

عنوان تحقیق: پردازش تصاویر پزشکی
رشته تحصیلی: مهندسی تکنولوژی نرم افزار کامپیوتر
استاد راهنما: دکتر مهدی رنجبر حسنی
دانشجو: دانیال کمالی

یادگیری عمیق برای قطعه‌بندی تصویر پزشکی با استفاده از ترکیب چندبُعدی

چکیده

چندبُعدی بودن، در تصویرپردازی پزشکی، بسیار استفاده می‌شود، زیرا میتواند اطلاعات متعددی را در مورد یک هدف (تومور، اندام یا بافت) ارائه دهد. قطعه‌بندی با استفاده از چندبُعد، شامل اطلاعات متعدد، برای بهبود قطعه‌بندی است. اخیراً، رویکردهای مبتنی بر یادگیری عمیق، در طبقه‌بندی تصویر، قطعه‌بندی، تشخیص اشیاء و کارهای ردیابی، عملکرد پیشرفت‌های را نشان داده‌اند. یادگیری عمیق، به دلیل توانایی تعمیم و خودآموزی‌اش روی حجم زیادی داده، اخیراً در زمینه‌ی قطعه‌بندی تصاویر پزشکی چندبُعدی، سهم زیادی کسب کرده است. در این مقاله، مروری بر رویکردهای مبتنی بر یادگیری عمیق، برای کار قطعه‌بندی تصویر پزشکی چندبُعدی، ارائه می‌دهیم. در ابتدا، اصل کلی یادگیری عمیق و قطعه‌بندی تصویر پزشکی چندبُعدی را معرفی می‌کنیم. سپس، معماری‌های مختلف شبکه‌ای یادگیری عمیق را ارائه داده و استراتژی‌های ترکیبی آنها را تحلیل می‌کنیم و نتایجشان را مقایسه می‌کنیم. ترکیب سریع و اولیه، معمولاً استفاده می‌شود، زیرا ساده است و بر روی معماری شبکه‌ی طبقه‌بندی متعاقب، تمرکز می‌کند. با این حال، ترکیب آخری یا متعاقب، توجه بیشتری به استراتژی ترکیب می‌کند تا به روابط پیچیده‌ی میان بُدهای مختلف پی ببرد. به‌طور کلی، اگر روش ترکیب، به اندازه‌ی کافی مؤثر باشد، در مقایسه با ترکیب اولیه، ترکیب آخر، نتیجه‌ی دقیق‌تری تولید می‌کند. ما در مورد مسائل رایج در طبقه‌بندی تصویر پزشکی نیز بحث می‌کنیم. در نهایت، چشم‌اندازهایی به تحقیقات آینده، ارائه می‌دهیم.

کلمات کلیدی: یادگیری عمیق، قطعه‌بندی تصویر پزشکی؛ ترکیب چندبُعدی؛ مرور

۱. مقدمه

با توسعه‌ی سیستم‌های اکتساب تصاویر پزشکی، قطعه‌بندی با استفاده از چندبُعد، بسیار مورد مطالعه قرار گرفته است. استراتژی‌های مختلف ترکیب تصویر، مانند نظریه‌ی احتمال، مفهوم فازی، توابع باور و یادگیری ماشین، با موفقیت توسعه پیدا کرده‌اند. برای روش‌های مبتنی بر نظریه‌ی احتمال و یادگیری ماشین، ابعاد مختلف داده، دارای خواص آماری مختلف هستند که مدل‌سازی آنها با استفاده از مدل‌های سطحی و کم‌عمق را دشوار می‌سازد. برای روش‌های مبتنی بر مفهوم فازی، معیار فازی، درجه‌ی عضویت را نسبت به یک تصمیم، برای هر منبع، تعیین می‌کند. ادغام چند منبع، با استفاده از عملگرهای فازی در مجموعه‌های فازی، به‌دست می‌آید. برای روش‌های مبتنی بر نظریه‌ی تابع باور، هر منبع، ابتدا توسط یک گروه شواهد مدل‌سازی می‌شود، سپس قانون DempsterShafer برای ترکیب کردن تمام منابع، استفاده خواهد شد. مشکل اصلی در استفاده از نظریه‌ی تابع باور و نظریه‌ی مجموعه فازی، به انتخاب گروه شواهد، معیار فازی و تابع اجتماع یا ترکیب فازی مربوط می‌شود. با این حال، یک شبکه‌ی مبتنی بر یادگیری عمیق، می‌تواند به‌طور مستقیم، نگاشت را رمزگذاری کند. بنابراین، روش مبتنی بر یادگیری عمیق، نسبت به روش‌های متعارف و معمولی، پتانسیل زیادی برای تولید نتایج ترکیب بهتر دارد. از سال ۲۰۱۲، چند مدل شبکه‌ی عصبی پیچشی یا کانولوشنال عمیق، پیشنهاد شده‌اند، از جمله AlexNet، ZFNet، VGG، GoogleNet، Residual Net، DenseNet، FCN و U-Net. این مدل‌ها، نه تنها عملکرد پیشرفته‌ای در طبقه‌بندی تصویر، تشخیص اشیاء، قطعه‌بندی و کارهای ردیابی، ارائه کرده‌اند، بلکه دیدگاه جدیدی برای ترکیب تصویر نیز فراهم می‌کنند. عمده‌تأ چهار دلیل وجود دارد که در موفقیت آنها نقش دارند: اول اینکه، دلیل اصلی موفقیت شگفت‌انگیز مدل‌های یادگیری ماشین فراسنتی یادگیری عمیق، پیشرفت در شبکه‌های عصبی است، ویژگی‌های سطح بالا را از داده‌ها، به‌صورت افزایشی یا تدریجی یاد می‌گیرد، که نیاز به تخصص دامنه و استخراج ویژگی سخت را حذف می‌کند. و مسئله را به شیوه‌ی انتها به انتها حل می‌کند. دوماً، پیدایش واحد پردازش گرافیکی یا GPU و کتابخانه‌های محاسباتی آن، باعث می‌شود، این مدل بتواند ۱۰ تا ۳۰ برابر، نسبت به CPU سریع‌تر آموزش ببیند. و بسته‌های نرم‌افزاری منبع باز، پیاده‌سازی

GPU کارآمد را فراهم می‌کنند. سوم اینکه، مجموعه داده‌های در دسترس عموم، مانند ImageNet، برای آموزش می‌تواند استفاده شود که به محققان اجازه می‌دهد، انواع جدید مدل‌های یادگیری عمیق را آموزش دهند و آزمایش کنند. در نهایت، چند تکنیک بهینه‌سازی کارآمد موجود، نیز به موفقیت نهایی یادگیری عمیق می‌کنند، مانند dropout، نرمال‌سازی دسته‌ای batch normalization، بهینه‌ساز Adam و غیره. با تابع فعال‌سازی ReLU و انواع آن، می‌توانیم وزن‌ها را به‌روزرسانی کنیم و به عملکرد بهینه دست پیدا کنیم.

به دلیل موفقیت یادگیری عمیق، محققان در زمینه‌ی تصاویر پزشکی، نیز تلاش کرده‌اند، روش‌های مبتنی بر یادگیری عمیق را به قطعه‌بندی تصویر پزشکی در مغز، ریه، لوزالمعده، پروستات و اندام‌های متعدد اعمال کنند. قطعه‌بندی تصویر پزشکی، در آنالیز تصاویر پزشکی، یک حوزه‌ی مهم است و برای تشخیص، نظارت و درمان، ضروری است. هدف این است که به هر پیکسل موجود در تصویر، برچسب اختصاص دهیم، این کار در کل شامل دو مرحله است، اولاً، بافت ناسالم یا نواحی موردنظر را تشخیص دهیم؛ دوماً، ساختار مختلف آناتومی یا نواحی موردنظر را کاهش دهیم. این روش‌های مبتنی بر یادگیری عمیق، در مقایسه با روش‌های معمولی، در زمینه‌ی قطعه‌بندی تصاویر پزشکی، به کارایی برتری دست یافته‌اند. به‌منظور به‌دست آوردن قطعه‌بندی دقیق‌تر برای تشخیص بهتر، استفاده از تصاویر پزشکی چندبعدی، به یک استراتژی با روند روبه‌رشد، تبدیل شده است.

تجزیه و تحلیل کامل مقالات موجود، با کلمات کلیدی «یادگیری عمیق»، «قطعه‌بندی تصاویر پزشکی» و «چندبعدی بودن» در موتور جستجوی گوگل اسکولار، در شکل ۱ انجام شده است در ۱۷ جولای ۲۰۱۹.

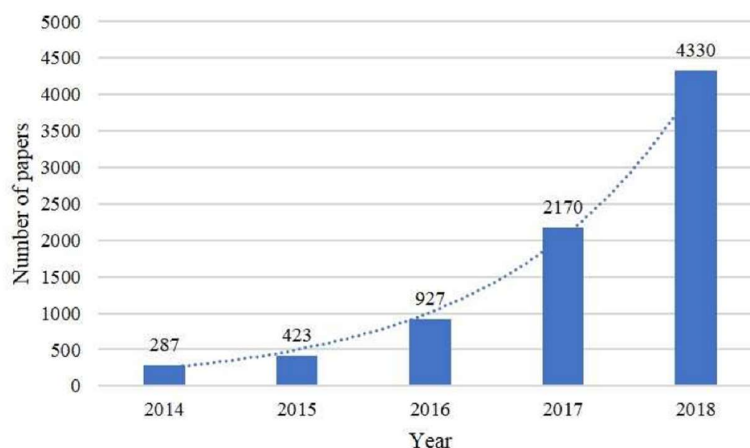
می‌توانیم مشاهده کنیم که تعداد مقالات، هر ساله از ۲۰۱۴ تا ۲۰۱۸، افزایش پیدا کرده است که بدین معناست که قطعه‌بندی تصاویر پزشکی چندبعدی در یادگیری عمیق، در سال‌های اخیر، توجه بیشتری کسب کرده است. برای اینکه ابعاد این زمینه‌ی پژوهشی را بهتر درک کنیم، تولید علمی جامعه‌ی قطعه‌بندی تصویر، جامعه‌ی قطعه‌بندی تصاویر پزشکی و قطعه‌بندی تصاویر پزشکی با استفاده از ترکیب چندبعدی را با و بدون یادگیری عمیق، در شکل ۲ مقایسه می‌کنیم. از این شکل می‌توانیم ببینیم که تعداد مقالات، در روش‌های بدون یادگیری عمیق، یک روند نزولی داشته است، اما مقالاتی که از روش یادگیری عمیق استفاده می‌کنند در هر زمینه‌ی

