LEARNING FOR CLINICAL OUTCOME PREDICTION FROM BIG MEDICAL DATA

by

JIAWEN YAO

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2019

To my parents and my wife, for their endless trust, support and love.

ACKNOWLEDGEMENTS

I would like to first express my sincere thanks to my supervising professor, Dr. Junzhou Huang for his constantly motivating and encouraging me, and also for his valuable advice during my 5-year study at UTA. He brought me into this challenging research field, but also had faith that I would be able to achieve highest of standards. His irreplaceable encouragement and supervision are the main reasons of the successful outcomes of my research. None of the work would have happened without him.

I sincerely express my gratitude to my thesis committee members Dr. Chris Ding, Dr. Jeff (Yu) Lei and Dr. Dajiang Zhu for serving on my committee. I wish to thank their interests in my research and valuable suggestions regarding my early proposal and this final thesis. I am very grateful to have each of them serve in my committees and really enjoy each time when we discuss on my research.

I also want to thank all my colleagues from the Scalable Modeling and Imaging and Learning Lab (SMILE Lab) of Computer Science and Engineering Department - Yeqing Li, Zheng Xu, Ruoyu Li, Xinliang Zhu, Sheng Wang, Zhifei Deng, Chaochao Yan, Ashwin Raju, Mohammad Minhazul Haq and many others. I feel very grateful to meet so many creative and nice people and all of you have been there to support me a lot.

My gratefulness also goes to my hosts and friends during my internships. Thanks Dr. Le Lu for offering me the position as research intern to solve real applications in medical imaging, and discussing with me on cutting edge research problems. Thanks Dr. Daguang Xu and Dr. Dong Yang at NVIDIA for being my mentor during the wonderful three months in Maryland. Thanks Dr. Dakai Jin at PAII for being my

mentor, and Dr. Shun Miao, Dr. Adam P. Harrison, Yirui Wang, Yizhi Chen for all the wonderful memories at Bethesda. I have been learning a lot from you through the collaborations.

Finally, I would like to express my earnest gratitude to my parents who raised me up and support me to have the best possible education. I also would like to give my special thank to my dear wife - Yafeng. Without her extreme supports and countless sacrifices, it would not have been possible to reach this stage in my career. The five years of study at UTA is my journey to meet not only the science, but also the love of my life. I promise to love you, to grow with you, and to have faith in our next journey together.

August 7, 2019

ABSTRACT

LEARNING FOR CLINICAL OUTCOME PREDICTION FROM BIG MEDICAL DATA

Jiawen Yao, Ph.D.

The University of Texas at Arlington, 2019

Supervising Professor: Dr. Junzhou Huang

With the advance of recent technological innovations, nowadays scientists can easily capture and store tremendous amounts of different types of medical data such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), big pathological images and high dimensional cell profiling data. Developing deep learning and machine learning to analyze such large-scale medical data sets for patient health care is an interesting but challenging problem. Inspired by the trend, in this dissertation, we focus on solving real-world problems, like survival analysis on image-omics data and reducing uncertainty from undersampled MRI.

Survival analysis is a crucial tool in the clinical study of cancer patients, as it allows clinicians to make early decisions in treatments. With respect to the problem of survival prediction using pathological image data, we first consider to develop a novel image-based pipeline for lung cancer patients. To deal with the subtype cell detection, we develop a deep learning-based detection approach to detect subtype cell locations in images. The proposed pipeline can extract subtype cellular features and

describe the tissue organization and structures more effectively than standard cellular imaging features.

With respect to the problem of multi-modality integration on image-omics data, the dissertation contributes a novel method for the integration. Previous work have suggested that complementary representation from different modalities provides important information for prognosis. However, due to the large discrepancy between different heterogeneous views, traditional survival models are unable to efficiently handle multiple modalities data as well as learn very complex interactions that can affect survival outcomes in various ways. To overcome these issues, we present a Deep Correlational Survival Model (DeepCorrSurv) for the integration of multi-view data. This results in a more accurate prediction compared with state-of-the-arts methods.

With respect to the problem of directly using Whole Slide Images (WSIs) for survival prediction, the dissertation proposes an attention guided deep multiple instance survival learning. Classical methods focus on manually selecting smaller "patches", which seek to represent the WSIs in order to reduce the computational burden. However, these patches are often unable to completely and properly reflect the patients' tumor morphology. Furthermore, the manual annotation work by medical experts required for these methods can often be infeasible to apply to large scale cancer datasets. State-of-the-art WSI-based survival models train patch-based CNN to learn features and then aggregate patch-level results to patient-level decision. However, those models are trained in a unified manner and the aggregation is not trainable. Our model can solve above issues and yield much better predictions than recent WSI-based learning models. Our results also demonstrate the effectiveness of the proposed method as a recommender system to provide personalized recommendations based on an individual's calculated risk.

Compared with pathological images, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans can be collected in much faster ways and thus are widely used for masses or tumors surveillance. Dynamic magnetic resonance imaging (dMRI) is one very important medical imaging technique that has been widely used for multiple clinical applications. To achieve clinical outcome prediction using dMRI, the reconstruction is a necessary first step as dMRI scans are originally undersampled. Without a high quality of reconstruction, it is impossible for later diagnosis. With respect to the problem of dynamic MRI reconstruction, the thesis contributes an efficient algorithm by solving a primal-dual form of the original problem. The convergence rate of the proposed algorithm can be theoretically proved. It is also very convenient to extend to parallel imaging which is more used in recent days. Extensive experiments on single-coil and multi-coil dynamic MR data demonstrate the superior performance of the proposed method in terms of both reconstruction accuracy and time complexity.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This thesis focuses on developing deep learning and machine learning techniques for the purpose of handling various medical data for prognosis, e.g. clinical outcome prediction, MRI reconstruction, etc.

1.1 Problem Statement

Predicting clinical outcomes, such as death of cancer patients, plays an important role in improving the performance of healthcare system and has huge impacts in precision medicine as lower healthcare costs improve quality of life. Survival analysis is a subfield of statistics where the goal is to model the data where the outcome is the time until the occurrence of an event of interest [2]. If we consider the event of interest as the death of patients and then survival analysis can provide a good solution to predict clinical outcomes.

For a survival problem, death time is known precisely only for those patients who have the event occurred during the study period. For other patients in the study, since we may lose track of them or their time to event is greater than the observation time. Those patients are considered to be **censored** instances in survival analysis. The censored time C may be the time of lost, withdrawn or the termination of the observation. In other words, either survival time T_i or censored time C_i can be observed for any given patient i. In survival analysis, a binary event indicator δ_i is usually used to indicate if a instance or patient is censored, i.e., $\delta_i = 1$ for an uncensored and $\delta_i = 0$ for a censored patient. If we use y_i as the observed time,

$y_i = T_i$ for an uncensored instance and $y_i = C_i$ for a censored instance. Furthermore, in the study, each patient usually comes with medical records or data and we denote those data as X_i. Thus, a given patient i will be represented by a triplet (X_i, y_i, δ_i) in the context of survival analysis. The goal of survival analysis is to estimate the time to death T_j for a new patient j with predictors denoted by X_j. In this dissertation, we consider X as image-omics data and develop several algorithms for survival analysis.

1.2 Motivation

In the current era of big data, nowadays scientists can easily capture and store tremendous amounts of different types of medical data such as 3D CT scans, MRI, big pathological images and high dimensional cell profiling data. Within this context, how to effectively extract knowledge and efficiently exploit those data is still an open problem. In this dissertation, we aim at providing effective deep learning and machine learning approaches for handling various medical data sets on solving real-world problems.

In this dissertation, we explore image-omics data for a very important clinical problem - survival prediction. To diagnose tumor, doctors usually take tumor samples in biopsy procedures like genetic and imaging tests which result in large-scale imaging and omics data which include pathology or radiology images, and genomics, proteomics or metabolomics, captured from the same patient. Those data can be used for tumor diagnosis. Compare with other modalities, pathological images can present tumor growth and morphology in extremely detailed, gigapixel resolution which are extremely useful for tumor prognosis. Tumor microenvironment is a complex milieu that includes not only the cancer cells but also the stromal cells and immune cells. All this "extra" genomic information may muddle results and therefore make molecular analysis a challenging task for cancer prognosis while imaging data can overcome such

issues by providing morphology information [3, 4]. Therefore, a lot of research interests have started to focus on developing image-based survival methods in recent years. Besides image-based survival models, there has been validated that complementary information from different modality from image-omics data can benefit the survival prediction. But since image-omics data are very heterogeneous, how to effectively integrate those multi-modalities data is another challenge that needs to be solved.

Secondly, we investigate the problem of handling Whole Slide Images (WSIs) for diagnosis. The most challenge one is that pathological images in real cancer dataset might be in terabytes (10^{12} pixels) level which makes most models computationally impossible. Another issue is that the label information in survival analysis is only given on patient-level while in the traditional supervised learning, each training instance is typically associated with one label. As the solid tumor may have a mixture of tissue architectures and structures, multiple WSIs from different parts of the patient's tissue are collected for diagnosis. Those terabyte-size large WSIs from one patient will share the survival label which will make the problem more challenging. Instead of handling original WSIs, state-of-the-art survival methods adopted several discriminative patches from manually annotated Region Of Interests (ROIs) and then extracted hand-crafted features for predictions [5, 1, 6]. However, a small set of image tiles cannot completely and properly reflect the patients' tumor morphology. These approaches have very high risks to lose survival-discriminative patterns if only select several tiles from very heterogeneous whole pathological image slides.

Thirdly, we investigate the problem of handling dynamic MRI reconstruction. Dynamic magnetic resonance imaging (dMRI) is an important medical imaging technique but the scanning is inherently a very slow process due to a combination of different constraints such as nuclear relaxation times and peripheral nerve stimulation. Since the speed of acquisition in dynamic MRI has physical limits, there

exists a trade-off between temporal and spatial resolution. Additionally, long scan durations can make patient uncomfortable and also increase the chance of motion artifacts. Hence, many approaches have been proposed to reduce scanning time by requiring partial k-space data for reconstruction instead of full sampling. However, most existing methods still suffer from expensive running time devoted to complex iterative learning. Therefore, we are trying to develop an approach to reduce the computational cost of the running time.

1.3 Our Techniques

Solid tumors are heterogeneous tissues composed of a mixture of cells and have special tissue architectures. However, cellular heterogeneity, the differences in cell types are generally not reflected in molecular profilers or in recent histopathological image-based analysis of lung cancer, rendering such information underused [5]. At first, we present the development of a computational approach in H&E stained pathological images to quantitatively describe cellular heterogeneity from different types of cells. In our work, a deep learning approach was first used for cell subtype classification. Then we introduced a set of quantitative features to describe cellular information. Several feature selection methods were used to discover significant imaging biomarkers for survival prediction. These discovered imaging biomarkers are consistent with pathological and biological evidence. Experimental results on two lung cancer data sets demonstrated that survival models built from the clinical imaging biomarkers have better prediction power than state-of-the-art methods using molecular profiling data and traditional imaging biomarkers.

Second, for the integration topic, we develop a Deep Correlational Survival Model (DeepCorrSurv) for the integration of multi-view data. This is based on the recent study that complementary representation from different modalities provides

important information for prognosis [7, 4]. However, due to the large discrepancy between different heterogeneous views, traditional survival models are unable to efficiently handle multiple modalities data as well as learn very complex interactions that can affect survival outcomes in various ways. In our work, The proposed network consists of two sub-networks, view-specific and common sub-network. To remove the view discrepancy, the proposed DeepCorrSurv first explicitly maximizes the correlation among the views. Then it transfers feature hierarchies from view commonality and specifically fine-tunes on the survival regression task. Extensive experiments on real lung and brain tumor data sets demonstrated the effectiveness of the proposed DeepCorrSurv model using multiple modalities data across different tumor types on small training samples.

Thirdly, for the Whole Slide Pathological Images (WSIs) topic, we take advantage of a deep multiple instance learning to encode all possible patterns from WSIs and view the problems of tumor patient survival analysis in a unified manner. Different from the existing works [8, 9] on learning using key patches or clusters from WSIs, We consider use a trainable attention-based pooling layer for efficient aggregation. Without annotated patch-level labeling, our model yields performance that is much better than state-of-the-art WSI-based survival learning models. More importantly, the proposed approach has good interpretability to locate important patterns that contribute more to predictions. We evaluate our model in its ability to predict patients' survival risks across the lung and colorectal tumor from two large whole slide pathological images datasets. The proposed framework can significantly improve the prediction performances compared with existing state-of-the-arts survival analysis approaches. Results also demonstrate the effectiveness of the proposed method as a recommender system to provide personalized recommendations based on an individual's calculated risk.

In this thesis, for the dynamic MRI topic, we propose an efficient algorithm for dynamic magnetic resonance (MR) image reconstruction. With the total variation (TV) and the nuclear norm (NN) regularization, the proposed TVNNR model can utilize both spatial and temporal redundancy in dynamic MR images. Such prior knowledge can help model dynamic MRI data significantly better than a low-rank or a sparse model alone. However, it is very challenging to efficiently minimize the energy function due to the non-smoothness and non-separability of both TV and NN terms. To address this issue, we propose an efficient algorithm by solving a primal-dual form of the original problem. We theoretically prove that the proposed algorithm achieves a convergence rate of $\mathcal{O}(1/N)$ for N iterations, which is much faster than $\mathcal{O}(1/\sqrt{N})$ by directly applying the black-box first-order method. In comparison with state-of-the-art methods, extensive experiments on single-coil and multi-coil dynamic MR data demonstrate the superior performance of the proposed method in terms of both reconstruction accuracy and time complexity.

1.4 Thesis Overview

Finally, we provide the overview of this thesis in summary. In Chapter 2, we present our pipeline for subtype imaging biomarkers for lung cancer survival prediction. Then, Chapter 3 presents the work to integrate multi-modality image-omics data for survival prediction on lung and brain tumor dataset. Chapter 4 shows the proposed deep multi-instance survival learning framework to process WSIs. Chapter 5 shows the proposed algorithm for dynamic MRI reconstruction which is a key step to acquire high-quality MRI images from under-sampled data for the further diagnosis purpose.

As the ending, Chapter 6 draws our conclusions of the thesis, where we summarize the presented deep learning techniques, highlight their contributions and provide future research directions for medical imaging applications.

CHAPTER 2

IMAGING BIOMARKER DISCOVERY FOR LUNG CANCER SURVIVAL PREDICTION

This chapter investigates the problem of survival prediction using pathological images. A novel pipeline is proposed to extract subtype cellular imaging biomarkers for prediction [5]. Experimental results on two lung cancer data sets demonstrated that survival models built from the clinical imaging biomarkers have better prediction power than state-of-the-art methods using molecular profiling data and traditional imaging biomarkers.

2.1 Introduction

Lung cancer is the second most common cancer in both men and women. The non-small cell lung cancer (NSCLC) is the majority (80-85%) of lung cancer and two major NSCLC types are Adenocarcinoma (ADC) (40%) and Squamous Cell Carcinoma (SCC) (25-30%).[1] The 5-year survival rate of lung cancer (19.4%) is still significantly lower than most other cancers.[2] Predicting clinical outcome of lung cancer is an active field in today's medical research.

Molecular profiling is a technique to query the expression of thousands of molecular data simultaneously. The information derived from molecular profiling can be used to classify tumors, and help to make clinical decisions [10, 11]. Many efforts have been made to search for biomarkers from molecule data that are significantly re-

[1] http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/

[2] http://seer.cancer.gov/statfacts/html/lungb.html

lated to patient death in lung cancer [12, 10, 11]. In 2008, Shedden et al. [10] showed that the gene expression signatures are able to predict lung cancer patients prognosis. However, their signature contained a large number of genes (over 1000 genes), which largely exceeded the 50 genes that most clinical assays can handle. Yuan et al. [11] investigated the benefit of integrating traditional clinical variables with diverse molecular data to predict patient survival. However, tumor microenvironment is a complex milieu that includes not only the cancer cells but also the stromal cells and immune cells. All this "extra" genomic information may muddle results and therefore make molecular analysis a challenging task for cancer prognosis [4].

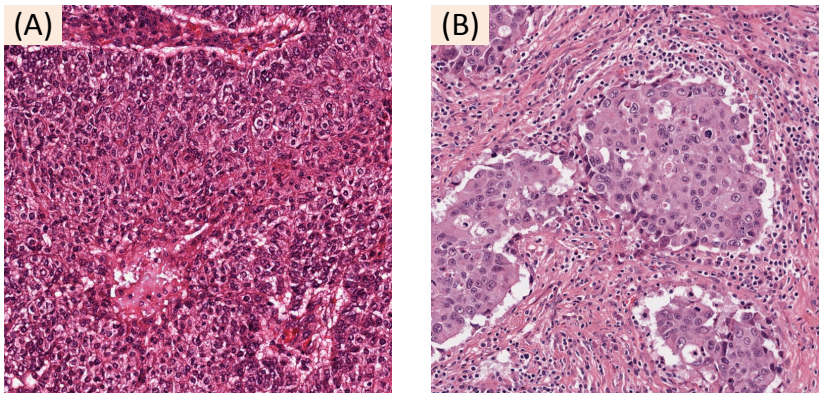

Figure 2.1: Tumor morphology are correlated with patient survival.

Recently, Arne Warth et al. [3] showed that there exists connections between lung tumor morphology and prognosis. Advances in imaging have created a good chance to study such information using hispathological images to help tumor diagnosis [4, 13, 14]. In general, a pathologist can visually examine stained slides of a tumor to discover imaging biomarkers that can be used for diagnosis. For example, Fig. 2.1-(A) shows two pathology images from ADC lung cancer patients. (A) is an image from one patient who had the worse survival outcome while (B) is captured from a

patient who lived longer. A distinct pattern can be found in Fig. 2.1-(A) as the more advanced tumor cells clustered in a larger more condensed area indicates a worse survival outcome than Fig. 2.1-(B) where tumor cells are scattered into a smaller region with lymphocytes and stromal cells nearby. However, the process of manually searching for such imaging biomarkers is very labor-intensive and cannot be easily scaled to large number of samples. Wang et al. [1] proposed an automated image analysis to help pathologists find imaging biomarkers that could identify lung cancer survival characteristics. They proposed a multi-scale distance map-based voting algorithm for cell detection and introduced an interactive scheme to form a repulsive balloon snake (RBS) model for touching cell segmentation. Based on cell detection and segmentation results, image morphometrics features are extracted for survival analysis. However, those traditional cell segmentation and detection approaches are unable to classify cell subtypes and achieve clinically interpretable imaging biomarkers in lung cancer.

In this chapter, we introduced a computational image analysis to discover clinically interpretable imaging biomarkers for lung cancer survival prediction. Experiments on two lung cancer cohorts demonstrate that: 1) Two major subtypes of NSCLC should be treated separately since they have different key imaging biomarkers. 2) Spatial distribution of subtype cells are informative imaging biomarkers for lung cancer survival prediction. 3) The proposed framework can better describe tumor morphology and can provide powerful survival analysis than the state-of-the-art method with molecular profiling data.

2.2 Methodology

An overview of our method is presented in Fig. 2.2. An expert pathologist first labels regions of tissues. Several image tiles are extracted from the interested

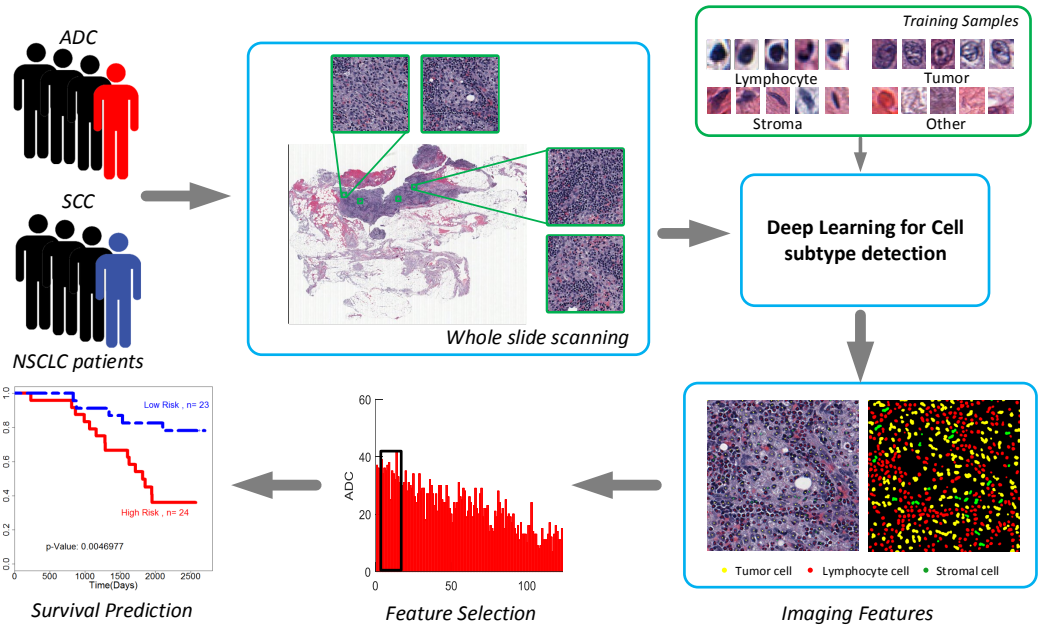

Figure 2.2: Overview of the proposed framework.

regions. Then a deep learning approach is applied to detect different types of cells (tumor, stroma and lymphocyte cells). A set of quantitative descriptors is used to cover granularity and subtype cellular heterogeneity. Our image analysis pipeline automatically segments H&E stained images, classifies cellular components into three categories (tumor, lymphocyte, stromal), and extracts features based on cell segmentation and detection results. Feature selection methods are used to find important features (image markers). These imaging biomarkers can then be applied for building survival models to predict patient clinical outcomes.

