Scientific Journal of Electronical & Cyber Defence
Vol. 10, No. 2, Summer 2022, Serial No. 38

Fast Detection of Vehicle Type and Position in Images Based on Deep Neural Network

M. Nasehi, M. Ashourian*, H. Emami

*Associate Professor, Faculty of Electrical Engineering, Shahr Majlesi Branch, Islamic Azad University, Isfahan, Iran

(Received: 22/08/2021, Accepted: 27/11/2021)

ABSTRACT

Today, large-scale vehicles are scattered in different parts of the city and therefore need to be controlled by programmed systems. Automatically finding vehicles in the images and categorizing them is complicated because vehicles come in so many different shapes, colors, and models, and their designs are so different. Therefore, different methods of image analysis have been proposed to solve this problem. But there are some challenges such as the multiplicity of images in a scene, the coherence of the image of the vehicle and the image background, the presence of noise in the images and the tolerance to changes in light. In recent years, the use of deep neural networks has been proposed as an effective tool in identification despite the diversity of environmental conditions and objects. But the challenge of using deep neural networks is their high computational load. In this paper, a new approach is used to identify the type of vehicles, which uses a combination of VGG neural network and the Yolo image separation and tracking algorithm. This method improves the challenges of the previous methods and also reduces the computational load. The images are taken from two databases, ImageNet and COCO, and these databases are used to train and test the neural network. The results show that the designed system solves many problems well, including the speed of vehicle detection and the problem of computational load. The detection accuracy has increased by 2 to 3% compared to previous systems and has reached about 98%. The advantages of this approach include high-quality image detection and the use of a YOLO algorithm with an acceptable speed in detecting the type of vehicle.

Keywords: : Real-time algorithm, Deep Convolutional Neural Networks (CNN), Neural Networks VGG, Vehicle Detection.

* Corresponding Author Email: Ashourian@iaumajlesi.ac.ir

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University © Authors

نشریه علمی "پدافند الکترونیکی و سایبری"

سال دهم، شماره ۲، تابستان ۱۴۰۱، ص ۱۲۷-۱۱۷

علمی – پژوهشی

شناسایی سریع مکان و نوع وسیله نقلیه در تصاویر با استفاده از روش یادگیری عمیق

مجتبی ناصحی[۱]، محسن عشوریان[۲*]، حسین امامی[۳]

۱- دانشجوی دکترا، ۲و۳- دانشیار، دانشکده مهندسی برق، واحد شهر مجلسی، دانشگاه آزاد اسلامی، اصفهان، ایران

(دریافت: ۱۴۰۰/۰۵/۳۱، پذیرش: ۱۴۰۰/۰۹/۰۶)

چکیده

امروزه وسایل نقلیه در مقیاس بالا، در قسمت‌های مختلف شهر پراکنده هستند و از این جهت احتیاج به کنترل توسط سامانه‌های برنامه‌ریزی شده دارند. پیدا کردن خودکار وسایل نقلیه در تصویر و دسته‌بندی نوع آن‌ها پیچیده است، زیرا وسایل نقلیه شکل‌ها، رنگ‌ها و مدل‌های بسیار متفاوتی دارند و طراحی‌شان با یکدیگر متفاوت است. از این رو روش‌های مختلف آنالیز تصاویر برای حل این مسئله مطرح گردیده است. اما بعضی از چالش‌ها مانند تعدد تصویر در یک صحنه، بهم پیوستگی تصویر وسیله نقلیه و زمینه تصویر، وجود نویز در تصاویر، تلرانس نسبت به تغییرات نور وجود دارد. در سال‌های اخیر استفاده از شبکه‌های عصبی عمیق به‌عنوان ابزاری کارآمد در شناسایی با وجود تنوع شرایط محیطی و اجسام مطرح شده‌اند. اما چالش استفاده از شبکه‌های عصبی عمیق بار محاسباتی بالای آن‌هاست. در این مقاله رویکرد جدیدی برای شناسایی نوع وسایل نقلیه استفاده می‌شود، این رویکرد از ترکیب شبکه عصبی VGG و الگوریتم تفکیک و دنبال کردن تصاویر Yolo استفاده کرده است. این روش باعث بهبود چالش‌های روش‌های پیشین می‌گردد و در ضمن باعث کاهش بار محاسباتی می‌گردد. تصاویر از دو پایگاه داده ImageNet و COCO گرفته شده و از این پایگاه‌ها به‌منظور آموزش و آزمون شبکه عصبی استفاده می گردد. نتایج نشان می‌دهد که سامانه طراحی شده بسیاری از مشکلات را به خوبی برطرف می‌نماید. دقت تشخیص در مقایسه با سامانه‌های قبلی ۲ الی ۳ درصد افزایش یافته است. از مزایا این رویکرد می‌توان به کیفیت بالا در آشکارسازی تصاویر و سرعت قابل قبول در تشخیص نوع وسیله نقلیه اشاره کرد.

کلیدواژه‌ها: تشخیص وسایل نقلیه، شبکه عصبی عمیق کانولوشن، شبکه عصبی VGG

۱- مقدمه

امروزه می‌بینیم وسایل نقلیه در مقیاس بالا، در قسمت‌های مختلف شهر پراکنده هستند و از این جهت احتیاج به کنترل توسط سامانه‌های برنامه‌ریزی شده دارند. از کاربردهای این سامانه می‌توان موارد زیر را نام برد [۱]:

۱. سامانه‌های کنترل ترافیک (برای مثال بررسی حجم ترافیک یا شناسایی متخلفین)؛

۲. برنامه‌ریزی شهری؛

۳. خودروهای بدون راننده؛

۴. مدیریت پارکینگ‌ها با اعلام ورود وسایل نقلیه به مکان مورد نظر توسط ماشین؛

۵. در صورتی که ویژگی‌های خودرو مانند رنگ، مدل، شماره پلاک و غیره را بتوان توسط ماشین شناسایی کرد می‌توان برای تشخیص خودروی سرقت شده، متخلف و یا هر موردی که در تعقیب است از آن استفاده کرد؛

۶. در عوارضی‌ها برای جلوگیری از تخلفات کارمندان.

تشخیص وسیله نقلیه در بسیار از موارد کاربرد دارد و روش‌های گوناگون جهت تشخیص وسایل نقلیه در تصویر و یا به‌طور کلی اشیاء در تصویر وجود دارد، که در ادامه آن‌ها را بررسی می‌شود. البته ابتدا باید محل وسیله نقلیه در تصویر تشخیص داده شود، سپس مراحل بعدی طی شود.

در صورتی که بتوان سامانه‌ای طراحی کرد که قادر به شناسایی نوع وسیله نقلیه از روی تصویر آن باشد؛ می‌توان با تلفیق این دو سامانه، از بسیاری از تخلف‌ها جلوگیری کرد. سامانه‌های تشخیص و شناسایی وسیله نقلیه به سه دسته کلی زیر قابل تقسیم می‌باشند:

* رایانامه نویسنده مسئول: Ashourian@iaumajlesi.ac.ir

* این مقاله یک مقاله با دسترسی آزاد است که تحت شرایط و ضوابط مجوز Creative Commons Attribution (CC BY) توزیع شده است.

ناشر: دانشگاه جامع امام حسین (ع)

© نویسندگان

۱. تشخیص وسیله نقلیه؛

۲. شناسایی دسته کلی وسیله نقلیه و

۳. شناسایی نوع وسیله نقلیه

پیدا کردن وسایل نقلیه در تصویر کمی پیچیده است، زیرا برای مثال صورت انسان ساختار ساده‌تر و انعطاف‌پذیری کمتری دارد، اما وسایل نقلیه شکل‌ها و رنگ‌ها مدل‌های بسیار متفاوتی دارند چون طراحی‌شان با یکدیگر متفاوت است.

