



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 عمادالدین فاطمی‌زاده

۱۴۰۰ اسفند

اللهُ أَكْبَرُ

## چکیده

نگارش پایان نامه علاوه بر بخش پژوهش و آماده سازی محتوا، مستلزم رعایت نکات فنی و نگارشی دقیقی است که در تهیه‌ی یک پایان نامه‌ی 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 پایان نامه‌ها یکسان است، با استفاده از نرم افزار حروف‌چینی زی‌تک و افزونه‌ی زی‌پرشین یک قالب استاندارد برای تهیه‌ی پایان نامه‌ها ارائه گردیده است. این قالب می‌تواند برای تهیه‌ی پایان نامه‌های 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:** پایان نامه، حروف‌چینی، قالب، زی‌پرشین

# فهرست مطالب

|       |  |    |
|-------|--|----|
| ۱     | روش‌های پیشنهادی و نتایج                                       | ۸  |
| ۱-۱   | مقدمه  | ۸  |
| ۱-۲   | روش‌های پیشنهادی قطعه‌بندی ساختارهای در ریسک                   | ۸  |
| ۱-۲-۱ | معرفی مجموعه دادگان و پیش‌پردازش                               | ۹  |
| ۱-۲-۲ | آموزش قطعه‌بندی با استفاده از متدهای چگالش دانش                | ۱۳ |
| ۱-۲-۳ | بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک          | ۲۷ |
| ۱-۲-۴ | معرفی تابع هزینه‌ی براساس شکل برای قطعه‌بندی ساختارهای در ریسک | ۳۲ |
| ۱-۳   | روش‌های پیشنهادی قطعه‌بندی تومور                               | ۳۴ |
| ۱-۳-۱ | معرفی دادگان و پیش‌پردازش                                      | ۳۴ |

# فهرست شکل‌ها

|   |    |
|---|----|
| ۱-۱ نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR در سه نمای اکسیال، کرونال و سجیتال .....   | ۱۰ |
| ۱-۲ نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش، در سه نمای اکسیال، کرونال و سجیتال .....                                    | ۱۱ |
| 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 | ۱۱ |
| ۱-۴ نمونه‌ی یک تصویر در نمای اکسیال با برچسب ساختار در ریسک متناظر (الف) تصویر اصلی (ب) تصویر تبدیل یافته برای افزایش دادگان .....                      | ۱۲ |
| ۱-۵ نمونه‌ی یک تصویر ام‌آرآی هیپوکامپ در نماهای مختلف با قطعه‌بندی (الف) تصویر اصلی (ب) تصویر پس از پیش‌پردازش .....                                    | ۱۳ |
| ۱-۶ خروجی‌های سخت و نرم برای طبقه‌بندی چهار کلاس [۱] .....  | ۱۵ |
| ۱-۷ عملکرد انتقال دانش توسط چارچوب چگالش دانش [۲] .....   | ۱۷ |
| ۱-۸ معماری شبکه‌ی UNet پیچیده، استفاده شده به عنوان آموزگار .....   | ۱۸ |
| ۱-۹ معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانش‌آموز اول .....   | ۱۹ |
| ۱-۱۰ معماری شبکه‌ی پیشنهادی بر اساس بلوک‌های Residual، استفاده شده به عنوان دانش‌آموز دوم .....   | ۱۹ |
| ۱-۱۱ آموزش شبکه‌های دانش‌آموز بر مبنای چارچوب چگالش دانش از آموزگار .....   | ۲۱ |

|   |    |
|---|----|
| ۱۲- تاثیر تغییر آستانه بر روی معیار Dice برای هر ساختار در ریسک . . . . .   | ۲۲ |
| ۱۳- خروجی مدل‌های مختلف به صورت آموزش داده شده با چارچوب چگالش دانش و تکی، (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش آموز Resnet (ت) پیش‌بینی دانش آموز UNet، (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی UNet ساده‌ی تنها . . . . . | ۲۴ |
| ۱۴- نمودار جعبه‌ای معیار Dice دادگان آزمایش برای چهار ساختار در ریسک و پنج شبکه‌ی آموزش داده شده . . . . .  | ۲۵ |
| ۱۵- نقشه‌های احتمالاتی خروجی شبکه‌ها قبل از آستانه‌گذاری (الف) آموزگار (ب) دانش آموز (پ) دانش آموز UNet (ت) Resnet تنها (ث) UNet تنها . . . . .   | ۲۶ |
| ۱۶- عملکرد چارچوب بازخورد خطای پیش‌بین برای قطعه‌بندی ساختارهای در ریسک . . . . .   | ۲۸ |
| ۱۷- معماری شبکه‌ی UNet استفاده شده برای آموزش با چارچوب بازخورد خطای پیش‌بین . . . . .  | ۳۰ |
| ۱۸- 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 . . . . .   | ۳۱ |

## فهرست جدول‌ها

|     |  |    |
|-----|--|----|
| ۱-۱ | معیار Dice برای مدل‌های مختلف  | ۲۳ |
| ۱-۲ | معیار Jaccard برای مدل‌های مختلف   | ۲۳ |
| ۱-۳ | فاصله‌ی Hausdorff برای مدل‌های مختلف   | ۲۳ |
| ۱-۴ | مقابسه‌ی میزان پیچیدگی و هزینه‌ی محاسباتی هر مدل   | ۲۷ |
| ۱-۵ | معیارهای ارزیابی قطعه‌بندی برای شبکه‌ی آموزش دیده شده به دو صورت تکی و چارچوب بازخورد خطای پیش‌بین | ۳۱ |

## 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 پرداخه می‌شود و در بخش دوم مطالعه همین روند برای قطعه‌بندی تومور طی خواهد شد.

#### ۱-۲ 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.

## 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<sup>۱</sup>. این مجموعه تصاویر، از ۴۰ بیمار دارای سرطان ریه و مری گرفته شده است و شامل قطعه‌بندی دستی چهار ساختار در ریسک قلب<sup>۲</sup>، آئورت<sup>۳</sup>، نای<sup>۴</sup> و مری<sup>۵</sup> است که به عنوان استاندارد مطلوب<sup>۶</sup> برای آموزش شبکه‌های عصبی تهیه شده است.

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

<sup>1</sup>Segmentation of THoracic Organs at Risk

<sup>2</sup>Heart

<sup>3</sup>Aorta

<sup>4</sup>Trachea

<sup>5</sup>Esophagus

<sup>6</sup>Ground Truth

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



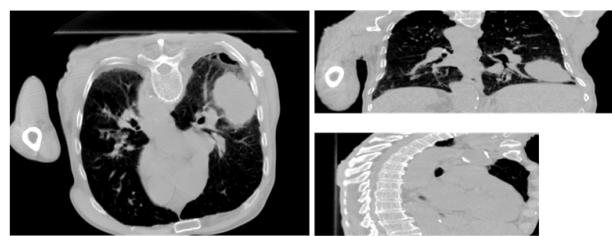
شکل ۱-۱: نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR در سه نمای اکسیال، کرونال و سجیتال

این تصاویر دارای ابعاد اولیه ( $150 \times 284 \times 512$ <sup>۷</sup>) و با فاصله‌ی واکسلی<sup>۸</sup> (~۰/۹۷) 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 لازم است.