2.2.1 Subtype Cell Detection

In cell biology and medicine, microscopy images analysis is a very popular topic and automatic cell detection is the basis. Different cell types (cancer cells, stromal cells, lymphocytes) play different roles in tumor growth and metastasis, and accurately classifying cell types is a critical step to better characterization of tumor

growth and outcome prediction [7, 4]. Due to the large appearance variation and high complexity of lung cancer tissues, traditional machine learning approaches do not clearly distinguish or define the different cell types. Most models focus on single cell detection and few works are proposed for automatic microscopic subtype cell analysis with deep convolution neural networks (DCNN). In this work, we introduce an accelerated deep convolution neural network [15] for subtype cell detection and the main architecture of the network can be seen in Fig.2.3.

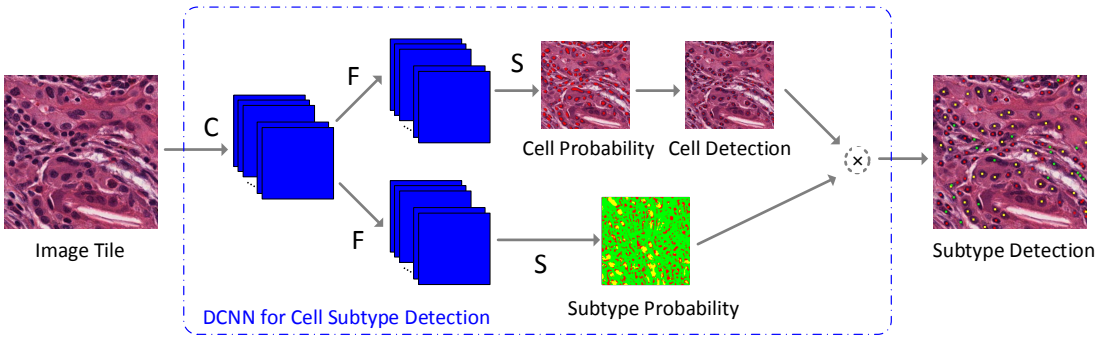

Figure 2.3: The architecture of DCNNs for cell type classification (C stands for the multiple shared convolution and pooling layers between two models. F stands for fully-connected layer and S stands for softmax layer).

Motivated by recent deep learning methods for cell detection [16, 17], the subtype cell detection network has a two partially shared-weighted deep convolution neural networks (DCNNs). One DCNN model is for cell/non-cell classification and the other is for subtype cell classification task. To train two partially share-weighted DCNN models for classification, cell patches are extracted according to their annotations. Sparse kernels [18] are applied in the two DCNN models to eliminate all the redundant calculations for acceleration. Sparse kernels are created by inserting all-zero rows and columns into original kernels to make every two original neighboring entries d-pixel away [19]. They are applied for all the convolution, pooling and

fully-connected layers in the network. Cell detection branch is for cell/non-cell task. After the softmax layer, the model outputs cell probability of the tile image. To get final cell detection result, the moment centroid based method is performed. The other branch is the DCNN for the subtype cell classification and it gives the probabilities of each pixel in the tile belongs to subtypes. In the final step, results of two DCNN models are merged by simple multiplying to achieve subtype cell detection.

2.2.2 Quantitative Imaging Feature Extraction

Motivated by [1, 20], three groups of cellular features were extracted using subtype cell detection results. These features cover cell-level information (e.g., appearance and shapes) of individual subtype cells and also texture properties of background tissue regions.

Group 1: Geometry Features. Geometry properties are calculated for each segmented subtype cell, including area, perimeter, circularity, major-minor axis ratio. Zernike moments were also applied on each type of cells. When combined with different tiles, we calculated mean, median and std. of each feature with a total of 564 features.

Group 2: Texture Features. This group of features contains Gabor "wavelet" features, co-occurrence matrix and granularity to measure texture properties of objects (e.g., cells and tissues), resulting in 1,685 texture features.

Group 3: Holistic Statistics. The four holistic statistics include overall information like the total area, perimeter, number and the corresponding ratio of each subtype cells.

2.2.3 Imaging Biomarkers Discovery

The objective of this step is to find important imaging biomarkers since not all features were highly correlated with patients' survival outcomes. Different from traditional applications, selecting features in survival analysis is a censoring problem (subjects are censored if they are not followed up or the study ends before they die). In this study, we built the predictive models using two well-established types of methods: (1) the multivariate Cox proportional hazards model with L1 penalized log partial likelihood (Lasso) [21] or component-wise likelihood based boosting (CoxBoost) [22] for feature selection, and (2) random survival forest (RSF) [23]. Because of the high dimension of the image features, we first applied univariate Cox regression and kept those with Wald test p value less than 0.05. Then we conducted the feature selection on a small candidates set for survival model to improve the speed.

2.3 Experimental Results

2.3.1 Materials

We focused on two widely used lung cancer dataset NLST (National Lung Screening Trial) [3] and TCGA Data Portal. Both dataset contains complete patients' pathology images with survival and clinical information while TCGA cohorts can provide additional molecular profiling data. In NLST, we collected 144 ADC and 113 SCC patients. In TCGA, we focused on SCC case and collected 106 patients with four types of molecular data including: Copy number variation (CNV), mRNA, microRNA and protein expression (RPPA). To examine whether imaging biomarkers from the proposed framework can achieve better predictions than traditional imaging

[3] https://biometry.nci.nih.gov/cdas/studies/nlst/

biomarkers and molecular profiling data (biomarkers), we evaluated with two state-of-the-arts framework in lung cancer [1, 11].

To evaluate the performances in survival prediction, we take the concordance index (C-index) as our evaluation metric [24]. The C-index quantifies the ranking quality of rankings and is calculated as follows

$$c = \frac{1}{n} \sum_{i \in \{1 \ldots N | \delta_i = 1\}} \sum_{t_j > t_i} I[f_i > f_j] \tag{2.1}$$

where n is the number of comparable pairs and $I[.]$ is the indicator function. $t.$ is the actual time observation. $f.$ denotes the corresponding risk. The C-index is a nonparametric measurement to quantify the discriminatory power of a predictive model: 1 indicates perfect prediction accuracy, and a C-index of 0.5 is as good as a random guess.

2.3.2 Imaging Biomarker Discovery for Survival Analysis

ADC vs SCC samples. In this experiment, we followed the framework in [1] and investigated differences in imaging biomarkers selecting from the set of ADC and SCC markers, and combining ADC and SCC markers together. To ensure the robustness of selection, we resampled the whole dataset with replacements and performed the boosting feature selection procedure [22] and calculated the frequency of choosing a variable. Fig. 2.4 shows that key features (high frequencies shown in the green rectangle) chosen from the combination set are very different from those of ADC and SCC, respectively. These differences convinced us the prognosis models for ADC and SCC should be developed separately. This discovery verified the evidence in lung cancer pathology, that lung cancer subtypes are highly heterogeneous and cannot be combined together.

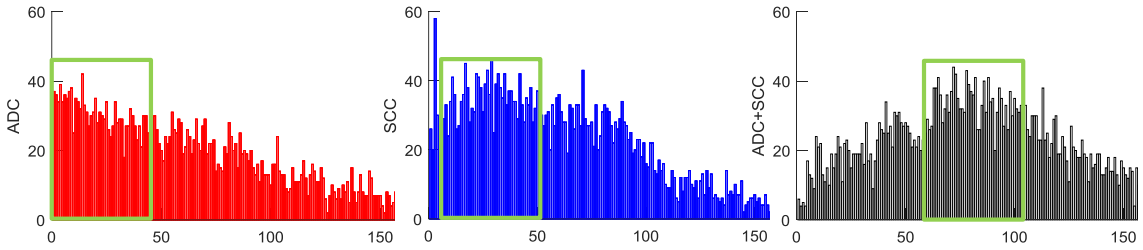

Figure 2.4: Frequencies of features on ADC, SCC and ADC+SCC set.

For ADC and SCC, selected features include information about suptype cell distributions, cell shape and granularity. Among them, subtype cell distributions and granularity have been confirmed to be associated with survival outcomes [25, 4]. To test these imaging biomarkers, we built multivariate Cox regression using the top 50 selected features on testing sets (47 for ADC and 37 for SCC). Fig. 2.5 presents the predictive power on a partitioning into two groups on testing set (a-b for ADC and c-d for SCC). A significant difference (Wald-Test) in survival times can be seen in Fig. 2.5-a,-c. It demonstrates that discovered imaging biomarkers which cover subtype cell distributions and granularity are more often associated with survival outcomes than traditional imaging biomarkers.

Then we randomly divided the whole set to 50 splits (2/3 for training, 1/3 for testing). Each feature selection method performed 10-fold cross validation for parameter optimization. Fig. 2.6 shows the concordance index (C-index) results of the two methods on ADC and SCC set. From Fig. 2.6, it can see the higher median C-index of the discovered imaging markers in both cases with different survival models. This illustrates the robustness of the proposed method since the discovered imaging biomarkers are highly associated with tumor growth and survival outcomes.

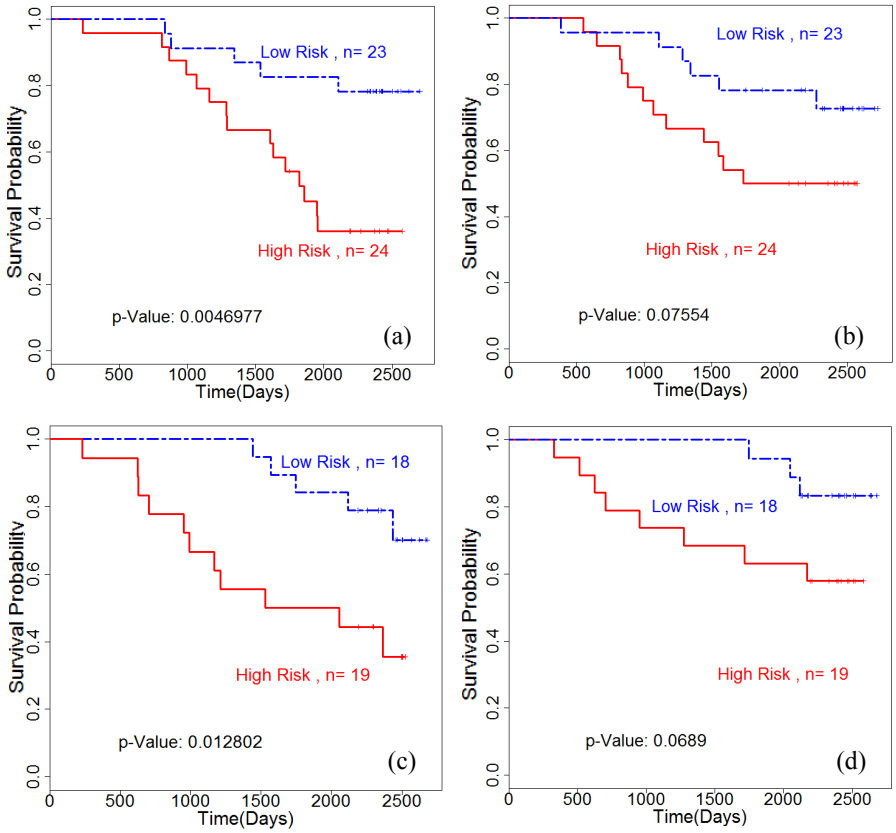

Figure 2.5: Kaplan-Meier survival curves of two groups on testing set. The x axis is the time in days and the y axis denotes the probability of overall survival. (a,c) are from the framework developed in this research, while (b,d) are using features from [1].

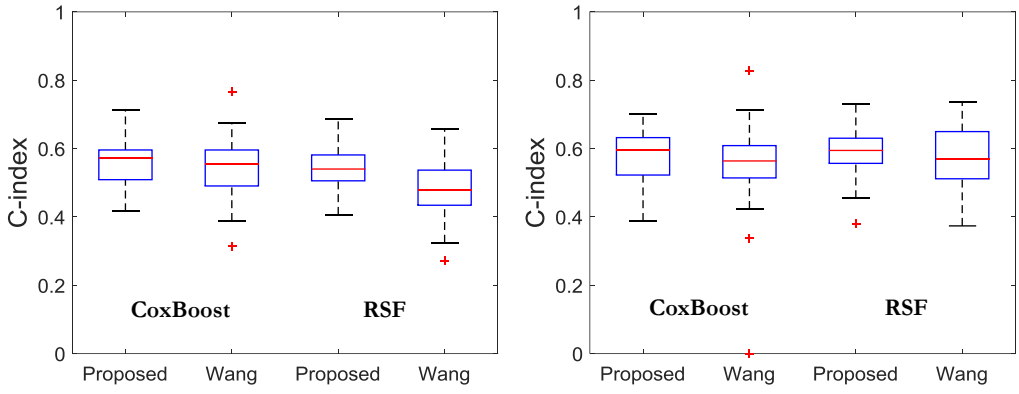

Figure 2.6: Boxplot of C-index distributions (Left: ADC, Right: SCC).

2.3.3 Comparison of Survival Model with Imaging and Molecular Data

To examine whether the proposed imaging biomarkers can provide better prediction power than traditional molecular data, we conducted experiments on TCGA LUSC cohort following the recent study [11]. We applied 50 random splits and assessed the C-index of a model built from the individual imaging and molecular data sets alone. Fig. 2.7-A presents the highest median C-index value of survival models built on the discovered imaging biomarkers. When each type of data integrates with clinical variables ("+" means the integration), all prediction accuracies increase while the proposed method still has the best results (Fig. 2.7-B). It verified the discovered imaging biomarkers can better describe tumor morphology which enabled the proposed framework to have the best predictions for survival analysis.

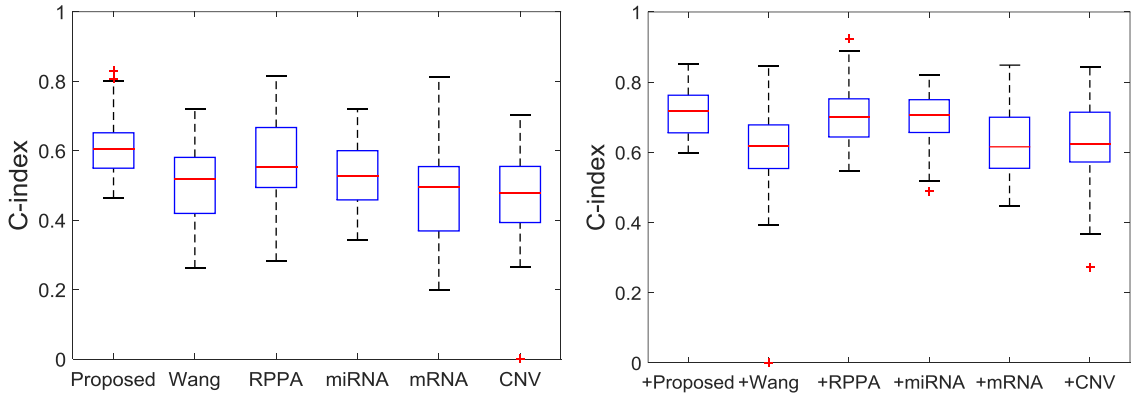

Figure 2.7: Comparison of the survival predictive power using Cox+Lasso model.

2.4 Conclusions

In this paper, we investigated subtype cell information and found that they have useful patterns for predicting patients survival. These results are consistent

with recent study in lung cancer pathology [3]. Extensive experiments have been conducted to demonstrate that imaging biomarkers from subtype cell information can better describe tumor morphology and provide more accurate prediction than state-of-the-art method using imaging and molecular profilers. In the future, we will try to find more quantitative measurements to better describe tumor morphology and further improve the prediction performances.

CHAPTER 3

DEEP CORRELATIONAL LEARNING FOR MULTI-MODALITY DATA

In this chapter, we develop a Deep Correlational Survival Model (DeepCorrSurv) for the integration of multi-view data. Extensive experiments on real lung and brain tumor data sets demonstrated the effectiveness of the proposed DeepCorrSurv model using multiple modalities data across different tumor types on small training samples [26].

3.1 Introduction

Survival analysis aims at modeling the time that elapses from the beginning of follow-up until a certain event of interest (e.g. biological death) occurs. The most popular survival model is Cox proportional hazards model [27]. However, the Cox model and recent approaches [28, 29, 30, 5] are still built based on the assumption that a patient's risk is a linear combination of covariates. Another limitation is that they mainly focus on one view and cannot efficiently handle multi-modalities data. Recently, Katzman *et al*. proposed a deep fully connected network (DeepSurv) to learn highly complex survival functions [31]. They demonstrated that DeepSurv outperformed the standard linear Cox proportional hazard model. However, DeepSurv cannot process pathological images and also is unable to handle multi-view data.

To integrate multiple modalities and eliminate view variations, a good solution is to learn a joint embedding space which different modalities can be compared directly. Such embedding space will benefit the survival analysis since recent study has suggested that common representation from different modalities provide important

information for prognosis [4, 14, 32]. To learn the embedding space, one very popular method is canonical correlation analysis (CCA) [33] which aims to learn features in two views that are maximally correlated. Deep canonical correlation analysis [34] has been shown to be advantageous and such correlational representation learning (CRL) methods provide a very good chance for integrating different modalities of survival data. However, since these CRL methods are unsupervised learning models, they still have the risk of discarding important markers that are highly associated with patients' survival outcomes.

In this chapter, we develop a Deep Correlational Survival Model (DeepCorrSurv) to integrate views of pathological images and molecular data for survival analysis. The proposed method first eliminates the view variations and finds the maximum correlated representation. Then it transfers feature hierarchies from such common space and specifically fine-tunes on the survival regression task. It has the ability to discover important markers that are not found by previous deep correlational learning which will benefit the survival prediction. The contribution of this paper can be summarized as: 1) DeepCorrSurv can model very complex view distributions and learn good estimators for predicting patients' survival outcomes with insufficient training samples. 2) It used CNNs to represent much more abstract features from pathological images for survival prediction. Traditional survival models usually adopted hand-crafted imaging features. 3) Extensive experiments on TCGA-LUSC and GBM demonstrate that DeepCorrSurv model outperforms those state-of-the-art methods and achieves more accurate predictions across different tumor types.

3.2 Methodology

Given two sets $\mathbf{X}, \mathbf{Y}$ with m samples, the i-th sample is denoted as $\mathbf{x}_i$ and $\mathbf{y}_i$. Survival analysis is about predicting the time duration until an event occurs, and in

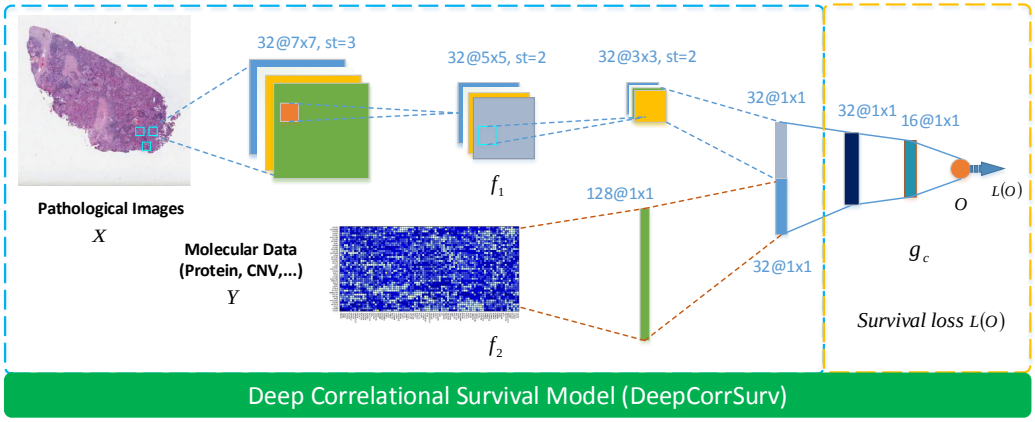

Figure 3.1: The architecture of the DeepCorrSurv. 'st' is short for 'stride'.

our case the event is the death of a cancer patient. In survival data set, patient i has observation time and the censored status, denoted as (t_i, δ_i). δ_i is the indicator: 1 is for a uncensored instance (the death event occurs during the study), and 0 is for a censored instance (the event is not observed). The observation time t_i is either a survival time (S_i) or a censored time (C_i) which is determined by the status indicator δ_i. If and only if $t_i = \min(S_i, C_i)$ can be observed during the study, the dataset is said to be right-censored which is the most common case in real world.

Figure 3.1 illustrates the pipeline of the proposed DeepCorrSurv. It consists of two sub-networks, view-specific sub-network f_1, f_2 and the common sub-network g_c. We proposed Convolutional Neural Networks (CNNs) as one image-view sub-network f_1 and Fully Connected Neural Networks (FCNs) as another view-specific sub-network f_2 to learn deep representations from pathological images and molecular profiling data, respectively. The sub-network f_1 consists of 3 convolutional layers, 1 max-pooling layer and 1 fully-connected layer. In each convolutional layer, we employ ReLU as the nonlinear activation function. The sub-network f_2 includes two fully connected layers with 128 and 32 neurons equipped with ReLU activation function.

3.2.1 Deep Correlational Model

For any sample $\mathbf{x}_i, \mathbf{y}_i$ passing through the corresponding view sub-network, its representation is denoted as $f_1(\mathbf{x}_i; \mathbf{w_x})$ and $f_2(\mathbf{y}_i; \mathbf{w_y})$ respectively where $\mathbf{w_x}, \mathbf{w_y}$ represent all parameters of two sub-networks. The outputs of two branches will be connected to a correlation layer to form the common representation.

Deep correlational model seeks pairs of projections that maximize the correlation of two outputs from each networks $f_1(\mathbf{x}_i; \mathbf{w_x}), f_2(\mathbf{y}_i; \mathbf{w_y})$. If $\mathbf{w_x}, \mathbf{w_y}$ mean all parameters of two networks, then the commonality is enforced by maximizing the correlation between two views as follows

$$L = corr(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{i=1}^{m}(f_1(\mathbf{x}_i) - \overline{f_1(\mathbf{X})})(f_2(\mathbf{y}_i) - \overline{f_2(\mathbf{Y})})}{\sqrt{\sum_{i=1}^{m}(f_1(\mathbf{x}_i) - \overline{f_1(\mathbf{X})})^2 \sum_{i=1}^{m}(f_2(\mathbf{y}_i) - \overline{f_2(\mathbf{Y})})^2}}, \tag{3.1}$$

where we omit networks' parameters $\mathbf{w_x}, \mathbf{w_y}$ in the loss function (3.1). We can maximize the correlation loss function to provide the shared representation indicating the most correlated features from two modalities.