پژوهشی، افزایش پیدا کرده است. مخصوصاً در زمینه‌ی قطعه‌بندی تصاویر پزشکی، به دلیل مجموعه داده‌های محدود، روش‌های کلاسیک هنوز، یک موقعیت بارزتر دارند، اما در روش‌هایی که از یادگیری عمیق استفاده می‌کنند، می‌توان یک روند افزایشی آشکار را مشاهده کرد. ابعاد اصلی در تحلیل تصاویر پزشکی، عبارتند از: توموگرافی (پرتونگاری) کامپیوتری (CT)، تصویربرداری با رزونانس مغناطیسی (MRI) و برشنگاری با گسیل پوزیترون (PET). در مقایسه با تصاویر واحد، تصاویر چندبعدی، به استخراج ویژگی‌ها از نماها و زوایای مختلف کمک می‌کنند، اطلاعات تکمیلی به همراه دارند و به بازنمایی بهتر داده‌ها و قدرت تمییز و افتراق شبکه، کمک می‌کنند. همان‌طور که در مرجع ذکر شده، تصویر CT می‌تواند اختلال ماهیچه و استخوان، مانند تومورها و شکستگی‌های استخوان را تشخیص دهد، درحالی‌که تصویر MR می‌تواند کنتراست یا تضاد بافت نرم خوب را بدون تشعشع، ارائه دهد. تصاویر عملکردی، مانند PET، فاقد خصوصیات آناتومیکی (اندامشناسی) هستند، درحالی‌که می‌توانند اطلاعات متابولیکی و عملکردی کمی در مورد بیماری را فراهم کنند. MRI به دلیل وابستگی‌اش به پارامترهای اکتساب متغیر، مانند: تصاویر T1 - وزندار (T1)، T1 - وزندار با کنتراست بهبودیافته (T1c)، T2 - وزندار (T2) و بازیابی وارونگی تضعیف مایع (Flair) می‌تواند اطلاعات تکمیلی ارائه دهد. T2 و Flair، برای تشخیص تومور با ورم صفاتی مناسب هستند، درحالی‌که T1 و T1c برای تشخیص هسته‌ی تومور بدون ورم صفاتی، مناسب هستند. بنابراین، استفاده از تصاویر چندبعدی، می‌تواند عدم قطعیت اطلاعات را کاهش دهد و تشخیص بالینی و دقت قطعه‌بندی را تشخیص دهد. چند تصویر پزشکی چندبعدی که بیشترین کاربرد را دارند در شکل ۳ نشان داده شده است. ترکیب اولیه و زودتر، ساده است و اکثر کارها از این استراتژی ترکیب برای انجام قطعه‌بندی استفاده می‌کنند، که روی طراحی معماری شبکه‌ی قطعه‌بندی پیچیده‌ی متعاقب، تمرکز می‌کند، اما روابط میان بعدهای مختلف را در نظر نمی‌گیرد و نحوه‌ی ترکیب اطلاعات ویژگی مختلف برای بهبود عملکرد قطعه‌بندی را تحلیل نمی‌کند. با این حال، ترکیب آخر و بعدی، توجه بیشتری به مسئله‌ی ترکیب می‌کند، زیرا هر بعد، به عنوان یک ورودی از یک شبکه استفاده می‌شود که می‌تواند اطلاعات ویژگی پیچیده و مکمل هر بعد را یاد بگیرد. به‌طور کلی، در مقایسه با ترکیب زودتر، این نوع ترکیب، اگر روش

ترکیب به قدر کافی مؤثر باشد، می‌تواند عملکرد قطعه‌بندی بهتری را به‌دست بیاورد. و انتخاب روش ترکیب، به مسئله‌ی خاص، بستگی دارد.

بررسی‌های دیگری هم روی آنالیز تصاویر پزشکی با استفاده از یادگیری عمیق، وجود دارد. باین‌حال، آنها روی استراتژی ترکیب تمرکز نمی‌کنند. برای مثال، لیتجنس^۱ و همکاران، مفاهیم اصلی یادگیری عمیق در تحلیل تصاویر پزشکی را بررسی کردند. برنل^۲ و همکاران در CNN عمیق برای تحلیل MRI، یک مرور انجام دادند. در این مقاله، روی روش‌های ترکیب تصاویر پزشکی چندبعدی برای قطعه‌بندی تصاویر پزشکی، تمرکز می‌کنیم. ادامه‌ی این مقاله، به‌شرح زیر است. در بخش ۲، اصول کلی یادگیری عمیق و قطعه‌بندی تصاویر پزشکی چندبعدی را معرفی می‌کنیم. در بخش ۳، نحوه‌ی آماده‌سازی داده‌ها قبل از تغذیه^۳ به شبکه را نشان می‌دهیم.

در بخش ۴، شبکه‌ی قطعه‌بندی چندبعدی دقیق بر مبنای استراتژی‌های ترکیب مختلف را توضیح می‌دهیم. در بخش ۵، تعدادی از مسائل رایج در این زمینه را بحث می‌کنیم. در نهایت، چشم‌انداز آینده، در زمینه‌ی قطعه‌بندی تصاویر پزشکی چندبعدی را مورد بحث قرار می‌دهیم.

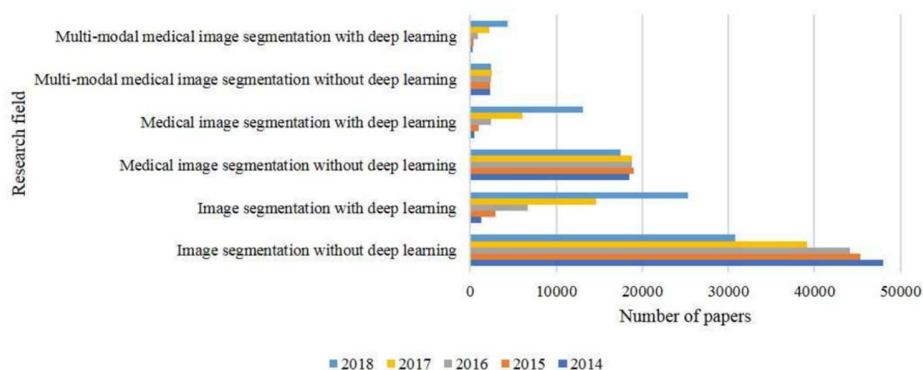


شکل ۱. گرایش قطعه‌بندی تصاویر پزشکی چندبعدی در یادگیری عمیق.

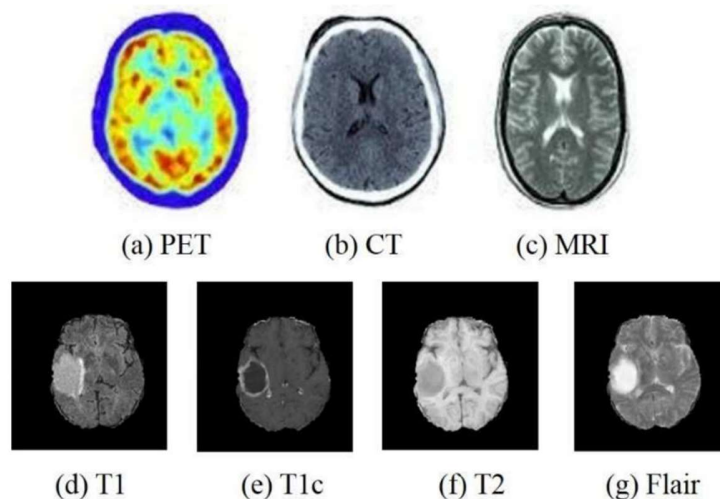
¹ Litjens

² Bernal

³ Feeding



شکل ۲. گرایش زمینه‌ی پژوهشی نسبی با / بدون یادگیری عمیق



شکل ۳. تصاویر پزشکی چندبعدی (a) – (c) تصاویر پزشکی چندبعدی هستند که به‌طور معمول استفاده می‌شوند و

(d) – (g) توالی‌های مختلف MRI مغز هستند.

۲. روش‌های مبتنی بر یادگیری عمیق

۱.۲. یادگیری عمیق

یادگیری عمیق به یک شبکه‌ی عصبی، با چند لایه واحد پردازشی غیرخطی اشاره دارد. هر لایه‌ی متوالی، از خروجی لایه‌ی قبلی‌اش، به‌عنوان ورودی، استفاده می‌کند. این شبکه با استفاده از این لایه‌ها، می‌تواند ویژگی‌های سلسله‌مراتبی پیچیده را از حجم زیادی داده استخراج کند. در سال‌های اخیر، یادگیری عمیق، در طبقه‌بندی تصویر، شناخت، تشخیص اشیاء و تحلیل تصاویر پزشکی، به بهبودهای

چشمگیری دست پیدا کرده است، جایی که نتایج بسیار خوبی تولید کرده‌اند که قابل مقایسه با متخصصین انسانی است یا حتی برتر از آن‌هاست. از میان الگوریتم‌های یادگیری عمیق، مانند خود رمزگذارهای انباشته یا پشته‌ای، ماشین‌های بولتزمن عمیق و شبکه‌های عصبی پیچشی، موفق‌ترین آن‌ها برای قطعه‌بندی تصاویر، شبکه‌های عصبی پیچشی (CNN) است. که برای اولین بار، در سال ۱۹۸۹ توسط لیچان^۴ پیشنهاد شد و اولین کاربرد موفق در دنیای واقعی، تشخیص رقم‌های دست‌نویس در سال ۱۹۹۸ توسط لیچان است، جایی که یک معماری کاملاً سازگار و پنج لایه پیشنهاد کرد. به دلیل نتایج دقت آن (نرخ خطای ۱ درصد و نرخ رد ۹ درصدی از مجموعه داده‌ی کاراکترهای دست‌نویس ۲۰۰۷)، این شبکه‌های عصبی می‌توانند به یک مسئله‌ی دنیای واقعی، اعمال شوند. با این حال، تا زمان انجام تحقیق کریجوسکی^۵ و همکاران، روی چالش ImageNet در سال ۲۰۱۲، توجه زیادی جلب نکرده بود. AlexNet. پیشنهادی، مشابه با LeNet اما عمیق‌تر، نسبت به تمام رقبای خود، بهتر عمل کرد و با کاهش خطای بالای ۵ (درصد نمونه‌های تستی که کلاس صحیح در ۵ کلاس پیشبینی شده‌ی اول، وجود نداشته است) از ۲۶ درصد به ۱۵/۳ درصد، برنده‌ی این چالش شد. در سال‌های بعدی، روشهای دیگری بر پایه‌ی معماری‌های CNN پیشنهاد شدند، از جمله VGGNet، GoogleNet، Residual Net و DenseNet، جدول ۱ جزئیات این معماری‌های شبکه را توضیح می‌دهد.

