U.S. National Library of Medicine

Lister Hill National Center for Biomedical Communications

Medical image analysis and medical information retrieval, with and without deep learning

Oge Marques, PhD

Professor

College of Engineering and Computer Science

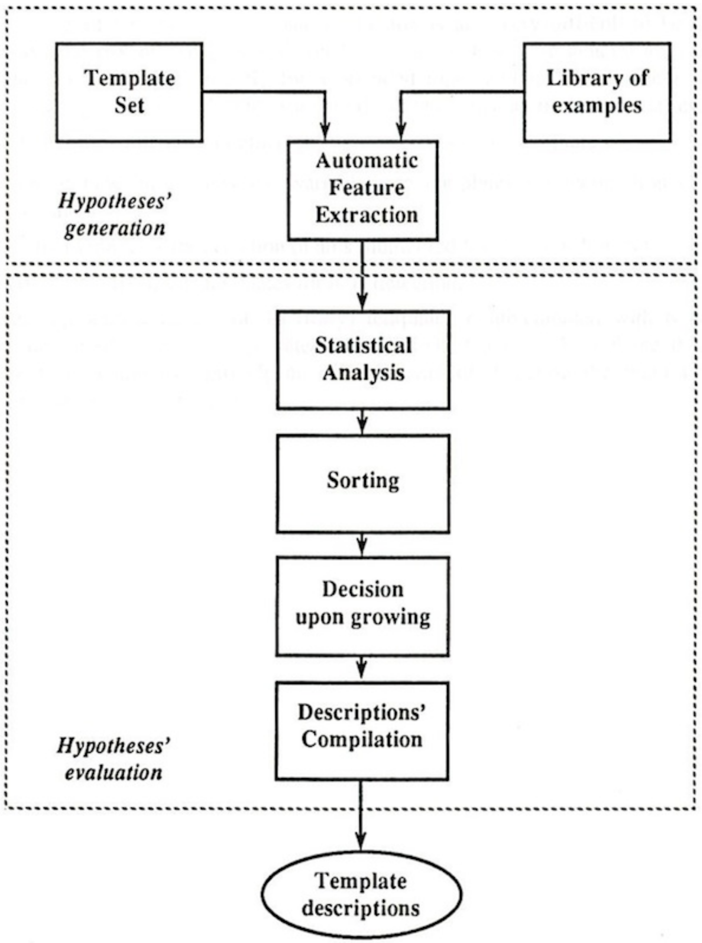
Florida Atlantic University

Outline

- Introduction / Background
- Deep Learning in Medical Image Analysis
- Examples of recent work
  - TBT classification
  - Skin lesion segmentation and classification
  - Medical Case Retrieval (MCR)
- At NIH: goals, ideas and plans for collaboration
- Concluding remarks

Brazil (1964-1988 and 1989-1997)

The Netherlands (1988-1989)

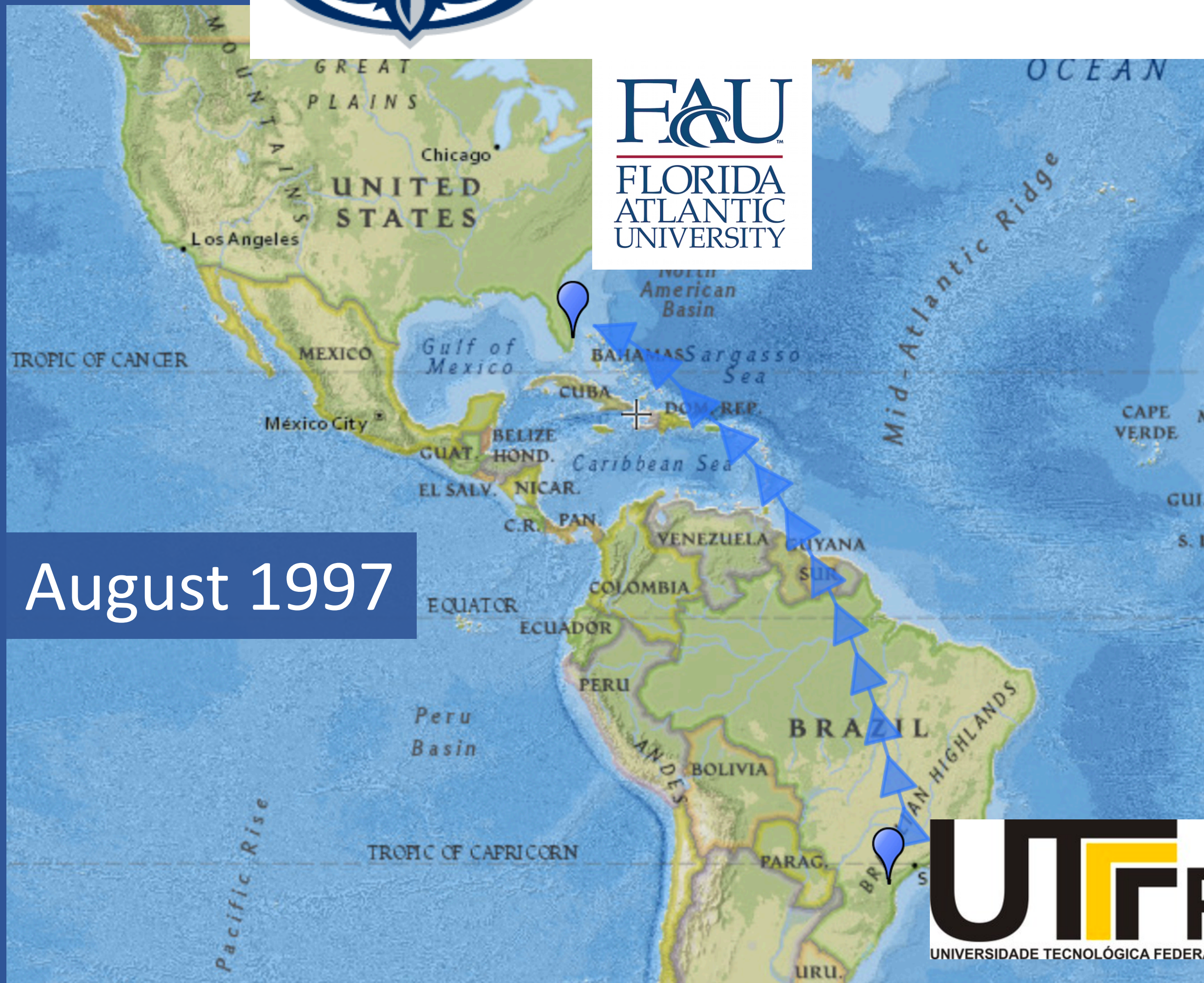


PHILIPS

MEE thesis: "Automatic Derivation of Template Descriptions for Character Recognition From a Set of Examples"

UNBRIDLED AMBITION®

FAU FLORIDA ATLANTIC UNIVERSITY



August 1997

UTFPR
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

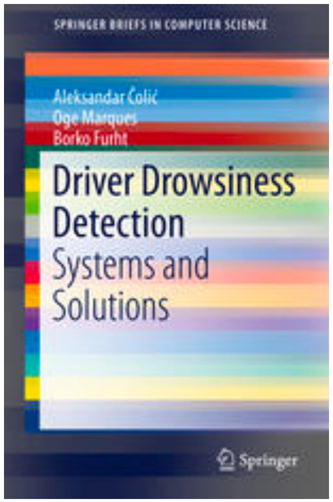
My research focus

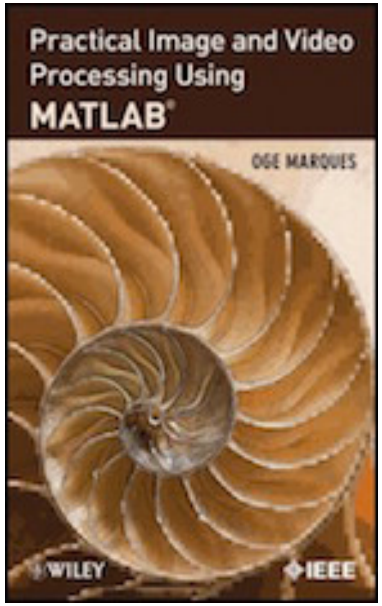
Intelligent processing of visual information

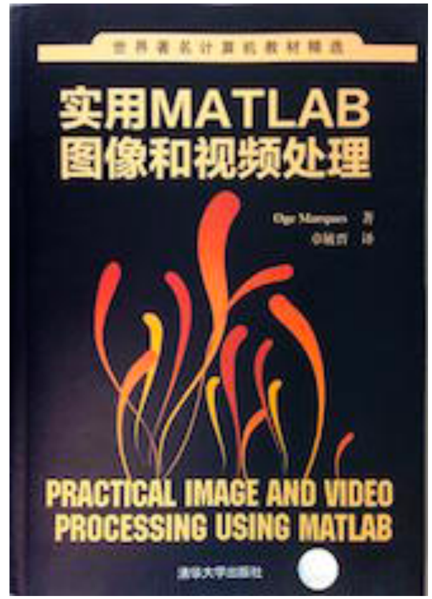
– image processing
– medical image analysis
– computer vision
– human vision
– artificial intelligence
– machine learning
– deep learning

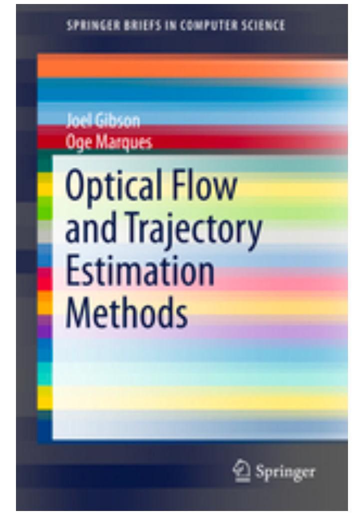
Image from: https://goo.gl/a42gD7

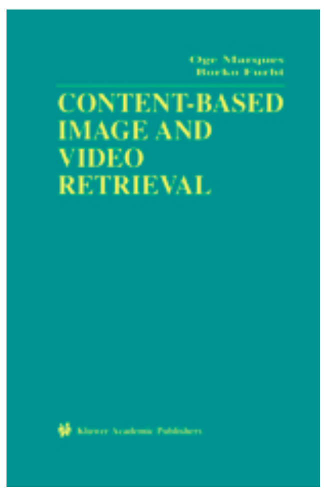
My work: selected books

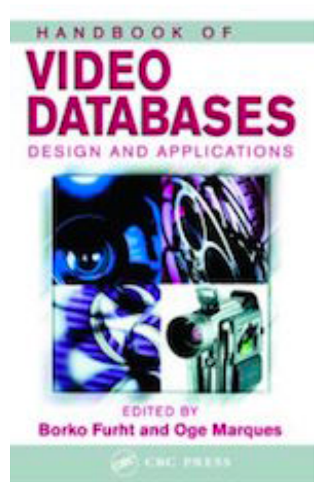


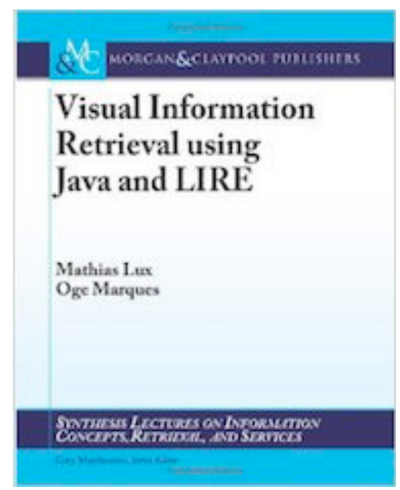








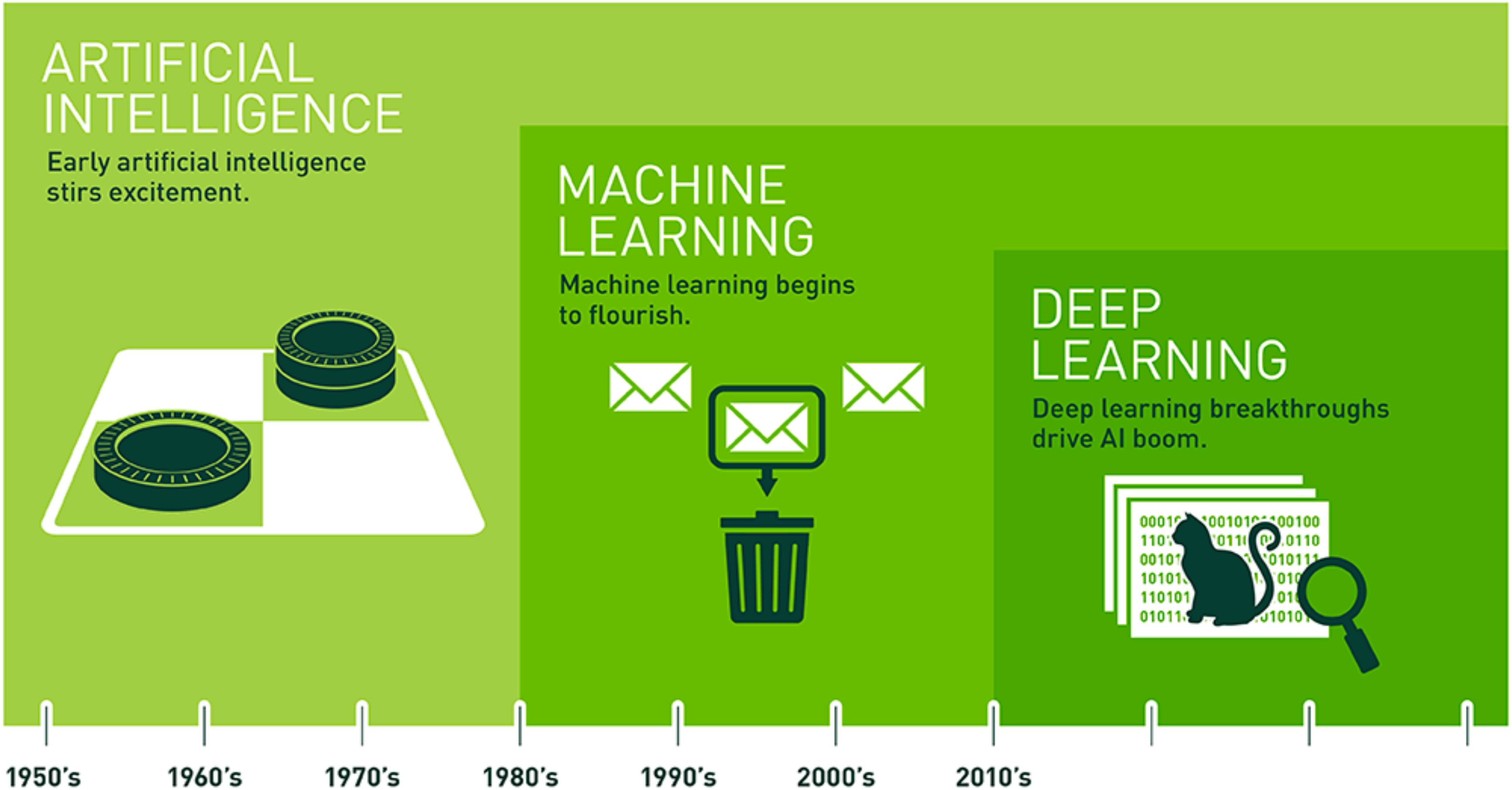




My work: highlights

- **Intelligent image analysis and classification system for tuberculosis diagnosis with visual question answering (VQA) capabilities** (FAU, *starting soon*)
- **Mobile app and machine learning solutions for early melanoma detection** (FAU, *ongoing*)
- **Skin lesion detection, segmentation and classification** (FAU with UPC Barcelona, *ongoing*)
  - **Tuberculosis type classification** (2nd place at 2018 ImageCLEFtuberculosis) (FAU) (2018)
  - **Medical Case Retrieval** (with AAU Klagenfurt) (2015-17)
  - **Machine learning strategies for childhood pneumonia diagnosis** (with UFG Brazil) (2013-14)
  - **Summarization of arthroscopic videos** (with AAU Klagenfurt) (2008-09)

