

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

شبکه SEGNET و نمونه های بهبود یافته آن

گزارش درس شبکه عصبی در رشته‌ی مهندسی کامپیوتر
گرایش هوش مصنوعی

مهدی فیروزبخت

۴۰۰۱۳۱۰۲۷

استاد درس:

دکتر رضا صفابخش

تابستان ۱۴۰۱

دانشگاه امیرکبیر

دپارتمان هوش مصنوعی

چکیده

شبکه عصبی SegNet که توسط الکس کندال ، وی جی بدریناریانان و روبرتو سیپولا که همگی از دانشگاه کمبریج هستند، توسعه یافته است، یک شبکه عصبی پیچشی است که برای برچسب‌گذاری عاقلانه پیکسلی استفاده می‌شود. این مسئله معمولاً تقسیم بندی معنایی نامیده می شود. در تصاویر کامپیوتری، واژه " تقسیم بندی تصویر " به تقسیم تصویر به گروهی از پیکسل براساس برخی معیارها اشاره دارد . یک الگوریتم تقسیم بندی تصویر به عنوان ورودی و خروجی مجموعه‌ای از مناطق را می‌گیرد. شبکه‌هایی که برای تقسیم‌بندی معنایی استفاده می‌شوند، معمولاً یک تصویر RGB را به عنوان داده ورودی می‌گیرند و دارای برچسبی هستند که یک تصویر دارای n کانال است، که در آن n تعداد برچسب‌های درگیر است. هر کانال مربوط به یک برچسب است، به عنوان مثال اتومبیل، جاده، و غیره، و هر پیکسل در یک کانال خاص بسته به اینکه آیا آن پیکسل به برچسب مربوط به آن کانال تعلق دارد، ۱ یا ۰ خواهد بود. SegNet از لایه هایی به نام رمزگذار و رمزگشا تشکیل شده است. به طور معمول رمزگذار یک شبکه از پیش آموزش دیده است . SegNet پارامترهای بسیار کمتری در مقایسه با مدل پیشرفته آن زمان دارد ، از M۱۳۴ پارامتر به M۱۴/۷ پارامتر کاهش یافته است

در این گزارش ، یک بررسی جامع از معماری شبکه Segnet و برخی از نمونه های بهبود یافته آن ارائه داده شده است.

واژه‌های کلیدی : شبکه عصبی SegNet ، شبکه عصبی پیچشی ، تقسیم بندی معنایی ، پیکسل ، رمز گذار ، رمز گشا

۱.....	فصل اول مقدمه
۳.....	فصل دوم : معماری Segnet
۳.....	۱-۲ مقدمه
۳.....	۲-۲ معرفی معماری
۴.....	۳-۲ معماری شبکه
۵.....	۴-۲ آموزش معماری
۶.....	۵-۲ تحلیل و بررسی
۹.....	۶-۲ نتیجه گیری
۱۰.....	فصل سوم معماری P_Segnet & PN_Segnet
۱۰.....	۱-۳ مقدمه
۱۰.....	۲-۳ معرفی موضوع
۱۱.....	۳-۳ ساختار معماری
۱۲.....	۴-۳ آموزش معماری
۱۳.....	۵-۳ آزمایشات صورت گرفته
۱۶.....	۶-۳ نتیجه گیری
۱۷.....	فصل چهارم معماری FL_Segnet
۱۷.....	۱-۴ مقدمه
۱۷.....	۲-۴ معرفی معماری
۱۸.....	۳-۴ آزمایشات صورت گرفته
۲۴.....	۴-۴ نتیجه گیری
۲۵.....	فصل پنجم معماری M_Segnet
۲۵.....	۱-۵ مقدمه
۲۵.....	۲-۵ معرفی موضوع
۲۶.....	۳-۵ معرفی معماری
۲۸.....	۴-۵ آزمایشات صورت گرفته
۳۰.....	۵-۵ نتیجه گیری
۳۱.....	فصل ششم معماری FA_Segnet
۳۱.....	۱-۶ مقدمه
۳۱.....	۲-۶ معرفی معماری
۳۵.....	۳-۶ آزمایشات صورت گرفته
۳۶.....	۴-۶ نتیجه گیری
۳۷.....	منابع و مراجع

صفحه

فهرست اشکال

تصویر ۱-۲ : خروجی معماری Segnet برای تصاویر جاده و تصاویر داخلی.....	۴
تصویر ۲-۲: معماری شبکه Segnet.....	۶
تصویر ۳-۲: مقایسه تصویر جاده ای در Segnet و معماری های دیگر.....	۸
تصویر ۱-۳: روند انجام آزمایشات مقاله	۱۱
تصویر ۲-۳: معماری شبکه NP_Segnet	۱۳
تصویر ۳-۳: مقایسه عملکرد سه شبکه مختلف شناسایی تصاویر ابری.....	۱۴
تصویر ۴-۳: تفاوت بین معماری شبکه کاملاً پیچشی و سه معماری.....	۱۴
تصویر ۵-۳: مقایسه Segnet و NP_Segnet	۱۵
تصویر ۱-۴: ساختار معماری Segnet	۱۸
تصویر ۲-۴: تصویر آزمایشات صورت گرفته در تصاویر تونل	۱۹
تصویر ۳-۴: منحنیهای دقت آموزشی و منحنیهای تلفات سه نوع روش.....	۲۰
تصویر ۴-۴: نتایج تشخیص ارائه شده توسط روشهای مختلف	۲۱
تصویر ۵-۴: نتایج معیارهای معماری های متفاوت	۲۱
تصویر ۶-۴: نتایج تشخیص ارائه شده توسط روشهای مختلف در همپوشانی ترک ها.....	۲۲
تصویر ۷-۴: نتایج معیارهای معماری های متفاوت در همپوشانی ترک ها.....	۲۲
تصویر ۸-۴: نتایج بررسی معماری های متفاوت تحت تأثیر نویز پس زمینه	۲۳
تصویر ۹-۴: نتایج بررسی معماری های متفاوت تحت شرایط نوری ناهموار.....	۲۳
تصویر ۱-۵: معماری کلی شبکه M_Segnet.....	۲۶
تصویر ۲-۵: معماری توجه جهانی	۲۷
تصویر ۳-۵: نتیجه بررسی در مجموعه داده OASIS.....	۲۹
تصویر ۴-۵: نتیجه بررسی در مجموعه داده IBSR.....	۲۹
تصویر ۵-۵: نتیجه بررسی در مجموعه داده OASIS.....	۳۰
تصویر ۶-۵: نتیجه بررسی در مجموعه داده IBSR.....	۳۰
تصویر ۱-۶: واحد توجه کانال فازی	۳۲
تصویر ۲-۶: واحد توجه فضایی فازی	۳۲
تصویر ۳-۶: ساختار معماری FA_Segnet	۳۳
تصویر ۴-۶: نتایج کیفی چهار روش مختلف.....	۳۵

صفحه

فهرست جداول

۶.....	جدول ۱-۲.....
۷.....	جدول ۲-۲.....
۷.....	جدول ۳-۲.....
۹.....	جدول ۲-۴.....
۱۲.....	جدول ۱-۳.....
۱۵.....	جدول ۲-۳.....
۱۹.....	جدول ۱-۴.....
۲۴.....	جدول ۲-۴.....
۳۶.....	جدول ۱-۶.....
۳۶.....	جدول ۲-۶.....

فصل اول

مقدمه

همانطور که میدانیم شبکه های پیچشی^۱ در پردازش تصاویر بسیار پرکاربرد هستند. این شبکه ها با ترکیب ویژگیهای سطح پایینتر در لایه های پایین، ویژگیهای سطح بالاتری بدست میاورند. در پردازش تصویر، تقسیمبندی تصویر فرآیند تقسیمبندی یک تصویر دیجیتال به بخشهای تصویر متعدد است که بهعنوان مناطق تصویر یا اشیاء تصویر نیز شناخته میشوند. هدف تقسیمبندی ساده سازی یا تغییر نمایش تصویر به چیزی است که معنادارتر و تحلیل آن آسان تر باشد. تقسیم بندی تصویر معمولاً برای تعیین موقعیت اشیا و مرزها در تصاویر استفاده می شود. به طور دقیق تر، تقسیمبندی تصویر، فرآیند تخصیص یک برچسب به هر پیکسل در یک تصویر است به طوری که پیکسلهایی با برچسب یکسان ویژگیهای خاصی را به اشتراک می گذارند.

تقسیمبندی معنایی^۲ طیف گسترده ای از کاربردها از درک صحنه^۳، استنباط روابط پشتیبانی بین اشیاء تا رانندگی خودمختار^۴ را شامل می شود. روش های اولیه که به نشانه های بینایی سطح پایین^۵ متکی بودند، به سرعت توسط الگوریتم های یادگیری ماشینی محبوب جایگزین شدند. به ویژه، یادگیری عمیق اخیراً در تشخیص رقم دست نویس، دسته بندی کل تصاویر و تشخیص اشیا در تصاویر، موفقیت چشمگیری داشته است.

در سال ۲۰۱۷ وی جی بدرینارایانان، الکس کندال و روبرتو سیپولا یک معماری جدیدی به نام SegNet معرفی کردند. این معماری طراحی شده است تا یک معماری کارآمد برای تقسیمبندی معنایی پیکسلی باشد. انگیزه اصلی آن برنامه های درک صحنه جاده است که به توانایی مدل سازی ظاهر (جاده، ساختمان)، تصویر (ماشین ها، عابران پیاده) و درک رابطه فضایی (زمینه) بین طبقات مختلف مانند جاده و پیاده روی نیاز دارد. در سال های بعد، افراد دیگری برای زمینه های متفاوتی معماری اصلی را تغییر داده و برای آن را بهبود داده اند. در سال ۲۰۱۹، از این معماری برای تشخیص تصاویر ابرها استفاده شده است. در این مقاله، ۲ معماری با نام های P-Segnet و NP-Segnet ارائه شده است. در مقاله دیگری در سال ۲۰۱۹، معماری با نام FL-Segnet که برای تشکیل آسیبهای ساختارهای تونل معرفی شده است. تشخیص آسیب های ساختاری پوشش تونل برای اطمینان از عملیات ایمن تونل حیاتی است. با این حال، تشخیص نقص به دلیل عدم تعادل اندازه بین ترک ها، پوسته شدن و پس زمینه، کار چالش برانگیزی است. در مقاله دیگری در سال ۲۰۲۱ معماری جدیدی با نام M-Segnet در تقسیمبندی تصاویر MRI مغزی استفاده شده است. در این معماری از مکانیسم توجه جهانی^۶ استفاده شده است.

-
1. Convolutional neural network
 2. Semantic segmentation
 3. Scene understanding
 4. Autonomous driving
 5. Low-level vision cues
 6. Magnetic resonance imaging
 7. Mechanism of global attention

مکانیسم توجه جهانی برای گرفتن اطلاعات متنی غنی در مرحله رمزگشا^۱ با ادغام ویژگی‌های محلی با وابستگی‌های جهانی مربوطه استفاده می‌شود. در مقاله دیگری در سال ۲۰۲۲ معماری جدیدی با نام FA-Segnet برای تقسیم بندی تصاویر MRI قلبی مورد استفاده قرار گرفته است. در این معماری از توجه فازی استفاده شده است. در این معماری یک واحد^۲ توجه کانال فازی^۳ اضافه شده است تا تشخیص مناطق هدف مختلف را تقویت کند.

مقالات مختلفی در زمینه معماری و ساختار این شبکه‌ها مطرح شده است که در این گزارش به مواردی از آنها می‌پردازیم.

ساختار این گزارش در فصل‌های بعد به صورت زیر است:

- فصل دوم: در این فصل معماری پایه و اصلی Segnet را معرفی کرده و آن را بررسی می‌کنیم.
- فصل سوم: در این فصل معماری P-Segnet و NP-Segnet را بررسی می‌کنیم.
- فصل چهارم: در این فصل معماری FL-Segnet را بررسی می‌کنیم.
- فصل پنجم: در این فصل معماری M-Segnet را بررسی می‌کنیم.
- فصل ششم: در این فصل معماری FA-Segnet را بررسی می‌کنیم.

-
1. Decoder
 2. Module
 3. Fuzzy

فصل دوم

معماری Segnet

(Badrinarayanan , Kendall & Cipolla , 2017)

۱-۲ : مقدمه

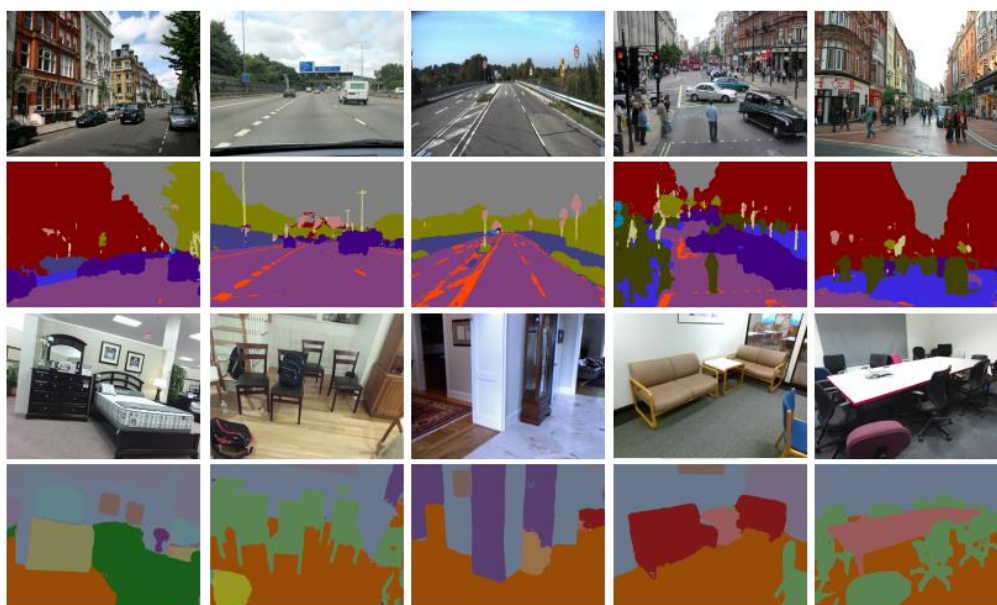
در این مقاله معماری شبکه عصبی کاملاً پیچیده عمیق و کاربردی جدید و کاربردی برای بخش‌بندی پیکسلی معنایی به نام SegNet ارائه شده است. این موتور بخش‌بندی قابل آموزش هسته‌ای از یک شبکه رمزگذار^۱، یک شبکه رمزگشای متناظر و به دنبال یک لایه طبقه‌بندی پیکسلی تشکیل شده است. معماری شبکه رمزگذار از نظر توپولوژیکی^۲ با ۱۳ لایه پیش‌بینی در شبکه ۱۶VGG یکسان است. نقش شبکه رمزگشا این است که نقشه‌های ویژگی رمزگذار با وضوح پایین را به نقشه‌های ویژگی وضوح ورودی کامل برای طبقه‌بندی پیکسلی نگاشت کند. تازگی SegNet در روشی است که در آن رمزگشا از نقشه(های) ویژگی ورودی با وضوح پایین تر نمونه برداری می‌کند. به طور خاص، رمزگشا از شاخص‌های ادغام محاسبه شده در مرحله حداکثر تجمع^۳ رمزگذار مربوطه برای انجام فرا نمونه برداری^۴ غیرخطی استفاده می‌کند. این امر نیاز به یادگیری فرا نمونه برداری را از بین می‌برد. نقشه‌های نمونه‌برداری شده پراکنده هستند و سپس با فیلترهای قابل آموزش برای تولید نقشه‌های ویژگی متراکم در هم می‌پیچند. در این مقاله در نهایت معماری جدید ارائه شده با معماری‌های FCN ، DeepLab-LargeFOV و DeconvNet مقایسه می‌شوند. این معماری به گونه‌ای طراحی شده است تا هم از نظر حافظه و هم از نظر زمان محاسباتی در طول استنتاج کارآمد باشد. همچنین از نظر تعداد پارامترهای قابل آموزش به میزان قابل توجهی نسبت به سایر معماری‌های رقیب کمتر است .

۲-۲ : معرفی معماری

SegNet، طراحی شده است تا یک معماری کارآمد برای تقسیم بندی معنایی پیکسلی باشد. انگیزه اصلی آن برنامه‌های درک صحنه جاده است که به توانایی مدل‌سازی ظاهر (جاده، ساختمان)، تصویر (ماشین‌ها، عابران پیاده) و درک رابطه فضایی (زمینه) بین طبقات مختلف مانند جاده و پیاده‌روی نیاز دارد. در صحنه‌های معمولی جاده، اکثر پیکسل‌ها به کلاس‌های بزرگی مانند جاده، ساختمان تعلق دارند و از این رو شبکه باید تقسیم‌بندی‌های صافی ایجاد کند. موتور همچنین باید توانایی ترسیم اجسام را بر اساس تصویر آنها با وجود اندازه کوچکشان داشته باشد. از این رو حفظ اطلاعات مرزی در نمایش تصویر استخراج شده مهم است. از منظر محاسباتی، لازم است شبکه از نظر حافظه و زمان محاسبات در طول استنتاج کارآمد باشد. توانایی آموزش انتها به انتها به منظور بهینه سازی مشترک همه وزنه‌ها در شبکه با استفاده از تکنیک به روز رسانی وزن کارآمد مانند نزول گرادیان تصادفی (SGD)^۵ یک مزیت اضافی است زیرا به راحتی قابل تکرار است. طراحی SegNet از نیاز به مطابقت با این معیارها ناشی شد. شبکه رمزگذار در SegNet از نظر توپولوژیکی با لایه های پیش‌بینی در ۱۶VGG یکسان است.

-
1. Magnetic resonance imaging
 2. Mechanism of global attention
 3. Decoder
 4. Module
 5. Fuzzy
 6. Stochastic Gradient Descent

جزء کلیدی SegNet شبکه رمزگشا است که از سلسله مراتبی از رمزگشاها تشکیل شده است که هر یک مربوط به هر رمزگذار است. از این میان، رمزگشاهای مناسب از شاخصهای حداکثر تجمع دریافت شده از رمزگذار مربوطه برای انجام فرا نمونه برداری غیرخطی از نقشه های ویژگی ورودی خود استفاده می کنند. استفاده مجدد از شاخصهای حداکثر تجمع در فرآیند رمزگشایی چندین مزیت عملی دارد. (۱) تعیین مرزها را بهبود می بخشد، (۲) تعداد پارامترها را کاهش می دهد که آموزش آنها به انتها را قادر می سازند، و (۳) این تصویر از نمونه برداری را می توان در هر معماری رمزگذار-رمزگشا گنجانند. نمونه ای از نتیجه معماری Segnet به روی تصاویر جاده ای و تصاویر داخلی را در تصویر ۱ مشاهده میکند.



تصویر ۱-۲ : خروجی معماری Segnet برای تصاویر جاده و تصاویر داخلی

۲-۳ : معماری شبکه

SegNet دارای یک شبکه رمزگذار و یک شبکه رمزگشای متناظر است که به دنبال آن یک لایه طبقه بندی پیکسلی نهایی قرار دارد. این معماری در تصویر ۲-۲ نشان داده شده است. شبکه رمزگذار از ۱۳ لایه پیچشی تشکیل شده است که با ۱۳ لایه پیچشی اول در شبکه ۱۶VGG طراحی شده برای طبقه بندی اشیاء مطابقت دارد. بنابراین، می توانیم فرآیند آموزش را از وزنه هایی که برای طبقه بندی در مجموعه های داده بزرگ آموزش داده شده اند، اولیه کنیم. همچنین می توانیم لایه های کاملاً متصل را به نفع حفظ نقشه های ویژگی با وضوح بالاتر در عمیق ترین خروجی رمزگذار کنار بگذاریم. این همچنین تعداد پارامترها را در شبکه رمزگذار SegNet به طور قابل توجهی کاهش می دهد (از 1134×10^6 به 114×10^6). هر لایه رمزگذار دارای یک لایه رمزگشای مربوطه است و از این رو شبکه رمزگشا دارای ۱۳ لایه است. خروجی رمزگشای نهایی به یک طبقه بندی کننده چند کلاسه داده می شود تا احتمالات کلاس برای هر پیکسل به طور مستقل تولید شود.

