CAP 6635 – Artificial Intelligence

Lecture 5: Netflix Prize and beyond (Challenges in AI) (Part 2)

Oge Marques, PhD
Professor
College of Engineering and Computer Science
College of Business

FLORIDA ATLANTIC UNIVERSITY

@ProfessorOge

ProfessorOgeMarques

Previously on CAP 6635…

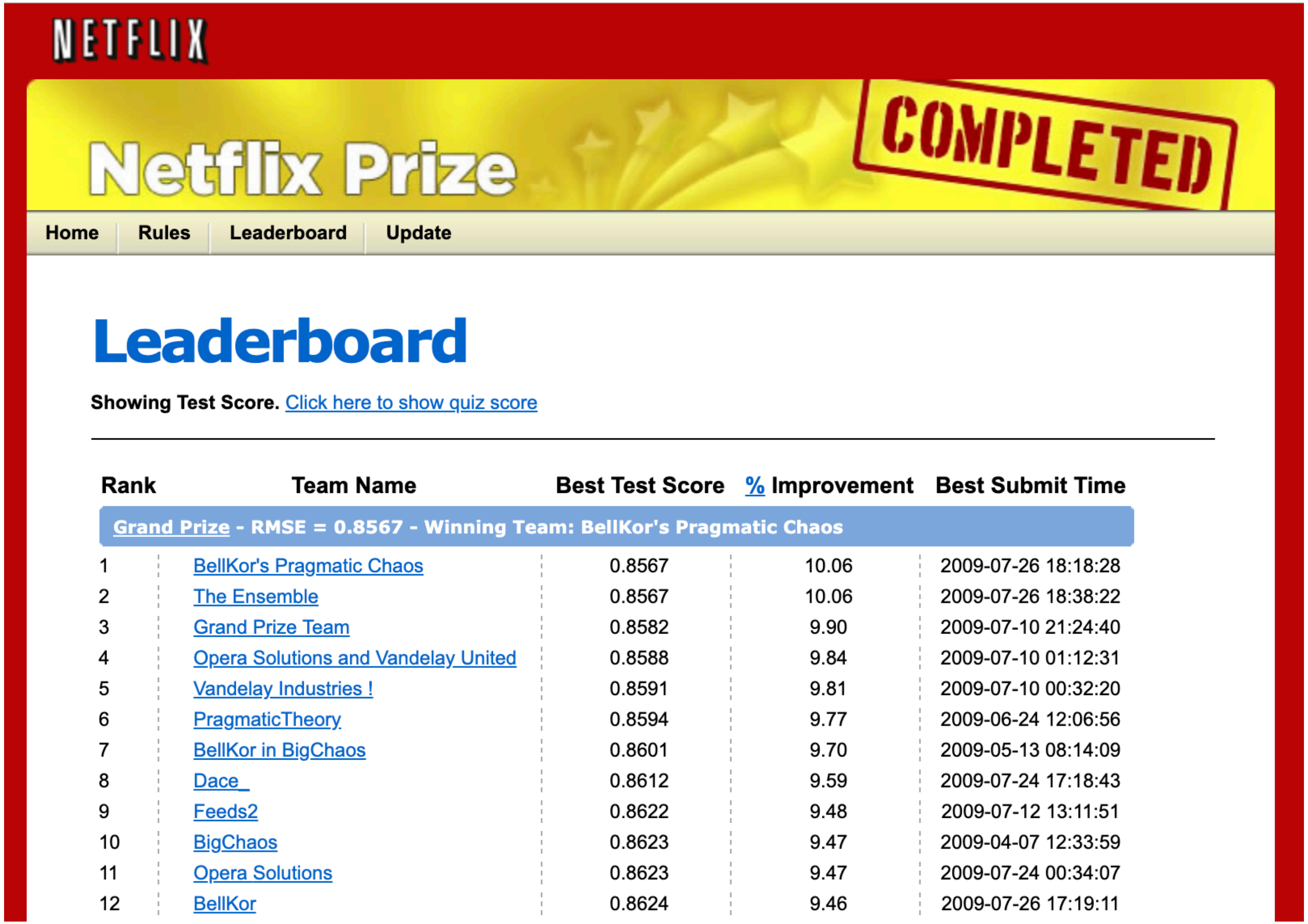
Take-home message

Progress in Artificial Intelligence (AI) (and related areas) **has been accelerated by challenges, prizes, hackathons, etc.**

These are also a **great source of learning opportunities.**

Netflix and the recommendation-engine challenge

https://www.netflixprize.com/



NETFLIX

Netflix Prize

COMPLETED

Home | Rules | Leaderboard | Update

Leaderboard

Showing Test Score. Click here to show quiz score

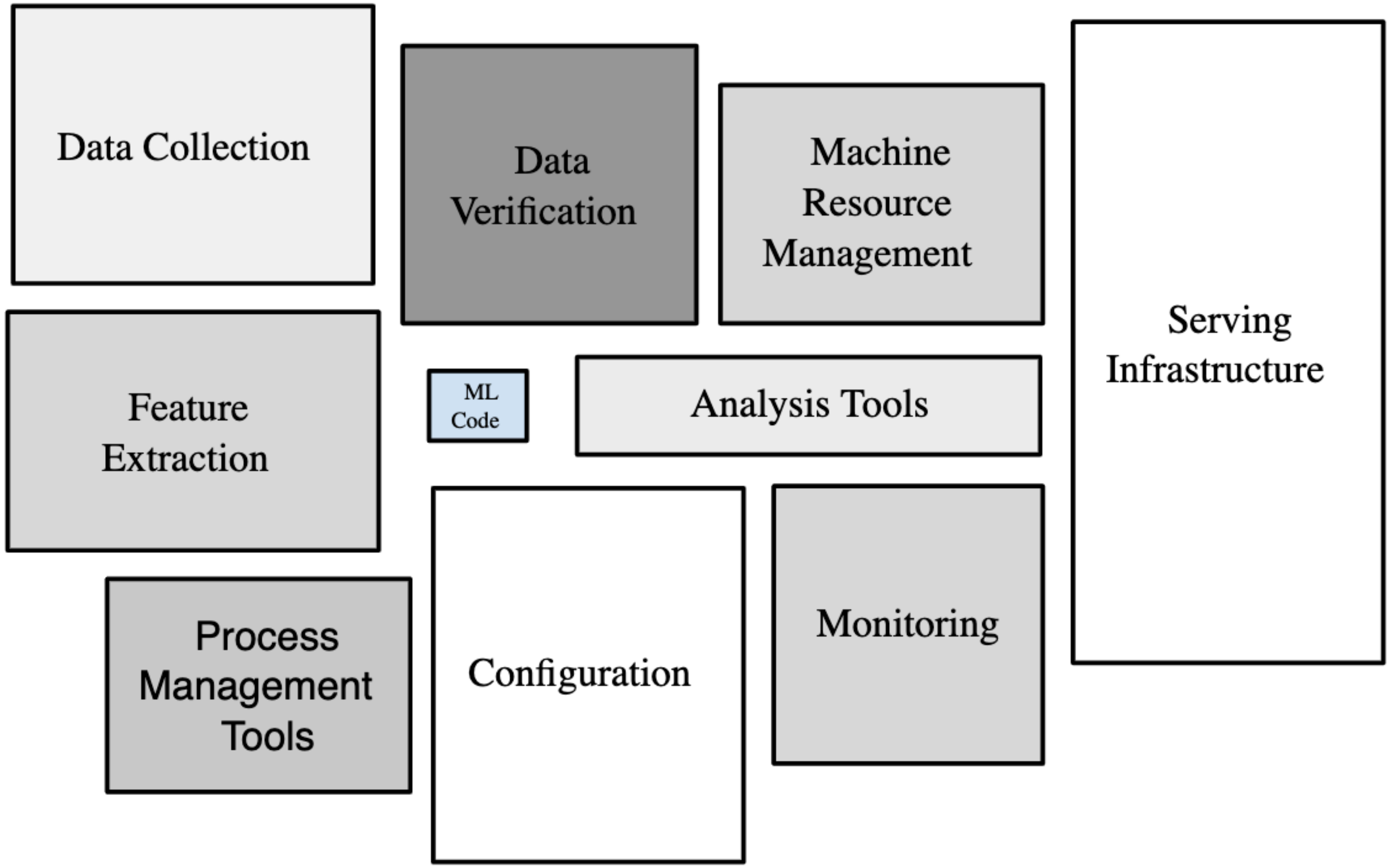
Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

Netflix prize: lessons

- Alliances among teams became essential to success
- It paid off (for Netflix) – many man-hours at reduced cost
- Scientific advancement – and the art of reporting "what didn't work"
- Timing is everything
- Winning the challenge is just the beginning… and it might not mean much beyond collecting the prize and bragging about it.

Real-world production ML system

"The ML code is at the heart of a real-world ML production system, but that box often represents only <u>5% or less of the overall code</u> of that total ML production system."