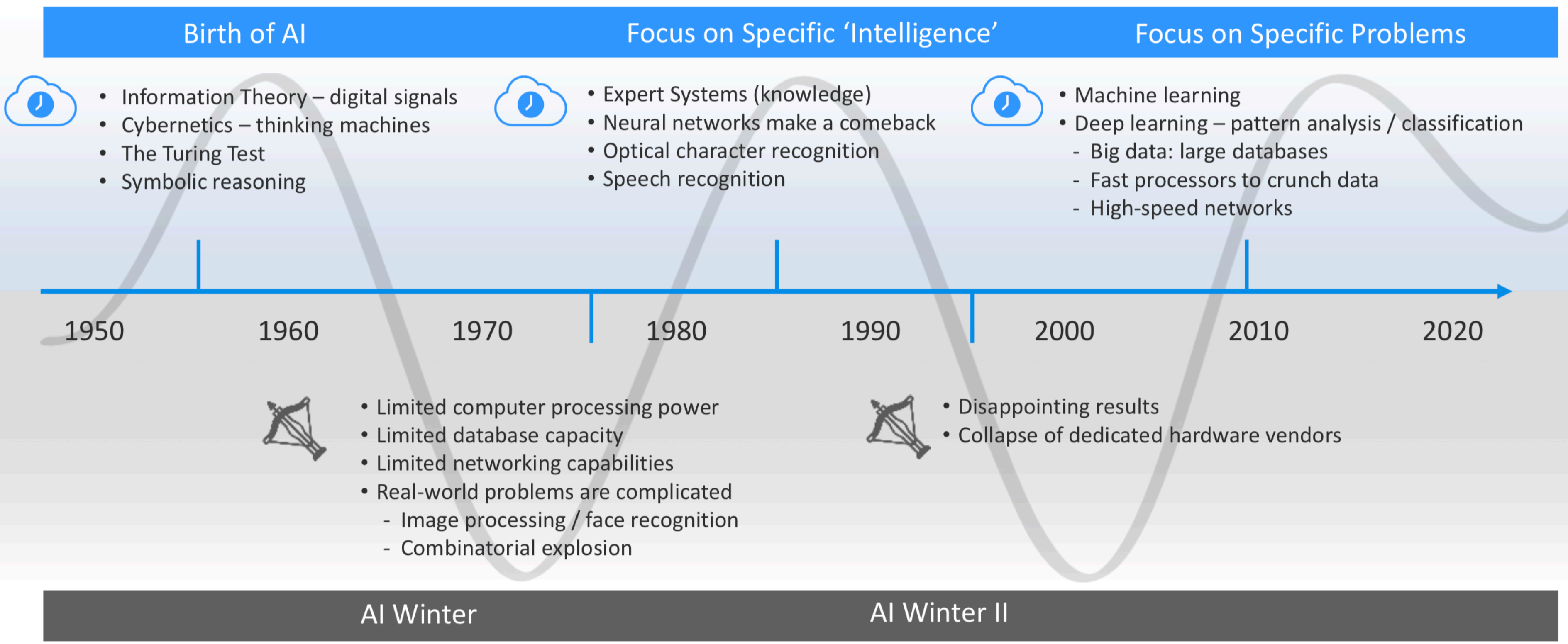
The impact of Deep Learning in Computer Vision and related areas



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

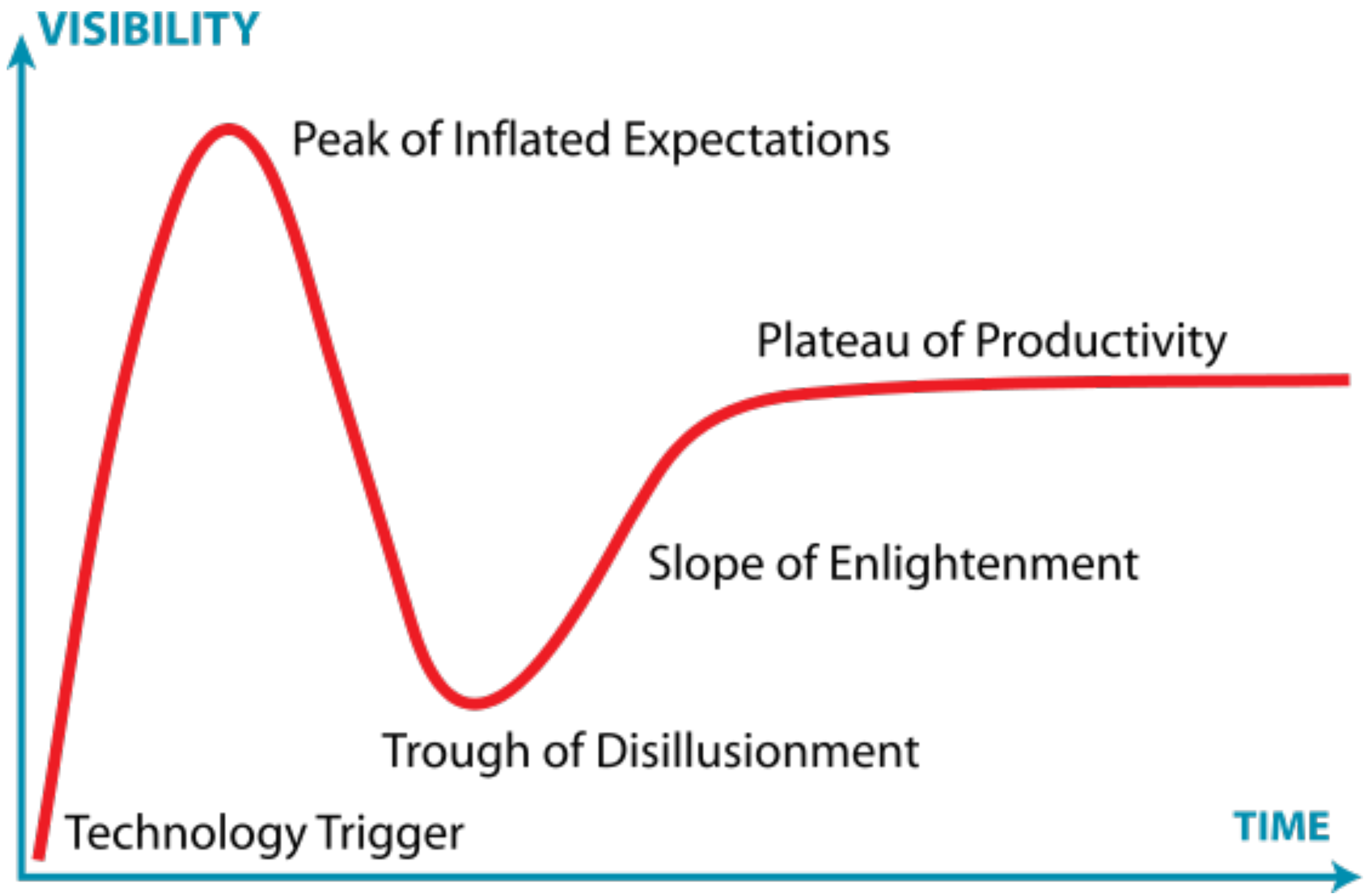
Source: NVIDIA Blog

An AI Timeline



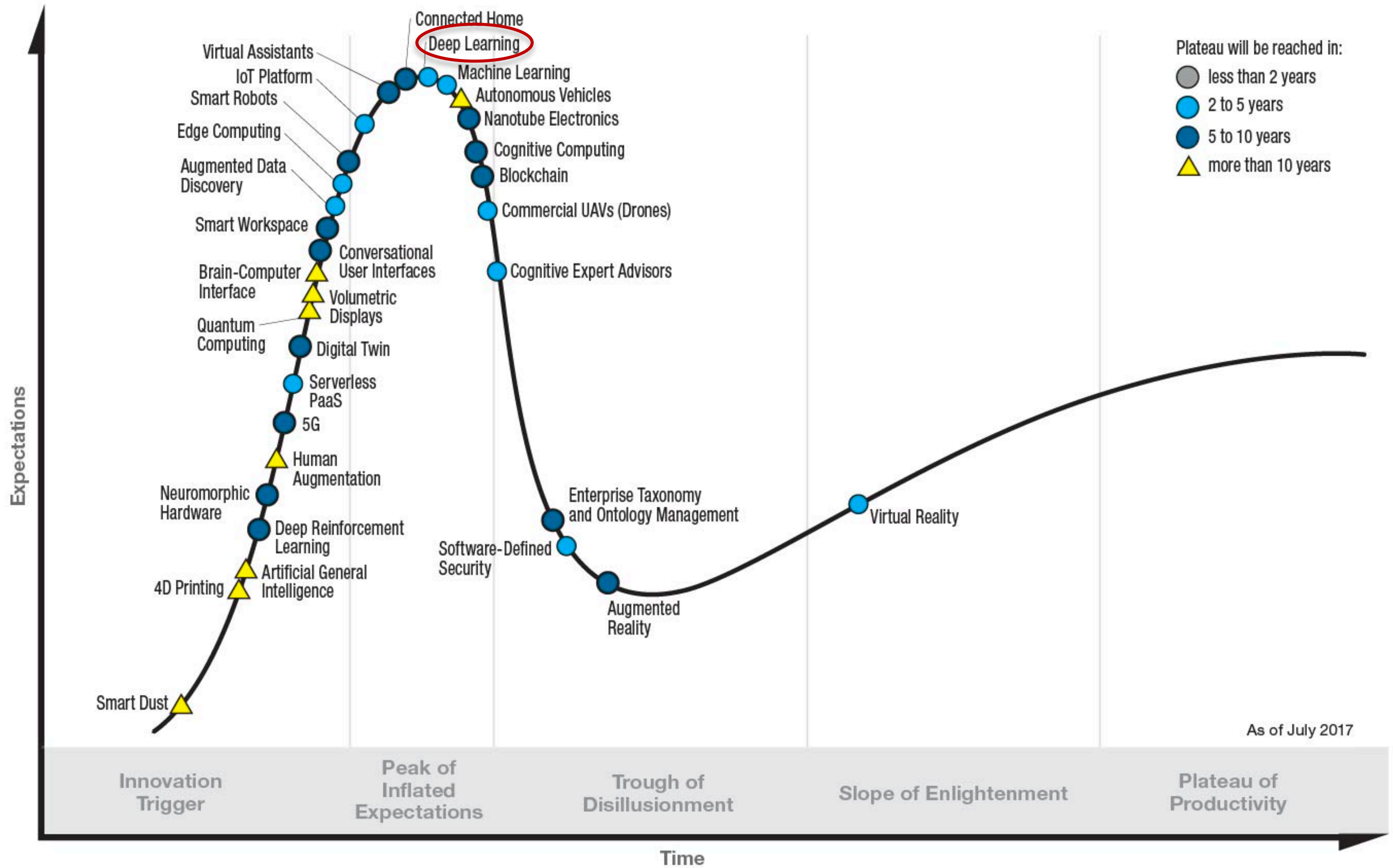
Source: https://www.slideshare.net/dlavenda/ai-and-productivity

The deep learning phenomenon



Source: (Wikimedia)

Gartner Hype Cycle for Emerging Technologies, 2017



gartner.com/SmarterWithGartner

Source: Gartner (July 2017)

Gartner

Do you *love* your child?

If you love your child, teach them deep learning.

Keras4Kindergartners

Mon. & Wed. 4PM-6PM

Learn Keras in just two afternoons a week!

Mommy and Me MXNet

Sat. 10AM-2:30PM

Learn about MXNet together with your tot!

Papa and Me PyTorch

Sun. 10AM-2:30PM

Spend Sunday together learning about PyTorch!

Price upon request.

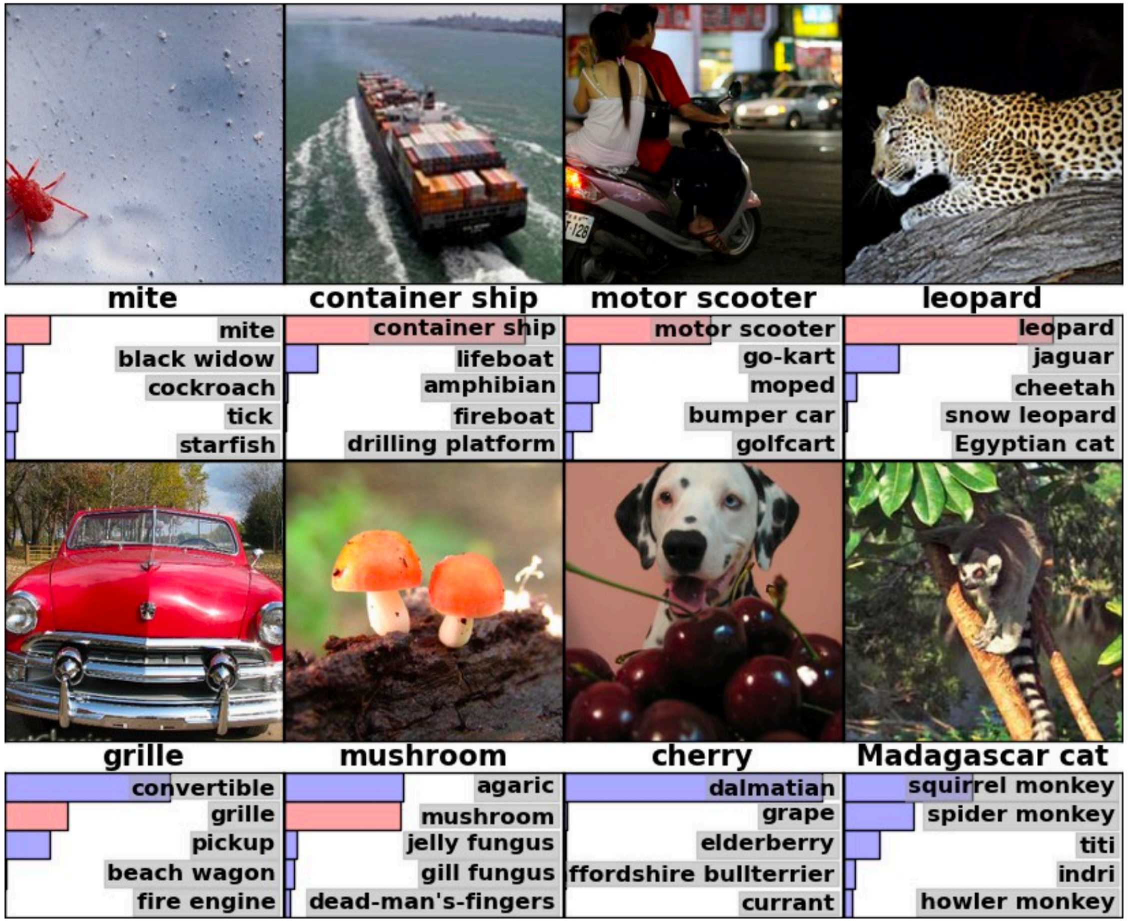
Spread the word! Put up this poster up at your local raw water shop.

