Machine Learning for Image Analysis and Treatment Planning

Lei Xing, PhD

Departments of Radiation Oncology & Electrical Engineering
Stanford University

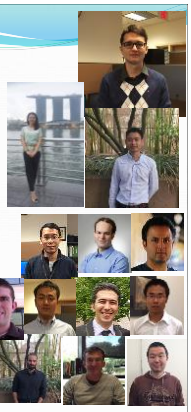
Stanford University
Department of Radiation Oncology
School of Medicine

Disclosures

- The Department of Radiation Oncology at Stanford University Hospital has a research agreement with Varian Medical Systems.
- Technology License Agreement with Varian. Dr. Lei Xing has received speakers honoraria from Varian.
- Dr. Lei Xing serves as advisory scientist in HuiyiHuiying Inc, MoreHealth and Zap Surgical.
- Research grants supports from NIH and Google

Acknowledgement

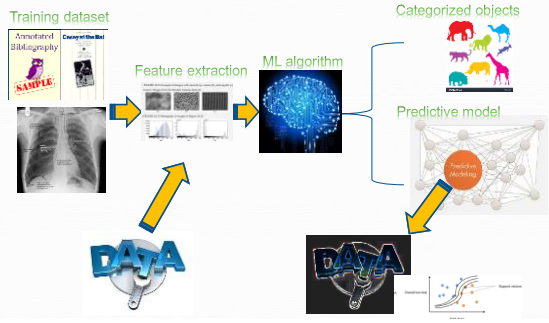
- Yixuan Yuan, Bulat Ibragimov, W. Qin, H. Liu, M. Korani, D. Li, K. Cheng, W. Zhao, C. Jenkins, S. Tzoumas, D. Vernekohl, S. Youselfi, , B. Ungan
- Y. Cui, J Wang, R. Li, P. Dong, B. Han, A. Koong, D. Chang, D. Toesca, S. Soltys
- Funding: NIH/NCI/NIBIB & Google



Outline

- Machine learning in medicine 101
- Image analysis & radiomics with machine learning
  - Image analysis in gastrointestinal tract.
  - Liver cancer imaging and analysis.
  - Brain tumor RT.
- Machine learning and autopiloted and/or knowledge-based treatment planning
- Clinical studies
- Future outlooks and trends

Machine learning 101



Go, a complex game popular in Asia, has frustrated the efforts of artificial-intelligence researchers for decades.

ARTIFICIAL INTELLIGENCE

Google masters Go

Deep-learning software excels at complex ancient board game.

BY ELIZABETH GIBNEY

reveals in research published in *Nature* on 27 January[1]. It also defeated its silicon-based famously beat grandmaster Garry Kasparov in 1997, was explicitly programmed to win at the

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 cancer, the most common human malignancy[1–3], is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. Deep convolutional neural networks (CNNs)[4,5] show potential for general and highly variable tasks across many fine-grained object categories[6–11]. Here we demonstrate classification of skin lesions using a single CNN, trained end-to-end from images directly, using only pixels and disease labels as inputs. We train a CNN using a dataset of 129,450 clinical images—two orders of magnitude larger than previous datasets[12]—consisting of 2,032 different diseases. We test its performance against 21 board-certified dermatologists on biopsy-proven clinical images with two critical binary classification use cases: keratinocyte carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi. The first case represents the identification of the most common cancers, the second represents the identification of the deadliest skin cancer. The CNN achieves performance on par with all tested experts across both tasks, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6.3 billion smartphone subscriptions will

images (for example, smartphone images) exhibit variability in factors such as zoom, angle and lighting, making classification substantially more challenging[12,13]. We overcome this challenge by using a data-driven approach—1.41 million pre-training and training images make classification robust to photographic variability. Many previous techniques require extensive preprocessing, lesion segmentation and extraction of domain-specific visual features before classification. By contrast, our system requires no hand-crafted features; it is trained end-to-end directly from image labels and raw pixels, with a single network for both photographic and dermoscopic images. The existing body of work uses small datasets of typically less than a thousand images of skin lesions[14,15,16,17,18,19], which, as a result, do not generalize well to new images. We demonstrate generalizable classification with a new dermatologist-labelled dataset of 129,450 clinical images, including 3,374 dermoscopy images.

