مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال نهم، شماره دوم، تابستان ۱۴۰۱

شناسایی خودروهای اضطراری مبتنی بر یادگیری عمیق به منظور استفاده در خودروهای بدون راننده

مریم اسدی[۱] و عبداله چاله چاله[۲]

چکیده

هدف از طراحی و ساخت خودروهای بدون راننده حذف عامل انسانی به منظور کاهش تلفات، هزینه‌ها و نیز افزایش ایمنی خودرو با جایگزینی تجهیزات هوشمند است. امروزه با بهره‌مندی از فناوری‌های هوش مصنوعی و یادگیری ماشین شاهد پیشرفت‌های چشم-گیری در صنعت حمل و نقل هوشمند به ویژه خودروهای تمام خودکار هستیم که با استفاده از حسگرهای پیشرفته و تکنیک بینایی ماشین قادر به تجزیه و تحلیل اطلاعات محیط پیرامون خود هستند. از چالش‌های مطرح در طراحی سیستم این نوع از خودروها، شناسایی درست سایر وسایل نقلیه‌ی اطراف مسیر حرکت خودرو است. در این مقاله، برای شناسایی خودروهای اضطراری یك روش مبتنی بر یادگیری عمیق ارائه شده است که فرایندهای استخراج ویژگی و طبقه‌بندی درآن به صورت همزمان انجام می‌شود. شبکه عمیق مورد استفاده در این پژوهش شبکه پیچشی می‌باشد. در شبکه‌های عصبی پیچشی دستیابی به نتایج قابل قبول وعملکرد مناسب، مستلزم در اختیار داشتن حجم عظیمی از داده‌ها برای آموزش شبکه می‌باشد. با توجه به محدود بودن تعداد تصاویر موجود در مجموعه داده مورد استفاده در این پژوهش و به منظور افزایش دقت شناسایی، از فرایند یادگیری انتقالی و شبکه پیش‌آموزش‌دیده VGG16 نیز استفاده شده است. برای این تحقیق دو مجموع داده جدید ایجاد و در کنار دو مجموعه داده دیگر مورد آزمایش قرار گرفت. روش پیشنهادی با چهار روش دیگر نیز مورد مقایسه قرار گرفت و نتایج به دست آمده نمایانگر کارایی بسیار خوب روش پیشنهادی است.

کلیدواژه‌ها: خودروهای بدون راننده، خودروهای اضطراری، یادگیری عمیق، پردازش تصویر

۱ مقدمه

از زمان تولید اولین خودرو تا به امروز، روشن شدن موتور، تعویض دنده، باز و بسته کردن درب‌ها و پنجره‌های خودرو دست‌خوش تغییرات بسیاری گردیده است و همواره تلاش‌هایی در جهت آسان نمودن و به نحوی خودکار نمودن همه این فرآیندها صورت پذیرفته است. اما عمل رانندگی، بسیار کم تغییر کرده است، خودکار شدن عمل رانندگی نشان‌دهنده یک پتانسیل ایمنی عظیم است، اما چالش‌های بسیاری را نیز به همراه دارد. رایانه‌ها می‌توانند مقادیر بیشتری از اطلاعات را برای تصمیم‌گیری سریع‌تر در مقایسه با انسان پردازش کنند. از سوی دیگر، توسعه این فناوری‌ها زمان‌بر است، زیرا محیط رانندگی پیچیده و هزینه‌های انسانی و اقتصادی خطاها زیاد است [۱].

بینایی ماشین، فناوری کلیدی است که در خودروهای بدون راننده از آن استفاده می‌شود. در این نوع از خودروها از دوربین‌ها، حسگرها و یک رایانه برای بررسی جاده و محیط اطراف استفاده می‌شود. عملیات پایه در خودروهای بدون راننده عبارت‌اند از: ثبت تصاویر توسط دوربین‌ها، پردازش تصاویر و در نهایت استخراج اطلاعاتی مانند علائم ترافیکی یا موانع در مسیر خودرو. عملیاتی که این نوع از ماشین‌ها انجام می‌دهند باید بر اساس قوانین راهنمایی و رانندگی باشد مانند رعایت سرعت مجاز و توجه به علائم راهنمایی. بنابراین مستلزم داشتن دانش بیشتری هستند که بصورت هوشمندانه عمل نموده و بتوانند در شرایط خاص عکس‌العمل درستی انجام دهند. به طور مثال هنگامی‌که چراغ سبز است اما یک عابر پیاده در تقاطع حضور دارد، خودرو باید بتواند توقف کامل نماید [۲]. جزییات بیشتری در خصوص ویژگی‌های ماشین خودران در مقاله [۳] آورده شده است.

[۱] دانشجوی کارشناسی ارشد مهندسی کامپیوتر (نرم افزار)، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه رازی
رایانامه: maryam.asadi9613@gmail.com
[۲] گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی، دانشگاه رازی
رایانامه: Chalechale@razi.ac.ir

نویسنده مسئول: عبداله چاله‌چاله
DOR: 20.1001.1.23831197.1401.9.2.2.5

طراحی سیستمی که بتواند خودروهای اضطراری را با دقت بالایی شناسایی کند بسیار حائز اهمیت است. روش‌های مختلفی برای حل این مسئله وجود دارد. یکی از جدیدترین روش‌ها استفاده از تکنیک یادگیری عمیق است، با توجه به حجم ورودی‌های سیستم در این نوع خودروها، از تکنولوژی یادگیری عمیق در شناسایی اشیاء استفاده می‌شود و بعد از شناسایی یك خودروی اضطراری در محیط اطراف اقدامات مناسب انجام می‌شود. با توجه به اهمیت رفت و آمد این نوع خودروها در خیابان‌ها و جاده‌ها، شناسایی آنها توسط خودروهای بدون راننده از اهمیت بسیار زیادی برخوردار است.

در این راستا، در این پژوهش روشی برای شناسایی سه نوع خودروی اضطراری شامل خودروی آمبولانس، خودروی آتش‌نشانی و خودروی پلیس از سایر خودروهای درون جاده ارائه می‌شود. در این مقاله، برای شناسایی خودروهای اضطراری یك چارچوب مبتنی بر یادگیری عمیق ارائه شده است که مراحل استخراج ویژگی و طبقه‌بندی در آن به صورت همزمان انجام می‌شود. شبکه عمیق مورد استفاده در این پژوهش شبکه پیچشی می‌باشد، در شبکه‌های عصبی پیچشی دستیابی به نتایج قابل قبول و عملکرد مناسب، مستلزم در اختیار داشتن حجم عظیمی از داده‌ها برای آموزش شبکه است. با توجه به محدود بودن تعداد تصاویر موجود در مجموعه داده مورد استفاده در این پژوهش و به منظور افزایش دقت شناسایی، از فرایند یادگیری انتقالی و شبکه پیش‌آموزش‌دیده VGG16استفاده شده است. بدیهی است با توجه به تنوع و همه‌منظوره بودن مجموعه داده آموزشی ImageNet [۴] بارگذاری وزن‌های یادگیری شده برای اهداف مختلف در مجموعه داده‌های دیگر امکان‌پذیر بوده و موجب صرفه‌جویی در زمان یادگیری شبکه و تسریع محاسبات خواهد شد. براي اين تحقيق دو مجموع داده جديد ايجاد و در كنار دو مجموعه داده ديگر مورد آزمايش قرار گرفت. روش پیشنهادی با چهار روش دیگر نیز مورد مقایسه قرار گرفت و نتایج به دست آمده نمایانگر کارایی بسیار خوب روش پیشنهادی است.

ادامه این مقاله به صورت زیر سازماندهی شده است: بخش دوم مروری است بر کارهای پیشین در شناسایی و طبقه‌بندی خودروها. در بخش سوم مفاهیم پایه از جمله یادگیری عمیق و الگوریتم‌های مطرح شده در این خصوص توضیح داده شده است. در بخش چهارم جزئیات روش پیشنهادی آورده شده و پس از آن معیارهای ارزیابی، ارزیابی روش پیشنهادی و تحلیل نتایج در بخش پنجم ذکر شده است. نهایتاً بخش آخر نیز به نتیجه‌گیری و پیشنهادهای آینده اختصاص داده شده است.

۲ پیشینه تحقیق

شناسایی نوع خودرو در جاده نقش مهمی در کاربردهای مربوط به سیستم‌های حمل و نقل هوشمند جهت افزایش امنیت و سهولت تردد وسایل نقلیه در جاده‌ها ایفا می‌کند. شناسایی نوع خودرو می‌تواند به دو شکل انجام شود: (۱) نوع خودرو تنها از نظر ابعاد و ویژگی‌های کلی و ظاهری آن تعیین شود (تقسیم‌بندی به گروه‌های سواری، اضطراری و غیره) و (۲) به طور جزئی‌تر، مدل و شرکت سازنده خودرو هم تعیین شود. در این مقاله روش پیشنهادی متمرکز بر شکل اول می‌باشد.

