ICWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

Transfer Learning Approach for Recognition of Iranian Traffic Signs in Self-Driving Cars

Mohammad Gorji*[1], Farhad Nasri[2], Azadeh Tabatabaei[3]

[1] Bachelor Student of Computer Engineering Department, Science and Culture University, Tehran, Iran
m.gorji@usc.ac.ir

[2] Bachelor of Computer Engineering Department, Science and Culture University, Tehran, Iran

f.nasri@usc.ac.ir

[3] Assistant Professor of Computer Engineering Department, Science and Culture University, Tehran, Iran

a.tabatabaei@usc.ac.ir

Abstract

In recent years, with the increasing advancement of artificial intelligence and machine learning, huge changes have occurred in the field of vehicles and self-driving cars. Recognition and interpretation of urban traffic signs, by machine vision systems, improve the safety of driving operations as one of the basic principles of self-driving vehicles. Due to the high interaction of self-driving vehicles with traffic signs during the movement, creating a system with high accuracy for interpretation and an immediate decision is a big challenge. In this research, with the use of convolutional neural networks, a system is designed that can recognize Iranian traffic signs. By applying the transfer learning approach, we train our model with a new collection of traffic signs images that reaches a high accuracy in optimal conditions.

Keywords: Traffic Sign Recognition, Convolutional Neural Networks, Transfer Learning, Machine Learning, Deep Learning

JCWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

رویکرد یادگیری انتقالی برای شناسایی تابلوهای ترافیک ایرانی در اتومبیل‌های خودران

محمد گرجی*[۱]، فرهاد نصری[۲]، آزاده طباطبائی[۳]

[۱] دانشجوی کارشناسی گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ ، تهران،
m.gorji@usc.ac.ir

[۲] کارشناسی گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ ، تهران،
f.nasri@usc.ac.ir

[۳] استادیار گروه مهندسی کامپیوتر دانشگاه علم و فرهنگ، تهران،
a.tabatabaee@usc.ac.ir

چکیده

در سال‌های اخیر با پیشرفت روزافزون هوش مصنوعی و یادگیری ماشین شاهد تحولات بزرگی در زمینه وسایل نقلیه و خودروهای خودران بوده‌ایم. تشخیص و تفسیر تابلوهای ترافیک شهری توسط سیستم‌های بینایی ماشین می‌تواند به عنوان یکی از اصول اولیه وسایل نقلیه خودران به بهبود ایمنی عملیات رانندگی کمک بسزایی نماید. تعامل بالای وسایل نقلیه خودران با تابلوهای ترافیک در طول حرکت، ایجاد سیستمی با دقت بالا برای تفسیر و تصمیم‌گیری بی‌درنگ را چالش برانگیز می‌کند. در این پژوهش سیستمی طراحی شده است که با استفاده از شبکه‌های عصبی پیچشی قادر به بازشناسی تابلوهای ترافیکی ایرانی خواهد بود. ما با استفاده از رویکرد یادگیری انتقالی، مدل خود را با یک مجموعه جدید از تصاویر تابلوهای ترافیک کشور ایران آموزش خواهیم داد که در شرایطی بهینه به دقت بالایی خواهد رسید.

کلمات کلیدی

بازشناسی تابلو ترافیک، شبکه‌های عصبی پیچشی، یادگیری انتقالی، یادگیری ماشین، یادگیری عمیق

۱- مقدمه

تابلوهای راهنمایی و رانندگی محور اساسی سیستم ترافیکی محسوب می‌شوند و نقش بسیار مهمی در حمل و نقل ایمن و کارا دارند. امنیت مسیر برای رانندگان و عابران نگرانی مهمی است و عوامل بسیاری مانند شرایط آب و هوایی، تابلوهای ترافیک و شرایط جاده بر آن تاثیر می‌گذارند که از مهم‌ترین آن‌ها تابلوهای ترافیک است. اگرچه تصاویر تابلوهای ترافیک توسط سیستم بصری انسان به آسانی تفسیر می‌شوند، اما دقت انسان به راحتی تحت تاثیر عواملی نظیر خستگی، خطای دید و غیره قرار خواهد گرفت. به علت جلوگیری از خطای انسانی در تفسیر تابلوهای ترافیک و افزایش ایمنی در امر رانندگی، توسعه سیستم‌هایی هوشمند با دقت و سرعت تشخیص بالا برای بازشناسی تابلوهای ترافیک از اهمیت ویژه‌ای برخوردار است. از طرف دیگر، تفسیر این تصاویر توسط سیستم‌های هوشمند نیز به دلایلی از قبیل انسداد، روشنایی، ظاهر بیرونی و غیره امری دشوار است.

