Novel Deep Learning Models for Medical Imaging Analysis

by

Fei Gao

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved September 2019 by the
Graduate Supervisory Committee:

Teresa Wu, Chair
Jing Li
Hao Yan
Bhavika Patel

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

Deep learning is a sub-field of machine learning in which models are developed to imitate the workings of the human brain in processing data and creating patterns for decision making. This dissertation is focused on developing deep learning models for medical imaging analysis of different modalities for different tasks including detection, segmentation and classification. Imaging modalities including digital mammography (DM), magnetic resonance imaging (MRI), positron emission tomography (PET) and computed tomography (CT) are studied in the dissertation for various medical applications. The first phase of the research is to develop a novel shallow-deep convolutional neural network (SD-CNN) model for improved breast cancer diagnosis. This model takes one type of medical image as input and synthesizes different modalities for additional feature sources; both original image and synthetic image are used for feature generation. This proposed architecture is validated in the application of breast cancer diagnosis and proved to be outperforming the competing models. Motivated by the success from the first phase, the second phase focuses on improving medical imaging synthesis performance with advanced deep learning architecture. A new architecture named deep residual inception encoder-decoder network (RIED-Net) is proposed. RIED-Net has the advantages of preserving pixel-level information and cross-modality feature transferring. The applicability of RIED-Net is validated in breast cancer diagnosis and Alzheimer’s disease (AD) staging. Recognizing medical imaging research often has multiples inter-related tasks, namely, detection, segmentation and classification, my third phase of the research is to develop a multi-task deep learning model. Specifically, a feature transfer enabled multi-task deep learning model (FT-MTL-Net) is proposed to transfer high-resolution features from segmentation task to low-resolution feature-based classification task. The application of FT-MTL-Net on breast cancer detection, segmentation and classification using DM images is studied. As a continuing effort on exploring the transfer learning in deep models for medical application, the last phase is to develop a deep learning model for both feature transfer and knowledge from pre-training age prediction task to new domain of Mild cognitive impairment (MCI) to AD conversion prediction task. It is validated in the application of predicting MCI patients' conversion to AD with 3D MRI images.

DEDICATION

To my dear wife, parents and brothers, who have always been emotionally supportive.

ACKNOWLEDGMENTS

To begin with, I would like to thank my advisor, Dr. Teresa Wu, who taught me how to work independently and who encouraged and supported me during my research and dissertation. Her attitude on life and research profoundly influenced me during the last several years and also will help me through my entire life. I would also like to specially thank Dr. Jing Li for providing valuable suggestions and kind guidance. It was a great pleasure working with Dr. Wu and Dr. Li at the ASU-Mayo Clinic Center for Innovative Imaging (AMCII). I would also like to express gratitude to members of my committee for their insightful comments, valuable interactions, and advice: Dr. Hao Yan and Dr. Bhavika Patel.

Thank you to all members of AMCII, past and present: Min Zhang, Na Zou, Can Cui, Shuluo Ning, Bing Si, Hyunsoo Yoon, Congzhe Su, Xiaonan Liu, Yinlin Fu, Nathan Gaw, Lujia Wang, Yanzhe Xu, Zhiyang Zheng, Jiajing Huang and Hope Lancaster. I would like to warmly acknowledge my parents, parents-in-law and brothers for unconditionally supporting me all days and nights.

I met and married my dear wife during my Ph.D. studies, I regard this as the most important achievement I have gained. She deserves the utmost appreciation for her constant and unwavering support.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

1.1. Background

Over the past few decades, various medical imaging modalities including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and digital mammography (DM) are invented and introduced into clinical applications for disease diagnosis, prognosis and treatment assessment. These images need domain experts such as radiologists and physicians for interpretation and clinical decisions. With the technological advancements in image processing and analytic modeling, automatic models and systems are proposed. For example, different imaging processing models are applied to extract imaging features to quantify the characteristics of the raw image or region of interests (ROIs). These features are used to train machine learning models for specific tasks such as tumor detection, segmentation and classification, image synthesis and automatic diagnosis.

In recent years, deep learning models, especially the convolutional neural network (CNN), as a class in machine learning models which uses a cascade of convolutional layers and non-linear processing units for feature extraction and transformation, has attracted great attentions. The success of CNNs is mainly due to their powerful learning capability behind large set of parameters and the ability to derive ‘optimal’ hierarchical features from raw images to serve different tasks. Motivated by the success of deep learning in various application domains (e.g., computer vision), this dissertation research focuses on the applications on medical images. Different imaging modalities such as digital mammography (DM), magnetic resonance imaging (MRI), positron emission tomography

(PET) and computed tomography (CT) are studied for various applications including tumor segmentation, detection and classification, cross-modality synthesis and automatic diagnosis.

1.2. State of the Art

The tasks of deep CNNs in medical image analysis generally fall into four major categories: classification, detection, segmentation and synthesis. Classification for medical images mainly focuses on the discrimination of malignant lesions from benign or the identification of certain disease. To address this task, raw images or extracted patches with assigned labels are fed into the CNNs, features from different levels of the convolutional layers are updated based on the prediction results of each training iteration. After training, the model will obtain the capability of mapping the input image into a binary (classification) variable or multiple binary variables (multiclass classification). As one of the earliest CNN applications, Sahiner et al. (1996) propose a deep CNN to make predictions on benign or malignant for mammography patches containing ROIs. Arevalo et al. (2017) address the breast lesion classification problem through a deep learning model combining imaging features defined by domain experts. In Huynh, Li and Giger (2016), a pre-trained CNN on natural image patches is studied to address the lesion classification task. Gao et al. (2016) use a deep CNN to make holistic classification patches from CT images of lung into six classes (normal, emphysema, ground glass, fibrosis, micronodules and consolidation).

The task of detection aims at localizing the abnormal structures from the provided images. This task is addressed through the prediction of the centers, boundary or bounding box that contains the abnormal region. Ciresan et al. (2013) use a deep CNN to detect mitosis in breast cancer histology images. Sirinukunwattana et al. (2016) implement a spatially constrained CNN (SC-CNN) to detect nuclei in histopathology images. Roth et al. (2015) train a deep network with 2D CT images to detect five different parts of the body such as neck, liver, pelvis, lungs and legs.

In the task of segmentation, a probability map is generated for each pixel in 2D or voxel in 3D within input image to quantify the probability it belongs to the associated object. For instance, Pereira et al. (2016) implement a deep CNN with small-sized kernel for brain tumor segmentation in MRI images. Ronneberger, Fischer and Brox (2015) study the breast and fibroglandular tissue segmentation task through a deep CNN architecture named U-Net. Kleesiek et al. (2016) implement a 3D CNN architecture for the skull segmentation and extraction from T1-weighted MR images.

In image synthesis, a deep model (e.g., CNN) is launched to capture the non-linear mapping between the input images and the output images. The trained model can be used to generate a virtual image in scenarios where the desired image modality is not accessible. The very first published literature may be from Li et al. (2014) where a 4-layer shallow network, is developed to map the Positron Emission Tomography (PET) images from MRI. Improved diagnosis accuracy is observed after using the combination of MRI and synthetic PET for Alzheimer's disease. Yang et al. (2017) design a 4-layer CNN to reconstruct dual-energy subtraction soft-tissue chest image from a multi-scale gradient imaging of the original chest radiograph image. Han (2017) borrows the 'copy and crop' idea from U-Net and implements a 27-layer sCT-DCNN to generate virtual CT images from co-registered MRI images.

Multi-task learning (MTL) (Caruana, 1997) which attempts to handle multiple tasks at the same time emerges and has shown great promises in natural language processing (Collobert & Weston, 2008), speech recognition (Deng, Hinton, & Kingsbury., 2013), and computer vision (Girshick, 2015; He, Gkioxari, Dollar, & Girshick, 2017). The advantage of MTL in reducing risk of overfitting (Baxter, 1997; Ruder, 2017) and improving learning efficiency and prediction accuracy makes it an ideal solution for medical applications. For example, Akselrod-ballin et al. (2016) propose a faster R-CNN for mass detection and classification simultaneously. Samala, Chan, Hadjiiski, Helvie, and Cha (2018) take mass classification from digital mammograms and digitalized screen-film mammograms as two separate tasks and tackle the two tasks using a single framework based on the Visual Geometry Group (VGG) model (Noh, Hong, & Han, 2015). The study from Liu, Zhang, Adeli, and Shen (2018) focuses on AD using neuroimaging to conduct classification and predict clinical scores. Feng, Nie, Wang, and Shen (2018) propose a multi-task residual fully convolutional network (FCN) to segment organs (e.g. bladder, prostate and rectum) and estimate the intensities.

MTL is to take advantage of the joint power from multiple tasks on the limited data available in medical applications, to some extent. To directly address the data limitation issue, transfer learning is extensively investigated. In transfer learning, the deep model is first pre-trained on large size of labelled dataset (e.g., natural images) to capture the features, and is then fine-tuned on the target dataset. For example, Hon et al. (2017) use the VGG16 and Google Inception v4 CNN model to pre-train images from the ImageNet Challenge dataset and fine-tune the last fully connected layer on the MRI images for the final AD diagnosis. Similarly, Hosseini-Asl et al. (2018) pre-train a 3D Convolutional

Auto-Encoder (CAE) to feature extractions and fine-tune on fully connected layers with a Softmax layer for AD diagnosis.

The success of deep learning in medical applications motivates this dissertation. We identify five major research challenges and issues as the focus of this research.

- **Challenge I: requiring domain knowledge for network architecture design, parameters tuning and human-defined feature selection.** Existing deep learning models in medical image analysis usually need to be trained from raw images. Various architecture options (e.g. pooling layers, activation function, shortcut connections) and hyper-parameters (e.g. learning rate, batch size) need to be set before the training, and the settings have significant impacts on the model's performance. In order to obtain the relatively 'optimal' settings, multiple training trails are conducted which is computationally costly. Incorporating domain knowledge into the deep learning model is highly desirable. In addition, as reviewed above, some CNNs are combined with human-defined features for improved performance; and the selection of features for specific task requires prior knowledge on both the task and the features. However, these two important prerequisites are not always available for researchers in medical area.
- **Challenge II: lack of consideration for the missing modality.** Different imaging modalities (e.g., MRI, PET) provide complementary information for the disease diagnosis and staging. However, not all imaging modalities readily available for the patients. We note most existing methods focus on developing

deep learning models on the modality of images available only. There is a need to address the “missing” modality for improved disease diagnosis.

- **Challenge III: lack of innovation and pertinence on architectures for emerging applications such as image synthesis.** Developing deep learning models for image synthesis is an emerging field. The existing models directly take the network architectures used in image segmentation for image synthesis task. While there are some similarities in these two different tasks, image synthesis is more complicated as the prediction for each pixel (voxel) is the actual intensity value instead of a label (e.g., 1 indicates within the object, 0 indicates outside the object). As a result, more information is needed for image synthesis, and a new model to tackle imaging synthesis is needed and is currently lacking.
- **Challenge IV: lack of consideration for the joint power from different tasks in the training procedure.** We contend that detection, segmentation and classification for the same image or ROI shall share some common information/knowledge. However, most existing research is to design and train different models for different tasks, separately, failed to recognize the complementary nature of the task and potential joint forces from each task for improved performances.
- **Challenge V: lack of considering the usage of knowledge captured from pre-training procedure in transfer learning.** Transfer learning methods usually include two separate procedures: pre-training and fine-tuning. Research thus far mainly focuses on transferring the features from the pre-training into

the fine-tuning and ignores the knowledge captured in the pre-training stage. This research is to address this challenge.

1.3. Expected Original Contribution

The overall objective of this dissertation is to develop innovative deep learning methods that overcome the aforementioned limitations and demonstrate their utility in the medical imaging applications, including detection, segmentation, classification and image synthesis. The expected original contributions include:

- **Development of a deep learning system integrated with free feature generator and synthetic image provider for improved diagnosis performance:** A deep learning system is proposed that adopts a deep CNN architecture (ResNet) pre-trained with natural images as a feature generator. The system is integrated with a shallow CNN to generate virtual advanced modality image as additional feature source and a gradient boosting model as classifier. The proposed system, shallow-deep convolutional neural network (SD-CNN), shows significant improvement on breast cancer diagnosis. Details are discussed in Chapter 2.
- **Development of an advanced deep learning architecture for improved image synthesis performance:** A new deep CNN architecture for image synthesis is proposed. The new architecture addresses the potential issues of losing pixel information and gradient vanishing problem faced with other state-of-art models. The goal is achieved through the novel design of encoder-decoder architecture and residual inception blocks. The proposed model is

validated using two datasets: digital mammography (for breast cancer) and MRI (for AD diagnosis). Details are discussed in Chapter 3.

- **Development of a multi-task CNN model to save training efforts and improve performance of individual task through combing features from parallel task:** A feature transfer enabled multi-task deep learning model (FT-MTL-Net) is developed which combines features from segmentation task to further improve classification accuracy. Three contributions come out of the FT-MTL-Net. First, to our best knowledge, it may be one of the first fully automatic deep learning systems in medical imaging that can be trained end-to-end through a unified cost function and solve the tasks of tumor detection, segmentation, and classification simultaneously. Second, it enables feature transfer from a segmentation task to a classification task. The features from both high resolution (transferred from segmentation) and low resolution (existing features) are adopted to help improve the classification accuracy. Third, the features transferred are re-weighted based on the prior knowledge from the segmentation probability map; As a result, information from irrelevant regions is excluded, and the feature map is representative of the tumor regions only Details are discussed in Chapter 4.
- **Development of a CNN model which trained through transfer learning and utilizes pre-training task result as additional biomarker for improved classification performance:** An age-adjust neural network (AD-Net) is proposed. In the AD-NET, we revisit the transfer learning to make the pre-trained model serves dual purpose: (1) feature transferring: similar to existing

research from literature, the pre-trained model without the last layer is used as feature extractor; (2) knowledge transferring: the whole pre-trained model is kept into the fine-tuning stage to transfer the knowledge captured in the age prediction process. Details are discussed in Chapter 5.

1.4. Dissertation Organization

The proposed dissertation research will be presented in the following four chapters. Specifically, Chapter 2 presents the development of topic (I): SD-CNN: a Shallow-Deep CNN for Improved Breast Cancer. Chapter 3 presents the development of topic (II): Deep Residual Inception Encoder-Decoder Network for Medical Imaging Synthesis. Chapter 4 presents the development of topic (III): A Feature Transfer Enabled Multi-Task Deep Learning Model on Medical Imaging. Chapter 5 presents the development of topic (IV): AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction.

CHAPTER 2

SD-CNN: A SHALLOW-DEEP CNN FOR IMPROVED BREAST CANCER DIAGNOSIS

2.1. Introduction

Although about 1 in 8 U.S. women (~12%) will develop invasive breast cancer over the course of her lifetime (U.S. Breast Cancer Statistics, 2018), breast cancer death rates have been steadily and/or significantly decreasing since the implementation of the population-based breast cancer screening program in late 1970s due to the early cancer detection and the improved cancer treatment methods (Rosenquist & Lindfors, 1998). Among the existing imaging modalities, full field digital mammography (FFDM) is the only clinically acceptable imaging modality for the population-based breast cancer screening, while Ultrasound (US) and Magnetic Resonance Imaging (MRI) are also used as adjunct imaging modalities to mammography for certain special subgroups of women (Lehrer et al., 2012). However, using FFDM is not an optimal approach in breast cancer screening due to its relatively low detection sensitivity in many subgroups of women. For example, although FFDM screening has an overall cancer detection accuracy of 0.75 to 0.85 in the general population, its accuracy in several subgroups of the high-risk women including those with positive BRCA (BReast CAncer) mutation or dense breasts decreases to 0.30 to 0.50 (Elmore, Armstrong, Lehman, & Fletcher, 2005). On the other hand, using dynamic contrast enhanced breast MRI can yield significantly higher cancer detection performance due to its ability to detect tumor angiogenesis through contrast enhancement and exclude suspicious dense tissues (Warner et al., 2004). Yet, its substantially higher cost, lower accessibility and longer imaging scanning time forbids breast MRI being used as a primary imaging modality in breast cancer screening and detection. In addition, lower image resolution of breast MRI is a disadvantage as comparing to FFDM.

In order to combine the advantages of both FFDM and MRI, a new novel imaging modality namely, contrast-enhanced digital mammography (CEDM) emerges and starts to attract broad research and clinical application interest. CEDM is a recent development of digital mammography using the intra-venous injection of an iodinated contrast agent in conjunction with a mammography examination. Two techniques have been developed to perform CEDM examinations: the temporal subtraction technique with acquisition of high-energy images before and after contrast medium injection and the dual energy technique with acquisition of a pair of low and high-energy images only after contrast medium injection. During the exam, a pair of low and high-energy images is obtained after the administration of a contrast medium agent. The two images are combined to enhance contrast uptake areas and the recombined image is then generated (Fallenberg et al., 2014). In CEMD, it has low energy (LE) imaging, which is comparable to routine FFDM and recombined imaging similar to breast MRI. Comparing to breast MRI, CEDM exam is about 4 times faster with only about 1/6 the cost (Patel et al., 2017). In addition, CEDM imaging has 10 times the spatial resolution of breast MRI. Therefore, CEDM can be used to more sensitively detect small residual foci of tumor, including calcified Ductal Carcinoma in Situ (DCIS), than using MRI (Patel et al., 2017). Several studies including prospective clinical trials conducted at Mayo Clinic have indicated that CEDM is a promising imaging modality that overcomes tissue overlapping (“masking”) occurred in FFDM, provides tumor neovascularity related functional information similar to MRI, while maintaining high image resolution of FFDM (Cheung et al., 2014; Fallenberg et al., 2014;

Gillman, Toth, & Moy, 2014; Luczyńska et al., 2014). Unfortunately, CEDM as a new modality is yet widely available in many other medical centers or breast cancer screening facilities in the U.S. and/or across the world limiting its broad clinical impacts.

In clinical breast imaging (US, MRI, FFDM and CEDM), reading and interpreting the images remains a difficult task for radiologists. Currently, breast cancer screening has high false positive recall rate (i.e., ≥ 10%). Computer-aided detection (CADe) and diagnosis (CADx) schemes (Tan et al., 2014; Carneiro et al., 2017; Gao et al., 2016; Muramatsu et al., 2016) have been developed and demonstrated the clinical potentials to be used as "the second reader" to help improve radiologists' performance in the diagnosis. In order to overcome the limitation of lower accessibility to CEDM systems and help radiologists more accurately conduct the diagnosis, this research proposes the development and validation of a new CADx scheme, termed Shallow-Deep Convolutional Neural Network (SD-CNN). SD-CNN combines image processing and machine learning techniques to improve the malignancy diagnosis using FFDM by taking advantages of information available from the CEDM.

CNN is a feed-forward artificial neural network that has been successfully implemented in the broad computer vision areas for decades (Lecun, Bengio, & Hinton, 2015; LeCun, Bottou, Bengio, & Haffner, 1998). As it evolves, different CNN models have been developed and implemented. The computational resource and devices available in recent years make the training of CNN with large number of layers (namely, the deep CNN) possible. Applying deep CNNs in image recognition was probably first demonstrated in ImageNet competition (Russakovsky et al., 2015) back in 2012. Since then, it has become a popular model for various applications ranging from natural language processing, image segmentation to medical imaging analysis (Cha et al., 2016; Tajbakhsh et al., 2016, Wang et al., 2017). The main power of a deep CNN lies in the tremendous trainable parameters in different layers (Eigen, Rolfe, Fergus, & LeCun, 2013; Zeiler & Fergus, 2014). These are used to extract discriminative features at different level of abstraction (Tajbakhsh et al., 2016). However, training a deep CNN often requires a large volume of labeled training data, which may not be easily available in medical applications. Secondly, training a deep CNN requires massive computational resources, as well as rigorous research in architecture design and hyper-parameters tuning. To address these challenges, a promising solution is transfer learning (Banerjee et al. 2017), that is, a deep CNN model is trained followed by a task-specific parameter fine-tuning process. The trained models are established by experienced researchers using publicly labeled image datasets. For a specific task, the model is often treated as a feature generator to extract features describing the images from abstract level to detailed levels. One can then develop classification models (SVMs, ANNs, etc.) using the derived features. Promising results have been reported in several medical applications, such as chest pathology identification (Bar, Diamant, Wolf, & Greenspan, 2015), breast mass detection and classification (Samala et al., 2016), just to name a few. While exciting, earlier CNN models such as AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), GoogLeNet (Simonyan & Zisserman, 2014) and VGGNet (Szegedy et al., 2014) are known to suffer from gradient vanishing when the number of layers increases significantly. A newer model, ResNet (Kaiming He, Zhang, Ren, & Sun, 2014) with a "short-cut" architecture is recently proposed to address the issue. The imaging competition results show the ResNet outperforms other CNN models by at least 44% in classification accuracy.

The potentials CNN brings to medical imaging research are not limited to deep CNN for imaging feature extraction. A second area that medical research can benefit is indeed using CNN for synthetic image rendering. Here an image is divided into a number of smaller patches fed into a CNN (e.g., 4-layer CNN in this research) as the input and the output is a synthetic image. The CNN is trained to learn the non-linear mapping between the input and output images. Several successful applications have been reported, such as synthesizing positron emission tomography (PET) imaging (Li et al., 2014) or CT image (Han, 2017; Nie et al., 2016) from MRI image, and from regular X-ray to bone-suppressed recombined X-ray (Yang et al., 2017).

Motivated by this two-fold applicability of CNN, this research proposes a Shallow-Deep CNN (SD-CNN) as a new CAD scheme to tackle the unique problem stemmed from the novel imaging modality, CEDM, for breast cancer diagnosis. My first hypothesis is that applying a deep CNN to CEDM is capable of taking advantage of recombined imaging for improved breast lesion classification due to the contribution from the tumor functional image features. Second, in order to expand the advantages of CEDM imaging modality to the regular FFDM modality, we hypothesize that a shallow CNN is capable to discover the nonlinear mapping between LE and recombined images to synthesize the “virtual” recombined images. As a result, traditional FFDM can be enriched with the “virtual” recombined images. The objective of this study is to validate these two hypotheses by using a unique study procedure and two imaging datasets of both CEDM and FFDM images. The details of the study procedures and experimental results are reported in the following section of this chapter.

2.2. Materials

In this research, two separate datasets are used, which include a dataset acquired from tertiary medical center (Mayo Clinic Arizona), and a public dataset from INbreast (Moreira et al., 2012).

2.2.1 Institutional Dataset from Mayo Clinic Arizona:

Based on Institutional Review Board (IRB) approved study and data collection protocol, we reviewed CEDM examinations performed using the Hologic Imaging system (Bedford, MA, USA) between August 2014 and December 2015. All patients undertaken CEDM had a BI-RADS (Breast Imaging Reporting and Data Systems) (Liberman, L. and Menell, J.H., 2002) rating of 4 and 5 in their original FFDM screening images. Due to the detection of highly suspicious breast lesions, CEDM was offered as an adjunct test to biopsy in a clinical trial environment. All CEDM tests were performed prior to the biopsies. In summary, the patient cohort in this clinical trial had the following criteria: 1) the diagnostic mammogram was rated BI-RADS 4 or 5, and 2) histopathology test result was available from surgical or image-guided biopsy. We limited the cohort to BIRADS 4 and 5 lesions because the analysis required the gold standard of lesion pathology. 49 cases were identified that met the above inclusion criteria, which include 23 benign and 26 cancer biopsy-proven lesions. We analyzed one lesion per patient. If a patient had multiple enhancing lesions, the annotating radiologist used the largest lesion to ensure best feature selection. In CEDM, there are cranial-caudal (CC) and mediolateral-oblique (MLO) views for both LE and recombined images. Figure 1 illustrates the example views on the LE and recombined images, respectively.

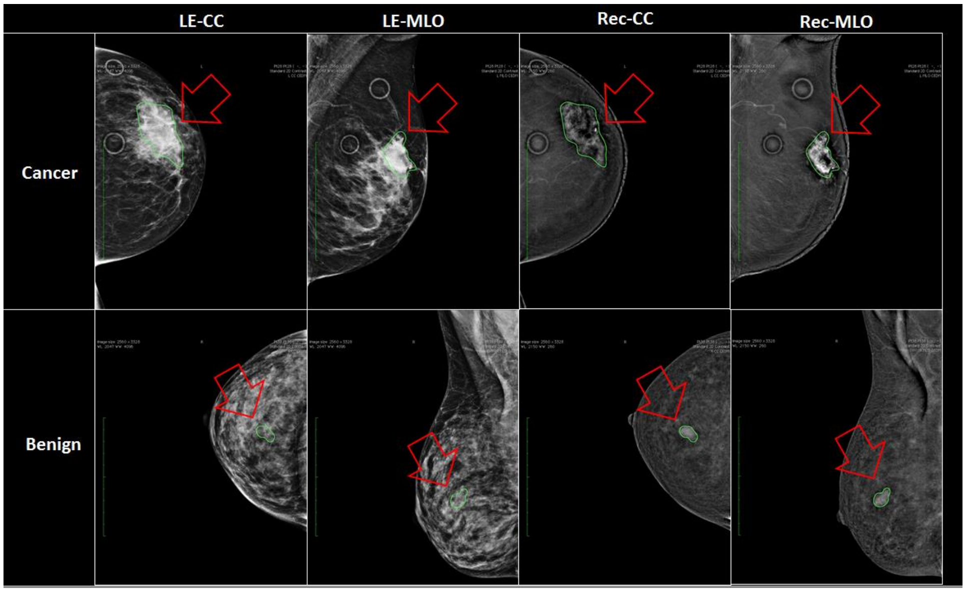


Figure 1 Example of breast images (Cancer and Benign) for LE and recombined (Rec) images with 2 views (CC and MLO) (Lesions are highlighted with green circle).

For the 49 cases, all CEDM images with DICOM format were de-identified and transferred from the clinical PACS to a research database and loaded into the open source image processing tool OsiriX (OsiriX foundation, Geneva, Switzerland) (Rosset, Spadola, & Ratib, 2004). DICOM images were anonymized and prepared for blinded reading by a radiologist. A fellowship trained breast radiologist with over 8 years of imaging experience interpreted the mammogram independently and used the OsiriX tool to outline lesion contours. Contours were drawn on recombined images (both CC and MLO views) for each patient on recombined images. These contours were then cloned onto LE images. All lesions were visible on both view CC and MLO views. This information is further used in the imaging pre-processing (see details in methodology section). Some examples LE and recombined images are shown in Figure 1. As observed, LE images are not as easy as recombined images to visualize the lesions for both cancerous and benign cases.

2.2.2 INbreast Public Dataset:

This dataset was obtained from INbreast, an online accessible full-field digital mammographic database (Moreira et al., 2012). INbreast was established by the researchers from the Breast Center in CHJKS, Porto, under the permission of both the Hospital's Ethics Committee and the National Committee of Data Protection. The FFDM images were acquired from the MammoNovation Siemens system with pixel size of 70 mm (microns), and 14-bit contrast resolution. For each subject, both CC and MLO view were available. For each image, the annotations of region of interests (ROIs) were made by a specialist in the field, and validated by a second specialist. The masks of ROIs were also made available. In this research, a dataset of 89 subjects was extracted by including subjects that have BI-RADS scores of 1, 2, 5 and 6. Subjects with BI-RADS 1 and 2 are regarded as benign tumor, and subjects with BI-RADS 5 and 6 are regarded as cancer. For each subject, images of CC and MLO view are used for feature extraction.

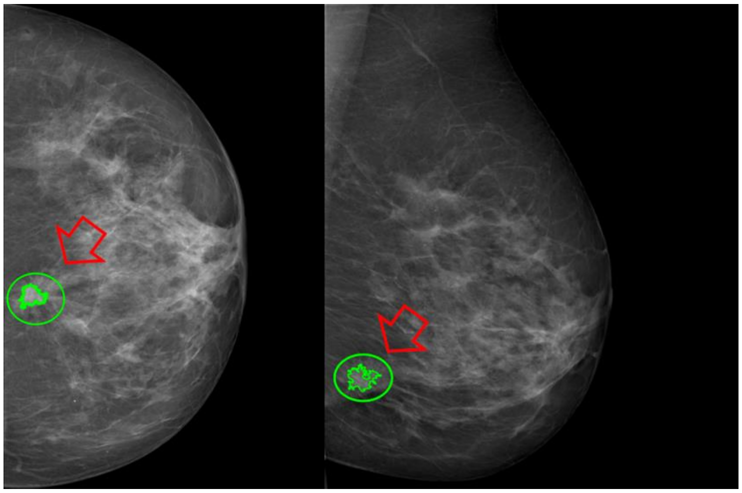


Figure 2 Example of breast images for FFDM images from INbrease dataset with 2 views (CC on left and MLO on right) (Lesions are highlighted with green circle).

2.3. Methodology

To fully explore the advantages of CNNs and CEDM in breast cancer research, a Shallow-Deep CNN (SD-CNN) is proposed (Figure 3). First, we develop a Shallow-CNN from CEDM to discover the relationships between LE images and recombined images. This Shallow-CNN is then applied to FFDM to render "virtual" recombined images. Together with FFDM, a trained Deep-CNN is introduced for feature extraction followed by classification models for diagnosis. Note for CEDM, we can start the workflow with the Deep-CNN directly.

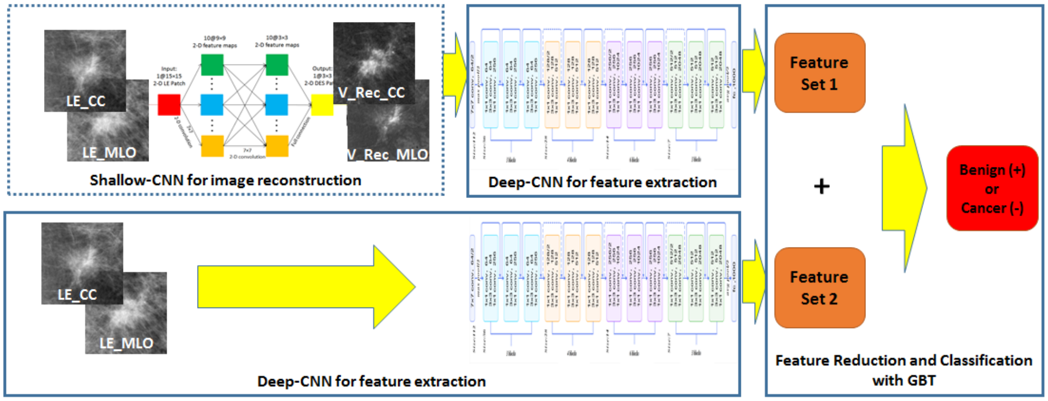


Figure 3 Architecture of Shallow-Deep CNN.

