نشریه علمی مهندسی فناوری اطلاعات مکانی

سال یازدهم، شماره نخست، بهار ۱۴۰۲

Vol.11, No.1, Spring 2023

۱۰۵-۱۲۳

مقاله پژوهشی

DOR: *20.1001.1.20089635.1402.11.1.6.7*

ارزیابی عملکرد سه مدل یادگیری عمیق در استخراج عوارض ساختمانی از تصاویر هوایی و ماهواره‌ای

نیما احمدیان[1*]، امین صداقت[2] ، نازیلا محمدی[3]

۱- دانشجوی کارشناسی ارشد سنجش‌ازدور، گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه تبریز

۲- دانشیار، گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه تبریز

۳- استادیار، گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه تبریز

تاریخ دریافت مقاله: ۱۴۰۲/۰۲/۱۱ تاریخ پذیرش مقاله: ۱۴۰۲/۰۳/۲۴

چکیده

ساختمان‌ها به‌عنوان یکی از مهم‌ترین عوارض دست‌ساز بشر، کاربردهای فراوانی در زمینه‌های مختلف داشته و ارزیابی و شناسایی آن‌ها با استفاده از تصاویر هوایی و ماهواره‌ای امری ضروری است. روش‌های مبتنی بر یادگیری عمیق، اخیراً به طور گسترده‌ای برای استخراج عوارض ساختمانی از تصاویر هوایی و ماهواره‌ای به‌صورت خودکار استفاده شده‌اند. شناخت خصوصیات روش‌های مختلف در مقایسه با یکدیگر و برای انواع مختلف از تصاویر با شرایط هندسی و روشنایی متفاوت ضروری است. بدین منظور، در این تحقیق عملکرد سه مدل یادگیری عمیق مطرح شامل *Mask-RCNN (Mask Region-based Convolutional Neural Network)*، *U-Net* و *MA-FCN* (*Multi-scale Aggregation Fully Convolutional Network*) در استخراج عوارض ساختمانی از سه مجموعه داده تصاویر ماهواره‌ای و هوایی با استفاده از معیارهای *IOU* (*Intersection Over Union*) و *F1-score* بررسی شده است. علاوه بر این در این تحقیق اثر استفاده از مدل رقومی سطح در فرآیند استخراج ساختمان توسط این الگوریتم‌ها نیز بررسی شده است. به طور کلی نتایج حاصل از این تحقیق نشان می‌دهد که علاوه بر نوع مدل، تعداد و کیفیت نمونه‌های آموزشی و استفاده از مدل رقومی سطح نیز در نتایج تأثیرگذار است. همچنین استفاده از مدل رقومی سطح در کنار تصاویر سه‌باندی روش مناسبی برای بهبود عملکرد مدل‌های یادگیری عمیق در استخراج عوارض ساختمانی است. مدل رقومی سطح نتایج حاصل از استخراج ساختمان‌ها را در مدل‌های *U-Net* و *MA-FCN* به‌ترتیب ٪ ۷٫۴۶ و ٪ ۵٫۷ در تصاویر ماهواره‌ای و ٪ ۳٫۶۱ و ٪ ۳٫۳۴ در تصاویر هوایی در معیار *IOU* بهبود داده است. مدل‌های *U-Net* و *MA-FCN* به‌دلیل ترکیب ویژگی‌های قسمت رمزگذار با ویژگی‌های قسمت رمزگشا، در مرز ساختمان‌ها دقیق‌تر هستند. مدل *Mask-RCNN* به‌دلیل دارا بودن ساختار *ResNet* در معماری خود به مسئله فرابرازش مقاوم‌تر است.

کلید واژه‌ها : یادگیری عمیق، ساختمان، مدل رقومی سطح، تصاویر ماهواره‌ای، *U-Net*.

[*] نویسنده مکاتبه‌کننده: گروه مهندسی نقشه‌برداری، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران

تلفن: ۰۹۱۴۶۵۱۵۳۹۱

Email: n.ahmadian99@ms.tabrizu.ac.ir

۱- مقدمه

استخراج عوارض ساختمانی از تصاویر هوایی و ماهواره‌ای اهمیت بالایی داشته و دارای کاربردهای فراوان در برنامه‌ریزی شهری، تشخیص تغییرات و به‌روزرسانی نقشه‌ها است [۱]. امروزه باوجود سنجنده‌های پیشرفته که می‌توانند تصاویر هوایی و ماهواره‌ای با قدرت تفکیک مکانی بالا را در اختیار ما قرار دهند، می‌توان با استفاده از این تصاویر عملیات استخراج ساختمان‌ها را با دقت بالایی انجام داد [۲]. همچنین با ظهور فناوری‌های جدید سنجش‌ازدور و علوم داده می‌توان عملیات استخراج ساختمان‌ها را به‌صورت خودکار انجام داد ولی به دلیل وجود ساختارهای پیچیده ساختمان‌ها و پس‌زمینه، این کار چالش بزرگی به‌شمار می‌رود [۳].

استخراج ساختمان‌ها به‌صورت دستی با توجه به ابعاد بزرگ تصاویر هوایی و ماهواره‌ای، امری زمان‌بر و پیچیده است. به همین دلیل توسعه روش‌هایی خودکار جهت استخراج ساختمان‌ها از این تصاویر ضروری است. تاکنون روش‌های خودکار و نیمه‌خودکار زیادی برای استخراج ساختمان‌ها ارائه شده است ولی به دلیل اینکه ساختمان‌ها می‌توانند اختلافات زیادی در ویژگی‌هایی همچون ظاهر، هندسه و طیف داشته باشند، فرآیند استخراج خودکار ساختمان‌ها از تصاویر هوایی و ماهواره‌ای مسئله‌ای چالش‌برانگیز است [۴].

روش‌های مرسوم استخراج ساختمان‌ها اغلب از ویژگی‌هایی همچون رنگ، طیف، شکل، بافت و سایه استفاده می‌کردند ولی با توجه به اینکه این ویژگی‌ها وابسته به شرایط نوری و اتمسفری، کیفیت سنجنده، مقیاس و نوع ساختمان‌ها هستند، نمی‌توان از این ویژگی‌ها برای استخراج خودکار ساختمان‌ها استفاده کرد [۳]. همچنین بعضی از روش‌های مرسوم از تکنیک حدآستانه‌گذاری برای استخراج ساختمان‌ها استفاده می‌کنند که نتایج آن‌ها به میزان زیادی وابسته به تنظیم پارامتر حد آستانه است [۵ و ۶].

هورتاس و همکاران (۱۹۸۸) از اشکال مستطیلی شکل و سایه‌ها برای استخراج ساختمان‌ها از تصاویر هوایی استفاده کرده‌اند [۷]. بدیهی است که در صورت عدم وجود سایه و وجود ساختمان‌هایی با اشکال پیچیده، این مدل کارایی نخواهد داشت. در تحقیقی دیگر از فیلتر *Top-hat* به همراه الگوریتم *k-means* برای استخراج ساختمان‌ها استفاده شده است که فیلتری بر اساس ریخت‌شناسی است [۸]. در این روش، ابتدا تصویر با استفاده از دو تبدیل *Top-hat*، به مناطق روشن و تاریک تقسیم شده و سپس عملیات بهبود کنتراست روی تصویر اعمال می‌شود. سپس با استفاده از الگوریتم *k-means*، مقادیر میانگین سه کلاس روشن، تاریک و میانی تعیین می‌شود. از میانگین کلاس میانی به‌عنوان حدآستانه برای تقسیم تصویر به دو منطقه روشن و تاریک استفاده می‌شود. درنهایت، برای هر دو منطقه فیلتر میانه با ابعاد ۳×۳ اعمال شده و ساختمان‌ها استخراج می‌شوند. در این روش، نتایج نهایی وابستگی زیادی به شرایط رادیومتریکی تصویر و حدآستانه موردنظر بستگی دارد. در تحقیقی دیگر شاخص ریخت‌شناسی ساختمان[1] که از ویژگی‌هایی همچون روشنایی، کنتراست و اندازه برای استخراج خودکار ساختمان‌ها استفاده می‌کند، توسعه داده شده است [۹]. به‌دلیل وجود شباهت‌هایی در ویژگی‌های مربوط به ساختمان‌ها با عوارضی دیگر همچون راه‌ها، این روش چندان قابل اعتماد نیست. در سالیان اخیر تحقیقات زیادی درباره استخراج ساختمان‌ها از تصاویر هوایی و ماهواره‌ای به‌صورت خودکار انجام گرفته است. با توجه به این تحقیقات می‌توان از تکنیک‌های یادگیری عمیق برای استخراج خودکار ساختمان‌ها استفاده کرد [۱۰]. مدل‌های یادگیری عمیق محدودیت‌های روش‌های مرسوم را ندارند و نتایج خوبی در فرآیند استخراج خودکار ساختمان‌ها از تصاویر هوایی و ماهواره‌ای داشته‌اند [۱۱].

در سالیان اخیر یادگیری عمیق به‌خصوص شبکه‌های

[1] *Morphological Building Index (MBI)*

عصبی کانولوشنی[1] به دلیل پیشرفت‌های قابل‌توجه به طور گسترده‌ای در کاربردهای مختلف در سنجش از دور استفاده شده و نتایج خوبی داشته‌اند. مزیت این روش‌ها این است که این شبکه‌ها می‌توانند اطلاعات عمیق و مفیدی را در سطوح مختلف از داده‌های ورودی یاد گرفته و برای حل مسائل مختلف به کار گیرند [۱۲]. در برخی از تحقیقات هم از شبکه‌های کانولوشنی عمیق برای استخراج ساختمان‌ها به‌صورت خودکار استفاده شده است [۱۳،۱۴و۱۵].

