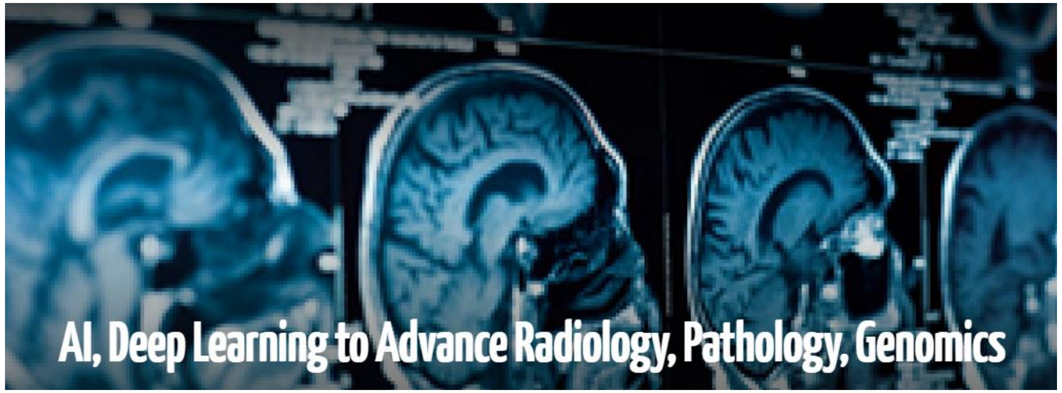
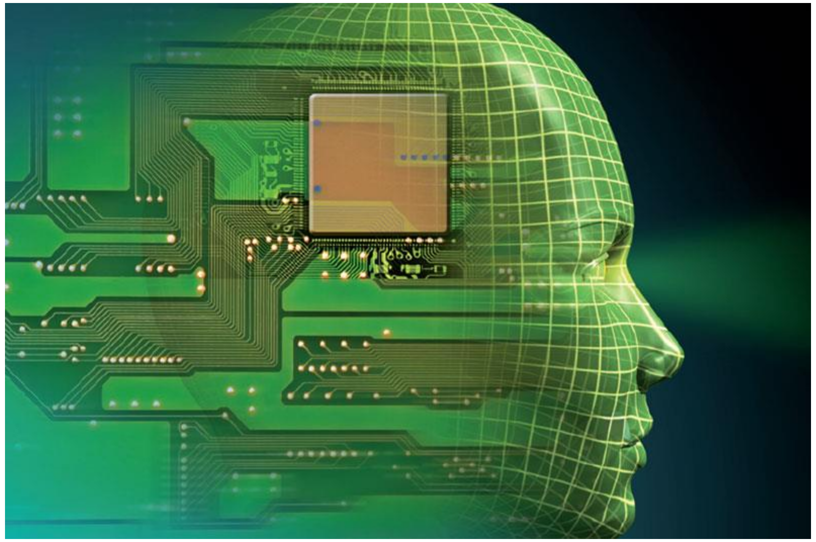
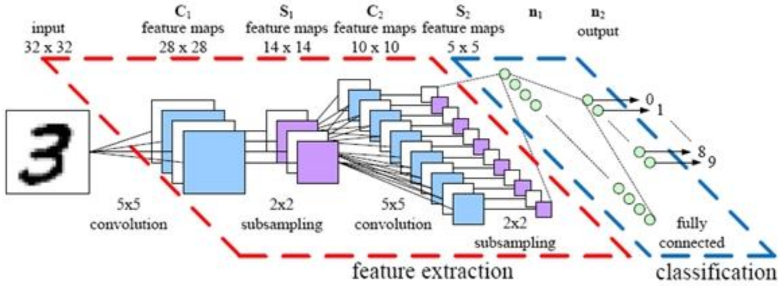
Deep Learning in Medical Image Analysis





Deep learning is a truly transformative technology and the longer-term impact on the radiology market should not be under-estimated. It's more a question of *when*, not *if*, machine learning will be routinely used in imaging diagnosis.

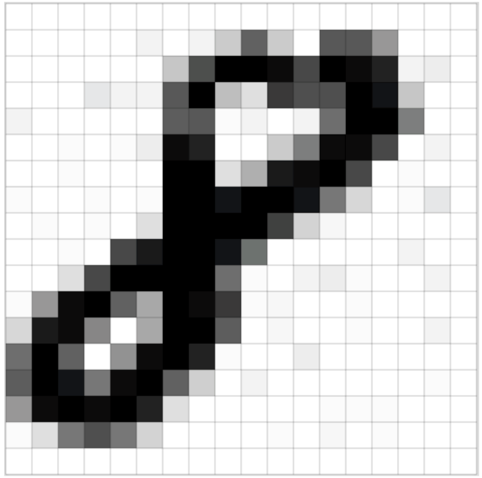
Jin Keun Seo
Computational Science & Engineering, Yonsei U.



Deep Learning

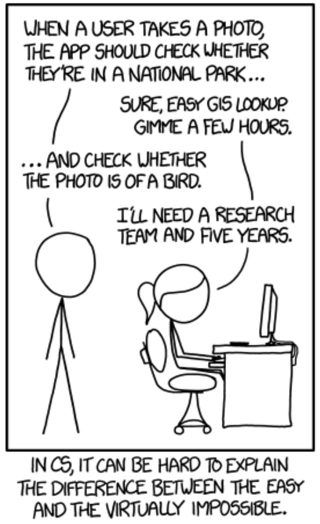
Example of handwritten digit recognition

What human sees



What machine sees

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

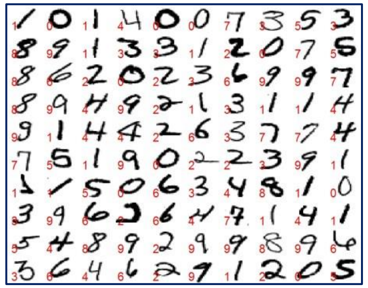


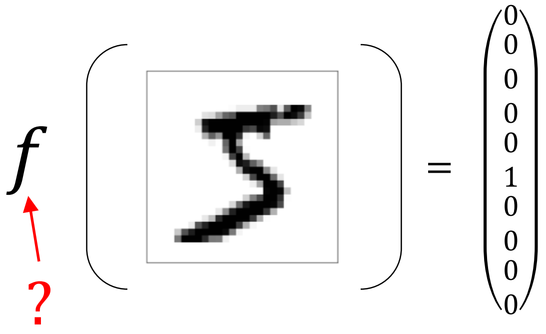
What is Machine Learning?

ML gives computers the ability to learn without being explicitly programmed.

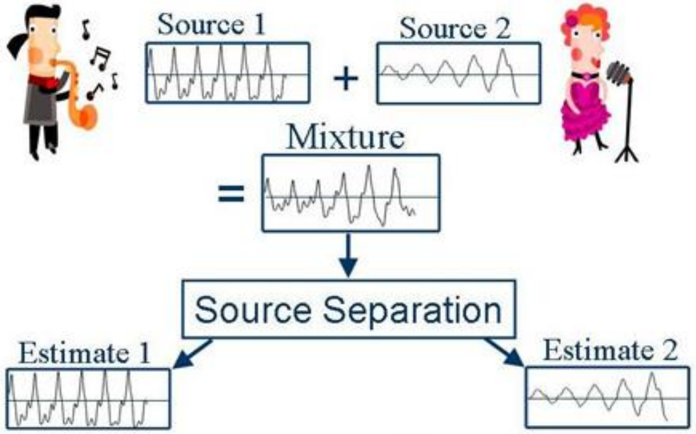
Supervised learning is the machine learning technique of finding a feed-forward function $f(\boldsymbol{x}) = \boldsymbol{y}$ from labeled training data, $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) : i = 1, \dots, m\}$, such that $f\left(\boldsymbol{x}^{(i)}\right) = \boldsymbol{y}^{(i)}$, $i = 1, \dots, m$.

Training data $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) : i = 1, \dots, m\}$



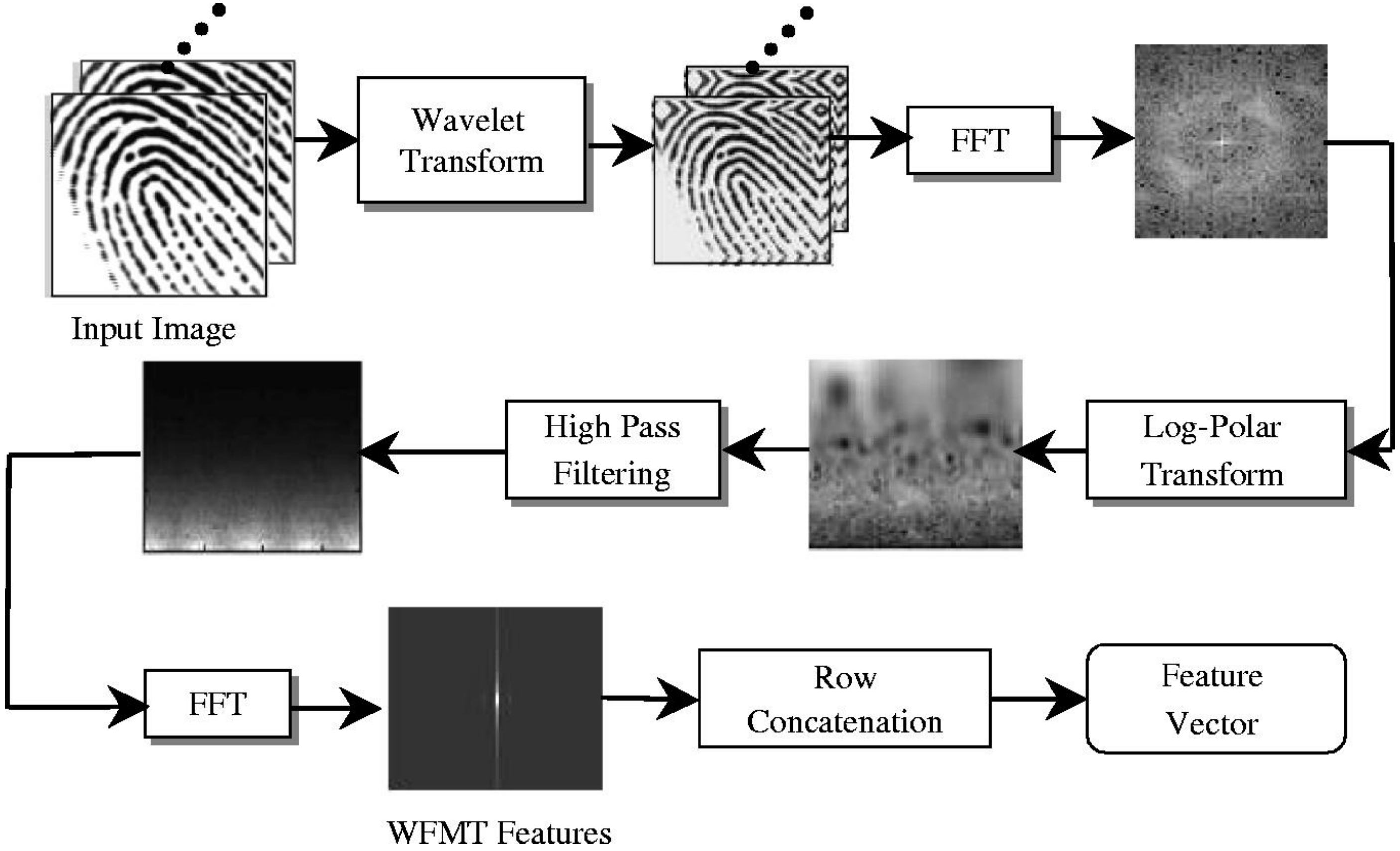


Unsupervised learning infer a function to describe hidden structure from unlabeled data leaving it on its own to find structure in its input.



Reinforcement learning concerned with how software agents ought to take a actions in an environment so as to maximize some notion of cumulative reward

Fingerprint recognition: Easy due to its three basic patterns: arch, loop and whorl.



Scientists have found that family members often share the same general fingerprint patterns, leading to the belief that these patterns are [inherited](#)

Face recognition

Question. Which one is Obama’s cartoon?

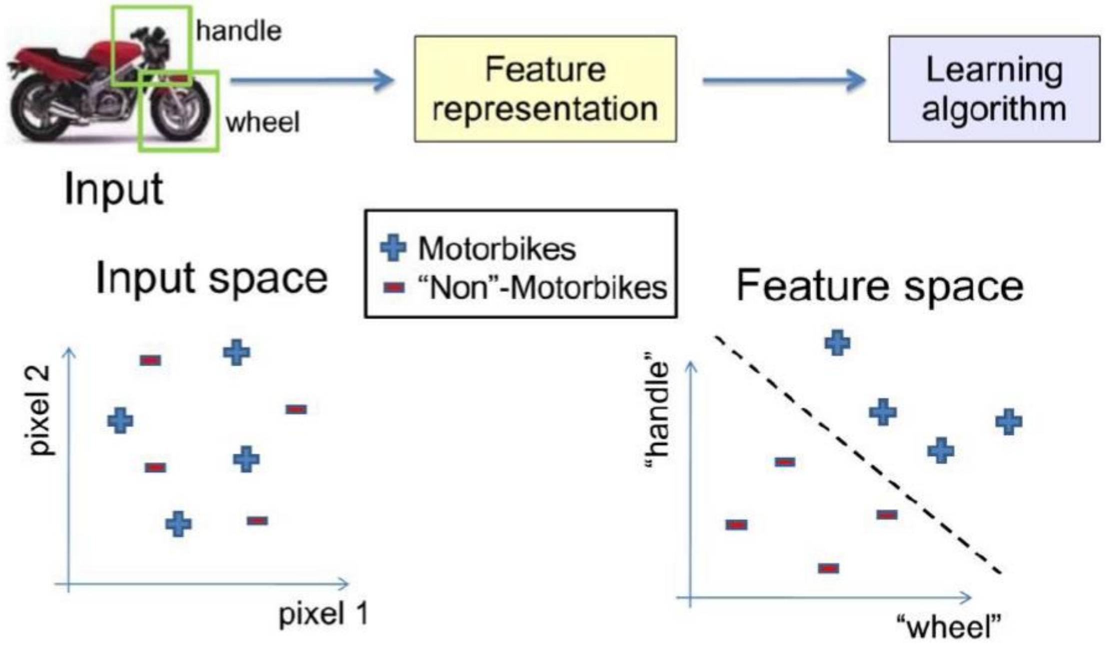
Any pattern?

