COMPUTER VISION 2.0

Grzegorz Chlebus

05.07.2018

Science Night @ SBMC 2018

VOL. 81 NO. 2

Radiology

AUGUST 1963

a monthly journal devoted to clinical radiology and allied sciences

PUBLISHED BY THE RADIOLOGICAL SOCIETY OF NORTH AMERICA, INC.

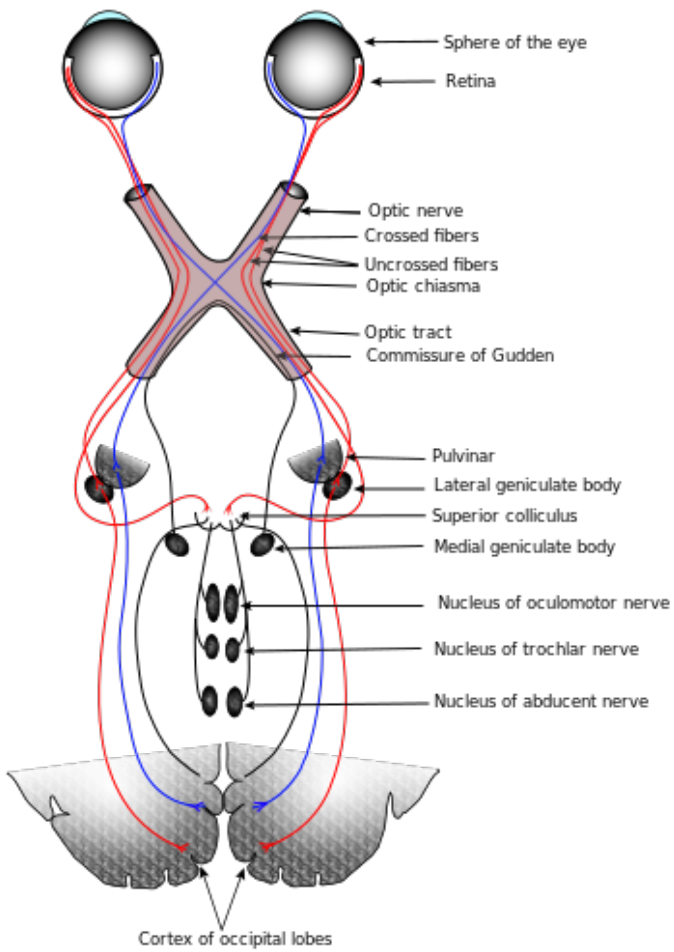
The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer[1]

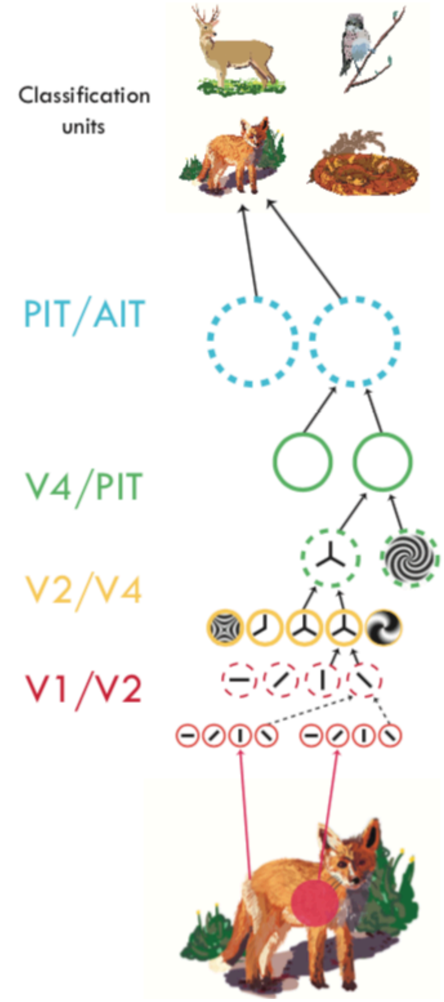
GWILYM S. LODWICK, M.D., THEODORE E. KEATS, M.D., and JOHN P. DORST, M.D.

THIS PAPER WILL DESCRIBE a concept of converting the visual images on roentgenograms into numerical sequences that can be manipulated and evaluated by the digital computer and will report the results of employing this system to determine the significance of certain radiographic findings in lung cancer. The development of such a coding system for conveying radiographic information into an electronic data-processing system makes cause, against a background of air density, the intimate details of the relationship between tumor and host may be faithfully reproduced roentgenographically. Parenthetically, it may be stated that similar density ranges exist in the relationships between bone and soft tissue and that an equally effective descriptive system has been evolved for bone sarcomas (1, 2).

For this study, the roentgenograms of 541 cases of primary lung cancer were made

Computer Vision 1.0 (before 1990)

