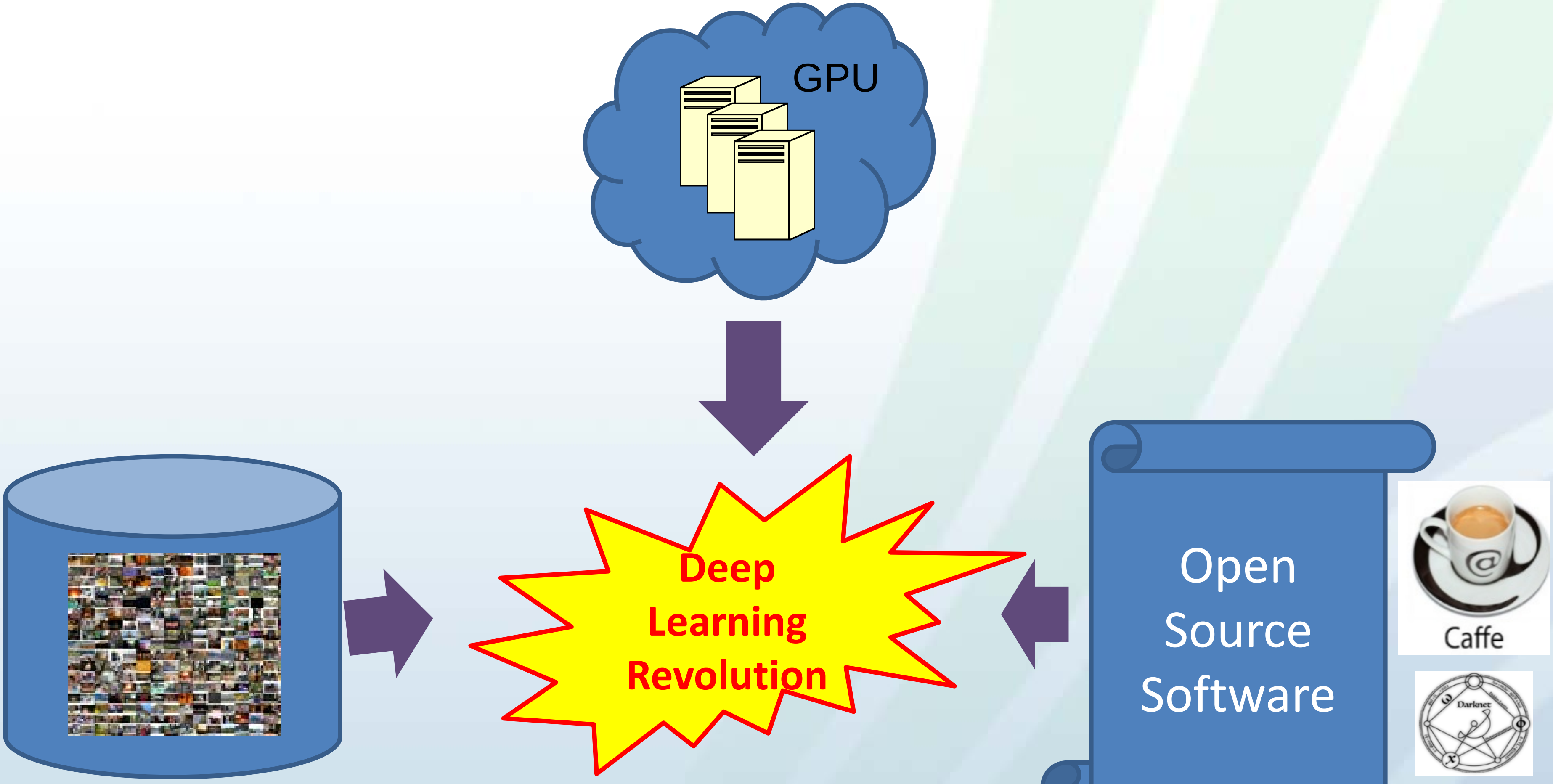
4

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. **ImageNet: A Large-Scale Hierarchical Image Database.** *IEEE Computer Vision and Pattern Recognition,* 2009.



A. Krizhevsky, I. Sutskever, and G. Hinton. [“ImageNet Classification with Deep Convolutional Neural Networks](ImageNet Classification with Deep Convolutional Neural Networks).” *Neural Information Processing Symposium*, 2012.

