نشریه علمی علوم و فنون نقشه برداری، دوره نهم، شماره ۲، آذر ماه ۱۳۹۸ – مقاله پژوهشی

ارائه روشی بهینه مبتنی بر یادگیری عمیق به منظور طبقه‌بندی طیفی‌مکانی تصاویر با قدرت تفکیک مکانی بالا در مناطق نیمه‌شهری

سید مهدی موسوی*[۱]، حمید عبادی[۲]، عباس کیانی[۳]

[۱] دانشجوی کارشناسی ارشد فتوگرامتری و سنجش‌ازدور - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
mousavi.j93@gmail.com

[۲] استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
ebadi@kntu.ac.ir

[۳] دانشجوی دکتری فتوگرامتری و سنجش‌ازدور - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
abbasekiani@yahoo.com

(تاریخ دریافت آبان ۱۳۹۷، تاریخ تصویب شهریور ۱۳۹۸)

چکیده

رشد و پیشرفت روزافزون در شهرسازی و تغییرات سریع در سطح زمین ضرورت بررسی مستمر این تغییرات را افزایش داده است. طبقه‌بندی تصاویرسنجش‌ازدوری با قدرت تفکیک بالا می‌تواند بهینه‌ترین راه ممکن در جهت نیل به این هدف باشد. طبقه‌بندی این تصاویر به دلیل شباهت‌های بین کلاسی موجود و هم‌چنین وجود تفاوت‌ها در یک کلاس، همواره با چالش‌هایی روبرو بوده است. وجود این نوع چالش‌ها لزوم به‌کارگیری روش‌های دقیق در زمینه‌ی طبقه‌بندی تصاویر را یادآوری می‌کند. در این مقاله از روش شبکه‌های عصبی کانولوشنی مبتنی بر یادگیری عمیق به منظور طبقه‌بندی تصاویر استفاده گردیده است. دلیل این انتخاب امکان استفاده از ویژگی‌های عمیق و فراگیر توسط روش نام‌برده می‌باشد. در این مقاله، هدف اساسی تعیین ساختاری مبتنی بر شبکه‌های عمیق برای کلاسه‌بندی بهینه‌ی تصاویر هوایی با قدرت تفکیک مکانی بالا است. برای رسیدن به این هدف، جزییات و رویکردهای در نظر گرفته شده برای شبکه از اهمیت بالایی برخوردار است. به همین منظور، ابتدا، شبکه‌ای عمیق به منظور استخراج ویژگی‌های عمیق و بهینه از تصویر هوایی طراحی گردیده است. سپس، برای ارزیابی تاثیرگذاری همسایگی‌های مختلف در تولید ویژگی‌های عمیق بهینه، استخراج ویژگی در پچ‌های تصویری با ابعاد متفاوت، مورد بررسی قرار گرفته است. در انتها، برای بررسی قابلیت طبقه‌بندی روش یادگیری عمیق، در رویکردی متفاوت، از روش ماشین بردار پشتیبان برای طبقه‌بندی براساس ویژگی‌های عمیق تولیدشده، استفاده گردیده است. بررسی و مقایسه نتایج حاصله، تصویر روشنی از قابلیت طبقه‌بندی در روش یادگیری عمیق به نسبت روش مرسوم ماشین بردار پشتیبان، در شرایط مشابه استفاده از ویژگی‌های عمیق ارائه کرده است. جهت ارزیابی روش، از داده‌های هوایی با قدرت تفکیک مکانی یک متر در منطقه des moines در ایالات متحده آمریکا و تصویری از منطقه‌ی رویان واقع در استان مازندران استفاده گردیده است. در نهایت نتایج ارزیابی‌ها، بهبود در سه معیار دقت precision، recall و f1-score را در رویکرد استفاده از پچ‌های تصویری بزرگ‌تر را نشان می‌دهد. هم‌چنین استفاده از روش‌های یادگیری عمیق به عنوان استخراج‌کننده ویژگی و طبقه‌بندی تصویر با استفاده از ویژگی‌های عمیق تولیدشده توسط ماشین بردار پشتیبان، در حالت کلی نتایج ارزیابی بهتری به نسبت تولید ویژگی و طبقه‌بندی به صورت یک‌پارچه توسط روش شبکه‌ی عصبی کانولوشنی داشته است.

واژگان کلیدی: کلاسه‌بندی تصویر، تصاویر با قدرت تفکیک بالا، استخراج ویژگی، یادگیری عمیق، شبکه عصبی کانولوشنال

* نویسنده رابط

۱- مقدمه

پیشرفت تکنولوژی تصویربرداری امکان فراهم آوردن تصاویر سنجش‌ازدور با قدرت تفکیک‌های متفاوت را به‌وجود آورده است. این امر امکان تفسیر هوشمند پوشش و کاربری سطح زمین را مهیا ساخته است [۱]. از طرفی با توجه به رشد و گسترش روزافزون محیطی، مدیریت و ساماندهی این نواحی امری ضروری محسوب می‌گردد. تصاویر سنجش‌ازدور با قدرت تفکیک مکانی بالا، به‌عنوان یک ابزار دقیق، سریع و اقتصادی، امکانات فوق‌العاده‌ای برای استخراج عوارض و تجزیه و تحلیل‌های مکانی نسبت به سایر روش‌ها در مناطق شهری فراهم کرده‌اند.

پیشرفت تکنولوژی تصویربرداری باعث افزایش قدرت تفکیک تصاویر شده است، بنابراین هر شی موجود در تصویر از چندین پیکسل تشکیل شده است و بررسی پیکسل‌ها به صورت مجزا روش مناسبی برای کلاسه‌بندی تصاویر نمی‌باشد، به همین دلیل ایده‌ی استفاده از الگوهای مکانی و پیکسل‌های همسایه شکل گرفت. در یکی از حالات تحلیل تصاویر بر مبنای شی، تفسیر تصاویر را از مرحله‌ی تحلیل پیکسل منفرد به مرحله‌ی تحلیل اشیای تصویری ارتقا داد [۲]. هم‌چنین، در سال‌های اخیر نظریه مبتنی بر یادگیری ماشین به صورت گسترده در زمینه‌ی پردازش تصاویر کاربرد داشته است.

در نظریه‌های یادگیری ماشین[۱] ، سیستم طراحی‌شده، توسط نمونه‌های آموزشی قدرت تفسیر را می‌آموزد و سپس بر روی داده‌های جدید به پیش‌بینی می‌پردازد. در واقع یادگیری ماشین به دنبال راهی برای ایجاد برنامه‌ای است که عملکرد خود را به صورت خودکار و با توجه به تجربیات خود الگوریتم، ارتقا دهد [۳]. در چند سال اخیر، یادگیری عمیق[۲] به عنوان یکی از موفق‌ترین روش‌های یادگیری ماشین معرفی گردیده است. روش‌های یادگیری عمیق معمولا از تعداد زیادی لایه تشکیل شده‌اند که تعدادی از این لایه‌ها تبدیلات غیرخطی را به داده‌های ورودی اعمال می‌کند. سه ویژگی که سبب برتری روش‌های یادگیری عمیق نسبت به روش‌های سنتی شده‌اند عبارتند از: ۱) توانایی این روش‌ها در یادگیری مستقیم از داده‌های خام، ۲) ساختاری سلسله‌مراتبی و عمیق و ۳) عمومی و بهینه بودن روش نسبت به روش‌های سنتی. در روش‌های یادگیری عمیق، ویژگی‌ها در ساختاری عمیق و سلسله‌مراتبی در حالتی خودکار تولید می‌شوند. از میان روش‌های مرسوم در یادگیری عمیق سه روش خودرمزنگارهای انباشته[۳]، شبکه‌های باور عمیق[۴] و شبکه‌های عصبی کانولوشنی[۵] برای کلاسه‌بندی تصاویر سنجش‌ازدور مورد استفاده بوده است [۴].

استفاده از روش‌های یادگیری عمیق برای کلاسه‌بندی تصاویر سنجش‌ازدور بسته به کاربرد فرآیند، به دو هدف کلی کلاسه‌بندی متراکم[۶] و کلاسه‌بندی صحنه‌های تصویری[۷] تقسیم می‌شود. در حالت کلاسه‌بندی متراکم، هر پیکسل به صورت جداگانه، با در نظر گرفتن ویژگی‌های مربوط به خود پیکسل و گاها پیکسل‌های همسایه، برچسب‌دهی می‌شود. این نوع از کلاسه‌بندی تصاویر توسط روش‌های یادگیری عمیق بسته به ویژگی‌های مورداستفاده برای کلاسه‌بندی در سه حالت استفاده از ویژگی‌های طیفی، ویژگی‌های مکانی و ویژگی‌های طیفی-مکانی انجام می‌پذیرد [۵]. بنابراین در روش‌های یادگیری عمیق امکان بهره‌مندی از ویژگی‌های معنایی نظیر ویژگی‌های طیفی-مکانی وجود دارد، در حالی که روش‌های استخراج ویژگی مبتنی بر پیکسل از لحاظ شناختی و معنایی بسیار ابتدایی هستند. از طرف دیگر در روش‌های یادگیری عمیق بر خلاف روش‌های مبتنی بر شی تصویری، امکان بهره‌مندی از ویژگی‌های موجود در پیکسل‌های همسایه وجود دارد. در روش‌های یادگیری عمیق استخراج ویژگی‌های طیفی برای کلاسه‌بندی از لحاظ مفهومی ساده و برای اجرا آسان است، اما این روش‌ها ویژگی‌های مکانی، که بخش عمده‌ی ویژگی‌های موجود در یک تصویر سنجش‌ازدور است، را در نظر نمی‌گیرند [۶]. در این حالت ویژگی‌های طیفی به صورت برداری وارد شبکه می‌گردد [۷, ۸]. در حالت استفاده از ویژگی‌های مکانی، یک همسایگی از پیکسل موردنظر به صورت دوبعدی برای استخراج اطلاعات مکانی در نظر گرفته می‌شود [۹]. در حالت استفاده از ویژگی‌های مکانی نیاز به روش‌های کاهش دامنه‌ی طیفی نظیر روش آنالیز مولفه اصلی[۸] می‌باشد.

---

۱ Machine learning
۲ Deep learning

۳ Stacked autoencoders
٤ Deep belief network
٥ Convolutional neural network
٦ Dense classification
٧ Scene classification
٨ Principal component analysis

در حالت سوم، از ویژگی‌های طیفی-مکانی استفاده می‌گردد. حالت استفاده از ویژگی‌های طیفی-مکانی شامل دو رویکرد می‌باشد. رویکرد اول، مدل‌های استخراج ویژگی طیفی و مکانی را، که به صورت جداگانه به استخراج ویژگی پرداخته‌اند، با هم ترکیب می‌کند. به عنوان مثال استفاده از شبکه‌ی کانولوشنی یک بعدی (استخراج اطلاعات طیفی) و دوبعدی (استخراج اطلاعات مکانی) و سپس ترکیب ویژگی‌های تولیدشده در مرحله‌ی پیش‌بینی [۱۰]. این روش‌ها به نسبت روش‌هایی که اطلاعات طیفی-مکانی را به صورت ترکیبی استخراج می‌کنند دارای مزیت کمتری هستند. در رویکرد دوم، ورودی به صورت سه‌بعدی وارد شبکه می‌شود که هر پیکسل را به عنوان یک همسایگی P×P و در B باند به عنوان ورودی در نظر گرفته می‌شود (P×P×B). در این حالت می‌توان از تمام پتانسیل روش‌های یادگیری عمیق در استخراج ویژگی‌های بهینه استفاده کرد [۱۱, ۱۲]. در هر دو تحقیق ارائه شده در [۸] و [۱۲] از روش شبکه‌های عصبی کانولوشنی سه‌بعدی برای یادگیری ویژگی‌های طیفی-مکانی عمیق استفاده شده است. به طور خاص در [۸] یک شبکه با مقیاس بزرگ طراحی شده است که ورودی سه‌بعدی با ابعاد ۲۷×۲۷ دارد. در [۱۲] از شبکه‌ای با ابعاد ورودی کوچک‌تر استفاده شده است. ورودی در این شبکه ۵×۵ است. هم‌چنین در تحقیق صورت‌گرفته توسط عبدی و صمدزادگان [۱۳] از شبکه‌ی عصبی خودرمزنگار و کانولوشن برای استخراج ویژگی‌های ژرف برای طبقه‌بندی اطلاعات سنجنده‌های چندگانه استفاده کرده است.

هدف کلی دیگر در به‌کارگیری روش‌های یادگیری عمیق به منظور طبقه‌بندی تصاویر، مربوط به طبقه‌بندی صحنه‌های تصویری می‌باشد. گستردگی زیادی که در الگوهای ساختاری و مکانی صحنه‌های تصویری وجود دارد، این فرآیند را به امری بسیار چالش‌برانگیز تبدیل کرده است [۱۴]. کارهای ابتدایی در زمینه کلاسه‌بندی صحنه‌های سنجش‌ازدور براساس ویژگی‌های دست‌ساز انجام گرفته است [۱۵-۱۷]. مهندسی ویژگی (نحوه‌ی استخراج و انتخاب ویژگی‌های بهینه) در این روش‌ها توسط فرد متخصص صورت می‌گیرد. متداول‌ترین این روش‌ها عبارت است از هیستوگرام رنگی [۱۸]، توصیفگرهای بافت [۱۹]، GIST [۲۰]، SIFT [۲۱] و HOG[۱] [۲۲]. اگرچه ترکیب ویژگی‌های اولیه استخراج‌شده از تصویر توسط روش‌های تولید ویژگی سنتی عملکرد مناسبی بر جای گذاشته است، اما هم‌چنان چگونگی ترکیب ویژگی‌های متفاوت برای دستیابی به یک ویژگی کلی مناسب، امری چالش‌برانگیز محسوب می‌گردد. علاوه بر این، کیفیت ویژگی‌های دست‌ساز بسیار به خلاقیت فرد وابسته می‌باشد. به خصوص در مواردی که کلاسه‌بندی با موارد چالش‌برانگیز روبرو باشد، ویژگی‌های تولیدشده توسط روش‌های اشاره‌شده محدود و ناتوان خواهند بود [۲].

