سنجش از دور
و
GIS ایران

سنجش از دور و GIS ایران سال سیزدهم، شماره دوم، تابستان ۱۴۰۰
Iranian Remote Sensing & GIS Vol.13, No. 2, Summer 2021

۳۹-۶۰

مقاله پژوهشی

# راهکاری مبتنی‌بر شبکه‌های عصبی کاملاً کانوولوشنی برای تشخیص هم‌زمان جاده‌ها و ساختمان‌ها در تصاویر هوایی

ناصر فرج‌زاده[1*] و هیوا ابراهیم‌زاده[2]

۱. دانشیار دانشکدهٔ فنّاوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز

۲. دانشجوی کارشناسی ارشد دانشکدهٔ فنّاوری اطلاعات و مهندسی کامپیوتر، دانشگاه شهید مدنی آذربایجان، تبریز

تاریخ دریافت مقاله: ۱۳۹۸/۱۰/۱۸ تاریخ پذیرش مقاله: ۱۳۹۹/۰۸/۰۵

## چکیده

توسعهٔ سیستم‌های خودکار تشخیص جاده و ساختمان در تصاویر هوایی همواره با چالش‌های مهمی مانند متفاوت‌بودن ظاهر ساختمان‌ها، تغییرات روشنایی، زاویهٔ تصویربرداری و فشرده و چگال‌بودن جاده‌ها و ساختمان‌ها در نواحی شهری روبه‌روست. در چند سال اخیر، استفاده از شبکه‌های عصبی مصنوعی چندلایه (شبکه‌های عصبی عمیق) مورد توجه بسیاری از پژوهشگران این حوزه (و حوزه‌های مشابه) قرار گرفته و نتایج خیره‌کننده‌ای با به‌کارگیری آنها حاصل شده است. باوجوداین، به‌دلیل استفاده از لایه‌های کاملاً متصل در راهکارهای داده‌شده، میانگین مدت زمان پردازش هنوز بسیار زیاد است و مدل ساخته‌شده نیز به‌سرعت دچار پدیدهٔ بیش‌برازش می‌شود. علاوه‌براین، در بیشتر روش‌های پیشنهادی، برای تفسیر تصاویر هوایی براساس چنین راهکاری از رویکرد تک‌کلاس استفاده شده است. به‌عبارتی، تشخیص جاده‌ها و ساختمان‌ها از عوارض طبیعی به‌طور هم‌زمان امکان‌پذیر نیست و لازم است مدل‌های جداگانه‌ای برای تشخیص هریک از آنها ایجاد شود. هدف اصلی، در این پژوهش، طراحی معماری جدیدی است که مدل ساخته‌شده با استفاده از آن بتواند، هم‌زمان، جاده‌ها و ساختمان‌ها را از عوارض طبیعی تشخیص دهد و به‌این‌ترتیب، پیچیدگی عمل طبقه‌بندی را به حداقل برساند. همچنین، در طراحی معماری پیشنهادی، حذف لایه‌های کاملاً متصل از معماری چندلایه‌ای مرسوم و در نتیجه، کاهش میانگین مدت زمان پردازش مورد توجه قرار گرفته است. نتایج آزمایش‌های انجام‌گرفته روی بانک تصاویر هوایی ماساچوست نشان می‌دهد عملکرد معماری پیشنهادی ۳۸٪ سریع‌تر از دیگر روش‌های مبتنی‌بر شبکه‌های عصبی چندلایه بوده است و دقت تشخیص را به‌طور میانگین، ۲٪ افزایش می‌دهد.

**کلیدواژه‌ها:** یادگیری عمیق، شبکه‌های عصبی مصنوعی، شبکه‌های عصبی کانوولوشنی، تصاویر هوایی، شناسایی جاده، شناسایی ساختمان، شناسایی عوارض طبیعی، هوش مصنوعی.

* نویسندهٔ مکاتبه‌کننده: تهران، تبریز، ۳۵کیلومتری جادهٔ تبریز- مراغه، دانشگاه شهید مدنی آذربایجان، دانشکدهٔ فنّاوری اطلاعات و مهندسی کامپیوتر، کد پستی: ۵۳۷۵۱۷۱۳۷۹.
تلفن: ۰۹۳۰۸۱۴۹۶۰۰

Email: n.farajzadeh@azaruniv.ac.ir

## ۱- مقدمه

«تفسیر تصاویر هوایی» به فرایند بررسی این تصاویر به‌منظور شناسایی اشیا و تعیین ویژگی‌های متفاوت اشیای شناسایی‌شده گفته می‌شود. شروع این فرایند به جنگ جهانی اول و زمانی بازمی‌گردد که عکس‌های گرفته‌شده با استفاده از هواپیماها به‌منظور شناسایی اهداف بررسی می‌شدند (Mayer, 1999). در ادبیات بینایی ماشین، تفسیر خودکار تصاویر هوایی معمولاً با عنوان برچسب‌گذاری پیکسل‌ها (با کلاس‌های مورد نظر) مطرح می‌شود. ازاین‌رو، هدف از تفسیر تصاویر هوایی تقطیعِ معنایی کامل آنها به قطعاتی مانند ساختمان، جاده، درختان، فضای سبز و آب (Kluckner & Bischof, 2009, Kluckner et al., 2009) و یا طبقه‌بندی دودویی تصاویر به دو کلاس خاص، مثل نواحی ساخت بشر و منابع طبیعی، است (فرج‌زاده و هاشم‌زاده، ۱۳۹۸).

تقطیع معنایی تصاویر هوایی اغلب با چندین چالش همچون تغییر در ظاهر شیء ناشی از تغییر در زاویهٔ دید، انسداد، شلوغی پس‌زمینه، تغییرات روشنایی، سایه و چگالی بالای جاده‌ها و ساختمان‌ها در نواحی شهری روبه‌روست. برای بررسی این چالش‌ها، مطالعهٔ موضوع شناسایی اشیای جغرافیایی، به‌صورت گسترده، از سال ۱۹۸۰ آغاز شده است. وضوح و کیفیت پایین تصاویر ماهواره‌ای اولیه، مانند لندست[1]، امکان شناسایی و تشخیص اشیای ساخت بشر از عوارض طبیعی را به پژوهشگران نمی‌داد؛ بنابراین، پژوهشگران اغلب روی استخراج ویژگی‌هایی از نواحی این تصاویر متمرکز شده بودند. با پیشرفت تکنولوژی سنجش از راه دور و پدیدآمدن ماهواره‌هایی با قابلیت ثبت تصاویر با وضوح بالا، مانند کوئیک‌برد[2]، اسپات-۵[3] و ایکونوس[4]، تصاویر ماهواره‌ای با اطلاعات بافتی و مکانی بیشتری برای پژوهشگران فراهم شد (Cheng & Han, 2016).

به‌دلیل اینکه استخراج اشیای مورد نظر در تصاویر هوایی به‌دست انسان کاری بسیار پرهزینه است، دادن راهکاری که بتواند حجم انبوهی از تصاویر هوایی را در مدت زمان اندک و به‌صورت خودکار برچسب‌گذاری کند همواره مورد توجه پژوهشگران بوده است.

در اغلب روش‌های مطرح‌شده برای خودکارسازی تفسیر تصاویر هوایی، استفاده از راهکارهای مبتنی‌بر ویژگی‌های محلی مورد توجه قرار گرفته است (فرج‌زاده و هاشم‌زاده ۱۳۹۸). در این راهکارها، معمولاً با استفاده از یک روش استخراج ویژگی مهندسی‌شده و یک الگوریتم یادگیری، برای ایجاد مدل طبقه‌بندی‌کننده- تشخیص‌دهنده اقدام می‌شود. به‌رغم آنکه این راهکارها نتایج درخور توجهی دارند، متأسفانه نیاز به مهیابودن دانش پیشین دربارهٔ ظاهر شیء یا نحوهٔ توزیع مقادیر پیکسل‌ها در هر کلاس یکی از نقاط ضعف آنها به‌شمار می‌رود.

در سال‌های اخیر، به‌لطف پیشرفت‌های چشمگیر در عرصهٔ تولید GPUهایی با توان پردازشی بسیار بالا، استفاده از معماری چندلایه در شبکه‌های عصبی مصنوعی طرفداران بسیاری پیدا کرده و به‌کارگیری آنها در حوزه‌های گوناگون هوش مصنوعی باعث شده است نتایج خیره‌کننده‌ای حاصل شود (Liu et al., 2017). دلیل اصلی موفقیت چنین شبکه‌هایی توانایی استخراج خودکار ویژگی‌ها با استفاده از ساختار چندلایه‌ای آنهاست. بااین‌حال، چنین شبکه‌هایی با دو چالش اصلی مواجه‌اند (Cheng et al., 2018): ۱. مدت زمان ساخت مدل با استفاده از این شبکه‌ها (حتی با بهره‌گیری از GPU) بسیار طولانی است؛ ۲. مدل ساخته‌شده همواره در معرض دچارشدن به پدیدهٔ بیش‌برازش[5] است و ازهمین‌رو، باید داده‌های بسیار زیادی برای آموزش آنها فراهم شود. شایان ذکر است که خاستگاه هر دو چالش مطرح‌شده وجود تعداد بسیار زیاد نورون‌ها[6] و ارتباط بین آنها در این معماری است که به معماری کاملاً متصل[7] نیز معروف است.

1. Landsat
2. Quickbird
3. Spot-5
4. Ikonos
5. Overfitting
6. Neurons
7. Fully Connected

برای غلبه بر چالش‌های یادشده در معماری کاملاً متصل، معماری جدیدی با نام معماری کانوولوشنی[1] معرفی شده که در آن تغییراتی در ارتباطات مابین نورون‌ها و ماهیت لایه‌ها داده شده است. این معماری در کاربردهای گوناگونی، به‌ویژه بینایی ماشین، کارآیی بسیار مطلوبی از خود نشان داده و تبدیل به یکی از رایج‌ترین معماری‌ها شده است (Aggarwal, 2018).

در زمینهٔ تشخیص سازه‌های ساخت بشر مانند جاده‌ها و ساختمان‌ها، در تصاویر هوایی با استفاده از شبکه‌های عصبی کانوولوشنی، پژوهش‌های متعددی انجام شده است (Mnih, 2013; Saito et al., 2016; Alshehhi et al., 2017; Panboonyuen et al., 2017; Hui et al., 2018). هریک از این روش‌ها سعی در عرضهٔ معماری جدیدی، با رویکرد کاهش میانگین زمان پردازش و تا حدی افزایش دقت تشخیص، داشته است. اما اغلب آنها، حتی با توسل به معماری کانوولوشنی هم، هزینهٔ بسیار زیاد محاسباتی دارند و دلیل آن وجود دست‌کم یک لایهٔ کاملاً متصل است که معمولاً در انتهای معماری گنجانده می‌شود. همچنین، در بیشتر روش‌ها تشخیص هم‌زمان جاده‌ها و ساختمان‌ها از عوارض طبیعی، با استفاده از یک مدل واحد، مقدور نیست و برای تشخیص جاده‌ها و یا ساختمان‌ها، باید مدل‌های جداگانه‌ای ساخته شوند. این موضوع نیز باعث افزایش پیچیدگی مسئلهٔ تشخیص خودکار سازه‌های ساخت بشر در تصاویر هوایی شده است.

در پژوهش پیش رو، با هدف تشخیص هم‌زمان جاده‌ها و ساختمان‌ها از عوارض طبیعی و در نتیجه، به‌حداقل‌رساندن میزان پیچیدگی مسئلهٔ تشخیص خودکار سازه‌های ساخت بشر در تصاویر هوایی، معماری جدیدی پیشنهاد می‌شود. چنین رویکردی با توجه به پرهزینه‌بودن سخت‌افزار مورد نیاز و زمان‌بربودن ساخت مدل‌های جداگانه برای هر کلاس، همواره مورد توجه پژوهشگران حوزهٔ بینایی ماشین بوده است. همچنین با هدف کاهش مدت زمان ساخت مدل و افزایش میانگین سرعت تشخیص و جلوگیری از بروز پدیدهٔ بیش‌برازش، در طراحی معماری پیشنهادی از هیچ لایهٔ کاملاً متصلی استفاده نشده است.

در ادامه، ادبیات و پیشینهٔ پژوهش مرور می‌شود. در بخش سوم، جزئیات معماری پیشنهادی توضیح داده می‌شود. در بخش چهارم، دربارهٔ نتایج آزمایش‌های انجام‌گرفته برای ارزیابی معماری پیشنهادی بحث می‌شود و در بخش پنجم هم، یافته‌های پژوهش و مسیرهای ممکن برای توسعه و مطالعهٔ بیشتر مطرح می‌شود.

