نشریه علمی مهندسی فناوری اطلاعات مکانی

سال یازدهم، شماره سوم، پاییز ۱۴۰۲
Vol.11, No.3, Autumn 2023
۴۳ -۵۷
مقاله پژوهشی

ارزیابی قابلیت شبکه رمزگذار-رمزگشای *DeepLabV3+* با پیچش‌های آتروس اصلاح شده (مطالعه موردی: قطعه بندی معنایی ساختمان)

محمدعرفان امتی[۱]، فاطمه طبیب محمودی[۲*]

۱- دانشجوی کارشناسی ارشد سنجش از دور، گروه مهندسی نقشه‌برداری، دانشکده عمران، دانشگاه تربیت دبیر شهید رجایی
۲- استادیار گروه مهندسی نقشه‌برداری، دانشکده عمران، دانشگاه تربیت دبیر شهید رجایی

تاریخ دریافت مقاله: ۱۴۰۲/۰۴/۲۲ تاریخ پذیرش مقاله: ۱۴۰۲/۰۷/۱۱

چکیده

قطعه‌بندی ساختمان‌ها به دلیل نیاز به ویژگی‌های معنایی غنی کار دشواری است. تفاوت در شکل، رنگ و اندازه ساختمان‌ها و نزدیکی آن‌ها به سایر عوارض مانند پارکینگ‌ها و خیابان‌ها تشخیص آنها را در تصاویر با وضوح زیاد با چالش‌هایی روبرو می‌سازد. در این تحقیق با هدف استخراج ساختمان از تصاویر با وضوح زیاد، از یک معماری شبکه عصبی پیچشی عمیق از نوع رمزگذار-رمزگشا مبتنی بر مدل اصلاح شده *DeepLabV3+* استفاده شده است. در ماژول آتروس این مدل اصلاح شده، لایه‌های پیچش با نرخ‌های کمتری در مقایسه با ماژول اصلی، اعمال شده و از پیچش گسترده به جای پیچش استاندارد استفاده گردید تا هدف دستیابی به قطعه‌بندی معنایی قدرتمندتر عوارض ساختمانی با اندازه کوچک و بزرگ محقق گردد. قابلیت اجرایی مدل پیشنهادی در این تحقیق با استفاده از دو مجموعه داده *WHU* و *INRIA* ارزیابی گردید و نتایج بدست آمده نشان داد که استفاده از نرخ‌های آتروس کمتر و تغییر آنها به ۴، ۸ و ۱۲به‌طور قابل‌توجهی عملکرد قطعه‌بندی را در هردو مجموعه داده بهبود بخشید. مدل اصلاح شده پیشنهادی توانست شاخص‌های بازخوانی، *IOU* و امتیاز-اف را در مجموعه داده *WHU* نسبت به سایر مدل‌های پیشرفته به ترتیب به میزان۰٫۳۳، ۰٫۳۹ و ۰٫۵۳ بهبود بخشد. به علاوه، روش اصلاح شده در مجموعه داده *INRIA* توانست شاخص‌های فوق را نسبت به این مدل‌ها به ترتیب به میزان ۱٫۲۲، ۰٫۳۵ و ۰٫۳۵ بهبود بخشد. مدل پیشنهادی دراین تحقیق براساس کاهش نرخ‌های آتروس به ۴، ۸ و ۱۲ و تغییر در لایه‌های *ResNet-50* توانست در استخراج عوارض ساختمانی به*IOU* برابر با ۸۹٫۵۱ در مجموعه داده *WHU* و ۷۶٫۶۴ در مجموعه داده *INRIA* دست یابد. در حالیکه، مدل *DeepLabV3+* اصلی با نرخ‌های آتروس ۶، ۱۲، ۱۸ و نسخه اصلی *ResNet-50*، مقدار*IOU* برابر با ۸۸٫۸۷ را در مجموعه داده *WHU* و مقدار *IOU* برابر با ۷۵٫۸۲ را در مجموعه داده *INRIA* برای قطعه‌بندی ساختمان‌ها به دست آورد.

کلید واژه‌ها : قطعه بندی معنایی، شبکه عصبی پیچشی عمیق، رمزگذار،رمزگشا، پیچش آتروس.

* نویسنده مکاتبه کننده: تهران، لویزان، دانشگاه تربیت دبیر شهید رجایی.
تلفن: ۰۲۱۲۲۹۷۰۰۲۱

Email: fmahmoudi@sru.ac.ir

[DOI: 10.61186/jgit.11.3.43]

۱- مقدمه

با توجه به توسعه سریع سنجنده‌های مختلف در سال-های اخیر، تصاویر سنجش از دور با وضوح زیاد در بسیاری از کاربردها از جمله تشخیص و طبقه‌بندی عوارض شهری، تولید مدل‌های سه بعدی شهر و ارزیابی مخاطرات در مدیریت بحران به طور گسترده مورد استفاده قرار گرفته‌اند[۱،۲،۳،۴،۵،۶و۷]. ساختمان‌ها از جمله مهم‌ترین عناصر شهری هستند و شناسایی آنها با استفاده از روش‌های سنتی مبتنی بر تعیین ویژگی‌های دستی یا استخراج خودکار امکان پذیر است[۸، ۹و۱۰]. استخراج دستی ویژگی‌های ساختمان نیازمند نیروی انسانی متخصص و هزینه‌های زیاد است. با این حال، روش‌های سنتی فقط می‌توانند به اطلاعات سطح پایین یا متوسط دسترسی داشته باشند، زیرا به اندازه کافی هوشمند نیستند[۱۱و ۱۲]. تلاش‌های قابل توجهی برای توسعه و گسترش روش‌های استخراج خودکار برای دستیابی به ویژگی‌های سطح بالا انجام شده است[۱۳]. اگرچه محدودیت‌های روش‌های سنتی در این رویکرد به طور قابل توجهی حذف شده، ولی استخراج ساختمان‌ها به طور دقیق و کارآمد از تصاویر با وضوح زیاد همچنان چالش برانگیز است.

تفاوت در اندازه، رنگ و شکل ساختمان‌ها و نزدیکی آن‌ها به سایر عوارض مانند خیابان‌ها، پارکینگ‌ها و درختان چالش‌ها و مشکلاتی را در مدل‌سازی دقیق ویژگی‌های آن‌ها با استفاده از تصاویر ماهواره‌ای یا هوایی با وضوح زیاد ایجاد می‌کند. بنابراین، در استخراج ساختمان‌ها، به دست آوردن ویژگی‌های متمایز بین کلاسی مرتبط با خصوصیات یک ساختمان و تمایز بین ویژگی‌های درون کلاسی حیاتی است[۱۴].

فناوری یادگیری عمیق مجموعه کاملی از ابزارها برای انجام پردازش‌های مختلف از جمله طبقه‌بندی، قطعه بندی معنایی و سایر کاربردها مانند تجزیه و تحلیل تصاویر سنجش از دور است [۱۵ و ۱۶]. شبکه‌های عصبی پیچشی عمیق(*DCNN*)[۱] تکامل یافته شبکه‌های عصبی مصنوعی هستند که در سال‌های اخیر، رایج‌ترین روش‌های قطعه‌بندی معنایی تصاویر با وضوح زیاد براساس این نوع از شبکه‌ها طراحی شده‌اند [۱۷، ۱۸، ۱۹، ۲۰و ۲۱]. مزیت اصلی این الگوریتم‌ها این است که می‌توانند ویژگی‌های سطح بالا را بیاموزند و استخراج کنند که این امر منجر به بهبود عملکرد آنها شده و دقت قطعه‌بندی معنایی را بهبود می‌بخشد [۱۸]. برای افزایش دامنه دریافت ورودی از تصویر اصلی، این شبکه‌ها عملیات ادغام[۲]و پیچش را تکرار می‌کنند. این فرآیند منجر به دستیابی به ویژگی‌های معنایی عمیق می‌شود. با این وجود، نمونه برداری کاهشی، ابعاد تصویر اصلی را کاهش می‌دهد و جزئیات مکانی مهم را حذف می‌کند. در نتیجه قطعه‌بندی نامناسب با مرزهای نادرست ایجاد می‌شود. محققان با هدف کاهش از دست دادن جزئیات مکانی و تولید نتایج قابل اعتماد، انواع معماری رمزگذار-رمزگشا را توسعه داده‌اند [۱۸و۱۹]. این تکنیک موثر است اما تعداد پارامترهای قابل یادگیری شبکه را افزایش می‌دهد.

