

بهبود شبکه‌های U-net برای قطعه‌بندی تصاویر پزشکی با اضافه کردن لایه‌های مکانیزم توجه

محمد رجب قانع^۱، عباس بحرالعلوم^۲ و مهدی افتخاری^۳

چکیده

قطعه‌بندی تصاویر پزشکی یکی از مهم‌ترین گام‌ها در تحلیل تصاویر پزشکی، جهت بهبود تشخیص و یافته‌ها است. یکی از متداول‌ترین روش‌های قطعه‌بندی در یادگیری عمیق، استفاده از شبکه‌های U-net است. وجود لایه‌های متراکم در قسمت رمزگشای U-net، اجازه استخراج اطلاعات از لایه‌های عمیق‌تر را نمی‌دهد؛ همچنین به علت محدودیت میدان دریافتی هسته‌های کانولوشن، اطلاعات و وابستگی‌های دوربرد به خوبی در نظر گرفته نمی‌شوند. در این مقاله، هدف طراحی یک ساختار در اتصالات پرش به منظور کاهش شکاف معنایی بین ناحیه رمزگذار و رمزگشا است. استخراج بهتر و تمرکز بیشتر بر روی ویژگی‌های محلی و سراسری در مجموعه داده‌های مختلف، از ویژگی‌های این ساختار است. همچنین یک ساختار توجه به منظور کاهش پارامترهای شبکه و بهبود نتایج، در گلوگاه شبکه طراحی شده است. این روش بر روی ۶ مجموعه داده پزشکی ارزیابی شده است که نتایج به دست آمده در دو معیار ارزیابی Dice و Iou نشان می‌دهد مدل پیشنهاد شده نتایج بهتری نسبت به U-net و روش‌های مبتنی بر آن دارد.

کلید واژه‌ها

قطعه‌بندی تصاویر پزشکی، شبکه‌های عصبی پیچشی، مکانیزم‌های توجه، یادگیری عمیق، U-net

۱ - مقدمه

تصویر، یک دسته‌بندی در سطح پیکسل است که علاوه بر دسته بندی پیکسل‌ها، تمایز بین این دسته‌ها برای جداسازی پس زمینه و نواحی هدف باید مشخص شود. این روش، نه تنها برای شناسایی نواحی غیرطبیعی در بافت ارگان‌های بدن استفاده می‌شود، بلکه می‌تواند میزان پیشرفت یک بیماری را نیز مشخص کند [۱]. قطعه‌بندی خودکار تصاویر پزشکی، یک امر حیاتی در حوزه پزشکی است و مزایای بسیاری دارد. این روش، زمان پردازش داده‌ها را کم و با اندازه‌گیری دقیق به پزشکان کمک می‌کند [۲].

تاکنون روش‌های سنتی زیادی مانند الگوریتم‌های فازی [۳]، آستانه‌گذاری [۴]، مبتنی بر ناحیه [۵] برای قطعه‌بندی تصاویر پزشکی ارائه شده است. در این روش‌ها، کارایی نتیجه تا حدود زیادی به کیفیت تصویر بستگی دارد. به عنوان مثال، نویز زیاد تصویر، اثر منفی در نتیجه قطعه‌بندی می‌گذارد؛ همچنین روش‌های

تصاویر پزشکی حاوی اطلاعات مهمی برای تشخیص و درمان بسیاری از بیماری‌ها هستند. این تصاویر باید با دقت بالایی تحلیل شوند اما به علت پیچیده بودن و تعداد بالای تصاویر، بررسی آن‌ها به صورت دستی زمان‌بر است و دقت لازم را ندارد. قطعه‌بندی

این مقاله تیرماه ۱۴۰۲ دریافت شد در مهرماه بازنگری و در آبان‌ماه پذیرفته شد.

^۱ دانشجوی کارشناسی ارشد هوش مصنوعی، دانشگاه شهید باهنر، کرمان

رایانامه: mohammad.rajabghane@eng.uk.ac.ir

^۲ دانشکده مهندسی کامپیوتر، دانشگاه شهید باهنر، کرمان

رایانامه: a.bahrololoum@uk.ac.ir

^۳ دانشکده مهندسی کامپیوتر، دانشگاه شهید باهنر، کرمان

رایانامه: m.eftkhari@uk.ac.ir

ماژول خود توجه چندسر^۷ استفاده شده است. ویژگی‌های سراسری و ارتباط میان آن‌ها در این بلاک استخراج می‌شوند. استخراج ویژگی‌های محلی در ماژول توجه محلی که در آن لایه‌های کانولوشن با اندازه هسته 1×1 تعبیه شده است، انجام می‌شود. در نهایت، ترکیب ویژگی‌های محلی و سراسری استخراج شده، در اختیار لایه‌های رمزگشا قرار می‌گیرد.

در این مقاله، برای بهبود بیشتر قطعه بندی و کاهش پارامترهای شبکه، یک ماژول در گلوگاه^۸ شبکه U-net، با ایده گرفتن از [۱۲] طراحی شده است. این ماژول برای برجسته‌تر شدن ویژگی‌های زمینه‌ای و فضایی استفاده شده است. همچنین این ساختار، عملیات ترکیب و استخراج ویژگی‌ها را در عمیق‌ترین سطح شبکه با تعداد پارامتر و محاسبات کمتری انجام می‌دهد. در ادامه مقاله، در بخش ۲ مروری بر کارهای پیشین، در بخش ۳ تشریح روش پیشنهادی، در بخش ۴ آزمایش‌ها و پیاده‌سازی‌ها و در بخش ۵ به نتیجه‌گیری پرداخته خواهد شد.

۲- مرور کارهای پیشین

در سال‌های اخیر با توجه به عملکرد فوق العاده شبکه‌های کانولوشن در پردازش تصویر، معماری‌های زیادی برای قطعه‌بندی تصاویر پزشکی ارائه شده است اما به دلیل اینکه هدف این مقاله بررسی شبکه‌های U شکل است، در ادامه این بخش، چند نمونه از مقالات در این حوزه را بررسی خواهیم کرد.

مدل‌های مبتنی بر ساختار U-net به طور گسترده در قطعه‌بندی تصاویر پزشکی توسعه یافتند. برای بهبود نتیجه قطعه‌بندی، محققان تمرکز خود را روی استخراج بهتر ویژگی‌ها در ناحیه رمزگذار و بهبود اتصالات پرش گذاشتند. در روش U-net++ [۱۰] یک طراحی جدید در قسمت اتصالات پرش به منظور کاهش شکاف معنایی بین لایه‌های رمزگذار و رمزگشا انجام شد. در این مقاله، اتصالات پرش متراکم^۹ به گونه‌ای بازطراحی شدند تا پردازش بیشتری روی ویژگی‌های ورودی از ناحیه رمزگذار انجام شود. این ساختار، با کاهش اختلاف سطح معنایی ویژگی‌های رمزگذار و رمزگشا باعث بهبود در نتیجه قطعه‌بندی شود. پیچیدگی زیاد شبکه بر اثر افزودن لایه‌های متراکم در اتصالات پرش، مهم‌ترین نقطه ضعف این روش است. در U-net3+ [۱۳] یک روش جدید برای ترکیب ویژگی‌ها از مقیاس‌های مختلف در لایه رمزگشا استفاده شد؛ یعنی هر سطح از لایه رمزگشا، علاوه بر دریافت ویژگی‌های هم سطح، از سایر سطوح رمزگذار نیز ویژگی‌ها را دریافت می‌کند تا شکاف معنایی بین رمزگذار و رمزگشا کاهش یابد. در سال‌های اخیر مکانیزم‌های توجه برای تمرکز در یک ناحیه خاص در پردازش داده‌های بزرگ مورد استفاده قرار گرفتند. در

سنتی در تصاویری که بافت‌های بدن در آن پیچیده هستند، نتیجه مناسبی نمی‌دهند. امروزه حجم عظیمی از داده‌ها با شیوه‌های متفاوت تصویر برداری [۶] مانند سی‌تی‌اسکن، تصویر برداری رزونانس مغناطیسی^۱ و X-ray در اختیار پزشکان قرار می‌گیرد.

