

Learning From Prior, Online/Incremental Learning

[Spring 2020 CS-8395 Deep Learning in Medical Image Computing]

Instructor: Yuankai Huo, Ph.D.
Department of Electrical Engineering and Computer Science
Vanderbilt University

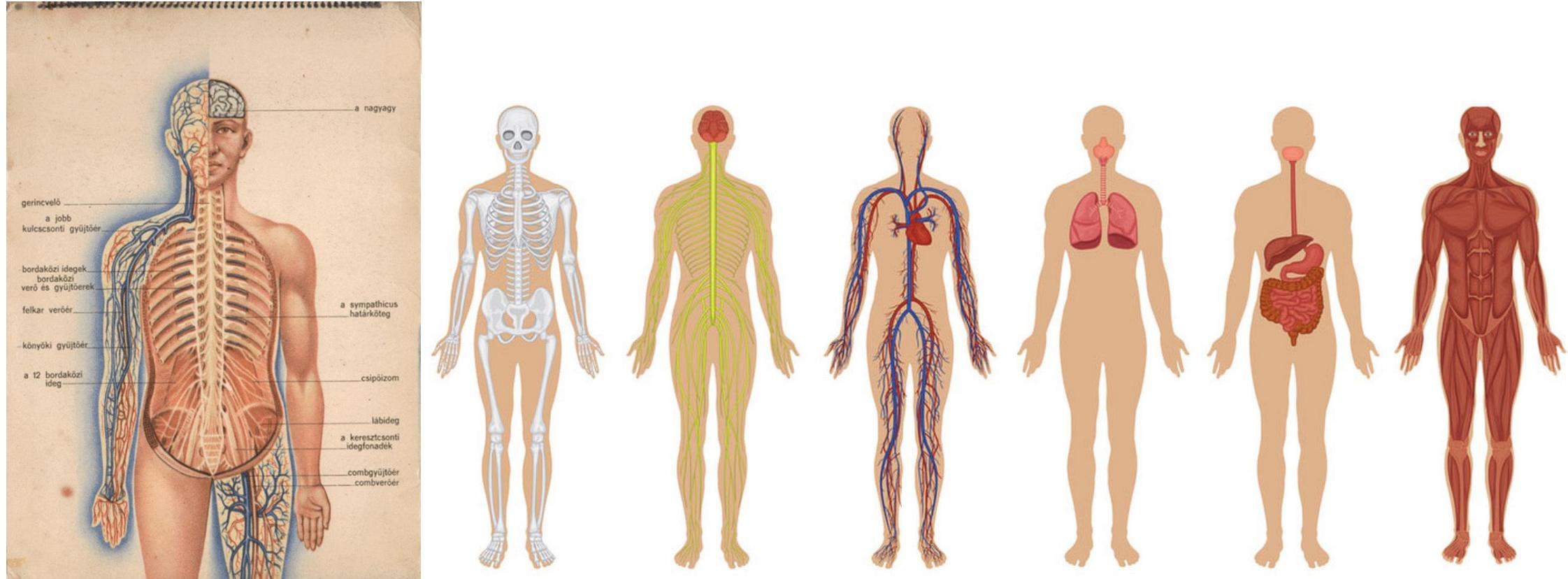
Advantages in Medical Imaging

Strong Prior

Great Traditional Methods

Self-supervised Learning

Prior



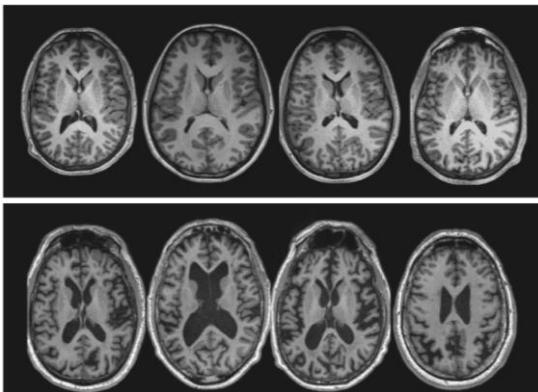
<https://giphy.com/gifs/health-human-body-anatomical-models-qqCsa4ARr6fRe>

<https://www.livescience.com/37009-human-body.html>

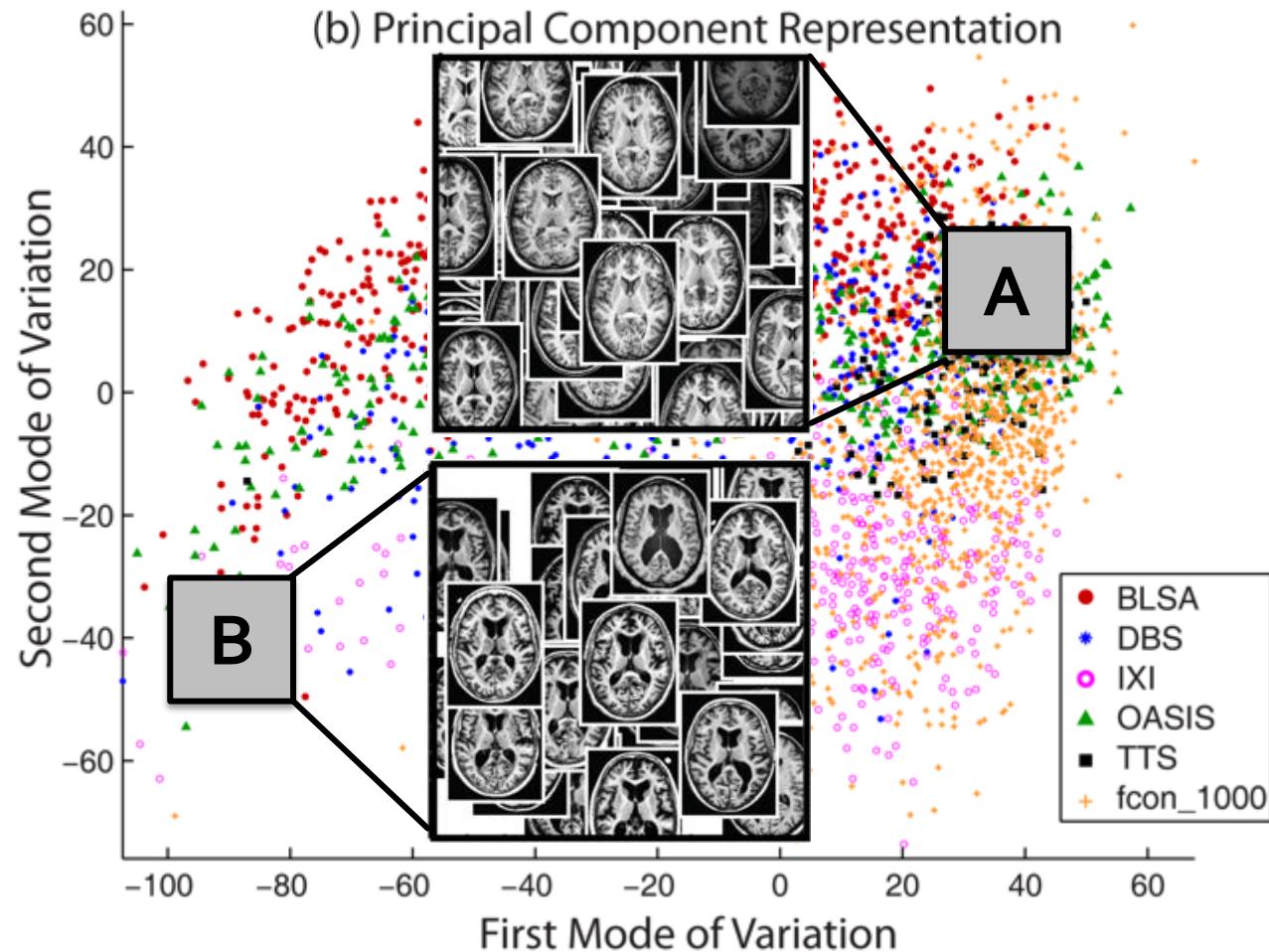
Incorporate strong prior



http://www.diegoantognini.com/projects/image_classification/



<http://www.mahoningvalleylanes.com/2425/car-picture-for-kids.html>



Accurate Detection of Inner Ears in Head CTs Using a Deep Volume-to-Volume Regression Network with False Positive Suppression and a Shape-Based Constraint

Dongqing Zhang^(✉), Jianing Wang, Jack H. Noble,
and Benoit M. Dawant

Department of Electrical Engineering and Computer Science,
Vanderbilt University, Nashville, TN 37235, USA
dongqing.zhang@vanderbilt.edu

Abstract. Cochlear implants (CIs) are neural prosthetics which are used to treat patients with hearing loss. CIs use an array of electrodes which are surgically inserted into the cochlea to stimulate the auditory nerve endings. After surgery, CIs need to be programmed. Studies have shown that the spatial relationship between the intra-cochlear anatomy and electrodes derived from medical images can guide CI programming and lead to significant improvement in hearing outcomes. However, clinical head CT images are usually obtained from scanners of different brands with different protocols. The field of view thus varies greatly and visual inspection is needed to document their content prior to applying algorithms for electrode localization and intra-cochlear anatomy segmentation. In this work, to determine the presence/absence of inner ears and to accurately localize them in head CTs, we use a volume-to-volume convolutional neural network which can be trained end-to-end to map a raw CT volume to probability maps which indicate inner ear positions. We incorporate a false positive suppression strategy in training and apply a shape-based constraint. We achieve a labeling accuracy of 98.59% and a localization error of 2.45 mm. The localization error is significantly smaller than a random forest-based approach that has been proposed recently to perform the same task.

Keywords: Cochlear implants · Landmark localization · 3D U-Net

Keywords: Cochlear implants · Landmark localization · 3D U-Net

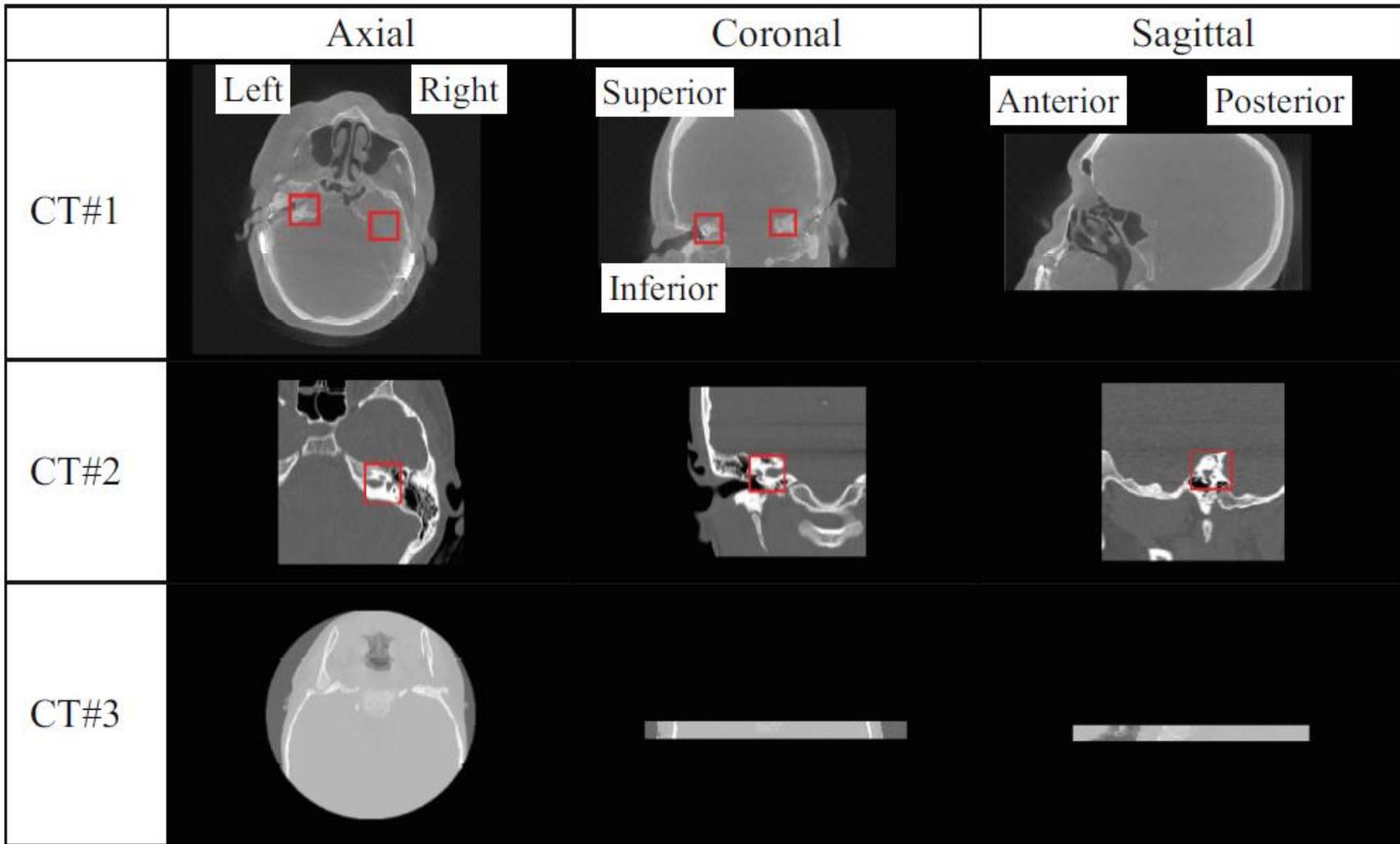


Fig. 1. Three exemplar CTs from our dataset. The inner ears are shown in red boxes if they are present in the volume and visible in the slice

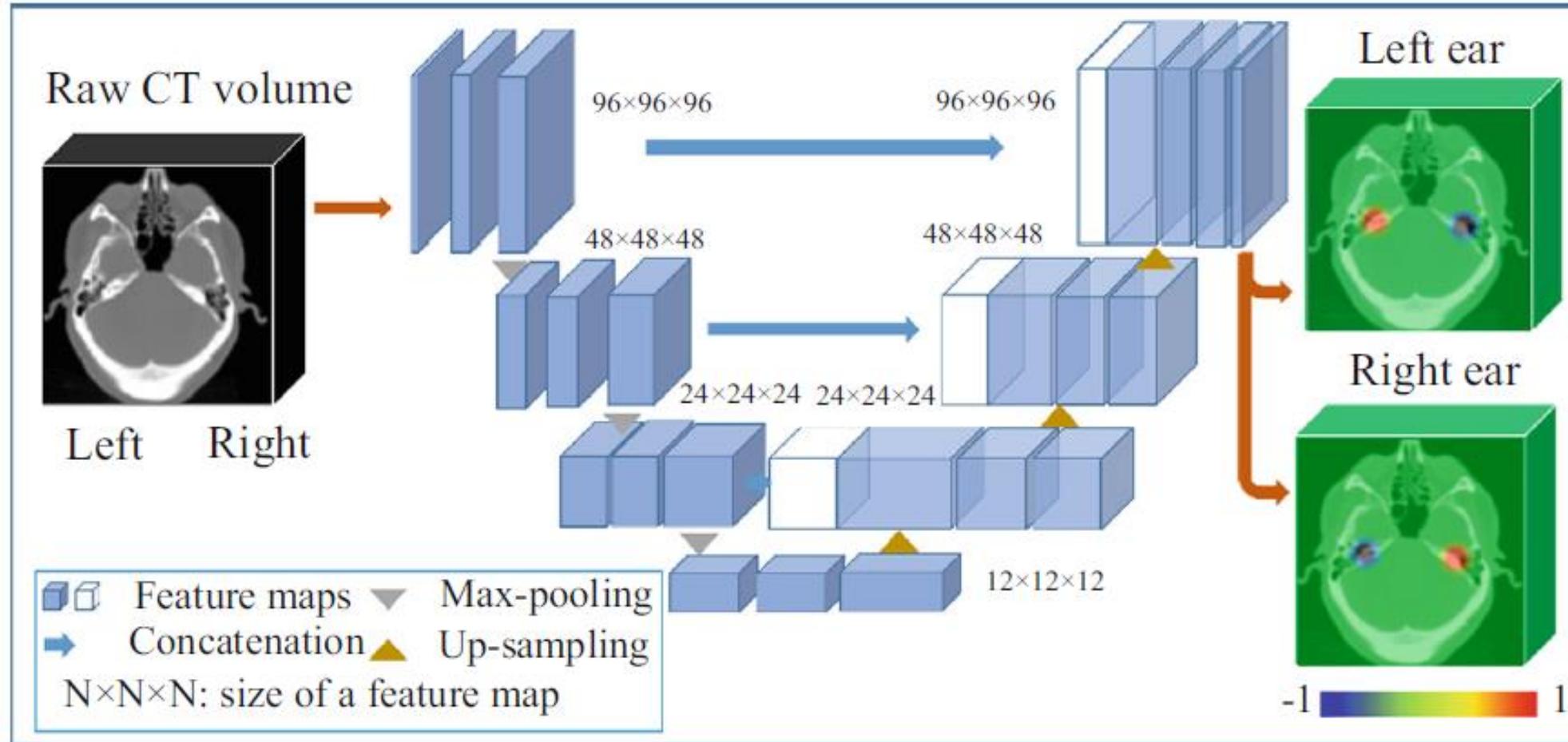


Fig. 2. Network architecture, input, and output for our approach

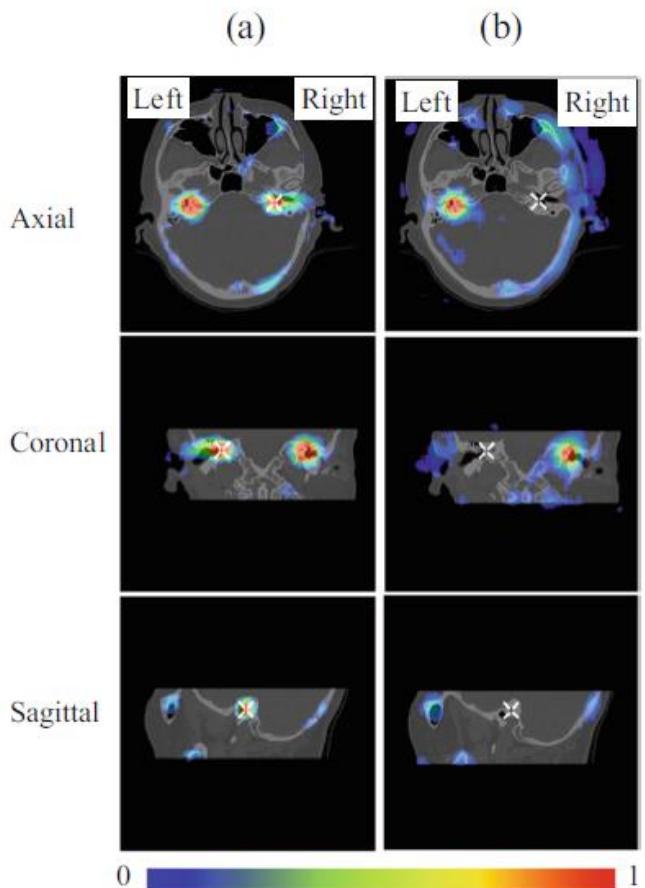


Fig. 3. The response maps of an input image containing a left ear. Column (a): before applying false positive suppression: the response at the right ear is also very high, (b): after applying false positive suppression, the false positive is eliminated (the location of the right ear is marked).

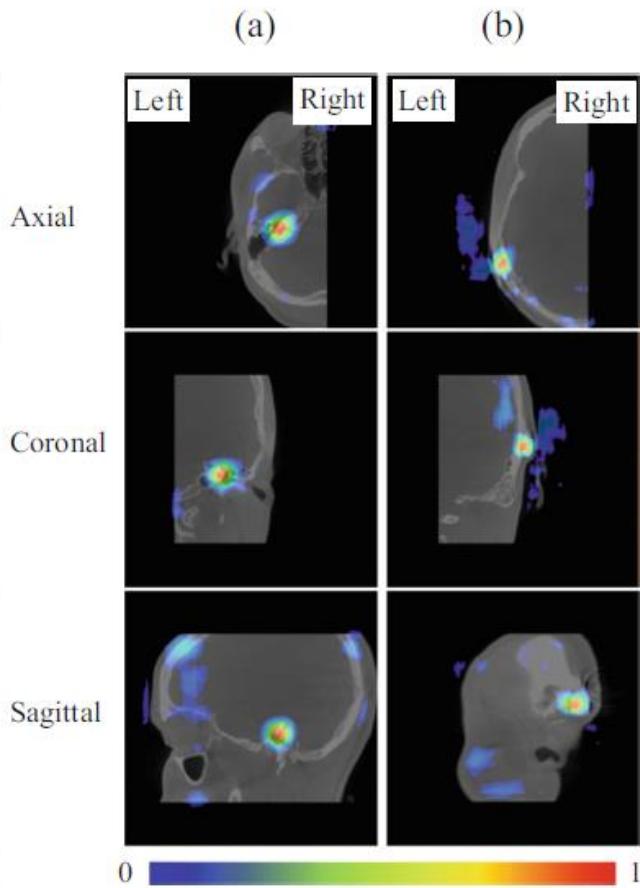


Fig. 4. The response maps of an input image that includes the left half of the head. Column (a): the response map associated with the left ear, (b): the response map associated with the right ear. The response at the location of the CI transmitter is so high that it is detected as an ear. It can be eliminated by the shape-based constraint.

Advantages in Medical Imaging

Strong Prior

Great Traditional Methods

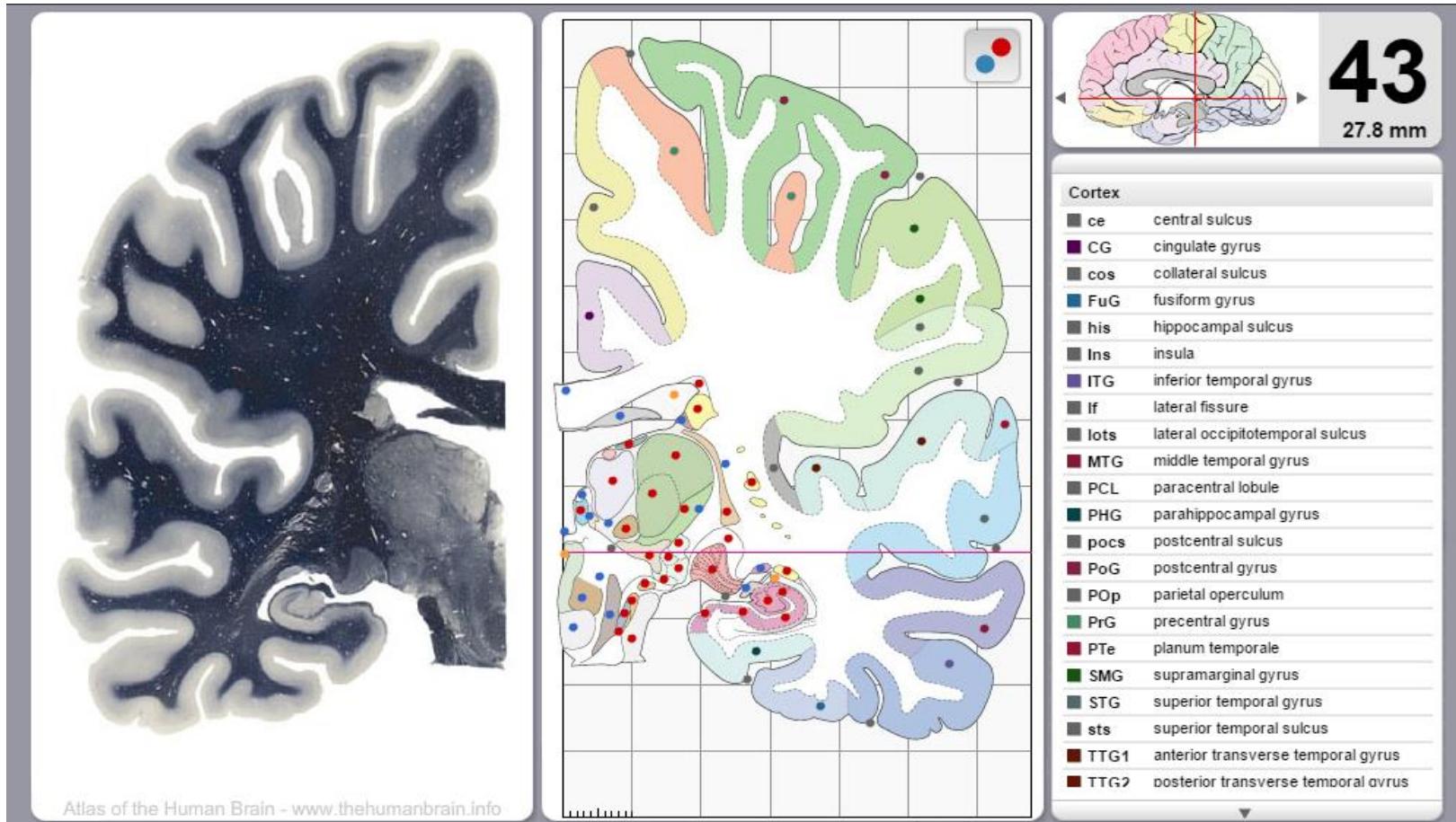
Atlas Segmentation

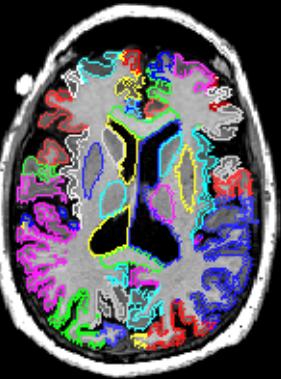
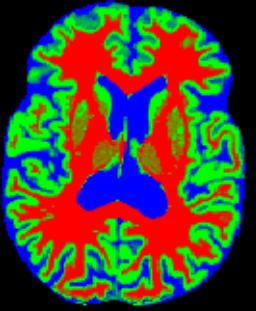


<http://jaysimons.deviantart.com/art/Map-of-the-Internet-1-0-427143215>

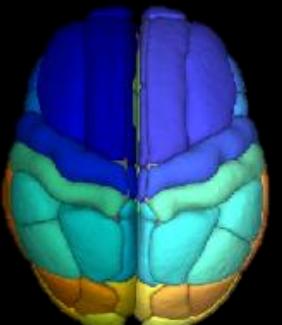


Segmentation





Segmentation Methods on Medical Images



Thresholding

Region Growing

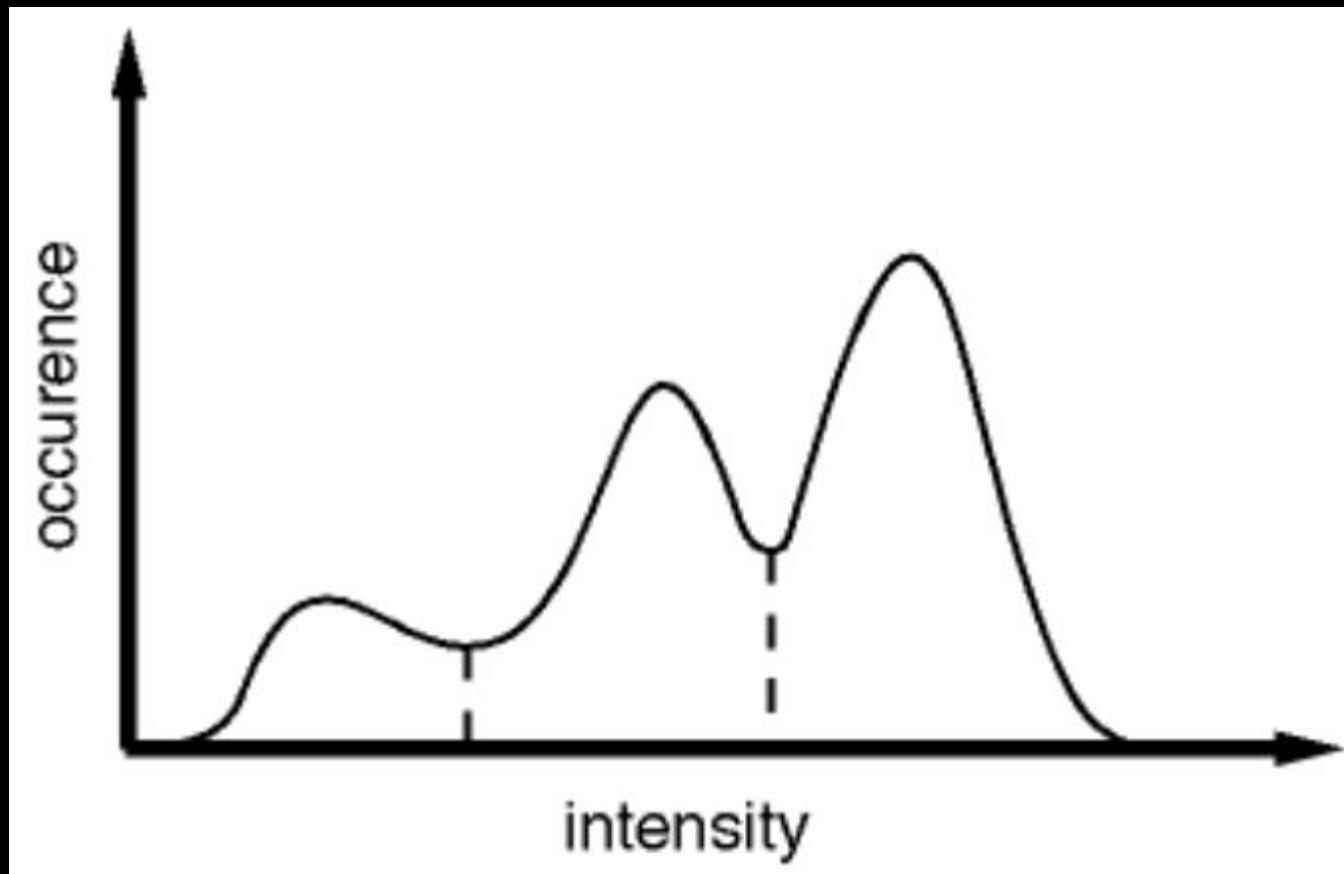
Classifiers

Clustering

Markov Random Field Models

Deformable Models

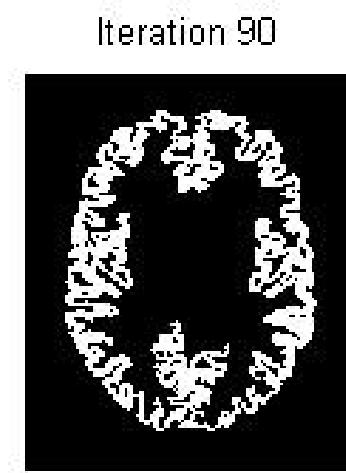
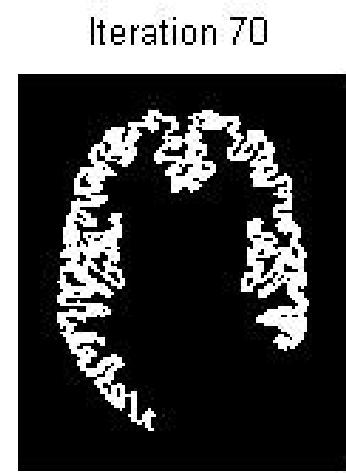
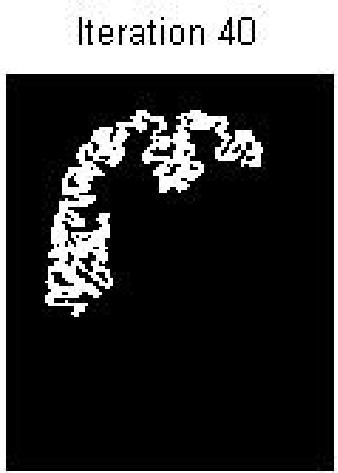
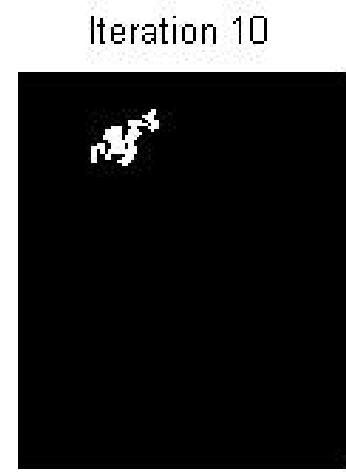
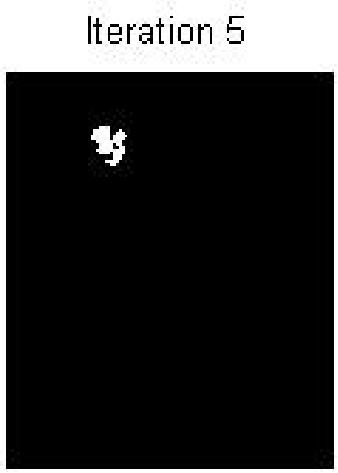
Atlas-based Segmentation