Source: https://developers.google.com/machine-learning/crash-course/production-ml-systems

Ongoing AI challenges

Examples of other challenges

Computer Vision Community

Middlebury Computer Vision

http://vision.middlebury.edu/

Middlebury

Microsoft Research

NSF

Pascal VOC

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

http://host.robots.ox.ac.uk/pascal/VOC/

Int J Comput Vis (2015) 111:98–136
DOI 10.1007/s11263-014-0733-5

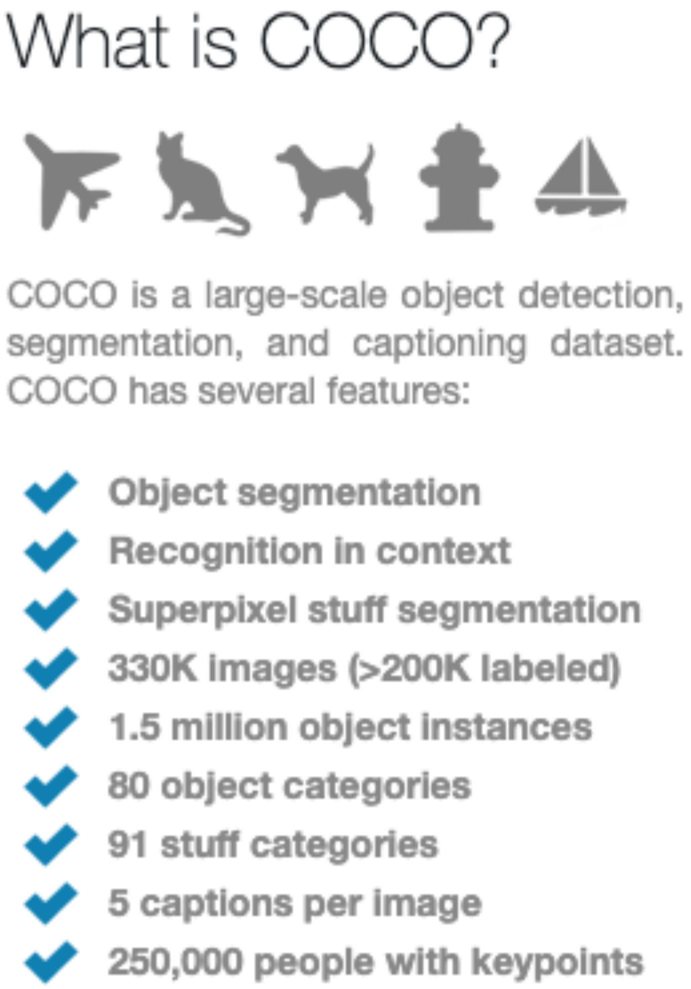
The PASCAL Visual Object Classes Challenge: A Retrospective

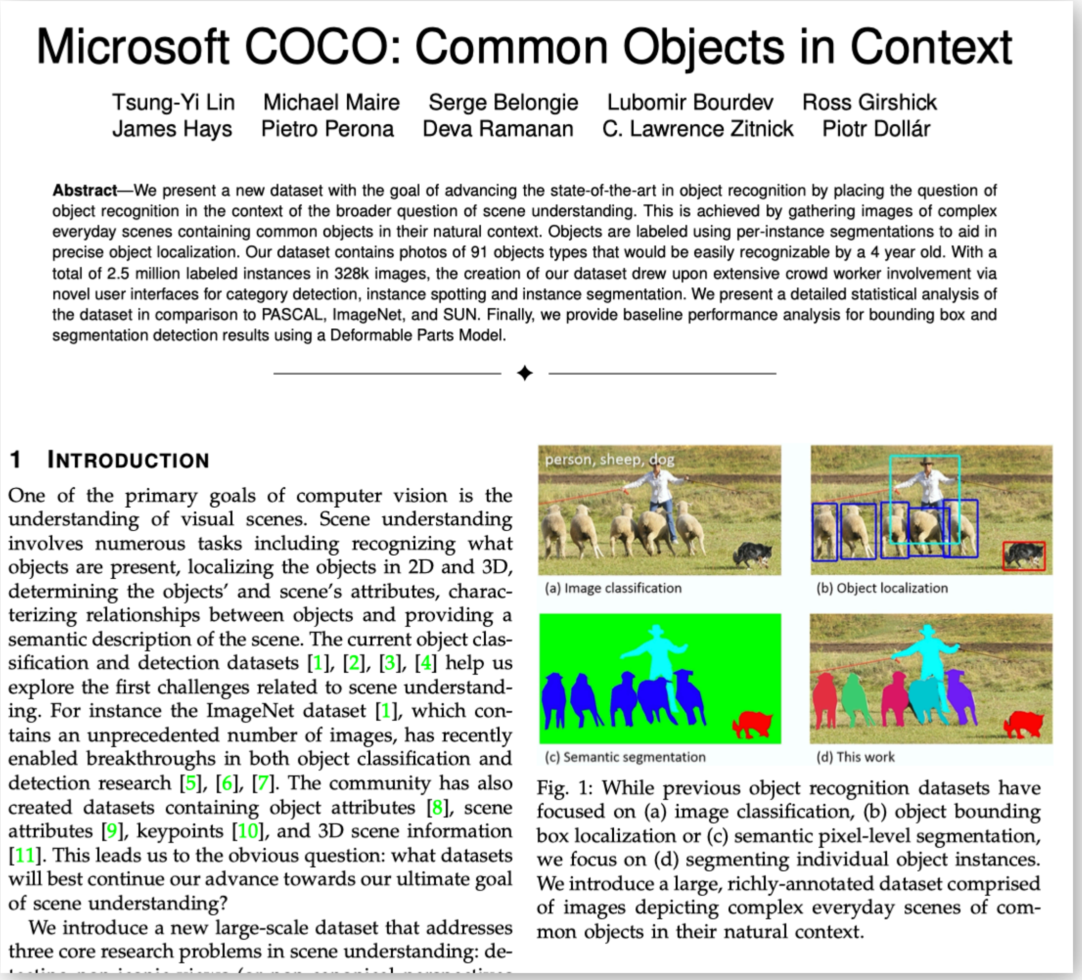
**Mark Everingham · S. M. Ali Eslami · Luc Van Gool ·
Christopher K. I. Williams · John Winn ·
Andrew Zisserman**

Available on Canvas

Microsoft COCO

http://cocodataset.org/





Microsoft COCO: Common Objects in Context

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár

Abstract—We present a new dataset with the goal of advancing the state-of-the-art in object recognition by placing the question of object recognition in the context of the broader question of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance segmentations to aid in precise object localization. Our dataset contains photos of 91 objects types that would be easily recognizable by a 4 year old. With a total of 2.5 million labeled instances in 328k images, the creation of our dataset drew upon extensive crowd worker involvement via novel user interfaces for category detection, instance spotting and instance segmentation. We present a detailed statistical analysis of the dataset in comparison to PASCAL, ImageNet, and SUN. Finally, we provide baseline performance analysis for bounding box and segmentation detection results using a Deformable Parts Model.

1 Introduction

One of the primary goals of computer vision is the understanding of visual scenes. Scene understanding involves numerous tasks including recognizing what objects are present, localizing the objects in 2D and 3D, determining the objects' and scene's attributes, characterizing relationships between objects and providing a semantic description of the scene. The current object classification and detection datasets [1], [2], [3], [4] help us explore the first challenges related to scene understanding. For instance the ImageNet dataset [1], which contains an unprecedented number of images, has recently enabled breakthroughs in both object classification and detection research [5], [6], [7]. The community has also created datasets containing object attributes [8], scene attributes [9], keypoints [10], and 3D scene information [11]. This leads us to the obvious question: what datasets will best continue our advance towards our ultimate goal of scene understanding?

We introduce a new large-scale dataset that addresses three core research problems in scene understanding: de-

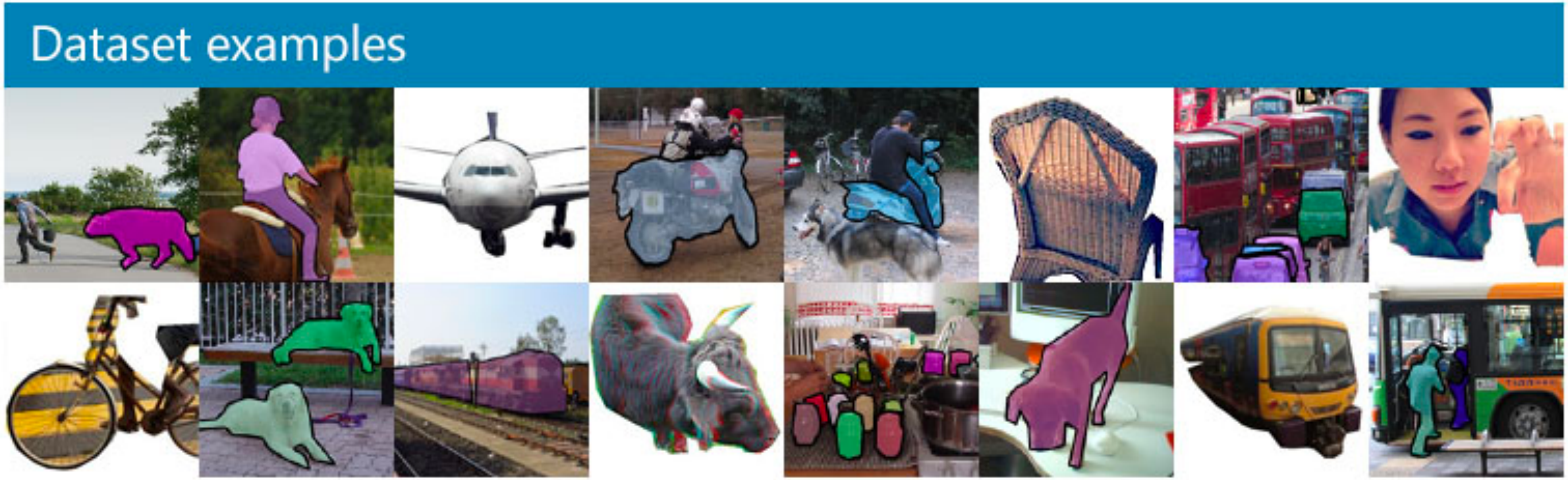
(a) Image classification (b) Object localization

(c) Semantic segmentation (d) This work

Fig. 1: While previous object recognition datasets have focused on (a) image classification, (b) object bounding box localization or (c) semantic pixel-level segmentation, we focus on (d) segmenting individual object instances. We introduce a large, richly-annotated dataset comprised of images depicting complex everyday scenes of common objects in their natural context.

Available on Canvas

Microsoft COCO http://cocodataset.org/



Caltech 101 and Caltech 256

- http://www.vision.caltech.edu/Image_Datasets/Caltech101/
- http://www.vision.caltech.edu/Image_Datasets/Caltech256/

New material starts here...

ImageNet

IMAGENET

14,197,122 images, 21841 synsets indexed

- **ImageNet** is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images.

