نشریه علمی مهندسی فناوری اطلاعات مکانی

سال نهم، شماره سوم، پاییز ۱۴۰۰
Vol.9, No.3, Autemn 2021
۱۰۹ -۱۳٤
مقاله پژوهشی
DOR: 20.1001.1.20089635.1400.9.3.6.7

ارزیابی و مقایسه عملکرد شبکه‌های عصبی عمیق *FCN* و *RDRCNN* به‌منظور شناسایی و استخراج عارضه راه شهری با استفاده از تصاویر سنتینل-۲ با قدرت تفکیک مکانی متوسط

سیدهدایت شیخ‌قادری[1*]، پرویز ضیائیان فیروزآبادی[2]، منوچهر کلارستاقی[3]

۱- کارشناس ارشد سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشکده علوم جغرافیایی، دانشگاه خوارزمی تهران
۲- دانشیار گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشکده علوم جغرافیایی، دانشگاه خوارزمی تهران
۳- استادیار گروه آموزشی مهندسی برق و کامپیوتر، دانشکده فنی و مهندسی، دانشگاه خوارزمی تهران

تاریخ دریافت مقاله: ۱۴۰۰/۰۵/۰۷ تاریخ پذیرش مقاله: ۱۴۰۰/۰۹/۲۲

چکیده

استخراج عارضه راه با استفاده از تصاویر سنجش‌ازدور طی سال‌های گذشته به‌عنوان یکی از موضوعات جذاب، موردتوجه محققین بوده است. اخیراً پیشرفت و توسعه شبکه‌های عصبی عمیق *(DNN)* در بخش تقسیم‌بندی معنایی به یکی از روش‌های مهم استخراج راه تبدیل‌شده است. در این میان اکثریت تحقیقات انجام‌شده در زمینه استخراج عارضه راه با بهره‌گیری از *DNN* در مناطق شهری و غیرشهری با استفاده از تصاویر با قدرت تفکیک مکانی بالا انجام‌گرفته است. در این تحقیق برای اولین بار جهت استخراج عارضه راه با استفاده *DNN*، از تصاویر با وضوح مکانی متوسط سنجنده سنتینل-۲ بهره گرفته شد، به این صورت که از تصویر شهر تهران به‌عنوان داده تست و از ۷ شهر دیگر (مشهد، اصفهان، شیراز، تبریز، کرمانشاه، ارومیه و بغداد) به‌عنوان داده‌های آموزش و اعتبارسنجی، استفاده گردید. در این میان، پس از آماده‌سازی و برچسب‌زنی همه پیکسل‌های مربوط به عارضه راه، تصاویر به قطعات ۲۵۶ در ۲۵۶ پیکسل تبدیل و پس از جداسازی قطعات نامناسب، برای داده‌های تست، آموزش و اعتبارسنجی به ترتیب ۱۳۵، ۱۵۰۰ و ۱۰۰ قطعه تصویر به دست آمد. درنهایت برای آموزش و استخراج عارضه راه، از شبکه‌های عصبی کانولوشن باقیمانده عمیق پالایش‌شده *(RDRCNN)* و *U-Net* که مبتنی بر شبکه‌های کاملاً کانولوشن *(FCN)* است، استفاده شد. نتایج حاصله حاکی از آن است که هر دو مدل *RDRCNN* و *FCN* در مقایسه با داده‌های واقعیت زمینی شبکه راه شهری تهران را از تصاویر سنتینل-۲ به‌خوبی شناسایی و استخراج کرده‌اند. در این میان مدل *FCN* هم ازنظر بصری و هم ازنظر متریک‌های ارزیابی صحت نسبت به مدل *RDRCNN* عملکرد بهتری داشته، به‌طوری‌که برای مدل *FCN*، معیارهای کامل‌بودن ۸۲,۹۲٪، صحت ۷۷,۶۷٪، امتیاز *F1* ۸۰,۲۰٪ و دقت کلی ۹۶,۳۰٪ و برای *RDRCNN* معیارهای کامل‌بودن ۸۰,۴۳٪، صحت ۷۱,۳۷، امتیاز *F1* ۷۷,۷۴٪ و دقت کلی ۹۵,۷۱٪ به دست آمد. به‌طورکلی یافته‌های این پژوهش پتانسیل استفاده از روش‌های *DNN* برای استخراج عارضه‌ی راه شهری با بهره‌گیری از تصاویر باقدرت تفکیک مکانی متوسط سنتینل ۲ را نشان می‌دهد.

کلید واژه‌ها : شبکه‌های عصبی عمیق، استخراج راه، *FCN*، *RDRCNN*، سنتینل-۲.

* نویسنده مکاتبه کننده: گروه سنجش‌ازدور و سیستم اطلاعات جغرافیایی، دانشکده علوم جغرافیایی، دانشگاه خوارزمی تهران.
تلفن: ۰۹۳۸۷۹۵۹۵۰۸

Email: hedayatp90@gmail.com

۱- مقدمه

امروزه برای انجام هرگونه برنامه‌ریزی و فعالیت اجرایی داشتن بانک اطلاعات مکانی و توصیفی بسیار مهم تلقی و داشتن منبع عظیمی از اطلاعات مکانی و بروز نمودن آن‌ها امری حیاتی است. در این میان، روش‌های تولید و استخراج اطلاعات مکانی در حوزه ژئوماتیک و در رأس آن‌ها فتوگرامتری و سنجش‌ازدور بیشترین سهم در تولید نقشه و اطلاعات مکانی بر عهده دارند. در میان مجموعه عوارض موجود در هر بانک اطلاعات مکانی، عارضه راه یکی از مهم‌ترین منابع اطلاعات مکانی محسوب می‌شود. راه‌ها در بسیاری از کاربردها ازجمله حمل‌ونقل، سیستم‌های هدایت خودکار وسایل نقلیه، مدیریت ترافیک، واکنش‌ها در مواقع وقوع بحران مورداستفاده قرار می‌گیرند *[۱، ۲ و ۳]*. شبکه‌های حمل‌ونقل شهری به‌طور گسترده به سمت توسعه در جریان هستند؛ بنابراین، نیاز مبرم به روش استخراج راه دقیق و سریع برای به هنگام کردن نقشه راه‌ها وجود دارد. روش‌های نقشه‌برداری سنتی پرزحمت و زمان‌گیر هستند *[۴]*. در سال‌های اخیر، روش‌های مختلفی به‌منظور استخراج عارضه راه توسعه‌یافته است. روش‌های پیشرفته را می‌توان به سه دسته معمول تقسیم‌بندی کرد: (۱) استخراج مرز. (۲) شناسایی اشیاء / طبقه‌بندی؛ و (۳) قطعه‌بندی تصویر *[۵، ۶ و ۷]*. همچنین در منابع دیگر انواع روش‌های استخراج راه را به روش‌های تعقیب راه، مورفولوژی، طبقه‌بندی و ناحیه‌بندی، منحنی فعال، چندمقیاسی و چند تفکیکی، آنالیز چندزمانه، تجزیه‌وتحلیل استرئوسکپی و غیره تقسیم‌بندی کرده‌اند *[۳]*. درواقع هرکدام از روش‌های یادشده بسته به نوع تصویر ورودی دارای معایبی هستند که به‌طورکلی می‌توان به محاسبات سنگین و زمان‌بر، اتوماتیک نبودن و وابستگی بیش‌ازحد به کاربر، نیاز به پس پردازش، عدم استفاده از ویژگی‌های هندسی تصاویر، کارایی پایین و گاها عدم کارایی آن‌ها بر روی تصاویر با قدرت تفکیک مکانی پایین، کمتر استفاده شدن آن‌ها در مناطق پیچیده‌ی شهری، عدم شناسایی برخی از قطعات راه‌ها و غیره اشاره نمود. در همین راستا، ظهور روش‌های یادگیری عمیق، در کنار دقت، سرعت پردازش بالا، آموزش خودکار ویژگی‌ها به‌صورت نظارت‌نشده، زمان آموزش پایین و کاهش حجم محاسبات، نتایج قابل قبولی نیز از خود نشان داده‌اند. یادگیری عمیق، جزء آن دسته از روش‌هایی است که با استفاده از مدل‌های مختلف، اقدام به مشخص کردن موقعیت عوارض موجود در تصاویر می‌کند. این روش، مفاهیم انتزاعی سطح بالا اشیاء را از تصاویر، دریافت و آن‌ها را به لایه‌هایی با مفاهیم سطح پایین‌تر تبدیل می‌کنند *[۸]*. با پیشرفت اخیر در هوش مصنوعی، به‌خصوص فن‌آوری یادگیری عمیق، ایده های جدیدی برای استخراج راه ها فراهم‌شده است [۹]. درواقع یادگیری عمیق از تحقیقات شبکه‌های عصبی توسعه داده‌شده و پیشرفت‌های قابل‌توجهی در بینایی رایانه‌ای و زمینه‌های دیگر داشته است [۱۱ و ۱۰]. یک از این موارد روش قطعه‌بندی معنایی است. این روش باهدف برچسب‌گذاری هر پیکسل با احتمال مالکیت در طبقات مختلف انجام می‌شود که کار چالش‌برانگیزی است [۱۲]. در این میان تحقیقات زیادی به‌منظور استخراج عارضه راه با استفاده از روش‌های یادگیری عمیق انجام‌گرفته است که در ادامه به تعدادی از مهم‌ترین آن‌ها پرداخته خواهد شد. گائو و همکاران (۲۰۱۹) در تحقیقی یک روش جدید برای استخراج راه‌ها از تصاویر ماهواره‌ای اپتیک با استفاده از یک شبکه عصبی کانولوشن باقیمانده عمیق پالایش‌شده (*RDRCNN*[۱]) با یک مرحله پس پردازش پیشنهاد کردند. *RDRCNN* پیشنهادی آن‌ها شامل یک واحد متصل باقی‌مانده (*RCU*[۲]) و یک واحد ادراک متسع‌شده (*DPU*[۳]) بود.

[۱] *Refined Deep Residual Convolutional Neural Network*

[۲] *Residual Connected Unit*

[۳] *Dilated Perception Unit*

[DOR: 20.1001.1.20089635.1400.9.3.6.7]
[DOI: 10.52547/jgit.9.3.109]

آن‌ها در تحقیق خود از مورفولوژی ریاضی و یک الگوریتم رأی‌دهی تانسور برای بهبود عملکرد *RDRCNN* در طول پس پردازش استفاده کردند. آزمایش‌ها بر روی دو مجموعه داده از تصاویر با وضوح‌بالا انجام گرفت تا عملکرد معماری‌های شبکه پیشنهادی را نشان داده و نتایج معماری های آن با معماری‌های شبکه دیگر مقایسه گردد که نتایج نشان‌دهنده کارایی مؤثر روش پیشنهادی برای استخراج راه‌ها از یک محیط پیچیده بود [۱۳]. لی و همکاران (۲۰۱۹) در مقاله‌ای، روشی به‌منظور بهبود شبکه‌های عصبی برای استخراج راه‌ها از تصاویر سنجش‌ازدوری پهپاد را پیشنهاد دادند. آن‌ها با تمرکز بر روی مسئله کارایی محاسباتی پایین *D - LinkNet*، به پیشرفت‌هایی رسیدند. درنهایت با توجه به برخی تغییرات، شبکه عصبی بهبودیافته دیگری بانام *B - D - LinknetPlus* ساخته و مقایسه بین شبکه‌های عصبی انجام دادند. نتایج نهایی نشان داد که شبکه‌های عصبی بهبودیافته در کاهش اندازه شبکه و ایجاد دقت موردنیاز برای استخراج راه مثمر ثمر است [۱۴]. سین و همکاران (۲۰۱۹) یک روش جدید برای استخراج شبکه جاده از تصاویر سنجش‌ازدور با استفاده از یک مدل *Dense UNet* با چند پارامتر و ویژگی‌های قوی ارائه کردند.

