

دانشگاه صنعتی خواجه نصیرالدین طوسی

پروژه درس یادگیری ماشین

علی صدیقی ارقون

استاد درس:

دکتر بشری پیشگو

بهمن ۱۴۰۴

چکیده

ناحیه‌بندی دقیق تصاویر سه‌بعدی مغز انسان برای تشخیص تومور، برنامه‌ریزی درمان و بسیاری از کاربردهای پزشکی مرتبط با سلامت انسان‌ها ضروری می‌باشد. طی سال‌های اخیر شبکه‌های عصبی عمیق به پیشرفته‌ترین روش‌های ناحیه‌بندی خودکار تبدیل شده‌اند. با این حال، نتایج مبتنی بر روش‌های خودکار ممکن است نیازمند اصلاح باشد تا به حدی دقیق شود که مورد استفاده بالینی قرار گیرد. ناحیه‌بندی تعاملی به کاربران اجازه می‌دهد تا با تعامل خود، نواحی مورد نظر خود را با دقت بیشتری استخراج کنند. اگر چه این عرصه توسط کارهای دیگری بررسی شده است، اما همچنان شکافی بین دنیای آکادمیک و جهان واقعی و صنعتی وجود دارد. روش‌های موجود، مناسب استفاده در سیستم‌های کم‌توان نمی‌باشند. همچنین این روش‌ها برای تولید خروجی مطلوب، نیازمند تلاش زیادی از سمت کاربر می‌باشند. روش استفاده شده در این پروژه هر دو مشکل را با به‌روزرسانی خروجی در نواحی محلی حل می‌کند. در این روش به جای ناحیه‌بندی تصویر به صورت کامل و یکجا، که نیازمند توان محاسباتی بالا می‌باشد، یک ناحیه‌بندی درشت دانه بر روی بُرش‌های کم وضوح از ناحیه مورد نظر کاربر صورت می‌گیرد. سپس یک اصلاح محلی صورت می‌گیرد تا وضوح از دست رفته ناحیه هدف بازیابی شود. همچنین برای جلوگیری از تولید مجدد خروجی برای تمامی نواحی تصویر، از یک تحلیل مورفولوژی استفاده می‌شود تا تنها نواحی نیازمند تغییر اصلاح شوند و جزئیات سایر نواحی با توجه به خروجی قبلی حفظ شود. در این پروژه، کارایی مناسب این روش تحت آزمایش‌های متعددی اثبات می‌شود و نتایج تجربی به دست آمده نشان دهنده دو نکته مثبت می‌باشد. نکته مثبت اول این است که، این روش در مقایسه با سایر روش‌ها، با وجود تعامل کمتر کاربر به نتایج دقیق‌تری دست می‌یابد. همچنین در مقایسه با روش‌های دیگر، این نتایج در زمان کمتری تولید می‌شود و نیازمند توان محاسباتی کمتری می‌باشد. نکته مثبت دوم تعمیم قابل قبول این روش در وظایفی است که در مرحله آموزش با آنها مواجه نشده است.

واژگان کلیدی: یادگیری عمیق، بینایی کامپیوتر، ناحیه‌بندی تعاملی تصویر، ناحیه‌بندی تومور مغزی، تصاویر سه‌بعدی پزشکی

فهرست مطالب

۱-۰	روش پیشنهادی	۱
۲-۰	مشکلات روش‌های قبلی	۲
۲-۰-۱	کارایی نامناسب بر روی دستگاه‌های با توان محاسباتی کم	۲
۲-۰-۲	عدم استفاده از ماسک از پیش موجود	۲
۳-۰	ویژگی‌های الگوریتم ارائه شده	۳
۴-۰	الگوریتم در یک نگاه	۴
۵-۰	جزئیات الگوریتم	۵
۵-۰-۱	بخش Target Crop	۶
۵-۰-۲	بخش Coarse Segmentation	۶
۵-۰-۳	بخش Focus Crop	۸
۵-۰-۴	بخش Local Refinement	۹
۵-۰-۵	بخش Progressive Merge	۱۰
۶-۰	تابع خطا	۱۱
۷-۰	ترکیب کردن مدالیته‌های مختلف MRI	۱۲
۸-۰	آماده‌سازی مجموعه داده BraTS	۱۴
۹-۰	توزیع دیتاست به مجموعه‌های آموزش، اعتبارسنجی و تست	۱۴
۱۰-۰	محیط و هایپرپارامترهای عمومی آموزش	۱۵
۱۱-۰	نحوه تعریف و اجرای آزمایش‌ها	۱۵
۱۲-۰	نکات کلی درباره نتایج آزمایش‌ها	۱۷
۱۳-۰	آزمایش ۱: HRNet18s-FLAIR-WT-Pretrained	۱۷

۲۰	۱۴-۰ آزمایش ۲: HRNet32-FLAIR-WT-Pretrained
۲۲	۱۵-۰ آزمایش ۳: SegFormerB3-FLAIR-WT-Pretrained
۲۴	۱۶-۰ آزمایش ۴: SegFormerB3-Mix-WT و ترکیب مدالیت‌ها
۲۶	۱۷-۰ آزمایش ۵: ارزیابی با تسک جدید دیده نشده در آموزش
۲۸	۱۸-۰ آزمایش ۶: ارزیابی با مدالیت و تسک جدید دیده نشده در آموزش
۳۱	۱۹-۰ مقایسه با نتایج کارهای مرتبط
۳۱	۱۹-۱ مقاله Enhanced U-Net
۳۲	۱۹-۲ مقاله DeepIGeoS
۳۳	۱۹-۳ مقاله MIDeepSeg
۳۴	۱۹-۴ روش‌های سنتی و غیر عمیق
۳۶	۲۰-۰ جمع‌بندی
۳۶	۲۱-۰ پیشنهادها و کارهای آینده
۳۶	۲۱-۱ مناسب‌سازی FocalClick برای ورودی سه‌بعدی
۳۷	۲۱-۲ استفاده از BRS در FocalClick
۳۷	۲۱-۳ استفاده از تبدیل فاصله ژئودزیک نمایی
۳۸	۲۱-۴ پیشنهادهای دیگر

۱-۰ روش پیشنهادی

در این قسمت، بر روی روش ناحیه‌بندی تعاملی مبتنی بر کلیک تمرکز می‌شود. تعامل مبتنی بر کلیک در واقع همان تعامل مبتنی بر نقاط Seed می‌باشد که در اینجا با نام کلیک به آن اشاره می‌شود. تحت این روش، کاربر به طور متوالی کلیک‌های مثبت مانند نقاط قرمز در تصویر ۱ را برای تعریف FG و کلیک‌های منفی مانند نقاط سبز در تصویر ۱ را برای تعریف BG قرار می‌دهد و مدل پس از هر کلیک کاربر، پیش‌بینی‌های جدید را خروجی می‌دهد.



شکل ۱: انواع کلیک [۲]

الگو اصلی برای ناحیه‌بندی تعاملی مبتنی بر کلیک، استفاده از نقشه‌ها یا دیسک‌های گاوسی^۱ برای نمایش کلیک‌های کاربر می‌باشد [۱۷]. در این روش یک نقشه گاوسی به مرکزیت کلیک کاربر و شعاع مشخص ایجاد می‌شود. در ادامه نقشه‌های کلیک با تصویر ورودی الحاق می‌شود، و وارد یک مدل ناحیه‌بند می‌شود تا ماسک اولیه پیش‌بینی شود.

بر اساس این روش، کارهای قبلی که مبتنی بر پارادایم‌های دیگر بود از جنبه‌های مختلفی بهبود می‌یابند. با این حال، هنگام اعمال این روش‌ها در سناریوهای عملی و صنعتی، نتیجه در جنبه‌هایی رضایت بخش نیست که در ادامه به این جنبه‌ها اشاره می‌شود.

¹Gaussian Maps

۲-۰ مشکلات روش‌های قبلی

۱-۲-۰ کارایی نامناسب بر روی دستگاه‌های با توان محاسباتی کم

انتظار می‌رود که یک ابزار ناحیه‌بند خوب، ماسک‌های خوبی را در زمان کوتاهی تولید کند. بیشتر کارهای قبلی فقط بر دقت تمرکز دارند و از مدل‌های بزرگ با ورودی‌های با وضوح بالا^۲ استفاده می‌کنند. با این حال، وقتی روی لپ‌تاپ‌های شخصی، دستگاه‌های لبه^۳ یا برنامه‌های مبتنی بر وب با درخواست‌های حجیم مستقر می‌شوند، با مشکل مواجه می‌شوند. همچنین هنگامی که سعی می‌شود اندازه ورودی برای عملکرد سریع‌تر کاهش داده شود، دقت این مدل‌ها به طور قابل توجهی کاهش می‌یابد.

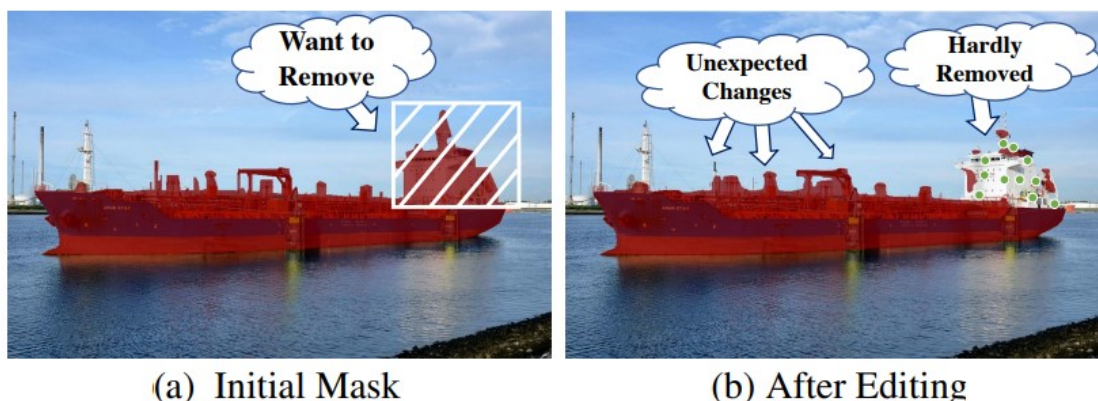
۲-۲-۰ عدم استفاده از ماسک از پیش موجود

در کاربردهای عملی و در دنیای واقعی، ممکن است تصویری که می‌خواهیم روی آن ناحیه‌بندی صورت گیرد، دارای یک ماسک اولیه ناحیه‌بندی شده باشد. این ماسک‌های موجود می‌تواند توسط یک مدل آفلاین قبلی یا سایر روش‌های پیش‌پردازش ارائه شده باشد. همچنین ماسک قبلی تولید شده توسط خود مدل را نیز می‌توان در نظر گرفت. استفاده از این ماسک‌های موجود و ایجاد تغییرات در آنها می‌تواند عمل ناحیه‌بندی را تسهیل کند. با این وجود، روش‌های قبلی موجود با ماسک‌های خارجی و از پیش موجود سازگار نمی‌باشند. بسیاری از این روش‌ها حتی وجود ماسک قبلی را به عنوان ورودی در نظر نمی‌گیرند. البته در بعضی کارها ماسک قبلی با تصویر ورودی و نقشه‌های به دست آمده از کلیک‌های کاربر به هم ملحق می‌شوند و به عنوان ورودی به مدل داده می‌شوند، اما این روش‌ها نیز دارای عملکرد خوبی در استفاده از ماسک از پیش موجود نیستند.

عدم استفاده از ماسک از پیش موجود باعث ایجاد مشکلات زیادی می‌شود. برای مثال در تصویر^۲ قصد کاربر حذف کردن کابین کشتی از ماسک قبلی می‌باشد، بنابراین تعدادی کلیک منفی (سبز) بر روی کابین کشتی ایجاد می‌کند تا این قسمت حذف شود. اما پس از اجرای دوباره مدل متوجه می‌شویم مدل علاوه بر حذف کابین، قسمت‌های دیگری که از کلیک‌های کاربر فاصله زیادی دارند و در خروجی قبلی دارای برجسب درستی بوده‌اند را هم حذف کرده است. یا حتی قسمت‌های جدیدی را به اشتباه به ماسک خروجی اضافه کرده است.

²High-Resolution

³Edge Devices



شکل ۲: تغییرات ناخواسته هنگام تولید خروجی جدید [۲]

در واقع هنگامی که کاربر بر روی یک ناحیه کلیک می‌کند، قصد دارد که مدل به آن ناحیه توجه کند. برای مثال در تصویر ۲ کاربر قصد دارد مدل به کابین کشتی توجه کند نه به محیط اطراف کشتی. اما بسیاری از روش‌های قبلی به قصد کاربر توجهی ندارند و در هنگام تولید خروجی جدید با تمامی پیکسل‌های تصویر به صورت برابر رفتار می‌کنند. این امر باعث ایجاد دو پدیده می‌شود. اول اینکه نواحی‌ای که قبلاً به صورت خوبی ناحیه‌بندی شده بودند نیازمند ناحیه‌بندی دوباره هستند که این پدیده افزونگی محاسبات^۴ نامیده می‌شود. دوم اینکه تمامی پیکسل‌های ماسک قبلی دوباره باید ناحیه‌بندی شوند و مدل اطلاعات قبلی موجود از آنها را از دست می‌دهد که این پدیده از دست دادگی اطلاعات^۵ نامیده می‌شود.