2.3.1 Image Pre-processing

Before the Deep-CNN and Shallow-CNN are employed, a four-step imaging pre-processing procedure is launched. First, for each image we identify a minimum-area bounding box that contains the tumor region. Specifically, for each tumor, we have a list of boundary points with coordinates in pair (x,y) available. The bounding box is decided using the (x_{min}, y_{min}) and (x_{max}, y_{max}) as the two diagonal corner points to ensure the box covers the whole tumor area. Note we have CC and MLO views for FFDM and we have

CC and MLO views for both LE and recombined images for CEDM. As a result, there are two images from FFDM and four images from CEDM. The bounding box size varies case by case due to different sizes of tumors (ranging from 65×79 to 1490×2137 in this study). Next, an enlarged rectangle that is 1.44 times (1.2 times in width and 1.2 times in height) the size of bounding box is obtained. The enlarged bounding box approach is to include sufficient neighborhood information proved to increase the classification accuracy (Lévy & Jain, 2016). In the second step, this 'enlarged' rectangle is extracted and saved as one image. The third step is to normalize the image intensity to be between 0 and 1 using the max-min normalization. In the last step, the normalized images are resized to 224×224 to fully take advantage of trained ResNet model. Here we take the patches that contain tumor instead of the whole image as input. This is because the focus of the study is on tumor diagnosis and we believe the features generated by the deep-CNN from the tumor region shall better characterize the tumor, especially for the cases where the tumor region is small.

2.3.2 Shallow-CNN: Virtual Image Rendering

Inspired by the biological processes (Elmore et al., 2005), CNNs use a variation of multilayer perceptions designed to require minimal preprocessing. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. This process is simulated through different layers (convolutional, pooling, fully connected). A CNN's capability is hidden behind the large amount trainable parameters which can be learned iteratively through gradient descent algorithms. In this research, a 4-layer CNN is implemented to model the latent relationship between the LE images (patches) and recombined images (patches). The model is then used to render “virtual” recombined images (patches) from FFDM images (patches).

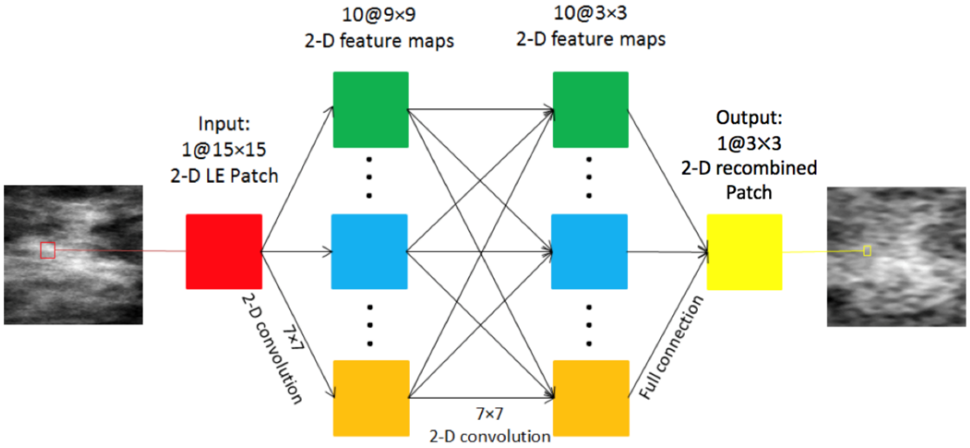


Figure 4 Architecture of 4-layer shallow-CNN for “virtual” recombined image rendering.

2.3.3 Deep-CNN: Feature Generation

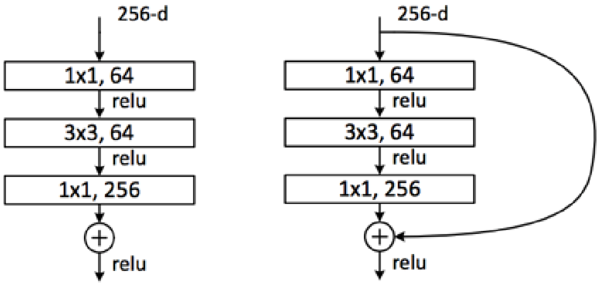


Figure 5 Building blocks for traditional CNNs (left) and ResNet (right) (He et al., 2014)

ResNet is a trained deep CNN developed in 2015 with a revolutionary architecture using the “short-cut” concept in the building block. As seen in Figure 5, the output of building blocks takes both final classification results and the initial inputs (the short-cut) when updating the parameters. As a result, it outperforms traditional deep-CNNs which are known to suffer from higher testing error since gradient tends to vanish as the number of layers increases (Kaiming He et al., 2014). ResNet has different versions with 50, 101 and 152 layers but all based on the same building blocks. In the ImageNet competition, ResNet-50, ResNet-101 and ResNet-152 have comparable performances (top 5 error: 5.25% vs. 4.60% vs. 4.49%), but with quite different numbers of parameter (0.85M vs. 1.7M vs. 25.5M). For the consideration of balance between computation efficiency and accuracy, especially for the limited computation resources, we adopt ResNet-50 in this research.

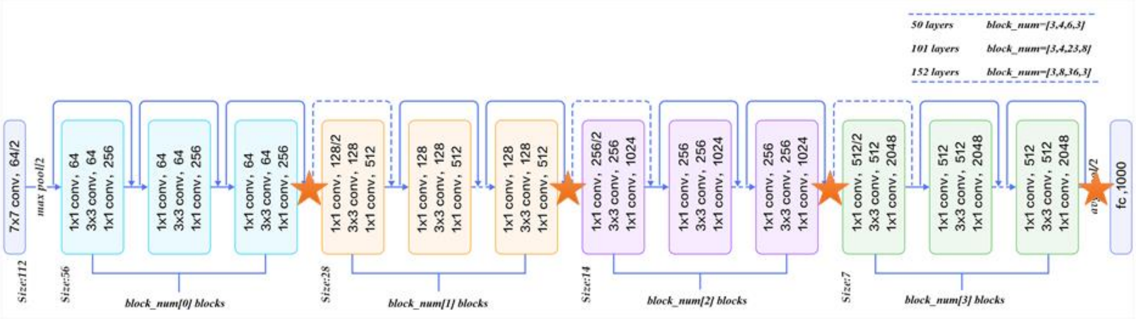


Figure 6 Architecture of ResNet (K He, Zhang, & Ren, 2016) (Red star are placed in layers where features are extracted; Dotted shortcuts increase feature dimensions by zero-padding; based on the output dimension of building blocks, the ResNet is divided into 4 different building blocks (BBs), they are shown with different colors in the figure (BB_1: blue, BB_2: orange, BB_3: purple, BB_4: green). Different version of ResNets vary in the number of BBs, for instance, the 50-layer version ResNet has 3 BB_1s, 4 BB_2s, 6 BB_3s and 3 BB_4s).

In general, ResNet consists of four types of buildings blocks. The CNN structures and the number of features for each block are shown in Figure 6. We mark them with different colors. For simplicity, let blue for block type 1, orange for block type 2, purple for block type 3 and green for block type 4. ResNet-50 is defined as [3, 4, 6, 3] meaning that it has 3 type 1 blocks, 4 type 2 blocks, 6 type 3 blocks and 3 type 4 blocks. The output features are extracted from the final layer of each block type, that is, layer 10, 22, 40 and

49. Since we have no prior knowledge about the feature performance, we decide to take the features from all four layers (10, 22, 40 and 49) for the classification model development. For each feature map, the mean value is calculated and used to represent the whole feature map. The number of features extracted from each layer is listed in Table 1. For each view, we have 3840 (256+512+1024+2048) total features.

Table 1 Number of features from each layer for one image.

Layer #	10	22	40	49
# of features	256	512	1024	2048

2.3.4 Classification

Boosting is a machine learning ensemble meta-algorithm aiming to reduce bias and variance (Bauer, Kohavi, Chan, Stolfo, & Wolpert, 1999). It converts weak learners to strong ones by weighing each training sample inversely correlated to the performance of previous weak learners. Gradient boosting trees (GBT) is one of the most powerful boosting ensemble decision trees used in regression and classification tasks (Yang et al., 2017). It builds the model in a stage-wise fashion, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. The nature of GBT makes it robust to overfitting by measuring the criterion it used when splitting the tree nodes. In addition, it provides the importance of each feature in the regression/classification for the ease of interpretation which is desirable in the medical applications. In GBT, the feature importance is related to the concept of Gini impurity (Rokach et akk., 2008). To compute

Gini impurity ($\mathrm{I_G}(p)$) for a set of items with $\boldsymbol{J}$ classes, suppose $i \in \{1,2,\dots,J\}$, and let p_i be the fraction of items labeled with i class in the set, we have:

$$\mathrm{I_G}(p) = 1 - \sum_{\mathrm{i=1}}^{\mathrm{J}} p_i^2 \tag{2.1}$$

When constructing each decision tree in the boosting classifier, a feature is used to divide the parent node into two children nodes based on a threshold. Since the decision tree is constructed with the goal being to minimize the overall Gini impurity, the post-splitting Gini impurity shall be smaller than the pre-splitting Gini impurity. The reduced Gini impurity thus can be used to as a measure of the contribution from the feature in the process of splitting the tree. The training procedure is to identify the optimal splitting features that offer the maximum impurity reduction (Yang et al., 2017) among the whole feature set. The process of building trees serves as feature selection and classification.

2.4. Experiments and Results

The overall objective of this research is to demonstrate the clinical utility of our novel SD-CNN approach for breast cancer diagnosis. Therefore, we conduct two sets of experiments. The first experiment is to validate the values from recombined images for improved breast cancer diagnosis. Deep CNN, ResNet is applied. The second experiment is to investigate the feasibility of applying SD-CNN to enrich the traditional FFDM for improved diagnosis. A public FFDM dataset from INbreast is used and the results are compared with six state-of-the-art algorithms.

2.4.1 Experiment I: Validating the Improved Accuracy in Breast Cancer Diagnosis on CEDM using Deep-CNN

The workflow of our first experiment is shown in Figure 7. Using 49 CEDM cases collected from Mayo Clinic Arizona, we first conduct the experiments using LE from CEDM images. For each subject in the dataset, LE images (both CC and MLO views) are processed through pre-processing procedure described in Section 3.3.1, after which 2 patches (224×224) are extracted. They are fed into the trained ResNet. As features from different layers of ResNet describe the image from different scales and aspects, in this research, we have all the features fed into the GBT to classify the case as cancer vs. benign. The procedures are implemented with a python library named “sklearn”. Different settings to prevent the model from overfitting are used. For example, we set maximum depth of individual tree to be 3, use early stopping strategy by setting number of decision tree to be 21, max number of features to be searched for each split is $\sqrt{N}$ (N is the number of features), the minimal number of samples falling in each leaf node is 2. Other settings are set to be default.

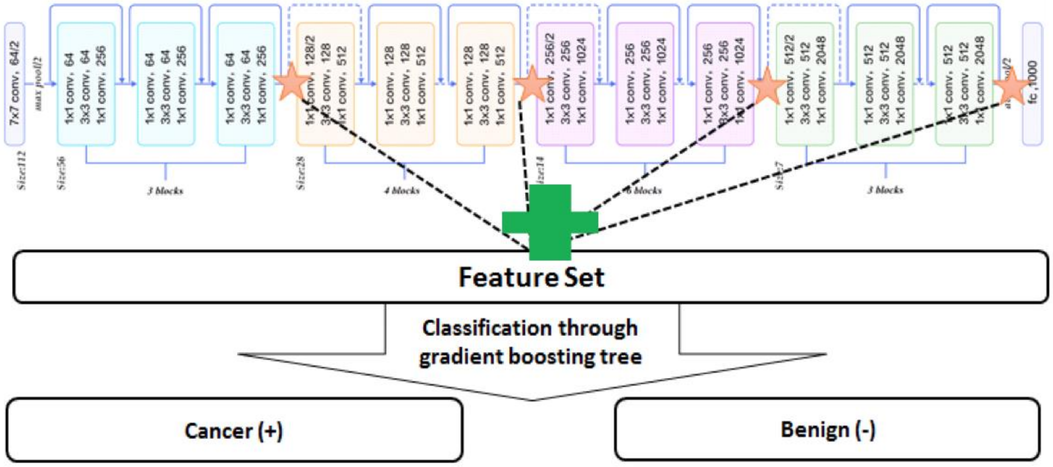


Figure 7 Workflow of Experiment I.

Next, we study the added values from recombined images for improved diagnosis. Specifically, CC and MLO view from recombined images are fed into the same pre-processing and feature generating procedure. The combination of LE and recombined image features are used in the classification model. Performance is measured based on leave-one-out cross validation to fully use the training dataset which is limited in size. Performance metrics are accuracy, sensitivity and specificity, and area under receiver operating characteristic curve (AUC) (see Table 2). The ROC curves for two models are shown in Figure 8. By using all the LE features generated by ResNet, we obtain the accuracy of 0.85 (Sensitivity=0.89 Specificity=0.80) and 0.84 for AUC. With additional features from recombined image, the model accuracy is improved to 0.89 (Sensitivity=0.93 Specificity=0.86) and AUC to 0.91.

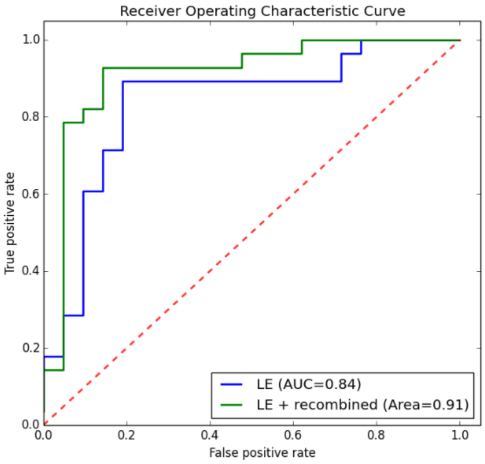


Figure 8 Receiver operating characteristic curve for the model using FFDM image only vs. FFDM and recombined image.

Table 2 Classification Performance of Experiment Using LE Images vs. LE and Recombined Images.

Metric	**LE**	**LE and Recombined images**
Accuracy	0.85	0.89
Sensitivity	0.89	0.93
Specificity	0.80	0.86
AUC	0.84	0.91

To explore the features contributing to the classification model, we calculated the contribution of each feature, and track the source image for each feature. The feature's importance score is measured through calculating the total impurity reduction when building the ensemble trees. (Note that the feature importance is calculated inside each leave-one-out loop, and the final result is the average for each feature among the loop). Table 3 summarizes the importance scores for the features from different sources (LE vs. Recombined Image). Here the scores are normalized by dividing individual score with summation of all scores. From Table 3, we observe among all the 99 features used in the model, 56 are from LE images which contribute 76.84% of the impurity reduction, 43 features are from recombined images which contribute to 23.16% in the modeling. The features from the recombined images help improve the accuracy of breast cancer diagnosis from 0.85 to 0.89.

Table 3 Contribution of features from different image sources.

Image Source	Number of features	Contribution of impurity reduction
LE image	56	76.84%
Recombined image	43	23.16%

2.4.2 Experiment II: Validating the Value of "Virtual" Recombined Imaging in Breast Cancer Diagnosis on FFDM Using SD-CNN

The improved performance by adding the features from recombined images motivates us to study the validity of constructing and using the "virtual" recombined images from FFDM images for breast cancer diagnosis.

Here we first develop a 4-layer shallow CNN that learns the nonlinear mapping between the LE and recombined images using the same 49 CEDM dataset. CC and MLO view images are regarded as separate training data, so a total of 98 images are used, in which 5 subjects (10 images) are selected as validation material, and the rest 44 subjects (88 images) are sued as training material. By randomly extracting 2500 pair of training samples within masked tumor from each LE (input) and recombined image (output), a training dataset of 220000 (88×2500) samples is generated. The input samples for the CNN are 15×15 patches from LE images, the same input size as in (Li et al., 2016). Considering the relatively small receptive field and complexity of a shallow CNN, we set the output samples size as 3×3, it is our intention to explore the impact of the different output patch size for the breast cancer diagnosis as one of our future tasks. The output patches from recombined image are centered in the same position as input patches from the LE image. The input and output samples are fed into the CNN framework implemented with package of “Keras”. The CNN has 2 hidden layers, with 10 7*7 filters in each layer. There are 5K trainable parameters through backpropagation with mini-batch gradient decent algorithm to increase learning speed. Batch size is set to be 128. The learning rate is set to be 0.01, ReLu activation function is used in all layers except the output layer, where activation function is not used. Other parameters are set to be default by “Keras” package. Finally, with the trained CNN and patches extracted from available modality, we can construct a “virtual” recombined image by assembling predicted patches into a whole image.

We use mean squared error (MSE) to evaluate the similarity between the “virtual” recombined image and the true recombined image for the 10 images in validation dataset. MSE measures the pairwise squared difference in intensity as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}|TRecombined(i) - VRecombined(i)|^2 \qquad (2.2)$$

Where N is total number of pixel in the selected patches, *TRecombined(i)* and *VRecombined (i)* are the intensity values for the same position in patches from the true recombined image and corresponding virtual recombined image.

For the 10 validation images, the MSE is 0.031 (standard deviation is 0.021). For illustration purpose, we choose four samples to demonstrate the resulting “virtual” recombined images vs. the true recombined images (Figure 9). As seen, the abstract features (e.g., shape) and some details of the tumor from true recombined images are restored by the “virtual” recombined images.

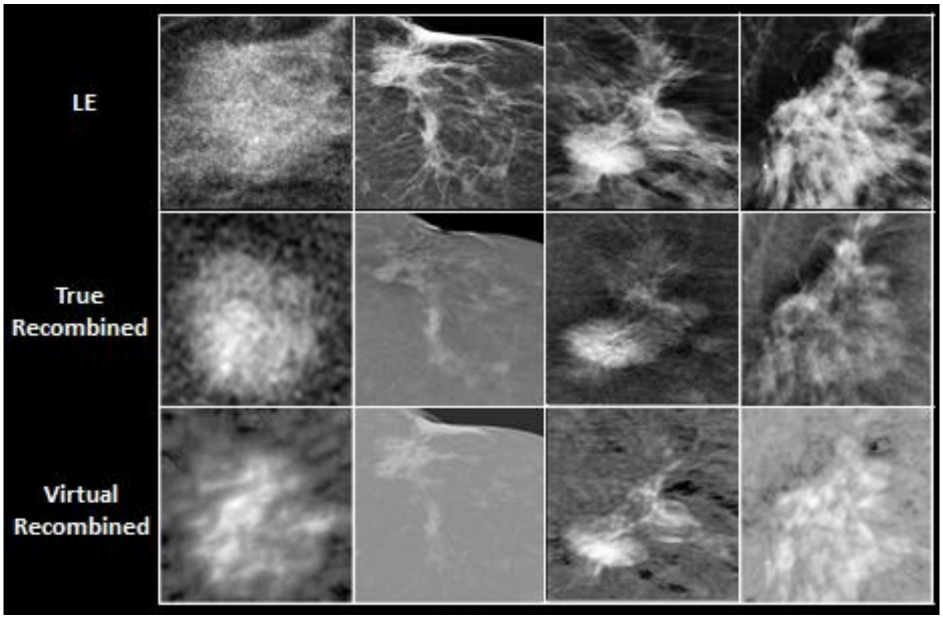


Figure 9 Sample images of LE image, true recombined image and its corresponding virtual recombined in dataset I. (from left to right: benign, cancer, cancer, cancer).

With this trained shallow-CNN, we used the 89 FFDM cases from INbreast dataset to render the "virtual" recombined images. Specifically, for each subject, we slide the 15×15 window from left to right, top to bottom (step size = 1) in FFDM image, to get the input patches. The input patches are fed into the trained 4-layer CNN, from which we get the predicted virtual recombined image patches (3×3) as outputs. The small patches are placed at the same position as their corresponding input patches in the "virtual" recombined images. For the position with overlapping pixels, the values are replaced with mean value for all overlapping pixels. At last, the "virtual" recombined images are rendered. Figure 10 illustrates some example FFDM images and their corresponding "virtual" recombined images. One clinical advantage of recombined image is it filters out dense tissues which often lead to false positive diagnosis. As seen from Figure 10, the "virtual" recombined images preserve this advantage. Specifically, dense tissues surrounding tumors are excluded in “virtual” recombined images, making the core region easier to be identified (left two cases in Figure 10). For the benign cases on the right, as the suspicious mass is mostly filtered out, it is mainly composed of dense tissues.

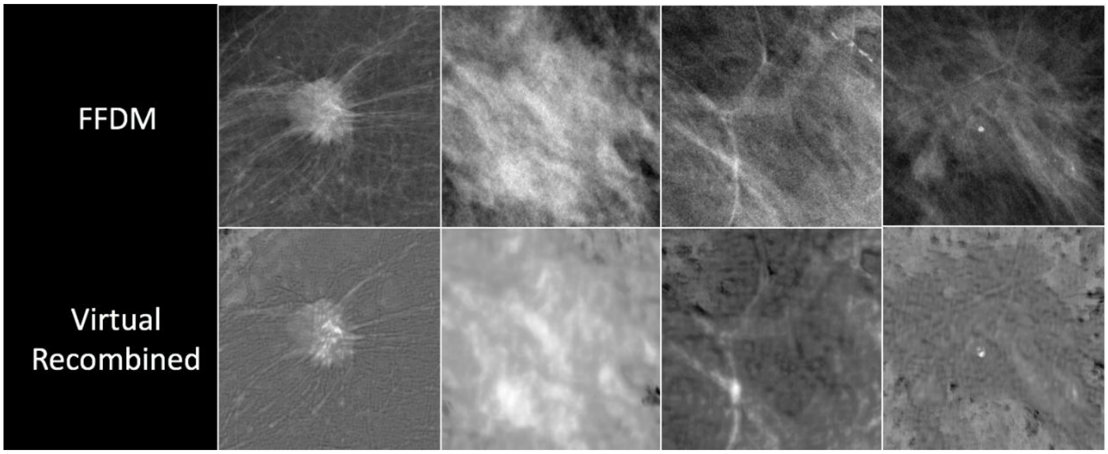


Figure 10 Sample images of FFDM in dataset II and its corresponding “virtual” recombined (Two cases on left are cancerous with BI-RADS = 5, two cases on right are benign with BI-RADS=2).

Next, following the same procedure as the first experiment, we apply the ResNet on the FFDM alone, and on both FFDM and “virtual” recombined images together. ResNet is used for feature extraction followed by the GBT ensemble classifiers. The parameter for GBT settings is further tuned since the training dataset is slightly imbalanced (benign: cancer = 30: 59). The training weights for benign and cancer are set to be 1 and 0.5. Numbers of trees set to be 31. Other parameter settings remain the same as the first experiment and 10-fold cross validation is used. Figure 11 shows the mean ROC curves for the model on FFDM alone vs. the model on FFDM and the “virtual” recombined image. The mean AUC for the classifier using FFDM features is 0.87 ± 0.12, while after adding the features from virtual recombined image, the AUC is increased to 0.92 ± 0.14. It is interesting to observe from Figure 11 that sensitivities (true positive rate) of the two models have similar performance, the specificities (1 – false positive rate) vary greatly. We want to highlight the importance of specificity as breast cancer screening has high false positive recall rate (i.e., ≥ 10%). One known fact is that the probability that a woman will have at least one false positive diagnosis at 10 years screening program is 61.3% with annual and 41.6% with biennial screening (Michaelson et al., 2016). This will lead to additional MR exams (extra cost) and even biopsy. Another side effect is the negative psychological impacts. In this research, the use of recombined images ("virtual" recombined images) shows the great potential to address these challenges by improving the specificity. In Table 4, we summarize the model performances in terms of accuracy, sensitivity and specificity (threshold is set to be 0.75). While we observe that the model on FFDM vs. the model on FFDM and "virtual" recombined image show no significant differences on accuracy, sensitivity and even AUC, the performance on specificity shows significant improvements ($p<0.05$).

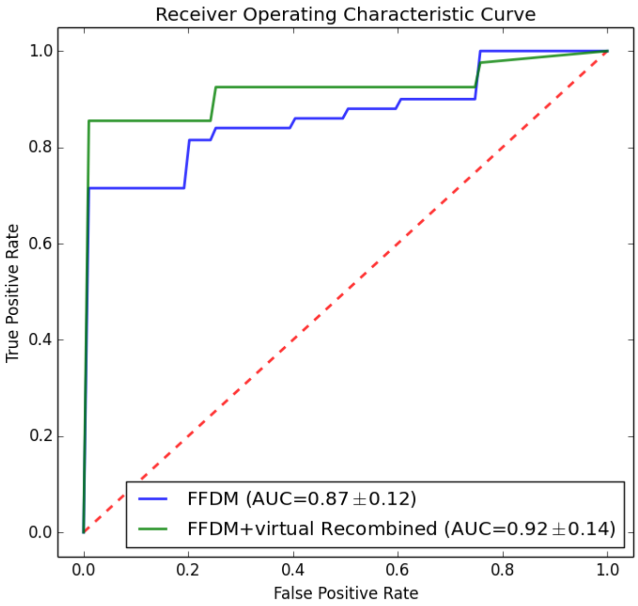


Figure 11 Receiver operating characteristic curve for the model using FFDM image only verse FFDM and virtual recombined.

Table 4 Classification Performance of Experiment Using FFDM Imaging vs. FFDM + Recombined imaging.

Metric	**FFDM**	**FFDM + Virtual Recombined**	**P value**
Accuracy	0.84 ± 0.09	0.90 ± 0.06	0.14
Sensitivity	0.81 ± 0.16	0.83 ± 0.16	0.91
Specificity	0.85 ± 0.12	0.94 ± 0.04	**< 0.05**
AUC	0.87 ± 0.12	0.92 ± 0.14	0.28

In looking into the contributions from the features (Figure 5), the use of “virtual” recombined imaging features improves the performances in terms of both accuracy and AUC. Calculation of contribution follows the same procedure as experiment I and is conducted inside each cross-validation loop. Among all the 154 features used in this experiment, 87 are from the “virtual” recombined image, which contribute 77.67% of the total impurity reduction. The rest 67 features are from LE images, and they contributed the rest 22.33% impurity reduction. It is interesting to observe from this experiment that the contributions from "virtual" recombined images are higher than the contributions from the true recombined images from the first experiment. One reason may be the second dataset has denser tissue cases and it is believed recombined images shall be more useful in diagnosing the dense breast cases. This is yet to be confirmed with the radiologists which is our immediate next step.

Table 5 Contribution of features from different image sources.

Image Source	**Number of features**	**Contribution of impurity reduction**
LE image	67	22.33%
Virtual Recombined image	87	77.67%

We further explore the state-of-the-art algorithms using the same INbreast dataset and compare our methods against the eight methods from the literature (see Table 6). As seen, our approach using "virtual" recombined image outperforms six algorithms in terms of both accuracy and AUC. We want to highlight that one of papers by Dhungel et al. (2017) proposes four approaches. Among the four, the best performer has a 0.95 in accuracy and 0.91 in AUC, and the second performer has a 0.91 in accuracy and 0.87 in AUC. We conclude our approach has better AUC (0.92) comparing to both while inferior in accuracy (0.90). We contend that indeed, AUC is a more robust metric in the medical research and it is considered to be more consistent and have better discriminatory power comparing to accuracy (Huang et al. 2005).

Table 6 Classification performance for using FFDM feature alone and using features from FFDM and "virtual" recombined and other state-of-the-art methods using INbreast dataset.

Method	Acc.	AUC
Random Forest on features from CNN with pre-training (Dhungel, Carneiro, & Bradley, 2017)	0.95±0.05	0.91±0.12
CNN + hand crafted features pre-training (Dhungel et al., 2017)	0.91±0.06	0.87±0.06
Random Forest + hand crafted features pre-training (Dhungel, Carneiro, & Bradley, 2015)	0.90±0.02	0.80±0.15
CNN without hand crafted features pre-training (Dhungel et al., 2017)(Dhungel et al., 2017)	0.72±0.16	0.82±0.07
Multilayer perceptron (Sasikala, 2016)	0.88	0.89
Lib SVM (Diz, Marreiros, & Freitas, 2016)	0.89	0.90
Multi-kernel classifier (Augusto, 2014)	NA	0.87
Linear Discriminant analysis (Domingues et al., 2012)	0.89	NA
Our proposed approach on FFDM only	0.84±0.09	0.87±0.12
Our proposed approach on both FFDM and "virtual" Recombined Image	0.90±0.06	0.92±0.14

2.5 Discussion and Conclusion

Differentiating benign cases from malignant lesions is one of the remaining challenges of breast cancer diagnosis. In this study, we propose a SD-CNN (Shallow-Deep CNN) to study the two-fold applicability of CNN to improve the breast cancer diagnosis. One contribution of this study is to investigate the advantages of recombined images from CEDM in helping the diagnosis of breast lesions using a Deep-CNN method. CEDM is a promising imaging modality providing information from standard FFDM combined with enhancement characteristics related to neoangiogenesis (similar to MRI). Based on our review of literature, no existing study has investigated the extent of CEDM imaging potentials using the deep-CNN. Using the state-of-art trained ResNet as a feature generator for classification modeling, our experiment shows the features from LE images can achieve accuracy of 0.85 and AUC of 0.84, adding the recombined imaging features, model performance improves to accuracy of 0.89 with AUC of 0.91.

