دانشگاه علم و صنعت ایران
یادگیری عمیق (بهار ۱۳۹۹)

استفاده از EfficientNet سه بعدی برای طبقه‌بندی تصاویر CT-scan ریه

رامین کمالی
دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران
ramin_kamali@comp.iust.ac.ir

محمدرضا داودی
دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران
davoodi_m@comp.iust.ac.ir

چکیده

طبقه‌بندی غده‌های ریه به دسته‌های خوش‌خیم و بدخیم در تصاویر CT موضوعی است که در این کار ارائه شده است. برای انجام این طبقه‌بندی از یادگیری عمیق استفاده شده که این روش بصورت خودکار ویژگی‌ها را از داده‌ها استخراج می‌کند. با توجه به سه بعدی بودن تصاویر CT-scan برای استخراج ویژگی از شبکه‌های کانولوشنی سه بعدی استفاده کرده‌ایم. ایده اصلی مطرح شده استفاده از مقیاس مختلف به منظور استخراج ویژگی‌های مطلوب‌تر است که به وسیله شبکه‌های مسیر دوگانه پیاده‌سازی(dual-path) شده است. در این شبکه‌ها از دو تصویر از هر غده استفاده شده است که یک تصویر کوچک که تنها غده و مرز آن در تصویر موجود است و یک تصویر بزرگتر که اطراف غده نیز در آن مشخص است. برای بهبود نتایج و کم شدن پارامترها بجای اینکه شبکه را تنها در یکی از فاکتورهای عمق عرض یا رزولوشن بزرگ‌تر کنیم بصورت ترکیبی بزرگ‌تر کرده‌ایم. در مرحله بعد از تابع فعال‌سازی Leaky-RELU برای بهبود نهایی استفاده شده است. شبکه دوم مطرح شده ازیک شبکه اشتراکی در هر دو مقیاس استفاده می‌کند تا تعداد پارامترها را کم کند. شبکه دیگری که مطرح شده بجای استفاده از بلوک‌های کانولوشنی معمولی از بلوک‌های متراکم کانولوشنی استفاده کرده است. این استفاده به‌دلیل حل مشکل ناپدید شدن گرادیان (vanishing-gradient) است. در آخرین معماری شبکه که مطرح شده است به‌جای استفاده از دو ورودی جداگانه برای حفظ چند مقیاسه بودن از چند فیلتر با اندازه‌های مختلف در ابتدای شبکه استفاده شده است.

۱ مقدمه

سرطان ریه یکی از کشنده‌ترین سرطان‌ها است و سالانه جان بسیاری از انسان‌ها را در سرتاسر دنیا می‌گیرد. مشاهدات پزشکی نشان می‌دهد که هر چه شناسایی و درمان این بیماری زودتر صورت بگیرد شانس زنده ماندن بیمار بیشتر می‌گردد و نرخ مرگ و میر کاهش می‌یابد. یکی از بهترین ابزار تشخیص غده‌های سرطانی استفاده از تصاویر سه بعدی CT-scan هستند که نسبت به تصاویر دو بعدی اشعه ایکس، اطلاعات بیشتری را منتقل می‌کنند. پزشکان با وجود استفاده از این تصاویر برای اطمینان بیشتر از آزمایش‌ها و نمونه برداری نیز استفاده می‌کنند که بسیار زمان‌گیر و پرهزینه هستند و در برخی از موارد برای بیمار حتی مضر هستند. از طرفی به‌دلیل ناهمگن بودن بافت تومورها ممکن است نتایج آزمایش درست نباشند و نیاز به نمونه گیری مجدد باشد. به همین دلیل توجه‌ها به سمت سیستم‌های تشخیصی به کمک رایانه (CAD) جمع شد. در این سیستم‌ها سعی بر انجام این کار با دقت و سرعت شده است. روش‌های اولیه مبتنی بر استخراج ویژگی بصورت دستی بودند که استخراج این ویژگی‌ها زمان‌گیر بود و این ویژگی‌ها از نظر مفهومی سطح بالا نبودند. ولی با این وجود همین روش‌ها نیز تا حدودی کارا واقع شدند و نیاز به آزمایش‌های ثانویه را کم کردند. روش‌های بعدی استفاده از یادگیری عمیق برای استخراج ویژگی‌های سطح بالاتر بود.

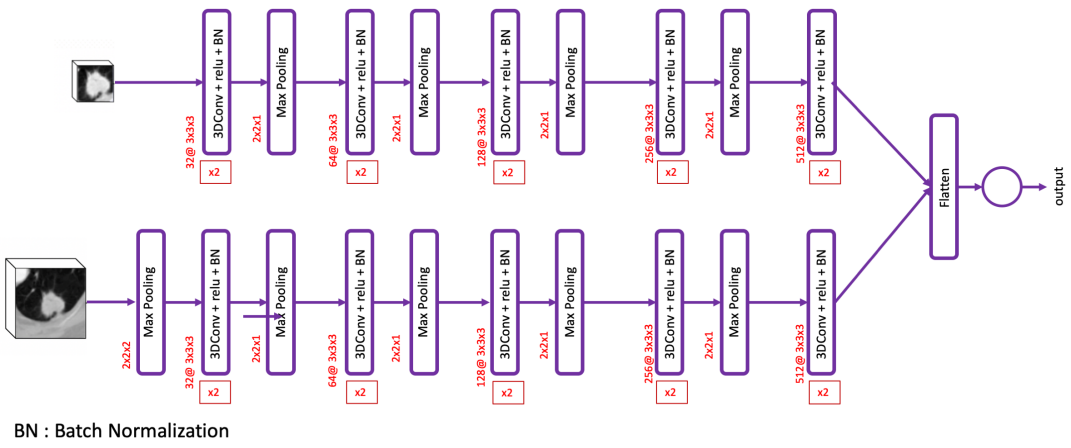
این روش‌ها نیز بهبود خوبی را نشان دادند ولی با توجه به سه بعدی بودن داده بخش مهمی از اطلاعات را استفاده نمی‌کردند. در ادامه پیشرفت‌ها، از شبکه‌های سه بعدی استفاده شد که نتایج را تا حد خیلی خوبی بهبود دادند.