برای رهایی از محدودیت‌های موجود در روش‌های استخراج ویژگی مهندسی انسانی[۲]، یادگیری ویژگی به صورت خودکار از تصاویر به عنوان روشی عملی موردتوجه قرار گرفت. با استفاده از ویژگی‌های آموزش‌دیده از تصویر می‌توان به ویژگی‌هایی که قدرت تمایز بیشتری دارند دست یافت. این روش‌ها به دو صورت نظارت‌شده و نظارت‌نشده قابل اجراست. روش‌های مطرح در زمینه‌ی یادگیری ویژگی نظارت‌نشده شامل روش کدگذاری متراکم[۳] [۲۳] و روش خودرمزنگار[۴] [۲۴] است. در سال‌های اخیر از روش کدگذاری متراکم به صورت گسترده در زمینه‌ی کلاسه‌بندی در تصاویر سنجش‌ازدور استفاده گردیده است [۲۵, ۲۶]. روش خودرمزنگار نیز به صورت موفق در زمینه‌ی کلاسه‌بندی صحنه‌های تصویری اعمال گردیده است [۲۷].

در کاربردهای واقعی، روش‌های یادگیری ویژگی نظارت‌نشده به نتایج خوبی در زمینه‌ی کلاسه‌بندی کاربری زمین دست یافته‌اند. اما استفاده نکردن از برچسب‌های کلاس‌های مختلف از بار محتوایی و معنایی روش‌های نظارت‌نشده کاسته است و تضمینی برای ایجاد بهترین تمایز بین کلاس‌ها وجود ندارد. برای کلاسه‌بندی بهتر همچنان به داده‌های دارای برچسب نیاز است [۲]. به همین دلیل بیشتر روش‌های نوین در زمینه‌ی کلاسه‌بندی شامل روش‌های نظارت‌شده می‌باشد.

---

۱ Histogram of oriented gradients
۲ Hand-crafted feature extraction
۳ Sparse coding
٤ Autoencoder

هم‌چنین در روش‌های یادگیری ویژگی کم‌عمق[1]، از ویژگی‌های سطح پایین[2] یا سطح متوسط برای کلاسه‌بندی تصاویر استفاده می‌گردد. به‌کارگیری ویژگی‌ها در سطوح پایین در مواجهه با شباهت‌های بین کلاسی و تفاوت‌های درون‌کلاسی موجود در تصاویر سنجش‌ازدور دارای محدودیت تصمیم‌گیری است. به همین سبب، در سال‌های اخیر تحقیقات در زمینه‌ی کلاسه‌بندی تصاویر به سمت استفاده از ویژگی‌های عمیق پیش رفته است. به طور خاص در سال ۲۰۰۶ موفقیتی چشم‌گیر در زمینه‌ی روش‌های یادگیری ویژگی عمیق توسط آقایان هینتون و سالاخودینو شکل گرفت [۲۴]. از آنجایی که هدف محققان جایگزینی روش‌های تولید ویژگی دست‌ساز به وسیله‌ی روش‌های تولید ویژگی خودکار بود، شبکه‌های چندلایه‌ی قابل آموزش می‌توانست گزینه‌ی مناسبی باشد. از سوی دیگر ویژگی‌های تولیدی در روش‌های یادگیری عمیق در ساختاری سلسله‌مراتبی و در سطوح مختلف انتزاعی امکان تولید دارند. تعداد زیادی از روش‌های یادگیری عمیق نتایج بسیار خوبی در زمینه‌ی کلاسه‌بندی تصاویر سنجش‌ازدور از خود بر جای گذاشتند [۲۸-۳۰]. در مقایسه با روش‌های سنتی که به دانش و تجربه‌ی بالای فرد متخصص نیاز داشت، روش‌های یادگیری عمیق به تولید ویژگی به صورت خودکار از داده‌های خام می‌پردازد و تنها نیاز به طراحی مدل شبکه و تعیین فراپارمترها[3] دارد. هم‌چنین در مقایسه با روش‌های نظارت‌نشده که از مدل‌های کم‌عمق برای طراحی آن‌ها استفاده شده است، شبکه‌های عمیق نظارت‌شده قابلیت داشتن چندین لایه با قابلیت یادگیری را دارا می‌باشد که این امر می‌تواند منجر به آموزش قوی‌تر شبکه و تولید ویژگی‌های متمایزکننده‌تر گردد [۳۱]. مدل‌های زیادی بر مبنای یادگیری عمیق امروزه مورد استفاده هستند که شبکه‌های باور عمیق [۳۲]، ماشین بولتزمن عمیق[4] [۳۳]، اتوانکودر انباشته[5] [۳۴]، شبکه‌های عمیق کانولوشنی [۳۵] از مهم‌ترین آن‌ها می‌باشد.

شبکه‌های عصبی کانولوشن به دلیل دارا بودن ساختاری مناسب برای کار با تصاویر، به عنوان روش منتخب بر مبنای یادگیری عمیق برای تولید ویژگی از تصاویر در نظر گرفته شده است. موضوعی که در این میان از اهمیت بالایی برخوردار است، تعیین ساختار شبکه متناسب با داده‌های ورودی به شبکه است. در این مقاله، از ساختاری مبتنی بر یادگیری عمیق استفاده شده است به گونه‌ای که معماری شبکه، فراپارمترهای موجود در شبکه و لایه‌های به کار رفته در آن به صورتی باشد که با ایجاد تعادلی میان عمق در نظر گرفته شده برای شبکه، که در واقع تعداد وزن‌های قابل تنظیم در فرآیند آموزش شبکه را نمایندگی می‌کند و محدودیت در تعداد نمونه‌های آموزشی، که ابزارهای آموزشی شبکه هستند، با در نظر گرفتن مشخصات تصاویر هوایی بتواند بهینه‌ترین حالت تولید ویژگی‌های عمیق و متمایزکننده از تصویر را داشته باشد و در عین حال از بیش‌برازش شبکه بر روی نمونه‌های آموزشی جلوگیری کند. نتایج به دست آمده نشان از کیفیت بالای ویژگی‌های تولیدی در ساختار پیشنهادی دارد. هم‌چنین تاثیر عامل اندازه‌ی ورودی شبکه که نشان‌دهنده‌ی میزان تاثیرگذاری ابعاد همسایگی در نظر گرفته شده در کلاسه‌بندی پیکسل است، مورد بررسی قرار گرفته است و تصویر روشنی از تاثیرگذاری عامل موردنظر ارائه گردیده است. در راستای جامعیت‌بخشی به بررسی‌های صورت گرفته، رویکرد استخراج ویژگی از طریق شبکه‌ی طراحی‌شده و کلاسه‌بندی توسط ماشین بردار پشتیبان با روش کلاسه‌بندی توسط شبکه‌ی عمیق سراسری مقایسه گردیده است تا علاوه بر بررسی و کیفیت‌سنجی ویژگی‌های عمیق تولیدی توسط روش پیشنهادی در کلاسه‌بندهای متفاوت، میزان توانایی شبکه‌های عمیق سراسری نیز در کاربرد کلاسه‌بندی مورد بررسی قرار گیرد.

۲- مبانی نظری تحقیق

روش‌های یادگیری عمیق از مجموع پیشرفت‌های شبکه‌ی عصبی مصنوعی[6] در طول چند دهه اخیر و به خصوص تکامل در ساختار و معماری این شبکه‌ها می‌باشد. نکته‌ی قابل ذکر در تفاوت ساختار شبکه‌های عصبی کانولوشنی نسبت به شبکه‌های عصبی مصنوعی، محدودیت‌های در نظر گرفته شده در ارتباطات بین نورون‌ها در شبکه‌های کانولوشنی است. اگرچه این محدودیت‌های اعمال‌شده عمومیت شبکه‌های کانولوشن را

[1] Shallow learning
[2] Low-level features
[3] Hyperparameters
[4] Deep Boltzmann machines
[5] Stacked autoencoder
[6] Artificial neural network

در مواجهه با داده‌های متنوع کاهش می‌دهد، اما ساختار این شبکه‌ها با توجه به قیود خاص آن کاملا منطبق و مناسب با ساختار تصاویر است. واضح است که به کار بردن ساختار مناسب برای مسئله موجب بهبود عملکرد در حل مسئله خواهد بود. شبکه‌های کانولوشن با توجه به محدودیت‌های اعمال‌شده در ساختارشان، میزان محاسبات برای حل مسئله کلاسه‌بندی را به نسبت شبکه‌های عصبی مصنوعی به میزان قابل‌توجهی کاهش می‌دهند که این امر امکان استفاده از این شبکه‌ها در حالت عمیق و با استفاده از لایه‌های متعدد را فراهم می‌آورد [۳۶]. بر این اساس، مقاله حاضر بر روی فرایندهای تولید ویژگی‌های عمیق و کلاسه‌بندی براساس آن متمرکز شده است که در ادامه به معرفی و بیان جزئیات روش‌های بکار رفته در هر مرحله، پرداخته شده است.

۲-۱- شبکه‌های عصبی کانولوشنی

یادگیری عمیق زیرمجموعه‌ی روش یادگیری ماشین محسوب می‌گردد. ویژگی برجسته و متمایزکننده‌ی روش‌های یادگیری عمیق ساختار سلسله‌مراتبی و چندلایه آن است. روش‌های یادگیری عمیق هم‌چنین در میان روش‌های یادگیری بازنمایندگی[۱] نیز دسته‌بندی می‌شود. یادگیری بازنمایندگی مجموعه‌ای از روش‌هاست که در آن سیستم با داده‌های خام تغذیه می‌گردد و به صورت خودکار یک مجموعه ویژگی بهینه را جهت تشخیص و طبقه‌بندی ارائه می‌دهد. روش‌های یادگیری عمیق درواقع روش‌های مبتنی بر یادگیری بازنمایندگی هستند که در سطوح مختلف به تولید ویژگی می‌پردازند. شبکه‌های عصبی کانولوشنی از مهم‌ترین و پرکاربردترین روش‌های یادگیری عمیق به طور خاص در در زمینه‌ی استخراج ویژگی و کلاسه‌بندی تصاویر سنجش‌ازدور محسوب می‌گردد [۳۷].

شبکه‌های عصبی کانولوشنی دارای لایه متعددی می‌باشند که هر یک از لایه‌ها یک سطح از ویژگی را نمایندگی می‌کند. این سطوح متفاوت نمایندگی، توسط کنار هم قرار گرفتن ساختارهای ساده و غیرخطی به دست می‌آید که هرکدام از ساختارها ویژگی‌ها را به سطح انتزاع بالاتری ارتقا خواهد داد. تفاوت اصلی میان روش شبکه‌ی عصبی کانولوشن و شبکه‌ی پرسپترون چندلایه این موضوع است که در شبکه‌ی کانولوشن در هر لایه از دیدگاه متفاوتی، از طریق لایه‌های کانولوشن به ویژگی‌ها نگاه می‌شود [۳۸]. به طور مثال برای یک داده‌ی تصویری، ویژگی‌های تولیدشده در لایه‌های ابتدایی معمولا نشان‌دهنده‌ی وجود یا عدم وجود لبه‌ها در جهتی خاص یا مکان مشخصی از تصویر به عنوان ویژگی‌های اولیه می‌باشد. لایه‌های میانی عمدتا با دید وسیع‌تری به تصویر نگاه می‌کنند. به عنوان مثال در لایه‌های میانی شکل عمده‌تری از ترکیب لبه‌های استخراج‌شده در لایه‌های ابتدایی، به عنوان ویژگی تولید می‌شود. در لایه‌های انتهایی عموما اشیا به عنوان ترکیبی از ویژگی‌هایی که در لایه‌های پیشین استخراج شده است، تشخیص داده می‌شود. نکته‌ی کلیدی در یادگیری عمیق این موضوع می‌باشد که لایه‌های مختلف تولید ویژگی توسط فرد متخصص مقداردهی و تعیین نمی‌گردد و شبکه از طریق داده‌های آموزشی و در فرآیندی خودکار به یادگیری و تنظیم می‌پردازد [۳۷]. این موضوع وابستگی ویژگی‌های تولیدشده به تخصص انسانی را کاهش می‌دهد و هم‌چنین روش در مواجهه با تصاویر با مشخصات درونی متفاوت تعمیم‌پذیری مطلوبی خواهد داشت.