Our second contribution lies in addressing the limited accessibility of CEDM and developing SD-CNN to improve the breast cancer diagnosis using FFDM in general. This the first study to develop a 4-layer shallow CNN to discover the nonlinear association between LE and recombined images from CEDM. The 4-layer shallow-CNN can be applied to render "virtual" recombined images from FFDM images to fully take advantage of the CEDM in improved breast cancer diagnosis. Our experiment on 89 FFDM dataset using the same trained ResNet achieves accuracy of 0.84 with AUC of 0.87. With the "virtual" recombined imaging features, the model performance is improved to accuracy of 0.90 with AUC of 0.92.

While promising, there is room for future work. First of all, the trained ResNet is a black-box feature generator and the features extracted may not be easy to be interpreted by the physicians. It is our intention to discover possible clinical interpretations from the features as one of our future tasks. For example, as the ResNet goes deeper, initial layers may represent the raw imaging characteristics as the first order statistics, the deeper layer of the features may represent the morphological characteristics (e.g., shape). This is yet to be explored. A second future work is related to the patch sizes. We plan to assess impacts of the different sized patches for both input and output images on the breast cancer diagnosis.

CHAPTER 3

DEEP RESIDUAL INCEPTION ENCODER-DECODER NETWORK FOR IMAGE SYNTHESIS

3.1. Introduction

During the last decade, precision medicine as an approach considering individual variability in the diagnosis and treatment has emerged as a novel paradigm for healthcare. One cornerstone for precision medicine is medical imaging. Tremendous efforts have been dedicated to medical imaging research which in general can be categorized in four areas: imaging-based classification, imaging object detection, imaging segmentation and imaging synthesizing. The emerging Convolutional Neural Network (CNN) has been successfully introduced into all these areas with different focuses (Greenspan, Ginneken, & Summers, 2016). Imaging classification and detection work on the object of interest (e.g., tumor). Specifically, classification is to categorize the object, for example, to be benign vs. malignant, in which the entire image or the extracted region of interest (ROI) is fed into a CNN, with one or more probabilities or class labels as the outputs. As early as 1996, a 4-layer CNN is implemented to classify regions of interest (ROIs) from mammogram as either biopsy-proven masses or normal tissues (Sahiner et al., 1996). Since then different CNNs have been introduced for various classification tasks including breast lesion (Araujo et al., 2017; Huynh, Li, & Giger, 2016a), lung pattern (Microbiana et al., 2016), skin lesion (Yap, Yolland, & Tschandl, 2018) or pulmonary peri-fissural nodules (Ciompi et al., 2015), just name a few. The task of detection is to derive an envelope box to enclose the object. In the area of detection, bounding boxes or patches centered on the candidate objects are identified and CNN-based detectors are trained to find boxes that truly contain desired objects. Applications include colonic polyps in CT images (Roth et al., 2016), cerebral microbleeds from MRI scans (Dou et al., 2016), and nuclei in histopathological images (Sirinukunwattana et al., 2016). Please note both classification and detection are interested in the objects thus the requirement on the pixel level details could be much relaxed.

There is another category of problems known as dense prediction. It requires the pixel-level specifics and that is the research focus from imaging segmentation and synthesis. In segmentation, a probability map that quantifies the likelihood of each pixel being within the imaging object (e.g., tumor) is generated. Successful implementations have been reported in brain tumor/structures segmentation (Havaei et al., 2017; Zhang et al., 2015; Zhao et al., 2018), epithelial tissue in prostatectomy (Bulten, Litjens, Hulsbergen-van de Kaa, & van der Laak, 2018), etc. In another application (Zhang et al., 2015), a four layer CNN is designed to take T1, T2 Magnetic Resonance images (MRI) and Fractional Anisotropy (FA) image as inputs and the outputs are the segmentation maps for three types of tissues, namely white matter, gray matter and cerebrospinal fluid. To do so, a local response normalization layer is implemented between the convolutional layer and the final fully connected layer to enforce competitions between features at the same spatial location across different feature maps and thus improve the segmentation results. A fully convolutional neural network (FCNN) collaborated with random fields in a unified framework is proposed to segment brain tumor regions in MRI images(Zhao et al., 2018). The same FCNN is introduced in another task of epithelial tissue segmentation (Bulten et al., 2018). In another application (Havaei et al., 2017), a two-pathway CNN architecture is proposed to harvest both local features and global contextual features simultaneously and improve the brain tumor segmentation result. As research on exploring CNN on segmentation progresses, a notable new architecture, U-Net (Ronneberger, Fischer, & Brox, 2015) emerges. One application of U-Net is to segment neuronal structures in electron microscopic stacks. The novel design of a contracting path to capture context and a symmetric expanding path to enable precise localization improve the segmentation performance significantly (Ronneberger et al., 2015). Following the success, U-Net and its variants are studied in a number of medical imaging segmentation problems. For instance, it is implemented for joint craniomaxillofacial bone segmentation and landmark digitization (Shen, Tang, Chen, J.Xia, & Shen, 2018). A 3D U-Net is proposed in volumetric imaging segmentation for Xenopus kidney (Liu, Li, Luo, Loy, & Tang, 2016). V-Net (Milletari, Navab, & Ahmadi, 2016), an extension of U-Nets with added shortcut connections between different layers, is introduced to segment prostate from 3D volumetric images.

Imaging systhtesis tackles a different dense prediction problem. It is to discover the pixel-wise nonlinear associations between the input images and the output images. Imaging synthesis has great potentials in medical applications, especially in the scenarios where some imaging modalities may be of limited access or missing due to various reasons such as cost (Litjens et al., 2017). As a new field, to the best of our knowledge, the very first published literature may be from Li (Li et al., 2014). To test the innovative idea, a 4-layer shallow network, is developed to map the Positron Emission Tomography (PET) images from MRI. Improved diagnosis accuracy is observed after using the combination of MRI and synthetic PET for Alzheimer’s disease. In another research (Yang et al., 2017), a 4-layer CNN is designed to reconstruct dual-energy subtraction soft-tissue chest image from a multi-scale gradient imaging of the original chest radiograph image. Another interesting effort is related to breast cancer research. Full Field Digital Mammography (FFDM) is the mainstay in breast cancer screening program but is known to suffer from diagnosis accuracy. Contrast Enhanced Digital Mammography (CEDM) is a recent development mammography which has a low energy imaging comparable to FFDM and recombined imaging taking advantage of high-energy images (Patel et al., 2017). While promising, as a new modality, CEDM has not been widely available in many medical centers in the U.S. and worldwide. To tackle this accessibility issue, a SD-CNN (Gao et al., 2018) is proposed to render synthetic recombined images from FFDM thus significantly improve the breast cancer diagnosis using FFDM. Similarly, a 4-layer CNN is implemented to map the low energy (FFDM) images to the recombined images (Gao et al., 2018). The research reviewed above is taking the proof-of-the-concept approach exploring the applicability of 4-layer network in imaging synthesis. The aforementioned 4-layer network is shallow and simpler compared to deep networks used in imaging classification, detection and segmentation. Therefore, most research only handles the images by taking small patches from the ROIs extracted through the images. For example, in the experiments of some research (Gao et al., 2018; Li et al., 2014), most ROIs are smaller than 400×600 pixels and the size of training patches is 15×15 pixels. We contend this approach may work well for smaller images or under the condition where ROI is provided. For the later cases, the involvement from domain experts (e.g. radiologist) is required. An ideal solution for synthetic imaging is a CNN capable to handle the whole image. A shallow network with limited learning power may suffer while a deep network may be the promising network to be explored. This is because a deep network has much more layers and trainable parameters, thus is better equipped to learn the complicated associations between input and output image at the whole image scale.

Given imaging segmentation and synthesis share the common interest on the pixel level details, the satisfying performance of U-Net in segmentation makes it a potential approach for the synthetic imaging research. There is an initial attempt in this direction. For instance, a 27-layer sCT-DCNN (Han, 2017a) borrowing the ‘copy and crop’ idea from U-Net is implemented to generate virtual CT images from MRI images of same subjects. Significantly improved synthetic results are achieved compared with the traditional atlas-based method. It is worth mentioning that in this research, 128×128×160 CT images are rendered from 256×256×256 MR images. The lowered synthetic imaging resolution makes the max pooling a viable approach. In the max pooling, each grid (e.g., a group of 4 neighboring pixels) is represented by a single value (maximum value) in its subsequent feature map. This maximization operation may keep the pixel-level specifics to some extent. In the application where the input images and output images are of similar resolutions, the performance of approach in (Han, 2017a) may not be guaranteed.

In this research, we propose a new deep CNN, named Residual Inception Encoder-Decoder Net (RIED-Net). Noting the max pooling generates one pixel (the max) from neighborhood pixels (e.g., 4) during the encoding process, we introduce convolutional layers to learn the “optimal” contributions from each pixel within the neighborhood in generating the next layer pixel. Similarly, during the decoding process, respective deconvolution layers are added to learn the “optimal” weights aligned from the pixel into the pixel neighbors in the next layers. By doing so, pixel-level information is preserved precisely by the learnable filters. While the convolution and deconvolution add the value in synthetic imaging, the added layers may lead to the issues of gradient vanishing or degradation, which is long being criticized from very deep networks (K He et al., 2016; Kaiming He, 2015; Srivastava, Greff, & Schmidhuber, 2015). Res-Net has been proposed to show promising results in building deep CNNs to avoid the aforementioned problems by its short-cut connection (K He et al., 2016). Motivated by this, a residual inception block is introduced to our deep network resulting RIED-Net. Two separate datasets are used to evaluate our proposed method, which include a CEDM dataset acquired from tertiary medical center (Mayo Clinic Arizona), and a public dataset from Alzheimer's Disease Neuroimaging Initiative (ADNI). We compare our proposed RIED-Net against two benchmark methods: shallow-CNN (Gao et al., 2018) and sCT-DCNN (Han, 2017a). Three metrics from the literature are adopted for the comparison: Structural Similarity Index (SSIM), Mean Absolute Error (MAE) and Peak Signal-To-Noise Ratio (PSNR). Experimental results show that RIED-Net outperforms the two competitors on both datasets.

3.2. Background

3.2.1. U-Net and Dense Prediction Problem

CNNs have been successfully implemented to tackle different machine learning and computer vision problems. Improved performance has been achieved in imaging classification and object detection tasks (Kaiming He et al., 2014; Simonyan & Zisserman, 2014; Szegedy et al., 2014). Researchers further extend this success to imaging segmentation, a dense prediction problem, and U-Net (Ronneberger et al., 2015) is a representative model. U-Net and its variants have been applied to various segmentation problems such as joint craniomaxillofacial bone segmentation and landmark digitization(Shen et al., 2018), volumetric imaging segmentation for Xenopus kidney (Liu et al., 2016) and segment prostate from 3D volumetric images (Milletari et al., 2016). Most recently, U-Net is introduced to the synthetic imaging (Han, 2017a). One example is sCT-DCNN (see Figure 12). It consists of an encoding path (left side) and a decoding path (right side), the contracting path (represented with black arrow from left pointing right) is added to transfer additional input features from encoding layer to corresponding decoding layers by copying and pasting the entire feature maps. During the encoding and decoding process, max pooling and unpooling are applied. Max pooling is a common approach to reduce the spatial resolution and increase the receptive fields in the CNN models. During the max pooling operation, the input representation's dimensionality is reduced by replacing each $n \times n$ matrix (n is the pooling size) with one single value (e.g., maximum value) in the output representation maps. After several iterations of pooling operations, the high dimensional input image is represented by a set of feature maps of reduced spatial resolution. Taking Figure 12 as an example, after the 4th max poling layer (the 4th red box from left), the original image (256×256×1) is compressed to a 16×16×512 feature maps, each pixel within the feature map representing a region of 16×16 (256/16) within the input image.

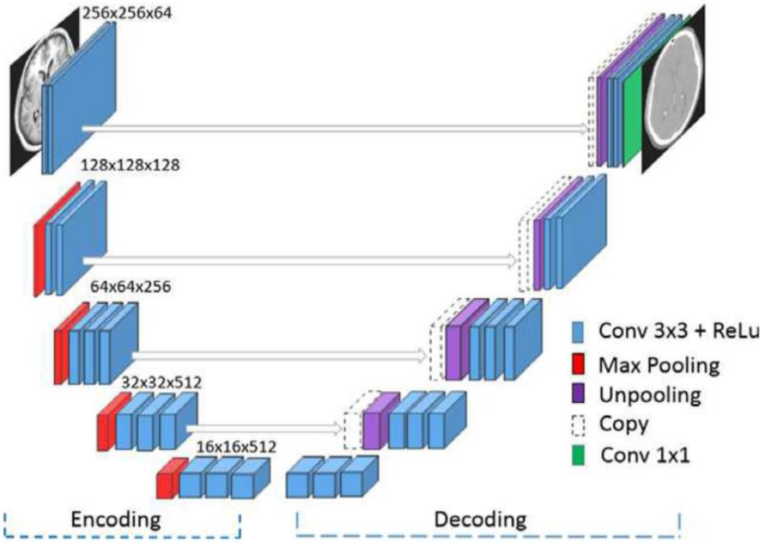


Figure 12 Architecture of sCT DCNN proposed in (Han, 2017a) (Each blue box represents a (3×3) convolutional operation (with a rectified linear unit (ReLu) as the activation function). Each red box denotes a max-pooling operation, and each purple box denotes an unpooling operation. Each white box denotes a copying layer. The sizes (width×height×number of channels) of the feature map (blue boxes) at each level are provided at the top of blue box in each level. The green box denotes the final 1×1 convolution operation that generates the output sCT prediction).

Max pooling may be desirable for imaging classification and detection problems where the outcome is a prediction on the interested object as a whole. As we discussed earlier, dense prediction problem differs as it requires preserving the pixel-level details (Chen et al., 2017). As a result, max pooling used in a dense prediction problem may face the challenges of losing pixel information. Recognizing this problem, fully convolutional networks (FCNs) (Long, Shelhamer, & Darrell, 2015) is proposed to enrage the feature maps through bilinear interpolation, and in (Noh, Hong, & Han, 2015) , unpooling layer is introduced. Specifically, when doing the max pooling operation within a grid, the location of pixel with maximum intensity is recorded. In the corresponding unpooling layer, output feature map is enlarged from the input map, the recorded position within output feature map is filled with corresponding values from input map, and the rest positions are placed with zeros (zero padding). As pointed out by (Chen et al., 2017; Liu et al., 2016), unpooling suffers from the loss of information due to the excessive use of dimension reduction and zero paddings. We want to highlight another potential issue of the max pooling and unpooling approach. That is, the max pooling operation keeps the location of the pixels with maximum contrast compared with its neighbors and the position the pixel back to the same location in the corresponding unpooling operation. The underlying assumption is that the pixels from the input image with high contrast remain at the same positions throughout different layered feature maps thus the output image. This may not be true in image synthesis where the input image and output image are from two different modalities, same region in location from two images may show different appearances (Morris, 2016). One possible solution is the use of convolutional and deconvolutional layers with the learnable filters to better record the compression information during the encoding process and de-compression information during the decoding process. This is reviewed in the next subsection.

3.2.2. Convolutional and Deconvolutional Layers

The convolutional layer is the core building block of a CNN. A set of learnable filters are included in the convolutional layer to compute the convolved value as the filters slide through all the pixels. Often, the filters slide a single pixel per step (stride = 1) to keep spatial resolution of input and output feature maps the same (Krizhevsky et al., 2012; Szegedy et al., 2014). By setting different strides, the filter can jump several pixels and thus result an output feature map of reduced spatial resolution such as the networks proposed in (Milletari et al., 2016; Sermanet et al., 2013). In parallel, as in (Noh et al., 2015), the deconvolution layer associates one single input with multiple outputs and is used as the reserve operation of convolution layer to enlarge and densify the outputs.

For illustration purpose, an example of convolution and deconvolution is shown in Figure 13. As seen, using convolution operation, the value of each pixel (e.g. C4’) in the output map equals to the convolve result of its corresponding area (C) in input map and a learnable filter (W). As a result, the value of each pixel in the output map is a weighted summation of all corresponding input pixels. In deconvolution operation, the values of an output region equal to the pairwise multiplication of its corresponding pixel (D3’) in input map with the filter (W’). By learning the optimal filters (W and W’) in training the network model, the pixel-information shall be better preserved in encoding and decoding process.

However, as the network is getting deeper with added convolution and deconvolution layers, potential issues such as gradient vanishing or degradation may emerges. We will review short-cut idea from Res-Net to address these problems in the next section.

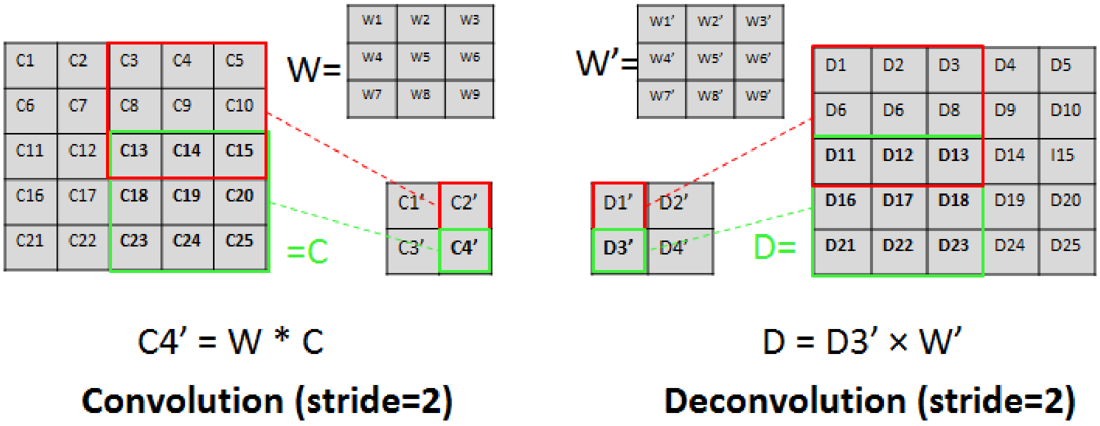


Figure 13 Illustration of convolution and deconvolution operations.

3.2.3. Residual Inception Short-cut Block

Deep networks integrate features from multiple levels and classifiers in an end-to-end multilayer fashion, and the levels of features are enriched by the number of stacked layers (depth). The stacked convolutional layers (e.g. 56-layer) tend to underperform its shallower counterparts (e.g. 20-layer) due to the gradient vanishing/exploding issue, as millions of parameters in deep networks are updated based on a single value of error gradient (K He et al., 2016). The error gradient, calculated based on the prediction result from the last layer and the earlier layers tends to be less sensitive to the error gradient, as the it gets smaller and less accurate when being referred backwards through the layers (Chollet, 2017; Szegedy, Ioffe, Vanhoucke, & Alemi, 2016a).

One interesting idea to preserve gradient over a deep network is residual shortcut connection (K He et al., 2016). Let *x* be the input image/feature map, in residual shortcut block, let *H(x)* denotes the desired non-linear mapping between the input and output of the residual block, instead of directly estimating *H*, the residual mapping *F(x)* is estimated by the learnable filters within the 2 convolutional layers, and the original mapping can be recast into *F(x)+ x*. Different experiments have been conducted to justify the advantages of this residual mapping in imaging classification problems (Chen et al., 2017; Xie, Girshick, Dollár, Tu, & He, 2016). Residual shortcut design also achieves extended success in segmentation (Fakhry, Zeng, & Ji, 2016; Milletari et al., 2016). But, the shortcut design requires the input (*x*) and output (*H(x)*) are of the same size for pixel-wise summation. The strict requirement limits its applications, especially for architecture such as U-Net and sCT-DCNN. In looking into deep learning models handling input and output with varied sizes, another interesting idea emerges, that is, inception (Szegedy et al., 2014, 2016a). Indeed, the primary goal of inception is to reduce the computational burden and improve classification accuracy. In the inception structure, different inception paths (the number of convolutional layers is different in each path) spread out from the same input, and then combined together to approximate a sparse CNN with normal dense construction. In its variants (Chollet, 2017; Szegedy et al., 2016a), the inception layer is combined with the residual short cut to improve performance. Realizing the potential from Inception in handling input-output with varied sizes, we propose to incorporate Inception to the residual shortcut model for imaging synthesis in this research.

In summary, U-Net and its variants have shown great performances in imaging segmentation and synthesis and thus having been the dominating network models. However, we argue the max pooling and unpooling layers leave it the risk of losing pixel information and impaired prediction accuracy. Additional layers such as convolutional and deconvolution layers to preserve the pixel specifics are needed. As the network gets deeper, it comes with challenges such as decreased accuracy because of gradient vanishing or degradation. A generalized shortcut model is of necessary. In this research, we propose an integrated deep model termed Residual Inception Encoder-Decoder Neural Network (RIED-Net) to serve the purpose.

3.3. RIED-Net

In our proposed Residual Inception Encoder-Decoder Neural Network (RIED-Net), the ‘copy and crop’ idea and a symmetric expanding path are added to capture the context features. Convolutional layers and deconvolutional layers are created as learnable filters so the pixel information can be traced in both the encoding and decoding procedure. In addition, the inception residual block is proposed to address issues raised from networks getting deeper.

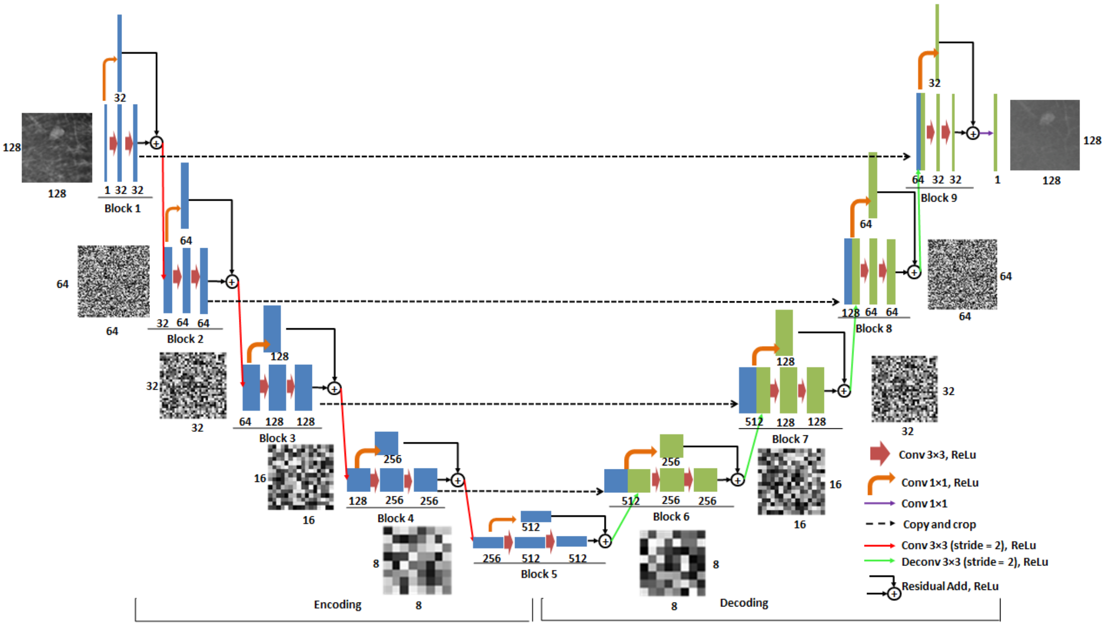


Figure 14 Architecture of RIED-Net.

As seen in Figure 14, each brown arrow represents a 3×3 convolutional operation (with a rectified linear unit (ReLu) as the activation function (Conv 3×3, ReLu)). Each red arrow denotes a 3×3 convolutional operation (stride = 2, with a ReLu as the activation function), each orange arrow denotes a 1×1 convolutional operation (with a ReLu as the activation function) and each green arrow denotes a 3×3 deconvolution operation (stride = 2, with a ReLu as the activation function). Each black dotted arrow denotes a copying operation. The final purple arrow denotes the final 1×1 convolution operation that generates the output of synthetic images. The depth (number of channels) of the feature map from each convolution layer is provided at the bottom of each blue box. Examples of feature maps from different levels are also displayed. There are 9 residual inception blocks (block 1- 9) in RIED-Net. The residual inception blocks take a new architecture (see Figure 15). It consists of one traditional convolutional path with two 3×3 convolutional layers as sCT-DCNN or U-Net, and a unique residual inception short-cut path with a 1×1 convolutional layer. The 1×1 convolutional layer is implemented to increase (during encoding) or decrease (during decoding) the filter depth and project the input feature map into the same space as output to ensure the pixel-wise summation. In the traditional 2-layer convolution block, given the input image/feature map x, assume the desired mapping fitted by stacked nonlinear layers fitting is *H(x)*. After introducing a residual inception shortcut with one single convolution layer, *H(x)* can be estimated as *F(x)* + *G(x)* in the proposed residual inception block. In this way, *H(x)* is estimated simultaneously using features from 2 different levels, which will improve the accuracy as more features are introduced (Szegedy et al., 2014). Besides, *G(x)* can be regarded as a projection/estimation of *x* (Chollet, 2017), following the same hypothesis that has been proven in (K He et al., 2016; Szegedy, Ioffe, Vanhoucke, & Alemi, 2016b), the residual mapping *F(x)* and projecting mapping *G(x)* are much easier to optimize and resulting in a more accurate results than the original mapping *H(x)*. Compared with traditional residual shortcut block, our proposed residual inception block deals with the problem that input feature map has different channel from the output feature maps. It is simpler and easier to deploy than other state-of-art residual inception designs.

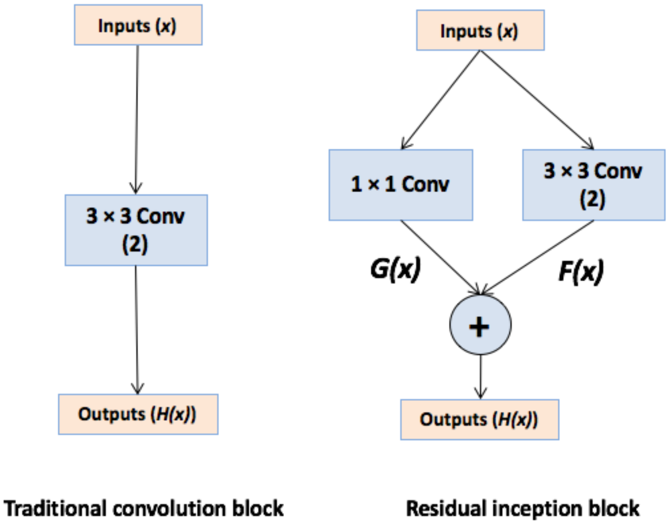


Figure 15 Schema for original convolution block and proposed residual inception (Note that in traditional convolution block, the input *x* and output *H(x)* has different number of channels which makes the directly residual shortcut inapplicable).

3.4. Experimental Validation

In this section, we conduct two experiments to validate the performance of RIED-Net using digital mammography dataset from Mayo Clinic and a public neuroimaging dataset from Alzheimer's disease Neuroimaging Initiative (ADNI)(Weiner et al., 2016)

3.4.1. Evaluation Metrics

Given the ground truth image $I_1^{m\times n}$ and its synthetic image $I_2^{m\times n}$ produced by a model, three commonly used metrics from literature (Chen et al., 2017; Han, 2017a) that quantify the similarity between the ground truth image and synthetic image are employed to evaluate the synthesis performance. These three metrics, i.e., mean absolute error (MAE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR).

3.4.1.1. Mean absolute error

Mean absolute error (MAE) is to measure pixel-wise intensity absolute difference between ground truth image and predicted image. It is also widely used as cost function for various regression models. The MAE between image I_1 and image I_2 is calculated through the following formula,

$$MAE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}|I_1(i,j) - I_2(i,j)| \tag{3.1}$$

with

$\boldsymbol{I_1(i,j)}$ the intensity value at position *(i, j)* of image *I₁*;

$\boldsymbol{I_2(i,j)}$ the intensity value at position *(i, j)* of image *I₂*;

m/n the width/height of image I_1 and I_2.

3.4.1.2. Structural similarity index

SSIM (Wang, Bovik, Sheikh, & Simoncelli, 2004) is a metric used for measuring the similarity between two images (ground truth image and predicted image). It compares the local patterns of pixels' intensity that have been normalized for luminance and contrast. A higher value means the higher similarity of the reconstruction. The SSIM between image I1 and image I2 can be calculated through the following formula:

$$SSIM = \frac{(2\mu_{I_1}\mu_{I_2}+c_1)(2\sigma_{I_1I_2}+c_2)}{(\mu_{I_1}^2+\mu_{I_2}^2+c_1)(\sigma_{I_1}^2+\sigma_{I_2}^2+c_2)} \tag{3.2}$$

with

μ_{I_1}: the average intensity of image I_1;

μ_{I_2}: the average intensity of image *I2*;

$\sigma_{I_1}^2$: the intensity variance of image *I1*;

$\sigma_{I_2}^2$: the intensity variance of image *I2*;

$\sigma_{I_1 I_2}$: the covariance between all intensity values in image *I1* and image *I2*;

$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$: two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel-value (typically this is $2^{\#\text{bits per pixel}} - 1$, in experiment I and II, we set L equals to 4095 and 255 respectively); $\mathrm{k_1}$=0.01, $\mathrm{k_2}$=0.03 are default values.

3.4.1.3. Peak signal-to-noise ratio

PSNR is a metric to assess image quality and distortion. It is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. A higher value usually indicates a better reconstruction. The PSNR between image I1 and image I2 is be calculated through the following formula:

$$PSNR = 20\, log_{10}(MAX_I) - 10\, log_{10}(MSE) \tag{3.3}$$

where

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_1(i,j) - I_2(i,j)]^2 \tag{3.4}$$

MAX_I: the maximum possible pixel value of image I_1 and image I_2.

We conduct two sets of experiments to compare against the methods from the literature in terms of these three metrics. The first experiment is on breast cancer diagnosis and the second is on Alzheimer Disease Staging.