اولین گام در پیش‌پردازش یکسان‌سازی فاصله‌ی واکسلی به  $2 \times 0/97 \times 0/97$  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 در تصاویر اصلی که در بازه‌ی (~۱۰۰۰ - ۱۳۰۰۰) بود مقادیری که در خارج از بازه‌ی (~۴۰۰ - ۸۰۰) 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  $512 \times 512 \times 384$  به  $288 \times 288 \times 384$  کاهش یافت. شکل زیر تصویر نشان داده شده در شکل ۱-۱ را بعد از پیش‌پردازش نشان می‌دهد.

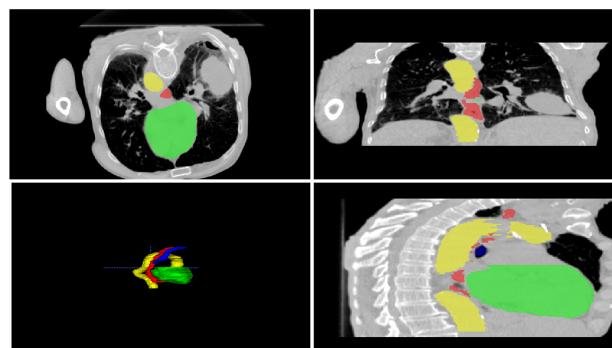
<sup>7</sup>Voxel Spacing

<sup>8</sup>Hounsfield Unit



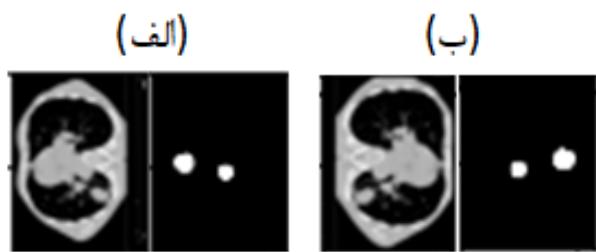
شکل ۱-۲: نمونه‌ی یک تصویر سی‌تی اسکن از مجموعه داده‌ی SegTHOR بعد از پیش‌پردازش، در سه نمای اکسیال، کرونال و سجیتال

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<sup>۴</sup> 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 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 (۵, -۵) درجه، بزرگ‌نمایی و کوچک‌نمایی با ضرایب ۱/۹ و ۰/۹ و قرینه‌ی افقی تعداد دادگان آموزش افزایش می‌یابد. در شکل ۱-۴ (الف) تصویر دو بعدی اصلی با برچسب متناظر نشان داده شده است و در (ب) قرینه‌ی افقی و بزرگ‌نمایی آن با ضرایب ۱/۱ 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<sup>۹</sup> 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.<sup>[۱۰]</sup>

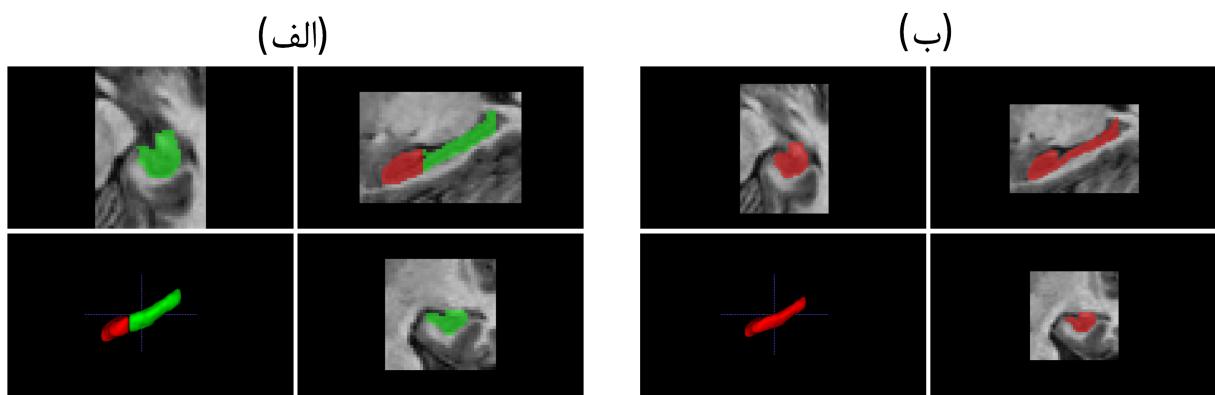
بنابراین این عضو نیز از ساختارهای مهم و حیاتی است و لازم است در طی پرتو درمانی تومورهای مغزی این ساختار نیز به عنوان ساختار در ریسک، قطعه‌بندی شود و از آسیب به آن جلوگیری شود. مجموعه داده‌ی عمومی قطعه‌بندی تصاویر پزشکی Decathlon [۵] شامل ۲۶۰ تصویر سه‌بعدی ام‌آرای T1-weighted است که دو بخش سر و بدن هیپوکامپ را به صورت دستی و به عنوان قطعه‌بندی مطلوب ارایه داده است. ابعاد تصاویر در بازه‌ی (۴۷ ~ ۵۹) \* (۴۰ ~ ۴۳) \* (۳۱ ~ ۶۴) و با فاصله‌ی واکسلی یک میلی‌متر هستند.

برای پیش‌پردازش این دادگان ابتدا برای همسان‌سازی اندازه، تمام ابعاد با اضافه کردن صفر<sup>۱۰</sup> به اندازه ۶۴ \* 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 این کار و استفاده نکردن از مقدار بیشینه مقاوم کردن نرم‌السازی نسبت به نویزهای با

<sup>9</sup>Hippocampus

<sup>10</sup>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). شنگل 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 متادانش چگالش

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<sup>۱۱</sup> از مدل پیچیده (آموزگار<sup>۱۲</sup>) به مدل ساده (دانش‌آموز<sup>۱۳</sup>) است که در ادامه قصد داریم این روش را در قطعه‌بندی ساختارهای در ریسک استفاده کنیم و مدل‌های ساده‌تر

<sup>11</sup>Knowledge Distillation

<sup>12</sup>Teacher

<sup>13</sup>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<sup>۱۴</sup> (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<sup>۱۵</sup> 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].

<sup>14</sup>Logits

<sup>15</sup>Soft 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 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<sup>۱۶</sup> ( $T$ ) 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<sup>۱۷</sup> نقش مهمی در آموزش دانش‌آموز دارند و با داشتن این مقادیر می‌توان تابع هزینه‌ی مربوط به چگالش دانش<sup>۱۸</sup> و تابع هزینه‌ی دانش‌آموز<sup>۱۹</sup> 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.