A Perfect Storm

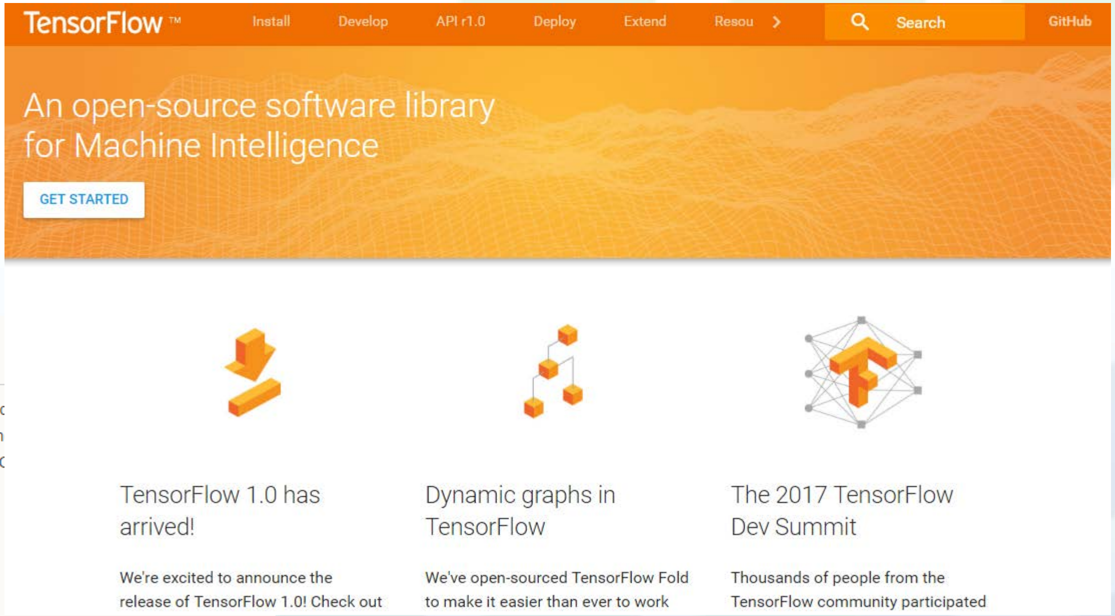


Kitware

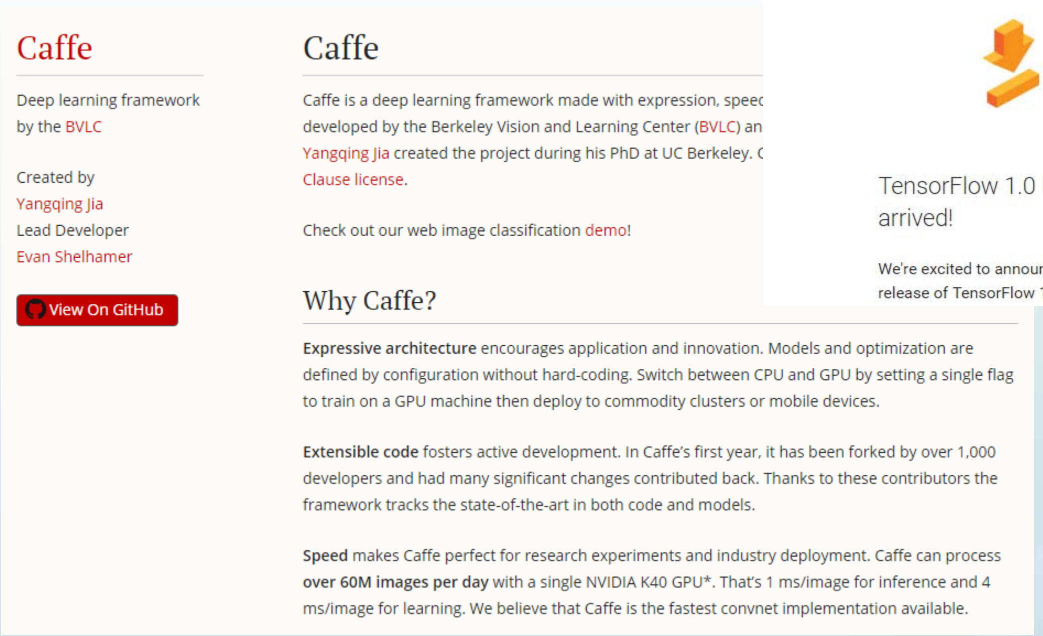
Kitware Open Source Platforms

- **KWIVER** Kitware Imagery and Video Exploitation and Retrieval
- **VTK** the visualization toolkit
- **ParaView** large data analysis & visualization application
- **ITK** insight image analysis toolkit
- **CMake** cross-platform build system
  - CDash, CTest, CPack, software process tools
- **Resonant/Girder** informatics and information visualization
- **Kiwi & VES** mobile visualization
- IGSTK, CTK, vxl, Open Chemistry Project, VolView, tubeTk, and more…
- **MIDAS** for computational scientific research, testing, and visualization

CNN Open Source Platforms



Google
https://www.tensorflow.org/



UC Berkeley
http://caffe.berkeleyvision.org/

CNN Open Source Models

Caffe
41 models

Model Zoo

Iacopo Masi edited this page 16 days ago · 111 revisions

Check out the model zoo documentation for details.

To acquire a model:

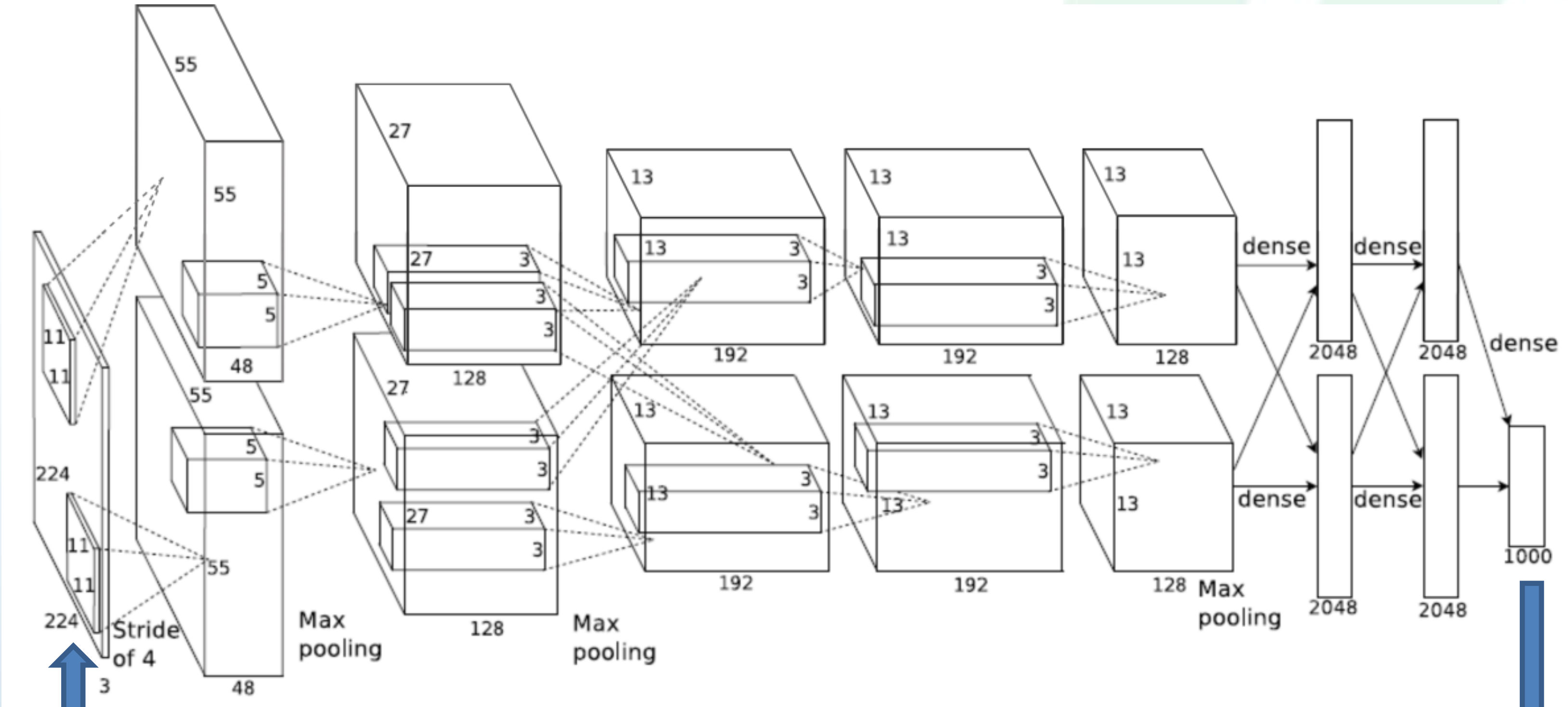
1. download the model gist by `./scripts/download_model_from_gist.sh <gist_id> <dirname>` to load the model metadata, architecture, solver configuration, and so on. (`<dirname>` is optional and defaults to caffe/models).
2. download the model weights by `./scripts/download_model_binary.py <model_dir>` where `<model_dir>` is the gist directory from the first step.

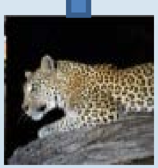
or visit the model zoo documentation for complete instructions.