Deep learning algorithms, powered by advances in computation and very large datasets[20], have recently been shown to exceed human performance in visual tasks such as playing Atari games[21], strategic board games like Go[22] and object recognition[4]. In this paper we outline the development of a CNN that matches the performance of dermatologists at three key diagnostic tasks: melanoma classification, melanoma classification using dermoscopy and carcinoma classification. We restrict the comparisons to image-based classification.

We utilize a GoogleNet Inception v3 CNN architecture[9] that was pre-trained on approximately 1.28 million images (1,000 object categories)

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]

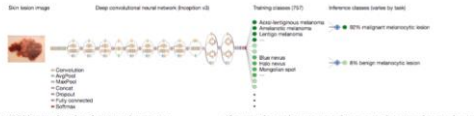


Figure 1 | Deep CNN layout. Our classification technique is a deep CNN. Data flow is from left to right: an image of a skin lesion (for example, melanoma) is sequentially warped into a probability distribution over clinical classes of skin disease using Google Inception v3 CNN architecture pretrained on the ImageNet dataset (1.28 million images over 1,000 generic object classes) and fine-tuned on our own dataset of 129,450 skin lesions comprising 2,032 different diseases. The 757 training classes are defined using a novel taxonomy of skin disease and a partitioning algorithm that maps diseases into training classes (for example, acrolentiginous melanoma, amelanotic melanoma, lentigo melanoma). Inference classes are more general and are composed of one or more training classes (for example, malignant melanocytic lesions—the class of melanomas). The probability of an inference class is calculated by summing the probabilities of the training classes according to taxonomy structure (see Methods). Inception v3 CNN architecture reprinted from https://research.googleblog.com/2016/03/train-your-own-image-classifier-with.html

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]

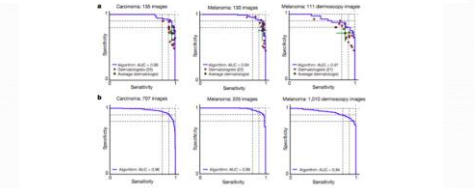


Figure 3 | Skin cancer classification performance of the CNN and dermatologists. a, The deep learning CNN outperforms the average of the dermatologists at skin cancer classification using photographic and dermoscopic images. Our CNN is tested against at least 21 dermatologists at keratinocyte carcinoma and melanoma recognition. For each test, previously unseen, biopsy-proven images of lesions are displayed, and dermatologists are asked if they would: biopsy/treat the lesion or reassure the patient. Sensitivity, the true positive rate, and specificity, the true negative rate, measure performance. A dermatologist outputs a single prediction per image and is thus represented by a single red point. The green points are the average of the dermatologists for each task, with error bars denoting one standard deviation (calculated from $n = 25$, 22 and 21 tested dermatologists for keratinocyte carcinoma, melanoma and melanoma under dermoscopy, respectively). The CNN outputs a malignancy probability P per image. We fix a threshold probability t such that the prediction $\hat{y}$ for any image is $\hat{y} = P \geq t$, and the blue curve is drawn by sweeping t in the interval 0–1. The AUC is the CNN's measure of performance, with a maximum value of 1. The CNN achieves superior performance to a dermatologist if the sensitivity–specificity point of the dermatologist lies below the blue curve, which most do. Epidermal test: 65 keratinocyte carcinomas and 70 benign seborrheic keratoses. Melanocytic test: 33 malignant melanomas and 97 benign nevi. A second melanocytic test using dermoscopic images is displayed for comparison: 71 malignant and 40 benign. The slight performance decrease reflects differences in the difficulty of the images tested rather than the diagnostic accuracies of visual versus dermoscopic examination. **b**, The deep learning CNN exhibits reliable cancer classification when tested on a larger dataset. We tested the CNN on more images to demonstrate robust and reliable cancer classification. The CNN's curves are smoother owing to the larger test set.

