



دانشگاه صنعتی شریف
دانشکده‌ی مهندسی برق

پایان‌نامه‌ی کارشناسی ارشد
بیوالکتریک

عنوان:

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

نگارش:

رضا کریمزاده

استاد راهنما:

دکتر عمادالدین فاطمی‌زاده

اسفند ۱۴۰۰

سلام افلا

چکیده

نگارش پایان‌نامه علاوه بر بخش پژوهش و آماده‌سازی محتوا، مستلزم رعایت نکات فنی و نگارشی دقیقی است که در تهیه‌ی یک پایان‌نامه‌ی موفق بسیار کلیدی و مؤثر است. از آن جایی که بسیاری از نکات فنی مانند قالب کلی صفحات، شکل و اندازه‌ی قلم، صفحات عنوان و غیره در تهیه‌ی پایان‌نامه‌ها یکسان است، با استفاده از نرم‌افزار حروف‌چینی زی‌تک و افزونه‌ی زی‌پرشین یک قالب استاندارد برای تهیه‌ی پایان‌نامه‌ها ارائه گردیده است. این قالب می‌تواند برای تهیه‌ی پایان‌نامه‌های کارشناسی و کارشناسی ارشد و نیز رساله‌ی دکتری مورد استفاده قرار گیرد. این نوشتار به طور مختصر نحوه‌ی استفاده از این قالب را نشان می‌دهد.

کلیدواژه‌ها: پایان‌نامه، حروف‌چینی، قالب، زی‌پرشین

فهرست مطالب

۷	۱ روش‌های پیشنهادی و نتایج
۷	۱-۱ مقدمه
۷	۱-۲ روش‌های پیشنهادی قطعه‌بندی ساختارهای در ریسک
۸	۱-۲-۱ معرفی مجموعه دادگان و پیش‌پردازش
۱۲	۱-۲-۲ آموزش قطعه‌بندی با استفاده از متد چگالش دانش
۱۹	۱-۳ روش‌های پیشنهادی قطعه‌بندی تومور
۱۹	۱-۳-۱ معرفی دادگان و پیش‌پردازش

فهرست شکل‌ها

- ۱-۱ نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR در سه نمای اکسیال، کرونال و سجیتال ۹
- ۲-۱ نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش، در سه نمای اکسیال، کرونال و سجیتال ۱۰
- ۳-۱ 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 ۱۰
- ۴-۱ نمونه‌ی یک تصویر در نمای اکسیال با برچسب ساختار در ریسک متناظر (الف) تصویر اصلی (ب) تصویر تبدیل یافته برای افزایش دادگان ۱۱
- ۵-۱ نمونه‌ی یک تصویر ام‌آرآی هیپوکامپ در نماهای مختلف با قطعه‌بندی (الف) تصویر اصلی (ب) تصویر پس از پیش‌پردازش ۱۲
- ۶-۱ خروجی‌های سخت و نرم برای طبقه‌بندی چهار کلاس [۱] ۱۴
- ۷-۱ عملکرد انتقال دانش توسط چارچوب چگالش دانش [۲] ۱۶
- ۸-۱ معماری شبکه‌ی UNet پیچیده، استفاده شده به عنوان آموزگار ۱۷
- ۹-۱ معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانش‌آموز اول ۱۸
- ۱۰-۱ معماری شبکه‌ی پیشنهادی بر اساس بلوک‌های Residual، استفاده شده به عنوان دانش‌آموز دوم ۱۸

فهرست جدول‌ها

فصل ۱

روش‌های پیشنهادی و نتایج

۱-۱ 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 پرداخته می‌شود و در بخش دوم مطالعه همین روند برای قطعه‌بندی تومور طی خواهد شد.

۲-۱ روش‌های پیشنهادی قطعه‌بندی ساختارهای در ریسک

همانطور که در فصل اول اشاره گردید، ضرورت قطعه‌بندی ساختارهای در ریسک، در تهیه نقشه‌ی 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 دقت 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 تسریع 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 مراکز 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.