هر رمزگذار در شبکه رمزگذار، پیچش را با یک بانک فیلتر انجام می دهد تا مجموعه ای از نقشه های ویژگی را تولید کند. سپس اینها به صورت دسته ای نرمال می شوند. سپس یک عنصر غیرخطی اصلاح شده $\max(\text{ReLU})$ اعمال می شود. پس از آن، حداکثر ادغام با یک پنجره 2×2 و گام ۲ (پنجره غیر همپوشانی) انجام می شود و خروجی حاصل با ضرب ۲ نمونه برداری می شود. در تصویر ورودی نمونه برداری فرعی منجر به یک زمینه تصویر ورودی بزرگ (پنجره فضایی) برای هر پیکسل در نقشه ویژگی می شود. در حالی که چندین لایه حداکثر ادغام و فرو نمونه گیری می توانند به عدم تغییر ترجمه بیشتری برای طبقه بندی قوی دست یابند، به همین ترتیب وضوح مکانی

نقشه‌های ویژگی از دست می‌رود. نمایش تصویر با تلفات فزاینده (جزئیات مرزی) برای تقسیم بندی که در آن تعیین مرز حیاتی است، مفید نیست. بنابراین لازم است قبل از انجام نمونه‌برداری فرعی، اطلاعات مرزی در نقشه‌های ویژگی رمزگذار ثبت و ذخیره شود. اگر حافظه در طول استنتاج محدود نباشد، تمام نقشه‌های ویژگی رمزگذار (پس از نمونه برداری فرعی) می‌توانند ذخیره شوند. این معمولاً در کاربردهای عملی صدق نمی‌کند و از این رو در این مقاله روشی کارآمدتر برای ذخیره این اطلاعات پیشنهاد شده است. این شامل ذخیره سازی تنها شاخص‌های حداکثر ادغام است، به عنوان مثال، مکان‌های حداکثر مقدار ویژگی در هر پنجره ادغام برای هر نقشه ویژگی رمزگذار به خاطر سپرده می‌شود. در اصل، این کار را می‌توان با استفاده از 2×2 بیت برای هر پنجره 2×2 ادغام انجام داد و بنابراین ذخیره سازی در مقایسه با به خاطر سپردن نقشه(ها) ویژگی با دقت شناور بسیار کارآمدتر است.

رمزگشای مناسب در شبکه رمزگشا، نقشه(های) ویژگی ورودی خود را با استفاده از شاخص‌های جمع آوری حداکثر از نقشه(های) ویژگی رمزگذار مربوطه نمونه برداری می‌کند. این مرحله نقشه(های) ویژگی پراکنده را تولید می‌کند. این تکنیک رمزگشایی SegNet در تصویر ۳ نشان داده شده است. این نقشه‌های ویژگی سپس با یک بانک فیلتر رمزگشا قابل آموزش برای تولید نقشه‌های ویژگی متراکم در هم می‌آمیزند. سپس یک مرحله عادی سازی دسته ای برای هر یک از این نقشه ها اعمال می‌شود.

توجه داشته باشید که رمزگشای مربوط به اولین رمزگذار (نزدیک ترین به تصویر ورودی) یک نقشه ویژگی چند کاناله تولید می‌کند، اگرچه ورودی رمزگذار آن دارای ۳ کانال (RGB) است. این برخلاف دیگر رمزگشاها در شبکه است که نقشه‌های ویژگی را با تعداد اندازه و کانال به اندازه ورودی رمزگذار خود تولید می‌کنند. نمایش ویژگی ابعادی بالا در خروجی رسیور نهایی به یک طبقه‌بندی کننده نرم‌افزار قابل آموزش داده می‌شود. این نرم افزار حداکثر هر پیکسل را به طور مستقل طبقه بندی می‌کند. خروجی طبقه‌بندی‌کننده یک تصویر کانال K از احتمالات است که در آن K تعداد کلاس‌ها است. تقسیم بندی پیش بینی شده مربوط به کلاس با حداکثر احتمال در هر پیکسل است. رمزگشای مناسب در شبکه رمزگشا، نقشه(های) ویژگی ورودی خود را با استفاده از شاخص‌های جمع آوری حداکثر از نقشه(های) ویژگی رمزگذار مربوطه فرا نمونه برداری می‌کند. این مرحله نقشه(های) ویژگی پراکنده را تولید می‌کند. این نقشه‌های ویژگی سپس با یک بانک فیلتر رمزگشا قابل آموزش برای تولید نقشه‌های ویژگی متراکم در هم می‌آمیزند. سپس یک مرحله عادی سازی دسته ای برای هر یک از این نقشه ها اعمال می‌شود.

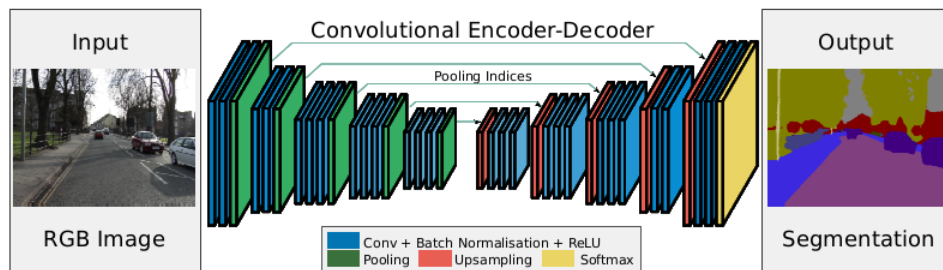
توجه داشته باشید که رمزگشای مربوط به اولین رمزگذار (نزدیک ترین به تصویر ورودی) یک نقشه ویژگی چند کاناله تولید می‌کند، اگرچه ورودی رمزگذار آن دارای ۳ کانال (RGB) است. این برخلاف دیگر رمزگشاها در شبکه است که نقشه‌های ویژگی را با تعداد اندازه و کانال به اندازه ورودی رمزگذار خود تولید می‌کنند. نمایش ویژگی ابعادی بالا در خروجی دریافت کننده نهایی به یک طبقه‌بندی کننده قابل آموزش داده می‌شود. این طبقه‌بندی کننده حداکثر هر پیکسل را به طور مستقل طبقه بندی می‌کند. خروجی طبقه‌بندی‌کننده یک تصویر کانال K از احتمالات است که در آن K تعداد کلاس‌ها است. تقسیم بندی پیش بینی شده مربوط به کلاس با حداکثر احتمال در هر پیکسل است.

۲-۴: آموزش معماری

در این مقاله از مجموعه داده صحنه‌های جاده CamVid برای محک زدن عملکرد انواع رمزگشا استفاده شده است. این مجموعه داده کوچک است و شامل ۳۶۷ آموزش و ۲۳۳ تصویر RGB آزمایشی (صحنه‌های روز و غروب) با وضوح 480×360 است. چالش تقسیم ۱۱ کلاس مانند جاده، ساختمان، اتومبیل‌ها، عابران پیاده، علائم، تیرها، پیاده روی و غیره است. برای آموزش از نزول گرادین تصادفی (SGD) با نرخ یادگیری ثابت 10^{-4} و تکانه 10^{-9} استفاده شده است. متغیرها تا زمانی که ضرر تمرینی همگرا شود، آموزش داده می‌شود. قبل از هر دوره، مجموعه آموزشی به هم ریخته می‌شود و سپس هر دسته کوچک (۱۲ تصویر) به ترتیب انتخاب می‌شود، بنابراین اطمینان حاصل می‌شود که هر تصویر فقط یک بار در یک دوره استفاده می‌شود. در این مقاله از تلفات آنتروپی^۲ متقابل به عنوان تابع هدف برای آموزش شبکه استفاده شده است.

۵-۲: تحلیل و بررسی

برای مقایسه عملکرد کمی انواع مختلف رمزگشا، از سه معیار عملکرد رایج استفاده شده است: دقت جهانی (G) که درصد پیکسل‌هایی را که به درستی در مجموعه داده طبقه‌بندی شده‌اند، اندازه‌گیری می‌کند، دقت متوسط کلاس (C) میانگین دقت پیش‌بینی‌کننده در تمام کلاس‌ها و میانگین تقاطع روی اتحادیه (mIoU) روی همه کلاس‌ها است که در VOC ۱۲ استفاده می‌شود. متریک mIoU معیار دقیق‌تری نسبت به دقت متوسط کلاس است زیرا پیش‌بینی‌های مثبت کاذب را جریمه می‌کند. با این حال، متریک mIoU مستقیماً از طریق از دست دادن متقابل آنتروپی متعادل کلاس بهینه نشده است. متریک mIoU که با نام شاخص Jacard شناخته می‌شود، بیشتر در محک زدن^۳ استفاده می‌شود. ارزیابی دیگری در کنار این معیارها انجام می‌شود. ایده کلیدی در محاسبه امتیاز کانتور معنایی، ارزیابی اندازه‌گیری ۱F است که شامل محاسبه مقادیر دقت و یادآوری بین مرز کلاس حقیقت پیش‌بینی شده و زمینی با توجه به فاصله تحمل پیکسل است.



تصویر ۲-۲: معماری شبکه Segnet

در جدول ۱ نتایج عددی تجزیه و تحلیل خود را گزارش شده است.

Variant	Params (M)	Storage multiplier	Infer time (ms)	Median frequency balancing						Natural frequency balancing							
				Test				Train		Test				Train			
				G	C	mIoU	BF	G	C	mIoU	G	C	mIoU	BF	G	C	mIoU
Fixed upsampling																	
Bilinear-Interpolation	0.625	0	24.2	77.9	61.1	43.3	20.83	89.1	90.2	82.7	82.7	52.5	43.8	23.08	93.5	74.1	59.9
Upsampling using max-pooling indices																	
SegNet-Basic	1.425	1	52.6	82.7	62.0	47.7	35.78	94.7	96.2	92.7	84.0	54.6	46.3	36.67	96.1	83.9	73.3
SegNet-Basic-EncoderAddition	1.425	64	53.0	83.4	63.6	48.5	35.92	94.3	95.8	92.0	84.2	56.5	47.7	36.27	95.3	80.9	68.9
SegNet-Basic-SingleChannelDecoder	0.625	1	33.1	81.2	60.7	46.1	31.62	93.2	94.8	90.3	83.5	53.9	45.2	32.45	92.6	68.4	52.8
Learning to upsample (bilinear initialisation)																	
FCN-Basic	0.65	11	24.2	81.7	62.4	47.3	38.11	92.8	93.6	88.1	83.9	55.6	45.0	37.33	92.0	66.8	50.7
FCN-Basic-NoAddition	0.65	n/a	23.8	80.5	58.6	44.1	31.96	92.5	93.0	87.2	82.3	53.9	44.2	29.43	93.1	72.8	57.6
FCN-Basic-NoDimReduction	1.625	64	44.8	84.1	63.4	50.1	37.37	95.1	96.5	93.2	83.5	57.3	47.0	37.13	97.2	91.7	84.8
FCN-Basic-NoAddition-NoDimReduction	1.625	0	43.9	80.5	61.6	45.9	30.47	92.5	94.6	89.9	83.7	54.8	45.5	33.17	95.0	80.2	67.8

جدول ۱

از جدول ۱، می‌بینیم که نمونه برداری مبتنی بر درونیایی دوخطی بدون هیچ یادگیری، بدترین عملکرد را بر اساس تمام معیارهای دقت دارد. همه روش‌های دیگری که یا از یادگیری برای نمونه‌برداری استفاده می‌کنند (FCN-Basic و انواع مختلف) یا یادگیری فیلترهای رمزگشا پس از نمونه‌برداری (SegNet-Basic و انواع آن) به طور قابل توجهی بهتر عمل می‌کنند. این امر بر نیاز به یادگیری رمزگشاها برای تقسیم بندی تأکید می‌کند. این همچنین توسط شواهد تجربی جمع‌آوری شده توسط سایر نویسندگان هنگام مقایسه شبکه کاملاً پیچشی با تکنیک‌های رمزگشایی SegNet پشتیبانی می‌شود.

1. Magnetic resonance imaging
2. Mechanism of global attention
3. Decoder

هنگامی که SegNet-Basic و FCN-Basic را با هم مقایسه می‌کنیم، می‌بینیم که هر دو در این تست در تمام معیارهای دقت عملکرد یکسانی دارند. تفاوت این است که SegNet از حافظه کمتری در حین استنتاج استفاده می‌کند زیرا فقط شاخص‌های حداکثر ادغام را ذخیره می‌کند. از سوی دیگر FCN-Basic رمزگذار ذخیره‌سازی نقشه‌ها را به طور کامل نشان می‌دهد که حافظه بسیار بیشتری (۱۱ برابر بیشتر) مصرف می‌کند. SegNet-Basic دارای یک رمزگشا با ۶۴ نقشه ویژگی در هر لایه رمزگشا است. در مقایسه FCN-Basic، که از کاهش ابعاد استفاده می‌کند، نقشه‌های ویژگی کمتری (۱۱) در هر لایه رمزگشا دارد. این باعث کاهش تعداد پیچشی‌ها در شبکه رمزگشا می‌شود و از این رو FCN-Basic در طول استنتاج (گذر به جلو) سریعتر است. از منظر دیگر، شبکه رمزگشا در SegNet-Basic آن را به طور کلی شبکه بزرگتر از FCN-Basic می‌کند. این به آن انعطاف پذیری بیشتری می‌بخشد و از این رو به دقت آموزشی بالاتری نسبت به FCN-Basic برای همان تعداد تکرار می‌رسد. به طور کلی می‌بینیم که SegNet-Basic نسبت به FCN-Basic زمانی که حافظه زمان استنتاج محدود است، اما زمانی که زمان استنتاج می‌تواند تا حدی به خطر بیفتد، مزیت دارد.

نتایج جدول ۲-۲ نشان می‌دهد که SegNet، SegNet-Basic در مقایسه با روش‌هایی که از CRF استفاده می‌کنند، نتایج رقابتی به دست می‌آورند. این توانایی معماری عمیق را برای استخراج ویژگی‌های معنی‌دار از تصویر ورودی و نگاشت آن به برجسب‌های بخش کلاس دقیق و صاف نشان می‌دهد.

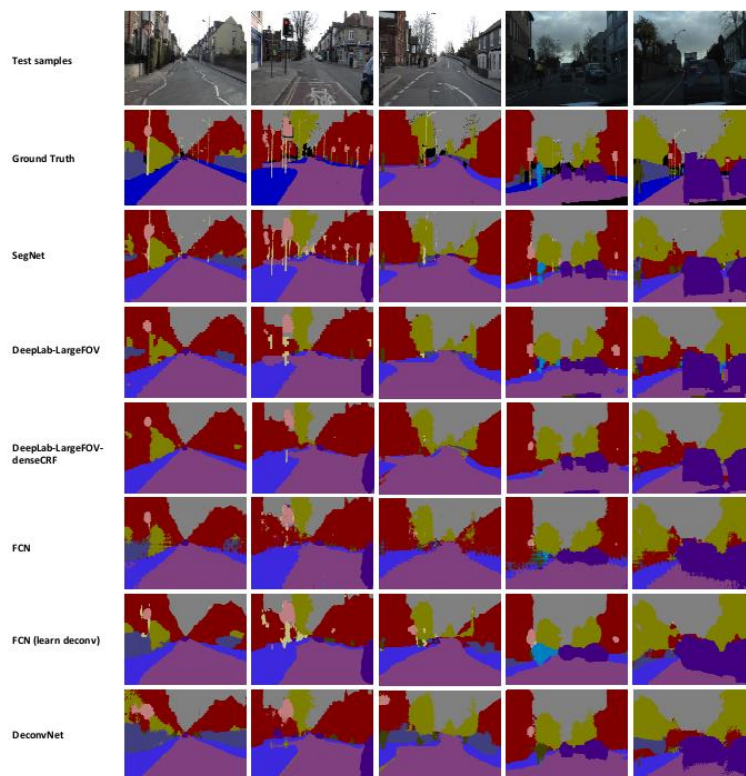
در جدول ۳-۲ عملکرد SegNet را با معماری‌های کاملاً پیچشی به‌طور گسترده برای تقسیم‌بندی مقایسه شده است. در مقایسه با آزمایش جدول ۲، برای آموزش هیچ یک از معماری‌های عمیق از جمله SegNet از هیچ کلاسی استفاده نشده است. این به این دلیل است که آموزش مدل‌های بزرگ‌تر مانند DeconvNet با متعادل‌سازی فرکانس میانه دشوار بود. عملکرد را در $K=40$ ، $K=80$ و بیش از $K=100$ تکرار کرده که با توجه به اندازه کوچک دسته‌ای و اندازه مجموعه آموزشی تقریباً با دوره‌های ۵۰، ۱۰۰ و بیش از ۱۰۰ مطابقت دارد. از جدول ۳-۲ بلافاصله می‌بینیم که SegNet، DeconvNet بالاترین امتیاز را در تمام معیارها در مقایسه با سایر مدل‌ها به دست می‌آورند. DeconvNet دقت ترسیم مرز بالاتری دارد اما SegNet در مقایسه با DeconvNet بسیار کارآمدتر است.

Method	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Side-walk	Bicyclist	Class avg.	Global avg.	mIoU	BF
SfM+Appearance [28]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53.0	69.1	n/a*	
Boosting [29]	61.9	67.3	91.1	71.1	58.5	92.9	49.5	37.6	25.8	77.8	24.7	59.8	76.4	n/a*	
Dense Depth Maps [32]	85.3	57.3	95.4	69.2	46.5	98.5	23.8	44.3	22.0	38.1	28.7	55.4	82.1	n/a*	
Structured Random Forests [31]					n/a							51.4	72.5	n/a*	
Neural Decision Forests [64]					n/a							56.1	82.1	n/a*	
Local Label Descriptors [65]	80.7	61.5	88.8	16.4	n/a	98.0	1.09	0.05	4.13	12.4	0.07	36.3	73.6	n/a*	
Super Parsing [33]	87.0	67.1	96.9	62.7	30.1	95.9	14.7	17.9	1.7	70.0	19.4	51.2	83.3	n/a*	
SegNet (3.5K dataset training - 140K)	89.6	83.4	96.1	87.7	52.7	96.4	62.2	53.45	32.1	93.3	36.5	71.20	90.40	60.10	46.84
CRF based approaches															
Boosting + pairwise CRF [29]	70.7	70.8	94.7	74.4	55.9	94.1	45.7	37.2	13.0	79.3	23.1	59.9	79.8	n/a*	
Boosting+Higher order [29]	84.5	72.6	97.5	72.7	34.1	95.3	34.2	45.7	8.1	77.6	28.5	59.2	83.8	n/a*	
Boosting+Detectors+CRF [30]	81.5	76.6	96.2	78.7	40.2	93.9	43.0	47.6	14.3	81.5	33.9	62.5	83.8	n/a*	

جدول ۲-۲

Network/Iterations	40K				80K				>80K				Max iter
	G	C	mIoU	BF	G	C	mIoU	BF	G	C	mIoU	BF	
SegNet	88.81	59.93	50.02	35.78	89.68	69.82	57.18	42.08	90.40	71.20	60.10	46.84	140K
DeepLab-LargeFOV [3]	85.95	60.41	50.18	26.25	87.76	62.57	53.34	32.04	88.20	62.53	53.88	32.77	140K
DeepLab-LargeFOV-denseCRF [3]	not computed								89.71	60.67	54.74	40.79	140K
FCN	81.97	54.38	46.59	22.86	82.71	56.22	47.95	24.76	83.27	59.56	49.83	27.99	200K
FCN (learned deconv) [2]	83.21	56.05	48.68	27.40	83.71	59.64	50.80	31.01	83.14	64.21	51.96	33.18	160K
DeconvNet [4]	85.26	46.40	39.69	27.36	85.19	54.08	43.74	29.33	89.58	70.24	59.77	52.23	260K

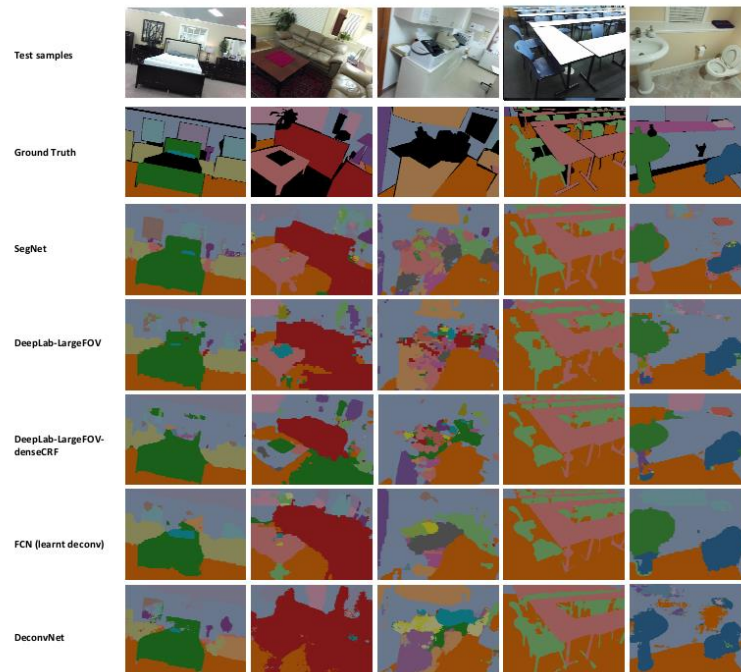
جدول ۳-۲



تصویر ۲-۳ : مقایسه تصویر جاده ای در Segnet و معماری های دیگر

تصاویر صحنه جاده تنوع محدودی دارند، هم از نظر طبقات مورد علاقه و هم از نظر ترتیبات فضایی آنها. هنگامی که از یک وسیله نقلیه در حال حرکت گرفته می شود که موقعیت دوربین تقریباً همیشه موازی سطح جاده است و تنوع در نقاط دید را محدود می کند. این امر باعث می شود شبکه های عمیق یاد بگیرند که آنها را به طور قوی تقسیم کنند. در مقایسه، تصاویر صحنه های داخلی پیچیده تر هستند، زیرا نقاط دید می توانند بسیار متفاوت باشند و نظم کمتری در تعداد کلاس های موجود در یک صحنه و آرایش فضایی آنها وجود دارد. متصویر دیگر به دلیل اندازه های بسیار متفاوت کلاس های شی در صحنه است. نتایج کیفی SegNet بر روی نمونه هایی از صحنه های داخلی از انواع مختلف مانند اتاق خواب، اتاق نشیمن، آزمایشگاه، اتاق جلسه، حمام در تصویر ۲-۴ نشان داده شده است.