پیچش‌های آتروس[۳]می‌توانند میدان دریافت شبکه را گسترش دهند و در عین حال تعداد پارامترهای قابل یادگیری را کنترل کنند[۲۰ و۲۱]. شبکه کاملاً پیچشی[۴] نمونه‌ای از این نوع شبکه‌هاست[۲۲و ۲۳]. در ساختار این شبکه‌ها، لایه‌های کاملاً متصل با لایه‌های کانولوشن جایگزین می‌شوند تا نقشه‌های ویژگی حاوی اطلاعات موقعیتی باشند. در قطعه‌بندی معنایی مبتنی بر *DCNN*، با تکرار عملیات پیچش و فرآیند نمونه‌برداری کاهشی، ویژگی‌های معنایی در لایه‌های عمیق‌تر شبکه به‌دست می‌آیند. با این وجود، این فرآیند

[۱] *Deep Convolutional Neural Network*
[۲] *Pooling*
[۳] *Atrous Spatial Pyramid Pooling (ASPP)*
[۴] *Fully Convolutional Network (FCN)*

[DOI: 10.61186/jgit.11.3.43]

منجر به از دست دادن ویژگی‌های مکانی مهم می‌شود. برای بهبود نتایج استخراج ساختمان، ویژگی‌های معنایی غنی در لایه‌های عمیق با ویژگی‌های مکانی غنی در لایه‌های کم عمق از طریق یک عملگر الحاق[1] ترکیب می‌شوند تا نقشه ویژگی نهایی را با دقت بیشتری پیش‌بینی کنند. شبکه‌های *U-Net* [۱۹]، شبکه‌های ساعت شنی [۲۴] و شبکه مرجع[2][۲۵] از این ساختار خاص استفاده می‌کنند. مجموعه‌ای از مطالعات براساس شبکه‌های *U-Net* برای استخراج ساختمان‌ها از تصاویر با وضوح زیاد انجام شده است[۲۶، ۲۷، ۲۸، ۲۹، ۳۰ و ۳۱].

پن و همکاران (۲۰۱۹) با درنظر گرفتن تأثیرات وضوح زیاد تصاویر سنجش از دور بر ایجاد ابهام در نتایج قطعه‌بندی معنایی، یک شبکه مولد با مکانیزم‌های مکانی با نام *(GAN-SCA)* برای انجام قطعه بندی روباست ساختمان‌ها پیشنهاد دادند [۲۶]. با هدف بررسی تأثیر وضوح مکانی تصویر بر نتایج قطعه بندی معنایی ساختمان ها، گو و همکاران (۲۰۲۳) با استفاده از نمونه برداری افزایشی و کاهشی، تصاویر با وضوح مختلف از یک منطقه مطالعاتی را در شبکه *U-Net* مورد مطالعه قرار دادند و تأثیر وضوح مکانی را در نتایج خود ارائه نمودند [۳۲].

هنگام استخراج اجسام با ابعاد مختلف، استراتژی دیگر شامل بزرگ‌کردن میدان‌های گیرنده از طریق عملیات کانولوشن با نرخ‌های مختلف است. به عنوان نمونه، می‌توان به شبکه‌های *DeepLabv3+* [۳۳] و *PSPNet*[۳۴] در این زمینه اشاره کرد. جی و همکاران (۲۰۱۸) یک شبکه مقاوم در مقیاس با استفاده از ساختارهای *ASPP* برای استخراج ساختمان از تصاویر هوایی و ماهواره‌ای پیشنهاد کرد [۳۵]. ژو (۲۰۲۱) در کار خود تکنیکی را پیشنهاد کرد که یک رمزگذار و رمزگشا را براساس شبکه *DeepLabv3*، که شامل یک ماژول *ASPP* است، ترکیب می‌کند تا اطلاعات مرز ساختمان را بازیابی کند [۳۶]. وانگ و همکاران (۲۰۲۳) یک شبکه با نام *AFF-Unet* را با هدف بهینه سازی عملکرد قطعه بندی معنایی پیشنهاد دادند. این شبکه مشتمل بر اتصالات پرش متراکم و یک ماژول ادغام ویژگی است که به طور تطبیقی سطوح مختلف نقشه‌های ویژگی را برای دستیابی به ترکیب ویژگی تطبیقی وزن دهی می‌کند[۳۷].

تحقیق انجام شده در این مقاله با استفاده از مفهوم مدل *DeepLabV3+*، یک نسخه بهبود یافته از آن را ارائه می‌دهد که منجر به افزایش دقت استخراج ساختمان‌ها می‌شود. معماری *ResNet-50* از پیش آموزش‌یافته مبتنی بر عملیات پیچش گسترده [3] به‌منظور استخراج اطلاعات زمینه از تصاویر سنجش از دوری در این مدل استفاده شده است. معماری *ResNet* از بلوک‌های باقیمانده برای غلبه بر مسئله ناپدیدشدن و انفجار گرادیان‌ها استفاده می‌کند. در لایه‌های آخر ساختار معماری *ResNet-50* از لایه‌های پیچش گسترده به جای لایه‌های پیچش استاندارد با هدف افزایش وضوح نقشه استفاده شده‌است. بعلاوه ما در ماژول *ASPP*، لایه‌های پیچشی با نرخ‌های کمتری را در مقایسه با ماژول اصلی اعمال کردیم تا به هدف قطعه‌بندی معنایی قدرتمندتر عوارض ساختمانی با اندازه کوچک و متغیر دست یابیم. به منظور مشاهده اثربخشی چارچوب اصلاح شده پیشنهادی، عملکرد مدل اصلی *DeepLabV3+* در استخراج ساختمان‌ها نیز مورد ارزیابی قرار خواهدگرفت و کارایی مدل پیشنهادی در مقایسه با معماری‌های *U-Net* ، *PSPNet* و *DeepLabV3+*بررسی می‌شود.

با توجه به موارد فوق الذکر در مورد نقاط قوت و ضعف شبکه‌های عصبی پیچشی عمیق در قطعه‌بندی معنایی ساختمان‌ها، اهداف اصلی و نوآوری‌های مطرح در این

[1] *Concatenation*
[2] *Ref-Net*
[3] *Dilated convolutional*

[DOI: 10.61186/jgit.11.3.43]

مقاله به شرح ذیل است:

۱) یک شبکه کارآمد براساس معماری رمزگذار-رمزگشا برای قطعه‌بندی معنایی ساختمان‌ها طراحی شده است که در آن کارایی الگوریتم *DeepLabV3+* اصلاح‌شده به طوریکه از نرخ‌های پایین‌تر در ماژول *ASPP* استفاده می‌کند و در آن از پیچش گسترده به جای پیچش استاندارد استفاده شده تا هدف دستیابی به قطعه‌بندی معنایی قدرتمندتر عوارض ساختمانی با اندازه کوچک و بزرگ محقق گردد.

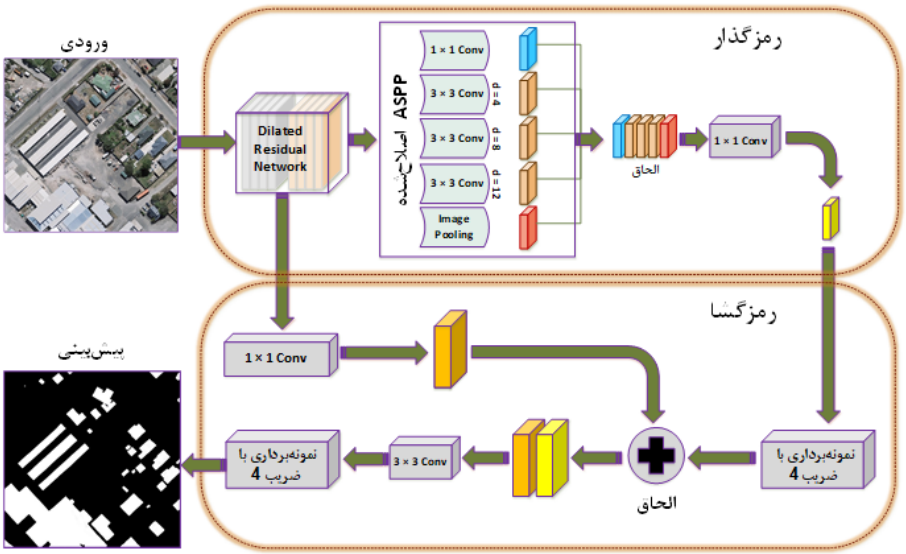
۲) اگرچه عملیات پیچش و ادغام در بخش رمزگذار اطلاعات غنی را فراهم می‌کنند، اطلاعات مرزی در مورد مناطق مورد نظر می‌تواند از بین برود. این مقاله نشان می‌دهد که رویکرد بهره‌وری از پیچش آتروس می‌تواند از دست دادن اطلاعات معنایی را به حداقل برساند.