چنین سیستمی قادر است به عنوان یک دستیار هوشمند در کنار راننده انسانی برای تشخیص و تفسیر تابلوهای ترافیکی به کار گرفته شود. همچنین مهمترین اصل در توسعه خودروهای خودران و نیمه خودران، چگونگی تعامل آن‌ها با محیط اطراف خودرو است. بدون وجود یک سیستم کارا برای بازشناسی بلادرنگ علائم راهنمایی، کارکرد وسایل نقلیه خودران با مشکل مواجه خواهند شد. اگرچه استفاده از اطلاعات نقشه‌های ترافیکی می‌تواند پاسخی برای تعامل سیستم‌های راهبری خودکار با علائم راهنمایی رانندگی باشد اما پوشش جغرافیایی و به‌روزرسانی این نقشه‌ها امری هزینه بر است. از طرفی اطلاعات جمع‌آوری شده توسط وسایل نقیه هوشمند، خود می‌تواند عامل بهینه‌ای در راستای گردآوری و به‌روزرسانی اطلاعات مسیرها و تابلوهای راهنمایی در نقشه‌های ترافیکی باشد.

یادگیری عمیق[1] به عنوان یکی از روش‌های زیر شاخه یادگیری ماشین، به صورت موفقیت آمیز در مسائل بینایی ماشین استفاده شده است. در میان مسائلی که توسط بینایی ماشین حل شده‌اند، بازشناسی تابلوهای ترافیک به طور گسترده مورد بحث محققان قرار گرفته است. روش‌های جدید مبتنی بر یادگیری عمیق با استفاده از شبکه‌های عصبی پیچشی[1]، ویژگی‌های تصویر را در لایه‌های پیچشی متفاوت استخراج و در انتها برای بازشناسی استفاده می‌کنند. پژوهش‌های اخیر بر مسئله بازشناسی تابلوهای ترافیک با مجموعه داده‌های متنوع از کشورهای متفاوت انجام شده است؛ در حالی که مطالعات انگشت شماری بر روی تابلوهای ایرانی گزارش شده است [2].

هدف این مقاله، پیاده سازی سیستمی است که تابلوهای ترافیکی ایرانی را با دقت خوبی بازشناسی کند. مدل پیشنهادی با تصاویر تابلوهای ترافیک کشور آلمان آموزش داده می‌شود و سپس با رویکرد یادگیری انتقالی، ویژگی‌های استخراج شده از مدل آموزش داده شده را با داده‌های کمتر در مجموعه داده ایرانی تنظیم دقیق[2] می‌کند. لازم به ذکر است که در این پژوهش برای اولین بار از مجموعه داده ایرانی ذکر شده استفاده شده است.

۲- پیشینه پژوهش

در برخی از رویکردهای اولیه در حل مسئله بازشناسی تابلوهای ترافیک، از روش‌های پردازش تصویر به منظور استخراج ویژگی و از مدل‌های ساده یادگیری ماشین برای دسته‌بندی استفاده می‌شد [3]–[6]. در روش‌های یادگیری ماشین سنتی استخراج دستی ویژگی‌ها اهمیت بالایی دارد اما استخراج ویژگی دستی، ممکن است منجر به از دست رفتن اطلاعات سودمند از داده‌ها شود. مدل‌های یادگیری عمیق می‌توانند به صورت خودکار ویژگی-های مورد نیاز برای مسئله محول شده را استخراج کنند [7]. مطالعات نشان داده است که استخراج ویژگی توسط این نوع از مدل‌ها به خوبی و یا بهتر از استخراج ویژگی دستی عمل می‌کند [8]. در سال‌های اخیر شبکه‌های عصبی عمیق در پژوهش‌های بازشناسی الگو و بینایی ماشین توجه زیادی به خود جلب کرده‌اند و به لطف مجموعه داده‌های عظیم با هزاران تصویر [9] به طور گسترده برای مسائل بازشناسی[10] و تشخیص اشیا[11] سازگار شده‌اند.

شبکه‌های عصبی عمیق در بازشناسی تابلوهای ترافیک بر روی مجموعه دادگان GTSRB [12] به دقت‌های بسیار بالایی رسیده‌اند. سرمانت و همکارانش با ارائه یک معماری چند مقیاسی (MSCNN) ، با استفاده از برقراری اتصالات از خروجی تمامی لایه‌ها به دو لایه نهایی دسته بند و استفاده از تصاویر سیاه و سفید در کنار تصاویر رنگی به دقت ۹۹.۱۷٪ رسیدند [13]. با ارائه شبکه عصبی عمیق چند ستونه (MCDNN)، چیرشان و همکارانش با رویکرد یادگیری گروهی از پنج شبکه عصبی عمیق با پنج ورودی پیش پردازش شده متفاوت استفاده کردند. آن‌ها با پیشی گرفتن از میانگین و بهترین دقت انسان (۹۸.۸۴٪ و ۹۹.۲۲٪) به دقت ۹۹.۴۶٪ رسیدند [14]. همچنین جین و همکارانش شبکه MCDNN را با تابع زیان هینگ[3] آموزش دادند و با استفاده از نرمال سازی پاسخ محلی[4]، به دقت ۹۹.۶۵٪ رسیدند [15]. در ادامه نیز شبکه عصبی عمیق با استفاده از ترنسفورمرهای فضایی[5] و روش‌های بهینه سازی تصادفی به دقت تقریبا کامل ۹۹.۷۱ رسید [16]. اگرچه این رویکردها به دقت‌های بالایی رسیدند، اما فرآیند آموزش آن‌ها زمان‌بر بوده و به یادگیری تعداد زیادی پارامتر نیاز دارند.