شبکه‌های عصبی مصنوعی مجموعه‌ای از نورون‌های متصل به هم می‌باشد که به یکدیگر پیغام می‌فرستند. در اصطلاح به شبکه‌های عصبی‌ای که پیغام‌ها در آن به صورت یک‌طرفه رو به انتهای شبکه ارسال می‌گردد، شبکه‌های پیشرو گفته می‌شود [۳۹]. این نوع شبکه‌ها رایج‌ترین حالت به کار رفته در بخش کار با تصاویر می‌باشد. هر نورون در این حالت یک بردار ورودی $x = x_1 \ldots x_n$ دریافت می‌کند و عملگر ساده‌ای را برای محاسبه‌ی خروجی محاسبه می‌کند.

$$\alpha = \sigma(wx + b) \qquad (۱)$$

بردار w بردار وزن، x بردار ورودی به هر نورون، b بایاس و σ تابع فعال‌سازی برای نورون می‌باشد. وزن‌ها و بایاس متغیرهای قابل‌آموزش برای نورون می‌باشند. هدف فرآیند آموزش شبکه‌های عصبی، تعیین متغیرهای قابل‌آموزش برای نورون‌های موجود در شبکه می‌باشد.

در شبکه‌ی عصبی کانولوشن نیز فرآیند پیشرو همانند شبکه‌های عصبی مصنوعی صورت می‌گیرد. با این تفاوت که در شبکه‌های کانولوشنال مشخصات و قیود خاصی برای

[۱] Representation learning

شبکه در نظر گرفته شده است. شبکه کانولوشنال دارای چهار ایده و مشخصه اصلی است که آن را از سایر روش‌ها متمایز ساخته است. این چهار ایده شامل ارتباطات محلی، اشتراک وزن، لایه‌ی ادغام و مشخصه‌ی عمق می‌باشد.

معماری متداول در شبکه‌های کانولوشنی بخش‌های متوالی را شامل می‌شود. بخش‌های اولیه معمولا متشکل از دو نوع لایه‌ی کانولوشن و ادغام است. واحدهای لایه‌ی کانولوشن در یک نقشه‌ی ویژگی سازماندهی می‌شوند که در آن هر واحد به قسمت‌هایی از نقشه‌ی ویژگی در لایه‌ی قبل از خود، توسط مجموعه‌ای از وزن‌ها که بانک فیلتر نامیده می‌شود، متصل می‌گردد. در شبکه‌های کانولوشنی مشخصه‌ی اتصال محلی در نظر گرفته شده است و هر نورون به یک محدوده مکانی خاص از ورودی خود متصل است. این امر باعث می‌گردد که تعداد متغیرهای قابل تنظیم در شبکه کاهش یابد و به طبع از میزان محاسبات شبکه نیز می‌کاهد. نتیجه‌ی اعمال لایه‌ی کانولوشن بر نقشه‌های ویژگی لایه‌ی ماقبل خود از یک تابع فعال‌ساز غیرخطی عبور می‌کند. خروجی α_{ij} مرتبط با موقعیت (i , j) در یک لایه‌ی کانولوشن در تصویر از طریق زیر به دست می‌آید.

$$\alpha_{ij} = \sigma\,(W * X)_{ij} + b\,) \qquad (۲)$$

در رابطه (۲)، W یک فیلتر با وزن‌های آموزش‌دیده، X ورودی لایه و عملگر '*' کانولوشن است. در رابطه (۲) برخلاف رابطه (۱) مشخصه اتصال محلی اعمال گردیده است. هم‌چنین فیلترهای آموزش‌دیده در این فرآیند به تمام موقعیت‌های مکانی در تصویر اعمال می‌گردد.

در شبکه‌های عصبی کانولوشنی معمولا تابع غیرخطی ReLu[۱] به عنوان تابع فعال‌ساز در نظر گرفته می‌شود [۳۷]. به این صورت، ویژگی‌های تصویری با گذر تصویر ورودی از لایه‌ی کانولوشن، با در نظر گرفتن اتصالات به صورت محلی، و تابع فعال‌ساز ویژگی استخراج می‌گردند. از فیلترهای تنظیم‌شده در مرحله آموزش، طبق قاعده‌ی اشتراک وزن، برای استخراج ویژگی در تمامی موقعیت‌های مکانی تصویر استفاده می‌گردد. دلیل این مورد، همبستگی بالای داده‌های همسایه و هم‌چنین یکسان بودن ماهیت ویژگی‌های دارای اهمیت در موقعیت‌های مختلف تصویر است. تابع ReLu به صورت زیر توصیف می‌گردد. که در آن x ورودی تابع ReLu می‌باشد.

$$Relu: max(0, x) \qquad (۳)$$

پس از لایه‌های کانولوشن عمدتا از لایه‌های کاهش نمونه[۲] استفاده می‌شود. لایه‌های ادغام معمول‌ترین لایه‌های در نظر گرفته‌شده برای کاهش نمونه در شبکه‌های کانولوشنی می‌باشد [۴۰]. در لایه‌ی ادغام به طور معمول از حالت‌های بیشترین مقدار و میانگین استفاده می‌گردد. در تابع ادغام بیشترین مقدار، گروهی از همسایگی در نقشه‌های ویژگی تولیدشده در لایه‌ی قبل به عنوان ورودی وارد لایه می‌شوند و بیشترین مقدار ورودی‌ها به عنوان ویژگی برجسته انتخاب شده و به عنوان خروجی لایه معرفی می‌گردد. به دلیل اینکه امکان دوران و جابجایی برای اشیای موجود در تصویر وجود دارد، لایه‌ی ادغام از طریق ادغام معنایی ویژگی‌های مشابه و برجسته‌سازی ویژگی‌های صریح در پچ‌های محلی تصاویر، امکان استخراج ویژگی‌های متمایزکننده در حالت مستقل از دوران و جابجایی‌های کوچک را فراهم می‌آورد. هم‌چنین این لایه‌ها با حرکت در سراسر تصویر و ادغام ویژگی‌های تصویری، باعث کوچک شدن ابعاد تصویر گردیده و نقش موثری در کاهش میزان محاسبات ایفا می‌کنند [۳۷].

معمولا در پی چند مرحله‌ی کانولوشن-تابع غیرخطی-ادغام که در ادامه‌ی یکدیگر قرار می‌گیرند، یک لایه‌ی تمام متصل[۳] قرار خواهد گرفت. در لایه‌های تمام متصل قاعده‌ی اتصال محلی نادیده گرفته می‌شود و نورون‌های موجود در این لایه به تمامی نورون‌های موجود در لایه‌ی ماقبل خود متصل هستند. درواقع با در نظر نگرفتن محدودیت‌های موجود در سایر لایه‌ها، می‌توان اطلاعات را با انتقال از سطوح پایین‌تر، در یک دید کلی‌تر به صورت خلاصه و چکیده‌ای از سطوح مختلف تبدیل کرد. هم‌چنین به دلیل اتصال کامل در این لایه، تعداد پارامترهای قابل تنظیم به طور قابل توجهی بیشتر از سایر لایه‌ها خواهد بود. پس از لایه‌های تمام‌متصل نیز به صورت معمول از یک فعال‌ساز استفاده می‌گردد. در لایه‌ی تمام متصل انتهایی خروجی به تعداد کلاس‌های موجود در تصویر خواهد بود و نتیجه‌ی آن امتیاز تعلق نمونه‌ی ورودی به هر کلاس را مشخص می‌کند.

[۱] Rectified linear unit

[۲] Pooling
[۳] Fully connected

برای لایه‌های تمام متصل موجود در شبکه نیز معمولا از تابع فعال‌ساز ReLu استفاده می‌شود. اما در این مورد لایه‌ی تمام متصل واقع در انتهای شبکه استثنا محسوب می‌گردد. برای این لایه به طور معمول از فعال‌ساز Softmax استفاده می‌گردد. این فعال‌ساز به گونه‌ای طراحی شده است که درجه عضویت یا احتمال تعلق به تمامی کلاس‌های موجود، برای نمونه‌ی موردنظر تعیین می‌کند. فعال‌ساز Softmax به صورت زیر توصیف می‌گردد.

$$P(y=j \mid x)=\frac{e^{x^t w_j}}{\sum_{k=1}^{K} e^{x^t w_j}} \qquad (۴)$$

در رابطه بالا، احتمال عضویت به کلاس j ام برای بردار نمونه x با بردار وزن w در میان K کلاس موجود محاسبه می‌گردد. لایه‌ی softmax توزیع احتمالاتی مربوط به کلاس‌های مختلف را در بازه‌ی ۰ تا ۱ نرمال می‌کند و امکان بررسی و مقایسه‌ی احتمالاتی بین کلاس‌های مختلف و محاسبه‌ی آسان‌تر مقدار تابع هزینه در هر تکرار آموزش را در اختیار قرار می‌دهد.

آموزش شبکه‌های عصبی کانولوشنی عمدتا تحت روش گرادیان نزولی تصادفی انجام می‌گیرد. در حالت پیشرو، ورودی که وارد شبکه می‌گردد، پس از گذر از لایه‌های موجود در شبکه تبدیل به یک خروجی می‌شود. اختلاف خروجی به‌دست‌آمده و مقدار واقعی توسط تابع هزینه محاسبه می‌گردد. در شبکه‌های عصبی کانولوشنال برای کلاسه‌بندی چندکلاسه به طور معمول از تابع هزینه‌ی Cross_entropy استفاده می‌گردد، که به صورت زیر تعریف می‌گردد.

$$H_{y'}(\mathrm{y}) = -\sum_{i} y'_i \log(y_i) \qquad (۵)$$

که در آن y' مقدار واقعی و y مقدار محاسبه‌شده برای نمونه‌ی شماره i می‌باشد. پس از محاسبه‌ی میزان خطا توسط تابع هزینه، مقدار تصحیح در هر پارامتر قابل تنظیم در شبکه، از طریق روش پس‌انتشار خطا و قاعده‌ی مشتق زنجیره‌ای محاسبه می‌گردد و مقدار تصحیح به پارامتر موردنظر اعمال می‌گردد. این فرآیند به صورت تکراری به تعداد دفعات مشخص‌شده برای شبکه انجام می‌گیرد.

$$\omega_i - \lambda \frac{\partial L}{\partial \omega_i} \rightarrow \omega_i \qquad (۶)$$

در رابطه (۶)، L تابع هزینه، ω_i وزن تنظیمی برای نمونه i- ام و λ نرخ آموزش تعیین‌شده برای شبکه است.

۲-۲- رویکردهای کاربردی متفاوت از شبکه‌های عصبی کانولوشنال

یکی از مراحل مهم در پیاده‌سازی موفق طبقه‌بندی تصویر، انتخاب ویژگی‌های مناسب می‌باشد. در بحث طبقه‌بندی، می‌توان از ویژگی‌هایی نظیر اثر طیفی، اطلاعات بافت و غیره برای ایجاد تمایز بین کلاس‌ها استفاده کرد. بکارگیری ویژگی‌های متنوع و متعدد در امر طبقه‌بندی، ممکن است به دلیل همبستگی بین آن‌ها، منجر به کاهش دقت طبقه‌بندی گردد؛ بنابراین، انتخاب ویژگی‌های مفید و مناسب برای تفکیک کلاس‌ها از یکدیگر، از اهمیت بالایی برخوردار است. ساختار سلسله‌مراتبی و لایه‌های تشکیل‌دهنده‌ی شبکه‌ی عصبی کانولوشنال علاوه بر امکان تولید و استخراج ویژگی‌های عمیق و بهینه، امکان برجسته‌سازی ویژگی‌های متمایزکننده و سرکوب ویژگی‌هایی با قدرت تمایز پایین را فراهم آورده است.

دسترسی به ویژگی‌های عمیق و بهینه نیازمند طراحی معماری متناسب با داده‌های در دسترس و امکانات سخت‌افزاری موجود برای آموزش شبکه می‌باشد. علاوه بر معماری مناسب، شبکه نیازمند تعیین فراپارامترهایی نظیر تعداد و ابعاد فیلترها در هر لایه، نرخ آموزش شبکه، تعداد نورون‌های موجود در لایه‌های تمام‌متصل و ... خواهد بود. در تعیین معماری و فراپارامترهای شبکه دانش و تجربه‌ی فرد متخصص نقش مهمی ایفا می‌کند.

با تعیین شبکه‌ی مناسب در جهت پردازش داده‌های موجود، دو راهکار متفاوت در جهت کاربرد شبکه‌های کانولوشنی در دسترس خواهد بود. در راهکار اول از شبکه‌ی کانولوشنی به صورت توام و یکپارچه به عنوان استخراج‌کننده ویژگی و کلاسه‌بند استفاده می‌گردد. در رویکرد دوم از شبکه به عنوان استخراج‌کننده ویژگی استفاده می‌گردد و ویژگی‌های تولیدشده به عنوان ورودی به یک کلاسه‌بند مجزا از شبکه وارد می‌گردد تا کلاسه‌بندی توسط ویژگی‌های تولیدی در شبکه‌های عصبی کانولوشنی ولی در خارج از شبکه اجرا گردد. نوگیرا و همکاران در [۴۱] به بررسی حالت‌های متفاوت کاربرد شبکه‌های کانولوشنی در کلاسه‌بندی صحنه‌های تصویری در تصاویر هوایی پرداخته‌اند.

