



ترکیب داده‌های لایدار و تصاویر هوایی بر مبنای شبکه‌های عصبی کانولوشن به منظور تشخیص مدل ساختمان‌ها

فاطمه علی دوست^۱، حسین عارفی^{۲*}

۱- دانشجوی دکتری فتوگرامتری، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی دانشگاه تهران
۲- استادیار دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی دانشگاه تهران

تاریخ دریافت مقاله: ۱۳۹۴/۱۲/۱۶ تاریخ پذیرش مقاله: ۱۳۹۵/۰۷/۱۸

چکیده

ساختمان‌ها یکی از مهمترین سازه‌های شهری هستند که در کاربردهای مختلف و در نقشه‌برداری شهری مورد استفاده قرار می‌گیرند. در سال‌های اخیر، با توسعه تکنولوژی اخذ داده‌ها با توان تفکیک بالا، روش‌ها و الگوریتم‌های مختلفی به منظور استخراج مدل‌های دقیق و بهنگام ساختمان‌ها ارائه شده است. در این مقاله، روشی نوین و مدل‌مبنا به منظور استخراج ساختمان‌ها و شناسایی اتوماتیک مدل سقف آنها از قبیل سقف مسطح، شیروانی، شیبدار و هرمی ارائه شده است که در آن از شبکه‌های عصبی با معماری عمیق به منظور یادگیری سلسله مراتبی ویژگی‌های استخراج شده از داده‌های لایدار و تصاویر ارتوفتو استفاده می‌شود. مهمترین مراحل این روش عبارتند از: آموزش مدل و یادگیری، بخش‌بندی تصویر، استخراج ویژگی، و برچسب زدن عوارض. کلیه این مراحل در یک ساختار نظارت شده و با استفاده از یک شبکه عصبی کانولوشن که از قبل آموزش دیده شده است، اجرا می‌شوند تا یک سیستم تشخیص الگوی اتوماتیک برای تشخیص انواع مختلف ساختمان‌ها در یک ناحیه شهری فراهم گردد. در این روش، اطلاعات ارتفاعی، تولیدکننده‌ی ویژگی‌های هندسی پایدار برای شبکه عصبی کانولوشن هستند که در تعیین موقعیت محدوده هر سقف به کار گرفته می‌شوند. شبکه عصبی کانولوشن یکی از انواع شبکه‌های عصبی رو به جلو و با مفهوم درک و فهم چندلایه‌ای است که شامل تعدادی لایه کانولوشن و نمونه‌برداری می‌باشد. از آنجایی که در روش پیشنهادی، مجموعه داده‌ی آموزشی یک کتابخانه کوچک از مدل‌های برچسب‌دار است، لذا زمان محاسباتی برای یادگیری با استفاده از مدل‌های از قبل آموزش دیده، به طور قابل توجهی کم و در حدود چند ساعت است. نتایج حاصله، نشان‌دهنده موثر بودن تلفیق داده‌های ارتفاعی و تصاویر رنگی با هم در یادگیری عمیق به منظور استخراج ساختمان‌ها و شناسایی مدل سقف آنها به صورت همزمان است به طوری که خطای حد بالای اول و دقت آموزش مدل حاصل از تلفیق این دو دسته داده به ترتیب حدود ۰/۰۵ و ۹۵ درصد است. همچنین، میزان موفقیت و صحت شناسایی ساختمان‌ها به ترتیب حدود ۹۷ و ۶۹ درصد است.

کلیدواژه‌ها: یادگیری عمیق، شبکه‌ی عصبی کانولوشن، تشخیص الگو، لایدار، مدل سه بعدی ساختمان.

*نویسنده مسئول: گروه فتوگرامتری، دانشکده مهندسی نقشه برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی دانشگاه تهران، تهران، ایران.

۱- مقدمه

از طرفی، با به کارگیری اسکنرهای لیزری، مختصات سه بعدی نقاط با سرعت و دقت بالا اخذ می‌شود، با این حال، این داده‌ها اطلاعات طیفی عوارض را شامل نمی‌شود. به علت نقاط قوت و ضعف هر یک از این نوع داده‌ها، ترکیب داده‌های تصویری دو بعدی و داده‌های لیزر سه بعدی، به عنوان اطلاعات مکمل به منظور شناسایی ساختمان‌ها استفاده می‌شود [۳، ۴، ۵ و ۶]

در سال‌های اخیر، روش‌های مختلفی به منظور شناسایی و تشخیص ساختمان‌ها توسعه یافته است که بر اساس موجود بودن دانش متنی، به سه دسته کلی روش‌های مدل مبنا، داده مبنا و ترکیبی تقسیم می‌شوند [۷ و ۸] در روش‌های مدل مبنا (بالا به پایین)، بر اساس دانش اولیه از شکل ساختمان‌ها در ناحیه مورد نظر، یک کتابخانه شامل مجموعه‌ای از مدل‌های پارامتریک از شکل سقف ساختمان‌ها تهیه و به نقاط استخراج شده مربوط به ساختمان‌ها برازش داده می‌شوند تا بهترین مدل سه بعدی محاسبه شود. یکی از مزیت‌های مهم روش‌های مدل مبنا پایداری آنها است به خصوص هنگامی که نقاط کافی از ساختمان‌ها ناشی از ناکامل بودن داده‌های خام وجود ندارد [۹] علاوه بر این، در این روش‌ها، مدل نهایی دارای توپولوژی صحیحی است [۸] با این حال، از آنجایی که در روش‌های مدل مبنا، کتابخانه مورد استفاده شامل کلیه مدل‌ها از سقف ساختمان‌ها با اشکال متنوع و مختلف نیست، لذا امکان مدل‌سازی کامل بعضی از ساختمان‌ها به ویژه ساختمان‌های پیچیده در این روش‌ها وجود ندارد. در روش‌های داده مبنا (پایین به بالا) نقاط مربوط به سقف ساختمان به کمک اندازه‌گیری خصوصیات ساختمان استخراج می‌شوند و سپس صفحات در برگیرنده سقف و توپولوژی دو بعدی محاسبه و برای مدل‌سازی سه بعدی مورد استفاده قرار می‌گیرند [۱۰] یکی از مزیت‌های مهم روش‌های داده مبنا این است که در این روش‌ها نیازی به دانش اولیه

امروزه بازسازی مدل سه بعدی ساختمان‌ها، یکی از موضوعات چالش‌برانگیز و در دست مطالعه در مراکز تحقیقاتی سنجش از دور و فتوگرامتری است. شناسایی و استخراج اتوماتیک ساختمان‌ها در نواحی شهری، با استفاده از داده‌های سنجش از دور یکی از گام‌های اساسی و پایه در بازسازی مدل سه بعدی است و در کاربردهای مختلف مانند نقشه‌برداری شهری، تولید شهر سه بعدی، به هنگام سازی سیستم اطلاعات مکانی و پایش و ناوبری مورد استفاده قرار می‌گیرد. تاکنون روش‌ها و تکنیک‌های متنوعی برای شناسایی و استخراج عوارض مختلف در تصاویر اپتیکی ارائه شده است که می‌توانند از جنبه‌های مختلف از قبیل داده‌ی مورد استفاده، دانش اولیه در دسترس و الگوریتم مورد استفاده مورد بررسی و تحلیل قرار گیرند.

در کاربردهای سنجش از دور و فتوگرامتری، دو منبع داده‌ای مهم که به منظور استخراج سه بعدی ساختمان‌ها مورد استفاده قرار می‌گیرند، داده‌های اسکنر لیزری و تصاویر اپتیکی هستند. اسکنرهای لیزری هوایی (لایدار)، ابر نقاط سه بعدی و تصاویر شدت عوارض زمینی را اخذ می‌کنند. از طرفی، سنجنده‌های اپتیکی هوایی و یا ماهواره‌ای، داده‌های رادیومتریکی دو بعدی را به صورت درجات خاکستری ثبت می‌کنند. هر یک از این دو تکنولوژی، دارای مزایا و معایب مخصوص به خود هستند [۱ و ۲] به‌عنوان نمونه، سنجنده‌های اپتیکی، تصاویر درجات خاکستری را با قدرت تفکیک و محتوای اطلاعاتی بالا و به‌صورت دو بعدی اخذ می‌کنند، و به منظور استخراج داده‌های سه بعدی از تصاویر دو بعدی، نیاز به حجم داده‌ی بالا، پردازش‌های سه بعدی سریع و دقیق مانند تکنیک‌های تناظریابی اتوماتیک و نیمه اتوماتیک و ویرایش‌های دستی توسط نیروی انسانی است و در نتیجه، استخراج داده‌های سه بعدی از این تصاویر مشکل است [۳]