## ۲- پژوهش‌های پیشین

تجزیه‌وتحلیل تصاویر هوایی برای تفسیر آنها، به‌دلیل گستردگی تفاوت در شکل ظاهری سازه‌های دست بشر و مشکلات تصویربرداری، به‌صورت ذاتی چالش‌برانگیز است. علاوه‌براین، حجم انبوهی از تصاویر گردآوری‌شده ازطریق ابزارهایی مانند پهپادها، که با هزینهٔ اندک در اختیار بیشتر سازمان‌های ذی‌نفع قرار می‌گیرد، سختی‌های این مسئله را دوچندان کرده است (فرجزاده و هاشم‌زاده، ۱۳۹۸). ازآن‌جاکه تجزیه‌وتحلیل این تصاویر نقش بسیار مهمی در طیف گسترده‌ای از برنامه‌های کاربردی ایفا می‌کند، مطرح‌کردن راهکاری کارآمد برای تشخیص خودکار جاده‌ها و ساختمان‌ها از عوارض طبیعی توجه بسیاری از پژوهشگران را، در سال‌های اخیر، به خود جلب کرده است.

بر اساس دسته‌بندی بیان‌شده در پژوهش چنگ و هان[2] (۲۰۱۶)، روش‌های مرسوم توسعه‌یافته برای تشخیص خودکار اشیا در تصاویر هوایی در چهار گروه دسته‌بندی می‌شوند: ۱. روش‌های تشخیص شیء براساس مطابقت الگو (Lefèvre & Weber, 2007; Leninisha & Vani, 2015; Lin et al., 2015)؛ ۲. روش‌های تشخیص شیء براساس دانش (Akçay & Aksoy, 2010; Clinton et al., 2010; Liu et al., 2013; Ok et al., 2013)؛

1. Convolutional Neural Networks
2. Cheng & Han

۳. روش‌های تشخیص شیء طبق آنالیز تصویر مبتنی‌بر شیء (Hay et al., 2003; Walker & Blaschke, 2008; Feizizadeh et al., 2014; Goodin et al., 2015; Contreras et al., 2016)؛ ۴. روش‌های تشخیص شیء براساس یادگیری ماشین (Song et al., 2004; Das et al., 2011; Arı et al., 2014; Li et al., 2015; Wang et al., 2015). شایان ذکر است که این چهار دسته ضرورتاً از هم جدا نیستند و در برخی پژوهش‌ها، ترکیبی از روش‌های گوناگون استفاده شده است (Wang et al., 2013; Yokoya & Iwasaki, 2015; Zhao et al., 2015; Chen et al., 2018).

با پیشرفت الگوریتم‌های یادگیری ماشین و نمایش‌دهنده‌های قدرتمند ویژگی‌ها، بسیاری از پژوهش‌های اخیر تشخیص شیء در تصاویر هوایی را نکته‌ای در طبقه‌بندی شمرده و پیشرفت چشمگیری در آن به‌دست آورده‌اند (فرج‌زاده و هاشم‌زاده، ۱۳۹۸).

تشخیص شیء ممکن است با استفاده از یک طبقه‌بند[1] و با یادگیری تغییرات در ظاهر و دید شیء از مجموعه داده‌ای آموزشی در چارچوبی با نظارت ضعیف، نیمه‌نظارتی یا کاملاً نظارتی صورت گیرد. ورودی طبقه‌بند (مدل) مجموعه‌ای از نواحی با نمایش ویژگی‌های مورد نظر و خروجی آن شامل برچسب‌های (شیء یا غیرشیء) پیش‌بینی‌شده ازطریق مدل است.

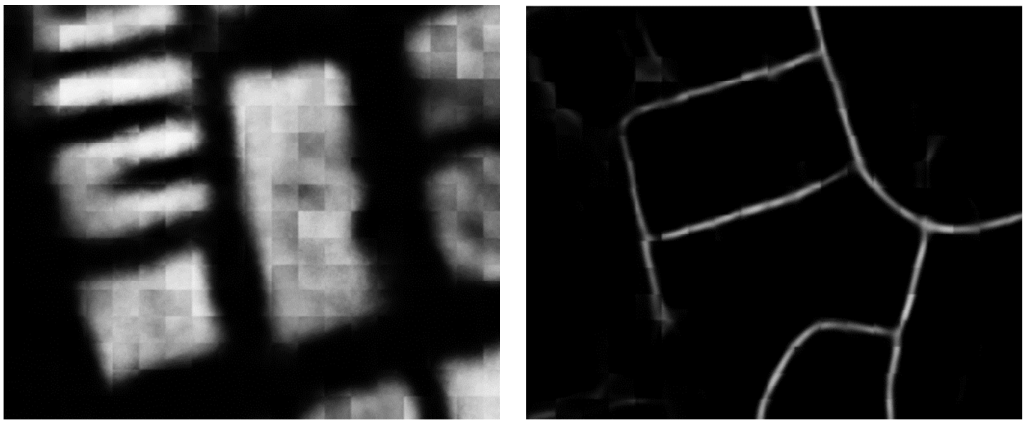
با توجه به پیشرفت چشمگیری که در حوزهٔ شبکه‌های عصبی چندلایه ازسوی هینتون و سالاخودینوف[2] (۲۰۰۶) صورت گرفت، مسئلهٔ بازنمایی ویژگی‌ها در تصاویر، به‌صورت خودکار، وارد عصر تازه‌ای شد؛ به‌طوری‌که در این روش، با بهره‌گیری از ساختار چندلایه‌ای از نورون‌ها، استخراج خودکار ویژگی‌ها امکان‌پذیر شد. به‌عبارتی، مسئولیت استخراج ویژگی‌ها، بدون توجه به ماهیت مسئلهٔ مفروض، به ساختار شبکهٔ عصبی چندلایه انتقال یافت. این درحالی بود که در روش‌های مرسوم پیشین، با اتکا به دانش قبلی دربارهٔ ماهیت مسئلهٔ مفروض، یکی از روش‌های استخراج ویژگی مانند HOG[3] (Tuermer et al., 2013)، BOW[4] (Sun et al., 2012; Bai et al., 2014)، Haar-like (Grabner et al., 2008; Leitloff et al., 2010) یا دیگر موارد انتخاب[5] (مهندسی)[6] و استفاده می‌شد.

یکی از پژوهش‌های پیشگام در زمینهٔ شناسایی سازه‌های ساخت بشر در تصاویر هوایی، به‌ویژه جاده‌ها، با استفاده از شبکه‌های عصبی چندلایه را مینهٔ[7] (۲۰۱۳) مطرح کرد. در این پژوهش، نخست تصاویر هوایی به قطعات کوچک (۶۴ × ۶۴) تقسیم و سپس از روش PCA[8] برای کاهش ابعادشان استفاده شد. در ادامه، بردارهای حاصل از مرحلهٔ قبل برای تنظیم دقیق وزن‌های شبکهٔ عصبی کانوولوشنی، که از پیش آموزش داده شده بود، به‌کار رفت.

این معماری، در مقایسه با روش‌های مرسوم مبتنی‌بر استخراج مهندسی‌شدهٔ ویژگی‌ها، به دقت بسیار خوب دست یافته است اما، به‌دلیل استفاده از لایه‌ای کاملاً متصل در معماری پیشنهادی، مدت زمان آموزش شبکه بسیار طولانی است و تشخیص با سرعت پایینی انجام می‌شود که مناسب کاربردهای بلادرنگ[9] و یا پردازش حجم انبوهی از تصاویر هوایی نیست.

اندازهٔ تصویر تولیدشده در این روش نیز، که حاوی برچسب جاده‌ها و یا عوارض طبیعی است، به‌دلیل استفاده از لایه‌هایی مانند لایهٔ ادغام[10] در معماری شبکه، کوچک‌تر از تصویر ورودی می‌شود. علاوه‌براین، حاشیه‌های جاده‌ها و ساختمان‌ها، در تصویر تولیدشده، صاف و متراکم نیستند و کیفیت مطلوبی ندارند (شکل ۱). برای جبران این مشکل، مینهٔ پیشنهاد داد از نورون‌های بیشتری در لایهٔ کاملاً متصل استفاده شود. بااین‌حال، نتایج چندان رضایت‌بخش نبود و افزایش تعداد نورون‌ها سبب افزایش مدت زمان پردازش شد.

1. Classifier
2. Hinton & Salakhutdinov
3. Histogram of Oriented Gradients
4. Bag-of-Words
5. Hand Crafted
6. Engineered
7. Mnih
8. Principal Component Analysis
9. Real-time
10. Pooling



(ب) (الف)

**شکل ۱.** نمونه‌ای از تولید قطعات برچسب‌دارشده به‌روش مینهٔ (۲۰۱۳). ملاحظه می‌شود که چگونه مرز نواحی برچسب‌دارشده متراکم نیست و پیوستگی مطلوبی ندارند: برچسب محور جاده‌ها (الف)؛ برچسب ساختمان‌ها (ب)

نقطه‌ضعف دیگر راهکار پیشنهادی مینهٔ ناتوانی مدل ایجادشده در تشخیص هم‌زمان جاده‌ها و ساختمان‌ها از عوارض طبیعی است. به‌عبارتی، مدل ایجاده‌شونده فقط می‌تواند جاده را از عوارض طبیعی تشخیص دهد و برای تشخیص ساختمان از عوارض طبیعی، باید مدل دیگری ساخته شود. سایتو[1] و همکارانش (۲۰۱۶) در پژوهشی به این موضوع توجه کردند. آنها با پیشنهاد معماری جدیدی، سعی کردند مدلی بسازند که بتواند هم‌زمان جاده‌ها و ساختمان‌ها را از عوارض طبیعی تشخیص دهد و در خروجی شبکه، برچسب‌های متناظر را تولید کند. متأسفانه راهکار پیشنهادی این پژوهشگران نیز، به‌دلیل استفاده از لایه‌های کاملاً متصل در معماری، تمامی مشکلات پژوهش مینهٔ (۲۰۱۳) را داراست.

الشِحی[2] و همکاران (۲۰۱۷) در پژوهش خود از رویکردی مشابه با مینهٔ (۲۰۱۳) بهره برده‌اند. روش پیشنهادی یک معماری کانوولوشنی مبتنی‌بر قطعه برای استخراج ساختمان‌ها از تصاویر هوایی با رزولوشن بالاست. در شبکهٔ کانوولوشنی پیشنهادی، لایهٔ کاملاً متصل با لایهٔ «انتخاب میانگین سراسری» (GAP)[3] جایگزین شده است. با بهره‌گیری از روش GAP، به‌جای استفاده از تمامی نقشه‌های لایهٔ قبلی در محاسبات، فقط میانگین نقشه‌هایی در لایهٔ قبلی محاسبه و استفاده می‌شود که انتظار می‌رود نتایج مشابهی برای قطعهٔ مفروض تولید کنند. با این رویکرد، حجم محاسبات لازم کاهش می‌یابد. این معماری، با اینکه نواحی ساختمان‌ها را به‌خوبی شناسایی می‌کند، پیوستگی و فشردگی اشکال ساختمان‌ها را که ممکن است باعث ایجاد خطوط نامنظم شوند، تضمین نمی‌کند. ازاین‌رو، در پژوهش یادشده، از یک گام پس‌پردازش برای کاهش نواحی نامساعد استفاده شده است (Achanta et al., 2012).

برای غلبه بر مشکل تعداد نورون‌ها و اتصالات مورد نظر در معماری شبکه، ماجیوری[4] و همکارانش (۲۰۱۷) استفاده از معماری کاملاً کانوولوشنی را برای تشخیص سازه‌های ساخت بشر در تصاویر هوایی پیشنهاد دادند. به‌دلیل حذف لایهٔ کاملاً متصل در این معماری، انتظار می‌رود که قدرت تعمیم این شبکه کاهش پیدا کند. به همین منظور، آن‌ها نخست مدل کاملاً کانوولوشنی خود را با استفاده از داده‌های گردآمده طی پروژهٔ اوپن‌استریت‌مپ[5] آموزش دادند و سپس وزن‌های آن را براساس داده‌های آموزشی مورد نظر، تنظیم کردند.