علاوه بر روش‌های ذکرشده، می‌توان از شبکه‌های کانولوشنی کامل که دارای چارچوب رمزگذار-رمزگشا و لایه‌های کانولوشنی، نمونه‌برداری کاهشی و نمونه برداری افزایشی هستند، استفاده کرد؛ زیرا این شبکه‌ها نیز می‌توانند نتایج خوبی در سطح پیکسل ارائه دهند [۱۶و۱۷]. در این شبکه ها یک خروجی باینری برای تصاویر ایجاد می‌شود که هم‌اندازه با تصاویر ورودی است، به‌طوری‌که پیکسل‌های سفید نشان‌دهنده ساختمان‌ها و پیکسل‌های سیاه نشان‌دهنده پس‌زمینه است [۱۸].

در تحقیقات پیشین، از شبکه های عصبی کانولوشنی برای پردازش تصاویر جهت آموزش و پیش بینی کلاس هر پیکسل استفاده شده است ولی معایب این شبکه ها در این است که خروجی آن ها توان تفکیک مکانی کمتری نسبت به تصاویر ورودی دارند [۱۹]. این شبکه ها قدرت تفکیک مکانی تصاویر ورودی را کاهش و بعد مربوط به ویژگی‌هایی که استخراج می‌کنند را افزایش می‌دهند [۲۰].

با توجه به اینکه *FCN*ها از *CNN*ها تشکیل شده‌اند، قدرت تفکیک مکانی خروجی آن ها از تصاویر ورودی پایین تر خواهد بود و استفاده از نمونه برداری افزایشی این اطلاعات را بازنمی گرداند [۲۰]. در *FCN*ها ابتدا نقشه‌های ویژگی‌ای با لایه‌های کانولوشنی و پولینگ و نمونه برداری کاهشی در قسمت رمزگذار تولید می شوند که اطلاعات زیاد ولی توان تفکیک مکانی پایینی دارند و برای افزایش توان تفکیک مکانی خروجی نهایی، ابعاد نقشه های ویژگی را با استفاده از لایه های کانولوشنی معکوس و نمونه‌برداری افزایشی افزایش داده و با نقشه‌های ویژگی هم‌اندازه موجود در قسمت رمزگذار شبکه ترکیب می کنند [۱۸]. به‌عنوان مثال، دو شبکه *SegNet* و *U-Net* از معروف ترین شبکه هایی هستند که از معماری رمزگذار-رمزگشا استفاده کرده و فرآیند ترکیب اطلاعات را در چند سطح انجام می‌دهند [۲۱و۲۲].

مدل‌های *SegNet* و *ResNet* از معروف‌ترین مدل‌هایی هستند که از شبکه های کانولوشنی کامل چند مقیاسه تشکیل شده‌اند و از تلفیق داده هایی مثل ابر نقطه لیدار یا داده چند-طیفی استفاده می کنند [۲۳]. به‌عنوان مثال از شاخص *NDVI* به همراه مدل رقومی سطح نرمال‌شده و مؤلفه اول *PCA* در کنار باندهای اصلی برای استخراج ساختمان‌ها و پوشش گیاهی با استفاده از *CNN* استفاده شده است [۲۴].

به دو روش می‌توان عملیات تلفیق را انجام داد: ترکیب در سطح داده و ترکیب در سطح تصمیم‌گیری [۲۵]. در تلفیق در سطح داده، اطلاعات طیفی مثل باندهای قرمز، سبز، آبی و مادون قرمز نزدیک با داده هایی ساختاری همچون مدل رقومی سطح ترکیب شده و به‌عنوان داده ورودی وارد شبکه موردنظر می‌شوند؛ درحالی‌که در تلفیق در سطح تصمیم‌گیری، داده‌های طیفی و ساختاری به شبکه هایی مختلف وارد شده و خروجی‌های آن‌ها با یکدیگر ترکیب می‌شوند [۲۵].

با توجه به گسترش روش های مبتنی بر یادگیری عمیق و افزایش کاربرد آنها در استخراج عوارض ساختمانی در سنجش از دور شناخت خصوصیات الگوریتم‌های موجود و بررسی عملکرد آنها در مقایسه با یکدیگر ضروری است. لین و همکاران عملکرد مدل های مختلف مبتنی بر یادگیری عمیق نظیر *GoogLeNet* [۲۶] و *VGG Networks* [۲۷] در استخراج عوارض ساختمانی از تصاویر سنجش‌ازدور را بررسی کرده و ویژگی‌های هر

[1] *Convolutional Neural Networks (CNNs)*

یک از مدل‌ها را بیان کرده‌اند [۲۸]. در تحقیق آنها، سه مجموعه داده مناسب برای آموزش مدل‌های یادگیری عمیق جهت استخراج عوارض ساختمانی معرفی شده است. با این وجود بعضی از روش‌های مطرح اخیر در تحقیق آنها بررسی نشده و علاوه بر این تأثیر استفاده از مدل رقومی ارتفاعی در نتایج نیز ارزیابی نشده است.

هدف از این تحقیق بررسی و مقایسه عملکرد سه مدل یادگیری عمیق در استخراج عوارض ساختمانی از تصاویر هوایی و ماهواره‌ای است که برای این منظور در این تحقیق از سه مدل یادگیری عمیق *MA-FCN*[1] [۱۰]، *Mask-RCNN*[2] [۲۹] و *U-Net* [۲۲] استفاده می‌شود. این سه مدل به‌دلیل دارا بودن ساختارهای متفاوت در معماری خود برای این تحقیق انتخاب شده‌اند. بر این اساس نوآوری‌های اصلی این تحقیق به ترتیب زیر قابل بیان هستند:

- معرفی، بررسی و مقایسه عملکرد سه مدل یادگیری عمیق مطرح شامل *Mask-RCNN*، *U-Net* و *MA-FCN* به منظور شناسایی مزایا و معایب هر یک از این مدل‌ها در شرایط مختلف.
- معرفی مجموعه داده‌های مختلف در استخراج ساختمان و استفاده از دو مجموعه داده که دارای نمونه‌های آموزشی با تعداد و توان تفکیک مکانی متفاوت هستند، برای ارزیابی روش‌های مورد مقایسه.
- بررسی اثر استفاده از مدل رقومی سطح در فرآیند آموزش و استخراج مدل‌های *MA-FCN* و *U-Net* به منظور بهبود عملکرد آنها. برای این منظور مدل‌های *MA-FCN* و *U-Net* یک‌بار با تصاویر سه باندی *R-G-B* و یک‌بار با تصاویر چهاربانده *R-G-B-DSM* آموزش و ارزیابی می‌شود.
- بحث و بررسی نتایج کمی مدل‌های مورد مقایسه در استخراج ساختمان و بیان نقاط ضعف و قوت هر یک از این مدل‌ها نسبت به شرایط مختلف در مقایسه با یکدیگر

در ادامه در بخش دوم روش تحقیق بیان شده و سپس فرآیند پیاده‌سازی و ارزیابی نتایج ارائه می‌شود. بعد از آن در بخش چهارم بحث و بررسی نتایج بیان شده و در نهایت در بخش پنجم نتیجه‌گیری و پیشنهادات بیان می‌شود.

۲- روش تحقیق

این تحقیق در دو مرحله معرفی و آموزش مدل‌های یادگیری عمیق موردنظر پیاده‌سازی می‌شود. در مرحله اول، ساختار و معماری شبکه‌های موردنظر معرفی شده و ویژگی‌های آن‌ها بررسی می‌شود. در مرحله دوم، شبکه‌های موردنظر با داده‌های موجود آموزش داده شده و پارامترهای تأثیرگذار در فرآیند آموزش بررسی می‌شود.

۲-۱- معرفی مدل‌های یادگیری عمیق موردنظر

همانطر که بیان شد در این تحقیق از سه مدل یادگیری عمیق *Mask-RCNN*، *MA-FCN* و *U-Net* استفاده شده و نتایج آن‌ها با یکدیگر مقایسه می‌شود. در مدل *Mask-RCNN* از تصاویر *RGB* استفاده خواهد شد ولی در مدل‌های *MA-FCN* و *U-Net* علاوه بر تصاویر *RGB*، از مدل رقومی سطح نیز در کنار این سه باند استفاده شده و تأثیرات آن در نتایج بررسی خواهد شد.

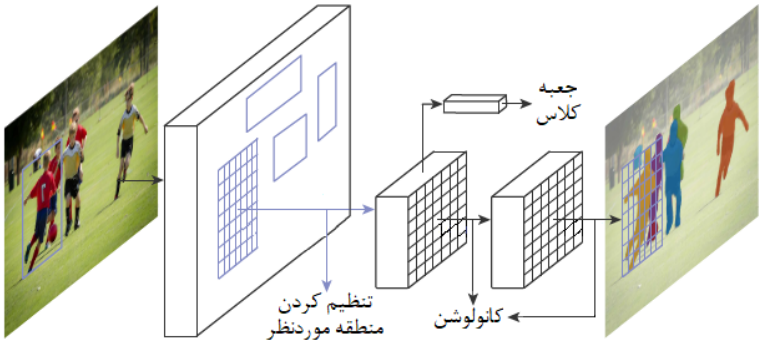
۲-۱-۱- مدل *Mask-RCNN*

Mask-RCNN، یک مدل تشخیص اشیاء از تصاویر است. این مدل، حالت پیشرفته‌ای از *Faster-RCNN* [۳۰] است که همانند *Faster-RCNN* می‌تواند نوع عارضه را تشخیص داده و یک جعبه محیطی نیز برای آن ایجاد کند. علاوه بر آن یک ماسک نیز برای آن عارضه اختصاص می‌دهد که این کار نیازمند استخراج اطلاعات دقیق‌تری از موقعیت عارضه است. این مدل از دو مرحله تشکیل شده است که در مرحله اول که *Region Proposal Network* (*RPN*) نام دارد، برای