هرکدام از رویکردهای استفاده از شبکه‌های کانولوشنی در کاربرد کلاسه‌بندی تصاویر سنجش‌ازدور دارای مزایا و

محدودیت‌هایی می‌باشد. در حالی که از یک کلاسه‌بند در حالت مجزا از شبکه استفاده گردد، ویژگی‌های استخراج‌شده توسط شبکه می‌تواند از هرکدام از لایه‌های شبکه استخراج گردد و به عنوان ورودی به یک کلاسه‌بند مجزا وارد گردد. در این حالت امکان ترکیب ویژگی‌های استخراج‌شده از شبکه با ویژگی‌های تولیدشده از سایر روش‌ها برای ورود به کلاسه‌بند مجزا وجود خواهد داشت. هم‌چنین این رویکرد قابلیت استفاده از کلاسه‌بندهای متفاوت را بسته به نیاز فرد در اختیار قرار خواهد داد. از طرفی استفاده از کلاسه‌بندهای مجزا نظیر ماشین بردار پشتیبان[1]، نیازمند تنظیم پارامترهای مرتبط با آن می‌باشد. در حالی که در شبکه‌های طراحی‌شده به عنوان کلاسه‌بند، نیاز به تعیین پارامترها فراتر از آنچه در شبکه آموزش می‌بیند و فراپارامترهای اولیه شبکه نیست. اگرچه در شبکه‌های عصبی کانولوشنی که به عنوان استخراج‌کننده ویژگی و کلاسه‌بند به صورت یکپارچه استفاده می‌شود، امکان استفاده از توابع هزینه مختلف نظیر Cross entropy و Hing برای کلاسه‌بندی وجود دارد، ولی این رویکرد هیچ امکانی جهت ترکیب ویژگی‌های تولیدشده با ویژگی‌های تولیدی توسط سایر روش‌ها را فراهم نخواهد کرد.

۲-۳- ارزیابی دقت طبقه‌بندی

ارزیابی دقت طبقه‌بندی، معمولاً از طریق ماتریس ابهام انجام‌گرفته و شاخص‌هایی نظیر دقت، Recall و F1-score را می‌توان محاسبه نمود. ماتریس ابهام، آرایه‌ای مربعی با ابعاد $n \times n$ است که در آن، n تعداد کلاس‌های موجود در تصویر است. این ماتریس رابطه بین دو مجموعه از نمونه‌ها را نشان می‌دهد که مجموعه نمونه اول (ستون‌های ماتریس) نماینده داده مرجع یا آزمون است و مجموعه نمونه دوم (سطرهای ماتریس) داده‌های برچسب‌گذاری شده توسط طبقه‌بندی کننده هستند. در این ماتریس عناصر روی قطر اصلی، تعداد نمونه‌های درست طبقه‌بندی‌شده و عناصر غیر قطری مجموعه اشتباهات را نشان می‌دهد [۴۲].

در این تحقیق برای ارزیابی روش از سه معیار دقت، recall و f1-score همانند تحقیق [۴۳] استفاده شده است. دقت یک کلاس، نسبت نمونه‌هایی است که به یک کلاس خاص تعلق داشته و درست پیش‌بینی شده است به تمام نمونه‌هایی که در فرآیند پیش‌بینی به آن کلاس خاص تعلق گرفته است. معیار ارزیابی Recall، نسبت نمونه‌هایی که به درستی متعلق به یک کلاس پیش‌بینی شده‌اند را به تعداد کل نمونه‌های حقیقی کلاس نشان می‌دهد. در واقع این معیار نشان می‌دهد کلاسه‌بند تا چه میزان از کل نمونه‌های متعلق به یک کلاس را به درستی به کلاس نسبت داده است. هم‌چنین معیار f1-score تابعی از دو معیار دقت و Recall است که درواقع میانگین وزن‌دار مرتبط‌کننده‌ی دو معیار نام‌برده محسوب می‌گردد. این معیار، معیاری شبیه به صحت کلی محسوب می‌گردد که مناسب برای کلاس‌هایی با توزیع نامساعد است.

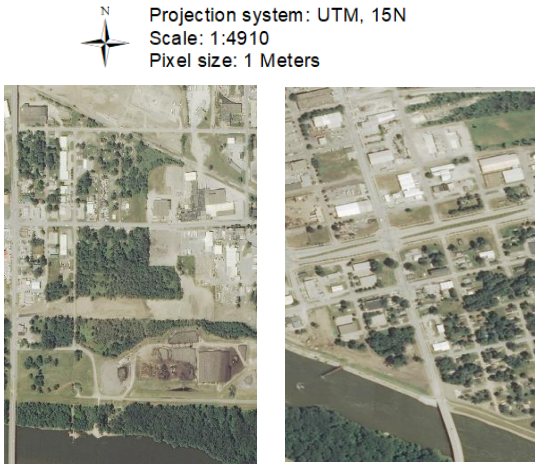
۳- روش پیشنهادی

در این بخش داده‌های مورداستفاده، روش پیشنهادی، آزمایشات صورت گرفته در ارائه‌ی ساختار شبکه‌ی عمیق و هم‌چنین نتایج این آزمایشات گردآوری شده است. ارزیابی‌های لازم جهت بررسی کارآمدی روش پیشنهادی به نسبت روش‌های متداول دیگر نیز مورد بررسی قرار گرفته است.

۱-۳- داده‌های مورداستفاده

داده‌ی مورداستفاده در این تحقیق مربوط به منطقه‌ی Des moines واقع در ایالت آیوآ در ایالات متحده‌ی آمریکا می‌باشد [۴۴]. این تصاویر هوایی که در ژوئیه سال ۲۰۱۴ اخذ گردیده دارای قدرت تفکیک مکانی ۱ متر است. برای اخذ تصاویر از سکوی هواپیما و سنجنده فعال استفاده شده است. تصاویر قائم مورداستفاده از ۴ باند مختلف قرمز ، سبز ، آبی و مادون قرمز تشکیل شده است. ابعاد تصویر اول ۵۷۹×۱۰۷۵ و ابعاد تصویر دوم ۸۸۴×۱۰۷۵ است. تصویر واقعیت زمینی از منطقه‌ی موردنظر به کمک اطلاعات موجود در نرم‌افزار Google earth برای ۵ کلاس مختلف آب، ساختمان، راه، درخت و فضای سبز تولید گردید تا برای ارزیابی عملکرد روش، مورداستفاده قرار گیرد. با توجه به تصویر واقعیت زمینی مبتنی بر ۵ کلاس نام‌برده، در تصویر اول حدود ۵۵ درصد و در تصویر دوم حدود ۴۷ درصد از مجموع پیکسل‌های تصویری دارای برچسب گردید. در شکل (۱) دو تصویر هوایی از محدوده‌ی مورد مطالعه نشان داده شده است.

[1] Support vector machine



تصویر اول تصویر دوم

شکل ۱- منطقه مطالعاتی

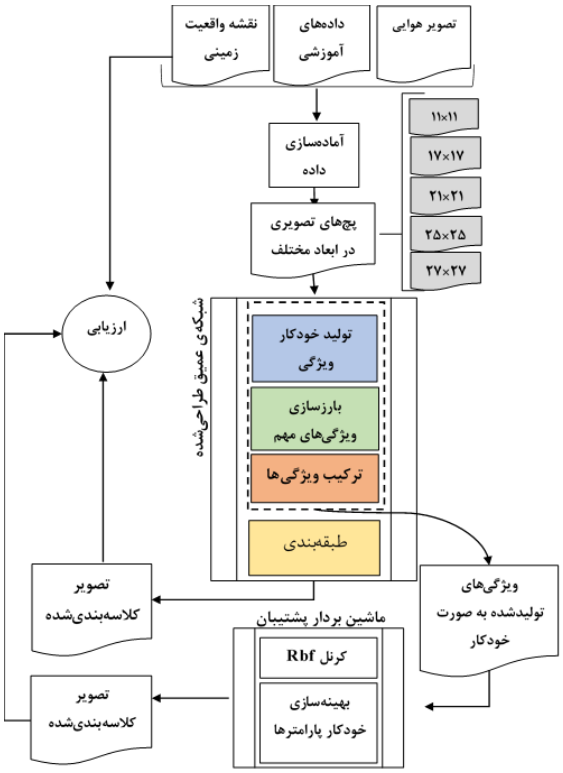
۳-۲- پیاده‌سازی روش

اساس کلی روش کلاسه‌بندی متراکم منطقه نیمه‌شهری در تصاویر هوایی بر پایه‌ی پچ‌های تصویری و با استخراج خودکار ویژگی توسط روش یادگیری عمیق است.

هدف اصلی در پیاده‌سازی روش، ابتدائا تعیین ساختار و معماری مناسبی از روش یادگیری عمیق برای استخراج ویژگی‌های بهینه در تصویر بوده است. سپس، تاثیرگذاری میزان مشارکت پیکسل‌های همسایه در در رسیدن به ویژگی‌های مطلوب مورد بررسی قرار گرفته است. در نهایت، به بررسی نتایج طبقه‌بندی ویژگی‌های استخراج‌شده از روش یادگیری عمیق در کلاسه‌بندی مجزا پرداخته شده است.

همانطور که در شکل (۲) دیده می‌شود، ابتدا پچ‌های تصویری از هرکدام از تصاویر استخراج می‌گردد. این پچ‌های تصویری در واقع یک همسایگی در اندازه‌های متفاوت از پیکسل‌های دارای برچسب در تصویر هستند. هرچه ابعاد پچ‌های تصویری افزایش یابد، درواقع بافت بیشتری در همسایگی پیکسل در نظر گرفته می‌شود و نتیجه‌ی حاصله از دید بزرگ‌تری در اطراف پیکسل مورد نظر متاثر خواهد بود. پچ‌های تصویری استخراج شده قبل از ورود به شبکه نرمال می‌شوند.

پچ‌های تصویری در ابعاد ۱۱×۱۱، ۱۷×۱۷، ۲۱×۲۱، ۲۵×۲۵ و ۲۷×۲۷ به جهت بررسی تاثیر گسترش ابعاد پچ‌های تصویری در کیفیت ویژگی‌های تولیدشده، در نظر گرفته شده است. این پچ‌های تصویری به عنوان ورودی به شبکه‌ی عمیق طراحی‌شده وارد می‌شوند. معماری در نظر گرفته‌شده برای شبکه‌ی عصبی کانولوشنال متناسب با داده‌ی هوایی موردنظر و به هدف تولید ویژگی‌های بهینه با توجه به مشخصات تصویر در نظر گرفته شده است. تولید ویژگی، بارزسازی ویژگی‌های قدرتمند، ترکیب ویژگی‌های استخراج‌شده و طبقه‌بندی در ساختار شبکه‌ی طراحی‌شده صورت می‌گیرد. با انجام این مرحله، توانایی شبکه در استخراج ویژگی‌های بهینه در مواجهه با تصویر هوایی، مورد ارزیابی قرار خواهد گرفت. هم‌چنین طبق دو رویکرد در نظر گرفته شده در طبقه‌بندی تصاویر در نظر گرفته شده است. در این رویکردها، ویژگی‌های عمیق تولیدشده توسط شبکه‌ی عمیق، یک بار توسط خود شبکه و بار دیگر توسط ماشین بردار پشتیبان طبقه‌بندی گردیده است و نتایج توسط سه معیار مورد اشاره در همه‌ی ابعاد در نظر گرفته شده، مورد بررسی قرار گرفته است.



شکل۲- روندنمای کلی

۳-۲-۱- پیش‌پردازش و آماده‌سازی داده‌ها

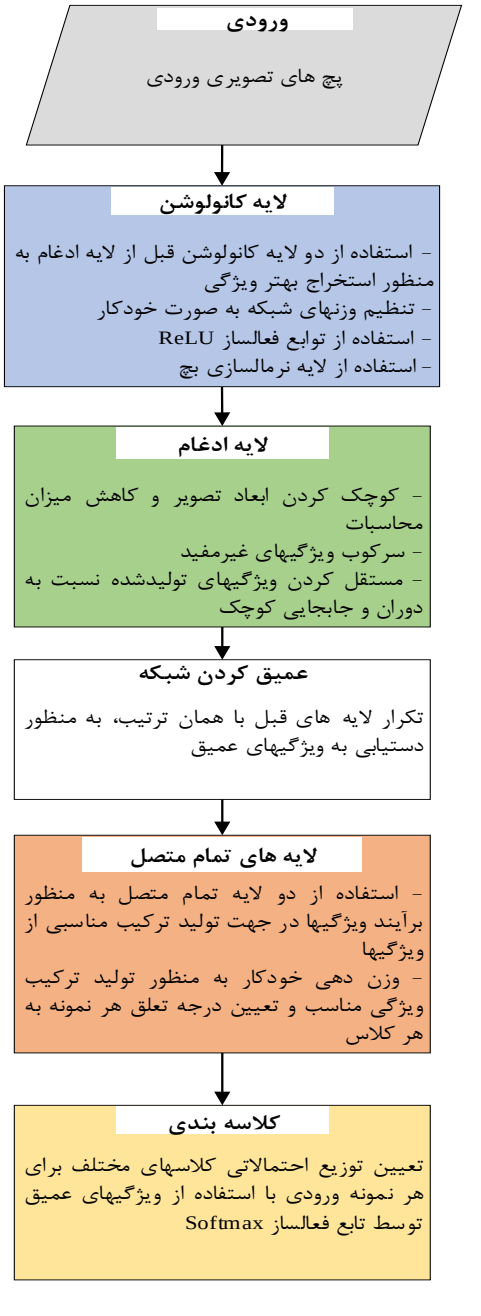
طرح کلی بر پایه‌ی به‌کارگیری پچ‌های تصویری برای کلاسه‌بندی پیکسل مرکزی در پچ شکل گرفته است. بنابراین برای آموزش شبکه، اعتبارسنجی و ارزیابی آن نیاز است تا از این نمونه‌ها در یک همسایگی معین، یک پچ تصویری برداشت گردد. به دلیل امکان استفاده از تمامی

نمونه‌های دارای برچسب، از پدینگ آینه‌ای به عنوان پیش‌پردازش برای تصاویر استفاده گردید. تعداد نمونه‌های آموزشی و اعتبارسنجی مورداستفاده برای هر کلاس در هر دو تصویر به ترتیب حدود ۱۰۰۰ و ۳۰۰ پیکسل بوده است. در واقع تعداد نمونه‌های آموزشی ۱/۴۲ درصد از کل نمونه‌های دارای برچسب در تصویر اول و ۱/۱۱ درصد در تصویر دوم بوده است. این میزان برای داده‌های اعتبارسنجی ۰/۴۴ و ۰/۳۳ درصد، تصویر اول و دوم بوده است. برای ارزیابی نهایی، شبکه پس از آموزش به پیش‌بینی تمام نمونه‌های دارای برچسب در تصاویر پرداخته است. هم‌چنین به عنوان پیش‌پردازش باند آبی به دلیل تاثیر پایین در عملکرد روش و کاهش میزان محاسبات از تصویر اصلی حذف گردید.