۱-۲-۱ 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

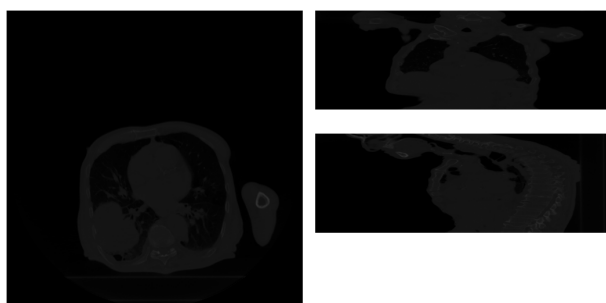
³Aorta

⁴Trachea

⁵Esophagus

⁶Ground Truth

(مانند مری) در هر شخص و نیز اختلاف شدت پایین در میان بافت‌های نرم در تصاویر سی‌تی اسکن، بسیار پرچالش است [۳]. شکل ۱-۱ یک نمونه از تصاویر اولیه‌ی موجود در این مجموعه داده را نشان می‌دهد که به علت کمبود اختلاف شده‌ها به سختی بدن بیمار قابل تشخیص است.



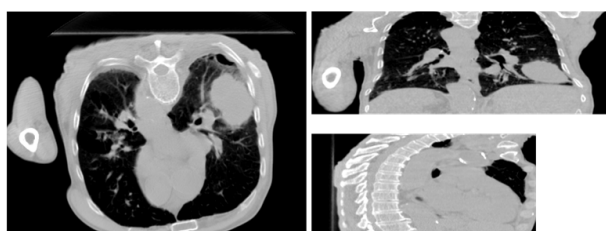
شکل ۱-۱: نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR در سه نمای اکسیال، کرونال و سجیتال

این تصاویر دارای ابعاد اولیه‌ی $(284 \sim 150) * 512 * 512$ و با فاصله‌ی واکسلی^۷، $(0.97 \sim 0.97)$ و $(2 \sim 2/5) * (0.97 \sim 1/36) * (1/36)$ در راستاهای x ، y و z است که حاکی از متفاوت بودن دستگاه‌های تصویر برداری در این مجموعه داده‌هاست. بنابراین لازم است برای تمایز هرچه بهتر بافت‌ها و یکسان‌سازی تصاویر دستگاه‌های مختلف در جهت آموزش شبکه‌ی عصبی عمیق، یک پیش‌پردازش مناسب لازم است.

اولین گام در پیش‌پردازش یکسان‌سازی فاصله‌ی واکسلی به $2 * 0.97 * 0.97$ است که با استفاده از درون‌یابی برای تصویر اصلی و تصویر قطعه‌بندی متناظر این کار انجام گردید و تا حدی تفاوت تصویر برداری 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^8 در تصاویر اصلی که در بازه‌ی $(13000 \sim -1000)$ بود مقادیری که در خارج از بازه‌ی $(400 \sim -800)$ 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 $512 * 512$ به $288 * 384$ کاهش یافت. شکل زیر تصویر نشان داده شده در شکل ۱-۱ را بعد از پیش‌پردازش نشان می‌دهد.

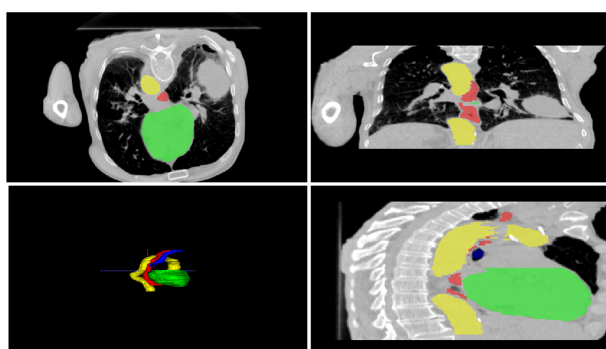
^۷Voxel Spacing

^۸Hounsfield Unit



شکل ۱-۲: 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 ۱-۳ نمونه‌ای از تصویر اصلی همراه با برچسب‌های چهار ساختار در ریسک را نشان می‌دهد.



شکل ۱-۳: 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 (شبکه‌های به صورت دوبعدی آموزش داده می‌شوند) و افزایش قدرت تعمیم‌پذیری آن از روش افزایش مجموعه داده‌گان 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 ۱-۴ (الف) تصویر دوبعدی اصلی با برچسب متناظر نشان داده شده است و در (ب) قرینه‌ی افقی و بزرگ‌نمایی آن با ضریب $1/1$ به تصویر کشیده شده است.