[1] *Multi-scale Aggregation Fully Convolutional Network*

[2] *Mask Region-based Convolutional Neural Network*

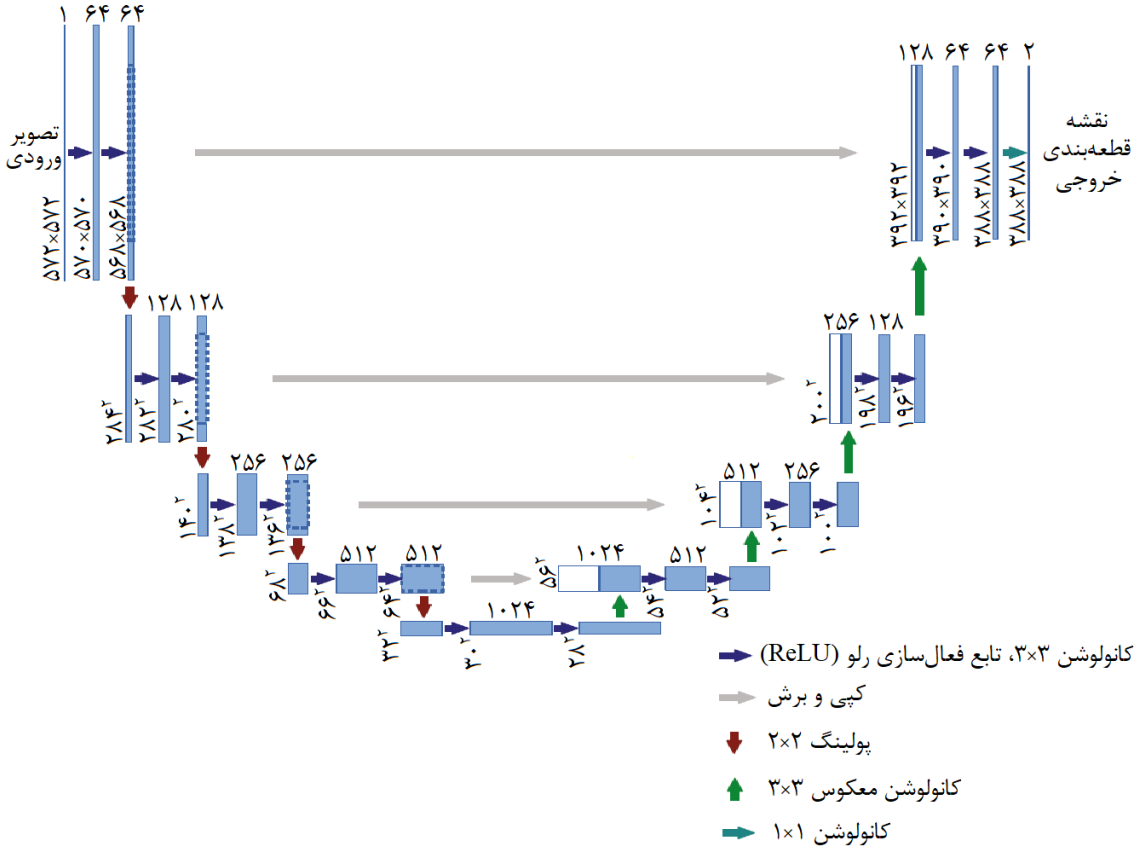
هر عارضه جعبه‌هایی محیطی استخراج می‌شود و در مرحله دوم عملیات استخراج ویژگی و طبقه‌بندی در داخل هر یک از این جعبه‌های محیطی صورت می‌گیرد. درنهایت، برای هر عارضه نوع آن، یک ماسک و یک جعبه محیطی اختصاص می‌یابد. از این مدل در تحقیقات زیادی استفاده شده و نتایج خوبی نیز به‌دست‌آمده است [۲و۳۱]. ساختار این مدل مطابق شکل(۱) است [۲۹].



شکل ۱: ساختار مدل *Mask-RCNN* [۲۹]

۲-۱-۲- مدل *U-Net*

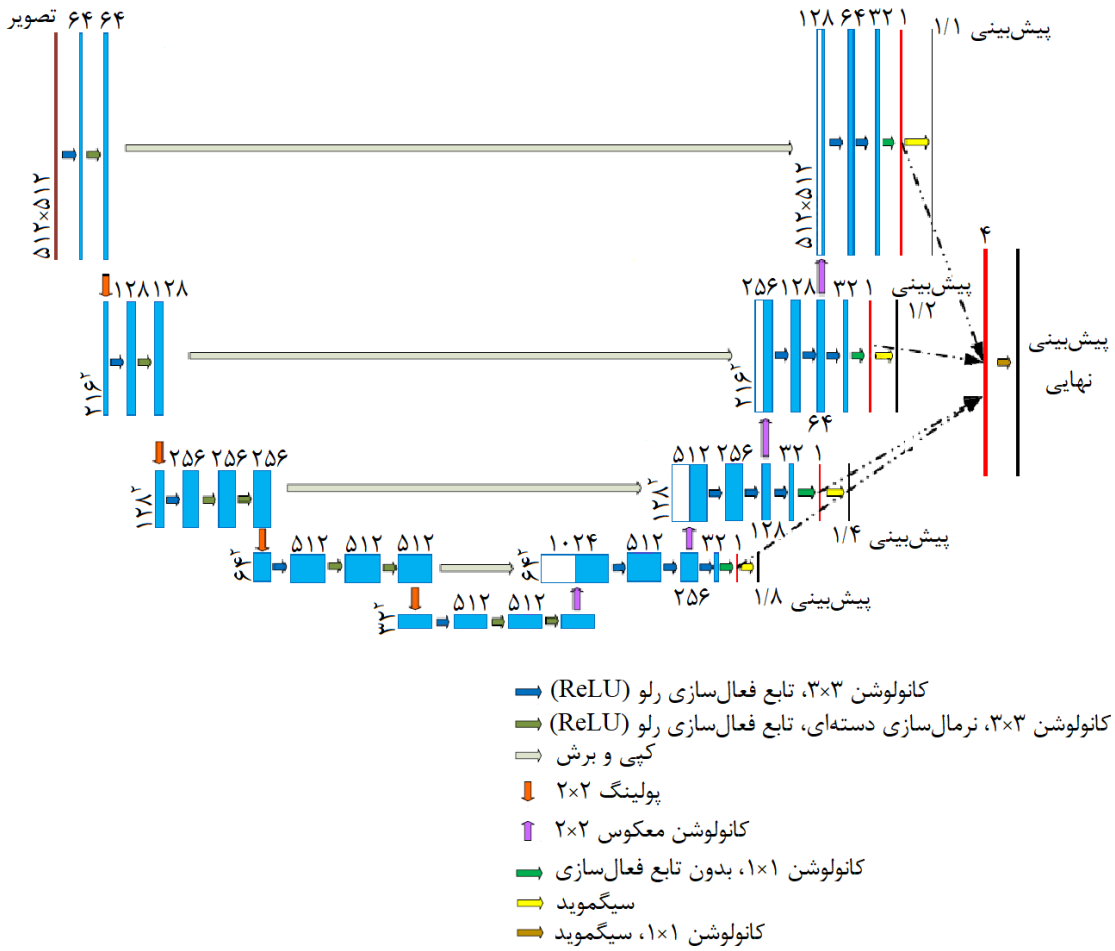
این مدل، دارای ساختار رمزگذار-رمزگشا به‌صورت متقارن است که درنهایت خروجی هم‌اندازه با تصویر ورودی ارائه می‌دهد. ابتدا تصاویر ورودی وارد قسمت رمزگذار شده و عملیات استخراج ویژگی با استفاده از شبکه‌های کانولوشنی انجام می‌شود به‌طوری‌که ابعاد خروجی هر بلوک ابتدا با استفاده از لایه پولینگ نصف شده و وارد بلوک بعدی می‌شود. بدین ترتیب خروجی قسمت رمزگذار دارای ابعاد تصویر کوچک‌تر و ابعاد ویژگی بیشتر است. برای ایجاد خروجی هم‌اندازه با تصاویر ورودی، ابعاد خروجی حاصل از قسمت رمزگذار با استفاده از لایه‌های کانولوشنی معکوس و نمونه‌برداری افزایشی در قسمت رمزگشا افزایش می‌یابد. در این حین، در هر مرحله از افزایش ابعاد تصویر، خروجی هم‌اندازه با آن از قسمت رمزگذار جهت بهبود اطلاعات مکانی تصویر، با یکدیگر ترکیب می‌شوند. ساختار کلی *U-Net* مطابق شکل(۲) است [۲۲].