Invariant theory

Machine learning-Feature Representation

Low level sensing-Preprocessing-Feature extraction-
-Feature selection-Inference, Prediction, Recognition

- **Most critical for accuracy**
- **Account for most of the computation for testing**
- **Most time-consuming in development cycle**
- **Often had-craft in practice**



Feature Learning: instead of designing features, let's design feature learners…

Supervised Learning

Given training data $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) : i = 1, \dots, m\}$, supervised learning is machine learning technique of finding a feed-forward function $f(\boldsymbol{x}) = \boldsymbol{y}$ such that $f\left(\boldsymbol{x}^{(i)}\right) = \boldsymbol{y}^{(i)},\ i = 1, \dots, m.$

Example of handwritten digit recognition

$$f\left(\begin{array}{c}\boldsymbol{x}\text{=image with}\\ n\text{ pixels}\end{array}\right) = \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{pmatrix}$$

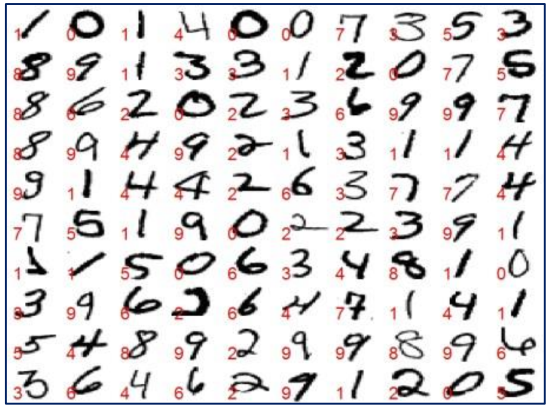
$y_1 = 1$ corresponds to $x = 0.$

$y_5 = 0$ corresponds to $x \neq 4.$

$f\Big($  $\Big) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$f\Big($  $\Big) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

Training data $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) : i = 1, \dots, m\}$



Deep learning: Convolution Neural Network

In deep convolution neural network the feed-forward function f is often given by

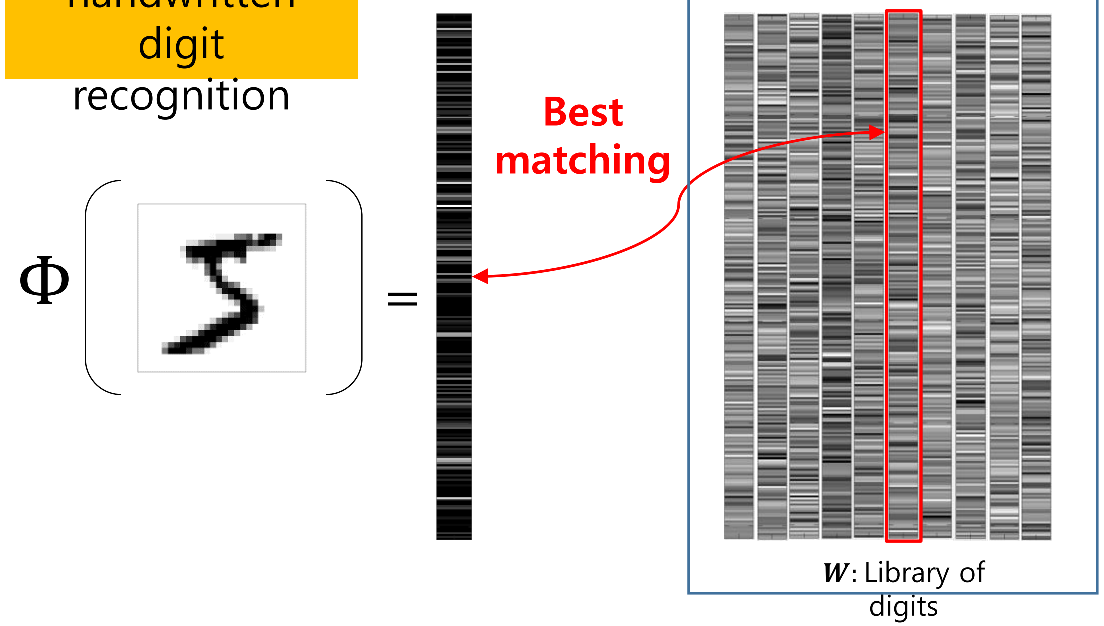
$$f(x) = \mathrm{f}_0(\Phi(x)) \qquad \Phi(x) = g\left(W^k g\left(W^{k-1} \dots \left(g\left(W^2\left(g(W^1 x)\right)\right)\right) \dots\right)\right)$$

g is a vector valued function. Ex) g is ReLU or pooling.

$x = (x, 1)$, $W =$ (weight & bias).

The mission of CNN is to determine $W^j, j = 1, \dots, k$, from the training data $\{(x^{(i)}, y^{(i)}) : i = 1, \dots, m\}$.

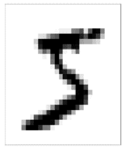
Example of handwritten digit recognition

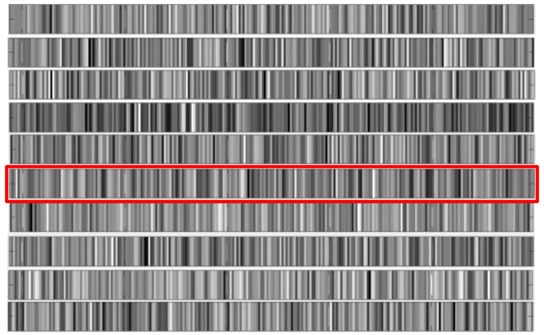


$f_0\Big(\Phi\Big($  $\Big)\Big) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

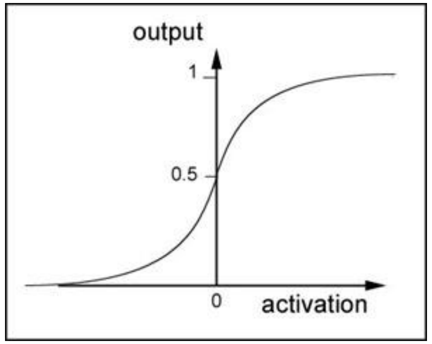
Deep learning: Convolution Neural Network

Example of handwritten digit recognition

f () $=$ f_0 (Φ ()) $=$ f_0 () $=$

$=\sigma$ ( ) $= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$$\sigma(s) = \frac{1}{1+e^{-s}}$$

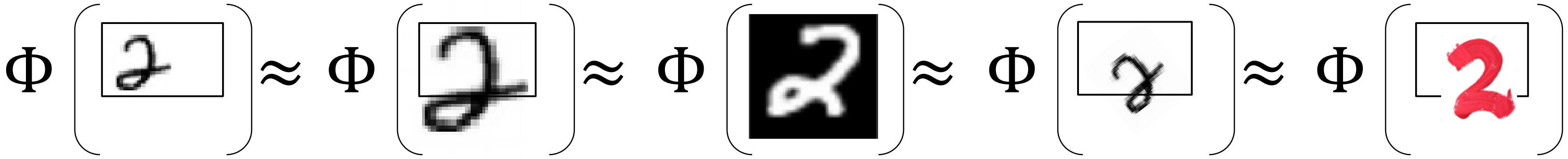


Finding the following Φ(x) is the dream of kernel Learning Algorithms (by S. Mallat)

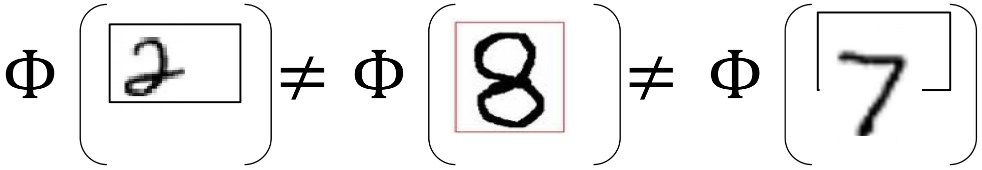
Φ is a contractive operator which reduces the range of variations of x, while still separating different values of f:

$$\Phi(x) \neq \Phi(x') \ \text{if} \ f(x) \neq f(x')$$

Example of handwritten digit recognition



Invariants: translations, diffeomorphisms (scaling, rotations,...)

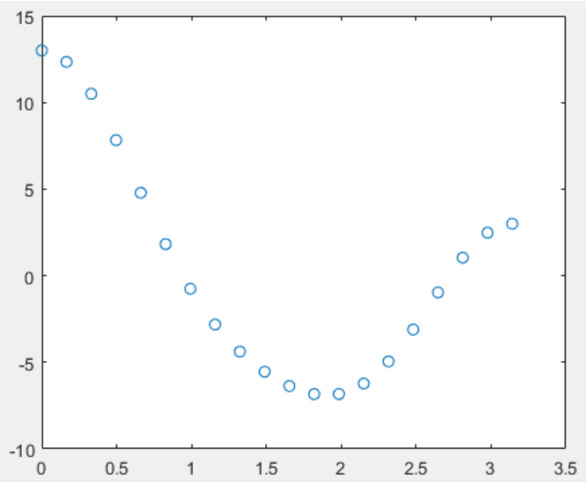


Example of Supervised learning

Given training data $\{(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) : i = 1, \ldots, m\}$, supervised learning is machine learning technique of finding a feed-forward function $f(\boldsymbol{x}) = \boldsymbol{y}$ such that $f(\boldsymbol{x}^{(i)}) = \boldsymbol{y}^{(i)},\ i = 1, \ldots, m$.

Assume $f : \mathbb{R}^3 \to \mathbb{R}^3$ is approximately linear. When training data is $\left\{ \left(\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 7 \\ 34 \\ -8 \end{pmatrix} \right), \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 10.5 \\ -4.9 \end{pmatrix} \right), \left(\begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -0.9 \\ -7 \\ -3 \end{pmatrix} \right), \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 16 \\ 4.1 \end{pmatrix} \right) \right\}$ then the feed-forward function $f(\boldsymbol{x}) = \boldsymbol{y}$ can be approximated as $\boldsymbol{f} \approx \begin{pmatrix} \mathbf{2} & \mathbf{1} & \mathbf{-1} \\ \mathbf{11} & \mathbf{7} & \mathbf{-2} \\ \mathbf{-5} & \mathbf{3} & \mathbf{6} \end{pmatrix}$.

Assume $f : \mathbb{R} \to \mathbb{R}$ is a function. The training data is given by



then the feed-forward function can be approximated as

$$\mathrm{f(x)} = 5\cos(x) + 7\cos(2x) + \cos(4x)$$

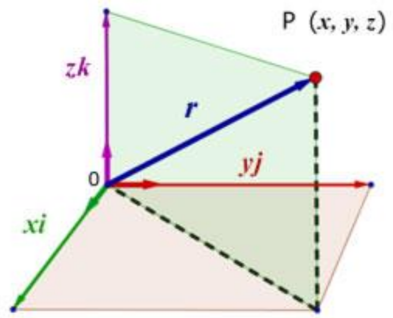
What is the algorithm to obtain f?

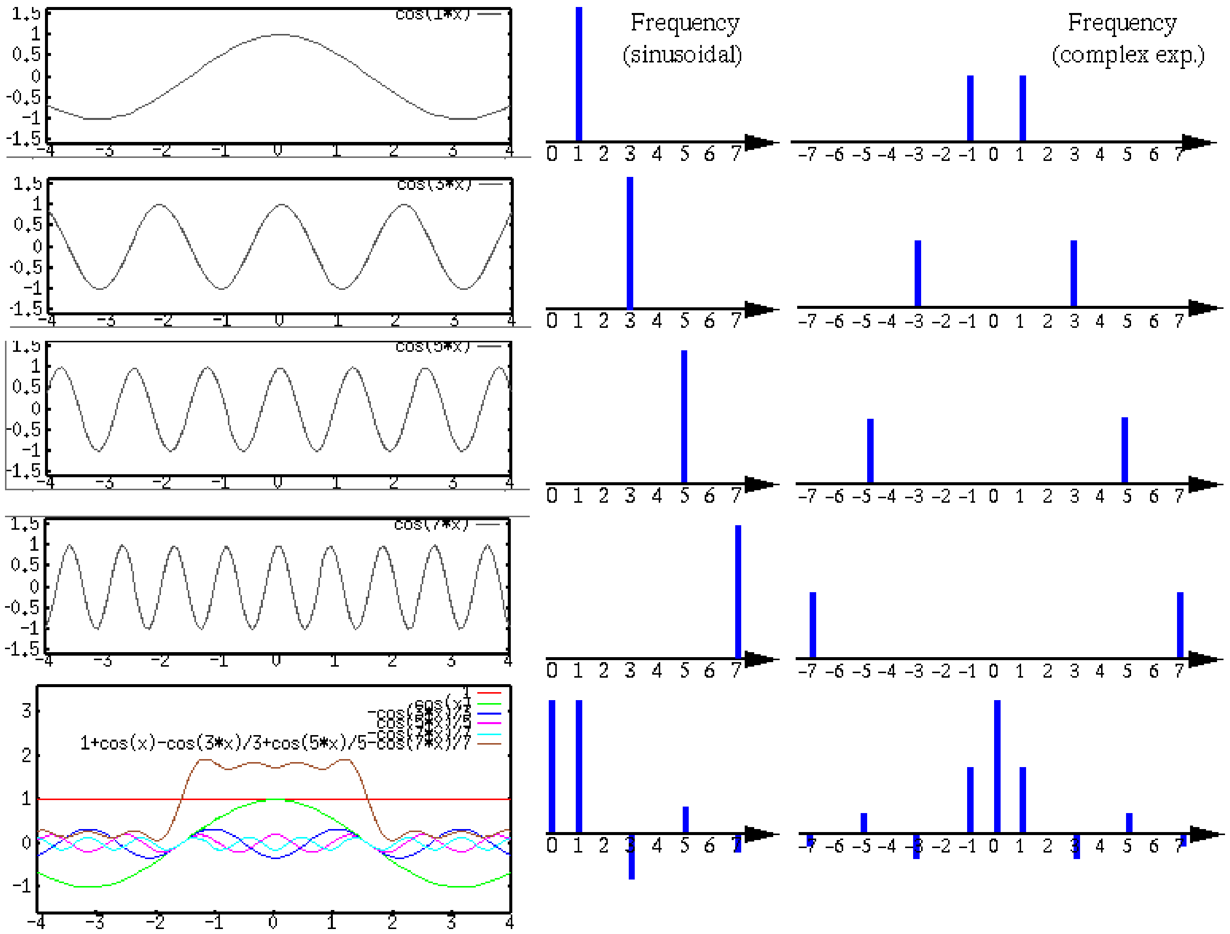
Basis: Fourier Transform

Every function can be expressed as a linear combination of basis functions $f = \sum c_j \phi_j$,

where $\{\phi_0, \phi_1, \cdots\}$ is a set of orthonormal basis $< \phi_n, \phi_m > = \begin{cases} 1 & if\ n = m, \\ 0 & if\ n \neq m. \end{cases}$

Basis





The Fourier transform of f is defined by $(\mathcal{F}f)(\xi) = \int_{-\infty}^{\infty} f(t)\, e^{-2\pi i \xi t}\, dt.$

Each fourier transform acts as a basis to demonstrate the ability to distinguish different signals.

