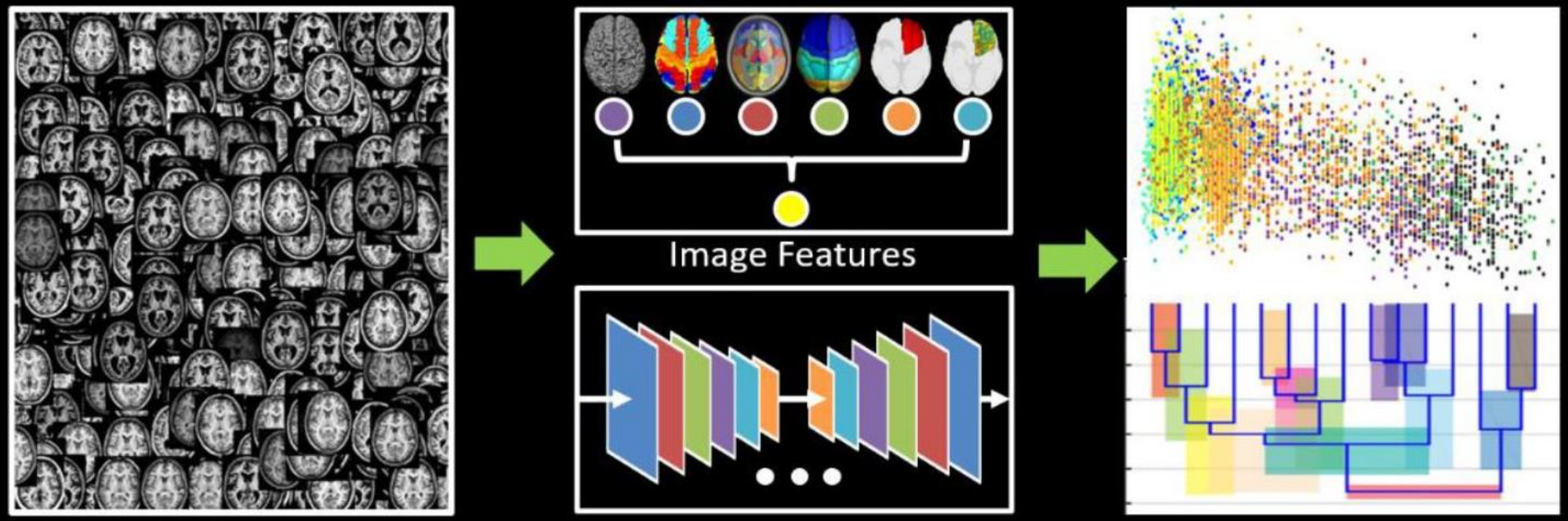


Multi-Modal Learning

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

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

Organization of Course

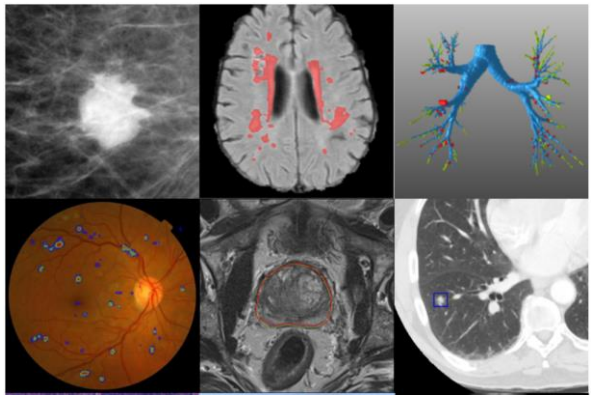
Overview

Overview of Deep Learning in Medical Image Computing
Neural Networks and CNN

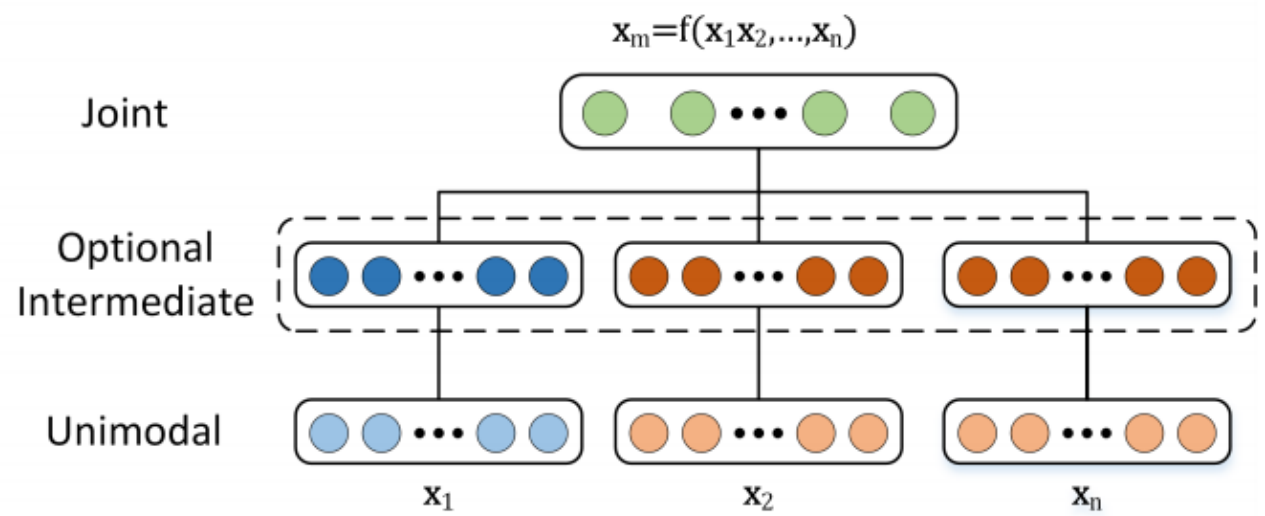
Key Techs

Classification (Medical Image Diagnosis)
Detection (Landmark Localization and Detection)
Segmentation (Medical Image Segmentation)
GAN (Medical Image Synthesis)

Topics in Medical Image Computing

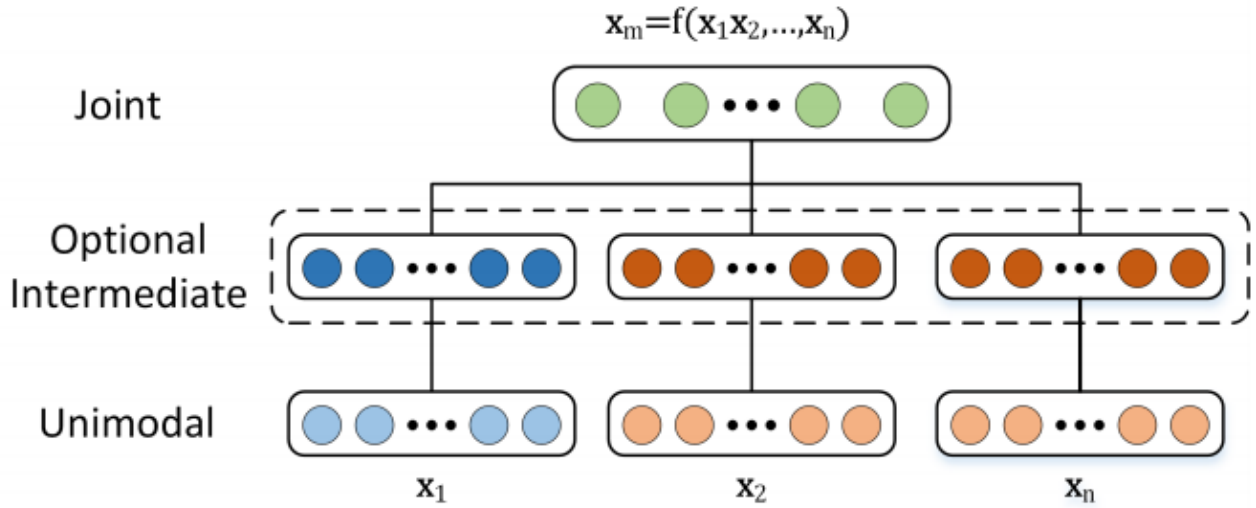


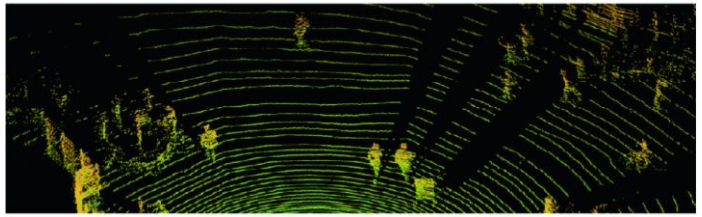
Multi-modal Learning

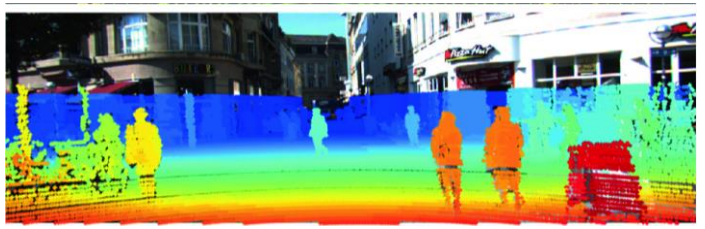


https://arxiv.org/pdf/1705.09406.pdf

Multi-modal Learning

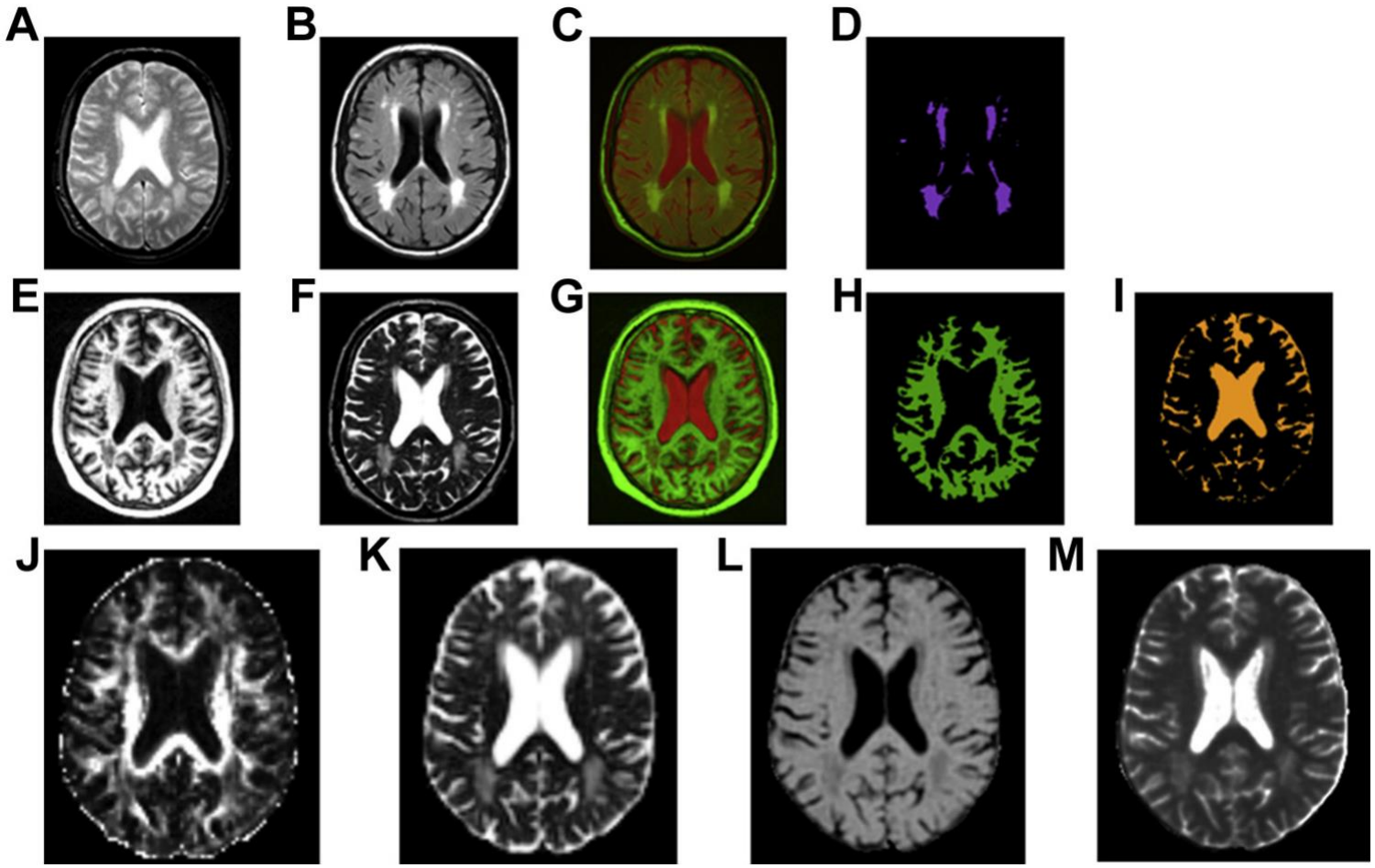






https://arxiv.org/pdf/1705.09406.pdf https://link.springer.com/chapter/10.1007/978-3-030-28603-3_17