ساختمان‌ها مورد استفاده قرار گیرند. بدین منظور ابتدا یک بخش‌بندی گراف مینا بر روی داده‌ها اعمال می‌شود و سپس با استفاده از یک الگوریتم طبقه‌بندی قانون مینا و نظارت شده، کلاس‌های مختلف عوارض از جمله ساختمان‌ها، درخت‌ها و جاده‌ها با نرخ شناسایی ۸۸ درصد استخراج می‌شوند [۲۰]. علاوه بر این روش‌ها، مساله شناسایی ساختمان‌ها به صورت مدل مینا می‌تواند با استفاده از مدل رقومی سطح زمین حاصل از تصاویر ماهواره‌ای و تکنیک‌های تناظریابی حل شود. بدین منظور، الگوریتم بخش‌بندی داخل مرز هر یک از ساختمان‌ها انجام شده و شکل‌های هندسی اولیه به صورت چهارگوش‌های منظم تولید می‌شود و سپس بهترین مدل برازش شده برای بازسازی ساختمان‌ها محاسبه می‌شود [۲۱]. یک روش داده مینا بر اساس الگوریتم پارتیشن‌بندی موسوم به گرَب کات^۳ می‌تواند به منظور استخراج اتوماتیک ساختمان‌ها از تصاویر ماهواره‌ای استفاده شود. موقعیت اولیه ساختمان‌ها بر اساس تحلیل سایه‌ها تخمین زده می‌شود و سپس پیکسل‌های مربوط به عوارض ساختمانی از سایر عوارض در یک روند بهینه‌سازی جدا می‌شوند [۲۲]. روش داده مینای دیگر به منظور شناسایی ساختمان‌ها با استفاده از داده‌های لایدار بر مبنای تجزیه چند مقیاسه داده‌ها انجام می‌شود. در این روش، نقاط موجود در صفحه با استفاده از الگوریتمی موسوم به سطح برآزش دهنده محلی (LoFS^۴) استخراج می‌شوند. سپس نواحی با خصوصیات هندسی و سطحی مشابه اما با الگوهای متنی مختلف با استفاده از تحلیل نواحی مسطح زمینی و غیر زمینی از هم جدا می‌شوند. در این روش، پارامترهای کامل بودن، صحت و کیفیت

در مورد ساختار ساختمان‌ها نیست، با این حال در این روش‌ها، تراکم بالای نقاط استخراج شده برای رسیدن به مدل سه بعدی صحیح، ضروری است [۹]. در گروه سوم از روش‌ها، ترکیبی از دو روش مدل مینا و داده مینا به کار گرفته می‌شود تا یک روش بهینه با توجه به مزایا و معایب هر دو گروه توسعه یابد.

بر طبق مطالعات انجام یافته در این تحقیق، روش‌ها و الگوریتم‌های توسعه یافته به منظور شناسایی و استخراج ساختمان‌ها می‌تواند به چهار گروه تقسیم شود که عبارتند از: روش‌های مبتنی بر برازش صفحه [۱۱]؛ روش‌های مبتنی بر فیلترسازی و حدآستانه‌گذاری [۵ و ۱۲] از قبیل روش‌های مورفولوژی [۱۳]؛ روش‌های مبتنی بر بخش‌بندی از قبیل روش‌های پارتیشن‌بندی باینری [۸]، روش‌های بخش‌بندی مبتنی بر استخراج سایه [۱۴] و [۱۵]، و روش‌های بخش‌بندی مبتنی بر رشد ناحیه‌ای [۱۶ و ۱۷]؛ و در نهایت روش‌های طبقه‌بندی نظارت شده [۵، ۱۸، ۱۹ و ۲۰].

یکی از روش‌های استخراج ساختمان‌ها، به کارگیری یک روش مدل مینا با ترکیب داده‌های لایدار و تصاویر هوایی می‌باشد. بدین منظور، یک کتابخانه از اشکال هندسی اولیه از ساختمان‌ها ایجاد می‌شود که شامل جهت شیب ساختمان‌ها است. سپس یک گراف به صورتی تولید می‌شود که نشان‌دهنده‌ی جهت شیب ساختمان و خصوصیات مربوط به شکل و نظم هندسی ساختمان‌ها باشد. با استفاده از این گراف، می‌توان ساختمان‌ها را با نرخ موفقیت ۹۳ درصد استخراج نمود [۶]. به جای استفاده از داده‌های لایدار، اطلاعات ارتفاعی عوارض از مدل‌های رقومی ارتفاعی زمین (DTM^۱) و سطح زمین (DSM^۲) می‌تواند استخراج شوند و برای استخراج مدل مینای

^۳ grab cut^۴ Local Fitting Surfaces^۱ Digital Terrain Model^۲ Digital Surface Model

بر روی داده‌های پاسکال^۶ نسخه ۲۰۰۷ و ۲۰۱۲ نشان دهنده‌ی بهبود قابل توجه در عملکرد شناسایی عوارض است [۴۵] روش تشخیص سه بعدی عوارض بر مبنای CNN با استفاده از داده‌های رنگی و عمق^۷ توسط [۲۹] در سال ۲۰۱۴ ارائه شد. در این روش یک شبکه عصبی کانولوشن برای هر باند تصویری (باندهای قرمز، سبز، آبی و عمق) به صورت جداگانه آموزش دیده است به صورتی که مقادیر اولیه برای بردار وزن در شبکه مربوط به هر باند تصویری با استفاده از مقادیر بردار وزن آموزش دیده در شبکه‌های باندهای تصویری دیگر محاسبه می‌شود. به عبارت دیگر، دانش موجود در باندهای مختلف داده‌های رنگی و عمق بین شبکه‌های عصبی کانولوشن مختلف منتقل می‌شود [۲۹] یک روش طبقه‌بندی نظارت شده بر مبنای CNN توسط [۱۸] در سال ۲۰۱۵ به منظور شناسایی ساختمان‌ها از تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا ارائه شده است. در این روش، بردار ویژگی‌های استخراج شده توسط CNN برای آموزش یک طبقه‌کننده باینری از نوع ماشین بردار پشتیبان استفاده می‌شود تا عوارض ساختمانی را از سایر عوارض جدا کند. سپس برچسب نهایی برای هر کلاس از عوارض با استفاده از یک مدل میدان تصادفی مارکوف (MRF)^۸ تعیین می‌شود [۱۸] به منظور استخراج ساختمان‌ها از تصاویر هوایی، با ترکیب ماتریس‌های ویژگی چند مرحله‌ای، یک شبکه CNN توسط [۳۰] طراحی شد. همچنین از تابع فاصله جهت‌دار^۹ به منظور نمایش مرز ساختمان‌ها استفاده کردند. سپس مدل CNN با مقادیر اولیه تصادفی برای ۲۰۰۰ تصویر شامل ۷۴۰۰ ساختمان آموزش دید. زمان و حجم محاسباتی روش

شناسایی ساختمان‌ها به ترتیب حدود ۹۴،۹۴/۳ و ۸۸/۹ درصد به دست آمده است [۱۱].

یکی از جدیدترین روش‌های طبقه‌بندی نظارت شده بر مبنای شبکه‌های عصبی کانولوشن (CNNs^۱) است که به طور وسیعی در کاربردهای بینایی کامپیوتر و یادگیری ماشین از قبیل طبقه‌بندی پیکسل مینا و تشخیص و شناسایی اتوماتیک عوارض به کار گرفته می‌شود [۲۳، ۲۴، ۲۵ و ۲۶] یک الگوریتم شناسایی مقیاس‌پذیر موسوم به شبکه عصبی کانولوشن ناحیه مینا (R-CNN^۲) در سال ۲۰۱۵ توسط [۲۸] توسعه یافته است که به منظور شناسایی اتوماتیک عوارض در تصاویر دو بعدی در زمینه بینایی کامپیوتر مورد استفاده قرار گرفته است. در این روش، از الگوریتم جستجوی انتخابی^۳ [۲۷] به منظور تعیین موقعیت عوارض و یافتن نواحی کاندید استفاده شده است. با استفاده از CNN، یک بردار ویژگی با طول ثابت برای هر یک از این نواحی استخراج می‌شود. سپس این بردارهای ویژگی با یک مدل از قبل آموزش دیده شده^۴ و بهبود یافته^۵ مقایسه شده و به هر ناحیه به ازای کلاس‌های مختلف یک امتیاز داده می‌شود و بر اساس این امتیازها، بهترین کلاس مربوط به هر ناحیه تعیین می‌شود [۲۸] در روش دیگر، دو الگوریتم جستجو به منظور تعیین موقعیت عوارض در تصاویر با دقت بالا توسط [۴۵] ارائه شده است که یکی بر مبنای تئوری بهینه‌سازی بی‌زین و دیگری بر مبنای معماری یادگیری عمیق است. یادگیری عمیق شامل یک تابع هدف مبتنی بر ماشین بردار پشتیبان و یک طبقه‌کننده مبتنی بر CNN است. نتایج این روش

^۶ PASCAL VOC

^۷ Depth

^۸ Markov Random Field

^۹ signed distance function

^۱ Convolutional Neural Networks

^۲ Region based Convolutional Neural Networks

^۳ Selective Search

^۴ Pre-trained model

^۵ Fine-tuned

غیر ساختمانی مقداردهی می‌شود و سپس با به‌کارگیری الگوریتم انتشار رو به عقب، کل مدل مجدداً آموزش می‌بیند. سپس مدل آموزش دیده شده، مجدداً به‌عنوان مقدار اولیه برای آموزش یک مدل CNN به ازای داده‌های ارتفاعی مورد استفاده قرار می‌گیرد. به‌منظور بررسی عملکرد مدل CNN آموزش دیده شده، از دو مجموعه داده تست استفاده و روند استخراج ساختمان‌ها و شناسایی مدل سقف آنها به‌صورت همزمان طی روند طبقه‌بندی با استفاده از مدل آموزش دیده شده اجرا می‌شود.