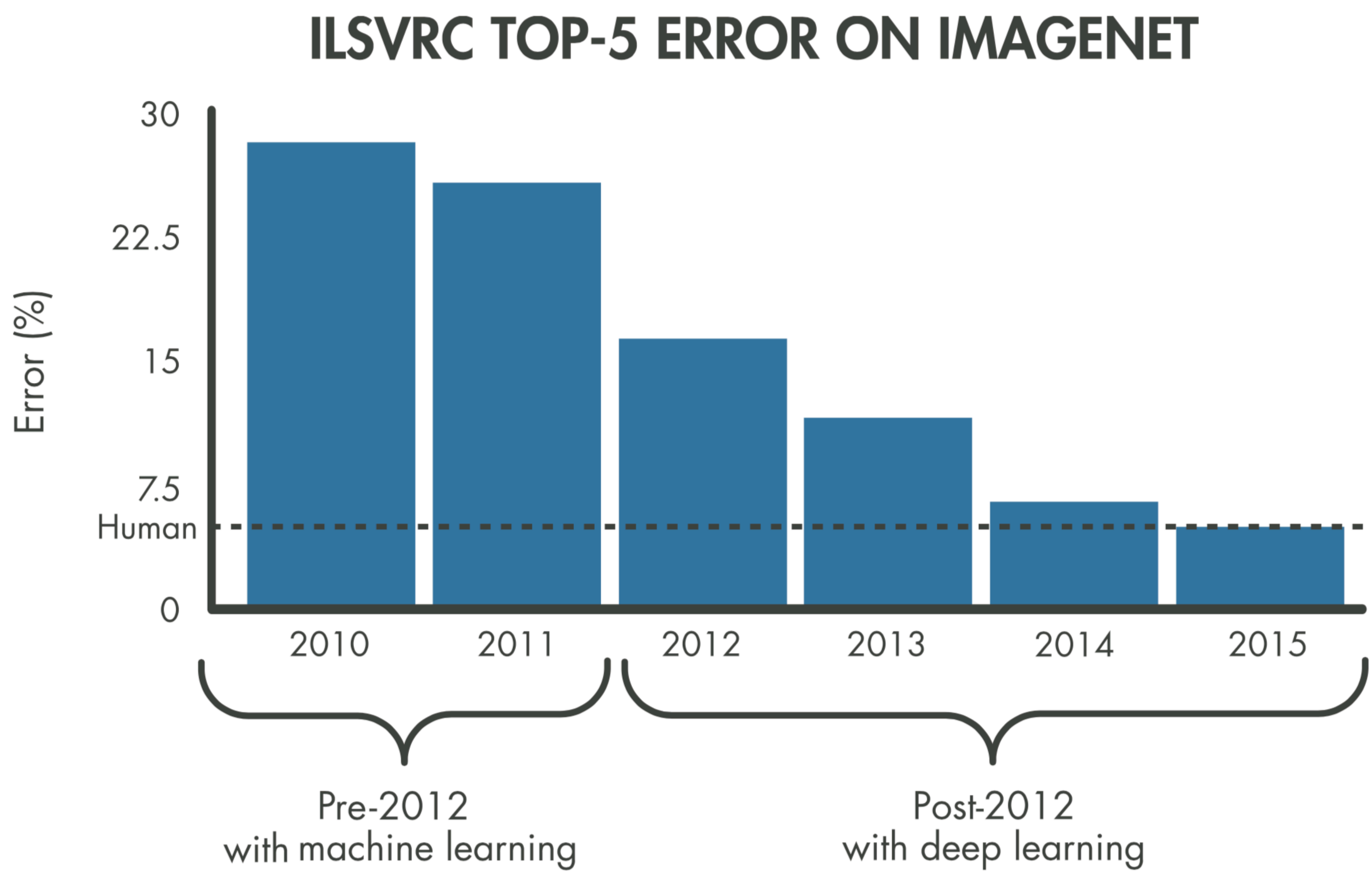
http://keras4kindergartners.com/

ImageNet LSVRC

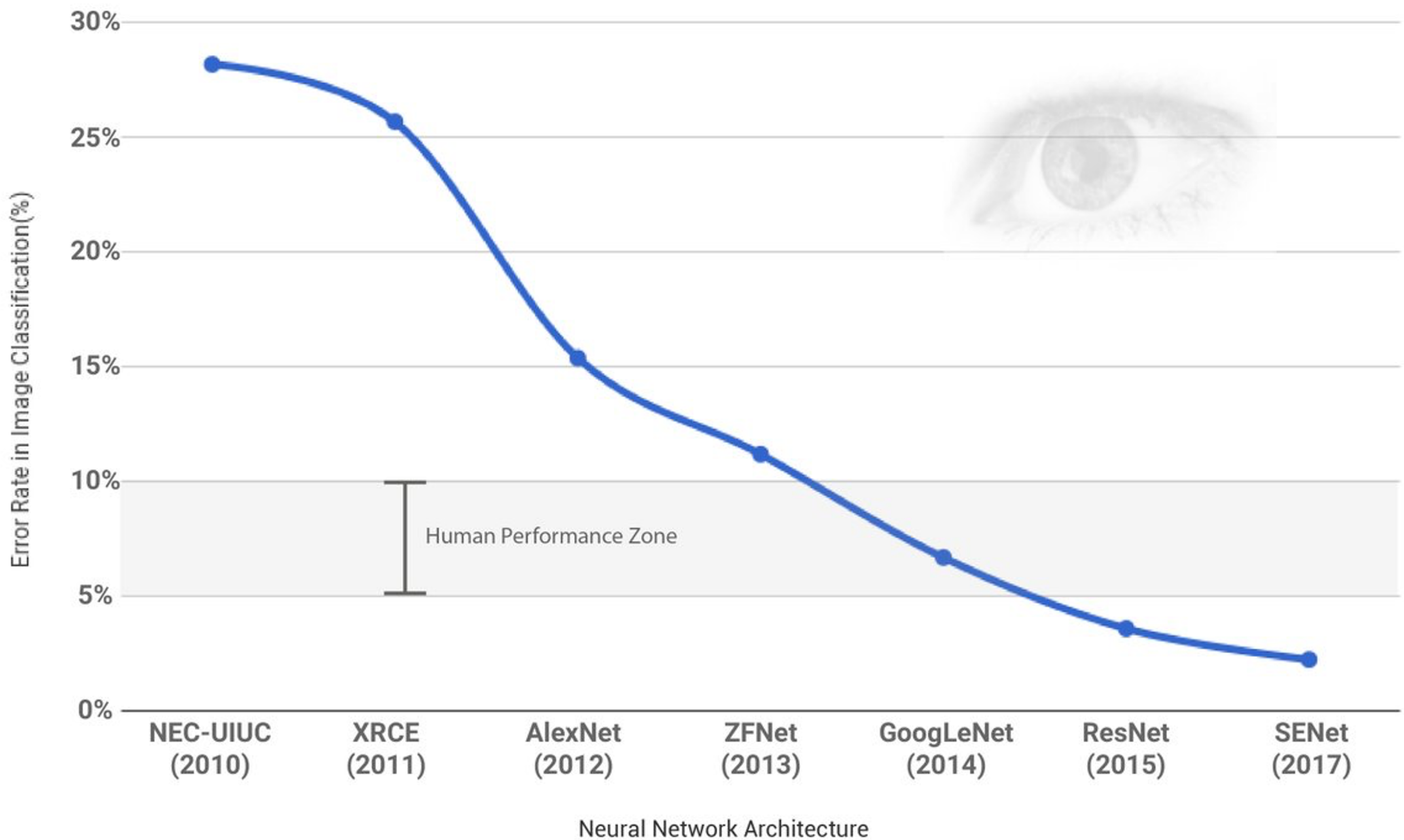
IMAGENET

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





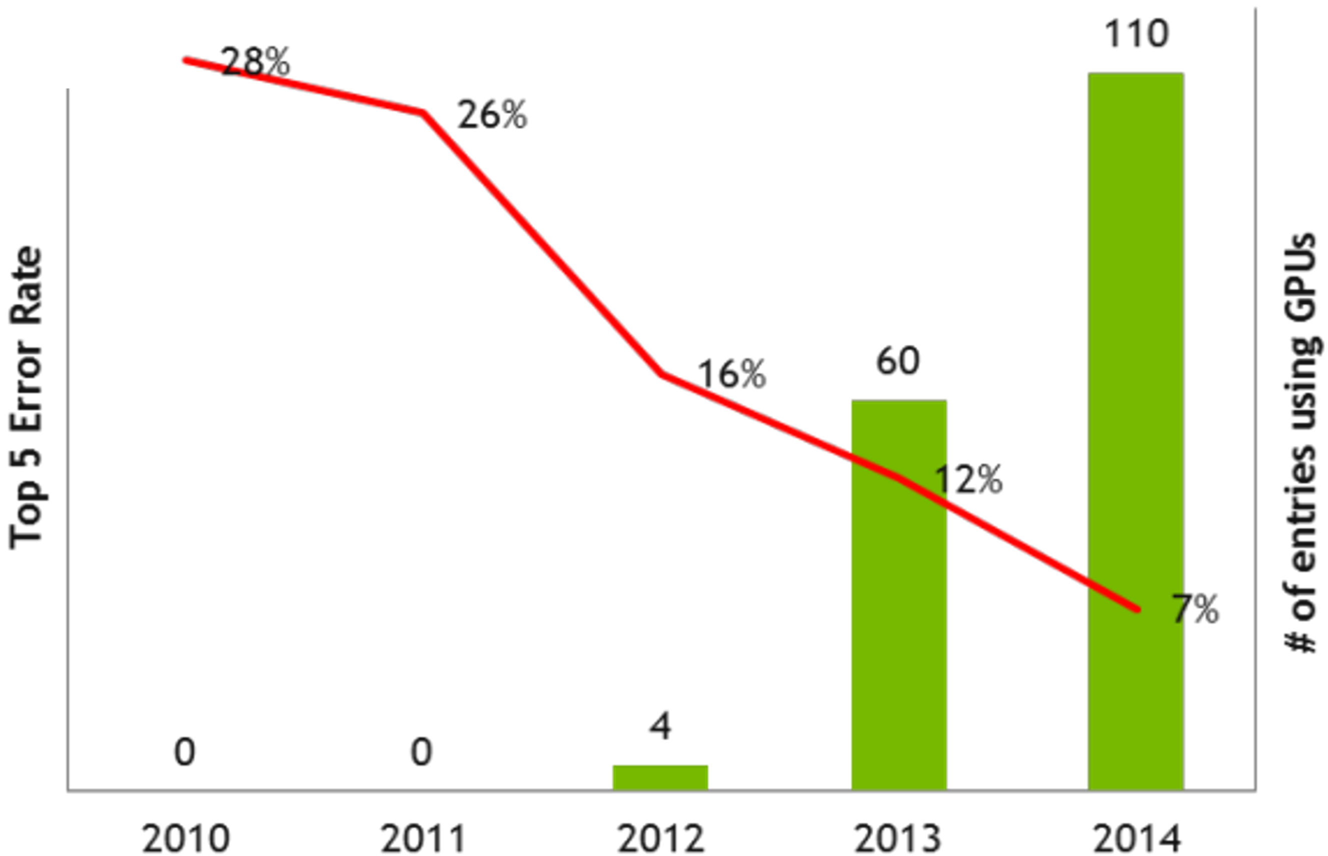
Source: (Mathworks 2017)



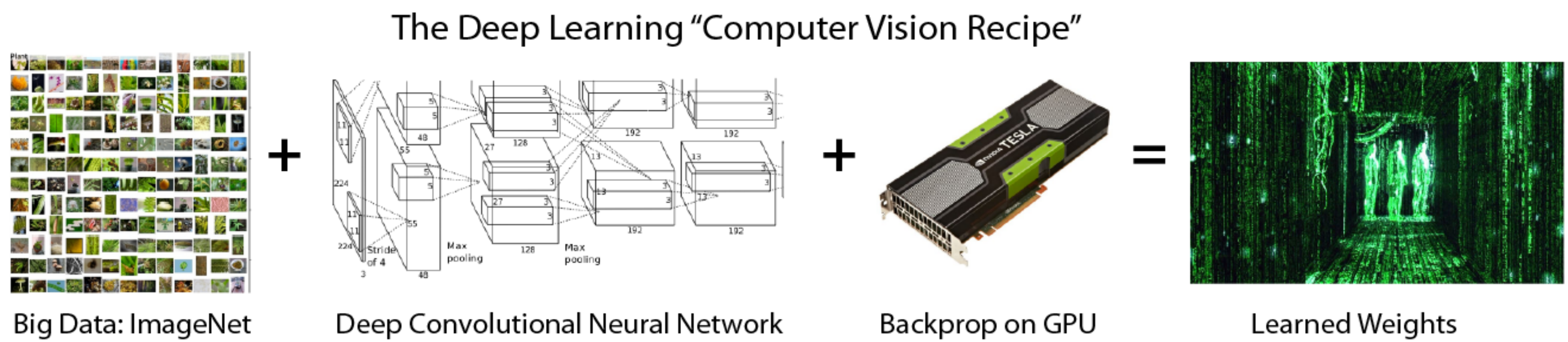
30%
25%
20%
15%
10%
5%
0%
Error Rate in Image Classification(%)
Human Performance Zone
NEC-UIUC (2010)
XRCE (2011)
AlexNet (2012)
ZFNet (2013)
GoogLeNet (2014)
ResNet (2015)
SENet (2017)
Neural Network Architecture


