



دانشگاه صنعتی شریف
دانشکده‌ی مهندسی برق

پایان‌نامه‌ی کارشناسی ارشد
بیوالکتریک

عنوان:

قطعه بندی ساختارهای در ریسک در تصاویر پزشکی با
استفاده از روش‌های یادگیری ماشین

نگارش:

رضا کریم‌زاده

استاد راهنما:

دکتر عماد الدین فاطمی‌زاده

استاد مشاور:

دکتر حسین عربی

اردیبهشت ۱۴۰۱

اللهُ أَكْبَرُ

به نام خدا

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی برق

پایان‌نامه‌ی کارشناسی ارشد

عنوان: قطعه بندی ساختارهای در ریسک در تصاویر پزشکی با استفاده از روش‌های
یادگیری ماشین

نگارش: رضا کریم‌زاده

کمیته‌ی ممتحنین

استاد راهنما: دکتر عmad الدین امضاء:
فاطمی‌زاده

استاد مشاور: دکتر حسین عربی امضاء:

استاد مدعو: استاد ممتحن امضاء:

تاریخ:

سپاس

از اساتید بزرگوار و با دانش، دکتر فاطمیزاده و دکتر عربی، که در این راه نه تنها به من درس علم آموختند، بلکه رسم زندگی و اخلاق را در منش این دو عزیز می‌توان دید و فراگرفت، بسیار سپاس‌گزارم.
امیدوارم این پایان‌نامه بتواند جیران زحمات ایشان باشد.

این پایان‌نامه را به پدر و مادرم تقدیم می‌کنم که در طی این سال‌ها پشت و پناه و یاورم بوده‌اند.

چکیده

برای رادیوتراپی و از بین بردن بافت‌های سرطانی لازم است پیش از درمان و تاباندن پرتوی پرانرژی، مکان تومور و ساختارهای آسیپ پذیر اطراف تومور با دقت بالایی مشخص شوند. برای این کار لازم است تصاویر دریافت شده از بیمار قطعه‌بندی شوند. این کار معمولاً به صورت دستی انجام می‌شود که علاوه بر زمان بر بودن هزینه‌ی بسیار زیادی نیز دارد.

روش‌های متنوعی برای قطعه‌بندی این تصاویر به صورت خودکار و نیمه-خودکار ارایه شده‌است که در میان آن‌ها روش‌های مبتنی بر یادگیری ماشین و یادگیری عمیق دقت بسیار بالاتری نسبت به سایر روش‌ها نشان داده‌اند. این روش‌ها علی‌رغم این برتری مشکلاتی از قبیل هزینه‌ها محاسباتی بالا، عدم توانایی در یادگیری شکل و ساختار در قطعه‌بندی بافت مورد نظر و عدم توانایی در قطعه‌بندی تومور به علت متغیر بودن اندازه، موقعیت و شکل آن را دارا می‌باشند.

در این مطالعه، برای حل مشکلات اشاره‌شده روش‌هایی ارائه گردیده است که با انتقال دانش از یک مدل پیچیده به یک مدل ساده‌تر و نیز ارائه‌ی یک چارچوب بر اساس بازخورد خطأ، بدون افزایش پیچیدگی و هزینه‌های محاسباتی در یک مدل ساده، دقت آن افزایش یابد. همچنین برای حل مشکل یادگیری شکل و ساختار برای قطعه‌بندی ساختارهای در ریسک، یک تابع هزینه بر اساس شکل پیشنهاد گردیده است که بر اساس فضای شکل تعریف شده، معتبر یا نامعتبر بودن شکل پیش‌بینی شده برای یک ساختار سه‌بعدی را ارزیابی می‌کند. در نهایت برای قطعه‌بندی تومور یک چارچوب بر اساس سازوکار توجه پیشنهاد شده است که از دو مازول توجه و قطعه‌بند تشکیل گردیده است. مازول توجه وظیفه‌ی یافتن موقعیت تومور و مازول قطعه‌بند با توجه به اطلاعات مازول توجه، قطعه‌بندی نهایی را انجام می‌دهد. ارزیابی‌های نهایی بر روی مجموعه دادگان مختلف نشان داد، روش‌های پیشنهادی دقت قطعه‌بندی را افزایش خواهند داد. به عنوان مثال در استفاده از تابع هزینه‌ی بر اساس شکل، معیار Dice شبکه برای دادگان آزمایش در قطعه‌بندی هیپوکامپ از $0/81$ برای شبکه‌ی آموزش داده شده با تابع هزینه‌ی BCE به $0/86$ برای شبکه‌ی آموزش داده شده با تابع هزینه‌ی پیشنهادی بر اساس شکل، رسید و با به‌کارگیری مازول توجه در قطعه‌بندی تومور، این معیار از $0/68$ برای شبکه‌ی بدون مازول توجه به $0/79$ برای شبکه‌ی با استفاده از مازول توجه، افزایش یافت.

کلیدواژه‌ها: قطعه‌بندی ساختارهای در ریسک، قطعه‌بندی تومور، شبکه‌های عمیق کانولوشنی

فهرست مطالب

۱۷	۱ مقدمه
۱۷	۱-۱ تعریف مسئله
۱۹	۲-۱ اهمیت موضوع
۱۹	۳-۱ ادبیات موضوع
۲۱	۴-۱ اهداف تحقیق
۲۱	۵-۱ ساختار پایاننامه
۲۲	۲ مفاهیم اولیه
۲۲	۲-۱ تصویربرداری پزشکی
۲۳	۲-۱-۱ تصویربرداری سی‌تی اسکن
۲۵	۲-۱-۲ تصویربرداری ام‌آرآی
۲۷	۲-۲ هوش مصنوعی
۲۷	۲-۲-۱ هوش مصنوعی
۲۹	۲-۲-۲ یادگیری ماشین
۳۱	۲-۲-۳ یادگیری عمیق
۳۵	۲-۳ شبکه‌های عصبی کانولوشنی عمیق
۳۶	۲-۳-۱ لایه‌های کانولوشنی

۳۸	۲-۳-۲ توابع فعالیت
۳۸	۳-۳-۲ لایه‌های ادغام
۳۹	۴-۳-۲ لایه‌های تمام متصل
۴۰	۴-۲ روش‌های تعمیم‌پذیری
۴۳	۵-۲ پردازش تصویر و بینایی کامپیوتر
۴۴	۶-۲ جمع‌بندی
۴۵	۳ مرور ادبیات
۴۵	۱-۳ مقدمه
۴۵	۲-۳ تعاریف و مفاهیم اولیه
۴۵	۱-۲-۳ تعریف قطعه‌بندی
۴۶	۲-۲-۳ اهداف قطعه‌بندی تصاویر پزشکی
۴۷	۳-۲-۳ چالش‌های قطعه‌بندی تصاویر پزشکی
۴۸	۲-۳ انواع قطعه‌بندی تصاویر پزشکی
۵۰	۲-۳-۵ ارزیابی قطعه‌بندی
۵۱	۲-۳-۶ معیارهای ارزیابی قطعه‌بندی
۵۴	۳-۳ روش‌های قطعه‌بندی
۵۴	۱-۳-۳ روش‌های آستانه‌گذاری
۵۵	۲-۳-۳ روش رشد ناحیه‌ای
۵۵	۳-۳-۳ روش‌های طبقه‌بندی
۵۷	۴-۳-۳ روش‌های خوش‌بندی
۵۸	۵-۳-۳ مدل‌های شکل‌پذیر
۶۰	۶-۳-۳ مدل‌های مبتنی بر اطلس

۶۰	۷-۳-۳ سایر روش‌های قطعه‌بندی	۳
۶۰	۴-۳ قطعه‌بندی با استفاده از شبکه‌های عصبی عمیق	۳
۶۱	۱-۴-۳ طبقه‌بندی پیکسل‌ها برای قطعه‌بندی	
۶۴	۲-۴-۳ شبکه‌های تمام کانولوشنی برای قطعه‌بندی	
۶۸	۳-۴-۳ شبکه‌های تمام کانولوشنی سه‌بعدی برای قطعه‌بندی	
۷۱	۵-۳ قطعه‌بندی تومور و ساختارهای در ریسک با یادگیری عمیق	
۷۴	۶-۳ جمع‌بندی	
۷۶	۴ روش‌های پیشنهادی و نتایج	
۷۶	۱-۴ مقدمه	
۷۶	۲-۴ روش‌های پیشنهادی قطعه‌بندی ساختارهای در ریسک	
۷۷	۱-۲-۴ معرفی مجموعه دادگان و پیش‌پردازش	
۸۱	۲-۲-۴ آموزش قطعه‌بندی با استفاده از متدهای چگالش دانش	
۹۵	۲-۳-۴ بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک	
۱۰۰	۲-۴-۴ معرفیتابع هزینه‌ی براساس شکل برای قطعه‌بندی ساختارهای در ریسک	
۱۱۵	۳-۴ روش پیشنهادی قطعه‌بندی تومور	
۱۱۶	۱-۳-۴ معرفی دادگان و پیش‌پردازش	
۱۱۷	۲-۳-۴ معرفی شبکه‌ی براساس ساز و کارتوجه برای قطعه‌بندی تومور	
۱۲۴	۴-۴ جمع‌بندی	
۱۲۵	۵ نتیجه‌گیری و کارهای آینده	
۱۲۹	آ مقالات ارایه شده در این مطالعه	

فهرست شکل‌ها

۱-۱ سه لایه‌ی مختلف از تصویر سی‌تی اسکن مغزی در نمای اکسیال. ردیف اول: ساخترهای در ریسک مغزی، ردیف دوم: تومور مغزی [۱]	۱۸
۱-۲ دستگاه سی‌تی اسکن [۲]	۲۴
۲-۱ نمونه‌ی تصویر سی‌تی اسکن از ناحیه‌ی شکم در سه نمای مختلف [۳]	۲۵
۲-۲ تصویر یک دستگاه ام‌آرآی (سمت چپ) و یک لایه از تصویر سه‌بعدی ام‌آرآی زانو (سمت راست) [۴]	۲۶
۲-۳ ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین، یادگیری عمیق و بینایی کامپیوتر .	۲۸
۲-۴ مقایسه‌ی الگوریتم‌های طبقه‌بندی و رگرسیون [۵]	۳۰
۲-۵ مقایسه‌ی یک نورون طبیعی و مدل‌سازی ساده‌ی آن [۶]	۳۲
۲-۶ مقایسه‌ی شبکه‌ی پرسپترون عمیق و کم عمق [۷]	۳۳
۲-۷ شبکه‌های مولد خصمانه [۸]	۳۴
۲-۸ ماتریس ویژگی‌های سه بعدی استخراج شده توسط یک شبکه‌ی کانولوشنی [۹]	۳۵
۲-۹ معماری یک شبکه‌ی کانولوشنی برای طبقه‌بندی اعداد دست‌نویس [۱۰]	۳۶
۲-۱۰ چگونگی حرکت فیلتر در یک لایه‌ی کانولوشنی [۱۰]	۳۷
۲-۱۱ روند محاسبات در یک لایه‌ی کانولوشنی [۱۰]	۳۷
۲-۱۲ چهار نمونه از پر استفاده‌ترین توابع فعالیت [۱۱]	۳۸

۳۹	۱۴-۲ ادغام میانگین‌گیر و ادغام بیشینه‌گیر [۱۰]
۴۲	۱۵-۲ توقف زود هنگام با توجه به خطای داده‌های ارزیابی [۱۲]
۴۲	۱۶-۲ چگونگی اعمال تکنیک dropout [۱۳]
۴۸	۱-۳ ناهمگنی شدت و میدان در یک تصویر ام‌آرای [۱۴]
۵۲	۲-۳ شهود عملکرد معیار Dice [۱۵]
۵۳	۳-۳ شهود عملکرد معیار Jaccard [۱۶]
۵۴	۴-۳ شهود عملکرد فاصله‌ی Hausdorff [۱۷]
۵۷	۵-۳ قطعه‌بندی تصویر ام‌آرای مغزی با روش K-means [۱۸]
۵۸	۶-۳ قطعه‌بندی تصویر ام‌آرای مغزی با روش Fuzzy C-means [۱۹]
۵۸	۷-۳ قطعه‌بندی با استفاده از مدل‌های شکل‌پذیر [۲۰]
۵۹	۸-۳ روش تکامل منحی و استخراج مجموعه-تراز صفر برای قطعه‌بندی [۲۱]
۶۰	۹-۳ روش مبتنی بر اطلس بری قطعه‌بندی تصویر مغزی [۲۲]
۶۲	۱۰-۳ العزاندن یک پنجره روی تصویر ام‌آرای مغزی و استخراج پچ [۲۳]
۶۲	۱۱-۳ شبکه‌ی کانولوشنی برای طبقه‌بندی پیکسل‌ها و سطی پچ‌ها [۲۳]
۶۳	۱۲-۳ معماری شبکه‌ی کانولوشنی برای طبقه‌بندی بر اساس پچ ساختارهای در ریسک سر و گردن [۲۴]
۶۳	۱۳-۳ نتایج بدست آمده در قطعه‌بندی ساختارهای در ریسک سر و گردن (قرمز: خروجی شبکه، سبز: قطعه‌بندی مطلوب) [۲۴]
۶۴	۱۴-۳ قطعه‌بندی تصاویر با استفاده از رویکرد شبکه‌های تمام کانولوشنی [۲۵]
۶۵	۱۵-۳ استخراج نقشه‌ی گرمایی وجود یک شی در تصویر با استفاده از ویژگی‌های استخراج شده توسط شبکه‌ی طبقه‌بند [۲۵]
۶۵	۱۶-۳ تغییر ابعاد ویژگی‌های استخراج شده بعد از لایه‌های ادغام به ابعاد تصویر اصلی [۲۵]
۶۶	۱۷-۳ معماری شبکه‌ی شامل کانولوشن و دیکانولوشن برای بهبود دقت قطعه‌بندی [۲۶]

۶۶	۳-۱۸ نمایش تفاوت عملکرد کانولوشن و دیکانولوشن [۲۶]
۶۷	۳-۱۹ معماری شبکه‌ی UNet [۲۷]
۶۸	۳-۲۰ معماری شبکه‌ی SegNet [۲۸]
۶۹	۳-۲۱ معماری شبکه‌ی UNet سه‌بعدی [۲۹]
۶۹	۳-۲۲ یک بلوک با اتصال residual [۳۰]
۷۰	۳-۲۳ کانولوشن‌های منبسط شده [۳۱]
۷۰	۳-۲۴ معماری شبکه‌ی سه‌بعدی شامل اتصالات residual و کانولوشن انساطی [۳۰]
۷۱	۳-۲۵ معماری شبکه‌ی VNet [۳۲]
۷۲	۳-۲۶ معماری روش پیشنهادی در [۳۳]
۷۳	۳-۲۷ معماری UNet پیشنهادی در [۳۴]
۷۳	۳-۲۸ نتایج قطعه‌بندی تومور مغزی در تصاویر ام‌آرآی در مرجع [۳۵]
۷۴	۳-۲۹ روش پیشنهادی در [۳۵] برای استفاده از پچ‌های سه‌بعدی با اندازه‌ی مختلف
۷۸	۴-۱ نمودار بلوکی گام‌های پیش‌پردازش بر روی دادگان
۷۹	۴-۲ نمونه‌ی یک تصویرسی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش، در سه نمای اکسیال، کرونال و سجیتال
۷۹	۴-۳ نمونه‌ی یک تصویرسی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش همراه با برچسب ساختار در ریسک، در چهار نمای اکسیال، کرونال، سجیتال و سه‌بعدی
۸۰	۴-۴ نمونه‌ی یک تصویر در نمای اکسیال با برچسب ساختار در ریسک متناظر (الف) تصویر اصلی (ب) تصویر تبدیل یافته برای افزایش دادگان
۸۱	۴-۵ نمونه‌ی یک تصویر ام‌آرآی هیپوکامپ در نماهای مختلف با قطعه‌بندی (الف) تصویر اصلی (ب) تصویر پس از پیش‌پردازش
۸۳	۴-۶ خروجی‌های سخت و نرم برای طبقه‌بندی چهار کلاس [۳۶]
۸۵	۴-۷ عملکرد انتقال دانش توسط چارچوب چگالش دانش [۳۷]

۸-۴ معماری شبکه‌ی UNet پیچیده، استفاده شده به عنوان آموزگار	۸۶
۹-۴ معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانش‌آموز اول	۸۷
۱۰-۴ معماری شبکه‌ی پیشنهادی بر اساس بلوک‌های Residual، استفاده شده به عنوان دانش‌آموز دوم	۸۷
۱۱-۴ آموزش شبکه‌های دانش‌آموز بر مبنای چارچوب چگالش دانش از آموزگار	۸۹
۱۲-۴ تاثیر تغییر آستانه بر روی معیار Dice برای هر ساختار در ریسک	۹۰
۱۳-۴ خروجی مدل‌های مختلف به صورت آموزش داده شده با چارچوب چگالش دانش و تکی، (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش‌آموز Resnet (ت) پیش‌بینی دانش‌آموز UNet، (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی UNet ساده‌ی تنها	۹۲
۱۴-۴ نمودار جعبه‌ای معیار Dice دادگان آزمایش برای چهار ساختار در ریسک و پنج شبکه‌ی آموزش داده شده	۹۳
۱۵-۴ نقشه‌های احتمالاتی خروجی شبکه‌ها قبل از آستانه‌گذاری (الف) آموزگار (ب) دانش‌آموز (پ) دانش‌آموز UNet (ت) Resnet تنها (ث) UNet تنها	۹۴
۱۶-۴ عملکرد چارچوب بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک . .	۹۶
۱۷-۴ معماری شبکه‌ی UNet استفاده شده برای آموزش با چارچوب بازخورد خطای پیش‌بین	۹۸
۱۸-۴ پیش‌بینی قطعه‌بندی ساختارهای در ریسک در نماهای مختلف (الف) خروجی مطلوب (ب) خروجی UNet تنها (پ) خروجی UNet آموزش دیده شده با چارچوب بازخورد خطای پیش‌بین	۹۹
۱۹-۴ استخراج بردارهای ویژه از تصاویر چهره و مقایسه با تصاویر غیر چهره (الف) صد نمونه از تصاویر چهره‌ی استفاده شده، (ب) صد چهره‌ی ویژه (بردارهای ویژه از دادگان چهره) و (پ) صد نمونه تصویر غیر چهره (دادگان FashionMNIST)	۱۰۳
۲۰-۴ بازسازی یک تصویر چهره با استفاده از چهره‌های ویژه، سمت چپ: تصویر اصلی، وسط: تصویر بازسازی شده با تمام چهره‌های ویژه، سمت راست: تصویر بازسازی شده با ۲۰ چهره‌ی ویژه‌ی اول	۱۰۴

- ۴-۲۱ بازسازی یک تصویر غیرچهره با استفاده از چهره‌های ویژه، سمت چپ: تصویر اصلی، سمت راست: تصویر بازسازی شده با تمام چهره‌های ویژه ۱۰۴
- ۴-۲۲ تمایز بین تصاویر چهره و غیرچهره بر اساس وزن‌های بازسازی (الف) هیستوگرام وزن‌های بازسازی تصاویر چهره (آبی) و غیرچهره (آجری)، (ب) کورتوسیس وزن‌های بازسازی چهره و غیرچهره برای مجموعه داده‌ی آموزش واستخراج آستانه‌ی طبقه‌بندی با روش اوتسو (پ) طبقه‌بندی کورتوسیس وزن‌های بازسازی برای مجموعه دادگان آزمایش ۱۰۵
- ۴-۲۳ ارگان‌های ویژه‌ی استخراج شده برای (الف) هیپوکامپ و (ب) قلب در سه نمای اکسیال، کرونال و سجیتال ۱۰۷
- ۴-۲۴ (الف) هیستوگرام وزن‌های بازسازی هیپوکامپ و غیر هیپوکامپ، (ب) هیستوگرام کورتوسیس وزن‌های بازسازی هیپوکامپ و غیر هیپوکامپ (تابع هزینه با خط مشکی نشان داده شده است)، (پ) هیستوگرام وزن‌های بازسازی قلب و غیر قلب، (ت) هیستوگرام کورتوسیس وزن‌های بازسازی قلب و غیر قلب (تابع هزینه با خط مشکی نشان داده شده است). در تصاویر واضح است که مقدار کورتوسیس برای ساختارهای صحیح بالاتر از ساختارهای با شکل ناقص است بنابراین تابع هزینه‌ی تعریف شده، برای جریمه‌ی شبکه به ایجاد ممان چهارم بیشتر برای وزن‌های بازسازی است. ۱۰۸
- ۴-۲۵ توصیف کلی عملکرد تابع هزینه‌ی پیشنهادی ۱۰۹
- ۴-۲۶ معماری شبکه‌ی کانولوشنی سه‌بعدی استفاده شده ۱۱۰
- ۴-۲۷ قطعه‌بندی تصاویر ام‌آرآی هیپوکامپ، (الف) قطعه‌بندی مطلوب (ب) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه BCE و (پ) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه SBCF در نماهای مختلف ۱۱۲
- ۴-۲۸ قطعه‌بندی تصاویر سی‌تی اسکن قلب، (الف) قطعه‌بندی مطلوب (ب) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه BCE و (پ) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه SBCF در نماهای مختلف ۱۱۲
- ۴-۲۹ نمودار جعبه‌ای معیارهای ارزیابی Dice ، Hausdorff و Jaccard برای (الف) قطعه‌بندی هیپوکامپ (ب) قطعه‌بندی قلب ۱۱۳

- ۴-۳۰ داده‌ی پرت در قطعه‌بندی هیپوکامپ با استفاده از تابع هزینه‌ی بر اساس شکل، (الف) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی BCE (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز)، (ب) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی SBCF (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز)
۱۱۴
- ۴-۳۱ هیستوگرام کورتوسیس پیش‌بینی شده توسط تابع هزینه‌ی BCE (رنگ سبز) و تابع هزینه‌ی SBCF (رنگ آجری) برای داده‌های آزمایش در کنار کورتوسیس‌های مطلوب (رنگ سبز). (الف) دادگان هیپوکامپ (ب) دادگان قلب
۱۱۴
- ۴-۳۲ چهار دنباله تصویر ام‌آرآی T₁ ، T₂ ، T_{1ce} و FLAIR برای یک بیمار خاص با قطعه‌بندی متناظر برای تومور (قرمز: هسته‌ی تومور، زرد: ناحیه‌ی پیش توموری و سبز: تومور در حال رشد)
۱۱۷
- ۴-۳۳ کلیت روش پیشنهادی برای قطعه‌بندی تومور بر اساس ساز و کار ماژول توجه و قطعه‌بند
۱۱۸
- ۴-۳۴ چارچوب پیشنهادی ماژول‌های توجه (قسمت نارنجی) و قطعه‌بند (قسمت سبز) برای قطعه‌بندی تومور به صورت دو بعدی. از معماری UNet به عنوان شبکه‌ی پایه استفاده گردید.
۱۱۹
- ۴-۳۵ قطعه‌بندی تومور در نماهای مختلف برای یک مورد از دادگان آزمایش، (الف) قطعه‌بندی مطلوب، (ب) خروجی ماژول توجه، (پ) خروجی ماژول قطعه‌بند بر اساس ساز و کار توجه و (ت) خروجی UNet بدون استفاده از ماژول توجه
۱۲۱
- ۴-۳۶ نمودار جعبه‌ای معیارهای ارزیابی Dice ، Jaccard و Hausdorff برای مجموعه دادگان آزمایش در دو شبکه Unet بر اساس ماژول توجه (Unet WA) و تنها Unet
۱۲۳

فهرست جدول‌ها

۱-۴ معیار Dice برای مدل‌های مختلف	۹۱
۲-۴ معیار Jaccard برای مدل‌های مختلف	۹۱
۳-۴ مقابسه‌ی میزان پیچیدگی و هزینه‌ی محاسباتی هر مدل	۹۵
۴-۴ معیارهای ارزیابی قطعه‌بندی برای شبکه‌ی آموزش دیده شده به دو صورت تکی و چارچوب بازخورد خطای پیش‌بین	۹۹
۵-۴ معیارهای ارزیابی برای مجموعه دادگان آزمایش برای قطعه‌بندی هیپوکامپ و قلب بر اساس رویکرد تابع هزینه‌ی بر اساس شکل و تابع هزینه‌ی BCE	۱۱۱
۶-۴ معیارهای ارزیابی برای مجموعه دادگان آزمایش برای قطعه‌بندی تومور در دو شبکه‌ی Unet بر اساس ماثول توجه و Unet تنها	۱۲۲

فصل ۱

مقدمه

یکی از موثرترین روش‌های معرفی شده در دهه‌های گذشته برای درمان سرطان، رادیوتراپی^۱ بوده است. در رادیوتراپی از پرتوهای پرانرژی مانند اشعه ایکس^۲ و گاما^۳، با دوز^۴ بالا برای از بین بردن سلول‌های سرطانی استفاده می‌شود. این تشعشعات پرانرژی با برخورد به بافت هدف، در برخی موارد، باعث تجزیه و نابودی ناگهانی سلول‌های سرطانی می‌شود. اما در اکثر موارد با آسیب رساندن به دیانای^۵ بافت سرطانی، مانع از تکثیر و درنتیجه نابودی آن می‌شود. بنابراین لازم است این پرتوها تنها به بافت هدف (تومور) برخورد کنند تا از آسیب به بافت‌های سالم جلوگیری شود [۳۸].

۱-۱ تعریف مسئله

همانطور که اشاره گردید در رادیوتراپی، پرتوهای پرانرژی بین بافت سالم و بافت سرطانی تفکیکی قابل نمی‌شوند و در صورت تابانده شدن پرتو به بافت‌های سالم، این بافت‌ها نیز دچار آسیب دیانای و در نهایت از بین می‌روند. بنابراین لازم است قبل از رادیوتراپی موقعیت بافت‌های سالم و بافت سرطانی با دقیق‌ترین تفکیک شود تا از آسیب به سلول‌های سالم و حیاتی جلوگیری شود.

به ساختارهای سالم اطراف تومور که باید از پرتوهای پرانرژی در هنگام رادیوتراپی محافظت شوند،

¹Radiotherapy

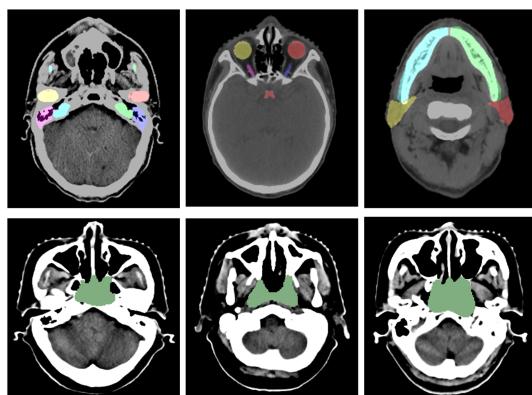
²X-ray

³Gamma

⁴Dose

⁵Deoxyribonucleic Acid (DNA)

ساختارهای در ریسک^۶ گفته می‌شود. برای شهود بهتر در تفاوت بین ساختارهای در ریسک و تومور شکل ۱-۱ را در نظر بگیرید. در ردیف اول این شکل نمونه‌هایی از ساختارهای در ریسک مغزی موجود در یک تصویر سی‌تی اسکن^۷ در نمای اکسیال^۸ مانند چشم‌ها، عصب‌های بینایی، ساقه‌ی مغزی، گوش‌های میانی و ... نشان داده شده است و در ردیف پایین نیز، تومور متناظر به نمایش گذاشته شده است.



شکل ۱-۱: سه لایه‌ی مختلف از تصویر سی‌تی اسکن مغزی در نمای اکسیال. ردیف اول: ساختارهای در ریسک مغزی، ردیف دوم: تومور مغزی [۱].

در رادیوتراپی معمولاً سه اصل اساسی در نظر گرفته می‌شود:

- افزایش دوز پرتوهای تابشی به تومور امکان بالا رفتن کنترل بر روی رشد تومور را میسر می‌کند.
- با بالا رفتن کنترل بر روی تومور درجه‌ی بهبود افزایش می‌یابد زیرا از گسترش بیشتر آن جلوگیری می‌شود.
- کم کردن میزان تابش پرتو به بافت‌های سالم اطراف تومور، عوارض جانبی را کاهش می‌دهد.

بنابراین برای بیشینه کردن دوز دریافتی تومور و در عین حال کمینه کردن دوز دریافتی بافت‌های سالم، قطعه‌بندی^۹ دقیق تومور و ساختارهای در ریسک در تصاویر پزشکی یک گام اساسی و حیاتی است [۳۹]. بنابراین در این پژوهش سعی شده است این قطعه‌بندی که اولین و اساسی‌ترین مرحله در رادیوتراپی است با پیشنهاد روش‌هایی بر اساس یادگیری ماشین^{۱۰} و یادگیری عمیق^{۱۱} با دقت بالایی انجام شود.

⁶Organs At Risks (OAR)

⁷ Computational Tomography (CT) Scan

⁸Axials

⁹Segmentation

¹⁰Machine Learning

¹¹Deep Learning

۱-۲ اهمیت موضوع

در رادیوتراپی تهیه‌ی یک نقشه‌ی درمان^{۱۲} که موقعیت تومور و ساختارهای در ریسک را مشخص می‌کند ضروری است که در آن میزان تابش پرتوهای پرانرژی به بافت‌های سالم و تومور تعیین می‌شود. ضرورت وجود این نقشه‌ی درمان به این دلیل است تا تابش حداکثری به بافت تومور و کمینه کردن آسیب به بافت‌های سالم صورت گیرد و در نهایت با تعیین توزیع پرتو، بین بافت سالم و بافت سرطانی می‌توان اطمینان یافت که سلول‌های سرطانی از بین رفته‌اند و میزان آسیب به ساختارهای در ریسک به چه اندازه بوده است [۴۰].

برای تهیه‌ی یک نقشه‌ی درمان، اولین مرحله، قطعه‌بندی تومور و ساختارهای در ریسک است. به طور معمول این قطعه‌بندی به صورت دستی و توسط یک پزشک متخصص صورت می‌گیرد که به عنوان مثال برای قطعه‌بندی تصویرسی‌تی اسکن سر و گردن یک بیمار و مشخص کردن ساختارهای در ریسک، به صورت تقریبی برای یک متخصص آموزش دیده، ۳ ساعت زمان خواهد برد [۴۱] که این عملیات در روند درمان، به علت تغییر شکل تومور باید چندین مرتبه تکرار گردد [۴۲]. مشکلات دیگری علاوه بر زمان بر بودن در قطعه‌بندی به صورت دستی وجود دارد، از جمله: امکان بروز خطای انسانی به علت مشغله‌ی کاری یا خستگی، خسته‌کننده بودن روند قطعه‌بندی و سلیقه‌ی شخص متخصص در طبقه‌بندی بافت‌ها در بین بیماران مختلف، می‌توان اشاره کرد. بنابراین ارائه‌ی یک الگوریتم خودکار^{۱۳} یا نیمه‌خودکار^{۱۴} برای تسريع روند قطعه‌بندی و نیز حل مشکلات ذکر شده ضروری است.

۱-۳ ادبیات موضوع

هدف از قطعه‌بندی در تصاویر، طبقه‌بندی پیکسل‌ها^{۱۵} در تصاویر دو بعدی^{۱۶} یا واکسل‌ها^{۱۷} در تصاویر سه بعدی^{۱۸} به یک طبقه‌ی خاص است که هر طبقه یک ساختار و یا ارگان خاصی را تعریف می‌کند.

¹²Treatment Plan

¹³Automated

¹⁴Semi-Automated

¹⁵Pixels

¹⁶2D images

¹⁷Voxels

¹⁸3D images

با وجود پیشرفت‌های چشمگیر در حوزه‌های پردازش تصویر^{۱۹} و بینایی کامپیوتر^{۲۰}، قطعه‌بندی خودکار و نیمه-خودکار، هنوز یک عملیات پرچالش در زمینه‌های تحلیل تصاویر پزشکی^{۲۱}، پردازش تصاویر طبیعی^{۲۲} و موتورهای جستجوگر^{۲۳} تصویر است [۴۵، ۴۴، ۴۲].

برای قطعه‌بندی تصاویر پزشکی الگوریتم‌های خودکار و نیمه-خودکار فراوانی در دهه‌های گذشته برای تسريع در روند تشخیص و درمان ارائه گردید. از جمله‌ی این روش‌ها می‌توان به موارد زیر اشاره کرد:

- الگوریتم‌های مبتنی بر شدت پیکسل‌ها (واکسل‌ها): در این نوع الگوریتم‌ها معمولاً چندین رویکرد در نظر گرفته می‌شود که شامل آستانه‌گذاری‌های سراسری و یا محلی^{۲۴} بر روی شدت پیکسل‌ها و الگوریتم مبتنی بر رشد ناحیه‌ای^{۲۵} در یک منطقه از تصویر با استفاده از یک بذر^{۲۶} اولیه که کاربر می‌دهد.

- الگوریتم‌های خوشبندی:^{۲۷} این نوع الگوریتم‌ها بر پایه‌ی روش‌های غیرنظرارتی^{۲۸} هستند و سعی در خوشبندی ساختارهای موجود در تصویر با توجه به شدت پیکسل‌ها دارند. از جمله‌ی این روش‌ها می‌توان به روش‌های Fuzzy Clustering C-means و K-means اشاره کرد.

- الگوریتم‌های مبتنی بر مدل‌های شکل‌پذیر:^{۲۹} در این روش قطعه‌بندی، یک کانتور^{۳۰} اولیه توسط کاربر تعیین می‌شود و با کمینه یک تابع هزینه، کانتور حرکت می‌کند و بر روی مرز ساختار مورد نظر قرار می‌گیرد. از جمله‌ی این روش‌ها می‌توان به الگوریتم Snake و Level-Set اشاره کرد.

علاوه بر روش‌های ذکر شده در بالا، روش‌های دیگری مانند روش‌های قطعه‌بندی مبتنی بر اطلس^{۳۱}، روش‌های مبتنی بر طبقه‌بندی^{۳۲} و ... نیز وجود دارد که در فصل‌های بعد به تفصیل مورد مطالعه و مرور

¹⁹Image Processing

²⁰Computer Vision

²¹Medical Images Analysis

²²Natural Images Processing

²³Search Engines

²⁴Global and/or Local Theresholding

²⁵Region Growing

²⁶Seed

²⁷Clustering

²⁸Unsupervised

²⁹Active Contours

³⁰Contour

³¹Atlas based Segmentation

³²Classification

قرار می‌گرند.

با ظهور و گسترش روش‌های یادگیری عمیق در بحث طبقه‌بندی تصاویر طبیعی و بدست آمدن نتایج بسیار خیره کننده‌ای این حوزه با استفاده از شبکه‌های کانولوشنی عمیق^{۳۳} بسیاری از توجهات برای استفاده از این ابزار در قطعه‌بندی تصاویر با استفاده از روش‌های یادگیری عمیق، جلب شد. با بهکار گرفتن ابزارهای یادگیری ماشین و یادگیری عمیق در بحث قطعه‌بندی نتایج بسیار بهتری نسبت به سایر الگوریتم‌های موجود بدست آمده است [۴۶].

۱-۴ اهداف تحقیق

در این پژوهش قصد داریم با استفاده از ابزارهای یادگیری ماشین و یادگیری عمیق روش‌هایی ارایه دهیم که علاوه بر بالا بردن دقت قطعه‌بندی تصاویر پزشکی، امکان قطعه‌بندی چند ساختار به صورت همزمان (ساختارهای در ریسک) وجود داشته باشد.

همانطور که پیش‌تر اشاره گردید، برای رادیوتراپی لازم است قطعه‌بندی تومور و ساختارهای در ریسک موجود باشد. عملیات قطعه‌بندی تومور به علت تغییر اندازه، موقعیت و شکل آن یک عملیات بسیار چالشی برای الگوریتم‌های تمام خودکار است. در این مطالعه، ما روشی ارایه داده‌ایم تا قطعه‌بندی تومور با دقت بالاتری با استفاده از روش‌های یادگیری عمیق صورت پذیرد و در برابر چالش‌های تغییر اندازه، موقعیت و شکل تومور قرار گیریم.

۱-۵ ساختار پایان‌نامه

این پایان‌نامه شامل پنج فصل است. فصل دوم در برگیرنده تعاریف اولیه‌ی مرتبط با پایان‌نامه است. در فصل سوم مسئله‌ی قطعه‌بندی تصاویر پزشکی و کارهای مرتبطی که در این زمینه انجام شده به تفصیل بیان می‌گردد. در فصل چهارم نتایج جدیدی که در این پایان‌نامه به دست آمده ارائه می‌گردد. در این فصل، روش‌های ارایه شده برای قطعه‌بندی ساختارهای در ریسک و تومور توضیح داده خواهد شد. فصل پنجم به نتیجه‌گیری و پیشنهادهایی برای کارهای آتی خواهد پرداخت.

³³Deep Convolutional Neural Networks

فصل ۲

مفاهیم اولیه

در این فصل به معرفی حوزه‌ها و مفاهیم اولیه مرتبط با این پژوهش پرداخته می‌شود. در ابتدا، انواع تصویربرداری پزشکی معرفی و بررسی خواهد شد و در ادامه به تعریف حوزه‌های مرتبط با پردازش تصاویر پزشکی و بینایی کامپیوتر پرداخته می‌شود.

۱-۲ تصویربرداری پزشکی

زمینه‌ی تصویربرداری پزشکی و سیستم‌های تصویرگر، یک زمینه‌ی بسیار گسترده و به نوبه‌ی خود پیچیده است که با استفاده از انواع پرتوهای ایکس، فراصوت، گاما، امواج الکترومغناطیس و ... با تاباندن به بدن بیمار و بازسازی تصویر از روی پرتوهای دریافتی، صورت می‌گیرد^۱. تصویربرداری پزشکی کمک بسیار زیادی به متخصصین حوزه‌ی سلامت در روند تشخیص و درمان صورت داده است، به طور کلی می‌توان روش‌های تصویربرداری پزشکی را به صورت زیر خلاصه کرد[۴۷]:

- رادیوگرافی^۲: که شامل تصویربردای های بر مبنای پرتوی ایکس است که از مهمترین آن‌ها می‌توان به فلوروسکوپی^۳ و رادیوگرافی‌های پروجکشنال^۴ اشاره کرد.

^۱ البته در تصویربرداری هسته‌ای، بر عکس سایر روش‌های تصویربرداری، منبع تابش پرتو درون بدن بیمار قرار می‌گیرد و با آشکارسازی پرتوهای دریافتی تصویر نهایی تشکیل می‌گردد.

²Radiography

³Fluoroscopy

⁴Projectional radiographs

- تصویربرداری هسته‌ای^۵: که با تزریق ایزوتوب‌های خاص و ذره‌های پرانرژی گسیل شده از مواد رادیواکتیو به بدن بیمار و آشکارسازی پرتوهای دریافتنی، تصویر نهایی ساخته می‌شود. از جمله روش‌های تصویربرداری در این حوزه میتوان به PET^۶ و SPECT^۷ اشاره کرد.
- تصویربرداری فراصوت^۸: در این نوع از تصویربرداری از کریستال‌های پیزوالکتریک^۹ برای تولید صوت با فرکانس بالا استفاده می‌شود این صوت به بافت بدن تابانده می‌شود و بازسازی تصویر از روی صوت بازگشته صورت می‌گیرد.
- تصویربرداری توموگرافی^{۱۰}: در تصویربرداری هدف ساخت یک تصویر سه‌بعدی است برای این منظور از لایه‌های مختلف یک شیء بدون آن که بریده شود تصویربرداری صورت می‌گیرد و این لایه‌ها برروی یکدیگر انباشت می‌شوند و در نهایت تصویر سه‌بعدی نهایی ساخته می‌شود. از جمله روش‌های موجود در این حوزه می‌توان به تصویربرداری سی‌تی اسکن و تصویربرداری بر مبنای تشدید مغناطیسی (ام‌آرآی)^{۱۱} اشاره نمود.
- انواع دیگر تصویربرداری مانند: تصویربرداری فوتواکوستیک^{۱۲}، تصویربرداری حرارتی^{۱۳} و ... نیز وجود دارد که به دلیل کاربردهای کیلینیکی کمتر از شرح آن‌ها خودداری می‌شود.

در عملیات قطعه‌بندی تصاویر پزشکی به دلیل آن که قطعه‌بندی به طور معمول از روی تصاویر توموگرافی سه‌بعدی صورت می‌گیرد، در ادامه به شرح و بسط بیشتر روش‌های تصویربرداری سی‌تی اسکن و ام‌آرآی پرداخته می‌شود.

۱-۱-۲ تصویربرداری سی‌تی اسکن

تصویربرداری سی‌تی اسکن یک تکنیک تصویربرداری پزشکی است که در رادیولوژی برای استخراج اطلاعات از بدن به صورت غیرتهاجمی استفاده می‌شود و روند تشخیص را سرعت می‌بخشد. برخلاف

⁵Nuclear Imaging

⁶Positron emission tomography

⁷single-photon emission computerized tomography

⁸Ultrasound

⁹Piezoelectric

¹⁰Tomography

¹¹Magnetic Resonance Imaging (MRI)

¹²Photoacoustic imaging

¹³Thermography

دستگاه‌های معمول تصویربرداری پرتوی ایکس، که از یک منبع ثابت پرتوی ایکس استفاده می‌کنند، در سی‌تی اسکن از یک منبع متحرک مجهز به موتور استفاده می‌شود که حول گانتری^{۱۴} دستگاه قابلیت چرخش دارد. در طی تصویربرداری، بیمار بر روی یک تخت قرار می‌گیرد و به آهستگی به داخل گانتری وارد می‌شود؛ در همین حین منبع پرتوی ایکس درون گانتری دور بدن بیمار می‌چرخد و باریکه‌ی پرتوهای اشعه‌ی ایکس از بدن بیمار عبور می‌کند. در سی‌تی اسکن از آشکارسازهای دیجیتال پرتوی ایکس استفاده می‌شود که دقیقاً در مقابل منبع پرتوی ایکس قرار گرفته‌اند و با آشکارسازی اشعه‌ی عبوری از بدن بیمار، یک سیگنال به کامپیوتر ارسال می‌شود. شکل ۲-۱ شمای کلی یک دستگاه سی‌تی اسکن را نشان می‌دهد.



شکل ۲-۱: دستگاه سی‌تی اسکن [۲]

با چرخش منبع پرتوی ایکس حول بدن بیمار و دریافت پرتوهای عبوری توسط آشکارساز سیگنال‌های آشکارساز توسط الگوریتم‌های پیچیده‌ی ریاضیاتی به تصویرهای دو بعدی از لایه‌های بدن بیمار تبدیل می‌شوند. ضخامت این لایه‌ها بستگی به نوع دستگاه سی‌تی اسکن دارد اما به طور معمول بین یک تا ده میلی‌متر از بافت بدن برای هر لایه است. وقتی بازسازی یک لایه به اتمام رسید بروی لایه‌های قبلی انباسته می‌شود و در نهایت یک تصویر سه‌بعدی ساخته می‌شود.

هر لایه از تصویر ساخته شده امکان نمایش به صورت مجزا و یا به صورت انباشه شده و سه‌بعدی را دارد که قابلیت نمایش، اسکلت، ساختارها و بافت‌های بدن و همچنین ناهنجاری‌های ایجاد شده در بدن را دارا می‌باشد و امکان تشخیص را برای پزشک مهیا می‌سازد. استفاده از تصاویر سی‌تی در روند تشخیص و درمان فواید زیادی دارد از جمله، توانایی چرخش و جابجایی بین لایه‌های مختلف تصویر که امکان مکانیابی موقعیت دقیق ناهنجاری را فراهم می‌سازد. شکل ۲-۲ یک نمونه تصویر از ناحیه‌ی شکم را نشان می‌دهد که در آن ساختارها و بافت‌های مختلف به راحتی قابل تفکیک است.

^{۱۴} Gantry به محفظه‌ی سیلندری شکل دستگاه سی‌تی اسکن گفته می‌شود که تیوب پرتوی ایکس درون آن قرار می‌گیرد.



شکل ۲-۲: نمونه‌ی تصویر سی‌تی اسکن از ناحیه‌ی شکم در سه نمای مختلف [۳]

از سی‌تی اسکن می‌توان در تشخیص بیماری و آسیب نواحی مختلف بدن استفاده نمود به عنوان مثال از سی‌تی در تشخیص تومور و غدد سرطانی و سایر ناهنجاری‌ها در نواحی شکم، سر و گردن و قفسه‌ی سینه بسیار استفاده می‌شود [۴۸].

۲-۱-۲ تصویربرداری ام‌آر‌آی

ام‌آر‌آی یا روش تصویربرداری با تشدید مغناطیس یکی از روش‌های پیشرفته تصویربرداری پزشکی است. با به‌کارگیری این روش می‌توان از بافت‌های درونی بدن تصویر تهیه نمود و از آن طریق مشکلات و بیماری‌های اعصابی بدن را تشخیص داد. همانطور که می‌دانیم در روش‌های تصویربرداری با اشعه‌ی ایکس مانند رادیوگرافی ساده و یا سی‌تی اسکن بدن تحت تابش مقدار معینی از اشعه یونیزه کننده قرار می‌گیرد که اگر از حد مشخصی بیشتر باشد می‌تواند موجب اشکالاتی در کارکرد سلول‌ها شود. ولی در ام‌آر‌آی از اشعه ایکس استفاده‌ای نمی‌شود و بنابراین نسبت به رادیوگرافی و سی‌تی اسکن بسیار کم ضررتر است.