3.2.2 Fine-tune with Survival Loss

Denote the shared representation from the two views as $\mathbf{Z}$. Denote $\mathbf{O} = [o_1, ..., o_m]^\top$ as the outputs of common sub-network $\mathbf{g}_c$, i.e., $o_i = \mathbf{g}_c(\mathbf{z}_i)$. Denote the label of the i-th patient as (t_i, δ_i) where t_i is the observed time, δ_i is the indicator, 1 is for a uncensored instance (death), and 0 is for a censored instance.

We assume that censoring data ($\delta = 0$, death not observed) is non-informative in that, given $\mathbf{x}_i$, the event and censoring time for the j-th patient are independent. Let $t_1 < t_2 < \cdots < t_D$ denote the ordered event times. The risk set $R(t_i)$ is the set of all individuals who are still under study at a time just prior to t_i. Conditioned

upon the existence of a unique event at some particular time t the probability that the death event occurs in the patient i is

$$L_i = \frac{\exp(\mathbf{o}_i)}{\Sigma_{j \in R(t_i)} \exp(\mathbf{o}_j)}, \tag{3.2}$$

Assuming the patients' events were statistically independent, the joint probability of all death events conditioned upon the existence of events at those times is the partial likelihood:

$$L = \prod_{i:\delta_i=1} \frac{\exp(\mathbf{o}_i)}{\Sigma_{j \in R(t_i)} \exp(\mathbf{o}_j)}, \tag{3.3}$$

The corresponding log partial likelihood is

$$\begin{aligned} l = log(L) &= \sum_{i:\delta_i=1} (\mathbf{o}_i - \log \sum_{j:R(t_i)} \exp(\mathbf{o}_j)) \\ &= \sum_i \delta_i (\mathbf{o}_i - \log \sum_{j:R(t_i)} \exp(\mathbf{o}_j)), \end{aligned} \tag{3.4}$$

The function can be maximized over the network parameters to produce maximum partial likelihood estimates. It is equivalent to minimize the negative log partial likelihood. We then use the negative log partial likelihood as the loss function in our model as shown in below

$$L(\mathbf{o}_i) = \sum_i \delta_i (-\mathbf{o}_i + \log \sum_{j:t_j >= t_i} \exp(\mathbf{o}_j)). \tag{3.5}$$

where j is from the set whose survival time is equal or larger than t_i ($t_j \geq t_i$). In a simplified view, the loss function contributes to overall concordance by penalizing any discordance in any values of higher risk patients if they are greater than lower those of lower risk. Different from Cox-based models which only handle the linear condition in the risk function, the proposed model can better fit realistic data and learn complex interactions using deep representation.

3.2.3 Discussions

Although different views of health data are very heterogeneous, they still do share common information for prognosis. Deep correlational learning is first trained to find such common representation using the correlation function (3.1). However, this procedure has a risk of discarding the discriminant markers for predicting patients' survival outcomes due to it belongs to unsupervised learning. To overcome this problem, the DeepCorrSurv transfers knowledge from the deep correlational learning and fine-tunes the network using the survival loss function (4.3). This will make DeepCorrSurv able to discover important markers that are ignored by correlational model and learn the best representation for survival prediction. Compared with the recent deep survival models [31, 35] which can only handle one specific view of data, the DeepCorrSurv can achieve more complex architecture for the integration of multimodalities data which can be used in the practical application on more challenging dataset.

3.3 Experiments

3.3.1 Dataset Description

We used a public cancer survival dataset TCGA (The Cancer Genome Atlas) project [36] which provides high resolution whole slide pathological images and molecular profiling data. We conducted experiments on two cancer types: glioblastoma multiforme (GBM) and lung squamous cell carcinoma (LUSC). For each cancer type, we adopted a core sample set from UT MD Anderson Cancer Center [11] in which each sample has information for the overall survival time, pathological images and molecular data related to gene expression.

- **TCGA-LUSC**: Non-Small-Cell Lung Carcinoma (NSCLC) is the majority of lung cancer. Lung squamous cell carcinoma (LUSC) is one major type in NSCLC. We collected 106 patients with pathological images and protein expression (reverse-phase protein array, 174 proteins).
- **TCGA-GBM**: Glioma is a type of brain cancer and it is the most common malignant brain tumor. 126 patients are selected from the core set with images and CNV data (Copy number variation, 106 dimension).

With the help of pathologists, we have annotations that locate the tumor regions in whole slide images (WSIs). We randomly extract patches of size 1024×1024 from the tumor regions. To analyze pathological images in comparison survival models, we calculated hand-crafted features using CellProfiler [37] which serves as a state-of-the-art medical image feature extracting and quantitative analysis tool. Similar to the pipeline in [20], a total of 1,795 quantitative features were calculated from each image tile. These types of image features include cell shape, size, texture of the cells and nuclei, as well as the distribution of pixel intensity in the cells and nuclei.

3.3.2 Comparison methods

We compare our DeepCorrSurv with five state-of-the-art survival models and three baselines deep survival models. Five survival methods include LASSO-Cox [12], Parametric censored regression models with components with Weibull, Logistic distribution [38], Boosting concordance index (BoostCI) [39] and Multi-Task Learning model for Survival Analysis (MTLSA) [40]. To demonstrate the effectiveness of the integration in our model, We adopted structured sparse CCA-based feature selection (SCCA) [41] to identify stronger correlation patterns from imaging genetic associations. Then we applied MTLSA using such associations for survival analysis.

Three baseline deep survival models are listed as follows: 1) CNN-Surv: CNN sub-network f_1 followed by survival loss [35]. 2)FCN-Surv: FCN sub-network f_2 followed by survival loss. It will use molecular profiling data for prediction. It can be also regarded as the DeepSurv [31] version on the dataset in this paper. 3)DeepCorr+DeepSurv: Since finding the common space by maximizing the correlation between two views belongs to unsupervised method, it cannot ensure that the embedding space is highly correlated with survival outcomes. We extract the shared representation by Deep correlational learning and feed them to another DeepSurv model.

Overall speaking, the DeepCorrSurv is optimized by the gradient descent following the chain rule, i.e., firstly compute the loss of objective, and then propagate the loss to each layer and finally employ gradient descent to update the whole network. These procedures can be automatically processed by Theano [42]. To make fair comparisons, the architectures of different deep survival models are kept the same with that corresponding parts in the proposed DeepCorrSurv. The source codes of MTLSA and SCCA are downloaded from the authors' websites. All other methods in our comparisons were implemented in R. LASSO-Cox and EN-Cox are built using the *cocktail* function from the *fastcox* package. The implementation of BoostCI can be found in the supplementary materials of [39]. The parametric censored regression are from the *survival* package.

3.3.3 Results and Discussion

In order to evaluate the proposed approach with other state-of-the-arts methods, we used a 5-fold cross-validation. For each of the 5 folds, models were established using the other 4 folds as the training subset, and performance was evaluated with the unused fold. To evaluate the performances in survival prediction, we take the

Table 3.1: Performance comparison of the proposed methods and other existing related methods using C-index values on TCGA-LUSC and GBM

Data	Model	LUSC	GBM
Images	LASSO-Cox [12]	0.5945 (0.1849)	0.5476 (0.0949)
	BoostCI [39]	0.5769 (0.2599)	0.5235 (0.1263)
	Weibull [38]	0.4988 (0.1924)	0.4885 (0.0127)
	Logistic [38]	0.4498 (0.1432)	0.4865 (0.0061)
	MTLSA [40]	0.6074 (0.1128)	0.6223 (0.1897)
	CNN-Surv [35]	0.5540 (0.2170)	0.5053 (0.0264)
Protein/CNV	LASSO-Cox [12]	0.5005 (0.1565)	0.5779 (0.0609)
	BoostCI [39]	0.4309 (0.1160)	0.4610 (0.1470)
	weibull [38]	0.4334 (0.1587)	0.5131 (0.0895)
	logistic [38]	0.5821 (0.1653)	0.5013 (0.1406)
	MTLSA [40]	0.5911 (0.2532)	0.6150 (0.1773)
	FCN-Surv [31]	0.5989 (0.1131)	0.5596 (0.0934)
Integration	SCCA [41] + MTLSA	0.5518 (0.0882)	0.5915 (0.1195)
	DeepCorr+DeepSurv	0.5760 (0.1645)	0.5842 (0.0450)
	DeepCorrSurv	**0.6287 (0.0596)**	**0.6452 (0.0389)**

concordance index (CI) as our evaluation metric. The C-index quantifies the ranking quality of rankings and is calculated as follows

$$c = \frac{1}{n} \sum_{i \in \{1 \ldots N | \delta_i = 1\}} \sum_{t_j > t_i} I[o_i > o_j] \tag{3.6}$$

where n is the number of comparable pairs and $I[.]$ is the indicator function. $t.$ is the actual observation and $o.$ represents the risk obtained from survival models.

Table 3.1 presents the C-index values by various survival regression methods on two datasets. Results of using each individual view in the table present that pathological images and molecule data can provide predictive powers while the integration of both modalities in the proposed DeepCorrSurv achieves the best performance for both lung and brain cancer. Because the proposed DeepCorrSurv can remove view discrepancy as well as learn the survival-related common representations from both views, it obtains the highest C-index with low standard variation. When looking at deep survival models, CNN-Surv cannot achieve good prediction using imaging data alone.

But when integrating with information from another view, DeepCorr+DeepSurv and the proposed DeepCorrSurv can achieve better performances than CNN-Surv using the same imaging data. This demonstrates that the common representation by maximizing the correlation between both views can benefit the survival analysis when the samples are not sufficient.

Another observation is DeepCorr+DeepSurv and SCCA+MTLSA cannot obtain a very good estimation compared with some predictions from one view. This demonstrates that the common representation by maximizing the correlation in an unsupervised manner still has the risk of discarding markers that are highly associated with survival outcomes. On the contrary, the DeepCorrSurv can consider discriminancy as well as view discrepancy which can ensure a representation that is robust to view discrepancy and also discriminant for survival prediction.

Results on TCGA-GBM dataset suggest that most models using CNV data can have better predictions than same models using imaging data. This observation is different from that in LUSC cohort. This reminds us, due to the heterogeneous of different tumor types, it is not easy to find a general model that can successfully estimate patients' survival outcomes across different tumor types using only one specific view. In addition, the original data in each view might contain variations or noises which are not survival-related and might affect the estimation of survival models. The proposed DeepCorrSurv can effectively integrate with two views and thus achieve good prediction performances across different tumor types.

3.4 Conclusion

In this paper, we proposed Deep Correlational Survival model (DeepCorrSurv) that is able to efficiently integrate multi-modalities censored data with small samples. One challenge is the view-discrepancy between different views in recent real cancer

database. To eliminate the view discrepancy between imaging data and molecular profiling data, deep correlational learning provides a good solution to maximize the correlation of two views and find the common embedding space. However, deep correlational learning is an unsupervised approach which cannot ensure the common space is suitable for survival prediction. In order to find the truly deep representations for prediction, the proposed DeepCorrSurv transfers knowledge from the embedding space and fine-tunes the whole network using survival loss. Experiments have shown that DeepCorrSurv can discover important markers that are ignored by correlational learning and extract the best representation for survival prediction. The results have shown that since DeepCorrSurv can model non-linear relationships between factors and prognosis, it achieved quite promising performances with improvements. In the future, we will extend the framework with other kinds of data sources.

CHAPTER 4

DEEP MULTI-INSTANCE SURVIVAL LEARNING FROM WHOLE SLIDE IMAGES

Above image-based survival models rely on discriminative patch labeling, which are both time consuming and infeasible to extend to large scale cancer datasets. The main challenge is that the gigapixel resolution of Whole Slide Pathological Images (WSIs) makes traditional approaches computationally impossible. Different from the existing works on learning using key patches or clusters from WSIs, in this chapter, we take advantage of a deep multiple instance learning to encode tissue patterns as instances from WSIs and view the problems of tumor patient survival analysis in a unified manner. Attention-based MIL pooling is performed to efficiently aggregate instance-level information to patient-level representation which is more flexible and adaptive than aggregation techniques in recent survival models.

4.1 Introduction

Recent technological innovations are enabling scientists to capture big whole slide images (WSIs) at increasing speed and resolution for diagnosis. The learning model is required to correctly predict the survival risk of each patient from his/her tumor tissue pathological images. The more precise is risk assessment for a cancer patient, the better the patient can be treated. Compared with other modalities, pathological images can present tumor growth and morphology in extremely detailed, gigapixel resolution which is extremely useful for tumor prognosis [3, 4]. The high resolution greatly benefits survival analysis with more precise information. However,

the diagnosis is extremely laborious and highly dependent on expertise which requires pathologists to carefully examine the biopsies under the microscope [43]. To reduce the risk of misdiagnosis, pathologists have to conduct a thorough inspection of the whole slide which make the diagnosis quite cumbersome. Automatic analysis of histology has become one of the most rapidly expanding fields in medical imaging. Computer aided diagnostics in digital pathology can not only alleviate pathologists' workload, but also help to reduce the chance of diagnosis mistakes. However, using WSIs for survival prediction is very challenging due to several reasons: 1) pathological images in real cancer dataset might be in terabytes (10^{12} pixels) level which makes most models computationally impossible. 2) the large variations of textures and biological structures from tumor heterogeneity, As the solid tumor may have a mixture of tissue architectures and structures, multiple WSIs from different parts of the patient's tissue are collected for diagnosis; 3) label on patient-level while each patient might have multiple WSIs for diagnosis. Those terabyte-size large WSIs from one patient will share the survival label which will make the problem more challenging.

4.1.1 Related Work

During recent years, many methods have been proposed for survival prediction using pathological slides. They can be categorized as ROI-based and WSI-based methods.

4.1.1.1 Region of Interest Analysis

Previously due to the lack of computational power, most of the literature focused on regions of interest (ROI) patches which are selected by pathologists from WSIs [44]. Instead of handling original WSIs, state-of-the-art survival methods extracted hand-crafted features from ROIs for predictions [4, 13, 14, 5, 1, 6, 45]. Wang

et al. [1] proposed a novel framework to first segment cells in annotated patches and then perform cellular morphological properties from those cells which result in 166 imaging features. Yu et al. [6] extract 9,879 quantitative image features from annotated regions of interest and results suggest that automatically derived image features can predict the prognosis of lung cancer patients and thereby contribute to precision oncology. Beyond classical cell detection, Yao et al. [5] used a deep subtype cell detection first to classify different cell subtypes and then extracted features from cellular subtype information. Cheng et al. [45] used a deep auto-encoder to cluster cell patches into different types and then extracted topological features to characterize cell type distributions from ROIs for prediction. These methods extracted hand-crafted features based on nuclei detection and segmentation and those features were considered to represent prior knowledge of boundary, region or shape. However, hand-crafted features are limited in representation power and capability.

Recently, with the advance of deep neural networks, deep learning-based survival models are proposed for seeking more powerful deep representation [31, 35, 26, 46]. Katzman *et al*. first proposed a deep fully connected network (DeepSurv) to represent the nonlinear risk function [31]. They demonstrated that DeepSurv outperformed the standard linear Cox proportional hazard model. Another improvement is deep convolutional survival learning (DeepConSurv) which is the first attempt to use pathological images in deep survival model [35]. Later, Yao et al. [26] integrated genome modality with DeepConSurv for survival prediction using multi-modality data. However, DeepConSurv is designed to use pre-selected ROI patches by pathologists from WSIs and then perform CNN based on those patches. A small set of image tiles might not completely and properly reflect the patients' tumor morphology. Also, those methods perform average pooling to achieve patient-wise predictions from patch-based results. Such combination cannot effectively aggregate predictions

from patch-level and needs further attention. Thus, it would be much helpful if we can facilitate knowledge discovery from big whole slide images.

4.1.1.2 WSI-based Analysis

Although recent deep neural networks have achieved very promising performances on various computer vision tasks [47, 48], they are unable to perform convolutional operations directly on whole slide images in gigapixel. With detailed and densely annotations on WSIs, nowadays a series of approaches whole-slide image analysis have been proposed for a variety of applications including classification, detection or segmentation [19, 43, 49, 50, 51]. However, the success of those applications is built on integrating detailed patch contents and using labor-extensive annotations which might not be applicable for survival prediction. Recent classification tasks actually are for slide-level decision making while survival prediction is based on patient-level analysis and one patient might have multiple whole slide images. How to make patient-level decision from slide-level results is not the target of those studies.

To achieve weakly-supervised survival prediction without annotations, Zhu et al. [8] proposed a patch-based two-stage framework to predict patients' survival outcomes. Patches are extracted from the WSIs and clustered to different patterns defined as "phenotypes" according to their visual appearances in the first stage. Then WSISA [8] adopted DeepConvSurv [35] to select important patch clusters and aggregated those clusters for final prediction. Although this framework has practical merits to consider important patch clusters, it is hard to incorporate it into state-of-the-art deep learning paradigm as the whole approach has separate steps. In addition, it is not a scalable solution because the first stage will be significantly inefficient if more patches are sampled. Moreover, it treats each cluster within a patient as independent of each other and doesn't consider any connections among clusters. One recent work

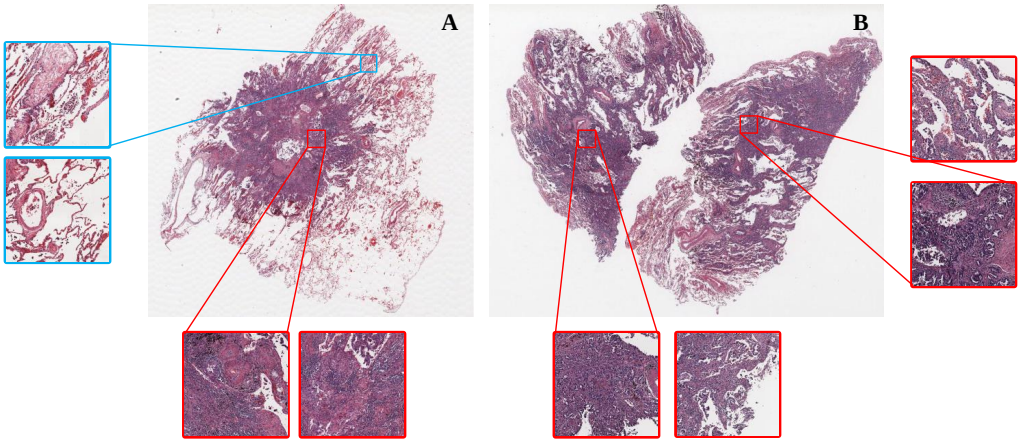

Figure 4.1: Gigapixel Whole Slide Histopathological Images of two lung cancer patients (best viewed in color). Patient B has worse clinical outcome than patient A. Patches shown in red are from lung tumor. Patches in blue are from low-grade tumor or non-tumor tissue regions. Discriminative patterns from both A and B are very similar but patient A has more non-tumor or low-grade tumor regions.

CapSurv [9] is proposed by introducing Capsule network [52]. However, CapSurv still has similar issues with WSISA as the main framework is following the WSISA pipeline.

The relationship of tissue patterns on WSIs is the great importance on survival analysis. Fig.4.1 shows whole slide images from two lung cancer patients. Patches framed in red represent tumor regions and those in blue show low-grade tumor or non-tumor regions. Patient B has much worse clinical outcome than patient A and one distinct pattern in patient A is that the biopsy sample has more non-tumor tissue regions. This observation reminds us that the joint effects of phenotypes could be used to better predict patients' survival outcomes. Li et al. [53] proposed a graph convolutional network (GCN) based method to consider such relationship of patches in the WSIs and then learn effective representation for survival prediction. However, this method requires detailed graph structure knowledge to construct a complete

graph representation for effective GCN training which is not flexible and needs prior knowledge.

4.1.2 Contributions

Though many works can be found on WSI analysis for segmentation, classification and detection, there were limited works on weakly-supervised learning for survival prediction. In this study, we propose a novel framework, referred to as Deep Attention Multiple-Instance Survival Learning (DeepAttnMISL) for Whole Slide Images. By viewing the problem of survival prediction in a unified manner as a form of the multi-instance learning, the proposed model can identify patients' survival outcomes from WSIs without any additional annotations. More specifically, DeepAttnMISL uses the siamese MI-FCN network to learn features from different phenotype clusters. Attention-based MIL pooling layer is added to perform a trainable weighted aggregation and generate the patient-level representation from all instance representations. The proposed framework can effectively highlight the prognosis-related clusters and extract image features from larger range in WSIs without small region limitations of ROIs. The contributions can be summarized as follows

- Different from recent WSI survival models [8] that treated and trained patch clusters independently from all patients, we formulate survival prediction in the Multiple instance learning (MIL) which increases the flexibility of the approach and allows to train the model by optimizing survival objective function.
- Each phenotype provides morphology-specific representation, the proposed DeepAttnMISL aggregates them using a trainable weighted average where weights can be fully parameterized by neural networks that corresponds to the attention mechanism which is much flexible than fixed pooling operators in recent work [8, 9].

- With the advantage of MIL and attention mechanism, the proposed has a good interpretability to find important patterns of patients are more likely to achieve better patient-level predictions.

To evaluate the performance of the proposed DeepAttnMISL model, one large WSI datasets on lung cancer is used and extensive experimental results verify the effectiveness —our method can efficiently exploit and utilize all discriminative patterns in whole slide pathological images to perform accurate patients' survival predictions. Additionally, we present results representing a patient's treatment group to illustrate how to view the proposed model as a treatment recommender system. Results validate that the proposed model can accurately model the risk functions of the population and thus guide treatment decisions for improving patient lifespan.