امروزه سیستم‌های نظارت بر ترافیک نقش مهمی را در سیستم‌های حمل و نقل هوشمند ایفا می‌کنند. معمولاً اولین مرحله در یک سیستم نظارتی این است که اشیاء را از تصاویر یا ویدئو تشخیص داده و آنها را به دسته‌های مختلف طبقه‌بندی کرد [۵]. در دهه‌های اخیر، از رویکردهای مبتنی بر سخت‌افزار و نرم‌افزار برای طبقه‌بندی وسایل نقلیه استفاده شده است. به عنوان مثال، تشخیص و طبقه‌بندی با استفاده ازسنسورهای مادون قرمز، اولتراسونیک و رادار مورد بررسی قرار گرفته است. روش‌های مبتنی بر سخت‌افزار معایبی همچون آسیب‌پذیری و هزینهٔ بالا برای نصب و پیاده‌سازی دارند. در مقابل روش‌های مبتنی بر پردازش تصویر نگهداری ساده داشته و نصب و پیاده‌سازی آنها نیز کم هزینه است. طبقه‌بندی براساس تصاویر در طی سالیان متمادی، انجام شده است. در حالت کلی، شناسایی نوع وسایل نقلیه بر اساس تصاویر را می‌توان به دو مرحله تقسیم کرد: ۱) استخراج ویژگی‌ها و ۲) طراحی طبقه‌بند. استخراج ویژگی نقش کلیدی در تشخیص و طبقه‌بندی وسیله نقلیه دارد. روش‌های متداول زیادی برای استخراج ویژگی‌های تصویر وسایل نقلیه وجود دارد. در روش‌های متعارف، ویژگی‌های استخراج شده شامل ویژگی‌های هندسی مانند طول و عرض [۶]، رنگ و ابعاد وسیله نقلیه می‌باشد. سپس برای طبقه‌بندی ویژگی‌های استخراج شده از روش‌هایی مانند ماشین بردار پشتیبان[1] (SVM) [۷]، K نزدیک‌ترین همسایه[2] (KNN) [۸] و درخت تصمیم استفاده می‌شود. در ادامه، پژوهشگران به بررسی روش‌های مختلف استخراج ویژگی و نمایش اشیاء پرداختند. به عنوان نمونه، در [۹] برای شناسایی ویژگی‌های وسایل نقلیه روشی مبتنی بر هریس[3] پیشنهاد شد. این رویکرد به صورت بازگشتی تصویر را به چهار قسمت تقسیم می‌کند، نقاط قوت ویژگی در این قسمت‌ها تجمیع و سپس با روش بازگشتی و سلسله‌مراتبی بصورت محلی نرمال‌سازی می‌شوند. در این پژوهش دو روش طبقه‌بندی مختلف بررسی شده است: KNN و Naive Bayes. سیستم پیشنهادی قادر به طبقه‌بندی وسایل نقلیه با دقت ۹۶٫۰٪ است. مطالعه [۱۰] استفاده از توصیف‌گر SURF[4] را برای تشخیص و شناسایی مدل وسیله نقلیه در جاده معرفی می‌کند.

[1] Support Vector Machines - SVMs
[2] k-Nearest Neighbors
[3] Harris
[4] Speeded Up Robust Features

همچنین از هیستوگرام گرادیان جهت‌دار[1] (HOG) برای استخراج ویژگی‌ها و از ماشین بردار پشتیبان (SVM) چندکلاسه برای شناسایی چهار کلاس تصویر استفاده شده است [۱۱]. در [۱۲]، استفاده از تبدیل موجک گابور[2] و هرم هیستوگرام گرادیان-های جهت‌دار[3] (PHOG) بررسی شده است.

در سال‌های اخیر، تکنولوژی یادگیری عمیق[4] به سرعت توسعه یافته است و برخلاف روش‌های سنتی که مبتنی بر ویژگی‌های دستی هستند، روش‌های مبتنی بر یادگیری عمیق قادر به یادگیری ویژگی‌های اشیاء در چندین مرحله و بصورت خودکار می‌باشند. این رویکرد در سالهای اخیر کاربردهای فراوانی در مسائلی مانند تشخیص چهره، تشخیص نقاط صورت، تشخیص عابر پیاده، ارزیابی کیفیت تصویر، طبقه‌بندی تصویر و ویدئو داشته است. برای طبقه‌بندی وسایل نقلیه نیز پژوهش‌هایی مبتنی بر شبکه‌های عصبی عمیق انجام شده است. ژن دونگ و همکاران [۱۳]، برای طبقه‌بندی نوع وسیله نقلیه یک شبکه عصبی پیچشی[5] نیمه نظارتی[6] نظارتی[6] پیشنهاد کردند که از تصاویر نمای جلوی خودرو استفاده می‌کند. به منظور کسب اطلاعات غنی و متمایز از وسایل نقلیه، در این روش از فیلتر لاپلاس[7] با مقادیر زیادی از داده‌های بدون برچسب استفاده شده است. به عنوان لایه خروجی شبکه، طبقه‌بند Softmax با یادگیری چندوظیفه‌ای با مقادیر کمی از داده‌های دارای برچسب آموزش داده می‌شود. در سال ۲۰۱۵، یک روش جدید مبتنی بر شبکه عصبی پیچشی و ناحیه‌بندی تصویر به نام Faster R-CNN ارائه شد [۱۴] که نتایج بسیار خوبی در شناسایی اشیاء از خود نشان داده است و در چند سال اخیر به عنوان یکی از روش‌های برجسته و پایه‌ای برای مسائل شناسایی به کار گرفته شده است [۱۵] تا [۱۶]. از این رو، در پژوهش [۱۷]، یک سیستم طبقه‌بندی نوع وسیله نقلیه در زمان واقعی براساس روش Faster R-CNN ارائه شده است. نتایج نشان می‌دهند که استفاده از این روش نه تنها باعث صرفه جویی در وقت بلکه دقت بالایی را نیز به همراه دارد. با توجه به اینکه روش‌های مبتنی بر یادگیری عمیق برای رسیدن به نتیجه مطلوب نیازمند داده‌های آموزشی بسیار زیادی هستند و در دنیای واقعی برای بسیاری از مسائل کاربردی داده‌های زیادی در دسترس نیستند میتوان از تکنیک‌هایی همچون تقویت داده [۱۸] و شبکه‌های پیش‌آموزش دیده استفاده کرد. لی ژو و همکاران در سال ۲۰۱۷ با استفاده از معماری گوگل‌نت که بر روی مجموعه داده ILSVRC-2012 آموزش دیده بود طبقه‌بندی وسایل نقلیه بر روی فیلم‌های نظارت بر ترافیک را انجام دادند [۱۹]. در مطالعه [۲۰] نیز از روش تقویت داده برای طبقه‌بندی داده‌های نامتوازن استفاده می‌شود. استخراج ویژگی مبتنی بر یادگیری عمیق و استفاده از روش‌های مبتنی بر یادگیری ماشین همچون ماشین بردار پشتیبان به عنوان طبقه‌بند نیز نتایج خوبی را نشان داده است [۲۱]. روش مورد استفاده در [۲۲] تا [۲۳] مبتنی بر متن هستند و با توجه به مشخصه‌های موجود بر بدنهٔ خودروهای اضطراری از جمله خودروی آمبولانس، شناسایی این نوع خودروها را مورد بررسی قرار می‌دهند. در پژوهش [۲۴] برای تشخیص خودروهای اضطراری از شبکه پیش‌آموزش‌دیده VGG16 [۲۵] و معماری Yolo [۲۶] استفاده شده است. در این مقاله برای شناسایی دقیقتر و سریعتر خوردوهای اضطراری از شبکه عمیق پیچشی استفاده شده است که کار استخراج ویژگی و طبقه‌بندی به صورت همزمان در یک مرحله انجام می‌شود.

۳ مفاهیم پایه

در این بخش، براي درک بهتر الگوریتم پیشنهادی، مفاهیم شبکه-هاي عصبی پیچشی، معماری VGGNet و فرایند یادگیری انتقالی بیان شده است.

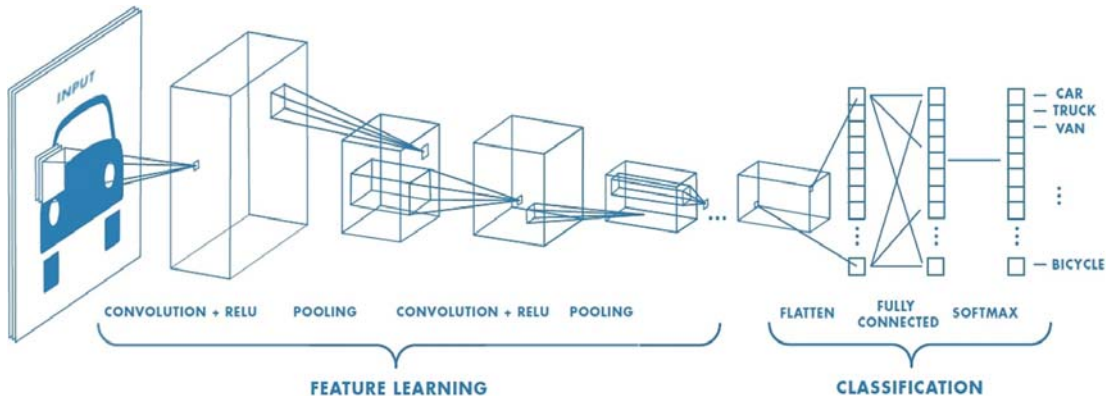
۳-۱ یادگیری عمیق

یادگیری عمیق از تکنیک‌های مبتنی بر شبکه‌های عصبی می‌باشد که به عنوان زیرمجموعه‌ای از یادگیری ماشین به شمار می‌رود. در سالهای اخیر معماریهای یادگیری عمیق متنوعی به وجود آمده‌اند که از جملهٔ آنها می‌توان به شبکه عصبی پیچشی [۲۷] (CNN)، شبکه باور عمیق[8] [۲۸] و شبکه عصبی بازگشت‌کننده[9] [۲۹] اشاره اشاره کرد. کارآمدی و موثر بودن معماریهای معرفی شده در حوزه-های متنوعی از جمله تشخیص اشیاء[10] [۳۰]، تشخیص خودکار گفتار[11] [۳۱]، شناسایی چهره[12] [۲۷] به اثبات رسیده است. هدف یادگیری عمیق، استخراج هوشمندانه ویژگیها طی چندین مرحله یادگیری است [۳۲].

در شبکه‌های مختلف ارائه شده مبتنی بر یادگیری عمیق، شبکه عصبی پیچشی نتایج بسیار خوبی در حوزه تصویر از خود نشان داده است. در این پژوهش روش ارائه شده، مبتنی بر شبکه عصبی پیچشی است. بنابراین در ادامه این بخش، مروری بر مفاهیم شبکه عصبی عمیق و تعدادی از معماری‌های برجسته آن خواهیم داشت.

شبکه عصبی پیچشی یک شبکه نظارت شده است که عملکرد بسیار خوبی را در بسیاری از کارهای طبقه‌بندی چندکلاسه نشان داده است.

