

ارائه یک الگوریتم ترکیبی مبتنی بر یادگیری عمیق و ماشین بولتزمان محدود با هدف قطعه‌بندی معنایی خودرو از تصاویر مادون‌قرمز حرارتی پهنپایه

مهدی خوش برش ماسوله^۱، رضا شاه حسینی^{۲*}

۱- دانشجوی دکتری، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

۲- استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران

تاریخ دریافت مقاله: ۱۳۹۷/۰۹/۲۴ تاریخ پذیرش مقاله: ۱۳۹۸/۰۳/۲۵

چکیده

امروزه پایش وسایل نقلیه زمینی با استفاده از روش‌های پردازش تصویر، یکی از حیطه‌های کاربردی در کنترل ترافیک هوشمند به شمار می‌آید. در این زمینه، به‌کارگیری تصاویر مادون‌قرمز حرارتی پهنپایه به دلیل قدرت تفکیک مکانی مناسب، مقرون به صرفه بودن و حجم کمتر تصاویر، یکی از گزینه‌های مطلوب برای هدف پایش وسایل نقلیه است. روش‌هایی که تا به حال برای استخراج وسایل نقلیه از تصاویر حرارتی ارائه شده‌اند، اغلب دارای مشکلاتی نظیر دقت پایین در شناسایی و قطعه‌بندی (مانند روش $HOG+SVM$) و نیاز به کلان داده‌های آموزشی (مانند روش‌های یادگیری عمیق) است. در تحقیق حاضر، یک مدل جدید با نام $SegRBM-Net$ بر اساس یادگیری عمیق و ماشین بولتزمان محدود ارائه شده است. از جمله ویژگی‌های مدل $SegRBM-Net$ ، افزایش دقت شناسایی و قطعه‌بندی وسایل نقلیه از تصاویر حرارتی با استفاده توأم از لایه‌های کانولوشنی و ویژگی‌های ماشین بولتزمان محدود گوسین - برنولی می‌باشد. این ساختار موجب شده است تا الگوریتم، هدف را با سرعت و دقت بیشتری نسبت به سایر روش‌های یادگیری عمیق پیدا کند. به‌منظور ارزیابی کارایی و دقت روش پیشنهادی، از چهار مجموعه داده مادون‌قرمز حرارتی پهنپایه با ویژگی‌هایی نظیر تراکم بالای وسایل نقلیه در صحنه و زاویه دید متنوع استفاده شده است. بر اساس نتایج این تحقیق، مدل $SegRBM-Net$ با دقت میانگین ۹۹ درصد و بهبود سرعت پردازش، نسبت به روش‌های مشابه دارای کارایی مناسبی می‌باشد.

کلیدواژه‌ها: شبکه عصبی کانولوشنی عمیق، ماشین بولتزمان محدود گوسین - برنولی، قطعه‌بندی معنایی، وسایل نقلیه زمینی، تصاویر مادون‌قرمز حرارتی پهنپایه.

* نویسنده مکاتبه کننده: گروه فتوگرامتری و سنجش از دور، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران.

۱- مقدمه

ارائه یک راهبرد هدفمند و علمی برای پایش عناصر طبیعی و مصنوعی در محیطزیست، همواره یکی از موضوعات مهم در برنامه‌ریزی‌های مدیران تصمیم‌گیرنده در کشورها محسوب می‌شود [۱]. بسیاری از تحقیق‌های صورت گرفته، از منابع داده سنجش‌ازدور با قدرت تفکیک مکانی پایین جهت پایش کاربردهای مختلفی چون مدیریت بحران‌های زلزله، سیل، آتش‌سوزی و یا تهیه نقشه‌های حرارتی و رطوبت خاک استفاده کرده‌اند [۲]؛ اما تنها تجزیه‌وتحلیل این محصولات با قدرت تفکیک مکانی پایین توانایی‌های لازم را برای تصمیم‌گیری نخواهد داشت. بسیاری از مخاطرات و نیازهای مدیران شهری در مقیاس‌های محلی، نیازمند اطلاع از جزئیات پدیده‌ها است [۳]. امروزه استفاده از پهپادهای تصویربردار راه‌حل مناسبی برای پایش پدیده‌های مختلف در مقیاس محلی به شمار می‌آیند [۴]. از طرفی استفاده از پهپادهای تصویربردار در کاربردهای مختلف نظامی و غیرنظامی گسترش پیدا کرده است، از طرف دیگر نیاز به پردازش تصاویر آن‌ها با قابلیت جلوگیری از افزونگی داده و سرعت بالا در تولید محصولات اطلاعات مکانی وجود دارد [۵].

مسئله قطع‌بندی معنایی وسایل نقلیه زمینی از تصاویر پهپاد در کاربردهایی نظیر پایش حجم ترافیک و تجزیه‌وتحلیل تصادف‌های درون‌شهری و برون‌شهری، کمک به کنترل خودروهای بدون سرنشین، مدیریت بر فعالیت وسایل نقلیه زمینی در کارگاه‌های عمرانی، مدیریت هوشمند شبکه راه‌های دسترسی با آنالیز حرکت وسایل نقلیه، تجزیه‌وتحلیل عبور و مرور وسایل نقلیه در مناطق ممنوعه، مدیریت پارکینگ‌های موجود در فضای باز و شناسایی وسایل نقلیه مفقودشده بر اثر سوانح طبیعی و غیرطبیعی (سرقت) دارای کاربردهای مهمی است [۶، ۷، ۸، ۹، ۱۰ و ۱۱]. اغلب روش‌های متداول برای انجام قطع‌بندی معنایی تصاویر با هدف کشف وسایل نقلیه مبتنی بر تصویربرداری از روی زمین

و با سکویهای زمینی می‌باشد. مهم‌ترین مشکل در این روش‌ها، پوشش محدود این دوربین‌ها و کاهش قدرت کنترل در ابعاد مناسب است. به همین جهت، به‌کارگیری تصاویر پهپاد یک راه‌حل کارآمد و کم‌هزینه برای پایش عناصر محلی به شمار می‌آید.

هدف از آشکارسازی وسایل نقلیه با استفاده از تصاویر پهپاد، مشخص نمودن موقعیت وسیله نقلیه بر روی تصویر با استفاده از یک مستطیل (جعبه) بوده که چهارگوشه آن دارای موقعیت محلی می‌باشند و بر اساس اندازه شی، این مستطیل کوچک یا بزرگ می‌شود [۱۲]. اما مدل جامع‌تر از آشکارسازی را می‌توان در قطع‌بندی معنایی تصویر جستجو کرد. در قطع‌بندی معنایی تصویر، علاوه بر مشخص نمودن موقعیت شی، می‌توان محدوده آن شی را نیز مشخص کرد. تعیین محدوده شی به‌صورت خودکار، فرآیند کنترل وسایل نقلیه برای جلوگیری از برخورد با سایر عناصر در خیابان‌های شلوغ را در تصاویر با قدرت تفکیک مکانی بالا تسهیل می‌کند [۱۳]. در شکل (۱) نمونه‌ای از آشکارسازی وسایل نقلیه از تصاویر رنگی پهپاد با استفاده از معماری یادگیری عمیق یولو^۱ [۱۴] (شکل ۱-ب) در مقایسه با قطع‌بندی معنایی وسایل نقلیه با استفاده از معماری یادگیری عمیق سگنت^۲ [۷] (شکل ۱-ج) نشان داده شده است. بر اساس این شکل می‌توان قابلیت‌های قطع‌بندی معنایی را بهتر درک کرد، به‌ویژه در مواردی که تراکم خودروها زیاد است (بردار مشکی) و صرفاً آشکارسازی آن‌ها با یک جعبه آشکارساز نمی‌تواند راه‌حل مناسبی باشد.

تلفیق فیلتر گابور^۳ جهت استخراج ویژگی برای برجسته نمودن لبه‌های وسایل نقلیه با استفاده از طبقه‌بندی‌کننده ماشین بردار پشتیبان^۴ جهت

^۱ You Only Look Once (YOLO)

^۲ SegNet

^۳ Gabor Filter

^۴ Support Vector Machine (SVM)

می آید [۱۵].

آشکارسازی وسایل نقلیه در تصاویری از خیابان‌ها، یکی از اولین تلاش‌ها برای آشکارسازی وسایل نقلیه به شمار



شکل ۱: دیدگاه‌های مختلف استخراج وسایل نقلیه از تصاویر پهپاد (جعبه زرد: موقعیت خودروها، ناحیه صورتی: قطعه‌بندی خودروها)

کانوولوشنی^۵ [۱۲] نیز از جمله راه‌حل‌های پیشنهادی است. به‌طورکلی، روش‌های پیشنهادی برای آشکارسازی و قطعه‌بندی معنایی وسایل نقلیه از تصاویر گوناگون پهپاد در سه گروه دسته‌بندی می‌شوند. دسته اول، روش‌های مبتنی بر استخراج ویژگی غیر خودکار مانند روش‌های تبدیل مستقل از مقیاس و ویژگی، ویژگی‌های تسریع شده مقاوم^۶، هیستوگرام گرادیان شیب گرا، لبه یابی، تعریف قواعد طیفی (در صورت چند طیفی بودن تصاویر) و استفاده ترکیبی از این تبدیل‌ها است [۲۰]. دسته دوم، به‌کارگیری روش‌های یادگیری ماشین مانند ماشین بردار پشتیبان و انواع شبکه‌های عصبی مصنوعی می‌باشد [۲۱]. دسته سوم، روش‌های مدرن جهت استخراج ویژگی با استفاده از الگوریتم‌های یادگیری عمیق است که می‌توان این روش را نوع توسعه‌یافته دو روش قبل دانست [۲۲]. در حال حاضر الگوریتم‌های یادگیری عمیق به دلیل قابلیت مناسب برای پردازش کلان داده‌ها در کاربردهای مختلف دانشگاهی و صنعتی مورد توجه است [۲۳]. اغلب روش‌های یادگیری ماشین برای پردازش کلان داده‌ها

استفاده از تصاویر رنگی و استخراج لبه اشیاء تصویری (با روش‌های متداول لبه یابی) راه‌حلی دیگر برای آشکارسازی وسایل نقلیه است [۱۶]. از جمله تحقیق‌های صورت گرفته در حیطه آشکارسازی و قطعه‌بندی معنایی وسایل نقلیه از تصاویر رنگی پهپاد می‌توان به مواردی چون تلفیق تبدیل مستقل از مقیاس و ویژگی^۱ به دلیل عدم تأثیرپذیری به مقیاس، دوران و شرایط نوردهی با مدل شکل ضمنی^۲ برای تولید توصیفگرهای ویژگی و طبقه‌بندی ویژگی‌ها با ماشین بردار پشتیبان اشاره کرد [۱۷]. الگوریتم دیگر، از روش ترکیبی ویولا-جونز^۳ و طبقه‌بندی کننده ماشین بردار پشتیبان با ویژگی‌های هیستوگرام گرادیان شیب گرا^۴ برای تصاویر با فواصل کم از وسیله نقلیه استفاده نموده است [۱۸]. بهره‌گیری از تصاویر چند طیفی پوشش‌دار با هدف تولید نقشه عمق و تمیز دادن وسیله نقلیه از سایر عوارض [۱۹] و به‌کارگیری تصاویر مادون قرمز حرارتی پهپاد با هدف آشکارسازی بی‌درنگ جعبه‌ای وسایل نقلیه با استفاده از شبکه‌های عصبی

^۱ Scalar Invariant Feature Transform (SIFT)

^۲ Implicit Shape Model (ISM)

^۳ Viola-Jones (V-J)

^۴ Histogram of Oriented Gradient (HOG)

^۵ Convolutional Neural Networks (CNNs)

^۶ Speeded-Up Robust Features (SURF)

حرارتی $NPU_CS_UAV_IR_DATA$ که دارای چالش‌هایی بیشتر در مقایسه با سایر منابع داده بوده و با هدف قطعه‌بندی معنایی وسایل نقلیه صورت می‌گیرد [۱۲].

