نشریه تحلیل فضایی مخاطرات محیطی، سال یازدهم، شماره ۱، بهار ۱۴۰۳

صفحات ۱ – ۱۵

ارزیابی استفاده از هوش مصنوعی در تکمیل داده‌های داوطلبانه مکانی. مطالعه موردی: داده‌های کاربری اراضی OSM

جواد سدیدی[1]؛ دانشیار گروه سنجش از دور و GIS, دانشکده علوم جغرافیایی, دانشگاه خوارزمی.

Email: jsadidi@khu.ac.ir

فاطمه تام نیا؛ کارشناس ارشد GIS, گروه سنجش از دور و GIS, دانشکده علوم جغرافیایی, دانشگاه خوارزمی

هانی رضائیان؛ استادیار گروه سنجش از دور و GIS, دانشکده علوم جغرافیایی, دانشگاه خوارزمی

دریافت مقاله :۱۴۰۲/۰۸/۱۳ پذیرش نهایی:۱۴۰۳/۰۶/۰۳

چکیده

یادگیری عمیق یک روش مدرن پردازش تصویر و تجزیه و تحلیل داده هاست که با داشتن نتایج امیدوار کننده و پتانسیل بالا وارد حوزه مدیریت شهری شده است. پروژه Open Steet Map(OSM) بزرگ‌ترین مجموعه داده های مکانی داوطلبانه است که در بسیاری از حوزه‌های کاربردی مختلف به عنوان مکمل یا جایگزین با داده‌های مرجع استفاده می‌شود. در بعضی از موارد در کشورهای پیشرفته دقت داده‌های داوطلبانه تولید شده توسط موبایل و دیگر ابزار توسط کاربران حتی بیش از داده ی مرجع دولتی می‌باشد. هدف از تحقیق حاضر ارزیابی استفاده از هوش مصنوعی در تکمیل داده های داوطلبانه در مناطق کمتر مشارکت شده توسط داوطلبان می باشد. ابتدا با استفاده از شبکه عصبی کانولوشنی Res_UNet کاربری اراضی با دقت ۸۳ درصد به دست آمد، سپس با توجه به پیش‌بینی انجام شده، از روش واحد مبنا جهت ارزیابی میزان کامل بودن داده‌های OSM استفاده شد. نتایج نشان می‌دهد میزان کامل بودن بلوک‌های ساختمانی OSM در کل منطقه مطالعاتی برابر با ۳/۶ درصد، جنگل‌ها۹/۷درصد، درخت‌های میوه ۹۰/۴ درصد و زمین‌های کشاورزی ۸۱/۸۸درصد می‌باشد. که نشان از نرخ پایین کامل بودن بلوک‌های ساختمانی و جنگل و نرخ بالای کامل بودن زمین‌های کشاورزی و درختان میوه می‌باشد. نتایج تحقیق بیانگر درصد مشارکت پایین داوطلبانه درتولید داده‌های مکانی می‌باشد. از طرفی دقت بالای تولید کاربری اراضی توسط هوش مصنوعی نتایج امیدوارکننده‌ای را در استفاده از هوش مصنوعی در تولید و تکمیل داده‌های داوطلبانه به جای نیروی انسانی بخصوص در کشورهای کمتر توسعه‌یافته یا مناطق با جمعیت داوطلب کمتر یا نقاط دورافتاده و صعب‌العبور ارائه میدهد.

واژه‌های کلیدی: هوش مصنوعی، کاربری اراضی، کامل بودن، یادگیری عمیق، OSM و واحد مبنا، کرج

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]

[DOR: 20.1001.1.24237892.1403.11.1.1.1]

[DOI: 10.61186/jsaeh.11.1.3]

مقدمه

در سال‌های اخیر تهیه و به‌روزرسانی نقشه‌ها با استفاده از روش‌های متداول نقشه برداری کندتر از قبل صورت می‌پذیرد. یکی از مهم‌ترین دلایل این امر بالارفتن هزینه تهیه نقشه‌ها از طریق روش‌های مرسوم است(Zielstra,2010). امروزه با به و جود آمدن فناوری‌های جدید خصوصاً وب ۲/۰[1]و فراگیری استفاده از آن امکان تولید و جمع‌آوری داده‌ها و اطلاعات توسط تمامی افراد حتی غیرمتخصص فراهم شده است. با توسعه سامانه اطلاعات مکانی و آشکارشدن اهمیت و کاربردهای آن، انگیزه زیادی برای تولید اطلاعات مکانی توسط مشارکت‌کنندگان، فراهم شده است(Cooper et all,2012). سامانه اطلاعات مکانی وب مبنا از این منظر ابزارهای توانمندی را برای برنامه‌ریزی‌های مشارکتی در اختیار افراد قرار می‌دهند. در همین راستا گود چایلد در سال ۲۰۰۷، با درک نقش انسان حتی با کمترین تخصص و دانش در تولید اطلاعات مکانی، مفهوم تازه‌ای به نام VGI[2] را مطرح کرد. او انسان‌ها را به حسگرهایی تشبیه می‌کند که می‌توانند محیط اطراف خود را حس کرده و اطلاعاتی را در خصوص این محیط تولید و منتشر کنند(Goodchild,2012). تولید نقشه‌های LULC معمولاً توسط افراد بسیار آموزش دیده و ماهر انجام می‌شود که داده‌ها را تفسیر و طبقه‌بندی می‌کنند. پردازش داده‌ها در قالب مناسب برای استخراج اطلاعات، تجزیه و تحلیل و طبقه‌بندی (شامل ارزیابی کیفیت و تولید محصول و مستندسازی) است (Caetano et all 2006)؛ بنابراین زمان و هزینه تولید افزایش می‌یابد. این محدودیت‌ها به این معنی است که تمرکز نقشه برداری LULC بر روی مضامین و مناطق مهم‌تر و برای استفاده در چندین برنامه کاربردی در نظر گرفته می‌شوند. این همچنین تأثیر منفی بر استراتژی‌های به‌روزرسانی دارد. در نتیجه، این پایگاه‌های اطلاعاتی در برخی مناطق به سرعت منسوخ می‌شوند(Cihlar ,2000) . سامانه اطلاعات جغرافیایی مبتنی بر مشارکت مردمی[3] به روش‌هایی اشاره دارد که اطلاعات مکانی را برای انواع برنامه‌ریزی‌های شهری و منطقه‌ای فراهم می‌کند(Peng,2003). داده‌های جغرافیایی داوطلبانه، داده‌های متن‌باز [4] هستند که توسط تعداد زیادی از کاربران غیرحرفه‌ای و عموم مردم ثبت شده است(Wang,2013). پروژه‌های شناخته شده‌ای مانندOSM ،Wikimapia ،Google mapmake(که بر اساس این ایده شکل‌گرفته‌اند) کاربران را قادر می‌سازد تا نقشی در تولید داده‌های مکانی داشته باشد. نرخ بالای افزایش کاربران چنین محیط هایی نشان از سطح محبوبیت این پروژه می‌باشد(Hashemi et all,2015). پروژه OSM به عنوان یکی از موفق‌ترین و بزرگ‌ترین پروژه‌ها در این زمینه می‌باشدOSM یک نقشه آنلاین، رایگان و قابل ویرایش توسط کاربران با پوشش جهانی تحت وب برای عموم افراد فراهم می‌آورد. با توجه به مشارکت تعداد زیاد کاربران، این مجموعه داده‌ها به صورت پویا به‌روزرسانی می‌شوند. مزایای فراوان اطلاعات مکانی داوطلبانه همچون پوشش جهانی حتی برای شهرهای کمتر توسعه یافته، قابلیت به‌روزرسانی سریع و هزینه پایین دسترسی به اطلاعات، همواره کاربران را به استفاده از اطلاعات مکانی داوطلبانه سوق داده است. به دلیل انعطاف‌پذیری و ساختار داده باز OSM امکان استفاده آن برای طیف گسترده‌ای از برنامه‌های کاربردی وجود خواهد داشت. سیستم‌های ناوبری و مسیریابی، بهبود کارتوگرافی توسعه خدمات مکان‌مبنا و توسعه مدل‌های شهری سه بعدی نمونه ای از طیف گسترده کاربردهای پیاده سازی شده بر داده‌های OSM هستند.