برای تشخیص وسیله نقلیه در تصویر چالش‌هایی که در پیش رو هستند عبارتند از:

- تصاویر بسیار متنوع از زوایای متفاوت از وسیله نقلیه؛
- گوناگونی در وسایل نقلیه به‌عنوان مثال خودرو سواری ، اتوبوس و ...؛
- رنگ‌ها و طرح‌های متفاوت وسایل نقلیه
- تصاویر در شب یا روز؛
- تصاویر در شرایط جوی متفاوت مانند برفی، بارانی یا آفتابی یا ابری.

در طی دو دهه گذشته روش‌های زیادی جهت شناسایی وسیله نقلیه مطرح شده است. با این حال داشتن سامانه‌ای با دقت خوب و سرعت عمل بالا و کاری در محیط‌های مختلف هنوز نیاز به کار دارد [۲].

در این مقاله، رویکردی جدید برای شناسایی نوع و مکان‌یابی وسایل نقلیه ارائه گشته است؛ این رویکرد با تکیه بر الگوریتم بلادرنگ و شبکه VGG سعی بر طبقه‌بندی وسایل نقلیه در چهار سطح سواری، شاسی بلند، اتوبوس و کامیون دارد، این روش نسبت به سایر روش‌های دیگر ترجیح داده شده است.

در بخش دوم مقاله، مروری بر کارهای انجام شده در این حوزه و مزایا و معایب هر یک صورت گرفته است. بخش سوم دربرگیرنده روش پیشنهادی است. نتیجه آزمایش‌ها در بخش چهارم ارائه گشته است و نتیجه‌گیری و جمع‌بندی مقاله نیز در بخش پنجم آورده شده است.

۲- مروری بر کارهای پیشین

روش‌هایی که برای شناسایی نوع و مکان‌یابی وسیله نقلیه وجود دارد به سه دسته تقسیم می‌شود.

۲-۱- روش سنتی و مبتنی بر استفاده از عملگرهای ساده تصویری

۲-۱-۱- تعیین مقدار آستانه بر روی تصاویر ترافیکی

استوارت و همکاران در سال ۱۹۹۴ روش تعیین مقدار آستانه بر روی تصاویر ترافیکی را پیشنهاد دادند. این روش ساده‌ترین روش در تشخیص وسیله نقلیه در حال حرکت است و بر اساس این حقیقت است که شی متحرک دارای شدت نور متفاوت از تصویر پس زمینه خواهد بود. با تعیین یک آستانه شدت نور در یک ناحیه کوچک از تصویر می‌توان خودرو را از تصویر پس زمینه تشخیص داد. این روش به شدت به مقدار آستانه وابسته است، که این مقدار نیز بر اساس چگالی نوری تصویر مربوط به یک وسیله نقلیه مشخص و مدل تصویر پس زمینه آن تعیین می‌شود [۴]. برای بهینه کردن مقدار آستانه می‌توان از شمارش تغییرات نوری استفاده کرد، اما آشکارسازی و تشخیص اشتباه وسایل نقلیه به دلیل سایه-های در حال حرکت همراه با وسیله نقلیه در طول روز، که دارای شدت نور یکسانی با محیط اطراف هستند، اجتناب‌ناپذیر است [۵].

محاسن آن عبارتند از:

- جدا کردن یک شی از پس زمینه؛
- متراکم کردن پیکسل‌های یک شیء؛
- محاسبه و استخراج ویژگی‌های یک شیء.

۲-۱-۲- آشکارسازی لبه‌های تصاویر ترافیکی

وون و همکاران در سال ۲۰۰۱ روش آشکارسازی لبه‌های تصاویر ترافیکی را پیشنهاد دادند. عملکرد این روش بر اساس ویژگی‌هایی است که در لبه‌های شیء مورد نظر در تصویر دیجیتالی گرفته شده از دوربین‌های نظارت تصویری وسیله نقلیه وجود دارد [۷]. این روش با توجه به اینکه، روش مقایسه‌ای بین دو تصویر مربوط به دو فریم نمی‌باشد، می‌تواند وسایل نقلیه متوقف شده را نیز تشخیص دهد. روش‌های آشکارسازی لبه‌های شیء متحرک مورفولوجیکال کاربردهای وسیعی دارند [۸]. در این روش لبه‌های وسایل نقلیه، نسبت به لبه‌های جاده برجسته نمایش داده می‌شوند. با استفاده از هیستگرام تصویر، به همراه یک سری عملیات‌های مورفولوژی می-توان ویژگی‌های لبه‌های تصویر را در حوزه مکانی به‌دست آورد. روش‌های آشکارسازی سه نوع گسستگی اساسی در شدت تصویر یعنی نقطه، خط و لبه در این روش قابل ارزیابی است. متداول‌ترین روش برای جستجوی گسستگی در این روش، اعمال ماسک به تصویر است.

یک لبه نوعی ممکن است مرز میان یک بخش قرمز رنگ و یک بخش سیاه رنگ باشد؛ حال آنکه یک خط می‌تواند تعداد کمی پیکسل‌های ناهم‌رنگ در یک زمینه یکنواخت باشد. در هر سوی خط یک لبه وجود خواهد داشت. لبه‌ها نقش مهمی در کاربردهای پردازش تصویر دارند.

۲-۱-۳- روش تجمیع ویژگی‌ها و دنبال کردن شیء متحرک

کیم و همکاران در سال ۲۰۰۱ روش تجمیع ویژگی‌ها و دنبال کردن شیء متحرک را پیشنهاد دادند. این روش بر اساس ابعاد و شکل وسیله نقلیه و یا به‌طور کلی ویژگی‌های هندسی وسیله نقلیه، تشخیص الگویی را سبب می‌شود. سپس کدگذاری‌هایی بر روی این ویژگی‌های هندسی وسیله نقلیه گذاشته خواهد شد [۹ و ۱۰]. این کدها می‌توانند بر اساس سیگنال زمانی شدت نور هر نقطه از تصویر ایجاد شوند. سپس با به‌کارگیری این کدها و ارتباط آن با مفاهیم فیزیکی حرکت می‌توان پیش بینی در زمینه تغییرات رفتار مدل به‌دست آمده از وسایل نقلیه را داشت [۱۱].

۲-۱-۴- روش جریان نوری

تِچمر و همکاران در سال ۲۰۰۱ روش جریان نوری را پیشنهاد دادند. رویکردهای موجود در روش جریان نوری از این امر بهره می‌جویند که راه هر شیء در حین حرکت اندکی تغییرات کوچک دارد، حال آنکه تغییر اساسی در گوشه‌هایی از تصویر رخ می‌دهد که شیء مورد نظر در حال حرکت به داخل و خارج از پس زمینه است [۱۲ و ۱۳]. جریان نوری در واقع نگاشت مکانی پیکسل تصویر در یک فریم و در زمان خاص است [۱۴ و ۱۵]. به عبارت دیگر، تغییرات موقتی ساختار سطح خاکستری را در سری تصویر نشان می‌دهد. همچنین این روش، اطلاعات مربوط به جابه‌جایی نسبی پیکسل‌ها و ساختار فضایی تصاویر را ارائه می‌دهد. رویکردهای گوناگونی برای تخمین صحیح و مؤثر جریان نوری ارائه شده‌اند [۱۶ و ۱۷].

