

سامانه هوشمند تشخیص نوع بدنه خودرو در کابین پاشش رنگ رباتیک با استفاده از یادگیری عمیق

میرمسعود حسینی کوکمری^۱، رضا آقائی زاده ظروفی^۲

چکیده

– شناسایی دقیق پلتفرم بدنه خودرو پیش از ورود به کابین پاشش رنگ، برای جلوگیری از خطاهای پرهزینه، بهینه‌سازی مصرف مواد و افزایش قابلیت اطمینان در خط تولید خودرو امری حیاتی است. سیستم‌های متداول مبتنی بر حسگرهای نوری به دلیل ناتوانی در تمایز بدنه‌های مشابه و پردازش بلادرنگ، فاقد دقت لازم در محیط پیچیده صنعتی هستند. این پژوهش به طراحی و پیاده‌سازی یک سامانه هوشمند مبتنی بر پردازش تصویر و یادگیری عمیق می‌پردازد که قادر به تشخیص خودکار و دقیق سه پلتفرم مختلف بدنه خودرو (SP100، X200 و X200 بهینه) است. در معماری پیشنهادی، از پایه ResNet همراه با انتقال یادگیری و تکنیک‌های پیشرفته افزایش داده برای تطبیق پذیری و مقاوم‌سازی در برابر شرایط متغیر محیط صنعتی (تغییرات نور و پیچیدگی پس‌زمینه) استفاده شد. مدل بر روی مجموعه داده‌ای جامع و واقعی، جمع‌آوری شده از خط تولید، آموزش و به‌طور گسترده ارزیابی گردید. نتایج تجربی نشان داد که مدل ResNet-50 به دقت تشخیص ۱۰۰٪ و سطح اطمینان متوسط بالای ۹۸٪ در محیط عملیاتی و پرچالش دست یافته است. این سامانه از پایداری و قابلیت اطمینان بالایی برخوردار است و پتانسیل کامل برای جایگزینی با سیستم‌های سنتی به منظور کاهش خطا، صرفه‌جویی در مصرف مواد و کاهش وابستگی به مداخله انسانی در خط رنگ خودرو را دارا می‌باشد.

کلید واژه‌ها

تشخیص نوع بدنه خودرو، یادگیری عمیق، پردازش تصویر، بینایی ماشین، کابین پاشش رنگ

۱- مقدمه

سامانه‌های فعلی تشخیص بدنه، عمدتاً مبتنی بر فناوری‌های صنعتی مانند پرده‌های نوری، سنسورهای تشخیص هویت با فرکانس رادیویی [۲] و حسگرهای سه‌بعدی [۳] هستند. این سامانه‌ها اگرچه در شرایط استاندارد عملکرد قابل قبولی دارند، اما در مواجهه با بدنه خودروهای مشابه با پلتفرم یکسان (مانند پلتفرم X200 خودرو کوئیک و پلتفرم X200 بهینه خودرو اطلس شرکت سایپا) یا ویژگی‌های پیچیده (مانند وجود سانروف) با محدودیت‌های جدی روبه‌رو می‌شوند. به‌عنوان مثال، سامانه پرده نوری مورد استفاده در شرکت سایپا که توسط شرکت سمز توسعه یافته، تنها قادر به اندازه‌گیری مقاطع طولی بدنه است و تفاوت‌های ظریف در نمای جلو، عقب یا سقف را تشخیص نمی‌دهد. این نقصان منجر به خطا در تنظیمات ربات‌های پاشش رنگ، اتلاف مواد و افزایش هزینه‌های تولید می‌شود. از سوی دیگر، روش‌هایی مانند تشخیص هویت با فرکانس رادیویی نیازمند زیرساخت‌های

در صنعت خودروسازی، فرآیند رنگ‌آمیزی بدنه خودرو یکی از مراحل حیاتی و پرهزینه است که نیازمند دقت بالا در شناسایی نوع بدنه و تطبیق پارامترهای پاشش رنگ با ویژگی‌های آن است [۱].

مقاله در تاریخ ۲۲ تیر ماه ۱۴۰۴ دریافت شد

^۱ کارشناس ارشد مهندسی برق، گروه خودروسازی سایپا، تهران

رایانامه: hosseini_mi@saipacorp.com

^۲ استاد، دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

رایانامه: zoroofi@ut.ac.ir

نویسنده مسئول: میرمسعود حسینی کوکمری

ناظمی و همکاران^{۱۳} نشان می‌دهد که تشخیص مدل خودرو^۶ به عنوان یک مسئله طبقه‌بندی دقیق^۷ نیازمند استخراج ویژگی‌های غیرخطی و مقاوم به شرایط محیطی است. نویسندگان با استفاده از کدگذاری خطی با محدودیت مکانی^۸ و یادگیری بدون نظارت، به دقت ۹۷/۵٪ در مجموعه داده‌های واقعی (شامل تصاویر با نور نامناسب و بازتاب نور خورشید) دست یافتند. این پژوهش تأکید می‌کند که روش‌های مبتنی بر ویژگی‌هایی مانند SIFT در مقایسه با معماری‌های عمیق، انعطاف‌پذیری کمتری در مواجهه با تغییرات محیطی دارند. از آنجایی که در پروژه ما، محیط پس زمینه نهایتاً یک محیط متغیر و دارای شرایط نوری غیر ثابت می‌باشد. برای مقابله با چالش شرایط نوری متغیر و پس‌زمینه‌ی غیریکنواخت، معماری ResNet انتخاب گردید. دلیل این انتخاب، توانایی ذاتی این معماری در یادگیری ویژگی‌های انتزاعی و غیرخطی از داده‌های خام است. به طور خاص، مکانیزم اتصالات پرسی در ResNet که مشکل گرادیان ناپدیدشونده را حل می‌کند، امکان آموزش شبکه‌های بسیار عمیق‌تر را فراهم می‌سازد. شبکه‌های عمیق‌تر قادر به یادگیری سلسله‌مراتب پیچیده‌تری از ویژگی‌ها هستند - از لبه‌ها و بافت‌های ساده تا اشکال و اجزای پیچیده - که نسبت به تغییرات ظاهری ناشی از نورپردازی متفاوت، مقاوم‌تر هستند. این توانایی در کنار تکنیک افزایش داده که طی آموزش اعمال شد، مدل را برای شرایط غیرایده‌آل محیط صنعتی آماده می‌کند.