۳-۰. ویژگی‌های الگوریتم ارائه شده

ویژگی‌های الگوریتم ارائه شده بایستی در راستا و متناظر با حل مشکلات اشاره شده باشد. توجه و تمرکز این الگوریتم باید بر روی نواحی محلی^۶ خاصی باشد. همچنین پیش‌بینی جدید تنها برای ناحیه‌هایی که نیازمند محاسبات دوباره هستند صورت گیرد و الگوریتم تنها خروجی ناحیه‌هایی را به‌روزرسانی کند که کاربر هم قصد به‌روزرسانی آن را داشته است. علاوه بر این موارد الگوریتم باید در نواحی دیگر خروجی قبلی را حفظ کند. در این پروژه با توجه به نکات، مشکلات و اهداف اشاره شده، از مدل و الگوریتم FocalClick [۲] برای

^۴Computational Redundancy

^۵Loss of Information

^۶Local

ناحیه‌بندی تعاملی تومور مغزی استفاده خواهیم کرد. در ادامه بخش‌های مختلف این الگوریتم طبق مقاله FocalClick [۲] توضیح داده می‌شود.

۴-۰ الگوریتم در یک نگاه

هنگامی که کاربر کلیک جدیدی انجام می‌دهد، ابتدا یک برش از تصویر اصلی با نام Target Crop ایجاد می‌شود. این برش با توجه به ماسک از پیش موجود ساخته می‌شود. در شکل ۳ این برش با یک Bounding Box زرد رنگ نمایش داده شده است.

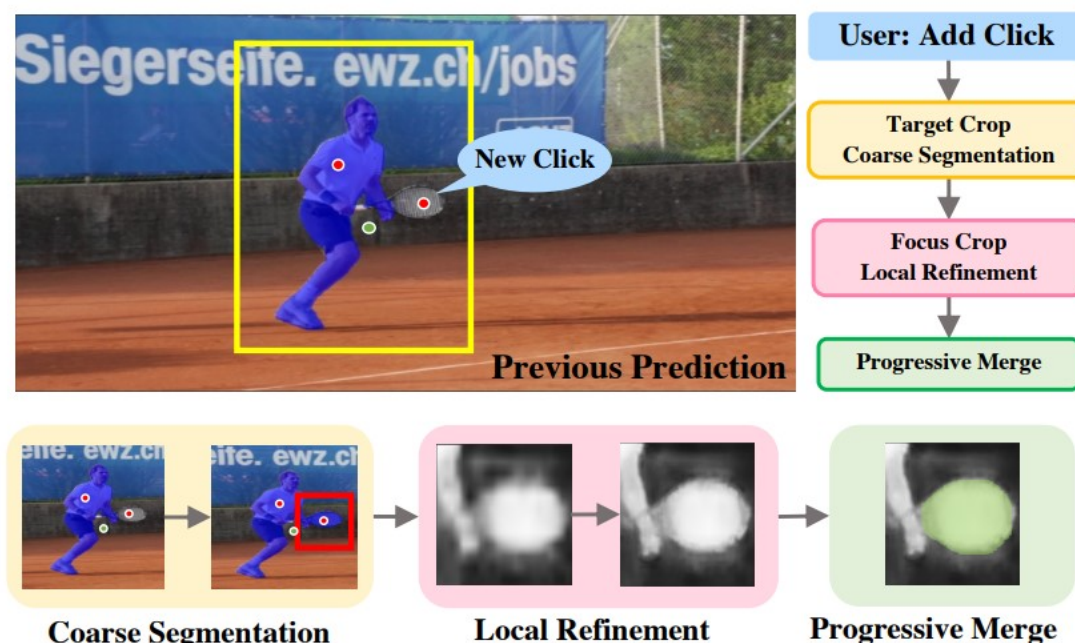
سپس این Target Crop دچار تغییر اندازه^۷ می‌شود. همچنین از وضوح آن نیز کاسته می‌شود. بر روی این تصویر کوچک و با وضوح کم یک ناحیه‌بندی درشت دانه صورت می‌گیرد که این عمل تحت نام Coarse Segmentaion شناخته می‌شود.

سپس یک ناحیه کوچک دیگر که کاربر قصد تغییر خروجی ناحیه‌بندی در آن جا را داشته، شناسایی می‌شود. با توجه به اینکه تمرکز مدل باید بر روی این ناحیه باشد، آن را Focus Crop می‌نامیم که در شکل ۳ با یک محیط قرمز رنگ مشخص شده است. سپس این ناحیه مورد توجه وارد قسمت پالایش یا Local Refinement می‌شود تا کیفیت و وضوح از دست رفته خود را بازیابی کند. بنابراین عملیات سنگین و زمان‌بر ناحیه‌بندی یک تصویر بزرگ، به دو عملیات محلی و با هزینه کم تبدیل می‌شود.

در نهایت ناحیه پالایش شده وارد قسمت ادغام پیش‌رونده یا Progressive Merge می‌شود تا با یک تحلیل مورفولوژیکی^۸ بر روی ماسک از پیش موجود و خروجی جدید، تصمیم بر به‌روزرسانی یا حفظ برچسب‌های قبلی گرفته شود. بنابراین نواحی‌ای که قبلاً درست پیش‌بینی شده بودند دچار تغییر نمی‌شوند.

^۷Resize

^۸Morphological



شکل ۳: نگاهی کلی به الگوریتم FocalClick [۲]

با توجه به توضیحات داده شده می‌توان متوجه شد که روش بالا بسیار متکی به عمل برش یا Crop است که این عمل در بعضی کارهای دیگر ناحیه‌بندی نیز دیده می‌شود. اما تفاوت این الگوریتم با سایر روش‌ها در این است که کارهای دیگر نیازمند حفظ وضوح تصویر برش یافته هستند و ناحیه‌بند آنها یک خروجی با جزئیات دقیق تولید می‌کند که این کار زمان‌بر است و با سرعت کمی صورت می‌گیرد. اما در این الگوریتم وضوح ناحیه برش یافته کاهش یافته و خروجی ناحیه‌بند سریع و درشت دانه است.

۵-۰ جزئیات الگوریتم

در این قسمت، بخش‌ها و مراحل الگوریتم با ذکر جزئیات مورد بررسی قرار می‌گیرد. ابتدا مدل تصویر ورودی، دو نقشه مربوط به کلیک‌های منفی و مثبت و ماسک از پیش موجود را به عنوان ورودی دریافت می‌کند. برای ایجاد نقشه مربوط به کلیک‌های کاربر، از یک دیسک دودویی^۹ با شعاع ۲ استفاده می‌شود. به عبارت دیگر ناحیه درون دایره‌ای با شعاع ۲ و مرکز کلیک برچسب ۱ و بیرون این دایره برچسب ۰ می‌گیرد. سپس Target

^۹Binary Disk

Crop پیرامون شی هدف ایجاد می شود و اندازه و وضوح آن کاهش می یابد. سپس این برش به شبکه ناحیه بند داده می شود تا خروجی درشت دانه (Coarse) ایجاد شود. ناحیه Focus Crop با محاسبه روی تفاوت نواحی میان ماسک قبلی و ناحیه بندی درشت دانه ایجاد می شود. از این برش برای پالایش و بازیابی جزئیات تصویر درشت دانه استفاده می شود. در نهایت بخش Progressive Merge ماسک قبلی و ناحیه پالایش شده را ادغام می کند. در ادامه جزئیات هر مرحله از این الگوریتم توضیح داده می شود.

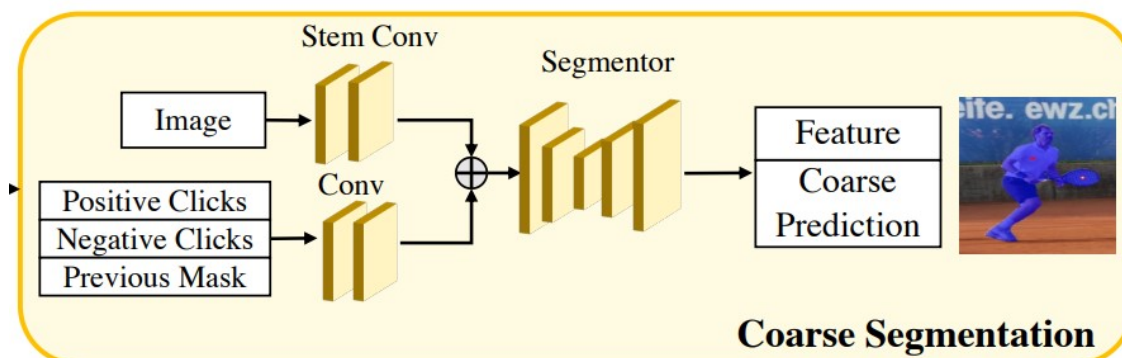
۱-۵-۰ بخش Target Crop

هدف این قسمت فیلتر کردن ناحیه BG از ناحیه مربوط به شی هدف است. برای ایجاد این ناحیه یک Bounding Box خارجی کمینه که شامل ماسک (شی هدف) از پیش موجود و کلیک جدید کاربر می باشد ایجاد می شود. در ادامه این Bounding Box با هایپرپارامتر ^{۱۰} ratio_TC بسط می یابد. مقدار این هایپرپارامتر برابر ۱/۴ در نظر گرفته شده است. همانطور که در تصویر ۱ مشاهده می شود مستطیل زرد رنگ شامل کلیک جدید کاربر و شی هدف می باشد و ابعاد آن کمی بسط یافته است. پس از گسترش این Bounding Box و به دست آمدن مختصات آن، سه تنسور مربوط به تصویر، ماسک قبلی و نقشه های کلیک کاربر با توجه به این Bounding Box برش می خورند. در نهایت وضوح و اندازه این برش های به دست آمده کم می شود تا آماده قسمت بعدی شود.

۲-۵-۰ بخش Coarse Segmentation

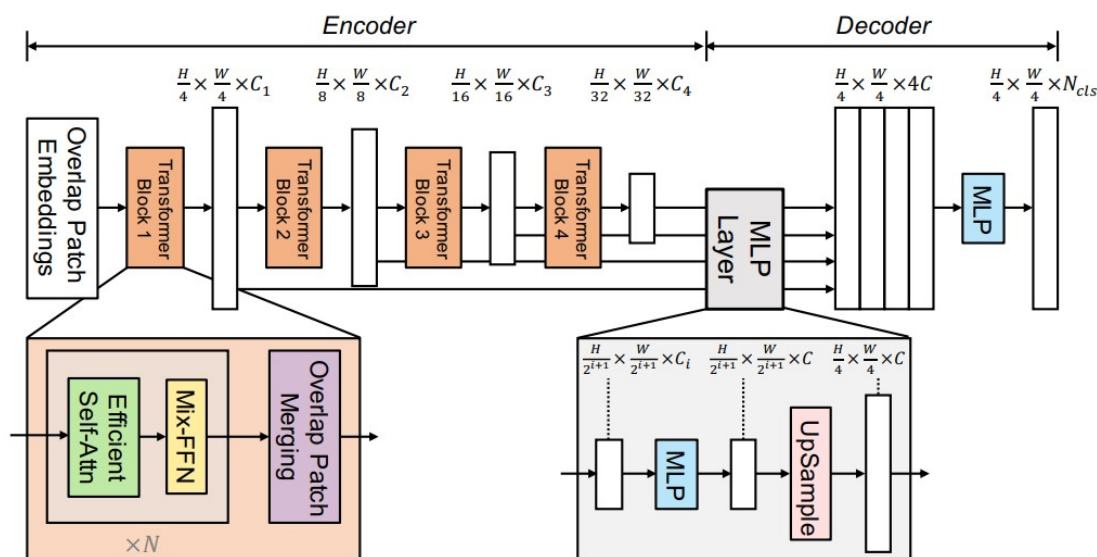
در این قسمت ورودی های دودویی مربوط به نقشه کلیک های منفی، کلیک های مثبت و ماسک از پیش موجود وارد دو لایه کانولوشنی می شوند تا در بعد کانال مشابه تصویر ورودی شوند. تصویر ورودی هم از دو لایه کانولوشنی عبور می کند تا ویژگی های سطح پایین آن به دست آید. سپس ویژگی های به دست آمده از تصویر و سه ورودی دودویی با هم ترکیب می شوند.