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

Editorial pages 2366 and 2368
Supplemental content

IMPORTANCE Deep learning is a family of computational methods that allow an algorithm to program itself by learning from a large set of examples that demonstrate the desired behavior, removing the need to specify rules explicitly. Application of these methods to medical imaging requires further assessment and validation.

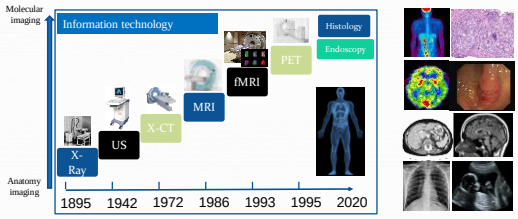
OBJECTIVE To apply deep learning to create an algorithm for automated detection of diabetic retinopathy and diabetic macular edema in retinal fundus photographs.

DESIGN AND SETTING A specific type of neural network optimized for image classification called a deep convolutional neural network was trained using a retrospective development data set of 128 175 retinal images, which were graded 3 to 7 times for diabetic retinopathy, diabetic macular edema, and image gradability by a panel of 54 US licensed ophthalmologists and ophthalmology senior residents between May and December 2015. The resultant algorithm was validated in January and February 2016 using 2 separate data sets, both graded by at least 7 US board-certified ophthalmologists with high intragrader consistency.

EXPOSURE Deep learning-trained algorithm.

ML for Medical Image Analysis

- Images are data!
- Imaging is one of the first choices for clinical diagnosis
- 70% clinical decisions depend on medical images



Clinical Application

Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks

Bulat Ibragimov[a] and Lei Xing
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305, USA

(Received 2 May 2016; revised 31 October 2016; accepted for publication 23 November 2016; published 13 February 2017)

Purpose: Accurate segmentation of organs-at-risks (OARs) is the key step for efficient planning of radiation therapy for head and neck (HaN) cancer treatment. In the work, we proposed the first deep learning-based algorithm, for segmentation of OARs in HaN CT images, and compared its performance against state-of-the-art automated segmentation algorithms, commercial software, and interobserver variability.

Methods: Convolutional neural networks (CNNs)—a concept from the field of deep learning—were used to study consistent intensity patterns of OARs from training CT images and to segment the OAR in a previously unseen test CT image. For CNN training, we extracted a representative number of positive intensity patches around voxels that belong to the OAR of interest in training CT images, and negative intensity patches around voxels that belong to the surrounding structures. These patches then passed through a sequence of CNN layers that captured local image features such as corners, endpoints, and edges, and combined them into more complex high-order features that can efficiently describe the OAR. The trai
image where the correspond
fication results by using M
component of the smoothed
to remove cavities of the co
Results: The performance
glands, submandibular glan
CT images. The obtained s

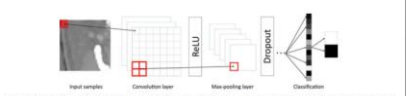


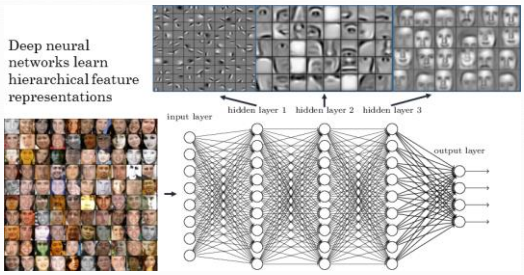
FIG. 1. A schematic illustration of the convolutional neural network architecture. Three orthogonal cross-sections around target voxel define the input of the network that consists of three stacks of convolution, ReLU, max-pooling layer, and dropout layers, fully connected and softmax layers. [Color figure can be viewed at wileyonlinelibrary.com]

