AIRI 지능정보기술연구원

Deep Learning Applications in Medicine:
Present and Future Perspectives Based on Experience

2017.10.31 NVIDIA Deep Learning Day

지능정보기술연구원 김휘영 PhD
astaria82@gmail.com

Contents

- MR Image based Pseudo-CT Image Synthesis using Conditional Generative Adversarial Network
- MR Image based Gleason Score Classification for Prostate Cancer Patients
- Deep Learning in Medicine

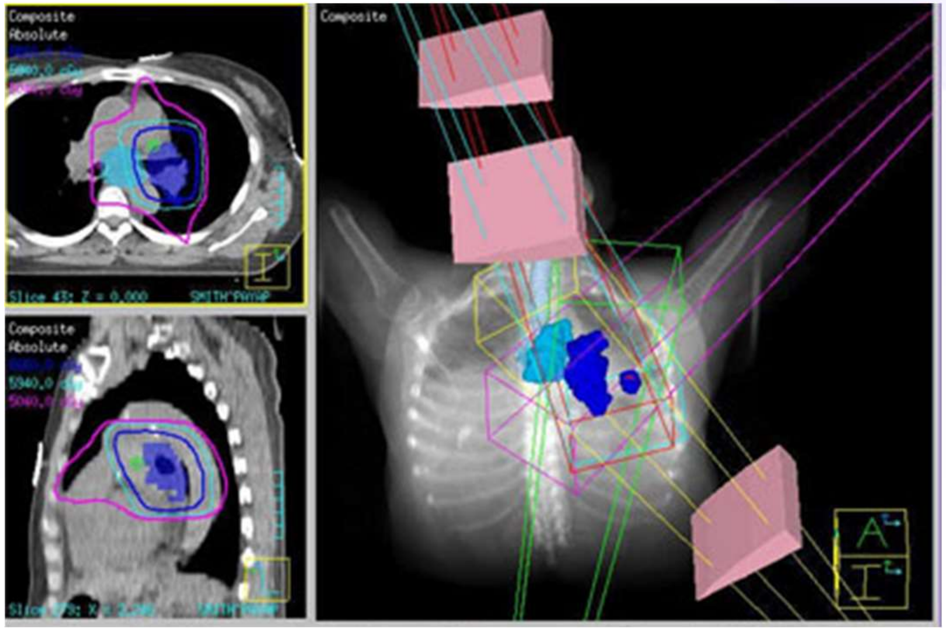
지능정보기술연구원

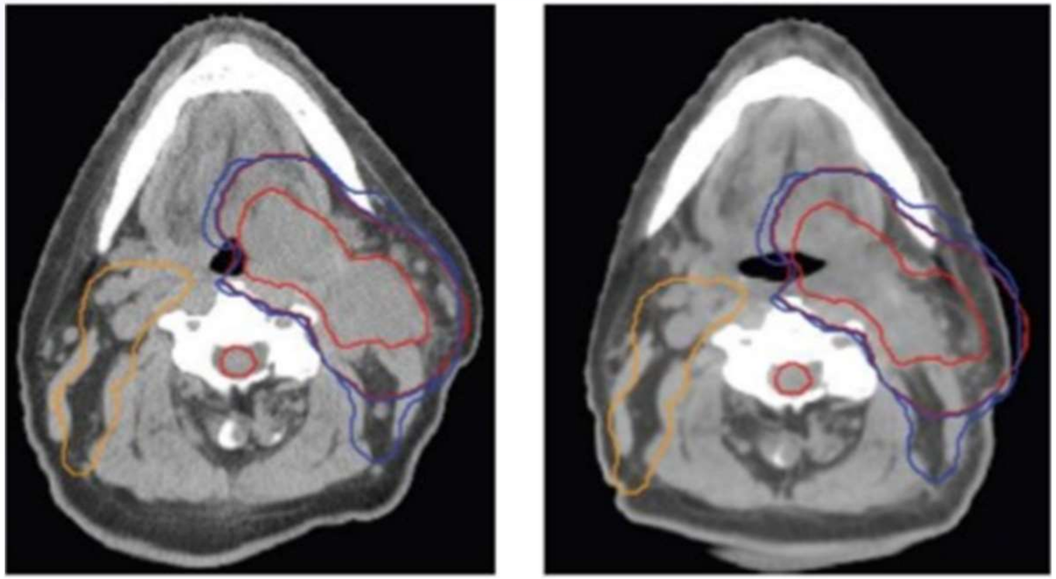
MR Image based Pseudo-CT Image Synthesis using Conditional Generative Adversarial Network

2017.10.31 NVIDIA Deep Learning Day

지능정보기술연구원 김휘영 PhD
astaria82@gmail.com

Image-guided RT Planning





Planning CT

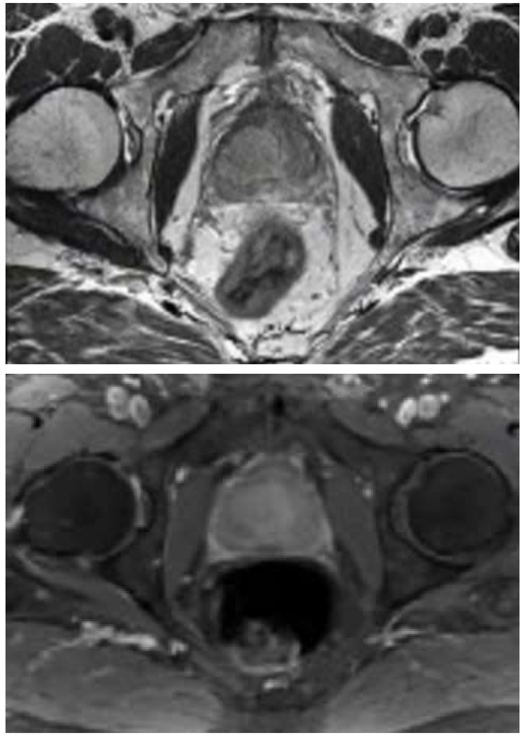
During Treatment

MR Image-based RT Planning

- Patients are exposed to radiation during CT imaging
  - 0.4 % of cancers were due to CT scanning (Hall *et al.* 2007)

  → not suitable for real-time/daily adaptive RT

- MR-based RT benefits:
  - Non-exposure of X-ray radiation
  - Superior and multiple tissue contrast compared w/ CT
  - Reduced examination time and cost
  - Benefits to MR-LINAC

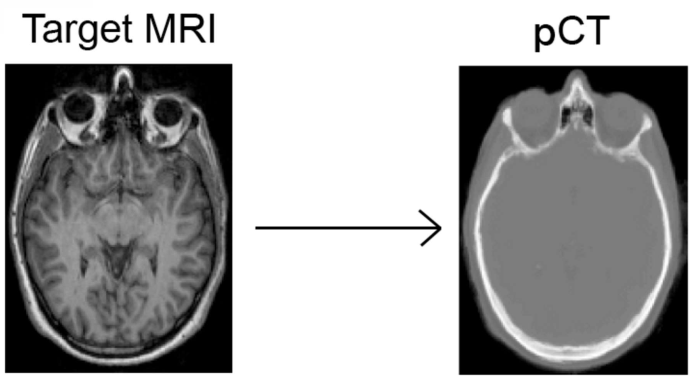


Need to estimate CT Image

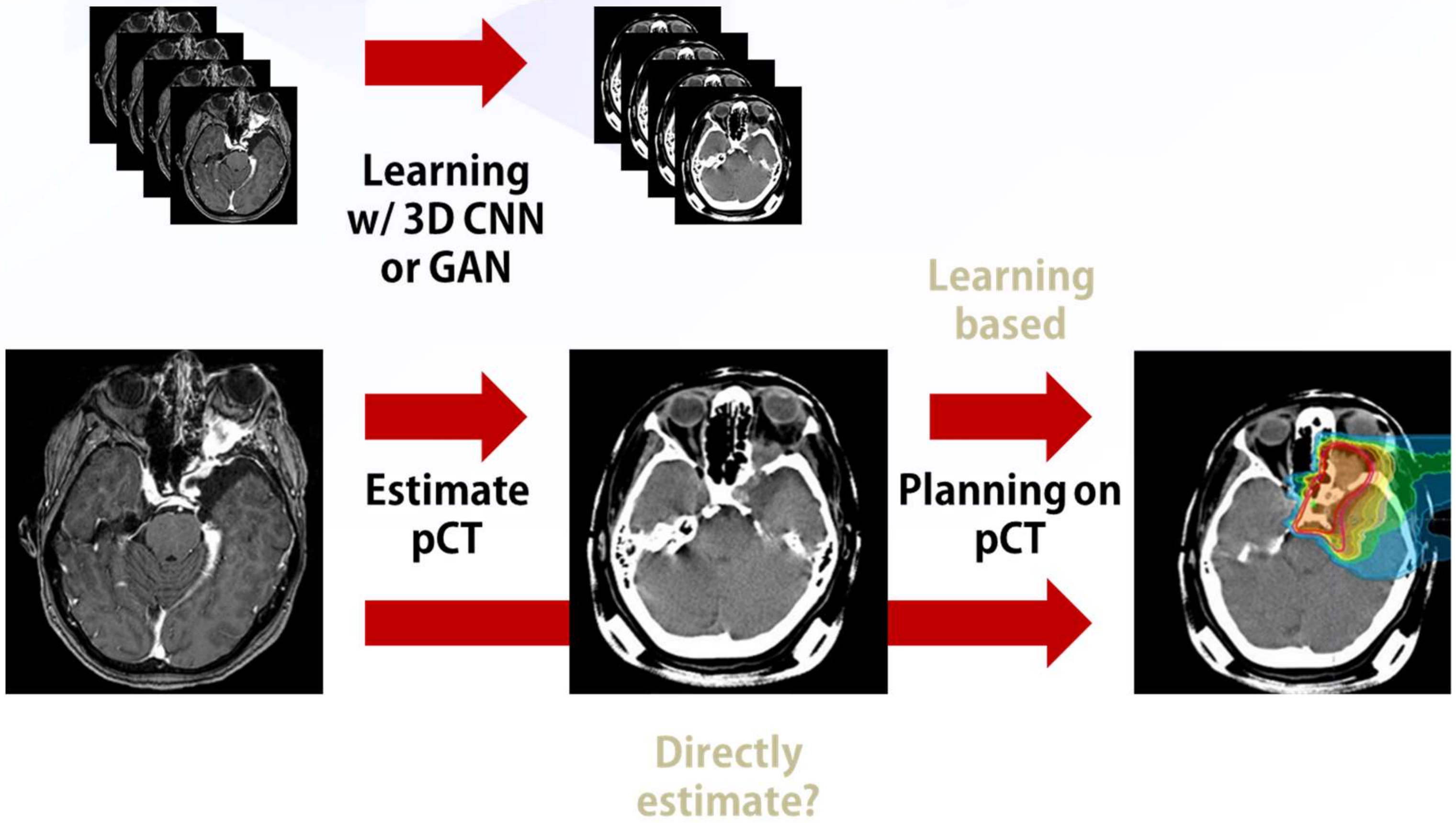
- CT scan provides Hounsfield units (HU)
  - A normalized value of the linear attenuation coefficient
  - Essential for dose calculation in radiation treatment planning system (RTPS)
  - To consider tissue inhomogeneity

- MRI itself is cannot be directly used for treatment planning
- **Estimate CT image (pseudo CT, pCT) from MR image**

- Intensity-based & Atlas-based



Objective


Learning
w/ 3D CNN
or GAN
Learning
based
Estimate
pCT
Planning on
pCT
Directly
estimate?

