

ارائه یک شبکه عصبی عمیق جدید برای استخراج چند مقیاسی ساختمان‌ها از تصاویر سنجش‌ازدوری

محمد پارسائی^۱، سعید نیازمردی^{۲*}، علی اسماعیلی^۳

^۱ دانش آموخته کارشناسی ارشد سنجش از دور - گروه مهندسی نقشه‌برداری - دانشکده عمران و نقشه‌برداری -

دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته

mohammadparsaei96@gmail.com

^۲ استادیار دانشکده عمران و نقشه‌برداری - گروه مهندسی نقشه‌برداری - دانشگاه تحصیلات تکمیلی صنعتی و

فناوری پیشرفته

s.niazmardi@kgut.ac.ir

^۳ استادیار دانشکده عمران و نقشه‌برداری - گروه مهندسی نقشه‌برداری - دانشگاه تحصیلات تکمیلی صنعتی و فناوری

پیشرفته کرمان

aliesmaeily@hotmail.com

(تاریخ دریافت: شهریور ۱۴۰۱، تاریخ تصویب: مرداد ۱۴۰۲)

چکیده

امروزه استخراج ساختمان یکی از مهم‌ترین نیازهای برنامه‌ریزی برای محیط‌های شهری و روستایی است. استخراج ساختمان‌ها از تصاویر سنجش از دوری به دلیل سرعت بالا و هزینه پائین تهیه و پردازش این نوع از داده به مهم‌ترین داده مورد استفاده در این حوزه تبدیل شده‌است. از بین روش‌های زیادی که برای استخراج ساختمان وجود دارد، شبکه‌های عصبی عمیق بیشتر مورد توجه محققین قرار گرفته‌اند. اما این شبکه‌ها دارای ضعف‌هایی نیز هستند. اکثر مدل‌های موجود، در استخراج چند مقیاسی ساختمان عملکرد ضعیفی دارند. بدین معنا که شبکه‌ها توانایی استخراج ساختمان‌هایی با اندازه‌های مختلف را ندارند. از اینرو، در این تحقیق، مدلی برای استخراج چند مقیاسی ساختمان طراحی و ارائه شده‌است. مدل پیشنهادی، چالش استخراج چند مقیاسی ساختمان را با استفاده از یک ماژول بهبود نتیجه رفع کرده‌است. ماژول طراحی شده با استفاده از پیچش گسترش یافته زمینه دریافت اطلاعات را افزایش می‌دهد که باعث می‌شود ناپیوستگی در خروجی ساختمان‌های بزرگ پدیدار نشود. استخراج صحیح ساختمان‌های بزرگ توسط ماژول بهبود و ساختمان‌های کوچک و متوسط توسط بدنه اصلی، مدل پیشنهادی را به یک مدل کارآمد تبدیل کرده‌است. برای بررسی عملکرد مدل پیشنهادی، از دو مجموعه داده Massachusetts و WHU استفاده شده‌است. با آزمایش این دو مجموعه، مدل پیشنهادی با کسب مقادیر ۰/۶۴۹۵ و ۰/۸۵۷۲ برای معیار IoU بهترین عملکرد را در مقایسه با مدل‌های FCN، USPP، U-Net و DeepLabV3+ نشان داد. بررسی عملکرد ماژول بهبود نشان داد که پس از اضافه کردن این ماژول، مدل پیشنهادی در معیار IoU به میزان ۰/۱۰۷۷ عملکرد بهتری نسبت به مدل بدون ماژول نشان می‌دهد. نتایج همچنین نشان داد که مدل پیشنهادی در استخراج ساختمان‌های بزرگ، برتری قابل توجهی نسبت به سایر مدل‌ها دارد که نشان از عملکرد موفق ماژول طراحی شده‌است.

واژگان کلیدی: استخراج ساختمان، تصاویر سنجش از دوری، یادگیری عمیق، پردازش چند مقیاسی

۱- مقدمه

ساختمان‌ها یکی از مهم‌ترین عوارض ساخته دست بشر هستند که به عنوان عنصر پایه در ساختار شهر و روستا از اهمیت ویژه‌ای برخوردارند. اطلاعات کمی و کیفی به دست آمده از ساختمان‌ها منابع بسیار ارزشمندی برای برنامه‌ریزی و تصمیم‌گیری‌های آینده هستند. این اطلاعات کاربردهایی نظیر تخمین جمعیت ساکن، کشف ساخت و سازها و تخریبات غیرمجاز، پی بردن به میزان و جهت گسترش مناطق مسکونی، تخمین میزان انرژی مصرفی و برنامه‌ریزی برای ارائه خدمات به مناطق مختلف را دارد [۱-۳]. علاوه بر این، کاربردهایی مثل مدیریت بحران در زمان حوادثی مانند آتش سوزی، سیل و زلزله نیازمند اطلاعات بروز از ساختمان‌ها هستند [۴، ۵].

در سال‌های اخیر، با توجه به افزایش دسترسی به تصاویر ماهواره‌ای با قدرت تفکیک بالا، روش‌های سنجش از دوری به یکی از سریع‌ترین و کم هزینه‌ترین روش‌های تشخیص ساختمان‌ها تبدیل شده‌اند [۶]. فرآیند استخراج ساختمان از تصاویر سنجش از دوری شامل دو مرحله است. در مرحله اول وجود ساختمان با توجه به ویژگی‌های تشخیص داده می‌شود و در مرحله بعد شکل ساختمان با بالاترین درصد تشابه با واقعیت، بازسازی می‌شود [۷]. پیچیدگی‌های متعددی در زمینه استخراج ساختمان وجود دارد. از مهم‌ترین چالش‌های موجود می‌توان به اندازه متفاوت ساختمان‌ها، وجود سایه ساختمان‌ها و عوارض تاثیرگذار مانند درختان در کنار ساختمان اشاره کرد. علاوه بر این، در برخی موارد شباهت طیفی بام ساختمان با عوارضی مانند آسفالت معابر می‌تواند نتیجه را تحت تأثیر قرار دهد [۷، ۸].

روش‌های زیادی برای استخراج ساختمان از تصاویر سنجش از دوری بر اساس ویژگی‌های ساختمان وجود دارد که از مهم‌ترین آن‌ها می‌توان به روش‌های مبتنی بر منحنی فعال [۷، ۹]، استخراج بر اساس کشف لبه [۱۰]، روش مبتنی بر طبقه‌بندی و قطعه‌بندی [۱۱، ۱۲]، روش شی گرای [۱۳]، تبدیل هاف^۱ [۱۴] و روش‌های مبتنی بر عملگرهای مورفولوژی [۱۵] اشاره کرد.

شبکه‌های عمیق یکی از بروزترین روش‌های استخراج ساختمان‌ها هستند، که مورد توجه بسیاری از محققین قرار گرفته‌اند و به دلیل نتایج بسیار خوب به یکی از متداول‌ترین روش‌ها برای استخراج ساختمان از تصاویر تبدیل شده‌اند [۸]. برای استخراج ساختمان، در سال ۲۰۱۹ یک مدل یادگیری عمیق سبک به نام USPP^۲ ارائه شده‌است که ماژول SPP^۳ را با یک ساختار رمزگذار-رمزگشا^۴ ادغام می‌کند [۱۶]. محققان از ماژول SPP به عنوان پلی بین رمزگذار و رمزگشا استفاده کردند. در این مدل، ساختار رمزگذار-رمزگشا برای بازیابی اطلاعات و ماژول SPP برای جمع‌آوری اطلاعات چند مقیاسی مورد استفاده قرار گرفته‌است. در سال ۲۰۱۹ محققان با بیان اینکه که در شبکه‌های پیچشی^۵ استفاده مجدد از ویژگی‌های قبلی به طور مستقیم و بدون پردازش می‌تواند عملکرد شبکه را کاهش دهد، یک شبکه جدید به نام RFA-UNet^۶ ارائه دادند که وزن دهی مجدد را بر اساس میزان اهمیت ویژگی اعمال می‌کند. برای رسیدن به این هدف، از یک ماژول جدید به نام RFA^۷ استفاده شده‌است تا وزن دهی مجدد را انجام دهد. ماژول پیشنهادی بر ویژگی‌های معنی‌دار تأکید می‌کند و ویژگی‌های ناچیز را کم اهمیت در نظر می‌گیرد [۱۷]. برای رفع ضعف مدل‌های موجود در استخراج یکپارچه ساختمان و تداوم فضایی آنها، محققان در سال ۲۰۱۹ مدل جدیدی به نام EU-Net^۸ ارائه کردند [۱۸]. EU-Net یک شبکه‌ی کاملاً کانولوشنی هست که از سه قسمت رمزگذار، ماژول DSPP^۹ و رمزگشا تشکیل می‌شود. این شبکه با استفاده از ساختار DSPP می‌تواند ویژگی‌های چند مقیاسی را به طور همزمان استخراج کند و ساختمان‌هایی از مقیاس‌های مختلف را بهتر استخراج کند. نتایج این شبکه نشان داد که EU-Net یک مدل کارآمد با میزان یادگیری و سرعت پردازش زیاد است که توانایی بالایی در استخراج ساختمان‌های با اندازه بزرگ دارد.