۲- روش پیشنهادی

در این قسمت، ابتدا مروری بر به‌کارگیری الگوریتم‌های یادگیری عمیق در قطعه‌بندی معنایی تصویر شده است، سپس ماشین بولتزمان محدود گوسین - برنولی با هدف استخراج مدل هندسی اشیاء در تصویر مورد بررسی قرار گرفته است. در نهایت، روش پیشنهادی معرفی و فرآیند پیاده‌سازی ارائه می‌گردد.

۲-۱- قطعه‌بندی معنایی تصویر با الگوریتم‌های یادگیری عمیق

اولین نمونه از مدل‌های شبکه عصبی کانولوشنی عمیق حدود دو دهه قبل ارائه گردیده است [۲۸]. در طی این سال‌ها موضوع ارائه یک معماری مناسب برای قطعه‌بندی معنایی اشیاء از تصاویر به‌عنوان یک چالش مهم در حیطه علوم یادگیری ماشین^۵ به‌حساب می‌آید [۲۹]. منطق ارائه معماری‌های یادگیری عمیق بر افزایش تعداد لایه‌های پنهان در شبکه عصبی مصنوعی بنا شده است. اما به‌مرور با ارائه لایه‌های کانولوشنی و تلفیق آن‌ها در شبکه‌های عصبی مصنوعی، استفاده از این مدل‌ها در پردازش تصاویر بسیار مؤثرتر شده است [۱۳]. افزایش تعداد لایه‌های پنهان در شبکه‌های عصبی موجب می‌شود تا ویژگی‌هایی از داده ورودی (به‌عنوان مثال تصاویر رنگی) در سطوح مختلف استخراج شود. این سطوح از ویژگی‌ها شامل مشخصه‌های سطح پایین مانند لبه‌ها تا مشخصه‌های سطح بالا مانند مدل هندسی اشیاء را شامل می‌شوند [۳۰]. شبکه‌های عصبی کانولوشنی عمیق به دلیل کاهش تعداد پارامترها (وزن و بایاس) دارای قدرت تعمیم‌پذیری بالایی نسبت به مدل‌هایی با لایه‌های تماماً متصل هستند؛ از همین رو

دچار مشکل اشباع‌شدگی^۱ می‌شوند و نمی‌توانند از یک حد مشخصی بیشتر یاد بگیرند؛ به همین دلیل می‌توان یادگیری عمیق را یک روش متحول‌کننده در پردازش تصویر دانست [۲۴]. امروزه با فراگیر شدن سنجنده‌های تصویربرداری و تولید مجموعه داده‌های بزرگ، مسأله تفسیر خودکار اشیاء از تصاویر سنجنش‌ازدور (ماهواره‌ای، هوایی و پهپاد) با عملکرد مناسب با استفاده از الگوریتم‌های یادگیری عمیق به‌عنوان یک موضوع چالشی مهم در پردازش تصاویر است. همچنین، استفاده از تصاویر مادون‌قرمز حرارتی پهپاد به دلیل قدرت تفکیک مکانی مناسب، مقرون‌به‌صرفه بودن و حجم کمتر تصاویر، یکی از گزینه‌های مطلوب برای هدف پایش وسایل نقلیه نسبت به سایر منابع داده به شمار می‌آید. تحقیق‌های اخیر در پردازش تصاویر نشان می‌دهد، شبکه‌های عصبی عمیق مبتنی بر لایه‌های کانولوشن دارای قابلیت‌های مناسبی در حل مسأله استخراج اشیاء از تصاویر است [۲۵].

تحقیق حاضر با هدف ارائه و آزمایش یک مدل جدید ترکیبی مبتنی بر یادگیری عمیق و ماشین بولتزمان محدود گوسین - برنولی جهت قطعه‌بندی معنایی وسایل نقلیه زمینی با استفاده از تصاویر مادون‌قرمز حرارتی پهپاد انجام شده است. مدل پیشنهادی در این تحقیق $SegRBM-Net$ نام‌گذاری شده است. طراحی و پیاده‌سازی مدل $SegRBM-Net$ بر اساس ترکیب ایده معماری‌های یادگیری عمیق رزنت^۲ [۲۶]، یونت^۳ [۲۷] و سگنت با ماشین بولتزمان محدود صورت گرفته است. در این مدل، ماشین بولتزمان محدود در نقش بلوک‌های باقی‌مانده در یک شبکه کدگذار-کدگشا^۴ عمیق ظاهر می‌شود. فرآیند آزمایش و ارزیابی مدل $SegRBM-Net$ بر روی مجموعه تصاویر مادون‌قرمز

^۱ Saturation

^۲ ResNet

^۳ U-Net

^۴ Encoder-Decoder

^۵ Machine Learning (ML)

اطلاعات تصویر پُر شود و ردیف‌ها و ستون‌های فرد با مقادیر صفر پُر می‌شود. حال برای مقدار دادن به این خانه‌های صفر، بایستی از روش درون‌یابی کنترل‌شده استفاده کنیم. محاسبات یک درون‌یابی کنترل‌شده جهت بدست آوردن مقادیر خانه‌های صفر با قرار دادن کرنل گوسی در مرکز خانه با ضرب آن در مقادیر غیر صفر و مقیاس دهی مناسب نتیجه به مقدار یک، برای مقدار خانه موردنظر محاسبه می‌شود. معماری سگت از نظر حافظه مصرفی، دقت تولیدی و همچنین کاهش پارامترهای شبکه، نسبت به معماری‌های دیگر، بسیار عملکرد مناسبی در مورد قطعه‌بندی تصویر دارد [۷].

۲) معماری رزنت بر اساس اضافه نمودن یک بردار از پیکسل‌های تصویر ورودی، بر خروجی‌های کدگذاری شده عمل می‌کند.

۳) معماری یونت مبتنی بر یک لایه کانولوشنی مابین بلوک کدگذار و کدگشا با هدف بهبود یادگیری ویژگی عمل می‌کند.

هر لایه کانولوشنی معماری‌های شبکه عصبی مصنوعی مشتمل بر یک وزن پنجره‌ای (به ازای هر باند طیفی در تصویر یک وزن پنجره‌ای می‌توان در نظر گرفت)، یک تابع نرمال ساز بسته‌ها^۴ و یک تابع فعال‌سازی می‌باشد. رابطه (۱) تعریف ریاضی یک نورون در معماری‌های کانولوشنی را بر اساس تصویر ورودی (P) و وزن پنجره‌ای (C) نشان می‌دهد [۳۰].

در رابطه (۱)، $(P \times C)_{ij}$ حرکت وزن پنجره‌ای به همسایگی (i, j) بر روی تصویر می‌باشد؛ $\sum m$ مقادیر سطر تصویر، $\sum n$ مقادیر ستون تصویر، $\sum c$ تعداد باند طیفی تصویر می‌باشد. در این تحقیق به دلیل استفاده از تصویر مادون‌قرمز حرارتی مقدار $\sum c$ برابر ۱ در نظر گرفته می‌شود و b بیانگر مقدار بایاس است که یک مقدار ثابت در نظر گرفته می‌شود.

در کاربردهایی پایه مانند آشکارسازی اشیاء [۳۱]، استخراج مشخصه [۳۲] و طبقه‌بندی تصاویر [۲۵] دارای قابلیت‌های مناسبی می‌باشند.

اغلب معماری‌های قطعه‌بندی معنایی تصویر با یادگیری عمیق بر اساس یک بلوک انقباض (کدگذار) و یک بلوک انبساط (کدگشا) طراحی شده‌اند [۷]. بلوک انقباض جهت تولید ویژگی‌های اشیاء تصویری و بلوک انبساط جهت بازسازی تصویر ورودی مورد استفاده قرار می‌گیرند [۳۳]. وجه تمایز معماری‌های گوناگون به‌کارگیری چندین لایه کانولوشنی و استفاده از توابع فعال‌سازی^۱ مختلف است [۱۱]. به‌صورت کلی، ارائه تعداد لایه کانولوشنی مناسب با پارامترهای بهینه نیازمند تئوری ریاضی قوی می‌باشد. از میان معماری‌های یادگیری عمیق که برای قطعه‌بندی معنایی تصویر ارائه شده‌اند می‌توان سه معماری تأثیرگذار سگت، رزنت و یونت را نام برد. برجسته‌ترین ویژگی‌های این معماری‌ها عبارت است از:

۱) معماری سگت اولین بار در سال ۲۰۱۵ با آزمایش بر روی یک مجموعه داده از تصاویر دوربین‌های عکاسی ارائه شد. توپولوژی معماری سگت بر اساس یک شبکه کدگذار با سیزده لایه کانولوشنی در شبکه VGG16 (یک معماری یادگیری عمیق با ساختار هرمی مبتنی بر شانزده لایه وزن کانولوشنی) طراحی شده است. بلوک کدگذار شامل پنج زیر بلوک و هر زیر بلوک شامل چند لایه کانولوشنی و یک لایه کاهش‌دهنده ابعاد^۲ می‌باشد. به‌طور متناظر در بلوک کدگشا نیز پنج زیر بلوک با لایه‌های معکوس کانولوشنی و یک لایه نمونه‌افزایی^۳ وجود دارد. نوآوری معماری سگت در به‌کارگیری لایه‌های نمونه‌افزا (بازسازی ابعاد اصلی تصویر) است [۳۳]. این‌گونه پردازش موجب می‌شود تا ابعاد تصویر منبسط شود و ردیف‌ها و ستون‌های زوج با

^۱ Activation Functions

^۲ Pooling

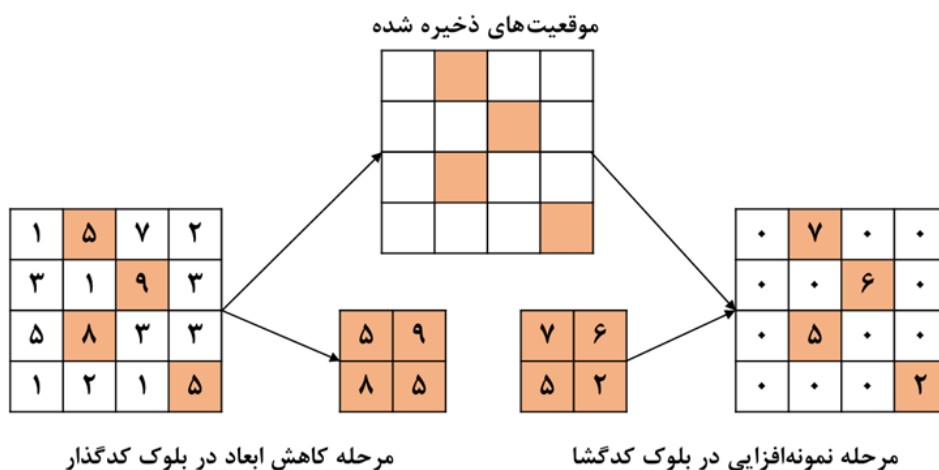
^۳ Upsampling

^۴ Batch Normalization (BN)

$$(P \times C)_{ij} = \sum_m \sum_n \sum_c P(i+m, j+n). C(-m, -n) + b \quad (1)$$

تابع ذخیره کردن موقعیت پیکسل با مقدار بیشینه و با همسایگی موردنظر (ابعاد متداول با همسایگی ۲×۲) ذخیره و سپس در مرحله نمونه افزایشی که دقیقاً عملکردی معکوس این تابع را دارد با مقادیر جدید جایگزین شود. این کار باعث می‌شود تا ویژگی‌های تولیدشده در طول شبکه عصبی کانولوشنی با محاسبات کمتری در موقعیت صحیح خود قرار بگیرند. شکل (۲) فرآیند اعمال لایه کاهش‌دهنده ابعاد و لایه نمونه‌افزا را بر روی یک همسایگی از پیکسل‌ها نمایش می‌دهد.