Approximation by 4 principal components (basis) only

$$\{x^{(1)}, \cdots, x^{(32)}\} \subset R^n$$

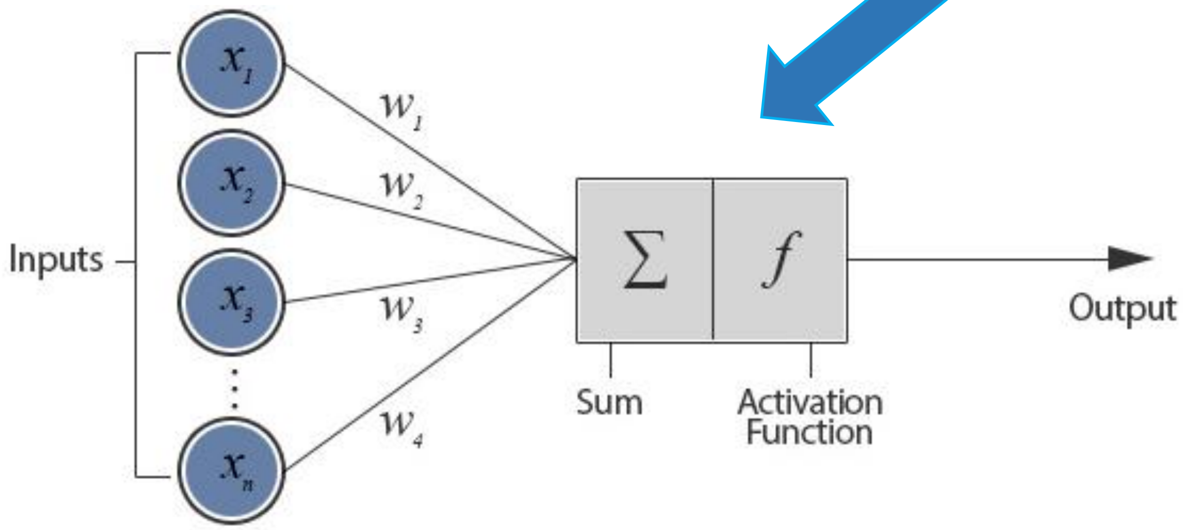
$\mathbf{n = 321 \times 261}$ ***dimensional images***

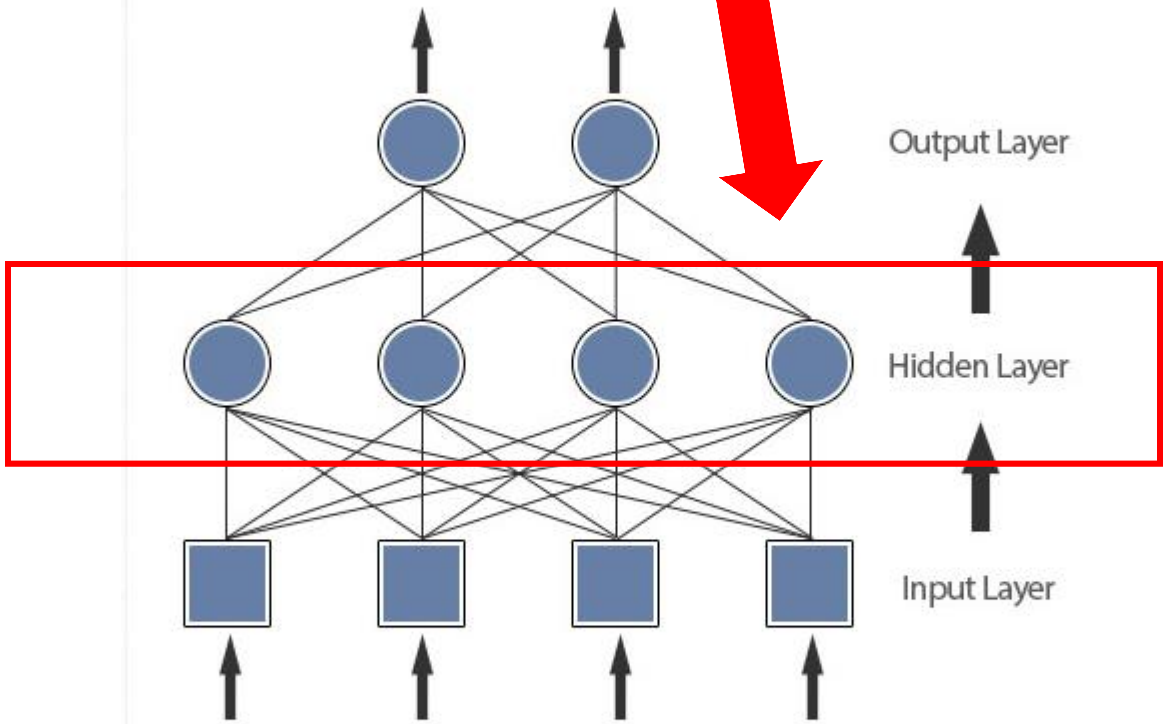
4 dimensional images

Slide Credit: Vaclav

What is Deep Learning?

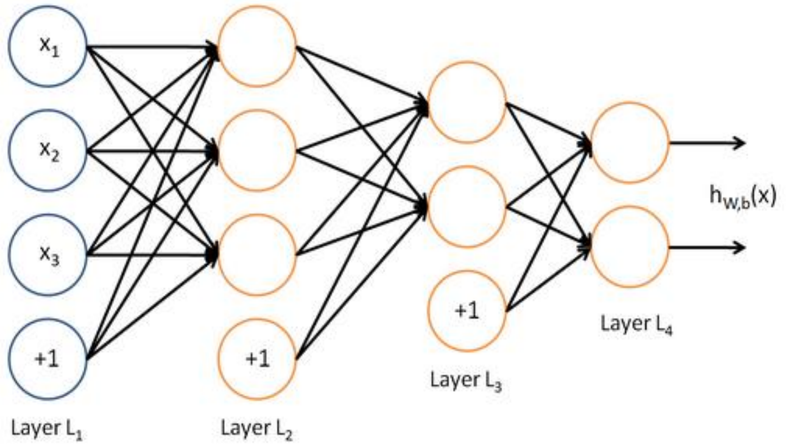
What is Shallow Learning?

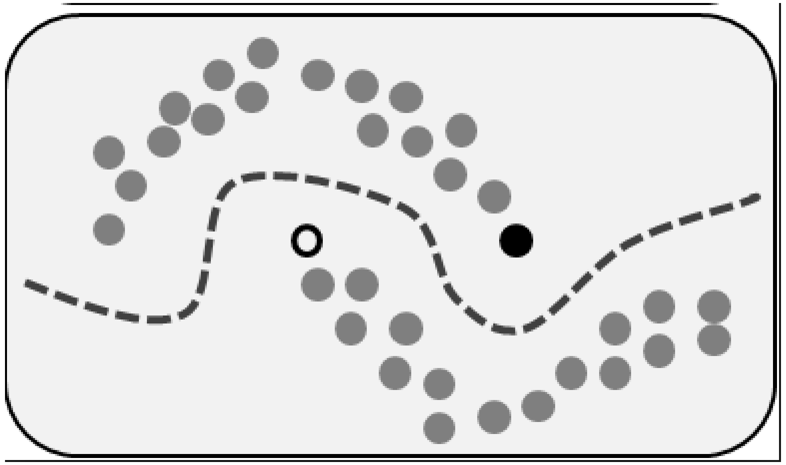




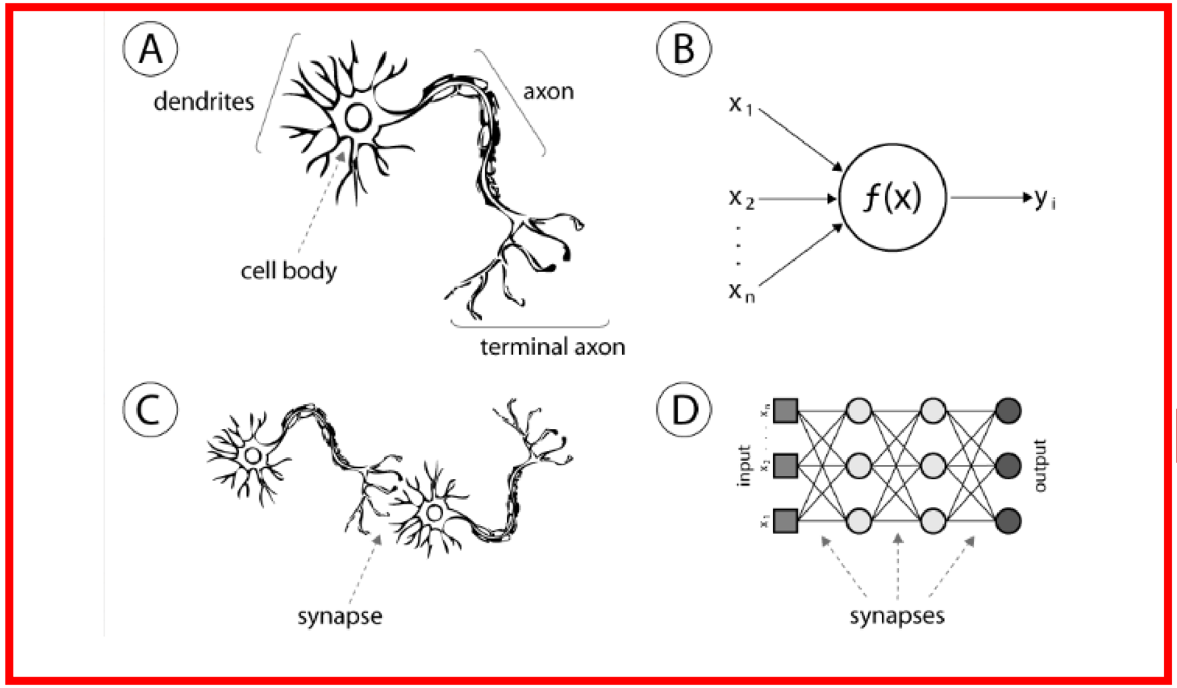
The name Deep Learning can mean different things for different people.

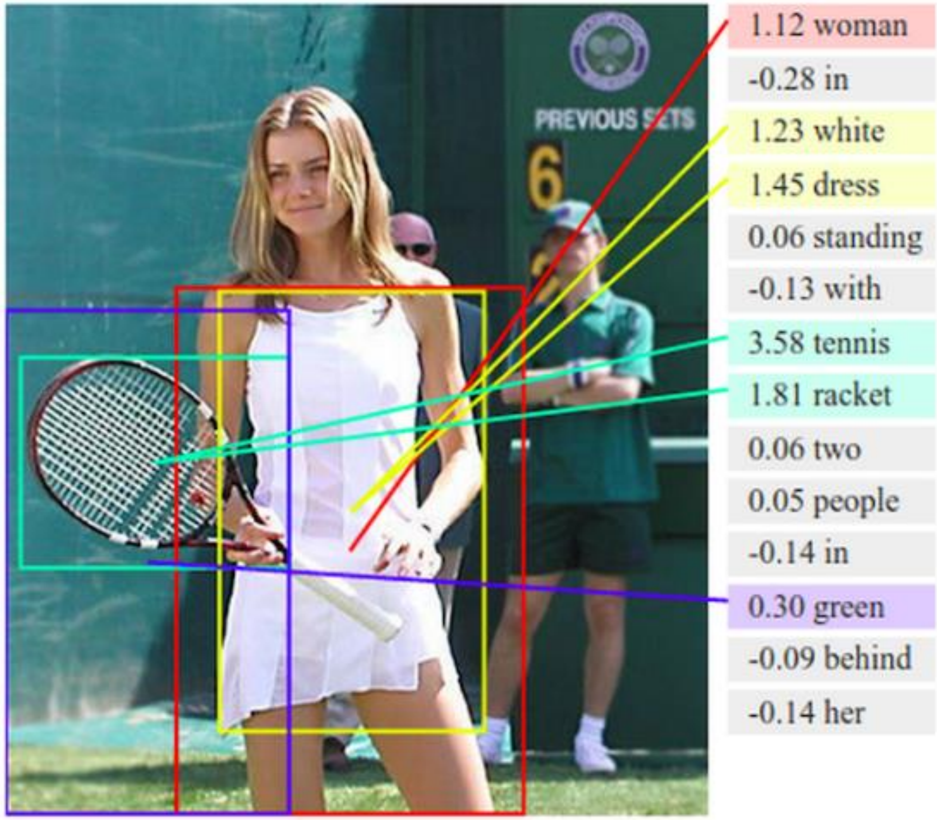
- For many researchers, the word "Deep" in "Deep Learning" means that the neural network has more than 2 layers. This definition reflects the fact that successful neural networks in speech and vision are both deeper than 2 layers.
- For many other researchers, the word "Deep" is also associated with the fact that the model makes use of unlabeled data.
- For many people that I talk to in the industry, the word "Deep" means that there is no need for human-invented features.





What is Neural Network?