<sup>16</sup>Temperature Factor<sup>17</sup>Ground Truth<sup>18</sup>Distillation Loss<sup>19</sup>Student 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 هزینه‌ی چگالش نسبت به لاجیت‌های دانش‌آموز می‌تواند به صورت زیر محاسبه گردد.

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

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.

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

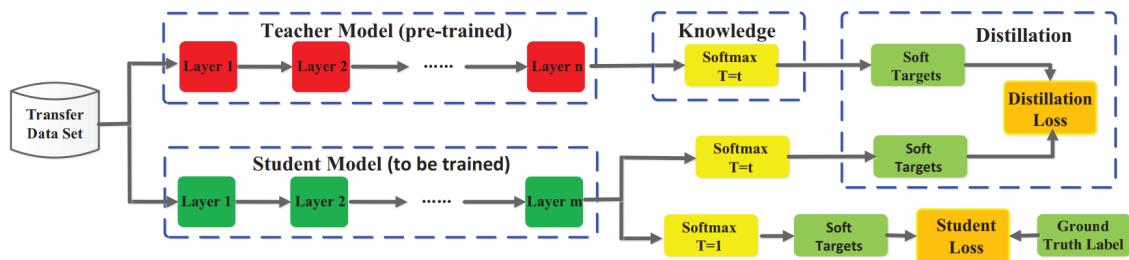
بنابراین با توجه به معادله‌ی ۶-۱ هدف تابع هزینه‌ی چگالش، کمینه کردن فاصله‌ی میان لاجیت‌های آموزگار و شاگرد است [۷]. تابع هزینه‌ی دانش‌آموز را نیز می‌توان به صورت Cross-Entropy بین خروجی مطلوب و لاجیت‌های دانش‌آموز به صورت زیر نوشت.

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

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 ۱-۸ تشکیل می‌شود که با ضریب  $\alpha$  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 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 ۱-۷ 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 [۲].

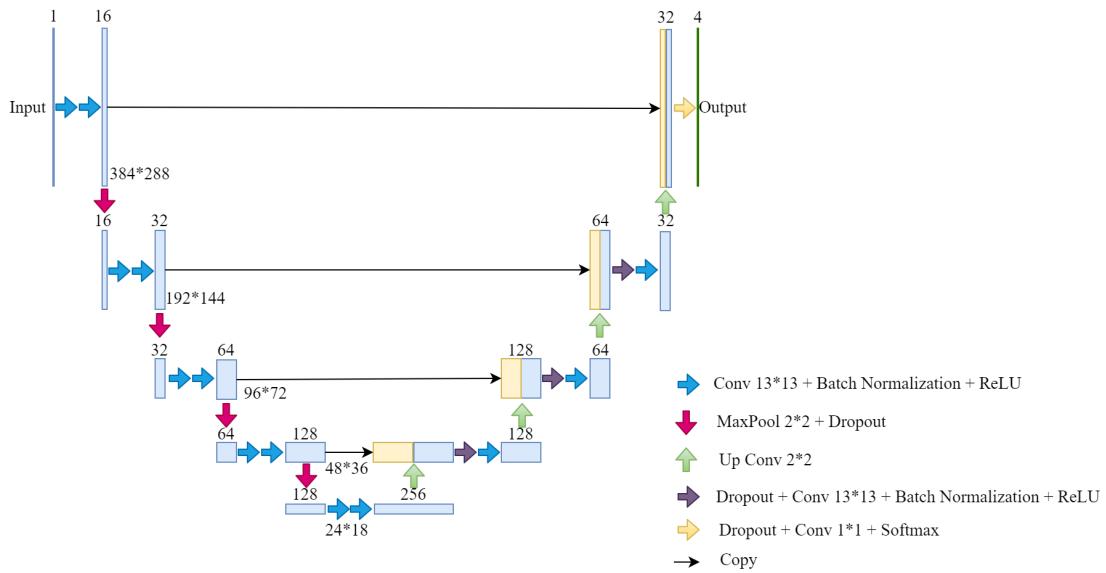


شکل ۱-۷: عملکرد انتقال دانش توسط چارچوب چگالش دانش [۲]

### 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 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.

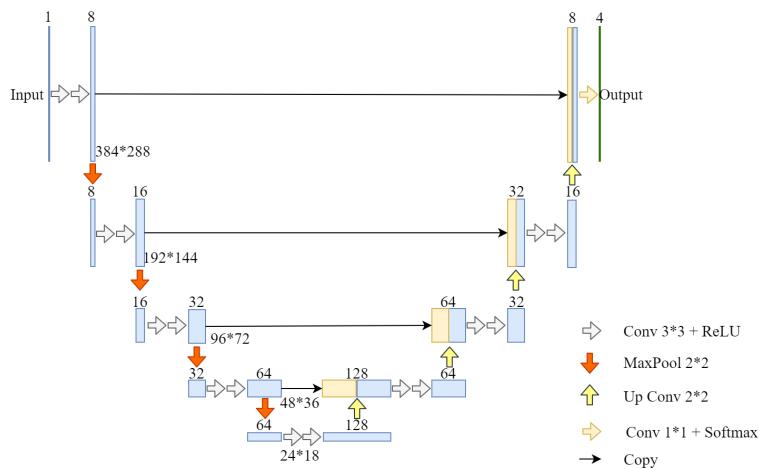
برای مدل آموزگار از معماری یک UNet پیچیده با تعداد پارامترهای قابل آموزش بالا و روش‌های تعمیم‌پذیری استفاده گردید. معماری این شبکه‌ی تمام کانولوشنی عمیق در شکل ۱-۸ قابل مشاهده است.