4.2 Methodolodgy

Given a set of N patients, $\{X_i\}, i = 1 \dots N$, each patient has the label (t_i, δ_i) indicating the survival status. The observation time t_i is either a survival time (O_i) or a censored time (C_i) for each data instance. If and only if $t_i = \min(O_i, C_i)$ can be observed during the study, the dataset is said to be right-censored [54]. δ_i is the indicator which is 1 for an uncensored instance (death occurs during the study) and 0 for a censored instance. Survival model predicts a value of a target variable y for a given patient where Y measures the hazard risk. As we discussed above, patient X_i will have multiple WSIs and our goal is to predict the corresponding target Y_i from those imaging data. As we don't have WSI-level annotations but only know patient-level information, this weakly-supervised learning can be solved by Multiple Instance Learning (MIL).

4.2.1 Multi-Instance Learning

In contrast to the standard supervised learning, multi-instance learning (MIL) considers a set of bags, each containing multiple feature vectors referred to as instances. The available label is only assigned to bag-level and labels of individual instances in the bag are not known. In MIL, not all the instances are necessarily relevant and some of them in the bag might not be relevant to certain labels. In the case of MIL problem, patient X is a bag of instances, $X = \{x_1, ..., x_K\}$ and K could vary for different bags. Furthermore, we assume that individual labels exist for the instances within a bag, i.e., $y_1, ..., y_K$ but those labels remain unknown during training. One very important assumption is that neither ordering nor dependency of instances within a bag and a MIL model must be permutation-invariant.

Based on the nature of MIL, it seems to perfectly fit medical imaging where medical domain often encounters annotations problem as labeling is much more expensive and requires long-time expertise training than that in the computer vision field. Dividing a medical image into smaller patches could be further considered as a bag with a single label. Recently, researchers have developed many MIL-based algorithms to medical images for segmentation [55], classification [56, 57] and gene annotation [58]. Hou et al. [56] proposed a MIL approach to train CNN to identify gigapixel resolution pathology images. It is the first work to use MIL method for WSIs classification. A more recent work [57] investigated a multi-instance based model on a multiple label classification task. However, survival prediction is more challenging than the segmentation or classification problem because it is a regression problem where the ranking of patients' prediction values matters [59] while in the tumor classification task, the prediction result of one patient's category is independent with others. More importantly, since the individual labels for instances are unknown, there is a threat that the instance-level classifier might be trained insufficient. Because

these methods aggregated instance-level to bag-level results by pre-defined pooling (e.g. max-pooling), it might introduce additional error to the final prediction which will be inappropriate for a more challenging task.

4.2.2 DeepAttnMISL

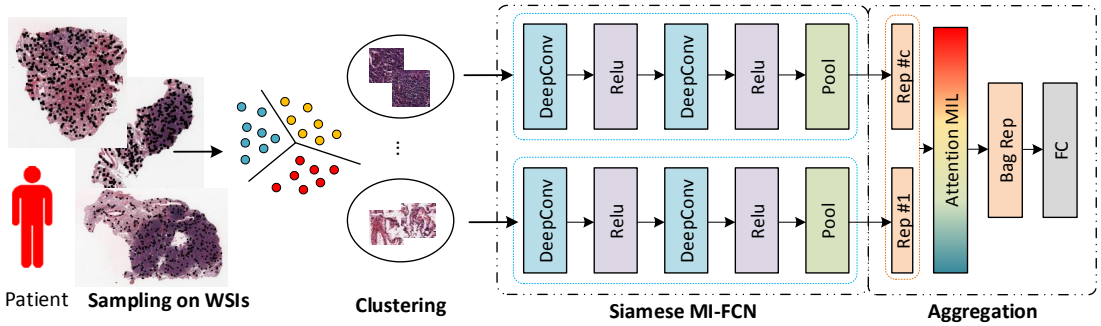

Figure 4.2: An overview of the proposed DeepAttnMISL model.

Fig.4.2 shows the overview of the proposed Deep Attention Multiple Instance Survival Learning (DeepAttnMISL) framework. Each patient X_i may contain multiple whole slides. To construct the bag in MIL, we choose phenotypes instead of sampling patches as instances within the bag because it will considerably reduce the complexity of the problem as there are a large number of heterogeneous patches. Phenotype patterns have been verified in state-of-the-art methods [8, 9] and are capable of representing different patterns in WSIs. By using phenotype patterns which are constructed by clustering, we can build the model for different types of tissue to extract morphology-specific features. To learn patient-level information from phenotype clusters, we design a multiple Multi-Instance Fully Convolutional Network (MI-FCN) running inside our deep learning architecture with weights being shared among them as in the Siamese architecture. To detect important phenotypes asso-

ciated with patients' clinical outcomes, attention-based MIL pooling layer is used to aggregate phenotype-level representation. The output is the hazard risk to represent how well for the patient behaves in the population of certain type of diseases.

4.2.2.1 Sampling and Clustering

At the first step, we extract patches from all WSIs which belong to the same patient and then cluster them into different phenotypes. To capture detailed information of the images, those patches are extracted from 20X (0.5 microns per pixel) objective magnifications and then fixed to $500 \times 500 \times 3$ size. We use the pre-trained VGG model from ImageNet [60] to extract features for each image patch ($d = 4096$) which have more representation power than smaller size (50×50) thumbnail images to represent their phenotypes [8]. Then we adopt K-means to cluster patches based on their VGG features. Notice that one patient might have multiple WSIs and we actually perform clustering on patient-level instead of the whole database. Fig.4.3 shows one patient's example. This patient has three WSIs that were sampled from different locations of the biopsy tissue. The corresponding phenotype clustering are shown in the right and each color means one type of phenotype clusters. In this example, we set the number of phenotypes C to 10. The results show the effectiveness of the clustering as we can see similar patches are grouped into the same cluster. This example demonstrates that VGG features are capable of identifying patterns of whole slide images and we would expect them to be distinctive and informative for survival learning task.

By clustering different patches from all WSIs of the patient into several distinguished phenotype groups, we will have different phenotype groups with various prediction power on this patients clinical outcome. The proposed DeepAttnMISL

takes phenotypes as multiple inputs and consider their connections for predicting survival status.

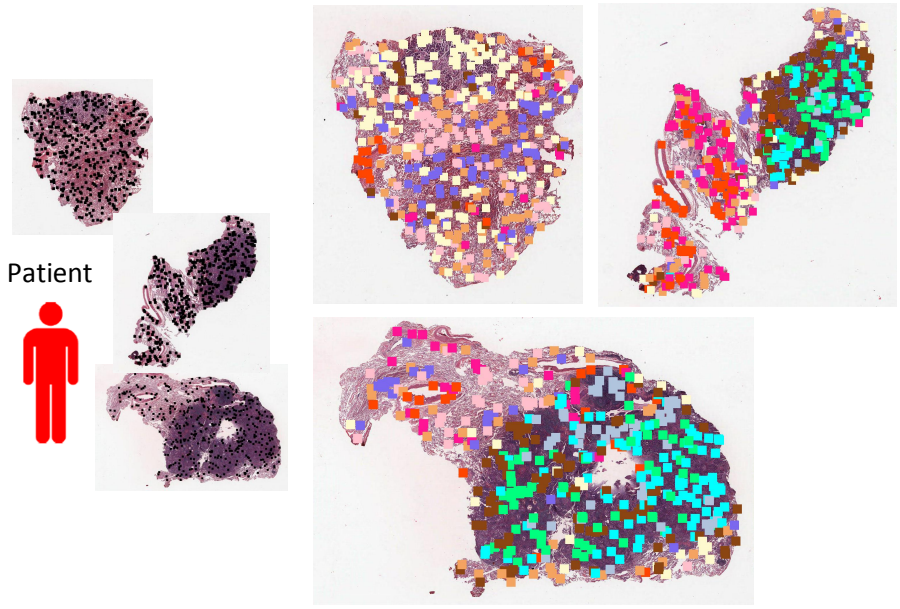

Figure 4.3: Phenotype patterns visualization after clustering on three WSIs belong to the same patient.

4.2.2.2 Siamese MI-FCN

After clustering, the patient is a set of phenotype clusters and we design a siamese MI-FCN to learn features from those patterns. Most existing well-known pre-trained models were trained based on single-instance bases, and the labels are associated with each image which is not the case of our problem. We embed multiple sub-networks running inside our deep learning architecture with weights being shared among them as in the Siamese architecture. Each sub-network is based on fully convolutional neural networks (FCN) that can learn informative representation for individual phenotype of the patient.

The architecture of each Multi-Instance Fully Convolutional Networks (MI-FCN) is shown in Fig.4.4. The combination of multiple layers of fully convolutional layers and non-linear activation functions has proven to be a powerful non-linear feature mapping in Multi-Instance problem [61]. The reason to use the fully convolutional networks (FCN) without including any fully connected layers is that FCN is more flexible and can handle any spatial resolution, which is needed for the considered problem since the number of patch samples in each phenotype varies. For each phenotype, the input is a set of features from m_i patches, can be organized as $1 \times m_i \times d$ (d is the feature dimension or channel). The network consists of several layer-pairs of 1×1 conv layer and ReLU layer (we show 2 pairs in Fig.4.4). The global pooling layer (e.g. average pooling) will be added at the end. For j-th phenotype, its representation is denoted as $\mathbf{r_j}$. The network receives one kind of phenotypes (tensor) as input and it can focus on local information and generate representation for the phenotype. Since the number of patches in each phenotype varies, the fully convolutional network is more flexible to handle this scenario.

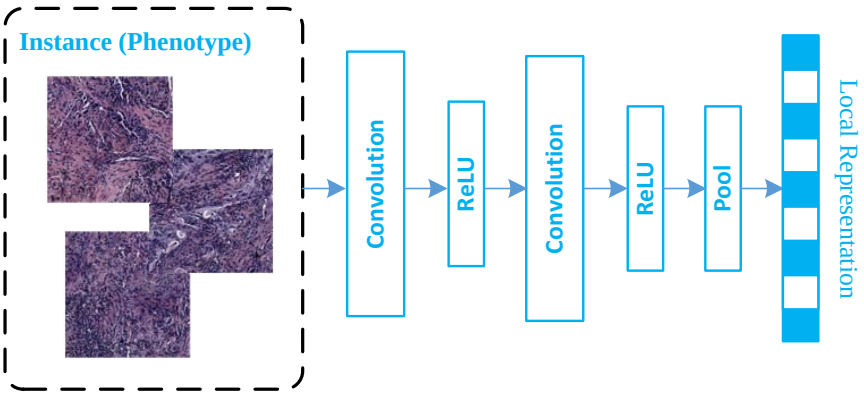

Figure 4.4: The network architecture in each MI-FCN.

4.2.2.3 Aggregation via Attention-based MIL pooling layer

Local representations from MI-FCN encode information of the corresponding phenotype clusters and how to aggregate them into patient-level representation is one necessary step. Let $R = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_C\}$ be one patient with C phenotype local representations and the goal is to get patient-level representation $\mathbf{z}$. The very straightforward choice is to use maximum or the mean operator, but drawbacks are very clear that they are pre-defined and non-trainable which might not be flexible and adjustable to the specific task. Previous work [8] used weighted average of features from clusters to get the patient feature but they performed such patient-level aggregation in a separate stage and the whole approach cannot be trained end-to-end from patient-level to instances-level. A better way to integrate phenotype-level information is to leverage an attention mechanism that considers the importance of each phenotype. In this paper, we propose to use the attention-based MIL pooling [62] for aggregation which is flexible and adaptive. By using such pooling operator, the patient-level representation can be calculated as

$$\mathbf{z} = \sum_{k=1}^{C} a_k \mathbf{r}_k, \tag{4.1}$$

where

$$a_k = \frac{\exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{h}_k^\top)\}}{\sum_{j=1}^{K} \exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{h}_j^\top)\}}. \tag{4.2}$$

In the weight a_k calculation, $\mathbf{w} \in \mathbb{R}^{L\times 1}$ and $\mathbf{V} \in \mathbb{R}^{L\times M}$ are trainable parameters. Tangent tanh(.) element-wise non-linearity is introduced both negative and positive values for proper gradient flow. The attention-based MIL pooling allows to assign different weights to phenotype clusters within one patient and hence the final patient-level representation could be highly informative for survival prediction. In other words, it should be able to locate key clusters and provide potential ROIs. Different from

traditional attention mechanism that all instances are sequentially dependent [63, 64], Multi-instance learning assumes all instances are independent. As phenotype in our problem is more natural to be independent to each other, attention mechanism used in MIL pooling will be beneficial to achieve good results.

4.2.2.4 Loss Function

We use the same negative log partial likelihood described in the previous chapter as the loss function in our model as shown in below

$$L(\mathbf{o}_i) = \sum_i \delta_i(-\mathbf{o}_i + \log \sum_{j:t_j>=t_i} \exp(\mathbf{o}_j)). \tag{4.3}$$

where j is from the set whose survival time is equal or larger than t_i ($t_j \geq t_i$). In a simplified view, the loss function contributes to overall concordance by penalizing any discordance in any values of higher risk patients if they are greater than lower those of lower risk. Different with other deep models used the same loss function [31, 35, 8], the proposed model can better fit realistic patients' whole slide imaging data and learn complex interactions using deep multi-instance representation that cover both holistic and local information. Since patient's risk is correlated with phenotypes from WSIs, the proposed framework can efficiently exploit phenotypes by deep multi-instance learning and attention mechanism for clinical outcome prediction at patient-level.

4.2.3 Discussion

There are several differences from the state-of-the-art method survival method using WSIs [8, 9]. First, clustering is performed on patient-wise while those approaches need to cluster on all patches from patients of the database. It is obvious that clustering patches within each patient will be more scalable as the number of

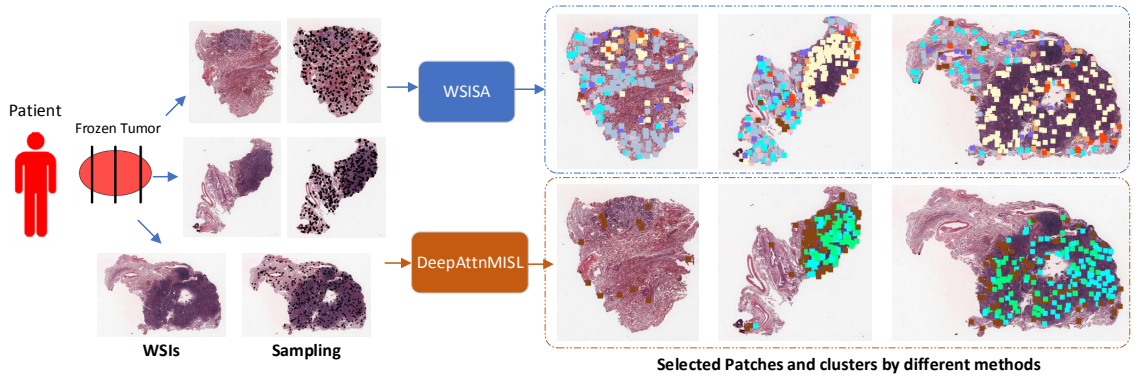

Figure 4.5: Visualization of selected patches and clusters from the proposed method and WSISA.

patients is much less than the number of patches. Because WSISA [8] needs independent DeepConvSurv to select important clusters and it has to divide the whole dataset into different types by clustering on all patches. With the advantage of MIL and attention mechanism, the proposed DeepAttnMISL can easily find important instances (clusters) within the bag are more likely to achieve better patient-level predictions. There is no need to perform clustering on the whole dataset. A trainable and adaptive attention-based MIL pooling in DeepAttnMISL can adjust to a task and data which could help succeed in calculating the better bag representation. Fig.4.5 shows one illustration example. This patient has three WSIs with two have clear tumor tissues. Black points on images record sampling positions. We present selected patches after WSISA and the proposed DeepAttnWISL in the right. The colors represent which clusters those patches belong to. Because clustering in WSISA is performed on the whole dataset, results will be easily biased by the heterogeneity of the dataset and the robustness of clustering algorithm. Results suggest selected patches from WSISA contain many non-tumor regions while the proposed DeepAttnMISL can better fo-

cus on tumor regions with attention mechanism which present better interpretability. More experimental results can be found in the next section.

4.3 Experiments

In this section, we will first describe data we use and then present performances of the proposed method on each dataset.

4.3.1 Dataset Description

To validate the performance of the proposed DeepAttnMISL, we use the very large dataset on lung cancer with high-resolution WSIs - the National Lung Screening Trial (NLST) [65]. NLST is a very large lung cancer dataset collected by the National Cancer Institute's Division of Cancer Prevention (DCP) and Division of Cancer Treatment and Diagnosis (DCTD). Clinical and pathological data were collected on all those cases, including 5-year follow-up data. In one whole slide image, usually about 50% of areas are background and it is easy to select regions to contain tissues rather than background or irregular regions according to pixel values. We extract patches from 20X objective magnifications and then fixed the size to 500×500. Even we only extract tissue patches and ignore background regions, it can still get tens of thousands of patches per WSI which will result in a huge number of images from the whole dataset. State-of-the-art WSI models [8, 9] need to control the scale of data as they will have significant computational issues on the very large number of patches. They sampled hundreds patches per WSI and collected around 20K-200K patches in total. One advantage of the proposed model is it has less computation issues because it uses MIL with attention to aggregate features from pre-trained models instead of training patch-based CNNs which is very time costly [66].In summary, the numbers of WSIs and patients in each dataset are shown in Table.4.1.

Table 4.1: The numbers of WSIs, patients, patches extracted.

Dataset	#patients	#WSIs	#patches	#patches/WSI
NLST	387	1,177	275,244	233

4.3.2 Implementation details

For training, we use Adam optimization with weight decay 5×10^{-4}. The learning rate is set to 10^{-4} and the training monitors the loss on validation dataset and it will early stop if the loss goes increased much. To select parameters of our model, we split the data into 80% training and 20% testing. 25% of training data will be used as validation data for achieving early stop training. Then we conducted 5-Fold cross-validation and report the average value on testing folds for more extensive evaluations.

To evaluate the performances in survival prediction, we take the concordance index (C-index) and area under curve (AUC) as our evaluation metrics [24]. The C-index quantifies the ranking quality of rankings and is calculated as follows

$$c = \frac{1}{n} \sum_{i \in \{1 \ldots N | \delta_i = 1\}} \sum_{t_j > t_i} I[f_i > f_j] \tag{4.4}$$

where n is the number of comparable pairs and $I[.]$ is the indicator function. $t.$ is the actual time observation. $f.$ denotes the corresponding risk. The value of C-index ranges from 0 to 1. The larger the value is, the better the model predicts.

4.3.3 Results

To compare with state-of-the-art image-based survival models, we make extensive experiments on NLST dataset as we have annotations that locate the tumor regions in whole slide images (WSIs) with the help of pathologists. Following the recent framework [6], we calculated hand-crafted features using CellProfiler [37] which serves as a state-of-the-art medical image feature extracting and quantitative analysis

tool. A total of 1,795 quantitative features were obtained from each image tile. Then we averaged those features across different patches for each patient. These types of image features include cell shape, size, texture of the cells and nuclei, as well as the distribution of pixel intensity in the cells and nuclei.

We compare our framework with several state-of-the-art survival models using pathological images. We can summarize the comparison methods into five categories as follows:

- **Cox models**: The Cox proportional hazards model is the most commonly used semi-parametric model in survival analysis. Two regularized Cox models l_1-norm (LASSO-Cox) [12] and boosting cox model (Cox-boost) [22] are compared in experiments.
- **Parametric censored regression models**: PCR models formulates the joint probability of the uncensored and censored instances as a product of death density function and survival functions, respectively [67]. We choose Weibull, Logistic distribution to approximate the survival data.
- **MTLSA**: Multi-Task Learning model for Survival Analysis (MTLSA) [40] reformulates the survival model into a multi-task learning problem.
- **WSISA**: WSISA can learn effective features from WSIs [8]. We then train LassoCox and MTLSA using WSISA learned features as they are top models based on their report.

4.3.3.1 Quantitative Results

We reported results from a few possible numbers of phenotypes, such as $\{6, 8, 10, 12\}$ on the testing dataset. From the Table 4.2, we can see models using fewer clusters are unable to achieve good results. The reason might be patches of lung cancer patients are very heterogeneous and it is relative difficult to learn survival-related

representations from fewer phenotypes. Results suggest the number of 10 achieves slightly better predictions which is consistent with findings in WSISA [8]. Thus, we decide to choose to cluster 10 phenotypes in our model. In each MI-FCN, we use one convolutional-ReLU layer pair with Global Average Pooling.

Table 4.2: Performances with different number of phenotypes.

No.	6	8	10	12
CI	0.6734	0.7691	0.7748	0.7417

Table 4.3 shows C-index and AUC values by various survival regression methods on 5-fold cross validation. It shows the prediction power of the proposed method compared with different survival models. One can see that the proposed method achieves both highest C-index and AUC values which present the best prediction performance among all methods. From the table, baseline models using hand-crafted features perform not well due to following reasons: 1) the limitation of local information provided by the patches extracted from the ROI using hand-crafted features; 2) the non-effective aggregation way to represent the heterogeneity of tumor and patient from patch-based results. Instead of using a small set of patches and human-designed features, the proposed method can effectively learn complex deep bag representation from phenotype patterns to predict patient survival outcomes.

WSISA [8] is the most representative WSI-based survival learning but it only extracts features from WSIs and needs a separate survival learning to get final predictions. We choose top survival models according to settings in WSISA [8], they are Lasso-Cox [12] and MTLSA [40]. WSISA achieves better results than baseline models which shows the good representative ability of features from WSISA. However, WSISA needs a separate stage to train several DeepConvSurv models independently

Table 4.3: Performance comparison of the proposed methods and other existing related methods using C-index values on NLST dataset. The larger C-index value is better.