برای جلوگیری از پدید بیش برازش (Overfitting) راه‌حل‌های مختلفی در شبکه‌های عصبی کانولوشنی عمیق پیشنهاد شده است. یکی از این راه‌حل‌ها استفاده از لایه‌های کاهش‌دهنده ابعاد است [۳۴]. این فرآیند با کدگذاری پیکسل‌ها در یک تصویر موجب می‌شود تا پارامترهای شبکه عصبی عمیق کاهش پیدا کرده و به‌موجب حجم محاسبات کمتر، فرآیند آموزش با دقت بیشتری انجام شود. توابع مختلفی برای انجام کاهش ابعاد مورد استفاده قرار می‌گیرند که از جمله توابع مناسب، تابع کاهش‌دهنده بیشینه است. اساس کار این



شکل ۲: فرآیند اعمال لایه کاهش ابعاد در بلوک کدگذار و لایه نمونه‌افزا در بلوک کدگشا

(باینری) می‌باشد. ماشین بولتزمان محدود یک نوع ماشین بولتزمان است که اتصالات واحدهای مشاهده‌پذیر و پنهان قطع شده است [۳۶]. ماشین بولتزمان محدود گوسین - برنولی^۲ شامل یک لایه از واحدهای مشاهده‌پذیر گوسین و یک لایه از واحدهای پنهان باینری می‌باشد [۳۷]. در این نوع ماشین بولتزمان هر واحد با سایر واحدهای لایه‌های دیگر

۲-۲- استخراج مدل هندسی اشیاء تصویری با ماشین بولتزمان محدود گوسین - برنولی
 ماشین بولتزمان یک نوع خاص از میدان تصادفی مارکوف می‌باشد که اولین بار در سال ۱۹۸۵ ارائه گردید [۳۵]. از جمله کاربردهای مهم ماشین بولتزمان در حل مسأله طبقه‌بندی، رگرسیون، کاهش ابعاد و استخراج ویژگی می‌باشد. این ماشین مشتمل بر تعدادی واحد مشاهده‌پذیر^۱ و پنهان^۲ به‌صورت دودویی

^۲ Hidden Units

^۳ Gaussian-Bernoulli Restricted Boltzmann Machine (GB-RBM)

^۱ Visible Units

رابطه (۲) تعریف می شود [۳۸].

$$E(v, h|\theta) = \sum_{i=1}^{n_v} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} w_{ij} h_j \frac{v_i}{\sigma_i^2} - \sum_{j=1}^{n_h} c_j h_j \quad \text{رابطه (۲)}$$

احتمالی که مدل یادگیری بولتزمان محدود به واحدهای مشاهده پذیر نسبت می دهد را می توان به صورت رابطه (۵) بیان نمود. در رابطه (۵)، $\mathcal{N}(\cdot|\mu, \sigma_i^2)$ تابع گوسین با میانگین μ و واریانس σ_i^2 است.

به طور مشابه برای واحدهای پنهان، احتمال با استفاده از رابطه (۶) محاسبه می شود. لایه های پنهان به دلیل عدم اتصال به یکدیگر مستقل از شرایط واحدهای مشاهده پذیر محاسبه می شوند.

$$P(v, h|\theta) = \frac{1}{Z(\theta)} \exp\{-E(v, h|\theta)\} \quad \text{رابطه (۳)}$$

$$Z = \sum_v \sum_h \exp\{-E(v, h|\theta)\} \quad \text{رابطه (۴)}$$

$$P(v_i = v|h) = \mathcal{N}(v|b_i + \sum_j w_{ij} h_j, \sigma_i^2) \quad \text{رابطه (۵)}$$

$$P(h_j = 1|v) = \text{sigmoid}(c_j + \sum_i w_{ij} \frac{v_i}{\sigma_i^2}) \quad \text{رابطه (۶)}$$

خطا؛ $KL(\rho||\hat{\rho}_k)$ واگرایی *Kullback-Leibler* بین ρ و $\hat{\rho}_k$ که به صورت رابطه (۸) محاسبه می شود. دلیل استفاده از تابع توزیع بولتزمان، ساختار دو قسمتی است که این شبکه دارد. شکل (۳) این ساختار دو قسمتی ماشین بولتزمان محدود گوسین - برنولی را برای استخراج مدل اولیه یک وسیله نقلیه از تصویر مادون قرمز حرارتی پهنای مبتنی بر واحدهای مشاهده پذیر و پنهان را نشان می دهد.

اتصال دارند اما اتصالات پنهان نیست. تابع انرژی در ماشین بولتزمان محدود گوسین - برنولی به صورت

در رابطه (۲)، v_i واحدهای مشاهده پذیر با تعداد n_v h_j واحدهای پنهان با تعداد n_h θ مشتمل بر وزن های w_{ij} که به واحدهای مشاهده پذیر و پنهان متصل می شوند؛ σ_i انحراف معیار همبسته به واحد مشاهده پذیر گوسین v_i ؛ b_i و c_j بایاس های هریک از واحدهای مشاهده پذیر و پنهان است. در این شبکه به هر مقدار از واحدهای مشاهده پذیر و پنهان با استفاده از رابطه (۱)، یک مقدار احتمال که به آن توزیع بولتزمان گفته می شود به صورت رابطه (۳) نسبت می دهد. در رابطه (۳)، Z ثابت نرمال سازی است و با استفاده از رابطه (۴) تعریف می شود.

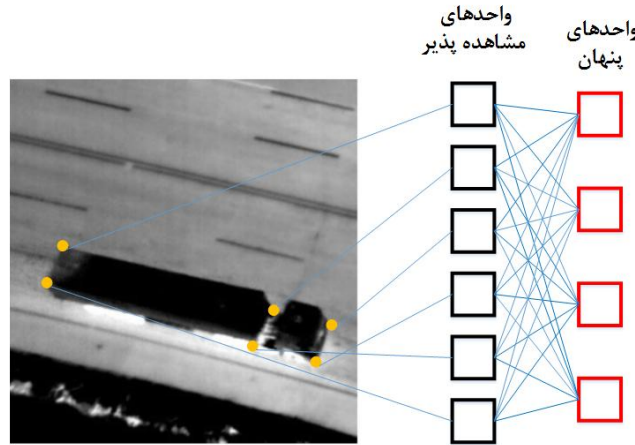
در رابطه (۶)، *sigmoid* نماد تابع سیگموئید می باشد، این تابع حقیقی، کران دار و مشتق پذیر می باشد و به ازای کلیه مقادیر حقیقی قابل تعریف بوده و دارای مشتق مثبت است. در نهایت برای رسیدن به انرژی کمتر برای مجموعه آموزشی و انرژی بیشتر برای بقیه مجموعه، از مدل پیشنهادی [۳۷] که خطاهای فاحش^۱ را مدل می کند، به صورت رابطه (۷) تعریف می شود.

در رابطه (۷)، ρ پارامتر تُنکی؛ $\hat{\rho}_k$ میانگین فعال سازی واحدهای پنهان در نمونه آموزشی k ؛ β وزن تُنک تابع

^۱ Gross Errors

$$\arg \min = \{\sum_{k=1}^m -\log \tilde{P}(v_k; \theta) + \beta KL(\rho || \hat{\rho}_k)\} \quad \text{رابطه (۷)}$$

$$KL(\rho || \hat{\rho}_k) = \rho \log \frac{\rho}{\hat{\rho}_k} + (1 - \rho) \log \frac{1-\rho}{1-\hat{\rho}_k} \quad \text{رابطه (۸)}$$



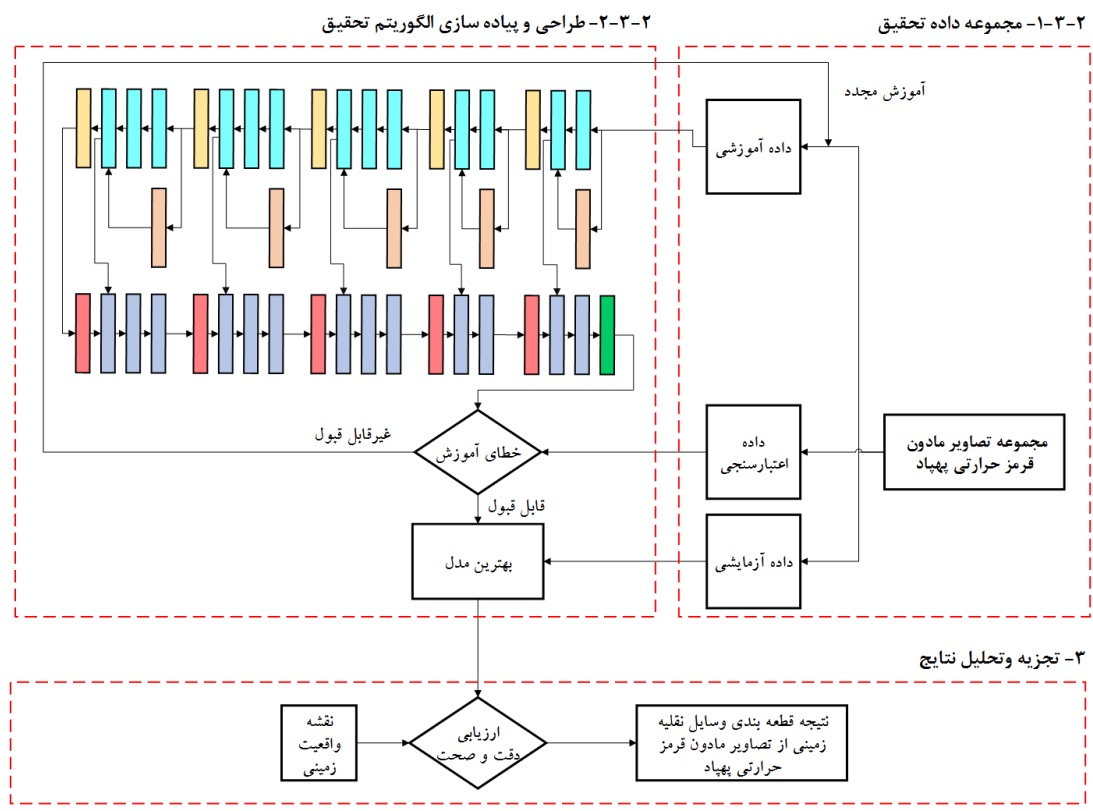
شکل ۳: استخراج مدل اولیه یک وسیله نقلیه از تصویر مادون قرمز حرارتی پهپاد مبتنی بر ساختار دو قسمتی ماشین بولتزمان محدود گوسین - برنولی