استفاده از شبکه‌های سه بعدی چالش‌های خاص خود را دارد. اولین چالش تعداد پارامترهای زیاد آن است. در شبکه‌های کانولوشنی سه بعدی، فیلترها سه بعدی هستند در نتیجه تعداد پارامترهای آنها زیادتر است. از طرفی مجموعه داده‌های تصاویر پزشکی معمولا به اندازه مجموعه داده‌های تصاویر طبیعی داده ندارند که همین موضوع باعث می‌شود تعداد پارامترهای زیاد، فرایند آموزش را دچار مشکل کنند. از طرفی استفاده از شبکه‌های سه بعدی نیازمند منابع زیادی است چون حجم محاسبات و داده‌ها در طول شبکه را بسیار زیاد می‌کند.

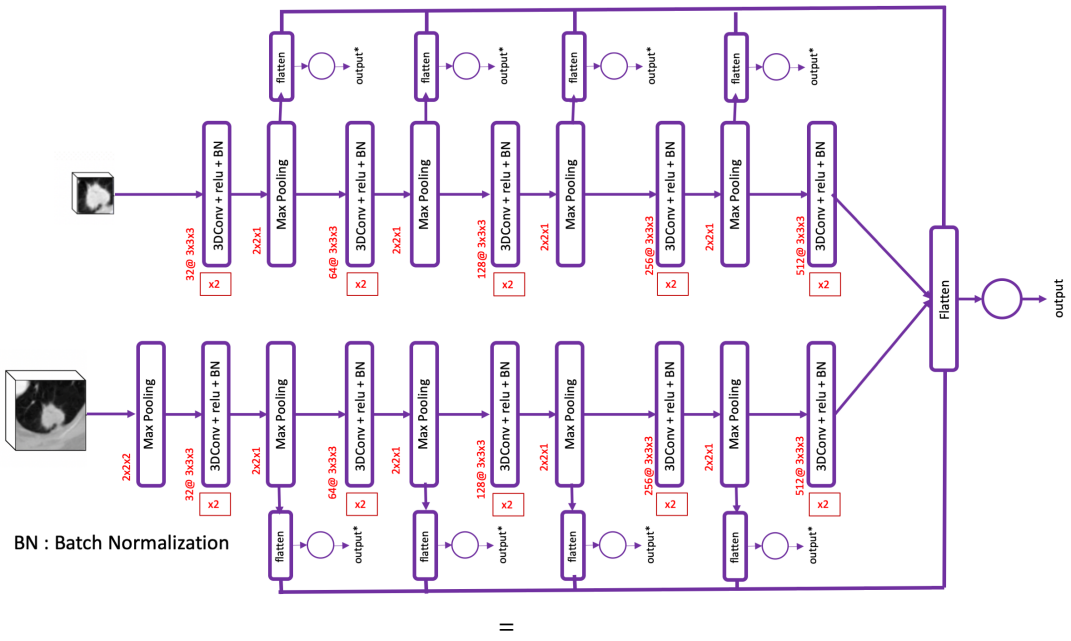
در این پروژه از مجموعه داده LIDC-IDRI استفاده شده است که از تصاویر CT-scan گرفته شده از ۱۰۱۰ بیمار تشکیل شده است. یکی از چالش‌های موجود در رابطه با این مجموعه داده تعیین برچسب داده است. برای تعیین کردن این برچسب از داده‌های استفاده شده که نظر چند پزشک راجع به تومور بیان کرده است. نظر هر پزشک با پزشک دیگر ممکن است تا حدودی متفاوت باشد. هر پزشک برای غده عددی بین ۰ تا ۵ نسبت می‌دهند. عدد ۰ بیانگر عدم نظر پزشک راجع به تومور است و عدد های ۱ تا ۵ میزان بدخیمی غده را بیان می‌کنند که هر چه عدد بیشتر باشد بیانگر بدخیمی بیشتر غده است. طبق یک روش پرکاربرد برای تعیین برچسب این داده‌ها، داده‌هایی را که حداقل سه پزشک راجع به بدخیمی آن‌ها نظر داده‌اند در نظر می‌گیریم و مابقی داده‌ها را دور می‌ریزیم. در مرحله عددی که پزشکان به آن نسبت داده‌اند را در نظر گرفته و میانه این عددها را بدست می‌آوریم. اگر این عدد بزرگ‌تر از ۳ بود این غده بدخیم است و اگر کوچکتر از ۳ بود خوش‌خیم است. اگر این عدد ۳ بدست آمده قابل نتیجه‌گیری نیست و آن را دور میریزیم. میانه مرکز غده را که پزشکان نشان‌گذاری کرده‌اند نیز موجود است. به‌دلیل اینکه این‌جا وظیفه تنها طبقه‌بندی است، از آن‌ها جهت مشخص کردن مختصات تومور و ورودی شبکه استفاده می‌کنیم.