Table of Contents

AlexNet

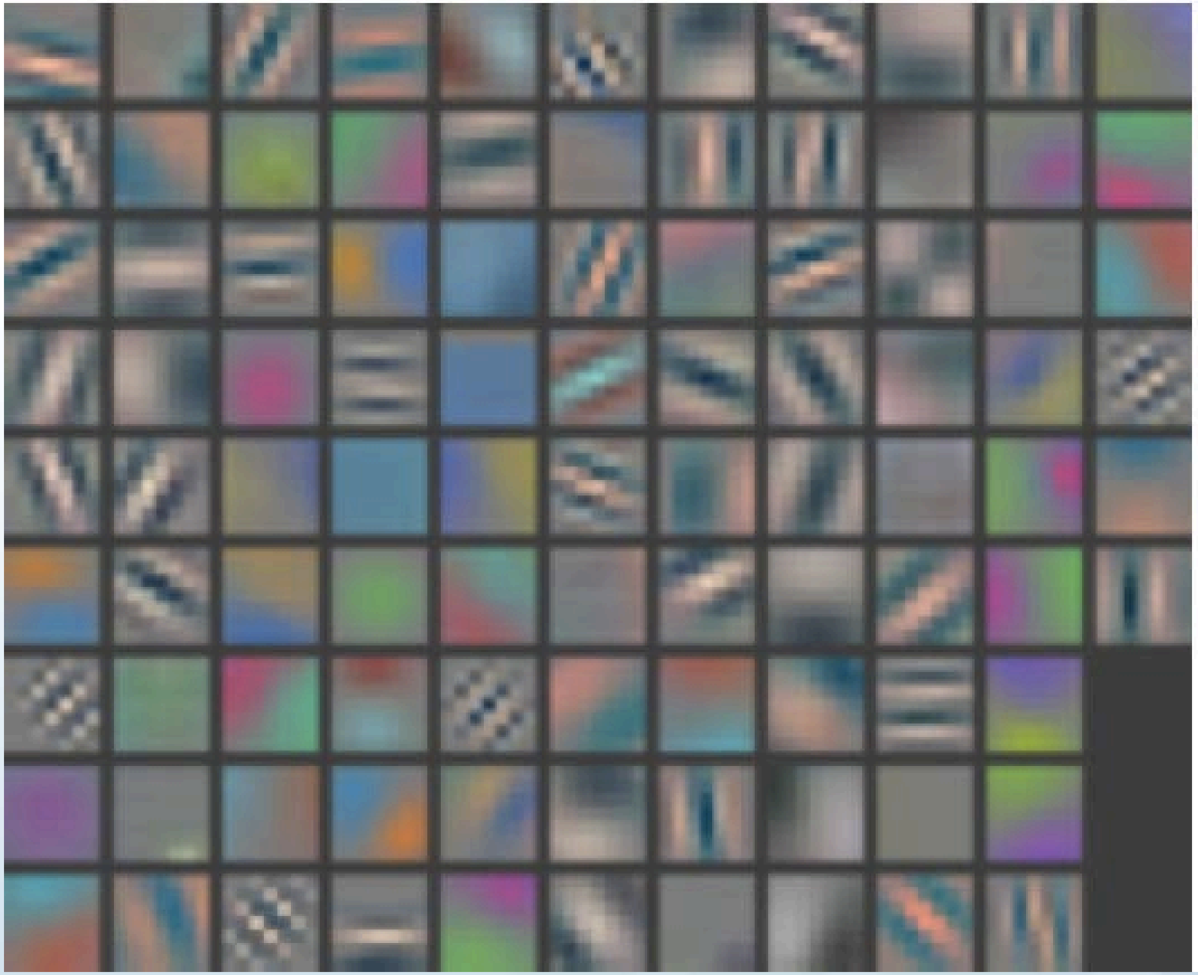
A. Krizhevsky, I. Sutskever, and G. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks.” *Neural Information Processing Symposium*, 2012.