Dataset

- CT-MR image pairs of 19 glioblastoma patients
  - From TCIA (the Cancer Imaging Archive) open medical database
  - http://www.cancerimagingarchive.net/

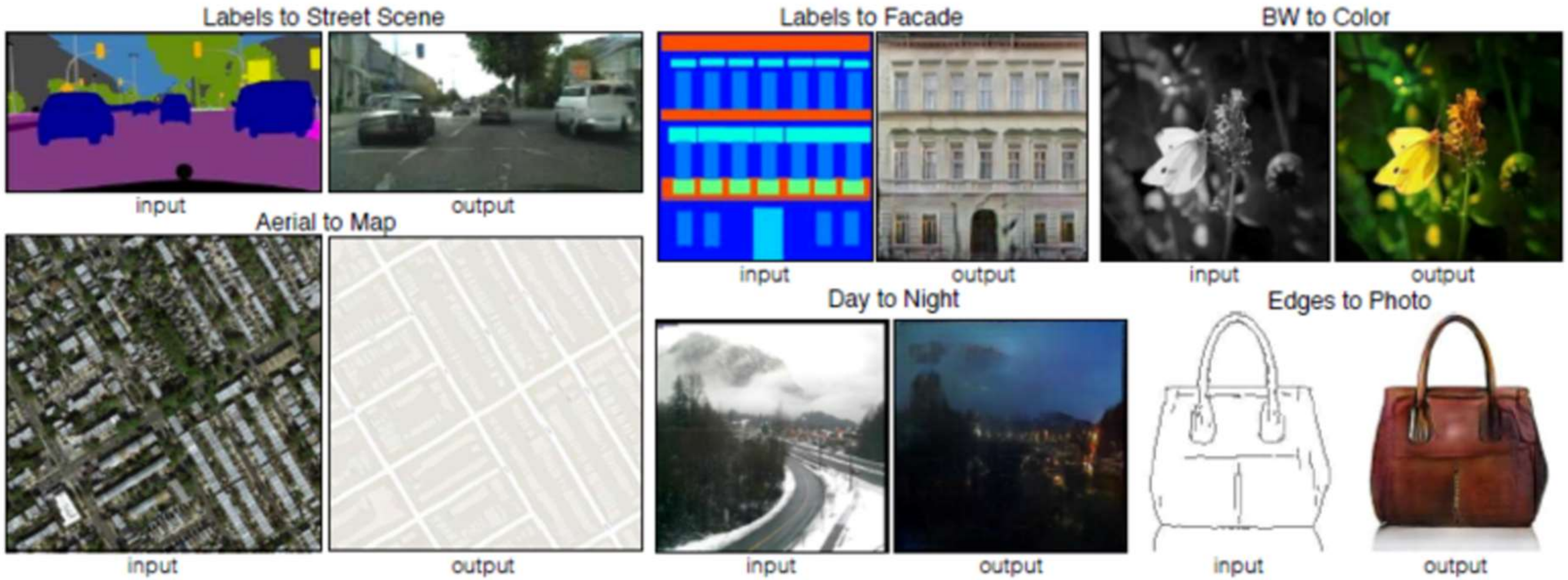
  - TCGA-GBM (The Cancer Genome Atlas Glioblastoma Multiforme)
    - Multi-institutional data (Henry Ford, UCSF, MDACC, Emory, Duke, …)
    - CT, MR, Pathologic sildes, Dx (with genomic data)
    - 262 patients, 575 studies w/ 481,158 images (73.5 GB)

  - Candidate selection criteria
    - Has CT and MR image pairs (interval within 1 month)
    - No significant noise/motion artifact

Scarpace, L., Mikkelsen, T., Cha, soonmee, Rao, S., Tekchandani, S., Gutman, D., … Pierce, L. J. (2016). Radiology Data from The Cancer Genome Atlas Glioblastoma Multiforme [TCGA-GBM] collection. The Cancer Imaging Archive

pix2pix

- P. Isola *et al.*, Image-to-Image Translation with Conditional Adversarial Networks, arXiv: 1611.07004v1



Conditional Generative Adversarial Network

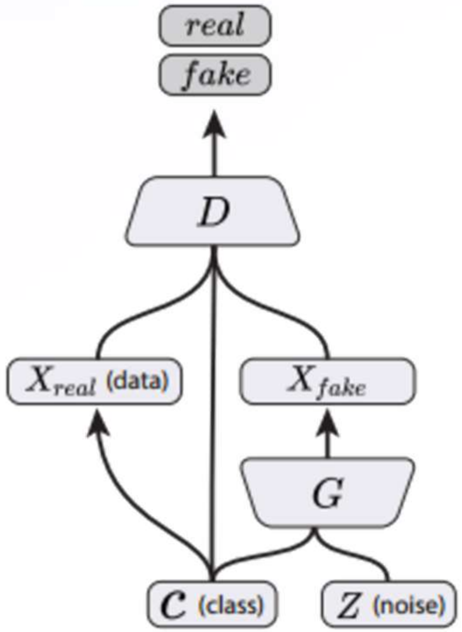
- **G (generator)** is a fully connected network
  - to generate pseudo CT images *G(C, z)*
    - from a random noise vector z
    - under the condition C (corresponding MR image)
- **D (discriminator)** is a convolutional neural network
  - discriminate btw/ an real CT image (ground truth) and an estimated pseudo-CT image.
- The G tries to minimize objective function against the D which tries to maximize it (min-max problem)

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x, y)] + \mathbb{E}_{x\sim p_{data}(x), z\sim p_z(z)}[\log(1 - D(x, G(x, z))], \quad (1)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y\sim p_{data}(x,y), z\sim p_z(z)}[\|y - G(x, z)\|_1].$$

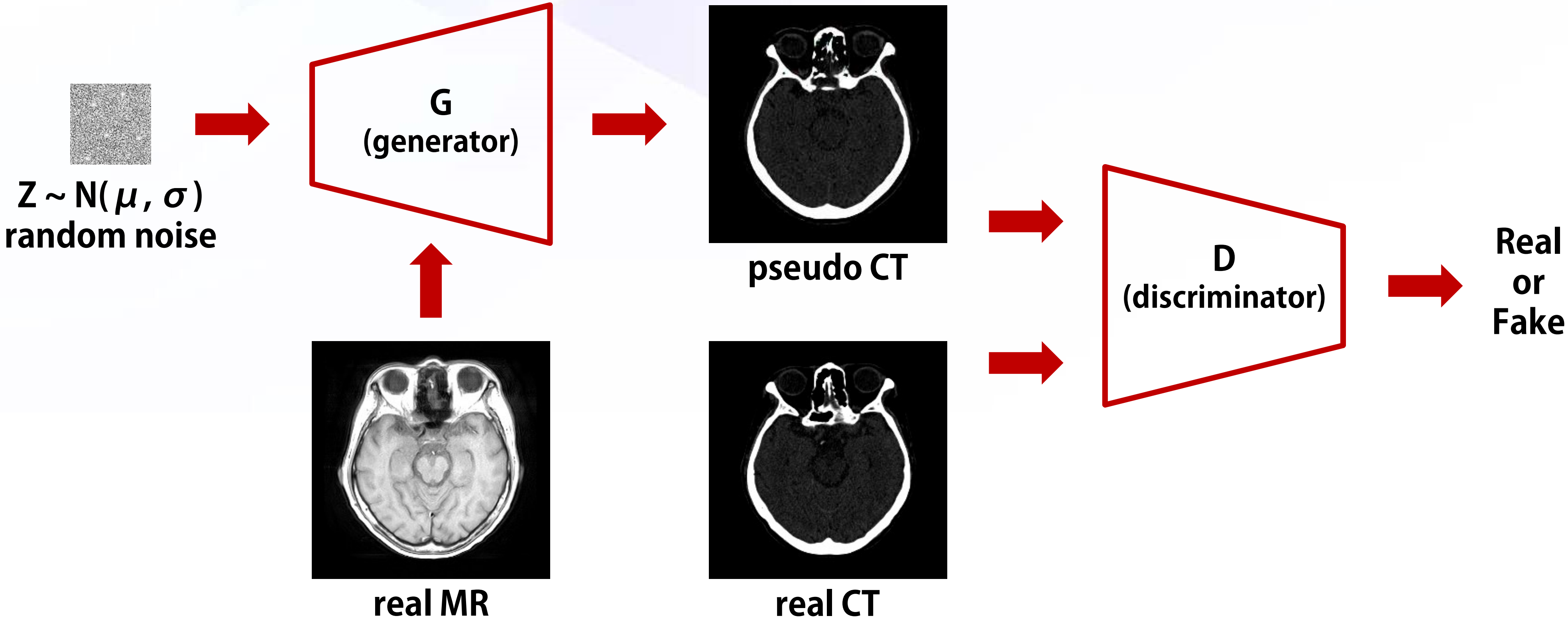
Our final objective is

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda\mathcal{L}_{L1}(G).$$

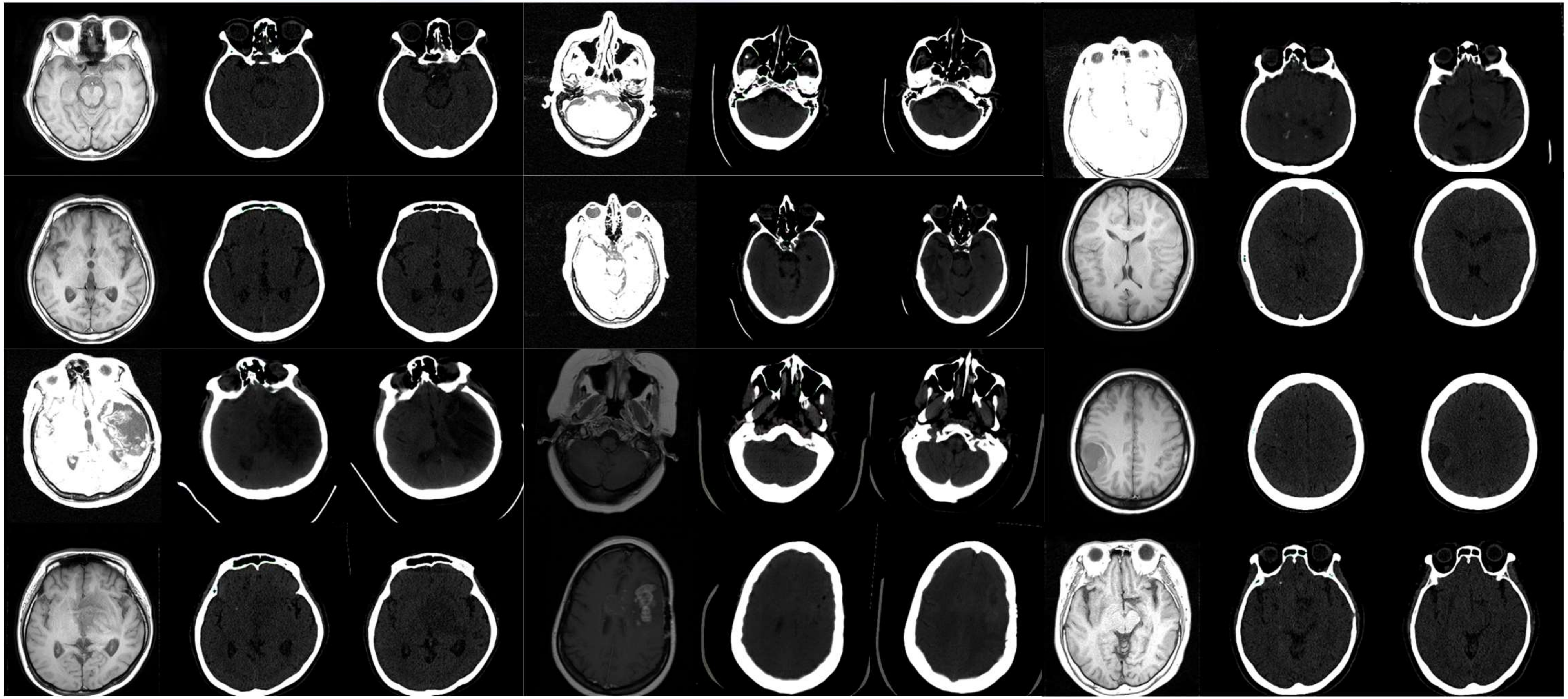


Conditional GAN
(Mirza & Osindero, 2014)