با وجود پیشرفت‌های اخیر در این حوزه و قابلیت بالای مدل‌های ارائه شده، به دلیل پیچیدگی مبحث استخراج ساختمان، هنوز ضعف‌هایی در این زمینه مشهود است. مقیاس و اندازه ساختمان بزرگترین مسئله هندسی است که بر عملکرد استخراج ساختمان با استفاده از الگوریتم‌های

۶ Residual Feature Map Attention- U networks

۷ Residual feature attention

۸ Efficient U-Net

۹ Dense Spatial Pyramid Pooling

۱ Hough Transform

۲ U-shape Structure With SPP

۳ Spatial Pyramid Pooling

۴ Encoder-Decoder

۵ Convolutional networks

یادگیری عمیق تأثیر می‌گذارد [۱۹]. نتایج نشان داده‌است که هر کدام از مدل‌های ارائه شده، یا در استخراج ساختمان‌های بزرگ توانایی بالایی دارند یا در استخراج ساختمان‌های کوچک. از مهم‌ترین ضعف‌های مدل‌های موجود، ناتوانی در استخراج چند مقیاسی ساختمان و عدم پیوستگی در خروجی ساختمان‌های بزرگ است.

در سال‌های اخیر مدل‌هایی برای استخراج چند مقیاسی ساختمان طراحی شده‌اند که به شیوه‌های متفاوت به دنبال حل این چالش بوده‌اند. یکی از پر استفاده‌ترین روش‌ها، استفاده از بلوک SPP به عنوان پل اتصال در میان رمزگذار و رمزگشای مدل است. اساس کار این بلوک استخراج ویژگی‌های متفاوت از شاخه‌های مختلف خود است. امروزه اشکال مختلف این بلوک در مدل‌هایی مانند DeepLab [۲۰]، USPP [۱۶] و WASPnet [۲۱] برای استخراج چند مقیاسی عوارض استفاده می‌شود. روش دیگری که محققان برای استخراج چند مقیاسی ساختمان استفاده می‌کنند، استخراج ویژگی در مسیرهای مختلف است. در این روش مدل دارای چند مسیر موازی است که هر کدام در مقیاسی متفاوت ویژگی‌های موجود را استخراج می‌کنند و در پایان از یک ماژول توجه^۱ برای ادغام بهینه ویژگی‌های استخراج شده استفاده می‌شود [۲۲، ۲۳]. روش سوم استخراج چند مقیاسی مربوط به ساختار رمزگشای مدل است. در این روش ویژگی‌های موجود از چندین سطح متفاوت بخش رمزگشا به طور مستقیم به اندازه نهایی upsample می‌شوند. ویژگی‌های به دست آمده از سطوح مختلف دارای ویژگی‌های متفاوتی هستند که با ادغام آن‌ها می‌توان ساختمان‌هایی با اندازه متفاوت را از تصاویر استخراج کرد [۲۴].

با توجه به موارد ذکر شده، طراحی مدلی کارآمد که قابلیت استخراج ساختمان با اندازه‌های متفاوت را داشته باشد از اهمیت بالایی برخوردار است. هدف این تحقیق ارائه شبکه‌ای برای استخراج چندمقیاسی ساختمان‌ها است. مدل پیشنهادی این چالش را به وسیله استفاده از یک ماژول بهبود نتیجه بر طرف می‌کند. اساس کار این ماژول، دریافت پیش-بینی اولیه بدنه اصلی مدل و بهبود نواقص مربوط به آن است. در ادامه این مقاله، روش پیشنهادی برای استخراج چند مقیاسی ساختمان‌ها در بخش دوم معرفی می‌شود. داده‌های استفاده شده و همچنین پیاده‌سازی‌های انجام شده در بخش

سوم ارائه شده‌اند. نتایج حاصل از پیاده‌سازی‌ها در بخش چهارم و در نهایت نتیجه‌گیری‌ها در پنجمین بخش آورده شده‌است.

۲- مبانی نظری روش پیشنهادی

در این تحقیق روشی متفاوت از موارد ذکر شده برای استخراج چند مقیاسی ساختمان ارائه شده‌است. روش پیشنهادی این چالش را به وسیله استفاده از یک ماژول بهبود نتیجه بر طرف می‌کند. این ماژول در انتهای مدل قرار می‌گیرد و ورودی آن خروجی تک کاناله بدنه اصلی مدل است. معمولاً در خروجی اولیه، ساختمان‌های بزرگ به صورت ناقص و ناپیوسته پدیدار می‌شوند؛ که این ضعف توسط ساختار ماژول بهبود طراحی شده برطرف می‌شود. جزئیات ماژول بهبود و مدل پیشنهادی در ادامه شرح داده می‌شود.

ماژول پیشنهاد شده در این تحقیق با استفاده از اطلاعات پیکسل‌های اطراف هسته پیچش، ناپیوستگی‌های موجود در خروجی اولیه را بر طرف می‌کند. توجه به پیکسل‌های بیشتر می‌تواند باعث بهبود بیشتر نتیجه شود. برای این کار از چندین لایه پیچشی گسترش یافته به صورت متوالی استفاده شده‌است. استفاده از شکل گسترش یافته لایه پیچش این امکان را فراهم می‌کند که بدون اضافه کردن پارامتر بیشتر، میدان دریافت اطلاعات برای این لایه افزایش پیدا کند. میدان دریافت گسترده لایه پیچش باعث می‌شود، این لایه از رابطه همسایگی پیکسل‌ها به صورت گسترده‌تری استفاده کند و از وجود رابطه‌ی آن‌ها بهره‌برد [۲۵]. این عمل با استفاده از پیچش‌های بزرگتر نیز امکان پذیر است با این تفاوت که پیچش با ابعاد بزرگتر با افزایش چشمگیر تعداد پارامتر همراه است.

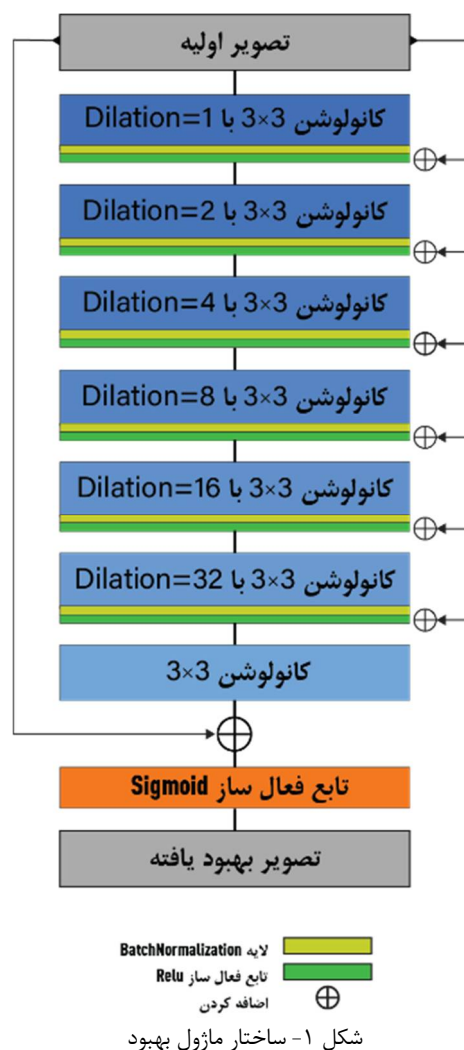
ماژول طراحی شده در این تحقیق از ۶ لایه پیچشی 3×3 با نرخ‌های گستردگی^۲ ۱، ۲، ۴، ۸، ۱۶ و ۳۲ تشکیل شده‌است که هر لایه دارای ۶۴ فیلتر است. در این ماژول پس از هر لایه پیچشی، لایه Batch Normalization و تابع فعال‌ساز ReLU^۳ قرار دارد. برای اینکه طی عملیات کانولوشن ویژگی‌های تصویر ورودی را از دست ندهیم، تصویر ورودی در هر مرحله به لایه‌های پیش‌رو اضافه می‌شود. در مسیر این ماژول نرخ گستردگی لایه کانولوشن به صورت تصاعدی

۳ Rectified Linear Unit

۱ Attention Modules

۲ Dilation Rates

افزایش پیدا می‌کند؛ تا جایی که درصد بالایی از یک ساختمان بزرگ را در خود جای دهد. در پایان یک کانولوشن 3×3 ساده با یک فیلتر اعمال می‌شود و برای خروجی نهایی تابع فعال‌ساز Sigmoid مورد استفاده قرار می‌گیرد. ساختار کلی ماژول پیشنهادی در شکل (۱) نمایش داده شده‌است.