Thresholding

Cons:

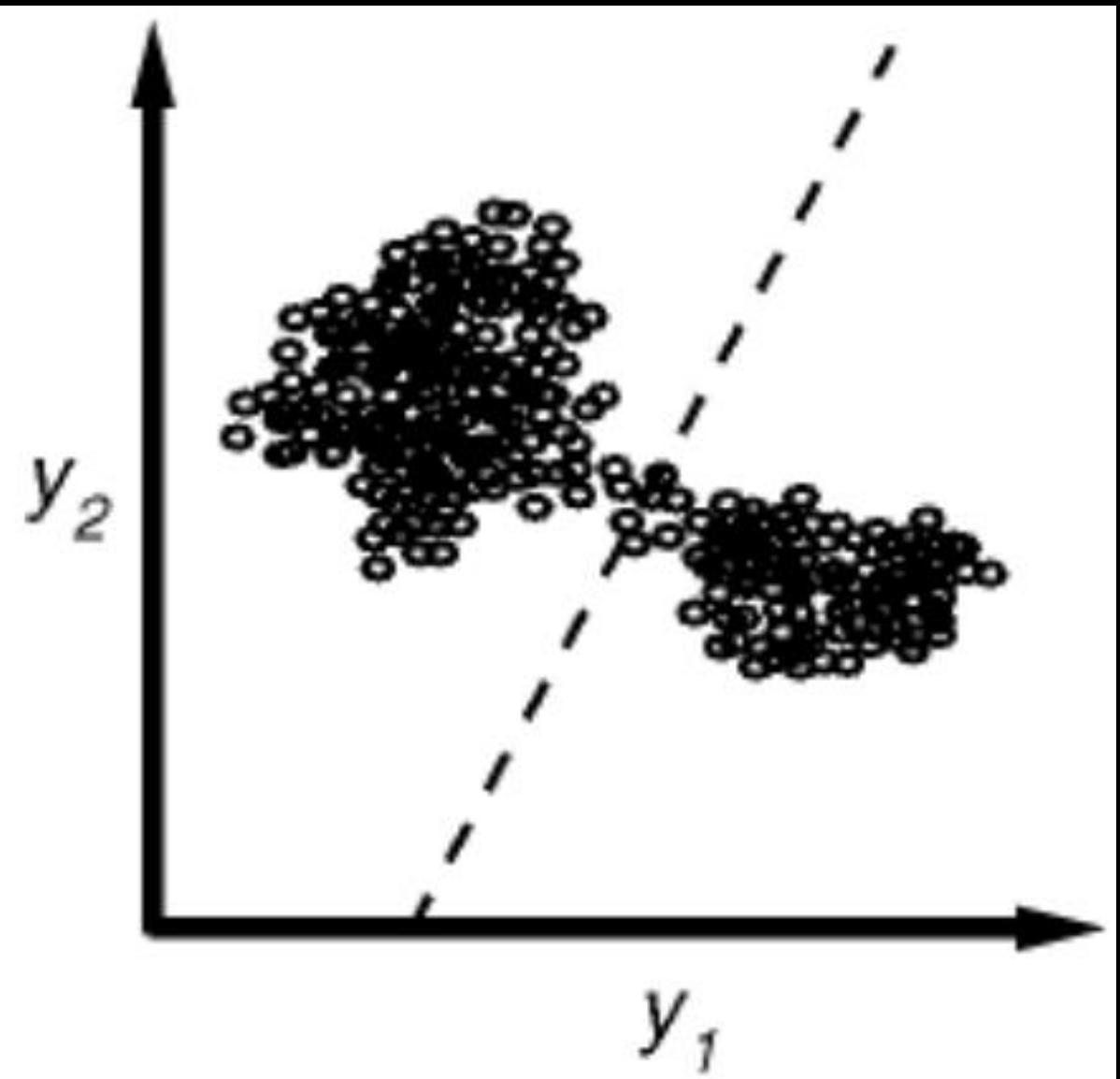
- Depends on absolute intensity



Cons:

- Depends on absolute intensity
- Semi-manual

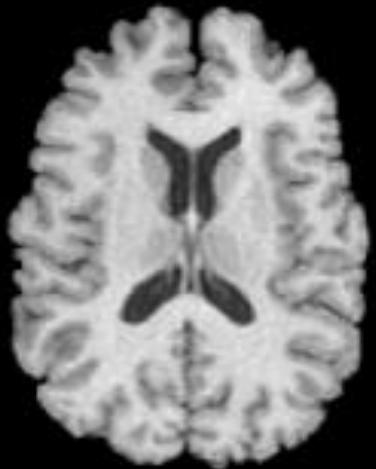
Region Growing



Classifier

Cons:

- Depends on absolute intensity



Clustering

Cons:

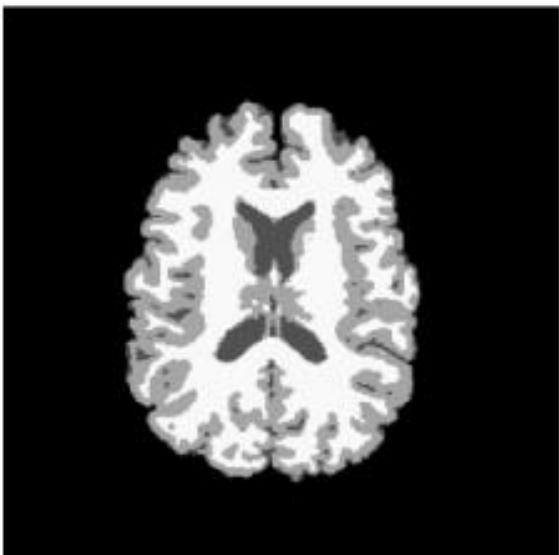
- Require hypothesis
- Depends on absolute intensity



(a)



(b)



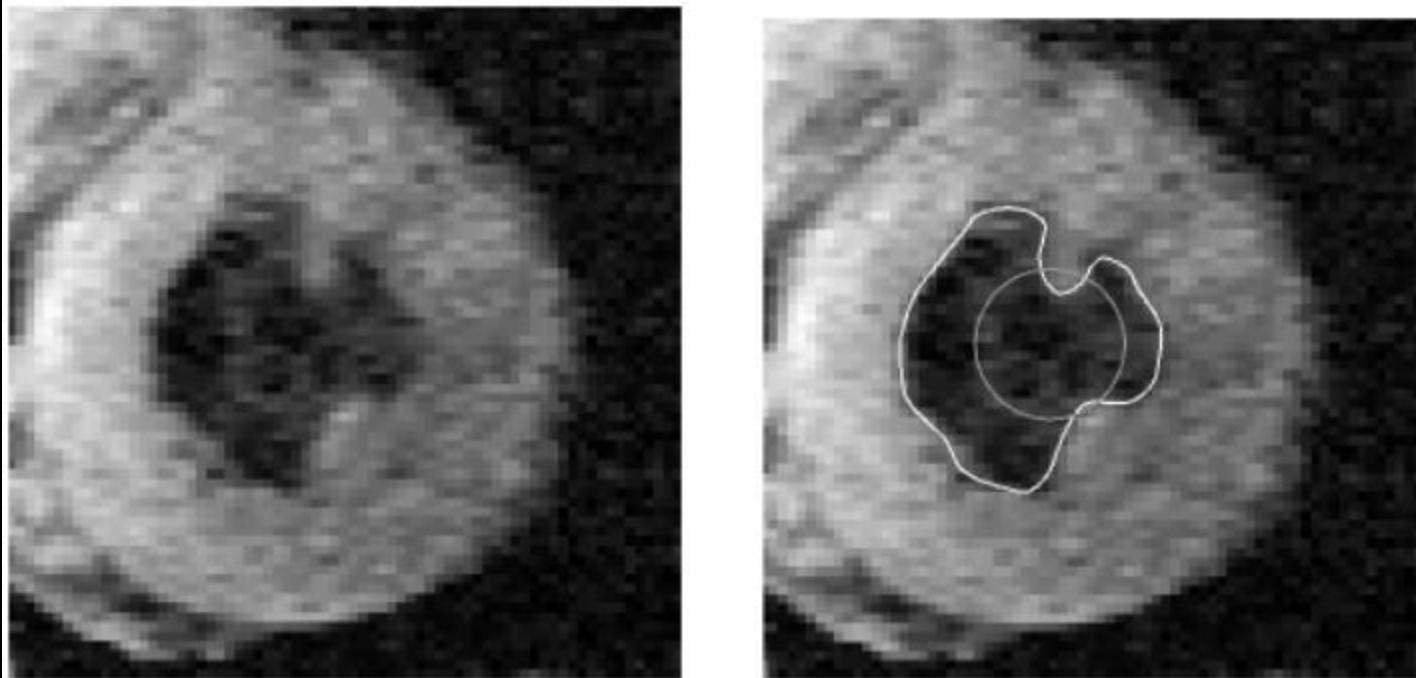
(c)

Markov Random Field

Pham et al., 2010

Cons:

- Not an segmentation method
- Sensitive to parameter setting



Deformable Model

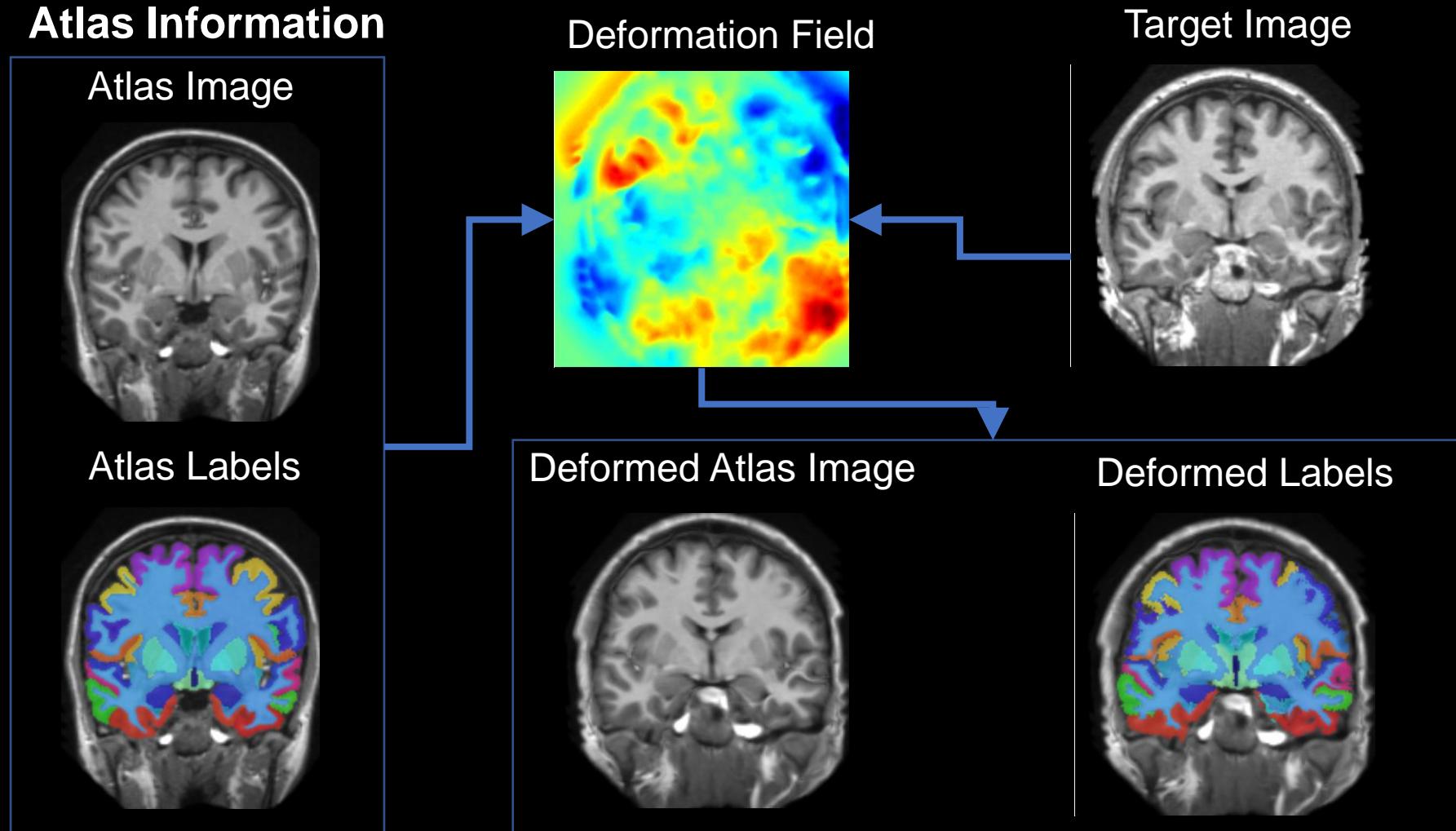
Cons:

- Manual Initialization
- Parameters



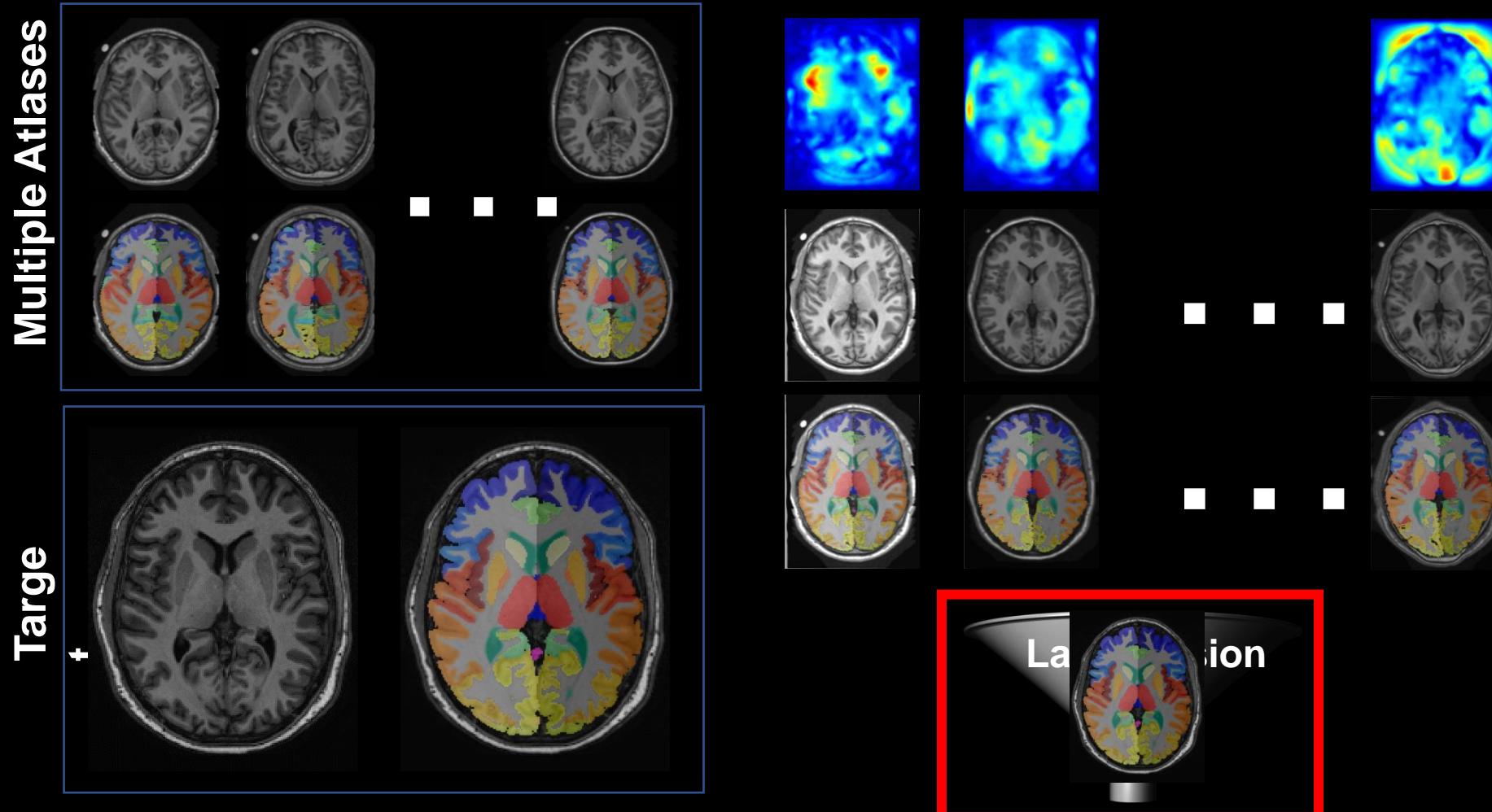
Learning by Example: Atlas-Based Segmentation

Gee et al. JCAT. 1993

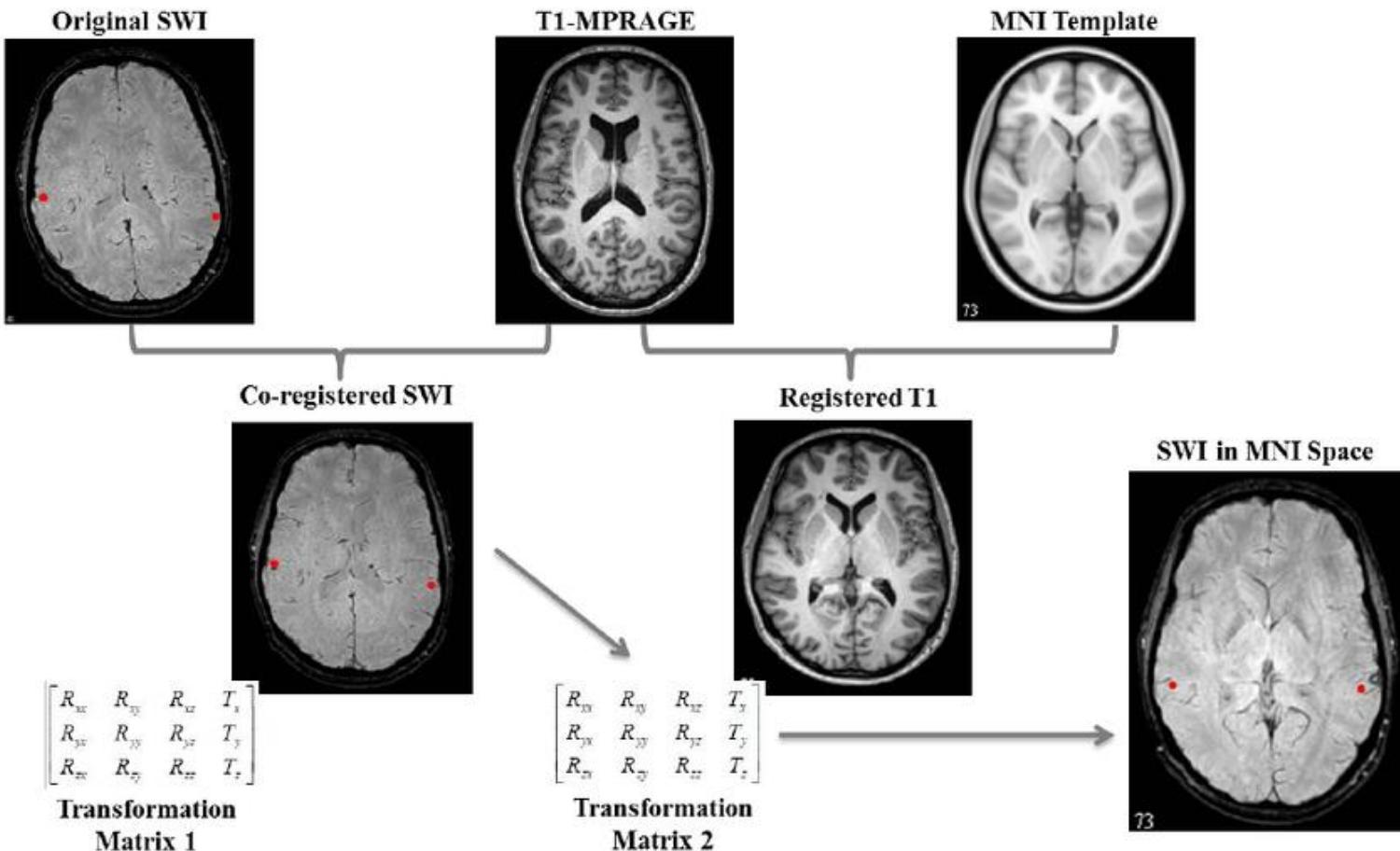


Multi-atlas Segmentation

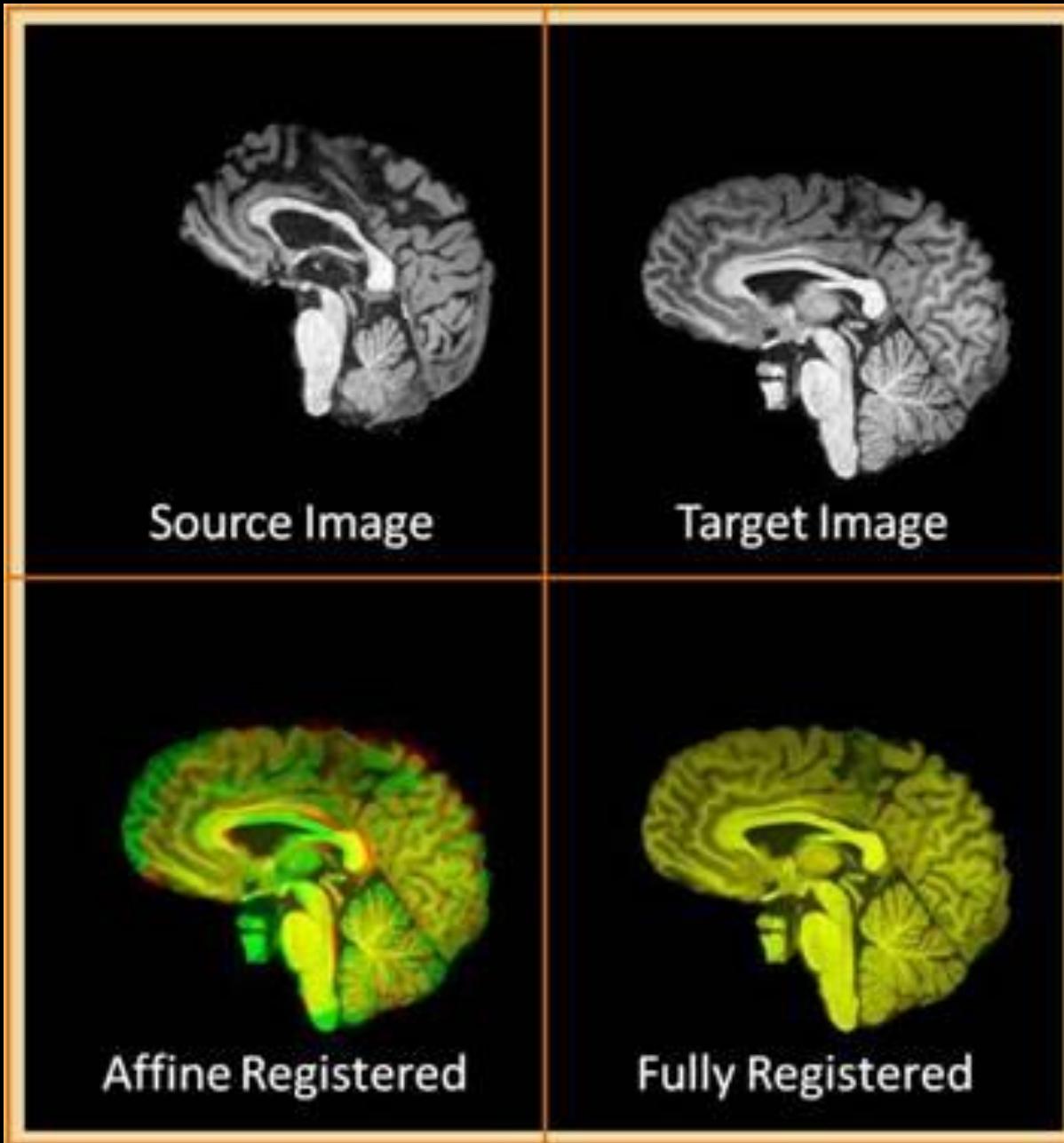
Learning by Example: Multi-atlas Segmentation

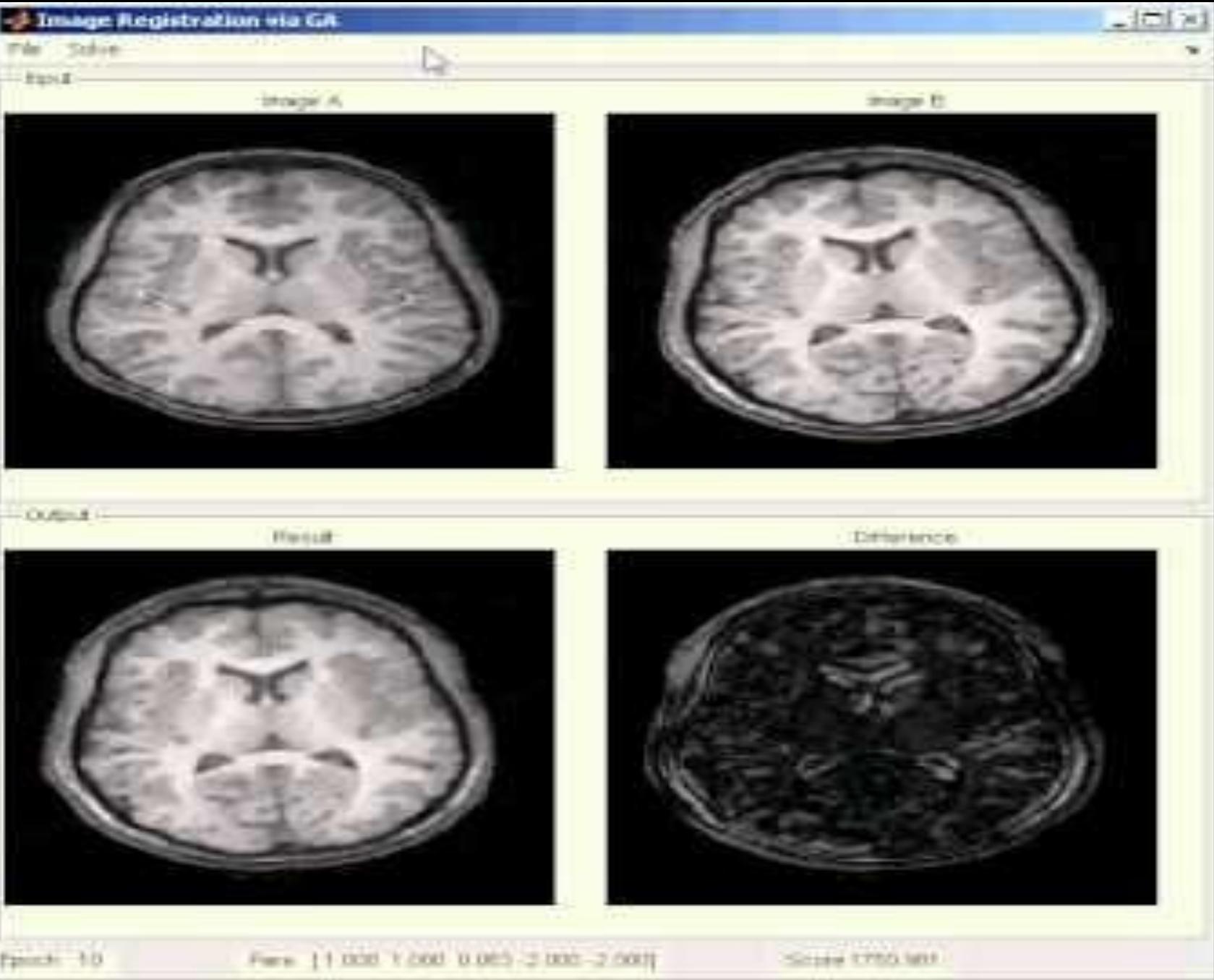


Registration



<https://www.researchgate.net/figure/Schematic-illustration-of-the-image-post-processing-framework-Red-dots->



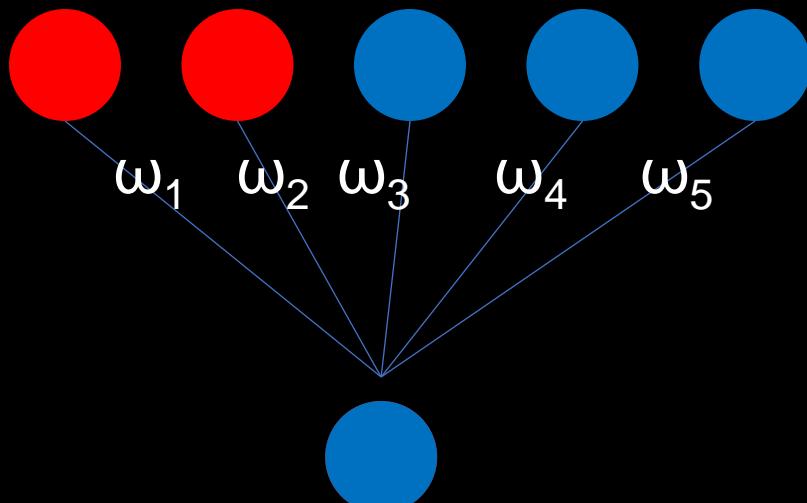


Label Fusion

Multi-atlas Label Fusion

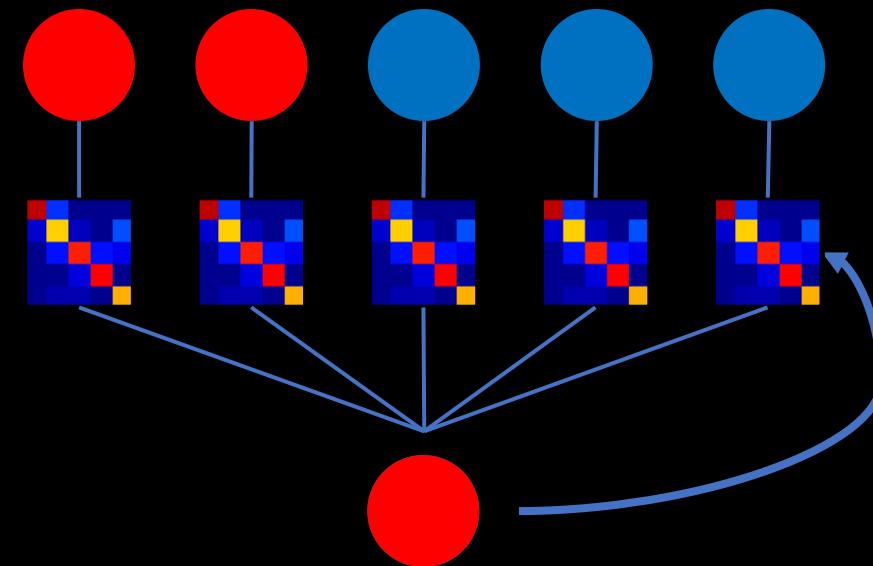
Iglesias et al. 2015 Asman et al. 2013

Voting Label Fusion



Majority Vote (MV)
Weighted Vote (WV)

Statistical Label Fusion



Expectation Maximization (EM)

Non-local Spatial STAPLE (NLSS)

Why multi-atlas label fusion works?