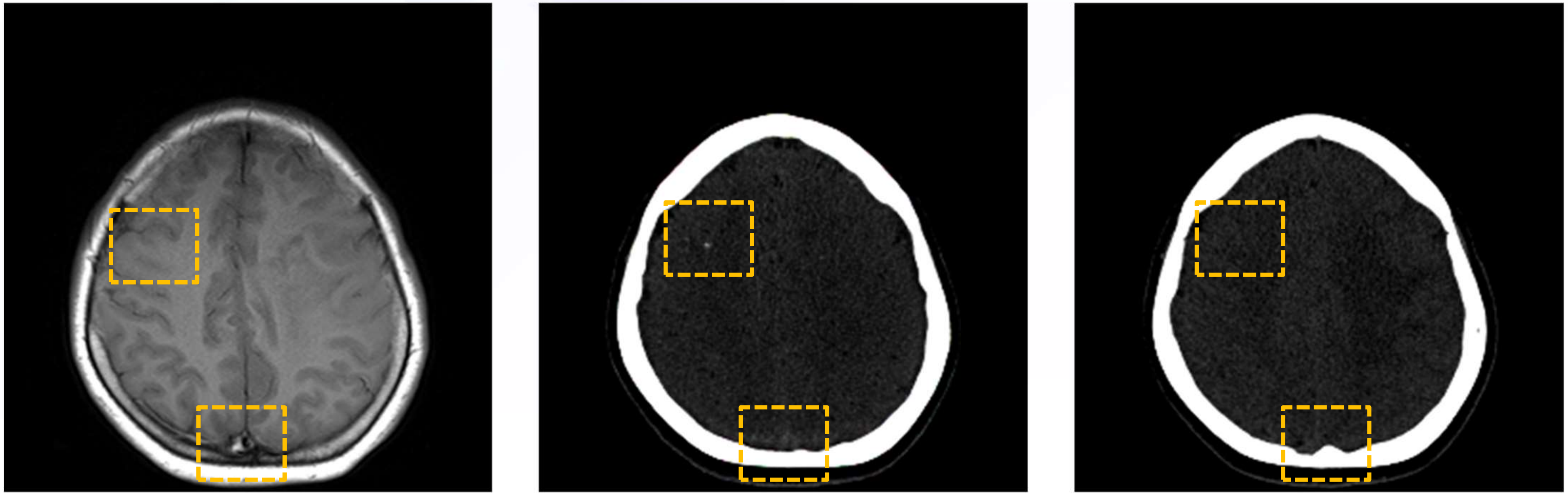
Conditional Generative Adversarial Network



Examples

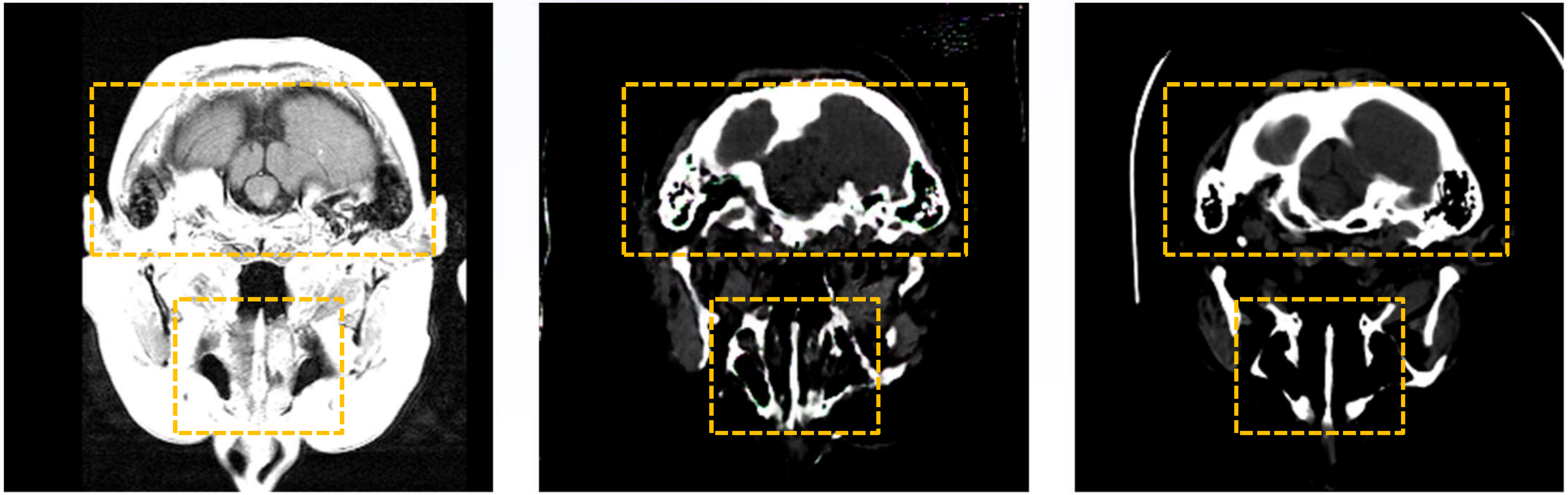


Best case (DSC = 0.986)



MR | pseudo CT | real CT (ground truth)

Worst case (DSC = 0.766)



MR

pseudo CT

real CT
(ground truth)

Discussion and Conclusions

- Generated pseudo-CT images well reconstructed anatomical boundaries but there were discrepancies in cavities and eye ball
- Training with more dataset will overcome overfitting and thus enhance the quality of pseudo-CT synthesis

- This results showed that our proposed method is feasible for predicting pseudo-CT images from their corresponding MR images.
- This technique has a potential to enable us to establish MR-based adaptive RT in clinic

Acknowledgement

- 서울대학교 융합과학기술대학원 예성준 교수님
- 분당서울대병원 영상의학과 이경준 교수님
- 지능정보기술연구원 이광희, 박대영 연구원

- The results shown here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/

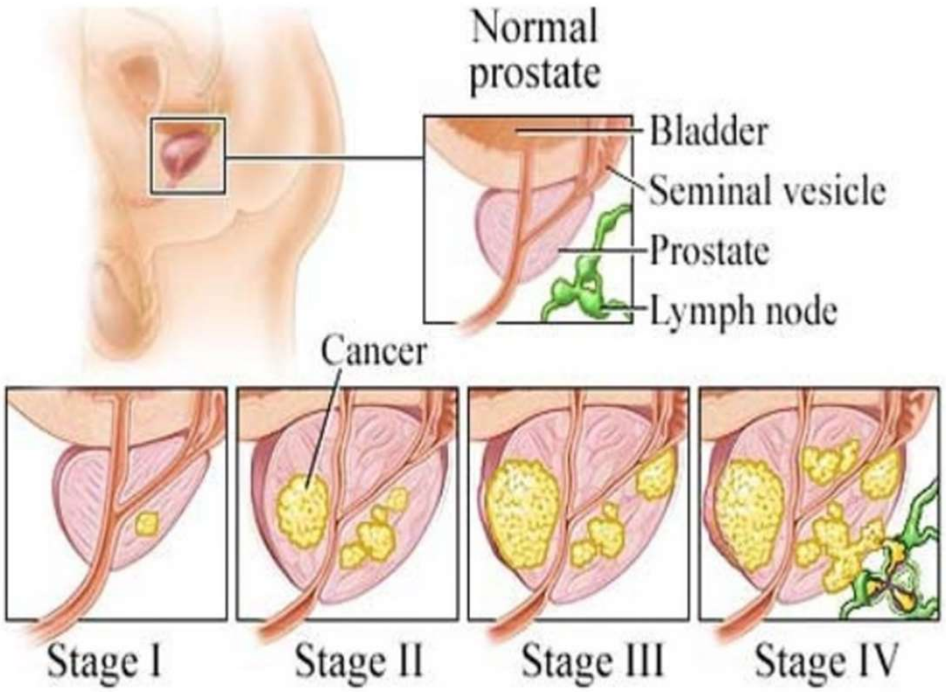
지능정보기술연구원

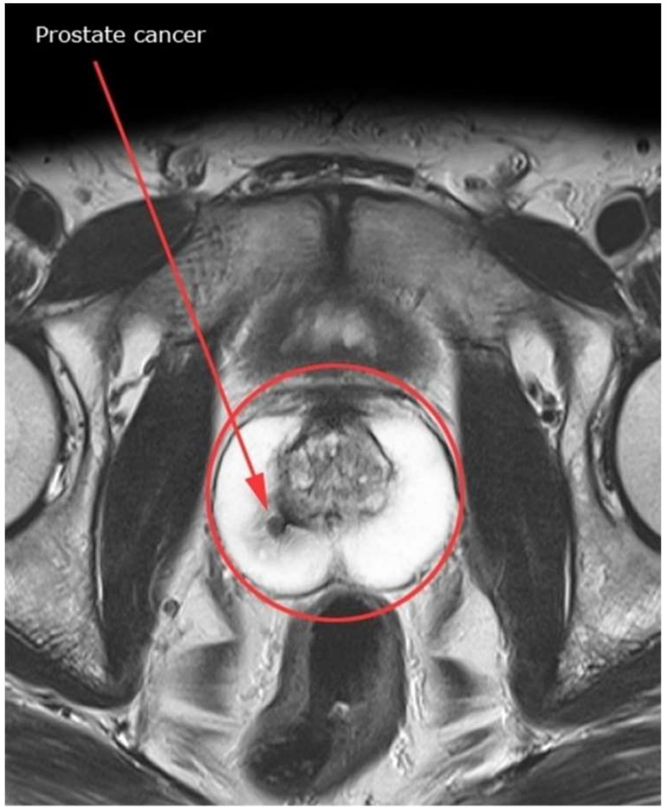
MR Image based Gleason Score Classification for Prostate Cancer Patients

2017.10.31 NVIDIA Deep Learning Day

지능정보기술연구원 김휘영 PhD
astaria82@gmail.com

Prostate Cancer Diagnosis

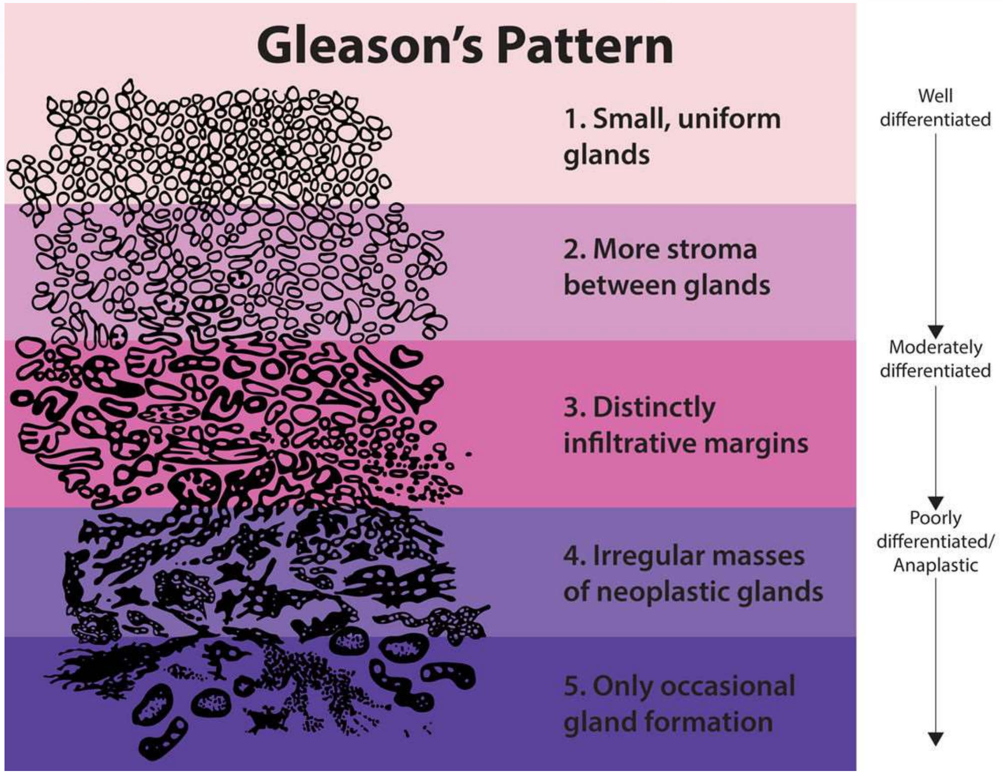




Gleason Score

- A system of grading prostate cancer tissue based on how it looks under a microscope (pathologic image)

- Range: 2~10
- Indicates how likely it is that a tumor will spread



Gleason Grade Group

Grade Group 1 (Gleason score ≤ 6): Only individual discrete well-formed glands

Grade Group 2 (Gleason score 3+4 = 7): Predominantly well-formed glands with lesser component of poorly-formed/fused/cribriform glands

Grade Group 3 (Gleason score 4+3 = 7): Predominantly poorly formed/fused/cribriform glands with lesser component of well-formed glands