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

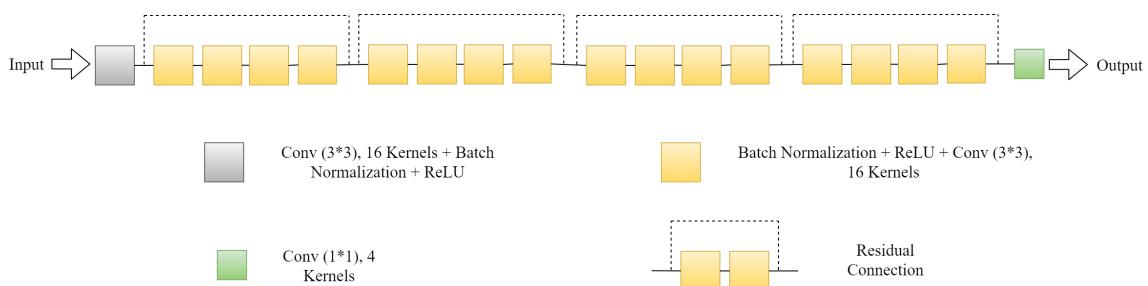
تفاوت این معماری با معماری UNet اولیه‌ی معرفی شده در [۸] استفاده از ابزارهای افزایش قدرت تعیین‌پذیری مانند dropout و Batch-normalization پس از لایه‌های کانولوشنی است. همچنین برای افزایش وسعت دید شبکه و استخراج ویژگی‌ها عمیق‌تر و کلی‌تر از سطح تصویر، اندازه‌ی فیلترهای کانولوشنی برابر با  $13 \times 13$  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 بک معماری بر اساس بلوک‌های Residual است. شکل ۹-۱ معماری یک UNet ساده شده را به عنوان دانش‌آموز اول نشان می‌دهد که در مقایسه با معماری آموزگار در شکل ۸-۱، ساختار ساده‌تری دارد که عبارتند از: کاهش ابعاد فیلترهای کانولوشنی به  $3 \times 3$  و در نتیجه کاهش وسعت دید هر فیلتر، استفاده نکردن از روش‌های تعیین‌پذیری dropout و Batch-normalization و در نهایت کاهش تعداد ویژگی‌های استخراج شده به نصف، نسبت به UNet پیچیده.



شکل ۱-۹: معماری شبکه‌ی UNet ساده، استفاده شده به عنوان دانشآموز اول

شکل ۱-۱۰، 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 ۱-۱۰: 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 ۱-۲-۱ 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، همانظور که در شکل‌های ۱-۸-۱ و ۱-۹ 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.

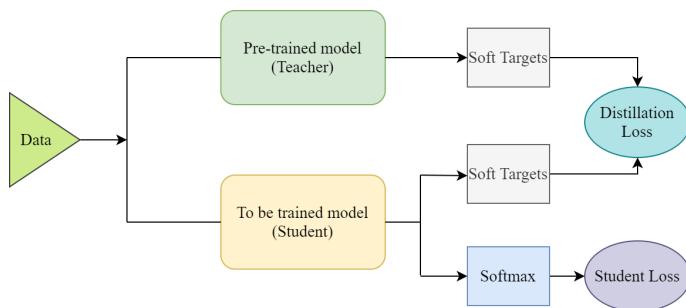
در این حالت، مشکلی که بوجود می‌آید عدم توانایی در استفاده از تابع هزینه‌ی Cross-Entropy است زیرا دیگر طبقه‌ی اضافی پس‌زمینه وجود ندارد. برای حل این مشکل از تابع هزینه‌ی Binary استفاده شد که به صورت زیر تعریف می‌شود:

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

که در آن  $y_i$  خروجی مطلوب که صفر یا یک است و  $p_i$  خروجی پیش‌بینی شده است که در بازه‌ی 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 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 ماقبی 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 ۱۱-۱ 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 ۱-۱۱: 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 loss divergence (KL) Kullback-Leibler می‌شود و سعی در کم کردن فاصله‌ی بین دو توزیع احتمال را دارد.

$$L_D(p(z_s), p(z_t)) = L_{KL}(p(z_s), p(z_t)) = \sum p(z_s) \log\left(\frac{p(z_s)}{p(z_t)}\right) \quad (10-1)$$

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.

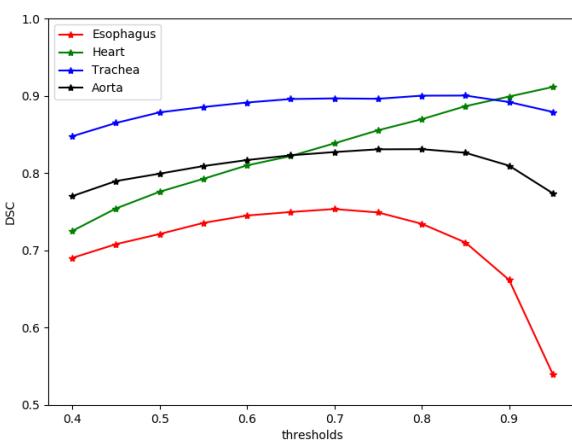
$$L_{total} = \alpha BCE(p(z_s), y) + (1 - \alpha)L_{KL}(p(z_s, T), p(z_t, T)) \quad (11-1)$$

با ملاحظات گفته شده در بالا و تنظیم پارامترهای  $T$  برابر با ۲ و  $\alpha$  برابر با  $0.5$  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.

## نتایج

برای محاسبه‌ی معیارهای ارزیابی باید برچسب‌های احتمالاتی پس از تابع فعالیت Softmax آستانه‌گذاری شوند و برخلاف تابع فعالیت Sigmoid که آستانه‌ی مشخص  $0.5$  را دارد، در این حالت باید آستانه‌های

بهینه برای بدست آوردن بالاترین مقدار در معیارهای قطعه‌بندی استخراج گردد. برای این‌کار، پیش‌بنی‌های احتمالاتی برای مدل آموزگار، بر روی مجموعه داده‌ی آموزش استخراج گردید و با آستانه‌گذاری‌های مختلف، معیار Dice برای دادگان آموزش و چهره ارگان در ریسک استخراج گردید. نمودار زیر تاثیر مقدار آستانه را بر روی مقدار معیار Dice برای هر ساختار نشان می‌دهد.



شکل ۱-۱۲: تاثیر تغییر آستانه بر روی معیار Dice برای هر ساختار در ریسک

همانطور که در شکل ۱-۱۲ 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، ۰/۹۳، قلب، ۰/۷۱، نای، ۰/۸۶ و آئورت، 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 سه معیار ارزیابی، Jaccard Dice و Hausdorff 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 ۱-۱: معیار Dice برای مدل‌های مختلف

| نام ساختار در ریسک |             |             |             |             |                    |
|--------------------|-------------|-------------|-------------|-------------|--------------------|
| نام مدل            | مری         | قلب         | نای         | آئورت       | نام ساختار در ریسک |
| UNet آموزگار       | ۰/۹۳ ± ۰/۰۳ | ۰/۹۵ ± ۰/۰۲ | ۰/۹۳ ± ۰/۰۸ | ۰/۸۶ ± ۰/۰۸ | نای                |
| Resnet دانش آموز   | ۰/۴۵ ± ۰/۱۱ | ۰/۶۲ ± ۰/۰۳ | ۰/۸۱ ± ۰/۰۶ | ۰/۶۳ ± ۰/۱۷ | آئورت              |
| Unet دانش آموز     | 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 | مری                |
| Unet تنها          | ۰/۵۴ ± ۰/۱۳ | ۰/۷۹ ± ۰/۰۹ | ۰/۸۸ ± ۰/۰۵ | ۰/۶۴ ± ۰/۱۷ | نای                |