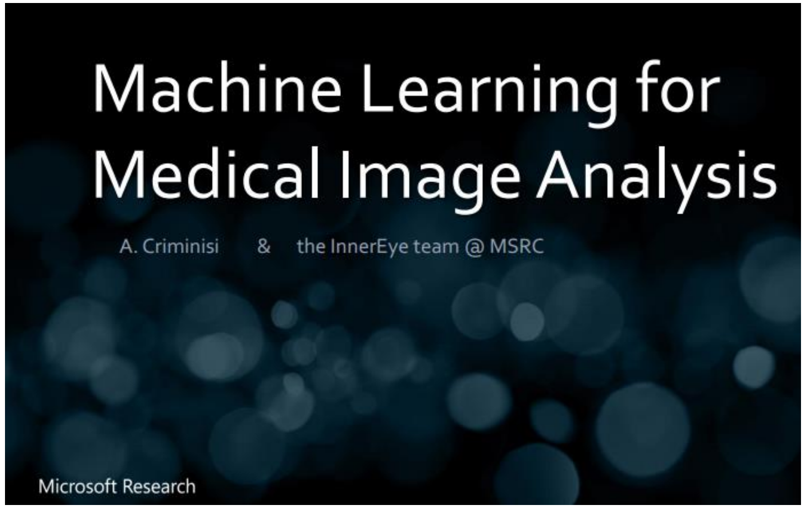
Generating Image Descriptions for unlabeled images

Artificial neural networks are relatively crude electronic networks of "neurons" based on the neural structure of the brain. They process records one at a time, and "learn" by comparing their classification of the record (which, at the outset, is largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm the second time around, and so on for many iterations.

Deep learning for time-saving diagnosis & improving work flow

- Machine learning techniques have received huge attention in recent years due to its remarkable ability of feature learning.
- DL is tackling problems by taking on many of the repetitive and time-consuming tasks performed by radiologists.
- ML can enhance its ability to make the best predictions (or decisions) when faced with new data.

To reduce the economic burden of health care, ...



By Jason Brownlee on July 14, 2016 in Deep Learning

8 Inspirational Applications of
Deep Learning

Beat the Math/Theory Doldrums and Start using Deep Learning in your own projects Today, without getting lost in "documentation hell".

1. Automatic Colorization of Black and White Images

Reduce many of the repetitive and time-consuming tasks

Colorization of Black and White Photographs
Image taken from Richard Zhang, Phillip Isola and Alexei A. Efros.

Reduce many of the repetitive and time-consuming tasks

2. Automatically Adding Sounds To Silent Movies

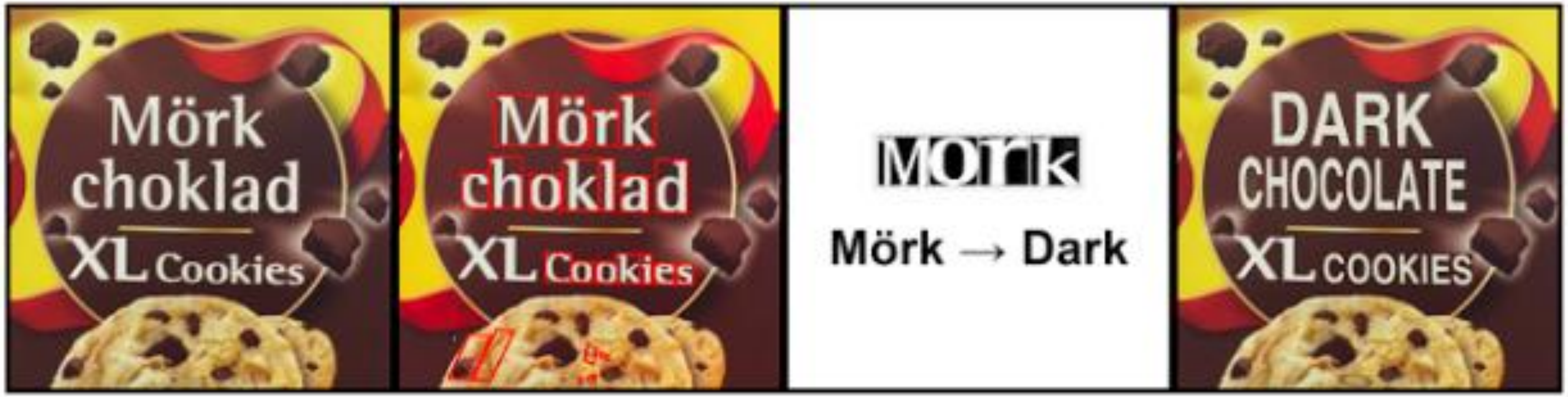
In this task the system must synthesize sounds to match a silent video.

Reduce many of the repetitive and time-consuming tasks

3. Automatic Machine Translation

This is a task where given words, phrase or sentence in one language, automatically translate it into another language.

Automatic machine translation has been around for a long time, but deep learning is achieving top results in two specific areas: Automatic Translation of Text. Automatic Translation of Images.



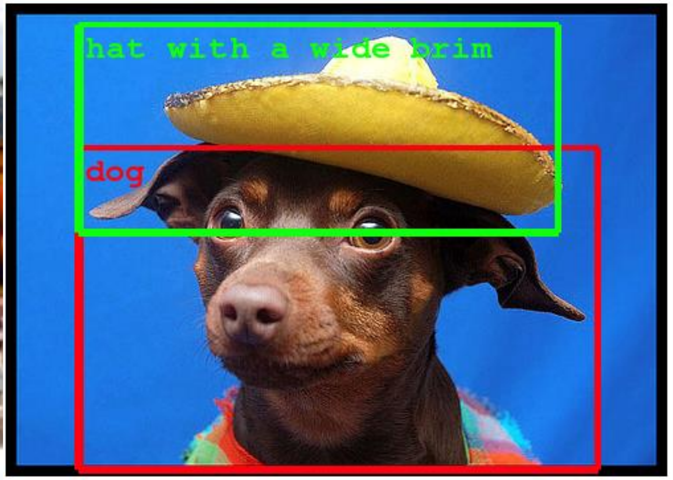
Instant Visual Translation

Example of instant visual translation, taken from the Google Blog.

Reduce many of the repetitive and time-consuming tasks

4. Object Classification and Detection in Photographs

This task requires the classification of objects within a photograph as one of a set of previously known objects.

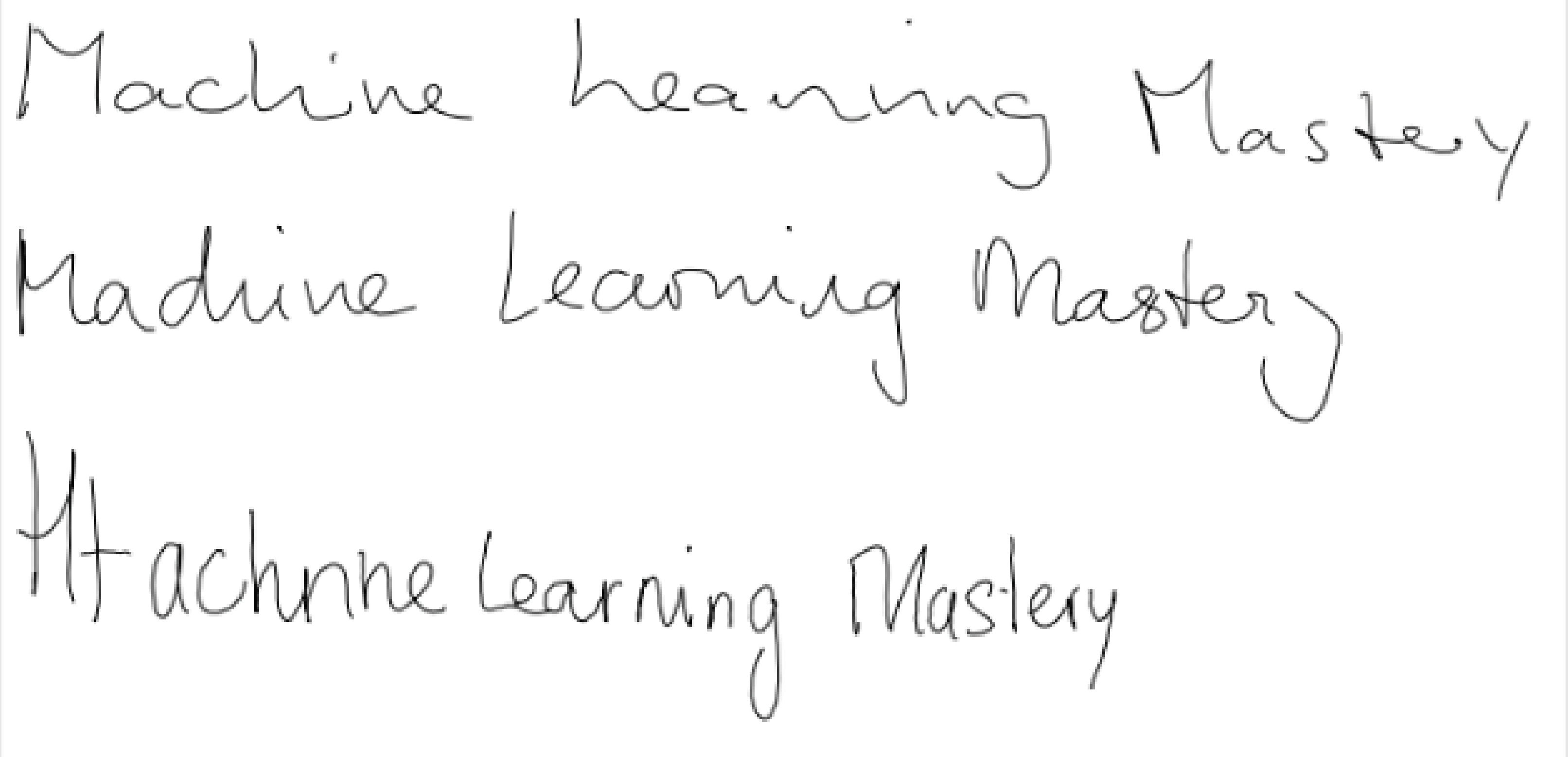


Example of Object Detection within Photographs
Taken from the Google Blog.

Example of Object Classification
Taken from ImageNet Classification with Deep Convolutional Neural Networks

Reduce many of the repetitive and time-consuming tasks

5. Automatic Handwriting Generation

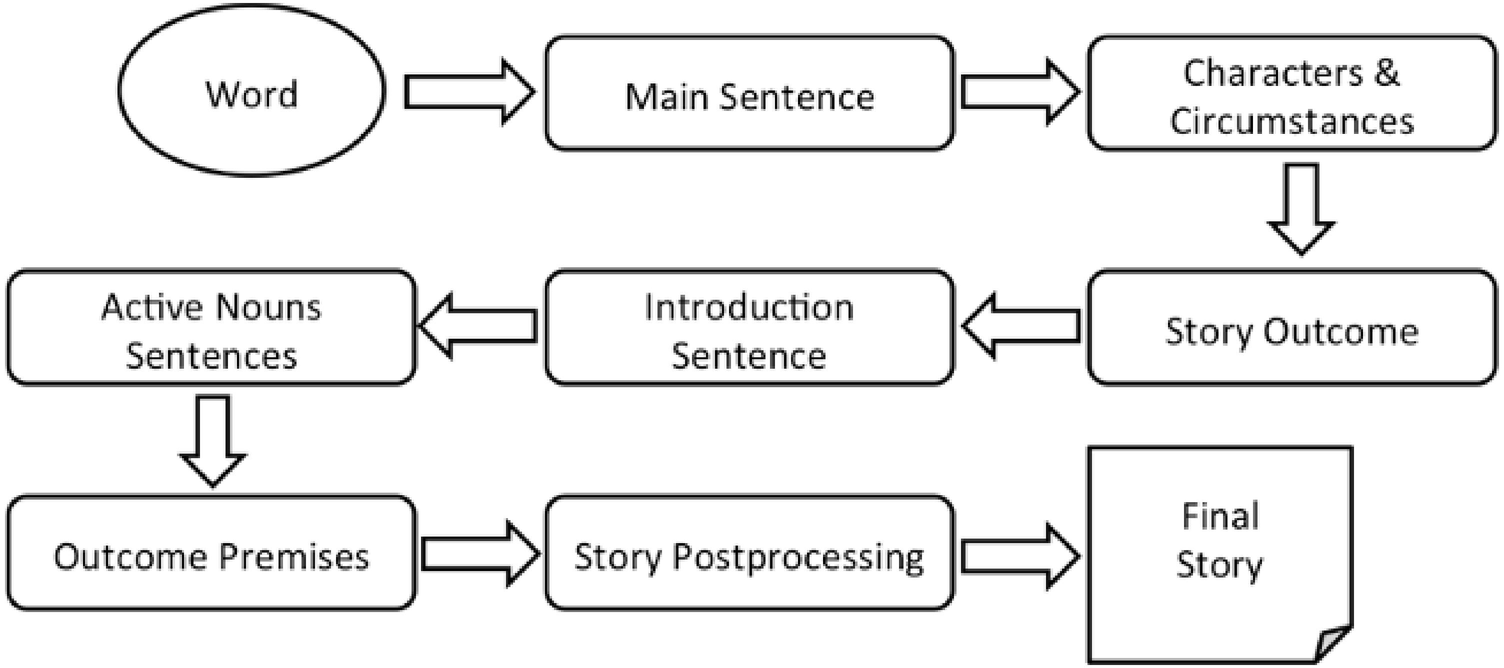


Sample of Automatic Handwriting Generation

Reduce many of the repetitive and time-consuming tasks

6. Automatic Text Generation

This is an interesting task, where a corpus of text is learned and from this model new text is generated, word-by-word or character-by-character.



Reduce many of the repetitive and time-consuming tasks

7. Automatic Image Caption Generation

Automatic image captioning is the task where given an image the system must generate a caption that describes the contents of the image.