برای طراحی بدنه اصلی مدل، ترکیبی از شبکه‌های Res-net و U-Net مورد استفاده قرار گرفته‌است. مدل پیشنهادی یک مدل رمزگذار-رمزگشا متقارن با ساختار U شکل است که هر سطح از بخش رمزگذار به وسیله یک مسیر اتصال مستقیم^۱ به بخش رمزگشا متصل می‌شود. در طراحی مدل‌های عمیق همواره با محدودیت افزایش تعداد لایه مواجه هستیم. یکی از مشکلاتی که با افزایش لایه‌ها در مدل به

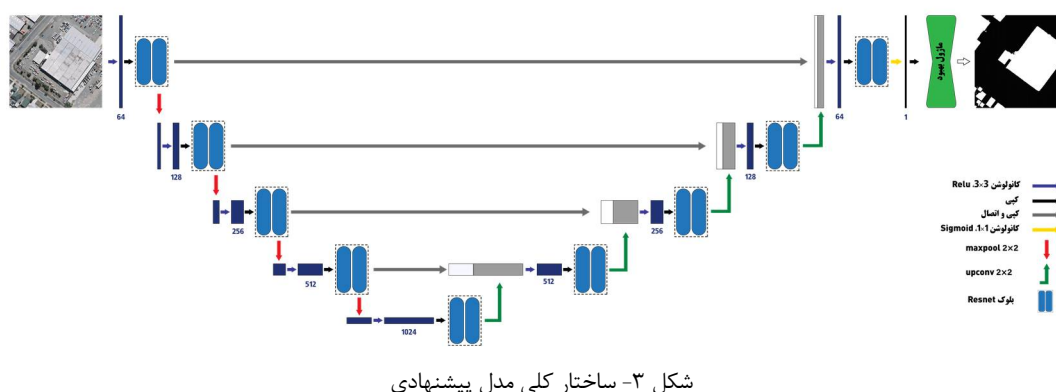
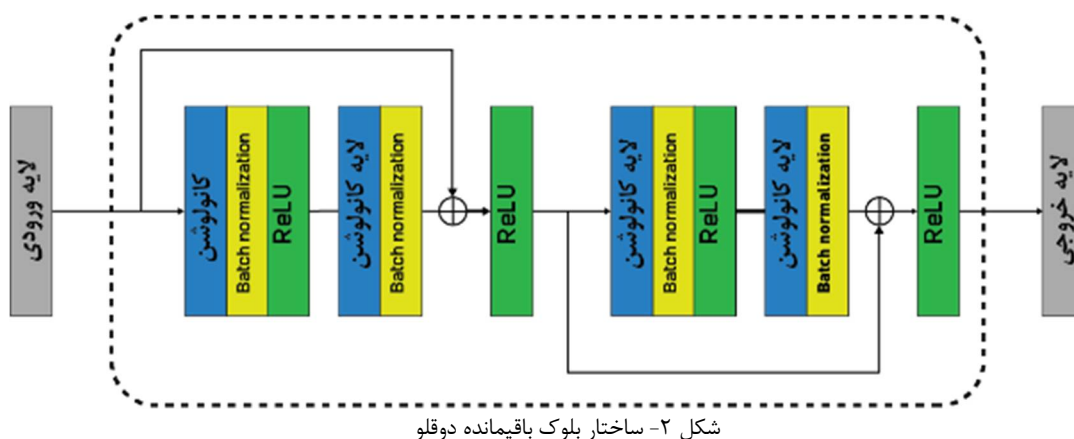
وجود می‌آید، محو شدگی گرادیان و در نتیجه آن عدم بروزسانی وزن‌های مدل و اختلال در یادگیری است. مدل پیشنهادی نیز با اضافه شدن لایه‌های بیشتر در نتیجه اتصال ماژول بهبود دچار این مشکل می‌شود. در مدل پیشنهادی این مشکل به دلیل استفاده مکرر از تابع فعال‌ساز Sigmoid تشدید می‌شود. برای حل این مشکل در مدل پیشنهادی، بلوک‌های باقیمانده^۲ جایگزین لایه کانولوشن معمولی شده‌اند. عنصر اصلی بلوک‌های باقیمانده، اتصال مستقیمی است که از لایه ورودی شروع می‌شود و لایه‌های میانی را رد می‌کند و به آخرین تابع فعال‌ساز بلوک متصل می‌شود. استفاده از این مسیر مستقیم باعث می‌شود تابع فعال‌ساز بر روی این اتصال اعمال نشود که در زمان پس انتشار مقدار گرادیان را خیلی کوچک و به صفر نزدیک کند. در واقع در زمان پس انتشار، گرادیان از این مسیر عبور می‌کند و تا رسیدن به اولین لایه، مقدار آن خیلی کوچک نمی‌شود؛ که باعث می‌شود وزن‌ها به خوبی بروزسانی شوند.

در مدل پیشنهادی به دلیل استفاده از بلوک‌های باقیمانده امکان استفاده از لایه‌های بیشتر امکان‌پذیر است. برای استخراج بهتر ویژگی‌های موجود، در این مدل از بلوک‌های باقیمانده به صورت دوقلو در کنار یکدیگر استفاده شده‌است. هر بخش این بلوک شامل دو لایه پیچشی است که پس از آن، لایه Batch normalization اعمال می‌شود. اتصال مستقیم این بلوک پس از دو لایه پیچشی وارد تابع فعال‌ساز می‌شود که همگی از نوع ReLU هستند. ساختار بلوک مورد نظر را در شکل (۲) قابل مشاهده است.

مدل پیشنهادی از چهار لایه پولینگ در طی قسمت رمزگذار استفاده می‌کند و با استفاده از چهار لایه افزایش بعد به اندازه اولیه می‌رسد و این دو بخش توسط اتصال مستقیم به یکدیگر متصل می‌شوند. در این مدل قبل از بلوک‌های دوقلو، یک پیچش 3×3 متناسب بر روی لایه قبل اعمال می‌شود و سپس خروجی آن وارد بلوک‌های دوقلو می‌شود. این لایه برای همسان‌سازی عمق داده‌ها است تا در استفاده از اتصال مستقیم بلوک باقیمانده مشکلی ایجاد نشود. مدل پیشنهادی از پنج سطح تشکیل شده‌است که لایه‌های پیچشی آن به ترتیب در سطح ۱ تا ۵ از ۶۴، ۱۲۸، ۲۵۶، ۵۱۲ و ۱۰۲۴ فیلتر استفاده می‌کنند. ساختار کلی مدل پیشنهادی در شکل (۳) نمایش داده شده‌است.

^۲ Residual Block

^۱ Skip Connection



۳- داده‌ها و پیاده‌سازی

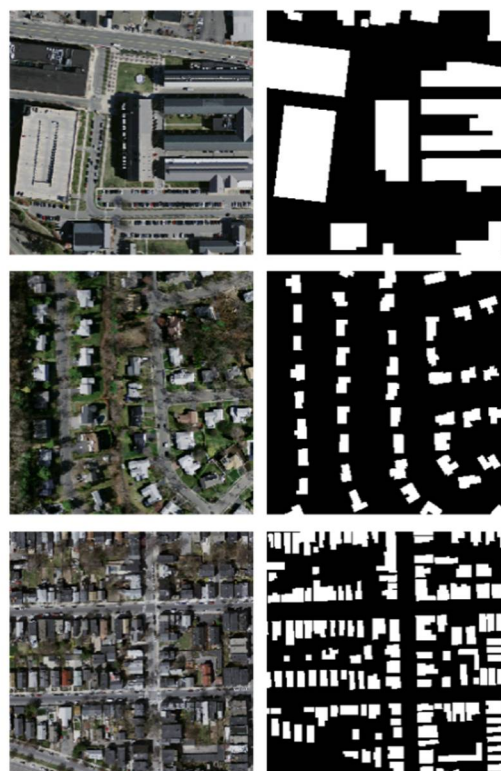
۳-۱- داده‌ها

می‌شود امکان استفاده مستقیم از آن‌ها برای آموزش وجود نداشته باشد. به دلیل محدودیت حافظه GPU، این تصاویر بزرگ به تکه‌های کوچک‌تر برش داده شده‌اند، تا امکان استفاده از آن‌ها فراهم شود. این برش به وسیله حرکت یک کرنل با ابعاد 256×256 پیکسل بر روی تصویر انجام شد. تکه‌های حاصل به اندازه هشت پیکسل با یکدیگر اشتراک دارند. در این فرایند تصاویر اصلی این مجموعه به ۴۹ تصویر کوچک‌تر تقسیم می‌شوند.

طی این مرحله تعداد زیادی از تصاویر تولید شده شامل مناطق غیر ساختمانی بودند؛ که تعداد بالای آن‌ها آموزش مدل را دچار مشکل می‌کند. با توجه به این مورد و ناقص بودن بخشی از تصاویر اصلی، تعدادی زیادی از تصاویر جدید در قسمت آماده‌سازی مجموعه حذف شدند. در نهایت تعداد ۳۵۸۲ تصویر با ابعاد 256×256 پیکسل برای قسمت آموزش تهیه شد. همچنین تعداد ۴۹۰ تصویر برای آزمایش و ۱۹۶

در این تحقیق برای ارزیابی مدل پیشنهادی از دو مجموعه داده استفاده شده‌است. مجموعه اول مجموعه داده ساختمان Massachusetts است در سال ۲۰۱۳ تهیه شده‌است [۲۶]. این مجموعه شامل ۱۵۱ تصویر هوایی RGB از منطقه بوستون در ایالت ماساچوست کشور آمریکا است. تصاویر موجود دارای ابعاد 1500×1500 پیکسل هستند و هر تصویر $2/25$ کیلومتر مربع را بر روی زمین پوشش می‌دهد. قدرت تفکیک مکانی تصاویر این مجموعه ۱ متر است که نسبتاً قدرت تفکیک کمی برای استخراج ساختمان است. این تصاویر هوایی به طور تصادفی به سه بخش آموزش، اعتبارسنجی و آزمایش تقسیم شده‌اند. سهم هر کدام از بخش‌ها به صورت ۱۳۷ تصویر برای قسمت آموزش، ۴ تصویر برای اعتبارسنجی و ۱۰ تصویر نیز برای قسمت آزمایش است. تهیه‌کنندگان این مجموعه، برای تهیه برچسب معنایی یا ماسک مورد نظر از اطلاعات پروژه OpenStreetMap استفاده کردند. ابعاد بزرگ تصاویر این مجموعه داده باعث

تصویر برای بخش اعتبارسنجی به دست آمد. نمونه‌هایی از تصاویر این مجموعه در شکل (۴) قابل مشاهده است.