Grade Group 4 (Gleason score 4+4 = 8; 3+5 = 8; 5+3 = 8): (1) Only poorly-formed/fused/cribriform glands or (2) predominantly well-formed glands and lesser component lacking glands or (3) predominantly lacking glands and lesser component of well-formed glands

Grade Group 5 (Gleason scores 9-10): Lacks gland formation (or with necrosis) with or without poorly formed/fused/cribriform glands

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, the Grading Committee. The 2014 International Society of Urologic Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of grading patterns and proposal for a new grading system. Am J Surg Pathol, (40)244-252, 2016

Objective

- Find quantitative multi-parametric MRI biomarkers for determination of Gleason Grade Group in prostate cancer

AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

Home | Directory | Career Services | Continuing Education | BBS | Contact

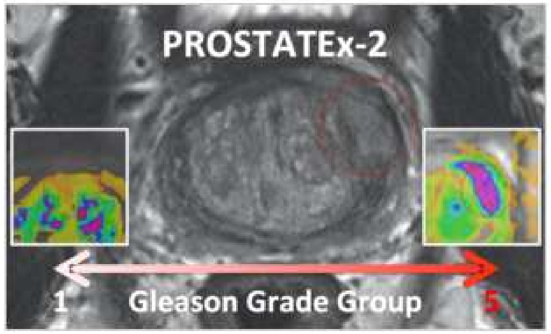
Advancing the Science, Education & Professional Practice of Medical Physics

- My AAPM
- AAPM
- Public & Media
- International
- Medical Physicist
- Members
- Students
- Meetings

PROSTATEx-2 CHALLENGE

SPIE-AAPM-NCI Prostate MR Gleason Grade Group Challenge

The American Association of Physicists in Medicine (AAPM), along with the SPIE (the international society for optics and photonics) and the National Cancer Institute (NCI), will conduct a part 2 "Grand Challenge" on the development of quantitative multi-parametric magnetic resonance imaging (MRI) biomarkers for the determination of Gleason Grade Group in prostate cancer. As part of the 2017 AAPM Annual Meeting, the PROSTATEx-2 Challenge will provide a unique opportunity for participants to compare their algorithms with those of others from academia, industry, and government in a structured, direct way using the same data sets.



Data

- Total 182 findings (lesions) from 162 cases
  - 112 for training set
  - 70 for test set (release date w/o truth: Jun 5, 2017)

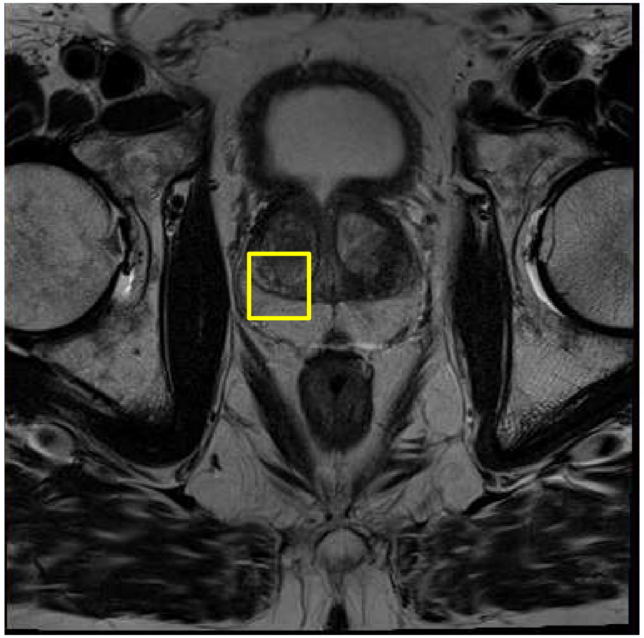
- 4 sets of MRI scan data
  - T2-weighted (axial and sagittal)
  - Dynamic contrast-enhanced (DCE)
  - Apparent diffusion coefficient (ADC)
  - Diffusion weighted imaging (DWI)

Group	1	2	3	4	5
N	36	41	20	8	7

Data imbalance

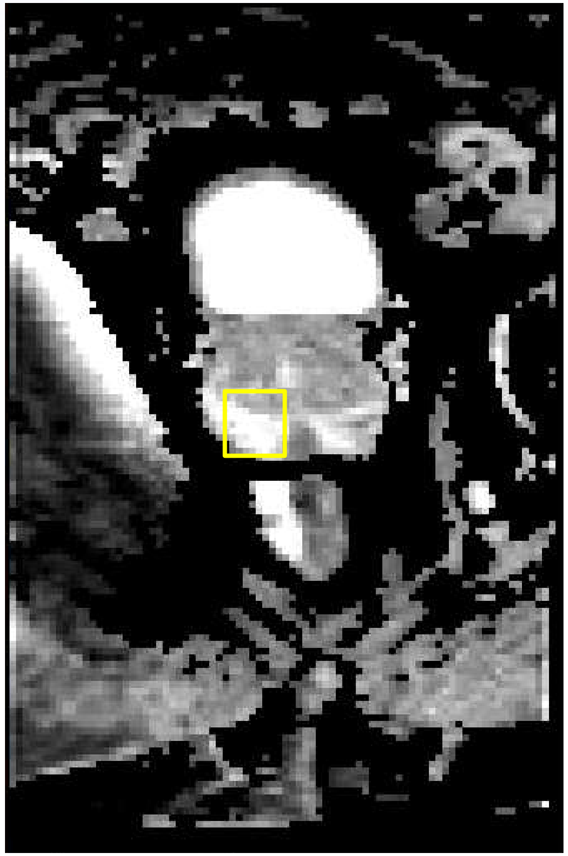
- w/ Lesion location (not a mask) and known GGG

MRI Sequences



T2-weighted (axial)

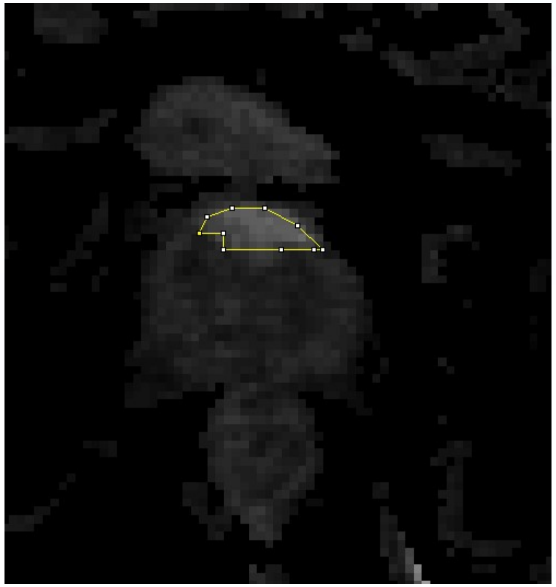
DCE

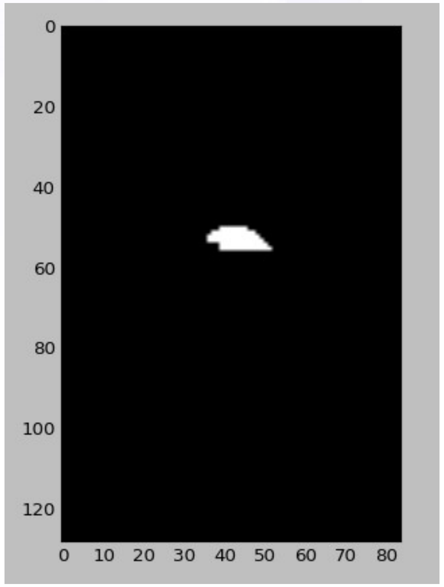


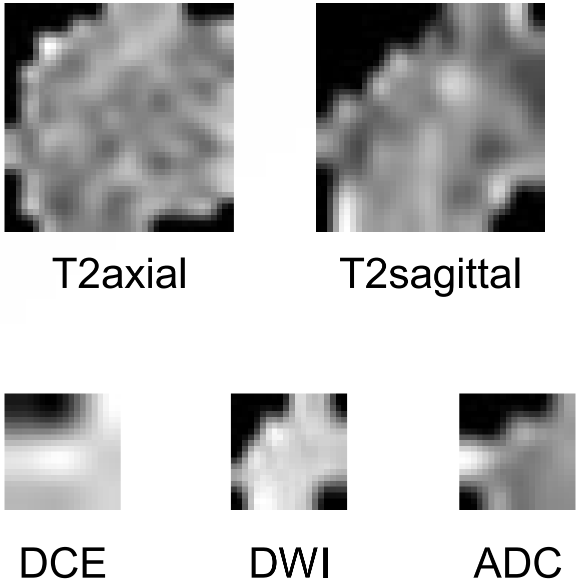
ADC

Masking

- Manual contouring (Courtesy of Dr. Woo)

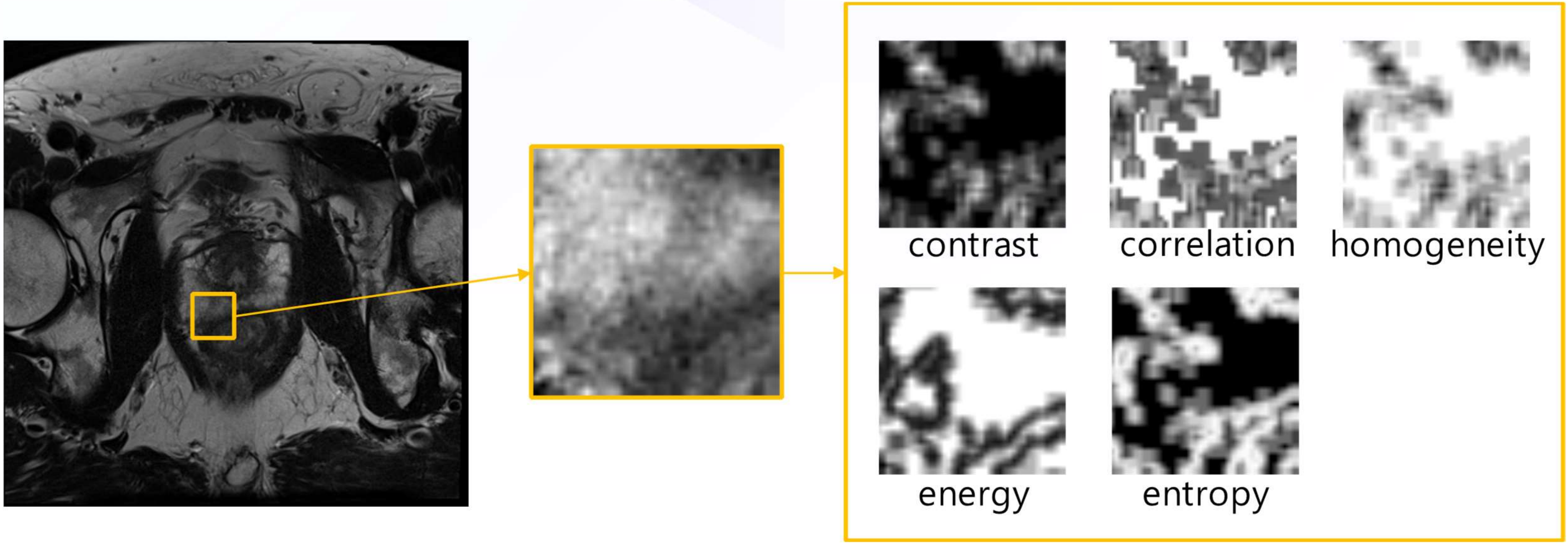






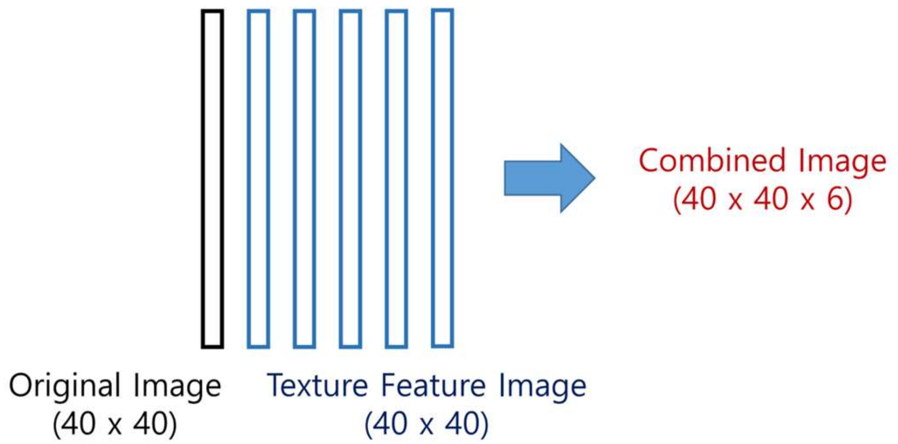
Textural Feature Map

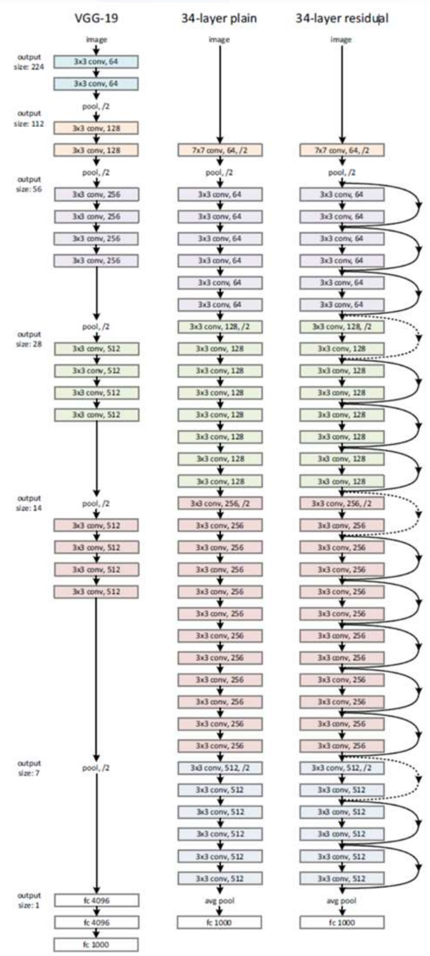
- Calculated textural features of 5x5 image patches for each voxel