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

정확한 계산에 의한 판세분석

8. Automatic Game Playing

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 529 > Issue 7587 > Articles > Article

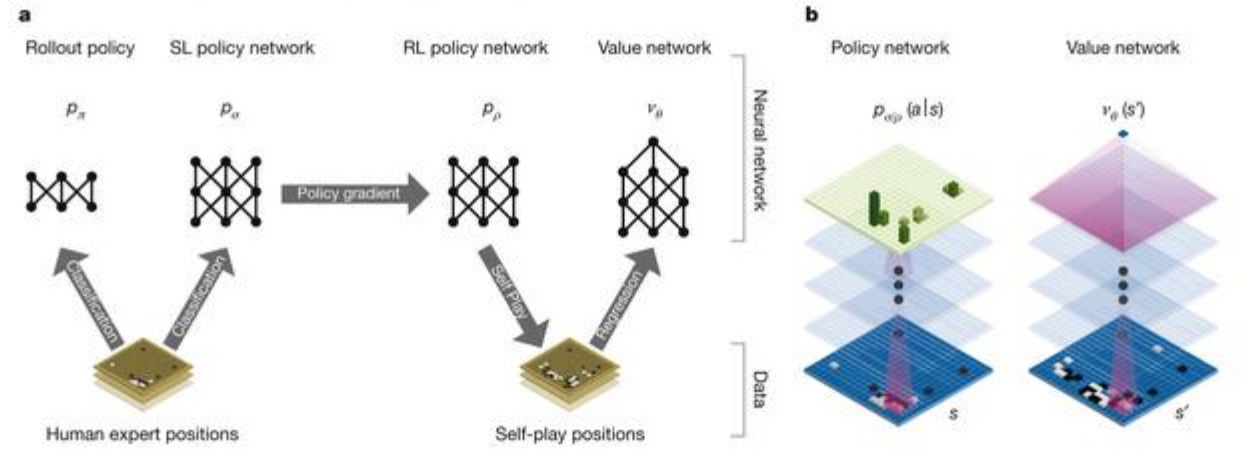
NATURE | ARTICLE

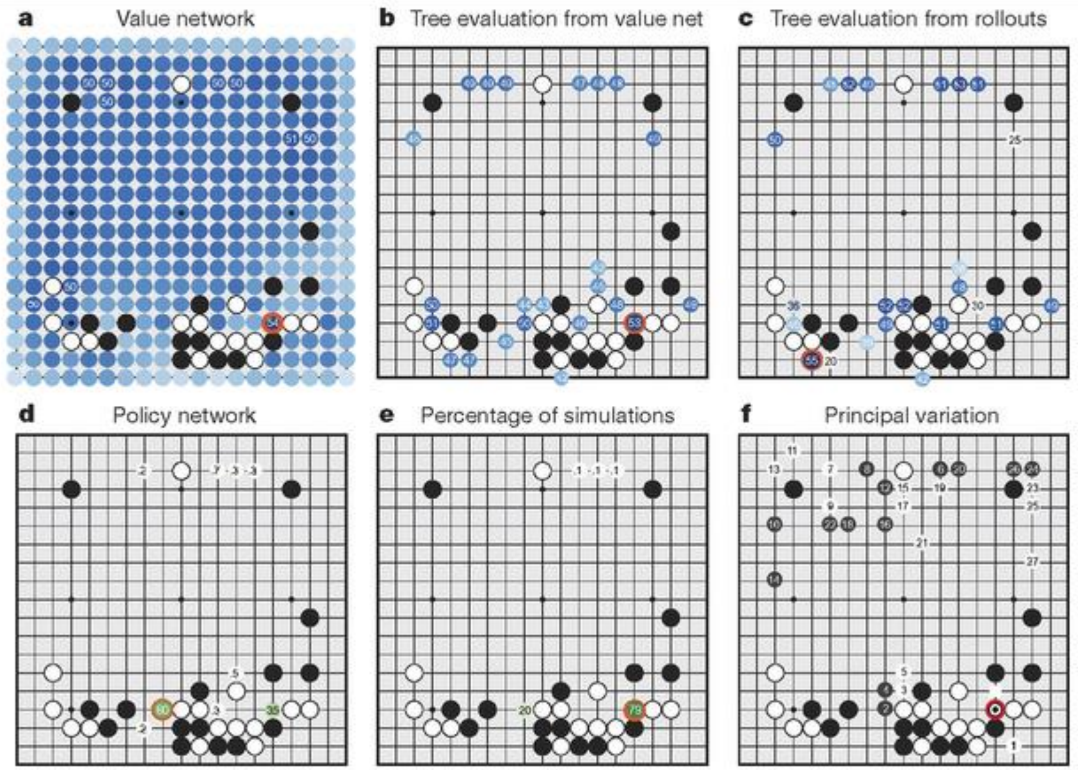
日本語要約

High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects

Benjamin P. Kleinstiver, Vikram Pattanayak, Michelle S. Prew, Shengdar Q. Tsai, Nhu T. Nguyen, Zongli Zheng & J. Keith Joung

Figure 1: Neural network training pipeline and architecture.





Deep Learning
for medical imaging analysis

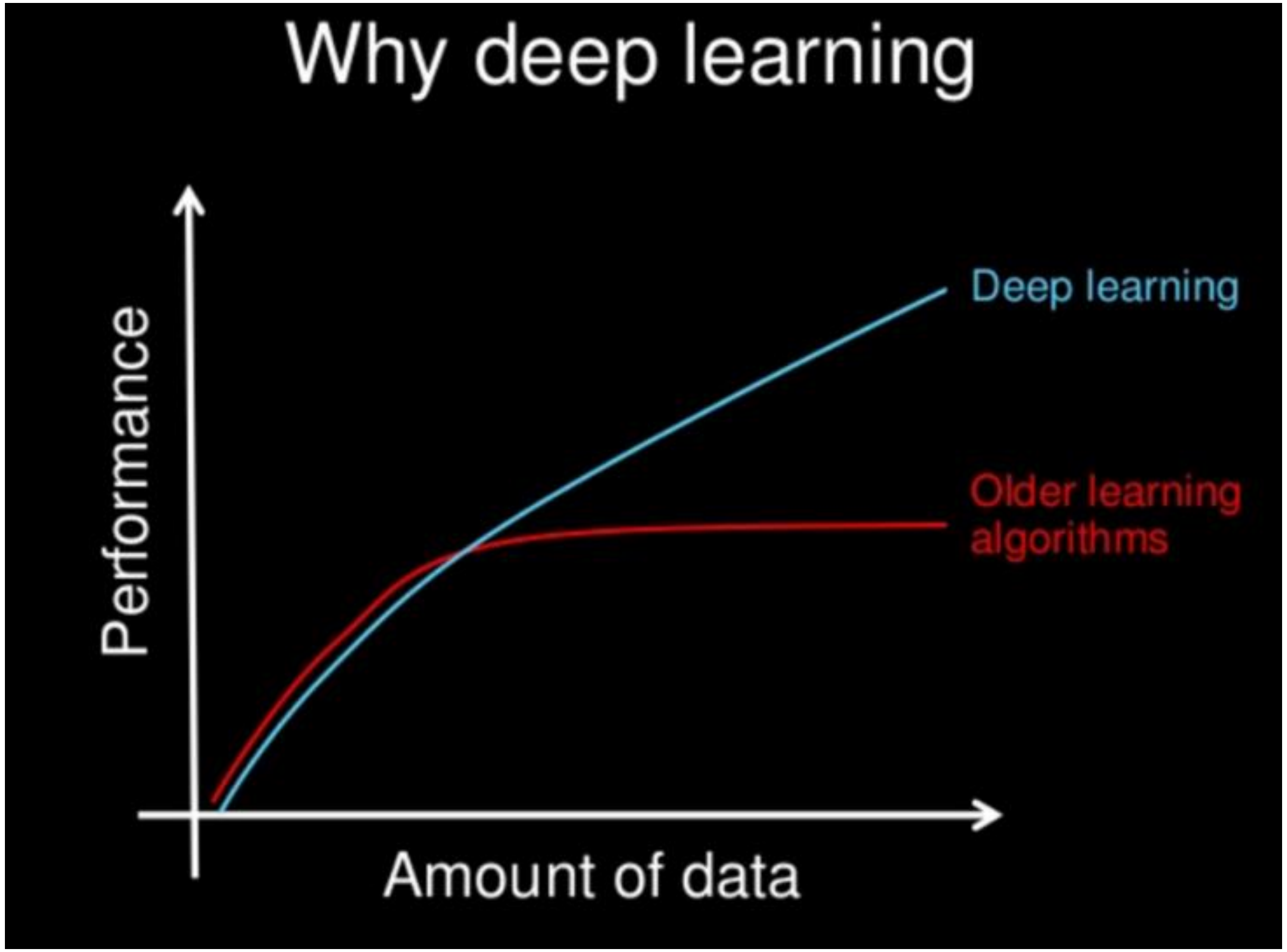
The accumulation of X-rays, CT scans and MRIs means that doctors face an enormous task of sifting through this medical data in order to reach diagnoses. But now, **thanks to advances in machine learning, this task is becoming much easier.**

Computing

Why IBM Just Bought Billions of Medical Images for Watson to Look At

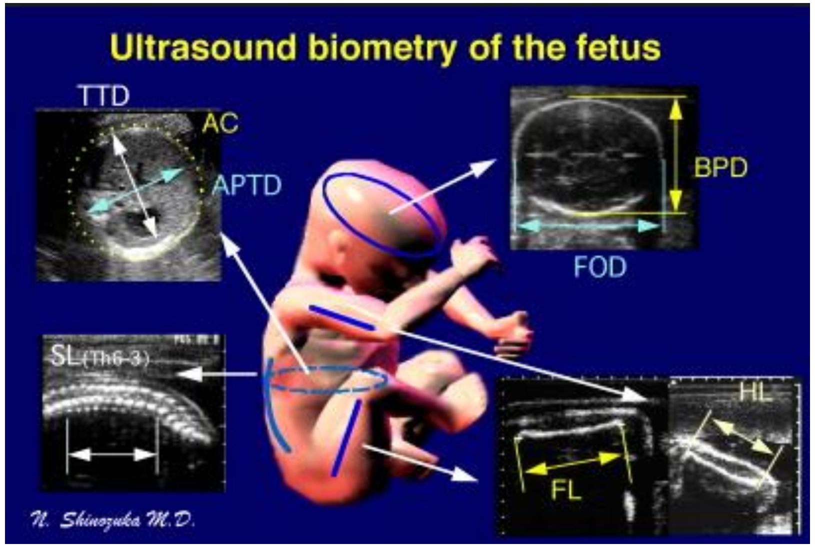
IBM seeks to transform image-based diagnostics by combining its cognitive computing technology with a massive collection of medical images.

by Mike Orcutt   August 11, 2015



Automated Fetal Biometry

to estimate fetal weight & fetal growth abnomalities



Goal: Improve clinical workflow

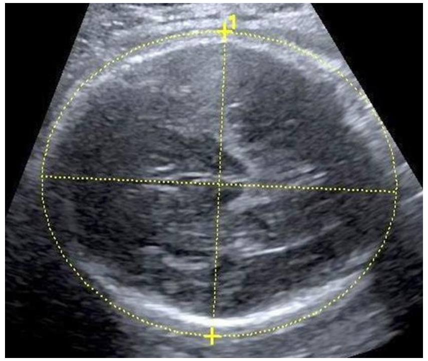
ergonomic stress 감소/업무process 간소화/ 문서작업 최소화

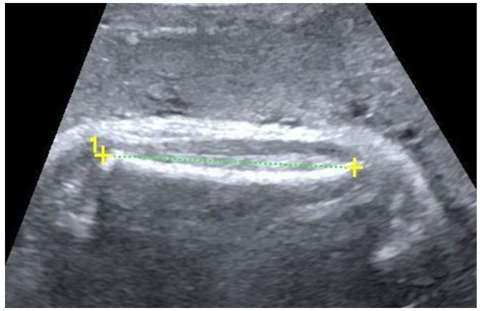
- Head (BPD, OFD, HC, Cephalic index)
- Thorax (Axis, Heart circumference Area, Thoracic circumference area, Head/Thorax ratio)
- Abdomen Circumference (AC),
- Bones: Femur (Femur length), Other Bones(optional)

Challenge (AC)

현존하는 방식 : 정확도 매우 낮음

Sub-challenges (HC, BPD, Femur length)





Deep learning Methodologies in Computed Tomography

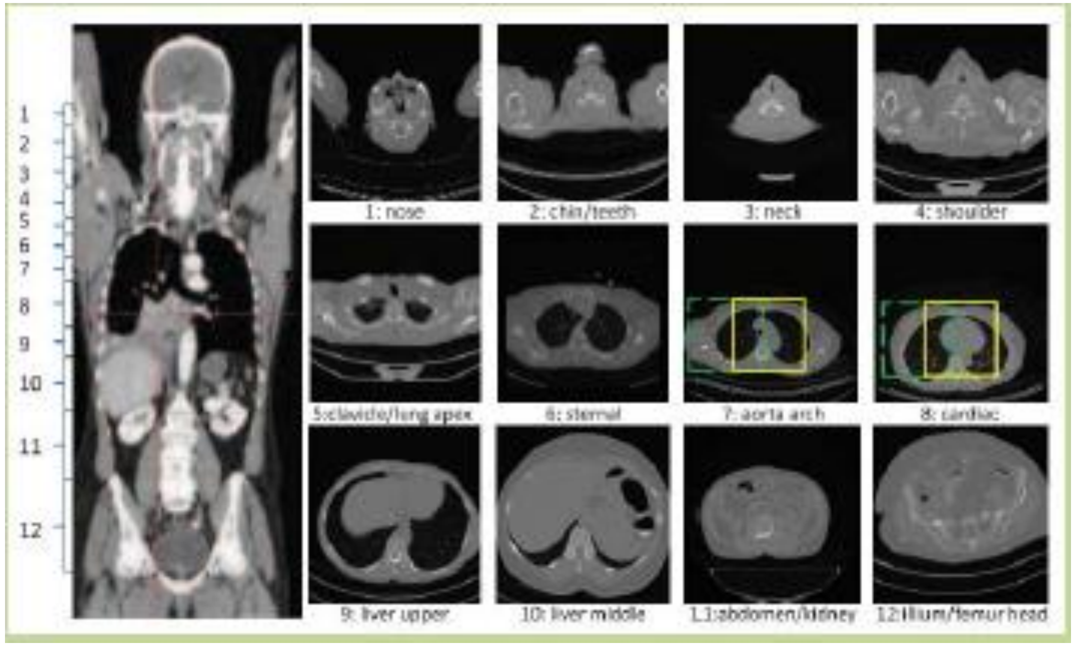
Deep learning has been applied to a range of problems in Computed Tomography including:

- **Anatomy recognition**
- **Organ segmentation;** Pancreas Segmentation, Unary Bladder segmentation
- **Image Registration (X-ray)**
- **Lung texture classification and airway detection**
- **Computed Aided Diagnosis**; Lymph node detection, Lung nodule detection and classification, Liver Lesion segmentation

Slide from **Sarfaraz Hussein** et al.