Architecture	Article	Rank on ILSVRC	Top-5 error rate	Number of parameters
LeNet[36]	LeCun et al. 1998	N/A	N/A	60 thousand
AlexNet[11]	Krizhevsky et al. 2012	1st	16.4%	60 million
ZFNet[12]	Zeiler et al. 2013	1st	11.7%	N/A
VGG Net[13]	Simonyan et al. 2014	2nd	7.3%	138 million
GoogleNet[14]	Szegedy et al. 2015	1st	6.7%	5 million (V1) & 23 million (V2)
ResNet[15]	He, Kaiming et al. 2016	1st	3.57%	25.6 million (ResNet-50)
DenseNet[16]	Huang et al. 2017	N/A	N/A	6.98 million (DenseNet-100, k=12)

جدول ۱. خلاصه‌ی معماری‌های شبکه‌ی یادگیری عمیق، ILSVRC: چالش تشخیص بصری مقیاس بزرگ ImageNet

^۴ LeCun

^۵ Krizhevsky

CNN، یک شبکه‌ی عصبی چندلایه است که شامل لایه‌های کاملاً متصل، پیچش، ادغام و فعال‌سازی است. لایه‌های پیچش، هسته‌ی CNN‌ها هستند و برای استخراج ویژگی مورد استفاده قرار می‌گیرند. عملگر پیچش، بسته به فیلترهای مورد استفاده، میتواند نگاشت‌های ویژگی مختلفی تولید کند. لایه‌ی ادغام، با استفاده از حداکثر یا میانگین همسایگی تعریف شده، به‌عنوان مقداری برای کاهش اندازه‌ی مکانی هر نگاشت ویژگی، یک عملیات کاهش نمونه^۷ را انجام می‌دهد. لایه‌ی یکسوسوده‌ی غیرخطی (ReLU) و تغییرات آن، مانند Leaky ReLU (ReLU سوراخدار)، از جمله توابع فعال‌سازی هستند که بیش‌ترین کاربرد را دارند که داده‌ها را با اختصار هر مقدار منفی به صفر تبدیل می‌کند و در عوض مقادیر ورودی مثبت، به خروجی انتقال داده می‌شوند. نورون‌ها در یک لایه‌ی کاملاً متصل، به تمام فعال‌سازی‌ها در لایه‌ی قبلی، کاملاً متصل هستند. آنها قبل از خروجی طبقه‌بندی یک CNN قرار می‌گیرند و برای مسطح کردن نتایج، قبل از انجام یک پیش‌بینی با استفاده از طبقه‌بندهای خطی، انجام می‌شود. در حین آموزش معماری CNN، مدل، نمرات کلاس تصاویر آموزشی را پیش‌بینی می‌کند، با استفاده از تابع زیان انتخابی، زیان را محاسبه می‌کند و در نهایت، وزن‌ها را با استفاده از روش نزول شیب یا گرادیان با روش پس انتشار، به‌روز رسانی می‌کند. زیان آنتروپی متقاطع، یکی از پرکاربردترین توابع زیان است و کاهش گرادیان تصادفی (SGD) محبوب‌ترین روش برای انجام کاهش گرادیان است.

Dataset	Train	Validation	Test	Segmentation Task	Modality	Image Size
Brats2012	35	N/A	15	Brain tumor	T1, T1C, T2, Flair	160 × 216 × 176 176 × 176 × 216
Brats2013	35	N/A	25	Brain tumor	T1, T1C, T2, Flair	160 × 216 × 176 176 × 176 × 216
Brats2014	200	N/A	38	Brain tumor	T1, T1C, T2, Flair	160 × 216 × 176 176 × 176 × 216
Brats2015	200	N/A	53	Brain tumor	T1, T1C, T2, Flair	240 × 240 × 155
Brats2016	200	N/A	191	Brain tumor	T1, T1C, T2, Flair	240 × 240 × 155
Brats2017	285	46	146	Brain tumor	T1, T1C, T2, Flair	240 × 240 × 155
Brats2018	285	66	191	Brain tumor	T1, T1C, T2, Flair	240 × 240 × 155
ISLES2015	28	N/A	36	Ischemic stroke lesion	T1, T2, TSE, Flair, DWI, TFE/TSE	230 × 230 × 154
	30	N/A	20		T1c, T2, DWI, CBF, CBV, TTP, Tmax	N/A
MRBrainS13	5	N/A	15	Brain Tissue	T1, T1_Imm, T1 IR, Flair	256 × 256 × 192 240 × 240 × 48
NeoBrainS12	20	N/A	5	Brain Tissue	T1, T2	384 × 384 × 50 512 × 512 × 110
						512 × 512 × 50
iSeg-2017	10	N/A	13	Brain Tissue	T1,T2	N/A
CHAOS	20	N/A	20	Abdominal Organs	CT, T1-DUAL, T2-SPIR	N/A
IVD	16	N/A	8	Intervertebral Disc	In-phase, Opposed-phase, Fat, Water	N/A

جدول ۲. خلاصه‌ی مجموعه داده‌های قطعه‌بندی پزشکی چندبعدی

Evaluation metric	Mathematical description
Dice score(DSC) DSC	$DSC = \frac{2TP}{2TP + FP + FN}$
Sensitivity	$Sensitivity = \frac{TP}{TP + FN}$
Specificity	$Specificity = \frac{TN}{TN + FP}$
Hausdorff distance(HD)	$HD = \max \{ \sup_{r \in \partial R} d_m(s, r), \sup_{s \in \partial S} d_m(s, r) \}$
Absolute relative volume difference(ARVD)	$ARVD(X, Y) = \left 100 \times \left(\frac{ X }{ Y } - 1 \right) \right $
Average boundary distance (ABD)	$ABD(X_s, Y_s) = \frac{1}{N_{X_s} + N_{Y_s}} \left(\sum_{x \in X_s} \min_{y \in Y_s} d(x, y) + \sum_{y \in Y_s} \min_{x \in X_s} d(y, x) \right)$

جدول ۳. خلاصه‌ی معیارهای ارزیابی مورد استفاده برای این مجموعه داده‌ها، با توجه به تعداد مثبت اشتباه (FP)، مثبت صحیح

(TP)، منفی اشتباه (FN) و منفی صحیح (TN)، ∂S و ∂R ، مجموعه‌ی پیکسل‌ها / واکسل‌های مرزی ضایعه برای قطعات

پیش‌بینی شده و واقعی هستند و $d_m(v, v)$ حداقل فاصله‌ی اقلیدسی میان یک واکسل v و واکسل‌های موجود در یک مجموعه V

می‌باشد. $|X|$ تعداد واکسل‌های موجود در قطعه‌بندی مرجع و $|Y|$ تعداد واکسل‌های موجود در قطعه‌بندی الگوریتم است، X_s و

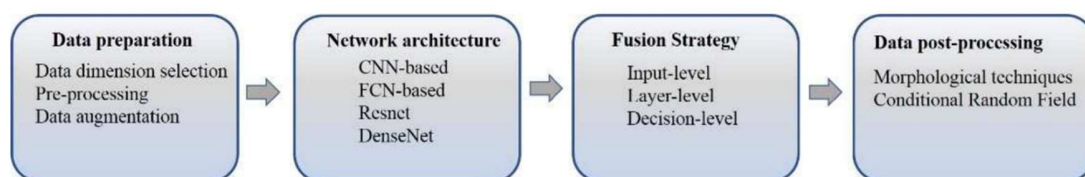
Y_s ، به ترتیب، مجموعه‌ی نقاط سطح قطعه‌بندی مرجع و الگوریتم هستند. عملگر d ، عملگر فاصله‌ی اقلیدسی است.

۲.۲. قطعه‌بندی تصاویر پزشکی چندبعدی

با توجه به اندازه، شکل و مکان متغیر بافت هدف، قطعه‌بندی تصویر پزشکی، یکی از چالش برانگیزترین کارها در زمینه‌ی تحلیل تصاویر پزشکی است. با وجود تنوع معماری‌های شبکه‌ی قطعه‌بندی پیشنهادی، مقایسه‌ی عملکرد الگوریتم‌های مختلف هنوز سخت است، زیرا اکثر این الگوریتم‌ها روی مجموعه داده‌های مختلف ارزیابی می‌شوند و در معیارهای مختلف گزارش می‌شوند. به‌منظور دستیابی به قطعه‌بندی دقیق و مقایسه‌ی جدیدترین روش‌های مختلف، تعدادی چالش معروف در قطعه‌بندی، ایجاد شده است، مانند قطعه‌بندی تومور مغز (BraTS)، قطعه‌بندی سگته مغزی ایسکمیک (ISLES)، قطعه‌بندی تصویر مغز MR (MRBrainS)، قطعه‌بندی مغز نوزادان (NeoBrainS)، قطعه‌بندی اندام شکمی سالم ترکیبی (CT-MR) (CHAOS) قطعه‌بندی MRI مغز کودک شش ماهه (Iseg – ۲۰۱۷) و محلی‌سازی و قطعه‌بندی دیسک بین مهره‌ای خودکار از تصاویر MR چندحالتی سه بعدی (IVDM3Seg). جدول ۲، اطلاعات دقیق مجموعه داده که در بالا ذکر شده‌اند را توضیح می‌دهد. جدول ۳ معیارهای ارزیابی اصلی در این مجموعه‌های داده را نشان می‌دهد. ما یک اصل قطعه‌بندی تصویر پزشکی چندبعدی بر اساس یادگیری عمیق را توضیح می‌دهیم که در شکل ۴ نشان داده شده است. این اصل شامل چهار قسمت است: آماده‌سازی داده، معماری شبکه، استراتژی ترکیب و پس پردازش داده. در مرحله‌ی آماده‌سازی داده، ابعاد داده در ابتدا انتخاب می‌شود، و از پس پردازش برای کاهش تنوع میان تغییرات، استفاده می‌شود و از استراتژی قطعه‌بندی نیز برای افزایش داده‌های آموزشی استفاده می‌شود تا از مشکل برازش بیش از حد^۶ اجتناب شود. در مراحل معماری شبکه و استراتژی ترکیب، استراتژی‌های تصاویر چندبعدی مفصل و شبکه‌ی پایه، برای آموزش شبکه‌ی قطعه‌بندی مورد استفاده قرار می‌گیرند. در مرحله‌ی پس پردازش داده، چند تکنیک پس

⁶ over-fit

پردازش ماندن روش‌های مورفولوژیکی و زمینه‌ی تصادفی شرطی، برای اصلاح نتیجه‌ی قطعه‌بندی نهایی، استفاده می‌شوند. در کار قطعه‌بندی تصویر پزشکی چندبعدی، ترکیب ابعاد مختلف، مسئله‌ی اصلی این کار است. با توجه به سطحی از معماری شبکه، که ترکیب در آن انجام می‌شود، استراتژی‌های ترکیب می‌توانند به سه گروه طبقه‌بندی شوند: ترکیب سطح ورودی، ترکیب سطح لایه و ترکیب سطح تصمیم، جزئیات به بخش ۴ برمی‌گردد.



شکل ۴. اصل قطعه‌بندی تصویر پزشکی چندبعدی براساس یادگیری عمیق.

۳. پردازش داده

این بخش، پردازش داده، از جمله تکنیک‌های انتخاب بُعد، پیش پردازش تصویر، افزایش داده‌ها و پس پردازش را توضیح خواهد داد. این مرحله در شبکه‌ی قطعه‌بندی مبتنی بر یادگیری عمیق، مهم است.

۳.۱. ابعاد داده

قطعه‌بندی تصاویر پزشکی، معمولاً با تصاویر سه بعدی سروکار دارد. برخی از مدل‌ها به‌طور مستقیم با تصاویر سه بعدی سروکار دارند تا مدل را آموزش دهند، درحالی‌که تعداد دیگری از مدل‌ها، تصاویر سه بعدی را تکه به تکه، پردازش می‌کنند. روش سه‌بعدی تصویر سه بعدی را به‌عنوان ورودی می‌گیرد و هسته‌ی پیچشی سه‌بعدی را برای بهره‌برداری از اطلاعات زمین‌های مکانی تصویر، استفاده می‌کند. اشکال اصلی، هزینه‌ی محاسباتی گران قیمت آن است. در مقایسه با استفاده از کل حجم تصویر، برای آموزش مدل، برخی از تکه‌های کوچک سه بعدی، می‌توانند برای کاهش هزینه‌ی محاسباتی، مورد استفاده قرار

بگیرند. برای نمونه، کامنیتساس^۷ و همکاران، 10k تکه‌ی تصادفی سه‌بعدی را در فواصل منظم، برای

آموزش قطعه‌بندی ضایعه‌ی مغزی، استخراج می‌کنند.