Multi-modal Learning

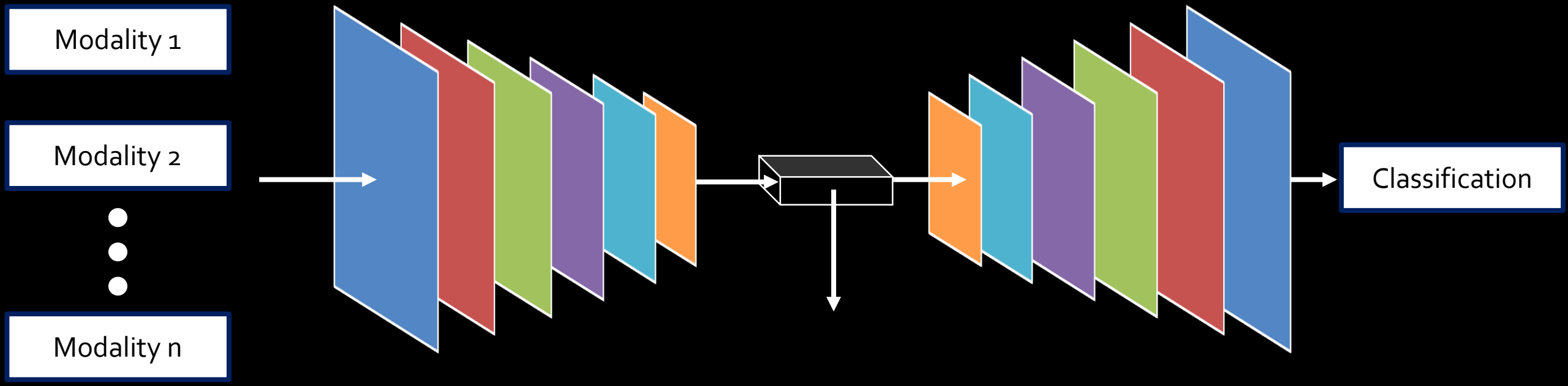


S.M. Maniega et al. / Neurobiology of Aging 36 (2015) 909e918

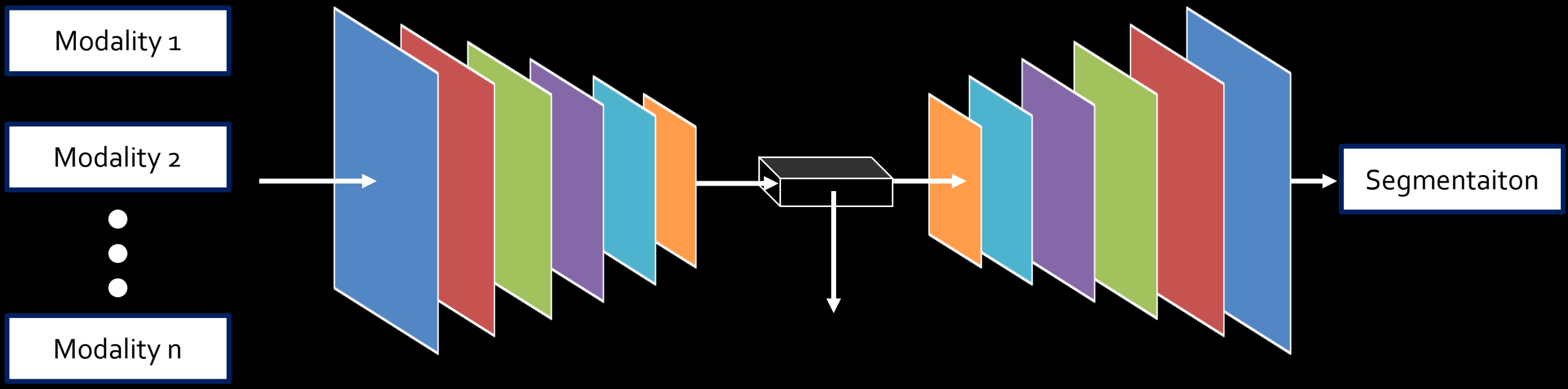
T2*W (A) and FLAIR (B) structural scans are combined in red-green color space (C) to facilitate the extraction of WMH voxels (D). T1W (E) and T2W (F) structural scans are combined in red-green color space (G) to facilitate the extraction of NAWM (H) and CSF (I) voxels; the latter is subtracted from the WMH and NAWM masks to avoid CSF partial volume averaging within the measurement masks. The last row shows reconstructed parametric images of MRI biomarkers: FA (J), MD (K), MTR (L) and T1 relaxation time (M)