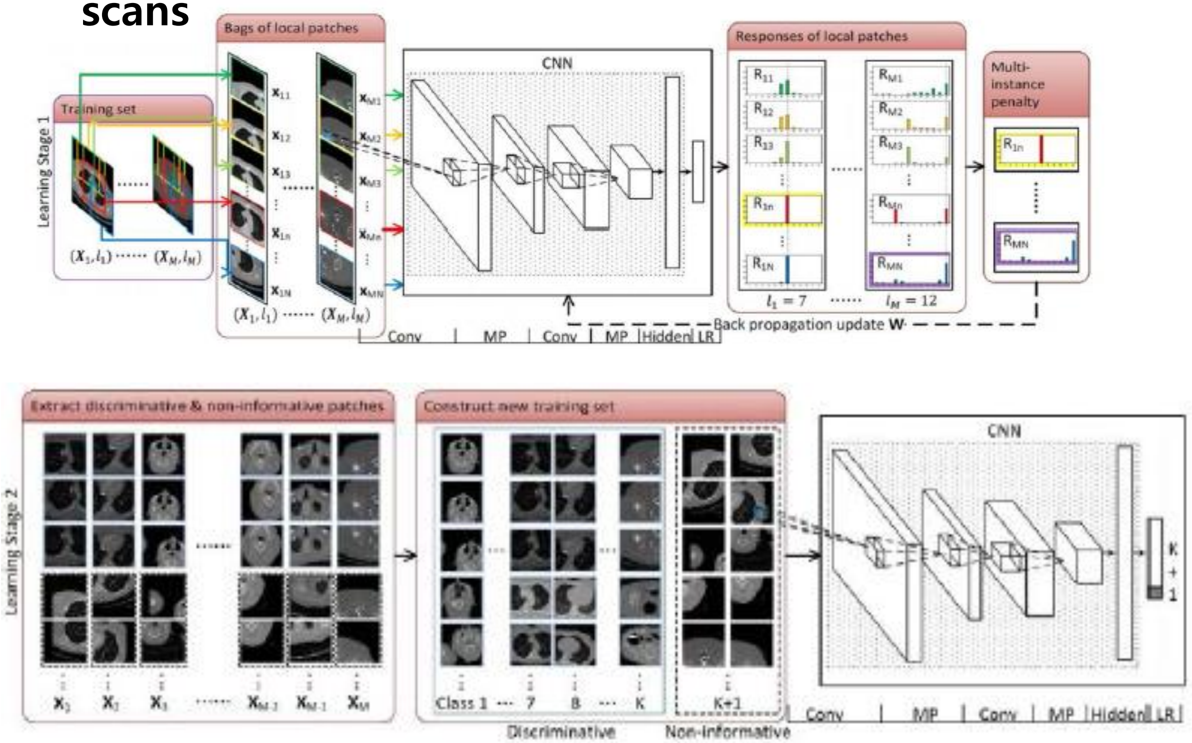
Anatomy Recognition

Objective: This paper aims at: 1) discover the local regions that are discriminative and non-informative to the image classification problem, and 2) learn a image-level classifier based on these local regions.



Method:

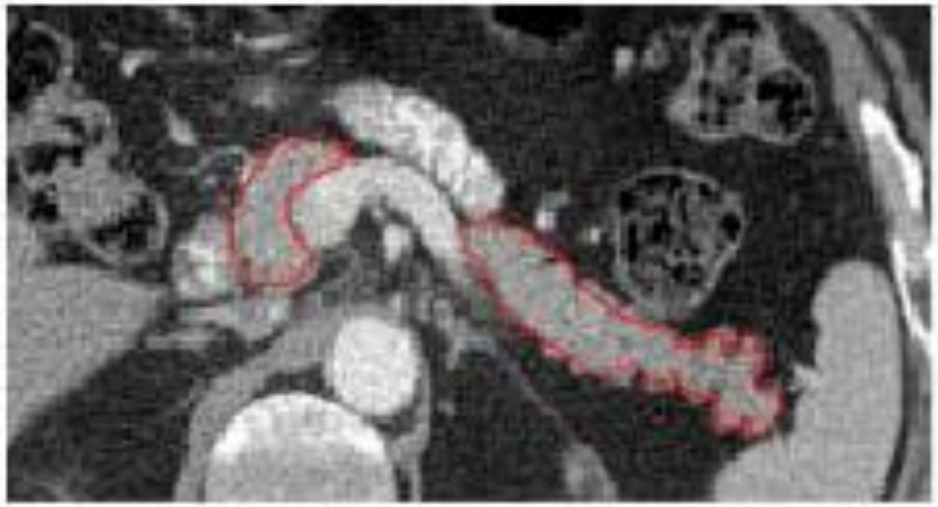
- **Training CNN in 2 stages:**
  - Multiple Instance learning based pre-training
  - CNN Boosting for improved recognition
- **Multiple instance learning is used to automatically learn the discriminative local patches from the CT scans**



Yan, Zhennan, et al. "Multi-Instance Deep Learning: Discover Discriminative Local Anatomies for Bodypart Recognition." *IEEE transactions on medical imaging* 35.5 (2016)

Pancreas Segmentation

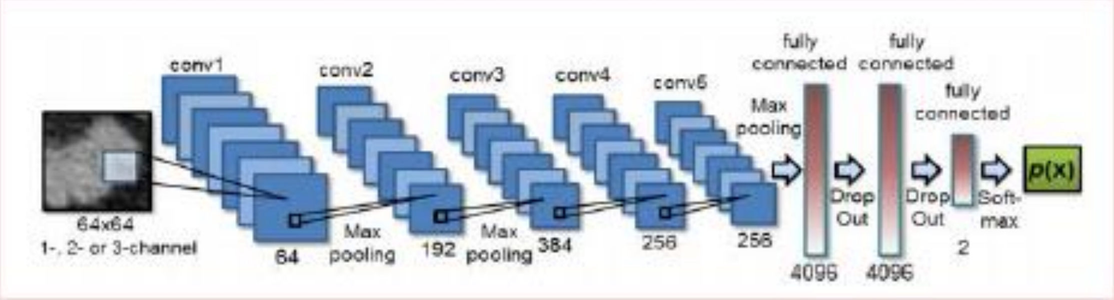
Objective: The pancreas is an abdominal organ with very high anatomical variability. This inhibits previous segmentation methods from achieving high accuracies, especially compared to other organs such as the liver, heart or kidneys. This paper presents a probabilistic bottom-up approach for pancreas segmentation in abdominal computed tomography (CT) scans, using multi-level deep convolutional networks



Roth, Holger R., et al. "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2015.

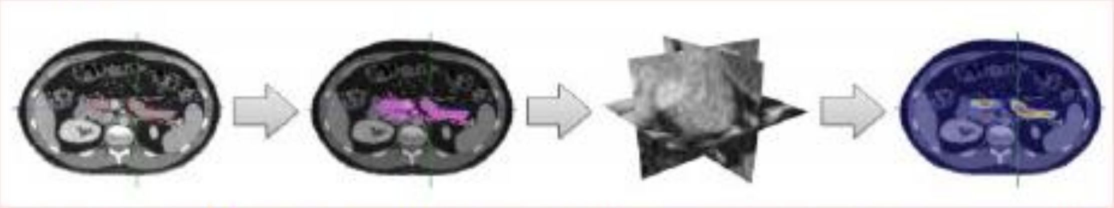
2. Method:

- Super-pixel segmentation on the CT image followed by Random Forest (RF) based classification to get initial pancreas segmentation.



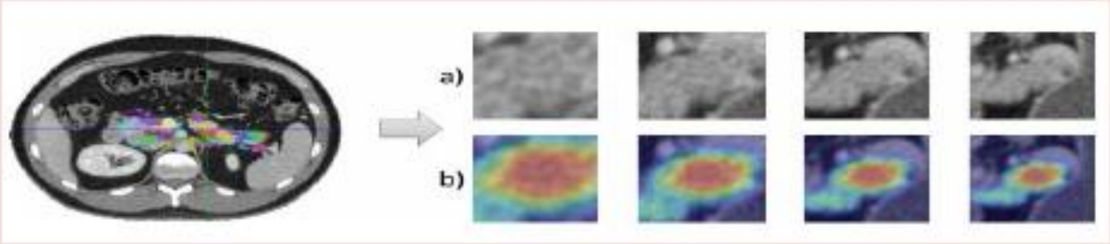
CNN architecture

- Classification of patches generated through sliding window using CNN (P-CNN)



(Left to Right) Red contour shows gold standard of segmentation, pink regions are classification results using RF, finally the probability map using P-CNN

- Region based classification (R-CNN) at different scales.

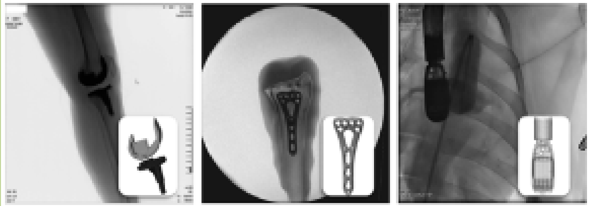


(a) Region based CNN on different scales, (b) Additional channel of input from P-CNN

2D/3D Registration

1. Objective:

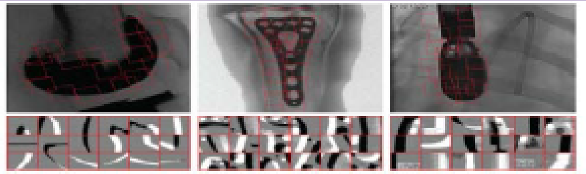
To perform realtime 2D/3D registration using CNN based regression



2D X-Ray image and a 3D model of the target object

3. Results

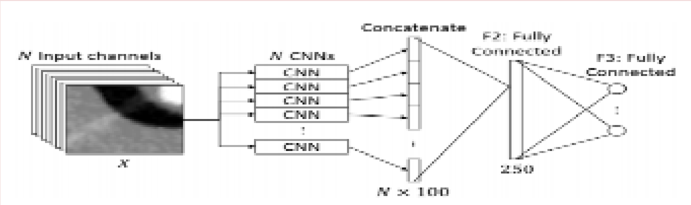
- Evaluation on knee-prosthesis, virtual implant system and X-ray echo fusion datasets
- Evaluation metric: Mean target registration error in the projection direction
- Significant improvement in performance as well as time.



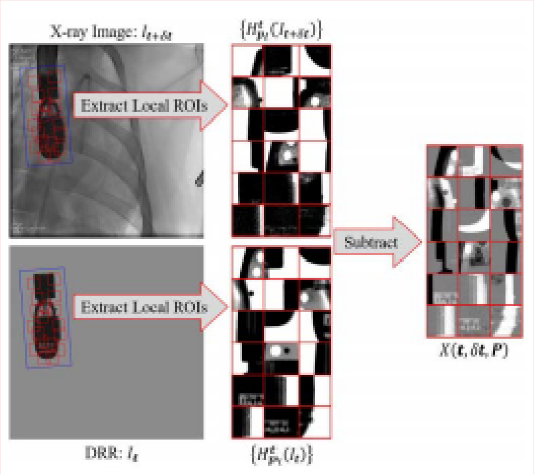
Examples of Region of Interest and Local Image Residuals from the 3 datasets

2. Method:

- The goal is to train a CNN regressor to map from 2D/3D image to their transformation parameter difference
- Local image residual features are extracted representing difference between rendered image and the X-ray image in local patches.
- Regression problem is simplified by partitioning the parameter space into 3 groups based on their difficulty
- The CNN architecture comprises of 2 convolutional, 2 maxpooling and 1 fully connected layers.



CNN Regression Model



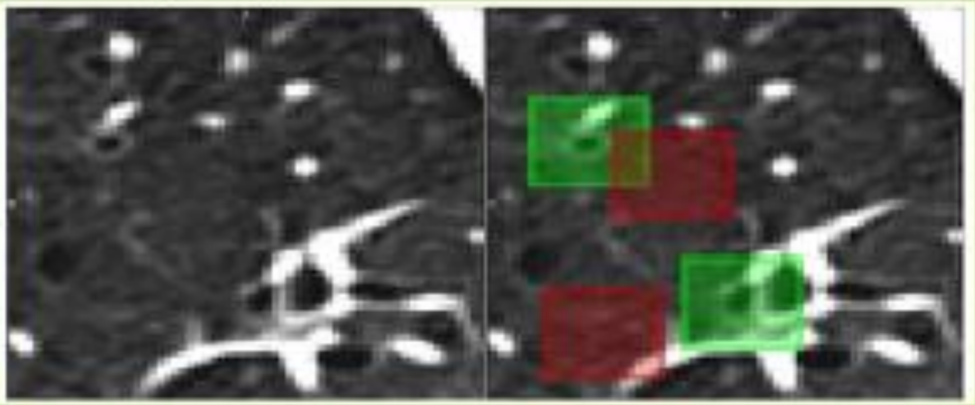
Local Image Residuals

Miao, Shun, Z. Jane Wang, and Rui Liao. "A CNN Regression Approach for Real-Time 2D/3D Registration." *IEEE transactions on medical imaging* 35.5 (2016)

Lung Texture Classification/Airway Detection

1. Objective:

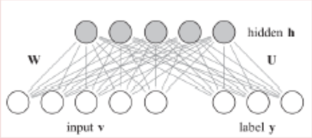
Lung texture classification and airway detection using Convolutional classification Restricted Boltzmann Machine (RBM)



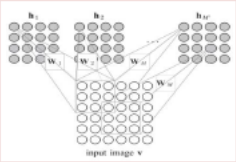
Airway dataset ; airway centerline (green) and non-airway (red)

2. Method:

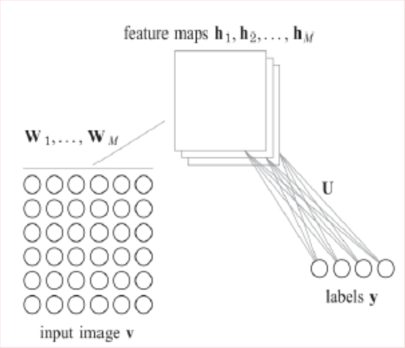
- A classification RBM is constructed by having a extra layer of labeled nodes to the visible layer.
- As in convolutional neural network, a convolutional RBM use the same weight sharing approach.
- A convolutional classification RBM (CC-RBM) have all visible, hidden and label layers
- A CC-RBM can be trained as a discriminative model and be tested to perform classification



A classification RBM



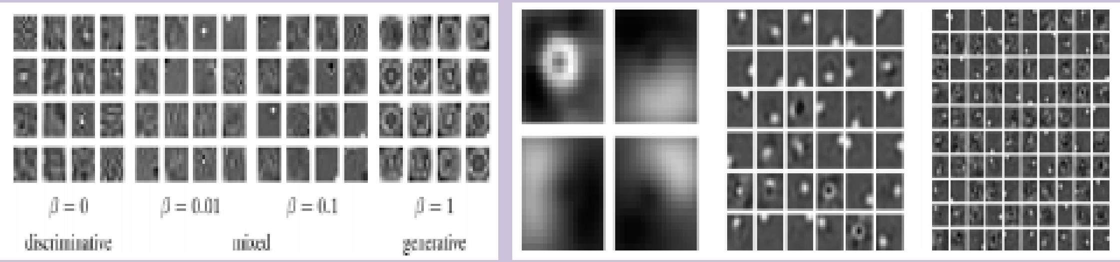
Convolutional RBM



Convolutional Classification RBM

3. Results

- Lung tissue classification on 73 scans, 40 scans for airway detection.
- A combination of generative and discriminative learning gives better classification accuracy than either of them alone.