شکل ۲: ساختار مدل *U-Net* [۲۲]

۲-۱-۳- مدل *MA-FCN*

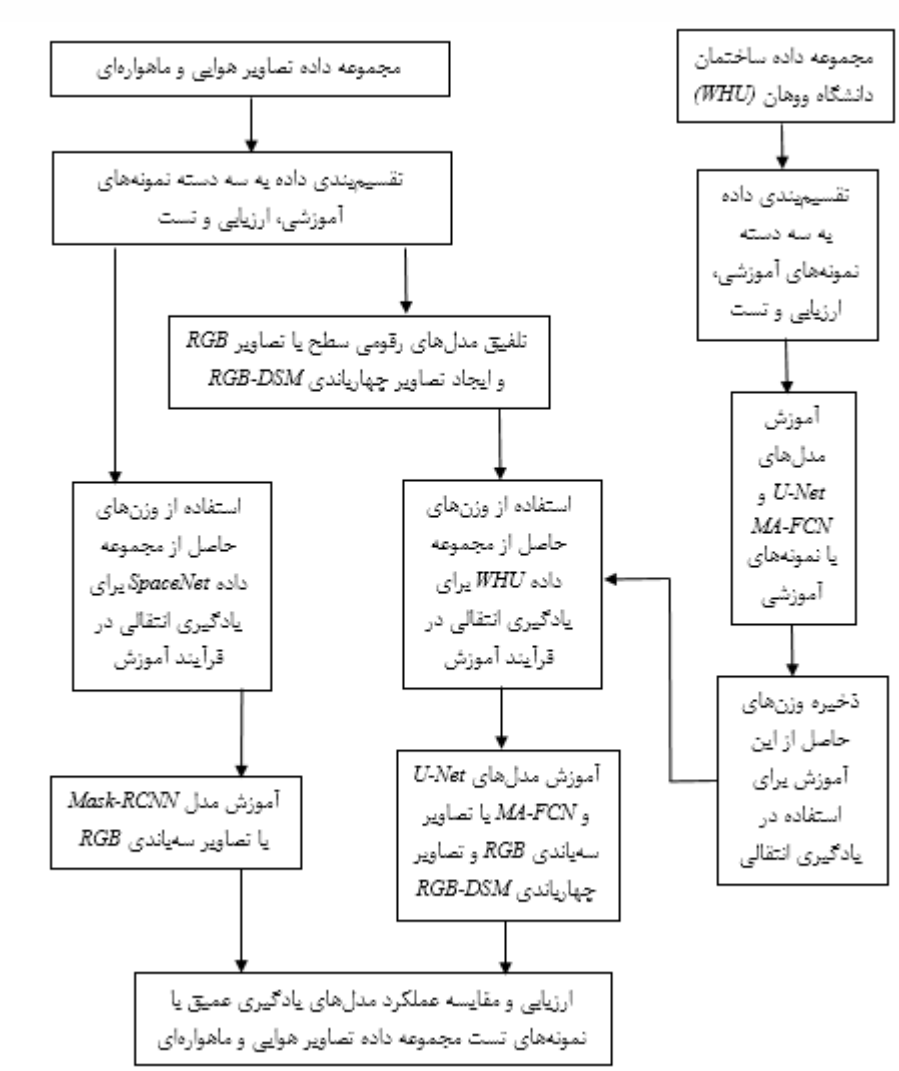
این مدل بر تأثیرات مقیاس در داده سنجش‌ازدور تمرکز می‌کند. این مدل دارای ساختار رمزگذار-رمزگشا چهار لایه *VGG*-۱۶ است که با یکدیگر در ارتباط هستند. ازآنجایی‌که در تصاویر، ساختمان‌هایی با اندازه‌های متفاوت وجود دارند، اطلاعات مربوط به مقیاس در سنجش‌ازدور مهم است. به همین دلیل در این مدل، خروجی نهایی از ترکیب خروجی‌هایی با اندازه‌های مختلف که هرکدام از سطح‌های مختلفی از مدل به دست می‌آیند، حاصل می‌شود. در قسمت رمزگذار ویژگی‌ها استخراج می‌شوند به‌طوری‌که هرچقدر از سطح‌های کم‌عمق به سطح‌های عمیق برویم، بعد مربوط به ویژگی‌ها افزایش و اندازه و توان تفکیک مکانی کاهش می‌یابد. پس خروجی‌های حاصل از سطح‌های کم‌عمق رمزگذار دارای ابعاد و توان تفکیک مکانی بالاتر ولی ابعاد ویژگی کمتر هستند و در مقابل، خروجی‌های حاصل از سطح‌های عمیق رمزگذار دارای ابعاد ویژگی بیشتر ولی ابعاد و توان تفکیک مکانی کمتر هستند. با توجه به اینکه خروجی نهایی هم شامل ویژگی‌های استخراج شده باشد و هم دارای توان تفکیک مکانی بالاتر بوده و درنهایت هم‌اندازه با تصویر ورودی باشد، در قسمت رمزگشا با استفاده از لایه‌های کانولوشنی معکوس و نمونه‌برداری افزایشی ابعاد خروجی حاصل از قسمت رمزگذار را افزایش می‌دهند. همچنین در این حین، در چند سطح خروجی‌های قسمت رمزگشا و رمزگذار را با یکدیگر ترکیب می‌کند و درنهایت خروجی نهایی از ترکیب این خروجی‌ها با استفاده از یک لایه کانولوشنی ۱×۱ حاصل می‌شود. ساختار این مدل مطابق شکل(۳) است [۱۰].



شکل ۳: ساختار مدل *MA-FCN* [۱۰]

۲-۲- آموزش شبکه‌های موردنظر

روند نمای کلی تحقیق در شکل(۴) نشان داده شده است. با توجه به اینکه هدف این تحقیق استخراج ساختمان‌ها از تصاویر هوایی و ماهواره‌ای به‌صورت خودکار با استفاده از یادگیری عمیق است، برای آموزش مدل‌های موردنظر نیاز به نمونه‌های آموزشی داریم. این نمونه‌های آموزشی بایستی شامل تصاویر هوایی و ماهواره‌ای به همراه تصاویر واقعیت زمینی باشند که این تصاویر واقعیت زمینی، تصاویری باینری هستند که ساختمان‌ها با رنگ سفید و پس‌زمینه با رنگ سیاه مشخص شده‌اند. در فرآیند آموزش، پارامترهای وزن و بایاس شبکه یادگیری عمیق طوری تغییر می‌کنند که از تصویر ورودی به تصویر واقعیت زمینی برسند. درنهایت، پس از اتمام آموزش، وزن‌های شبکه در حالتی هستند که می‌توانند برای تصویر ورودی‌ای که در فرآیند آموزش نبوده، تصویر واقعیت زمینی خروجی ایجاد کنند که در آن ساختمان‌ها دارای رنگ سفید و پس‌زمینه دارای رنگ سیاه است. سپس می‌توان با مقایسه این تصویر پیشنهادی از مدل و تصویر واقعیت زمینی اصلی مربوط به تصویر ورودی به ارزیابی عملکرد مدل پرداخت.



شکل ۴: روندنمای کلی تحقیق

برای آموزش این مدل‌ها به نمونه‌های آموزشی زیادی نیاز است زیرا این مدل‌ها دارای پارامترهای زیادی هستند و از طرفی هرچقدر تعداد نمونه‌های آموزشی زیاد باشد، برای آموزش مدل بهتر است [۳۲]. درصورتی‌که تعداد نمونه‌های آموزشی از منطقه موردنظر کم باشد، می‌توان از تکنیک یادگیری انتقالی[1] استفاده کرد. در یادگیری انتقالی، ابتدا مدل موردنظر با یک دسته داده مشابه که دارای نمونه‌های آموزشی با تعداد کافی است، آموزش داده می‌شود تا وزن‌های شبکه ایجاد شوند. سپس با استفاده از نمونه‌های آموزشی منطقه موردنظر، چندین لایه بالایی مدل آموزش داده می‌شود تا شبکه موجود بتواند نتایج مناسبی در پیش‌بینی منطقه موردنظر داشته باشد. درواقع ایده اصلی یادگیری انتقالی این است که با توجه به اینکه در قسمت‌های پایینی مدل، فرآیند استخراج ویژگی انجام می‌شود و این فرآیند استخراج ویژگی برای داده‌های مشابه (که در این تحقیق تصاویر مربوط ساختمان‌ها است)، تقریباً یکسان است، می‌توان

[1] *Transfer Learning*

وزن‌های قسمت عمده شبکه را با مجموعه داده‌ای که مشابه داده اصلی است ولی تعداد نمونه‌های آموزشی بیشتری دارد، آموزش داد و برای اینکه مدل موردنظر بتواند پیش‌بینی‌های خوبی در منطقه موردنظر داشته باشد، قسمت‌های بالایی مدل را با نمونه‌های آموزشی منطقه موردنظر آموزش داد.

در این تحقیق، برای آموزش هر سه مدل *Mask-RCNN*، *MA-FCN* و *U-Net* از یادگیری انتقالی استفاده شده است. همچنین تصاویر واقعیت زمینی منطقه موردنظر به‌صورت دستی ایجاد شده و سپس به سه دسته آموزش، ارزیابی و تست تقسیم‌بندی شده است. برای آموزش *Mask-RCNN*، ابتدا وزن‌هایی را که از آموزش این مدل برای استخراج ساختمان‌ها وجود دارد را در مدل بارگذاری کرده، سپس با استفاده نمونه‌های آموزشی منطقه موردنظر، مدل را آموزش می‌دهیم.

برخلاف *Mask-RCNN* برای مدل‌های *MA-FCN* و *U-Net* وزن‌های از پیش آماده در دسترس نبوده و به همین دلیل ابتدا این مدل‌ها را با مجموعه داده *WHU* [۳] که تعداد نمونه‌های آموزشی مناسبی دارد، آموزش می‌دهیم. سپس با بارگذاری این وزن‌ها در این مدل‌ها، دوباره آن‌ها را یک‌بار با نمونه‌های آموزشی سه باندی منطقه موردنظر و بار دیگر با نمونه‌های آموزشی چهار باندی منطقه موردنظر که شامل مدل رقومی سطح منطقه هستند، آموزش می‌دهیم و نتایج را بررسی می‌کنیم.

۳- پیاده‌سازی و ارزیابی نتایج

در این بخش، سه مجموعه داده که برای آموزش و ارزیابی عملکرد مدل‌های *Mask-RCNN*، *U-Net* و *MA-FCN* مورد استفاده قرار می‌گیرند، معرفی می‌شود. همچنین نتایج حاصل از مدل‌های موردنظر در این سه مجموعه داده بررسی می‌شود.