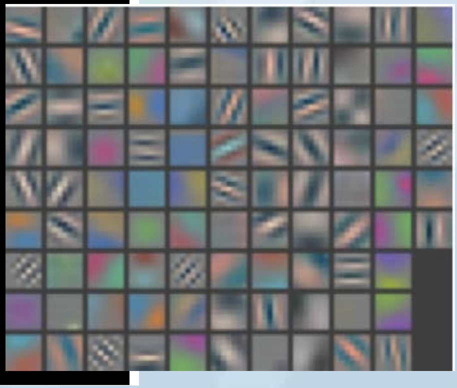
65M parameters

Layer 1 Filters

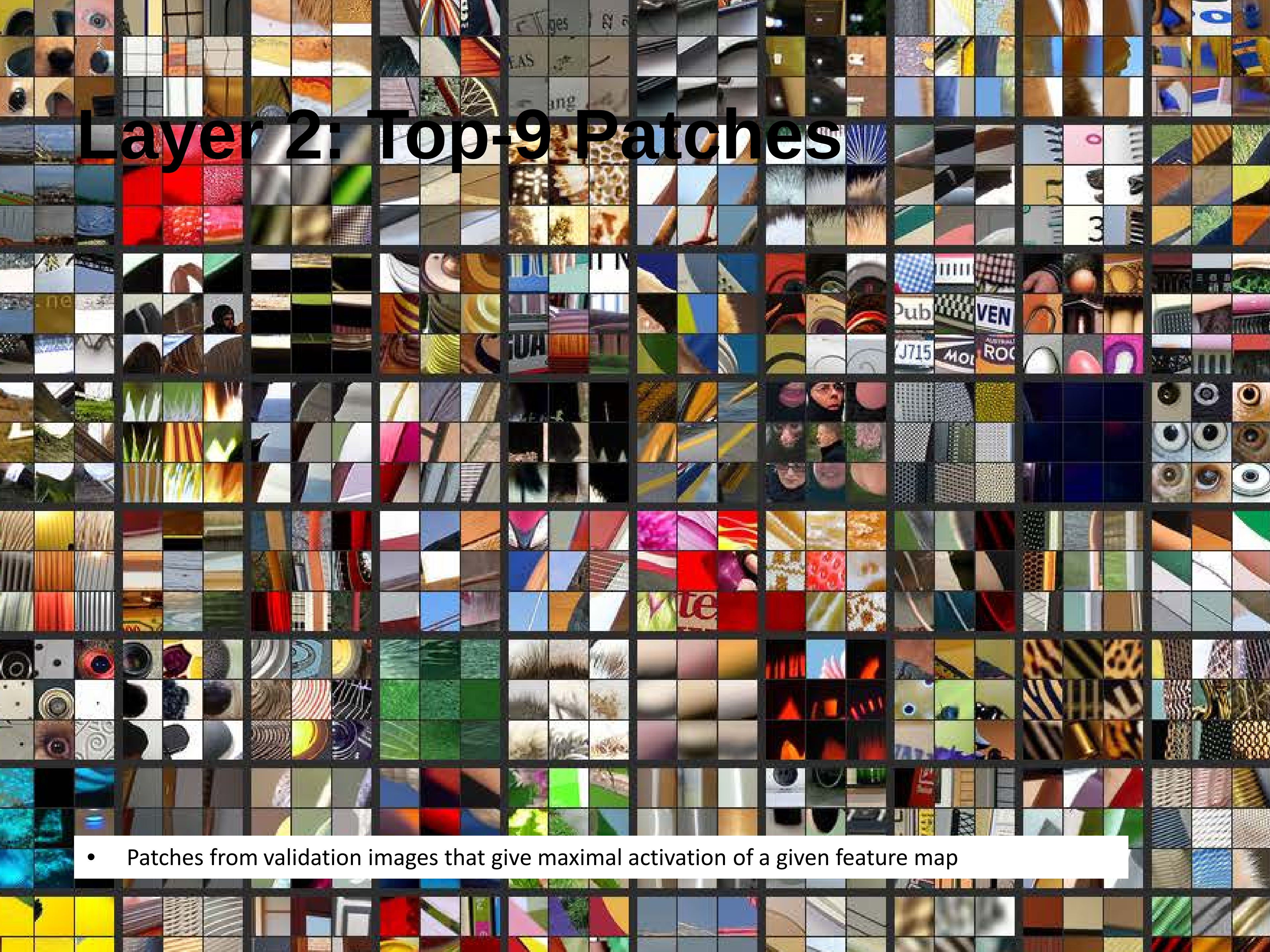


Slide credit: Yann LeCun

Layer 1: Top-9 Patches



Layer 2: Top-9 Patches



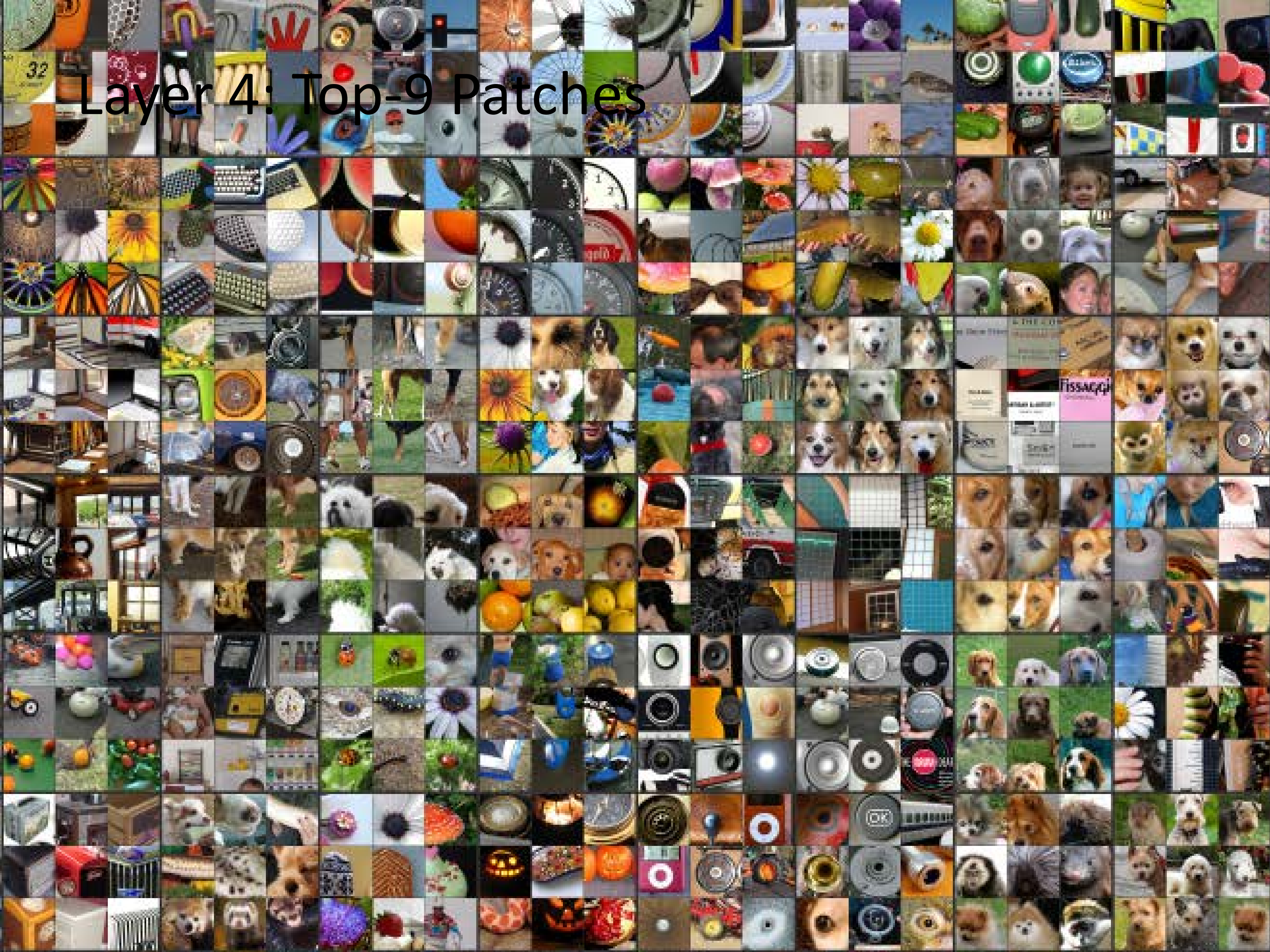
- Patches from validation images that give maximal activation of a given feature map

Layer 2: Top-9 Patches

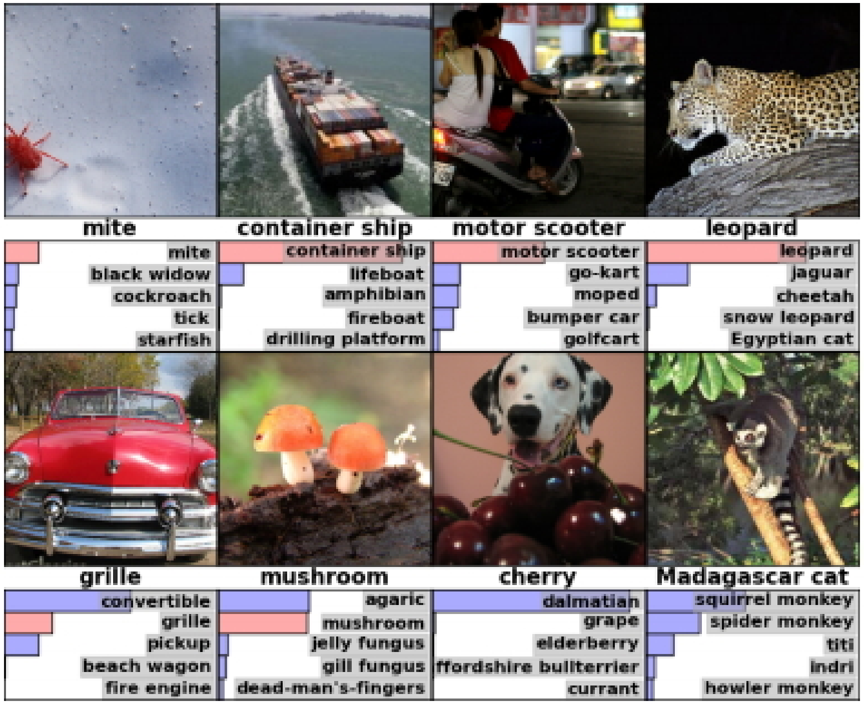
Layer 3: Top-9 Patches

Layer 3: Top-9 Patches

Layer 4: Top-9 Patches



AlexNet Results



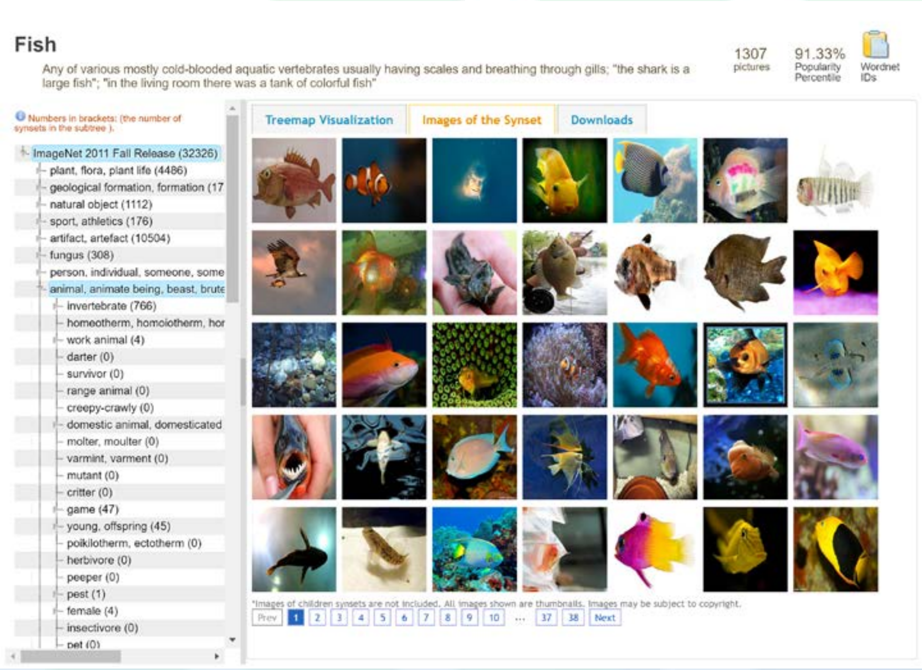
Top 5 Classes

Training Images

Kitware

Training Deep Networks: ImageNet

- Many deep networks for image recognition are trained on ImageNet
- ImageNet contains a large number of training images with wide diversity
- Using 1M+ images for training is typical
- Days worth of training time
- ***What can you do if your dataset is different from ImageNet content?***
  - ***Similar datasets do not exist for aerial / overhead / ISR data***