پس از دستیابی مدل‌های یادگیری عمیق به دقت‌های بالا در حل این مسئله، توجه پژوهشگران بیش از پیش به مسئله بهینگی مدل‌های پیشنهادی و کارایی میدانی آن‌ها معطوف شد [17]–[19]. در این راستا ژانگ و همکارانش از رویکرد تقطیر دانش[6] در شبکه‌های عصبی پیچشی استفاده کردند و دانش یک مدل پایه بزرگ (مدل معلم) را به یک مدل کوچکتر (مدل دانش‌آموز) با ۰.۸ میلیون پارامتر انتقال دادند که به دقت ۹۹.۶۱٪ رسیدند [20].

۳- روش پیشنهادی

به مانند اغلب مسائل یادگیری ماشین آموزش یک مدل برای بازشناسی تابلوهای ترافیک نیز دارای سه مرحله اصلی خواهد بود. ما در این پژوهش در قدم اول به پیش پردازش دادگان خواهیم پرداخت و بعد معماری مدل عصبی پیچشی را تعیین خواهیم کرد. در انتها نیز پس از آموزش مدل با مجموعه داده آموزشی، دقت مدل را با مجموعه داده آزمایشی خواهیم سنجید. همچنین ما در دو فاز شبکه پیچشی مد نظر را آموزش خواهیم داد. در فاز اول مدل را با مجموعه داده GTSRB آموزش خواهیم داد و در فاز دوم شبکه آموزش یافته را با مجموعه داده جدید Persian Traffic Sign Dataset (PTSD) تنظیم دقیق کرده و آموزش را با این مجموعه ادامه خواهیم داد.

۳-۱- شبکه‌های عصبی پیچشی

دسته بندی تصویر به انتساب یک یا چند کلاس به یک تصویر اطلاق می-شود. این امر از طریق تعیین برچسبی که گویای دسته غالب اشیا موجود در تصویر است صورت می‌پذیرد. به عنوان مثال در تصاویر مربوط به اشیا ترافیکی، هر تصویر می‌تواند به دسته‌ای به مانند تابلوها، موانع، چراغ‌های راهنمایی و غیره تعلق داشته باشند. از طرف دیگر مسئله شناسایی زیردسته-های یک دسته از تصاویر، به مانند تعیین نوع هر یک از تابلوهای ترافیک یک مسئله بازشناسی خواهد بود. اگرچه این دو مسئله از دیدگاه ذکر شده و نام تفاوت دارند، اما بازشناسی تابلوهای ترافیک می‌تواند به عنوان یک مسئله دسته بندی نیز بررسی شود.

شبکه‌های عصبی پیچشی یک رویکرد محبوب و کارا برای بسیاری از مسائل بینایی رایانه از جمله دسته بندی تصویر بوده است. یک شبکه عصبی پیچشی از تعدادی لایه پیچشی که خود دارای اجزای مختلف است تشکیل شده است. لایه پیچشی وظیفه پردازش و استخراج ویژگی‌های مختلفی از تصویر را بر عهده دارد که اینکار با محاسبه یک تابع خطی به عنوان فیلتر بر قسمت‌های مختلف تصویر و اعمال یک تابع غیر خطی بر نتایج حاصل صورت می‌پذیرد [1].

برای انجام محاسبات توسط یک لایه پیچشی، فیلتری با پارامترهایی که از طریق الگوریتم پس انتشار خطا قابل آموزش هستند، گام به گام بر روی تصویر حرکت کرده و با پیکسل‌های تصویر در حوزه تاثیر خود ضرب نقطه‌ای می‌شوند. حاصل بدست آمده پس از اعمال یک تابع غیر خطی به عنوان یک پیکسل از نقشه ویژگی[7] جدید در نظر گرفته می‌شود. لازم به ذکر است که

تصویر و فیلتر می‌تواند دو یا سه بعد داشته باشد چنانکه در هر لایه پیچشی چندین عملیات پیچشی انجام می‌شود و به تعداد فیلترها نقشه ویژگی بدست می‌آید. نقشه‌های ویژگی بدست آمده در قالب یک تنسور سه بعدی واحد و به عنوان ورودی لایه پیچشی بعدی در نظر گرفته خواهد شد [21].

برای عدم کاهش ابعاد تصویر بعد از عملیات پیچشی از حاشیه گذاری استفاده می‌شود، همچنین برای کاهش محاسبات و استخراج ویژگی‌های کاراتر، فیلترهایی از نوع ادغام بر نقشه‌های ویژگی اعمال می‌گردد. این فیلترها پارامتری برای یادگیری ندارند و با حرکت در تصویر، با جایگذاری حداکثر یا میانگین از مقادیر پیکسل‌های حوزه تاثیر خود باعث کاهش ابعاد و در نتیجه کاهش محاسبات می‌شوند. علاوه بر اندازه ابعاد هر فیلتر، طول حرکت فیلترها بر تصویر در هرگام حرکت نیز ابرپارامتری قابل تعیین است. تعیین جامع معماری شبکه عصبی پیچشی، لایه‌ها و ابرپارامترهای ذکر شده خود چالش بزرگی برای حل مسائل مختلف توسط شبکه‌های عصبی پیچشی است.