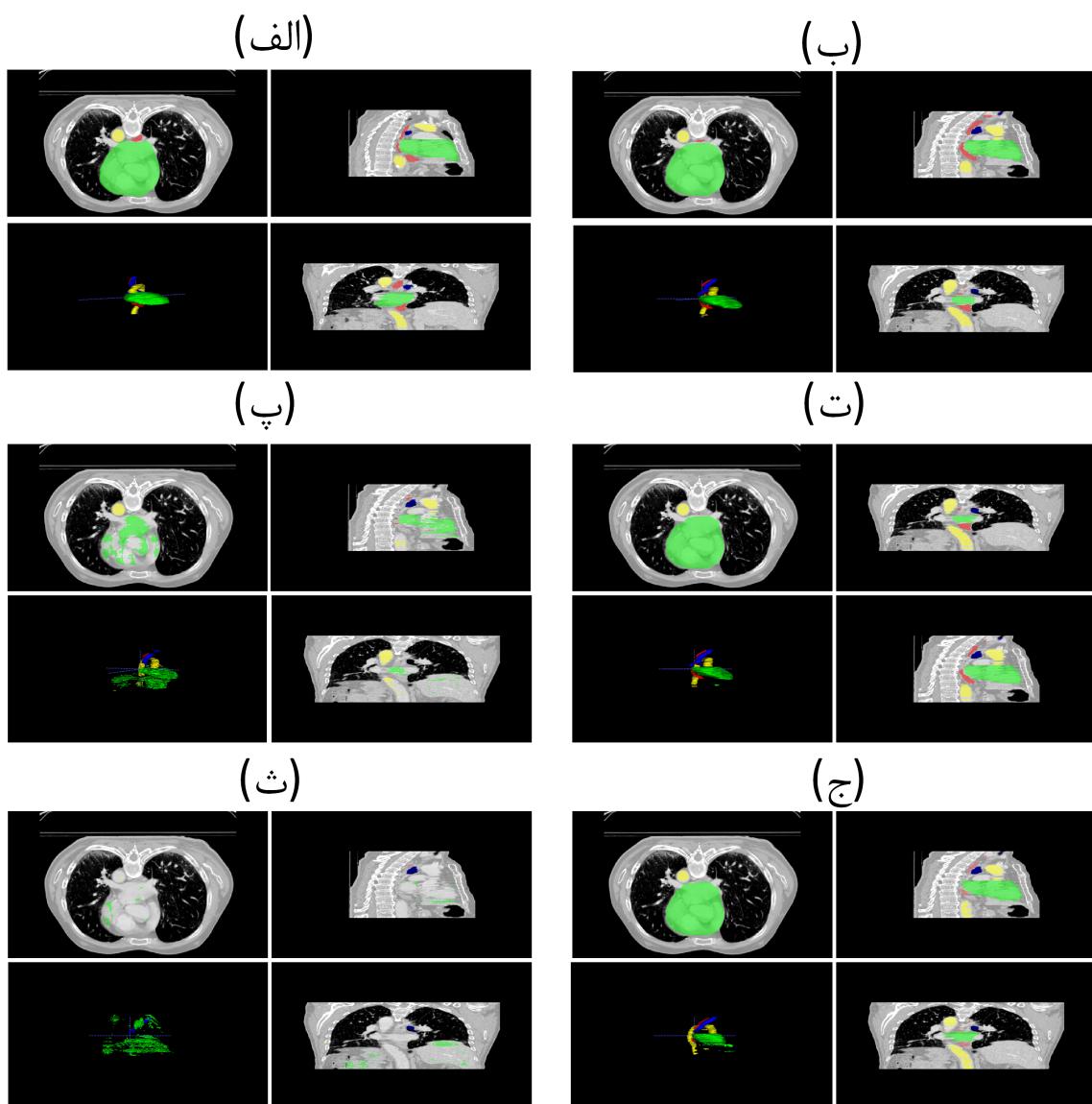
جدول ۱-۲: معیار Jaccard برای مدل‌های مختلف

| نام ساختار در ریسک |             |             |             |             |                    |
|--------------------|-------------|-------------|-------------|-------------|--------------------|
| نام مدل            | مری         | قلب         | نای         | آئورت       | نام ساختار در ریسک |
| UNet آموزگار       | ۰/۶۴ ± ۰/۱۲ | ۰/۹۳ ± ۰/۰۳ | ۰/۹۰ ± ۰/۰۹ | ۰/۸۲ ± ۰/۰۹ | نای                |
| Resnet دانش آموز   | ۰/۳۸ ± ۰/۰۹ | ۰/۶۰ ± ۰/۰۳ | ۰/۷۷ ± ۰/۰۷ | ۰/۵۶ ± ۰/۱۶ | آئورت              |
| Unet دانش آموز     | 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 | مری                |
| Unet تنها          | ۰/۴۵ ± ۰/۱۱ | ۰/۷۷ ± ۰/۰۹ | ۰/۸۵ ± ۰/۰۶ | ۰/۵۶ ± ۰/۱۵ | نای                |

جدول ۱-۳: فاصله‌ی Hausdorff برای مدل‌های مختلف

| نام ساختار در ریسک |             |              |             |             |                    |
|--------------------|-------------|--------------|-------------|-------------|--------------------|
| نام مدل            | مری         | قلب          | نای         | آئورت       | نام ساختار در ریسک |
| UNet آموزگار       | ۴/۵۴ ± ۰/۶۱ | ۴/۱۱ ± ۱/۴۸  | ۸/۱۱ ± ۰/۴۴ | ۶/۱۲ ± ۰/۶۱ | نای                |
| Resnet دانش آموز   | ۴/۷۴ ± ۰/۵۳ | ۱۰/۳۴ ± ۰/۵۶ | ۴/۵۲ ± ۰/۵۴ | ۷/۶۴ ± ۱/۱۶ | آئورت              |
| Unet دانش آموز     | ۴/۶۰ ± ۰/۵۴ | ۸/۰۲ ± ۱/۴۹  | ۴/۴۵ ± ۰/۲۸ | ۶/۰۲ ± ۰/۷۷ | قلب                |
| Resnet تنها        | ۴/۷۷ ± ۰/۴۸ | ۱۲/۸۰ ± ۱/۱۵ | ۴/۳۲ ± ۰/۳۶ | ۹/۰۰ ± ۰/۸۳ | مری                |
| Unet تنها          | ۴/۷۰ ± ۰/۵۶ | ۹/۵۰ ± ۱/۴۹  | ۴/۵۲ ± ۰/۱۸ | ۹/۶۳ ± ۰/۸۶ | نای                |