14M+ Images
21K+ Categories

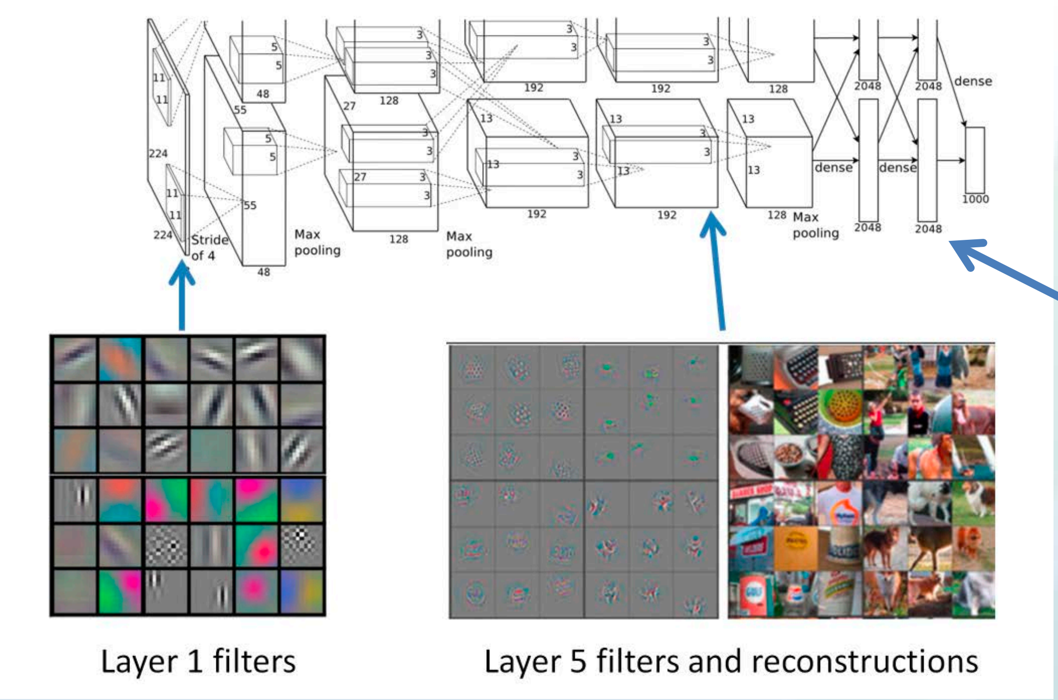
Training Alternatives

- **Train a network from scratch**
  - Requires large training volume, significant ground truth, computation time (days), experimental iteration
- **Refine an existing network**
  - Still requires potentially large training volume with significant ground truth
  - Relies on visual features being similar across datasets – open question
- **Simulation**
  - Simulated scenes can provide both training data and labels (known from underlying model)
  - Required level of fidelity is unknown
- **Generative models**
  - Training process tries to reproduce the input imagery
  - Hopefully produces features useful for discrimination

Training deep networks is still an art

Deep Learning Image Descriptors

- AlexNet or any CNN can be used as a generic image descriptor



Layer 7
Fully connected
4096 dimensions

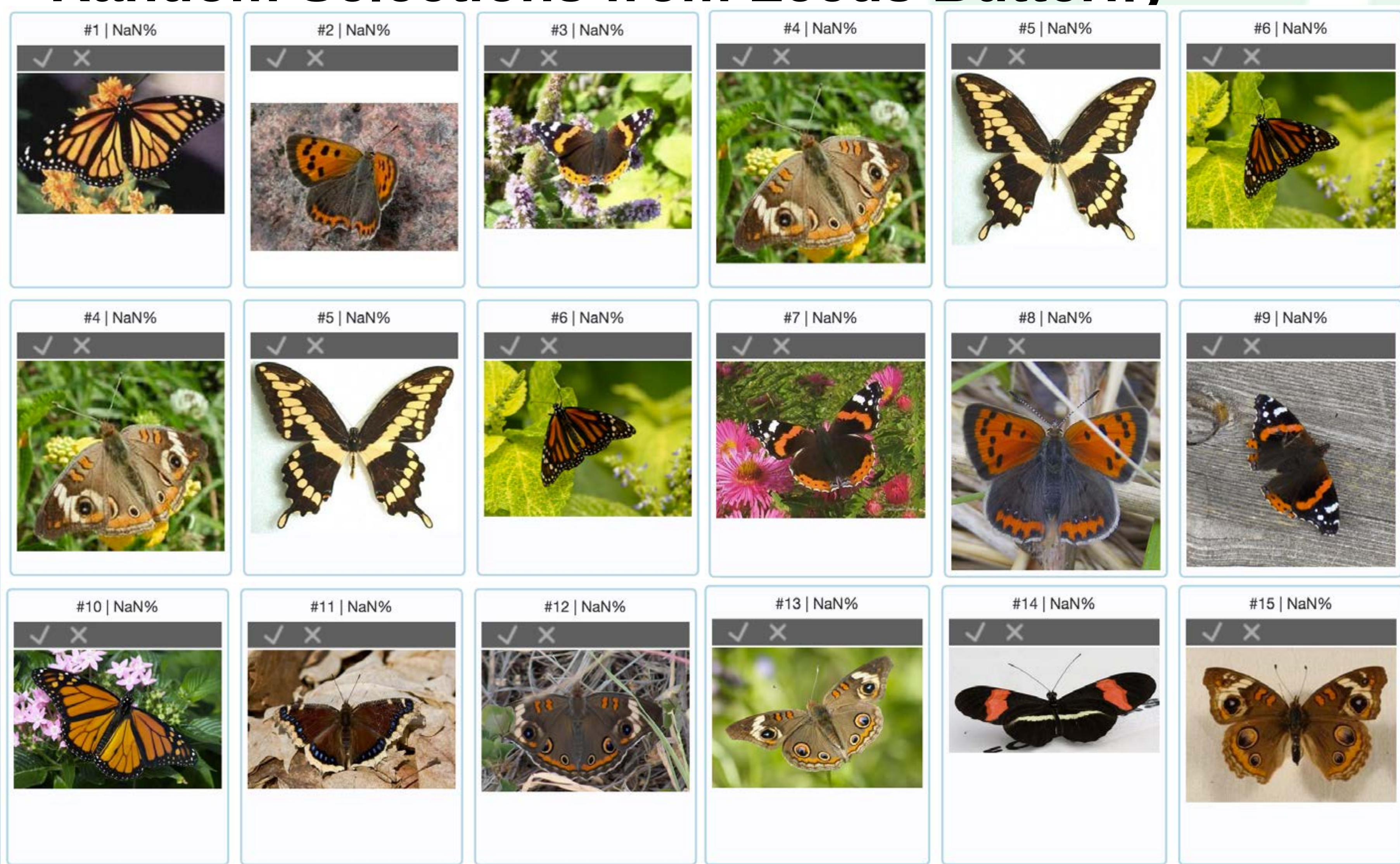
Image Query

Dataset contains 832 images with 55-100 images per type.