A **boring** experiment was conducted by Nashville Public Radio on 2015

How Much Does This Cow Weigh?



The cow's name is Penelope.

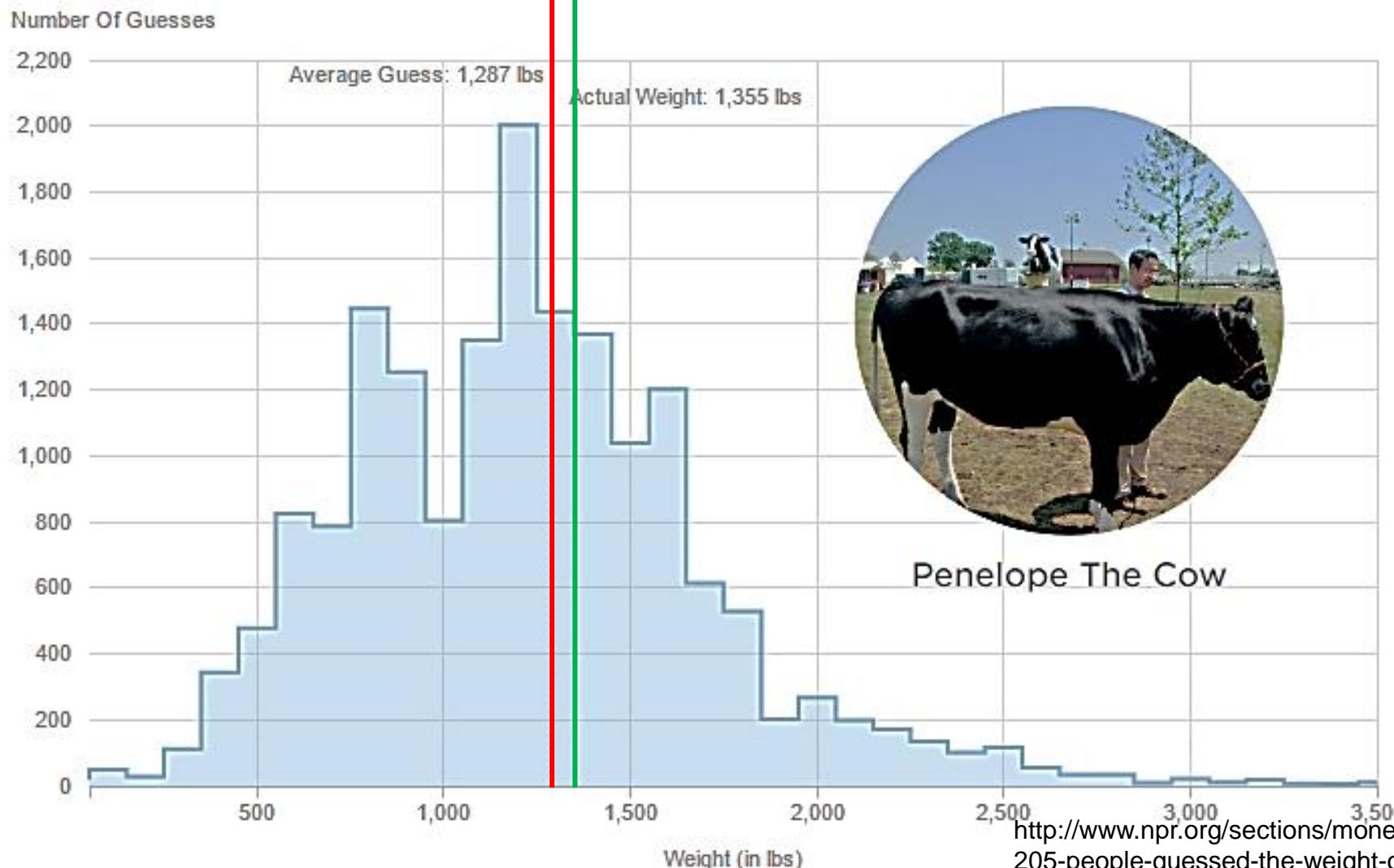
Let's Play the Game

Everyone guess the weight of Penelope



After a few weeks 17,205 boring people guessed the weight of a cow.

average guess
1,287 pounds



actual weight
1,355 pounds

68 pounds in difference
(about 5 percent)

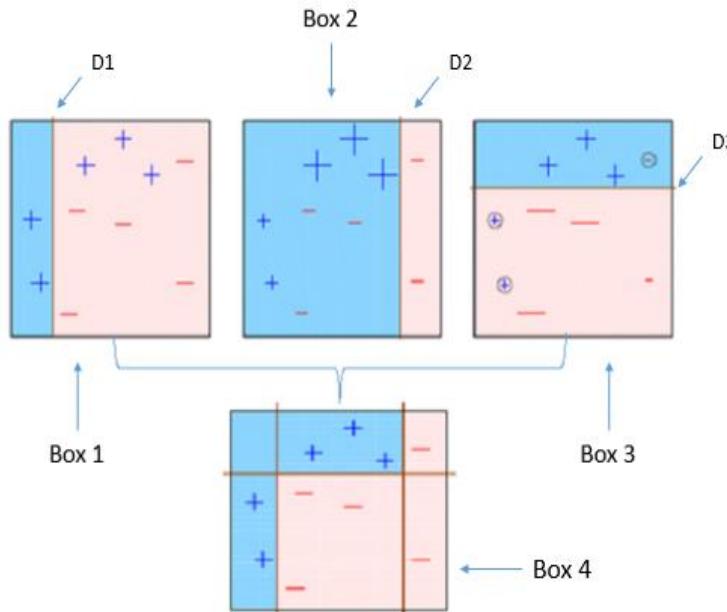


Penelope The Cow

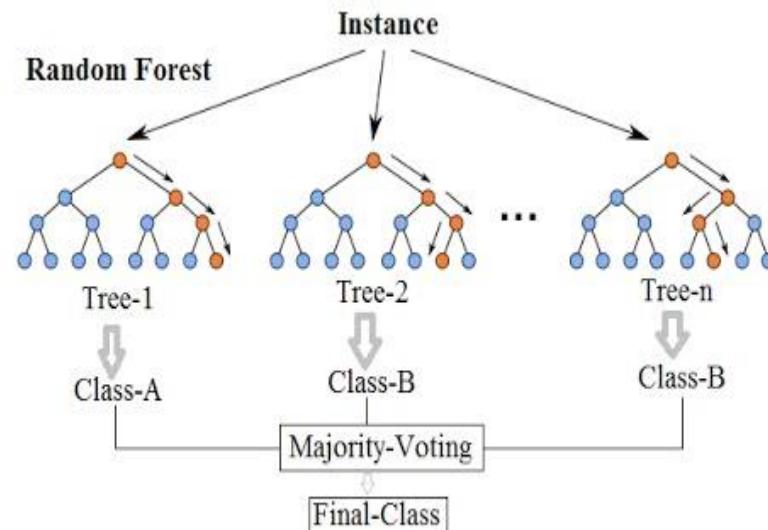
Boosting

“A set of **weak** learners create a single **strong** learner.”

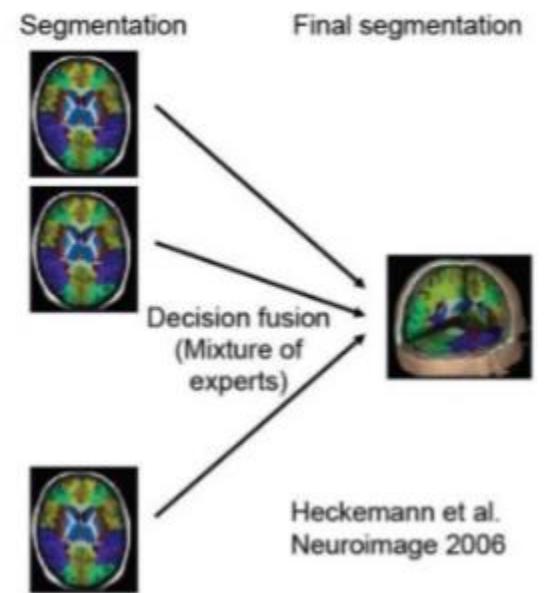
-- Robert Schapire 1990



AdaBoost



Random Forest



Multi-atlas Label Fusion

AtlasNet: Multi-atlas Non-linear Deep Networks for Medical Image Segmentation

M. Vakalopoulou^{1(✉)}, G. Chassagnon^{1,2}, N. Bus⁴, R. Marini⁴,
E. I. Zacharaki³, M.-P. Revel², and N. Paragios^{1,4}

¹ CVN, CentraleSupélec, Université Paris-Saclay, Paris, France
maria.vakalopoulou@centralesupelec.fr

Groupe Hospitalier Cochin-Hôtel Dieu, Université Paris Descartes, Paris, France

³ University of Patras, Patras, Greece
⁴ TheraPanacea, Paris, France

Abstract. Deep learning methods have gained increasing attention in addressing segmentation problems for medical images analysis despite the challenges inherited from the medical domain, such as limited data availability, lack of consistent textural or salient patterns, and high dimensionality of the data. In this paper, we introduce a novel multi-network architecture that exploits domain knowledge to address those challenges. The proposed architecture consists of multiple deep neural networks that are trained after co-aligning multiple anatomies through multi-metric deformable registration. This multi-network architecture can be trained with fewer examples and leads to better performance, robustness and generalization through consensus. Comparable to human accuracy, highly promising results on the challenging task of interstitial lung disease segmentation demonstrate the potential of our approach.

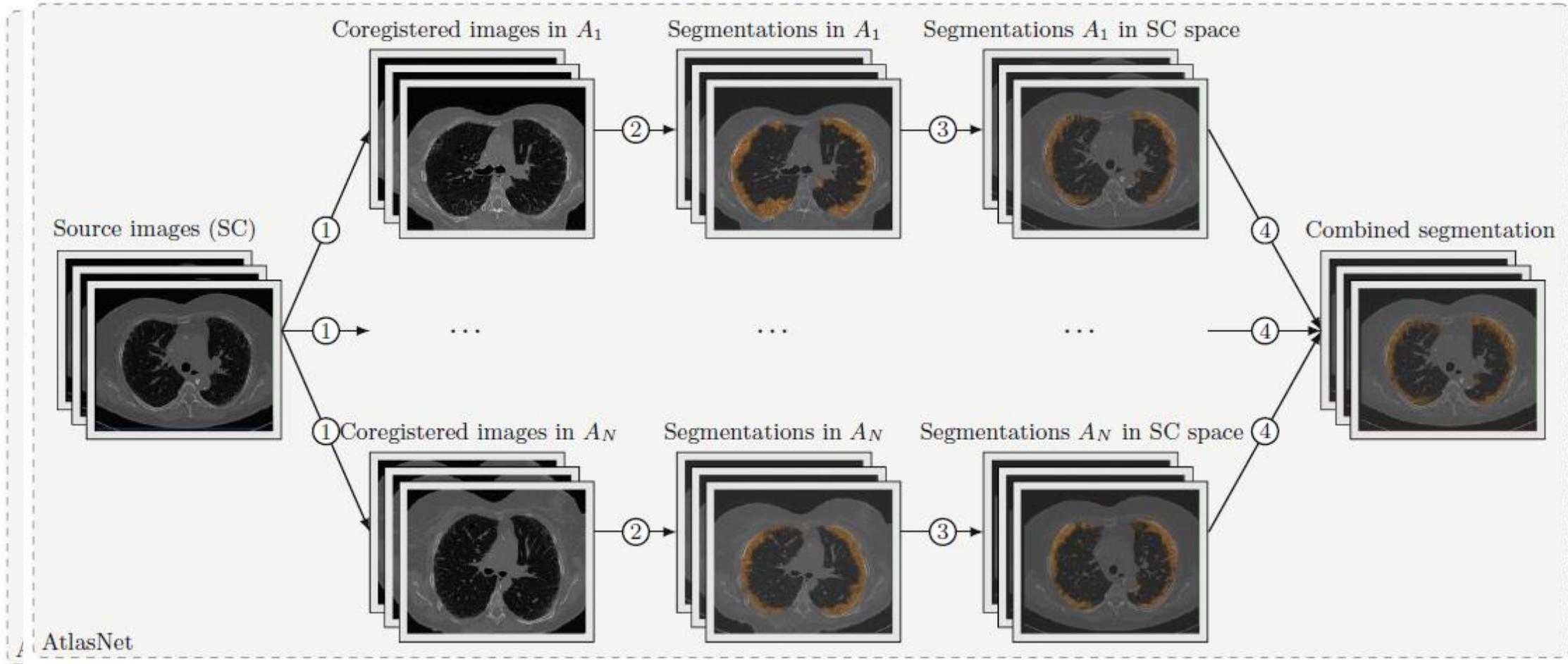


Fig. 1. The proposed AtlasNet framework. A_i indicates atlas i .

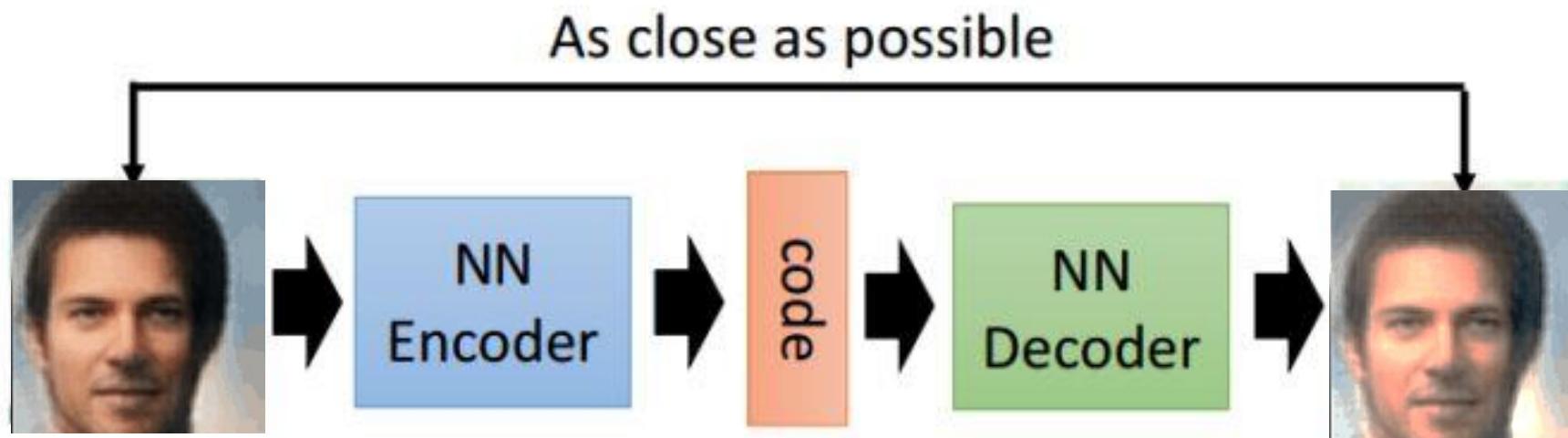
Good Things in Traditional Medical Imaging

BME 357 Advanced Image Processing (Dr. Benoit Dawant)

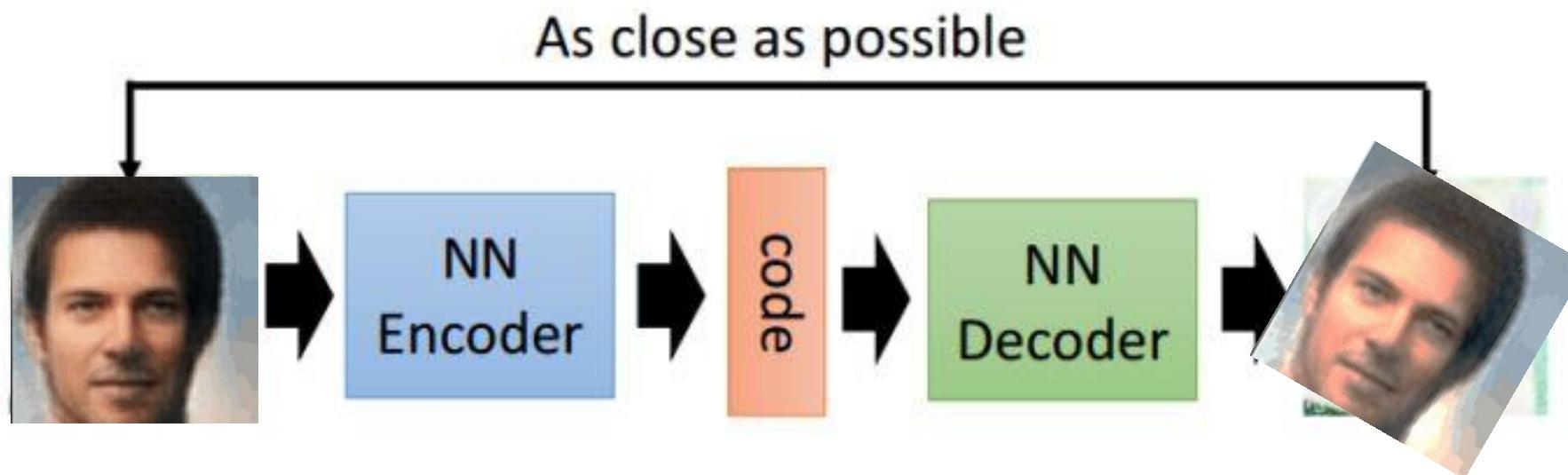
BME 358 Quantitative and Functional Imaging (Dr. Bennett Landman)

BME 377 Quantitative and Functional Imaging (Dr. Adam Anderson)

Self-supervised Learning



Self-supervised Learning



Self-supervised Learning

Self-supervised Learning

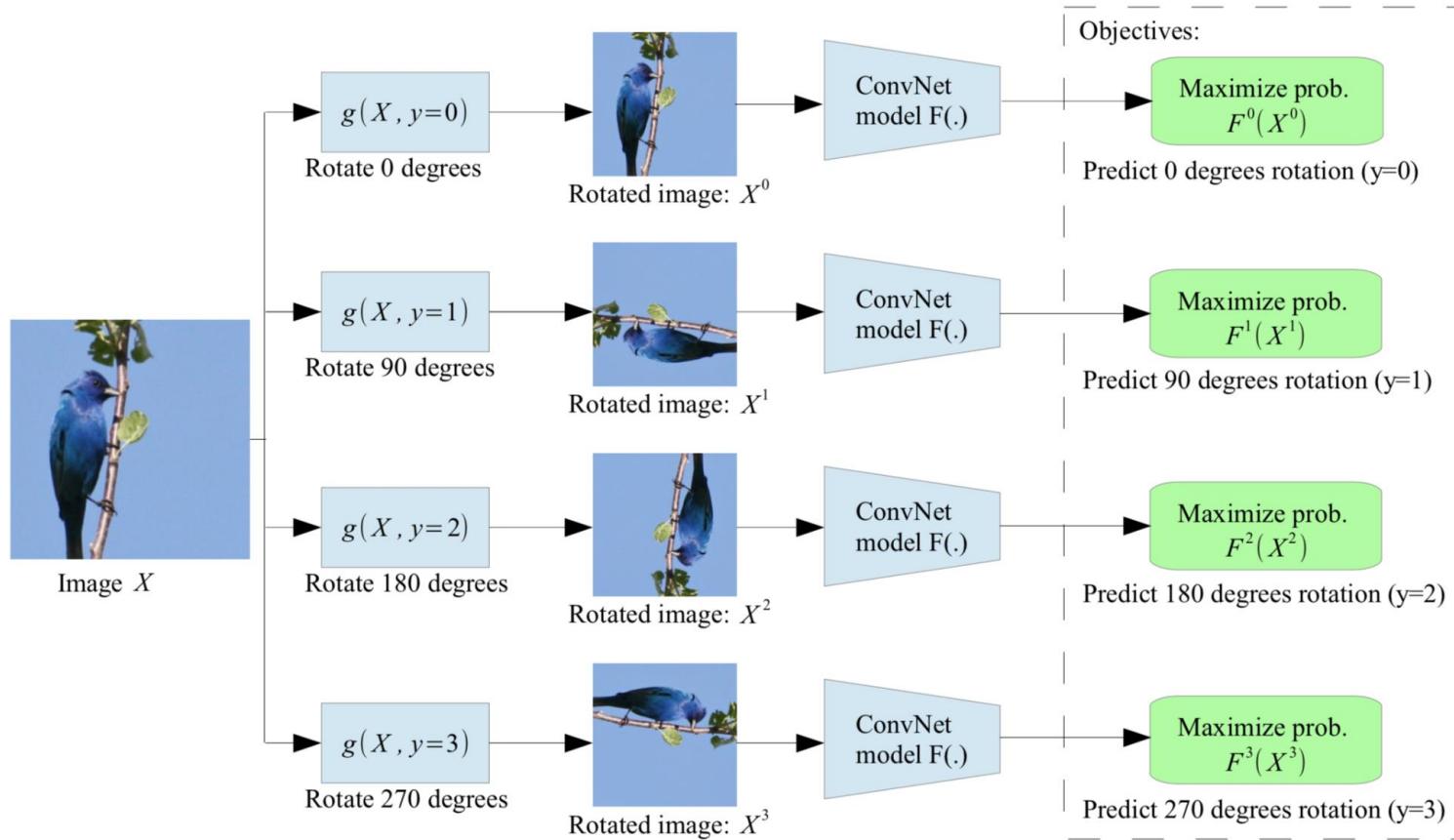
Distortion

Patches

Colorization

Generative Modeling

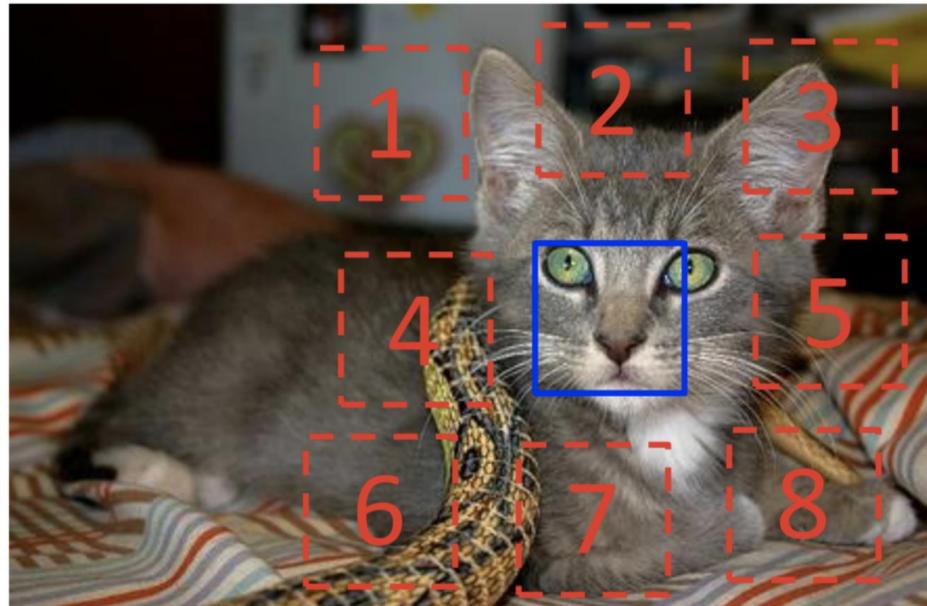
Distortion



predicting the degree

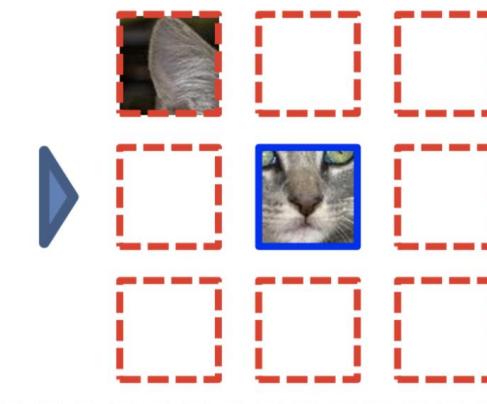
Gidaris et al. 2018

Patches



$$X = (\text{cat face}, \text{cat ear}); Y = 3$$

Example:



Question 1:



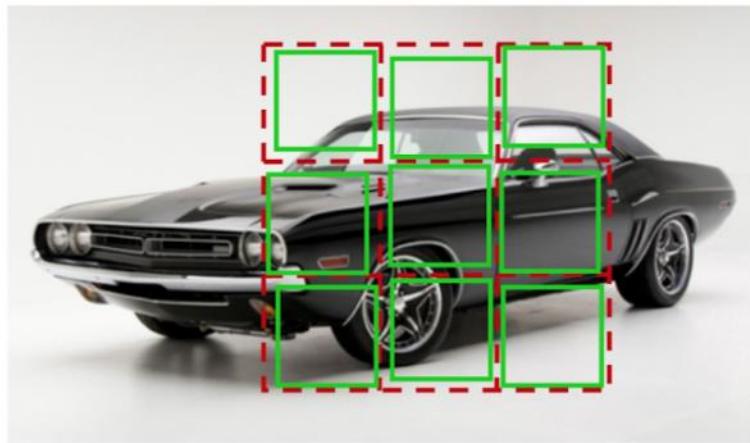
Question 2:



predicting the relative position

Doersch et al., 2015

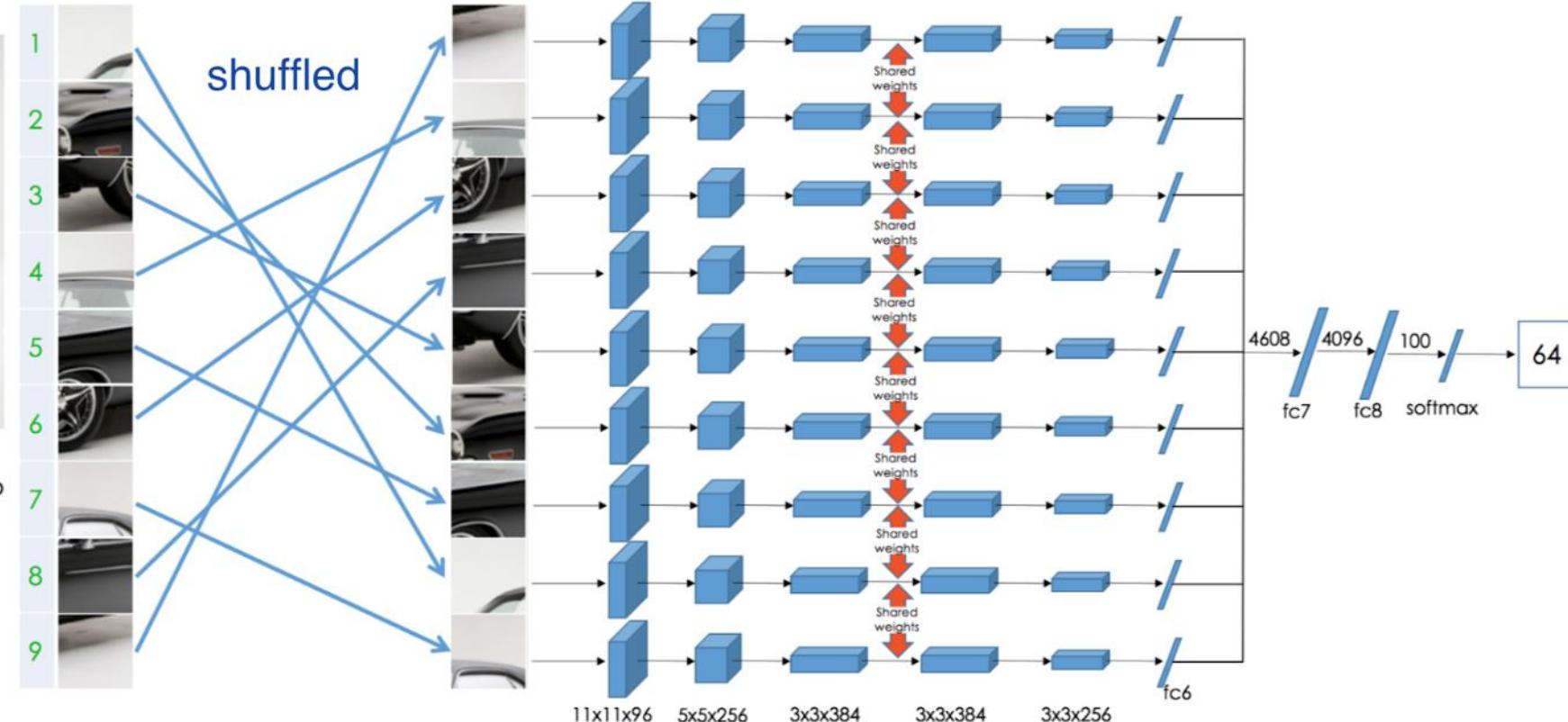
Patches



Permutation Set

index	permutation
64	9,4,6,8,3,2,5,1,7

Reorder patches according to the selected permutation



predicting the puzzle

Colorization

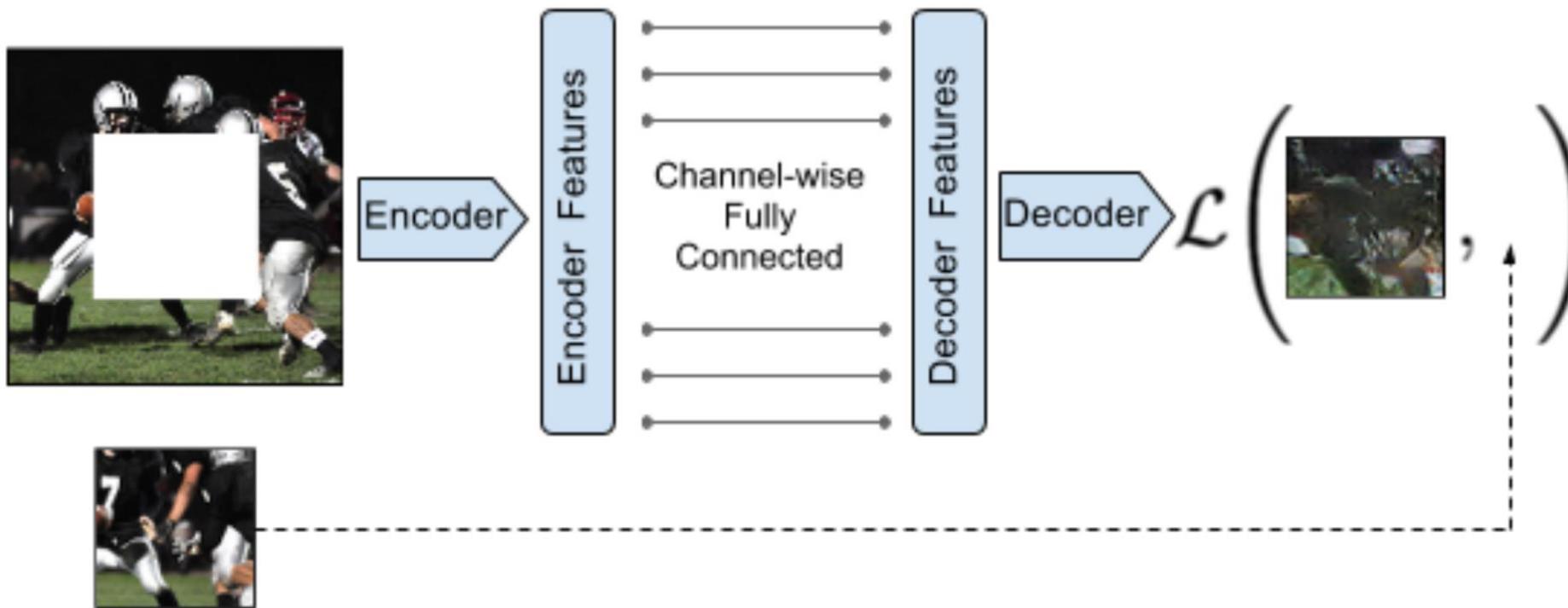


Fig. 1. Example input grayscale photos and output colorizations from our algorithm. These examples are cases where our model works especially well. Please visit <http://richzhang.github.io/colorization/> to see the full range of results and to try our model and code. Best viewed in color (obviously).

predicting the color

<https://arxiv.org/pdf/1603.08511.pdf>

Generative Model



predicting the context

<https://arxiv.org/pdf/1603.08511.pdf>

Topics

online machine learning is a method of [machine learning](#) in which data becomes available in a sequential order and is used to update our best predictor for future data at each step, as opposed to batch learning techniques which generate the best predictor by learning on the entire training data set at once.

We usually distinguish 3 optimization modes in machine learning:

- 1) Off-line / Batch
- 2) On-line,
- 3) Incremental

Off-line or Batch

An overview of gradient descent optimization algorithms*

Sebastian Ruder

Insight Centre for Data Analytics, NUI Galway

Aylien Ltd., Dublin

ruder.sebastian@gmail.com

Abstract

Gradient descent optimization algorithms, while increasingly popular, are often used as black-box optimizers, as practical explanations of their strengths and weaknesses are hard to come by. This article aims to provide the reader with intuitions with regard to the behaviour of different algorithms that will allow her to put them to use. In the course of this overview, we look at different variants of gradient descent, summarize challenges, introduce the most common optimization algorithms, review architectures in a parallel and distributed setting, and investigate additional strategies for optimizing gradient descent.

Off line Learning

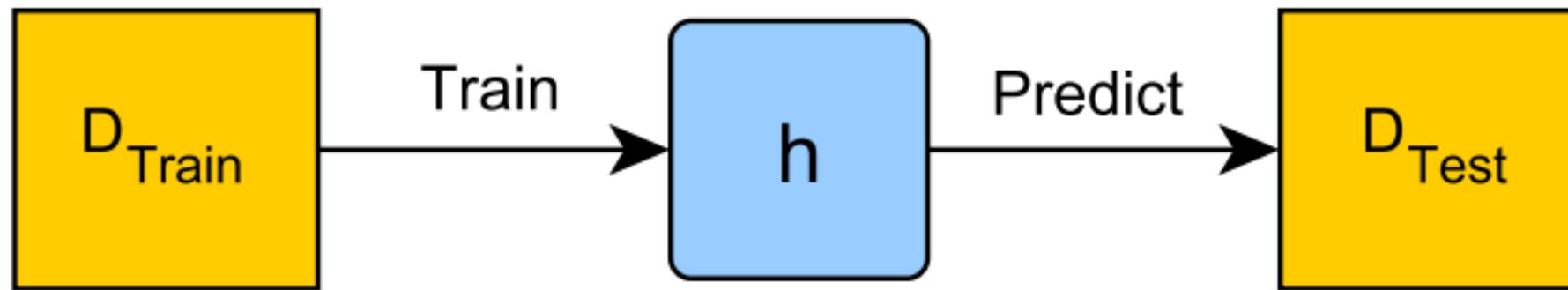


Figure 1: Classical scheme of evaluating a batch algorithm in off-line mode.

Online Learning

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 17, NO. 6, NOVEMBER 2006

A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks

Nan-Ying Liang, Guang-Bin Huang, *Senior Member, IEEE*, P. Saratchandran, *Senior Member, IEEE*, and
N. Sundararajan, *Fellow, IEEE*

Online Learning

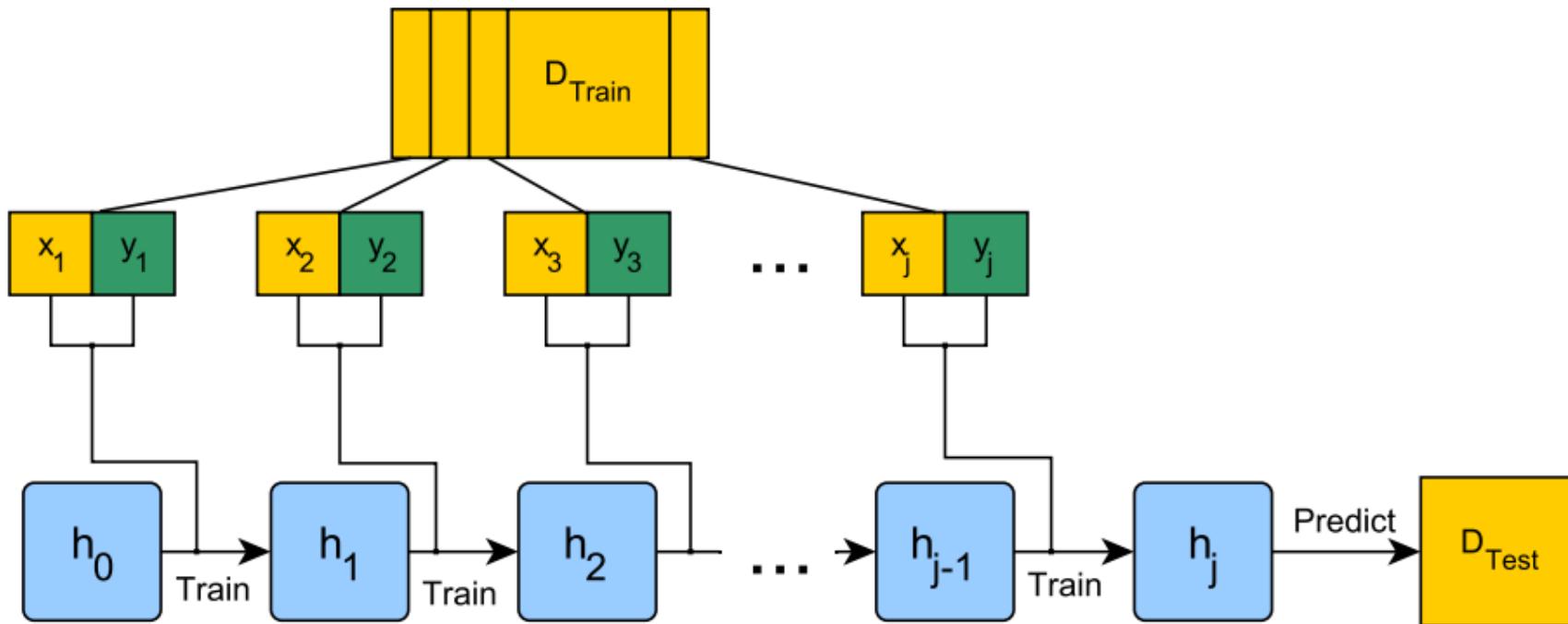


Figure 2: The process of testing an incremental algorithm in the off-line setting. Noticeably, only the last constructed model is used for prediction. All data used during training (x_i, y_i) is obtained from the training set D_{train}

Incremental Learning

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/6928613>

Universal Approximation Using Incremental Constructive Feedforward Networks With Random Hidden Nodes

Article *in* IEEE Transactions on Neural Networks · July 2006

DOI: 10.1109/TNN.2006.875977 · Source: PubMed

Incremental Learning

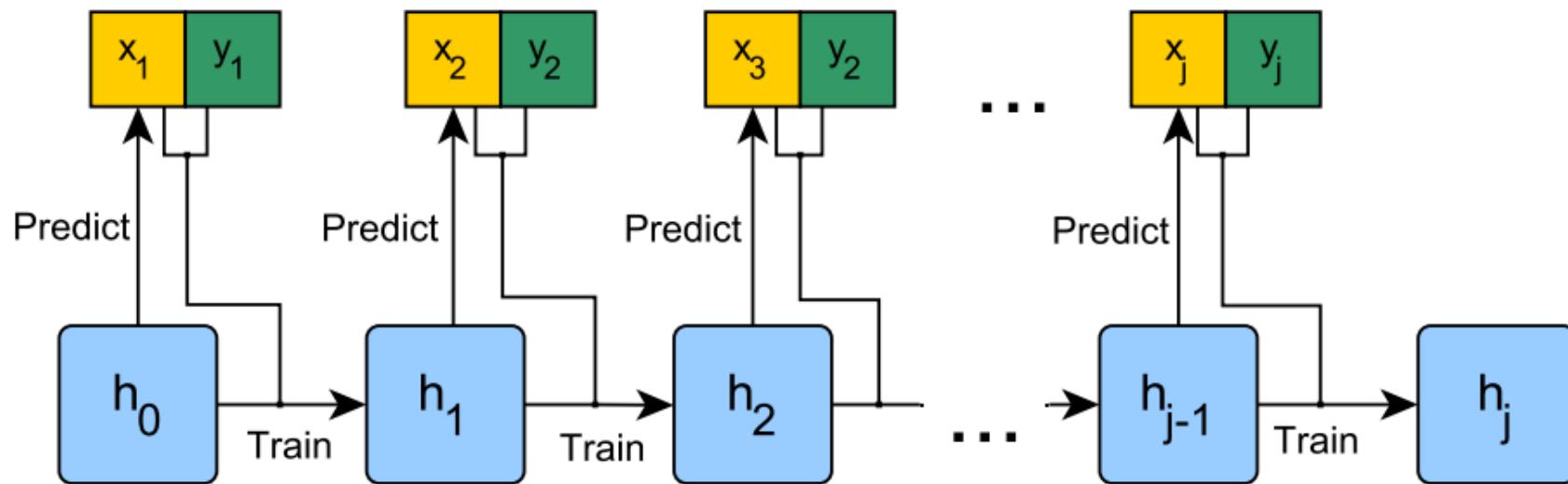


Figure 3: The online-learning scheme. Data is not split into training- and testing set. Instead, each model predicts subsequently one example, which is afterwards used for the construction of the next model.

Learn the New, Keep the Old: Extending Pretrained Models with New Anatomy and Images

Firat Ozdemir¹ , Philipp Fuernstahl², and Orcun Goksel¹ 

¹ Computer-assisted Applications in Medicine, ETH Zurich, Zurich, Switzerland
ozdemirf@vision.ee.ethz.ch

² CARD Group, University Hospital Balgrist, University of Zurich,
Zurich, Switzerland

Abstract. Deep learning has been widely accepted as a promising solution for medical image segmentation, given a sufficiently large representative dataset of images with corresponding annotations. With ever increasing amounts of annotated medical datasets, it is infeasible to train a learning method always with all data from scratch. This is also doomed to hit computational limits, e.g., memory or runtime feasible for training. Incremental learning can be a potential solution, where new information (images or anatomy) is introduced iteratively. Nevertheless, for the preservation of the collective information, it is essential to keep some “important” (i.e., representative) images and annotations from the past, while adding new information. In this paper, we introduce a framework for applying incremental learning for segmentation and propose novel methods for selecting representative data therein. We comparatively evaluate our methods in different scenarios using MR images and validate the increased learning capacity with using our methods.

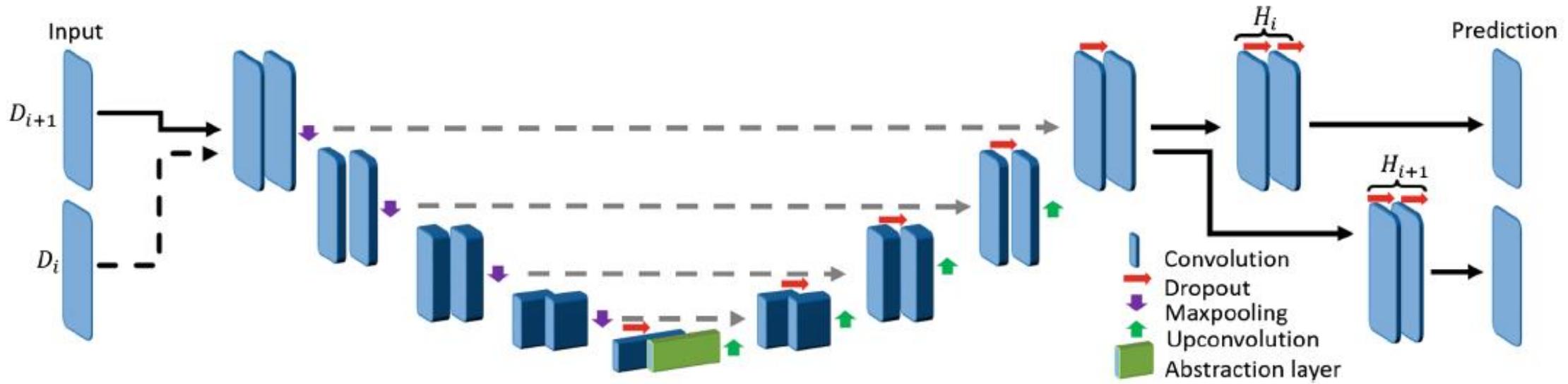


Fig. 1. Schematic of the proposed convolutional network at incremental step $i + 1$. Additional layers (“Head”) at step $i + 1$ are shown with H_{i+1} . Second layer at coarsest level is called *abstraction layer*. D_i denotes the *exemplar data* in Sect. 2.2.

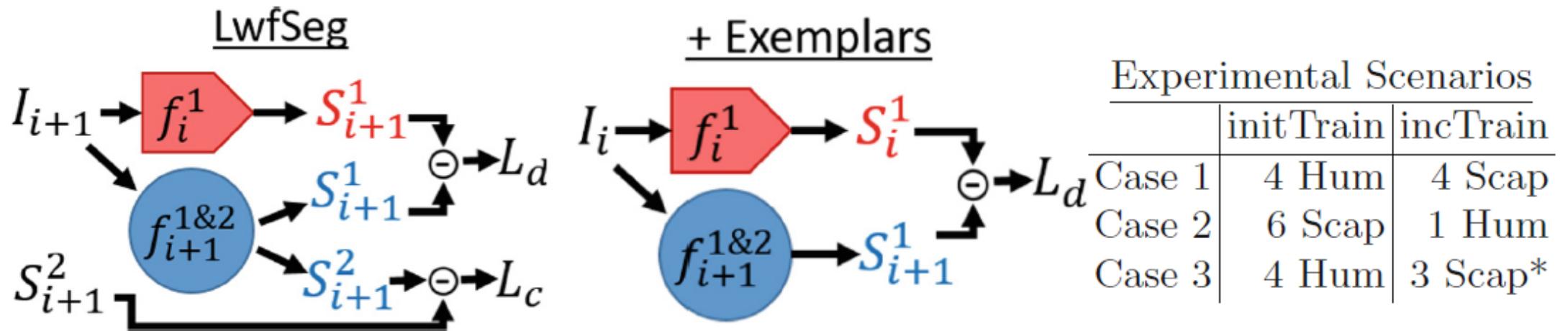


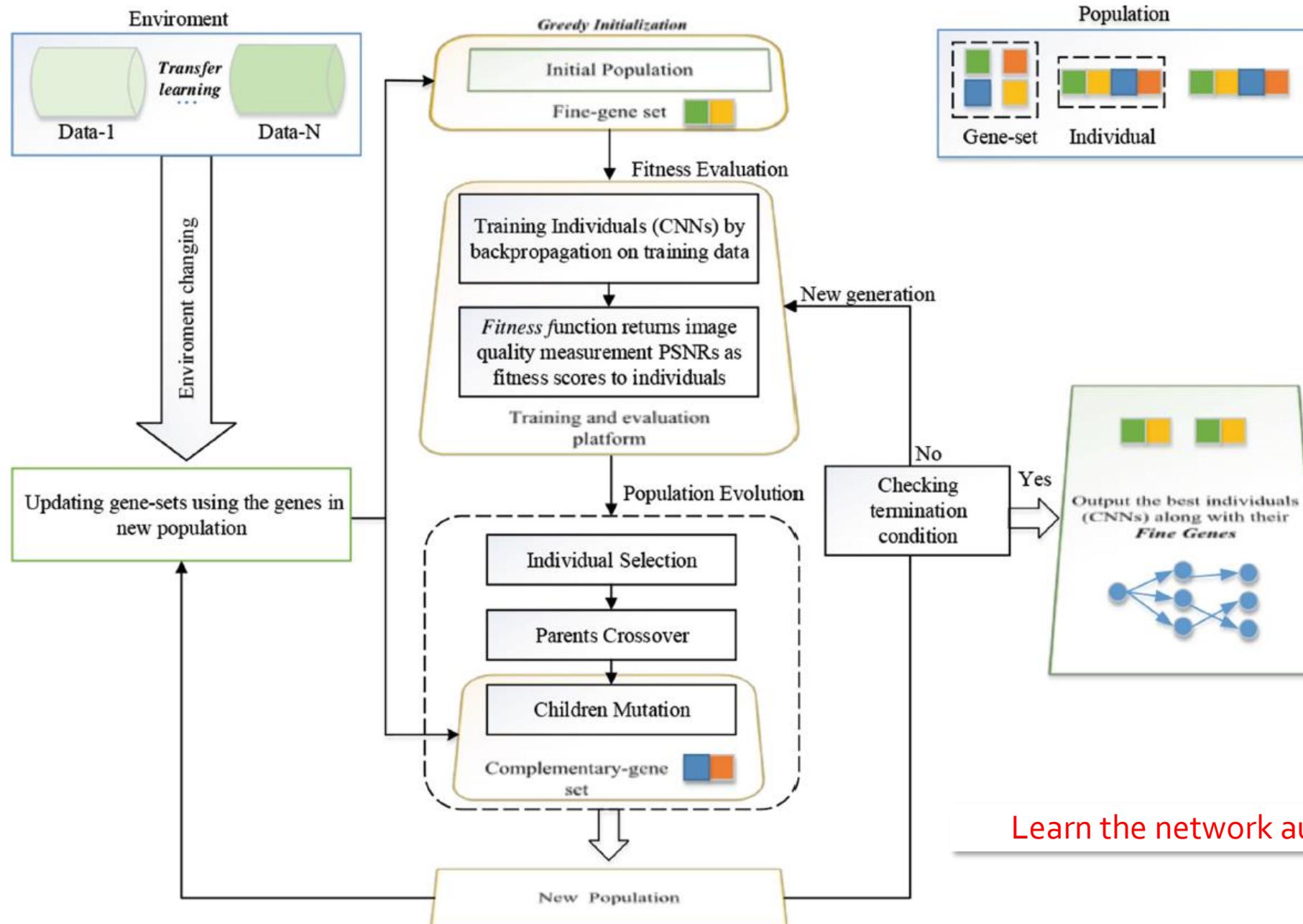
Fig. 2. Learning an incremental network $f_{i+1}^{1\&2}$ for classes 1&2 with new images I_{i+1} and annotations of new structures S_{i+1}^2 , given a pre-trained and frozen network f_i^1 . Left: representation of LwfSeg. Middle: additional loss in AeISeg and CoRiSeg for augmenting the new network (left) with exemplar images I_i . Right: experimental scenarios depicting initial (init) and incremental (inc) datasets for humerus (Hum) and scapula (Scap). Case 3 incTrain (*) was conducted on a different MR sequence (water-saturated Dixon).

Neural Network Evolution Using Expedited Genetic Algorithm for Medical Image Denoising

Peng Liu, Yangjunyi Li, Mohammad D. El Basha, and Ruogu Fang^(✉)

J. Crayton Pruitt Family Department of Biomedical Engineering,
University of Florida, Gainesville, FL, USA
Ruogu.Fang@bme.ufl.edu

Abstract. Convolutional neural networks offer state-of-the-art performance for medical image denoising. However, their architectures are manually designed for different noise types. The realistic noise in medical images is usually mixed and complicated, and sometimes unknown, leading to challenges in creating effective denoising neural networks. In this paper, we present a Genetic Algorithm (GA)-based network evolution approach to search for the fittest genes to optimize network structures. We expedite the evolutionary process through an experience-based greedy exploration strategy and transfer learning. The experimental results on computed tomography perfusion (CTP) images denoising demonstrate the capability of the method to select the fittest genes for building high-performance networks, named EvoNets, and our results compare favorably with state-of-the-art methods.



A Lifelong Learning Approach to Brain MR Segmentation Across Scanners and Protocols

Neerav Karani^(✉), Krishna Chaitanya, Christian Baumgartner,
and Ender Konukoglu

Computer Vision Lab, ETH Zurich, Zurich, Switzerland
nkarani@vision.ee.ethz.ch

Abstract. Convolutional neural networks (CNNs) have shown promising results on several segmentation tasks in magnetic resonance (MR) images. However, the accuracy of CNNs may degrade severely when segmenting images acquired with different scanners and/or protocols as compared to the training data, thus limiting their practical utility. We address this shortcoming in a lifelong multi-domain learning setting by treating images acquired with different scanners or protocols as samples from different, but related domains. Our solution is a single CNN with shared convolutional filters and domain-specific batch normalization layers, which can be tuned to new domains with only a few (≈ 4) labelled images. Importantly, this is achieved while retaining performance on the older domains whose training data may no longer be available. We evaluate the method for brain structure segmentation in MR images. Results demonstrate that the proposed method largely closes the gap to the benchmark, which is training a dedicated CNN for each scanner.

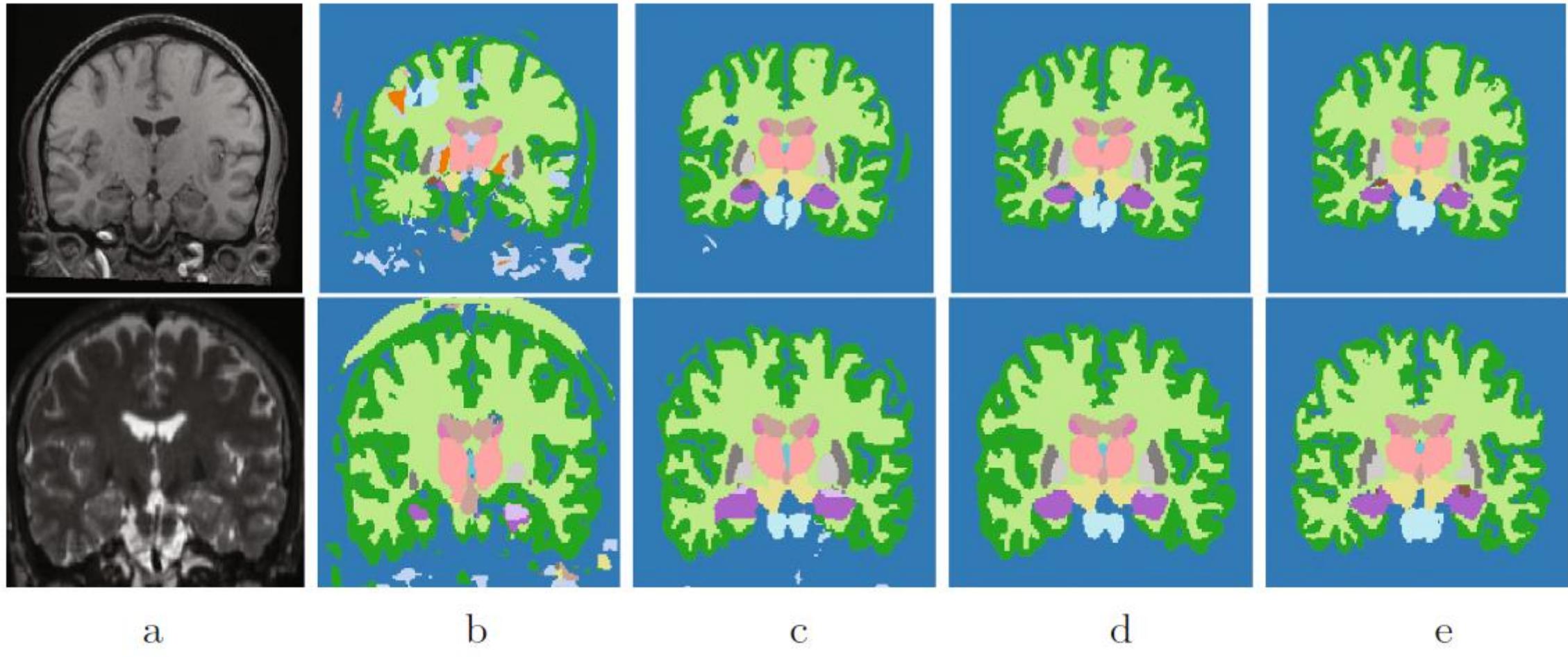


Fig. 2. Qualitative results: (a) images from domains D_d , segmentations predicted by (b) N_{123}^{bn} , bn_{k^*} , (c) $N_{123,k^*\rightarrow d}^{bn}$, bn_d , (d) N_d and (e) ground truth annotations, with $\{d, k^*\}$ as $\{4, 3\}$ (top) and $\{5, 2\}$ (bottom).

Keep and Learn: Continual Learning by Constraining the Latent Space for Knowledge Preservation in Neural Networks

Hyo-Eun Kim^(✉), Seungwook Kim, and Jaehwan Lee

Lunit Inc., Seoul, South Korea
`{hekim, swkim, jhlee}@lunit.io`

Abstract. Data is one of the most important factors in machine learning. However, even if we have high-quality data, there is a situation in which access to the data is restricted. For example, access to the medical data from outside is strictly limited due to the privacy issues. In this case, we have to learn a model sequentially only with the data accessible in the corresponding stage. In this work, we propose a new method for preserving learned knowledge by modeling the high-level feature space and the output space to be mutually informative, and constraining feature vectors to lie in the modeled space during training. The proposed method is easy to implement as it can be applied by simply adding a reconstruction loss to an objective function. We evaluate the proposed method on CIFAR-10/100 and a chest X-ray dataset, and show benefits in terms of knowledge preservation compared to previous approaches.

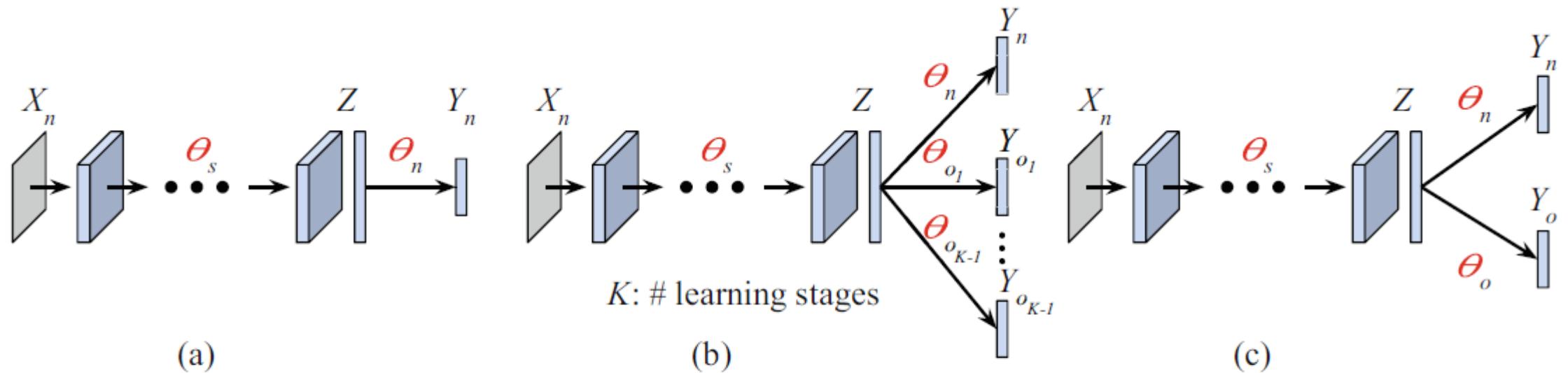


Fig. 1. Model architectures: (a) FT/EWC, (b) LwF, and (c) modified LwF (LwF+).