با افزایش تعداد لایه‌های پیچشی و عمیق سازی شبکه، ویژگی‌های سطح بالای مورد نیاز مسئله از تصاویر استخراج می‌شود. برای مسئله دسته بندی تصویر با مسطح سازی خروجی آخرین لایه پیچشی بردار ویژگی تصویر به دست می‌آید و با عرضه این بردار ویژگی به یک یا چند لایه شبکه عصبی پیشخور کاملا متصل[8]، عملیات دسته بندی تصویر صورت می‌پذیرد. اخرین لایه از شبکه پیشخور کاملا متصل به تعداد دسته‌های مورد نظر مسئله گره محاسباتی دارد و مقدار هر یک از گره‌ها، پس از اعمال تابع بیشینه سازی هموار گویای احتمال تعلق تصویر به دسته نظیر خود است.

۳-۲- معماری مدل پیشنهادی

ما در طراحی شبکه عصبی پیچشی از دو بلوک اصلی پیچشی استفاده کردیم. هر بلوک در این معماری شامل دو لایه پیچشی با فیلترهایی با ابعاد ۳×۳ با تابع فعال ساز رلو[9]، یک لایه فیلتر ادغام بیشینه با ابعاد ۲×۲ و یک لایه نرمال سازی دسته‌ای[10] است. نرمال سازی دسته‌ای باعث بهبود جریان گرادیان و در نتیجه آموزش بهتر مدل می‌شود [22]. همچنین عدم استفاده از حاشیه گذاری و استفاده از لایه ادغام بیشینه باعث کاهش تعداد پارامترهای مدل می‌شود. ما در چهار لایه پیچشی استفاده شده به ترتیب به تعداد ۱۶، ۳۲، ۶۴ و ۱۲۸ فیلتر تعیین کردیم تا با کاهش عرض شبکه به عمق آن افزوده شود و پردازش‌های بیشتری به تناسب بالا بودن سطح ویژگی‌های بدست آمده صورت گیرد.

در بلاک سوم و غیر پیچشی عملیات دسته بندی انجام می‌شود. در این بلاک خروجی بلاک پیچشی دوم مسطح شده و بردار ویژگی تصویر بدست می‌آید. این بردار پس از گذر از دو لایه شبکه عصبی پیشخور کاملا متصل با ۵۱۲ و ۶۴ گره محاسباتی و یک لایه نرمال سازی دسته‌ای به لایه کاملا متصل نهایی ورود پیدا می‌کند. همانطور که گفته شد لایه نهایی بایستی به تعداد دسته‌ها گره محاسباتی داشته و برای محاسبه احتمال هر کلاس از تابع فعال ساز بیشینه هموار[11] استفاده کند. ما همچنین برای جلوگیری از بیش برازش و از دست دادن قابلیت تعمیم دهی مدل بر مجموعه آزمایشی، از یک لایه حذف تصادفی[12] با نرخ ۰.۵ پیش از لایه نهایی استفاده کردیم. در جدول (۱) معماری مدل به همراه جزئیات لایه‌ها توصیف شده‌است.

جدول (۱) : جزئیات معماری مدل پیچشی پیشنهادی

تعداد پارامتر	ابعاد خروجی	تعداد فیلتر/گره	ابعاد فیلتر	نوع لایه
	۳۰ × ۳۰ × ۱			ورودی
۴۴۸	۲۸ × ۲۸ × ۱۶	۱۶	۳ × ۳	پیچشی
۴,۶۴۰	۲۶ × ۲۶ × ۳۲	۳۲	۳ × ۳	پیچشی
	۱۳ × ۱۳ × ۳۲		۲ × ۲	ادغام حداکثر
۱۲۸	۱۳ × ۱۳ × ۳۲			نرمال سازی دسته‌ای
۱۸,۴۹۶	۱۱ × ۱۱ × ۶۴	۶۴	۳ × ۳	پیچشی
۷۳,۸۵۶	۹ × ۹ × ۱۲۸	۱۲۸	۳ × ۳	پیچشی
	۴ × ۴ × ۱۲۸		۲ × ۲	ادغام حداکثر
۵۱۲	۴ × ۴ × ۱۲۸			نرمال سازی دسته‌ای
	۱ × ۱ × ۲۰۴۸			مسطح سازی
۱,۰۴۹,۰۸۸	۱ × ۱ × ۵۱۲	۵۱۲		پیشخور کاملا متصل
۳۲,۸۳۲	۱ × ۱ × ۶۴	۶۴		پیشخور کاملا متصل
۲۵۶	۱ × ۱ × ۶۴			نرمال سازی دسته‌ای
	۱ × ۱ × ۶۴			حذف تصادفی (۵۰٪)
۲,۷۹۵	۱ × ۱ × ۴۳	۴۳		پیشخور کاملا متصل