۳) با استفاده از لایه‌های پیچش گسترده به جای پیچش استاندارد در لایه‌های آخر *ResNet50* و به دنبال آن تغییر ضریب نمونه‌برداری از ۳۲ به ۱۶، اطلاعات دقیق‌تری در مورد عوارض ساختمانی کوچکتر بدست می‌آید.

۴) ارزیابی شبکه پیشنهادی بر روی دو مجموعه داده برچسب‌گذاری شده ساختمان و مقایسه با سایر مدل‌های قطعه‌بندی محبوب از نظر عملکرد.

۲- روش پیشنهادی

ساختار کلی رویکرد پیشنهادی در این تحقیق با هدف استخراج ساختمان‌ها از تصاویر با وضوح زیاد در شکل(۱) نشان داده شده‌است. مرحله استخراج ساختمان شامل معماری *backbone* به عنوان استخراج کننده ویژگی، ماژول *ASPP* و رمزگشاست. معماری *ResNet-50* [۳۸] در بخش رمزگذار شبکه استفاده می‌شود که وظیفه استخراج ویژگی‌های عمیق از تصویر ورودی را بر عهده دارد. از آنجایی که تخمین پیکسل متراکم نیازی به لایه‌های کاملاً متصل ندارد، این لایه‌ها حذف می‌شوند و یک بخش رمزگشا برای بازسازی تصویر ورودی و اختصاص یک برچسب به هر پیکسل، به شبکه اضافه می‌شود. شبکه‌های باقیمانده عمیق اثربخشی خود را در پرداختن به موضوع تخریب گرادیان ثابت کرده‌اند و در طول آموزش، توزیع مناسب‌تری از مقادیر گرادیان را تضمین می‌کنند. جزئیات مربوط به معماری پیشنهادی رمزگذار-رمزگشا در بخش‌های ۲-۱ و ۲-۲ مورد بحث قرار خواهد گرفت.



شکل۱: چارچوب مدل پیشنهادی برای قطعه‌بندی معنایی تصاویر با رزولوشن زیاد

[DOI: 10.61186/jgit.11.3.43]

۲-۱- جزئیات ماژول رمزگذار

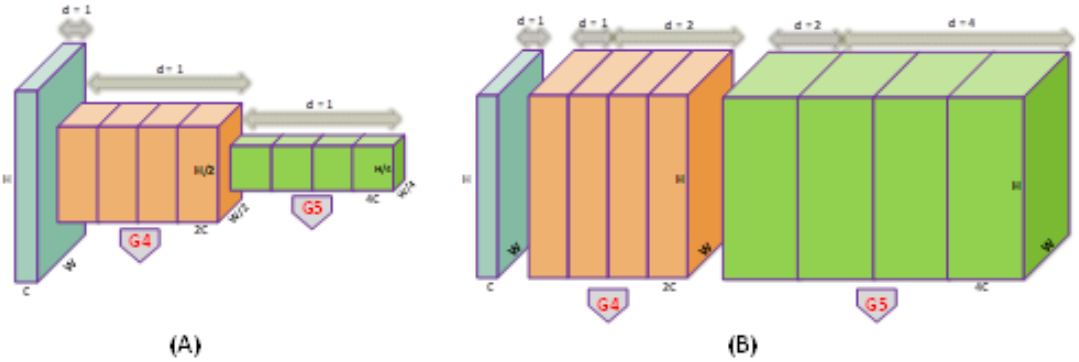
رمزگذار شبکه مسئول کدگذاری ویژگی‌هاست. علیرغم معماری عمیق‌تر، هزینه محاسباتی کمتر و عملکرد قوی، انگیزه اصلی استفاده از معماری *ResNet* در این مطالعه است. مطابق شکل (۲)، رمزگذار از یک بلوک کانولوشن و چهار بلوک باقیمانده تشکیل شده‌است که هر کدام با علامت *Conv(n, m)* نشان داده می‌شوند. به ترتیب، "*n*" به شماره بلوک، و مقدار "*m*" به تعداد لایه‌های کانولوشن باقی‌مانده در هر بلوک اشاره دارد. استفاده از چندین لایه کانولوشن در یک بلوک به شبکه اجازه می‌دهد تا الگوهای پیچیده‌تری را از داده‌های ورودی بیاموزد. نسبت اندازه نقشه ویژگی نهایی تولید شده توسط رمزگذار به اندازه تصویر ورودی، ضریب نمونه برداری کاهشی (*Down Sampling Factor*) یا گام خروجی را نشان می‌دهد[۳۹]. زمانی که از شبکه باقی‌مانده اصلی استفاده شود؛ در صورتیکه ابعاد تصاویر ورودی به شبکه را $L \times H$ در نظر بگیریم؛ ابعاد فضای ویژگی حاصل برای هر بلوک برابر با مقادیر *(L/2*H/2)*، *(L/4*H/4)*،*(L/8*H/8)*،*(L/16*H/16)*،*(L/32*L/32)* خواهد بود. بنابراین، نقشه‌های ویژگی با ضریب نمونه‌برداری ۳۲ تولید می‌شوند.

شکل۲: (الف) خط لوله کلی ساختار اصلی شبکه با ضریب نمونه‌برداری ۳۲، (ب) خط لوله کلی ساختار پیشنهادی با ضریب نمونه‌برداری ۱۶

در مدل پیشنهادی در این مقاله، به‌منظور افزایش وضوح نقشه‌های ویژگی، عملیات نمونه‌برداری کاهشی از بلوک‌های آخر حذف شده که البته منجر به کاهش میدان پذیرایی می‌شود. به همین دلیل، به‌منظور برقراری تعادل بین میدان پذیرنده و وضوح نقشه ویژگی، از کانولوشن گسترده با نرخ‌های متفاوت استفاده می‌کنیم. پیچش‌های آتروس امکان استخراج ویژگی‌ها را به صورت موثر فراهم می‌کنند. اعمال لایه‌های کانولوشن گسترده، میدان دریافت بزرگ‌تری را بدون افزایش تعداد پارامترها فراهم می‌کند و به دنبال آن اطلاعات غنی برای پیش‌بینی جزئیات تصاویر بدست می‌آید. ما در این مطالعه با اعمال کانولوشن گسترده به جای کانولوشن استاندارد، ضریب نمونه برداری یا همان گام خروجی را کاهش دادیم تا در

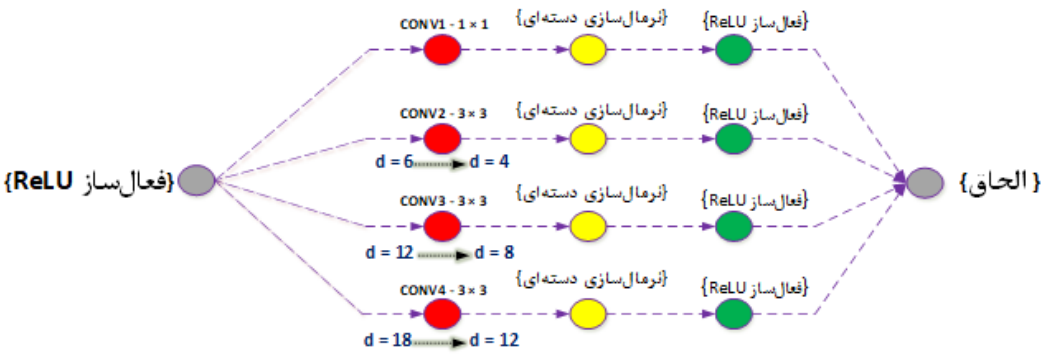
[DOI: 10.61186/jgit.11.3.43]

نهایت به مقدار $OS = ۱۶$ برسد. بدین ترتیب می‌توان اطلاعات دقیق‌تری در مورد عوارض کوچکتر بدست آورد. بنابراین، نقشه‌های ویژگی بزرگ‌تر ایجاد شده در بخش رمزگذار می‌توانند نقش مؤثری را در تولید نقشه‌های قطعه‌بندی ایفا کنند. شکل (۳) نحوه استفاده از کانولوشن گسترده را به جای کانولوشن استاندارد در چارچوب *ResNet* نشان می‌دهد.