۲- شبکه‌های عصبی کانولوشن

خلاصه‌ای از مفاهیم پایه یادگیری عمیق و شبکه‌های کانولوشن در این بخش ارائه می‌شود. یادگیری عمیق (DL) شاخه‌ی جدیدی از علم یادگیری ماشین و شبکه‌های عصبی مصنوعی است که دارای مفهوم درک و فهم چند لایه‌ای و ساختار یادگیری سلسله‌مراتبی است [۲۳، ۲۴ و ۲۵] این نوع یادگیری مبتنی بر آموزش سطوح مختلف نمایش یک عارضه است و یک راه حل کارا تر و موثرتر از روش‌های قدیمی ارائه می‌دهد. در حال حاضر، یادگیری عمیق در کاربردهای مختلفی نظیر شناسایی عوارض [۲۸]، طبقه‌بندی تصاویر [۳۱]، برچسب گذاری معنایی [۳۲]، تشخیص صدا [۳۳] و پردازش زبان طبیعی (NLP^۲) [۳۴] مورد استفاده قرار می‌گیرد. شبکه‌های عصبی کانولوشن نوعی از یادگیری عمیق هستند که شامل تعداد زیادی لایه‌های کانولوشن و نمونه‌برداری^۳ در یک ساختار منطبق بر سیستم بصری انسان می‌باشند [۳۵] لایه ورودی CNN معمولاً یک ماتریس تصویری با ابعاد دلخواه و خروجی آن، یک بردار ویژگی با تفکیک پذیری بالا و متناظر با کلاس‌های مختلف عوارض است. لایه‌های پنهان

مورد استفاده به دلیل استفاده از مقادیر اولیه تصادفی بالا است [۳۰].

به علت تغییرات زیاد در ظاهر و حتی موقعیت ساختمان‌ها در نواحی شهری، توسعه یک روش اتوماتیک برای شناسایی و تشخیص انواع مختلف ساختمان‌ها با دقت و صحت بالا دارای اهمیت است. تا کنون، مطالعات زیادی بر روی شناسایی و تشخیص سه بعدی ساختمان‌ها بر مبنای شبکه‌های عصبی کانولوشن و یا سایر الگوریتم‌های یادگیری عمیق و با استفاده از تصاویر سنجش از دور ارائه نشده است. در این تحقیق، استخراج ساختمان‌ها و تشخیص مدل سه بعدی سقف آنها از قبیل سقف مسطح، شیروانی، شیبدار و هرمی با استفاده از DSM حاصل از داده‌های لایدار و تصاویر ارتوی هوایی بررسی شده است. با توجه به اینکه طبق تحقیقات انجام شده قبلی، تاکنون از الگوریتم CNN به‌منظور شناسایی ساختمان‌ها و تشخیص نوع مدل سقف آنها استفاده نشده است، در این تحقیق، پتانسیل این الگوریتم در حوزه مذکور مورد بررسی و ارزیابی قرار می‌گیرد. همچنین در زمینه یادگیری عمیق، به جای استفاده از یک مجموعه داده‌ی آموزشی خیلی بزرگ، از یک مجموعه داده‌ی آموزشی کوچک و شیوه انتقال دانش بین مدل‌های CNN به منظور آموزش مدل مربوط به مساله‌ی شناسایی و تشخیص ساختمان‌ها بهره گرفته می‌شود و پارامترهای اولیه در مدل CNN در ۱۹ لایه پنهان، با به‌کارگیری یک کتابخانه‌ی کوچک حاوی کلاس‌های مختلف از عوارض ساختمانی و غیر ساختمانی که به‌صورت دستی تولید شده است، تنظیم و بهینه می‌شوند. بدین منظور، یک مدل CNN که از قبل و به ازای مجموعه داده‌ی بسیار بزرگی آموزش دیده شده است، به عنوان مقدار اولیه برای آموزش یک مدل CNN به ازای داده‌های اپتیکی در نظر گرفته می‌شود. در روند آموزش، لایه پنهان آخر در مدل CNN با مقادیر تصادفی و متناسب با کلاس‌های عوارض ساختمانی و

¹ Deep Learning

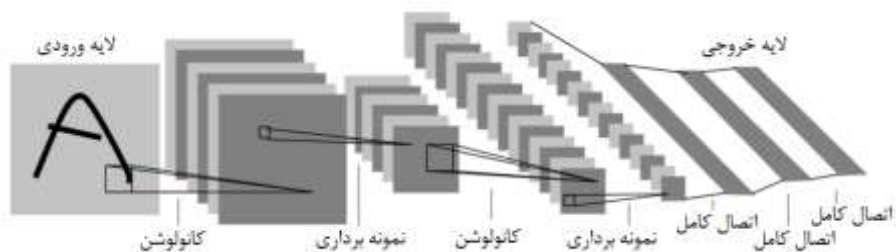
² Natural Language Processing

³ pooling

متصل است [۳۶].

به‌طور کلی، دو روش معمول به منظور آموزش یک مدل CNN وجود دارد که عبارتند از آموزش از ابتدا و با مقادیر اولیه و تصادفی برای بردار وزن‌ها، و بهبود یک مدل از قبل آموزش دیده شده [۲۸ و ۳۸] در روش اول، مجموعه داده‌ی بسیار بزرگی در کلاس‌های مختلف از عوارض نیاز است تا میلیون‌ها پارامتر موجود در مدل CNN آموزش ببینند و چنین روشی معمولاً به حافظه زیاد، زمان پردازشی بالا و امکانات سخت‌افزاری پیشرفته نیاز دارد تا بتواند همگرا شود و به دقت موردنظر برسد. از آنجایی که در اغلب کاربردها، داده‌های آموزشی کم و ناکافی موجود است، مدل‌های CNN نمی‌توانند از ابتدا و بر اساس مقادیر اولیه تصادفی آموزش ببینند. در چنین شرایطی، یک مدل از قبل آموزش دیده شده و موجود به‌عنوان نقطه ابتدایی برای آموزش یک مجموعه داده کوچک و با تعداد کلاس‌های دلخواه استفاده می‌شود.

در این شبکه شامل لایه کانولوشن، لایه نمونه‌برداری و لایه اتصال کامل است [۲۳ و ۳۶] ساختار معمول CNN در شکل (۱) نمایش داده شده است. لایه کانولوشن شامل وزن‌ها و بایاس‌های قابل آموزشی است که به صورت فیلتر با ابعاد و اعماق مختلف بر روی پنجره‌های نمونه‌ای از لایه ورودی اعمال می‌شوند و یک نقشه ویژگی به ازای هر نمونه و هر فیلتر تولید می‌شود. با اتصال این نقشه‌های ویژگی به یکدیگر، یک لایه کانولوشن تشکیل می‌شود [۳۶] لایه نمونه‌برداری یک تابع نمونه‌برداری غیر خطی به همراه کاهش مقیاس است که می‌تواند توابعی مانند ماکزیمم‌گیری، میانگین‌گیری و حتی نرم‌کمترین مربعات باشد. اعمال این لایه بر روی لایه ورودی، باعث می‌شود که ابعاد لایه ورودی به تدریج کاهش یابد [۳۶]. لایه اتصال کامل، لایه نهایی و دارای ویژگی‌های سطح بالا است و هر نرون در این لایه به یکی از نقشه‌های ویژگی در لایه قبلی



شکل ۱: ساختار یک CNN موسوم به LeNet-5 [۲۷]

این تحقیق، تصاویر دو بعدی با اطلاعات ارتفاعی ساختمان‌ها در یک ساختار CNN ترکیب شده تا ساختمان‌ها در یک منطقه شهری به صورت اتوماتیک و پایدار شناسایی شوند. بدین منظور، یک مدل اولیه و موجود که به ازای مجموعه داده‌ی آموزشی بزرگی تولید شده است، در نظر گرفته می‌شود. این مدل، به‌عنوان مقدار اولیه برای آموزش یک مدل CNN به ازای داده‌های اپتیکی به کار گرفته می‌شود، به عبارت دیگر، پارامترهای وزن و بایاس موجود در این مدل به ازای داده‌های اپتیکی مجدداً

در این روش از تابع کاهش گرادیان (SGD^1) [۲۸] به‌عنوان الگوریتم انتشار رو به عقب استفاده می‌شود تا مقادیر بردارهای وزن و بایاس متناسب با مجموعه آموزشی موردنظر آموزش یابند.

۳- روش پیشنهادی

تشخیص الگوی سقف ساختمان‌ها، مساله مهمی در تولید مدل سه بعدی و دقیق از ساختمان‌ها است. در

¹ Stochastic Gradient Descent

الف) آماده سازی داده‌ها

ب) آموزش داده‌ها

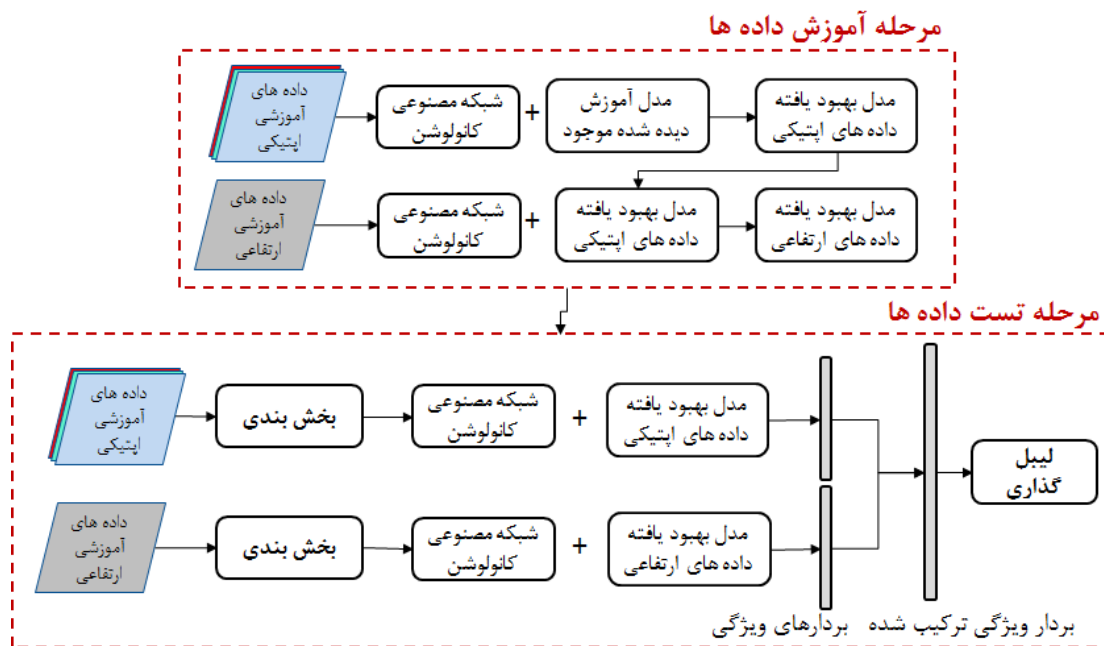
ج) تست داده‌ها

جزئیات هر مرحله در ادامه ارائه می‌شود.