کاربرد: برآورد حرکت و فشرده سازی ویدئو به‌عنوان جنبه‌های اصلی تحقیق در زمینه شار نوری توسعه یافته‌اند. میدان شار نوری به‌صورت محسوس به یک میدان حرکت متراکم شباهت دارد که از روش‌های برآورد حرکت استخراج می‌شود. این در حالی است که شار نوری نه تنها شامل مطالعه میدان شار نوری می‌شود، بلکه از آن در برآورد ماهیت سه بعدی و ساختار صحنه و همچنین حرکت سه بعدی اشیاء و ناظر نسبت به صحنه استفاده می‌شود. در اغلب موارد بیان شده از اصل ژاکوبین استفاده می‌شود.

در روش سنتی به دلیل طراحی ساده‌ای که داشت، سال‌ها در بسیاری از امور مورد استفاده قرار می‌گرفت، اما چالش‌هایی مانند شرایط آب و هوایی و تغییرات شب و روز که باعث تغییر نور می‌گردد، ایجاد مانع کوچک در برابر شیء مورد نظر و چسبیده بودن پیکسل‌های یک تصویر باعث شد که از روش‌های سنتی کمتر استفاده گردد [۱۸].

۲-۲- روش‌های مبتنی بر استخراج ویژگی‌های از تصویر و یک سامانه دسته‌بندی هوشمند

۲-۲-۱-روش آنالیز داده

لی و همکاران در سال ۲۰۱۵ روش روش آنالیز داده را پیشنهاد دادند. این روش با آنالیز داده و استفاده از تصاویر زیاد به الگوریتم آموزش داده می‌شود که در کدام تصاویر وسایل نقلیه وجود دارد یا وجود ندارد. سپس ویژگی‌های پرمعنا از تصویر استخراج می‌شود و ماشین تشخیص می‌دهد که وسیله نقلیه کجا قرار دارد یا اینکه اصلاً وجود ندارد [۱۸].

اجزای یک سامانه بینایی ماشین عبارتند از:

- یک یا چند دوربین دیجیتال و آنالوگ؛
- واسطه‌ای که عکس را برای پردازش آماده سازد (برای دوربین آنالوگ این واسط دیجیتال کننده عکس می‌باشد که به آن **frame grabber** می‌گویند.)؛
- یک پردازشگر مانند **DSP**؛
- نرم‌افزار بینایی ماشین؛
- سخت‌افزار ورودی و خروجی برای گزارش نتایج.

۲-۲-۲- تحلیل دنباله‌ای از تصاویر

هادی و همکاران در سال ۲۰۱۵ روش تحلیل دنباله‌ای از تصاویر را پیشنهاد دادند. در این روش فرض می‌شود صحنه پویاست. ماشین برای یافتن وسیله نقلیه سه راه دارد: ۱) با متدهای مشخص زمینه را حذف کند؛ ۲) ویژگی‌ها را با متدهایی استخراج کند؛ ۳) با استفاده از متدهایی، قالب‌های متفاوت را در تصویر از یکدیگر جدا کند و اجسام متحرک را در تصویر تشخیص دهد.

در این روش تصاویر مربوط به دوربین و نقاط سه بعدی صحنه به‌صورت هم‌زمان در قالب یک بردار حالت تخمین زده می‌شوند، بدین صورت که ابتدا زمینه تصویر حذف سپس ویژگی‌های تصویر استخراج شده و موقعیت سه بعدی تصویر روی خطی که از مرکز دوربین به آن نقاط متصل شده‌اند و با یک عدم قطعیت بالا در عمق ۰/۵ تا ۵ متر با توزیع یکنواخت تعیین می‌-

شوند، سپس در هر فریم جدید با تکرار ویژگی‌ها، میزان عدم قطعیت کاسته می‌گردد [۱۸].

۲-۲-۳- استفاده از ویژگی Haar-like

اولیویرا و همکاران در سال ۲۰۰۸ روش استفاده از ویژگی هارلایک را پیشنهاد دادند. در این روش ایده اصلی آن به این شرح است، که با توجه به ویژگی‌های ساده‌ای که طبقه‌بندی کننده در چند مرحله ساماندهی کرده است یک شیء به‌صورت آبشاری تعریف می‌کنند (یا به‌صورت پله پله). سامانه پیشنهادی بر اساس شبکه عصبی مصنوعی کار می‌کند. این روش تصویر را به‌صورت طبقه‌بندی آبشاری ساماندهی می‌کند و سپس از یک سری فیلتر عبور می‌دهد تا تصویر اصلی به‌دست آید [۱۹].

روش‌های مبتنی بر استخراج ویژگی تا حدود سال ۲۰۱۵ میلادی مورد استفاده قرار می‌گرفت و دقتی حدود ۷۰ درصد داشت، اما چالش‌هایی مانند تعدد تصویر در ورودی، تغییرات نسبت به نور، خطا در تشخیص اهدافی که دارای سرعت بالا داشتند و قدرت کم در تشخیص اجزای تصویر باعث شد از این روش کمتر استفاده گردد [۱۸].

۲-۳- روش‌های مبتنی بر شبکه عصبی کانولوشن

۲-۳-۱- شناسایی نوع وسیله نقلیه بر اساس شبکه R-FCN

ژیانگ و همکارن در سال ۲۰۱۸ این روش را پیشنهاد دادند. ایده این روش مبتنی بر شبکه یادگیری عمیق و کاملاً مجزا می‌باشد و از چارچوب R-FCN و ترکیب بانک اطلاعاتی وسایل نقلیه در ImageNet استفاده می‌کند. در این روش، از استخراج نمونه آنلاین سخت (OHEM) برای بهینه‌سازی پارامترهای شبکه استفاده می‌شود. پس از تکرارهای مکرر جهت آموزش شبکه، سرانجام مدل R-FCN از تشخیص هدف وسیله نقلیه به‌دست می‌آید. با توجه به نتایج شبیه‌سازی این روش ۸۷/۴۸ به‌دست می‌آید [۲۰].

۲-۳-۲- طبقه بندی نوع وسیله نقلیه بر اساس شبکه عصبی پیچشی

مینگلان و همکاران در سال ۲۰۱۸ این روش را پیشنهاد دادند. در این روش طبقه بندی نوع وسیله نقلیه بر اساس روش یادگیری عمیق ارائه می‌گردد. این سامانه از دو مرحله تشکیل شده است. در مرحله اول، برای افزایش مشکل مجموعه داده و عدم تعادل، از تقویت داده استفاده می‌شود. در مرحله دوم، از یک شبکه عصبی پیچشی استفاده کرده است [۲۱].

CNN[1] یا شبکه عصبی پیچشی مدل و معماری‌های مختلفی دارد. این شبکه با استفاده از پارامترهایی که از مجموعه داده‌ها به‌دست می‌آورد، آموزش داده می‌شود. این سامانه بخشی از یک برنامه یکپارچه است که امکان مدیریت خودکار سیگنال راهنمایی و رانندگی را بر اساس نوع تشخیص خودکار وسیله نقلیه فراهم می‌کند. در این روش با توجه به استفاده از شبکه پیچشی دقت حدود ۹۲٪ می‌گردد [۲۲ و ۲۳].

۲-۳-۳- تشخیص و طبقه بندی نوع وسیله نقلیه با استفاده از شبکه عصبی کانولوشن

مینگلان و همکاران در سال ۲۰۱۸ این روش را پیشنهاد دادند. در این مقاله با استفاده شبکه‌های عصبی روشی سریع جهت پاسخگویی با استفاده از شبکه‌های RCNN، AlexNet استفاده شده است. این الگوریتم با مجموعه داده‌های واقعی آموزش داده شده است و مجموعه داده‌ها با افزونگی شده می‌تواند در تعیین کیفیت نوع مدل وسیله نقلیه دقت و صحت بیشتری را در حدود ۹۱٪ به‌دست آورد [۲۴ و ۲۵].