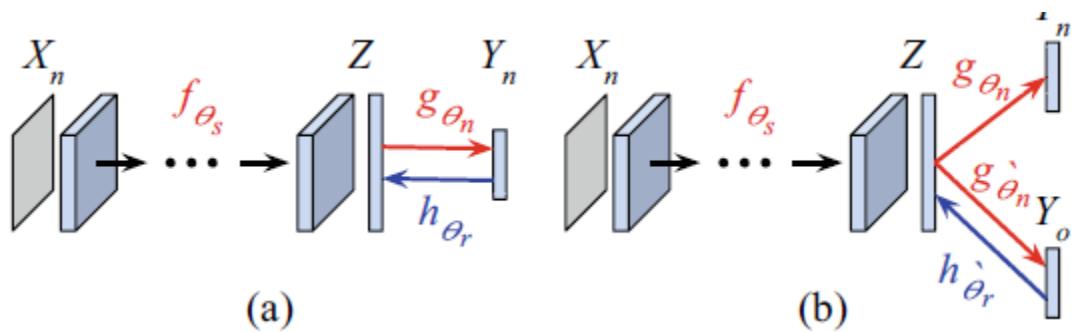


Fig. 2. Proposed model architecture: (a) the first learning stage and (b) the following learning stages.

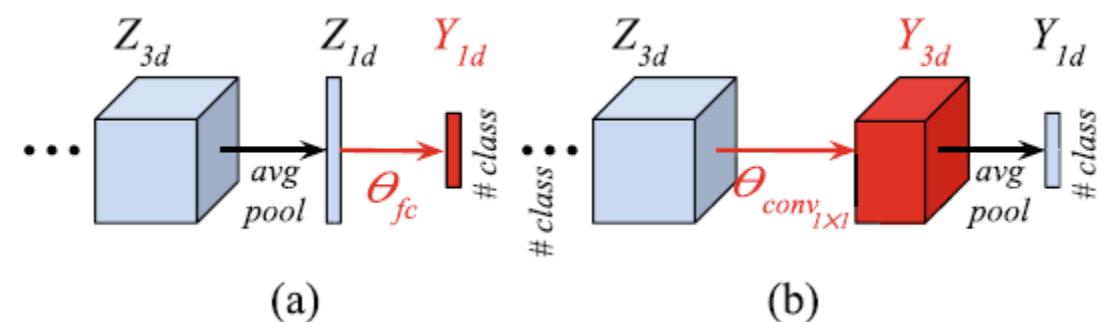


Fig. 3. Top layers of ResNet: based on (a) fc layer or (b) $conv_{1\times 1}$ layer. Both are functionally equivalent.

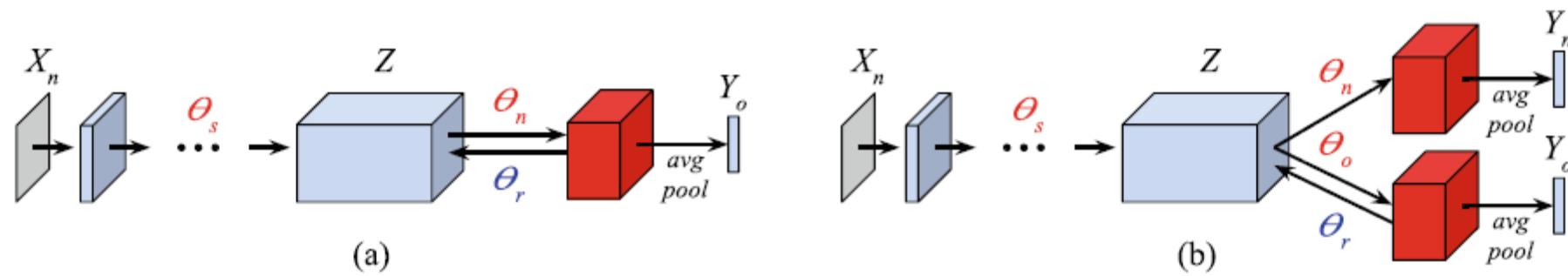


Fig. 4. Proposed model described in Fig. 2 based on the modified ResNet in Fig. 3.

Invasive Cancer Detection Utilizing Compressed Convolutional Neural Network and Transfer Learning

Bin Kong¹, Shanhui Sun²(✉), Xin Wang², Qi Song², and Shaoting Zhang¹(✉)

¹ Department of Computer Science, UNC Charlotte, Charlotte, NC, USA
s Zhang 16@uncc.edu

² CuraCloud Corporation, Seattle, WA, USA
shanhuis@curacloudcorp.com

Abstract. Identification of invasive cancer in Whole Slide Images (WSIs) is crucial for tumor staging as well as treatment planning. However, the precise manual delineation of tumor regions is challenging, tedious and time-consuming. Thus, automatic invasive cancer detection in WSIs is of significant importance. Recently, Convolutional Neural Network (CNN) based approaches advanced invasive cancer detection. However, computation burdens of these approaches become barriers in clinical applications. In this work, we propose to detect invasive cancer employing a lightweight network in a fully convolution fashion without model ensembles. In order to improve the small network's detection accuracy, we utilized the “soft labels” of a large capacity network to supervise its training process. Additionally, we adopt a teacher guided loss to help the small network better learn from the intermediate layers of the high capacity network. With this suite of approaches, our network is extremely efficient as well as accurate. The proposed method is validated on two large scale WSI datasets. Our approach is performed in an average time of 0.6 and 3.6 min per WSI with a single GPU on our gastric cancer dataset and CAMELYON16, respectively, about 5 times faster than Google Inception V3. We achieved an average FROC of 81.1% and 85.6% respectively, which are on par with Google Inception V3. The proposed method requires less high performance computing resources than state-of-the-art methods, which makes the invasive cancer diagnosis more applicable in the clinical usage.

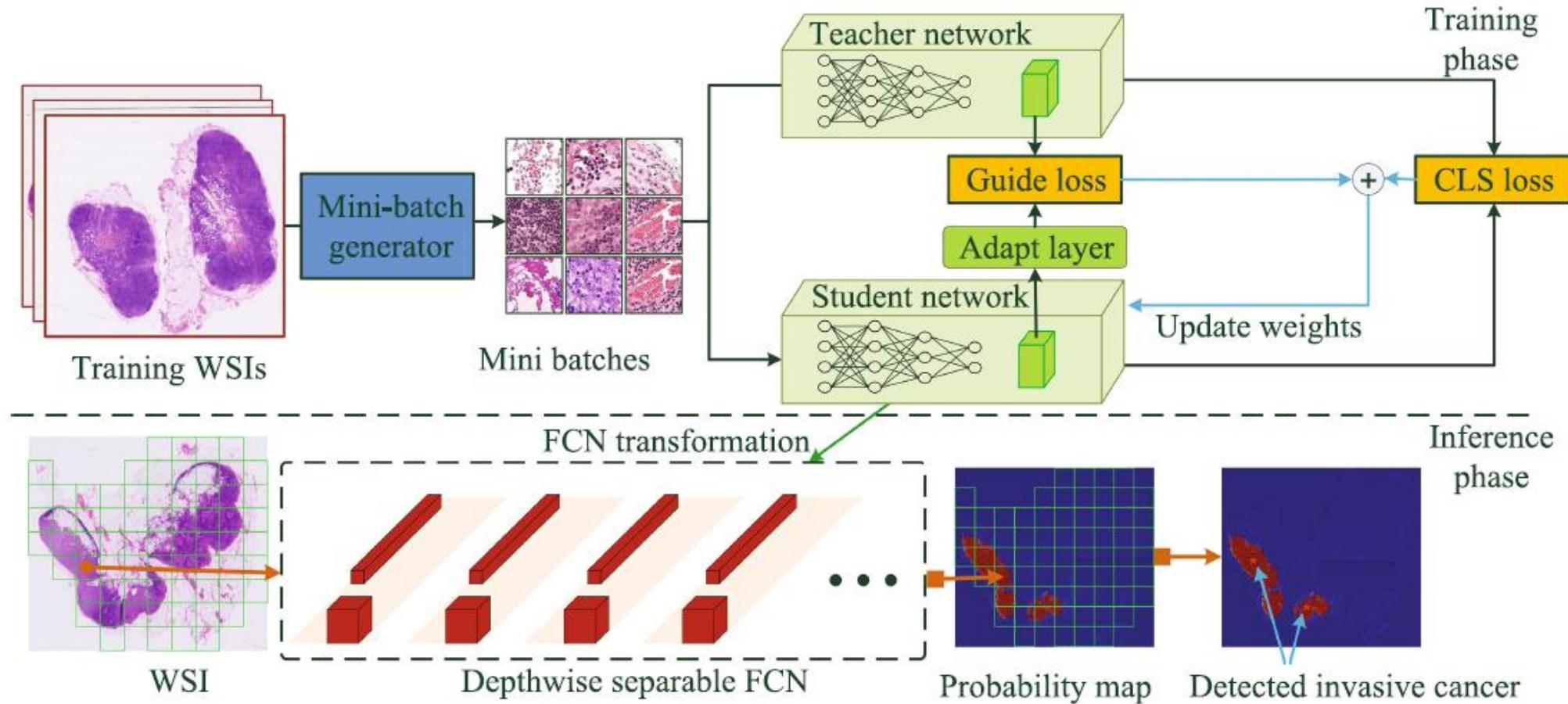


Fig. 1. Overview of the proposed framework. The above part indicates the training phase and below part indicates inference phase. Note that we only illustrate proposed transfer learning method in the training phase.

Methods		I	IF	S	SF	DSF	DSFG
Gastric cancer	Time (mins.)	3.8	2.3	1.5	0.6	0.6	0.6
	Ave. FROC	0.806	0.813	0.768	0.773	0.801	0.811
CAMELYON16	Time (mins.)	17.0	9.1	7.8	3.6	3.6	3.6
	Ave. FROC	0.857	0.859	0.809	0.815	0.847	0.856

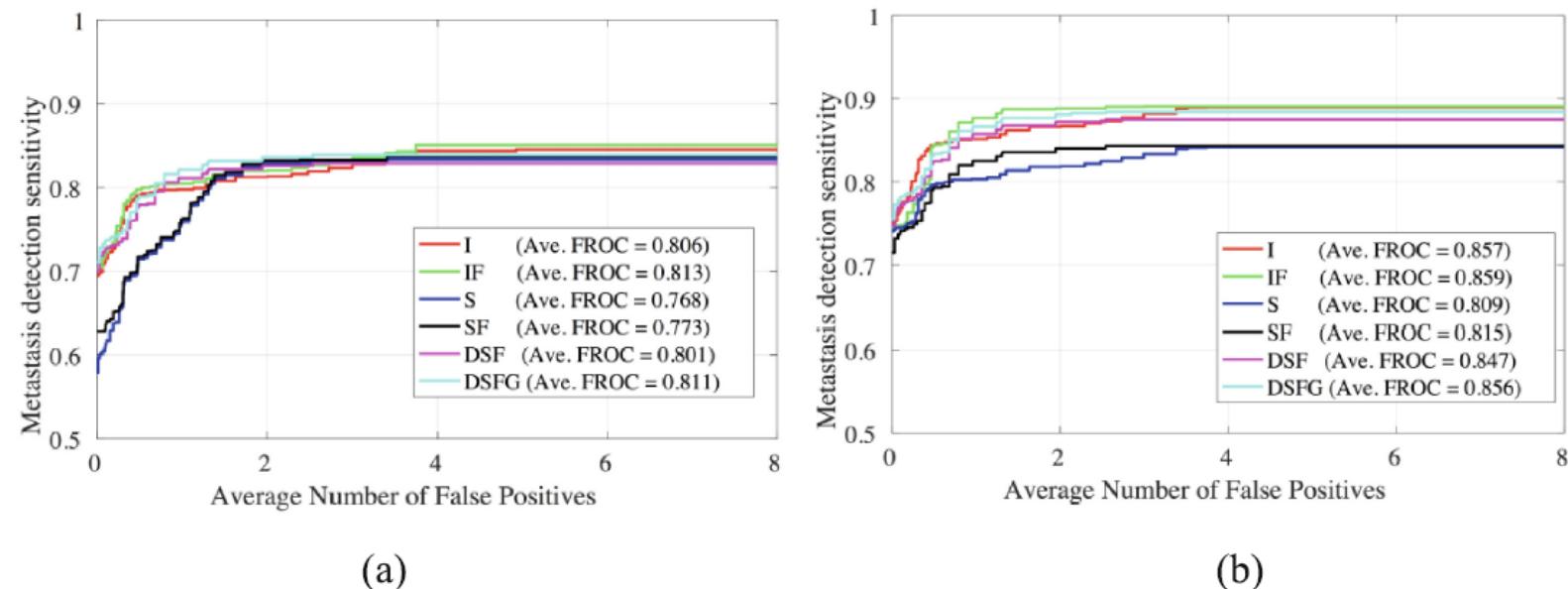


Fig. 2. Experimental results of the methods on the (a) gastric cancer and (b) CAMELYON16 datasets

Adversarial Domain Adaptation for Classification of Prostate Histopathology Whole-Slide Images

Jian Ren^{1(✉)}, Ilker Hacihaliloglu², Eric A. Singer³, David J. Foran³,
and Xin Qi³

¹ Department of Electrical and Computer Engineering, Rutgers University,
Piscataway, USA

jian.ren0905@rutgers.edu

Department of Biomedical Engineering, Rutgers University, Piscataway, USA

³ Rutgers Cancer Institute of New Jersey, New Brunswick, USA

Abstract. Automatic and accurate Gleason grading of histopathology tissue slides is crucial for prostate cancer diagnosis, treatment, and prognosis. Usually, histopathology tissue slides from different institutions show heterogeneous appearances because of different tissue preparation and staining procedures, thus the predictable model learned from one domain may not be applicable to a new domain directly. Here we propose to adopt unsupervised domain adaptation to transfer the discriminative knowledge obtained from the source domain to the target domain without requiring labeling of images at the target domain. The adaptation is achieved through adversarial training to find an invariant feature space along with the proposed Siamese architecture on the target domain to add a regularization that is appropriate for the whole-slide images. We validate the method on two prostate cancer datasets and obtain significant classification improvement of Gleason scores as compared with the baseline models.

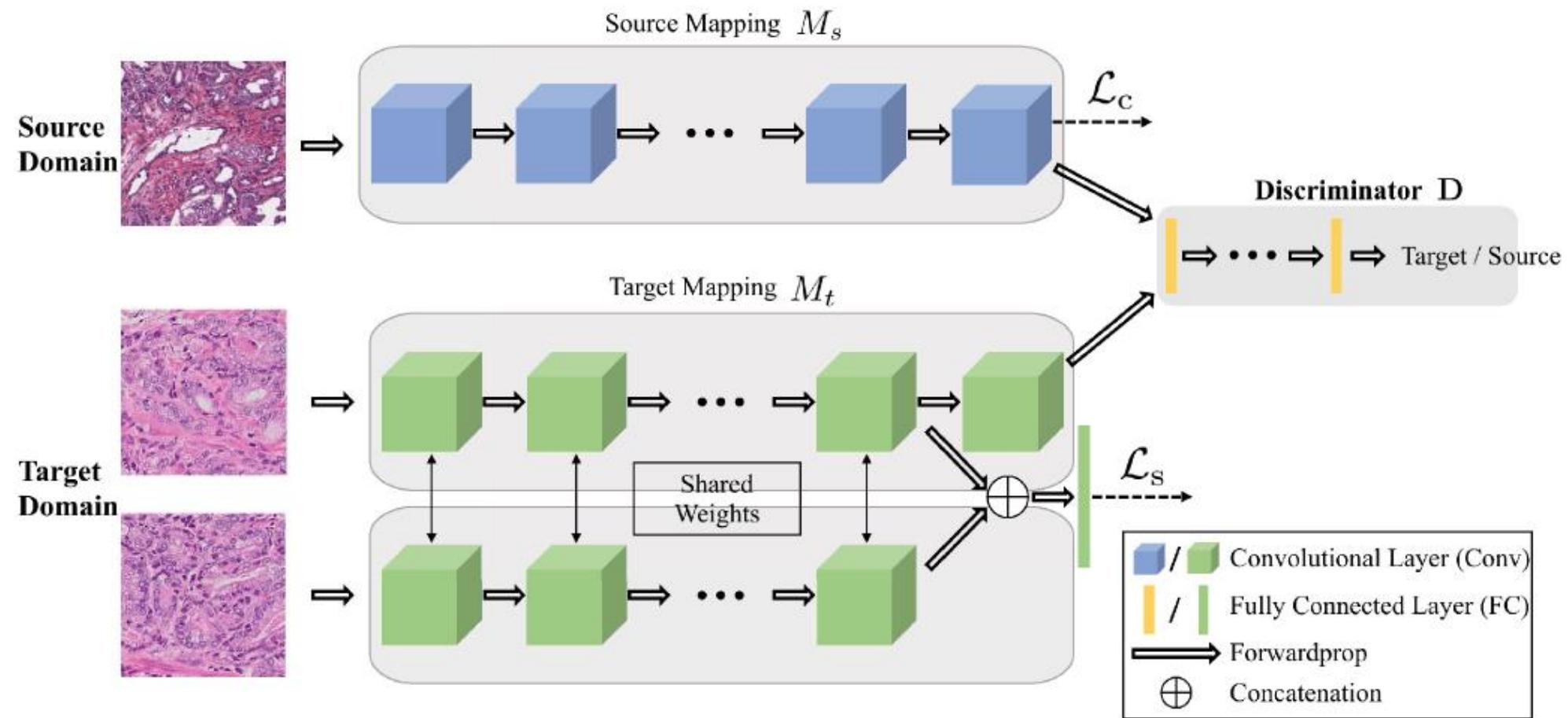


Fig. 1. The architecture of the networks for the unsupervised domain adaptation. The source network and the target network map the input samples into the feature space. The adaptation is accomplished by jointly training the discriminator and target network using the GAN loss to find the domain invariant feature. A Siamese network at target domain adds constraints for the WSIs.

Multiple Instance Learning for Heterogeneous Images: Training a CNN for Histopathology

Heather D. Couture^{1(✉)}, J. S. Marron^{2,3}, Charles M. Perou^{2,4},
Melissa A. Troester^{2,5}, and Marc Niethammer^{1,6}

¹ Department of Computer Science, University of North Carolina,
Chapel Hill, NC, USA
`heather@cs.unc.edu`

² Lineberger Comprehensive Cancer Center, University of North Carolina,
Chapel Hill, NC, USA

³ Department of Statistics and Operations Research, University of North Carolina,
Chapel Hill, NC, USA

⁴ Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

⁵ Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

⁶ Biomedical Research Imaging Center, University of North Carolina,
Chapel Hill, NC, USA

Abstract. Multiple instance (MI) learning with a convolutional neural network enables end-to-end training in the presence of weak image-level labels. We propose a new method for aggregating predictions from smaller regions of the image into an image-level classification by using the quantile function. The quantile function provides a more complete description of the heterogeneity within each image, improving image-level classification. We also adapt image augmentation to the MI framework by randomly selecting cropped regions on which to apply MI aggregation during each epoch of training. This provides a mechanism to study the importance of MI learning. We validate our method on five different classification tasks for breast tumor histology and provide a visualization method for interpreting local image classifications that could lead to future insights into tumor heterogeneity.

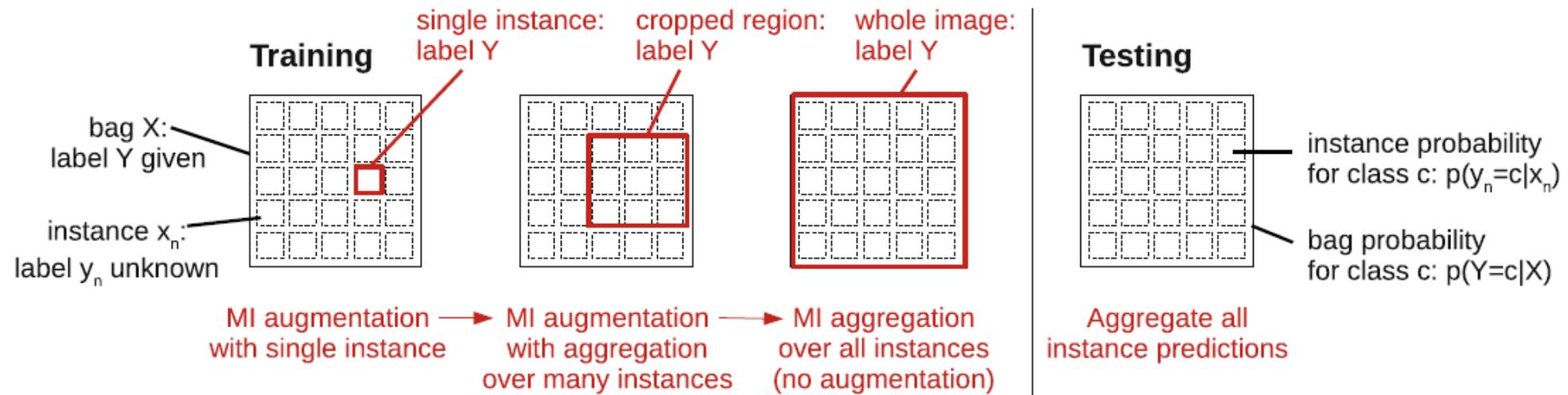


Fig. 1. In MI learning, each bag contains one or more instances. Labels are given for the bag, but not the instances. MI augmentation is a technique to provide additional training samples by randomly selecting a cropped image region and the instances within it. When the bag label is applied to a small number of instances, it is weak because this small region may not be representative of the bag class. Applying the bag label to larger cropped regions provides a stronger label, while still providing benefit from image augmentation. Training with the whole image maximizes the opportunity for MI learning, but restricts the benefits of image augmentation. *At test time, the whole image is processed and the predictions from all instances are aggregated into a bag prediction.*

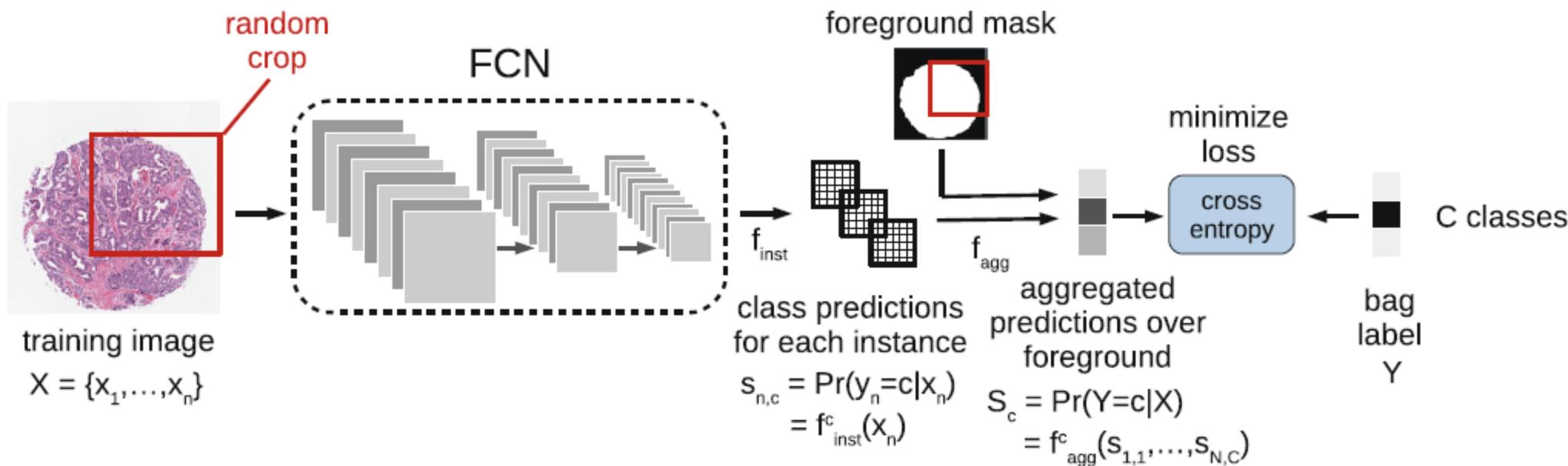


Fig. 2. During training, a cropped region of a given size is randomly selected. An FCN is applied to predict the class, producing a grid of instance predictions. The instance predictions are aggregated over the foreground of the image (as indicated by the foreground mask) using quantile aggregation to predict the class of the cropped image region. With a cross entropy loss applied, backpropagation then learns the FCN and aggregation function weights. At test time, the whole image is used.

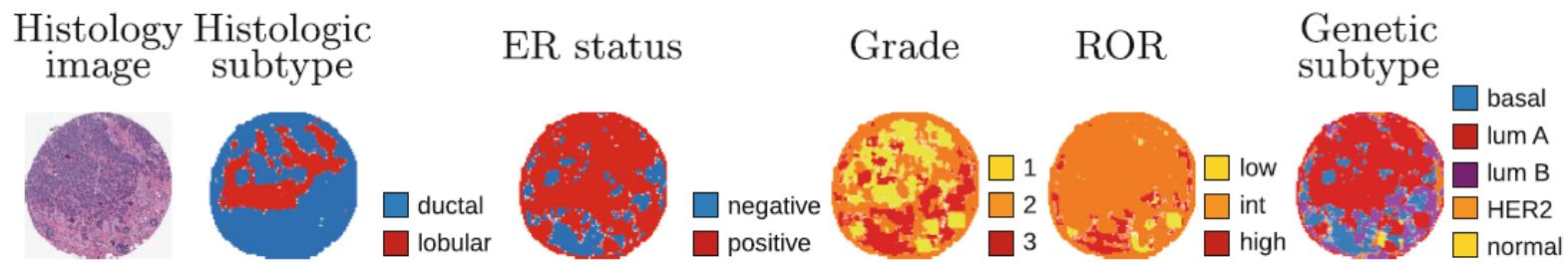


Fig. 4. Visualization of instance predictions for a sample with ground truth labels of ductal, ER positive, grade 1, low ROR, and luminal A.

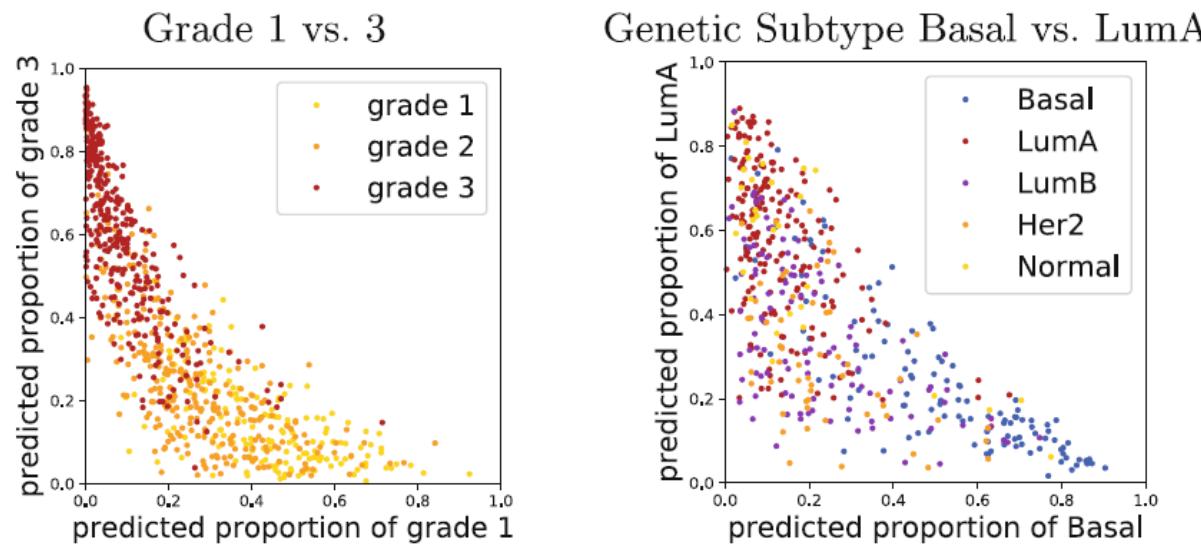


Fig. 5. Predicted heterogeneity for grade 1 vs. 3 and genetic subtype basal vs luminal A. The predicted proportion for each class is calculated as the proportion of instances in the sample predicted to be from each class. Test samples for all classes are plotted.

Spatially Localized Atlas Network Tiles Enables 3D Whole Brain Segmentation from Limited Data

Yuankai Huo¹(✉), Zhoubing Xu¹, Katherine Aboud¹, Prasanna Parvathaneni¹, Shunxing Bao¹, Camilo Bermudez¹, Susan M. Resnick², Laurie E. Cutting¹, and Bennett A. Landman¹

¹ Vanderbilt University, Nashville, TN, USA
yuankai.huo@vanderbilt.edu

² Laboratory of Behavioral Neuroscience, National Institute on Aging,
Baltimore, MD, USA

Abstract. Whole brain segmentation on a structural magnetic resonance imaging (MRI) is essential in non-invasive investigation for neuroanatomy. Historically, multi-atlas segmentation (MAS) has been regarded as the *de facto* standard method for whole brain segmentation. Recently, deep neural network approaches have been applied to whole brain segmentation by learning random patches or 2D slices. Yet, few previous efforts have been made on detailed whole brain segmentation using 3D networks due to the following challenges: (1) fitting entire whole brain volume into 3D networks is restricted by the current GPU memory, and (2) the large number of targeting labels (e.g., >100 labels) with limited number of training 3D volumes (e.g., <50 scans). In this paper, we propose the spatially localized atlas network tiles (SLANT) method to distribute multiple independent 3D fully convolutional networks to cover overlapped sub-spaces in a standard atlas space. This strategy simplifies the whole brain learning task to localized sub-tasks, which was enabled by combining canonical registration and label fusion techniques with deep learning. To address the second challenge, auxiliary labels on 5111 initially unlabeled scans were created by MAS for pre-training. From empirical validation, the state-of-the-art MAS method achieved mean Dice value of 0.76, 0.71, and 0.68, while the proposed method achieved 0.78, 0.73, and 0.71 on three validation cohorts. Moreover, the computational time reduced from >30 h using MAS to ≈15 min using the proposed method. The source code is available online (<https://github.com/MASILab/SLANTbrainSeg>).