۳-۱- آماده سازی داده‌ها

به‌منظور شناسایی و طبقه‌بندی ساختمان‌های مختلف در یک ناحیه شهری، یک کتابخانه از انواع مختلف سقف‌ها بر مبنای داده واقعیتهای زمینی و به‌صورت دستی تولید می‌شود به طوری که شامل کلاس‌های مختلف از سقف ساختمان‌ها به ازای دو داده DSM و تصاویر اپتیکی باشد. هر کلاس شامل تعداد زیادی تصویر است که در مرکز هر تصویر، نمای بالا از سقف ساختمان‌ها در ابعاد و زوایای دوران مختلف دیده می‌شود.

محاسبه می‌شوند و سپس از این مدل، به عنوان مقدار اولیه برای آموزش یک مدل CNN به ازای داده‌های ارتفاعی استفاده می‌شود. پس از اطمینان از دقت مدل آموزش دیده شده، از آن برای تعیین نوع کلاس ساختمان و نوع کلاس سقف آن به طور همزمان استفاده می‌شود. بدین منظور، ابتدا داده‌ی مورد نظر، بخش‌بندی می‌شود و قطعات تصویری حاوی عوارض مختلف از جمله ساختمان و درختان استخراج می‌شود. سپس با بکارگیری مدل آموزش دیده شده در مرحله قبل، بردار ویژگی برای هر یک از این قطعات محاسبه شده و بر اساس یک روند امتیازدهی، کلاس مربوط به ساختمان و نوع سقف آن به طور همزمان تعیین می‌شود. طبق شکل (۲)، مراحل اصلی این روش عبارتند از:



شکل ۲: مراحل اصلی در شناسایی و تشخیص مدل سقف ساختمان‌ها

کلاس غیر ساختمان شامل عوارض غیر سقف مانند ماشین و درختان و غیره نیز در نظر گرفته می‌شود. کلیه تصاویر نیاز است قبل از ورود به CNN پردازش شوند. سه نوع پردازش معمول عبارتند از: تغییر ابعاد تصویر، کم کردن تصویر میانگین، و نرمال کردن مقادیر [۲۸]، سپس کتابخانه نهایی به عنوان مجموعه داده

علاوه بر کلاس‌های مربوط به سقف‌های مسطح، شیبدار، شیروانی، هرمی، مانسارد^۱، دو کلاس به نام کلاس سقف پیچیده شامل سقف با شکل پیچیده و

^۱ mansard

بهبود می‌یابند. سپس مقادیر پارامترهای وزن و بایاس موجود در مدل بهبود یافته، به عنوان مقادیر اولیه برای آموزش مدل CNN به ازای داده‌های DSM استفاده می‌شود. به عبارت دیگر، دانش از شبکه مربوط به داده‌های اپتیکی به شبکه مربوط به داده‌های ارتفاعی منتقل می‌شود. سپس دو شبکه CNN حاصل برای تخمین کلاس عوارض شناسایی شده در مرحله بعد مورد استفاده قرار می‌گیرند.

۳-۳- تست داده‌ها

پس از آموزش یک شبکه عمیق با استفاده از کتابخانه سقوف ساختمان‌ها، یک داده تست به‌منظور ارزیابی عملکرد مدل آموزش دیده انتخاب می‌شود. بدین منظور سه مرحله اجرا می‌شود که عبارتند از:

الف) بخش‌بندی^۵ داده‌های تصویری

ب) استخراج ویژگی‌های CNN

ج) برچسب‌گذاری عوارض استخراج شده

جزئیات هر مرحله در ادامه ارائه می‌شود.

۳-۳-۱- بخش‌بندی داده‌های تصویری

بر خلاف طبقه‌بندی تصویر، به‌منظور شناسایی عوارض، نیاز به تعیین موقعیت آنها در یک تصویر است [۲۸] با استفاده از مدل DSM در ناحیه شهری، محدوده سقف‌ها می‌توانند بر مبنای حد‌آستانه‌گذاری و تولید ماسک‌های باینری برای مقادیر ارتفاعی مختلف استخراج شوند. سپس داده‌های تست شامل تصاویر اپتیکی و مدل DSM با استفاده از محدوده‌های استخراج شده، بُرش خورده و مجموعه‌ای از تصاویر کاندید برای CNN آماده می‌شوند. به‌منظور بهبود تصاویر کاندید و حذف عوارض غیر سقف، از تکنیک بازسازی تصویر بر اساس روش ارائه شده توسط عارفی و سایرین (۲۰۱۳) استفاده می‌شود [۴۴] همچنین مرحله پیش پردازش داده‌ها شامل تغییر ابعاد

آموزشی برای آموزش مدل CNN مورد استفاده قرار می‌گیرد.

۳-۲- آموزش داده‌ها

یک نمایش عمیق از CNN با استفاده از یک مجموعه داده بزرگ (برای مثال چندین میلیون تصویر) و با استفاده از برنامه‌های مختلف نظیر کافه^۱ [۳۹]، تیانو^۲ [۴۰]، و یا شبکه کانولوشن مبتنی بر متلب (MatConvNet^۳) [۴۱] در طی چندین هفته و با استفاده از پردازشگرهای قوی مانند واحد پردازش گرافیکی (GPU^۴) آموزش می‌بیند. به چنین شبکه‌ای، مدل از قبل آموزش دیده گفته می‌شود که انواع آنها به صورت آنلاین در اینترنت و در کتابخانه کافه موجود هستند [۳۹] از طرفی، آموزش یک شبکه عمیق برای یک کاربرد خاص با داده‌های آموزش ناکافی و کلاس‌های مختلف از عوارض، تنها با استفاده از مقادیر اولیه تصادفی دارای توجیه نیست و نمی‌تواند دقت مورد نظر را حاصل نماید. بنابراین یک مدل از قبل آموزش دیده شده که توسط یک مجموعه داده بزرگ آموزش دیده است، می‌تواند به عنوان یک توصیف‌گر تصویری قوی برای سایر مجموعه داده‌های کوچکتر استفاده شود [۴۲] در این روش، آخرین لایه اتصال کامل در مدلی که از قبل آموزش دیده شده است، حذف شده و یک لایه با بردار وزن تصادفی و بر اساس کلاس‌های عوارض در کاربرد مورد نظر جایگزین آن می‌شود. سپس وزن‌های موجود در این شبکه با استفاده از الگوریتم انتشار رو به عقب مبتنی بر تابع کاهش گرادینان به‌هنگام شده و بهبود می‌یابند. در این تحقیق، یک مدل از قبل آموزش دیده شده برای آموزش مجدد شبکه CNN استفاده می‌شود و مقادیر پارامترهای موجود در این مدل، به ازای داده‌های اپتیکی محاسبه و

¹ Caffe

² Theano

³ MATLAB based Convolutional Network

⁴ Graphics Processing Unit

⁵ Segmentation

تصاویر هوایی و داده‌های لایدار برای شهر اشتوتگارت در آلمان به مساحت ۱۲ کیلومترمربع است (شکل ۳)، تصویر الف). یک مدل DSM با استفاده از داده‌های بازگشت اول لایدار و با قدرت تفکیک یک متر تولید و سپس از آن برای تولید تصاویر ارتوی حقیقی با قدرت تفکیک ۲۰ سانتیمتر استفاده شده است. علاوه بر این، بر اساس تنوع ساختمان‌ها در منطقه مورد مطالعاتی، شش کلاس از انواع سقف مسطح، شیبدار، شیروانی، هرمی، مانسارد و پیچیده برای ایجاد کتابخانه سقف در نظر گرفته می‌شود. همچنین یک کلاس اضافی به نام کلاس "غیر سقف" برای طبقه‌بندی عوارضی که سقف نیستند، در نظر گرفته می‌شود.