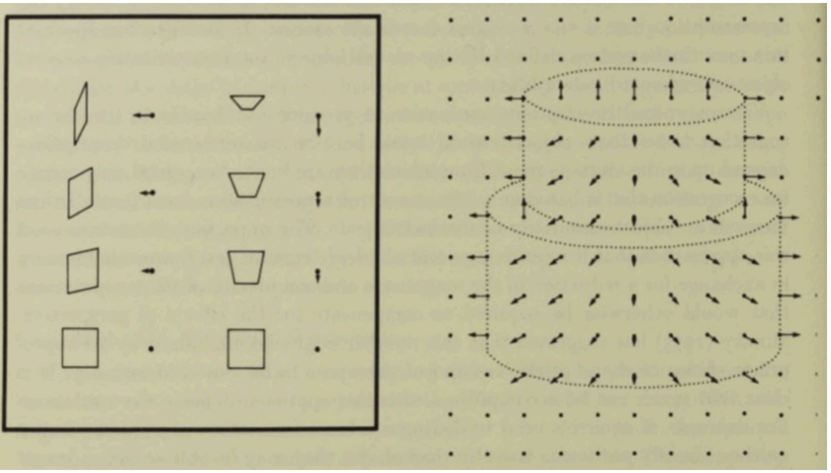


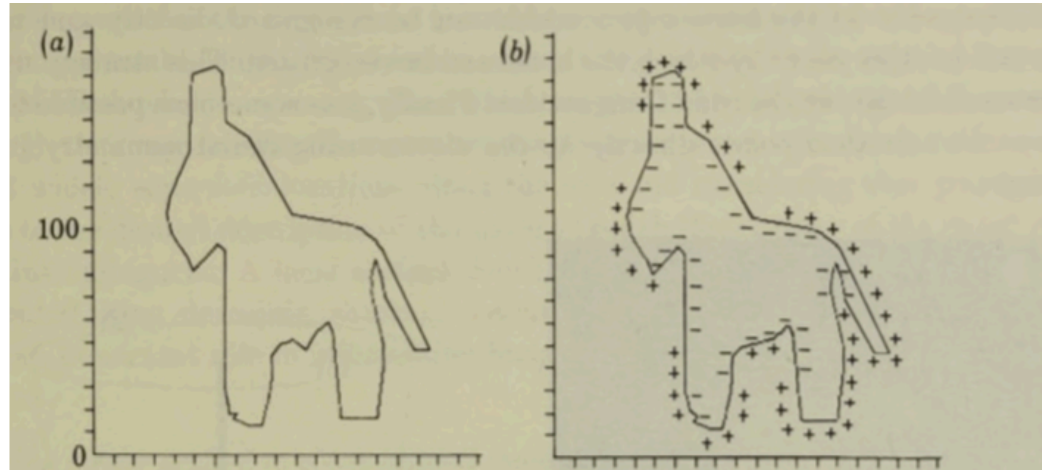


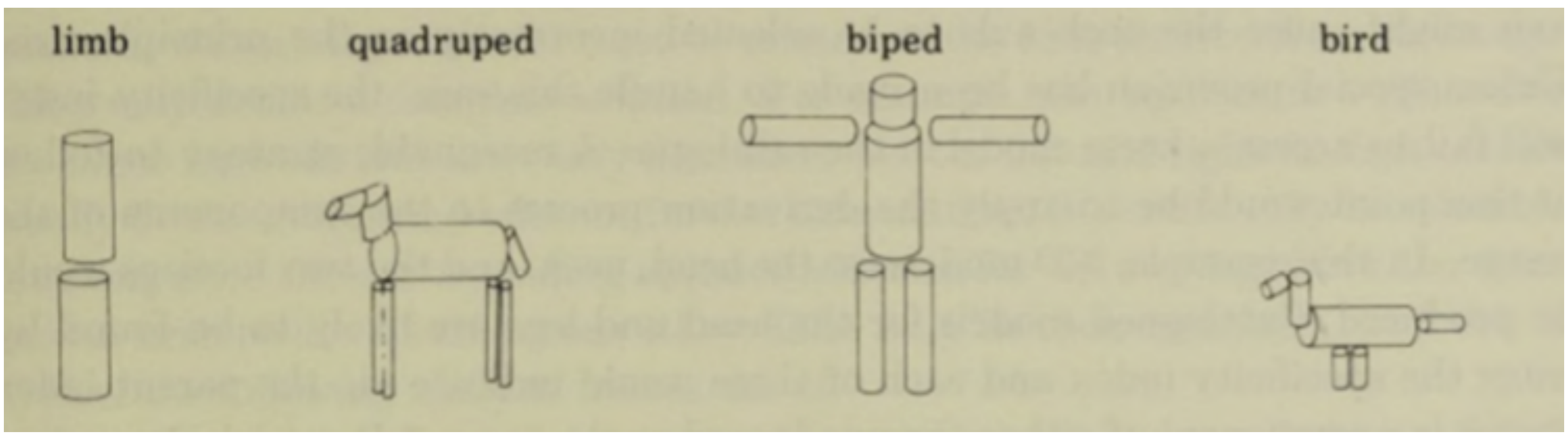
From: https://en.wikipedia.org/wiki/Visual_system

From: Serre T., "Hierarchical Models of the Visual System".

Computer Vision 1.0 (before 1990)







From: Marr D. “Representation and recognition of the spatial organization of three-dimensional shapes”, 1978.

Computer Vision 1.5 (1990-2012)

From:
https://commons.wikimedia.org/wiki/File:Cat03.jpg

Computer Vision 1.5 (1990-2012)

Vector describing image properties

From: https://commons.wikimedia.org/wiki/File:Cat03.jpg

Feature Extraction

From: http://areshmatlab.blogspot.com/2010/04/feature-detection-using-scale-invariant.html

Computer Vision 1.5 (1990-2012)

From:
https://commons.wikimedia.org/wiki/File:Cat03.jpg

Vector describing image properties

Feature Extraction

Classifier

From:
http://areshmatlab.blogspot.com/2010/04/feature-detection-using-scale-invariant.html

SVMs

Random Forests

K-Means

Computer Vision 1.5 (1990-2012)

From:
https://commons.wikimedia.org/wiki/File:Cat03.jpg

Vector describing image properties

Feature Extraction → Classifier →

Class scores

Cat: **0.95**

Tiger: 0.04

Dog: 0.01

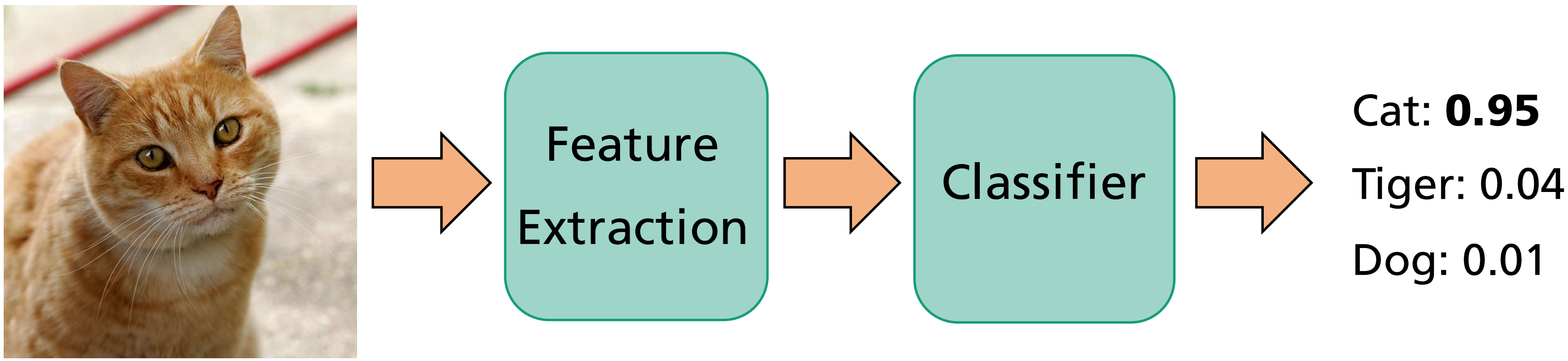
From:
http://areshmatlab.blogspot.com/2010/04/feature-detection-using-scale-invariant.html