[1] Histogram of Oriented Gradients
[2] Gabor Wavelet Transform
[3] Pyramid Histogram of Oriented Gradients
[4] Deep Learning
[5] Convolution Neural Network (CNN)
[6] Semi Supervised
[7] Laplace
[8] Deep Belief Network
[9] Recurrent Neural Network
[10] Object Detection
[11] Automatic Speech Recognition
[12] Face Recognition



شکل ۱ نمایی از یک معماری CNN برای مسئله طبقه‌بندی [۳۳]

در مقایسه با روش‌هاي سنتی استخراج ویژگی، با استفاده از CNN ویژگي‌های هر کلاس را مي‌توان به طور خودکار از تصاویر ورودي بدست آورد و نتایج طبقه‌بندي مستقیماً بدست مي‌آید. ویژگی‌های استخراج شده خاصیت سلسله‌مراتبی دارند به این معنی که در لایه‌های ابتدایی، بطور مثال گوشه‌ها، خطها و لبه‌ها، یاد گرفته می‌شوند و در لایه‌های عمیق‌تر به ترتیب ویژگی‌هایی با سطوح بالاتر، همانند اشیاء درون تصویر استخراج می‌شوند. علاوه بر این، استخراج ویژگی و طبقه‌بندی در همان شبکه پیاده‌سازی می‌شود. یک ساختار از CNN در شکل ۱ نشان داده شده است که از سه لایه تشکیل شده است: لایه کانولوشن[۱]، لایه ادغام[۲] و لایه تماماً متصل[۳] [۳۴].

شبکه عصبی پیچشی از دو بخش اصلی تشکیل شده است: بخش یادگیری ویژگی‌ها و بخش طبقه‌بندی. CNN به صورت یک سری لایه‌های بهم پیوسته اجرا می‌شود که این لایه‌ها از بلوک‌های پی در پی کانولوشن، ReLU و لایه‌های ادغام تشکیل شده‌اند.

ReLU لایه غیرخطی را به شبکه اضافه می‌کند تا شبکه را قادر سازد ترکیب‌های غیرخطی از ورودی‌های اصلی را که به آن استخراج ویژگی گفته می‌شود بیاموزد. این ویژگی‌های آموخته شده، به عنوان یک لایه فعال‌سازی شناخته می‌شوند و به عنوان ورودی به لایه بعدی وارد می‌شوند. لایه‌های ادغام از ورودی‌های خود نمونه‌برداری می‌کنند و به تلفیق ویژگی‌های تصویر محلی کمک می‌کنند. سرانجام، ویژگی‌های آموخته شده به ورودی‌های طبقه‌بندی شده یا تابع رگرسیون در انتهای شبکه تبدیل اعمال می‌شوند. برای مسائل طبقه‌بندی تصویر، لایه آخر یک طبقه‌بند است [۳۲]. در شبکه‌های عصبی پیچشی برای طبقه‌بندی چندکلاسه به طور معمول از تابع هزینه Cross_entropy استفاده می‌گردد که به صورت رابطه ۱ تعریف می‌گردد.

$$H_{y'}(y) = -\sum y_i' \log(y_i) \qquad (۱)$$

که در آن y' مقدار واقعی و y مقدار محاسبه شده برای نمونهٔ شماره i می‌باشد. پس از محاسبهٔ میزان خطا توسط تابع هزینه، مقدار تصحیح در هر پارامتر قابل تنظیم در شبکه، از طریق روش پس‌انتشار خطا و قاعدهٔ مشتق زنجیره‌ای محاسبه می‌گردد و مقدار تصحیح به پارامتر مورد نظر اعمال می‌گردد. این فرایند به صورت تکراری به تعداد دفعات مشخص شده برای شبکه انجام می‌گیرد.

در رابطه ۲، L تابع هزینه، ω_i وزن تنظیمی برای نمونه i ام و λ نرخ آموزش تعیین شده برای شبکه است.

$$\omega_i - \lambda \frac{\partial L}{\partial \omega_i} \rightarrow \omega_i \qquad (۲)$$

۲-۳ معماری VGGNet

از جمله معماری‌های مفید در حوزه پردازش و استخراج ویژگی از تصاویر، معماری VGGNet است که مبتنی بر شبکه عصبی پیچشی عمیق است و توسط سیمونیان[۴] و اندروزیسزمن[۵] توسعه داده شده است. این معماری به عنوان دومین شبکه پیشنهادی برنده مسابقات ILSVRC2014 در طبقه‌بندی تصاویر[۶] اما با بهترین عملکرد در عملیات تشخیص اشیاء نسبت به دیگر معماری‌های ارائه شده در آن رقابت انتخاب شده است [۲۵]. نسخه از پیش-آموزش‌دیده شده از این شبکه، توسط ۱,۲ میلیون تصویر موجود در مجموعه داده آموزشی ImageNet با رزولوشن بالا و اندازه ۲۲۷ × ۲۲۷ در ۱۰۰۰ کلاس پیاده‌سازی شده است. حداقل عمق این شبکه دارای ۱۱ لایه است که شامل هشت لایه پیچشی و سه لایه تماماً متصل می‌باشد و حداکثر عمق آن ۱۹ لایه است که دارای

[1] Convolutional
[2] Pooling
[3] Fully-Connected
[4] Karen Simonyan
[5] Andrew Zisserman
[6] Image Classification

۱۶ لایه پیچشی و سه لایه تماماً متصل است. ویژگی شاخص و متمایز این معماری بالا بودن عمق شبکه به عنوان مولفه‌ای حیاتی است که موجب عملکرد مناسب این شبکه شده است. بهترین نسخه این شبکه که شامل ۱۶ لایه پیچشی/تماماً متصل بوده، در این پژوهش مورد استفاده قرار می‌گیرد.

این شبکه در وظایف یادگیری انتقالی از بسیاری از شبکه‌ها از جمله گوگل‌نت[1] بهتر عمل می‌کند. شبکه VGGNet یکی از محبوب‌ترین شبکه‌ها برای استخراج ویژگی از تصاویر می‌باشد. از این رو در این پژوهش، با استفاده از ویژگی‌ها و برچسب‌های استخراج شده از شبکه VGGNet-16 براي طبقه‌بندی تصاویر استفاده می‌شود. در ادامه به معرفی مختصر دو معماری [۳۵] ResNet و گوگل نت [۳۶] که در چند سال اخیر پیشنهاد شده‌اند، می‌پردازیم:

- **معماری ResNet**

در سال ۲۰۱۵ تعدادی از محققین مایکروسافت معماری ResNet را ارائه دادند که خطا را به ۳/۶ درصد کاهش داد. این معماری از ۱۵۲ لایه تشکیل شده است که یک رکورد در هر سه بخش تشخیص اشیاء، طبقه‌بندی و مکان‌یابی با استفاده از یک مدل بجا گذاشت. این مدل در حقیقت قدرت انسان در تشخیص اشیاء را جا گذاشت و با دقت به مراتب بالاتر از انسان اشیاء را تشخیص داد.

ایده اصلی یک بلوک residual این است که ورودی توسط یک لایه پیچشی، لایه ReLU و یک لایه پیچشی پردازش می‌شود، نتیجه این تبدیلات یک تابع F(x) تولید می‌کند. نتیجه این تابع با ورودی لایه قبل جمع می‌شود [۳۵].

- **معماری گوگل‌نت**

این معماری در سال ۲۰۱۴ توسط گوگل معرفی شد و در طبقه‌بندی تصویر، شناسایی افراد و حوزه‌های دیگر بسیار مورد استفاده قرار گرفت و نتایج بسیار خوبی را منجر شد. شبکه عصبی پیچشی گوگل‌نت به عنوان عمیق‌ترین شبکه CNN در بین مدل‌های رایج بوده و مهمترین وجه تمایز این شبکه نسبت به مدل‌های قبلی استفاده از ماژول Inception در ساختار خود است. به طور کلی یک شبکه گوگل‌نت دارای دو لایه پیچشی، دو لایه ادغام و نه لایه Inception بوده که هر لایه Inception خود دارای شش لایه پیچشی و یک لایه ادغام است. مدل Inception V3 به عنوان یک مدل پیشرفته از شبکه عصبی پیچشی گوگل‌نت اولین بار در سال ۲۰۱۵ معرفی شد. در این مدل اندازه فیلترهای کانولوشنالی ۱×۱، ۳×۳ و ۵×۵ در یک ماژول یکسان و بر روی یک ورودی اعمال می‌شوند و خروجی تمامی این فرایندها با یکدیگر الحاق شده تا بعنوان ورودی لایه بعدی در نظر گرفته شوند [۳۶].

۳-۳ یادگیری انتقالی

یادگیری انتقالی[2] به معنای استفاده از یک مدل از پیش‌آموزش‌دیده در یک کاربرد جدید است. این تکنیک امروزه در یادگیری عمیق بسیار مورد توجه است، زیرا امکان آموزش شبکه‌های عصبی عمیق را با داده‌های نسبتاً کمی فراهم می‌کند. آموزش با داده کم بسیار ارزشمند است، زیرا در اکثر مسائل دنیای واقعی، داده‌های برچسب‌دار حجیم برای آموزش مدل‌های پیچیده یادگیری عمیق وجود ندارد. در این پژوهش از شبکه پیش‌آموزش‌دیده VGG16 که بر روی مجموعه داده آموزشی ImageNet آموزش دیده است، استفاده می‌شود.

۴-۳ تنظیم دقیق

در روش تنظیم دقیق[3]، ابتدا یک شبکه آموزش‌دیده که نتایج بسیار خوبی بر روی مجموعه داده تصاویر مطرح بدست آورده، انتخاب می‌گردد. سپس لایه تماماً متصل نهایی با لایه تماماً متصل دیگری که قابلیت طبقه‌بندی به تعداد کلاس‌های مورد نظر را دارد، جایگزین می‌شود. در این فرایند وزن‌های لایه‌های اولیه شبکه اصلی بدون تغییر وارد شبکه می‌شوند و لایه‌های نهایی شبکه دوباره آموزش می‌بینند تا شبکه مورد نظر قابلیت طبقه‌بندی تصاویر جدید را داشته باشد [۳۷]. در این روش، لایه‌های اولیه ثابت باقی می‌مانند و فقط چند لایه آخر با استفاده از داده‌های جدید دوباره آموزش داده می‌شوند. ثابت بودن لایه‌های اولیه به این دلیل است که در چند لایه اول ویژگی‌های سطح پایین مانند لبه‌ها، شکل‌ها، چرخ‌ها و غیره را از قبل یاد گرفته‌اند و بنابراین نیازی به آموزش مجدد ندارند.