Medical Physics, 44 (2), February 2017

14

Convolutional neural networks for segmentation

Deep neural networks learn hierarchical feature representations



Taken from here: http://www.rsipvision.com/exploring-deep-learning/

Stanford MEDICINE Radiation Oncology Medical Physics

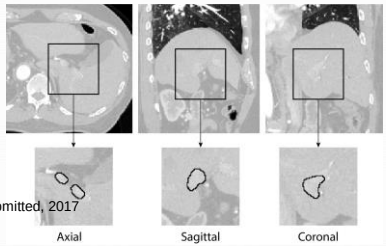
Stanford ENGINEERING Mechanical Engineering

Clinical Application

Image database

72 pre-treatment CT images:

- PV with contrast
- Stents
- Tumors are close to PV



B Ibragimov et al, submitted, 2017

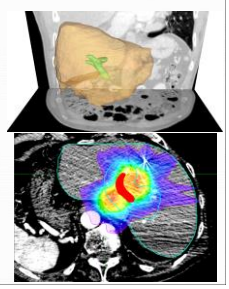
16

Central liver toxicity - B. Ibrimbrov, D. Toesca, A Koong, D. Chang, L XIng

Irradiation of hepatobiliary tract will likely result in central liver toxicity if the isotropic 15mm expansion of portal vein (PV) receives:

- $V_{BED10}30 > 45$ cc
- $V_{BED10}40 > 37$ cc

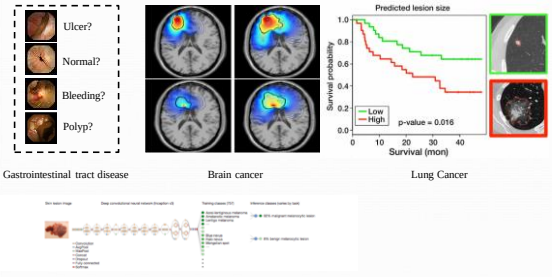
Can we predict such toxicity without manually annotating portal vein?



Stanford MEDICINE Radiation Oncology Medical Physics

Clinical Applications

Diagnosis | Treatment planning | Prognosis



Gastrointestinal tract disease | Brain cancer | Lung Cancer

ML for Medical Image Analysis

Variability
- **M**agnetic **R**esonance **I**maging
- Computed **T**omography
- Ultrasound
- Endoscope images

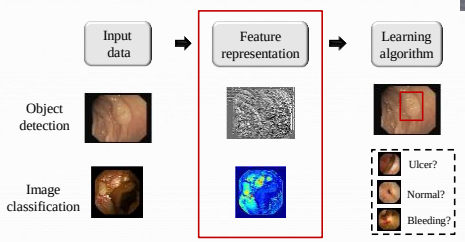
Large scale
- MRI images 1000+
- CT images 1000+
- Endoscope images 50K+

Medical Image + ML for substantially improved image analysis

Heterogeneity
- Different appearances
- Characteristics vary in different modalities

Subjective
- Clinicians' experience
- Time-consuming and tedious

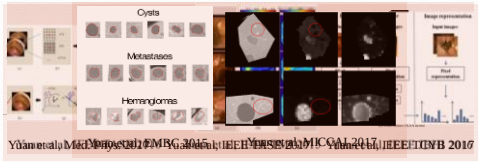
- Machine learning with hand-crafted features



Research on Medical Image Analysis

Integrative Biomedical Image Analysis

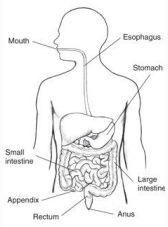
- Diseases in Gastrointestinal Tract
- Liver Cancers with CT images
- Prostate & Brain Cancers with MRI images

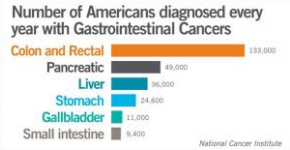


Yuan et al., Med Phys 2017 | Yuan et al., IEEE JBHI 2017 | Yuan et al., MICCAI 2017 | Yuan et al., IEEE TCYB 2017

Diseases in Gastrointestinal Tract

- Importance
  - Gastrointestinal (GI) tract - 30 feet long structure
  - 2nd commonest cancer





Wireless Capsule Endoscopy (WCE)

- Wireless capsule endoscopy
  - Introduced by Iddan et al. in 2000
  - Approved by the U.S Food & Drug Administration in 2001
  - 11x26mm