SVMs

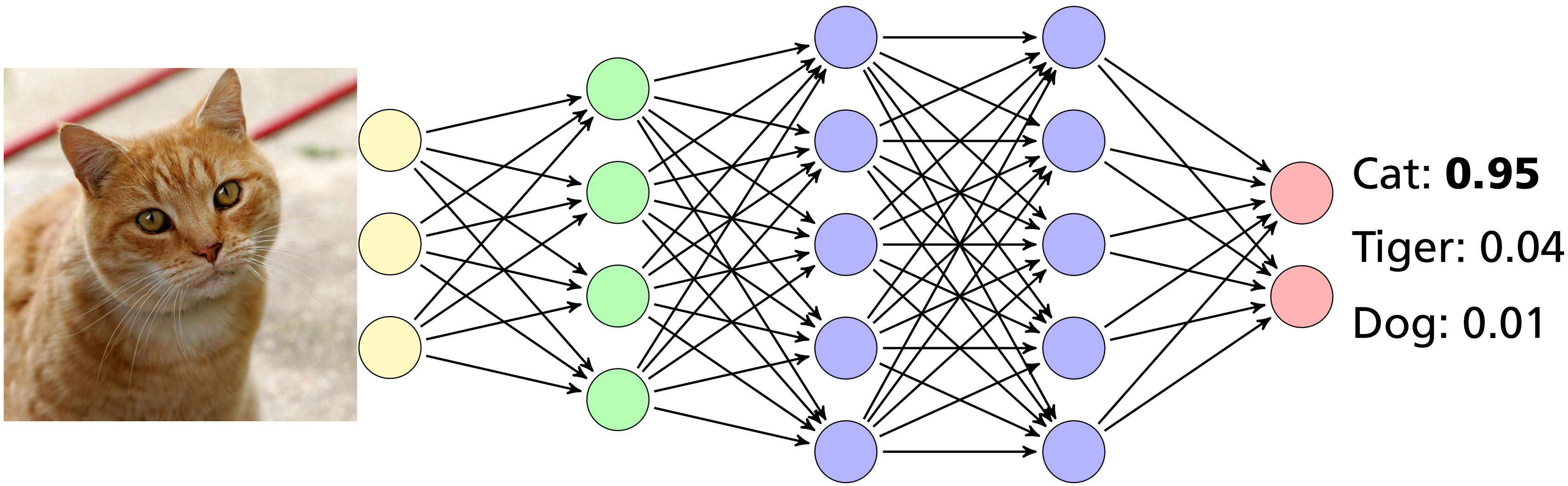
Random Forests

K-Means

Computer Vision 1.5 (1990-2012)



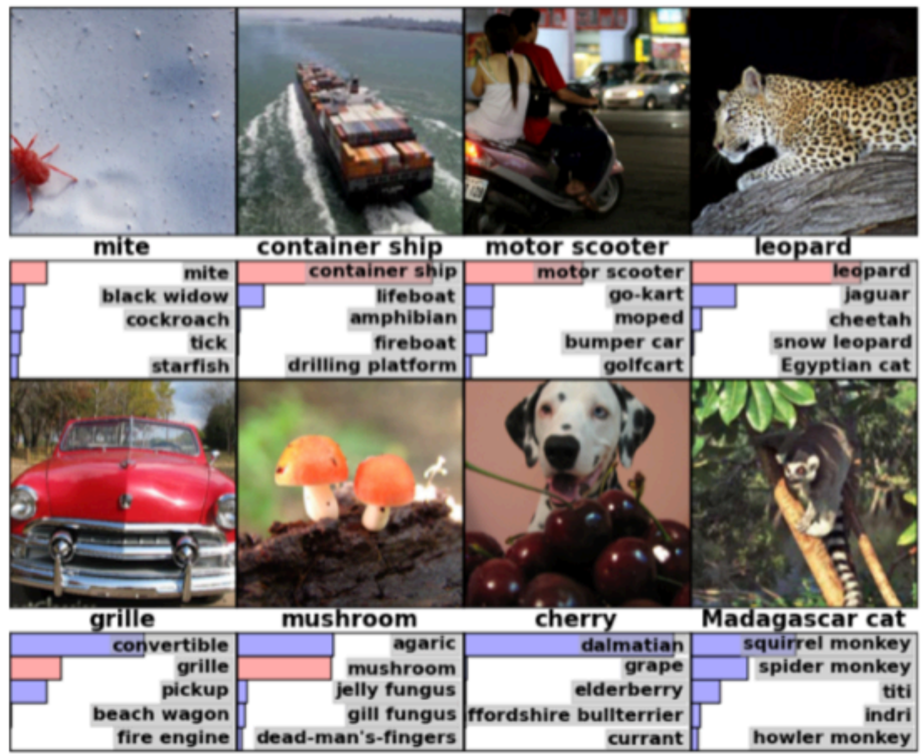
Computer Vision 2.0



Deep neural network

Computer Vision 2.0 beats humans

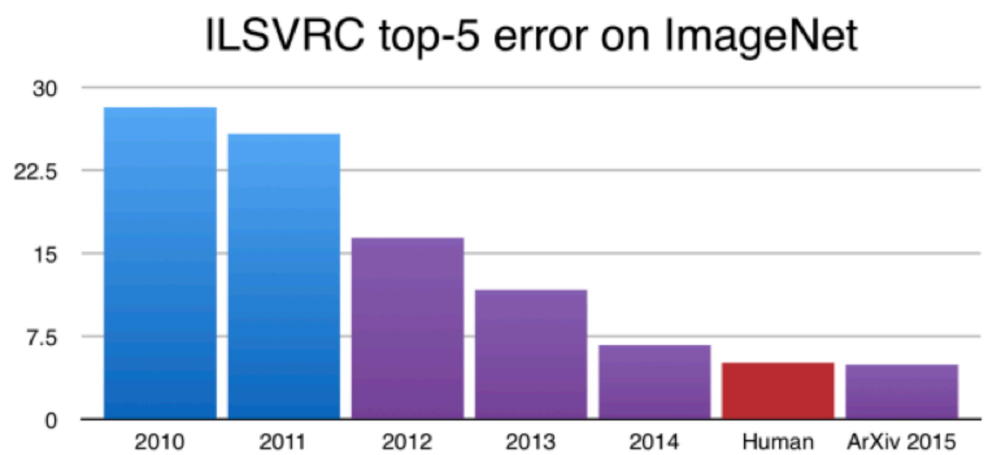
- ImageNet Large Scale Visual Recognition Competition
  - 1000 image classes
  - 1.2 M training images
  - 100k test images



From: https://blog.acolyer.org/2016/04/20/imagenet-classification-with-deep-convolutional-neural-networks/

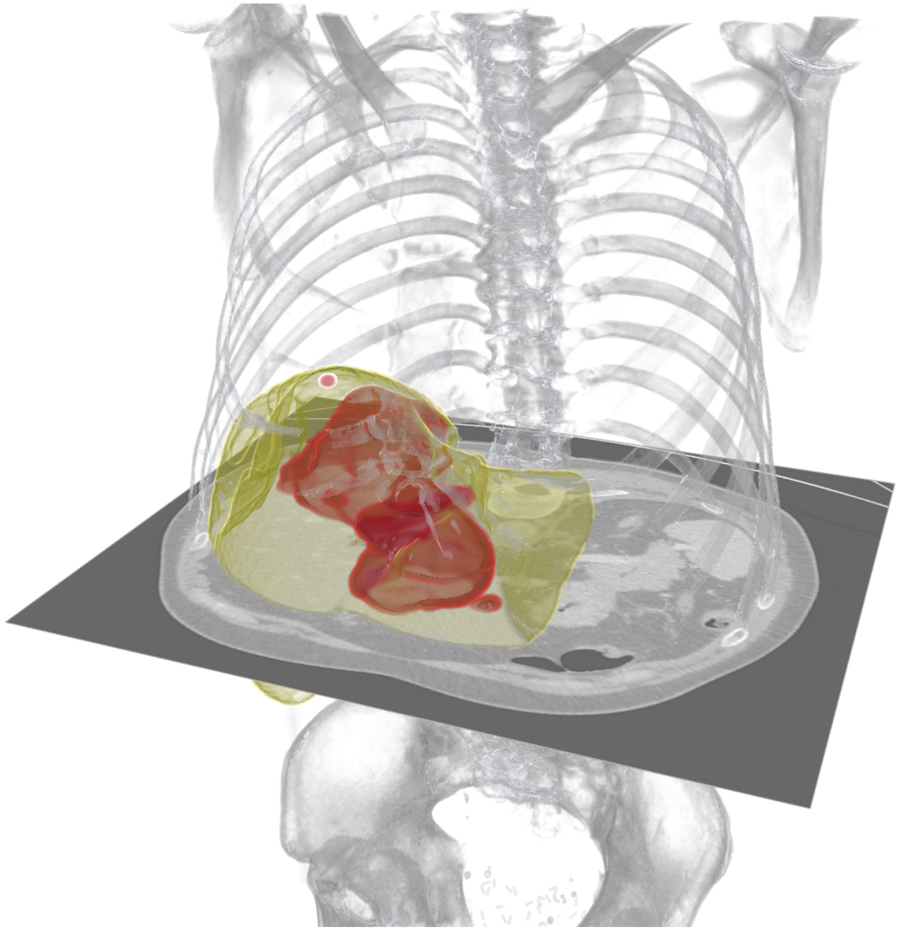
Computer Vision 2.0 beats humans

- ImageNet Large Scale Visual Recognition Competition
  - 1000 image classes
  - 1.2 M training images
  - 100k test images