۳-۱- مجموعه داده

مجموعه داده مورداستفاده در یادگیری عمیق از مهم‌ترین عوامل تأثیرگذار در عملکرد مدل است. داده مورد استفاده علاوه بر اینکه باید دارای تعداد نمونه آموزشی کافی و مناسب برای آموزش شبکه موردنظر باشد، بایستی بیانگر ویژگی‌های موردنظر با کیفیت مطلوب باشد. در شبکه‌های عمیق‌تر با توجه به اینکه تعداد پارامترها افزایش می‌یابد، مجموعه داده از اهمیت بسیار بالایی برخوردار است. در این تحقیق به منظور انجام یک ارزیابی کامل از مجموعه داده ساختمان دانشگاه ووهان[1] *(WHU)*، مجموعه داده مسابقه تلفیق داده سال ۲۰۱۹ سازمان مهندسین برق و الکترونیک *(IEEE)*[2] و مجموعه داده منطقه جینگهای چین[3] استفاده شده است.

۳-۱-۱- مجموعه داده *WHU*

مجموعه داده *WHU* از تصاویر هوایی که شامل ساختمان‌ها هستند، تشکیل شده است [۳]. در این مجموعه داده، تصاویر هوایی از ساختمان‌هایی با اندازه، رنگ و شکل‌های مختلف وجود دارند که آن را به داده‌ای مناسب جهت آموزش و ارزیابی مدل تبدیل می‌کند. در این مجموعه داده، ۸۱۸۸ تصویر هوایی با ابعاد ۵۱۲×۵۱۲ پیکسل وجود دارد که هر تصویر دارای یک تصویر باینری هم‌اندازه با آن به‌عنوان تصویر واقعیت زمینی است که در آن ساختمان‌ها با رنگ سفید و پس‌زمینه با رنگ سیاه مشخص شده‌اند. نمونه‌ای از این مجموعه داده در شکل(۵) قابل‌مشاهده است [۳].

---

[1] *Wuhan University Building Dataset (WHU)*

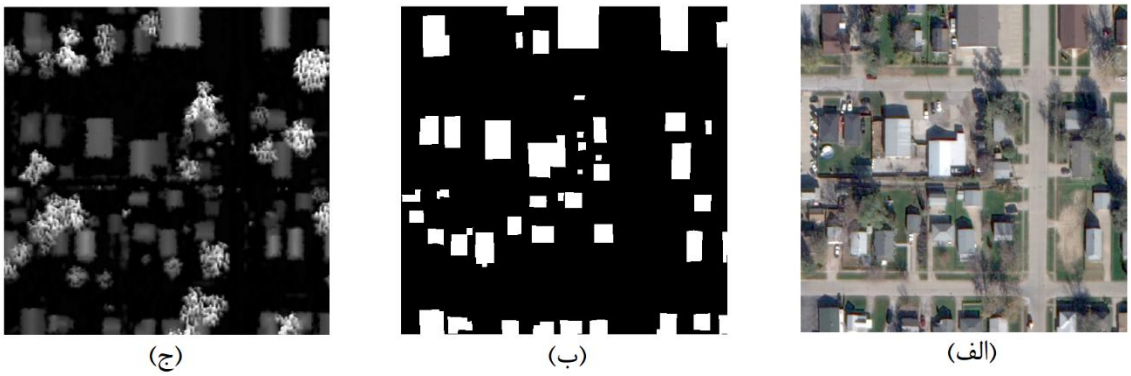
[2] *Institute of Electrical and Electronics Engineers Data Fusion Contest 2019*

[3] *Jinghai District*

شکل۵: یک تصویر نمونه به همراه تصویر واقعیت زمینی آن در دسته داده *WHU*، الف) تصویر هوایی نمونه، ب) تصویر واقعیت زمینی

۳-۱-۲- تصاویر ماهواره‌ای مجموعه داده مسابقه تلفیق داده سال ۲۰۱۹ سازمان *IEEE*

این مجموعه داده توسط سازمان *IEEE* در سال ۲۰۱۹ با هدف بازسازی مدل‌های سه‌بعدی و تولید نقشه‌های موضوعی ایجاد شده است [۳۳]. این مجموعه داده شامل تصاویر ماهواره‌ای و ابرنقطه لیدار است. با استفاده از ابرنقطه لیدار، مدل رقومی سطح برای این مجموعه داده ایجاد شده است. از این مجموعه داده، ۸۳ تصویر ماهواره‌ای با ابعاد ۱۰۲۴×۱۰۲۴ برای این تحقیق انتخاب شده است. تصاویر واقعیت زمینی این مجموعه داده به‌صورت دستی ایجاد شده است. از این مجموعه داده، ۶۶ تصویر برای آموزش، ۷ تصویر به‌عنوان داده‌های ارزیابی و ۱۰ تصویر برای تست انتخاب شده است. نمونه‌ای از این مجموعه داده در شکل(۶) نشان داده شده است.

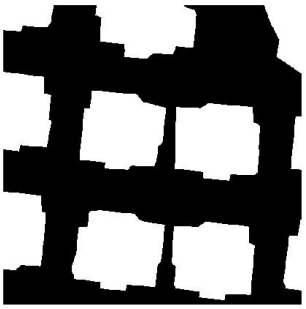


شکل ۶: یک تصویر نمونه به همراه تصویر واقعیت زمینی آن در مجموعه داده مسابقه تلفیق داده سال ۲۰۱۹ سازمان *IEEE*، الف) تصویر ماهواره‌ای سنجنده ، ب) تصویر واقعیت زمینی، ج) مدل رقومی سطح

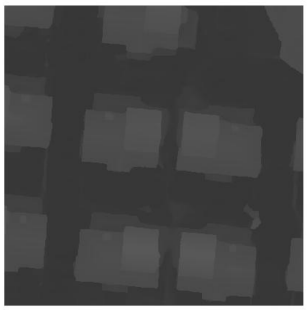
۳-۱-۳-تصاویر هوایی منطقه جینگهای در چین

مساحت این منطقه در حدود ۴ کیلومترمربع بوده و دارای اختلافات ارتفاعی ۱۰۸ متری است [۲]. این منطقه دارای ساختمان‌هایی صنعتی و مسکونی با اندازه، شکل و ارتفاع‌های مختلف است. داده‌های موجود از این منطقه، شامل تصاویر هوایی با قدرت تفکیک مکانی ۰/۰۵ متری به همراه مدل رقومی سطح هستند. از این منطقه در کل ۲۲۴ تصویر هوایی با ابعاد ۱۰۲۴×۱۰۲۴ وجود دارد که برای آموزش و ارزیابی عملکرد مدل‌های موردنظر استفاده می‌شود. از ۲۲۴ تصویر موجود، ۱۸۰ تصویر برای آموزش، ۲۲ تصویر به‌عنوان داده‌های ارزیابی و ۲۲ تصویر برای تست انتخاب شده است. نمونه‌ای از این مجموعه داده به همراه تصویر واقعیت زمینی و مدل رقومی سطح در شکل(۷) قابل‌مشاهده است [۲].

(الف)



(ب)



(ج)

شکل۷: یک تصویر نمونه به همراه تصویر واقعیت زمینی و مدل رقومی سطح آن در دسته داده جینگهای، الف) تصویر هوایی، ب) تصویر واقعیت زمینی، ج) مدل رقومی سطح

۳-۲-معیارهای ارزیابی

برای ارزیابی عملکرد مدل‌های آموزش‌دیده از مجموعه داده تست استفاده می‌شود. برای این منظور، بایستی تصاویر تست که در مرحله آموزش وجود نداشتند وارد مدل شود که مدل با استفاده از وزن‌هایی که در حین آموزش تعیین شده‌اند، عملیات پیش‌بینی را روی تصاویر تست انجام دهد. درنهایت با مقایسه این پیش‌بینی‌های حاصل از مدل با تصاویر واقعیت زمینی اصلی تصاویر تست، فرآیند ارزیابی مدل انجام می‌شود. برای ارزیابی عملکرد مدل‌ها از معیارهای ارزیابی متداول برای استخراج ساختمان شامل صحت[۱]، خطا[۲]، *IOU*[۳] و نمره *F*۱[۴] استفاده می‌شود[۳۴].

۴-نتایج و بحث

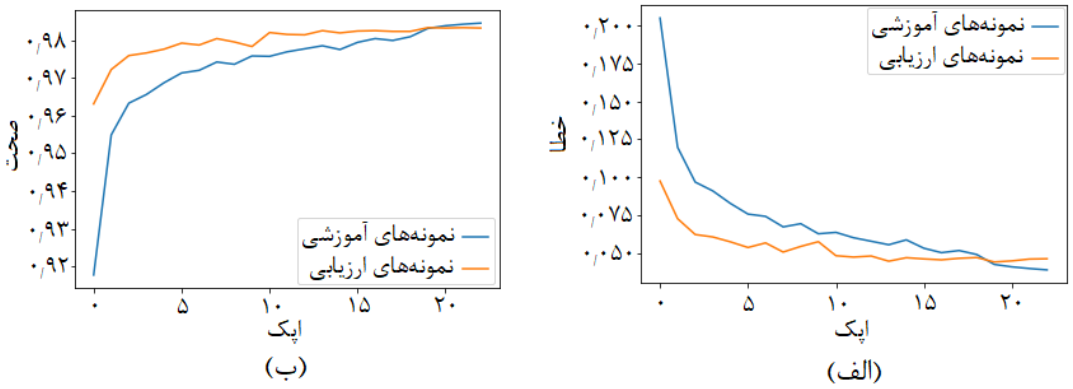
در این بخش، نتایج حاصل از مدل‌های مورد مقایسه در داده‌های مختلف به همراه پیش‌بینی‌های حاصل از این مدل‌ها در تصاویر تست ارائه و بررسی خواهد شد. همچنین نقاط ضعف و قوت مدل‌های موردنظر و اثرات نوع داده مورداستفاده در این مدل‌ها بررسی می‌شود.