Start IQR with a single positive examplar

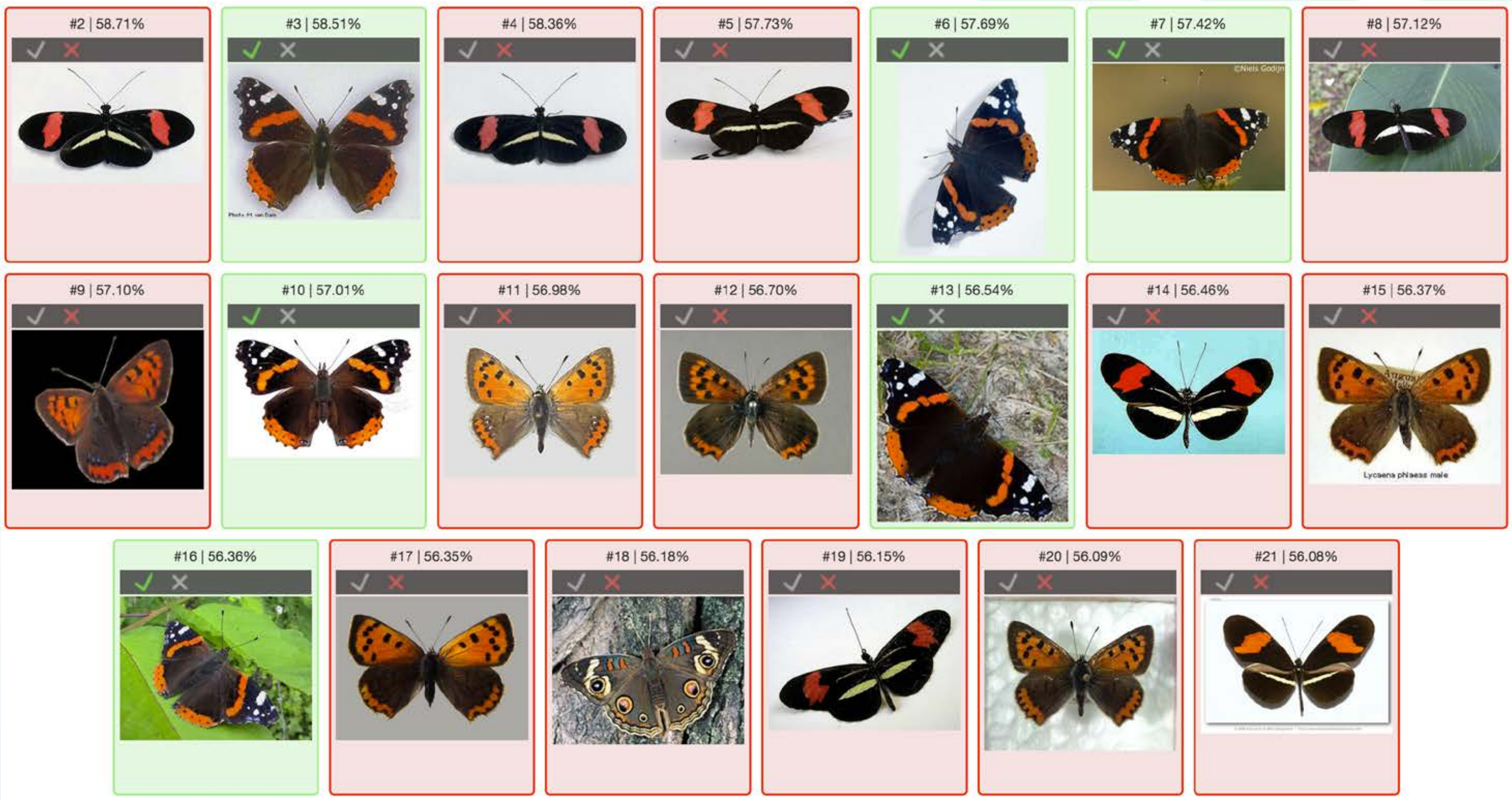
Use CAFFE AlexNet Layer 7 as an image descriptor

Random Selections from Leeds Butterfly

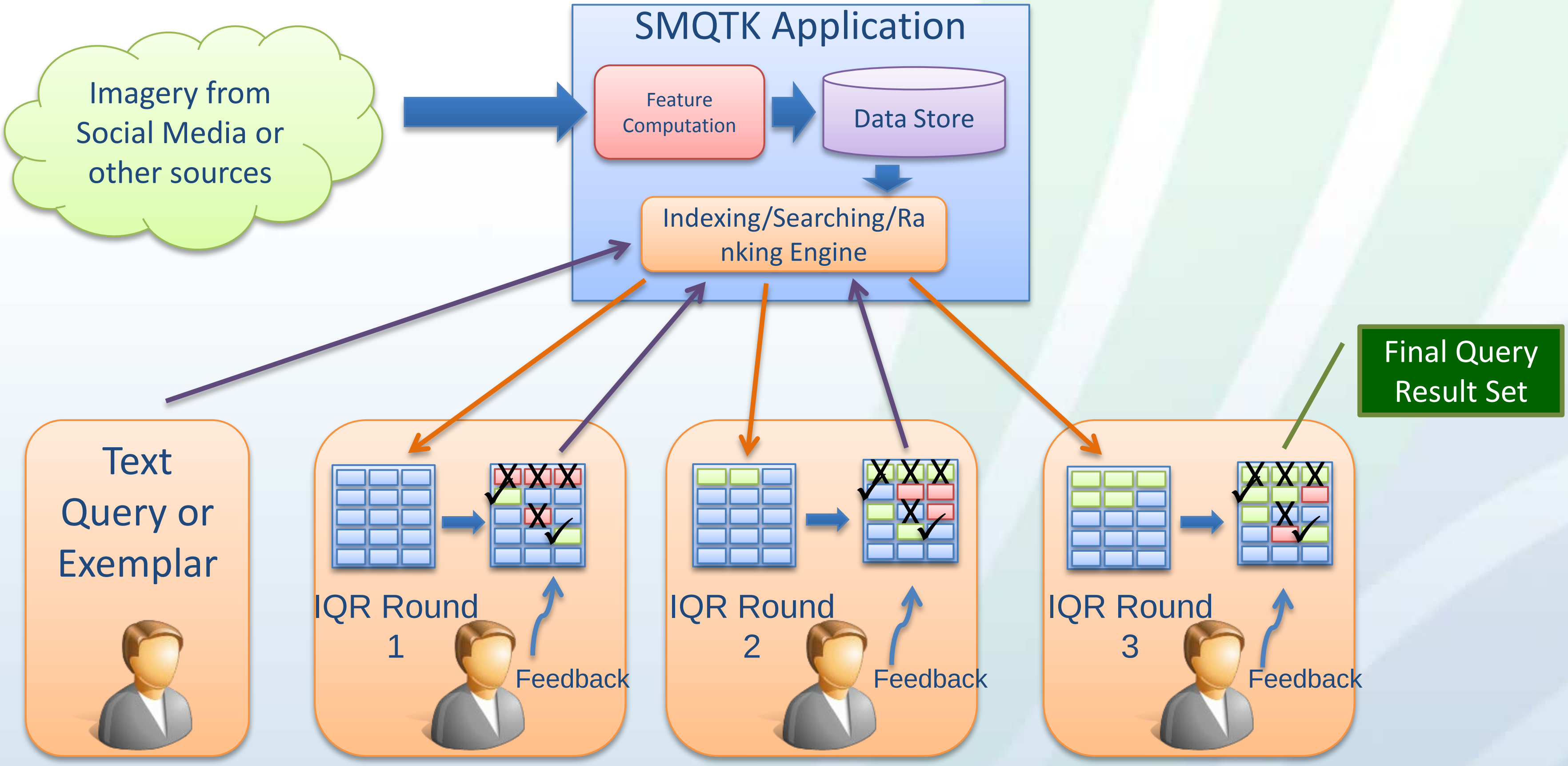


Josiah Wang, Katja Markert, and Mark Everingham
Learning Models for Object Recognition from Natural Language Descriptions
In *Proceedings of the 20th British Machine Vision Conference (BMVC2009)*