GEFORCE GTX
GTX 1080

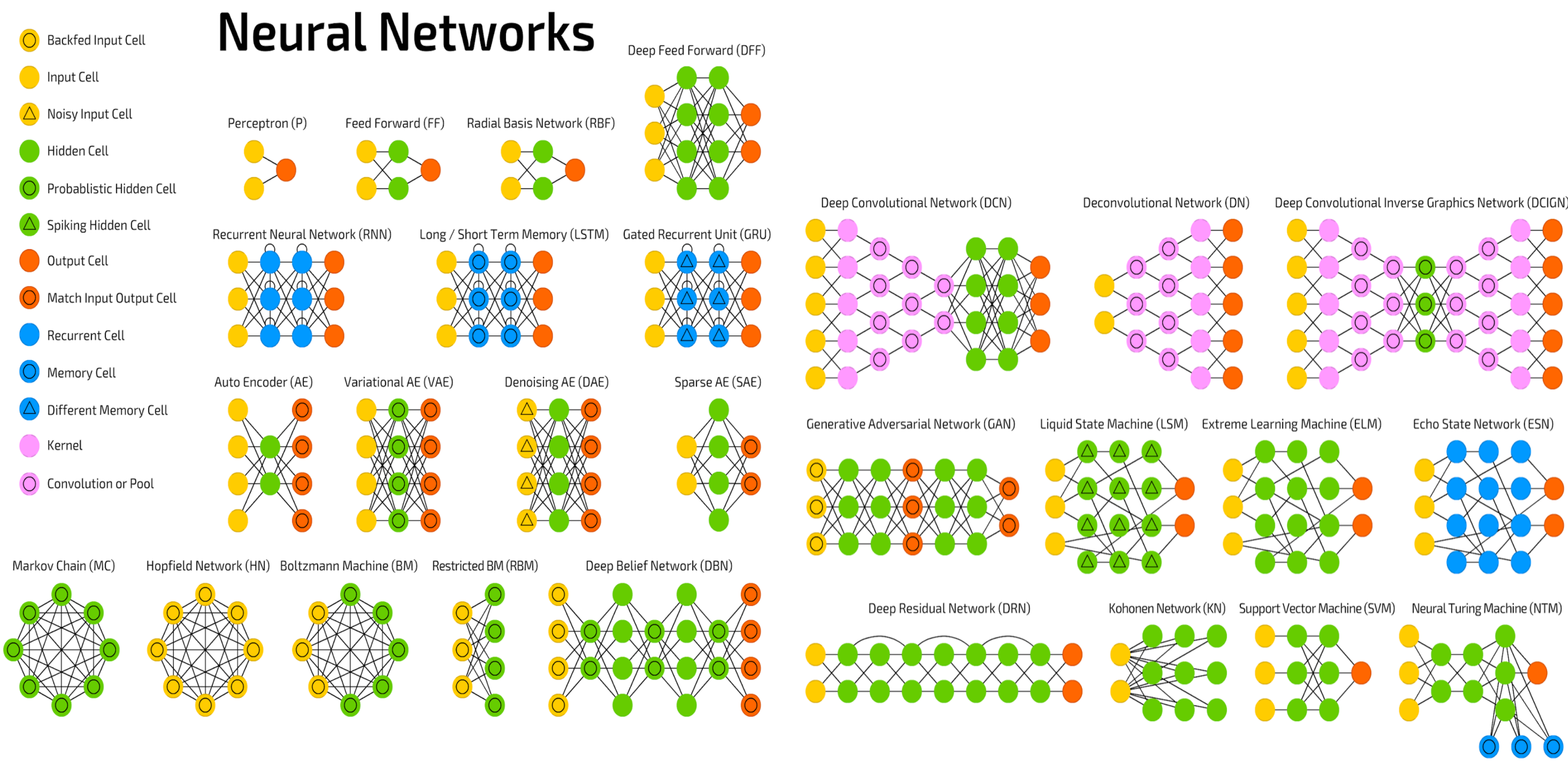
IMAGENET


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

The Deep Learning recipe for computer vision

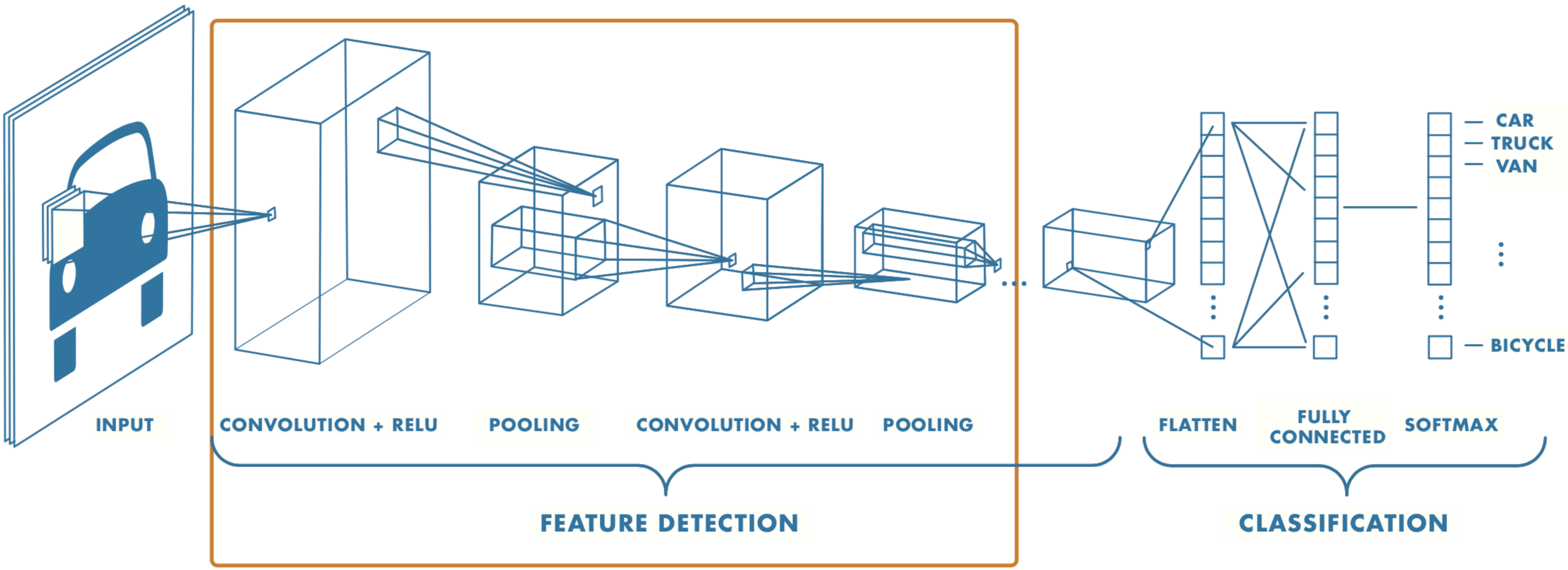


Source: http://www.computervisionblog.com/2015/05/deep-learning-vs-big-data-who-owns-what.html



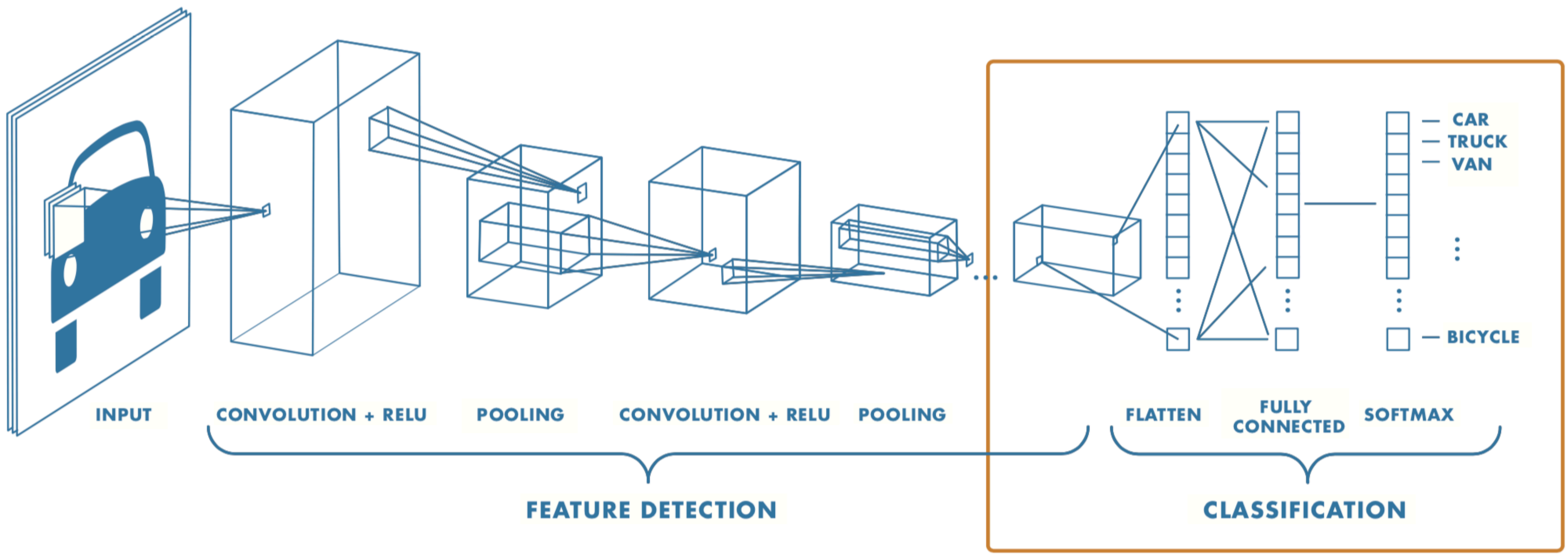
https://www.datasciencecentral.com/profiles/blogs/concise-visual-summary-of-deep-learning-architectures

CNNs



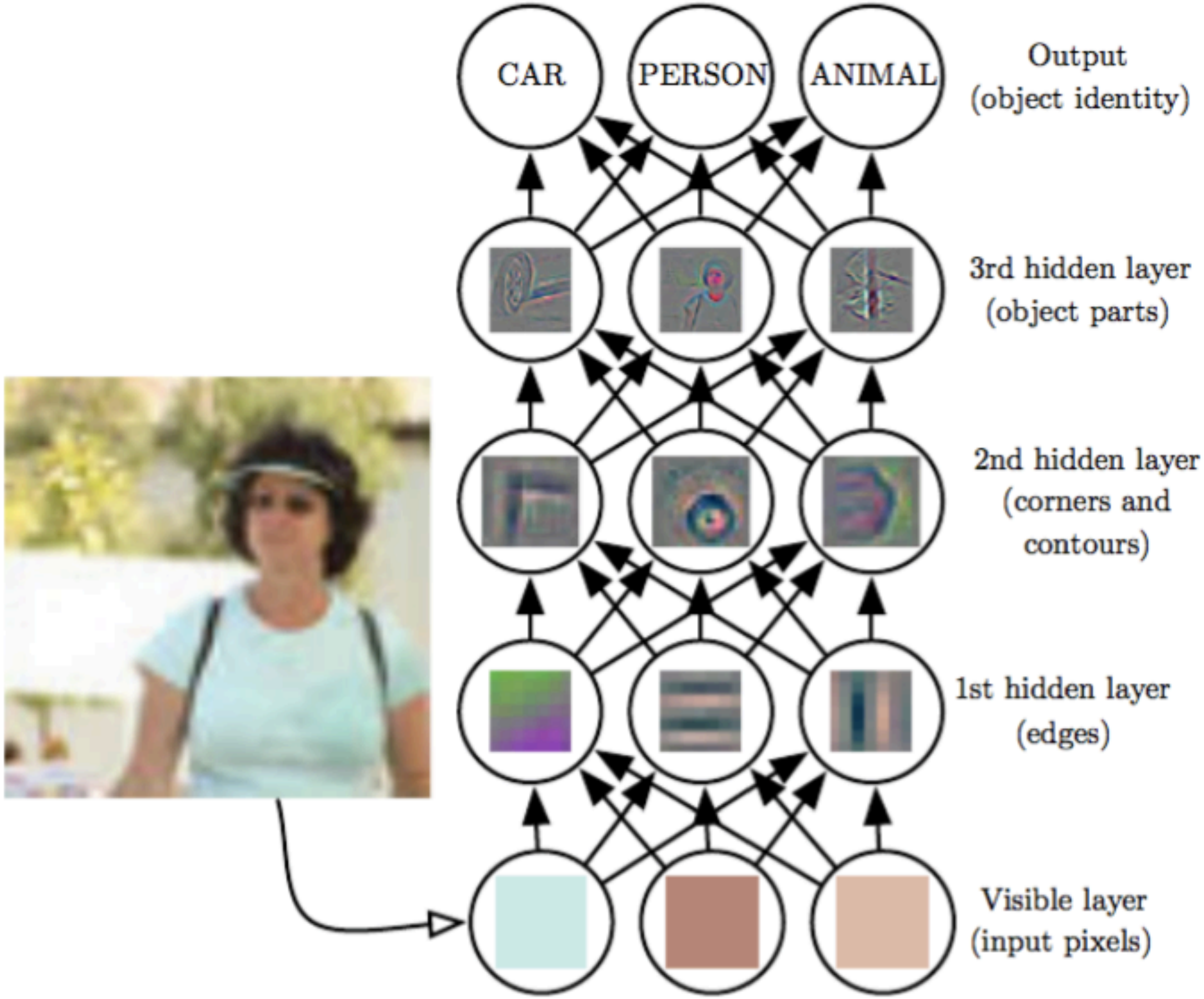
Source: (Mathworks 2017)

CNNs



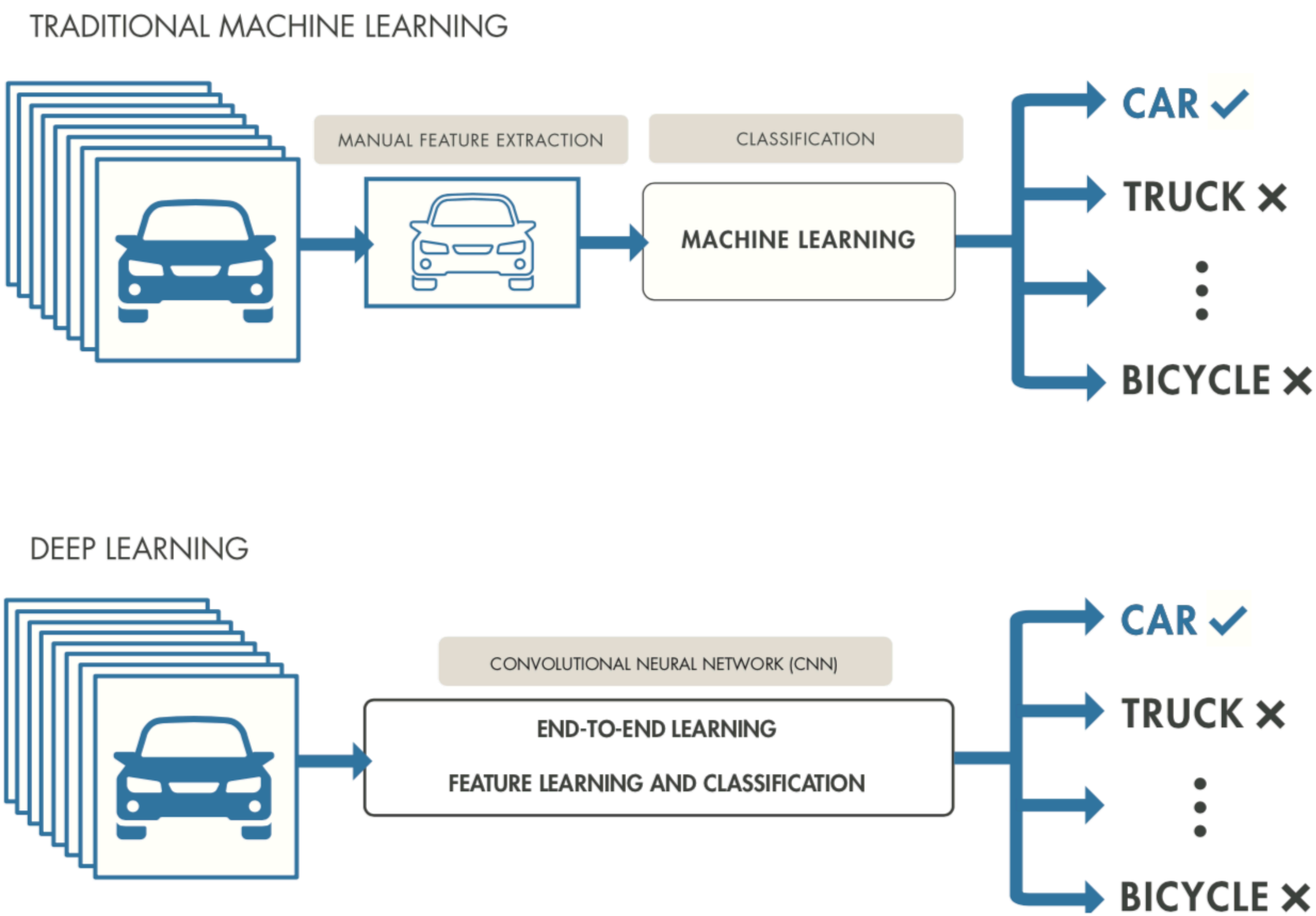
Source: (Mathworks 2017)