شکل ۱-۴: 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 اعماق 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 تثبیت‌کننده میان حافظه کوتاه‌مدت و بلندمدت است و مغز قدیمی را از آزموده‌های گذشته ما آگاه می‌کند. این مجموعه خاطرات گذشته را به شکل کوتاه‌مدت یا درازمدت حفظ می‌کند [۴].

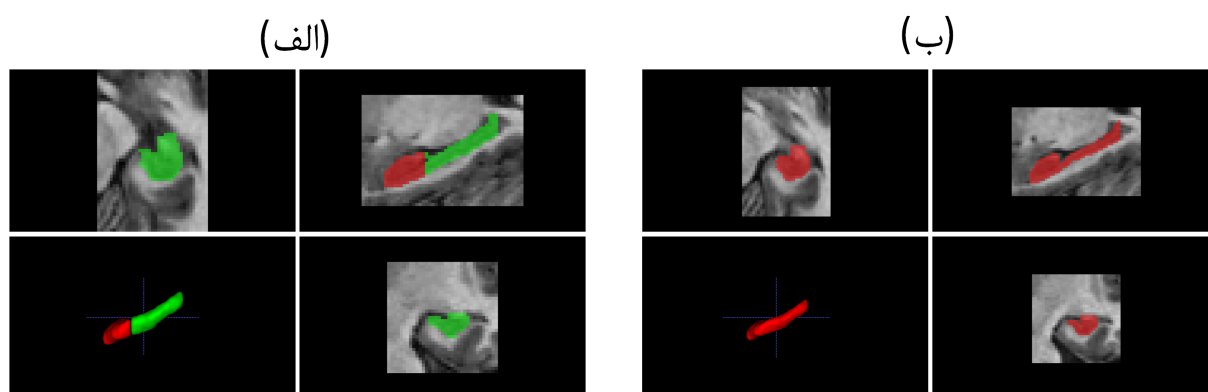
بنابراین این عضو نیز از ساختارهای مهم و حیاتی است و لازم است در طی پرتو درمانی تومورهای مغزی این ساختار نیز به عنوان ساختار در ریسک، قطعه‌بندی شود و از آسیب به آن جلوگیری شود. مجموعه داده‌ی عمومی قطعه‌بندی تصاویر پزشکی Decathlon [۵] شامل ۲۶۰ تصویر سه‌بعدی ام‌آرای T₁-weighted است که دو بخش سر و بدن هیپوکامپ را به صورت دستی و به عنوان قطعه‌بندی مطلوب ارایه داده است. ابعاد تصاویر در بازه‌ی (۲۴ ~ ۴۷) * (۴۰ ~ ۵۹) * (۳۱ ~ ۴۳) و با فاصله‌ی واکسلی یک میلی‌متر هستند.

برای پیش‌پردازش این دادگان ابتدا برای همسان‌سازی اندازه، تمام ابعاد با اضافه کردن صفر^{۱۰} به اندازه‌ی ۴۸ * ۶۴ * ۴۸ 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 این‌کار و استفاده نکردن از مقدار بیشینه مقاوم کردن نرمال‌سازی نسبت به نویزهای با

^۹Hippocampus

^{۱۰}Zero 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 پیوست باشد). شکل ۱-۵ تصویر قبل (الف) و بعد از پیش‌پردازش (ب) هیپوکامپ را با قطعه‌بندی متناظر نشان می‌دهد.



شکل ۱-۵: نمونه‌ی یک تصویر ام‌آرآی هیپوکامپ در نماهای مختلف با قطعه‌بندی (الف) تصویر اصلی (ب) تصویر پس از پیش‌پردازش

۱-۲-۲ آموزش قطعه‌بندی با استفاده از متد چگالش 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 [۶].

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

¹¹Knowledge Distillation

¹²Teacher

¹³Student

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 دانش‌آموز 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 ۱-۶ 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 [۱].

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 ۱-۶: خروجی‌های سخت و نرم برای طبقه‌بندی چهار کلاس [۱]

برای بدست آوردن احتمال حضور در امین^{۱۶} طبقه، لاجیت‌ها را از یک تابع فعالیت Softmax عبور می‌دهند تا این احتمال با مقدار p_i مشخص شود. معادله‌ی ۱-۱ تابع فعالیت Softmax را نشان می‌دهد.