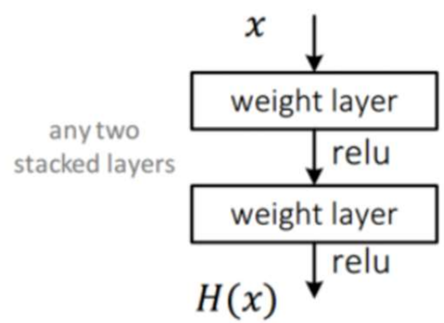


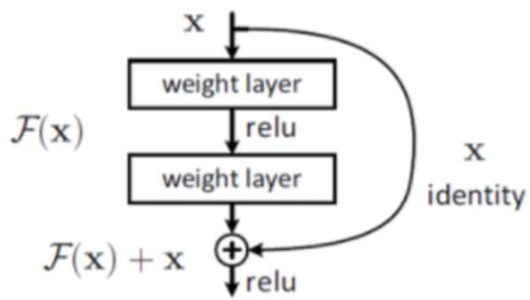
Experiment #1 (ResNet50)

- Trained w/ 7 GTX 1080





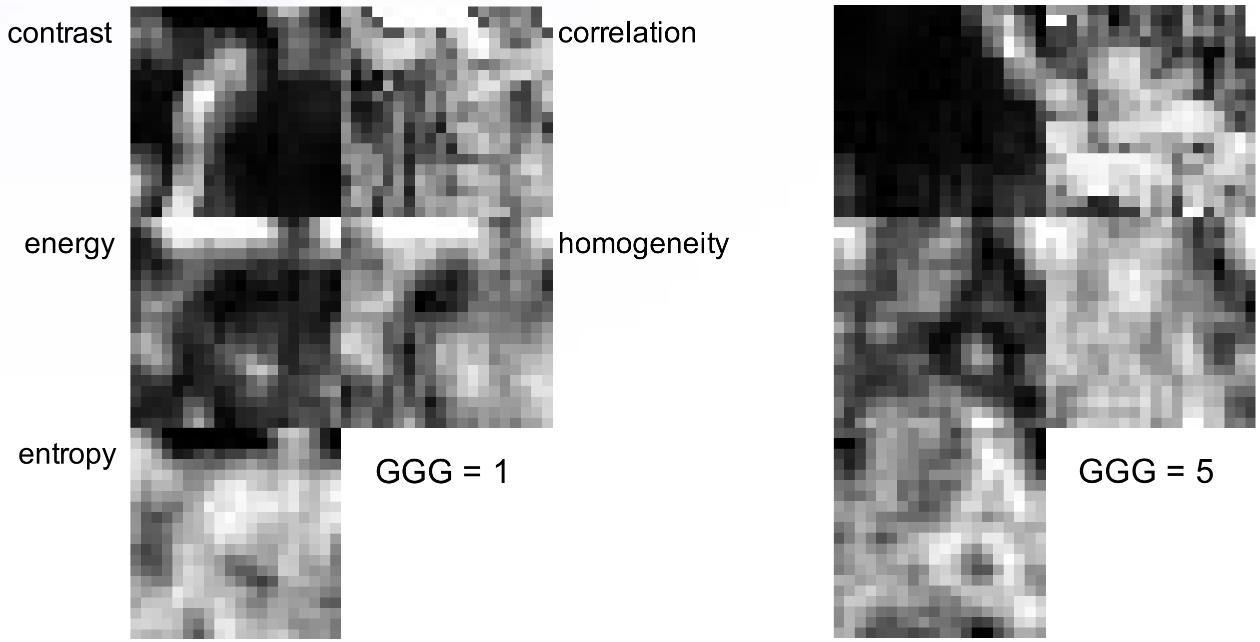




Result #1

- Validation set test accuracy = ~25 %

- Lack of training data
- Data imbalance



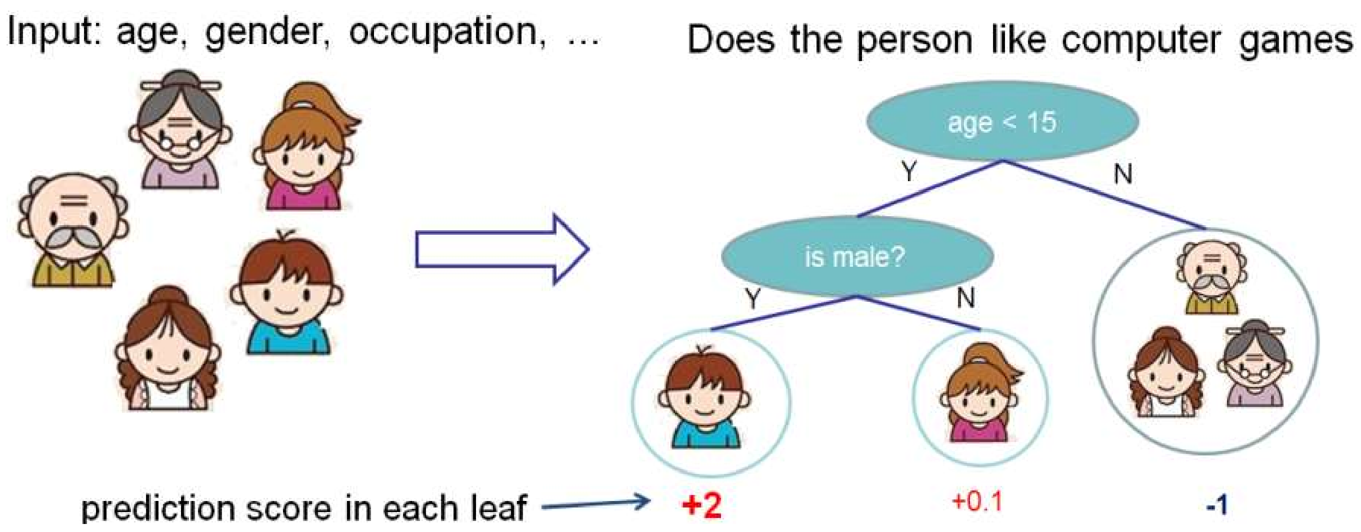
Experiment #2 (XGboost)

- Tree-ensemble method
  - w/ gradient boosting

Training loss

$$\sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_k \Omega(f_k),\ f_k \in \mathcal{F}$$

Complexity of the tree
# of nodes, depth, L2 norm of leaf weights



1^{st} Order Features

- w/ original images of each MRI sequences
- Derived w/ Intensity volume histogram

$$P(I) = \frac{number\ of\ pixels\ with\ gray\ level\ I}{total\ number\ of\ pixels\ in\ the\ region\ of\ interest}$$

  - Mean

$$mean = E[I] = \sum_{I=0}^{N_g-1} I\ P(I),\ where\ N_g\ is\ the\ number\ of\ exist\ gray\ levels$$

  - SD

$$SD = \frac{1}{N_g}\sqrt{(I - mean)^2}$$

  - Mean skewness

$$Skewness = \frac{1}{SD^3}\sum_{I=0}^{N_g-1}(I - mean)^3 P(I)$$

  - Mean kurtosis

$$Kurtosis = \frac{1}{SD^4}\sum_{I=0}^{N_g-1}(I - mean)^4 P(I)$$

Result #2

- Validation set test accuracy: Around 80%
  - (Validation set = 33 % of whole training set)

  - w/ all sequences: 31.65 %
  - T2ax: 79.35 %
  - T2sag: 82.76 %
  - ADC: 73.40 %
  - DWI: 80.67 %
  - DCE: 85.71 %

```
XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=1,
       gamma=0, learning_rate=0.3, max_delta_step=0, max_depth=10,
       min_child_weight=1, missing=None, n_estimators=1000, nthread=4,
       objective='binary:logistic', reg_alpha=0, reg_lambda=1,
       scale_pos_weight=1, seed=0, silent=True, subsample=0.8)
[ 1.  4.  1.  3.  2.  1.  2.  3.  3.  1.  1.  5.  2.  3.  1.  1.  5.  1.
  5.  3.  1.  3.  2.  2.  3.  4.  5.  2.  2.  2.  5.  3.  1.  2.  1.  3.
  1.  3.  3.  2.  1.  3.  5.  1.  1.  4.  3.  3.  3.  2.  3.  3.  1.  1.
  3.  5.  1.  2.  1.  5.  1.  2.  2.  3.  3.  1.  5.  1.  3.  2.  4.  4.
  5.  2.  2.  1.  2.  2.  1.  2.  2.  1.  1.  3.  1.  1.  2.  3.  3.  4.
  2.  3.  2.  1.  2.  2.  2.  2.  1.  1.  3.  2.  1.  2.  2.  1.  1.  3.
  1.  5.  5.  1.  2.  4.  2.  2.  2.  1.  1.  2.  2.  3.  2.  4.  2.  5.
  1.  5.  2.  3.  2.  3.  2.  1.  1.  3.  4.  2.  2.  2.]
[1.0, 2.0, 1.0, 3.0, 2.0, 1.0, 2.0, 3.0, 3.0, 1.0, 1.0, 1.0, 2.0, 3.0, 3.0
0, 2.0, 3.0, 3.0, 2.0, 2.0, 2.0, 2.0, 5.0, 3.0, 1.0, 2.0, 1.0, 3.0, 1.0, 3
4.0, 3.0, 3.0, 3.0, 2.0, 3.0, 3.0, 1.0, 1.0, 3.0, 5.0, 1.0, 2.0, 1.0, 5.0,
, 2.0, 2.0, 4.0, 2.0, 5.0, 2.0, 2.0, 1.0, 1.0, 2.0, 1.0, 2.0, 2.0, 1.0, 3.
.0, 3.0, 2.0, 1.0, 2.0, 2.0, 2.0, 1.0, 1.0, 1.0, 3.0, 2.0, 2.0, 2.0, 2.0,
 4.0, 2.0, 2.0, 2.0, 3.0, 1.0, 2.0, 2.0, 3.0, 2.0, 4.0, 2.0, 1.0, 1.0, 5.0
0, 4.0, 1.0, 2.0, 2.0]
Accuracy all: 31.65%
Accuracy T2ax: 79.35%
Accuracy T2sag: 82.76%
Accuracy ADC: 73.40%
Accuracy DWI: 80.67%
Accuracy DCE: 85.71%
```

- Final test result: 0.1022 (quadratic weighted kappa, 0.2772 for 1st place team)
- Achieved 8th highest score (out of 143 participants)

Discussion and Conclusions

- Data preprocessing was done successfully.
- Data augmentation was needed because of lack of data and imbalance among groups.