۳-۲-۲- طراحی شبکه‌ی عمیق

معماری شبکه به گونه‌ای طراحی گردیده است که قادر باشد ویژگی‌های بهینه و متمایزکننده از صحنه‌های تصویری استخراج کند (شکل ۳). مطابق آن‌چه در شکل (۳) مشاهده می‌گردد، در طراحی شبکه مجموعا از ۴ لایه‌ی کانولوشن استفاده گردیده است. این تعداد به منظور دستیابی به ویژگی‌های عمیق و سطح بالا تعیین گردیده است. در دو لایه‌ی اول از دو لایه‌ی پیاپی کانولوشن استفاده شده تا قبل از ادغام ویژگی‌های توسط لایه‌ی ادغام، امکان استخراج ویژگی‌ها متنوع فراهم گردد. پس از دو لایه‌ی کانولوشن ابتدایی، به منظور افزایش عمومیت‌پذیری شبکه، از لایه‌ی بچ‌نرمالیزاسیون استفاده شده است [۴۵].

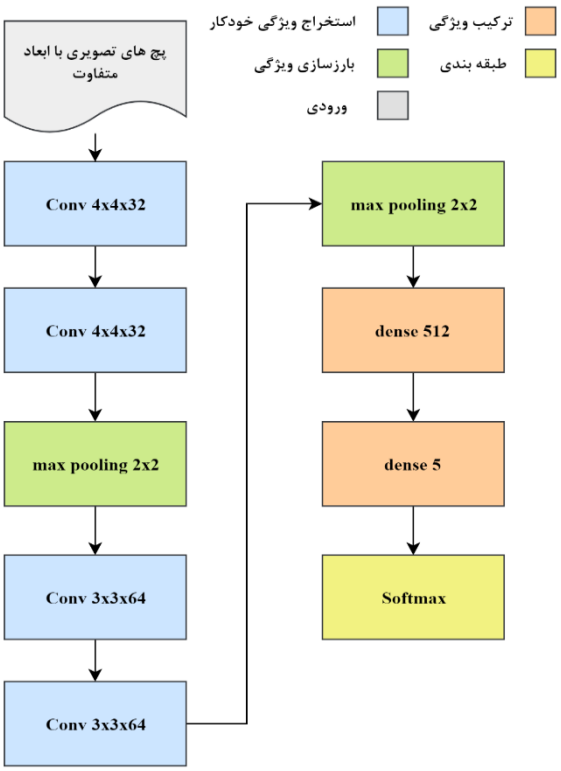
به دلیل تسریع در آموزش شبکه، تابع فعال‌ساز در نظر گرفته شده برای تمامی لایه‌های کانولوشن، تابع فعال‌ساز ReLU است [۴۵]. پس از لایه‌های کانولوشن ابتدایی، از یک لایه‌ی ادغام به منظور بارزسازی ویژگی‌ها و از بین بردن وبژگی‌های ناکارآمد استفاده شده است. تکرار روند کانولوشن-کانولوشن-ادغام به منظور عمیق‌سازی شبکه و استخراج ویژگی‌های سطح بالا انجام گرفته شده است. در نهایت پس از تولید و بارزسازی ویژگی، از دو لایه‌ی تمام‌متصل به منظور وزن‌دهی خودکار به ویژگی‌های تولیدی و محاسبه‌ی امتیاز تعلق هر نمونه ورودی به کلاس‌های مختلف استفاده می‌شود. در شکل (۴) جزئیات طراحی شبکه‌ی عمیق نشان داده شده است. در طراحی لایه‌های کانولوشن، در لایه‌های با عمق کم‌تر از ابعاد بزرگ‌تر و تعداد فیلتر کم‌تری استفاده گردیده است. این در حالی است که در لایه‌های عمیق، ابعاد کوچک‌تر و تعداد فیلترهای تعیین‌شده بیشتر گردیده است.



شکل۳- ساختار کلی شبکه‌ی عمیق طراحی‌شده

دلیل این امر، محدود و کلی بودن ویژگی‌هایی است که در سطوح پایین امکان استخراج دارند. در حالی که هرچه به عمق بالاتر برویم امکان تولید ویژگی‌های جزئی

به تعداد بیش‌تر فراهم خواهد بود [۳۷]. به منظور بارزسازی ویژگی از لایه‌های ادغام بیشترین مقدار استفاده گردیده است. دلیل انتخاب لایه‌ی ادغام بیشترین مقدار، تسریع در همگرایی شبکه، انتخاب ویژگی‌های برجسته‌ی مستقل و بهبود عمومیت‌پذیری شبکه است. هم‌چنین لایه‌ی ادغام ماکزیمم شبکه را نسبت به تغییرات محلی موقعیت مستقل می‌سازد و موجب کوچک شدن اندازه‌ی تصویر ورودی خواهد شد. در انتهای شبکه نیز از دو لایه‌ی تمام متصل به منظور ایجاد ترکیب‌های بهینه از ویژگی‌های تولیدشده، استفاده گردیده است. در لایه‌ی تمام متصل اول تعداد زیادی نورون به منظور ایجاد امکان ترکیب غیرخطی ویژگی‌ها برای دسترسی به برآیند ویژگی متمایزکننده قرار داده شده است.



شکل۴- ساختار کلی شبکه‌ی عمیق طراحی‌شده

در لایه‌ی تمام متصل دوم به منظور محاسبه‌ی امتیاز تعلق نهایی نمونه به هر کلاس از تعداد نورون‌های برابر با تعداد کلاس‌های موجود در تصویر استفاده می‌گردد. تابع فعال‌سازی لایه‌ی تمام متصل اول تابع ReLU و لایه‌ی تمام متصل دوم تابع Softmax است.

برای آموزش شبکه از شیوه پس‌انتشار خطا و روش گرادیان نزولی تصادفی [۴۶] استفاده شده است. تعداد اپک برابر ۵۰ و نرخ آموزش برابر ۰/۰۱ تعیین گردیده است. به منظور ایجاد تعادل در نرخ آموزش، مقدار کاهشی (نرخ آموزش / اپک) به آن در هر اپک اعمال می‌گردد. مقدار ممان آموزش برابر ۰/۹ و از روش شتاب‌دهنده‌ی گرادیان نستروف [۴۶] برای تنظیم ممان آموزش استفاده گردیده است. اندازه بچ ورودی به شبکه نیز در هر تکرار برابر ۶۴ تنظیم گردیده است. اندازه بچ ورودی به شبکه در هر تکرار معمولا بین ۵۰ تا ۲۵۶ متناسب با کاربرد شبکه در نظر گرفته می‌شود [۴۶].

فرآیند آموزش شبکه طبق معماری و فراپارامترهای تعیین‌شده برای شبکه انجام می‌گیرد. سپس دو رویکرد متفاوت در استفاده از شبکه‌ی آموزش‌دیده، بررسی می‌گردد. رویکرد اول از شبکه به صورت یکپارچه برای استخراج ویژگی و در ادامه کلاسه‌بندی تصاویر بهره می‌برد و رویکرد دوم از شبکه تنها به عنوان استخراج‌کننده ویژگی استفاده می‌کند و کلاسه‌بندی توسط روش ماشین بردار پشتیبان صورت می‌پذیرد. نتایج حاصله از هر رویکرد برای هر کدام از تصاویر در ادامه بررسی خواهد شد.

۳-۴- ارزیابی و تحلیل نتایج

در مرحله‌ی ارزیابی و تحلیل نتایج، نتایج حاصل از در نظر گرفتن شعاع‌های متفاوت برای صحنه‌های تصویری و هم‌چنین رویکردهای متفاوت در استفاده از شبکه‌های عمیق به دست آمده است. نتایج حاصله به صورت مجزا مورد بررسی قرار گرفته و تحلیل‌های مربوطه ارائه گردیده است.

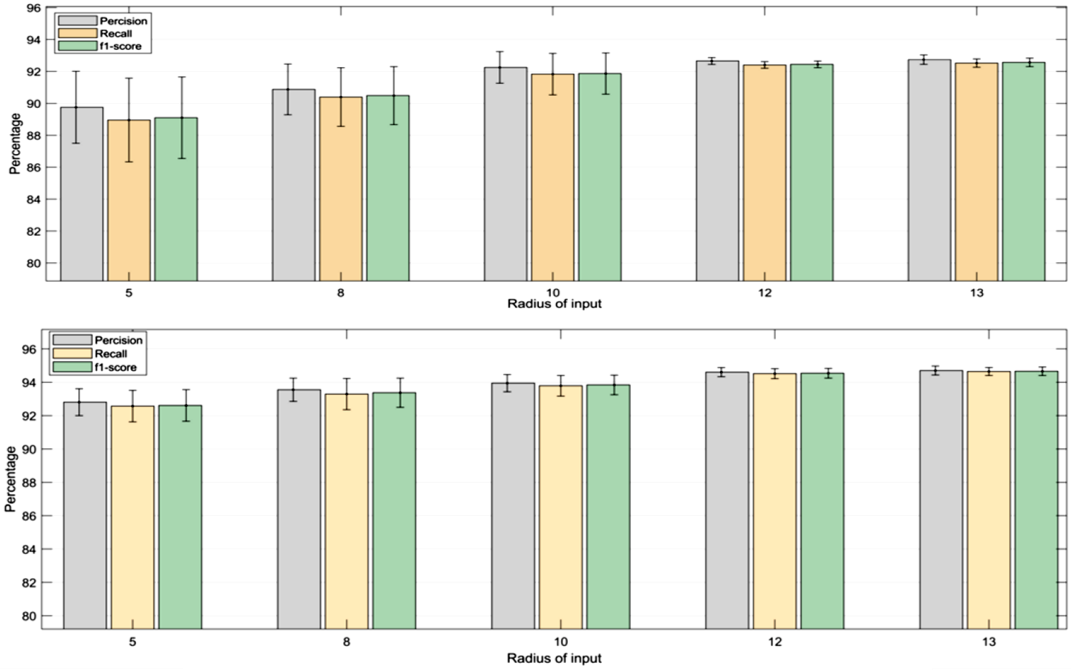
۳-۴-۱- بررسی توانایی شبکه‌ی طراحی‌شده و تاثیر ابعاد پچ‌های تصویری

در این مورد، پچ‌های تصویری با در نظر گرفتن ابعاد متفاوت به عنوان ورودی به شبکه داده می‌شود. ابعاد در نظر گرفته شده برای پچ‌های تصویری ۱۱×۱۱، ۱۷×۱۷، ۲۱×۲۱، ۲۵×۲۵ و ۲۷×۲۷ است. آموزش شبکه توسط پچ‌های تصویری با ابعاد متفاوت، برای هر دو تصویر انجام گرفته و نتایج حاصل از کلاسه‌بندی تصاویر به دست آمده است.نتایج نمایش داده شده، میانگین و انحراف معیار حاصل از سه بار آموزش شبکه و پیش‌بینی توسط آن را نشان می‌دهد.