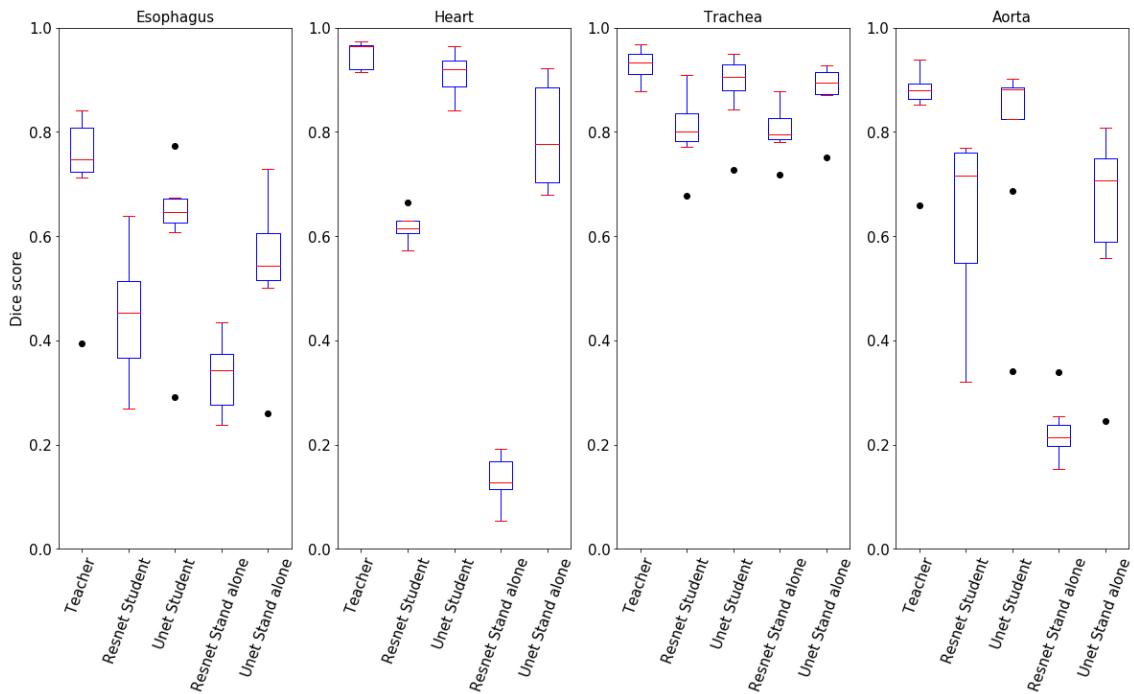
یک نمونه از خروجی قطعه‌بندی تصاویر سی‌تی اسکن سه‌بعدی دادگان آزمایش، در چهار نمای مختلف اکسیاب، سجیتال، کرونال و سه‌بعدی برای مدل‌های مختلف در شکل ۱۳-۱ نشان داده شده است که در آن (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش‌آموز UNet (ت) پیش‌بینی دانش‌آموز، (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی Resnet ساده‌ی تنها هستند.



شکل ۱۳-۱: خروجی مدل‌های مختلف به صورت آموزش داده شده با چارچوب چگالش دانش و تکی، (الف) قطعه‌بندی مطلوب، (ب) پیش‌بینی مدل آموزگار، (پ) پیش‌بینی دانش‌آموز UNet (ت) پیش‌بینی دانش‌آموز، UNet (ث) پیش‌بینی Resnet تنها و (ج) پیش‌بینی Resnet ساده‌ی تنها

پُرچہ گیری و نتیجہ

همانطور که اشاره گردید، هدف از چارچوب چگالش دانش، انتقال دانش یادگیری شده از آموزگار به دانشآموز است که این روند باعث بهبود دقت و یادگیری بهتر در دانشآموز می‌شود. همانطور که در جدول‌های ۱-۱، ۲-۱ و ۳-۱ و شکل ۱۳-۱ 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<sup>۲۰</sup> معيار Dice دادگان آزمایش برای چهار ساختار در ریسک و برای پنج شبکه‌ی آموزش داده شده به صورت یک‌جا در شکل ۱۴-۱ رسم گردید. همانطور که ملاحظه می‌شود شبکه‌هایی دانشآموز که یکبار به صورت تکی و بار دیگر با چارچوب چگالش دانش آموزش دیده‌اند، عملکردهای بهبود یافته است.

