



تشخیص اشکال سه بعدی ساختمان از داده های LiDAR با استفاده از مدل یادگیری

عمیق Mask R-CNN

مهدی بخشی^{1*}، امیر پیشوا¹¹. کارشناسی ارشد دانشگاه خوارزمی تهران، سنجش از دور، سیستم اطلاعات جغرافیایی

چکیده

تشخیص اشیاء یک موضوع مهم در بینایی کامپیوتر است. روش های مبتنی بر شبکه عصبی Region-Based Convolutional Neural Network (R-CNN) اغلب برای تشخیص اشیاء استفاده می شوند. برای اینکه سیستم های هوشمند و مستقل در محیط های پویا و بدون ساختار مانند سایت های ساخت و ساز به طور مؤثر عمل کنند باید بتوانند مدل معنایی محیط را استنتاج یا بدست آورند. تا بتوان امکانات جدیدی را برای کاربرد های وسیعی مانند جهت دهی خودروهای خودران و نقشه برداری سه بعدی فراهم ساخت. در این پژوهش به تشخیص و تولید اشکال سه بعدی ساختمان پرداخته شده است که هدف از آن ارزیابی دقت مدل Mask R-CNN در تشخیص خودکار اشکال سه بعدی ساختمان می باشد. از داده های LiDAR که از ویژگی های Intensity، Return Beam و مؤلفه های (X Y Z) بهره می برد استفاده شده است. برای آموزش شبکه Convolutional Mask R-CNN و مجموعه داده های COCO مورد استفاده قرار گرفته است و در راستای بهبود عملکرد مدل، تست های Hyper parameters مختلفی از جمله Learning Rate، Batch Size و Data Augmentation انجام شده است. برای ارزیابی دقت تشخیص مدل از معیار Mean Average Precision (mAP) استفاده شده است که نتایج نشان می دهد مدل با دقت mAP 93 Score برای تشخیص ساختمان های منطقه مورد مطالعه مناسب می باشد.

کلمات کلیدی: یادگیری عمیق، تشخیص اشیاء، لیزر LiDAR، اشکال سه بعدی، Mask R-CNN

1- مقدمه

امروزه تشخیص مدل سه بعدی ساختمان‌ها یکی از چالش برانگیزترین موضوعات تحقیقاتی سنجش از دور و فتوگرامتری می باشد که تشخیص خودکار اشکال سه بعدی ساختمان‌ها در مناطق شهری با استفاده از داده های سنجش از دور، می تواند در کاربردهای مختلف مانند نقشه برداری شهری، تولید شهر سه بعدی، به روز رسانی سیستم اطلاعات جغرافیایی و نظارت بر ناو بری حائز اهمیت باشد [1]. هدف از تشخیص شیء¹ در یک تصویر تعیین موقعیت و دسته بندی هر شیء موجود در تصویر می باشد، به این صورت که با تشخیص هدف، ما قادر خواهیم بود بفهمیم که هر شیء در تصویر کجا قرار گرفته است و آن را در یک دسته مشخص قرار دهیم، و هدف از Instance Segmentation شناسایی تمام Pixels متعلق به هر شیء مورد نظر است [2]. موفقیت پردازش تصویر و CNN² های مبتنی بر منطقه (RCNN³) به دلیل روشهای شبکه عصبی Convolutional (CNN) است [3]، در مطالعات متعددی از این روش‌ها به طور گسترده استفاده شده، علاوه بر این CNN نسبت به روش های مرسوم که بر استخراج دستی ویژگی ها متکی هستند موثر تر است [4, 5, 6]. علی رغم پیشرفت های سریع در علم اطلاعات جغرافیایی استخراج ساختمان‌ها در مقیاس بزرگ یک چالش اساسی در نقشه برداری این حوزه است [7, 8, 9]. در اواخر دهه 1990 بود که استفاده از CNN دوباره شروع شد و یک چالش در تشخیص بصری در مقیاس بزرگ، مشخص شد که CNN در مقایسه با رقبای خود بهترین عملکرد را دارد [1, 10]. تشخیص اشیاء گام مهمی در تولید اشکال سه بعدی است، که وظیفه تشخیص شیء عمدتاً بر روی اطلاعات مکانی انواع اشیاء در تصویر متمرکز است، مانند علامت گذاری موقعیت شیء شناسایی شده در تصویر با یک کادر مستطیلی می باشد [11]. مدل Mask R-CNN در کار تشخیص شیء بر روی مجموعه داده COCO در مقایسه با دیگر روش‌ها از نظر دقت تشخیص برتری دارد [12]. Mask R-CNN یک روش یادگیری عمیق بسیار محبوب برای تشخیص اشیا و Semantic Segmentation است که به نتایج فوق العاده ای در مجموعه داده COCO دست یافته است. هر چند مدل های دیگری توانسته اند از مدل Mask R-CNN در عملکرد معیار mAP پیشی بگیرند اما معمولاً بر اساس معماری Mask R-CNN هستند [13]. در این پژوهش از مدل Mask R-CNN برای تشخیص ساختمان و از مجموعه داده COCO⁴ برای آموزش استفاده شده است.