می‌شود. مرحله اول، شامل جمع‌آوری مجموعه تصاویر مادون قرمز حرارتی پهپاد و آماده‌سازی داده‌ها برای آموزش الگوریتم پیشنهادی است (بخش ۲-۳-۱). مرحله دوم، شامل طراحی معماری پیشنهادی تحقیق با نام *SegRBM-Net* مبتنی بر یک ساختار کدگذار-کدگشا که هر زیربلوک کدگذار شامل یک لایه مشتمل بر ماشین بولتزمان محدود گوسین - برنولی با هدف بهبود ویژگی‌های دقیق هندسی از وسایل نقلیه استفاده می‌شود. سپس با نتیجه هر زیر بلوک جمع می‌شوند. به دلیل هم‌جنس بودن نتایج، سازگاری بین نتایج لایه ماشین بولتزمان محدود گوسین - برنولی و هریک از زیربلوک‌های کدگذار وجود دارد. این فرآیند برای هریک از پنج زیربلوک کدگذار انجام می‌شود. در این مرحله، میزان خطای آموزش برای نتایج محاسبه می‌شود و سپس با استفاده از داده‌های اعتبارسنجی، مورد ارزیابی قرار می‌گیرند. معیار توقف آموزش تعداد تکرار ۱۰۰ مرتبه می‌باشد. در نهایت بهترین مدل بر مبنای کمترین خطای اعتبارسنجی تولید و ذخیره می‌شود (بخش ۲-۳-۲). در مرحله سوم، فرآیند ارزیابی مدل پیشنهادی با استفاده از داده‌های آزمایشی صورت

فرآیند پردازش ماشین بولتزمان محدود بر مبنای یافتن ویژگی‌های منظم مثل شکل هندسی اشیا تصویری انجام می‌شود. مطابق شکل (۳)، هدف از تولید بردار ویژگی‌های هندسی محدوده وسیله نقلیه (شکل هندسی) بدون یادگیری قبلی با استفاده از تابع انرژی (رابطه ۲) می‌باشد. ویژگی اصلی ماشین بولتزمان محدود گوسین - برنولی در پردازش تصویر، استخراج مدل هندسی شیء در تصویر است [۳۹]. این قابلیت موجب می‌شود تا اگر هندسه شیء در تصویر ضعیف (مثلاً تصاویر درجه خاکستری محدوده طیف حرارتی) باشد؛ با استفاده از این ماشین، هندسه را بهبود داد. در تحقیق حاضر روشی ارائه می‌شود تا با استفاده از ویژگی‌های تولیدشده با استفاده از این ماشین در یک معماری کدگذار-کدگشا از شبکه‌های عصبی کانوولوشنی عمیق، فرآیند استخراج وسایل نقلیه از تصاویر مادون قرمز حرارتی پهپاد با دقت بالا و قابلیت انعطاف‌پذیری برای قطعه‌بندی معنایی خودکار وسایل نقلیه زمینی ایجاد شود.

۲-۳- مدل پیشنهادی *SegRBM-Net* مطابق شکل (۴)، روش پیشنهادی در سه مرحله اجرا

گرفته و سپس با به کارگیری معیارهای متداول دقت و صحت از نظر عددی مورد تجزیه و تحلیل قرار می گیرد (بخش ۳).



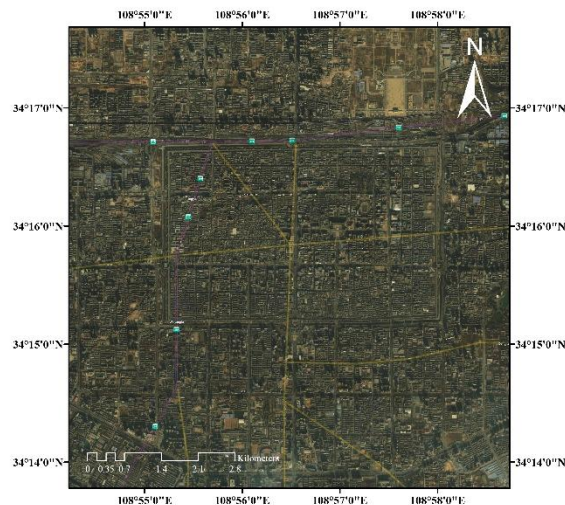
شکل ۴: الگوریتم پیشنهادی تحقیق

۲-۳-۱- مجموعه داده تحقیق

تصاویر مادون قرمز حرارتی اطلاعات مفیدی در مورد توزیع سطوح مختلف انرژی تابشی مادون قرمز از سطح جسم قابل اندازه گیری از محدوده ۸ تا ۱۵ میکرومتر در طیف الکترومغناطیس را به دست می دهد [۴۰]. در مسأله آشکارسازی خودرو استفاده از سنجنده های حرارتی این قابلیت را ایجاد می کند تا فاصله ۵۰۰ متر وسایل نقلیه در شرایط مختلف نوری و هوایی قابل تشخیص باشند [۴۱]. بعلاوه قابلیت پایش عوارض در روز و شب مستقل از نور طبیعی یا مصنوعی ایجاد می شود [۴۲]. همچنین تصاویر مادون قرمز حرارتی پهپاد دارای سه ویژگی مهم است: اول، قدرت تفکیک مکانی بالا. دوم، قدرت تفکیک زمانی بالا و سوم، امکان

اخذ تصاویر پوشش دار است. وجود این ویژگی ها موجب افزایش اهمیت پردازش تصاویر مادون قرمز حرارتی در کاربردهای مختلف و روزمره طی چند سال اخیر شده است. مجموعه داده مورد استفاده در این تحقیق با نام *NPU_CS_UAV_IR_DATA* شامل یک مجموعه از چهار دسته تصویربرداری مادون قرمز حرارتی به وسیله سنجنده *FLIR TAU2* و با پهپاد *DJI-MATRICE 100* از خیابان های مختلف شهر شیآن^۱ در کشور چین در سال ۲۰۱۷ میلادی می باشد. در شکل (۵) موقعیت جغرافیایی منطقه مورد مطالعه به همراه نقشه اطلاعات مکانی راه های دسترسی، نشان داده شده است.

^۱ Xi'an



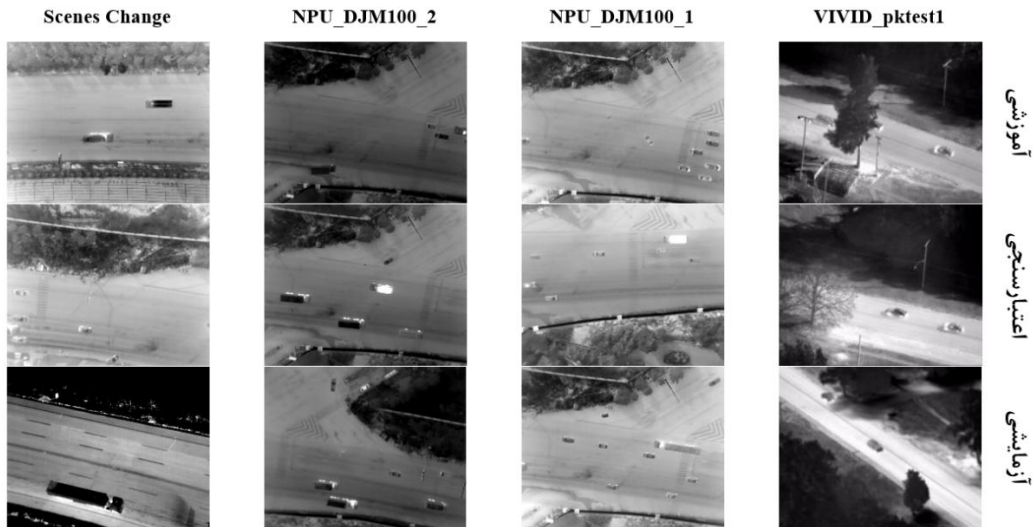
شکل ۵: منطقه مورد مطالعه در این تحقیق (بزرگراه با رنگ بنفش و خیابان‌های اصلی با رنگ زرد در تصویر نشان داده شده است)

مجموعه داده در جدول (۱) ارائه شده است. شکل (۶)، نمونه‌هایی از تصاویر آموزشی، اعتبارسنجی و آزمایشی را برای هر یک از مجموعه تصاویر نشان می‌دهد. این مجموعه داده در [۴۳] قابل دسترسی است.

سنجنده *FLIR TAU2* با وزن کمتر از ۷۰ گرم و با بیشینه حد تفکیک‌پذیری 640×512 پیکسل دارای دقت متوسط ± 5 درجه سانتی‌گراد می‌باشد، این ویژگی‌ها موجب شده است تا گزینه‌ای مناسب برای استفاده از این سنجنده بر روی پهپاد باشد. جزئیات این

جدول ۱: مشخصات مجموعه داده تحقیق

ارتفاع پرواز (متر)	ابعاد تصاویر بر حسب پیکسل	تصاویر آزمایشی (تعداد فریم)	تصاویر اعتبارسنجی (تعداد فریم)	تصاویر آموزشی (تعداد فریم)	نام مجموعه داده
۸۰	320×240	۵۰	۳۱	۱۷۰	<i>VIVID_pktest1</i>
۱۲۰	640×512	۵۰	۳۱	۱۵۱	<i>NPU_DJM100_1</i>
۱۲۰	640×512	۵۰	۴۰	۴۱۶	<i>NPU_DJM100_2</i>
۸۰	640×512	۲۰	۱۰	۷۰	<i>Scenes Change</i>
		۱۷۰	۱۱۲	۸۰۷	مجموع



شکل ۶: نمونه‌هایی از تصاویر آموزشی، اعتبارسنجی و آزمایشی به تفکیک هر یک از مجموعه تصاویر

۲-۳-۲- طراحی و پیاده‌سازی الگوریتم تحقیق

با توجه به اینکه هدف این تحقیق قطعه‌بندی معنایی وسایل نقلیه از تصاویر مادون قرمز حرارتی پهپاد با یادگیری عمیق است، سه موضوع زیر در طراحی الگوریتم پیشنهادی مورد توجه است:

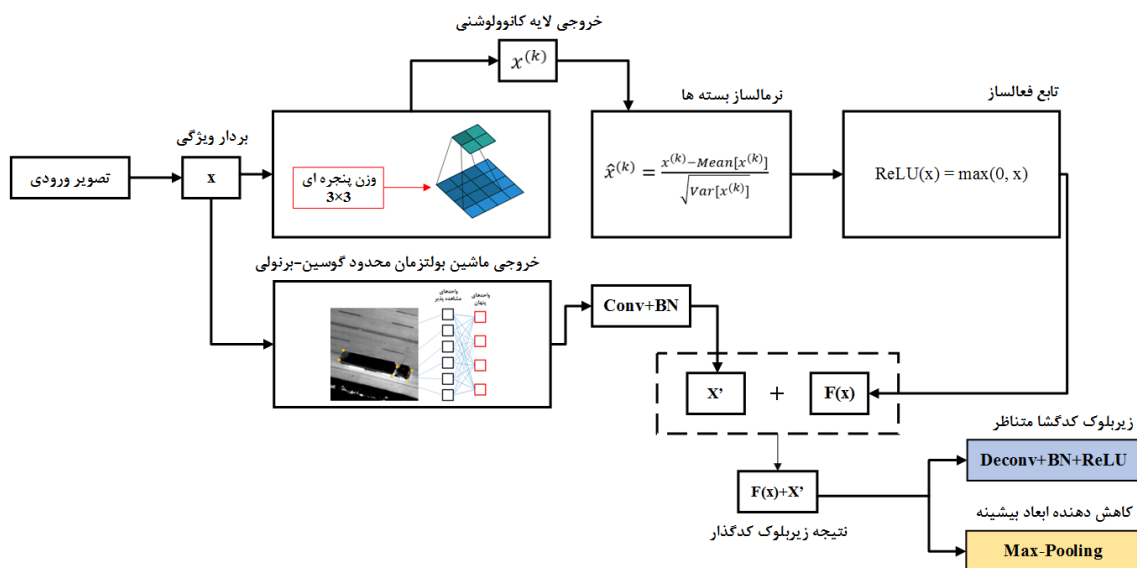
(۱) مجموعه تصاویر مورد استفاده تک بانده هستند، پس استخراج یک هندسه مناسب از وسایل نقلیه از نمای بالا دارای کاربرد مناسبی می‌باشد. به همین علت از ماشین بولتزمان محدود گوسین - برنولی جهت استخراج هندسه وسایل نقلیه استفاده شد. فرآیند استخراج اطلاعات هندسی به عنوان بلوک باقی‌مانده در سراسر شبکه مورد استفاده قرار گرفته است. علت اصلی این کار محدود بودن داده آموزشی (تصاویر حرارتی با یک باند طیفی) است که با این کار علاوه بر استفاده از داده‌های آموزشی می‌توان از یک بردار ویژگی (حاوی اطلاعات هندسی وسایل نقلیه) نیز استفاده نمود. بلوک باقی‌مانده موجب می‌شود تا بردار ویژگی اصلی (تصاویر حرارتی) در طول آموزش حفظ شود.