3.4.2. Experiment I: Case Study on Breast Cancer

Breast cancer is a worldwide leading type of cancer in women accounting for 25% of all cancer cases. In 2012, it resulted in 1.68 million new cases and over 0.52 million deaths. According to the U.S. Breast Cancer Statistics 2018, about 1 in 8 U.S. women (~12%) will develop invasive breast cancer over the course of her lifetime. Full field digital mammography (FFDM) is the only clinically acceptable imaging modality for the population-based breast cancer screening among existing imaging modalities (Lehrer et al., 2012). However, using FFDM is not an optimal approach in breast cancer screening due to its relatively low detection sensitivity in many subgroups of women (Elmore et al., 2005). Using dynamic contrast enhanced breast MRI may yield significantly higher cancer detection sensitivity, but its substantially higher cost, lower accessibility and longer imaging scanning time forbids breast MRI being used as a primary imaging modality in breast cancer screening and detection (Warner et al., 2004). In addition, lower image resolution of breast MRI is a disadvantage as comparing to FFDM.

To combine the advantages of both FFDM and MRI, a new novel imaging modality namely, contrast-enhanced digital mammography (CEDM) emerges which uses the intra-venous injection of an iodinated contrast agent in conjunction with a mammography examination. CEDM includes low energy (LE) imaging, which is comparable to routine FFDM (Francescone et al., 2014) and recombined imaging similar to breast MRI.

Comparing to breast MRI, CEDM exam is about 4 times faster with only about 1/6 the cost (Patel et al., 2017), and has 10 times the spatial resolution of breast MRI. Several studies including prospective clinical trials conducted at Mayo Clinic have indicated that CEDM is a promising imaging modality that overcomes tissue overlapping ("masking") occurred in FFDM, provides tumor neovascularity related functional information similar to MRI, while maintaining high image resolution of FFDM (Cheung et al., 2014; Fallenberg et al., 2014; Gillman et al., 2014; Luczyńska et al., 2014).

In this experiment, we evaluate the performance of REID-Net in mapping the LE images to the recombined images. Its performance is compared with other two state-of-art methods in medical image synthesis.

3.4.2.1. Dataset

Table 7 Detailed imaging features for the CEDM dataset.

Feature Name	Value (Pixels)
Width	2560
Height	3328
Intensity Range	0-4095

Based on Institutional Review Board (IRB) approved study and data collection protocol, we review CEDM examinations performed using the Hologic Imaging system (Bedford, MA, USA) between August 2014 and December 2015. A total of 139 subjects are collected. In CEDM dataset for each subject, there are cranial-caudal (CC) and mediolateral-oblique (MLO) views for both LE and recombined images. Examples for the images are shown in Figure 16. More details about the images can be found in Table 7. Among the dataset, 112 (80%) subjects are randomly selected as training dataset, the rest

27 (20%) subjects are used for blind test. For each subject, CC view and MLO view images are treated as two separate training images, which results in a dataset of 224 (112×2) training images and 54 testing images (27×2).

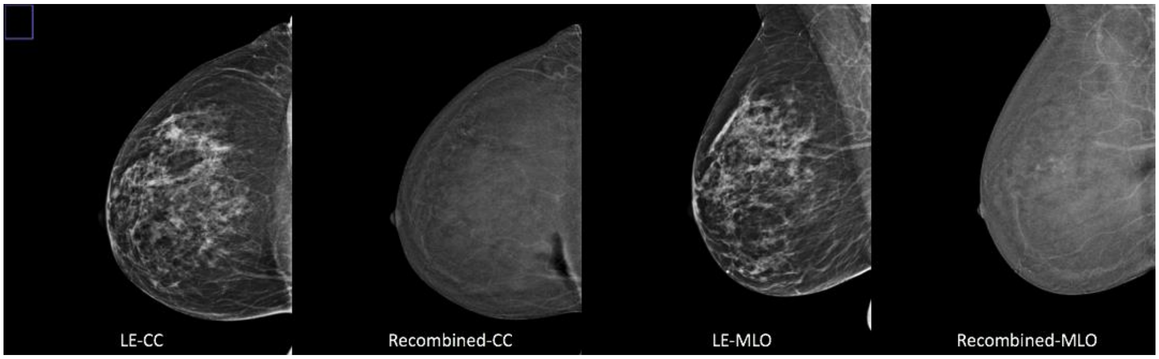


Figure 16 Examples of images in CEDM dataset.

3.4.2.2. Image processing

It is a common approach to extract patches from images as training samples to address the shortage of training dataset (Gao et al., 2018; Han, 2017a). However, the size of patches varies case by case. Larger patches require more memory for calculation, while small patches allow the network to see only little context. In the experiment, we want to make the training patches as large as possible in the range of GPU memory, and the largest patches we afford is 128×128, in alignment with dataset size. After patches size is set, training samples are extracted from the images in the step size of 8 in each dimension, and patches outside the breast boundary are excluded. As a result, a dataset of 65800 patches are obtained from the 112 training subjects. Among these 65800 patches, 59220 (90%) are used as training samples, and the rest 6580 (10%) are used as validation samples to tune the parameters. An ‘optimal’ parameter setting is decided based on the best validation result. Specifically, the overall architecture is implemented with programming language Python, and libraries including Keras and Tensorflow. Mean absolute error (MSE) is used as loss function and Adam (Kingma & Ba, 2014) is used as the optimizer. Learning rate is set to be 0.002 with learning rate decay equals to 0.005. Training batch size is set to be 64 and training iteration is set to be 80. We use the default settings of Keras for all the other parameters. For the two comparing models, the optimal parameters reported in the proposing articles are used.

3.4.2.3. Experimental Results and Comparison

The comparison of performance for different models is conducted on the reserved testing dataset of 54 images (27 subjects). For each image, we slide the 128×128 window from left to right, top to bottom (step size = 2) in LE image, to get the input patches. The input patches are fed into the trained model, from which we get the predicted virtual recombined image patches (128×128) as outputs. The output patches are placed at the same position as their corresponding input patches in the "virtual" recombined images. For the position with overlapping pixels, the values are replaced with mean value for all overlapping pixels. At last, the "virtual" recombined images are rendered. Our ultimate goal is synthesizing the whole image, so it is more desirable to evaluate metrics based on the predicted complete image and its corresponding ground truth image instead of individual testing patch. As a result, to quantify the synthesis performance for our proposed model, a set of 54 synthetic recombined images are generated for each LE image in the testing dataset with the trained model. Each individual synthetic recombined image is then compared with its corresponding ground truth image, and 3 evaluation metrics (MAE, SSIM, PSNR) introduced in section 4.4.1 are calculated to measure the similarity between the synthetic image and ground truth image.

In terms of each evaluation metric, the mean value and standard deviation across the 54 pairs of synthetic-ground truth image are reported in

Table 8, where the results of the 2 state-of-the-art models (Shallow CNN (Gao et al., 2018) and sCT-DCNN (Han, 2017a)) implemented exactly the same procedure are added for comparison. In order to further explore the robustness of the results, t-tests are performed for each metric between any pair of the three methods. The details are shown in Table 9.

Table 8 Performance of different models on the CEDM testing dataset.

Method	**MAE**	**SSIM**	**PSNR**
Shallow CNN (Gao et al., 2018)	219.753(±21.563)	0.793(±0.023)	29.224(±1.462)
sCT-DCNN (Han, 2017a)	11.502 (±2.187)	0.958(±0.013)	43.346 (±1.462)
RIED-Net	**11.277 (±2.112)**	**0.962(±0.012)**	**43.450(±1.423)**

Table 9 P-values of t-tests on pairwise comparison: (a) MAE, (b) SSIM, (c) PSNR.

(a)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.003
RIED-Net	<0.001	0.003	-

(b)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.015
RIED-Net	<0.001	0.015	-

(c)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.004
RIED-Net	<0.001	0.004	-

From Table 8 and Table 9, we have two conclusions. First, shallow CNN significantly underperforms both sCT-DCNN and our proposed RIED-Net on all three metrics. This confirms our argument that deep models with more trainable parameters may outperform shallow network in imaging synthesis problem. Comparing to sCT-DCNN, RIED-Net shows the marginal performance advantages (11.277 vs. 11.502 in MAE, 0.962 vs. 0.958 in SSIM, 43.450 vs. 43.346 in PSNR). RIED-Net has small standard deviation indicating it is a deep model with robust performance. To justify the marginal outperformance, we delve in details on the case by case bases. As seen in Figure 17, among all the 54 images, our proposed RIED-Net has higher SSIMs (the higher the better) on 38 images (70.4%), higher PSNRs (the higher the better) on 36 images (66.7%), smaller MAE (the smaller the better) on 39 images (72.2%). In looking at all three metrics together, RIED-Net outperforms sCT-DCNN on 36 cases (>66.7%).

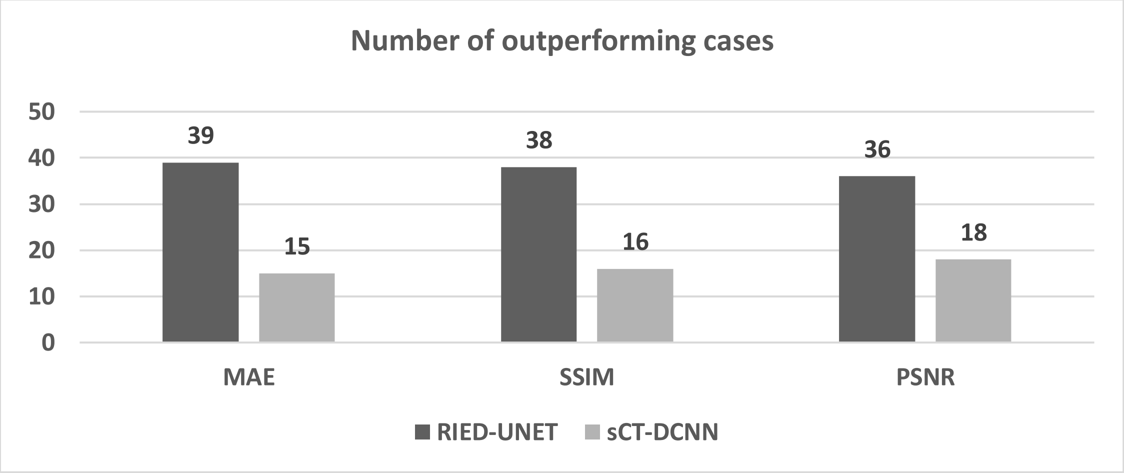


Figure 17 Distribution of outperforming cases for MAE, SSIM, PSNR on CEDM testing dataset.

For illustration purpose, we include one image from each model (see Figure 18). Figure 18A is ground truth recombined image. Figure 18B, C and D are predicted images of Shallow CNN, sCT-DCNN and RIED-Net respectively. The error maps of each output image are shown in Figure 19. Within the error map, the value of a pixel is the absolute value of difference between the intensities of two pixels at the same location in ground truth image and synthetic image. Each value is then divided by the same normalizer (normalizer value =15). The values greater than 1 are assigned with 1s. The aim of this procedure is to normalize the range of difference map into between 0 and 1, while excluding the effects of outlier pixels.

First, as expected, limited by the learning capability, there is a very significant gap between the output of the 4-layer shallow CNN and ground truth image (high MAE values). We can focus on the comparison between sCT-DCNN and our proposed model. Comparing output images C and D in Figure 18, we can observe that Figure 18C is coarser within the breast region, especially in the region close to boundary, while in Figure 18D, these regions are sharper and clearer. This is because, in these regions, the dense tissue is interlaced with other parts such vessels or fat, the differences among pixels from different parts are large. sCT-DCNN with max pooling loses the pixel information and the unpooling layers fail to restore such information, as a result, these pixels cannot be differentiated well and tend to give the similar predictions. The advantages of RIED-Net in this scenario clearly show. In looking at the error maps in Figure 19 (B and C), the red bounding boxes in Figure 19B has larger high-error regions comparing to Figure 19C. This may be because in sCT-DCNN, during the prediction, if a single pixel is estimated with high error, it will first affect its 3 neighboring pixels after unpooling layers, and this effect tends to expand to more pixels after more unpooling layers. In RIED-Net, the succeeding pixels after deconvolutional layers depend not only on that specific preceding pixel, but also the trainable parameters within the deconvolutional layers. In this way, even if a pixel is estimated with high error, the resu1lts of its following neighboring pixels can be relieved through the deconvolutional layer, thus the region of high-error in Figure 19C tends to be small and isolated regions. We conclude RIED-Net has satisfying performance on this imaging synthesis problem for breast cancer research on Digital Mammography (DM) Modality. Next, we will explore its applicability to an Alzheimer disease dataset across two imaging modalities: PET and MRI.

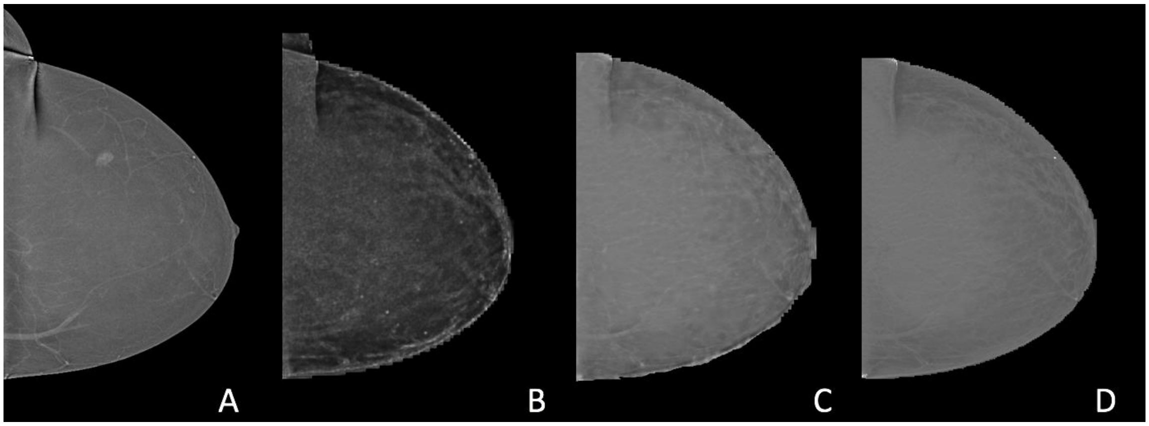


Figure 18 Sample of one ground truth recombined image (A), output 'virtual' re-combined images of Shallow-CNN (B), SCT-DCNN(C), our proposed model (D).

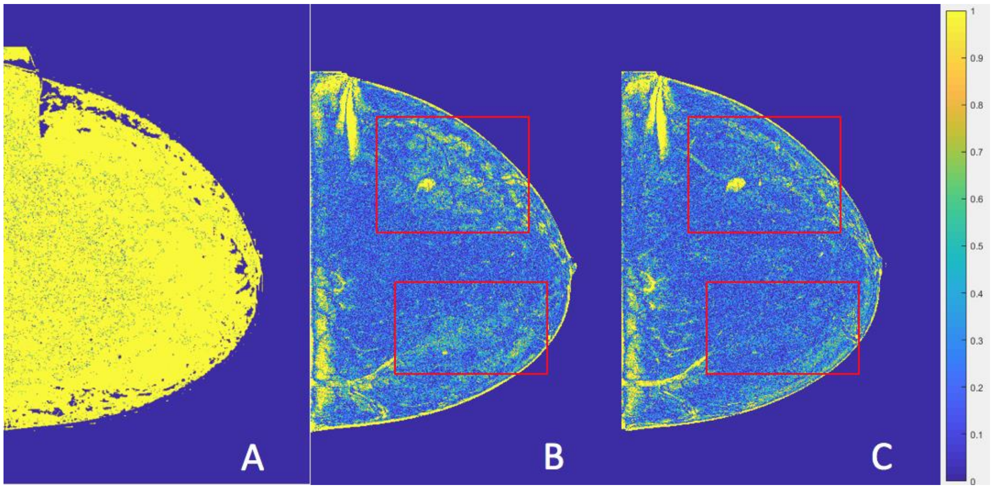


Figure 19 Error maps of output images for Shallow-CNN (A), sCT-DCNN(B), our proposed model (C).

3.4.3. Experiment II: Case Study on Alzheimer Disease

Alzheimer’s disease (AD) is a progressively neurodegenerative disease which is the most frequent type among elderly dementia patients. In the U.S., approximately 5.2 million people over 60 are afflicted by AD (Alzheimer’s Association, 2008). This drives a great amount of research investigating ways to slow down the AD progression and detect AD at early stage for better treatment or even prevent the disease. Mild cognitive impairment (MCI) is a syndrome defined as cognitive decline greater than expected for individuals during the course of aging but that does not interfere notably with activities of daily life (Gauthier et al., 2006). It is an intermediate stage between normal aging with mild cognitive decline and dementia where cognitive impairment is more severe even impacting daily function. Though it is distinct from dementia, MCI patients with memory complaints and deficits (amnestic mild cognitive impairment) have high risks of progression to AD (Castro & Smith, 2015; Gauthier et al., 2006). The early diagnosis of MCI stage is becoming essential when the interventional strategies may be more effective.

For the early diagnosis and prognosis of AD, the use of imaging has been highlighted by multiple expert consensus groups nationally and internationally, such as the working group convened by National Institute of Aging (NIA) and the Alzheimer's Association (AA) (Carrillo et al., 2013) and the International Working Group (Dubois et al., 2014). It has been widely-recognized that imaging of different modalities, including but not limited to structural MRI, FDG-PET, and amyloid-PET, play important and often complementary roles. However, it is difficult for a single modality to serve all the purposes as each modality has unique strength and weakness. Combining different imaging modalities is vitally important to make accurate and early diagnosis and prognosis, a prerequisite to develop effective disease-modifying therapies. But, patients may not have all imaging modalities available due to various reasons. In this experiment, the proposed architecture is to learn the non-linear mapping between PET images and MRI images. It will be trained to render 'virtual' PET images given MRI images as input. Its performance is compared with the same two methods mentioned in experiment I.

3.4.3.1. Dataset

The ADNI is launched aiming at finding the relationship between progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) and biomarkers, MRI, PET or clinical and neuropsychological assessments. ADNI enrolls a large cohort (>800) of participants (Weiner et al., 2016), for each subject, PET, MRI images, as well as clinical information are available. In this experiment, 14 subjects are downloaded and used in the experiment. The size of raw MRI images is 256×256×170, while the PET images are of the size 128×128×90. Among these 14 subjects, 10 subjects are used as training data; the rest 4 subjects are kept as dataset for blind testing. Other detailed information for ADNI dataset is shown in Table 10. Three sample images from different slices are shown in Figure 20.

Table 10 Detailed imaging features for the ADNI dataset.

Modality	**Feature Name**	**Before Co-registration** Value (Pixels)	**After Co-registration** Value (Pixels)
MRI	Width	256	79
	Height	256	79
	Slice	170	91
	Intensity Range	0-255	0-255
PET	Width	128	79
	Height	128	79
	Slice	90	91
	Intensity Range	0-255	0-255

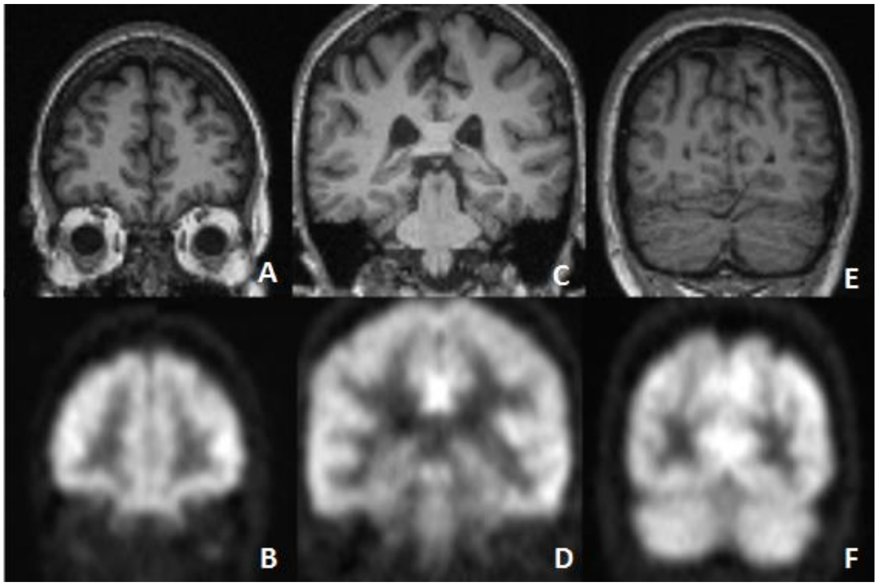


Figure 20 Examples of MRI images (A/C/E) in ADNI dataset and their corresponding PET images (B/D/F) (The images are extracted from different slices).

3.4.3.2. Image processing

The MRI and PET images are firstly spatially normalized into a same template space to make them rigidly aligned with each other. This process is known as image co-registration, which is conducted through a Matlab based library named Statistical Parametric Mapping (SPM 12 https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). After co-registration, the size of PET and MRI become 79×79×91 (limited by the resolution of atlas used in SPM 12). In this experiment, we set the input and output patches to be 64×64. Training samples are extracted from each slice of the 3D image of each subject, in order to exclude slices with poor quality and limited region of brain, slice 1-10 and slice 82-91 are excluded. Patches are extracted at step size of 4 in each dimension. As a result, a dataset of 34790 (10 × (91-20) × 7 × 7) patches are obtained from the 10 training subjects. The parameter settings for the 3 models are the same as experiment I.

3.4.3.3. Experimental Results and Comparison

The comparison of performance for different models is conducted on the reserved validation dataset of 284 images (4×71). In this experiment, the input and output patch sizes are set to be 64×64; all the other settings and procedure are the same as experiment I. The performances of 3 different models are reported in Table 11 and pair t-test results are summarized in Table 12.

Table 11 Performance of different models on the ANDI testing dataset.

Method	**MAE**	**SSIM**	**PSNR**
Shallow CNN (Gao et al., 2018)	24.018 (±3.051)	0.860 (±0.031)	17.852 (±0.886)
sCT-DCNN (Han, 2017a)	14.466 (±3.452)	0.945 (±0.022)	22.087 (±2.039)
Proposed RIED-Net	**13.412 (±3.278)**	**0.957 (±0.019)**	**22.813 (±2.125)**

Table 12 P-values of t-tests on pairwise comparison: (a) MAE, (b) SSIM, (c) PSNR.

(a)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.013
RIED-Net	<0.001	0.013	-

(b)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.004
RIED-Net	<0.001	0.004	-

(c)	**Shallow CNN**	**sCT-DCNN**	**RIED-Net**
Shallow CNN	-	<0.001	<0.001
sCT-DCNN	<0.001	-	0.008
RIED-Net	<0.001	0.008	-

Similar to the first experiment, from Table 11 and Table 12., we conclude shallow CNN underperforms sCT-DCNN and RIED-Net and RIED-Net significantly outperforms sCT-DCNN in terms of all the three metrics. Again, we compare the two deep network models on the case by case bases (Figure 21). Among the 284 test cases, RIED-Net has higher SSIMs and PSNR on 226 (79.6%) and 229 (80.6%) testing images respectively, lower MAEs on 232 images (81.7%).

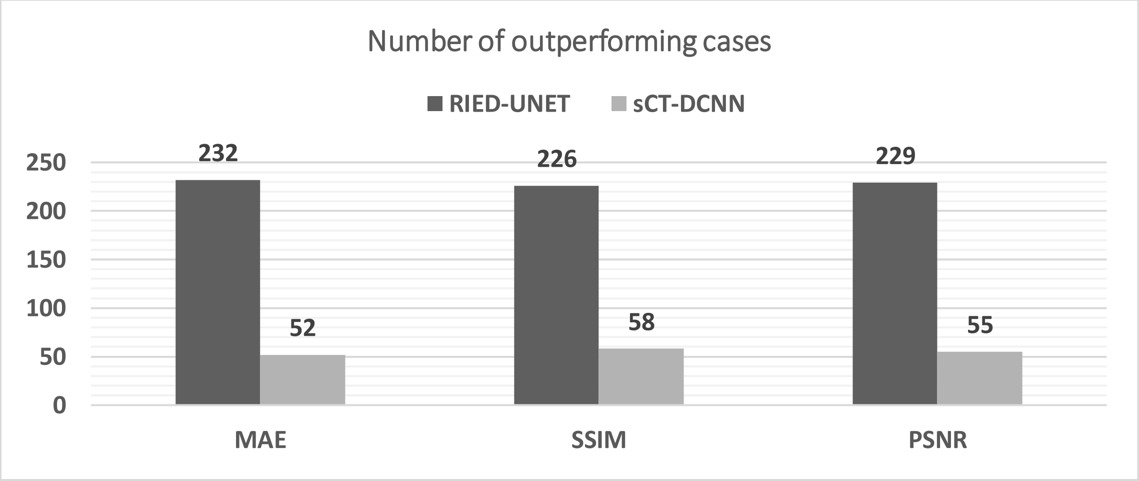


Figure 21 Distribution of outperforming cases for MAE, SSIM, PSNR on ADNI testing dataset.

Figure 22 is the illustrative figure showing one image from each of the three models with Figure 22A is ground truth PET image, B, C and D are output images of shallow CNN, sCT-DCNN and RIED-Net. In Figure 23, A, B and C are the error maps for the outputs from 3 models. The error maps are generated through the same procedure as experiment I.

As seen in Figure 22, the output of shallow CNN (Figure 22 B) roughly restores the layout of ground true PET image (Figure 22 A) while with significant errors in details (Figure 23A). In the error maps in Figure 23, we can locate several regions where sCT-

DCNN have higher errors in prediction, for examples, the two regions highlighted with red bounding box in Figure 23B, while RIED-Net shows lower errors in the same locations. We conclude RIED-Net has satisfying performance on synthesizing images across modalities.

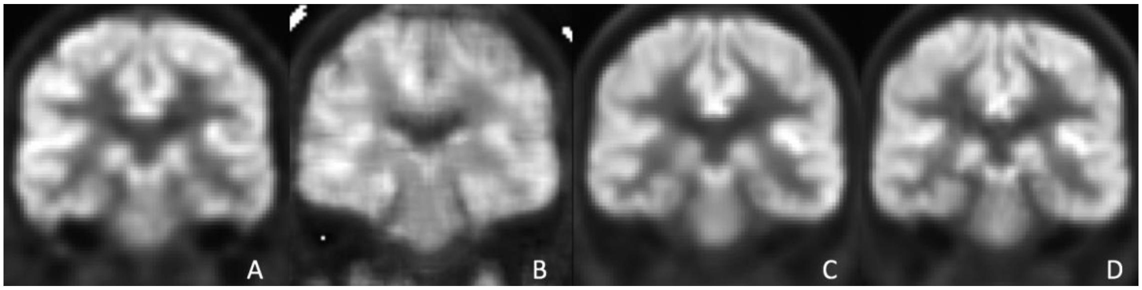


Figure 22 Sample of one ground truth PET image (A), output 'virtual' PET images of Shallow-CNN (B), sCT-DCNN(C), our proposed model (D).

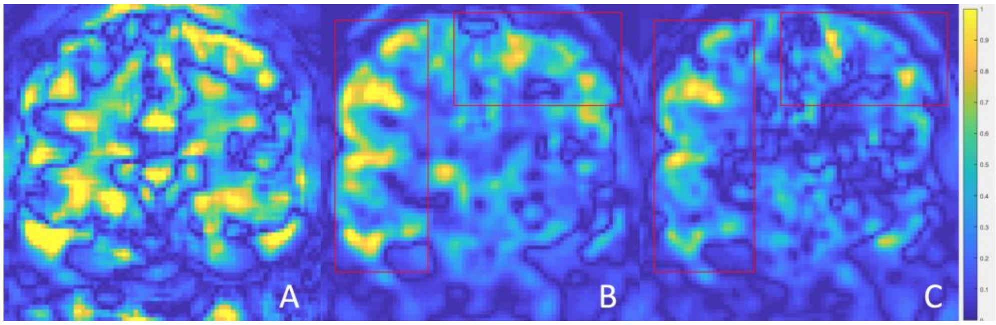


Figure 23 Error maps of output images for Shallow-CNN (A), sCT-DCNN (B), our proposed model (C).

3.5. Discussion and Conclusion

Image synthesis is becoming an important field in medical images research, especially for the scenario where some image modalities maybe missing. These days, CNNs has shown its promises in medical imaging research mostly in imaging classification, detection and segmentation. In this study, we propose a novel residual inception encoding-decoding network (RIED-Net) to tackle this image synthesis problem. There are two main contributions. First, the convolutional layers are introduced to reserve pixel information during the encoding process when the feature map size is reduced to increase receptive field size; and deconvolutional layers are implemented to restore pixel information within the decoding process. Second, residual inception shortcut block is designed to handle the gradient vanishing issues. The performance of our proposed architecture is evaluated using two datasets. Comparison experiments confirm the outperformances of the proposed network mode.

While promising, there is room for future work. For example, as we may observe in Figure 18 and Figure 19 from the breast cancer study, all the models perform poorly on small region of interest (e.g. suspicious tumor), this is because the ROI region is relatively small compared with the whole breast, and the models fail to pay more attention to such regions during the training. It is our plan to investigate strategies to this type of problems.

CHAPTER 4

A FEATURE TRANSFER ENABLED MULTI-TASK DEEP LEARNING MODEL ON MEDICAL IMAGING

4.1.Introduction

During the last decade, precision medicine - an approach that considers individual variability in diagnosis and treatment - has emerged as a novel paradigm for healthcare. One cornerstone for precision medicine is medical imaging. Tremendous resources and manpower have been directed towards research in medical imaging, and this domain of study can be broadly divided into three categories: object detection, image segmentation, and imaging-based classification. Object detection aims to derive an envelope encircling the object of interest or the center points of those objects of interest. Segmentation generates a probability map that quantifies the likelihood that each pixel/voxel is within the region of interest (e.g., tumor). Imaging-based classification primarily identifies the object of interest to be malignant or benign. Most recently, deep learning (Lecun, Bengio, & Hinton, 2015) has gained great success in performing all three tasks (Affonso Carlos, Renato, & Marques, 2015; He et al, 2016; Khatami et al., 2018; Szegedy et al., 2015).