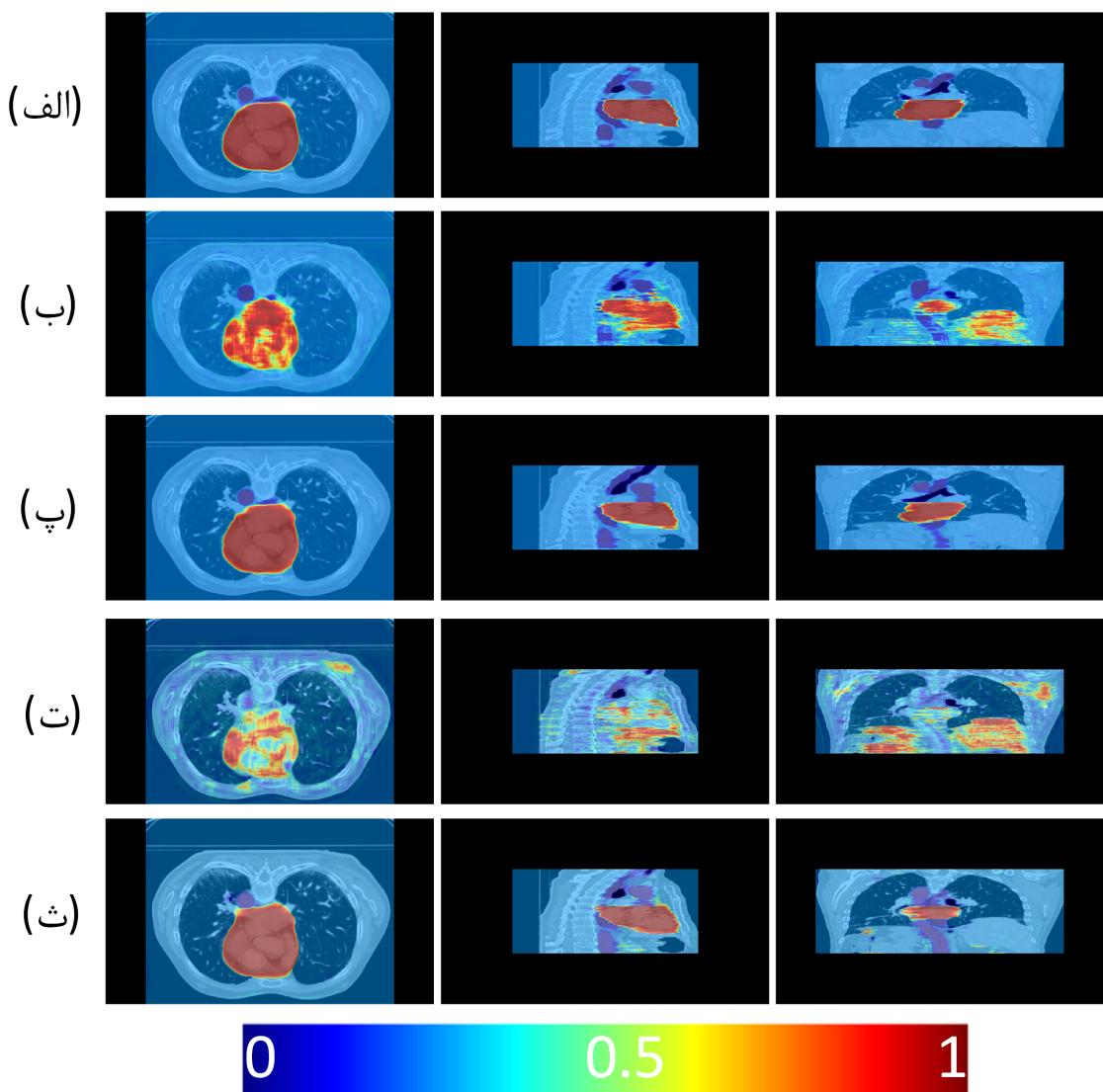


شکل ۱-۱۴: نمودار جعبه‌ای معیار Dice دادگان آزمایش برای چهار ساختار در ریسک و پنج شبکه‌ی آموزش داده شده

برای شهود بیشتر عملکرد چارچوب چگالش دانش، شکل ۱۵-۱ 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 در این نقشه ظاهر شده است و برای ساختارهای دیگر

20 Box Plot

مانند نای، مری و آئورت، 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.



شکل ۱-۱۵: نقشه‌های احتمالاتی خروجی شبکه‌ها قبل از آستانه‌گذاری (الف) آموزگار (ب) دانشآموز (پ) دانشآموز UNet (ت) Resnet UNet تنها (ث) Resnet

هدف چگالش دانش افزایش دقت در مدل‌های ضعیفتر است. برای مقایسه میزان پیچیدگی مدل‌ها از سه معیار تعداد پارامترهای قابل یادگیری (بر حسب میلیون)، تعداد عملیات اعشاری در واحد

زمان <sup>۲۱</sup> (بر حسب میلیارد) و متوسط زمان اجرا بر روی تصاویر سه‌بعدی (بر حسب ثانیه) در جدول ۱-۴ 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 دقیق آن را به دقت بالاتری رسانید و از هزینه‌های محاسباتی بالا و زمان پردازش زیاد جلوگیری نمود.

جدول ۱-۴: مقابسه‌ی میزان پیچیدگی و هزینه‌ی محاسباتی هر مدل

| نام شبکه     | تعداد پارامترها (میلیون) | FLOPs(G) | زمان اجرا (ثانیه) |
|--------------|--------------------------|----------|-------------------|
| Unet آموزگار | ۱۴/۲                     | ۱۸۳      | ۳۳/۵۴۲            |
| Resnet       | ۳/۵۴                     | ۱۸/۱     | ۰/۰۸۳             |
| Unet ساده    | ۲/۲۶                     | ۲/۷۱     | ۰/۰۴۰             |

### ۱-۲-۳ 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 <sup>۲۲</sup> بر اساس شبکه‌های بازگشتی <sup>۲۳</sup>، یک مسیر بازخورد از لایه‌های بالاتر به لایه‌ها پایین‌تر که حاوی پیش‌بینی عملکرد ناحیه‌های پایینی کورتکس است، متصل می‌شود و در مسیر روبره‌جلو، خطای پیش‌بینی شده جریان می‌یابد [۹].

با الگوبرداری از این مدل، در این قسمت قصد داریم با اضافه کردن یک مسیر پیش‌بین به شبکه‌های روبره‌جلوی کانولوشنی و بازخورد دادن خطای پیش‌بینی، یک چارچوب برای افزایش دقیق شبکه‌های 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 بر خلاف 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.

<sup>21</sup>Floating Point Operations Per Second (FLOPS)

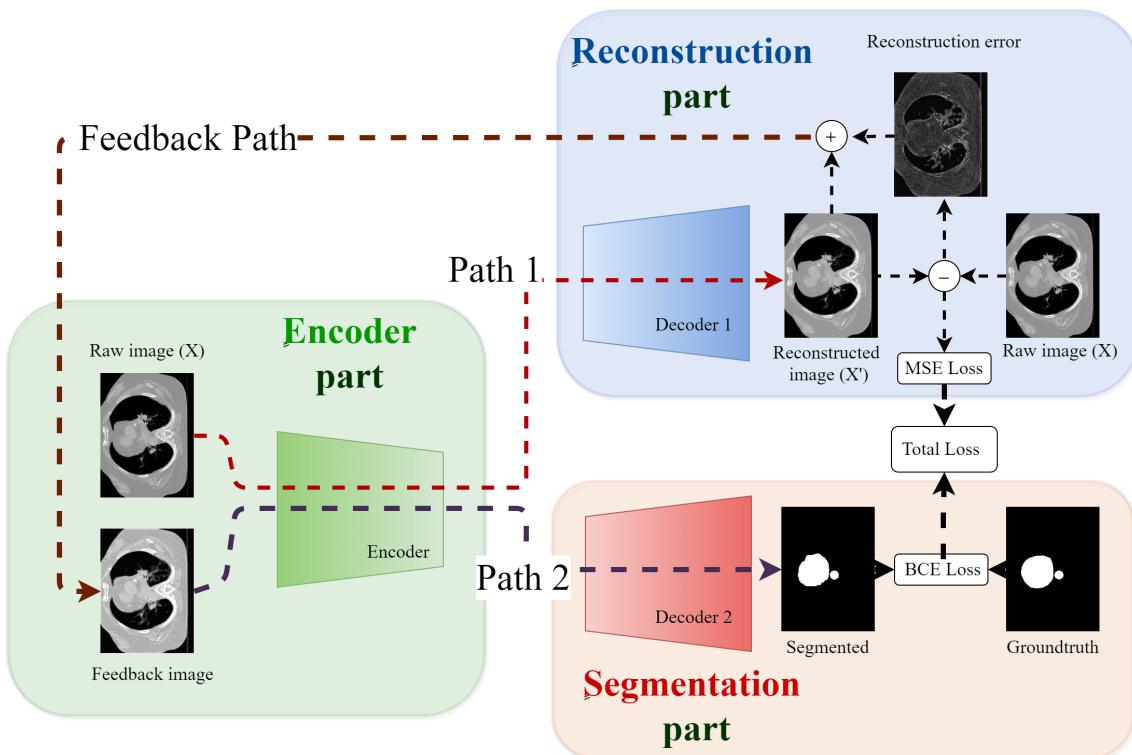
<sup>22</sup>Predictive Coding Model

<sup>23</sup>Recurrent 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 ۱-۱۶ 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 ۱-۱۶: 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:

$$X' = \text{Decoder}_1(\text{Encoder}(X)) \quad (12-1)$$

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

که در آن  $X$  تصویر ورودی،  $X'$  تصویر بازسازی شده،  $Z$  تصویر شامل خطای پیش‌بینی در مسیر بازخورد و  $\beta$  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.