در روش‌های مبتنی بر شبکه عصبی کانولوشن دقت تشخیص شیء به بالای ۹۰ درصد می‌رسد و تا حدودی چالش‌های روش‌های گذشته رفع می‌گردد، اما بار محاسباتی افزایش می‌یابد که سرعت تشخیص شیء را کندتر می‌کند [۱۸].

۳- روش پیشنهادی جهت شناسایی و مکان‌یابی وسایل نقلیه

استفاده از روش شبکه عصبی کانولوشن اگر چه گام مهمی در شناسایی وسایل نقلیه به شمار می‌آید، اما بهبود عملکرد پردازش محاسابات باید مورد توجه قرار بگیرد. در روش پیشنهادی شناسایی و مکان‌یابی وسایل نقلیه مراحل مختلفی استفاده می‌گردد، که عبارتند از: پیش پردازش، مکان‌یابی (شناسایی ناحیه)، تشخیص نوع وسیله نقلیه، طبقه بندی و ارزیابی.

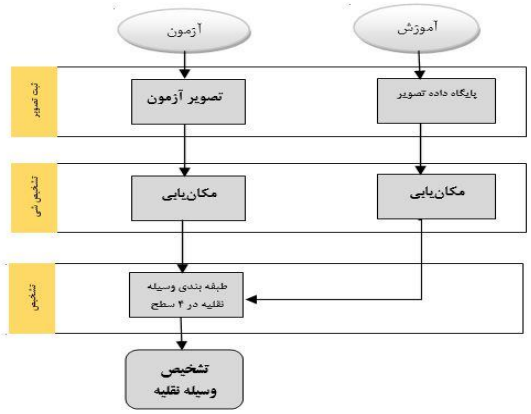
شکل (۱) فرآیند کلی آموزش و آزمون در این الگوریتم را نشان می‌دهد.

در پیش پردازش حذف نویز انجام می‌شود و با فرآیندهایی بهبود کیفیت تصویر انجام می‌پذیرد. سپس برای استخراج ویژگی وسایل نقلیه از روش یادگیری عمیق استفاده می‌گردد.

یادگیری عمیق روشی دارای ویژگی‌های زیر می‌باشد:

- یادگیری عمیق با استفاده از شبکه عصبی به تقلید از هوش جانداران می‌پردازد.

[1] Convolution neural network



شکل (۱): مراحل تشخیص و مکان‌یابی وسایل نقلیه در چهار سطح

- آموزش شبکه عصبی هوش مصنوعی با استفاده از یادگیری عمیق به داده‌های زیادی نیاز دارد.
- استفاده از یادگیری عمیق مستلزم برخورداری از یک واحد پردازش گرافیکی بسیار قدرتمند است.
- یادگیری عمیق با استفاده از مدل‌ها و الگوریتم‌های متفاوتی در تلاش است تا منطق بشر را شبیه‌سازی کند.

متداول‌ترین الگوریتم‌های یادگیری عمیق که برای تشخیص نوع وسیله نقلیه مورد استفاده قرار می‌گیرند عبارتند از: YOLO، SSD، صورت کوچک و روش‌های بخش‌بندی نمونه کوچک از جمله U-Net و R-CNN.

اما در میان روش‌های موجود، الگوریتم YOLO با توجه به سرعت بالا در تشخیص شیء، مورد توجه قرار گرفته است.

این الگوریتم یک سامانه تشخیص اشیاء است. کامل‌ترین سامانه بلادرنگ[1] در یادگیری عمیق و حل مسائل تشخیص تصویر است. این الگوریتم ابتدا تصویر را به بخش‌های مختلف تقسیم می‌کند و هر بخش را علامت‌گذاری می‌کند، سپس الگوریتم شناسایی را به‌صورت موازی برای تمامی این بخش‌ها اجرا می‌کند تا ببیند هر بخش به کدام دسته‌بندی تعلق می‌گیرد. بعد از شناسایی کامل اشیاء، آن‌ها را به هم متصل می‌کند تا دو هر شیء اصلی یک جعبه باشد.

همه این کارها به‌صورت موازی انجام می‌شوند؛ در نتیجه به‌صورت بلادرنگ است و می‌تواند تا ۴۰ تصویر را در یک ثانیه پردازش کند. با اینکه این مدل کمی نسبت به RCNN عملکرد ضعیف‌تری دارد، ولی به دلیل بلادرنگ بودن آن، می‌توان از آن برای حل مشکلات روزانه استفاده کرد.

نحوه تشخیص شیء در تصویر توسط الگوریتم yolo:

در شبکه YOLO تصاویر به‌صورت فرضی به N در N ناحیه تقسیم می‌شوند. به ازای هر کدام از این نواحی فرض می‌شود که K شیء امکان حضور دارند. شبکه به‌صورتی طراحی شده است که درنهایت به ازای تمامی نواحی امکان وجود اشیاء در آن ناحیه‌ها، دسته آن شیء و همچنین مختصات کادر دور شیء را پیش‌بینی می‌کند.

با توجه به اینکه کادر دور هر شیء ۴ مختصات دارد و همچنین دسته مورد نظر شیء یک بردار C تایی هست و همین‌طور یک عدد برای احتمال وجود شیء دارید. پس به ازای هر شیء یک تانور C+1+4 تایی مورد نیاز است.

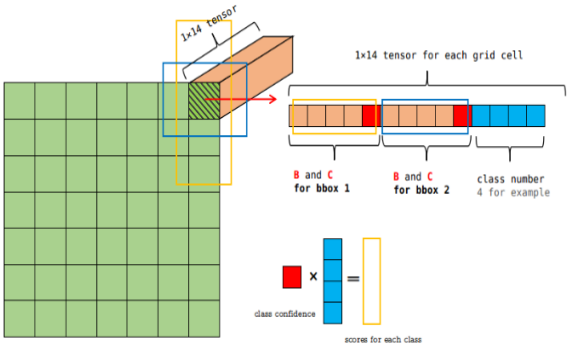
با توجه به اینکه فرض شد در هر تصویر N در N ناحیه وجود دارد و هر ناحیه احتمال حضور K شیء به‌صورت بیشینه وجود دارد و هر شیء نیز تانور C+1+4 نیاز دارد پس در انتهای شبکه تانور KNN(C+1+4) وجود دارد که به پیش‌بینی کادر اشیاء و احتمال آن‌ها و دسته آن‌ها می‌پردازد.

در نسخه نخست این شبکه پس از لایه‌های کانولوشن ویژگی‌ها خطی شده و به لایه‌های تمام متصل تحویل داده شده و درنهایت پیش‌بینی‌ها انجام می‌شوند. در حالی که در نسخه‌های بعدی به جای ایجاد تانور به این ابعاد حجیم یک کرنل کانولوشن با ابعاد K(C+1+4) به نحوی که بر روی ویژگی‌های آخرین لایه کانولوشن کانوالو می‌شود که ناحیه‌های ادراکی آن (Receptive Field) بر روی هر یک از N×N ناحیه از تصور ورودی قرار بگیرند و عملیات پیش‌بینی انجام شود. ساختار الگوریتم در شکل (۲) آورده شده است.

تفاوت عمده این شبکه با شبکه RCNN این است که این شبکه به‌صورت سراسری (End-to-End) کار می‌کند و بحث تشخیص مکان اشیاء را بدون نیاز به الگوریتم‌های بیرونی در RCNN حل می‌کند و سرعت بسیار بالایی دارد و تماماً با استفاده از شبکه‌های عصبی پیاده‌سازی شده و در نتیجه قابلیت موازی‌سازی بسیار مناسبی بر روی پردازنده‌هایی مثل GPUها و FPGAها را داراست.