شکل ۴- نمونه‌ای از تصاویر مجموعه Massachusetts

مجموعه داده دوم استفاده شده در این مقاله، مجموعه داده ساختمان^۱ WHU است که در سال ۲۰۱۸ تهیه و ارائه شد [۱۹]. تصاویر موجود در این مجموعه شامل تصاویر هوایی و ماهواره‌ای است. تصاویر اولیه این مجموعه طی یک مرحله تصویربرداری از ۴۵۰ کیلومتر مربع از شهر Christchurch کشور نیوزلند به دست آمد. تصاویر اولیه دارای قدرت تفکیک مکانی ۰/۰۷۵ متر هستند، اما محققان با بیان اینکه تصاویر با قدرت تفکیک مکانی کمتر از ۰/۳ متر باعث بهبود عملکرد شبکه‌های FCN^۲ نمی‌شوند، تصاویر اولیه را به این میزان نمونه‌برداری کردند. برای تهیه این مجموعه داده، بخشی از محدوده تصویربرداری شده شامل ۱۸۷۰۰۰ ساختمان انتخاب شد و در نهایت ۸۱۸۹ تصویر با اندازه ۵۱۲×۵۱۲ پیکسل از این منطقه تهیه شد. برای ایجاد ماسک مربوط به تصاویر، محققان اطلاعات اولیه را از سازمان اطلاعات زمین کشور نیوزلند دریافت کردند و به دلیل تناقضات موجود، اطلاعات دریافتی در نرم افزار ArcGIS ویرایش شد. برای آموزش، تصاویر این مجموعه به سه بخش آموزش،

اعتبارسنجی و آزمایش تقسیم شده‌اند که ۴۷۳۶ تصویر برای بخش آموزش، ۱۰۳۶ تصویر برای اعتبارسنجی و ۲۴۱۶ تصویر برای آزمایش مدل در نظر گرفته شده‌است. چند نمونه از تصاویر این مجموعه به همراه ماسک مورد نظر آن در شکل (۵) قابل مشاهده است.



شکل ۵- نمونه‌ای از تصاویر مجموعه WHU

تصاویر موجود در این دو مجموعه مربوط به مناطق روستایی و شهری هستند که ساختمان‌هایی با رنگ و اندازه متفاوت همراه با کاربری‌های مسکونی، تجاری، فرهنگی، صنعتی و... را شامل می‌شوند.

۳-۲- پیاده‌سازی

برای بررسی توانایی مدل پیشنهادی نیاز است تا این مدل با چندین مدل شاخص این حوزه مقایسه شود. برای این کار مدل‌های U-Net [۲۷]، DeepLabV3+ [۲۰] و USPP [۱۶] مورد استفاده قرار گرفتند. همچنین دو مدل FCN^۲ و U-Net با قسمت رمزگذار VGG16 در کنار سایر مدل‌ها به عنوان شاخص ارزیابی استفاده شدند. در این تحقیق با هدف مدیریت منابع پردازشی، در نظر داریم از تعداد دوره محدودی

^۲ Fully convolutional network

^۱ Wuhan University Buildings data set

در بخش بعد، تمام ۴۹۰ تصویر قسمت آزمایش این مجموعه داده مورد ارزیابی قرار گرفتند که نتیجه در جدول (۱) ارائه شده‌است.

جدول ۱- ارزیابی مدل‌های انتخابی در تمامی تصاویر آزمایش مجموعه‌داده Massachusetts

مدل	معیار ارزیابی		
	Precision	Recall	F1-Score
FCN	۰.۷۴۴۱	۰.۷۰۲۱	۰.۷۱۴۵
U-Net	۰.۸۲۵۶	۰.۷۰۲۳	۰.۷۵۰۹
DeepLabV3+	۰.۸۳۳۲	۰.۶۷۹۴	۰.۷۴۱۸
USPP	۰.۸۰۱	۰.۷۲۳۳	۰.۷۵۲۲
VGG-UNET	۰.۸۱۱۵	۰.۶۸۴۳	۰.۷۳۰۷
مدل پیشنهادی	۰.۷۹۱۹	۰.۷۸۷۸	۰.۷۸۲۳

نتایج حاصل از جدول (۱) نشان داد تمامی مدل‌ها بر روی مجموعه‌داده Massachusetts عملکرد ضعیفی دارند. نتایج نسبتاً ضعیف این مجموعه دو دلیل اصلی دارد. دلیل اول پیچیدگی کلی صحنه‌های موجود است. تصاویر این مجموعه دارای پیچیدگی در ساختار ساختمان‌ها و وجود درختان و سایه ساختمان‌های بلند هستند. دلیل دوم قدرت تفکیک مکانی پایین تصاویر این مجموعه است که نتایج را تحت تأثیر قرار داده‌است. با این حال مدل پیشنهادی توانست بهترین عملکرد را برای تشخیص و استخراج ساختمان نشان دهد.

جدول (۱) نشان داد مدل پیشنهادی بیشترین مقادیر را در معیارهای IoU و Recall کسب کرده است، که نشان از آن است که این مدل نسبت به سایر مدل‌ها درصد بیشتری از انواع ساختمان‌ها را به طور صحیح استخراج کرده‌است.

در ادامه پنج تصویر با خصوصیات متفاوت از این مجموعه انتخاب شد تا مقایسه‌ای با جزئیات بیشتر انجام شود. شکل (۷) نمایش بصری ساختمان‌های تشخیص داده شده توسط مدل‌های مختلف را نشان می‌دهد. نمونه‌های از ساختمان‌های بزرگ برای مقایسه بهتر عملکرد مدل‌ها در این شکل با مربعی قرمز رنگ مشخص شده‌اند. برای تفسیر نتایج به دست آمده نیاز است که خروجی مدل‌ها بر اساس معیارهای ذکر شده ارزیابی شوند. جدول (۲) مقایسه خروجی شکل (۷) بر اساس ۴ معیار تعریف شده‌است.

(۱۰ دوره) برای آموزش مدل‌ها استفاده کنیم و عملکرد مدل‌های ذکر شده را با میزان آموزش کم بسنجیم. با توجه به این مورد مشخصات آموزش مدل‌ها با توجه به مجموعه داده انتخابی به صورت مشابه انتخاب شدند. برای آموزش تمامی مدل‌ها از تابع آنتروپی متقاطع باینری^۱ به عنوان تابع هزینه استفاده شد. با توجه به ابعاد تصاویر موجود، اندازه دسته^۲ برای مجموعه Massachusetts میزان هشت تصویر و برای مجموعه WHU میزان یک تصویر در نظر گرفته شد. به عنوان الگوریتم بهینه‌سازی^۳، برای تمام مدل‌ها الگوریتم Adam^۴ با نرخ یادگیری^۵ ۰/۰۰۰۱ مورد استفاده قرار گرفت [۲۸]. برای پیاده‌سازی بر روی مجموعه داده‌های ذکر شده از داده افزایی^۶ چرخش و قرینه‌سازی با احتمال ۵۰ درصد در هر دور استفاده شد. مدل‌های مورد نظر از پارامترهای از قبل آموزش دیده استفاده نمی‌کنند و تمامی مدل‌ها به صورت طبیعی و بدون مقداردهی اولیه آموزش دیدند. برای مقایسه نتایج به دست آمده نیاز است که خروجی مدل‌ها بر اساس معیارهای مناسب ارزیابی شوند. برای این کار از چهار معیار Precision، Recall، F1-Score و IoU استفاده شد. علاوه بر ارزیابی تمام تصاویر بخش آزمایش، دقت برای گروهی از تصاویر به صورت مجزا محاسبه شده‌است. برای بررسی بیشتر، در این مرحله تصاویری انتخاب شدند که دارای پیچیدگی در پس‌زمینه باشند و یا شامل ساختمان‌هایی باشند که اندازه، رنگ و بافت متفاوتی دارند.

علاوه بر پیاده‌سازی‌های انجام شده برای مقایسه روش پیشنهادی با سایر روش‌های موجود، با هدف ارزیابی کارایی ماژول پیشنهادی، دقت مدل در صورت حذف این ماژول نیز محاسبه و گزارش شده‌است.

۴- نتایج

۴-۱- مقایسه عملکرد مدل پیشنهادی با سایر مدل‌ها

پس از آموزش مدل‌ها، در ابتدا عملکرد مدل پیشنهادی بر مجموعه Massachusetts بررسی شد. شکل (۶) نمایش بصری خروجی مدل پیشنهادی نسبت به تعدادی از تصاویر این مجموعه است.