Sep 25	Multi-model Learning

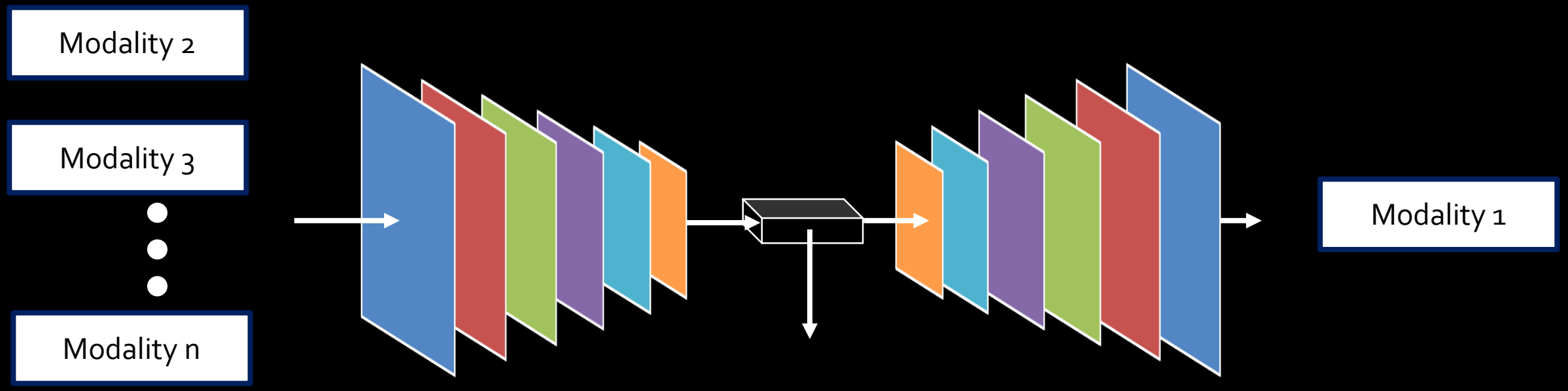
Multi-modal Learning



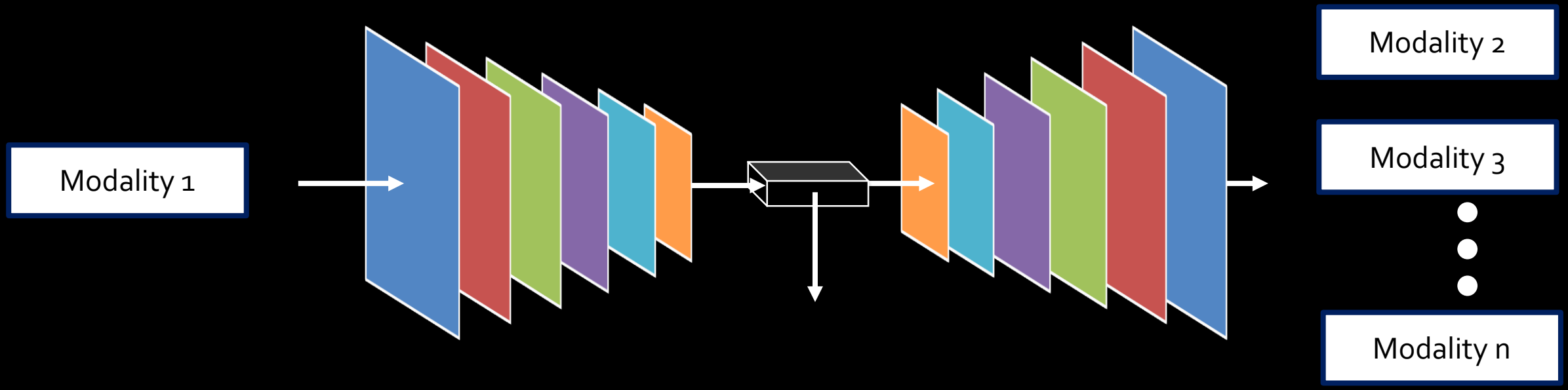
Multi-modal Learning



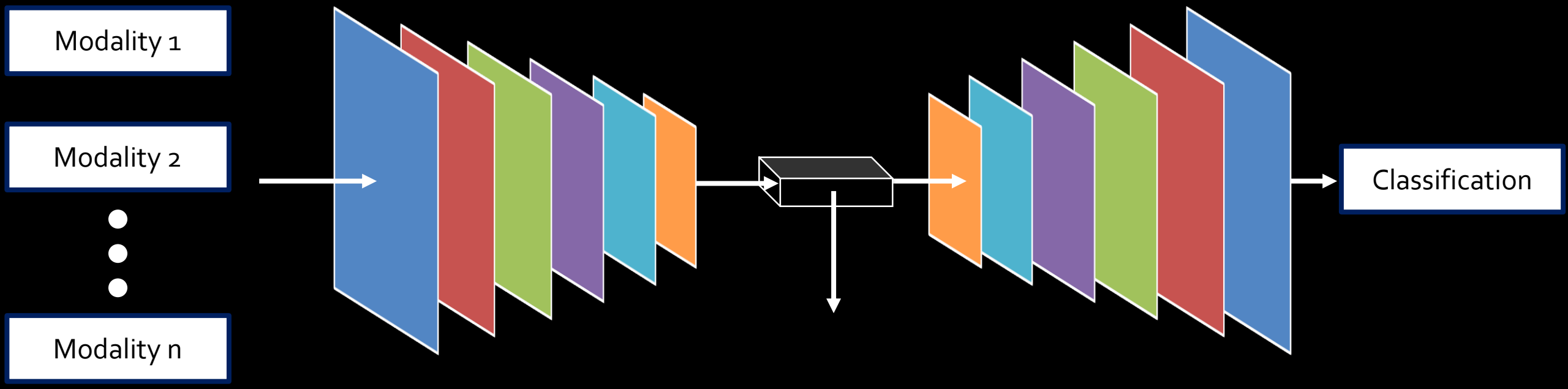
Multi-modal Learning



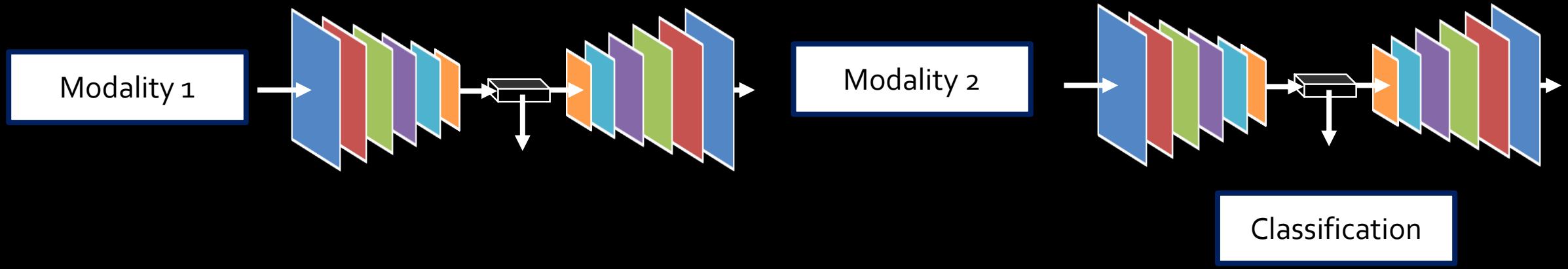
Multi-modal Learning



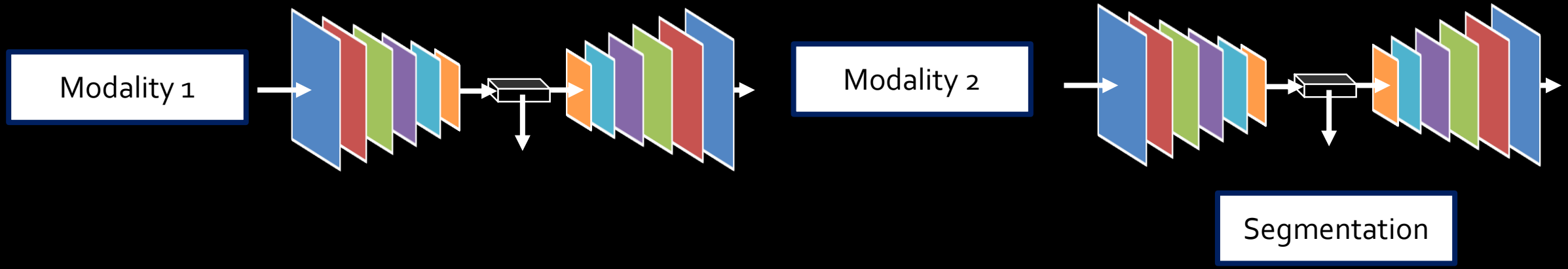
Multi-modal Learning



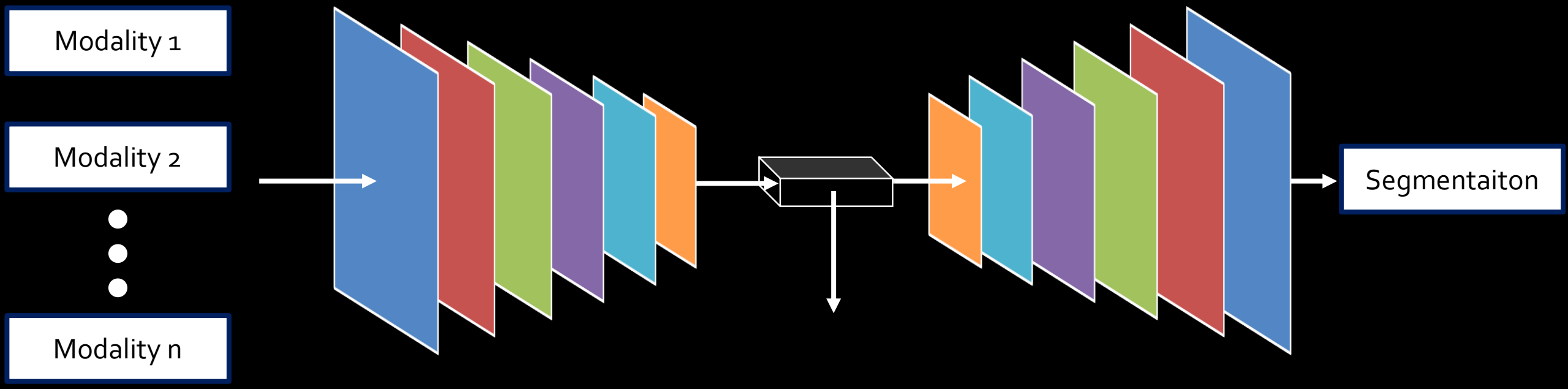
Multi-modal Learning



Multi-modal Learning



Multi-modal Learning



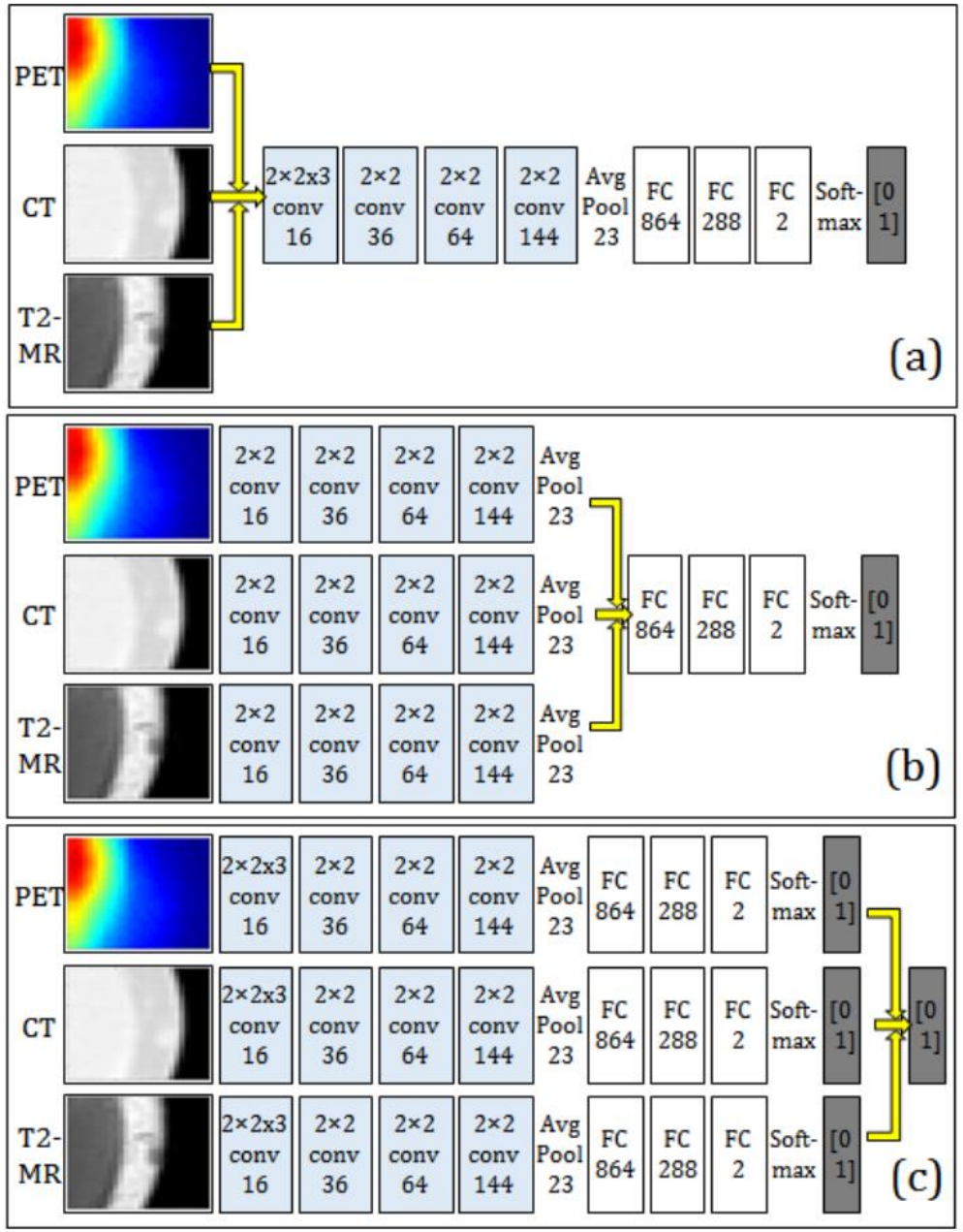
Medical Image Segmentation Based on Multi-Modal Convolutional Neural Network: Study on Image Fusion Schemes

Zhe Guo*[1], Xiang Li*[2], Heng Huang[3], Ning Guo[1], and Quanzheng Li[1]

[1]Massachusetts General Hospital
[2]Beijing Institute of Technology
[3]University of Pittsburgh *Joint first authors

November 3, 2017

Image analysis using more than one modality (i.e. multi-modal) has been increasingly applied in the field of biomedical imaging. One of the challenges in performing the multi-modal analysis is that there exist multiple schemes for fusing the information from different modalities, where such schemes are application-dependent and lack a unified framework to guide their designs. In this work we firstly propose a conceptual architecture for the image fusion schemes in supervised biomedical image analysis: fusing at the feature level, fusing at the classifier level, and fusing at the decision-making level. Further, motivated by the recent success in applying deep learning for natural image analysis, we implement the three image fusion schemes above based on the Convolutional Neural Network (CNN) with varied structures, and combined into a single framework. The proposed image segmentation framework is capable of analyzing the multi-modality images using different fusing schemes simultaneously. The framework is applied to detect the presence of soft tissue sarcoma from the combination of Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron Emission Tomography (PET) images. It is found from the results that while all the fusion schemes outperform the single-modality schemes, fusing at the feature level can generally achieve the best performance in terms of both accuracy and computational cost, but also suffers from the decreased robustness in the presence of large errors in any image modalities.


PET
CT
T2-MR
2×2x3 conv 16
2×2 conv 36
2×2 conv 64
2×2 conv 144
Avg Pool 23
FC 864
FC 288
FC 2
Soft-max
[0 1]
(a)
(b)
(c)

PIMMS: Permutation Invariant Multi-Modal Segmentation

Thomas Varsavsky[1], Zach Eaton-Rosen[1,2], Carole H. Sudre[2,1,3], Parashkev Nachev[4] and M. Jorge Cardoso[2,1]

[1] CMIC, University College London, UK
[2] School of Biomedical Engineering and Imaging Sciences, Kings College London, UK
[3] Dementia Research Centre, University College London, UK
[4] Institute of Neurology, University College London, UK

Abstract. In a research context, image acquisition will often involve a pre-defined static protocol and the data will be of high quality. If we are to build applications that work in hospitals without significant operational changes in care delivery, algorithms should be designed to cope with the available data in the best possible way. In a clinical environment, imaging protocols are highly flexible, with MRI sequences commonly missing appropriate sequence labeling (e.g. T1, T2, FLAIR). To this end we introduce PIMMS, a Permutation Invariant Multi-Modal Segmentation technique that is able to perform inference over sets of MRI scans without using modality labels. We present results which show that our convolutional neural network can, in some settings, outperform a baseline model which utilizes modality labels, and achieve comparable performance otherwise.

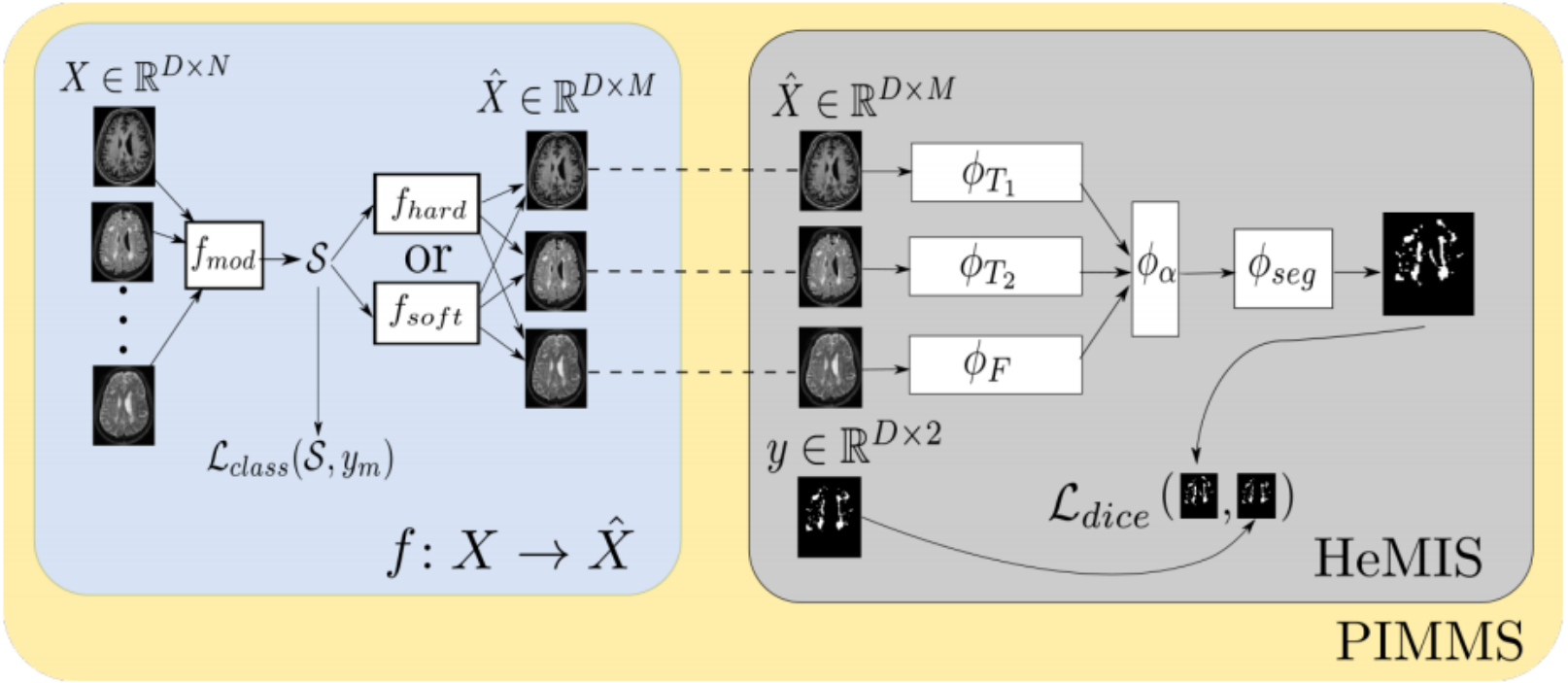
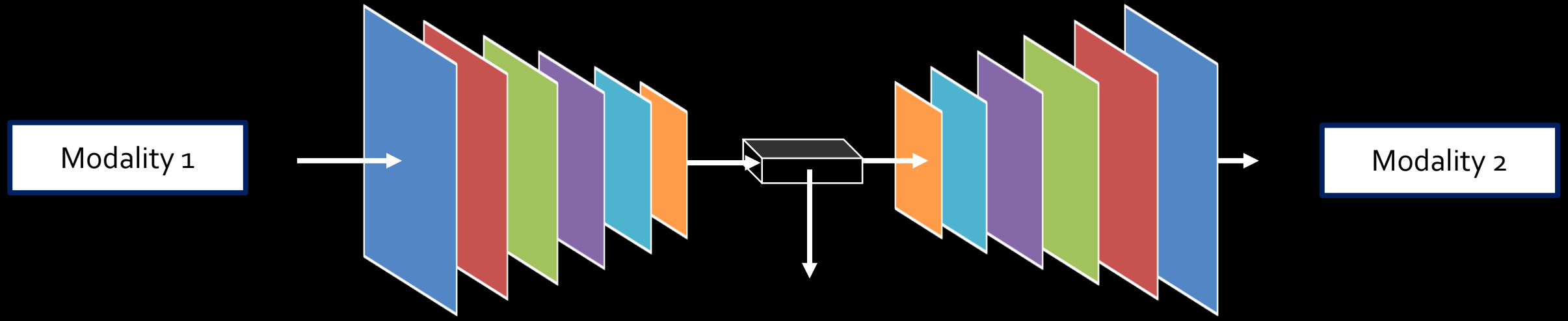


Fig. 1: Diagram showing the network architecture. During training the inputs are $X \in \mathbb{R}^{D \times N}$ and the corresponding ground truth binary segmentation $y \in \mathbb{R}^{D \times 2}$. A function f_{mod} takes each scan as input and outputs a modality score $\mathcal{S}$ which produces the representation $\hat{X} \in \mathbb{R}^{D \times M}$. The weights of ϕ_{T_1}, ϕ_{T_2}, ϕ_F and ϕ_{seg} are learned by differentiating with respect to $\mathcal{L}_{seg}$ and the weights of f are learned by differentiating with respect to $\mathcal{L}_{class}$. y_m is a one-hot encoded modality label.

Multi-modal Learning



Ultra-Fast T2-Weighted MR Reconstruction Using Complementary T1-Weighted Information

Lei Xiang[1], Yong Chen[2], Weitang Chang[2], Yiqiang Zhan[1], Weili Lin[2], Qian Wang[1(✉)], and Dinggang Shen[2(✉)]

[1] Institute for Medical Imaging Technology, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
wang.qian@sjtu.edu.cn

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
dgshen@med.unc.edu

Abstract. T1-weighted image (T1WI) and T2-weighted image (T2WI) are the two routinely acquired Magnetic Resonance Imaging (MRI) protocols that provide complementary information for diagnosis. However, the total acquisition time of ~10 min yields the image quality vulnerable to artifacts such as motion. To speed up MRI process, various algorithms have been proposed to reconstruct high quality images from under-sampled k-space data. These algorithms only employ the information of an individual protocol (e.g., T2WI). In this paper, we propose to combine complementary MRI protocols (i.e., T1WI and under-sampled T2WI particularly) to reconstruct the high-quality image (i.e., fully-sampled T2WI). To the best of our knowledge, this is the first work to utilize data from different MRI protocols to speed up the reconstruction of a target sequence. Specifically, we present a novel deep learning approach, namely Dense-Unet, to accomplish the reconstruction task. The Dense-Unet requires fewer parameters and less computation, but achieves better performance. Our results have shown that Dense-Unet can reconstruct a 3D T2WI volume in less than 10 s, i.e., with the acceleration rate as high as 8 or more but with negligible aliasing artefacts and signal-noise-ratio (SNR) loss.

