

## ارائه یک الگوریتم ترکیبی مبتنی بر یادگیری عمیق و ماشین بولتزمان محدود با هدف قطعه‌بندی معنایی خودرو از تصاویر مادون قرمز حرارتی پهپاد

مهدي خوش برش ماسوله<sup>۱</sup>، رضا شاه حسیني<sup>\*۲</sup>

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

### چکیده

امروزه پایش وسائل نقلیه زمینی با استفاده از روش‌های پردازش تصویر، یکی از حیطه‌های کاربردی در کنترل ترافیک هوشمند به شمار می‌آید. در این زمینه، به کارگیری تصاویر مادون قرمز حرارتی پهپاد به دلیل قدرت تفکیک مکانی مناسب، مقرن به صرفه بودن و حجم کمتر تصاویر، یکی از گزینه‌های مطلوب برای هدف پایش وسائل نقلیه است. روش‌هایی که تا حال برای استخراج وسائل 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 وسائل 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 مادون قرمز حرارتی پهپاد با ویژگی‌های نظری تراکم بالای وسائل 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.

کلید واژه‌ها : شبکه عصبی کانولوشنی عمیق، ماشین بولتزمان محدود گوسین - برنولی، قطعه‌بندی معنایی، وسائل نقلیه زمینی، تصاویر مادون قرمز حرارتی پهپاد.

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

تلفن: ۶۱۱۱۴۵۲۷

## ۱- مقدمه

و با سکوهای زمینی می‌باشد. مهم‌ترین مشکل در این روش‌ها، پوشش محدود این دوربین‌ها و کاهش قدرت کنترل در ابعاد مناسب است. به همین جهت، به کارگیری تصاویر پهپاد یک راه حل 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 چهارگوش 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 قطعه‌بندی معنایی تصویر جستجو کرد. در قطعه‌بندی معنایی تصویر، علاوه بر مشخص نمودن موقعیت شی، می‌توان محدوده آن شی را نیز مشخص کرد. تعیین محدوده شی به صورت 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<sup>۱</sup> [۱۴] (شکل ۱-ب) در مقایسه با قطعه‌بندی معنایی وسایل نقلیه با استفاده از معماری یادگیری عمیق سگنت<sup>۲</sup> [۷] (شکل ۱-ج) نشان داده شده است. بر اساس این شکل می‌توان قابلیت‌های قطعه‌بندی معنایی را بهتر درک کرد، به ویژه در مواردی که تراکم خودروها زیاد است (بردار مشکی) و صرفاً آشکارسازی آن‌ها با یک جعبه آشکارساز نمی‌تواند راه حل مناسبی باشد.

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

ارائه یک راهبرد هدفمند و علمی برای پایش عناصر طبیعی و مصنوعی در محیط‌بیست، همواره یکی از موضوعات مهم در برنامه‌ریزی‌های مدیران تصمیم‌گیرنده در کشورها محسوب می‌شود [۱۱]. بسیاری از تحقیق‌های صورت گرفته، از منابع داده سنجش‌ازدor 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 راه حل 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].

مسأله قطعه‌بندی معنایی وسایل نقلیه زمینی از تصاویر پهپاد در کاربردهایی نظیر پایش حجم ترافیک و تجزیه‌وتحلیل تصادفاتی 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 قطعه‌بندی معنایی تصاویر با هدف کشف وسایل نقلیه مبتنی بر تصویربرداری از روی زمین

<sup>۱</sup> You Only Look Once (YOLO)

<sup>۲</sup> SegNet

<sup>۳</sup> Gabor Filter

<sup>۴</sup> Support Vector Machine (SVM)

می آید [۱۵].

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



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

کانون‌لوشنی<sup>۵</sup> [۱۲] نیز از جمله راه حل‌های پیشنهادی است.

به طور کلی، روش‌های پیشنهادی برای آشکارسازی و قطعه‌بندی معنایی وسائل 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 ویژگی، ویژگی‌های تسریع شده مقاوم<sup>۶</sup>، هیستوگرام گرادیان شبیگرا، لبه تصاویر) و استفاده ترکیبی از این تبدیل‌ها است [۲۰]. دسته دوم، به کارگیری 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 توسعه یافته 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 موردنظر است [۲۳]. اغلب روش‌های یادگیری ماشین برای پردازش کلان داده‌ها

استفاده از تصاویر رنگی و استخراج لبه اشیاء تصویری (با روش‌های متداول لبه یابی) راه حلی دیگر برای آشکارسازی وسائل نقلیه است [۱۶]. از جمله تحقیق‌های صورت گرفته در حیطه آشکارسازی و قطعه‌بندی معنایی وسائل نقلیه از تصاویر رنگی پهپاد می‌توان به مواردی چون تلفیق تبدیل مستقل از مقیاس ویژگی<sup>۷</sup> به دلیل عدم تأثیرپذیری به مقیاس، دوران و شرایط نوردهی با مدل شکل ضمنی<sup>۸</sup> برای تولید توصیفگرهای ویژگی و طبقه‌بندی ویژگی‌ها با ماشین بردار پشتیبان اشاره کرد [۱۷]. الگوریتم دیگر، از روش ترکیبی ویولا-جونز<sup>۹</sup> و طبقه‌بندی کننده ماشین بردار پشتیبان با ویژگی‌های هیستوگرام گرادیان شبیگرا<sup>۱۰</sup> 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 به کارگیری تصاویر مادون‌قرمز حرارتی پهپاد با هدف آشکارسازی بی‌درنگ جعبه‌ای وسائل نقلیه با استفاده از شبکه‌های عصبی

<sup>۱</sup> Scalar Invariant Feature Transform (SIFT)

<sup>۲</sup> Implicit Shape Model (ISM)

<sup>۳</sup> Viola-Jones (V-J)

<sup>۴</sup> Histogram of Oriented Gradient (HOG)

<sup>۵</sup> Convolutional Neural Networks (CNNs)

<sup>۶</sup> 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 وسائل نقلیه صورت می‌گیرد [۱۲].

## ۲- روش پیشنهادی

در این قسمت، ابتدا مروری بر به کارگیری 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 به عنوان یک چالش مهم در حیطه علوم یادگیری ماشین<sup>۵</sup> به حساب می‌آید [۲۹]. منطق ارائه معماری‌های یادگیری عمیق بر افزایش تعداد لایه‌های پنهان در شبکه عصبی مصنوعی بنا شده است. اما به مرور 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 (به عنوان 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 اشیا 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<sup>۱</sup> 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 تصویربردار و تولید مجموعه داده‌های بزرگ، مسئله تفسیر خودکار اشیاء از تصاویر سنجش از دور (ماهواره‌ای، هوایی و پهپاد) با عملکرد مناسب با استفاده از الگوریتم‌های یادگیری عمیق به عنوان یک موضوع چالشی مهم در پردازش تصاویر است. همچنین، استفاده از تصاویر مادون‌قرمز حرارتی پهپاد به دلیل قدرت تفکیک مکانی مناسب، مقرر به صرفه بودن و حجم کمتر تصاویر، یکی از گزینه‌های مطلوب برای هدف پایش وسائل نقلیه نسبت به سایر منابع داده به شمار می‌آید. تحقیقاتی اخیر در پردازش تصاویر نشان می‌دهد، شبکه‌های عصبی عمیق مبتنی بر لایه‌های کانولوشن دارای قابلیت‌های مناسبی در حل مسئله 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 وسائل 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 رزن特<sup>۲</sup> [۲۶]، سگنت<sup>۳</sup> [۲۷] و سگنت با ماشین بولتزمان محدود صورت گرفته است. در این مدل، ماشین بولتزمان محدود در نقش بلوک‌های باقی‌مانده در یک شبکه کدگذار-کدگشای<sup>۴</sup> عمیق ظاهر می‌شود. فرآیند آزمایش و ارزیابی مدل *SegRBM-Net* بر روی مجموعه تصاویر مادون‌قرمز