Features learnt from Lung texture (left) and Airway (right) datasets.

van Tulder, Gijs, and Marleen de Bruijne. "Combining Generative and Discriminative Representation Learning for Lung CT Analysis With Convolutional Restricted Boltzmann Machines." *IEEE transactions on medical imaging* 35.5 (2016)

Lymph Node Detection

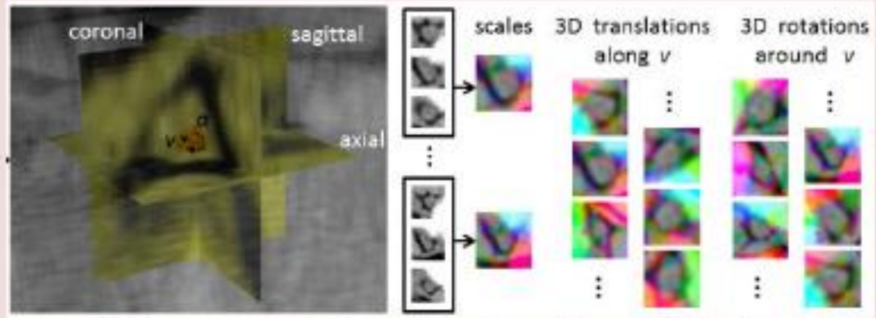
1. Objective:

Lymph node (LN) detection in 176 CT scans using 2.5D Convolutional Neural Network.

Lymph node in an axial CT slice marked in green

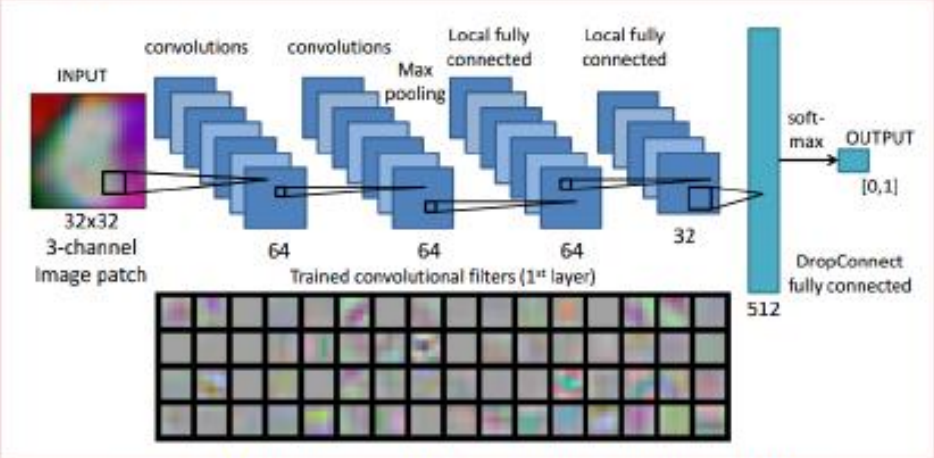
2. Method:

- The 3 views of CT are considered as different channels (RGB) of an image.
- Data augmentation is performed using random translation and rotation



Data augmentation by random translation and rotation

- 2 convolution, a max pooling, 2 locally fully connected and one drop-connect layers.



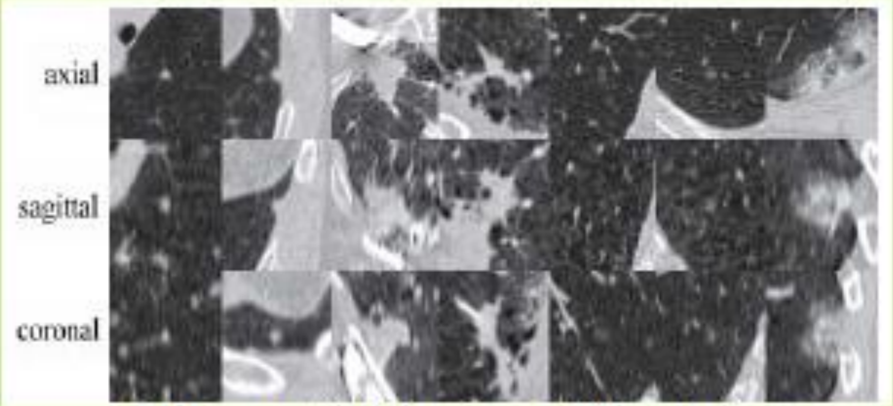
CNN Network and the learnt features from the first layer

Roth, Holger R., et al. "A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations."*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2014.

Lung Nodule Detection

1. Objective:

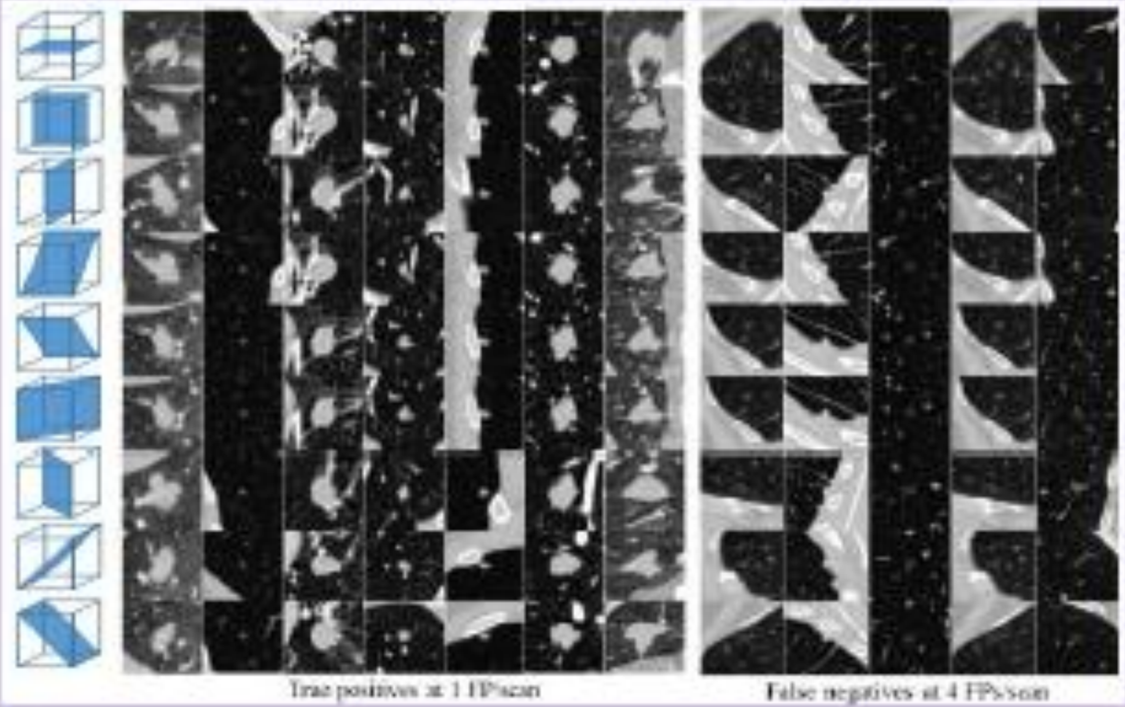
Detection of lung nodules in low dose CT images with CNN based False Positive rejection



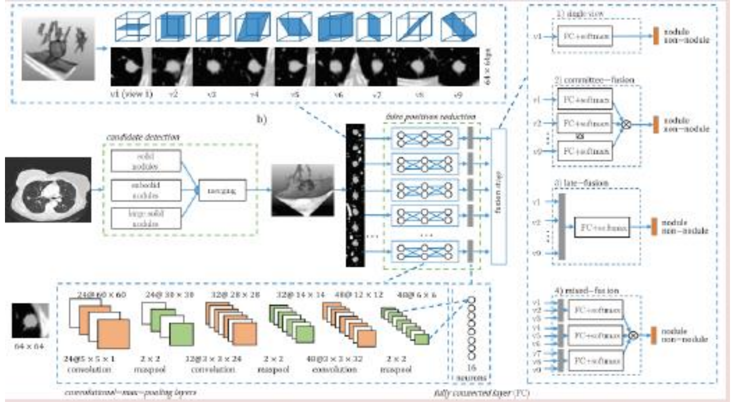
Pulmonary Nodules across different views

3. Results

- Evaluations on 3 Low Dose CT datasets with 1018, 55 and 612 scans
- Sensitivity of 90.1% with 4 False Positives per scan



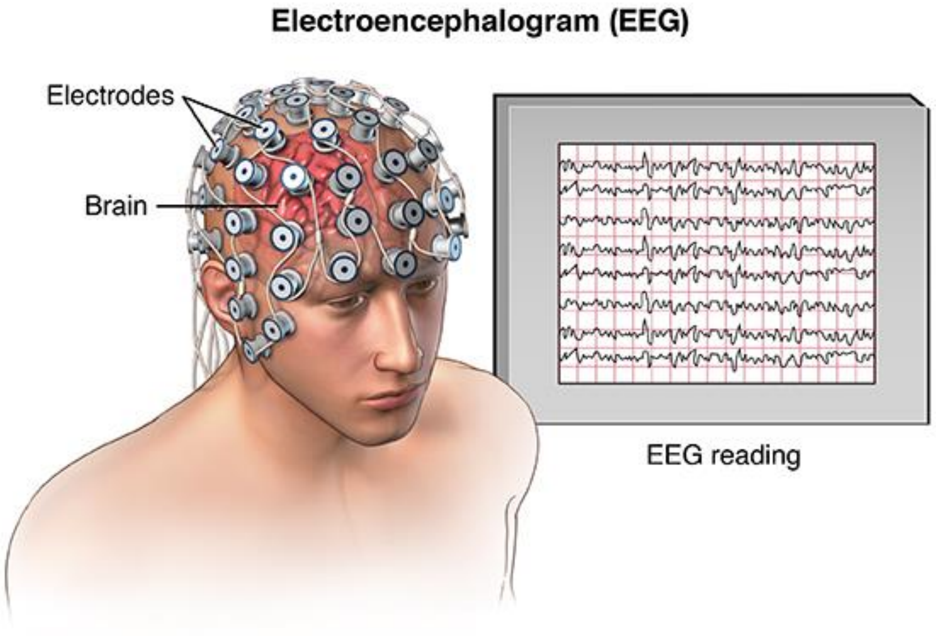
Method:



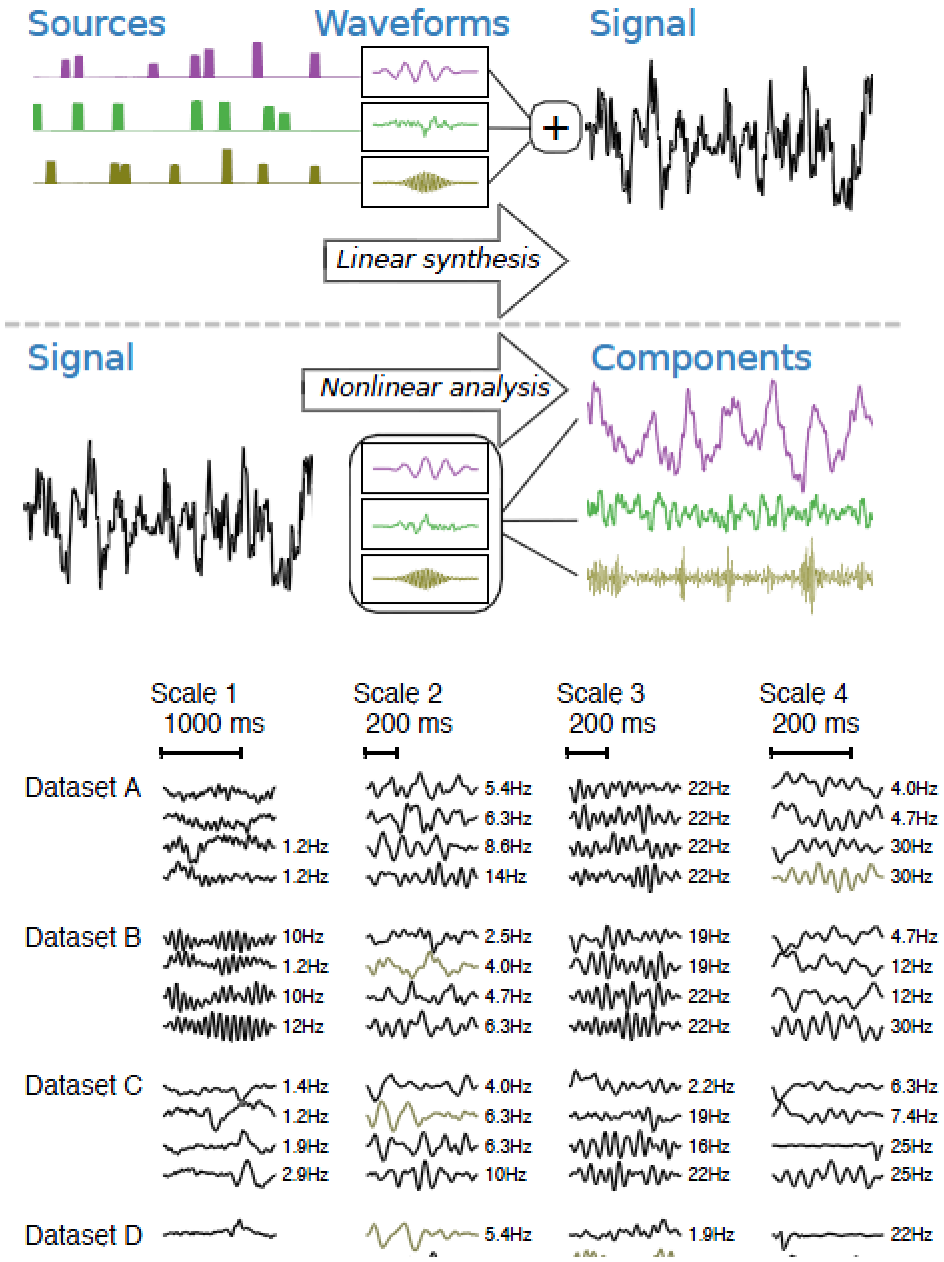
Setio, Arnaud Arindra Adiyoso, et al. "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks."*IEEE transactions on medical imaging* 35.5 (2016)