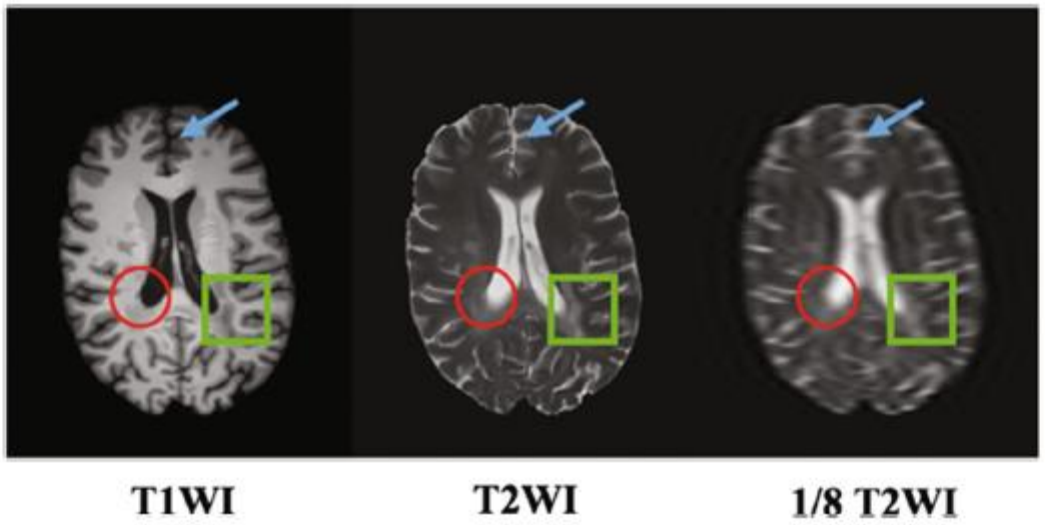


Fig. 1. Examples of the pairs of T1WI, T2WI and 1/8 under-sampled T2WI data from the same patient. Multiple sclerosis lesions are marked by circles and boxes in the figure.

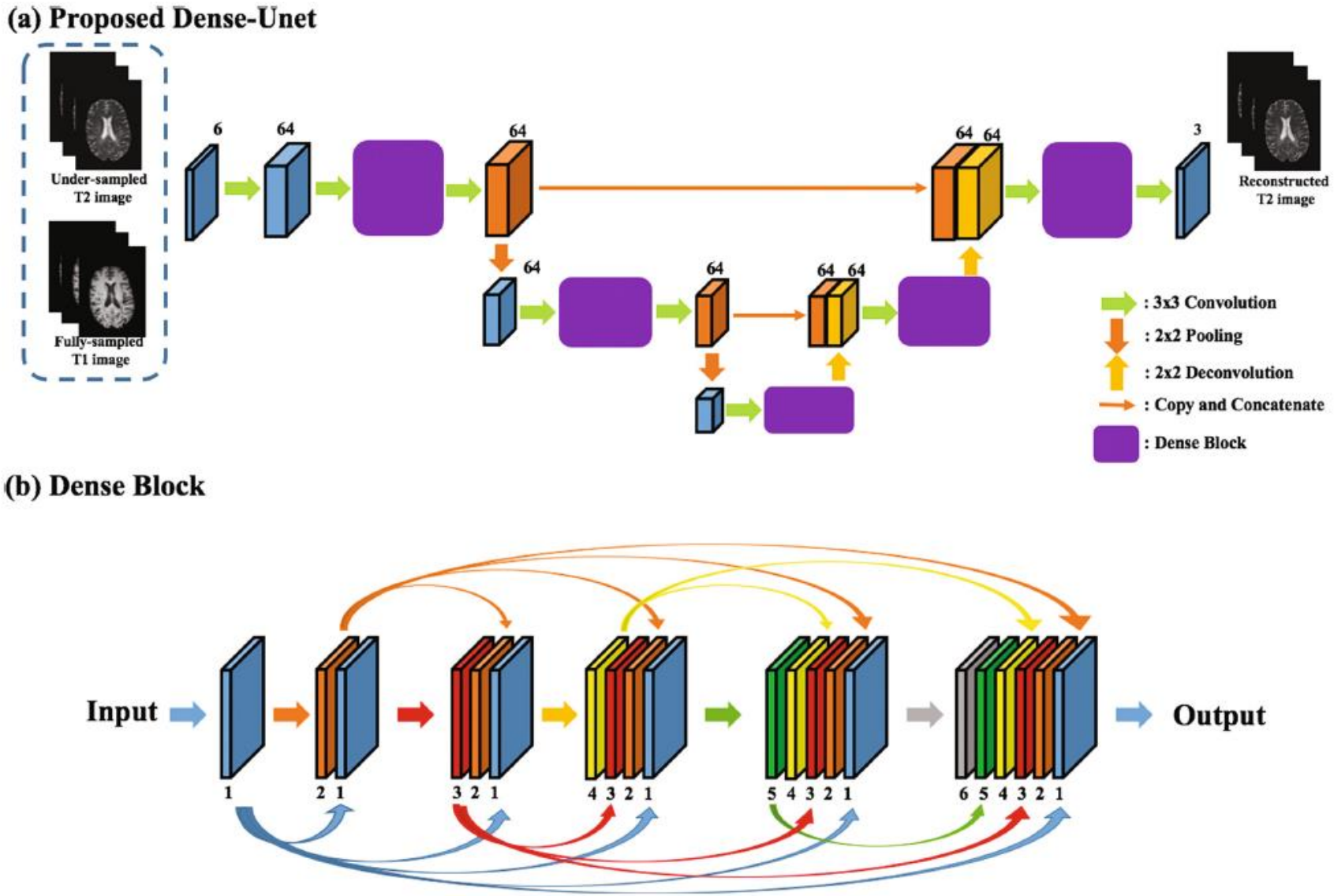


Fig. 2. Illustration of (a) the framework for T2WI reconstruction with T1WI and under-sampled T2WI and (b) the detailed configuration of dense block.

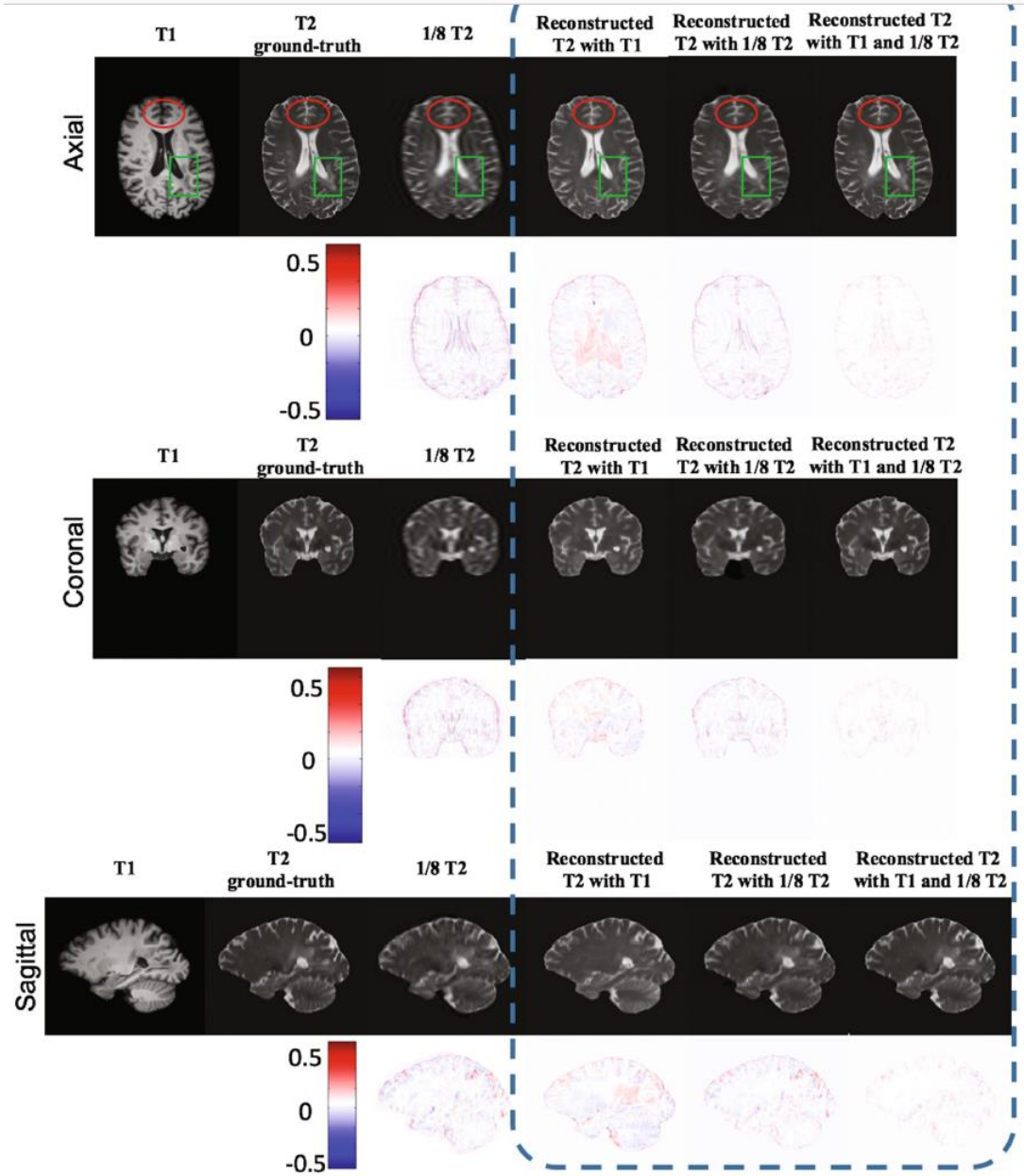


Fig. 3. Visual examples of using multi-inputs for T2WI reconstruction.

Conditional Generative Adversarial Networks for Metal Artifact Reduction in CT Images of the Ear

Jianing Wang(✉), Yiyuan Zhao, Jack H. Noble,
and Benoit M. Dawant

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

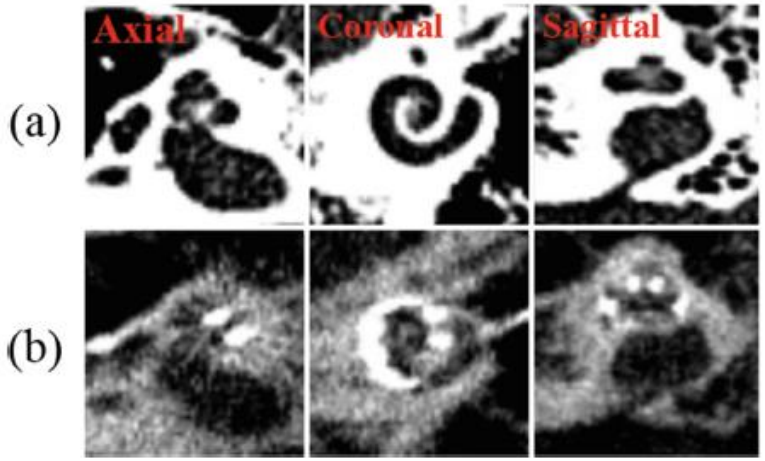


Fig. 2. Three orthogonal views of (a) the Pre-CT and (b) the Post-CT of an example ear.

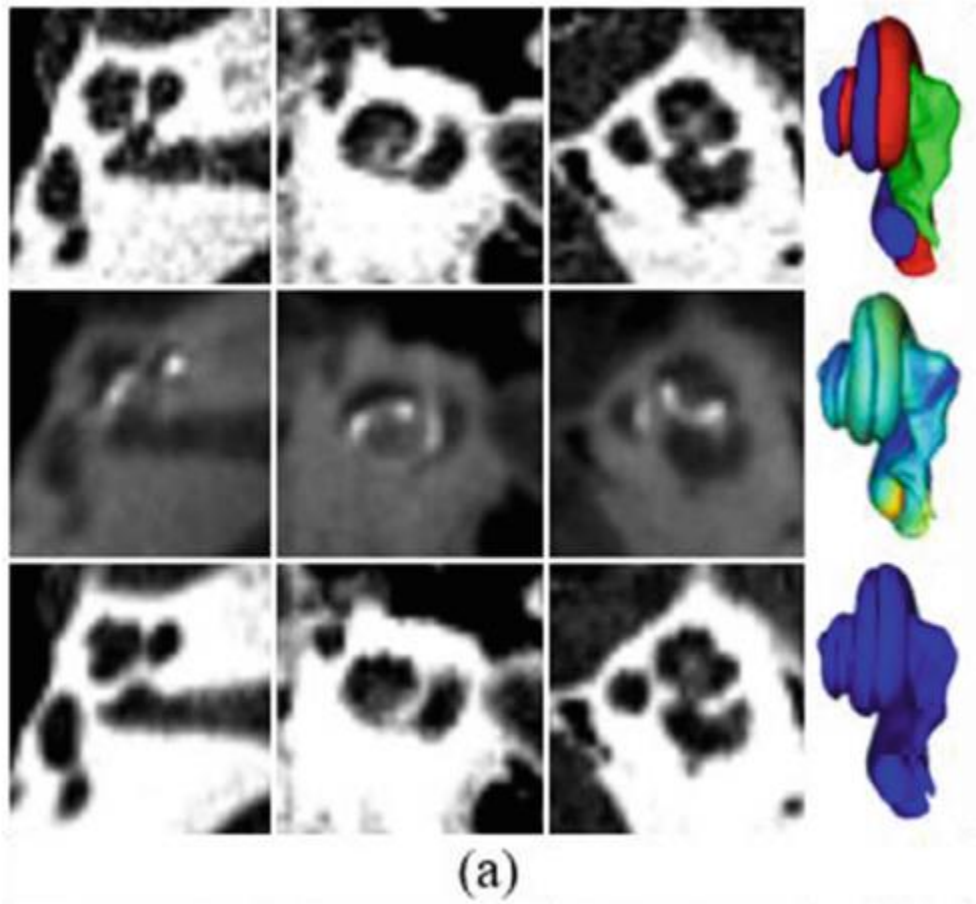
4.1 cGAN

In this work we rely on the cGAN framework proposed by Isola *et al.* [4]. A cGAN consists of a generator G and a discriminator D. The total loss can be expresses as

$$L = arg\ \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L_1}(G) \tag{1}$$

wherein

$$L_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log(D(x, y))] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{2}$$



Towards Generating Personalized Volumetric Phantom from Patient's Surface Geometry

Yifan Wu[1], Vivek Singh[1(✉)], Brian Teixeira[1], Kai Ma[1], Birgi Tamersoy[1], Andreas Krauss[2], and Terrence Chen[1]

[1] Medical Imaging Technologies, Siemens Healthineers, Princeton, NJ, USA
vivek-singh@siemens-healthineers.com
[2] Siemens Healthcare GmbH, Forchheim, Germany

Abstract. This paper presents a method to generate a volumetric phantom with internal anatomical structures from the patient's skin surface geometry, and studies the potential impact of this technology on planning medical scans and procedures such as patient positioning. Existing scan planning for imaging is either done by visual inspection of the patient or based on an ionizing scan obtained prior to the full scan. These methods are either limited in accuracy or result in additional radiation dose to the patient. Our approach generates a "CT"-like phantom, with lungs and bone structures, from the patient's skin surface. The skin surface can be estimated from a 2.5D depth sensor and thus, the proposed method offers a novel solution to reduce the radiation dose. We present quantitative experiments on a dataset of 2045 whole body CT scans and report measurements relevant to the potential clinical use of such phantoms. (This feature is based on research, and is not commercially available. Due to regulatory reasons its future availability cannot be guaranteed.)