نتایج کمی در جدول ۲-۴ نشان می دهد که تمام معماری های عمیق، معیارهای mIoU و مرزی پایین را به اشتراک می گذارند. میانگین جهانی و کلاس (همبستگی خوبی با mIoU) نیز کوچک است. SegNet از همه روش های دیگر از نظر معیارهای G، C، BF بهتر عمل می کند و نسبت به DeepLab-LargeFOV، mIoU کمی کمتر دارد. یکی از دلایل عملکرد ضعیف کلی، تعداد زیاد کلاس ها در این کار تقسیم بندی است، که بسیاری از آنها بخش کوچکی از تصویر را اشغال می کنند و به ندرت ظاهر می شوند. یکی دیگر از دلایل عملکرد ضعیف می تواند در عدم توانایی این معماری های عمیق (همه بر اساس معماری VGG هستند) در تنوع زیاد در صحنه های داخلی باشد.



تصویر ۴-۲ : مقایسه تصویر داخلی در Segnet و معماری های دیگر

Network/Iterations	40K				80K				>80K				Max iter
	G	C	mIoU	BF	G	C	mIoU	BF	G	C	mIoU	BF	
SegNet	88.81	59.93	50.02	35.78	89.68	69.82	57.18	42.08	90.40	71.20	60.10	46.84	140K
DeepLab-LargeFOV [3]	85.95	60.41	50.18	26.25	87.76	62.57	53.34	32.04	88.20	62.53	53.88	32.77	140K
DeepLab-LargeFOV-denseCRF [3]	not computed								89.71	60.67	54.74	40.79	140K
FCN	81.97	54.38	46.59	22.86	82.71	56.22	47.95	24.76	83.27	59.56	49.83	27.99	200K
FCN (learnt deconv) [2]	83.21	56.05	48.68	27.40	83.71	59.64	50.80	31.01	83.14	64.21	51.96	33.18	160K
DeconvNet [4]	85.26	46.40	39.69	27.36	85.19	54.08	43.74	29.33	89.58	70.24	59.77	52.23	260K

جدول ۴-۲

۶-۲ : نتیجه گیری

در این مقاله SegNet را ارائه شده است ، یک معماری شبکه پیچیده عمیق برای بخش‌بندی معنایی. انگیزه اصلی SegNet نیاز به طراحی یک معماری کارآمد برای درک صحنه جاده و داخل ساختمان بود که هم از نظر حافظه و هم از نظر زمان محاسباتی کارآمد باشد. همچنین SegNet تجزیه و تحلیل شده و آن را با سایر انواع مهم مقایسه شده است تا مبادلات عملی مربوط به طراحی معماری برای بخش‌بندی، به ویژه زمان آموزش، حافظه در مقابل دقت آشکار شود. معماری‌هایی که نقشه‌های ویژگی شبکه رمزگذار را به طور کامل ذخیره می‌کنند بهترین عملکرد را دارند اما در طول زمان استنتاج حافظه بیشتری مصرف می‌کنند. از طرف دیگر SegNet کارآمدتر است زیرا فقط شاخص‌های حداکثر تجمع نقشه‌های ویژگی را ذخیره می‌کند و از آنها در شبکه رمزگشای خود برای دستیابی به عملکرد خوب استفاده می‌کند. در مجموعه داده‌های بزرگ و شناخته شده SegNet به صورت رقابتی عمل می‌کند و به نمرات بالایی برای درک صحنه جاده می‌رسد.

فصل سوم

معماری P_Segnet and NP_Segnet

(Lu , Wang , Zhu &... , 2019)

۱-۳ : مقدمه

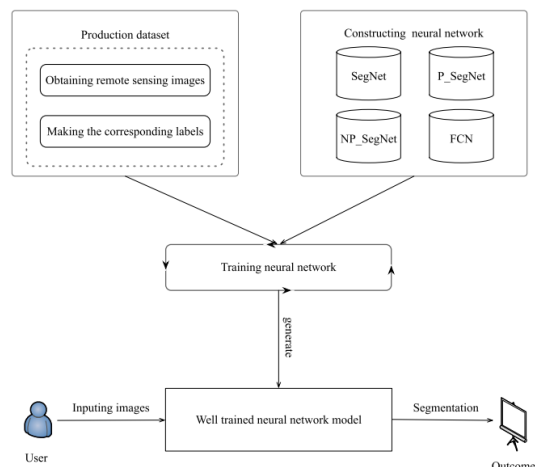
در سال های اخیر، تصاویر سنجش از دور نقش مهمی در پایش محیطی، پایش بینی آب و هوا و برنامه ریزی کشاورزی ایفا کرده اند. با این حال، تصاویر سنجش از راه دور اغلب حاوی تعداد زیادی لایه ابر هستند. این ابرها می توانند حجم زیادی از اطلاعات سطحی را پوشش دهند. بنابراین، تعداد فزاینده ای از روش های تشخیص ابر برای کاهش تاثیر پوشش ابر پیشنهاد شده است. در روش های سنتی تشخیص ابر متصویرات زیادی وجود دارد. به عنوان مثال، روش آستانه مبتنی بر ویژگی های طیفی، دقت تشخیص ابر را بهبود می بخشد، اما اغلب منجر به حذف یا قضاوت نادرست در تشخیص ابر می شود و به دانش قبلی بستگی دارد. برای بهبود دقت و کارایی تشخیص ابر، از یادگیری عمیق برای رسیدگی به متصویرات تشخیص ابر در تصاویر سنجش از دور استفاده می کنیم. در این مقاله مجموعه ای از روش ها را از کسب و تولید مجموعه داده های آموزشی تا آموزش شبکه های عصبی و برنامه های کاربردی تشخیص ابر را پیشنهاد شده است. این مقاله تحقق شناخت ابری تصاویر سنجش از دور^۱ بر اساس معماری SegNet را توصیف می کند. در این مقاله دو معماری به نام های P_SegNet و NP_SegNet را پیشنهاد شده است که از SegNet اصلاح شده اند. برخی از ساختارهای موازی نیز در معماری SegNet به کار گرفته شد تا دقت تشخیص ابر را بهبود بخشد. با توجه به این تغییرات، این مقاله همچنین تاثیر ساختار شبکه تقارن بر دقت نهایی را مورد بحث قرار می دهد. روش پیشنهادی با رویکرد شبکه عصبی کاملاً پیچیده (FCNN) مقایسه شده است. نتایج امکان سنجی و کاربردی بودن استفاده از رویکرد یادگیری عمیق را برای تشخیص ابر در تصاویر سنجش از دور نشان داده است.

۲-۳ : معرفی موضوع

رویکردهای سنتی تشخیص ابر، مانند روش آستانه^۲ و تحلیل بافت^۱، با مسائل عمده زیر روبرو هستند: (۱) کارایی تشخیص ابر کم است. (۲) این روش ها صرفاً بین ابرها و برجسته کردن اشیاء تمایز قائل می شوند. و (۳) برخی از روش ها بیش از حد به تجربه انسانی متکی هستند که منجر به ذهنیت می شود. برای رسیدگی به این متصویرات، برخی از محققان شروع به استفاده از الگوریتم های هوشمند می کنند. پیشرفت های سریع اخیر در هوش مصنوعی به محققان این امکان را داده است که از یادگیری عمیق برای تشخیص ابر در تصاویر سنجش از دور استفاده کنند. برای مثال، اقدام و هراوی یک الگوریتم طبقه بندی تصویر طبیعی را بر اساس شبکه عصبی پیچیده عمیق پیشنهاد کردند که به دقت طبقه بندی بالایی در مجموعه داده ImageNet دست یافت. تصویر ۱-۳ روند اصلی آزمایش را نشان می دهد. ابتدا مجموعه داده های آموزشی و آزمایشی تولید شده که هر کدام شامل دو مرحله است. اولین گام بدست آوردن داده های اصلی، تصاویر سنجش از راه دور است. مرحله دوم ساخت برچسب های مربوطه است که زمان بر اما حیاتی است. بعد، تصمیم گرفته شده که از کدام معماری شبکه عصبی استفاده شود. سپس این شبکه های عصبی را با مجموعه داده های آموزشی از قبل آماده شده آموزش داده شده است. در نهایت، از مدل های شبکه عصبی آموزش دیده برای بخش بندی مجموعه داده آزمایشی استفاده شد.

1. Remote Sensing

2. Threshold method



تصویر ۳-۱ : روند انجام آزمایشات مقاله

هدف اصلی استفاده از یادگیری عمیق، بهبود دقت و سرعت پردازش تشخیص ابر در تصاویر سنجش از راه دور است. مدل شبکه عصبی موجود، SegNet، مبنایی برای ارزیابی عملکرد پردازش تشخیص ابر ارائه می‌دهد. با اصلاح معماری آن، می‌توان عواملی را که بر دقت تشخیص ابر و میزان تبعیض اشیاء بسیار مشابه تأثیر می‌گذارند، بررسی کرد. در تشخیص ابر، اشیاء بسیار مشابه شامل برف، بزرگراه‌ها، اشیاء برجسته و غیره است. برای نشان دادن کارایی تشخیص ابر با این معماری، تقریباً ۱۰۰۰ تصویر سنجش از دور را جمع‌آوری کرده و به صورت دستی آنها برچسب گذاری شده‌اند. برچسب‌ها به سه دسته تقسیم شدند: ابر، برف و پس زمینه.

۳-۳ : ساختار معماری

معماری SegNet داده‌های ورودی را از طریق تعداد زیادی از لایه‌های پیچشی پردازش می‌کند و ساختار شبکه عصبی کل فرآیند پیچشی - ادغام - نمونه برداری - دکانولوشن یک ساختار متقارن است. به دلیل وجود لایه‌های نمونه‌برداری، اندازه تصویر ورودی با تنظیم دستی طول و عرض شبکه عصبی SegNet محدود می‌شود.

به منظور بهبود دقت تشخیص ابر، ساختار SegNet اصلاح شد. همانطور که در تصویر ۳-۲ نشان داده شده است، در این مقاله مدل اولیه GoogLeNet با استفاده از یک بانک فیلتر چند مقیاسی برای تولید مجموعه‌ای از نقشه‌های ویژگی در یک لایه خاص از شبکه تقلید شده است.

P_SegNet یک نسخه از SegNet است. ساختار آن تقریباً مشابه SegNet است. شبکه رمزگذار P_SegNet از ۱۳ لایه پیچشی تشکیل شده است، مانند Seg-Net. در این مقاله سعی شده است با گسترش عرض شبکه عصبی عملکرد معماری را بهبود بخشیده شود. چندین هسته پیچشی با اندازه‌های مختلف می‌توانند انعطاف‌پذیری یک شبکه را افزایش دهند. در این مدل، اولین لایه‌های پیچشی در شبکه رمزگذار دارای دو اندازه فیلتر 3×3 و 5×5 هستند. همه آنها پیچیدگی یکسانی دارند که به این معنی است که نقشه‌های ویژگی که آنها تولید می‌کنند می‌توانند مستقیماً توسط لایه اتصال به هم متصل شوند. استفاده از اندازه‌های مختلف هسته‌های پیچشی به معنای اندازه‌های مختلف زمینه‌های پذیرنده است، و اتصال نهایی به معنای ادغام ویژگی‌های مقیاس مختلف است. برای حل متصور پارامترهای اضافی که باعث بزرگتر شدن مدل و کاهش بازدهی می‌شود، از واحد اولیه با کاهش ابعاد استفاده می‌کنیم. با استفاده از 1×1 لایه‌های پیچشی، تعداد پارامترها و بیش از حد اتصالات کاهش می‌یابد. ما همچنین لایه نرمال‌سازی دسته‌ای را در پشت هر لایه پیچشی اضافه کردیم. نرمال‌سازی دسته‌ای (BN) یک روش منظم‌سازی بسیار موثر است. این می‌تواند آموزش شبکه‌های پیچشی بزرگ را سرعت بخشد و دقت طبقه بندی پس از همگرایی را می‌توان تا حد زیادی بهبود بخشید. کاهش کانال‌ها در هر نقشه ویژگی انتظار نمی‌رود.

رفت، بنابراین نقشه های ویژگی دوخته شده دو برابر بزرگتر از نقشه های ویژگی تولید شده توسط لایه های مربوطه از شبکه SegNet بود. مدل های آغازین تنها بر روی دو لایه پیچشی اول اضافه شدند. اضافه کردن آنها به تمام لایه ها بار زیادی را به لایه موازی اضافه می کرد. ساختارهای موازی اولیه فقط در شبکه رمزگذار P_SegNet اضافه شد. در نتیجه، ما همچنین NP_SegNet را طراحی کردیم که ساختار موازی دارای شبکه رمزگذار و رمزگشا را برای کشف تأثیر تقارن شبکه عصبی بر دقت تشخیص ابر اضافه کرده است.

جدول ۱-۳ پارامترهای هر لایه از شبکه عصبی را نشان می دهد "maxpool" نشان دهنده استفاده از بزرگترین لایه جمع آوری است، اندازه آن ۲، اندازه مرحله آن ۲ است. "فرا نمونه گیری" نشان دهنده لایه فرا نمونه گیری است، اندازه هسته آن ۲ است، طول و عرض داده ها به دو برابر بزرگتر از قبل بازیابی می شوند. ستون "num" نشان دهنده چند بار تکرار هر لایه است. در ساختار موازی، R به این معنی است که قبل از عملیات پیچشی، یک لایه پیچشی ۱×۱ وجود دارد تا تعداد پارامترها کاهش یابد.

Configuration			
num	SegNet	P_SegNet	NP_SegNet
input(480 × 360 RGB image)			
2	conv3-64 pad 1	R conv3-32 pad 1 R conv5-32 pad 2	R conv3-32 pad 1 R conv5-32 pad 2
maxpool			
2	conv3-128 pad 1	R conv3-64 pad 1 R conv5 64 pad 2	R conv3-64 pad 1 R conv5 64 pad 2
maxpool			
3	conv3-256 pad 1	conv3-256 pad 1	conv3-256 pad 1
maxpool			
3	conv3-512 pad 1	conv3-512 pad 1	conv3-512 pad 1
maxpool			
3	conv3-512 pad 1	conv3-512 pad 1	conv3-512 pad 1
3	deconv3-512 pad 1	deconv3-512 pad 1	deconv3-512 pad 1
upsample			
3	deconv3-512 pad 1	deconv3-512 pad 1	deconv3-512 pad 1
upsample			
3	deconv3-256 pad 1	deconv3-256 pad 1	deconv3-256 pad 1
upsample			
2	deconv3-128 pad 1	deconv3-128 pad 1	R deconv3-64 pad 1 R deconv5-64 pad 2
upsample			
2	deconv3-64 pad 1	deconv3-64 pad 1	R deconv3-32 pad 1 R deconv5-32 pad 2
SoftmaxWithLoss			

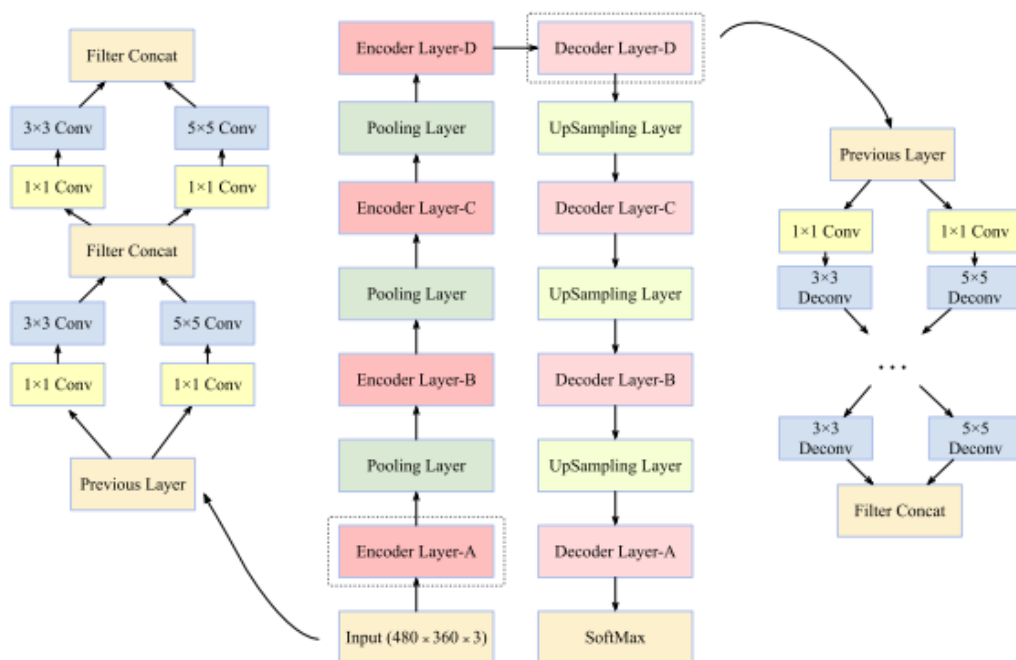
جدول ۱-۳

P_SegNet ساختار موازی را فقط در لایه های رمزگذار اضافه کرده است. NP_SegNet ساختار موازی را در هر دو لایه رمزگذار و رمزگشا اضافه کرده است. ساختار موازی دارای دو اندازه مختلف هسته پیچشی است که ۳×۳ و ۵×۵ هستند. برای جلوگیری از بزرگ بودن شبکه عصبی، ساختارهای موازی فقط در چند لایه اول رمزگذار و رمزگشا اضافه شدند. تصویر ۲-۳ نشاندهنده معماری شبکه NP-Segnet میباشد.