¹⁰Hyperparameter

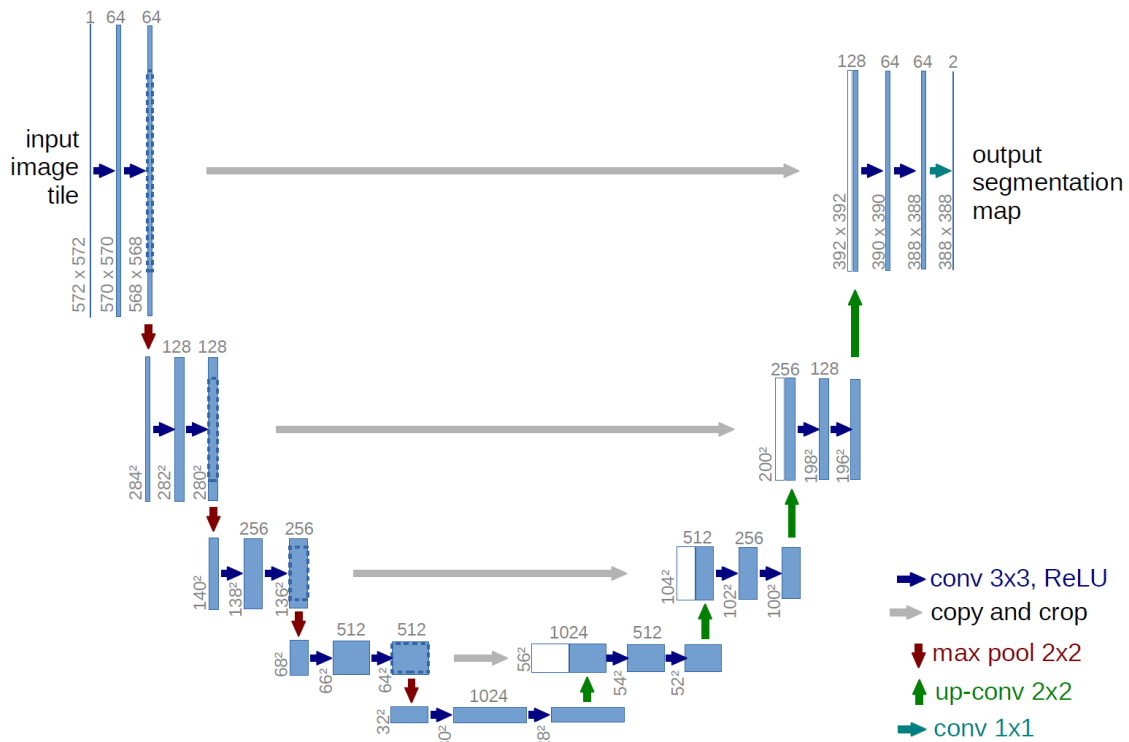


شکل ۴: معماری قسمت Coarse Segmentation [۲]

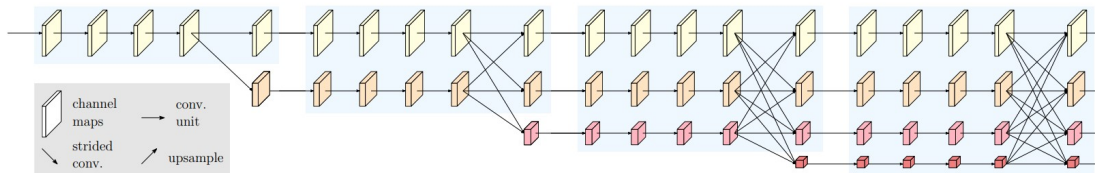
این ویژگی‌های ترکیب شده وارد یک شبکه ناحیه‌بند می‌شوند. این شبکه می‌تواند هر شبکه ناحیه‌بندی مانند U-Net [۱۱]، HRNet [۱۵] (مناسب تصاویر با وضوح بالا) و یا شبکه‌های ناحیه‌بند مبتنی بر Transformer مانند SegFormer [۱۶] باشد. ویژگی مشترک این شبکه‌ها وجود معماری Encoder-Decoder و همچنین اتصالات پرشی برای حفظ اطلاعات تصویر ورودی می‌باشد. در نهایت خروجی ناحیه‌بند Coarse تولید می‌شود.



شکل ۵: معماری شبکه SegFormer [۱۶]



شکل ۶: معماری شبکه U-Net [۱۱]

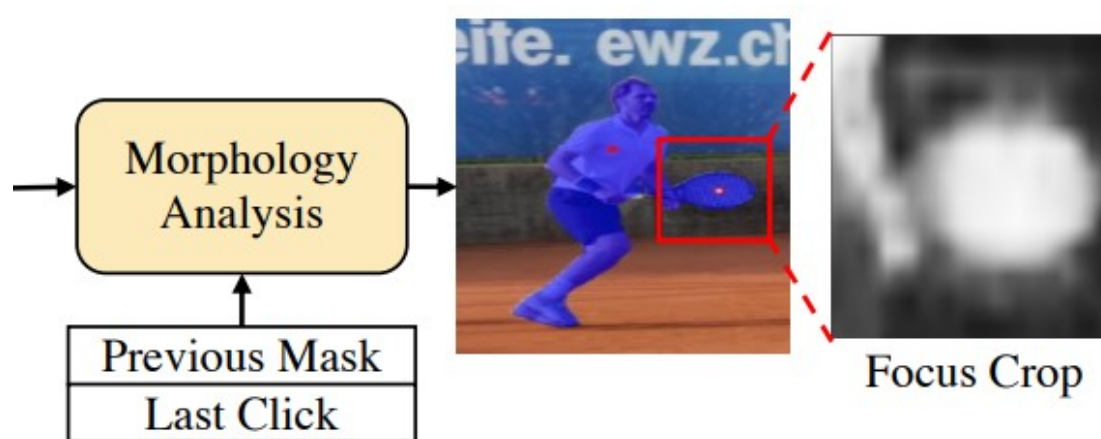


شکل ۷: معماری شبکه HRNet [۱۵]

۳-۵-۰ بخش Focus Crop

این قسمت ابتدا تفاوت ماسک از پیش موجود و خروجی قسمت Coarse Segmentation را محاسبه می‌کند. این مقایسه به وسیله تابع منطقی XOR صورت می‌گیرد و نتیجه آن یک ماسک جدید با نام Difference Mask (DM) می‌باشد. سپس بیشینه ناحیه متصل درون این ماسک که شامل کلیک جدید کاربر می‌باشد محاسبه می‌شود. سپس یک Bounding Box دور بیشینه ناحیه متصل ایجاد می‌شود. مشابه قسمت Target Crop، این Bounding Box به دست آمده با هایپرپارامتر ratio_FC بسط می‌یابد. مقدار این هایپرپارامتر را هم برابر ۱/۴

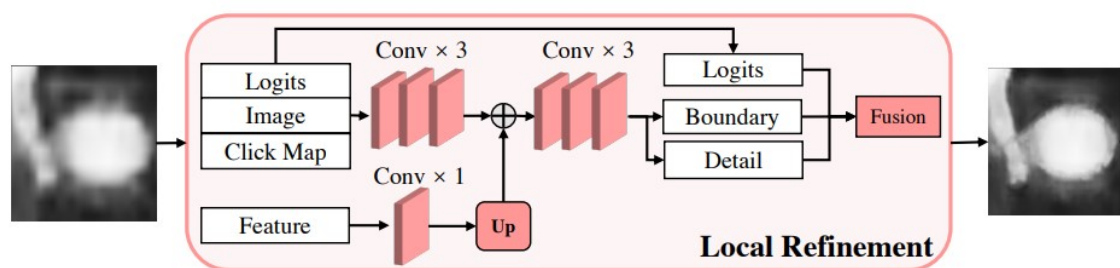
در نظر می‌گیریم. در نهایت Focus Crop که در تصویر ۳ با رنگ قرمز نمایش داده شده است به دست می‌آید. مطابق Focus Crop به دست آمده، تصویر ورودی و نقشه‌های کلیک کاربر نیز Crop می‌شوند. همچنین از ماژول RoiAlign [۵] برای ایجاد برش روی Featureها و Logitهای خروجی ناحیه‌بند استفاده شده است.



شکل ۸: معماری قسمت Focus Crop [۲]

۴-۵-۰ بخش Local Refinement

این قسمت، جزئیات پیش‌بینی درشت دانه صورت گرفته بر روی Focus Crop را بازبینی و پالایش می‌کند. ابتدا ویژگی سطح پایین با استفاده از لایه‌های کانولوشنی Xception [۳] از تانسورهای برش یافته استخراج می‌شود. در همین حال، تعداد کانال‌های مربوط به ویژگی‌های RoiAlign تنظیم می‌شود و سپس با ویژگی‌های سطح پایین به دست آمده ادغام می‌شود.



شکل ۹: معماری قسمت Local Refinement [۲]

برای به دست آوردن پیش‌بینی‌های پالایش شده، از دو Head برای پیش‌بینی نقشه جزئیات M_d و نقشه مرزی M_b استفاده می‌کنیم. سپس نقشه پیش‌بینی پالایش شده M_r با استفاده از رابطه زیر به دست می‌آید. در این رابطه M_l بیانگر Logit‌های پیش‌بینی شده توسط شبکه Coarse Segmentor می‌باشد.

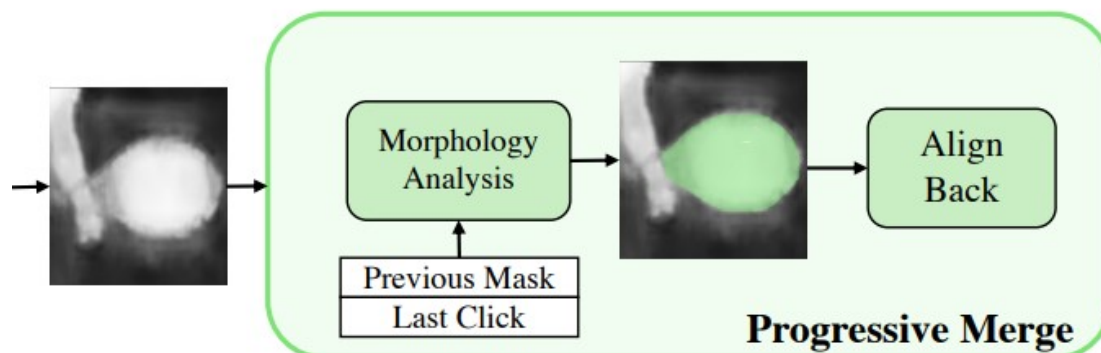
$$M_r = \text{Sigmoid}(M_b) * M_d + (1 - \text{Sigmoid}(M_b)) * M_l \quad (۱)$$

۵-۵-۰ بخش Progressive Merge

هنگام برچسب‌گذاری یا ویرایش ماسک‌ها، انتظار نداریم که مدل پس از هر کلیک، ماسک را برای همه پیکسل‌ها به‌روز کند. در غیر این صورت تمامی جزئیاتی که الگوریتم به خوبی برچسب زده دوباره به‌روزرسانی می‌شود و حتی ممکن است برچسب صحیح قبلی از بین برود. کاربر فقط می‌خواهد در مناطق محدودی که قصد اصلاح آن‌ها را دارد به‌روزرسانی صورت گیرد. در قسمت Progressive Merge نیز مشابه قسمت‌های Target Crop و Focus Crop از آنالیز مبتنی بر مورفولوژی استفاده می‌شود تا ناحیه‌ای که نیازمند اصلاح است به دست آید.

پس از وارد شدن یک کلیک کاربر، به سادگی ماسک پیش‌بینی شده جدید با آستانه ^{۱۱} برابر ۰/۵ باینری می‌شود و ناحیه متفاوتی بین پیش‌بینی جدید و ماسک موجود محاسبه می‌شود. سپس بیشینه ناحیه متصل که حاوی کلیک جدید است به عنوان منطقه به‌روز رسانی انتخاب می‌شود (ناحیه سبز رنگ در تصویر ۱۰). در این ناحیه، ماسک جدید پیش‌بینی شده روی ماسک قبلی قرار داده می‌شود و به‌روز می‌شود. همچنین ماسک قبلی در نواحی دیگر دست نخورده باقی می‌ماند.

¹¹Threshold



شکل ۱۰: معماری قسمت Progressive Merge [۲]

هنگامی که الگوریتم با یک ماسک از قبل موجود اجرا می‌شود یا یک ماسک اولیه از سایر ابزارهای ناحیه‌بندی وجود دارد، قسمت Progressive Merge برای حفظ جزئیات صحیح اعمال می‌شود. اما هنگام برچسب‌گذاری اولیه و بدون وجود ماسک قبلی از ابتدا، قسمت Progressive Merge پس از ۱۰ کلیک فعال می‌شود.

۶-۰ تابع خطا

با توجه به این که این الگوریتم دارای قسمت‌های مختلف می‌باشد، برای هر کدام از این بخش‌ها یک تابع خطا در نظر گرفته می‌شود و با ترکیب این توابع، تابع نهایی خطا به دست می‌آید و الگوریتم طبق نظارت^{۱۲} این تابع آموزش داده می‌شود. تابع خطای نقشه مرزی M_b از ۸ مرتبه Down Sample کردن ناحیه‌بندی مربوط به Ground Truth و سپس Up Sample کردن آن به اندازه قبلی به دست می‌آید. در این حالت پیکسل‌های تغییر یافته می‌توانند نمایانگر ناحیه‌هایی باشند که نیازمند دقت و جزئیات بیشتری هستند. از تابع خطای Binary Cross Entropy (BCE) برای نظارت روی Boundary Head استفاده می‌شود. همچنین برای نظارت بر روی خروجی ناحیه‌بند Coarse از تابع خطا Normalized Focal Loss (NFL) [۱۲] استفاده می‌شود. برای نظارت بر روی قسمت Local Refinement نیز از تابع خطا NFL وزن‌دار استفاده می‌شود. این وزن برابر $b = 1.5$ در نظر گرفته می‌شود و تابع جدید وزن‌دار به اختصار BNFL نامیده می‌شود. در نهایت تابع خطا نهایی کل شبکه از جمع این ۳ تابع خطا به دست می‌آید.

¹²Supervision

$$L = L_{BCE} + L_{NFL} + L_{BNFL} \quad (۲)$$

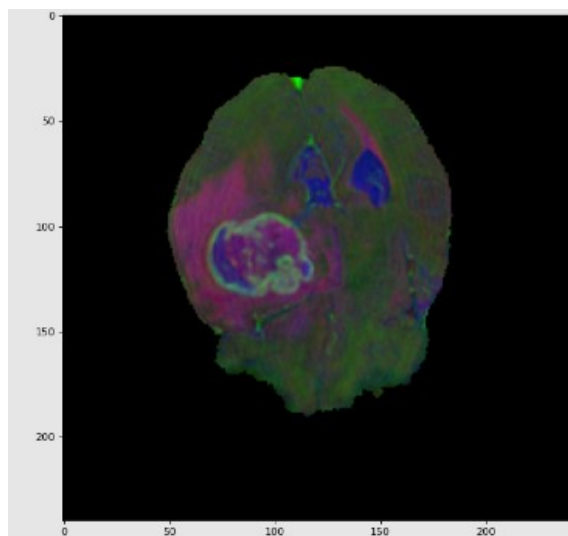
۷-۰ ترکیب کردن مدالیت‌های مختلف MRI

همانطور که در فصل دوم اشاره شد، هر نمونه تصویر MRI مجموعه داده BraTS در چهار مدالیت FLAIR، T1، T2 و T1ce موجود است. همچنین هر کدام از این مدالیت‌ها بخش‌های خاصی از مغز یا تومور مغزی را بهتر برجسته می‌کنند. این مدالیت‌ها در قالب نقشه‌های Grayscale وجود دارند و هر پیکسل آنها دارای شدتی بین ۰ تا ۱ می‌باشد. در کارهای دیگر که بر روی ناحیه‌بندی تصاویر MRI مغزی تمرکز دارند معمولاً از یکی از این مدالیت‌ها استفاده می‌شود و مدالیت استفاده شده اغلب از نوع FLAIR یا T1ce می‌باشد [۸، ۱۴].