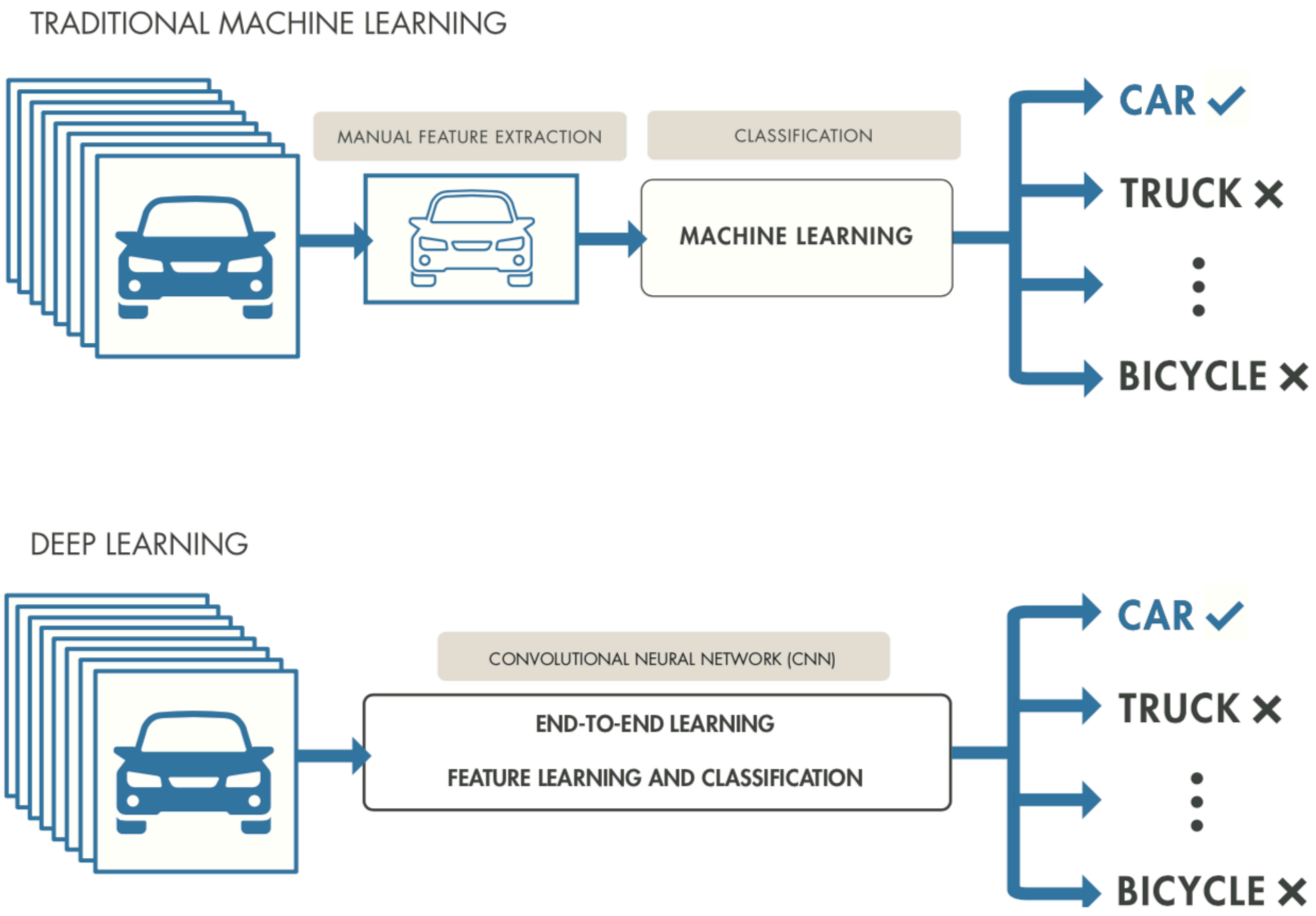
- WordNet® is a large lexical database of English.
  - Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (*synsets*), each expressing a distinct concept.
  - Synsets are interlinked by means of conceptual-semantic and lexical relations.
  - The resulting network of meaningfully related words and concepts can be navigated with the browser.