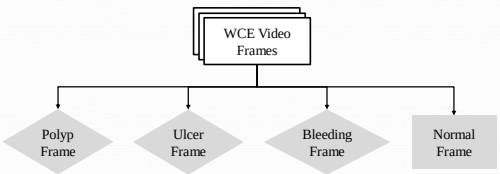
- Examination procedure
  - Swallowed by patients
  - Propelled by peristalsis
  - Send images to data-recording device
  - Downloaded for reviewing

Illustration of WCE

- Enable direct, non-invasive and visual examination of the GI tract

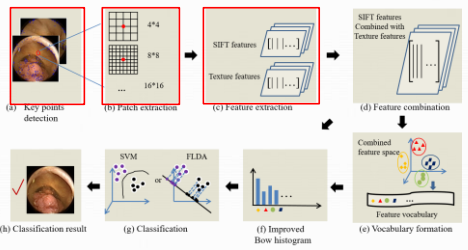
Research on WCE

- To automatically recognize abnormality for clinicians



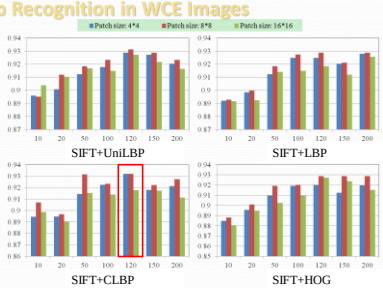
Polyp Recognition in WCE Images

- Proposed method: improved bag of words for polyp detection



- Experiment results

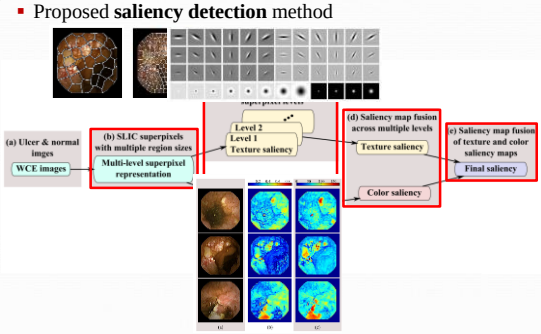
Polyp Recognition in WCE Images



Yuan et al. "Improved Bag of Feature for Automatic Polyp Detection," IEEE Transactions on Automation Science and Engineering (TASE), 2016.

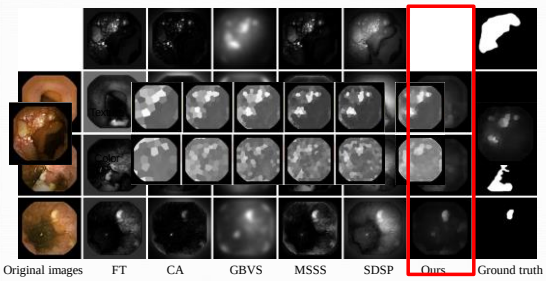
Ulcer Recognition in WCE Images

- Proposed **saliency detection** method



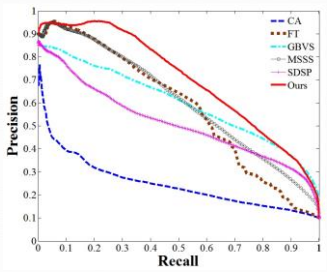
Ulcer Recognition in WCE Images

- Experiment results for **saliency detection**



Ulcer Recognition in WCE Images

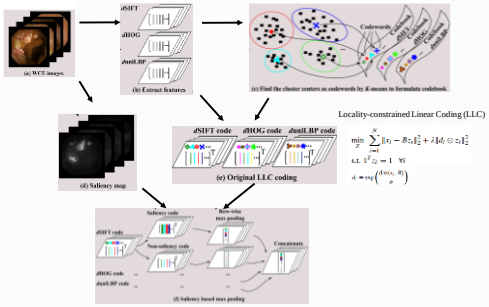
- Experiment results for **saliency detection**