۴ Adaptable Moment estimation

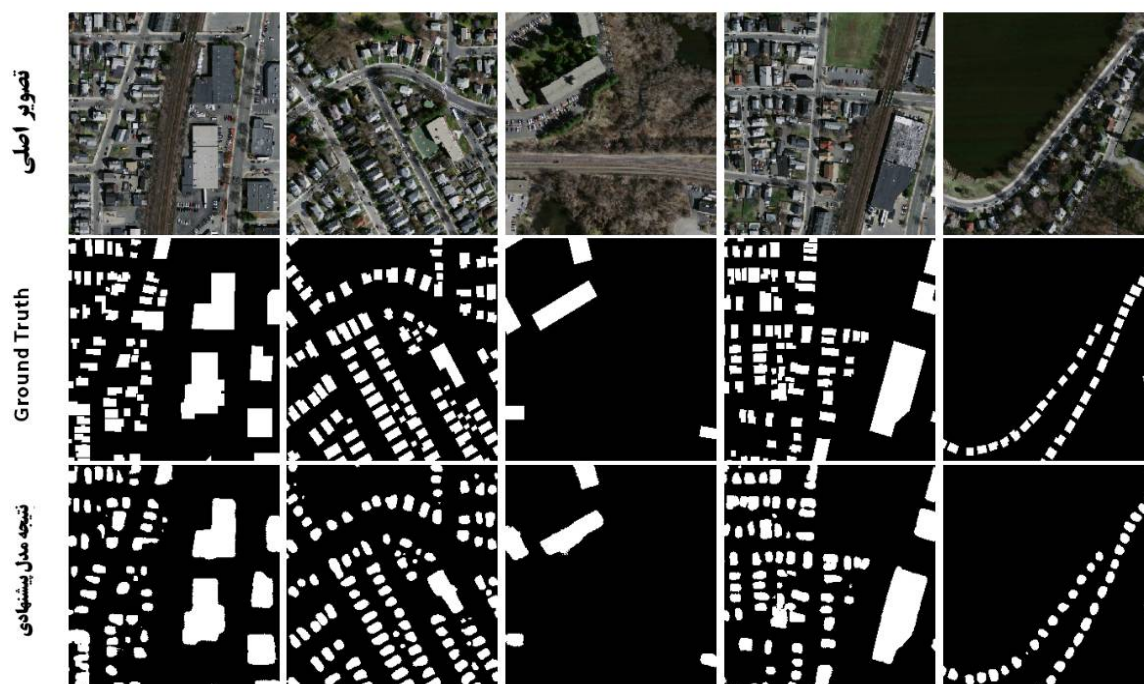
۵ Learning Rate

۶ Data Augmentation

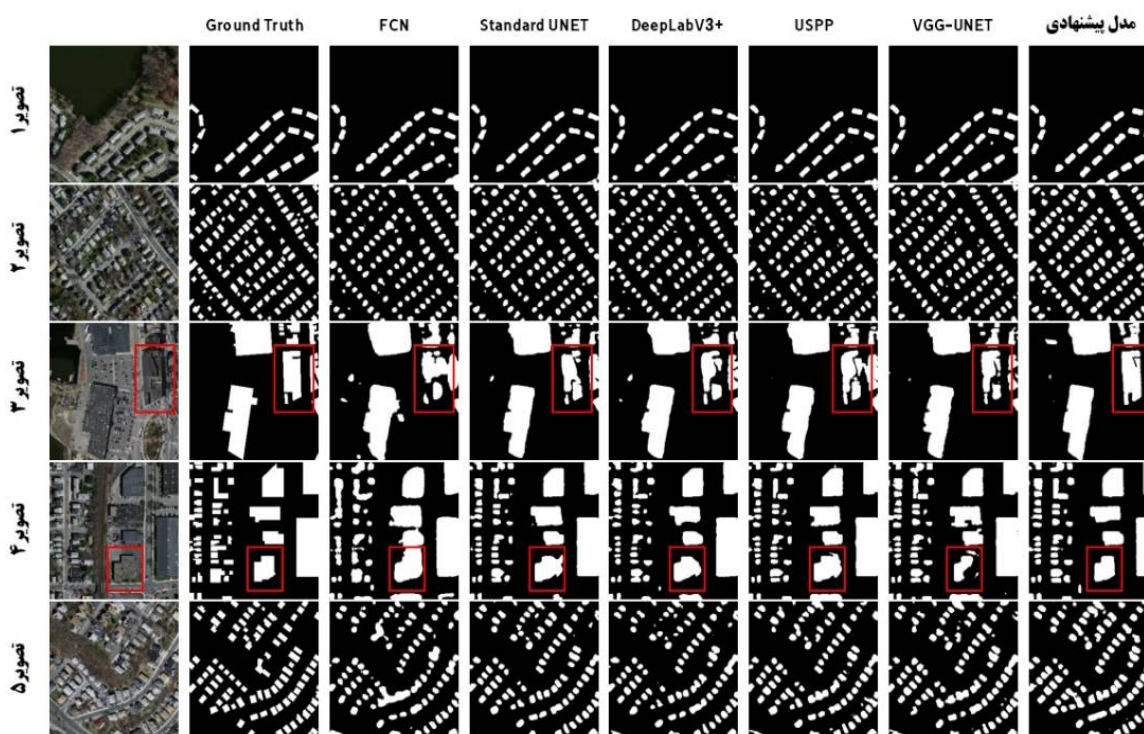
۱ Binary Cross Entropy

۲ Batch Size

۳ Optimization Algorithm



شکل ۶- نتایج تشخیص ساختمان در تصاویر Massachusetts با استفاده از مدل پیشنهادی



شکل ۷- نتایج تشخیص ساختمان در تصاویر مجموعه Massachusetts با استفاده از مدل‌های مختلف، نمونه‌ای از ساختمان‌هایی که مدل‌ها عملکرد متفاوتی نشان داده‌اند با مربع قرمز رنگ مشخص شده‌است.

جدول ۲- ارزیابی خروجی مدل‌ها در تصاویر انتخابی مجموعه Massachusetts

تصویر	معیار	FCN	U-Net	DeepLabV3+	USPP	VGG -UNet	مدل پیشنهادی
تصویر ۱	Precision	۰.۷۸۳۶	۰.۸۱۸۵	۰.۸۵۱۱	۰.۸۳۳۹	۰.۷۸۱۳	۰.۷۷۵۱
	Recall	۰.۷۶۹۱	۰.۷۱۲۶	۰.۶۹۵	۰.۷۳۳۱	۰.۷۲۷	۰.۷۵۰۵
	F1-Score	۰.۷۷۶۳	۰.۷۶۱۹	۰.۷۶۵۲	۰.۷۸۰۲	۰.۷۵۳۲	۰.۷۶۲۶
	IoU	۰.۶۳۴۴	۰.۶۱۵۳	۰.۶۱۹۶	۰.۶۳۹۷	۰.۶۰۴۱	۰.۶۱۶۳
تصویر ۲	Precision	۰.۸۴۶۵	۰.۸۷۴۵	۰.۸۷۶۲	۰.۸۷۶۷	۰.۸۵۶	۰.۸۱۹۸
	Recall	۰.۶۹۶۴	۰.۷۶۷۳	۰.۷۴	۰.۷۵۴۸	۰.۷۸۴۶	۰.۸۵۱۶
	F1-Score	۰.۷۶۴۲	۰.۸۱۷۴	۰.۸۰۲۴	۰.۸۱۱۲	۰.۸۱۸۸	۰.۸۳۵۴
	IoU	۰.۶۱۸۳	۰.۶۹۱۱	۰.۶۷	۰.۶۸۲۴	۰.۶۹۳۱	۰.۷۱۷۴
تصویر ۳	Precision	۰.۸۱۹۱	۰.۹۰۹۲	۰.۸۹۵۹	۰.۸۸۵۱	۰.۹۲۲۴	۰.۸۷۵۴
	Recall	۰.۸۹۱۱	۰.۸۹۱۲	۰.۸۸۱۸	۰.۹۰۹۱	۰.۸۶۹۹	۰.۹۵۱۵
	F1-Score	۰.۸۵۳۶	۰.۹۰۰۱	۰.۸۸۸۸	۰.۸۹۶۹	۰.۸۹۵۴	۰.۹۱۱۸
	IoU	۰.۷۴۴۶	۰.۸۱۸۴	۰.۷۹۹۸	۰.۸۱۳۱	۰.۸۱۰۶	۰.۸۳۸
تصویر ۴	Precision	۰.۷۷۴	۰.۸۹۳۲	۰.۸۹۷۳	۰.۸۷۴۵	۰.۸۷۳۶	۰.۹۲۱۸
	Recall	۰.۸۸۸۷	۰.۸۸۴۱	۰.۸۷۲۸	۰.۸۸۵	۰.۸۷۱۵	۰.۹۰۱۶
	F1-Score	۰.۸۲۷۴	۰.۸۸۸۶	۰.۸۸۴۹	۰.۸۷۹۷	۰.۸۷۲۵	۰.۹۱۱۶
	IoU	۰.۷۰۵۶	۰.۷۹۹۵	۰.۷۹۳۵	۰.۷۸۵۲	۰.۷۷۳۹	۰.۸۳۷۶

با توجه به نتایج جدول (۲) و خروجی شکل (۷)، استخراج ساختمان‌های کوچک و متوسط توسط تمام مدل‌ها تا حدود زیادی به طور مشابه انجام گرفته‌است. نقطه قوت مدل پیشنهادی در نتایج حاصل از تصاویر ۳ و ۴ نمایان می‌شود. در این تصاویر که شامل ساختمان‌های بزرگ هستند، مدل پیشنهادی بیشترین مقادیر را در معیار IoU نسبت به سایر مدل‌ها کسب کرد. ساختمان‌هایی که با کادر قرمز رنگ مشخص شده‌اند نشان از استخراج بهتر مرزها و پیوستگی بیشتر در خروجی مدل پیشنهادی نسبت به سایر مدل‌های انتخابی است. تصویر ۴ موجود در شکل (۷)، شامل تعداد زیادی ساختمان با اندازه مختلف است که نتایج حاصل از معیارهای ارزیابی نشان از اختلاف زیاد مدل

پیشنهادی با سایرین دارد. این اختلاف نشان از توان بالای مدل پیشنهادی در استخراج چند مقیاسی ساختمان از تصاویر سنجش از دوری است.