شکل ۳: مقایسه *ResNet* اصلی (ستون A) با *ResNet* گسترده (ستون B). در اینجا، H، W و C به ترتیب ارتفاع، عرض و تعداد کانال‌های یک نقشه ویژگی میانی را نشان می‌دهند. $d = 1$ نشان دهنده یک کانولوشن استاندارد است. سایر مقادیر d نشان‌دهنده نرخ‌های مختلف مورد استفاده در کانولوشن است.

در ادامه، از ماژول *ASPP* استفاده می‌شود تا الگوهای چندمقیاسی و اطلاعات زمینه‌ای را بدون اضافه‌کردن بار محاسباتی زیاد، استخراج کند. عملکرد این ماژول به این صورت است که از نقشه ویژگی رمزگذار، اطلاعات موجود از چند مقیاس جمع‌آوری می‌شوند و در نهایت با یکدیگر ترکیب می‌شوند. این ماژول متشکل از چهار فیلتر کانولوشن با نرخ‌های متفاوت و یک لایه *Pooling* است. معماری *DeepLabV3+* اصلی پیشنهاد می‌کند از نرخ‌های آتروس ۶، ۱۲ و ۱۸ در ماژول *ASPP* استفاده شود. با این حال، اندازه‌های کوچک و متغیر عوارض ساختمانی در مقایسه با کل تصویر ممکن است مانع از اثربخشی نرخ‌های آتروس اولیه شود. بنابراین، رویکرد خود را بهره‌وری از نرخ‌های پایین‌تر به‌منظور قطعه‌بندی قوی الگوهای ساختمانی قرار می‌دهیم. هدف نیز بدست آوردن جزئیات مکانی کافی از منطقه مورد نظر است. جزئیات اصلاحات صورت گرفته بر روی ماژول *ASPP* در این مقاله در شکل (۴) نشان داده شده است.



شکل ۴: چارچوب دقیق ماژول اصلاح شده *ASPP*

[DOI: 10.61186/jgit.11.3.43]

۲-۲- جزئیات ماژول رمزگشا

در رمزگشای پیشنهادی، ویژگی‌های ماژول *ASPP* ابتدا با ضریب ۴ به صورت دوخطینمونه‌برداری می‌شوند. سپس، ویژگی‌های نمونه‌برداری شده با ویژگی‌های سطح پایین به هم متصل می‌شوند. ویژگی‌های سطح پایین قبل از الحاق به‌منظور یک فرآیند آموزشی مؤثر، از لایه کانولوشن ۱ × ۱ عبور می‌کنند. ترکیب ویژگی‌های سطح پایین دارای اطلاعات مکانی غنی با ویژگی‌های سطح بالا، دقت قطعه‌بندی را بهبودمی-بخشد. سپس فیلتر کانولوشن ۳ × ۳ اعمال می‌شود و در نهایت نمونه برداری با ضریب ۴ به‌منظور پیش‌بینی قطعه‌بندی نهایی همانطور که در شکل (۱) نشان داده شده است، اعمال می‌شود.

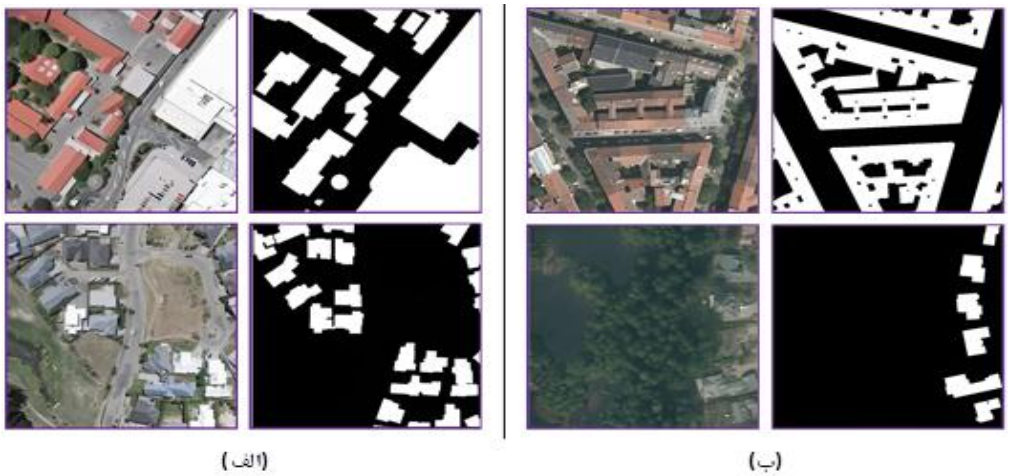
۳- نتایج اجرایی

در این بخش پس از معرفی داده‌های مورد استفاده در این پژوهش، معیارهای ارزیابی، تنظیمات اجرایی و نتایج حاصل از اعمال روش پیشنهادی بر روی هر یک از دو مجموعه داده، ارائه می‌گردد.

۳-۱- داده‌های مورد استفاده

در این تحقیق، کارایی شبکه قطعه‌بندی را با استفاده از دو مجموعه داده ساختمانی متمایز ارزیابی کردیم: مجموعه داده هوایی *WHU* [۴۰] و مجموعه داده *INRIA* [۴۱]. اولین مجموعه داده شامل زمینی به مساحت ۴۵۰ کیلومتر مربع در شهر کرایست چرچ در نیوزیلند است. زیرمجموعه‌ای از مجموعه داده *WHU* برای ارزیابی کارایی شبکه پیشهادی انتخاب شد که در آن ساختمان ها دارای ابعاد و ظواهر متفاوت هستند (شکل (۵-الف)). مشخصات این زیرمجموعه دارای وضوح فضایی ۰٫۳ متر و اندازه هر تصویر ۵۱۲ در ۵۱۲ پیکسل است. مجموعه داده شامل تعداد ۴۷۳۶ موزائیک تصویر برای آموزش، تعداد ۲۴۱۶ تصویر برای تست، و تعداد ۱۰۳۶ تصویر برای اعتبارسنجی است.

مجموعه داده دوم، پنج شهر (آستین، شیکاگو، کیتساپ، تیرول و وین) را پوشش می‌دهد. مجموعه داده ساختمان‌های *INRIA* مجموعه‌ای چالش برانگیز از ۳۶۰ تصویر هوایی است. این تصاویر دارای وضوح مکانی ۰٫۳ متر و اندازه آن ۵۰۰۰ در ۵۰۰۰ پیکسل است. ما تصاویر بزرگ‌تر را به بخش‌های کوچک‌تر تقسیم کردیم تا استفاده از آن‌ها ساده‌تر شود، بدین ترتیب هر تصویر ۵۱۲ * ۵۱۲ پیکسل دارد. در نهایت، موزائیک‌های تصویرکه شامل ساختمان نیستند، برای آموزش کارآمد حذف می شوند. در نتیجه، تعداد ۹۷۳۵ و ۱۹۴۰ موزائیک‌های تصویری به ترتیب برای آموزش و اعتبار سنجی استفاده می شوند. شکل (۵-ب) تصاویر مربوط به این مجموعه داده را نشان می‌دهد.



شکل۵: (الف) مجموعه داده هوایی *WHU*، (ب) مجموعه داده *INRIA*

[DOI: 10.61186/jgit.11.3.43]

۳-۲- معیارهای ارزیابی

جنبه‌های مختلف اثربخشی رویکرد پیشنهادی با استفاده از چهار معیار ارزیابی زیر مورد بررسی قرار می‌گیرد:

- معیار *IOU*: همپوشانی بین مناطق ساختمانی پیش‌بینی‌شده و زمینی را با تقسیم تقاطع دو منطقه بر اتحاد آنها بر طبق رابطه(۱) محاسبه می‌کند.