¹ - Object detection

² - Convolutional neural network

³ - Region convolutional neural network

⁴ - Common Objects In Context

لذا هدف این تحقیق، استفاده از این مدل برای بالا بردن دقت آموزش بر روی مجموعه داده های COCO برای تشخیص و تولید اشکال سه بعدی ساختمان از داده های LiDAR¹ می باشد.

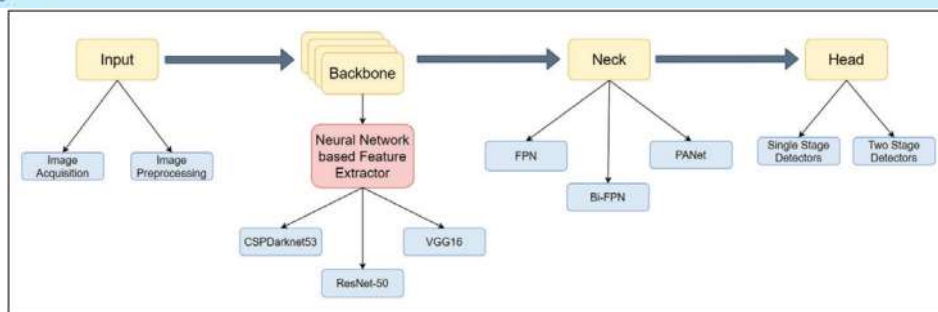
تشخیص شیء Object Detection

تشخیص اشیاء یک کار بینایی کامپیوتری مهم است که با شناسایی نمونه هایی از اشیاء بصری کلاس مشخص (مانند انسان، حیوانات یا اتومبیل) در تصاویر رقومی سروکار دارد. دو معیار مهم برای تشخیص اشیاء، دقت (از جمله دقت طبقه بندی و دقت محلی سازی) و سرعت هستند [14]. تشخیص اشیاء به تعیین موقعیت و چیدمان یک شیء در تصویر، تشخیص وسیله نقلیه، نظارت و غیره کمک می کند. تفاوت بین Algorithms تشخیص اشیاء و Algorithms طبقه بندی این است که در Algorithms تشخیص، یک Bonding Box در اطراف شیء مورد نظر ترسیم می شود تا آن را در تصویر مشخص کند [3]. هر کادر همچنین دارای یک امتیاز اطمینان^۲ است که نشان می دهد مدل چقدر مطمئن است یک شیء خاص قابل شناسایی در مکان مشخصی از تصویر وجود دارد. به عنوان یک مرحله پس از پردازش، کادر هایی که نمره آن ها زیر یک آستانه مشخصی قرار دارد با استفاده از Non Algorithm Maximum Suppression شناسایی و حذف می شود. هر تکنیک تشخیص تصویر مبتنی بر CNN عملکردهای اساسی مشترکی برای انجام دارد زیرا در نهایت، همه این تکنیک ها باید به یک هدف مشترک برسند که طبقه بندی و تشخیص اشیاء موجود در تصویر است [15]. دو نوع تشخیص شیء مبتنی بر CNN وجود دارد، مدل تشخیص شیء Single Shot یا One-Stage و مدل تشخیص شیء Two-Stage یا Region Proposal-based. در مدل های Two-Stage تصویر بر اساس ROI³ (منطقه مورد علاقه) تقسیم می شود و دقت بالایی در محلی سازی و تشخیص اشیاء دارد. هر مرحله نیاز به توجه ویژه ای دارد زیرا فرآیندهای پیچیده درگیر هستند. هر جزء باید به صورت جداگانه آموزش داده شود، بنابراین بهینه سازی این مدل ها دشوار است. به دلیل محاسبات پیچیده، مدل های Two-Stage در مقایسه با مدل های One-Stage به دقت بیشتری دست می یابند اما کند هستند و شامل مدل های R-CNN، Fast R-CNN، Faster R-CNN و Mask R-CNN می باشد [16].