<sup>۱</sup> Saturation

<sup>۲</sup> ResNet

<sup>۳</sup> U-Net

<sup>۴</sup> Encoder-Decoder

<sup>۵</sup> 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 کنترل شده استفاده کنیم. محاسبات یک درون‌یابی کنترل شده 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. معماري سگنت از نظر حافظه مصرفی، دقت تولیدی و همچنین کاهش پارامترهای شبکه، نسبت به معماري‌های دیگر، بسیار عملکرد مناسبی در مورد قطعه‌بندی تصویر دارد [۷].

۲) معماري رزن特 بر اساس اضافه نمودن یک بردار از پیکسل‌های تصویر ورودی، بر خروجی‌های کدگذاری شده عمل می‌کند.

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

هر لایه کانولوشنی معماري‌های 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<sup>۴</sup> و یک تابع فعال‌سازی می‌باشد. رابطه (۱) تعریف ریاضی یک نورون در معماري‌های 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 ( $j$ ,  $i$ ) 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.

<sup>۴</sup> Batch Normalization (BN)

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

اغلب معماري‌های 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 معماري‌های گوناگون به کارگیری چندین لایه کانولوشنی و استفاده از توابع فعال‌سازی<sup>۱</sup> مختلف است [۱۱]. بهصورت کلی، ارائه تعداد لایه کانولوشنی مناسب با پارامترهای بهینه نیازمند تئوری ریاضی قوی می‌باشد. از میان معماري‌های یادگیری عمیق که برای قطعه‌بندی معنایی تصویر ارائه شده‌اند می‌توان سه معماري تأثیرگذار سگنت، رزن特 و یونت را نام برد. برجسته‌ترین ویژگی‌های این معماري‌ها عبارت است از:

۱) معماري سگنت اولین بار در سال ۲۰۱۵ با آزمایش بر روی یک مجموعه داده از تصاویر دوربین‌های عکاسی ارائه شد. توپولوژی معماري سگنت بر اساس یک شبکه کدگذار با سیزده لایه کانولوشنی در شبکه VGG16 (یک معماري 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<sup>۲</sup> می‌باشد. بهطور متناظر در بلوک کدگشا نیز پنج زیر بلوک با لایه‌های معکوس کانولوشنی و یک لایه نمونه افزایی<sup>۳</sup> وجود دارد. نواوری معماري سگنت در به کارگیری لایه‌های نمونه افزا (بازسازی ابعاد اصلی تصویر) است [۳۳]. این‌گونه پردازش موجب می‌شود تا ابعاد تصویر منبسط شود و ردیف‌ها و ستون‌های زوج با

<sup>۱</sup> Activation Functions

<sup>۲</sup> Pooling

<sup>۳</sup> Upsampling

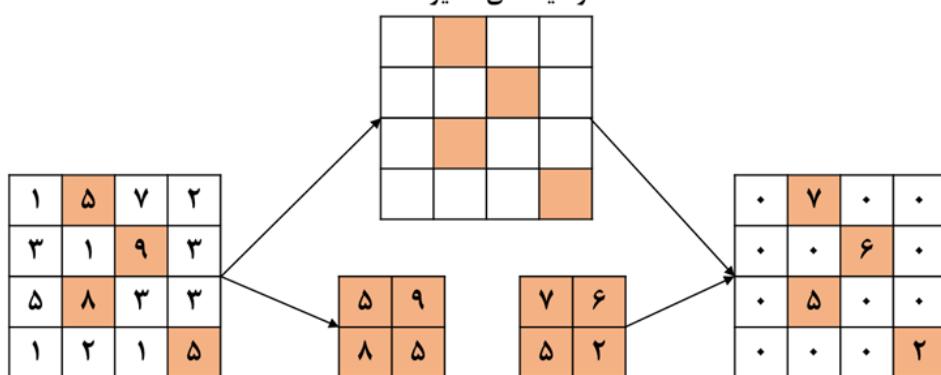
$$(P \times C)_{ij} = \sum_m \sum_n \sum_c P(i+m, j+n) \cdot C(-m, -n) + b$$

تابع ذخیره کردن موقعیت پیکسل با مقدار بیشینه و با همسایگی موردنظر (ابعاد متداول با همسایگی  $2 \times 2$ ) ذخیره و سپس در مرحله نمونه افزایی 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 نمونه افزایی را بر روی یک همسایگی از پیکسل‌ها نمایش می‌دهد.

رابطه (۱)

برای جلوگیری از پدید بیش برازش (*Overfitting*) راه حل‌های مختلفی در شبکه‌های عصبی کانولوشنی عمیق پیشنهاد شده است. یکی از این راه حل‌ها 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 آموخت 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

موقعیت‌های ذخیره شده



مرحله کاهش ابعاد در بلوک کدگذار

مرحله نمونه افزایی در بلوک کدگشای

شکل ۲: فرآیند اعمال لایه کاهش ابعاد در بلوک کدگذار و لایه نمونه افزایی در بلوک کدگشای

(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<sup>۱</sup> 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 مسئله طبقه‌بندی، رگرسیون، کاهش ابعاد و استخراج ویژگی می‌باشد. این ماشین مشتمل بر تعدادی واحد مشاهده‌پذیر<sup>۲</sup> و پنهان<sup>۳</sup> به صورت دودویی

<sup>۱</sup> Hidden Units<sup>۲</sup> Gaussian-Bernoulli Restricted Boltzmann Machine (GB-RBM)<sup>۳</sup> Visible Units

رابطه(۲) تعریف می‌شود [۳۸].

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

احتمالی که مدل یادگیری بولتزمان محدود به واحدهای مشاهده‌پذیر نسبت می‌دهد را می‌توان به صورت رابطه(۵) بیان نمود. در رابطه(۵)،  $\mathcal{N}(\mu, \sigma_i^2)$  تابع گوسین با میانگین  $\mu$  و واریانس  $\sigma_i^2$  است.

به طور مشابه برای واحدهای پنهان، احتمال با استفاده از رابطه(۶) محاسبه می‌شود. لایه‌های پنهان به دلیل عدم اتصال به یکدیگر مستقل از شرایط واحدهای مشاهده‌پذیر محاسبه می‌شوند.