Type	method	C-index	AUC
Deep Learning	DeepAttnMISL	**0.6829** (0.0385)	**0.7143** (0.0541)
	WSISA-LassoCox	0.5996 (0.0750)	0.5957 (0.0674)
	WSISA-MTLSA	0.6305 (0.0575)	0.6479 (0.0936)
Cox-based	Lasso-Cox	0.4842 (0.0508)	0.4903 (0.1011)
	Cox-boost	0.5474 (0.0370)	0.5271 (0.0386)
Parametric models	Logistic	0.4998 (0.0881)	0.5013 (0.1146)
	Weibull	0.5577 (0.0395)	0.5618 (0.0976)
Multi-task based	MTLSA	0.5053 (0.0509)	0.5362 (0.0416)
Ranking based	BoostCI	0.5595 (0.0610)	0.5487 (0.0532)

and will discard some phenotypes in the final stage, the performance actually depends on how well to select important clusters and WSISA still has the chance to lose in selecting survival-related clusters for a good final survival prediction. Instead of selecting phenotypes, the proposed model is designed to consider all possible survival-related patterns and uses more flexible attention mechanism to learn more informative and discriminate patterns. This architecture of multiple instance makes the proposed method can better learn heterogeneous information encoded in WSIs from large number of patches which will make it more practical in real applications.

4.3.3.2 Personalized Recommendations

Given the trained survival models, we can use the estimated testing risk scores to classify patients into low or high-risk group for personalized treatments. Two groups are classified by the median of predicted risk scores. We evaluate if those models can correctly classify death patients (uncensored data) into two groups since uncensored data is more informative. Patients with longer survival time should be classified into low risk group and vice versa. If the model cannot correctly distinguish high and low

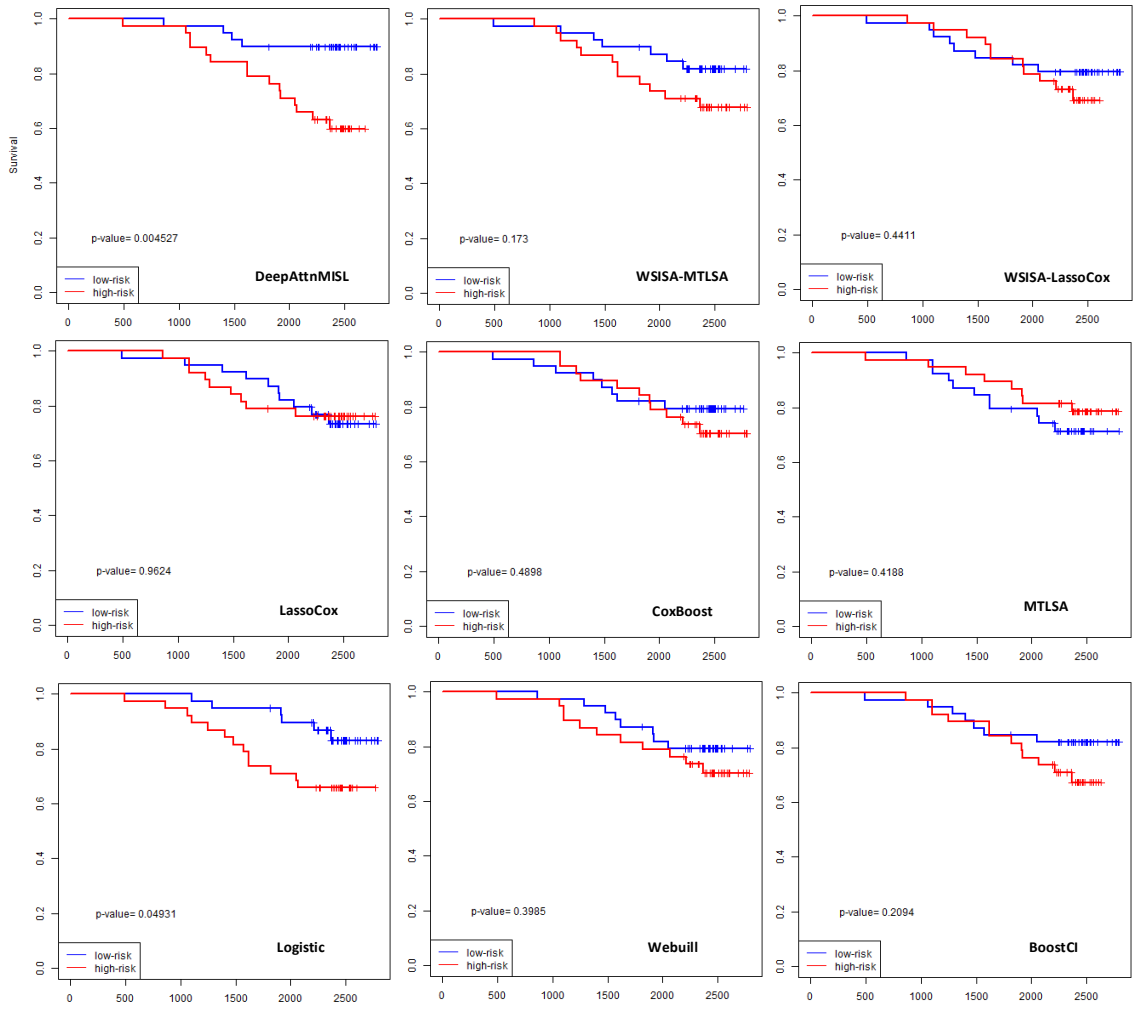

Figure 4.6: Kaplan-Meier survival curves of different models for one testing fold. High risk (great than median) groups are plotted as green lines, and low risk (less than or equal to median) groups are plotted as red lines. The x axis shows the time in days and y axis presents the probability of overall survival. Log rank p value is shown on each figure. "+" means the censored patient.

risk death patients, two average death times should be very close. We plot Kaplan-Meier survival curves in Fig.4.6. From the figure, one can see that the proposed model can more successfully group testing death patients into two groups than other methods in all datasets. The log rank test is conducted to test the difference of two curves.

It is shown that the proposed method can achieve the most significant log rank test outcome (p-value = $4.527e^{-3}$) while some of others do not reach statistical significances. Kaplan-Meier curves suggest that the proposed comprehensive prediction model can offer personalized risk scores which can better group individuals into two groups. The proposed model has a significant impact on population survival times. It can be used as a recommendation system for offering personalized treatments by determining the relationship between a patient's whole slide pathological images and his or her risk of an event (death).

4.3.3.3 High-risk Regions Localization

The most important advantage of DeepAttnMISL is its good interpretability and we create a heatmap by showing the corresponding attention weight of each phenotype cluster. Red color indicates the highest attention weight while blue means the lowest values. From the obtained heatmap, we can see the proposed approach can identify higher risk regions properly because most of patches with high attention weights are from tumor regions. When we look at selected patches from WSISA, we can observe that many patches from non-tumor regions are also selected. That is because WSISA selects clusters based on patches from the whole database and thus it cannot guarantee reliable selection on the specific patient due to the heterogeneity across patients.

We pick one patient as the example to show visualization results. Fig.4.7 presents this patient's all WSIs and the corresponding tumor region annotations. We present the usefulness of the attention mechanism of the proposed model in providing high-risk regions compare with WSISA. Fig.4.8-4.9 shows results WSIs of the same patient. This patient has three whole slide images and two of them have clear tumor regions. Fig.4.8 shows results from the proposed model. The first row shows heatmaps

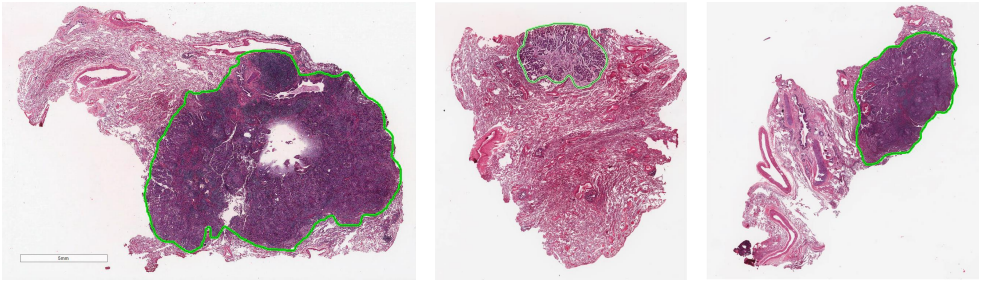

Figure 4.7: WSI Annotations of one example patient.

and the second row shows phenotype pattern distributions on original WSIs. The bottom presents randomly selected patches with higher attention weights (patches with red colors in heatmaps). We can see interested patches are all from tumor regions of WSIs. The comparison visualization from WSISA can be seen in Fig.4.9. Patches from cancerous regions can be grouped in similar clusters but not all of them will be selected as the selection is performed via DeepConvSurv on all patches of the database. Selected phenotypes are more likely discriminated for the whole database with all patients and they are not well interpreted for the specific patient.

Fig.4.10 shows annother whole slide image and the first row presents results from the proposed model. Fig.4.10-(a) visualizes phenotype patterns after clustering on original WSI. Each color represents one type of phenotypes. We plot the corresponding heatmap in (b) and red color means the highest attention weight. The second row shows results from WSISA and Fig.4.10-(d) shows selected pattern after WSISA. It is clear to see that the proposed model can localize tumor regions correctly especially more sampling patches in this case are normal tissues. WSISA cannot properly identifies patches from tumor and normal region. It will treat tumor and normal region patches equally for prediction while our model can assign different weights on them, in which will result in better prediction.

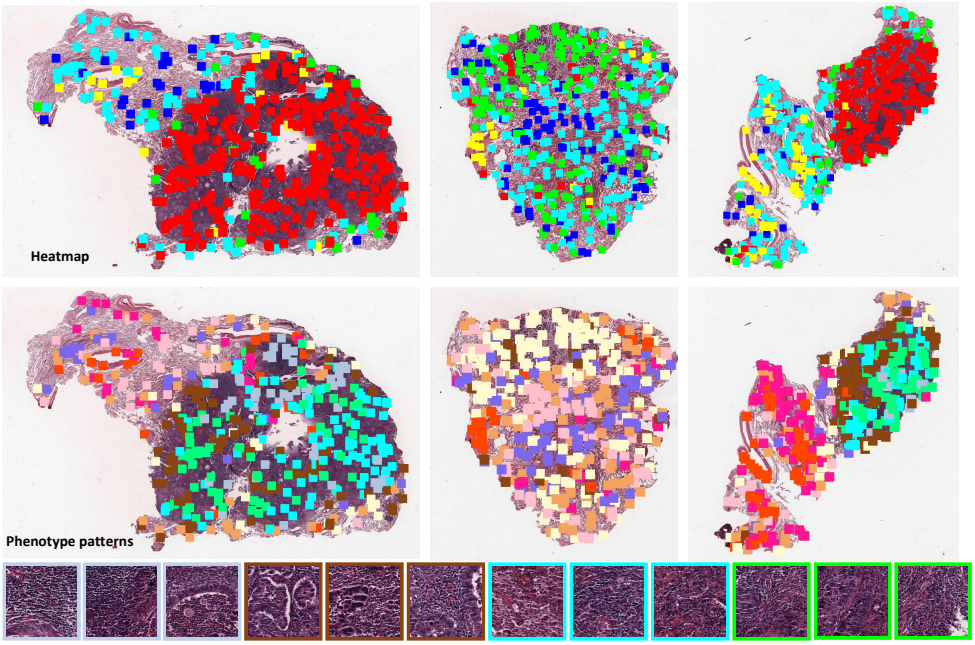

Figure 4.8: Phenotype patterns distribution and the corresponding heatmaps from the proposed model on three WSIs of the same patient. The bottom shows patches from phenotypes with high attention values.

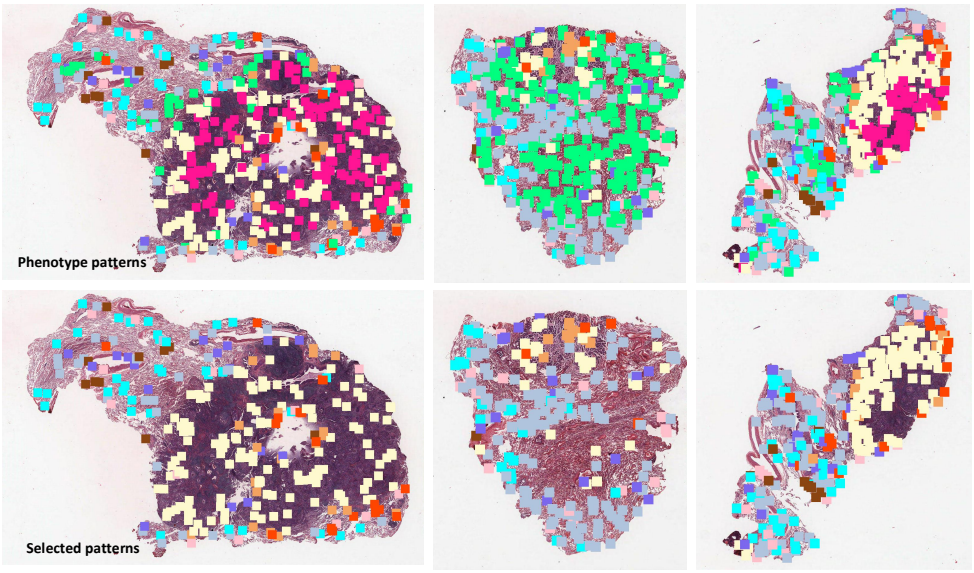

Figure 4.9: Phenotype patterns distribution and selected patterns from WSISA. Missing tumor patches can be observed from selected patterns.

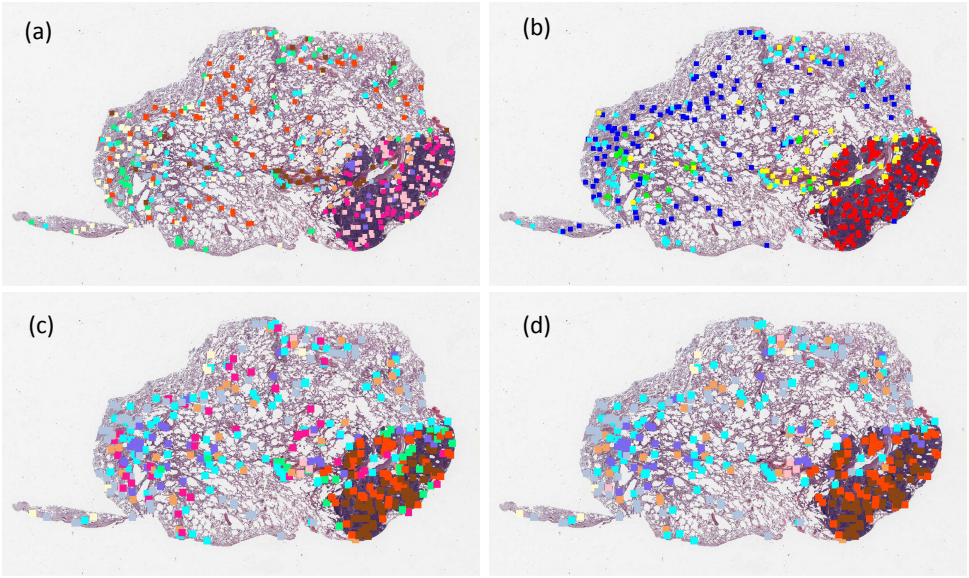

Figure 4.10: (a) Phenotype patterns distribution from our model; (b) Generated heatmap; (c) Phenotype patterns distribution from WSISA; (d) Selected patches from WSISA.

4.4 Conclusion

In this paper, we proposed a deep multi-instance model to directly learn survival patterns from gigapixel images without annotations which make it more easily applicable in large scale cancer dataset. Compared to existing image-based survival models, the developed framework can handle various numbers and sizes whole slide images among different patients. It can learn holistic information of the patient using bag representations and achieve much better performance compared to the ROI patch based methods. Moreover, the flexible and interpretable attention-based MIL pooling can overcome drawbacks from fixed aggregation techniques in state-of-the-art survival learning models. We showed that our approach provides an interpretation of the clinical outcome prediction by presenting reasonable ROIs which is very important in such practical application. Additionally, We illustrated the proposed method can provide personalized treatment for patients and can be used by doctors to guide

their treatment decisions for improving patient lifespan. With future research and development, the proposed approach has the potential to be applied in other tumor types.

CHAPTER 5

AN EFFICIENT ALGORITHM FOR DYNAMIC MRI RECONSTRUCTION

Compared with pathological images, computed tomography (CT) and MRI scans have excellent availability for organs or tumors surveillance. Dynamic magnetic resonance imaging (dMRI) is an important medical imaging technique that has been widely used for multiple clinical applications. However, dynamic MRI is inherently a very slow process due to a combination of different constraints such as nuclear relaxation times and peripheral nerve stimulation. Before using dMRI for clinical applications, a good reconstruction is necessary. In this chapter, we study the problem of dynamic MRI reconstruction. We propose an efficient algorithm for dynamic magnetic resonance (MR) image reconstruction. In comparison with state-of-the-art methods, extensive experiments on single-coil and multi-coil dynamic MR data demonstrate the superior performance of the proposed method in terms of both reconstruction accuracy and time complexity [68, 69].

5.1 Introduction

Since the speed of acquisition in dynamic MRI has physical limits, there exists a trade-off between temporal and spatial resolution. Long scan durations can make patient uncomfortable and also increase the chance of motion artifacts. Hence, many approaches have been proposed to reduce scanning time by requiring partial k-space data for reconstruction instead of full sampling. Popular techniques are echo planar imaging [70] and parallel MR imaging [71, 72, 73, 74, 75] with multiple receiver coils.

In general, when k-space is under-sampled, the Nyquist criterion is violated and the inverse Fourier transform will exhibit aliasing artifacts. Fortunately, it has recently received interest due to the development of Compressive Sensing (CS) theory [76, 77]. CS studies the topic of signal reconstruction from incomplete measurements using the fact that the signal of interest is sparse in its original representation or another domain after applying certain transformations. By incorporating prior information, researchers have proposed different transformations to represent the MR signal [78, 79, 80, 81, 82]. For example, it is possible to reconstruct high quality MR images with the sparsity-induced regularization such as Wavelets [80] or Total Variation [81, 82].

CS-MRI reconstructions typically suffer from artifacts at high undersampling factors with fixed, non-adaptive signal models like wavelets [83]. Therefore, there has been interest in image reconstruction methods where the dictionary is adapted to provide highly sparse representation of data. Recent research has shown benefits for such adaptation of dictionaries in dynamic MRI [83, 84, 85, 86, 87]. These models jointly estimate the image and dictionary for the image patches from under-sampled k-space data. They assume that unknown image patches can be well approximated by a sparse linear combination of the atoms of a learned dictionary. Although these models improve image reconstructions with dictionaries, they are harder than conventional compressed sensing dynamic MRI approaches which take much more time to process. For example, DLTG [87] usually takes much time to process one real dynamic MRI images.

Various alternative models have been explored for dynamic data in recent years. They used one important property that dynamic MRI provides redundant temporal information because it records motions of organ(s). Since the changes of the same organ(s) are subtly slow, dynamic MR frames actually are temporally correlated through

the whole image sequence. Such high correlation in the temporal domain becomes one piece of important prior knowledge for guiding dynamic MRI reconstruction. To use such correlation, Chen et al. [88] applied a sparsity constraint in the temporal domain and proposed Dynamic Total Variation (DTV). Several work have demonstrated the efficacy of low-rank models for dynamic MRI reconstruction [89, 90, 91]. There has been growing interest in decomposing the data into the sum of a low-rank (L) and a sparse (S) component (L+S) [92, 93, 94, 95]. Some other related work consider modeling the dynamic image sequence as both low-rank and sparse (L&S) [96]. In dynamic MRI, since these methods collect the data from all frames in the reconstruction, they can exploit the redundancies of the whole dataset and reconstruct accurate results. However, when the acquired data are contaminated with noise, the sparse prior cannot exploit the local spatial consistency of dynamic MR images and thus make them sensitive to noise and unable to recover clean images.

The limitation of the low-rank regularization in dynamic MR image reconstruction could be remedied by incorporating the piecewise smoothness which can enforce the local spatial consistency during the optimization. One possible choice is total variation (TV) [97] which has been widely used in CS-MRI as the piecewise smoothness constraint of MR images [98, 81] and Dynamic MRI [99, 100]. The joint TV/NN minimization problem may be efficiently solved by popular optimization techniques known as the Fast Composite Splitting Algorithm (FCSA) [81] and Alternating Direction Method of Multipliers (ADMM) [101]. FCSA has been successfully applied in CS-MRI applications, e.g., multi-contrast MRI [102], CS-MRI with tree sparsity [103]. ADMM has been applied for dynamic MRI in k-t SLR [99]. Although the idea of combining low-rank and total variation in a unified framework is intuitive and has been explored in the literature [104, 99], the problem is very difficult to solve because of the non-separability and non-smoothness of the TV and NN term and there

still lack of efficient algorithms to provide theoretical guarantee for dynamic MRI reconstruction.

In this chapter, we propose a Fast algorithm for Total Variation and Nuclear Norm Regularization for dynamic MRI reconstruction (FTVNNR). In our TVNNR model, nuclear norm (NN) exploits the low-rank property of dynamic MR images, while total variation encourages each MR frame's intensities to be locally consistent, which can enforce the piecewise smoothness constraint and make reconstruction more robust to noise. The intuition of combining both TV and NN terms is simple, but the joint TV/NN minimization problem is actually difficult to solve because of the non-separability and non-smoothness of the two terms. A fast algorithm (FTVNNR) is then proposed to efficiently solve this problem. It can obtain a $\mathcal{O}(1/N)$ convergence rate for N iterations. Our approach 1) exploits redundancies in both temporal and spatial domains, 2) has an explicit solution in each step which can be solved inexpensively, and 3) has a theoretically proved convergence rate. Extensive experiments on dynamic MR data demonstrate its superior performance over all previous methods in terms of both reconstruction accuracy and computational complexity.

The rest of this chapter is organized as follows. In Section II, we will give a brief review of the widely used dMRI reconstruction models. The motivation of this work and details can be found in Section III. Experiments on dynamic MR images of both single-coil and parallel imaging can be found in Section IV.