1. Saito
2. Alshehhi
3. Global Average Pooling
4. Maggiori
5. OpenStreetMap (www.openstreetmap.org)

یکی از نقاط ضعف این معماری ناتوانی آن در تشخیص هم‌زمان جاده و ساختمان در تصاویر است و بیشتر تلاش نویسندگان این مقاله معطوف به تشخیص دقیق ساختمان‌ها در تصاویر هوایی شده است. همچنین استفاده از راهکار دومرحله‌ای پیش‌آموزش و سپس، ساخت مدل نهایی باعث افزایش پیچیدگی آن شده است.

پنبونیون[1] و همکارانش (۲۰۱۷)، با تمرکز بر استخراج جاده، معماری کاملاً کانوولوشنی را پیشنهاد کردند که در آن، به‌جای استفاده از تابع نگاشت مرسوم ReLU[2]، از تابع نگاشت دیگری به نام ELU[3] استفاده شده است. به‌تازگی بسیاری از پژوهشگران به این تابع توجه داشته‌اند. این تابع نگاشت، برخلاف ReLU، دچار مشکل پاسخ‌ندادن نورون‌های دارای مقادیر منفی نمی‌شود. این محققان، برای افزایش دقت خروجی، از ویژگی ظاهری جاده‌ها و روش CRF[4] (Chen et al., 2014) استفاده کرده‌اند. در پژوهشی دیگر هیوئی[5] و همکارانش (۲۰۱۸)، با استفاده از معماری یو ـ نت[6] (Ronneberger et al., 2015) و ترکیب آن با معماری اِکسپشن[7] (Chollet, 2017) و تنظیم وزن‌های شبکه با داده‌های تصاویر هوایی، اقدام به تشخیص ساختمان‌ها کردند. نوآوری اصلی این شیوه ترکیب ماژول‌های از پیش آموزش‌دیده با یکدیگر است. اما به‌دلیل استفاده از معماری پیچیدهٔ یو ـ نت و ترکیب آن با اِکسپشن، مدت زمان پردازشی و نیز حافظهٔ مورد نیاز این روش بسیار زیاد است.

با مرور پژوهش‌های انجام‌گرفته در تشخیص جاده‌ها و ساختمان‌ها در تصاویر هوایی، می‌توان به این نتیجه رسید که روش‌های مبتنی‌بر استخراج ویژگی‌های مهندسی‌شده به‌همراه یک الگوریتم یادگیری، به‌دلیل وابسته‌بودن به دانش پیشین در زمینهٔ عوارض مورد مطالعه، کارآیی کمتری در مقایسه با روش‌های مبتنی‌بر شبکه‌های چندلایه دارند. در مقابل، روش‌های مبتنی‌بر شبکه‌های چندلایه سه ضعف مهم دارند: ۱. وجود لایه‌ای کاملاً متصل در معماری باعث افزایش بسیار مدت زمان برای ساخت مدل و میانگین زمان تشخیص می‌شود؛ ۲. در مرز نواحی برچسب‌گذاری‌شده، نواحی غیرواقعی تولید می‌شوند؛ و ۳. مطرح‌کردن معماری‌ای که بتواند هم‌زمان جاده، ساختمان و عوارض طبیعی را از هم تشخیص دهد چالش‌برانگیز است و ساخت چنین مدل چندکلاسی باعث کاهش کارآیی تشخیص شبکه می‌شود.

در این پژوهش، با انگیزهٔ کاهش زمان پردازش، افزایش دقت تشخیص و تولید نقشهٔ خروجی دقیق از جاده‌ها و ساختمان‌های تشخیص‌داده‌شده، قصد داریم شبکهٔ عصبی کاملاً کانوولوشنی را مطرح کنیم که به‌صورت خودکار، تصاویر ورودی را هم‌زمان در سه برچسب جاده، ساختمان و پس‌زمینه طبقه‌بندی کند. علاوه‌براین، فاز آموزش در معماری پیشنهادی، برخلاف پژوهش‌های مشابه مانند هیوئی و همکارانش (۲۰۱۸) که از شبکه‌های پیش‌آموزش‌دیده همچون یو ـ نت و اِکسپشن برای غلبه بر کاهش کارآیی شبکه استفاده کرده‌اند، تک‌مرحله‌ای است و صرفاً با استفاده از داده‌های مربوط به عوارض زمین ساخته می‌شود.

## ۳- روش پیشنهادی

در این بخش، روش پیشنهادی برای تشخیص جاده‌ها و ساختمان‌ها از عوارض طبیعی بیان می‌شود. این روش از شبکهٔ عصبی چندلایه‌ای با معماری کاملاً کانوولوشنی بهره می‌برد. نخست، چارچوب کلی شبکه‌های عصبی کانوولوشنی بررسی و سپس معماری روش پیشنهادی، به‌همراه جزئیات لازم، مطرح می‌شود.

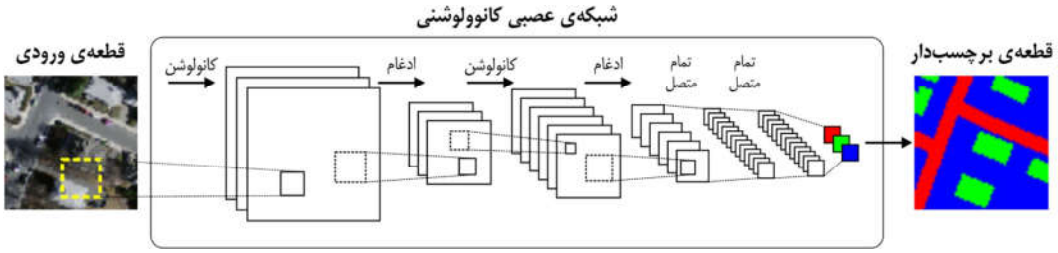
1. Panboonyuen
2. Rectified Linear Unit
3. Exponential linear unit
4. Conditional Random Field
5. Hui
6. U-Net
7. Xception

### ۳-۱- معماری عمومی شبکه‌های عصبی کانوولوشنی

نمای معماری عمومی شبکه‌های عصبی کانوولوشنی، با تأکید بر کاربرد آنها در تشخیص جاده، ساختمان و عوارض طبیعی از همدیگر، در شکل ۲ نمایش داده شده است. به‌طورکلی، هر شبکهٔ عصبی کانوولوشنی مجموعه‌ای از سه لایهٔ کانوولوشن، ادغام و به‌تمامی متصل است که هریک از این لایه‌ها وظایف متفاوتی انجام می‌دهد. معمولاً چیدمان لایه‌ها به‌صورتی است که پس از هر لایهٔ کانوولوشنی، یک لایهٔ ادغام قرار می‌گیرد و در انتها، تعدادی لایهٔ کاملاً متصل وجود دارند که وظیفهٔ آنها در معماری عمل طبقه‌بندی است.

**لایهٔ ادغام:** در این لایه، عمل نمونه‌برداری و کاهش ابعاد ماتریس‌های نگاشت ویژگی انجام می‌شود. این کار معمولاً با استفاده از توابع از پیش تعریف‌شده‌ای، همچون تابع میانگین[5] و تابع بیشینه[6]، انجام می‌شود. دلیل استفاده از چنین لایه‌ای افزایش مقاومت شبکه در مقابل تغییرات مکانی[7] ورودی‌هاست (Nogueira et al., 2017).

**لایهٔ کاملاً متصل:** در این لایه، تمامی نورون‌های لایهٔ پیشین به کل نورون‌های این لایه وصل می‌شوند. وظیفهٔ این لایه تولید بردار احتمالات رخداد هریک از کلاس‌ها (دسته‌ها) در خروجی است. برای انتخاب یکی




**شکل ۲.** معماری عمومی یک شبکهٔ عصبی کانوولوشنی برای تشخیص جاده و ساختمان از عوارض طبیعی

**لایهٔ کانوولوشن:** این لایه معمولاً از دو عملگر کانوولوشن و نگاشت غیرخطی تشکیل شده است. با اعمال عملگر کانوولوشن روی تصویر ورودی، یک (یا چندین) تصویر (ماتریس) کوچک‌تر به نام «نقشهٔ ویژگی»[1] ایجاد می‌شود. هر درایهٔ ماتریس نقشهٔ ویژگی از ضرب نقطه‌ای عناصر محلی (مشخص‌شده با نقطه‌چین زرد در قطعهٔ ورودی، شکل ۲) با مجموعه‌ای از وزن‌ها که با نام فیلتر یا هسته[2] شناخته می‌شوند، حاصل می‌شود. سپس با استفاده از تابعی[3] غیرخطی مانند ReLU یا tanh[4]، مقادیر هریک از درایه‌ها به مقدار جدیدی نگاشت می‌شود. دلیل استفاده از چنین نگاشت غیرخطی‌ای افزایش قابلیت یادگیری و قدرت تعمیم شبکه برای ورودی‌های پیچیده است (Aggarwal, 2018).

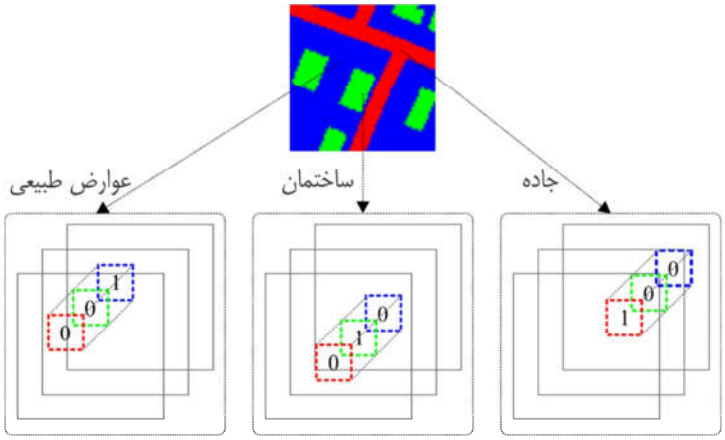
از چند کلاس براساس احتمالات رخداد، از یک تابع نگاشت استفاده می‌شود. معمول‌ترین تابع نگاشت مورد استفاده در این لایه، سافت‌مکس[8] نام دارد (Ibid.). معمولاً در این لایه از راهکار حذف تصادفی[9] برای کاهش تعداد اتصالات و نیز جلوگیری از بروز پدیدهٔ بیش‌برازش در زمان آموزش استفاده می‌شود.

1. Feature Map
2. Kernel
3. Activation function
4. Hyperbolic tangent
5. Average pooling
6. Max pooling
7. Spatial invariance
8. Softmax
9. Dropout

### ۳-۲- معماری پیشنهادی

**برچسب‌گذاری:** فرض می‌شود که مجموعه‌ای از $N$ تصویر هوایی $A = (A^{(1)}, \dots, A^{(N)})$ و تصاویر برچسب‌دار متناظرشان $L = (\tilde{L}^{(1)}, \dots, \tilde{L}^{(N)})$ در دسترس است. همچنین، فرض می‌شود که تصاویر هوایی $m \times m$ و دارای C کانال[1] هستند و $\tilde{L}^{(n)}$ تصویر برچسب متناظر تصویر $n$اُم است که ازلحاظ اندازه، با تصویر $A^{(n)}$ برابر است. برچسب‌گذاری دودویی تصویر $\tilde{L}^{(n)}$ با مقدار ۰ و یا ۱ خواهد بود که عدد ۱ نشان‌دهندهٔ وجود و ۰ نشان‌دهندهٔ فقدان شیء کلاس مد نظر است. با توجه به اینکه هدف از این پژوهش تشخیص هم‌زمان جاده و ساختمان از عوارض طبیعی است، تصویر برچسب‌دار شامل سه نقشهٔ جاده، ساختمان و پس‌زمینه (عوارض طبیعی) خواهد بود. بنابراین، در تصویر برچسب‌دارشده، هر پیکسل یک بردار سه‌مؤلفه‌ای است. شکل ۳ نمونه‌ای از چنین فرموله‌سازی را نشان می‌دهد.