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

در رابطه(۲)،  $v_i$  واحدهای مشاهده‌پذیر با تعداد  $n_v$  واحدهای پنهان با تعداد  $n_h$  مشتمل بر وزن‌های  $w_{ij}$  که به واحدهای مشاهده‌پذیر و پنهان متصل می‌شوند؛  $\sigma_i$  انحراف معیار همبسته به واحد مشاهده‌پذیر گوسین  $c_j$  و  $b_i$  بایاس‌های هریک از واحدهای مشاهده‌پذیر و پنهان است. در این شبکه به هر مقدار از واحدهای مشاهده‌پذیر و پنهان با استفاده از رابطه(۱)، یک مقدار احتمال که به آن توزیع بولتزمان گفته می‌شود به صورت رابطه(۳) نسبت می‌دهد. در رابطه(۳)،  $Z$  ثابت نرمال‌سازی است و با استفاده از رابطه(۴) تعریف می‌شود.

$$P(v, h|\theta) = \frac{1}{Z(\theta)} \exp\{-E(v, h|\theta)\} \quad (3)$$

$$Z = \sum_v \sum_h \exp\{-E(v, h|\theta)\} \quad (4)$$

$$P(v_i = v|h) = \mathcal{N}(v|b_i + \sum_j w_{ij} h_j, \sigma_i^2) \quad (5)$$

$$P(h_j = 1|v) = \text{sigmoid}(c_j + \sum_i w_{ij} \frac{v_i}{\sigma_i^2}) \quad (6)$$

خطا؛  $KL(\rho||\hat{\rho}_k)$  و اگرایی *Kullback-Leibler* که به صورت رابطه(۸) محاسبه می‌شود. 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 مادون‌قرمز حرارتی پهپاد مبتنی بر واحدهای مشاهده‌پذیر و پنهان را نشان می‌دهد.

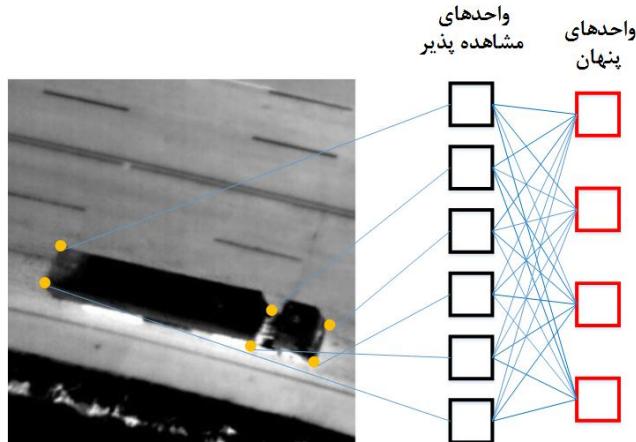
در رابطه(۶)،  $\text{sigmoid}$  نماد تابع سیگموید می‌باشد، این تابع حقیقی، کران‌دار و مشتق‌پذیر می‌باشد و به ازای کلیه مقادیر حقیقی قابل تعریف بوده و دارای مشتق مثبت است. در نهایت برای رسیدن به انرژی کمتر برای مجموعه آموزشی و انرژی بیشتر برای بقیه مجموعه، از مدل پیشنهادی [۳۷] که خطاهای فاحش<sup>۱</sup> را مدل می‌کند، به صورت رابطه(۷) تعریف می‌شود.

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

<sup>۱</sup> 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

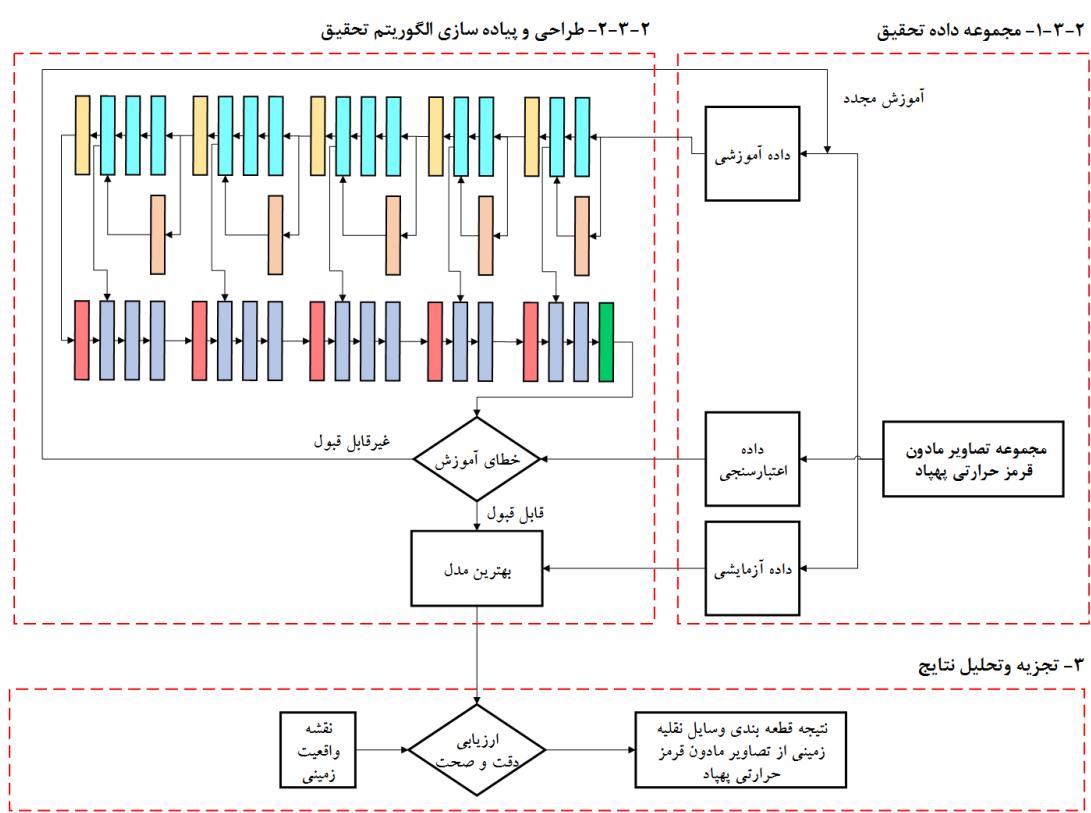
می‌شود. مرحله اول، شامل جمع‌آوری مجموعه تصاویر مادون قرمز حرارتی پهپاد و آماده‌سازی داده‌ها برای آموزش الگوریتم پیشنهادی است (بخش ۱-۳-۲). مرحله دوم، شامل طراحی معماری پیشنهادی تحقیق با نام *SegRBM-Net* مبتنی بر یک ساختار کدگذار-کدگشا که هر زیربلوک کدگذار شامل یک لایه مشتمل بر ماشین بولتزمان محدود گوسین - برنولی با هدف بهبود ویژگی‌های دقیق هندسی از وسایل نقلیه استفاده می‌شود. سپس با نتیجه هر زیربلوک 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 ازتابع انرژی (رابطه ۲) می‌باشد. ویژگی اصلی ماشین بولتزمان محدود گوسین - برنولی در پردازش تصویر، استخراج مدل هندسی شی در تصویر است [۳۹]. این قابلیت موجب می‌شود تا اگر هندسه شی در تصویر ضعیف (مثلًا 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 کانولوشنی 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.