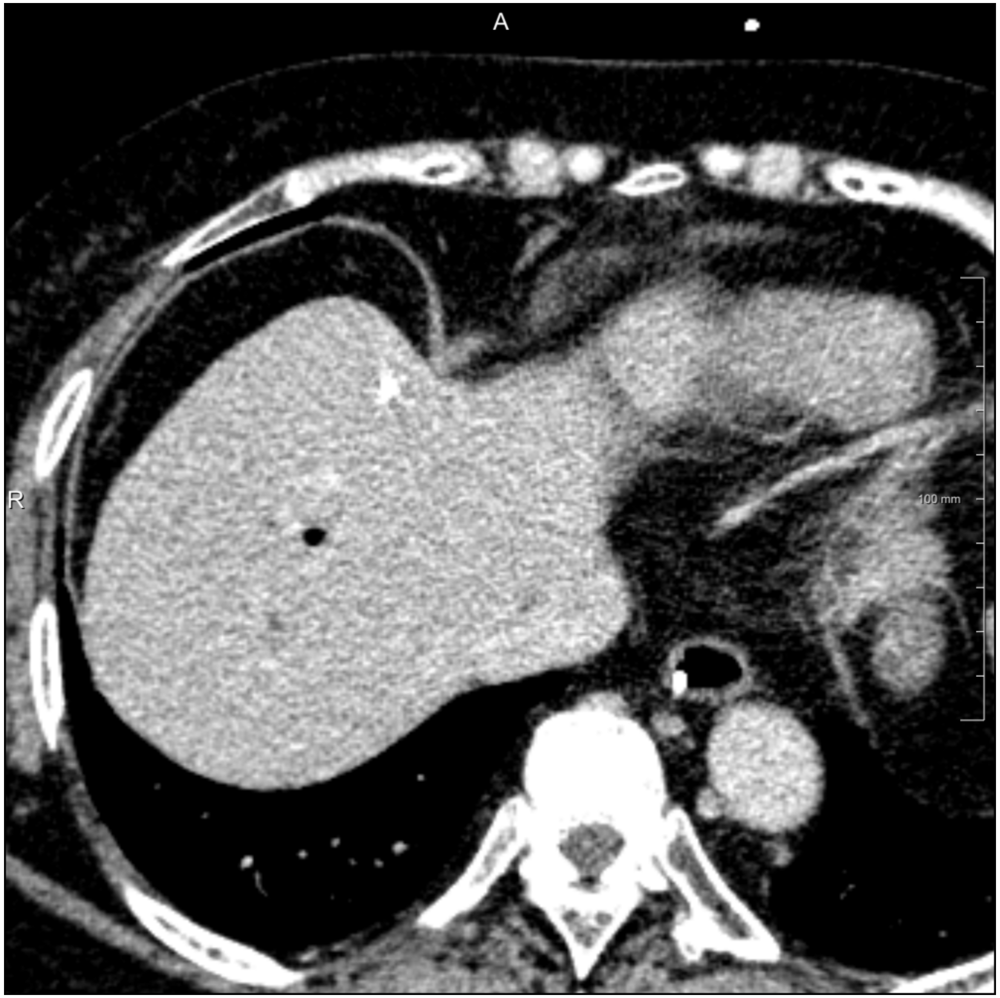


From: https://devblogs.nvidia.com/mocha-jl-deep-learning-julia/image1/

Example: CT Liver Segmentation



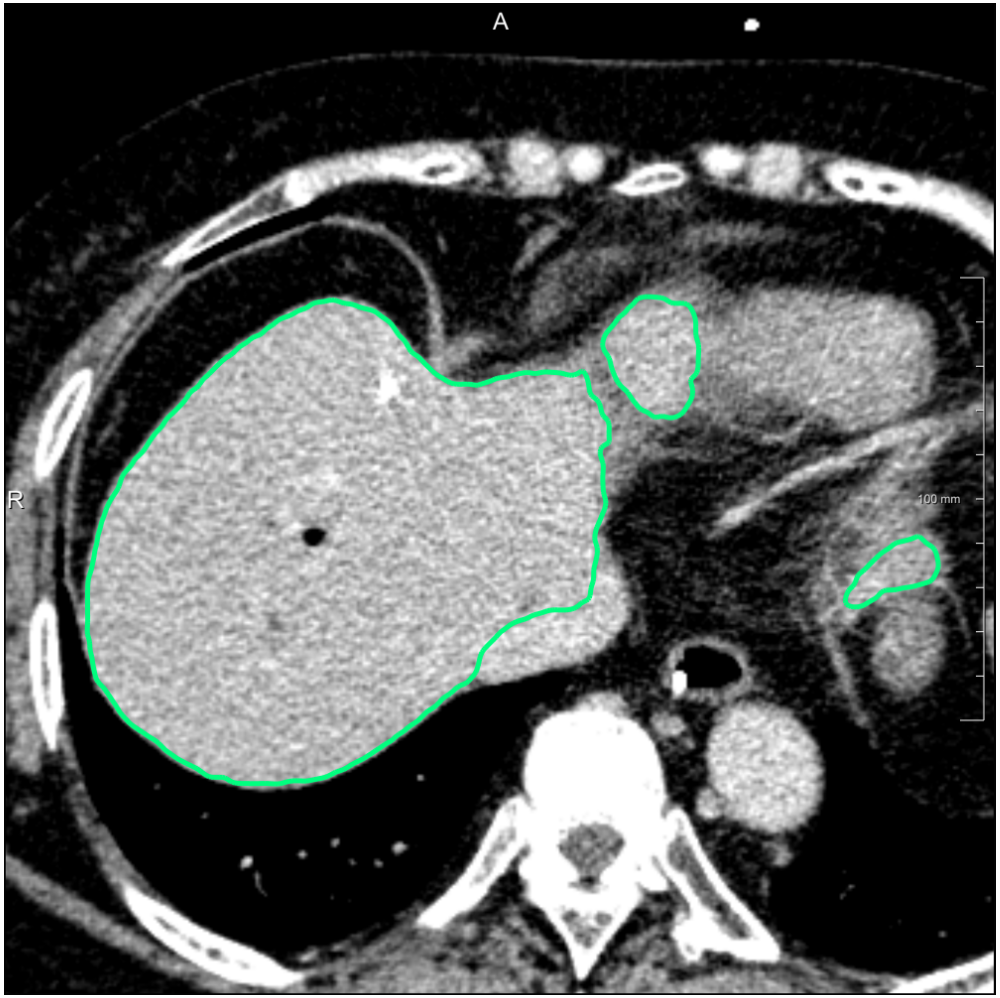
Example: CT Liver Segmentation



Medical Knowledge Through Research

Fraunhofer MEVIS

Example: CT Liver Segmentation



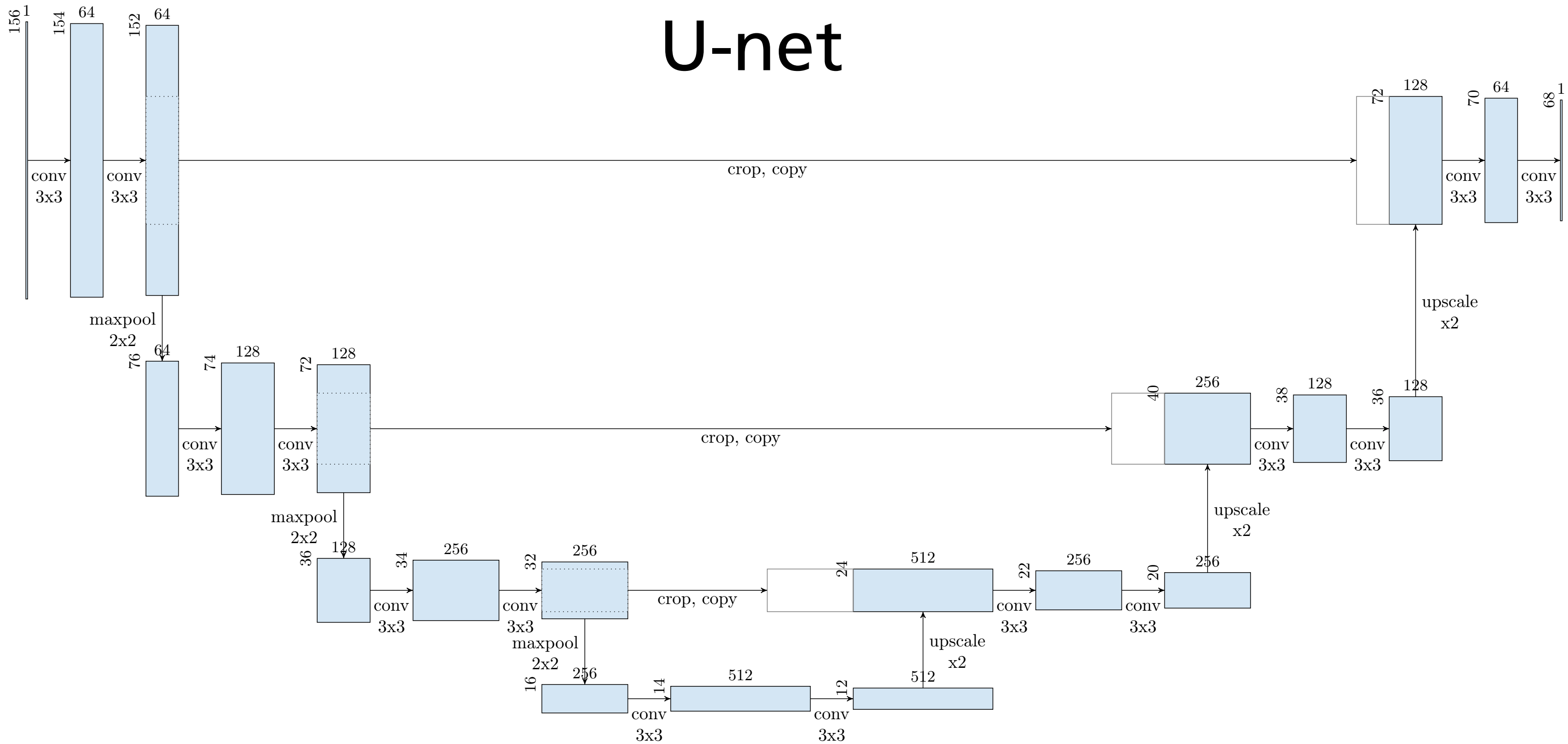
Medical Knowledge Through Research

Fraunhofer
MEVIS

Example: CT Liver Segmentation

Neural Network Architecture for Semantic Segmentation

U-net



Ronneberger O. et al., "U- Net: Convolutional Networks for Biomedical Image Segmentation", 2015.

Training

```
1. Training images with reference labels REF
2. Initialize neural network NN parameters randomly
3. DO
4.    Apply NN to a batch of training images → OUTPUT
5.    Compute the difference between OUTPUT and REF → LOSS
6.    Compute LOSS derivatives w.r.t. NN parameters → GRADIENTS
7.    Apply GRADIENTS to update NN parameters
8. UNTIL convergence
```