$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP} + \mathrm{FN}} \quad \text{رابطه(۱)}$$

- معیار دقت : مطابق با رابطه(۲)، به نسبت نمونه‌های مثبت به درستی شناسایی شده در مقایسه با تعداد کل پیش‌بینی‌های مثبت اشاره دارد. دقت بالا نشان‌دهنده نرخ مثبت کاذب پایین است، که نشان می‌دهد مدل در شناسایی دقیق نمونه‌های مثبت برتری دارد.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP}} \quad \text{رابطه(۲)}$$

در رابطه (۲)، مثبت واقعی *(TP)* به مواردی اشاره دارد که مدل به درستی یک ساختمان را شناسایی می‌کند. منفی واقعی *(TN)* مربوط به مواردی است که مدل به درستی یک منطقه غیرساختمانی را شناسایی می‌کند. پیش‌بینی مدل با حقیقت واقعی در هر دوی این موارد مطابقت دارد. مثبت کاذب *(FP)* زمانی رخ می‌دهد که مدل به اشتباه یک منطقه غیرساختمانی را به عنوان یک ساختمان شناسایی کند. منفی کاذب *(FN)* زمانی ایجاد می‌شود که مدل به اشتباه یک منطقه ساختمانی را به عنوان غیرساختمانی طبقه‌بندی کند. در هر دو موقعیت، پیش‌بینی مدل با حقیقت واقعی در تضاد است.

- معیار بازخوانی[۱]: این معیار به شناسایی صحیح موارد مثبت واقعی نسبت به تعداد کل موارد مثبت واقعی در مجموعه آزمایشی اشاره دارد (رابطه(۳)). مقادیر زیاد این معیار نشان‌دهنده نرخ منفی کاذب پایین است.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{FN} + \mathrm{TP}} \quad \text{رابطه(۳)}$$

- معیار امتیاز-اف[۲]: نشان‌دهنده میانگین متعادل بین دقت و بازخوانی است که به عنوان میانگین هارمونیک محاسبه می‌شود و بر طبق رابطه(۴) محاسبه می‌شود.

$$\mathrm{F_{score}} = \frac{2*(\mathrm{Precision}*\mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}} \quad \text{رابطه(۴)}$$

۳-۳- تنظیمات اجرایی

بسته برنامه نویسی پایتورچ[۳] بر روی یک *GPU GeForce RTX 3090* در تمام آزمایشات برای اجرای استراتژی پیشنهادی استفاده شد. مقادیر پیکسل تصاویر قبل از استفاده برای آموزش، از محدوده اصلی [۰، ۲۵۵] تا محدوده [۰، ۱] نرمال می‌شوند. در هر دو مجموعه داده، ابعاد ورودی را ۵۱۲×۵۱۲ در نظر می‌گیریم. معماری‌های تاثیرگذار و کارآمد *U-Net*، *PSPNet* و *DeepLabv3+* برای مقایسه با مدل پیشنهادی و ارزیابی اثربخشی آن انتخاب شده‌اند. تحت شرایط مشابه، هر چهار مدل بر روی دو مجموعه داده فوق الذکر آزمایش می‌شوند. ما از بهینه ساز *Adam* [۴۲] با نرخ یادگیری اولیه ۰٫۰۰۰۱ برای آموزش شبکه استفاده کردیم که پس از هر ۵ اپوک با مقدار ۰٫۸۵ کاهش یافت. فرآیند آموزش شامل ۵۰ اپوک بود. بعلاوه، رویکردهای تقویت داده‌ها مانند چرخش و وارونگی تصادفی تصویر، با هدف افزایش تعمیم‌پذیری مدل و مقابله با خطرات مرتبط با بیش‌برازش انجام می‌شود.

۴- مقایسه و تجزیه و تحلیل نتایج

در ادامه نتایج مقایسه بر روی هر مجموعه داده به تفکیک تشریح شده است.

۴-۱- نتایج مقایسه بر روی مجموعه داده ساختمان هوایی *WHU*

جدول(۱) مقایسه‌های عددی به‌دست‌آمده از معیارهای ارزیابی را نشان می‌دهد. روش پیشنهادی در این مقاله

[۱] *Recall*

[۲] *F Score*

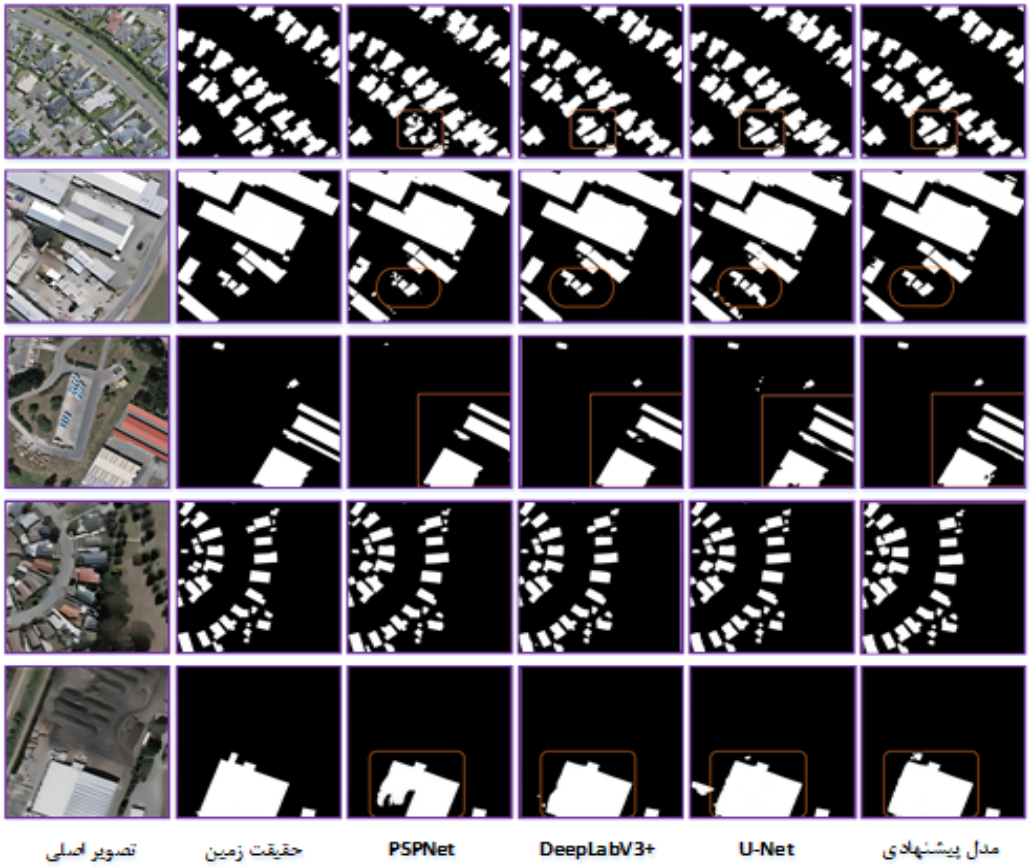
[۳] *PyTorch*

[DOI: 10.61186/jgit.11.3.43]

نسبت به سه روش دیگر در متریک‌های *IoU* و امتیـاز-اف عملکرد بهتـری دارد و بـه ترتیـب مقـادیر ۸۹٫۵۱ و ۹۴٫۶۵ را به دست می‌آورد. تـاثیر اسـتفاده از نـرخ‌هـای آتروس کمتر در مقایسه با مـدل *Deeplabv3+* نیـز در کمی‌سازی معیارهای ارزیابی جـدول مربوطـه مشـخص است. جهت بررسی عملکـرد مـدل پیشـنهادی در ایـن مقاله، نتیجه حاصل را بـا خروجـی مـدل‌هـای انتخـابی مقایسه کردیم. برای اعتبـار بیشـتر، تصـاویری انتخـاب شدند که اشکال مختلـف سـاختمان را شـامل شـوند و پیچیدگی‌های متفاوتی در پس‌زمینه آنها وجـود داشـته باشـد. شـکل(۶) نتـایج بصـری بـه‌دسـت‌آمـده توسـط مدل‌های مختلف بر روی مجموعه داده ساختمان هوایی *WHU* را نشان می‌دهد. نتایج بصری نشان می‌دهـد کـه ساخت یک مدل قطعه بندی قـوی بـا اعمـال تغییـرات مختلف در شبکه قابل دستیابی است.