- Explore ImageNet: http://image-net.org/explore

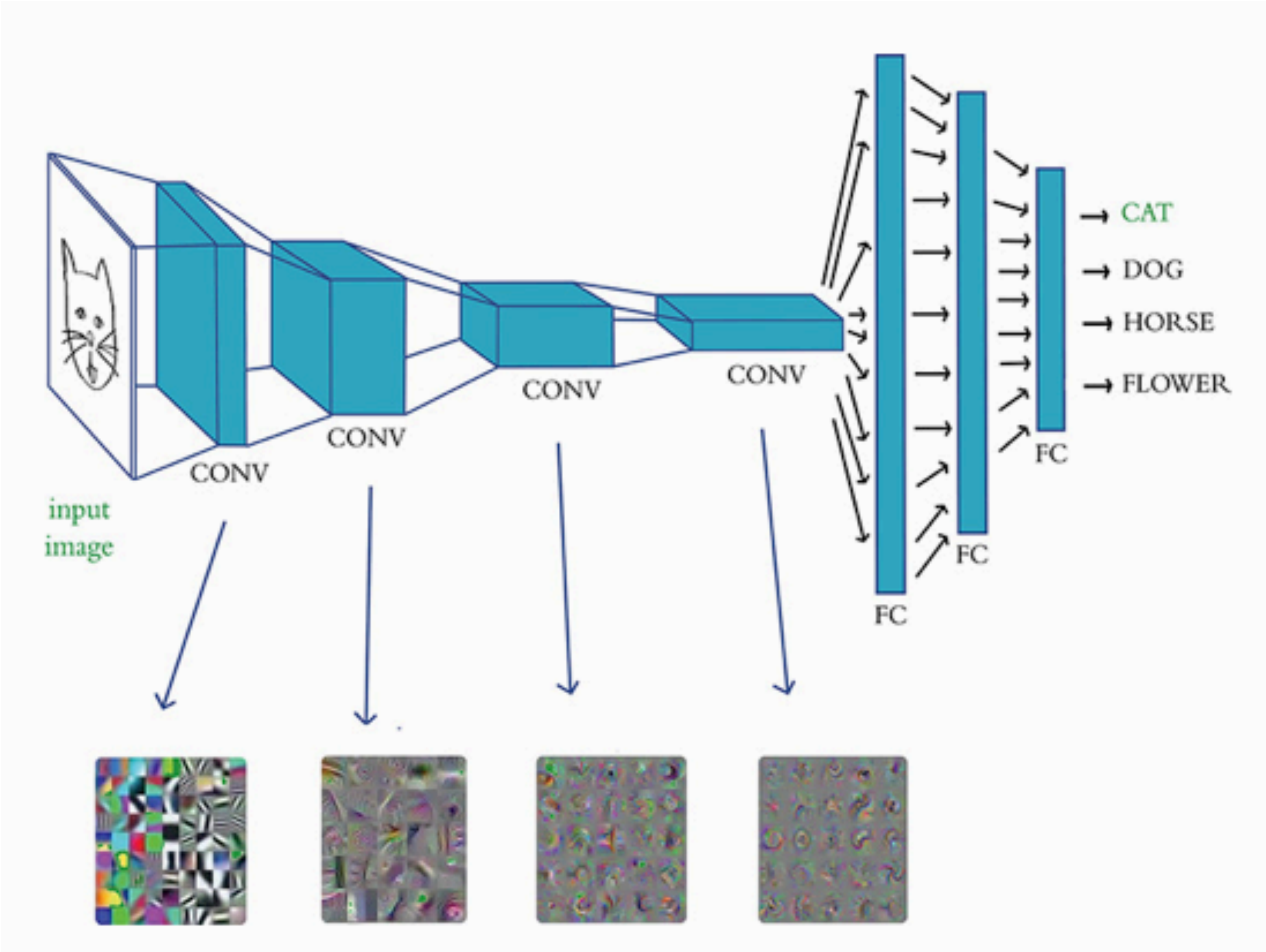
ImageNet LSVRC

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

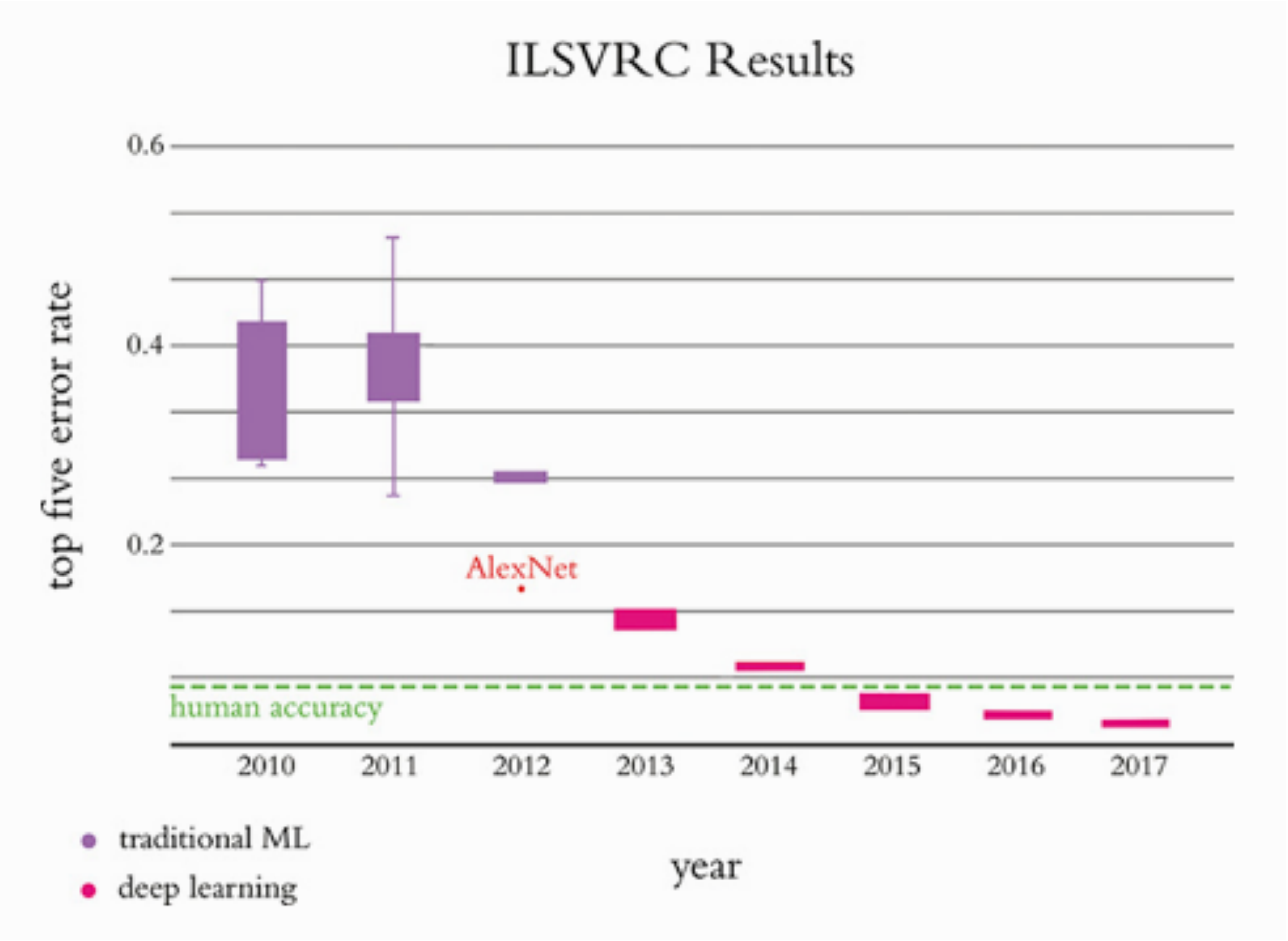
Deep learning vs. machine learning

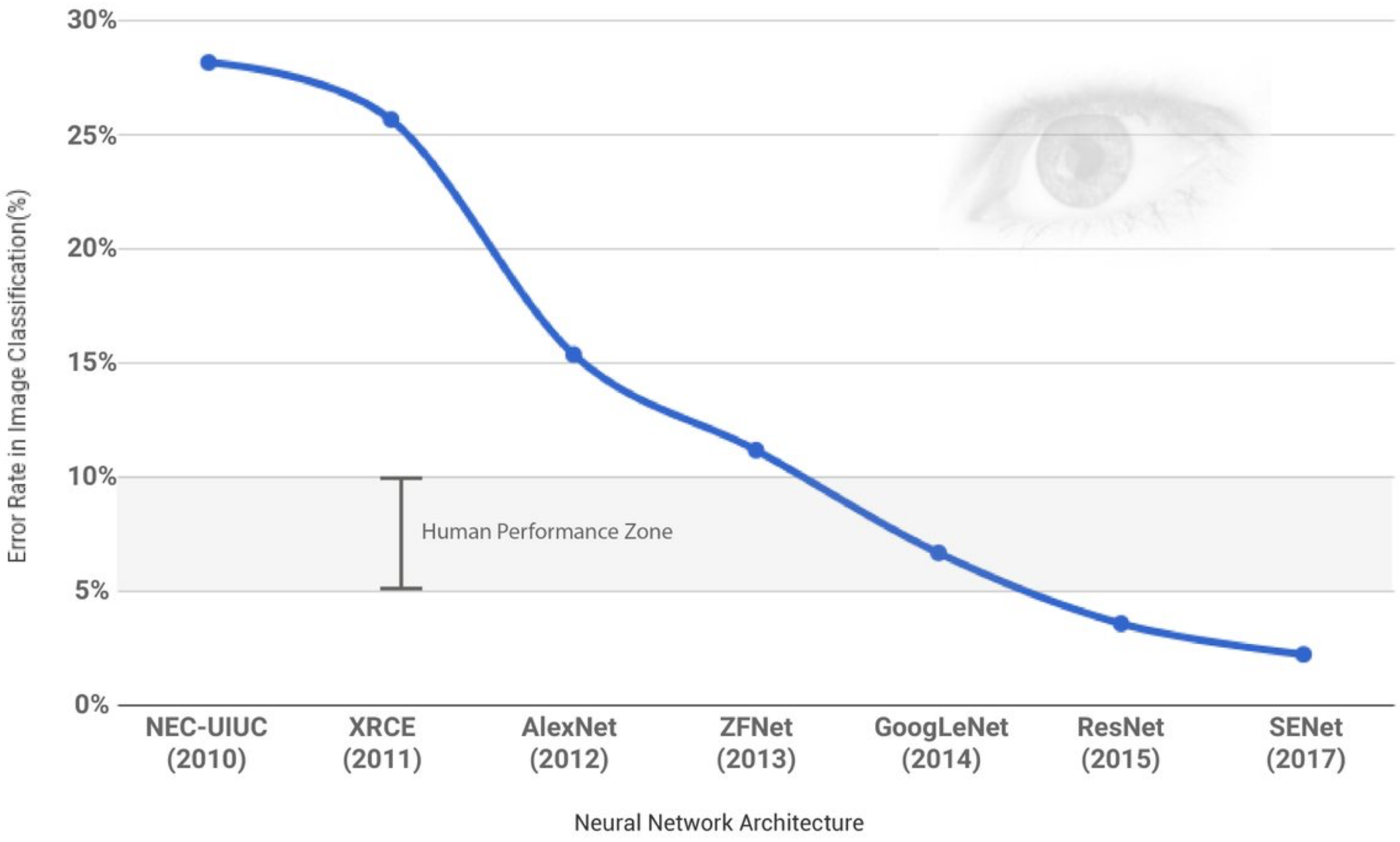


Source: (Mathworks 2017)



input image
CONV
CONV
CONV
CONV
FC
FC
FC
CAT
DOG
HORSE
FLOWER

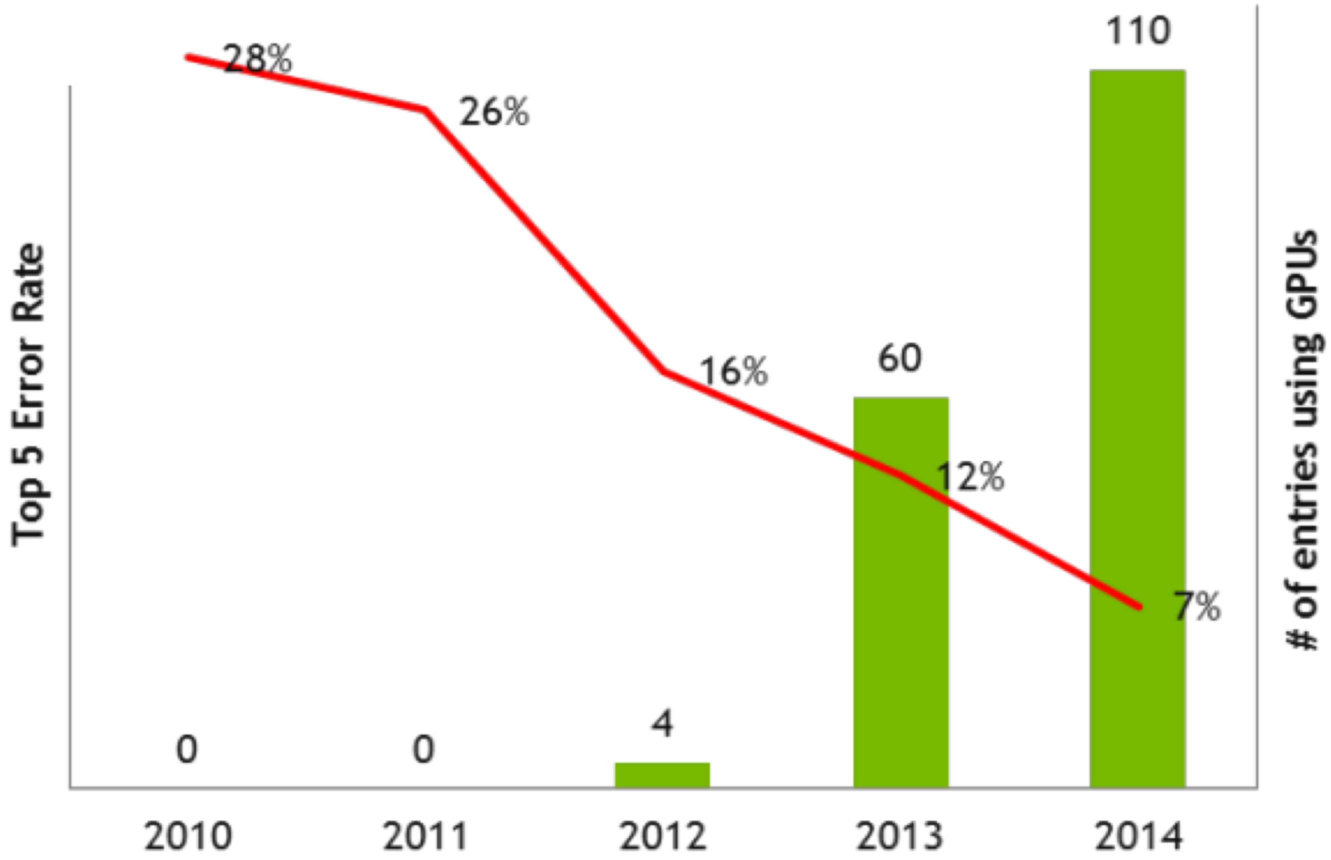

ILSVRC Results
0.6
0.4
0.2
top five error rate
AlexNet
human accuracy
2010
2011
2012
2013
2014
2015
2016
2017
year
traditional ML
deep learning



See also: https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/


GEFORCE GTX
GTX 1080

IMAGENET


Top 5 Error Rate
# of entries using GPUs
28%
26%
16%
12%
7%
0
0
4
60
110
2010
2011
2012
2013
2014

NVIDIA

248.56 USD −1.02 (0.41%)

Closed: Mar 21, 7:59 PM EDT · Disclaimer

After hours 248.78 +0.22 (0.089%)

1 day | 5 days | 1 month | 1 year | **5 years** | Max

Open	249.32	Div yield	0.24%
High	252.00	Prev close	249.58
Low	247.33	52-wk high	254.50
Mkt cap	150.38B	52-wk low	95.49
P/E ratio	53.87		