بخش بعد مربوط به نتایج مربوط به مجموعه WHU است. با توجه به اینکه تصاویر این مجموعه از لحاظ قدرت تفکیک مکانی، متفاوت از تصاویر مجموعه Massachusetts هستند، نتایج حاصل می‌تواند مقایسه‌ای از جنبه دیگر باشد. همچنین با توجه به کیفیت بالاتر این تصاویر از لحاظ قدرت تفکیک مکانی، نتایج حاصل از اعتبار بیشتری برخوردارند. پس از آموزش طبق مشخصات ذکر شده، ابتدا تمام تصویر قسمت آزمایش این مجموعه داده توسط

مدل‌های موجود مورد ارزیابی قرار گرفتند که نتیجه در جدول (۳) ارائه شده است.

برای نمایش نتایج طبقه‌بندی مدل پیشنهادی تعدادی از تصاویر این مجموعه توسط این مدل مورد آزمایش قرار گرفتند که نتیجه در شکل (۸) قابل مشاهده است.

جدول ۳- ارزیابی مدل‌های انتخابی در تمامی تصاویر آزمایش مجموعه داده WHU

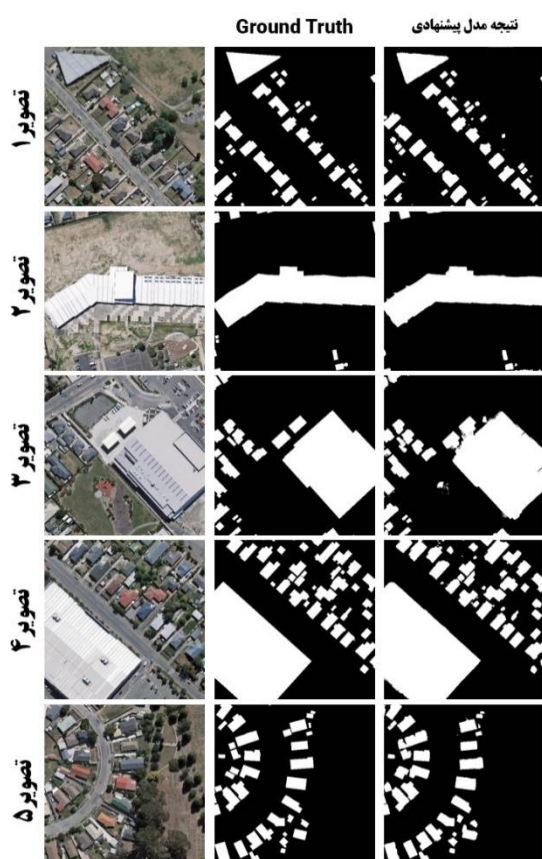
مدل	معیار ارزیابی			
	Precision	Recall	F1-Score	IoU
FCN	۰.۹۰۰۳	۰.۸۷۸۹	۰.۸۸۵۴	۰.۸۰۱
U-Net	۰.۸۸۹	۰.۸۷۹۹	۰.۸۷۹۷	۰.۷۹۷۹
DeepLabV3+	۰.۹۱۳۶	۰.۹۱۸۵	۰.۹۱۳۶	۰.۸۴۶۲
USPP	۰.۹۰۰۶	۰.۹۱۶۹	۰.۹۰۵۵	۰.۸۳۳۲
VGG-UNET	۰.۸۸۱۵	۰.۸۷۹	۰.۸۷۵۱	۰.۷۹۲۳
مدل پیشنهادی	۰.۹۰۴۱	۰.۹۱۹۷	۰.۹۰۸۳	۰.۸۵۷

نتایج جدول (۳) و شکل (۸) نشان از قابلیت بالای مدل پیشنهادی در استخراج ساختمان از تصاویر سنجش از دوری است. در نتایج موجود در جدول (۳)، مدل پیشنهادی به همراه مدل DeepLabV3+ بهترین عملکرد را نسبت به مجموعه WHU نشان دادند.

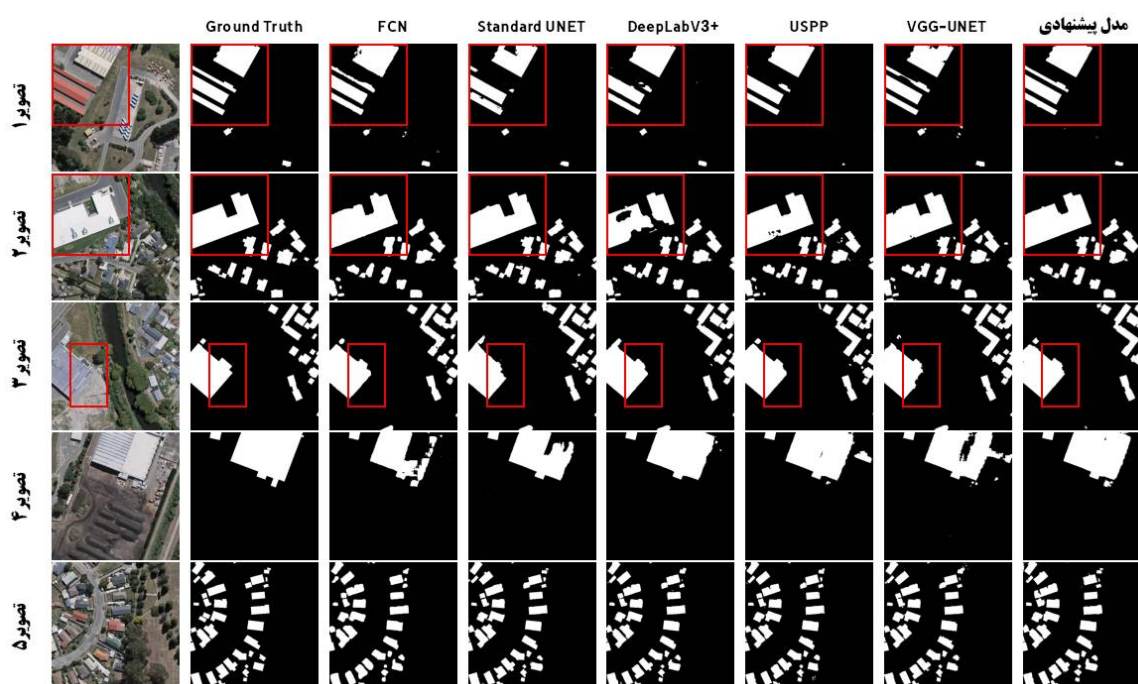
برای مقایسه با جزئیات بیشتر، ۵ تصویر متفاوت از مجموعه WHU انتخاب شد تا پیشبینی مدل‌ها نسبت به شرایط مختلف موجود در محیط بررسی شود. شکل (۹) نمایش بصری خروجی ساختمان‌های تشخیص داده شده توسط مدل‌های مختلف نسبت به تصاویر انتخابی را نشان می‌دهد. نمونه‌هایی از ساختمان‌های بزرگ در این شکل نیز با مربع قرمز نشان داده شده‌اند. برای تفسیر نتایج به دست آمده، خروجی مدل‌های مورد نظر بر اساس معیارهای ارزیابی محاسبه شد که در جدول (۴) قابل مشاهده است.

اشکال (۸) و (۹) تشکیل شده از تصاویری هستند که دارای صحنه‌های پیچیده شامل ساختمان‌های با ابعاد و اشکال گوناگون هستند. نتیجه مدل نسبت به این تصاویر نشان از عملکرد بالای مدل پیشنهادی و قابلیت استخراج چند مقیاسی آن است. در مواردی که صحنه تماماً شامل

ساختمان‌های کوچک و متوسط است، تمامی مدل‌ها عملکردی شبیه به هم داشته‌اند. در مناطقی که به وسیله کادر قرمز رنگ مشخص شده‌اند؛ قابلیت بیشتر مدل پیشنهادی در استخراج پیوسته و صحیح ساختمان‌های بزرگ مشخص است. طبق جدول (۴)، در مواردی مثل تصاویر ۱ و ۴ که ساختمان‌های بزرگ بخش عمده محیط را تشکیل داده‌اند؛ مدل پیشنهادی نسبت به سایر مدل‌ها عملکرد بهینه‌تری دارد. طبق جدول (۳) که نتایج ارزیابی نسبت به تمام تصاویر آزمایش این مجموعه است برتری با مدل‌هایی است که با هدف استخراج چند مقیاسی طراحی شده‌اند؛ این مدل‌ها شامل مدل پیشنهادی، DeepLabV3+ و USPP است. در این ارزیابی مدل پیشنهادی با کسب مقادیر ۰/۸۵۷۲ و ۰/۹۱۹۷ در معیارهای IoU و Recall، موفق‌ترین مدل در بیشترین میزان استخراج صحیح ساختمان است.