1.Web2.0
2. Geographic Volunteered
3. People's participation
4. Open source data

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

با وجود مزایای فراوان محیط OSM، کیفیت داده‌های این محیط بزرگ‌ترین محدودیت آن است. یکی از دلایل اصلی این موضوع، عدم اطمینان در مورد نحوه جمع‌آوری این داده‌ها می‌باشد علاوه بر این، فقدان استانداردهای یکپارچه، عدم وجود مکانیزم‌های کنترل کیفیت و استفاده از فناوری‌ها و ابزارهای مختلف در محیط OSM، کیفیت ناهمگن در این محیط را افزایش می‌دهد (Feick et all,2013). بنابراین مهم‌ترین چالش پیش روی این محیط‌ها تردید درمورد کیفیت داده‌های گردآوری شده می‌باشد. از این رو لازم است که کیفیت این اطلاعات بررسی و تضمین شود تا بتوان از آنها در مقابل داده های رسمی در کاربردهای مختلف استفاده کرد. سازمان بین المللی استاندارد در قالب سری ISO/TC211 دستورالعمل هایی را در زمینه کیفیت داده های مکانی تعریف کرده است. بر اساس استاندارد ISO 19157 کیفیت داده های مکانی شامل المان های مختلفی می باشد که عبارت اند از: دقت موقعیتی، کامل بودن، سازگاری منطقی، دقت معنایی و دقت زمانی(Hecht et all,2013). در این مطالعه کامل بودن که یکی از مهمترین عناصر کیفیت ذکر شده به شمار می رود، هدف بررسی این مطالعه در نظر گرفته شده است. شبکه عصبی کانولوشن(CNN) به عنوان نماینده ی شبکه عصبی عمیق ، برای طبقه بندی پردازش و تجزیه و تحلیل تصاویر بزرگ مقیاس محلی و جهانی طراحی شده است(Lecun et all,2015). شبکه های عصبی کانولوشن(CNN) زمینه جدید و موفقی برای طبقه بندی کاربری زمین (LU) شهری مورد استفاده قرار گرفته است(Zhang et all, 2018). CNN عملکرد بسیار خوبی در شناسایی و طبقه بندی تصویر دارد و دقت بالاتری نسبت به روش های طبقه بندی استاندارد، مانند SVM، جنگل تصادفی (RF) ، رگرسیون لجستیک و سایر روش های مشابه ارائه می دهد(Chen et all, 2018; Marmanis et all, 2016). استفاده از داده‌های مکانی داوطلبانه، برای کاربری‌ها نیازمند تخمین کیفیت این اطلاعات است. بیشتر این اطلاعات توسط داوطلبان با مهارت کم یا بدون مهارت در ترسیم نقشه تولید شده و می‌شود و این موضوع باعث عدم گسترش استفاده از این داده‌ها در پردازش‌های رسمی تا زمان تخمین کیفیت اطلاعات می‌گردد. در نتیجه، ارزیابی سطح کیفی این اطلاعات به یکی موضوعات اساسی برای استفاده از این اطلاعات تبدیل شده است. کیفیت داده‌های Open Street Map در چین مورد بررسی قرار گرفته و نشان داد که ۶۶٪ از داده‌های OSM کشور چین دقیق هستند(Zheng,2014). XIE و همکارانش با استفاده از تصاویر ماهواره ای با وضوح بالا و یادگیری عمیق به ارزیابی کامل بودن داده های OSM در شهر لاس وگاس ایالت نوادا آمریکا پرداختند. تحقیق آنها نشان داد که روش پیشنهادی میتواند OSM را به طور موثر و دقیق ارزیابی کند و نشان داد که ساختمان های استخراج شده با روش پیشنهادی میتواند به عنوان یک منبع جدید برای ارزیابی کیفیت OSM تلقی شود و همچنین میتواند برای برنامه ریزی شهری در مناطقی که OSM فاقد داده های ساختمانی است مورد استفاده قرار بگیرد (Xie et all,2019). Längkvist و همکاران در سال(۲۰۱۶) به ارزیابی کارای شبکه عصبی عمیق برای تشخیص عوارض در پنج دسته (پوشش گیاهی، زمینی، جاده ای، ساختمان ها و آب) از تصاویر ماهواره ای پرداختند و دقت طبقه بندی با سایر کارهای طبقه بندی در هر پیکسل در سایر مناطق زمین که انتخاب مشابهی از دسته ها دارند مقایسه کردند. نتایج آن ها نشان داد طبقه بندی و تقسیم بندی کامل در بخش های انتخاب شده، که CNN ابزاری مناسب برای تقسیم بندی و تشخیص داده های سنجش ازدور است (Längkvist et all. 2016).کامل بودن داده‌های OSM در مناطق روستایی نسبت به مناطق شهری کمتر است(Zipf and zielstro,2010). ارزیابی کیفیت دادهای OSM در ایران به صورت محدود مورد مطالعه قرار گرفته است و هنوز باید بررسی‌های دقیق‌تری انجام شود. عکس‌های RGB از Google Maps به همراه داده‌های آموزشی با برچسب ضعیف به طور خودکار از داده‌های