۳-۳- آموزش بر مجموعه دادگان کشور آلمان

ما در فاز اول آموزش از مجموعه GTSRB استفاده کردیم. این مجموعه شامل ۳۹,۲۰۹ تصویر برای آموزش و ۱۲,۶۳۰ تصویر برای آزمایش است که تمامی آن‌ها متعلق به تابلوهای ترافیکی کشور آلمان بوده و در ۴۳ گروه دسته بندی می‌شوند. یکی از چالش‌های اصلی در یادگیری عمیق نیاز این نوع از شبکه‌ها به تعداد بسیار زیادی از نمونه‌های آموزشی است. اگرچه استفاده از داده بیشتر برای آموزش شبکه‌های عصبی، اغلب نتایج بهتری را حاصل می‌کند، اما جمع آوری و برچسب گذاری داده‌ها امری هزینه بر بوده و زمان زیادی را طلب می‌کند. یکی از اهداف این پژوهش بررسی نتایج مدلی است که با دادگان به نسبت کمتری آموزش یافته باشد. برای این منظور ما به تعداد حداکثر ۲۴۰ نمونه از هرکلاس و در مجموع از ۱۰,۲۳۰ نمونه برای آموزش و اعتبار سنجی مدل استفاده کردیم که تقریبا یک چهارم نمونه‌های موجود در مجموعه داده آموزشی GTSRB است.

علاوه بر محدودیت داده یکی دیگر از چالش‌های بازشناسی تابلوهای ترافیک در عمل، ابعاد بسیار کوچکتر محدوده تابلوها به نسبت کل تصویر ضبط شده توسط دوربین خودرو است. همچنین فاصله موثر برای عکس العمل به تابلو و وضوح پایین بعضی از دوربین‌ها ایجاب می‌کند تا مدل استفاده شده قادر به تشخیص تصویر تابلوها با وضوح پایین باشد. بکارگیری تصاویری با وضوح بالا باعث افزایش دقت مدل می‌شود اما ما برای تحقق هدف ذکر شده، در مرحله پیش پردازش اندازه تصاویر را به ابعاد ۳۰×۳۰ کاهش دادیم. این کار همچنین باعث کاهش ابعاد ورودی و در نتیجه کاهش پارامترهای مدل خواهد شد.

ما در مرحله آموزش، مدل را در ۲۰ دوره[13] با دسته‌های ۳۲ تایی و با بهینه ساز آدام[14] آموزش دادیم. همچنین نرخ یادگیری را 10^{-3} تعیین کردیم و این مقدار را با افزایش دوره‌های آموزش کاهش دادیم. در انتها نیز با آزمایش مدل با مجموعه داده آزمایشی GTSRB به دقت ۹۶.۹۵٪ رسیدیم. اگرچه دقت این مدل قابل افزایش است، اما گمان می‌رود افزایش دقت موجب

سوگیری شدید مدل به مجموعه داده خود شده و کارایی آن برای استفاده به عنوان مدل پایه در رویکرد یادگیری انتقالی و آموزش بر مجموعه داده دیگر کاهش یابد.

۳-٤- آموزش بر مجموعه دادگان کشور ایران

با استفاده از رویکرد یادگیری انتقالی، دانشی که یک مدل برای یادگیری یک مسئله خاص بدست می‌آورد، می‌تواند در مسئله مشابه دیگری بکار گرفته شود. به طور خاص در شبکه‌های عصبی پیچشی، لایه‌های پیچشی اولیه قادر به استخراج ویژگی‌های سطح پایین به مانند خطوط، بافت و اشکال ساده هندسی هستند. با حرکت در عمق شبکه، لایه‌های پیچشی می‌توانند ویژگی‌های سطح بالاتری از تصویر به مانند یک بخش از یک شئ موجود در تصویر را استخراج کنند [21]. بسیاری از الگوهای استخراجی در یک مجموعه تصویر خاص می‌تواند در مجموعه دیگری نیز وجود داشته باشد. از طرفی تابلوهای ترافیک فارغ از اینکه متعلق به چه کشوری باشند اغلب از شکل‌های هندسی، حروف، ارقام و علائم جهت دار تشکیل شده اند. بدین علت محتمل است که رویکرد یادگیری انتقالی بتواند در حل مسئله بازشناسی تابلوهای ترافیک کارایی زیادی داشته باشد.

مجموعه داده PTSD شامل ۱۴,۴۰۵ تصویر برای آموزش و ۲,۴۲۱ تصویر برای آزمایش است که تمامی آن‌ها متعلق به تابلوهای ترافیکی کشور ایران بوده و همانند مجموعه داده GTSRB در ۴۳ گروه دسته بندی می‌شوند. این مجموعه از همتای آلمانی خود تعداد کمتری نمونه دارد و استفاده از آن برای مسئله‌ای نظیر بازشناسی تصویر به علت وجود نوشته‌های فارسی و تفاوت در علائم، چالش برانگیز است. ما در مرحله پیش پردازش از تکنیک تقویت تصویر[15] به منظور جلوگیری از بیش برازش و مقاوم سازی مدل نسبت به چرخش و کنتراست استفاده کردیم. اگرچه این تکنیک باعث داده افزایی نیز می‌شود، اما با استفاده از رویکرد یادگیری انتقالی، بخش قابل توجهی از آموزش شبکه در فاز اول صورت می‌پذیرد و در نتیجه نیاز به تعداد بالای نمونه آموزشی در فاز دوم کاهش می‌یابد [23].