Training

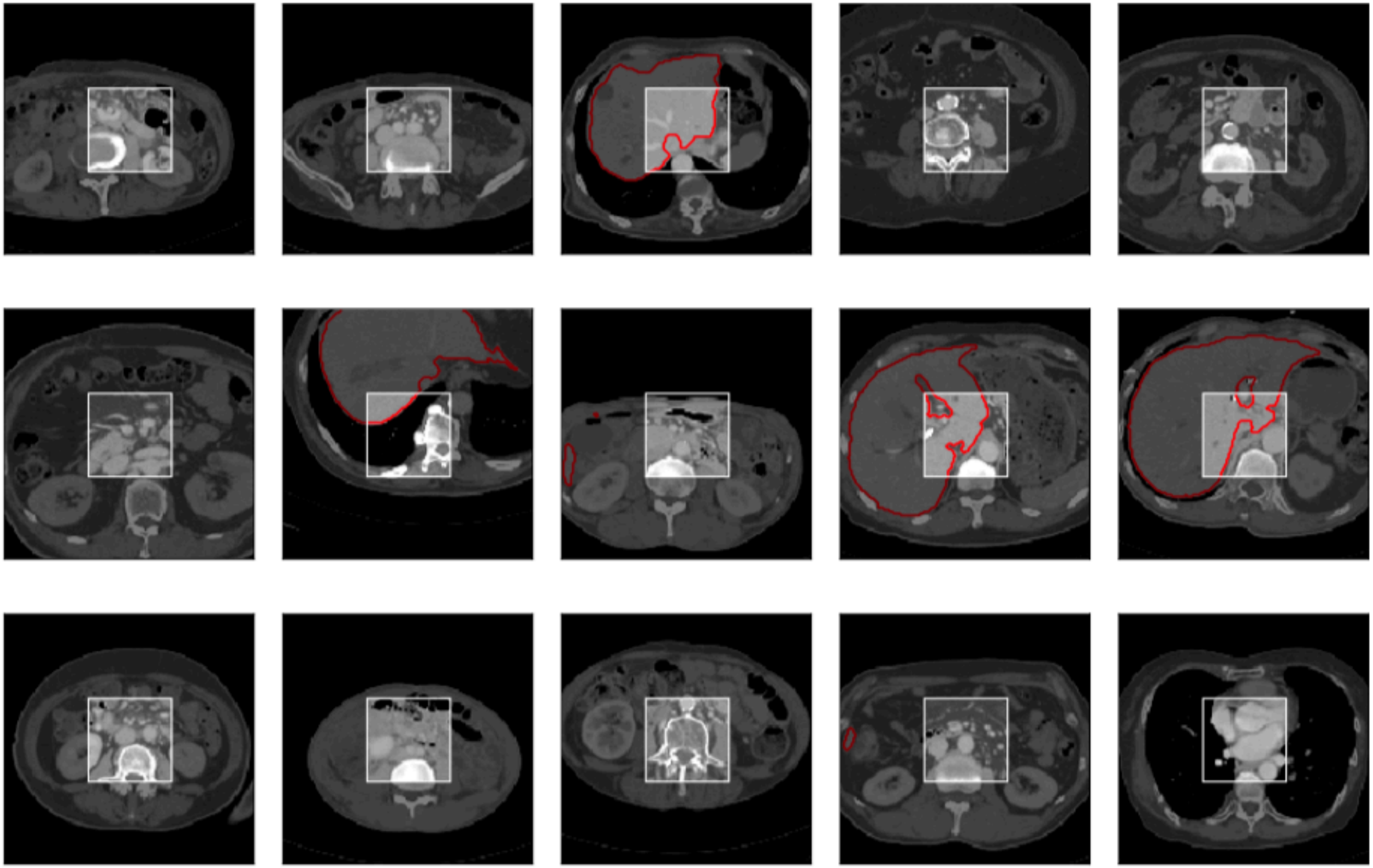
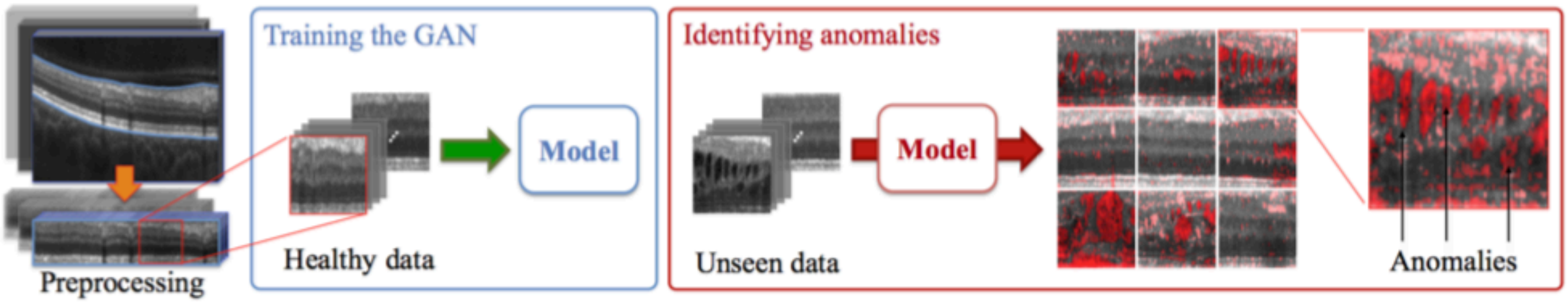


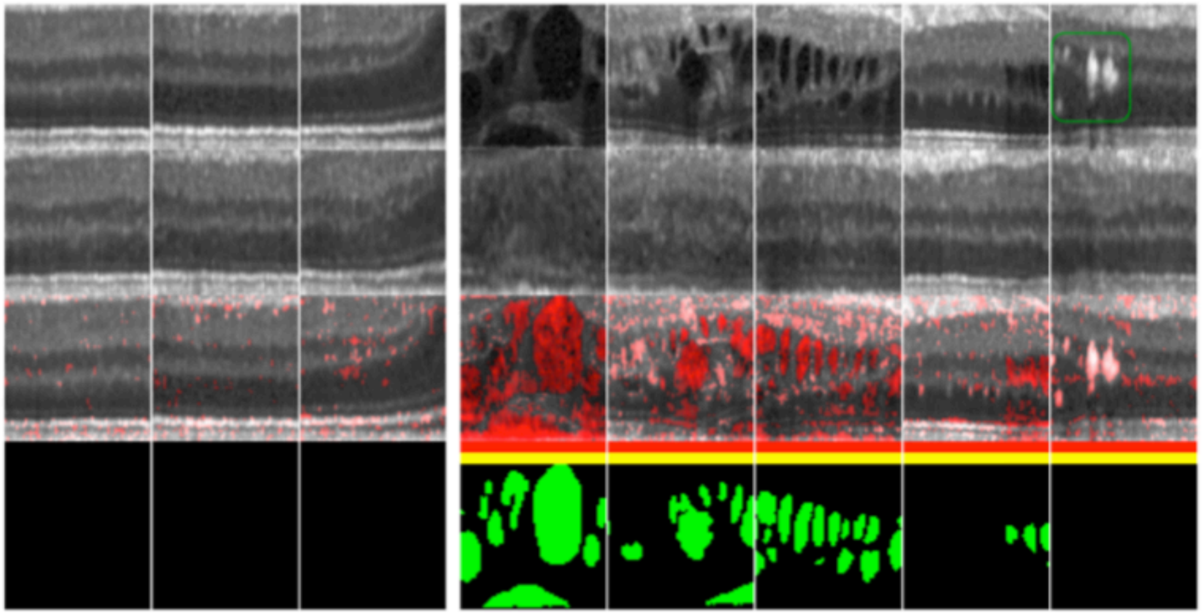
Image courtesy of Dr. Hans Meine


Liver Training Videos

Selected Deep Learning Applications in the Medical Context

Anomaly Detection





Schlegl T. et al., “Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery”, 2017.

Deep MR to CT Synthesis

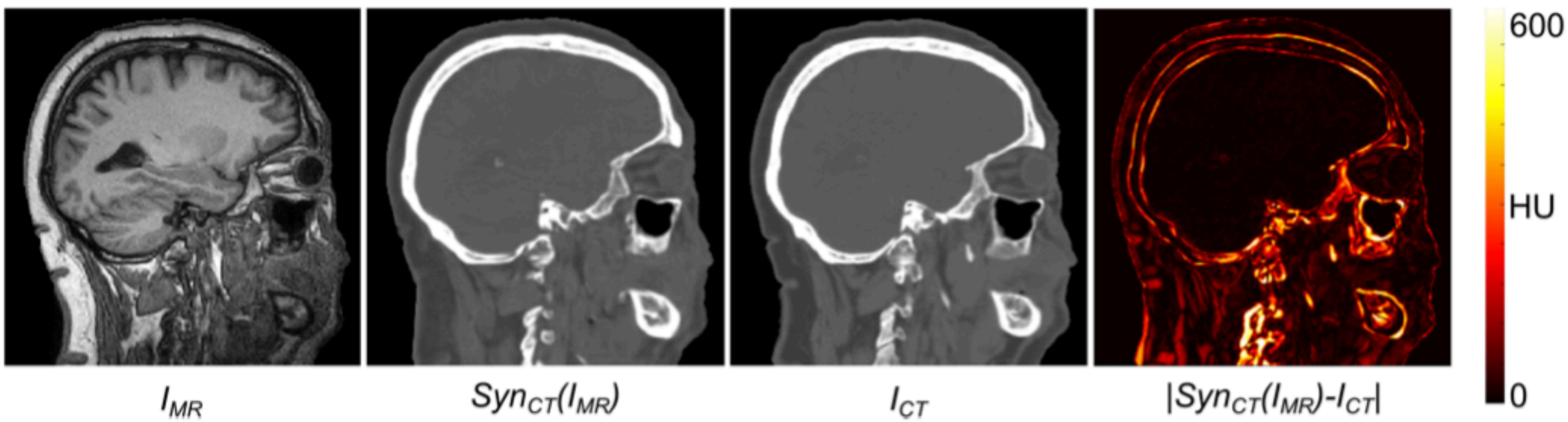


Fig. 4: *From left to right* Input MR image, synthesized CT image, reference real CT image, and absolute error between real and synthesized CT image.