امواج مورد استفاده در ام‌آر‌آی از جنس امواج رادیویی و مغناطیسی هستند که ضرری برای بدن ندارند.

ام آرآی از این واقعیت فیزیکی استفاده می‌کند که پروتون‌هایی که در هسته‌ی اتم‌ها قرار گرفته‌اند مانند کره‌ی زمین در حول محور خود با سرعت زیادی اسپین دارند و در نتیجه یک میدان معناطیسی در اطراف خود تشکیل می‌دهند. در ام آرآی بیمار در یک میدان معناطیسی بسیار قوی قرار می‌گیرد. این میدان موجب می‌شود محور اسپین پروتون‌های هسته‌ی اتم‌ها در تمام بافت‌های بدن (بخصوص پروتون‌هایی که در هسته مولکول آب قرار دارند) در امتداد خطوط میدان معناطیسی ام آرآی قرار گیرند. سپس امواج رادیویی خاصی به سوی بدن بیمار تابانده می‌شود. این امواج که بصورت پالس فرستاده می‌شوند موجب می‌گردند تا محور اسپین پروتون‌ها کمی تغییر کند. با اتمام پالس رادیویی، محور چرخش پروتون دوباره در امتداد خطوط میدان معناطیسی برمی‌گردد. این برگشت موجب ایجاد یک موج رادیویی (الکترومعناطیسی) جدید می‌شود. سپس این امواج رادیویی ثانویه که از تک تک پروتون‌ها ساطع می‌شوند، توسط گیرنده‌های دستگاه ام آرآی دریافت شده و به کامپیوتر آن ارسال می‌گردد. کامپیوتر ام آرآی بسیار پرقدرت و با توان محاسباتی بالا است. در این کامپیوتر امواج دریافت شده بسرعت تحلیل شده و سپس تصاویری براساس این تحلیل‌ها ساخته می‌شود که پژوهش آن‌ها را بر روی مانیتور دستگاه می‌بیند و در صورت لزوم آنها را چاپ می‌کند.

ام آرآی یک روش تصویربرداری دقیق و پرقدرت برای تشخیص مشکلات و بیماری‌های بافت‌های بدن است. یکی از نقاط تمایز این روش با سی‌تی اسکن در این است که در ام آرآی تصاویر بافت‌های نرم مانند غضروف، تاندون، عصب و رگ‌ها بسیار واضح و دقیق دیده می‌شوند و این روش تصویربرداری بخصوص برای تشخیص بیماری‌های این بافت‌ها مفید است. تصویر حاصل از ام آرآی مانند تصویر سی‌تی اسکن، سه‌بعدی است و امکان نمایش لایه‌های مختلف وجود دارد. شکل ۲-۳ در سمت چپ یک دستگاه ام آرآی آر و در سمت راست یک نمونه تصویر دریافت شده از زانو را نشان می‌دهد که با جزئیات بالایی بافت‌ها را از یکدیگر تفکیک کرده است [۴].



شکل ۲-۳: تصویر یک دستگاه ام آرآی (سمت چپ) و یک لایه از تصویر سه‌بعدی ام آرآی زانو (سمت راست) [۴]

۲-۲ هوش مصنوعی

در ادامه‌ی این قسمت به بررسی و تعریف حوزه‌های مرتبط با هوش مصنوعی^{۱۵} پرداخته خواهد شد.

۲-۲-۱ هوش مصنوعی

هوش مصنوعی به عنوان یک رشته‌ی دانشگاهی در دهه‌ی ۵۰ میلادی پایه‌گذاری شد. در واقع اصطلاح هوش مصنوعی توسط جان مک‌کارتی^{۱۶}، دانشمند علوم کامپیوتر آمریکایی، ابداع شد که طبق تعریف او، هوش مصنوعی عبارت است از علم ساخت ماشین‌های هوشمند به خصوص برنامه‌های کامپیوتری هوشمند.

اگرچه از تعریف هوش مصنوعی چندین دهه می‌گذرد اما رشد و گسترش آن تنها در دو دهه‌ی اخیر به علت در دسترس بودن داده‌های عظیم^{۱۷} و بالا رفتن توان کامپیوتری میسر شده است. هوش مصنوعی بهترین عملکرد خود را با در کنار هم قرار دادن حجم بالای دادگان و الگوریتم‌های هوشمند بدست می‌آورد که باعث یادگیری ویژگی‌ها و الگوهای درون دادگان به صورت خودکار می‌شود. از جمله کاربردهای روز افزون هوش مصنوعی که تقریباً با زندگی همه‌ی ما آمیخته شده است می‌توان به: ماشین‌های خودران، دستیاران صوتی، تشخیص هوشمند چهره برای شناسایی افراد، پروژه‌های وابسته به ژن‌های انسان و ... اشاره نمود.

اگرچه بسیاری از افراد اصطلاحات هوش مصنوعی، یادگیری ماشین و یادگیری عمیق را به جای یکدیگر استفاده می‌کنند اما در واقع این اصطلاحات تعاریف دقیق‌تر و جداگانه‌ای از یکدیگر دارند. همانطور که پیشتر اشاره شد هوش مصنوعی علم ایجاد ماشین‌ها و برنامه‌های کامپیوتری هوشمند است که همانطور که از این تعریف نیز پیداست حوزه‌ی بسیار وسیعی را دربر می‌گیرد که امروزه تقریباً هر شرکتی ادعای استفاده از هوش مصنوعی در محصولات خود را دارد. بنابراین یادگیری ماشین در زیرمجموعه‌ی هوش مصنوعی قرار می‌گیرد و شامل تکنیک‌های و مدل‌های بسیار پیشرفته‌تر که کامپیوترها را قادر به استخراج ویژگی‌های دقیق‌تر از دادگان می‌سازد و به الگوریتم هوش مصنوعی تحويل می‌دهد. به طور خلاصه می‌توان گفت یادگیری ماشین علم قرار دادن کامپیوترها در عمل است بدون آن که به طور صریح برنامه‌ریزی شده باشند.

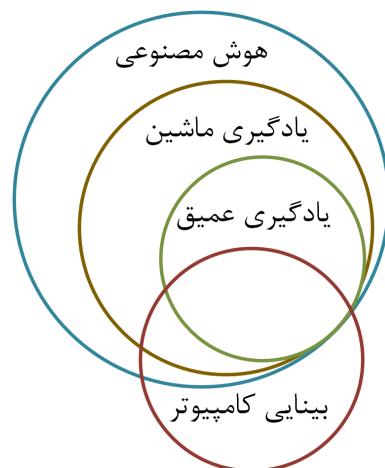
¹⁵ Artificial Intelligence

¹⁶ John McCarthy

¹⁷ Big Data

در نهایت، یادگیری عمیق یک حوزه جدید یادگیری ماشین است که از شبکه‌های عصبی عمیق و چند لایه برای استخراج ویژگی و تصمیم‌گیری استفاده می‌کند که این کار موجب استخراج الگوها و ویژگی‌های مرتبه‌ی بالاتر و تفکیک‌گر بهتر می‌شود و در نتیجه دقت الگوریتم افزایش پیدا خواهد کرد. از جمله کاربردهای یادگیری عمیق می‌توان به طبقه‌بندی اشیا در تصاویر^{۱۸}، شناسایی اشیا در تصاویر^{۱۹}، ترجمه از یک زبان به زبان دیگر^{۲۰}، تولید تصاویر ساختنگی^{۲۱} و ... اشاره نمود [۴۹].

به طور کلی می‌توان ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین و یادگیری عمیق را به صورت شکل ۲-۴ نشان داد. همانطور که ملاحظه می‌شود حوزه‌های یادگیری ماشین و یادگیری عمیق در زیر مجموعه‌ی هوش مصنوعی واقع می‌شود و حوزه‌ی بینایی کامپیوتر علاوه بر داشتن مفاهیم خاص خود در سال‌های اخیر اشتراک نسبتاً زیادی با روش‌های هوش مصنوعی پیدا کرده است (در ادامه به بررسی حوزه‌ی بینایی کامپیوتر پرداخته خواهد شد).



شکل ۲-۴: ارتباط حوزه‌های هوش مصنوعی، یادگیری ماشین، یادگیری عمیق و بینایی کامپیوتر

¹⁸Object Classification

¹⁹Object Recognition

²⁰Language Translation

²¹Fake Image Generation

۲-۲-۲ یادگیری ماشین

یادگیری ماشین شامل سه نوع حوزه‌ی یادگیری است که عبارتند از: یادگیری با سرپرست^{۲۲}، یادگیری بدون سرپرست^{۲۳} و یادگیری تقویتی^{۲۴}، که هدف نهایی این الگوریتم‌ها فراگیری ویژگی‌ها و الگوها از دادگان است بدون آنکه به صورت صریح به این ویژگی و الگوها اشاره شده باشد.

یادگیری با سرپرست

در یادگیری با سرپرست، خروجی مطلوب یا داده‌های Ground Truth در دسترس هستند و در نتیجه با تشکیل یک تابع هزینه بین خروجی مطلوب و خروجی پیش‌بینی شده، با استفاده از روش‌های آموزش مدل‌های یادگیری ماشین الگوریتم آموزش می‌بینند. خروجی مطلوب تعریف شده بسته به وظیفه‌ی مورد نظر می‌تواند گسسته (شماره‌ی مربوط به کلاس‌های مختلف دادگان) و یا یک خروجی پیوسته باشد. بر این اساس یادگیری با سرپرست را می‌توان به دو گروه طبقه‌بندی^{۲۵} و رگرسیون^{۲۶} تقسیم نمود.

- طبقه‌بندی: در الگوریتم‌های طبقه‌بند هدف نسبت دادن یک مشاهده به کلاس‌های تعریف شده است به عنوان مثال طبقه‌بندی تصاویر تومور پوستی به دو گروه خوش‌خیم و بدخیم یک طبقه‌بندی دو کلاسه است و یا طبقه‌بندی ایمیل‌های دریافتی به دو کلاس ایمیل‌های هرز و ایمیل‌های سالم نیز نمونه‌ای از انجام طبقه‌بندی با روش‌های یادگیری ماشین با سرپرست است.
- رگرسیون: در این نوع مسایل به پیش‌بینی یک مقدار پیوسته نیاز است که در نتیجه خروجی مدل باید یک مقدار پیوسته باشد، به عنوان مثال پیش‌بینی قیمت خانه از روی متراژ، سال ساخت، منطقه‌ی ساخت و ... یک مساله‌ی رگرسیون است.

برای مقایسه‌ی بهتر این دو گروه، شکل ۲-۵ تصویر کاملی از تفاوت هدف این دو الگوریتم را به نمایش کشیده است. برای طبقه‌بندی، هدف یافتن مرزی است که کلاس‌های مختلف را از هم جدا می‌کند اما در مساله‌ی رگرسیون هدف پیدا کردن یک منحنی است که بهترین تخمین از دادگان را داشته باشد.

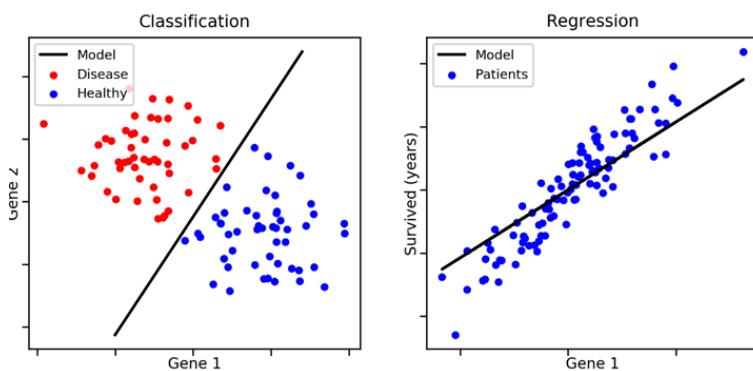
²²Supervised Learning

²³Unsupervised Learning

²⁴Reinforcement Learning

²⁵Classification

²⁶Regression



شکل ۲-۵: مقایسه‌ی الگوریتم‌های طبقه‌بندی و رگرسیون [۵]

یادگیری بدون سرپرست

در این نوع از الگوریتم‌های یادگیری، خروجی مطلوب در دسترس نیست و یک یادگیری بدون بازخورد است. این نوع یادگیری برای استخراج الگوی و پیوندهای خاص نامعلوم بین دادگان در دسترس، استفاده می‌شود. با تعریف ارایه شده می‌توان این نوع یادگیری را به دو کلاس خوش‌بندی^{۲۷} و تشخیص رابطه^{۲۸} تقسیم نمود.

- خوش‌بندی: در این نوع یادگیری بدون سرپرست هدف پیدا کردن خوش‌هایی بین دادگان موجود است به طوریکه داده‌های قرار گرفته در این خوش‌ها دارای ویژگی‌های یکسان باشند. به عنوان مثال با در دست داشتن کل اخبار، خبرهای حوزه‌های مختلف از یکدیگر تفکیک داده شوند.
- تشخیص رابطه: هدف در تشخیص رابطه پیدا کردن قانونی است که اکثر دادگان در دسترس، از آن پیروی می‌کنند. به عنوان مثال الگوریتم‌های توصیه کننده در خریدهای اینترنتی محصولاتی که احتمال خرید بیشتری بر اساس جستجوهای قبلی را دارند، نمایش می‌دهند.

یادگیری تقویتی

در این نوع یادگیری بهجای استفاده از داده‌های علامت‌گذاری شده و مطلوب از الگوریتم‌های مبتنی بر پاداش و مجازات استفاده می‌شود. به عنوان مثال یک کودک را در نظر بگیرید که برای اولین بار قصد

²⁷Clustering

²⁸Association

راه رفتن دارد. این کودک اگر به زمین بیفتد ممکن است آسیب ببیند ولی اگر بتواند راه برود با تشویق والدین مواجه می‌شود بنابراین با سعی در بیشینه کردن پاداش خود که همان تشویق و آغوش والدین است و با سعی و خطأ اقدام به یادگیری می‌کند. با الگو برداری از چنین روندی الگوریتم‌های یادگیری تقویتی توسعه پیدا کرده‌اند که با سعی و خطأ و مشاهدات فراوان، هدف بیشینه کردن پاداش است که در نهایت به یادگیری الگوریتم برای انجام یک عملیات می‌انجامد.

۳-۲-۲ یادگیری عمیق

همانند یادگیری ماشین، یادگیری عمیق نیز یک متادادگیری آماری است که وظیفه‌ی استخراج ویژگی از دادگان خام را دارد. تفاوت اصلی یادگیری عمیق با یادگیری ماشین در استخراج ویژگی توسط شبکه‌ها عصبی چند لایه با تعداد زیاد لایه‌های میانی^{۲۹} است که پشت سر هم قرار گرفته‌اند. همچنین الگوریتم‌های یادگیری عمیق نسبت به یادگیری ماشین پیچیده‌تر و نیازمند قدرت پردازش بالاتری هستند. بنابراین همانطور که پیشتر اشاره گردید، با افزایش تعداد دادگان در دسترس و بالا رفتن قدرت پردازش این حوزه از هوش مصنوعی در دهه‌ی اخیر رشد بسیار چشمگیری داشت.

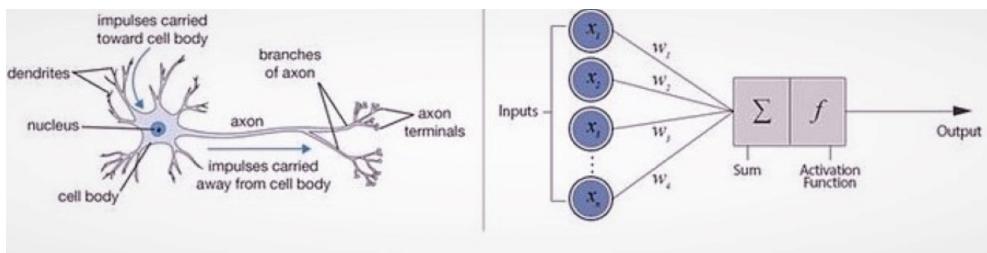
در تعاریف یادگیری عمیق دیدیم که این حوزه زیر مجموعه‌ی یادگیری ماشین قرار می‌گیرد بنابراین تمام متدها و حوزه‌های یادگیری، در یادگیری ماشین، در یادگیری عمیق نیز کاربرد دارد (یادگیری با و بدون سرپرست و یادگیری تقویتی). در ادامه به معرفی چند مدل پر استفاده در حوزه‌ی یادگیری عمیق پرداخته می‌شود.

شبکه‌های عصبی پرسپترون چند لایه

چگونه یک فرد تفاوت میان سگ و گربه را درک می‌کند؟ برای پاسخ به این سوال باید به سراغ بیولوژی شبکه‌ی نورونی داخل مغز برویم. درون مغز انسان به صورت تقریبی حدود ۱۰ میلیارد نورون وجود دارد که هریک از این نورون‌ها به حدود ۱۰ هزار نورون همسایه‌ی خود متصل‌اند. این اتصالات بسیار زیاد و تعداد هنگفت نورون‌ها باعث استخراج ویژگی و یادگیری الگو در داده‌های ورودی حسگرهای بدن می‌شود و بنابراین یک شخص با این ابزار می‌تواند تفاوت میان سگ و گربه را بفهمد و این دو را تفکیک کند.

²⁹HIDDEN LAYERS

هر یک از این نورون‌ها سیگنال‌های الکتروشیمیایی را توسط دندrit‌های^{۳۰} خود از سایر نورون‌ها دریافت می‌کند و در صورتی که ولتاژ غشای نورون از یک آستانه عبور کند سیگنال‌های دریافتی را توسط آکسون^{۳۱} خود به سایر نورون‌ها انتقال می‌دهد. برخلاف تلاش‌های فراون صورت گرفته برای شناخت عملکرد مغز، هنوز مدل دقیقی برای مطالعه‌ی عملکرد این مجموعه نورونی عظیم ارایه نشده است اما مدل‌سازی یک تک نورون توسط هاچکین-هاکسلی [۵۰] با دقت خوبی ارایه شد. در حوزه‌ی شبکه‌های عصبی مصنوعی از یک مدل بسیار ساده‌تر از مدل هاچکین-هاکسلی استفاده می‌شود که به آن پرسپترون^{۳۲} گفته می‌شود. در شکل ۲-۶ یک نورون بیولوژیکی در کنار یک مدل بسیار ساده از نورون نشان داده شده است.



شکل ۲-۶: مقایسه‌ی یک نورون طبیعی و مدل‌سازی ساده‌ی آن [۵۰]

در این مدل پرسپترون چند نکته حائز اهمیت وجود دارد:

- ورودی: در گره‌های ورودی پرسپترون که به نوعی دندrit‌های نورون هستند مقادیر حقیقی به عنوان میزان عملکرد نورون‌های همسایه به ورودی داده می‌شود. این مقادیر به عنوان مثال می‌تواند شدت روشنایی پیکسل‌های یک تصویر باشد.
- وزن‌ها: هر ارتباط که از ورودی به این پرسپترون می‌رسد، دارای یک وزن خاص است که می‌تواند هر مقدار حقیقی‌ای باشد. در واقع در روند یادگیری این وزن‌ها هستند که تغییر می‌کنند و نیازمند محاسبات هستند و در نهایت با بدست آمدن وزن‌های بهینه وظیفه‌ی مورد نظر توسط شبکه عصبی انجام می‌شود.
- جمع وزن‌دار: پس از آنکه ورودی به پرسپترون داده می‌شود یک جمع وزن‌دار با توجه به وزن ارتباطات نورونی صورت می‌گیرد و یک مقدار حقیقی بدست می‌آید.

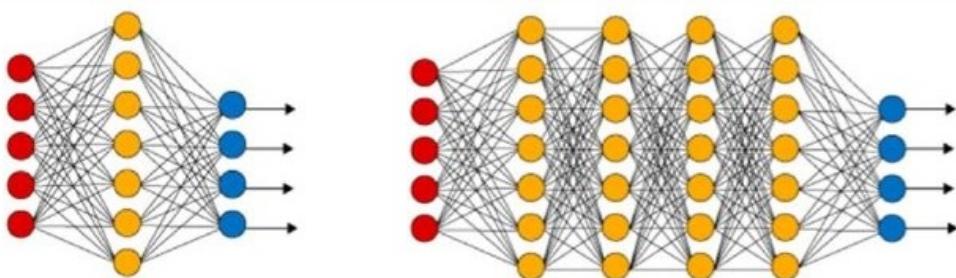
³⁰Dendrites

³¹Axon

³²Perceptron

- تابع فعالیت^{۳۳}: همانند یک نورون طبیعی که باید ولتاژ آن از یک آستانه عبور کند تا به اصطلاح آتش کند، یک تابع فعالیت بر سر راه جمع وزن دار قرار می‌گیرد تا این مدلسازی صورت گیرد.
- خروجی: در نهایت پس از اعمال تابع فعالیت خروجی نهایی پرسپترون ایجاد می‌شود.

با قرار دادن تعدادی از این مدل‌های پرسپترون در کنار یکدیگر می‌توان یک شبکه‌ی عصبی ساخت. اگر تعداد لایه‌های زیاد باشد به آن شبکه عصبی عمیق گفته می‌شود. شکل ۲-۷ دو شبکه‌ی تمام متصل کم عمق (سمت چپ) و عمیق (سمت راست) را نشان می‌دهد.



شکل ۲-۷: مقایسه‌ی شبکه‌ی پرسپترون عمیق و کم عمق [۵]

شبکه‌های عصبی کانولوشنی

این نوع از شبکه‌ها از جمله محبوب‌ترین مدل‌ها یادگیری عمیق برای پردازش تصویر و ویدیو در حوزه‌های پردازش تصویر و بینایی کامپیوتر هستند. به دلیل آنکه در این پروژه از این نوع شبکه‌ها استفاده‌ی بسیار زیادی شده است در بخش بعدی با جزئیات بررسی خواهد شد.

شبکه‌های عصبی بازگشتی

به زبان ساده، شبکه‌های عصبی بازگشتی^{۳۴} برای پیش‌بینی وقایع و استخراج ویژگی از ورودی‌های قبلی استفاده می‌کنند. به عنوان مثال اگر بخواهیم حرف آخر یک کلمه را پیش‌بینی کنیم منطقی است که حروف‌های ابتدایی آن کلمه را در حافظه قرار دهیم و با استفاده از اطلاعات آن‌ها اقدام به پیش‌بینی نهایی کنیم. بنابراین شبکه‌های عصبی بازگشتی امکان به خاطر سپاری ورودی‌های گذشته را فراهم می‌سازد که

³³Activation Function

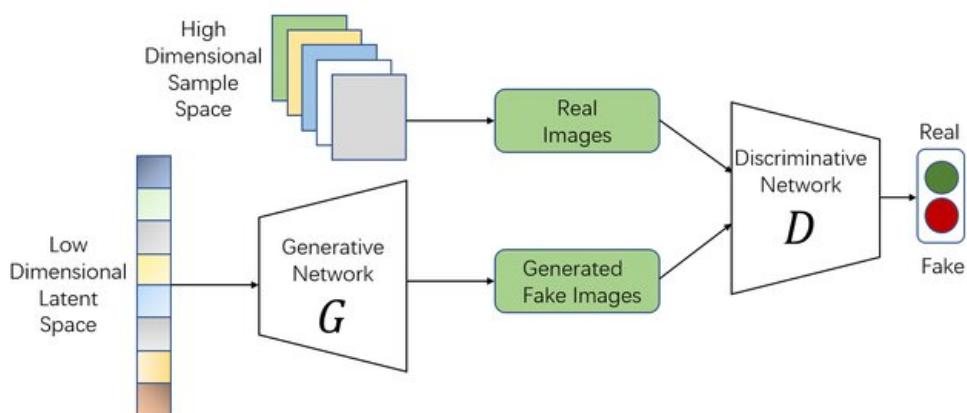
³⁴Recurrent Neural Networks

این ویژگی این نوع شبکه‌ها را از دیگر مدل‌ها متمایز می‌سازد. کاربرد این مدل‌ها برای داده‌های متوالی مانند حوزه‌ی پردازش زبان طبیعی^{۳۵}، پیش‌بینی سیگنال‌های بازار سرمایه، تولید موسیقی و ... می‌باشد.

شبکه‌های مولد خصم‌مانه

شبکه‌های عصبی مولد خصم‌مانه^{۳۶} در سال ۲۰۱۴ توسط گودفلو [۵۱] ارایه شد. این نوع مدل‌ها برخلاف مدل‌های دیگر که فقط از دادگان یادگیری، ویژگی استخراج می‌کردند، امکان خلق کردن بر اساس دادگان ایجاد گردید. این نوع مدل‌ها از دو شبکه‌ی عصبی تشکیل شده‌اند: شبکه‌ی مولد^{۳۷} و شبکه‌ی نقاد^{۳۸}. شبکه‌ی مولد یک نویز در ورودی می‌گیرد و با سعی در تخمین توزیع داده‌های آموزش، یک داده‌ی جعلی ایجاد می‌کند. در دست دیگر شبکه‌ی نقاد سعی در شناسایی داده‌های جعلی از داده‌های حقیقی را دارد. بنابراین یک رقابت بین مولد برای فریب دادن نقاد و نقاد برای تفکیک داده‌های جعلی و حقیقی صورت می‌گیرد.

شکل ۲-۸ چگونگی رقابت بین مولد و نقاد را نشان می‌دهد. همان‌طور که ملاحظه می‌شود، شبکه‌ی نقاد از دو سمت ورودی دریافت می‌کند (داده‌های جعلی و داده‌های حقیقی). در نهایت با تکمیل روند آموزش، شبکه‌ی مولد آنقدر حرفه‌ای می‌شود که با دقت بالایی می‌تواند داده‌هایی تولید کند که نقاد قادر به تشخیص آن از داده‌های حقیقی نباشد. کاربرد این مدل‌ها در خلق دادگان جدید مانند ایجاد تصاویر ساختگی، تولید موسیقی، تولید صدای طبیعی انسان و ... است [۵۲].



شکل ۲-۸: شبکه‌های مولد خصم‌مانه [۸]

³⁵Natural Language Processing

³⁶Generative Adversarial Networks (GAN)

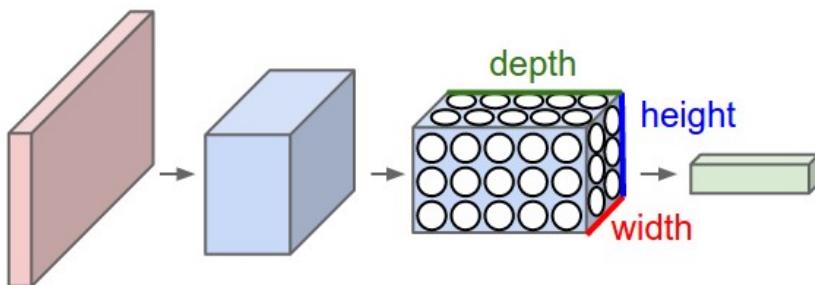
³⁷Generator

³⁸Discriminator

۳-۲ شبکه‌های عصبی کانولوشنی عمیق

شبکه‌های عصبی کانولوشنی بسیار مشابه با شبکه‌های تمام متصل و پرسپترون چند لایه هستند که از نورون‌های شامل وزن و بایاس تشکیل شده‌اند. هر نورون یک ورودی دریافت می‌کند و سپس با اعمال یک ضرب داخلی بین ورودی و وزن‌های خود و در ادامه اضافه کردن یک تابع غیرخطی (تابع فعالیت) خروجی خود را تولید می‌کند. در این حالت، تمام شبکه به عنوان یک تابع مشتق‌پذیر عمل می‌کند و در نهایت با ایجاد خروجی و تشکیل تابع هزینه با استفاده از الگوریتم Back Propagation شبکه وظیفه‌ی مورد نظر را یاد می‌گیرد.

در شبکه‌های تمام متصل، به تعداد المان‌های ورودی نورون در لایه‌ی ورودی قرار می‌گیرد و در ادامه چندین لایه‌ی مخفی که تمامی نورون‌های هر لایه به لایه‌ی قبل و بعد خود متصل‌اند، بر روی این ورودی فعالیت می‌کنند تا خروجی نهایی تولید گردد. بنابراین با افزایش اندازه‌ی ورودی و نیز تعداد لایه‌ها، تعداد وزن‌هایی که باید در آن‌ها یادگیری صورت گیرد بسیار زیاد خواهد شد. با ظهور شبکه‌های کانولوشنی علاوه بر رفع مشکل تغییر تعداد پارامترها با تغییر اندازه‌ی ورودی، پیش‌بینی نهایی نسبت به جابجایی و دیگر انتقالات درون تصویر، مقاوم شد. در شبکه‌های کانولوشنی برخلاف شبکه‌های تمام متصل، نورون‌ها در سه بعد، طول^{۳۹} و عرض^{۴۰} و عمق^{۴۱} چیده شده‌اند و با حرکت بر روی تصویر و استخراج ویژگی از آن خروجی نهایی را ایجاد می‌کنند. شکل ۲-۹ ویژگی‌های استخراج شده از یک تصویر ورودی (حجم قرمز رنگ در تصویر) را نشان می‌دهد که چون نورون‌ها در سه بعد چیده شده‌اند ویژگی‌های استخراج شده نیز سه‌بعدی هستند. با مقایسه‌ی این شکل و شکل ۲-۷ تفاوت دو شبکه‌ی کانولوشنی و تمام متصل در ابعاد ویژگی‌های استخراج شده مشخص است.



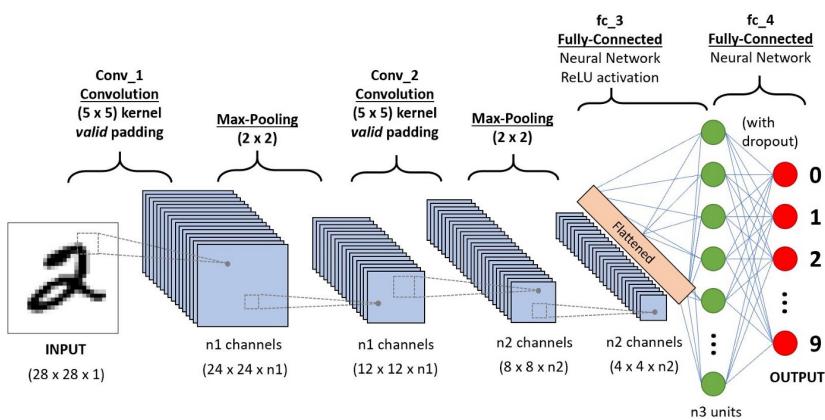
شکل ۲-۹: ماتریس ویژگی‌های سه بعدی استخراج شده توسط یک شبکه‌ی کانولوشنی [۹]

³⁹Height

⁴⁰Width

⁴¹Depth

اولین کاربرد شبکه‌های کانولوشنی در طبقه‌بندی تصاویر صورت گرفت و پر استفاده‌ترین شبکه‌ها در این حوزه از معماری با لایه‌های کانولوشن، توابع فعالیت، لایه‌های ادغام^{۴۲} و لایه‌های تمام متصل برای طبقه‌بندی استفاده می‌شود. در شکل ۱۰-۲ معماری یک شبکه‌ی طبقه‌بند برای طبقه‌بندی اعداد دست‌نویس مشاهده می‌شود که در ادامه به توضیح مختصر هر یک از لایه‌های استفاده شده می‌پردازیم.



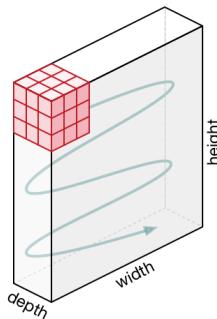
شکل ۲-۱۰: معماری یک شبکه‌ی کانولوشنی برای طبقه‌بندی اعداد دست‌نویس [۱۰]

۲-۳-۱ لایه‌های کانولوشنی

لایه‌های کانولوشنی شامل یکسری فیلترهایی هستند که با عمق ماتریس ویژگی برابری می‌کنند اما طول و عرض این فیلترها کوچکتر است و بنابراین با لغزاندن بر روی ماتریس ویژگی و در واقع محاسبه‌ی همبستگی^{۴۳} به جای کانولوشن، ماتریس ویژگی جدید محاسبه می‌گردد. به عنوان مثال، اگر ابعاد تصویر ورودی $3 * 100 * 100$ باشد حتماً عمق فیلتر باید با عمق تصویر برابر باشد بنابراین فیلتر می‌تواند ابعاد $3 * 5 * 5$ داشته باشد. شکل ۱۱-۲ یک ماتریس ویژگی و حرکت یک فیلتر بر روی آن را در یک لایه‌ی کانولوشنی را نشان می‌دهد

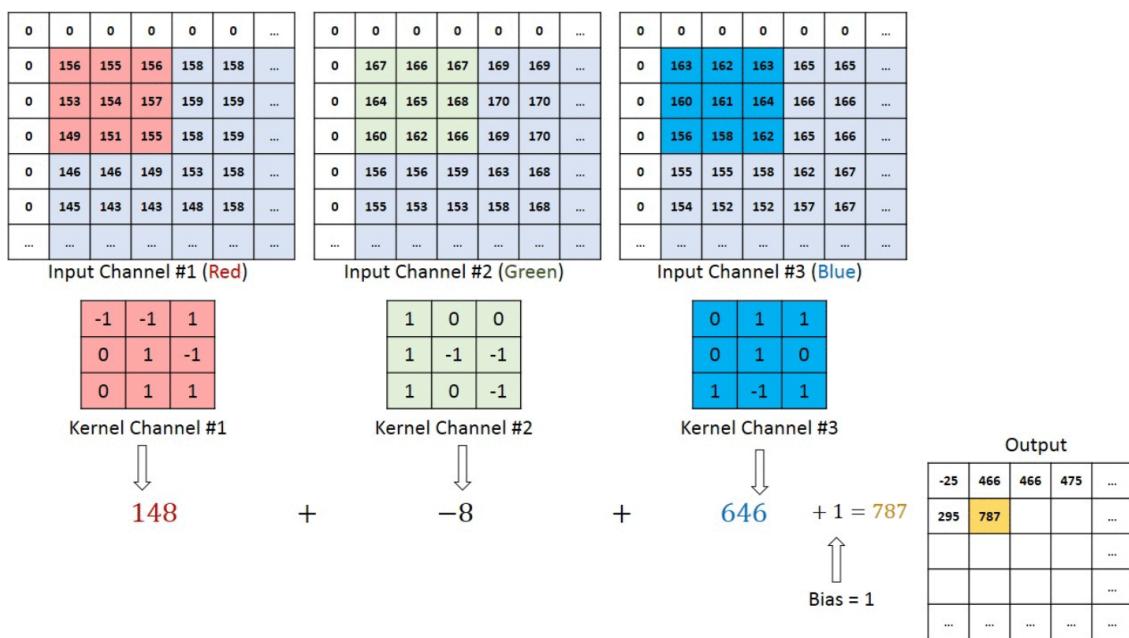
⁴²Pooling Layers

⁴³Correlation



شکل ۲-۱۱: چگونگی حرکت فیلتر در یک لایه کانولوشنی [۱۰]

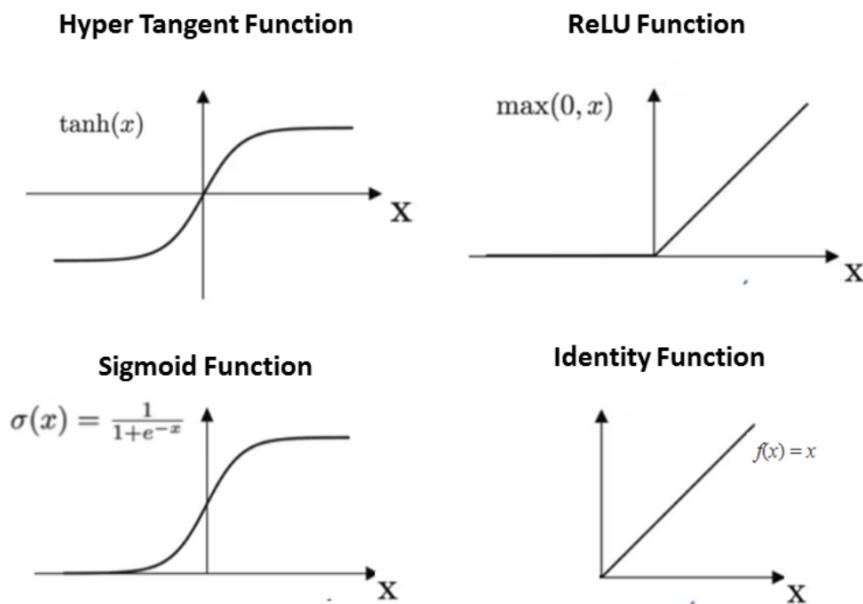
برای درک بهتر چگونگی محاسبات در یک لایه کانولوشنی، شکل ۲-۱۲ را در نظر بگیرید. همانطور که پیشتر گفته شد، عمق فیلتر و عمق تصویر باید برابر باشد (در اینجا ۳) سپس با لغزیده شدن فیلتر بر روی تصویر مشابه شکل قبل و انجام یک ضرب نقطه‌ای و مجموع‌گیری بین سه کanal و در نهایت اضافه کردن بایاس به مقدار بدست آمده، مقدار خروجی فیلتر محاسبه می‌شود.



شکل ۲-۱۲: روند محاسبات در یک لایه کانولوشنی [۱۰]

۲-۳-۲ توابع فعالیت

همانطور که در بخش‌های قبل گفته شد، یک نورون طبیعی با رسیدن به ولتاژ آستانه آتش می‌کند بنابراین برای شبیه‌سازی این رفتار از یک تابع فعالیت استفاده می‌شود. تابع فعالیت در واقع تصمیم می‌گیرد که نورون مدل‌سازی شده با چه مقادیری فعالیت کند و میزان فعالیت آن در شکل‌گیری خروجی نهایی چه مقدار باشد. همیچنین تابع فعالیت چون غیر خطی است، امکان تخمین و استخراج ویژگی‌های غیر خطی توسط شبکه فراهم می‌شود. شکل ۲-۱۳^{۴۴} چهار نمونه از پر استفاده‌ترین توابع فعالیت در شبکه‌های عمیق را نشان می‌دهد که عبارتند از: ReLU، Tanh، Sigmoid و Identity. به دلیل آنکه در شبکه‌های عمیق با مشکل محو شدن گرادیان مواجهیم بنابراین استفاده از ReLU برای رفع این مشکل بسیار بیشتر است.



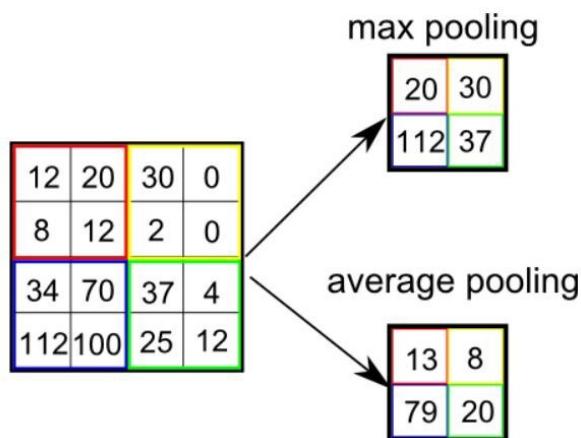
شکل ۲-۱۳: چهار نمونه از پر استفاده‌ترین توابع فعالیت [۱۱]

۲-۳-۳-۲ لایه‌های ادغام

لایه‌های ادغام وظیفه‌ی کاهش ابعاد ماتریس‌های ویژگی استخراج شده توسط لایه‌های کانولوشنی را دارند که این کار باعث کاهش حجم محاسبات مورد نیاز و افزایش سرعت خواهد شد. علاوه بر این این لایه‌ها ویژگی‌های غالب را استخراج می‌کنند که نسبت به دوران و جابجایی مقاوم هستند. نحوه‌ی

^{۴۴}Rectified Linear Unit

عملکرد لایه‌های ادغام مانند لایه‌های کانولوشنی است و بر روی ماتریس ویژگی لغزانیده می‌شود، با این تفاوت که در لایه‌های ادغام محاسبات نسبت به لایه‌های کانولوشنی متفاوت است. در شکل ۱۴-۲ دو نوع از پر استفاده‌ترین لایه‌های ادغام نشان داده شده است، ادغام میانگین‌گیر^{۴۵} و ادغام بیشینه‌گیر^{۴۶}. اندازه‌ی فیلترهای ادغام $2 * 2$ است و با گام ۲ بر روی ماتریس ویژگی لغزانیده شده‌اند. در ادغام میانگین‌گیر، بین ۴ عدد موجود در فیلتر میانگین‌گیری می‌شود و در ادغام بیشینه‌گیر، بیشینه به خروجی می‌رود.



شکل ۲-۱۴: ادغام میانگین‌گیر و ادغام بیشینه‌گیر [۱۰]

۴-۳-۲ لایه‌های تمام متصل

لایه‌های تمام متصل در بخش شبکه‌های پرسپترون چند لایه معرفی شد. به طور معمول این لایه‌ها برای طبقه‌بندی با استفاده از ویژگی‌های استخراج شده از دادگان استفاده می‌شوند. پیش از ظهور شبکه‌های عصبی عمیق که خودشان ویژگی‌های لازم را استخراج می‌کنند، در روش‌های یادگیری ماشین، ویژگی‌های یک مجموعه داده به صورت دستی و یا با روش‌های کلاسیک استخراج می‌شد و با استفاده از شبکه‌های تمام متصل وظیفه‌ی طبقه‌بندی تکمیل می‌شد. در شبکه‌های کانولوشنی این ویژگی‌ها توسط لایه‌های کانولوشن و ادغام استخراج می‌شود و در نهایت با دادن این ویژگی‌ها به لایه‌های تمام متصل کار طبقه‌بندی انجام می‌شود^{۴۹}. لایه‌های آخر شکل ۲-۱۰ این لایه را نشان می‌دهد.

از جمله شبکه‌های موفق کانولوشنی در حوزه‌ی تصاویر طبیعی می‌توان به AlexNet که

⁴⁵Average Pooling

⁴⁶Max Pooling

در سال ۲۰۱۲ ارایه شد و انقلاب شبکه‌های عمیق را آغاز کرد، اشاره نمود [۵۳]. پس از ارایه‌ی این مدل، معماری‌های جدید و روش‌های تعمیم‌پذیری^{۴۷} زیادی ارایه گردید که نتایج را در زمینه‌ی طبقه‌بندی تصاویر حتی از انسان نیز بهتر کرد. از جمله‌ی این معماری‌ها می‌توان به معماری‌های VGGNet، ResNet، GoogleNet و ... اشاره نمود [۵۴].

۴-۲ روش‌های تعمیم‌پذیری

به هرگونه تکنیکی که قدرت تعمیم مدل را افزایش می‌دهد و خطای تعمیم‌پذیری را کم می‌کند، روش تعمیم‌پذیری گفته می‌شود. از جمله‌ی این روش‌ها می‌توان به موارد زیر اشاره نمود:

- زود هنگام^{۵۱} جریمه‌ی وزن‌های شبکه^{۴۸}
- Dropout افزایش مجموعه دادگان^{۴۹}
- روش مقاومت در برابر نویز^{۵۰} Batch-Normalization

در ادامه به معرفی مختصر هر یک از این روش‌ها پرداخته خواهد شد.

جریمه‌ی وزن‌های شبکه

برای جلوگیری از یادگیری بیش از حد^{۵۲} بر روی داده‌های آموزش یکی از روش‌ها، جریمه‌ی وزن‌های شبکه است به این صورت که با اضافه کردن یک جمله بهتابع هزینه‌ی شبکه، که آن جمله می‌تواند نرم اول یا دوم وزن‌های شبکه باشد؛ از بزرگ شدن بیش از حد وزن‌ها جلوگیری می‌کند و همچنین با کاهش پیچیدگی مدل، قدرت تعمیم‌پذیری آن بالا می‌رود.

⁴⁷Regularization Terms

⁴⁸Parameter (weights) Penalties

⁴⁹Dataset Augmentation

⁵⁰Noise Robustness (input/output)

⁵¹Early Stopping

⁵²Overfitting

افزایش مجموعه دادگان

در مدل‌های عصبی، هرچقدر تنوع و تعداد دادگان بیشتر باشد در نهایت مدل بهتر آموزش می‌بیند بنابراین روش‌هایی برای افزایش تعداد دادگان پیشنهاد شده است که به عنوان مثال می‌توان به دوران با زوایای تصادفی، چرخش عمودی و افقی، تغییرات شدت روشنایی و ... اشاره نمود.