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

(بخش ۳).

گرفته و سپس با به کارگیری معیارهای متداول دقت و صحت از نظر عددی مورد تجزیه و تحلیل قرار می‌گیرد

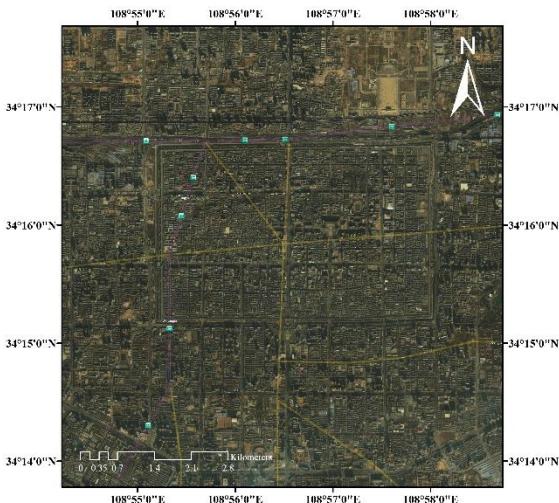


شکل ۴: الگوریتم پیشنهادی تحقیق

اخذ تصاویر پوشش‌دار است. وجود این ویژگی‌ها موجب افزایش اهمیت پردازش تصاویر مادون‌قرمز حرارتی در کاربردهای مختلف و روزمره طی چند سال اخیر شده است. مجموعه داده مورداستفاده در این تحقیق با نام *NPU\_CS\_UAV\_IR\_DATA* شامل یک مجموعه از چهار دسته تصویربرداری مادون‌قرمز حرارتی به وسیله *DJI-MATRICE 100* و با پهپاد *FLIR TAU2* و *Xi'an*<sup>۱</sup> در سال ۲۰۱۷ میلادی می‌باشد. در شکل(۵) موقعیت جغرافیایی منطقه موردمطالعه به همراه نقشه اطلاعات مکانی راه‌های دسترسی، نشان داده شده است.

### ۲-۱-۳- مجموعه داده تحقیق

تصاویر مادون‌قرمز حرارتی اطلاعات مفیدی در مورد توزیع سطوح مختلف انرژی تابشی مادون‌قرمز از سطح جسم قابل اندازه‌گیری از محدوده ۸ تا ۱۵ میکرومتر در طیف الکترومغناطیس را به دست می‌دهد [۴۰]. در مسئله آشکارسازی خودرو استفاده از سنجنده‌های حرارتی این قابلیت را ایجاد می‌کند تا فاصله ۵۰۰ متر وسایل نقلیه در شرایط مختلف نوری و هوایی قابل تشخیص باشند [۴۱]. علاوه قابلیت پایش عوارض در روز و شب مستقل از نور طبیعی یا مصنوعی ایجاد می‌شود [۴۲]. همچنین تصاویر مادون‌قرمز حرارتی پهپاد دارای سه ویژگی مهم است: اول، قدرت تفکیک مکانی بالا. دوم، قدرت تفکیک زمانی بالا و سوم، امکان



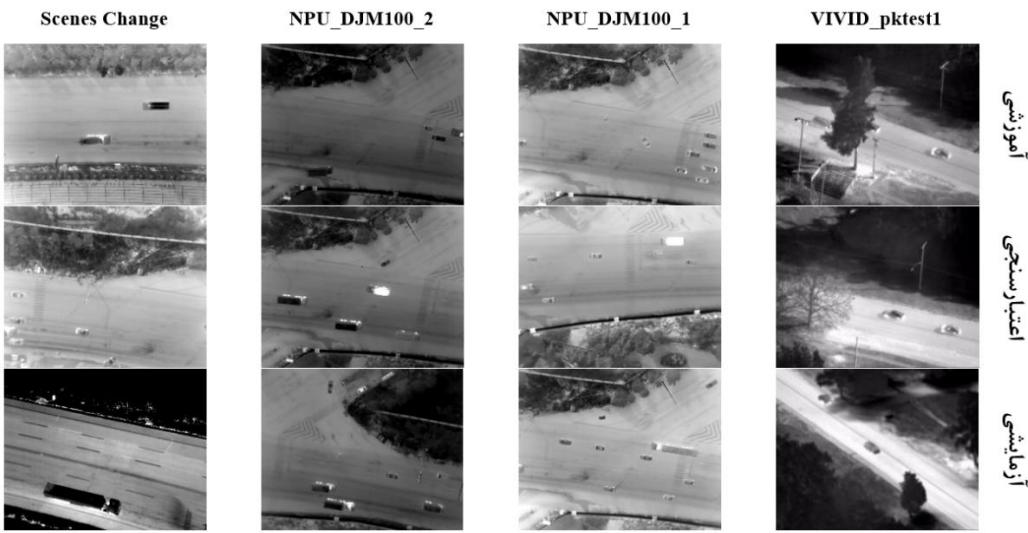
شکل ۵: منطقه مورد مطالعه در این تحقیق (بزرگراه با رنگ بنفش و خیابان‌های اصلی با رنگ زرد در تصویر نشان داده شده است)

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

سنجدنه FLIR TAU2 با وزن کمتر از ۷۰ گرم و با بیشینه حد تفکیک‌پذیری  $640 \times 512$  پیکسل دارای دقیق متوسط  $\pm 5$  درجه سانتی‌گراد می‌باشد، این ویژگی‌ها موجب شده است تا گزینه‌های مناسب برای استفاده از این سنجدنه بر روی پهپاد باشد. جزئیات این

جدول ۱: مشخصات مجموعه داده تحقیق

نام مجموعه داده	تصاویر آموزشی (تعداد فریم)	تصاویر اعتبارسنجی (تعداد فریم)	تصاویر آزمایشی (تعداد فریم)	ابعاد تصاویر بر حسب پیکسل	ارتفاع پرواز (متر)
VIVID_pktestI	۱۷۰	۳۱	۵۰	$320 \times 240$	۸۰
NPU_DJM100_I	۱۵۱	۳۱	۵۰	$640 \times 512$	۱۲۰
NPU_DJM100_2	۴۱۶	۴۰	۵۰	$640 \times 512$	۱۲۰
Scenes Change	۷۰	۱۰	۲۰	$640 \times 512$	۸۰
مجموع	۸۰۷	۱۱۲	۱۷۰		



شکل ۶: نمونه هایی از تصاویر آموزشی، اعتبارسنجی و آزمایشی به تفکیک هوبک از مجموعه تصاویر