مجموعه داده آموزشی شامل کل منطقه نشان داده شده در شکل ۳)، تصویر الف است که مناطق در نظر گرفته شده برای تست مدل (شکل ۳)، تصویر ب و ج) در آن وجود ندارد. این مجموعه داده آموزشی به منظور ایجاد یک کتابخانه شامل هفت کلاس استفاده می‌شود. در این تحقیق، این کتابخانه به صورت دستی تولید می‌شود. به ازای هر کلاس در کتابخانه، حدود ۱۰۰ تصویر از منظره بالای سقف و به صورت تصادفی از مجموعه داده‌ی آموزشی انتخاب می‌شود. در نتیجه، مجموعه داده آموزشی در مجموع دارای ۷۰۰ داده‌ی تصویری است که ۷۰ درصد آن برای آموزش و ۳۰ درصد آن به منظور اعتبار سنجی مدل آموزش دیده مورد استفاده قرار می‌گیرد. همچنین دو تصویر در ناحیه مورد نظر و خارج از مجموعه داده‌ی آموزشی برای تست مدل در نظر گرفته می‌شود (شکل ۳)، تصاویر ب و ج). در نتیجه، مجموعه داده‌ی آموزشی و تست دارای سه باند رنگی قرمز، سبز و آبی به ازای داده‌های ایتیکی و سه باند تکراری شامل مدل DSM به ازای داده‌های ارتفاعی می‌باشند.

تصاویر، کم کردن تصویر میانگین و نرمال کردن بر روی هر یک از تصاویر اجرا شده و تصاویر مناسب برای ورود به CNN تهیه می‌شوند.

۳-۳-۲- استخراج ویژگی‌های CNN

با استفاده از مدل CNN بهبود یافته برای تصاویر ایتیکی و مدل DSM، بردار ویژگی با ابعاد ۱ در ۷ (۷ تعداد کلاس‌های موجود در این تحقیق است) برای هر یک از تصاویر کاندید و به ازای هر مدل استخراج می‌شود. بدین منظور از ابزار MatConvNet [۴۱] استفاده می‌شود.

۳-۳-۳- برچسب‌گذاری عوارض استخراج شده

بردار ویژگی‌های استخراج شده شامل امتیاز هر تصویر کاندید به ازای هر کلاس از سقف است. در صورتی که تصویر کاندید متعلق به کلاس خاصی باشد، امتیاز آن کلاس، مقدار حداکثر را خواهد داشت. در این تحقیق دو بردار ویژگی استخراج شده به ازای تصاویر ایتیکی و مدل DSM با یکدیگر ترکیب می‌شوند تا حداکثر امتیاز یک تصویر کاندید محاسبه و برچسب بهینه مربوط به هر تصویر پیدا شود.

۴- پیاده‌سازی و ارزیابی نتایج

در این تحقیق، یک روش دانش مبنای به منظور شناسایی ساختمان‌ها و تشخیص مدل سه بعدی انواع مختلف سقف ساختمان‌ها در یک منطقه شهری ارائه شده است که بر مبنای یادگیری سلسله مراتبی ترکیبی از ویژگی‌های طیفی و هندسی است. بدین منظور، یک ساختار جدید برای شبکه عصبی کانولوشن ارائه شده است که با استفاده از تصاویر ایتیکی و مدل DSM به یادگیری ویژگی‌ها و سپس تخمین برچسب هر کلاس می‌پردازد.

۴-۱- مورد مطالعاتی

مجموعه داده اصلی و مورد استفاده در این تحقیق،



شکل ۳: الف: منطقه مورد مطالعه، ب: منطقه تست اول، ج: منطقه تست دوم

۴-۲- آموزش مدل

شبکه عصبی پایه و اولیه‌ی مورد استفاده در این تحقیق، شبکه CNN [۴۳] است که بر روی مجموعه داده‌ی بزرگی به نام شبکه‌ی داده‌های تصویری بزرگ مقیاس برای تشخیص بصری (ILSVRC1) [۲۴] و برای ۱۰۰۰ کلاس از عوارض مختلف با بیش از چند میلیون تصویر در حوزه کامپیوتر ویژن، با استفاده از ابزار MatConvNet [۴۱] آموزش دیده شده است. مقادیر پارامترهای موجود در این شبکه، به‌عنوان مقادیر اولیه برای آموزش شبکه CNN به ازای کتابخانه حاوی داده‌های اپتیکی مورد استفاده قرار می‌گیرد. از آنجایی که در ساختار CNN مورد استفاده، داده‌های ورودی باید به ابعاد $۲۲۴ \times ۲۲۴ \times ۳$ باشد، ابعاد کلیه ۷۰۰ تصویر موجود در کتابخانه سقف تغییر کرده و سپس تصویر میانگین از این ۷۰۰ تصویر محاسبه و از هر یک از آنها کم می‌شود. سپس برای سازگاری شبکه CNN با ۷ کلاس مورد نظر، عملیات بهبود مدل ILSVRC برای ۲۰۰ تکرار و با نرخ آموزشی ۰/۰۱ اجرا می‌شود.

مقدار بهینه برای تعداد تکرار و مقدار نرخ آموزشی بر اساس نمودار همگرایی خطا در آزمون‌های مختلف و به‌صورت سعی و خطا تعیین می‌شود. سیستم پردازشی مورد استفاده در این تحقیق، یک پردازشگر سی پی یو با پنج هسته و با هشت گیگابایت حافظه داخلی است.

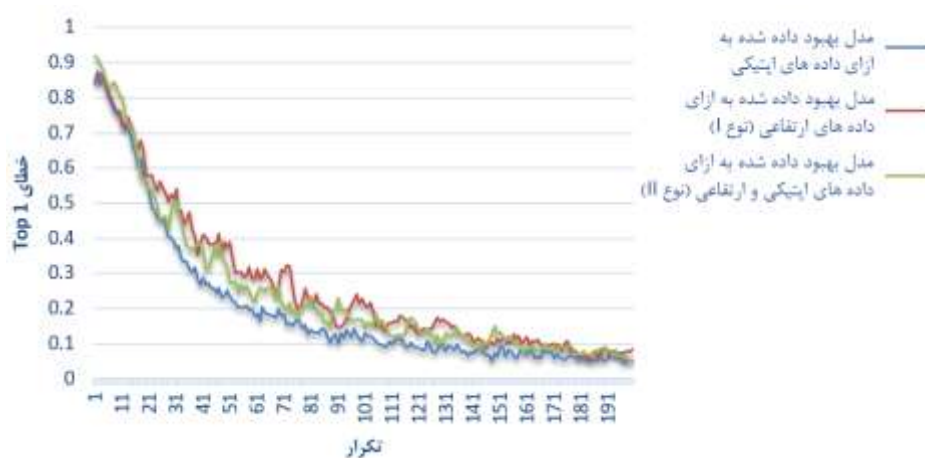
در این تحقیق، داده‌های ارتفاعی یکبار بر اساس مدل ILSVRC (مدل نوع I در جدول (۱)) و یکبار بر اساس مدل بهبود یافته‌ی مربوط به داده‌های اپتیکی (مدل نوع II در جدول (۱)) آموزش دیده‌اند. پارامترهای مقایسه برای سه مدل بهبود یافته به ازای تصاویر اپتیکی و داده‌های ارتفاعی (مدل DSM) از قبیل مدت زمان آموزشی، خطای حد بالای اول^۲ و دقت مدل آموزش دیده شده در تکرار آخر در جدول (۱) ارائه شده است. خطای حد بالای اول به این معنا است که بالاترین امتیاز تا چه اندازه اشتباه تخمین زده شده است [۴۳].

² top1 error

¹ ImageNet Large Scale Visual Recognition Challenge

جدول ۱: پارامترهای مقایسه برای دو مدل بهبود یافته

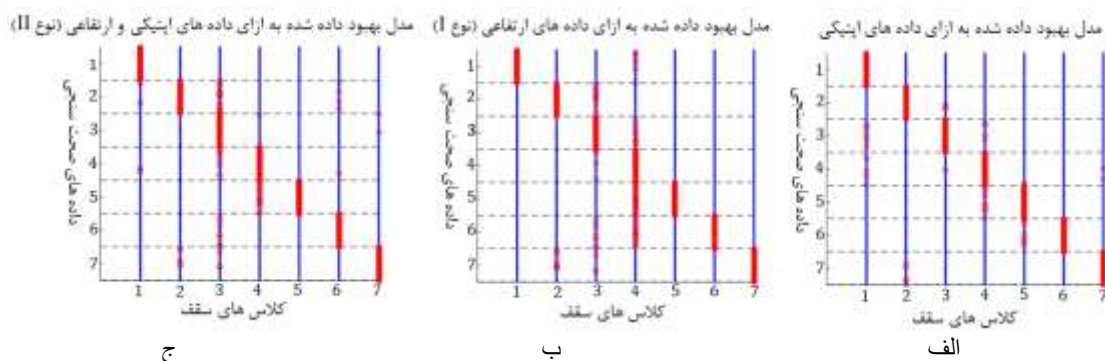
داده ورودی	مدل اولیه	زمان (ساعت)	خطای حد بالای اول	دقت (%)	مدل نهایی
تصاویر اپتیکی	مدل ILSVRC	۴/۳	۰/۰۵۲۹	۹۴/۷	مدل بهبود یافته به ازای تصاویر اپتیکی
داده‌های ارتفاعی	مدل ILSVRC	۸/۲	۰/۰۸۴۳	۹۱/۶	مدل بهبود یافته به ازای داده‌های ارتفاعی (نوع I)
داده‌های ارتفاعی	مدل بهبود یافته به ازای تصاویر اپتیکی	۶/۶	۰/۰۴۸۶	۹۵/۱	مدل بهبود یافته به ازای داده‌های ارتفاعی و اپتیکی (نوع II)



شکل ۴: نرخ همگرایی مدل آموزش دیده شده

حالتی است که از مدل ILSVRC به‌عنوان مقدار اولیه برای داده‌های ارتفاعی استفاده شود. به‌منظور ارزیابی بهتر عملکرد آموزش داده‌ها، حداکثر امتیاز داده‌های مربوط به صحت‌سنجی در هر کلاس با سایر کلاس‌ها و به ازای سه مدل آموزش دیده شده مطابق با شکل (۵) مقایسه شده است. در شکل (۵)، کلاس اول، سقف پیچیده، کلاس دوم، سقف مسطح، کلاس سوم، سقف شیبدار، کلاس چهارم، سقف شیروانی، کلاس پنجم، سقف هرمی، کلاس ششم، سقف مانسارد و کلاس هفتم، عوارض غیر سقف است.