اما الگوریتم باید بر روی شبکه عصبی پیاده‌سازی گردد، برای اجرای این عمل از شبکه عصبی کانولوشن استفاده می‌گردد. متداولی شبکه‌های عصبی کانولوشن که وجود دارند عبارتند از: شبکه ImageNet، شبکه LeNet، شبکه AlexNet، شبکه ZFNet، شبکه VGGNet، شبکه GoogleNet و شبکه ResNet.

[1] Real Time



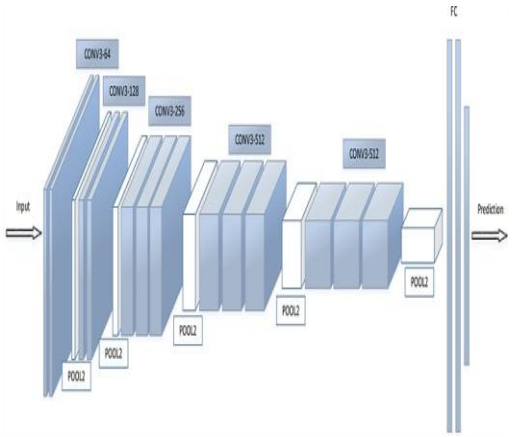
شکل (۲): ساختار YOLO

اما در میان شبکه‌های عصبی بیان شده شبکه VGG در حال حاضر محبوب‌ترین انتخاب برای Feature extraction از تصاویر است. به‌طور خاص مدل از پیش آموزش داده شده آن‌ها برای استفاده در کتابخانه Caffe وجود دارد. یک کاستی این شبکه این است که حافظه مصرفی و تعداد پارامتر بسیار زیادی دارد.

شبکه VGG[1] در دو معماری مختلف با عنوان‌های VGG 16و VGG 19 ارائه شده است. ابتدا شبکه VGG 16 پیشنهاد شد و بعدها با تغییراتی جزئی در شبکه VGG 16، شبکه VGG 19 مطرح گردید. شبکه VGG 16 همان‌طور که در شکل (۳) نشان داده شده، شامل ۱۶ لایه کانولوشنی یا ۱۶ لایه پارامتری است. شبکهVGG 16 ، شامل دو لایه کانولوشنی با ۶۴ فیلتر ۳×۳ هست که پشت سر هم قرار گرفته‌اند. سپس، یک لایه ماکس‌پولینگ ۲×۲ با پرش (Stride) به اندازه ۲ قرار گرفته است. این لایه ماکس‌پولینگ علاوه بر نمونه‌برداری، وظیفه کاهش بعد ویژگی‌ها به نصف را هم دارد. در ادامه، دو لایه کانولوشنی دیگر با ۱۲۸ فیلتر ۳×۳ و یک لایه ماکس‌پولینگ ۲×۲ و پرش ۲ قرار گرفته‌اند. به‌طور مشابه، سه لایه کانولوشنی با ۲۵۶ فیلتر ۳×۳ و یک لایه ماکس‌پولینگ ۲×۲ با پرش ۲ قرار گرفته‌اند. لایه کانولوشنی با ۵۱۲ فیلتر ۳×۳ و یک لایه ماکس‌پولینگ ادامه این شبکه هست که البته دو بار تکرار می‌شود. درنهایت، ویژگی‌ها تبدیل به یک بردار ویژگی می‌شوند تا در اختیار لایه‌های نورونی یا تمام اتصال[2] قرار گیرند. دو لایه نورونی به ابعاد ۴۰۹۶ پشت سر هم قرار گرفته‌اند. درنهایت، یک لایه نورونی به ابعاد ۱۰۰۰ که متناظر با تعداد کلاس‌های کاربرد ما هست، درنظر گرفته شده است. در تمامی لایه‌های کانولوشنی و لایه‌های نورونی از تابع فعال‌ساز[3] به نام RELU استفاده شده است [۳۰].

همان‌طور که در شبکه‌های کانولوشنی متداول است، می‌بینید که بیشترین حافظه مصرفی در لایه‌های کانولوشن ابتدایی مصرف شده‌اند و بیشترین تعداد پارامترها نیز در لایه‌های تماماً متصل آخر وجود دارند. در این مورد خاص، اولین لایه تماماً متصل دارای ۱۰۰ میلیون وزن از ۱۴۰ میلیون وزن موجود در سراسر شبکه می‌باشد [۶].

با توجه به مزایایی که ارائه گردید، استفاده از ترکیب شبکه عصبی VGG با الگوریتم YOLO یک راه حل مناسب جهت افزایش سرعت پردازش (کاهش بار محاسباتی) و بهبود عملکرد چالش‌هایی همچون شرایط آب و هوایی، گوناگونی در انواع وسایل نقلیه، تصاویر بسیار متنوع از زوایای متفاوت از وسایل نقلیه می‌باشد، که در نتیجه باعث افزایش دقت (تشخیص شیء) می‌گردد. در ضمن این سامانه می‌تواند باعث تسریع عمل شناسایی وسایل نقلیه در اتوبان‌ها و در هنگام عبور از ایست و بازرسی شود.



شکل (۳): شبکه عصبی VGG

۴- نتایج

۴-۱- مجموعه داده‌ها

در این مقاله از چندین پایگاه داده استفاده شده است، که همه آن‌ها آموزش دیده شده‌اند که در جدول (۱) مقدار هر کلاس آورده شده است.

جدول (۱): مجموع داده‌های تصاویر

Dataset/Class	سواری	شاسی بلند	کامیون	ون
VEDAI	۷۴۰	۶۰۰	۳۰۰	۱۰۰
NZAM/ONERA Christchurch	۱۱۶۷	۱۱۵۰	۷۳	۱۲۰
ISPRS Potsdam	۱۱۸۷	۷۵۳	۳۳	۱۸۱

[1] Very Deep Convolutional Networks for Large-Scale Image Recognition
[2] Fully Connected
[3] Activation Function

این دیتاست‌های مختلف با کلاس‌های مختلفی جمع‌آوری گردیده شده که هر کدام از پایگاه داده‌های مختلفی از قبیل ImageNet و COCO جمع‌آوری شده است و در شبیه‌سازی از آن‌ها استفاده شده است.

ImageNet یک پایگاه داده تصویری است که طبق سلسله مراتب WordNet سازمان یافته است، که در آن هر گروه توسط صدها و هزاران تصویر تشکیل شده است.

COCO یک پایگاه داده تصویری است که جهت تشخیص شیء مورد استفاده قرار می‌گیرد.

مجموع تصاویر استفاده ۶۴۰۴ تصویر می‌باشد، از این مجموع حدود ۷۰ درصد تصاویر جهت آزمون (حدود ۴۴۸۳ تصویر) و مابقی که حدود ۳۰درصد (۱۹۲۱ تصویر) تصاویر در بخش آزمایش مورد استفاده قرار می‌گیرد. شکل۴و۵ نمونه‌ای از تصاویر دیتاست که از پایگاه داده‌های مختلف جمع‌آوری شده است.

شکل (۴): نمونه‌هایی از دیتاست‌های تک شیء (از دیتاست مربوط به پایگاه ImageNet)

شکل (۵): نمونه‌هایی از دیتاست‌های چند وسیله نقلیه در تصویر (از دیتاست مربوط به پایگاه ImageNet)

۴-۲- نتایج شبیه‌سازی

به‌منظور پیاده‌سازی روش پیشنهادی از نرم‌افزار متلب استفاده شده است. تمامی اجرا را نیز بر روی یک نسخه 2018b سیستم عامل Window10، پردازنده Intel core i7، کارت حافظه ۱۶ GB of DDR3 memory running at 799.5 MHz و کارت گرافیک NVIDIA GeForce GTX 750 Ti انجام شده است. دیگر تنظیمات به شرح زیر است:

- معماری شبکه مورد استفاده شبکه عصبی کانولوشنی VGG و
- برای بهینه‌سازی وزن‌های شبکه از الگوریتم بلادرنگ YOLO استفاده شده است.