بولتzman، می توانیم حاصل مجموع خروجی هریک از زیر بلوک های شبکه عصبی کانولوشنی را با ویژگی هایی که توسط ماشین بولتzman به دست آمده جمع کنیم و سپس نتایج را وارد زیربلوک بعدی کنیم. این کار موجب می شود تا ویژگی های بهینه ای که از هر زیربلوک شبکه عصبی کانولوشنی ایجاد شده است؛ تحت تأثیر ویژگی های به دست آمده از ماشین بولتzman قرار بگیرند تا علاوه بر اطلاعات طیفی، خصوصیات هندسی وسایل نقلیه تقویت گردد؛ به این ترتیب روند آموزش تقویت می شود. برای جمع کردن نتیجه ماشین بولتzman محدود و شبکه عصبی از یک فیلتر کانولوشنی و یک تابع نرمال ساز بسته ها استفاده می کنیم. با این کار ابعاد بردارها را با یکدیگر هماندازه کدگذار در معماری *SegRBM-Net* مشتمل بر لایه های کانولوشنی به ابعاد  $3 \times 3$  و با وزن دهی تصادفی، برای هر لایه کانولوشنی یک تابع نرمال ساز بسته ها جهت جلوگیری از پدیده بیش برازش و یک تابع فعال سازی رلو (این تابع به دلیل اینکه در نواحی مثبت دارای مشتق صفر است هیچ گاه به اشباع نمی رسد) می باشد. شکل (۷) فرآیند تلفیق لایه های کانولوشنی با ویژگی های تولید شده از ماشین بولتzman محدود

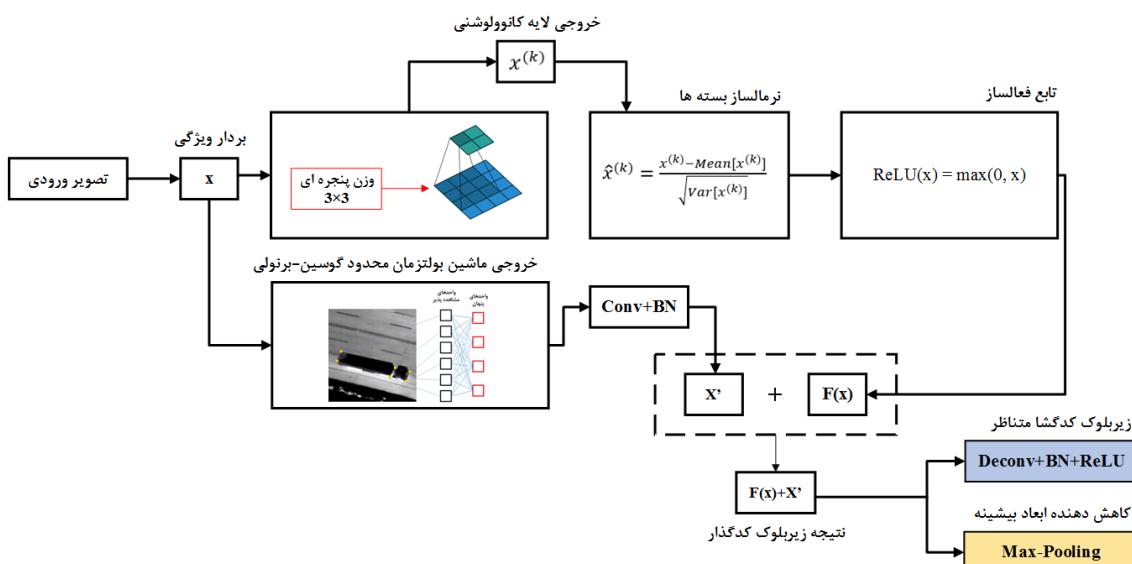
**۲-۳-۲- طراحی و پیاده سازی الگوریتم تحقیق**  
با توجه به اینکه هدف این تحقیق قطعه بندی معنایی وسایل نقلیه از تصاویر مادون قرمز حرارتی پهپاد با یادگیری عمیق است، سه موضوع زیر در طراحی الگوریتم پیشنهادی مورد توجه است:

(۱) مجموعه تصاویر مورد استفاده تک باندی هستند، پس استخراج یک هندسه مناسب از وسایل نقلیه از نمای بالا دارای کاربرد مناسبی می باشد. به همین علت از ماشین بولتzman 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 داده های آموزشی می توان از یک بردار ویژگی (حاوی اطلاعات هندسی وسایل نقلیه) نیز استفاده نمود. بلوک باقیمانده موجب می شود تا بردار ویژگی اصلی (تصاویر حرارتی) در طول آموزش حفظ شود.

(۲) نحوه به کار گیری ماشین بولتzman محدود گوسین - برنولی در یک ساختار شبکه عصبی کانولوشنی عمیق است. برای چنین عملکردی با توجه به سازگاری خروجی زیربلوک ها با ویژگی های تولیدی توسط ماشین

ویژگی‌های 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 مورد نظر 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 کانولوشنی ما با یک مقدار  $F(x)$  مواجه هستیم که با توجه به ویژگی‌های تولیدشده در یک زیربلوک کدگذار دارای مقادیر مشخصی هستند. از طرفی به صورت مستقل  $x$  را به عنوان ورودی ماشین بولتزمان محدود استفاده می‌کنیم و در این حالت نیز

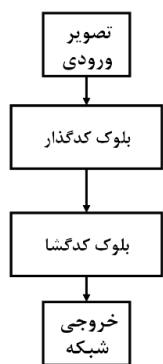


شکل ۷: الگوی تلفیق خروجی ماشین بولتزمان محدود گوسین-برنولی و معماری پیشنهادی در زیربلوک‌های کدگذار و کدگشا

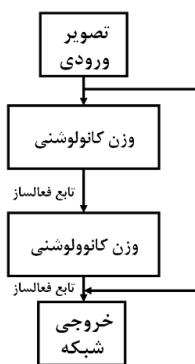
معماری رزنت (شکل ۸-ب) و به کارگیری ایده معماری سگمنت (شکل ۸-ج)، مدل کلی معماری پیشنهادی این تحقیق طراحی شده است.

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

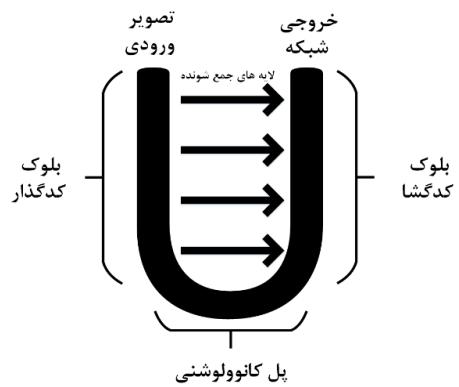
(ج)



(ب)



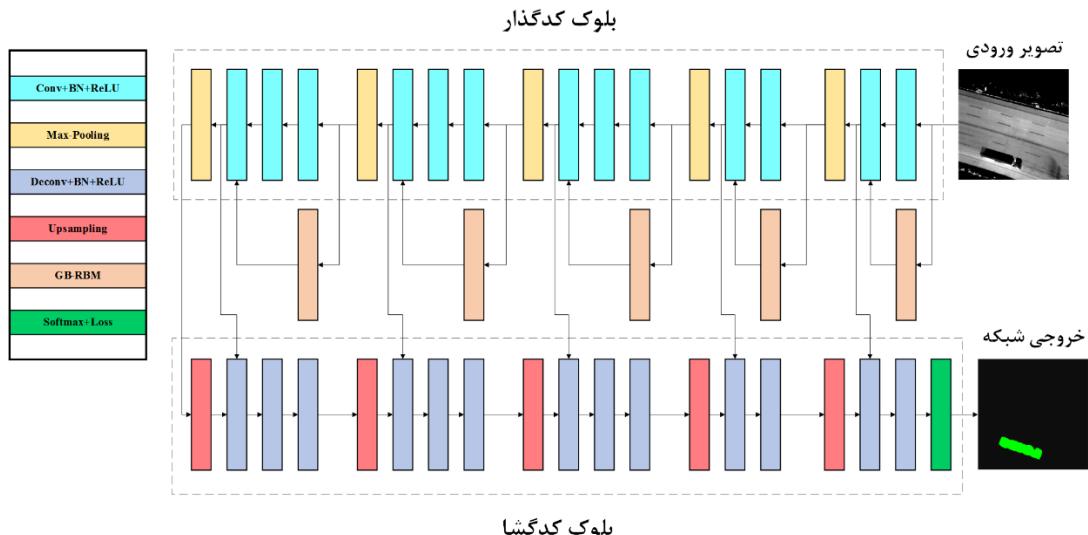
(الف)