شبکه‌های عصبی عمیق یک نوع شبکه عصبی مصنوعی با لایه‌های متعدد ورودی و خروجی، الگوهایی از داده‌ها را یادگیری می‌کنند. اگرچه روش‌های یادگیری عمیق جدید نیستند اما با ظهور واحدهای پردازشگر گرافیکی برای تسریع در اجرای شبکه‌های عصبی، موفقیت دوچندانی کسب کردند. یکی از مهم‌ترین شبکه‌های عصبی عمیق، شبکه عصبی پیچشی یا کانولوشن (CNN) است که کاربرد زیادی در پردازش تصویر دارد. شبکه‌های کانولوشن با استفاده از استخراج ویژگی‌ها موفق به قطعه‌بندی تصاویر شده‌اند به نحوی که نیاز به ویژگی‌های دست‌ساز برای قطعه بندی از بین می‌رود. چالش اصلی شبکه‌های عصبی کانولوشن، در نظر نگرفتن وابستگی‌های دوربرد به علت محدودیت اندازه هسته لایه‌های کانولوشن است. این چالش در قطعه بندی تصاویر معادل نادیده گرفتن همبستگی میان پیکسل‌ها است. در سال ۲۰۱۵، شبکه کاملاً پیچشی^۲ [۷] با قراردادن چندین بلاک شامل لایه‌های کانولوشن، توابع فعال‌ساز^۳ و ادغام^۴، تصویر ورودی را در ابعاد دلخواه دریافت و پس از رمزگذاری در اختیار مسیر رمزگشا قرار می‌داد تا پس از چندلایه، به تدریج نقشه ویژگی با وضوح بالا و ابعاد اولیه حاصل شود.

با الهام گرفتن از معماری رمزگذار-رمزگشا^۵ مدل‌های زیادی براساس آن توسعه یافتند. یکی از محبوب‌ترین معماری‌های رمزگذار-رمزگشا در قطعه‌بندی تصاویر، U-net [۸] است. این شبکه مانند شبکه کاملاً پیچشی، شامل دو قسمت رمزگذار و رمزگشا متقارن است؛ نکته کلیدی در این روش، کپی کردن نقشه ویژگی حاصل شده در هر سطح رمزگذار به همان سطح در رمزگشا از طریق اتصالات پرش^۶ است. این طراحی، اطلاعات معنایی را که بر اثر اعمال لایه‌های کانولوشن و ادغام از دست رفته‌اند و در مراحل بازیابی اطلاعات به آن‌ها احتیاج است، جبران می‌کند. این ایده، ساختار پایه بسیاری از روش‌ها قرار گرفت. در همین راستا تعداد زیادی مدل برای بهبود اتصالات پرش مانند Attention U-net [۹]، U-net++ [۱۰] و Densely Connections [۱۱] توسعه یافتند. در این مقاله، هدف ما طراحی یک ساختار در ناحیه اتصالات پرش به منظور ترکیب ویژگی‌های سراسری و محلی و توجه بیشتر به ویژگی‌های مهم‌تر است. در این روش برای استخراج ویژگی‌های سراسری و در نظر گرفتن وابستگی‌های دوربرد بین پیکسل‌ها، از

¹ Magnetic resonance imaging (MRI)

² Fully Convolution Network (FCN)

³ Activation Functions

⁴ Pooling

⁵ Encoder-Decoder

⁶ Skip Connection

^۷ Milti Head Self Attention

^۸ Bottleneck

^۹ Dense Skip Connections

TransAttUnet [۲۳] انجام شد و نتایج خوبی صورت گرفت. Swin-Unet دارای یک ساختار U شکل است که قسمت رمزگذار و رمزگشا در آن از بلاک‌های ترانسفورمری ساخته شده است. در این بلاک‌ها، استفاده از پنجره‌های شیفت داده شده و نقشه ویژگی-های سلسله مراتبی با وضوح چندگانه باعث بهبود نتیجه می‌شود. مکانیزم خود توجه در ترانسفورمرها [۲۴] نقص مدلسازی شبکه‌های کانولوشن در ویژگی‌های سراسری را جبران کرد و توانست همبستگی میان پیکسل‌ها را به خوبی محاسبه کند این ماژول، ویژگی‌های مختلف را در چندین زیرفضای متفاوت یاد می‌گیرد. در UTNet [۲۲] به منظور رفع برخی نواقص شبکه‌های کانولوشن در وابستگی‌های دوربرد و محاسبه همبستگی بین پیکسل‌ها، از ساختار خودتوجه استفاده شده است. در این روش، ساختاری برپایه خودتوجه، در رمزگذار و رمزگشا طراحی شده است. همچنین یک رمزگذار موقعیتی در کنار ساختار خودتوجه طراحی شده است تا سربار محاسباتی کاهش یابد. کندبودن یادگیری به علت استفاده زیاد از ساختارهای خودتوجه و عدم توجه کافی به ویژگی‌های محلی از جمله نقاط ضعف این روش است.

در روش TransAttUnet یک شبکه جدید هدایت شونده توجه مبتنی بر ترانسفورمرها پیشنهاد شد که در آن توجه هدایت شده چند سطحی و اتصالات پرش چند مقیاسی، برای بهبود عملکرد قطعه بندی معنایی پیشنهاد شده است. در ساختار TransAttUnet با الهام از ترانسفورمرها، سه ماژول توجه خودآگاه^{۱۴}، خودتوجه و توجه فضایی سراسری^{۱۵} گنجانده شده است تا تعاملات غیرمحلی بین ویژگی‌های رمزگذار را به طور مؤثر یاد بگیرد. علاوه بر این، از اتصالات پرش چند مقیاسی اضافی بین بلوک‌های رمزگشا برای جمع‌آوری ویژگی‌های نمونه برداری شده با مقیاس‌های معنایی مختلف استفاده شده است تا توانایی نمایش اطلاعات چند مقیاسی برای تولید ویژگی‌های متمایز تقویت شود. این روش می‌تواند به طور موثری از دست دادن جزئیات ظریف ناشی از انباشته شدن لایه‌های کانولوشن و عملیات نمونه برداری متوالی را کاهش دهد و در نهایت کیفیت قطعه بندی تصاویر پزشکی را بهبود بخشد. تعمیم این روش بر روی مجموعه داده‌های تصویر پزشکی مختلف از نقاط قوت این روش است.

مرور بر کارهای پیشین روش Unet نشان می‌دهد که هنوز جای تحقیقات بسیاری در زمینه قطعه بندی تصاویر برای دستیابی به نتایج مطلوب‌تر وجود دارد. یکی از مهم‌ترین چالش‌های روش‌های ارائه شده، پیچیدگی محاسباتی زیاد و کند بودن یادگیری است. همچنین اغلب روش‌های پیشنهاد شده در این زمینه، قابلیت تعمیم

Attention Unet [۹] دروازه‌های توجه^{۱۰} به اتصالات پرش اضافه شدند. هدف این دروازه‌ها، سرکوب کردن و وزن دهی کم‌تر به مناطق غیرضروری است؛ درحالی که مناطق مهم‌تر را برجسته تر می‌سازند. خروجی هر دروازه توجه حاصل ضرب مؤلفه ای ویژگی‌های ورودی در ضرایب توجه است. نکته مهم در این روش، افزایش نیافتن زیاد پارامترهای شبکه به دلیل استفاده از ساختار توجه در اتصالات پرش است. لازم به ذکر است افزایش نیافتن پارامترهای شبکه یک مزیت بزرگ در طراحی شبکه‌های عصبی عمیق محسوب می‌شود.