شکل (۴) میزان بهبود دقت آموزش هر مدل را در تکرارهای بالا نشان می‌دهد. مطابق با جدول (۱)، آموزش مجموعه داده‌های ارتفاعی با استفاده از مقدار اولیه مبتنی بر داده‌های اپتیکی (نوع II در شکل (۴)) بهتر از آموزش آنها با مدل ILSVRC (نوع I در شکل (۴)) است به طوری که خطای حد بالای اول و دقت آن به ترتیب $۰/۰۴۸۶$ و $۹۵/۱$ درصد به‌دست آمده است که این اعداد، نشان‌دهنده بهبود دقت مدل آموزش دیده شده در هنگام استفاده از داده‌های اپتیکی و ارتفاعی در کنار یکدیگر است. علاوه بر این، زمان مورد نیاز برای آموزش مدل در این حالت کمتر از



شکل ۵: مقایسه حداکثر امتیازها برای هر مجموعه داده در هر کلاس سقف (۱: پیچیده؛ ۲: مسطح؛ ۳: شیبدار؛ ۴: شیروانی؛ ۵: هرمی؛ ۶: مانسارد؛ ۷: غیر سقف) بر اساس مدل‌های آموزش دیده شده، الف: به ازای داده‌های اپتیکی، ب: به ازای داده‌های ارتفاعی (مدل نوع I)، ج: به ازای داده‌های اپتیکی و ارتفاعی (مدل نوع II)

هر ناحیه کاندید اعمال می‌شود. سپس نتیجه حاصل وارد شبکه CNN شده تا بردارهای ویژگی برای هر ناحیه کاندید استخراج شود. متریک‌های استاندارد کیفیت از قبیل کامل بودن، صحت و کیفیت شناسایی [۱۸] به ازای هر مجموعه داده مطابق با رابطه (۱) محاسبه شده است.

$$\text{Coml.} = \frac{TP}{TP + FN};$$

$$\text{Corr.} = \frac{TP}{TP + FP};$$

$$\text{Qual.} = \frac{TP}{TP + FN + FP}$$

رابطه (۱)

در رابطه (۱)، پارامتر TP نمونه‌ی مثبت و صحیح (بیانگر عوارضی که به درستی شناسایی شده‌اند)، FP نمونه مثبت و اشتباه (بیانگر عوارضی که به اشتباه شناسایی شده‌اند)، و FN نمونه منفی و اشتباه (بیانگر عوارضی که ساختمان بوده‌اند و شناسایی نشده‌اند)، است. در جدول‌های (۲ و ۳) و نمودارهای شکل (۶) مقادیر متریک‌های استاندارد برای دو مجموعه داده تست و برای سه مدلی که با استفاده از فقط داده‌های اپتیکی و یا فقط داده‌های ارتفاعی و همچنین با هر دو داده‌ی اپتیکی و ارتفاعی آموزش دیده‌اند، ارائه شده است، همچنین شکل (۷) مرزهای استخراج شده و برجسب‌های نهایی برای هر سقف شناسایی شده نمایش داده شده است.

بر اساس ارزیابی بصری نمودارهای شکل (۵)، تصاویر مربوط به داده‌های اپتیکی (تصویر الف) تفکیک‌پذیری بیشتری در هر کلاس دارند با این حال همچنان خطاهای طبقه‌بندی در بعضی از کلاس‌ها از جمله کلاس سقف پیچیده و شیروانی مشاهده می‌شود. هنگامی که داده‌های ارتفاعی بر اساس مدل اپتیکی بهبود می‌یابد (تصویر ج)، عملکرد طبقه‌بندی به خصوص در کلاس سقف‌های شیروانی (کلاس چهارم) نسبت به حالتی که از تصاویر اپتیکی در کنار داده‌های ارتفاعی استفاده نشده است (تصویر ب) بهبود می‌یابد. با توجه به نتایج حاصل از این مرحله، در مرحله بعد، مدل‌های CNN مبتنی بر داده‌های اپتیکی، داده‌های ارتفاعی و همچنین داده‌های ارتفاعی که بر مبنای داده‌های اپتیکی آموزش یافته است، برای استخراج ویژگی‌های نواحی کاندید مورد استفاده قرار می‌گیرند.

۴-۳- تست مدل

به‌منظور تست مدل‌های آموزش دیده، دو مجموعه داده خارج از ناحیه‌ی مربوط به داده‌های آموزشی و شامل انواع سقف ساختمانی و سایر عوارض انتخاب می‌شود. مرز هر ناحیه کاندید بر اساس نتایج حاصل از مرحله بخش‌بندی مبتنی بر حدآستانه گذاری تولید می‌شود. داده‌ی تست به ازای هر مرز برش خورده و مشابه با مرحله آموزش مدل، عملگرهای پیش پردازشی بر روی

جدول ۲: ارزیابی کمی نتایج شناسایی برای مجموعه داده تست اول

مجموعه داده اول	مقدار TP	مقدار FP	مقدار FN	درصد کامل بودن شناسایی	درصد صحت شناسایی	درصد کیفیت شناسایی
تست مدل آموزش دیده بر اساس داده‌های اپتیکی	۱۵	۱۶	۲	۸۸/۲	۴۸/۴	۴۵/۵
تست مدل آموزش دیده بر اساس داده‌های ارتفاعی (نوع I)	۱۱	۲۰	۲	۸۴/۶	۳۵/۵	۳۳/۳
تست مدل آموزش دیده بر اساس داده‌های اپتیکی و ارتفاعی (نوع II)	۲۲	۱۰	۱	۹۵/۶	۶۸/۷	۶۶/۷

جدول ۳: ارزیابی کمی نتایج شناسایی برای مجموعه داده تست دوم

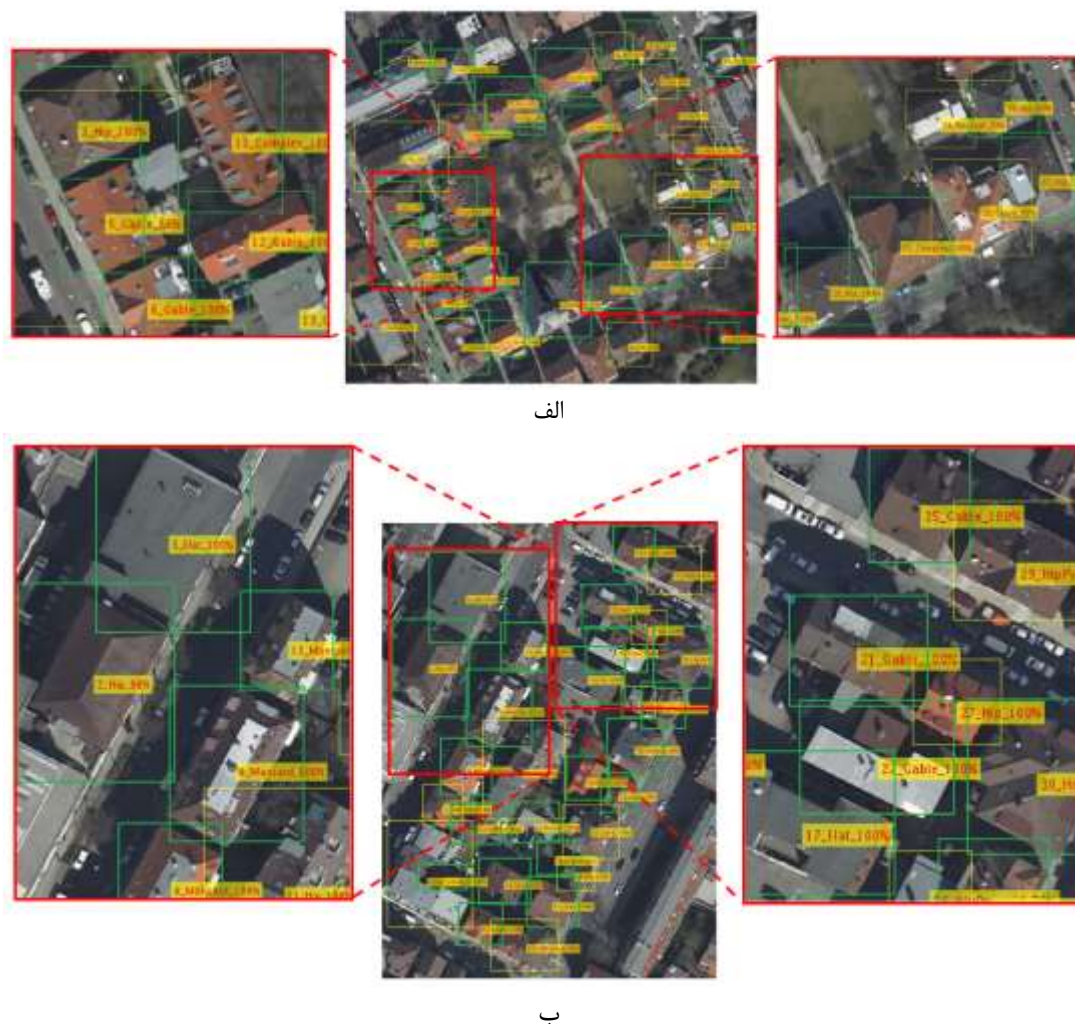
مجموعه داده اول	مقدار TP	مقدار FP	مقدار FN	درصد کامل بودن شناسایی	درصد صحت شناسایی	درصد کیفیت شناسایی
تست مدل آموزش دیده بر اساس داده‌های اپتیکی	۱۵	۱۲	۳	۸۳/۳	۵۵/۶	۵۰
تست مدل آموزش دیده بر اساس داده‌های ارتفاعی (نوع I)	۲	۲۷	۱	۶۶/۷	۷	۷
تست مدل آموزش دیده بر اساس داده‌های اپتیکی و ارتفاعی (نوع II)	۲۱	۹	۰	۱۰۰	۷۰	۷۰