۴-۳: آموزش معماری

مجموعه داده مورد استفاده برای آموزش شبکه عصبی نسبتاً کوچک بود. این شامل ۶۷۰ تصویر آموزشی و ۳۴۲ تصویر RGB آزمایشی و برچسب های تصویر مربوط به آنها بود. برای این مجموعه داده، انتظار می رفت که شبکه عصبی سه کلاس را تقسیم کند: ابر، برف و پس زمینه. از دست دادن آنتروپی متقاطع به عنوان تابع هدف در شبکه استفاده شد. در فرآیند آموزش کل مجموعه داده به طور مستقیم، دقت تقسیم بندی ابر و برف به مقدار بسیار پایینی نزدیک به صفر رسید و مقدار تلفات مربوطه بدون توجه به اینکه چند دور تمرین انجام شد، بدون تغییر باقی ماند. برعکس، با استفاده از یک مجموعه آموزشی از تنها ۱۰ تصویر، شبکه عصبی همچنان می تواند به سرعت با نرخ تشخیص بسیار بالایی برای ابرها و برف همگرا شود. به منظور جلوگیری از عدم همگرایی، ۵۰ تصویر بسیار

نماینده در مجموعه داده را برای تشکیل یک زیر مجموعه انتخاب شده است. از زیر مجموعه برای آموزش شبکه عصبی با نرخ یادگیری ۰/۱ در طول پیش‌آموزش استفاده شده است. پس از ۱۰۰۰ تکرار، نرخ یادگیری را به ۰/۰۱ کاهش داده شده و شبکه عصبی را با کل مجموعه داده آموزش داده‌اند.

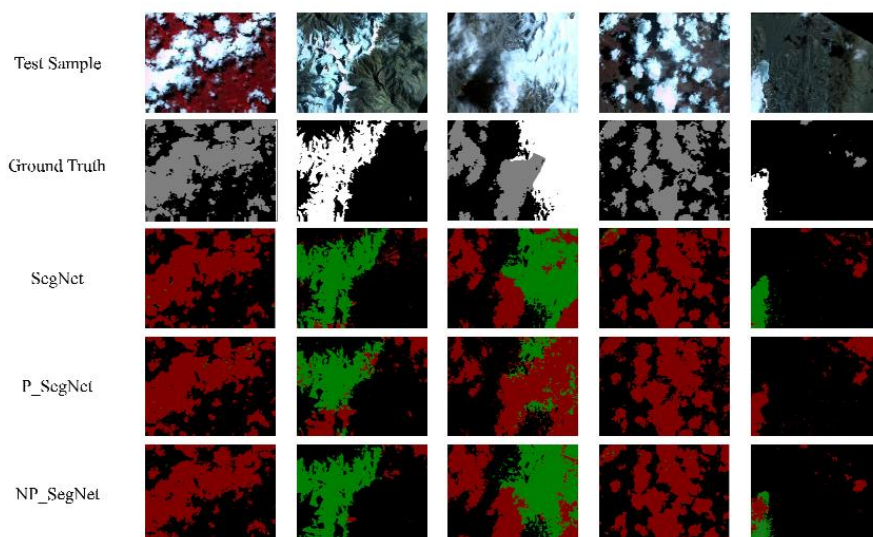


تصویر ۳-۲: معماری شبکه NP-SegNet

۳-۵: آزمایشات صورت گرفته

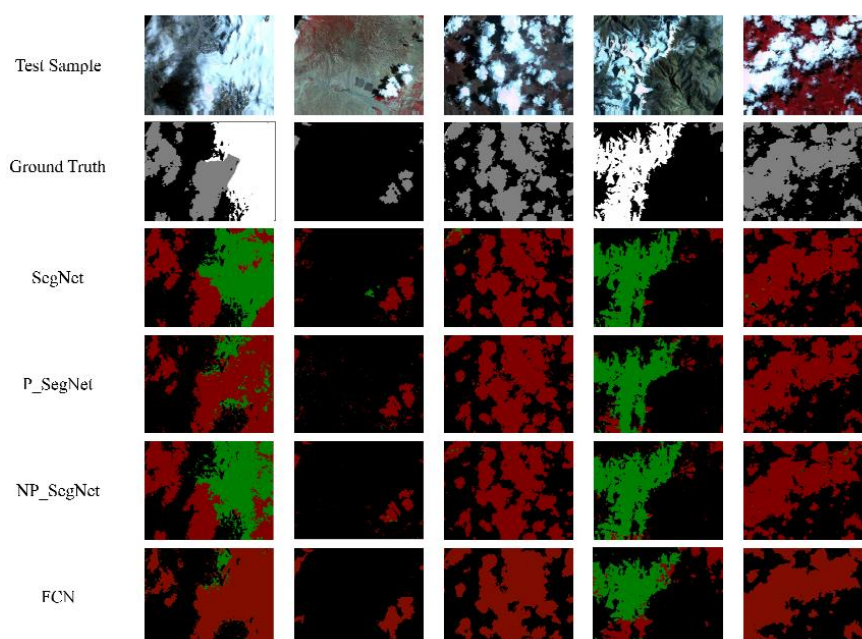
تصاویر ورودی که برای آزمایش عملکرد استفاده می‌شوند را می‌توان به چهار دسته تقسیم کرد: (۱) تصاویر ابری، (۲) تصاویر برفی، (۳) ترکیبی از ابرها و تصاویر برفی، و (۴) تصاویر ابری و برفی کمتر. تصاویر ورودی RGB در مجموعه داده آزمایشی ۴۸۰×۳۶۰ بود. سه معماری شبکه عصبی مختلف (SegNet، P_SegNet و NP_SegNet) برای تعیین عواملی که بر عملکرد تشخیص ابر تأثیر می‌گذارند، آزمایش شدند. معماری‌های NP_SegNet و P_SegNet نیز مستقیماً برای تعیین میزان تأثیر ساختارهای موازی بر دقت تشخیص ابر مقایسه شدند. همچنین عملکرد بین سه شبکه و معماری سنتی، شبکه کاملاً پیچشی را مقایسه شده است. نرخ یادگیری ۰/۰۰۰۱ با تکانه ۰/۹ استفاده شد. همچنین از همان حل کننده SGD برای تمام تست ها استفاده شده است. به منظور جلوگیری از برآزش بیش از حد شبکه های عصبی در آموزش، از تکنیک حذف تصادفی^۱ استفاده کردیم. میزان حذف تصادفی ۰/۵ بود. یک گام اولیه حیاتی، تعریف عملیاتی بود که باید معنای تصاویر ابری، برفی و ترکیبی ابر-برف را تعریف کرد. در این مقاله تصمیم گرفته شده است که اگر ابرها بیش از ۴۰ درصد تصویر را تشکیل دهند، تصویری ابری در نظر گرفته می شود. به همین ترتیب، اگر برف بیش از ۴۰ درصد تصویر را تشکیل می داد، تصویر برفی در نظر گرفته می شد. اگر تصویری حاوی بیش از ۲۰ درصد از ابرها و برف باشد، تصویر ترکیبی ابر و برف خواهد بود. مقایسه عملکرد سه شبکه مختلف شناسایی تصاویر ابری را می توان در تصویر ۳-۳ مشاهده کرد. این نشان می دهد که SegNet و اشکال تغییر یافته آن می توانند به خوبی تشخیص ابر را اداره کنند. به ویژه، همه آنها توانایی قوی برای شناسایی مرزها داشتند. با مقایسه نتایج تشخیص P_SegNet و NP_SegNet، می‌توان دریافت که ساختار شبکه عصبی با

ساختار متقارن برای تمایز ابر-برف بهتر است. P_SegNet تمایل داشت اجسام درخشان را به عنوان ابر تشخیص دهد به طوری که میزان دقت تشخیص برف بسیار کاهش یافت.



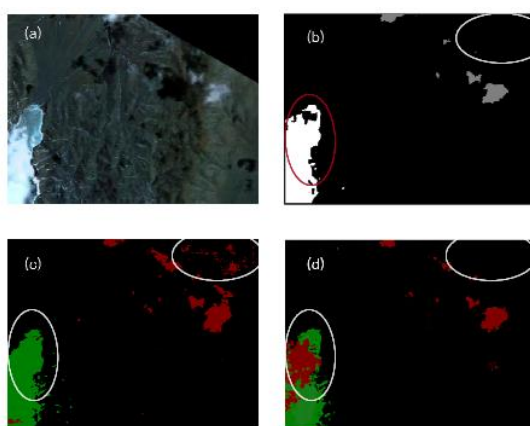
تصویر ۳-۳ : مقایسه عملکرد سه شبکه مختلف شناسایی تصاویر ابری

تصویر ۴-۳ تفاوت بین معماری شبکه کاملاً پیچشی و سه معماری SegNet را برای تشخیص ابر نشان می دهد. در مقایسه، می توانیم ببینیم که معماری شبکه کاملاً پیچشی در مورد مرزها مبهم تر است. سه شبکه عصبی SegNet مرزها را با دقت و هموارتر طبقه بندی کردند. واضح است که دقت تشخیص شبکه کاملاً پیچشی برای برف بسیار پایین بوده است.



تصویر ۴-۳ : تفاوت بین معماری شبکه کاملاً پیچشی و سه معماری SegNet

اگرچه میزان دقت تشخیص NP_SegNet برای ابر ۲/۹٪ بیشتر از SegNet است، اما میزان دقت تشخیص برای برف ۷/۸٪ کمتر از SegNet است. نتیجه نشان می‌دهد که عملکرد SegNet در میزان تمایز ابر-برف کمی بهتر از NP_SegNet است. با این حال، تصویر ۵-۳ نشان می‌دهد که احتمال شناسایی اشتباه SegNet که غیر ابری یا غیر برفی را به عنوان ابر تشخیص می‌دهد، بیشتر از NP_SegNet بود. هنگامی که ابر بخش بزرگی از تصویر سنجش از راه دور را به خود اختصاص می‌دهد، نرخ دقت تشخیص NP_SegNet بسیار بیشتر از SegNet بود. در جدول ۲-۳، دقت کلاس متوسط چهار معماری (FCN، NP_SegNet، P_SegNet، SegNet) را در تکرارهای مختلف مقایسه شده است. تکرارهای آموزشی K۵۰، K۱۰۰ و K۵۰۰ جداگانه آزمایش شده است. در آموزش، از یک دسته کوچک با اندازه ۲ استفاده شده است. در استنتاج، اندازه دسته ۱ بود. به منظور مقایسه عملکرد کمی هر معماری، از دقت میانگین کلاس و نمره کلان IF برای اندازه‌گیری توانایی هر معماری در طبقه‌بندی استفاده شده است. دقت تشخیص با ۱۳۳ تصویر سنجش از راه دور انتخاب شده در معماری‌های مختلف در K۵۰، K۱۰۰ و K۵۰۰ تکرارهای آموزشی آزمایش شد.



تصویر ۵-۳ : مقایسه Segnet و NP-Segnet

Network / Iteration	50K				100K				500K				Macro-F1
	Cloud	Snow	Background	Class avg.	Cloud	Snow	Background	Class avg.	Cloud	Snow	Background	Class avg.	
SegNet	82.6	75.5	86.0	81.4	88.6	79.5	89.5	85.9	89.9	88.7	90.6	89.7	0.707
P_SegNet	81.3	49.3	92.1	74.2	90.7	55.3	90.8	78.9	93.4	58.7	92.1	81.0	0.586
NP_SegNet	80.5	70.6	90.5	80.5	91.2	70.3	92.4	84.6	92.8	80.8	93.5	89.0	0.712
FCN	75.9	19.3	80.9	58.7	80.4	19.5	87.6	62.5	81.7	20.7	85.5	62.63	-

جدول ۲-۳

موارد شدیدی وجود داشت که می‌توانست بر نتایج نهایی تأثیر بگذارد. به عنوان مثال، در هیچ تصویر برفی، اگر مدل قسمت کوچکی از یک شی را به عنوان برف تشخیص دهد، دقت تشخیص این کلاس صفر می‌شود و بر نرخ تشخیص کلی تأثیر می‌گذارد. برای جلوگیری از چنین نتایج آزمایش شدیدی که بر دقت میانگین کلاس تأثیر می‌گذارد، مجموعه تصاویر آزمایشی فقط تصاویر ترکیبی ابر-برف بودند. همانطور که در جدول ۲-۳ نشان داده شده است، سه معماری SegNet در مدت زمان کوتاهی حدود ۱۰۰ هزار تکرار آموزشی به نتیجه بهتری دست می‌یابند. تحت شرایط آموزشی مشابه، نرخ‌های تشخیص ابر و برف معماری شبکه کاملاً پیچشی کمتر از سه معماری SegNet بود، به‌ویژه نرخ تشخیص برف. معماری‌های SegNet مانند برای تشخیص ابر مناسب‌تر بودند. با افزایش تکرارهای آموزشی، شبکه عصبی مرزهای ابر را با دقت بیشتری شناسایی کرد و نرخ شناسایی نادرست کمتر است. SegNet و NP_SegNet امتیازهای IF ماکرو مشابهی دارند و بهتر از P_SegNet هستند. این نشان می‌دهد که شبکه‌های عصبی متقارن قابلیت تشخیص ابر بهتری دارند. نتایج تجربی نشان داد که معماری SegNet مانند برای تشخیص ابر مناسب‌تر از معماری شبکه کاملاً پیچشی است. نتایج NP_SegNet و P_SegNet نشان می‌دهد که شبکه عصبی

پیچشی که متقارن است، اهداف بسیار مشابه را بهتر متمایز می‌کند. اگرچه نرخ تمایز ابر-برف SegNet کمی بیشتر از NP_SegNet است، اما نرخ مثبت کاذب آن بیشتر از NP_SegNet است. عملکرد NP_SegNet در تصاویر سنجش از راه دور بدون برف بهتر از SegNet اصلی بود. اگر تصاویر سنجش از راه دور آزمایش بیشتر با ابرها و برف مخلوط شده بود، SegNet برای تشخیص ابر مناسب تر بود.

۳-۶ : نتیجه گیری

در این مقاله، مجموعه‌ای از روش‌ها از اکتساب مجموعه‌های آموزشی تا آموزش شبکه‌های عصبی را برای دستیابی به هدف استفاده از یادگیری عمیق برای انجام کارآمد تشخیص ابر تصویر سنجش از دور پیشنهاد شده است. SegNet به عنوان معماری شبکه اصلی برای شناسایی ابر استفاده شد. تغییراتی را در معماری ارائه شد تا عوامل اساسی را که بر عملکرد در تشخیص ابر تأثیر می‌گذارند، یافته شود. همچنین سه معماری SegNet مانند را با معماری‌های کاملاً پیچیده سنتی مقایسه شده است. در این آزمایش تعدادی از مزایا و معایب معماری شبکه‌های عصبی مختلف را تحت شرایط مختلف شناسایی شده است. نتایج تجربی نشان داد که استفاده از یادگیری عمیق برای تشخیص ابر در تصویر سنجش از دور پتانسیل بیشتری نسبت به روش‌های سنتی دارد. با توسعه بیشتر شبکه‌های عصبی، یادگیری عمیق می‌تواند به طور کامل جایگزین روش‌های سنتی تشخیص ابر شود.

فصل چهارم

معماری FL-Segnet

(DONG, Wang , ZHANG&... , 2019)

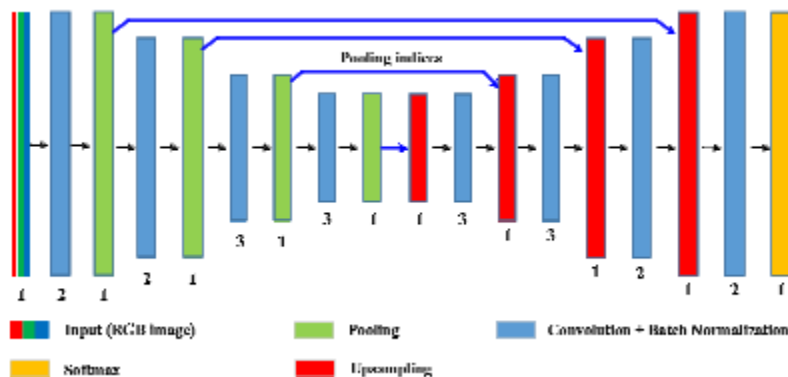
۴-۱ : مقدمه

تشخیص آسیب های ساختاری پوشش تونل برای اطمینان از عملیات ایمن تونل حیاتی است. با این حال، تشخیص نقص چندانکه به دلیل عدم تعادل اندازه بین ترک ها، پوسته شدن و پس زمینه، کار چالش برانگیزی است. در حال حاضر، روش های مبتنی بر یادگیری عمیق برای نقص های چندانکه به شبکه های چند مرحله ای وابسته هستند که مقیاس پذیری و فرآیندهای پیچیده کار فریم آن ها را محدود کرده اند. برای تشخیص دقیق نقص چندانکه در سطح پیکسل تنها با استفاده از شبکه های یک مرحله ای، روش جدیدی پیشنهاد شد که SegNet اصلی را با یک تابع از دست دادن کانونی یکپارچه می کند و به عنوان روش FL-SegNet نامیده می شود. تابع از دست دادن کانونی برای رسیدگی به متصور عدم تعادل اندازه با کاهش وزن تلفات اختصاص داده شده به نمونه های طبقه بندی شده مناسب اتخاذ شد و سپس آموزش بر روی نمونه های سخت متمرکز شد. علاوه بر این، آزمایش های مقایسه ای برای ارزیابی عملکرد روش های مختلف انجام شد. نتایج تجربی نشان داد که روش FL-SegNet قادر به پیش بینی دقیق پروفیل ترک های با اندازه کوچک و آسیب های همپوشانی حتی در شرایط مختلف نویز بود و با موفقیت از روش دو جریانی و روش اصلی SegNet در این زمینه بهتر عمل کرد. معیارهای عملکرد (MPA و MloU) روش FL-SegNet به طور قابل توجهی بالاتر از سایر روش های تشخیص عیب چندانکه در سناریوهای مختلف بود (تصاویر با آسیب های کوچک به ۸۱/۵۳٪ و ۶۹/۸۶ درصد نسبت به روش دو جریانی ۱۱/۹۹ درصد و ۴/۸۸ درصد افزایش و نسبت به SegNet پایه ۱۷/۷۸ درصد و ۷/۶۹ درصد افزایش داشته است. بنابراین، این مقاله یک راه حل موثر برای تشخیص آتی عیب های متعدد در پوشش تونل ارائه می دهد.

۴-۲ : معرفی معماری

به دلیل ویژگی های رطوبت و اسید-باز محیط، پوشش های ضد حریق روی سطح پوشش تونل تغییر رنگ داده، پوشش داده می شوند و پوسته پوسته می شوند. مقاومت ناکافی بتن باعث ترک خوردن و پوسته شدن بتن می شود. به آن پدیده ها ریزش می گویند. روش پیشنهادی FL-SegNet یک SegNet پایه را با از دست دادن کانونی برای رسیدگی به تصاویر اندازه های نامتعادل که در طی فرآیندهای شناسایی آسیب های متعدد با آن مواجه می شوند، یکپارچه کرده بود. این روش یک شبکه انتها به انتها، پیکسل به پیکسل برای پیش بینی کلاس های آسیب در هر پیکسل بود، که می توانست نمایه آسیب ها را بازسازی کند.

شبکه SegNet برای شناسایی همزمان ترک ها و آسیب های پوسته شدن در لاینینگ تونل استفاده شد. ورودی شبکه تصاویر اصلی بود و خروجی برچسب گذاری پیکسلی پیش بینی بود که با تصاویر مطابقت داشت. شبکه SegNet از یک شبکه رمزگذار و یک شبکه رمزگشای متناظر تشکیل شده بود که سپس توسط یک لایه طبقه بندی پیکسلی نهایی دنبال شد. شبکه رمزگذار شامل پیچشی با یک بانک فیلتر، نرمال سازی دسته ای و حداکثر ادغام برای به دست آوردن نقشه های ویژگی بود. رمزگشا نقشه های ویژگی ورودی را با استفاده از شاخص های جمع آوری حداکثری حفظ شده به منظور بازسازی تصاویر ورودی نمونه برداری کرد. معماری شبکه SegNet در تصویر ۴-۱ به تفصیل توضیح داده شده است.



تصویر ۴-۱ : معماری Segnet

معماری این شبکه همان معماری Segnet است که از تابع ضرر کانونی استفاده کرده است.