شکل ۸: بلوک‌های سازنده معماری‌های (الف) یونت؛ (ب) رزنت؛ (ج) سگنت

و جدول(۳) ساختار بلوک کدگشا در معماری SegRBM-Net ارائه شده است.

طراحی معماری پیشنهادی تحقیق بر اساس شکل(۹) صورت گرفته است. در جدول(۲) ساختار بلوک کدگذار



شکل ۹: معماری پیشنهادی تحقیق با نام SegRBM-Net (در خروجی شبکه ناحیه سبز: قطعه‌بندی 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	
$m_1 \times n_1 \times 64$	1	$3 \times 3 / 64$	<i>Conv1+BN+ReLU</i>	1	بلوک کدگذار	
$m_2 \times n_2 \times 64$	1	$3 \times 3 / 64$	<i>Conv2+BN+ReLU+GBRBM</i>			
-	2	$2 \times 2$	<i>Max_Pooling</i>	2		
$m_3 \times n_3 \times 128$	1	$3 \times 3 / 128$	<i>Conv1+BN+ReLU</i>			
$m_4 \times n_4 \times 128$	1	$3 \times 3 / 128$	<i>Conv2+BN+ReLU+ GBRBM</i>	3		
-	2	$2 \times 2$	<i>Max_Pooling</i>			
$m_5 \times n_5 \times 256$	1	$3 \times 3 / 256$	<i>Conv1+BN+ReLU</i>	4		
$m_6 \times n_6 \times 256$	1	$3 \times 3 / 256$	<i>Conv2+BN+ReLU</i>			
$m_7 \times n_7 \times 256$	1	$3 \times 3 / 256$	<i>Conv3+BN+ReLU+ GBRBM</i>			
-	2	$2 \times 2$	<i>Max_Pooling</i>			
$m_8 \times n_8 \times 512$	1	$3 \times 3 / 512$	<i>Conv1+BN+ReLU</i>	5	بلوک کدگذار	
$m_9 \times n_9 \times 512$	1	$3 \times 3 / 512$	<i>Conv2+BN+ReLU</i>			
$m_{10} \times n_{10} \times 512$	1	$3 \times 3 / 512$	<i>Conv3+BN+ReLU+ GBRBM</i>			
-	2	$2 \times 2$	<i>Max_Pooling</i>			
$m_{11} \times n_{11} \times 1024$	1	$3 \times 3 / 1024$	<i>Conv1+BN+ReLU</i>			
$m_{12} \times n_{12} \times 1024$	1	$3 \times 3 / 1024$	<i>Conv2+BN+ReLU</i>			
$m_{13} \times n_{13} \times 1024$	1	$3 \times 3 / 1024$	<i>Conv3+BN+ReLU+ GBRBM</i>			
-	2	$2 \times 2$	<i>Max_Pooling</i>			

واحدهای پردازش گرافیکی (*GeForce GTX 1080 Ti*) و مرکزی (*Intel core i7-3770 CPU*) و سیستم عامل ویندوز ۷ صورت گرفته است.

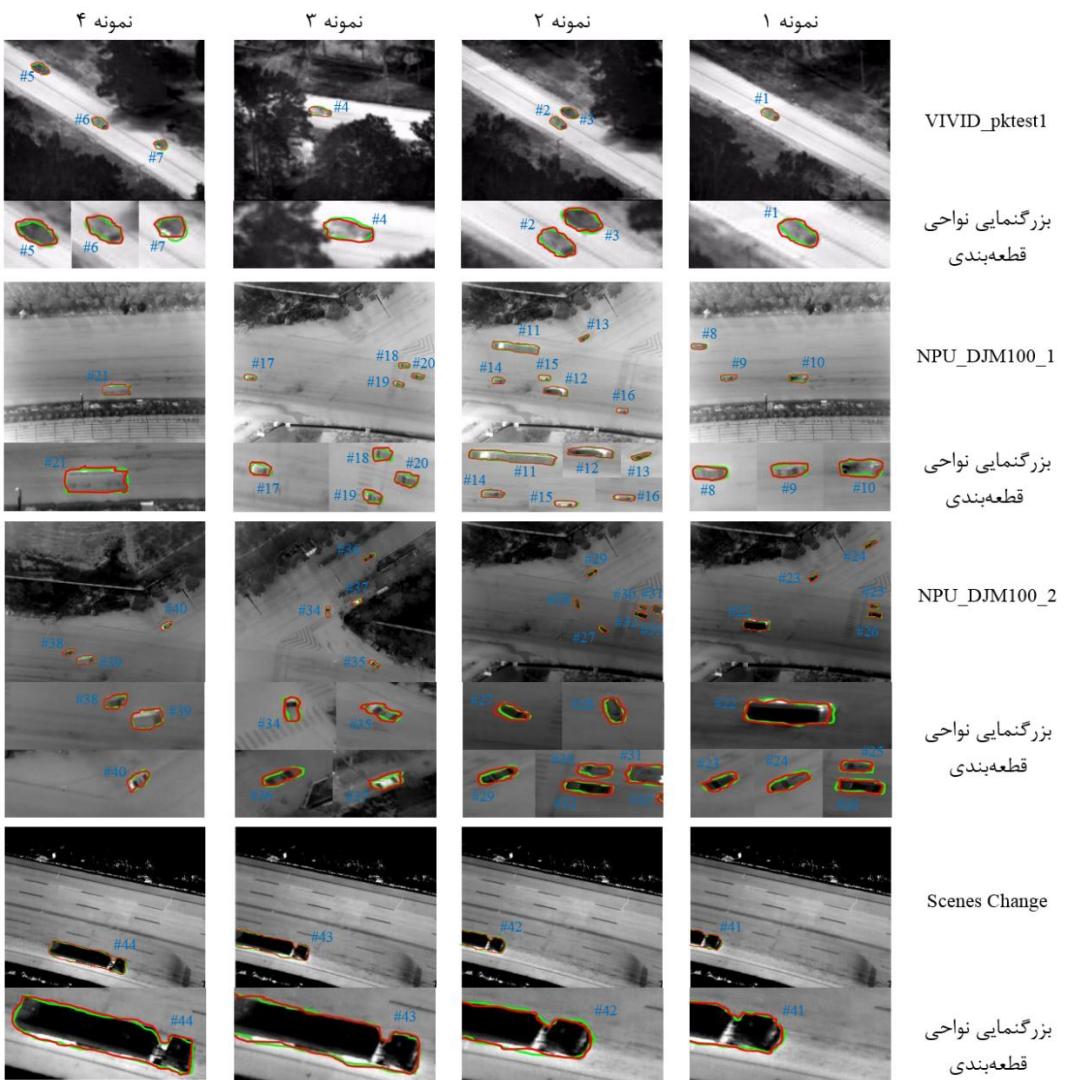
شکل (۱۰)، نتیجه آزمایش الگوریتم پیشنهادی تحقیق بر روی چهار نمونه از هر مجموعه داده با استفاده از روش پیشنهادی تحقیق را نشان می‌دهد. برای انتخاب این نمونه‌ها تنوع در جهت قرار گرفتن وسایل نقلیه، ابعاد و همچنین تفاوت در میزان انرژی تابشی وسایل نقلیه موردنظر بوده است.