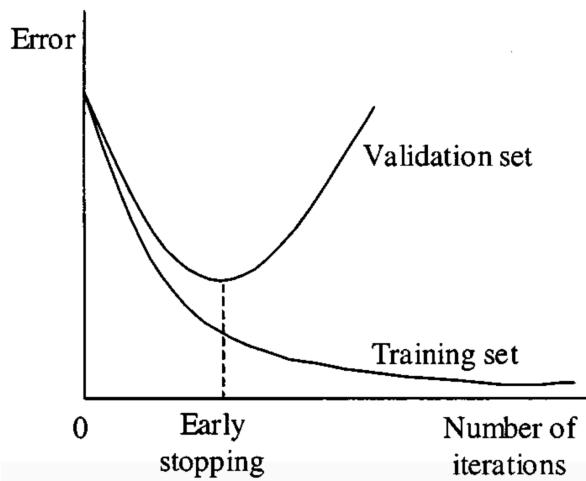
مقاومت در برابر نویز

با اضافه کردن نویز به ورودی یا خروجی مدل و بالا بردن عدم قطعیت، شبکه سعی در استخراج ویژگی‌های مقاوم‌تری می‌کند و در نتیجه مدل نهایی قدرت تعمیم‌پذیری بالاتری دارد [۵۵].

توقف زود هنگام

اگر دادگان آموزش را به دو مجموعه داده‌ی آموزش و ارزیابی تقسیم کنیم (برروی دادگان آموزش وزن‌های شبکه تغییر می‌کنند و دادگان ارزیابی در تغییر وزن‌ها نقشی ندارند) و در انتهای هر epoch^{۵۳} مقدار خطای شبکه را برای این دو گروه محاسبه کنیم نموداری مشابه شکل زیر بدست می‌آید که نشان می‌دهد پس از مدتی، شبکه در حال حفظ کردن دادگان آموزش است و خطای دادگان ارزیابی افزایش می‌یابد بنابراین قدرت تعمیم‌پذیری شبکه در حال کم شدن است. به توقف روند آموزش بعد از مشاهده روند افزایشی خطای تعمیم‌پذیری، توقف زود هنگام گفته می‌شود.

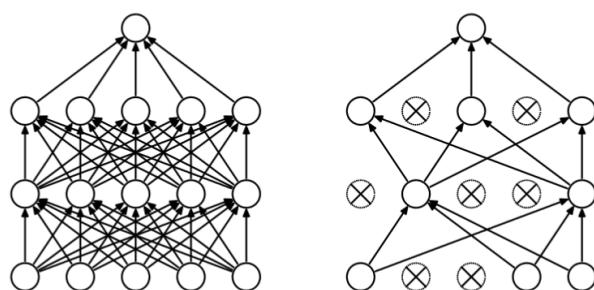
^{۵۳} به هر بار آموزش یک مدل بر روی کل مجموعه دادگان یک epoch گفته می‌شود.



شکل ۲-۱۵: توقف زود هنگام با توجه به خطای داده‌های ارزیابی [۱۲]

روش Dropout

اگر در یک مدل عصبی دو نورون ویژگی‌هایی که استخراج می‌کنند مشابه باشند و یا همبستگی بالایی داشته باشند این شبکه به صورت کارآمد عمل نمی‌کند. برای رفع این مشکل تکنیک dropout پیشنهاد شد که با یک درصد احتمال در هر تکرار آموزش، تعدادی از نورون‌ها را غیر فعال می‌کند بنابراین نورون‌های دخیل در روند آموزش مجبور به یادگیری ویژگی‌های مستقل از هم و کارآمدتر هستند و در نتیجه قدرت تعمیم‌پذیر مدل افزایش پیدا می‌کند. در شکل ۲-۱۶ یک شبکه با و بدون استفاده از dropout نشان داده شده است که در استفاده از dropout تعدادی از نورون‌ها حذف شده‌اند و نورون‌های باقی مانده مجبور به استخراج ویژگی‌های کاراتر برای پیش‌بینی نتیجه‌ی مطلوب هستند.



شکل ۲-۱۶: چگونگی اعمال تکنیک dropout [۱۳]

روش Batch-Normalization

این متدها مشکل ناهمگونی تابع چگالی احتمال داده‌ها در لایه‌های میانی را حل می‌کند. این مشکل به این صورت پدید می‌آید که در هر batch از داده‌های ورودی، در روند آموزش یک توزیع احتمال پدید می‌آید که این توزیع‌های احتمال با یکدیگر متفاوت‌اند. مشکل مهم‌تر آنکه، این ناهمگونی توزیع‌ها در خروجی لایه‌های شبکه نیز پدید می‌آید. برای رفع این مشکل تکنیک Batch-Normalization معرفی گردید که ابتدا تمام توزیع‌ها را با توجه به میانگین و انحراف معیارشان نرمال‌سازی می‌کند و سپس با یک سری پارامترهای قابل یادگیری این توزیع‌ها را تغییر می‌دهند تا همسان‌سازی نهایی صورت پذیرد [۵۶].

۵-۲ پردازش تصویر و بینایی کامپیوتر

چشم انسان از تعداد بسیار زیادی سلول حساس به نور تشکیل شده‌است که با دریافت نور محیطی و پردازش آن باعث ایجاد یک درک از محیط و اشیای اطراف می‌شود. این روند علی‌رغم اینکه برای انسان بسیار ساده است اما برای یک ماشین تا بتواند درک از تصویر دریافت شده داشته باشد، بسیار پیچیده است. بنابراین زمینه‌های مختلفی برای پردازش و ایجاد درک از تصویر دیجیتال برای ماشین‌ها ایجاد شده است.

در بینایی کامپیوتر مشابه قدرت بینایی انسان، هدف ایجاد درک و فهم عمیق یک تصویر یا یک ویدیو است اما پردازش تصاویر دیجیتال، پردازش تصاویر با استفاده از یک کامپیوتر است. این پردازش‌ها می‌توانند شامل بهبود کیفیت تصویر و حذف نویز با فیلتر کردن تصویر و دستکاری شدت‌های آن باشد. بنابراین زمینه‌ی پردازش تصویر زیر مجموعه‌ی بینایی کامپیوتر قرار می‌گیرد.

حوزه‌ای بینایی کامپیوتر مرتبط با مبحث هوش مصنوعی و یادگیری ماشین است که با هدف ایجاد چارچوب‌های لازم برای پیاده‌سازی قابلیت بینایی در کامپیوترها و سیستم‌های کامپیوتری شکل گرفته است. دانشمندان فعل در حوزه بینایی کامپیوتر، سعی در تولید تکنیک‌ها و روش‌هایی دارند که مفهوم دیدن را برای یک کامپیوتر تعریف می‌کنند. از این طریق، کامپیوترها توانایی شناسایی و درک محتوای موجود در تصاویر دیجیتال و ویدیوها را پیدا می‌کنند [۵۷].

۶-۲ جمع‌بندی

در این فصل با مفاهیم اولیه‌ی این مطالعه آشنایی پیدا شد. در ابتدا مفاهیم تصویربرداری پزشکی معرفی گردید و دو نوع تصویربرداری پرکاربرد (ام‌آرآی و سی‌تی اسکن) برای تشخیص و درمان به صورت دقیق‌تر بررسی شد. در ادامه مفاهیم هوش مصنوعی، یادگیری ماشین و یادگیری عمیق معرفی شدند که با توجه به رشد روز افزون این حوزه‌ها و مدل‌های پرکاربرد، استفاده‌ی این مدل‌ها در بینایی کامپیوتر و پردازش تصویر بسیار رشد نموده است. بنابراین در این پژوهه در نقطه‌ی مشترک حوزه‌های معرفی شده قرار گرفته‌ایم و قصد برآن است با استفاده از مدل‌های یادگیری عمیق در حوزه‌ی پردازش تصاویر پزشکی، قطعه‌بندی تومور و ساختارهای در ریسک را با دقت و سرعت بالاتر نسبت به الگوریتم‌های پیشین، با ابزارهای معرفی شده، انجام دهیم. در فصل بعد به بررسی و مرور سوابق کارهای انجام شده در حوزه‌ی قطعه‌بندی تصاویر پزشکی پرداخته خواهد شد.

فصل ۳

مرور ادبیات

۱-۳ مقدمه

در این فصل سوابق و مطالعات انجام شده برای قطعه‌بندی تصاویر پزشکی بررسی خواهند شد. در ابتدا مفاهیم مقدماتی و تعاریف مرتبط با این حوزه معرفی می‌شوند و سپس روش‌های پیشنهاد شده، برای قطعه‌بندی تصاویر پزشکی مورد مطالعه قرار خواهند گرفت. در نهایت رویکردهای قطعه‌بندی تومور و ساختارهای در ریسک با استفاده از روش‌های یادگیری عمیق معرفی و بررسی می‌شوند.

۲-۳ تعاریف و مفاهیم اولیه

در این قسمت تعاریف و مفاهیم اولیه‌ی مرتبط با قطعه‌بندی ارایه می‌گردد. در ابتدا مفهوم قطعه‌بندی معرفی می‌شود و سپس اهداف و چالش‌های قطعه‌بندی بررسی می‌شود، سپس انواع دسته‌بندی روش‌های قطعه‌بندی معرفی می‌گردد و در نهایت معیارهای ارزیابی قطعه‌بندی مطالعه خواهد شد.

۳-۱ تعریف قطعه‌بندی

همانطور که در فصل دوم اشاره گردید، یک تصویر پزشکی می‌تواند دو بعدی (تصاویر رادیوگرافی) یا سه بعدی (تصاویر ام‌آرآی و سی‌تی اسکن) که به عناصر سازنده‌ی این تصاویر بر اساس این ابعاد، پیکسل

و یا واکسل گفته می‌شود. برای قطعه‌بندی تصاویر تعاریف متنوعی ارایه شده است که در زیر به بررسی هر یک پرداخته می‌شود.

- به جزء‌بندی یک تصویر به مناطق غیر هم‌پوشان همگن که یک ویژگی یا یک معیار مشترک مانند شدت نوری^۱، شکل^۲ و یا بافت^۳ همسان دارند قطعه‌بندی گفته می‌شود.
- به روند نسبت دادن یک برجسب با مفهوم بیولوژیکی خاص به پیکسل‌ها و یا واکسل‌ها که این پیکسل‌ها و واکسل‌ها یک ویژگی خاص مشترک دارند و یا متعلق به یک آناتومی مشترک هستند.
- جداسازی ساختارهای مدنظر از یکدیگر و از پس زمینه را قطعه‌بندی می‌گویند [۲۰].

با توجه به تعاریف انجام شده، می‌توان یک تعریف ریاضی برای این عملیات ارایه داد. اگر تصویر را I بنامیم، آنگاه قطعه‌بندی یعنی مشخص کردن مجموعه‌هایی (S_k) در تصویر که $I \subset S_k$ به صورتی که اعضای این مجموعه‌ها دارای یک خاصیت مشترک باشند و این مجموعه‌ها هیچ اشتراکی با یکدیگر نداشته باشند:

$$\begin{aligned} \text{Property}(S_k) &= \text{True} \quad \forall x \in S_k \\ S_k \cap S_j &= \emptyset \quad \text{for } k \neq j \end{aligned} \tag{۱-۳}$$

به صورت ایده‌آل قطعه‌بندی به صورتی انجام می‌شود که هر یک از S_k ‌ها یک ساختار و آناتومی خاص را نشان می‌دهد. شرط دیگر در این فرمول‌بندی متصل بودن هریک از این S_k ‌هاست که برای ساختارهای آناتومیکی یکپارچه این شرط باید برقرار شود. اگر از این شرط چشم‌پوشی کنیم قطعه‌بندی در سطح پیکسل (واکسل) صورت می‌گیرد و در مواردی استفاده می‌شود که ساختارهای مورد نظر از یکدیگر گستته هستند اما خواص مشابه دارند [۵۸].

۲-۲-۳ اهداف قطعه‌بندی تصاویر پزشکی

در روند قطعه‌بندی تصاویر پزشکی، جزئیات مورد نظر در استخراج قطعه‌ها، بسیار به کاربرد کیلینیکی بستگی دارد. هدف قطعه‌بندی بهبود نمایش تصاویر و اجزای مختلف آن برای تشخیص و درمان کارآمدتر است. به طور کلی از مهم‌ترین اهداف قطعه‌بندی می‌توان به موارد زیر اشاره کرد:

¹Intensity

²Shape

³Texture

- کمی‌سازی پیشرفت بیماری
- تشخیص بیماری
- استخراج ویژگی
- مطالعه‌ی پیشرفت بیماری
- موقعیت یابی تومور
- تهیه‌ی نقشه‌ی درمان
- تطبیق تصاویر
- تخمین حجم تومور

به طور کلی قطعه‌بندی تصاویر پزشکی سهم عمدہ‌ای در تحلیل و پردازش تصاویر پزشکی دارد که امکان بررسی مشکلات عملکردی ساختارهای آناتومیکی مختلف را فراهم می‌سازد و تمام فاکتورهای لازم برای تشخیص، کمی‌سازی و مطالعه‌ی پیشرفت بیماری را میسر می‌سازد [۵۹، ۶۰].

۳-۲-۳ چالش‌های قطعه‌بندی تصاویر پزشکی

قطعه‌بندی تصاویر پزشکی، با مشکلات زیادی مواجه است که کیفیت قطعه‌بندی نهایی به دلیل وجود برخی اشکالات در تصویر اصلی پدید می‌آید. از جمله‌ی این اشکالات می‌توان به موارد زیر اشاره نمود [۶۱].

نویز

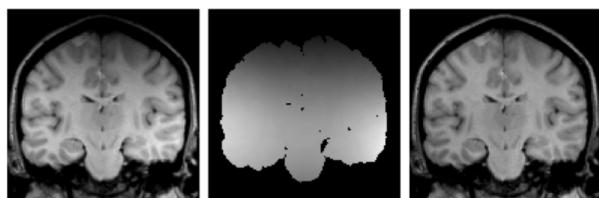
با حضور نویز در تصویر عدم قطعیت بالا می‌رود و مقدار شدت پیکسل‌ها در تصویر تغییر می‌کند که این تغییر باعث دگرگونی ساختار همگن بافت‌ها می‌شود و در نتیجه قطعه‌بندی دشوار می‌گردد [۶۲].

ناهمگنی شدت

ناهمگنی شدت^۴ به تغییرات جعلی و هموار در شدت‌های داخل تصویر گفته می‌شود که نمایانگر یک بافت با یک ویژگی مشترک هستند اما به دلیل تغییرات در میدان دستگاه تصویرگر، پدید می‌آیند. این مشکل اغلب در تصاویر ام‌آرای به دلیل تغییر میدان مغناطیسی آن بوجود می‌آیند. شکل ۱-۳ در سمت چپ یک تصویر مغزی دچار این مشکل نشان داده شده است و در شکل وسط، ناهمگنی میدان دستگاه

⁴Intensity In-homogeneity

تصویربرداری استخراج شده و در نهایت در شکل راست با تقسیم تصویر مشکل دار بر ناهمگنی میدان، تصویر ترمیم شده بدست آمده است [۱۴].



شکل ۳-۱: ناهمگنی شدت و میدان در یک تصویر ام‌آرای [۱۴]

اثر حجم جزیی

اثر حجم جزیی^۵ به این دلیل بوجود می‌آید که هر واکسل یک حجم خاص از فضا را با یک مقدار شدت تعريف می‌کند حال اگر دو بافت مختلف در یک واکسل قرار گیرد، مقدار شدت این واکسل ترکیبی از جمع وزن دار شدتهای این دو بافت است و بنابراین اثر حجم جزیی لبه‌های تصویر را هموارتر و محoter می‌کند [۶۳].

علاوه بر چالش‌های ذکر شده، چالش‌های دیگری مانند: حرکت بیمار در طی تصویربرداری، شدتهای یکسان در بافت‌های مختلف، لبه‌های ضعیف در بافت‌های نرم و ... باعث دشوارتر شدن قطعه‌بندی تصاویر پزشکی می‌شوند.

۴-۲-۳ انواع قطعه‌بندی تصاویر پزشکی

انواع قطعه‌بندی تصاویر پزشکی را بر اساس بعد تصویر و چالش‌های موجود در قطعه‌بندی، می‌توان به دسته‌های مختلفی تقسیم‌بندی کرد. در زیر به بررسی انواع دسته‌بندی روش‌های قطعه‌بندی پرداخته خواهد شد.

⁵Partial Volume Effect

قطعه‌بندی بر اساس بعد تصویر

همانطور که پیش‌تر اشاره گردید، تصاویر پزشکی می‌تواند دو بعدی یا سه‌بعدی باشند بنابراین روش‌های مختلفی بر اساس بعد تصاویر ارایه شده است که می‌تواند در تصاویر دو بعدی یا سه‌بعدی قطعه‌بندی را انجام دهند. همچنین برای تصاویر سه‌بعدی می‌توان الگوریتم قطعه‌بند دو بعدی را در لایه‌های مختلف تصویر اجرا کرد و نتیجه را پشت سر هم قرار داد تا قطعه‌بندی تصویر سه‌بعدی انجام شود. مزیت این روش کاهش هزینه‌های محاسباتی، پیاده‌سازی ساده‌تر و کاهش حافظه‌ی مورد نیاز است. روش‌های سه‌بعدی برای کاربردهای خاص مانند استخراج رویه و سطح ساختارهای سه‌بعدی ارایه شده‌اند که در این حالت نمی‌توان از روش‌های دو بعدی استفاده نمود [۶۴].

قطعه‌بندی نرم و سخت

قطعه‌بندی سخت^۶ برای یک پیکسل (واکسل) به صورت صفر و یک تعریف می‌شود، یا پیکسل متعلق به کلاس مورد نظر است یا نیست. تعریف ریاضیاتی قطعه‌بندی سخت را می‌توان به صورت فرمول ۲-۳ ارایه داد که در آن $m_k(x_{i,j})$ تابع عضویت پیکسل $x_{i,j}$ در موقعیت j, i تصویر به کلاس m است. در این نوع قطعه‌بندی امکان حل مشکل اثر حجم جزئی وجود ندارد.

$$m_k(x_{i,j}) = \begin{cases} 1, & x_{i,j} \subset S_k \\ 0, & \text{otherwise} \end{cases} \quad (2-3)$$

در دست دیگر، قطعه‌بندی نرم^۷ قرار دارد که امکان طبقه‌بندی پیکسل‌ها به کلاس‌های مختلف را با یک احتمال تعلق نسبت به تمام کلاس‌ها می‌دهد. فرمول‌بندی این نوع قطعه‌بندی را می‌توان در معادلات زیر ملاحظه نمود. که در آن K تعداد کل کلاس‌های موجود است.

$$\begin{cases} 0 \leq m_k(x_{i,j}) \leq 1 \\ \sum_{k=1}^K m_k(x_{i,j}) = 1 \end{cases} \quad (3-3)$$

با استفاده از قطعه‌بندی نرم، می‌توان اثرات همپوشانی پیکسل‌ها و اثر حجم جزئی را حل نمود. قطعه‌بندی نرم را با یک آستانه‌گذاری می‌توان به قطعه‌بندی سخت تبدیل کرد [۶۵].

⁶Hard Segmentation

⁷Soft Segmentation

قطعه‌بندی گسسته و پیوسته

در قطعه‌بندی گسسته، خروجی نهایی یک نقشه‌ی گسسته، مانند تصویر اولیه است و بر اساس فاصله‌ی فضایی پیکسل‌ها (واکسل‌ها) این قطعه‌بندی صورت می‌گیرد. در سمت دیگر، قطعه‌بندی پیوسته، در یک فضای پیوسته عمل می‌کند و امکان قطعه‌بندی به صورت زیرپیکسل^۱ وجود دارد. به عنوان مثال در مدل‌های شکل پذیر قطعه‌بندی نهایی به صورت پیوسته است [۶۶].

قطعه‌بندی بر اساس تعامل کاربر

انواع روش‌های قطعه‌بندی را بر اساس تعامل کاربر می‌توان به دسته‌های زیر تقسیم‌بندی نمود:

- **روش‌های دستی:** ئر این گونه روش‌ها، متخصص به صورت دستی و تعیین طبقه‌ی هر یک از پیکسل‌ها (واکسل‌ها) اقدام به قطعه‌بندی می‌کنند که یک عملیات بسیار زمان‌بر و خسته‌کننده برای متخصص و همچنین هزینه‌بر برای استخدام متخصص است.
- **روش‌های نیمه خودکار:** در این نوع روش‌ها، متخصص با ایجاد یک نقطه‌ی اولیه برای شروع الگوریتم، الگوریتم را راه‌اندازی می‌کند و در نهایت خروجی قطعه‌بندی ایجاد می‌شود.
- **روش‌های خودکار:** در این نوع روش‌ها عملیات قطعه‌بندی کاملاً به صورت خودکار انجام می‌شود و در نتیجه می‌توان تعداد بسیار زیادی تصویر را با یک دقت خاص، بدون دخالت فرد، بدست آورد. از دیگر مزیت‌های این روش‌ها نسبت به روش‌های قبل سرعت بالاتر است [۶۷].

۵-۲-۳ ارزیابی قطعه‌بندی

برای کمی‌سازی عملکرد روش قطعه‌بندی روش‌های ارزیابی بسیار حیاتی هستند. روش‌های ارزیابی را می‌توان به دو دسته‌ی عمدۀ تقسیم‌بندی کرد:

- **مقایسه با قطعه‌بندی متخصص:** این روش پر استفاده‌ترین روش برای ارزیابی الگوریتم‌های قطعه‌بندی است که در آن از قطعه‌بندی شخص متخصص به عنوان استاندارد استفاده می‌شود و در نهایت با معیارهای ارزیابی بین خروجی مدل قطعه‌بند و استاندارد موجود، عملکرد قطعه‌بندی

^۱Subpixel

کمی‌سازی می‌گردد. از جمله مشکلات این روش می‌توان به، زمان‌بر و هزینه‌بر بودن قطعه‌بندی توسط متخصص و همچنین اختلاف نظر بین متخصصین مختلف اشاره کرد.

- استفاده از فانتوم‌های فیزیکی و ریاضیاتی: با استفاده از فانتوم‌های^۹ ریاضیاتی و فیزیکی، اجزای مختلف فانتوم به طور کامل از یکدیگر تفکیک‌پذیر هستند و یک تصویرسازی دقیق از روند تصویربرداری ایجاد می‌کنند. بنابراین با قرار دادن قطعه‌بندی فانتوم به عنوان استاندارد، می‌توان الگوریتم قطعه‌بندی را ارزیابی نمود [۶۸].

با توجه به تعریف‌های فوق ملاحظه گردید که لازم است برای کمی‌سازی ارزیابی روش قطعه‌بندی معیارهایی تعریف گردد. در ادامه به معرفی این معیارها پرداخته می‌شود.

۳-۲-۶ معیارهای ارزیابی قطعه‌بندی

برای کمی‌سازی عملکرد قطعه‌بندی معیارهای بسیار زیادی ارایه گردیده است که از جمله مهمترین آن‌ها می‌توان به موارد زیر اشاره نمود:

- اختصاص^{۱۴}
- ماتریس درهم آمیختگی^{۱۰}
- معيار Dice
- دقت^{۱۱}
- معيار Jaccard
- حساسیت^{۱۲}
- فاصله‌ی Hausdorff
- صحت^{۱۳}

معیارهای پر استفاده و شناخته‌شده در قطعه‌بندی تصاویر شامل سه معيار آخر (معیار Dice، معيار Jaccard و فاصله‌ی Hausdorff) هستند که در ادامه به بررسی بیشتر آن‌ها پرداخته می‌شود.

⁹Phantom

¹⁰Confusion Matrix

¹¹Accuracy

¹²Sensitivity

¹³Precision

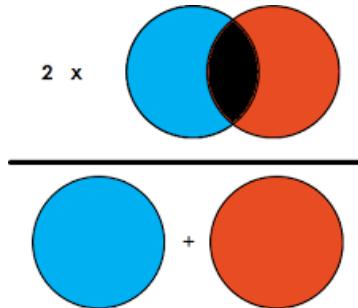
¹⁴Specificity

Dice معيار

F1-Score یک نشانگر آماری برای سنجش شباهت میان دو نمونه است. این معيار با نام نیز شناخته می‌شود. تعریف ریاضیاتی این معيار به صورت زیر است که A و B در بحث قطعه‌بندی خروجی الگوریتم و استاندارد مورد نظر هستند.

$$DSC(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (4-3)$$

شکل ۲-۳ شهود عملکرد این معيار را نشان می‌دهد. در واقع هرچقدر همپوشانی دو ناحیه بیشتر باشد این معيار به یک نزدیکتر می‌شود و اگر دو ناحیه به طور کامل از یکدیگر جدا باشند مقدار این معيار برابر صفر می‌گردد.



[۱۵] شهود عملکرد معيار Dice

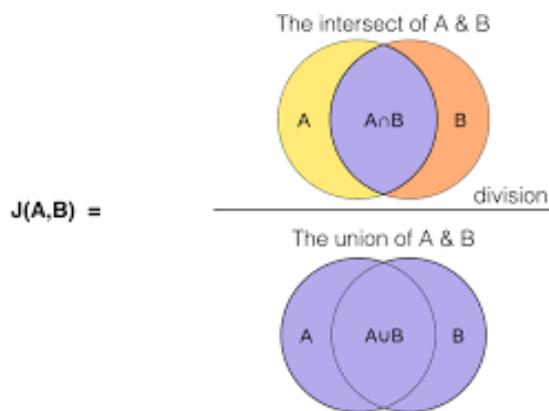
Jaccard معيار

معیار Jaccard نیز مانند معيار Dice یک نشانگر آماری است که میزان مشابهت و تفاوت را میان مجموعه‌ی نمونه‌ها نشان می‌دهد. این معيار با نام اشتراک بر اجتماع^{۱۵} نیز شناخته می‌شود. فرمول‌بندی این معيار به صورت زیر است.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (5-3)$$

برای شهود بیشتر، شکل ۳-۳ را در نظر بگیرید. هرچه میزان همپوشانی دو ناحیه بیشتر شود این معيار به یک نزدیکتر می‌شود و اگر این دو ناحیه از هم جدا باشند، مقدار این معيار برابر با صفر خواهد شد.

¹⁵Intersection over Union



شکل ۳-۳: شهود عملکرد معیار Jaccard [۱۶]

فاصله‌ی Hausdorff

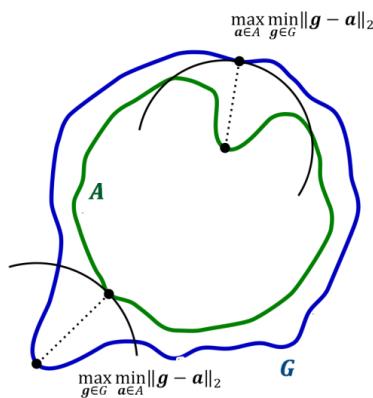
فاصله‌ی Hausdorff در واقع معیاری است که مشخص می‌کند دو مجموعه داده چه مقدار از یکدیگر دوراند. اگر نقاط متناظر در دو مجموعه داده نزدیک به یکدیگر باشند، این فاصله به صفر نزدیک می‌شود و هر چه این نقاط متناظر فاصله‌ی بیشتری بگیرند، مقدار این معیار بزرگتر خواهد شد. تعریف ریاضیاتی این فاصله به صورت زیر است:

$$HD(G, A) = \max(hd(G, A), hd(A, G)) \quad (6-3)$$

$$hd(G, A) = \max_{g \in G} \min_{a \in A} \|g - a\|_2$$

$$hd(A, G) = \max_{a \in A} \min_{g \in G} \|g - a\|_2$$

که در آن، A خروجی الگوریتم قطعه‌بندی و G استاندارد مطلوب و $\|\cdot\|_2$ نشان دهنده نرم اقلیدسی است. در فاصله Hausdorff به صورت شهودی مطابق شکل ۴-۳ بیشینه‌ی کمینه فاصله‌های لبه‌های خروجی الگوریتم با لبه‌های استاندارد مطلوب محاسبه می‌شود.



[۱۷] شهود عملکرد فاصله‌ای Hausdorff

۳-۳ روش‌های قطعه‌بندی

با نگاه در طول زمان، روش‌ها و تکنیک‌های پیشنهاد شده‌ی قطعه‌بندی با گذشت زمان پیشرفت چشمگیری داشته‌اند. روش‌ها و الگوریتم‌های معرفی شده در این حوزه بسیار فراوان هستند. برای بررسی این روش‌ها در این مطالعه، بر اساس تمرکز روش ارایه شده بر یک ویژگی غالب در تصویر یا یک زمینه‌ی خاص، دسته‌بندی صورت گرفته است و سپس در هر دسته این روش‌ها معرفی شده‌اند. در ادامه به بررسی روش‌های موجود در زمینه‌ی قطعه‌بندی پرداخته می‌شود.

۱-۳-۳ روش‌های آستانه‌گذاری

آستانه‌گذاری^{۱۶} یکی از پر استفاده‌ترین روش‌های قطعه‌بندی است که بر اساس شدت پیکسل‌ها عمل می‌کند. با اعمال یک آستانه به شدت پیکسل‌های تصویر می‌توان پیش‌زمینه را از پس‌زمینه تفکیک داد و یک قطعه‌بندی دو کلاس صورت داد. معادله‌ی ۷-۳ توصیف ریاضیاتی قطعه‌بندی (S) یک تصویر دو بعدی X را در مختصات j ، با آستانه‌ی T نشان می‌دهد.

$$S(i, j) = \begin{cases} 1 & X(i, j) \geq T \\ 0 & X(i, j) < T \end{cases} \quad (7-3)$$

¹⁶Thresholding

روش آستانه‌گذاری را می‌توان با تعیین چندین آستانه، به یک قطعه‌بند چند کلاس تعمیم داد. نکته‌ی مهم در این روش، پیدا کردن آستانه بهینه است. برای این منظور روش‌های زیادی ارایه گردیده است که از جمله این روش‌ها می‌توان به موارد زیر اشاره نمود:

- **تئوری تصمیم بیز^{۱۷}**: در این روش، در ابتدا با متدهای تخمین تابع توزیع (مانند مدل مخلوط گاوی^{۱۸})، توزیع شدت پیکسل‌ها استخراج می‌شود. در نهایت روش بیز با کمینه کردن مجموع مقدار مثبت و منفی کاذب^{۱۹} مقدار آستانه‌ی بهینه را استخراج می‌کند [۶۹].
- **روش آستانه‌گذاری اوتسو^{۲۰}**: در این روش با بیشینه‌سازی یک تابع هزینه، مقدار آستانه‌ی بهینه بدست می‌آید. این تابع هزینه از دو قسمت واریانس بین کلاس‌ها و واریانس درون کلاسی تشکیل شده است که با بیشینه کردن تابع هزینه، واریانس بین کلاس‌ها بیشینه و واریانس درون کلاس‌ها کمینه می‌شود و در نهایت آستانه‌ی مطلوب بدست می‌آید [۷۰].

۲-۳-۳ روش رشد ناحیه‌ای

رشد ناحیه‌ای یک روش قطعه‌بندی تعاملی است که یک هسته‌ی اولیه توسط کاربر برای قطعه‌بندی انتخاب می‌شود. سپس با مقایسه‌ی شدت روشنایی هسته‌ی اولیه با پیکسل‌های کناری، اگر این اختلاف کوچکتر از یک آستانه بود این ناحیه توسعه پیدا می‌کند تا در نهایت به پیکسل‌های با شدت‌های متفاوت برسد. مشکل این روش قطعه‌بندی نیمه-خودکار بودن آن است و لازم است هر بار توسط یک متخصص برای هر ناحیه، مقدار دهی اولیه صورت گیرد. همچنین قبل از قطعه‌بندی با این روش باید پیش پردازش‌هایی برای بهبود و همنواخت کردن شدت پیکسل‌های متعلق به یک بافت صورت گیرد [۷۱].

۳-۳-۳ روش‌های طبقه‌بندی

روش‌های مبتنی بر طبقه‌بندی برای قطعه‌بندی تصاویر، معمولاً با استخراج نقشه‌ی ویژگی^{۲۱} از تصویر و دادن این ویژگی‌ها به یک طبقه‌بند برای تصمیم‌گیری نهایی صورت می‌گیرد. این نقشه‌ی ویژگی می‌تواند

¹⁷Bayes Decision Theorem

¹⁸Gaussian Mixture Models (GMM)

¹⁹False Positive and False Negative

²⁰Otsu's Thresholding

²¹Feature Map

شامل ویژگی‌های سطح پایین مثل شدت روشنایی و یا پاسخ تصویر نسبت به یک فیلتر تا ویژگی‌های سطح بالاتر استخراج شده توسط الگوریتم‌های دیگر باشد^{۲۲}. روش‌های طبقه‌بندی معمولاً از مدل‌های شناخت الگو بهره می‌برند و به نحوی می‌توان این روش‌ها را یادگیری با سرپرست نامید که نتیجه‌ی مطلوب و استاندارد مورد نظر در دسترس است^[۷۲]. روش‌های طبقه‌بندی را می‌توان به تکنیک‌های زیر تقسیم‌بندی نمود.

روش بیشینه‌کردن شباهت

در روش بیشینه‌کردن شباهت^{۲۳} هدف ارایه دادن یک طبقه‌بند بر اساس تابع احتمال با کمترین خطای ممکن است. در این تکنیک، نقشه‌ی ویژگی استخراج شده باید خاصیت احتمالاتی ویژگی‌ها را داشته باشد تا این الگوریتم به درستی عمل کند^[۷۳].

نزدیک‌ترین همسایگی

روش نزدیک‌ترین همسایگی^{۲۴} یک روش غیر پارامتری^{۲۵} است که با هدف طبقه‌بندی ارایه شده است. در این روش با بررسی ویژگی یک نقطه در نقشه‌ی ویژگی و مقایسه با نقاط نزدیک در نقشه‌ی ویژگی برچسب مطلوب اعمال می‌گردد.

شبکه‌های عصبی

استفاده از شبکه‌های عصبی برای طبقه‌بندی یک روش پر استفاده است که در این تکنیک ابتدا ویژگی‌های مورد نظر از تصویر استخراج می‌گردد و با رویکرد یادگیری با سرپرست در صورت وجود خروجی مطلوب و یا یادگیری بدون سرپرست در صورت نبود استاندارد مورد نظر، شبکه نسبت به طبقه‌بندی یادگیری را صورت می‌دهد. توجه شود که در این نوع روش‌ها معمولاً ویژگی‌ها از پیش استخراج شده‌اند و عملکرد شبکه‌ی عصبی فقط طبقه‌بندی است^[۷۴].

^{۲۲}مانند الگوریتم‌های SURF ، SIFT و ...

²³Maximum Likelihood

²⁴Nearest neighbor

²⁵Non-Parametric

درخت تصمیم

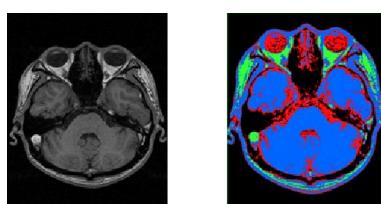
درخت تصمیم^{۲۶} یک تکنیک ساده و پر استفاده برای طبقه‌بندی است. طبقه‌بند درخت تصمیم مجموعه‌ای از پرسش‌ها و ارزیابی‌ها را در یک شکل درختی ساختار می‌دهد و با بررسی ویژگی‌ها در هر شاخه در نهایت به یک تصمیم بله یا خیر می‌رسد [۷۵].

۴-۳-۳ روش‌های خوشبندی

روش‌های خوشبندی مانند روش‌های طبقه‌بندی یک هدف را دنبال می‌کنند اما در این رویکرد بر عکس طبقه‌بندی از یادگیری بدون سرپرست استفاده می‌شود و داده‌ها بدون حضور استاندارد مطلوب، قطعه‌بندی می‌شوند. دو نوع روش خوشبندی متداول در حوزه‌ی قطعه‌بندی بسیار استفاده می‌شوند که در ادامه به معرفی هر یک پرداخته می‌شود.

K-means

در تکنیک K-means، ابتدا تعداد خوشبندی مدنظر توسط کاربر مشخص می‌شود و سپس به ازای هر خوشبندی یک مرکز خوشبندی به صورت تصادفی در فضای ویژگی قرار می‌گیرد. با محاسبه‌ی فاصله‌ی هر ویژگی نسبت به هر خوشبندی سپس نسبت دادن نزدیکترین مرکز خوشبندی به آن ویژگی، مرکز خوشبندی‌ها به روز رسانی می‌شوند. این روند آنقدر تکرار می‌شود تا در نهایت مرکز خوشبندی‌ها ثابت شوند و دیگر تغییری نداشته باشند [۷۶]. شکل ۳-۵ در سمت چپ یک تصویر ام‌آرآی را نشان می‌دهد که در سمت راست با روش K-means قطعه‌بندی شده است. نکته‌ی قابل توجه در این قطعه‌بندی، طبقه‌بندی به صورت سخت و قاطع است که هر پیکسل تنها متعلق به یک کلاس خاص است.

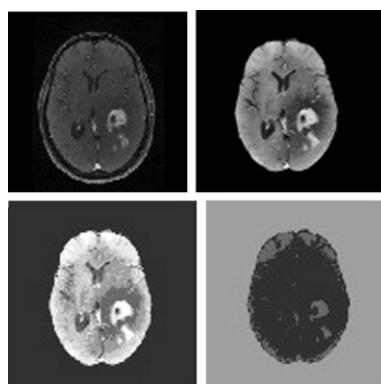


شکل ۳-۵: قطعه‌بندی تصویر ام‌آرآی مغزی با روش K-means [۱۸]

²⁶Decision Tree

Fuzzy C-means

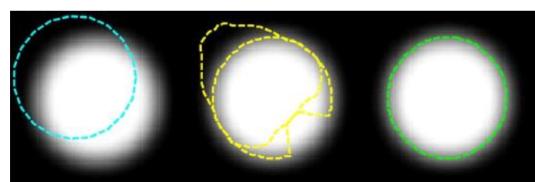
این تکنیک بر اساس تئوری فازی بنا شده است که بر عکس K-means که به صورت قطعه‌بندی سخت عمل می‌کند، در اینجا شاهد یک قطعه‌بندی نرم هستیم که امکان محاسبه احتمال تعلق به هر کلاس را بوجود می‌آورد [۷۷]. شکل ۳-۶ در سمت چپ-بالا یک تصویر ام‌آرآی مغزی را نشان می‌دهد که توسط روش Fuzzy C-means نقشه‌ی احتمال تعلق هر پیکسل به هر کلاس نشان داده شده است.



شکل ۳-۶: قطعه‌بندی تصویر ام‌آرآی مغزی با روش Fuzzy C-means [۱۹]

۳-۵-۳ مدل‌های شکل‌پذیر

یک مدل شکل‌پذیر می‌تواند یک منحنی و یا یک رویه باشد که امکان تغییر آن در طول زمان وجود دارد. تغییرات این نوع مدل‌ها بر اساس تعریفتابع هزینه‌ی خود مدل و اطلاعات درون تصویر صورت می‌گیرد. این نیروهای داخلی و خارجی در نهایت خم شکل‌پذیر را به سمت لبه‌های تصویر و یا یک ویژگی مطلوب در تصویر سوق می‌دهند. این نوع مدل‌ها در دسته‌ی نیمه-خودکار قرار می‌گیرند و لازم است، یک خم اولیه برای شروع الگوریتم بر روی تصویر تعریف کند. شکل ۳-۷ تکامل یک خم اولیه طی چندین تکرار برای قطعه‌بندی را نشان می‌دهد.



شکل ۳-۷: قطعه‌بندی با استفاده از مدل‌های شکل‌پذیر [۲۰]

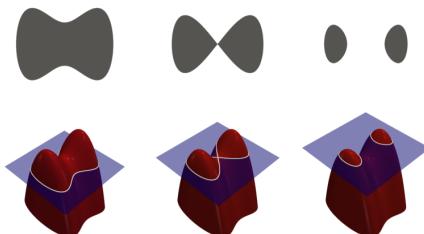
مدل‌های شکل‌پذیر را می‌توان به دو دسته‌ی عمدی مدل‌های شکل‌پذیر پارامتری (صریح)^{۲۷} و مدل‌های شکل‌پذیر هندسی (ضمی)^{۲۸} تقسیم نمود.

مدل‌های شکل‌پذیر پارامتری

به این نوع مدل‌ها، کانتورهای فعال یا مدل‌های snake نیز گفته می‌شود. ایده‌ی اصلی در طراحی این مدل‌ها یافتن مرزها با تطابق مدل شکل‌پذیر و منحنی‌های درون تصویر با استفاده از نیروهای داخلی و خارجی است که با بهینه کردن یک تابع هزینه بدست می‌آیند. نیروهای داخلی شامل قابلیت ارتجاع^{۲۹} و سختی^{۳۰} مدل است و نیروهای خارجی باعث کشش کانتور به سمت لبه‌های تصویر می‌شود [۷۸].

مدل‌های شکل‌پذیر هندسی

مدل‌های شکل‌پذیر هندسی و یا مجموعه-تراز^{۳۱} از تئوری تکامل منحنی و مجموعه-تراز به جای پارامتری کردن منحنی استفاده می‌کنند. عملکرد این مدل‌ها بر اساس یافتن مجموعه-تراز صفر در یک رویه با بعد بالاتر است که قطعه‌بندی نهایی را توصیف می‌کند [۷۹]. شکل ۳-۸ روند تکامل منحنی و استخراج مجموعه-تراز صفر برای قطعه‌بندی یک ساختار دو بعدی را نشان می‌دهد.



شکل ۳-۸: روش تکامل منحنی و استخراج مجموعه-تراز صفر برای قطعه‌بندی [۲۱]

²⁷Parametric (explicit) deformable model

²⁸Geometric (implicit) deformable model

²⁹Elasticity

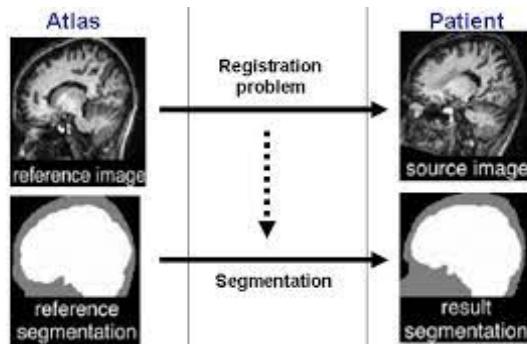
³⁰Rigidity

³¹Level Set

۳-۳-۶ مدل‌های مبتنی بر اطلس

روش‌های مبتنی بر اطلس زمانیکه یک اطلس و یا قالب استاندارد در دسترس باشد یکی از قدرتمندترین ابزارها برای قطعه‌بندی تصاویر پزشکی هستند. اطلس با گردآوری اطلاعات مربوط به آناتومی که نیاز به قطعه‌بندی دارد، تولید می‌شود و سپس به عنوان مرجع برای قطعه‌بندی استفاده می‌گردد.

به اینگونه روش‌ها به عنوان یک مساله‌ای انطباق تصویر نگاه می‌شود که در ابتدا یک نگاشت یک به یک که تصویر اطلس از پیش قطعه‌بندی شده را بر روی تصویر هدف منطبق می‌کند، پیدا می‌شود و سپس قطعه‌بندی اطلس با همین نگاشت تبدیل داده می‌شود تا قطعه‌بندی هدف حاصل شود. این نوع روش‌ها برای قطعه‌بندی ساختارهایی بهتر نتیجه می‌دهند که تغییر شکل ندارند مانند استخوان‌ها [۸۰]. شکل ۹-۳ روند کلی قطعه‌بندی بر اساس اطلس را نشان می‌دهد.



شکل ۳-۹: روش مبتنی بر اطلس برای قطعه‌بندی تصویر مغزی [۲۲]

۳-۳-۷ سایر روش‌های قطعه‌بندی

علاوه بر آنچه اشاره گردید، روش‌های دیگری نیز برای قطعه‌بندی مانند روش‌های مبتنی بر گراف، روش‌ها هیبرید (استفاده توام از چند روش قطعه‌بندی)، روش‌های مبنی بر معادلات دیفرانسیل جزئی و ... نیز ارایه شده است [۸۱] که به دلیل کمتر استفاده شدن، از توضیح آن‌ها خودداری می‌شود.

۴-۳ قطعه‌بندی با استفاده از شبکه‌های عصبی عمیق

از یک الگوریتم قطعه‌بندی تصاویر پزشکی مناسب انتظار می‌رود موارد زیر را ارضاء کند [۶۰].

- مقاوم در برابر نویز
- دقت بالا
- کمترین تعامل با کاربر
- قابلیت اطمینان بالا
- سرعت بالا
- قابلیت تکرار

تمامی الگوریتم‌های کلاسیکی که پیش‌تر بررسی شد، سعی در برآورده کردن موارد بالا داشتند که تا حد قابل قبولی پیش رفته‌اند و توانسته‌اند اعتماد پزشکان و متخصصان را به عنوان یک الگوریتم تشخیص کمکی، جلب کنند. با ظهور روش‌های یادگیری عمیق و شکست دادن انسان [۸۲] در بسیاری از کارها مانند طبقه‌بندی تصاویر طبیعی [۵۴]، بازی‌های فکری و ... توجه بسیاری برای استفاده از این روش‌ها در قطعه‌بندی تصاویر، جلب شد. با استفاده از متدهای یادگیری عمیق در قطعه‌بندی تصاویر پزشکی، دقت و قابلیت اعتماد بالاتری نسبت به روش‌های کلاسیک بدست آمد و امروزه تقریباً در اکثر نرم‌افزارهای قطعه‌بندی، این روش‌ها جایگزین روش‌های کلاسیک شده‌اند. در ادامه به بررسی مقالات و روش‌ها ارایه شده در این حوزه پرداخته می‌شود.