¹ - Light Detection and Ranging

³ - Confidence Score

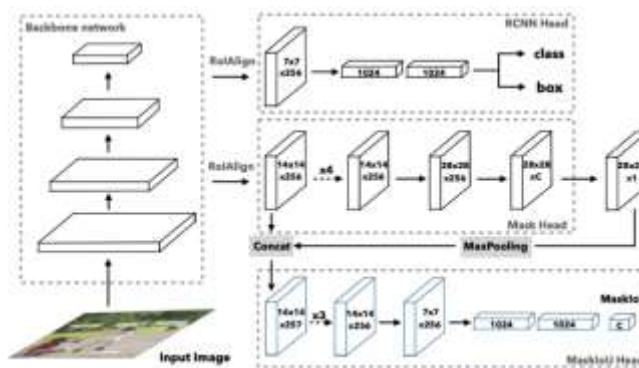
³ - Region of Interest



شکل 1- چارچوب کلی تشخیص شیء [15]

مدل Mask R-CNN

مدل شبکه عصبی Convolutional Mask R-CNN با هدف انجام تشخیص شیء به صورت Pixel طراحی شده است. این مدل یک نمونه اصلاح شده مدل Faster R-CNN می باشد که از یک مرحله اضافی به نام Masking برخوردار است. در این مدل، هر Pixel اسکن شده و پیش‌بینی می شود که آیا یک قسمت از شیء است یا خیر. اگر داده‌های آموزش با موقعیت Pixel هر شیء در تصویر برجسب‌گذاری شده باشند، مدل Mask R-CNN می تواند با استفاده از این برجسب‌های دقیق، دقت تشخیص شیء را به طور کارآمدی افزایش دهد. در این مدل، لایه RoI pooling با لایه RoI alignment جایگزین شده است. لایه RoI alignment دسته‌بندی و کادرهای محدود کننده RoI را پیش‌بینی می کند و امکان استفاده از یک شبکه Convolutional کاملاً توسعه یافته را برای پیش‌بینی موقعیت Pixel شیء فراهم می کند، زیرا نقشه ویژگی همه RoI ها را با همان شکل خروجی می دهد [17]. در این تحقیق از مدل Mask R-CNN برای تشخیص ساختمان ها استفاده شده است.





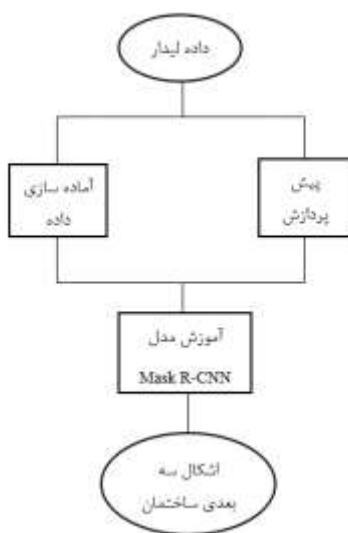
ISC_{CM}
International Science Citation Center

OxfordCert
UNIVERSAL

شکل 2- معماری مدل Mask R-CNN

2- مواد و روش ها

انجام کلی این تحقیق که در چندین مرحله می باشد در شکل 3 نمایش داده شده است.