با توجه به عدم تعادل اندازه‌های ترک، پوسته پوسته شدن و پس‌زمینه روی تصاویر مشابه، دقت تشخیص آسیب‌های متعدد با استفاده از شبکه‌های یک مرحله‌ای نسبتاً پایین بود. مشاهده شده است که این امر به ویژه در مورد مرزها و آسیب‌های در مقیاس کوچک صادق است. برای مثال، خطوط مرزی توصیف شده کاملاً مبهم بوده و آسیب‌های در مقیاس کوچک، مانند ترک‌های ریز، قبلاً به سختی قابل تشخیص بوده‌اند. بنابراین، عدم تعادل اندازه ترک‌ها، پوسته شدن و پس‌زمینه‌ها، موانع اصلی جلوگیری از پیشی گرفتن روش‌های تشخیص آسیب چندگانه مبتنی بر روش شبکه‌های یک مرحله‌ای از روش‌های مبتنی بر مدل دو جریانی بوده است. بنابراین، در مطالعه حاضر، به منظور پرداختن به این متصور، از تابع از دست دادن کانونی به جای تابع آنتروپی متقابل سنتی (CE) به منظور تمرکز بر یادگیری مثال‌های سخت و کاهش وزن منفی‌های متعدد استفاده شد. عملکرد از دست دادن کانونی (FL) به شرح زیر بود:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

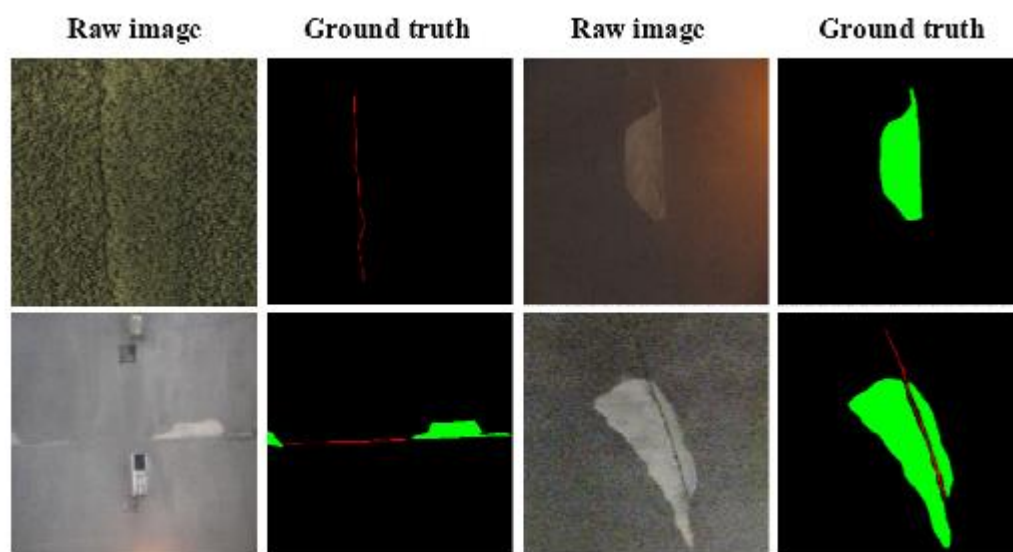
$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise.} \end{cases}$$

که در آن y کلاس حقیقت پایه است و p نشان دهنده برچسب پیش بینی با کلاس حقیقت پایه ۱ است. γ (۱ - p_t) به عنوان پارامتر تمرکز به عنوان یک عامل تعدیل کننده برای کاهش اتلاف نسبی برای نمونه‌های طبقه بندی شده خوب ($p_t > 0.5$) استفاده می‌شود، و بیشتر بر روی نمونه‌های طبقه بندی اشتباه (سخت) تمرکز می‌شود. γ افزایش می‌یابد اثر عامل تعدیل نیز افزایش می‌یابد. به این ترتیب، مشارکت ترک‌ها، پوسته شدن و پس‌زمینه‌ها را می‌توان در عملکرد از دست دادن متعادل کرد، که نه تنها کارایی را بهبود می‌بخشد، بلکه دقت شناسایی را نیز افزایش می‌دهد.

۴-۳ : آزمایشات صورت گرفته

در این مقاله، آزمایش‌هایی برای تأیید برتری روش پیشنهادی یک مرحله‌ای مبتنی بر تشخیص آسیب‌های متعدد در شناسایی ترک‌ها و پوسته شدن در پوشش‌های تونل انجام شد. مجموعه داده‌های مورد استفاده در این مطالعه از تونل نانشیبی، که یک تونل جداگانه در بزرگراه داگوانگ در استان جیانگشی است، گرفته شده است. طول کل تونل ۹۲۴ متر است. این تونل بین دو مخزن واقع شده و با برهنه شدن ساختاری، فرسایش تکنیکی و تصویر

های زمین حل شده مشخص می شود. تعداد زیادی دندان سنگی، شیاریهای محلول و غارهای کارستی در سطح زمین بیرون زده و عمده ترین پدیده نامطلوب زمین شناسی آن کارست و گسل است. بنابراین، عیوب بسیاری در ساختار پوشش تونل وجود دارد. تعداد کل تصاویر خام ۱۸۷۹ بود که شامل ۱۸۰۷ تصویر با ۳۴۵۶×۲۳۰۴ پیکسل بود. ۱۸ تصویر با ۴۰۹۶×۴۰۹۶ پیکسل. و ۵۴ تصویر سه بعدی ۱۰۳۱۵×۴۲۶۸ پیکسل. سپس ۲۹۸ تصویر با ابعاد ۳۰۰×۳۰۰ پیکسل از تصاویر خام استخراج شد که شامل ۸۳ تصویر ترک، ۳۵ تصویر پوسته پوسته شدن و ۱۸۰ تصویر حاوی هر دو ترک و پارگی بود. کل مجموعه داده ها به دو بخش به شرح زیر تقسیم شدند: مجموعه های آموزشی (۲۲۶ تصویر) و مجموعه های تست (۷۲ تصویر). علاوه بر این، هیچ یک از مجموعه های تست برای آموزش استفاده نشد. تصاویر جمع آوری شده و حقایق زمینی مربوطه در تصویر ۲-۴ نشان داده شده است.



تصویر ۲-۴: تصویر آزمایشات صورت گرفته در تصاویر تونل

علاوه بر این، برای اثبات عدم تطبیق بیش از حد مدل، تصاویر آموزشی به طور تصادفی به پنج قسمت (چهار قسمت برای آموزش، و یک قسمت برای اعتبارسنجی) برای آزمایش های اعتبارسنجی متقابل پنج برابری این مطالعه تقسیم شدند. مقادیر دقت و تلفات برای هر مرحله از آموزش و اعتبارسنجی در جدول ۲ نشان داده شده است که در آن "دقت و ضرر" به ترتیب نشان دهنده دقت تمرین و از دست دادن تمرین است و Val_acc و Val_loss نشان دهنده دقت و از دست دادن اعتبار هستند. واضح بود که مقادیر Val_acc و Val_loss رضایت بخش برای هر مرحله به دست آمده بود، که نشان می دهد مدل در مجموعه داده های مورد استفاده بیش از حد مطابقت ندارد.

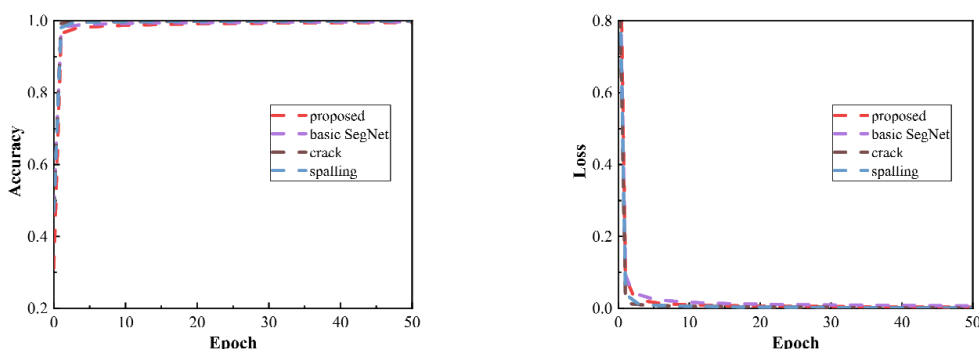
Step	Accuracy	Loss	Val acc	Val loss
1	0.9952	0.0034	0.9867	0.0768
2	0.9950	0.0040	0.9886	0.0450
3	0.9950	0.0039	0.9871	0.0516
4	0.9951	0.0037	0.9850	0.0641
5	0.9952	0.0036	0.9875	0.0449
Mean	0.9951	0.0037	0.9870	0.0565

جدول ۱-۴

ابعاد دقیق هر لایه و عملیات موجود در روش پیشنهادی FL-SegNet، ۴ لایه حداکثر جمع آوری؛ ۴ لایه نمونه برداری بالا؛ یک لایه ورودی؛ و یک لایه فعال سازی Softmax. Adadelta به عنوان بهینه ساز انتخاب شد که نیازی به تنظیم پیش از نرخ یادگیری نداشت.

تنظیم پارامترهای α و γ بسیار مهم بود. در این مطالعه، γ برای متعادل کردن مثال‌های مثبت/منفی ۲ تنظیم شد. متغیر وزن $\alpha \in [0, 1]$ سپس می‌تواند توسط فرکانس کلاس معکوس به منظور رسیدگی به عدم تعادل کلاس تنظیم شود. در کار تقسیم‌بندی معنایی این مطالعه، از آن برای پرداختن به عدم تعادل اندازه استفاده شد که به صورت عدم تعادل در تعداد پس‌زمینه، ترک و پیکسل‌های پراکنده بیان می‌شود. سپس برای تعیین مقادیر مناسب α ، مقادیر $[0.1, 0.3]$ ، $[0.1, 0.3]$ و $[0.1, 0.3]$ برای مقاصد مقایسه تنظیم شدند. مطابق با این اصل که سهم تعداد بیشتری از نمونه‌ها در تابع ضرر کوچک بوده و مجموع مشارکت همه نمونه‌ها در تابع ضرر ۱ است، عملیات زیر روی α انجام شد. برای ارزیابی دقت انواع مختلف روش‌ها برای تقسیم‌بندی معنایی، دو معیار ارزیابی به شرح زیر اتخاذ شد: میانگین دقت پیکسل (MPA) و میانگین تقاطع روی اتحاد (MIoU).

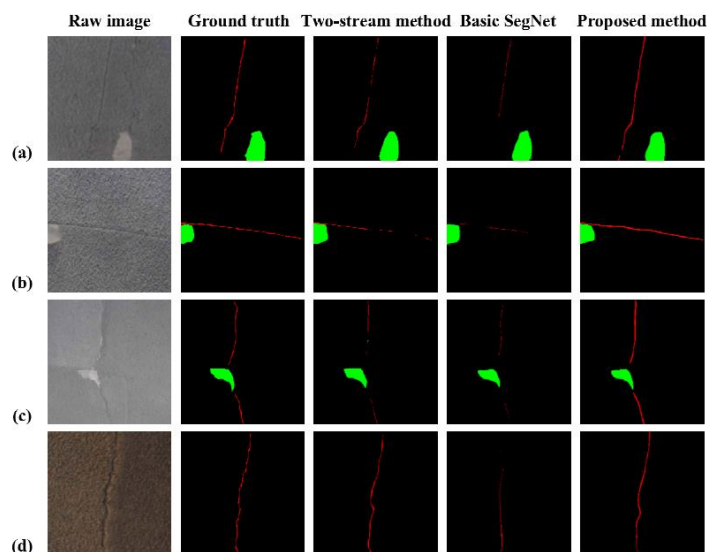
تصویر ۳-۴ منحنی‌های دقت آموزشی و منحنی‌های تلفات سه نوع روش را روی یک مجموعه داده به منظور شناسایی آسیب‌های متعدد در سطح پیکسل نشان می‌دهد. برای روش اصلی SegNet، حداکثر دقت و حداقل تلفات به ترتیب به ۹۹/۶۸٪ و ۰/۰۰۷ در دوره پنجاهم رسیده بود. ضمناً برای روش دو جریانی که از مدل ترک و مدل اسپلینگ تشکیل شده بود، دو مدل به طور جداگانه آموزش داده شد. مدل ترک حداقل اتلاف ۰/۰۰۲ خود را در دوره ۵۰ با دقت ۹۹/۹۱ درصد به دست آورد. همچنین حداکثر دقت و حداقل تلفات برای مدل پوسته شدن به ترتیب ۹۹/۹۲ درصد و ۰/۰۰۲ تعیین شد که در دوره ۵۰ تثبیت شده بود. در مقایسه با نتایج دو شبکه فوق‌الذکر، FL-SegNet پیشنهادی در دوره ۳۵ همگرا شده بود و حداکثر دقت و حداقل تلفات به ترتیب ۹۹/۵۲٪ و ۰/۰۰۳ بود.



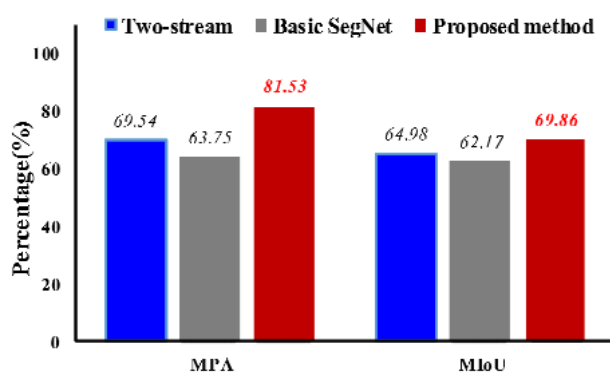
تصویر ۳-۴: منحنی‌های دقت آموزشی و منحنی‌های تلفات سه نوع روش

یکی از مشارکت‌های اصلی مطالعه حاضر این بود که روش پیشنهادی می‌تواند به دقت کلاس‌های آسیب‌دیده در هر پیکسل را از آسیب‌های متعدد در اندازه‌های بسیار متفاوت تشخیص دهد. به عنوان مثال، با توجه به ترک‌ها و آسیب‌های ریزش در تونل‌ها، اندازه‌های ترک‌ها بسیار کوچک‌تر از پوسته شدن و پس‌زمینه روی همان تصویر تعیین شد که قبلاً منجر به تصاویر بزرگ عدم تعادل اندازه شده بود. بنابراین، این مطالعه ابتدا آسیب‌های متعدد با ترک‌های کوچک را با استفاده از روش پیشنهادی شناسایی کرد تا عملکرد برتر روش در پرداختن به تصاویر عدم تعادل اندازه را تایید کند. تصویر ۴-۴ برخی از نتایج تشخیص ارائه شده توسط روش‌های مختلف شناسایی آسیب‌های چندگانه را هنگام بررسی داده‌های تصویر با ترک‌های کوچک، پوسته شدن و پس‌زمینه‌های پیچیده ارائه می‌کند. واضح بود که روش پیشنهادی در مقایسه با حقایق پایه از روش‌های دیگر بهتر عمل کرده است. روش اصلی SegNet مستعد تداخل عدم تعادل اندازه است و بسیاری از پیکسل‌های ترک به اشتباه شناسایی شده‌اند. روش دو جریانی نسبتاً بهتر از روش اصلی SegNet عمل کرد زیرا به وضوح مرزهای آسیب‌های پوسته شدن را تشخیص

داد. با این حال، همانطور که در تصویر نشان داده شده است، همچنان بسیاری از نتایج نادرست پیش بینی شده را برای ترک های پوشش تونل ارائه می کند. $bF-4$ و $cF-4$. در مقابل، FL-SegNet، پیشنهادی، نمایه های اصلاح شده ترک ها را بازسازی کرد، و همچنین به پیش بینی های پیکسلی قوی کمک کرد، حتی اگر برخی از ترک ها در مقایسه با پس زمینه های پرپشت و پیچیده تصاویر بسیار کوچک بودند.

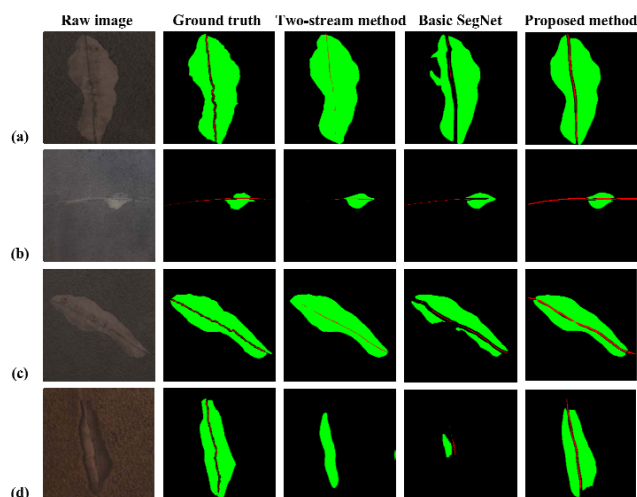


تصویر ۴-۴: نتایج تشخیص ارائه شده توسط روش های مختلف

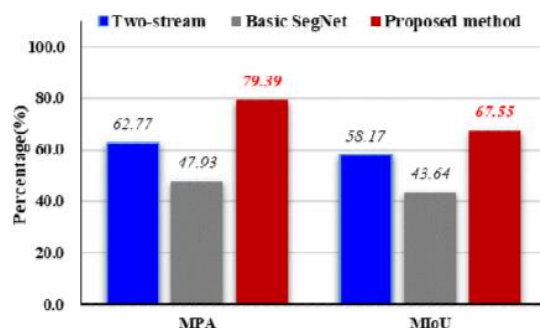


تصویر ۴-۵: نتایج معیارهای معماری های متفاوت

یکی از چالش های شناخته شده برای تشخیص آسیب چندگانه پوشش های تونل، همپوشانی ترک ها و پوسته شدن است. پیکسل های متضاد نواحی همپوشانی باعث می شود که تشخیص آسیب های متعدد دشوارتر شود. برای تأیید استحکام روش FL-SegNet در پرداختن به متصویر همپوشانی، برخی از نتایج پیش بینی شده حاوی آسیب های متعدد همپوشانی در فهرست کوتاه قرار گرفتند، همانطور که در تصویر $F-6$ نشان داده شده است. مشاهده شد که در برخی موارد، همپوشانی ترک ها و پوسته شدن به طور دقیق توسط هر سه روش بررسی شده، همانطور که در تصویر $bF-6$ شرح داده شده است، شناسایی شده بود. با این وجود، برای اکثر موارد، روش دو جریانی و روش اصلی SegNet نتوانستند به طور رضایت بخشی آسیب های متعدد همپوشانی را تشخیص دهند.



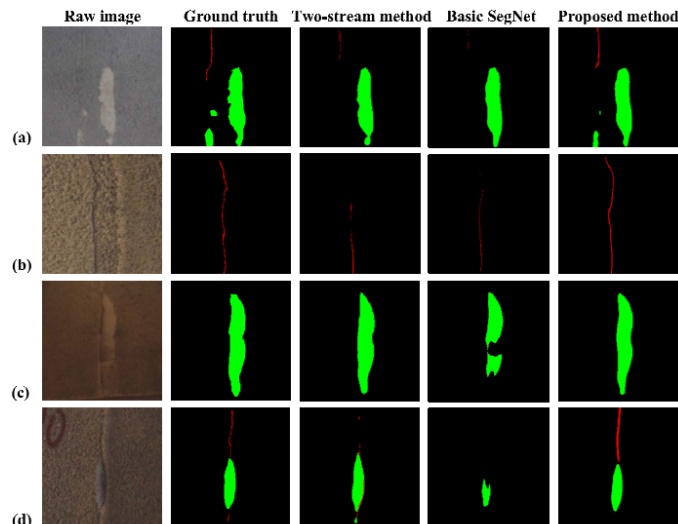
تصویر ۴-۶: نتایج تشخیص ارائه شده توسط روش‌های مختلف در همپوشانی ترک‌ها



تصویر ۴-۷: نتایج معیارهای معماری‌های متفاوت در همپوشانی ترک‌ها

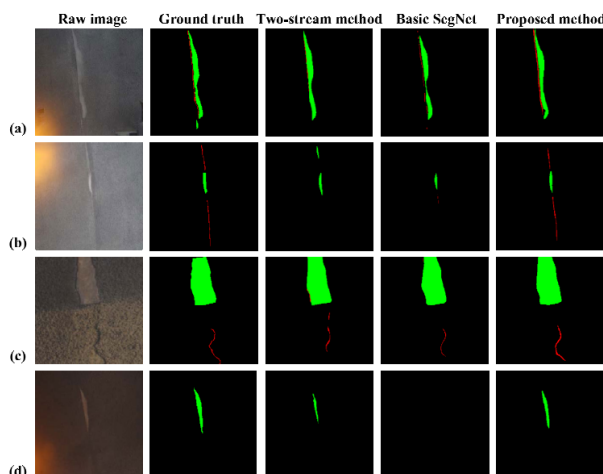
تداخل پس‌زمینه و شرایط نوری ناهموار دو تصویر اصلی نویز در تونل‌ها هستند و بر شناسایی آسیب‌های متعدد تأثیر می‌گذارند. نتایج تجربی این مطالعه در این بخش به منظور اعتبارسنجی استحکام و سازگاری روش پیشنهادی FL-SegNet در محیط‌های تونلی که دو منبع اصلی نویز فوق‌الذکر وجود دارد، نشان داده شد.