۴-۱-نتایج حاصل از مجموعه داده *WHU*

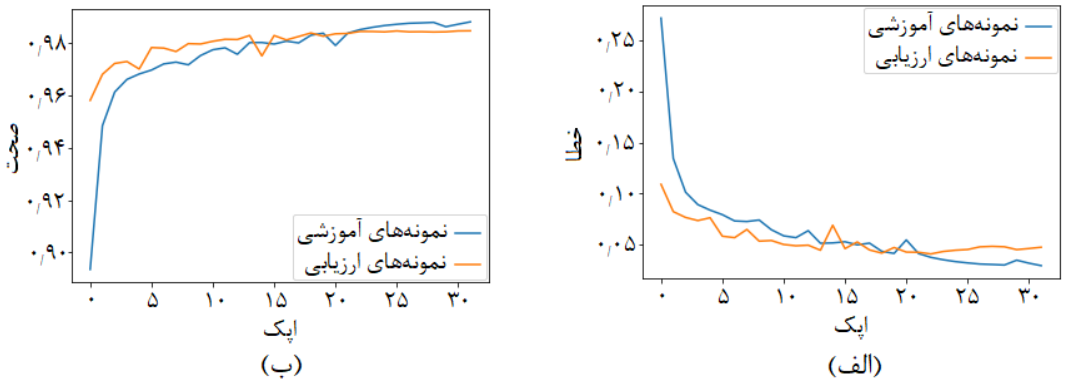
مدل‌های *U-Net* و *MA-FCN* ابتدا با استفاده از مجموعه داده *WHU* آموزش داده شده است تا از این وزن‌ها در آموزش این مدل‌ها با استفاده از داده‌های ماهواره‌ای و هوایی معرفی‌شده استفاده شود. نمودارهای مربوط به خطا و صحت آموزش مدل‌های *U-Net* و *MA-FCN* با استفاده از مجموعه داده *WHU* به‌ترتیب

[1] *accuracy*
[2] *loss*
[3] *Intersection over Union*
[4] *F1-score*

در شکل(۸) و شکل(۹) نشان داده شده است. با توجه به اینکه مدل *U-Net* پارامترهای کمتری نسبت به مدل *MA-FCN* دارد، در تعداد مراحل آموزشی کمتری به صحت موردنظر دست‌یافته است.



شکل ۸: الف) نمودار خطا در آموزش *U-Net* با مجموعه داده *WHU*، ب) نمودار صحت در آموزش *U-Net* با مجموعه داده *WHU*

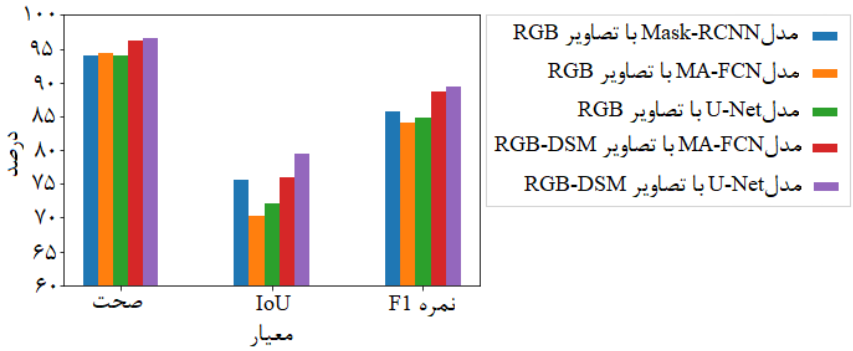


شکل ۹: الف) نمودار خطا در آموزش *MA-FCN* با مجموعه داده *WHU*، ب) نمودار صحت در آموزش *MA-FCN* با مجموعه داده

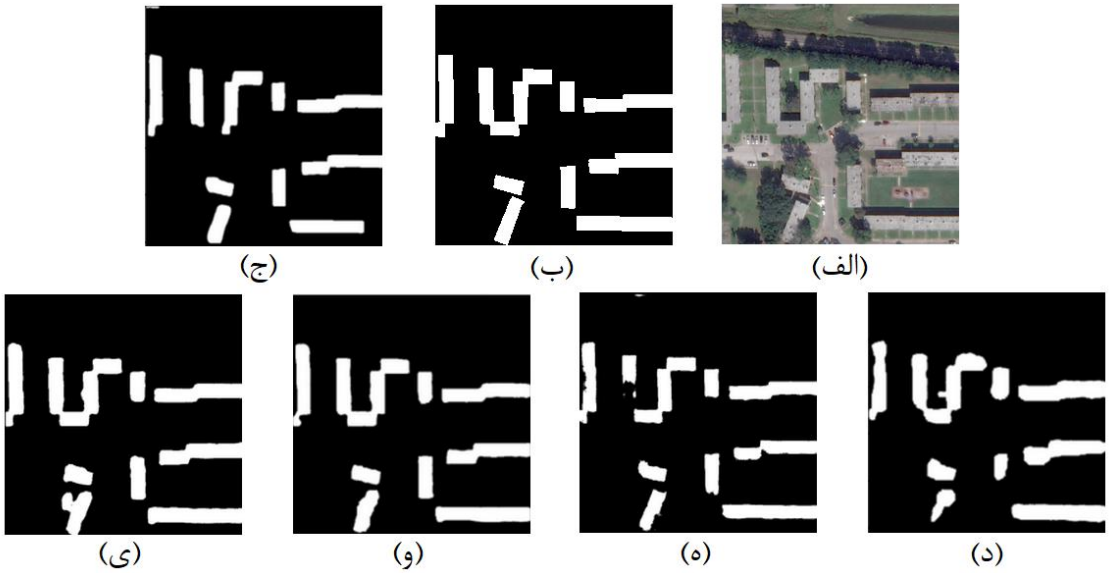
۴-۲-نتایج حاصل از مجموعه تصاویر ماهواره‌ای

همانطور که پیشتر بیان شد، در آموزش هر سه مدل موردنظر با استفاده از این مجموعه داده از تکنیک یادگیری انتقالی استفاده شده است. وزن‌های اولیه مورداستفاده برای یادگیری انتقالی در مدل‌های *U-Net* و *MA-FCN* با استفاده از مجموعه داده *WHU* به‌دست‌آمده است. همچنین برای مدل *Mask-RCNN* از وزن‌های آماده حاصل از مجموعه داده *SpaceNet* برای یادگیری انتقالی استفاده شده است[۳۵].

مدل *Mask-RCNN* با استفاده از تصاویر سه‌باندی و مدل‌های *U-Net* و *MA-FCN* هم با تصاویر سه‌باندی و هم با تصاویر چهار باندی، که باند چهارم را مدل رقومی سطح تشکیل می‌دهد، آموزش داده شده است. نتایج و پیش‌بینی‌های حاصل از مدل‌های موردنظر بر روی تصویر تست ماهواره‌ای به همراه تصویر واقعیت زمینی به‌ترتیب در شکل(۱۰) و شکل(۱۱) نشان داده شده است.



شکل ۱۰: نتایج سه مدل موردنظر در مجموعه داده تست تصاویر ماهواره‌ای



شکل ۱۱: الف) تصویر تست ماهواره‌ای، ب) تصویر واقعیت زمینی، ج) پیش‌بینی مدل *Mask-RCNN*، د) پیش‌بینی مدل *MA-FCN* با تصویر سه‌باندی، ه) پیش‌بینی مدل *U-Net* با تصویر سه‌باندی، و) پیش‌بینی مدل *MA-FCN* با تصویر چهار باندی، ی) پیش‌بینی مدل *U-Net* با تصویر چهار باندی

در بین مدل‌های معرفی‌شده، مدل *Mask-RCNN* در تصاویر سه‌باندی نتایج بهتری نسبت به مدل‌های *U-Net* و *MA-FCN* داشته است. بدیهی است که استفاده از مدل رقومی سطح که نشان‌دهنده ارتفاع نقاط بوده و ویژگی مناسبی برای تمایز ساختمان‌ها از دیگر عوارض است، نتایج را نسبت به تصاویر سه‌باندی بهبود داده است. استفاده از مدل رقومی سطح در کنار تصاویر سه‌باندی نتایج حاصل از دو مدل *MA-FCN* و *U-Net* را به‌ترتیب ٪ ۵٫۷ و ٪ ۷٫۴۶ در معیار *IOU* و ٪ ۴٫۴۸ و ٪ ۴٫۵۷ در معیار *F1-score* بهبود داده است.

۴-۲-۱-بحث

در مجموعه داده ماهواره‌ای به دلایل مختلف نظیر کم بودن تعداد نمونه‌های آموزشی، تنوع زیاد ساختمان‌ها نسبت به تعداد تصاویر و همچنین وجود خطای هم‌مرجع‌سازی بین مدل‌های رقومی سطح با تصاویر ماهواره‌ای، داده‌ای چالش‌برانگیز و پیچیده برای مدل‌های مختلف یادگیری عمیق جهت استخراج ساختمان‌ها است.

در یادگیری عمیق اگر تعداد نمونه‌های آموزشی کم باشد، مهم‌ترین مسئله‌ای که بایستی به آن توجه شود، مسئله فرابرازش[1] مدل است [۲۸]. از بین مدل‌های موردنظر، مدل *Mask-RCNN* به‌دلیل استفاده از ساختار *ResNet*۱۰۱ در معماری خود، نسبت به مسئله فرابرازش مقاوم‌تر است چون ساختار *ResNet* از مسئله محوشدگی گرادیان‌ها[2] جلوگیری می‌کند. همچنین این مدل برخلاف مدل‌های *U-Net* و *MA-FCN* که برای کل تصویر عملیات تشخیص ساختمان و ایجاد ماسک را انجام می‌دهند، به‌صورت عارضه مبنا کار می‌کند. یعنی ابتدا ساختمان‌ها را به‌صورت جداگانه تشخیص داده و برای هرکدام یک ماسک ایجاد می‌کند. چون در این مجموعه داده، تعداد نمونه‌های آموزشی کم است ممکن است مدل‌های *U-Net* و *MA-FCN* فرصت کافی برای یادگیری اطلاعات لازم جهت استخراج ساختمان‌ها به‌صورت یکجا نداشته باشند. به همین دلیل این مدل عملکرد بهتری نسبت به مدل‌های *MA-FCN* و *U-Net* در این مجموعه داده داشته است.