رویکرد ۲ بعدی، تکه یا برش تصویر که از تصاویر سه بعدی استخراج شده را به‌عنوان ورودی می‌گیرد و هسته‌ی پیچشی دوبعدی را اعمال می‌کند، روش دوبعدی می‌تواند به‌طور مؤثر، هزینه‌ی محاسباتی را کاهش

دهد، درحالی‌که اطلاعات مکانی تصویر در بعد Z را نادیده می‌گیرد. برای مثال، ژائو^۸ و همکاران،

ابتدا، FCNN ها را با استفاده از تکه تصاویر آموزش دادند و سپس CFR ها به‌عنوان شبکه‌های عصبی

بازگشتی با استفاده از تکه تصاویر با پارامترهای FCNN ثابت، آموزش دیدند و در نهایت، FCNN ها و

CFR-RNN با استفاده از تکه تصاویر، به‌طور دقیق تنظیم شدند. برای بهره‌برداری از اطلاعات ویژگی

تصویر دوبعدی و سه‌بعدی، املیانرسکی^۹ و همکاران یک مدل مبتنی بر CNN را برای قطعه‌بندی تومور

مغز توضیح دادند، که ابتدا ویژگی‌های دوبعدی تصویر را از نمای محوری، کروئال و سهمی شکل استخراج

می‌کند و سپس، آنها را به‌عنوان ورودی مدل مبتنی بر CNN سه‌بعدی، انتخاب می‌کند. این می‌تواند

اطلاعات ویژگی غنی را در سه بعد یاد بگیرد که به عملکرد خوبی با نمرات تاس متوسط ۰/۹۱۸ (تومور

کامل)، ۰/۸۸۳ (هسته‌ی تومور) و ۰/۸۵۴ (هسته‌ی پیشرفته یا افزایشی) دست پیدا می‌کند.

۳.۲. پیش‌پردازش

پیش‌پردازش، نقش مهمی در کار قطعه‌بندی بعدی دارد، مخصوصاً برای قطعه‌بندی تصاویر پزشکی

چندبعدی، زیرا در این تصاویر، شدت، کنتراست و نویز گوناگون وجود دارد. بنابراین، برای اینکه تصاویر،

با شباهت بیشتری ظاهر شوند و آموزش شبکه مسطح و قابل تعیین باشد، قبل از تغذیه‌ی شبکه‌ی

قطعه‌بندی، چند تکنیک پیش‌پردازش، مورد استفاده قرار می‌گیرند.

⁷ Kamnitsas

⁸ Zhao

⁹ Mlynarski

تکنیک‌های معمولی پیش‌پردازش، شامل ثبت تصویر، تصحیح میدان بایاس و نرمال‌سازی شدت است. برای مجموعه داده‌ی BraTS، ثبت تصویر، قبل از اینکه در اختیار عموم قرار بگیرد، انجام شده است. از روش N4ITK برای اصلاح انحراف داده‌های MRI استفاده کردند.

برای نرمال‌سازی هر بُعد از هر یک از بیماران، به‌طور مستقل با تفریق میانگین و تقسیم بر انحراف استاندارد ناحیه‌ی مغز پیشنهاد شدند.

۳.۳. افزایش داده

اکثر اوقات، تعداد زیادی برچسب برای آموزش، به دلایل متعدد در دسترس نیستند. برچسب‌گذاری مجموعه داده، به یک متخصص در این حوزه نیاز دارد که هم پرهزینه و هم زمانبر است. هنگام آموزش شبکه‌های عصبی بزرگ، از داده‌های آموزشی محدود و کم، مسئله‌ی برازش بیش از حد، باید در نظر گرفته شود. افزایش داده‌ها، روشی برای کاهش برازش بیش از حد و افزایش حجم داده‌های آموزشی است. که تصاویر جدید را با تبدیل (چرخش، تفسیر، مقیاس‌گذاری، انتقال، تحریف و افزودن نویزهایی مانند نویز گاوسین) تصاویر موجود در مجموعه داده‌ی آموزشی، ایجاد می‌کند. هم تصویر اصلی و هم تصاویر ساخته‌شده، به شبکه‌ی عصبی خورانده می‌شوند. برای مثال، آیسنس^{۱۰} و همکاران حل مشکل برازش بیش از حد را با استفاده از انواع زیادی از روش‌های افزایش داده مانند چرخش تصادفی، مقیاس‌گذاری تصادفی، تغییر شکل الاستیک تصادفی، افزایش اصلاح گاما و تکنیک قرینه‌سازی در طی آموزش، پیشنهاد دادند.

۳.۴. پس پردازش

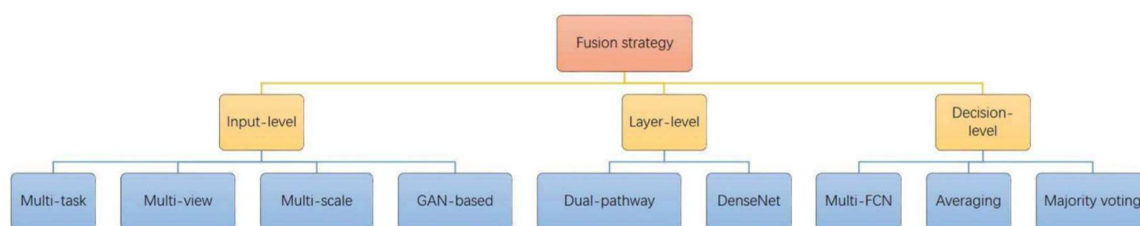
پس پردازش، برای تصحیح نتیجه‌ی نهایی در شبکه‌ی قطعه‌بندی مورد استفاده قرار می‌گیرد. برچسب‌های قطعه‌بندی جداگانه با اندازه‌ی کوچک، مستعد اثرات مصنوعی هستند و بزرگترین حجم،

¹⁰ Isensee

معمولاً در قطعه‌بندی نهایی، استفاده می‌شود. در این حالت، تکنیک‌های مورفولوژیکی (ساختاری) برای حذف قطعات کوچک نادرست و حفظ بزرگترین حجم، ترجیح داده می‌شوند. و تعدادی از روش‌های پس پردازش، می‌توانند مطابق با ساختار ناحیه‌ی شناسایی شده، طراحی شوند. برای مثال، با توجه به اینکه بیماران LGG ممکن است تومور پیشرفته نداشته باشند، آیسنس و همکاران، پیشنهاد دادند که اگر تعداد تومورهای پیشرفته کمتر از یک حد آستانه باشد، واکسل‌های تومور پیشرفته با نویز جایگزین شوند. از آنجایی که اگر یک واکسل مثبت اشتباه در قطعه‌بندی پیش‌بینی شده وجود داشته باشد که در آن نبود تومور پیشرفته در واقعیت، منجر به یک نمره‌ی تاس ۰ خواهد شد. مورد دیگر در، یک میدان تصادفی شرطی کاملاً متصل سه بعدی (CRF) برای پس پردازش استفاده می‌شود که به‌طور مؤثر، موارد مثبت اشتباه را برای تصحیح نتیجه‌ی قطعه‌بندی، حذف می‌کند.

۴. شبکه‌های قطعه‌بندی چندبعدی

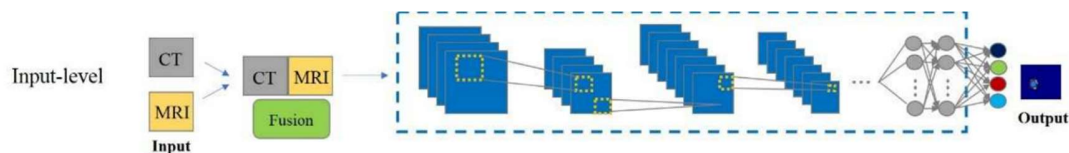
در طی سال‌ها، تکنیک‌های خودکار و شبه خودکار مختلف، برای قطعه‌بندی تصاویر پزشکی چندبعدی، با استفاده از روش‌های مبتنی بر یادگیری عمیق پیشنهاد شده‌اند، مانند CCN و FCN و به‌ویژه U-Net. مطابق با استراتژی‌های ترکیب چندبعدی، معماری‌های شبکه را به شبکه‌های ترکیب سطح ورودی، سطح لایه و سطح تصمیم، دسته‌بندی می‌کنیم، و برای هر کدام، روش‌های رایج را نشان می‌دهیم که در شکل ۵ نشان داده شده است.



شکل ۵. دسته‌بندی کلی استراتژی ترکیب.

۴.۱. شبکه‌ی ترکیب سطح ورودی

در استراتژی ترکیب سطح ورودی، تصاویر چندبعدی، به صورت ورودی‌های چندکاناله برای یادگیری یک‌بازنمایی ویژگی ترکیبی، کانال به کانال، ترکیب می‌شوند، و سپس شبکه‌ی طبقه‌بندی را آموزش می‌دهد. اکثر شبکه‌های قطعه‌بندی تصاویر پزشکی چندبعدی موجود، استراتژی ترکیب سطح ورودی را انتخاب می‌کنند که به طور مستقیم تصاویر چندبعدی را در فضای ورودی اصلی، ادغام می‌کند. شکل ۶، معماری شبکه‌ی عمومی متعلق به شبکه‌ی قطعه‌بندی ترکیب سطح ورودی را توضیح می‌دهد. ما CT و MRI را به عنوان دو بعد یا حالت در نظر می‌گیریم، شبکه‌ی عصبی پیچشی را به عنوان شبکه‌ی قطعه‌بندی و قطعه‌بندی تومور مغز را به عنوان کار قطعه‌بندی، انتخاب می‌کنیم. با استفاده از استراتژی ترکیب سطح ورودی، اطلاعات ویژگی غنی از حالت‌های مختلف را می‌توان به طور کامل در تمام لایه‌ها، استفاده کرد. این نوع ترکیب، معمولاً از چهار تکنیک استفاده می‌کند که عبارتند از: قطعه‌بندی چندکاره، قطعه‌بندی چندنما، قطعه‌بندی چندمقیاس و قطعه‌بندی مبتنی بر GAN.



شکل ۶. معماری کلی شبکه‌ی ترکیب سطح ورودی.