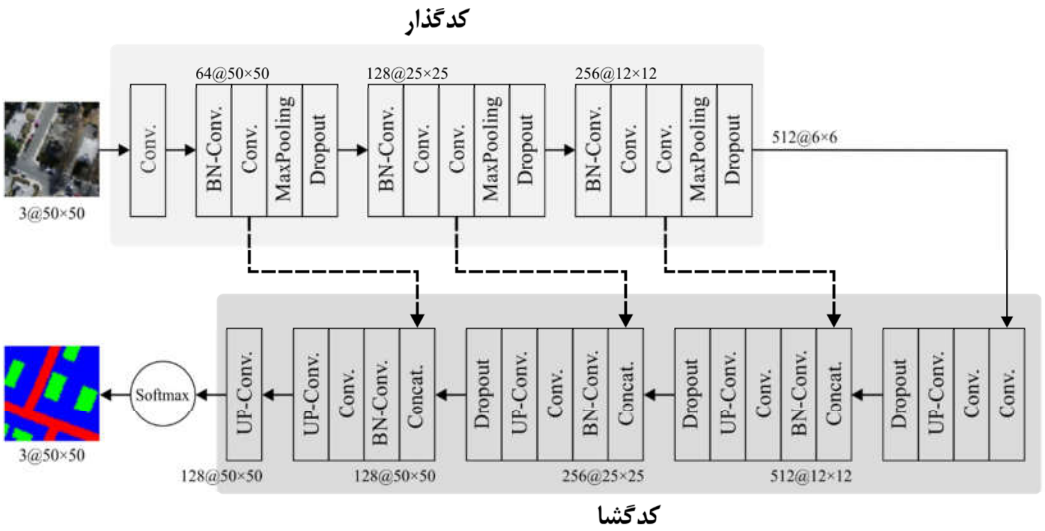
**معماری:** در معماری پیشنهادی به‌منظور حذف لایه‌های کاملاً متصل، با الهام از پژوهش لانگ[2] و همکارانش (۲۰۱۵)، شبکه‌ای کاملاً کانوولوشنی طراحی می‌شود. در حقیقت، شبکهٔ عصبی کاملاً کانوولوشنی یک شبکهٔ عصبی کانوولوشنی معمولی است که آخرین لایهٔ (کاملاً متصل) آن با یک لایهٔ کانوولوشنی ۱×۱ جایگزین شده است. این لایه نقش دریافت‌کنندهٔ بزرگ سیگنال (تصویر) را ایفا می‌کند و ایدهٔ اصلی استفاده از آن، در معماری، رسیدن به چارچوبی سراسری از سیگنال (تصویر) ورودی است که در آن بتوان مکان تقریبی اشیا در تصویر را شناسایی کرد. برخلاف شبکهٔ عصبی کانوولوشنی معمولی، با استفاده از این رویکرد، علاوه‌بر طبقه‌بندی اشیا می‌توان عمل تقطیع تصویر را نیز انجام داد.




**شکل ۳.** فرموله‌سازی چندکلاسی برای تشخیص هم‌زمان جاده و ساختمان از عوارض طبیعی

۱. در بیشتر کاربردهای بینایی ماشین، تصاویر معمولاً با سه کانال رنگی قرمز، سبز و آبی نشان داده می‌شوند. شایان ذکر است که این کانال‌های رنگی نباید با رنگ‌های استفاده‌شده برای برچسب‌گذاری پیکسل‌ها اشتباه گرفته شوند (شکل‌های ۳ و ۴).

2. Long




**شکل ۴.** معماری پیشنهادی کاملاً کانوولوشنی برای تشخیص هم‌زمان جاده و ساختمان از عوارض طبیعی. ورودی شبکه قطعه‌ای تصویر ۵۰×۵۰ با ۳ کانال رنگی (۵۰ × ۵۰@۳) و خروجی آن نیز تصویری هم‌اندازه با تصویر ورودی است. پیکسل با برچسب قرمز معرف نقطه‌ای روی جاده، پیکسل دارای برچسب سبز معرف نقطه‌ای روی ساختمان و پیکسل با برچسب آبی معرف نقطه‌ای روی عوارض طبیعی است

به‌طورکلی، معماری این نوع شبکه‌ها از دو بخش کدگذار[1] و کدگشا[2] تشکیل شده است. هر بخش نیز بلوک‌هایی از لایه‌های کانوولوشنی و ادغام را دربر می‌گیرد. شکل ۴ نمای معماری پیشنهادی را نشان می‌دهد. در این شکل، ورودی هر بلوک (یا خروجی بلوک قبلی) با نماد $c@m \times m$ نشان داده شده که $c$ و $m \times m$ به‌ترتیب نمایانگر تعداد نقشه‌های ویژگی (ماتریس‌ها) و ابعاد آنهاست.

### ۳-۳- کدگذاری و کدگشایی

یکی از مزایای استفاده از لایه‌های ادغام افزایشِ میدان دید[3] و در نتیجه، افزایش میزان تجمیع اطلاعات در شبکه است. در مقابل، استفاده از این لایه باعث ازبین‌رفتن اطلاعات مکانی پیکسل‌ها می‌شود و این درحالی است که در کاربرد حال حاضر، داشتن اطلاعات مکانی دقیق پیکسل‌ها برای ساخت نقشهٔ برچسب ضروری است و اطلاعات مکانی (مختصات) هریک از آنها باید حفظ شود.

برای فائق‌آمدن بر این مشکل، می‌توان از شبکه‌های کاملاً کانوولوشنی با معماری کدگذار- کدگشا استفاده کرد (Seeliger et al., 2018). بخش کدگذار به‌تدریج ابعاد فضایی نقشه‌های ویژگی را با استفاده از لایه‌های ادغام، کاهش می‌دهد و در مقابل، بخش کدگشا جزئیات اشیا و ابعاد فضایی را که در نقشه‌های ویژگی مستترند، به‌مرور، بازیابی می‌کند.

بازیابی با استفاده از راهکار درون‌یابی مقادیر پیکسل‌های همسایه صورت می‌گیرد. این کار «ترانهادهٔ کانوولوشن»[4] یا «نمونه‌سازی»[5] نامیده می‌شود. همچنین، با هدف کمک به بازیابی بهتر جزئیات در بخش کدگشا، معمولاً اتصال‌های میان‌بری از بخش کدگذار به بخش کدگشا وجود دارند (این اتصالات، در شکل ۴، با پیکان‌های نقطه‌چین مشخص شده‌اند). به‌عبارتی، اتصالات میان‌بر نتایج حاصل از بخش کدگذار (به‌دست‌آمده ازطریق لایه‌های ادغام و لایه‌های کانوولوشن) را به‌منزلهٔ پشتیبان

1. Encode
2. Decode
3. Receptive Fields
4. Transpose Convolution
5. Up-sampling

برای افزایش دقت بازسازی[1]، به بلوک‌های بخش کدگشا ارسال می‌کنند (Ronneberger et al., 2015).

**بخش کدگذار:** این بخش شامل چهار بلوک است. در بلوک نخست، فقط یک لایهٔ کانوولوشن[2] وجود دارد که تصویر با سه کانال رنگی و ابعاد $m \times m$ را به‌منزلهٔ ورودی، دریافت می‌کند و عملیات کانوولوشن را با فیلترهای ۳ × ۳، با گام‌هایی[3] به‌طول ۱، روی تصاویر اعمال می‌کند. خروجی این بلوک ۶۴ نقشهٔ ویژگی با ابعاد $m \times m$ خواهد بود (Mnih, 2013).

بلوک دوم شامل چهار لایه است. در لایهٔ نخست[4] این بلوک، ابتدا نرمال‌سازی دسته‌ای[5] انجام می‌شود. در ادامه، عملیات کانوولوشن با فیلتر ۳ × ۳ با گامی[6] برابر با ۱، بدون گسترش مرز[7]، صورت می‌گیرد. توضیح اینکه نرمال‌سازی دسته‌ای با هدف افزایش سرعت آموزش در نتیجهٔ کاهش تغییرات در لایهٔ قبلی (Ioffe & Szegedy, 2015) انجام می‌گیرد. خروجی این لایه ۱۲۸ نقشهٔ ویژگی با ابعاد $m \times m$ است (Mnih, 2013). لایهٔ دوم در این بلوک لایه‌ای کانوولوشنی، با پارامترها و ابعاد خروجی مشابه با لایهٔ کانوولوشنی قبلی است. در لایهٔ سوم[8]، عملیات ادغام براساس بیشترین مقدار برای کاهش ابعاد نقشه‌های ویژگی به $\frac{m}{2} \times \frac{m}{2}$ روی خروجی لایهٔ قبل اعمال می‌شود. در آخرین لایه از این بلوک، از راهکار حذف تصادفی[9] برای جلوگیری از بیش‌برازش استفاده شده است. در این راهکار، تعدادی از خروجی‌های یک لایه (در زمان آموزش شبکه) به‌صورت تصادفی انتخاب می‌شوند و از آنها به‌منزلهٔ ورودی لایهٔ بعدی استفاده نمی‌شود.

بلوک سوم و چهارم، هر دو، شامل پنج لایه‌اند که شباهت بسیاری به بلوک دوم دارند. تنها تفاوت این بلوک‌ها با بلوک دوم در تعداد و ابعاد لایه‌های کانوولوشنی است. خروجی بلوک سوم و چهارم به‌ترتیب ۲۵۶ و ۵۱۲ نقشهٔ ویژگی، با ابعاد به‌ترتیب $\frac{m}{4} \times \frac{m}{4}$ و $\frac{m}{8} \times \frac{m}{8}$ است.

**بخش کدگشا:** این بخش نیز شامل چهار بلوک می‌شود. نخستین بلوک این بخش از چهار لایه تشکیل شده است. لایهٔ اول یک لایهٔ کانوولوشن است که ورودی آن همان خروجی بخش کدگذار است و عملیات کانوولوشن با فیلتر ۳ × ۳ و گام ۱، بدون گسترش مرز، روی این ورودی‌ها اعمال می‌شود. خروجی این لایه ۵۱۲ نقشهٔ ویژگی با ابعاد $\frac{m}{8} \times \frac{m}{8}$ است. همانند لایهٔ نخست، لایهٔ دوم این بلوک نیز لایه‌ای کانوولوشنی است.

وظیفهٔ لایهٔ سوم[10] بازسازی نقشه‌های ویژگی است که با نمونه‌سازی مبتنی‌بر درون‌یابی پیکسل‌های همسایه انجام می‌شود (Long et al., 2015). این عملیات با فیلترهای ۳ × ۳ و گام ۲ روی داده‌های به‌دست‌آمده از لایهٔ قبل صورت می‌گیرد. خروجی این لایه نقشه‌های ویژگی با ابعاد $\frac{m}{4} \times \frac{m}{4}$ است. در ادامه، عملیات حذف تصادفی روی نتایج لایهٔ سوم اعمال می‌شود.

بلوک دوم از بخش کدگشا شامل پنج لایه است. در لایهٔ اول[11] از این بلوک، نتایج بخش کدگذار با نتایج حاصل از بلوک‌های بخش کدگشا، با هدف افزایش بهبود نتایج، ادغام می‌شوند. سپس در لایهٔ دوم، نرمال‌سازی دسته‌ای و بعد، عملیات کانوولوشن روی ورودی‌ها اعمال می‌شود. در لایهٔ سوم نیز، عملیات کانوولوشن روی نتایج به‌دست‌آمده از لایهٔ پیشین اجرا می‌شود.

در لایهٔ چهارم بار دیگر عملیات بازسازی، به‌منظور افزایش ابعاد نقشه‌های ویژگی، اعمال می‌شود و لایهٔ پنجم عملیات حذف تصادفی را روی نتایج لایهٔ قبل اعمال می‌کند. خروجی این بلوک ۲۵۶ نقشهٔ ویژگی با ابعاد $\frac{m}{2} \times \frac{m}{2}$ است.

معماری بلوک سوم مشابه با معماری بلوک دوم است و همان لایه‌ها، با عملکردی مشابه، روی داده‌ها اعمال می‌شوند. خروجی بلوک سوم ۲۵۶ نقشهٔ ویژگی با ابعاد $m \times m$ است. با اینکه ابعاد خروجی (طول و عرض) این بلوک با ابعاد تصویر ورودی برابر است؛ برای

1. Reconstruction
2. Conv.
3. Stride
4. BN-Conv.
5. Batch Normalization (BN)
6. Stride
7. Padding
8. MaxPooling
9. Dropout
10. UP-Conv.
11. Concat.

بهبود دقت و کیفیت برچسب‌گذاری، دو بلوک دیگر نیز در معماری پیشنهادی لحاظ شده است.

در لایهٔ نخست بلوک چهارم، ابتدا نتایج لایه‌های قبلی با نقشه‌های به‌دست‌آمده از قسمت کدگذار ترکیب می‌شوند و سپس در لایه‌های بعدی، عملیات کانوولوشن روی داده‌ها اعمال می‌شود. لایهٔ آخر لایهٔ بازسازی با گام ۱ است. نتایج حاصل، به‌منزلهٔ ورودی آخرین بلوک، به یک لایهٔ کانوولوشن با اندازهٔ فیلتر ۱ × ۱ و سه نقشهٔ ویژگی داده می‌شوند تا بازسازی نهایی انجام شود. در آخرین گام، از یک تابع نگاشت برای به‌دست‌آوردن نتیجهٔ نهایی و برچسب‌های پیش‌بینی‌شده استفاده می‌شود.