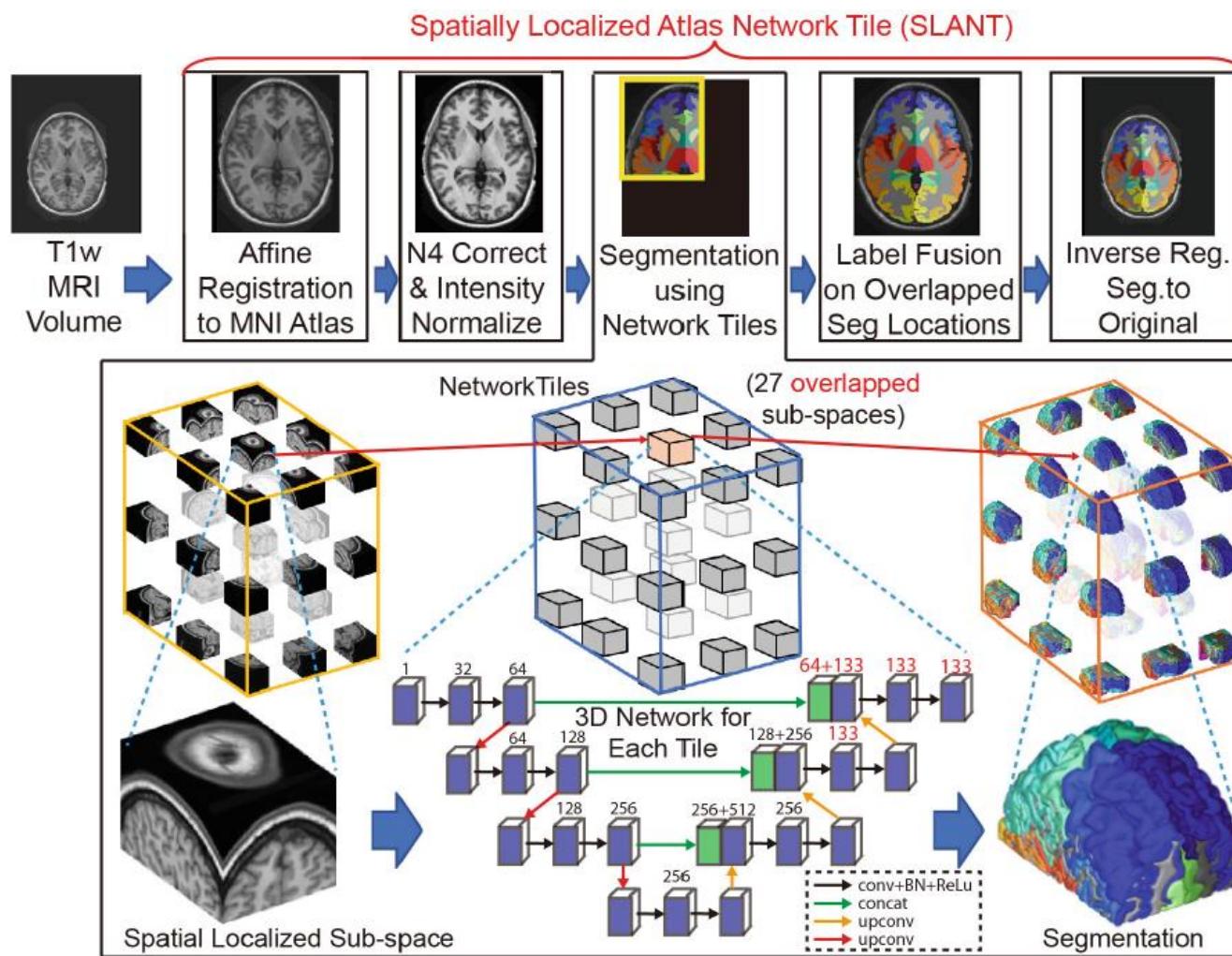


Fig. 1. The proposed SLANT-27 (27 network tiles) method is presented, which combines canonical medical image processing methods (registration, harmonization, label fusion) with 3D network tiles. 3D U-Net is used as each tile, whose deconvolutional channel numbers are modified to 133. The tiles are spatially overlapped in MNI space.

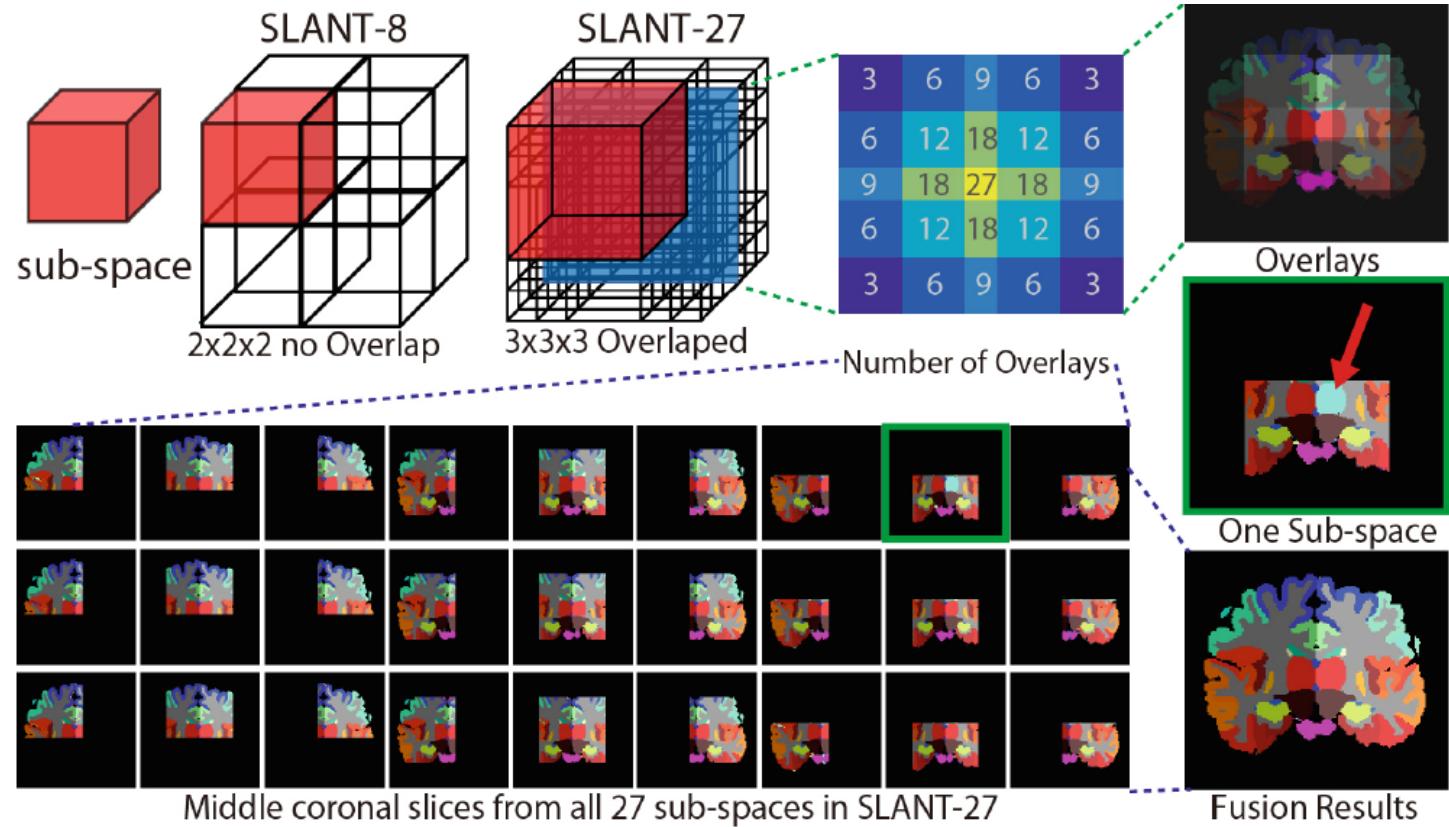


Fig. 2. SLANT-8 covered eight non-overlapped sub-spaces in MNI, while SLANT-27 covered 27 overlapped sub-spaces in MNI. Middle coronal slices from all 27 sub-spaces were visualized (lower left panel). The number of overlays, as well as sub-spaces overlays, were showed (upper right panel). The incorrect labels (red arrow) in one sub-space were corrected in final segmentation by performing majority vote label fusion.

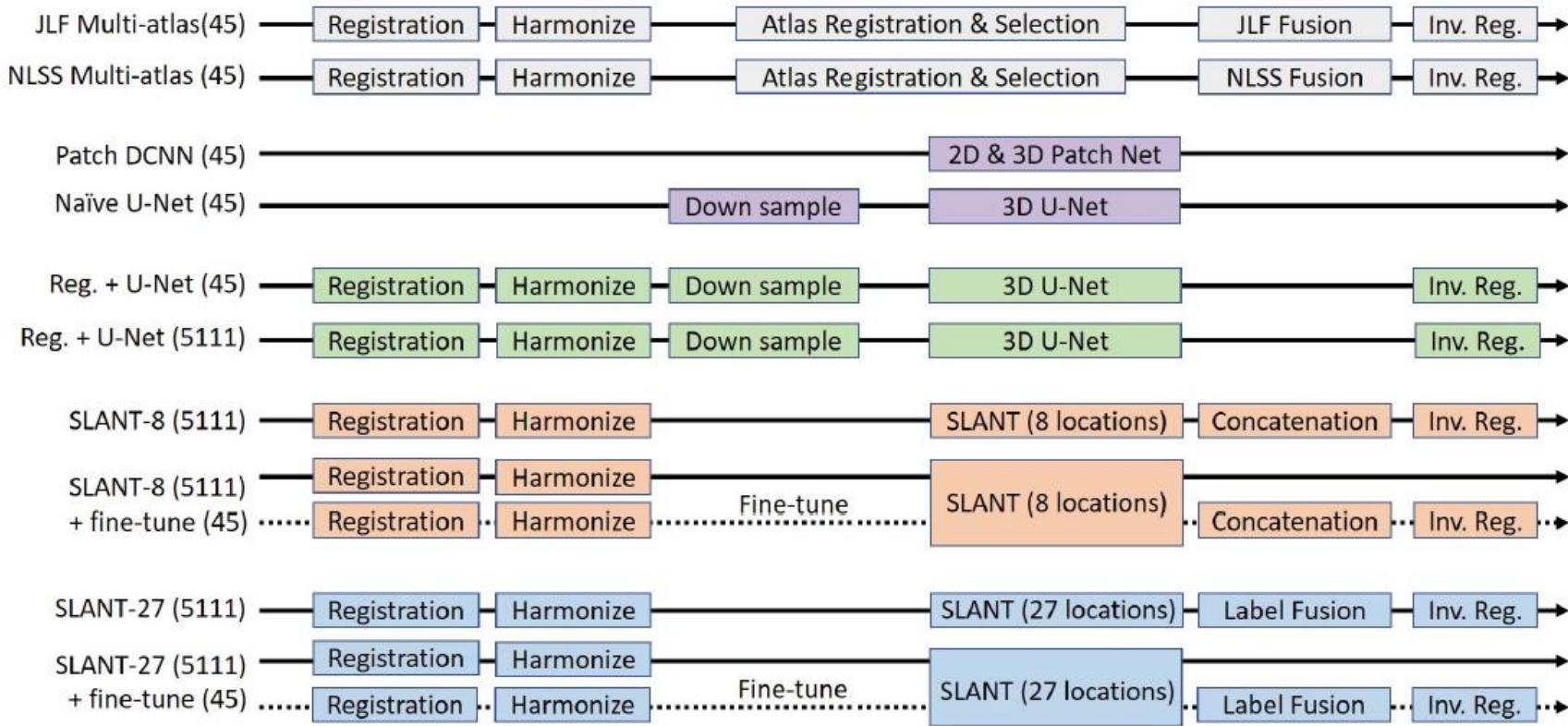


Fig. 3. This figure demonstrates the major components of different segmentation methods. (45) indicated the 45 OASIS manually traced images were used in training, while (5111) indicated the 5111 auxiliary label images were used in training.

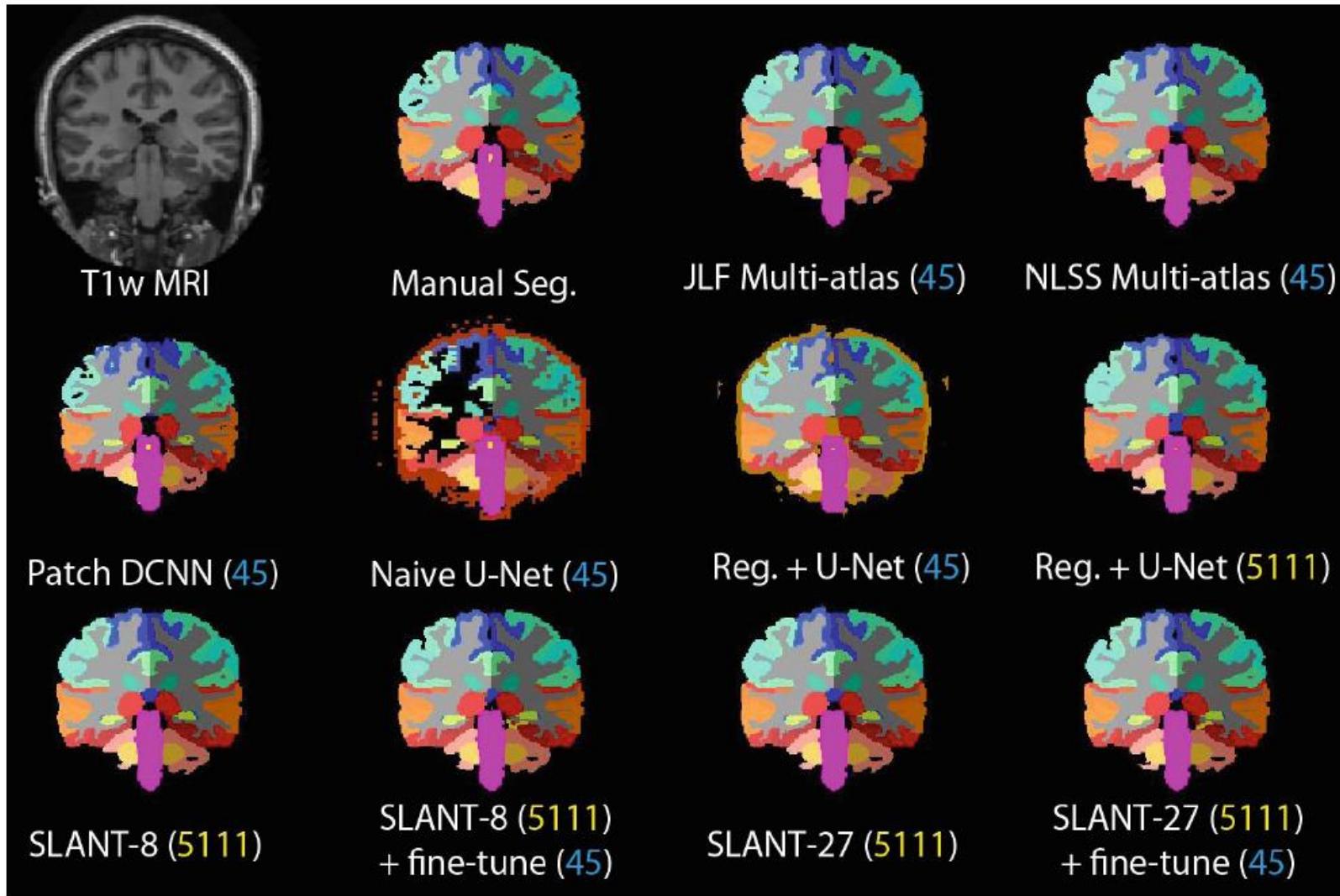


Fig. 4. Qualitative results of manual segmentation, MAS methods, patch-based DCNN method, U-Net approaches and proposed SLANT methods.

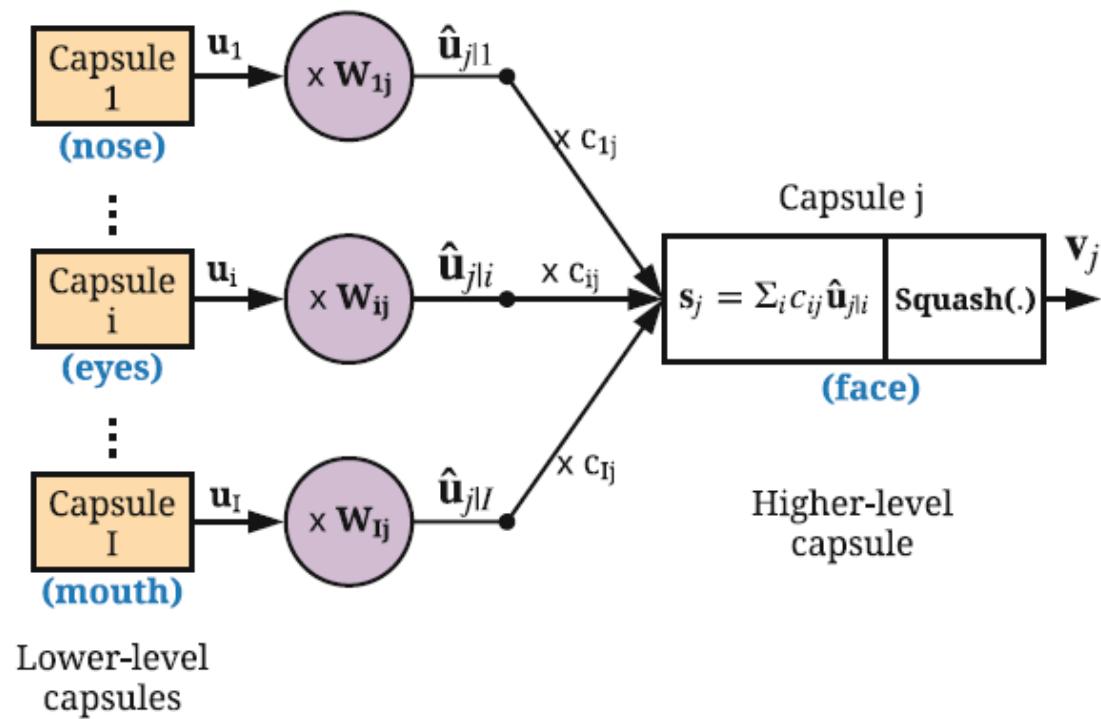
Fast CapsNet for Lung Cancer Screening

Aryan Mobiny and Hien Van Nguyen^(✉)

Department of Electrical and Computer Engineering, University of Houston,
Houston, USA
hvnguy35@central.uh.edu

Abstract. Lung cancer is the leading cause of cancer-related deaths in the past several years. A major challenge in lung cancer screening is the detection of lung nodules from computed tomography (CT) scans. State-of-the-art approaches in automated lung nodule classification use deep convolutional neural networks (CNNs). However, these networks require a large number of training samples to generalize well. This paper investigates the use of capsule networks (CapsNets) as an alternative to CNNs. We show that CapsNets significantly outperforms CNNs when the number of training samples is small. To increase the computational efficiency, our paper proposes a consistent dynamic routing mechanism that results in $3\times$ speedup of CapsNet. Finally, we show that the original image reconstruction method of CapNets performs poorly on lung nodule data. We propose an efficient alternative, called convolutional decoder, that yields lower reconstruction error and higher classification accuracy.

Abstract. Lung cancer is the leading cause of cancer-related deaths in the past several years. A major challenge in lung cancer screening is the detection of lung nodules from computed tomography (CT) scans. State-of-the-art approaches in automated lung nodule classification use deep convolutional neural networks (CNNs). However, these networks require a large number of training samples to generalize well. This paper investigates the use of capsule networks (CapsNets) as an alternative to CNNs. We show that CapsNets significantly outperforms CNNs when the number of training samples is small. To increase the computational efficiency, our paper proposes a consistent dynamic routing mechanism that results in $3\times$ speedup of CapsNet. Finally, we show that the original image reconstruction method of CapNets performs poorly on lung nodule data. We propose an efficient alternative, called convolutional decoder, that yields lower reconstruction error and higher classification accuracy.

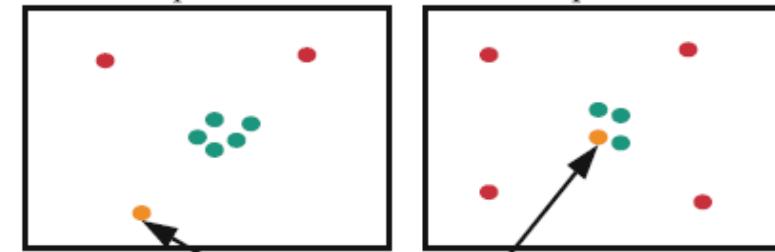


Higher-level
capsules

Capsule j

Capsule A

Capsule B



Capsule i
Lower-level capsule

Fig. 1. Left: connections between the lower and higher-level capsules, **Right:** dynamic routing for sending information from a lower-level capsule to higher-level ones.

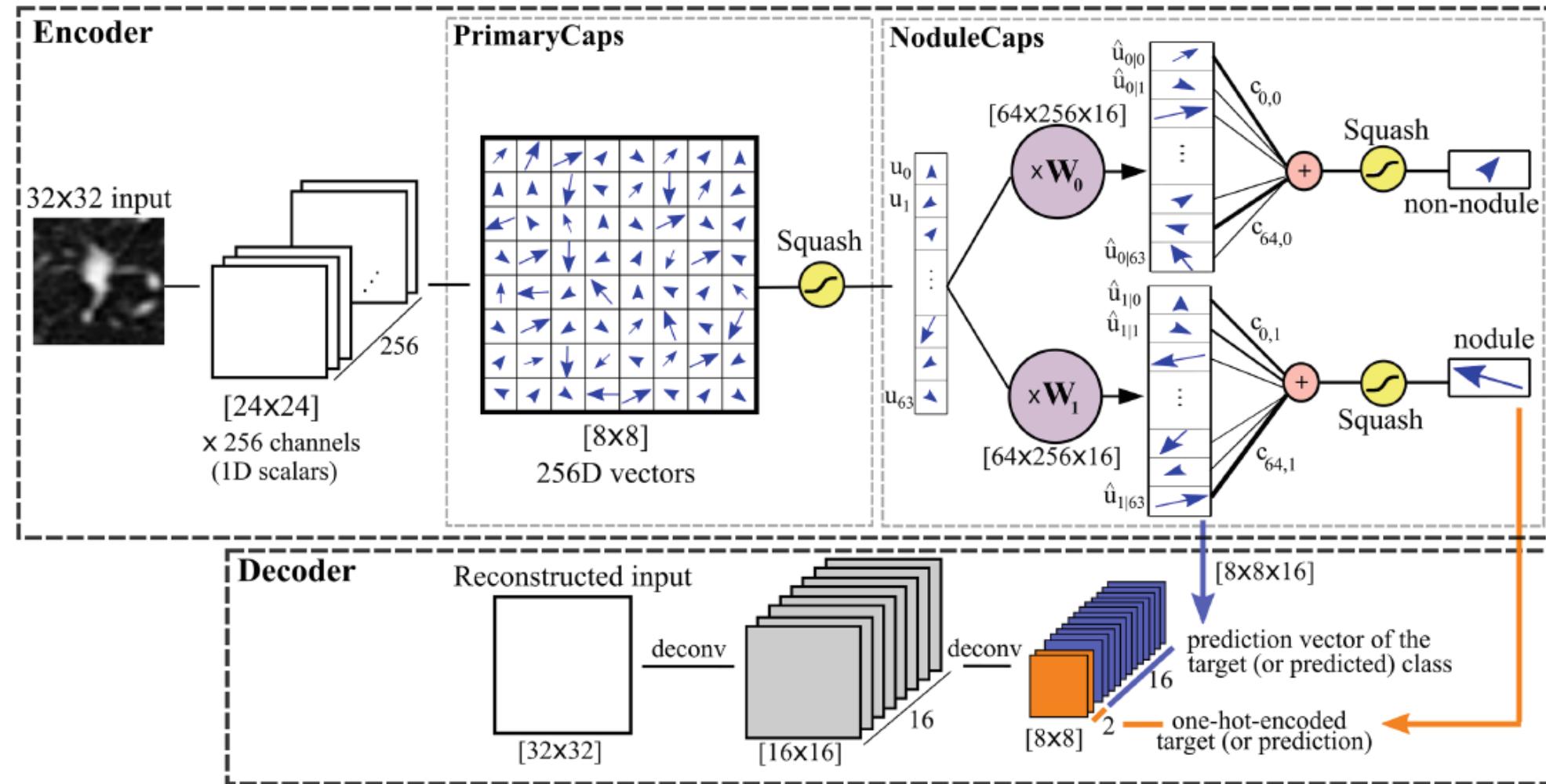


Fig. 2. Visual representation of Fast Capsule Network

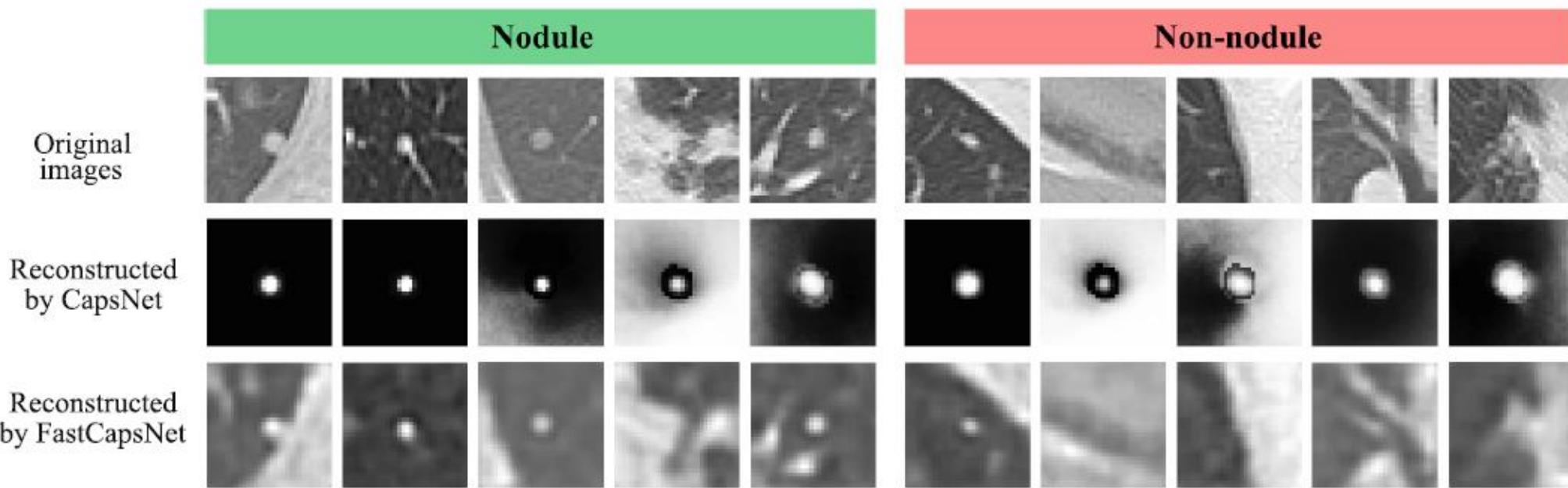


Fig. 3. Sample images of nodules (left) and non-nodules (right) and their reconstructions using the original CapsNet (middle row) and the FastCapsNet (last row)

Automated Pulmonary Nodule Detection: High Sensitivity with Few Candidates

Bin Wang^{1,3}, Guojun Qi², Sheng Tang^{1(✉)}, Liheng Zhang², Lixi Deng^{1,3},
and Yongdong Zhang¹

Key Lab of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, China

ts@ict.ac.cn

² University of Central Florida, Orlando, FL, USA

³ University of the Chinese Academy of Sciences, Beijing, China

Abstract. Automated pulmonary nodule detection plays an important role in lung cancer diagnosis. In this paper, we propose a pulmonary detection framework that can achieve high sensitivity with few candidates. First, the Feature Pyramid Network (FPN), which leverages multi-level features, is applied to detect nodule candidates that cover almost all true positives. Then redundant candidates are removed by a simple but effective Conditional 3-Dimensional Non-Maximum Suppression (Conditional 3D-NMS). Moreover, a novel Attention 3D CNN (Attention 3D-CNN) which efficiently utilizes contextual information is proposed to further remove the overwhelming majority of false positives. The proposed method yields a sensitivity of 95.8% at 2 false positives per scan on the Lung Nodule Analysis 2016 (LUNA16) dataset, which is competitive compared to the current published state-of-the-art methods.