یکی از نوآوری‌های اثربخش به کار گرفته شده در این پروژه استفاده همزمان از سه مدالیت FLAIR، T1ce و T2 می‌باشد. برای انجام این کار، مشابه تصاویر معمولی RGB که از ادغام سه کانال قرمز، سبز و آبی شکل گرفته اند عمل شده است. یعنی سه مدالیت FLAIR، T1ce و T2 با هم Concatenate شده و یک تصویر سه کاناله ایجاد شده است. برای اشاره راحت‌تر به این حالت آن را حالت Mix می‌نامیم.

هر چند تصویر تولید شده از ادغام این سه مدالیت که نمونه‌ای از آن در تصویر ۱۱ آورده شده است برای انسان دارای معنی و اطلاعات خاصی نیست و حتی باعث اشتباه می‌شود، اما مدل مبتنی بر یادگیری عمیق می‌تواند از هر کانال، ویژگی‌های مورد نیاز خود را استخراج کند تا خروجی بهتری تولید کند. همچنین مزیت دیگر این کار این است که تصویر ورودی به مدل مشابه تصاویر RGB می‌باشد که در این صورت می‌توان از مدل‌های از پیش آموخته^{۱۳} شده که بر روی مجموعه داده‌های بزرگی مثل ImageNet [۴] آموزش دیده شده اند راحت‌تر استفاده کرد. پیش از این برای استفاده از مدل‌های از پیش آموخته بایستی مدالیت انتخاب شده سه بار تکرار می‌شد که نه تنها برای مدل اطلاعات جدیدی نداشت، بلکه تعداد پارامترهای مدل را هم بسیار زیادتر می‌کرد.

¹³Pretrained



شکل ۱۱: نمایش تصویر ایجاد شده در حالت ترکیب مدالیت‌ها

۸-۰ آماده‌سازی مجموعه داده BraTS

الگوریتم FocalClick مناسب تصاویر رنگی طبیعی با وضوح بالا می‌باشد و ورودی قسمت‌های Refiner و Segmentor آن حتی پس از انجام عمل Crop به صورت 256×256 می‌باشد. همچنین این الگوریتم مناسب تصاویر دوبعدی RGB می‌باشد. در حالی که مجموعه داده BraTS شامل تصاویر سه‌بعدی در چهار مدالیته با اندازه $240 \times 240 \times 155$ می‌باشد. بنابراین منطبق کردن این الگوریتم و این مجموعه داده نیازمند تغییراتی در هر دو می‌باشد. ابتدا اندازه ورودی قسمت‌های Refiner و Segmentor را از 256×256 به 96×96 تبدیل می‌کنیم تا مناسب اندازه تصاویر BraTS شود. همچنین با توجه به این که پردازش تصاویر و آموزش آن‌ها به صورت سه‌بعدی نیازمند قدرت محاسباتی بسیار قوی و حافظه زیادی می‌باشد، و همچنین عدم وجود ماژول‌های سه‌بعدی نظیر RoiAlign در ابزارهای یادگیری عمیق مثل PyTorch و TensorFlow، مجموعه داده BraTS را به صورت دوبعدی وارد شبکه می‌کنیم. برای این کار هر اسلایس مربوط به Depth این مجموعه داده را یک نمونه جدید در نظر می‌گیریم. از مدالیته‌های FLAIR، T1ce و حالت ابداعی Mix نیز در آزمایش‌ها استفاده شده است.

۹-۰ توزیع دیتاست به مجموعه‌های آموزش، اعتبارسنجی و تست

با توجه به در دسترس عموم نبودن مجموعه تست BraTS، از مجموعه داده آموزشی BraTS سال ۲۰۲۰ برای آموزش، اعتبارسنجی و تست الگوریتم استفاده شده است. در این نسخه از مجموعه داده ۳۶۹ نمونه سه‌بعدی در چهار مدالیته وجود دارد. از این ۳۶۹ نمونه، ۷۵ درصد برای آموزش مدل، ۲۰ درصد برای اعتبارسنجی مدل و ۵ درصد برای تست مدل جدا شده است. در جدول ۱ خلاصه‌ای از تعداد نمونه‌ها در حالت دوبعدی و سه‌بعدی مربوط به مجموعه‌های آموزش، اعتبارسنجی و تست آورده شده است.

جدول ۱: توزیع مجموعه‌های آموزش، اعتبارسنجی و تست

	Percentage	3D Cases	2D Cases
Training Set	75 %	276	42780
Validation Set	20 %	73	11315
Testing Set	5 %	20	3100
Total	100 %	369	57195

۱۰-۰. محیط و هایپرپارامترهای عمومی آموزش

تعدادی از هایپرپارامترها و مشخصه‌ها در تمامی آزمایش‌های صورت گرفته ثابت می‌باشد که در این قسمت به این موارد اشاره می‌شود. تمامی مدل‌های برای ۲۰ اپیک ۱۴ آموزش داده شده اند. همچنین نمونه‌ها در Batchهایی با اندازه ۳۲ وارد مدل می‌شوند تا مدل قادر باشد عملیات را موازی‌سازی کند. استفاده از Batchهای بزرگتر باعث پر شدن حافظه و توقف الگوریتم می‌شود. همچنین Crop و کاهش وضوح به نحوی انجام می‌شود که نواحی Target Crop و Focus Crop دارای اندازه 96×96 شوند. هایپرپارامترهای ratio_FC و ratio_TC که برای بسط دادن ابعاد Bounding Box زرد و قرمز بود نیز برابر $1/4$ در نظر گرفته شده است. آزمایش‌ها در دو محیط Google Colab با پردازشگر گرافیکی Nvidia K80 / T4 و محیط Kaggle با پردازشگر گرافیکی Nvidia P100 انجام شده است. لینک این محیط‌ها، نتایج آزمایش، تصاویر خروجی مدل در هر مرحله از آموزش (Visualization)، وزن مدل‌ها و پیاده‌سازی الگوریتم در قسمت پیوست‌ها آورده شده است.

۱۱-۰. نحوه تعریف و اجرای آزمایش‌ها

برای تعریف و اجرای آزمایش‌ها، کافی است یکی از نوتبوک‌های Google Colab یا Kaggle اجرا شود. طراحی این نوتبوک‌ها به گونه‌ای می‌باشد که با تغییر متغیرهای موجود، می‌توان آزمایش‌های مختلفی انجام داد. برای مثال می‌توان از شبکه‌های ناحیه‌بند موجود یکی را به عنوان Backbone انتخاب کرد. همچنین امکان انتخاب

¹⁴Epoch

مدالیه‌های گوناگون هم در فاز آموزش و هم در فاز ارزیابی وجود دارد. تسک مورد نظر کاربر نیز قابل تعریف می‌باشد و کاربر می‌تواند ناحیه و ماسک مطلوب خود را تعریف کند. جزییات مربوط به نحوه تعریف آزمایش در تصویر ۱۲ و تصویر ۱۳ نمایش داده شده است. علاوه بر موارد بالا، امکان آموزش مدل بر روی چندین GPU به صورت همزمان نیز وجود دارد.

```
MODE = "FocalClick" # CDNet, FocalClick
BACKBONE = "segformerB3" # cdnet_res34, hrnet18s, hrnet32, segformerB3

# CDNet-ResNet34, FocalClick-HRNet18s, FocalClick-HRNet32, FocalClick-SegFormerB3
EXP_NAME = "FC-SF-Mix-WT"

CHANNEL = "mix" # flair, t1, t1ce, t2, mix
LABEL = "wt" # net, ed, et, wt, tc
EPOCHS = 20
BATCH_SIZE = 32
NGPUS = 1
WORKERS = 2

MODEL_FILE = f"{REPO_PATH}/models/brats/{BACKBONE}.py"
MODEL_DIR = f"{REPO_PATH}/experiments/brats/{BACKBONE}/000_{EXP_NAME}/checkpoints/"

# Set --pretrained in below cell
```

شکل ۱۲: نحوه تعریف متغیرهای فاز آموزش

```
MODE = "FocalClick" # CDNet, FocalClick
BACKBONE = "segformerB3" # cdnet_res34, hrnet18s, hrnet32, segformerB3
EXP_NAME = "000_FocalClick-SegFormerB3"
CHECKPOINT = "epoch-19-val-loss-1.28.pth"

CHANNEL = "t1ce" # flair, t1, t1ce, t2, mix
LABEL = "tc" # net, ed, et, wt, tc

MODEL_DIR = f"{REPO_PATH}/experiments/brats/{BACKBONE}/{EXP_NAME}/checkpoints/"
```

شکل ۱۳: نحوه تعریف متغیرهای فاز ارزیابی

۱۲-۰ نکات کلی درباره نتایج آزمایش‌ها

همواره معیار Dice در مقایسه با معیار IoU دارای مقدار بیشتری می‌باشد که این اتفاق به دلیل وجود ضریب ۲ در رابطه Dice می‌باشد. در تحلیل و تفسیر نتایج نیز تمرکز بیشتر بر روی معیار Dice می‌باشد و معیار IoU جهت مقایسه نتایج این الگوریتم با سایر مقالاتی که نتایج خود را فقط در قالب IoU ارائه کرده اند، می‌باشد. همچنین زیاد بودن معیار NoF به این دلیل است که بسیاری از اسلایس‌های تصویر سه‌بعدی تبدیل به دوبعدی شده فاقد پیکسل FG می‌باشند. در این مواقع، الگوریتم این نمونه‌ها را از نوع Failure شمرده است. در تصویر اول ارائه شده برای هر آزمایش، در سمت بالا و با شروع از سمت چپ هر قسمت به ترتیب بیانگر Full Image سپس Full Ground Truth سپس Full Prediction و در سمت راست Previous Full Prediction می‌باشد. در پایین تصویر و با شروع از سمت چپ نیز هر قسمت به ترتیب بیانگر Refine Image سپس Refine Focus Prediction و در سمت راست Trimap Prediction (ترکیب نقشه‌های کلیک منفی و مثبت) می‌باشد.

در تصویر دوم ارائه شده برای هر آزمایش، ناحیه و IoU به دست آمده به ازای ۲۰ کلیک نشان داده شده است. در این تصویر محل کلیک‌های مثبت و منفی نیز نمایش داده شده است. همچنین محل مربوط به Target Crop و Focus Crop با Bounding Box‌های زرد و قرمز مشخص شده است. علاوه بر موارد بالا تصویر Ground Truth و خروجی نهایی مدل و خروجی ماقبل نهایی مدل نیز آورده شده است. در تصویر سوم و آخر ارائه شده برای هر آزمایش نیز، ناحیه Focus Crop به دو صورت Coarse و Refined آورده شده است.

۱۳-۰ آزمایش ۱: HRNet18s-FLAIR-WT-Pretrained

در این آزمایش، الگوریتم FocalClick با استفاده از Backbone و ناحیه‌بند HRNet18s که از پیش بر روی مجموعه داده ImageNet آموزش داده شده است، اجرا می‌شود. همچنین از مدالیت FLAIR برای ایجاد مجموعه داده استفاده شده است. تسک و هدف مورد نظر آزمایش نیز پیدا کردن ناحیه مربوط به کل تومور یا Whole Tumor (WT) می‌باشد. پس از اینکه آموزش مدل کامل شد، ارزیابی مجموعه تست، بر روی همان مدالیت و همان تسک مرحله آموزش صورت می‌گیرد.