## ۴- آزمایش‌ها

برای ارزیابی کارآیی روش پیشنهادی، آزمایش‌های متعددی روی مجموعه‌ای از تصاویر هوایی انجام شده و کارآیی روش پیشنهادی نیز، با دیگر روش‌های مشابه، مقایسه شده است. در این بخش، نحوهٔ انجام‌شدن آزمایش‌ها و نتایج تجربی به‌دست‌آمده بیان می‌شود.

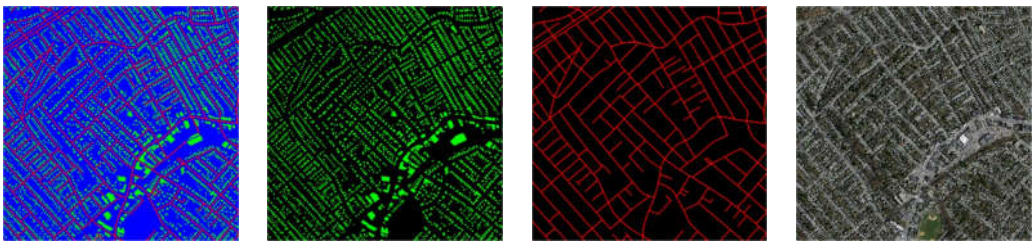
### ۴-۱- بانک تصاویر

در راستای ارزیابی معماری پیشنهادشده، از بانک تصاویر هوایی جاده‌ها و ساختمان‌های ماساچوست[1] استفاده می‌شود. این داده‌ها را مینهٔ (۲۰۱۳) گردآوری کرده و در اختیار عموم پژوهشگران قرار داده است. اندازهٔ تمامی تصاویر گردآمده ۱۵۰۰×۱۵۰۰ پیکسل و وضوح آنها ۱ مترمربع بر پیکسل (پیکسل‌های یک‌متری) است. هر تصویر تقریباً ۲.۲۵ کیلومترمربع از شهر بوستون را پوشش می‌دهد. تعداد تصاویر برداشت‌شده از جاده‌ها و ساختمان‌ها به‌ترتیب برابر با ۱۱۷۱ و ۱۵۱ است.

ازآن‌جاکه هدف، در معماری پیشنهادی، تشخیص هم‌زمان جاده و ساختمان و عوارض طبیعی است؛ مجموعه تصاویر جاده‌ها و ساختمان‌ها باید با هم ادغام شوند (Saito et al., 2016). اختلاف تعداد تصاویر در این دو مجموعه چالشی در برابر ادغام شمرده می‌شود. برای حل این چالش، می‌توان مجموعه تصاویر ساختمان‌ها را به‌منزلهٔ مبنا در نظر گرفت و تصاویر جادهٔ متناظر با پایگاه دادهٔ ساختمان‌ها را انتخاب کرد (Ibid.). به‌این‌ترتیب، از ۱۱۷۱ تصویر جاده، ۱۵۱ تصویر انتخاب می‌شوند که همتای آنها در مجموعه تصاویر ساختمان‌ها موجود است. در جدول ۱، جزئیات شیوهٔ تقسیم‌بندی بانک تصاویر برای انجام‌دادن آزمایش‌ها ارائه شده است. شکل ۵ نیز نمونه‌هایی از تصاویر و برچسب متناظرشان را که به‌صورت دستی ایجاد شده‌اند (Mnih, 2013)، نشان می‌دهد.

**جدول ۱.** جزئیات تقسیم‌بندی بانک تصاویر ماساچوست

| نام مجموعه | تعداد تصاویر آموزشی | تعداد تصاویر آزمایشی | تعداد تصاویر اعتبارسنجی |
|---|---|---|---|
| جاده | ۱۱۰۸ | ۴۹ | ۱۴ |
| ساختمان | ۱۳۷ | ۱۰ | ۴ |
| جاده و ساختمان | ۱۳۷ | ۱۰ | ۴ |



(الف) (ب) (پ) (ت)

**شکل ۵.** نمونه‌ای از تصاویر جاده‌ها و ساختمان‌های شهر بوستون در بانک تصاویر هوایی ماساچوست. تصویر هوایی(الف)؛ برچسب جاده‌ها (ب)؛ برچسب ساختمان‌ها (پ)؛ برچسب جاده‌ها و ساختمان‌ها (ت)

1. Massachusetts

**۴-۲- سخت‌افزار**

برای انجام‌دادن آزمایش‌ها، از سیستمی رایانه‌ای با پردازندهٔ Intel Xeon E5 2620 V4، شانزده گیگابایت حافظهٔ اصلی و پردازندهٔ گرافیکی GTX 1070 با هشت گیگابایت حافظه استفاده شده است. زبان برنامه‌نویسی به‌کاررفته پایتون و کتابخانهٔ استفاده‌شده برای یادگیری عمیق Keras[1] است.

**۴-۳- تنظیمات**

مطابق با پژوهش‌های مشابه (Miikkulainen et al., 2019)، تابع نگاشت به‌کاررفته در تمامی لایه‌ها ReLU است. مزیت اصلی این تابع، بر دیگر توابع نگاشت، جلوگیری از فعال‌شدن هم‌زمان همهٔ نورون‌هاست و نتایج تجربی بسیاری بیان می‌کند استفاده از این تابع عملکرد شبکه را، در مقایسه با توابع دیگر، بهبود چشمگیری می‌بخشد (Eckle & Schmidt-Hieber, 2019). اندازهٔ فیلتر ادغام ۲ × ۲ در نظر گرفته شده و از تابع هزینهٔ آنتروپی متقاطع[2] به‌همراه تابع نگاشت سافت‌مکس در لایهٔ خروجی استفاده شده است (Kozma et al., 2018).

**تابع سافت‌مکس:** این تابع نگاشت طبق رابطه (۱) تعریف می‌شود.

رابطه (۱) $$p_k(i,j) = \frac{\exp(a_k(i,j))}{\sum_{k'=1}^{K}\exp(a_{k'}(i,j))}$$

در رابطه (۱)، $a_k(i,j)$ مقدار پیکسل در موقعیت $(i,j)$ کانال $k$اُم نقشهٔ ویژگی تولیدشده در آخرین لایهٔ کانوولوشن است. K تعداد کلاس‌ها و $p_k(i,j)$ احتمال محاسبه‌شده در موقعیت $(i,j)$ کانال $k$ است.

**تابع آنتروپی متقاطع:** این تابع انحراف و تغییر $p_k(i,j)$ از مقدار درست آن در تصویر را براساس رابطه (۲) محاسبه می‌کند.

رابطه (۲) $$L = -\sum_{i=1}^{h}\sum_{j=1}^{w}\sum_{k=1}^{K} l_k(i,j)\ln\big(p_k(i,j)\big)$$

در رابطه (۲)، $w$ و $h$ به‌ترتیب عرض و طول تصویر، $l_k(i,j)$ مقدار پیکسل در موقعیت $(i,j)$ کانال $k$اُم تصویر برچسب مبنا[3] و $p_k(i,j)$ هم مقدار پیکسل $(i,j)$ کانال $k$اُم تصویر برچسب پیش‌بینی‌شده هستند.

در طول فرایند آموزش شبکه، وزن‌ها با استفاده از الگوریتم بهینه‌ساز RMSprop[4] (Aggarwal, 2018) به‌روز می‌شوند. مقادیر اَبرپارامترها (نرخ یادگیری، نرخ حذف تصادفی و کاهش نرخ یادگیری)، در هر به‌روزرسانی، براساس مقادیر پیش‌فرض پیشنهادی در کتابخانهٔ Keras انتخاب شده‌اند؛ نرخ یادگیری برابر با $۱۰^{-۳}$، نرخ حذف تصادفی برابر با ۰.۵ و کاهش نرخ یادگیری برابر با $۱۰^{-۵}$ در نظر گرفته می‌شوند.

**۴-۴- معیارهای ارزیابی سیستم**

دقت[5] (رابطه (۳)) و صحت[6] (رابطه (۴)) جزء معیارهای رایج برای ارزیابی نتایج تشخیص جاده و ساختمان در تصاویر هوایی محسوب می‌شوند. در روابط (۳) و (۴)، TP[7] و FP[8] و FN[9] به‌ترتیب نمایانگر مثبت‌های صحیح، مثبت‌های کاذب و منفی‌های کاذب‌اند. در تشخیص جاده یا ساختمان در تصاویر هوایی، «دقت» عبارت است از نسبت پیکسل‌های درست جاده یا ساختمان در تصاویرِ برچسب به تعداد پیکسل‌های تشخیص‌داده‌شده به‌منزلهٔ جاده یا ساختمان در تصاویرِ پیش‌بینی‌شده و «صحت» عبارت است از نسبت کل پیکسل‌های تشخیص‌داده‌شده به پیکسل‌های صحیح.

رابطه (۳) $$\mathcal{P} = \frac{TP}{TP+FP}$$

رابطه (۴) $$\mathcal{R} = \frac{TP}{TP+FN}$$

در پژوهش‌های مینهٔ (۲۰۱۳) و سایتو و همکاران (۲۰۱۶)، به‌جای استفاده از دقت و صحت برای ارزیابی نتایج آزمایش‌ها، عبارت دقت سست[10] و صحت سست[11]

1. https://keras.io/
2. Cross entropy
3. Ground truth
4. Root Mean Square Propagation
5. Precision
6. Recall
7. True Positive
8. False Positive
9. False Negative
10. Relaxed precision
11. Relaxed recall

به‌کار رفته است. «دقت سست» به بخشی از پیکسل‌های شناسایی‌شده در تصویر پیش‌بینی‌شده گفته می‌شود که در شعاع $\rho$ پیکسل برچسب قرار دارند و «صحت سست» به بخشی از پیکسل‌های صحیح اطلاق می‌شود که در شعاع $\rho$ پیکسل شناسایی‌شده‌اند. طبق پژوهش مینهٔ (۲۰۱۳)، مقدار $\rho$ در تمامی آزمایش‌ها برابر با ۳ پیکسل در نظر گرفته شده است.

یکی از شیوه‌های رایج برای ارزیابی کارآیی تشخیص نیز استفاده از منحنی دقت‌ـ صحت است که رابطهٔ بین دقت و صحت را برای آستانه‌های متفاوت، نشان می‌دهد (Mnih, 2013). به‌عبارتی، برای رسم این منحنی، نقاط گوناگون به‌منزلهٔ مجموعه‌ای از مقادیر دقت و صحت با آستانهٔ $t \in [0,1]$ محاسبه می‌شوند. سپس این منحنی با نقطهٔ سربه‌سر[1]، که در آن مقدار دقت و صحت با هم برابرند، خلاصه می‌شود. در این پژوهش نیز، از نقطهٔ سربه‌سر به‌منزلهٔ معیاری برای ارزیابی و مقایسهٔ کارآیی روش پیشنهادی استفاده می‌شود.

### ۴-۵- نتایج

برای نشان‌دادن کارآیی روش پیشنهادی و مقایسهٔ آن با کارهای انجام‌شدهٔ پیشین، شبکه یک‌بار به‌صورت مجزا (تک‌کلاس) با مجموعه تصاویر جاده، یک‌بار به‌صورت مجزا (تک‌کلاس) با مجموعه تصاویر ساختمان و بار دیگر، به‌صورت یک‌جا (چندکلاس) با مجموعه تصاویر جاده و ساختمان ارزیابی می‌شود (جدول ۱).