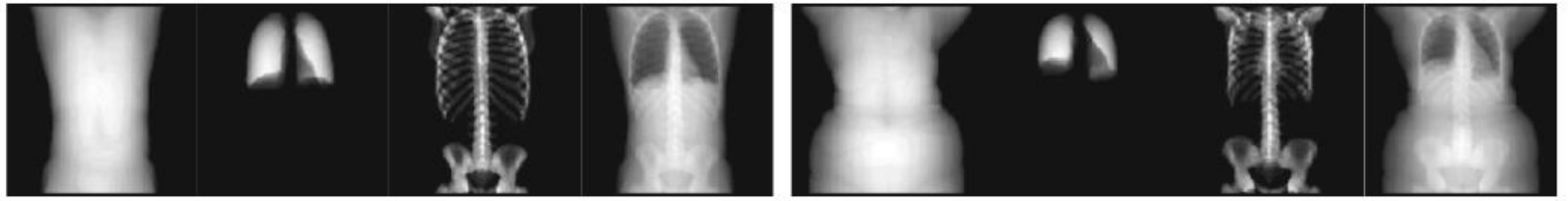


Fig. 1. Illustration of data. From left to right we show the patient's body surface mask, lungs mask, bone mask and phantom respectively, for 2 different patients. All masks and phantoms are volumetric, images displayed here are orthographic projections (averaged along the AP axis).

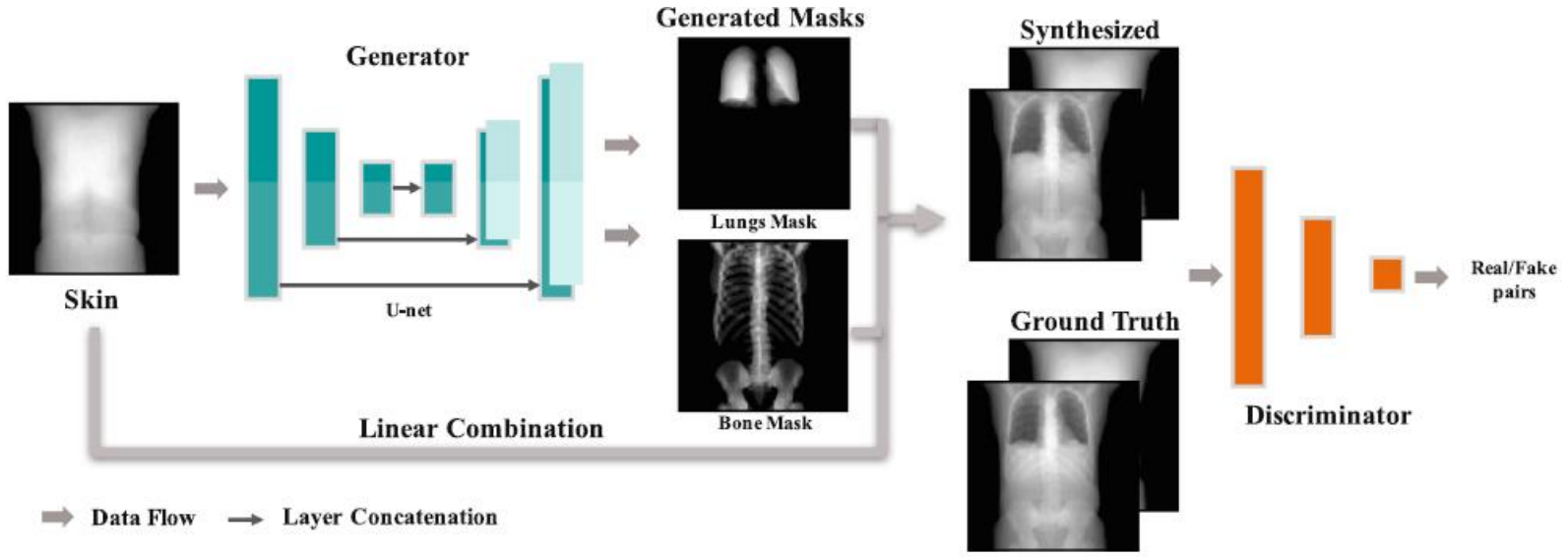
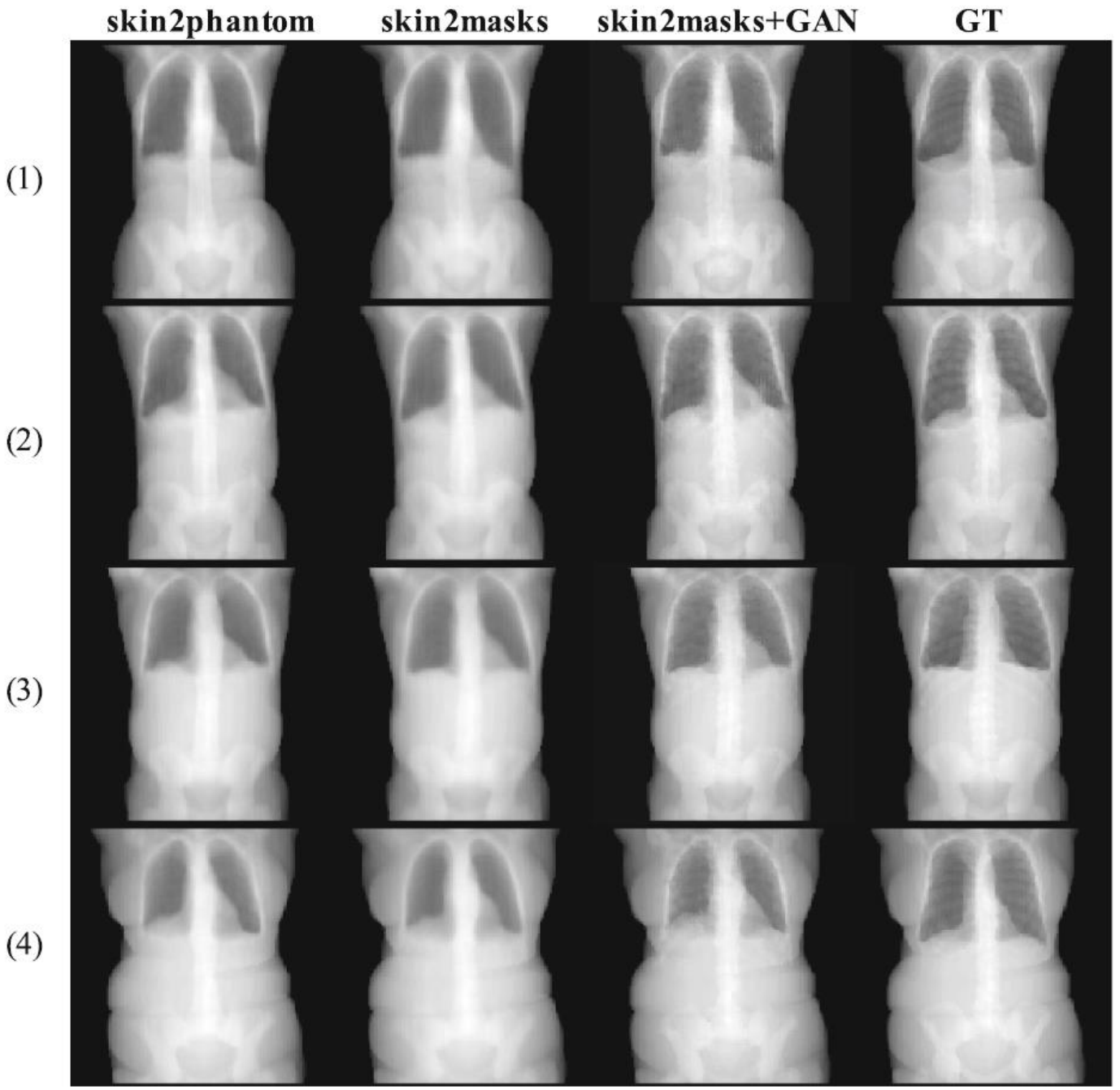
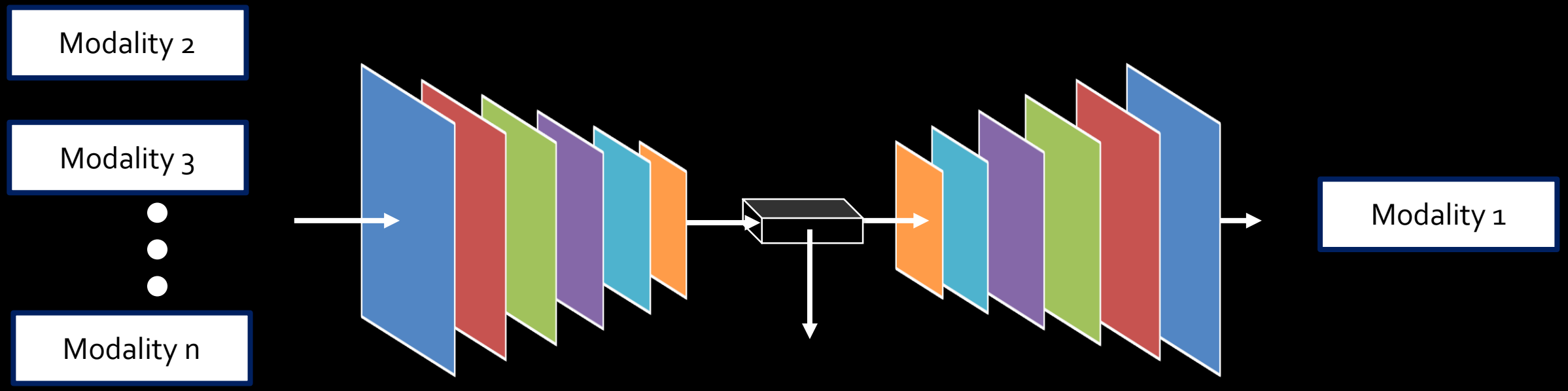


Fig. 2. Overview of the proposed framework for phantom generation. Images displayed here are orthographic projections (averaged along the AP axis).


skin2phantom
skin2masks
skin2masks+GAN
GT
(1)
(2)
(3)
(4)

Multi-modal Learning



Learning Myelin Content in Multiple Sclerosis from Multimodal MRI Through Adversarial Training

Wen Wei[1,2,3(✉)], Emilie Poirion[3], Benedetta Bodini[3], Stanley Durrleman[2,3], Nicholas Ayache[1], Bruno Stankoff[3], and Olivier Colliot[2,3]

[1] UCA, Inria, Epione project-team, Sophia Antipolis, France
wen.wei@inria.fr
[2] Inria, Aramis project-team, Paris, France
[3] Sorbonne Université, Inserm, CNRS, Institut du cerveau et la moelle (ICM), AP-HP-Hôpital Pitié-Salpêtrière, Boulevard de l'hôpital, Paris, France

Abstract. Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS). A reliable measure of the tissue myelin content is therefore essential to understand the physiopathology of MS, track progression and assess treatment efficacy. Positron emission tomography (PET) with [^{11}C]PIB has been proposed as a promising biomarker for measuring myelin content changes in-vivo in MS. However, PET imaging is expensive and invasive due to the injection of a radioactive tracer. On the contrary, magnetic resonance imaging (MRI) is a non-invasive, widely available technique, but existing MRI sequences do not provide, to date, a reliable, specific, or direct marker of either demyelination or remyelination. In this work, we therefore propose Sketcher-Refiner Generative Adversarial Networks (GANs) with specifically designed adversarial loss functions to predict the PET-derived myelin content map from a combination of MRI modalities. The prediction problem is solved by a sketch-refinement process in which the sketcher generates the preliminary anatomical and physiological information and the refiner refines and generates images reflecting the tissue myelin content in the human brain. We evaluated the ability of our method to predict myelin content at both global and voxel-wise levels. The evaluation results show that the demyelination in lesion regions and myelin content in normal-appearing white matter (NAWM) can be well predicted by our method. The method has the potential to become a useful tool for clinical management of patients with MS.