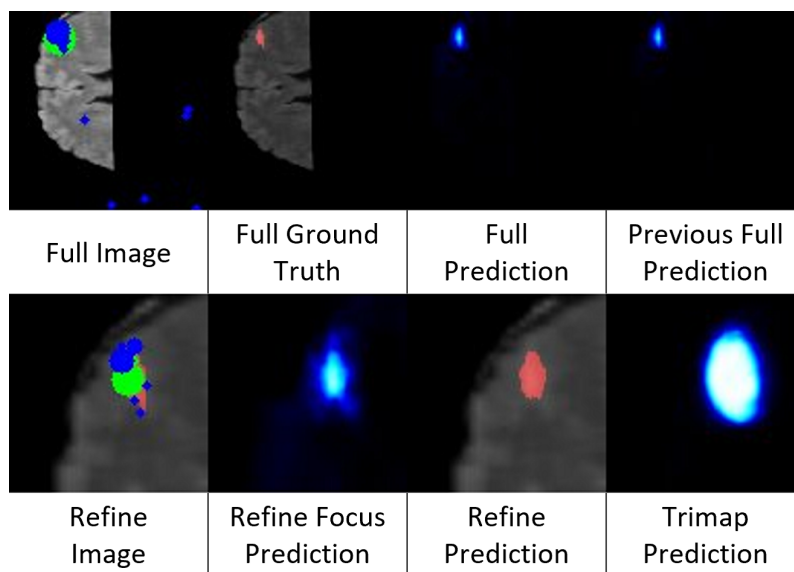
جدول ۲: معیار NoC و NoF در آزمایش ۱

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	3.47	4.31	6.00	325	568	0.061	0:56:38
Dice	2.50	2.92	3.72	180	259	0.061	0:56:38

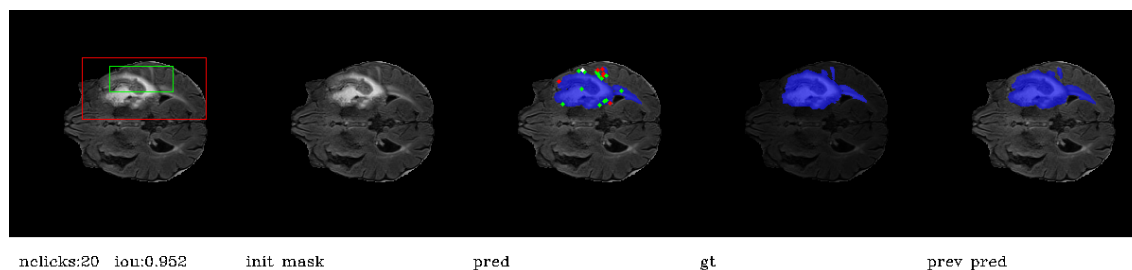
جدول ۳: معیار mIoU و mDice در آزمایش ۱

Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	84.27 %	85.77 %	86.59 %	88.08 %	89.79 %	90.94 %
mDice	88.39 %	89.67 %	90.28 %	91.36 %	92.63 %	93.55 %

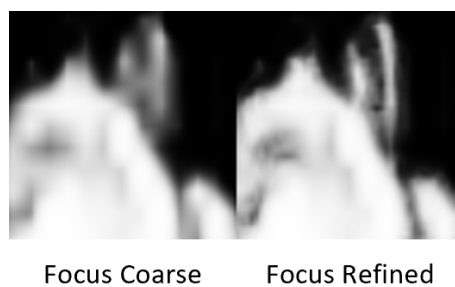
با توجه به اینکه ناحیه‌بند HRNet18s یک ناحیه‌بند سبک با تعداد پارامتر کم می‌باشد، الگوریتم پس از دریافت کلیک جدید توانسته در زمان خوبی خروجی جدید را تولید کند. همچنین برای رسیدن معیار Dice به مقدار ۹۰ درصد به صورت میانگین ۳/۷۲ کلیک مورد نیاز است که تا حد خوبی قابل قبول می‌باشد. الگوریتم با دریافت ۲۰ کلیک نیز توانسته به میانگین Dice برابر با ۹۳/۵۵ درصد برسد که باز هم نشان دهنده عملکرد خوب الگوریتم می‌باشد. همچنین هرچه تعداد کلیک بیشتر می‌شود عملکرد مدل نیز بهتر می‌شود، که نشان دهنده تاثیر کلیک‌های کاربر و استفاده مناسب مدل از این کلیک‌ها می‌باشد.



شکل ۱۴: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۱



شکل ۱۵: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۱



شکل ۱۶: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۱

۱۴-۰ آزمایش ۲: HRNet32-FLAIR-WT-Pretrained

در این آزمایش، الگوریتم FocalClick با استفاده از Backbone و ناحیه‌بند HRNet32 که از پیش بر روی مجموعه داده ImageNet آموزش داده شده است، اجرا می‌شود. همچنین از مدالیت FLAIR برای ایجاد مجموعه داده استفاده شده است. تسک و هدف مورد نظر آزمایش نیز پیدا کردن ناحیه مربوط به کل تومور یا Whole Tumor (WT) می‌باشد. پس از اینکه آموزش مدل کامل شد، ارزیابی مجموعه تست، بر روی همان مدالیت و همان تسک مرحله آموزش صورت می‌گیرد.

جدول ۴: معیار NoC و NoF در آزمایش ۲

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	4.23	5.11	6.53	515	697	0.103	1:35:37
Dice	3.02	3.53	4.52	326	443	0.103	1:35:37

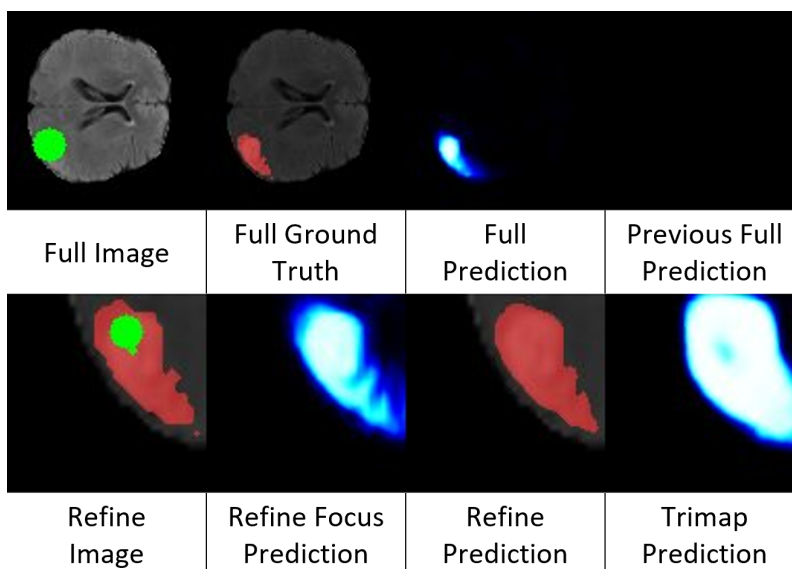
جدول ۵: معیار mIoU و mDice در آزمایش ۲

Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	85.58 %	79.77 %	79.02 %	79.39 %	80.66 %	81.69 %
mDice	89.03 %	82.69 %	81.84 %	81.65 %	82.46 %	83.34 %

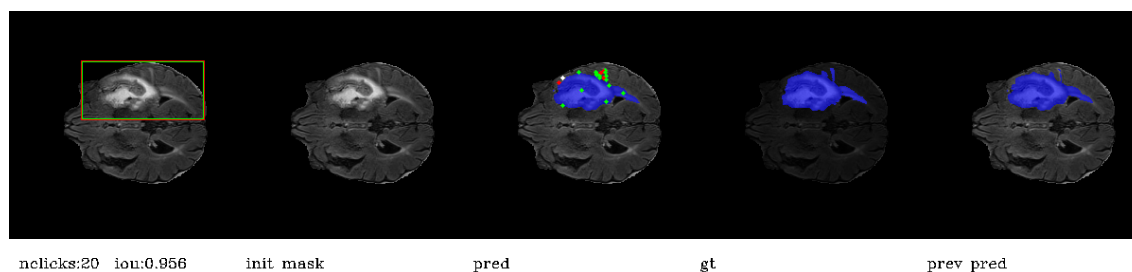
با توجه به این نکته که شبکه HRNet32 یک شبکه بزرگ با تعداد پارامتر بسیار زیاد می‌باشد و همچنین مناسب بودن این شبکه برای تصاویر با وضوح بالا، مشاهده می‌شود که زمان پاسخ الگوریتم برابر ۱/۰۳ ثانیه شده است که نسبت به آزمایش‌های دیگر افزایش دو برابری داشته است. همچنین این شبکه در تمامی معیارهای موجود نسبت به آزمایش قبلی عملکرد بدتری داشته است. با توجه به اینکه شبکه HRNet32 نسخه بزرگ‌تر HRNet18s می‌باشد، می‌توان حدس زد که مدل در این آزمایش دچار Overfit شده است.

نکته جالب دیگر این است که با افزایش تعداد کلیک‌ها، مقادیر IoU و Dice بدتر می‌شوند. این بدان معنی است که شبکه نتوانسته از تعاملات کاربر استفاده مطلوب کند و این راهنمایی‌ها حتی باعث بدتر شدن نتیجه شده است. با بررسی تصاویر خروجی مدل در هر مرحله می‌توان متوجه شد که با توجه به ابعاد مدل، ابتدا یک

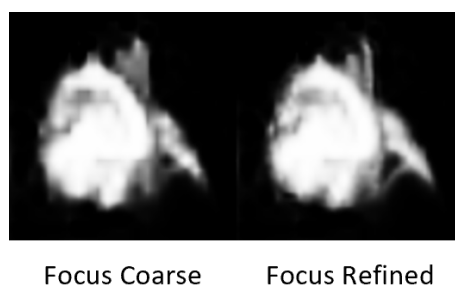
ناحیه‌بندی مناسب ایجاد می‌شود، اما با گذشت زمان پدیده Overfit بر کلیک‌های کاربر غلبه می‌کند و نتیجه با گذشت زمان و افزایش تعداد کلیک بدتر می‌شود.



شکل ۱۷: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۲



شکل ۱۸: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۲



شکل ۱۹: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۲

۱۵-۰ آزمایش ۳: SegFormerB3-FLAIR-WT-Pretrained

در این آزمایش، الگوریتم FocalClick با استفاده از Backbone و ناحیه‌بند SegFormerB3 که از پیش بر روی مجموعه داده ImageNet آموزش داده شده است، اجرا می‌شود. همچنین از مدالیت FLAIR برای ایجاد مجموعه داده استفاده شده است. تسک و هدف مورد نظر آزمایش نیز پیدا کردن ناحیه مربوط به کل تومور یا Whole Tumor (WT) می‌باشد. پس از اینکه آموزش مدل کامل شد، ارزیابی مجموعه تست، بر روی همان مدالیت و همان تسک مرحله آموزش صورت می‌گیرد.

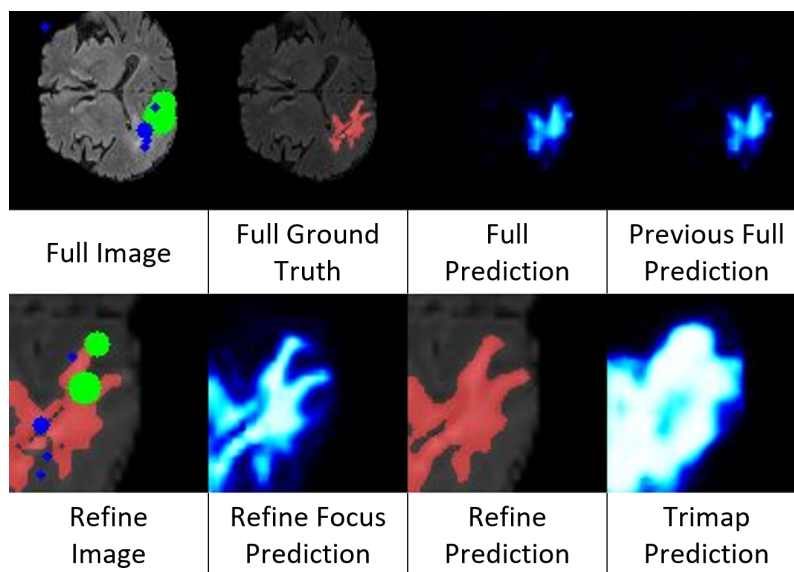
جدول ۶: معیار NoC و NoF در آزمایش ۳

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	3.28	4.03	5.75	320	530	0.059	0:55:07
Dice	2.45	2.81	3.48	176	246	0.059	0:55:07

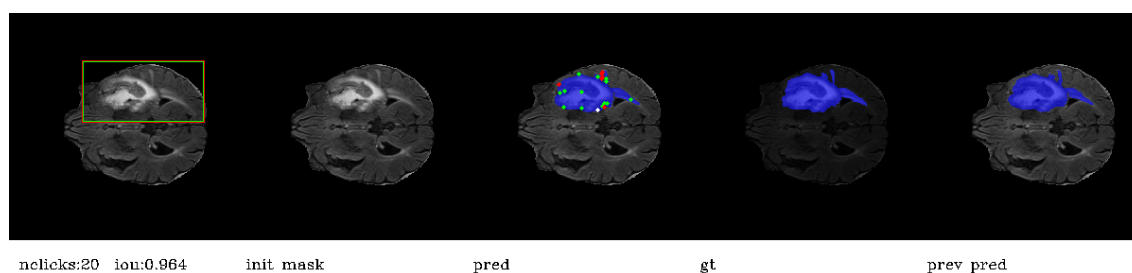
جدول ۷: معیار mIoU و mDice در آزمایش ۳

Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	85.35 %	85.30 %	86.46 %	88.37 %	90.29 %	91.75 %
mDice	88.96 %	88.84 %	89.75 %	91.16 %	92.78 %	94.12 %

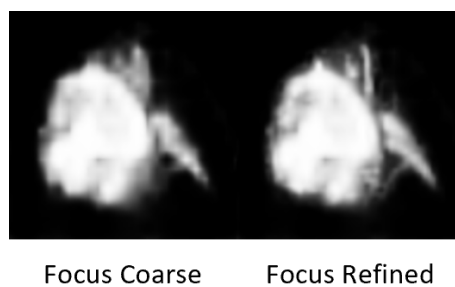
با توجه به اینکه ناحیه‌بند SegFormer یک ناحیه‌بند سبک‌تر با تعداد پارامتر کمتر نسبت به HRNet32 می‌باشد، الگوریتم پس از دریافت کلیک جدید توانسته در زمان خوبی (تقریباً مشابه آزمایش ۱ اما کمی بهتر) خروجی جدید را تولید کند همچنین برای رسیدن معیار Dice به مقدار ۹۰ درصد به صورت میانگین ۳/۴۸ کلیک مورد نیاز است که نسبت به دو آزمایش قبلی بهبود خوبی داشته است. الگوریتم با دریافت ۲۰ کلیک نیز توانسته به میانگین Dice برابر با ۹۴/۱۲ درصد برسد که باز هم نشان دهنده عملکرد خوب الگوریتم نسبت به دو آزمایش قبلی می‌باشد. همچنین هرچه تعداد کلیک بیشتر می‌شود عملکرد مدل نیز بهتر می‌شود، که نشان دهنده تاثیر کلیک‌های کاربر و استفاده مناسب مدل از این کلیک‌ها می‌باشد. به صورت خلاصه می‌توان گفت استفاده از ناحیه‌بند SegFormerB3 باعث بهبود عملکرد مدل در تمامی جنبه‌ها شده است.