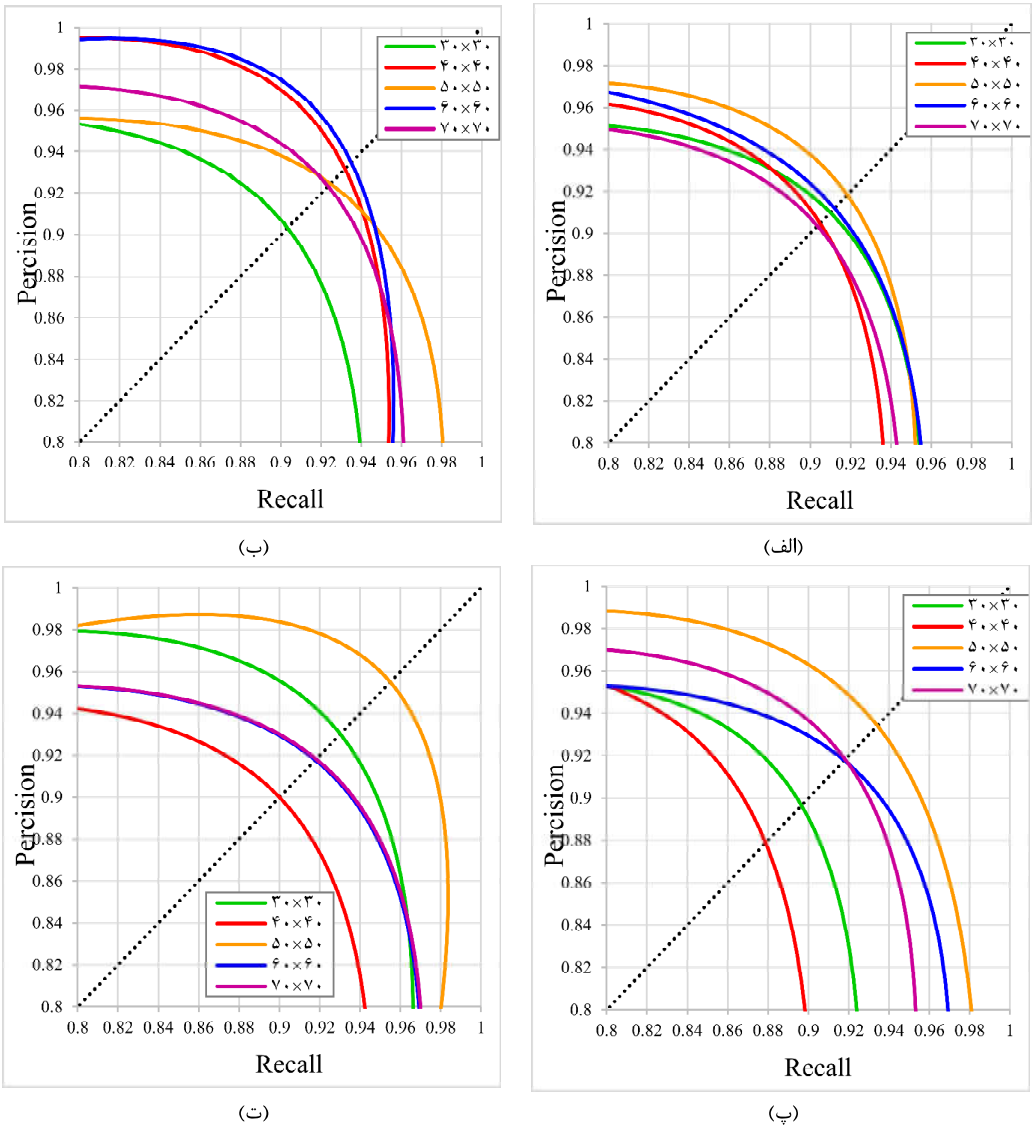
با توجه به اینکه راهکار پیشنهادی مبتنی‌بر استفاده از بلوک‌بندی تصویر ورودی به مجموعه‌ای از قطعات است؛ بنابراین، لازم است اندازهٔ مناسب قطعات در معماری پیشنهادی پیدا شود. به همین منظور، آزمایش‌های گوناگونی با اندازهٔ قطعات متفاوت (۳۰ × ۳۰، ۴۰ × ۴۰، ۵۰ × ۵۰، ۶۰ × ۶۰ و ۷۰ × ۷۰) انجام شده است. توضیح اینکه بلوک‌بندی تصاویر به‌روش غیرهم‌پوشانی انجام می‌گیرد؛ به‌این‌ترتیب، قطعات ایجادشده هیچ بخش مشترکی با هم ندارند. جدول ۲ نتایج به‌دست‌آمده از این آزمایش‌ها را نشان می‌دهد. در این جدول، بهترین نتایج در هر ستون با قلم ضخیم مشخص شده‌اند. همچنین، منحنی‌های دقت‌ـ صحت متناظر نیز در شکل ۶ نشان داده شده‌اند.

بررسی نتایج جدول ۲ نشان می‌دهد استفاده از قطعاتی با ابعاد ۵۰×۵۰ و ۶۰×۶۰، در روش پیشنهادی، کارآیی بهتری در مقایسه با دیگر ابعاد دارد. در این میان، با توجه به اینکه کارآیی روش پیشنهادی با قطعاتی با ابعاد ۵۰×۵۰ اغلب بیشتر از قطعات دارای ابعاد ۶۰×۶۰ است، می‌توان نتیجه گرفت اندازهٔ مناسب قطعه‌بندی، در روش پیشنهادی برای تصاویر هوایی ماساچوست با تراکم ۱ مترمربع بر پیکسل، ۵۰×۵۰ است.

**جدول ۲.** ارزیابی دقت ($\mathcal{P}$)، صحت ($\mathcal{R}$) و نقطهٔ سربه‌سر ($\mathcal{B}$) روش پیشنهادی برای تشخیص جاده و ساختمان از عوارض طبیعی، با درنظرگرفتن ابعاد گوناگون قطعه‌بندی

| اندازهٔ قطعه‌ها | تک‌کلاس | | | | | | چندکلاس | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | جاده | | | ساختمان | | | جاده | | | ساختمان | | |
| | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ |
| ۳۰ × ۳۰ | ۹۰.۹ | ۸۹.۱ | ۹۰.۱ | ۹۱.۲ | ۸۹.۵ | ۹۰.۳ | ۸۲.۶ | ۸۷.۸ | ۸۷.۹ | ۸۵.۲ | ۸۷.۳ | ۸۶.۶ |
| ۴۰ × ۴۰ | ۹۱.۲ | ۹۱.۱ | ۹۴.۰ | ۹۳.۰ | ۹۰.۹ | ۹۳.۲ | ۹۲.۰ | ۹۲.۳ | ۹۱.۶ | ۹۳.۹ | ۸۸.۳ | ۹۲.۴ |
| ۵۰ × ۵۰ | **۹۳.۱** | ۹۰.۸ | **۹۴.۹** | ۹۶.۲ | **۹۴.۰** | **۹۵.۶** | **۹۴.۸** | ۹۲.۱ | **۹۳.۵** | **۹۶.۶** | **۹۲.۳** | **۹۳.۲** |
| ۶۰ × ۶۰ | ۹۱.۱ | **۹۱.۲** | ۹۴.۲ | **۹۶.۶** | ۹۱.۳ | ۹۳.۵ | ۹۰.۹ | **۹۲.۸** | ۹۱.۹ | ۹۶.۳ | ۹۱.۶ | ۹۱.۷ |
| ۷۰ × ۷۰ | ۹۰.۸ | ۹۰.۰ | ۹۲.۳ | ۹۳.۰ | ۹۱.۸ | ۹۲.۳ | ۹۲.۹ | ۹۰.۹ | ۹۱.۹ | ۹۵.۹ | ۹۰.۱ | ۹۱.۹ |

1. Breakeven Point




**شکل ۶.** منحنی‌های دقت‌ـ صحت حاصل از روش پیشنهادی به‌ازای اندازهٔ قطعات متفاوت: جاده‌ـ تک‌کلاس (الف)؛ ساختمان‌ـ تک‌کلاس (ب)؛ جاده‌ـ چندکلاس (پ)؛ ساختمان‌ـ چندکلاس (ت)

همچنین، ملاحظه می‌شود که دقت بازشناسایی ساختمان، در هر دو حالت تک‌کلاس و چندکلاس، بیشتر از دقت تشخیص جاده است. دلیل این موضوع را می‌توان ابعاد بزرگ‌تر ساختمان‌ها در تصاویر دانست که باعث افزایش قدرت تفکیک‌پذیری آنها از عوارض طبیعی (مانند چمن‌زارها، درخت‌ها، رودخانه‌ها) می‌شود. نکتهٔ مهم دیگری که از بررسی این جدول و منحنی‌های دقت‌ـ صحت حاصل می‌شود این است که میانگین درصد دقت بیشتر از میانگین درصد صحت است (۲.۸٪ بزرگ‌تر، با درنظرگرفتن قطعات دارای ابعاد ۵۰×۵۰) و این نشان می‌دهد که کیفیت شناسایی جاده‌ها و ساختمان‌ها از عوارض طبیعی مطلوب است.

بررسی نقاط سربه‌سر در حالت تک‌کلاس و چندکلاس نیز نشان می‌دهد کارآیی معماری، در حالت

چندکلاس، اندکی کمتر از حالت تک‌کلاس است؛ ۱.۴٪ کاهش در تشخیص جاده و ۲.۴٪ کاهش در تشخیص ساختمان. دلیل چنین کاهشی را می‌توان در دو عامل جستجو کرد. نخست اینکه تعداد تصاویر آموزشی در این آزمایش، به‌نسبت تعداد تصاویر آموزشی در حالت تک‌کلاس، به‌دلیل یکسان‌سازی، کاهش چشمگیری داشته و در نتیجه، قدرت یادگیری شبکه کاهش یافته است. دوم اینکه فرموله‌سازی مسئله به‌شکل چندکلاسی باعث می‌شود احتمال قرارگیری پیکسل‌های یک کلاس در کلاس‌های دیگر افزایش پیدا کند (Nogueira et al., 2017).

شکل ۷ نمونه‌ای از نتایج به‌دست‌آمده از تشخیص جاده‌ها و ساختمان‌ها در روش پیشنهادی را نشان می‌دهد. طبق شکل ۷-الف، به‌دلیل مساحت کمی که جاده‌ها در کل تصویر اشغال کرده‌اند، معماری پیشنهادی در فاز کدگشایی و حین بازسازی نقشه‌های ویژگی، اندکی از اطلاعات مربوط به جاده‌ها را از دست داده و در مواقعی، به‌اشتباه، عوارض طبیعی و ساختمان‌ها را به‌منزلهٔ جاده شناسایی کرده است. اما چنانکه شکل ۷- ب نشان می‌دهد، این مشکل به‌ندرت در تشخیص ساختمان‌ها رخ داده است. شکل‌های ۷- پ و ۷- ت نیز عملکرد روش پیشنهادی را در تشخیص هم‌زمان جاده‌ها و ساختمان‌ها از عوارض طبیعی نشان می‌دهند. نکتهٔ مهم در این شکل‌ها آن است که، حتی با توجه به استفاده از معماری یکپارچه، قابلیت تشخیص هم‌زمان جاده و ساختمان به‌شکل مطلوبی حفظ شده و سیستم، فقط در تراکم شدید و حضور سایه، دچار اشتباه شده است.

نکتهٔ درخور توجه دیگری که در نتایج نشان‌داده‌شده در شکل ۷ وجود دارد عبارت است از اینکه مرزهای نواحی برچسب‌دارشده در روش پیشنهادی چگال‌اند و پیوستگی بسیار مطلوبی دارند (برای مقایسه، رک. شکل ۱). دلیل چنین خروجی مطلوبی را می‌توان در آرایش مناسب لایه‌های بخش کدگشای معماری پیشنهادی دانست که با آزمایش‌های متعددی، حاصل شده است.

### ۴-۵- مقایسه با سایر روش‌ها

در این بخش، کارآیی روش پیشنهادی با دیگر روش‌های شناخته‌شدهٔ تشخیص جاده و ساختمان، در تصاویر هوایی، مقایسه و ارزیابی می‌شود. ملاک انتخاب روش‌های مورد مقایسه، نخست، بانک تصاویر هوایی است که در همگی آنها، به‌طور مشترک، از بانک تصاویر ماساچوست استفاده شده است و دوم اینکه تمامی این روش‌ها از معماری کانوولوشنی استفاده کرده‌اند.