Information Retrieval Community

ImageCLEF / LifeCLEF - Multimedia Retrieval in CLEF

https://www.imageclef.org/

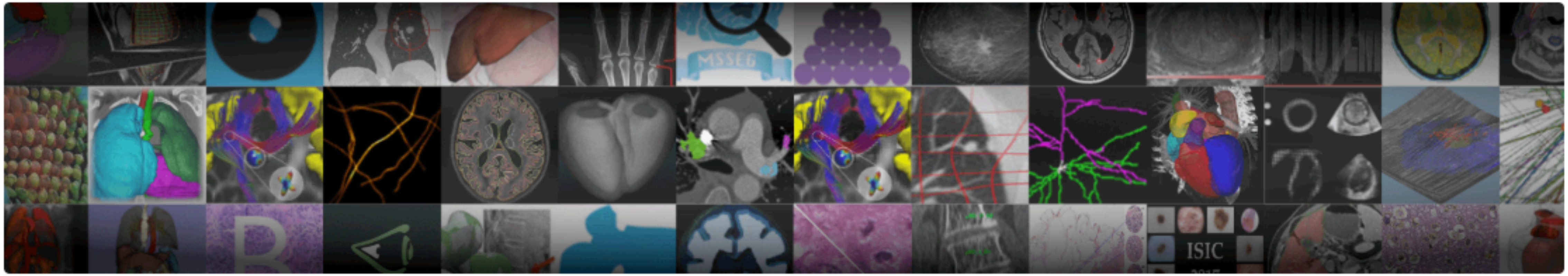
ImageCLEF 2020

- ImageCLEFlifelog
- ImageCLEFcoral
- ImageCLEFdrawnUI
- ImageCLEFmedical

LifeCLEF 2020

- BirdCLEF 2020
- GeoLifeCLEF 2020
- PlantCLEF 2020
- SnakeCLEF 2020

Medical Image Analysis Community



Grand Challenges in Biomedical Image Analysis

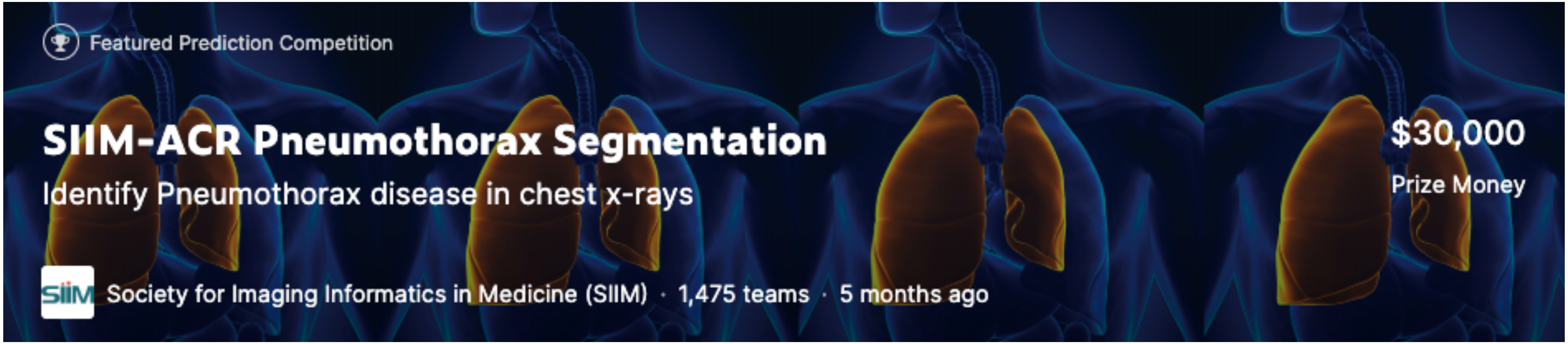
- Statistics (as of Jan 2020)
  - 198 projects
  - 5 active (due in 2020)

https://grand-challenge.org/

40 SIIM
1980-2020

ACR®
AMERICAN COLLEGE OF
RADIOLOGY
QUALITY IS OUR IMAGE

SIIM-ACR Pneumothorax Segmentation

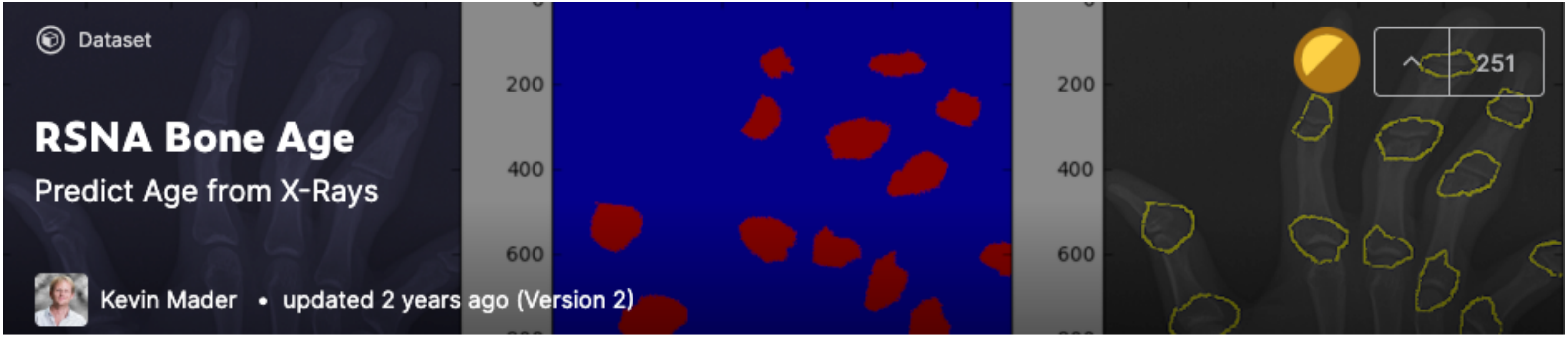


https://siim.org/page/pneumothorax_challenge

RSNA®

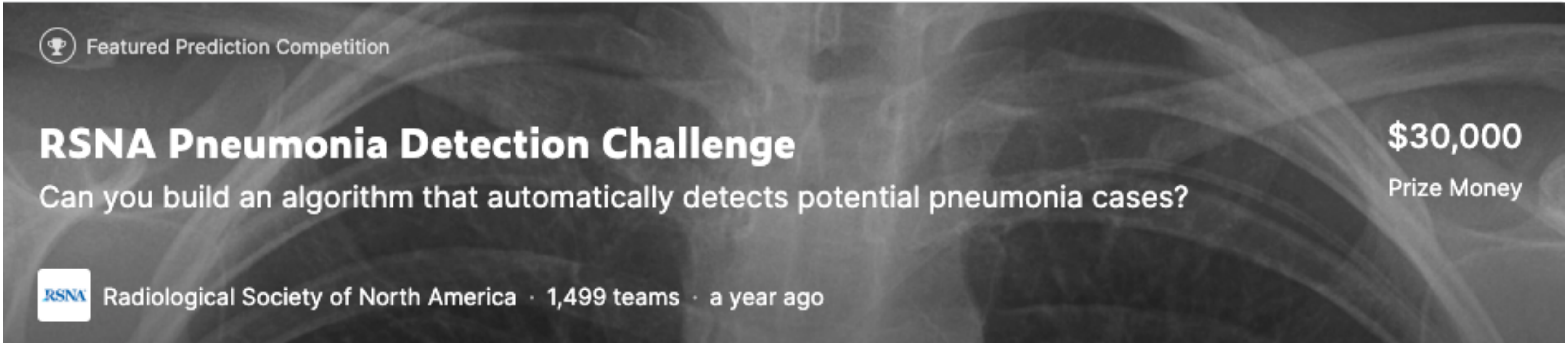
Radiological Society
of North America

RSNA 2017



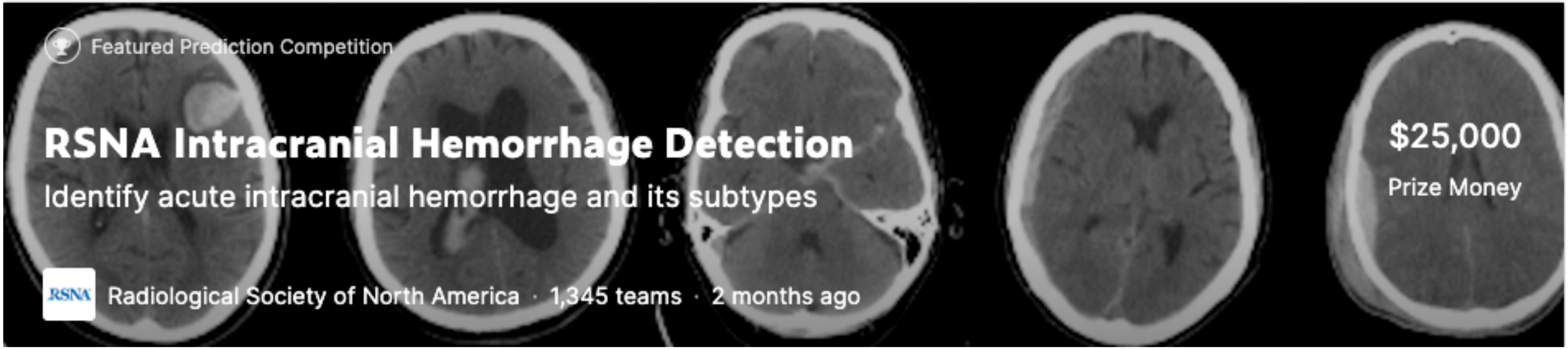
https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/RSNA-Pediatric-Bone-Age-Challenge-2017

RSNA 2018



https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018

RSNA 2019



https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge

Machine Learning Community

Challenges in AI and ML

Kaggle

https://www.kaggle.com/

13 Active Competitions

Deepfake Detection Challenge
Identify videos with facial or voice manipulations
Featured · Code Competition · 2 months to go · video data, online video
$1,000,000
1,206 teams