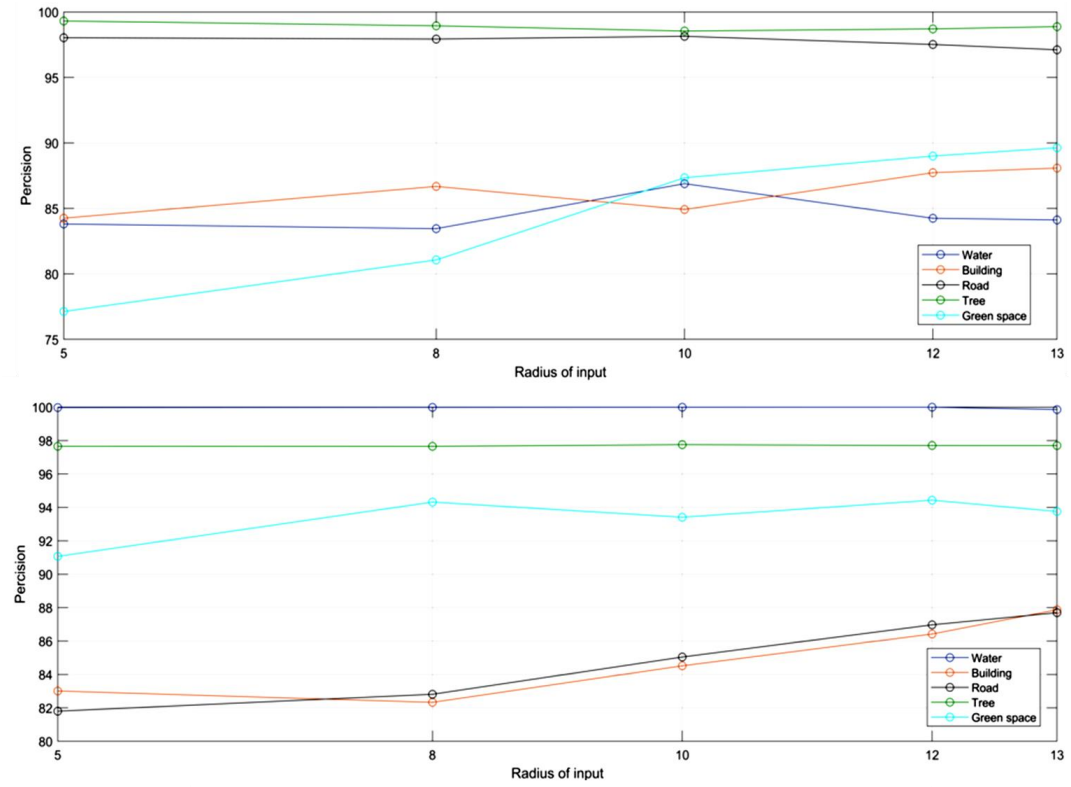
نتایج حاصله برای هر دو تصویر، نشان از افزایش دقت میانگین برای هر سه معیار ارزیابی در ازای افزایش ابعاد

پچ ورودی دارد. این امر نشان‌دهنده تاثیر مثبت در نظر گرفتن ابعاد بزرگ‌تر بافت در نتایج نهایی کلاسه‌بندی است. هم‌چنین با افزایش ابعاد در پچ‌های تصویری ورودی، انحراف معیار سه بار آموزش شبکه در نتایج به دست آمده برای هر دو تصویر با کاهش همراه بوده است که نشان‌دهنده‌ی آموزش بهتر شبکه در این حالت است.

در شکل (۵) میانگین و انحراف معیار دقت به دست آمده برای دو تصویر نشان داده شده است. رفتار کلاس‌های مختلف به ازای ابعاد متفاوت در پچ‌های تصویری ورودی به شبکه در هر دو تصویر مورد آزمایش، در شکل (۶) قابل مشاهده است.



شکل ۵- نمودار دقت میانگین و انحراف معیار در کلاسه‌بندی تصویر اول(بالا) و تصویر دوم(پایین) به ازای ابعاد متفاوت پچ‌های ورودی



شکل ۶- نمودار دقت میانگین کلاس‌های مختلف در کلاسه‌بندی تصویر اول (بالا) و تصویر دوم(پایین) به ازای ابعاد متفاوت پچ‌های ورودی

در تصویر اول، کلاسه‌بندی کلاس فضای سبز با بیشترین بهبود دقت در ازای افزایش ابعاد پچ‌های ورودی همراه بوده است. این کلاس از دقت حدود ۷۵ درصد در ابعاد ۱۱×۱۱ به دقت حدود ۹۰ درصد در ابعاد ۲۷×۲۷ رسیده است. کلاس ساختمان نیز در حالت کلی با افزایش دقت کلاسه‌بندی همراه بوده است. در سه کلاس دیگر به ازای افزایش در ابعاد پچ‌های ورودی، میزان دقت کلاسه‌بندی کلاس موردنظر در حالت کلی تغییر قابل ملاحظه‌ای نداشته است.

در تصویر دوم، کلاسه‌بندی کلاس‌های ساختمان و راه با افزایش ابعاد در پچ‌های ورودی، به صورت آشکار با بهبود دقت همراه بوده‌اند. کلاس فضای سبز نیز در حالت کلی با افزایش ابعاد، بهتر کلاسه‌بندی شده و نتایج دقیق‌تری به همراه داشته است. کلاس‌ها آب و درخت در ابعاد مختلف در نظر گرفته شده برای صحنه‌های ورودی، دقت کلاسه‌بندی بالایی داشته‌اند و میزان این دقت در روند افزایش ابعاد تغییر چندانی نداشته است. در مجموع، با در نظر گرفتن این موضوع که تعدادی از کلاس‌ها با توجه به ابعاد جسم در سطح زمین در اندازه‌ی مشخصی از ابعاد ورودی به دقت بهینه رسیده‌اند، در حالت کلی می‌توان مشاهده نمود که ابعاد بزرگ‌تر برای پچ‌های ورودی منجر به بهبود وضعیت گردیده است.

با افزایش ابعاد در پچ‌های تصویری ورودی، اشتباهات شبکه در طبقه‌بندی کلاس‌های مشابه کاهش یافته است. در تصویر دوم، در ابعاد ۱۱×۱۱ بیشترین اشتباه در پیش‌بینی شبکه مربوط به طبقه‌بندی نادرست کلاس ساختمان به جای کلاس راه بوده است. با در نظر گرفتن ابعاد کوچک برای صحنه‌های ورودی درواقع توانایی شبکه در تولید ویژگی‌های متمایزکننده بین کلاس‌های مشابه محدود گردیده است.

وقتی ابعاد صحنه‌های تصویری ورودی به شبکه در تصویر دوم به ۲۷×۲۷ افزایش می‌یابد، میزان اشتباه در طبقه‌بندی کلاس ساختمان به جای راه در تصویر دوم به کمتر از نصف کاهش یافته است. در واقع با در نظر گرفتن همسایگی بزرگتر در اطراف پیکسل مرکزی این توانایی در هنگام آموزش در شبکه ایجاد گردیده است که با دید بزرگ‌تری در اطراف پیکسل بتواند ویژگی‌های متنوع و متمایز را از بافت اطراف کلاس تولید نماید تا از این طریق دقت طبقه‌بندی به خصوص در کلاس‌هایی که شباهت بین‌کلاسی بالایی دارند، افزایش یابد.

۳-۴-۲- بررسی رویکردهای متفاوت در استفاده از شبکه‌های عصبی کانولوشنال

در رویکرد رویکرد اول شبکه‌ی عصبی کانولوشنی که وزن‌های قابل تنظیم در آن به روش نظارت‌شده تنظیم گردیده است به عنوان استخراج‌کننده ویژگی و کلاسه‌بند مورد استفاده قرار می‌گیرد. در این حالت از تابع Softmax برای تعیین احتمال عضویت نمونه ورودی به هر کدام از کلاس‌های موجود در تصویر استفاده می‌گردد و شبکه یکپارچه خواهد بود.

در رویکرد دوم شبکه‌های عصبی کانولوشنی به عنوان استخراج‌کننده ویژگی مورداستفاده قرار می‌گیرند. به این صورت که، در اپک نهایی فرآیند آموزش شبکه، ویژگی‌های تولیدشده از لایه‌ی آخر استخراج گردیده و برای کلاسه‌بندی به ماشین بردار پشتیبان انتقال داده می‌شوند.

ماشین بردار پشتیبان ویژگی‌های تولیدشده توسط شبکه‌ی عصبی کانولوشنی را دریافت کرده و از روش نظارت‌شده به پیش‌بینی کلاس نمونه‌های تست می‌پردازد. کرنل در نظر گرفته شده برای ماشین بردار پشتیبان، کرنل rbf بوده و دو پارامتر C و گاما توسط روش جستجوی شبکه‌ای برای این کلاسه‌بند تعیین گردیده است.

در جدول (۱) نتایج حاصل از اعمال دو رویکرد متفاوت برای تمامی ابعاد پچ‌های تصویری اشاره‌شده در بخش ۳-۳-۱ گردآوری گردیده است. برای این منظور، با در نظر گرفتن ابعاد مختلف برای پچ‌های تصویری، ابتدا شبکه‌ی عصبی کانولوشنی، طبق معماری و فراپارامترهای معرفی‌شده، آموزش داده شده است. سپس این شبکه‌ی آموزش‌دیده، یک بار به صورت یکپارچه کلاسه‌بندی تصویر را انجام می‌دهد و بار دیگر از همین شبکه، بدون آموزش مجدد، تنها به عنوان استخراج‌گر ویژگی استفاده می‌گردد و ویژگی‌های تولیدی برای کلاسه‌بندی به کلاسه‌بند مجزا وارد می‌شود. از این منظر فرآیند آموزش شبکه نمی‌تواند تاثیری در عملکرد هر کدام از دو رویکرد متفاوت به کار گرفته داشته باشد.

با توجه به استفاده از شبکه‌های آموزش‌دیده یکسان برای مقایسه‌ی دو رویکرد، تنها از یک بار اجرا برای مقایسه استفاده گردیده است. به همین سبب نتایج نشان

داده شده در جدول (۱) صرفا معیار ارزیابی برای رویکردهای موردنظر با توجه به استفاده از شبکه‌های آموزش‌دیده یکسان برای مقایسه‌ی دو رویکرد، تنها از یک بار اجرا برای مقایسه استفاده گردیده است. به همین سبب نتایج نشان داده شده در جدول (۱) صرفا معیار ارزیابی برای رویکردهای موردنظر می‌باشد. هم‌چنین، برای ارزیابی نتایج سه معیار ارزیابی Precision، recall و f1-score محاسبه گردیده است.

جدول ۱- ارزیابی دقت نتایج رویکردهای ارائه‌شده

روش‌های مختلف	معیار دقت	تصویر تست اول					تصویر تست دوم				
		ابعاد پچ ورودی به شبکه									
		۱۱×۱۱	۱۷×۱۷	۲۱×۲۱	۲۵×۲۵	۲۷×۲۷	۱۱×۱۱	۱۷×۱۷	۲۱×۲۱	۲۵×۲۵	۲۷×۲۷
CNN	Precision (%)	۹۰,۶۸	۸۹,۷۳	۹۲,۵۷	۹۲,۸۹	۹۲,۶۰	۹۳,۸۴	۹۳,۹۱	۹۴,۰۴	۹۴,۶۳	۹۴,۶۳
	Recall (%)	۹۰,۱۶	۸۹,۱۶	۹۲,۲۵	۹۲,۶۷	۹۲,۳۲	۹۳,۷۲	۹۳,۷۶	۹۳,۹۳	۹۴,۵۴	۹۴,۵۳
	F1-score (%)	۹۰,۲۶	۸۹,۳۳	۹۲,۳۰	۹۲,۷۱	۹۲,۳۷	۹۳,۷۵	۹۳,۸۱	۹۳,۹۶	۹۴,۵۷	۹۴,۵۶
CNN+SVM	Precision (%)	۹۰,۸۲	۹۰,۴۳	۹۲,۵۶	۹۳,۰۰	۹۲,۵۳	۹۳,۹۶	۹۴,۰۶	۹۴,۰۸	۹۴,۴۵	۹۴,۵۸
	Recall (%)	۹۰,۴۰	۸۹,۹۴	۹۲,۲۲	۹۲,۸۰	۹۲,۲۵	۹۳,۸۷	۹۳,۹۳	۹۳,۹۸	۹۴,۳۱	۹۴,۴۸
	F1-score (%)	۹۰,۴۹	۹۰,۰۷	۹۲,۲۷	۹۲,۸۳	۹۲,۲۹	۹۳,۹۱	۹۳,۹۷	۹۴,۰۲	۹۴,۳۶	۹۴,۵۲

با توجه به نتایج ارائه‌شده در جدول (۱)، در اکثر موارد آزمایش‌شده، نتایج در حالت استخراج ویژگی توسط شبکه‌های عصبی کانولوشنی و کلاسه‌بندی توسط ماشین بردار پشتیبان، با بهبود در دقت کلاسه‌بندی همراه بوده است. این بهبود در دقت در حالت کلی برای هر دو تصویر اتفاق افتاده است. اگرچه تاثیر مثبت در استفاده از ماشین بردار پشتیبان لزوما در همه‌ی موارد صورت نگرفته است، اما در حالت کلی با در نظر گرفتن تمام موارد آزمایش‌شده بهبود دقت در نتایج صورت پذیرفته است. با توجه به آزمایشات صورت‌گرفته و نتایج آن، کلاسه‌بندی تصاویر هوایی مربوط به مناطق نیمه‌شهری توسط ویژگی‌های استخراج‌شده از شبکه‌ی عمیق و کلاسه‌بندی ماشین بردار پشتیبان به نسبت شبکه‌ی کانولوشنی یک‌پارچه در حالت کلی، اما نه لزوما در تمام موارد، موجب بهبود در دقت حاصله گردیده است.

۳-۴-۳- مقایسه روش پیشنهادی با روش‌های مطرح در حوزه یادگیری عمیق

برای ارزیابی نهایی روش، نتایج به دست آمده توسط شبکه‌ی طراحی‌شده و با در نظر گرفتن بهترین حالات نتایج به‌دست آمده در آزمایش تعیین ابعاد مناسب برای ورودی و بهترین رویکرد استفاده از شبکه‌های عصبی کانولوشنی با دو روش مرسوم در زمینه‌ی یادگیری عمیق شامل روش‌های شبکه‌ی عصبی مصنوعی چندلایه و استفاده از شبکه‌های پیش‌آموزش برای دو تصویر موردمطالعه در مقاله مقایسه گردیده است.