Stanford University

Ulcer Recognition in WCE Images

- Proposed **ulcer recognition** method



Ulcer Recognition in WCE Images

- Experiment results for **ulcer recognition**

Comparison with state-of-the-art methods

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
Li et al., 2009	89.49± 0.12	87.06± 0.38	91.91± 0.13
Charisis et al., 2013	90.74± 0.07	86.62± 0.2	94.85± 0.15
Eid et al., 2013	85.44± 0.18	86.03± 0.65	84.84± 0.4
Yu et al., 2014	82.35± 0.93	91.18± 0.05	73.53± 3.89
Ours	**92.65± 1.23**	**94.12± 2.47**	**91.18± 0.91**

Yuan et al. "Saliency based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis," IEEE Transaction on Medical Imaging (TMI), 2015.

Multi-abnormalities Classification

- Experiment results

Comparison with state-of-the-art methods

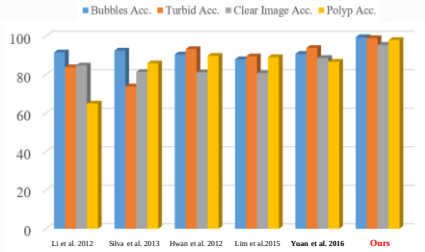
Method	Overall Accuracy	Bleeding Accuracy	Polyp Accuracy	Ulcer Accuracy	Normal Accuracy
Hwang et al. 2013	83.69 ± 2.42	90.95 ± 2.60	81.33 ± 4.71	78.33 ± 1.08	83.56 ± 3.48
Nawarathna et al. 2015	87.62 ± 1.24	93.33± 1.23	84.00 ± 2.02	83.33 ± 0.98	89.33 ± 2.03
Yao et al. 2016	89.46 ± 1.08	95.24 ± 1.36	87.78 ±1.28	83.33 ± 0.83	90.67 ± 1.06
Ours	**90.78 ± 0.53**	**97.14 ± 0.21**	**86.84 ± 0.61**	**86.67 ± 0.82**	**92.11 ± 0.28**

Yuan et al. "Discriminative Joint-feature Topic Models with Dual Constraints for WCE Classification," Accepted by IEEE Transaction on Cybernetics (TCYE), 2017.

Stanford University

Polyp Recognition with Deep Learning

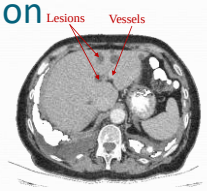
- Experiment results



Yuan et al. "Deep Learning for Polyp Recognition in Wireless Capsule Endoscopy Images," Medical Physics, 2017.

Stanford University

Liver Lesion Detection

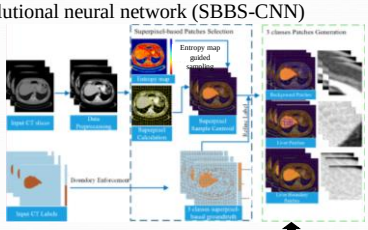


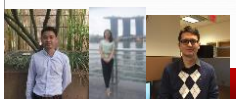
- Importance
  - Important functions
  - 4th common lesion in the world
  - Computed Tomography (CT) scans
- Challenges
  - Low imaging resolution (≈1 × 1 × 1mm³)
  - Complex structure information
  - Some lesions are small
  - Vessels exist in the liver
- Our solution
  - Segment the liver → detect liver lesion

Stanford University

Liver Segmentation

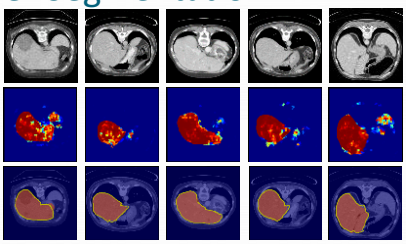
- Proposed method: superpixel-based boundary-sensitive convolutional neural network (SBBS-CNN)