شکل ۲۰: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۳



شکل ۲۱: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۳



شکل ۲۲: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۳

۱۶-۰ آزمایش ۴: *SegFormerB3-Mix-WT* و ترکیب مدالیت‌ها

در این آزمایش، الگوریتم FocalClick با استفاده از Backbone و ناحیه‌بند SegFormerB3 و بدون اینکه از پیش چیزی آموخته باشد، استفاده شده است. همچنین از حالت Mix مدالیت‌ها یعنی ترکیب FLAIR, T1ce, T2 برای ایجاد مجموعه داده استفاده شده است. هدف در این آزمایش این است که ببینیم استفاده همزمان از مدالیت‌ها در کنار هم چه تاثیری بر عملکرد مدل دارد. تسک مورد نظر آزمایش نیز پیدا کردن ناحیه مربوط به کل تومور یا Whole Tumor (WT) می‌باشد. پس از اینکه آموزش مدل کامل شد، ارزیابی مجموعه تست، بر روی همان مدالیت‌ها و همان تسک مرحله آموزش صورت می‌گیرد.

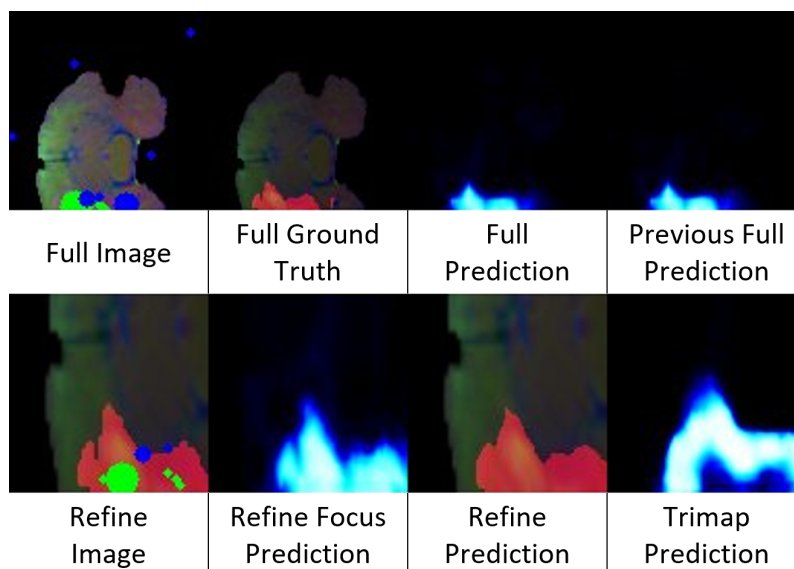
جدول ۸: معیار NoC و NoF در آزمایش ۴

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	3.24	4.15	5.63	319	533	0.061	0:56:54
Dice	2.21	2.66	3.53	138	233	0.061	0:56:54

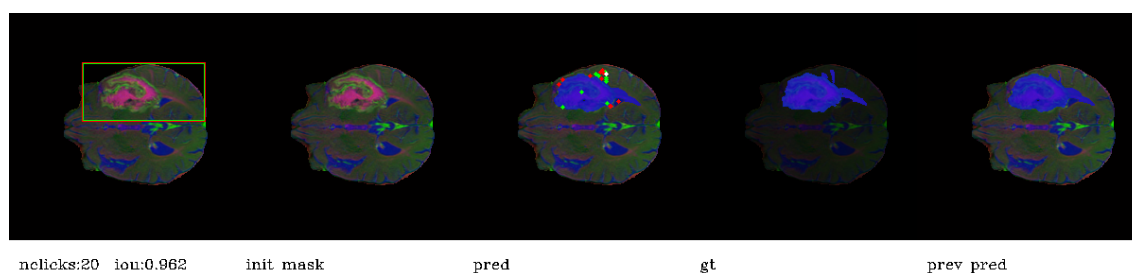
جدول ۹: معیار mIoU و mDice در آزمایش ۴

Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	86.17 %	87.31 %	88.36 %	89.73 %	91.74 %	93.07 %
mDice	89.57 %	90.86 %	91.74 %	92.82 %	94.49 %	95.56 %

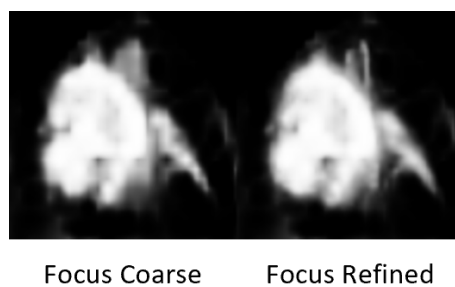
با توجه به این که در این آزمایش مشابه با آزمایش ۳ از ناحیه‌بند SegFormerB3 استفاده شده است و در تعداد پارامترهای مدل تفاوتی ایجاد نشده است، زمان پاسخ و اجرای مدل مشابه آزمایش ۳ بوده است که زمان بسیار مناسبی است. همچنین از مقایسه نتایج مربوط به NoC و mDice این آزمایش با آزمایش قبلی متوجه می‌شویم که مدل در هر دو معیار به صورت کلی بهبود یافته است. این امر ثابت کننده این است که استفاده همزمان و ترکیبی از مدالیت‌ها اثربخش می‌باشد. برای مثال در آزمایش ۳ برای رسیدن معیار IoU به مقدار ۹۰ درصد مدل به صورت میانگین به ۵/۷۵ کلیک نیاز داشته، در حالی که این مقدار در این آزمایش برابر ۵/۶۳ می‌باشد. همچنین در آزمایش ۳ با دریافت ۲۰ کلیک مقدار mIoU برابر ۹۱/۷۵ درصد شده است. در حالی که در این آزمایش این مقدار بهبود یافته است و برابر ۹۳/۰۷ درصد می‌باشد.



شکل ۲۳: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۴



شکل ۲۴: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۴



شکل ۲۵: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۴

۱۷-۰. آزمایش ۵: ارزیابی با تسک جدید دیده نشده در آموزش

در این آزمایش قصد داریم عملکرد مدل را بر روی تسکی که در مرحله آموزش با آن مواجه نشده^{۱۵} است بررسی کنیم. در این راستا از وزن‌های مدل آموزش داده شده در آزمایش ۳ برای ارزیابی تسکی جدید استفاده می‌کنیم. در آزمایش ۳ تسک تعریف شده پیدا کردن ناحیه مربوط به کل تومور (WT) بود و مدل بر اساس آن تسک آموزش داده شده بود. در حالی که در این آزمایش تسک مورد نظر پیدا کردن ناحیه مربوط به هسته تومور (TC) می‌باشد. با تحلیل و تفسیر نتایج این آزمایش می‌توان متوجه شد که کلیک‌های کاربر با وجود تغییر تسک، چقدر برای مدل مفید بوده است تا مدل را در راستای تسک جدید راهنمایی کند.

جدول ۱۰: معیار NoC و NoF در آزمایش ۵

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	5.96	6.39	6.82	667	790	0.059	0:54:53
Dice	4.93	5.49	6.13	456	596	0.059	0:54:53

جدول ۱۱: معیار mIoU و mDice در آزمایش ۵

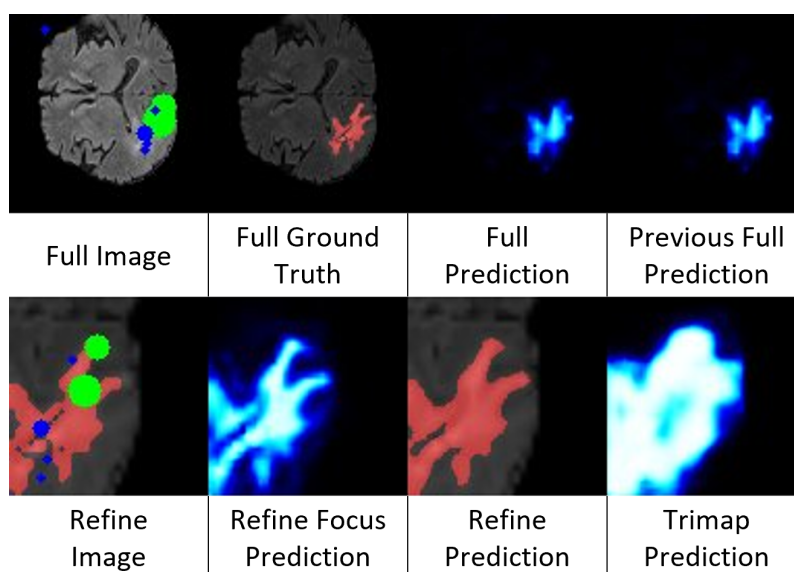
Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	78.04 %	78.29 %	78.96 %	80.30 %	82.85 %	86.92 %
mDice	81.76 %	82.12 %	82.71 %	83.83 %	86.09 %	89.75 %

ابتدا باید گفت ناحیه مربوط به TC اندازه کوچک‌تری نسبت به ناحیه WT دارد و همین موضوع ناحیه‌بندی قسمت TC را دشوارتر می‌کند. بنابراین کم‌تر بودن مقادیر IoU و Dice نسبت به آزمایش‌های قبلی امری طبیعی است. در واقع حتی اگر مدل روی تسک TC هم آموزش ببیند باز هم مقادیر IoU و Dice کمتری خواهد داشت.

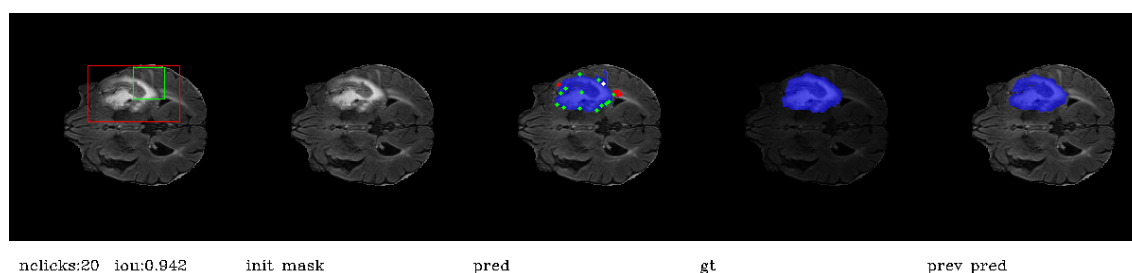
با مشاهده نتایج به دست آمده متوجه می‌شویم در ابتدا که مدل تنها یک کلیک دریافت کرده است، دارای عملکرد بسیار بدی بوده است و مقدار mDice آن برابر ۸۱/۷۶ درصد بوده است. در واقع این اتفاق مربوط به تغییر تسک مدل می‌باشد. اما با افزایش تعداد کلیک متوجه می‌شویم مدل توانسته متوجه تسک جدید

¹⁵Unseen Task

شود و مطابق آن عمل کند. به طوری که با رسیدن تعداد کلیک به عدد ۲۰ مقدار mDice به ۸۹/۷۵ درصد رسیده است که برای تسک TC مقدار مناسبی است. همچنین در این آزمایش تاثیر تعداد کلیک‌های کاربر بر مقادیر mIoU و mDice نسبت به آزمایش‌های قبلی مشهودتر می‌باشد. از تصاویر خروجی تولید شده در مراحل مختلف الگوریتم نیز می‌توان متوجه این نکته شد که هنگامی که تعداد کلیک کاربر کم است، مدل ناحیه مربوط به WT را خروجی می‌دهد و با افزایش تعداد کلیک متوجه می‌شود که یک ناحیه دیگر یعنی TC مورد نظر کاربر بوده است.



شکل ۲۶: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۵ یا همان ۳



شکل ۲۷: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۵



Focus Coarse Focus Refined

شکل ۲۸: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۵

۱۸-۰. آزمایش ۶: ارزیابی با مدالیته و تسک جدید دیده نشده در آموزش

در این آزمایش قصد داریم به صورت همزمان عملکرد مدل را بر روی مدالیته و تسکی که در مرحله آموزش با آن مواجه نشده^{۱۶} است بررسی کنیم. در این راستا از وزن‌های مدل آموزش داده شده در آزمایش ۳ برای ارزیابی مدالیته و تسکی جدید استفاده می‌کنیم. در آزمایش ۳ تسک تعریف شده پیدا کردن ناحیه مربوط به کل تومور (WT) از روی مدالیته FLAIR بود و مدل بر اساس آن مدالیته و تسک آموزش داده شده بود. در حالی که در این آزمایش تسک مورد نظر پیدا کردن ناحیه مربوط به هسته تومور (TC) با استفاده از مدالیته T1ce می‌باشد. با تحلیل و تفسیر نتایج این آزمایش می‌توان متوجه شد که کلیک‌های کاربر با وجود تغییر مدالیته و تسک، چقدر برای مدل مفید بوده است تا مدل را در راستای مدالیته و تسک جدید راهنمایی کند.