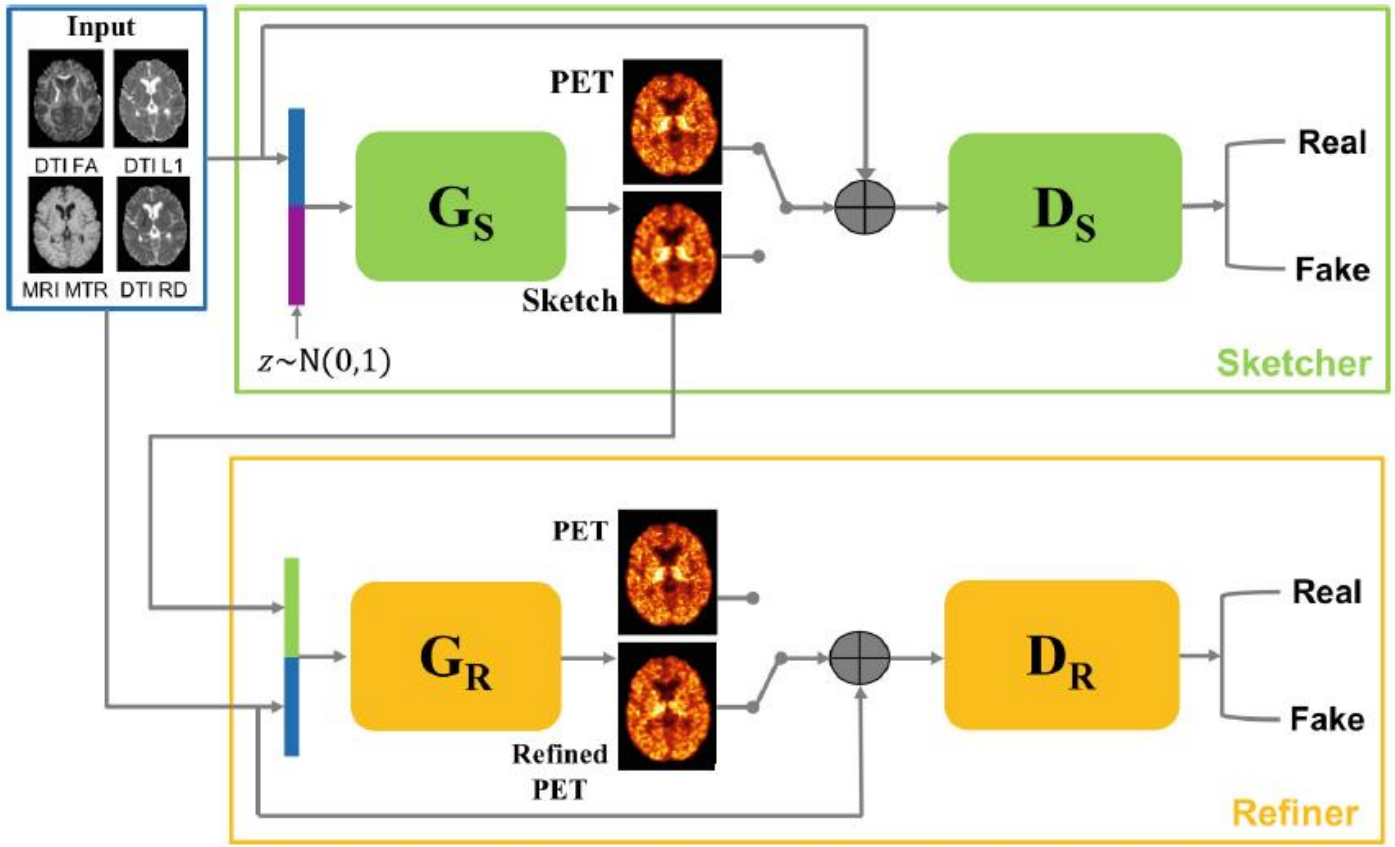


Fig. 1. The proposed sketcher-refiner GANs. The sketcher receives MR images and generates the preliminary anatomy and physiology information. The refiner receives MR images I_M and the sketch I_S. Then it refines and generates PET images.

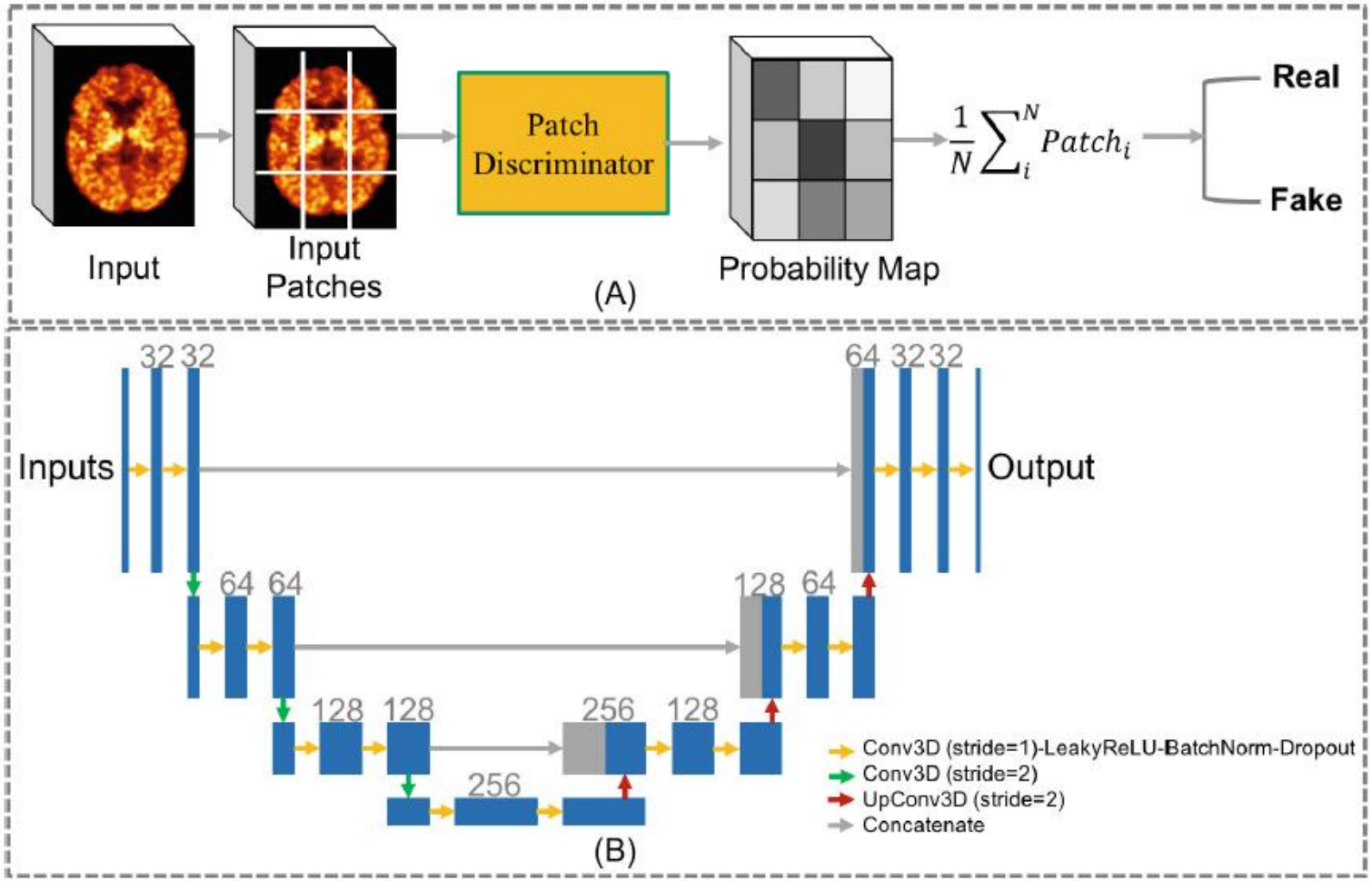


Fig. 2. The D and the G in our GANs. (A) The proposed 3D patch discriminator which takes all the patches and classifies them separately to output a final loss. (B) The 3D U-Net shaped generator with implementation details shown in the image.

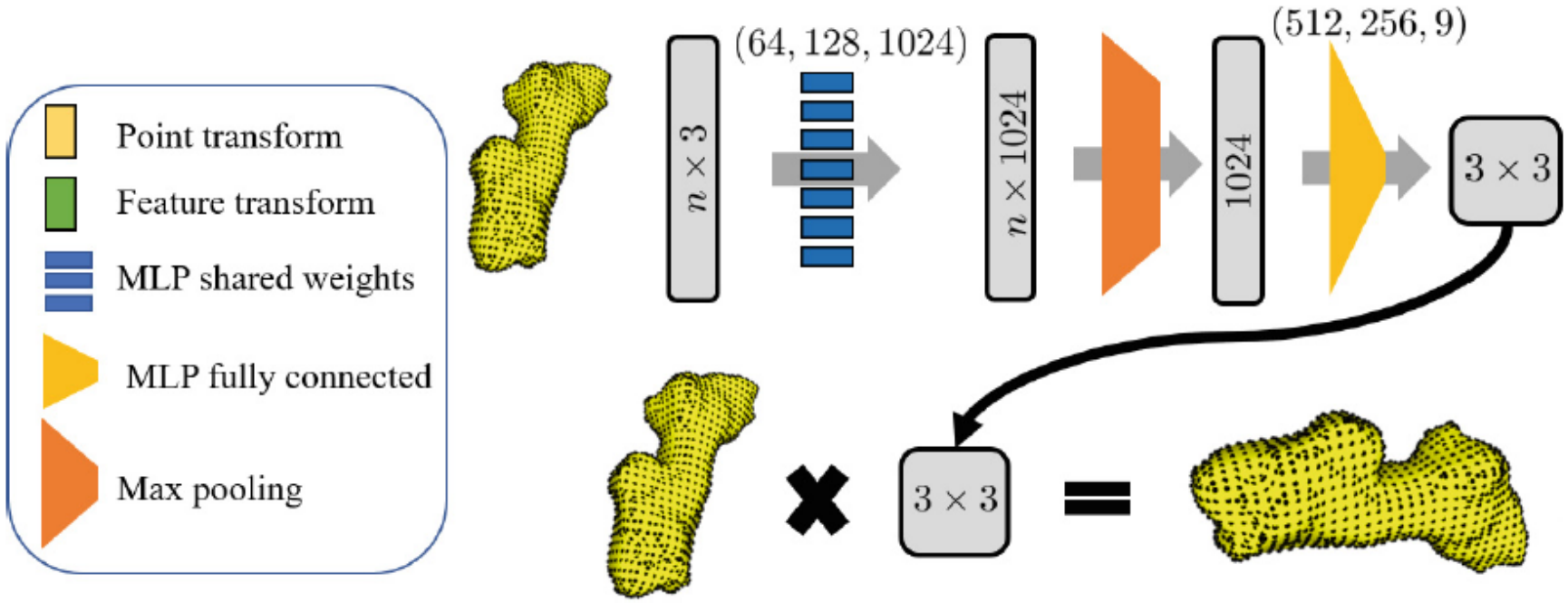


Fig. 2. Transformation network (T-Net) for predicting a transformation matrix to map a point cloud to canonical space before processing. A similar network is used to transform the features; the only difference is that the output corresponds to a 64×64 matrix.

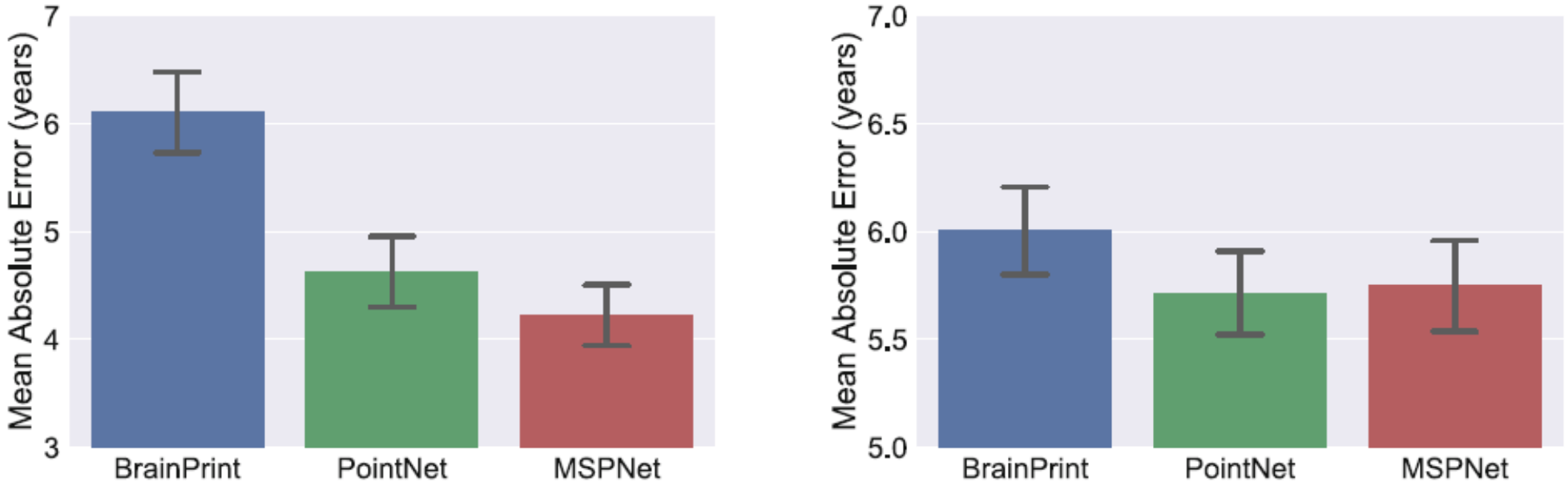
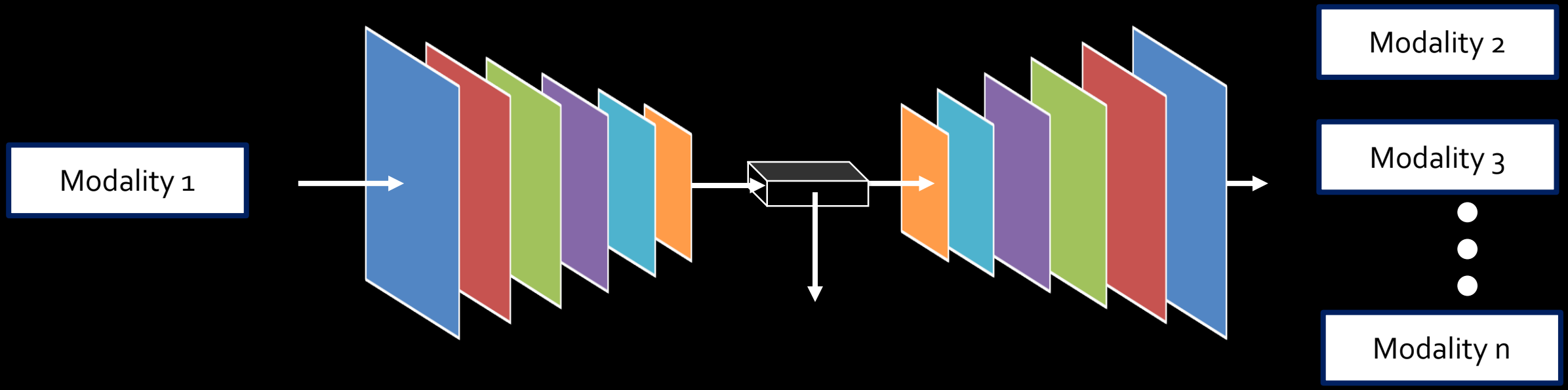