مثلاً، وانگ و همکاران، یک شبکه‌ی قطعه‌بندی چندبعدی را با استفاده از مجموعه داده‌ی BraTS پیشنهاد دادند که تومور مغز را به سه زیرناحیه‌ی کل تومور، هسته‌ی تومور و هسته‌ی تومور پیشرفته، تقسیم می‌کند. و از تکنیک‌های چند وظیفه و چندنما استفاده می‌کند. به منظور دستیابی به یک مجموعه ویژگی متحد و یکپارچه، به طور مستقیم، چهار حالت (T1, T1c, T2 و Flair متعلق به MRI) را به عنوان ورودی‌های چندکاناله در فضای ورودی، ادغام می‌کند. سپس، با توجه به ساختار سلسله‌مراتبی تومور مغز، قطعه‌بندی چند کلاسه‌ی پیچیده را به چند وظیفه‌ی قطعه‌بندی ساده‌تر تقسیم می‌کند. کل تومور در ابتدا تقسیم می‌شود و سپس کادر محصور کننده‌ی تومور کل، برای

قطعه‌بندی هسته‌ی تومور، استفاده می‌شود. بر اساس کادر محصور کننده‌ی حاصل از هسته‌ی تومور، هسته‌ی تومور پیشرفته در نهایت، قطعه قطعه می‌شود. علاوه بر این، برای بهره‌گیری از اطلاعات زمینه‌ای سه بعدی، برای هر وظیفه‌ی فردی، با میانگین گیری نتایج بیشینه‌ی هموار وظیفه‌ی فردی، از نتایج قطعه‌بندی از سه نمای متعامد مختلف (محوری، کرونال و سهمی)، استفاده کردند. آزمایشات با مجموعه آزمایش داده‌های BraTS ۲۰۱۷، نشان می‌دهد که روش پیشنهادی، برای هسته تومور پیشرفته، تومور کل و هسته‌ی تومور، به ترتیب، میانگین نمرات تاس ۰/۷۸۳۱، ۰/۸۷۳۹ و ۰/۷۷۴۸ را بدست آورده که در چالش BraTS 2017، دومین جایگاه را دارد. قطعه‌بندی چند وظیفه‌ای، وظیفه‌ی پیچیده‌ی قطعه‌بندی چند کلاس را به وظایف قطعه‌بندی ساده‌تر تقسیم کرده و از مزایای ساختار سلسله مراتبی زیرنواحی تومور استفاده می‌کند تا دقت قطعه‌بندی را بهبود دهد.

ژائو و همکاران، نیز یک شبکه‌ی قطعه‌بندی چندوظیفه روی مجموعه داده‌ی BraTS پیشنهاد می‌دهد، که تصاویر MR چندبعدی را در فضای ورودی، کانال به کانال ترکیب می‌کند، تا یک بازنمایی ویژگی ترکیبی را یاد بگیرد. در مقایسه با قطعه‌بندی مرجع که مشکل پیچیدگی شبکه دارد و همبستگی میان وظایف قطعه‌بندی ترتیبی را نادیده می‌گیرد، این روش، قطعه‌بندی تومور مغز را به سه وظیفه‌ی مختلف اما مرتبط، تجزیه می‌کند. هر وظیفه دارای یک لایه‌ی پیچشی مستقل، یک لایه‌ی طبقه‌بندی، یک لایه‌ی زیان و داده‌های ورودی مختلف است. بر اساس یادگیری برنامه‌ی درسی، که به معنای افزایش تدریجی دشواری وظایف آموزشی است، آنها برای بهبود کیفیت هم‌گرایی مدل، با آموزش اولین وظیفه، تا زمانی که منحنی زیان مسطح شود، از یک استراتژی مؤثر، استفاده کردند، سپس داده‌ی اول و دوم، همراه با بعد دسته، به‌عنوان ورودی به وظیفه‌ی دوم، متصل شدند. عملکرد وظیفه‌ی سوم نیز مانند وظیفه‌ی دوم است. در این روش، نه تنها پارامترهای مدل، بلکه داده‌های آموزشی هم از یک وظیفه‌ی ساده‌تر به یک وظیفه‌ی دشوارتر، منتقل می‌شوند. رویکرد پیشنهادی در ابتدا روی مجموعه آزمایشی

BRATS 2015، دسته‌بندی انجام می‌دهد و در مجموعه داده‌ی BRATS 2017 به عملکرد بالایی

دست پیدا می‌کند.

هنگام قطعه‌بندی نواحی مختلف در یک تصویر، احتمال دارد به میدان ورودی مختلفی نیاز باشد. برای مثال، نواحی بزرگ، ممکن است به یک میدان ورودی بزرگ به ضرر جزئیات دقیق نیاز داشته باشد، در حالی که نواحی کوچک ممکن است نیازمند اطلاعات محلی با رزولوشن بالا باشند. کین^{۱۱} و همکاران، با استفاده از پردازش چند مقیاسی، لایه‌ی پیچشی خودتمرکز را برای بهبود توانایی‌های شبکه‌های عصبی، پیشنهاد دادند. آنها، بعد از ادغام تصاویر چندبعدی در فضای ورودی، لایه‌ی پیچشی خودتمرکز را با استفاده از لایه‌های پیچشی متعدد با نرخ انبساط مختلف اعمال کردند، تا اندازه‌ی میدان ورودی را تغییر دهند. لایه‌ی پیچشی خودتمرکز، می‌تواند اهمیت هر مقیاس را به هنگام پردازش مکان‌های مختلف یک تصویر، نشان دهد. برای انتخاب مقیاس بهینه از یک مکانیسم توجه استفاده کردند. لایه‌ی خودتمرکز پیشنهادی می‌تواند به راحتی، در شبکه‌های موجود ادغام شود تا عملکرد یک مدل را بهبود ببخشد. روش پیشنهادی روی کارهای چالش برانگیز قطعه‌بندی اندام‌های متعدد در CT لگنی و قطعه‌بندی تومور مغز در MRI، به عملکرد امیدبخشی دست پیدا کرده است.

به دلیل موفقیت شبکه‌ی مولد تخصصی (GAN) که یک بازی حداقل - حداکثر یا مین - مکس را بین تولیدکننده و تشخیص دهنده، شبیه‌سازی می‌کند، روش‌هایی وجود دارند که استفاده از تشخیص-دهنده را به صورت یک محدودیت اضافی پیشنهاد می‌دهند که عملکرد طبقه‌بندی را بهبود می‌بخشد. با ترکیب تصاویر چندبعدی، به عنوان ورودی‌های چندکاناله، دو شبکه‌ی جداگانه را آموزش دادند: یک U-Net باقی‌مانده به عنوان شبکه‌ی مولد و یک شبکه‌ی تشخیص دهنده، شبکه‌ی قطعه‌بندی یک قطعه‌بندی تولید خواهد کرد، در حالی که شبکه‌ی تشخیص دهنده، میان قطعه‌بندی‌های تولیدشده و

¹¹ Qin

ماسک‌های حقیقی، تمییز قائل می‌شود. تشخیص‌دهنده، یک شبکه‌ی کم عمق است که حاوی سه بلوک کانولوشن سه‌بعدی است که به دنبال هر یک از این بلوک‌ها، یک لایه‌ی حداکثر – ادغام می‌آید. به‌منظور دستیابی به یک قطعه‌بندی قوی، از طریق مرزها یا کانتورها محدودیت‌هایی اضافی به مدل اضافه کردند. فاصله‌ی هاسدورف^{۱۲} در بین کانتورهای حقیقی و کانتورهای پیشبینی، به‌عنوان معیاری برای عدم شباهت، مورد استفاده قرار می‌گیرد. روش پیشنهادی، روی مجموعه داده‌ی BraTS 2018 ارزیابی شد و نتایج قابل رقابتی حاصل شد که نشان می‌دهد نتایج قطعه‌بندی خام، می‌تواند با گنجاندن محدودیت‌های اضافی در کانتورها و آموزش تخصصی، بهبود پیدا کنند. هوو^{۱۳} و همکاران، از PatchGAN، به‌عنوان یک تشخیص‌دهنده‌ی اضافی برای نظارت بر روند آموزش شبکه، استفاده کردند. روش مبتنی بر GAN می‌تواند به دلیل محدودیت اضافی تشخیص‌دهنده، به قطعه‌بندی قوی دست پیدا کند، اما آموزش این تشخیص‌دهنده، حافظه‌ی بیشتری می‌طلبد.

استراتژی ترکیب سطح ورودی، می‌تواند حداکثر اطلاعات تصویر اصلی را حفظ کند و ویژگی ذاتی تصویر را یاد بگیرد. استفاده از شبکه‌های قطعه‌بندی ترتیبی، امکان استفاده از استراتژی‌های مختلف را فراهم می‌کند، مثلاً قطعه‌بندی مبتنی بر GAN، چندوظیفه‌ای، چندنما و چندمقیاسی، تا به‌طور کامل، بازنمایی ویژگی از تصاویر چندبعدی را مورد بهره‌برداری قرار دهد.

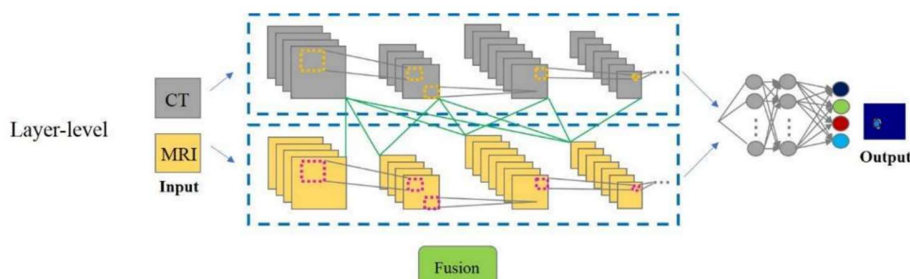
۴.۲. شبکه‌ی ترکیب سطح لایه

در استراتژی ترکیب سطح لایه، تصاویر یک یا دو بعدی، به‌عنوان تنها ورودی برای آموزش شبکه‌ی قطعه‌بندی فردی، مورد استفاده قرار می‌گیرند و سپس این بازنمایی ویژگی‌های فردی آموخته شده، در لایه‌های شبکه ترکیب خواهند شد، در نهایت، نتایج ترکیب‌شده به لایه‌ی تصمیم داده می‌شود تا نتیجه‌ی نهایی طبقه‌بندی به دست بیاید. شبکه‌ی ترکیب سطح لایه، می‌تواند به‌طور مؤثر، تصاویر

¹² Hausdorff

¹³ Huo

چندبعدی را ادغام کرده و کاملاً مورد استفاده قرار دهد. شکل ۷، معماری کلی شبکه‌ی کار قطعه‌بندی ترکیب سطح لایه را توضیح می‌دهد.



شکل ۷. معماری کلی شبکه‌ی ترکیب سطح لایه.

به طور مثال، قطعه‌بندی تومور مغز را در چند توالی MRI انتخاب می‌کنیم تا این نوع ترکیب را نشان دهیم. مشخص است که T1 MRI و T1c و T2 و Flair، برای قطعه‌بندی هسته‌ی تومور بدون ورم صفاتی مناسب هستند، در حالی که T2 و Flair، برای قطعه‌بندی ورم صفاتی مناسب هستند. چن و همکاران، یک شبکه‌ی قطعه‌بندی تومور مغزی چندبعدی با گذرگاه دوگانه را پیشنهاد می‌دهند. اولین گذرگاه، از T2 و Flair برای استخراج ویژگی نسبی، برای قطعه‌بندی کل تومور از پس‌زمینه استفاده می‌کند و دومین گذرگاه، از T1 و T1c برای آموزش همان شبکه‌ی قطعه‌بندی استفاده می‌کند تا بازنمایی ویژگی نسبی دیگر را بیاموزد و سپس، ویژگی‌های هر دو مسیر، ترکیب می‌شوند و در نهایت، به یک طبقه‌بند بیشینه‌ی هموار چهار کلاس، خورانده می‌شوند تا پس‌زمینه، ED، ET و NCR/NET را قطعه‌بندی کند. شبکه‌ی قطعه‌بندی مسیر دوگانه، می‌تواند اطلاعات ویژگی مؤثری را از حالت‌ها و ابعاد مختلف استخراج کند و به یک نتیجه‌ی قطعه‌بندی دقیق، دست پیدا کند.