یکی از کارهای انجام شده در زمینه استفاده از ساختارهای توجه در معماری رمزگذار-رمزگشا، روش DCSAU-Net [۳۰] با قابلیت تعمیم بر روی مجموعه داده‌های مختلف است. در بخش رمزگذار این مدل، از استراتژی حفظ ویژگی^{۱۱} استفاده شده است تا علاوه بر کاهش پارامترها و محاسبات، اطلاعات فضایی دوربرد در لایه‌های عمیق‌تر محاسبه شود. ویژگی‌های حاصل شده از این مسیر، در اختیار ساختار طراحی شده با عنوان CSA^{۱۲} در لایه‌های رمزگذار و رمزگشا قرار می‌گیرد تا با استفاده از ساختار توجه چندگانه که شامل لایه‌های کانولوشن متفاوت است، ترکیبی از نقشه‌های ویژگی با میدان پذیرش متفاوت را خروجی دهد.

اگر چه روش‌های مبتنی بر شبکه‌های عصبی کانولوشن عملکرد خوبی در قطعه بندی تصاویر پزشکی دارند، اما با توجه به اینکه اندازه هسته لایه‌های کانولوشن محدود است، عموماً در به دست آوردن زمینه کلی تصویر و اطلاعات دوربرد دچار مشکل هستند و تمرکز خود را در نواحی محلی تصویر می‌گذارند [۱۵]؛ به همین منظور، برخی روش‌ها مانند مکانیزم خود توجه براساس ویژگی کانولوشن [۱۶] و ویژگی‌های هرمی [۱۷] ارائه شدند. با این حال این روش‌ها فقط فیلد دریافتی شبکه را گسترش می‌دادند اما نمی‌توانستند به خوبی مدل‌سازی ویژگی‌های سراسری به خصوص در تصاویر پیچیده را انجام دهند.

ترانسفورمرها در پردازش زبان طبیعی^{۱۳} به طور گسترده استفاده می‌شوند. توانایی فوق‌العاده آن‌ها در یادگیری دنباله‌های ویژگی و مدل سازی وابستگی‌های دوربرد باعث شد تا در زمینه بینایی ماشین معرفی شوند [۱۸]. با توسعه VIT [۱۹] تحقیقات درباره استفاده از ترانسفورمرها در قطعه بندی تصاویر پزشکی شروع شد. در ابتدا TransUnet [۲۰] نتیجه خوبی در قطعه بندی تصاویر قلب و اندام‌های شکمی به دست آورد. در این روش از ماژول خود توجه به جای لایه‌های کانولوشن استفاده شد. پس از این مقاله، کارهای زیادی در رابطه با قطعه بندی تصاویر پزشکی براساس این ساختار مانند: Swin-Unet [۲۱]، UTNet [۲۲] و

¹⁰ Attention Gates

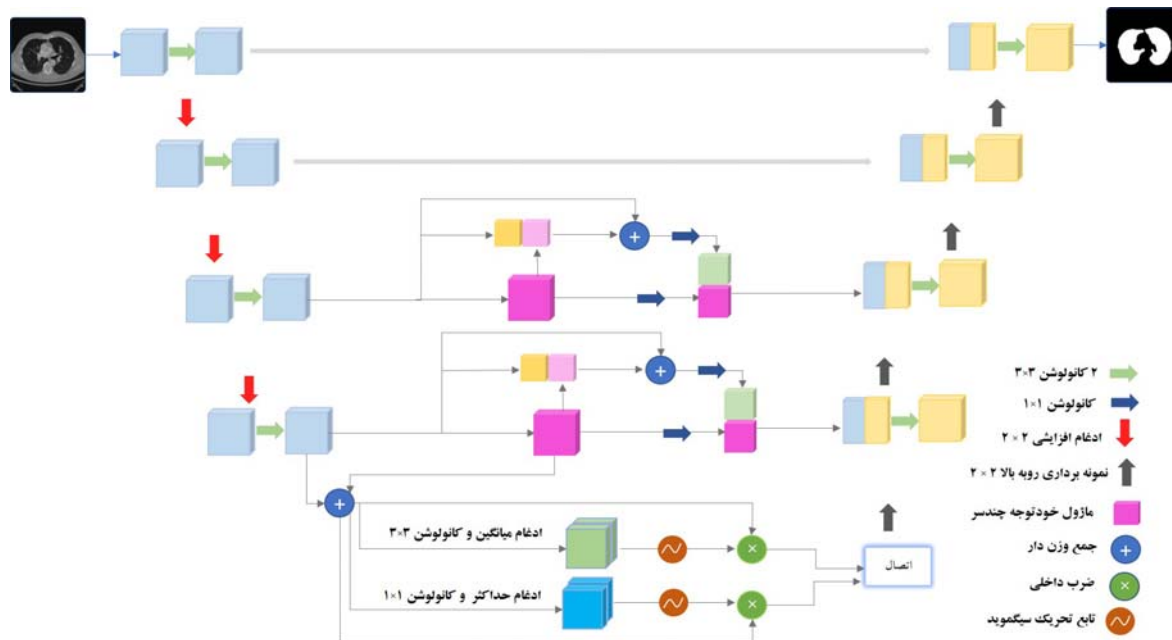
¹¹ primary feature conservation

¹² Compact Split Attention

¹³ Natural Language Processing

¹⁴ Self-aware Attention

¹⁵ Gloal Spatial Attention

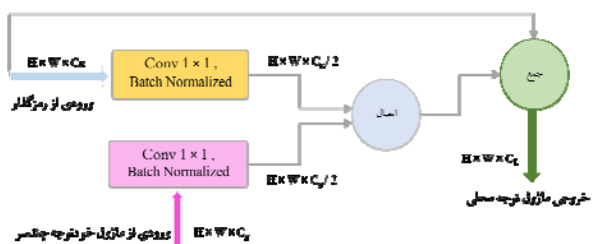


شکل (۱): معماری کلی شبکه طراحی شده، در قسمت اتصالات پرش ماژول‌های توجه محلی و سراسری استفاده شده است و در قسمت گلوگاه ماژول توجه فضایی به کانال‌ها تعبیه شده است.

را در اتصالات پرش ندارند. به همین منظور یک ماژول ترکیب ویژگی‌های سراسری و محلی را در اتصالات پرش سطوح ۳ و ۴ طراحی شده است. علاوه بر این، به منظور ترکیب بهتر ویژگی‌های فضایی و عمقی یک ماژول دیگر در قسمت گلوگاه شبکه طراحی شده است تا علاوه بر استخراج بهتر ویژگی‌ها، محاسبات را با تعداد پارامتر کمتری انجام دهد.

۳-۲ ماژول توجه محلی

هدف این قسمت، طراحی یک ساختار برای استخراج و وزن‌دهی بیشتر به ویژگی‌های محلی است (شکل ۲). در این ماژول، از لایه‌های کانولوشن با اندازه هسته 1×1 جهت کاهش تعداد کانال‌های نقشه ویژگی استفاده شده است. این کانولوشن‌ها شاید در نگاه اول به دلیل واحد بودن هسته پیش‌همانند یک ضرب ساده به نظر برسند، اما در حقیقت یک جمع وزن‌دار از کانال‌های نقشه ویژگی به ما می‌دهند [۲۵]. این کانولوشن‌ها به دلیل واحد بودن هسته پیش‌همانند تعداد پارامترهای شبکه را افزایش قابل توجهی نمی‌دهند.