Fig. 3. Mean absolute error for the age prediction experiment on healthy subjects (left) and on all subjects, including MCI and AD (right)

Multi-modal Learning



Retinal Image Understanding Emerges from Self-Supervised Multimodal Reconstruction

Álvaro S. Hervella[1,2(✉)], José Rouco[1,2], Jorge Novo[1,2], and Marcos Ortega[1,2]

[1] CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain
[2] Department of Computer Science, University of A Coruña, A Coruña, Spain
{a.suarezh,jrouco,jnovo,mortega}@udc.es

Abstract. The successful application of deep learning-based methodologies is conditioned by the availability of sufficient annotated data, which is usually critical in medical applications. This has motivated the proposal of several approaches aiming to complement the training with reconstruction tasks over unlabeled input data, complementary broad labels, augmented datasets or data from other domains. In this work, we explore the use of reconstruction tasks over multiple medical imaging modalities as a more informative self-supervised approach. Experiments are conducted on multimodal reconstruction of retinal angiography from retinography. The results demonstrate that the detection of relevant domain-specific patterns emerges from this self-supervised setting.

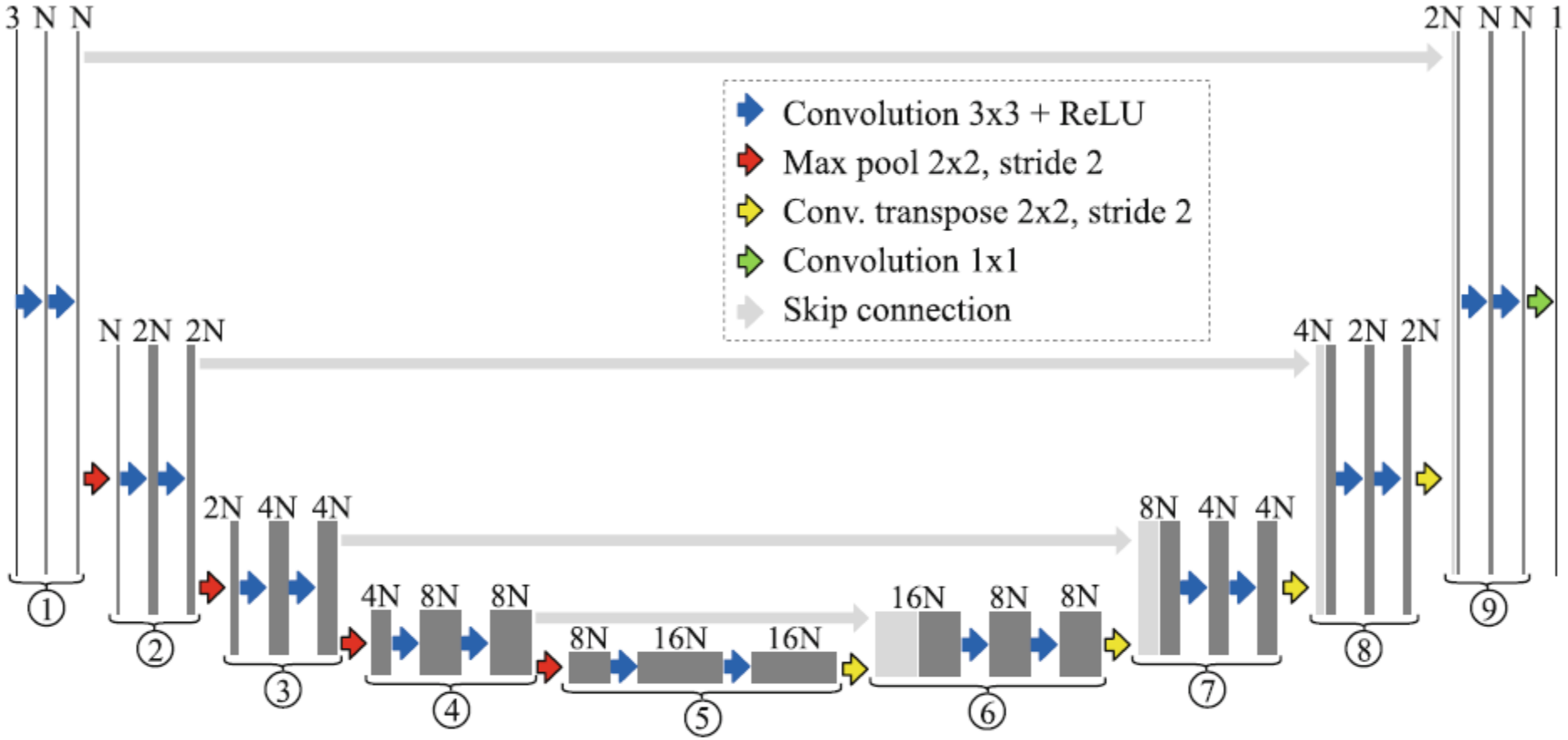


Fig. 1. U-Net architecture, where N is the number of base channels.

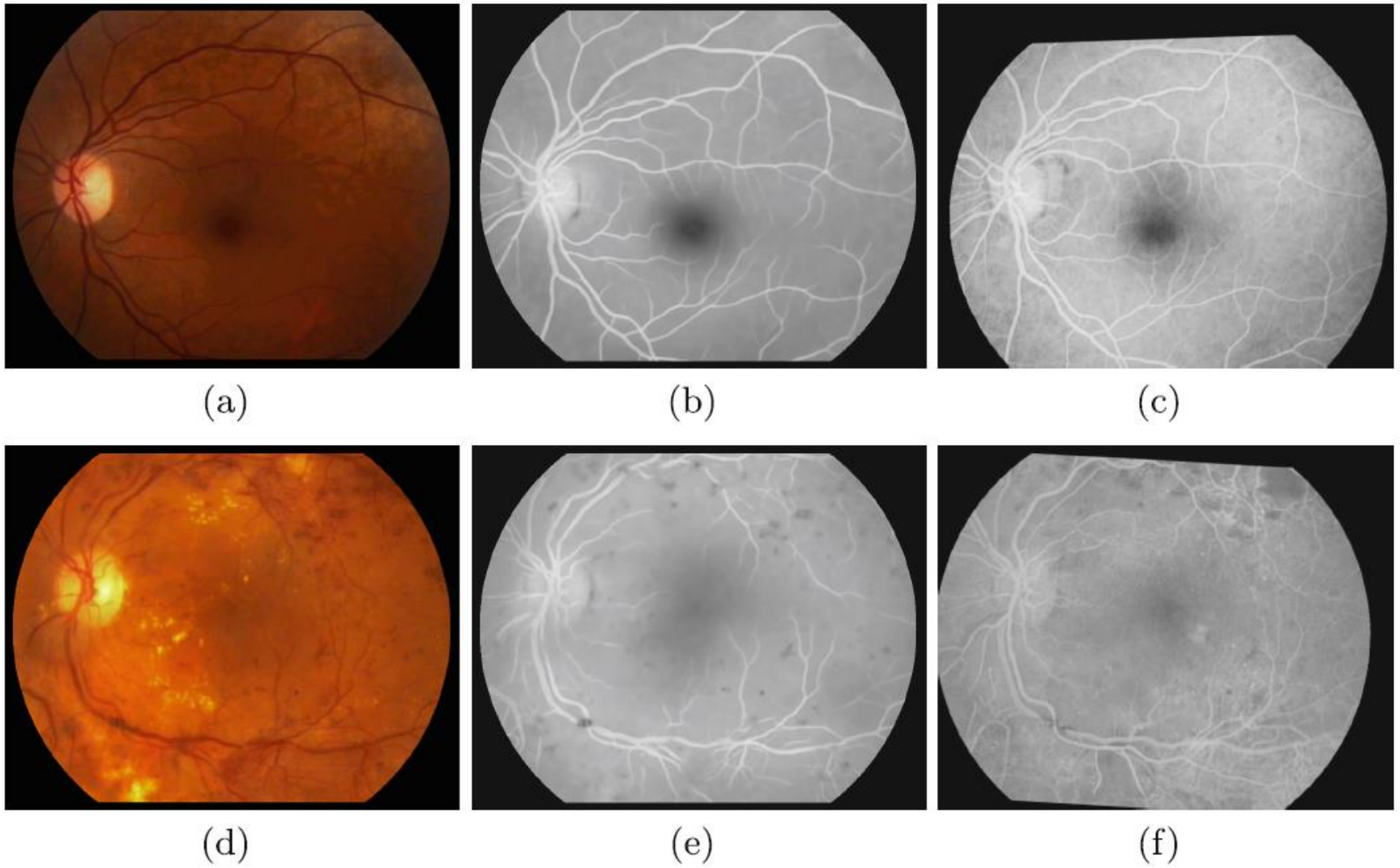
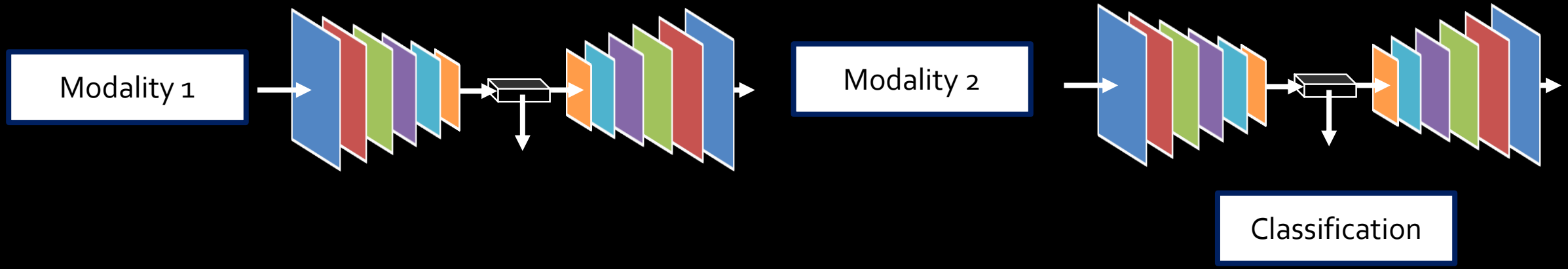


Fig. 3. Two examples of pseudo-angiography generation after training with SSIM: (a), (d) retinography; (b), (e) pseudo-angiography; (c), (f) registered angiography.

Multi-modal Learning



Volume-Based Analysis of 6-Month-Old Infant Brain MRI for Autism Biomarker Identification and Early Diagnosis

Li Wang[1(✉)], Gang Li[1], Feng Shi[2], Xiaohuan Cao[1], Chunfeng Lian[1], Dong Nie[1], Mingxia Liu[1], Han Zhang[1], Guannan Li[1], Zhengwang Wu[1], Weili Lin[1], and Dinggang Shen[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA
{li_wang, dinggang_shen}@med.unc.edu
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. Autism spectrum disorder (ASD) is mainly diagnosed by the observation of core behavioral symptoms. Due to the absence of early biomarkers to detect infants either *with* or *at-risk of* ASD during the first postnatal year of life, diagnosis must rely on behavioral observations long after birth. As a result, the window of opportunity for effective intervention may have passed when the disorder is detected. Therefore, it is clinically urgent to identify imaging-based biomarkers for early diagnosis and intervention. In this paper, *for the first time*, we proposed a volume-based analysis of infant subjects with risk of ASD at very early age, i.e., as early as at 6 months of age. A critical part of volume-based analysis is to accurately segment 6-month-old infant brain MRI scans into different regions of interest, e.g., white matter, gray matter, and cerebrospinal fluid. This is actually very challenging since the tissue contrast at 6-month-old is extremely low, caused by inherent ongoing myelination and maturation. To address this challenge, we propose an anatomy-guided, densely-connected network for accurate tissue segmentation. Based on tissue segmentations, we further perform brain parcellation and statistical analysis to identify those significantly different regions between autistic and normal subjects. Experimental results on National Database for Autism Research (NDAR) show the advantages of our proposed method in terms of both segmentation accuracy and diagnosis accuracy over state-of-the-art results.