(۲) نحوه به کارگیری ماشین بولتزمان محدود گوسین - برنولی در یک ساختار شبکه عصبی کانولوشنی عمیق است. برای چنین عملکردی با توجه به سازگاری خروجی زیربلوک‌ها با ویژگی‌های تولیدی توسط ماشین

بولتزمان، می‌توانیم حاصل مجموع خروجی هر یک از زیر بلوک‌های شبکه عصبی کانولوشنی را با ویژگی‌هایی که توسط ماشین بولتزمان به دست آمده جمع کنیم و سپس نتایج را وارد زیربلوک بعدی کنیم. این کار موجب می‌شود تا ویژگی‌های بهینه‌ای که از هر زیربلوک شبکه عصبی کانولوشنی ایجاد شده است؛ تحت تأثیر ویژگی‌های به دست آمده از ماشین بولتزمان قرار بگیرند تا علاوه بر اطلاعات طیفی، خصوصیات هندسی وسایل نقلیه تقویت گردد؛ به این ترتیب روند آموزش تقویت می‌شود. برای جمع کردن نتیجه ماشین بولتزمان محدود و شبکه عصبی از یک فیلتر کانولوشنی و یک تابع نرمال ساز بسته‌ها استفاده می‌کنیم. با این کار ابعاد بردارها را با یکدیگر هم‌اندازه می‌کنیم تا بتوانند با یکدیگر جمع شوند. یک بلوک کدگذار در معماری SegRBNet شامل بر لایه‌های کانولوشنی به ابعاد 3×3 و با وزن دهی تصادفی، برای هر لایه کانولوشنی یک تابع نرمال ساز بسته‌ها جهت جلوگیری از پدیده بیش برآزش و یک تابع فعال‌سازی رلو (این تابع به دلیل اینکه در نواحی مثبت دارای مشتق صفر است هیچ‌گاه به اشباع نمی‌رسد) می‌باشد. شکل (۷) فرآیند تلفیق لایه‌های کانولوشنی با ویژگی‌های تولیدشده از ماشین بولتزمان محدود

ویژگی‌هایی تولید می‌شود که به آن x' می‌گوییم، در نهایت با جمع کردن نتایج هریک از زیربلوک‌ها با یکدیگر به مقدار $F(x)+x'$ که ویژگی‌هایی قوی‌تر تولید می‌کند دست پیدا می‌کنیم. علت اصلی برای تولید این ویژگی‌ها تلفیق اطلاعات طیفی و هندسی می‌باشد که در طراحی این شبکه موردنظر می‌باشد.

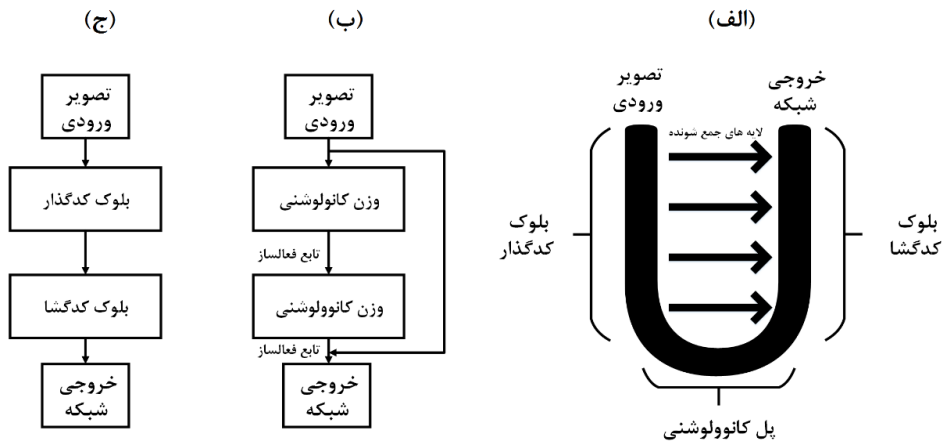
گوسین-برنولی را نشان می‌دهد. بر این اساس اگر بردار ویژگی تصویر ورودی را برابر x در نظر بگیریم، با عبور از یک لایه کانوولوشنی ما با یک مقدار $F(x)$ مواجه هستیم که با توجه به ویژگی‌های تولیدشده در یک زیربلوک کدگذار دارای مقادیر مشخصی هستند. از طرفی به‌صورت مستقل x را به‌عنوان ورودی ماشین بولتزمان محدود استفاده می‌کنیم و در این حالت نیز



شکل ۷: الگوی تلفیق خروجی ماشین بولتزمان محدود گوسین - برنولی و معماری پیشنهادی در زیربلوک‌های کدگذار و کدگشا

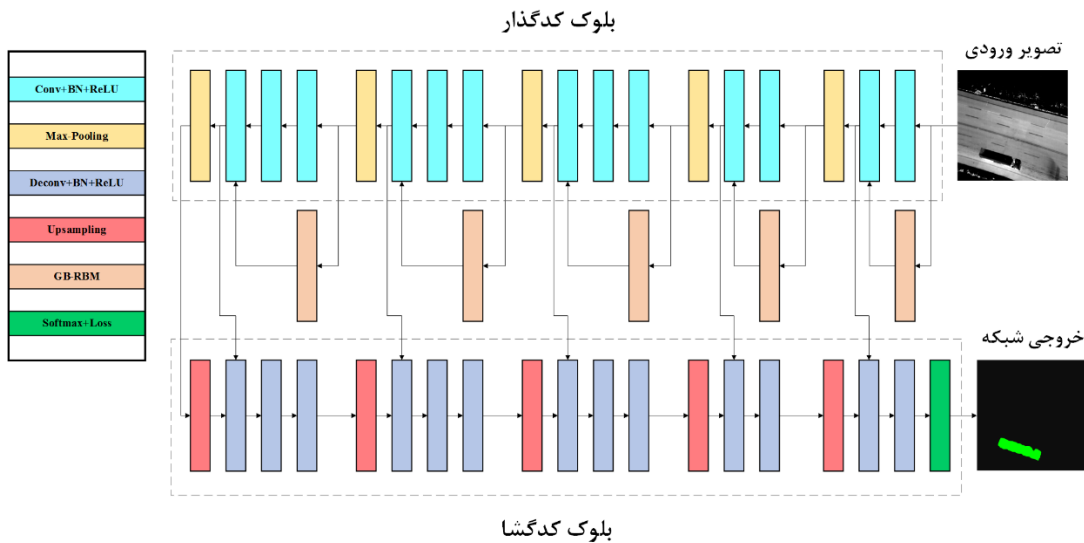
معماری رزنت (شکل ۸-ب) و به‌کارگیری ایده معماری سنگت (شکل ۸-ج)، مدل کلی معماری پیشنهادی این تحقیق طراحی شده است.

۳) ساختار شبکه عصبی کانوولوشنی عمیق می‌بایست از یکی از معماری‌های آزمایش‌شده پیروی نماید تا روند آموزش تا حد زیادی کنترل شده باشد. به همین علت با ایده گرفتن از ساختار معماری یونت (شکل ۸-الف)،



شکل ۸: بلوک های سازنده معماری های (الف) یونت؛ (ب) رزنت؛ (ج) سگنت

طراحی معماری پیشنهادی تحقیق بر اساس شکل (۹) و جدول (۳) ساختار بلوک کدگشا در معماری صورت گرفته است. در جدول (۲) ساختار بلوک کدگذار *SegRBM-Net* ارائه شده است.



شکل ۹: معماری پیشنهادی تحقیق با نام *SegRBM-Net* (در خروجی شبکه ناحیه سبز: قطعه بندی وسیله نقلیه، ناحیه مشکی: پس زمینه)

جدول ۲: ساختار بلوک کدگذار در معماری SegRBM-Net

ابعاد خروجی	گام	ابعاد فیلتر	لایه کانولوشنی	زیربلوک	-
$m_{input} \times n_{input} \times 1$	-	-	-	-	تصویر ورودی
$m_1 \times n_1 \times 64$	۱	$3 \times 3 / 64$	Conv1+BN+ReLU	۱	بلوک کدگذار
$m_2 \times n_2 \times 64$	۱	$3 \times 3 / 64$	Conv2+BN+ReLU+GBRBM		
-	۲	2×2	Max_Pooling		
$m_3 \times n_3 \times 128$	۱	$3 \times 3 / 128$	Conv1+BN+ReLU	۲	
$m_4 \times n_4 \times 128$	۱	$3 \times 3 / 128$	Conv2+BN+ReLU+GBRBM		
-	۲	2×2	Max_Pooling		
$m_5 \times n_5 \times 256$	۱	$3 \times 3 / 256$	Conv1+BN+ReLU	۳	
$m_6 \times n_6 \times 256$	۱	$3 \times 3 / 256$	Conv2+BN+ReLU		
$m_7 \times n_7 \times 256$	۱	$3 \times 3 / 256$	Conv3+BN+ReLU+GBRBM		
-	۲	2×2	Max_Pooling	۴	
$m_8 \times n_8 \times 512$	۱	$3 \times 3 / 512$	Conv1+BN+ReLU		
$m_9 \times n_9 \times 512$	۱	$3 \times 3 / 512$	Conv2+BN+ReLU		
$m_{10} \times n_{10} \times 512$	۱	$3 \times 3 / 512$	Conv3+BN+ReLU+GBRBM		
-	۲	2×2	Max_Pooling		
$m_{11} \times n_{11} \times 1024$	۱	$3 \times 3 / 1024$	Conv1+BN+ReLU	۵	
$m_{12} \times n_{12} \times 1024$	۱	$3 \times 3 / 1024$	Conv2+BN+ReLU		
$m_{13} \times n_{13} \times 1024$	۱	$3 \times 3 / 1024$	Conv3+BN+ReLU+GBRBM		
-	۲	2×2	Max_Pooling		

واحدهای پردازش گرافیکی (GeForce GTX 1080 Ti) و مرکزی (Intel core i7-3770 CPU) و سیستم عامل ویندوز ۷ صورت گرفته است. شکل (۱۰)، نتیجه آزمایش الگوریتم پیشنهادی تحقیق بر روی چهار نمونه از هر مجموعه داده با استفاده از روش پیشنهادی تحقیق را نشان می دهد. برای انتخاب این نمونه ها تنوع در جهت قرار گرفتن وسایل نقلیه، ابعاد و همچنین تفاوت در میزان انرژی تابشی وسایل نقلیه مورد نظر بوده است.