۲ کارهای مرتبط / پیش‌زمینه

در سیستم‌های CAD اولیه فقط از ویژگی‌های که بصورت دستی استخراج می‌شدند، استفاده می‌گردید. این ویژگی‌ها سطح پایین بودند و محاسبه‌ی آن‌ها زمان‌گیر بود. به مرور زمان روش‌های مبتنی بر یادگیری عمیق که در کابردهای تصاویر طبیعی موفق بودند، وارد این عرصه شد. در از برش‌هایی از داده‌های سه بعدی استفاده می‌شد و بعد سوم آن نادیده گرفته می‌شد. نادیده گرفتن بعد سوم موجب از دست رفتن اطلاعات مهمی از می‌گردد. بنابراین در مقالات جدیدتر، بیشتر از شبکه‌های کانولوشن سه بعدی برای استخراج ویژگی استفاده می‌شود. تومورها اندازه، جهت، مکان و بافت متفاوتی دارند بنابراین طبقه‌بندی آن‌ها کار آسانی نخواهد بود. به همین دلیل ایده‌ای مطرح شده که تومورها بصورت چند مقیاسی مورد بررسی قرار بگیرند تا هم ویژگی تومورهای کوچک و هم ویژگی‌های مطلبوب تومورهای بزرگ‌تر به‌خوبی استخراج شوند. یکی از راه‌های پیاده‌سازی این ایده، این است که تصاویر با اندازه‌های مختلف را به عنوان ورودی به شبکه داده و برای تمام اندازه‌ها استخراج ویژگی را انجام داده و بر أساس تمام این ویژگی‌ها تصمیم نهایی گرفته می‌شود که این روش در (۱) آمده است. در این کار انجام شده ورودی‌های شبکه دو قطعه ۱۰۰ در ۱۰۰ و ۱۰ در ۵۰ و ۵۰ در ۵ است که تصویر اول ناحیه اطراف تومور را نیز بررسی می‌کند و تمرکز قطعه اول بیشتر بر غده است. سپس ویژگی‌های هر یک از این تصاویر با استفاده از شبکه کانولوشنی سه بعدی استخراج می‌شوند و در انتها به هم الحاق می‌گردند تا طبقه بندی نهایی صورت گیرد. شبکه اولیه تعریف شده برای این مساله از دو زیر شبکه سه بعدی برای استخراج ویژگی‌های هر یک از دو تصویر تشکیل شده است. برای این‌که تا حد امکان از محو شدن گرادیان دوری کنیم در مدل دیگر از طبقه‌بندهای اضافی بهره می‌بریم تا گرادیان به ابتدای شبکه تزریق شوند. این دو شبکه در شکل۱ و شکل ۲ موجود می‌باشند.



شکل ۱: شبکه کانولوشن سه بعدی معمولی

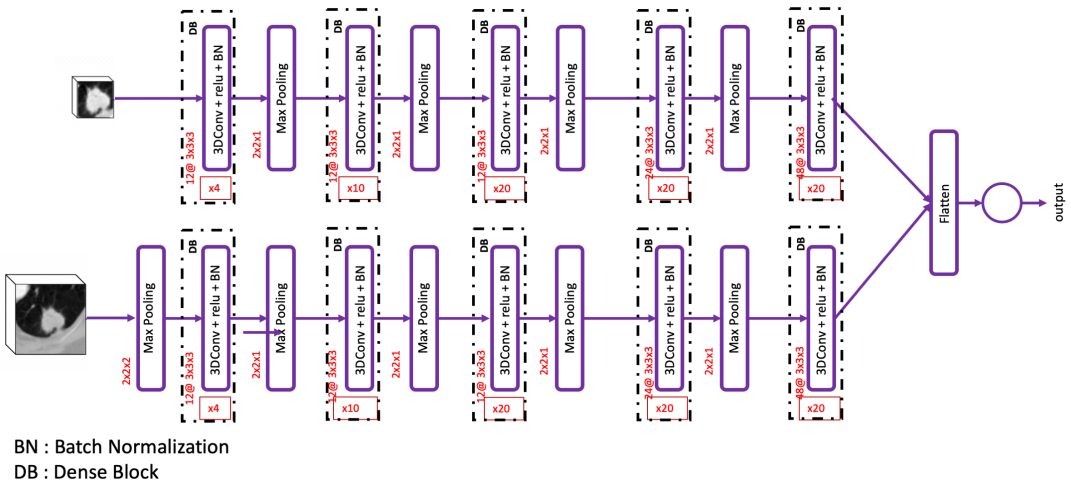


=

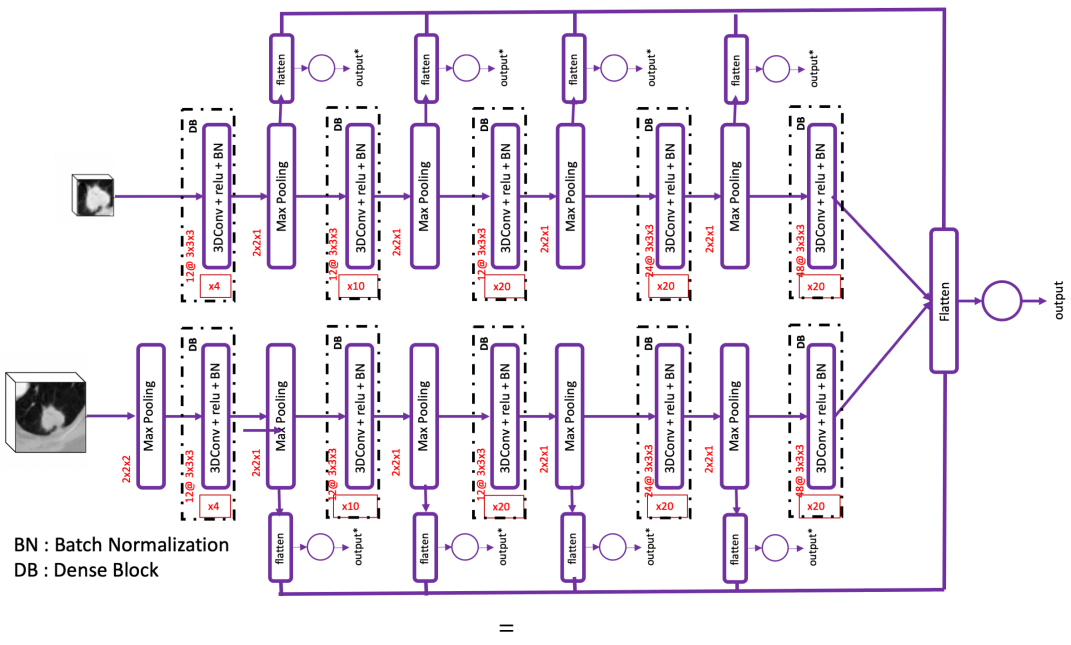
شکل ۲: کانولوشن سه بعدی معمولی با طبقه‌بند اضافی