جدول (۲): نتایج دقت تشخیص وسایل نقلیه در شبکه‌های مختلف عصبی عمیق

دقت تشخیص وسایل نقلیه	
Alex Net	٪۸۵/۳۲
VGG Net	٪۹۳/۳۲
Google Net	٪۹۱/۰۹
Res Net	٪۹۱/۹۹

با توجه به نتایج جدول (۲) بهترین انتخاب شبکه عصبی، شبکه عصبی VGG می‌باشد. شبکه عصبی VGG، به دلیل این که تعدادHyper Parameter ها را کاهش داده، پیچیدگی کمتری نسبت به AlexNet و ZFNet دارد و به دلیل همین سادگی، از محبوبیت بالایی برخوردار است و دارای دقت بالایی هم می‌باشد. اما الگوریتم YOLO، کامل‌ترین سامانه بلادرنگ در یادگیری عمیق و حل مسائل تشخیص تصویر است. این الگوریتم ابتدا تصویر را به بخش‌های مختلف تقسیم می‌کند و هر بخش را علامت‌گذاری می‌کند، سپس الگوریتم شناسایی را به‌صورت موازی برای تمامی بخش‌ها اجرا می‌کند تا ببیند هر بخش به کدام دسته‌بندی تعلق می‌گیرد. بعد از شناسایی کامل اشیاء، آن‌ها را به هم متصل می‌کند. همه این کارها به‌صورت موازی انجام می‌شوند؛ در نتیجه به‌صورت بلادرنگ است و می‌تواند تا ۴۰ شیء در یک تصویر را در یک ثانیه پردازش کند.

YOLO در دو مرحله مجزا انجام می‌شود. ابتدا، شبکه YOLOبا پایگاه داده ImageNet برای عمل کلاس‌بندی آموزش داده شده است. در این فرآیند آموزش، از ۲۰ لایه کانولوشنی ابتدایی YOLO استفاده شده است. در انتهای این ۲۰ لایه، یک

لایه پولینگ میانگین[1] و یک لایه فولی‌کانکتد قرار گرفته است. تصاویر ورودی در اندازه ۳×۲۲۴×۲۲۴ به شبکه داده شده‌اند.

در مرحله دوم، برای کار تشخیص اشیاء در ساختار مدل تغییراتی ایجاد شده است. تغییرات به این صورت است که چهار لایه کانولوشنی و دو لایه فولی کانکتد با وزن‌های تصادفی به انتهای ۲۰ لایه شبکه اضافه شده است. در کار تشخیص اشیاء اغلب به اطلاعات با جزئیات بیشتری نیاز است، به همین دلیل رزولوشن ورودی شبکه از ۳×۲۲۴×۲۲۴ به ۳×۴۴۸×۴۴۸ افزایش داده شده است. هدف از افزایش اندازه ورودی، بهره‌گیری از جزئیات بیشتر در تصویر است.

سپس به بررسی خروجی شبکه پرداخته می‌شود. اندازه خروجی شبکه ۳۰×۷×۷ است. ابتدا از اندازه ۷×۷ شروع کنید؛ تصاویر ورودی به یک شبکه ۷×۷ تقسیم‌بندی می‌شوند. بنابراین، خروجی ۷×۷ متناظر با تصویر شبکه شده ورودی است. هر درایه در ۷×۷ خروجی، متناظر با یک سلول در تصویر شبکه شده ورودی است. هر درایه از این ماتریس ۷×۷ خروجی، یک بردار به طول ۳۰ دارد. این بردار به طول ۳۰ شامل اطلاعات پیش‌بینی احتمال‌ها و مختصات باکس است. هر سلول از این آرایه ۷×۷ دو باکس می‌تواند رسم کرد. برای رسم هر باکس به ۵ پارامتر (x,y,w,h,confidence)نیاز است. پارامترهای x و y، مختصات سطر و ستون مبدأ باکس (مرکز باکس) را نشان می‌دهند. مختصات w و h به ترتیب متناظر با پهنا و ارتفاع باکس هستند. با این چهار پارامتر می‌توان باکس را ترسیم کرد. پارامتر پنجم confidenceهست؛ یک پارامتر احتمالاتی با مقدار بین ۰ تا ۱، این باکس شامل شیء هست یا اینکه پس‌زمینه تصویر است.

با توجه به مزایای که گفته شد با استفاده از نرم‌افزار متلب شبیه‌سازی انجام می‌شود. نتایج شبیه‌سازی در جدول (۳) آورده شده است. نتایج شبیه‌سازی نشان می‌دهد که در محیط پیچیده، حتی اگر چندین نوع وسیله نقلیه هم‌زمان در تصویر وجود داشته باشد، روش پیشنهادی می‌تواند طبقه‌بندی دقیق‌تری از روش‌های پیشین انجام دهد. بازدهی در مجموع چهار کلاس حدود ٪۹۷/۵۸ شده است و سرعت نرخ یادگیری (سرعت پردازش) حدود ۰/۰۰۱۰ می‌باشد که هر ۵۰ دور یک بار نرخ یادگیری بر ۱۰ تقسیم می‌گردد و به مرور زمان کوچک‌تر می‌شود. نرخ یادگیری[2] اغلب با نماد αو گاهی اوقات با نماد η نمایش داده می‌شود و بیانگر سرعت (گام) به‌روزرسانی وزن‌ها است که می‌تواند مقداری ثابت یا به سازگار شونده تغییر کند. در این روش یادگیری به‌صورت دسته‌ای انجام می‌پذیرد، یعنی به جای اینکه تمام داده‌های آموزشی یک جا به شبکه اعمال گردند، به‌صورت بسته‌ای و به نوبت به شبکه عصبی وارد می‌شوند. سرعت پردازش بالا می‌باشد و دلیل آن استفاده از الگوریتم YOLO می‌باشد. این الگوریتم باعث شده نرخ یادگیری در فرآیند آموزش، زمان آموزش را کاهش دهد و سرعت پردازش را بهبود بخشد، که در نتیجه باعث کاهش بار محاسبات و افزایش دقت تشخیص شیء نسبت به روش‌های پیشین گردد. نتایج شبیه‌سازی بر روی مجموعه داده‌های چهار کلاس سواری، شاسی بلند، اتوبوس و کامیون در جدول (۳) آورده شده است.

جدول (۳): نتایج شبیه‌سازی با استفاده از نرم‌افزار متلب

نوع وسیله نقلیه	اتلاف	دقت شبکه عصبی	سرعت نرخ یادگیری
سواری	۰/۳۷۴۹	٪۹۶/۹۷	۰/۰۰۱۰
کامیون	۰/۳۶۴۱	٪۹۷/۲۲	۰/۰۰۱۰
شاسی بلند	۰/۳۶۸۹	٪۹۸/۵۵	۰/۰۰۱۰
اتوبوس	۰/۳۶۷۸	٪۹۷/۶۳	۰/۰۰۱۰

جدول (۴): نتایج محاسبات با استفاده از نرم‌افزار متلب

RMSE	Time Elapsed	Iteration	Epoch
میانگین مربع خطاها	زمان سپری شده	تکرار آموزش	دوره
۰/۰۵	۰۰:۰۰:۰۴	۱	۱
	۰۰:۰۱:۳۰	۲۰۰	۱
	۰۰:۰۲:۵۶	۴۰۰	۲
	۰۰:۰۴:۲۰	۶۰۰	۳
	۰۰:۰۵:۴۳	۸۰۰	۳
۰/۱۵	۰۰:۰۷:۰۷	۱۰۰۰	۴
۰/۱۱	۰۰:۰۸:۳۰	۱۲۰۰	۵
	۰۰:۰۹:۵۰	۱۴۰۰	۵
۰/۱۴	۰۰:۱۱:۱۰	۱۶۰۰	۶
۰/۱۴	۰۰:۱۲:۳۰	۱۸۰۰	۷
	۰۰:۱۳:۵۰	۲۰۰۰	۷
	۰۰:۱۴:۱۵	۲۰۶۵	۷

منظور از RMSE روشی برای برآورد میزان خطاست که در واقع تفاوت بین مقادیر تخمینی و آنچه تخمین زده شده است. RMSE به دو دلیل تقریباً همه جا مثبت است (صفر نیست) یک اینکه تصادفی است و دوم به این دلیل که تخمین‌گر اطلاعاتی که قابلیت تولید تخمین دقیق‌تری دارد را حساب نمی‌کند. پس این

[1] Average Pooling
[2] Learning Rate

شاخص که مقداری همواره نامنفی دارد، هر چقدر مقدار آن به صفر نزدیک‌تر باشد، نشان دهنده میزان کمتر خطاست.