شکل (۲): ماژول توجه محلی

بر روی مجموعه داده‌های مختلف را ندارند. همچنین ساختارهای جدید معرفی شده برای استخراج بهتر ویژگی‌ها از جمله ترانسفورمرها، عیب‌هایی از جمله پیچیدگی محاسباتی زیاد و در نظر نگرفتن ویژگی‌های محلی دارند. در این مقاله سعی شده است که تاثیر ساختارهای توجه محلی در کنار توجه سراسری مورد بررسی قرار گیرد. همچنین با تغییر ساختار گلوگاه شبکه U-net، به کاهش سربار محاسباتی و پارامترهای شبکه کمک شود. در ادامه این مقاله، روش پیشنهادی خود را جهت دستیابی به نتایج مطلوب تر در قطعه‌بندی تصاویر پزشکی، بررسی خواهیم کرد.

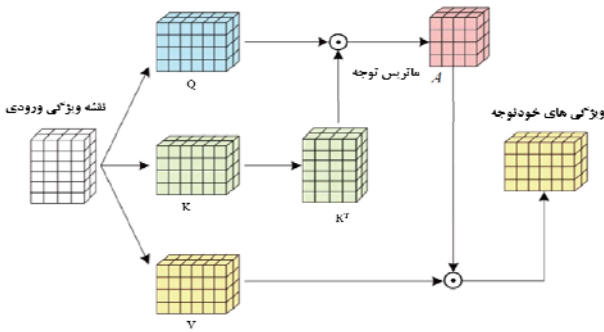
۳-۱ روش پیشنهادی

در این قسمت، معماری کلی شبکه در بخش ۱-۳، در بخش ۲-۳ معماری طراحی شده در قسمت اتصالات پرش برای توجه محلی، در بخش ۳-۳ ماژول خودتوجه چندسری برای توجه سراسری و سپس در بخش ۳-۴ ساختار طراحی شده در قسمت گلوگاه بررسی می‌شود.

۳-۱ معماری کلی شبکه

معماری روش پیشنهاد شده برای بهبود اتصالات پرش و گلوگاه در شکل ۱ نشان داده شده است. مدل ما براساس روش U-net، دارای ۴ لایه کاهش نمونه برداری^{۱۶} برای به دست آوردن اطلاعات غنی‌تر و نظیر آن ۴ لایه افزایش نمونه برداری^{۱۷} برای بازیابی نقشه ویژگی‌ها به اندازه اصلی است. روش U-net در مقایسه با روش پیشنهاد شده، توانایی ترکیب ویژگی‌های چندمقیاسی و توجه به ویژگی‌ها

¹⁶ Downsampling
¹⁷ Upsampling



شکل (۳): ماژول خودتوجه برای استخراج ویژگی های سراسری و در نظر گرفتن وابستگی های دوربرد میان پیکسل ها - در این شکل، یک بعد (سر) از ماژول خودتوجه نشان داده شده است.

۳-۴ ماژول توجه فضایی

این ماژول برای برجسته‌تر کردن ویژگی‌های معنایی به صورت فضایی و کانالی، در آخرین لایه رمزگذار شبکه U-net که اصطلاحاً گلوگاه نامیده می‌شود، تعبیه شده است. در ساختار اولیه U-net، گلوگاه شبکه به علت متقارن بودن ساختار رمزگذار-رمزگشا و استفاده از ویژگی‌های عمیق‌تر، با اعمال دو لایه کانولوشن 3×3 تعداد کانال‌های نقشه ویژگی را به ۲ برابر افزایش می‌دهد. این عملیات به دلیل اینکه در عمیق‌ترین سطح شبکه انجام می‌شود، سربار محاسباتی زیادی را به شبکه تحمیل می‌کند. محاسبات زیاد در این ناحیه باعث کندشدن یادگیری خواهد شد. به همین دلیل ساختار متفاوتی برای این ناحیه طراحی شده است تا علاوه بر ترکیب و توجه بیشتر به ویژگی‌های مکانی و محلی، محاسبات کمتری انجام شود. همانطور که در شکل ۴ مشاهده می‌شود، ساختار پیشنهاد شده برای گلوگاه، یک ورودی از رمزگذار سطح ۴ و ورودی دیگری از ماژول خودتوجه چندسری که در قسمت اتصالات پرش وجود دارد، دریافت می‌کند. از جمع این ورودی‌ها، نقشه ویژگی $X \in \mathbb{R}^{H \times W \times C}$ محاسبه می‌شود. طبق رابطه ۲، کانولوشن 1×1 همراه با ادغام حداکثر و بار دیگر کانولوشن 3×3 همراه با ادغام میانگین روی ورودی اعمال می‌شود تا ویژگی‌های فضایی و مکانی استخراج شوند.

$$C2 = X \cdot \sigma(\text{Conv}(\text{MaxPool}(X))) \quad (2)$$

$$C1 = X \cdot \sigma(\text{Conv}(\text{AvgPool}(X)))$$

$$C = \text{concat}(C1, C2)$$

مطابق رابطه ۲، به منظور نرمالیزه کردن و فعال سازی نواحی مهم، تابع تحریک سیگموئید را روی $C1 \in \mathbb{R}^{H \times W \times C/2}$ و $C2 \in \mathbb{R}^{H \times W \times C/2}$ اعمال می‌شود. سپس نقشه ویژگی‌ها به یکدیگر متصل و در اختیار لایه رمزگشا قرار داده می‌شود.

بعد از هربلاک کانولوشن، یک نرمال‌ساز دسته ای^{۱۸} برای سریع‌تر و جلوگیری از بیش برازش^{۱۹} شبکه طراحی شده است. این ماژول دو نقشه ویژگی به عنوان ورودی دریافت می‌کند. نقشه ویژگی اول، $X \in \mathbb{R}^{H \times W \times C}$ از لایه رمزگذار به طور مستقیم به بلاک کانولوشن 1×1 وارد می‌شود و پس از اعمال فیلترهای کانولوشنی و نرمال‌ساز دسته‌ای، تعداد کانال‌ها نصف می‌شود. نقشه ویژگی دوم، $G \in \mathbb{R}^{H \times W \times C}$ از ماژول خودتوجه چندسری به عنوان ورودی وارد بلاک کانولوشن 1×1 می‌شود و تعداد کانال‌ها نصف و نرمال‌ساز دسته ای روی آن اعمال می‌شود. سپس $X \in \mathbb{R}^{H \times W \times C/2}$ و $G \in \mathbb{R}^{H \times W \times C/2}$ به یکدیگر متصل می‌گردند و تابع سیگموئید به عنوان فعال ساز روی نقشه ویژگی به دست آمده، اعمال می‌شود. این نقشه ویژگی با نقشه ویژگی اولیه‌ای که از لایه رمزگذار دریافت شده است، جمع خواهد شد. بنابراین نقشه ویژگی بدست آمده در این مرحله حاوی ویژگی‌های محلی خواهد بود که در مرحله بعد با نقشه ویژگی به دست آمده از ماژول خود توجه ترکیب خواهد شد.

۳-۳ ماژول خودتوجه چندسری

خودتوجهی^{۲۰} یکی از انواع مکانیزم توجه است که با محاسبه ماتریس همبستگی، وابستگی دو به دو را بین تمام موجودیت‌ها محاسبه می‌کند و قابلیت درک وابستگی‌های دوربرد را دارد. خود توجهی برای اولین بار در پردازش زبان طبیعی و سپس در بینایی ماشین مورد استفاده قرار گرفت. در این مقاله، از ماژول خودتوجه چند سری به منظور استفاده عمیق‌تر از فضای ویژگی موجود استفاده شده است (شکل ۳).

در رابطه ۱، V, Q, K از اعمال لایه‌های کانولوشن با اندازه هسته 1×1 روی ورودی $X \in \mathbb{R}^{H \times W \times C}$ محاسبه می‌شوند. به این ترتیب $Q, K, V \in \mathbb{R}^{H \times W \times d}$ هر کدام دارای بعد d خواهند بود. d در اینجا بعد تعبیه شده در هر سر است. در این مقاله، از ۴ سر برای ماژول خودتوجه استفاده شده است.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

در رابطه ۱، سه مقدار K, Q, V مقادیر تبدیل یافته از یک ورودی یکسان می‌باشند. در این رابطه، ابتدا ماتریس توجه از ضرب داخلی Q و ترانهاده K به دست می‌آید. سپس ماتریس توجه در V ضرب داخلی می‌شود و ویژگی‌های خودتوجه حاصل خواهند شد. در اینجا چون از ماژول توجه چند سر استفاده می‌کنیم، ویژگی‌های به دست آمده در هر بعد، به یکدیگر متصل خواهند شد.