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

برای استخراج خروجی‌های نرم، با اضافه کردن یک فاکتور دما^{۱۶} (T) به معادله‌ی 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. معادله‌ی زیر چگونگی اعمال فاکتور دما در معادله‌ی Softmax را نشان می‌دهد.

$$p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (2-1)$$

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

¹⁶Temperature Factor

¹⁷Ground Truth

¹⁸Distillation Loss

¹⁹Student Loss

$$L_D(p(z_t, T), p(z_s, T)) = -\sum_i p_i(z_{ti}, T) \log(p_i(z_{si}, T)) \quad (۳-۱)$$

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

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{p(z_t, T) - p(z_s, T)}{T} = \frac{1}{T} \left(\frac{\exp(z_{si}/T)}{\sum_j \exp(z_{sj}/T)} - \frac{\exp(z_{ti}/T)}{\sum_j \exp(z_{tj}/T)} \right) \quad (۴-۱)$$

اگر فاکتور دما (T) بسیار بزرگتر از مقدار لاجیت‌ها باشد آنگاه با استفاده از بسط تیلور معادله‌ی ۴-۱ را می‌توان به صورت زیر بازنویسی نمود.

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{1}{T} \left(\frac{1 + \frac{z_{si}}{T}}{N + \sum_j \frac{z_{sj}}{T}} - \frac{1 + \frac{z_{ti}}{T}}{N + \sum_j \frac{z_{tj}}{T}} \right) \quad (۵-۱)$$

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

$$\frac{\partial L_D(p(z_t, T), p(z_s, T))}{\partial z_{si}} = \frac{1}{NT^2} (z_{si} - z_{ti}) \quad (۶-۱)$$

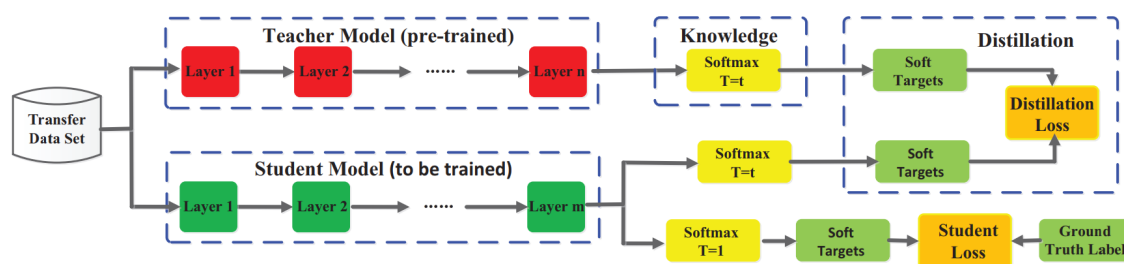
بنابراین با توجه به معادله‌ی ۶-۱ هدف تابع هزینه‌ی چگالش، کمینه کردن فاصله‌ی میان لاجیت‌های آموزگار و شاگرد است [۷]. تابع هزینه‌ی دانش‌آموز را نیز می‌توان به صورت Cross-Entropy بین خروجی مطلوب و لاجیت‌های دانش‌آموز به صورت زیر نوشت.

$$L_S(y, p(z_s, T)) = -\sum_i y \log(p_i(z_{si}, T)) \quad (۷-۱)$$

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 ۸-۱ تشکیل می‌شود که با ضریب α مجموع این دو تابع به صورت وزن‌دار محاسبه می‌شود.

$$L_{TotalDistillation} = \alpha L_D(p(z_t, T), p(z_s, T)) + (1 - \alpha) L_S(y, p(z_s, T)) \quad (۸-۱)$$

فاکتور دمای T در تابع هزینه دانش آموز برابر با یک و در تابع هزینه چگالش برابر یا بزرگتر از یک در نظر گرفته می‌شود. بنابراین به صورت کلی، چارچوب چگالش دانش را می‌توان به صورت شکل ۷-۱ بیان نمود که در آن، آموزگار یک مدل از پیش آموزش دیده شده است و دانش آموز به واسطه خروجی‌های نرم و مقادیر مطلوب که دو تابع هزینه را تشکیل می‌دهند، آموزش داده می‌شود [۲].