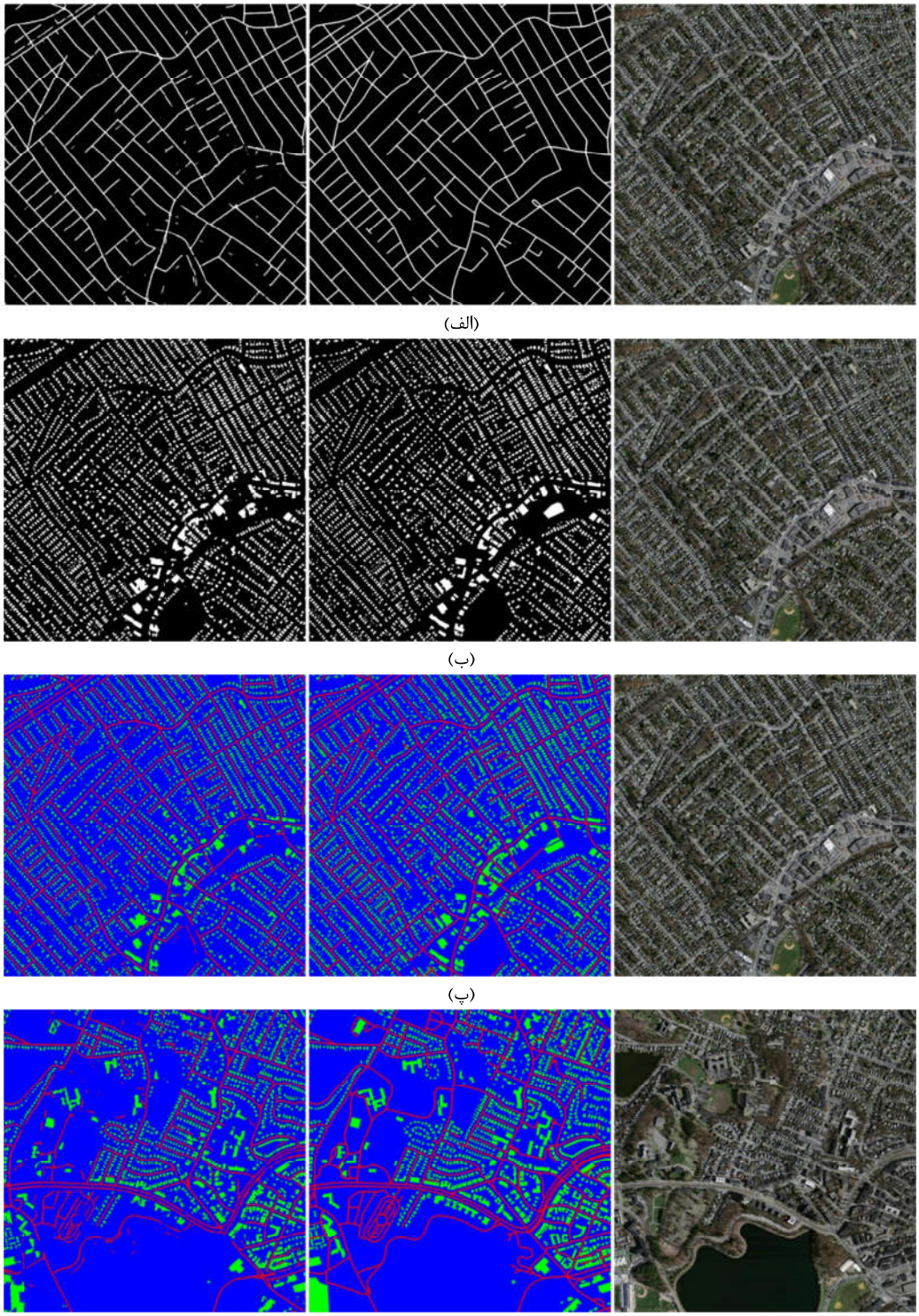
در بیشتر روش‌ها، به‌دلیل ناتوانی در تشخیص هم‌زمان جاده و ساختمان از عوارض طبیعی، نتایج با درنظرگرفتن فقط یک کلاس (جاده یا ساختمان و عوارض طبیعی) مطرح شده است. باید اشاره کرد که در برخی از این پژوهش‌ها، از نقطهٔ سربه‌سر به‌منزلهٔ معیار ارزیابی استفاده نشده است؛ در این‌گونه موارد، معیارهای دقت و صحت گزارش‌شده مد نظر قرار گرفته‌اند.

شایان ذکر است که برای مقایسهٔ عادلانهٔ مدت زمان آموزش معماری پیشنهادی با روش‌های دیگر، پژوهشگران این مقاله تمامی الگوریتم‌های مورد مقایسه را روی سخت‌افزاری یکسان (رک. بخش ۴-۲) اجرا کرده و نتایج را در جدول ۳ گزارش داده‌اند[۱].

با مطالعهٔ ستون رویکرد در جدول ۳، می‌توان به این نتیجه رسید که هیچ‌یک از روش‌های مورد مقایسه هم‌زمان از رویکرد چندکلاس و کاملاً کانوولوشنی برای تشخیص بهره نبرده است. در حقیقت، در معماری پیشنهادی، با ترکیب رویکرد چندکلاس و معماری کاملاً کانوولوشنی، مدل یکپارچه‌ای به‌دست آمده که هم‌زمان قادر به تشخیص جاده و ساختمان از عوارض طبیعی است. نتیجهٔ مستقیم این معماری کاهش پیچیدگیِ مسئلهٔ تشخیص و نیز کاهش مدت زمان آموزش و ساخت مدل یکپارچه با یک‌بار آموزش است و این درحالی است که در روش‌های تک‌کلاس، لازم است یک مدل مجزا برای هر کلاس ساخته شود.

---

۱. در بیشتر موارد، در مکاتبه با نویسندگان مقالات مورد نظر، کدهای الگوریتم‌های مورد مقایسه در اختیار پژوهشگران این مقاله قرار گرفته است. درمواردی که این کدها در دسترس نبود، پژوهشگران این مقاله آنها را اعمال کرده‌اند.



(الف)

(ب)

(پ)

(ت)

**شکل ۷.** نمونه‌ای از عملکرد روش پیشنهادی در تشخیص جاده‌ها و ساختمان‌های بانک تصاویر هوایی ماساچوست: تشخیص جاده‌ها (تک‌کلاس) (الف)؛ تشخیص ساختمان‌ها (تک‌کلاس) (ب)؛ تشخیص جاده‌ها و ساختمان‌ها (چندکلاس) (پ و ت)

با مطالعهٔ مدت زمان ساخت مدل، این نتیجه به‌دست می‌آید که زمان صرف‌شده برای این کار، براساس معماری پیشنهادی، تقریباً برابر با سایر معماری‌های کاملاً کانوولوشنی است و این نشان می‌دهد طراحی معماری پیشنهادی باعث به‌وجودآمدن هزینهٔ محاسباتی معتنابهی نشده است.

مقایسهٔ مدت زمان ساخت مدل با استفاده از معماری پیشنهادی با رویکردهای چندکلاسی، که از لایهٔ کاملاً متصل بهره می‌برند (Saito et al., 2016; Alshehhi et al., 2017)، نشان می‌دهد زمان ساخت مدل با معماری پیشنهادی تقریباً ۳۸٪ از شیوهٔ سایتو و همکاران (۲۰۱۶) و ۳۶٪ از روش اَلشِحی و همکاران (۲۰۱۷) سریع‌تر است.

مقایسهٔ نقطهٔ سربه‌سر روش‌های مطرح‌شده در دو پژوهش سطرهای پیشین نیز نشان می‌دهد که معماری پیشنهادی تقریباً ۲٪ این معیار را افزایش داده است. علاوه‌براین، مقایسهٔ دقت و صحت معماری پیشنهادی، با درنظرگرفتن فقط یک کلاس، نیز نشان می‌دهد روش پیشنهادی، با وجود استفاده از رویکرد چندکلاسی، دقت و صحت روش مورد اشاره در تحقیق پنبونیون و همکاران (۲۰۱۷) را به‌ترتیب ۷.۲٪ و ۱.۴٪ افزایش داده و دقت آن، در تشخیص جاده، فقط ۱.۱٪ کمتر از معماری مورد نظر پژوهش هوئی و همکاران (۲۰۱۸) است.

## ۵- نتیجه‌گیری و محورهای مطالعه و توسعهٔ بیشتر

تا کنون پژوهش‌های بسیاری، به‌منظور خودکارسازی تشخیص جاده‌ها و ساختمان‌ها در تصاویر هوایی، صورت گرفته است. در پژوهش‌های اخیر، با توجه به نتایج خیره‌کنندهٔ به‌دست‌آمده ازطریق شبکه‌های عصبی کانوولوشنی در حوزهٔ بینایی ماشین، استفاده از این شبکه‌ها بیشتر مورد توجه قرار گرفته اما در اغلب این کارها، از رویکرد تشخیص مجزای جاده‌ها و ساختمان‌ها استفاده شده است. به‌عبارتی، مدل‌های مجزایی برای تشخیص جاده و ساختمان از عوارض طبیعی ساخته می‌شوند. همچنین، رویکرد اصلی در این پژوهش‌ها استفاده از معماری‌های کانوولوشنی با یک یا چند لایهٔ کاملاً متصل است. در این پژوهش، با هدف کاهش پیچیدگی مدل و افزایش سرعت ساخت آن، شبکه‌ای کاملاً کانوولوشنی مورد توجه قرار گرفت که قادر به تشخیص هم‌زمان جاده و ساختمان است. روش پیشنهادی، با اندازهٔ قطعات متفاوت، روی بانک تصاویر هوایی ماساچوست آزموده شد و نتایج نشان داد این روش، با قطعات به‌ابعاد ۵۰×۵۰، کارآیی بهتری دارد. مقایسه و ارزیابی معماری پیشنهادی با سایر معماری‌ها نیز حاکی از این بود که پیچیدگی معماری پیشنهادی به‌طور مطلوبی کم است و مدت زمان ساخت مدل براساس آن تقریباً ۳۸٪ کمتر از دیگر معماری‌های

**جدول ۳.** مقایسهٔ کارآیی روش پیشنهادی با دیگر روش‌های مبتنی‌بر معماری کانوولوشنی، در تشخیص ساختمان‌ها و جاده‌ها از عوارض طبیعی، با استفاده از بانک تصاویر ماساچوست

| پژوهش | سال | جاده | | | ساختمان | | | جاده و ساختمان | زمان (ساعت) | رویکرد |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ | $\mathcal{P}$ | $\mathcal{R}$ | $\mathcal{B}$ | $\mathcal{B}$ | | |
| مینهٔ، ۲۰۱۳ | ۲۰۱۳ | – | – | ۹۰.۰ | – | – | ۹۲.۰ | – | ۱۰۸ | تک‌کلاس- کانوولوشنی با لایهٔ کاملاً متصل |
| سایتو و دیگران، ۲۰۱۶ | ۲۰۱۶ | – | – | ۹۰.۰ | – | – | ۹۴.۳ | ۹۳.۳ | ۱۰۱ | چندکلاس- کانوولوشنی با لایهٔ کاملاً متصل |
| پنبونیون و دیگران، ۲۰۱۷ | ۲۰۱۷ | ۸۵.۸ | ۸۹.۴ | – | – | – | – | – | ۶۵ | تک‌کلاس- کاملاً کانوولوشنی |
| اَلشِحی و دیگران، ۲۰۱۷ | ۲۰۱۷ | – | – | – | – | – | – | ۹۳.۲ | ۹۸ | چندکلاس- کانوولوشنی با لایهٔ کاملاً متصل |
| هوئی و دیگران، ۲۰۱۸ | ۲۰۱۸ | ۹۴.۲ | – | – | – | – | – | – | ۵۹ | تک‌کلاس- کاملاً کانوولوشنی |
| روش پیشنهادی | | ۹۳.۱ | ۹۰.۸ | ۹۴.۹ | ۹۶.۲ | ۹۴.۰ | ۹۵.۶ | ۹۵.۱ | ۶۳ | چندکلاس- کاملاً کانوولوشنی |

چندکلاسی است. علاوه‌براین، نشان داده شد که بهبود دقت مدل ساخته‌شده در مقایسه با مدل‌های چندکلاسی، ۲٪ بیشتر است.

یافته‌ها و نتایج این پژوهش نویسندگان را بر آن داشته است به موارد زیر به‌صورت نقشهٔ راه پژوهش‌های آتی خود، با هدف بهبود دقت و تسریع تشخیص اشیا در تصاویر هوایی، توجه داشته باشند:

- بررسی کارآیی توابع نگاشت متفاوت و پیشنهاد تابع نگاشت جدیدی متناسب با مسئلهٔ مفروض؛
- استفاده از قطعه‌بندی با درنظرگرفتن هم‌پوشانی، برای جلوگیری از حذف اطلاعات لبه در تصاویر.

## ۶- منابع


فرج‌زاده، ن.، هاشم‌زاده، م.، ۱۳۹۸، **تشخیص سازه‌های ساخت بشر در تصاویر هوایی با استفاده از ویژگی‌های آماری مبتنی‌بر رنگ و یادگیری ماشین**، سنجش از دور و GIS ایران، سال یازدهم، شمارهٔ ۳، صص. ۴۲-۲۱.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Süsstrunk, S., 2012, **SLIC Superpixels Compared to State-of-the-Art Superpixel Methods**, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), PP. 2274-2282.