شبکه عصبی به دنبال یک سری وزن‌ها (W) و انحراف (b) هست تا بتواند خطوطی مرزی بین انواع مختلف (مثلاً پراید و اتوبوس) رسم کرده و به وسیله آن‌ها، تفاوت‌ها را یاد گرفته و درک کند. این کار با استفاده از قانون ضرب و توابع فعال‌سازی [1] انجام می‌شود.

توابع فعال‌سازی در انتهای واحد پنهان برای معرفی پیچیدگی غیر خطی به مدل استفاده می‌شوند.

تابع اتلاف در واقع میزان خطا در هر بار اجرای شبکه عصبی را برای داده‌های آموزشی نمایش می‌دهد. تابع اتلاف از طریق فرمول زیر محاسبه می‌گردد:

$$\text{Loss}=1/N \ \sum_{i=1}^{N}(y_P - y_t)^2 \qquad (۱)$$

خروجی مطلوب یعنی y_t از خروجی شبکه یعنی y_p کم شده سپس به توان ۲ رسانده شده است. به ازای هر نمونه، یک خروجی (y_p) دارید. به ازای هر خروجی، یک خطا خواهید داشت. سپس مقدار خطاهای همه نمونه‌ها با هم جمع شده و خطای میانگین شبکه محاسبه شده است. در واقع کار اصلی شبکه عصبی کم کردن مقدار تابع ضرر تا حد ممکن است. طبیعتاً ایده‌آل برای شبکه عصبی مقدار ۰ است.

روش پیشنهادی میزان خطای کمتری نسبت به روش‌های معرفی شده دارد و این به دلیل دقت روش پیشنهادی استفاده شده در آن می‌باشد.

دقت مجموعه‌ای از اندازه‌گیری‌ها، به‌صورت میزان نزدیک بودن نتایج آن اندازه‌گیری‌ها به یکدیگر، تعریف می‌شود. در واقع، هر چه نتایج به‌دست آمده فاصله کمتری با یکدیگر داشته باشند، دقت آن سامانه بیشتر است. دقت شبکه عصبی از طریق فرمول زیر محاسبه می‌گردد:

$$\text{Accuracy}=TR/TS \qquad (۲)$$

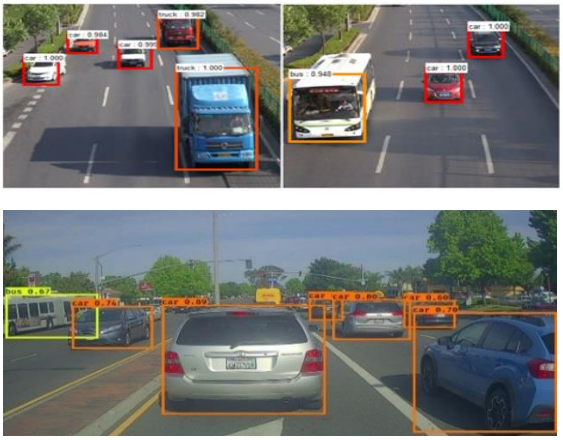
در روش پیشنهادی به دلیل اینکه از شبکه عصبی کانولوشن VGG به همراه الگوریتم YOLO استفاده شده است، سرعت پردازش محاسبات افزایش یافته و در نتیجه باعث بهبود عملکرد دقت (تشخیص شیء) نسبت به روش‌های پیشین گردد. معیار صحت [2] معیاری است که به ما می‌گوید الگوریتم چند درصد «بلی»هایش درست بوده است و از طریق فرمول زیر محاسبه می‌گردد:

$$\text{Precision}=TR/TR+FR \qquad (۳)$$

در روش پیشنهادی الگوریتم YOLO، از هر ۱۰۰ تصویر حدود ۹۷ تصویر را به درستی تشخیص داده است، در واقع معیار صحت حدود ۹۷ درصد می‌باشد. بنابراین تشخیص شیء مورد نظر در روش پیشنهادی نسبت به روش‌های پیشین افزایش یافته است.

در روش‌های پیشین جداسازی وسایل نقلیه از پس‌زمینه تصاویر خطایی در حدود ۱۰درصد داشت، اما در رویکرد ارائه شده جداسازی وسایل نقلیه از پس‌زمینه تصاویر خطایی حدود ۳ درصد دارد که نسبت به بقیه روش‌ها خطای خیلی کمتری دارد.

خروجی نتایج شبیه‌سازی در شکل (۶) آمده است.



شکل (۶): نتایج شبیه‌سازی (از دیتاست مربوط به پایگاه ImageNet)

شکل (۷): نتایج تصاویر در شب و تصاویر با نویز (از دیتاست مربوط به پایگاه ImageNet)

[1] Activation Function
[2] Precision

در این روش حدود ۷۷۳ تصویر در شب و ۳۶۱ تصـویر در حالـت نویز از دو پایگاه داده ImageNet و COCO استفاده شده است، که با توجه به نتایج شکل (۷) مشاهده می‌گردد که رویکرد ارائه شده جهت تشخیص وسایل نقلیه در شب یا تصـاویری کـه دارای نـویز هستند، دارای دقت بالای ۹۶٪ می‌باشد. اما دارای اتلافی نیز می-باشد که علت اتلاف مشابه بودن تصاویر چهار کلاس می‌باشد کـه باعث به اشتباه افتادن شبکه عصبی مـی‌گـردد، امـا در قیـاس بـا روش‌های معرفی شده این روش کارایی بالاتری را در چالش‌هـای معرفی شده داراست و دقت در شناسایی نوع و مکان‌یـابی وسـایل نقلیه افزایش و بار محاسباتی کاهش یافته است که باعث سـرعت پردازش در شناسایی وسایل نقلیه می‌گردد.

۵- نتیجه‌گیری

تشخیص وسیله نقلیه در بسیار از موارد همچون سامانه‌های کنترل ترافیک (برای مثال بررسی حجم ترافیک یا شناسایی متخلفین)، برنامه‌ریزی شهری، خودروهای بدون راننده، مدیریت پارکینگ‌ها با اعلام ورود وسایل نقلیه به مکان مورد نظر توسط ماشین، در صورتی که ویژگی‌های خودرو مانند رنگ، مدل، شماره پلاک و غیره را بتوان توسط ماشین شناسایی کرد می‌توان برای تشخیص خودروی سرقت شده، متخلف و یا هر موردی که در تعقیب است از آن استفاده کرد، در عوارضی‌ها برای جلوگیری از تخلفات کارمندان کاربرد دارد.