¹⁸ Batch Normalization

¹⁹ Over-Fitting

²⁰ Self Attention

۲-۴ معیارهای ارزیابی

میانگین ضریب شباهت دایس^{۲۲} و میانگین اشتراک اتحادها^{۲۳} از متداول‌ترین معیارهای ارزیابی در قطعه‌بندی تصاویر هستند. ضریب شباهت دایس (رابطه ۳) بیانگر ۲ برابر همپوشانی داده مبنا و نتیجه پیش بینی شده است. به این معنی که پیکسل‌های نتیجه پیش‌بینی شده و داده مبنا با یکدیگر مقایسه می‌شوند و ۲ برابر تعداد پیکسل‌هایی که همپوشانی داشته باشند در صورت رابطه قرار خواهد گرفت. در قسمت مخرج رابطه ۳ نیز، مجموع تعداد پیکسل‌های نتیجه پیش‌بینی شده و داده مبنا قرار می‌گیرد.

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

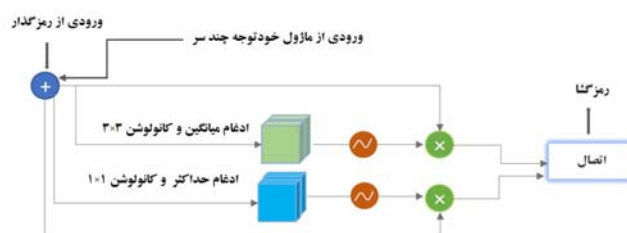
میانگین اشتراک اتحادها (IoU) از همپوشانی داده مبنا و نتیجه پیش‌بینی شده بر اجتماع پیکسل‌های داده مبنا و پیش‌بینی محاسبه می‌شود. IoU، Dice در بیشتر موارد همبستگی مثبت دارند. به عنوان مثال اگر Dice یک پایگاه داده از پایگاه داده دیگر بیشتر باشد، IoU آن نیز بیشتر خواهد بود.

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

برای مقایسه روشن‌تر نتایج، دو معیار Precision و Recall نیز محاسبه شده است.

۳-۴ جزئیات پیاده‌سازی

تمام آزمایش‌ها در این مقاله در چارچوب کتابخانه پایتورچ^{۲۴}، پردازنده گرافیکی GPU Tesla p100 و ۱۲ گیگابایت حافظه انجام شده است. ۸۰ درصد داده‌ها برای آموزش، ۱۰ درصد برای ارزیابی و ۱۰ درصد برای آزمون استفاده شده است. پارامترها و توابع تنظیم شده در پیاده‌سازی بهینه ساز آدام^{۲۵}، تابع هزینه دایس، نرخ یادگیری^{۲۶} ۱۰^{-۴} است. همچنین تعداد ۲۰۰ تکرار برای آموزش تنظیم شده است. اندازه دسته‌ها^{۲۶} با توجه به محدودیت سخت‌افزاری و تنوع داده‌ها در پایگاه داده‌های 2018-DSB، Cvc-clinicDB و Kvasir برابر ۸ و برای سایر پایگاه‌ها ۲ در نظر گرفته شده است.



شکل (۴): ماژول توجه فضایی در گلوگاه شبکه

۴- پیاده‌سازی و نتایج

۱-۴ پایگاه داده‌های آموزشی

در این مقاله برای اثبات کارآمدی نتیجه، از ۶ پایگاه داده تصویر پزشکی ۲ بعدی استفاده شده است. پایگاه داده اول (LUNA)^{۲۱} [۲۶] شامل ۲۶۷ تصویر سی تی اسکن با ابعاد ۵۱۲×۵۱۲ از ریه بیماران مشکوک به سرطان است.

پایگاه داده دوم Kaggle 2018 data science bowl [۲۷] شامل ۶۷۰ تصویر سلول میکروسکوپی که به ابعاد ۲۵۶×۲۵۶ تغییر اندازه داده شده است. ۸۰ درصد تصاویر به عنوان داده آموزشی، ۱۰ درصد برای ارزیابی و ۱۰ درصد برای آزمون به صورت تصادفی تقسیم شده‌اند. در پایگاه داده‌های سوم و چهارم Kvasir [۲۸] و CVC-ClinicDB [۲۹] به ترتیب شامل ۱۰۰۰ و ۶۱۲ تصویر پولیپ روده و داده مبنا متناظر آن قرار دارد. مجموعه داده پنجم، مجموعه داده‌ی تجزیه و تحلیل لکه‌های پوستی ISIC2017 است که در سال ۲۰۱۷ منتشر شده است. این مجموعه شامل ۲۰۰۰ تصویر آموزشی است که ما به دلیل محدودیت‌های سخت‌افزاری ۸۰۰ تصویر را به صورت تصادفی از انتخاب کرده ایم. مجموعه داده ششم شامل ۲۰ تصویر از عروق خونی شبکیه چشم است که با چرخش آن‌ها به ۴ جهت، داده‌ها به ۸۰ تصویر افزایش داده شده‌اند. جزئیات بیشتر پایگاه داده‌های آموزشی را در جدول ۱ نشان داده شده است.

جدول ۱: جزئیات پایگاه داده‌های آموزشی، تعداد و ابعاد آن‌ها

پایگاه داده	تعداد	آزمون	ارزیابی	ابعاد تصویر
Kvasir	۱۰۰۰	۱۰۰	۱۰۰	۲۵۶×۲۵۶
Cvc-ClinicDB	۶۱۲	۶۱	۶۱	۳۸۴×۲۸۸
2018-DSB	۶۷۰	۶۷	۶۷	۲۵۶×۲۵۶
ISIC2017	۸۰۰	۱۵۰	۱۵۰	۵۱۲×۵۱۲
Drive-2004	۸۰	۱۰	۱۰	۵۱۲×۵۱۲
LUNA	۳۲۱	۵۴	۵۴	۵۱۲×۵۱۲

²² Mean Dice Similarity Coefficient (Dice)

²³ Intersection Over Union (IoU)

²⁴ Pytorch

²⁵ Adam optimiser

²⁶ Batch size

²¹ Lung Nodule Analysis

۴-۵ نتایج پیاده‌سازی

به ترتیب ۳٪ و ۳٫۸٪ در معیارهای Dice و Iou پیشرفت نتایج را نسبت به روش Unet داشته‌ایم. همچنین روش پیشنهاد شده، اما Precision بالاتری در مقایسه با سایر روش‌ها کسب کرده است. اما Recall روش پیشنهادی، نسبت به روش TransAttUnet کمتر است. روش پیشنهاد شده در این مجموعه داده در حد قابل قبولی نتایج را بهبود داده است. نتایج روش پیشنهادی روی مجموعه داده 2018-DSB و مقایسه با سایر روش‌ها، در جدول ۳ نشان داده شده است.

۴-۵-۳ مجموعه داده CVC-ClinicDB

پولیپ‌ها و توده‌های موجود در سیستم گوارش انسان، یکی از اصلی‌ترین راه‌های تشخیص سرطان در روده‌ها هستند که با تشخیص سریع می‌توان از پیشرفت آن‌ها جلوگیری کرد. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰٫۸۸۷ و در معیار Iou برابر ۰٫۸۲۵ است، که توانسته تنها در دو معیار Iou و Precision نتایج را نسبت به روش‌های مقایسه شده بهبود دهد. در این مجموعه، به ترتیب ۴٫۶٪ و ۳٫۶٪ در معیارهای Iou و Dice پیشرفت نتایج را نسبت به روش Unet داشته‌ایم. روش پیشنهاد شده اگر چه توانسته است با پیچیدگی محاسباتی کمتری برخی معیارها را بهبود دهد اما در برخی معیارهای دیگر، مانند Dice نتایج کمتری نسبت به روش‌های مقایسه شده، کسب کرده است. بنابراین روش پیشنهاد شده در این مجموعه داده، هنوز جای پیشرفت دارد. نتایج این آزمایش در جدول ۴ نمایش داده شده است.