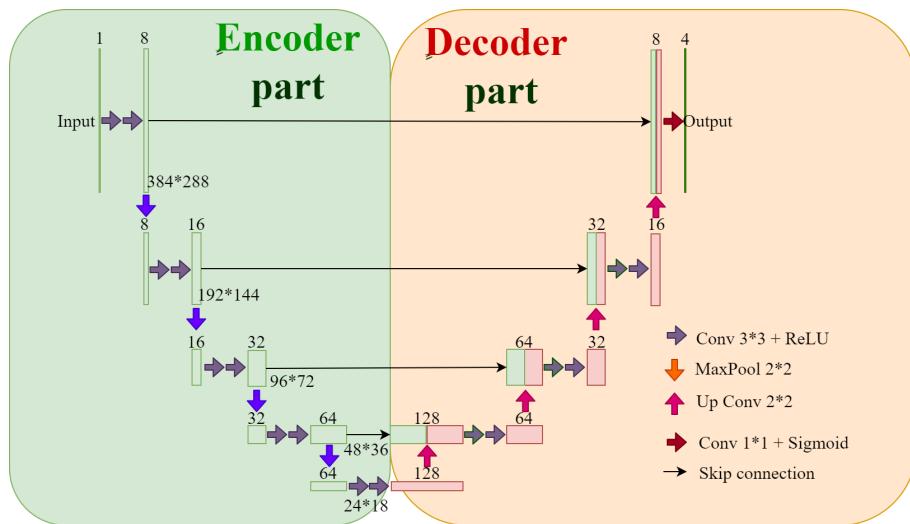
$$P = \text{Decoder}_2(\text{Encoder}(Z)) \quad (13-1)$$

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 خطای<sup>۲۴</sup> (MSE) و برای تابع هزینه‌ی قطعه‌بندی از BCE استفاده کرد. جمع وزن‌دار این دوتابع هزینه، هزینه‌ی نهایی را ایجاد می‌کند. معادلات ۱۴-۱ تابع هزینه‌ی این چارچوب را به صورت فرمول‌بندی نشان می‌دهد.

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

<sup>24</sup>Mean Square Error

برای ارزیابی این چارچوب از معماری یک UNet مطابق با شکل ۱۷-۱ استفاده گردید. برای شبکه‌های رمزگذار و رمزگشای معرفی شده در شکل ۱۶-۱، 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 همچنان در چارچوب معرفی شده برای استفاده از شبکه‌ی UNet همچنان پا بر جاست.



شکل ۱-۱۷: معماری 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  $\alpha$  و  $\beta$  در معادلات ۱۴-۱ و ۱۲-۱ به ترتیب برابر با  $0/2$  و  $2$  در روند آموزش قرار داده شد.

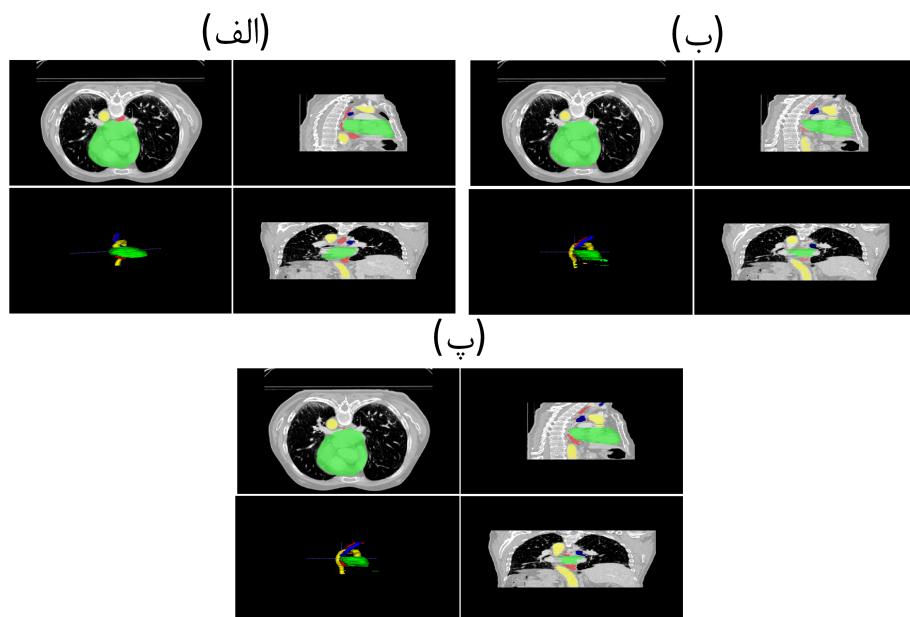
## نتایج و ملاحظات

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 ۱-۵: 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 ۱-۱۸-۱ 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 ۱-۱۸-۱: 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

همانطور که در شکل ۱-۱۸-۱ و جدول ۱-۵ 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 ۰/۱۱ رشد داشته است). علت این بهبود در استفاده از این چارچوب را می‌توان در دو دلیل توضیح داد:

- استفاده از یک Encoder مشترک در پیش‌بینی ورودی و قطعه‌بندی آن با توابع هزینه‌های 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 بازنورد خطاب برای افزایش دقت قطعه‌بندی ارایه گردید و ملاحظه شد، با آموزش یک UNet با استفاده از این چارچوب دقت به صورت قابل ملاحظه‌ای افزایش یافت.

## ۱-۲-۴ معرفی تابع هزینه‌ی بر اساس 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 [۱۰].

- 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 ۱۳-۱ و ۱۸-۱ 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 [۱۱]. 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<sup>۲۵</sup> برای قطعه‌بندی و یا روش‌های پس‌پردازش<sup>۲۶</sup> ارایه شده است که از جمله‌ی آن‌ها می‌توان به فیلد‌های تصادفی شرطی/مارکوف<sup>۲۷</sup> و مدل‌های کانتور فعال<sup>۲۸</sup> اشاره کرد [۱۲].

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 [۱۳] 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 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 از 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

<sup>25</sup>Prior Knowledge

<sup>26</sup>Post-Processing

<sup>27</sup>Conditional/Markov Random Fields

<sup>28</sup>Active Contour Models

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

## ۱-۳ روش‌های پیشنهادی قطعه‌بندی تومور

### ۱-۳-۱ معرفی دادگان و پیش‌پردازش

# 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] S. Bohlender, I. Oksuz, and A. Mukhopadhyay. A survey on shape-constraint deep learning for medical image segmentation. *arXiv preprint arXiv:2101.07721*, 2021.
- [13] 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.

## **Abstract**

We present a standard template for typesetting theses in Persian. The template is based on the `XEPersian` package for the `LATEX` typesetting system. This write-up shows a sample usage of this template.

**Keywords:** Thesis, Typesetting, Template, `XEPersian`



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