جدول ۱۲: معیار NoC و NoF در آزمایش ۶

Metric	NoC80%	NoC85%	NoC90%	NoF85%	NoF90%	TPC	Time
IoU	4.56	4.87	5.42	480	576	0.058	0:54:22
Dice	4.20	4.38	4.68	421	458	0.058	0:54:22

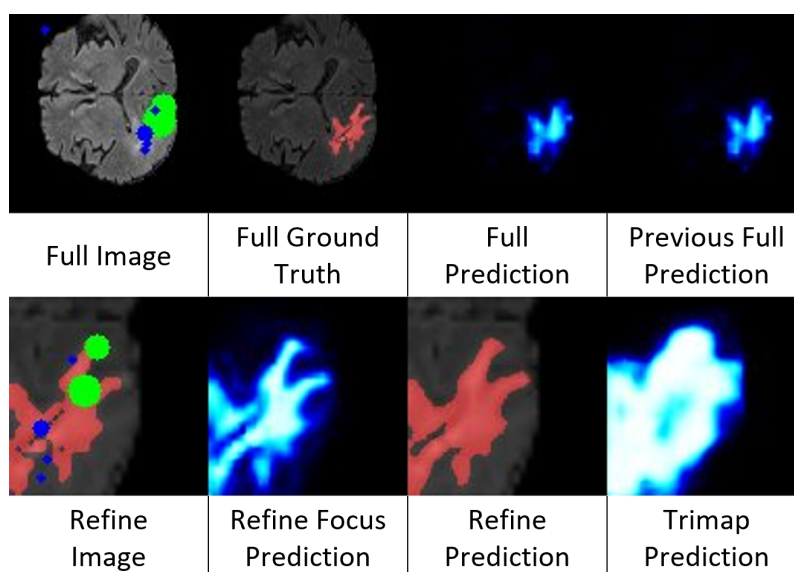
¹⁶Unseen Modality and Task

جدول ۱۳: معیار mIoU و mDice در آزمایش ۶

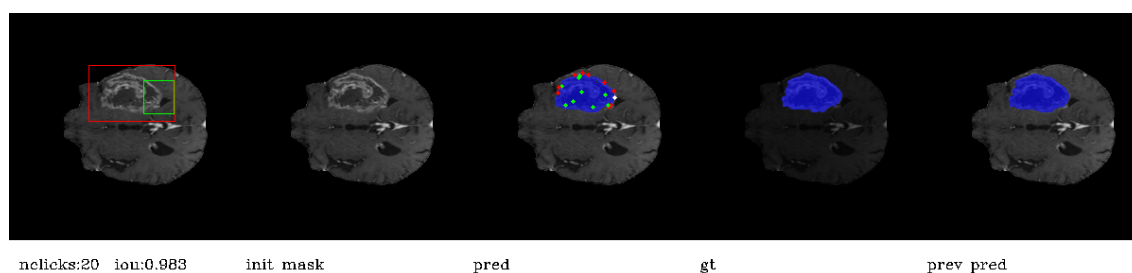
Metric	Clicks=1	Clicks=2	Clicks=3	Clicks=5	Clicks=10	Clicks=20
mIoU	79.80 %	79.27 %	80.06 %	81.31 %	82.96 %	85.49 %
mDice	81.52 %	80.77 %	81.47 %	82.58 %	84.21 %	86.77 %

ابتدا باید گفت ناحیه مربوط به TC اندازه کوچکتری نسبت به ناحیه WT دارد و همین موضوع ناحیه‌بندی قسمت TC را دشوارتر می‌کند. بنابراین کم‌تر بودن مقادیر IoU و Dice نسبت به آزمایش‌های قبلی امری طبیعی است. در واقع حتی اگر مدل روی تسک TC هم آموزش ببیند باز هم مقادیر IoU و Dice کمتری خواهد داشت. علاوه بر مورد بالا، همانطور که در فصل ۲ اشاره شد مدالیت FLAIR ناحیه مربوط به Edema را برجسته‌تر می‌کند. ناحیه مربوط به کل تومور شامل ناحیه Edema می‌باشد اما ناحیه مربوط به هسته تومور شامل آن نمی‌باشد. مدالیت T1ce نیز بخش فزاینده تومور را برجسته‌تر می‌کند که این ناحیه در هر دو ماسک WT و TC وجود دارد. همه این موارد باعث می‌شود مدل با یک تسک چالش برانگیز و سخت مواجه شود. با مشاهده نتایج به دست آمده متوجه می‌شویم در ابتدا که مدل تنها یک کلیک دریافت کرده دارای عملکرد بسیار بدی بوده است و مقدار mDice آن برابر ۸۱/۵۲ درصد بوده است. در واقع این اتفاق مربوط به تغییر مدالیت و تسک مدل می‌باشد. اما با افزایش تعداد کلیک متوجه می‌شویم مدل توانسته متوجه مدالیت و تسک جدید شود و مطابق آن عمل کند. به طوری که با رسیدن تعداد کلیک به عدد ۲۰ مقدار mDice به ۸۶/۷۷ درصد رسیده است. این مقدار نسبت به آزمایش ۵ که مدالیت تغییر نکرده بود کمتر می‌باشد و تغییر مدالیت اثر مخرب خود را داشته است. با این حال در نهایت مدل به ازای تعداد کلیک زیاد به نتیجه بدی نرسیده است. البته یکی از اهداف این پروژه کم کردن کار مورد نیاز از سمت کاربر می‌باشد که نتیجه این آزمایش در تضاد با آن هدف می‌باشد و با تغییرات مدالیت و تسک، کاربر برای رسیدن به نتیجه مطلوب خود باید تلاش زیادی کند.

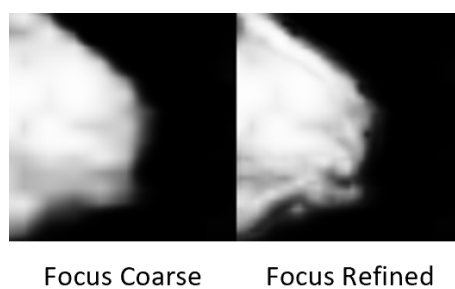
همچنین در این آزمایش تاثیر تعداد کلیک‌های کاربر بر مقادیر mIoU و mDice نسبت به آزمایش‌های قبلی مشهودتر می‌باشد. از تصاویر خروجی تولید شده در مراحل مختلف الگوریتم نیز می‌توان متوجه این نکته شد که هنگامی که تعداد کلیک کاربر کم است، مدل ناحیه مربوط به WT را خروجی می‌دهد و با افزایش تعداد کلیک متوجه می‌شود که یک ناحیه دیگر یعنی TC مورد نظر کاربر بوده است.



شکل ۲۹: خروجی الگوریتم در یکی از مراحل فاز آموزش آزمایش ۶ یا همان ۳



شکل ۳۰: خروجی الگوریتم در یکی از نمونه‌های ارزیابی آزمایش ۶



شکل ۳۱: بازیابی جزئیات تصویر یکی از نمونه‌های ارزیابی در آزمایش ۶

۱۹-۰ مقایسه با نتایج کارهای مرتبط

وجود ۴ نکته باعث شده امکان مقایسه قطعی و دقیق میان کارهای مختلف بر روی مجموعه داده BraTS با مشکل همراه شود. این چهار نکته عبارت اند از:

۱. وجود نسخه‌های سالانه از مجموعه داده BraTS (از سال ۲۰۱۲ تا ۲۰۲۲) و استفاده مقالات از نسخه‌های متفاوت

۲. رفتار متفاوت مقالات در توزیع مجموعه داده BraTS به مجموعه‌های آموزشی، اعتبارسنجی و تست

۳. وجود چهار مدالیته برای هر تصویر و استفاده مقالات از مدالیته‌های مختلف

۴. وجود چندین ماسک مختلف و استفاده مقالات از ماسک‌های مختلف این مجموعه داده

با وجود این مشکلات همچنان می‌توان یک مقایسه تقریبی میان نتیجه روش پیشنهادی ما و سایر روش‌ها ارائه کرد. در این قسمت نتایج روش پیشنهادی ما با نتایج سه مقاله [۱۰] Enhanced U-Net، [۱۰] DeepIGeoS و [۱۴] MIDeepSeg [۸] مقایسه می‌شود. در آخر هم یک مقایسه اجمالی با نتایج مقالات قدیمی‌تر و غیر عمیق صورت می‌گیرد.

۱۹-۰ مقاله Enhanced U-Net

مقاله [۱۰] Enhanced U-Net یک روش غیر تعاملی ناحیه‌بندی را ارائه می‌کند. در این روش از شبکه U-Net برای ناحیه‌بندی استفاده شده است. همچنین از نسخه‌های سالانه ۲۰۱۷ تا ۲۰۲۰ این مجموعه داده به صورت مجزا استفاده شده است. این مقاله نسخه ۲۰۱۹ این مجموعه داده را به صورت ۶۸ درصد آموزش، ۲۰ درصد اعتبارسنجی و ۱۲ درصد تست تفکیک کرده است. همچنین در این مقاله از مدالیته‌های FLAIR و T1ce استفاده شده است. همانطور که در جدول ۱۴ مشاهده می‌شود این مقاله به صورت تقریبی به Dice برابر با ۸۴ درصد رسیده است که نسبت به روش پیشنهادی و تعاملی ما دارای مقدار کمتری می‌باشد. اختلاف زیاد نتیجه این مقاله با روش ما به دلیل غیر تعاملی بودن این مقاله می‌باشد و از این نکته می‌توان به اهمیت و تاثیر تعامل کاربر پی برد.

جدول ۱۴: نتایج مقاله Enhanced U-Net [۱۰]

Dataset	Loss	Accuracy	Mean IoU	Precision	Sensitivity	Specificity	Dice Score
BraTS 2017	0.0056	0.9980	0.9637	0.9973	0.9970	0.9972	0.8453
BraTS 2018	0.0057	0.9979	0.8927	0.9972	0.9970	0.9940	0.8160
BraTS 2019	0.0054	0.9981	0.9130	0.9974	0.9971	0.9991	0.8409
BraTS 2020	0.0056	0.9980	0.8935	0.9973	0.9970	0.9983	0.8300

۱۹-۰-۲ مقاله DeepIGeoS

مقاله DeepIGeoS [۱۴] یک روش تعاملی ناحیه‌بندی را ارائه می‌کند. در این روش از فاصله ژئودزیک^{۱۷} برای تبدیل کلیک‌های کاربر به نقشه‌های تعامل استفاده شده است. همچنین در این روش از شبکه‌های P-Net [۱۳] و R-Net [۱] و CRF-Net [۷] به صورت همزمان برای ناحیه‌بندی استفاده شده است.

در این مقاله از نسخه ۲۰۱۵ مجموعه داده BraTS استفاده شده است. این مقاله نسخه ۲۰۱۵ این مجموعه داده را به صورت ۸۵ درصد آموزش و ۱۵ درصد اعتبارسنجی و تست تفکیک کرده است. همچنین در این مقاله فقط از مدالیته FLAIR استفاده شده است. تسک تعریف شده نیز بر روی ماسک WT می‌باشد. همانطور که در جدول ۱۵ مشاهده می‌شود این مقاله در بهترین حالت خود به Dice برابر با ۸۹/۹۳ درصد رسیده است که نسبت به روش پیشنهادی ما دارای مقدار کمتری می‌باشد.

¹⁷Geodesic

جدول ۱۵: نتایج مقاله DeepIGeoS [۱۴]

Method	Dice (%)	ASSD (pixels)
Before refinement	87.55 ± 6.72	2.04 ± 1.70
Min-cut user-editing	88.41 ± 7.05	1.74 ± 1.53
3D R-Net (Euc)	88.82 ± 7.68	1.60 ± 1.56
3D R-Net	89.30 ± 6.82	1.52 ± 1.37
3D R-Net (Euc) + CRF-Net (fu)	89.27 ± 7.32	1.48 ± 1.22
3D R-Net + CRF-Net (fu)	89.93 ± 6.49	1.43 ± 1.16

۱۹-۰-۳ مقاله MIDeepSeg

مقاله MIDeepSeg [۸] یک روش تعاملی ناحیه‌بندی را ارائه می‌کند. در این روش از فاصله ژئودزیک نمایی^{۱۸} برای تبدیل کلیک‌های کاربر به نقشه‌های تعامل استفاده شده است. همچنین در این روش از شبکه U-Net [۱۱] برای ناحیه‌بندی اولیه استفاده شده است. برای اعمال کلیک‌های کاربر بر روی ناحیه‌بندی اولیه تولید شده نیز از روش غیر عمیق Graph Cut استفاده شده است.

در این مقاله از نسخه ۲۰۱۸ مجموعه داده BraTS استفاده شده است. این مقاله نسخه ۲۰۱۸ این مجموعه داده را به صورت ۷۸ درصد آموزش و ۲۲ درصد اعتبارسنجی و تست تفکیک کرده است. همچنین در این مقاله از مدل‌های FLAIR و T1ce به صورت مجزا استفاده شده است. تسک تعریف شده نیز بر روی ماسک TC و WT می‌باشد.

جدول ۱۶: توزیع مجموعه داده در مقاله MIDeepSeg [۸]

Object	Modality	N_{train}	N_{test}	Dataset
Tumor core	MRI (T1ce)	170	47	BraTS2018
Whole tumor	MRI (FLAIR)	0	60	BraTS2018

این مقاله نیز مشابه روش پیشنهادی ما عملکرد خود را بر روی مدل‌های و تسک دیده نشده می‌سنجد.

¹⁸Exponentialized Geodesic Distance (EGD)

همانطور که در جدول ۱۷ مشاهده می‌شود این مقاله در بهترین حالت خود به Dice برابر با ۸۸/۷۱ درصد رسیده است که نسبت به روش پیشنهادی ما دارای مقدار کمتری می‌باشد.