۱) تداخل پس‌زمینه: در محیط تونل، تأثیرات نویز پس‌زمینه، مانند بافت‌های پس‌زمینه، سایه‌زنی لامپ و علائم دستی روی سطوح پوشش‌های تونل، تنوع داده‌ها را افزایش می‌دهد و چالش‌های عمده‌ای را برای شناسایی دقیق آسیب‌ها ایجاد می‌کند. تصویر ۴-۸ تصاویری را بیان می‌کند که تحت تأثیر نویز پس‌زمینه در مطالعه حاضر قرار گرفته بودند. ۴-۸a، ۴-۸b، ۴-۸c و ۴-۸d نشان دادند که روش پیشنهادی نتایج پیش‌بینی عالی را برای تصاویر با بافت‌های پس‌زمینه متغیر و علامت‌های دستی نشان داده است. با این حال، مشاهده شد که روش اصلی SegNet نمی‌تواند مشخصات ترک‌ها و پوسته شدن را به طور کامل شناسایی کند و همچنین نتوانسته بود ترک‌هایی را که به تفصیل در تصویر ۴-۸d شناسایی شده‌اند.



تصویر ۴-۸: نتایج بررسی معماری های متفاوت تحت تأثیر نویز پس زمینه

۲) شرایط نوری ناهموار : شرایط نور در تونل ها ممکن است منجر به ایجاد لکه ها، سایه ها و سایر نویزها در پس زمینه اطراف شود که بر عملکرد روش های تشخیص آسیب تأثیر منفی می گذارد. در این مطالعه، برخی از تصاویر معرف با متصویرات تداخل نوری ناهموار در لیست کوتاه قرار گرفتند و عملکرد سه روش شناسایی آسیب به طور مفصل مورد بحث قرار گرفت. همانطور که در تصویر ۴-۹ مشاهده می شود، ترک ها و آسیب های پوسته شدن با روش پیشنهادی به طور رضایت بخشی شناسایی شدند. ۴-۹ b و ۴-۹ d، می توان مشاهده کرد که روش اصلی SegNet به اشتباه ترک ها و پوسته شدن را تحت تأثیر شرایط نوری ناهموار تشخیص داده است، و پیش بینی ترک روش پایه SegNet به اشتباه ترک های ناپیوسته و ناقص را نشان می دهد. ۴-۹ a و ۴-۹ c نشان دادند که سه روش نسبتاً مشابه بودند. با این حال، FL-SegNet توانایی هایی را برای بازسازی دقیق نمایه های تمام آسیب ها و همچنین مرزهای همه آسیب ها نشان داد.



تصویر ۴-۹: نتایج بررسی معماری های متفاوت تحت شرایط نوری ناهموار

جدول ۴-۲ مقایسه شاخص های میانگین در فرآیندهای تست روش دو جریانی، روش SegNet پایه و روش پیشنهادی FL-SegNet را توضیح می دهد. می توان مشاهده کرد که روش پیشنهادی بالاترین امتیازات را برای MPA و MIoU تحت دو نوع تداخل بررسی شده به دست آورده است.

Method	Background Interference		Uneven Lighting	
	MPA(%)	MIoU(%)	MPA(%)	MIoU(%)
Two-stream	61.52	58.23	66.48	62.73
Basic SegNet	63.77	62.32	56.89	55.64
Proposed method	83.35	75.06	75.67	69.15

جدول ۴-۲

۴-۴ : نتیجه گیری

در مطالعه حاضر، یک روش مبتنی بر DL برای شناسایی همزمان آسیب‌های متعدد در پوشش‌های تونل با استفاده از شبکه‌های یک مرحله‌ای پیشنهاد شد. روش پیشنهادی، که در این مطالعه به عنوان FL-SegNet نامیده می‌شود، که شبکه اصلی SegNet را با یک تابع از دست دادن کانونی برای تمرکز بر یادگیری مثال‌های سخت و کاهش وزن منفی‌های آسان متعدد یکپارچه می‌کند. مشخص شد که روش پیشنهادی قادر به تشخیص ترک‌های با مقیاس کوچک و سایر جزئیات از آسیب‌های پوخته شدن مقیاس بزرگ است. همچنین قادر بود به طور همزمان چندین آسیب را حتی با وجود تفاوت در اندازه یا همپوشانی تشخیص دهد. این روش با استفاده از داده‌های تصویر آسیب تونل عملی با موفقیت تأیید شد. مقادیر بهینه α و γ در تابع از دست دادن کانونی به ترتیب برای مدل آموزش دیده شده به ترتیب [۱۰، ۱، ۳] و ۲ تعیین شد. عملکردهای روش پیشنهادی FL-SegNet به صورت تجربی در سناریوهای مختلف، از جمله آسیب‌های کوچک، ترک‌ها و پوخته شدن، تداخل پس‌زمینه و شرایط نوری ناهموار مورد بحث قرار گرفت. عملکرد روش پیشنهادی با دو روش شناسایی آسیب چندگانه مقایسه شد: یک روش دو جریانی و یک روش پایه SegNet. نتایج تجربی نشان داد که روش پیشنهادی FL-SegNet هم از روش دو جریانی و هم از روش اصلی SegNet برای هر چهار مورد بهتر عمل کرده است. MPA و MIOU روش پیشنهادی FL-SegNet به ترتیب ۱۱/۹۹٪ و ۴/۸۸٪ برای شناسایی آسیب چندگانه نامتعادل اندازه در مقایسه با روش دو جریانی افزایش یافت. بنابراین، از دیگر الگوریتم‌های عالی تشخیص آسیب چندگانه بهتر عمل کرد. در همین حال، با توجه به همپوشانی آسیب چندگانه، MPA روش FL-SegNet از ۶۲/۷۷٪ به ۷۹/۳۹٪ بهبود یافت و MIOU آن از ۵۸/۱۷٪ به ۶۷/۵۵٪ افزایش یافت. علاوه بر این، معیارهای عملکرد FL-SegNet بالاترین امتیاز را برای شناسایی آسیب‌های چندگانه تحت شرایط تداخل پس زمینه (MPA: 83.35٪ و MIOU: 75.06٪) و روشنایی ناهموار (MPA: 75.67٪ و MIOU: 69.15٪) به دست آوردند. در نتیجه، نتایج این مطالعه نشان داد که روش پیشنهادی FL-SegNet توانایی اجرای همزمان تشخیص عیب چندگانه را با دقت بالاتری دارد.

فصل پنجم

معماری M-Segnet

(Yamanakkanavar & Lee , 2019)

۱-۵ : مقدمه

در این مقاله، یک معماری جدید M-SegNet با توجه جهانی برای تقسیم‌بندی تصویربرداری تشدید مغناطیسی مغز (MRI) پیشنهاد شده است. معماری پیشنهادی شامل یک شبکه عمیق چند مقیاسی در سمت رمزگذار، نظارت عمیق در سمت رمزگشا، یک مکانیسم توجه جهانی، اندازه‌های مختلف هسته‌های پیچشی، و اتصالات ترکیبی با اتصالات پرش و شاخص‌های ادغام است. لایه‌های ورودی جانبی چند مقیاسی برای پشتیبانی از لایه‌های عمیق برای استخراج اطلاعات متمایز استفاده می‌شوند و لایه فرا نمونه‌گیری در سمت رمزگشا نظارت عمیقی را ارائه می‌دهد که متصور گرادیان را کاهش می‌دهد. مکانیسم توجه جهانی برای گرفتن اطلاعات متنی غنی در مرحله رمزگشا با ادغام ویژگی‌های محلی با وابستگی‌های جهانی مربوطه استفاده می‌شود. علاوه بر این، هسته‌های چند مقیاسی پیچشی با اندازه‌های مختلف برای استخراج ویژگی‌های معنایی فراوان از اسکن‌های MRI مغز در مازول‌های رمزگذار و رمزگشا استفاده شد. علاوه بر این، از اتصالات ترکیبی برای انتقال ویژگی‌ها از رمزگذار به مسیر رمزگشا برای بازیابی اطلاعات فضایی از دست رفته در طول فرو نمونه برداری استفاده شد و باعث می‌شود مدل سریع‌تر همگرا شود. علاوه بر این، تکه‌های ورودی یکنواخت بدون همپوشانی برای تمرکز بر جزئیات دقیق برای تقسیم‌بندی MRI مغز اتخاذ شده است. این معماری پیشنهادی در مجموعه داده‌های در دسترس عموم برای وظیفه تقسیم‌بندی MRI مغز تأیید شده است. نتایج تجربی نشان می‌دهد که مدل پیشنهادی با دستیابی به میانگین امتیاز ضریب شباهت تاس ۰/۹۶ از روش‌های مرسوم بهتر عمل می‌کند.

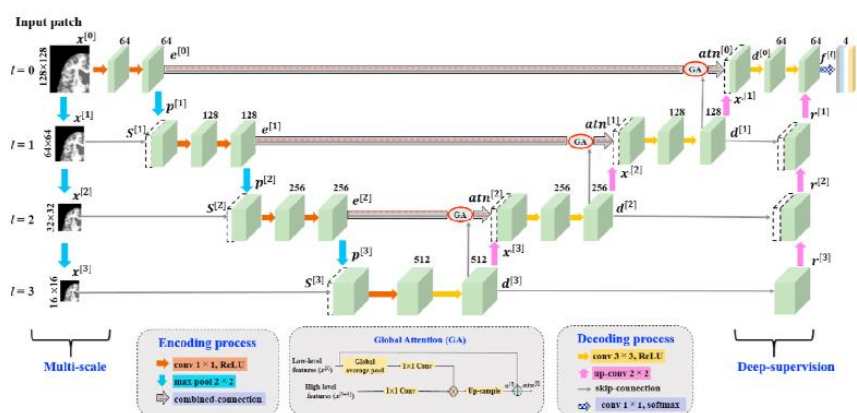
۲-۵ : معرفی موضوع

در این مقاله یک معماری جدید M-SegNet با توجه جهانی برای بخش بندی MRI مغز پیشنهاد شده است. معماری M-SegNet پیشنهادی شامل ورودی چند مقیاسی (پای چپ) و نظارت عمیق (پای راست) است، به ویژه با مکانیزم توجه جهانی جدید، اندازه‌های مختلف هسته‌های پیچشی در بلوک‌های رمزگذار-رمزگشا، و اتصالات ترکیبی برای تقسیم‌بندی بافت های مغز در MRI ورودی نمونه برداری شده در مسیر سمت چپ، اطلاعات چند مقیاسی را استخراج می‌کند و نمایش ویژگی‌ها را در لایه رمزگذار بهبود می‌بخشد. خروجی لایه رمزگشای نمونه برداری شده در مسیر سمت راست به مدل کمک می‌کند تا همگرایی را سرعت بخشد و متصور گرادیان ناپدید شدن را برطرف کند. بنابراین، هر دو لایه جانبی در شبکه پیشنهادی اطلاعات متمایز را استخراج می‌کنند، که منجر به مشتق کلی بالاتری از لایه شبکه می‌شود که برای کاهش متصور گرادیان مفید است. یک مکانیسم توجه جهانی جدید در مرحله رمزگشا برای ارائه زمینه جهانی ویژگی‌های سطح بالا به عنوان افزونه‌ای به ویژگی‌های سطح پایین برای دستیابی به محلی‌سازی دسته‌بندی کلاس اعمال می‌شود. علاوه بر این، از اندازه‌های مختلف هسته‌های پیچشی در بلوک‌های رمزگذار-رمزگشا برای استخراج ویژگی‌های معنایی فراوان از اسکن‌های MRI مغز استفاده می‌شود. برای بازیابی اطلاعات فضایی از دست رفته در طول نمونه برداری پایین، از اتصالات ترکیبی برای انتقال ویژگی‌ها از رمزگذار به مسیر رمزگشا استفاده می‌شود. این به مدل اجازه می‌دهد تا با کاهش پارامترهای قابل یادگیری در شبکه پیشنهادی سریع‌تر همگرا شود. علاوه بر این، ما از تکه‌های یکنواخت از کل بخش ورودی MRI برای آموزش مدل پیشنهادی استفاده کردیم. چنین تقسیم‌بندی به تکه‌های یکنواخت، محلی‌سازی در MRI مغز را بهبود می‌بخشد و به مدل آموزش‌دیده اجازه می‌دهد تا بر روی جزئیات دقیق‌تر در هر وصله تمرکز کند، همانطور که در مطالعه قبلی ما تأیید شد. در شبکه‌های عصبی عمیق، مکانیسم توجه بر مناطق کلیدی محلی در نقشه‌های ویژگی تأکید می‌کند

و داده‌های نامربوط را فیلتر می‌کند. اخیراً، بسیاری از محققان بر مکانیسم‌های توجه متمرکز شده‌اند، زیرا می‌توانند به مدل‌سازی وابستگی‌های دوربرد کمک کنند و در عین حال عملکرد محاسباتی و آماری را حفظ کنند.

۳-۵ : معرفی معماری

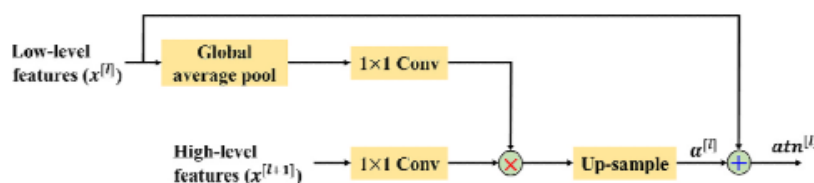
SegNet، U-net و M-net دارای محدودیت‌های مشترکی هستند، مانند از دست دادن اطلاعات همسایه یا محلی و عملیات ادغام متوالی یا گام‌های پیچشی، که وضوح نقشه ویژگی را کاهش می‌دهد. برای رسیدگی به این متصورات و بهبود دقت بخش‌بندی MRI مغز، معماری M-SegNet را با توجه جهانی پیشنهاد شده است که از مسیرهای جانبی، اتصالات ترکیبی، مکانیزم توجه جهانی، اندازه‌های مختلف هسته‌های پیچشی در رمزگذار استفاده می‌کند. و بلوک‌های رمزگشا و وصله‌های ورودی یکنواخت. معماری پیشنهادی ما از شبکه M-net الهام گرفته شده است، که در ابتدا برای حذف نویز و رنگ آمیزی تصاویر اثر انگشت به طور همزمان استفاده می‌شد. تصویر ۱-۵ معماری کلی شبکه پیشنهادی را نشان می‌دهد. همانطور که نشان داده شده است، شبکه پیشنهادی دارای یک ساختار مبتنی بر رمزگشا-رمزگشا است که از یک ویژگی کلی M-net، مانند ورودی چند مقیاسی و نظارت عمیق، مکانیزم توجه جهانی جدید، اتصالات ترکیبی، و اندازه‌های مختلف هسته‌های پیچشی تشکیل شده است. در بلوک‌های رمزگذار و رمزگشا.



تصویر ۱-۵ : معماری کلی شبکه M-Segnet

بلوک رمزگذار: هر لایه در بلوک رمزگذار از دو بلوک پیچشی متوالی با اندازه هسته 1×1 تشکیل شده است که هر بلوک پیچشی توسط یک تابع ReLU فعال می‌شود. این بلوک‌ها ویژگی‌های بسیار محلی را استخراج می‌کنند و به گرفتن اطلاعات کوچکتر و پیچیده در تصویر کمک می‌کنند. در معماری M-SegNet پیشنهادی ما، بلوک‌های پیچشی شبکه‌های پیش‌خور هستند که خروجی لایه (۱-۱) -امین را به عنوان ورودی به لایه ۱-ام متصل می‌کنند. تابع فعال سازی ReLU از واحد پیچشی ۱ است. خروجی واحد پیچشی با استفاده از عملیات ادغام حداکثر 2×2 با گام دو به صورت پایین نمونه برداری می‌شود که در بیان می‌شود. در سمت چپ معماری، تصویر ورودی با استفاده از عملیات ادغام حداکثر 2×2 با گام دو نمونه برداری می‌شود و از طریق اتصالات پرش جانبی به لایه رمزگذار مربوطه متصل می‌شود. ویژگی به دست آمده توسط دو بلوک پیچشی متوالی در لایه رمزگذار با استفاده از اتصالات ترکیبی به لایه رمزگشای مربوطه منتقل شد. این اتصالات برای انتقال ویژگی‌ها از رمزگذار به مسیر رمزگشا برای بازیابی اطلاعات فضایی از دست رفته در حین فرو نمونه برداری و ایجاد همگرایی سریع‌تر مدل استفاده می‌شوند. همانطور که در تصویر ۱-۵ نشان داده شده است، اتصالات ترکیبی مورد استفاده در شبکه پیشنهادی با فلش‌های قرمز خاکستری و نقطه چین نشان داده شده است.

توجه جهانی (GA): مکانیسم توجه یادگیری عمیق بسیار شبیه توجه بصری انسان است. هدف اصلی طبقه‌بندی مهم‌ترین جزئیاتی است که برای تقسیم‌بندی بافت از مقدار زیادی داده لازم است. علاوه بر این، مکانیسم توجه با حذف ویژگی‌های نامربوط، کارایی شبکه را افزایش می‌دهد. برای درک این موضوع، ما یک GA جدید با مکانیزم توجه به خود پیشنهاد می‌کنیم که تقسیم بندی را به طور موثر انجام می‌دهد. GA به زمینه جهانی به عنوان راهنمایی از ویژگی‌های سطح پایین برای تعیین جزئیات محلی سازی کلاس اجازه می‌دهد. تصویر ۵-۲ معماری GA را نشان می‌دهد که با معماری تقسیم بندی مغز پیشنهادی ادغام شده است.



تصویر ۵-۲ : معماری توجه جهانی

از $x^{[l+1]}$ یک مقیاس درشت‌تر جمع‌آوری می‌شود و به عنوان یک بردار سیگنال دروازه‌ای عمل می‌کند. برای انتخاب ناحیه فوکوس روی هر پیکسل اعمال شد. $\alpha^{[l]}$ ضریب توجهی است که با سرکوب پاسخ‌های ویژگی نامربوط، فعال سازی مرتبط با وظیفه هدف را حفظ می‌کند. خروجی GA افزودن عنصری نقشه ویژگی از لایه رمزگذاری l با ضریب توجه است.

ادغام میانگین جهانی اطلاعات زمینه جهانی را به عنوان راهنمایی برای ویژگی‌های سطح پایین فراهم می‌کند. این ویژگی‌های محلی را با زمینه جهانی ادغام می‌کند. علاوه بر این، اطلاعات کلی تولید شده از ویژگی سطح بالا به عنوان ورودی به پیچیدگی 1×1 با تابع فعال سازی ReLU تغذیه می‌شود. سپس با 1×1 ویژگی‌های سطح پایین پیچیده ضرب می‌شود تا ویژگی‌های وزنی سطح پایین استخراج شود. ما از توجه ضربی برای به دست آوردن ضریب توجه استفاده کردیم، که سپس نمونه‌برداری می‌شود و با ویژگی‌های سطح پایین اضافه می‌شود تا محلی‌سازی پیکسلی خاص برای دسته کلاس نقشه ویژگی سطح بالا استخراج شود.

بسیاری از روش‌های مرسوم از اتصالات پرش برای انتقال نقشه‌های ویژگی از بلوک رمزگذار به بلوک رمزگشا برای بازسازی بهتر در خروجی شبکه استفاده می‌کنند. با این حال، تمایز بین اطلاعات مفید مورد نیاز برای انتقال از رمزگذار به بلوک‌های رمزگشا و اطلاعات غیر مفید دشوار است. برای غلبه بر این متصور، ما مکانیزم GA را پیشنهاد کردیم که برای تولید یک تصویر با استخراج نقشه‌های ویژگی مهم مطابق با زمینه تصویر مفید است. نقشه‌های ویژگی به دست آمده از واحد توجه حاوی اطلاعات زمینه‌ای هستند و با نقشه ویژگی لایه رمزگشایی مربوطه که اتصالات ترکیبی مبتنی بر توجه را تشکیل می‌دهند، الحاق می‌شوند. این اتصالات ترکیبی از هر دو ویژگی با وضوح بالا و پایین استفاده می‌کنند و در حین انجام عملیات نمونه برداری بر روی مرتبط ترین اطلاعات تمرکز می‌کنند.