Deep Learning: learning hierarchical representations



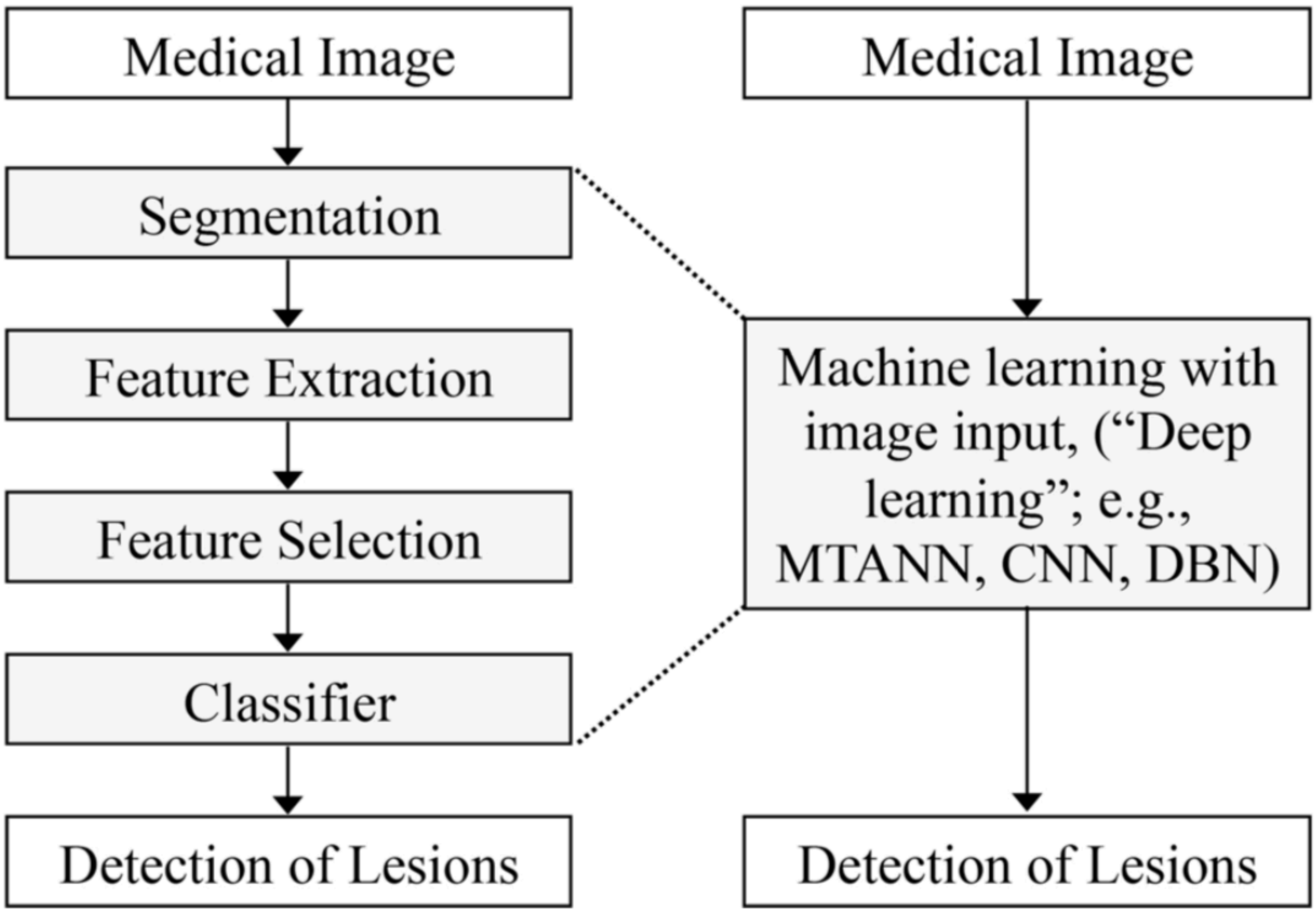
Source: (Goodfellow et al. 2016)

Deep learning vs. machine learning



Source: (Mathworks 2017)

Medical image analysis: before and after DL



Source: Suzuki (2017)

Medical Image Analysis 42 (2017) 60–88

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

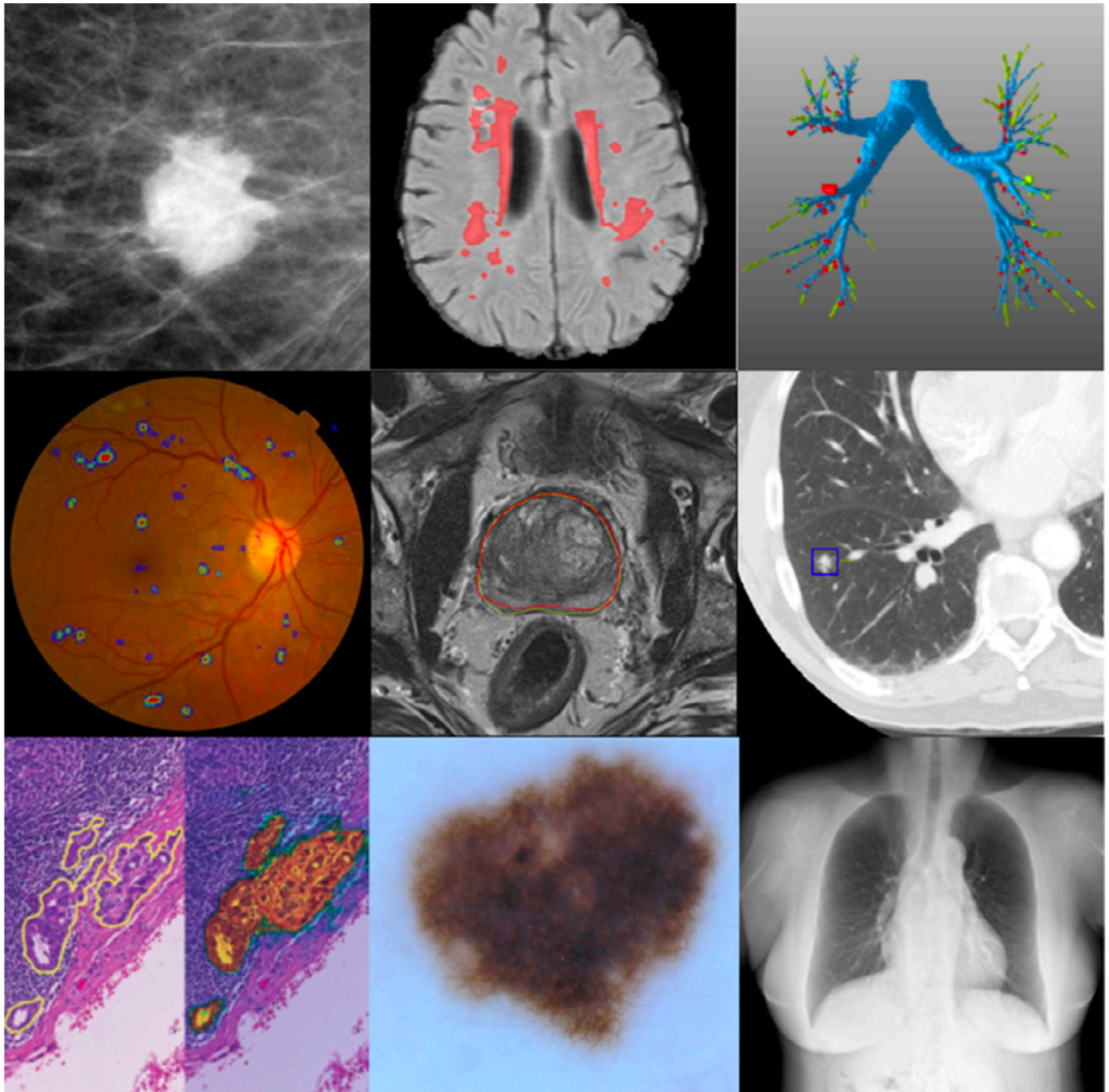
Survey Paper

A survey on deep learning in medical image analysis

Geert Litjens*, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez

Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, The Netherlands

Examples of medical imaging applications in which deep learning has achieved state-of-the-art results



G. Litjens et al./Medical Image Analysis 42 (2017) 60–88

From top-left to bottom-right:

- mammographic mass classification (2016)
- segmentation of lesions in the brain (2016)
- leak detection in airway tree segmentation (2017)

- diabetic retinopathy classification (2015-2016)
- prostate segmentation (2016)
- nodule classification (2017)

- breast cancer metastases detection in lymph nodes (2016)
- human expert performance in skin lesion classification (2017)
- state-of-the-art bone suppression in x-rays (2016)

Deep Learning: limitations and criticism

- Lack of theory surrounding the methods
  - “alchemy” before “chemistry”
- Data hungry
- Computationally intensive / slow to train
- Not sufficiently transparent / “black box”
- Time-consuming and ad-hoc (hyperparameter) optimization
- Lack of explanatory power
- Problematic software development pipeline
- Fixed architectures
- Security concerns (e.g., adversarial examples)
- … (many more)

Tuberculosis Type (TBT) Classification

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)

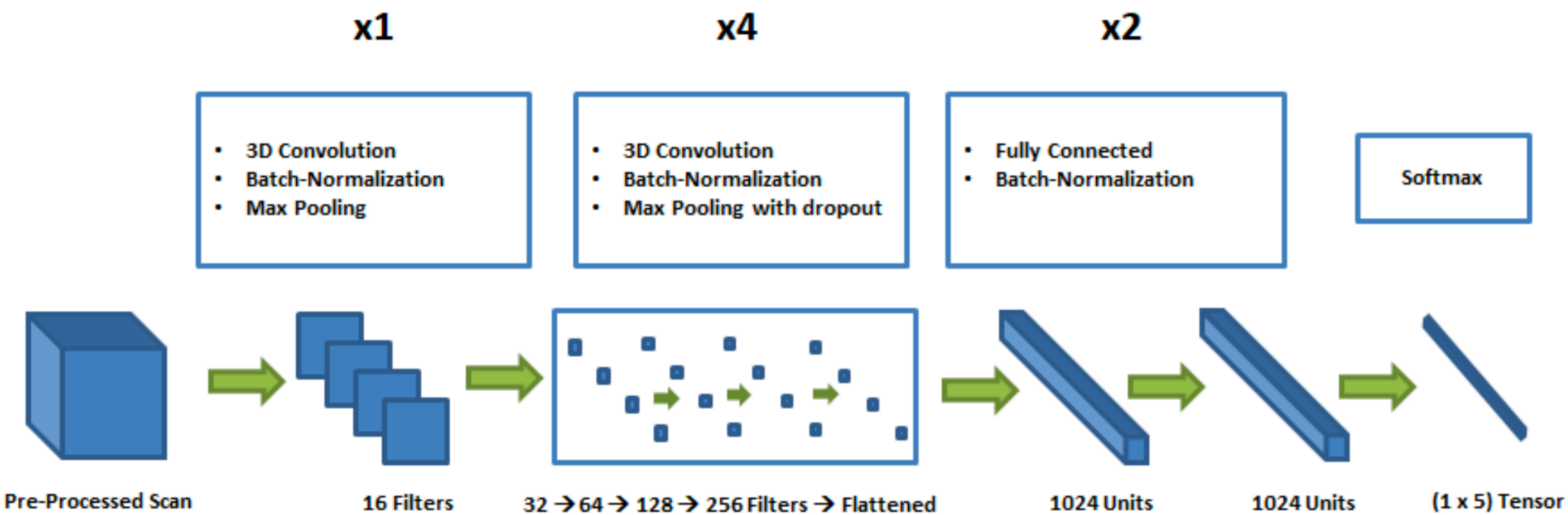
Class	Train Patients (Scans)	Test Patients (Scans)
Infiltrative (1)	228 (376)	89 (176)
Focal (2)	210 (273)	80 (115)
Tuberculoma (3)	100(154)	60 (86)
Miliary (4)	79(106)	50 (71)
Fibro-cavernous (5)	60 (99)	38 (57)
Total	677 (1008)	317 (505)

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

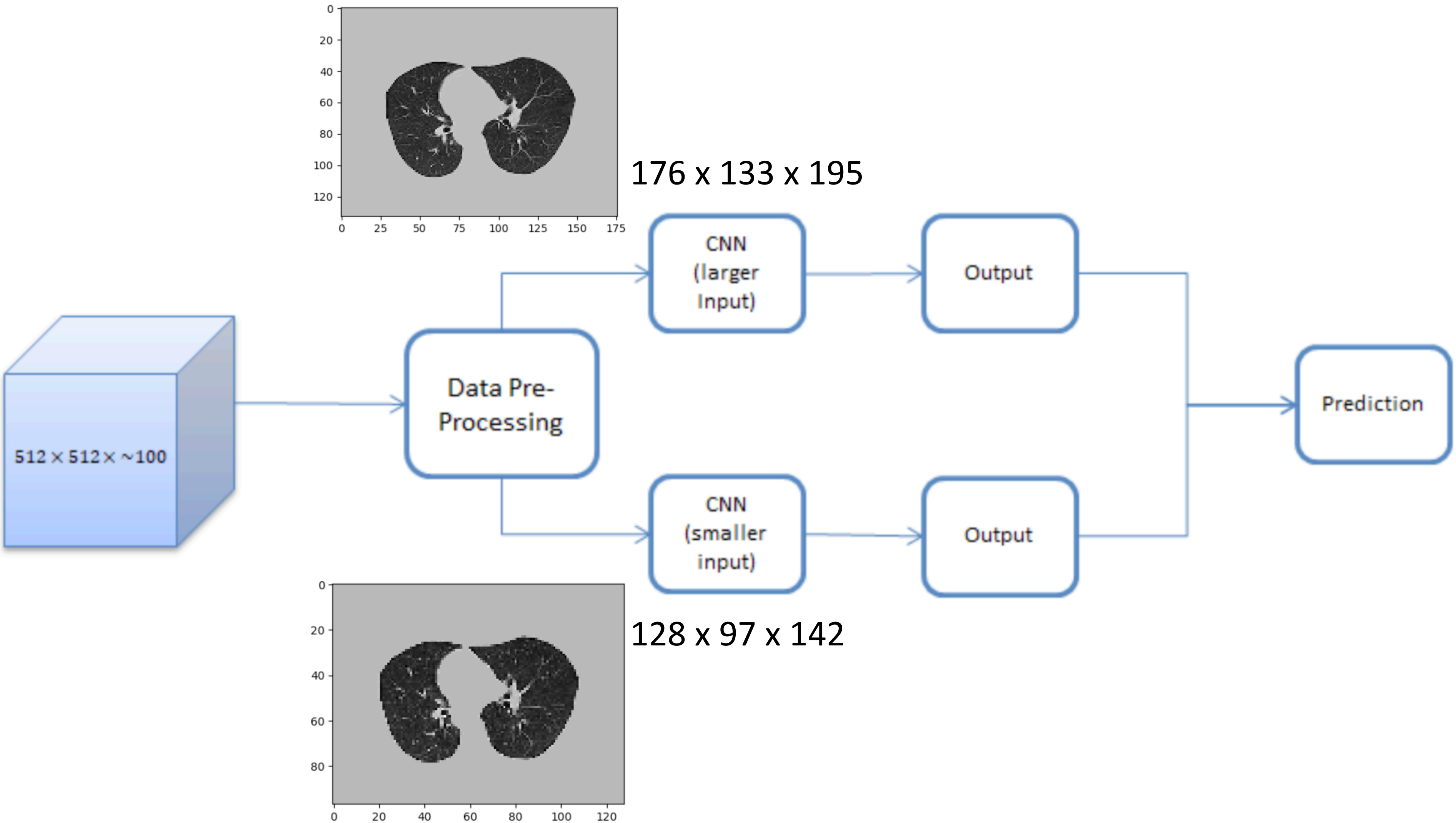
Pre-processing pipeline

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

Skin lesion detection, segmentation, and classification

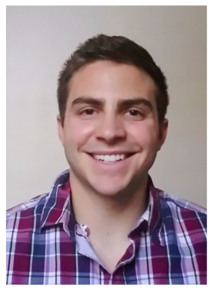
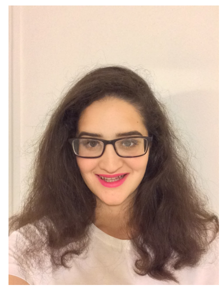
MIDDLE Research Group

Image Processing Group

Signal Theory and

Communications Department



Oge Marques Borko Furht Jack Burdick Janet Weinthal Adrià Romero López Xavier Giró-i.Nieto