5

شکل 3- فلوچارت کلی نحوه انجام تحقیق

منطقه مورد مطالعه

Beverly Hills یک روستا در ایالات متحده آمریکا است که در شهرستان Michigan واقع شده است. این روستا 1041 کیلومتر مربع مساحت و 222 متر بالاتر از سطح دریا در موقعیت جغرافیایی (83 درجه، 14 دقیقه، 32 ثانیه غربی) و (42 درجه، 31 دقیقه، 19 ثانیه شمالی) قرار گرفته است.

داده LiDAR



برای انجام این پژوهش از داده های لیزر LiDAR برای منطقه مورد مطالعه استفاده شده است، داده هایی که توسط شرکت ESRI به صورت ابرنقطه ای بوده اند با استفاده از ابزار Geoprocessing به یک لایه Raster تک باند با وضوح 0.2 متر مربع در هر Pixel تبدیل شده است.

آماده سازی داده و پیش پردازش

در این بخش، داده های آموزشی با فرمت TIFF و با ابعاد 256 در 256 تولید شده است. برای تهیه این داده ها، تصاویر LiDAR اصلی به قسمت های کوچک تر تقسیم شده اند که هر کدام از آن ها یک Chip تصویر را شامل می شود. همچنین، اطلاعات ابر داده نیز به صورت فایل های جداگانه ایجاد و ذخیره شده اند که دارای اطلاعاتی از هر Chip تصویر شامل ویژگی های خاصی از داده لیزر LiDAR می باشد. برای اینکه تمامی داده ها مقیاس یکسانی داشته باشند از روش نرمال سازی Z-Score Normalization که باعث بهبود عملکرد مدل می شود، و همچنین از روش های Augmentation مانند Rotation و Brightness برای افزایش تنوع و تشویق مدل به یادگیری ویژگی های مختلف استفاده شده است. برای بهبود تشخیص، تمیزتر دیده شدن اشیاء و مقابله با نوع متغیر نوری محیط از داده های Multispectral با نوع MS¹ استفاده شده است. انواع پیش پردازش های ذکر شده سیستم را قادر می سازد تا مدل بهتری را با داده های محدود آموزش دهد و همچنین از برآزش بیش از حد مدل جلوگیری کند.

آموزش مدل و تولید اشکال سه بعدی

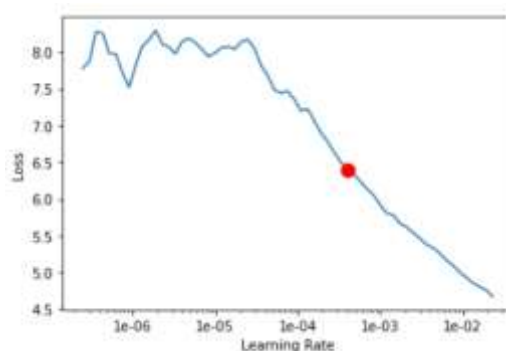
در این مرحله داده هایی که در مرحله قبل تولید شده است را به همراه داشته و مدل شروع به آموزش می کند. مدل Mask R-CNN به همراه مجموعه داده COCO با تنظیم مشخصه های مختلفی برای تشخیص و قرار دادن قطعات سقف داخل Bonding Box آموزش داده شده است. از جمله این تنظیم، اندازه سلول برای تشخیص ویژگی ها به منظور استخراج اطلاعات سه بعدی از داده LiDAR تعیین می شود. همچنین اندازه Tile و Stride نیز به ترتیب مقدارهای 256 و 128 تنظیم شده اند که منجر به تقسیم بهتر و همگن Chip تصویر می شوند. از پارامتر Intensity یا شدت نور لیزر LiDAR به عنوان یک کانال اضافی به داده های LiDAR اضافه شده است. در فرآیند آموزش مدل، Intensity به صورت یک مقدار عددی برای هر نقطه در داده های

¹ - Map Scale

LiDAR در نظر گرفته شده است و مدل به گونه ای آموزش دیده است که از این ویژگی نیز برای تشخیص ساختمان ها استفاده کند. و همچنین، از پارامتر Return Beam، مؤلفه های (X Y Z) و Raster های DSM برای بدست آوردن ارتفاع پایه برای هر بخش ساختمان و محاسبه جهت پشته سقف استفاده شده است. مرحله آخر استفاده از ابزار GP¹ "ویژگی‌های قوانین City Engine" می باشد تا بازسازی نهایی ساختمان‌های سه بعدی اعمال شود تا خواسته نهایی که تولید اشکال سه بعدی ساختمان های منطقه مورد مطالعه می باشد محقق شود.