برای طراحی و پیاده سازی مدل پیشنهادی موسوم به SegRBM-Net از بستر کد نویسی متلب^۱ ۲۰۱۷ و با بهره گیری از جعبه ابزار مت-کانونت^۲ که به طراحی مناسب تر شبکه کمک می کند استفاده شده است. جهت ارتقاء نقشه واقعیت زمینی با استفاده از ابزار برچسب دهی به تصاویر در نرم افزار متلب مرزهای وسایل نقلیه از روی تصاویر استخراج و ذخیره شد. همچنین جهت اصلاح وزن های شبکه پیشنهادی از الگوریتم بهینه سازی گرادیان نزولی تصادفی^۳ جهت بهبود همگرایی الگوریتم استفاده شده است. پیکره بندی سخت افزار مورد استفاده بر اساس پردازش موازی بر

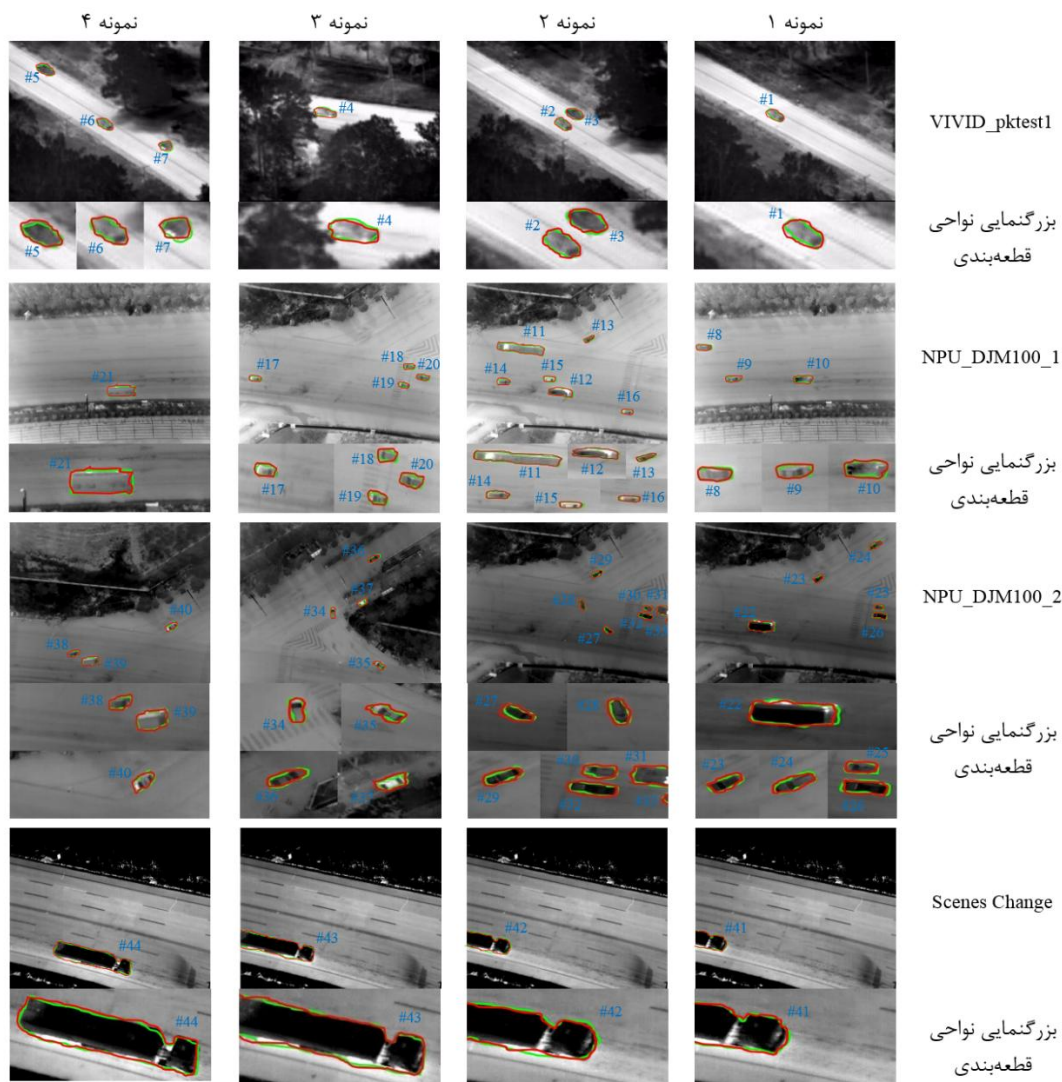
^۱ MATLAB

^۲ MatConvNet

^۳ Stochastic Gradient Descent (SGD)

جدول ۳: ساختار بلوک کدگشا در معماری SegRBM-Net

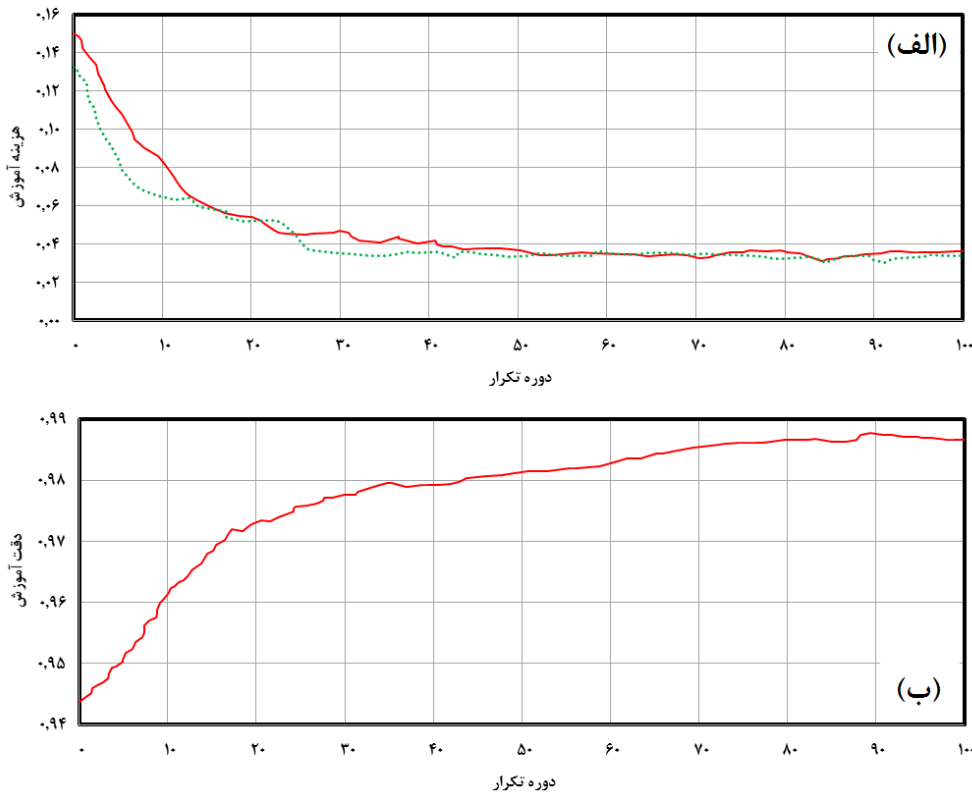
ابعاد خروجی	گام	ابعاد فیلتر	لایه کانولوشنی	زیر بلوک	-
-	۲	۲×۲	Upsampling	۱	بلوک کدگشا
$m_1 \times n_1 \times 1024$	۱	۳×۳/۱۰۲۴	Deconv1+BN+ReLU+ GBRBM		
$m_2 \times n_2 \times 1024$	۱	۳×۳/۱۰۲۴	Deconv2+BN+ReLU		
$m_3 \times n_3 \times 512$	۱	۳×۳/۵۱۲	Deconv3+BN+ReLU	۲	
-	۲	۲×۲	Upsampling		
$m_4 \times n_4 \times 512$	۱	۳×۳/۵۱۲	Deconv1+BN+ReLU+ GBRBM		
$m_5 \times n_5 \times 512$	۱	۳×۳/۵۱۲	Deconv2+BN+ReLU	۳	
$m_6 \times n_6 \times 256$	۱	۳×۳/۲۵۶	Deconv3+BN+ReLU		
-	۲	۲×۲	Upsampling		
$m_7 \times n_7 \times 256$	۱	۳×۳/۲۵۶	Deconv1+BN+ReLU+ GBRBM	۴	
$m_8 \times n_8 \times 256$	۱	۳×۳/۲۵۶	Deconv2+BN+ReLU		
$m_9 \times n_9 \times 128$	۱	۳×۳/۱۲۸	Deconv3+BN+ReLU		
-	۲	۲×۲	Upsampling	۵	
$m_{10} \times n_{10} \times 128$	۱	۳×۳/۱۲۸	Deconv1+BN+ReLU+ GBRBM		
$m_{11} \times n_{11} \times 64$	۱	۳×۳/۶۴	Deconv2+BN+ReLU		
-	۲	۲×۲	Upsampling	۵	
$m_{12} \times n_{12} \times 64$	۱	۳×۳/۶۴	Deconv1+BN+ReLU+ GBRBM		
$m_{13} \times n_{13} \times 2$	۱	۳×۳/۲	Deconv2+BN+ReLU		
-	-	-	Softmax+Loss	-	خروجی شبکه
$m_{Output} \times n_{Output} \times 1$	-	-	-	-	-



شکل ۱۰: نمونه‌های آزمایشی قطعه‌بندی وسایل نقلیه زمینی با استفاده از روش پیشنهادی بر روی مجموعه تصاویر مادون قرمز حرارتی پهپاد. سبز: نقشه واقعیت زمینی، قرمز: روش پیشنهادی

۱۰۰ مرتبه تکرار نشان می‌دهد. بهترین مدل آموزش‌دیده در مرتبه ۹۰ تکرار به دست آمده است. معیار این انتخاب بر اساس منحنی تغییرات اعتبارسنجی تعیین می‌گردد.

شکل (۱۱)، منحنی تغییرات هزینه آموزش، اعتبارسنجی و دقت آموزش مدل پیشنهادی برای قطعه‌بندی وسایل نقلیه زمینی با استفاده از مجموعه داده آموزشی تصاویر مادون قرمز حرارتی پهپاد را در



شکل ۱۱: (الف) منحنی تغییرات هزینه آموزش (منحنی قرمز) و اعتبارسنجی (منحنی سبز). (ب) تغییرات دقت آموزش

۳- تجزیه و تحلیل نتایج

برای ارزیابی نتایج تحقیق و بررسی عملکرد الگوریتم *SegRBM-Net* از دو الگوریتم کارآمد یادگیری عمیق با نام سگنت و رز-یونت و یک روش متداول بر اساس ماشین بردار پشتیبان به نام *HOG+SVM* [۱۸] جهت مقایسه نتایج استفاده شده است. شکل (۱۲)، چهار نمونه از قطعه‌بندی وسایل نقلیه را بر اساس معماری‌های معرفی شده نشان می‌دهد. همچنین در شکل (۱۳) نتایج عددی ارزیابی دقت شناسایی و قطعه‌بندی به صورت نمودار میله‌ای برای این نمونه‌ها ارائه شده است. بر همین اساس از معیارهای دقت (رابطه ۹)، یادآوری (رابطه ۱۰) و نمره *FI* (رابطه ۱۱) برای ارزیابی قدرت شناسایی قطعه‌بندی وسایل نقلیه استفاده شده است. همچنین با توجه به اینکه مجموعه داده آموزشی به چهار دسته طبقه‌بندی می‌شوند، زمان پردازش نیز بر حسب ثانیه و به تفکیک الگوریتم‌های

مورد استفاده و مجموعه تصاویر مادون قرمز حرارتی مورد مقایسه قرار گرفته‌اند. برای ارزیابی کلی قطعه‌بندی از ضریب همبستگی متیو^۱ (رابطه ۱۲) و شاخص تشابه ژاکارد^۲ (رابطه ۱۳) استفاده شده است [۶]. در روابط ۹ تا ۱۳، *TP* تعداد پیکسل‌های هدف در کلاس هدف، *FP* تعداد پیکسل‌های غیر هدف در کلاس هدف، *FN* تعداد پیکسل‌های هدف در کلاس غیر هدف و *TN* تعداد پیکسل‌های غیر هدف در کلاس غیر هدف قرار دارند و مقدار بهینه برای روابط (۹) تا (۱۳) برابر ۱ می‌باشد.