Deep learning owes its success largely to the fact that its models are capable of learning and reproducing an extensive range of parameters from the layers. These parameters are utilized to extract features from images to achieve good performance with respect to the tasks (Litjens et al., 2017). As one of the first deep learning techniques, convolutional neural networks (CNNs) have been extensively investigated (Greenspan, Ginneken, & Summers, 2016). For object detection, CNN-based detectors are trained to find “bounding boxes” on the desired object(s). Example applications are colonic polyps in computed tomography (CT) images (Roth et al., 2016), cerebral microbleeds in Magnetic Resonance Imaging (MRI) (Dou et al., 2016), and breast and lung cancers in Ultrasound images (Lee & Chen, 2015). For segmentation, successful implementations of CNN have been reported in segmenting brain tumors (Havaei et al., 2017; Zhang et al., 2015; Zhao et al., 2018), joint craniomaxillofacial bone and landmark digitization (Zhang et al., 2018) and epithelial tissue in prostatectomy (Bulten, Hulsbergen-van de Kaa, van der Laak & Litjens, 2018), just to name a few. For classification, CNN often takes an extracted region of interest (ROI) as the input, and the outputs are different class labels on the ROIs. The first application can be traced back to 1996, when a four-layer CNN was employed to classify the ROIs into biopsy-proven masses and normal tissues from mammogram images (Sahiner et al., 1996). Since then, different CNNs have been introduced for various medical classification applications including breast lesions (Araujo et al., 2017; Huynh, Li, & Giger, 2016), lung nodules (Shen, Han, Aberle, Bui, & Hsu, 2019), skin lesions (Yap, Yolland, & Tschandl, 2018) and pulmonary peri-fissural nodules (Ciompi et al., 2015), etc. Though commendable classification results have been reported, they are limited to scenarios where manually labeled tumors (ROIs) are provided.

The research reviewed above separately focuses on each individual task, namely detection, segmentation, or classification. Recognizing the inter-dependencies of these tasks, researchers have started to explore utilizing the joint powers from multiple tasks. The first attempt is to integrate multiple serially conducted tasks together as a pipeline-based approach (Al-antari et al., 2018; Al-Masni et al., 2017). Most recently, an emerging field from CNN is developing new deep learning models to conduct tasks in parallel, a method termed Multi-Task Learning (MTL) (Ruder, 2017). The core of existing MTL models is separate deep models for each individual task, ending with one joint cost function (He et al., 2017; Redmon et al., 2016; Ren et al., 2017). We contend that the features from different tasks may benefit other tasks in the training process. Therefore, we propose a new MTL architecture, feature transfer MTL neural network (FT-MTL-Net), to utilize the features from parallel tasks.

As the initial step to validate the idea of feature transfer in MTL architecture, we explore transferring features from a segmentation task to a classification task. This is because: 1) the goal of most medical imaging applications is to accurately diagnose/stage the disease - a classification problem; 2) though segmentation and detection are both closely tied to classification, the features used in segmentation, detection, and classification differ. Specifically, classification and detection require features of low resolution for the abstracted representation (Szegedy et al., 2015; Wu, Zhong, & Liu, 2017), while segmentation needs high resolution features for the pixel/voxel-wise prediction (Badrinarayanan, Kendall, & Cipolla, 2017; Shelhamer, Long, & Darrell, 2017). Moreover, given that the segmentation task has already delineated the candidate areas through the output masks, these areas can be taken as prior knowledge to guide the feature generation procedure focusing on the candidate areas. Motivated by these two aspects, our proposed FT-MTL-Net is designed to transfer segmentation features from candidate regions to the classification task. Three contributions come out of this novel design. First, to our best knowledge, it may be one of the first fully-automatic deep learning systems in medical imaging that can be trained end-to-end through a unified cost function and solve the tasks of tumor detection, segmentation, and classification simultaneously. Second, it enables feature transfer from a segmentation task to a classification task. The features from both high resolution (transferred from segmentation) and low resolution (existing features) are adopted to help improve the classification accuracy. Third, the features transferred are re-weighted based on the prior knowledge from the segmentation probability map. As a result, information from irrelevant regions is excluded, and the feature map is representative of the tumor regions only. Such design only requires ~700 added parameters which is negligible compared to ~2M parameters from Mask-RCNN (He, Gkioxari, Dollar, & Girshick, 2017) and thus is at a comparable scale of computational complexity to existing MTL models.

We evaluate the proposed FT-MTL-Net in the Full Filed Digital Mammogram (FFDM), a publicly available dataset published in INbreast (Moreira et al., 2012). The performance is measured based on five-fold cross validation. For the classification task, FT-MTL-Net is compared with eight methods (four are manual and four are automated) using the performance metric area under curve (AUC). Experimental results indicate FT-MTL-Net outperforms all eight competing methods with an AUC of 0.92 (± 0.02). For the detection task, FT-MTL-Net outperforms four competing methods with a true positive rate of 0.91 (± 0.05) at an average of 3.67 false positives per image. For the segmentation task, FT-MTL-Net is compared with three existing methods and achieves a comparable result of average dice index of 0.76 ± 0.03.

4.2.Related Work

4.2.1 Integration of Multiple Tasks as Pipeline Systems

Recognizing the inter-dependences of detection, segmentation, and classification tasks in medical applications, researchers often develop pipeline systems to tackle each task one at a time and connect the tasks as a whole system. In such systems, automatic detection and segmentation are often the first steps before classification. For instance, Al-Masni et al. (2017) propose a regional convolutional neural network (R-CNN) for mass detection, followed by a fully-connected CNN-based classifier for “benign versus malignant” prediction. Dhungel, Carneiro, and Bradley (2017a) develop a three-step pipeline for mass detection, segmentation, and classification. In this research, raw images are fed into a CNN model for mass detection, which is refined through a random forest classifier on hand-crafted features. The refined boxes containing candidates are then segmented through a Conditional Random Fields (CRF) model (Lafferty, McCallum & Pereira, 2001) followed by an active contour model (Jorstad & Fua, 2015). A mixture model combining a CNN model and random forest is trained with bounding boxes extracted from the detection step. The classification results are further finetuned through hand-crafted features extracted from both bounding boxes and segmentation outputs from detection. ‘User intervention’ is introduced where the false positive detections are manually excluded, to get an accurate training dataset for the following segmentation and classification tasks. In the research proposed by Al-antari et al. (2018), a fully automatic system is designed for detection, segmentation, and classification - all deploying deep learning models. You-Only-Look-Once (YOLO) (Redmon, Divvala, Girshick, & Farhadi, 2016) is implemented for mass detection, followed by a Full resolution Convolutional Network (FrCN) for segmentation, and finally a traditional CNN for classification.

Although these approaches are more advanced in terms of automation/semi-automation with satisfying results on diagnosis, serial-type pipeline systems come with a set of disadvantages. First, the design and implementation of a deep learning model for each task is complicated and time-consuming; a large amount of effort and computing resources are needed for model design, training, testing, and tuning. Second, the relatively limited medical imaging dataset for training could potentially lead to overfitting (Litjens et al., 2017). To address these issues, multi-task learning (MTL) (Caruana, 1997) has emerged and shown great potential in natural language processing (Collobert & Weston, 2008), speech recognition (Deng, Hinton, & Kingsbury, 2013), and computer vision (Girshick, 2015; He et al., 2017). One advantage of MTL is saving computational resources by sharing convolutional layers (features maps) amongst separate tasks. MTL also may reduce the risk of overfitting through learning a more generalized feature map for each task (Baxter, 1997; Ruder, 2017). In addition, MTL improves learning efficiency and prediction accuracy for the task-specific models (Caruana, 1997). Different deep multi-task learning methods in medical applications are reviewed in the following section.

4.2.2 Deep Multi-Task Learning

Deep Multi-Task Learning develops deep learning architectures to conduct multiple tasks in a parallel fashion. Current deep MTL research is dominated by the direct parameters sharing approach (Ruder, 2017). The models employ “1-m-1” structure. The first “1” is a main shared deep CNN architecture (a.k.a. backbone). The “m” refers to multiple separate subnetworks (a.k.a. head architecture) for different tasks (He et al., 2017). These “m” head architectures share the feature maps from the backbone and make predictions individually. The second “1” is a cost function. During the training, the parameters from the backbone and the heads are updated simultaneously based on this single cost function in the form of a linear combination of each individual task’s cost.

Following this “1-m-1” structure, several methods have been proposed for natural image analysis (Redmon et al., 2016; Ren et al., 2017). The success of MTL on natural images is naturally extended to the medical imaging applications. For instance, in the search conducted by Akselrod-ballin et al. (2016), a faster R-CNN is introduced for detection and classification of mass regions simultaneously. In this architecture, a single ResNet model (He et al., 2016) is implemented to provide mass candidates and feature maps which are shared by the tasks of localization and classification. Samala, Chan, Hadjiiski, Helvie, and Cha (2018) take mass classification from digital mammograms and digitalized screen-film mammograms as two separate tasks and address these two tasks by a single framework based on the Visual Geometry Group (VGG) model (Noh, Hong, & Han, 2015). The study from Liu, Zhang, Adeli, and Shen (2018) focuses on neuroimaging for Alzheimer’s disease to diagnose classification and predict clinical scores. Feng, Nie, Wang, and Shen (2018) propose a multi-task residual fully convolutional network (FCN) to segment organs (e.g. bladder, prostate, and rectum) and estimate the intensities. While “1-m-1” approaches aim to handle multiple tasks from one model, the backbone needs to be carefully designed to include most if not all the features, which must be shared. Moreover, “1-m-1” models fail to consider the potential contributions from the head-features to the tasks, individually and jointly. As medical applications have unique challenges of potential overfitting due to limited training dataset size, sharing head-features may help address this issue.

When first proposed, transfer learning was interested in the problems from different data sources. Here the data source is known as the domain. Transfer learning integrates knowledge gained from source domains with the data in target domains to help overcome data shortages in the target domain. The existing transfer learning methods fall into three major categories: instance transfer, parameter transfer, and feature transfer (Pan & Yang, 2010). Instance transfer reuses data from the source domains to augment the data in the target domains. Although it is intuitive, instance transfer may be questioned for its validity when source and target domains differ greatly. Parameter transfer assumes that closely related tasks should have similar parameters in their respective models and encourages source and target domains to share some model parameters. Yet, it is challenging to appropriately utilize parameters from source domains and tune hyperparameters for the target domain. Feature transfer aims to identify a joint feature map shared by the source and target domains. Because multiple sources and target domains have shared knowledge and representations, features transferred from the source domains may enhance the generalizability of the model with reduced risk of overfitting. However, both parameter and feature transfer face the major obstacle of negative transfer (Pan & Yang, 2010; Yoon & Li, 2019). That is, when domain discrepancy exists, the transferred knowledge may damage instead of helping the predictive power of the models. Fortunately, this research is interested in multiple tasks from the same domain. Considering the feature map from each task is one view of the domain, and the domain discrepancy from the cross-domain transferring is not of concern, so the performance of an individual task shall be improved by cross-view feature transferring. Therefore, we propose FT-MTL-Net, an MTL with cross-view feature transferring. It is novel especially for applications in medicine. This is because object detection, segmentation, and classification are three essential and inter-related tasks in medical imaging analysis. Represented joint feature maps from the cross-view feature transferring will take advantage of the complementary power of the features from different tasks without having a domain discrepancy issue. As a result, the generalizability of the target task is enhanced on the medical dataset even with limited samples.

4.3.Proposed FT-MTL-Net

The architecture of our proposed FT-MTL-Net is shown in Figure 24. The first part of FT-MTL-Net is the backbone architecture. Similar to Mask-RCNN, the backbone consists of shared convolution layers (Conv layer) for feature map generation and a region proposal network (RPN) (Ren et al., 2017) for candidate region detection. Raw images are fed into the shared convolution layers to generate feature representations for all subsequent tasks (e.g., detection, segmentation, and classification). RPN uses bounding boxes with pre-defined sizes to search entire raw images and outputs a set of rectangular candidate regions. Each candidate region is treated as an ROI candidate with a corresponding area within the feature map to describe it. Feature maps for ROI candidates are resized to be the same through a bilinear interpolation (ROI-align (He et al., 2017)) to be fed into the head structures. Following the backbone, three head architectures are proposed to focus on these ROI candidates and make ROI-oriented predictions. Specifically, the detection head refines the ROI candidates for an accurate bounding box. The segmentation head generates masks for each ROI candidate. The classification head predicts whether the ROI candidates are benign or malignant.

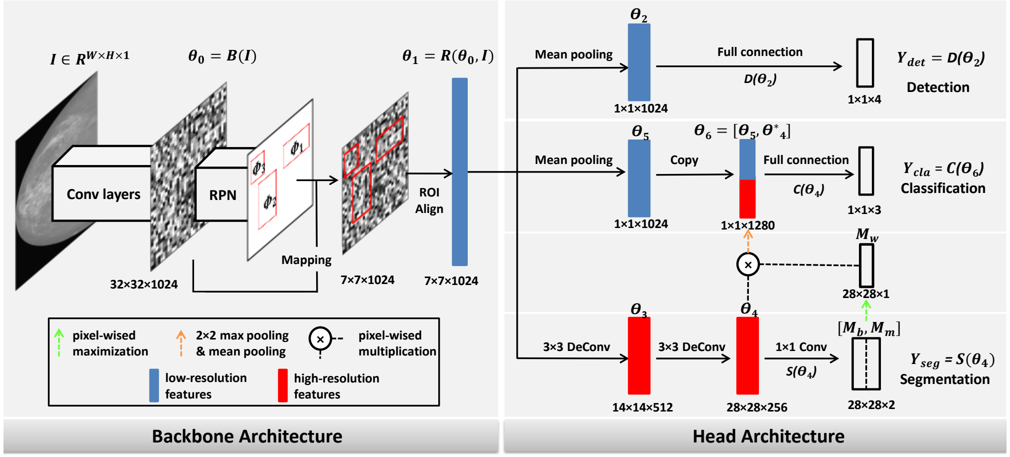


Figure 24 Architecture of proposed FT-MTL-Net.

4.3.1 Backbone Architecture

4.3.1.1. Shared Convolution layers for Feature Generation

The first part of the backbone is sharing convolution layers to render feature maps. Note we use 2D images (pixel) in the following discussions for simplicity, and the same methodology applies to 3D images (voxel). Given the grayscale input image $I \in \mathbb{R}^{W\times H\times 1}$, a feature map $\theta_0 = B(I)$ is generated by mapping $B(\cdot)$ conducted by the shared convolution layers. In this research, ResNet (He, Zhang, Ren, & Sun, 2015) is adopted to serve this purpose. ResNet is a well-known deep CNN architecture with the novel design of a ‘short cut’ connection in the building block. Compared to traditional deep-CNNs, this design helps improve the performance in avoiding the problem of gradient vanishing (Drozdzal, Vorontsov, Chartrand, Kadoury, & Pal, 2017; He et al., 2016). Since inception, ResNet has been implemented in various computer vision tasks including medical applications (Fakhry, Zeng, & Ji, 2016; Gao et al., 2018). For the consideration of the balance between computation efficiency and accuracy with the limited computation resources, we use ResNet-50. The last fully-connected layer originally designed for classification is removed. Note that ResNet has 4 max-pooling layers. Let the original input image be $W \times H \times 1$ (width × height × channel; the following notations of feature map/image size follow this same format, if the channel number equals to 1, it will be omitted), the output of ResNet-50 is a feature map of $w \times h \times 1024$ ($w = W/16$ and $h = H/16$). In this study, the image resolution is 512 × 512. As a result, the feature map θ_0 is 32 × 32 × 1024.

4.3.2.1 Region Proposal Network for ROI Proposal Detection

Taking feature map θ_0 from ResNet-50 and raw image I as inputs, RPN (Ren et al., 2017) predicts object bounds and objectness at each position. The objectness score is a probability measure of an object within this specific patch. The outputs are a set of indicators for rectangular candidates (a.k.a. ROI proposals), denoted as $\Phi = \{\mathit{\Phi}_1, \mathit{\Phi}_2, \dots, \mathit{\Phi}_n\}$. Since the targeting object in the raw image can be at any location with arbitrary sizes, searching the whole raw images for regions of all possible sizes and locations is computationally prohibitive. In RPN, the candidates in P are searched on the feature map using a sliding window. A sliding window runs spatially on the feature map at a pre-defined step size *s*. For each pixel in the center, ROI candidates with pre-defined sizes are generated and mapped back to raw images. For candidate i, let $\mathit{\Phi}_i = \left(a_{iw}, a_{ih}, a_{ix}, a_{iy}\right)$, where a_{iw} denotes the width, a_{ih} denotes the height, and $\left(a_{ix}, a_{iy}\right)$ denotes the center's coordinate. If $\mathit{\Phi}_i$ has an overlap with the ground truth mask that is greater than a pre-defined threshold, it is taken as a positive ROI candidate. Otherwise, it is negative. Each $\mathit{\Phi}_i$ is represented by a 1-dimensional array of features, which is the mean value of each channel on the feature map (θ_0). These features are used to predict the objectness for each Φ_i. After training, the RPN outputs a set Φ containing ROI candidates with higher objectness scores than a predefined threshold (e.g., 0.5).

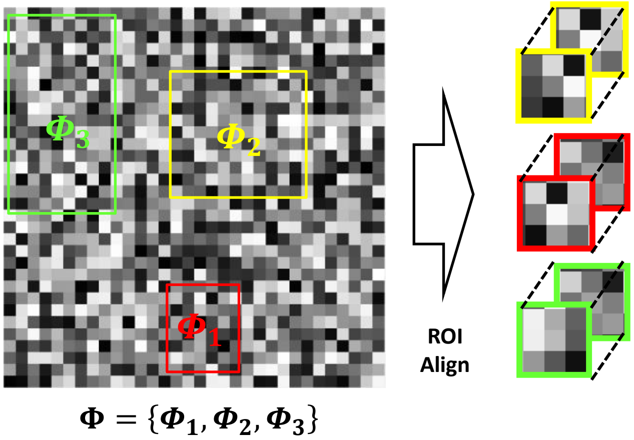

Φ_3
Φ_2
Φ_1
ROI
Align
$\boldsymbol{\Phi} = \{\boldsymbol{\Phi_1}, \boldsymbol{\Phi_2}, \boldsymbol{\Phi_3}\}$

Figure 25 Illustration of bounding boxes being resized to same size through ROI-Align.

For $\Phi_i \in \Phi$, the associated bounding boxes on the feature map vary in sizes. Therefore, the candidates are resized to the same size (7×7 in this study) through ROI align layer (He et al., 2017), a linear interpolation procedure. Next, the ROI candidates within Φ are represented with its associated feature map θ_1 of the same size (as shown in Figure 25), and shared by the head architectures (see section 4.3.2).

4.3.2 Multi-Task Head Architecture

4.3.2.1. Head Architecture for Detection Task

The detection subnetwork follows the same design by Ren et al. (2017) where a mean pooling layer is implemented to reduce the feature map resolution to one dimension. It is fully connected to the output layer of bounding box regression. The output value is associated with the corresponding ROI candidate $\Phi_i = (a_{iw}, a_{ih}, a_{ix}, a_{iy})$ before the resizing procedure. Let $T = (a_{tw}, a_{th}, a_{tx}, a_{ty})$ be the target candidate, in which

(a_{tx}, a_{ty}) denotes the predicted center coordinate and a_{tw} and a_{th} denote the predicted width and height, respectively. Assume the targeting outputs for ground truth bounding box is $\Upsilon = (a_{vw}, a_{vh}, a_{vx}, a_{vy})$, where (a_{vx}, a_{vy}) denotes the ground truth bounding box's center coordinate, a_{vw} and a_{vh} denote the width and height, respectively. The cost function for regression task is as follows:

$$L_{reg}(T, \Upsilon) = Smooth_{L1}(f(T, \Phi_i) - f(\Upsilon, \Phi_i)) \tag{4.1}$$

where,

$$Smooth_{\mathrm{L1}}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{4.2}$$

$$f(T, \Phi_i) = \left(\log\left(\frac{a_{tw}}{a_{iw}}\right), \log\left(\frac{a_{th}}{a_{ih}}\right), \frac{a_{tx} - a_{ix}}{a_{iw}}, \frac{a_{ty} - a_{ix}}{a_{ih}}\right) \tag{4.3}$$

$$f(\Upsilon, \Phi_i) = \left(\log\left(\frac{a_{vw}}{a_{iw}}\right), \log\left(\frac{a_{vh}}{a_{ih}}\right), \frac{a_{vx} - a_{ix}}{a_{iw}}, \frac{a_{vy} - a_{ix}}{a_{ih}}\right) \tag{4.4}$$

The detection head will refine the sizes and locations of ROI candidates and output the final predictions on the bounding boxes.

4.3.2.2. Head Architecture for Segmentation Task

In the segmentation subnetwork, two deconvolutional layers are introduced to increase the resolution of the feature maps for segmentation and derive task-specific feature maps (θ_3 and θ_4). Following the deconvolutional layers, one 1×1 convolutional layer is added for the final output. Per-pixel sigmoid function is applied to this final output to obtain two probability maps (M_b and M_m). Since the candidate Φ_i from RPN has 7×7, the resolution is increased by 2x2 (two deconvolution layers) resulting in M_b and M_m sized β

$\times\ \gamma$ ($\beta = \gamma = 28$). M_b and M_m describe the probabilities that each pixel is within the benign and malignant tumors independently.

The last feature map (θ_4) before the final segmentation output provides high-resolution information for each pixel along the 256 channels. The features (θ_4) are different from those from the detection task (θ_2) and classification task (θ_5, discussed in Section 3.2.3). Both θ_2 and θ_5 are abstracted features of lower resolution (He et al., 2017; Ronneberger, Fischer, & Brox, 2015; Shelhamer et al., 2017). We hypothesize that the high-resolution features from the segmentation shall help improve the classification (discussed in Section 3.2.3) greatly, thus they are transferred. Transferring high-resolution feature maps to low-resolution feature maps requires certain operations. One example is max pooling or average pooling (He et al., 2016; Szegedy et al., 2015) where the maximum or the mean values of the features are derived and transferred. Yet, such an approach treats all features inside and outside ROIs equally. Knowing medical imaging analysis mostly focuses on tumorous areas (such as in this study), we propose a prior knowledge guided feature generation method: feature values representing different regions are re-weighted based on the probability maps. A weight map M_w of size $\beta \times \gamma$ is generated based on the outputs of segmentation masks M_b and M_m:

$$M_{w_{i,j}} = \max\left(M_{b_{i,j}}, M_{m_{i,j}}\right) \text{ for } i \in [1, \beta], j \in [1, \gamma] \tag{4.5}$$

where $M_{w_{i,j}}$ is combined with the feature map θ_4 (of size $\beta \times \gamma \times \delta$) to generate a prior knowledge guided feature map $\theta_4^* = P(\theta_4, M_w)$ of the same size:

$$\theta^*_{4_{i,j,k}} = \theta_{4_{i,j,k}} \times M_{w_{i,j}} \text{ for } i \in [1, \beta], j \in [1, \gamma], k \in [1, \delta] \tag{4.6}$$

In order to generate compressed features that can be directly used by the classification task, the resolution of the feature map θ_4^* is reduced from 28×28 to 1×1 through a max pooling layer and a global mean pooling layer (similar procedure as in (Noh et al., 2015)).

For the cost function of segmentation, assume the output prediction map is $s^{\beta \times \gamma}$ of resolution $\beta \times \gamma$, the cost function for segmentation is the average cross-entropy over all the pixels within $s^{\beta \times \gamma}$ and ground truth mask $m^{\beta \times \gamma}$ (resized to resolution $\beta \times \gamma$), which can be calculated as follows:

$$L_{seg}(s, m) = (\frac{1}{\beta \times \gamma}) \sum_{i=1}^{\beta} \sum_{j=1}^{\gamma} CrossEntropy(s_{ij}, m_{ij}) \tag{4.7}$$

in which

$$CrossEntropy(y^*, y) = -y \log(y^*) - (1 - y) \log (1 - y^*) \tag{4.8}$$

The segmentation head outputs two individual probability maps that measure the likelihood of each pixel being within benign and malignant tumor respectively. Following the same setting as Mask-RCNN (He et al., 2017) to solve the overlapping issue of different types of tumors, a final mask is selected based on the output of the classification task.

4.3.2.3. Head Architecture for Classification Task

In the classification subnetwork, the feature map θ_6 for the final classification layer is of size $1 \times 1 \times 1280$. Among these 1280 feature channels, 1024 are obtained from a shared feature map θ_1 provided by the backbone through a global mean pooling; the rest 256 channels come from θ_4', which are used as an addition of pixel-wise information. The feature channels from two sources are combined and fully connected to the final classification layer with 3 outputs (background, benign and malignant), and a corresponding probability array $\mathrm{P} = (p_0, p_1, p_2)$ is computed over the 3 outputs by a softmax activation function (Krizhevsky, Sutskever, & Hinton, 2012). The cost function for the classification task is the log loss function for its corresponding class u (u = 0, 1 or 2) where 0 for background, 1 for benign and 2 for malignant.

$$L_{cls}(p, u) = -log(p_u) \tag{4.9}$$

The ROIs with high probabilities of being benign or malignant tumors are investigated for final prediction using “malignant-veto” logic described in Section 4.3.3.

4.3.3 Model Training and Inference

Table 13 Training procedure details.

Step 1. Initialize the ResNet-50 with the weights trained using natural images from the dataset of ImageNet, which is made available online by the developers of ResNet (He et al., 2016).
Step 2. Initialize the weights of all other layers through a normal distribution with mean = 0 and standard deviation = 0.05.
Step 3. Fine-tune end-to-end for the region candidate task using cost function L_{prop}.
Step 4. Keep the weights within shared layers and RPN layer fixed, tune the weights within subnetworks alone with cost function L_{uni}.
Step 5. Keep tuning the weights within shared layers and subnetworks together with cost function L_{uni}.

In the training procedure, all three tasks are trained simultaneously with one combined loss function:

$$L_{uni} = \lambda_1 L_{cls} + \lambda_2 L_{reg} + \lambda_3 L_{seg} \tag{4.10}$$

where λ_1, λ_2, and λ_3 are weights for each individual cost function. In this study, λ_1, λ_2, and λ_3 are all set to be 1 treating all three tasks equally important. A 5-step training procedure (see Table 13 Training procedure details) following the same logic in (Ren et al., 2017) is adopted. Once the training process is completed, the model is ready to make inferences for testing images.

There are two major differences between the inference workflow and the training procedure. The first difference is sequential execution vs. parallel training. That is, in inference, it follows (Step 1) ROI candidates are obtained from the backbone; (Step 2) the detection task is conducted to provide accurate bounding box predictions; (Step 3) the segmentation task is triggered to generate mask predictions and features based on the bounding boxes; (Step 4) features from segmentation are transferred and joined for classification. The second difference is an added "malignant-veto" logic motivated by the medical practices in the inference workflow. As expected, each medical case often may have multiple bounding boxes and thus ROIs to be investigated. We define the "malignant-veto" logic as if one bounding box is predicted as malignant, this mass will be predicted as malignant with a score equaling the maximum score among all these boxes indicating malignancy; if none of the bounding boxes indicates malignancy, it gets a malignancy score $[1 - Sb_{max}]$, where Sb_{max} is the maximum score among all the bounding boxes assigned with a benign score.

4.4.Experiment and Results

4.4.1 Dataset

The dataset used in this study is obtained from INbreast, an online accessible full-field digital mammographic (FFDM) database (Moreira et al., 2012). INbreast was established by the researchers from the Breast Center in CHJKS, Porto, under the permission of both the Hospital's Ethics Committee and the National Committee of Data

Protection. The FFDM images were acquired from the MammoNovation Siemens system with a pixel size of 70 mm (microns), and 14-bit contrast resolution. The resolution of each image is 2560 × 3328. For each subject, both CC and MLO views are available. For each image, the annotations of region of interests (ROIs) (ground truth masks) were made by a specialist in the field and were validated by a second specialist. The ROI masks were also made available through the attached XML file.

In this research, 108 subjects with labeled masses are selected. Each mass is assigned with a Breast Imaging Reporting and Data System (BI-RADS) (Eberl, Fox, Edge, Carter, & Mahoney, 2015) score ranging from 2 to 6. Following the same definition in (Dhungel et al., 2017a), the masses with BI-RADS score=2, 3 are treated as benign and the remaining cases (BI-RADS=4, 5, 6) are labeled as malignant. There are 37 benign subjects and 71 malignant subjects.

4.4.2 Data pre-processing

For cases with multiple masses in one image, each individual mass and its corresponding bounding mask is extracted and saved as a new data sample. As a result, the total number of cases in the dataset increases to 115 (41 benign vs. 74 malignant). For each mass, a bounding box is computed as the minimal rectangle in the image that contains the whole mass. In the second step, for each breast image, a rectangle that contains the entire breast is obtained, and the region outside of this bounding box is excluded. This step is to exclude the background region in each image and reduce search space and computational burden during the training process.

Five-fold cross validation is adopted, and data augmentation is implemented to enrich the training dataset. Specifically, within each fold, the training dataset (80%) is augmented by randomly selecting 2 to 5 options from the operations including rotating, flipping, zooming in/out, cropping, contrasting enhancement and Gaussian smoothing. The image, mask and bounding box will go through the same procedure. Considering the imbalance of benign cases vs. malignant cases, each benign sample is augmented 150 times, and each malignant sample is augmented 75 times, so the ratio of benign and malignant cases is roughly 1:1. The final training dataset has 9360 images (4920 benign vs 4440 malignant).

4.4.3 Experimental Setup

The experiments are conducted on a Windows desktop with 32G RAM and an Intel 16-core CPU. The model is trained using one single NVIDIA Titan XP GPU with 12G memory. Both the data processing procedure and the architecture are developed with Python and deep learning libraries (e.g., Keras and TensorFlow). The whole architecture is built upon the MASK RCNN package downloaded through the open-source website GitHub (https://github.com/matterport/Mask_ RCNN). Details of tuned parameters are: (1) training iterations for the 4 training steps are set to be 10; (2) the learning rate for each step is set to be 0.005 with a momentum equal to 0.9; (3) the training batch size is set to be 8 to satisfy the GPU memory; (4) other parameters are set with default values provided by Keras or the downloaded Mask RCNN package.