5.2 Related Work

5.2.1 Compressed Sensing Dynamic MRI Reconstruction Approaches

In this section, we describe how recent methods reconstruct dMRI images from a minimum number of samples. At first, we denote one image at time t as $x_t \in \mathbb{C}^{m \times n}$

and $\mathbf{X} = [x_1, x_2, ..., x_T]$ denotes the whole T images. The acquisition domain for MR data is k-space, which is equivalent to the Fourier domain. The dMRI sequence in image space x_t is related to the k-space data by $\hat{x_f} = Fx_t + \epsilon$, where F performs a 2D Discrete Fourier Transform (DFT) on each temporal frame and $\epsilon \in \mathbb{C}^{m \times n}$ is additive white Gaussian acquisition noise. The only data available for reconstruction is under-sampled k-space data, which is a subset Ω of k-space, refereed to $b_t = R_t\hat{x_f}$. R_t denotes the undersampling operator to acquire only a subset of k-space, which contains the rows from the identity matrix that corresponds to the samples of $\hat{x_f}$ that are in Ω. Since this problem is ill-posed and requires regularization, many CS-based methods were proposed to exploit the temporal correlation in dMRI reconstruction. It can be formulated as:

$$\min \Phi(\mathbf{X}) \text{ s.t. } \sum_{t=1}^{T} ||R_t F x_t - b_t||_2^2 \leq \varepsilon \tag{5.1}$$

where Φ denotes the regularization term. Based on Φ, here we review some of the widely used approaches.

Temporal Fourier transform. Temporal Fourier transform is proposed to sparsify periodic motions [79]. That is $\Phi(\mathbf{X}) = ||F_t\mathbf{X}||_1$, where F_t denotes the Fourier transform along the temporal direction, $||\cdot||_1$ denotes the vector ℓ_1 norm. This technique was used in many later works, e.g. [105][106].

Temporal total variation. It assumes that the images change smoothly along the temporal direction [107]. Therefore the gradient along the temporal direction should be small: $\Phi(\mathbf{X}) = ||\nabla_t\mathbf{X}||_1$. In order to achieve the online scheme, Chen et al. [88] extended the temporal TV to dynamic TV by using a reference image x_1 (e.g. the first frame): $\Phi(x_t) = ||x_t - x_1||_{TV}$.

Low rank approximation. Recently, researchers observed that the matrix $\mathbf{X}$ may be usually rank deficient due to the high correlation among different frames. Based on low rank assumption, some methods are proposed in dynamic MRI [91, 99, 100, 93, 92]. To achieve the rank deficient solution, the non-convex Schatten p-norm is used in k-t SLR [99] and locally low rank method [100]. Another type of work [93, 92] focused on the nuclear norm as the convex envelope of rank operator. In this case, $\Phi(\mathbf{X})$ can be defined as $||\mathbf{X}||_*$ where $||\cdot||_*$ denotes the nuclear norm and means the sum of singular value of X.

5.3 Method

5.3.1 Framework

Following the previous notations, we have the undersampling k-space data at time t as

$$b_t = R_t F x_t + \epsilon_t, \tag{5.2}$$

where b_t is the measurement vector which may contain noise (ϵ_t represents noise in k-space).

With prior knowledge in the temporal and spatial domains, it is possible to reconstruct x_t with fewer k-space measurements b_t. Based on a batch scheme, $\mathbf{X} = [\mathbf{Vec}(x_1), \mathbf{Vec}(x_2), ..., \mathbf{Vec}(x_t)] \in \mathbb{C}^{P\times T}$ denotes the whole dynamic MR images. Since dynamic MRI data are complex-valued and we first give the definition of matrix inner product on complex space as $< \mathbf{A}, \mathbf{B} >= tr(\mathbf{A}^\mathsf{H}\mathbf{B})$ where $\mathbf{A}^\mathsf{H}$ denotes the Hermitian transpose of $\mathbf{A}$. The Frobenius norm now is defined as $\|A\|_F = \sqrt{tr(\mathbf{A}^\mathsf{H}\mathbf{A})}$ and thus $\|A\|_F^2 = tr(\mathbf{A}^\mathsf{H}\mathbf{A})$.

The proposed TVNNR model for dMRI reconstruction is defined as follows

$$\min_{\mathbf{X}} \frac{1}{2}\|RF\mathbf{X} - \mathbf{B}\|_F^2 + \lambda_1\|\mathbf{X}\|_{TV} + \lambda_2\|\mathbf{X}\|_*. \tag{5.3}$$

where $\|.\|_*$ is the nuclear norm—the sum of singular values of the matrix $\mathbf{X}$. $\|.\|_{TV}$ denotes the anisotropic total variation of the matrix $\mathbf{X}$. It is defined as $\sum_{t=1}^{T}\sum_{ij}(|\nabla_1 x_{i,j,t}|+|\nabla_2 x_{i,j,t}|)$ where ∇_1 and ∇_2 denote the forward finite difference operators on the first and second coordinates, respectively. If we define $\nabla = [\nabla_1, \nabla_2]$, $\|\mathbf{X}\|_{TV}$ can be simplified as $\|\nabla\mathbf{X}\|_1$. $\mathbf{B} = [b_1, b_2, ..., b_T]^T$, which represents the collection of all the measurements. In (5.3), the Nuclear Norm regularization considers the global information of the sequence, while Total Variation minimization encourages each frame to be locally consistent. The proposed TVNNR model (5.3) combines both types of prior information by exploiting spatial and temporal redundancy to achieve more robust performance.

5.3.2 Optimization

Instead of directly solving the primal problem, we propose to solve a primal-dual form [108, 109] of the original problem (5.3). Motivated by recent algorithms [110, 111] to solve TV regularization using its dual form, we can have the primal-dual form of the primal problem (5.3) by the Legendre-Fenchel transformation of total variation [112, Example.3.26, p. 93] as

$$\min_{\mathbf{X}}\max_{\mathbf{Y}}\frac{1}{2}\|RF\mathbf{X}-\mathbf{B}\|_F^2+\lambda_2\|\mathbf{X}\|_*+\lambda_1\Re\{\langle\nabla\mathbf{X},\mathbf{Y}\rangle\}-I_{B_\infty}(\mathbf{Y}), \tag{5.4}$$

where $\mathbf{Y}$ is the dual variable and $I_{B_\infty}(\mathbf{Y})$ is the indicator function of the ℓ_∞ unit norm ball

$$I_{B_\infty}(\mathbf{Y}) = \begin{cases} 0 & \|\mathbf{Y}\|_\infty \le 1, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.5}$$

First, we denote RF as $\mathcal{A}$. Then we can get

$$\min_{\mathbf{X}}\max_{\mathbf{Y}}\frac{1}{2}\|\mathcal{A}\mathbf{X}-\mathbf{B}\|_F^2+\lambda_2\|\mathbf{X}\|_*+\lambda_1\Re\{\langle\nabla\mathbf{X},\mathbf{Y}\rangle\}-I_{B_\infty}(\mathbf{Y}), \tag{5.6}$$

The min-max problem (5.6) can be solved by a splitting scheme [109] as

$$\mathbf{X}^{n+1} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{t_1}{2}\|\mathcal{A}\mathbf{X} - B\|_F^2 + t_1\lambda_1\Re\{\langle\nabla\mathbf{X}, \mathbf{Y}^n\rangle\} + t_1\lambda_2\|\mathbf{X}\|_* \tag{5.7}$$

$$\mathbf{Y}^{n+1} = \arg\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{Y} - \mathbf{Y}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}) - t_2\lambda_1\Re\{\langle\nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y}\rangle\}, \tag{5.8}$$

where $\mathbf{X}^n, \mathbf{Y}^n$ are the primal and dual variables in the n-th iteration, respectively, and t_1, t_2 denote the corresponding iteration step sizes.

To simplify (5.7), one widely used technique in many similar methods is to approximate the least squares term [113, 114]. Let $f(\mathbf{X}) = \frac{1}{2}\|\mathcal{A}\mathbf{X} - \mathbf{B}\|_F^2$. One can easily verify that $\nabla f(\mathbf{X}) = \mathcal{A}^{\mathsf{H}}(\mathcal{A}\mathbf{X} - B)$ where $\mathcal{A}^{\mathsf{H}}$ is the adjoint operator of $\mathcal{A}$. The (smallest) Lipschitz constant L is given by $L = \lambda_{max}(\mathcal{A}^{\mathsf{H}}\mathcal{A})$ where $\lambda_{max}(.)$ denotes the largest eigenvalue of a linear operator [114].

Following the similar relaxation [114], we can relax (5.7) to

$$\begin{aligned}\mathbf{X}^{n+1} = &\arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}^n\|_F^2 + \frac{t_1}{2}\|\mathcal{A}\mathbf{X}^n - B\|_F^2 + t_1\lambda_1\Re\{\langle\nabla\mathbf{X}, \mathbf{Y}^n\rangle\} \\ &+ \frac{t_1 L}{2}\|\mathbf{X} - \mathbf{X}^n\|_F^2 + t_1\lambda_2\|\mathbf{X}\|_* + t_1\Re\{\langle\mathcal{A}^{\mathsf{H}}(\mathcal{A}\mathbf{X}^n - B), \mathbf{X} - \mathbf{X}^n\rangle\},\end{aligned} \tag{5.9}$$

Omitting the constant term $\frac{t_1}{2}\|\mathcal{A}\mathbf{X}^n - B\|_F^2$ and combining least square terms, it can become

$$\begin{aligned}\mathbf{X}^{n+1} = &\arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - (\mathbf{X}^n - \frac{t_1}{1 + t_1 L}\mathcal{A}^{\mathsf{H}}(\mathcal{A}\mathbf{X}^n - B))\|_F^2 \\ &+ \frac{t_1\lambda_1}{1 + t_1 L}\Re\{\langle\nabla\mathbf{X}, \mathbf{Y}^n\rangle\} + \frac{t_1\lambda_2}{1 + t_1 L}\|\mathbf{X}\|_*.\end{aligned} \tag{5.10}$$

So far, the closed-form solution of (5.10) is still unclear. To continue simplifying the problem, we introduce the adjoint operator of the difference operator. By reformulating the inner product term to its adjoint one, we can convert the problem into a

nuclear norm regularized de-noising problem. First, we revisit the forward difference operator denoted by $\nabla\mathbf{X}$. It is written as

$$\nabla\mathbf{X} = (P, Q),$$

where $P \in \mathbb{C}^{(m-1)\times n}$ and $Q \in \mathbb{C}^{n\times(m-1)}$ are the matrix defined by

$$P_{i,j} = x_{i,j} - x_{i+1,j},$$
$$Q_{i,j} = x_{i,j} - x_{i,j+1}.$$

Thus the dual variable $\mathbf{Y}$ is constructed by the matrix pair (P, Q). By definition, the adjoint operator of ∇ denoted by ∇^{H} satisfies

$$\langle \nabla\mathbf{X}, \mathbf{Y}\rangle = \langle \mathbf{X}, \nabla^{\mathsf{H}}\mathbf{Y}\rangle,$$

where

$$(\nabla^{\mathsf{H}}\mathbf{Y})_{i,j} = (\nabla^{\mathsf{H}}(P, Q))_{i,j} = P_{i,j} + Q_{i,j} - P_{i-1,j} - Q_{i,j-1}. \tag{5.11}$$

Following (5.11), we could simplify problem (5.10) to the de-noising problem:

$$\mathbf{X}^{n+1} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \bar{\mathbf{X}}^n\|_F^2 + \lambda\|\mathbf{X}\|_*, \tag{5.12}$$

where

$$\bar{\mathbf{X}}^n = \mathbf{X}^n - \frac{t_1}{1+t_1L}\mathcal{A}^{\mathsf{H}}(\mathcal{A}\mathbf{X}^n - B) - \frac{t_1\lambda_1}{1+t_1L}\nabla^{\mathsf{H}}\mathbf{Y}^n, \tag{5.13}$$

$\lambda = \frac{t_1\lambda_2}{1+t_1L}$ and $L = \lambda_{max}(\mathcal{A}^{\mathsf{H}}\mathcal{A})$. That's problem (5.7) in this paper. It is not hard to find that the problem has a closed-form solution by *Matrix Shrinkage Operator* [115]. Suppose that $\bar{\mathbf{X}}^n = \mathbf{U}diag(\sigma(\bar{\mathbf{X}}^n))\mathbf{V}^{\mathsf{H}}$ is any singular value decomposition of $\bar{\mathbf{X}}^n$. Then the solution of (5.12) can be obtained by the matrix shrinkage operator as $\mathbf{X}^{n+1} = S_\lambda(\bar{\mathbf{X}}^n) = \mathbf{U}diag(\overline{\sigma}_\lambda(\bar{\mathbf{X}}^n))\mathbf{V}^{\mathsf{H}}$ where $\overline{\sigma}_\lambda(\bar{\mathbf{X}}^n) = max(\sigma(\bar{\mathbf{X}}^n) - \lambda, 0)$.

Then we consider the $\mathbf{Y}$ **subproblem** in (5.8)

$$\mathbf{Y}^{n+1} = \arg\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{Y} - \mathbf{Y}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}) - t_2\lambda_1 \Re\{\langle \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n), \mathbf{Y}\rangle\}, \quad (5.14)$$

After simplification, it becomes

$$\mathbf{Y}^{n+1} = \arg\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{Y} - \bar{\mathbf{Y}}^n\|_F^2 + I_{B_\infty}(\mathbf{Y}), \quad (5.15)$$

where

$$\bar{\mathbf{Y}}^n = \mathbf{Y}^n + t_2\lambda_1 \nabla(2\mathbf{X}^{n+1} - \mathbf{X}^n),$$

The solution of (5.15) can be obtained by the Euclidean projection of $\bar{\mathbf{Y}}^n$ onto a ℓ_∞ unit ball, which can be evaluated by

$$\mathbf{Y}^{n+1} = sgn(\bar{\mathbf{Y}}^n) \cdot \min(|\bar{\mathbf{Y}}^n|, 1). \quad (5.16)$$

where $sgn(x)$ is the sign function; it outputs 1 if $x > 0$, -1 if $x < 0$ and zero otherwise. All the operations in (5.16) are element-wise.

According to the notation, the dimension of input data $\mathbf{X}$ is $P \times T$.

- In the Step 1, the dominate operations are matrix multiplication $\mathcal{A}^{\mathsf{H}}(\mathcal{A}\mathbf{X}^n - B)$ and $\nabla^{\mathsf{H}}\mathbf{Y}^n$. In practice, the operator $\mathcal{A}$ is the partial Fourier transform and performed on $\mathbf{X}$ at every time step, so the cost of this operation is $\mathcal{O}(TP\log P)$ when the Fast Fourier Transform (FFT) is applied. The cost of second linear operation is $\mathcal{O}(TP)$.
- In the Step 2, matrix shrinkage operator requires SVD computation, and its complexity is $\mathcal{O}(T^2P)$ because $P > T$ in our case.
- The Step 3 and 4 include linear and project operations where each has the cost of $\mathcal{O}(TP)$.

Considering the computational cost of each step, the main cost of FTVNNR should be $\mathcal{O}(T^2P)$ in each iteration. A key feature of the FTVNNR is its fast con-

vergence performance and the ergodic convergence rate of FTVNNR is $\mathcal{O}(1/N)$ for N iteration [69].

5.4 Experimental Results

In this section, we first compare the convergence performance of FTVNNR with two very popular algorithms - FCSA and ADMM. Then the proposed method is compared extensively with state-of-the-art schemes using real single coil and multi-coil dynamic MRI. All experiments were conducted with MATLAB R2015a on a standard PC using a single thread of an Intel core i7 4770 3.4GHz CPU and 16.0 GB RAM.

5.4.1 Convergence Performance

The experiments were tested using the simulation data from 2013 ISMRM Challenge[1] Sample case (256×256, 20 frames). This is a test dataset provided for method development and debugging. Fig.5.1(a) shows one frame from the data. In this experiment, we use Cartesian mask with 25% sampling ratio. The stopping criteria for all algorithms is $\|\mathbf{X}^{n+1} - \mathbf{X}^n\|_F / \|\mathbf{X}^n\|_F < 10^{-4}$ with a maximum iteration number of 200. Two parameters are set as $\lambda_1 = 0.01$ and $\lambda_2 = 1$. Two metrics were chosen for quantitative evaluation against fully-sampled reference images: the peak signal-to-noise ratio (PSNR) and high frequency error norm (HFEN) which was used to evaluate the reconstruction of edges and fine structures [83, 84]. In HFEN, the kernel size is 15×15 pixels and the standard deviation is 1.5 pixels.

5.4.1.1 Results

In FCSA, it is time-consuming to solve TV subproblem to achieve good results. We tune the iteration numbers in TV subproblem from 50 to 10 to see if we can reach

[1] http://challenge.ismrm.org/node/53

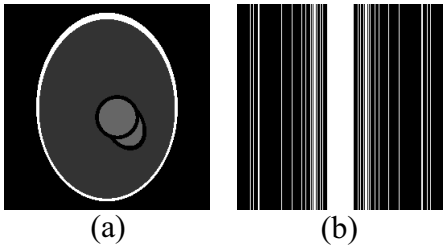

Figure 5.1: One frame of Sample case from 2013 ISMRM Challenge (a). The undersampling mask (b) was applied in k-space.

reasonable results using the minimum iterations. k-t SLR [104] introduced ADMM to solve the TV/NN problem. It splits the joint minimization to five subproblems by Augmented Lagrangian (AL) scheme and needs a conjugate gradient (CG) to exactly solve the first subproblem. We keep the default CG solver parameters in k-t SLR for experiments.

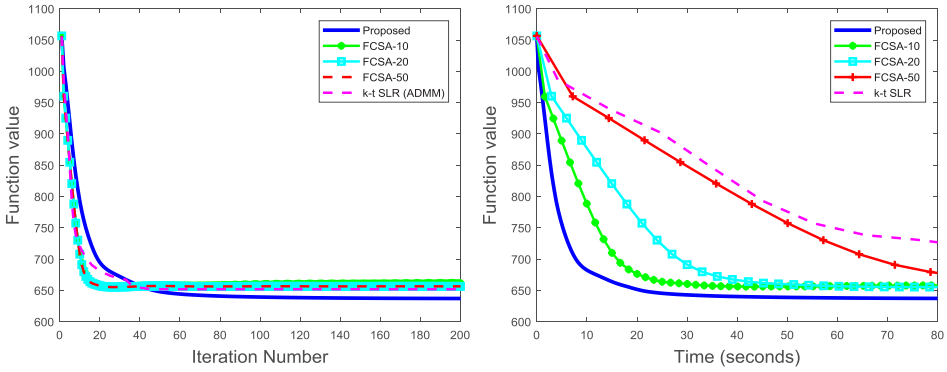

Figure 5.2: The convergence speeds of FCSA, k-t SLR and FTVNNR. Left: Function Value vs. Iteration Number. Right: Function Value vs. CPU Time (s)

Fig.5.2 presents their convergence performances. "FCSA-50" in the figure refers to results from the FCSA using 50 iterations. It can be seen that function values of FCSA and k-t SLR decrease slightly faster than that of FTVNNR in the early stage (fewer than 40 iterations). However, the complexity of each iteration in FCSA and

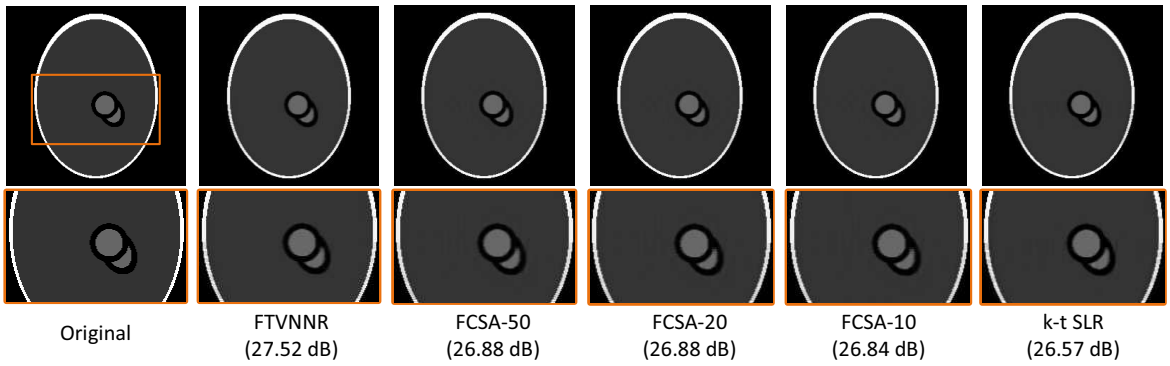

Figure 5.3: The first row shows the reconstructed results, and the second row shows the close-up views of the selected regions.

k-t SLR is much higher than the cost of FTVNNR. The right plot in Fig.5.2 shows the decrease of function values for each method until 80 seconds. We can see that the computational cost of k-t SLR is higher than FTVNNR and FCSA. Even the iteration number is set to 10, FCSA is still slower than the proposed FTVNNR. The proposed method converges much faster than FCSA and ADMM using much smaller computational time. After convergence, the energy function value of the proposed algorithm is smaller than that of FCSA and k-t SLR.

Fig.5.3 presents visual comparisons of the reconstructed 14th frame using different algorithms. It can be seen that even though FCSA and k-t SLR are solving the same optimization, they still cannot achieve better results than FTVNNR. That's because the main subproblem might not be solved exactly while FTVNNR has closed-form solution for each subproblem. From the close-up views of selected regions, one can clearly see that artifacts exist in results from FCSA and k-t SLR while the image from FTVNNR is clean and perfect.