ما در مرحله تعیین مدل، مدل آموزش یافته بر مجموعه GTSRB در فاز اول را بارگذاری کردیم و با حذف بلاک سوم یا بلاک دسته بند از دو بلاک پیچشی با وزن‌های آموزش یافته به عنوان بخش پیچشی مدل جدید خود استفاده کردیم. برای بلاک دسته بند مدل جدید نیز پس از لایه مسطح ساز، از دو لایه شبکه عصبی پیشخور کاملا متصل با ابعاد ۵۱۲ و ۶۴ به همراه تابع فعال ساز رلو استفاده کردیم. بعد از این سه لایه، همانند مدل فاز اول از یک لایه نرمالسازی دسته‌ای، یک لایه حذف تصادفی با نرخ ۰.۵ و یک لایه شبکه عصبی پیشخور کاملا متصل با ابعاد ۴۳ و با تابع فعال ساز بیشینه هموار استفاده کردیم.

به علت تصادفی بودن وزن‌های بلاک دسته بند در ابتدای آموزش، جریان گرادیان تولیدی توسط الگوریتم پس انتشار خطا باعث می‌شود تا وزن‌های بلاک‌های پیچشی که از پیش آموزش یافته‌اند دچار تغییرات ناگهانی شوند و اثر پیش آموزش از بین برود. ما برای حل این مشکل، با ثابت نگه داشتن مقادیر پارامترهای بلاک‌های پیچشی مدل را ۵ دوره آموزش دادیم. پس از آن، کل مدل را ۶۵ دوره با تنظیماتی همانند مرحله آموزش فاز اول، اما با دسته‌های ۶۴ تایی آموزش دادیم و پس از آزمایش مدل با مجموعه آزمایشی PTSD به دقت ۹۸.۵۱ رسیدیم. لازم به ذکر است که این معماری ۱,۱۸۳,۰۵۱ پارامتر داشته و به تناسب مسئله، مدلی بهینه بوده که قابلیت دارد بدون استفاده از واحدهای پردازش گرافیکی در مدت زمان قابل قبولی آموزش یابد.

٤- نتایج

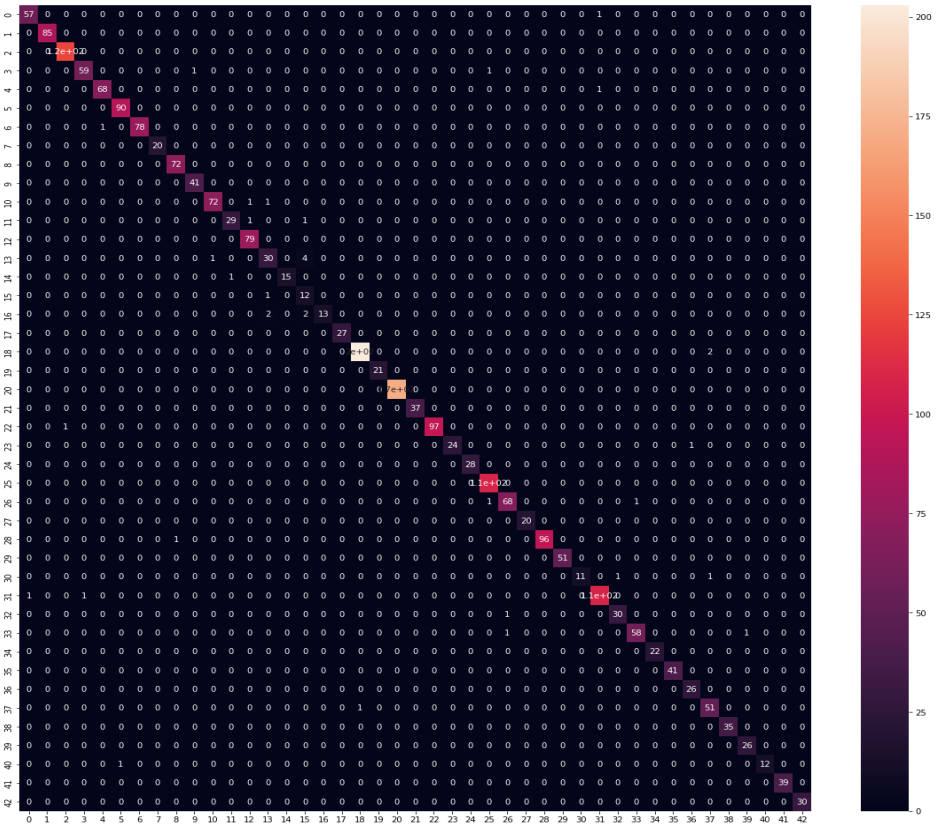
مدل پیچشی پیشنهادی ما با ۱,۱۸۳,۰۵۱ پارامتر و تابع هزینه آنتروپی متقابل پراکنده[16] و بهینه‌ساز آدام در ۷۰ دوره، دقت ۹۸.۵۱٪ را بدست آورد. دقت معیار مهمی است که در مسائل دسته بندی کاربرد دارد اما همیشه تصویر کاملی از کارایی مدل ارائه نمی‌دهد. در برخی از مسائل ممکن است بازشناسی صحیح نمونه‌های مربوط به یک دسته اهمیت بیشتری داشته باشد و یا تعداد نمونه‌های آزمایشی برای هر کلاس یکسان نباشد. بدین منظور در مسائل دسته‌بندی معیارهایی همچون صحت[17]، فراخوانی[18] و امتیاز F1 بر اساس ماتریسی درهم ریختگی[19] تعریف می‌شوند. در جدول (۲) نتایج مدل بر اساس این معیارها ارائه شده است.