۴ روش پیشنهادی

در این مقاله چارچوبی برای شناسایی خودروهای اضطراری مبتنی بر یادگیری عمیق با استفاده از شبکه عصبی پیچشی ارائه شده است. نکته حائز اهمیت این است که در شبکه‌های عصبی پیچشی دستیابی به نتایج قابل قبول و عملکرد مناسب، مستلزم در اختیار داشتن حجم عظیمی از داده‌ها برای آموزش شبکه است. با توجه به محدود بودن تعداد تصاویر موجود در مجموعه داده مورد استفاده در این پژوهش، و به منظور افزایش دقت شناسایی، از فرایند یادگیری انتقالی و شبکه پیش‌آموزش‌دیده VGG16 استفاده می‌شود. بدیهی است با توجه به تنوع و همه‌منظوره بودن مجموعه داده آموزشی ImageNet بارگذاری وزن‌های یادگیری شده برای اهداف مختلف در مجموعه داده‌های دیگر امکان‌پذیر بوده و موجب صرفه‌جویی در زمان یادگیری شبکه و تسریع محاسبات خواهد شد. در ادامه این بخش مرحله تقویت داده و سپس معماری پیشنهادی شرح داده خواهد شد.

[1] GoogleNet

[2] Transfer Learning

[3] Fine Tuning

۴-۱ تقویت داده

تقویت داده[1] یک روش کارآمد در کاهش بیش‌برازش[2] و آموزش بهتر مدل می‌باشد. هنگامی که امکان به دست آوردن نمونه‌های واقعی بیشتر وجود ندارد، از تقویت داده‌ها برای تولید نمونه‌های بیشتر استفاده می‌شود. در این تکنیک عملگرهای مختلفی مانند چرخش[3]، مقیاس‌گذاری[4]، جابه‌جایی[5]، برش[6] و افزایش نویز [۳۸] بر روی تصاویر اولیه اعمال شده و با تولید تصاویر جدید، تعداد تصاویر را افزایش می‌دهند. برای طبقه‌بندی با استفاده از شبکه‌های پیچشی، این مسأله را می‌توان با تغییر شکل تصاویر ورودی که کلاس‌های شیء مشاهده شده را تغییر نمی دهند، بدست آورد. این روش در زمان آموزش شبکه انجام می‌شود، هنگامی که داده‌ها بارگذاری می‌شوند، تغییرات اعمال شده و هر بار تصویر متفاوتی ایجاد می‌شود. در این پژوهش از چند نوع تقویت داده که در جدول ۱ آورده شده، استفاده شده است.

جدول ۱ نوع و مقادیر تقویت داده مورد استفاده در آموزش شبکه

نوع تقویت داده	مقادیر
rotation_range	۴۰° تا -۴۰°
width_shift_range	۰/۲
height_shift_range	۰/۲
zoom_range	۰/۲
horizontal_flip	True
fill_mode	nearest

۴-۲ معماری روش پیشنهادی

در اینجا، چارچوبی برای شناسایی خوردروهای اضطراری با استفاده از یادگیری عمیق و یادگیری انتقالی معرفی می‌گردد. نمای کلی روش پیشنهادی در شکل ۲ آورده شده است. عملکرد روش پیشنهادی شامل مراحل زیر است: ابتدا بر روی مجموعه داده پیش‌پردازش انجام شده و سپس عملیات استخراج ویژگی و طبقه‌بندی در یک شبکه عصبی پیچشی به صورت پشت سرهم و همزمان صورت می‌گیرد و در انتها با استفاده از معیارهای ارزیابی، شناسایی انواع خودروهای اضطراری انجام می‌شود. در ادامه این بخش پیش‌پردازش داده‌ها، معرفی معماری مدل یادگیری عمیق و نحوه آموزش مدل شناسایی نوع تصاویر خودروهای اضطراری با استفاده از داده‌های آموزشی به تفصیل بیان می‌گردد.

پیش‌پردازش

پیش‌پردازش داده‌ها، یکی از حیاتی‌ترین مراحل پیاده‌سازی سیستم‌های مبتنی بر یادگیری ماشین محسوب می‌شود. به دلیل اندازه متفاوت تصاویر موجود در مجموعه داده خودروهای اضطراری، پیش از اجرای قسمت‌های اصلی الگوریتم تمام تصاویر به ابعاد یکسان تغییر اندازه پیدا می‌کنند تا روند الگوریتم بهتر انجام شود. علاوه بر یکسان‌سازی ابعاد، عملیات زیر نیز انجام می‌شود:

- تبدیل داده‌ها به تنسور
- نرمال‌سازی داده‌ها بین مقادیر ۰ و ۱

معماری مدل یادگیری عمیق

مرحله دوم روش پیشنهادی در فرایند پیاده‌سازی مدل یادگیری عمیق، استفاده از معماری مدل یادگیری عمیق جهت شناسایی تصاویر است. به طور خلاصه، شبکه شناسایی نوع خودروی اضطراری از شش لایه تشکیل شده است: ۴ لایه کانولوشن و ۲ لایه تماماً متصل. از تابع Softmax در لایه آخر برای طبقه‌بندی استفاده می‌شود.

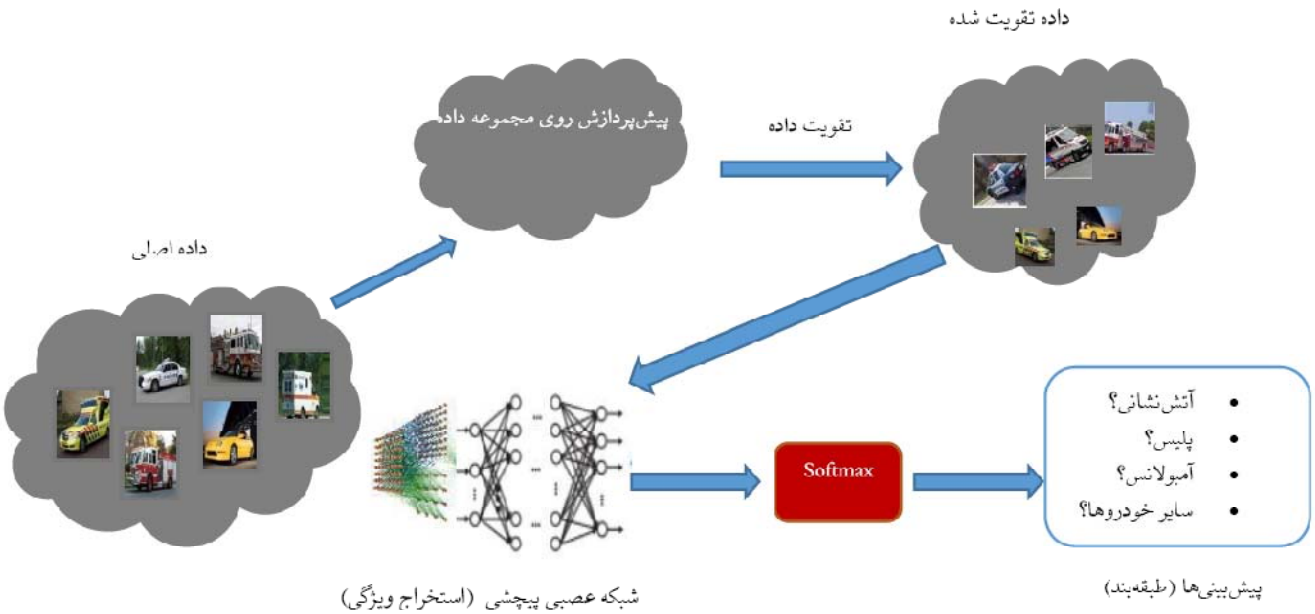
آموزش مدل طبقه‌بندی تصاویر با استفاده از داده‌های آموزشی

در این مرحله شبکه با مدل تعریف شده آموزش داده می‌شود. با توجه به تعداد کم تصاویر اولیه، مسئله بیش‌برازش رخ می‌دهد. برای رفع این مسئله از دو تکنیک تقویت داده و Dropout [۳۹] استفاده شده است. استفاده از تکنیک تقویت داده باعث می‌شود در شبکه هیچ تصویری دوبار مشاهده نشود. در نتیجه استفاده از تقویت داده مشکل بیش‌برازش را به مقدار قابل توجهی حل می‌کند. در ادامه برای بهبود بیشتر از Dropout با نرخ ۰,۵ استفاده شد. این تکنیک از آموزش همه گره‌های شبکه جلوگیری می‌کند و منجر به بهبود سرعت و عملکرد آموزش می‌شود. این روش می‌تواند روی لایهٔ ورودی و لایه‌های مخفی اعمال شود. استفاده از این دو تکنیک علاوه بر حذف مشکل بیش‌برازش باعث بهبود دقت شده است.

در اینجا برای دستیابی به دقت بهتر از شبکه پیش‌آموزش‌دیده VGG16 استفاده می‌شود. برای استفاده از شبکه پیش‌آموزش‌دیده دو راه کلی وجود دارد: (۱) استفاده از شبکه‌های پیش‌آموزش‌دیده به عنوان ابزاری برای استخراج ویژگی و (۲) ایجاد تغییراتی در شبکه VGG16 برای استخراج ویژگی بهتر. در این پژوهش از هر دو راه حل برای کسب نتیجه بهتر و نیز مقایسه، استفاده گردیده است. در راه حل اول اگر لایهٔ آخر VGG16 را با لایه‌ای جایگزین کنیم که به جای ۱۰۰۰ احتمال فقط ۴ (تعداد کلاس‌های ما ۴ است) احتمال را در نظر بگیرد، می‌توانیم از تمام دانشی که VGG16 در ImageNet آموزش‌دیده برای حل مسئله جاری استفاده کنیم. بنابراین ابتدا باید مرحله طبقه‌بندی را تغییر داد تا لایهٔ آخر شامل ۴ نورون شود.