^۱ Matthews Correlation Coefficient (MCC)

^۲ Jaccard Similarity Index (JSI)

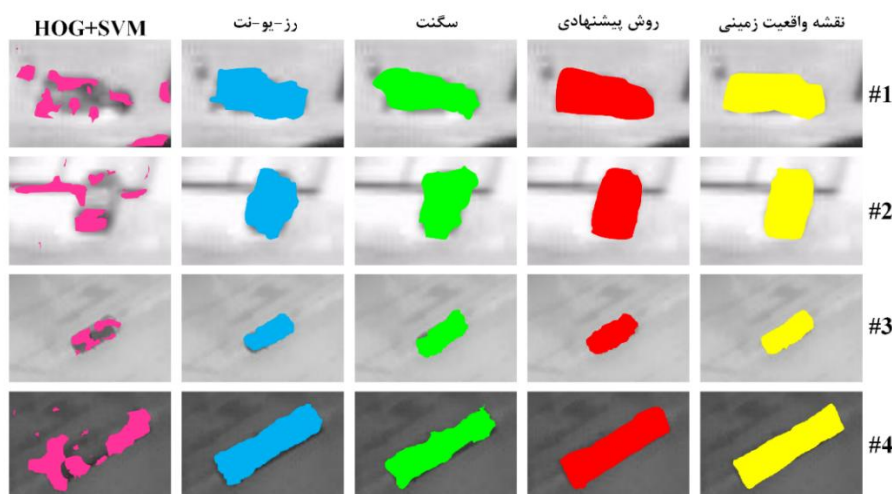
$$Precision = \frac{TP}{TP+FP} \quad \text{رابطه (۹)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{رابطه (۱۰)}$$

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad \text{رابطه (۱۱)}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \text{رابطه (۱۲)}$$

$$JSI = \frac{TP}{(TP+FN+FP)} \quad \text{رابطه (۱۳)}$$



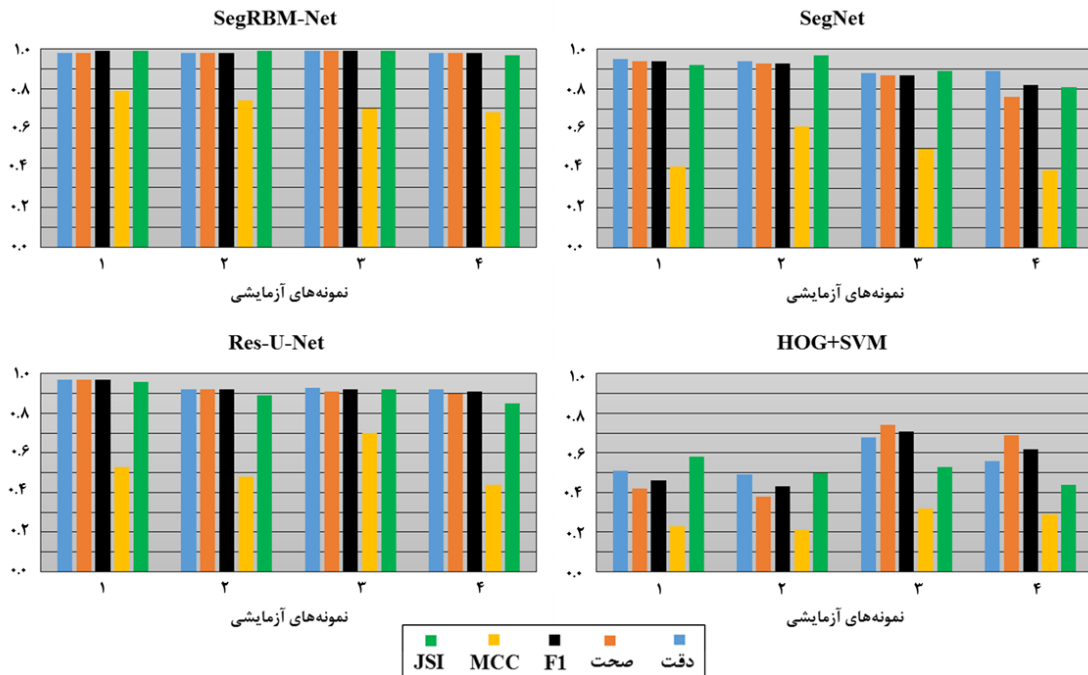
شکل ۱۲: مقایسه روش پیشنهادی تحقیق با الگوریتم‌های مشابه. زرد: نقشه واقعیت زمینی، قرمز: روش پیشنهادی، سبز: روش سگنت، آبی: روش رز-یو-نت، صورتی: روش *HOG+SVM*

به توضیح است که تمامی معماری‌های بررسی شده دارای نرخ همگرایی نسبتاً مناسبی هستند، اما معماری سگنت به دلیل عدم استفاده از لایه‌های بازگشتی دارای سرعت آموزش مطلوب‌تری می‌باشد. به‌علاوه نزدیک بودن نتایج عددی بیانگر این موضوع است که الگوریتم‌های یادگیری عمیق دارای قدرت یادگیری بازنمایی از داده‌های آموزشی بوده و می‌توانند نتایجی با دقت‌های مناسب نسبت به استفاده از روش‌هایی مرسوم مانند *HOG+SVM* که فرآیند استخراج ویژگی به‌صورت قوی صورت نمی‌گیرد، ارائه دهند.

برای بررسی الگوریتم پیشنهادی در مقایسه با الگوریتم‌های سگنت، رز-یو-نت و *HOG+SVM*، نتایج ارزیابی به تفکیک تمام داده‌های آزمایشی در جدول (۴) ارائه شده است. بر اساس جدول (۴) دو نتیجه مهم حاصل می‌شود که عبارت‌اند از:

(۱) الگوریتم *SegRBM-Net* به لحاظ قطعه‌بندی وسایل نقلیه زمینی از تصاویر مادون‌قرمز حرارتی پهن‌بند عملکرد مطلوب‌تری نسبت به سه روش سگنت، رز-یو-نت و *HOG+SVM* داشته است.

(۲) مسأله سرعت پردازش در الگوریتم پیشنهادی است که در این مورد نیز روش *SegRBM-Net* دارای سرعت پردازش مناسبی در مقایسه با سایر روش‌ها است. لازم



شکل ۱۳: عملکرد مدل پیشنهادی SegRBM-Net در مقایسه با سایر روش‌های مشابه بر روی چهار نمونه آزمایشی

جدول ۴: ارزیابی عملکرد معماری پیشنهادی با روش‌های مشابه بر روی مجموعه داده‌های آزمایشی. بهترین نتایج به صورت پررنگ در جدول نشان داده شده‌اند. (الف) مدل SegRBM-Net، (ب) مدل SegNet، (ج) مدل Res-U-Net و (د) مدل HOG+SVM

مجموعه	روش	دقت	صحت	F1	زمان (ثانیه)
VIVID_pktst1	الف	۰٫۹۹۶	۰٫۹۹۸	۰٫۹۹۷	۶٫۳۱
	ب	۰٫۹۹۴	۰٫۹۹۶	۰٫۹۹۵	۶٫۳۷
	ج	۰٫۹۹۵	۰٫۹۹۸	۰٫۹۹۷	۱۶٫۵۵
	د	۰٫۵۲۹	۰٫۶۵۳	۰٫۵۸۴	۱۲۸٫۲۱
NPU_DJM100_1	الف	۰٫۹۹۵	۰٫۹۹۸	۰٫۹۹۶	۲۶٫۱۸
	ب	۰٫۹۹۲	۰٫۹۹۷	۰٫۹۹۴	۲۷٫۰۵
	ج	۰٫۹۹۲	۰٫۹۹۷	۰٫۹۹۵	۴۲٫۱۱
	د	۰٫۵۷۵	۰٫۵۹۷	۰٫۵۸۶	۱۵۴٫۸۸
NPU_DJM100_2	الف	۰٫۹۸۹	۰٫۹۹۷	۰٫۹۹۳	۲۱٫۴۹
	ب	۰٫۹۸۵	۰٫۹۹۵	۰٫۹۹۰	۲۴٫۳۷
	ج	۰٫۹۸۶	۰٫۹۹۵	۰٫۹۹۰	۵۲٫۸۴
	د	۰٫۵۵۹	۰٫۹۵۳	۰٫۷۰۵	۳۸۲٫۹۱
Scenes Change	الف	۰٫۹۷۷	۰٫۹۹۱	۰٫۹۸۴	۱۱٫۲۶
	ب	۰٫۹۶۶	۰٫۹۸۳	۰٫۹۷۵	۱۳٫۳۰
	ج	۰٫۹۶۴	۰٫۹۸۴	۰٫۹۷۴	۳۲٫۰۸
	د	۰٫۶۰۰	۰٫۹۲۸	۰٫۷۲۹	۲۴۴٫۵۴

۴- نتیجه‌گیری

تحقیق حاضر با هدف پیشنهاد یک الگوریتم جدید بر اساس معماری‌های قطعه‌بندی تصویر در یادگیری عمیق و بهبود فرآیند آموزش با استفاده از ماشین بولتزمان محدود گوسین-برنولی بر روی مجموعه تصاویر مادون‌قرمز حرارتی پهباد با تمرکز بر روی قطعه‌بندی وسایل نقلیه زمینی صورت گرفته است. ایده اصلی تحقیق به‌کارگیری هدفمند یادگیری عمیق در بحث قطعه‌بندی وسایل نقلیه از تصاویر حرارتی با قدرت تفکیک مکانی بالا می‌باشد. در همین راستا از جمله چالش‌های به‌کارگیری الگوریتم‌های یادگیری عمیق، مسأله داده‌های کلان برای آموزش الگوریتم مطرح است. در این مقاله، برای بهبود فرآیند آموزش در حل مسأله قطعه‌بندی تصاویر حرارتی، یک ساختار جدید یادگیری عمیق مبتنی بر لایه‌های کانولوشنی و ماشین بولتزمان محدود (برای تقویت یادگیری هندسه اشیا تصویری) ارائه شد.

نتایج این تحقیق در سه دسته اصلی قابل ارائه هستند. (۱) طراحی یک معماری با قابلیت قطعه‌بندی با دقت و سرعت پردازش مناسب سهم بسیار زیادی در علوم وابسته به پردازش تصویر دارد که در این تحقیق با بررسی یک کاربرد مهم یعنی قطعه‌بندی وسایل نقلیه زمینی، الگوریتم پیشنهادی مورد آزمایش قرار گرفت. (۲) تصاویر مادون‌قرمز حرارتی با قدرت تفکیک مکانی بالا از جمله داده‌های بسیار با اهمیتی هستند که تحقیقات چندانی در زمینه پردازش این تصاویر صورت نگرفته است، به همین علت در این تحقیق استفاده از این داده‌ها مورد نظر قرار گرفته است.