جدول۱: ارزیابی کمی عملکرد مدل پیشنهادی در مجموعه داده *WHU*

مدل شبکه	دقت	بازخوانی	امتیاز-اف	*IOU*
PSPNet	۹۳٫۵۱	۹۴٫۱۹	۹۳٫۸۵	۸۸٫۴۳
DeepLabV3+	۹۴٫۲۸	۹۳٫۸۳	۹۴٫۰۵	۸۸٫۸۷
U-Net	۹۴٫۹۹	۹۳٫۲۷	۹۴٫۱۲	۸۹٫۱۲
مدل پیشنهادی	۹۴٫۷۷	۹۴٫۵۲	۹۴٫۶۵	۸۹٫۵۱



شکل۶: مقایسه بصری نتایج مدل پیشنهادی و سایر مدل ها بر مجموعه داده *WHU*

[DOI: 10.61186/jgit.11.3.43]

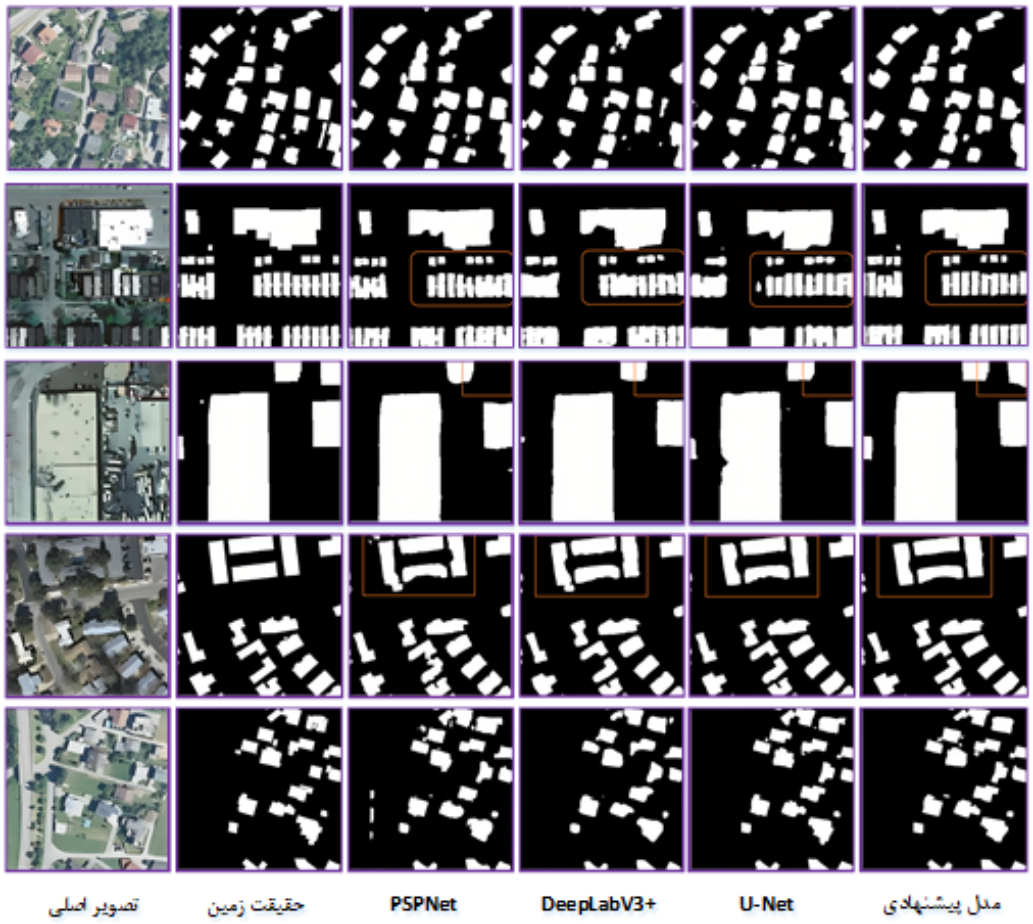
۴-۲- نتایج مقایسه بر روی مجموعه داده *inria*

جدول(۲) مقایسه‌های عددی به‌دست‌آمده از معیارهای ارزیابی را بر روی مجموعه داده *INRIA* نشان می‌دهد. در ارزیابی این مجموعه داده نیز، روش پیشنهادی این مقاله از سه روش دیگر در متریک‌های *IOU* و امتیاز-اف بهتر عمل می‌کند و به ترتیب مقادیر ۷۶٫۶۴ و ۸۷٫۱۴ را به دست می‌آورد.

شکل (۷) نتایج بصری به‌دست‌آمده توسط مدل‌های مختلف بر روی مجموعه داده *INRIA* را نشان می‌دهد.

جدول۲: ارزیابی کمی عملکرد مدل پیشنهادی در مجموعه داده *INRIA*

مدل شبکه	دقت	بازخوانی	امتیاز-اف	*IOU*
PSPNet	۸۴٫۷۵	۸۷٫۶۷	۸۶٫۱۹	۷۵٫۶۱
DeepLabV3+	۸۵٫۱۱	۸۷٫۸۶	۸۶٫۴۶	۷۵٫۸۲
U-Net	۸۴٫۹۵	۸۸٫۷۲	۸۶٫۷۹	۷۶٫۲۹
مدل پیشنهادی	۸۴٫۵۰	۸۹٫۹۴	۸۷٫۱۴	۷۶٫۶۴



شکل۷: مقایسه بصری نتایج مدل پیشنهادی و سایر مدل ها بر مجموعه داده *INRIA*

[DOI: 10.61186/jgit.11.3.43]

۴-۳- تجزیه و تحلیل نتایج

در مدل پیشنهادی در این تحقیق، با استفاده از لایه‌های کانولوشن گسترده به جای کانولوشن استاندارد در لایه‌های آخر *ResNet50* و به دنبال آن تغییر ضریب نمونه‌برداری از ۳۲ به ۱۶، اطلاعات دقیق‌تری در مورد عوارض ساختمانی کوچکتر بدست آمد. همان‌طور که در اشکال (۶) و (۷) مشخص است، تصاویر سطر اول و دوم هر دو مجموعه داده متشکل از عوارض کوچک ساختمانی هستند و مدل پیشنهادی ما در مقایسه با شبکه‌های ذکرشده، دقیق‌تر این موارد را شناسایی کرده‌است. به‌طور کلی، نتیجه قطعه‌بندی روش پیشنهادی ما پیشرفت نسبتاً خوبی در تشخیص ساختمان‌های بزرگ و کوچک و همچنین اطلاعات لبه دارد.

از آنجایی که مجموعه داده‌های *WHU* و *INRIA* محیطی بدون ساختار و شامل اشیایی با اندازه‌های مختلف و چگالی بالا هستند؛ از نرخ‌های آتروس پایین‌تر برای مناطق متراکم استفاده کردیم. اعمال نرخ‌های آتروس پایین‌تر در ماژول *ASPP* و تغییر آنها به ۴، ۸ و ۱۲به‌طور قابل‌توجهی عملکرد بخش‌بندی را در هردو مجموعه داده بهبود بخشید. به‌طور کلی در ساختمان‌های بزرگ، شبکه با ماژول ویژگی‌های چندمقیاسی، عملکرد بهتری دارد چرا که تجمیع ویژگی‌های چند مقیاسی منجر به دریافت اطلاعات زمینه‌ای وسیع‌تر می‌شود.