3- بحث و نتایج

آموزش مدل Mask R-CNN جهت تشخیص ساختمان و تولید تصویر سه بعدی آنها باید از دو جهت مورد بررسی قرار بگیرد یکی دقت طبقه بندی و دیگری دقت Bonding Box یا جعبه که در مدل Mask R-CNN وجود دارد. برای پیدا کردن بهترین نرخ یادگیری از تابعی استفاده شده است که مقدار بهینه نرخ یادگیری را با استفاده از روش تست کاهش نرخ یادگیری^۲ پیدا می کند. این روش در شروع با نرخ یادگیری کوچک و افزایش آن تا یک مقدار بزرگ، مدل را آموزش می دهد و میزان دقت یا تابع هزینه را برای هر نرخ یادگیری ثبت می کند. سپس با توجه به نمودار به دست آمده، مقدار بهینه ی نرخ یادگیری برای آموزش مدل انتخاب می شود.



شکل 4- نمودار انتخاب نرخ یادگیری

¹- Genetic Programming

²- Learning Rate Range Test

شکل (4) که خروجی انتخاب بهترین نرخ یادگیری می باشد نشان می دهد که چگونه تابع هزینه¹ به تغییر نرخ یادگیری واکنش نشان می دهد. در این نمودار محور X نشان دهنده نرخ یادگیری و محور Y نشان دهنده مقدار تابع هزینه در طول آموزش مدل است. ابتدا، نرخ یادگیری بسیار کوچک است و مقدار تابع هزینه نیز کم است. سپس با افزایش نرخ یادگیری، مقدار تابع هزینه به طور کلی کاهش می یابد. چون مدل بهتر شده و پیش بینی های بهتری ارائه می دهد. اما در این فرآیند در یک نقطه مشخص، مقدار تابع هزینه به طور ناگهانی افزایش می یابد و بالا می رود. این نقطه برابر با مقدار بهینه ی نرخ یادگیری است که باید برای آموزش مدل استفاده شود. علاوه بر نمودار یک عدد هم به ما نمایش می دهد که نشان دهنده مقدار بهینه نرخ یادگیری است که در آن تابع هزینه به مقدار کمینه ی خود می رسد. در این پژوهش این مقدار **0.0003981071705534973** می باشد به این معنی که در این نقطه، مدل به بهترین شکل ممکن آموزش دیده است و می توان از آن برای آموزش مدل نهایی استفاده کرد.

نتایج آموزش مدل

8

در این مرحله مدل با دوره هایی (Epoch) که مشخص شده است و نرخ یادگیری که از مرحله قبل به آن منتقل شده آموزش داده می شود. به این ترتیب، مدل با استفاده از مقدار بهینه نرخ یادگیری که پس از اجرای روش تست کاهش نرخ یادگیری به دست آمده است، آموزش داده می شود. در هر دوره از آموزش مدل با داده های آموزشی به روز رسانی می شود تا تابع هزینه کمینه شود. این به معنی بهبود عملکرد مدل در پیش بینی داده های آموزشی است با انجام آموزش بر روی داده های آموزشی، مدل به تدریج به یادگیری الگوهای موجود در داده ها می پردازد و پیش بینی دقیقی برای داده های جدید ارائه می دهد.