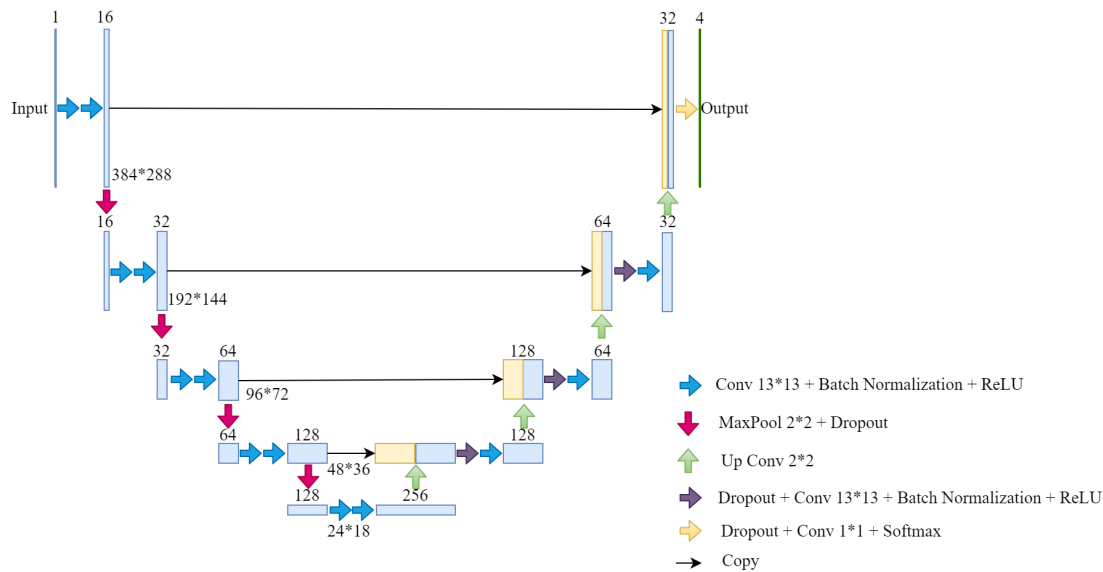


شکل ۷-۱: عملکرد انتقال دانش توسط چارچوب چگالش دانش [۲]

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 این کار نیاز به مدل‌های آموزگار و دانش آموز 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.

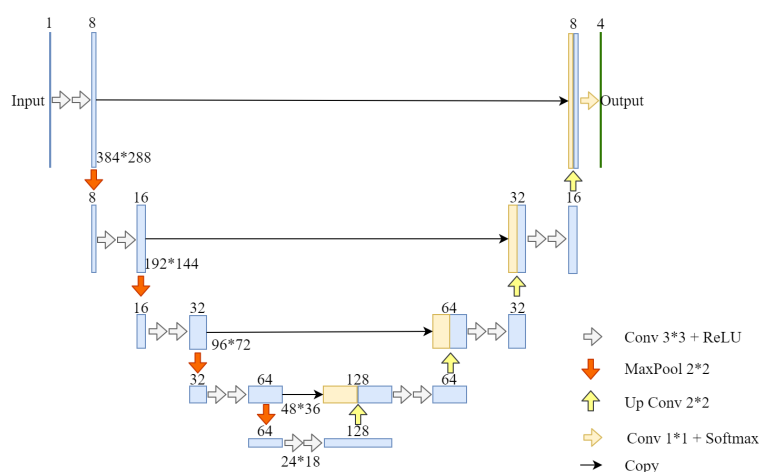
برای مدل آموزگار از معماری یک UNet پیچیده با تعداد پارامترهای قابل آموزش بالا و روش‌های تعمیم‌پذیری استفاده گردید. معماری این شبکه‌ی تمام کانولوشنی عمیق در شکل ۸-۱ قابل مشاهده است.