Dense UNet پیشنهادی آن‌ها از واحدهای اتصال متراکم تشکیل‌شده بود که بااتصال در لایه‌های مختلف شبکه، ادغام مقیاس‌های مختلف را تقویت می‌کرد. عملکرد روش پیشرفته در دو مجموعه داده با وضوح‌بالا در مقایسه با سه قطعه‌بندی معنایی کلاسیک تائید شده ارزیابی شد. به‌طورکلی نتایج تجربی حاکی از آن بود که روش فوق می‌تواند برای استخراج راه در مناطق پیچیده مورداستفاده قرار گیرد [۱۵]. ژو و همکاران (۲۰۱۸) مقاله‌ای با عنوان استخراج راه از تصاویر ماهواره‌ای با قدرت تفکیک بالا با استفاده یادگیری عمیق به انجام رساندند. مدل پیشنهادی باهدف استخراج اطلاعات محلی و جهانی جاده‌ها در تصاویر سنجش‌ازدور و بهبود دقت استخراج شبکه راه‌ها انجام گرفت. به‌طورکلی نتایج تحقیق نشان‌دهنده آن بود که مدل پیشنهادی آن‌ها یعنی *GL - U - U – Net* خوب عمل کرده و راه‌هایی که با مقیاس مختلف در تصاویر سنجش‌ازدور برچسب‌زنی شده بودند، با موفقیت از طریق شبکه عصبی عمیق کانولوشن استخراج‌شده‌اند. [۱۶]. هی و همکاران (۲۰۱۹) در تحقیقی با تلفیق حداکثر تجمع هرم مکانی[۱] *(ASPP)* با شبکه رمزگذار - رمزگشای[۲]، عملکرد شبکه استخراج جاده را بهبود بخشیدند. روش پیشنهادی از توانایی *ASPP* در استخراج ویژگی‌های چند مقیاس و توانایی شبکه رمزگذار - رمزگشای برای استخراج ویژگی‌های دقیق بهره می‌برد. آن‌ها برای اولین بار، از شباهت ساختاری[۳] (*ssim*) به‌عنوان یک تابع زیان برای استخراج راه استفاده کردند. نتایج تجربی با استفاده از مجموعه داده جاده ماساچوست نشان داد که روش پیشنهادی آن‌ها امتیاز F_1 ٪۸۳٫۵ و *ssim*٪۸۹۳ را کسب کرده است که در مقایسه با *U-net* نرمال، روش آن‌ها امتیاز F_1 ۲٫۶- و *ssim* را ۱۸٪ بهبود بخشیده است. به‌طورکلی نتایج روش پیشنهادی آن‌ها راه‌ها را از تصاویر سنجش‌ازدور به‌طور مؤثرتر و روشن‌تر از دیگر روش‌های مورد مقایسه استخراج کرد [۱۷]. ژانگ و همکاران (۲۰۱۹) در مقاله‌ای، یک روش استخراج راه جدید از تصاویر هوایی بر اساس یک شبکه معکوس تولیدی بهبودیافته[۴] ارائه کردند که روش فوق دارای یک چارچوب پایان به پایان و تنها نیازمند چند نمونه برای آموزش بود. نتایج تجربی در مورد مجموع داده جاده‌های ماساچوست نشان داد که روش پیشنهادی، عملکرد بهتری نسبت به چندین حالت از تکنیک‌های هنری درزمینه‌ی دقت تشخیص، کامل‌بودن، صحت و امتیاز F_1 ارائه

[۱] *Atrous spatial pyramid pooling*

[۲] *Encoder - Decoder*

[۳] *structural similarity(ssim)*

[۴] *improved generative adversarial network*

می‌دهد[۱۸]. سنتی‌لنز و همکاران (۲۰۲۰) در پژوهشی یک روش جدید متمرکز که شامل یک یادگیری انتقال عمیق با طبقه‌بندی گروهی[۱] *(TEC)* بود برای استخراج راه با استفاده از تصاویر پهپاد استفاده کردند. *TEC* عمیق پیشنهادی، استخراج راه بر روی تصاویر پهپاد را در دو مرحله انجام می‌داد، یعنی یادگیری انتقال عمیق و طبقه‌بندی گروهی. روش *TEC* عمیق دارای کیفیت متوسط ٪۷۱ به دست آمد که ٪۱۰ بالاتر از بهترین روش‌های استاندارد یادگیری عمیق بود. همچنین *TEC* عمیق سطح بالاتری از معیارهای ارزیابی مانند معیارهای کامل‌بودن، صحت و امتیاز F_1 را نشان داد. به‌طورکلی نتایج به‌دست‌آمده حاکی از آن بود که *TEC* عمیق در استخراج شبکه‌های جاده‌ای در یک منطقه شهری کارآمد است[۱۹]. عبدالهی و همکاران (۲۰۲۰) در تحقیقی بر اساس چهار نوع اصلی از روش‌های یادگیری عمیق، یعنی مدل *gans*، شبکه‌های عمیق کانولوشن[۲] *(CNN)*، شبکه کاملاً کانولوشنال یا پیچشی[۳] *(FCN)* و مدل‌های *CNN* مبتنی بر وصله کردن[۴] در مجموعه داده‌های تصاویر با وضوح‌بالای گوگل ارث و پهپاد جهت استخراج عارضه راه اعمال کردند. نتایج نشان داد که بهترین عملکرد گزارش‌شده مربوط به شبکه‌های عمیق کانولوشن است و معیار امتیاز F_1 برای مدل‌های معکوس مولد، روش *DenseNet* و *FCN - 32* اعمال‌شده بر تصاویر پهپاد و گوگل ارث زیاد است و به ترتیب ٪۹۶,۰۸، ٪۹۵,۷۲ و ٪۹۴,۵۹ به‌دست‌آمده است. [۲۰]. رن و همکاران در پژوهشی در سال ۲۰۲۰ یک کپسول جدید *(DA - CapsUNet)* با توجه به روش *U-Net* برای استخراج عارضه راه با ترکیب ویژگی‌های مفید نمایش کپسول و ویژگی‌های قدرتمند مکانیسم دقت پیشنهاد دادند. نتایج حاصله حاکی از آن بود که معیارهای کامل‌بودن، صحت و امتیاز *F1* به ترتیب ٪۹۴,۸۶، ٪۹۵,۲۳ و ٪۹۵,۰۴ به‌دست‌آمده است. ارزیابی‌های کمی در یک مجموعه داده بزرگ نشان داد که $DA - CapsUNet$ در مطالعه مقایسه‌ای با هشت روش یادگیری عمیق که اخیراً ایجادشده بودند در استخراج راه برتر و کاربردی‌تر است [۲۱]. در این میان، علاوه بر تحقیقات فوق که به آن اشاره شد مطالعات دیگری هم به‌منظور استخراج عارضه راه در مناطق شهری [۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱ و ۳۲] و غیرشهری [۲۲، ۳۳، ۳۴ و ۳۵] با استفاده از شبکه‌های عصبی عمیق[۵] *(DNN)* انجام‌گرفته است.

همان‌طور که از ادبیات تحقیق قابل استناد است، اکثر تحقیقات و مطالعات انجام‌گرفته درزمینه‌ی استخراج عارضه راه در مناطق پیچیده شهری با استفاده از تصاویر باقدرت تفکیک مکانی بالا انجام‌گرفته و تصاویر باقدرت تفکیک مکانی متوسط مورداستفاده قرار نگرفته است. از طرف دیگر نه‌تنها از تصاویر با وضوح مکانی متوسط تا آلان استفاده نگردیده، بلکه از تصاویر نوری سنتینل-۲ هم به‌منظور استخراج عارضه راه با بهره از الگوریتم‌های یادگیری عمیق استفاده‌نشده است. لذا در این تحقیق نه‌تنها برای اولین بار جهت استخراج عارضه راه در شهر تهران با استفاده از روش‌های یادگیری عمیق از تصاویر نوری سری سنتینل-۲ استفاده می‌گردد، بلکه برای نخستین بار هم از تصاویر شهرهای مختلف برای آموزش و اعتبارسنجی شبکه استفاده خواهد شد؛ بنابراین از دو الگوریتم یادگیری عمیق *FCN* و *RDRCNN* به علت تشابه در وجود یک لایه کانولوشن کامل و همچنین دقت و سرعت‌بالای آن‌ها در مقایسه با دیگر مدل‌های موجود جهت استخراج عارضه راه درون‌شهری استفاده می‌گردد.

[۱] *deep transfer learning with ensemble classifier (TEC)*

[۲] *Convolutional Neural Neworks*

[۳] *Fully Convolutional Network*

[۴] *patch*

[۵] *Deep neural networks*

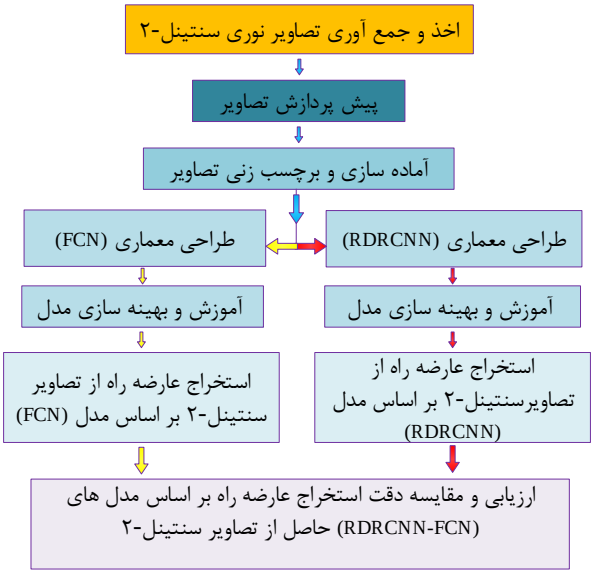
[DOR: 20.1001.1.20089635.1400.9.3.6.7]
[DOI: 10.52547/jgit.9.3.109]

۲- مواد و روش

در این تحقیق از تصاویر نوری سنتینل-۲ شهر تهران برای تست، تصاویر ۷ شهر دیگر که در ادامه به آن اشاره خواهد شد برای آموزش و اعتبارسنجی مدل‌ها استفاده می‌گردد. همچنین از شبکه‌ها یا مدل‌های *FCN* و *RDRCNN* به‌منظور آموزش و استخراج عارضه راه از تصاویر فوق بهره گرفته می‌شود. به این صورت که شبکه‌های فوق روی تصاویر اجراشده و نتایج باهم مقایسه خواهند شد. به‌طورکلی روند استخراج عارضه راه در شکل (۱) نشان داده‌شده است.

۲-۱-پیش‌پردازش و آماده‌سازی تصاویر نوری سنتینل-۲

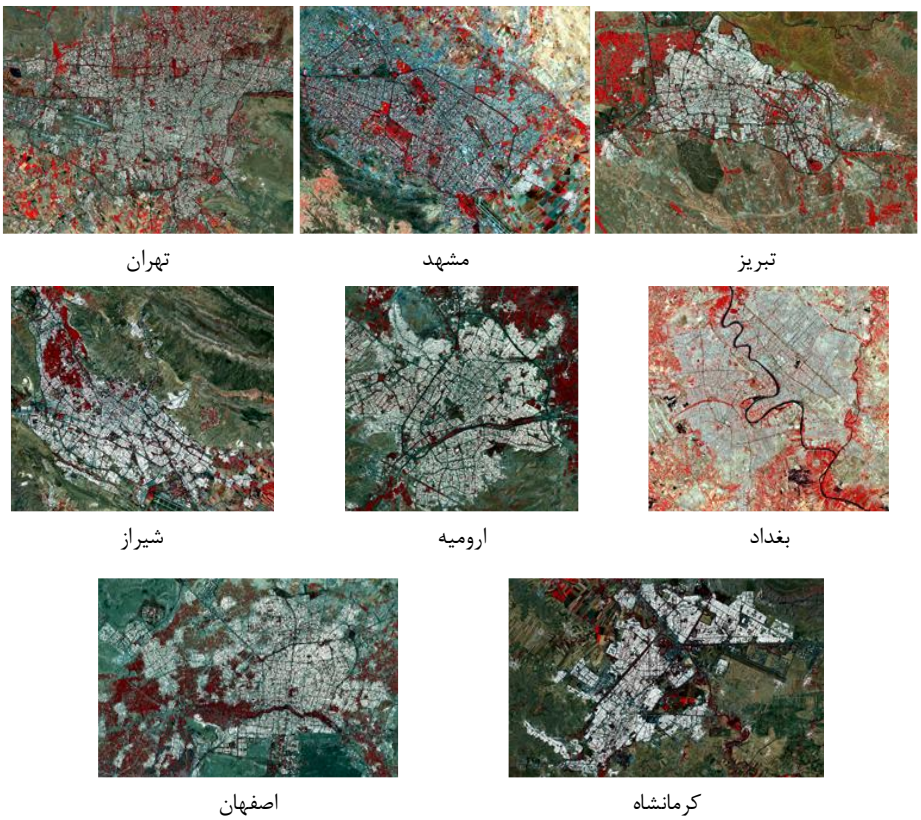
همان‌طور که در جدول (۱) آورده شده است، برای تصاویر نوری سنتینل-۲ از ۸ فریم کامل نسخه *MSIL2A* که دارای تمام تصحیحات و کالیبراسیون‌ها هستند و تقریباً هم‌زمان باهم تصویربرداری شده‌اند، استفاده می‌گردد. در این پژوهش از ۳ باند مادون‌قرمز نزدیک (۸)، قرمز (۴) و سبز (۳) که توان تفکیک مکانی ۱۰ متر دارند، استفاده می‌شود به‌طوری‌که از راست به چپ به رنگ‌های قرمز، سبز و آبی مانیتور اختصاص داده‌شده و یک ترکیب رنگی کاذب را تشکیل می‌دهند. در شکل(۲) تصاویر مورد استفاده نشان داده شده است.