دولز^{۱۴} و همکاران، یک شبکه‌ی عصبی پیچشی سه بعدی بر اساس DenseNets پیشنهاد می‌دهند که تعریف اتصال متراکم را به قطعه‌بندی چندبعدی، بسط می‌دهد. هر حالت تصویر، دارای یک مسیر است و اتصال متراکم هم در لایه‌هایی با مسیر یکسان و هم لایه‌های دارای مسیرهای متفاوت، وجود دارد.

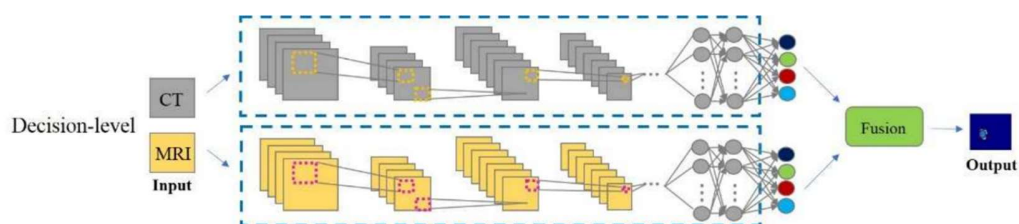
¹⁴ Dolz

بنابراین، شبکه‌ی پیشنهادی، می‌تواند بازنمایی‌های ویژگی پیچیده‌تری را در میان ابعاد، بیاموزد. نتایج آزمایشات گسترده، روی دو چالش قطعه‌بندی بافت مغز چندبعدی بسیار رقابتی و متفاوت: iSEG 2017 و MRBrainS 2013، نشان می‌دهند که روش پیشنهادی، در مقایسه با جدیدترین شبکه‌های قطعه‌بندی، بهبود چشمگیری به دست آورده و روی هر دو معیار، رتبه‌ی بالایی کسب می‌کند.

با الهام از مرجع، دولز و همکاران یک معماری برای محلی‌سازی و قطعه‌بندی IVD (دیسک بین مهرهای) در MRI چندحالتی، پیشنهاد می‌دهد. هر بعد یا حالت MRI در یک مسیر متناظر پیشنهاد می‌شود تا از بازنمایی ویژگی آن استفاده کند. این شبکه هم درون هر یک از مسیرها و هم در عرض مسیرهای مختلف، به‌طور متراکم متصل می‌شود، سپس به مدل این اختیار انتخاب جایی که می‌آموزد و نحوه‌ی پردازش و ترکیب ابعاد مختلف را می‌دهد. همچنین، ماژول‌های استاندارد U-Net را با گسترش ماژول‌های آغازین، با استفاده از دو بلوک پیچشی با کانولوشن‌های منبسط شده از یک مقیاس مختلف بهبود می‌دهد. از این طریق، به کنترل اطلاعات زمین‌های چندمقیاسی کمک می‌کند. به‌طور خلاصه، در شبکه‌ی قطعه‌بندی ترکیب سطح لایه، DenseNets پرکاربردترین شبکه‌ها هستند که سه مزیت زیر را دارند. ابتدا، ارتباطات مستقیم میان تمام لایه‌ها به بهبود جریان اطلاعات و گرادین‌های کل شبکه، کمک می‌کند که موجب کاهش مسئله‌ی شیب نزولی می‌شود. دوم، مسیرهای کوتاه به تمام نگاشت‌های ویژگی در این معماری، نظارت عمیق ضمنی را وارد می‌کنند. سوم اینکه، اتصالات متراکم، یک اثر تنظیم‌کننده دارند که ریسک برازش بیش از حد روی وظایفی با مجموعه داده‌های آموزشی کوچک‌تر را کاهش می‌دهد. بنابراین، DenseNets به بهبود اثربخشی و کارایی در شبکه‌ی قطعه‌بندی ترکیب سطح لایه، کمک می‌کند. در شبکه‌ی قطعه‌بندی ترکیب سطح لایه، ارتباط میان لایه‌های مختلف، می‌تواند روابط پیچیده‌ی میان حالت‌ها و ابعاد را بدست بیاورد که از بازنمایی ویژگی تصاویر چندبعدی، استفاده‌ی کامل را می‌کند.

۴.۳. شبکه‌ی ترکیب سطح تصمیم

در شبکه‌ی قطعه‌بندی ترکیب سطح تصمیم، مانند ترکیب سطح لایه، از هر یک از تصاویر هر بعد، به‌عنوان تنها ورودی شبکه‌ی قطعه‌بندی استفاده می‌شود. شبکه‌ی واحد می‌تواند اطلاعات منحصر به فرد حالت متناظر را بهتر مورد استفاده قرار دهد. سپس خروجی‌های شبکه‌های فردی ادغام می‌شوند تا نتیجه‌ی قطعه‌بندی نهایی به‌دست بیاید. شبکه‌ی قطعه‌بندی ترکیب سطح تصمیم، به این منظور طراحی می‌شود که به طور مستقل، اطلاعات تکمیلی از ابعاد مختلف را بیاموزد، زیرا تصاویر چندحالتی، به دلیل تکنیک‌های اکتساب تصویر مختلف، در فضای تصویر اصلی خود، اطلاعات مکمل مستقیم کمی دارند. شکل ۸ معماری کلی شبکه‌ی قطعه‌بندی ترکیب سطح لایه را نشان می‌دهد.



شکل ۸: معماری کلی شبکه‌ی ترکیب سطح تصمیم

به‌عنوان مثال، برای استفاده‌ی مؤثر چندبعد از T_1 ، T_2 و حالت ناهمسانگردی تابعی (FA)، نی و همکاران، یک معماری شبکه‌ی چند FCN جدید برای قطعه‌بندی بافت مغز نوزادان پیشنهاد دادند (ماده‌ی سفید (WM)، ماده‌ی خاکستری (GM) و مایع مغزی نخاعی (CSF)). به جای فقط ترکیب داده‌ی سه بعد از فضای ورودی، آنها یک شبکه را برای هر بعد آموزش دادند و سپس ویژگی‌های متعدد حالت را از لایه‌ی بالای هر شبکه، ترکیب کردند. نتایج نشان داده که روش پیشنهادی، از نظر دقت، بسیار بهتر از روش‌ها قبلی عمل کرده است.

برای ترکیب سطح تصمیم، استراتژی‌های ترکیب زیادی، پیشنهاد شده‌اند. اکثر آنها، بر اساس رأی‌گیری اکثریت و میانگین‌گیری هستند. در استراتژی میانگین‌گیری، کامنیتاس و همکاران، سه شبکه را به طور مجزا آموزش می‌دهند و سپس، میانگین اعتماد شبکه‌های فردی را محاسبه می‌کنند. قطعه‌بندی

نهایی با تخصیص هر واکسل به بیش‌ترین اعتماد حاصل می‌شود. برای استراتژی رأی‌گیری اکثریت، برچسب نهایی یک واکسل، به اکثریت برچسب‌های شبکه‌های فردی، بستگی دارد.

خصوصیات آماری حالت‌های مختلف، متفاوت است که باعث شده برای یک مدل واحد، مستقیماً پیدا کردن روابط میان ابعاد، سخت شود. بنابراین، در شبکه‌ی قطعه‌بندی ترکیب سطح تصمیم، شبکه‌های قطعه‌بندی متعدد می‌توانند استفاده شوند تا از ویژگی‌های چند حالت استفاده‌ی کامل شود. ایگن و همکاران روش‌های ترکیب مختلف را روی مسئله‌ی قطعه‌بندی تومور مغزی، از نظر حافظه و عملکرد بررسی می‌کنند. از نظر استفاده‌ی حافظه، استراتژی‌های ترکیب سطح تصمیم به حافظه‌ی بیشتری نیاز دارند، زیرا مدل، ویژگی‌ها را بعداً اضافه می‌کند و لایه‌ها به پارامترهای بیشتری نیاز دارند تا عملگر پیچش و سایر عملگرها را انجام دهند. باین‌حال، ترکیب آخر یا متعاقب، می‌تواند عملکرد بهتری کسب کند، زیرا هر بعد به‌عنوان ورودی یک شبکه استفاده می‌شود که در مقایسه با شبکه‌ی ترکیب سطح ورودی، می‌تواند اطلاعات ویژگی پیچیده و مکملی را یاد بگیرد.

۵. مسائل رایج

۵.۱. بیش‌برازش

یکی از محدودیت‌های قطعه‌بندی تصاویر پزشکی، کمیابی داده و اطلاعات است که معمولاً منجر به مسئله‌ی بیش‌برازش می‌شود که اشاره به مدلی دارد که روی داده‌های آموزشی عملکرد خوبی دارد اما برای داده‌های جدید، خیلی خوب کار نمی‌کند. اکثر اوقات، تعداد زیادی از برچسب‌های آموزشی، برای تحلیل تصویر پزشکی، موجود نیستند، زیرا برچسب‌گذاری مجموعه داده نیازمند افراد متخصص در این حوزه است و زمان بر و گاهی اوقات مستعد خطاست. هنگام آموزش شبکه‌های عصبی پیچیده با داده‌های آموزشی محدود، باید بسیار مراقب باشیم که مسئله‌ی بیش‌برازش بوجود نیاید. پیچیدگی یک شبکه‌ی عصبی، با هم ساختار و هم پارامترهای آن، تعریف می‌شود. بنابراین، می‌توانیم پیچیدگی

معماری شبکه را با کاهش لایه‌ها یا پارامترها کم کنیم یا روی روش‌های تمرکز کنیم که به جای تغییر معماری شبکه، تعداد داده‌های آموزشی را به طور مصنوعی، افزایش می‌دهند. مورد معمولاً برای تولید تصاویر مصنوعی جدید استفاده می‌شود که تبدیلات داده و واقعیت متناظر با آنها را انجام می‌دهد که شامل عملگرهای مقیاس‌گذاری، چرخش، تفسیر، تغییر روشنایی، تغییر شکل الاستیک، چرخش افقی و وارون کردن است (برای جزئیات بیشتر به بخش ۳.۳ افزایش داده‌ها، رجوع کنید).

۵.۲. عدم تعادل کلاس

یکی از چالش‌های مهم در تحلیل تصاویر پزشکی، کار با داده‌های نامتعادل است. در زمینه‌ی تصویربرداری پزشکی، این مسئله حتی پررنگ‌تر است. برای مثال، یک قطعه‌بندی تومور مغزی یا قطعه‌بندی ضایعه‌ی ماده‌ی سفید، ناحیه‌ی طبیعی مغز بزرگتر از ناحیه‌ی غیرطبیعی است. آموزش با داده‌های نامتعادل، می‌تواند منجر به یک شبکه‌ی قطعه‌بندی ناپایدار شود که به سمت کلاس‌هایی با یک ناحیه‌ی بزرگ، بایاس پیدا کرده است. جدول ۴، کلاس‌های موجود در داده‌های آموزشی BraTS 2017 را نشان می‌دهد، تعداد موارد مثبت (ED, NEC/NET, و ET) و موارد منفی (پس‌زمینه) بسیار متعادل هستند و پس‌زمینه بسیار برجسته است. در نتیجه، انتخاب توابع زیان در شبکه‌های قطعه‌بندی بسیار مهم است، مخصوصاً زمانی که با مسائل بسیار نامتعادل کار می‌کنیم. ما چند نوع تابع زیان را نشان می‌دهیم که به‌طور فردی یا در ترکیب با شبکه‌های قطعه‌بندی تصاویر پزشکی، بسیار استفاده می‌شوند. از سطح داده، این مسئله را می‌توان با نمونه‌گیری مجدد فضای داده، حل کرد. سه رویکرد اصلی وجود دارد: نمونه‌برداری کند از کلاس منفی یا نمونه‌برداری زیاد از کلاس منفی و SMOTE (تکنیک نمونه‌گیری بسیار زیاد اقلیت مصنوعی) که نمونه‌های مصنوعی را در امتداد پاره‌خط تولید می‌کند که نمونه‌های کلاس اقلیت را متصل می‌کند. دنبال کردن این روش‌ها ساده است اما ممکن است داده‌های مهمی را حذف کنند یا داده‌هایی اضافی به مجموعه آموزشی، اضافه کنند.