FAU
COLLEGE OF ENGINEERING & COMPUTER SCIENCE
Florida Atlantic University

NSF

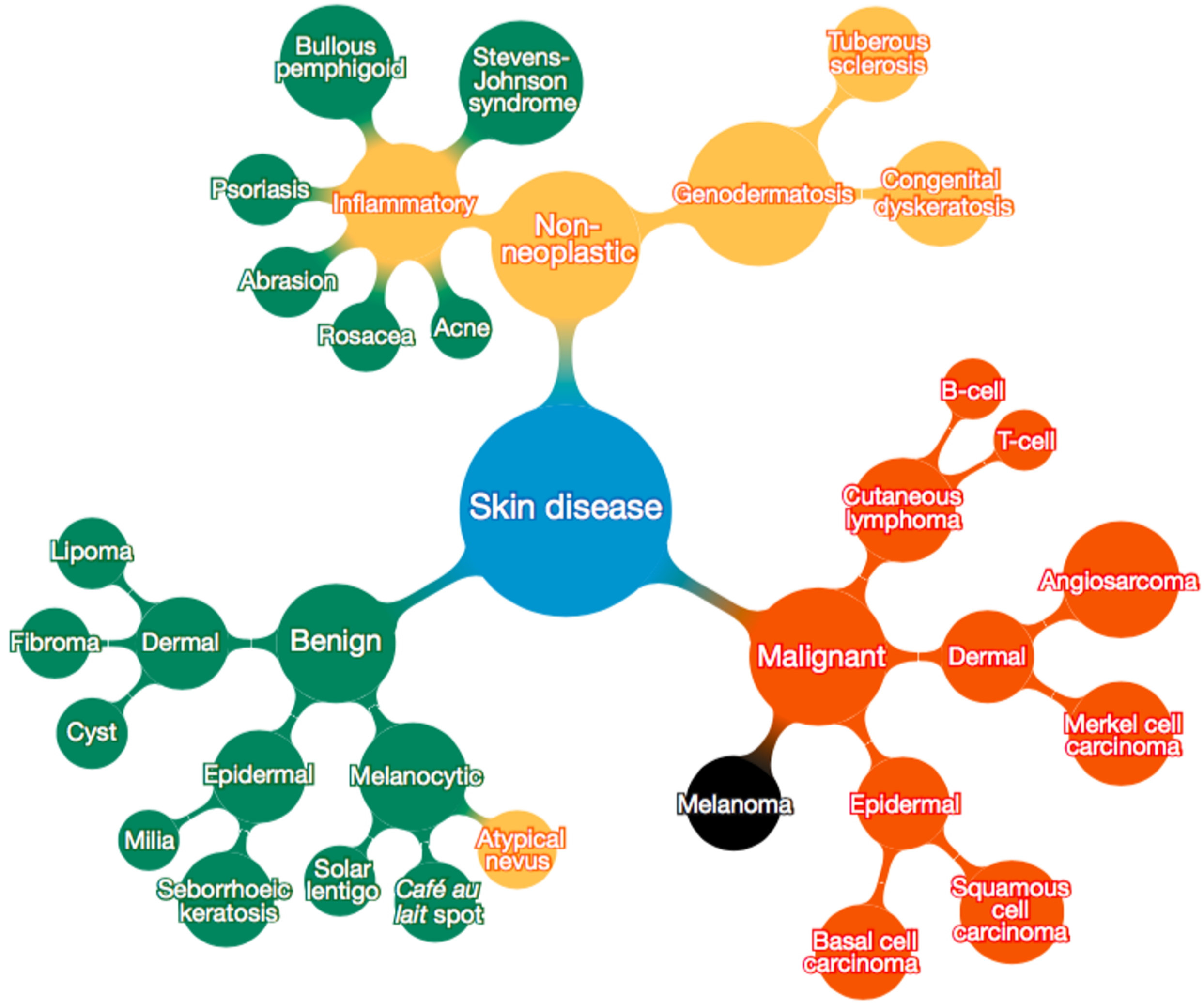
UPC
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

NSF Award No. 1464537, I/UCRC Phase II under NSF 13-542

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]

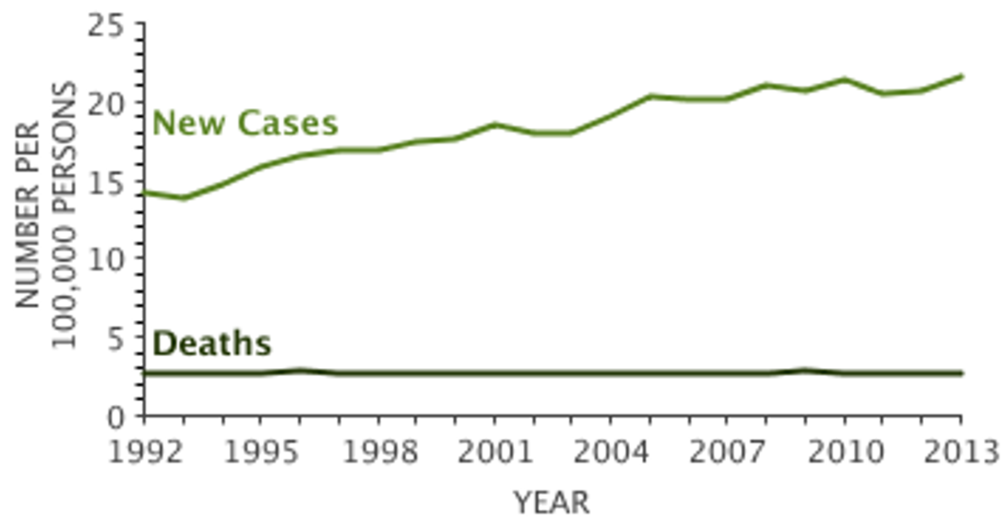
Skin Disease: An Illustrated Taxonomy



[Source: Esteva et al., Nature (2017)]

Melanoma Facts and Figures

Estimated New Cases in 2016	76,380
% of All New Cancer Cases	4.5%
Estimated Deaths in 2016	10,130
% of All Cancer Deaths	1.7%



- Melanoma is a **deadly** form of skin cancer without early detection and diagnosis
- **99%** survival rate in Stage I vs.
  14% survival rate in Stage II

[Source: https://seer.cancer.gov/statfacts/html/melan.html]

The ABCDE Rule

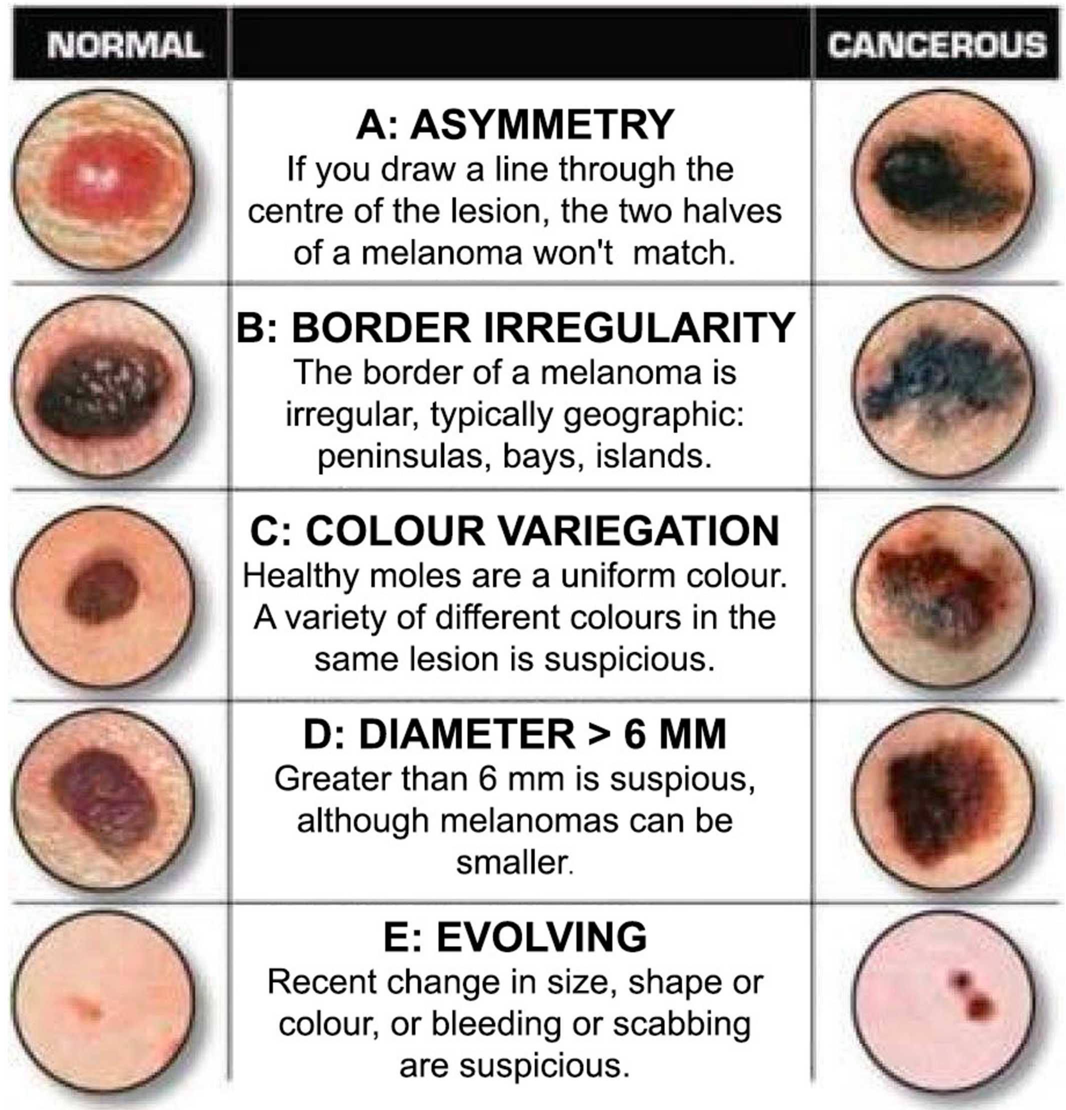
CASH (Henning et al., 2007)

Color

Architecture

Symmetry

Homogeneity



[Source: Center For Excellence In Dermatology - Kennewick, WA]

A Challenging Problem

Success rate (sensitivity)

Physicians, as low as:

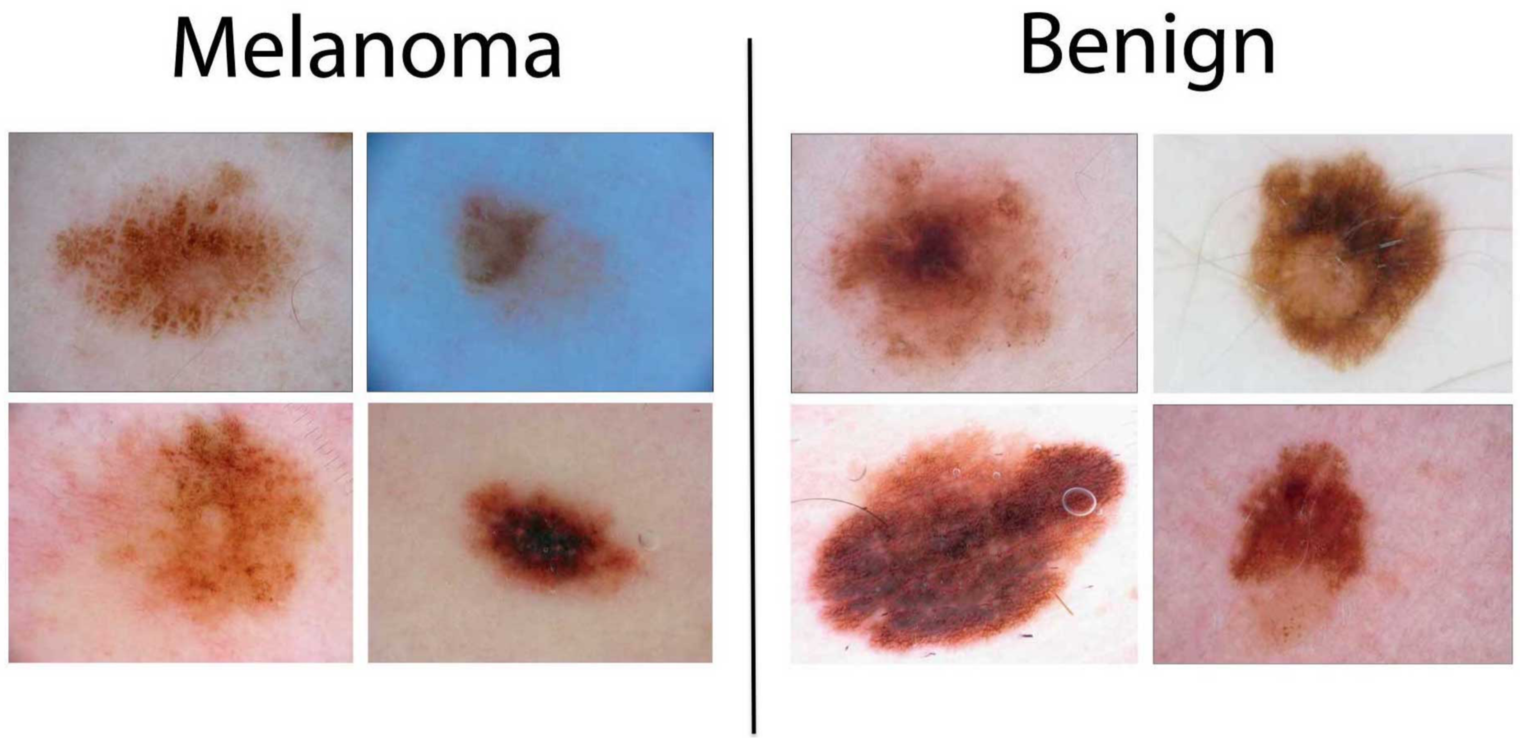
- 43 % - naked eye
- 79 % - dermoscopy

(Vestergaard et al, 2008)

Deep learning based methods, as high as: 94.83 % (Jafari et al, 2016; Premaladha and Ravichandran 2016)

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives\ +\ number\ of\ false\ negatives}$$

A Challenging Problem



Early work

- ISBI 2016 Challenge / ISIC Archive Dataset

	Class		
	Benign	Malignant	**Total Images**
Training subset	727	173	**900**
Testing subset	304	75	**379**