شکل ۱: مدل مفهومی پژوهش

جدول ۱: اطلاعات مربوط به تصاویر مورداستفاده

تصاویر	باندها	تاریخ	قدرت تفکیک مکانی
تهران	۸-۴-۳	۲۰۲۰/۶	۱۰
مشهد	۸-۴-۳	۲۰۲۰/۶	۱۰
شیراز	۸-۴-۳	۲۰۲۰/۶	۱۰
تبریز	۸-۴-۳	۲۰۲۰/۶	۱۰
اصفهان	۸-۴-۳	۲۰۲۰/۶	۱۰
ارومیه	۸-۴-۳	۲۰۲۰/۶	۱۰
کرمانشاه	۸-۴-۳	۲۰۲۰/۶	۱۰
بغداد	۸-۴-۳	۲۰۲۰/۶	۱۰



شکل ۲: تصاویر سنتینل-۲ مورداستفاده از شهرهای مختلف

پس از آماده‌سازی تصاویر، عملیات برچسب‌گذاری عارضه راه با بهره‌گیری از نرم‌افزار متلب انجام می‌گیرد. در این میان شیوه‌ی نمونه‌برداری به این صورت است که تمام پیکسل‌هایی که متعلق به عارضه راه هستند برچسب راه به خود اختصاص خواهند داد؛ لذا برای این منظور یعنی تشخیص و شناسایی عارضه راه در داخل تصاویر از قطعه‌بندی معنایی استفاده می‌شود. به‌طورکلی فرمت تصاویر موردنظر به‌صورت فایل تصویری برچسب‌زده شده[1] (*Tiff*) و فرمت فایل‌های برچسب گذاری‌شده که به‌عنوان واقعیت زمینی[2] شناخته و مستقیماً از روی تصاویر برداشت‌شده‌اند با فرمت گرافیک شبکه قابل حمل[3] (*PNG*) ذخیره‌سازی و در ادامه به مدل‌های مورداستفاده معرفی خواهند شد.

طبق تئوری یادگیری عمیق اندازه قطعات تصویر روی نتایج نهایی مؤثر است [۴]؛ بنابراین در این تحقیق از قطعات یکسان با اندازه‌ی سطر و ستون ۲۵۶×۲۵۶ پیکسل استفاده می‌شود. همان‌طور که در جدول(۲) مشاهده می‌شود از کل ۸ فریم تصاویر مربوطه، تعداد ۲۸۶۴ قطعه با اندازه‌ی ۲۵۶×۲۵۶ استخراج شدند که پس از جداسازی قطعات نامناسب نمونه‌های آموزشی، اعتبارسنجی و تست به دست آمدند. درواقع پس از جداسازی و آماده‌کردن قطعات نهایی، از تصویر شهر تهران به‌عنوان داده‌های تست، در میان ۷ فریم تصویر باقی‌مانده تعداد ۱۰۰ قطعه تصویر که با دقت بالاتری برچسب‌گذاری شده بودند به‌عنوان داده‌های اعتبارسنجی و درنهایت بقیه‌ی قطعات باقی‌مانده هم به‌عنوان داده‌های آموزش به شبکه معرفی می‌شوند.

[1] *Tagged Image File Format*
[2] *Ground True*
[3] *Portable Network Graphic*

جدول ۲: تعداد قطعات مناسب به‌دست‌آمده پس از جداسازی قطعات نامناسب

تصاویر	اعتبارسنجی	آموزش	تست
سنتینل-۲	۱۰۰	۱۵۰۰	۱۳۵

۲-۲- استخراج عارضه راه از تصاویر سنتینل-۲ با استفاده از الگوریتم‌های یادگیری عمیق

ازآنجاکه تاکنون اکثر تحقیقات انجام‌گرفته به‌منظور استخراج عارضه راه با استفاده از تصاویر باقدرت تفکیک مکانی بالا انجام‌گرفته است، بنابراین انتخاب مدل شبکه عصبی عمیق مناسب جهت استخراج عارضه راه از تصاویر ماهواره‌ای باقدرت تفکیک مکانی متوسط می‌تواند تا حدی چالش‌برانگیز باشد. لذا پس از بررسی تحقیقات و مطالعات پیشین و همچنین انجام سعی و خطا، دو مدل *RDRCNN* و *FCN(U-Net)* پیشنهاد داده می‌شود. به‌طورکلی در کنار سعی و خطا و مطالعه در مورد توانایی و کارکرد دیگر مدل‌های یادگیری عمیق؛ سرعت‌بالا در شناسایی و استخراج راه به علت وجود یک‌لایه کانولوشن کامل در هر دو مدل علی‌الخصوص در مدل *FCN*، کاهش زمان آموزش نسبت به دیگر مدل‌های موجود، وجود *DPU* ها در معماری *RDRCNN* که باعث حفظ وضوح تصویر و تأثیرگذاری آن بر روی تصاویر با قدرت تفکیک مکانی متوسط، نتایج مناسب مدل *RDRCNN* در تنها مرجع موجود و نهایتاً عملکرد مناسب هر دو مدل روی تصاویر سنتینل-۲ عمده دلایل استفاده از دو مدل فوق می‌باشد. در ادامه به تشریح هریک از مدل‌های فوق پرداخته خواهد شد.

۲-۲-۱- شبکه عصبی کانولوشن باقیمانده عمیق پالایش‌شده *(RDRCNN)*

مدل *RDRCNN* برای اولین بار توسط گائو و همکاران در سال ۲۰۱۹ به‌منظور استخراج عارضه راه با استفاده از تصاویر باقدرت تفکیک مکانی بالا، پیشنهاد داده شد [۱۳]. معماری *RDRCNN* یک ساختار آموزش متقارن پایان به پایان جهت پیش‌بینی سطح پیکسل است. این معماری از نتایج حاصل از مدل‌های *ResNet*، *U Net* و *Deeplab* الهام گرفته‌شده است [۳۶، ۳۷ و ۳۸]. *RDRCNN* از دو هسته مرکزی واحد متصل باقی‌مانده *(RCU)* و واحد ادراک گسترش‌یافته *(DPU)* و به دنبال آن از یک‌لایه کانولوشن کامل تشکیل‌شده است [۱۳].

این معماری از سه قسمت طراحی‌شده است. قسمت اول برای استخراج ویژگی‌ها با استفاده از برخی *RCU* تا با ساختار کوچک‌تر از طریق برخی از عملگرهای حداکثر تجمع ایجادشده‌اند. قسمت دوم یعنی *DPU* ها برای بزرگ کردن میدان دید[۱] (*FOV*) بدون از دست دادن وضوح تصویر با استفاده از واحدهای متصل متوالی در چند مقیاس است. قسمت سوم یک ساختار گسترده برای تولید نقشه استخراج راه است که به همان اندازه ورودی می‌باشد [۱۳].لذا برای بهبود قابلیت بیان، عدم کاهش وضوح مکانی تصاویر در فضای ویژگی و جلوگیری از موانع و مشکلات مربوط به گرادیان افزایشی و گرادیان کاهشی (گرادیان انفجاری و شیب محوشده) از *RCU* و *DPU* چندمقیاسی استفاده می‌شود [۱۳ و ۳۹]. در ادامه به توضیح *RCU* و *DPU* پرداخته خواهد شد.

۲-۲-۱-۱- واحد متصل باقی‌مانده *(RCU)*

یک شبکه عصبی باقی‌مانده (*ResNet*) یک شبکه عصبی مصنوعی (*ANN*) است که بر اساس ساختارهای شناخته‌شده از سلول‌های هرمی در پوسته مغزی ساخته‌شده است. شبکه‌های عصبی باقیمانده این کار را با استفاده از اتصالات عرضی یا میان‌بر برای پریدن روی بعضی از لایه‌ها انجام می‌دهند. مدل‌های معمولی *ResNet* با پرش دو یا سه لایه که حاوی غیرخطی بودن[۲] (*ReLU*) و نرمال‌سازی دسته‌ای (*BatchNormalization*) دراین‌بین هستند، اجرا می

[۱] *Field of view*

[۲] *Rectified linear Unit*

شوند [۳۶].

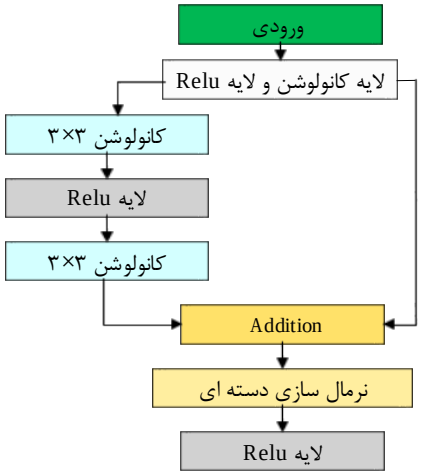
در این میان پیشنهاد نگاشت یادگیری و شناسایی باقی‌مانده توسط مسیرهای میان‌بر برای اولین بار توسط هی و همکاران در سال ۲۰۱۵ پیشنهادشده است [۳۶]. *RCU* یک شبکه عمیق است که در آن بخشی از داده‌های ورودی به لایه‌های بعدی پاس داده می‌شوند. این ویژگی، به این شبکه‌ها این امکان را می‌دهد که حقیقتاً عمیق باشند، اما درواقع، نوعی از شبکه‌های پیچشی عمیق بدون تأخیر صریح هستند. لذا *RCU* ها عملکرد خوبی را در حوزه بینایی کامپیوتری نشان می‌دهند [۱۳ و ۳۶]. ساختمان شبکه فوق توسط رابطه(۱) قابل‌بیان است:

$$xi+1 = wi+1^{\sigma}(wixi) + xi \quad \text{رابطه(۱)}$$

در رابطه(۱)، $xi+1$ برادرهای خروجی و xi بردارهای ورودی لایه‌های موردنظر را نشان می‌دهند. همچنین تابع $\sigma(.)$ نشان‌دهنده‌ی *ReLU* یا واحدهای خطی اصلاح‌شده است. درواقع رابطه (۱) به‌وسیله‌ی یک اتصال میان‌بر و جمع عنصر انجام می‌شود و سپس دومین تابع فعال هم به آن اضافه می‌شود [۱۳ و ۳۶].

هرچند هی و همکاران (۲۰۱۵) مطالب زیادی درباره تأثیرات ترکیبات مختلف *RCU* ارائه داده‌اند [۳۶]، اما در این مقاله، همانند تحقیق گائو و همکارانش (۲۰۱۹) از واحد متصل میانبر همان‌طور که در شکل(۳) نشان داده شده است، بدون تغییر استفاده می‌شود. همان‌طور که در شکل(۳) قابل مشاهده است برای جلوگیری از فاصله و شکاف بیش‌ازحد، یک‌لایه نرمال دسته‌ای به انتهای واحد پایه اضافه شده است [۳۹].

جزئیات بیشتر در مورد آن در جدول (۳) نشان داده شده است.



شکل ۳: ساختار واحد متصل باقی‌مانده.

جدول ۳: ساختار واحد متصل باقی‌مانده

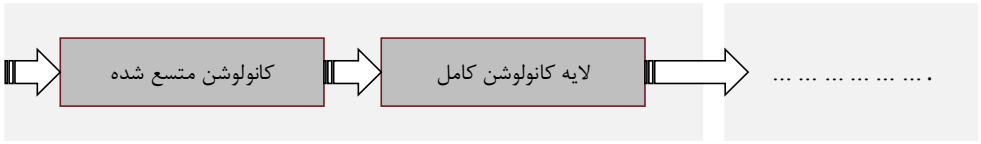
بخش	لایه‌ها	سایز کرنل
RCU-۱	لایه کانولوشن و لایه *Relu*	۱×۱
RCU-۲	لایه کانولوشن و *Relu*	۳×۳
RCU-۳	کانولوشن	۳×۳
RCU-۴	*Addition*	-
RCU-۵	نرمال‌سازی دسته‌ای	-
RCU-۶	لایه *Relu*	-

۲-۲-۱-۲- واحد ادراک گسترش‌یافته *(DPU)*

به‌منظور پذیرنده بودن و تفکیک مکانی بالا از *DPU* ها استفاده می‌شود [۴۰]. درواقع این معماری نه‌تنها قدرت تفکیک مکانی را حفظ می‌کند بلکه ورودی دریافت شبکه را هم بزرگ‌تر در نظر می‌گیرد [۴۱].

همان‌طور که در شکل (۴) قابل‌مشاهده است، یک *DPU* از یک‌لایه کانولوشن گسترش‌یافته و یک‌لایه کانولوشن کامل تشکیل‌شده است که کانولوشن گسترش‌یافته از کرنل‌های خاص با وزن‌های کم برای بزرگ کردن میدان دید (*FOV*) استفاده می‌کند و کانولوشن کامل روابط همسایگی میان آن‌ها را حفظ می‌کند. هم‌اندازه کرنل و هم فاصله وزن‌های کم، با افزایش ضریب انبساط به‌صورت نمایی افزایش می‌یابند؛ بنابراین با افزایش ضریب *DPU*، *FOV* نیز به‌صورت نمایی افزایش می‌یابد [۱۳ و ۴۲].

در این تحقیق، همچون مطالعات گائو و همکاران (۲۰۱۹) و یو و کولتون (۲۰۱۵) چهار مقیاس برای *DPU* طراحی خواهد شد [۱۳ و ۴۰]. جزئیات پارامترهای مورداستفاده در جدول (۴) آورده شده است.