به دلیل اینکه عارضه موردنظر برای استخراج در این تحقیق ساختمان است، استخراج دقیق مرز ساختمان‌ها اهمیت بسیار بالایی دارد. از بین مدل‌های موردنظر، دو مدل *U-Net* و *MA-FCN* به دلیل اینکه قبل از انجام پیش‌بینی روی تصویر موردنظر، در چند سطح ویژگی‌های قسمت رمزگذار، که اطلاعات مکانی بیشتری دارند، را با ویژگی‌های قسمت رمزگشا ترکیب می‌کنند، اطلاعات مکانی بهتری دارند و نسبت به مدل *Mask-RCNN* در مرز ساختمان‌ها دقیق‌تر هستند.

با توجه به اینکه تعداد نمونه‌های آموزشی در این مدل کم است از بین دو مدل *U-Net* و *MA-FCN*، مدل *U-Net* به دلیل اینکه ساختار ساده‌تری داشته و دارای پارامترهای کمتری است، نسبت به مدل *MA-FCN* بهتر عمل کرده است.

۴-۳-نتایج حاصل از مجموعه تصاویر هوایی

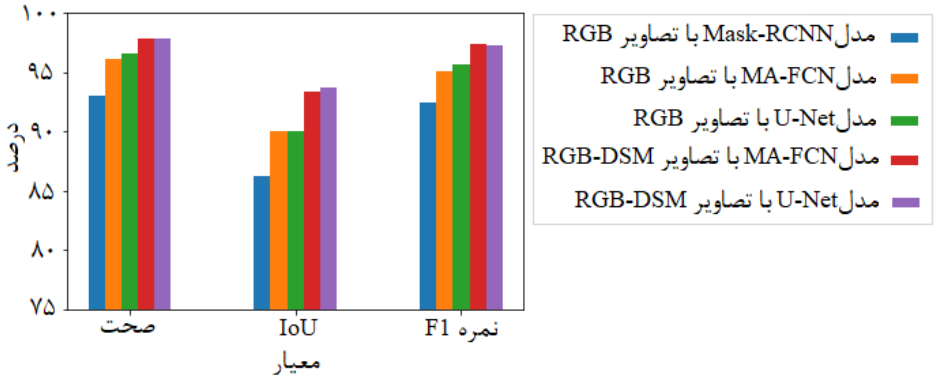
در آموزش هر سه مدل موردنظر با استفاده از این مجموعه داده نیز از تکنیک یادگیری انتقالی استفاده شده است. مدل *Mask-RCNN* با استفاده از تصاویر سه‌باندی و مدل‌های *U-Net* و *MA-FCN* هم با تصاویر سه‌باندی و هم با تصاویر چهار باندی، که باند چهارم را مدل رقومی سطح تشکیل می‌دهد، آموزش داده شده است. نتایج و پیش‌بینی‌های حاصل از مدل‌های موردنظر بر روی تصویر تست هوایی به همراه تصویر واقعیت زمینی به‌ترتیب در شکل(۱۲) و شکل(۱۳) نشان داده شده است.

در این مجموعه داده، در تصاویر سه‌باندی مدل *U-Net* نتایج بهتری نسبت به سایر مدل‌ها داشته است. بدیهی است که استفاده از مدل رقومی سطح به دلیل داشتن ارتفاع که ویژگی مناسبی برای تمایز ساختمان‌ها از دیگر عوارض است،

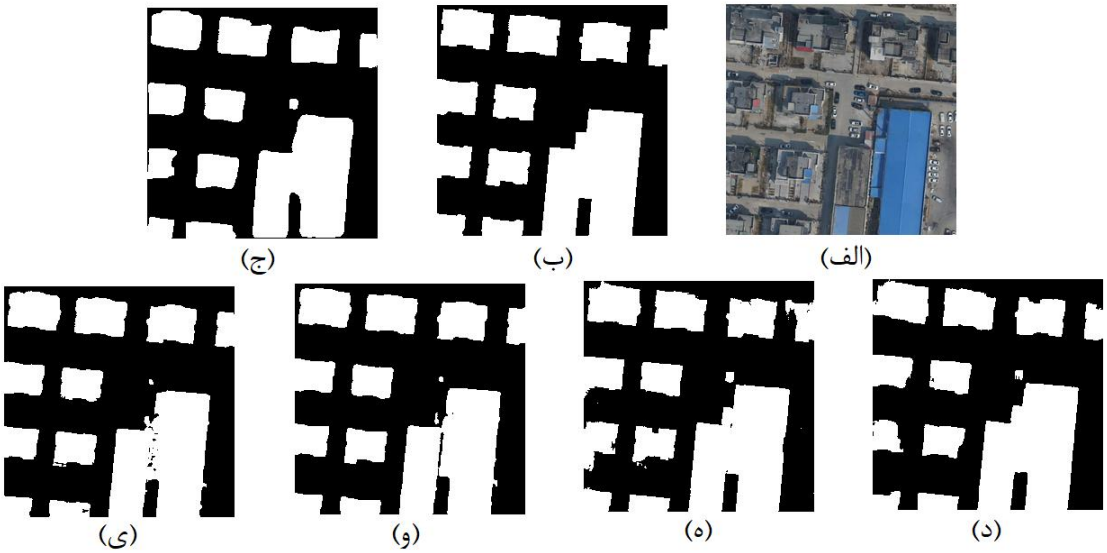
نتایج را نسبت به تصاویر سه‌باندی بهبود داده است. استفاده از مدل رقومی سطح در کنار تصاویر سه‌باندی نتایج حاصل از دو مدل *MA-FCN* و *U-Net* را به‌ترتیب ٪ ۳٫۳۴ و ٪ ۳٫۶۱ در معیار *IOU* و ٪ ۲٫۳۱ و ٪ ۱٫۵۶ در معیار *F۱-score* بهبود داده است.

[1] *Overfitting*

[2] *Vanishing Gradients*



شکل ۱۲: نتایج سه مدل موردنظر در مجموعه داده تست تصاویر هوایی



شکل ۱۳: الف) تصویر تست هوایی، ب) تصویر واقعیت زمینی، ج) پیش‌بینی مدل *Mask-RCNN*، د) پیش‌بینی مدل *MA-FCN* با تصویر سه‌باندی، ه) پیش‌بینی مدل *U-Net* با تصویر سه‌باندی، و) پیش‌بینی مدل *MA-FCN* با تصویر چهار باندی، ی) پیش‌بینی مدل *U-Net* با تصویر چهار باندی

۴-۳-۱-بحث

در این مجموعه داده، مدل‌های یادگیری عمیق مورد مقایسه به دلیل وجود نمونه‌های آموزشی بیشتر، توان تفکیک مکانی بالاتر نسبت به تصاویر ماهواره‌ای و هم‌مرجع بودن دقیق مدل‌های رقومی سطح با تصاویر، نتایج بهتری نسبت به مجموعه داده قبلی دارند.

در این مجموعه داده، هر سه مدل موردنظر نتایج بهتری نسبت به مجموعه داده قبلی دارند که نشان می‌دهد تعداد و توان تفکیک نمونه‌های آموزشی در عملکرد مدل‌ها تأثیرگذار است به‌طوری‌که هر چه تعداد نمونه‌های آموزشی بیشتر و توان تفکیک مکانی تصاویر بالاتر باشد، مدل‌های موردنظر اطلاعات بیشتر و دقیق‌تری یاد گرفته و پیش‌بینی‌های بهتری ارائه می‌دهند.

در این مجموعه داده به دلیل اینکه تعداد نمونه‌های آموزشی نسبت به مجموعه داده قبلی بیشتر است،

احتمال وقوع فرابرازش کمتر است و به همین دلیل مزیت مدل *Mask-RCNN* نسبت به دو مدل دیگر، که دارا بودن ساختار *ResNet*۱۰۱ و مقاوم‌تر بودن به مسئله فرابرازش است، در این مجموعه داده نمایان نشده است. نقاط ضعف مدل *Mask-RCNN* در این است که از تعداد لایه‌های کمی برای ایجاد ماسک استفاده می‌کند که باعث می‌شود نقشه باینری حاصل از آن نسبت به مدل‌های *U-Net* و *MA-FCN* دقت کمتری داشته باشد. همچنین به دلیل اینکه پیش‌بینی آن با تصویر ورودی هم‌اندازه باشد، فرآیند نمونه‌برداری افزایشی را بدون استفاده از ترکیب اطلاعات مکانی انجام می‌دهد که باعث می‌شود میزان اطلاعات از دست رفته بیشتر شده و در مرز ساختمان‌ها دقیق نباشد. به همین دلیل در این مجموعه داده مدل *Mask-RCNN* نسبت به سایر مدل‌ها ضعیف‌تر عمل کرده است.

نتایج حاصل از دو مدل *U-Net* و *MA-FCN* در این مجموعه داده هم در تصاویر سه‌باندی و هم در تصاویر چهاربانده نزدیک به هم هستند. این دو مدل ساختار شبیه به هم دارند با این تفاوت که مدل *MA-FCN* شبکه‌های کانولوشنی بیشتر و عمیق‌تری دارد و برای ایجاد پیش‌بینی نهایی، از چند سطح قسمت رمزگشا خروجی گرفته و آن‌ها را با یکدیگر ترکیب می‌کند و به این دلایل پارامترهای بیشتری نسبت به مدل *U-Net* دارد. با توجه به اینکه مدل *MA-FCN* مدلی عمیق‌تر با پارامترهایی بیشتر نسبت به مدل *U-Net* است، ممکن است در مجموعه داده‌ای که تعداد نمونه‌های آموزشی به‌اندازه کافی بالا باشد، به دلیل دارا بودن ساختاری عمیق‌تر می‌تواند اطلاعات بیشتری استخراج کرده و نتایج بهتری داشته باشد.