جدول 1- خروجی نتایج آموزش مدل

دوره	خطای آموزش	خطای صحت سنجی	زمان (برحسب دقیقه و ثانیه)
0	1.606340	1.920804	8:51
1	1.740587	1.903324	8:50
2	1.532840	1.865784	8:53
3	1.610261	1.812871	8:58

¹- Loss Function



4	1.379416	1.732130	9:01
5	1.425437	1.664806	9:03
6	1.470336	1.636349	9:06
7	1.341473	1.595248	9:05
8	1.412700	1.571953	9:02
9	1.265569	1.567657	9:07

در جدول (1) شاهد افزایش خطای صحت سنجی و خطای آموزش در بعضی از دوره ها می باشیم، اما این لزوماً به معنای بد شدن نتیجه نیست. از آنجایی که داده‌های آموزشی بسیاری از ساختمان‌ها را نیز از دست داده است، از دست دادن گاهی اوقات عملکرد واقعی مدل را نشان می‌دهد، بنابراین در عوض می‌توان به نتایج واقعی نگاه کرد.

9

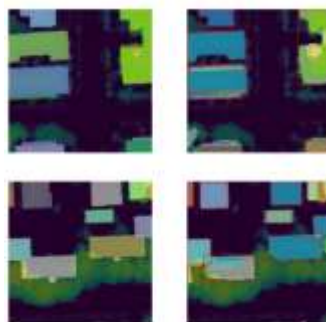
نمایش تجسم¹ تصویری نتایج

نمایش تجسمی نتایج آموزش مدل، یکی از روش‌های مفید برای ارزیابی عملکرد مدل است. در واقع، با استفاده از نمایش تجسمی به بررسی دقیق تر نتایج آموزش پرداخته شده است.



شکل 5- تصویری از چند Chip بعد از آموزش مدل به صورت تصادفی

¹ - Visualize



شکل 6- تصویری از چند Chip با حد آستانه بالای 50 درصد بعد از Bonding Box

شاخص mAP برای هر یک از کلاس‌های داده، دقت تشخیص اشیاء^۱ را اندازه‌گیری می‌کند. برای این منظور، با تعریف یک مقدار آستانه^۲، می‌توان تصمیم گرفت که آیا مدل برای یک شیء خاص در تصویر آن را شناسایی کرده است یا خیر. در اینجا، آستانه به عنوان یک مقدار ثابت تعیین شده است که اگر احتمال شناسایی شیء توسط مدل برای آن شیء بیشتر از آستانه باشد، آن شیء به عنوان شناسایی شده در نظر گرفته می‌شود. سپس، برای هر کلاس، از دو معیار Precision و Recall برای اندازه‌گیری دقت شناسایی استفاده شده است. در این روش، دقت به عنوان تعداد اشیاء که به درستی شناسایی شده است، تقسیم بر تعداد کل اشیاء شناسایی شده توسط مدل در نظر گرفته شده است. به عبارت دیگر، دقت نشان می‌دهد که از اشیاء که مدل آن‌ها را شناسایی کرده است، چه تعدادی به درستی شناسایی شده‌اند. همچنین، Recall به عنوان تعداد اشیاء که به درستی شناسایی شده‌اند تقسیم بر تعداد کل اشیاء موجود در داده‌های آموزشی یا تست در نظر گرفته شده است. به عبارت دیگر، Recall نشان می‌دهد که از تمام اشیاء موجود در تصویر، چه تعدادی به درستی شناسایی شده است. سپس، با استفاده از این دو معیار، برای هر کلاس دو منحنی Precision و Recall و تعیین بهترین آستانه به منظور انتخاب بهترین آستانه برای آن کلاس به دست آمده است. سپس،

^۱- Object Detection

^۲- Threshold

میانگین Precision و Recall برای هر کلاس به عنوان مقدار دقت مثبت (AP^1) برای آن کلاس در نظر گرفته شده است. در نهایت، میانگین این مقادیر برای تمام کلاس‌ها به عنوان میانگین دقت مثبت (mAP) مدل در مسئله شناسایی اشیاء در تصاویر استفاده شده است. به طور خلاصه، شاخص mAP برای ارزیابی عملکرد مدل‌های شناسایی اشیاء بر اساس Precision و Recall استفاده می‌شود. با محاسبه این شاخص، میانگین دقت عملکرد مدل در شناسایی اشیاء در تصاویر بدست آمده است. در عمل، این شاخص به کمک دو منحنی Precision و Recall و تعیین بهترین آستانه، ارزیابی دقیق و جامعی از عملکرد مدل در شناسایی اشیاء در تصاویر ارائه می‌دهد. در این پژوهش از این شاخص برای ارزیابی دقت مدل استفاده شده است که دقت 0.93 به دست آمده است که نشان می‌دهد میانگین دقت مدل در تشخیص سقف ساختمان‌ها در منطقه مورد مطالعه مطلوب است.