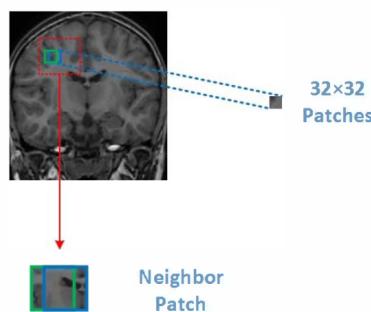
در فصل دوم جزیيات و ابزارهای مربوط به روش‌های یادگیری ماشین و یادگیری عمیق بررسی گردید. در این فصل استفاده ای این ابزارها در قطعه‌بندی تصاویر پزشکی و ایده‌هایی که باعث بهبود این روش‌ها شده‌اند بررسی می‌شوند.

۱-۴-۱ طبقه‌بندی پیکسل‌ها برای قطعه‌بندی

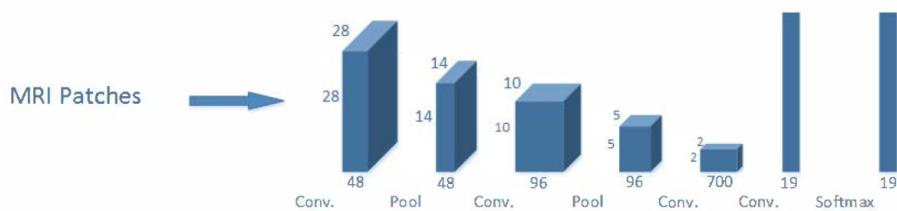
با ظهور و معرفی شبکه‌های عصبی عمیق در سال ۲۰۱۲ برای طبقه‌بندی تصاویر طبیعی [۵۳] و یک جهش بزرگ در دقت طبقه‌بندی، استفاده از این روش‌ها در قطعه‌بندی رونق گرفت. اولین و ساده‌ترین ایده برای ارتباط حوزه‌ی طبقه‌بندی و قطعه‌بندی، همانطور که پیش‌تر اشاره شد، استفاده از شبکه‌های عمیق برای طبقه‌بندی در سطح پیکسل‌هاست. یعنی با استخراج خودکار ویژگی از نواحی اطراف پیکسل با شبکه‌های عصبی عمیق و طبقه‌بندی پیکسل مدل نظر به طبقه‌ی مطلوب است.

در [۲۳] برای قطعه‌بندی تصاویر ام‌آرآی از شبکه‌های کانولوشنی عمیق استفاده شده است که شامل چند لایه پشت سرهم لایه‌های کانولوشنی و ادغام بیشینه‌گیر است که ویژگی‌ها تصویر را استخراج می‌کند. رویکرد این مطالعه استفاده از روش‌های طبقه‌بندی برای قطعه‌بندی است به این صورت که یک پنجره

با ابعاد 32^3 بر روی تصویر لغزانده می‌شود و یک پچ^{۳۲} از تصویر استخراج می‌شود، با استفاده از این پچ پیکسل وسط برچسبگذاری می‌شود که برای این‌کار ابتدا توسط شبکه‌ی کانولوشنی ویژگی‌های پچ استخراج می‌شود و سپس این ویژگی‌ها به یک شبکه‌ی تمام متصل برای طبقه‌بندی هر پیکسل داده می‌شود. شکل ۳-۱۰ چگونگی استخراج پچ‌ها از تصویر ام‌آرآی مغزی را نشان می‌دهد و سپس در شکل ۳-۱۱ این پچ‌ها به یک شبکه‌ی کانولوشنی برای استخراج ویژگی و در نهایت طبقه‌بندی توسط یک شبکه‌ی تمام متصل برای قطعه‌بندی ساختارهای مغزی استفاده می‌شود.



شکل ۳-۱۰: لغزاندن یک پنجره روی تصویر ام‌آرآی مغزی و استخراج پچ [۲۳]

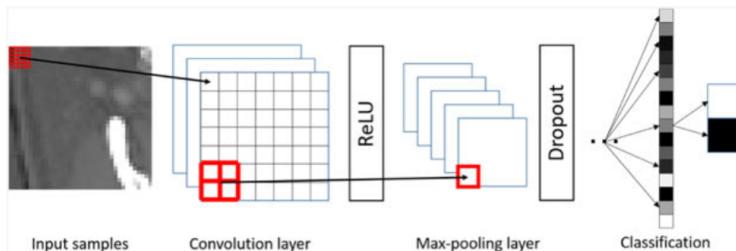


شکل ۳-۱۱: شبکه‌ی کانولوشنی برای طبقه‌بندی پیکسل‌ها وسطی پچ‌ها [۲۳]

ابراجیموف و زینک [۲۴]، ۱۳ ساختار در ریسک را در ۵۰ تصویر سی‌تی اسکن سروگردان قطعه‌بندی کرده‌اند. رویکرد آن‌ها استفاده از شبکه‌های کانولوشنی عمیق است که شدت‌های یکسان در بافت‌های همسان را استخراج کند و بتواند در تصاویر دیده نشده، این ساختارهای در ریسک را قطعه‌بندی کند. برای آموزش شبکه، پچ‌های مثبت که شامل شدت‌های مثبت و اطراف واکسل ساختار در ریسک هستند و همچنین پچ‌های منفی که شامل واکسل‌های ساختارهای کناری هستند استخراج می‌شود. سپس این پچ‌ها به شبکه‌ی کانولوشنی داده می‌شوند که با ترکیب ویژگی‌های سطح پایین مانند لبه‌ها و نقاط کلیدی با ویژگی‌های سطح بالاتر، به طور موثری ساختارهای در ریسک را قطعه‌بندی نموده است. در شکل زیر

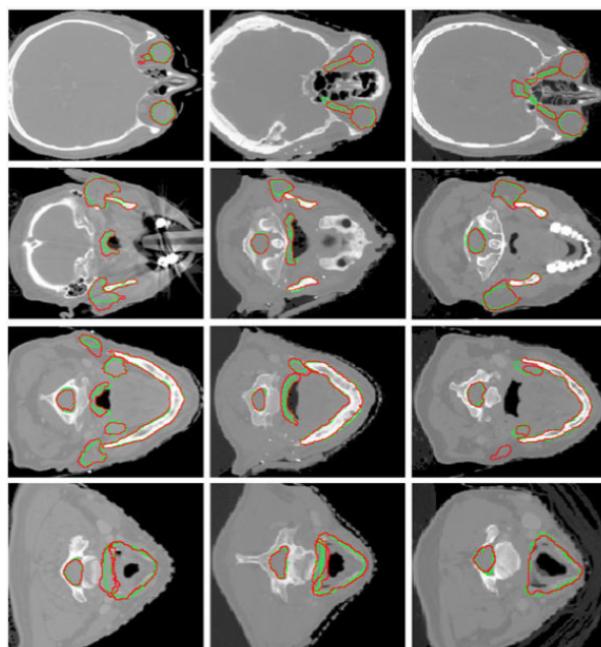
³²Patch

معماری شبکه‌ی استفاده شده که شامل لایه‌های کانولوشنی، ادغام بیشینه‌گیر و dropout است، نشان داده شده است. در نهایت ویژگی‌های استخراج شده به یک شبکه‌ی تمام متصل برای طبقه‌بندی داده می‌شود.



شکل ۳-۱۲: معماری شبکه‌ی کانولوشنی برای طبقه‌بندی بر اساس پنج ساختارهای در ریسک سر و گردن [۲۴]

شکل ۳-۱۳ خروجی شبکه (رنگ قرمز) و قطعه‌بندی مطلوب (رنگ سبز) را برای ساختارهای در ریسک سر و گردن مانند چشم‌ها، فک و ... نشان می‌دهد.



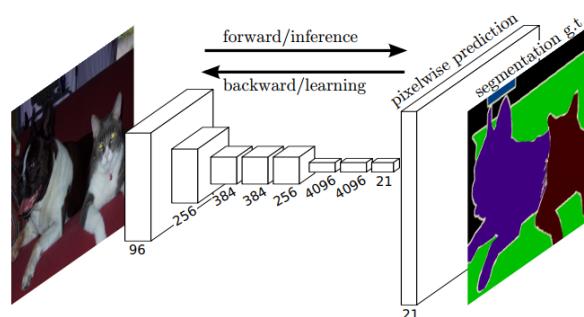
شکل ۳-۱۳: نتایج بدست آمده در قطعه‌بندی ساختارهای در ریسک سر و گردن (قرمز: خروجی شبکه، سبز: قطعه‌بندی مطلوب) [۲۴]

روش‌های قطعه‌بندی بر اساس پنج، علی‌رغم استفاده‌ی زیاد در مواردی که ابعاد تصاویر برای آموزش

بسیار بزرگ باشد و نیازمند حافظه‌ی زیاد باشد و نیز تعداد دادگان برای آموزش کم باشد (با این روش تعداد دادگان چندین برابر می‌شود) استفاده دارد [۸۳]. اما معایبی از جمله، از بین رفتن اطلاعات گستردگی و مربوط به بافت^{۲۳} تصویر، ایجاد قطعه‌هایی در جاهایی که مرتبط با بافت هدف برای قطعه‌بندی نیست و همچنین تصمیم‌گیری تنها بر اساس پیکسل‌های کناری است که از استخراج ویژگی‌های سطح بالاتر پیشگیری می‌شود. یکی دیگر از معایب آن به علت وجود شبکه‌ی تمام متصل، داشتن ورودی با ابعاد خاص است زیرا با تغییر ورودی ابعاد ویژگی استخراج شده تغییر می‌کند و در نتیجه باید ابعاد شبکه‌ی تمام متصل تغییر کند (مانند شبکه‌های طبقه‌بند). بنابراین محققان سعی در ارائه مدل‌هایی برای حل این مشکلات داشتند که در ادامه به بررسی این تکنیک‌ها پرداخته می‌شود [۸۴].

۲-۴-۳ شبکه‌های تمام کانولوشنی برای قطعه‌بندی

ایده‌ی اصلی برای ساخت شبکه‌های تمام کانولوشنی، همانطور که در بالا اشاره شد، قطعه‌بندی بر اساس اطلاعات کل تصویر است. بنابراین با دادن یک تصویر به عنوان ورودی، با همان ابعاد، نقشه‌ی قطعه‌بندی خروجی ایجاد می‌شود [۲۵]. شکل ۱۴-۳ معماری اینگونه شبکه‌ها را نشان می‌دهد که در ابتدا ویژگی‌هایی از کل تصویر توسط لایه‌های کانولوشنی استخراج می‌شود و سپس این ویژگی‌ها با افزایش اندازه و اعمال کانولوشن بر روی ویژگی‌های تغییر اندازه یافته، نقشه‌ی قطعه‌بندی نهایی ایجاد می‌شود.



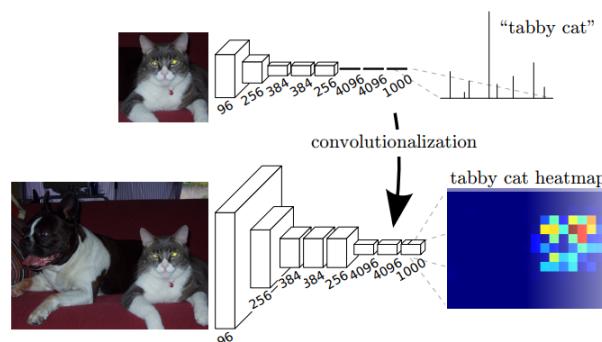
شکل ۳-۱۴: قطعه‌بندی تصاویر با استفاده از رویکرد شبکه‌های تمام کانولوشنی [۲۵]

لانگ و همکاران [۲۵]^{۳۴} دریافتند که امکان استخراج نقشه‌ی حرارتی^{۳۴} وجود یک شی در یک تصویر، با استفاده از بخش استخراج ویژگی شبکه‌های طبقه‌بند، وجود دارد. در شکل ۱۵-۳ وجود یک گربه

³³Texture

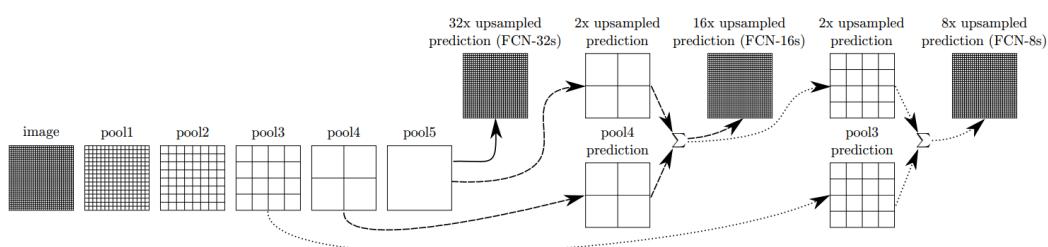
³⁴Heatmap

در تصویر توسط شبکه‌های طبقه‌بند پیش‌بینی شده است و با استخراج نقشه‌ی گرمایی از ویژگی‌های استخراج شده، مکان تقریبی آن قابل دستیابی است. بنابراین با این رویکرد و استفاده از معماری شکل بالا، این گروه توانستند قطعه‌بندی را با استفاده از شبکه‌های تمام کانولوشنی صورت دهند.



شکل ۳-۱۵: استخراج نقشه‌ی گرمایی وجود یک شی در تصویر با استفاده از ویژگی‌های استخراج شده توسط شبکه‌ی طبقه‌بند [۲۵]

مشکل موجود در چنین معماری‌هایی بازگرداندن ابعاد ویژگی‌های استخراج شده به اندازه‌ی تصویر اصلی است که برای آن روش‌های زیادی پیشنهاد شده است که در ادامه بررسی خواهند شد. در مقاله‌ی لانگ و همکاران برای حل این مشکل از upsample کردن ویژگی‌های استخراج شده بعد از لایه‌های ادغام و ترکیب ویژگی‌های با ابعاد مختلف با یکدیگر برای رسیدن به ابعاد تصویر اولیه است. روند این کار در شکل ۱۶-۳ نشان داده شده است.

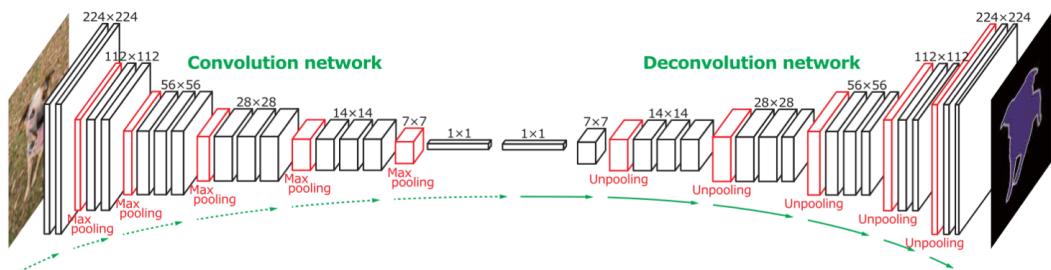


شکل ۳-۱۶: تغییر ابعاد ویژگی‌های استخراج شده بعد از لایه‌های ادغام به ابعاد تصویر اصلی [۲۵]

در مرجع [۲۶] برای حل مشکل upsampling به جای روش‌های معمول درون‌یابی، فیلترهایی با وزن‌های قابل یادگیری ارایه دادند که باعث بهبود دقت قطعه‌بندی گردید. شبکه‌ی پیشنهادی این گروه طبق شکل ۱۷-۲ از دو بخش کانولوشن و دیکانولوشن^{۳۵} تشکیل شده است که در بخش کانولوشن

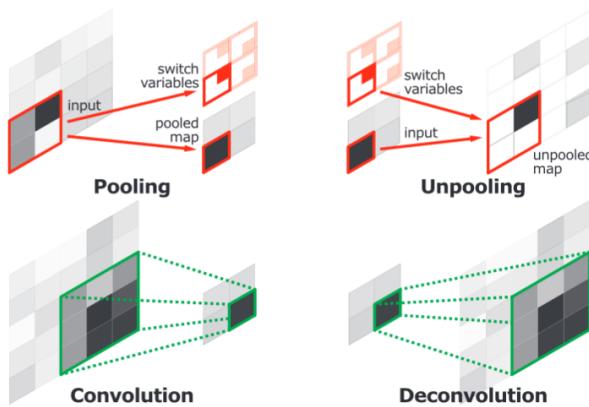
^{۳۵}Deconvolution

ویژگی‌های مطلوب توسط لایه‌های کانولوشن و ادغام استخراج می‌شود و سپس بخش دیکانولوشن این ویژگی‌ها را به ابعاد تصویر اصلی تغییر می‌دهد تا نقشه‌ی قطعه‌بندی حاصل شود.



شکل ۳-۱۷: معماری شبکه‌ی شامل کانولوشن و دیکانولوشن برای بهبود دقت قطعه‌بندی [۲۶]

برای تفهیم ایده‌ی لایه‌های دیکانولوشن ارایه شده در این شبکه، شکل ۳-۱۸ را در نظر بگیرید. در سمت بالا-چپ، عملکرد لایه‌ی ادغام بیشینه‌گیر نشان داده شده است که مقدار بزرگترین درایه و موقعیت آن را ذخیره می‌کند. در سمت بالا-راست، لایه‌ی بر عکس لایه‌ی ادغام نشان داده شده است که با دانستن موقعیت بزرگترین درایه در هنگام upsampling آن را در موقعیت پیشین خود قرار می‌دهد. در سمت پایین-چپ، لایه‌ی کانولوشنی نشان داده شده است که در فصل دوم عملکرد آن بررسی گردید و در نهایت در سمت پایین-راست، عملکرد لایه‌ی دیکانولوشن توضیح داده شده است که یک فیلتر با وزن‌های قابل یادگیری در یک مقدار ضرب می‌شود و به ابعاد بزرگتری تغییر اندازه می‌دهد، با لغزاندن این فیلتر و جمع کردن مقادیر همپوشان با یکدیگر، ابعاد ماتریس ویژگی بزرگتر می‌شود.

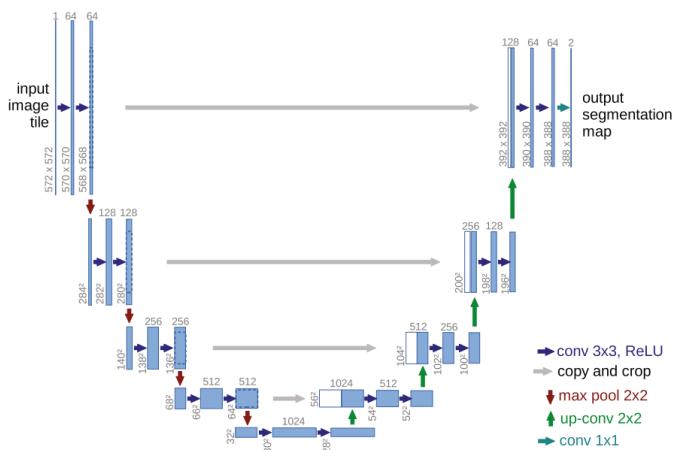


شکل ۳-۱۸: نمایش تفاوت عملکرد کانولوشن و دیکانولوشن [۲۶]

رونبرگر و همکاران [۲۷] برای قطعه‌بندی تصاویر پزشکی یک معماری جدید و پر استفاده به نام UNet ارایه دادند. شکل ۳-۱۹ این معماری را نشان می‌دهد که از دو مسیر انقباضی و انبساطی

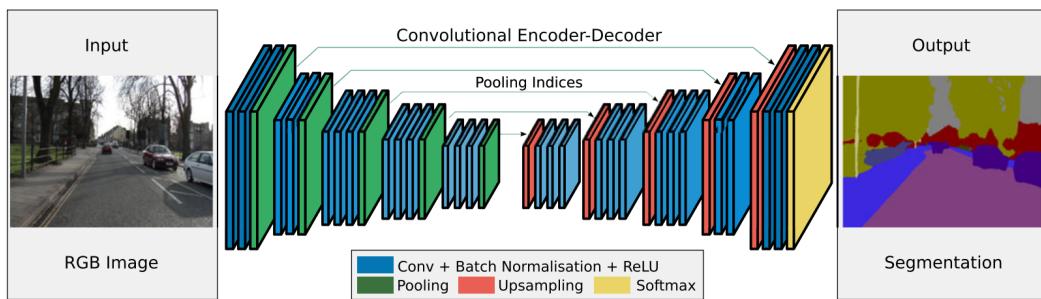
تشکیل شده است. در مسیر انقباضی روند معمول در شبکه های کانولوشنی پیاده سازی شده است که شامل، لایه های کانولوشنی 3×3 که بعد از آنها از تابع فعالیت ReLU استفاده شده است و در نهایت یک لایه ای ادغام بیشینه گیر 2×2 برای کاهش ابعاد ماتریس ویژگی به نصف ابعاد اولیه قرار داده شده است. در هر قدم از کاهش ابعاد تعداد کanal های ویژگی ها دو برابر شده است.

در مسیر انبساطی، بعد از هر بار upsampling یک کانولوشن 2×2 قرار داده شده است که در تصویر با نام 'Up-convolution' نشان داده شده و در طی این عملیات تعداد کانال‌های ویژگی نصف می‌شود. نکته‌ی مهم در این معماری قرار دادن مسیرهای اتصالی از مسیر انقباضی به مسیر انبساطی است. این کار باعث حل شدن مشکل محو شدن گرادیان طی بهروز رسانی وزن‌ها می‌شود و همچنین اطلاعات و جزئیاتی که به دلیل استفاده از لایه‌های ادغام از بین می‌روند دوباره در این مسیر احیا می‌شود و باعث بهبود دقت قطعه‌بندی می‌گردد. در نهایت بر روی کانال‌های ویژگی بدست آمده یک کانولوشن 1×1 برای تغییر عمق کانال‌ها زده می‌شود و نقشه‌های قطعه بندی مطلوب با کمینه کردن تابع هزینه ایجاد می‌شوند.



شکل ۳-۱۹: معماری شبکه UNet

برای بهبود عملکرد upsampling در معماری UNet همانند آنچه در شکل ۱۸-۳ بالا-راست، مشاهده شد، معماری جدیدی به نام SegNet معرفی گردید [۲۸] که مختصات بزرگترین درایه‌ها را در لایه‌های ادغام مسیر انقباضی ذخیره می‌کند و سپس برای upsample کردن، مقادیر را دقیقاً در همان درایه‌های متناظر قرار می‌دهد و بر عکس روش‌های دیگر، مابقی درایه‌ها برابر با صفر قرار داده می‌شوند. شکل ۲۰-۳ معماری این شبکه را نشان می‌دهد که علاوه بر مسیرهای اتصالی، درایه‌های لایه‌های ادغام نیز عبور داده می‌شود.



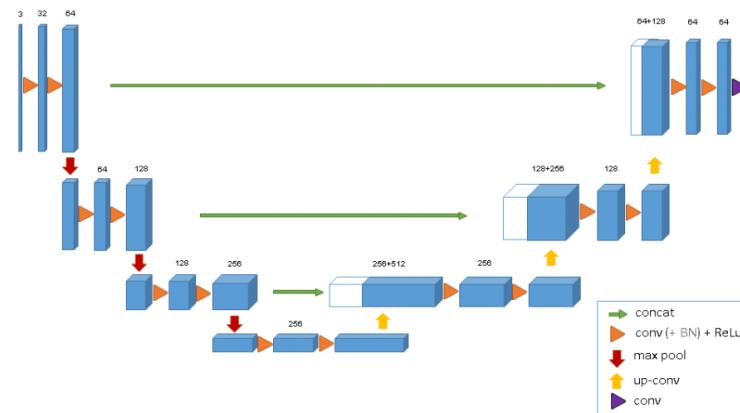
[۲۰] شکل ۳-۲۰: معماری شبکه‌ی SegNet

۳-۴-۳ شبکه‌های تمام کانولوشنی سهبعدی برای قطعه‌بندی

با گسترش استفاده‌ی شبکه‌های کانولوشنی دوبعدی برای پردازش تصاویر دوبعدی و حاصل شدن نتایج خیره کننده، استفاده‌ی این نوع شبکه‌ها برای پردازش ویدیو نیز مورد توجه قرار گرفت. در دنباله‌های ویدیویی، علاوه بر ابعاد طول و عرض تصویر بعد زمان نیز اضافه می‌شود و یک رشته‌ی سهبعدی برای پردازش پدید می‌آید. بنابراین با توسعه‌ی شبکه‌های کانولوشنی دوبعدی به سهبعدی امکان پردازش ویدیو نیز مهیا شد [۸۵، ۸۶].

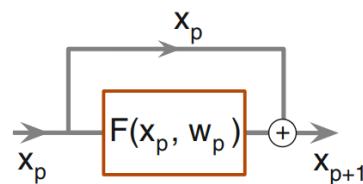
با این توسعه، پردازش تصاویر سهبعدی پژوهشی مانند ام آر آی و سی‌تی اسکن، نیز مورد توجه قرار گرفت که بحث قطعه‌بندی این نوع تصاویر را نیز شامل می‌شود. بنابراین با تعمیم شبکه‌های دوبعدی به سهبعدی تمام کانولوشنی، می‌توان مساله‌ی قطعه‌بندی را با تصویر سهبعدی انجام داد. در ادامه به بررسی شبکه‌های سهبعدی ارایه شده برای قطعه‌بندی تصاویر سه بعدی پرداخته خواهد شد.

شبکه‌ی UNet سهبعدی یا 3D UNet مشابه معماری UNet دوبعدی از مسیرهای انقباضی (بخش رمزگذار 'Encoder' و انساطی (بخش رمزگشای 'Decoder') تشکل شده‌است با این تفاوت که تمامی محاسبات کانولوشن و ادغام در فضای سهبعدی انجام می‌شود [۲۹]. معماری این شبکه در شکل زیر نشان داده شده‌است.



شکل ۳-۲۱: معماری شبکه‌ی UNet سه‌بعدی [۲۹]

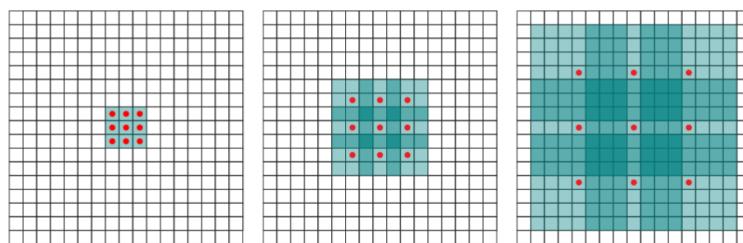
بلوک‌های residual توسط هی و همکاران [۸۷] برای آموزش موثر شبکه، ارایه شدند. ایده‌ی اصلی این بلوک‌ها ایجاد یک تطابق یک به یک برای میانبر کردن لایه‌های شامل پارامتر و تجمیع ورودی و خروجی در انتهای این بلوک است. این بلوک‌ها انتشار اطلاعات را هموارتر می‌کنند و باعث آموزش بهتر شبکه می‌شوند. در شکل ۲۴-۳ یک بلوک دارای اتصالات residual نشان داده شده است.



شکل ۳-۲۲: یک بلوک با اتصال residual [۳۰]

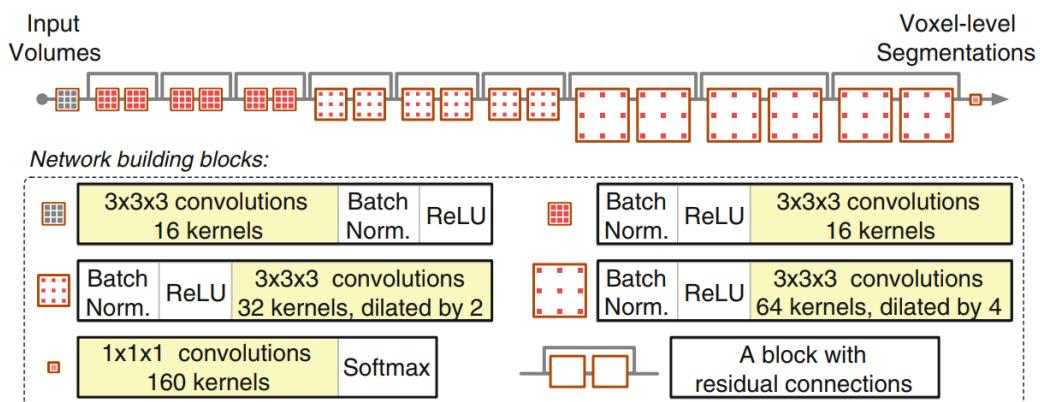
کانولوشن‌های منبسط شده^{۳۶} همانند کانولوشن‌های عادی عمل می‌کنند با این تفاوت که بین المان‌های سازنده‌ی آن‌ها بر حسب ضریب انبساط، صفر قرار می‌گیرد. به این ترتیب بدون افزایش تعداد پارامترهای فیلتر، میدان دید بزرگتر و استخراج ویژگی از ناحیه‌ی بزرگتری صورت می‌پذیرد که باعث استخراج ویژگی‌های عمیق‌تری می‌شود. شکل ۲۳-۳، میدان دید کانولوشن‌های منبسط شده با ضریب انبساط به ترتیب از چپ به راست ۱، ۲ و ۴ را نشان می‌دهد.

^{۳۶}Dilated Convolution



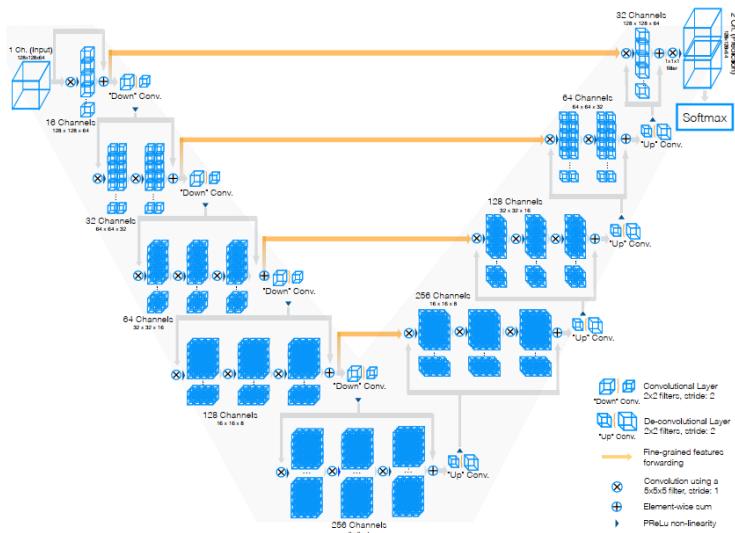
شکل ۳-۲۳: کانولوشن‌های منبسط شده [۳۱]

با استفاده از دو ایده معرفی شده در بالا (بلوک‌های residual و کانولوشن‌های منبسط شده) لی و همکاران یک شبکه سه‌بعدی برای قطعه‌بندی تصاویر ام‌آرآی مغزی ارایه کردند. این شبکه شامل ۲۰ لایه کانولوشنی است. در ۷ لایه اول، از کانولوشن‌های $3 \times 3 \times 3$ استفاده شده است که وظیفه آن‌ها استخراج ویژگی‌های سطح پایین مانند لبه‌ها و گوش‌های است. در لایه‌های بعدی ضریب انبساط برای کانولوشن‌های منبسط شده، ۲ و ۴ است. این لایه‌ها وظیفه استخراج ویژگی‌های سطح بالاتر را دارند. بلوک‌های residual به هر جفت لایه کانولوشنی برای بهبود عملکرد اضافه گردیده است. همچنین از روش Batch-normalization برای کمک به فرآیند تعمیم پذیری استفاده شده است. شکل زیر معماری این شبکه را نشان می‌دهد.



شکل ۳-۲۴: معماری شبکه سه‌بعدی شامل اتصالات residual و کانولوشن انبساطی [۳۰]

شبکه VNet نیز مانند UNet سه‌بعدی از کانولوشن و محاسبات سه‌بعدی استفاده می‌کند. تفاوت این دو معماری مطابق شکل ۲۵-۳، اضافه شدن بلوک‌های residual به معماری UNet است. این کار باعث جلوگیری از محو شدن گرادیان در لایه‌ها نخستین و همچنین عمیق‌تر کردن شبکه برای استخراج ویژگی‌های سطح بالا می‌شود [۳۲].



شکل ۳-۲۵: معماری شبکه‌ی VNet [۳۲]

علی‌رغم وجود مزیت‌های فراوان استفاده از شبکه‌های سه‌بعدی در استخراج ویژگی‌های فضایی بهتر، این شبکه‌ها با مشکلاتی از قبیل، هزینه‌ی محاسباتی بالا و حافظه‌ی مورد نیاز بالا برای استفاده و آموزش رو برو هستند [۸۸].

۳-۵ قطعه‌بندی تومور و ساختارهای در ریسک با یادگیری عمیق

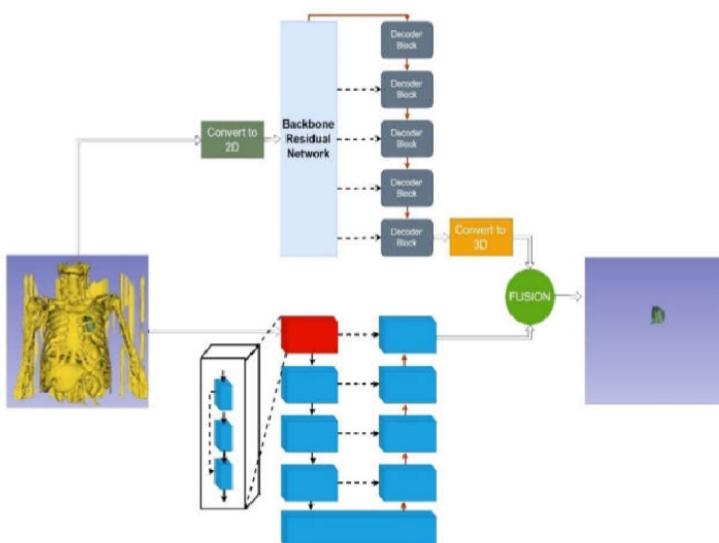
در قسمت قبل، مدل‌ها و روش‌های پایه‌ای قطعه‌بندی با استفاده از شبکه‌های عصبی عمیق بررسی گردید. در این قسمت چند نمونه از مطالعاتی که با استفاده از مدل‌های یادگیری عمیق قطعه‌بندی تومور و ساختارهای در ریسک را انجام داده‌اند، مطالعه خواهد شد. نکته‌ی قابل توجه آن‌که، مبنا و شاکله‌ی اصلی این روش‌های یادگیری عمیق، ابزارهایی هستند که در قسمت قبل و در فصل دوم بررسی گردید و این مطالعات با اضافه‌کردن ایده‌ها و نوآوری‌هایی به این مدل‌های پایه، قطعه‌بندی مورد نظر را بهبود بخشیده‌اند.

به دلیل آن‌که اغلب دادگان پزشکی موجود، تعدا کمی دارند، آموزش شبکه‌های عمیق به دلیل کمبود داده با مشکل مواجه است. بنابراین با روش‌های مانند افزایش تعداد دادگان، این کمبود جبران می‌شود. یکی دیگر از این روش‌ها استفاده از مدل‌های از پیش آموزش دیده شده بر روی یک مجموعه داده‌ی بزرگتر برای استخراج ویژگی‌های سطح بالاتر است. در مطالعه‌ی [۳۳] برای قطعه‌بندی تومور ریوی از

تصاویر سی‌تی اسکن، از این رویکرد استفاده شده است.

روش کار این پژوهش استفاده از یک شبکه‌ی Resnet از پیش آموزش دیده شده بر روی مجموعه داده‌ی ImageNet است [۸۷]. از این شبکه، به عنوان Encoder شبکه‌ی UNet استفاده شده است که ویژگی‌های عمیق و سطح بالا را از تصاویر دو بعدی سی‌تی اسکن (لایه‌های یک تصویر سه بعدی) استخراج می‌کند و سپس با مسیرهای ارتباطی به قسمت Decoder که قابلیت آموزش دارد متصل می‌کند تا تومور را در تصاویر دو بعدی شناسایی و قطعه‌بندی کند.

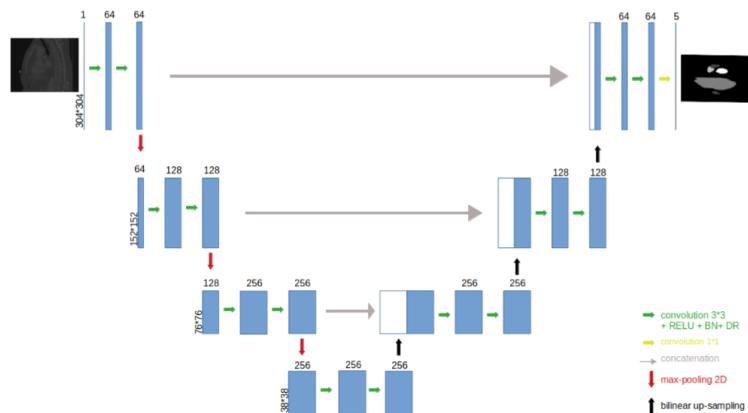
ایده‌ی دیگر این مطالعه برای از دست نرفتن اطلاعات حجمی تصاویر سی‌تی اسکن، استفاده از یک UNet سه بعدی دیگر است که مستقیماً تصویر سه بعدی را می‌گیرد و یک قطعه‌بندی تومور سه بعدی انجام می‌دهد. در نهایت با ترکیب خروجی این دو شبکه با استخراج کمترین مقدارها از خروجی دو شبکه، قطعه‌بندی نهایی صورت می‌گیرد. شکل ۲۶-۳ معماری این شبکه‌ها را نشان می‌دهد.



شکل ۲۶-۳: معماری روش پیشنهادی در [۳۳]

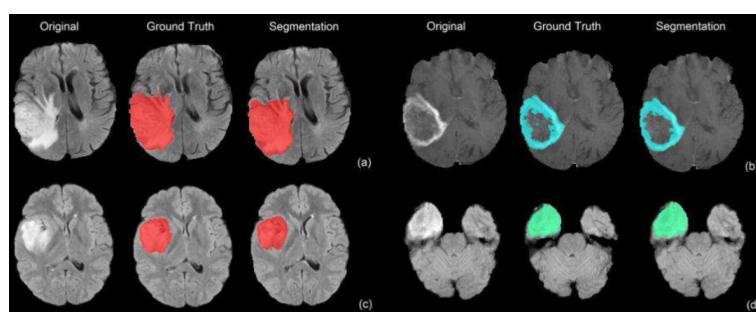
لامبرت و همکاران [۳۴] بر روی قطعه‌بندی چهار ساختار در ریسک ۵۰ تصویر سی‌تی اسکن قفسه‌ی سینه مطالعه کردند. روش پیشنهادی آن‌ها به دلیل کمبود دادگان استفاده از یک UNet ساده شده و همچنین افزایش سه برابری دادگان با اعمال تبدیلات خطی و غیرخطی بر روی تصویر است. UNet ساده شده‌ی آن‌ها برخلاف UNet عادی یک مرحله لایه‌ی ادغام کمتر دارد و در نتیجه تعداد ویژگی‌های استخراج شده و لایه‌های کانولوشنی کمتر از UNet معمول است. در بخش Decoder آن به جای استفاده از دیکانولوشن، از روش درون‌یابی سه خطی استفاده شده است تا تعداد پارامترهای آموزش

کاهش یابد. در نهایت تعداد پارامترهای آموزشی از $7/2 \times 64 \times 64$ میلیون برای UNet عادی به $4/8 \times 64 \times 64$ میلیون برای مدل پیشنهادی کاهش یافته است. برای افزایش قدرت تعمیم‌پذیر شبکه از روش‌های Batch-Dropout و normalization بعد از لایه‌های کانولوشنی استفاده شده است. همچنین ایده‌ی دیگر آن‌ها استفاده ازتابع هزینه Dice برای بهبود دقت قطعه‌بندی است. شکل زیر معماری شبکه‌ی پیشنهادی این مطالعه را نشان می‌دهد.



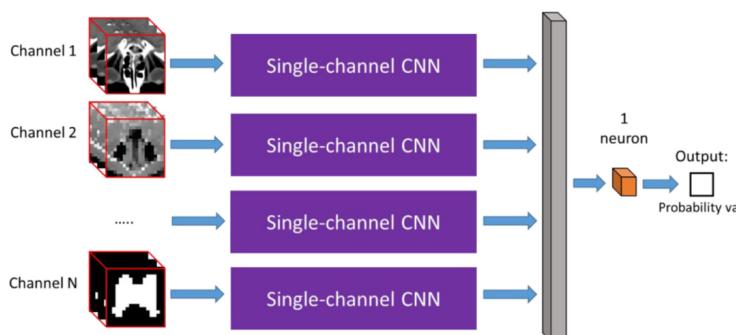
شکل ۳-۲۷: معماری UNet پیشنهادی در [۳۴]

دانگ و همکاران [۳۵] برای قطعه‌بندی تومورهای تصاویر ام‌آرآی مغزی، از شبکه‌ی تمام کانولوشنی UNet استفاده کرده‌اند. برای بهبود عملکرد شبکه‌ی UNet از تکنیک Dropout و جریمه‌ی وزن‌های شبکه استفاده شده است. همچنین از روش‌های افزایش دادگان مانند دوران، انتقال، بزرگنمایی، تغییر شدت روشناکی و ... برای افزایش دقت بهره‌گیری شده است. در نهایت این شبکه با استفاده ازتابع هزینه Dice آموزش دیده است. شکل ۳-۲۹ چهار نمونه تصویر را در مجموعه‌های سه‌تایی برای بررسی عملکرد شبکه‌ی پیشنهادی را نشان می‌دهد. مجموعه‌های سه‌تایی از چپ به راست عبارتند از: تصویر اصلی، قطعه‌بندی مطلوب و خروجی شبکه.



شکل ۳-۲۸: نتایج قطعه‌بندی تومور مغزی در تصاویر ام‌آرآی در مرجع [۳۵]

در این مرجع [۴۶] یکی از ایرادات اساسی قطعه‌بندی ساختارهای در ریسک مورد مطالعه قرار گرفته است. همانطور که پیش‌تر اشاره گردید ساختارهای در ریسک اندازه و خواص مختلفی دارند و برخی از این ساختارها مانند اعصاب بینایی یا ساقه‌ی مغزی، به دلیل آنکه بسیار کوچک‌اند و اختلاف شدت آن‌ها با بافت‌ها اطراف کم است بنابراین یافتن آن‌ها دشوار است و اکثر روش‌های ارایه شده این ساختارهای ریز را با دقت کمی مشخص می‌کنند. بنابراین رویکرد این مطالعه به این صورت است که با استفاده از یک شبکه‌ی کانولوشنی چند کanal که هر کanal پچ‌های با ابعاد مختلف اطراف یک بافت را می‌گیرد، ساختارهای ریز را قطعه‌بندی کند. همانطور که در شکل زیر مشاهده می‌شود این شبکه با ورودی‌های با سایزهای مختلف ساختار مورد نظر را قطعه‌بندی می‌کند. ادعایی که در این مطالعه شده بود این بود که پچ‌های با سایز بزرگ اطلاعات کلی و موقعیت کلی بافت مورد نظر را به شبکه می‌دهد و هرچه پچ‌ها کوچک‌تر شوند این اطلاعات ریزتر و قطعه‌بندی دقیق‌تر انجام می‌شود.



شکل ۳-۲۹: روش پیشنهادی در [۳۵] برای استفاده از پچ‌های سه‌بعدی با اندازه‌ی مختلف

۳-۶ جمع‌بندی

در این فصل به معرفی تعاریف، اهداف و چالش‌های قطعه‌بندی پرداخته شد و انواع دسته‌بندی روش‌های قطعه‌بندی معرفی گردید. معیارهای ارزیابی قطعه‌بندی مطالعه شد و در ادامه روش‌های قطعه‌بندی کلاسیک، قبل از ظهور یادگیری عمیق معرفی اجمالی گردید. سپس دیدیم با پیدا شدن روش‌های یادگیری عمیق در قطعه‌بندی، اکثر روش‌های کلاسیک با این روش‌ها جایگزین شدند. قطعه‌بندی با استفاده از شبکه‌های عمیق عصبی امکان قطعه‌بندی تصاویر دو بعدی و سه‌بعدی، امکان قطعه‌بندی نرم و سخت، امکان قطعه‌بندی تمام خودکار و یا نیمه‌خودکار را می‌دهد بنابراین نسبت به اکثر روش‌های کلاسیک ارجحیت دارد.

در نهایت، معماری و روش‌های پیشنهاد شده برای استفاده از یادگیری عمیق در قطعه‌بندی تومور و ساختارهای در ریسک بررسی گردید که با اضافه کردن خلاقیت‌هایی به معماری‌های پایه و استفاده از روش‌های تعمیم‌پذیری دقت شبکه افزایش داده می‌شد. در فصل بعد به معرفی روش‌های پیشنهادی بر مبنای شبکه‌های عصبی عمیق برای قطعه‌بندی ساختارهای در ریسک و تومور و بررسی نتایج استفاده از این روش‌ها پرداخته خواهد شد.