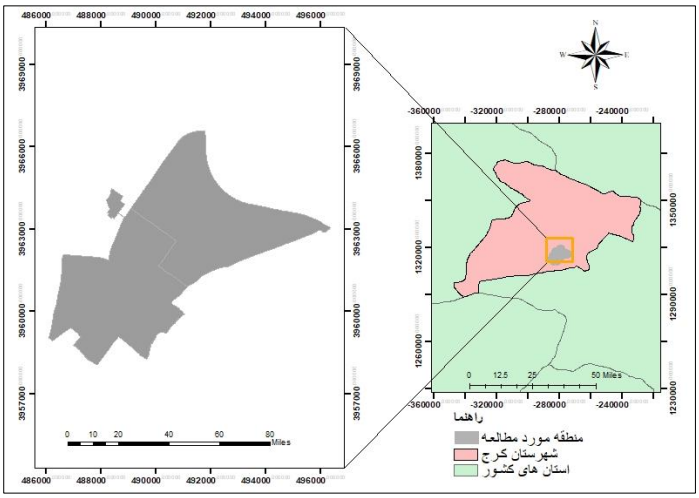
[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

Open Street Map برای شناسایی ساختمان‌ها، جنگل‌ها، زمین‌های کشاورزی و.... را به طور همزمان فراهم نمی‌کند. یکی از مشکلات داده‌های داوطلبانه استخراج اطلاعات به روش سنتی می‌باشد که از کیفیت پایینی برخوردار می‌باشند؛ لذا در این تحقیق با استفاده از هوش مصنوعی اقدام به تهیه و تکمیل داده های داوطلبانه جهت افزایش کامل بودن آنها شده است.

مواد و روش‌ها

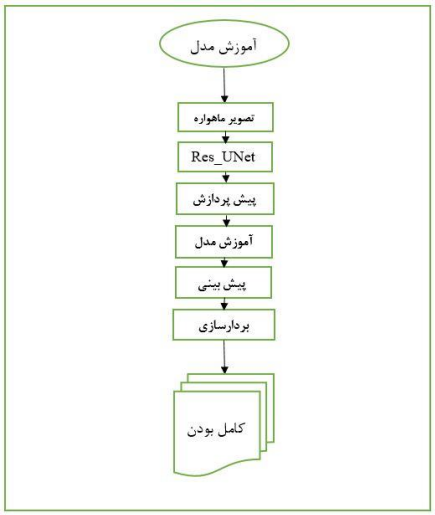
منطقه موردمطالعه

شهر کرج مرکز استان البرز، مرکز شهرستان کرج و یکی از کلان شهرهای ایران می باشد. این شهر با ۱/۵۹۲/۴۹۲تن جمعیت، چهارمین شهر پرجمعیت ایران و دومین شهر مهاجرپذیر ایران پس از تهران است. و به همین دلیل به آن لقب «ایران کوچک» داده شده است. همچنین جمعیت این شهر نسبت به سایر شهرهای بزرگ ایران جوان‌تر می باشد لذا پتانسیل بهتری جهت مشارکت داوطلبانه در تولید اطلاعات مکانی از آن انتظار می رود.



شکل ۱.نقشه موردمطالعه

روش کلی انجام تحقیق در شکل (۲) نشان‌داده شده است.



شکل ۲.فلوچارت کلی نحوه انجام تحقیق

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

مجموعه‌داده

تصاویر ماهواره‌ای

برای انجام این پژوهش، آموزش مدل و صحت سنجی مدل از تصاویر(RGB) سنجنده Google با قدرت تفکیک مکانی ۲۹/۸ سانتی‌متر مربوط به سال ۲۰۲۰ استفاده شد کاربری اراضی از روی تصویر سنجنده Goolge به صورت برداری جهت ایجاد ماسک ترسیم شدند.

برچسب‌گذاری تصاویر

بعد از مرحله ی استخراج اقدام به برچسب‌دهی عوارض شده است. تصویر ماهواره‌ای و ماسک تولید شده به‌صورت کامل قابل آموزش نیستند؛ لذا تصویر و ماسک تولید شده باید به chip هایی با اندازه معین تقسیم شوند، تصاویر به‌اندازه ۲۵۶*۲۵۶ برش داده شده است.



شکل ۳: نمونه مجموعه‌داده‌های استخراج شده. تصویر برش خورده تصویر ماهواره‌ای، ماسک دودویی داده‌های کاربری اراضی

تقسیم‌بندی مجموعه‌داده

در این تحقیق مجموعاً ۱۱۷۹۹ مجموعه داده تهیه شدکه به سه دسته ۱) آموزش مدل، ۲) اعتبار سنجی مدل، ۳) داده آزمایشی، تقسیم شدند. 80 درصد داده‌های ایجاد شده برای آموزش مدل، 10 درصد مجموعه داده برای صحت سنجی و 10 درصد برای آزمایش مدل مورد استفاده قرار گرفتند. جدول ۱. به صورت تفکیک شده تعداد مجموعه داده آموزشی ، صحت سنجی و مجموعه داده آزمایشی را نشان می‌دهد.

جدول ۱.تقسیم‌بندی مجموعه‌داده ها

مجموعه‌داده	
مجموعه‌داده آموزشی	۱۰۳۳۹
مجموعه‌داده صحت سنجی	۱۲۳۰
مجموعه‌داده آزمایشی	۱۲۳۰

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

پیش‌پردازش داده‌ها

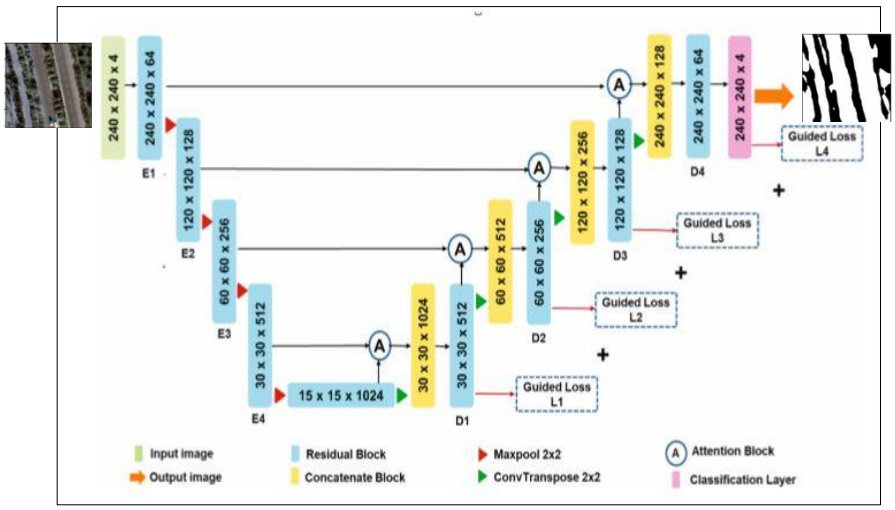
در این مرحله نرمال‌سازی و reshape داده‌ها جهت ورود به معماری مدل مورداستفاده در یادگیری عمیق و جهت آموزش مدل انجام شده است.