4- نتیجه گیری

11

در این مقاله، به تولید تشخیص اشکال سه بعدی ساختمان‌ها از روی داده‌های LiDAR با استفاده از یادگیری عمیق پرداخته شده است. برای این منظور، ابتدا داده‌های LiDAR را از سطح ساختمان استخراج کرده و سپس با استفاده از یادگیری عمیق و شبکه‌های مولد مبتنی بر یادگیری عمیق، تلاش شده است تا اشکال سه بعدی ساختمان‌ها با دقت بالا تولید شود. برای ارزیابی روش پیشنهادی، از داده‌های LiDAR با کیفیت بالا استفاده شده است. نتایج به دست آمده نشان می‌دهد که روش پیشنهادی می‌تواند با دقت بالا (0.93 درصد) اشکال سه بعدی ساختمان‌ها را تولید کند. علاوه بر تولید اشکال سه بعدی ساختمان‌ها، سقف‌های ساختمان‌های منطقه مورد مطالعه تشخیص و با استفاده از مدل Mask R-CNN و مجموعه داده COCO آموزش داده شده است و هر ساختمان داخل یک Bonding Box با دقت بالا قرار داده شده است. با توجه به دقت به دست آمده می‌توان گفت مدل Mask R-CNN کارایی و عملکرد خوبی از خود در تشخیص ساختمان نشان داده است.

منابع

1. Alidoost, F., & Arefi, H. (2017). Comparison of UAS-based photogrammetry software for 3D point cloud generation: a survey over a historical site. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 55-61.
2. Ashwani Kumar, SS Sai Satyanarayana Reddy, and Vivek Kulkarni. An object detection technique for blind people in real-time using deep neural network. In 2019 Fifth International Conference on Image Information Processing (ICIIP), pages 292-297. IEEE, 2019.

¹- Average Precision



3. Bi, Q., Qin, K., Zhang, H., Zhang, Y., Li, Z., & Xu, K. (2019). A multi-scale filtering building index for building extraction in very high-resolution satellite imagery. *Remote Sensing*, 11(5), 482.
4. Cai, Z., Ma, H., & Zhang, L. (2019). A building detection method based on semi-suppressed fuzzy C-means and restricted region growing using airborne LiDAR. *Remote Sensing*, 11(7), 848.
5. Chua, L. O., & Roska, T. (1993). The CNN paradigm. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 40(3), 147-156.
6. Dalal, A.A., Shao, Y., Alalimi, A. and Abdu, A., 2020. Mask R-CNN for geospatial object detection. *International Journal of Information Technology and Computer Science (IJITCS)*, 12(5), pp.63-72.
7. Dollár, K. H. G. G. P., & Girshick, R. (2017, October). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
8. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
9. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
10. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
11. Rottensteiner, F., Trinder, J., Clode, S., & Kubik, K. (2007). Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2), 135-149.
12. Shetty, A. K., Saha, I., Sanghvi, R. M., Save, S. A., & Patel, Y. J. (2021, April). A review: Object detection models. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-8). IEEE.
13. Shetty, J., & Jogi, P. S. (2019). Study on different region-based object detection models applied to live video stream and images using deep learning. In *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)* (pp. 51-60). Springer International Publishing.
14. Xiong, H., Wu, J., Liu, Q., & Cai, Y. (2020). Research on abnormal object detection in specific region based on Mask R-CNN. *International Journal of Advanced Robotic Systems*, 17(3), 1729881420925287.
15. Zhao, Z. B., Zhang, L., Qi, Y. C., & Shi, Y. Y. (2017). A generation method of insulator region proposals based on edge boxes. *Optoelectronics Letters*, 13(6), 466-470.
16. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature genetics*, 51(1), 12-18.
17. Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*.