فصل ۴

روش‌های پیشنهادی و نتایج

۱-۴ مقدمه

در این فصل به معرفی روش‌های ارایه شده در این مطالعه، برای قطعه‌بندی ساختارهای در ریسک و تومور با استفاده از شبکه‌های عصبی عمیق پرداخته خواهد شد. مطالعه شامل دو بخش است که در بخش اول به قطعه‌بندی ساختارهای در ریسک، چالش‌های موجود، راهکارهای پیشنهادی و ارزیابی روش‌های پیشنهادی پرداخه می‌شود و در بخش دوم مطالعه همین روند برای قطعه‌بندی تومور طی خواهد شد.

۲-۴ روش‌های پیشنهادی قطعه‌بندی ساختارهای در ریسک

همانطور که در فصل اول اشاره گردید، ضرورت قطعه‌بندی ساختارهای در ریسک، در تهیی نفشهی درمان برای رادیوتراپی است. این نقشهی درمان کمک می‌کند تا به بافت‌های سالم کنار تومور (ساختارهای در ریسک) کمترین آسیب و تومور بیشترین دوز پرتو را دریافت کند. بنابراین لازم است این ساختارها با دقیقت و سرعت بالا قطعه‌بندی شوند و در روند درمان قرار گیرند. این قطعه‌بندی به صورت دستی زمان‌بر، پرهزینه (هزینه‌ی استخدام شخص متخصص)، خسته‌کننده و متغیر بر اساس دیدگاه هر متخصص است.

بنابراین برای تسريع این روند و حل مشکلات ذکر شده، روش‌های خودکار و نیمه- خودکار فراوانی ارایه گردید که در فصل قبل به بررسی هر یک پرداخته شد. ملاحظه گردید که الگوریتم‌های کلاسیک

ارایه شده، در اکثر موارد توانایی قطعه‌بندی چند ساختار به صورت همزمان را ندارند و علاوه بر آن در اکثر موارد نیاز به یک مقدار (کانتور) اولیه دارند که این کاستی‌ها در کنار دقت پایین باعث محبوبیت پایین این روش‌ها شده است.

در دست دیگر، روش‌های قطعه‌بندی مبتنی بر یادگیری عمیق بررسی شد که با آموزش این مدل‌ها بر اساس یک مجموعه داده، علاوه بر سرعت و دقت بالا نسبت به روش‌های قطعه‌بندی کلاسیک و دستی، امکان قطعه‌بندی خودکار چند ساختار نیز وجود دارد. بنابراین شاهد استفاده‌ی روز افزون این روش‌ها و افزایش اعتبار و محبوبیت آن‌ها برای استفاده در نرم‌افزارهای قطعه‌بندی ساختارهای در ریسک در مرکز پرتو درمانی هستیم.

در این بخش به توضیح روش‌های پیشنهادی برای قطعه‌بندی ساختارهای در ریسک بر اساس مدل‌های یادگیری عمیق پرداخته می‌شود. ابتدا مجموعه دادگان و پیش‌پردازش‌های انجام شده، توضیح داده خواهد شد و در ادامه، روش‌های پیشنهادی معرفی می‌شوند و نتایج این روش‌ها بررسی خواهد شد.

۴-۲-۱ معرفی مجموعه دادگان و پیش‌پردازش

در این قسمت، دو مجموعه داده برای قطعه‌بندی ساختارهای در ریسک معرفی می‌شود که در ادامه به بررسی هر یک و پیش‌پردازش‌های انجام شده، پرداخته خواهد شد.

مجموعه داده‌ی SegTHOR

مجموعه داده‌ی SegTHOR شامل قطعه‌بندی چهار ساختار در ریسک قفسه‌ی سینه در تصاویر سی‌تی اسکن است^۱. این مجموعه تصاویر، از ۴۰ بیمار دارای سرطان ریه و مری گرفته شده است و شامل قطعه‌بندی دستی چهار ساختار در ریسک قلب^۲، آئورت^۳، نای^۴ و مری^۵ است که به عنوان استاندارد مطلوب^۶ برای آموزش شبکه‌های عصبی تهیه شده است.

این مجموعه داده برای قطعه‌بندی ساختارهای در ریسک به دلیل تغییر شکل ساختارهای در ریسک

¹Segmentation of THoracic Organs at Risk

²Heart

³Aorta

⁴Trachea

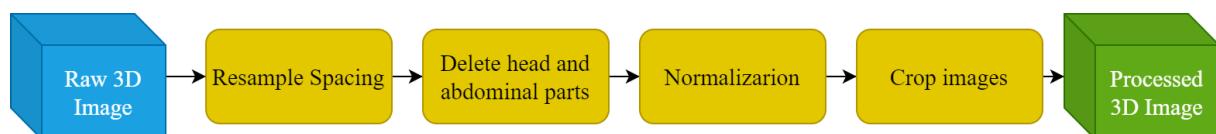
⁵Esophagus

⁶Ground Truth

(مانند مری) در هر شخص و نیز اختلاف شدت پایین در میان بافت‌های نرم در تصاویر سی‌تی اسکن، بسیار پر چالش است [۳۴].

این تصاویر دارای ابعاد اولیه‌ی $(150 \times 284 \times 512)$ و با فاصله‌ی واکسلی^۷ ~ 0.97 در راستاهای x، y و z است که حاکی از متفاوت بودن دستگاه‌های تصویربرداری در این مجموعه دادگان است. بنابراین لازم است برای تمایز هرچه بهتر بافت‌ها و یکسان‌سازی تصاویر دستگاه‌های مختلف در جهت آموزش شبکه‌ی عصبی عمیق، یک پیش‌پردازش مناسب صورت گیرد.

اولین گام در پیش‌پردازش یکسان‌سازی فاصله‌ی واکسلی به $0.97 \times 0.97 \times 0.97$ است که با استفاده از درون‌یابی برای تصویر اصلی و تصویر قطعه‌بندی متناظر این کار انجام گردید و تا حدی تفاوت تصویربرداری در بین دستگاه‌های مختلف یکسان‌سازی شد. در گام دوم، ناحیه‌ی شامل ساختارهایی که باید قطعه‌بندی شوند از تصویر اصلی استخراج گردید و سایر نواحی مانند قسمت‌های شکمی و سر و گردن حذف شد. در گام سوم مقدار HU^۸ در تصاویر اصلی که در بازه‌ی $(-1000 \sim 13000)$ بود مقادیری که در خارج از بازه‌ی $(-800 \sim 400)$ بود، به سقف و کف این بازه تبدیل داده شد و مقادیر این بازه در تصاویر اصلی بین صفر و یک نرمالیزه گردید. در نهایت فضاهای خالی اطراف تصویر به مرکزیت آن بریده شد و ابعاد $512 \times 384 \times 288$ به $512 \times 512 \times 384$ کاهش یافت. نمودار بلوکی این روند در شکل ۴ نشان داده شده است.

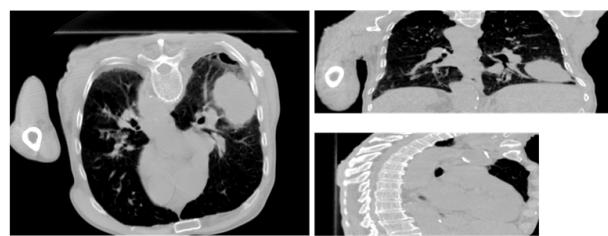


شکل ۴ - ۱: نمودار بلوکی گام‌های پیش‌پردازش بر روی دادگان

شکل زیر یک نمونه تصویر از مجموعه دادگان را بعد از پیش‌پردازش نشان می‌دهد.

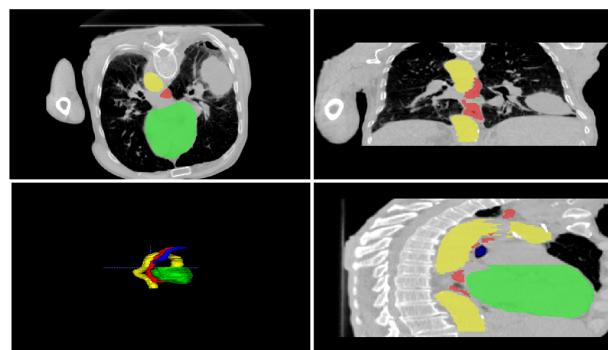
⁷Voxel Spacing

⁸Hounsfield Unit



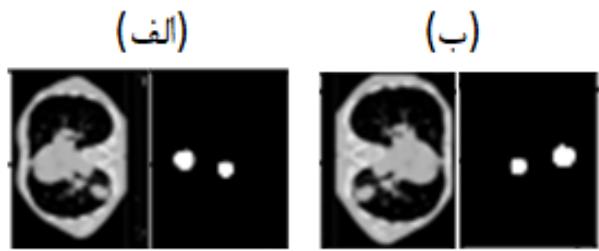
شکل ۴-۲: نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش، در سه نمای اکسیال، کرونال و سجیتال

برای حفظ تناظر یک به یک بین تصاویر اصلی و قطعه‌بندی استاندارد، تغییرات و پردازش‌های اعمال شده در بالا، مطابق با ماهیت برچسب‌ها، اعمال گردید تا دوباره ماسک‌های قطعه‌بندی شده در روند آموزش قابل استفاده گردند. شکل ۴-۳ نمونه‌ای از تصویر اصلی همراه با برچسب‌های چهار ساختار در ریسک را نشان می‌دهد.



شکل ۴-۳: نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش همراه با برچسب ساختار در ریسک، در چهار نمای اکسیال، کرونال، سجیتال و سه‌بعدی

برای آموزش شبکه (شبکه‌ها به صورت دو بعدی آموزش داده می‌شوند) و افزایش قدرت تعمیم‌پذیری آن از روش افزایش مجموعه دادگان استفاده شد که در اینجا به معرفی جزئیات آن پرداخته می‌شود. هر تصویر به صورت دو بعدی و از نمای اکسیال از تصویر سه‌بعدی پردازش شده با برچسب ساختارهای در ریسک متناظر استخراج می‌گردد و با اعمال تبدیلات دوران تصادفی بین زاویه‌ی ($-5, 5$) درجه، بزرگ‌نمایی و کوچک‌نمایی با ضرایب $1/1$ و $0/9$ و قرینه‌ی افقی تعداد دادگان آموزش افزایش می‌یابد. در شکل ۴-۴ (الف) تصویر دو بعدی اصلی با برچسب متناظر نشان داده شده است و در (ب) قرینه‌ی افقی و بزرگ‌نمایی آن با ضرایب $1/1$ به تصویر کشیده شده است.



شکل ۴-۴: نمونه‌ی یک تصویر در نمای اکسیال با برچسب ساختار در ریسک متناظر (الف) تصویر اصلی (ب) تصویر تبدیل یافته برای افزایش دادگان

مجموعه داده‌ی قطعه‌بندی هیپوکامپ

هیپوکامپ^۹ ساختمان عصبی خمیده‌ای است در مغز که در میانه‌ی بطن‌های طرفی مغز قرار دارد. هیپوکامپ در اعمق لوب گیجگاهی جای گرفته‌است و از دو شاخ منحنی‌وار تشکیل شده‌است که از بخش‌های مهم مغز پستانداران است. حافظه افرادی که هیپوکامپ آنان آسیب دیده یا با جراحی برداشته شده، دچار اختلال جدی می‌شود. هیپوکامپ تشییت‌کننده میان حافظه کوتاه‌مدت و بلندمدت است و مغز قدامی را از آزموده‌های گذشته ما آگاه می‌کند. این مجموعه خاطرات گذشته را به شکل کوتاه‌مدت یا درازمدت حفظ می‌کند [۱۰].

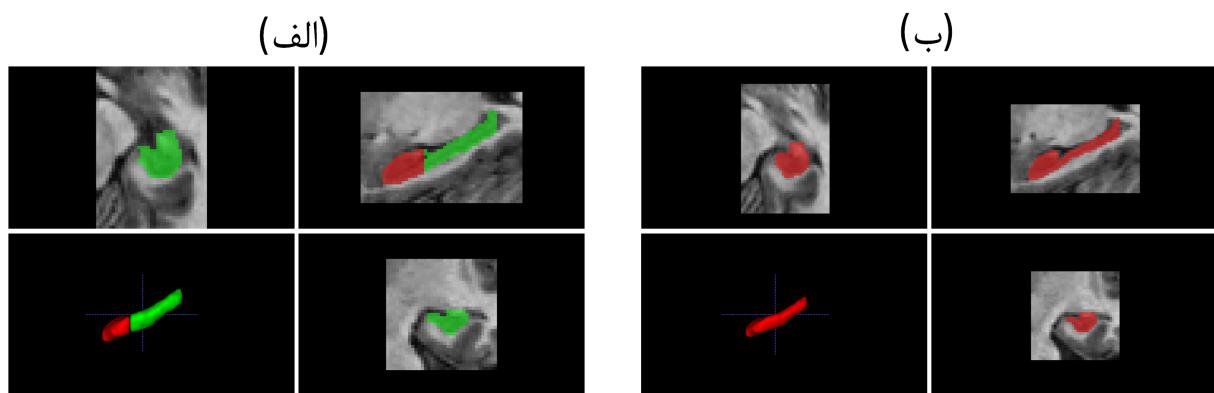
بنابراین این عضو نیز از ساختارهای مهم و حیاتی است و لازم است در طی پرتو درمانی تومورهای مغزی، این ساختار نیز به عنوان ساختار در ریسک، قطعه‌بندی شود و از آسیب به آن جلوگیری شود. مجموعه داده‌ی عمومی قطعه‌بندی تصاویر پزشکی Decathlon [۱۰] شامل ۲۶۰ تصویر سه‌بعدی ام‌آرای T1-weighted است که دو بخش سر و بدن هیپوکامپ را به صورت دستی و به عنوان قطعه‌بندی مطلوب ارایه داده است. ابعاد تصاویر در بازه‌ی (۳۱ ~ ۴۳) × (۴۰ ~ ۵۹) × (۲۴ ~ ۴۷) متر با فاصله‌ی واکسلی یک میلی‌متر هستند.

برای پیش‌پردازش این دادگان ابتدا برای همسان‌سازی اندازه، تمام ابعاد با اضافه کردن صفر^{۱۰} به اندازه ۴۸ × ۶۴ × ۶۴ تغییر داده شد. در گام بعدی برای نرمالیزه کردن تصاویر از هیستوگرام تصویر کمک گرفته شد و با تقسیم مقادیر به شدت ۹۵ درصد هیستوگرام تجمعی این نرمال‌سازی صورت گرفت. دلیل این کار و استفاده نکردن از مقدار بیشینه مقاوم کردن نرمال‌سازی نسبت به نویزهای با

⁹Hippocampus

¹⁰Zero Padding

مقدار بالا در تصاویر ام‌آرآی است. در گام آخر پیش‌پردازش برچسب‌های دو قسمت هیپوکامپ، برای ایجاد یک برچسب در قطعه‌بندی با یکدیگر ترکیب شدند (در ادامه خواهیم دید از این برچسب‌ها برای یک تابع هزینه‌ی خاص استفاده شده است و لازم است ساختار پیوسته باشد). شکل ۴-۵ تصویر قبل (الف) و بعد از پیش‌پردازش (ب) هیپوکامپ را با قطعه‌بندی متناظر نشان می‌دهد.



شکل ۴-۵: نمونه‌ی یک تصویر ام‌آرآی هیپوکامپ در نماهای مختلف با قطعه‌بندی (الف) تصویر اصلی (ب) تصویر پس از پیش‌پردازش

۴-۲-۴ آموزش قطعه‌بندی با استفاده از متادنگالش دانش

با ظهور روش‌های یادگیری عمیق و افزایش سرعت پردازش‌گرها، امکان تعریف مدل‌های بسیار پیچیده با تعداد پارامترهای قابل یادگیری بسیار زیاد فراهم گردید. بنابراین، این مدل‌های پیچیده، توانایی استخراج ویژگی‌های سطح بالا و در نتیجه مستعد اخذ دقت بالاتر هستند. اما از طرفی گران بودن دستگاه‌های با قدرت پردازش بالا و نیز محدودیت استفاده‌ی آن‌ها توجهات را به سمت انتقال دانش از مدل‌های پیچیده به سمت مدل‌های ساده‌تر جلب کرد. این کار باعث می‌شود، ویژگی‌هایی که مدل بسیار پیچیده توانایی استخراج آن‌ها را دارد به مدل ساده‌تر منتقل شود و علاوه بر بدست آوردن دقت نزدیک به مدل پیچیده، هزینه‌های محاسباتی نیز کمتر شود [۹۱].

یکی از این روش‌ها چگالش دانش^{۱۱} از مدل پیچیده (آموزگار^{۱۲}) به مدل ساده (دانش‌آموز^{۱۳}) است که در ادامه قصد داریم این روش را در قطعه‌بندی ساختارهای در ریسک استفاده کنیم و مدل‌های ساده‌تر

^{۱۱}Knowledge Distillation

^{۱۲}Teacher

^{۱۳}Student

برای این قطعه‌بندی را به دقت‌های بالاتر بدون تغییر هزینه‌های محاسباتی برسانیم.

چگالش دانش

چگالش دانش از یادگیری انسان اقتباس شده است که یک آموزگار که در یک موضوع مسلط است به دانش آموز آموزش می‌دهد. بنابراین چارچوب چگالش دانش را می‌توان شامل یک یا چند مدل بزرگ از پیش آموزش دیده شده و یک مدل ضعیف تعریف کرد که ایده‌ی اصلی آن آموزش مدل ضعیفتر (دانش‌آموز) با نظارت مدل پیچیده (آموزگار) برای رسیدن به دقت قابل مقایسه با آموزگار است.

سیگنال نظارتی که از مدل آموزگار به دانش‌آموز می‌رسد را دانش می‌نامیم که قبلاً توسط آموزگار یادگرفته شده است و دانش‌آموز سعی در تقلید رفتار آموزگار در یادگیری دانش را دارد. به عنوان مثال در یک مساله‌ی طبقه‌بندی تصاویر، لاجیت‌ها^{۱۴} (خروجی آخرین لایه در شبکه‌های عصبی عمیق) به عنوان حامل‌های دانش از مدل آموزگار به مدل دانش‌آموز استفاده می‌شود که این دانش توسط برچسب‌های صحیح مطلوب تامین نمی‌شود. برای فهم بیشتر این مطلب، فرض کنید یک مساله‌ی طبقه‌بندی بین چهار طبقه‌ی گاو، سگ، گربه و ماشین وجود دارد. در نهایت بعد از آموزش یک مدل بر روی این دادگان دو نوع خروجی نرم و سخت^{۱۵} در دسترس است. خروجی‌های سخت مربوط به برچسب‌های مطلوب است و خروجی‌های نرم مربوط به احتمالات پیش‌بینی شده برای هر طبقه، توسط مدل است. با توجه به شکل ۶-۴ و مقادیری که برای پیش‌بینی یک تصویر به عنوان سگ شده است می‌توان دریافت علاوه بر پیش‌بینی با احتمال بالا برای کلاس صحیح، میزان مشابهت با کلاس‌های دیگر نیز در خروجی‌های نرم وجود دارد. به عنوان مثال به علت شباهت زیاد سگ و گربه مقدار ۰/۱ برای این شباهت بدست آمده است اما در مقایسه با کلاس‌های ماشین و گاو، کلاس گاو نسبت به کلاس ماشین به علت حیوان و چهارپا بودن هر دو (سگ و گاو) احتمال بیشتری نسبت داده شده است. بنابراین در این نوع خروجی‌ها اطلاعات بیشتری نسبت به خروجی‌های سخت وجود دارد [۳۶].

¹⁴Logits

¹⁵Soft and Hard Targets

cow	dog	cat	car	
0	1	0	0	Hard targets
10^{-6}	0.9	0.1	10^{-9}	Soft targets

شکل ۴-۶: خروجی‌های سخت و نرم برای طبقه‌بندی چهار کلاس [۳۶]

برای بدست آوردن احتمال حضور در n امین طبقه، لاجیت‌ها را از یک تابع فعالیت Softmax عبور می‌دهند تا این احتمال با مقدار p_i مشخص شود. معادله ۱-۴ تابع فعالیت Softmax را نشان می‌دهد.

$$p_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (1-4)$$

برای استخراج خروجی‌های نرم، با اضافه کردن یک فاکتور دما^{۱۶} (T) به معادله Softmax می‌توان این خروجی‌ها را استخراج کرد. با افزایش مقدار فاکتور دما میزان نرم شدن خروجی‌ها بیشتر می‌شود و با کاهش آن مقدار خروجی‌ها به خروجی‌های سخت نزدیک‌تر می‌شود. بنابراین با این فاکتور می‌توان میزان اهمیت هر برچسب را نرم را کنترل کرد. معادله زیر چگونگی اعمال فاکتور دما در معادله Softmax را نشان می‌دهد.

$$p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (2-4)$$

خروجی‌های نرم در مدل‌های آموزگار و دانش آموز و نیز خروجی‌های مطلوب^{۱۷} نقش مهمی در آموزش دانش آموز دارند و با داشتن این مقادیر می‌توان تابع هزینه‌ی مربوط به چگالش دانش^{۱۸} و تابع هزینه‌ی دانش آموز^{۱۹} را تعریف نمود. تابع هزینه‌ی چگالش دانش را می‌توان به صورت زیر میان لاجیت‌های آموزگار و دانش آموز به صورت یک تابع هزینه‌ی Cross-Entropy نوشت.

¹⁶Temperature Factor

¹⁷Ground Truth

¹⁸Distillation Loss

¹⁹Student Loss

$$L_D(p(z_t, T), p(z_s, T)) = -\sum_i p_i(z_{ti}, T) \log(p_i(z_{si}, T)) \quad (3-4)$$

که در آن z_t و z_s به ترتیب، لاجیت‌های آموزگار و دانش‌آموز هستند. گرادیان تابع هزینه‌ی چگالش نسبت به لاجیت‌های دانش‌آموز می‌تواند به صورت زیر محاسبه گردد.

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{p(z_t, T) - p(z_s, T)}{T} = \frac{1}{T} \left(\frac{\exp(z_{si}/T)}{\sum_j \exp(z_{sj}/T)} - \frac{\exp(z_{ti}/T)}{\sum_j \exp(z_{tj}/T)} \right) \quad (4-4)$$

اگر فاکتور دما (T) بسیار بزرگتر از مقدار لاجیت‌ها باشد آنگاه با استفاده از بسط تیلور معادله‌ی ۴-۴ را می‌توان به صورت زیر بازنویسی نمود.

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{1}{T} \left(\frac{1 + \frac{z_{si}}{T}}{N + \sum_j \frac{z_{sj}}{T}} - \frac{1 + \frac{z_{ti}}{T}}{N + \sum_j \frac{z_{tj}}{T}} \right) \quad (5-4)$$

با فرض میانگین صفر بودن لاجیت‌های آموزگار و دانش‌آموز معادله‌ی فوق را می‌توان ساده‌تر نمود.

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{1}{NT^2} (z_{si} - z_{ti}) \quad (6-4)$$

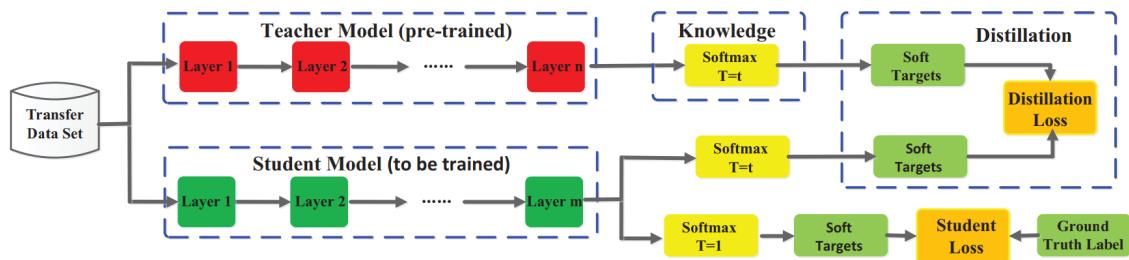
بنابراین با توجه به معادله‌ی ۶-۴ هدف تابع هزینه‌ی چگالش، کمینه کردن فاصله‌ی میان لاجیت‌های آموزگار و شاگرد است [۹۲]. تابع هزینه‌ی دانش‌آموز را نیز می‌توان به صورت Cross-Entropy بین خروجی مطلوب و لاجیت‌های دانش‌آموز به صورت زیر نوشت.

$$L_S(y, p(z_s, T)) = -\sum_i y \log(p_i(z_{si}, T)) \quad (7-4)$$

در نهایت، با کنار هم قرار دادن تابع هزینه‌ی چگالش و تابع هزینه‌ی دانش‌آموز، تابع هزینه‌ی کلی برای چارچوب چگالش دانش از آموزگار به شاگرد طبق معادله‌ی ۸-۴ تشکیل می‌شود که با ضریب α مجموع این دو تابع به صورت وزن‌دار محاسبه می‌شود.

$$L_{TotalDistillation} = \alpha L_D(p(z_t, T), p(z_s, T)) + (1 - \alpha) L_S(y, p(z_s, T)) \quad (8-4)$$

فاکتور دمای T در تابع هزینه‌ی دانش‌آموز برابر با یک و در تابع هزینه‌ی چگالش برابر یا بزرگتر از یک در نظر گرفته می‌شود. بنابراین به صورت کلی، چارچوب چگالش دانش را می‌توان به صورت شکل ۴-۷^{۳۷} بیان نمود که در آن، آموزگار یک مدل از پیش آموزش دیده شده است و دانش‌آموز به واسطه‌ی خروجی‌های نرم و مقادیر مطلوب که دو تابع هزینه را تشکیل می‌دهند، آموزش داده می‌شود [۳۷].

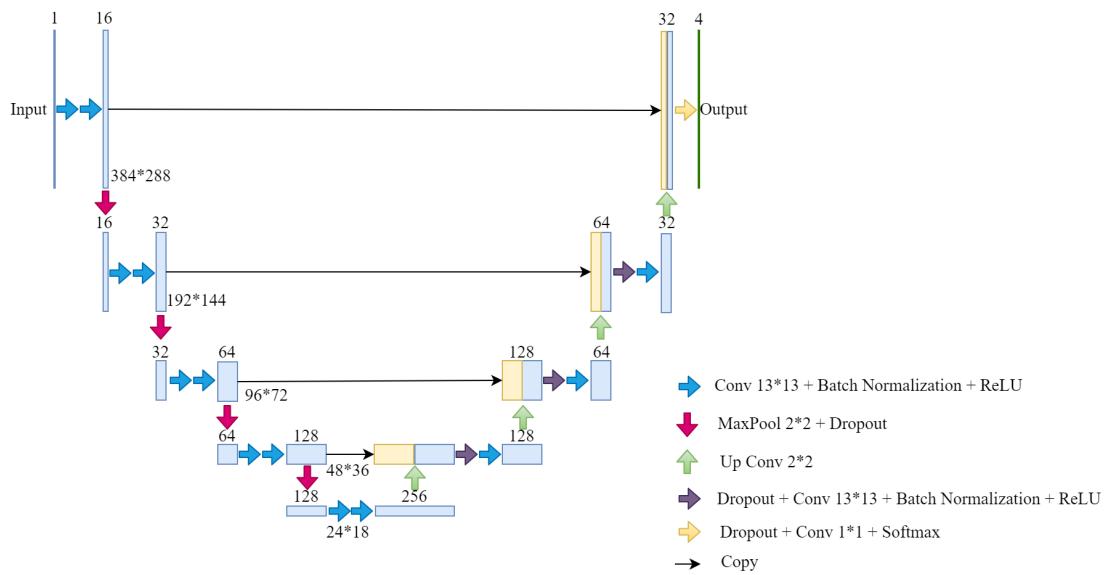


شکل ۴-۷: عملکرد انتقال دانش توسط چارچوب چگالش دانش [۳۷]

معماری شبکه‌های قطعه‌بند

همانطور که پیش‌تر اشاره گردید، در این قسمت، قصد داریم با استفاده از چارچوب چگالش دانش در بحث قطعه‌بندی ساختارهای در ریسک استفاده کنیم. برای این‌کار نیاز به مدل‌های آموزگار و دانش‌آموز است که در ابتدا یک مدل پیچیده با توانایی بالا بر روی دادگان به عنوان آموزگار آموزش می‌بیند و سپس مدل ساده‌تر (دانش‌آموز) با استفاده از چارچوب چگالش دانش، از دانش مدل آموزگار برای بهبود عملکرد بهره می‌برد.

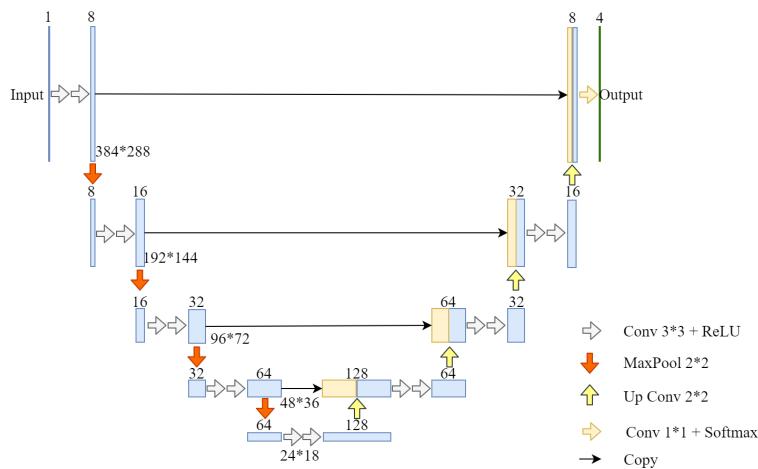
برای مدل آموزگار از معماری یک UNet پیچیده با تعداد پارامترهای قابل آموزش بالا و روش‌های تعمیم‌پذیری استفاده گردید. معماری این شبکه‌ی تمام کانولوشنی عمیق در شکل ۸-۴ قابل مشاهده است.



شکل ۴-۸: معماری شبکه‌ی UNet پیچیده، استفاده شده به عنوان آموزگار

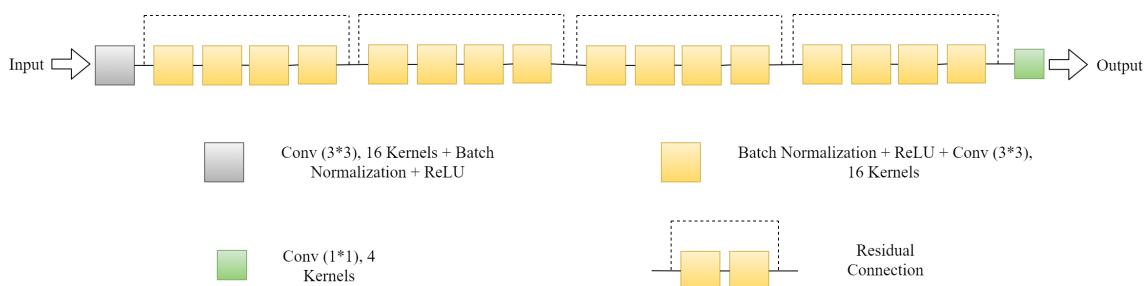
تفاوت این معماری با معماری UNet اولیه‌ی معرفی شده در [۲۷] استفاده از ابزارهای افزایش قدرت تعمیم‌پذیری مانند dropout و Batch-normalization پس از لایه‌های کانولوشنی است. همچنین برای افزایش وسعت دید شبکه و استخراج ویژگی‌ها عمیق‌تر و کلی‌تر از سطح تصویر، اندازه‌ی فیلترهای کانولوشنی برابر با 13×13 تعریف گردید.

در ادامه، برای تعریف مدل دانش‌آموز، از دو معماری ساده‌تر استفاده شد. دانش‌آموز اول یک مدل UNet بسیار ساده‌تر با تعداد پارامترهای بسیار کمتر از UNet پیچیده‌ی آموزگار و دانش‌آموز دوم یک معماری بر اساس بلوک‌های Residual است. شکل ۹-۴ معماری یک UNet ساده شده را به عنوان دانش‌آموز اول نشان می‌دهد که در مقایسه با معماری آموزگار در شکل ۸-۴، ساختار ساده‌تری دارد که عبارتند از: کاهش ابعاد فیلترهای کانولوشنی به 3×3 و در نتیجه کاهش وسعت دید هر فیلتر، استفاده نکردن از روش‌های تعمیم‌پذیری dropout و Batch-normalization و در نهایت کاهش تعداد ویژگی‌های استخراج شده به نصف، نسبت به UNet پیچیده.



شکل ۴-۹: معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانشآموز اول

شکل ۴-۱۰، معماری دانشآموز دوم را که بر اساس بلوک‌های Residual طراحی شده است را نشان می‌دهد. که دوباره، نسبت به مدل آموزگار، دارای تعداد ویژگی نصف در هر لایه است منتها برای آموزش بهتر آن از روش Batch-normalization نیز استفاده شده است.



شکل ۴-۱۰: معماری شبکه‌ی پیشنهادی بر اساس بلوک‌های Residual، استفاده شده به عنوان دانشآموز دوم

آموزش شبکه‌های قطعه‌بند

برای آموزش شبکه‌های معرفی شده در بالا، از مجموعه دادگان ساختارهای در ریسک قفسه‌ی سینه‌ی SegTHOR معرفی شده در بخش ۱-۲-۴ استفاده گردید. همانظور که گفته شد، این مجموعه داده قطعه‌بندی چهار ساختار در ریسک را در ۴۰ تصویر سی‌تی اسکن قفسه‌ی سینه تهیه نموده است. بنابراین برای قطعه‌بندی این چهار ساختار، همانطور که در شکل‌های ۴-۴ و ۴-۹ نشان داده شده است، در لایه‌ی خروجی این معماری‌ها چهار خروجی گرفته شده است که هر یک از آن‌ها مسؤول قطعه‌بندی

یکی از ساختارهای.

برای آموزش بر اساس چارچوب چگالش دانش لازم است تابع فعالیت خروجی برای کلاس‌ها Softmax باشد که با توجه به چهار ساختار در ریسک قلب، آئورت، نای و مری یک کلاس پس‌زمینه نیز به این چهار کلاس اضافه می‌شود تا بتوان از تابع هزینه‌ی Cross-Entropy استفاده نمود. اما با آزمایش‌های اولیه‌ای که صورت گرفت، ملاحظه گردید در نظر نگرفتن پس‌زمینه به عنوان یک کلاس مجزا و فقط قطعه‌بندی چهار ساختار در ریسک، دقت بالاتری را بوجود می‌آورد بنابراین معماری شبکه‌ها برای ایجاد خروجی، تنها برای ساختارهای در ریسک طراحی گردید.

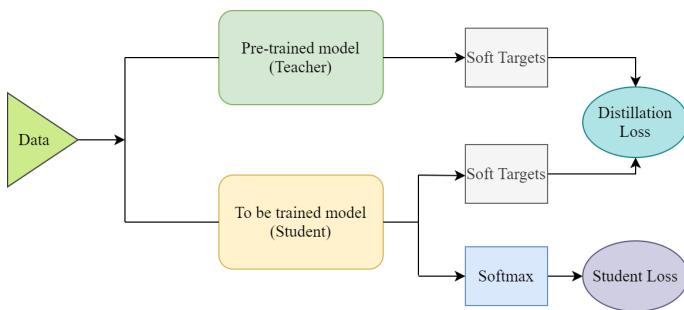
در این حالت، مشکلی که بوجود می‌آید عدم توانایی در استفاده از تابع هزینه‌ی Cross-Entropy است زیرا دیگر طبقه‌ی اضافی پس‌زمینه وجود ندارد. برای حل این مشکل از تابع هزینه‌ی Binary استفاده شد که به صورت زیر تعریف می‌شود:

$$BCE = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \sum_{l=1}^L y_i(k, l) \log(p_i(k, l)) + (1 - y_i(k, l)) \log(1 - p_i(k, l)) \quad (9-4)$$

که در آن y_i خروجی مطلوب که صفر یا یک است و p_i خروجی پیش‌بینی شده است که در بازه‌ی صفر و یک است، N تعداد نمونه‌ها و K و L ابعاد تصاویر هستند. فرض کنید اگر خروجی مطلوب برابر با صفر باشد، ترم اول در معادله‌ی ۹-۴ حذف می‌شود و ترم دوم سعی می‌کند با کمینه‌کردن مقدار هزینه، مقدار پیش‌بینی شده را به صفر نزدیک کند. بنابراین با صفر بودن دایمی یکی از کلاس‌ها با استفاده از این تابع هزینه، مشکلی در یادگیری بوجود نمی‌آید و با حل کردن این مساله‌ی بهینه‌سازی برای کلاس پس‌زمینه، مقدار ۰/۲۵ در هر یک از نقشه‌های احتمال چهار ساختار در ریسک، بدست می‌آید.

برای آموزش شبکه‌ها به صورت تصادفی ۸۰ درصد دادگان (۳۲ تصویر سی‌تی اسکن) به عنوان دادگان آموزش و مابقی دادگان به عنوان دادگان آزمایش جدا شدند. بنابراین با توجه به موارد گفته شده، در ابتدا مدل آموزگار با استفاده از تابع هزینه‌ی Binary Cross-Entropy و روش بهینه‌سازی Adam و با روش افزایش داده‌ی توضیح داده شده، آموزش داده شد. سپس با توجه به چارچوب چگالش دانش، دو مدل دانش‌آموز با همان دادگان آموزشی آموزگار، آموزش دیدند.

برای آموزش، دو مدل دانش‌آموز مطابق شکل ۱۱-۴ دادگان آموزش ورودی به مدل از پیش آموزش دیده‌ی آموزگار و دانش‌آموز وارد می‌شود و با دریافت لاجیت‌های خروجی هر مدل تابع هزینه‌ی چگالش و دانش‌آموز تعریف خواهد شد.



شکل ۴-۱۱: آموزش شبکه‌های دانش‌آموز بر مبنای چارچوب چگالش دانش از آموزگار

به طور معمول برای تعریف تابع هزینه‌ی چگالش بین خروجی مدل آموزگار و دانش‌آموز از فاصله‌ی KL^{۲۰} استفاده می‌شود [۳۷] که با معادله‌ی زیر توصیف می‌شود و سعی در کم کردن فاصله‌ی بین دو توزیع احتمال را دارد.

$$L_D(p(z_s), p(z_t)) = L_{KL}(p(z_s), p(z_t)) = \sum p(z_s) \log\left(\frac{p(z_s)}{p(z_t)}\right) \quad (10-4)$$

بنابراین با استفاده از فاصله‌ی KL به عنوان تابع هزینه‌ی چگالش دانش و تابع هزینه‌ی BCE به عنوان تابع هزینه‌ی میان‌بینی دانش‌آموز و مقادیر مطلوب، هزینه‌ی کلی به صورت زیر تعریف گردید.

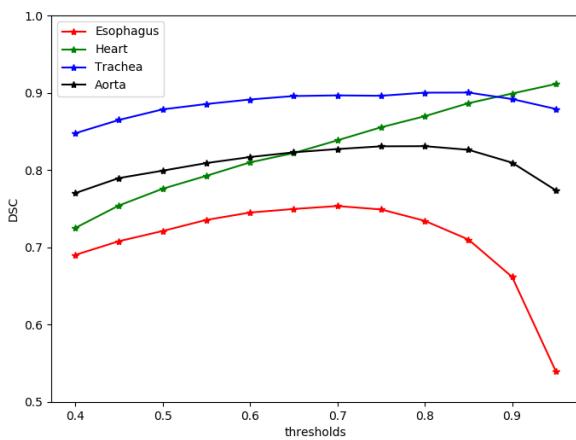
$$L_{total} = \alpha BCE(p(z_s), y) + (1 - \alpha) L_{KL}(p(z_s, T), p(z_t, T)) \quad (11-4)$$

با ملاحظات گفته شده در بالا و تنظیم پارامترهای T برابر با ۲ و α برابر با $0/2$ آموزش دو شبکه‌ی دانش‌آموز با چارچوب چگالش دانش از آموزگار صورت گرفت. برای مقایسه‌ی عملکرد این چارچوب، مدل‌های دانش‌آموز یک بار دیگر با همان دادگان آموزش برای آموزگار و شرایط یکسان منتها به صورت معمول و تنها تابع هزینه‌ی BCE آموزش داده شدند. در ادامه به بررسی نتایج حاصل پرداخته خواهد شد.

²⁰Kullback-Leibler (KL) divergence loss

نتایج

برای محاسبهٔ معیارهای ارزیابی باید برچسب‌های احتمالاتی پس از تابع فعالیت Softmax آستانه‌گذاری شوند و برخلاف تابع فعالیت Sigmoid که آستانه‌ی مشخص ۰/۵ را دارد، در این حالت باید آستانه‌های بهینه برای بدست آوردن بالاترین مقدار در معیارهای قطعه‌بندی استخراج گردد. برای این‌کار، پیش‌بینی‌های احتمالاتی برای مدل آموزگار، بر روی مجموعه داده‌ی آموزش استخراج گردید و با آستانه‌گذاری‌های مختلف، معیار Dice برای دادگان آموزش و چهار ارگان در ریسک استخراج گردید. نمودار زیر تاثیر مقدار آستانه را بر روی مقدار معیار Dice برای هر ساختار نشان می‌دهد.



شکل ۴-۱۲: تاثیر تغییر آستانه بر روی معیار Dice برای هر ساختار در ریسک

همانطور که در شکل ۴-۱۲ ملاحظه می‌شود مقدار آستانه بهینه برای هر ساختار متفاوت است. برای استخراج این آستانه‌ی بهینه از روش آستانه‌گذاری اوتسو، که در فصل قبل توضیح داده شد، استفاده گردید، به این ترتیب آستانه‌ی بهینه برای مری، ۰/۷۱، قلب، ۰/۹۳، نای، ۰/۸۶ و آورت، ۰/۸۲ بدست آمد. بنابراین با این آستانه‌گذاری ماسک‌های صفر و یک، برای چهار ارگان در ریسک ذکر شده بر روی دادگان آزمایش برای تمامی مدل‌ها بدست آمد. جدول‌های زیر دو معیار ارزیابی Dice و Jaccard را برای چهار ساختار در ریسک، نشان می‌دهد. در سه ردیف نخست، ارزیابی عددی برای شبکه‌های آموزگار، دانش‌آموز Resnet (شبکه‌ی مبتنی بر بلوک‌های Residual) و دانش‌آموز UNet ساده قرار دارند. در دو ردیف انتهایی، دو مدل دانش‌آموز به صورت تنها و بدون استفاده از چهارچوب چگالش دانش، ارزیابی شده‌اند.