۵- بحث و نتیجه گیری

در استخراج ساختمان‌ها، قطعه‌بندی معنایی قوی، درک بهتر مورفولوژی ساختمان‌ها و شناسایی صحیح مرزها را امکان‌پذیر می‌کند. قطعه‌بندی ساختمان‌ها به دلیل نیاز به ویژگی‌های معنایی غنی کار دشواری است. تفاوت در شکل، رنگ و اندازه ساختمان‌ها و نزدیکی آن‌ها به سایر عوارض مانند پارکینگ‌ها وخیابان‌ها تشخیص آنها را در تصاویر با وضوح زیاد دشوار می‌کند. با توجه به عدم تعادل در توزیع بین کلاس‌ها، تمرکز بر انتخاب مناسب‌ترین معیارها برای ارزیابی بهبود عملکرد مدل پیشنهادی بسیار مهم است. بنابراین، عملکرد و دقت بالا می‌تواند گمراه کننده باشد. در این مقاله، معیارهای امتیاز-اف و *IOU* در ارزیابی عملکرد شبکه قطعه‌بندی مورداستفاده قرار گرفتند. یافته‌های بصری نیز نشان می‌دهد که این دو معیار، شاخص‌های قابل اعتمادتری هستند.

بسیاری از معماری‌های قطعه‌بندی برای انجام اصلاحات مختلف با هدف بهبود عملکرد بخش‌بندی و امکان‌سنجی سخت‌افزار مناسب هستند. برای پیاده‌سازی سخت افزاری، می‌توان یک مدل سبک وزن با معماری مناسب در رمزگذار ساخت. همچنین، ساختارهای مورد استفاده در رمزگذار مستقیماً بر دقت قطعه‌بندی تأثیر می‌گذارند. تغییرات اضافی در پارامترهایی مانند اندازه ادغام، تعداد فیلتر کانولوشن، نرخ آتروس و ضریب نمونه برداری در ماژول‌های دیگر به جز رمزگذار می‌تواند عملکرد قطعه‌بندی را بهبود بخشد. مدل پیشنهادی در این تحقیق با اصلاح لایه‌های آخر معماری *ResNet-50* و اعمال پیچش گسترده به جای پیچش استاندارد در این لایه‌ها که تغییر ضریب نمونه‌برداری یا همان گام خروجی از ۳۲ به ۱۶ را به همراه دارد، توانست اطلاعات غنی برای پیش‌بینی جزئیات تصاویر بدست آورد. علاوه براین، با اصلاح و کاهش نرخ‌های آتروس ماژول *ASPP*، اطلاعات مکانی بیشتری از چندمقیاس جمع‌آوری می‌شود. مدل *DeepLabV3+* اصلی با نرخ‌های آتروس ۶، ۱۲، ۱۸ با *ResNet-50*، مقدار*IOU* برابر با ۸۸/۸۷ را در مجموعه داده *WHU* و مقدار*IOU* برابر با ۷۵/۸۲ را در مجموعه داده *INRIA* برای قطعه‌بندی ساختمان‌ها به دست آورد. در حالیکه، مدل پیشنهادی دراین تحقیق براساس کاهش نرخ‌های آتروس به ۴، ۸ و ۱۲ و تغییر در لایه‌های *ResNet-50* در عملیات استخراج عوارض ساختمانی مؤثرتر عمل کرد زیرا به*IOU* برابر با ۸۹/۵۱ در مجموعه داده *WHU* و ۷۶/۶۴ در مجموعه داده *INRIA* دست یافت. علاوه بر این، یک مقایسه جامع با سایر روش‌های قطعه‌بندی پیشرفته نیز برای مشاهده

[DOI: 10.61186/jgit.11.3.43]

اثربخشی مدل پیشنهادی انجام شد که برتری این مدل را نشان داد. به‌طور کلی چارچوب‌های مبتنی بر یادگیری عمیق نتایج قطعه‌بندی قوی را ارائه می‌دهند اما تعداد زیاد پارامترها در اکثر این مدل‌ها حائز اهمیت است. محدودیت اصلی در مدل پیشنهادی در این تحقیق نیز تعداد زیاد پارامترها در *DeepLabV3+* اصلاح‌شده است که لازم است کاهش یابد.

تقدیر و تشکر

این پژوهش با حمایت مالی دانشگاه تربیت دبیر شهید رجائی طبق ابلاغ گرنت شماره ۴۹۴۳ مورخ۱۴۰۲/۰۳/۰۶ انجام گردیده است.

مراجع

[1] Lin, J., Jing, W., Song, H., Chen, G. "ESFNet: Efficient Network for Building Extraction From High-Resolution Aerial Images," in IEEE Access, vol. 7, pp. 54285-54294, 2019, doi: 10.1109/ACCESS.2019.2912822.

[2] Musse, M.A., Barona, D.A., Rodriguez, L.M.S. "Urban environmental quality assessment using remote sensing and census data," International Journal of Applied Earth Observation and Geoinformation, vol. 71, PP. 95-108, 2018, https://doi.org/10.1016/j.jag.2018.05.010.

[3] Agarwal, L., Rajan, K. S. 2015. Fast ICA based algorithm for building detection from VHR imagery. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). 10.1109/IGARSS.2015.7326162.

[4] Feng, W., Sui, H., Hua, L.,Xu, C., Ma, G., Huang, W. 2020. Building extraction from VHR remote sensing imagery by combining an improved deep convolutional encoder-decoder architecture and historical land use vector map. International Journal of Remote Sensing, 41(17), 6595-6617. https://doi.org/10.1080/01431161.2020.1742944.

[5] Huang, J., Xia, G.S., Hu, F., Zhang, L. 2018. Accurate building detection in VHR remote sensing images using geometric saliency. IGRASS'18 conference paper. https://doi.org/10.48550/arXiv.1806.00908.

[6] Wang, X., Li, P. Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data, ISPRS Journal of Photogrammetry and Remote Sensing, vol.159, 2020, pp. 322-336, https://doi.org/10.1016/j.isprsjprs.2019.11.028.

[7] You, Y., Wang, S., Ma, Y., Chen, G., Wang, B., Shen, M., Liu, W. 2018. Building Detection from VHR Remote Sensing Imagery Based on the Morphological Building Index. Journal of remote sensing, 10, 1287, https://doi:10.3390/rs10081287.

[8] Li, J., Huang, X., Tu, L., Zhang, T. and Wang, L. 2022. A review of building detection from very high resolution optical remote sensing images. Journal of GISCIENCE & REMOTE SENSING, 59(1), 1199–1225. https://doi.org/10.1080/15481603.2022.2101727.

[9] Wang, S., Hou, X., Zhao, X. "Automatic Building Extraction From High-Resolution Aerial Imagery via Fully Convolutional Encoder-Decoder Network With Non-Local Block," in IEEE Access, vol. 8, pp. 7313-7322, 2020, doi: 10.1109/ACCESS.2020.2964043.

[10] Liu, Y., Zhou, J., QI, W., Li, X. et al. "ARC-Net: An Efficient Network for Building Extraction From High-Resolution Aerial Images," in IEEE Access, vol. 8, pp. 154997-155010, 2020, doi: 10.1109/ACCESS.2020.3015701.

[11] Bittner, K., Adam, F., Cui, S., Korner, M., Reinartz, P. 2018. Building Footprint Extraction From VHR Remote Sensing Images Combined With Normalized DSMs Using Fused Fully Convolutional Networks. IEEE Journal of Selected Topics

[DOI: 10.61186/jgit.11.3.43]

in Applied Earth Observation and Remote Sensing, 11(8), 2615-2629, https://doi.org/10.1109/JSTARS.2018.2849363.

[12] Zeng, Y., Guo, Y., Li, J. 2022. Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning. Journal of Neural Computing and Applications, 34, 2691-2706. https://doi.org/10.1007/s00521-021-06027-1.

[13] Zhu, Q., Liao, C., Hu, H., Mei, X., Li, H. "MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction From Remote Sensed Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 7, pp. 6169-6181, July 2021, doi: 10.1109/TGRS.2020.3026051.

[14] Liao, C.m Hu, H., Li, H., Ge, X., Chen, M., Li, C., Zhu, Q. "Joint Learning of Contour and Structure for Boundary-Preserved Building Extraction," Remote Sens. 2021, 13, 1049. https://doi.org/10.3390/rs13061049.

[15] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

[16] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A. " Deep learning in remote sensing applications: A meta-analysis and review," ISPRS Journal of Photogrammetry and Remote Sensing 152, pp. 166-177, 2019. DOI:10.1016/j.isprsjprs.2019.04.015.