در روش پشنهادی برای مقایسه، برای هر دو تصویر ابعاد ۲۵×۲۵ و رویکرد استفاده از شبکه‌های یک‌پارچه برای استخراج ویژگی و کلاسه‌بندی در نظر گرفته شده است. مطابق جدول (۱) در این حالت بهترین دقت حاصله برای دو تصویر به دست آمده است.

هم‌چنین در روش شبکه‌های عصبی چندلایه، در طی چند مرحله آزمون و خطا سعی شده است بهترین حالت استفاده از این شبکه‌ها در نظر گرفته شود. نرخ آموزش در شبکه ۰,۱ تنظیم گردیده است. برای استفاده از این شبکه‌ها از روش آموزش گرادیان نزولی تصادفی استفاده شده است و برای شبکه، تابع فعال‌ساز tanh در نظر گرفته شده است. سه لایه با تعداد نورون ۵۰، ۳۰۰ و ۵۰ به ترتیب به عنوان سه لایه‌ی میانی در تعیین معماری شبکه در نظر گرفته شده است. از آنجایی که ورودی شبکه‌های عصبی چندلایه بر خلاف روش شبکه‌های عصبی کانولوشن برداری می‌باشد، با در نظر گرفتن یک همسایگی ۷×۷، پچ تصویری به دست‌آمده به شکل بردار تبدیل و به عنوان ورودی به شبکه وارد شده است. لازم به ذکر است در آزمایشات متفاوت صورت گرفته، ابعاد ۷×۷ بهترین نتایج را در حالت استفاده از شبکه‌های عصبی چندلایه به همراه داشته است. در این حالت از هر ۴ باند تصویر استفاده شده است.

در ارزیابی دیگر، از شبکه‌های پیش‌آموزش برای مقایسه‌ی نتایج با شبکه‌ی عصبی کانولوشنی طراحی‌شده استفاده گردیده است. شبکه‌ی پیش‌آموزش vgg16 به دلیل تناسب ابعاد ورودی شبکه با ساختار مسئله، برای این کار در نظر گرفته شده است.

وزن‌های موجود در شبکه در حالت آموزش بر روی مجموعه داده‌ی imagenet تعیین گردیده است. برای تطبیق بهتر روش با نمونه‌های آموزشی، در طراحی ساختار شبکه، لایه‌ی آخر موجود در شبکه که لایه‌ی کلاسه‌بندی در آزمایش بر روی داده‌های آزمایشی imagenet است حذف گردیده و به جای آن دو لایه‌ی تمام متصل با تعداد نورون ۲۵۶ و ۵ به شبکه اضافه گردیده است. در فرآیند آموزش شبکه توسط نمونه‌های آموزشی فقط لایه‌ی آخر موجود در ساختار شبکه و دو لایه‌ی اضافه شده به ساختار شبکه شرکت داده شده‌اند و بقیه لایه‌ها دارای وزن‌های ثابت هستند. این امر به دلیل عمومیت نسبی ویژگی‌های تولیدی در لایه‌های ابتدایی و میانی و هم‌چنین عدم تاثیرگذاری در وزن‌های موجود در این لایه‌ها با تعداد کم نمونه‌های آموزشی موجود در شبکه است. مشخصات آموزش در شبکه برای لایه‌های انتهایی تعیین شده به عنوان لایه‌های قابل‌آموزش، با مشخصات آموزشی در نظر گرفته شده برای شبکه‌های عصبی کانولوشنی در بخش ۲-۲-۳ یکسان است. برای ورودی شبکه پچ به ابعاد ۵۰×۵۰ در نظر گرفته شده است. شبکه‌های پیش‌آموزش با توجه به ابعاد تصاویری که توسط آن‌ها آموزش دیده‌اند، ابعاد خاص ورودی برای شبکه دارند، از این رو پچ‌های تصویری با ابعاد پایین‌تر از ۴۷×۴۷ برای شبکه vgg16 امکان‌پذیر نمی‌باشد. ابعاد لایه‌های موجود در شبکه متناسب با ابعاد ورودی است. در موارد آزمایشی به کار گرفته شده در این مقاله، که شامل دو تصویر هوایی با قدرت تفکیک مکانی ۱ متر از یک منطقه نیمه‌شهری بوده است و با در نظر گرفتن ۵ کلاس محدود شامل کلاس‌های ساختمان، راه، آب، فضای سبز و درخت، نتایج به دست‌آمده در جدول (۲) برتری روش پیشنهادی را نسبت به روش‌های شبکه‌ی عصبی چندلایه و استفاده از شبکه‌های پیش‌آموزش در کلاسه‌بندی هر دو تصویر تست در هر سه معیار ارزیابی نشان می‌دهد.

جدول ۲- قیاس با روش‌های مطرح

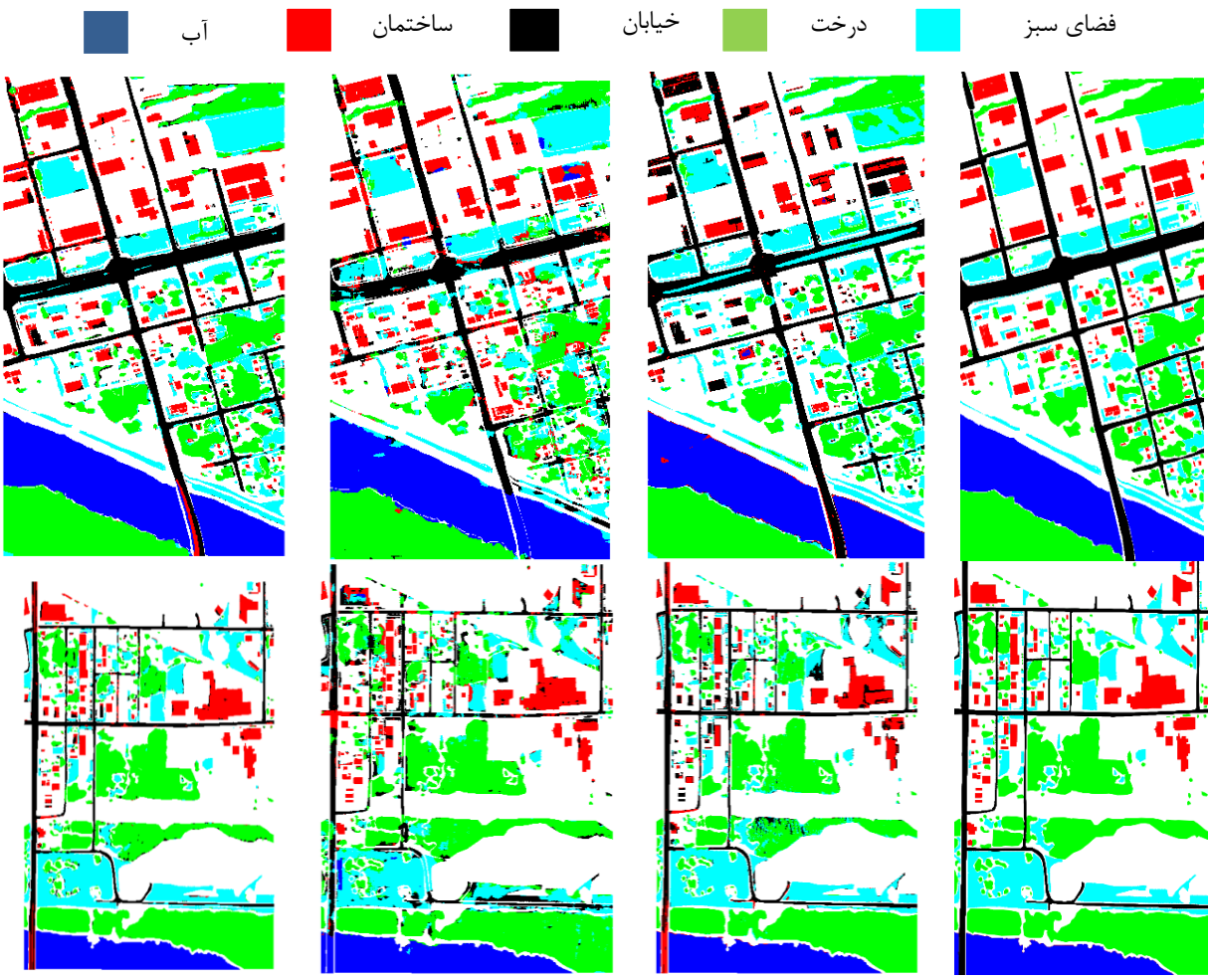
روش‌های مختلف	معیار دقت (درصد)	تصویر تست اول	تصویر تست دوم
روش پیشنهادی	Precision	۹۲,۸۹	۹۴,۶۳
	Recall	۹۲,۶۷	۹۴,۵۴
	F1-score	۹۲,۷۱	۹۴,۵۷
MLP	Precision	۸۶,۱۸	۹۰,۴۵
	Recall	۸۵,۵۲	۸۹,۸۹
	F1-score	۸۵,۱۲	۹۰,۰۷
VGG16	Precision	۸۱,۱۶	۸۲,۲۵
	Recall	۸۰,۲۷	۸۱,۹۷
	F1-score	۸۰,۳۲	۸۱,۹۲

۳-۴-۴- مقایسه نسبی و کیفی

برای ارزیابی روش شبکه‌ی عصبی کانولوشنی با ساختار پیشنهادی نسبت به روش‌های مرسوم در یادگیری عمیق، از آزمایش مک‌نمار[۱] استفاده شده است. این آزمون جهت بررسی اعتبار قیاس عددی نتایج و عدم وابستگی آنهاست [۴۷]. معیار zb در این آزمون برای روش پیشنهادی به روش mlp برای تصویر اول ۱۰۶,۴۵ و برای تصویر دوم ۱۰۵,۷۱ بوده است. هم‌چنین این معیار برای روش پیشنهادی به شبکه‌ی پیش‌آموزش vgg16 برای تصویر اول ۱۶۴,۵۱ و برای تصویر دوم ۱۹۹,۷۴ بوده است. نتایج نشان از استقلال آماری نتایج و در نتیجه تایید برتری کمی ذکر شده روش پیشنهادی به دو روش دیگر دارد.

در شکل (۷) نتایج بصری به دست آمده از روش‌های مطرح‌شده برای دو تصویر نمایش داده شده است. همانطور که در شکل قابل مشاهده است تفکیک کلاس‌های مشابه نظیر درخت و فضای سبز و هم‌چنین کلاس‌های راه و ساختمان و تعیین کلاس درست در مناطقی که دو کلاس روی هم قرار گرفته‌اند (جاده‌های پوشیده‌شده توسط درختان) چالش‌های عمده پیش‌رو بوده است. روش شبکه‌ی عصبی کانولوشنی با ساختار پیشنهادی نسبت به روش‌های mlp و شبکه‌ی پیش‌آموزش vgg16 بهتر عمل کرده است.

۱ Mcnemar test



الف) داده واقعیت زمینی ب) MLP د) شبکه پیش آموزش و) روش پیشنهادی

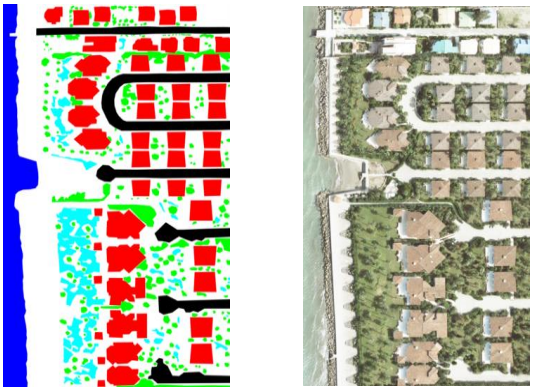
شکل ۷- نتایج طبقه‌بندی مناطق تست با استفاده از روش پیشنهادی تحقیق در مقایسه با سایر روش‌ها

۳-۴-۵- اجرای روش پیشنهادی بر روی داده داخل کشور و مقایسه با روش ماشین بردار پشتیبان

به منظور بومی‌سازی روش و ایجاد تنوع در داده‌های مورداستفاده در مقاله، الگوریتم پیشنهادی برای کلاسه‌بندی تصویری از منطقه رویان در استان مازندران به کار گرفته شده است. تصویر مورد اشاره قدرت تفکیک مکانی بالاتری نسبت به تصویر دس موینس استفاده شده در مقاله دارد، از این رو می‌توان ارزیابی خوبی نسبت به توانایی الگوریتم در تصاویر با قدرت تفکیک مکانی بالاتر به دست آورد.

هم‌چنین تصویر مورداستفاده دارای ۴ باند قرمز، سبز، آبی و مادون قرمز است. ابعاد تصویر ۳۸۱۸×۲۱۱۱ می‌باشد و کلاسه‌بندی برای ۵ کلاس آب، ساختمان، راه، درخت و فضای سبز انجام گرفته است. حدود ۱,۳ درصد از پیکسل‌های مورد آزمایش در تصویر به عنوان نمونه‌ی آموزشی و ۰,۱۳ درصد به عنوان نمونه‌ی ارزیابی در نظر گرفته شده است.

در این بخش به مقایسه روش پیشنهادی با روش ماشین بردار پشتیبان در زمینه‌ی کلاسه‌بندی پرداخته شده است. در روش پیشنهادی پچ‌های ورودی با ابعاد ۲۷×۲۷ در نظر گرفته شده است و ویژگی‌های تولیدشده برای کلاسه‌بندی به ماشین بردار پشتیبان وارد می‌گردد.