شکل ۶: نمودار میله‌ای نتایج شناسایی برای دو مجموعه داده تست

صحت و کیفیت شناسایی به طور میانگین حدود ۶۷ درصد در مجموعه داده اول و حدود ۷۰ درصد در مجموعه داده دوم است. علاوه بر این، طبق شکل (۶)، اطلاعات ارتفاعی به تنهایی قادر نخواهد بود به‌منظور شناسایی نوع سقف ساختمان‌ها به کار رود به طوری که درصد صحت و کیفیت شناسایی در این حالت خیلی پایین و در مجموعه داده اول حدود ۳۴ درصد و در مجموعه داده دوم حدود ۷ درصد است.

مطابق با ارقام ارائه شده در جدول‌های (۲ و ۳)، هنگامی که از اطلاعات ارتفاعی در کنار اطلاعات اپتیکی برای آموزش مدل استفاده می‌شود، درصد کامل بودن، صحت و کیفیت شناسایی به مراتب بالاتر از حالتی است که فقط از اطلاعات اپتیکی استفاده می‌شود. برای مثال، هنگامی که از اطلاعات ارتفاعی در کنار اطلاعات اپتیکی استفاده می‌شود، درصد کامل بودن شناسایی حدود ۹۵/۶ درصد در مجموعه داده اول و ۱۰۰ درصد در مجموعه داده دوم و



- شکل ۷: نتیجه نهایی شناسایی و تشخیص سقف ساختمان‌ها، الف: برای مجموعه داده اول، ب: برای مجموعه داده دوم
- مطابق با جداول (۲ و ۳)، هنگامی که از اطلاعات ارتفاعی در کنار اطلاعات اپتیکی به منظور آموزش مدل CNN استفاده می‌شود، میزان صحت شناسایی سقف در کلاس‌های مختلف به طور میانگین حدود ۶۹ درصد و کیفیت شناسایی حدود ۶۸ درصد برای هر دو مجموعه داده تست به دست آمده است که می‌توان عواملی از قبیل عوامل زیر را در پایین بودن صحت و کیفیت شناسایی موثر دانست (در شکل ۸)، تاثیر هر یک از این عوامل بر روی تعدادی از نمونه‌هایی که به اشتباه شناسایی شده‌اند، ارائه شده است:
- سایه ساختمان‌ها: نواحی پنهان ناشی از
 - سایه ساختمان‌های بلند بر روی یک سقف، باعث اشتباه شدن نتایج طبقه‌بندی می‌شود، چرا که سایه ساختمان‌های بلند بر روی ساختمان‌های کوتاه‌تر باعث ایجاد لبه‌های مصنوعی در تصاویر اپتیکی شده و ویژگی‌های استخراج شده از سقفی که شامل سایه است، متفاوت با ویژگی‌های آموزش دیده شده در مدل CNN خواهند بود. مطابق شکل (۸)، در تصویر (د)، سایه ساختمان، باعث ایجاد یک لبه مصنوعی بر روی ساختمان کوچکتر شده است و تشخیص مدل آن به اشتباه انجام شده است.
 - جزئیات موجود بر روی سقف:

روش‌های مدل مبنای این است که بعضی از مدل‌های پیچیده در کتابخانه سقوف در نظر گرفته نشده است لذا کل سقوف پیچیده یا طبقه‌بندی نمی‌شود و یا به طور اشتباه طبقه‌بندی می‌شود. از طرفی، ساختمان‌های پیچیده، معمولاً ترکیبی از اشکال ساده‌تر هستند، برای تجزیه چنین سقوفی به اشکال پایه، مرحله بخش‌بندی تصویر باید با دقت بالاتری انجام شود. مطابق شکل (۸)، در تصاویر (ه) و (و)، ساختمان‌های ترکیبی از چند نوع سقوف وجود دارند که نوع پایه سقوف، قابل شناسایی نبوده است و نیاز است که مرحله سگمنت‌بندی تصویر با دقت بیشتری انجام شود.



ج



ب



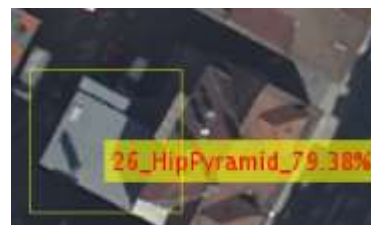
الف



و



ه



د

شکل ۸: تاثیر عوامل مختلف در عدم موفقیت در شناسایی ساختمان‌ها، الف: وجود جزئیات بر روی سقوف، ب و ج: تاثیر تضاريس موجود در لبه ساختمان‌ها، د: تاثیر عامل سایه، ه و و: پیچیده بودن ساختمان

موجود در لبه سقوف، باعث شبیه شدن سقوف Gable به Hip و تشخیص نادرست سقوف شده است.

علاوه بر نتایج فوق، می‌توان نتیجه گرفت که داده‌های ارتفاعی به عنوان مکملی برای داده‌های اپتیک‌ی باعث غنی‌تر شدن دانش مورد نیاز از نوع ساختمان‌ها و دقیق‌تر شدن نتایج شناسایی و تشخیص مدل‌ها می‌شود، اگر چه نتایج حاصل از پیش پردازش داده‌ها و مرحله بخش‌بندی نیز بر روی دقت مورد نظر

ساختارهای کوچک و جزئی بر روی سقوف ساختمان‌ها از قبیل دودکش و پنجره می‌تواند باعث تغییر ویژگی‌های استخراج شده برای سقوف نسبت به حالتی که سقوف بدون ساختارهای جزئی است، بشود و در نتیجه در صورتی که مدل آموزش دیده شده، انعطاف کافی و لازم را نداشته باشد، نتایج شناسایی ساختمان‌ها با ساختارهای کوچک دارای صحت و دقت کافی نخواهند بود. مطابق شکل (۸)، در تصویر (الف) و (ب)، وجود پنجره کوچک بر روی سقوف Gable باعث تشخیص اشتباه نوع سقوف شده است.

• سقوف با اشکال خیلی پیچیده: یکی از معایب

• نویز موجود در داده‌ها: به علت تضاريس موجود در لبه بعضی از ساختمان‌ها در تصویر ارتو که ناشی از مدل DSM است، مدل بعضی از ساختمان‌ها برای مثال از مدل شیب‌دار به مدل شیروانی تغییر می‌کند که این مورد تاثیر قابل توجهی بر روی دقت شناسایی دارد.

مطابق شکل (۸)، در تصویر (ب) و (ج)، تضاريس

موثر هستند.

۵- نتیجه گیری

یکی از مهمترین مسائل موجود در تولید مدل سه بعدی شهری، شناسایی اتوماتیک ساختمان‌ها و تشخیص انواع سقف ساختمانی از داده‌های سنجش از دور است. در این تحقیق، بر اساس ترکیب اطلاعات طیفی (تصاویر اپتیکی) و اطلاعات هندسی (مدل DSM)، یک شبکه مصنوعی کانولوشن طراحی و آموزش داده شد. بدین منظور ابتدا داده‌های اپتیکی به کمک یک مدل از قبل آموزش داده شده موسوم به ILSVRC آموزش یافته و سپس از آن برای آموزش داده‌های ارتفاعی استفاده شده است. سپس دو مدل بهبود یافته به منظور استخراج بردارهای ویژگی نواحی کاندید به کار گرفته شده و در نهایت برچسب بهینه

مراجع

- [1] N. Haala, and C. Brenner, "Extraction of buildings and trees in urban environments", *ISPRS Journal of Photogrammetry & Remote Sensing*, pp. 130-137, 1999.
- [2] C. Kim, and A. Habib, "Object-Based Integration of Photogrammetric and LiDAR Data for Automated Generation of Complex Polyhedral Building Models", *Journal of Sensors*, Vol.9, pp. 5679-5701, 2009.
- [3] T. T. Vu, F. Yamazaki, M. Matsuoka. "Multi-scale solution for building extraction from LiDAR and image data", *International Journal of Applied Earth Observation and Geoinformation*, pp. 281-289, 2009.
- [4] M. Salah, J. Trinder, A. Shaker, "Evaluation of the self-organizing map classifier for building detection from LiDAR data and multispectral aerial images", *Journal of Spatial Science*, Vol.54, Issue 2, pp. 15-34, 2009.
- [5] T. Hermosilla, L. A. Ruiz, J. A. Recio, J. Estornell, "Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data", *Journal of Remote Sensing*, Vol.3, Issue 6, pp. 1188-1210, 2011.
- [6] G. Q. Zhou, and X. Zhou, "Seamless fusion of LiDAR and aerial imagery for building extraction". *IEEE Journal of Transactions on Geoscience and Remote Sensing*, Vol.52, No.11, pp. 7393-7407, 2014.
- [7] E. Schwalbe, H-G. Maas, F. Seidel, "3D building generation from airborne laser scanner data using 2D GIS data and orthogonal point cloud projections", presented at the ISPRS workshop of laser scanning, the Netherlands, 2005.
- [8] A. Wichmann, J. Jung, G. Sohn, M. Kada, M. Ehlers, "Integration of building knowledge into binary space partitioning for the reconstruction of regularized building models", presented at ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, France, 2015.
- [9] G. Sohn, and I. Dowman, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction", *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol.62, pp. 43-63, 2007.