تعداد داده‌های استفاده شده برای آموزش، اعتبار سنجی و آزمایش بر روی هر یک از مجموعه داده‌ها در جدول (۲) نشان داده شده است. اطلاعات موجود در سطر اول نشان می‌دهد که برای پیش آموزش مدل پیشنهادی، تقریبا از ۲۱ درصد مجموعه آموزشی دادگان آلمانی استفاده شد و دقت ۹۶.۹۵ درصد را در برداشت. استفاده از داده‌های بیشتر برای رسیدن به دقت بالاتر ثمر بخش است؛ اما به علت سوگیری شدید مدل به ویژگی‌های استخراج شده در مجموعه‌داده آلمانی، استفاده از داده‌های بیشتر انجام نشد.

برای آموزش اصلی، در ابتدا استفاده از ۲,۰۴۹ داده به منظور تنظیم دقیق بر روی مجموعه داده ایران ما را به دقت ۹۵.۱۶ رساند. بررسی ماتریس درهم ریختگی نتایج پس از این آموزش نشان داد که به علت پایین بودن کیفیت تصاویر آموزشی در تمایز دادن تابلو ورود ممنوع و ایست دچار مشکل است. به منظور افزایش دقت با توجه به سطر آخر جدول (۲) از تمامی داده‌های مجموعه داده ایران به همراه تکنیک تقویت داده استفاده شد و با افزایش تعداد داده‌ها و دوره‌های آموزش و همچنین با استفاده از زمان بندی کاهش نرخ یادگیری، به دقت ۹۸.۵۱ دست یافتیم. برای بررسی دقیق تر کارکرد مدل بر مجموعه داده PTSD، ماتریس درهم ریختگی مربوط به کارکرد مدل در تصاویر آموزشی این مجموعه در شکل (۱) آورده شده است.

جدول (۲) : نتایج بدست آمده به تفکیک مجموعه داده

مجموعه داده	آموزشی	اعتبار سنجی	آزمایشی	نوع آموزش
GTSRB	۸۱۸۴	۲۰۴۶	۱۲۶۳۰	پیش آموزش
PTSD	۱۴۴۰۵	۲۸۸۱	۲۴۲۱	آموزش اصلی
مجموعه داده	دقت(٪)	صحت(٪)	فراخوانی(٪)	F1(%)
GTSRB	۹۶.۹۵	۹۷.۱۳	۹۶.۹۵	۹۶.۹۳
PTSD	۹۸.۵۱	۹۸.۶۱	۹۸.۵۱	۹۸.۵۳



شکل (۱) : ماتریس درهم ریختگی کارکرد مدل، بر مجموعه داده PTSD

اگرچه هدف از این پژوهش رسیدن به دقت بالایی در مجموعه داده شناخته شده GTSRB نیست؛ اما مقایسه عملکرد مدل پیشنهادی در فاز پیش آموزش با روش‌های دیگران می‌تواند مفید باشد. همانطور که در جدول (۳) مشاهده می‌شود، مدل پیشنهادی ۲.۸ درصد از بهترین نتیجه موجود دقت کمتری دارد. این در حالی است که مدل ما در این فاز، تنها با یک پنجم مجموعه GTSRB و در ۲۰ دوره به صورت ابتدائی آموزش دیده است. همچنین ابعاد تصویر ورودی در این مدل نسبت به دیگر مدل‌ها کمتر است که معیار مهمی برای بکارگیری عملیاتی آن است.

جدول (۳) : مقایسه دقت مدل پیشنهادی در فاز پیش آموزش با روش-های دیگر برای مجموعه داده GTSRB

روش	ابعاد ورودی	دقت(٪)
پیشنهادی (پیش آموزش)	۳ × ۳۰ × ۳۰	۹۶.۹۵
MCDNN [14]	۳ × ۴۸ × ۴۸	۹۹.۴۶
Student Model [20]	۳ × ۳۲ × ۳۲	۹۹.۶۱
HLSGD [15]	۳ × ۴۷ × ۴۷	۹۹.۶۵
Spatial Transformer [16]	۳ × ۴۸ × ۴۸	۹۹.۷۱
میانگین دقت انسان [24]	–	۹۸.۸۴
بهترین دقت انسان [24]	–	۹۹.۲۲

۵- نتیجه

در این مقاله، روشی مبتنی بر شبکه‌های عصبی پیچشی برای امر بازشناسی تابلوهای ترافیکی کشور ایران با استفاده از یک مجموعه داده جدید ارائه شد. ما با فرض کارایی بالای رویکرد یادگیری انتقالی در این مسئله، مدل پیشنهادی خود را با مجموعه دادگان کشور آلمان آموزش دادیم؛ و با استفاده از تصاویر تابلوهای ترافیک کشور ایران، از رویکرد یادگیری انتقالی برای تنظیم دقیق این مدل بهره بردیم. ما در این پژوهش با تاکید بر کارایی میدانی سیستم، از تصاویر با ابعاد ۳۰×۳۰ برای آموزش مدل خود استفاده کردیم و برای بررسی عملکرد کیفیت یادگیری این معماری، مدل را با یک پنجم مجموعه داده آلمانی آموزش دادیم و به دقت قابل قبولی در فاز آموزش پایه دست یافتیم.