2019 Data Science Bowl
Uncover the factors to help measure how young children learn
Featured · Code Competition · 2 hours to go · children, learning, education, video games
$160,000
3,518 teams

TensorFlow 2.0 Question Answering
Identify the answers to real user questions about Wikipedia page content
Featured · Code Competition · 2 hours to go · text data, text mining
$50,000
1,240 teams

Google QUEST Q&A Labeling
Improving automated understanding of complex question answer content
Featured · Code Competition · 19 days to go · text data, nlp
$25,000
1,072 teams

Real or Not? NLP with Disaster Tweets
Predict which Tweets are about real disasters and which ones are not
Getting Started · Ongoing · text data, binary classification
$10,000
1,795 teams

Bengali.AI Handwritten Grapheme Classification
Classify the components of handwritten Bengali
Research · Code Competition · 2 months to go · multiclass classification, image data
$10,000
727 teams

Challenges in AI and ML

Kaggle

https://www.kaggle.com/

Digit Recognizer
Learn computer vision fundamentals with the famous MNIST data
Getting Started · Ongoing · image data, tabular data, multiclass classification, object identification
Knowledge
2,302 teams

Titanic: Machine Learning from Disaster
Start here! Predict survival on the Titanic and get familiar with ML basics
Getting Started · Ongoing · binary classification, tabular data, tutorial
Knowledge
15,702 teams

House Prices: Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting
Getting Started · Ongoing · tabular data, regression
Knowledge
5,151 teams

IMAGENET

ImageNet Object Localization Challenge
Identify the objects in images
Research · 10 years to go · object detection, image data
Knowledge
62 teams

Predict Future Sales
Final project for "How to win a data science competition" Coursera course
Playground · a year to go
Kudos
5,536 teams

Categorical Feature Encoding Challenge II
Binary classification, with every feature a categorical (and interactions!)
Playground · 2 months to go · categorical data, table games, binary classification
Swag
309 teams

Connect X
Connect your checkers in a row before your opponent!
Getting Started · Ongoing
Knowledge
315 teams

facebook Microsoft

https://ai.facebook.com/blog/deepfake-detection-challenge/

Motivating factors

Motivating factors

Money

Bragging rights / reputation

Relevant problems ➡️ impactful solutions

Visibility

Strategies

Strategies

- Teaming up (Netflix)
- Ensemble approaches
- Learning from past/closed/similar challenges
- Others (not all of them legit)

Indirect Benefits

Indirect benefits

- Access to **datasets / benchmarks**
- Connecting with **key people** in the field
- Reading **latest papers** etc.
- Promoting **reproducible research**
- Learning from (and helping to solve) **“real world” problems**

Radiology: Artificial Intelligence

SPECIAL REPORT

Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis Competitions

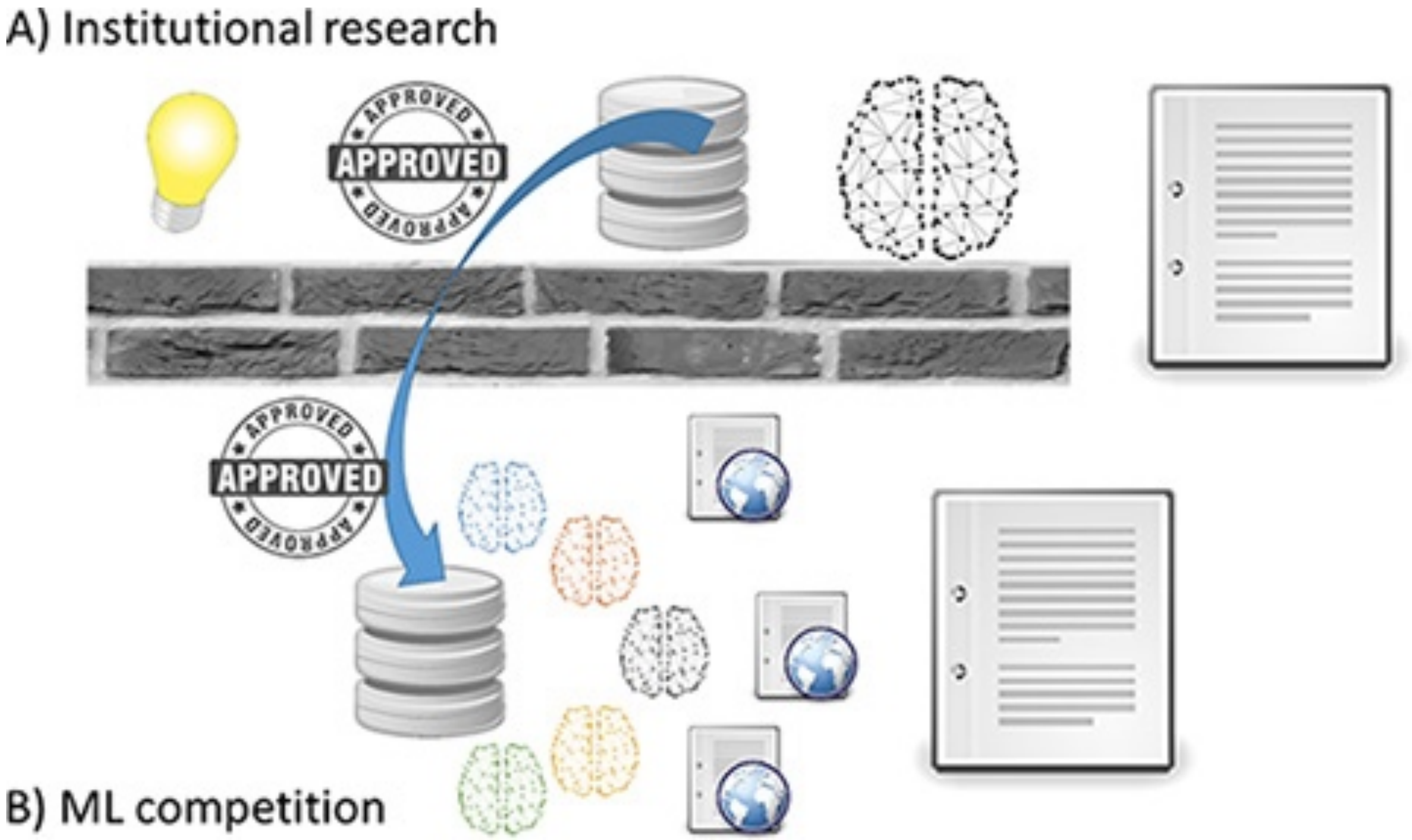
Luciano M. Prevedello, MD, MPH • Safwan S. Halabi, MD • George Shih, MD, MS • Carol C. Wu, MD • Marc D. Kohli, MD • Falgun H. Chokshi, MD • Bradley J. Erickson, MD, PhD • Jayashree Kalpathy-Cramer, PhD • Katherine P. Andriole, PhD • Adam E. Flanders, MD

From the Department of Radiology, The Ohio State University Wexner Medical Center, 395 West 12th Ave, 4th Floor, Room 422, Columbus, OH 43210 (L.M.P.); Department of Radiology, Stanford University School of Medicine, Stanford, Calif (S.S.H.); Department of Radiology, Weill Cornell Medical College, New York, NY (G.S.); Department of Diagnostic Radiology, University of Texas–MD Anderson Cancer Center, Houston, Tex (C.C.W.); Department of Radiology and Biomedical Imaging, University of California–San Francisco, San Francisco, Calif (M.D.K.); Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Ga (F.H.C.); Department of Radiology, Mayo Clinic, Rochester, Minn (B.J.E.); Department of Radiology and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, Mass (J.K.C.); Department of Radiology, Brigham and Women's Hospital, Massachusetts General Hospital and BWH Center for Clinical Data Science, Boston, Mass (K.P.A.); and Department of Radiology, Thomas Jefferson University Hospital, Philadelphia, Pa (A.E.F.). Received September 7, 2018; revision requested October 16; revision received December 6; accepted December 21. **Address correspondence to** L.M.P. (e-mail: *Luciano.Prevedello@osumc.edu*).

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2019; 1(1):e180031 • **https://doi.org/10.1148/ryai.2019180031** • **Content codes:** IN RS •

Available on Canvas



- Traditional model: hypothesis formulation, institutional review board approval, data acquisition, knowledge discovery, and manuscript submission
- Image competitions promote a parallel approach to knowledge discovery and dissemination in which multiple participants of varying disciplines find solutions to the same problem and openly share their discoveries on the web or manuscripts in a short time period.

Prevedello et al. RSNA:AI (2019)

Prevedello et al. (2019) : key takeaways

- Activities such as the competitions organized by RSNA may prove to be an **important way to address current roadblocks in applying AI to medical imaging** and to **increase the dialogue among radiologists and data scientists**, which serves to guide and move the field forward.

- Although competitions may help move research forward, **the field should still rely on standard rigorous scientific methodology to ensure safe and clinically relevant outcomes**.