---

[1] Data Augmentation
[2] Overfitting
[3] Rotating
[4] Scaling
[5] Translating
[6] Cropping



شکل ۲ چارچوب روش پیشنهادی

سپس باید شبکه را مجدداً آموزش داد که امکان تغییر وزن لایه‌های تماماً متصل فراهم شود. به این منظور، مقداردهی اولیه به وزن‌های شبکه با استفاده از وزن حاصل از ImageNet انجام می‌شود. پس از آن، همه لایه‌های پیچشی و ادغام فریز[1] می‌شوند تا وزن آنها تغییر پیدا نکنند. در چنین شرایطی، فقط لایه‌های تماماً متصل قابل بروزرسانی هستند و به محض اینکه این کار به انجام رسید، می‌توان پروسه آموزش مجدد را آغاز نمود. پس می‌توان از مرحله استخراج ویژگی در شبکه به خوبی استفاده و فقط طبقه‌بند نهایی را به‌طور دقیق تنظیم کرد تا بتواند به شکل بهتری با مجموعه دادهٔ جدید کار کند. این فرایند را یادگیری انتقالی می‌گویند؛ فرایندی که در آن از دانش به دست آمده از حل یک مسأله در مسأله متفاوت دیگری اما مرتبط، استفاده می‌شود. هرچقدر میزان شباهت مجموعه داده جدید و مجموعه‌داده‌ای که در ابتدا، شبکه به وسیله آن آموزش‌دیده بیشتر باشد، امکان حصولِ نتایج بهتر نیز افزایش پیدا می‌کند. اما اگر مجموعه دادهٔ مورد استفاده ارتباطی با ImageNetنداشته باشد یا بخواهیم نتایج را بهبود ببخشیم، لازم است که راهکار جدیدی پیشنهاد نمود. در این تحقیق از روش تنظیم دقیق که در بخش ۳-۴ توضیح داده شد برای این منظور استفاده شده است. این راهکار، همانگونه که در بخش ۵ خواهیم دید، منجر به حصول نتایج بهتری هم می‌شود.

به عبارت دیگر، در راه حل اولیه، فقط لایه‌های مرحلهٔ طبقه‌بندی تغییر داده می‌شود، اما دانشی که شبکه در هنگام استخراج خصوصیات (الگوها) در کار قبلی بدست آورده بود را تغییر نداده و وزن‌ها با توجه به کار قبلی بارگذاری می‌شوند. در راه حل مبتنی بر تنظیم دقیق، علاوه بر آموزش مجددِ مرحله طبقه‌بندی (لایه‌های تماماً متصل)؛ مرحله استخراج ویژگی (شامل لایه‌های پیچشی و ادغام) را هم مجدداً آموزش می‌دهیم.

باید به این نکتهٔ مهم نیز اشاره کرد که معمولاً لایه‌های نخست در شبکه‌های عصبی به شناساییِ مجموعه داده‌های کلی‌تر و ساده‌تر می‌پردازند و هر چه بیشتر در مراحل معماری پیش برویم، الگوهای پیچیده‌تری مورد شناسایی قرار می‌گیرند.

در نهایت همان‌طور که اشاره شد در راهکار پیشنهادی از تنظیم دقیق نیز استفاده شد. در نتیجه و بطور خلاصه، مدل نهایی تشکیل شده است از لایه‌های پیچشی شبکه VGG16 که به عنوان مدل پایه استفاده شده ولی لایه‌های بلوک آخر آن از حالت فریز خارج شده و به دو لایه تماماً متصل جدید مرتبط شده‌اند. این دو لایه جدید شامل ۲۵۶ نود و ۴ نود که به ترتیب در آنها از تابع فعال‌ساز Relu برای شناسایی و Softmax برای طبقه‌بندی استفاده شده است.

۵ ارزیابی نتایج

برای شبیه‌سازی و آموزش شبکه عصبی از کتابخانه کراس [۴۰] در زبان برنامه‌نویسی پایتون [۴۱] استفاده شده است. کلیه مراحل آموزش و آزمون شبکه‌ها بر روی یک رایانه با پردازنده intel core i7-6500u CPU شامل هفت هسته محاسباتی، فرکانس 2.60GHz و حافظه فیزیکی ۸ گیگابایت انجام گرفته است.

۵-۱ مجموعه دادگان خودرو

اولین پایگاه داده مورد استفاده در این مقاله، شامل ۴۰۰۰ تصویر از خودورهای اضطراری است. برای شناسایی تصاویر خودروهای اضطراری از ۴ کلاس استفاده شده است. این کلاس‌ها عبارتند از: خودروی آمبولانس، خودروی آتش‌نشانی، خودروی پلیس و خودروهای معمولی. برای جمع‌آوری خودروی معمولی، از تصاویر مجموعه داده‌های خودروهای دانشگاه استنفورد [۳۶] استفاده شده است. این مجموعه داده شامل ۱۶۱۸۵ تصویر از خودروهای

[1] Freeze

معمولی، خودروهای مسابقه، کامیون،[1] SUV و غیره است که از میان آنها ۱۰۰۰ تصویر انتخاب شده‌اند. برای جمع‌آوری تصاویر ۳ کلاس دیگر خودروهای اضطراری، موتور جستجوی گوگل به کار گرفته شده است. هر یک از سه کلاس مورد استفاده نیز شامل ۱۰۰۰ تصویر می‌باشند. تصاویر موجود در مجموعه داده در شرایط محیطی مختلف (باران، برف، روز و شب) و از زوایای گوناگون انتخاب شده است. شکل ۳ چند نمونه از تصاویر مجموعه داده مورد استفاده را نشان می‌دهد.

شکل ۳ نمونه‌هایی از تصاویر مجموعه داده مورد استفاده

به علاوه، در قسمت ارزیابی این مقاله از مجموعه داده ککِل[2] [۴۲] نیز استفاده شده است. این مجموعه داده توسط کیشور کومار جها از کشور هندوستان جمع‌آوری شده و شامل سه گروه وسایل نقلیه اضطراری (خودروی پلیس، آمبولانس و آتش‌نشانی) می‌باشد. این مجموعه داده شامل ۷۲۹۲ تصویر بوده که از این تعداد ۶۵۸۵ تصویر برای آموزش و ۷۰۷ تصویر برای آزمون در نظر گرفته شده است. این مجموعه داده در سال ۲۰۲۰ ایجاد شده است.

همچنین از مجموعه داده MIO-TCD [۴۳] نیز برای ارزیابی روش پیشنهادی استفاده شده است. این مجموعه در کل شامل ۷۸۶۷۰۲ تصویر می‌باشد که ۶۴۸۹۵۹ تصویر آن برای طبقه‌بندی و ۱۳۷۷۴۳ تصویر برای مکان‌یابی در نظر گرفته شده است. این مجموعه داده در ساعات مختلف روز و فصل‌های مختلف سال توسط هزاران دوربین ترافیکی مستقر در سراسر کانادا و ایالات متحده به دست آمده است.

از تصاویر موجود در این مجموعه داده برای پوشش طیف وسیعی از چالش‌های ترافیکی استفاده می‌شود و شامل داده‌هایی است که در سناریوهای مختلف ترافیک شهری ثبت شده‌اند. این مجموعه داده با هدف فراهم آوردن امکان دقیق محک زدن برای آموزش و آزمایش الگوریتم‌های موجود و جدید برای طبقه‌بندی و مکان‌یابی وسایل نقلیه در حال حرکت در صحنه‌های ترافیکی جمع‌آوری شده است. در اینجا ۱۱ کلاس مربوط به چالش طبقه-بندی تصاویر عبارتند از {کامیون، دوچرخه، اتوبوس، ماشین سواری، موتورسیکلت، وسایل نقلیه غیر موتوری، عابر پیاده، وانت باربری، کامیون تک چرخ، ون و پس زمینه}.

۲-۵ معیارهای ارزیابی

در این بخش معیارهای ارزیابی مورد استفاده در این مقاله را معرفی می‌کنیم. یکی از مهم‌ترین مراحل پس از طراحی و ساخت یک مدل یا یک الگوریتم، ارزیابی کارایی[3] آن است. حساسیت[4] (نرخ پاسخ‌های مثبت درست) و اختصاصیت یا تشخیص-پذیری[5] (نرخ پاسخ‌های منفی درست) دو شاخص مهم برای ارزیابی آماری عملکرد نتایج طبقه‌بندی هستند. زمانی که بتوان داده‌ها را به دو گروه مثبت و منفی تقسیم کرد، عملکرد نتایج یک آزمایش که اطلاعات را به این دو دسته تقسیم می‌کند با استفاده از شاخص‌های حساسیت و تشخیص‌پذیری قابل اندازه‌گیری و توصیف است. برای اندازه‌گیری کارایی ۴ پارامتر نیاز است:

مثبت صحیح (True Positive): درست شناسایی شده است.

مثبت کاذب (False Positive): اشتباه شناسایی شده است

منفی صحیح (True Negative): به درستی رد شده است

منفی کاذب (False Negative): به اشتباه رد شده است

معیار صحت[6]، بیان‌کننده تعداد «پیش‌بینی‌های صحیح انجام شده» توسط طبقه‌بند، تقسیم بر تعداد «کل پیش‌بینی‌های انجام شده» توسط همان طبقه‌بند است. رابطه ۳ چگونگی محاسبهٔ صحت را نشان می‌دهد:

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \qquad (۳)$$

معیار دیگر برای ارزیابی روش پیشنهادی معیار دقت[7] است. معیار دقت نسبت «تعداد پیش‌بینی‌های صحیح انجام شده» برای نمونه‌های یک کلاس خاص، به «تعدادکل پیش‌بینی‌ها» برای نمونه-های همان کلاس خاص را ارائه می‌کند. مقادیر بالا برای معیار

[1] Sports Utility Vehicle
[2] Kaggle
[4] Performance
[5] Sensitivity
[6] Specificity
[7] Accuracy
[8] Precision

دقت، بیانگر تعداد کم داده‌هایی است که به اشتباه در کلاس خاص طبقه‌بندی شده‌اند. این معیار به صورت رابطه ۴ محاسبه می‌شود:

$$\text{Precision} = \frac{\text{TP}}{\text{Tp+FP}} \qquad (۴)$$

معیار بعدی برای ارزیابی فراخوانی[1] است. این معیار بیان‌کننده نسبت «تعداد داده‌های درست طبقه‌بندی شده» در یک کلاس خاص، به «تعداد کل داده‌هایی است که باید در همان کلاس خاص طبقه‌بندی شوند» می‌باشد. مقادیر بالا برای این معیار، بیانگر تعداد کم داده‌هایی است که به اشتباه، در آن کلاس خاص طبقه‌بندی نشده‌اند و به صورت ۵ محاسبه می‌شود:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (۵)$$

معیار F1، نیز یک معیار مناسب برای ارزیابی دقت یک آزمایش است. این معیار پارامترهای دقت و فراخوانی را با هم ترکیب می‌کند تا مشخص شود یک مدل طبقه‌بند تا چه حد عملکرد خوبی از خود نشان می‌دهد. معیار F1 در بهترین حالت، یک و در بدترین حالت صفر است. چگونگی محاسبهٔ این معیار در رابطهٔ ۶ آورده شده است.

$$\text{F1}_{\text{Measure}} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \qquad (۶)$$

برای آموزش مدل، از الگوریتم بهینه‌ساز آدام[2] [۴۴] استفاده شده است. بجای استفاده از نرخ یادگیری ثابت، آدام از میزان یادگیری تطبیقی استفاده می‌کند. این الگوریتم بهینه‌سازی از دو پارامتر beta1 و beta2 برای کنترل میزان یادگیری استفاده می‌کند. پارامترهای بهینه‌ساز در جدول ۲ آورده شده است.