بردارسازی و یکسان‌سازی مجموعه‌داده

مجموعه داده‌ها قبل از عبور از مدل شبکه عصبی کانولوشنی، بردارسازی و سپس نرمال‌سازی شدند تا مقادیر پیکسل تمام تصاویر، در یک بازه مشخص قرار گیرند. سپس برای اطمینان از اندازه آن‌ها به‌اندازه ۲۵۶*۲۵۶ تغییر اندازه داده شدند و با توجه به اینکه تصاویر مورد استفاده ۸ بیتی هستند، نرمال‌سازی به صورت تقسیم هر تصویر بر ۲۵۵ انجام شد.

مدل Res_ Unet

معماری شبکه مدل Res_Unet شامل ۲۰ بلوک کانولوشن می‌باشد ۱۵ بلوک در قسمت رمزگذار و ۴ بلوک در قسمت رمزگشا همراه با یک بلوک لایه اتصال. معماری Res_Unet به صورت تدریجی لایه‌های کانولوشن را افزایش می‌دهد و اندازه تصویر را کاهش می‌دهد. لایه اول کانولوشن استفاده شده ۶۴ بود که به صورت تدریجی با افزایش لایه‌های کانولوشن به ۵۱۲ لایه کانولوشن افزایش پیدا می کند. در این مدل در قسمت رمزگذار خروجی بلوک قبلی از طریق عملیات پریدن با ورودی بلوک بعدی جمع می‌شود. قسمت رمزگذار با استفاده از لایه الحاق به صورت متناظر با بلوک متناظر خود جمع می‌شود. شکل شماره ۴ به صورت تفصیلی معماری مورد استفاده جهت آموزش شبکه عصبی را نشان میدهد.



شکل ۴- معماری مدلRes_Unet

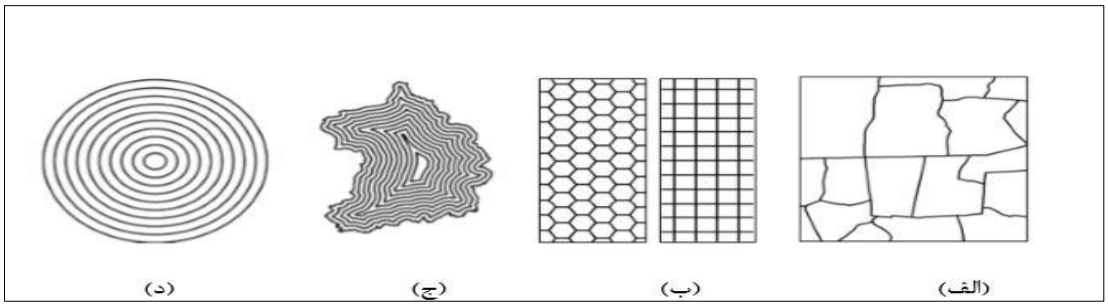
کامل‌بودن

رویکرد واحد مبنا

برخلاف مجموعه داده‌های رسمی، در مجموعه داده داوطلبانه پارامترهای کیفیت به صورت ناهمگن در سراسر منطقه تغییر می‌کند. از این رو ارزیابی پارامتر کیفیت برای کل ناحیه مورد بررسی نمی‌تواند میزان قابل قبولی باشد. برای این مشکل رویکرد واحد مبنا معرفی شده است که پارامترهای کیفیت را در یک واحد مکانی مشخص با مساحت کوچک‌تر محاسبه می‌کند. در بیشتر تحقیقات از الگوهای شبکه شش‌ضلعی منتظم و یا مرز مربوط به مناطق شهری استفاده شده

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

است(Brouelli,2016) . در پژوهش حاضر با استفاده از مرز مناطق شهری میزان کامل بودن داده‌ها مورد بررسی قرار گرفته است. از مزایای استفاده از این روش‌ها می‌توان به محاسبات کمتر و دقت بیشتر اشاره نمود.



شکل ۵: (الف) مرزهای محلی و اداری (ب)شبکه منظم (مربعی - شش‌ضلعی) (ج) حریم مرزهای شهری (د) دوایر هم مرکز



شکل ۶. روند کلی رویکرد واحد مبنا جهت ارزیابی کامل بودن

حذف خطای توپولوژیک

مرحله حذف خطاهای توپولوژی بر روی هر دو مجموعه داده OSM و رسمی باید انجام گیرد. با توجه به تمرکز این تحقیق بر داده‌های چندضلعی این خطاها عبارت‌اند از: عوارض تکراری، همپوشانی عوارض و پلیگون‌های زائد. این‌گونه خطاها باعث اختلال در نتایج ارزیابی می‌گردد؛ به همین دلیل نیاز است قبل از استفاده از داده‌ها، خطاها شناسایی و حذف شوند.

تعیین ابعاد مناسب سلول

تعیین الگو و ابعاد مناسب سلول‌ها در عملیات ارزیابی واحد مبنا فرایندی کاملاً تجربی است و رابطه ریاضی مشخصی برای آن وجود ندارد. در تحقیقات مختلف این مقدار در بازه عددی ۱۰۰ تا ۵۰۰ متر انتخاب شده است(Brouelli,2016). در این پژوهش از مرز مناطق شهری استفاده شد و نیازی به تعیین سلول نیست.

محاسبه کامل‌بودن

جهت ارزیابی میزان کامل‌بودن داده‌ها با استفاده از روش واحد مبنا اقدام به ارزیابی میزان کامل بودن داده‌ها شده است، در رویکرد واحد مبنا می‌توان از مقایسه مساحت و تعداد عوارض، کامل بودن عوارض چند ضلعی را محاسبه کرد. در ادامه هریک از معیارهای اندازه‌گیری کامل بودن به تفصیل تشریح شده‌اند (رقیه ادبی و همکاران،۱۴۰۰).

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

معیار مساحت

در بیشتر تحقیقات از شاخص نسبت مجموع مساحت عوارض در دو مجموعه داده برای محاسبه کامل بودن عوارض چندضلعی استفاده شده است(Törnros,2015) . در این روش کامل بودن داده‌های کاربری اراضی OSM نسبت به داده‌های رسمی پس از محاسبه مجموع مساحت هر عارضه از طریق معادله شماره (۱) قابل محاسبه است.