- [10] K. Zhang, Z. Cui, P. A. Houle, "Airborne LiDAR remote sensing and its applications", in *Advances in Mapping from Remote Sensor Imagery, Techniques and Applications*. X. Yang, and J. Li, CRC Press, 2012.
- [11] D. Mongus, N. Lukac, B. Zalik, "Ground and building extraction from LiDAR data based on differential morphological profiles and locally fitted surfaces", *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 145–156, 2014.
- [12] E. Maltezos, and C. Ioannidis, "Automatic detection of building points from LiDAR and dense image matching point clouds", presented at *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, France, 2015.
- [13] B. Yu, H. Liu, J. Wu, Y. Hu, L. Zhang, "Automated derivation of urban building density information using airborne LiDAR data and object-based method", *Landscape and Urban Planning*, Vol.98, Issue 3-4, pp. 210–219, 2010.
- [14] T. Ngo, C. Collet, V., Mazet, "Automatic rectangular building detection from VHR aerial imagery using shadow and image segmentation", presented at *IEEE International Conference on Image Processing*, 2015.
- [15] G. Singh, M. Jouppi, Z. Zhang, A. Zakhor, "Shadow Based Building Extraction from Single Satellite Image", In *Proceedings of SPIE, the International Society for Optical Engineering*, Vol.9401, 2015.
- [16] L. Matikainen, J. Hyypä, H., Kaartinen, "Automatic detection of changes from laser scanner and aerial image data for updating building maps", *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol.35, pp. 434–439, 2004.
- [17] M. Awrangjeb, C. Zhang, C. S. Fraser, "Automatic extraction of building roofs using LIDAR data and multispectral imagery", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol.83, pp. 1–18, 2013.
- [18] M. Vakalopoulou, K. Karantza, N. Komodakis, N. Paragios, "Building detection in very high resolution multispectral data with deep learning features", *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1873 – 1876, 2015.
- [19] L. Guo, N. Chehata, C. Mallet, S. Boukir, "Relevance of airborne LiDAR and multispectral image data for urban scene classification using Random Forests", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol.66, Issue 1, pp. 56–66, 2011.
- [20] K. Karantza, P. Koutsourakis, I. Kalisperakis, L. Grammatikopoulos, "Model based building detection from low-cost optical sensors on-board unmanned aerial vehicles", presented at the *International Conference on Unmanned Aerial Vehicles in Geomatics*, Toronto, Canada, 2015.
- [21] T. Partovi, T. Krauß, H. Arefi, M. Omidalizandi, P. Reinartz, "Model-driven 3D building reconstruction based on integration of DSM and spectral information of satellite images", presented at the *Geoscience and Remote Sensing Symposium (IGARSS)*, Quebec, Canada, pp. 3168-3171, 2014.
- [22] M. Khurana, and V., Wadhwa, "Automatic building detection using modified grab cut algorithm from high resolution satellite image", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol.4, Issue 8, 2015.
- [23] T. Liu, S. Fang, Y. Zhao, P. Wang, J. Zhang, "Implementation of training convolutional neural networks", *Computer Vision and Pattern Recognition*, 2015.
- [24] J. Deng, W. Dong, R., Socher, L.-J. Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database", In *Proc. CVPR*, 2009.
- [25] Y. Bengio, "Learning Deep Architectures for AI", *Foundations and Trends in Machine*

- Learning, Vol.2, Issue 1, pp. 1-127, 2009.
- [26] M. D. Zeiler, and R. Fergus, "Visualizing and understanding convolutional networks", in Computer Vision, ECCV 2014. Vol.8689 of the series Lecture Notes in Computer Science, pp. 818-833, 2014.
- [27] J. Uijlings, K. van de Sande, T. Gevers, A. Smeulders, "Selective search for object recognition", International Journal of Computer Vision, Vol.104, Issue 2, pp. 154-171, 2013.
- [28] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Region-based convolutional networks for accurate object detection and semantic segmentation", the IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
- [29] L. A. Alexandre, "3D object recognition using convolutional neural networks with transfer learning between input channels", presented at the 13th International Conference on Intelligent Autonomous Systems, Italy, 2014.
- [30] J. Yuan, "Automatic Building Extraction in Aerial Scenes Using Convolutional Networks", 2016.
- [31] Q. Zou, L. Ni, T. Zhang, Q. Wang, "Deep learning based feature selection for remote sensing scene classification", IEEE Geoscience and Remote Sensing Letters, Vol.12, No.11, 2015.
- [32] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation", In Proc. CVPR, 2015.
- [33] G. E. Dahl, D. Yu, L. Deng, A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition", IEEE Transactions on Audio, Speech, and Language Processing, Vol.20, Issue 1, pp. 30-42, 2012.
- [34] R. Socher, Y. Bengio, C., Manning, "Deep learning for NLP. Tutorial at Association of Computational Logistics (ACL)", 2013, <http://www.socher.org/index.php/DeepLearningTutorial>.
- [35] S. Filipe, and L. A. Alexandre, "From the human visual system to the computational models of visual attention: A survey. Artificial Intelligence Review", IEEE Transactions On Geoscience And Remote Sensing, Vol.52, No.11, p.p. 1-47, 2014.
- [36] S. L. Phung, and A. Bouzerdoum, "MATLAB Library for Convolutional Neural Networks, Technical Report", Visual and Audio Signal Processing Lab, University of Wollongong, 2009.
- [37] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, Vol.86, No.11, 1998.
- [38] D. S. Maitra, U. Bhattacharya, S. K. Parui, "CNN based common approach to handwritten character recognition of multiple scripts", presented at the 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, 2015.
- [39] Y. Jia, "Caffe: An open source convolutional architecture for fast feature embedding", 2013, <http://caffe.berkeleyvision.org/>.
- [40] F. Bastien, P. Lamblin, R. Pascanu, J. I. Bergstra, J. Goodfellow, A. Bergeron, N. Bouchard, Y. Bengio, "Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning", presented at the NIPS Workshop, 2012.
- [41] A. Vedaldi, and K. Lenc, "MatConvNet: Convolutional Neural Networks for MATLAB", proceeding of the ACM International Conference on Multimedia, 2015, <http://www.vlfeat.org/matconvnet/>
- [42] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets", In Proc. BMVC, 2014.
- [43] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks", NIPS, pp. 1106-1114, 2012.

- [44] H. Arefi, and P. Reinartz, "Building reconstruction using DSM and orthorectified images", *Remote Sensing*, pp. 1681-1703, 2013.
- [45] Y. Zhang, K. Sohn, R. Villegas, G. Pan, H. Lee, "Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction", presented at IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015.



Synergistic Use of LiDAR Data and Aerial Image based on Convolutional Neural Networks for Building Model Recognition

Fateme Alidoost ¹, Hossein Arefi ^{*2}

1- PhD student of photogrammetry in School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran
 2- Assistant professor in School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran

Abstract

Buildings are one of the most important urban structures that are used for various applications and urban mapping. In recent years, with the development of the high resolution data acquisition technologies, many different approaches and algorithms have been presented to extract the accurate and timely updated 3D models of buildings as a key element of city structures for numerous applications in urban mapping. In this paper, a novel and model-based approach is proposed for automatic recognition of buildings' roof models such as flat, gable, hip, and pyramid hip roof models based on deep structures for hierarchical learning of features that are extracted from both LiDAR and aerial orthophotos. The main steps of this approach include building segmentation, feature extraction and learning, and finally building roof labelling in a supervised pre-trained Convolutional Neural Network (CNN) framework to have an automatic recognition system for various types of buildings over an urban area. In this framework, the height information provides invariant geometric features for CNN network to localize the boundary of each individual roofs. CNN is a kind of feed-forward neural network with the multilayer perceptron concept that consists of a number of convolutional and sub-sampling layers in an adaptable structure which is widely used in pattern recognition and object detection applications. Since the training dataset is a small library of labelled models for different shapes of roofs, via using the pre-trained models, the computation time of learning can decrease significantly. The experimental results highlight the effectiveness of the deep learning approach to detect and extract the pattern of buildings' roofs automatically considering the complementary nature of height and RGB information. Based on the training results, the top 1 error and accuracy of training are about 0.05 and 95 %, respectively. Moreover, the average of correctness and completeness rates are about 97 % and 69 %, respectively.

Key words: Pattern Recognition, Deep Learning, 3D Modelling, Convolutional Neural Network, LiDAR.

Correspondence Address : Photogrammetry Group, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
Tel : +98 21 61114523.
Email : hossein.arefi@ut.ac.ir