در دیگر شبکه ارائه شده به‌جای استفاده از بلوک‌های معمولی کانولوشن سه بعدی، از بلوک‌های کانولوشنی سه بعدی که اتصالات dense دارند استفاده می‌کنیم. منظور از اتصالات dense این است که در هر بلوک، ورودی هر لایه کانولوشن به تمام خروجی‌های لایه‌های کانولوشنی که قبل از آن در بلوک آمده‌اند، متصل است. استفاده از این بلوک‌ها منجر به استخراج ویژگی‌های بهتر از تصویر می‌گردد و درعین حال منجر به کاهش محوشدن گرادیان می‌شود. در زمینه تصاویر پزشکی این بلوک‌های dense بسیار پرکاربرد هستند. استفاده از این بلوک‌ها موجب یادگیری بهتر مدل می‌گردد. در(۱) شبکه‌ی مشابه با شبکه ذکر شده آورده شده

است که معماری آن مانند شکل ۳ است. مدل دیگر مطرح شده، مشابه مدل قبل است که در آن از طبقه‌بندهای اضافی استفاده شده است ومعماری آن در شکل ۴ قابل مشاهده است.



شکل ۳: شبکه کانولوشن سه بعدی با بلوک‌های dense

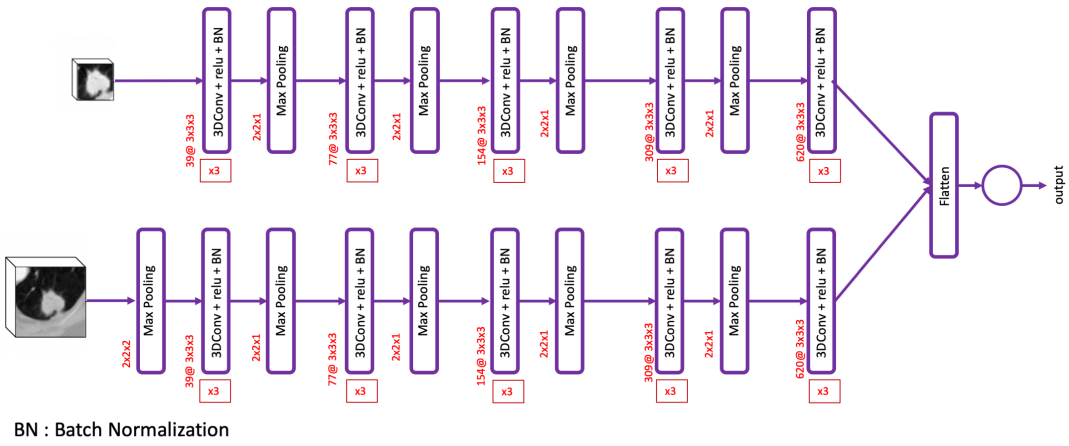


=

شکل ۴: کانولوشن سه بعدی با بلوک‌های dense و طبقه‌بند اضافی

۳ مدل پشینهاد شده

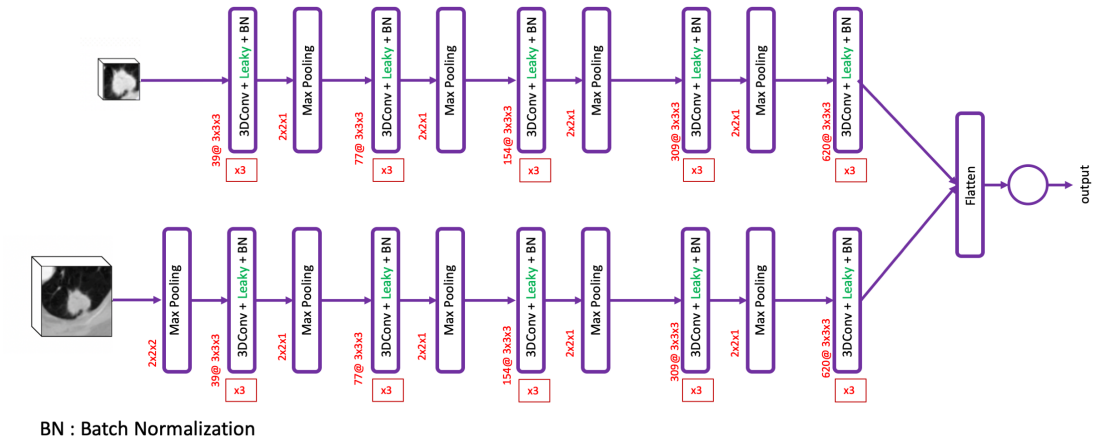
با توجه به تعداد تصاویر محدود CT-scan نمی‌توان مدل را خیلی عمیق در نظر گرفت. از طرف دیگر هر چه عمق شبکه بیشتر باشد، شبکه توانایی بیشتری برای استخراج ویژگی‌های سطح بالاتر دارد. این ویژگی‌های سطح بالا در تفکیک دو کلاس خیلی می‌توانند کمک کننده باشند. در (۲) با شواهد نشان داده شده است که افزایش عمق تا حدی تاثیر دارد و افزایش بیشتر آن خیلی در بازدهی مدل تاثیر ندارد. از طرفی در همین منبع بیان شده است که افزایش عرض شبکه یا تعداد feature-map ها نیز می‌تواند منجر به استخراج ویژگی‌های بهتر شود. همچنین افزایش رزولوشن یا تفکیک پذیری تصویر ورودی نیز می‌تواند به استخراج الگوهای مفید از تصویر کمک شود. روشی در این منبع برای شبکه‌های دو بعدی بیان شده است که از آن به مقیاس گذاری ترکیبی یاد شده است. در این روش به‌جای گسترش شبکه در تنها یکی از موارد عمق، عرض و رزولوشن به مقدار زیاد، بصورت ترکیبی این گسترش انجام می‌گیرد تا با توجه به ظرفیت سیستم، بیشترین مقدار بازدهی را با تعداد پارامتر کمتر داشته باشیم. برای بدست آوردن ضرایب گسترش مقیاس در هر یک از موارد بالا، از روابط موجود در همین منبع استفاده شده و با استفاده از یک جستجوی grid ساده، بهترین مقادیری که در این رابطه صدق می‌کنند بدست آورده می‌شوند. این ضرایب برای یک شبکه دوبعدی با تصاویر طبیعی در عمق برابر با ۱.۲، در عرض برابر با ۱.۱ و در رزولوشن برابر ۱.۱۵است. در کاربرد مورد استفاده ما رزولوشن ورودی تاثیر زیادی ندارد چون ورودی تصاویر سه بعدی با رزولوشن بالا هستند و تنها محدوده اطراف تومور برای ما اهمیت دارد، افزایش رزولوشن کل تصویر می‌تواند برای ما اثر منفی هم داشته باشد. این اثر به دلیل این است که اندازه تومورها از ۳ تا ۳۰ میلی‌متر متغیر است با تغییر رزولوشن تصویر ورودی، عملا اندازه تومور را از دست می‌دهیم و ممکن است اندازه یک تومور بیش از اندازه ورودی تصویر شبکه شود که موجب عدم استخراج ویژگی‌های مناسب از آن می‌گردد و اندازه تومور اثر زیادی در نتیجه طبقه‌بندی دارد. از طرفی اگر محدوده بیشتری از اطراف تومور را استخراج کنیم ناحیه‌های از ورودی را که تاثیری در تصمیم‌گیری ندارند را به شبکه داده‌ایم که منجر به کاهش کیفیت طبقه بندی نهایی ما می‌گردد. به همین دلیل ضریب گسترش رزولوشن را یک در نظر می‌گیریم یعنی رزولوشن ورودی را تغییر نمی‌دهیم. سپس از این ضرایب استفاده کرده و شبکه موجود در شکل ۱ را گسترش می‌دهیم. با انجام این تغییرات روی شبکه با بلوک‌های کانولوشنی معمولی که در قسمت قبل توضیح داده شد، معماری جدید ساخته می‌شود که در شکل ۵ قابل رویت است.