- In case of ResNet50 model, prediction accuracy was quite low.
  - Lack of data and time
- Using XGboost method, we got meaningful prediction accuracy around 80 % but not for test data (top 8th)
- For further research, various data augmentation methods and other deep learning models can be applied in this problem.

Acknowledgement

- 서울대학교 융합과학기술대학원 예성준 교수님
- 서울대병원 영상의학과 조정연 교수님
- 국군대전병원 영상의학과 우성민 선생님
- 서울대학교 융합과학기술대학원 이지민 연구원
- 조형주 연구원

지능정보기술연구원

Deep Learning in Medicine

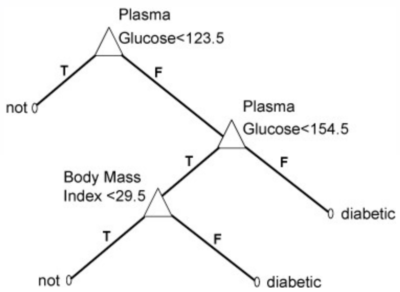
2017.10.31 NVIDIA Deep Learning Day

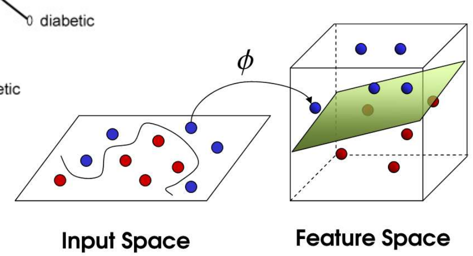
지능정보기술연구원 김휘영 PhD
astaria82@gmail.com

Traditional Machine Learning

Make a Program to get desired output from corresponding input data

INPUT → Feature Extraction → Classifier → OUTPUT

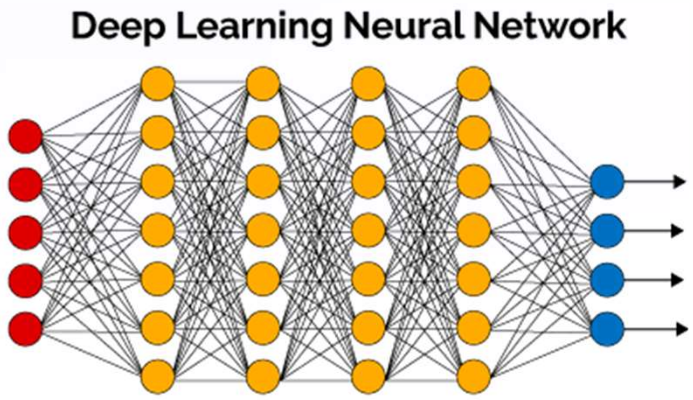




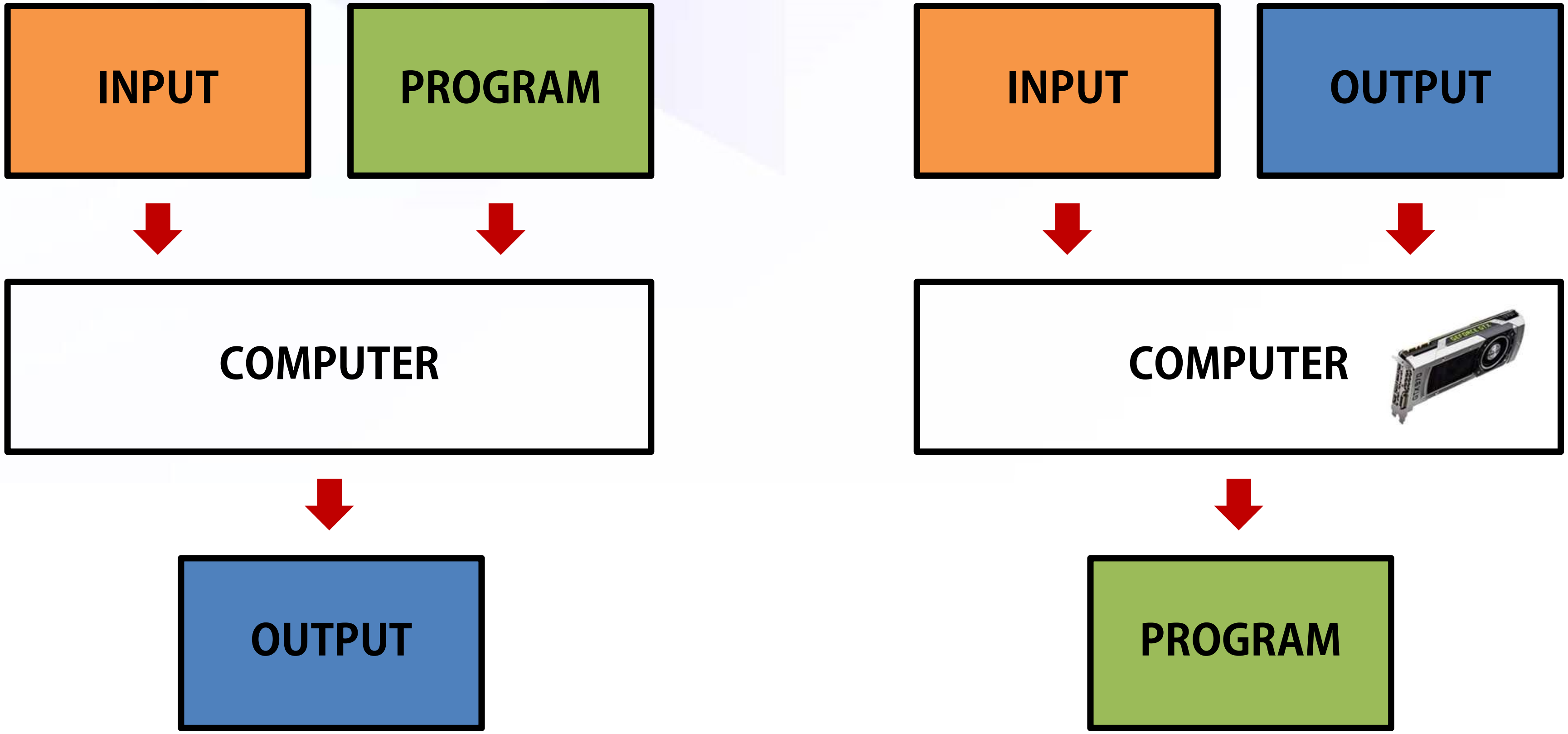
Paradigm shift

Simply GET a Program (automatically!) with your data

INPUT → OUTPUT



Data is ALL we need



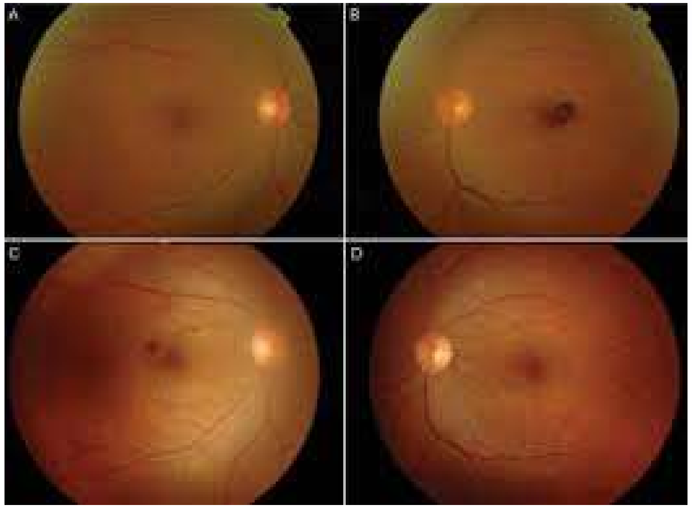

Google DeepMind
Challenge Match
8 - 15 March 2016
ALPHAGO
01:40:56
LEE SEDOL
01:35:33
AlphaGo
Lee Sedol

AI beats medical doctors

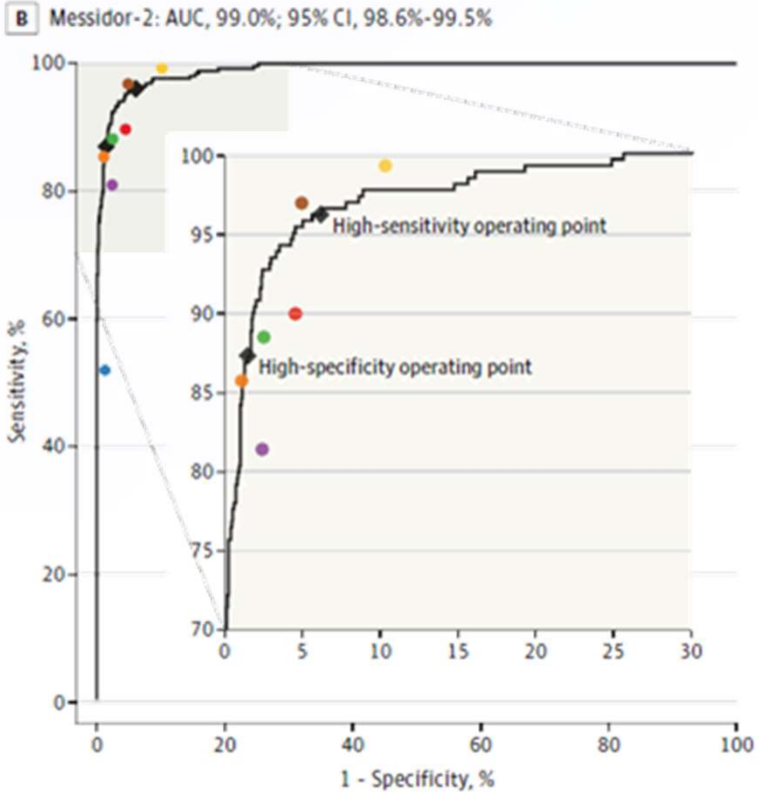
JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD



Seo *et al.*, Kor J Ophthalmol (2009)



Gulshan *et al.*, JAMA (2016)

Will AI replace doctors?

중앙일보 경제

경제일반 재테크 증권 부동산 IT/과학 경제칼럼

[단독] 의사가 하는 일 70% … 2030년엔 AI도 한다

[중앙일보] 입력 2017.08.29 02:30 수정 2017.08.29 03:15 | 종합 1면 지면보기▸

임미진 기자 최현주 기자 정선언

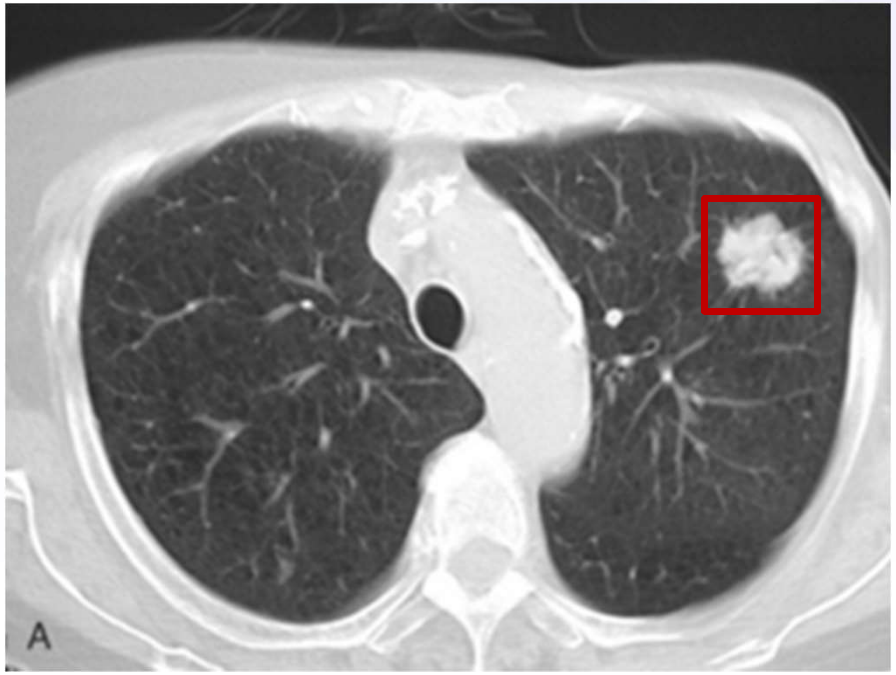
MGH 1811 MASSACHUSETTS GENERAL HOSPITAL

NVIDIA

NVIDIA, Massachusetts General Hospital Use Artificial Intelligence to Advance Radiology, Pathology, Genomics