Region	Background	NET/NCR	Edema	Enhancing tumor
Percentage	99.12	0.28	0.40	0.20

جدول ۴. توزیع کلاس‌های روی مجموعه آموزشی ۲۰۱۷ NET.BraTS: تومور غیرپیشرفته، NCR: نکروزی

روش مبتنی بر نمونه‌گیری دسته‌ای، می‌تواند مسئله‌ی داده‌ی متعادل را هم کاهش دهد. برای مثال، کامنیتاس و همکاران، برای کاهش مسئله‌ی عدم تعادل کلاس، یک استراتژی متعادل را پیشنهاد می‌دهند. آنها تکه‌های آموزشی را با احتمال متمرکز شدن ۵۰ درصد، یا روی واکسل‌های سالم یا ضایعه، استخراج می‌کنند. کریگس و همکاران از استراتژی ضایعه محور استفاده می‌کنند، که در آن، تمام تکه‌های آموزشی، از ناحیه‌ی متمرکز در یک واکسل ضایعه استخراج می‌شود. علاوه بر این، یک آفست (حاشیه جبران) تصادفی به یک نمونه ۴۷۵ نقطه‌ای اضافه می‌شود تا از بایاس اجتناب شود، جایی که یک واکسل ضایعه، همیشه در مرکز دسته انتظار می‌رود و سپس به افزایش داده نیز کمک می‌کند.

برای سطح الگوریتم، هاوای و همکاران، یک روش آموزشی دو مرحله‌ای پیشنهاد می‌دهند. که ابتدا یک مجموعه داده‌ی دسته‌ای می‌سازد به طوری که، با در نظر گرفتن تنوع در تمام کلاس‌ها، تمام برچسب‌ها به احتمال یکسان رخ می‌دهند و سپس تنها لایه‌ی خروجی، حفظ می‌شود تا احتمالات خروجی را به درستی، تنظیم کند. به این ترتیب، مسئله‌ی عدم تعادل کلاس، حل می‌شود. رویکرد دیگر شامل استفاده از یک قطعه‌بندی چند وظیفه است که وظیفه‌ی قطعه‌بندی چند کلاسه‌ی پیچیده را به وظایف ساده تجزیه می‌کند، زیرا هر وظیفه‌ی آموزشی، فقط یک ناحیه را قطعه‌بندی می‌کند، که در آن توزیع برچسب، نسبت به کلاس‌های متعدد قطعه‌بندی در یک زمان، عدم تعادل کمتری خواهد داشت. چند رویکرد وجود دارد که مسئله‌ی عدم تعادل کلاس را از طریق یک یادگیری گروهی، با ترکیب طبقه‌بندهای یکسان یا مختلف، حل می‌کند تا توانایی تعمیم آنها را بهبود دهد. در غیر اینصورت، توابع زیان می‌توانند، این مسئله را با تغییر توزیع داده‌های آموزشی، کم کنند. ما آنها را در زیر نشان می‌دهیم.

زیان آنتروپی متقابل (CE): آنتروپی متقابل، برای کار طبقه‌بندی تصاویر، بیش‌ترین استفاده را دارد. که با معادله‌ی (۱) محاسبه می‌شود. از آنجایی‌که، زیان آنتروپی متقابل، به‌طور فردی و مجزا، پیش‌بینی کلاس برای هر بردار پیکسل را ارزیابی می‌کند و سپس میانگین تمام پیکسل‌ها را محاسبه می‌کند، اگر یک بازنمایی کلاس نامتعادل در تصویر موجود باشد، این کار می‌تواند منجر به خطا شود. لانگ و همکاران، وزن‌دهی یا نمونه‌گیری تابع زیان را برای هر کانال خروجی پیشنهاد می‌دهند تا مسئله‌ی عدم تعادل کلاس را کم کرده باشند.

$$Loss_{CE} = - \sum_{i \in N} \sum_{l \in L} y_i^{(l)} \log \hat{y}_i^{(l)} \quad (1)$$

که در آن، N مجموعه‌ی تمام نمونه‌ها و L مجموعه‌ی تمام برچسب‌هاست، $y_i^{(l)}$ رمزگذاری یک بارز (۰ یا ۱) برای نمونه و برچسب است، $\hat{y}_i^{(l)}$ احتمال پیش‌بینی برای زوج نمونه/برچسب است. آنتروپی متقابل وزندار (WCE): از آنجایی‌که نواحی پس‌زمینه، بر مجموعه آموزشی غالب هستند، منطقی است وزن‌های کلاس‌های متعدد را در آنتروپی متقابل، قرار دهیم که به صورت زیر تعریف می‌شود:

$$Loss_{WCE} = - \sum_{i \in N} \sum_{l \in L} w_i y_i^{(l)} \log \hat{y}_i^{(l)} \quad (2)$$

که در آن w_i بیانگر وزن تخصیص داده شده به i امین برچسب است.

زیان Dice یا تاس (DL): زیان Dice یک تابع زیان محبوب در زمینه‌ی قطعه‌بندی تصاویر پزشکی است که معیاری از هم‌پوشانی میان نمونه‌ی پیش‌بینی شده و نمونه‌ی واقعی است. این معیار بین ۰ و ۱ متغیر است که در آن امتیاز ۱، بیانگر همپوشانی کامل است و به‌صورت زیر تعریف می‌شود:

$$Loss_{DL} = 1 - 2 \frac{\sum_{l \in L} \sum_{i \in N} y_i^{(l)} \hat{y}_i^{(l)} + \epsilon}{\sum_{l \in L} \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \epsilon} \quad (3)$$

که در آن ϵ یک ثابت کوچک است تا از تقسیم بر صفر جلوگیری شود.

Dice تعمیم یافته (GDL): Sudre و همکاران استفاده از خواص تعادل مجدد هم پوشانی Dice

تعمیم یافته را پیشنهاد دادند که در مرجع (۴) تعریف شده است، یک تابع زیان یادگیری عمیق قوی و دقیق برای کارهای عدم تعادل است. نویسندگان رفتار توابع زیان Dice، زیان آنتروپی متقابل و زیان dice تعمیم یافته را در وجود نرخ های مختلف عدم تعادل در کارهای قطعه بندی دوبعدی و سه بعدی، بررسی می کنند. نتایج نشان می دهد که GDL نسبت به سایر توابع زیان، قویتر است.

$$Loss_{GDL} = 1 - 2 \frac{\sum_{l \in L} w_l \sum_{i \in N} y_i^{(l)} \hat{y}_i^{(l)} + \epsilon}{\sum_{l \in L} w_l \sum_{i \in N} (y_i^{(l)} + \hat{y}_i^{(l)}) + \epsilon} \quad (4)$$

تابع زیان کانونی (FL): زیان کانونی، در اصل برای کار تشخیص معرفی شده است. که مدل را ترغیب می کند به مثال های آسان وزن کمی دهد و روی آموزش موارد منفی سخت، تمرکز کند. به طور رسمی، این زیان، با تعریف یک عامل تعدیل کننده به تابع زیان آنتروپی متقابل و یک پارامتر برای توازن کلاس، تعریف می شود:

$$Loss_{FL}(p_t) = -\alpha_t(1 - p_t)^y \log(p_t) \quad (5)$$

$$p_t = \begin{cases} 1 & \text{if } y = 1 \\ 1 - p_t & \text{otherwise} \end{cases} \quad (6)$$

که در آن $y \in \{-1, +1\}$ ، کلاس واقعی است و $p_t \in [0, 1]$ احتمال برآورد شده برای کلاسی با برچسب $y=1$ است. پارامتر تمرکز γ به طور هموار نرخی را تنظیم می کند که در آن وزن نمونه های آسان، کم می شود، تنظیم $\gamma < 0$ می تواند زیان نسبی را برای نمونه های به خوبی طبقه بندی شده، کاهش می دهد، تمرکز را روی نمونه های سخت و نمونه هایی قرار می دهد که درست طبقه بندی نشده اند، تمرکز کانونی برابر با زیان آنتروپی متقابل اصلی است برای زمانی که $\gamma=0$.

Article	Pre-processing	Data	Network	Fusion level	Results (DSC)	Database
[49]	Normalization	3D	CNN	Input	whole/core/enhanced	BraTS15
	Bias Field Correction	Patch	CRF		0.84/0.66/0.63	
[20]	Normalization	2D	CNN	Input	whole/core/enhanced	BraTS13
	Bias Field Correction	Patch			0.84/0.71/0.57	
[41]*	Normalization	3D	U-Net	Input	whole/core/enhanced	BraTS15
	Data Augmentation	Patch	ResNet		0.85/0.74/0.64	BraTS17
					0.85/0.77/0.64	
[50]	Normalization	3D	FCN	Input	whole/core/enhanced	BraTS13
	Bias Field Correction	Patch	CRF		0.86/0.73/0.62	BraTS15
			RNN		0.84/0.73/0.62	BraTS16
					4/3/2(rank)	
[57]	Normalization	3D	U-Net	Input	whole/core/enhanced	BraTS15
			ResNet		0.87/0.75/0.64	
[48]	Normalization	2D	U-Net	Input	whole/core/enhanced	BraTS17
	Bias Field Correction	Slice	ResNet		0.87/0.77/0.78	
[42]	Normalization	3D	U-Net	Input	whole/core/enhanced	BraTS18
	Data Augmentation	Patch	ResNet		0.87/0.80/0.77	
[45]	Normalization	2D	CNN	Input	whole/core/enhanced	BraTS15
	Data Augmentation	Patch	FCN		0.89/0.77/0.80	
	Bias Field Correction					
[20]	Normalization	3D	CNN	Input	whole/core/enhanced	BraTS13
	Bias Field Correction				0.88/0.81/0.76	
[75]	Normalization	2D	FCN	Input	whole/core/enhanced	BraTS16
	Data Augmentation				0.87/0.81/0.72	
[52]*	Normalization	3D	U-Net	Input	whole/core/enhanced	BraTS17
	Bias Field Correction		FCN		0.88/0.78/0.72	
			DeepMedic			
[55]*	Normalization	3D	U-Net	Input	whole/core/enhanced	BraTS18
	Data Augmentation		VAE		0.88/0.81/0.76	
[76]	N/A	2D	CNN	Input	0.9112	IVD
		Slice	ResNet			
[56]*	Normalization	3D	U-Net	Input	0.59 ± 0.31	ISLES15
		Patch	ResNet		0.84 ± 0.10	(SISS/SPES)
[77]*	Normalization	3D	SVM	Input	CSF/WM/GM	MRBrainS13
		Patch			0.78 0.88 0.84	
[44]*	N/A	3D	CNN	Layer	CSF/WM/GM	iSEG-2017
		Patch	DenseNet		0.95/0.91/0.90	MRBrainS13
					0.84/0.90/0.86	
[78]*	Normalization	3D	DenseNet	Layer	CSF/WM/GM	iSEG-2017
		Patch			0.96/0.91/0.90	
[46]	N/A	2D	U-Net	Layer	0.9191 ± 0.0179	IVD
		Slice	DenseNet			
[47]	N/A	2D	FCN	Decision	CSF/WM/GM	Private data

جدول ۵. خلاصه‌ای از روش‌های یادگیری عمیق برای قطعه‌بندی تصاویر پزشکی چندبعدی، موارد

پررنگ نشانگر بهترین عملکرد در این چالش است. کلمات اختصاری در نتایج عبارتند از: مایع مغزی نخاعی (CSF)، ماده‌ی خاکستری (GM)، ماده‌ی سفید (WM)، نماد * روشی را نشان می‌دهد که کد آن موجود است.