برای نشان دادن بهبود نتایج، مدل پیشنهادی با روش‌های AttentionUnet, UTNet, Unet, TransAttUnet, Unet3+, Unet++ , DCSAU-Net مقایسه شده است. جدول ۱ مقایسه بین مدل‌های دیگر و روش پیشنهاد شده در پارامترها و سال انتشار را نشان می‌دهد. ساختار طراحی شده در اتصالات پرش، به دلیل استفاده از مکانیزم‌های توجه و لایه‌های کانولوشن با اندازه هسته ۱×۱، پارامتر زیادی به شبکه وارد نمی‌کند (حدود ۱ میلیون پارامتر محاسباتی). روش پیشنهاد شده در این مقاله، به دلیل ساختار طراحی شده در گلوگاه، پارامترهای شبکه را در حدود ۱۰ میلیون نسبت به روش Unet کاهش داده است. تعداد پارامترها و پیچیدگی محاسباتی روش پیشنهادی، نسبت به سایر روش‌های مقایسه شده در این مقاله، کاهش چشمگیری داشته است. در ادامه این بخش، نتایج بدست آمده روش پیشنهادی با سایر روش‌ها مقایسه خواهد شد.

جدول ۱ مقایسه تعداد پارامترهای روش پیشنهاد شده با سایر روش‌ها

روش	سال انتشار	پارامترها
Unet	۲۰۱۵	۳۱،۰۴
Unet++	۲۰۱۸	۳۶،۶۲
Unet3+	۲۰۲۰	۲۶،۹۷
Attention Unet	۲۰۱۸	۳۴،۸۷
UTNet	۲۰۲۱	۵۷،۵۰
DCSAU-Net	۲۰۲۳	۲۵،۹۸
TransAttUnet	۲۰۲۳	۲۵،۹۴
پیشنهادی	۲۰۲۳	۲۱،۰۳

۴-۵-۱ مجموعه داده LUNA

تشخیص سرطان ریه در مراحل ابتدایی از اهمیت بسزایی برخوردار است. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰٫۹۸۸ و در معیار Iou برابر ۰٫۹۷۶ است که توانسته بهترین نتایج را نسبت به روش‌های مقایسه شده در این معیارها کسب کند. در این مجموعه، به ترتیب ۲٫۱٪ و ۳٫۳٪ در معیارهای Dice و Iou پیشرفت نتایج را نسبت به روش Unet داشته‌ایم. همچنین روش پیشنهاد شده، توانسته Recall بالاتری در مقایسه با سایر روش‌ها کسب کند. بنابراین روش پیشنهاد شده در این مجموعه داده، نتایج بهتری را نسبت به روش‌های مشابه کسب کرده است. نتایج روش پیشنهادی روی این مجموعه داده و مقایسه آن با سایر روش‌ها، در جدول ۲ نشان داده شده است.

جدول ۲ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده LUNA

روش	Dice	Iou	Precision	Recall
Unet(2015)	۰٫۹۶۷	۰٫۹۳۵	۰٫۹۷۱	۰٫۹۶۹
Unet++(2018)	۰٫۹۵۶	۰٫۹۲۳	۰٫۹۷۴	۰٫۹۶۹
Unet3+(2020)	۰٫۹۲۵	۰٫۹۰۱	۰٫۹۶۸	۰٫۹۳۹
Attention Unet(2018)	۰٫۹۶۲	۰٫۹۴۹	۰٫۹۷۹	۰٫۹۷۰
UTNet(2021)	۰٫۹۷۶	۰٫۹۵۶	۰٫۹۹۱	۰٫۹۶۷
DCSAU-Net(2023)	۰٫۹۷۴	۰٫۹۵۹	۰٫۹۷۷	۰٫۹۷۴
TransAttUnet(2023)	۰٫۹۷۹	۰٫۹۶۸	۰٫۹۸۴	۰٫۹۸۲
پیشنهادی	۰٫۹۸۸	۰٫۹۷۶	۰٫۹۹۰	۰٫۹۸۶

جدول ۳ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده 2018-DSB

روش	Dice	Iou	Precision	Recall
Unet(2015)	۰٫۹۰۸	۰٫۸۳۷	۰٫۹۲۰	۰٫۹۰۹
Unet++(2018)	۰٫۸۷۶	۰٫۸۱۵	۰٫۸۷۲	۰٫۹۱۸
Unet3+(2020)	۰٫۸۹۳	۰٫۸۳۴	۰٫۸۸۹	۰٫۹۰۷
Attention Unet(2018)	۰٫۸۸۹	۰٫۸۲۵	۰٫۹۲۳	۰٫۸۸۸
UTNet(2021)	۰٫۸۹۲	۰٫۸۱۶	۰٫۹۳۱	۰٫۸۷۵
DCSAU-Net(2023)	۰٫۸۸۲	۰٫۸۰۲	۰٫۹۲۶	۰٫۸۵۷
TransAttUnet(2023)	۰٫۹۱۷	۰٫۸۵۱	۰٫۹۱۷	۰٫۹۲۴
پیشنهادی	۰٫۹۲۸	۰٫۸۷۵	۰٫۹۳۳	۰٫۹۱۲

۴-۵-۲ مجموعه داده 2018-DSB

بررسی دقیق مولکول‌های خونی و تغییر شکل آن‌ها، مرحله مهمی در تشخیص بسیاری از بیماری‌ها است. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰٫۹۲۸ و در معیار Iou برابر ۰٫۸۷۵ است، که توانسته بهترین نتایج را نسبت به روش‌های مقایسه شده در این معیارها کسب کند. در این مجموعه،

Iou برابر ۰,۶۱۹ است، که بهترین نتایج را نسبت به روش‌های مقایسه شده در این معیارها توانسته است کسب کند. در این مجموعه، به ترتیب ۱۰,۵٪ و ۱۰,۸ در معیارهای Dice و Iou پیشرفت نتایج را نسبت به روش Unet داشته ایم. همچنین روش پیشنهاد شده، Recall و Percision بالاتری در مقایسه با سایر روش‌ها کسب کرده است. به دلیل رزولوشن و تعداد بالای تصاویر، تار و نازک بودن مرزهای نواحی قطعه‌بندی شده و محدودیت‌های سخت‌افزاری در اختیار ما، نتایج قطعه‌بندی انجام شده در این مجموعه داده به طور کلی ضعیف بودند اما روش پیشنهاد شده روی مجموعه داده در حد قابل قبولی توانسته است نتایج را بهبود دهد. نتایج روش پیشنهادی روی مجموعه داده DRIVE و مقایسه با سایر روش‌ها، در جدول ۷ نشان داده شده است.