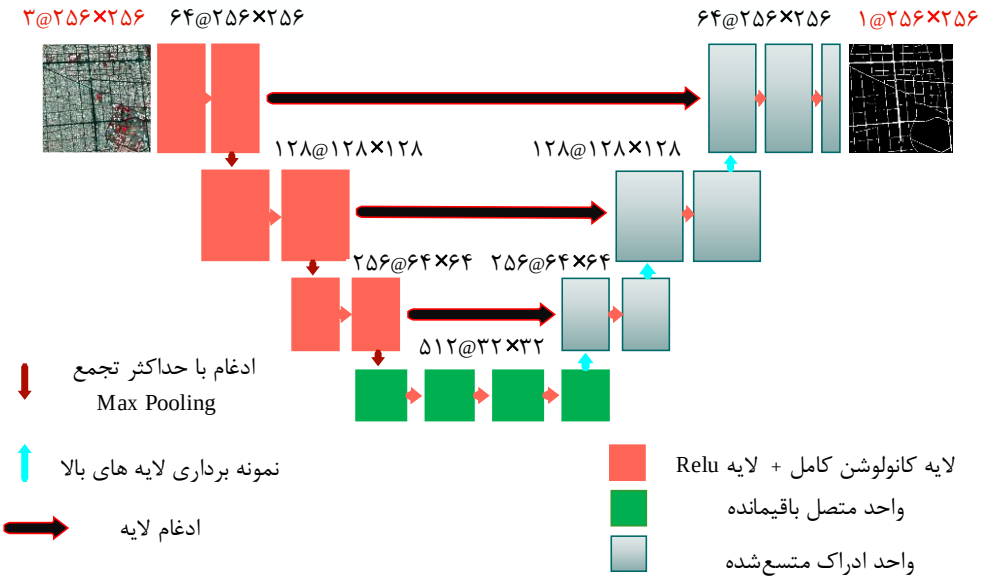


شکل ۴: ساختار واحد ادراک متسع‌شده

جدول ۴: جزئیات واحدهای کانولوشن متسع‌شده با مقیاس‌های مختلف

بخش	سایز کرنل	سایز تصویر	مقیاس متسع‌شده
DPU-۱	۳×۳	۵۱۲	۶
DPU-۲	۳×۳	۵۱۲	۱۲
DPU-۳	۳×۳	۵۱۲	۱۵
DPU-۴	۳×۳	۵۱۲	۲۴

در تحقیق حاضر به‌منظور استخراج عارضه راه با استفاده از تصاویر سنتینل-۲ بر مبنای تحقیق گائو و همکاران (۲۰۱۹)، به‌وسیله‌ی معماری *RDRCNN* خواهد بود. با این تفاوت که آن‌ها در تحقیق خویش، جهت استخراج هرچه بهتر تقاطع راه‌ها و شکستگی‌ها که توسط سایه‌ها یا درختان در خروجی ایجادشده بودند از یک مرحله‌ی پس پردازش برای کاهش مناطق شکستگی و بهبود توپولوژی استفاده کردند؛ این در حالی است که در این پژوهش به علت استفاده از تصاویر باقدرت تفکیک مکانی پایین‌تر، از مرحله پس پردازش فوق استفاده نخواهد شد. معماری شبکه *RDRCNN* در شکل (۵) نشان داده‌شده است.



شکل ۵: معماری مدل *RDRCNN*

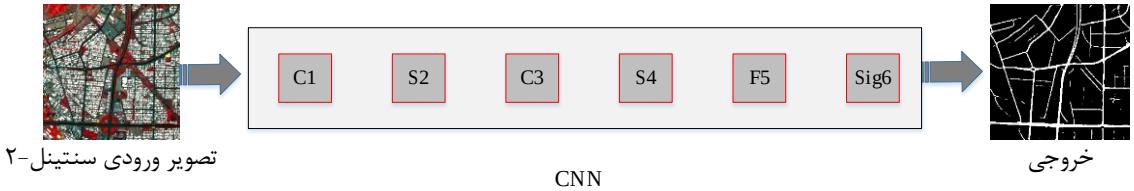
۲-۲-۲- مدل *FCN (U-NET)*

در تحقیق حاضر از یک شبکه *CNN* توسعه‌یافته به نام *FCN* که رونبرگر و همکاران در سال ۲۰۱۵ به‌منظور استخراج سلول‌ها در تصاویر پزشکی تنظیم نمودند و در سال ۲۰۱۹ توسط ژانگ و همکاران برای استخراج عارضه راه از تصاویر سنتینل-۱ استفاده شد، جهت استخراج عارضه راه از تصاویر سنتینل-۲ استفاده می‌گردد [۴ و ۳۷].

۲-۲-۲-۱ خلاصه‌ای از معماری شبکه *CNN*

در این تحقیق همچون مطالعه ژانگ و همکاران ابتدا مدل *DNN* از معماری *CNN* روی‌داده‌های سنتینل-۲ پیاده‌سازی می‌شود. درواقع معماری *CNN* مورداستفاده از روش سنتی *LeNet – 5* الهام گرفته‌شده و در طبقه‌بندی ارقام در تصاویر کارایی دارد. در مقاله حاضر، یک‌لایه کاملاً متصل به‌منظور استخراج راه استفاده می‌شود. *CNN* طراحی‌شده برای استخراج راه از تصاویر نوری که از انباشته شدن دولایه کانولوشن درهم‌آمیخته با دولایه حداکثر تجمع (*max-pooling*) و همچنین یک‌لایه کاملاً متصل که با عملیات کانولوشن در *FCN* جایگزین شده است، به دست می‌آید. درنهایت راه در تصاویر سنتینل-۲ توسط یک طبقه‌بندی کننده سیگموئید طبقه‌بندی می‌شود. معماری شبکه *CNN* مورد استفاده در شکل(۶) نشان داد شده است.

در شکل (۶)، *C* نشان‌دهنده توابع کانولوشن است، درحالی‌که *S* به عملیات نمونه‌برداری پایین اشاره دارد، به این معنی که *C1* و *C3* هر دو کانولوشن هستند، درحالی‌که *S2* و *S4* نمونه‌گیری پایین را انجام می‌دهند. *F5* تنها لایه کاملاً متصل است. *Sig6* نشان‌دهنده عملکرد سیگموئید مورداستفاده است.



شکل ۶: معماری شبکه *CNN*

لایه‌های کانولوشن با به دست آوردن ویژگی‌های محلی، کانولوشن تصاویر ورودی را با یک دسته فیلتر محاسبه می‌کنند. اندازه‌ی کرنل مورداستفاده باید متناسب باشد به‌طوری‌که انتخاب اندازه‌ی بزرگ یا اندازه‌ی کوچک می‌تواند باعث به وجود آمدن خطاهایی شود که فرایند آموزش را با دشواری مواجه کند [۴]. لذا در مطالعه حاضر، همانند تحقیق ژانگ و همکاران اندازه کرنل‌ها متناسب با شبکه تنظیم شدند، همچنین برای جلوگیری از بیش‌برازش[۱] شدن، قسمت‌هایی از شبکه، بعضی از نورون‌ها کنار گذاشته می‌شوند[۲]. درواقع *Overfit* شدن به معنای این است که الگوریتم فقط داده‌هایی که در مجموعه آموزشی یاد گرفته است را می‌تواند به‌درستی پیش‌بینی کند ولی اگر داده‌ای کمی از مجموعه‌ی آموزشی فاصله داشته باشد، الگوریتمی که *Overfit* شده باشد، نمی‌تواند به‌درستی پاسخی برای این داده‌های جدید پیدا کند و آن‌ها را با اشتباهِ زیادی طبقه‌بندی می‌کند. واحدهای خطی اصلاح‌شده (*ReLU*) به‌عنوان یک تابع فعال‌سازی در لایه کاملاً متصل انتخاب‌شده‌اند [۴ و ۴۳].

لایه *C1* شامل چهار کرنل است درحالی‌که *C3* شامل هشت کرنل است. برای هر دولایه از اندازه‌ی کرنل ۵×۵ در واحد پیکسل استفاده شد. برای لایه‌های *S2* و *S4*، پنجره ۳×۳ برای *max-pooling* انتخاب شده است. در این مطالعه برای حذف انحراف و اعوجاج تصویر از روش *max-pooling* به‌عنوان روش نمونه‌برداری فرعی استفاده می‌شود. لایه *F5* شامل ۱۲۸ نورون است. ازآنجاکه توابع سیگموئید به علت ماهیت منفرد و یکنواختی که دارند معمولاً به‌عنوان توابع آستانه در شبکه‌های عصبی استفاده می‌شوند و درواقع باعث می‌شود که متغیرها بین ۱ و ۰ قرار گیرند [۴].

لذا لایه *Sig6* یک تابع سیگموئید است که توسط رابطه(۲) تعریف می‌شود:

$$S(x)=1/(1+e^{-1}) \quad \text{رابطه(۲)}$$

همان‌طور که توضیح داده‌شده *FCN* شاخه‌ای از شبکه *CNN*است که با جایگزینی یک‌لایه کانولوشن با یک‌لایه به‌طور کامل متصل ایجاد می‌شود [۴]. *FCN* روشی برای از نو طرح کردن مدل‌های *CNN* به‌صورت شبکه‌های تمام کانولوشنی ارائه می‌کند و قادر به تولید خروجی‌هایی با اندازه متناسب به‌صورت بهینه است. گرچه *FCN* عمدتاً برای قطعه‌بندی معنایی ارائه‌شده است اما این تکنیک را می‌توان در کاربردهای دیگر نظیر دسته‌بندی تصاویر، لبه یابی و غیره نیز مورداستفاده قرارداد [۴۴ و ۴۵]. برای سری‌تر شدن و عبور از تنگنای اندازه ورودی *CNN*، یک‌لایه کاملاً متصل برای جایگزینی عملیات کانولوشن مورداستفاده قرار می‌گیرد [۴]. *FCN* به‌طورکلی ازنظر استخراج عارضه راه از پیکسل، به‌عنوان بهترین روش محسوب می‌شود [۴]. در این تحقیق از معماری *FCN* مبتنی بر تحقیقات اخیر ژانگ و همکاران (۲۰۱۹) و همچنین ژو و همکارانش (۲۰۱۷) به‌منظور استخراج عارضه راه از تصاویر سنتینل-۲ استفاده می‌شود [۴۳ و ۴]. *U-Net* با پیاده‌سازی پیچیده‌تر معماری *FCN*، لایه‌های نمونه‌برداری را برای بازسازی تصویر اضافه می‌کند. [۴].

در تحقیق حاضر به‌منظور هر چه‌بهتر، دقیق‌تر و سریع‌تر استخراج عارضه راه از تصاویر نوری مبنای کار بر اساس تحقیقات ژانگ و همکاران (۲۰۱۹) و همچنین ژو و همکاران (۲۰۱۷) خواهد بود، به این صورت یک ساختار *U – Net* مستخرج از تحقیقات فوق، طراحی و راه‌ها بر اساس معماری ادغامی از دو مطالعه‌ی یادشده استخراج می‌گردند [۴ و ۴۳].

در بخش بعدی به معماری تغییریافته حاصل از دو تحقیق موردنظر پرداخته خواهد شد.

۲-۲-۲-۲- معماری *FCN (U-NET)*

همان‌طور که اشاره شد، مدل *FCN (U-NET)* مورداستفاده در این پژوهش تلفیقی از دو مرجع موردنظر یعنی ژانگ و همکاران (۲۰۱۹) و همچنین ژو

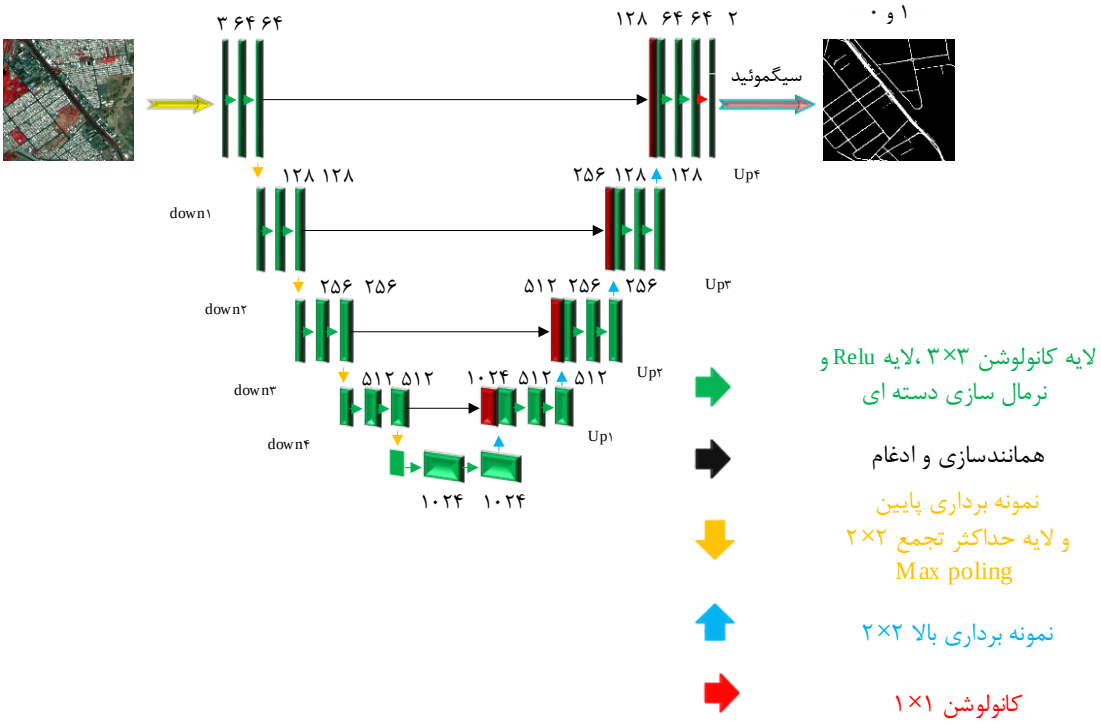
[۱] *Overfit*

[۲] *dropout*

و همکارانش (۲۰۱۷) می‌باشد و در این راستا تغییراتی در معماری آن نسبت به مدل‌های مبنا انجام می‌گیرد [۴ و ۴۳]. به‌طورکلی در تحقیق حاضر با حفظ ماهیت معماری اصلی، تغییراتی در بعضی از لایه‌های مدل جهت بهبود دقت، افزایش سرعت آموزش و نرخ یادگیری اعمال خواهد شد. به این صورت که اولاً برخلاف تحقیق ژانگ و همکارانش در معماری حاضر در کانولوشن اول یعنی کانولوشن ۶۴ در ۶۴ و کانولوشن چهارم یعنی کانولوشن ۱۶ در ۱۶ از استراتژی *dropout* برای جلوگیری از بیش‌برازش و افزایش سرعت آموزش استفاده می‌شود. تغییر مهم دیگر در معماری حاضر در قسمت کانولوشن ۳۲ در ۳۲ لحاظ گردید و برخلاف پژوهش ژانگ و همکاران (۲۰۱۹) به‌جای پریدن روی دولایه، کانولوشن ۳۲ در ۳۲ یعنی دو کانولوشن ۱۰۲۴ باهم در نظر گرفته‌شده و به هم متصل می‌گردند. درواقع در پژوهش ژانگ و همکاران (۲۰۱۹)، کانولوشن ۵۱۲ مستقیماً از روی دو کانولوشن ۱۰۲۴ عبور و دو کانولوشن ۳۲ در ۳۲ در نظر گرفته نشده است. به‌طورکلی با اعمال تغییرات در معماری شبکه نسبت به حالت مبنا (ژانگ و همکاران، ۲۰۱۹)، دقت مدل هم در سرعت آموزش شبکه و هم در دقت نهایی راه‌های استخراجی حاصل از تصاویر سنتینل-۲ بهبود پیدا کرد. در شکل (۷) معماری شبکه *FCN (U-Net)* نشان داده‌شده است.