بلوک رمزگشا: هر لایه در بلوک‌های رمزگشا از دو بلوک پیچشی متوالی با اندازه هسته 3×3 تشکیل شده است. هسته 3×3 در سراسر تصویر پخش می‌شود و ویژگی‌های عمومی بیشتری را استخراج می‌کند و اجزای اصلی تصویر را به تصویر می‌کشد. یک لایه unpooling جایگزین عملیات ادغام حداکثر می‌شود و نقشه‌های ویژگی ورودی را بدون استفاده از هیچ پارامتر قابل یادگیری اضافی نمونه برداری می‌کند. لایه unpooling از شاخص‌های ذخیره شده استفاده می‌کند در حالی که حداکثر ادغام بلوک رمزگذار مربوطه را انجام می‌دهد تا بعد فضایی نقشه ویژگی ورودی را نمونه برداری کند.

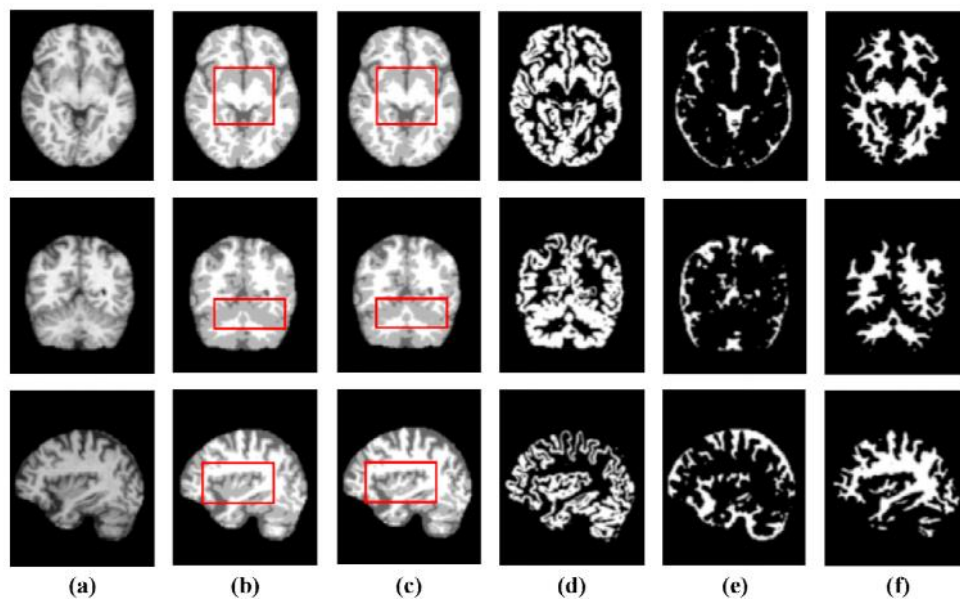
نقشه‌های ویژگی جمع‌نشده به واحد GA منتقل می‌شوند و با نقشه‌های ویژگی رمزگذار با ابعاد فضایی یکسان با استفاده از اتصالات ترکیبی ترکیب می‌شوند. این اتصالات ترکیبی اطلاعات زمینه‌ای بالایی را برای پشتیبانی از بخش‌بندی و ایجاد یک مسیر بدون مقاومت برای عبور گرادیان‌ها از مناطق عمیق‌تر به مناطق کم‌عمق‌تر ارائه می‌کنند که روند آموزش شبکه را بهبود می‌بخشد.

لایه طبقه‌بندی: لایه رمزگشای نهایی از یک لایه پیچشی 1×1 با فعال سازی softmax تشکیل شده است که نقشه تقسیم‌بندی بازسازی شده را پیش‌بینی می‌کند. خروجی شامل چهار کلاس پیش‌بینی شده است: WM، GM، CSF، و پس زمینه. مدل پیشنهادی تصویر ورودی را می‌پذیرد و نمایش آموخته شده مربوطه را تولید می‌کند. بر اساس این نمایش ویژگی، تصویر ورودی به هر یک از چهار کلاس خروجی طبقه‌بندی شد. برای بهینه سازی تابع هزینه مدل شبکه پیشنهادی، از تابع تلفات متقاطع آنتروپی استفاده کردیم. علاوه بر این، ما از لایه طبقه‌بندی‌کننده softmax استفاده کردیم که نمایش را از رمزگشا یاد می‌گیرد و آنها را به کلاس خروجی تفسیر می‌کند. امتیاز احتمال $f[i]$ به کلاس خروجی برای تقسیم‌بندی بافت‌های مغز اختصاص داده شد.

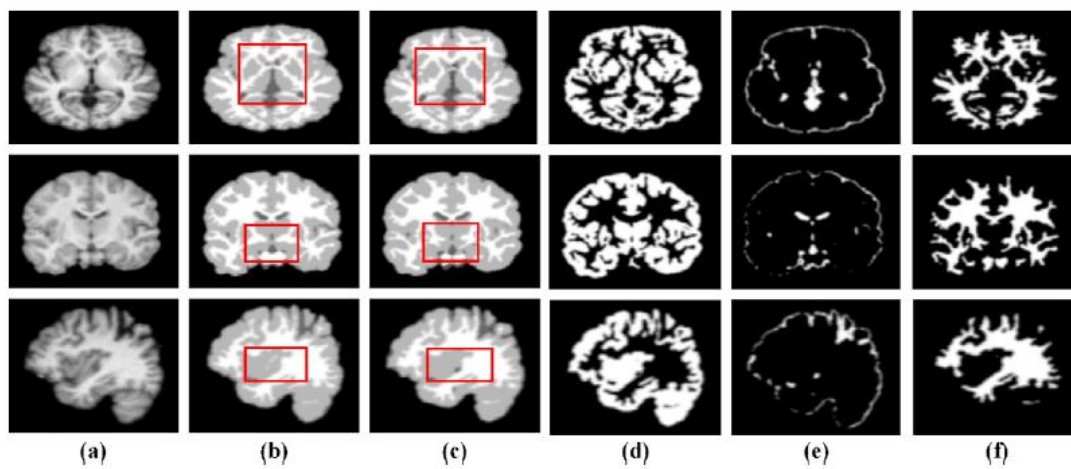
۴-۵: آزمایشات صورت گرفته

ما روش پیشنهادی خود را بر روی دو مجموعه داده تقسیم‌بندی MRI مغز ارزیابی کردیم: مجموعه‌ای از مطالعات تصویربرداری با دسترسی باز و مجموعه‌های داده مخزن تقسیم‌بندی مغز اینترنتی. ابرپارامترها و طرح‌ها دقیقاً برای هر دو مجموعه داده انتخاب شدند. در طول آموزش، ما نزول گرادیان تصادفی را با تکانه 0.001 ، نرخ یادگیری 0.001 ، تقسیم اعتبار 0.2 ، و تعدادی دوره از 10 اتخاذ کردیم. برای ارزیابی اثربخشی روش پیشنهادی، از ضریب تشابه دایس، شاخص جاکارد و فاصله هاسدورف برای مقایسه خروجی‌های تقسیم‌بندی با حقایق زمینی استفاده کردیم. نتایج تقسیم‌بندی برای مجموعه داده‌های OASIS و IBSR در تصویرها نشان داده شده است. به ترتیب تصویر ۵-۳ و ۴-۵. همانطور که نشان داده شد، رویکرد پیشنهادی عملکردهای تقسیم‌بندی خوبی برای WM، GM، و CSF MRI مغز در هر دو مجموعه داده به دست می‌آورد.

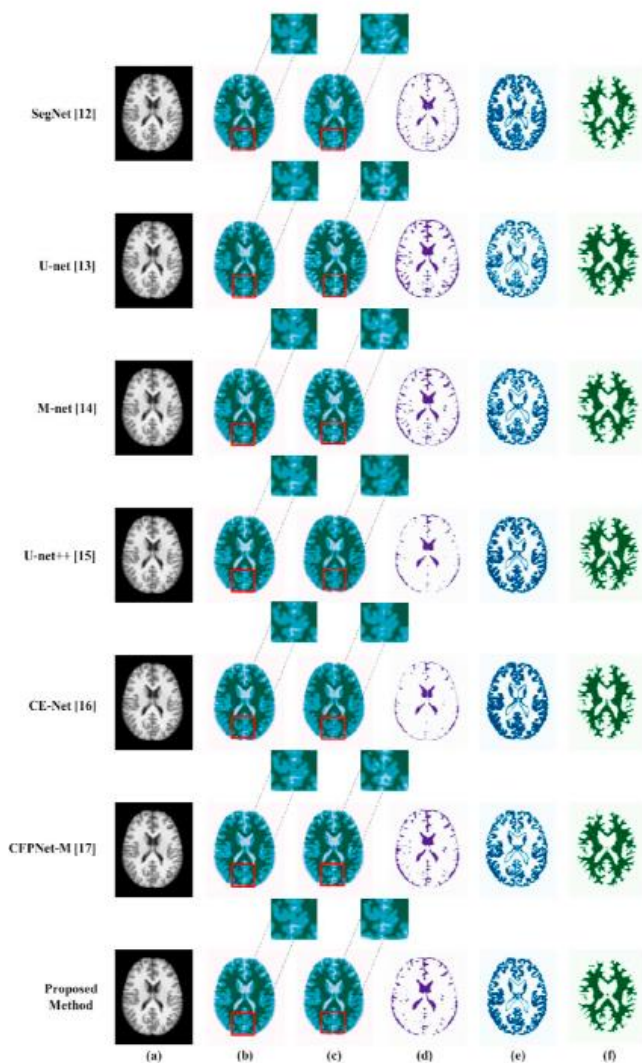
برای مقایسه نتایج تقسیم‌بندی معماری‌های شبکه مختلف، آزمایش‌هایی را بر روی مدل‌های SegNet، U-net، M-net، CE-Net، U-net++، و CFPNet-M تحت شرایط آزمایشی مشابه انجام دادیم. ۵-۵ و ۶-۵ مقایسه‌ای از نتایج تقسیم‌بندی را نشان می‌دهد. همانطور که نشان داده شده است، روش پیشنهادی نتایج برتر از نظر کیفیت نقشه تقسیم‌بندی را در مقایسه با سایر روش‌های مرسوم نشان می‌دهد. جداول ۵-۱ و ۵-۲ تجزیه و تحلیل کمی روش پیشنهادی را در مقایسه با روش‌های SegNet، U-net، M-net، U-net++، CE-Net، و CFPNet-M در مجموعه داده‌های OASIS و IBSR نشان می‌دهد. می‌توان مشاهده کرد که روش پیشنهادی در مقایسه با روش‌های موجود از نظر معیارهای DSC، JI و HD به دقت تقسیم‌بندی بهتری دست می‌یابد.



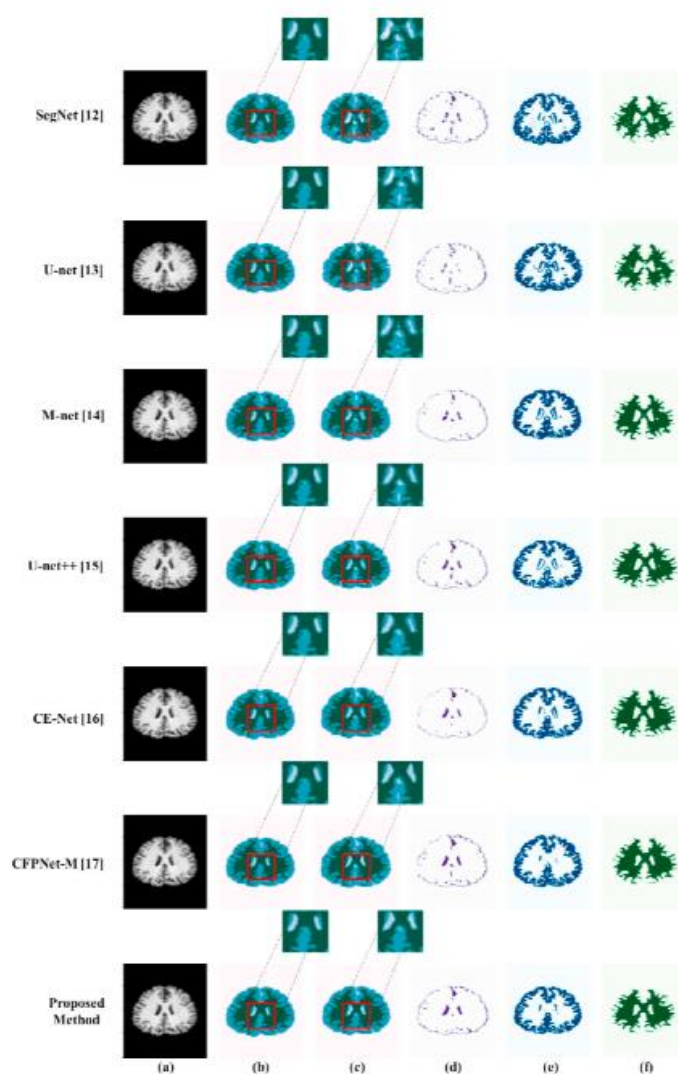
تصویر ۳-۵ : نتیجه بررسی در مجموعه داده OASIS



تصویر ۴-۵ : نتیجه بررسی در مجموعه داده IBSR



تصویر ۵-۵ : نتیجه بررسی در مجموعه داده OASIS



تصویر ۵-۶ : نتیجه بررسی در مجموعه داده IBSR

۵-۵ : نتیجه گیری

در این مقاله، ما M-SegNet را با معماری GA برای تقسیم‌بندی MRI مغز پیشنهاد کردیم. در معماری پیشنهادی، از لایه‌های ورودی جانبی چند مقیاسی برای پشتیبانی از لایه‌های عمیق برای استخراج اطلاعات متمایز استفاده می‌شود و لایه فرا نمونه‌گیری در سمت رمزگشا نظارت عمیق را فراهم می‌کند. طرح GA پیشنهادی یک زمینه جهانی را به عنوان مرجعی برای ویژگی‌های سطح پایین برای انتخاب اطلاعات محلی‌سازی دسته‌ها فراهم می‌کند. علاوه بر این، هسته‌های پیچشی گشاد شده چند مقیاسی در بلوک‌های رمزگذار-رمزگشا اطلاعات متمایز را در مقیاس‌های مختلف استخراج می‌کنند. اتصالات ترکیبی برای انتقال ویژگی‌ها از رمزگذار به مسیر رمزگشا برای بازیابی اطلاعات مکانی از دست رفته در حین نمونه برداری پایین استفاده می‌شود. علاوه بر این، استفاده از وصله‌های ورودی یکنواخت به ثبت جزئیات دقیق محلی کمک می‌کند. در نهایت، روش ما با توجه به معیارهای ارزیابی مانند DSC و JI برای تقسیم‌بندی MRI مغز به مناطق CSF، GM، و WM بهبود قابل‌توجهی نشان داد و دارای مقادیر متوسط DSC و JI به ترتیب ۰/۹۶ و ۰/۹۲ بود.

فصل ششم

معماری FA-Segnet

(Yang, YU , Yin&... , 2022)

۱-۶ : مقدمه

با هدف تقسیم‌بندی تصاویر پزشکی با تشخیص کم و نویز پس‌زمینه بالا، یک مدل تقسیم‌بندی تصویر شبکه عصبی پیچشی عمیق بر اساس مکانیسم توجه فازی^۱ پیشنهاد شده است که FA-SegNet نامیده می‌شود. SegNet را به عنوان چارچوب اصلی می‌گیرد. در واحد فرو نمونه برداری برای استخراج ویژگی تصویر، یک واحد توجه کانال فازی اضافه شده است تا تشخیص مناطق هدف مختلف را تقویت کند. در واحد فرا نمونه‌برداری برای بازیابی اندازه تصویر و ادغام ویژگی‌های چند مقیاسی، یک واحد توجه فضایی فازی اضافه می‌شود تا از دست رفتن جزئیات تصویر را کاهش دهد و حوزه دریافتی را گسترش دهد. در این مقاله، شناخت فازی به ترکیب ویژگی‌های شبکه عصبی پیچشی معرفی می‌شود. بر اساس مکانیسم توجه، عضویت فازی برای کالیبراسیون مجدد اهمیت مقدار پیکسل در مناطق محلی استفاده می‌شود. این می‌تواند توانایی تمایز ویژگی‌های تصویر، و توانایی ترکیب اطلاعات متنی را تقویت کند، که دقت تقسیم‌بندی مناطق هدف را بهبود می‌بخشد. با در نظر گرفتن قطعه بندی MRI به عنوان یک مثال تجربی، اهداف متعددی مانند بطن چپ، بطن راست و میوکارد بطن چپ به عنوان اهداف تقسیم بندی انتخاب می شوند. دقت پیکسل ها ۹۲/۴۷ درصد، میانگین تقاطع تا اتحاد ۸۶/۱۸ درصد و ضریب تاس ۹۲/۴۴ درصد است که در مقایسه با روش های دیگر بهبود یافته اند. این دقت و کاربرد روش پیشنهادی را برای تقسیم‌بندی تصاویر پزشکی، به‌ویژه اهدافی با تشخیص کم و انسداد جدی تأیید می‌کند.

۲-۶ : معرفی معماری

منطق فازی با تقلید از طرز تفکر و استدلال نامطمئن مغز انسان و با کمک مفاهیمی مانند تابع عضویت، به وابستگی و همبستگی بین روابط فازی می پردازد. برای تقسیم‌بندی تصویر MRI، واحد توجه فازی بر روی کانال ویژگی در مرحله فرو نمونه برداری (رمزگذاری) و ناحیه فضایی در مرحله فرا نمونه برداری (رمزگشایی) تعبیه شده است و اهمیت ویژگی‌ها مجدداً کالیبره می‌شود.

واحد توجه فازی در کانال های ویژگی لایه نمونه برداری معرفی شده است. ابتدا موقعیت جاسازی لایه فازی تعیین می شود. بگذارید بعد داده های ورودی (C, W, H, N) باشد و دو عملیات اتصال کامل را در واحد توجه کانال ویژگی انجام دهید. برای کنترل وزن پارامتر همبستگی بین کانال های مشخصه خروجی از طریق درجه عضویت تعیین شده توسط محاسبه منطق فازی، لایه مکانیزم فازی بین دو لایه اتصال کامل تعبیه شده است، یعنی خروجی اولین اتصال کامل به عنوان ورودی لایه فازی و خروجی لایه فازی به عنوان ورودی دومین لایه اتصال کامل استفاده می شود. تابع نوع S با ویژگی عضویت فازی به عنوان تابع عضویت لایه فازی پذیرفته شده است:

$$\mu(x) = \frac{1}{1 + e^{ax+b}}.$$

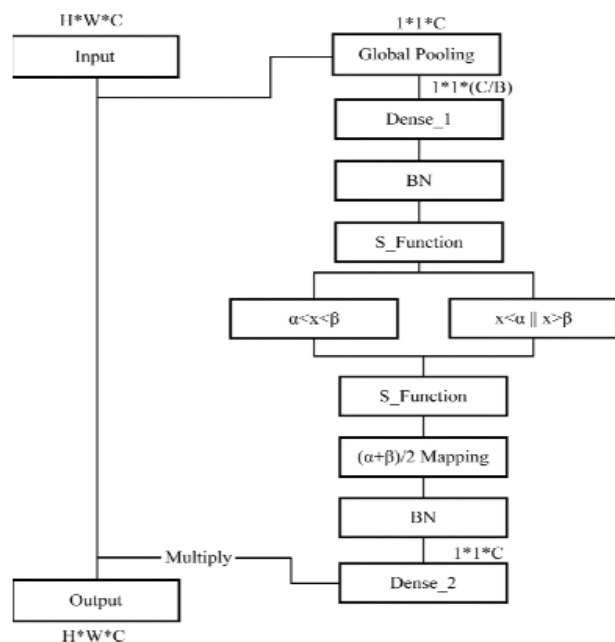
وزن کانال های ویژگی را با قطعیت بالا تقویت کنید، در حالی که وزن های قطعیت پایین را بدون تغییر نگه دارید تا میزان محاسبه کاهش یابد. با در نظر گرفتن بهبود توانایی غیرخطی عدم دقت تابع عضویت، نتیجه فرمول بالا دوباره وارد تابع عضویت نوع S می شود تا یک مقدار عضویت جدید بدست آید که سپس با ۵٪ تعریف می شود.