جدول ۱۷: نتایج مقاله MIDeepSeg [۸]

Method	Dice (%)	ASSD (mm)	Time (s)
3D Graph Cuts	78.91 ± 14.98	3.46 ± 5.10	99.4 ± 36.7
ITK-SNAP	82.34 ± 11.42	1.99 ± 1.31	173.0 ± 75.5
DeepIGeoS	82.47 ± 10.78	3.34 ± 3.81	82.2 ± 44.7
DeepGrabCut	82.58 ± 11.79	2.89 ± 2.37	10.5 ± 8.3
DIOS	83.34 ± 10.47	2.57 ± 1.79	67.5 ± 23.6
DEXTR	86.39 ± 9.03	1.59 ± 1.11	34.7 ± 18.6
MIDeepSeg	88.71 ± 7.00	1.24 ± 0.88	28.6 ± 12.2

۱۹-۰-۴ روش‌های سنتی و غیر عمیق

در این قسمت مقدار Dice برای روش‌های سنتی و غیر عمیق آورده شده است. با توجه به تعداد زیاد این روش‌ها، جزئیات آنها ذکر نمی‌شود. نکته قابل توجه اختلاف زیاد مقدار Dice در این روش‌ها نسبت به روش‌های عمیق می‌باشد. طوری که برای تسک WT مقدار Dice بین ۴۴ تا ۸۲ درصد می‌باشد. همچنین برای تسک TC نیز این معیار بین ۱۷ تا ۷۰ درصد می‌باشد.

جدول ۱۸: نتایج روش‌های قدیمی و غیر عمیق [۹]

	whole		core		active	time (min) (arch)
Dice (in %)	<i>LG/HG</i>		<i>LG/HG</i>			
Bauer	68	49/74	48	30/54	57	8 (CPU)
Buendia	57	19/71	42	8/54	45	0.3 (CPU)
Cordier	68	60/71	51	41/55	39	20 (Cluster)
Doyle	74	63/78	44	41/45	42	15 (CPU)
Festa	62	24/77	50	33/56	61	30 (CPU)
Geremia	62	55/65	32	34/31	42	10 (Cluster)
Guo	74	71/75	65	59/67	49	< 1 (CPU)
Hamamci	72	55/78	57	40/63	59	20 (CPU)
Meier	69	46/77	50	36/55	57	6 (CPU)
Menze (D)	78	81/76	58	58/59	54	20 (CPU)
Menze (G)	69	48/77	33	9/42	53	10 (CPU)
Reza	70	52/77	47	39/50	55	90 (CPU)
Riklin Raviv	74	<i>na</i> /74	50	<i>na</i> /50	58	8 (CPU)
Shin	30	28/31	17	22/15	5	8 (CPU)
Subbanna	75	55/82	70	54/75	59	70 (CPU)
Taylor	44	24/51	28	11/34	41	1 (Cluster)
Tustison	75	68/78	55	42/60	52	100 (Cluster)
Zhao (I)	82	78/84	66	60/68	49	15 (CPU)
Zhao (II)	76	67/79	51	42/55	52	20 (CPU)
Zikic	75	62/80	47	33/52	56	2 (CPU)
Best Combination	88	86/89	78	66/82	71	
Fused	82	68/87	73	62/77	65	

۰-۲۰. جمع‌بندی

در این پروژه روش و الگوریتم FocalClick [۲] به عنوان یک روش ناحیه‌بندی تعاملی عمیق مبتنی بر کلیک استفاده شد تا تصاویر پزشکی MRI از مغز را ناحیه‌بندی و نواحی دارای تومور را مشخص کند. آزمایش‌های صورت گرفته نشان دادند که الگوریتم ارائه شده عملکرد بسیار بهتری نسبت به الگوریتم‌های غیر تعاملی و تعاملی قبلی دارد. این آزمایش‌ها بر روی مجموعه داده BraTS سال ۲۰۲۰ صورت گرفت و در تمامی آزمایش‌ها نتیجه به دست آمده بهتر از روش‌های پیشین بود. این عملکرد بهتر به دلیل وجود کلیک‌های کمکی کاربر بود که با ارائه روشی برای بهره‌برداری مناسب از این کلیک‌ها، ثابت شد که وجود تعامل کاربر تا چه میزان کمک کننده است. همچنین عملکرد مناسب الگوریتم FocalClick در معیارهایی نظیر TPC و NoC نشان دهنده این است که این الگوریتم بر خلاف روش‌های پیشین کاملاً مناسب استفاده در صنعت و دنیای واقعی می‌باشد. الگوریتم FocalClick به طور قابل توجهی کارایی Pipeline‌های موجود را بهبود می‌بخشد و امکان استقرار آنها در دستگاه‌های کم مصرف را فراهم می‌کند. علاوه بر موارد بالا، نتایج آزمایش‌ها نشان داد که الگوریتم مورد نظر بر روی تسک‌های متفاوت به خوبی تعمیم^{۱۹} می‌یابد. این تعمیم به گونه‌ای است که مدل با آموزش بر روی یک تسک، می‌تواند نتیجه مطلوبی در تسک دیگر دیده نشده داشته باشد. در ادامه به کارهایی که می‌تواند قدم‌های بعدی پروژه باشد، اشاره می‌شود.

۰-۲۱. پیشنهادها و کارهای آینده

۰-۲۱-۱ مناسب‌سازی FocalClick برای ورودی سه‌بعدی

در این پروژه به دلیل وجود نداشتن ماژول‌های مورفولوژی سه‌بعدی و همچنین تعریف نشدن قسمت‌های مختلف الگوریتم FocalClick برای حالت سه‌بعدی و چالش برانگیز بودن طراحی سه‌بعدی این قسمت‌ها، مجموعه داده سه‌بعدی BraTS با انجام عمل Slicing بر روی Depth تبدیل به حالت دوبعدی شد. این کار می‌تواند باعث از دست رفتن اطلاعات مکانی^{۲۰} و در نهایت باعث کاهش عملکرد شود. البته به دلیل محدودیت‌های حافظه‌ای و بزرگ بودن هر نمونه سه‌بعدی، امکان ورودی دادن این نمونه‌ها به صورت کامل به مدل وجود ندارد. برای حل این مشکل می‌توان به جای عمل Crop که در FocalClick از آن استفاده می‌شود،

¹⁹Generalization²⁰Spatial

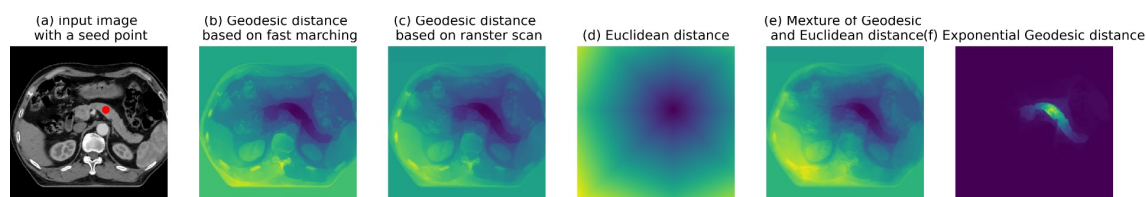
از ترکیب دو عمل Crop و Slice استفاده کرد. در این صورت به جای اینکه هر نمونه مجموعه داده که دارای ابعاد $155 \times 240 \times 240$ می باشد به ابعاد 96×96 تبدیل شود و بعد سوم خود را کاملاً از دست دهد، می توان آن نمونه را به ابعاد $96 \times 96 \times 48$ تبدیل کرد تا اطلاعات مکانی از بعد سوم نیز وجود داشته باشد. این عمل در مقالات DeepIGeoS [۱۴] و MIDeepSeg [۸] انجام شده است و نتیجه بخش بوده است.

۰-۲۱-۲ استفاده از BRS در FocalClick

در مقاله BRS [۶] تطبیق پذیری و تاثیر مثبت روش BRS بر روی چندین شبکه CNN نشان داده می شود و ثابت می شود که استفاده از BRS باعث کاهش معیار NoC می شود. کم کردن تعداد کلیک مورد نیاز و همچنین کاهش تلاش کاربر، یکی از مهم ترین اهداف تعریف شده برای طراحی مدل های ناحیه بندی تعاملی می باشد. یکی از کارهای آینده در این پژوهش می تواند به کار گیری روش BRS در الگوریتم FocalClick باشد. البته این کار نیازمند تغییرات زیادی در الگوریتم موجود می باشد.

۰-۲۱-۳ استفاده از تبدیل فاصله ژئودزیک نمایی

در الگوریتم استفاده شده در این پروژه، کلیک های کاربر به نقشه های Binary تبدیل می شوند. به عبارت دیگر هر کلیک کاربر به یک دایره با شعاع ۲ تبدیل می شود. این تبدیل جزو ساده ترین تبدیل های موجود می باشد و حتی نسبت به سایر روش های معمول دیگر مانند فاصله اقلیدسی یا گاوسی اطلاعات کمتری را برای مدل فراهم می کند. در مقاله DeepIGeoS از فاصله ژئودزیک برای تبدیل کلیک های کاربر به نقشه های تعامل استفاده شده است. در ادامه مقاله MIDeepSeg از حالت بهبود یافته این فاصله یعنی فاصله ژئودزیک نمایی (EGD) استفاده کرده است. در تصویر ۳۲ می توان نقشه تعامل ساخته شده در روش های مختلف را مشاهده کرد. همانطور که مشاهده می شود تبدیل EGD بهترین نقشه تعامل را برای مدل فراهم کرده است و تقریباً ناحیه مورد نظر کاربر را به خوبی نمایش داده است. استفاده از تبدیل EGD را می توان به عنوان یکی از کارهای آینده در نظر گرفت.



شکل ۳۲: مقایسه تبدیل‌های مختلف برای کلیک‌های کاربر

۰-۲۱-۴ پیشنهادهای دیگر

علاوه بر موارد بالا، استفاده از نسخه ۲۰۲۲ مجموعه داده BraTS که مجموعاً دارای ۵۸۸۰ (تقریباً ۱۰ برابر نسخه ۲۰۲۰) اسکن MRI می‌باشد می‌تواند برای آموزش مدل بسیار مناسب باشد. در این پروژه از روش‌های داده‌افزایی بسیار ساده نظیر Flip و Rotaion استفاده شده است. در آینده می‌توان از روش‌های بیشتری در قسمت داده‌افزایی استفاده کرد. استفاده از ناحیه‌بندها و Backbone‌های دیگر هم می‌تواند جزو کارهای آینده باشد.

- [1] CHEN, S., CHEN, Z., YAN, L., AND GUO, W. R-net: A novel fully convolutional network–based infrared image segmentation method for intelligent human behavior analysis. *Infrared Physics & Technology* 123 (04 2022), 104164.
- [2] CHEN, X., ZHAO, Z., ZHANG, Y., DUAN, M., QI, D., AND ZHAO, H. Focalclick: Towards practical interactive image segmentation. in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA, jun 2022), IEEE Computer Society, pp. 1290–1299.
- [3] CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA, jul 2017), IEEE Computer Society, pp. 1800–1807.
- [4] DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K., AND FEI-FEI, L. Imagenet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255.
- [5] HE, K., GKIOXARI, G., DOLLÁR, P., AND GIRSHICK, R. Mask r-cnn. in *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2980–2988.
- [6] JANG, W.-D., AND KIM, C.-S. Interactive image segmentation via backpropagating refinement scheme. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019).
- [7] LI, H., AND PEERS, P. Crf-net: Single image radiometric calibration using cnns. in *Proceedings of the 14th European Conference on Visual Media Production (CVMP 2017)* (12 2017), pp. 1–9.

- [8] LUO, X., WANG, G., SONG, T., ZHANG, J., AERTSEN, M., DEPREST, J., OURSELIN, S., VERCAUTEREN, T., AND ZHANG, S. Mideepseg: Minimally interactive segmentation of unseen objects from medical images using deep learning. *Medical Image Analysis* 72 (2021), 102102.
- [9] MENZE, B. H. E. A. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging* 34, 10 (2015), 1993–2024.
- [10] NASIM, M. A. A., MUNEM, A., ISLAM, M., PALASH, M. A. H., HAQUE, M. M. A., AND SHAH, F. M. Brain tumor segmentation using enhanced u-net model with empirical analysis. *ArXiv abs/2210.13336* (2022).
- [11] RONNEBERGER, O., FISCHER, P., AND BROX, T. U-net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Cham, 2015), Springer International Publishing, pp. 234–241.
- [12] SOFIHUK, K., PETROV, I. A., AND KONUSHIN, A. Reviving iterative training with mask guidance for interactive segmentation. in *2022 IEEE International Conference on Image Processing (ICIP)* (2022), pp. 3141–3145.
- [13] SONG, W., YU, H., AND WU, J. Patch network for medical image segmentation, 2023.
- [14] WANG, G., ZULUAGA, M., LI, W., AUGHWANE, R., PATEL, P., AERTSEN, M., DOEL, T., DAVID, A., DEPREST, J., OURSELIN, S., AND VERCAUTEREN, T. Deepigeos: A deep interactive geodesic framework for medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence PP* (07 2017).
- [15] WANG, J., SUN, K., CHENG, T., JIANG, B., DENG, C., ZHAO, Y., LIU, D., MU, Y., TAN, M., WANG, X., LIU, W., AND XIAO, B. Deep high-resolution representation learning for visual recognition, 2019.
- [16] XIE, E., WANG, W., YU, Z., ANANDKUMAR, A., ALVAREZ, J. M., AND LUO, P. Segformer: Simple and efficient design for semantic segmentation with transformers. in *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)* (12 2021).
- [17] XU, N., PRICE, B., COHEN, S., YANG, J., AND HUANG, T. Deep interactive object selection. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA, USA, jun 2016), IEEE Computer Society, pp. 373–381.