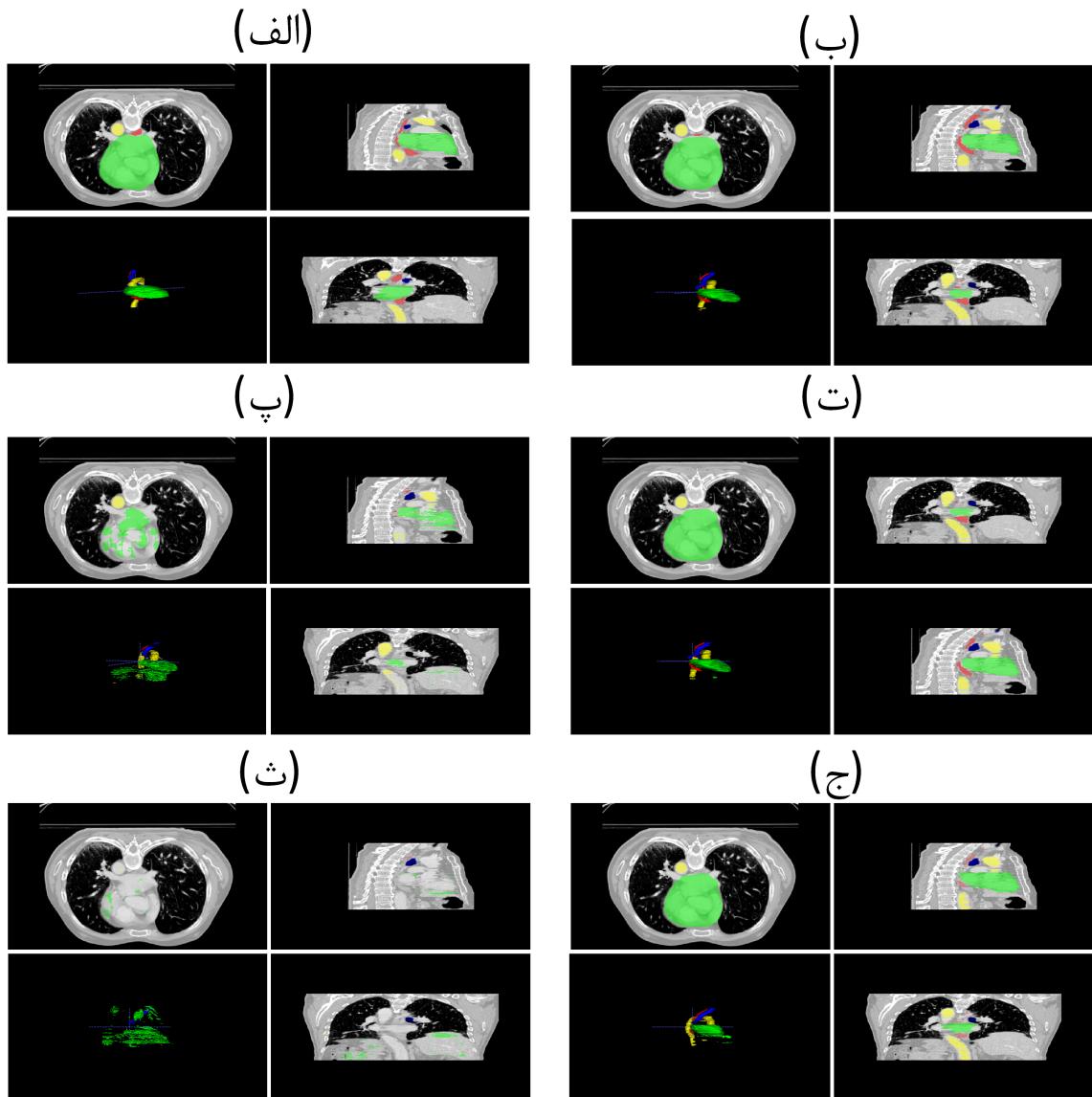
جدول ۴-۱: معیار Dice برای مدل‌های مختلف

نام ساختار در ریسک					
نام مدل	مری	قلب	نای	آئورت	نام ساختار در ریسک
UNet آموزگار	۰/۷۳ ± ۰/۱۳	۰/۹۵ ± ۰/۰۲	۰/۹۳ ± ۰/۰۳	۰/۸۶ ± ۰/۰۸	آئورت
Resnet دانش‌آموز	۰/۴۵ ± ۰/۱۱	۰/۶۲ ± ۰/۰۳	۰/۸۱ ± ۰/۰۶	۰/۶۳ ± ۰/۱۷	نای
Unet دانش‌آموز	۰/۶۲ ± ۰/۱۳	۰/۹۱ ± ۰/۰۴	۰/۸۹ ± ۰/۰۷	۰/۷۹ ± ۰/۱۸	قلب
Resnet تنها	۰/۳۳ ± ۰/۰۷	۰/۱۳ ± ۰/۰۴	۰/۸۰ ± ۰/۰۵	۰/۲۲ ± ۰/۰۵	آئورت
Unet تنها	۰/۵۴ ± ۰/۱۳	۰/۷۹ ± ۰/۰۹	۰/۸۸ ± ۰/۰۵	۰/۶۴ ± ۰/۱۷	نای

جدول ۴-۲: معیار Jaccard برای مدل‌های مختلف

نام ساختار در ریسک					
نام مدل	مری	قلب	نای	آئورت	نام ساختار در ریسک
UNet آموزگار	۰/۶۴ ± ۰/۱۲	۰/۹۰ ± ۰/۰۳	۰/۹۳ ± ۰/۰۳	۰/۸۲ ± ۰/۰۹	آئورت
Resnet دانش‌آموز	۰/۳۸ ± ۰/۰۹	۰/۶۰ ± ۰/۰۳	۰/۷۷ ± ۰/۰۷	۰/۵۶ ± ۰/۱۶	نای
Unet دانش‌آموز	۰/۵۳ ± ۰/۱۱	۰/۸۹ ± ۰/۰۴	۰/۸۶ ± ۰/۰۷	۰/۷۴ ± ۰/۱۹	قلب
Resnet تنها	۰/۲۹ ± ۰/۰۵	۰/۰۹ ± ۰/۰۳	۰/۷۶ ± ۰/۰۵	۰/۲۲ ± ۰/۰۴	آئورت
Unet تنها	۰/۴۵ ± ۰/۱۱	۰/۷۷ ± ۰/۰۹	۰/۸۵ ± ۰/۰۶	۰/۵۶ ± ۰/۱۵	نای

یک نمونه از خروجی قطعه‌بندی تصاویر سی‌تی اسکن سه‌بعدی دادگان آزمایش، در چهار نمای مختلف اکسیاب، سجیتال، کرونال و سه‌بعدی برای مدل‌های مختلف در شکل ۱۳-۴ نشان داده شده است که در آن (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش‌آموز (ت) پیش‌بینی دانش‌آموز، UNet، (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی Resnet ساده‌ی تنها هستند.

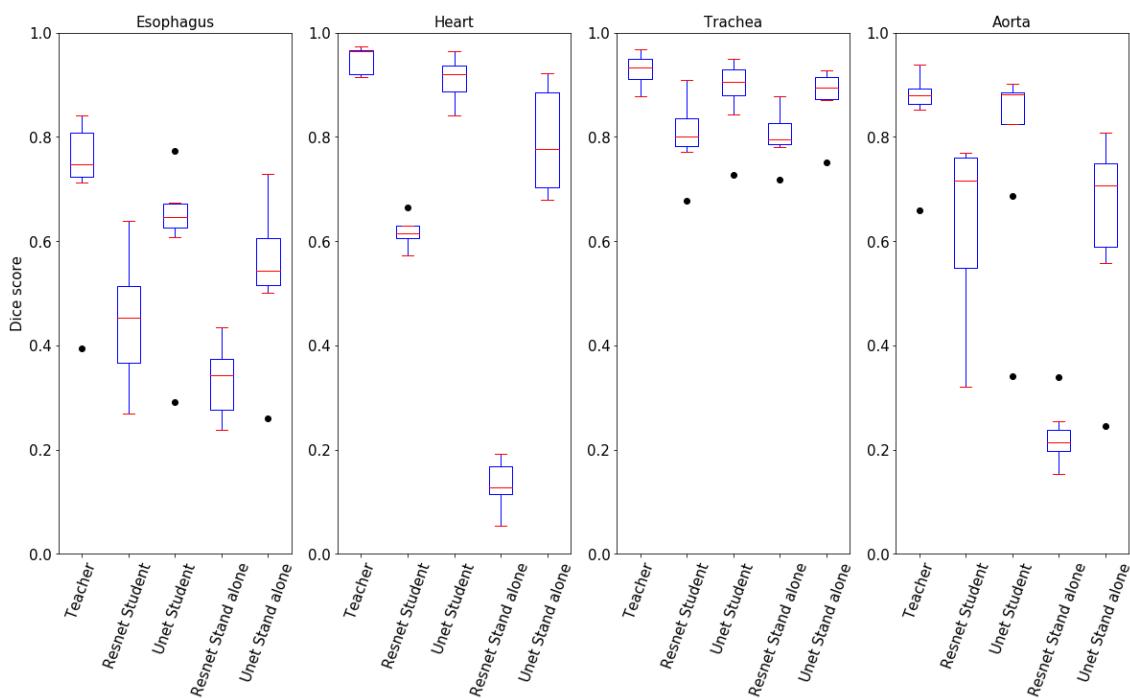


شکل ۴-۱۳: خروجی مدل‌های مختلف به صورت آموزش داده شده با چارچوب چگالش دانش و تکی، (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش‌آموز Resnet (ت) پیش‌بینی دانش‌آموز UNet، (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی UNet ساده‌ی تنها

بحث و نتیجه‌گیری

همانطور که اشاره گردید، هدف از چارچوب چگالش دانش، انتقال دانش یادگیری شده از آموزگار به دانش‌آموز است که این روند باعث بهبود دقیق و یادگیری بهتر در دانش‌آموز می‌شود. همانطور که در

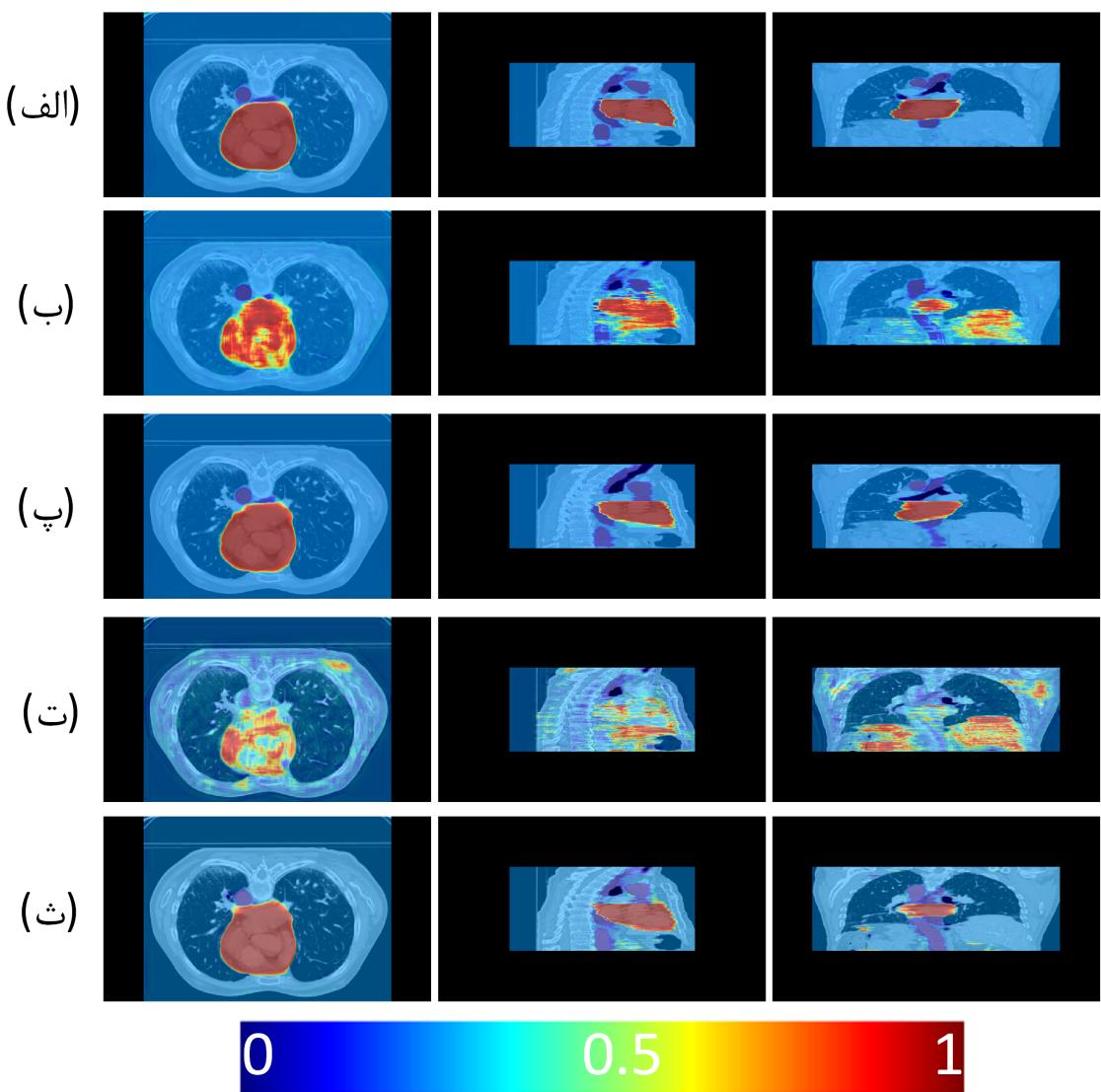
جدول‌های ۱-۴ و ۲-۴ و شکل ۱۳-۴ ملاحظه گردید، استفاده از این چارچوب باعث افزایش دقت عملکرد در قطعه‌بندی ساختارهای در ریسک بدون تغییر مدل و اضافه کردن هزینه‌ی محاسباتی می‌شود. برای مقایسه‌ی بهتر بهبود عملکرد نمودار جعبه‌ای^{۲۱} معیار Dice دادگان آزمایش برای چهار ساختار در ریسک و برای پنج شبکه‌ی آموزش داده شده به صورت یکجا در شکل ۱۴-۴ رسم گردید. همانطور که ملاحظه می‌شود شبکه‌هایی دانشآموز که یکبار به صورت تکی و بار دیگر با چارچوب چگالش دانش آموزش دیده‌اند، عملکردشان بهبود یافته است.



شکل ۴-۱۴: نمودار جعبه‌ای معیار Dice دادگان آزمایش برای چهار ساختار در ریسک و پنج شبکه‌ی آموزش داده شده

برای شهود بیشتر عملکرد چارچوب چگالش دانش، شکل ۱۵-۴ را در نظر بگیرید. در این شکل، نقشه‌ی احتمالاتی خروجی شبکه‌ها قبل از آستانه‌گذاری برای قلب نشان داده شده است. ملاحظه می‌شود که طبق انتظار، پس زمینه‌ی تصویر با احتمال ۰/۲۵ احتمال با احتمال بسیار نزدیک به صفر است. نکته‌ی قابل توجه در نقشه‌ی احتمالاتی دانشآموزان با و بدون استفاده از چگالش دانش است که در در Resnet احتمال بالایی برای حضور قلب در سمت راست بدن ایجاد شده است اما بعد از چگالش دانش این مشکل رفع شده است.

²¹Box Plot



شکل ۴-۱۵: نقشه‌های احتمالاتی خروجی شبکه‌ها قبل از آستانه‌گذاری (الف) آموزگار (ب) دانش‌آموز (پ) دانش‌آموز UNet (ت) Resnet (ث) Resnet

هدف چگالش دانش افزایش دقت در مدل‌های ضعیفتر است. برای مقایسه‌ی میزان پیچیدگی مدل‌ها از سه معیار تعداد پارامترهای قابل یادگیری (بر حسب میلیون)، تعداد عملیات اعشاری در واحد زمان^{۲۲} (بر حسب میلیارد) و متوسط زمان اجرا بر روی تصاویر سه‌بعدی (بر حسب ثانیه) در جدول ۴-۳ نشان داده شده است که حاکی از پیچیدگی بسیار زیاد مدل آموزگار نسبت به مدل‌های دانش‌آموز است. اما با استفاده از روش چگالش دانش، ملاحظه گردید که بدون افزایش پیچیدگی مدل می‌توان

^{۲۲}Floating Point Operations Per Second (FLOPS)

دقت آن را به دقت بالاتری رسانید و از هزینه‌های محاسباتی بالا و زمان پردازش زیاد جلوگیری نمود.

جدول ۴-۳: مقابسه‌ی میزان پیچیدگی و هزینه‌ی محاسباتی هر مدل

نام شبکه	تعداد پارامترها (میلیون) (G)	FLOPs	زمان اجرا (ثانیه)
Unet آموزگار	۱۴/۲	۱۸۳	۳۳/۵۴۲
Resnet	۳/۵۴	۱۸/۱	۰/۰۸۳
Unet ساده	۲/۲۶	۲/۷۱	۰/۵۴۰

۴-۲-۳ بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک

همانطور که می‌دانید، عملکرد شبکه‌های عمیق علی‌الخصوص شبکه‌های کانولوشنی عمیق بر اساس عملکرد کورتکس بینایی انسان طراحی شده است. از بین مدل‌هایی که برای عملکرد سیستم عصبی ارایه شده است، مدل کدگذاری پیش‌بین^{۲۳} بر اساس شبکه‌های بازگشتی^{۲۴}، یک مسیر بازخورد از لایه‌های بالاتر به لایه‌ها پایین‌تر که حاوی پیش‌بینی عملکرد ناحیه‌های پایینی کورتکس است، متصل می‌شود و در مسیر روبه‌جلو، خطای پیش‌بینی شده جریان می‌یابد [۹۳].

با الگوبرداری از این مدل، در این قسمت قصد داریم با اضافه کردن یک مسیر پیش‌بین به شبکه‌های روبه‌جلوی کانولوشنی و بازخورد دادن خطای پیش‌بینی، یک چارچوب برای افزایش دقیق قطعه‌بند برای قطعه‌بندی ساختارهای در ریسک ارایه دهیم.

پیش‌تر اشاره گردید که مفهوم و عملکرد شبکه‌های کانولوشنی عمیق از عملکرد مغز الگو برداری شده است اما برخلاف مدل کدگذاری پیش‌بین، فقط یک مسیر رو به جلو در طراحی این شبکه‌ها لحاظ شده است در صورتی که در عملکرد کورتکس بینایی مسیرهای بازگشتی برای بازخورد خطای از لایه‌های بالاتر به لایه‌های پایین‌تر نیز وجود دارد. برای رفع این مشکل، چارچوب بازخورد خطای پیش‌بین در این مطالعه ارایه گردیده است.

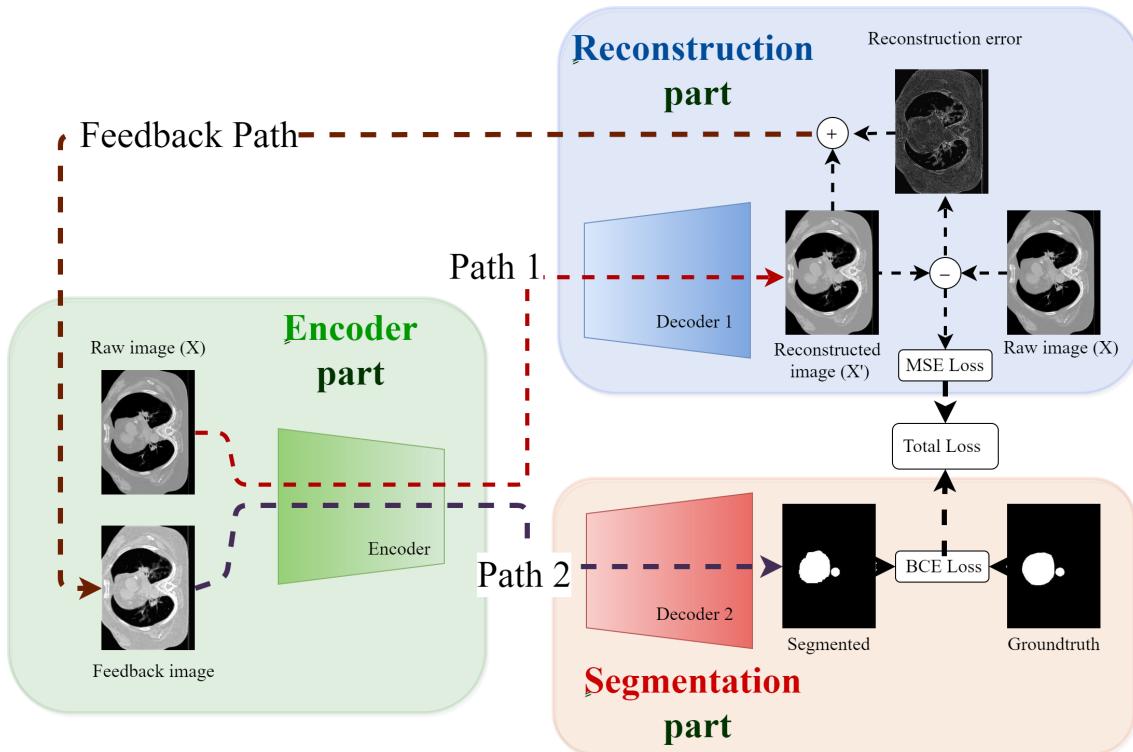
²³Predictive Coding Model

²⁴Recurrent Networks

معماری چارچوب و آموزش مدل

چارچوب بازخورد خطای پیش‌بین، از یک شبکه‌ی رمزگذار (Encoder) و دو شبکه‌ی رمزگشا (De-coder) تشکیل شده است. برای شهود بیشتر شکل ۴-۱۶ را در نظر بگیرید که در آن قصد، قطعه‌بندی ساختارهای در ریسک لایه‌های دو بعدی تصاویر سه بعدی سی‌تی اسکن است. این چارچوب از سه مسیر که در شکل نشان داده شده‌اند، تشکل شده است:

- مسیر اول یا مسیر بازسازی ورودی
- مسیر بازخورد خطای پیش‌بین
- مسیر دوم یا مسیر پیش‌بینی قطعه‌بندی



شکل ۴-۱۶: عملکرد چارچوب بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک

در مسیر اول ابتدا تصویر ورودی به شبکه‌ی کدگذار وارد می‌شود و بردارهای ویژگی استخراج شده از آن، به شبکه رمزگشای اول (قسمت بازسازی) وارد می‌شود تا از تصویر ورودی یک بازسازی حاصل شود. سپس این بازسازی با تصویر اولیه مقایسه می‌شود تا نواحی‌ای که به اشتباه بازسازی شده‌اند مشخص

شوند. این اختلاف بازسازی با یک ضریب با تصویر پیش‌بینی شده، جمع می‌شود و از طریق مسیر بازخورد به قسمت شبکه‌ی رمزگذار بازخورد داده می‌شود. تا این قسمت از روند را می‌توان به صورت زیر فرمول‌بندی نمود:

$$X' = \text{Decoder}_1(\text{Encoder}(X)) \quad (12-4)$$

$$Z = X' + \beta |X' - X|$$

که در آن X تصویر ورودی، X' تصویر بازسازی شده، Z تصویر شامل خطای پیش‌بینی در مسیر بازخورد و β ضریب ثابت برای جمع وزن‌دار تصویر پیش‌بینی با خطای بازسازی است.

در گام بعدی از مسیر دوم، مجموع وزن‌دار خطای پیش‌بینی و تصویر پیش‌بینی شده وارد شبکه‌ی کدگذار و سپس از قسمت کدگشای شبکه‌ی قطعه‌بند عبور می‌کند تا قطعه‌بندی نهایی تولید گردد. عملکرد این قسمت نیز به شکل زیر فرمول‌بندی می‌شود.

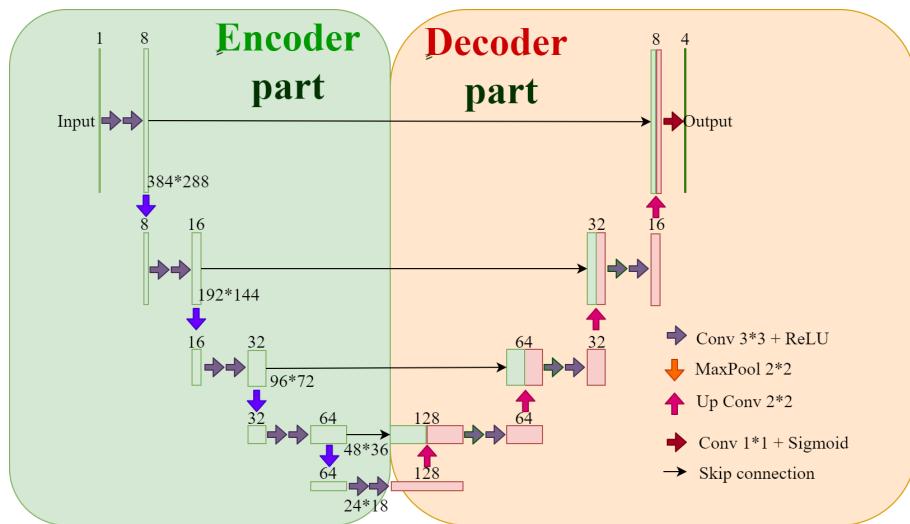
$$P = \text{Decoder}_2(\text{Encoder}(Z)) \quad (13-4)$$

که در آن P خروجی قطعه‌بندی نهایی است. برای آموزش یک شبکه توسط این چارچوب، لازم است دوتابع هزینه تعریف گردد، تابع هزینه‌ی شبکه بازسازی کننده که باید فاصله‌ی میان تصویر بازسازی شده و تصویر اولیه را کمینه کند و تابع هزینه‌ی شبکه قطعه‌بندی که هزینه‌ی بین ماسک قطعه‌بندی خروجی و قطعه‌بندی مطلوب را کمینه می‌کند. برای تابع هزینه‌ی بازسازی می‌توان از میانگین مربعات خطأ^{۲۵} (MSE) و برای تابع هزینه‌ی قطعه‌بندی از BCE استفاده کرد. جمع وزن‌دار این دوتابع هزینه، هزینه‌ی نهایی را ایجاد می‌کند. معادلات ۱۴-۴ تابع هزینه‌ی این چارچوب را به صورت فرمول‌بندی نشان می‌دهد.

$$\begin{aligned} MSE(X, X') &= \frac{1}{N} \sum_{i=1}^N (X_i - X'_i)^2 \\ Loss &= \alpha MSE(X, X') + BCE(Y, P) \end{aligned} \quad (14-4)$$

²⁵Mean Square Error

برای ارزیابی این چارچوب از معماری یک UNet مطابق با شکل ۱۷-۴ استفاده گردید. برای شبکه‌های رمزگذار و رمزگشای معرفی شده در شکل ۱۶-۴، از قسمت‌های Encoder و Decoder و معماري شبکه‌ی زیر استفاده گردید. که Encoder برای دو شبکه‌ی Decoder به صورت مشترک عمل می‌کند. مسیرهای اتصالی از Decoder به همان در چارچوب معرفی شده برای استفاده از شبکه‌ی UNet پا برجا هستند.



شکل ۱۷-۴ : معماري شبکه‌ی UNet استفاده شده برای آموزش با چارچوب بازخورد خطای پیش‌بین

برای آموزش این شبکه، از مجموعه دادگان SegTHOR با شرایط گفته شده در بخش‌های قبل استفاده گردید. پارامترهای α و β در معادلات ۱۴-۴ و ۱۲-۴ به ترتیب برابر با $0/2$ و 2 در روند آموزش قرار داده شد.

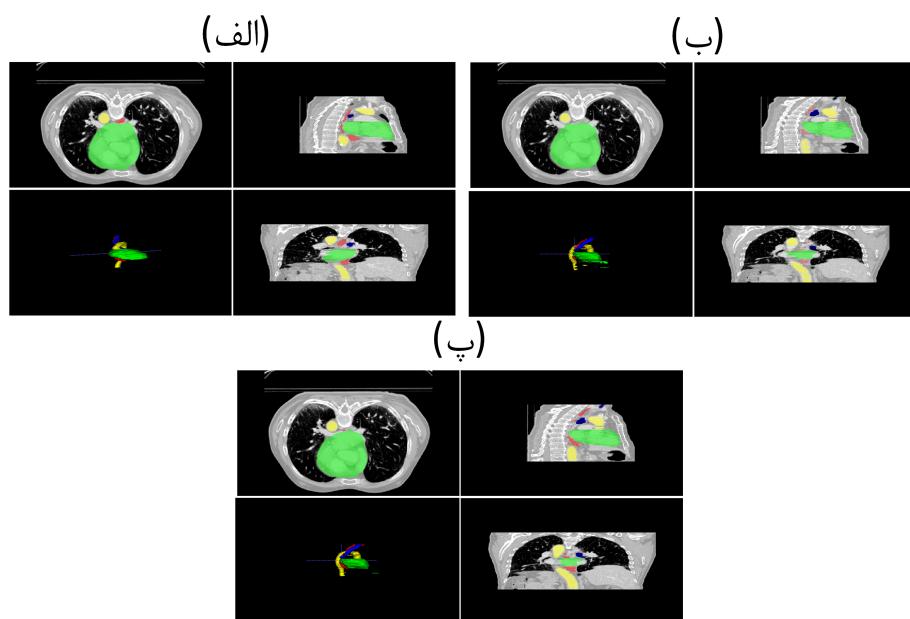
نتایج و ملاحظات

شبکه‌ی UNet یک مرتبه به صورت تکی و یک بار با استفاده از چارچوب بازخورد خطای پیش‌بین، آموزش داده شد. معیارهای ارزیابی برای این دو شبکه به صورت زیر بر روی دادگان آزمایش استخراج گردید.

جدول ۴-۴: معیارهای ارزیابی قطعه‌بندی برای شبکه‌ی آموزش دیده شده به دو صورت تکی و چارچوب بازخورد خطای پیش‌بین

میانگین	نام ساختار در ریسک			معیار	نام شبکه
	آئورت	نای	قلب		
۰/۸۲ ± ۰/۰۷	۰/۸۲ ± ۰/۱۳	۰/۸۸ ± ۰/۰۳	۰/۸۸ ± ۰/۰۴	۰/۶۸ ± ۰/۰۸	Dice
۵/۹۶ ± ۰/۸۰	۶/۱۶ ± ۱/۰۱	۴/۶۷ ± ۰/۴۰	۸/۳۷ ± ۱/۰۶	۴/۶۳ ± ۰/۷۳	Hausdorff با چارچوب
۰/۷۷ ± ۰/۰۷	۰/۷۶ ± ۰/۱۴	۰/۸۵ ± ۰/۰۴	۰/۸۷ ± ۰/۰۴	۰/۵۸ ± ۰/۰۸	Jaccard
۰/۷۱ ± ۰/۱۱	۰/۶۴ ± ۰/۱۷	۰/۸۸ ± ۰/۰۵	۰/۷۹ ± ۰/۰۹	۰/۵۴ ± ۰/۱۳	Dice
۶/۳۴ ± ۰/۷۷	۶/۶۳ ± ۰/۸۶	۴/۵۲ ± ۰/۱۸	۹/۵۰ ± ۱/۴۹	۴/۷۰ ± ۰/۵۶	Hausdorff تنها
۰/۶۶ ± ۰/۱۰	۰/۵۶ ± ۰/۱۵	۰/۸۵ ± ۰/۰۶	۰/۷۷ ± ۰/۰۹	۰/۴۵ ± ۰/۱۱	Jaccard

شکل ۱۸-۴ پیش‌بینی دو شبکه‌ی UNet آموزش دیده شده به صورت تکی (ب) و با چارچوب بازخورد خطای پیش‌بین (پ) را در کنار قطعه‌بندی مطلوب (الف) در نماهای مختلف نشان می‌دهد.



شکل ۱۸-۴: پیش‌بینی قطعه‌بندی ساختارهای در ریسک در نماهای مختلف (الف) خروجی مطلوب (ب) خروجی UNet تنها (پ) خروجی UNet آموزش دیده با چارچوب بازخورد خطای پیش‌بین همانطور که در شکل ۱۸-۴ و جدول ۴-۴ ملاحظه می‌شود با استفاده کردن از چارچوب بازخورد

خطای پیشین، دقت قطعه‌بندی برای ساختارهای در ریسک افزایش خواهد یافت (میانگین معیار Dice برای چهار ارگان در ریسک ۱۱/۰ رشد داشته است). علت این بهبود در استفاده از این چارچوب را می‌توان در دو دلیل توضیح داد:

- استفاده از یک Encoder مشترک در بازسازی ورودی و قطعه‌بندی آن با توابع هزینه‌ی مختلف، باعث هدایت کردن Encoder به سمت استخراج ویژگی‌های سطح بالا و در نتیجه بهبود عملکرد قسمت قطعه‌بندی می‌شود.
- با جمع کردن خطای پیش‌بینی شده با تصویر بازسازی شده و عبور دادن این تصویر برای قطعه‌بندی، ناحیه‌هایی که شبکه توانایی بازسازی آن‌ها نداشته است مورد تأکید قرار می‌گیرند و در مسیر دوم باعث بهبود دقت قطعه‌بندی خواهند شد.

بنابراین در این قسمت یک چارچوب برای آموزش شبکه‌های عصبی عمیق با رویکرد بازخورد خطابای افزایش دقت قطعه‌بندی ارایه گردید و ملاحظه شد، با آموزش یک UNet با استفاده از این چارچوب دقت به صورت قابل ملاحظه‌ای افزایش یافت.

۴-۲-۴ معرفی تابع هزینه‌ی براساس شکل برای قطعه‌بندی ساختارهای در ریسک

ابزارهای یادگیری عمیق نسبت به ابزارهای کلاسیک دیگر در حوزه‌ی قطعه‌بندی تصاویر پزشکی، دقت بسیار بالاتری نشان داده است. عناصر سازنده‌ی یک الگوریتم یادگیری عمیق از موارد زیر تشکیل شده است [۹۴].

- معماری شبکه با تعداد پارامتر قابل یادگیری مناسب برای تشکیل نگاشت غیرخطی بین ورودی و خروجی
- تابع هزینه‌ی مناسب برای آموزش شبکه
- دادگان آموزش مناسب
- الگوریتم بهینه‌سازی برای تنظیم وزن‌های شبکه

هر گونه ضعف در یکی از موارد بالا باعث ایجاد خروجی نامطلوب در پیش‌بینی شبکه‌ی عصبی عمیق می‌شود. یکی از موارد ضعف این مدل‌ها که در شکل‌های ۱۳-۴ و ۱۸-۴ مشاهده می‌شود وجود جزیره‌ای از قطعه‌بندی‌های نامطلوب یا سوراخ‌هایی در یک بافت یکپارچه است. مطالعات اخیر نشان می‌دهد، برخلاف توانمندی بسیار بالای مدل‌های یادگیری عمیق در استخراج ویژگی‌های سطح بالا، این مدل‌ها توانایی یادگیری شکل و ساختار بافت‌ها را ندارند [۹۵]. علت این کاستی را در تابع هزینه‌ی طراحی شده برای شبکه‌های عصبی می‌توان یافت. این توابع هزینه معمولاً به صورت پیکسل به پیکسل عمل می‌کنند (مانند BCE) و ویژگی‌های مربوط به شکل و ساختار عضو مورد مطالعه را در نظر نمی‌گیرند.

بنابراین این مشکل در خروجی قطعه‌بندی باعث پدید آمدن کاستی‌هایی مانند، تکه‌تکه شدن ساختارهای یکپارچه، پدید آمدن جزیره‌های واکسلی و ناهمگونی شکل در ساختار قطعه‌بندی شده، می‌شود. روش‌های بسیاری برای حل این مشکل از طریق ایجاد یک دانش اولیه^{۲۶} برای قطعه‌بندی و یا روش‌های پس‌پردازش^{۲۷} ارایه شده است که از جمله‌ی آن‌ها می‌توان به فیلد‌های تصادفی شرطی/مارکوف^{۲۸} و مدل‌های کانتور فعل^{۲۹} اشاره کرد [۹۶].

همانطور که گفته شد این روش‌ها با تاثیر بر روی ورودی یا خروجی نهایی شبکه‌ی عصبی سعی در حل مشکلات ذکر شده دارند. اما هنوز مشکل حل نشده باقی مانده است و شبکه ویژگی‌های مربوط به شکل و ساختار را یاد نگرفته است. تا کنون تلاش‌هایی برای اجبار شبکه به یادگیری شکل شده است. محققی و همکاران [۹۷] در مطالعه‌ای برای قطعه‌بندی سه‌بعدی کبد با شبکه‌های عصبی عمیق، یک چارچوب برای استخراج ویژگی‌های شکل معرفی کردند. رویکرد آن‌ها استفاده از یک شبکه‌ی دیگر که ماسک قطعه‌بندی را بازسازی می‌کند و ویژگی‌های مربوط به این ماسک قطعه‌بندی شده‌ی سه‌بعدی را استخراج می‌کند سپس در شبکه‌ی قطعه‌بند یک تابع هزینه برای کمینه‌کردن فاصله‌ی ویژگی‌های استخراج شده از قطعه‌بندی مطلوب و قطعه‌بندی پیش‌بینی شده، اضافه نموده است. با وجود بهبود دقت در قطعه‌بندی، در این چارچوب، هنوز اطمینانی وجود ندارد که شبکه‌ی استخراج کننده‌ی ویژگی، ویژگی‌های مربوط به شکل را استخراج کرده است. بنابراین در این مطالعه سعی داریم تابع هزینه‌ای پیشنهاد کنیم که با اطمینان، ویژگی‌ها شکل و ساختار مورد نظر را استخراج و مقایسه کند.

ساختارهای در ریسک موجود در بدن، معمولاً شکل و ساختار خاصی دارند و در بین بیماران مختلف

²⁶Prior Knowledge

²⁷Post-Processing

²⁸Conditional/Markov Random Fields

²⁹Active Contour Models

تغییرات نسبتاً کمی دارند. بنابراین با اجبار شبکه به یادگیری شکل این ساختارها می‌توان قطعه‌بندی با دقیق‌تری برای این ساختارها تولید نمود.

استخراج ویژگی‌های تمایز پذیر بر اساس شکل

برای ایجاد ویژگی‌های تمایز پذیر بر اساس شکل و ساختار، لازم است ابتدا فضای شکل^{۳۰} تعریف شود تا بتوان ویژگی‌های مشترک میان اعضای سازنده‌ی آن فضا را استخراج نمود. برای این کار از ایده‌ی معرفی شده در مقاله‌ی Eigenface [۹۸] می‌توان استفاده کرد که با استخراج بردارهای ویژه از فضای برداری شده‌ی تصاویر چهره، امکان بازسازی تصویر اولیه را با مجموع دار بردارهای ویژه فراهم می‌سازد.

شکل ۱۹-۴ قسمت (الف) صد نمونه از تصاویر خاکستری چهره، با ابعاد 32×32 را نشان می‌دهد. برای استخراج بردارهای ویژه، ابتدا این تصاویر برداری‌سازی^{۳۱} شده و به بردار با ابعاد 1024×1 تغییر اندازه داده می‌شود. سپس ماتریس کوواریانس از این مجموعه بردارها، که فضای شکل را می‌سازند، استخراج می‌شود و با اعمال الگوریتم PCA^{۳۲} بردارهای ویژه (چهره‌های ویژه) و مقادیر ویژه استخراج می‌شود. پس از استخراج چهره‌های ویژه، با تغییر ابعاد مجدد به 32×32 می‌توان این تصاویر را در شکل ۱۹-۴ قسمت (ب) مشاهده نمود. فرمول‌بندی این روند را می‌توان به صورت زیر بیان نمود:

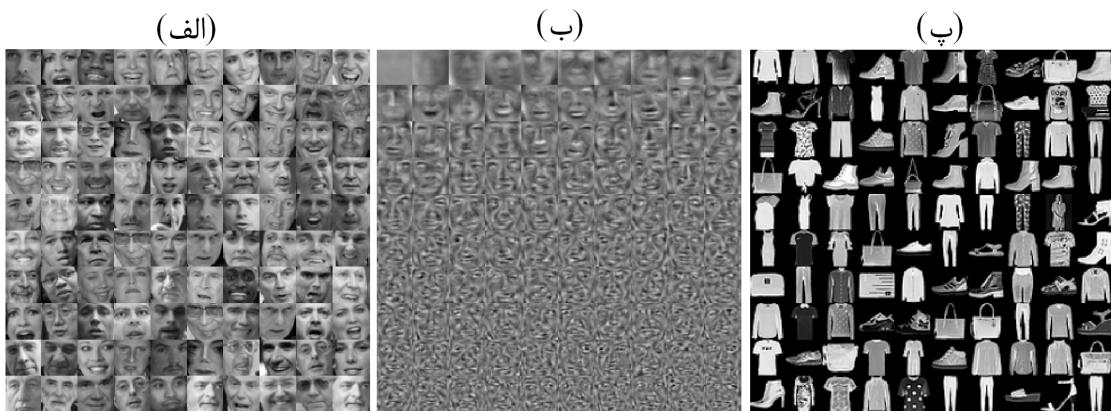
$$\begin{aligned} \theta_i &= \mathbf{L}_i - \mu & \mu &= \frac{1}{n} \sum_{i=1}^n \mathbf{L}_i \\ C &= \frac{1}{n} \sum_{i=1}^n \theta_i \theta_i^T = \frac{1}{n} \Theta \Theta^T & \Theta &= [\theta_1, \theta_2, \dots, \theta_n] \\ C \mathbf{u}_i &= \frac{1}{n} \Theta \Theta^T \mathbf{u}_i = \lambda_i \mathbf{u}_i & (15-4) \end{aligned}$$

که در آن، \mathbf{L}_i تصویر برداری شده‌ی شماره‌ی i ، μ میانگین تصاویر برداری شده، C ماتریس کوواریانس، \mathbf{u}_i بردار ویژه و λ_i مقدار ویژه است.

³⁰Shape-Space

³¹Vectorize

³²Principal Component Analysis



شکل ۴-۱۹: استخراج بردارهای ویژه از تصاویر چهره و مقایسه با تصاویر غیر چهره (الف) صد نمونه از تصاویر چهره استفاده شده، (ب) صد چهره‌ی ویژه (بردارهای ویژه از دادگان چهره) و (پ) صد نمونه تصویر غیر چهره (دادگان FashionMNIST) بازسازی شده.

با جمع دار این چهره‌های ویژه، می‌توان یک تصویر را بازسازی نمود. شکل ۴-۲۰ بازسازی یک تصویر چهره (سمت چپ) را با استفاده از چهره‌های ویژه را نشان می‌دهد که با جمع وزن‌دار تمام این چهره‌های ویژه می‌توان دقیقاً به تصویر اصلی رسید (وسط) اما با کم کردن چهره‌های ویژه، کلیت چهره قابل تشخیص است اما دقیقاً مانند تصویر اصلی نیست (سمت راست). وزن‌های بازسازی از فرمول‌های زیر استخراج می‌شود.

$$\mathbf{w} = \mathbf{L}_i \times U \quad U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n] \quad (16-4)$$

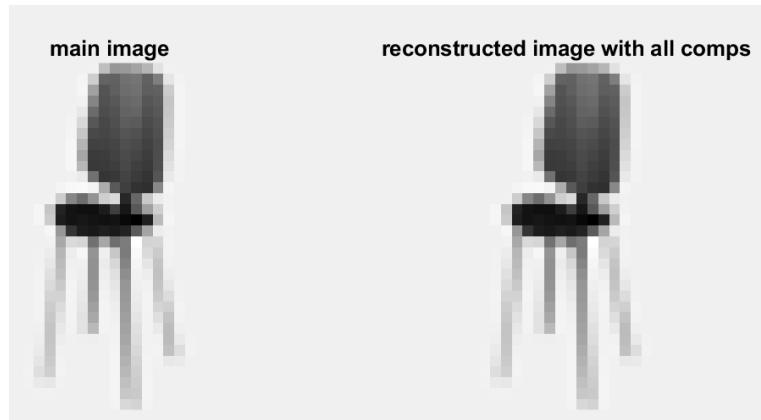
$$\mathbf{L}'_i = U \mathbf{w}^T \quad (17-4)$$

که \mathbf{L}'_i تصویر برداری بازسازی شده، U ماتریس بردارهای ویژه و \mathbf{w} بردار وزن‌های بازسازی است.



شکل ۴-۲۰: بازسازی یک تصویر چهره با استفاده از چهره‌های ویژه، سمت چپ: تصویر اصلی، وسط: تصویر بازسازی شده با تمام چهره‌های ویژه، سمت راست: تصویر بازسازی شده با ۲۰ چهره‌ی ویژه اول

حال سوالی که پیش می‌آید اینست که آیا می‌توان با استفاده از این چهره‌های ویژه تصاویر دیگر را بازسازی کرد؟ شکل ۴-۲۱ نشان می‌دهد با استفاده از تمام بردارهای ویژه می‌توان یک تصویر غیرچهره (صندلی) را به طور کامل بازسازی نمود. بنابراین ویژگی‌ای که باعث پدید آمدن تصاویر مختلف در طی بازسازی می‌شود، وزن‌های بازسازی است. در نتیجه می‌توان از این وزن‌ها برای تفکیک تصاویر چهره و غیرچهره استفاده کرد.

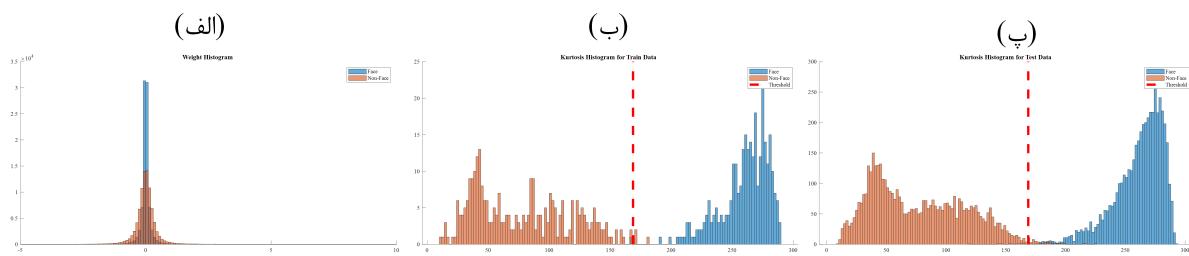


شکل ۴-۲۱: بازسازی یک تصویر غیرچهره با استفاده از چهره‌های ویژه، سمت چپ: تصویر اصلی، سمت راست: تصویر بازسازی شده با تمام چهره‌های ویژه

برای بررسی چگونگی تفکیک‌پذیری تصاویر چهره و غیر چهره، ۳۰۰ تصویر چهره و ۳۰۰ تصویر غیر چهره (دادگان آموزش) مطابق شکل ۴-۱۹، (الف) و (پ) استفاده گردید. ابتدا با به‌کارگیری چهره‌های ویژه، بردار وزن‌های بازسازی برای هر دو گروه چهره و غیرچهره استخراج گردید و هیستوگرام این وزن‌ها رسم شد. شکل ۴-۲۲ (الف)، هیستوگرام بازسازی وزن‌های دو گروه چهره (رنگ آبی) و

غیرچهره (رنگ آجری) را نشان می‌دهد. همانطور که در این شکل مشاهده می‌شود هیستوگرام وزن‌های تصاویر چهره نسبت به غیرچهره، تنکتر^{۳۳} است. دلیل این مشاهده را می‌توان به دلیل همبستگی بالای چهره‌های ویژه و تصاویر چهره دانست که با ضرایب بسیار کمتری می‌توان یک چهره را با این بردارهای ویژه بازسازی کرد، دانست. بر عکس در تصاویر غیر چهره همبستگی کمتری میان چهره‌های ویژه و این تصاویر وجود دارد و بنابراین هیستوگرام غیرتنکی بدست می‌آید. با این مشاهده و استدلال، اگر از بردار وزن‌های بازسازی تصویر، ممان چهارم (کورتوسیس^{۳۴}) طبق فرمول زیر، گرفته شود، انتظار می‌رود وزن‌های بازسازی چهره، کورتوسیس بالاتری داشته باشند.

$$Kurt(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] \quad (18-4)$$



شکل ۴-۲۲: تمایز بین تصاویر چهره و غیرچهره براساس وزن‌های بازسازی (الف) هیستوگرام وزن‌های بازسازی تصاویر چهره (آبی) و غیرچهره (آجری)، (ب) کورتوسیس وزن‌های بازسازی چهره و غیرچهره برای مجموعه داده‌ی آموزش و استخراج آستانه‌ی طبقه‌بندی با روش اوتسو (پ) طبقه‌بندی کورتوسیس وزن‌های بازسازی برای مجموعه دادگان آزمایش

در شکل ۴-۲۲ (ب)، کورتوسیس وزن‌های بازسازی برای ۳۰۰ داده‌ی آموزش چهره و غیرچهره رسم شده است که طبق انتظار، بردارهای وزن‌های بازسازی چهره، کورتوسیس بالاتری دارند. با استفاده از روش آستانه‌گذاری اوتسو، آستانه‌ی بهینه برای طبقه‌بندی این دو گروه از یکدیگر پیدا شد. در نهایت در قسمت (پ) شکل ۴-۲۲ برای ۴۷۰۰ تصویر آزمایش، با استفاده از چهره‌های ویژه‌ی استخراج شده در ۳۰۰ تصویر آموزش، کورتوسیس وزن‌های بازسازی محاسبه گردید و با آستانه‌ی بدست آمده در دادگان آموزش، طبقه‌بندی به دو گروه چهره و غیره‌چهره انجام گردید. دقت این طبقه‌بندی برابر با ۹۹/۳۷ درصد بدست آمد.