1. Fuzzy attention mechanism

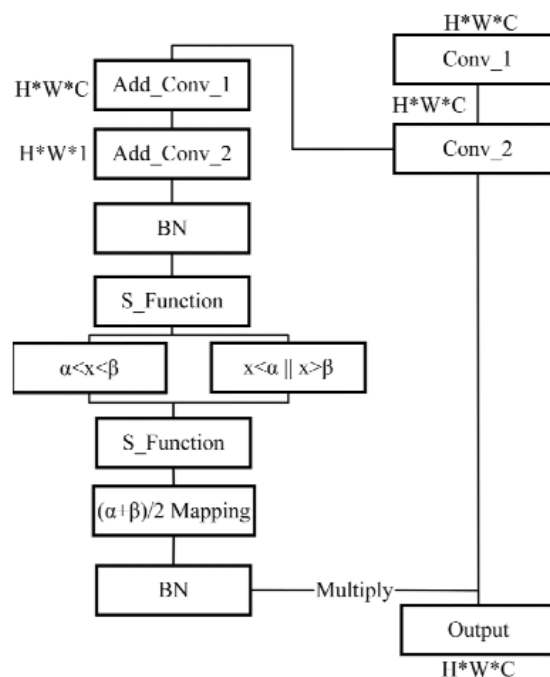
مقدار عضویت نزدیک به ۵/۰ قطعیت پائینی است. از طریق مجموعه ای از تغییر تصویر، وزن حاشیه نویسی کمتری از اهمیت کانال ویژگی دارد. دو طرف دور از ۵/۰ بسیار قطعی هستند، که باعث می شود وزن حاشیه نویسی اهمیت کانال ویژگی بیشتر شود. در نهایت، داده های وزن حاشیه نویسی با اندازه گیری عضویت با روش نرمال سازی دسته ای و خروجی نرمال می شوند. در واحد فرو نمونه برداری، یک واحد توجه کانال فازی برای تقویت تمایز مناطق مختلف هدف اضافه شده است. در تصویر ۶-۱ نشان داده شده است.

برای واحد توجه فضایی لایه فرا نمونه برداری، عدم قطعیت عضویت به عنوان وزن برای ارزیابی اهمیت پیکسل ها استفاده می شود. وزن پیکسل های با عدم قطعیت کم نقش مهمتری ایفا می کند، در حالی که وزن پیکسل های با عدم قطعیت بالا را کاهش می دهد. در جریان پردازش، لایه توجه فضایی فازی پس از آخرین عملیات پیچشی معماری فرا نمونه برداری و قبل از ضرب عددی با ماتریس نقشه ویژگی اصلی تعبیه شده است. در واحد توجه فضایی فازی، ویژگی های کم عمق فرو نمونه برداری با ویژگی های عمیق فرا نمونه برداری ترکیب می شوند تا به طور مؤثرتری ویژگی های مناطق مختلف را متمایز کنند.

در واحد فرا نمونه برداری، یک واحد توجه فضایی فازی اضافه می شود تا از دست رفتن ویژگی های جزئیات تصویر را کاهش دهد و حوزه دریافتی را گسترش دهد. مکانیسم واحد توجه فضایی فازی در تصویر ۶-۲ نشان داده شده است. در بخش بندی معنایی تصویر بر اساس FA-Seg-Net، برای نقشه ویژگی با اندازه یکسان، باید از دو مکانیسم توجه مختلف عبور کند تا همبستگی بین ویژگی ها را محدود کند. متفاوت از مدل اصلی SegNet، که تنها ویژگی ها را از طریق فعال سازی پیچشی برای ارائه عملیات غیرخطی استخراج می کند، مکانیسم توجه فازی نه تنها می تواند کانال های مشخصه نامشخص و وزن های حاشیه نویسی منطقه فضایی را ارائه کند، بلکه اطلاعات زمینه فعال شده با پیچیدگی اصلی را نیز حفظ می کند. این می تواند اطلاعات ویژگی های منطقه و لبه را در وظیفه تقسیم بندی بهتر استخراج کند و تداخل اطلاعات مستقل از کار را سرکوب کند.



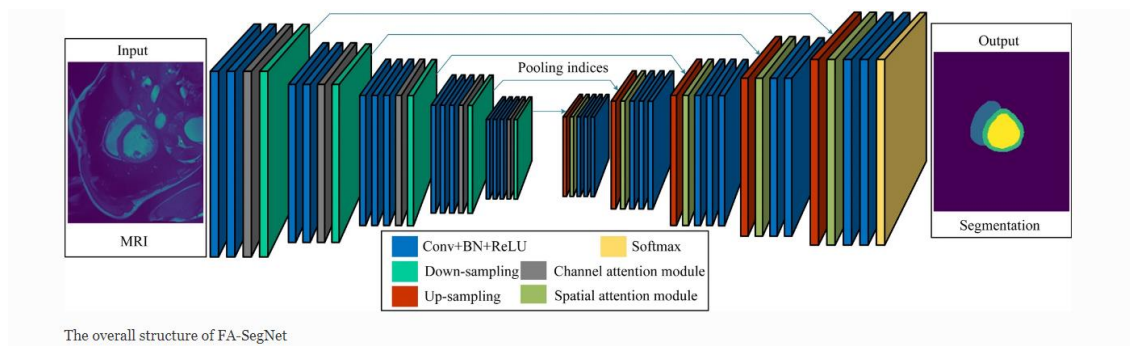
تصویر ۶-۱: واحد توجه کانال فازی



تصویر ۶-۲: واحد توجه فضایی فازی

بر اساس مکانیسم توجه، عضویت فازی برای کالبراسیون مجدد اهمیت مقدار پیکسل هر ناحیه محلی استفاده می شود. این می تواند توانایی تشخیص ویژگی های هدف تصویر، و همچنین توانایی ترکیب اطلاعات متنی و

محتوای متنوع تصاویر را تقویت کند و دقت تقسیم بندی مدل از مناطق هدف را بهبود بخشد. ساختار کلی FA-SegNet در تصویر ۳-۶ نشان داده شده است.



تصویر ۳-۶ : ساختار کلی FA-SegNet

جریان پردازش اطلاعات مدل FA-SegNet به شرح زیر است:

(۱) لایه ورودی: تصویر C-channel با اندازه $H \times W$. (۲) استخراج ویژگی تصویر پایین نمونه، یعنی مرحله رمزگذاری. در مرحله رمزگذاری، ویژگی های تصویر با پیچشی استخراج می شوند، از جمله ۵ واحد رمزگذار. هر واحد رمزگذار شامل لایه پیچشی، واحد توجه کانال فازی و لایه ادغام است. همه لایه های پیچشی یک پیچشی را اتخاذ می کنند، یعنی اندازه اصلی تصویر پس از عملیات پیچشی حفظ می شود و عملیات استانداردسازی دسته ای پس از هر عملیات پیچیدگی انجام می شود و توسط تابع ReLU فعال می شود. اجازه دهید نقشه ویژگی ورودی لایه پیچشی $X = [x_1, x_2, \dots, x_n]$ و هسته پیچشی $V = [v_1, v_2, \dots, v_C]$ و سپس خروجی $U = [u_1, u_2, \dots, u_C]$ عملیات پیچشی عبارت است

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s, \quad \text{از:}$$

که در آن $*$ عمل پیچشی است و v_c^s هسته پیچشی s مربوط به ورودی X است.

هر رمزگشا شامل ۲ یا ۳ لایه پیچشی متوالی است و نقشه ویژگی پس از عملیات پیچشی به واحد توجه کانال فازی وارد می شود. اول، اطلاعات کلی هر کانال با ادغام میانگین جهانی نقشه ویژگی، یعنی 1×1 عملیات پیچشی برای هر کانال به دست می آید. به طور رسمی، یک آماره $z \in \mathbb{R}^C$ با کوچک کردن U از طریق ابعاد فضایی $H \times W$ ایجاد

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j).$$

می شود، که در آن عنصر c -امین z به صورت زیر محاسبه می شود:

سپس، میانگین اطلاعات جهانی ادغام شده باید به ترتیب دو بار به طور کامل متصل شود و عملیات فازی سازی بین دو اتصال کاملاً انجام شود. عدم قطعیت تعیین شده توسط تابع عضویت منطق فازی، خروجی وزن پارامتر همبستگی بین کانال های ویژگی را کنترل می کند، یعنی خروجی اولین اتصال کامل به عنوان ورودی لایه فازی استفاده می شود، خروجی لایه فازی استفاده می شود. به عنوان ورودی دومین لایه اتصال کامل. فرم خاص این

$$s = \sigma(\text{Fuzzy}(g(z, W))) = \sigma(W_2 \text{Fuzzy}(\delta(W_1 z))),$$

است:

که در آن δ تابع ReLU است، فازی عملیات فازی سازی را نشان می دهد و σ تابع Sigmoid است. $W1 \in \mathbb{R}^{C \times C}$ و $W2 \in \mathbb{R}^{C \times C_r}$ به ترتیب مجموعه وزنی دو لایه کاملاً اتصال است که r نشان دهنده ضریب کاهش است که برای کاهش تعداد گره ها در لایه اتصال کامل و مقدار پارامترهای مدل استفاده می شود.

برای فرآیند فازی سازی، ابتدا داده های ورودی فازی می شوند و مقدار عضویت از طریق تابع عضویت نوع S محاسبه می شود. محدوده مقدار (0، 1) است. برای بهبود توانایی غیر خطی عدم دقت تابع عضویت، نتیجه فرمول فوق مجدداً وارد تابع عضویت نوع S می شود تا مقدار عضویت جدید به دست آید. در نهایت، داده های وزن حاشیه نویسی کانال ویژگی با اندازه گیری عضویت با روش نرمال سازی دسته ای و خروجی نرمال می شود. فرمول محاسبه به صورت زیر است:

$$z'' = \text{BN}(S(S(\text{BN}(z')))),$$

که در آن BN مخفف عملیات عادی سازی Batch Normalization و S مخفف S-type تابع است.

پس از دو عملیات اتصال کامل و فازی سازی، در نهایت وزن s هر کانال از نقشه ویژگی، یعنی درجه توجه به دست می آید. در مقادیر نقشه ویژگی مربوطه ضرب می شود و در نهایت اهمیت هر کانال ویژگی با وزن های مختلف

$$\tilde{x}_c = s_c \cdot u_c,$$

دوباره برچسب گذاری می شود. فرمول محاسبه خاص این است:

لایه Pooling در انتهای هر رمزگذار اضافه می شود و حداکثر ادغام با اندازه 2×2 انتخاب می شود که وضوح نقشه ویژگی را به نصف اصلی کاهش می دهد. حداکثر فرمول ادغام لایه:

$$z_p^l(i, j) = \max(a_p^{l-1}(2i - u, 2j - v)) \quad u, v \in 0, 1.$$

بخش رمزگشایی نیز شامل پنج واحد مربوط به مرحله رمزگذاری است. هر واحد رمزگشا شامل لایه فرا نمونه برداری، واحد توجه فضایی فازی و لایه پیچشی است. ورودی نقشه ویژگی به رسیور از طریق یک لایه نمونه برداری با اندازه 2×2 از نظر وضوح بهبود یافته است. با مطابقت با مقدار شاخص ثبت شده در حین نمونه برداری پایین در رمزگذار، داده های نقشه ویژگی ورودی در موقعیت اصلی قرار می گیرند و سایر موقعیت ها با 0 پر می شوند. عملیات نمونه برداری در لایه 1 به صورت زیر تعریف می شود:

$$z_p^l(i, j) = \begin{cases} a_p^{l-1}(\tilde{i}, \tilde{j}) & (i, j) \in \text{indices} \\ 0 & (i, j) \notin \text{indices} \end{cases}$$

نقشه ویژگی پس از فرا نمونه برداری، برای فازی سازی و تنظیم توجه فضایی به واحد توجه فضایی فازی وارد می شود. متفاوت از واحد توجه کانال، لایه فازی پس از آخرین عملیات پیچشی و قبل از ضرب عددی با ماتریس نقشه ویژگی اصلی تعبیه شده است. ابتدا، نقشه ویژگی به طور متوسط در جهت کانال جمع می شود، یعنی عملیات پیچیدگی $H \times W \times 1$ ، برای به دست آوردن اطلاعات کانال سراسری نقشه ویژگی.

$$z(i, j) = \frac{1}{C} \sum_{c=1}^C u_c(i, j).$$

سپس، نقشه ویژگی تلفیقی به همان روشی که در بخش قبل انجام شد فازی می شود تا ماتریس وزن اهمیت منطقه فضایی نهایی به دست آید. مقادیر در ماتریس درجات مختلفی از توجه به پیکسل ها در هر موقعیت مکانی

را نشان می دهد. در نهایت، ماتریس توجه ناحیه فضایی در ورودی اصلی در مرحله رمزگشایی ضرب می شود تا مجدداً اهمیت مقدار موقعیت پیکسل اصلی در موقعیت پیکسل مربوط به منطقه فضایی تکمیل شود. فرمول محاسبه به صورت زیر است:

$$\tilde{x}(i, j) = u(i, j) \times z(i, j).$$

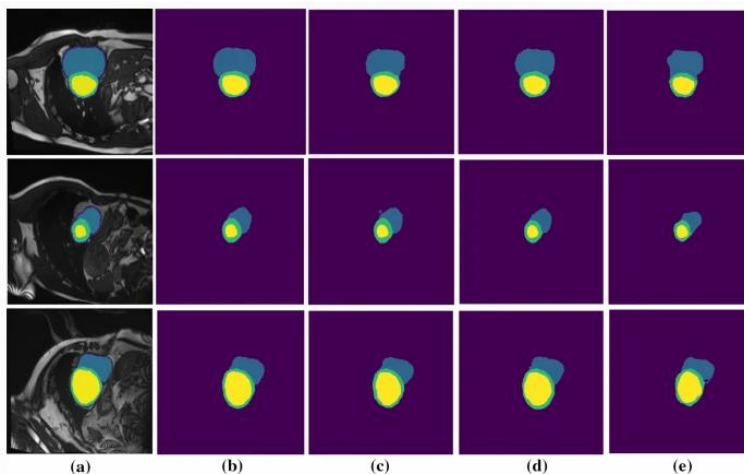
قسمت آخر یک لایه Softmax است که احتمال تعلق هر پیکسل تصویر به هر دسته را محاسبه می کند. دسته بندی مربوط به حداکثر مقدار، برچسب پیکسل است که طبقه بندی تصویر در سطح پیکسل را کامل می کند. تابع فعال سازی لایه خروجی Softmax است:

$$a_i^L = \text{softmax}(x_i^L) = \frac{e^{x_i^L}}{\sum_{k=1}^n e^{x_k^L}}.$$

در آموزش FA-SegNet از تلفات آنروپی متقاطع به عنوان تابع ضرر و از الگوریتم انتشار برگشتی برای یادگیری پارامترهای شبکه استفاده می شود.

۳-۶ : آزمایشات صورت گرفته

هنگام ارزیابی عملکرد FA-SegNet پیشنهاد شده در این مقاله، معیارهای ارزیابی تقسیم بندی تصویر عبارتند از: دقت پیکسل (PA)، دقت پیکسل کلاس (CPA)، دقت میانگین پیکسل (MPA)، ضریب تاس، تقاطع بر روی اتحاد (IOU) فاصله هاسدورف (HD). نتایج کیفی چهار روش مختلف در تصویر ۴-۶ نشان داده شده است که در آن (الف) تصویر اصلی MRI با علائم تقسیم بندی استاندارد است. (ب) نتیجه تقسیم بندی FA-SegNet است که در این مقاله پیشنهاد شده است. (ج) نتیجه تقسیم بندی مدل اصلی SegNet است. (د) نتیجه تقسیم بندی مدل U-net است. (ه) نتیجه تقسیم بندی مدل شبکه کاملاً پیچشی است.



تصویر ۴-۶ : نتایج کیفی چهار روش مختلف

از تصویر ۴-۶، از مقایسه بین چهار روش و نتایج تقسیم بندی با علائم تقسیم بندی استاندارد می توان دریافت که در اغلب موارد، نتایج مدل تقسیم بندی پیشنهادی در این مقاله با نتایج تقسیم بندی استاندارد مطابقت خوبی دارد. نتایج در جدول ۱-۶ نشان داده شده است.

Evaluating indicator	PA	Dice coefficient	IoU	Hausdorff distance/mm
Left ventricle	0.9727	0.9580	0.9203	9.151
Left ventricular myocardial	0.9207	0.9098	0.8359	9.824
Right ventricle	0.8808	0.9055	0.8291	11.365
Average	0.9247	0.9244	0.8618	9.78

جدول ۱-۶

از هر شاخص ارزیابی، می توان دریافت که قطعه بندی FA-SegNet بر روی تصویر MRI قلب نتایج خوبی به دست آورده است. نتایج شاخص های مختلف ارزیابی در جدول ۲-۶ نشان داده شده است.

Model	PA	MPA	MIoU
FCN	0.8740	0.8534	0.6340
U-net	0.9372	0.9216	0.8460
SegNet	0.9022	0.9157	0.8165
Proposed	0.9475	0.9247	0.8618

جدول ۲-۶

از نتایج تجربی جدول ۲-۶ می توان دریافت که FA-SegNet بهترین نتایج را در PA، MPA و MIoU در مقایسه با سایر الگوریتم ها به دست می آورد. به طور خاص، در مقایسه با روش SegNet اصلی، مدل پیشنهادی ۰/۰۴، ۰/۰۱ و ۰/۰۵ به ترتیب در PA، MPA و MIoU بالاتر است، که ثابت می کند که واحد توجه فازی پیشنهادی تأثیر خوبی در استخراج ویژگی عمیق دارد.

۴-۶ : نتیجه گیری

در این مقاله، با هدف مسئله تقسیم بندی تصویر پزشکی، مدل تقسیم بندی FA-SegNet که مکانیسم توجه فازی را به SegNet معرفی می کند، بر اساس ایده علوم شناختی و یادگیری عمیق پیشنهاد می شود. با اعمال شناخت فازی به واحد مکانیسم توجه، مفاهیم معنایی مانند تابع عضویت و عدم قطعیت با درجه توجه کانال های ویژگی و مناطق فضایی در مکانیسم توجه ترکیب می شوند تا محدودیت های دانش حوزه بر روی نقشه های ویژگی یا تقسیم بندی پیکسل را درک کنند. نتایج تجربی نشان می دهد که FA-SegNet تا حد زیادی اثر بخش بندی و دقت مدل را بهبود می بخشد و کاربرد خوبی برای اهداف کوچک و بزرگ دارد. در مقایسه با سایر مدل ها، شاخص های ارزیابی بسیار بهبود یافته اند. FA-SegNet از مزایای مکانیسم توجه فازی و شبکه پیچشی عمیق استفاده جامع می کند. در مکانیزم پردازش اطلاعات، می تواند اطلاعات زمینه فعال شده توسط عملیات پیچشی را حفظ کند، اطلاعات ویژگی های منطقه ای و لبه ای را در وظیفه تقسیم بندی بهتر استخراج کند، و تداخل اطلاعات مستقل از کار را سرکوب کند. می تواند یک روش یادگیری عمیق جدید برای تقسیم بندی تصویر ارائه دهد و چشم انداز خوبی برای محبوبیت و کاربرد دارد. برای کارهای آینده، این مطالعه را می توان گسترش داد و برای پردازش تصاویر پزشکی سه بعدی استفاده کرد که می تواند تغییرات در ساختار و مورفولوژی اندام ها و بافت ها را در کلینیک بهتر منعکس کند.

منابع و مراجع

[١]. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", 2017

[٢]. "P_Segnet and NP_Segnet: New Neural Network Architectures for Cloud Recognition of Remote Sensing Images", 2019

[٣]. "A Deep-Learning-Based Multiple Defect Detection Method for Tunnel Lining Damages", 2019

[٤]. "A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI ",Yamanakkanavar and Lee, 2021

[٥]. " An FA-SegNet Image Segmentation Model Based on Fuzzy Attention and Its Application in Cardiac MRI Segmentation." , Yang, R., Yu, J., Yin, J. et al. , 2022