Electroencephalogram

Objective: demonstrate an algorithm to automatically learn the time-limited waveforms associated with phasic events that repeatedly appear throughout an electroencephalogram



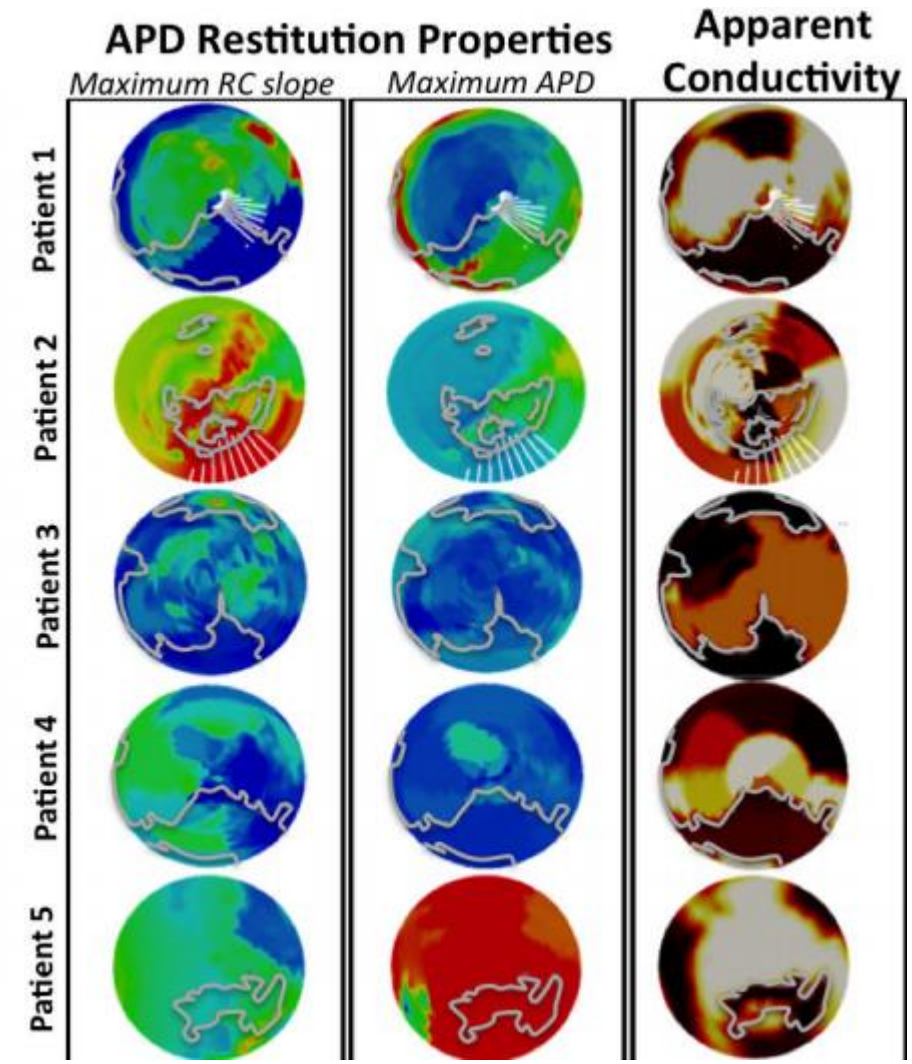
Method:

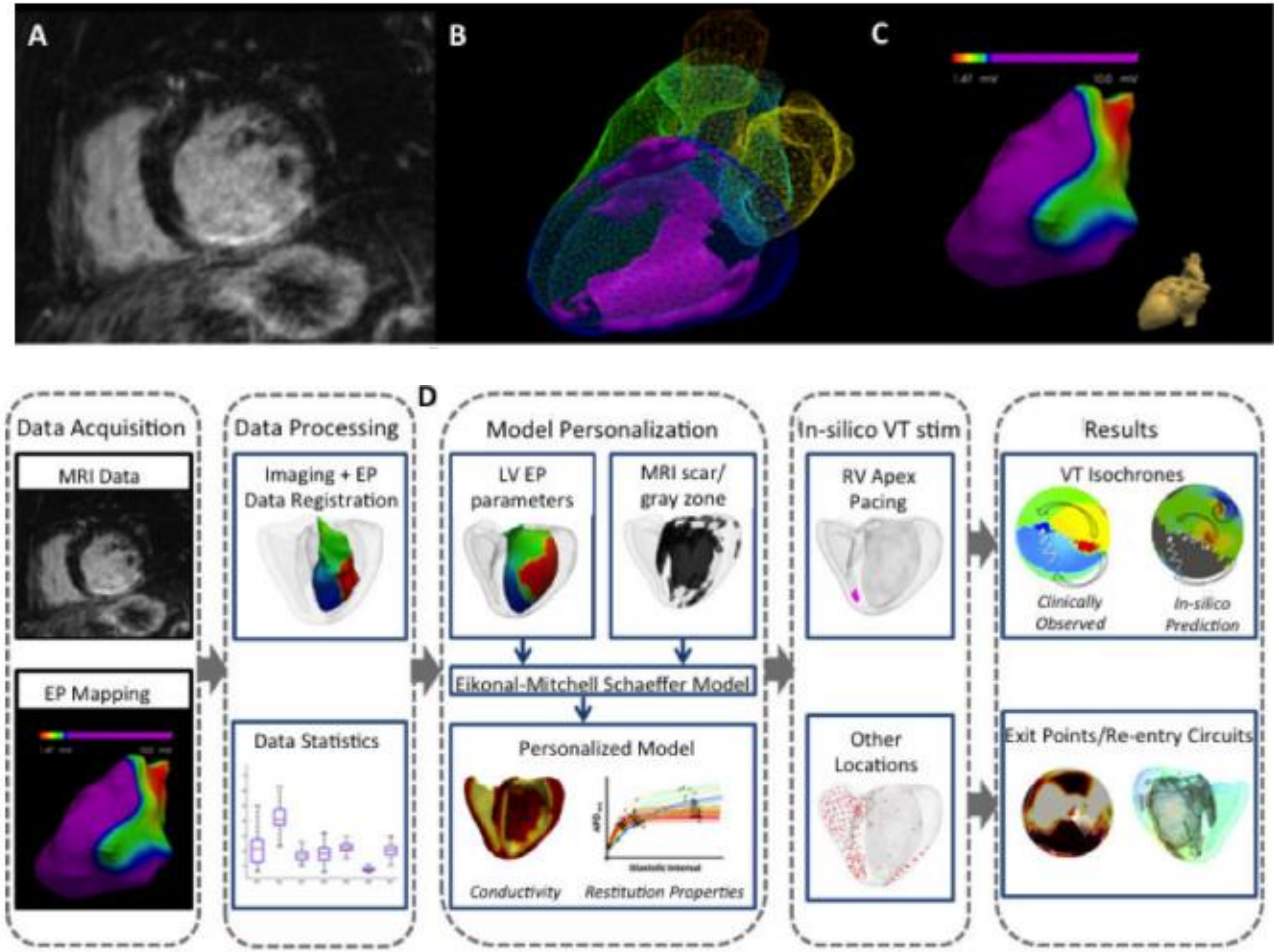


Austin J. Brockmeier et al. 'Learning Recurrent Waveforms within EEGs', TBME

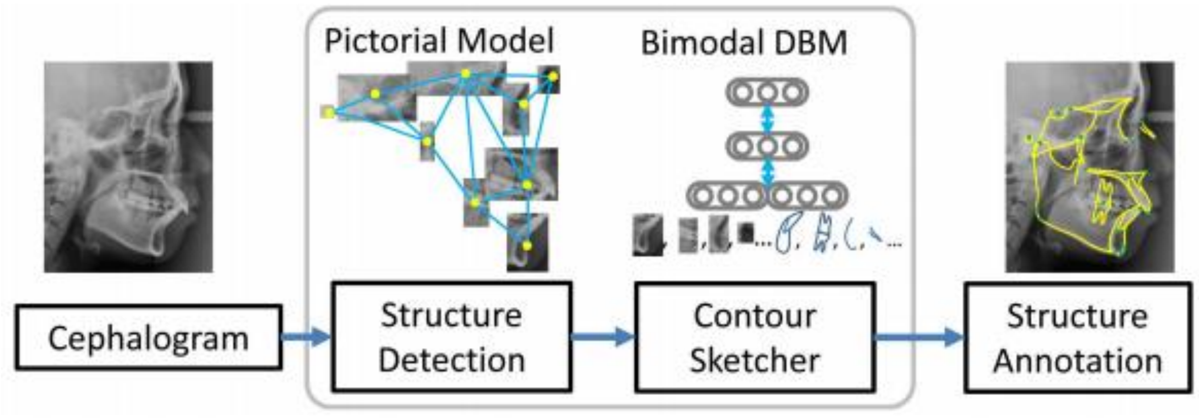
Electrophysiological mapping

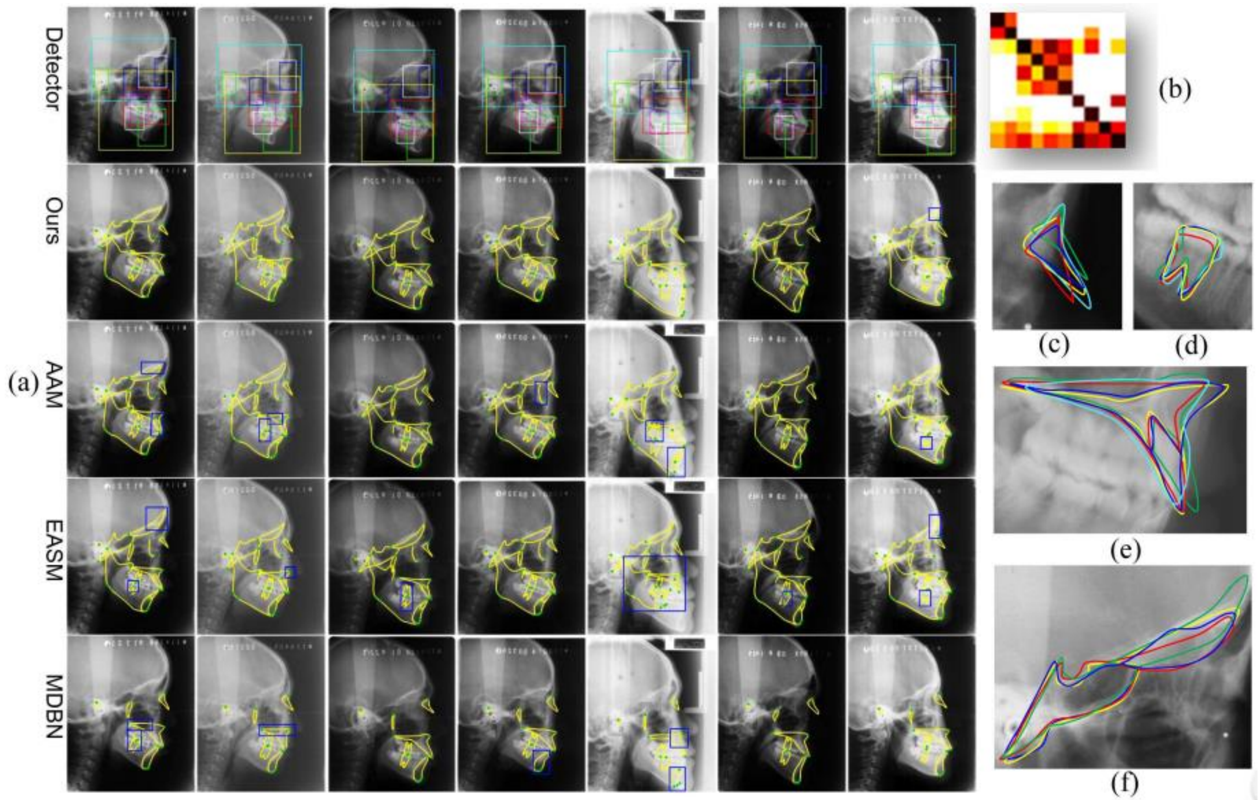
https://hal.inria.fr/tel-01206478/document/
http://www-sop.inria.fr/members/Nicholas.Ayache





Anatomical Structure Sketcher for Cephalograms
by Bimodal Deep Learning





The lateral cephalogram is a commonly used medium to acquire patient-specific morphology for diagnose and treatment planning in clinical dentistry. The robust anatomical structure detection and accurate annotation remain challenging considering the personal skeletal variations and image blurs caused by device-specific projection magnification, together with structure overlapping in the lateral cephalograms.