شکل ۸- نتایج تشخیص ساختمان در تصاویر WHU با استفاده از مدل پیشنهادی



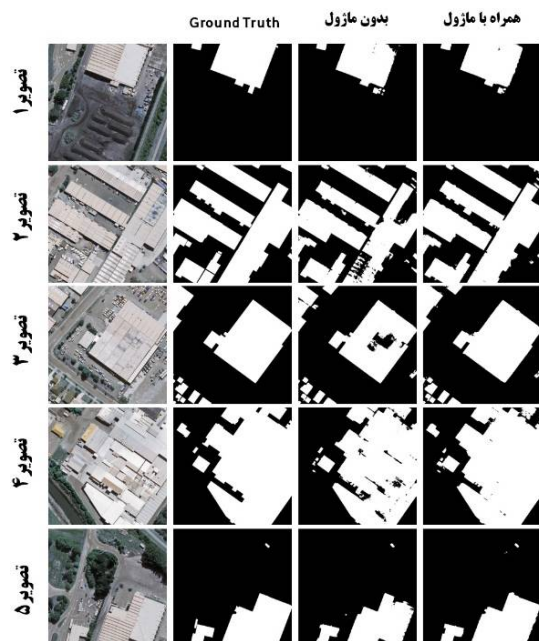
شکل ۹- نتایج تشخیص ساختمان در تصاویر مجموعه WHU با استفاده از مدل‌های مختلف، نمونه‌ای از ساختمان‌هایی که مدل‌ها عملکرد متفاوتی نشان داده‌اند با مربع قرمز رنگ مشخص شده‌است.

جدول ۴- ارزیابی خروجی مدل‌ها در تصاویر انتخابی مجموعه Massachusetts

تصویر	معیار	FCN	U-Net	DeepLabV3+	USPP	VGG-UNet	مدل پیشنهادی
تصویر ۱	Precision	۰.۹۷۶۰	۰.۹۶۸۷	۰.۹۸۶۸	۰.۹۹۸۲	۰.۹۹۰۰	۰.۹۹۲۷
	Recall	۰.۹۵۶۹	۰.۸۵۱۴	۰.۹۰۵۷	۰.۸۷۶۵	۰.۹۳۶۸	۰.۹۵۴۳
	F1-Score	۰.۹۶۶۳	۰.۹۰۶۳	۰.۹۴۴۵	۰.۹۳۳۴	۰.۹۶۲۷	۰.۹۷۳۲
	IoU	۰.۹۳۴۹	۰.۸۲۸۶	۰.۸۹۴۸	۰.۸۷۵۱	۰.۹۲۸۰	۰.۹۴۷۷
تصویر ۲	Precision	۰.۹۵۴۴	۰.۹۵۰۰	۰.۹۵۹	۰.۹۵۸۲	۰.۹۴۷۳	۰.۹۴۴
	Recall	۰.۹۴۳۵	۰.۹۶۱۰	۰.۸۴۵۰	۰.۹۵۰۲	۰.۹۶۳۷	۰.۹۶۸۷
	F1-Score	۰.۹۴۸۹	۰.۹۵۵۵	۰.۸۹۸۴	۰.۸۵۴۲	۰.۹۵۵۴	۰.۹۵۶۲
	IoU	۰.۹۰۲۸	۰.۹۱۴۷	۰.۸۱۵۶	۰.۹۱۲۴	۰.۹۱۴۶	۰.۹۱۶۱
تصویر ۳	Precision	۰.۹۲۴۴	۰.۸۹۷۹	۰.۹۱۷۱	۰.۹۲۹۶	۰.۹۱۶۴	۰.۹۲۳۱
	Recall	۰.۹۵۴۳	۰.۹۳۴۹	۰.۹۷۶۹	۰.۹۶۷۶	۰.۹۴۸۸	۰.۹۸۱۱
	F1-Score	۰.۹۳۹۱	۰.۹۱۶۰	۰.۹۴۶۰	۰.۹۴۸۲	۰.۹۳۲۳	۰.۹۴۹۶
	IoU	۰.۸۸۵۳	۰.۸۴۵۱	۰.۸۹۷۶	۰.۹۰۱۵	۰.۸۷۳۲	۰.۹۰۴
تصویر ۴	Precision	۰.۹۹۰۷	۰.۹۹۹۵	۰.۹۹۵۸	۰.۹۵۵۲	۰.۹۳۷۳	۰.۹۹۴۷
	Recall	۰.۶۷۹۵	۰.۷۹۰۰	۰.۹۲۶۳	۰.۹۴۳۱	۰.۸۶۵۱	۰.۹۳۷۶
	F1-Score	۰.۸۰۶۱	۰.۸۸۲۵	۰.۹۵۹۸	۰.۹۴۹۱	۰.۸۹۹۷	۰.۹۶۵۳
	IoU	۰.۶۷۵۲	۰.۷۸۹۷	۰.۹۲۲۷	۰.۹۰۳۲	۰.۸۱۷۸	۰.۹۳۲۹
تصویر ۵	Precision	۰.۹۱۹۴	۰.۹۴۵۰	۰.۹۴۵۴	۰.۹۴۶۶	۰.۹۴۸۹	۰.۹۱۹۵
	Recall	۰.۹۲۸۰	۰.۹۶۶۷	۰.۹۵۹۵	۰.۹۵۹۲	۰.۹۶۸۲	۰.۹۶۰۱
	F1-Score	۰.۹۲۳۷	۰.۹۵۵۷	۰.۹۵۲۴	۰.۹۵۲۸	۰.۹۵۸۵	۰.۹۳۹۴
	IoU	۰.۸۵۸۱	۰.۹۱۵۲	۰.۹۰۹۱	۰.۹۰۹۹	۰.۹۲۰۳	۰.۸۸۵۷

۴-۲- بررسی عملکرد ماژول بهبود دهنده

برای بررسی عملکرد ماژول طراحی شده، مدل پیشنهادی با حذف این ماژول نیز آموزش داده شد. مشخصات آموزش مشابه سایر مدل‌ها و طبق موارد ذکر شده اعمال شد. جهت بررسی تأثیر این ماژول، ۵ تصویر دارای ساختمان‌های بزرگ مقیاس از مجموعه WHU آزمایش شدند. شکل (۱۰) و جدول (۵) نتایج حاصل از این مقایسه را نشان می‌دهند. نتایج نشان دهنده عملکرد بهینه ماژول طراحی شده است. این ماژول با ساختار خود می‌تواند خروجی ناقص ساختمان‌های بزرگ را تا حد زیادی بهبود دهد. جدول (۵) نشان‌دهنده ارزیابی تصاویر شکل (۱۰) بر اساس معیارهای ذکر شده است. با آزمایش تمام تصاویر مجموعه WHU مشخص شد پس از اضافه شدن ماژول بهبود معیارهای IoU، Recall، F1-Score و Precision به ترتیب به میزان ۰/۱۳۱۹، ۰/۱۲۶۰ و ۰/۱۱۲۲ عملکرد بهتری نشان دادند.



شکل ۱۰- تفاوت نتایج طبقه‌بندی مدل پیشنهادی همراه با ماژول بهبود و بدون آن

جدول ۵- بررسی تأثیر ماژول بهبود توسط معیارهای ارزیابی

	معیار	تصویر ۱	تصویر ۲	تصویر ۳	تصویر ۴	تصویر ۵
بدون ماژول	Precision	۰/۹۵۹۹	۰/۹۸۳۷	۰/۹۹۵۰	۰/۹۸۷۲	۰/۹۸۹۸
	Recall	۰/۹۲۳۹	۰/۸۹۲۵	۰/۹۰۰۳	۰/۹۱۱۵	۰/۹۰۶۲
	F1-Score	۰/۹۴۱۶	۰/۹۳۵۹	۰/۹۴۵۳	۰/۹۴۷۸	۰/۹۴۶۱
	IoU	۰/۸۸۹۶	۰/۸۷۹۶	۰/۸۹۶۲	۰/۹۰۰۸	۰/۸۹۷۸
همراه با ماژول	Precision	۰/۹۶۰۱	۰/۹۷۴۳	۰/۹۹۲۲	۰/۹۷۴۶	۰/۹۹۴۷
	Recall	۰/۹۹۱۲	۰/۹۸۶۷	۰/۹۶۹۴	۰/۹۷۷۷	۰/۹۳۷۶
	F1-Score	۰/۹۷۵۴	۰/۹۸۰۵	۰/۹۸۰۷	۰/۹۷۶۱	۰/۹۶۵۳
	IoU	۰/۹۵۲۰	۰/۹۶۱۷	۰/۹۶۲۱	۰/۹۵۳۴	۰/۹۳۲۹