همان‌طور که در شکل (۷) مشاهده می‌شود، فلش‌های سبز پررنگ نشان‌دهنده توابع کانولوشن و فعال‌سازی هستند که بعد از هر لایه کانولوشن یک تابع غیرخطی و یک نرمال‌سازی دسته‌ای هم قرار می‌گیرد. فلش‌های سیاه‌رنگ نمایانگر همانندسازی و ادغام لایه‌ها باهم می‌باشند. فلش‌های نارنجی‌رنگ و آبی‌رنگ به ترتیب نشان‌دهنده‌ی نمونه‌برداری پایین و نمونه‌برداری بالا هستند که نمونه‌برداری‌های رو به بالا از روش نزدیک‌ترین همسایه استفاده می‌کنند؛ و درنهایت فلش قرمزرنگ، لایه‌ی یک عمل کانولوشن با دو هسته ۱×۱ را نشان می‌دهد.

جزئیات مربوط به معماری مورداستفاده در جدول (۵) نشان داده شده است.



شکل ۷: معماری مدل *U-Net (FCN)*

جدول ۵: جزئیات کامل مربوط به معماری *U-Net (FCN)*

نام لایه	سایز تصویر	کرنل
ورودی	۲۵۶×۲۵۶	-
کانولوشن (۱)	۲۵۶×۲۵۶	۳×۳
ادغام ۱	۱۲۸×۱۲۸	۲×۲*Max pooling*
کانولوشن (۲)	۶۴×۶۴	۳×۳
ادغام ۲	۶۴×۶۴	۲×۲*Max pooling*
کانولوشن (۳)	۳۲×۳۲	۳×۳
ادغام ۳	۳۲×۳۲	۲×۲*Max pooling*
کانولوشن (۴)	۱۶×۱۶	۳×۳
ادغام ۴	۱۶×۱۶	۲×۲
متصل‌شدن دو لایه	۳۲×۳۲	۲×۲
نمونه‌برداری بالا ۱	۳۲×۳۲	۲×۲
نمونه‌برداری بالا ۲	۶۴×۶۴	۲×۲
نمونه‌برداری بالا ۳	۱۲۸×۱۲۸	۲×۲
نمونه‌برداری بالا ۴	۲۵۶×۲۵۶	۲×۲
کانولوشن (۵)	۲۵۶×۲۵۶	۱×۱
خروجی	-	-

۳- پیاده‌سازی و ارزیابی نتایج

ارزیابی یک شبکه، نیازمند ارزیابی دقت آن مدل بر روی کلاس و عارضه موردمطالعه است. برای برچسب-گذاری عارضه و کلاس راه موردنظر در این تحقیق از تصاویر نوری (سنتینل-۲) استفاده شد. به‌طوری‌که از انواع مختلف راه‌ها ازنظر پهنا، اندازه و اشکال گوناگون در ۸ شهر مختلف استفاده گردید.

۳-۱- بهینه‌سازی مدل‌های *RDRCNN* و *FCN*

در تحقیق حاضر به‌منظور آموزش و پیاده‌سازی مدل-های موردنظر از زبان برنامه‌نویسی پایتون و کتاب‌خانه متن‌باز تنسرفولو[۱] استفاده می‌شود. [۴۶ و ۴]. به‌طورکلی تمام مراحل آموزش، بهینه‌سازی، ارزیابی و استخراج راه در محیط سامانه‌ی ابری گوگل کولب[۲] انجام خواهد شد.

برای استخراج عارضه راه، تابع ضرر یا تابع هزینه[۳] با استفاده از وزن‌های مختلف مربوط به هر کلاس (راه و غیر راه) از رابطه(۳) محاسبه می‌شوند. برای بهینه‌سازی فرایند آموزش مدل با تابع هزینه از الگوریتم آدام[۴] استفاده می‌گردد. [۴۷].

رابطه(۳)

$$loss(y\hat{},y) = -\frac{1}{N}\sum_{i=1}^{N} y^{(i)} log\, y\hat{}^{(i)} + (1-y^{(i)}) log(1-y\hat{}^{(i)}))$$

[۱] *TensorFlow*

[۲] *Google Colab*

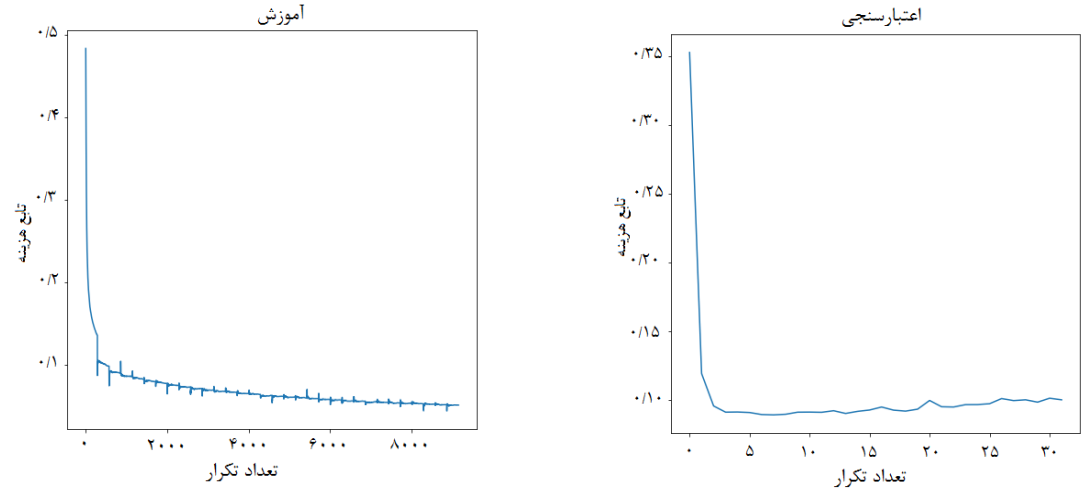
[۳] *loss function (cost function)*

[۴] *Adam*

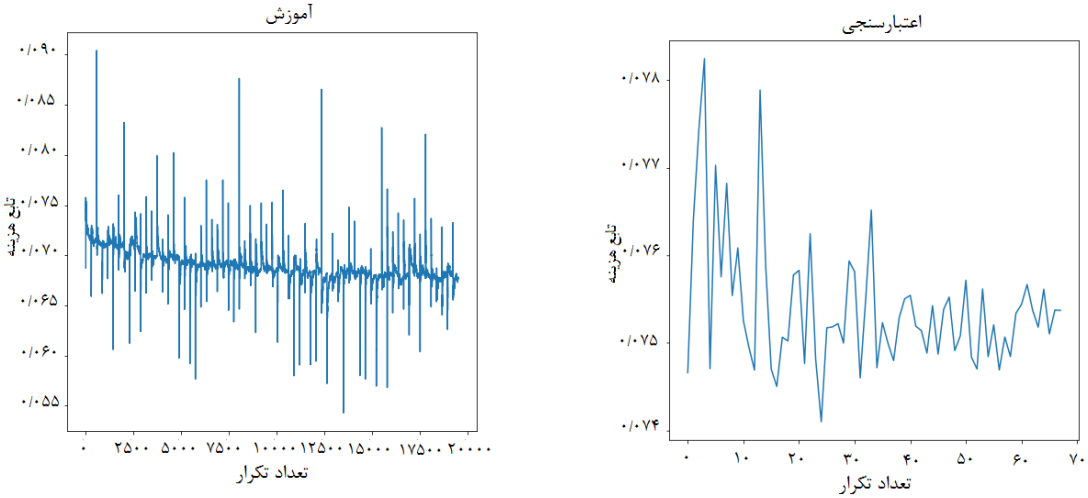
در رابطه(۳)، $\hat{y}$ برابر با مقداریست که شبکه تشخیص داده است، y برابر با مقدار واقعی (پیکسل‌های برچسب‌گذاری شده) و N تعداد کل تصاویر آموزشی است.

تابعِ ضرر یا تابعِ هزینه درواقع میزان خطا در هر بار اجرای شبکه‌ی عصبی را برای داده‌های آموزشی نمایش می‌دهد. به این صورت که شبکه با مشاهده مقدارِ ضرری که در هر بار اجرا داشته است، یاد می‌گیرد که چه مقدار بایستی وزن‌ها و انحراف را به‌هنگام یا آپدیت کند؛ بنابراین برای رسیدن به‌دقت بالا و همگرا شدن مدل و شبکه، لازم است تا شبکه با تعداد تکرار مکرر آموزش داده شود و پارامترهای دخیل در آن با استفاده از بهینه‌ساز موردنظر برآورد و نهایتاً تطبیق و تنظیم داده شوند. نتایج حاصل از الگوریتم بهینه‌سازی *Adam* برای تصاویر نوری سنتینل-۲ مربوط به مدل‌های *RDRCNN* و *FCN* در شکل (۸) نشان داده‌شده است. همان‌طور که در شکل (۸) قابل‌مشاهده است، تغییرات بر روی مجموعه داده‌های تحقیق (آموزش و اعتبارسنجی) با افزایش تعداد تکرار را نشان می‌دهد. درواقع بهترین و مناسب‌ترین مدل زمانی به دست می‌آید که در هنگام آموزش تابع هزینه کمترین مقدار را داشته باشد.

[DOR: 20.1001.1.20089635.1400.9.3.6.7]
[DOI: 10.52547/jgit.9.3.109]



مدل FCN



مدل RDRCN

شکل ۸: منحنی تکرار بر روی مجموعه تصاویر مورداستفاده، شکل راست: خطای اعتبارسنجی و شکل چپ: خطای آموزش (محور افقی: تعداد تکرار، محور عمودی: اختلاف اندازه‌گیری)

۳-۲- معیارهای ارزیابی

برای نمایش میزان دقت الگوریتم‌های موردنظر از معیارهای کامل‌بودن[۱] *(R)*، صحت[۲] *(P)* ، امتیاز[۳] *F1* و دقت کلی[۴] *(OA)* استفاده می‌گردد که در این بین امتیاز F_1 ترکیبی از دو معیار R و P است [۴۸و۴۹]. کامل‌بودن و صحت دو شاخص مهم برای ارزیابی آماری عملکرد نتایج آزمون‌های طبقه‌بندی باینری هستند که در آمار به‌عنوان توابع طبقه‌بندی شناخته می‌شوند.

[۱] *Recall*

[۲] *Precision*

[۳] *F1 score*

[۴] *overall accuracy*

نتایج به‌دست‌آمده بعد از تحلیل به دو گروه داده مثبت و منفی تقسیم‌بندی می‌شود.

بنابراین، برای ارزیابی معیارهای یادشده، تمام اهداف موردنظر (در این تحقیق عارضه راه) در تصاویر مشخص‌شده و شبکه برای هر هدف، مقادیری با عناوین *TP, FP, TN, FN* اختصاص داده می‌شود که بیانگر موارد زیر هستند:

۱) مثبت درست[۱] که یک معیار مثبت است؛ یعنی راه‌هایی که به‌درستی توسط شبکه تشخیص داده‌شده است.

۲) مثبت غلط[۲] که یک معیار منفی است؛ یعنی به نادرستی توسط شبکه، کلاس راه تشخیص داده‌شده است.

۳) درست منفی[۳] که یک معیار مثبت است؛ یعنی کلاس غیر راهی که به‌درستی توسط شبکه رد شده است.

۴) غلط منفی[۴] که یک معیار منفی است؛ یعنی کلاس راهی که به نادرستی توسط شبکه، غیر راه شناسایی‌شده است.