[17] Yoo, H.J. "Deep Convolution Neural Networks in Computer Vision: a Review," IEIE Transactions on Smart Processing and Computing, vol. 4, no. 1, 2015, pp. 35-43. http://dx.doi.org/10.5573/IEIESPC.2015.4.1.035.

[18] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J. "A Review on Deep Learning Techniques Applied to Semantic Segmentation ," Computer visión and Pattern recognition, 2017. https://doi.org/10.48550/arXiv.1704.06857.

[19] Ronneberger, O., Fischer, P., Brox, T. " U-Net: Convolutional Networks for Biomedical Image Segmentation," Springer International Publishing Switzerland 2015 N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 234–241, 2015. DOI: 10.1007/978-3-319-24574-4_2.

[20] Hamaguchi, R., Fujita, A., Nemoto, K., Imaizumi, T., Hikosaka, S. " Effective Use of Dilated Convolutions for Segmenting Small Object Instances in Remote Sensing Imagery," Computer visión and Pattern recognition, 2017. HTTPS://DOI.ORG/10.48550/ARXIV.1709.00179.

[21] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Transactions on Pattern Analysis and Machine Intelligence PP(99), 2016. DOI:10.1109/TPAMI.2017.2699184.

[22] Shao, Z., Tang, P., Wang, Z., Saleem, N., Yam, S., Sommai, C. "BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images," Remote Sensing 12(6):1050, 2020. DOI:10.3390/rs12061050.

[23] Shelhamer, E., Long, J., Darrell, T. "Fully Convolutional Networks for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, https://doi.org/10.48550/arXiv.1411.4038.

[24] Liu, Y., Minh Nguyen, D., Deligiannis, N., Ding, W., Munteanu, A. "Hourglass-shape network based semantic segmentation for high resolution aerial imagery," Remote Sens. 2017, 9, 522.

[25] Lin, G., Milan, A., Shen, C., Reid, I. "RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic

[DOI: 10.61186/jgit.11.3.43]

Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 5168-5177. DOI 10.1109/CVPR.2017.549.

[26] Pan, X., Yang, F., Gao, L., Chen, Z., Zhang, B., Fan, H., Ren, J. " Building Extraction from High-Resolution Aerial Imagery Using a Generative Adversarial Network with Spatial and Channel Attention Mechanisms," Remote Sens. 2019, 11, 917; doi:10.3390/rs11080917.

[27] Xu, Y., Wu, L., Xie, Z., Chen, Z. "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters," Remote Sens. 2018, 10, 144; doi:10.3390/rs10010144.

[28] Wang, S., Hou, X., Zhao, X. "Automatic Building Extraction From High-Resolution Aerial Imagery via Fully Convolutional Encoder-Decoder Network With Non-Local Block," in IEEE Access, vol. 8, pp. 7313-7322, 2020, doi: 10.1109/ACCESS.2020.2964043.

[29] Wei, S., Ji, S., Lu, M. "Toward Automatic Building Footprint Delineation From Aerial Images Using CNN and Regularization," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 3, pp. 2178-2189, March 2020, doi: 10.1109/TGRS.2019.2954461.

[30] Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., and Zhao, Z. " Semantic Segmentation of Urban Buildings from VHR Remote Sensing Imagery Using a Deep Convolutional Neural Network," Remote Sens. 2019, 11, 1774; doi:10.3390/rs11151774.

[31] Yang, H. L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A., Bhaduri, B. "Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 8, pp. 2600-2614, Aug. 2018, doi: 10.1109/JSTARS.2018.2835377.

[32] GUO, Z., SHI, X., ZHANG, H., HUANG, D., SONG, X., YAN, J., SHIBASAKI, R. "Enhancing Building Semantic Segmentation Accuracy with Super Resolution and Deep Learning: Investigating the Impact of Spatial Resolution on Various Datasets," Computer Vision and Pattern Recognition, 2023. HTTPS://DOI.ORG/10.48550/ARXIV.2307.04101

[33] Chen, L.C., Papandreou, G., Schroff, F., Adam, H. "Rethinking Atrous Convolution for Semantic Image Segmentation," Computer Vision and Pattern Recognition, 2017. https://doi.org/10.48550/arXiv.1706.05587.

[34] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. "Pyramid Scene Parsing Network," Computer Vision and Pattern Recognition, 2016. https://doi.org/10.48550/arXiv.1612.01105.

[35] Ji,Sh., Wei, Sh., Lu, M. "A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery," International Journal of Remote Sensing, 2018, DOI: 10.1080/01431161.2018.1528024.

[36] Zeyu, X., Zhanfeng, Sh., Yang, L., Lifang, Z., Yingming, K., Lingling, L., Qi, W. "Classification of high-resolution remote sensing images based on Enhanced DeepLab algorithm and adaptive loss function," Journal of Remote Sensing, 2021, DOI: 10.11834/jrs.20209200.

[37] Wang, X., Hu, Z., Shi, S. et al. "A deep learning method for optimizing semantic segmentation accuracy of remote sensing images based on improved UNet," Sci Rep 13, 7600 (2023). https://doi.org/10.1038/s41598-023-34379-2.

[38] He, K., Zhang, X. Ren, S., Sun, J. "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[39] Chen, LC., Zhu, Y., Papandreou, G.,

[DOI: 10.61186/jgit.11.3.43]

Schroff, F., Adam, H. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_49.

[40] Ji, S., Wei, S., Lu, M. "Fully Convolutional Networks for Multisource Building Extraction From an Open Aerial and Satellite Imagery Data Set," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 1, pp. 574-586, Jan. 2019, doi: 10.1109/TGRS.2018.2858817.

[41] Maggiori, E., Tarabalka,Y.,Charpiat, G., Alliez, P. "Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark," IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), Jul 2017, Fort Worth, United States. ffhal-01468452.

[42] Kingma, D., Ba, J. "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations. 2014.

[DOI: 10.61186/jgit.11.3.43]

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.11, No.3, Autumn 2023

Research Paper

Evaluating the Capabilities of DEEPLABV3+ Encoder-Decoder Network with Modified Atrous Convolutions (Case Study: Deep Semantic Building Segmentation)

Mohammad Erfan Omati [1], Fatemeh Tabib Mahmoudi [2]*

1- Ms.c student of remote sensing in Department of Geomatics, Faculty of Civil, Shahid Rajaee Teacher Training University
2- Assisstant professor in Department of Geomatics, Faculty of Civil, Shahid Rajaee Teacher Training University

Abstract

Building segmentation is a difficult task due to the need for rich semantic features. Differences in the shape, color and size of buildings and their proximity to the other features such as parking lots and streets make it challenging to recognize them in high resolution images. In this research, with the aim of extracting buildings from high-resolution images, the deep convolutional neural network architecture of the encoder-decoder type which is based on the modified DeepLabV3+ model, has been used. In the Atrous module of this modified model, in order to achieve the goal of performing a more powerful semantic segmentation of small and large building objects, the convolution layers are applied with lower rates compared to the original module. The performance of the proposed model in this research was evaluated using two data sets, WHU and INRIA, and the results showed that using lower Atrous rates and changing them to 4, 8, and 12 improved the segmentation performance in both data sets significantly. Compared to the other advanced models in the WHU data set, the proposed modified model was able to improve the IOU and F-Score indices by 0.39 and 0.53 respectively. In addition, the modified method in the INRIA dataset improved both of the above indices by 0.35. The proposed model in this research, that was based on the reduction of Atros rates to 4, 8 and 12 and the change in ResNet-50 layers, was able to achieve an IOU equal to 89.51 in the WHU dataset and 76.64 in the INRIA one in the extraction of construction charges whereas the original DeepLabV3+ model with the Atros rates of 6, 12 and 18 and the original ResNet-50 achieved an IOU of 88.87 in the WHU dataset and 75.82 in the INRIA one for building segmentation.

Key words: *Semantic Segmentation, Deep Convolutional Neural Network, Encoder-Decoder, Atrous Convolution.*

Correspondence Address: *Remote Sensing Group, Department of Geomatics, Faculty of Civil, Shahid Rajaee Teacher Training University, Tehran, Iran.*
Tel : *+98 21 22970021.*
Email*: fmahmoudi@sru.ac.ir*

[DOI: 10.61186/jgit.11.3.43]