۵-نتیجه‌گیری

یادگیری عمیق روشی مناسب برای استخراج ساختمان‌ها از تصاویر هوایی و ماهواره‌ای به‌صورت خودکار است. در این تحقیق عملکرد سه مدل یادگیری عمیق در استخراج عوارض ساختمانی از تصاویر هوایی و ماهواره‌ای ارزیابی شد. در مجموعه داده‌ای با تعداد نمونه‌های آموزشی کم، مسئله فرابرازش مهم بوده و استفاده از مدلی که به فرابرازش مقاوم‌تر است، بهتر است. درصورتی‌که تعداد نمونه‌های آموزشی به‌اندازه کافی بالا باشد، مسئله فرابرازش کم‌رنگ‌تر می‌شود و می‌توان از مدل‌هایی عمیق‌تر استفاده کرد. همچنین در موضوع استخراج ساختمان‌ها بهتر است از مدل‌هایی استفاده شود که خروجی آن‌ها دارای اطلاعات مکانی بهتری بوده و از دست رفتگی اطلاعات در آن‌ها کم است. به همین دلیل، جهت استخراج دقیق‌تر مرز ساختمان‌ها، استفاده از مدل‌هایی که برای بازیابی اطلاعات مکانی از ترکیب ویژگی‌های قسمت‌های رمزگذار و رمزگشا در پیش‌بینی نهایی استفاده می‌کنند، بهتر است. درنهایت علاوه بر مدل، تعداد نمونه‌های آموزشی، تنوع ساختمان‌های موجود در تصاویر، توان تفکیک مکانی تصاویر و کیفیت مدل رقومی سطح مورداستفاده نیز در نتایج تأثیرگذار هستند.

در نهایت موارد زیر به عنوان تحقیقات آتی این تحقیق پیشنهاد می‌شود:

- ارزیابی الگوریتم‌های مورد مقایسه در انواع دیگر از تصاویر با شرایط مختلف نظیر تصاویر با سطوح نویز متفاوت و یا تصاویر راداری
- ترکیب انواع مختلف منابع داده
- مقیاسه دیگر الگوریتم‌ها موجود
- بررسی اثر خطای هم مرجع سازی میان منابع مختلف داده در کیفیت ساختمان های مستخرج
- توسعه مدل‌های جدید یادگیری عمیق برای غلبه بر مشکلات الگوریتم‌های موجود

مراجع

[1] Z. Li, Q. Xin, Y. Sun, and M. Cao, "A Deep Learning-Based Framework for Automated Extraction of Building Footprint Polygons from Very High-Resolution Aerial Imagery," Remote Sensing, vol. 13, no. 18, p. 3630, 2021.

[2] D. Yu, S. Ji, J. Liu, and S. Wei, "Automatic 3D building reconstruction from multi-view aerial images with deep learning," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 171, pp. 155-170, 2021.

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

[4] K .Rastogi, P. Bodani, and S. A. Sharma, "Automatic building footprint extraction from very high-resolution imagery using deep learning techniques," Geocarto International, vol. 37, no. 5, pp. 1501-1513, 2022/03/04 2022

[5] S. Ji, S. Wei, and M. Lu, "A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery," International Journal of Remote Sensing, vol. 40, no. 9, pp. 3308-3322, 2019/05/03 2019

[6] J. Yuan, "Learning Building Extraction in Aerial Scenes with Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 11, pp. 2793-2798, 2018

[7] A. Huertas and R. Nevatia, "Detecting buildings in aerial images," Computer vision, graphics, and image processing, vol. 41, no. 2, pp. 131-152, 1988.

[8] N. L. Gavankar and S. K. Ghosh, "Automatic building footprint extraction from high-resolution satellite image using mathematical morphology," European Journal of Remote Sensing, vol. 51, no. 1, pp. 182-193, 2018.

[9] X. Huang, L. Zhang, and T. Zhu, "Building Change Detection From Multitemporal High-Resolution Remotely Sensed Images Based on a Morphological Building Index," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 1, pp. 105-115, 2014

[10] S. Wei, S. Ji, and M. Lu, "Toward Automatic Building Footprint Delineation From Aerial Images Using CNN and Regularization," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 3, pp. 2178-2189, 2020

[11] S. Du, F. Zhang, and X. Zhang, "Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 105, pp. 107-119, 2015/07/01/ 2015

[12] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 105-109, 2016

[13] W. Feng, H. Sui, L. Hua, C. Xu, G. Ma, and W. Huang, "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, vol. 41, no. 17, pp. 6595-6617, 2020/09/01 2020

[14] J. Ma, L. Wu, X. Tang, F. Liu, X. Zhang, and L. Jiao, "Building Extraction of Aerial Images by a Global and Multi-Scale Encoder-Decoder Network," Remote Sensing, vol. 12, no. 15, p. 2350, 2020

[15] Z. Shao, P. Tang, Z. Wang, N. Saleem,

S. Yam, and C. Sommai, "BRRNet: A Fully Convolutional Neural Network for Automatic Building Extraction From High-Resolution Remote Sensing Images," Remote Sensing, vol. 12, no .6 ,p. 1050, 2020

[16] D. Marmanis, J. D. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla, "Semantic segmentation of aerial images with an ensemble of CNSS," ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, vol. 3, pp. 473-480, 2016.

[17] X. Liu, M. Chi, Y. Zhang, and Y. Qin, "Classifying High Resolution Remote Sensing Images by Fine-Tuned VGG Deep Networks," in IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 22-27 July 2018 2018, pp. 7137-7140

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7-12 June 2015 2015, pp. 3431-3440

[19] W. Zhao and S. Du, "Learning multiscale and deep representations for classifying remotely sensed imagery," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 113, pp. 155-165, 2016/03/01/ 2016

[20] A. Milosavljević, "Automated Processing of Remote Sensing Imagery Using Deep Semantic Segmentation: A Building Footprint Extraction Case," ISPRS International Journal of Geo-Information, vol. 9, no. 8, p. 486, 2020

[21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 2017

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015// 2015: Springer International Publishing, pp. 234-241 .

[23] N. Audebert, B. L. Saux ,and S. Lefèvre, "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks," in Asian conference on computer vision, 2016: Springer, pp. 180-196 .

[24] Y. Xu, L. Wu, Z. Xie, and Z. Chen, "Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters," Remote Sensing, vol. 10, no. 1, p. 144, 2018

[25] H. Hosseinpour, F. Samadzadegan, and F. D. Javan, "CMGFNet: A deep cross-modal gated fusion network for building extraction from very high-resolution remote sensing images," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 184, pp. 96-115, 2022/02/01/ 2022

[26] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9 .

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[28] L. Luo, P. Li, and X. Yan, "Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review," Energies, vol. 14, no. 23, p. 7982, 2021

[29] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on

computer vision, 2017, pp. 2961-2969 .

[30] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448 .

[31] K. Zhao, J. Kang, J. Jung, and G. Sohn, "Building extraction from satellite images using mask R-CNN with building boundary regularization," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 247-251 .

[32] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.", 2022.

[33] Bertrand Le Saux, Naoto Yokoya, Ronny Hänsch, Myron Brown, October 31, 2019, "Data Fusion Contest 2019 (DFC2019)", IEEE Dataport

[34] A. Abdollahi, B. Pradhan, S. Gite, and A. Alamri, "Building Footprint Extraction from High Resolution Aerial Images Using Generative Adversarial Network (GAN) Architecture," IEEE Access, vol. 8, pp. 209517-209527, 2020

[35] Van Etten, A., Lindenbaum, D., & Bacastow, T.M. (2018). SpaceNet: A Remote Sensing Dataset and Challenge Series. ArXiv, abs/1807.01232 .

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

Journal of Geospatial Information Technology
Vol.11, No.1, Spring 2023

Research Paper

The performance evaluation of three deep learning models in building footprint extraction from aerial and satellite images

Nima Ahmadian *[1], Amin Sedaghat [2], Nazila Mohammadi [3]

1- Ms.c student of remote sensing in Department of Geomatics, Faculty of Civil Engineering, University of Tabriz
2- Associate professor in Department of Geomatics, Faculty of Civil Engineering, University of Tabriz
3 - Assistant professor in Department of Geomatics, Faculty of Civil Engineering, University of Tabriz

Abstract

Buildings as one of the important man-made objects have various applications and need to be observed and detected with aerial and satellite images. Deep learning models have recently been used to automatically extract building footprints from aerial and satellite images. It is essential to evaluate and compare the features of different deep learning models in images with geometric and brightness variations. For this purpose, in this research the performance of three deep learning models called Mask-RCNN (Mask Region-based Convolutional Neural Network), U-Net and MA-FCN (Multi-scale Aggregation Fully Convolutional Network) in building footprint extraction from two aerial and satellite datasets is evaluated with F1-score and IOU metrics. The results of this research indicate that the model, quantity and quality of training samples and digital surface model affect the performance of these models. Furthermore, using digital surface models along with the 3 band RGB images is an effective way of improving the building footprint extraction with deep learning models. By using the digital surface model, the IOU results of U-Net and MA-FCN models in building footprint extraction increased 7.46% and 5.7% in satellite dataset and 3.61% and 3.34% in aerial dataset, respectively. U-Net and MA-FCN are more precise in building boundaries since they concatenate feature maps of encoder and decoder parts in producing the final segmentation maps. Mask-RCNN is stable to overfitting because of using ResNet in its architecture.

Key words : *Deep Learning, Buildings, Digital Surface Model, Satellite Imagery, U-Net.*

Correspondence Address*: Department of Geomatics Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran.*
Tel*: : +989146515391*
Email : *n.ahmadian99@ms.tabrizu.ac.ir*