Wolterink J.M. et al., "Deep MR to CT Synthesis using Unpaired Data", 2017.

Thank you for your attention ☺

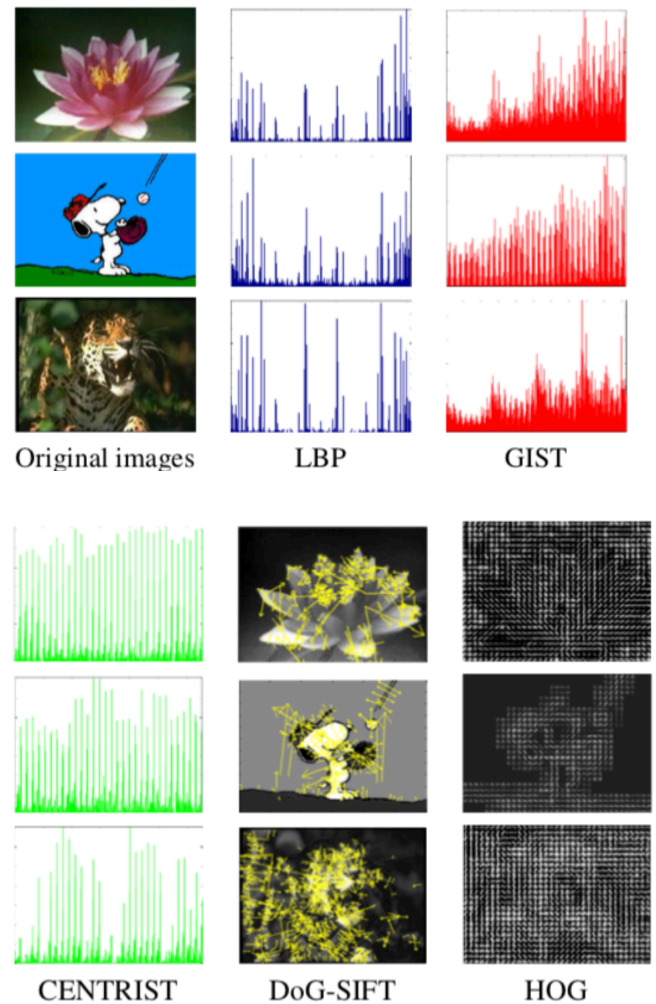
Questions?

Computer Vision 1.5 (1990-2012)

2.1. Image Feature Descriptors

In this paper, we unsupervised integrate five renowned descriptors, including DoG-SIFT [14], LBP [17], GIST [18], CENTRIST [22], and HOG [5]. Figure 1 demonstrates the visual patterns of each descriptor for sample images. Each representation corresponds to a single descriptor.

DoG-SIFT is originally designed for recognizing the same object appearing under different conditions and has been widely used in computer vision and image content retrieval. As a local descriptor, it is invariant to image rotation as well as scale. It is also robust across a substantial range of affine noise and change in illumination. There are several variations of SIFT descriptors (*e.g.* Dense SIFT [1]) in literature. In order to fairly compare our method to existing unsupervised scene categorization methods [11, 7], we use DoG-SIFT to be consistent with their selection. **LBP** is a powerful texture feature based on occurrence histogram of local binary patterns. It emphasizes the local structure and is famous for its robustness to rotation and non-uniform illumination. **HOG** is a good local descriptor to describe the shape information of the image. Differing to SIFT which describes the feature at the candidate location (keypoint), HOG describes the feature over the given region. **GIST** encodes rough geometry and spatial structures within an image and suppress detailed texture focusing on the holistic information. It achieved high accuracy in recognizing natural scene categories, *e.g.* mountain and coast. But it often fails to recognize images from indoor environments. **CENTRIST** is a holistic descriptor to capture the the stable spatial structure within images that reflects the functionality of the location, and especially suitable for indoor environment



Cai X. et al. "Heterogeneous Image Feature Integration via Multi-Modal Spectral Clustering", 2011.

Training

- Depicted: Tiling of CT slice → patches
- Patches randomly selected from all cases / volumes
  - random cases
  - random slices
  - random tile from slice
- Batch consists of several slices
  - next up: batch with 15 patches