جدول ۲ پارامترهای بهینه‌ساز آدام

پارامتر	مقادیر
نرخ یادگیری	۴e-۱
Beta 1	۰/۹
Beta 2	۰/۹۹۹

۳-۵ ارائه و تحلیل نتایج

در این بخش ابتدا نتایج روش پیشنهادی بررسی می‌شود و پس از آن، نتیجه روش پیشنهادی با تعدادی از روش‌های پیشین مقایسه می‌گردد. در ابتدا یك شبکه عصبی پیچشی ایجاد کرده و با تعدادی اندازه مختلف برای تصاویر، معیار صحت طبقه‌بندی اندازه‌گیری می‌شود. نتایج بدست آمده در جدول ۳ آورده شده است.

جدول ۳ معیارهای ارزیابی در اندازه‌های ورودی مختلف بر روی مجموعه داده جمع‌آوری شده توسط نویسندگان

اندازه تصویر ورودی	صحت (%)	دقت (%)	فراخوانی (%)	F1-score (%)
۳×۶۴×۶۴	۹۴	۹۴/۱۱	۹۴	۹۳/۹۹
۳×۱۵۰×۱۵۰	۹۶,۲۴	۹۶/۳۰	۹۶/۲۴	۹۶/۲۵
۳×۲۲۴×۲۲۴	۹۷,۲۴	۹۷/۲۹	۹۷/۲۴	۹۷/۲۴

در مجموع با توجه به نتایج به دست آمده، اندازه تصاویر با مقدار ۲۲۴ مناسب دیده شد. برای بهبود نتیجه به دست آمده از شبکه VGG16 در دو حالت مختلفی که در بخش ۴ توضیح داده شد، استفاده می‌شود. نتایج در جدول شماره ۴ ارائه شده است. همانگونه که از نتایج مشاهده می‌شود با تنظیم دقیق شبکه و آموزش آن، صحت بالاتری به دست می‌آید.

در جدول شماره ۵ معیارهای بدست آمده از آموزش شبکه با استفاده از روش تنظیم دقیق بر روی شبکه VGG16 بر روی هر یک از ۴ کلاس به صورت جداگانه نشان داده شده است.

جدول ۴ نتایج مربوط به شبکهٔ VGG16 قبل از تنظیم دقیق و پس از آن بر روی مجموعه داده جمع‌آوری شده توسط نویسندگان

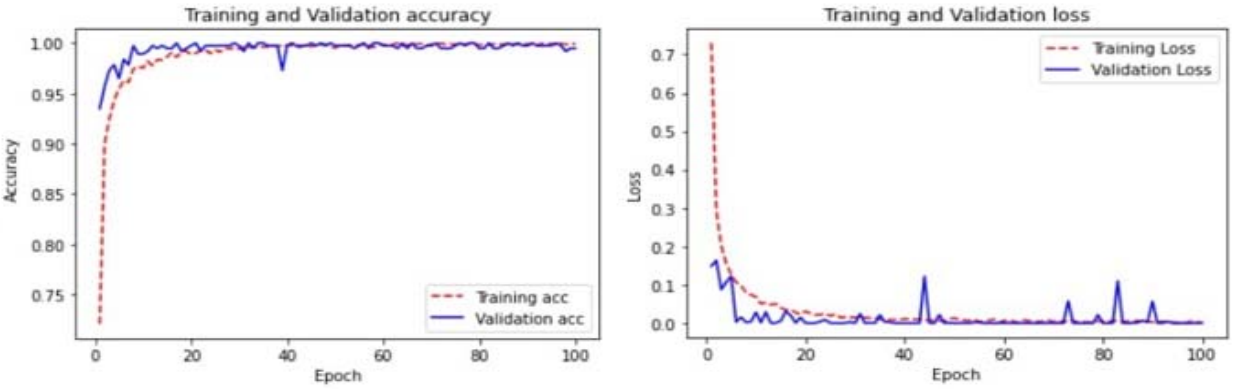
مدل	صحت (%)	دقت (%)	فراخوانی (%)	F1-score (%)
بدون استفاده از تنظیم دقیق	۹۸,۵۰	۹۸/۵۲	۹۸/۵۰	۹۸/۴۹
با استفاده از تنظیم دقیق	۹۹,۷۵	۹۹/۷۵	۹۹/۷۵	۹۹/۷۴

نمودار تابع هزینه و نمودار صحت به ترتیب در شکل ۴ (الف) و (ب) ترسیم شده است. همانطورکه در شکل ۴ مشاهده می‌شود شبکه با تغییراتی که بر روی VGG16 انجام گرفته شده است به خوبی آموزش دیده است. و هم بر روی مجموعه داده آموزشی و هم بر روی مجموعه داده اعتبار سنجی به صحت بالایی رسیده است. نمای کلی شبکه VGG16 و بلوک‌هایی که تکنیک تنظیم دقیق در آنها اعمال شده است، در شکل ۵ نشان داده شده است.

جدول ۵ نتایج طبقه‌بندی بر روی هر یک از کلاس‌های مجموعه داده جمع‌آوری شده توسط نویسندگان به صورت جداگانه

نوع خودرو	تعداد نمونه	صحت (%)	دقت (%)	فراخوانی (%)	F1-score (%)
آمبولانس	۱۰۰۰	۱۰۰	۹۹	۱۰۰	۹۹/۵۰
آتش‌نشانی	۱۰۰۰	۹۹	۱۰۰	۹۹	۹۹/۴۹
پلیس	۱۰۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
خودروی معمولی	۱۰۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰

[1] Recall
[2] Adam

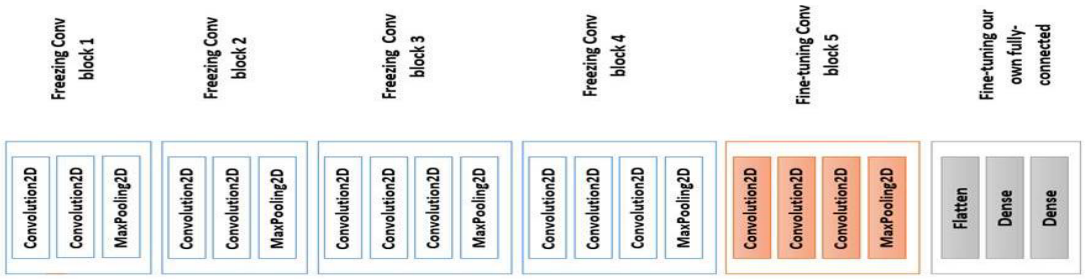


الف

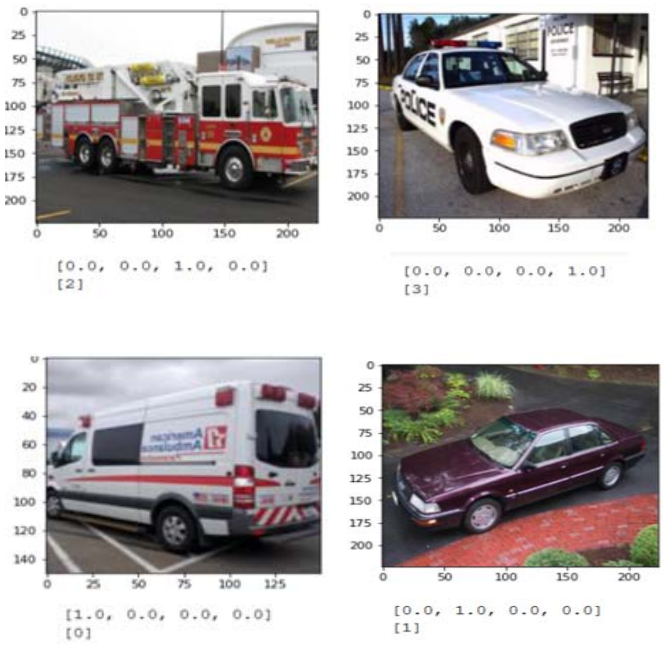
ب

شکل ۴ الف: نمودار خطای آموزش و اعتبارسنجی

ب: نمودار دقت آموزش و اعتبارسنجی



شکل ۵ معماری مدل VGG-16. "Freezing" نشان دهنده بخشی از مدل است که در طول دوره آموزش تغییر نکرده است [۴۵]



شکل ۶ نمونه‌هایی از تصاویر طبقه‌بندی شده توسط رویکرد پیشنهادی این پژوهش

در شکل ۶ نیز چهار مورد از خروجی‌های شبکه، که به درستی طبقه‌بندی شده‌اند، نشان داده شده است. این تصاویر به صورت تصادفی انتخاب و به شبکه داده شده است. اعداد ۰ تا ۳ به ترتیب نشان‌دهندهٔ کلاس‌های آمبولانس، خودروی معمولی، آتش‌نشانی و پلیس هستند.همان طور که در شکل۷ مشاهده می‌شود، یک تصویر به اشتباه پیش‌بینی شده است. برچسب تصویری که اشتباه پیش‌بینی شده است با رنگ قرمز نشان داده شده است. تصویر مورد نظر یک ماشین آمبولانس می‌باشد اما مدل به اشتباه آن را ماشین آتش-نشانی تشخیص داده است. علت آن نیز شباهت بسیار زیاد برخی ماشین‌های آتش‌نشانی با ماشین آمبولانس است.