معادله شماره (۱)

$$C_{erea} = \frac{\sum Area_{osm}}{\sum Area_{Ref}} * 100$$

در رابطه (۱) $\sum Area_{OSM}$ مجموع مساحت هرعارضه OSM و $\sum Area_{OSM}$ مجموعه مساحت عارضه داده های رسمی می باشد.

برخلاف کاربرد وسیع مقایسه مساحت دو مجموعه داده در ارزیابی کامل بودن، برخی خطاها در مدل‌سازی می‌توان تأثیر نادرست بر ارزیابی نتایج داشته باشند. شکل (۷) نمونه‌ای از خطای تأثیرگذار بر معیار نسبت مساحت را نشان می‌دهد. برای مثال ترسیم ساده ساختمان‌های پیچیده در OSM منجر به سطح بالاتری از کامل بودن می‌شود. این خطا در مدل‌سازی توسط دو مفهوم مثبت صحیح [۱] (مساحت عوارض مرجع که به درستی در OSM ترسیم شده‌اند)، منفی کاذب[۲](مساحت عوارض مرجع که در OSM ترسیم نشده‌اند) بررسی می‌گردد.

بر اساس دو مفهوم ذکر شده، می‌توان تخمین دقیقی از پارامترهای کیفیت را به‌صورت معادله شماره(2) ،(3) محاسبه کرد.

معادله شماره (۲)

$$TP\frac{TP}{TP + FN} * 100$$

معادله شماره (۳)

$$FN\frac{FN}{TP + FN} * 100$$

نرخ مثبت صحیح (TP) ، نمایانگر صحت طبقه‌بندی مجموعه داده OSM است. نرخ منفی کاذب(FN) ، نشانگر داده‌هایی است که در OSM موجود نیست.



شکل ۷.نمونه‌ای از خطای تأثیرگذار بر معیار مساحت

[1] True Positive
[2] False Negetive

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

معیار تعداد عوارض

کامل بودن داده‌های چند ضلعی را می‌توان از نسبت تعداد عوارض چند ضلعی مجموعه داده داوطلبانه به رسمی بررسی کرد(Fram,2015) . در این روش پس از تقسیم ناحیه مورد بررسی به زیر نواحی کوچک‌تر، کامل بودن داده‌های کاربری اراضی OSM نسبت به داده‌های رسمی از طریق رابطه (۴) محاسبه می‌گردد.

معادله شماره (۴)

$$C_{ecount} = \frac{\sum T_{osm}}{\sum T_{Ref}} * 100$$

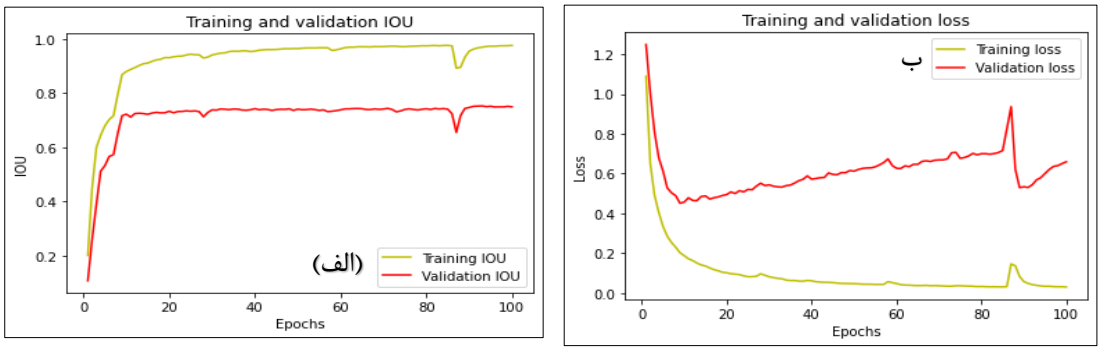
مجموعه تعداد هر عارضه در $\sum T_{Ref}$ و OSM مجموعه تعداد هر عارضه در $\sum T_{osm}$در رابطه(۴) داده‌های رسمی است.

نتایج و بحث

در تحقیق حاضر ابتدا کاربری اراضی با استفاده از هوش مصنوعی تولید و دقت آن محاسبه و سپس با داده‌های تولید شده توسط داوطلبان در سایت OSM جهت ارزیابی میزان تکمیل بودن این داده‌ها مقایسه شد.

نمودار تغییرات میزان خطا و میزان IOU دو مجموعه آموزش و اعتبارسنجی

در هوش مصنوعی مدلی کارآمد است که میزان خطا و میزان دقت مجموعه داده آموزشی و مجموعه داده اعتبارسنجی منطبق بر هم باشند. روند تابع زیان داده‌های آموزشی و مجموعه داده اعتبارسنجی روند کاهشی به خود بگیرد و میزان دقت داده‌های آموزشی و اعتبارسنجی روند افزایشی داشته باشند. مدل Res_unet دارای دو مرحله می‌باشد. مرحله اول پیش‌بینی عوارض تصویر و مرحله دوم پیش‌بینی با نمودار. در مرحله پیش‌بینی با تصویر همه عوارض به طور مطلوبی شناسایی شده و برای ارزیابی میزان کامل بودن قابل قبول می‌باشند، اما به لحاظ نمودار معمولاً دچار overfit می‌شود. یکی از راه‌های جلوگیری از overfit شدن استفاده از داده‌های آموزشی زیاد است. داده زیاد به مدل کمک می‌کند شناخت بهتری از سیگنال ورودی به دست آورد. داده خوراک الگوریتم ماشین است و هرچقدر بیشتر یعنی دانش بیشتر و شناخت بیشتر مدل از مسئله، داده زیاد به مدل کمک می‌کند تا درک بهتری از سیگنال ورودی به دست آورد. البته باید مدام تابع زیان و دقت مورد نظارت قرار گیرد گاهی داده‌های زیاد هم باعث بیش برازش شده و سیستم دیداش محدود به داده‌های آموزشی می‌شود و جنرالایز خوبی به دست نمی‌دهد.