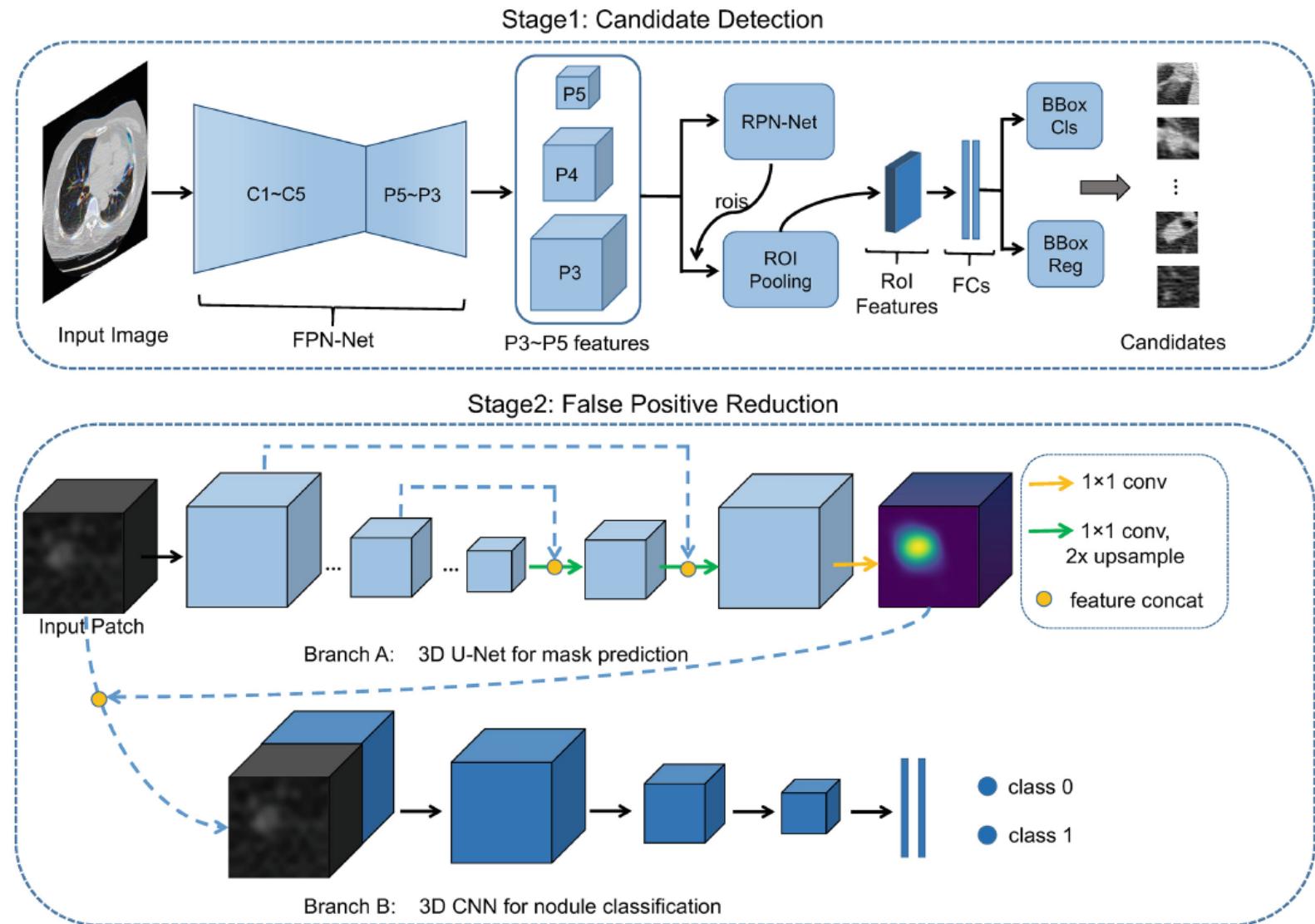


Fig. 1. The framework of proposed pulmonary nodule detection system.

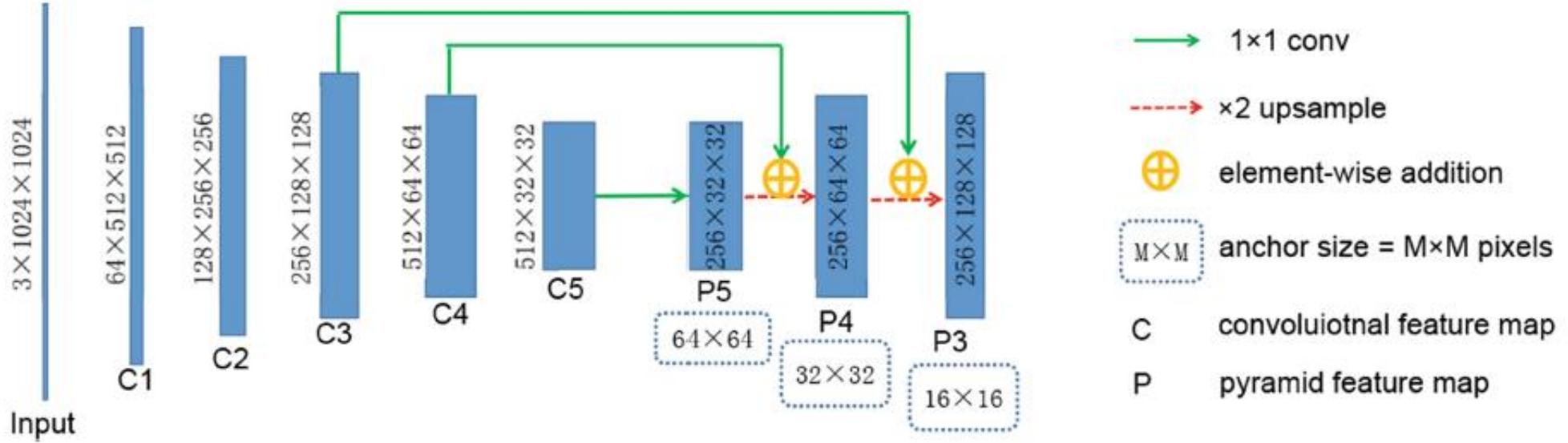


Fig. 2. Feature pyramid networks in our candidate detection architecture.

Deep Learning from Label Proportions for Emphysema Quantification

Gerda Bortsova¹(✉), Florian Dubost¹, Silas Ørting², Ioannis Katramados³, Laurens Hogeweg³, Laura Thomsen⁴, Mathilde Wille⁵, and Marleen de Bruijne^{1,2}

¹ Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, The Netherlands
gerdabortsova@gmail.com

² Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

³ COSMONiO, Groningen, The Netherlands

⁴ Department of Respiratory Medicine, Hvidovre Hospital, Hvidovre, Denmark

⁵ Department of Diagnostic Imaging, Bispebjerg Hospital, Copenhagen, Denmark

Abstract. We propose an end-to-end deep learning method that learns to estimate emphysema extent from proportions of the diseased tissue. These proportions were visually estimated by experts using a standard grading system, in which grades correspond to intervals (label example: 1–5% of diseased tissue). The proposed architecture encodes the knowledge that the labels represent a volumetric proportion. A custom loss is designed to learn with intervals. Thus, during training, our network learns to segment the diseased tissue such that its proportions fit the ground truth intervals. Our architecture and loss combined improve the performance substantially (8% ICC) compared to a more conventional regression network. We outperform traditional lung densitometry and two recently published methods for emphysema quantification by a large margin (at least 7% AUC and 15% ICC), and achieve near-human-level performance. Moreover, our method generates emphysema segmentations that predict the spatial distribution of emphysema at human level.

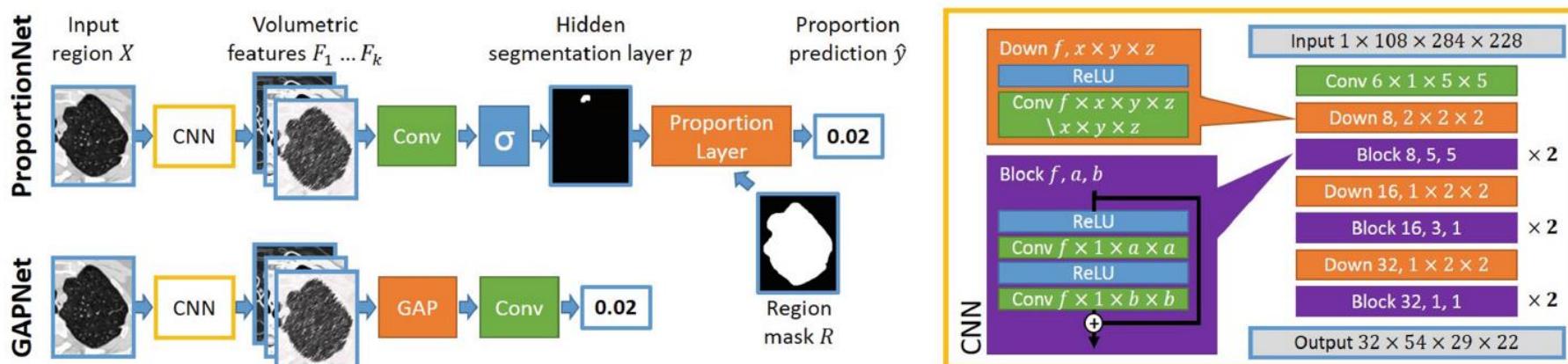


Fig. 1. (a) “Conv”: $1 \times 1 \times 1$ convolution with one output feature; “GAP”: global average pooling; σ : sigmoid. (b) “Conv”: valid convolutions with parameters “{# of output features}, {kernel size} / {stride}”; “Block”: residual blocks [5].

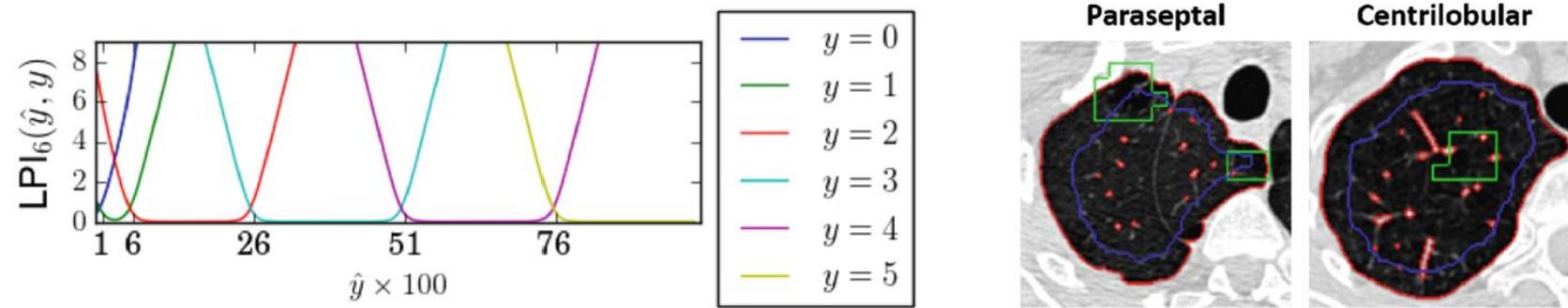


Fig. 2. *Left:* LPI₆ loss with all $w_c = 1$, $\alpha = 120$ and $\text{thresh} = (0, 0.005, 0.055, 0.255, 0.505, 0.755, 1)$. *Right:* images with different predominant emphysema patterns. Green: ProportionNet segmentations; red: region mask; blue: the 10px margin for separating near-boundary detections from the rest.

S4ND: Single-Shot Single-Scale Lung Nodule Detection

Naji Khosravan^(✉) and Ulas Bagci

Center for Research in Computer Vision (CRCV), School of Computer Science,
University of Central Florida, Orlando, FL, USA
naji.khosravan@gmail.com

Abstract. The most recent lung nodule detection studies rely on computationally expensive multi-stage frameworks to detect nodules from CT scans. To address this computational challenge and provide better performance, in this paper we propose S4ND, a new deep learning based method for lung nodule detection. Our approach uses a single feed forward pass of a single network for detection. The whole detection pipeline is designed as a single 3D Convolutional Neural Network (CNN) with dense connections, trained in an end-to-end manner. S4ND does not require any further post-processing or user guidance to refine detection results. Experimentally, we compared our network with the current state-of-the-art object detection network (SSD) in computer vision as well as the state-of-the-art published method for lung nodule detection (3D DCNN). We used publicly available 888 CT scans from LUNA challenge dataset and showed that the proposed method outperforms the current literature both in terms of efficiency and accuracy by achieving an average FROC-score of 0.897. We also provide an in-depth analysis of our proposed network to shed light on the unclear paradigms of tiny object detection.

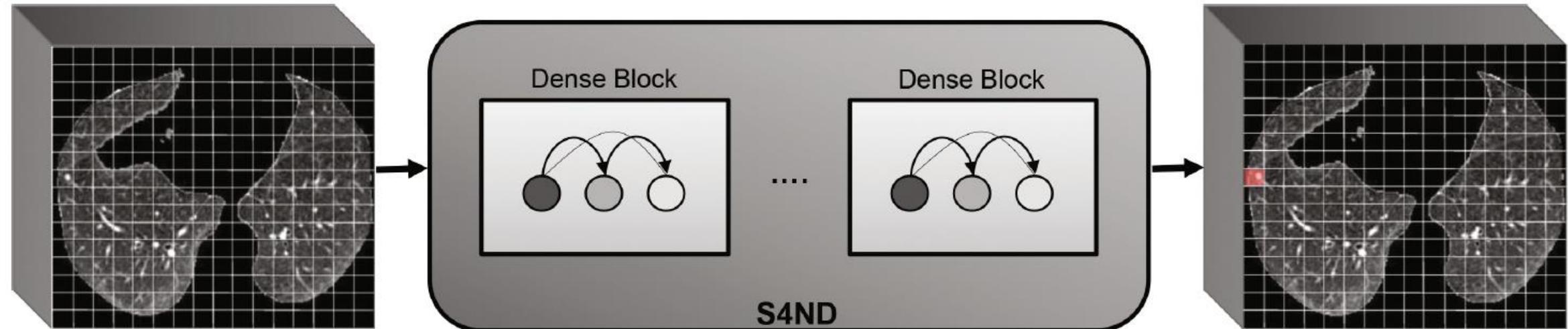


Fig. 1. Our framework, named S4ND, models nodule detection as a cell-wise classification of the input volume. The input volume is divided by a $16 \times 16 \times 8$ grid and is passed through a newly designed 3D dense CNN. The output is a probability map indicating the presence of a nodule in each cell.

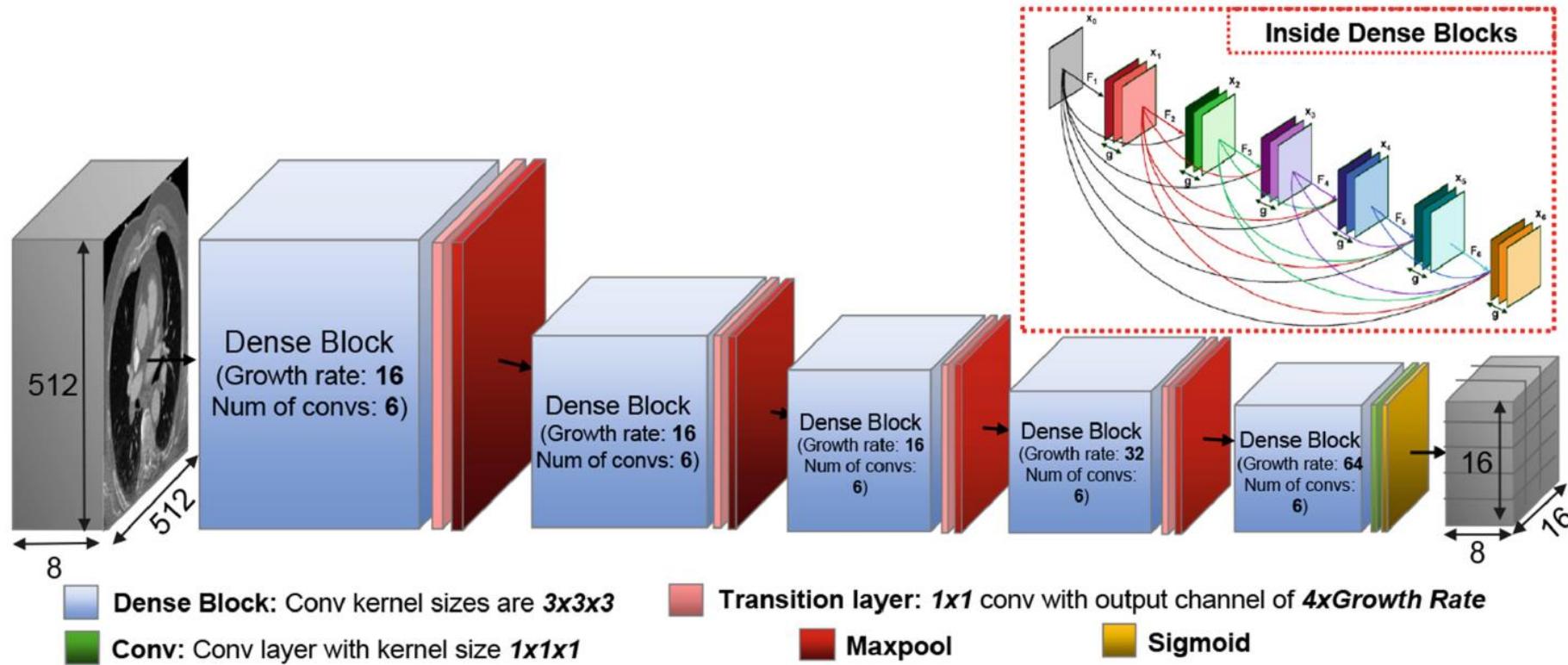


Fig. 2. Input to the network is a $512 \times 512 \times 8$ volume and output is a $16 \times 16 \times 8$ probability map representing likelihood of nodule presence. Our network has 5 dense blocks each having 6 conv. layers. The growth rates of blocks 1 to 5 is 16, 16, 16, 32, 64 respectively. The network has 4 transition layers and 4 max-pooling layers. The last block is followed by a convolution layer with kernel size $1 \times 1 \times 1$ and output channel of 1 and a sigmoid activation function.

Brain Biomarker Interpretation in ASD Using Deep Learning and fMRI

Xiaoxiao Li^{1(✉)}, Nicha C. Dvornek⁴, Juntang Zhuang¹, Pamela Ventola⁵,
and James S. Duncan^{1,2,3,4}

¹ Biomedical Engineering, Yale University, New Haven, CT, USA
xiaoxiao.li@yale.edu

² Electrical Engineering, Yale University, New Haven, CT, USA

³ Department of Statistics & Data Science, Yale University, New Haven, CT, USA
Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, US

⁵ Child Study Center, Yale School of Medicine, New Haven, CT, USA

Abstract. Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder. Finding the biomarkers associated with ASD is extremely helpful to understand the underlying roots of the disorder and can lead to earlier diagnosis and more targeted treatment. Although Deep Neural Networks (DNNs) have been applied in functional magnetic resonance imaging (fMRI) to identify ASD, understanding the data driven computational decision making procedure has not been previously explored. Therefore, in this work, we address the problem of interpreting reliable biomarkers associated with identifying ASD; specifically, we propose a 2-stage method that classifies ASD and control subjects using fMRI images and interprets the saliency features activated by the classifier. First, we trained an accurate DNN classifier. Then, for detecting the biomarkers, different from the DNN visualization works in computer vision, we take advantage of the anatomical structure of brain fMRI and develop a frequency-normalized sampling method to corrupt images. Furthermore, in the ASD vs. control subjects classification scenario, we provide a new approach to detect and characterize important brain features into three categories. The biomarkers we found by the proposed method are robust and consistent with previous findings in the literature. We also validate the detected biomarkers by neurological function decoding and comparing with the DNN activation maps.

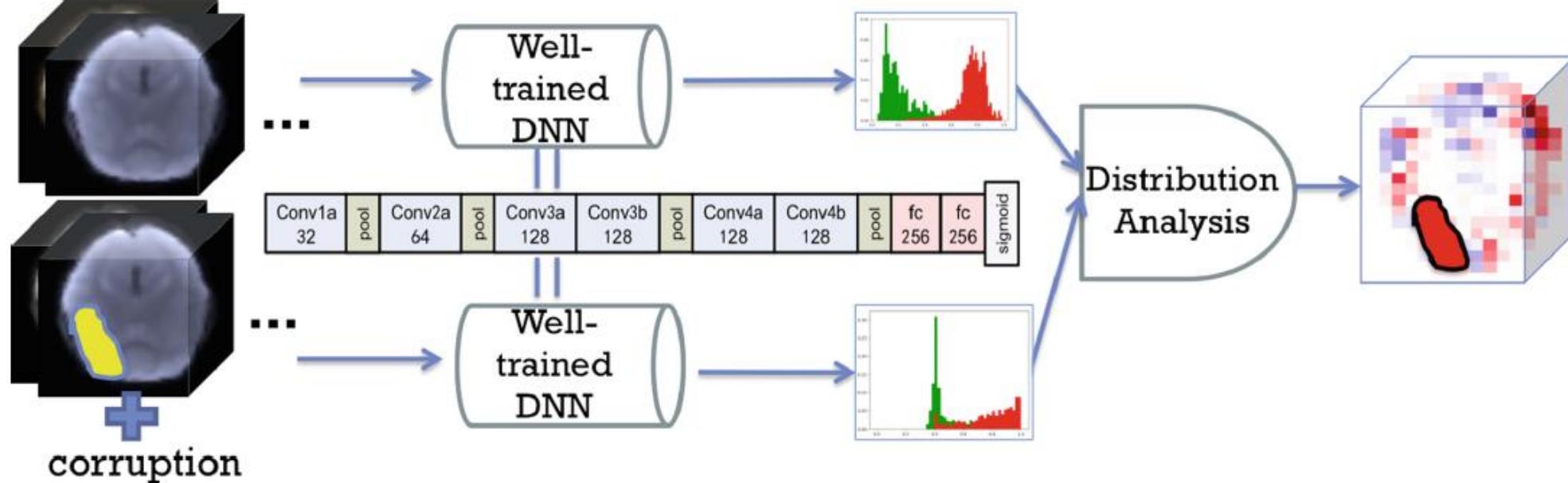
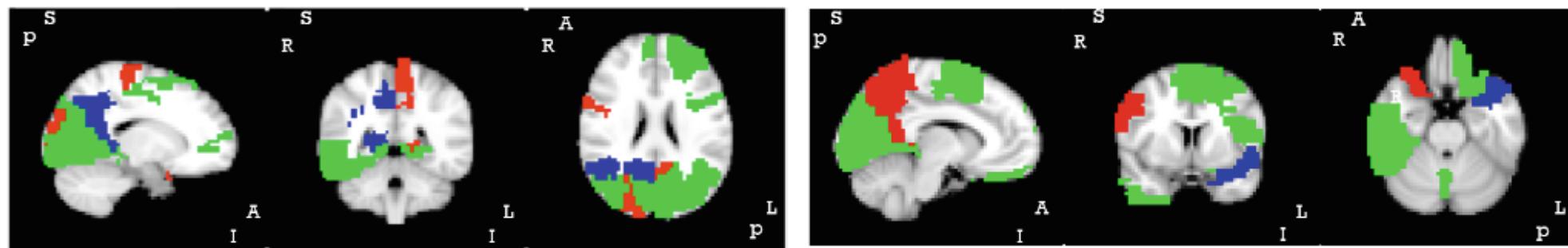
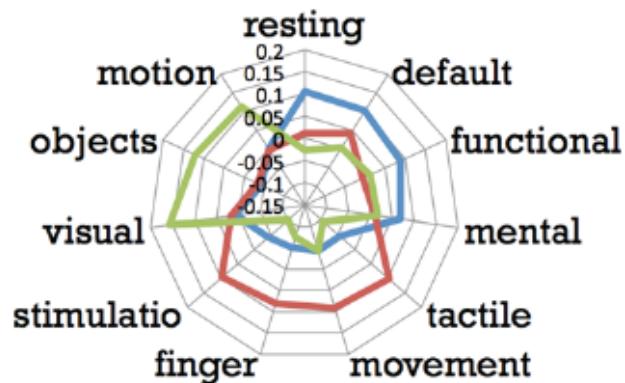


Fig. 1. Pipeline for interpreting important features from a DNN

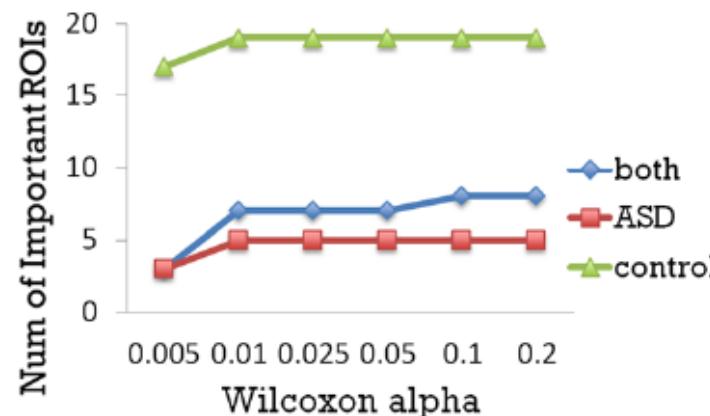


(a) The three groups of saliency regions (biomarkers) from 2 views ($\alpha_{JSD} = 0.05$, $\alpha_W = 0.05$)

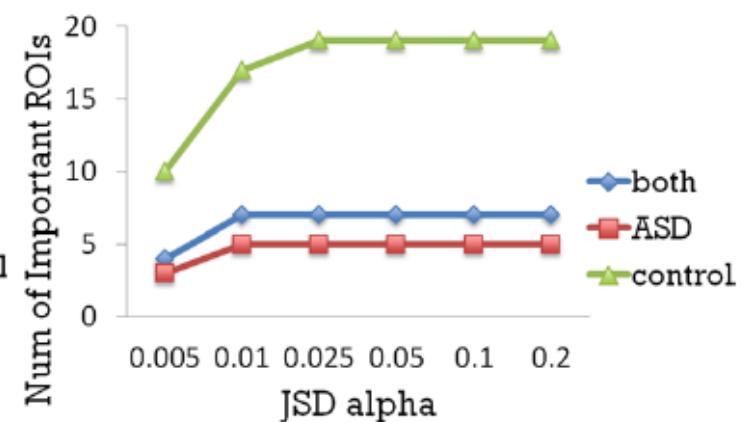
— both — ASD — control



(b) The correlation between the brain region groups and functional keywords decoded by Neurosynth

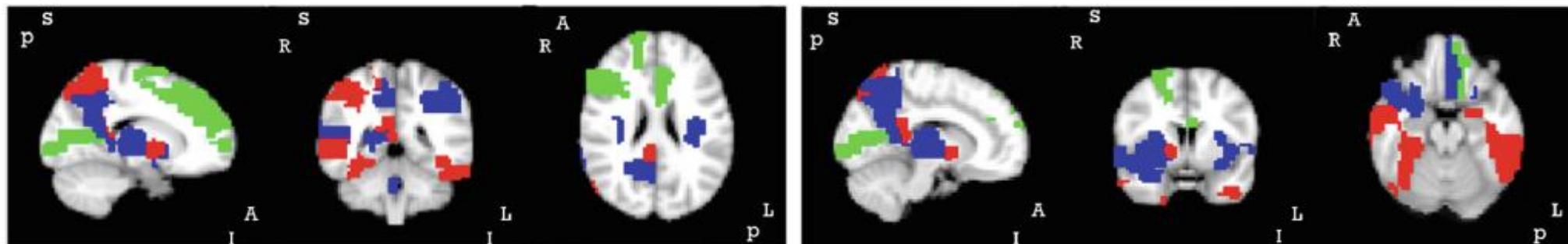


(c) Number of important regions in each group when α_W increases ($\alpha_{JSD} = 0.05$)

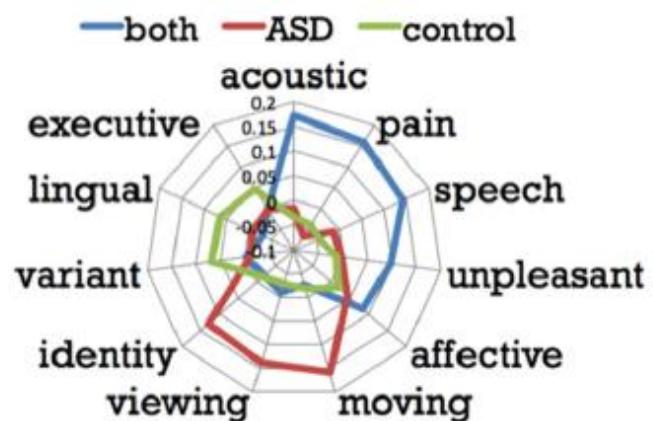


(d) Number of important regions in each group when α_{JSD} increases ($\alpha_W = 0.05$)

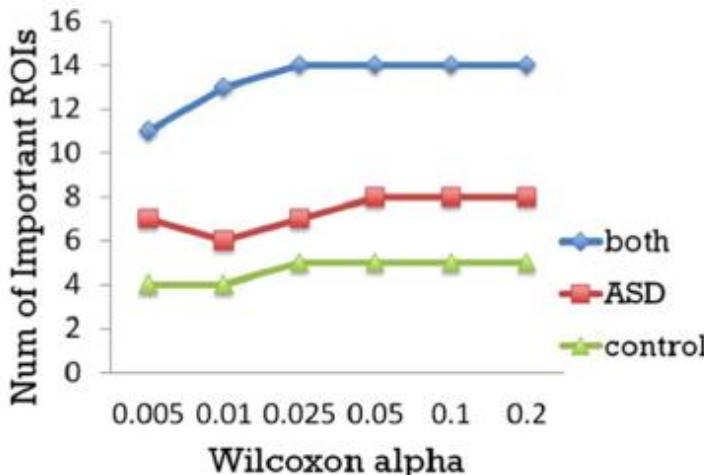
Fig. 3. Important biomarkers detected in biopoint dataset



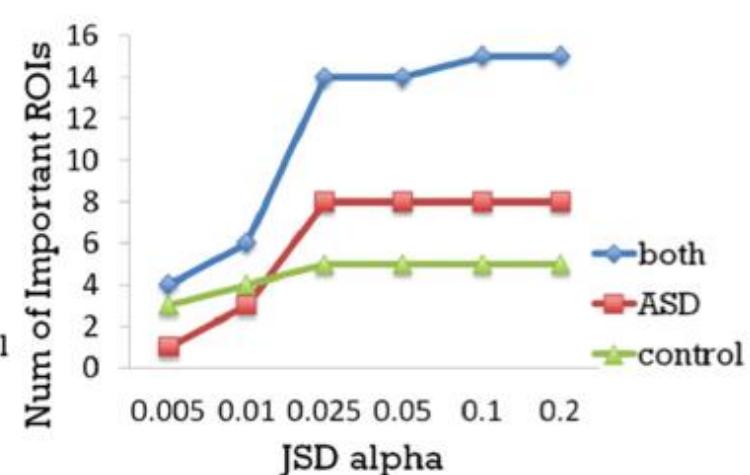
(a) The three groups of saliency regions (biomarkers) from 2 views ($\alpha_{JSD} = 0.05$, $\alpha_w = 0.05$)



(b) The correlation between the brain region groups and functional keywords decoded by Neurosynth



(c) Number of important regions in each group when α_w increases ($\alpha_{JSD} = 0.05$)



(d) Number of important regions in each group when α_{JSD} increases ($\alpha_w = 0.05$)

Fig. 4. Important biomarkers detected in ABIDE dataset

Deep Multi-structural Shape Analysis: Application to Neuroanatomy

Benjamín Gutiérrez-Becker^(✉) and Christian Wachinger

Artificial Intelligence in Medical Imaging (AI-Med), KJP, LMU München
Munich, Germany

benjamin.gutierrez-becker@med.uni-muenchen.de

Abstract. We propose a deep neural network for supervised learning on neuroanatomical shapes. The network directly operates on raw point clouds without the need for mesh processing or the identification of point correspondences, as spatial transformer networks map the data to a canonical space. Instead of relying on *hand-crafted* shape descriptors, an optimal representation is learned in the end-to-end training stage of the network. The proposed network consists of multiple branches, so that features for multiple structures are learned simultaneously. We demonstrate the performance of our method on two applications: (i) the prediction of Alzheimer’s disease and mild cognitive impairment and (ii) the regression of the brain age. Finally, we visualize the important parts of the anatomy for the prediction by adapting the occlusion method to point clouds.