شکل ۱-۸: معماری شبکه‌ی UNet پیچیده، استفاده شده به عنوان آموزگار

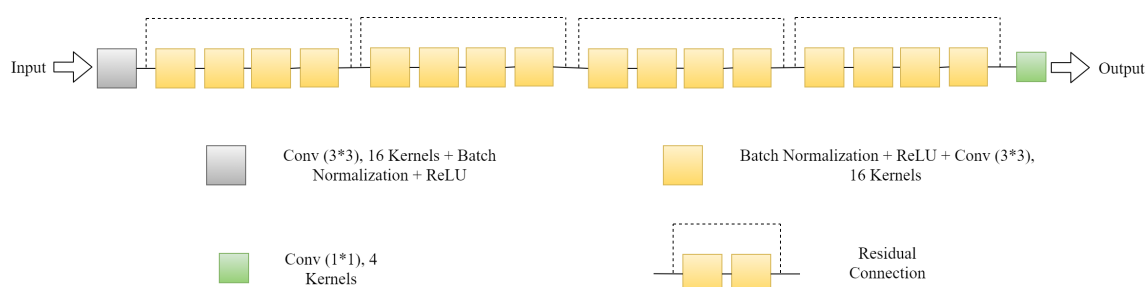
تفاوت این معماری با معماری UNet اولیه‌ی معرفی شده در [۸] استفاده از ابزارهای افزایش قدرت تعمیم‌پذیری مانند dropout و Batch-normalization پس از لایه‌های کانولوشنی است. همچنین برای افزایش وسعت دید شبکه و استخراج ویژگی‌ها عمیق‌تر و کلی‌تر از سطح تصویر، اندازه‌ی فیلترهای کانولوشنی برابر با 13×13 تعریف گردید.

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 ۱-۹ معماری یک UNet ساده شده را به عنوان دانش‌آموز اول نشان می‌دهد که در مقایسه با معماری آموزگار در شکل ۱-۸، 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 ۹-۱: معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانش‌آموز اول

شکل ۱۰-۱، 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 ۱۰-۱: 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 ۱-۲-۱ استفاده گردید. همانطور 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 ۸-۱ و ۹-۱ نشان داده شده است، در لایه‌ی خروجی این معماری‌ها چهار خروجی گرفته شده است که هر یک از آن‌ها مسوول قطعه‌بندی

یکی از ساختارهاست.

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 هزینه 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 هزینه Cross-Entropy است زیرا دیگر طبقه‌ای اضافی پس‌زمینه وجود ندارد. برای حل این مشکل از تابع هزینه Binary Cross-Entropy استفاده شد که به صورت زیر تعریف می‌شود:

$$BCE = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (9-1)$$

که در آن y_i خروجی مطلوب که صفر یا یک است و p_i خروجی پیش‌بینی شده است که در بازه صفر و یک است. فرض کنید اگر خروجی مطلوب برابر با صفر باشد، ترم اول در معادله‌ی ۹-۱ حذف می‌شود و ترم دوم سعی می‌کند با کمینه‌کردن مقدار هزینه، مقدار پیش‌بینی شده را به صفر نزدیک کند. بنابراین با صفر بودن دائمی یکی از کلاس‌ها با استفاده از این تابع هزینه، مشکلی در یادگیری بوجود نمی‌آید و با حل کردن این مسأله‌ی بهینه‌سازی برای کلاس پس‌زمینه، مقدار ۰/۲۵ در هر یک از نقشه‌های احتمال چهار ساختار در ریسک، بدست می‌آید.

نتایج

۳-۱ روش‌های پیشنهادی قطعه‌بندی تومور

۱-۳-۱ معرفی دادگان و پیش پردازش

مراجع

- [1] 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.
- [2] J. Gou, B. Yu, S. J. Maybank, and D. Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.
- [3] 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.
- [4] J. Martin. Lymbic system and cerebral circuits for emotions, learning, and memory. *Neuroanatomy: text and atlas (third ed.)*. McGraw-Hill Companies, page 382, 2003.
- [5] 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.
- [6] 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.
- [7] G. Hinton, O. Vinyals, J. Dean, et al. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2(7), 2015.

-
- [8] 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.

Abstract

We present a standard template for typesetting theses in Persian. The template is based on the X_YTeX Persian package for the L^AT_EX typesetting system. This write-up shows a sample usage of this template.

Keywords: Thesis, Typesetting, Template, X_YTeX Persian



Sharif University of Technology
Department of Electrical Engineering

M.Sc. Thesis

Organs at Risk (OAR) segmentation using machine learning methods

By:

Reza Karimzadeh

Supervisor:

Dr. Emad Fatemizadeh

February 2022