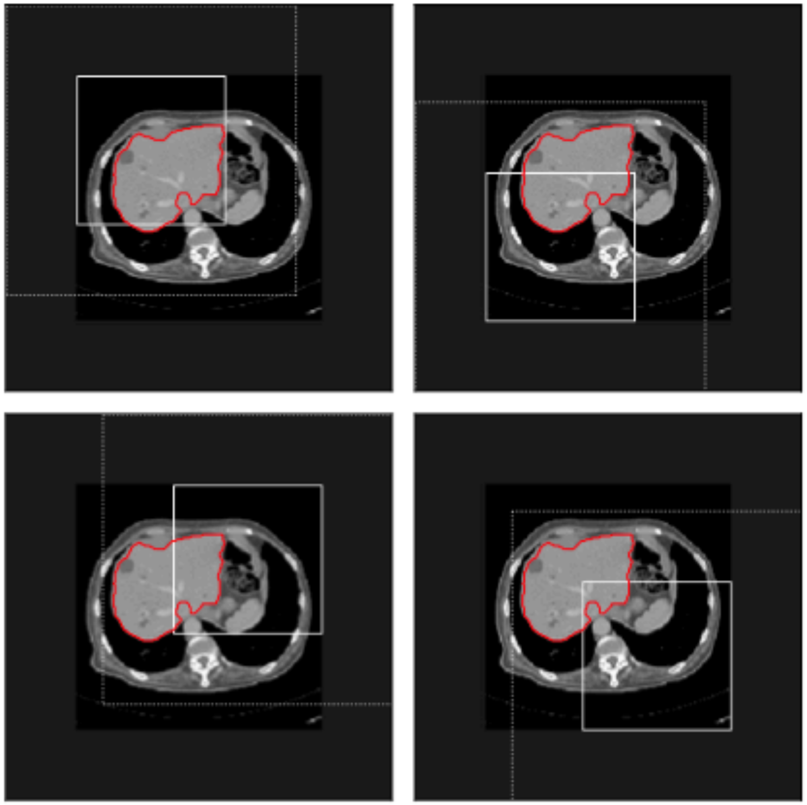
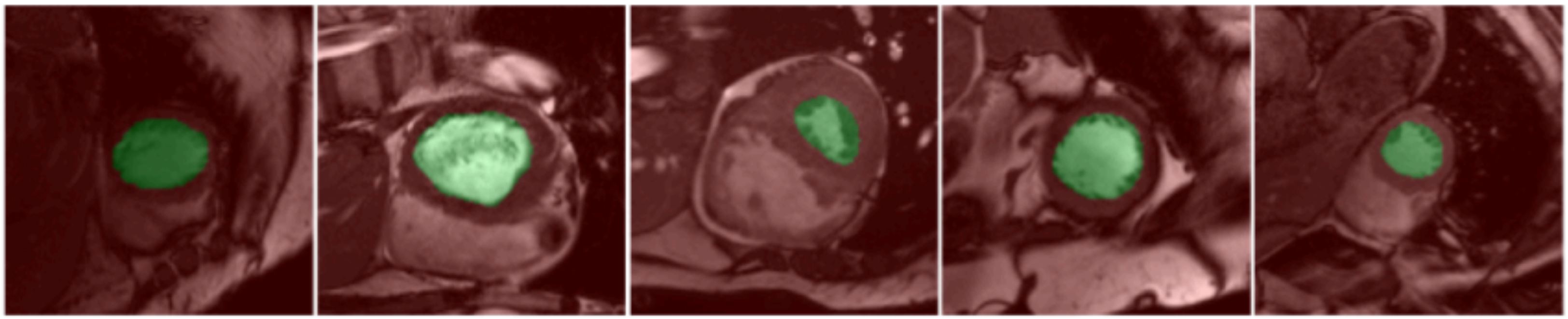
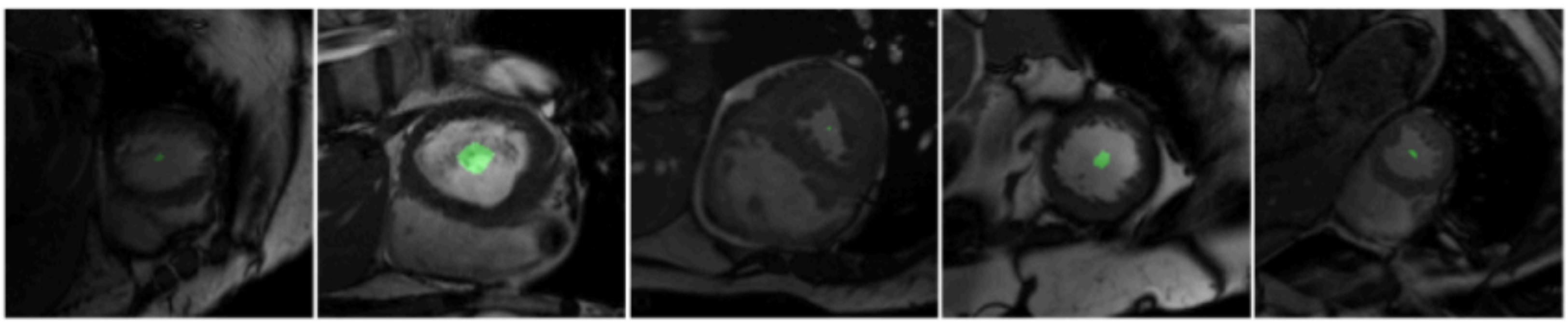


Image courtesy of Dr. Hans Meine

Weakly Supervised Training

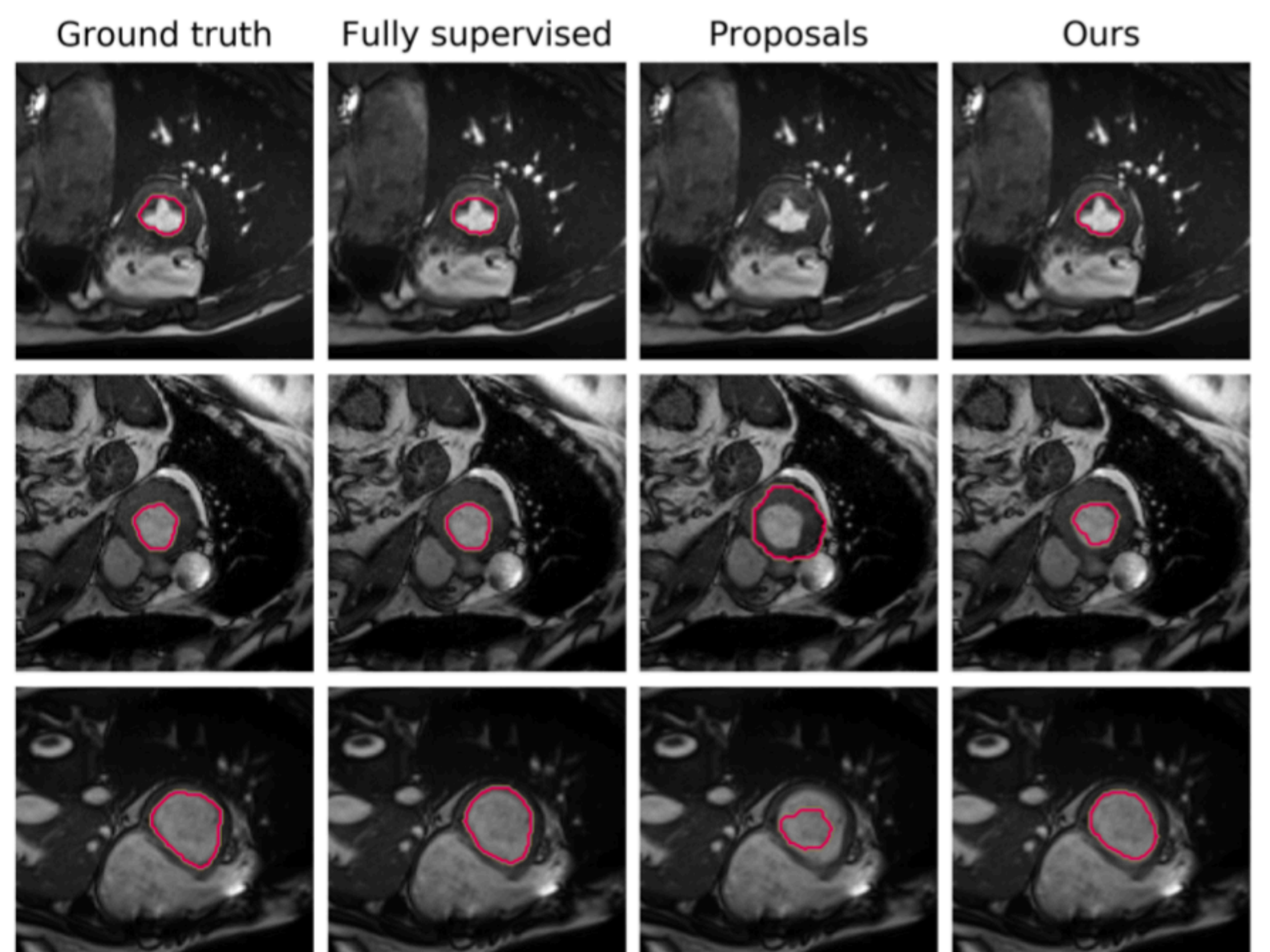


a) Fully supervised labels



b) Weakly supervised labels

Kervadec H. et al., "Constrained-CNN losses for weakly supervised segmentation", 2018.


Ground truth
Fully supervised
Proposals
Ours