شکل ۵: مدل بهبود یافته با استفاده از ایده مقاله EfficientNet

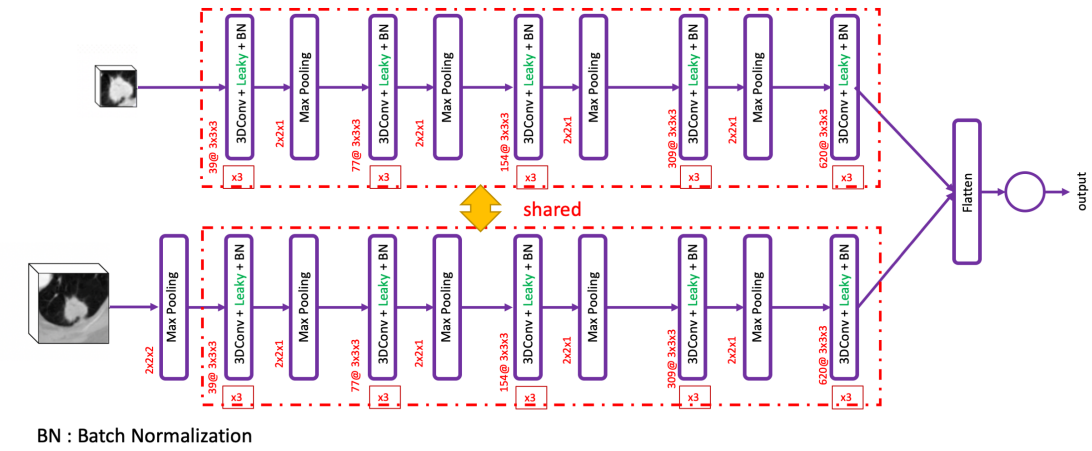
استفاده از تابع فعال‌سازی Relu در مثال قبل موجب به صفر شدن خروجی و گرادیان واحدهایی که خروجی آن‌ها مقدار منفی است، در شبکه می‌گردد. برای حل این مشکل تابع فعالسازی Leaky-Relu پیشنهاد می‌گردد که ضمن حفظ غیرخطی بودن تابع اعمال شده، از صفر شدن خروجی در آن واحد جلوگیری می‌کند. به‌دلیل اینکه تعداد تصاویر ما محدود است، برای بهبود

نتایج بدست آمده در مدل ارائه شده توابع فعالسازی آن را به Leaky-Relu تغییر می‌دهیم و شکل ۶ بیانگر معماری این شبکه است.



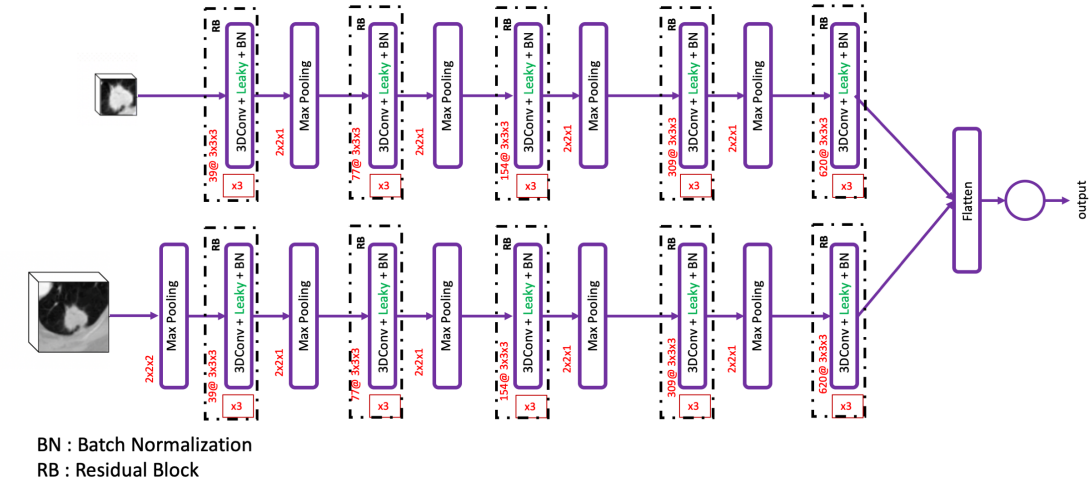
شکل ۶: مدل بهبود یافته با ایده EfficientNet و تابع فعالسازی Leaky-Relu