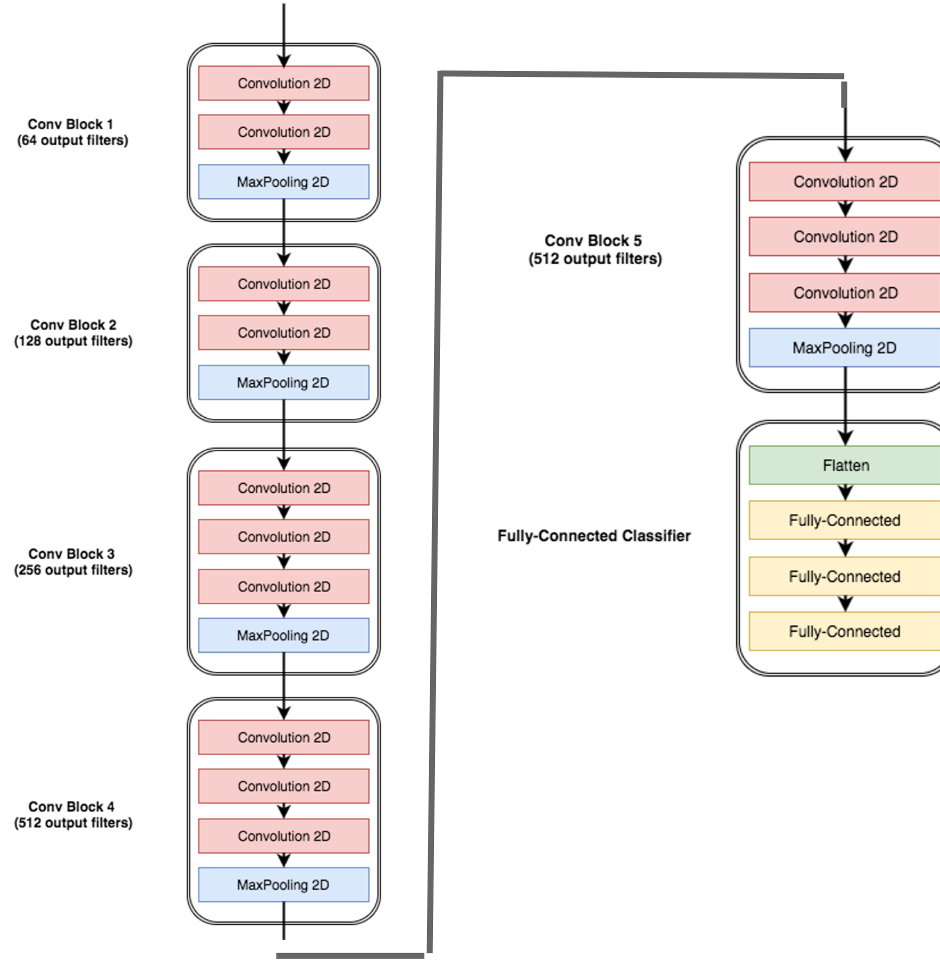
A. Romero Lopez, X. Giro-i-Nieto, J. Burdick, and **O. Marques**, "Skin lesion classification from dermoscopic images using deep learning techniques", *IASTED International Conference on Biomedical Engineering*, Innsbruck, Austria, February 2017. DOI: 10.2316/P.2017.852-053

One Problem, Three Possible Solutions

1. Training the VGG-16 from **scratch**

2. Using the VGG-16 as a **feature extractor**

3. Fine-tuning the VGG-16 architecture

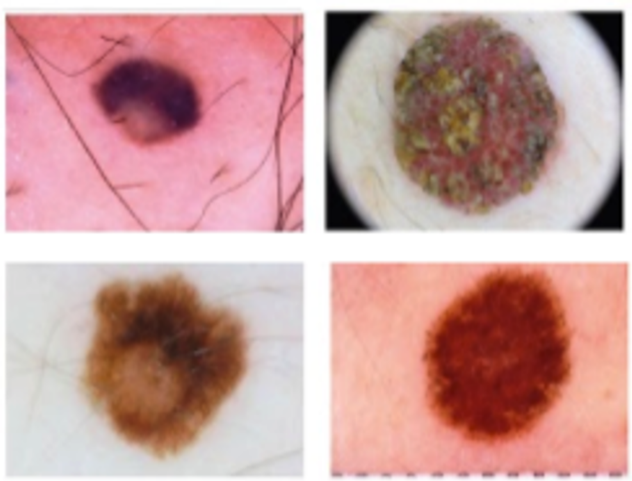


A. Romero Lopez, X. Giro-i-Nieto, J. Burdick, and **O. Marques**, "Skin lesion classification from dermoscopic images using deep learning techniques", *IASTED International Conference on Biomedical Engineering*, Innsbruck, Austria, February 2017. DOI: 10.2316/P.2017.852-053

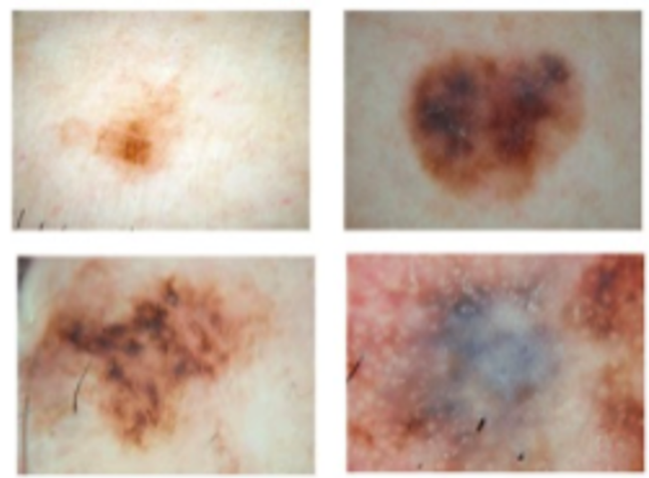
Classification Evaluation on Test Set

Model	Loss	Accuracy	Sensitivity	Precision
1. From scratch	0.6743	66.00 %	0.5799	0.6777
2. As feature extractor	1.0306	68.67 %	0.3311	0.4958
3. Fine-tuning	**0.4337**	**81.33 %**	**0.7866**	**0.7974**

False Positives



False Negatives

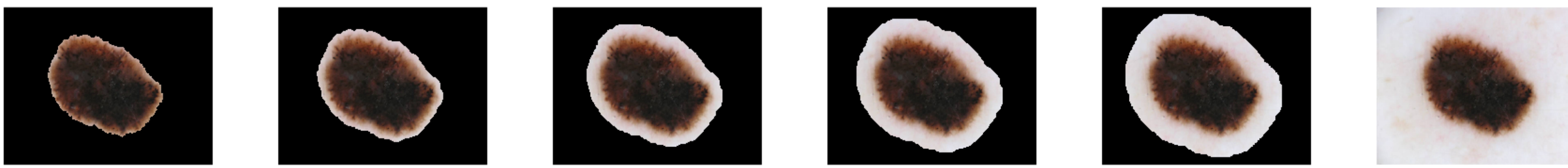


A. Romero Lopez, X. Giro-i-Nieto, J. Burdick, and **O. Marques**, "Skin lesion classification from dermoscopic images using deep learning techniques", *IASTED International Conference on Biomedical Engineering*, Innsbruck, Austria, February 2017. DOI: 10.2316/P.2017.852-053

Rethinking Skin Lesion Segmentation in a Convolutional Classifier

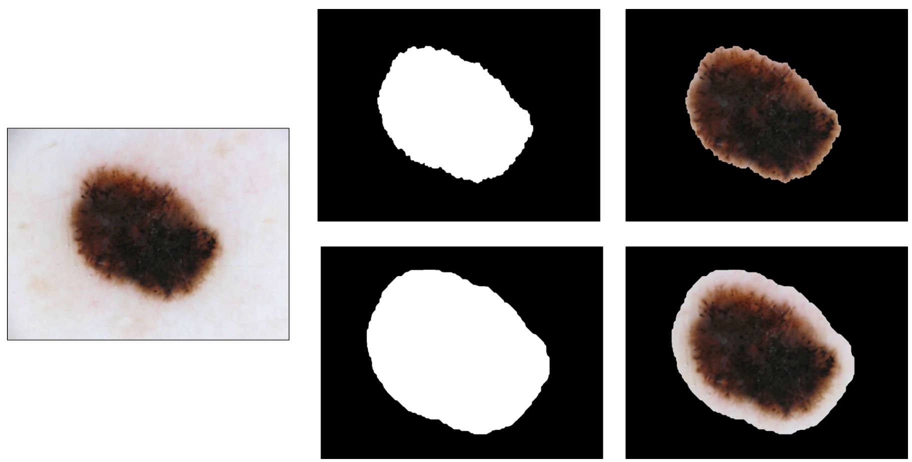
Jack Burdick[1] · Oge Marques[1] · Janet Weinthal[1] · Borko Furht[1]

J Digit Imaging (2018) 31:435–440
https://doi.org/10.1007/s10278-017-0026-y



Our Hypothesis

- Image segmentation improves the performance of skin lesion classifiers using convolutional neural networks.



[Source:International Skin Imaging Collaboration Archive]

J. Burdick, **O. Marques**, J. Weinthal, and B. Furht, "Rethinking Skin Lesion Segmentation in a Convolutional Classifier", *Journal of Digital Imaging* (2017) https://doi.org/10.1007/s10278-017-0026-y

Approach and results

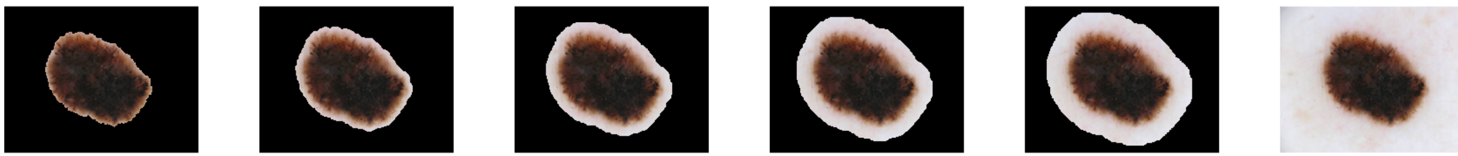
- VGG16 (Simonyan & Zisserman, 2014) + Transfer Learning

	Sensitivity	Accuracy	AUC
Perfect Segmentation	45.3%	58.7%	62.2%
Partial Segmentation	**56.0%**	**60.7%**	**62.6%**
Unsegmented	24.0%	51.3%	53.2%

J. Burdick, **O. Marques**, J. Weinthal, and B. Furht, "Rethinking Skin Lesion Segmentation in a Convolutional Classifier", *Journal of Digital Imaging* (2017) https://doi.org/10.1007/s10278-017-0026-y

Further Investigation

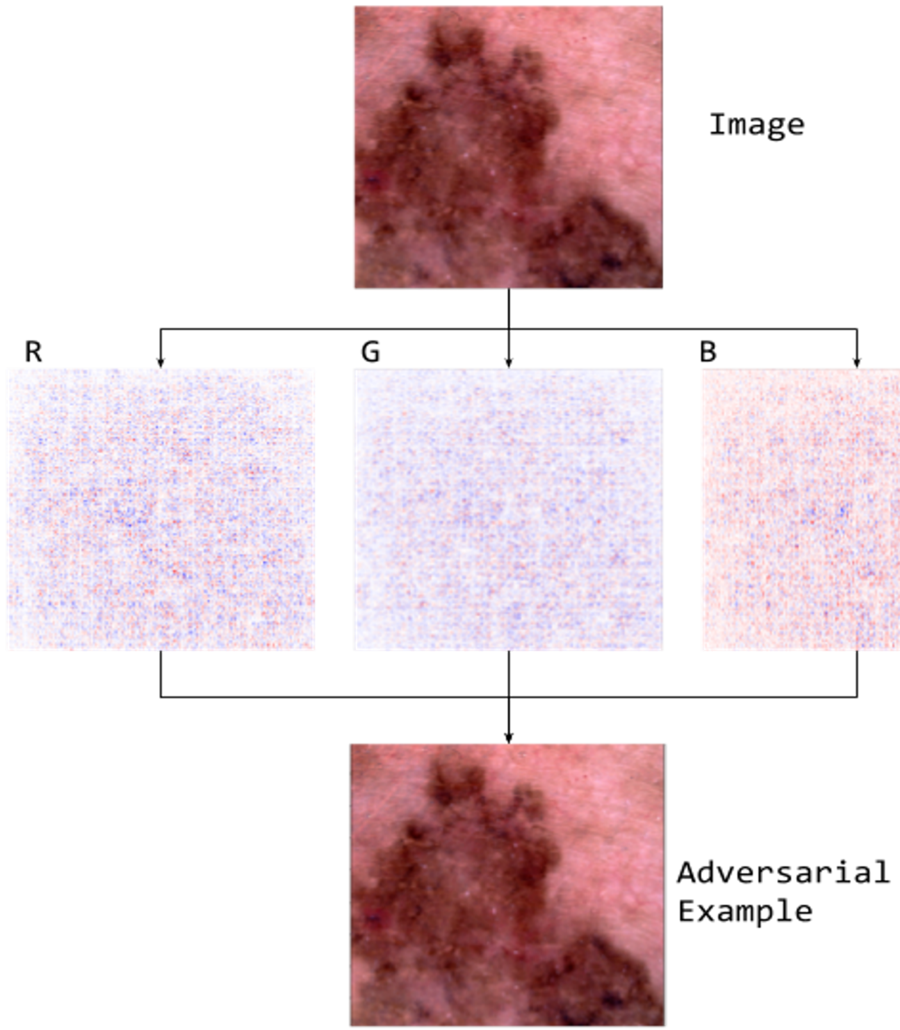
- What if we vary the degree of border expansion?



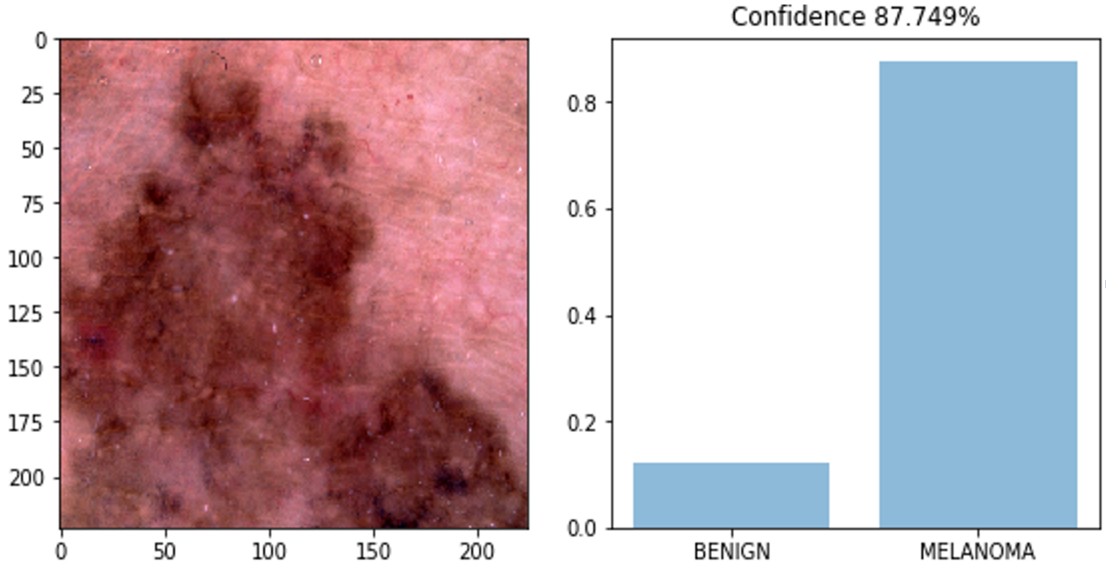
	Sensitivity	Accuracy	AUC
Perfect Segmentation	45.3%	58.7%	62.2%
+25	53.3%	61.3%	64.2%
+50	56.0%	60.7%	62.6%
+75	57.3%	59.3%	60.8%
+100	34.7%	55.3%	57.9%
Unsegmented	24.0%	51.3%	53.2%

J. Burdick, **O. Marques**, J. Weinthal, and B. Furht, "Rethinking Skin Lesion Segmentation in a Convolutional Classifier", *Journal of Digital Imaging* (2017) https://doi.org/10.1007/s10278-017-0026-y