درواقع بر اساس موارد یادشده یک ماتریس ابهام به‌صورت جدول (۶) شکل خواهد گرفت:

جدول ۶: ماتریس ابهام

	راه	غیر راه
راه	*TP*	*FP*
غیر راه	*FN*	*TN*

لذا بر اساس تعاریف فوق روابط زیر مطرح خواهد بود:

کامل‌بودن *(R)* را می‌توان در قالب تعداد مثبت‌های بازگردانده شده توسط مدل تعریف کرد. با کمک رابطه(۴) و با استفاده از ماتریس سردرگمی می‌توان آن را به‌سادگی محاسبه کرد.

رابطه(۴)
$$\mathrm{Re}call = \frac{TP}{TP + FN}$$

صحت *(P)* استفاده‌شده در بازیابی اسناد می‌تواند به‌صورت تعداد اسناد صحیح بازگردانده شده توسط مدل، تعریف شود. در رابطه (۵) نحوه‌ی محاسبه‌ی آن آورده شده است.

رابطه(۵)
$$\mathrm{Pr}ecision = \frac{TP}{TP + FP}$$

امتیاز F_1، میانگین هارمونیک از کامل‌بودن و صحت را به ما می‌دهد. به بیان ریاضی، امتیاز F_1 میانگین وزن‌دار از کامل‌بودن و صحت *(R* و *P)* است. بهترین مقدار برای امتیاز F_1، ۱ و بدترین مقدار صفر است. با کمک رابطه(۶) می‌توان امتیاز F_1 را محاسبه کرد.

رابطه(۶)
$$F1 = 2 \times \frac{\mathrm{Pr}ecision \times \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call}$$

دقت کلی *(OA)*، متداول‌ترین معیار کارایی برای الگوریتم‌های طبقه‌بندی است. می‌توان آن را در قالب تعداد پیش‌بینی‌های صحیح صورت گرفته به نسبت همه پیش‌بینی‌های صورت گرفته، تعریف کنیم. در رابطه (۷) نحوه محاسبه دقت کلی نشان داده شده است.

رابطه(۷)
$$OA = \frac{TP + TN}{TP + TN + FN + FP}$$

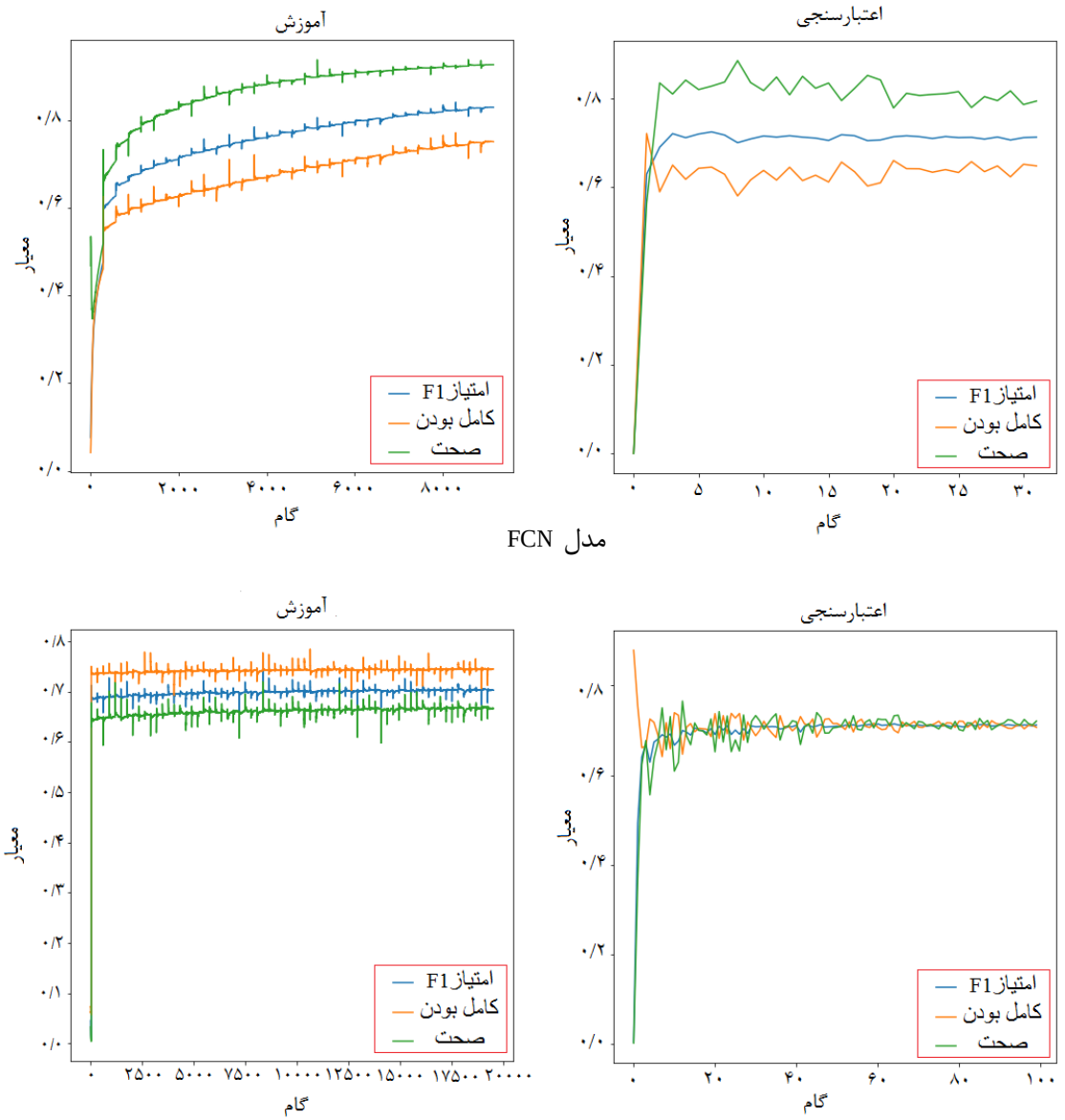
روند آموزش و اعتبارسنجی مدل‌های *RDRCNN* و *FCN* با استفاده از معیارهای کامل‌بودن، صحت و امتیاز F_1 برای تصاویر سنتینل-۲ در شکل (۹) نشان داده‌شده است.

[۱] *True Positive*

[۲] *False Positive*

[۳] *True Negative*

[۴] *False Negative*



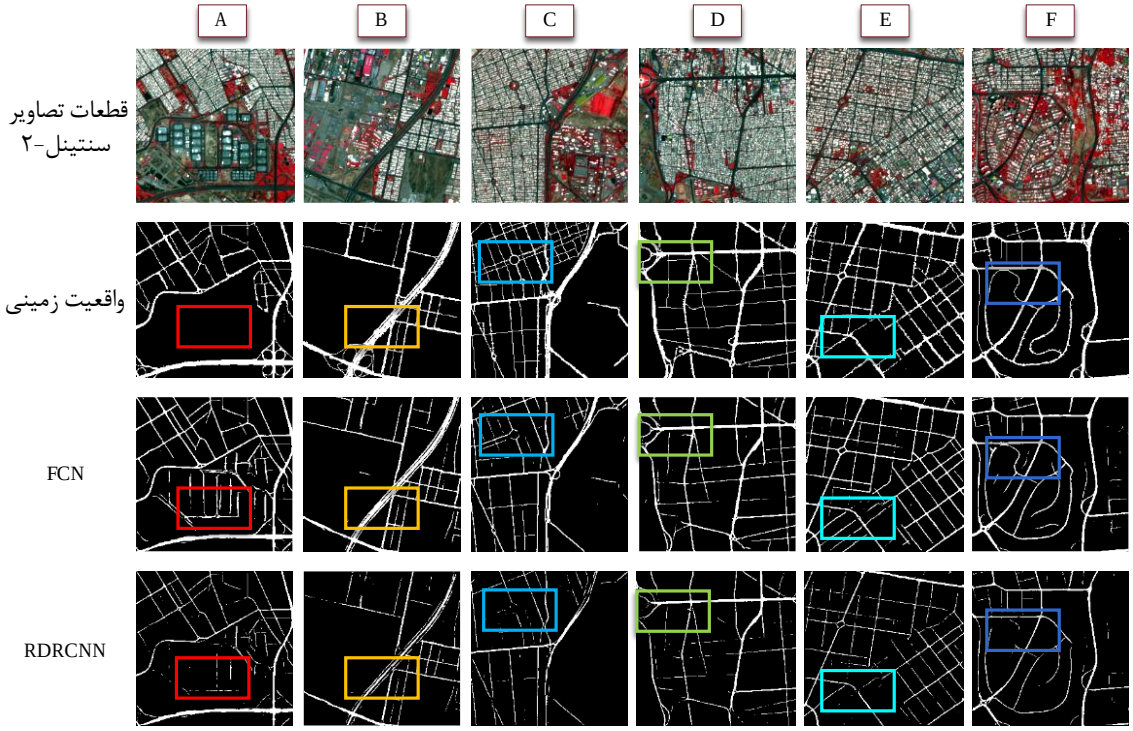
شکل ۹: نتایج شبکه با استفاده از متریک‌های ارزیابی صحت بر روی داده‌های اعتبارسنجی (سمت راست) و داده‌های آموزشی (سمت چپ)

همان‌طور که در شکل(۹) مشاهده می‌شود، روند داده‌های آموزش و اعتبارسنجی در مدل *FCN* و *RDRCNN* و ارزیابی آن‌ها با معیارهای اعتبارسنجی به این صورت است که هر چه روند متریک‌ها در طول دوره با معیارهای یادشده مقدار بالاتری داشته باشد، دقت مدل هم بیشتر خواهد بود. همان‌طور که مشاهده می‌شود در مدل *RDRCNN* معیارهای ارزیابی در هر دو داده‌ی آموزش و اعتبارسنجی همبستگی بالاتری نسبت مدل *FCN* دارند.

۴- بحث و نتایج

پس از آموزش شبکه با مدل‌های *FCN* و *RDRCNN* به‌منظور استخراج عارضه راه از تصاویر سنتینل-۲ برای محدوده‌ی شهر تهران و انجام سعی و خطا درروند آموزش شبکه نهایتاً خروجی هرکدام از شبکه‌های

موردنظر برای تصاویر به دست آمد. در این میان ۶ قطعه از تصاویر به همراه داده‌های واقعیت زمینی و خروجی‌های دو مدل مورداستفاده برای مقایسه به‌صورت تصادفی از بین ۱۳۵ قطعه انتخاب شدند. نتایج بصری برای تصاویر سنتینل-۲ در شکل (۱۰) نشان داده‌شده است.



شکل ۱۰: نتایج حاصل مدل *FCN* و *RDRCNN* بر روی قطعات منتخب تصاویر سنتینل-۲ شهر تهران

همان‌طور که در شکل (۱۰) نشان داده‌شده است، مدل *FCN* با دقت بالاتری توانسته شبکه راه را نسبت به مدل *RDRCNN* استخراج کند؛ به این صورت که تا حد زیادی متناسب با برچسب‌های ورودی (تصاویر واقعیت زمینی) است. در این میان مدل *FCN* پیکسل‌هایی از تصویر ورودی که برچسب راه به آن‌ها تعلق نگرفته است را توانسته به‌عنوان عارضه راه شناسایی کند. هرچند این قضیه در خروجی‌های مدل *RDRCNN* هم تا حدی قابل‌مشاهده است (قطعه *A*). از طرفی هر دو مدل در شناسایی مناطقی که روند و مسیر عارضه راه پیچیده بوده تشخیص مناسبی نداشته‌اند؛ اما باز مدل *FCN* متناسب با داده‌های واقعیت زمینی توانسته پیکسل‌های مربوط به عارضه راه را شناسایی و استخراج کند (قطعات *B-C-D*). با تجزیه‌وتحلیل نتایج تصاویر سنتینل-۲، واضح است که مدل *FCN* قادر به استخراج اطلاعات راه بیشتری نسبت به مدل *RDRCNN* است. از طرفی هر دو مدل تا حدی در استخراج راه‌های عریض شبیه به هم عمل کرده‌اند، اما در استخراج راه‌های کم‌عرض مدل *FCN* نتایج بهتری داشته است. در این میان هر دو مدل آزادراه‌ها، بزرگراه‌ها و راه‌های اصلی که عرض بالای ۳ پیکسل دارند را با دقت بالایی تشخیص داده‌اند و در استخراج راه‌های فرعی هم تا حد زیادی موفق عمل کرده‌اند؛ به‌طورکلی راه‌های فرعی که عرض ۲ تا ۳ پیکسل داشته‌اند نتایجشان مناسب بوده است؛ اما در بعضی از قسمت‌ها هر دو مدل خصوصاً مدل *RDRCNN* با چالش مواجه بوده و راه‌های فرعی که عرض کمتر از ۲ پیکسل داشته‌اند را به‌صورت بریده‌بریده و قطعه‌قطعه شناسایی نموده‌اند (قطعات *A*-

C-E). در همین راستا هر دو مدل در فرایند تشخیص راه در بعضی از مناطق پیچیده شهری مثلاً مناطقی که دارای میدان یا دورهای برگردان بوده‌اند دقت نسبتاً پایینی از خود نشان داده‌اند و این مورد در نتایج مدل *RDRCNN* بیشتر مشهود است (قطعات *A-C*).