Results from Single Exemplar

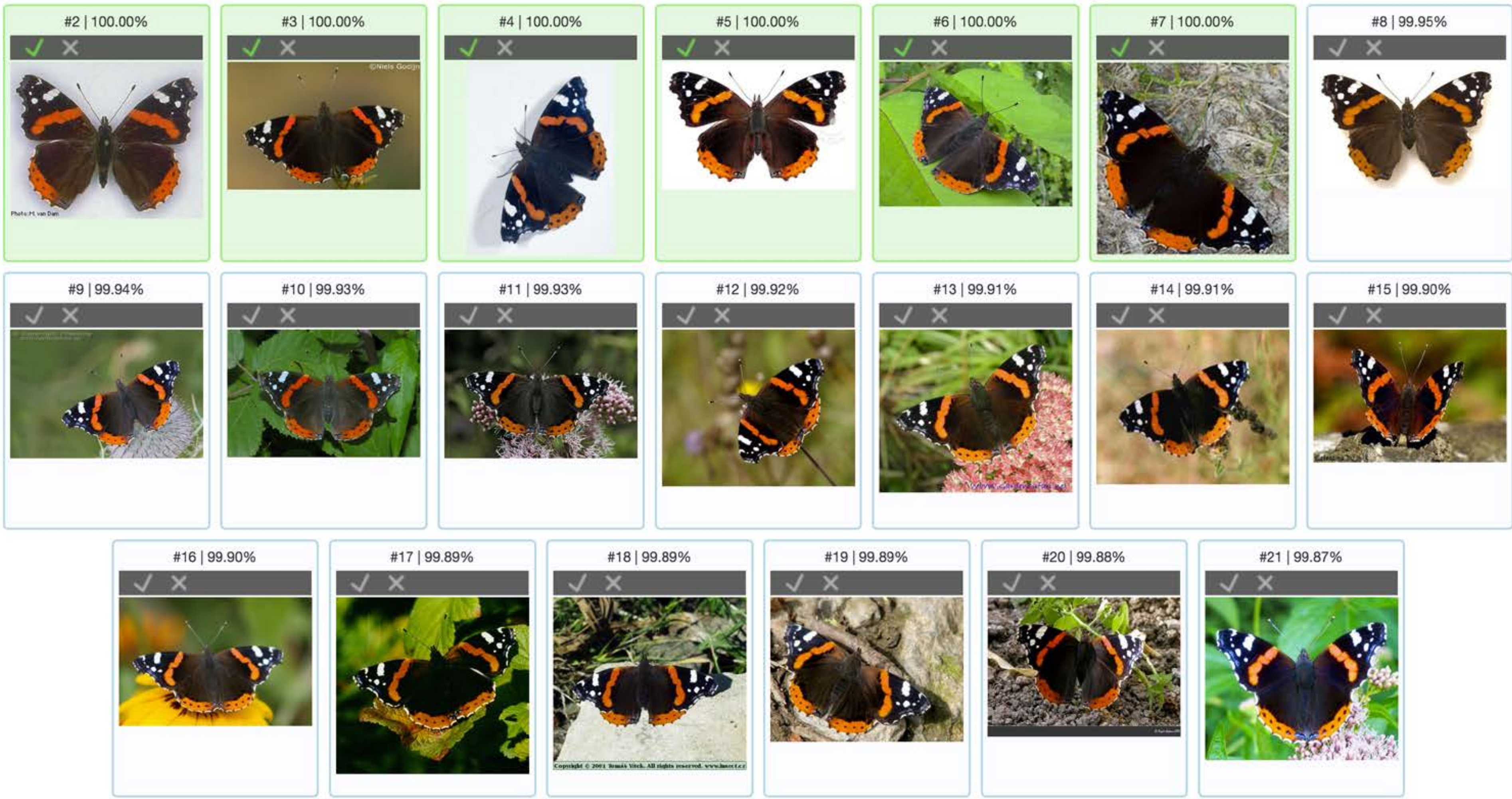


Interactive Query Refinement

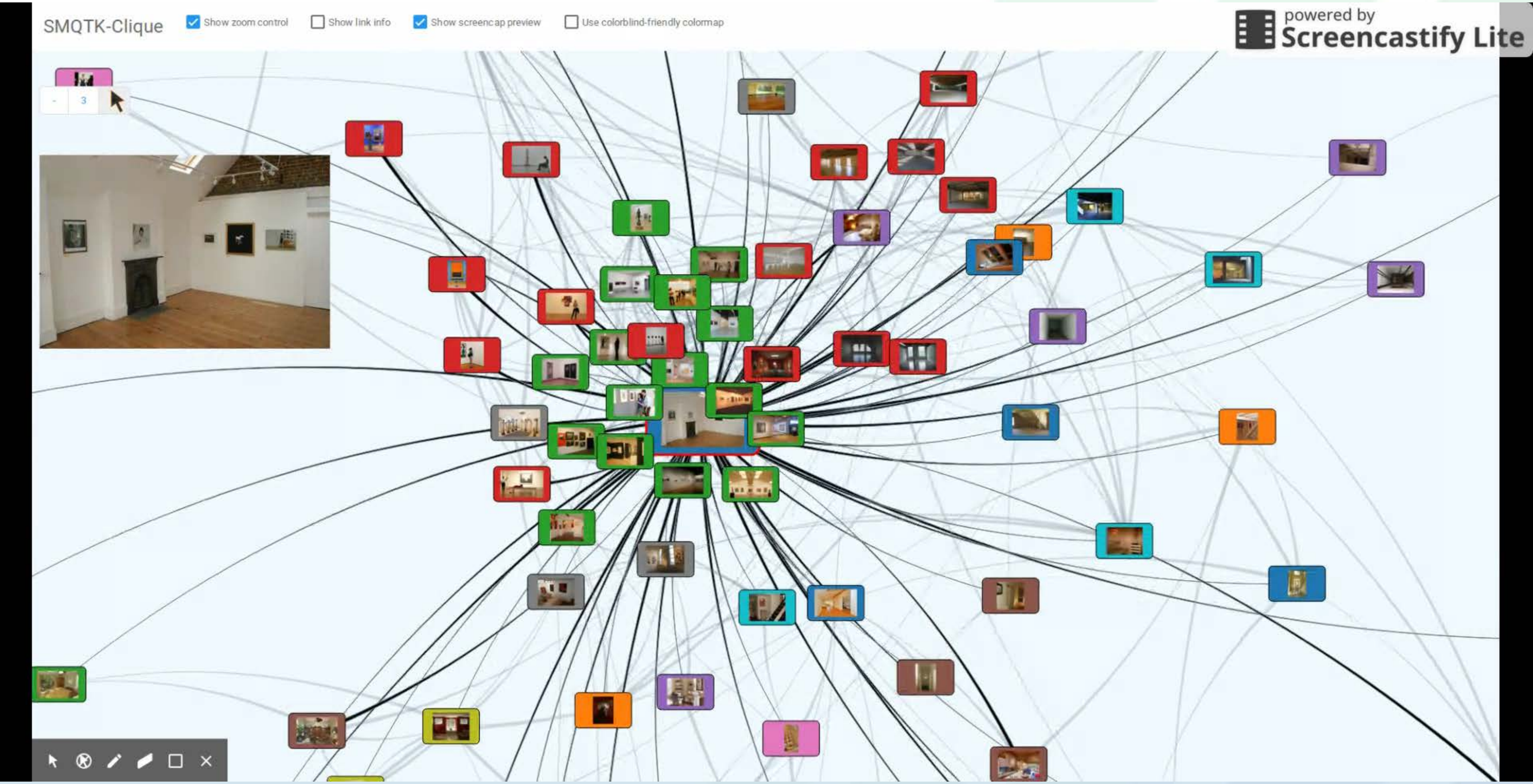


Kitware

One Refinement Based on Adjudications from



Visualization of Image-Induced Networks



Social Multimedia Query ToolKit (SMQTK)

- Indexing, Searching and Query Refinement on any images
- Rapid query times from ITQ-based indexing
- Plugin based architecture allows rapid prototyping and experimentation
- Open source at KWIVER.org
- Web Based Sample Apps

FMV vs. AlexNet data

ImageNet “automobile”

AlexNet: 224x224 chips

A.P. Hill “vehicle”