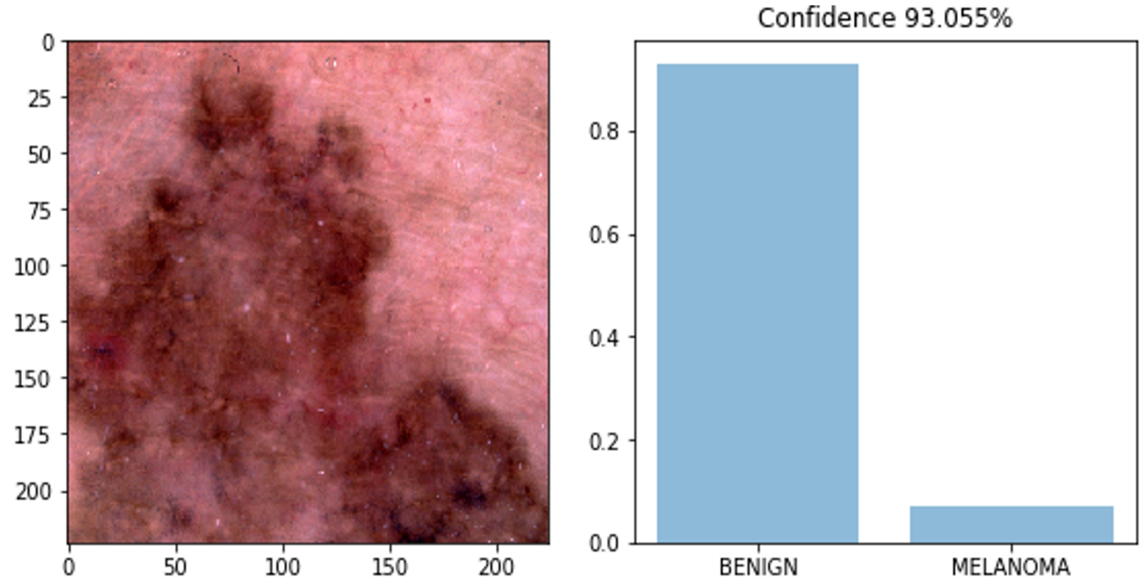
Adversarial Example



Adversarial Example



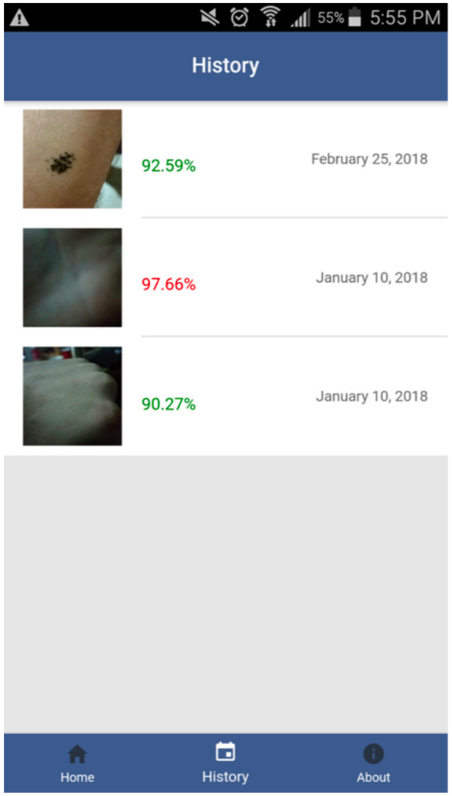
Unaltered Image



Adversarial Example

Ongoing Work

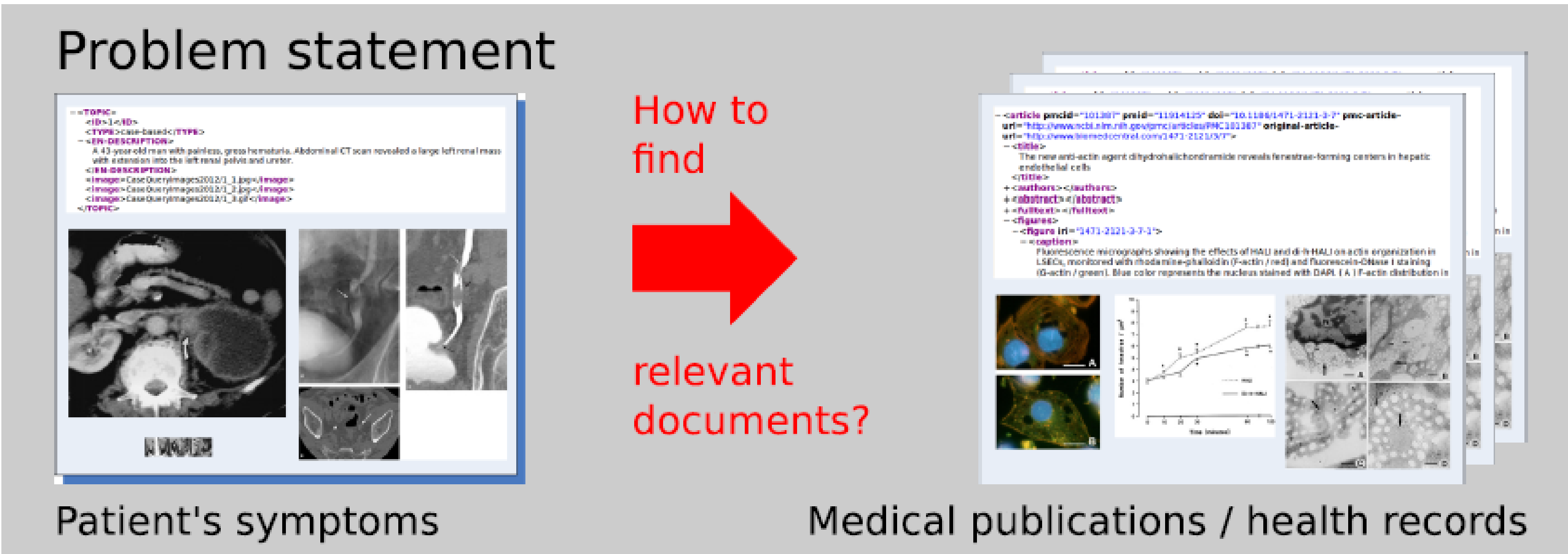
- Additional / larger / more challenging datasets
- Partnerships and collaborations
- Mobile app



Medical Case Retrieval (MCR)

The following slides are courtesy of **Dr. Mario Taschwer** (Alpen-Adria Universität, Klagenfurt, Austria)

Medical Case Retrieval (MCR)



- Major component of medical decision support systems based on case-based reasoning
- Solution may help to generate datasets for medical education and research

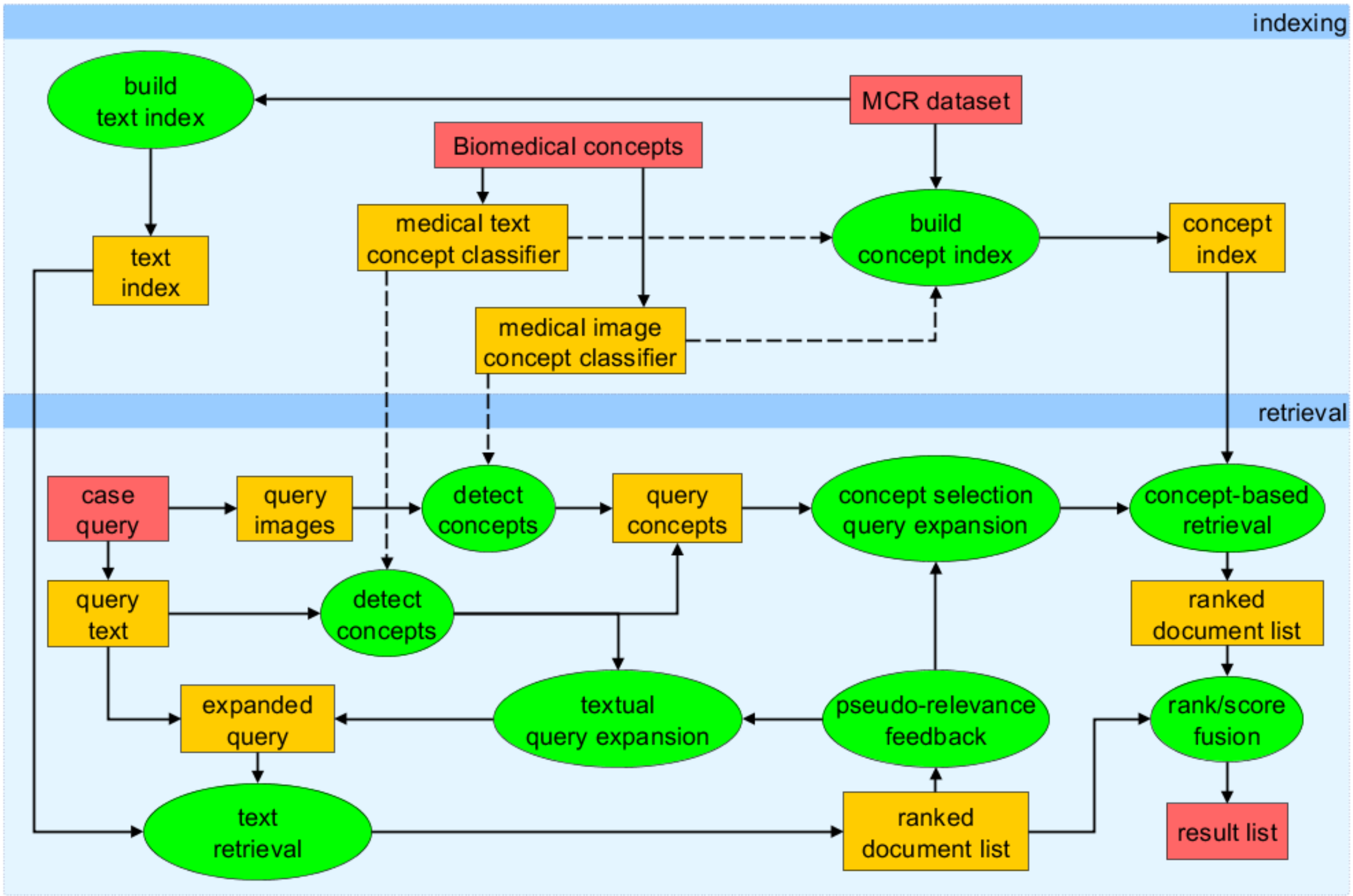
Problem Statement

- State of the art for MCR on general datasets:
  - Best systems employ purely textual techniques
- Main research problem:
  - How to improve MCR methods using textual and visual information?
- Hypothesis:
  - Biomedical concepts may help – with techniques:
  - Query or document expansion for text retrieval
  - Concept-based retrieval
  - Fusion of text and concept-based retrieval

CONTRIBUTIONS OF PHD THESIS

- Novel automatic methods for compound figure classification and separation
- Evaluation of concept mapping techniques:
  - New and existing methods of mapping text or images to biomedical concepts
- Comparison of query and document expansion by biomedical concepts for text-based MCR
- Novel framework combining text and concept-based retrieval, improving over state of the art

RETRIEVAL FRAMEWORK



Further Work

- Concept mapping:
  - Extended evaluation of string matching and image-to-concept mapping algorithms
  - Utilize other biomedical vocabularies and ontologies
  - Evaluate concept mapping by multi-view learning
  - Perform a study of manual MeSH annotations
  - Acquire an MCR dataset with more complete ground-truth MeSH annotations and relevance judgments
  - Apply deep learning to concept mapping (recent advances in image caption generation)

Further Work

- Learning from medical expert users:
  - Use relevance feedback for short-term or long-term learning
  - Apply transductive (semi-supervised) techniques for long-term learning, e.g. manifold-ranking
  - Consider active learning approaches to cope with the small sample size problem for long-term learning

At NIH

(Visiting Research Scientist, Oct-Nov 2018)

(Biomedical) video summarization

2017 IEEE 30th International Symposium on Computer-Based Medical Systems

Novel Method for Storyboarding Biomedical Videos for Medical Informatics

Sema Candemir, Sameer Antani, Zhiyun Xue, George Thoma
Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine, NIH, Bethesda, MD, USA
(sema.candemir, sameer.antani, xuez, george.thoma)@nih.gov

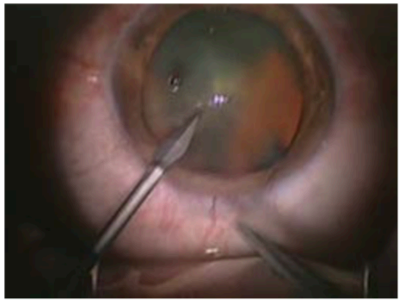
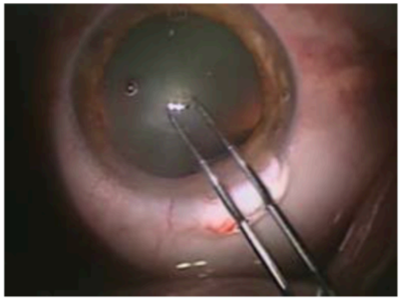
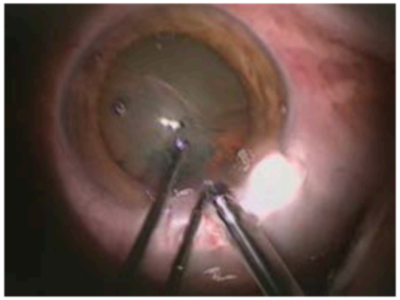
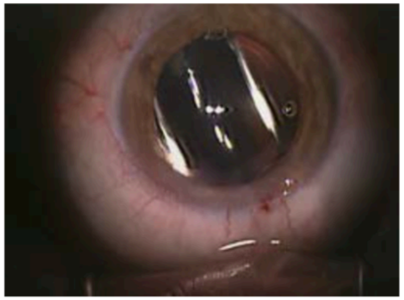
(Biomedical) video summarization

- Questions
  - What makes biomedical video different?
  - How could this be inspired by existing tools (e.g., OSUM)?
  - How could this enrich the functionality of existing tools (e.g., Open-i)?

- Goals and planned deliverables
  - Survey of video summarization / abstraction in the "deep learning era"
  - Detailed outline of working plan for future months/years
  - Possible collaborative work: MS/PhD/Post-docs, grant proposals, publications.

Possible partnership

- OVID Project
  - Relevance Detection in Ophthalmic Surgery Videos (Oct 2018 – Oct 2021)
  - First dataset (*Cataract-101 Video Dataset*) publicly available https://zenodo.org/record/1220951

Possible partnership

- KISMET: Knowledge & Information Sharing in Medical Expert Teams (2015-2019)
  - Focus: endoscopy in gynecology, particularly endometriosis
  - Datasets (e.g., LapGyn4) and tools (e.g., ECAT)
    - **The ITEC LapGyn4 Gynecologic Laparoscopy Image Dataset**
    - 500+ gynecologic laparoscopic surgeries
    - Four collections: general surgical actions, anatomical structures, actions on specific anatomy, and examples of differing amounts of visible instruments

A. Leibetseder, S. Petscharnig, M. J. Primus, S. Kletz, B. Münzer, K. Schoeffmann, J. Keckstein. 2018. Lapgyn4: a dataset for 4 automatic content analysis problems in the domain of laparoscopic gynecology. In Proceedings of the 9th ACM Multimedia Systems Conference (MMSys '18). ACM, New York, NY, USA, 357-362.

ECAT (Endoscopic Concept Annotation Tool)

a) editing window
b) statistics tab
c) case overview
d) uniform sampled grid tab
e) classification feedback visualization.

B. Münzer, A. Leibetseder, S. Kletz, K. Schöffmann, ECAT - Endoscopic Concept Annotation Tool, Demo Paper at 24th ACM Intl. Conference on Multimedia Modeling (MMM) 2019.

Concluding remarks

FAU FLORIDA ATLANTIC UNIVERSITY

Oge Marques, Ph.D.
Professor

Computer & Electrical Engineering and Computer Science (CEECS)
777 Glades Road
Boca Raton, FL 33431-0991
tel: 561.297.3857
fax: 561.297.2800
email: omarques@fau.edu
Skype: ProfessorOge

ProfessorOgeMarques

@ProfessorOge