۶. بحث و نتیجه‌گیری

در بخش‌های بالا، یک مجموعه‌ی بزرگ از جدیدترین شبکه‌های قطعه‌بندی تصویر پزشکی چندبعدی که بر پایه‌ی یادگیری عمیق هستند را ارائه دادیم. این موارد در جدول ۵ خلاصه شده‌اند. برای چالش BraTS، این روش‌ها از سال ۲۰۱۳ گردآوری شده‌اند، زیرا روش‌های یادگیری ماشین از آن زمان استفاده شده‌اند. مجموعه داده‌های تصاویر پزشکی چندبعدی یا چندحالتی برای کار قطعه‌بندی، کمیاب هستند، پرکاربردترین مجموعه داده، BraTS است که از سال ۲۰۱۲ پیشنهاد شده است. برای قطعه‌بندی آنها، بهترین روش فعلی در گزارش شده است، آنها از استراتژی سطح ورودی برای ادغام مستقیم حالات مختلف در فضای ورودی، استفاده می‌کنند، ساختار رمزگذار - رمزگشای CNN را در ترکیب با یک شعبه VAE اضافی (رمزگذار خودکار متغیر) به بخش رمزگذار، اعمال می‌کنند. شعبه‌ی VAE می‌تواند تصویر ورودی را بازسازی کند و ویژگی‌های نقطه انتهایی رمزگذار را بهتر استخراج کند. همچنین، یک راهنمایی و تنظیم مجدد برای بخش رمزگذار، فراهم می‌کند. نویسندگان نشان می‌دهند که تکنیک‌های افزایش داده‌ی پیچیده‌تر، تکنیک‌های پس پردازش داده یا شبکه‌ی عمیق‌تر، عملکرد شبکه را بیشتر بهبود خواهد داد که بدان معناست که معماری شبکه، نقش مهمی در شبکه‌ی قطعه‌بندی ایفا می‌کند تا سیار عملیات پردازش داده.

برای قطعه‌بندی تصاویر پزشکی چندبعدی، استراتژی ترکیب، نقش مهمی در دستیابی به یک نتیجه‌ی قطعه‌بندی دقیق‌تر ایفا می‌کند. استراتژی ترکیب تصویر پیچشی، یک نگاشت مستقیم میان تصاویر مبدأ و تصاویر هدف، یاد می‌گیرد، استراتژی ترکیب شامل دو مرحله اصلی است: اندازه‌گیری سطح فعالیت و قانون ترکیب. اندازه‌گیری سطح فعالیت، با طراحی فیلترهای محلی، برای استخراج جزئیات فرکانس بالا، اجرا می‌شود و سپس اطلاعات تصاویر مبدأ مختلف، با استفاده از قواعد طراحی شده، مقایسه می‌شوند تا یک تصویر واضح به دست آید. برای دستیابی به عملکرد بهتر، این مسائل پیچیده‌تر می‌شوند، بنابراین، ارائه‌ی یک استراتژی ترکیب ایده‌آل به‌صورت دستی، که کاملاً به مسائل مهم توجه کند، دشوار است. برای این منظور، یک شبکه‌ی مبتنی بر یادگیری عمیق، مستقیماً نگاشت را رمزگذاری می‌کند. روش‌های مبتنی بر یادگیری

عمیق، از سه‌نظر بهتر عمل می‌کنند. اول اینکه، شبکه‌های مبتنی بر یادگیری عمیق، یک بازنمایی ویژگی سلسله‌مراتبی پیچیده و انتزاعی برای داده‌های تصویر، یاد می‌گیرند تا بر مشکل طراحی ویژگی دستی، غلبه کنند. دوماً، شبکه‌های مبتنی بر یادگیری عمیق، می‌توانند روابط پیچیده‌ی میان حالت‌های مختلف را با استفاده از لایه‌ی شبکه‌ی سلسله‌مراتبی نشان دهند، مانند استراتژی ترکیب سطح لایه. سوم، استراتژی ترکیب و تبدیل تصویر، در استراتژی ترکیب پیچشی، می‌تواند به‌طور مشترک، به‌وسیله‌ی آموزش یک مدل یادگیری عمیق، تولید شود، به این ترتیب، تعدادی از معماری‌های شبکه‌ی یادگیری عمیق را می‌توان با طراحی یک استراتژی ترکیب تصویر مؤثر، بررسی کرد. بنابراین، روش مبتنی بر یادگیری عمیق، در تولید نتایج ترکیب بهتر، نسبت به روش‌های پیچشی، پتانسیل بالایی دارند.

انتخاب یک استراتژی ترکیب یادگیری عمیق، هنوز مسئله‌ای مهم است. در چالش BraTS سال‌های ۲۰۱۳ تا ۲۰۱۸، تمام روش‌ها، از ترکیب سطح ورودی استفاده کردند تا تصاویر MR مختلف را به‌طور مستقیم در فضای ورودی، ادغام کنند، که کار ساده‌ای است و می‌تواند ویژگی ذاتی تصویر را حفظ کند و به روش موردنظر اجازه می‌دهد روی طرح‌های معماری شبکه‌ی قطعه‌بندی متعاقب، تمرکز کند، مثلاً استراتژی‌های مبتنی بر GAN، چندمقیاس، چندنما، چندوظیفه. درحالی‌که این استراتژی فقط حالت‌ها و ابعاد موجود در فضای ورودی را متراکم می‌کند، روابط میان حالات مختلف را مورد بهره‌برداری قرار نمی‌دهد. برای ترکیب سطح لایه، با اتصال متراکم میان لایه‌ها، استراتژی ترکیب DenseNet را به‌عنوان شبکه‌ی پایه انتخاب می‌کند. اتصال میان لایه‌های مختلف، می‌تواند روابط پیچیده‌ی میان ابعاد را بدست بیاورد که در مقایسه با ادغام ابعاد مختلف در فضای ورودی، می‌تواند به شبکه‌ی قطعه‌بندی کمک کند تا اطلاعات ارزشمندتری به دست بیاورد و عملکرد بهتری کسب کند. استراتژی ترکیب سطح تصمیم، در مقایسه با ترکیب سطح ورودی، می‌تواند عملکرد بهتری کسب کند، زیرا از هر بُعد برای آموزش یک شبکه‌ی واحد استفاده می‌شود تا بازنمایی ویژگی مستقل را بیاموزد، درحالی‌که این امر، مستلزم حافظه و زمان محاسباتی زیاد است. در مقایسه با دو استراتژی ترکیب آخر، استراتژی ترکیب لایه، ظاهراً بهتر است، زیرا از اتصال متراکم میان لایه‌ها

می‌توان برای بهره‌برداری اطلاعات تکمیلی و پیچیده‌تر استفاده کرد که به بهبود آموزش شبکه کمک می‌کند، درحالی‌که ترکیب سطح تصمیم، فقط بازنمایی ویژگی مستقل در یک حالت را یاد می‌گیرد. از آنجایی که نتایج این سه استراتژی ترکیب، از داده‌های یکسانی حاصل نشده‌اند، مقایسه‌ی آنها از نظر عملکرد، دشوار است. از نظر متدولوژیکی، هر روش، مزایا و معایب خودش را دارد.

اگرچه، مزایای این استراتژی‌های ترکیب مبتنی بر یادگیری عمیق، بر اساس کارهای قبلی را مشاهده کردیم، هنوز می‌توانیم ببینیم که موانعی در زمینه‌ی قطعه‌بندی تصویر پزشکی چندبعدی بر پایه‌ی یادگیری عمیق وجود دارد. مشخص است که شبکه‌های ترکیب چندبعدی، معمولاً بهتر از شبکه‌ی تک حالت هستند. مسئله این است که چگونه حالات یا ابعاد مختلف را ترکیب کنیم تا بهترین سازش را برای یک قطعه‌بندی دقیق، به دست بیاوریم. از این رو، نحوه‌ی طراحی شبکه‌های چندبعدی برای ترکیب مؤثر ابعاد مختلف، نحوه‌ی بهره‌برداری از روابط پنهان میان ابعاد مختلف و نحوه‌ی ادغام اطلاعات متعدد در شبکه‌ی قطعه‌بندی، برای بهبود عملکرد قطعه‌بندی، می‌تواند موضوعاتی برای پژوهش‌های آینده باشد.

مسئله‌ی دیگر، مربوط به داده‌هاست. اول اینکه، از آنجایی که دستیابی به تعداد زیادی داده‌ی تصویر پزشکی، سخت است، داده‌های آموزشی محدود می‌تواند منجر به تنظیم بیش از حد یا همان بیش برازش شود. برای مقابله با این مسئله، کاهش پیچیدگی معماری شبکه یا افزایش تعداد داده‌های آموزشی، از این مسئله می‌کاهد.

دوم اینکه، آموزش با داده‌های نامتعادل می‌تواند منجر به یک شبکه‌ی قطعه‌بندی ناپایدار، مخصوصاً با قطعه‌بندی ساختار یا ضایعه‌ی کوچک شود. نمونه‌گیری مجدد فضای داده، با استفاده از روش آموزش دو مرحله‌ای، نمونه‌گیری دقیق از مسیر و تابع زیان مناسب، استراتژی‌های پیشنهادی برای غلبه بر این مشکل هستند. سوم، همانند مشکلات رایج در یادگیری عمیق، آموزش یک شبکه‌ی عمیق با داده‌های محدود اصلی بدون افزایش داده یا سایر تکنیک‌های بهینه، دشوار است. بنابراین، طراحی روش‌های سریع‌تر برای انجام روش‌های بهینه‌سازی مناسب و پیچش، می‌تواند به آموزش یک شبکه‌ی قطعه‌بندی مؤثر، کمک کند. در

حوزه‌ی بینایی ماشین و رایانه، انتشار کدهای منبع در دسترس عموم، به روشی گسترده تبدیل شده است. در جدول ۵، کد موجود را نشان داده‌ایم. این روش، به تسریع پژوهش‌های این حوزه کمک می‌کند. روش رایج دیگر، اعتبارسنجی مدل روی مجموعه داده‌های مختلف است که می‌تواند دری به روی طراحی یک مدل قوی باز کند که بتوان به مجموعه داده‌های کاربردهای مشابه اعمال کرد.