در مدل‌های قسمت قبل و همچنین مدل‌های ارائه شده، با اینکه شبکه‌های موجود در هر دو مسیر از نظر معماری یکسان هستند (به جز لایه maxpooling اول در مسیر دوم) ولی طبق گفته منبع مذکور، وزن‌های شبکه مشترک نیستند. استفاده از وزن‌های اشتراکی می‌تواند موجب کاهش قابل ملاحظه تعداد پارامترها گردد. به همین دلیل معماری جدیدی که پیشنهاد می‌گردد، از دو شبکه با وزن‌های مشترک در هر دو مسیر استفاده شده است. شکل ۷ بیانگر این موضوع است.



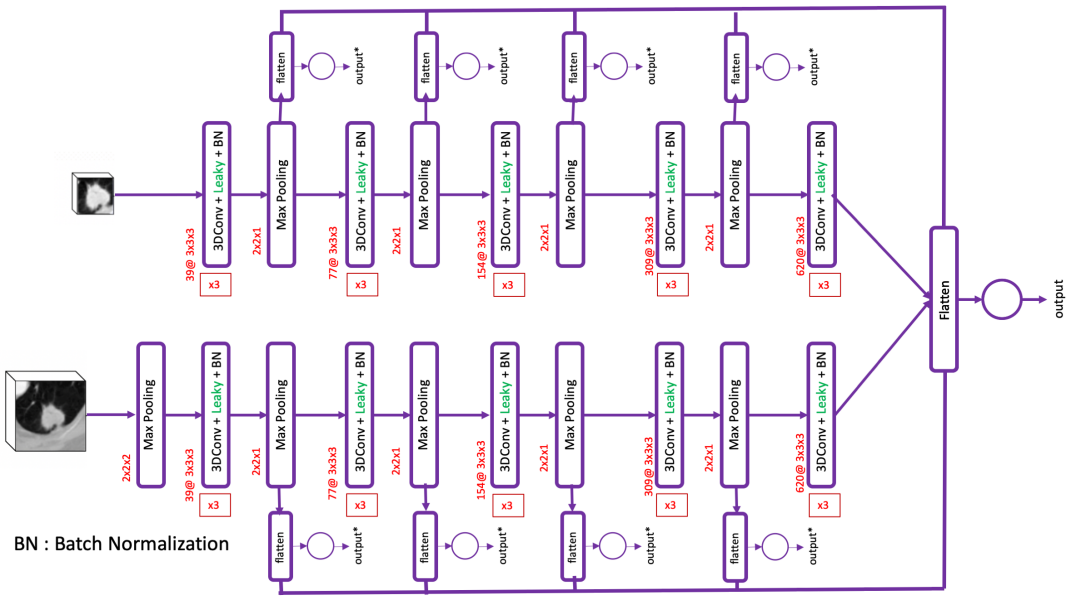
شکل ۷: استفاده از مدل بهبود یافته بصورت اشتراکی

یکی از مشکلات شبکه‌های عمیق، محوشدن گرادیان است. یکی از راه‌های بهبود این مشکل همان‌گونه که در (۳) ذکر شده، استفاده از اتصالات residual در شبکه است. در مدل‌های جدیدی که در این‌جا ارائه کردیم، عمق شبکه‌ها بیشتر شده پس بنابراین احتمال محوشدن گرادیان بیشتر می‌گردد. در اینجا مدل جدیدی معرفی می‌گردد تا این مشکل را حل کند. بلوک‌های ارائه شده در این مدل مانند مدل ارائه شده قبلی است با این تفاوت که اتصالات residual نیز به آن اضافه شده است. شکل ۸ معماری این شبکه را نشان می‌دهد.



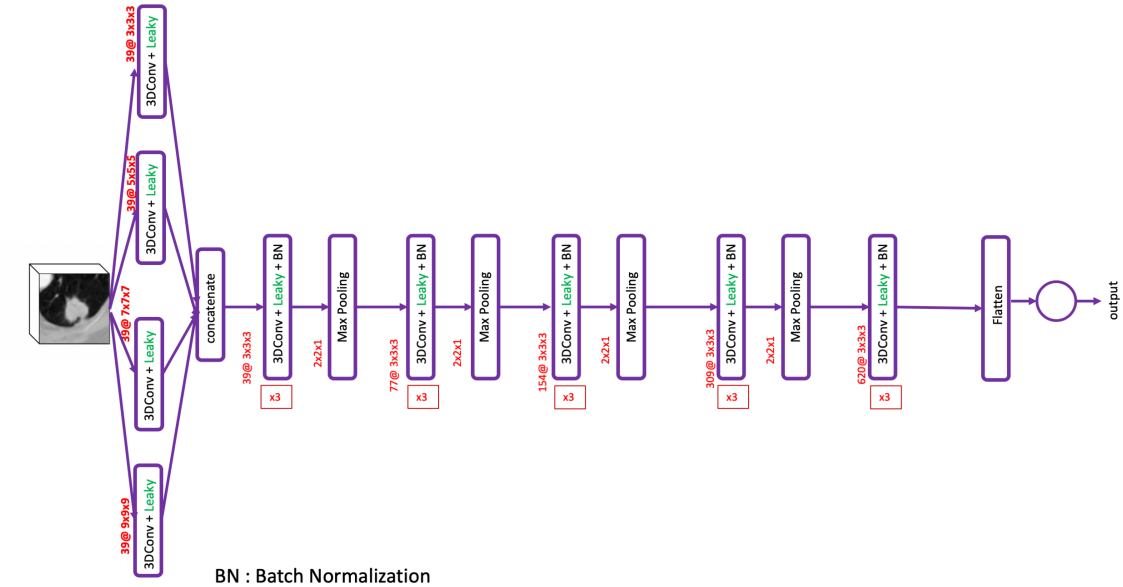
شکل ۸: اعمال ایدهResNet در مدل بهبود یافته