Aggarwal, C.C., 2018, **Neural Networks and Deep Learning**, Springer.

Akçay, H.G. & Aksoy, S., 2010, **Building Detection Using Directional Spatial Constraints**, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, IEEE.

Alshehhi, R., Marpu, P.R., Woon, W.L., Dalla Mura, M. & Sensing, R., 2017, **Simultaneous Extraction of Roads and Buildings in Remote Sensing Imagery with Convolutional Neural Networks**, ISPRS Journal of Photogrammetry and Remote Sensing, 130, PP. 139-149.

Arı, Ç., Aksoy, S. & Sensing, R., 2014, **Detection of Compound Structures Using a Gaussian Mixture Model with Spectral and Spatial Constraints**, IEEE Transactions on Geoscience and Remote Sensing, 52(10), PP. 6627-6638.

Bai, X., Zhang, H. & Zhou, J., 2014, **VHR Object Detection Based on Structural Feature Extraction and Query Expansion**, IEEE Transactions on Geoscience and Remote Sensing, 52(10), PP. 6508-6520.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A., 2014, **Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs**.

Chen, L., Zhu, Q., Xie, X., Hu, H. & Zeng, H., 2018, **Road Extraction from VHR Remote-Sensing Imagery via Object Segmentation Constrained by Gabor Features**, ISPRS Int. J. Geo-Inf., 7(9), P. 362.

Cheng, G. & Han, J., 2016, **A Survey on Object Detection in Optical Remote Sensing Images**, ISPRS Journal of Photogrammetry and Remote Sensing, 117, PP. 11-28.

Cheng, Y., Wang, D., Zhou, P. & Zhang, T., 2018, **Model Compression and Acceleration for Deep Neural Networks: The Principles, Progress, and Challenges**, IEEE Signal Processing Magazine, 35(1), PP. 126-136.

Chollet, F., 2017, **Xception: Deep Learning with Depthwise Separable Convolutions**, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Clinton, N., Holt, A., Scarborough, J., Yan, L. & Gong, P., 2010, **Accuracy Assessment Measures for Object-Based Image Segmentation Goodness**, Photogramm. Eng. Remote Sens, 76(3), PP. 289-299.

Contreras, D., Blaschke, T., Tiede, D., Jilge, M.J.C. & Science, G.I., 2016, **Monitoring Recovery after Earthquakes through the Integration of Remote sensing, GIS, and Ground Observations: The Case of L'Aquila (Italy)**, Cartography and Geographic Information Science, 43(2), PP. 115-133.

Das, S., Mirnalinee, T., Varghese, K. & Sensing, R., 2011, **Use of Salient Features for the Design of a Multistage Framework to Extract Roads from High-Resolution Multispectral Satellite Images**, IEEE Transactions on Geoscience and Remote Sensing, 49(10), PP. 3906-3931.

Eckle, K. & Schmidt-Hieber, J., 2019, **A Comparison of Deep Networks with ReLU Activation Function and Linear Spline-Type Methods**, Neural Networks, 110, PP. 232-242.

Feizizadeh, B., Tiede, D., Rezaei Moghaddam, M.H. & Blaschke, T., 2014, **Systematic Evaluation of Fuzzy Operators for Object-Based Landslide Mapping**, South-Eastern European Journal of Earth Observation and Geomatics, 3(2s), PP. 219-222.

Goodin, D.G., Anibas, K.L. & Bezymennyi, M., 2015, **Mapping Land Cover and Land Use from Object-Based Classification: An Example from a Complex Agricultural Landscape**, International Journal of Remote Sensing, 36(18), PP. 4702-4723.

Grabner, H., Nguyen, T.T., Gruber, B. & Bischof, H., 2008, **On-Line Boosting-Based Car Detection from Aerial Images**, ISPRS Journal of Photogrammetry and Remote Sensing, 63(3), PP. 382-396.

Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard A., 2003, **A Comparison of Three Image-Object Methods for the Multiscale Analysis of Landscape Structure**, ISPRS Journal of Photogrammetry and Remote Sensing, 57(5-6), PP. 327-345.

Hinton, G.E. & Salakhutdinov, R.R., 2006, R**educing the Dimensionality of Data with Neural Networks**, Science, 313(5786), PP. 504-507.

Hui, J., Du, M., Ye, X., Qin, Q. & Sui, J., 2018, **Effective Building Extraction From High-Resolution Remote Sensing Images With Multitask Driven Deep Neural Network**, IEEE Geoscience and Remote Sensing Letters, 16(5), PP. 786-790.

Ioffe, S. & Szegedy, C., 2015, **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**.

Kluckner, S. & Bischof, H., 2009, **Semantic Classification by Covariance Descriptors within a Randomized Forest**, Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, IEEE.

Kluckner, S., Mauthner, T., Roth, P.M. & Bischof, H., 2009, **Semantic Classification in Aerial Imagery by Integrating Appearance and Height Information**, Asian Conference on Computer Vision, Springer.

Kozma, R., Alippi, C., Choe, Y. & Morabito, F.C., 2018, **Artificial Intelligence in the Age of Neural Networks and Brain Computing**, Academic Press.

Lefèvre, S. & Weber, J., 2007, **Automatic Building Extraction in VHR Images Using Advanced Morphological Operators**, Urban Remote Sensing Joint Event, 2007, IEEE.

Leitloff, J., Hinz, S. & Stilla, U., 2010, **Vehicle Detection in Very High Resolution Satellite Images of City Areas**, IEEE Transactions on Geoscience and Remote Sensing, 48(7), PP. 2795-2806.

Leninisha, S. & Vani, K., 2015, **Water Flow Based Geometric Active Deformable Model for Road Network**, ISPRS Journal of Photogrammetry and Remote Sensing**,** 102, PP. 140-147.

Li, E., Femiani, J., Xu, S., Zhang, X. & Wonka, P., 2015, R**obust Rooftop Extraction from Visible Band Images Using Higher Order CRF**, IEEE Transactions on Geoscience and Remote Sensing, 53(8), PP. 4483-4495.

Lin, Y., He, H., Yin, Z. & Chen, F., 2015, **Rotation-Invariant Object Detection in Remote Sensing Images Based on Radial-Gradient Angle**, IEEE Geoscience and Remote Sensing Letters, 12(4), PP. 746-750.

Liu, G., Sun, X., Fu, K. & Wang, H., 2013, **Aircraft Recognition in High-Resolution Satellite Images Using Coarse-to-Fine**

**Shape Prior**, IEEE Geoscience and Remote Sensing Letters, 1(3), PP. 573-577.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. & Alsaadi, F.E., 2017, **A Survey of Deep Neural Network Architectures and their Applications**, Neurocomputing, 234, PP.11-26.

Long, J., Shelhamer, E. & Darrell, T., 2015, **Fully Convolutional Networks for Semantic Segmentation**, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Maggiori, E., Tarabalka, Y., Charpiat, G. & Alliez, P., 2017, **Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification**, IEEE Transactions on Geoscience and Remote Sensing, 55(2), PP. 645-657.

Mayer, H., 1999, **Automatic Object Extraction from Aerial Imagery—A Survey Focusing on Buildings**, Computer Vision and Image Understanding, 74(2), PP. 138-149.

Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A. & Duffy, N., 2019, **Evolving Deep Neural Networks**, Artificial Intelligence in the Age of Neural Networks and Brain Computing, Elsevier, PP. 293-312.

Minh, V., 2013, **Machine Learning for Aerial Image Labeling**, University of Toronto (Canada).

Nogueira, K., Penatti O.A.B. & dos Santos, J.A., 2017, **Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification**, Pattern Recognition, 61, PP. 539-556.

Ok, A.O., Senaras, C. & Yuksel, B., 2013, **Automated Detection of Arbitrarily Shaped Buildings in Complex Environments from Monocular VHR Optical Satellite Imagery**, IEEE Transactions on Geoscience and Remote Sensing, 51(3), PP. 1701-1717.

Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S., Srestasathiern, P. & Vateekul, P., 2017, **Road Segmentation of Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields**, Remote Sensing, 9(7), P. 680.

Ronneberger, O., Fischer, P. & Brox, T., 2015, **U-net: Convolutional Networks for Biomedical Image Segmentation**, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer.

Saito, S., Yamashita, T. & Aoki, Y., 2016, **Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks**, Journal of Imaging Science and Technology, 60(1).

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J.-M., Bosch, S. & van Gerven, M.J.N., 2018, **Convolutional Neural Network-Based Encoding and Decoding of Visual Object Recognition in Space and Time**, NeuroImage, 180, PP. 253-266.

Song, M., Civco, D. & Sensing, R., 2004, **Road Extraction Using SVM and Image Segmentation**, American Society for Photogrammetry and Remote Sensing, 70(12), PP. 1365-1371.

Sun, H., Sun, X., Wang, H., Li, Y. & Li, X., 2012, **Automatic Target Detection in High-Resolution Remote Sensing Images Using Spatial Sparse Coding Bag-of-Words Model**, IEEE Geoscience and Remote Sensing Letters, 9(1), PP. 109-113.

Tuermer, S., Kurz, F., Reinartz, P. & Stilla, U., 2013, **Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(6), PP. 2327-2337.

Walker, J. & Blaschke, T., 2008, **Object Based Land Cover Classification for the Phoenix Metropolitan Area: Optimization vs. Transportability**, International Journal of Remote Sensing, 29(7), PP. 2021-2040.

Wang, H., Nie, F., Huang, H. & Ding, C., 2013, **Heterogeneous Visual Features Fusion via Sparse Multimodal Machine**, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Wang, J., Song, J., Chen, M. & Yang, Z., 2015, **Road Network Extraction: A Neural-Dynamic Framework Based on Deep Learning and a Finite State Machine**, International Journal of Remote Sensing, 36(12), PP. 3144-3169.

Yokoya, N. & Iwasaki, A., 2015, **Object Detection Based on Sparse Representation and Hough Voting for Optical Remote Sensing Imagery**, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(5), PP. 2053-2062.

Zhao, Y.-Q. & Yang, J., 2015, **Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint**, IEEE Transactions on Geoscience and Remote Sensing, 53(1), PP. 296-308.

سنجش از دور
و
GIS ایران

سنجش از دور و GIS ایران سال سیزدهم، شماره دوم، تابستان ۱۴۰۰
Iranian Remote Sensing & GIS Vol.13, No. 2, Summer 2021
39-60

# A Fully Convolutional Neural Network-Based Approach for Detecting Simultaneously Roads and Buildings in Aerial Imagery

Farajzadeh N.[1*] and Ebrahimzadeh H.[2]

1. Associate Prof., Faculty of IT and Computer Engineering Dep., Azarbaijan Shahid Madani University, Tabriz
2. M.Sc. Student, Faculty of IT and Computer Engineering Dep., Azarbaijan Shahid Madani University, Tabriz

**Abstract**

The development of automatic road and building detection systems in aerial imagery are always faced with challenges such as the appearance of buildings, illumination changes, imaging angles, and the density of roads and buildings in urban areas, to name a few. In recent years, employing multi-layered approach in artificial neural networks, known as deep neural networks, has attracted many researchers in this field (and the other fields alike), achieving stunning results. However, the use of fully connected layers in this approach, significantly increases the average processing time and results in an overfitted model. In addition, in most of these methods, a single-class approach has been considered. That is, detecting the roads and the buildings from natural scenes is not possible at the same time, and therefore, it is necessary to build separate binary models for each of them. The main goal of this research is to design a new architecture by which the produced model can be able to simultaneously detect roads and buildings from natural scenes, and thus minimizing the complexity of the classification process. In addition, in the proposed architecture, excluding all fully connected layers from the traditional multi-layered architectures is considered in order to reduce the average processing time. The results of the experiments performed on the Massachusetts dataset, show that the proposed architecture performs 38% faster than the other deep neural network-based methods, and also increases the accuracy by an average of 2%.

**Keywords:** Deep learning, Artificial neural networks, Convolutional neural networks, Aerial imagery, Road detection, Building detection, Natural scene detection, Artificial intelligence.

* Correspondence Address: Faculty of IT & Computer Engineering, Azarbijan Shahid Madani University, Tabriz, Post Code: 5375171379. Tel: +98 930 814 9600
Email: n.farajzadeh@azaruniv.ac.ir