FMVNet: 96x96 chips

http://image-net.org/synset?wnid=n02958343#

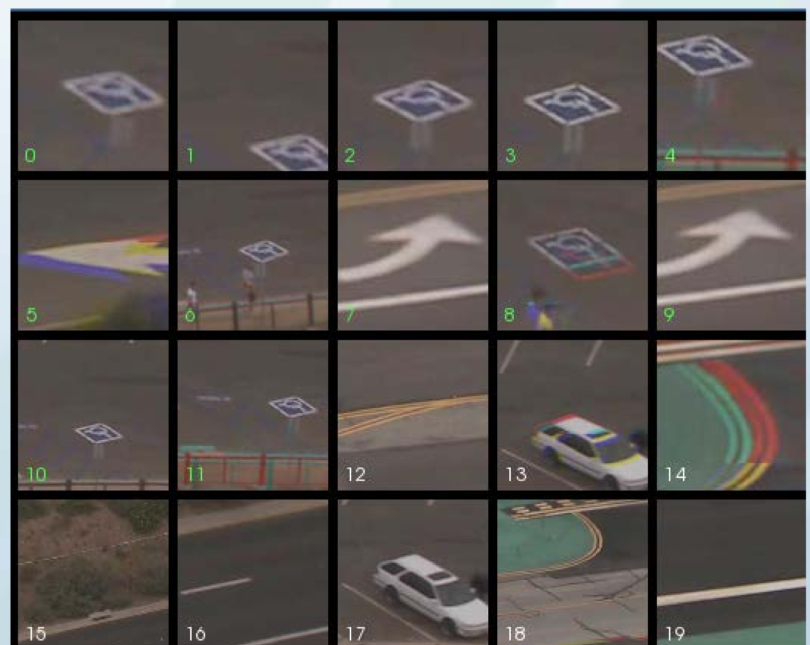
Neovision data w/ VIRAT indexing & IQR

Neovision data ingested into VIRAT framework, using FMVNet FC7 descriptors and motion & saliency detectors.

Query

(no IQR)

Query



(w/ IQR)

Data: Neovision
CNN: FMVNet
Detector: Motion & Saliency
Descriptors: CNN FC7
Indexing: VIRAT

Kitware

Neovision in SMQTK

Query

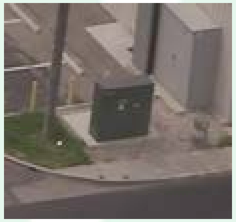
Results after a few rounds of IQR and a re-query of the database

Data: Neovision
CNN: AlexNet
Detector: Windowing
Descriptors: AlexNet
Indexing: SMQTK

Neovision in SMQTK

- Results obtained with IQR but no re-querying of the database.
- Several rounds were required to train the model away from cars and windows.

Query



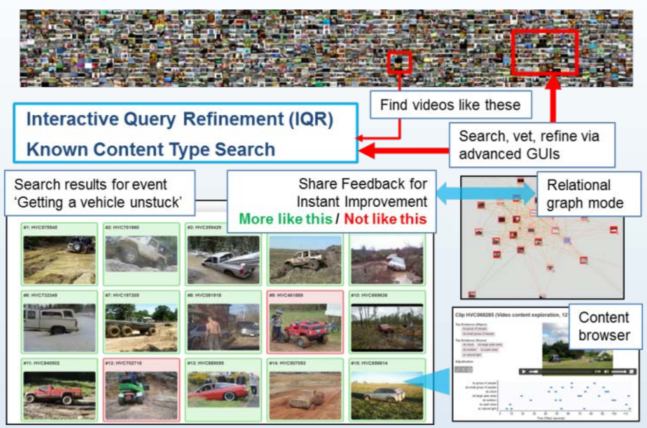
Data: Neovision
CNN: AlexNet
Detector: Windowing
Descriptors: AlexNet
Indexing: SMQTK

KWIVER.org

Kitware Image and Video Exploitation and Retrieval Toolkit

An Open Source, production-quality video analytics toolkit

Social Multimedia Query ToolKit



VIBRANT: *Video and Image-Based Retrieval and Analysis Toolkit*

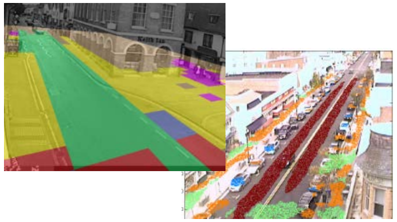
Motion-imagery Aerial Photogrammetry Toolkit

Kitware

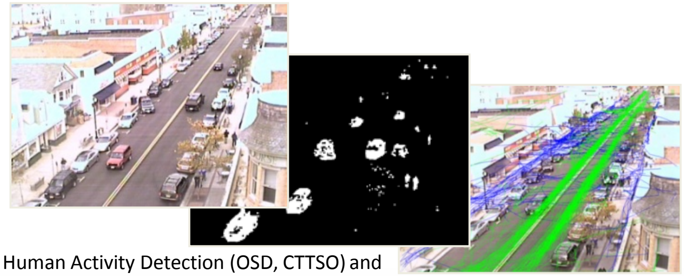
Summary

- Dramatic, disruptive advances in deep learning for computer vision are fueled by:
  - Big Data
  - GPU computation
  - Open source software
- Various tricks can greatly reduce training data requirements
- The ISR community should rapidly adopt deep learning for sensor exploitation problems
- To learn more, come to Honolulu on July 21-26 for the IEEE/CVF Conference on Computer Vision and Pattern Recognition

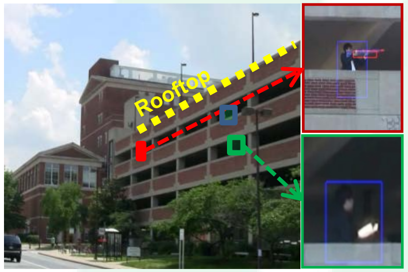
Questions?



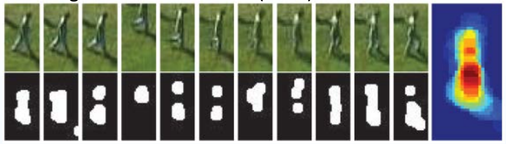
Object and Building Recognition by Function (DARPA)



Human Activity Detection (OSD, CTTSO) and Tracking in Wide-Area Video (AFRL)

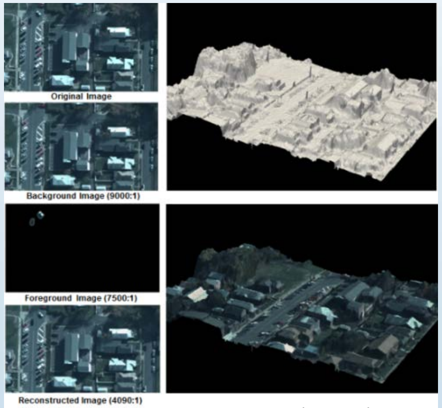


Threat Detection in Video (DARPA)

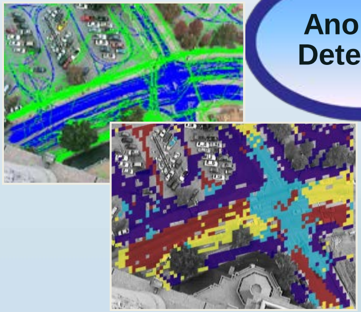


Content-based Video Retrieval by Actions (DARPA VIRAT)

Complex Event Recognition in Internet Videos (GENIE)

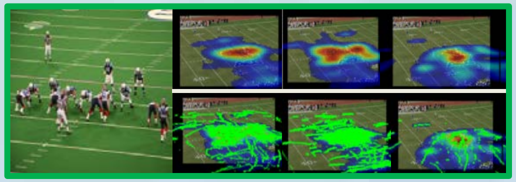


3D model-based video compression (DARPA) and super-resolved 3D reconstruction (DARPA)



Normalcy Modeling and Anomaly Detection (DARPA PANDA and PerSEAS)

Football Play Recognition (DARPA CARVE)



Wide-Area Motion Imagery Event, Anomaly and Activity Detection (OSD Data to Decisions, DARPA PerSEAS)