Words of caution

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-07619-7 OPEN

Why rankings of biomedical image analysis competitions should be interpreted with care

Lena Maier-Hein et al.#

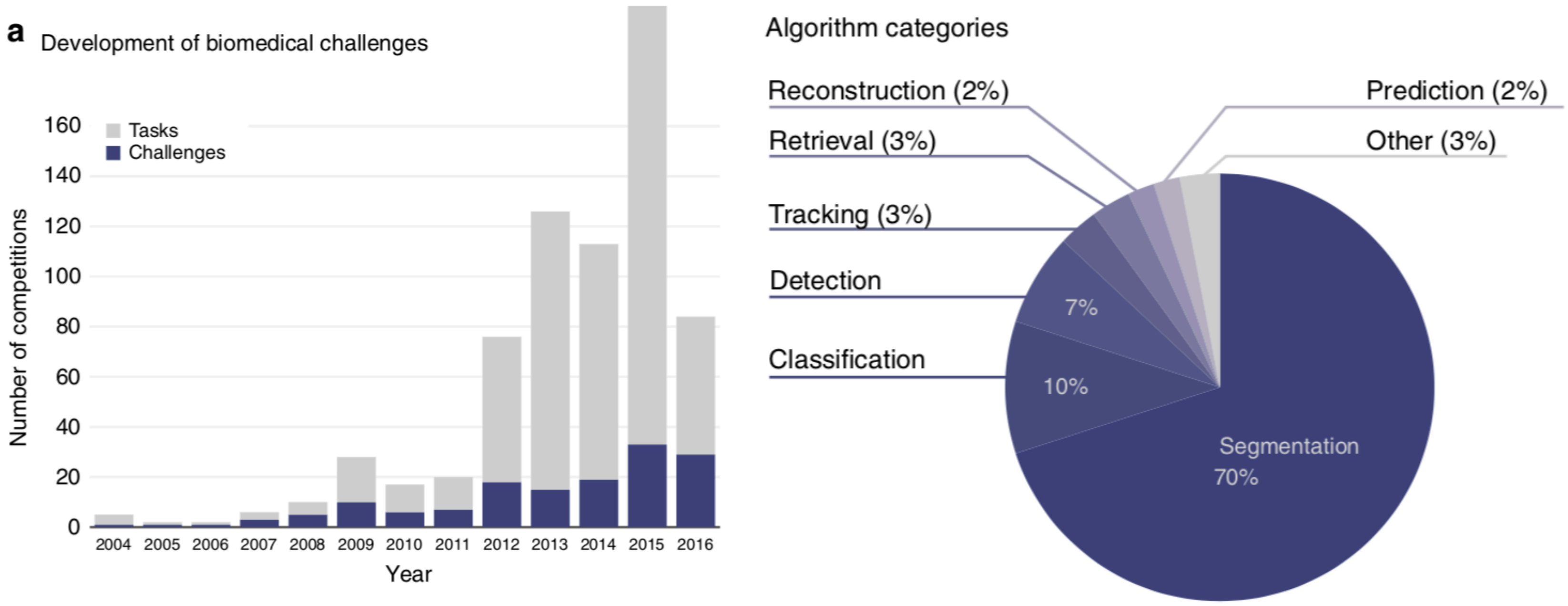
Available on Canvas

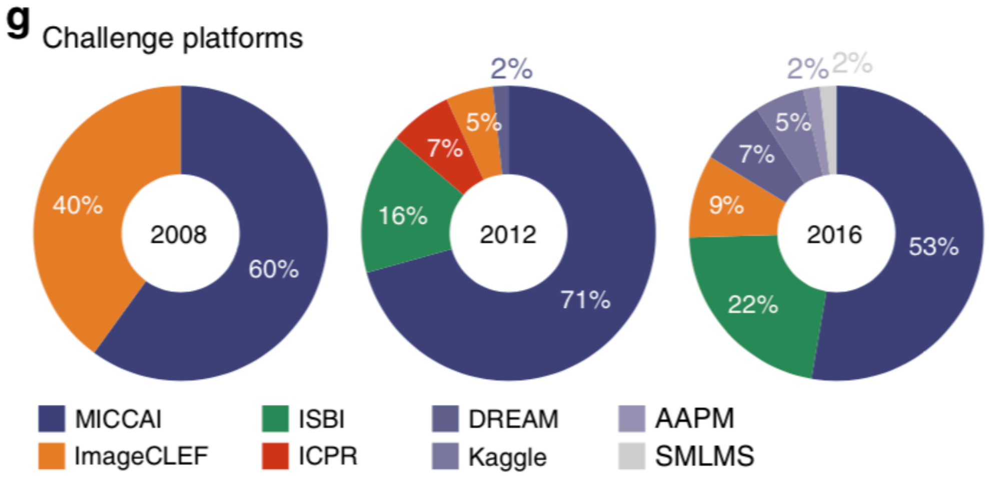
Why rankings of biomedical image analysis competitions should be interpreted with care

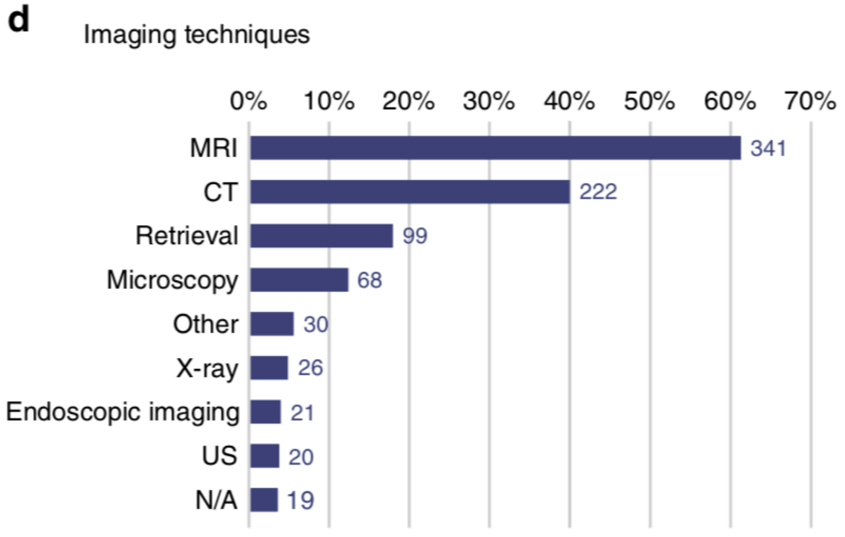
- The first comprehensive evaluation of biomedical image analysis challenges based on **150 challenges** conducted up until the end of 2016
- A total of **549 different image analysis tasks**
- 57% of these challenges (75% of all tasks) published their results in journals or conference proceedings

Maier-Hein et al., 2018

Why rankings of biomedical image analysis competitions should be interpreted with care







Maier-Hein et al., 2018

Why rankings of biomedical image analysis competitions should be interpreted with care

- We demonstrate the importance of challenges and show that the **lack of quality control has critical consequences**:
  - **Reproducibility** and **interpretation** of the results **is often hampered** as only a fraction of relevant information is typically provided.
  - The rank of an algorithm is generally **not robust** to a number of variables such as the test data used for validation, the ranking scheme applied and the observers that make the reference annotations.
- To overcome these problems, we recommend **best practice guidelines** and define open research questions to be addressed in the future.

Maier-Hein et al., 2018

Our work

Our work: selected examples

- M. Lux, M. Taschwer, and **O. Marques**, "Classification of Photos based on Good Feelings", Multimedia Grand Challenge Solution paper, ACM Multimedia Conference, Nara (Japan), October 2012.
- M. Taschwer and **O. Marques**, "AAUITEC at ImageCLEF 2015: Compound Figure Separation", Conference and Labs of the Evaluation Forum (CLEF) 2015, Toulouse, France, September 2015.
- A. Ishay and **O. Marques**, "ImageCLEF 2018 Tuberculosis Task: Ensemble of 3D CNNs with Multiple Inputs for Tuberculosis Type Classification", Conference and Labs of the Evaluation Forum (CLEF) 2018, Avignon, France, September 2018.
- M. Taschwer, M.J. Primus, K. Schöffmann, and **O. Marques**, "Early and Late Fusion of Classifiers for the MediaEval Medico Task", MediaEval'18 (Multimedia Evaluation Workshop), Sophia Antipolis, France, 29-31 October 2018.

Tuberculosis

Type (TBT) Classification from CT scans

ImageCLEF 2018 Tuberculosis Task: Ensemble of 3D CNNs with Multiple Inputs for Tuberculosis Type Classification

Adam Ishay[1] and Oge Marques[2]

Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, 33431 Boca Raton FL
{aishay,omarques}@fau.edu

ImageCLEFtuberculosis (2nd edition) 2018

Motivation: need for quick cheap methods of drug resistance (DR) detection based on Computed Tomography (CT) image analysis.

Subtask #2: TBT classification
The goal of this subtask is to automatically categorize each TB case into one of the following five types:
(1) Infiltrative, (2) Focal, (3) Tuberculoma, (4) Miliary, (5) Fibro-cavernous.

Dataset (each scan ~ 100 512x512 slices)

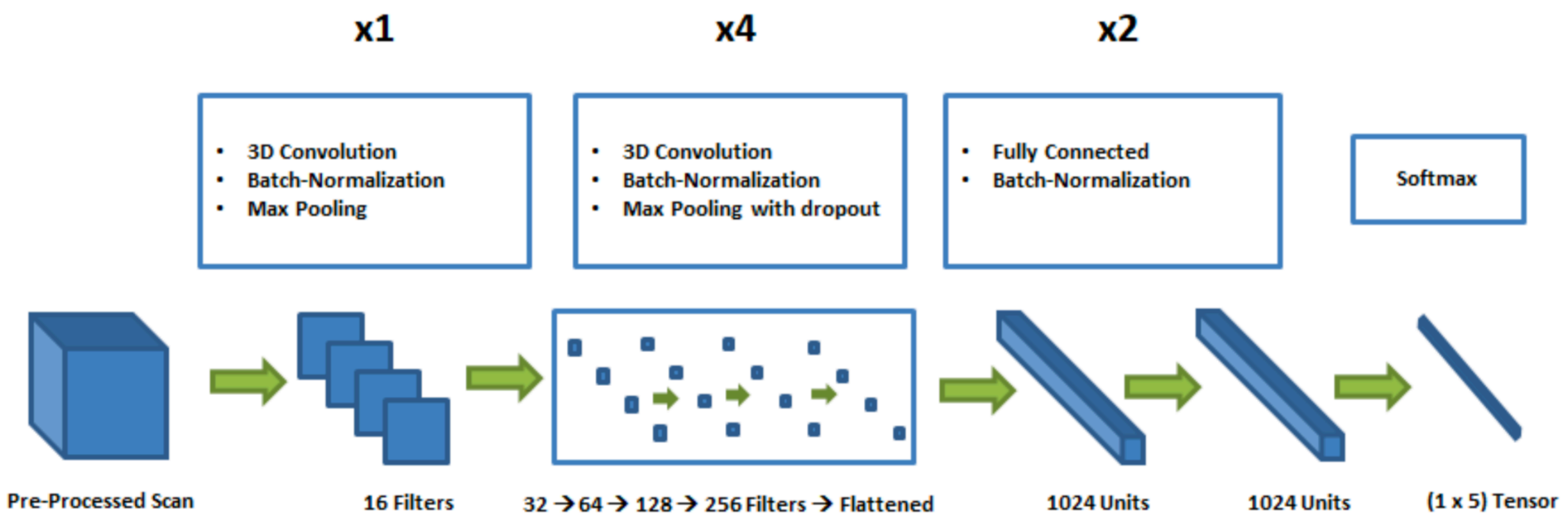
Class	Train Patients (Scans)	Test Patients (Scans)
Infiltrative (1)	228 (376)	89 (176)
Focal (2)	210 (273)	80 (115)
Tuberculoma (3)	100(154)	60 (86)
Miliary (4)	79(106)	50 (71)
Fibro-cavernous (5)	60 (99)	38 (57)
Total	677 (1008)	317 (505)