در سال‌های اخیر معماری‌های مختلفی برای طبقه‌بندی اشیاء پیشنهاد شده است که در اینجا برای ارزیابی دقیق‌تر، روش پیشنهادی با تعدادی از معماری‌های برجسته مقایسه شده است. جدول ۶ نتایج بدست آمده بر روی مجموعه داده تصاویر ایجاد شده در این پژوهش و چهار معماری دیگر را نشان می‌دهد. همان-طور که مشاهده می‌شود روش پیشنهادی نسبت به معماری‌های دیگر به نتایج بهتری رسیده است.

در این راستا و برای ارزیابی دقیق‌تر روش ارائه شده در این پژوهش، از مدل مورد استفاده در [۴۶] نیز بر روی مجموعه داده جمع‌آوری شده توسط نویسندگان پژوهش حاضر استفاده شد. پژوهش [۴۶] در رابطه با طبقه‌بندی وسایل نقلیه کوچک، بر اساس شکل ظاهری، مدل و سال ساخت یک وسیله نقلیه است.

در این تحقیق برای طبقه‌بندی وسایل نقلیه کوچک چندین معماری برجسته اخیر مبتني بر شبکه عصبی پیچشی (CNN) را که نتایج بهتری داشته‌اند، مورد استفاده قرار گرفته است. به علاوه، عملکرد شبکه‌های VGG16، ResNets، Inception، DenseNets و MobileNet مقایسه شده و از مجموعه داده Stanford Cars-196 که دارای ۱۹۶ نوع مختلف وسیله نقلیه است استفاده شده است. مشابه با روش ما، نویسندگان [۴۶] نیز از تکنیک‌هایی همچون افزایش داده‌ها و تنظیم مجدد استفاده کرده و در نهایت به دقت طبقه‌بندی ۹۴٫۶٪ رسیده‌اند.

در ادامه، برای ارزیابی بیشتر، روش پیشنهادی برروی سه مجموعه داده که در بخش ۵-۱ معرفي شده‌اند مورد آزمایش قرار گرفت. علاوه بر این، با بهره‌گیری از روش‌های تغییر تصویر تعداد تصاویر مجموعه داده جمع‌آوری شده از ۴۰۰۰ به ۲۰۰۰۰ افزایش داده شد و مجموعه داده جدیدی ایجاد و مورد آزمایش و ارزیابی قرار گرفت. معیارهای ارزیابی به دست آمده در جدول ۷ آورده شده است. همچنین در این جدول نتایج چند پژوهش که در آن‌ها دیتاست MIO-TCD مورد استفاده قرار گرفته است، جهت ارزیابی و مقایسه نتایج کارهای مختلف با هم ارائه شده است.

جدول ۶ مقایسه معیارهای ارزیابی مدل پیشنهادی با سایر مدل‌ها بر روی مجموعه داده جمع‌آوری شده توسط نویسندگان

مدل	اندازه ورودی	صحت (%)	دقت (%)	فرخوانی (%)	F1-score (%)
MobileNetV2 [47]	۲۲۴×۲۲۴×۳	۹۵	۹۵/۷۶	۹۵	۹۴/۹۵
ResNet152V2 [35]	۲۲۴×۲۲۴×۳	۹۶,۷۵	۹۷/۰۳	۹۶/۷۵	۹۶/۷۴
Inception-V3 [48]	۲۲۴×۲۲۴×۳	۹۹,۵۰	۹۹/۵۰	۹۹/۵۰	۹۹/۴۹
Recent Deep Learning Architectures [46]	۲۲۴×۲۲۴×۳	۹۹,۷۵	۹۹,۷۷	۹۹,۷۳	۹۹,۷۴
VGG16 [ours]	۲۲۴×۲۲۴×۳	۹۹,۷۵	۹۹/۷۵	۹۹/۷۵	۹۹/۷۴

لازم به ذکر است که علت کمتر شدن معیارهای ارزیابی روش پیشنهادی برروی پایگاه داده MIO-TCD-Dataset، اولاً پایین بودن کیفیت تصاویر این مجموعه و ثانیاً زیاد بودن تعداد تصاویر و کلاس‌ها (۱۱ کلاس) است که برای اجرای مدل نیاز به سخت افزار بسیار قوی برای اجرای مدل می‌باشد.

شکل۷ نمونه‌هایی از پیش‌بینی مدل بر روی مجموعه داده آزمون

۶ نتیجه‌گیری و پیشنهادها

با توجه به اهمیت رفت و آمد انواع خودروهای اضطراری در خیابان‌ها و جاده‌ها، شناسایی آنها توسط خودروهای بدون راننده از اهمیت بسیار زیادی برخوردار است. ارائه راهکاری که بتواند شناسایی این خودروها را با دقت بالا انجام داده بسیار اهمیت دارد. در سال‌های اخیر استفاده از شبکه‌های یادگیری عمیق در مسائل طبقه‌بندی نتایج بسیار خوبی داشته است. بنابراین در این پژوهش از تکنیک شبکه عصبی پیچشی که یکی از قدرتمندترین تکنیک‌های یادگیری عمیق در مسائل طبقه‌بندی تصاویر می‌باشد در مسئله شناسایی خودروهای اضطراری در خودروهای بدون راننده استفاده شده است که با توجه به ارزیابی‌ها نتایج بسیار خوبی بدست آمده است. با توجه به محدود بودن تعداد تصاویر در این پژوهش از فرآیند یادگیری انتقالی و شبکه پیش‌آموزش‌دیده VGG16 برای رسیدن به دقت مطلوب استفاده شده است. بدیهی است که استفاده از شبکه عصبی پیچشی و فرآیند یادگیری انتقالی تأثیر قابل توجهی در افزایش دقت طبقه‌بندی در این مسئله داشته است. در تحقیقات آینده می‌توان از شبکه‌های مولد تخاصمی برای تولید تصاویر بیشتر استفاده کرد. همچنین می‌توان کلاسی برای موتورهای پلیس نیز در نظر گرفت. علاوه بر این موارد، می‌توان به عنوان یک تحقیق جدید شناسایی خودروهای اضطراری در ترافیک بالا را نیز مورد بررسی قرار داد.

جدول ۷ معیارهای ارزیابی بدست آمده از روش پیشنهادی برروی دیتاست‌های مختلف

الگوریتم	دیتاست مورد استفاده	تعداد تصاویر	صحت (%)	دقت (%)	فراخوانی (%)	F1-score (%)
Liu et al[20]	MIO-TCD	۶۴۸۹۵۹	-	۰/۹۲	۰/۸۸	۰,۸۹
Lee and Chung [49]	MIO-TCD	۶۴۸۹۵۹	۹۷,۹۲	۹۲,۹۸	۹۰,۲۸	۹۰,۹۸
Liu et al[50]	MIO-TCD	۶۴۸۹۵۹	-	۰,۹۳	۰,۹۰	۹۱,۴۷
Wang et al[51]	MIO-TCD	۶۴۸۹۵۹	۰,۹۷	-	-	-
Tanveer et al [52]	MIO-TCD	۶۴۸۹۵۹	۹۸,۳۴	-	-	-
. HEDEYA et al[53]	MIO-TCD	۶۴۸۹۵۹	۰,۹۷	۰,۹۲	۰,۹۱	۰,۹۱
روش پیشنهادی	Kaggle[۴۱]	۴۰۰۰	۹۲,۲۵	۹۲/۴۲	۹۲/۲۵	۹۲/۲۷
روش پیشنهادی	مجموعه داده جمع‌آوری شده توسط نویسندگان	۴۰۰۰	۹۹,۷۵	۹۹/۷۵	۹۹/۷۵	۹۹/۷۴
روش پیشنهادی	مجموعه داده جمع‌آوری شده همراه با افزایش تصاویر	۲۰۰۰۰	۹۹,۹۵	۹۹/۹۵	۹۹/۹۵	۹۹/۹۴
روش پیشنهادی	MIO-TCD	۲۰۳۰۶	۸۵,۵۸	۸۶,۳۰	۸۵,۱۹	۸۵,۷۳

تقدیر و تشکر

نگارندگان بر خود لازم می‌دانند که از سرکار خانم مریم تقی‌زاده، دانشجوی دکتری دانشگاه رازی، به خاطر کمک ایشان در انجام این پژوهش تشکر نمایند.

مراجع

[1] Teoh, ER., and Kidd, DG., "Rage against the machine? Google's self-driving cars versus human drivers", Journal of safety research, Vol. 63, pp. 57-60, 2017.

[۲] نصر، علی و مجیدفر، فرزان، **"آینده پژوهی نوآوری‌های بنیان شکن در بنگاه‌های بزرگ: مورد کاوی فناوری خودروهای بدون راننده در صنایع خودروسازی جهان و ایران"**، پنجمین کنفرانس بین المللی و نهمین کنفرانس ملی مدیریت فناوری، تهران، ۱۳۹۴.

[۳] اسدی، مریم و چاله چاله، عبداله، **"مروری بر توانایی‌ها، ملزومات و امکانات مورد نیاز در ماشین‌های خودران"**، ششمین کنگره سالانه ملی عمران معماری و توسعه شهری، تهران، ۱۳۹۷.

[4] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., "Image Net: A Large-Scale Hierarchical Image Database", IEEE conference on computer vision and pattern recognition, pp. 248-255, 2009.

[5] Hsieh, J.W., Yu, S.H., Chen, Y.S. and Hu, W.F., "Automatic Traffic Surveillance System for Vehicle Tracking and Classification", Transactions on Intelligent Transportation Systems, Vol. 2, pp.175-187, 2006.

[6] Avery, R.P., Wang, Y. and Rutherford, G.S., "Length-Based Vehicle Classification Using Images from Uncalibrated Video Cameras", The 7th International IEEE Conference on Intelligent Transportation Systems, pp. 737-742, 2004.

[7] Chen, Z., Pears, N., Freeman, M. and Austin, J., "Road Vehicle Classification Using Support Vector Machines". International Conference on Intelligent Computing and Intelligent Systems, Vol. 4, pp. 214-218, 2009.