روش‌های گوناگون جهت تشخیص وسایل نقلیه در تصویر و یا به‌طور کلی اشیاء در تصویر وجود دارد. این روش‌ها به سه گروه سنتی، استخراج ویژگی‌ها و استفاده از شبکه عصبی تقسیم شده-اند.

با استفاده از روش‌های بالا چالش‌هایی همچون: تعدد تصویر، تغییرات نور، گوناگونی در انواع وسایل نقلیه، تصاویر بسیار متنوع از زوایای متفاوت، ایجاد مانع در برابر شیء و خطا در تشخیص اهدافی که دارای سرعت بالایی هستند، تقریباً رفع گردید. اما چالش بار محاسباتی هنوز باقی ماند. با ارائه رویکرد جدید که در این مقاله به آن پرداخته شد، تقریباً تمامی چالش‌های باقی مانده رفع گردید و بار محاسباتی نیز کاهش یافت.

در رویکرد ارائه شده به دلیل استفاده از شبکه عصبی VGG و الگوریتم YOLO، بار محاسباتی کاهش یافته، که باعث افزایش سرعت شناسایی شیء می‌گردد. در نتایج شبیه‌سازی بهبود عملکرد پردازش محاسابات نشان دهنده یکی از جنبه‌های نوآوری در رویکرد ارائه شده می‌باشد. در ضمن تشخیص وسایل نقلیه‌ای که در شب وجود دارند و بازیابی تصاویری وسایل نقلیه‌ای که دارای نویز هستند، یکی از محاسن رویکرد ارائه شده می‌باشد. با توجه به نتایج شبیه‌سازی دقت تشخیص وسیله نقلیه نیز نسبت به بقیه روش‌های معرفی شده افزایش یافته و حدود ۹۷ درصد رسیده است.

۶- مراجع

[1] S. Yamaguchi, "A Car Detection System Using the Neocognitron," IEEE International Joint Conference on, Singapore, 2002.

[2] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle Detection in Aerial Images Based on Region Convolutional Neural Networks and Hard Negative Example Mining," Sensor, 2017.

[3] B. D. Stewart, I. Reading, M. S. Thomson, T. D. Binnie, K. W. Dickinson, and C. L. Wan, "Adaptive Lane Finding in Road Traffic Image Analysis," Proceedings of Seventh International Conference on Road Traffic Monitoring and Control, IEEE, London, 1994.

[4] W. Enkelmann, "Obstacle Detection by Evaluation of Optical Flow Field from Image Sequences," Proceedings of European Conference on Computer Vision, Antibes, France 427, pp. 134–138, 1990.

[5] Y. Park, "Shape-resolving Local Thresholding for Object Detection," Pattern Recognition Letters 22, pp. 883–890, 2001.

[6] J. M. Blosseville, C. Krafft, F. Lenoir, V. Motyka, and S. Beucher, "New Traffic Measurements by Image Processing," IFAC Transportation Systems, Tianjin, Proceedings, 1994.

[7] Y. Won, J. Nam, and B. H. Lee, "Image Pattern Recognition in Natural Environment Using Morphological Feature Extraction", In: F.J. Ferri (Ed.), SSPR&SPR 2000, Springer, Berlin, pp.806–815, 2001.

[8] J. B. Kim, H. S. Park, M. H. Park, and H. J. Kim, "A Real-time Region-based Motion Segmentation Using Adaptive Thresholding and K-means Clustering," In: M. Brooks, D. Corbett, M. Stumptner (Eds.), AI 2001, Springer, Berlin, pp. 213–224, 2001.

[9] M. Dubuisson and A. Jain, "Contour Extraction of Moving Objects in Complex Outdoor Scenes", International Journal of Computer Vision, vol. 14, pp. 83–105, 1995.

[10] H. Bensedik, A. Azough and M. Meknasssi, "Vehicle Type Classification Using Convolutional Neural Network", IEEE 5th International Congress on Information Science and Technology (CiSt), 2018.

[11] Y. Li, B. Song, X .Kang , X .Du ,and M .Guizani , "Vehicle-type Detection Based on Compressed Sensing and Deep Learning in Vehicular Networks," Sensors, ۲۰۱۸.

[12] A. Techmer, "Real-time Motion Based Vehicle Segmentation in Traffic Lanes", In: B. Radig, S. Florczyk (Eds.), DAGM, Springer, Berlin, pp. 202–207, 2001.

[13] M. A. Ali, H. E. Abd El Munim,. A. H. Yousef,and Sh. Hammad, "A Deep Learning Approach for Vehicle Detection," 13th International Conference on Computer Engineering and Systems (ICCES), 2018.

[14] A. Giachetti, M. Campani, and V. Torre, (). "The Use of Optical Flow for Road Navigation," IEEE Transactions on Robotics and Automation, vol. 14, no. 1, 2000.

[15] X. Li, Z. Q. Liu, and K. M. Leung, "Detection of Vehicles from Traffic Scenes Using Fuzzy Integrals," Pattern Recognition, vol. 35, pp. 967–980, ۲۰۰۲.

[16] M. Oliveria and V. Santos, "Automatic Detection of Cars in Real Roads Using Haar-like Features," Department of Mechanical Engineering, University of Aveiro, Portugal, 2008.

[17] M. Nasehi, M. Ashourian, and P. Moalem, "An Overview of the Type of Vehicle Detection Techniques," Majlesi Journal of Telecommunication Devices, vol. 9, no. 3, 2020.

[18] G. D. Sullivan, K. D. Baker, A. D. Worrall, C. I. Attwood, and P. M. Remagnino, () "Model-based Vehicle Detection and Classification Using Orthographic Approximations," Image and Vision Computing, vol. 15, pp. 649–654, 2004.

[19] Z .Zhigang ,L .Huan ,D. Pengcheng, Z. Guangbing, W. Nan , and Z. Wei-Kun., "Vehicle Target Detection Based on R-FCN," Chinese Control And Decision Conference (CCDC), 2018.

[20] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang."Real-time Vehicle Type Classification with Deep Convolutional Neural Networks," Journal of Real-Time Image Processing, 2019.

[21] Z. Zhang, C. Xu, and W. Feng , "Road Vehicle Detection and Classification Based on Deep Neural Network"7th IEEE International Conference on Software Engineering and Service Science (ICSESS), 2016.

[22] Sh. Yu, Y. Wu, W. Li, Z. Song, and W. Zeng,"A Model for Fine-Grained Vehicle Classification Based on Deep Learning", Neurocomputing, , 2017.

[23] L. Suhao,, L. Jinzhao, L. Guoquan, B. Tong, W. Huiqian, and P. Yu, "Vehicle Type Detection Based on Deep Learning in Traffic Scene," Procedia Computer Science, 2018.

[24] M. Sheng, Ch. Liu, Q. Zhang, L. Lou, and Y. Zheng, "Vehicle Detection and Classification Using Convolutional Neural Networks," IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), 2018.

[25] A. Murali,, B. B Nair, and S. N. Rao, "Comparative Study of Different CNNs for Vehicle Classification," IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.

[26] H. Moon, R. Chellapa, and A. Rosenfeld, "Performance Analysis of a Simple Vehicle Detection Algorithm", Image and Vision Computing, vol. 20, pp. 1–13, 2003.

[27] T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields", Signal Processing: Image Communication, vol. 7, pp. 147–160, 1995.

[28] J. Lan, J. Li, G. Hu, B. Ran, and L. Wang, "Vehicle Speed Measurement Based on Gray Constraint Optical Flow Algorithm," Optic, vol. 125, pp. 289-295, 2014.

J. Li, F. Yang, M. Tomizuka, and C. Choi "Evolvegraph: Multi-agent trajectory Prediction with Dynamic Relational Reasoning," Proceedings of the Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada. 6–12 December 2022.