پس از استخراج راه‌ها از تصاویر تست موردنظر (تهران) با استفاده از مدل‌های پیشنهادی در تحقیق، در مرحله‌ی بعد لازم است که نتایج به‌دست‌آمده با داده‌های مرجع (واقعیت زمینی) مقایسه و مورد ارزیابی قرار گیرند. لذا به‌منظور این کار همان‌طور که در بخش‌های قبل به آن اشاره شد، نتایج تحقیق با تصاویر مرجع و استفاده از ماتریس خطا برای عوارض راه درست و نادرست استخراج‌شده، ارزیابی خواهد شد. جدول ارزیابی صحت برای شکل (۱۰) در جدول (۷) آورده شده است.

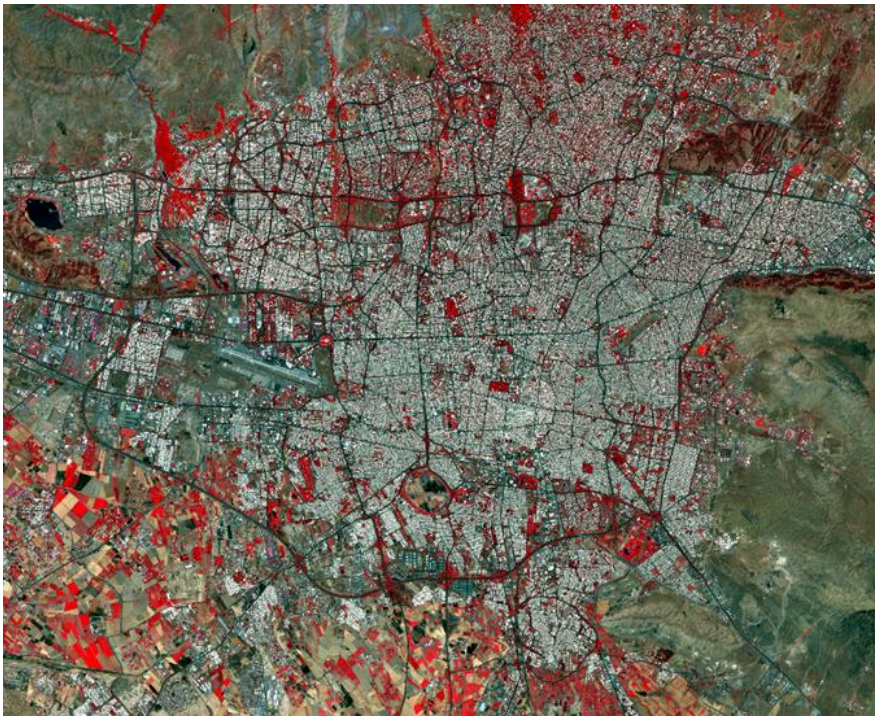
در جدول (۷) نتایج کمیتی حاصل از پارامترهای آماری برای مدل‌های *FCN* و *RDRCNN* نشان داده شده است.

جدول ۷: ارزیابی صحت راه استخراجی با استفاده از پارامترهای آماری مربوط به خروجی‌های شکل(۹)

قطعات تصاویر سنتینل-۲	*FCN*				*RDRCNN*			
	P (%)	*R (%)*	*F_1 (%)*	*OA (%)*	*P (%)*	*R (%)*	*F_1 (%)*	*OA (%)*
A	۷۷٫۱۵	۸۳٫۲۷	۸۰٫۱۳	۹۶٫۱۶	۷۳٫۲۷	۸۰٫۸۱	۷۶٫۸۵	۹۴٫۹۴
B	۷۹٫۱۴	۸۳٫۷۱	۸۱٫۳۶	۹۶٫۴۱	۷۳	۷۹٫۷۸	۷۶٫۲۳	۹۵
C	۷۷٫۵۲	۸۲٫۵۵	۷۹٫۹۷	۹۵٫۶۴	۷۴٫۵۶	۸۱٫۵۳	۷۸٫۱۶	۹۵٫۲۳
D	۷۸٫۱۳	۸۱٫۸۳	۷۹٫۹۳	۹۵٫۷۱	۷۴٫۱۱	۸۰٫۴۱	۷۷٫۱۳	۹۴٫۸۷
E	۷۹٫۵۹	۸۳٫۶۷	۸۱٫۵۷	۹۶٫۲۲	۷۵٫۳۳	۸۲٫۲۵	۷۸٫۶۳	۹۴٫۳۹
F	۸۰٫۱۹	۸۲٫۳۵	۸۱٫۲۵	۹۶٫۳۹	۷۵٫۹۲	۸۰٫۰۷	۷۷٫۹۳	۹۵٫۸۳

همان‌طور که در جدول (۷) مشاهده می‌شود، پارامترهای آماری به‌منظور ارزیابی دقت، حاکی از آن است که مدل *FCN* نسبت به مدل *RDRCNN* دقت بالاتری داشته است؛ به‌طوری‌که تابع F_1 و P در مدل *FCN* به‌صورت چشم‌گیری بهبودیافته است. در این میان هرچند معیارهای R و OA در هر دو مدل نزدیک به هم عمل کردند، اما باز برتری محسوسی در مدل *FCN* نسبت به مدل *RDRCNN* وجود دارد. لذا به‌طورکلی در استخراج عارضه راه از شهر تهران با استفاده از تصاویر نوری سنتینل-۲، مدل *FCN* نتایج مناسب‌تری داشته است.

پس از آموزش مدل‌ها و استخراج عارضه راه از قطعات تصاویر، نهایتاً با همان مختصات ورودی، قطعات تجمیع شده و نقشه راه شهر تهران برای تصاویر و مدل‌های مورداستفاده آماده‌سازی گردید. در شکل (۱۱) تصویر ورودی سنتینل-۲ از شهر تهران به‌عنوان داده‌ی تست و در شکل‌های (۱۲) و (۱۳) خروجی‌های نهایی شبکه راه استخراج‌شده به ترتیب برای مدل‌های *FCN* و *RDRCNN* نشان داده‌شده است.



شکل ۱۱: تصویر سنتینل-۲ از شهر تهران

شکل ۱۲: نتایج نهایی استخراج عارضه راه بر روی تصاویر سنتینل-۲ شهر تهران، مدل *(FCN)*

شکل ۱۳: نتایج نهایی استخراج عارضه راه بر روی تصاویر سنتینل-۲ شهر تهران، مدل *(RDRCNN)*

پس از تجمیع قطعات تصاویر و خروجی‌ها، جهت به دست آوردن میانگین دقت خروجی‌های موردنظر اقدام شد. در جدول (۸) میانگین دقت برای هرکدام از مدل-ها قابل مشاهده است.

جدول ۸: میانگین دقت عارضه راه استخراج‌شده از شهر تهران

مدل	*P (%)*	*R (%)*	*F_1 (%)*	*OA (%)*
FCN	۷۷٫۶۷	۸۲٫۹۲	۸۰٫۲۰	۹۶٫۳۰
RDRCNN	۷۱٫۳۷	۸۰٫۴۳	۷۷٫۷۴	۹۵٫۷۱

با توجه به جدول (۸) نتایج نهایی حاکی از آن است که مدل *FCN* نسبت به مدل *RDRCNN* با دقت بالاتری شبکه‌ی راه شهری تهران را استخراج کرده است.

درواقع استخراج راه با استفاده از روش‌های یادگیری عمیق در تصویر ماهواره‌ای به‌شدت تحت تأثیر طیف‌ها، بافت‌ها و موقعیت‌های مختلف قرار می‌گیرد. در همین حال، ویژگی‌های هندسی جاده‌ها نظیر تغییر انحنا و عرض نیز چالش‌هایی را برای استخراج راه به ارمغان می‌آورد. لذا تفاوت در ماهیت تصاویر با وضوح‌بالا در تحقیقات پیشین و تصاویر با وضوح متوسط در این پژوهش نتایج متفاوتی را رقم زد. در جریان استخراج عارضه راه از تصاویر نوری سنتینل-۲ مدل *FCN* نسبت به مدل *RDRCNN* دقت بالاتری داشت؛ و برعکس مطالعه‌ی گائو و همکاران (۲۰۱۹) که از مدل *RDRCNN* جهت استخراج عارضه راه از تصاویر نوری باقدرت تفکیک مکانی بالا استفاده کرده بودند، در این تحقیق مدل فوق نسبت به مدل *FCN* هم ازنظر بصری و هم ازنظر متریک‌های صحت‌سنجی دقت پایین‌تری داشت. هرچند هر دو مدل *FCN* و *RDRCNN* جهت استخراج راه از تصاویر سنتینل-۲ دقت بالایی داشتند،

اما با این وجود در قسمت‌هایی که شبکه راه دارای پیچیدگی بود، درروند شناسایی پیکسل‌های مربوط به کلاس راه دچار چالش شده‌اند. به‌طورکلی وجود پیکسل‌های مخلوط و قدرت تفکیک مکانی پایین تصاویر سنتینل-۲ نسبت به تحقیقات دیگر که از تصاویر با وضوح‌بالا استفاده کرده‌اند، فرایند برچسب-گذاری را با چالش و نهایتاً باعث پایین آمدن دقت مدل‌های مورداستفاده شده است. به صورتی که مشکلات فوق روی راه‌های فرعی با عرض کمتر از ۲ پیکسل به علت ادغام طیفی بیشتر تأثیرگذار بوده و روی نتایج نهایی تأثیر گذاشته است.

ازآنجاکه اخذ و دریافت تصاویر با قدرت تفکیک مکانی بالا مستلزم پرداخت هزینه‌های زیادی است و این تصاویر کمتر در اختیار کاربران و محققین قرار می-گیرد؛ این پژوهش با هدف ارزیابی عملکرد تصاویر با قدرت تفکیک مکانی متوسط سنتینل-۲ جهت استخراج عارضه راه درون‌شهری تهران با بهره‌گیری از الگوریتم‌های یادگیری عمیق انجام گرفت. اگرچه داده‌های ماهواره‌ای با وضوح مکانی متوسط سنتینل-۲ مطمئناً مزایای زیادی همچون منبع‌باز و دسترسی رایگان، پوشش دادن بخش زیادی از سطح زمین و قدرت تفکیک زمانی مناسب دارند و این موارد می‌توانند در مقیاس متوسط برای شناسایی و استخراج عارضه راه خصوصاً در مناطق شهری تا حد زیادی مناسب و کاربردی باشد؛ اما با استفاده از این تصاویر فقط ویژگی-های کلی راه‌ها همچون الگوی راه‌ها قابل‌شناسایی بوده و در این راستا استخراج اطلاعات با دقت بالا از راه‌ها میسر نخواهد بود. هرچند استخراج راه‌ها با استفاده از روش‌های یادگیری عمیق در تصاویر با وضوح مکانی متوسط چالش‌ها و مشکلات زیادی خصوصاً در جریان برچسب‌زنی به علت وجود پیکسل‌های مخلوط دارند؛ اما به‌طورکلی از پیچیدگی کمتری نسبت به تصاویر با قدرت تفکیک مکانی بالا برخوردار هستند؛ به این صورت که در استخراج راه درون‌شهری با بهره‌گیری از تصاویر با قدرت تفکیک مکانی بالا در کنار دقت بالای نتایج، مشکلاتی همچون حضور وسایل نقلیه، سایه ساختمان‌های بلند، تاج پوشش متراکم درختان در مجاورت راه‌ها، علائم و خط‌کشی‌های بستر راه ناپیوستگی‌های متعددی را به وجود می‌آورد. این در حالی است که در تصاویر با قدرت تفکیک مکانی متوسط مشکلات و پیچیدگی‌های یادشده وجود ندارد. از طرف دیگر هرچند تاکنون مطالعه‌ای تطبیقی به-منظور ارزیابی و مقایسه دقت روش‌های یادگیری عمیق و دیگر روش‌های متداول جهت استخراج عارضه راه درون‌شهری از تصاویر با وضوح مکانی متوسط خصوصاً تصاویر سنتینل-۲ انجام نگرفته است؛ اما با مقایسه مواردی همچون زمان محاسبات، نیاز به کاربر و پس پردازش و... که در مطالعات مختلف به آن اشاره شده است [۵۰]، می‌توان به این نتیجه رسید که استفاده از روش‌های یادگیری عمیق عملکرد بهتری دارند. در این راستا یافته‌های این پژوهش نشان‌دهنده‌ی این است که بهره‌گیری از روش‌های یادگیری عمیق نه‌تنها دقت بالایی دارند؛ بلکه با ایجاد یک بانک اطلاعاتی از داده-های آموزشی این امکان فراهم می‌شود که بتوان با سرعت بیشتر، بدون نیاز به کاربر و همچنین استفاده از دیگر شبکه‌های عصبی عمیق نقشه‌ی راه‌ها را به هنگام سازی نمود.