Posted on December 25, 2016 by Aaron

What we expected



Malignancy: high
Spiculation: yes
Calcification: none
…

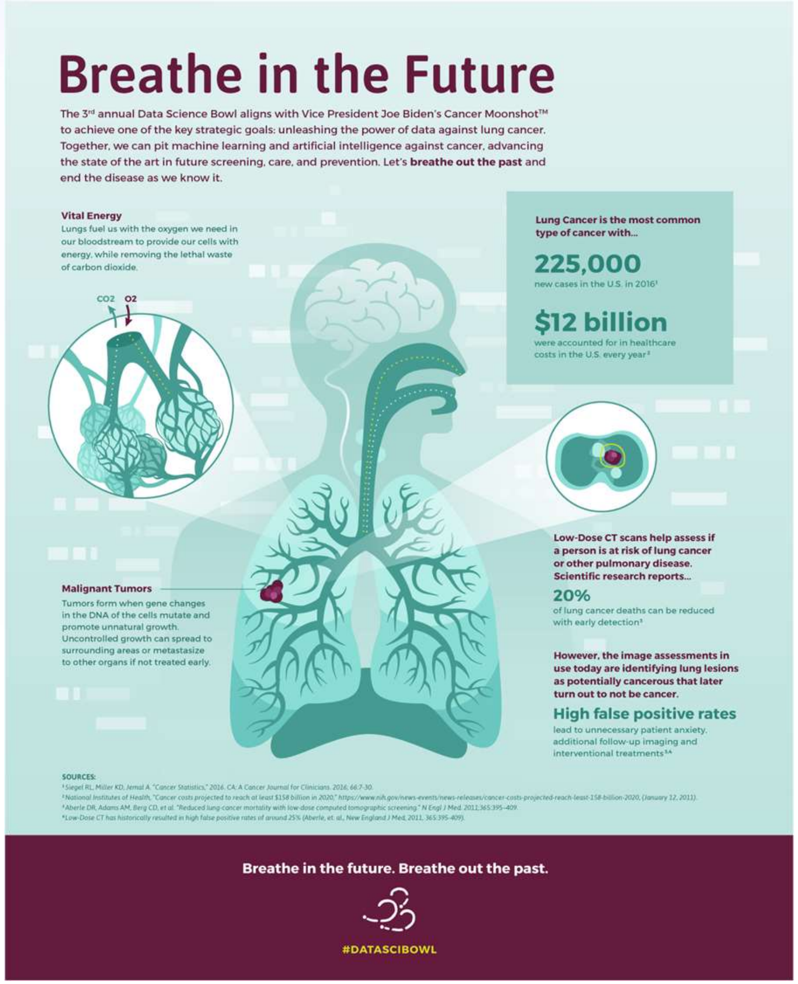
Probability of being diagnosed with cancer within a 12 month time frame is 85%

Deep Learning can do this !!

Data Science Bowl 2017

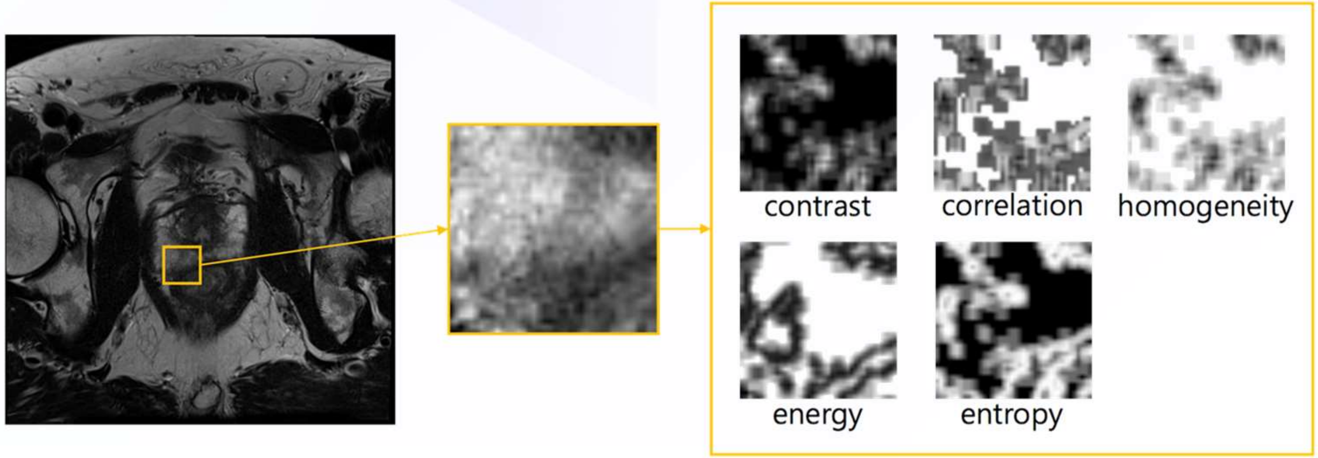
- Kaggle grand challenge
  - Make better lung cancer detection model
  - 2017.01.12 ~2017.04.12
  - Total **100,000 $** prizes

- Ranked in the top of 2.5 %
  - out of 1972 teams
  - Accuracy: around 77 %



ProstatEX2

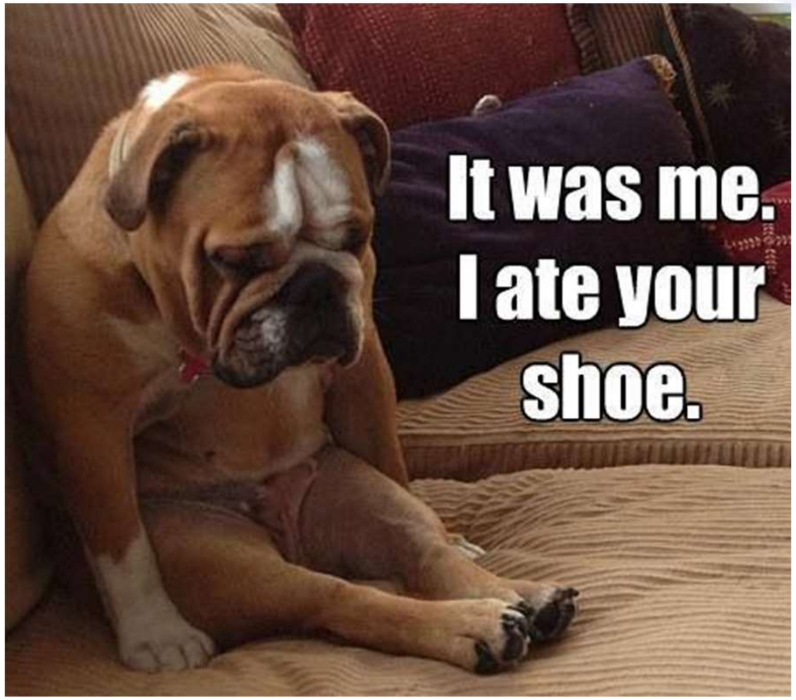
- Find quantitative multi-parametric MRI biomarkers for determination of Gleason Grade Group in prostate cancer



- Validation set test accuracy: Around 80% (w/ DCE sequence)
- Final test result: 0.1022 (quadratic weighted kappa, 0.2772 for 1st place team)
- Achieved 8th highest score (out of 143 participants)

Being a detective

- Prediction of being diagnosed w/ cancer within a year
  - ONLY with imaging diagnosis



It is a dog

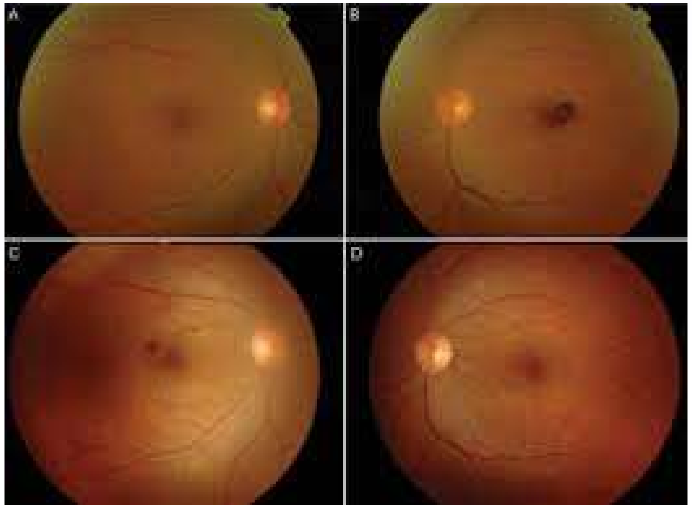
It is a dog, a bulldog. I can see dirt on the dog. (S)he will probably get sick if the dog mistakenly swallows something while (s)he rolled in a dirty spot.

AI can do what humans do

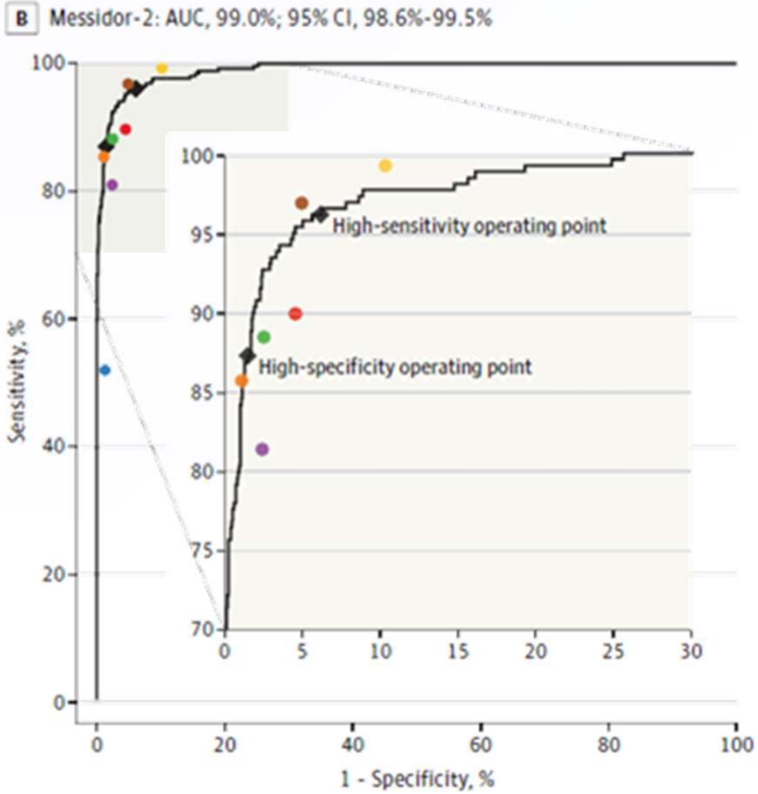
JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD



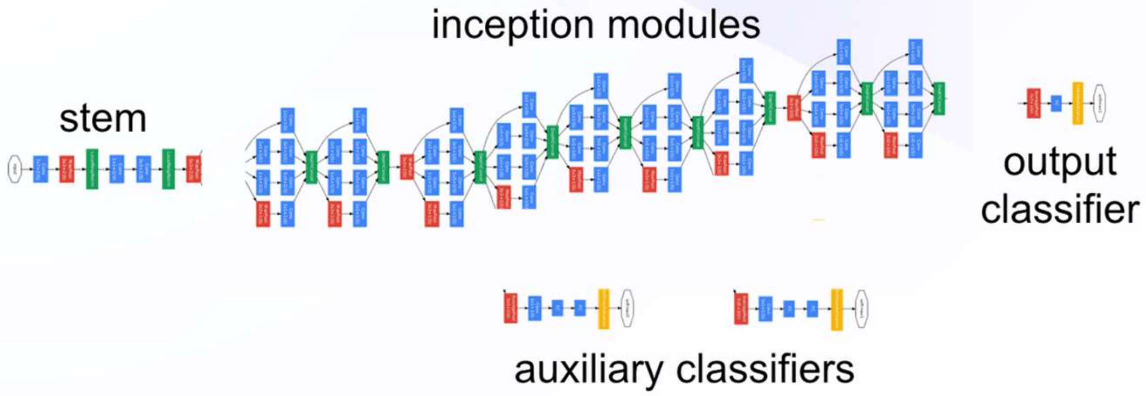
Seo *et al.*, Kor J Ophthalmol (2009)



Gulshan *et al.*, JAMA (2016)

Can we beat Google?