4.4.4 Experimental Results

FT-MTL-Net is designed for three inter-related tasks in medical applications: classification, object detection, and segmentation. High-resolution features from the segmentation task are transferred to the classification task for improving performance. In the comparison study, we decide to compare the proposed FT-MTL-Net with methods in classification, object detection, and segmentation, respectively. These include some methods that only focus on one of the three tasks, e.g., classification, as well as methods handling multiple tasks. To the best of our knowledge, Mask-RCNN (He et al., 2017) may be the only method that addresses all three tasks jointly for medical applications. We include Mask-RCNN in the comparison on all three tasks with the competing methods. In addition, detailed comparison analysis between FT-MTL-Net and Mask-RCNN is provided.

4.4.4.1. Classification Task

A response operating characteristic (ROC) curve is commonly used to evaluate the classification performance, especially in medical imaging applications. ROC is a function of true positive rate (TPR) with respect to 1- false positive rate (1-FPR). The area under the ROC curve (AUC) is used as a metric to evaluate the classification performance of a model.

Table 14 summarizes the comparison results. The first three methods take manually delineated ROIs from domain experts as inputs and focus on the classification task only. The AUC ranges from 0.86 to 0.91. The following four pipelined systems are automated systems taking the whole images detecting the objects and classifying them. Here we take the classification results for comparison, and the AUC ranges from 0.76 to 0.86. It is not surprising the AUC performances from the pipelined system are not as good as that from the one-task approaches as the later heavily involves the domain experts to provide accurate segmentations. However, the delineation of the ROIs by experts is time-consuming and may not always be available. In looking at the multi-task category, we observe the approaches in the category outperform most one-task and pipelined systems. Though Mask-RCNN has an AUC of 0.89, lower than that from Random Forest on CNN (0.91), Mask-RCNN has a much smaller standard deviation, 0.02 compared to 0.12 from the Random Forest, indicating the robustness of the model.

Table 14 Comparison between our proposed model and eight competing methods on mass classification on INBreast Dataset.

Method	**Configuration**	**AUC**
Transfer learning from deep CNNs + ensembled classifiers (Huynh et al., 2016)	one task	0.86 ± 0.01
Lib SVM (Diz, Marreiros, & Freitas, 2016)	one task	0.90
Random Forest on CNN with pre-training (Dhungel et al., 2017a)	one task	0.91 ± 0.12
Random Forest on CNN with pre-training (Dhungel et al., 2017a)	pipelined system	0.76 ± 0.23
Multi-view Residual Network (Dhungel, Carneiro, & Bradley, 2017b)	pipelined system	0.80 ± 0.04
Deep learning through unregistered views (Carneiro, Nascimento, & Bradley, 2017)	pipelined system	0.78 ± 0.09
Pre-trained CNNs + multiple instance learning (Zhu, Lou, Vang, & Xie, 2017)	pipelined system	0.86 ± 0.03
Mask-RCNN (He et al., 2017)	multi-task	0.89 ± 0.02
Proposed FT-MTL-Net	multi-task	**0.92 ± 0.01**

In comparing our proposed FT-MTL-Net with Mask-RCNN (see Figure 26), the ROC curve from FT-MTL-Net, in general, dominates that from Mask-RCNN. FT-MTL-Net has AUC 0.92 ± 0.01 compared to Mask-RCNN with 0.89 ± 0.02. A paired t-test gives $p<0.01$ indicating FT-MTL-Net significantly outperforms Mask-RCNN on AUC value.

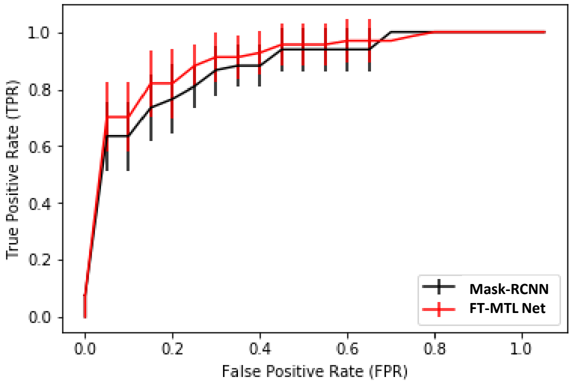


Figure 26 ROC curves for Mask-RCNN and our proposed FT-MTL-Net model on test dataset (vertical line denotes 2×TPR std across 5 folds).

From this comparison, three conclusions are drawn: (1) the FT-MTL-Net outperforms both pipelined approaches and traditional one-task approaches in terms of AUC. This indicates joint advantages of multiple tasks; (2) One task approaches need time-consuming manual processing, which requires expert knowledge and manual steps to find the suspicious regions, whilst the FT-MTL-Net is an automated end-to-end approach; (3) the FT-MTL-Net outperforms Mask-RCNN with statistical significance.

Please note as the first attempt into MTL, our current design of FT-MTL-Net only transfers the segmentation features into the classification. Because MTL approaches like the one we propose can improve multiple tasks in general, we are still interested in exploring the performance of the detection and segmentation tasks with respect to the competing methods. This is discussed in the following two sections.

4.4.4.2. Detection Tasks

For the detection experiment, we first present the comparison results in mean true positive rate (TPR) across 5 folds and false positive rates per image (FPI) (see Table 15).

Since the literature reports the TPRs under different FPIs, for a comprehensive and fair comparison, we derive two sets of TPRs under different FPI settings: FPI = 3.67 and/or 5. Standard deviation across 5 folds is reported. As seen from Table 15, multi-task learning approaches (Mask-RCNN and FT-MTL-Net) have comparable detecting power as traditional one-task detection models and pipelined systems. It should be noted that the multi-view residual network (Dhungel et al., 2017b) achieves the best performance (0.96±0.03@0.8). This is because, after the detection module, a specifically designed cluster method is implemented to remove overlapping for both true positives and false positives. We intend to further improve the detection performance by adopting some new postprocessing methods such as those proposed by Dhungel et al. (2017b), as a future study.

Table 15 Comparison between our proposed FT-MTL-Net model and other competing methods on mass detection on INBreast dataset.

Method	**Configuration**	**TPR@FPI**
Adaptive thresholding + machine learning (Kozegar, Soryani, Minaei, & Domingues, 2013)	one task	0.84@3.67
Cascaded Deep Learning +Random Forests (Dhungel, Carneiro, & Bradley, 2015)	one task	0.78@3.67
Random Forest on CNN with pre-training (Dhungel et al., 2017a)	pipelined system	0.87@5
Multi-view Residual Network (Dhungel et al., 2017b)	pipelined system	0.96±0.03@0.8
Deep learning through unregistered views (Carneiro et al., 2017)	pipelined system	N.A.
Pre-trained CNNs + multiple instance learning (Zhu et al., 2017)	pipelined system	N.A.
Mask-RCNN (He et al., 2017)	multi-task	0.85 ±0.07@3.67 0.85 ± 0.07@5
Proposed FT-MTL-Net	multi-task	**0.91 ±0.05 @3.67** **0.91 ± 0.05@5**

Next, we compare FT-MTL-Net with Mask-RCNN. We use the free response operating characteristic (FROC) curve to present its performance. FROC is a function of true positive rate (TPR) with respect to false positive rate per image (FPI). Following the same standard in experiment conducted by Dhungel et al. (2017a), we define: if the intersection of union (IoU) between predicted bounding boxes and ground truth is greater than 0.2, this bounding box is regarded as true positive, otherwise, it will be regarded as false negative. From Figure 27, we observe that FT-MTL-Net achieves a TPR of 0.91 with a standard deviation of 0.05 (TPR = 0.91 ± 0.05) at FPI = 3.67 on the testing dataset. In fact, this TPR (0.91) tends to be stable for FPIs that are greater than 1.5. The Mask-RCNN obtains a TPR = 0.85 ± 0.07 at FPI = 3.67. A t-test is conducted on the TPR values obtained among the 5 folds for Mask-RCNN and our proposed FT-MTL-Net. With a p-value < 0.05, we conclude FT-MTL-Net outperforms Mask-RCNN. One may be surprised to observe such performance as our FT-MTL-Net indeed takes the same architecture as proposed by Ren et al. (2017) for the detection task. This may be explained as follows: in the testing stage, each detected bounding box uses the probabilities (background vs. benign tumor vs. malignant tumor) from the classification task as its objectness score. FT-MTL-Net has a classification head architecture with enhanced capability that is not only better at differentiating benign tumors from malignant tumors, but also better at classifying tumors from background regions. This capability, in turn, helps improve the detection task indirectly. To measure the robustness of the detection results on different IoU thresholds, the average precision curve is shown in Figure 28. It is a function of true positive rate against the different IoUs. It is noted for values where IoU <= 0.4, the TPR remains stable and consistently is above 0.9. The TPR starts to decrease if IoU is greater than 0.4. As a result, we set IoU = 0.4 as our threshold to define whether a mass is detected by the predicted bounding box for the following two tasks. The performances for segmentation and classification are evaluated only on the detected mass which takes an average of 95% of the testing dataset according to the curve. In integrated systems such as the one proposed by Dhungel et al. (2017a), similar approaches are implemented by manually excluding all false positives.

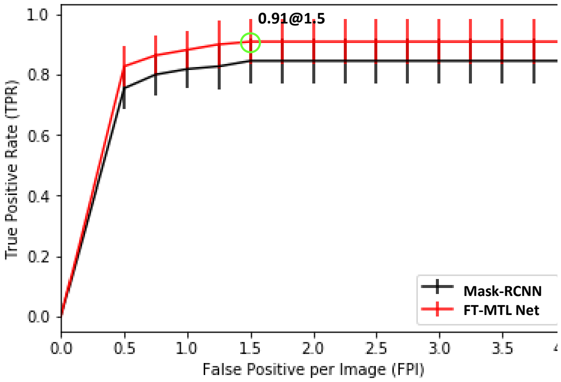


Figure 27 FROC curves for Mask-RCNN and our proposed FT-MTL-Net model (IoU > 0.2, vertical line denotes 2×TPR std across 5 folds).

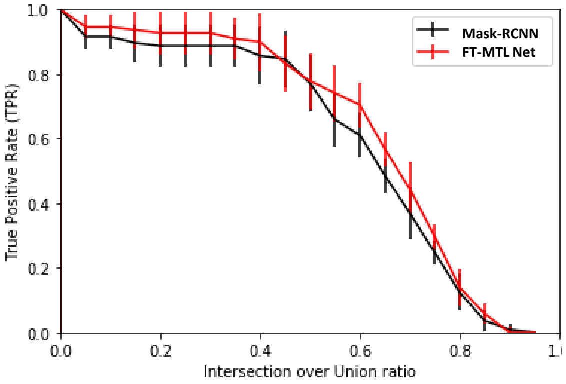


Figure 28 Average detection precisions under different IoU settings on the testing dataset for Mask-RCNN and our proposed FT-MTL-Net model (vertical line denotes 2×TPR std across 5 folds).

In summary, from this comparison experiment, we derive at two conclusions: (1) FT-MTL-Net outperforms most of the competing methods (see Table 15) except Multi-view Residual Network (Dhungel et al., 2017b) which has 0.96±0.03@0.8. This is because Multi-view Residual Network has a post-process procedure to remove the overlapping for both true positives and false positives, which helps improve the performance of the detection task. (2) Compared to Mask-RCNN, FT-MTL-Net significantly outperforms. FT-MTL-Net has a classification head architecture with enhanced capability from the classification task as its objectness score. FT-MTL-Net is not only better at differentiating benign tumors from malignant tumors, but also better at classifying tumors from background regions. This capability, in turn, helps significantly improve the detection task.

4.4.4.3. Segmentation Tasks

The segmentation performance is quantified with Dice similarity index (Dice, 1945). Let A be the predicted mask, and B be the ground truth mask, Dice can be calculated through the following equation:

$$Dice(A, B) = \frac{2(A \cap B)}{A \cup B} \tag{4.11}$$

Where

$A \cap B$ counts the number of pixels that are labeled with 1s in both masks A and B.

$A \cup B$ counts the number of pixels that are labeled with 1s in either mask A or B.

We compare FT-MTL-Net with Mask-RCNN, 1 one-task method, and the same four pipelined systems. From Table 16, we observe these two MTL models underperform the other competing methods to a certain degree. The reason may be that, in the methods proposed by Dhungel et al. (2017a) and Al-antari et al. (2018), the input training images are outputs from a former detection procedure, there is a ‘manual intervention’ procedure that will exclude all the false positive detections and this helps improve the performance of segmentation results. The MTL models are fully automatic model without any user intervention. The segmentation network is trained with both true positive and false positive detections from the RPN, and the false positive detections have a negative influence on segmentation results. Another reason may come from an architecture aspect: the feature maps used for segmentation are highly reduced in spatial resolution compared with the original masks. Before the segmentation network, 4 max-pooling layers are implemented within the shared convolutional layers, in which important pixel information for segmentation is lost (Chen et al., 2017). Such lost information is difficult (if not impossible at all) to retrieve through the subsequent layers. With limited pixel information, the segmentation network may suffer from low accuracy. Noting this, our plan for the next steps is to improve FT-MTL-Net with a focus on segmentation improvement. For example, we may add a connecting path from high-resolution features to enrich feature sets as those in U-Net (Gao et al., 2019; Ronneberger et al., 2015) and SegNet (Badrinarayanan et al., 2017).

Table 16 Comparison between our proposed FT-MTL-Net model and other competing methods on segmentation with INBreast Dataset.

Method	**Configuration**	**DICE index**
FrCNN (Al-antari et al., 2018)	one task	92.67
Random Forest on CNN with pre-training (Dhungel et al., 2017a)	pipelined system	0.85 ± 0.02
Multi-view Residual Network (Dhungel et al., 2017b)	pipelined system	N.A.
Deep learning through unregistered views (Carneiro et al., 2017)	pipelined system	N.A.
Pre-trained CNNs + multiple instance learning (Zhu et al., 2017)	pipelined system	N.A.
Mask-RCNN (He et al., 2017)	multi-task	0.79 ± 0.02
Proposed FT-MTL-Net	multi-task	**0.76 ± 0.03**

From Table 16, four conclusions can be drawn from this comparison experiment: (1) MTL models in the segmentation task underperform the other competing methods to a certain degree. It is not surprising that the performance on the segmentation task of the FT-MTL-Net is not as good as those from one task and pipelined task approaches. This is because in the experiments conducted by Dhungel et al. (2017a) and Al-antari et al. (2018), an extra procedure is introduced to manually exclude all the false positive detections, and this helps improve the performance of segmentation results; (2) the feature maps in MTL used for segmentation are highly reduced in spatial resolution compared with the original masks. With limited pixel information, the segmentation network may suffer from low accuracy in the MTL framework; (3) two competing methods require manual configuration, but the MTL approaches is an automated end-to-end approach; (4) compared to the multi-task approaches, FT-MTL-Net shows inferior results (0.76±0.03) to that of Mask-RCNN (0.79±0.02). For conclusion #4, we conduct further investigation to understand the performance. We conclude FT-MTL-Net underperforms Mask-RCNN is because the reported segmentation results are based on the detection outcome (one of the reasons why multi-task frameworks are needed for medical applications). Among the 115 images, ROIs within 97 images are correctly detected by both methods. ROIs within 7 images are missed by both; ROIs from 3 images are detected by Mask-RCNN only, and ROIs from 8 images are detected by FT-MTL-Net only. This is supported by the detection metric (0.91 from FT-MTL-Net vs. 0.85 from Mask-RCNN). The DICE from Mask-RCNN is derived from the 100 cases (97+3) while the DICE from FT-MTL-Net is derived from the 105 cases (97+8). For illustrative purposes, we have the 3 images and the 8 images shown in Figure 29 and Figure 30, respectively. By visually checking these images, the 8 cases handled by FT-MTL-Net show smaller ROIs, and some (e.g., the first case on the top left) have very irregular shapes. As a result, FT-MTL-Net has lower DICE than that from Mask-RCNN.

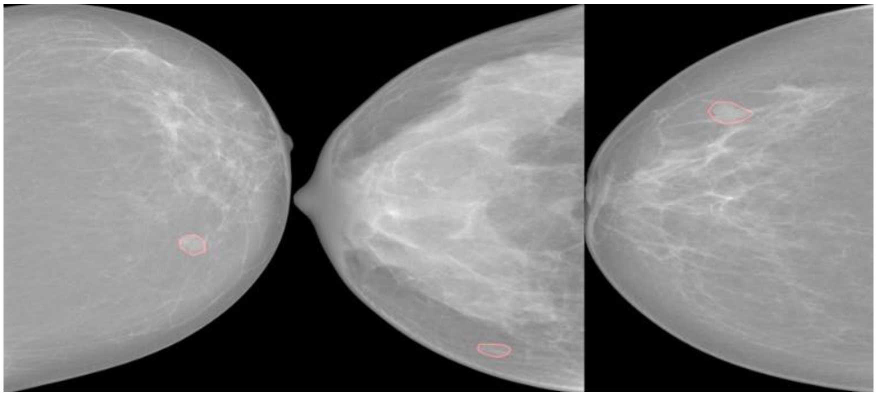


Figure 29 Three images with ROIs detected by Mask-RCNN only (red contour denotes boundary of ground true mask).

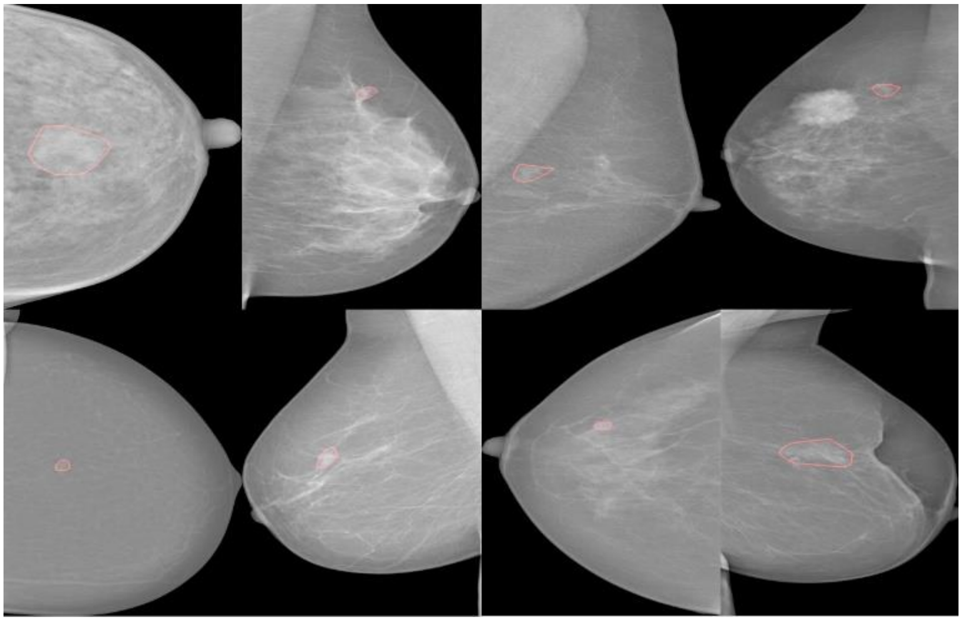


Figure 30 Eight images with ROIs detected by FT-MTL-Net only (red contour denotes boundary of ground true mask).

4.4.4.4. Illustration

To demonstrate the functions of FT-MTL-Net, we present the prediction results from the two cases and their corresponding outputs after different steps in Figure 31. As shown, each raw image is fed into the trained model. After the backbone architecture, several candidates (marked with yellow dashed bounding box) of pre-defined size and with objectness score (O score) greater than 0.5 are detected (the above case has two candidates and the bottom case has only one). These candidates are resized to the same size and fed into the head architecture. Through the head architecture, each candidate’s bounding box (dashed bounding boxes) are refined by the detection task; the mask (solid contour region) are predicted through the segmentation task; the classification task assigns each candidate a probability of being malignant or benign (M score/B score). These predicted results are finalized through the “malignant-veto” logic introduced above to reduce the overlapping detections. The illustration shows FT-MTL-Net accurately identifies suspicious regions within breast images, makes good predictions on the suspicious regions’ categories, and outputs segmentation masks with reasonable accuracy.

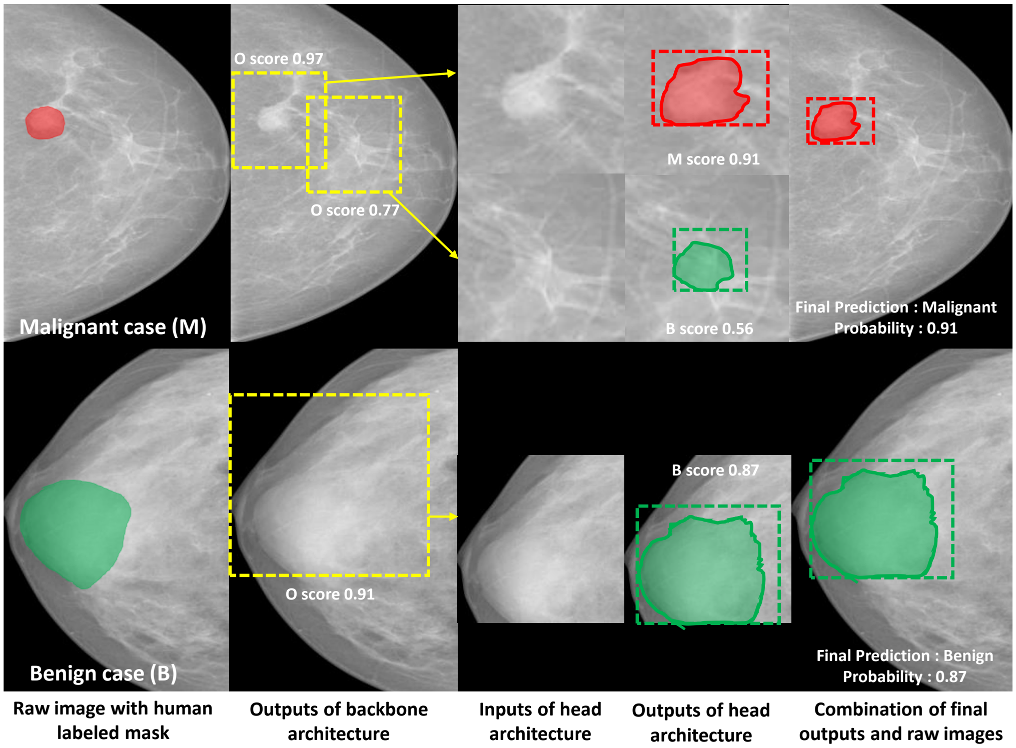

O score 0.97
O score 0.77
M score 0.91
B score 0.56
Malignant case (M)
Final Prediction : Malignant
Probability : 0.91
O score 0.91
B score 0.87
Benign case (B)
Final Prediction : Benign
Probability : 0.87
Raw image with human labeled mask
Outputs of backbone architecture
Inputs of head architecture
Outputs of head architecture
Combination of final outputs and raw images

Figure 31 Examples of two cases (malignant case on top and benign case on bottom) and their corresponding outputs from different steps.

4.5.Conclusion and Discussion

Most image analysis applications are related to one or more tasks in object detection, segmentation, and classification. Multi-task deep learning thus becomes a viable solution to tackle these tasks together, as it provides the advantages of both multi-task learning and deep neural networks. While the success from Mask-RCNN (He et al., 2017), a pioneering research in MTL field is acknowledged, we recognize the core of existing MTL models (including Mask-RCNN) is separate deep models for each individual task. Under the assumption that the features from one deep model (for one specific task) will be valuable to a different model (a different task), we propose a new MTL architecture, FT-MTL-Net, enabled by the features transferring in between the tasks.

The advantages of our approach are four-fold: firstly, the FT-MTL-Net does not need manual configuration for each task. To the best of our knowledge, our proposed FT-MTL-Net may be one of the first fully automatic systems that addresses detection, segmentation and classification of tumors in medical imaging, and FT-MTL-Net can simultaneously be trained end-to-end. Second, unlike most MTL models—which focus on the unified cost function at the end—FT-MTL-Net restructures the models by transferring the features from one task model to a different task model. Specifically, the FT-MTL-Net improves the classification task by utilizing the features from low pixel-wise prediction in the segmentation task. Third, the features transferred are from the same domain but different tasks. Considering each task is a different view of the same domain, this cross-view feature domain is free from negative transfer issues. Lastly, the features transferred are re-weighted based on the targeted ROIs, resulting in ~700 parameters being added to the new model. Compared to the ~2M parameters in Mask-RCNN, the added computational burden is negligible.

As for the future direction, we plan to explore the features transferred across all three tasks to improve the performance of all three tasks together. The computing burden with the added parameters to enable the cross-task features shall be evaluated. Next, we plan to further validate our proposed FT-MTL-Net in other clinical applications (e.g., brain tumor), and with different imaging modalities (e.g., MR).

CHAPTER 5

AD-NET: AGE-ADJUST NEURAL NETWORK FOR IMPROVED MCI TO AD CONVERSION PREDICTION

5.1 Introduction

Alzheimer's disease (AD) is one of the most common progressive neurodegenerative diseases in elderly patients. Over 5.5 million Americans presently suffer from AD, and the number is expected to increase to 16 million by 2050 with projected healthcare cost reaching to $1.2 trillion (Gaugler, James, Johnson, Scholz, & Weuve, 2016). Early detection is critical for AD because that is when the intervention can be more effective before irreversible brain damage occurs. Thus, mild cognitive impairment (MCI), a pre-dementia stage, has been of great interest in both AD research and clinical practices. MCI is the stage when the individual has greater cognitive decline than expected from the normal aging but has not shown noticeably interruptions from the daily activities (Selkoe, 1997). Studies show that MCI patients with memory complaints and deficits (amnestic mild cognitive impairment) have higher risk of progression to AD (Gauthier et al., 2006). This calls for a deeper investigation to classify the MCI patients to be a converter (who will progress to AD) vs. a non-converter (who will remain at a stable stage). This is a non-trivial task. Fortunately, recent studies have demonstrated that medical images can more sensitively and consistently measure the disease progression than cognitive assessment (F. Li & Liu, 2018). Imaging biomarkers as the objective and quantitative criteria thus have been intensively studied as potential means for AD early detection.

Most research on AD imaging biomarkers focuses on discovering the features directly measured from the images, structural images (e.g., MR) and functional images (e.g., PET). Some structural imaging-based biomarkers show great promises as diagnostic criterions for AD. For instance, the atrophy rate per year (Fox, Cousens, Scahill, Harvey, & Rossor, 2003), hippocampal volume (Chupin et al., 2009), derived from the serial structural MRI are found to be able to differentiate AD patients vs. healthy individuals. The patterns of cortical thickness and cortical regions measured from structural MRI show the potential to discriminate the MCI converters vs. MCI non-converters (Eskildsen et al., 2013), non-converters are the subjects who remain stable for three years). Alternatively, functional imaging has been explored for AD diagnosis and early detection. PET with ^{18}F-fluorodeoxyglucose (^{18}F-FDG-PET) (Filippi et al., 2012) and PET with Pittsburgh compound B (PIB-PET) (Pike et al., 2007) are clinically mature functional imaging-based biomarkers to detect the early-stage of AD. They are becoming essential to monitor the progression of AD. Biomarkers from resting-state functional MRI (fMRI) has also be studied for the same purpose (H. J. Li et al., 2015; Yamada et al., 2017).

Biomarker discovery requires joint efforts from predictive modeling and medicine domain knowledge. Earlier works on modeling have been mainly related to machine learning pipeline, where feature extraction and selection are usually the first steps. Hu et al. (2015) use wavelet transform method to extract multi-scale features from the preprocessed structural MRI followed by a Support Vector Machine (SVM) to differentiate MCI-converters and MCI non-converters. Hojjati et al. (2017) introduce graph theory to extract features from resting-state fMRI where features are treated as a graph by constructing a brain connectivity matrix. Multiple features selection methods (e.g. Chi-square, Gini, Fisher, et al.) are employed to identify an optimal feature set as the input to SVM to classify MCI-converters vs. non-converters. Westman et al. (2012) extract features from MRI images and cerebrospinal fluid (CSF), followed by a multivariate model on the combined features to differentiate AD vs. healthy control, MCI vs. healthy control, and MCI-converter vs. non-converter. Young et al.(2013) extract features from multi-source data (MRI, FDG-PET, cerebrospinal fluid, and APOE genotype) and develop a Gaussian process classifier to predict MCI-converter. It is noted that combining data from multiple sources demonstrates improved discriminatory power than using imaging features alone. But, it may result in high-dimensional feature set which makes machine learning models prone to overfitting. A standard technique to prevent overfitting is regularization. Ye et al. (2012) employ a sparse regularized logistic regression model with a stability strategy to guarantee the model's regularization ability. The model is then tested on features extracted from MRI, demographic, genetic and cognitive measurements for classifying MCI-converters and MCI non-converters.