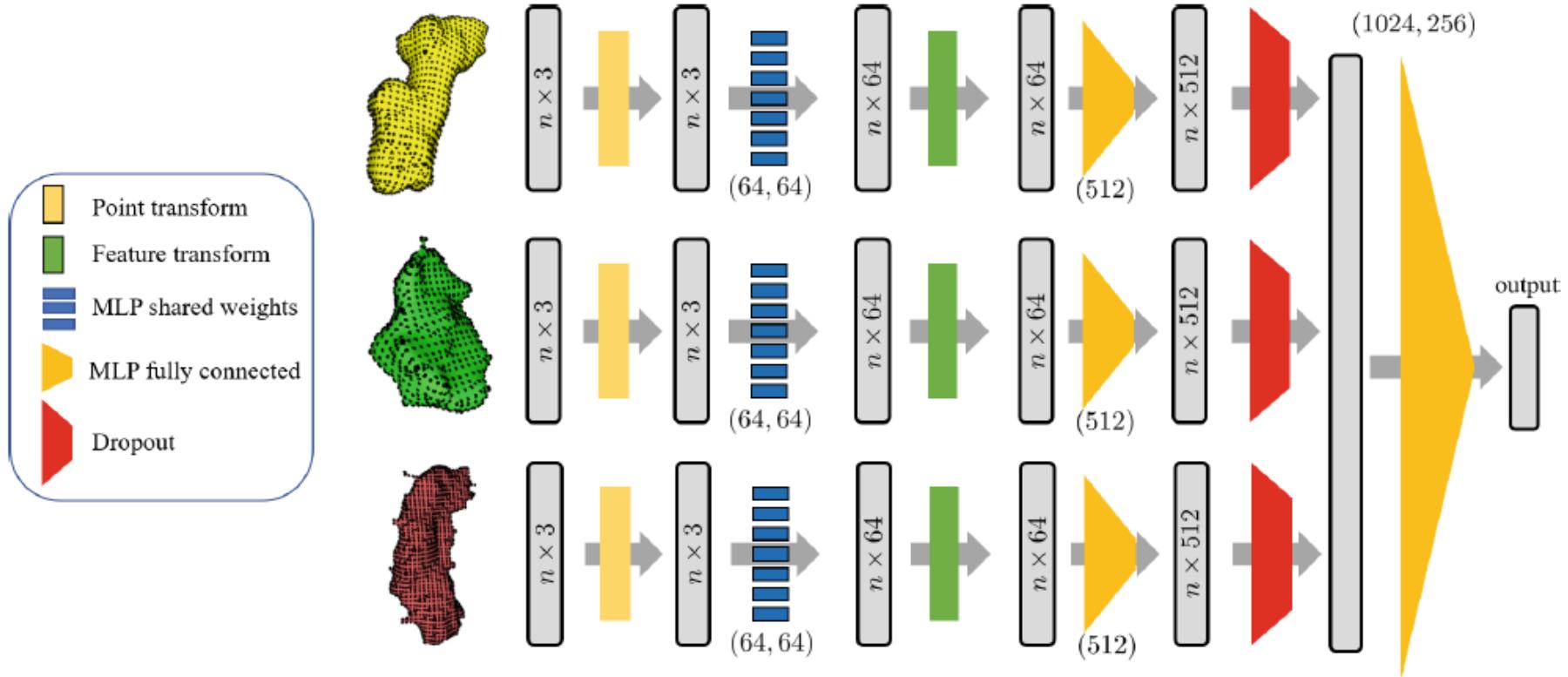


Fig. 1. MSPNet Architecture. The network consists of one branch per structure (illustrated for three structures), which are fused before the final multilayer perceptron (MLP). Each structure is represented by a point cloud with n points that pass through transformer networks and multilayer perceptrons of the individual branch. Numbers in brackets are layer sizes.

CompNet: Complementary Segmentation Network for Brain MRI Extraction

Raunak Dey^(✉) and Yi Hong

Computer Science Department, University of Georgia, Athens, USA
rd31879@uga.edu

Abstract. Brain extraction is a fundamental step for most brain imaging studies. In this paper, we investigate the problem of skull stripping and propose complementary segmentation networks (CompNets) to accurately extract the brain from T1-weighted MRI scans, for both normal and pathological brain images. The proposed networks are designed in the framework of encoder-decoder networks and have two pathways to learn features from both the brain tissue and its complementary part located outside of the brain. The complementary pathway extracts the features in the non-brain region and leads to a robust solution to brain extraction from MRIs with pathologies, which do not exist in our training dataset. We demonstrate the effectiveness of our networks by evaluating them on the OASIS dataset, resulting in the state of the art performance under the two-fold cross-validation setting. Moreover, the robustness of our networks is verified by testing on images with introduced pathologies and by showing its invariance to unseen brain pathologies. In addition, our complementary network design is general and can be extended to address other image segmentation problems with better generalization.

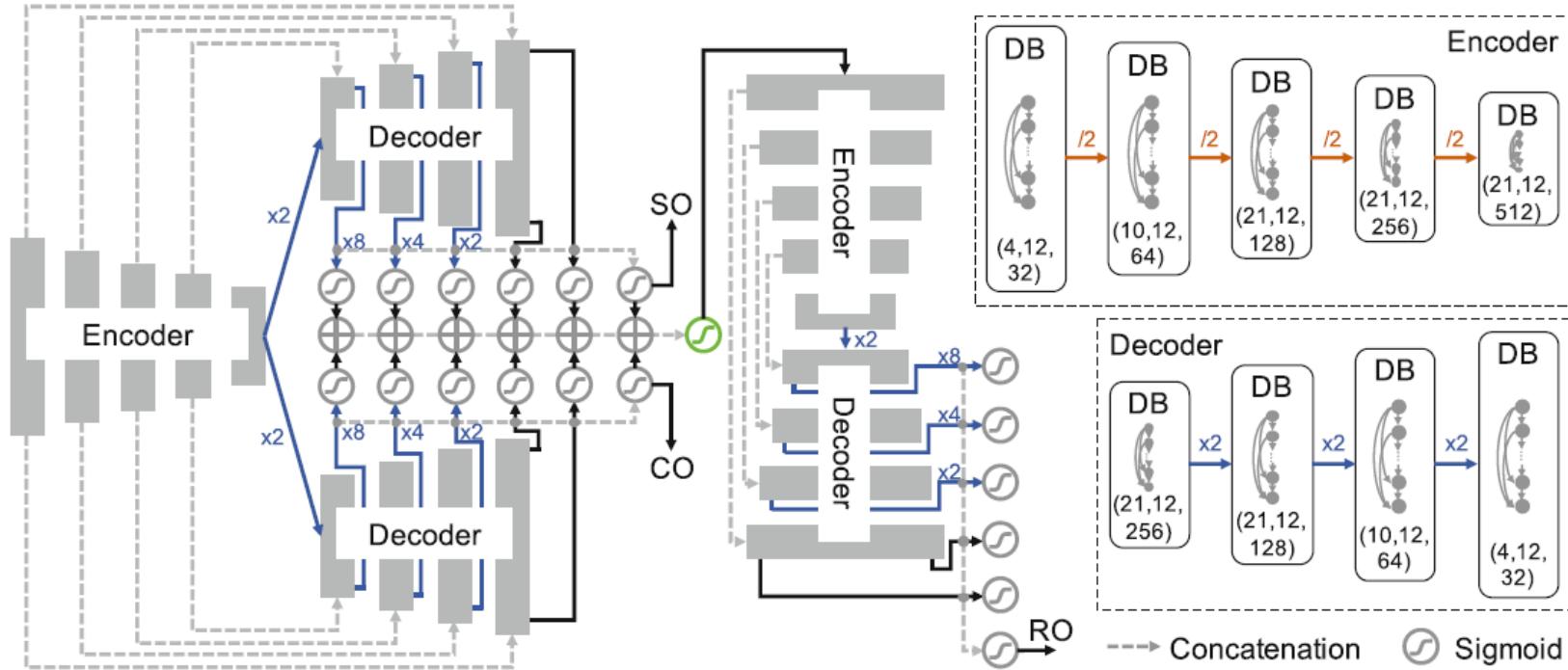


Fig. 1. Architecture of our complementary segmentation network, the optimal CompNet. The dense blocks (DB), corresponding to the gray bars, are used in each encoder and decoder. The triple (x, y, z) in each dense block indicates that it has x convolutional layers with a kernel size 3×3 ; each layer has y filters, except for the last one that has z filters. SO: segmentation output for the brain mask; CO: complementary segmentation output for the non-brain mask; RO: reconstruction output for the input image. These three outputs produced by the Sigmoid function are the final predictions; while all other Sigmoids produce intermediate outputs, except for the green one that generates input for the image reconstruction sub-network. Best viewed in color.

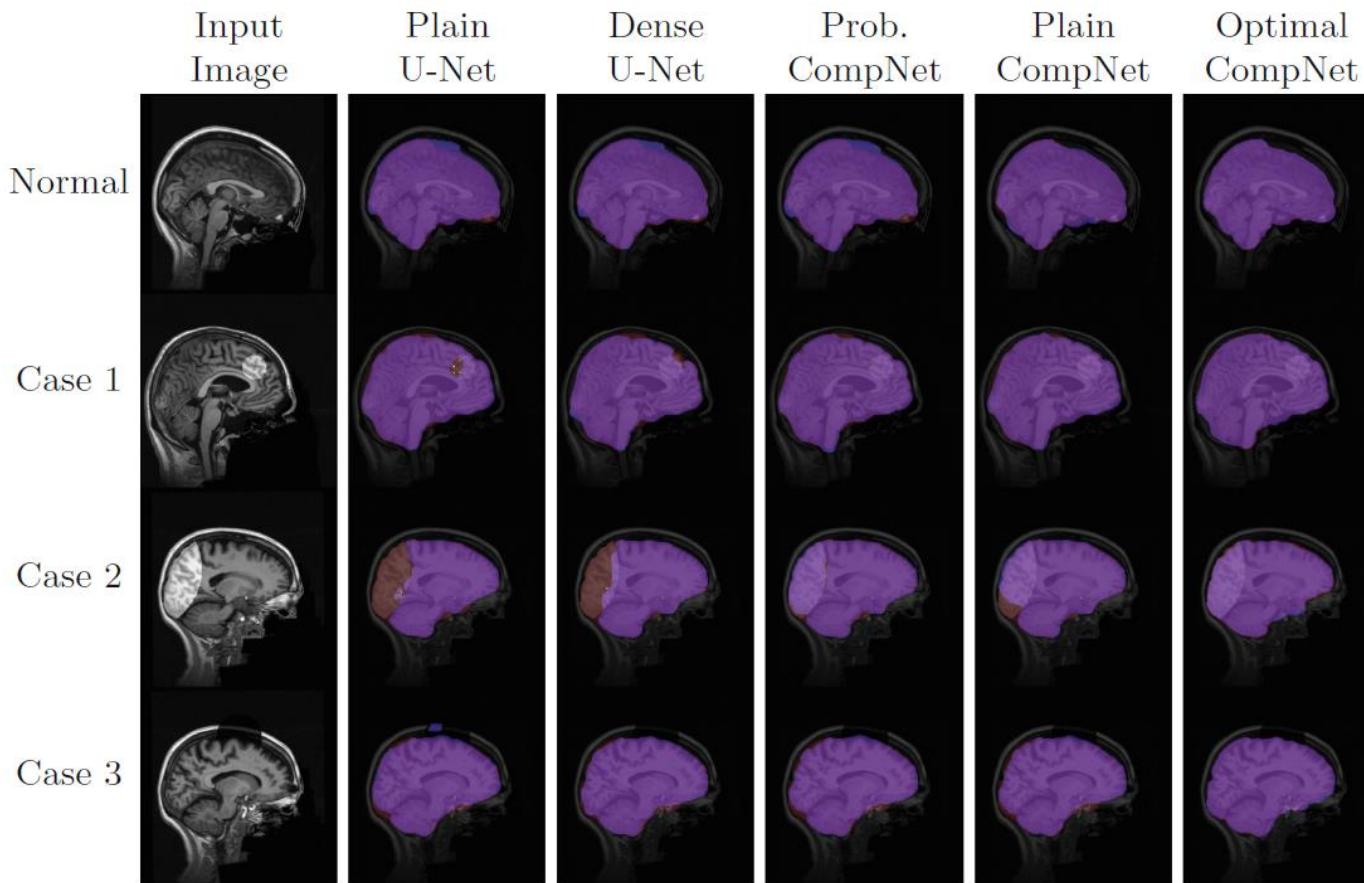
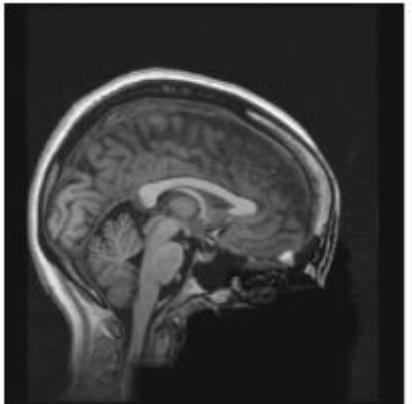


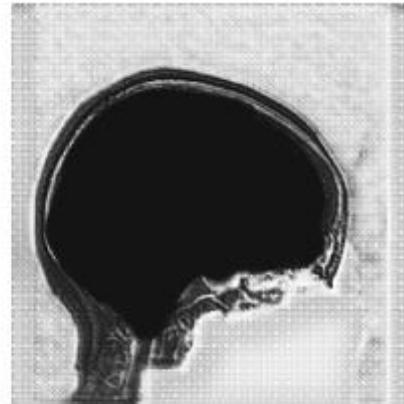
Fig. 2. Qualitative comparison among five networks, plain and dense U-Nets, probability, plain, and optimal CompNets, on four image samples: a normal one, one with pathology inside of the brain (case 1), one with pathology on the boundary of the brain (case 2), and one with a damaged skull (case 3). The true (red) and predicted (blue) masks are superimposed over the original images. The purple color indicates a perfect overlap between the ground truth and the prediction. Best viewed in color.



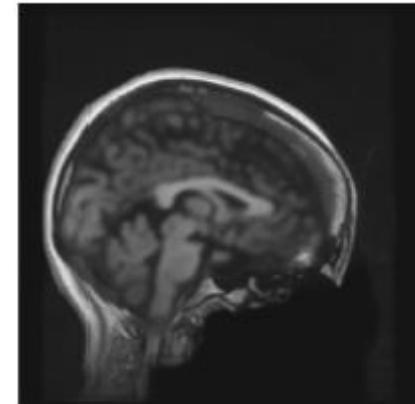
(a) Input



(b) Brain mask



(c) Complement



(d) Reconstruction

Fig. 3. Three outputs (b–d) of our optimal CompNet for an input brain scan (a).

One-Pass Multi-task Convolutional Neural Networks for Efficient Brain Tumor Segmentation

Chenhong Zhou¹, Changxing Ding^{1(✉)}, Zhentai Lu², Xinchao Wang³,
and Dacheng Tao⁴

¹ School of Electronic and Information Engineering,
South China University of Technology, Guangzhou, China
chxding@scut.edu.cn

² Guangdong Provincial Key Laboratory of Medical Image Processing,
School of Biomedical Engineering, Southern Medical University, Guangzhou, China
³ Department of Computer Science, Stevens Institute of Technology, Hoboken, USA
⁴ UBTECH Sydney AI Centre, SIT, FEIT, University of Sydney, Sydney, Australia

Abstract. The model cascade strategy that runs a series of deep models sequentially for coarse-to-fine medical image segmentation is becoming increasingly popular, as it effectively relieves the class imbalance problem. This strategy has achieved state-of-the-art performance in many segmentation applications but results in undesired system complexity and ignores correlation among deep models. In this paper, we propose a light and clean deep model that conducts brain tumor segmentation in a single-pass and solves the class imbalance problem better than model cascade. First, we decompose brain tumor segmentation into three different but related tasks and propose a multi-task deep model that trains them together to exploit their underlying correlation. Second, we design a curriculum learning-based training strategy that trains the above multi-task model more effectively. Third, we introduce a simple yet effective post-processing method that can further improve the segmentation performance significantly. The proposed methods are extensively evaluated on BRATS 2017 and BRATS 2015 datasets, ranking first on the BRATS 2015 test set and showing top performance among 60+ competing teams on the BRATS 2017 validation set.

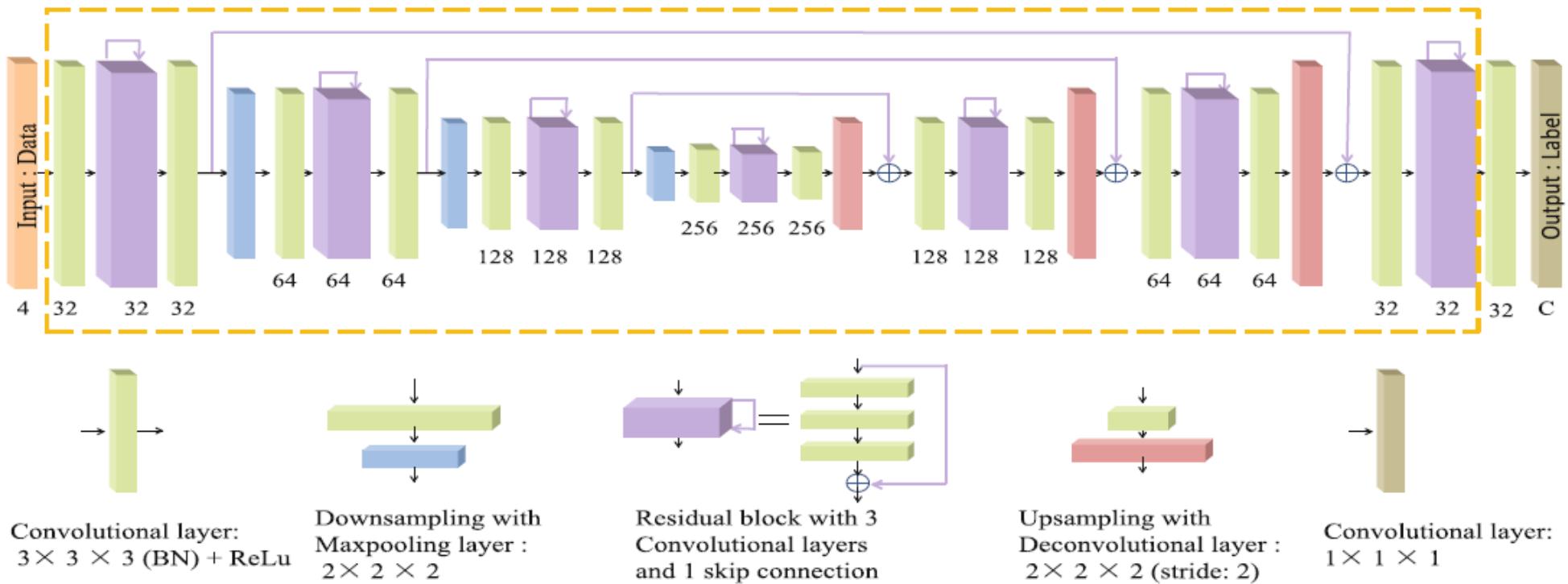


Fig. 1. Network architecture used in each task. The building blocks are represented by colored cubes with numbers below being the number of feature maps. C equals to 5, 5, and 2 for the first, second, and third task, respectively. (Best viewed in color)

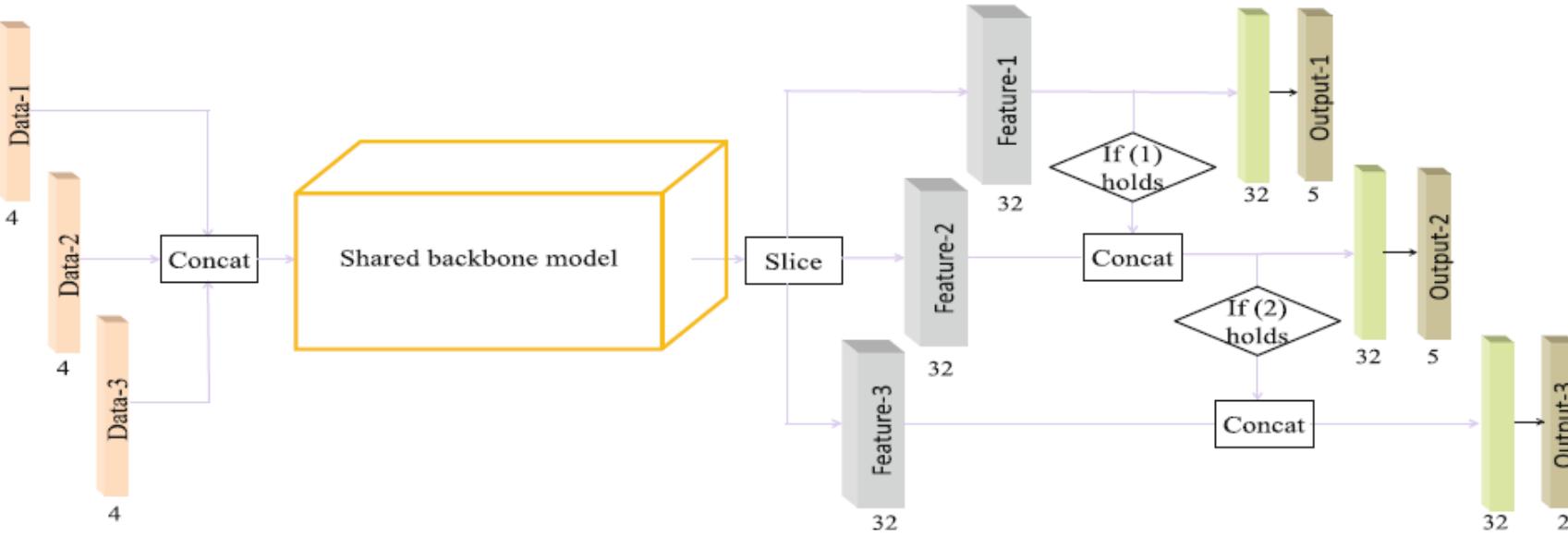


Fig. 2. Architecture of OM-Net. Data-*i*, Feature-*i*, and Output-*i* denote training data, feature, and classification layer for the *i*-th task, respectively. The shared backbone model refers to the network layers outlined by the yellow dashed line in Fig. 1.

Semi-automatic RECIST Labeling on CT Scans with Cascaded Convolutional Neural Networks

Youbao Tang^{1(✉)}, Adam P. Harrison³, Mohammadhadi Bagheri², Jing Xiao⁴,
and Ronald M. Summers¹

¹ Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, National Institutes of Health Clinical Center, Bethesda, MD 20892, USA
youbao.tang@nih.gov

² Clinical Image Processing Service, National Institutes of Health Clinical Center, Bethesda, MD 20892, USA

³ NVIDIA, Santa Clara, CA 95051, USA

⁴ Ping An Insurance Company of China, Shenzhen 510852, China

Abstract. Response evaluation criteria in solid tumors (RECIST) is the standard measurement for tumor extent to evaluate treatment responses in cancer patients. As such, RECIST annotations must be accurate. However, RECIST annotations manually labeled by radiologists require professional knowledge and are time-consuming, subjective, and prone to inconsistency among different observers. To alleviate these problems, we propose a cascaded convolutional neural network based method to semi-automatically label RECIST annotations and drastically reduce annotation time. The proposed method consists of two stages: lesion region normalization and RECIST estimation. We employ the spatial transformer network (STN) for lesion region normalization, where a localization network is designed to predict the lesion region and the transformation parameters with a multi-task learning strategy. For RECIST estimation, we adapt the stacked hourglass network (SHN), introducing a relationship constraint loss to improve the estimation precision. STN and SHN can both be learned in an end-to-end fashion. We train our system on the DeepLesion dataset, obtaining a consensus model trained on RECIST annotations performed by multiple radiologists over a multi-year period. Importantly, when judged against the inter-reader variability of two additional radiologist raters, our system performs more stably and with less variability, suggesting that RECIST annotations can be reliably obtained with reduced labor and time.

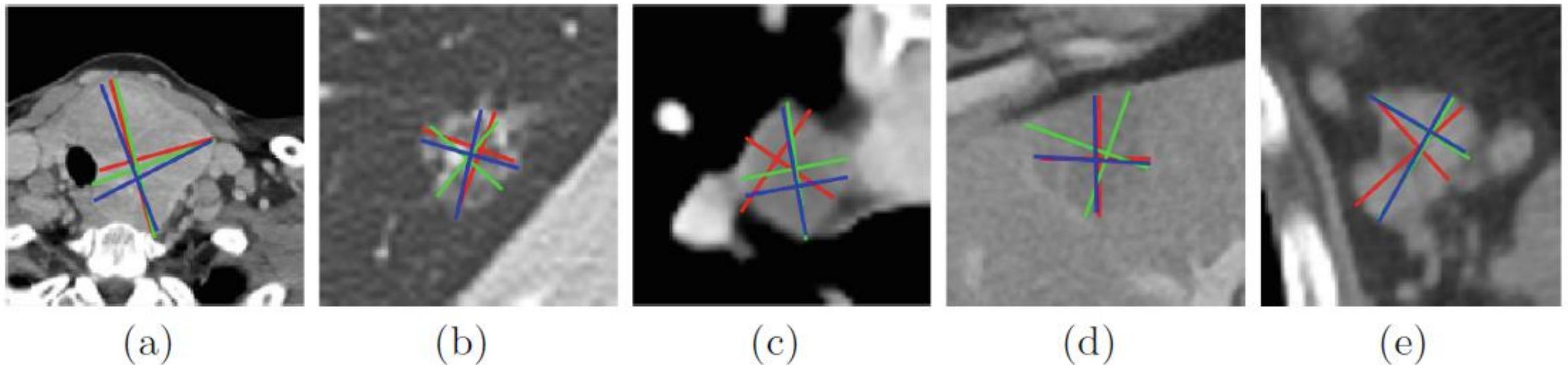


Fig. 1. Five examples of RECIST annotations labeled by three radiologists. For each image, the RECIST annotations from different observers are indicated by diameters with different colors. Better viewed in color.

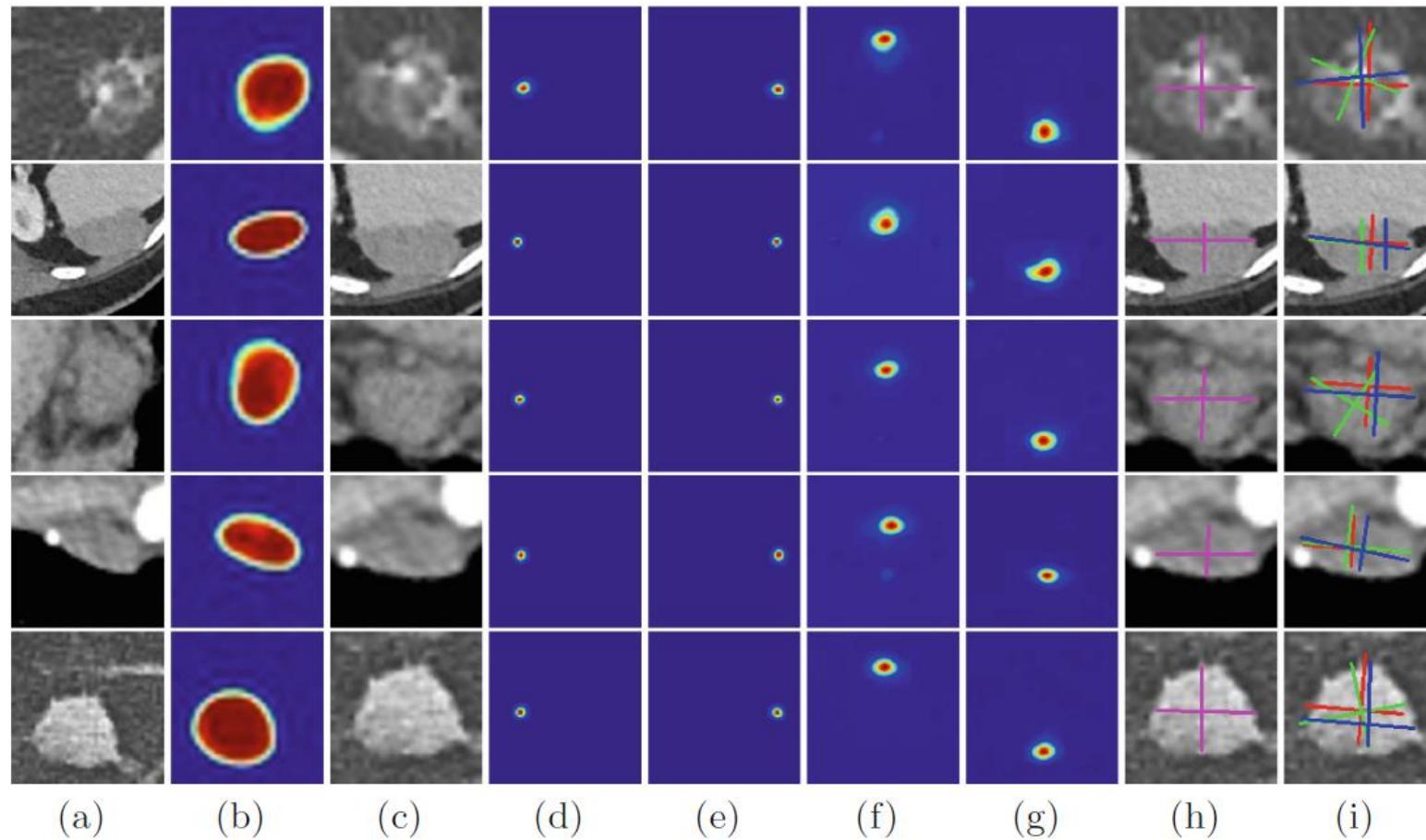


Fig. 3. Given the input test image (a), we can obtain the predicted lesion mask (b), the transformed image (c) from the STN, and the estimated keypoint heatmaps (d)–(g) from the SHN. From (d)–(g), we obtain the estimated RECIST (h), which is close to the annotations (i) labeled by radiologists. Red, green, and blue marks denote DL, R1, and R2 annotations, respectively.