Quantitative evaluations for selected regions on the whole sequence (20 frames) can be seen in Fig.5.4. The proposed FTVNNR outperforms all other comparisons in terms of PSNR and HFEN. All experiments clearly illustrate that the proposed

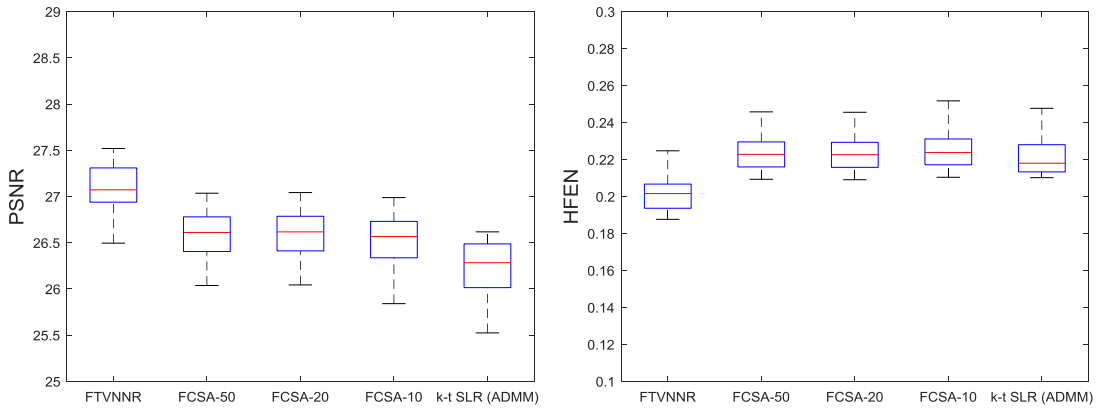

Figure 5.4: Boxplot of PSNR and HFEN results.

algorithm can more solve the TVNNR model much better than other optimization techniques in terms of both efficiency and effectiveness.

Table 5.1 summarizes the computational time and performances of each method. From the table, we can reduce the iteration number to 20 to have best reconstruction in FCSA but the proposed FTVNNR can still reconstruct higher quality images and it is approximately 7 times faster than FCSA.

Table 5.1: Performances of different algorithms (Time: Seconds)

	Proposed	FCSA-50	FCSA-20	FCSA-10	k-t SLR
Time	93	1434	613	339	872
PSNR	27.08(0.30)	26.56(0.29)	26.57(0.29)	26.48(0.35)	26.20 (0.34)
HFEN	0.202(0.011)	0.225(0.011)	0.224(0.011)	0.227(0.012)	0.223(0.011)

5.4.2 Real Data Evaluation

We then explored our method on one real publicly available dataset from [84]. The myocardial perfusion MRI data was acquired using a saturation recovery FLASH sequence (three slices, TR/TE = 2.5/1.5 ms, sat.recovery time = 100 ms, phase ×

frequency encodes × time = 190×90×70). To test the robustness of our method, the k-space data is corrupted with additional complex Gaussian white noises with varying standard deviation. The most practical Cartesian masks with varying sampling ratios were used as the undersampling mask in our experiments. We compared our method with four state-of-the-art methods, the undersampled (k,t)-Space via low-rank plus sparse prior (ktRPCA) [93], blind compressive sensing (BCS) [84], dictionary learning based method DLTG [87] and k-t SLR [99]. The source codes for these methods are downloaded from each author's website. BCS is implemented with both 50 inner and outer iterations. The rest of the parameters in each method is tuned for each dataset separately to achieve the best performance. Similarly, the regularization parameters (λ_1, λ_2) were selected empirically by examining the reconstruction results over a range of possible values. The effect of varying the parameters is discussed later. We choose λ_1=0.03 and λ_2=100 by exploiting the best performances from parameter optimization.

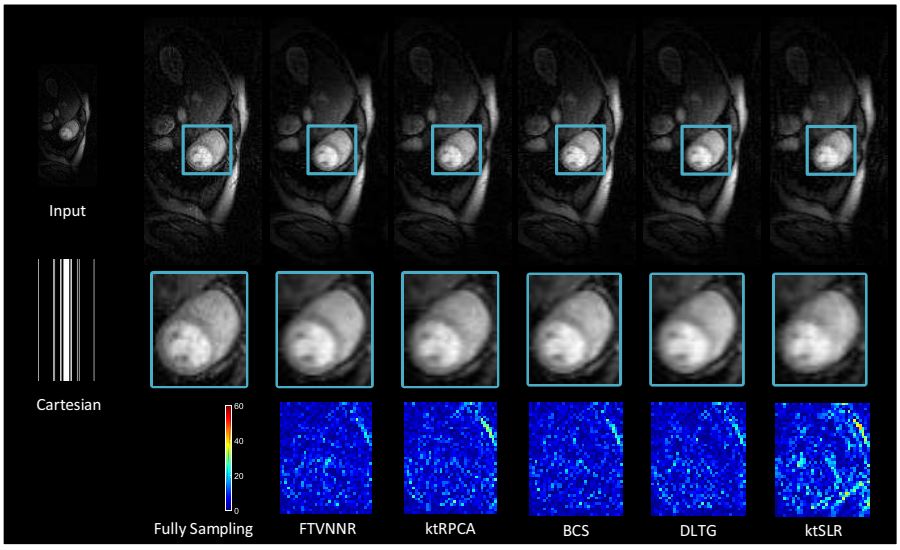

Figure 5.5: Results of the 29th frame of the perfusion sequence at 20% sampling ratio.

Fig.5.5 presents the 29th reconstructed frame of the myocardial perfusion data with 1/5 sampling ratio. Metrics were computed within the manually defined region of interest. For each method, the reconstructed image is presented together with its error. Clear visible artifacts can be observed on the image by k-t SLR. Our approach achieves the lowest reconstruction error among all rest methods.

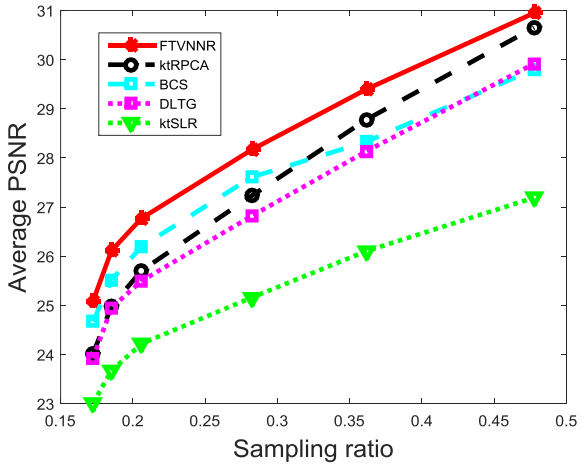

Figure 5.6: Average PSNR with different levels of under-sampling.

Fig.5.6-5.7 present PSNR and HFEN measurements for all methods while changing the sampling ratio from 0.17 to 0.47. It is obvious that the proposed FTVNNR outperforms all other comparison methods in all undersampling cases for both PSNR and HFEN. Compared with the other four methods, the proposed FTVNNR can achieve the best reconstruction with different levels of under-sampling. From the result, it is also observed that our approach is more robust to the changes of sampling ratios, compared to BCS.

To test the reconstruction performance to noise, we added Gaussian white noise with standard deviation $\sigma = \{0.01, 0.03, 0.05, 0.07, 0.09, 0.1\}$ and applied the undersampling mask with 20% ratio. Since DLTG requires much more time (1-2 hours) and

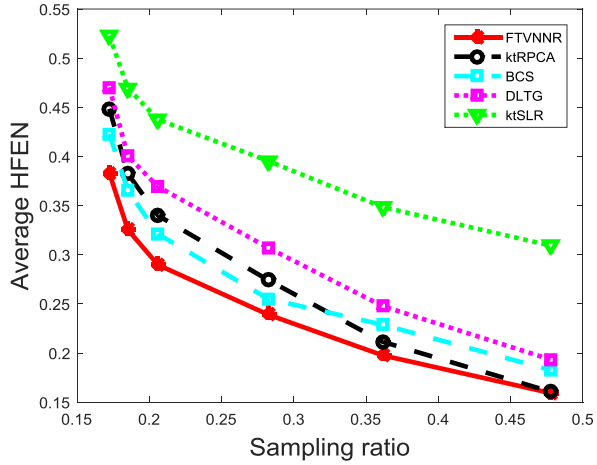

Figure 5.7: Average HFEN with different levels of under-sampling.

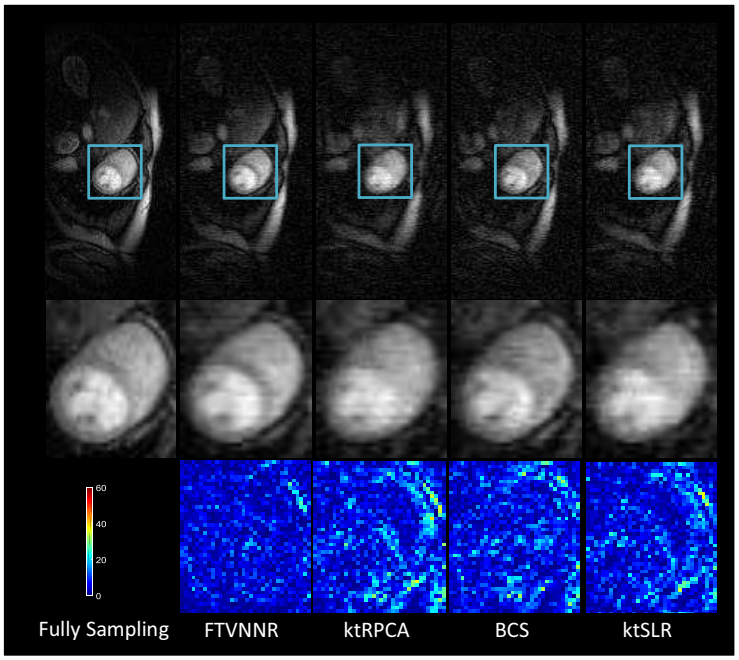

Figure 5.8: Results of 29th frame with $\sigma = 0.05$.

thus we only compare the proposed with other comparison methods. Fig.5.8 shows visual comparisons when using noisy data at $\sigma = 0.05$. It is evident from the error that our method achieves superior visual reconstruction quality. The interest region is zoomed up for better visual inspection. Compared to the original image, results

of ktRPCA and ktSLR appear blurry. BCS provides better reconstruction while the proposed method shows more fine and clear details. From the figures, it can be seen that FTVNNR better preserves the various details in the images including edges and boundaries.

All metrics among timeframes can be found in Fig.5.9. The proposed method outperforms others almost every frame in both PSNR and HFEN. It can be seen that ktRPCA is unable to perform well on noisy data since the sparsity constraint cannot exploit the local spatial consistency or piece-wise smoothness of dynamic MR images.

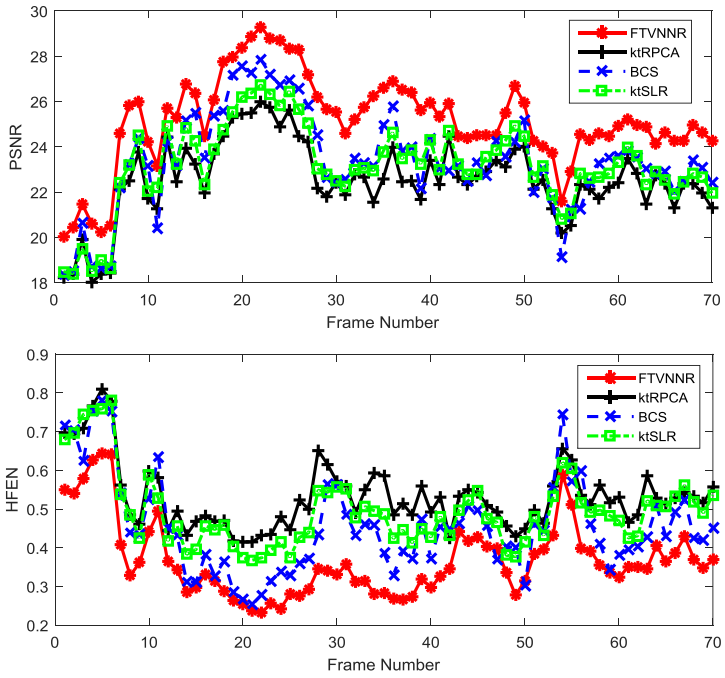

Figure 5.9: PSNR and HFEN metrics among all timeframes

Fig.5.10 demonstrates the results when using noisy data changing σ from 0.01 to 0.1. Performance reduces when noise level increases while the proposed FTVNNR still achieves best results than all comparison methods. That's because the FTVNNR

can utilize the local consistency in the spatial domain which makes it more robust to noise.

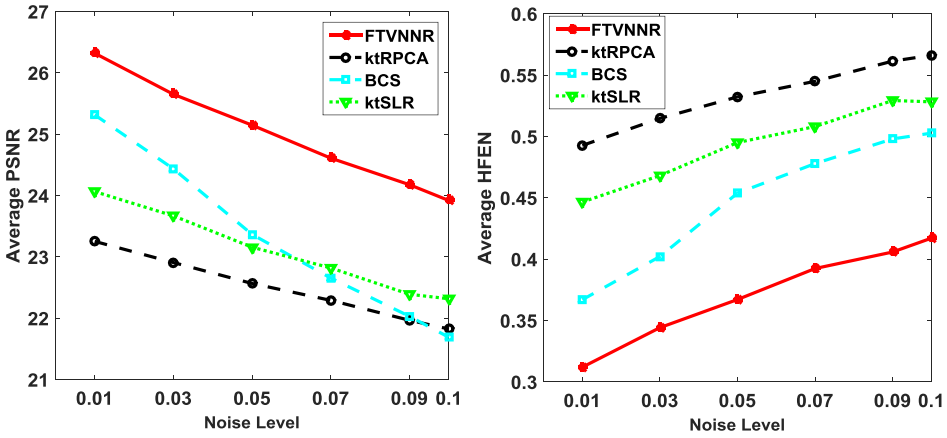

Figure 5.10: Average PSNR and HFEN with different levels of noise.

Running Time: Time usages of different methods in the case of no noise and with noise can be seen in Fig.5.11 and Fig.5.12. Table 5.2 summarizes the execution time of the methods in all cases. We recorded the mean and standard deviation of the different running times for each method.

One can see that DLTG requires nearly 1-2 hours for processing. The proposed method is significantly efficient over other methods, which is almost at least 4 times faster than state-of-the-arts algorithms. Therefore, the proposed method outperforms others in terms of both accuracy and efficiency.

5.4.3 Parallel Imaging

Although the problem (5.3) is the single coil case, it has the potential to process multi-coil parallel MRI data. When the coil sensitivities are available, it can be combined with SENSE in the k-t SPARSE-SENSE framework [106] by multiplying

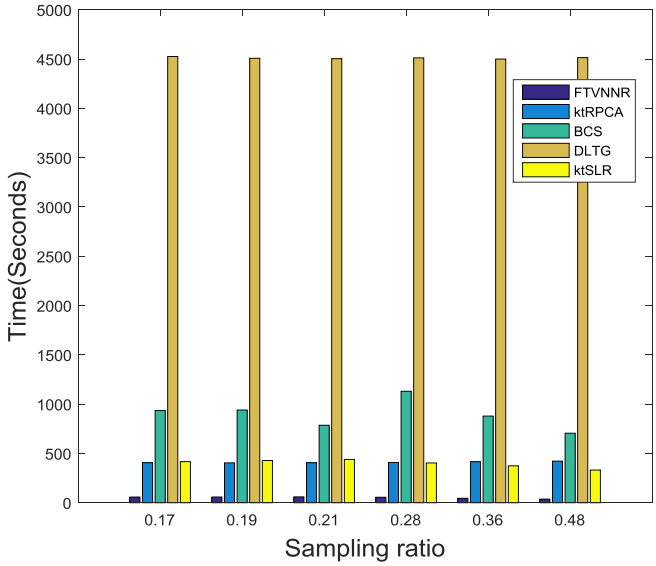

Figure 5.11: CPU Time for each method with different sampling ratios.

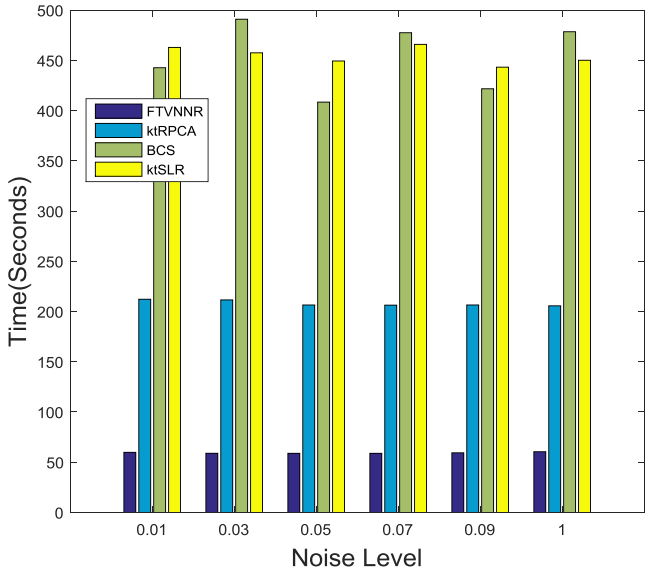

Figure 5.12: CPU Time for each method with different noise levels.

coil sensitivities E after the undersampled Fourier transform, which means the least square term in (5.3) will be $||RFE\mathbf{X} - \mathbf{B}||_F^2$.

Table 5.2: The average time cost of different methods (Seconds)

Methods	No Noise	Noise
FTVNNR	52.64 ± 9.17	59.70 ± 0.99
ktRPCA	410.64 ± 7.14	208.26 ± 2.70
BCS	896.53 ± 146.41	509.34 ± 151.41
DLTG	4511.1 ± 8.98	–
ktSLR	399.26 ± 39.89	453.03 ± 9.32

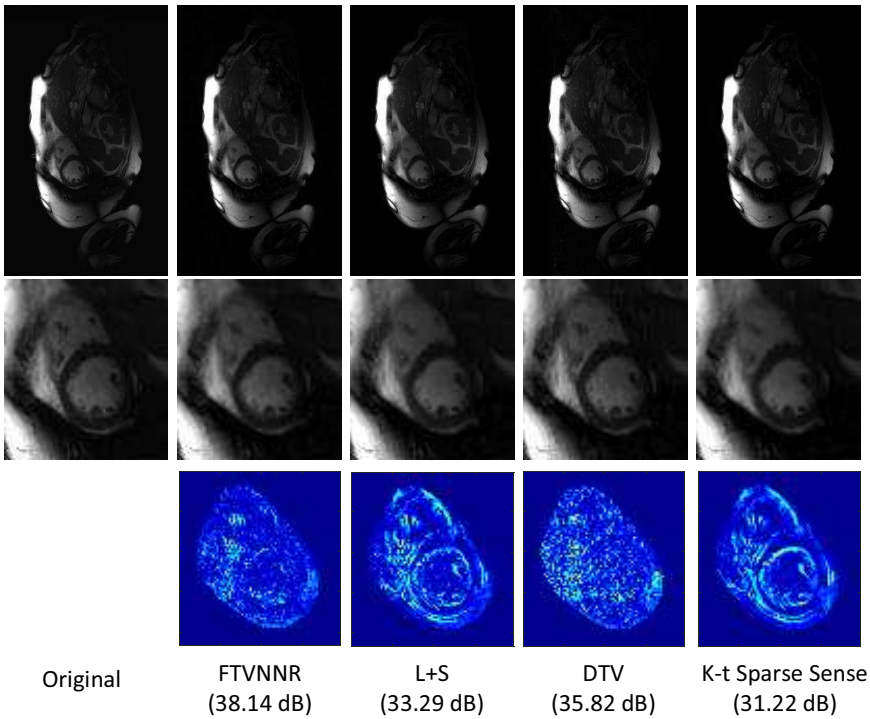

Figure 5.13: Comparison of the reconstruction results from the 3rd frame. The radial mask with the sampling rate of 0.10 is used. The first row shows whole images. The second row shows images from ROIs and the third row shows the corresponding error images.

To further evaluate performances, we used one fully-sampled cardiac cine data distributed by the 2013 ISMRM Recon Challenge committee[2]. The data was collected using a 2D cine breath-held bSSFP sequence with 32-channel cardiac receiver coils. Scan parameters were spatial resolution $1 \times 1mm^2$, matrix size $346 \times 210 \times 27$. The

[2]http://challenge.ismrm.org/node/66

data was retrospectively under-sampled using Cartesian golden-angle radial sampling patterns with the acceleration factors ranged from 5 to 30 (sampling ratio from 1/5 to 1/30). We compared the proposed method with three state-of-the-art parallel MRI approaches including low-rank plus sparse reconstruction (L+S) [92], dynamic Total Variation (DTV) [88] and k-t SPARSE-SENSE [106]. For all methods, we tune parameters to achieve the best result under the 1/30 sampling rate and then perform on other cases using these parameters. The stopping criteria for all methods is 10^{-4} with a maximum iteration number of 50. All quantitative evaluations are calculated within the Region of Interest (ROI).

Reconstruction results at the sampling ratio 10% are shown in Fig.5.13. When looking at details of the cardiac region, it can be observed that FTVNNR presents less noisy and more clear results because it can utilize the local consistency in the spatial domain while the temporal FFT in k-t SPARSE-SENSE and sparse prior in L+S cannot exploit the spatial sparsity. PSNR value of each time frame can be seen in Fig.5.14. It can be seen that the proposed FTVNNR outperforms other state-of-the-arts parallel dynamic MRI methods in each time frame.

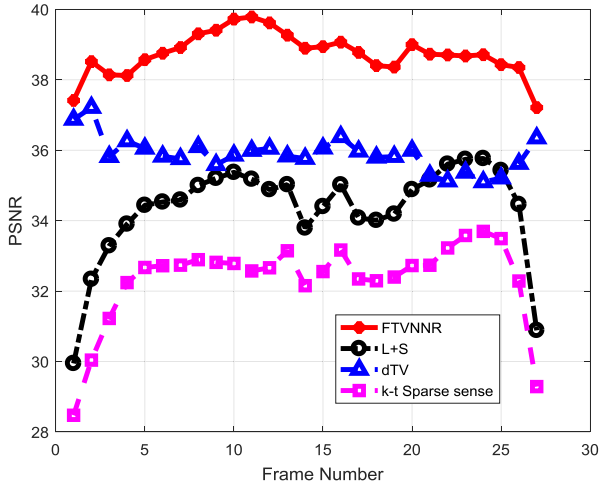

Figure 5.14: Results of every frame at the 10% sampling rate.

Fig.5.15 depicts the PSNR of the reconstructed images at different sampling rates. DTV performs the worst when the sampling ratio is very low. That's because DTV needs a relative high sampling rate at the first frame to reconstruct the reference image. If the high quality reference image cannot be guaranteed, it will not produce satisfactory dynamic MR sequence.