Most recently, deep learning is introduced to AD research. Deep Neural Network (DNN) model is an artificial neural network with multiple layers. It has been successfully implemented in the broad computer vision domains for decades (Lecun, Bengio, & Hinton, 2015; LeCun, Bottou, Bengio, & Haffner, 1998). In related to AD, most efforts are to take the deep learning model as a feature extractor where generic (low-level) and/or problem specific (high-level) features are extracted from layer to layer. It is noted that the earlier layers of a deep model contain more generic features that could be used for many domains and the features from later layers are more domain specific (Nogueira, Penatti, & dos Santos, 2017). The features are then used in different machine learning models for AD diagnosis. One example is from Shen et al. (2013) in which the low-level features (e.g. gray matter tissue volume, mean intensity, et al.) from structural MRI and PET images and

CSF is first extracted. A Stacked Auto-Encoder (SAE) model is constructed as an unsupervised pre-training model to learn the latent or hidden representation (high-level features) from those low-level features. Upon features elected from a multi-task learning model, a multi-kernel SVM model is developed to classify AD vs. MCI patients. Motivated by the success from SAE model, Shi et al. (2018) design a Stacked Deep Polynomial Network (SDPN) model to learn the features from both structural MRI and PET images. To save the preprocessing step, Suk et al. (2014) develop a Deep Boltzmann Machine (DBM) model to capture the high-level features directly from the raw images and apply a weighted ensemble SVM classifier to differentiate AD vs. MCI patients. Other than implementing different machine learning models on the features extracted from a deep model, researchers append the deep model with one last layer as a classifier for AD diagnosis and staging. For example, Basaia et al (2019) build a simplified Convolutional Neural Network (CNN) without the need of activation layer for AD diagnosis. Spasov et al. (2019) design a parameters-efficient multi-task CNN model for increased generalizability to predict MCI-Converter. Lee et al. (2019) apply a Recurrent Neural Network (RNN), to learn from multi-source data (demographic information, neuroimaging phenotypes measured by MRI, cognitive performance, and CSF measurements) to identify the person with higher risk of developing AD.

While deep learning opens great opportunities in medical research, its potential is compromised by the limited data available in the domain. Unlike natural images, medical images are rarely available in large quantities. As a result, overfitting is a major obstacle faced by deep learning research community (Lever, Krzywinski, & Altman, 2016; Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). One solution is transfer learning, that is, the deep model is first pre-trained on a large labelled dataset (e.g., natural images) to capture the features from images in general. The model is then fine-tuned on the targeted image dataset to extract specific features related to medical images. Therefore, the earliest attempts take a network model as two parts: (1) the first *N* layers are for high-level feature extraction, and (2) the last layer is a classifier. We category them as "*N*+1" models. The whole network (*N*+1) is pre-trained on the source domain. In the fine-tuning procedure, the last one layer is replaced with the appropriate classification structure tied to the target problem. For example, Hon et al.(2017) use the VGG16 and Google Inception v4 CNN model to pre-train images (>1 millions) from the ImageNet Challenge dataset and fine-tune the last fully connected layer with a Softmax layer (essentially a multiclass logistic classifier) on the MRI images for the final AD diagnosis. Similarly, Hosseini-Asl et al. (2018) pre-train a 3D Convolutional Auto-Encoder (CAE) and fine-tune the fully connected layers with a Softmax layer for AD diagnosis. This approach may work well if the data from the source domain and target domain have higher degree of similarity, e.g., all images are of same modality. In case they differ greatly, researchers decide to further divide the first *N* layers into (1) earlier layers for low-level feature exaction; (2) middle layers for high-level feature exaction. While the pre-training is still conducted on the whole network model, fine-tuning on the target domain would involve the middle and last layer of the network. Some example efforts in this direction include Cheng et al. (2017), Lu et al. (2018). The research reviewed above takes pre-training and fine-tuning as two independent procedures. Lately, researcher start to explore integrating not only the feature extracted from the pre-training, but additional features from different sources, into the fine-tuning procedure for improved performance. For example, Liu et al. (2017) fuse the features extracted from a pre-trained VGG model with several texture features into a feature pool. Zhang et al. (2018) combine the features extracted from pretrained CNN model with handcrafted visual features (e.g. Bag-of-Features, Local Binary Pattern et al.) to classify the type of different medical images (e.g. CT, MRI, Ultrasound). Song et al. (2017) generate Fisher Vector (FV) descriptors integrating the features from DBN model, CNN model in an unsupervised manner~~.~~

Please note most existing efforts on transfer learning focus on extracting and transferring features from the pre-training procedure. The outcome (a.k.a. knowledge) from the pre-training process thus is ignored. Here we hypothesize that transferring knowledge from the pre-training to the fine-tuning may benefit the target problem solving. The knowledge of particular interest in this research is related to a new AD surrogate biomarker (Cole et al., 2017). The researchers train a deep learning model on MRI neuroimages from healthy subjects to predict each subject's biological age (B-Age) (Cole et al., 2017). The trained model is then used to predict the BAs for the subjects with brain disease. Under the assumption that BA shall align well with chronological age (C-Age) for heathy subject and the BA and CA for unhealthy individuals shall present notable differences, the difference (termed Δ_{age}) is used to detect group differences between diseased cohort vs. healthy cohort (Cole et al., 2017). Motivated by this knowledge related to B-Age vs. C-Age, we propose a new deep learning model named Age-adjust neural network (AD-NET) to predict MCI converter vs. non-converter on individual bases. In the AD-NET, we revisit the transfer learning and propose dual purposes from the pre-trained model: (1) feature transferring: similar to existing research from literature, the pre-trained model without the last layer is used as feature extractor; (2) knowledge transferring: the whole pre-trained model is kept into the fine-tuning stage to transfer the knowledge captured in the age prediction process. Instead of simply appending the Δ_{age} as an additional feature to CNN model, we propose to adjust the prediction based on both Δ_{age} and the correlation between Δ_{age} and MCI-converter. Experiments are conducted using two public brain imaging datasets (IXI ("IXI Dataset," n.d.) and ADNI (F. Li & Liu, 2018)). We compare our proposed AD-NET with 8 existing methods including logistic regression, SVM and deep learning models, our AD-NET achieved the best AUC of 0.81 (±0.05) and comparable accuracy, sensitivity and specificity, which are 0.76 (±0.03), 0.77 (±0.07) and 0.76 (±0.09) respectively.

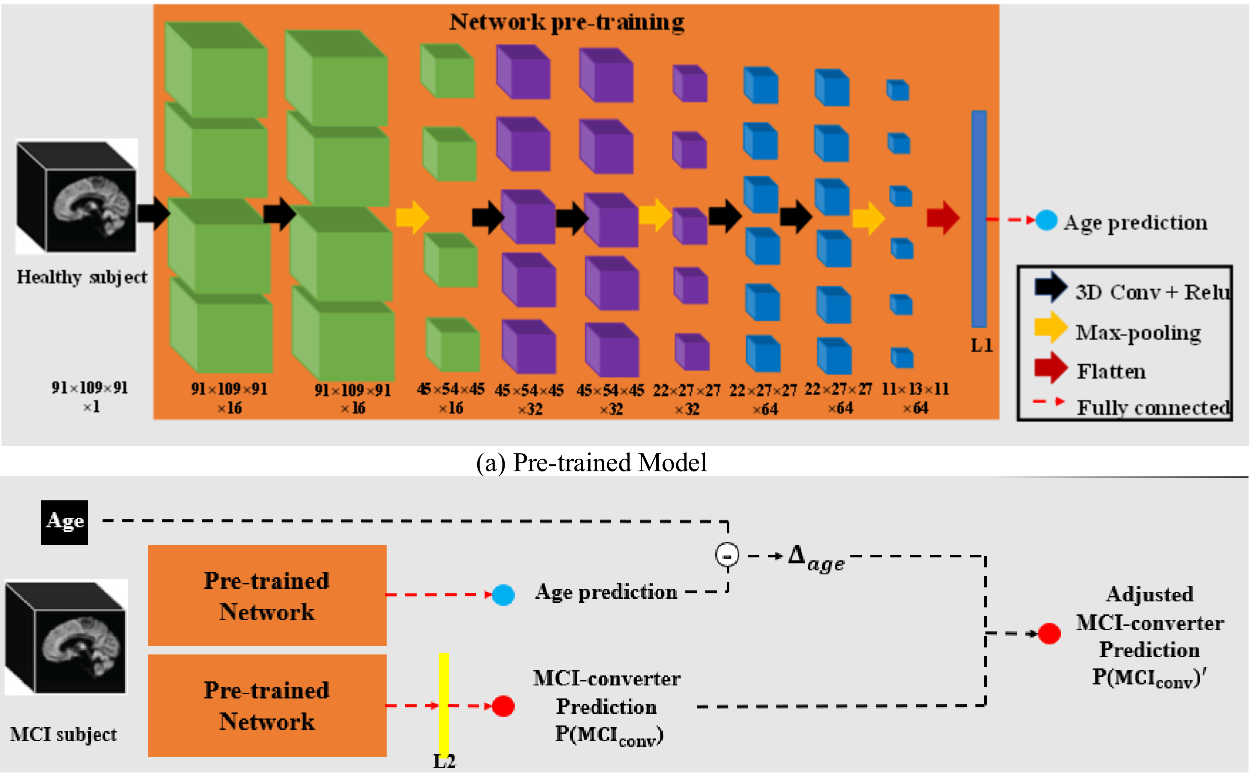
5.2 Methodology

5.2.1 Architecture and training strategy

The schematic illustration of proposed AD-NET architecture is shown in Figure 32. It contains two separate parts: (1) a pre-trained network for feature extraction and age prediction; and (2) a fine-tuned network to transfer both features and knowledge in age prediction for MCI converter prediction.

Figure 32a is the pre-trained network. It takes 3D MR images from healthy subjects as inputs and predicts age and extracts related features. The size of input 3D MRI is 91×109×91. It contains repeated 3 blocks, within each block, there are two (3×3×3) convolutional layers and one max-pooling layer; each convolutional layer is followed by a rectified linear unit (ReLU) layer. The number of feature channels is set to be sixteen for the first block and is doubled for each subsequent block. The output of last block is flattened into one dimension (layer L1 colored with blue in Figure 32). This layer is fully connected to one single output with linear activation function. A dropout layer with rate equals to 0.2 (as in (Gao et al., 2019; He, Zhang, Ren, & Sun, 2016)) is added to avoid potential overfitting.

The overall architecture of the fine-tuned model is shown in Figure 32b. Specifically, the L1 layer is fully connected (with dropout rate = 0.2) to L2 layer, which is connected (with dropout rate = 0.2) the final single output with sigmoid activation function for MCI conversion prediction. L2 layer is added to make proper feature transformation from age prediction task and produce the initial output of MCI-Converter prediction task ($\mathrm{P(MCI_{conv})}$). To serve the knowledge transfer purpose, the whole pre-trained model is kept (including L1) to predict the age (Δ_{age}) used to adjust MCI prediction $\mathrm{P(MCI_{conv})'}$.



(a) Pre-trained Model

(b) Fine-tuned Model

Figure 32 Architecture of the proposed AD-Net. 3D boxes represent input and feature maps. The arrows represent network operations: black arrow indicates 3D convolutional operation followed by a rectified linear unit (ReLU) activation function; orange arrow represents max-pooling operations; red arrow represents the flatten operation; dotted red arrow represents fully connected layers; purple square represents the regression outputs for predicted brain age; blue square represents classification outputs for MCI-Converter probability; layers within dotted square forms a building block, and there are 3 repeating blocks (block×3) for feature extraction before flatten layer.

For the AD-Net, in the pre-training procedure, the parameters within 3D blocks, layer L1 and age prediction are trained through the age prediction task. In this procedure, a dataset of 900 3D MRI images from health subjects are used. In the fine-tuning procedure, the parameters within pre-trained network are kept fixed, 200 MRI 3D images from MCI patients are used to tune only parameters within L2 layer to transfer features learned by age prediction task for the MCI-converter prediction task with. In addition, Δ_{age} is incorporated in the fine-tuning procedure.

5.2.2 Aging adjustment in fine-tuning procedure

Given the 3D image I_i for a specific MCI patient i, AD-NET outputs the risk of the patient to be an MCI-converter or a non-converter, denoted as $P(MCI_{conv}^{i})$ and $P(MCI_{non-conv}^{i})$. We have

$$P(MCI_{conv}^{i}) + P(MCI_{non-conv}^{i}) = 1 \tag{5.1}$$

For patient *i*, the chronological age (C-Age) y_{age}^{i} is available. One output from the AD-NET is biological age (B-Age) prediction, that is, $\hat{y}_{age}^{i}$. The difference between predicted B-Age and C-Age is Δ_{age}:

$$\Delta_{age}^{\mathrm{i}} = \hat{y}_{age}^{i} - y_{age}^{i} \tag{5.2}$$

Under the assumption that Δ_{age} is strongly positively correlated to the risk of developing brain disease (Cole et al., 2017) , we adjust the probability of a MCI subject i converting to AD ($P(MCI_{conv}^{i})$) with $\mathit{\Delta}_{age}^{i}$. The basic idea is, for any subject i, (1) if the predicted B-Age is greater than its C-Age, that is, $\Delta_{age}^{\mathrm{i}} > 0$, this subject has a higher risk to convert to AD. We will increase the MCI conversion probability ($P(MCI_{conv}^{i})$) with respect to the magnitude of $\mathit{\Delta}_{age}^{i}$; (2) If the predicted B-Age is less than its C-Age, that is, $\Delta_{age}^{\mathrm{i}} < 0$, this subject will has less risk to convert to an AD. We decrease $P\left(MCI_{conv}^{i}\right)$ accordingly. To model this idea, we have

$$P'(MCI_{conv}^{i}) = \frac{\left(0.5 + w^{i}r\right)P(MCI_{conv}^{i})}{(0.5 + w^{i}r)P(MCI_{conv}^{i}) + (0.5 - w^{i}r)P(MCI_{non-conv}^{i})} \tag{5.3}$$

where,

$$w^{i} = \frac{1}{2m} max\ (-m, min\ (\mathit{\Delta}_{age}^{i}, m)\ )$$

m: pre-defined normalizer to filter outlier impact *r*: correlation between all Δ_{age} and MCI-Converter labels.

In equation (3), each subject's $P'(MCI_{age}^{i})$ is obtained by adjusting the $P(MCI_{conv})$ with respect to two scalars: a global scalar *r* and a subject-dependent scalar w^{i}. During the cross-validation process, for the training folds, we have the Δ_{age} and patient status (MCI converter vs. non-converter). Global scalar *r* is derived as the correlation between the Δ_{age} with the patient status, where $r \in [-1, 1]$. A total positive linear correlation exists for *r* being 1, and total negative linear correlations for *r* being -1, no correlation for *r* being 0. In this study, we would expect to have *r* being positive value to describe the general relationship between the Δ_{age} and the patient status on the group bases. Scalar w^{i} is to measure normalized deviation level of Δ_{age}^{i} for subject *i*. w^{i} is proportional to Δ_{age}^{i}, and it is normalized to the range of -0.5 to 0.5 by a pre-defined normalizer *m*. We adopt *m* here to avoid potential issue from outliers with extreme large Δ_{age}.

To better illustrate the effects of w^{i} and *r* in adjusting $P'(MCI_{age}^{i})$ in equation (3), we plot 4 curves for $\mathrm{P(MCI_{conv})'}$ vs. $\mathrm{P(MCI_{conv})}$ under different settings of w^{i} and *r* (see Figure 33). Here we only discuss the scenario where *r* is positive (same holds true when *r* is negative), and w^{i} can be both negative and positive. From Figure 2, we observe three properties:

(1) Under the same setting of *r*, for positive $w^{i}(w^{i} > 0)$, $P'(MCI_{conv})$ increases as w^{i} increases. The larger the $w^{i}(w^{i} > 0)$ is, the greater adjustment made from $P(MCI_{conv})$ to $P'(MCI_{conv})$. For negative $w^{i}(w^{i} < 0)$, $P'(MCI_{conv})$ decreases as w^{i} decreases. The smaller the w^{i} (w^{i} <0) is, the greater adjustment made from

$P(MCI_{conv})$ to $P'(MCI_{conv})$. This is consistent with our earlier discussion, that is, w^i is proportional to Δ^i_{age} and Δ^i_{age} is positively correlated with the AD conversion risk.

(2) Under the same setting of w^i, larger the *r* is, the greater adjustment made from $P(MCI_{conv})$ to $P'(MCI_{conv})$. This is a desirable property since *r* measures the correlations between Δ_{age} and MCI conversion risk. The larger the *r* is, the higher risk one would convert to AD.

(3) The adjustment has more effects for subjects with $P(\mathrm{MCI}_{\mathrm{conv}})$ falling in the middle of the distribution (e.g., 0.4 - 0.6) than that at the two sides (e.g. 0-0.1 and 0.9-1.0). We believe this is a desirable property indicating the adjustments can strengthen the differentiation power for the subjects who were not certain on determining the conversion risks.

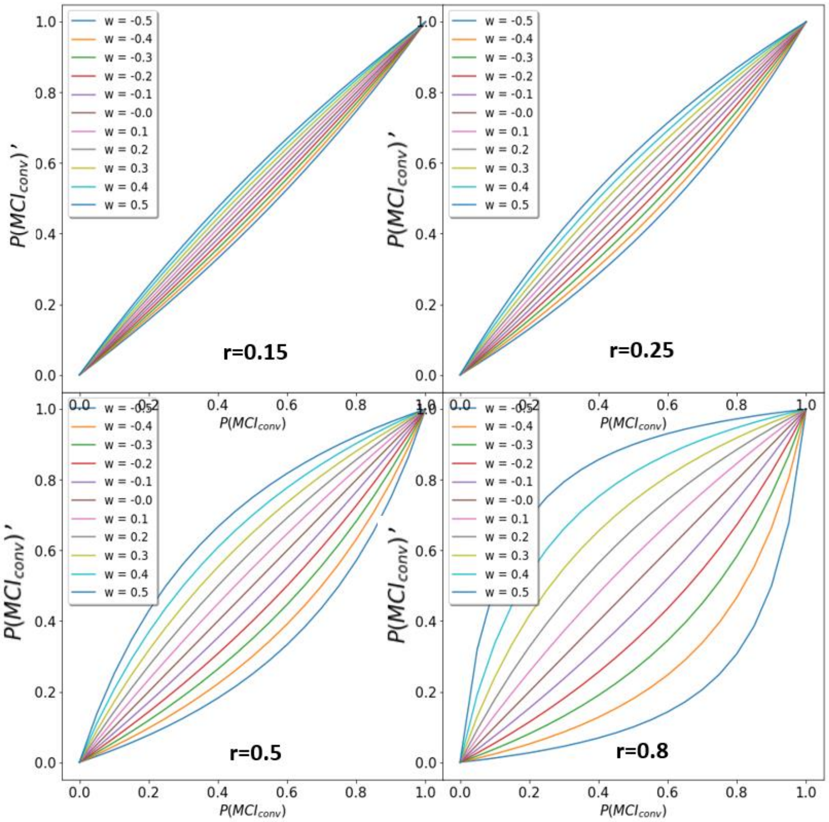


Figure 33 Curves for $P(MCI_{conv})'$ vs. $P(MCI_{conv})$ under different settings of w^i and r.

In the fine-tuning model, the age-related information from the pre-training is transferred. Together with the features from the pre-training model, the risk of the subject converting to AD is predicted. A comprehensive comparison experiment is conducted and is discussed in the next section.

5.3 Dataset and Image Pre-processing

All neuroimaging data used in the study are T1-weighted MRI. The datasets used in pre-training and fine-tuning procedure are obtained from different cohorts, and we conducted pre-processing procedure to ensure consistency among images from different cohorts.

5.3.1 Data

5.3.1.1 Dataset I for age prediction

The dataset used in pre-training procedure for age prediction task includes 847 subjects (male/female = 395/452, mean age = 56.86 ± 18.34, age range 18–94 years). Among the whole dataset, 253 are healthy controls obtained from Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset (F. Li & Liu, 2018)), the ages range from 56-89. The ADNI is launched aiming at finding the relationship between progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD) and biomarkers, magnetic resonance imaging (MRI), positron emission tomography (PET) or clinical and neuropsychological assessments. ADNI enrolls a large cohort (>800) of participants (Weiner et al., 2015), for each subject, PET, MRI images, as well as clinical information (including age) are available. The selected 253 subjects include all healthy subject in ADNI dataset.

In order to increase the size of training dataset and widen the age range to ensure a more accurate and robust age prediction model, we obtain additional 581 healthy subjects from Information eXtraction from Images (IXI) public dataset (“IXI Dataset,” n.d.). The subjects from IXI dataset are obtained from 3 different hospitals in London: Hammersmith Hospital (Philips 3T system), Guy’s Hospital (Philips 1.5T system) and Institute of Psychiatry (GE 1.5T system). For each subject, personal information such as sex, height, weight, occupation and age are included.

5.3.1.2 Dataset II for MCI-conversion prediction

The dataset used in fine-tuning procedure for MCI conversion prediction task is obtained from ADNI dataset, we exclude some special MCI cases who returns to normal stage. As a result, all subject has the status as being either converter or non-converter. The dataset includes a total of 297 subjects (male/female = 121/172, mean age = 74.62±7.30, age range 55–88 years). These 297 subjects are diagnosed as MCI when their first image is obtained (baseline diagnosis). Among the 297 subjects, 168 are MCI-Converters and the rest 129 subjects are MCI non-converter. The MCI-converter and MCI non-converter subjects are labeled through the following logic: a subject is labeled as MCI-converter if the subject was diagnosed as MCI and converted to AD during a three-year follow-up; and a subject is labeled as MCI non-converter if the subject was diagnosis as MCI at both baseline and 36 months. Those subjects whose diagnosis was missing at 36 months were excluded in the dataset.

5.3.2 Pre-processing

We convert DICOM files to Nifti format and register the raw Nifti files to MNI152 (VS Fonov, AC Evans, RC McKinstry, CR Almli, 2009) space to ensure consistency of position and orientation. The images were resampled using cubic spline interpolation, to transfer data acquired from different studies into the same voxel sizes and dimensions (1mm^3, 182×218×182). Examples of the different data used in the study are shown in Figure 34.

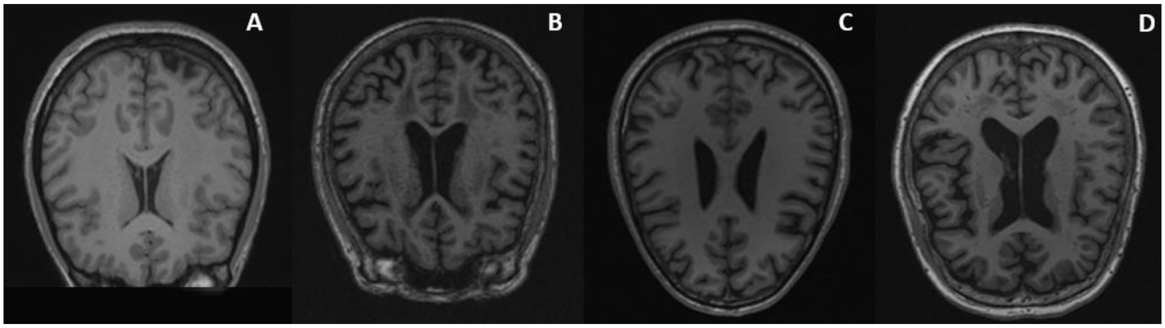


Figure 34 Examples input T1-weighted MRI imaging after the minimal pre-processing procedure. A) healthy subject from IXI dataset. B) healthy subject from ADNI dataset. C) MCI Non-Converter subject from ADNI dataset. D) MCI-Converter subject from ADNI dataset.

5.4 Experimental Results

5.4.1 Experiment I: Pre-training and Age Prediction Task

In this experiment, 84 (10%) subjects are randomly selected from dataset I as blind testing dataset, the remaining 763 subjects are used as training dataset. The proposed AD-NET is trained using mean squared error (MSE) as loss function, Adam (Kingma & Ba, 2014) is used as the optimizer to solve the problem. The parameter settings are: learning rate is 0.01; learning rate decay equals to 0.005; training batch is 16 and training iteration is 200. The model achieves MSE of 187.16 and mean absolute error (MAE) of 11.17 on the training dataset. The Pearson correlation (pc) between C-Age and predicted B-Age is 0.75. On the testing dataset, we have MSE=196.42, MAE=12.28, pc=0.67. For illustration purpose, we include the plot of C-Age vs. predicted B-Age for both training dataset and testing dataset in Figure 35. From the result and figure, we conclude that after pre-training, the AD-NET for age prediction can learn the mapping between raw MRI image and C-Age with good accuracy among healthy subjects.

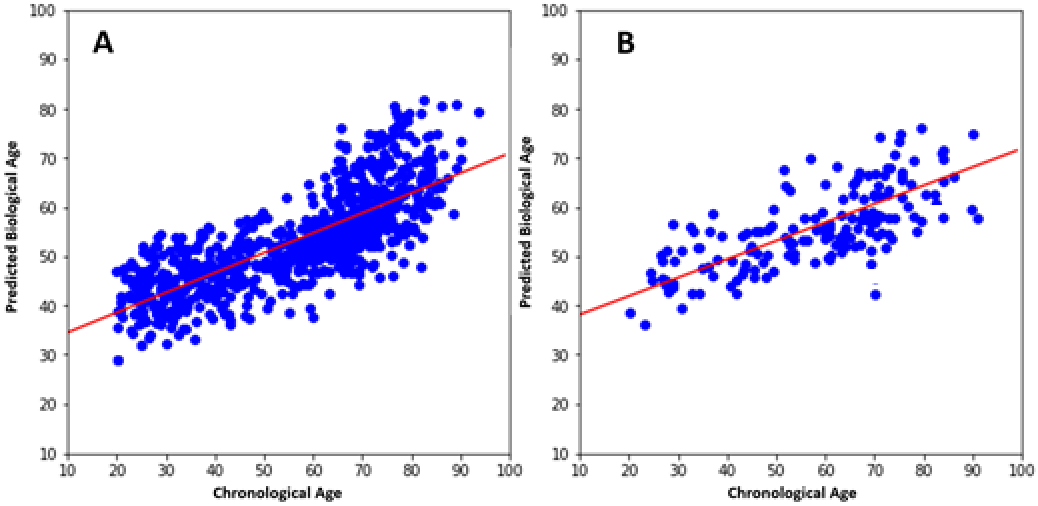


Figure 35 Plot of chronological age (C-Age) vs. predicted biological age (B-Age): A) training dataset B) testing dataset. Red lines are the fitted linear regression respectively.

We do recognize that the model performance maybe not optimal compared with (Cole et al., 2017) and there is potential space for improvement. Given the focus of this study is to demonstrate the advantages of surrogate biomarker from age for MCI converter prediction, we decide to leave the age prediction model improvement as a future research effort. Here we feed all subjects in dataset II into the pre-trained model and obtained predicted age for each MCI subject. Figure 36 shows the plot of C-age vs. predicted B-Age for all subjects in dataset II.

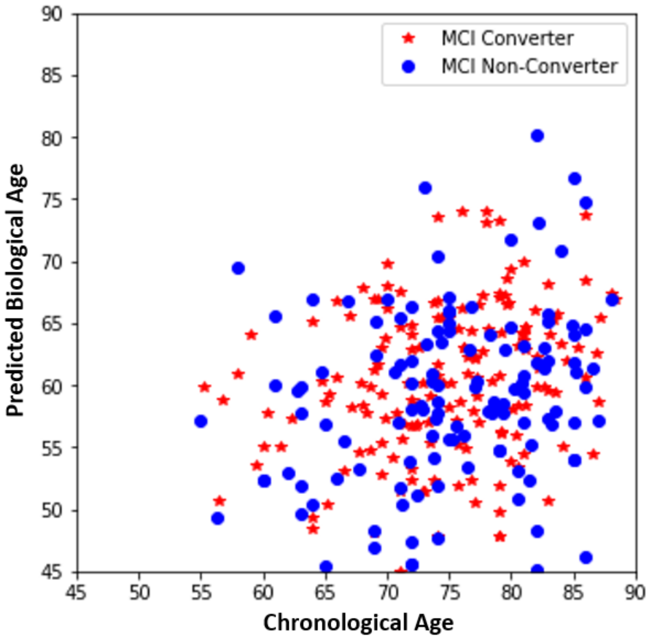


Figure 36 Plot of chronological age (C-Age) vs. predicted biological age (B-Age for subjects in dataset II.

The Δ_{age} for each subject is derived using equation (2). We conduct t-test between the MCI-converter and MCI non-converter groups. The p-value is 0.021, indicating the significant difference on Δ_{age} between MCI-converter group and MCI non-converter group. Next, we determine the hyper-parameters settings for equation (3). The distribution of different Δ_{age} values is shown in Figure 37. It should be noted that the mean Δ_{age} for all subjects in dataset II is -16.64, this is because of the bias from the pre-trained model on healthy subjects. Here we do observe the group difference in the distribution: the MCI-Converter groups tend to have more subjects with larger Δ_{age} values compared with MCI Non-Converter group. Ideally, we would like the mean Δ_{age} close to zero, to utilize positive or negative symbol of Δ_{age} value as direction to increase or decrease the value of $P(MCI_{conv})$. Here we subtract average Δ_{age} (-16.64) from Δ_{age} of each individual subject to normalize. From Figure 37, we observe that the distribution of Δ_{age} follows Gaussian distribution, it is common practice to use n times standard deviation to exclude outliers (Ben-Gal, 2005). In this experiment, we follow the same practice, and set m is to be 17, which is 2 (n=2) times of normalized Δ_{age} standard deviation. r is set to be 0.15, which is the Pearson correlation (Benesty, Chen, Huang, & Israel Cohen, 2009) between Δ_{age} and MCI-Converter labels (p=0.04).

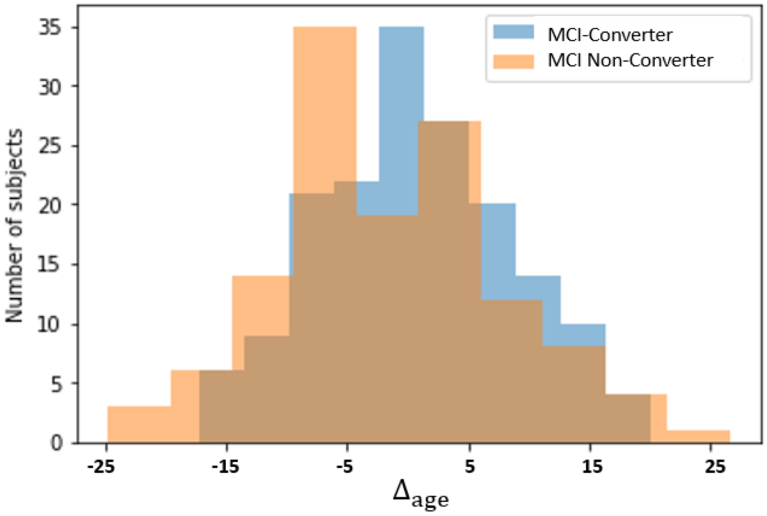


Figure 37 Distribution normalized Δ_{age} values for MCI-Convertor and MCI Non-Converter groups

In this experiment, the accuracy of AD-NET in age prediction task and the potential of biomarker Δ_{age} in differentiating MCI-converter vs. MCI non-converter are both validated. Next, we conduct the second experiment on MCI conversion prediction.