ایده دیگر مطرح شده برای حل مشکل محو شدن گرادیان، استفاده از طبقه‌بندهای کمکی است. همان‌گونه که در مدل‌های پایه توزیع داده شده، این طبقه بندها بعد از هر بلوک آمده و موجب تزریق گرادیان در شبکه می‌گردند. معماری شکل ۹ مربوط به این قسمت است.



شکل ۹: استفاده از طبقه‌بندهای کمکی در مدل پیشنهاد شده

شبکه‌هایی که تا این‌جا ارائه شده اند از دو ورودی استفاده می‌کردند. ایده مطرح شده در(۴) استفاده از چند لایه کانولوشنی با اندازه‌های متفاوت در ابتدای شبکه و الحاق کردن پاسخ‌ها به هم را در شبکه‌های کانولوشن دوبعدی بررسی می‌کند. این تغییر در ابتدای شبکه رخ می‌دهد و بقیه شبکه مانند قبل است. در این‌جا دیگر از دو ورودی استفاده نشده و ایده چند مقیاس بودن از طریق کانولوشن‌ها با اندازه‌های متغیر پیاده‌سازی می‌گردد. این ایده مشابه ایده موجود در (۵) است که با نام inception مطرح شده است. با استفاده از فیلترهای متفاوت، چند مقیاسه بودن را در شبکه اعمال می‌کنیم. شکل ۱۰ مدل پیشنهاد شده در این قسمت است.

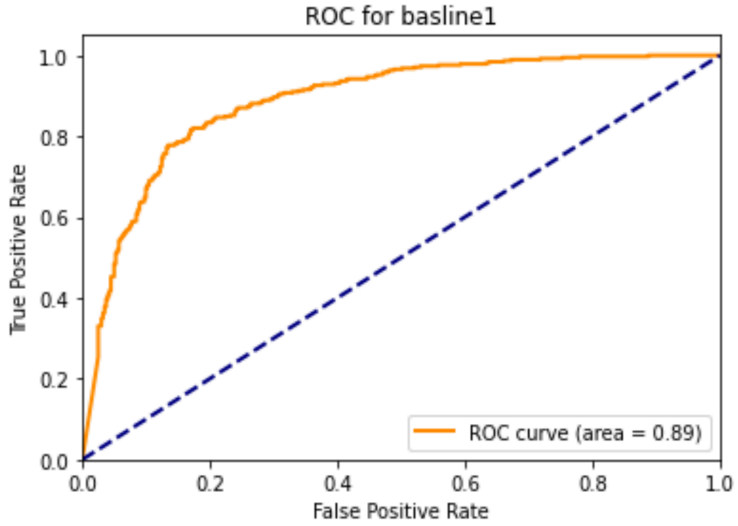


شکل ۱۰: ترکیب مدل بهبود داده شده با ایده multi-convolution

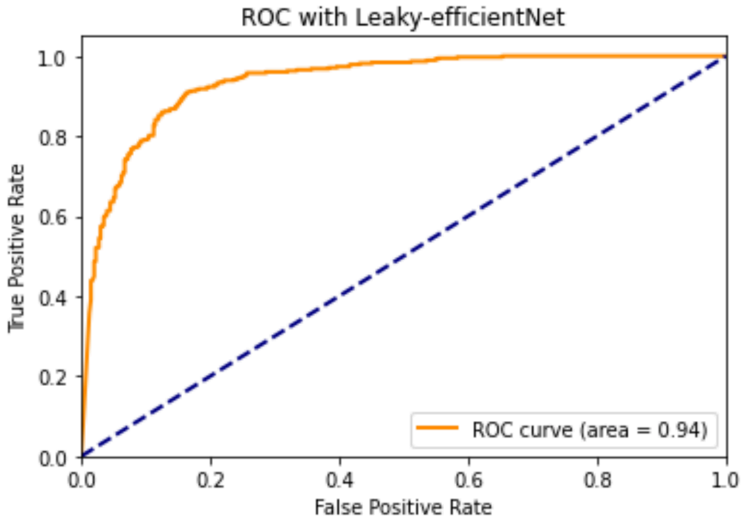
۴ نتایج

نتایج بدست آمده از این آزمایش بصورت براینِدی از ارزیابی متقابل ۵ قسمت است. چون داده ارزیابی مقاله بصورت خصوصی بود، داده اصلی را در ۵ نوبت به قسمت‌های آموزش و ارزیابی تقسیم کردیم و نتایج حاصل برآیندی از عملکرد مدل آموزش دیده شده است.

با توجه به قسمت‌های قبل، ۴ مدل پایه همرا ه با ۶ مدل به منظور بهبود مدل‌های پایه ارائه شدند. خروجی این مدل‌ها در جدول ۱ آمده است. چهار مدل اول مربوط به مدل‌های پایه هستند و شش مدل بعدی، بهبود یافته این مدل‌ها هستند. برای ارزیابی مدل‌ها از معیارهای AUC، TPR، TNR PPV و دقت استفاده می‌گردد. معیار TPR همان نرخ true-positive است ، TNR نیز مربوط به نرخ true-negative است و PPV بیانگر precision می‌باشد. چون مجموعه داده نامتوازن است، از تابع خطا وزن‌دار استفاده شده است. بهترین نمودار ROC مربوط به مدل‌های پایه، متعلق به مدل اول است که نمودار آن در شکل ۱۱ رسم شده است. بهترین نمودار ROC بدست آمده از مدل‌های بهبود یافته شده نیز مربوط به مدل مبتنی بر efficientNet است که در آن از تابع فعالسازی Leaky-Relu استفاده شده است و در شکل ۱۲ رسم شده است.