تصویر از منطقه رویان داده واقعیت زمینی

شکل ۸- منطقه مطالعاتی رویان

نتایج حاصل از این روش با نتایج به دست آمده توسط روش ماشین بردار پشتیبان در حالت ورود مقادیر پیکسلی به عنوان ویژگی‌های ورودی مقایسه گردیده است. با این توضیح که برای هر دو حالت استفاده از ماشین بردار پشتیبان از کرنل rbf و دو پارامتر C و گاما توسط روش جستجوی شبکه‌ای برای این کلاسه‌بند تعیین گردیده است.

نتایج به دست آمده در کلاسه‌بندی تصویر مربوط به منطقه‌ی رویان در روش پیشنهادی در جدول (۳)، نشان از برتری این روش نسبت به روش متداول ماشین بردار پشتیبان دارد که نشان‌دهنده‌ی تاثیر قابل توجه ویژگی‌های عمیق تولیدشده توسط مدل عمیق طراحی شده در بهبود هر سه معیار ارزیابی است.

جدول ۳- نتایج به دست آمده برای داده رویان

نتایج تصویر رویان	معیار دقت	روش‌های مختلف
۹۵,۶۷	Precision (%)	روش پیشنهادی (CNN+SVM)
۹۶,۱۷	Recall (%)	
۹۶,۲۵	F1-score (%)	
۸۸,۸۰	Precision (%)	SVM
۹۰,۸۰	Recall (%)	
۸۹,۶۲	F1-score (%)	

۴- نتیجه‌گیری

کلاسه‌بندی تصاویر هوایی همواره با چالش‌های زیادی روبرو بوده است. از طرفی با پیشرفت تکنولوژی و افزایش قدرت تفکیک مکانی تصاویر، با وجود افزایش کلی دقت، چالش‌های مربوط به کلاسه‌بندی این تصاویر نیز تشدید گردیده است. تفاوت در نمونه‌های مختلف متعلق به یک کلاس و شباهت‌های موجود در اشیای متعلق به کلاس‌های متفاوت، عملکرد نسبی الگوریتم‌های کلاسه‌بندی را تحت تاثیر قرار داده است. در تصاویر در نظر گرفته شده، تفکیک کلاس‌های مشابه (نظیر درخت و فضای سبز)، تعیین کلاس درست در مناطقی که دو کلاس روی هم قرار گرفته‌اند (جاده‌های پوشیده‌شده توسط درختان) و همچنین سایه‌های موجود در تصویر چالش‌های عمده پیش‌رو بوده است.

در ساختار طراحی‌شده با استفاده از روش‌های یادگیری عمیق در حالت استفاده از پچ‌های تصویری، ویژگی‌های عمیق و متمایزکننده از تصاویر استخراج گردیده است. ویژگی‌های استخراج‌شده از تصاویر به صورت خودکار و در ساختاری سلسله‌مراتبی تولید شده است. در قدم اول، هدف تعیین ساختار و معماری عمیق به منظور استخراج بهینه ویژگی‌ها از تصویر است. برای این منظور در تعیین ساختار لایه‌ای شبکه، قیود خاصی جهت استخراج، بارزسازی و ترکیب ویژگی‌ها در حالت بهینه و متناسب با ساختار داده‌های تصاویر هوایی با قدرت تفکیک مکانی بالا در نظر گرفته شده است. ساختار نهایی تعیین‌شده برای شبکه به دقت ۹۲,۶۰ درصد برای تصویر تست اول و ۹۴,۶۳ درصد برای تصویر تست دوم از داده دس موینس و به دقت ۹۵,۶۷ درصد برای داده‌ی رویان رسیده است که در مقایسه با روش‌های مورد آزمایش بهترین نتیجه را داشته است. این نتایج برای تصاویر مورد آزمایش با شرایط مطرح‌شده به دست آمده است.

ابعاد پچ‌های تصویری ورودی به شبکه از عوامل تاثیرگذار در عملکرد استخراج ویژگی توسط شبکه می‌باشد. در واقع ابعاد پچ‌های ورودی برای مشارکت در تصمیم‌گیری کلاس پیکسل موردنظر تعیین می‌گردد. برای بررسی میزان و نحوه‌ی تاثیرگذاری ابعاد همسایگی برای هر دو تصویر در نظر گرفته‌شده از منطقه des moines، ابعاد ۱۱×۱۱، ۱۷×۱۷، ۲۱×۲۱، ۲۵×۲۵ و ۲۷×۲۷ برای پچ‌های ورودی در نظر گرفته شد و شبکه توسط این ابعاد متفاوت به صورت جداگانه آموزش دید و به صورت یک‌پارچه کلاسه‌بندی داده‌های آزمایشی توسط شبکه‌ی آموزش‌دیده انجام گرفت. سه معیار precision، recall و f1-score برای ارزیابی نتایج حاصله به دست آمده است. نتایج به صورت میانگین و انحراف معیار برای سه بار اجرای برنامه تهیه شده است. در تصویر اول بیشترین میزان دقت در ابعاد ۲۷×۲۷ به میزان ۹۲,۷۴ درصد و کمترین میزان ۸۹,۷۵ درصد و مربوط به ابعاد ورودی ۱۱×۱۱ می‌باشد. در تصویر دوم نیز بیشترین میزان دقت ۹۴,۷۰ درصد در ابعاد ورودی ۲۷×۲۷ و کمترین میزان دقت ۹۲,۸۰ درصد و در ابعاد ۱۱×۱۱ بوده است. این آزمایشات برای دو تصویر از منطقه دس موینس صورت گرفت. هم‌چنین، انحراف معیار در سه بار اجرا برای ابعاد ورودی گوناگون، با افزایش ابعاد کاهش یافته که خود نشان از تولید ویژگی‌های مستحکم در حالت ابعاد ورودی بزرگ‌تر و آموزش بهتر شبکه در این حالت دارد. نتایج آزمایش نشان می‌دهد که افزایش همسایگی مشارکت داده شده موجب بهبود عملکرد شبکه شده است. این بهبود عملکرد با افزایش فاصله از پیکسل مرکزی کم‌تر شده است.

هم‌چنین دو رویکرد متفاوت در زمینه استفاده از شبکه‌های عصبی کانولوشنی در کلاسه‌بندی متراکم تصاویر هوایی به کار گرفته شده است. در رویکرد اول از این شبکه‌ها در حالت یکپارچه برای استخراج ویژگی و در ادامه کلاسه‌بندی استفاده شده است و در رویکرد دوم ویژگی‌های استخراج‌شده توسط شبکه‌های عمیق به عنوان ورودی به ماشین بردار پشتیبان برای کلاسه‌بندی وارد می‌گردد. آزمایشات برای هر دو رویکرد در همه‌ی ابعاد در نظر گرفته شده برای پچ‌های ورودی در مرحله‌ی قبل اجرا گردیده است. نتایج در حالت کلی نشان از بهبود دقت در رویکرد دوم دارد.

قابل ذکر است همه‌ی نتایج به دست آمده در مقاله، برای دو تصویر هوایی با قدرت تفکیک مکانی یک متر در یک منطقه نیمه‌شهری در منطقه‌ی دس موینس و هم‌چنین یک تصویر از منطقه‌ی رویان واقع در استان مازندران و برای پنج کلاس ساختمان، راه، آب، درخت و فضای سبز به دست آمده است. در تحقیقات آینده بررسی‌های صورت گرفته در این مقاله برای شرایط تصویری متفاوت اعم از تصاویر با قدرت تفکیک مکانی، رادیومتریکی و طیفی متفاوت، تصاویر نویزدار، تصاویری که دارای سایه ابر است، تصاویر مایل و برای کلاس‌های متفاوت بررسی خواهد گردید.

مراجع

[1] Gómez-Chova, L., et al., Multimodal classification of remote sensing images: A review and future directions. Proceedings of the IEEE, 2015. 103(9): p. 1560-1584.

[2] Cheng, G., J. Han, and X. Lu, Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE, 2017. 105(10): p. 1865-1883.

[3] Michalski, R.S., J.G. Carbonell, and T.M. Mitchell, Machine learning: An artificial intelligence approach. 2013: Springer Science & Business Media.

[4] Li, Y., et al., Deep learning for remote sensing image classification: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018: p. e1264.

[5] Paoletti, M., et al., A new deep convolutional neural network for fast hyperspectral image classification. ISPRS Journal of Photogrammetry and Remote Sensing, 2017.

[6] Li, M., et al., A review of remote sensing image classification techniques: The role of spatio-contextual information. European Journal of Remote Sensing, 2014. 47(1): p. 389-411.

[7] Ghamisi, P., et al., Advanced spectral classifiers for hyperspectral images: A review. IEEE Geoscience and Remote Sensing Magazine, 2017. 5(1): p. 8-32.

[8] Chen, Y., et al., Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing, 2016. 54(10): p. 6232-6251.

[9] Vetrivel, A., et al., Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS journal of photogrammetry and remote sensing, 2018. 140: p. 45-59.

[10] Zhang, H., et al., Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. Remote Sensing Letters, 2017. 8(5): p. 438-447.

[11] He, K., et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. in European conference on computer vision. 2014. Springer.

[12] Li, Y., H. Zhang, and Q. Shen, Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sensing, 2017. 9(1): p. 67.

[13] abdi, G. and F. samadzadegan, Classification of Aerial Visible-Thermal Data Based on Deep Learning Models. Journal of geomatic science and technology, 1397.

[14] Zhu, X.X., et al., Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 2017. 5(4): p. 8-36.

[15] Cheng, G., et al. Learning coarse-to-fine sparselets for efficient object detection and scene classification. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[16] Penatti, O.A., K. Nogueira, and J.A. dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? in Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2015.

[17] Cheng, G., et al., Object detection in remote sensing imagery using a discriminatively trained mixture model. ISPRS Journal of Photogrammetry and Remote Sensing, 2013. 85: p. 32-43.

[18] Swain, M.J. and D.H. Ballard, Color indexing. International journal of computer vision, 1991. 7(1): p. 11-32.

[19] Ojala, T., M. Pietikainen, and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 2002. 24(7): p. 971-987.

[20] Oliva, A. and A. Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of computer vision, 2001. 42(3): p. 145-175.

[21] Lowe, D.G., Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004. 60(2): p. 91-110.

[22] Dalal, N. and B. Triggs. Histograms of oriented gradients for human detection. in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. 2005. IEEE.

[23] Olshausen, B.A. and D.J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision research, 1997. 37(23): p. 3311-3325.

[24] Hinton, G.E. and R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks. science, 2006. 313(5786): p. 504-507.

[25] Qi, K., et al., Sparse coding-based correlaton model for land-use scene classification in high-resolution remote-sensing images. Journal of Applied Remote Sensing, 2016. 10(4): p. 042005.

[26] Dai, D. and W. Yang, Satellite image classification via two-layer sparse coding with biased image representation. IEEE Geoscience and Remote Sensing Letters, 2011. 8(1): p. 173-176.

[27] Othman, E., et al., Using convolutional features and a sparse autoencoder for land-use scene classification. International Journal of Remote Sensing, 2016. 37(10): p. 2149-2167.

[28] Zhao, W. and S. Du, Scene classification using multi-scale deeply described visual words. International Journal of Remote Sensing, 2016. 37(17): p. 4119-4131.

[29] Längkvist, M., et al., Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing, 2016. 8(4): p. 329.

[30] Maggiori, E., et al., Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 2017. 55(2): p. 645-657.

[31] Y., Y. Bengio, and G. Hinton, Deep learning. nature 521 (7553): 436. Google Scholar, 2015.

[32] Hinton, G.E., S. Osindero, and Y.-W. Teh, A fast learning algorithm for deep belief nets. Neural computation, 2006. 18(7): p. 1527-1554.

[33] Hinton, G.E. and R.R. Salakhutdinov. A better way to pretrain deep boltzmann machines. in Advances in Neural Information Processing Systems. 2012.

[34] Vincent, P., et al., Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research, 2010. 11(Dec): p. 3371-3408.

[35] Szegedy, C., et al. Going deeper with convolutions. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[36] Maggiori, E., et al., High-resolution semantic labeling with convolutional neural networks. arXiv, 2016. arXiv preprint arXiv:1611.01962.

[37] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. nature, 2015. 521(7553): p. 436.

[38] C., et al., A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. ISPRS Journal of Photogrammetry and Remote Sensing, 2018. 140: p. 133-144.

[39] C.M., Neural networks for pattern recognition. 1995: Oxford university press.

[40] Boureau, Y.-L., J. Ponce, and Y. LeCun. A theoretical analysis of feature pooling in visual recognition. in Proceedings of the 27th international conference on machine learning (ICML-10). 2010.

[41] Nogueira, K., O.A. Penatti, and J.A. dos Santos, Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 2017. 61: p. 539-556.

[42] Richards, J.A. and J. Richards, Remote sensing digital image analysis. Vol. 3. 1999: Springer.

[43] Tang, T., et al., Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining. Sensors, 2017. 17(2): p. 336.

[44] https://geodata.iowa.gov/dataset/2016-aerial-imagery-iowa/resource/32b0610f-2608-4fb5-99ed-6195dce1425a.

[45] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in Advances in neural information processing systems. 2012.

[46] Ruder, S., An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016.

[47] Mushore, T.D., et al., Assessing the potential of integrated Landsat 8 thermal bands, with the traditional reflective bands and derived vegetation indices in classifying urban landscapes. Geocarto international, 2017. 32(8): p. 886-899.