³³Sparse

³⁴Kurtosis

بنابراین در این بخش توانستیم بر اساس ویژگی شکل و ساختار تصویر یک معیار ارایه دهیم که بین دو گروه تصویر که ساختار ناهمسان دارند تفکیک قابل شود. بنابراین می‌توان از این روند برای تمایز میان شکل‌های معتبر و نامعتبر با توجه به مجموعه دادگان مورد نظر، استفاده نمود.

دادگان و تعریف تابع هزینه‌ی بر اساس شکل

همانطور که اشاره گردید، اکثر ساختارهای در ریسک، معمولاً در میان افراد مختلف شکل و ساختمان یکسانی دارند، حال اگر شبکه مجبور به یادگیری شکل این ساختار شود، قطعه‌بندی آن با دقت بالاتری صورت می‌پذیرد. برای این‌کار از دو مجموعه داده‌ی سی‌تی اسکن قلب و ام‌آرآی هیپوکامپ استفاده گردید. برای تصاویر سی‌تی اسکن قلب از مجموعه داده‌ی SegTHOR استفاده شد و تصویر قلب و قطعه‌بندی متناظر پس از پیش‌پردازش‌های گفته شده در قسمت ۱-۲-۴ با بریدن آن در اندازه‌ی $64 \times 64 \times 144 \times 160$ برای ۴۰ تصویر استخراج گردید. از این تصاویر ۸۰ درصد آن به عنوان مجموعه‌ی آموزش و مابقی برای مجموعه‌ی آزمایش به صورت تصادفی انتخاب گردید.

برای دادگان هیپوکامپ از دادگان و پیش‌پردازش‌های معرفی شده در قسمت ۱-۲-۴ استفاده شد و در نهایت ۹۰ درصد (۲۳۴ مورد) آن به عنوان مجموعه‌ی آموزش و مابقی برای مجموعه‌ی آزمایش (۲۶ مورد) به صورت تصادفی انتخاب گردید.

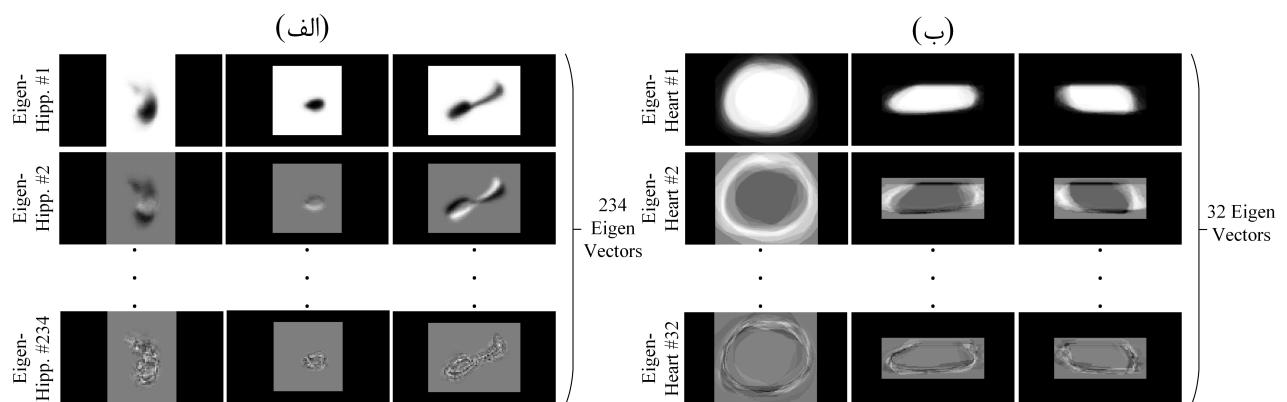
حال، با توجه به ویژگی استخراج شده در بخش قبل که بر اساس شکل و ساختار میان تصاویر چهره و غیر چهره تمایز قابل می‌شد، این قابلیت تمایز، برای برچسب‌های پیش‌بینی شده‌ی قطعه‌بندی ساختارها می‌تواند مورد استفاده قرار گیرد. برای این‌کار طبق آنچه گفته شد، ابتدا تصاویر سه‌بعدی برچسب‌ها، برداری می‌شوند و سپس بردارهای ویژه استخراج می‌گردد. اما در این‌جا یک مشکل وجود دارد. به عنوان مثال، ابعاد برچسب‌های قطعه‌بندی قلب دارای ابعاد $144 \times 64 \times 160$ هستند که با برداری کردن این ابعاد به یک بردار با ابعاد 1×1474560 تبدیل می‌شود و با قرار دادن آن در فرمول ۱۵-۴ ماتریس کوواریانس C با ابعاد 1474560×1474560 بدست می‌آید که عملاً در حافظه‌ی رایانه‌های عادی، نمی‌توان آنرا ذخیره کرد. به این ترتیب با تغییر در فرمول‌بندی معادلات ۱۵-۴ می‌توان این مشکل را رفع کرد.

$$\frac{1}{n} \Theta^T \Theta \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\left(\frac{1}{n} \Theta^T \Theta \right) \Theta \mathbf{v}_i = \lambda_i (\Theta \mathbf{v}_i) \quad \mathbf{u}_i = \Theta \mathbf{v}_i \quad (19-4)$$

با ایجاد این تغییر به جای محاسبه $\Theta^T \Theta$ که Θ یک ماتریس $1474560 \times n$ است (n) تعداد دادگان سه‌بعدی مجموعه آموزش است)، $\Theta^T \Theta$ محاسبه می‌شود که خروجی آن یک ماتریس $n \times n$ است و قابلیت کنترل‌پذیری بالاتری دارد. در نهایت با استفاده از معادله ۱۹-۴ بردارهای ویژه \mathbf{u}_i موردنظر استخراج می‌گردند.

در نهایت با توصیفات ارایه شده برای برچسب‌های قطعه‌بندی دو مجموعه داده روند توضیح داده شده، اجرا گردید. شکل ۲۳-۴ بردارهای ویژه تغییر اندازه یافته به اندازه‌ی تصاویر اصلی را در سه نمای اکسیال، کرونال و سجیتال برای دو مجموعه داده‌ی قلب (تصویر راست) و هیپوکامپ (سمت چپ) برای ارگان‌های ویژه^{۳۵} اول و دوم و آخر نشان می‌دهد.



شکل ۴-۲۳: ارگان‌های ویژه استخراج شده برای (الف) هیپوکامپ و (ب) قلب در سه نمای اکسیال، کرونال و سجیتال

پس از استخراج ارگان‌های ویژه از تصاویر مجموعه‌ی آموزش، امکان استخراج بردار وزن‌های بازسازی طبق معادله ۱۶-۴ فراهم می‌شود. برای داده‌های غیر ارگان با همان ابعاد برچسب‌های اصلی، برچسب‌های خراب شده با افزایش^{۳۶} و کاهش^{۳۷} و ایجاد سوراخ^{۳۸} در برچسب‌های اصلی و

^{۳۵} در این مطالعه، بردارهای ویژه استخراج شده از ساختارهای بدن را ارگان ویژه می‌نامیم.

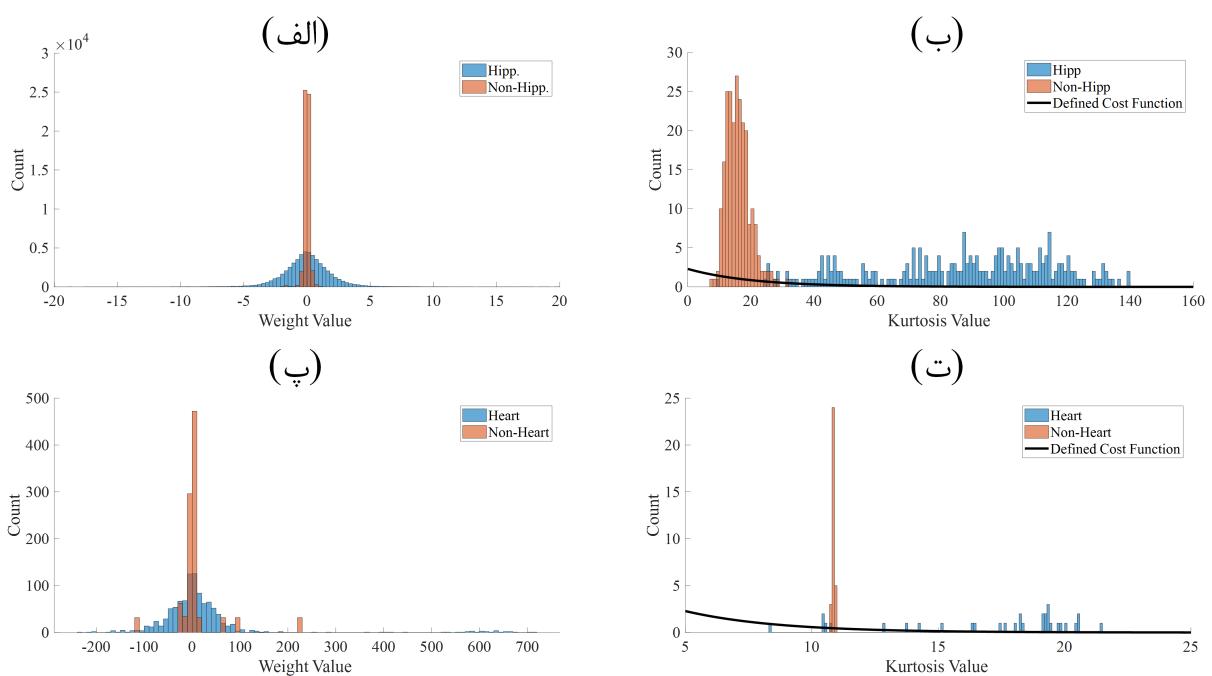
^{۳۶}Dilation

^{۳۷}Erosion

^{۳۸}Holes

نیز برچسب‌های نویزی، ایجاد گردید. در نهایت وزن‌های بازسازی برای داده‌های ارگان و غیر ارگان، استخراج شد و از این وزن‌ها طبق معادله ۱۸-۴ کورتوسیس گرفته شد.

شکل ۲۴-۴، هیستوگرام وزن‌های بازسازی و هیستوگرام کورتوسیس وزن‌های بازسازی برای دادگان هیپوکامپ و قلب را نشان می‌دهد. همانطور که مشاهده می‌شود، مقدار کورتوسیس برای داده‌ها با شکل‌های حقیقی و صحیح بیشتر است بنابراین با تعریف یکتابع هزینه برای جریمه‌ی شبکه، در ایجاد کورتوسیس وزن‌های بازسازی بالاتر، شکل‌های صحیح ایجاد می‌گردد و در این حالت شبکه قادر به یادگیری ویژگی‌های شکل و ساختار می‌شود.



شکل ۲۴-۴: (الف) هیستوگرام وزن‌های بازسازی هیپوکامپ و غیر هیپوکامپ، (ب) هیستوگرام کورتوسیس وزن‌های بازسازی هیپوکامپ و غیر هیپوکامپ (تابع هزینه با خط مشکی نشان داده شده است)، (پ) هیستوگرام وزن‌های بازسازی قلب و غیر قلب، (ت) هیستوگرام کورتوسیس وزن‌های بازسازی قلب و غیر قلب (تابع هزینه با خط مشکی نشان داده شده است). در تصاویر واضح است که مقدار کورتوسیس برای ساختارهای صحیح بالاتر از ساختارهای با شکل ناقص است بنابراین تابع هزینه‌ی تعریف شده، برای جریمه‌ی شبکه به ایجاد ممان چهارم بیشتر برای وزن‌های بازسازی است.

برای تعریف تابع هزینه، با توجه به هیستوگرام کورتوسیس‌ها از یکتابع نمایی کاوهنده استفاده گردید که در کورتوسیس‌های بالاتر مقدار کمتر و در کورتوسیس‌های پایین، مقدار بیشتری دارد. بنابراین تابع

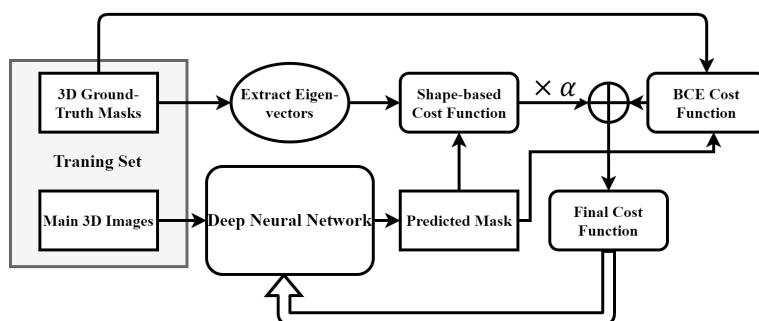
هزینه‌ی بر اساس شکل ۲۹ را می‌توان با فرمول زیر بیان نمود.

$$SBCF = \beta \exp\left(-\frac{x - \gamma m}{\delta \sigma}\right) \quad (20-4)$$

که در آن x مقدار کورتوسیس خروجی قطعه‌بندی شبکه، m و σ میانگین و انحراف معیار کورتوسیس مجموعه داده‌ی آموزش، β ، γ و δ پارامترهای تنظیم برای تابع نمایی کاهنده هستند. برای دو مجموعه داده‌ی قلب و هیپوکامپ، این تابع هزینه با منحنی مشکی در شکل ۲۴-۴ نشان داده شده است. در نهایت، تابع هزینه کلی برای آموزش مجموع وزن‌دار تابع هزینه BCE و SBCF است. که یکی در سطح واکسل‌ها شبکه را جریمه می‌کند و دیگری بر اساس شکل و ساختار تولیدی. معادله‌ی ۲۱-۴ تابع هزینه‌ی نهایی را نشان می‌دهد.

$$TotalLoss = \alpha SBCF(p) + BCE(p, y) \quad (21-4)$$

شکل ۲۵-۴ توصیف کلی از ساختار تابع هزینه‌ی پیشنهادی برای آموزش یک شبکه را نشان می‌دهد که از برچسب‌های قطعه‌بندی مجموعه دادگان آموزش، ارگان‌های ویژه استخراج می‌شوند و سپس بر اساس معادلات معرفی شده، کورتوسیس وزن‌های بازسازی پیش‌بینی شبکه محاسبه می‌شود و به تابع BCE هزینه‌ی بر اساس شکل داده می‌شود در نهایت مجموع وزن‌دار (با ضریب α) تابع هزینه‌ی شکل و وزن‌های شبکه را به روزرسانی می‌کند.

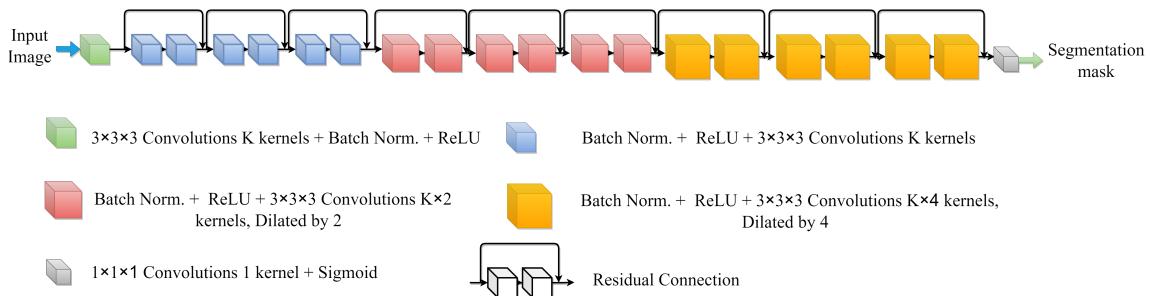


شکل ۲۵-۴: توصیف کلی عملکرد تابع هزینه‌ی پیشنهادی

³⁹Shape-Based Cost Function (SBCF)

آموزش شبکه‌ی سه‌بعدی قطعه‌بند برای یادگیری ویژگی‌های شکل

برای استفاده از تابع هزینه‌ی ارایه شده، لازم است یک شبکه‌ی کانولوشنی عمیق سه‌بعدی برای پیش‌بینی قطعه‌بندی به صورت حجمی، تعریف گردد. برای این شبکه‌ی سه‌بعدی از معماری پیشنهادی در [۳۰] استفاده گردید. که معماری آن در شکل ۲۶-۴ نشان داده شده است.



شکل ۴-۴: معماری شبکه‌ی کانولوشنی سه‌بعدی استفاده شده

در این معماری K تعداد ماتریس‌های ویژگی را نشان می‌دهد که برای آموزش با مجموعه دادگان هیپوکامپ این مقدار برابر با ۱۶ و برای مجموعه دادگان قلب به علت بزرگتر بودن ابعاد تصویر برابر با ۳ قرار داده شد (مقدارهای بیشتر باعث سرریز از حافظه و توقف آموزش می‌شد). ضریب α در تابع هزینه‌ی نهایی برای هیپوکامپ $0/1$ و برای قلب ۲ در نظر گرفته شد. برای بهینه‌سازی وزن‌های شبکه از الگوریتم بهینه‌سازی Adam استفاده شد و آموزش پس از رسیدن تابع هزینه به حالت ثابت متوقف گردید.

برای ارزیابی عملکرد تابع هزینه‌ی بر اساس شکل، شبکه یک مرتبه با تابع هزینه‌ی پیشنهادی در معادله‌ی ۲۱-۴ آموزش داده شد و بار دیگر تنها از تابع هزینه‌ی BCE استفاده گردید. در ادامه به بررسی نتایج بدست آمده پرداخته می‌شود.

نتایج

برای ارزیابی عملکرد، همانطور که اشاره گردید، با استفاده از داده‌های سه‌بعدی ام‌آرآی هیپوکامپ و سی‌تی اسکن قلب، شبکه‌ی تمام کانولوشنی سه‌بعدی با دو تابع هزینه‌ی مختلف آموزش دیده شد. جدول ۵-۴ معیارهای ارزیابی Dice، Jaccard و Hausdorff را برای مجموعه دادگان آزمایش نشان می‌دهد. برای مقایسه‌ی آماری بهبود عملکرد تابع هزینه‌ی پیشنهادی با تابع هزینه‌ی BCE از معیار P-

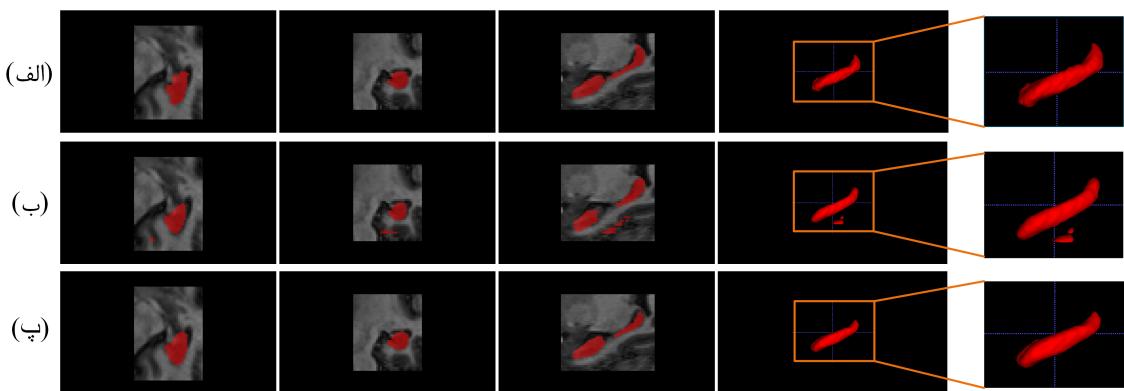
value (ردیف سوم در جدول) استفاده گردید که در صورت کمتر بودن این معیار از مقدار ۰/۰۵ رشد چشمگیری در روش ارایه شده نسبت به روش قبلی داشته‌ایم.

جدول ۴-۵: معیارهای ارزیابی برای مجموعه دادگان آزمایش برای قطعه‌بندی هیپوکامپ و قلب بر اساس رویکرد تابع هزینه‌ی بر اساس شکل و تابع هزینه‌ی BCE

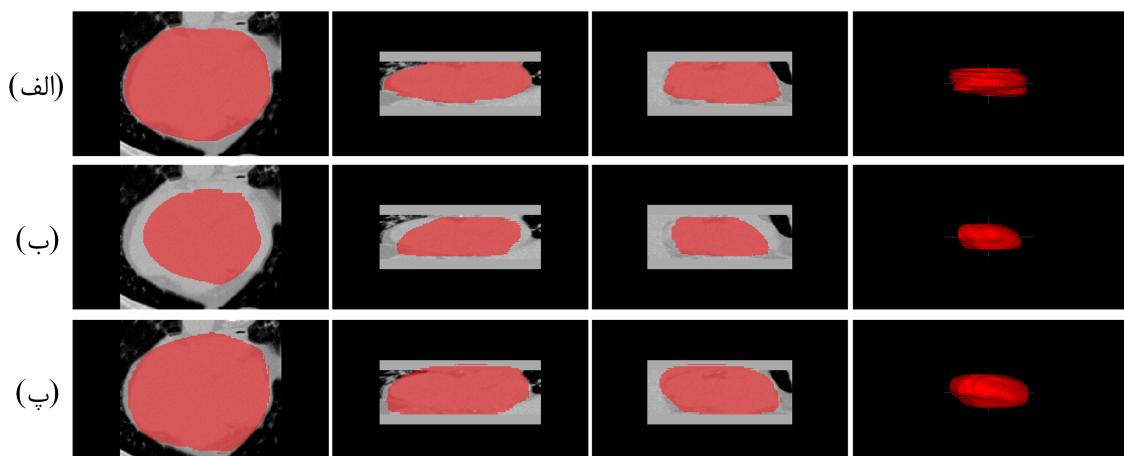
نام دادگان	روش	معیار	Hausdorff	Jaccard	Dice
قطعه‌بندی هیپوکامپ	BCE	$3/10 \pm 0/44$	$0/69 \pm 0/05$	$0/81 \pm 0/03$	
SBCF (روش ما)		$2/58 \pm 0/35$	$0/75 \pm 0/05$	$0/86 \pm 0/03$	
P-value		$0/000007$	$0/000003$	$0/000006$	
قطعه‌بندی قلب	BCE	$9/31 \pm 1/70$	$0/59 \pm 0/08$	$0/74 \pm 0/07$	
SBCF (روش ما)		$8/60 \pm 1/39$	$0/77 \pm 0/08$	$0/87 \pm 0/05$	
P-value		$0/33$	$0/0026$	$0/0032$	

شکل‌های ۲۷-۴ و ۲۸-۴ قطعه‌بندی یک نمونه از تصاویر آزمایش هیپوکامپ و قلب را به ترتیب نشان می‌دهند که با استفاده از دو مدل کانولوشنی سه‌بعدی آموزش دیده شده با تابع هزینه‌های مختلف، ایجاد شده‌اند. در هر دو تصویر، ردیف اول قطعه‌بندی مطلوب، ردیف دوم، خروجی شبکه‌ی قطعه‌بندی آموزش دیده شده بر اساس تابع هزینه‌ی BCE و ردیف سوم، خروجی شبکه‌ی قطعه‌بندی آموزش دیده شده بر اساس تابع هزینه‌ی پیشنهادی در نماهای اکسیال، کرونال، سجیتال و سه‌بعدی، نشان داده شده است.

با توجه در شکل‌ها خروجی و جدول فوق، علاوه بر افزایش دقت قطعه‌بندی، مشکلاتی از قبیل، تکه‌تکه شدن ساختار، جزیره‌های واکسلی و ناهمگونی شکل با استفاده کردن از تابع هزینه‌ی SBCF از بین رفته است. با دقت در تصویر ۲۷-۴ و در قسمت‌های بزرگ‌نمایی شده، می‌توان دریافت که استفاده از تابع هزینه‌ی BCE باعث بوجود آمدن جزیره‌های واکسلی و تکه‌تکه شدن ساختار می‌شود اما با استفاده کردن از تابع هزینه‌ی SBCF این مشکل رفع شده است و شکل و ساختار معتبری برای قطعه‌بندی هیپوکامپ در کنار افزایش دقت قطعه‌بندی، ایجاد گردیده است. به صورت مشابه در شکل ۲۸-۴ مشکل ناهمگونی شکل (ساختار پیش‌بینی شده برای قلب ساختار معتری نیست) با استفاده از تابع هزینه‌ی پیشنهادی رفع شده است.

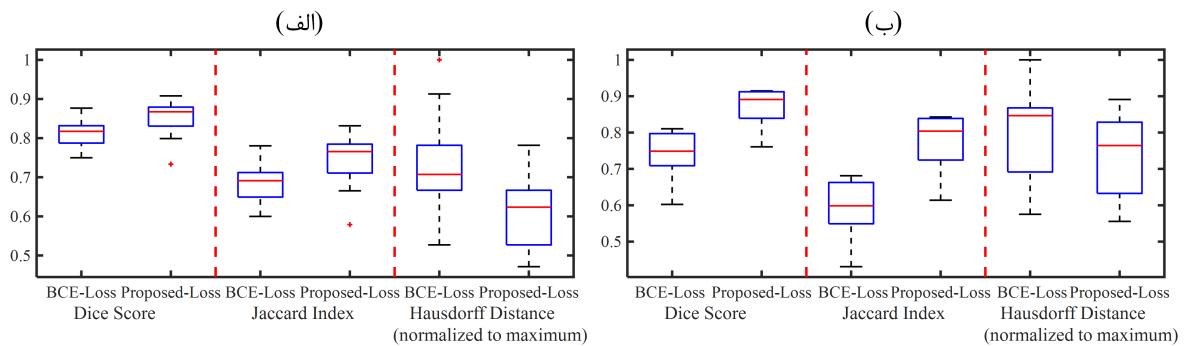


شکل ۴-۲۷: قطعه‌بندی تصاویر ام‌آرآی هیپوکامپ، (الف) قطعه‌بندی مطلوب (ب) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه BCE و (پ) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه SBCF در نماهای مختلف



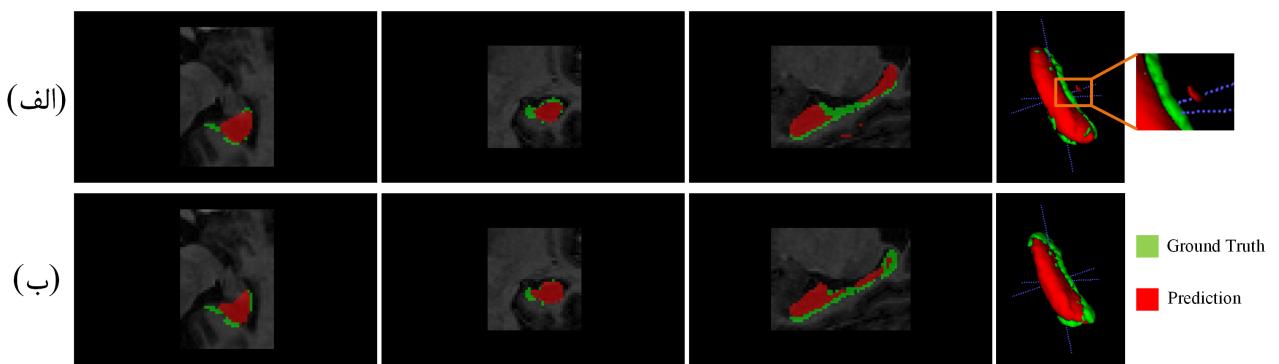
شکل ۴-۲۸: قطعه‌بندی تصاویر سی‌تی اسکن قلب، (الف) قطعه‌بندی مطلوب (ب) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه BCE و (پ) خروجی قطعه‌بندی شبکه‌ی آموزش دیده شده با تابع هزینه SBCF در نماهای مختلف

در شکل ۴-۲۹ نمودار جعبه‌ای معیارهای ارزیابی Hausdorff و Jaccard برای دو تابع هزینه استفاده شده در قطعه‌بندی هیپوکامپ (قسمت الف) و قطعه‌بندی قلب (قسمت ب) نشان داده شده است. برای نمایش بهتر، فاصله‌ی Hausdorff در نمودارهای جعبه‌ای به بیشینه‌ی خود تقسیم شد ۴/۲۴ برای هیپوکامپ و ۱۱/۶۶ برای قلب).



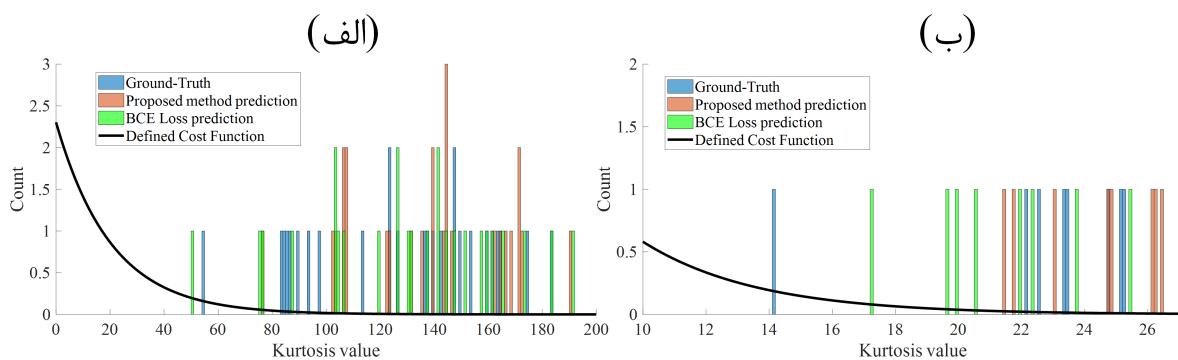
شکل ۴-۲۹: نمودار جعبه‌ای معیارهای ارزیابی Dice، Jaccard و Hausdorff برای (الف) قطعه‌بندی هیپوکامپ (ب) قطعه‌بندی قلب

همانطور که در جدول ۴-۵ و شکل ۴-۲۹، مشاهده شد، استفاده از تابع هزینه‌ی بر اساس شکل دقت قطعه‌بندی را افزایش می‌دهد. در نمودارهای جعبه‌ای، یک داده‌ی پرت مشاهده می‌شود که بر اساس تابع هزینه‌ی پیشنهادی بوجود آمده است. این مورد در تصویر ۴-۳۰ نشان داده شده است. در قسمت (الف) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی BCE (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز)، و در قسمت (ب) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی SBCF (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز) نشان داده شده است. مقدار Dice برای این مورد در تصویر (الف) برابر با 0.77 و در تصویر (ب) برابر با 0.73 است که فاصله‌ی نسبتاً کمی است. اما علی‌رغم بالاتر بودن این معیار، جزیره‌های واکسلی ناخواسته در قطعه‌بندی نهایی BCE وجود دارد (در تصویر (الف) بزرگنمایی شده است)، که این مشکل در تصویر (ب) مشاهده نمی‌شود.



شکل ۴-۳۰: داده‌ی پرت در قطعه‌بندی هیپوکامپ با استفاده از تابع هزینه‌ی بر اساس شکل، (الف) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی BCE (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز)، (ب) پیش‌بینی شبکه‌ی آموزش دیده شده با تابع هزینه‌ی SBCF (با رنگ قرمز) در کنار قطعه‌بندی مطلوب (رنگ سبز)

شکل ۴-۳۱ هیستوگرام کورتوسیس دادگان آزمایش در قطعه‌بندی‌های پیش‌بینی شده توسط شبکه‌های آموزش دیده شده با دو تابع هزینه‌ی مختلف برای (الف) هیپوکامپ و (ب) قلب، در کنار هیستوگرام کورتوسیس‌های مطلوب (رنگ آبی) نشان داده شده است. با توجه به تابع هزینه‌ی تعریف شده برای هر مجموعه داده (منحنی‌های مشکی رنگ)، شبکه‌ی آموزش دیده شده با تابع هزینه‌ی SBCF سعی در بیشینه کردن کورتوسیس‌ها و ایجاد شکل و ساختار معتر برای قطعه‌بندی مورد نظر داشته است.



شکل ۴-۳۱: هیستوگرام کورتوسیس پیش‌بینی شده توسط تابع هزینه‌ی BCE (رنگ سبز) و تابع هزینه‌ی SBCF (رنگ آجری) برای داده‌های آزمایش در کنار کورتوسیس‌های مطلوب (رنگ سبز).
 (الف) دادگان هیپوکامپ (ب) دادگان قلب

جمع‌بندی و نتیجه‌گیری

در این مطالعه دیدیم شبکه‌های عمیق، نقطه‌ی ضعفی در استخراج ویژگی‌های شکل و یادگیری ساختار دارند. تلاش‌هایی در این زمینه با ایجاد دانش اولیه برای شبکه و نیز پس‌پردازش‌ها برای پالایش خروجی و ایجاد ساختار معتبر شده بود اما هنوز مشکل یادگیری ساختار در این شبکه‌ها وجود داشت.

برای حل این مشکل، در این پژوهش از رویکرد ایجاد فضای شکل و استخراج ارگان‌های ویژه برای سنجش میزان معتبر بودن ساختار پیش‌بینی شده توسط شبکه، استفاده گردید. روش پیشنهادی استخراج وزن‌های بازسازی ماسک قطعه‌بندی، با توجه به ارگان‌های ویژه است. سپس با محاسبه کورتوسیس این وزن‌ها ملاحظه گردید ساختارهایی که شکل معتبرتری دارند کورتوسیس بالاتری ایجاد می‌شود. بر همین مبنای تابع هزینه‌ی بر اساس شکل پیشنهاد گردید.

در نهایت با آموزش یک شبکه با و بدون استفاده از این تابع هزینه بر روی دو مجموعه داده‌ی قطعه‌بندی ساختارهای سه‌بعدی در تصاویر سی‌تی اسکن و ام‌آرآی، ملاحظه گردید که نقص‌های شبکه در پیش‌بینی ساختار ناهمگون و ایجاد جزیره‌های واکسلی، برطرف گردید.

از این تابع هزینه، در مواردی که تغییرات ساختار مورد نظر کم باشد می‌توان استفاده کرد و نتایج قطعه‌بندی را بهبود بخشد اما برای قطعه‌بندی ساختارهایی که شکل و موقعیت گوناگونی در تصاویر مختلف دارند (مانند توموگرافی) استفاده از این رویکرد توصیه نمی‌شود.

۳-۴ روش پیشنهادی قطعه‌بندی تومور

قطعه‌بندی تومور در کنار قطعه‌بندی ساختارهای در ریسک، یک نیاز حیاتی است که لازم است با دقیق و سرعت بالایی انجام شود تا در روند تشخیص و درمان مورد استفاده قرار گیرد. قطعه‌بندی تومور نیز یک روند زمان‌بر و خسته‌کننده برای شخص متخصص است (به خصوص آنکه این قطعه‌بندی در طی درمان باید چندین مرتبه تکرار گردد). بنابراین خودکار کردن این روند در کلینیک‌های درمانی ضروری است.

قطعه‌بندی تومور به علت تغییرات اندازه، موقعیت و شکل آن، بسیار چالش برانگیزتر از قطعه‌بندی سایر ساختارهای بدن است که شکل و موقعیت خاصی دارند. بر همین اساس این قطعه‌بندی برای روش‌های خودکار نسبت به روش‌های دستی و نیمه-خودکار بسیار پرچالش‌تر است. روش‌های یادگیری عمیق نسبت به سایر روش‌های خودکار در قطعه‌بندی تومور دقت بالاتری نشان داده‌اند بنابراین ذکر ادامه

قصد داریم روشی مبتنی بر روش‌های یادگیری عمیق برای قطعه‌بندی خودکار تومور ارایه دهیم.

۴-۳-۱ معرفی دادگان و پیش‌پردازش

قطعه‌بندی تومور مغزی یک مرحله‌ی مهم، برای تشخیص و درمان سرطان مغزی است. تصاویر ام‌آرآی به علت ایجاد تضاد بالا در بافت‌های نرم و رزلوشن بالایی که برای تصویربرداری تامین می‌کنند، در تحلیل و بررسی ساختارهای مغزی، مطالعه‌ی تومور مغزی و نظارت و تهیه‌ی نقشه‌ی درمان، کاربرد فراوانی دارند. ام‌آرآی برخلاف روش تصویربرداری سی‌تی اسکن از پرتوهای یونیزه شده استفاده نمی‌کند و از این جهت خطرهای پرتوهای پر انرژی را ندارد.

در کاربردهای کلینیکی، برای تشخیص بهتر، تصویرهای ام‌آرآی مکمل سه‌بعدی دیگری از جمله، T1 Fluid Attenuation Inversion و T1-weighted، T2-weighted، contrastenhanced (T1ce) برای بهبود تضاد، بین بافت‌های نرم و همچنین تومور و پس زمینه، تهیه می‌شوند [۹۹].

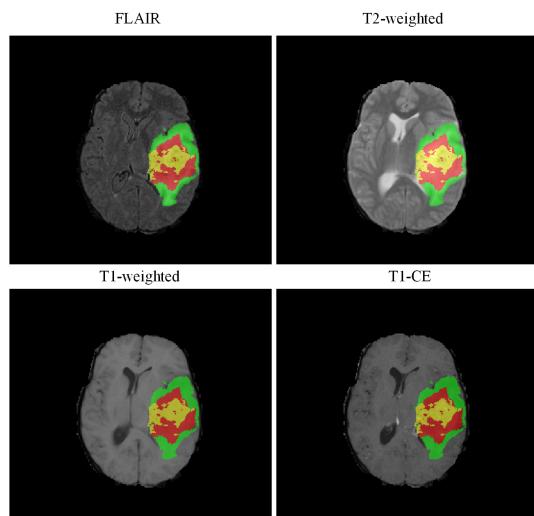
مجموعه دادگان BraTS^{۴۰} شامل چهار دنباله تصویر ام‌آرآی T1، T2، T1ce و FlAIR برای هر بیمار است که قطعه‌بندی مطلوب متناظر برای هر بیمار توسط متخصص قطعه‌بندی شده است. در هر قطعه‌بندی سه ناحیه برای تومور مشخص شده است که عبارتند از: هسته‌ی تومور^{۴۱}، ناحیه‌ی پیش توموری^{۴۲} و تومور در حال رشد^{۴۳}. این مجموعه دادگان از ۳۶۹ مورد تشکیل شده ازت که ابعاد تصویر ام‌آرآی در هر مورد $155 \times 240 \times 240$ است و فاصله واکسلی برابر با $1 \times 1 \times 1$ میلی‌متر مکعب است [۱۰۰، ۱۰۱]. شکل ۴-۲۲ چهار دنباله‌ی تصویر و قطعه‌بندی متناظر برای نواحی مختلف تومور را نشان می‌دهد.

⁴⁰Brain Tumor Segmentation Dataset

⁴¹Tumor Core

⁴²Peritumoral Edema

⁴³Enhancing Tumor



شکل ۴-۳۲: چهار دنباله تصویر ام آر آی T1، T2، T1ce و FLAIR برای یک بیمار خاص با قطعه‌بندی متناظر برای تومور (قرمز: هسته‌ی تومور، زرد: ناحیه‌ی پیش توموری و سبز: تومور در حال رشد)

در این مطالعه، هرچند استفاده از چهار دنباله‌ی تصویر، دقت قطعه‌بندی را افزایش می‌دهد اما تنها از یک دنباله‌ی تصویر FLAIR استفاده گردید. دلیل این کار به این علت است که در مراکز درمانی معمولاً یک دنباله‌ی تصویر در دسترس است و طبق مطالعات انجام شده دنباله‌ی FLAIR نسبت به سایر دنباله‌ها، تضاد بهتری بین تومور و بافت اطراف فراهم می‌کند [۱۰۲]. برای پیش‌پردازش این مجموعه تصاویر ابتدا کناره‌های تصویر به مرکزیت آن بریده شد و ابعاد به $155 \times 176 \times 144$ تغییر کرد. سپس مقدار شدت‌ها به مقدار ۹۰ درصد هیستوگرام تجمعی برای نرمال‌سازی صورت گرفت. برای قطعه‌بندی تومور سه برچسب با یکدیگر ترکیب گردید تا قطعه‌بندی تمام تومور انجام گیرد. در نهایت ده درصد از دادگان به عنوان مجموعه‌ی آزمایش و مابقی برای آموزش شبکه استفاده گردید.

۴-۳-۲ معرفی شبکه‌ی بر اساس ساز و کار توجه برای قطعه‌بندی تومور

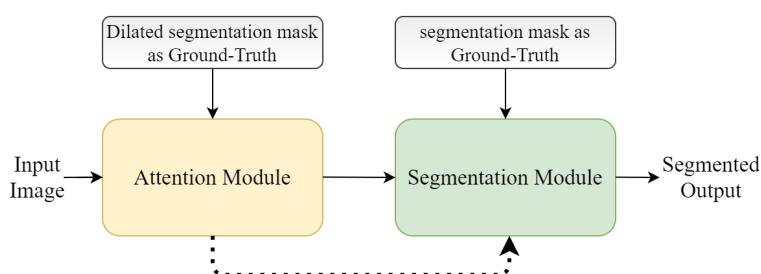
چالش اصلی قطعه‌بندی تومور در تصاویر سه بعدی پزشکی، تغییرات اندازه و موقعیت تومور و در برخی موارد تضاد پایین میان تومور و بافت‌های کناری است، که این قطعه‌بندی را دشوار می‌کند. روش پیشنهادی ما برای حل این مشکل استفاده از ساز و کار توجه^{۴۴} است که باعث می‌شود شبکه، به ناحیه‌هایی که در آن‌ها احتمال حضور تومور وجود دارد، توجه بیشتر کند. برای این منظور، دو مأذول توجه^{۴۵} و

⁴⁴Attention Mechanism

⁴⁵Attention Module

قطعه‌بندی^{۴۶} در این مطالعه ارایه گردید که هدف ماژول توجه یافتن موقعیت و ناحیه‌ای است که در آن تومور حضور دارد و سپس ماژول قطعه‌بند با استفاده از اطلاعات ماژول توجه، قطعه‌بندی تومور را به صورت دقیق انجام می‌دهد.

روش پیشنهادی ما، بر اساس معماری شبکه‌ی UNet دو بعدی، برای ماژول توجه و قطعه‌بند، پایه‌گذاری شده است. ایده‌ی اصلی این مطالعه در شکل ۴-۳۳ نشان داده شده است. این ساختار به صورت سریال عمل می‌کند. در ابتدا ماژول توجه یک قطعه‌بندی تقریبی با نرخ مثبت صحیح^{۴۷} بالا ایجاد می‌کند که برای آموزش این ماژول از ماسک‌های قطعه‌بندی افزایش^{۴۸} یافته استفاده می‌شود که وظیفه‌ی آن یافتن موقعیت تومور با یک تابع هزینه‌ی مناسب است. در گام بعدی ویژگی‌های استخراج شده توسط ماژول توجه، در کنار تصویر اصلی به ماژول قطعه‌بند داده می‌شود و قطعه‌بندی نهایی ایجاد می‌گردد که برای آموزش این ماژول از ماسک‌های قطعه‌بندی اصلی (بدون افزایش) استفاده می‌شود.