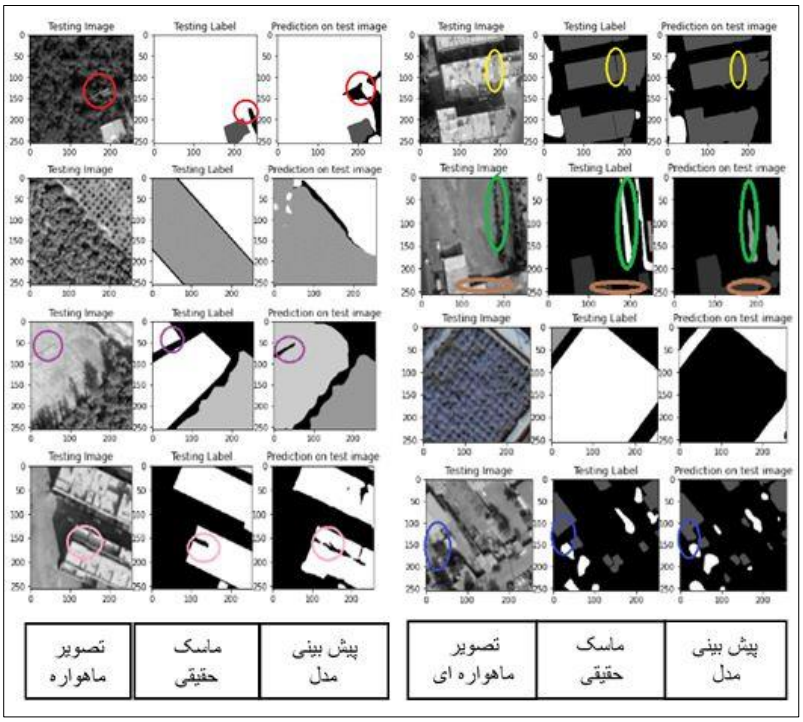


شکل ۸.(الف)نمودار تابع دقت مدل و (ب) نمودار تابع زیان مدل Res_Unet

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

بررسی بصری نتایج

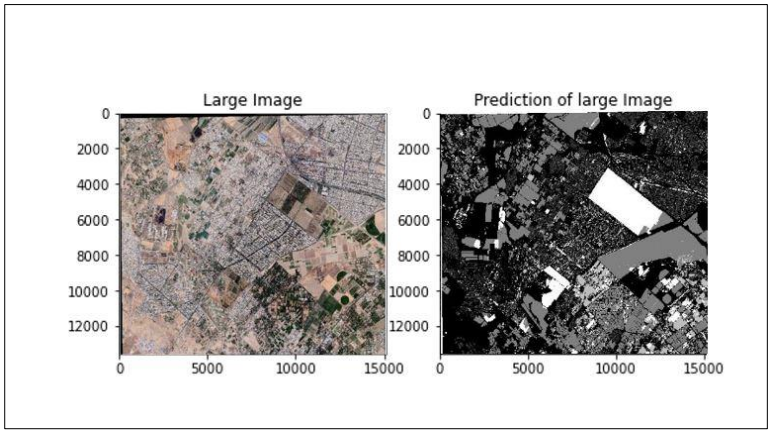
به منظور بررسی بصری نتایج حاصل از ارزیابی صحت مدل نمونه‌های پیش‌بینی شده برای مجموعه داده آزمون بررسی شدند. با توجه به بررسی بصری که انجام شد، در داده های تهیه شده برخی عوارض ترسیم نشده‌اند؛ اما مدل توانسته به خوبی عوارض را تشخیص دهد و مرز آن‌ها را تعیین کند. نواحی که با رنگ قرمز نشان داده شده‌اند،کاملاً به عنوان یک کلاس در نظر گرفته شده بودند، اما مدل به خوبی عوارض را از یکدیگر تفکیک کرده، نواحی که با رنگ زرد نشان داده شده‌اند هیچ مرزی برای ساختمان وجود نداشته و مدل این را به خوبی تشخیص داده و مرز را برداشته و همینطور نواحی که با رنگ صورتی و بنفش نشان داده شده‌اند، نشان دهنده مطلوب عمل‌کردن مدل می‌باشند. نواحی به رنگ آبی، سبز و نارنجی نیز عوارضی می‌باشند که در داده های آموزش ترسیم نشده‌اند؛ اما مدل به خوبی عوارض را شناسایی کرده است و این نشان‌دهنده این است که مدل با دقت بالایی آموزش دیده است.



شکل ۹.بررسی بصری مدل Res_unet پیش‌بینی مدل در تصویر بزرگ

جهت ارزیابی میزان کامل بودن داده‌های OSM نیاز به خروجی با اندازه منطقه موردمطالعه است و از آنجا که CNN به دلیل محدودیت‌های محاسباتی فعلی که دارد نمی‌توان تصویر بزرگ را جهت پیش‌بینی وارد شبکه عصبی کانولوشن کرد؛ لذا با توجه به این محدودیت تصویر منطقه موردمطالعه به اندازه‌های ۲۵۶*۲۵۶ برش داده شد و برای پیش‌بینی به مدل وارد شد و بعد از پیش‌بینی مدل و حذف خطاهای توپولوژیک در کنار یکدیگر چسبانده شدند، تا بتوان دید کلی از منطقه داشت.

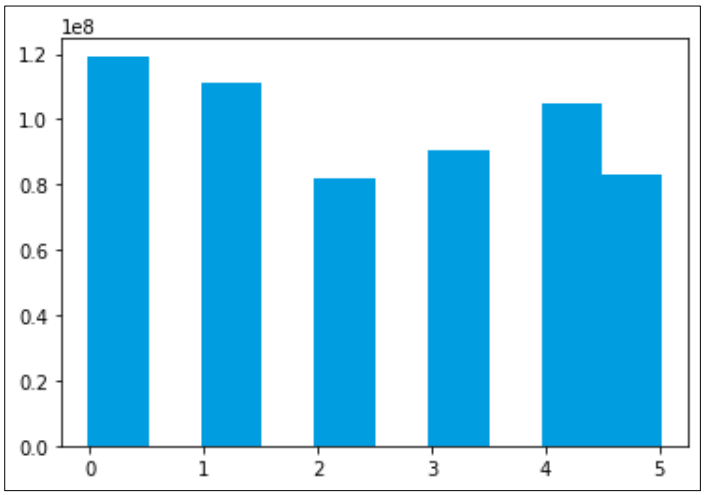
[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]



شکل ۱۰-تصویر، سمت چپ تصویر ماهواره‌ای بزرگ، سمت راست نتیجه پیش‌بینی

بررسی دقت تشخیص عوارض

همان‌طور که از نمودار مشخص می شود میزان تشخیص عوارض به خوبی صورت‌گرفته است. بهترین تشخیص برای مناطق بدون عارضه، بلوک‌های ساختمانی و زمین‌های کشاورزی می‌باشد. تشخیص سایر عوارض هم در حد مطلوبی است. و می‌توان ارزیابی خوبی از کاربری اراضی را با مدل Res_Unet داشت.