برای طراحی و پیاده‌سازی مدل پیشنهادی موسوم به *SegRBM-Net* از بستر کد نویسی متلب<sup>۱</sup> و با بهره‌گیری از جعبه‌بازار مت-کانوونت<sup>۲</sup> که به طراحی مناسب‌تر شبکه کمک می‌کند استفاده شده است. جهت ارتقاء نقشه واقعیت زمینی با استفاده از ابزار برچسب‌دهی به تصاویر در نرمافزار متلب مزه‌های وسایل نقلیه از روی تصاویر استخراج و ذخیره شد. همچنین جهت اصلاح وزن‌های شبکه پیشنهادی از الگوریتم بهینه‌سازی گرادیان نزولی تصادفی<sup>۳</sup> جهت بهبود همگرایی الگوریتم استفاده شده است. پیکربندی سخت‌افزار مورد استفاده بر اساس پردازش موازی بر

<sup>۱</sup> MATLAB<sup>۲</sup> MatConvNet<sup>۳</sup> Stochastic Gradient Descent (SGD)

جدول ۳: ساختار بلوک کدگشا در معماری *SegRBM-Net*

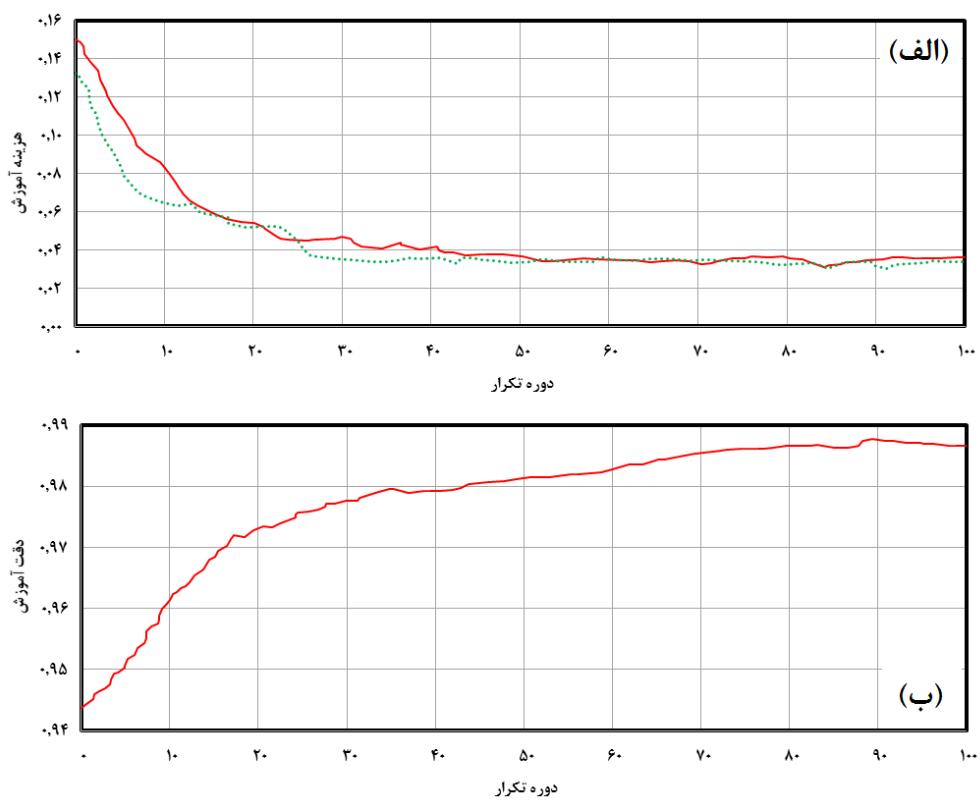
ابعاد خروجی	گام	ابعاد فیلتر	لایه کانولوشنی	زیربلوک	-
-	۲	۲×۲	<i>Upsampling</i>	۱	بلوک کدگشا
$m_1 \times n_1 \times 1024$	۱	۳×۳/۱۰۲۴	<i>Deconv1+BN+ReLU+ GBRBM</i>		
$m_2 \times n_2 \times 1024$	۱	۳×۳/۱۰۲۴	<i>Deconv2+BN+ReLU</i>		
$m_3 \times n_3 \times 512$	۱	۳×۳/۵۱۲	<i>Deconv3+BN+ReLU</i>		
-	۲	۲×۲	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



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 آموزش دیده 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

مورد استفاده و مجموعه تصاویر مادون قرمز حرارتی مورد مقایسه قرار گرفته‌اند. برای ارزیابی کلی قطعه‌بندی از ضریب همبستگی متیو<sup>۱</sup> (رابطه ۱۲) و شاخص تشابه ژاکارد<sup>۲</sup> (رابطه ۱۳) استفاده شده است [۶]. در روابط ۹ تا ۱۳،  $TP$  تعداد پیکسل‌های هدف در کلاس هدف،  $FP$  تعداد پیکسل‌های هدف در کلاس 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.

### 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-یو-نت و یک روش متدالوی بر اساس ماشین بردار پشتیبان به نام *HOG+SVM*<sup>۱۸</sup> [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

<sup>۱</sup> Matthews Correlation Coefficient (MCC)

<sup>۲</sup> Jaccard Similarity Index (JSI)

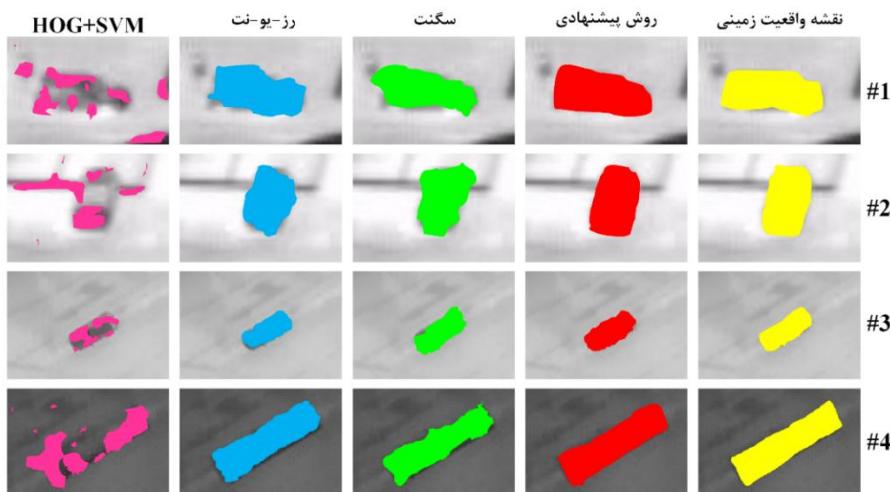
$$Precision = \frac{TP}{TP+FP} \quad \text{رابطه (۹)}$$