مراجع

- [1] M. Gähler, "Remote sensing for natural or man-made disasters and environmental changes," *Environ. Appl. Remote Sens.*, Jun. 2016.
- [2] J. R. B. Fisher, E. A. Acosta, P. J. Dennedy-Frank, T. Kroeger, and T. M. Boucher, "Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality," *Remote Sens. Ecol. Conserv.*, vol. 4, no. 2, pp. 137-149, 2017.

(۳) ارائه راه‌حلی جدید برای افزایش کارایی الگوریتم‌های یادگیری عمیق در پردازش تصاویر حرارتی با استفاده از ماشین بولتزمان محدود گوسین-برنولی، یکی دیگر از دستاوردهای این تحقیق است.

در این تحقیق با استفاده از قابلیت‌های ماشین بولتزمان محدود گوسین-برنولی در یک چارچوب شبکه عصبی کانولوشنی عمیق که آن را با نام *SegRBM-Net* معرفی کردیم، تلاش شد تا برای یک فرآیند قطعه‌بندی وسایل نقلیه زمینی، راهکاری توانمند ارائه شود که می‌توان برای سایر منابع داده و با کاربردهای متنوع مورد استفاده قرار داد. اما باید به این نکته توجه کرد که مدل *SegRBM-Net* دارای یک محدودیت مهم می‌باشد. این محدودیت در انتخاب داده مناسب به عنوان ورودی مدل *SegRBM-Net* است. با توجه به اینکه در این معماری از بلوک‌های باقی‌مانده که مشتمل بر ماشین بولتزمان محدود می‌باشد؛ استفاده شده است و با توجه به اینکه استفاده از ماشین بولتزمان محدود نیز دارای هزینه محاسبات است، به‌کارگیری تصاویر چند طیفی یا فراطیفی موجب می‌شود تا ابعاد بردار ویژگی افزایش یابد و موجب کاهش عملکرد مدل در فرآیند آموزش شود. به همین علت پیشنهاد می‌شود در پژوهش‌های آتی به مسأله استفاده از ماشین بولتزمان محدود در بستر مدل پیشنهادی این تحقیق جهت کاهش ابعاد تصاویر چندطیفی یا فراطیفی و همچنین پردازش مجموعه‌های بزرگ تصاویر حرارتی با قدرت تفکیک مکانی بالا با تمرکز بر قطعه‌بندی سایر عناصر شهری از جمله ساختمان‌ها به دلیل اهمیت ویژه این عنصر در شهرها پرداخته شود.

- [3] A. Nassar, K. Amer, R. ElHakim, and M. ElHelw, "A deep CNN-based framework for enhanced aerial imagery registration with applications to UAV geolocalization," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2018.
- [4] N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan, and M. Zuair, "Deep learning approach for car detection in UAV imagery," *Remote Sens.*, vol. 9, no. 4, p. 312, Mar. 2017.
- [5] G. V. Konoplich, E. O. Putin, and A. A. Filchenkov, "Application of deep learning to the problem of vehicle detection in UAV images," in *2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM)*, St. Petersburg, Russia, 2016, pp. 4–6.
- [6] L. Mou and X. X. Zhu, "Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network," *IEEE Trans. Geosci. and Remote Sens.*, vol. 56, no. 11, pp. 6699–6711, Nov. 2018.
- [7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [8] J. Lee, J. Wang, D. Crandall, S. Sabanovic, and G. Fox, "Real-time, cloud-based object detection for unmanned aerial vehicles," in *2017 First IEEE International Conference on Robotic Computing (IRC)*, Taichung, Taiwan, 2017, pp. 36–43.
- [9] D. Wu, Y. Zhang, Y. Chen, and S. Zhong, "Vehicle detection in high-resolution images using superpixel segmentation and CNN iteration strategy," *IEEE Geosci. and Remote Sens. Letters*, vol. 16, no. 1, pp. 105–109, Jan. 2019.
- [10] B. Kellenberger, D. Marcos, and D. Tuia, "Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning," *Remote Sens. Environ.*, vol. 216, pp. 139–153, Oct. 2018.
- [11] J. E. Ball, D. T. Anderson, and C. S. Chan, "A comprehensive survey of deep learning in remote sensing: theories, tools and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 04, p. 1, Sep. 2017.
- [12] X. Liu, T. Yang, and J. Li, "Real-Time Ground vehicle detection in aerial infrared imagery based on convolutional neural network," *Electronics*, vol. 7, no. 6, p. 78, May 2018. Available online: <https://pan.baidu.com/s/1d0BMIwp1EUF C4eievhInmw#list/path=%2F>
- [13] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.
- [15] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," *2002 14th Int. Conf. Digit. Signal Process. Proc. DSP 2002 Cat No02TH8628*.
- [16] Luo-Wei Tsai, L.-W. Tsai, J.-W. Hsieh, and K.-C. Fan, "Vehicle detection using normalized color and edge map," *IEEE Int. Conf. Image Process.* 2005, 2005.
- [17] X. Chen and Q. Meng, "Vehicle detection from UAVs by using SIFT with implicit shape model," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 3139–3144.

- [18] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on Viola-Jones and HOG+SVM from uav images," *Sensors*, vol. 16, no. 8, p. 1325, 2016.
- [19] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arab. J. Sci. Eng.*, vol. 41, no. 3, pp. 1043–1051, 2015.
- [20] W. Shao, W. Yang, G. Liu, and J. Liu, "Car detection from high-resolution aerial imagery using multiple features," *2012 IEEE Int. Geosci. Remote Sens. Symp.*, 2012.
- [21] Z. Chen, C. Wang, H. Luo, H. Wang, Y. Chen, C. Wen, Y. Yu, L. Cao, and J. Li "Vehicle detection in high-resolution aerial images based on fast sparse representation classification and multiorder feature," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2296–2309, 2016.
- [22] S. L. Yu, T. Westfechtel, R. Hamada, K. Ohno, and S. Tadokoro, "Vehicle detection and localization on bird's eye view elevation images using convolutional neural network," *2017 IEEE Int. Symp. Saf. Secur. Rescue Robot. SSR*, 2017.
- [23] K. Bittner, S. Cui, and P. Reinartz, "Building extraction from remote sensing data using fully convolutional networks," *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-1/W1, pp. 481–486, May 2017.
- [24] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.
- [25] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sens.*, vol. 9, no. 5, p. 498, May 2017.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*, 2016.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 9351, pp. 234–241, 2015.
- [28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [29] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [30] F. Chollet, *Deep learning with Python*. Shelter Island, New York: Manning Publications Co, 2018.
- [31] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.
- [32] K. Chen, K. Fu, X. Gao, M. Yan, X. Sun, and H. Zhang, "Building extraction from remote sensing images with deep learning in a supervised manner," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 2017, pp. 1672–1675.
- [33] M. K. Masouleh and R. Shah-Hosseini, "Fusion of deep learning with adaptive bilateral filter for building outline extraction from remote sensing imagery," *J. Appl. Remote Sens.*, vol. 12, no. 04, p. 1, Nov. 2018.

- [34] N. Audebert, A. Boulch, H. Randrianarivo, B. Le Saux, M. Ferecatu, S. Lefevre, and R. Marlet, "Deep learning for urban remote sensing," in *2017 Joint Urban Remote Sensing Event (JURSE)*, Dubai, United Arab Emirates, 2017, pp. 1–4.
- [35] R. Salakhutdinov and H. Larochelle, "Efficient learning of deep Boltzmann machines," in *Proceeding of the Thirteenth International Conference on Artificial Intelligence and Statistics*, p. 8, 2010.
- [36] Y. Sakai and K. Yamanishi, "Data fusion using Restricted Boltzmann machines," *2014 IEEE Int. Conf. Data Min.*, 2014.
- [37] S. Keronen, K. Cho, T. Raiko, A. Ilin, and K. Palomaki, "Gaussian-Bernoulli restricted Boltzmann machines and automatic feature extraction for noise robust missing data mask estimation," *2013 IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013.
- [38] Z. Li, X. Cai, and T. Liang, "Gaussian-Bernoulli based convolutional restricted Boltzmann machine for images feature extraction," *Lecture Notes in Computer Science*, pp. 593–602, 2016.
- [39] H. Zhang, S. Zhang, K. Li, and D. N. Metaxas, "Robust shape prior modeling based on Gaussian-Bernoulli restricted Boltzmann Machine," *2014 IEEE 11th Int. Symp. Biomed. Imaging ISBI*, 2014.
- [40] Y. Qiao, Z. Wei, and Y. Zhao, "Thermal Infrared Pedestrian Image Segmentation Using Level Set Method," *Sensors*, vol. 17, no. 8, p. 1811, Aug. 2017.
- [41] "Tau 2 Longwave Infrared Thermal Camera Core | FLIR Systems." [Online]. Available: <https://www.flir.com/products/tau-2/>. [Accessed: 05-Dec-2018].
- [42] M. S. Jadin and S. Taib, "Infrared image enhancement and segmentation for extracting the thermal anomalies in electrical equipment," *Electron. Electr. Eng.*, vol. 120, no. 4, Apr. 2012.
- [43] VE_UAV_TIR_Data. Available online: <https://1drv.ms/u/s!Aj2qslPUOabbVHUH9J4E0s1yFE>



A Hybrid Algorithm based on Deep Learning and Restricted Boltzmann Machine for Car Semantic Segmentation from Unmanned Aerial Vehicles (UAVs)-based Thermal Infrared Images

Mahdi Khoshboresh Masouleh ¹, Reza Shah Hosseini ^{2*}

1- PhD Student., School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

2- Assistant Professor., School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

Abstract

Nowadays, ground vehicle monitoring (GVM) is one of the areas of application in the intelligent traffic control system using image processing methods. In this context, the use of unmanned aerial vehicles based on thermal infrared (UAV-TIR) images is one of the optimal options for GVM due to the suitable spatial resolution, cost-effective and low volume of images. The methods that have been proposed for vehicle extraction from thermal infrared imaging often experience problems such as low accuracy in detection, segmentation (e.g. HOG+SVM) and also the need for big data training (e.g. deep learning methods). In the present study, a new model, called SegRBM-Net, based on deep learning (DL) and the restricted Boltzmann machine (RBM) is being presented. One of the features of the SegRBM-Net model is the improving accuracy of vehicle detection and segmentation from thermal infrared images by using both convolutional layers and the features of the Gaussian-Bernoulli restricted Boltzmann machine. This structure has led the algorithm to find the target faster and more accurately than other DL methods. To examine the performance of the proposed method, we performed a controlled benchmark (e.g. high density of vehicles scene, and difference in viewing angle) of SegRBM-Net and other DL models on four UAV-TIR image datasets. The results showed that the SegRBM-Net model with a mean accuracy of 99% and improved processing speed compared with similar methods have a good performance.

Key words: Convolutional Neural Networks, Gaussian-Bernoulli Restricted Boltzmann Machine, Semantic Segmentation, Ground Vehicle, Thermal Infrared UAV Imagery

Correspondence Address : Photogrammetry & Remote Sensing Group, School of Surveying & Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.

Tel: +98 21 61114527.

Email: rshahosseini@ut.ac.ir