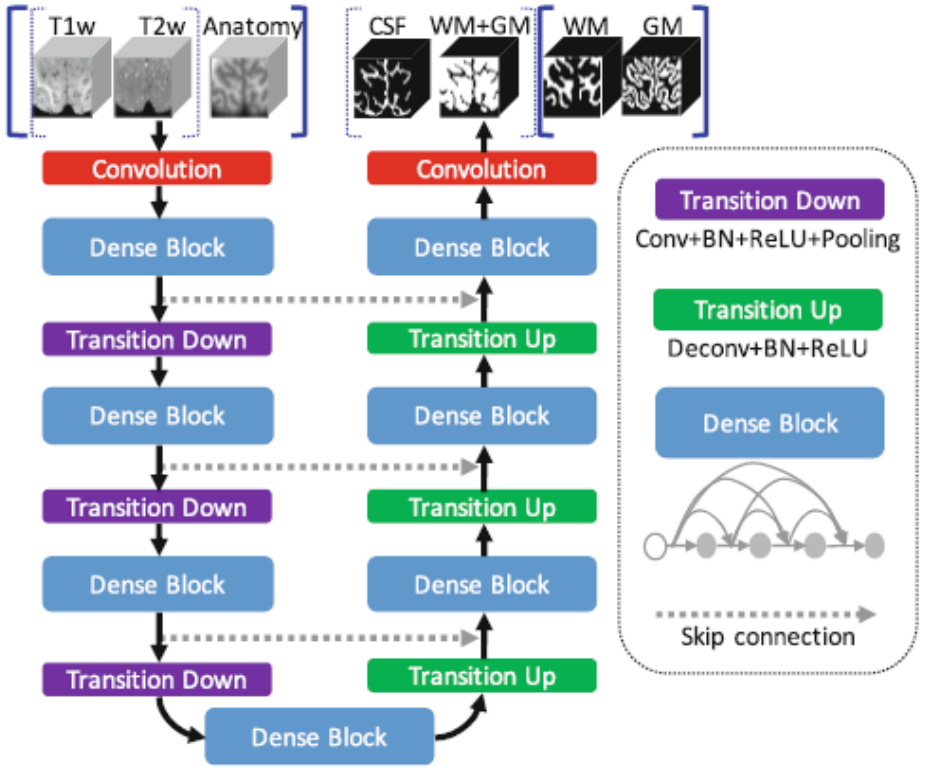


Fig. 2. Diagram of our architecture for segmentation. Input 1: T1w and T2w images for (CSF, WM+GM) segmentation to construct anatomy guidance; Input 2: T1w and T2w images and anatomy guidance for (WM, GM) segmentation.

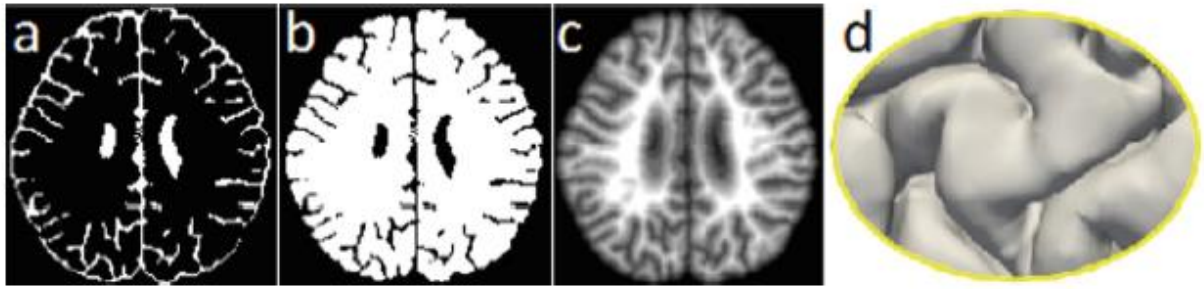
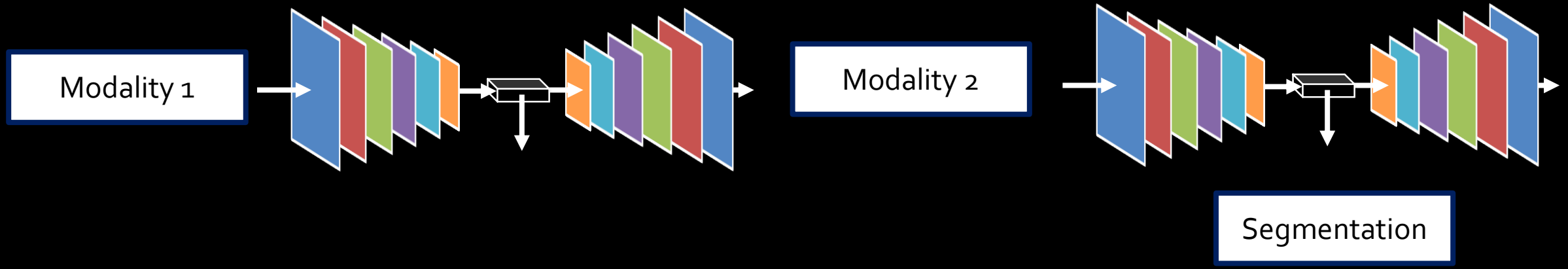


Fig. 3. (a) and (b) show the estimated CSF and WM + GM segmentations for a testing image in Fig. 1. (c) illustrates the signed distance function with respect to the outer surface shown in (d).

Multi-modal Learning



More Knowledge Is Better: Cross-Modality Volume Completion and 3D+2D Segmentation for Intracardiac Echocardiography Contouring

Haofu Liao[1(✉)], Yucheng Tang[3], Gareth Funka-Lea[2], Jiebo Luo[1], and Shaohua Kevin Zhou[4]

[1] Department of Computer Science, University of Rochester, Rochester, USA
hliao6@cs.rochester.edu
[2] Medical Imaging Technologies, Siemens Healthineers, Princeton, USA
[3] Department of Electrical and Computer Engineering, New York University, New York, USA
[4] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Abstract. Using catheter ablation to treat atrial fibrillation increasingly relies on intracardiac echocardiography (ICE) for an anatomical delineation of the left atrium and the pulmonary veins that enter the atrium. However, it is a challenge to build an automatic contouring algorithm because ICE is noisy and provides only a limited 2D view of the 3D anatomy. This work provides *the first automatic solution* to segment the left atrium and the pulmonary veins from ICE. In this solution, we demonstrate the benefit of building a *cross-modality framework* that can leverage a database of diagnostic images to supplement the less available interventional images. To this end, we develop a novel deep neural network approach that uses the (i) *3D geometrical information* provided by a position sensor embedded in the ICE catheter and the (ii) *3D image appearance information* from a set of computed tomography cardiac volumes. We evaluate the proposed approach over 11,000 ICE images collected from 150 clinical patients. Experimental results show that our model is significantly better than a direct 2D image-to-image deep neural network segmentation, especially for less-observed structures.

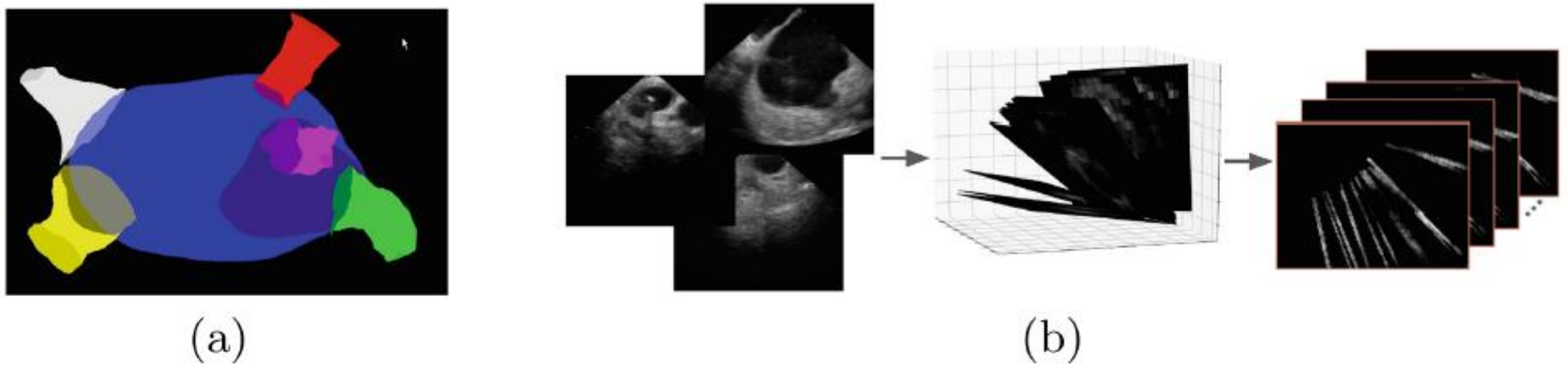


Fig. 1. (a) Graphical illustration of LA and its surrounding structures: blue-LA, green-left atrial appendage (LAA), red-left inferior pulmonary vein (LIPV), purple-left superior pulmonary vein (LSPV), white-right inferior pulmonary vein (RIPV), yellow-right superior pulmonary vein (RSPV). (b) 3D sparse ICE volume generation using the location information associated with each ICE image.

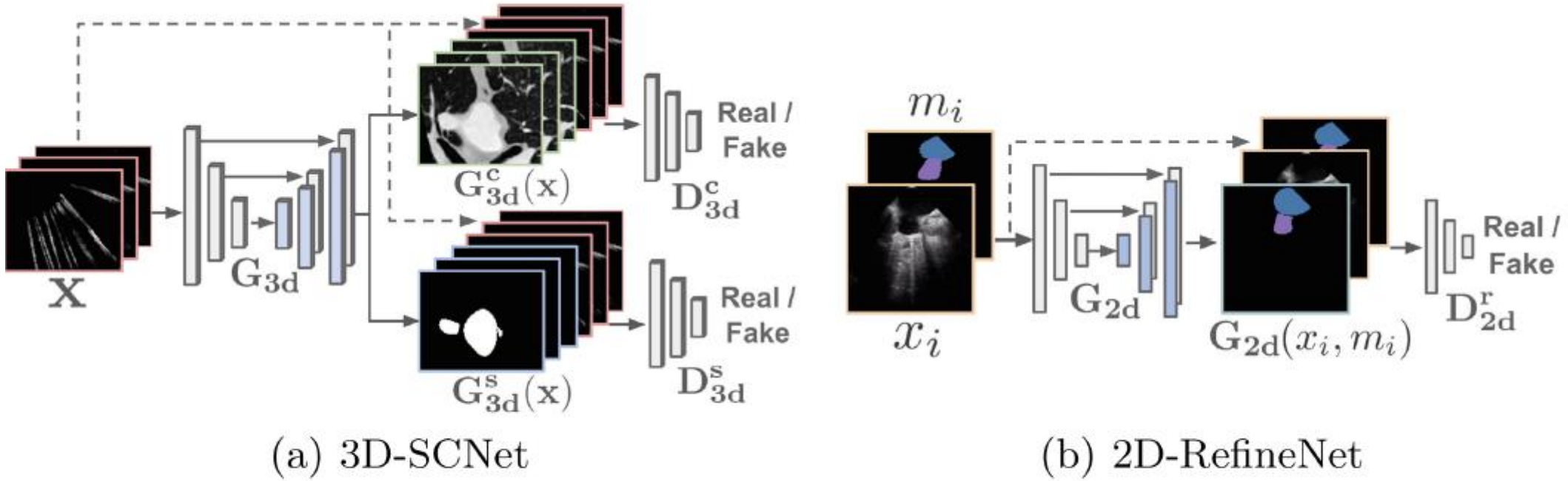


Fig. 2. The network architectures of the proposed method.

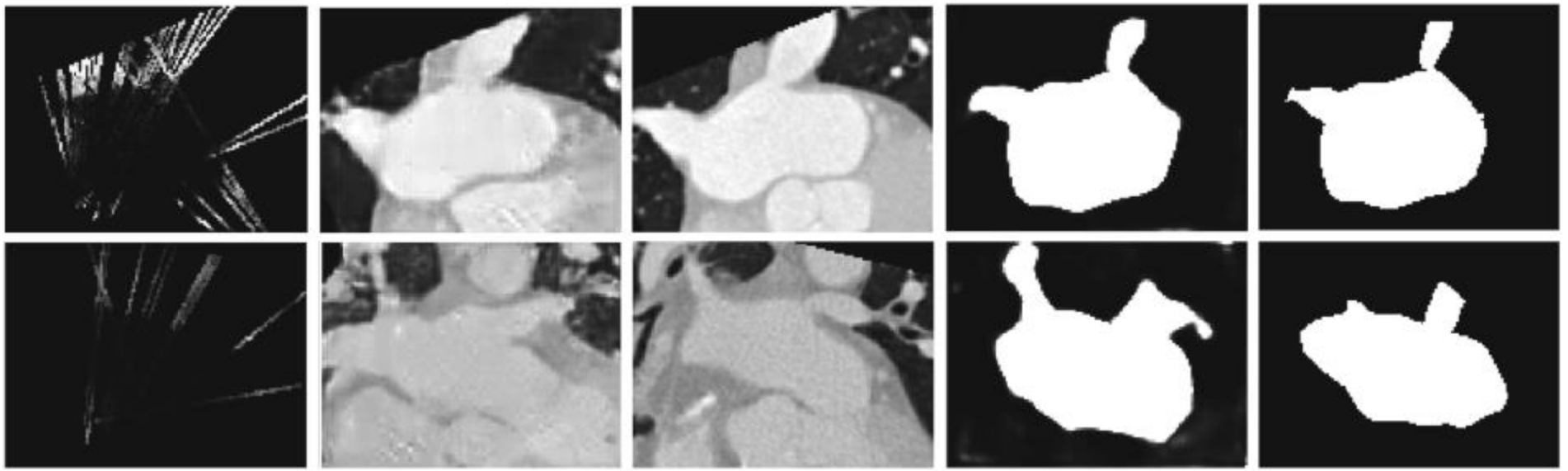


Fig. 3. Sparse volume segmentation and completion results for 2 cases. (a) Sparse ICE volume; (b) Completed CT volume; (c) the paired "ground truth" CT volume; (d) Predicted and (e) Ground truth 3D segmentation map.