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

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

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

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



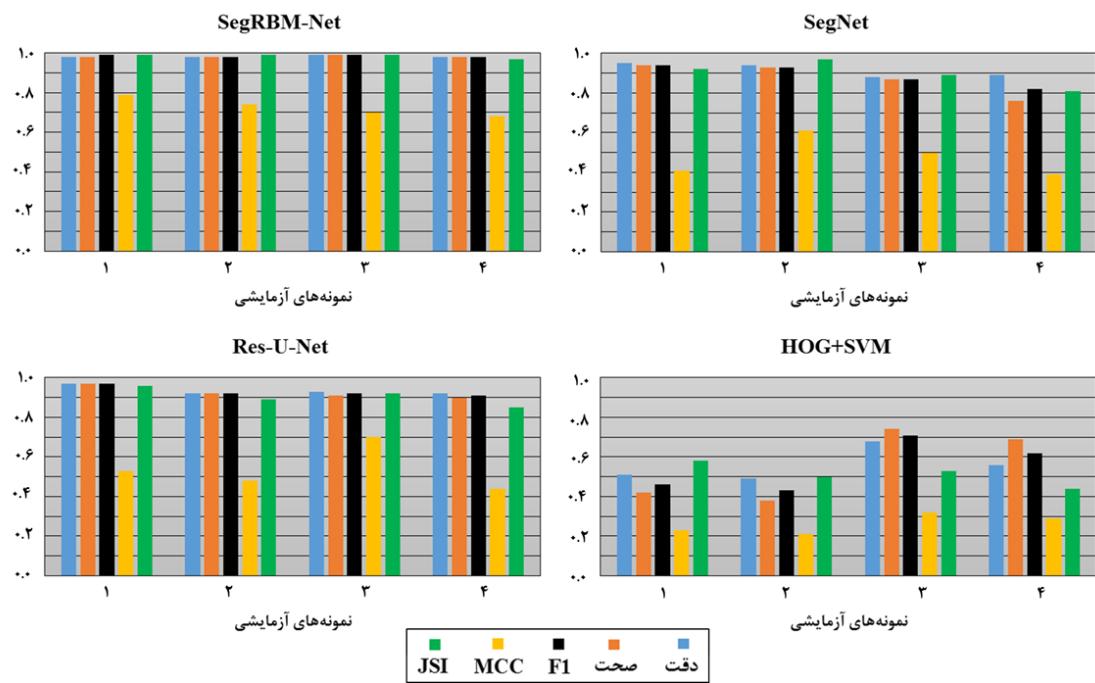
شکل ۱۲: مقایسه روش پیشنهادی تحقیق با الگوریتم‌های مشابه. زرد: نقشه واقعیت زمینی، قرمز: روش پیشنهادی، سبز: روش سگنت، آبی: روش رز-یو-نت، صورتی: روش *HOG+SVM*

به توضیح است که تمامی معماری‌های بررسی شده دارای ترخ همگرایی نسبتاً مناسبی هستند، اما معماری سگنت به دلیل عدم استفاده از لایه‌های بازگشتی دارای سرعت آموزش مطلوب‌تری می‌باشد. به علاوه 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 به صورت 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 و سایل نقلیه زمینی از تصاویر مادون‌قرمز حرارتی پهپاد عملکرد مطلوب‌تری نسبت به سه روش سگنت، رز-یو-نت و *HOG+SVM* داشته است.

۲) مسئله 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 پرنگ در جدول نشان داده شده‌اند. (الف) مدل *SegRBM-Net* (ب) مدل *SegNet* (ج) مدل *Res-U-Net* و (د) مدل *HOG+SVM*

زمان (ثانیه)	<i>F1</i>	صحت	دقت	روش	مجموعه
۶,۳۱	۰,۹۹۷	۰,۹۹۸	۰,۹۹۶	الف	VIVID_pktest1
۶,۳۷	۰,۹۹۵	۰,۹۹۶	۰,۹۹۴	ب	
۱۶,۵۵	۰,۹۹۷	۰,۹۹۸	۰,۹۹۵	ج	
۱۲۸,۲۱	۰,۵۸۴	۰,۶۵۳	۰,۵۲۹	د	
۲۶,۱۸	۰,۹۹۶	۰,۹۹۸	۰,۹۹۵	الف	NPU_DJM100_1
۲۷,۰۵	۰,۹۹۴	۰,۹۹۷	۰,۹۹۲	ب	
۴۲,۱۱	۰,۹۹۵	۰,۹۹۷	۰,۹۹۲	ج	
۱۵۴,۸۸	۰,۵۸۶	۰,۵۹۷	۰,۵۷۵	د	
۲۱,۴۹	۰,۹۹۳	۰,۹۹۷	۰,۹۸۹	الف	NPU_DJM100_2
۲۴,۳۷	۰,۹۹۰	۰,۹۹۵	۰,۹۸۵	ب	
۵۲,۸۴	۰,۹۹۰	۰,۹۹۵	۰,۹۸۶	ج	
۳۸۲,۹۱	۰,۷۰۵	۰,۹۵۳	۰,۵۵۹	د	
۱۱,۲۶	۰,۹۸۴	۰,۹۹۱	۰,۹۷۷	الف	Scenes Change
۱۳,۳۰	۰,۹۷۵	۰,۹۸۳	۰,۹۶۶	ب	
۳۲,۰۸	۰,۹۷۴	۰,۹۸۴	۰,۹۶۴	ج	
۲۴۴,۵۴	۰,۷۲۹	۰,۹۲۸	۰,۶۰۰	د	

## ۴- نتیجه‌گیری

تحقيق 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 پهپاد با تمرکز بر روی قطعه‌بندی وسایل نقلیه زمینی صورت گرفته است. ایده اصلی تحقیق به کارگیری هدفمند یادگیری عمیق در بحث قطعه‌بندی وسایل نقلیه از تصاویر حرارتی با قدرت تفکیک مکانی بالا می‌باشد. در همین راستا از جمله چالش‌های به کارگیری الگوریتم‌های یادگیری عمیق، مسئله داده‌های کلان برای آموزش الگوریتم مطرح است. در این مقاله، برای بهبود فرآیند آموزش در حل مسئله 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 موردنظر قرار گرفته است.

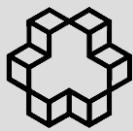
## مراجع

- [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.

- [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/1d0BM1wp1EUF 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. SSRR*, 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!Aj2qsxlPUOabbVHUH9J4E0sIyFE>



## 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 <sup>1</sup>, Reza Shah Hosseini <sup>2\*</sup>

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: rshahhosseini@ut.ac.ir