جدول ۴ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده CVC-ClinicDB

روش	Dice	Iou	Percision	Recall
Unet(2015)	۰,۸۵۱	۰,۷۷۹	۰,۸۸۱	۰,۸۵۶
Unet++(2018)	۰,۸۹۲	۰,۸۲۲	۰,۹۰۶	۰,۹۰۴
Unet3+(2020)	۰,۸۲۹	۰,۷۴۶	۰,۸۷۷	۰,۸۴۷
Attention Unet(2018)	۰,۸۲۴	۰,۷۶۳	۰,۸۸۶	۰,۸۵۸
UTNet(2021)	۰,۸۸۳	۰,۸۲۰	۰,۹۰۹	۰,۸۸۱
DCSAU-Net(2023)	۰,۸۳۸	۰,۷۵۱	۰,۸۹۸	۰,۸۲۱
TransAttUnet(2023)	۰,۸۸۱	۰,۸۱۹	۰,۸۷۰	۰,۹۱۰
پیشنهادی	۰,۸۸۷	۰,۸۲۵	۰,۹۱۳	۰,۸۸۹

جدول ۵ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده Kvassir

روش	Dice	Iou	Percision	Recall
Unet(2015)	۰,۷۹۲	۰,۷۰۸	۰,۸۵۲	۰,۸۰۱
Unet++(2018)	۰,۷۸۰	۰,۶۸۵	۰,۸۵۹	۰,۷۸۲
Unet3+(2020)	۰,۷۷۲	۰,۶۷۳	۰,۷۶۶	۰,۷۸۸
Attention Unet(2018)	۰,۸۰۵	۰,۷۱۶	۰,۸۶۲	۰,۸۳۱
UTNet(2021)	۰,۸۱۶	۰,۷۳۵	۰,۸۶۵	۰,۸۳۰
DCSAU-Net(2023)	۰,۷۶۱	۰,۶۷۴	۰,۸۳۰	۰,۷۶۳
TransAttUnet(2023)	۰,۸۱۵	۰,۷۲۲	۰,۸۶۵	۰,۸۱۲
پیشنهادی	۰,۸۲۲	۰,۷۳۲	۰,۸۷۶	۰,۸۲۵

جدول ۶ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده DRIVE

روش	Dice	Iou	Percision	Recall
Unet(2015)	۰,۷۵۳	۰,۶۰۵	۰,۸۰۵	۰,۷۷۳
Unet++(2018)	۰,۸۰۰	۰,۶۶۵	۰,۸۱۰	۰,۷۹۶
Unet3+(2020)	۰,۵۶۴	۰,۴۱	۰,۴۷۲	۰,۷۹۰
Attention Unet(2018)	۰,۸۰۱	۰,۶۶۷	۰,۸۱۹	۰,۷۸۷
UTNet(2021)	۰,۸۰۰	۰,۶۷۱	۰,۷۹۹	۰,۸۰۷
DCSAU-Net(2023)	۰,۷۷۹	۰,۶۳۹	۰,۸۳۲	۰,۷۳۷
TransAttUnet(2023)	۰,۷۸۷	۰,۶۵۰	۰,۸۳۷	۰,۷۴۹
پیشنهادی	۰,۷۹۶	۰,۶۶۹	۰,۷۹۶	۰,۷۸۷

۴-۵-۴ مجموعه داده Kvassir

این مجموعه داده، همانند مجموعه CVC-ClinicDB یکی از پرکاربردترین پایگاه داده‌ها در تشخیص پولیپ‌ها و توده‌های موجود در سیستم گوارش انسان است. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰,۸۲۲ و در معیار Iou برابر ۰,۷۳۲ است، که توانسته بهترین نتایج را نسبت به روش‌های مقایسه شده در این معیارها کسب کند. در این مجموعه، به ترتیب ۳٪ و ۲,۴٪ در معیارهای Dice و Iou پیشرفت نتایج را نسبت به روش Unet داشته ایم. همچنین روش پیشنهاد شده، Precision بالاتری در مقایسه با سایر روش‌ها کسب کرده است. اما Recall روش پیشنهادی، نسبت به روش Attention Unet کمتر است. روش پیشنهادی نتایج قابل قبولی نسبت به سایر روش‌های مقایسه شده، کسب کرده است. نتایج انجام شده روی مجموعه داده Kvassir، در جدول ۵ نشان داده شده است.

۴-۵-۵ مجموعه داده DRIVE

مجموعه داده DRIVE، یکی از مجموعه داده‌های متداول در زمینه بررسی نتایج قطعه‌بندی استفاده است. قطعه‌بندی عروق و تعریف صفات عروق خونی شبکه‌ی چشم مانند طول، عرض و پیچیدگی برای تشخیص، درمان و ارزیابی بیماری‌های عروقی و چشمی مانند دیابت و فشار خون استفاده می‌شود. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰,۷۹۶ و در معیار Iou برابر ۰,۶۶۹ است. در این مجموعه، به ترتیب ۴,۳٪ و ۶,۴٪ در معیارهای Dice و Iou پیشرفت نتایج را نسبت به روش Unet داشته ایم. اگر چه روش پیشنهادی توانسته است نتایج کسب شده در Unet را بهبود دهد اما در مقایسه با روش‌های دیگر، نتایج کمی ضعیف‌تر کسب کرده است. برای این مجموعه داده، ساختارهای اختصاصی زیادی برای بررسی قطعه‌بندی در آن معرفی شده است اما در ساختارهای پیشنهاد شده عمومی، به علت نازک بودن عروق خونی و عدم تشخیص پیکسل‌ها در نواحی بسیار نازک، نتیجه مناسبی کسب نمی‌شود. با این حال روش پیشنهاد شده در این مقاله توانسته است قطعه‌بندی را با تعداد پارامتر کمتری نسبت به روش Unet بهبود دهد. نتایج روش پیشنهادی روی مجموعه داده DRIVE و مقایسه با سایر روش‌ها، در جدول ۶ نشان داده شده است.

۴-۵-۶ مجموعه داده ISIC2017

مجموعه داده ISIC2017، شامل تصاویری از نواحی مختلف بدن از جمله صورت، سینه و سایر نواحی بدن است. این تصاویر برای تشخیص بیماری‌های پوستی مانند مول، سرطان پوستی و بیماری‌های پوستی دیگر استفاده می‌شود. روش پیشنهاد شده بر روی این مجموعه داده، در معیار Dice برابر ۰,۷۲۲ و در معیار

مراجع

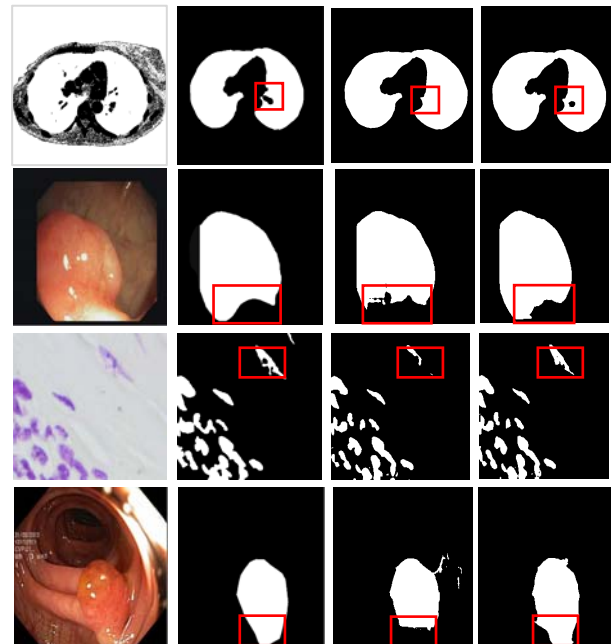
- [1] Vedaldi A, Lenc K (2015) Matconvnet: convolutional neural networks for matlab. In: Proceedings of the 23rd ACM international conference on Multimedia. ACM, pp 689–692, 2015.
- [2] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers et al., "The medical segmentation decathlon," Nature communications, vol. 13, no. 1, pp. 1–13, 2022.
- [3] H.R. Tizhoosh, Image thresholding using type II fuzzy sets, Pattern Recognit., vol. 38, no. 12, pp. 2363–2372, 2005.
- [4] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. vol. 9, no. 1, pp. 62–66, 1979.
- [5] Yuanzhi Cheng, Xin Hu, Ji Wang, Yadong Wang, Shinichi Tamura, Accurate vessel segmentation with constrained B-snake, IEEE Trans. Image Process. vol. 24, no. 8, pp. 2440–2455, 2015.
- [6] Guo Y, Ashour A (2019) Neutrosophic sets in dermoscopic medical image segmentation. Neutrosophic Set Med Image Anal. vol.11, no. 4, pp. 229–243, 2019.
- [7] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440, 2015
- [8] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 234–241, 2015.
- [9] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999, 2018.
- [10] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, pp. 3–11, 2018.
- [11] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," IEEE Trans. Med. Imag., vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