[8] Clady, X., Negri, P., Milgram, M. and Poulenard, R., "Multi-class Vehicle Type Recognition System", IAPR Workshop on Artificial Neural Networks in Pattern Recognition, pp. 228-239, 2008.

[9] Pearce, G. and Pears, N., "Automatic Make and Model Recognition from Frontal Images of Cars", 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance, pp. 373-378, 2011.

[10] Hsieh, J.W., Chen, L.C., Chen, D.Y. and Cheng, S.C., "Vehicle Make and Model Recognition Using Symmetrical SURF", 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 472-477, 2013.

[11] Ng, L.T., Suandi, S.A. and Teoh, S.S., "Vehicle Classification Using Visual Background Extractor and Multi-class Support Vector Machines", Springer Science+Business Media Singapore, pp. 221-227, 2014.

[12] Zhang B., "Reliable Classification of Vehicle Types Based on Cascade Classifier Ensembles", Transactions on intelligent transportation systems, Vol. 14, pp. 322-332, 2013.

[13] Dong, Z., Wu, Y., Pei, M. and Jia, Y., "Vehicle Type Classification Using a Semisupervised Convolutional Neural Network", Transactions on intelligent transportation systems, Vol. 4, pp. 2247-2256, 2015.

[14] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-with Region Proposal Networks CNN: Towards Real-Time Object Detection", IEEE Transactions on pattern analysis and machine intelligence, Vol. 39, pp. 1137-1149, 2017.

[15] Rosati, R., Romeo, L., Silvestri, S., Marcheggiani, F., Tiano, L. and Frontoni, E., "Faster R-CNN approach for detection and quantification of DNA damage in Comet Assay Images", Computers in Biology and Medicine, Vol. 123, p.103912, 2020.

[16] Wu, W., Yin, Y., Wang, X. and Xu, D., "Face Detection with Different Scales Based on Faster R-CNN", IEEE Transactions on cybernetics, Vol. 11, pp. 4017-4028, 2018.

[17] Wang, X., Zhang, W., Wu, X., Xiao, L., Qian, Y. and Fang, Z., "Real-Time Vehicle Type Classification with Deep Convolutional Neural Networks", Journal of Real-Time Image Processing, Vol. 1, pp. 5-14, 2017.

[18] Wang, J. and Perez, L., "The Effectiveness of Data Augmentation in Image Classification using Deep Learning", Convolutional Neural Networks Vis. Recognit, 2017.

[19] Zhuo, L., Jiang, L., Zhu, Z., Li, J., Zhang, J. and Long, H, "Vehicle Classification for Large-Scale Traffic Surveillance Videos Using Convolutional Neural Networks", Machine Vision and Applications, Vol. 28, pp. 793–802, 2017.

[20] Liu, W., Zhang, M., Luo, Z. and Cai, Y., "An Ensemble Deep Learning Method for Vehicle Type Classification on Visual Traffic Surveillance Sensors", EEE Access, Vol. 5, pp. 24417-24425, 2017.

[21] CHEN, W., SUN, Q., WANG, J., DONG, J.J. and XU, C., "A Novel AdaBoost and CNN Based for Vehicle Classification", IEEE, pp. 2169-3536, 2018.

[22] Gowtham, P., P. Eswari, and V. P. Arunachalam. "An Investigation Approach Used for Pattern Classification and Recognition of an Emergency Vehicle", ICSNS, pp. 1-7, 2018.

[23] Deepa, Kumari, N., Manisha, K., Manu, M., Kshama S., "Smart Detection of Emergency Vehicles in Traffic", International Journal of Current Engineering and Scientific Research, Vol. 5, pp. 2394-0697, 2018.

[24] Roy, S. and Rahman, M.S., "Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network", International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-6, 2019.

[25] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition", Computer Vision and Pattern Recognition, 2015.

[26] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., "You Only Look Once: Unified, Real-Time Object Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[27] Lawrence, S., Giles, C.L., Tsoi, A.C. and Back, A.D., "Face Recognition: A Convolutional Neural-Network Approach: A Convolutional Neural-Network Approach", IEEE transactions on neural networks, Vol. 8, pp. 98-113, 1997.

[28] Hinton, G.E., "Deep Belief Network", Scholarpedia, Vol. 4, p. 5947, 2009.

[29] Kombrink, S., Mikolov, T., Karafiát, M. and Burget, L., "Recurrent Neural Network Based Language Model", Twelfth annual conference of the international speech communication association, Vol. 2, p. 3, 2010.

[30] Girshick, R., Donahue, J., Darrell, T. and Malik, J., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation", IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

[31] Hinton, G., et. al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", IEEE Signal processing magazine, Vol. 29, pp. 82–97, 2012.

[۳۲] قنبری سرخی، علی و فاتح، منصور و حسن پور، حمید، **"تشخیص و فیلترینگ هوشمند تصاویر نامتعارف به کمک شبکه-های عصبی عمیق"**، مجله پردازش علائم و داده‌ها، جلد ۳۶، شماره ۲، ص ۵۵ تا ۶۸، ۱۳۹۷.

[33] Patrick, M.K., Adekoya, A.F., Mighty, A.A. and Edward, B.Y., "Capsule Networks – A Survey", Journal of King Saud University-Computer and Information Sciences, 2019.

[34] Simhambhatla, R., Okiah, K., Kuchkula, S. and Slater, R., "Self-Driving Cars: Evaluation of Deep Learning Techniques for Object Detection in Different Driving Conditions", SMU Data Science Review, p. 23, 2019.

[35] He, K., Zhang, X., Ren, S. and Sun, J., "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[36] Szegedy, C., et. Al., "Going Deeper with Convolutions", IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015.

[۳۷] کارساز، علی و محمدیان روشن، صبورا، **" پردازش تصاویر پزشکی با استفاده از شبکه‌های عصبی کانولوشن عمیق"**، فصل نامه علمی ترویجی عصر برق.; جلد ۱۱، شماره ۳ ، ص ۲۸-۲۳، ۱۳۹۷.

[38] Stenroos, O., "Object Detection from Images Using Convolutional Neural Networks", Master Thesis, Aalto University, Espoo, Finland, 2017.

[39] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., "Dropout: A Simple Way to Prevent Neural Networks from Overtting", Journal of Machine Learning Research, Vol. 15, pp. 1929-1958, 2014.

[40] Chollet, Francis, et al. "Keras", 2015, https://github.com/keras-team/keras.

[41] Foundation Python Software, "Python 2.7.13 Documantation".

[42] https://www.kaggle.com/parthplc/emergency-vs-nonemergency-vehicle-classification.

[43] Luo, Z., B.-Charron, F., Lemaire, C., K Konrad, J., Li, S., Mishra, A.,Achkar, A., Eichel,J., Jodoin,P, "MIO-TCD: A new benchmark dataset for vehicle classification and localization", Journal Of Latex Class Files,2017.

[44] Kingma, D.P. and Ba, J., "Adam: a Method for Stochastic Optimization", 3rd International Conference for Learning Representations, 2017.

[45] Zrira, N., Ahmad Khan, H., Bouyakhf, E., "Discriminative Deep Belief Network for indoor environment classi_cation using global visual features", Springer Verlag, 2018.

[46] Valev, K., Schumann, A.,Sommer,L., Beyerer,J, ""A Systematic Evaluation of Recent Deep Learning Architectures for Fine-Grained Vehicle Classification",

Proceedings Volume 10649, Pattern Recognition and Tracking, 2018.

[47] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., "MobileNetV2: Inverted Residuals and Linear Bottlenecks". EEE conference on computer vision and pattern recognition, pp. 4510-4520, 2018.

[48] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., "Rethinking the Inception Architecture for Computer Vision", IEEE conference on computer vision and pattern recognition. pp. 2818-2826, 2015.

[49] Lee, J., Chung, Y, "Deep Learning-based Vehicle Classification using an Ensemble of Local Expert and Global Networks, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.

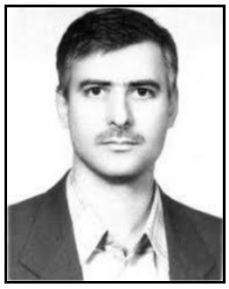
[50] Liu,W, Luo,Z., Li,S, "Improving Deep Ensemble Vehicle ClassiÞcation by Using Selected Adversarial Samples", Knowledge-Based Systems,2018

[51] WANG, Y., BAN, X., WANG, H., WU, D., WANG, H., YANG, S., LIU, S., LAI, J, "Detection and Classification of Moving Vehicle From Video Using Multiple Spatio-Temporal Features", IEEE Access, Volume: 7, 2019.

[52] Tanveer, M.S., Khan, M.U.K., Kyung, Ch.," Fine-Tuning DARTS for Image Classification", Computer Vision and Pattern Recognition 2020.

[53] Hedeya,M., EID, A., Abdel-Kader,R, "A Super-Learner Ensemble of Deep Networks for Vehicle-Type Classification", IEEE Access, Volume: 8,2020.

مریم اسدی دارای مدرک کارشناسی مهندسی تکنولوژی نرم‌افزار از موسسه غیر انتفاعی جهاد دانشگاهی کرمانشاه و دانشجوی کارشناسی ارشد مهندسی کامپیوتر (گرایش نرم افزار) از دانشگاه رازی، کرمانشاه، ایران است. علایق تحقیقاتی وی شامل بینایی ماشین، یادگیری عمیق و پردازش تصویر است.



عبداله چاله چاله متولد کرمانشاه، ایران بوده و مدرک کارشناسی مهندسی برق (سخت‌افزار) و کارشناسی‌ارشد مهندسی کامپیوتر (نرم‌افزار) خود را از دانشگاه صنعتی شریف، تهران، ایران دریافت کرده است. وی مدرک دکترای مهندسی کامپیوتر خود را از دانشگاه ولونگونگ استرالیا در سال ۲۰۰۵ میلادی دریافت و در حال حاضر عضو هیأت علمی دانشگاه رازی در کرمانشاه است. علایق تحقیقاتی وی شامل هوش مصنوعی، پردازش تصویر و ویدئو، سیستم‌های توزیع شده و اینترنت اشیا است.