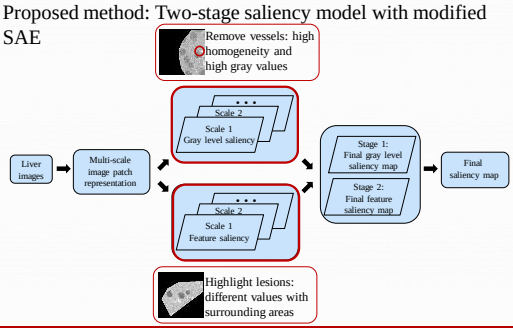
- Experiment results

Liver Segmentation



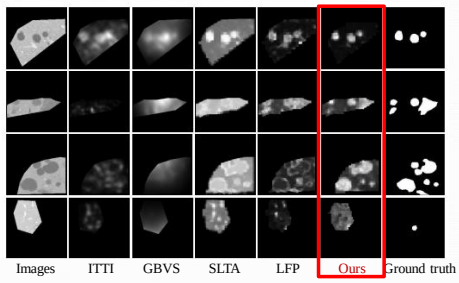
Liver Lesion Detection

- Proposed method: Two-stage saliency model with modified SAE



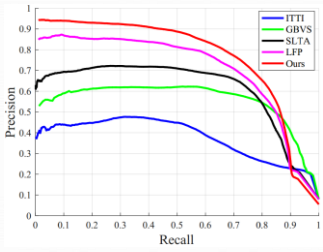
Liver Lesion Detection

- Experiment results



Liver Lesion Detection

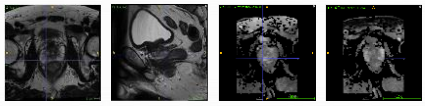
- Experiment results



Yixuan Yuan, et al. " Liver Lesion Detection based on Two-Stage Saliency Model with Modified Sparse Autoencoder," Accepted by MICCAI 2017

Prostate Cancer Classification

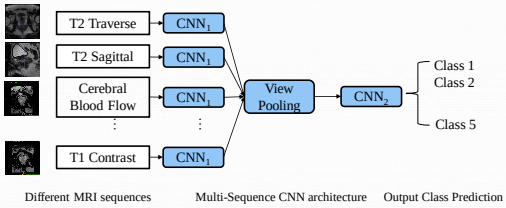
- Importance
  - Different cancer levels (gleason score) lead to different therapy
  - Reduce the core needle biopsy
- Modality for diagnosis
  - Magnetic Resonance Imaging (MRI)

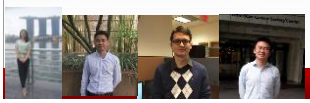


T2-weighted images (transaxial) | T2-weighted images (sagittal) | **A**pparent **D**iffusion **C**oefficient images | T1-weighted Contrast images

Prostate Cancer Classification

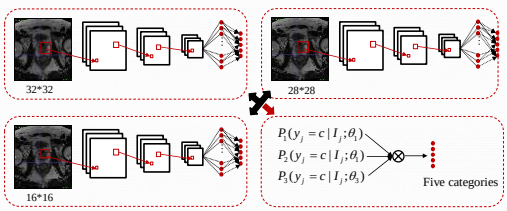
- Multi-sequence CNN architecture





Prostate Cancer Classification

- Multi-scale CNN



Prostate Cancer Classification

- Experiment results

Comparison with different methods

Method	Accuracy	Precision	Sensitivity
Traditional features	83.69 ± 2.42	90.95 ± 2.60	81.33 ± 4.71
SAE	87.62 ± 1.24	93.33± 1.23	84.00 ± 2.02
Multi-scale	89.46 ± 1.08	95.24 ± 1.36	87.78 ±1.28
Multi-sequence	90.78 ± 0.53	97.14 ± 0.21	86.84 ± 0.61

Yixuan Yuan, et al. AAPM 2017 Challenge

Radiology

ORIGINAL RESEARCH ■ NEURORADIOLOGY

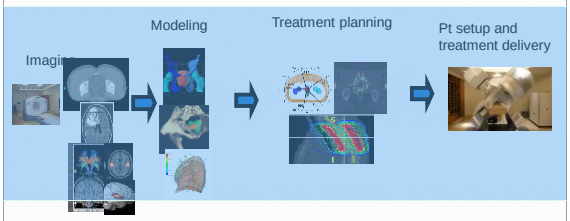
Prognostic Imaging Biomarkers in Glioblastoma: Development and Independent Validation on the Basis of Multiregion and Quantitative Analysis of MR Images[1]