شکل ۱۱. نمودار میله‌ای میزان صحت سنجی عوارض کامل بودن بر اساس رویکرد واحد مبنا

روش واحد مبنا در دو مرحله انجام می‌شود،1.معیار مساحت2.معیار تعداد عوارض. جهت انجام روش واحد مبنا ابتدا باید منطقه مورد مطالعه به زیر نواحی تقسیم شود، تقسیم منطقه مورد مطالعه به روش‌های مختلفی صورت می‌گیرد. همان طور که گفته شد، در این پژوهش از مرز مناطق شهری برای تقسیم بندی منطقه مورد مطالعه استفاده شد.

کامل بودن بر اساس معیار مساحت

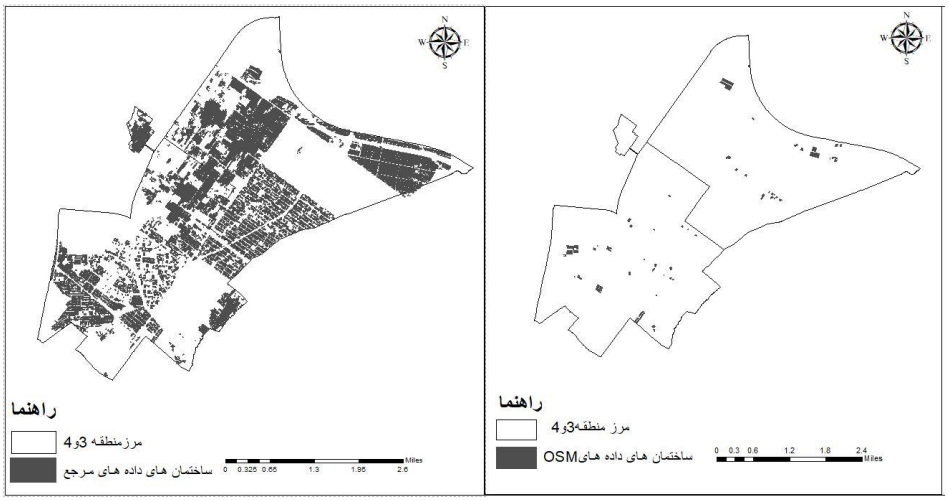
در این روش کامل بودن عوارض بر اساس نسبت مساحت عوارض داوطلبانه به مساحت عوارض رسمی در هر ناحیه محاسبه شد. نتایج حاصل از محاسبه کامل بودن عوارض بر اساس مرز مناطق شهری می‌باشد. همان‌طور که در شکل (۱۲) مشاهده می‌شود داده‌های رسمی نسبت به داده‌های OSM بیشترین مقدار کامل بودن را دارند. جهت ارزیابی

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

دقیق‌تر از رابطه (۱) برای برآورد میزان کامل بودن از طریق مساحت اقدام شد و نتایج نشان داد که بلوک‌های ساختمانی ۳/۶، جنگل ۹/۷، درختان میوه ۹۰/۴ و زمین‌های کشاورزی ۸۱/۸۸ درصد از کامل بودن را نشان می‌دهد.



[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]



شکل ۱۲. مقایسه مساحت داده‌های OSM و مجموعه‌داده‌های مرجع

همان‌طور که بیان شد، برخلاف کاربرد وسیع مقایسه مساحت دو مجموعه داده در ارزیابی کامل بودن، برخی خطاها در مدل‌سازی می‌تواند تأثیر نادرست بر نتایج ارزیابی داشته باشد. در نتیجه با استفاده از دو مفهوم نرخ مثبت صحیح (صحت شناسایی عوارض در OSM)، نرخ منفی کاذب (نشانگر داده‌های است که در OSM موجود نیست) می‌توان تخمین دقیق‌تری از کامل بودن را برآورد کرد. نتایج حاصل از این دو مفهوم در جدول ۲ نشان داده شده است. نتایج نشان می‌دهد که داده‌های اضافی (FP) در OSM موجود نیست همچنین نرخ مثبت صحیح در ساختمان ۳/۸ ، جنگل‌ها ۷/۶، درختان میوه،۹۰/۱۱، زمین‌های کشاورزی ۸۱/۵۰ می‌باشد. نرخ منفی کاذب نشانگر داده‌های است که در OSM موجود نیست. نرخ منفی کاذب در بلوک‌های ساختمانی ۹۶/۱۱، جنگل‌ها ۹۲/۳۸، درختان میوه ۹/۸۸ و زمین‌های کشاورزی ۱۸/۴۹می‌باشد. نتایج نرخ پایین کامل بودن بلوک‌های ساختمانی و جنگل و نرخ بالای کامل بودن زمین‌های کشاورزی و درختان میوه را نشان می‌دهد.

جدول ۲. میزان کامل بودن داده‌های OSM Landuse با استفاده از مثبت صحیح و مثبت کاذب

	ساختمان	جنگل	درختان میوه	زمین کشاورزی
TP	۳/۸	۷/۶۱	۹۰/۱۱	۸۱/۵۰
FP	۹۶/۱۱	۹۲/۳۸	۹/۸۸	۱۸/۴۹

کامل بودن با استفاده از معیار تعداد عوارض

در پژوهش حاضر، نتایج حاصل از معیار نسبت تعداد عوارض ($\frac{\text{تعداد در داده های } \boldsymbol{OSM}}{\text{تعداد در داده های مرجع}}$) بیانگر نرخ پایین کامل بودن بلوک های ساختمانی و جنگل و نرخ بالای کامل بودن زمین های کشاورزی و درختان میوه را نشان می دهد.