برای آموزش اصلی با مجموعه داده ایرانی نیز، با بهره‌گیری از رویکرد تقویت تصویر و کاهش نرخ یادگیری شاهد پیشرفت دقت بودیم. همانطور که پیشتر اشاره شد، بهینگی تعداد پارامترهای آموزشی یک مدل یادگیری عمیق برای استفاده در سیستم‌های توکار و بلادرنگ به مانند خودروهای خودران چالش برانگیز است. ما با تاکید بر این نکته، مدل پیشنهادی را با ۱.۱ میلیون پارامتر توسعه دادیم که در مجموعه آزمایشی تابلوهای ترافیک ایرانی به دقت ۹۸.۵۱ دست یافت. با توجه به تلاش‌های اخیر برای بهینه‌سازی مدل‌های یادگیری عمیق، استفاده از سازوکارهای تقطیر دانش و هرس مدل[۲۰]، می‌تواند در توسعه مدلی کاراتر بر پایه مدل توسعه داده شده، مفید واقع شود.

سپاسگزاری

با سپاس از خانم سارا پارسا سرشت، دانشجوی کارشناسی ارشد دانشگاه تبریز به علت بیش از یکسال تلاش بی وقفه ایشـان در جمـع آوری مجموعـه داده تابلوهای ترافیکی ایران (PTSD).

مراجع

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] S. Kheirinejad, N. Riaihi, and R. Azmi, "Persian Text Based Traffic sign Detection with Convolutional Neural Network: A New Dataset," in *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2020, pp. 60–64.

[3] B. Hoferlin and K. Zimmermann, "Towards reliable traffic sign recognition," in *2009 IEEE Intelligent Vehicles Symposium*, 2009, pp. 324–329.

[4] F. Moutarde, A. Bargeton, A. Herbin, and L. Chanussot, "Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular Traffic Signs Recognition system," in *2007 IEEE intelligent vehicles symposium*, 2007, pp. 1122–1126.

[5] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, 2005, pp. 255–260.

[6] A. de La Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE transactions on industrial electronics*, vol. 44, no. 6, pp. 848–859, 1997.

[7] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.

[8] G. Antipov, S.-A. Berrani, N. Ruchaud, and J.-L. Dugelay, "Learned vs. hand-crafted features for pedestrian gender recognition," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1263–1266.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[12] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," in *The 2011 international joint conference on neural networks*, 2011, pp. 1453–1460.

[13] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *The 2011 International Joint Conference on Neural Networks*, 2011, pp. 2809–2813.

[14] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*, 2012, pp. 3642–3649.

[15] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE transactions on intelligent transportation systems*, vol. 15, no. 5, pp. 1991–2000, 2014.

[16] Á. Arcos-García, J. A. Alvarez-Garcia, and L. M. Soria-Morillo, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods," *Neural Networks*, vol. 99, pp. 158–165, 2018.

[17] F. Franzen, C. Yuan, and Z. Li, "Traffic sign recognition with neural networks in the frequency domain," in *Journal of physics: Conference series*, 2020, vol. 1576, no. 1, p. 012015.

[18] C. Lin, L. Li, W. Luo, K. C. P. Wang, and J. Guo, "Transfer learning based traffic sign recognition using inception-v3 model," *Periodica Polytechnica Transportation Engineering*, vol. 47, no. 3, pp. 242–250, 2019.

[19] W. A. Haque, S. Arefin, A. S. M. Shihavuddin, and M. A. Hasan, "DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements," *Expert Systems with Applications*, vol. 168, p. 114481, 2021.

[20] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7, pp. 369–379, 2020.

[21] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*, 2017, pp. 1–6.

[22] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger, "Understanding batch normalization," *Advances in neural information processing systems*, vol. 31, 2018.

[23] L. Jöckel, M. Kläs, and S. Martínez-Fernández, "Safe traffic sign recognition through data augmentation for autonomous vehicles software," in *2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 2019, pp. 540–541.

[24] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural networks*, vol. 32, pp. 323–332, 2012.

زیرنویس‌ها

[1] Convolutional neural networks
[2] Fine-tuning
[3] Hinge loss
[4] Local response normalization
[5] Spatial transformers
[6] Knowledge distillation
[7] Feature map
[8] Dense feed-forward neural network
[9] RELU: Rectified linear unit
[10] Batch normalization

11 Softmax activation function
12 Dropout
13 Epoch
14 Adam optimizer
15 Image augmentation
16 Cross-entropy loss
17 Precision
18 Recall
19 Confusion matrix
20 Model pruning