۵- نتیجه‌گیری

نیز هستند. از مهم‌ترین ضعف‌های مدل‌های موجود، ناتوانی در استخراج صحیح ساختمان با اندازه‌های متفاوت است. در این مقاله مدلی را ارائه کردیم که مشکل استخراج چند مقیاسی ساختمان‌ها را رفع می‌کند. مدل پیشنهادی این چالش را توسط یک ماژول بهبود نتیجه از بین می‌برد. اساس کار مدل پیشنهادی، استخراج ساختمان‌ها توسط ساختار رمزگذار و رمزگشا و رفع نواقص آن توسط ماژول بهبود است. ماژول پیشنهادی با افزایش میدان دریافت اطلاعات توسط پیچش گسترش یافته از روابط همسایگی پیکسل‌ها به طور مؤثر استفاده می‌کند. این ماژول با استفاده از این ساختار عدم پیوستگی ساختمان‌های بزرگ در خروجی را برطرف می‌کند. نتایج نشان‌دهنده قابلیت بالای مدل پیشنهادی در استخراج

امروزه یکی از نیازهای برنامه‌ریزی شهری، محیط زیست و مدیریت حوادث، استخراج ساختمان است. اهمیت این موضوع زمانی بیشتر می‌شود که بسیاری از سازمان‌ها و نهادها، نیازمند بروزرسانی مداوم نقشه‌ها و پایگاه داده‌های خود هستند. با توجه به موارد ذکر شده، استفاده از تصاویر سنجنش از دوری و تلفیق آن با روش‌های خودکار، اهمیت ویژه‌ای پیدا می‌کند؛ که باعث افزایش سرعت و دقت و کاهش هزینه می‌شود. در سال‌های اخیر در میان روش‌های موجود استخراج ساختمان از تصاویر سنجنش از دوری، شبکه‌های عصبی عمیق عملکرد بهتری نشان داده‌اند. اما این شبکه‌ها دارای ضعف‌هایی

چند مقیاسی ساختمان است. نسبت به تصاویر هر دو مجموعه استفاده شده، مدل پیشنهادی توانست بهترین عملکرد را نشان دهد. استخراج ساختمان از تصاویر دارای قدرت تفکیک مکانی متفاوت و دارای محیط‌های پیچیده نشان از توانایی بالای مدل پیشنهادی دارد.

به عنوان تحقیقات آتی، تهیه مدل‌های عمیقی برای استخراج ساختمان از تصاویر اخذ شده از مناطق شهری ایران را باید در نظر داشت. مناطق شهری ایران از ساختار و الگوی

متفاوتی نسبت به مناطق شهری سایر کشورها تشکیل شده‌اند. ساختمان‌های ایران مجاور هم و با استفاده از مصالح یکسانی برای سقف‌ها ساخته می‌شوند. بنابراین، مدل‌هایی که با استفاده از نمونه تصاویر خارج از ایران آموزش دیده باشند، قادر به تفکیک ساختمان‌ها در تصاویر اخذ شده از مناطق شهری ایران نیستند. بنابراین، نیاز است که مجموعه داده‌ای منحصراً از ساختمان‌های مناطق شهری ایران تهیه و مدل‌ها با استفاده از آن آموزش ببینند.

مراجع

- [۱] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 60-77, 2018/11/01/ 2018, doi: <https://doi.org/10.1016/j.isprsjprs.2018.04.014>.
- [۲] F. Mohammadimanesh, B. Salehi, M. Mahdianpari, E. Gill, and M. Molinier, "A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 151, 04/02 2019, doi: 10.1016/j.isprsjprs.2019.03.015.
- [۳] M. Vakalopoulou, K. Karantzaos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features", in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 26-31 July 2015 2015, pp. 1873-1876, doi: 10.1109/IGARSS.2015.7326158.
- [۴] X. Tong *et al.*, "Use of shadows for detection of earthquake-induced collapsed buildings in high-resolution satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 79, pp. 53-67, May 01, 2013 2013, doi: 10.1016/j.isprsjprs.2013.01.012.
- [۵] M. Rezaei, H. Arefi, H. Rastiveis, and M. Sajadian, "Building Extraction and Modeling Using LiDAR Point Clouds Imaging on Two-Dimensional Surface," (in eng), *Journal of Geomatics Science and Technology*, Research vol. 7, no. 3, pp. 139-150, 2018. [Online]. Available: <http://jgst.issge.ir/article-1-361-fa.html>.
- [۶] M. Khoshboresh Masouleh, R. Shah-Hosseini, and A. R. Safari, "Integration of Deep Learning Algorithms and Bilateral Filters with the Purpose of Building Extraction from Mono Optical Aerial Imagery," (in eng), *Journal of Geospatial Information Technology*, Research vol. 7, no. 2, pp. 241-26, ۲۰۱۹, ۲, doi: 10.29252/jgit.7.2.241.
- [۷] R. Yazdan, M. J. Valadan Zoj, H. Ebadi, and A. Mohammadzadeh, "Semi-Automatic Building Extraction Using Snake Models from High Resolution Aerial Images," (in 2), *Journal of Geomatics Science and Technology*, Research vol. 4, no. 2, pp. 179-188, 2014. [Online]. Available: <http://jgst.issge.ir/article-1-253-fa.html>.
- [۸] D. Tiede, G. Schwendemann, A. Alobaidi, L. Wendt, and S. Lang, "Mask R-CNN-based building extraction from VHR satellite data in operational humanitarian action: An example related to Covid-19 response in Khartoum, Sudan," *Transactions in GIS*, vol. 25, 05/01 2021, doi: 10.1111/tgis.12766.
- [۹] K. Karantzaos and N. Paragios, "Recognition-Driven Two-Dimensional Competing Priors Toward Automatic and Accurate Building Detection," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, pp. 133-144, 02/01 2009, doi: 10.1109/TGRS.2008.2002027.
- [۱۰] K. Khoshelham and Z. Li, "A model - based approach to semi - automated reconstruction of buildings from aerial images," (in en), *Photogrammetric record*, vol. 19, no. 108, pp. 342 - 359, 2004 2004, doi: urn:nbn:nl:ui:28-1530137e-e9f8-43df-8885-f3ea99f3a19b.
- [۱۱] G. Sohn and I. Dowman, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 1, pp. 43-63, 2007/05/01/ 2007, doi: <https://doi.org/10.1016/j.isprsjprs.2007.01.001>.
- [۱۲] Z. J. Liu, J. Wang, and W. P. Liu, "Building extraction from high resolution imagery based on multi-scale object oriented classification and probabilistic Hough transform," in *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, 29-29 July 2005 2005, vol. 4, pp. 2250-2253, doi: 10.1109/IGARSS.2005.1525421.
- [۱۳] R. Attarzadeh and M. Momeni, "Object-Based Building Extraction from High Resolution Satellite Imagery," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XXXIX-B4, pp. 57-60, 2012, doi: 10.5194/isprarchives-XXXIX-B4.۲۰۱۲-۵۷-

- [۱۴] S. Cui, Q. Yan, and P. Reinartz, "Complex building description and extraction based on Hough transformation and cycle detection," *Remote Sensing Letters*, vol. 3, no. 2, pp. 151-159, 2012/03/01 2012, doi: 10.1080/01431161.2010.548410.
- [۱۵] X. Huang and L. Zhang, "Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 161-172, 2012, doi: 10.1109/JSTARS.2011.2168195.
- [۱۶] Y. Liu, L. Gross, Z. Li, X. Li, X. Fan, and W. Qi, "Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder With Spatial Pyramid Pooling," *IEEE Access*, vol. 7, pp. 128, ۲۰۱۹, ۱۲۸۷۸۶-۷۷۴doi: 10.1109/ACCESS.2019.2940527.
- [۱۷] Z. Ye, Y. Fu, M. Gan, J. Deng, A. Comber, and K. Wang, "Building Extraction from Very High Resolution Aerial Imagery Using Joint Attention Deep Neural Network," *Remote Sensing*, vol. 11, no. 24, p. 2. ۲۰۱۹, ۹۷۰[Online]. Available: <https://www.mdpi.com/2072-4292/11/24/2970>.
- [۱۸] W. Kang, Y. Xiang, F. Wang, and H. You, "EU-Net: An Efficient Fully Convolutional Network for Building Extraction from Optical Remote Sensing Images," *Remote Sensing*, vol. 11, no. 23, p. 2813, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/23/2813>.
- [۱۹] 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, doi: 10.1109/TGRS.2018.2858817.
- [۲۰] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834-848, 2017.
- [۲۱] B. Artacho and A. Savakis, "Waterfall Atrous Spatial Pooling Architecture for Efficient Semantic Segmentation," *Sensors*, vol. 19, no. 24, p. 5361, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/24/5361>.
- [۲۲] Q. Zhu, C. Liao, H. Hu, X. Mei, and H. Li, "MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction From Remote Sensed Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6169-6181, 2021, doi: 10.1109/TGRS.2020.3026051.
- [۲۳] J. Cai and Y. Chen, "MHA-Net: Multipath Hybrid Attention Network for Building Footprint Extraction From High-Resolution Remote Sensing Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5807-5817, 2021, doi: 10.1109/JSTARS.2021.3084805.
- [۲۴] 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. [Online]. Available: <https://www.mdpi.com/2072-4292/12/15/2350>.
- [۲۵] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [۲۶] G. E. Hinton and V. Mnih, "Machine learning for aerial image labeling," 2013 .
- [۲۷] 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 .
- [۲۸] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 4-6 June 2018 2018, pp. 1-2, doi: 10.1109/IWQoS.2018.8624183 .