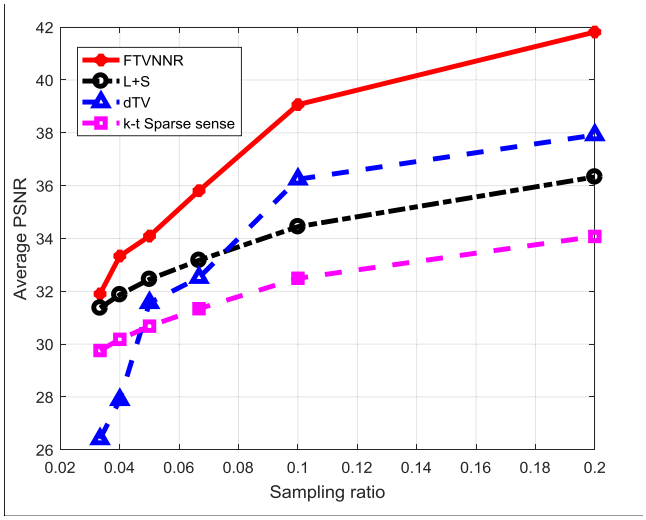

Figure 5.15: Results with different levels of under-sampling.

The average running time of all methods under different sampling rates can be found in Table 5.3. One can see that the proposed method has the fastest reconstruction speed compared to others, due to its fewer iterations and faster convergence.

Table 5.3: The average time cost on different sampling ratios. "ktSS" is k-t SPARSE-SENSE

Time (Seconds)	Proposed	L+S	ktSS	DTV
Data ($346 \times 210 \times 27$)	511.33	539.36	1298.6	960.81

5.5 Conclusion

We have proposed an efficient algorithm for dynamic MRI. The contributions of our work are as follows. First, the proposed FTVNNR can achieve lower computation cost at each iteration than other popular optimization methods such as FCSA and ADMM. The convergence rate can be theoretically proved as $\mathcal{O}(1/N)$. Second, the proposed FTVNNR achieves the best reconstruction performance when compared to state-of-the art methods. Also, experiments demonstrate that it is faster than other dMRI methods. These properties make the proposed method more powerful than conventional dMRI methods in terms of both accuracy and time efficiency. Moreover, the proposed method can be easily extended to parallel MRI. The parallel version of FTVNNR can also share good properties like fast convergence. Numerous experiments were conducted to show its better performance. In our future work, we will investigate on using reconstructed dynamic MRI for clinical outcome prediction.

CHAPTER 6

CONCLUSION AND FUTURE WORK

The aim of this thesis is to present learning techniques for large-scale medical data. We investigated several typical types of medical data in important applications including 1) survival analysis; 2) image reconstruction.

We have demonstrated our deep learning and machine learning approaches formed effective and efficient solutions with clear improvements in extensive experiments on different data types. Specifically, we have developed the following methods:

Imaging biomarker discovery for survival prediction. We investigated subtype cellular information and proposed a pipeline for predicting patients survival. We adopted a deep learning based subtype cell detection method to detect cells into different types. Then subtype cellular information and features are collected and imaging biomarkers are searched with traditional survival regression methods. As subtype cellular information can better describe tumor morphology, our results have shown that those imaging biomarkers can provide more accurate prediction than state-of-the-art method using traditional imaging and molecular profilers. In the future, we will try to find more quantitative measurements to better describe tumor morphology and further improve the prediction performances.

Deep correlational learning for integrating multi-modality data. A deep correlational survival model (DeepCorrSurv) is proposed to efficiently integrate multi-modalities censored data with small samples. In the literature, one challenge is the view-discrepancy between different views in recent real cancer databases. To eliminate the view discrepancy between imaging data and molecular profiling data,

deep correlational learning provides a good solution to maximize the correlation of two views and find the common embedding space. The proposed DeepCorrSurv transfers knowledge from the embedding space and fine-tunes the whole network using survival loss. Experiments have shown that DeepCorrSurv can discover important markers that are ignored by correlational learning and extract the best representation for survival prediction. In the future, this framework will be introduced for other kinds of data.

Attention Guided Multi-instance networks for survival prediction using Whole Slide Images. Above two presented works need image patches extracted from ROIs. Nowadays, weakly-supervised learning using Whole Slide Images (WSIs) for survival prediction attracts much attentions. This kind of methods don't need to use ROI annotations but it should have the ability to fuse results from sampling patches. We proposed a attention guided deep multi-instance model to directly learn survival patterns from gigapixel images without annotations which make it more easily applicable in large scale cancer dataset. The flexible and interpretable attention-based MIL pooling can overcome drawbacks from fixed aggregation techniques in state-of-the-art survival learning models. We showed that our approach provides an interpretation of the clinical outcome prediction by presenting reasonable ROIs which is very important in such practical application.

Dynamic MRI reconstruction using total variation and nuclear norm regularizations. Dynamic MRI is one of the most widely utilized data source for organ surveillance owing to its high diagnostic performance and excellent availability. Reducing the number of k-space measurements is a standard way of speeding up the dynamic MRI examination time. However, undersampled k-space often exhibit blur or aliasing effects and this will make them unsuitable for clinical use. The goal of the reconstruction is to restore a high fidelity image from partially observed measurements

for future monitoring. We proposed an efficient dynamic MRI reconstruction that can both keep the low-rank property and piece-wise smoothness of dynamic MRI images making reconstruction more robust to noise. Experiments have shown that the proposed algorithm can reconstruct very high quality images with much fewer time compared with recent reconstruction algorithms. The reconstructed MRI images can be further used for monitoring.

REFERENCES

[1] H. Wang, F. Xing, H. Su, A. Stromberg, and L. Yang, "Novel image markers for non-small cell lung cancer classification and survival prediction," *BMC Bioinformatics*, vol. 15, no. 1, p. 310, 2014. [Online]. Available: http://www.biomedcentral.com/1471-2105/15/310

[2] P. Wang, Y. Li, and C. K. Reddy, "Machine learning for survival analysis: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, p. 110, 2019.

[3] A. Warth, T. Muley, M. Meister, A. Stenzinger, M. Thomas, P. Schirmacher, P. A. Schnabel, J. Budczies, H. Hoffmann, and W. Weichert, "The novel histologic international association for the study of lung cancer/american thoracic society/european respiratory society classification system of lung adenocarcinoma is a stage-independent predictor of survival," *Journal of clinical oncology*, pp. JCO–2011, 2012.

[4] Y. Yuan, H. Failmezger, O. M. Rueda, H. R. Ali, S. Gräf, S.-F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell, *et al.*, "Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling," *Science translational medicine*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.

[5] J. Yao, S. Wang, X. Zhu, and J. Huang, "Imaging biomarker discovery for lung cancer survival prediction," in *MICCAI*. Springer International Publishing, 2016, pp. 649–657.

[6] K.-H. Yu, C. Zhang, G. J. Berry, R. B. Altman, C. R, D. L. Rubin, and M. Snyder, "Predicting non-small cell lung cancer prognosis by fully automated micro-

scopic pathology image features," *Nature Communications*, vol. 7, no. 12474, 2016.

[7] A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn, and D. Koller, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Science translational medicine*, vol. 3, no. 108, pp. 108ra113–108ra113, 2011.

[8] X. Zhu, J. Yao, F. Zhu, and J. Huang, "WSISA: Making survival prediction from whole slide histopathological images," in *CVPR*, 2017, pp. 7234–7242.

[9] B. Tang, A. Li, B. Li, and M. Wang, "Capsurv: Capsule network for survival analysis with whole slide pathological images," *IEEE Access*, 2019.

[10] K. Shedden, J. M. Taylor, S. A. Enkemann, M.-S. Tsao, T. J. Yeatman, W. L. Gerald, S. Eschrich, I. Jurisica, T. J. Giordano, D. E. Misek, *et al.*, "Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study," *Nature medicine*, vol. 14, no. 8, pp. 822–827, 2008.

[11] Y. Yuan, E. M. Van Allen, L. Omberg, N. Wagle, A. Amin-Mansour, A. Sokolov, L. A. Byers, Y. Xu, K. R. Hess, L. Diao, *et al.*, "Assessing the clinical utility of cancer genomic and proteomic data across tumor types," *Nature biotechnology*, vol. 32, no. 7, pp. 644–652, 2014.

[12] R. Tibshirani *et al.*, "The lasso method for variable selection in the cox model," *Statistics in medicine*, vol. 16, no. 4, pp. 385–395, 1997.

[13] J. Barker, A. Hoogi, A. Depeursinge, and D. L. Rubin, "Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles," *Medical Image Analysis*, vol. 30, pp. 60 – 71, 2016.

[14] X. Zhu, J. Yao, X. Luo, G. Xiao, Y. Xie, A. Gazdar, and J. Huang, "Lung cancer survival prediction from pathological images and genetic data - an integration study," in *ISBI*, 2016, pp. 1173–1176.

[15] S. Wang, J. Yao, Z. Xu, and J. Huang, "Subtype cell detection with an accelerated deep convolution neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 640–648.

[16] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *MICCAI 2015*, ser. Part III. LNCS, N. Navab, J. Hornegger, M. W. Wells, and F. A. Frangi, Eds. Heidelberg: Springer, 2015, vol. 9351, pp. 358–365.

[17] Z. Xu and J. Huang, "Efficient lung cancer cell detection with deep convolution neural network," in *Patch-Based Techniques in Medical Imaging*, G. Wu, P. Coupé, Y. Zhan, B. Munsell, and D. Rueckert, Eds. Heidelberg: Springer, 2015, vol. 9467, pp. 79–86.

[18] H. Li, R. Zhao, and X. Wang, "Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification," *arXiv preprint arXiv:1412.4526*, 2014.

[19] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.

[20] J. Yao, D. Ganti, X. Luo, G. Xiao, Y. Xie, S. Yan, and J. Huang, "Computer-assisted diagnosis of lung cancer using quantitative topology features," in *Machine Learning in Medical Imaging*, ser. Lecture Notes in Computer Science, L. Zhou, L. Wang, Q. Wang, and Y. Shi, Eds. Springer International Publishing, 2015, vol. 9352, pp. 288–295.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[22] H. Binder and M. Schumacher, "Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models," *BMC Bioinformatics*, vol. 9, no. 1, pp. 1–10, 2008.

[23] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests," *The annals of applied statistics*, pp. 841–860, 2008.

[24] P. J. Heagerty and Y. Zheng, "Survival model predictive accuracy and roc curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, 2005.

[25] W. D. Travis and C. Harris, "Pathology and genetics of tumours of the lung, pleura, thymus and heart," 2004.

[26] J. Yao, X. Zhu, F. Zhu, and J. Huang, "Deep correlational learning for survival prediction from multi-modality data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 406–414.

[27] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 187–220, 1972.

[28] E. Bair and R. Tibshirani, "Semi-supervised methods to predict patient survival from gene expression data," *PLoS Biol*, vol. 2, no. 4, p. E108, 2004.

[29] E. Bair, T. Hastie, D. Paul, and R. Tibshirani, "Prediction by supervised principal components," *Journal of the American Statistical Association*, vol. 101, no. 473, 2006.

[30] H. M. Bøvelstad, S. Nygård, H. L. Størvold, M. Aldrin, Ø. Borgan, A. Frigessi, and O. C. Lingjærde, "Predicting survival from microarray dataa comparative study," *Bioinformatics*, vol. 23, no. 16, pp. 2080–2087, 2007.

[31] J. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "Deep survival: A deep cox proportional hazards network," *arXiv preprint arXiv:1606.00931*, 2016.

[32] X. Zhu, J. Yao, G. Xiao, Y. Xie, J. Rodriguez-Canales, E. R. Parra, C. Behrens, I. I. Wistuba, and J. Huang, "Imaging-genetic data mapping for clinical outcome prediction via supervised conditional gaussian graphical model," in *Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on*. IEEE, 2016, pp. 455–459.

[33] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[34] G. Andrew, R. Arora, J. A. Bilmes, and K. Livescu, "Deep canonical correlation analysis." in *ICML*, 2013, pp. 1247–1255.

[35] X. Zhu, J. Yao, and J. Huang, "Deep convolutional neural network for survival analysis with pathological images," in *BIBM*. IEEE, 2016, pp. 544–547.

[36] C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, *et al.*, "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013.

[37] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat, *et al.*, "Cellprofiler: image analysis software for identifying and quantifying cell phenotypes," *Genome biology*, vol. 7, no. 10, p. R100, 2006.

[38] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data*. John Wiley & Sons, 2011, vol. 360.

[39] A. Mayr and M. Schmid, "Boosting the concordance index for survival data–a unified framework to derive and evaluate biomarker combinations," *PloS one*, vol. 9, no. 1, p. e84483, 2014.

[40] Y. Li, J. Wang, J. Ye, and C. K. Reddy, "A multi-task learning formulation for survival analysis," in *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, 2016.

[41] L. Du, H. Huang, J. Yan, S. Kim, S. L. Risacher, M. Inlow, J. H. Moore, A. J. Saykin, L. Shen, A. D. N. Initiative, *et al.*, "Structured sparse canonical correlation analysis for brain imaging genetics: an improved graphnet method," *Bioinformatics*, vol. 32, no. 10, p. 1544, 2016.

[42] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints*, vol. abs/1605.02688, May 2016. [Online]. Available: http://arxiv.org/abs/1605.02688

[43] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[44] M. N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, p. 147, 2009.

[45] J. Cheng, X. Mo, X. Wang, A. Parwani, Q. Feng, and K. Huang, "Identification of topological features in renal tumor microenvironment associated with patient survival," *Bioinformatics*, vol. 34, no. 6, pp. 1024–1030, 2017.

[46] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. V. Vega, D. J. Brat, and L. A. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[49] B. Kong, X. Wang, Z. Li, Q. Song, and S. Zhang, "Cancer metastasis detection via spatially structured deep network," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 236–248.

[50] Y. Li and W. Ping, "Cancer metastasis detection with neural conditional random field," in *Medical Imaging with Deep Learning*, 2018.

[51] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P.-A. Heng, "Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE transactions on medical imaging*, 2019.

[52] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in neural information processing systems*, 2017, pp. 3856–3866.

[53] R. Li, J. Yao, X. Zhu, Y. Li, and J. Huang, "Graph cnn for survival analysis on whole slide pathological images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 174–182.

[54] C. K. Reddy and Y. Li, "A review of clinical prediction models," in *Healthcare Data Analytics*. Chapman and Hall/CRC, 2015, pp. 343–378.

[55] Y. Xu, J.-Y. Zhu, E. Chang, and Z. Tu, "Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering," in *CVPR*. IEEE, 2012, pp. 964–971.

[56] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification." in *CVPR*, 2016, pp. 2424–2433.

[57] C. Mercan, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 316–325, 2018.

[58] T. Zeng and S. Ji, "Deep convolutional neural networks for multi-instance multi-task learning," in *ICDM*. IEEE, 2015, pp. 579–588.

[59] H. Steck, B. Krishnapuram, C. Dehing-oberije, P. Lambin, and V. C. Raykar, "On ranking in survival analysis: Bounds on the concordance index," in *Advances in neural information processing systems*, 2008, pp. 1209–1216.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[61] H. Yang, J. T. Zhou, J. Cai, and Y. S. Ong, "MIML-FCN+: Multi-instance multi-label learning via fully convolutional networks with privileged information," in *CVPR*, 2017, pp. 1577–1585.

[62] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," *arXiv preprint arXiv:1802.04712*, 2018.

[63] Z. Lin, M. Feng, C. N. d. Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding," *arXiv preprint arXiv:1703.03130*, 2017.

[64] C. Raffel and D. P. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," *arXiv preprint arXiv:1512.08756*, 2015.

[65] N. L. S. T. R. Team *et al.*, "The national lung screening trial: overview and study design," *Radiology*, 2011.

[66] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Efficient multiple instance convolutional neural networks for gigapixel resolution image classification," *arXiv preprint arXiv:1504.07947*, p. 7, 2015.

[67] E. T. Lee and J. Wang, *Statistical methods for survival data analysis*. John Wiley & Sons, 2003, vol. 476.

[68] J. Yao, Z. Xu, X. Huang, and J. Huang, "Accelerated dynamic MRI reconstruction with total variation and nuclear norm regularization," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*. Springer International Publishing, 2015, pp. 635–642.

[69] ——, "An efficient algorithm for dynamic mri using low-rank and total variation regularizations," *Medical image analysis*, vol. 44, pp. 14–27, 2018.

[70] P. Mansfield, "Multi-planar image formation using nmr spin echoes," *Journal of Physics C: Solid State Physics*, vol. 10, no. 3, p. L55, 1977. [Online]. Available: http://stacks.iop.org/0022-3719/10/i=3/a=004

[71] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays," *Magnetic Resonance in Medicine*, vol. 38, no. 4, pp. 591–603, 1997.

[72] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger, *et al.*, "Sense: sensitivity encoding for fast mri," *Magnetic resonance in medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[73] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (grappa)," *Magnetic resonance in medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.

[74] D. J. Larkman and R. G. Nunes, "Parallel magnetic resonance imaging," *Physics in medicine and biology*, vol. 52, no. 7, p. R15, 2007.

[75] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling

for fast and flexible dynamic volumetric MRI," *Magnetic resonance in medicine*, vol. 72, no. 3, pp. 707–717, 2014.

[76] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[77] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[78] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008.

[79] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proceedings of the Annual Meeting of ISMRM*, 2006, p. 2420.

[80] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[81] J. Huang, S. Zhang, and D. Metaxas, "Efficient MR image reconstruction for compressed MR imaging," *Medical Image Analysis*, vol. 15, no. 5, pp. 670–679, 2011.

[82] J. Huang, S. Zhang, H. Li, and D. Metaxas, "Composite splitting algorithms for convex optimization," *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1610–1622, 2011.

[83] S. Ravishankar and Y. Bresler, "Mr image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.

[84] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 6, pp. 1132–1145, 2013.

[85] S. Ravishankar and Y. Bresler, "Multiscale dictionary learning for mri," in *Proc. ISMRM*, 2011, p. 2830.

[86] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X. P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing mri," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5007–5019, Dec 2014.

[87] J. Caballero, A. N. Price, D. Rueckert, and J. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 979–994, 2014.

[88] C. Chen, Y. Li, L. Axel, and J. Huang, "Real time dynamic MRI with dynamic total variation," in *MICCAI 2014*, ser. Part I. LNCS, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds. Springer, Heidelberg, 2014, vol. 8673, pp. 138–145.

[89] Z.-P. Liang, "Spatiotemporal imagingwith partially separable functions," in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2007, pp. 988–991.

[90] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t pca: Temporally constrained k-t blast reconstruction using principal component analysis," *Magnetic resonance in medicine*, vol. 62, no. 3, pp. 706–716, 2009.

[91] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*. IEEE, 2010, pp. 996–999.

[92] R. Otazo, E. Cands, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[93] B. Trémoulhéac, N. Dikaios, D. Atkinson, and S. Arridge, "Dynamic MR image reconstruction-separation from under-sampled (k,t)-space via low-rank plus sparse prior," *IEEE Transactions on Medical Imaging*, vol. 33, no. 8, pp. 1689–1701, 2014.

[94] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2502–2514, 2015.

[95] J. Yao, X. Liu, and C. Qi, "Foreground detection using low rank and structured sparsity," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–6.

[96] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled-space data with joint partial separability and sparsity constraints," *IEEE transactions on medical imaging*, vol. 31, no. 9, pp. 1809–1820, 2012.

[97] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[98] F. Shi, J. Cheng, L. Wang, P.-T. Yap, and D. Shen, "Lrtv: MR image super-resolution with low-rank and total variation regularizations," *Medical Imaging, IEEE Transactions on*, vol. 34, no. 12, pp. 2459–2466, 2015.

[99] S. G. Lingala, Y. Hu, E. Dibella, and M. Jacob, "Accelerated first pass cardiac perfusion mri using improved k- t slr," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1280–1283.

[100] X. Miao, S. G. Lingala, Y. Guo, T. Jao, M. Usman, C. Prieto, and K. S. Nayak, "Accelerated cardiac cine MRI using locally low rank and finite difference constraints," *Magnetic resonance imaging*, vol. 34, no. 6, pp. 707–714, 2016.

[101] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[102] J. Huang, C. Chen, and L. Axel, "Fast multi-contrast MRI reconstruction," *Magnetic resonance imaging*, vol. 32, no. 10, pp. 1344–1352, 2014.

[103] C. Chen and J. Huang, "Compressive sensing mri with wavelet tree sparsity," in *Advances in neural information processing systems*, 2012, pp. 1115–1123.

[104] S. G. Lingala, Y. Hu, E. Dibella, and M. Jacob, "Accelerated first pass cardiac perfusion mri using improved k- t slr," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on.* IEEE, 2011, pp. 1280–1283.

[105] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t focuss: A general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009.

[106] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 767–776, 2010.

[107] J. Caballero, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity in dynamic MRI," in *Proceedings of MICCAI*, 2012, pp. 256–263.

[108] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[109] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 119–149, 2012.

[110] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.

[111] A. Chambolle and T. Pock, "On the ergodic convergence rates of a first-order primal-dual algorithm." Sept. 2015, to appear in Math. Programm. A. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01151629

[112] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[113] Y. Nesterov, *Introductory lectures on convex optimization*. Springer Science & Business Media, 2004, vol. 87.

[114] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[115] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

BIOGRAPHICAL STATEMENT

Jiawen Yao received his Ph.D. in Computer Science from the University of Texas at Arlington at 2019. Prior to the Ph.D. program in UTA, He received his M.Eng in Signal Processing and B.Eng degree in Information Engineering from Xi'an Jiaotong University, China in 2014 and 2011, respectively. His main research interests are deep learning, machine learning, and medical image analysis. During his Ph.D. study, he has published more than 19 papers in the world leading conferences and journals in the literature such as the Annual International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), IEEE Conference on Computer Vision and Pattern Recognition (CVPR) and Medical Image Analysis (MedIA).