شکل ۴-۳۳: کلیت روش پیشنهادی برای قطعه‌بندی تومور بر اساس ساز و کار ماژول توجه و قطعه‌بند

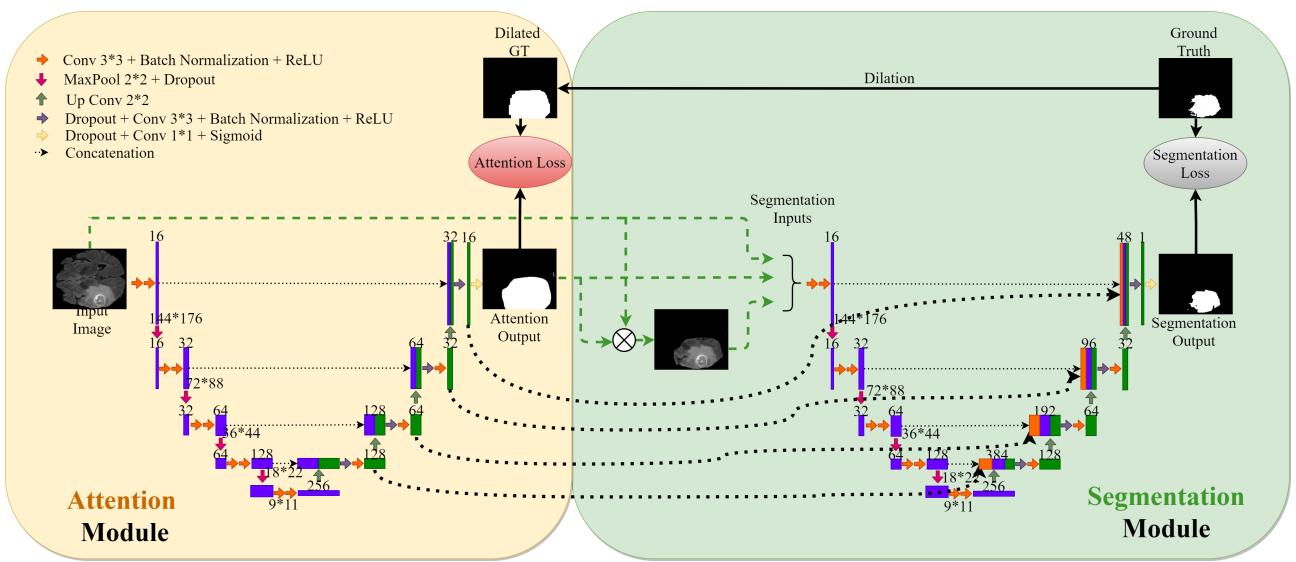
ماژول توجه

همانطور که اشاره گردید شبکه‌ی پایه‌ی مورد استفاده، هم برای ماژول توجه و هم برای ماژول قطعه‌بندی، شبکه‌ی UNet می‌باشد که با جزئیات روند آموزش در شکل ۴-۳۴ نشان داده شده است. هدف ماژول توجه ارایه‌ی یک قطعه‌بندی به صورت تقریبی و یافتن موقعیت تومور است که نرخ مثبت صحیح بالایی داشته باشد و کل تومور را در بر بگیرد. برای این منظور ماسک‌های قطعه‌بندی افزایش یافته به عنوان قطعه‌بندی مطلوب در روند آموزش این ماژول استفاده گردید. علاوه بر این برای اطمینان از بالا بودن نرخ مثبت صحیح، یک تابع هزینه‌ی جدید معرفی گردید که هدف آن بالا بردن این نرخ باشد.

⁴⁶Segmentation Module

⁴⁷True Positive Rate

⁴⁸Dilated



شکل ۴-۳۴: چارچوب پیشنهادی مارژول‌های توجه (قسمت نارنجی) و قطعه‌بند (قسمت سبز) برای قطعه‌بندی تومور به صورت دو بعدی. از معماری UNet به عنوان شبکه‌ی پایه استفاده گردید.

معادله‌ی اینتابع هزینه را می‌توان به صورت زیرنوشت که در آن TP ، مثبت صحیح، FN ، منفی غلط، TN ، منفی صحیح و FP ، مثبت غلط است.

$$\begin{aligned} SensitivityLoss &= 1 - \frac{TP}{TP + FN} \\ SpecificityLoss &= 1 - \frac{TN}{TN + FP} \\ AttentionLoss &= SensitivityLoss + \alpha \times SpecificityLoss \end{aligned} \quad (۲۲-۴)$$

وظیفه‌ی اینتابع هزینه بیشینه کردن مقدار TP است که با توجه به معادلات این کار در بخش انجام می‌شود اما اگر این قسمت به تنها یی استفاده شود خروجی نهایی یک تصویر تمام یک است. برای حل این مشکل و مصالحه میان مثبت صحیح (TP) و منفی صحیح (TN) یک قسمت دیگر (SpecificityLoss) اضافه گردید تا این مصالحه صورت پذیرد. در نهایت جمع وزن دار (با ضریب α) این دو قسمت تابع هزینه‌ی نهایی برای مارژول توجه را تشکیل می‌دهد.

ماژول قطعه‌بند

قسمت سبز رنگ در شکل ۴-۳۴، ماژول قطعه‌بند را نشان می‌دهد. در این قسمت برای آموزش از قطعه‌بندی‌های مطلوب بدون انجام هیچ پردازش (بدون افزایش)، استفاده گردید. برای تابع هزینه‌ی این قسمت از تابع BCE میان پیش‌بینی شبکه و خروجی مطلوب استفاده گردید.

به قسمت ورودی این ماژول سه تصویر دو بعدی داده می‌شوند که عبارتند از: الف) تصویر دو بعدی ام‌آرآی، ب) خروجی ماژول توجه و پ) ضرب دو مورد قبل در یکدیگر. این سه تصویر در کنار هم قرار می‌گیرند و به ورودی شبکه داده می‌شوند. علاوه بر این ورودی‌ها ویژگی‌های استخراج شده از قسمت Decoder ماژول توجه در کنار ویژگی‌های ماژول قطعه‌بند قرار داده می‌شوند و با اعمال کانولوشن Decoder ماژول قطعه‌بند، خروجی قطعه‌بندی نهایی ایجاد می‌گردد. این روند قسمت قطعه‌بند را مجبور به توجه بیشتر به نواحی دارای بافت تومور می‌کند.

جزیيات پیاده‌سازی

با توجه به شکل ۴-۳۴ دو شبکه‌ی مورد استفاده از معماری UNet تبعیت می‌کنند و دارای دو بخش Encoder و Decoder با لایه‌های کانولوشنی، ادغام و دیکانولوشن هستند. برای آموزش شبکه‌ها از مجموعه دادگان قطعه‌بندی تومور مغزی در تصاویر ام‌آرآی (BraTS) به صورت دو بعدی با پیش‌پردازش‌های توضیح داده شده در ابعاد 144×176 استفاده گردید.

برای آموزش ماژول توجه، ماسک‌های قطعه‌بندی آموزش، با ضریب ۲۱ افزایش داده شد. از تابع هزینه‌ی ۴-۲۲ با قرار دادن ضریب α برابر با $0/5$ برای تاکید بیشتر در ایجاد نرخ مثبت صحیح، در روند آموزش استفاده گردید. در نهایت این مدل آموزش دیده شد و به عنوان یک مدل از پیش‌آموزش داده شده در روند آموزش ماژول قطعه‌بندی استفاده شد.

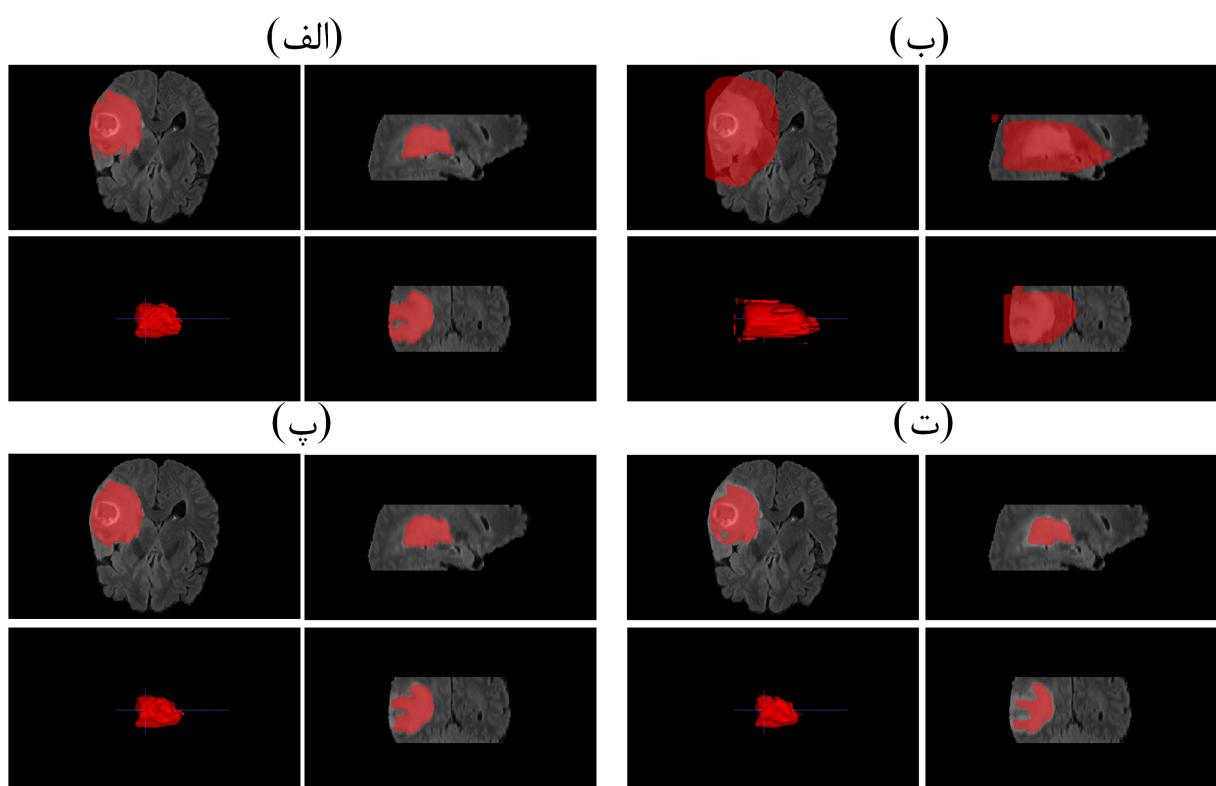
در طی آموزش ماژول قطعه‌بند، خروجی و ویژگی‌های تولید شده در ماژول توجه برای هر مورد استخراج گردید و به ماژول قطعه‌بند طبق شکل ۴-۳۴ داده شد. تابع هزینه‌ی این ماژول از BCE میان خروجی مطلوب و پیش‌بینی شبکه ایجاد گردید. برای بهینه‌سازی وزن‌ها از الگوریتم بهینه‌سازی Adam استفاده گردید. برای مقایسه‌ی نتایج، UNet استفاده شده در قسمت ماژول قطعه‌بندی، یکبار بدون استفاده از ماژول توجه و یکبار با استفاده از این ماژول آموزش داده شد. در ادامه به بررسی نتایج پرداخته می‌شود.

نتایج

برای ارزیابی عملکرد مژول توجه، همانطور که اشاره شد، هدف بیشینه کردن نرخ صحیح مثبت است که طبق معادله‌ی زیر این معیار محاسبه می‌شود.

$$TPR = \frac{TP}{TP + FN} \quad (23-4)$$

این معیار ارزیابی برای مجموعه دادگان آزمایش برابر با 0.998 ± 0.003 بدست آمد که نشان می‌دهد موقعیت تومور به طور صحیح و تقریباً کامل پیدا شده است. شکل ۴-۳۵ قسمت (ب) خروجی مژول توجه را نشان می‌دهد که در مقایسه با قطعه‌بندی مطلوب (قسمت (الف)), تومور را به صورت نادقيق اما با نرخ مثبت صحیح بالا قطعه‌بندی نموده است.

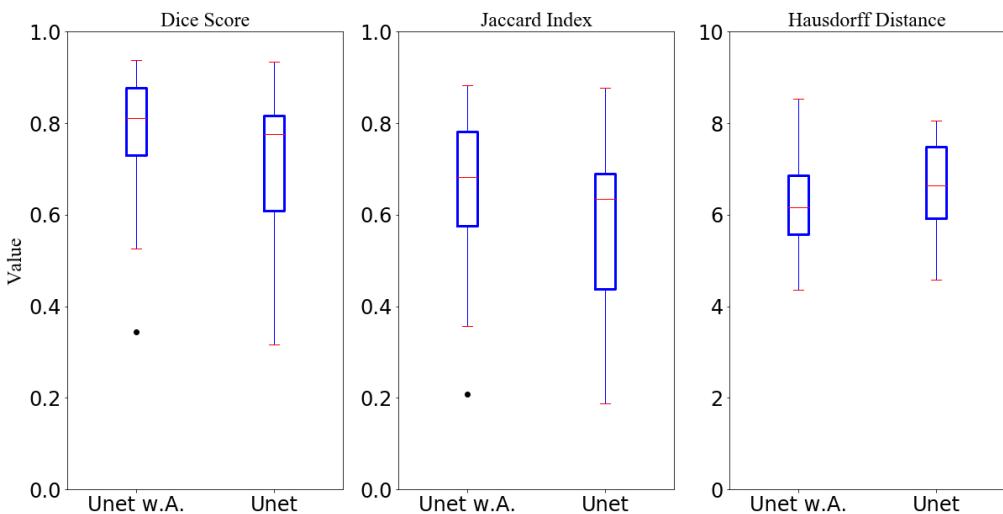


شکل ۴-۳۵: قطعه‌بندی تومور در نماهای مختلف برای یک مورد از دادگان آزمایش، (الف) قطعه‌بندی مطلوب، (ب) خروجی مژول توجه، (پ) خروجی مژول قطعه‌بند بر اساس ساز و کار توجه و (ت) خروجی UNet بدون استفاده از مژول توجه

برای ارزیابی عملکرد ماثول قطعه‌بند از معیارهای Hausdorff و Jaccard Dice استفاده گردید. این معیارها برای دو شبکه‌ی Unet آموزش دیده شده به صورت تکی و با استفاده از ماثول توجه، استخراج گردید. با توجه به جدول ۴-۶ میانگین معیار Dice با استفاده کردن از ماثول توجه از ۰/۶۸ به ۰/۷۹ افزایش یافته است و انحراف معیار آن از ۰/۱۸ به ۰/۱۲ کاهش یافت. برای مقایسه‌ی معنادار بودن این بهبود از معیار P-value استفاده گردید که در سطر آخر جدول نشان داده شده است. کمتر بودن این مقدار از ۰/۰۵ نشانگر پیشرفت چشمگیر است که نشان می‌دهد استفاده از ماثول توجه باعث افزایش دقت قطعه‌بندی می‌شود. شکل ۴-۶ نمودارهای جعبه‌ای معیارهای ارزیابی را برای دو شبکه نشان می‌دهد. واضح است میانگین و انحراف معیارهای این معیارها با به کارگیری ماثول توجه بهبود می‌یابد.

جدول ۴-۶: معیارهای ارزیابی برای مجموعه دادگان آزمایش برای قطعه‌بندی تومور در دو شبکه‌ی Unet بر اساس ماثول توجه و Unet تنها

نام معیار	نام شبکه		
Hausdorff	Jaccard	Dice	
۰/۶ ± ۱/۰	۰/۵۷ ± ۰/۱۷	۰/۶۸ ± ۰/۱۸	Unet تنها
۰/۲ ± ۱/۰	۰/۶۹ ± ۰/۱۵	۰/۷۹ ± ۰/۱۲	Unet با ماثول توجه
۰/۰۶	۰/۰۲	۰/۰۳	P-value



شکل ۴-۳۶: نمودار جعبه‌ای معیارهای ارزیابی Dice، Jaccard و Hausdorff برای مجموعه دادگان آزمایش در دو شبکه Unet بر اساس مازول توجه (Unet WA) و Unet تنها

نتیجه‌گیری

در این مطالعه، یک مازول توجه برای قطعه‌بندی تومور مغزی ارایه گردید. در این مازول یکتابع هزینه‌ی جدید برای بیشینه کردن نرخ صحیح مثبت پیشنهاد شد و برای آموزش آن از قطعه‌بندی‌های افزایش (Dilated) داده شده استفاده گردید. این مازول، در واقع، یک قطعه‌بندی نادقيق اما در برگیرنده‌ی تمام تومور پیش‌بینی می‌کند و می‌تواند موقعیت تومور را با دقیق بسیار بالایی استخراج نماید.

در نهایت برای ارزیابی این مازول توجه، یک UNet یک مرتبه با بهکار گیری این مازول و بار دیگر به صورت تکی بر روی دادگان یکسان، آموزش دیده شد و با استخراج معیارهای ارزیابی، ملاحظه گردید، با بهکار گیری این مازول دقیق چشمگیری در قطعه‌بندی نهايی خواهیم داشت.

۴-۴ جمع‌بندی

در این فصل، به معرفی روش‌های پیشنهادی و نتایج بدست آمده پرداخته شد. در ابتدا سه روش برای قطعه‌بندی ساختارهای در ریسک معرفی و ارزیابی شد که در روش اول هدف انتقال دانش از یک مدل پیچیده‌تر به یک مدل ساده‌تر بود که در عین ثابت نگه داشتن پیچیدگی مدل ساده‌تر، دقت آنرا افزایش می‌دهد. در روش دوم، یک چارچوب بر اساس بازخورد خطای پیش‌بین معرفی گردید که با بازسازی ورودی در یک قسمت، پیش‌بینی قطعه‌بندی در قسمت دیگر و بازخورد دادن خطای بازسازی به ورودی مسیر قطعه‌بندی، قطعه‌بندی ساختارهای در ریسک انجام می‌شد. در روش سوم، یک تابع هزینه‌ی بر اساس شکل پیشنهاد گردید که هدف آن استفاده از ویژگی‌های ساختاری اعضای بدن برای قطعه‌بندی است. با استفاده کردن از این تابع هزینه، مشکلات مربوط به عدم یادگیری شکل در شبکه‌های یادگیری عمیق پاسخ داده شد.

در قسمت نهایی به معرفی یک روش برای قطعه‌بندی پرچالش تومور در تصاویر ام‌آرآی مغزی پرداخته شد. قطعه‌بندی تومور به علت داشتن اندازه، شکل و موقعیت متغیر، نسبت به قطعه‌بندی ساختارهای در ریسک، روندی چالش برانگیزتر است. رویکرد پیشنهادی برای غلبه بر این چالش‌ها، استفاده از ساز و کار توجه برای یافتن موقعیت تومور در یک مازول توجه بود و در نهایت با راهنمایی‌های مازول توجه، مازول قطعه‌بند، قطعه‌بندی دقیق نهایی را انجام می‌داد.

فصل ۵

نتیجه‌گیری و کارهای آینده

در این پایان‌نامه بر روی قطعه‌بندی ساختارهای در ریسک و تومور با استفاده از روش‌های یادگیری ماشین و یادگیری عمیق مطالعه صورت گرفت. پیش از ظهور یادگیری عمیق روش‌های بسیار متنوعی برای قطعه‌بندی تصاویر طبیعی و پزشکی ارائه شده بود که در فصل سوم این روش‌ها معرفی شدند. اما با پیدایش روش‌های یادگیری عمیق و با به کارگیری آن‌ها در قطعه‌بندی تصاویر پزشکی، دقت قطعه‌بندی نسبت به روش‌های پیشین افزایش یافت و تقریباً تمام روش‌های کلاسیک جای خود را به روش‌های مبتنی بر هوش مصنوعی دادند.

در این مطالعه، پس از معرفی ابزارهای موجود در یادگیری عمیق در فصل دوم، اقدام به پیشنهاد روش‌هایی برای قطعه‌بندی ساختارهای در ریسک و تومور گردید که قطعه‌بندی این ساختارها یک گام ضروری و اساسی در پرتو درمانی برای تهیه نقشه‌ی دوز تابشی به بدن بیمار است و جلوی آسیب به ساختارهای سالم اطراف تومور (ساختارهای در ریسک) را کمینه می‌کند و میزان تابش به بافت تومور را در حالت بیشینه قرار می‌دهد.

در این پایان‌نامه، چهار روش قطعه‌بندی بر اساس ابزارهای یادگیری عمیق پیشنهاد گردید که در سه روش اول به قطعه‌بندی ساختارهای در ریسک پرداخته شد و در روش آخر قطعه‌بندی تومور مورد مطالعه قرار گرفت. در ادامه روش‌های ارایه شده به طور اجمالی مرور خواهد شد.

- قطعه‌بندی ساختارهای در ریسک با چارچوب چگالش دانش: هدف در چارچوب چگالش دانش، انتقال دانش از یک مدل بسیار پیچیده (آموزگار) که دقت بالایی در کنار هزینه‌های محاسباتی

بالا به یک مدل ساده‌تر (دانش‌آموز) است که بدون افزایش هزینه‌های محاسباتی، دقتی مشابه با آموزگار پیدا کند. بنابراین با توجه به این چارچوب، یک مدل بسیار پیچیده، با تعداد پارامتر و هزینه‌ی محاسباتی بالا، بر روی دادگان آموزش قطعه‌بندی ساختارهای در ریسک آموزش داده شد. سپس دو مدل ضعیفتر که به عنوان دانش‌آموز معرفی شدند، پیاده‌سازی و یک مرتبه بدون چارچوب چگالش دانش و یک مرتبه با چگالش دانش از آموزگار، آموزش دیدند. ملاحظه گردید که با استفاده کردن از این چارچوب بدون افزودن به پیچیدگی مدل و هزینه‌های محاسباتی، دقت آن را افزایش داد.

- **قطعه‌بندی ساختارهای در ریسک با چارچوب بازخورد خطای پیش‌بین:** این روش بر اساس مدل کدگذار پیش‌بین برای مدل‌سازی عملکرد نورون‌های مغزی در تعامل با یکدیگر، پایه‌گذاری شد. در مدل کدگذار پیش‌بین، نورون‌های لایه‌های بالایی مغز، سعی در پیش‌بینی خروجی نورون‌های سطح پایینی مغزی را دارند و با ارتباط‌های بازخوردی، مقادیر پیش‌بینی شده از لایه‌های بالاتر به لایه‌های پایین‌تر عبور داده می‌شوند. بر همین اساس یک چارچوب پیشنهاد گردید که از یک Encoder و دو Decoder تشکیل شده بود. Decoder اول بازسازی تصویر ورودی و Decoder دوم، قطعه‌بندی ساختارهای در ریسک را انجام می‌دهند. بنابراین با بازخورد مجموع تصویر بازسازی شده و خطای بازسازی به مسیر قطعه‌بندی، نواحی‌ای که اطلاعات ریزتری دارند و در قسمت بازسازی، به صورت دقیق بازسازی نشده‌اند، تاکید بیشتری می‌شود. دیگر مزیت این روش، استفاده از یک Encoder مشترک در بازسازی و قطعه‌بندی است که باعث استخراج ویژگی‌های سطح بالاتر در شبکه می‌شود. استفاده از این چارچوب، به دلایل گفته شده، باعث افزایش دقت در قطعه‌بندی ساختارهای در ریسک گردید.

- **قطعه‌بندی ساختارهای در ریسک به صورت سه‌بعدی با پیشنهاد یکتابع هزینه بر اساس شکل:** در این مطالعه سعی شد، مساله‌ی عدم توانایی شبکه‌های عصبی عمیق در استخراج ویژگی‌های مربوط به شکل مورد مطالعه قرار گیرد. رویکرد ما در این مطالعه ایجاد یک فضای شکل بر اساس برچسب‌های قطعه‌بندی و استخراج بردارهای ویژه بود. سپس با توجه به وزن‌های بازسازی میان شکل‌های معتبر و نامعتبر تمایز قابل شد و در نهایت بر این اساس یکتابع هزینه برای اجبار شبکه به تولید شکل‌های معتبر و جلوگیری از ایجاد جزیره‌های واکسلی، تکه‌تکه شدن ساختار و وجود سوراخ در قطعه‌بندی ساختار یکپارچه، می‌توان جلوگیری کرد. اکثر ساختارهای بدن در افراد مختلف معمولاً یک شکل دارند و می‌توان با آموزش شبکه با این تابع هزینه، از ایجاد شکل‌های

نامعتبر در قطعه‌بندی ساختارهای در ریسک جلوگیری کرد و دقت قطعه‌بندی را بهبود داد.

- قطعه‌بندی تومور با استفاده از سازوکار توجه در شبکه‌های عمیق: قطعه‌بندی تومور در روش‌های قطعه‌بندی خودکار به علت متغیر بودن شکل، موقعیت و اندازه‌ی تومور بسیار چالش برانگیز است. بنابراین برای حل این مشکلات در شبکه‌های عصبی عمیق بهتر است از سازوکار توجه استفاده گردد تا توجه شبکه به نواحی‌ای که در آنها احتمال حضور تومور وجود دارد جلب گردد. برای این‌کار در این مطالعه، از دو مازول توجه و قطعه‌بند استفاده گردید. در مازول توجه، هدف یافتن موقعت تومور به نحوی که نرخ مثبت صحیح بالایی داشته باشد و کل تومور را پیدا کند. برای این‌کار یک تابع هزینه‌ی جدید معرفی گردید که این خواسته را به خوبی برآورده نمود. سپس خروجی مازول توجه و ویژگی‌های استخراج شده توسط این مازول به مازول قطعه‌بند داده می‌شود که این‌کار باعث اجبار شبکه به توجه نمودن به نواحی شامل تومور خواهد شد. در نهایت ارزیابی‌ها نشان داد استفاده از این مازول توجه دقت قطعه‌بندی را به صورت چشمگیر افزایش می‌دهد.

با وجود روش‌های خودکار ارایه شده بر اساس یادگیری عمیق برای قطعه‌بندی تصاویر پزشکی و علی‌رغم افزایش سرعت و دقت نسبت به روش‌های کلاسیک و دستی (به عنوان مثال برای قطعه‌بندی تصویر سی‌تی اسکن سر و گردن یک بیمار توسط متخصص حدوداً ۳ ساعت زمان نیاز است که در طول درمان با تغییر شکل تومور و ساختارهای اطراف چندین بار باید تکرار شود [۴۱، ۴۲]) اما ۶۸ درصد متخصصین هنوز به این روش‌ها اعتماد ندارند و قطعه‌بندی به صورت دستی را ترجیح می‌دهند [۴۲]. بنابراین یک مرحله‌ی رابط کاربری به عنوان نظارت متخصص در قطعه‌بندی پس از قطعه‌بندی به صورت اتوماتیک لازم است تا اعتمادسازی صورت گیرد. برای این‌کار نرم‌افزارهای زیادی وجود دارد که می‌توان پیش‌بینی قطعه‌بندی توسط شبکه را در آن وارد نمود و توسط متخصص، پس‌پردازش‌های لازم انجام شود. از جمله‌ی این نرم‌افزارها می‌توان به 3D Slicer، Osirix، ITK Library ... اشاره نمود.

پیشنهادهایی که برای کارهای آینده در قطعه‌بندی تومور و ساختارهای در ریسک می‌توان داد به شرح زیر است:

- افزودن الگوریتم‌های یادگیری فعال^۱ برای دریافت بازخورد از متخصص و بهبود دقت مدل قطعه‌بند در هربار استفاده.

^۱ Active Learning

- استفاده از روش‌های Few Shot Learning به علت کمبودن داده‌های قطعه‌بندی پزشکی می‌تواند در بهبود دقت و سرعت یادگیر مفید باشد.
- تعمیم تابع هزینه‌ی پیشنهاد شده بر اساس شکل، در این مطالعه، برای قطعه‌بندی و در نظر گرفتن چند ساختار به صورت همزمان.
- ایجاد یک رابط کاربری و قرار دادن الگوریتم‌های مختلف قطعه‌بندی تومور و ساختارهای در ریسک در کنار یکدیگر و قطعه‌بندی همزمان چندین ساختار و تومور به صورت همزمان.

آپیوست

مقالات ارایه شده در این مطالعه

ماحصل این پایاننامه یک مقاله‌ی بین‌المللی که در حال ارسال آن به مجله هستیم و سه مقاله‌ی ارایه شده‌ی کنفرانسی (یک کنفرانس داخلی (28th National and 6th International Iranian Conference on Nuclear Science) و دو کنفرانس خارجی (The 2021 Conference on Biomedical Engineering Symposium (NSS) and Medical Imaging Conference (MIC) زیر قابل دسترس است.

Journal Paper: A novel shape-based loss function for machine learning-based seminal organ segmentation in medical imaging

ICBME paper: Attention-based deep learning segmentation: Application to brain tumor delineation

MIC 1st paper: Knowledge Distillation: A Strategy to Enhance the Performance of Deep Learning-based Seminal Segmentation

MIC 2nd paper: Prediction Error Propagation: A Novel Strategy to Enhance Performance of Deep Learning Models in Seminal Segmentation

علی‌رغم پذیرش و ارایه‌ی مقالات در کنفرانس MIC ، در حال حاضر این مقالات توسط کنفرانس چاپ نشده است و در اختیار عموم قرار نگرفته است.

مراجع

- [1] Structseg2019 - grand challenge, <https://structseg2019.grand-challenge.org/dataset/>. Accessed: 2022-01-15.
- [2] Pgi gets state-of-the-art ct scan machine, <https://www.tribuneindia.com/news/punjab/pgi-gets-state-of-the-art-ct-scan-machine-179634>. Accessed: 2022-01-21.
- [3] Ct of a normal abdomen and pelvis, thumbnail, <https://commons.wikimedia.org/wiki/>. Accessed: 2022-01-21.
- [4] Magnetic resonance imaging (mri), <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>. Accessed: 2022-01-22.
- [5] How to use data science to understand what makes wine taste good, <https://www.freecodecamp.org/news/using-data-science-to-understand-what-makes-wine-taste-good-669b496c67ee>. Accessed: 2022-03-29.
- [6] B. M. Njogholo. *Investigating keystroke dynamics as a two-factor biometric security*. PhD thesis, Strathmore University, 2018.
- [7] J. M. Johnson and T. M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54, 2019.
- [8] L. Cai, Y. Chen, N. Cai, W. Cheng, and H. Wang. Utilizing amari-alpha divergence to stabilize the training of generative adversarial networks. *Entropy*, 22(4):410, 2020.
- [9] Convolutional neural networks, <https://cs231n.github.io/convolutional-networks/convert>. Accessed: 2022-04-02.

- [10] A comprehensive guide to convolutional neural networks — the eli5 way, <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. Accessed: 2022-04-03.
- [11] Activation functions, <https://www.datacamp.com/community/tutorials/neural-network-models-r>. Accessed: 2022-04-03.
- [12] R. Gençay and M. Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12(4):726–734, 2001.
- [13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [14] U. Vovk, F. Pernus, and B. Likar. A review of methods for correction of intensity inhomogeneity in mri. *IEEE transactions on medical imaging*, 26(3):405–421, 2007.
- [15] Metrics to evaluate your semantic segmentation model, <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2>. Accessed: 2022-04-12.
- [16] Understand jaccard index, jaccard similarity in minutes, <https://medium.com/data-science-bootcamp/understand-jaccard-index-jaccard-similarity-in-minutes-25a703fbf9d7>. Accessed: 2022-04-12.
- [17] Hausdorff distance, <https://en.wikipedia.org/wiki/hausdorffdistance>. Accessed: 2022-04-12.
- [18] H. Ng, S. Ong, K. Foong, P.-S. Goh, and W. Nowinski. Medical image segmentation using k-means clustering and improved watershed algorithm. In *2006 IEEE southwest symposium on image analysis and interpretation*, pages 61–65. IEEE, 2006.
- [19] S. Madhukumar and N. Santhiyakumari. Evaluation of k-means and fuzzy c-means segmentation on mr images of brain. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(2):475–479, 2015.

- [20] slides of medical image analysis and processing (miap) course, segmentation chapter, tutor: Emad fatemizadeh, sharif university of technology,. Accessed: 2020.
- [21] Level set, <https://en.wikipedia.org/wiki/level-setmethod>. Accessed: 2022-04-18.
- [22] S. Bourouis, R. Alroobaea, S. Rubaiee, and A. Ahmed. Toward effective medical image analysis using hybrid approaches—review, challenges and applications. *Information*, 11(3):155, 2020.
- [23] Z. Cui, J. Yang, and Y. Qiao. Brain mri segmentation with patch-based cnn approach. In *2016 35th Chinese Control Conference (CCC)*, pages 7026–7031. IEEE, 2016.
- [24] B. Ibragimov and L. Xing. Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks. *Medical physics*, 44(2):547–557, 2017.
- [25] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [26] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015.
- [27] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [28] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [29] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

- [30] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In *International conference on information processing in medical imaging*, pages 348–360. Springer, 2017.
- [31] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.
- [32] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
- [33] M.-T. Tran, S.-H. Kim, H.-J. Yang, and G.-S. Lee. Lung tumor segmentation by fusing 2d and 3d models. Technical report, EasyChair, 2020.
- [34] Z. Lambert, C. Petitjean, B. Dubray, and S. Kuan. Segthor: Segmentation of thoracic organs at risk in ct images. In *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2020.
- [35] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.
- [36] X. Liu, X. Wang, and S. Matwin. Improving the interpretability of deep neural networks with knowledge distillation. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 905–912. IEEE, 2018.
- [37] J. Gou, B. Yu, S. J. Maybank, and D. Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.
- [38] P. NAGPAL. Basics of oncology techniques for undergraduates techniques of radiotherapy. *National Medicos Organisation Journal (ISSN-2348-3806)*, 13(01):54–59, 2019.
- [39] A. Ramkumar, J. Dolz, H. A. Kirisli, S. Adebahr, T. Schimek-Jasch, U. Nestle, L. Massoptier, E. Varga, P. J. Stappers, W. J. Niessen, et al. User interaction in semi-automatic segmentation of organs at risk: a case study in radiotherapy. *Journal of digital imaging*, 29(2):264–277, 2016.

- [40] C. Wang, X. Zhu, J. C. Hong, and D. Zheng. Artificial intelligence in radiotherapy treatment planning: present and future. *Technology in cancer research & treatment*, 18:1533033819873922, 2019.
- [41] A. A. Qazi, V. Pekar, J. Kim, J. Xie, S. L. Breen, and D. A. Jaffray. Auto-segmentation of normal and target structures in head and neck ct images: A feature-driven model-based approach. *Medical physics*, 38(11):6160–6170, 2011.
- [42] L. J. Stapleford, J. D. Lawson, C. Perkins, S. Edelman, L. Davis, M. W. McDonald, A. Waller, E. Schreibmann, and T. Fox. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 77(3):959–966, 2010.
- [43] F. Lateef and Y. Ruichek. Survey on semantic segmentation using deep learning techniques. *Neurocomputing*, 338:321–348, 2019.
- [44] B. Kayalibay, G. Jensen, and P. van der Smagt. Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*, 2017.
- [45] H. Arabi and H. Zaidi. Comparison of atlas-based techniques for whole-body bone segmentation. *Medical image analysis*, 36:98–112, 2017.
- [46] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, and Q. Wang. Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images. *Medical physics*, 45(5):2063–2075, 2018.
- [47] A. Elangovan and T. Jeyaseelan. Medical imaging modalities: a survey. In *2016 International Conference on emerging trends in engineering, technology and science (ICETETS)*, pages 1–4. ieee, 2016.
- [48] Computed tomography (ct), <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>. Accessed: 2022-01-21.
- [49] J. M. Helm, A. M. Swiergosz, H. S. Haeberle, J. M. Karnuta, J. L. Schaffer, V. E. Krebs, A. I. Spitzer, and P. N. Ramkumar. Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13(1):69–76, 2020.

- [50] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500, 1952.
- [51] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [52] Notes on artificial intelligence, machine learning and deep learning for curious people, <https://towardsdatascience.com/notes-on-artificial-intelligence-ai-machine-learning-ml-and-deep-learning-dl-for-56e51a2071c2>. Accessed: 2022-03-28.
- [53] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [54] W. Wang, Y. Yang, X. Wang, W. Wang, and J. Li. Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*, 58(4):040901, 2019.
- [55] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.
- [56] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.
- [57] Difference between image processing and computer vision, <https://www.geeksforgeeks.org/difference-between-image-processing-and-computer-vision/>: :text=image Accessed: 2022-04-05.
- [58] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(1):315–337, 2000.
- [59] T. Zuva, O. O. Olugbara, S. O. Ojo, and S. M. Ngwira. Image segmentation, available techniques, developments and open issues. *Canadian Journal on Image Processing and Computer Vision*, 2(3):20–29, 2011.

- [60] N. Sharma and L. M. Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.
- [61] V. Shrimali, R. Anand, and V. Kumar. Current trends in segmentation of medical ultrasound b-mode images: a review. *IETE technical review*, 26(1):8–17, 2009.
- [62] S. S. Al-Amri, N. V. Kalyankar, and S. D. Khamitkar. A comparative study of removal noise from remote sensing image. *arXiv preprint arXiv:1002.1148*, 2010.
- [63] J. Tohka. Partial volume effect modeling for segmentation and tissue classification of brain magnetic resonance images: A review. *World journal of radiology*, 6(11):855, 2014.
- [64] N. Makris, J. Kaiser, C. Haselgrove, L. J. Seidman, J. Biederman, D. Boriel, E. M. Valera, G. M. Papadimitriou, B. Fischl, V. S. Caviness Jr, et al. Human cerebral cortex: a system for the integration of volume-and surface-based representations. *Neuroimage*, 33(1):139–153, 2006.
- [65] P. Sinha, M. Tuteja, and S. Saxena. Medical image segmentation: hard and soft computing approaches. *SN Applied Sciences*, 2(2):1–8, 2020.
- [66] E. Debreuve, M. Gastaud, M. Barlaud, and G. Aubert. Using the shape gradient for active contour segmentation: from the continuous to the discrete formulation. *Journal of Mathematical Imaging and Vision*, 28(1):47–66, 2007.
- [67] N. Ironside, C.-J. Chen, S. Mutasa, J. L. Sim, D. Ding, S. Marfatia, D. Roh, S. Mukherjee, K. C. Johnston, A. M. Southerland, et al. Fully automated segmentation algorithm for perihematomal edema volumetry after spontaneous intracerebral hemorrhage. *Stroke*, 51(3):815–823, 2020.
- [68] B. Alfano, M. Comerci, M. Larobina, A. Prinster, J. P. Hornak, S. E. Selvan, U. Amato, M. Quarantelli, G. Tedeschi, A. Brunetti, et al. An mri digital brain phantom for validation of segmentation methods. *Medical image analysis*, 15(3):329–339, 2011.

- [69] Z.-K. Huang and K.-W. Chau. A new image thresholding method based on gaussian mixture model. *Applied mathematics and computation*, 205(2):899–907, 2008.
- [70] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
- [71] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6):641–647, 1994.
- [72] J. C. Bezdek, L. Hall, L. Clarke, et al. Review of mr image segmentation techniques using pattern recognition. *MEDICAL PHYSICS-LANCASTER PA-*, 20:1033–1033, 1993.
- [73] L. Le Cam. Maximum likelihood: an introduction. *International Statistical Review/Revue Internationale de Statistique*, pages 153–171, 1990.
- [74] T. Song, M. M. Jamshidi, R. R. Lee, and M. Huang. A modified probabilistic neural network for partial volume segmentation in brain mr image. *IEEE Transactions on Neural Networks*, 18(5):1424–1432, 2007.
- [75] R. B. Bhatt, G. Sharma, A. Dhall, and S. Chaudhury. Efficient skin region segmentation using low complexity fuzzy decision tree model. In *2009 Annual IEEE India Conference*, pages 1–4. IEEE, 2009.
- [76] M.-N. Wu, C.-C. Lin, and C.-C. Chang. Brain tumor detection using color-based k-means clustering segmentation. In *Third international conference on intelligent information hiding and multimedia signal processing (IIH-MSP 2007)*, volume 2, pages 245–250. IEEE, 2007.
- [77] W. Wiharto and E. Suryani. The comparison of clustering algorithms k-means and fuzzy c-means for segmentation retinal blood vessels. *Acta Informatica Medica*, 28(1):42, 2020.
- [78] L. He, Z. Peng, B. Everding, X. Wang, C. Y. Han, K. L. Weiss, and W. G. Wee. A comparative study of deformable contour methods on medical image segmentation. *Image and vision computing*, 26(2):141–163, 2008.
- [79] Y. Jiang, M. Wang, and H. Xu. A survey for region-based level set image segmentation. In *2012 11th International Symposium on Distributed Computing*

- and Applications to Business, Engineering & Science*, pages 413–416. IEEE, 2012.
- [80] H. Kalinic. Atlas-based image segmentation: A survey. *Croatian Scientific Bibliography*, pages 1–7, 2009.
- [81] S. Masood, M. Sharif, A. Masood, M. Yasmin, and M. Raza. A survey on medical image segmentation. *Current Medical Imaging*, 11(1):3–14, 2015.
- [82] Q. Yu, Y. Yang, F. Liu, Y.-Z. Song, T. Xiang, and T. M. Hospedales. Sketch-a-net: A deep neural network that beats humans. *International journal of computer vision*, 122(3):411–425, 2017.
- [83] T. B. Sekou, M. Hidane, J. Olivier, and H. Cardot. From patch to image segmentation using fully convolutional networks—application to retinal images. *arXiv preprint arXiv:1904.03892*, 2019.
- [84] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):640–651, 2016.
- [85] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928. IEEE, 2015.
- [86] Q. Chen, Z. Liu, Y. Zhang, K. Fu, Q. Zhao, and H. Du. Rgb-d salient object detection via 3d convolutional neural networks. *arXiv preprint arXiv:2101.10241*, 2021.
- [87] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [88] T. Lei, R. Wang, Y. Wan, X. Du, H. Meng, and A. K. Nandi. Medical image segmentation using deep learning: A survey. 2020.
- [89] J. Martin. Lymbic system and cerebral circuits for emotions, learning, and memory. *Neuroanatomy: text and atlas (third ed.)*. McGraw-Hill Companies, page 382, 2003.

- [90] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.
- [91] Y. Cheng, D. Wang, P. Zhou, and T. Zhang. Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35(1):126–136, 2018.
- [92] G. Hinton, O. Vinyals, J. Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.
- [93] R. P. Rao and D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87, 1999.
- [94] L. Fidon, S. Ourselin, and T. Vercauteren. Generalized wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: Brats 2020 challenge. In *International MICCAI Brainlesion Workshop*, pages 200–214. Springer, 2020.
- [95] I. S. Klyuzhin, Y. Xu, A. Ortiz, J. L. Ferres, G. Hamarneh, and A. Rahmim. Testing the ability of convolutional neural networks to learn radiomic features. *Computer Methods and Programs in Biomedicine*, 219:106750, 2022.
- [96] S. Bohlender, I. Oksuz, and A. Mukhopadhyay. A survey on shape-constraint deep learning for medical image segmentation. *arXiv preprint arXiv:2101.07721*, 2021.
- [97] S. Mohagheghi and A. H. Foruzan. Incorporating prior shape knowledge via data-driven loss model to improve 3d liver segmentation in deep cnns. *International Journal of Computer Assisted Radiology and Surgery*, 15(2):249–257, 2020.
- [98] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587. IEEE Computer Society, 1991.

-
- [99] A. Myronenko. 3d mri brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop*, pages 311–320. Springer, 2018.
 - [100] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.
 - [101] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Segmentation labels for the pre-operative scans of the tcga-gbm collection (2017), 2017.
 - [102] C. Bergamino, S. Hoey, K. Waller, and C. Skelly. Comparison of t1wflair and t1wtse sequences in imaging the brain of small animals using high-field mri. *Irish veterinary journal*, 72(1):1–10, 2019.

Abstract

For radiotherapy and removal of cancerous tissues, it is necessary to determine the location of the tumor and the vulnerable structures around the tumor before treating and irradiating the high-energy beam. To do this, the images received from the patient need to be segmented. This is usually done manually, which is not only time consuming but also very expensive.

Various methods for segmenting these images are presented automatically and semi-automatically, among which methods based on machine learning and deep learning have shown much higher accuracy than other methods. Despite this superiority, these methods have problems such as high computational costs, inability to learn the shape and structure of the tissue and the inability to segment the tumor due to the variability of its size, position and shape.

In this study, to solve the mentioned problems, methods have been proposed that increase its accuracy by transferring knowledge from a complex model to a simpler model and also providing a framework based on error feedback, without increasing complexity and computational costs in a simple model. Also, to solve the shape and structure learning problem for segmenting organs at risk, a shape-based cost function is proposed, which evaluates the validity or invalidity of the predicted shape for a three-dimensional structure based on the defined shape space. Finally, a framework based on the attention mechanism has been proposed for tumor segmentation, which consists of two modules of attention and segmentation. Attention module The task of finding the position of the tumor and the segmentation module performs the final segmentation according to the information of the module of attention. Final evaluations of the various datasets showed that the proposed methods would increase the accuracy of segmentation. For example, in using the shape-based cost function, the network Dice metric, for test data in hippocampus segmentation was increased from 0.81 for a network trained with the BCE cost function to 0.86 for a network trained with the proposed cost function based on the shape, and by using Attention module in brain tumor segmentation, this criterion increased from 0.68 for a network without an attention module to 0.79 for a network using an attention module.

Keywords: Organs at Risk Segmentation, Tumor Segmentation, Deep Learning, Deep Convolutional Neural Networks



Sharif University of Technology
Department of Electrical Engineering

M.Sc. Thesis

Organs at Risk (OAR) segmentation using machine learning methods

By:

Reza Karimzadeh

Supervisors:

Dr. Emad Fatemizadeh,

Dr. Hossein Arabi

May 2022