Technique?



DATA!!!

Characteristics	Development Data Set	EyePACS-1 Validation Data Set
No. of images	128 175	9963
No. of ophthalmologists	54	8
No. of grades per image	3-7	8
Grades per ophthalmologist, median (interquartile range)	2021 (304-8366)	8906 (8744-9360)

Dr. Hyunseok Min's PPT

Can we make Watson for oncology?

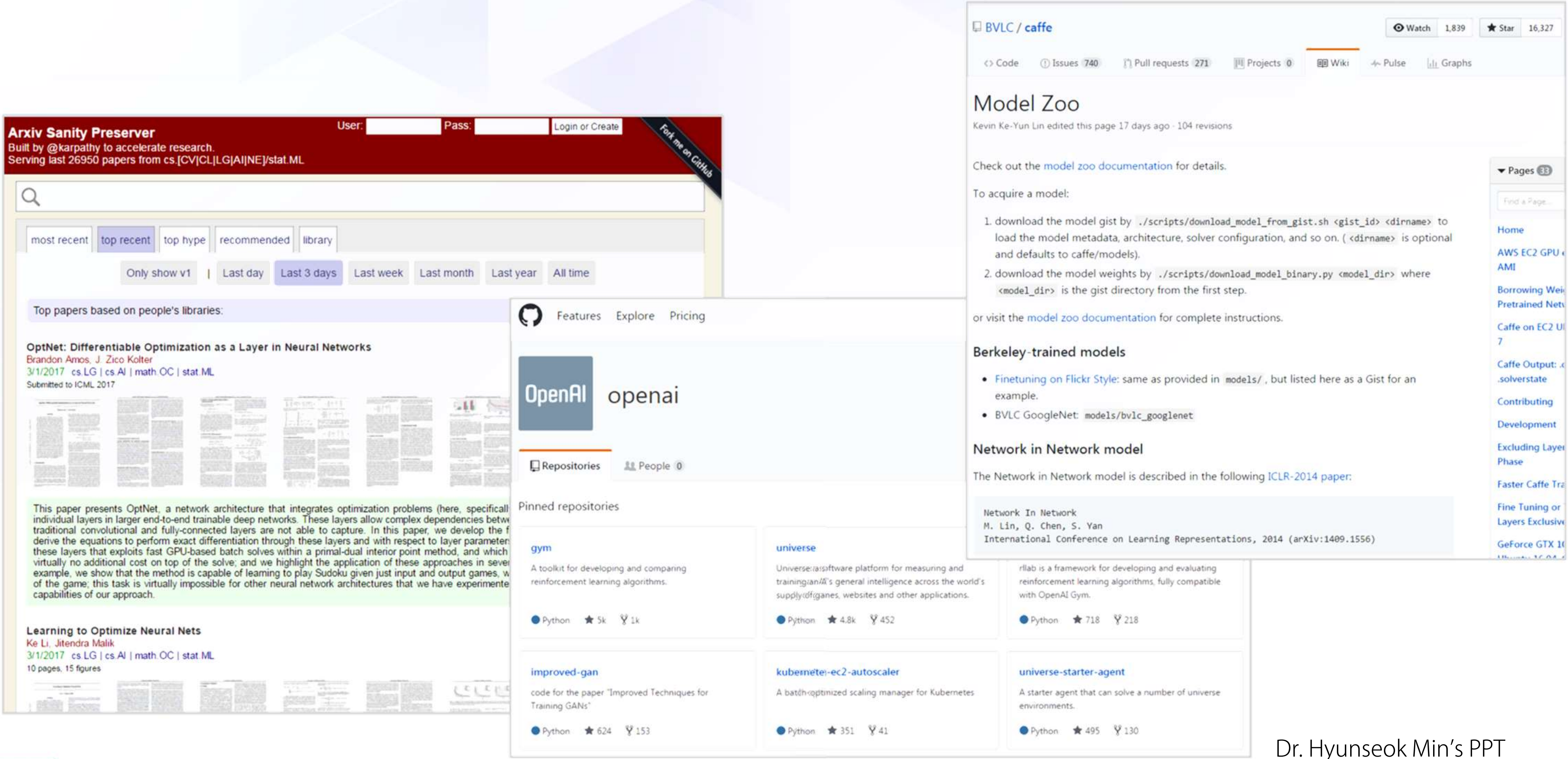
50억원 / year

News room > News releases >

IBM Closes Deal to Acquire Merge Healthcare

0.7 billion $ (약 8천억)

Open Innovation



Dr. Hyunseok Min's PPT

Again, DATA is all we need

Image data → **INPUT**

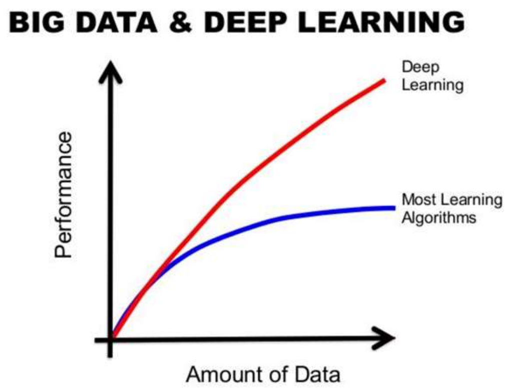
OUTPUT ← **Ground truth (Label)**

1300 patients

COMPUTER

PROGRAM

Feature scores of each solitary pulmonary nodules By 4 radiologists



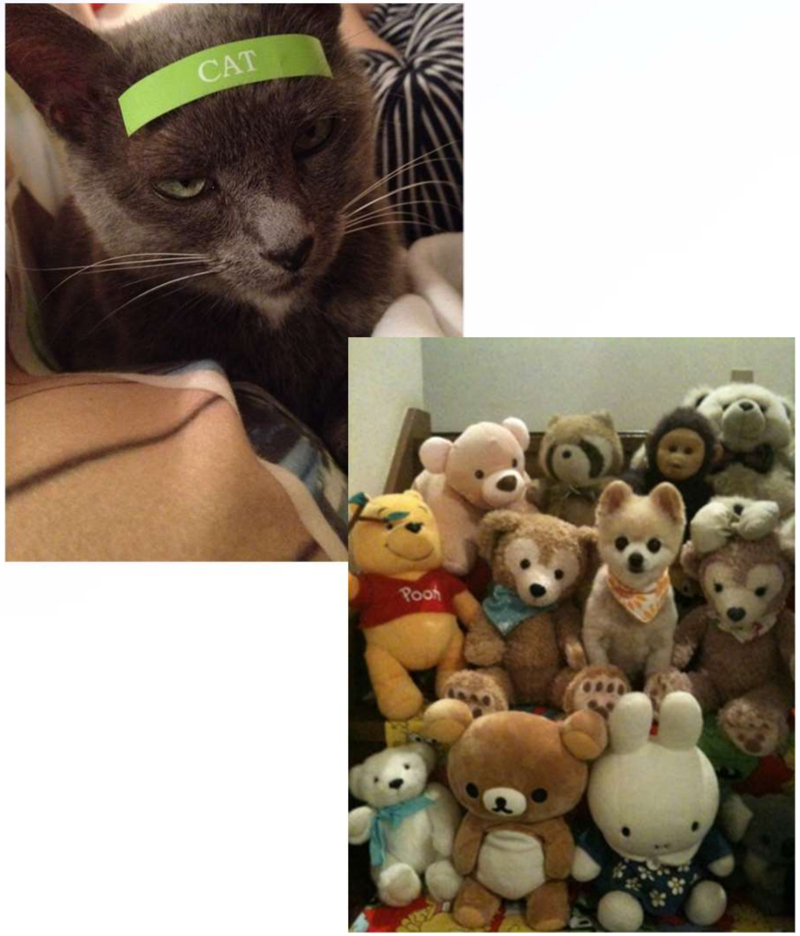
Lack of training data

Medical data

Labeled data

Cleaned data

Abnormal	Normal



Will AI replace doctors?

- **I will say NO**

- The value of high-quality data will rapidly increase
  - Only physicians can make high-quality (reliable) medical data

- AI extends the extent of expertize and gives new perspectives
  - w/ commoditization of deep learning tools (like MS words and excel)

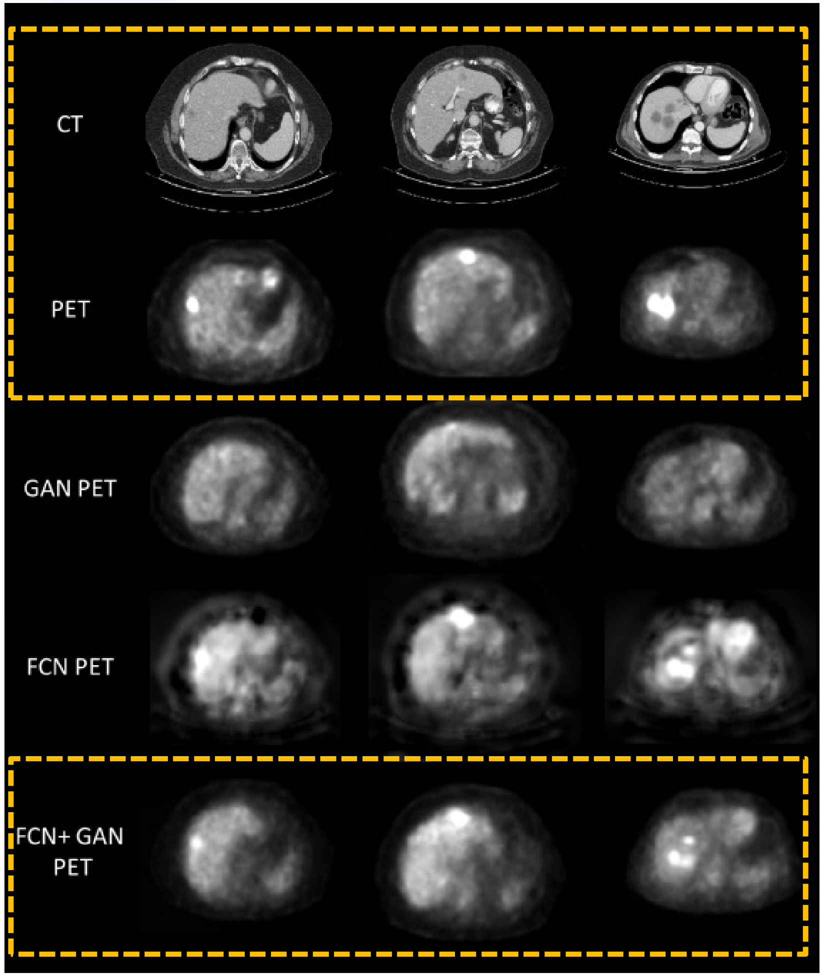
Structural Image → Functional Image ?!?

Virtual PET Images from CT Data Using Deep Convolutional Networks: Initial Results

Avi Ben-Cohen[1], Eyal Klang[2], Stephen P. Raskin[2], Michal Marianne Amitai[2], and Hayit Greenspan[1]

[1]Tel Aviv University, Faculty of Engineering, Department of Biomedical Engineering, Medical Image Processing Laboratory, Tel Aviv 69978, Israel
[2]Sheba Medical Center, Diagnostic Imaging Department, Abdominal Imaging Unit, affiliated to Sackler school of medicine Tel Aviv University, Tel Hashomer 52621, Israel

Paper accepted
@ SASHIMI2017 workshop, MICCAI 2017



AI has no responsibility

음악을 들으면 그에 해당하는 개인의 추억을 되새겨주는(영상이나 글로) 인공지능을 만들어봐도 재밌겠네요.
좋아요 · 답글 달기 · 1 · 7월 10일 오전 9:23 · 수정됨

만약 아내와 음악을 듣는데 옛노래에 연관된 이미지로 전여친 사진이 뜬다면?!?
좋아요 · 답글 달기 · 3 · 7월 10일 오전 9:47

인공지능은 책임을 지지 않습니다 ㅎㅎ
좋아요 · 답글 달기 · 7월 10일 오후 12:41

AI RI 지능정보기술연구원

Thank you for your attention

2017.10.31 NVIDIA Deep Learning Day

지능정보기술연구원 김휘영 PhD
astaria82@gmail.com