Yi Cui, PhD
Khin Khin Tha, MD, PhD
Shunsuke Terasaka, MD, PhD
Shigeru Yamaguchi, MD, PhD
Jeff Wang, BA
Kohsuke Kudo, MD, PhD
Lei Xing, PhD
Hiroki Shirato, MD, PhD
Rujiang Li, PhD

Purpose: To develop and independently validate prognostic imaging biomarkers for predicting survival in patients with glioblastoma on the basis of multiregion quantitative image analysis.

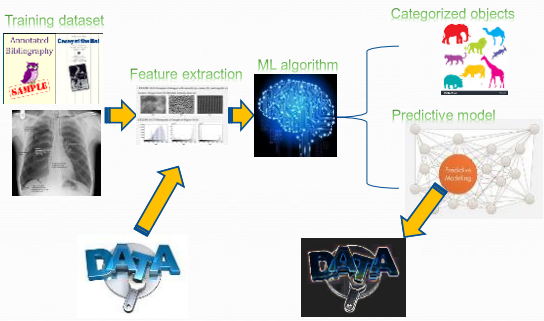
Materials and Methods: This retrospective study was approved by the local institutional review board, and informed consent was waived. A total of 79 patients from two independent cohorts were included. The discovery and validation cohorts consisted of 46 and 33 patients with glioblastoma from the Cancer Imaging Archive (TCIA) and the local institution, respectively.

Radiation Therapy workflow



- Automation
- Artificial intelligence
  - Data, imaging, image guidance & integration

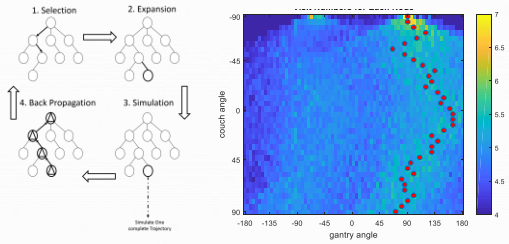
Machine learning 101



Machine learning 101

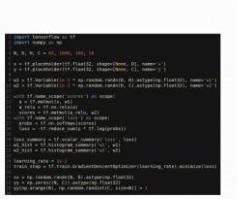
- Supervised learning
- Unsupervised learning
- Reinforcement learning

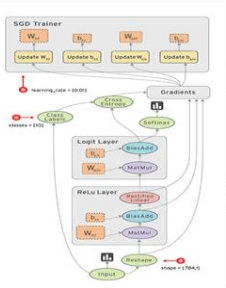
Artificial Intelligence (AI)-Based Non-Coplanar Rotational Arc Trajectory Design



Work supported by a Faulty Research Award from Google Inc.

TensorFlow framework for machine learning (Google, 2011)





TensorFlow

- Google open-sourced API (Python, C++)
  - Expressing large-scale machine learning algorithms
  - Implementation for executing such algorithms
- Heterogeneous execution (E.G., mobile phones, CPU, multi-GPU)
- Flexible modeling and parallelism for deep neural networks
- Powerful visualization tool (Tensorboard)
  - Tracks the graph evolution and model statistics

Abadi, Martın, et al. "Tensorflow: Large-scale machine learning on heterogeneous distri… *arXiv preprint arXiv:1603.04467* (2016).

Stanford MEDICINE | Radiation Oncology Medical Physics

Stanford Mechanical Engi… ERING

Future Work

Medical Image Analysis
- Feature extraction with domain knowledge
- Multi-modalities data analysis
- Deep learning application and modification

Biomedical Informatics
- High-level semantic feature extraction
- Feature fusion with image features
- RNN application

Medical Videos
- IGRT
- 3D reconstruction
- Surgical navigation
- Precision localization / Medical SLAM

Stanford University

Future Work

- Big imaging data in medicine