جدول ۳ میزان کامل بودن با استفاده از معیار تعداد عوارض

ساختمان	جنگل	درختان میوه	کشاورزی
۵/۷٪	۲٪	۷۲٪	۷۵٪

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

نتیجه‌گیری

پایگاه‌داده OSM دارای جنبه‌های مثبتی بوده که می‌توان به در دسترس بودن اطلاعات، قابلیت به‌روزرسانی و هزینه پایین دسترسی به اطلاعات اشاره کرد. با این‌حال کیفیت این داده‌ها به عنوان یک چالش مهم می‌باشد. در این پژوهش کامل بودن که یکی از روش‌های ارزیابی کیفیت داده‌های OSM است مورد بررسی قرار گرفت. با توجه به نرخ مثبت صحیح، ساختمان‌ها ٪۳/۸، جنگل ٪۷/۶۱، درختان میوه ٪۹۰/۱۱، زمین‌های کشاورزی ۸۱/۵۰ درصد می‌باشند. نتایج نشان می‌دهد جنگل‌ها، ساختمان‌ها، از صحت شناسایی کمتری نسبت به درختان میوه و زمین‌های کشاورزی برخوردار هستند. با این وجود کیفیت برچسب‌دهی عوارض ضعیف است، با توجه به این که عوارض درست شناسایی شده‌اند؛ اما مرزها به درستی ترسیم نشده‌اند و این نشان دهنده عدم تخصص کاربران در تولید داده‌های داوطلبانه می‌باشد. از هوش مصنوعی جهت کمک به انسان در تکمیل داده‌های داوطلبانه مکانی جهت تولید داده‌های با کیفت بالا از مدل شبکه عصبی Res_Unet استفاده شد. طبق نتایج این تحقیق میزان شاخص IOU روش Res_Unet ۰/۸۳۰ می‌باشد. همچنین نتایج بصری این تحقیق نشان داد، مدل Res_Unet توانسته عوارضی که در ماسک حقیقت زمینی ترسیم نشده‌اند یا مرز بین آن‌ها به‌اشتباه ترسیم شده است را به درستی تشخیص دهد که نشان دهنده توانایی بالای آن در تولید و تکمیل اطلاعات مکانی می‌باشد. نتایج حاصل از این مطالعه نشان داده که درختان میوه ۹۰/۴ درصد و زمین‌های کشاورزی ۸۱/۸۸ درصد از میزان کامل بودن بیشتری نسبت به بلوک‌های ساختمانی با ۳/۶ و جنگل‌ها با ۹/۷ درصد برخوردار می‌باشند. در نهایت داده‌های تولید شده باهوش مصنوعی به داده‌های داوطلبانه OSM اضافه شده و با دقت به دست آمده داده تقریباً کاملی در اختیار داوطلبان جهت استفاده قرار می‌دهد.

پیشنهادات

با توجه به بررسی‌های انجام شده در تحقیق حاضر، داده‌های داوطلبانه از کیفیت پایینی برخوردار هستند؛ لذا نیاز است با دقت بیشتری و توسط افراد متخصص کامل شوند، و تحقیقات بیشتری در زمینه کیفیت داده‌ها صورت گیرد و اصلاحاتی در داده های تکمیل شده انجام پذیرد. مدل‌های کانولوشن علاوه بر نوین بودن از دقت بالای برخوردار هستند. در نتیجه با استفاده از این روش‌ها می‌توان عارضه‌ها را شناسایی و نسبت به کامل بودن آن‌ها اقدام کرد.

منابع

۱. ادبی، رقیه؛ رحیم علی عباسپور، علیرضا چهرتان. ۱۴۰۰.سنجش میزان کامل‌بودن داده‌های مکانی داوطلبانه با تمرکز بر داده‌های بلوک‌های ساختمانی (مطالعه موردی: کلان‌شهر تهران). نشریه علمی مهندسی فناوری اطلاعات مکانی:۴۶

2. Brovelli,M.;M.Minghini,M. Molinari, and G. Zamboni," Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection:The method and a case study." 2016.

3. Chen,J.; Y.Zhou,A.Zipf,and H.Fan. 2018.Deep learning from multiple crowds:A case study of humanitarian mapping.IEEE Transactions on Geoscience and Remote Sensing,57(**3**),1713-1722.

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]

4. Cooper,A.;S.Coetzee,D.kourie,and I.kaczmarek.2012.Volunteered geographical information-the challenges.
5. Cihlar,J.Land cover mapping of large areas from satellites: Status and research priorities.2000
6. Fram,C.; K.Chistopoulou,and C.Ellul,'Assessing the quality of Open street map building data and searching for a proxy variable to estimate OSM building data completeness,'In: proceedings of the 23 rd GIS Research UK (GISRUK) conference, Leeds,UK,PP.17-15,2015.
7. Feick,R.;and S.Roche. 2013.Understandi g the Value of VGI. In Crowdsourcing geographi knowledge(pp.15 2).Springer Dordrecht.
8. Goodchild,M.;L.Li. 2012.Assuring the quality of volunteered geographic information Spatial statistics,1,110-120.
9. Hashemi,P.;R. Ali Abbaspour. 2015.Assment of logical consistency in Open street map based on the spatial similarity concept.In OpenStreet Map in giscience(pp.19-36).Springer,Cham.
10. Haklay,M. 2010.How good is volunteered geographical information? A comparative study,OpenSteetMap Ordnance Survey datasets.Environment and planning B:Planning and design,37(**4**),682-703.
11. Hecht, R.;C.Kunze, and S.Hahmann. 2013. Measuring completeness of building footprints in OpenStreetMap over space and time. ISPRS International Journal of Geo-Information, 2(4), 1066-1091.
12. LeCun,Y.;Y.Bengio,G. Hinton.2015.Deep learning.Nature521.436-444.
13. Längkvist, M., Kiselev, A., Alirezaie, M., & Loutfi, A. (2016). Classification and segmentation of satellite orthoimagery using convolutional neural networks. *Remote Sensing*, *8*(4), 329.
14. Molinar,E., i, and Zamboni,G., " Positional accuracy assessment of the OpenStreetMap buildings layer.
15. Peng,Z.;M.Tsou,M.Goodchild.2003. Internet,GIS:distributed geographic information services for the internet and wireless networks. John Wiley & Sons.
16. Törnros,T.;H.Dorn,S.Hahmann,andZipf,A.2015.Uncertainties.of completeness,measures,in, OpenStreetMap–A case study for buildings in a medium-sized German city. ISPRS ann als of the photogrammetry, remote sensing and spatial information sciences, 2, 353.
17. Wang, M.;Q.Li,Q.Hu,M. Zhou.2013.Quality analysis of open street map data. International archives of the photogrammetry, remote sensing and spatial information sciences, 2(W1).
18. Xie, X., Zhou, Y., Xu, Y., Hu, Y., & Wu, C. (2019). OpenStreetMap data quality assessment via deep learning and remote sensing imagery. IEEE Access, 7, 176884-176895.
19. Zhang, C.; I.Sargent,X. Pan, H.Li, A. Gardiner, J. Hare, and P. Atkinson. 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote sensing of environment, 216, 57-70.
20. Zheng, S.;and J. Zheng. 2014. Assessing the completeness and positional accuracy of OpenStreetMap in China. In Thematic cartography for the society (pp171-189).Springer,Cham.
21. Zielstra, D.;and A. Zipf.2010.A comparative study of proprietary geodata and volunteered geographic information for Germany. In 13th AGILE international conference on geographic information science (Vol. 2010, 1-15).

[Downloaded from c4i2016.khu.ac.ir on 2026-10-30]
[DOR: 20.1001.1.24237892.1403.11.1.1.1]
[DOI: 10.61186/jsaeh.11.1.3]