Source: https://www.imageclef.org/2018/tuberculosis

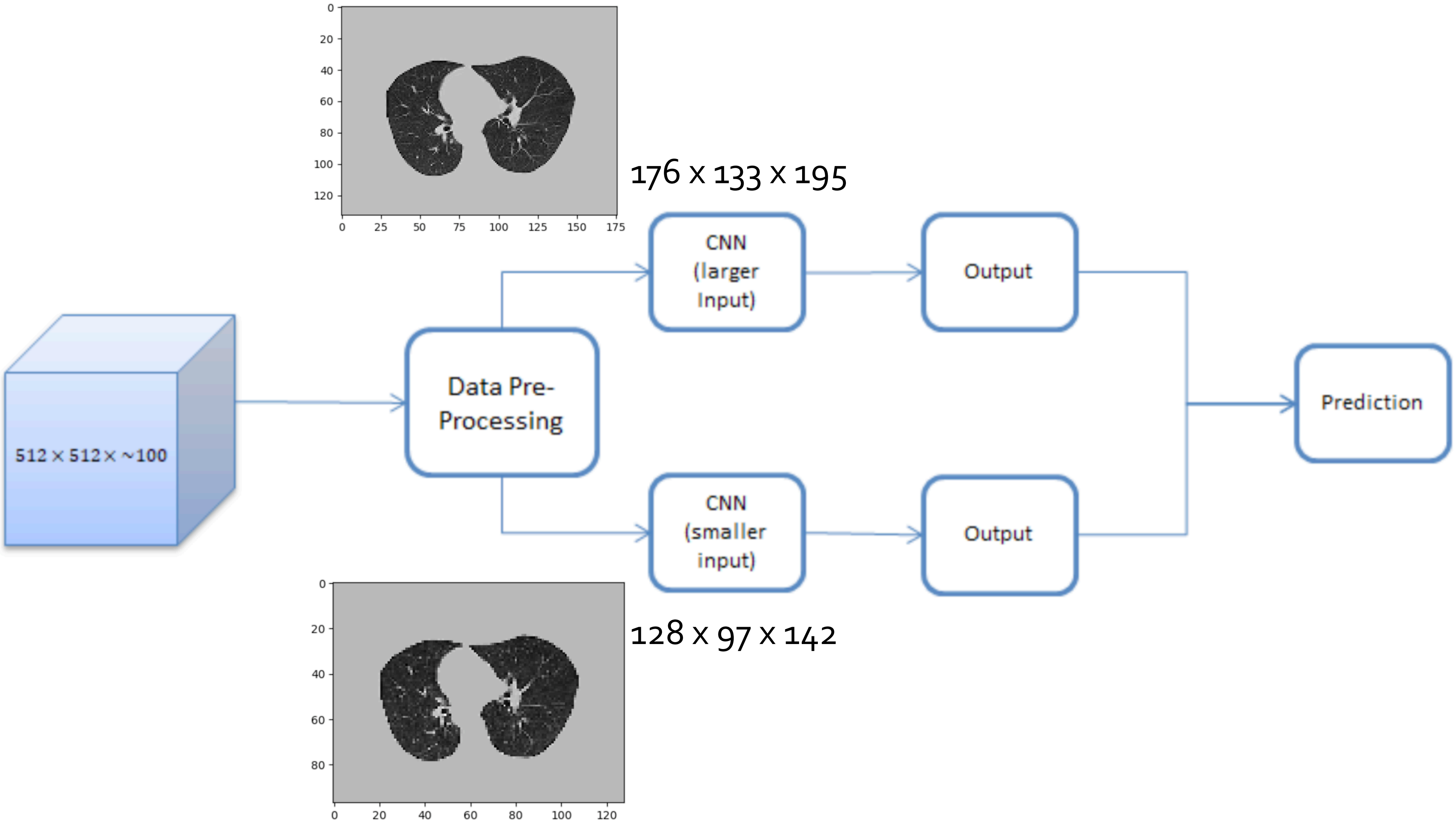
Pre-processing pipeline

Image → Mask →Resample → Cut → Normalize → Pad → Zero-center → Resize

3D CNNs used for training



Pipeline for predicting labels of test scans



ImageCLEF 2018 Tuberculosis Task: Ensemble of 3D CNNs with Multiple Inputs for Tuberculosis Type Classification

Adam Ishay[1] and Oge Marques[2]

Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, 33431 Boca Raton FL
{aishay,omarques}@fau.edu

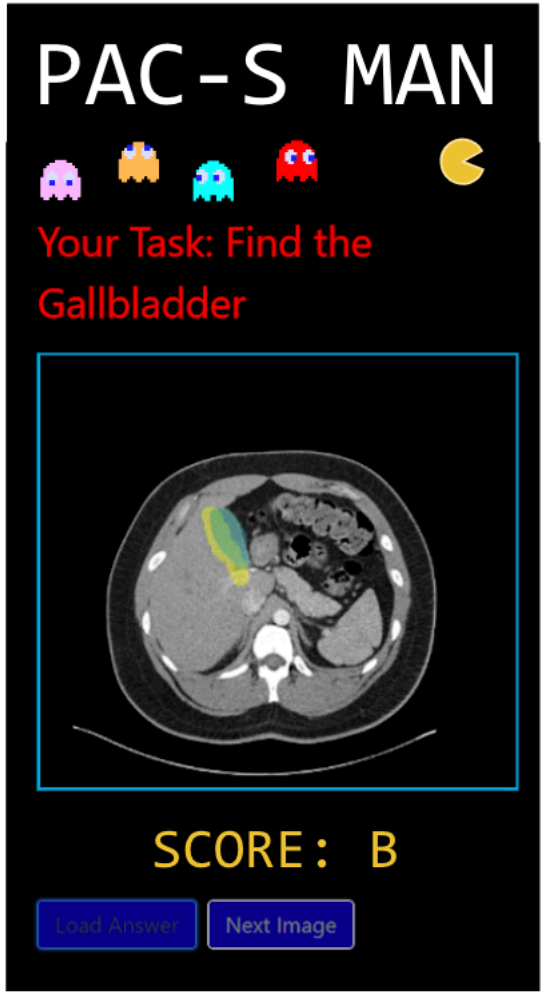
Subtask #2: TBT classification

Subtask 2 - Tuberculosis type classification

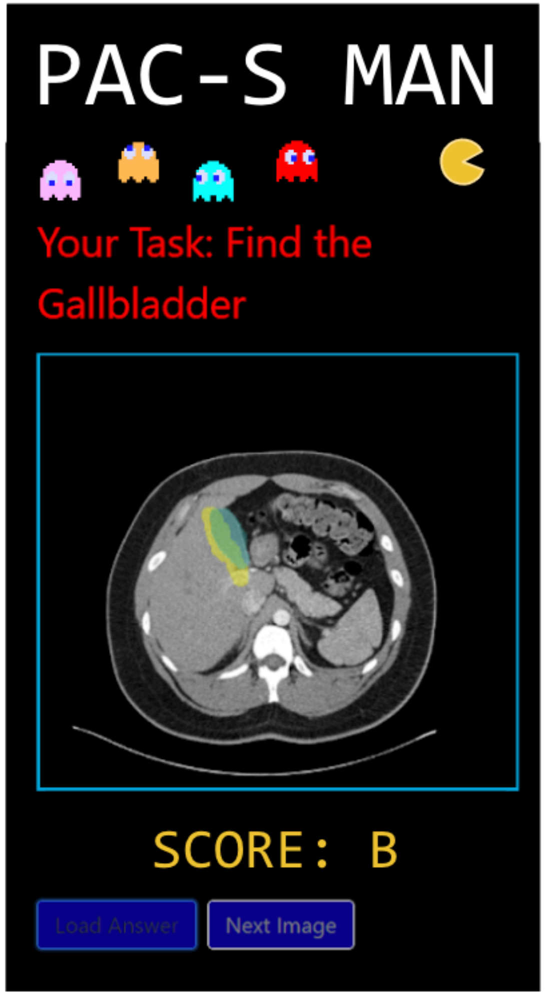
Group Name	Run	Kappa	Rank_Kappa	Accuracy	Rank_Acc
UIIP_BioMed	TBT_run_TBdescs2_zparts3_thrprob50_rf150.csv	0.2312	1	0.4227	1
fau_ml4cv	TBT_m4_weighted.txt	0.1736	2	0.3533	10
MedGIFT	TBT_AllFeats_std_euclidean_TST.csv	0.1706	3	0.3849	2
MedGIFT	TBT_Riesz_AllCols_euclidean_TST.csv	0.1674	4	0.3849	3

Source: https://www.imageclef.org/2018/tuberculosis

Our work: SIIM 2019 Hackathon



Our work: SIIM 2019 Hackathon



Learn more about it!
https://siim.org/page/hacking_healthcare
2019 highlights video: https://youtu.be/h9mHDw6JV2Q

Join the team!
Contact me

Concluding remarks

Challenges in AI

Good

- Low barrier to entry
- Standardized datasets
- Comparable results
- Reproducible research

Bad

- “Best results” don’t always mean “best algorithms”
- Competition often becomes the goal (rather than ‘means to an end’)
- Rules can be frustrating
- Annotations / ground truth not always 100% correct