5.4.2 Experiment II: MCI-Converter Prediction Task

In this experiment, 5-fold cross-validation is conducted to evaluate AD-NET's performance on MCI-Converter prediction problem. The parameters obtained from pre-training procedure are kept fixed, in order to get stable age prediction from AD-NET. Parameters within L2 layer are trained to make proper feature transformation from age prediction to MCI converter prediction. In this experiment, the proposed AD-NET is fine-tuned using cross-entropy as loss function and Adam optimizer (Kingma & Ba, 2014). Other parameters are selected based on the best performance: learning rate is 0.01; learning rate decay equals to 0.005; training batch is 16 and training iteration is 50. Area under receiver operating characteristic curve (AUC), accuracy (ACC.), sensitivity (SEN.) and specificity (SPE.) are calculated to measure the prediction power of our model from different aspects.

For comparison purpose, we implement two competing methods, which are pre-trained through the same procedure as AD-NET: Transfer learning CNN model (TL-CNN) and Transfer learning CNN model with Δ_{age} as additional features (TL-CNN-Δ_{age}). The architecture of TL-CNN is the same as our-proposed AD-NET, the only difference is that during the fine-tuning procedure, neither C-Age information nor predicted B-Age from pre-training procedure is included. This deep learning architecture is well-studied in a number of medical image applications such as age prediction (Cole et al., 2017), breast cancer classification (Gao et al., 2018) and medical imaging synthesis (R. Li et al., 2014). This competing method is selected to validate the novelty of AD-NET in adding Δ_{age} as a biomarker to provide additional information for improved classification performance. In TL-CNN-Δ_{age}, the Δ_{age} for each subject is calculated the after the pre-training procedure. During the fine-tuning procedure, it is added as one single input along with last layer (layer L2 in Figure 1). This competing method is selected to validate the novelty of AD-NET in adjusting $\text{P}(\text{MCI}_{\text{conv}})$ with Δ_{age}. In addition, six existing methods from literature using the same ADNI dataset are chosen for comparison, both traditional machine leaning models (e.g. logistic regression and SVM) and deep learning models are included. The detailed results of all eight methods are included in Table 17.

Table 17 AUC values for AD-NET and competing methods.

Methods	**Data**	**AUC**	**Acc.**	**Sen.**	**Spe.**	**Cate.**
Logistic/Cox regression (Ewers et al., 2012)	Structural MRI+CSF+ Neuropsychological testing	N.A.	0.77	**0.82**	0.73	ML
Orthogonal partial least squares (Westman et al., 2012)	Structural MRI + CSF	0.76	0.69	0.74	0.63	ML
Gaussian Process (Young et al., 2013)	Structural MRI + CSF + PET + APOE	0.80	0.74	0.79	0.66	ML
SVM (F. Liu, Wee, Chen, & Shen, 2014)	Structural MRI + PET	0.70	0.68	0.65	0.70	ML
SAE + Logistic regression (S. Liu et al., 2015)	Structural MRI + PET	N.A.	0.54	0.52	**0.87**	ML
Deep polynomial network +SVM (Shi et al., 2018)	Structural MRI	0.80	**0.79**	0.68	**0.87**	**DL**
TL-CNN (Cole et al., 2017)	Structural MRI	0.76 ±0.06	0.73 ±0.04	0.68 ±0.09	0.77 ±0.09	DL
TL-CNN-Δ_{age}	Structural MRI + Age	0.77 ±0.05	0.77 ±0.02	0.80 ±0.04	0.73 ±0.05	DL
AD-NET	Structural MRI + Age	**0.81 ±0.05**	0.76 ±0.03	0.77 ±0.07	0.76 ±0.09	DL

From this table, we have four conclusions. First, traditional machine learning models usually require additional information (e.g. clinical testing scores, APOE) to achieve comparable performance as deep learning models which takes only images data. We conclude this demonstrates the advantage of deep learning models. Second, with Δ_{age} added, the TL-CNN- Δ_{age} marginally outperforms TL-CNN in terms of overall performance metrics (accuracy and AUC). This demonstrates the advantage of Δ_{age} as a surrogate marker for the MCI conversion prediction problem. However, TL-CNN-Δ_{age} underperforms the other two deep learning models which are specifically designed in architecture and enhanced with traditional models (logistic regression and SVM). Third, AD-NET and TL-CNN-Δ_{age} achieve improved sensitivity compared with other deep leaning models, which is more desirable in clinical application since sensitivity is more important than specificity (in this study, early detect the converter for effective interventions). However, TL-CNN-Δ_{age} sacrifices specificity while the proposed AD-NET achieves a comparable specificity as TL-CNN. One reason may be, in dataset II, there are several MCI-Converter subjects with larger positive Δ_{age} (Higher B-Age than C-Age). The Δ_{age} helps differentiate such subjects from MCI Non-Converter subjects (increase sensitivity). However, for both MCI-Converter and MCI Non-Converter subjects, they all unlikely to have younger C-Age than their C-Age, since they have already been diagnosed with cognitive impairment. As result, the specificity is not improved. Last and most importantly, our proposed AD-NET outperforms all competing methods in terms of AUC, which is a robust metric in the medical researches, and it is more consistent and have better discriminatory power comparing to accuracy. This can be explained through the meaning of AUC and logic behind age adjustment procedure: since we are using Δ_{age} as a prior knowledge to gain more confidence on MCI-converter vs. MCI non-converter classification by increasing or decreasing the corresponding probabilities, thus improving the model's discriminatory power, especially for the cases which the original model is uncertain with.

5.5 Discussion and Conclusion

Alzheimer's disease (AD) is one of the most common progressive neurodegenerative diseases in elderly patients. It is critical for AD being detected early so that more effective intervention can be conducted. Mild cognitive impairment (MCI), a pre-dementia stage, has been of great interest in both AD research and clinical practices as MCI patients have higher risk of progression to AD. This calls deep investigation to classify the MCI patients to be a converter vs. a non-converter. To address this problem, different biomarkers are proposed by researchers from both predictive modeling and medicine domain, trying to quantify the disease from different aspects. Moreover, researchers have introduced deep learning to this area, with the hope to take the advantages of its powerful classification and feature extraction capability.

In this study, to address challenging problem of MCI conversion prediction, we propose an AD-NET (Age-adjust neural network) to study the applicability of transfer learning and biomarker Δ_{age} to improve the MCI-Converter prediction problem. One contribution of this study is to transfer learning the knowledge captured in the pre-training to the fine-tuning procedure. The knowledge-based transfer learning not only saves training resources but also improves prediction accuracy. Our second contribution lies in proposing a novel age adjust procedure where Δ_{age} is introduced as a risk factor for MCI converter prediction. With these contributions, our proposed model AD-NET achieved the best AUC of 0.81 (±0.05) compared with all eight competing models. As for the future direction, we plan to further improve the age prediction results with more training dataset and different parameter settings. We expect these explorations will further improve the MCI conversion predictions. In addition, we plan to expand our proposed architecture to other clinical applications (e.g., migraine prediction), and with different imaging modalities (e.g., MR and PET).

CHAPTER 6

CONCLUSIONS AND FUTUER WORK

The overall objective of this research is to develop novel deep learning models for various medical imaging applications. In Chapter 2, I develop a Shallow-Deep Convolutional Neural Network (SD-CNN) with demonstrated performance improvement of mass classification for breast mammography by the combination of a pre-trained deep CNN architecture and synthetic advanced imaging modality (CEDM). Motivated by the success, I propose an advanced deep learning architecture named encoder-decoder residual inception network (REID-Net) to further extend the application of image synthesis on complete image instead of extracted patches in Chapter 3. Its capability of imaging synthesizing is demonstrated with digital mammography and neuro imaging datasets. In Chapter 4, I focus on addressing multiple tasks together through deep multi-task learning and improving the performance of individual task within an MTL architecture by referring features as additional information from parallel task. The proposed feature transferring multi-task learning network (FT-MTL-Net) is evaluated with digital mammography data on tasks of breast cancer detection, segmentation and classification. Transferred features from segmentation task help the proposed model obtain improved classification. Finally, in Chapter 5, I focus on applying transfer learning in the training procedure of deep learning models. The novity of my proposed age-adjustment neural network (AD-Net) lies in the transfer of both features and knowledge from pre-training task to the fine-tuning task aiming at reducing computation cost and improving the model's performance in fine-tuning task. The advantage of this model is demonstrated in the task of MCI to AD conversion prediction task.

For the further work, I would like to consider an extension of FT-MTL-Net to enable feature transferring between different tasks. As in the initial study of Chapter 5, the current model obtained improved classification capability in the single task by taking segmentation features; its performance in segmentation or detection should also be improved if additional features are introduced. With feature transferring between multiple tasks, the model's performance on multiple tasks should be improved simultaneously. Lastly, we introduced AD-Net in Chapter 6, in this model predicted age obtained from pre-training task is introduced as addition knowledge to the fine-tuning MCI conversion prediction task through a proposed equation. The equation is proposed based on our prior knowledge about predicted age and MCI conversion, however such kind of prior knowledge is not always clear; we may dive deep into the methods which enable the model to learn such knowledge automatically.

REFERENCES

Affonso, C., Renato, J. S., & Marques, B. R. (2015). Biological image classification using rough-fuzzy artificial neural network. Expert Systems with Applications, 42(24), 9482–9488.

Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. International Journal of Medical Informatics, 117(April), 44–54.

Al-masni, M.A., Al-antari, M.A., Park, J.M., Gi, G., Kim, T.Y., Rivera, P., Valarezo, E., Han, S.M. & Kim, T.S. (2017). Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS (pp. 1230–1233).

Akselrod-ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., & Barkan, E. (2016). A Region Based Convolutional Network for Tumor Detection and Classification in Breast Mammography. Deep Learning and Data Labeling for Medical Applications., 197–205.

Araujo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., … Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. PLoS ONE, 12(6), 1–14.

Augusto, G. B. (2014). Multiple Kernel Learning for Breast Cancer Classification.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.

Bar, Y., Diamant, I., Wolf, L. and Greenspan, H., 2015, March. Deep learning with non-medical training used for chest pathology identification. In Medical Imaging 2015: Computer-Aided Diagnosis (Vol. 9414, p. 94140V). International Society for Optics and Photonics..

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., & Filippi, M. (2019). Automated classification of Alzheimer’s disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage: Clinical, 21(December 2018), 101645.

Bauer, E., Kohavi, R., Chan, P., Stolfo, S., & Wolpert, D. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. Machine Learning, 36(August), 105–139.

Baxter, J. (1997). A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling. Machine Learning, 28(1), 7–39.

Ben-Gal, I. (2005). Outlier detection. Springer.

Benesty, J., Chen, J., Huang, Y., & Israel Cohen. (2009). Pearson correlation coefficient. Noise Reduction in Speech Processing, 1–4.

Bulten, W., Hulsbergen-van de Kaa, C. A., van der Laak, J., & Litjens, G. J. (2018, March). Automated segmentation of epithelial tissue in prostatectomy slides using deep learning. In Medical Imaging 2018: Digital Pathology (Vol. 10581, p. 105810S). International Society for Optics and Photonics.

Carneiro, G., Nascimento, J., & Bradley, A. P. (2017). Automated Analysis of Unregistered Multi-View Mammograms with Deep Learning. IEEE Transactions on Medical Imaging, 36(11), 2355–2365.

Carrillo, M. C., Dean, R. A., Nicolas, F., Miller, D. S., Berman, R., Khachaturian, Z., … Knopman, D. (2013). Revisiting the framework of the National Institute on Aging-Alzheimer’s Association diagnostic criteria. Alzheimer’s and Dementia, 9(5), 594–601.

Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41–75.

Castro, M., & Smith, G. E. (2015). Mild cognitive impairment and Alzheimer’s disease. In APA handbook of clinical geropsychology, Vol. 2: Assessment, treatment, and issues of later life. (pp. 173–207). Washington: American Psychological Association

Cha, K. H., Hadjiiski, L., Samala, R. K., Chan, H.-P., Caoili, E. M., & Cohan, R. H. (2016). Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. Medical Physics, 43(4), 1882–1896.

Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., … Wang, G. (2017). Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN). IEEE Transactions on Medical Imaging, 36(12), 2524–2535.

Cheng, D., Liu, M., Fu, J., & Wang, Y. (2017, July). Classification of MR brain images by combination of multi-CNNs for AD diagnosis. In Ninth International Conference on Digital Image Processing (ICDIP 2017) (Vol. 10420, p. 1042042). International Society for Optics and Photonics.

Cheung, Y. C., Lin, Y. C., Wan, Y. L., Yeow, K. M., Huang, P. C., Lo, Y. F., … Chang, C. J. (2014). Diagnostic performance of dual-energy contrast-enhanced subtracted mammography in dense breasts compared to mammography alone: interobserver blind-reading analysis. European Radiology, 24(10), 2394–2403.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., ... & Colliot, O. (2009). Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus, 19(6), 579-587.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016, October). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.

Ciompi, F., de Hoop, B., van Riel, S. J., Chung, K., Scholten, E. T., Oudkerk, M., … van Ginneken, B. (2015). Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Medical Image Analysis, 26(1), 195–202.

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D., & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage, 163(July), 115–124.

Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.

Deng, L., Hinton, G., & Kingsbury, B. (2013, May). New types of deep neural network learning for speech recognition and related applications: An overview. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 8599-8603). IEEE.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2015, November). Automated mass detection in mammograms using cascaded deep learning and random forests. In 2015 international conference on digital image computing: techniques and applications (DICTA) (pp. 1-8). IEEE.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017a). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Medical Image Analysis, 37, 114–128.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017b). Fully automated classification of mammograms using deep residual neural networks. In Proceedings of IEEE 14th international symposium on biomedical imaging, 310-314.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. Ecology.

Diz, J., Marreiros, G., & Freitas, A. (2016). Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. Journal of Medical Systems, 40(9).

Domingues, I., Sales, E., Pereira, W. C. A., Tec, I., Faculdade, P., Porto, U., … Janeiro, R. De. (2012). Inbreast-Database Masses Characterization, (January), 1–5.

Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., … Heng, P. A. (2016). Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks. IEEE Transactions on Medical Imaging, 35(5), 1182–1195.

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2017). The importance of skip connections in biomedical image segmentatio. Deep Learning and Data Labeling for Medical Applications, 2, 179–187.

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., … Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. The Lancet Neurology, 13(6), 614–629.

Eberl, M. M., Fox, C. H., Edge, S. B., Carter, C. a, & Mahoney, M. C. (2015). BI-RADS classification for management of abnormal mammograms. Journal of the American Board of Family Medicine, 19(2), 161–164.

Eigen, D., Rolfe, J., Fergus, R., & LeCun, Y. (2013). Understanding deep architectures using a recursive convolutional network. arXiv preprint arXiv:1312.1847.

Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). CLINICIAN ' S CORNER Screening for Breast Cancer. Journal of the American Medical Association, 293(10), 1245–1256.

Eskildsen, S. F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J. C., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2013). Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage, 65, 511-521.

Ewers, M., Walsh, C., Trojanowski, J. Q., Shaw, L. M., Petersen, R. C., Jack Jr, C. R., ... & Vellas, B. (2012). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiology of aging, 33(7), 1203-1214.

Fakhry, A., Zeng, T., & Ji, S. (2016). Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation. IEEE Transactions on Medical Imaging, 0062(c), 1–1.

Fallenberg, E. M., Dromain, C., Diekmann, F., Engelken, F., Krohn, M., Singh, J. M., … Renz, D. M. (2014). Contrast-enhanced spectral mammography versus MRI: Initial results in the detection of breast cancer and assessment of tumour size. European Radiology, 24(1), 256–264.

Feng, Z., Nie, D., Wang, L., & Shen, D. (2018, April). Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 885-888). IEEE.

Filippi, M., Agosta, F., Barkhof, F., Dubois, B., Fox, N. C., Frisoni, G. B., ... & Scheltens, P. (2012). EFNS task force: the use of neuroimaging in the diagnosis of dementia. European Journal of Neurology, 19(12), 1487-1501.

Fox, N. C., Cousens, S., Scahill, R., Harvey, R. J., & Rossor, M. N. (2000). Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. Archives of neurology, 57(3), 339-344.

Francescone, M. A., Jochelson, M. S., Dershaw, D. D., Sung, J. S., Hughes, M. C., Zheng, J., … Morris, E. A. (2014). Low energy mammogram obtained in contrast-enhanced digital mammography (CEDM) is comparable to routine full-field digital mammography (FFDM). European Journal of Radiology, 83(8), 1350–1355.

Gao, F., Wu, T., Chu, X., Yoon, H., Xu, Y., & Patel, B. (2019). Deep Residual Inception Encoder-Decoder Network for Medical Imaging Synthesis. IEEE Journal of Biomedical and Health Informatics.

Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., & Patel, B. (2018). SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Computerized Medical Imaging and Graphics, 70, 53–62.

Gao, F., Zhang, M., Wu, T., & Bennett, K. M. (2016). 3D small structure detection in medical image using texture analysis. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 6433–6436.

Gaugler, J., James, B., Johnson, T., Scholz, K., & Weuve, J. (2016). 2016 Alzheimer’s disease facts and figures. Alzheimer’s and Dementia.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., ... & Cummings, J. L. (2006). Mild cognitive impairment. The lancet, 367(9518), 1262-1270.

Gillman, J., Toth, H. K., & Moy, L. (2014). The Role of Dynamic Contrast-Enhanced Screening Breast MRI in Populations at Increased Risk for Breast Cancer. Women's Health, 10(6), 609–622.

Girshick, R. (2015). Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1440–1448).

Greenspan, H., Ginneken, B. van, & Summers, R. M. (2016). Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Transactions on Medical Imaging, 35(5), 1153–1159.

Han, X. (2017). MR-based synthetic CT generation using a deep convolutional neural network method: Medical Physics, 44(4), 1408–1419.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., … Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis, 35, 18–31.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2980–2988).

He, K., & Sun, J. (2015). Convolutional Neural Networks at Constrained Time Cost. In IEEE conference on computer vision and pattern recognition.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., Babajani-Feremi, A., & Alzheimer's Disease Neuroimaging Initiative. (2017). Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM. Journal of neuroscience methods, 282, 69-80.

Hon, M., & Khan, N. M. (2017, November). Towards Alzheimer's disease classification through transfer learning. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1166-1169). IEEE.

Hosseini-Asl, Ghazal M, Mahmoud A, Aslantas A, Shalaby AM, Casanova MF, Barnes GN, Gimel'farb G, Keynton R, E.-B. A. (2018). Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. Frontiers in Bioscience.

Hu, K., Wang, Y., Chen, K., Hou, L., & Zhang, X. (2016). Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. Neurocomputing, 175, 132-145.

Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging, 3(3), 034501.

IXI Dataset. Retrieved April 6, 2019, from https://www.nitrc.org/projects/ixi_dataset/

Jorstad, A. & Fua, P., 2014, September. Refining mitochondria segmentation in electron microscopy imagery with active surfaces. In Proceedings of european conference on computer vision (pp. 367-379). Springer, Cham.

Khatami, A., Babaie, M., Tizhoosh, H. R., Khosravi, A., Nguyen, T., & Nahavandi, S. (2018). A sequential search-space shrinking using CNN transfer learning and a radon projection pool for medical image retrieval. Expert Systems with Applications, 100, 224–233.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. ArXiv Preprint ArXiv:1412.6980.

Kozegar, E., Soryani, M., Minaei, B., & Domingues, I. (2013). Assessment of a novel mass detection algorithm in mammograms. Journal of Cancer Research and Therapeutics, 9(4), 592.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems, 1–9.

Lafferty, J., McCallum, A. & Pereira, F.C. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2323.

Lee, H., & Chen, Y. P. P. (2015). Image based computer aided diagnosis system for cancer detection. Expert Systems with Applications, 42(12), 5356–5365.

Lee, G., Nho, K., Kang, B., Sohn, K. A., & Kim, D. (2019). Predicting Alzheimer's disease progression using multi-modal deep learning approach. Scientific reports, 9(1), 1952.

Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., Mahoney, M. C., Iii, W. P. E., … Adams, A. (2012). Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single ScreeningMRI toMammography inWomen With Elevated Breast Cancer Risk, 307(13).

Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. Nature Methods.

Lévy, D., & Jain, A. (2016). Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks, (Nips). arXiv preprint arXiv:1612.00542

Li, F., & Liu, M. (2018). Alzheimer’s disease diagnosis based on multiple cluster dense convolutional networks. Computerized Medical Imaging and Graphics, 70, 101–110.

Li, H. J., Hou, X. H., Liu, H. H., Yue, C. L., He, Y., & Zuo, X. N. (2015). Toward systems neuroscience in mild cognitive impairment and Alzheimer's disease: A meta‐analysis of 75 fMRI studies. Human Brain Mapping, 36(3), 1217-1232.

Li, R., Zhang, W., Suk, H., Wang, L., Li, J., Shen, D., & Ji, S. (2014). Deep learning based imaging data completion for improved brain disease diagnosis. In International Conference on Medical Image Computing and Computer-Assisted Intervention. (pp. 305–312).

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

Liu, F., Wee, C. Y., Chen, H., & Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. NeuroImage, 84, 466-475.

Liu, M., Zhang, J., Adeli, E., & Shen, D. (2018). Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis. IEEE Transactions on Biomedical Engineering, 66(5), 1195-1206.

Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., ... & Fulham, M. J. (2014). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Transactions on Biomedical Engineering, 62(4), 1132-1140.

Liu, T., Xie, S., Yu, J., Niu, L., & Sun, W. (2017, March). Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 919-923). IEEE.

Long, J., Shelhamer, E., & Darrell, T. B. T.-C. V. and P. R. (2015). Fully convolutional networks for semantic segmentation. In IEEE conference on computer vision and pattern recognition. (pp. 3431–3440).IEEE

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., Beg, M. F., & Alzheimer's Disease Neuroimaging Initiative. (2018). Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. Medical image analysis, 46, 26-34.

Luczyńska, E., Heinze-Paluchowska, S., Dyczek, S., Blecharz, P., Rys, J., & Reinfuss, M. (2014). Contrast-enhanced spectral mammography: Comparison with conventional mammography and histopathology in 152 women. Korean Journal of Radiology, 15(6), 689–696.

Michaelson, J. S., Shih, Y. T., Walter, L. C., Church, T. R., Flowers, C. R., Lamonte, S. J., … Otis, W. (2016). Guideline Update from the American Cancer Society, 314(15), 1599–1614.

Microbiana, B., Hidalgo, D., Anthimopoilos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. Ieee Transactions on Medical Imaging, 35(5), 1207–1216.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In In 3D Vision (3DV), 2016 Fourth International Conference on (pp. 565–571).

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: Toward a Full-field Digital Mammographic Database. Academic Radiology, 19(2), 236–248.

Morris, E. A. (2016). Contrast-enhanced digital mammography. Diseases of the Brain, Head and Neck, Spine 2016-2019: Diagnostic Imaging, 69, 339–342.

Muramatsu, C., Hara, T., Endo, T., & Fujita, H. (2016). Breast mass classification on mammograms using radial local ternary patterns. Computers in Biology and Medicine, 72, 43–53.

Nogueira, K., Penatti, O. A., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61, 539-556.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE international conference on computer vision (pp. 1520-1528).

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

Patel, B. K., Garza, S. A., Eversman, S., Lopez-Alvarez, Y., Kosiorek, H., & Pockaj, B. A. (2017). Assessing tumor extent on contrast-enhanced spectral mammography versus full-field digital mammography and ultrasound. Clinical Imaging, 46, 78–84.

Pike, K. E., Savage, G., Villemagne, V. L., Ng, S., Moss, S. A., Maruff, P., ... & Rowe, C. C. (2007). β-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. Brain, 130(11), 2837-2844.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Rosenquist, C. J., & Lindfors, K. K. (1998). Screening mammography beginning at age 40 years: a reappraisal of cost-effectiveness. Cancer, 82(11), 2235–2240.

Rosset, A., Spadola, L., & Ratib, O. (2004). OsiriX: An open-source software for navigating in multidimensional DICOM images. Journal of Digital Imaging, 17(3), 205–216.

Roth, H., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., … Summers, R. (2016). Improving Computer-aided Detection using Convolutional Neural Networks and Random View Aggregation. IEEE Transactions on Medical Imaging, PP(99), 1.

Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. ArXiv Preprint ArXiv:1706.05098, (May).

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211–252.

Sahiner, B., Chan, H. P., Petrick, N., Wei, D., Helvie, M. A., Adler, D. D., & Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. IEEE Transactions on Medical Imaging, 15(5), 598–610.

Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Wei, J., & Cha, K. (2016). Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Medical Physics, 43(12), 6654–6666.

Samala, R. K., Chan, H., Hadjiiski, L. M., Helvie, M. A., & Cha, K. H. (2018). Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms. Physics in Medicine & Biology, 62(23), 8894–8908.

Selkoe, D. J. (1997). Alzheimer's Disease--Genotypes, Phenotype, and Treatments. Science, 275(5300), 630–631.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229.

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. IEEE transactions on pattern analysis and machine intelligence. 34(4), 640-651.

Shi, J., Zheng, X., Li, Y., Zhang, Q., & Ying, S. (2017). Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE journal of biomedical and health informatics, 22(1), 173-183.

Shen, S., Han, S., Aberle, D., Bui, A. A., & Hsu, W. (2019). An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. Expert Systems with Applications, 128, 84–95.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv Preprint ArXiv:1409.1556, 1–14.

Sirinukunwattana, K., Raza, S. E. A., Tsang, Y.-W., Snead, D. R. J., Cree, I. A., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE TRANSACTIONS ON MEDICAL IMAGING, 35(5), 1196–1206.

Song, Y., Li, Q., Huang, H., Feng, D., Chen, M., & Cai, W. (2017). Low dimensional representation of fisher vectors for microscopy image classification. IEEE transactions on medical imaging, 36(8), 1636-1649.

Spasov, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., & Alzheimer's Disease Neuroimaging Initiative. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage, 189, 276-287.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958..

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. arXiv preprint arXiv:1505.00387.

Suk, H. I., Lee, S. W., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2014). Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage, 101, 569-582.

Suk, H. Il, & Shen, D. (2013). Deep learning-based feature representation for AD/MCI classification. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-First AAAI Conference on Artificial Intelligence.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Transactions on Medical Imaging, 35(5), 1299–1312.

VS Fonov, AC Evans, RC McKinstry, CR Almli, D. C. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage, 47, S102.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612.

Warner, E., Plewes, D. B., Hill, K. a, Causer, P. a, Jong, R. a, Cutrara, M. R., … Narod, S. a. (2004). Surveillance of BRCA1 and BRCA2 Mutation Carriers With Magnetic Resonance Imaging, Ultrasound, Mammography, and Clinical Breast Examination. Jama, 292(11), 1317–1325.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Nigel, J., Cedarbaum, J., … Trojanowski, J. Q. (2015). Impact of the Alzheimer’s Disease Neuroimaging Initiative, 2004 to 2014. Alzheimer’s & Dementia, 11(7), 865–884.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Nigel, J., Cedarbaum, J., … Trojanowski, J. Q. (2016). Impact of the Alzheimer’s Disease Neuroimaging Initiative, 2004 to 2014. Alzheimer’s & Dementia, 11(7), 865–884.

Westman, E., Muehlboeck, J. S., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage, 62(1), 229-238.

Wu, S., Zhong, S., & Liu, Y. (2017). Deep residual learning for image steganalysis. Multimedia Tools and Applications, 1–17.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1492-1500).

Yamada, T., Hashimoto, R. I., Yahata, N., Ichikawa, N., Yoshihara, Y., Okamoto, Y., ... & Kawato, M. (2017). Resting-state functional connectivity-based biomarkers and functional MRI-based neurofeedback for psychiatric disorders: a challenge for developing theranostic biomarkers. International Journal of Neuropsychopharmacology, 20(10), 769-781.

Yang, W., Chen, Y., Liu, Y., Zhong, L., Qin, G., Lu, Z., … Chen, W. (2017). Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. Medical Image Analysis, 35, 421–433.

Yap, J., Yolland, W., & Tschandl, P. (2018). Multimodal Skin Lesion Classification using Deep Learning. Experimental Dermatology, 0–1.

Ye, J., Farnum, M., Yang, E., Verbeeck, R., Lobanov, V., Raghavan, N., ... & Narayan, V. A. (2012). Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data. BMC neurology, 12(1), 46.

Yoon, H., & Li, J. (2019). A novel positive transfer learning approach for telemonitoring of Parkinson’s disease. IEEE Transactions on Automation Science and Engineering, 16(1), 180–191.

Young, J., Modat, M., Cardoso, M. J., Mendelson, A., Cash, D., Ourselin, S., & Alzheimer's Disease Neuroimaging Initiative. (2013). Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. NeuroImage: Clinical, 2, 735-745.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. Computer Vision–ECCV 2014, 8689, 818–833.

Zhang, J., Xia, Y., Xie, Y., Fulham, M., & Feng, D. D. (2017). Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE journal of biomedical and health informatics, 22(5), 1521-1530.

Zhang, Jun, Liu, M., Wang, L., Chen, S., Yuan, P., Li, J., … Shen, D. (2018). Joint Craniomaxillofacial Bone Segmentation and Landmark Digitization by Context-Guided Fully Convolutional Networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention. (Vol. 3, pp. 720–728).

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage, 108, 214–224.

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Medical Image Analysis, 43, 98–111.

Zhu, W., Lou, Q., Vang, Y. S., & Xie, X. (2017). Deep multi-instance networks with sparse label. In Proceedings of international conference on medical image computing and computer-assisted intervention. (pp. 603-611). Springer, Cham.