شکل ۱۱: ROC مربوط به بهترین مدل پایه



شکل ۱۲: ROC مربوط به بهترین مدل بهبود داده شده

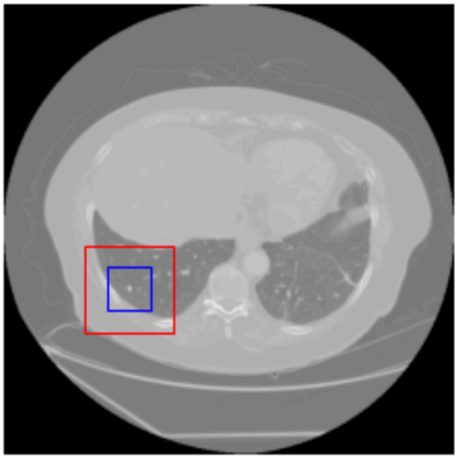
جدول ۱: نتایج به تفکیک مدل

Model	AUC	TPR	TNR	PPV	ACC
3D ConvNet	0.89	0.84	0.79	0.79	0.81
Multiple outputs 3D ConvNet	0.88	0.78	0.87	0.85	0.82
3D ConvNet with Dense Blocks	0.85	0.76	0.79	0.77	0.78
Multiple outputs 3D ConvNet with Dense Blocks	0.83	0.86	0.79	0.79	0.82
3D EfficientNet	0.93	0.86	0.85	0.84	0.85
3D EfficientNet with Leaky Relu	0.94	0.87	0.83	0.83	0.85
shared 3D EfficientNet with Leaky Relu	0.90	0.78	0.82	0.81	0.80
3D EfficientNet with Leaky Relu and Residual Blocks	0.93	0.86	0.84	0.84	0.85
Multiple outputs 3D EfficientNet with Leaky Relu	0.93	0.80	0.81	0.84	0.82
3D Multi-ConvNet	0.93	0.86	0.84	0.80	0.85

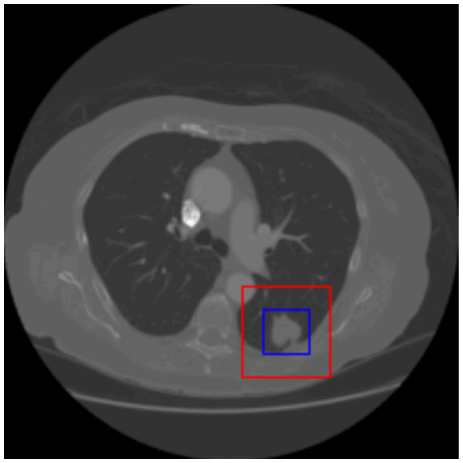
۵ تحلیل

همان‌گونه که از نتایج بدست آمده از جدول ۱ مشخص است، استفاده از ایده مطرح شده در effnet باعث بهبود عمده در مدل‌های ارائه شده است. تمامی مدل‌هایی که از ایده efficientNet استفاده کردند، نتیجه‌گیری به نسبت بهتری بدست آوردند. دلیل این امر همان‌طور که قبل هم گفته شد، این است که تعداد داده‌های پزشکی موجود نسبت به تعداد تصاویر طبیعی خیلی کمتر هستند پس با افزایش عمق به تنهایی نمی‌توان دقت بالایی روی داده تست بدست آورد. به همین منظور با افزایش عمق و عرض شبکه بصورت متناسب پاسخ خوبی را بدست آوردیم. زیاد شدن feature-map ها نیز تاثیر مهمی در این امر داشت که باعث استخراج ویژگی‌های بیشتر و بهتر شد. استفاده از بلوک Residual باعث بهبود TPR شد که این فاکتور بسیار مهمی است زیرا اگر افراد بیمار به اشتباه سالم تشخیص داده شوند، ریسک این مدل به شدت بالا می‌رود و این تشخیص غلط از دیگر تشخیص‌های دیگر خطرناک‌تر است. این بهبود نیز به دلیل غلبه نسبی بر مشکل محوشدن گرادیان است. استفاده از تابع فعالسازی Leaky-Relu نیز بصورت خیلی اندک موجب بهبود مدل شد که دلیل آن می‌تواند زمانی باشد که خروجی تابع فعالسازی منفی شده باشد. در این صورت گرادیان عددی غیر صفر می‌گردد. حسن این روش‌ها نسبت به قبل این است که میانگین کارایی روی همه‌ی معیارها، بیشتر شده است.

استفاده از ایده شبکه با وزن مشترک برای دو ورودی موجب بهبود عملکرد نشد. دلیل این امر می‌تواند این باشد که ورودی کوچک دنبال استخراج ویژگی‌های تومور است در حالی که وظیفه ورودی بزرگتر استخراج ویژگی در مرز تومور و اطراف آن است و این دو وظیفه با هم جمع پذیر نیستند. ایده دیگری که شکست خورد، ایده طبقه‌بند اضافی بود. دلیل این امر نیز می‌تواند زیاد شدن پارامترهای شبکه باشد که موجب بیش برازش رو داده‌های آموزش می‌گردد.



شکل ۱۳: نمونه شناسایی تومور خوش‌خیم



شکل ۱۴: نمونه شناسایی تومور بدخیم

منابع

[1] Y. H. Raunak Dey, Zhongjie Lu, "Diagnostic classification of lung nodules using 3d neural networks," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018.

[2] Q. V. L. Mingxing Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[3] e. a. Kaiming He, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[4] e. a. Guokai Zhang, "Multi-scale pulmonary nodule classification with deep feature fusion via residual network," *Journal of Ambient Intelligence and Humanized Computing*, 2018.

[5] e. a. Christian Szegedy, "Going deeper with convolutions," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.