جدول ۷ مقایسه نتایج روش پیشنهادی با سایر روش‌ها در پایگاه داده ISIC2017

روش	Dice	Iou	Precision	Recall
Unet(2015)	۰٫۶۱۷	۰٫۵۱۱	۰٫۶۷۹	۰٫۷۷۰
Unet++(2018)	۰٫۶۳۰	۰٫۵۲۸	۰٫۷۳۲	۰٫۷۴۱
Unet3+(2020)	۰٫۵۶۶	۰٫۴۶۴	۰٫۶۳۲	۰٫۷۲۱
Attention Unet(2018)	۰٫۶۲۷	۰٫۵۲۳	۰٫۶۸۴	۰٫۷۷۶
UTNet(2021)	۰٫۶۹۱	۰٫۵۸۳	۰٫۷۷۱	۰٫۷۸۰
DCSAU-Net(2023)	۰٫۷۱۳	۰٫۶۰۸	۰٫۷۸۰	۰٫۷۸۸
TransAttUnet(2023)	۰٫۷۰۵۲	۰٫۵۹۹	۰٫۷۱۷	۰٫۸۲۳
پیشنهادی	۰٫۷۲۲	۰٫۶۱۹	۰٫۸۳۰	۰٫۷۳۲

۵- نتیجه‌گیری و بحث

قطعه بندی تصاویر پزشکی یک گام مهم در تحلیل و پردازش داده‌ها در حوزه پزشکی است. در این مقاله، سعی شده است تا یک مدل براساس ساختار Unet، جهت بهبود قطعه‌بندی تصاویر پزشکی، ارائه شود. در این مقاله، ساختار جدیدی برای ترکیب و استخراج ویژگی‌های محلی و سراسری در ناحیه اتصالات پرش طراحی شده است. همچنین در قسمت گلوگاه شبکه، یک ساختار برای توجه به ویژگی‌های فضایی و کانالی طراحی شده است تا علاوه بر کاهش پیچیدگی و پارامترهای محاسباتی در این ناحیه، قطعه‌بندی مطلوب‌تری را نتیجه دهد. آزمایش‌ها نشان می‌دهد روش پیشنهادی، باعث بهبود نتیجه در قطعه‌بندی تصاویر پزشکی می‌شوند. کارهای آینده ما تمرکز بر استخراج و یادگیری بهتر ویژگی‌ها در ناحیه رمزگذار و همچنین ارائه یک ساختار جدید برای متصل کردن ویژگی‌ها در ناحیه رمزگشا است.



شکل (۴): نتایج بدست آمده از قطعه بندی تصاویر (الف) تصویر اصلی، (ب) داده مبنا، (ج) روش Unet، (د) روش پیشنهادی ما

- segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 61–71, 2021.
- [23] B. Chen, Y. Liu, Z. Zhang, G. Lu and A. W. K. Kong, "TransAttUnet: Multi-Level Attention-Guided U-Net With Transformer for Medical Image Segmentation," in IEEE Transactions on Emerging Topics in Computational Intelligence, doi: 10.1109/TETCI.2023.3309626.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30, 2017.
- [25] M. Lin, Q. Chen, and S. Yan. Network in network. arXiv:1312.4400,2013.
- [26] Adiyoso Arnaud Arindra Setio, et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge, Med. Image Anal. pp. 1–13, 2017
- [27] J.C. Caicedo, A. Goodman, K.W. Karhohs, B.A. Cimini, J. Ackerman, M. Haghighi, C. Heng, T. Becker, M. Doan, C. McQuin, et al., Nucleus segmentation across imaging experiments: The 2018 data science bowl, Nature Methods.vol. 16, no. 12, pp. 1247–1253, 2019.
- [28] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., Lange, T. d., Johansen, D., and Johansen, H. D., "Kvasirseg:A segmented polyp dataset," in [International Conference on Multimedia Modeling (MMM)],Springer,pp. 451–462, Dec 2019.
- [29] Bernal, J., S´anchez, F. J., Fern´andez-Esparrach, G., Gil, D., Rodr´ıguez, C., and Vilari˜no, F., "Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," Comput. Med. ImagingGraph (CMIG), vol. 43,pp. 99–111, July 2015.
- [30] Xu, Q., Ma, Z., He, N., Duan, W. DCSAU-Net: A deeper and more compact splitattention U-net for medical image segmentation. Comput. Biol. Med. 154, 106626. 2023
<http://dx.doi.org/10.1016/j.combiomed.2023.106626>
- [12] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 421–429, 2018.
- [13] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation, in: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, pp. 1055–1059, 2020
- [14] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: hybrid densely connected unet for liver and tumor segmentation from CT volumes, IEEE Trans.Med. Imaging. vol. 37, no. 12, pp. 2663–2674, 2018.
- [15] M.M. Naseer, K. Ranasinghe, S.H. Khan, M. Hayat, F. Shahbaz Khan, M.H. Yang, Intriguing properties of vision transformers, Adv. Neural Inf. Process. Syst. 34 23296–23308, 2021
- [16] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, L.C. Chen, Axial-deeplab: Standalone axial-attention for panoptic segmentation, in: European Conference on Computer Vision, Springer, pp. 108–126, 2020
- [17] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890, 2017.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, pp. 5998–6008, 2017.
- [19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
- [20] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou,Transunet: Transformers make strong encoders for medical image segmentation,2021,
<http://dx.doi.org/10.48550/arXiv.2102.04306>, arXiv preprint arXiv:2102.04306.
- [21] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swinunet: Unet-like pure transformer for medical image segmentation, arXiv preprint arXiv:2105.05537, 2021.
- [22] Y. Gao, M. Zhou, D.N. Metaxas, Utnet: A hybrid transformer architecture for medical image



محمد رجب قانع مدرک کارشناسی خود را در سال ۱۴۰۰ و کارشناسی ارشد خود را در سال ۱۴۰۲ در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه شهید باهنر کرمان دریافت کرده است. زمینه پژوهشی وی پردازش تصویر و بینایی ماشین می‌باشد.



عباس بحرالعلوم مدرک کارشناسی خود را در رشته مهندسی کامپیوتر از دانشگاه اصفهان و مدرک کارشناسی ارشد خود را در سال ۱۳۷۸ در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه شیراز اخذ کرد. ایشان مدرک دکترای خود را در سال ۱۳۹۵ در رشته مخابرات از دانشگاه باهنر کرمان دریافت نمود.

زمینه پژوهشی ایشان پردازش تصویر، بینایی ماشین و رایانش نرم است.



مهدی افتخاری مدرک کارشناسی خود را در سال ۱۳۷۹ در رشته مهندسی کامپیوتر گرایش سخت افزار از دانشگاه شیراز و مدرک کارشناسی ارشد و دکترای خود را به ترتیب در سال‌های ۱۳۸۲ و ۱۳۸۶ در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از همان دانشگاه اخذ کرد. وی از سال ۱۳۸۶ تاکنون عضو هیأت علمی بخش

مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان است و در سال ۱۴۰۰ به مرتبه استادی ارتقاء پیدا کرده است. حوزه‌های تخصصی پژوهش ایشان یادگیری ماشین، یادگیری عمیق و مجموعه‌ها و سیستم‌های فازی است. وی تاکنون بیش از ۱۴۰ مقاله علمی در نشریات و کنفرانس‌های معتبر داخلی و خارجی به چاپ رسانیده است.