۵- نتیجه‌گیری

در این تحقیق یک طرح مبتنی بر *DNN* برای استخراج عارضه راه از تصاویر نوری (سنتینل-۲) با استفاده قطعه‌بندی معنایی ارائه شد. تمرکز این مطالعه برای اولین بار بر روی تصاویر باقدرت تفکیک مکانی متوسط به‌منظور استخراج عارضه راه در شهر تهران بود. در مقاله‌ی حاضر از مدل‌های *RDRCNN* و *FCN* به‌منظور استخراج عارضه راه از تصاویر سنتینل-۲ شهر تهران استفاده شد. نتایج حاصله حاکی از آن بود که هر دو مدل مورداستفاده با دقت قابل قبولی عارضه راه را از تصاویر فوق استخراج کرده‌اند و به‌طورکلی مدل *FCN* نسبت به مدل *RDRCNN* بهتر عمل کرده است. در این تحقیق مشخص شد که با استفاده از تصاویر باقدرت

[DOR: 20.1001.1.20089635.1400.9.3.6.7]
[DOI: 10.52547/jgit.9.3.109]

تفکیک مکانی متوسط هم می‌توان عارضه راه را با بهره-مندی از الگوریتم‌های یادگیری عمیق استخراج کرد. هرچند با توجه به ماهیت و قدرت تفکیک مکانی پایین تصاویر سنتینل-۲ و چالش‌های ناشی از آن‌که در هنگام برچسب‌گذاری و آموزش شبکه به وجود آمد، اما در مقایسه با تحقیقات پیشین که غالباً با استفاده از تصاویر باقدرت تفکیک مکانی بالا و خیلی بالا انجام‌گرفته نتایج استخراج عارضه راه در شهر تهران با استفاده از تصاویر نوری سنتینل-۲ قابل‌قبول است؛ و در این میان پیشنهاد می‌شود که در تحقیقات آتی علاوه بر استفاده از این تصاویر برای شهرهای مختلف و استفاده از دیگر الگوریتم‌های یادگیری عمیق، از روش-های مرسوم و متداول دیگر استخراج راه هم بهره گرفته‌شده و با نتایج حاصل از این پژوهش مقایسه گردد.

مراجع

[1] *R. C. Estoque, & Y. Murayama, "Monitoring surface urban heat island formation in a tropical mountain city using Landsat data (1987–2015)". ISPRS Journal of Photogrammetry and Remote Sensing, 133, pp 18-29, 2017.*

[2] *A. A. Matkan, M. Hajeb, & S. Sadeghian, "Road extraction from lidar data using support vector machine classification". Photogrammetric Engineering & Remote Sensing, 80(5), pp. 409-422, 2014.*

[3] *J. B. Mena, "State of the art on automatic road extraction for GIS update: a novel classificationP. Pattern recognition letters, 24(16), pp. 3037-3058, 2003.*

[4] *Q. Zhang, Q. Kong, C. Zhang, S. You, H. Wei, R. Sun, & L.Li, "A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network". European Journal of Remote Sensing, 52(1), pp. 572-582, 2019.*

[5] *P. Isola, J. Y. Zhu, T. Zhou, & A. A. Efros, "Image-to-image translation with conditional adversarial networks". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134, 2017.*

[6] *M. M. U. Rathore, A. Paul, A. Ahmad, B. W. Chen, B. Huang, & W. Ji, "Real-time big data analytical architecture for remote sensing application". IEEE journal of selected topics in applied earth observations and remote sensing, 8(10), pp. 4610-4621, 2015.*

[7] *M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, & Y. Zhu, "Big data for remote sensing: Challenges and opportunities". Proceedings of the IEEE, 104(11), pp. 2207-2219, 2016.*

[8] *Y. LeCun, Y. Bengio, & G. Hinton, "Deep learning", nature, 521(7553), pp. 436-444, 2015.*

[9] *R. Lian, W. Wang, N. Mustafa, & L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, pp. 5489-5507, 2020.*

[10] *J. Lu, J. Hu &, J. Zhou, "Deep metric learning for visual understanding: An overview of recent advances". IEEE Signal Processing Magazine, 34(6), pp. 76-84, 2017.*

[11] *K. Fu, J. Peng, Q. He, & H. Zhang, "Single image 3D object reconstruction based on deep learning: A review". Multimedia Tools and Applications, 80(1), pp. 463-498, 2021.*

[12] *W. Xia, Y. Z. Zhang, J. Liu, L. Luo, & K. Yang, "Road extraction from high resolution image with deep convolution network—A case study of GF-2 image". In Multidisciplinary Digital Publishing Institute Proceedings, Vol. 2, No. 7, p. 325, 2018.*

[13] L. Gao. W. Song. J. Dai. & Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network". Remote Sensing, 11(5), p. 552, 2019.

[14] Y. Li, B. Peng, L. He, K. Fan, Z. Li, & L. Tong, "Road extraction from unmanned aerial vehicle remote sensing images based on improved neural networks". Sensors, 19(19), p. 4115, 2019.

[15] J. Xin, X. Zhang, Z. Zhang, & W. Fang, "Road extraction of high-resolution remote sensing images derived from DenseUNet". Remote Sensing, 11(21), p. 2499, 2019.

[16] Y. Xu, Z. Xie, Y. Feng, & Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning". Remote Sensing, 10(9), p. 1461, 2018.

[17] H. He, D. Yang, S. Wang, S. Wang, & Y. Li, "Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss". Remote Sensing, 11(9), p. 1015, 2019.

[18] X. Zhang, X. Han, C. Li, X. Tang, H. Zhou, & L. Jiao, "Aerial image road extraction based on an improved generative adversarial network". Remote Sensing, 11(8), p. 930, 2019.

[19] J. Senthilnath, N. Varia, A. Dokania, G. Anand, & J. A. Benediktsson, "Deep TEC: Deep transfer learning with ensemble classifier for road extraction from UAV imagery". Remote Sensing, 12(2), p. 245, 2020.

[20] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, & A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review". Remote Sensing, 12(9), p. 1444, 2020.

[21] Y. Ren, Y. Yu, & H. Guan, "DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery". Remote Sensing, 12(18), p. 2866, 2020.

[22] S. Wang, H. Yang, Q. Wu, Z. Zheng, Y. Wu, & J. Li, "An improved method for road extraction from high-resolution remote-sensing images that enhances boundary information". Sensors, 20(7), p. 2064, 2020.

[23] X. Gao, X. Sun, Y. Zhang, M. Yan, G. Xu, H. Sun, ... & K. Fu, "An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network". IEEE Access, 6, pp. 39401-39414, 2018.

[24] Z. Hong, D. Ming, K. Zhou, Y. Guo, & T. Lu, "Road extraction from a high spatial resolution remote sensing image based on richer convolutional features". IEEE Access, 6, pp. 46988-47000, 2018.

[25] A. Abdollahi, B. Pradhan, & A. Alamri, "VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data". IEEE Access, 8, pp. 179424-179436, 2020.

[26] S. Wu, C. Du, H. Chen, Y. Xu, N. Guo, & N. Jing, "Road extraction from very high resolution images using weakly labeled OpenStreetMap centerline". ISPRS International Journal of Geo-Information, 8(11), p. 478, 2019.

[27] T. Sun, Z. Chen, W. Yang, & Y. Wang, "Stacked u-nets with multi-output for road extraction". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 202-206, 2018.

[28] Z. Zhang, Q. Liu, & Y. Wang, "Road extraction by deep residual u-net". IEEE Geoscience and Remote Sensing Letters, 15(5), pp. 749-753, 2018.

[29] Y. Wei, Z. Wang, & M. Xu, "Road structure refined CNN for road extraction in aerial image". IEEE Geoscience and Remote Sensing Letters, 14(5), pp. 709-713, 2017.

[30] A. Wulamu., Z. Shi., D. Zhang., & Z. He. (2019). "Multiscale road extraction in remote sensing images". Computational

intelligence and neuroscience, 2019.

[31] B. Shan, & Y. Fang, "A cross entropy based deep neural network model for road extraction from satellite images". Entropy, 22(5), p. 535, 2020.

[32] R. Lian, W. Wang, N. Mustafa, & L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review". IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, pp. 5489-5507, 2020.

[33] C. Tao, J. Qi, Y. Li, H. Wang, & H. Li, "Spatial information inference net: Road extraction using road-specific contextual information". ISPRS Journal of Photogrammetry and Remote Sensing, 158, pp. 155-166, 2019.

[34] L. Zhou, C. Zhang, & M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 182-186, 2018.

[35] A. Buslaev, S. Seferbekov, V. Iglovikov, & A. Shvets, "Fully convolutional network for automatic road extraction from satellite imagery". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 207-210, 2018.

[36] K. He, X. Zhang, S. Ren, & J. Sun, "Deep residual learning for image recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[37] O. Ronneberger, P. Fischer, & T. Brox, "U-net: Convolutional networks for biomedical image segmentation". In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241, 2015.

[38] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, & 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, 40(4), pp. 834-848, 2017.

[39] S. Ioffe, & C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In International conference on machine learning, pp. 448-456, 2015.

[40] F. Yu, & V. Koltun, "Multi-scale context aggregation by dilated convolutions". arXiv preprint arXiv:1511.07122, 2015.

[41] P. Krähenbühl, & V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials". Advances in neural information processing systems, 24, pp. 109-117, 2011.

[42] R. Hamaguchi, A. Fujita, K. Nemot, T. Imaizumi, & S. Hikosaka, "Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery". In 2018 IEEE winter conference on applications of computer vision (WACV), pp. 1442-1450, 2018.

[43] Z. Xu, R. Wang, H. Zhang, N. Li, & L. Zhang, "Building extraction from high-resolution SAR imagery based on deep neural networks". Remote Sensing Letters, 8(9), pp. 888-896, 2017.

[44] D. Yoo, S. Park, J. Y. Lee, & So I. Kweon, "Multi-scale pyramid pooling for deep convolutional representation". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 71-80, 2015.

[45] S. Xie, & Tu, Z, "Holistically-nested edge detection". In Proceedings of the IEEE international conference on computer vision, pp. 1395-1403, 2015.

[46] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, ... & X. Zheng, "Tensorflow: A system for large-scale machine learning". In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), pp. 265-

283, 2016.

[47] D. P. Kingma, & J. Ba, "Adam: A method for stochastic optimization". arXiv preprint arXiv:1412.6980, 2014.

[48] H. G. Akçay, & S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations". IEEE Transactions on Geoscience and Remote Sensing, 46(7), pp. 2097-2111, 2008.

[49] Y. Li, X. Sun, H. Wang, H. Sun, & X. Li, "Automatic target detection in high-resolution remote sensing images using a contour-based spatial model". IEEE Geoscience and Remote Sensing Letters, 9(5), pp. 886-890, 2012.

[50] S.A. Kianejad Tejnaki, "Automatic Road Detection in Urban Areas Using LIDAR and Multi-Spectral Imagery Based on Object-Oriented Method". Ph.D. Thesis. K. N. Toosi University of Technology, 2019.

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

Journal of Geospatial Information Technology
Vol.9 No.3, Autumn 2021

Research Paper

Evaluation and comparison performance of deep neural networks FCN and RDRCNN in order to identify and extract urban road using images of Sentinel-2 with medium spatial resolution

Seyed Hedayat Sheikhghaderi[1]*, Parviz Zeaiean firouzabadi[2], Manoochehr Kelarestaghi[3]

1- Maste of Remote Sensing and Geographic Information System, Faculty of Geographical Sciences, Kharazmi University, Tehran
2- Associate Professor, Department of Remote Sensing and Geographic Information System, Faculty of Geographical Sciences, Kharazmi University, Tehran
3- Assistant Professor, Department of Electrical and Computer Engineering, Faculty of technical and Engineering, Kharazmi University, Tehran

Abstract

Road extraction, using remote sensing images, has been one of the most interesting topics for researchers in recent years.. Recently, the development of deep neural networks (DNNs) in the field of semantic section has become one of the important methods of Road extraction. The majority of the researches in the field of road extraction using DNN in urban and non-urban areas has been carried out using images with high spatial resolution. In this research, for the first time, to extract the road using DNN, the images with medium spatial resolution of Sentinel-2 sensor were used, that is the image of Tehran was used as test data and the images of 7 other cities (Mashhad, Isfahan, Shiraz, Tabriz, Kermanshah, Urmia and Baghdad) were used as training and validation data. In the Meanwhile, after preparing and labeling all the pixels related to the road surface, the images were converted into 256 × 256 pieces, and after separating the unsuitable parts, for test, training and validation data, 135, 1500 and 100 image pieces were obtained respectively. Finally, deep refined residual convolution neural networks (RDRCNN) and U-Net, that are based on fully convolutional networks (FCN), were used to train and extract the road complication. The results show that both RDRCNN and FCN models have well identified and extracted Tehran urban road network from Sentinel 2 images in comparison with the ground reality data. Meanwhile, the FCN model performed better than the RDRCNN model both visually and in terms of accuracy assessment metrics, so that for the FCN model, the criteria Recall 82.92%, accuracy 77.67%, F1 score 77.53 and overall accuracy 96. 30% and for RDRCNN the criteria Recall 80.43%, accuracy 71.37, F1 score 72.14% and overall accuracy 95.71% were obtained. In general, the findings of this study show the potential of using DNN methods to extract urban roads using images with medium spatial resolution of Sentinel-2.

Key words : *Deep Neural Networks(DNN), Road Extraction, RDRCNN, U-net, FCN, Sentinel-2, Medium Spatial Resolution.*

Correspondence Address*: Department of Remote Sensing and Geographic Information System, Faculty of Geographical Sciences, Kharazmi University, Tehran.*
Tel*: : +98 9387959508.*
Email : *hedayatp90@gmail.com*

[DOR: 20.1001.1.20089635.1400.9.3.6.7]
[DOI: 10.52547/jgit.9.3.109]