در این پژوهش، یک سامانه هوشمند مبتنی بر پردازش تصویر و شبکه عصبی ResNet برای تشخیص دقیق بدنه خودرو در کابین پاشش رنگ طراحی شده است. خط تولیدی که این سامانه به آن اضافه شده کابین پاشش رنگ رویه رباتیک سالن رنگ ۱ شرکت سایپا می‌باشد. نوآوری اصلی این تحقیق، ترکیب انتقال یادگیری با تکنیک‌های افزایش داده برای تطبیق مدل با شرایط محیطی پیچیده است. داده‌های آموزشی از تصاویر بدنه‌های خودروهای X200 (خودرو کوییک)، SP100 (خودرو شاهین) و X200 بهینه (خودرو اطلس) در دو محیط داخل کابین پاشش رنگ (با نور کنترل‌شده و پس‌زمینه ساده) و خارج کابین پاشش رنگ (با نور طبیعی و پس‌زمینه پیچیده و متغیر) جمع‌آوری شدند. سپس، با استفاده از معماری‌های ResNet-18 و ResNet-50، دو مدل آموزش داده شدند و عملکرد آن‌ها بر اساس معیارهای دقت^۹، اطمینان^{۱۰} مورد ارزیابی قرار گرفت.

ساختار این مقاله به صورت زیر است: در بخش مواد و روش‌ها، مراحل جمع‌آوری داده‌ها، پیش‌پردازش تصاویر و آموزش مدل‌ها تشریح شده است. در بخش نتایج، عملکرد سامانه توسعه داده شده با شبکه‌های عصبی عمیق بکار رفته با استفاده از جداول

پیچیده و نگهداری مداوم هستند، در حالی که حسگرهای سه‌بعدی با وجود دقت بالا، هزینه‌های اجرایی سنگینی را تحمیل می‌کنند. با گسترش فناوری‌های بینایی ماشینی و یادگیری عمیق، امکان جایگزینی سامانه‌های سنتی با راهکارهای هوشمندتر فراهم شده است. [۴] این فناوری‌ها با تحلیل تصاویر بدنه خودرو و استخراج ویژگی‌های ظریف، قادرند حتی تفاوت‌های جزئی بین مدل‌های مشابه را شناسایی کنند. در سال‌های اخیر، مطالعات متعددی از شبکه‌های عصبی پیچیده مانند ResNet و YOLO برای تشخیص اشیاء در محیط‌های صنعتی استفاده کرده‌اند. کاربرد معماری‌های عمیق در محیط صنعتی در مطالعات متعددی اشاره شده است. برای نمونه، کاربرد مدل‌های YOLO در کارهایی مانند بازرسی [۵] و تشخیص نقص [۶، ۷] موفقیت‌آمیز بوده است. در همین راستا، معماری‌های مبتنی بر ResNet نیز به دلیل توانایی در استخراج ویژگی‌های پیچیده، به طور گسترده در تشخیص اشیاء صنعتی [۸] و حتی در حوزه‌های خاص مانند تولید الوار [۹] به کار گرفته شده‌اند. علاوه بر این، رویکردهای پیشرفته‌تری مانند شبکه‌های کانولوشنی عمیق چندمسیره نیز برای تشخیص دقیق مدل خودرو [۱۰] توسعه یافته‌اند که نشان‌دهنده پتانسیل بالای این فناوری است. با این حال، چالش اصلی در به‌کارگیری این مدل‌ها، سازگاری آن‌ها با شرایط واقعی خطوط تولید (مانند نور متغیر، پس‌زمینه شلوغ و محدودیت‌های ایمنی) است. در سال‌های اخیر، روش‌های شناسایی دقیق نوع خودرو به‌ویژه در کاربردهایی نظیر سامانه‌های حمل‌ونقل هوشمند، پارکینگ‌های خودکار و خطوط تولید خودکار، مورد توجه بسیاری از محققین قرار گرفته‌اند.

وَنگ و همکاران^{۱۱} با معرفی معماری چندمسیره^۲ نشان دادند که استفاده همزمان از نواحی خاصی مانند نمای جلو و لوگوی خودرو، به شکل چشمگیری می‌تواند دقت تشخیص را در مواجهه با خودروهای مشابه افزایش دهد. این رویکرد با یافته‌های عملی ما در محیط صنعتی تطابق داشته و تأکید مجددی بر اهمیت زاویه دید مناسب و استخراج ویژگی‌های متمایز از نمای جلوی خودرو دارد. با این حال شرایط ما به گونه‌ای است که به دلیل داشتن بدنه مونتاژ نشده، شرایط تشخیص برخی ویژگی‌ها مانند لوگوی خودرو برای کاربرد ما، امکان‌پذیر نیست.

مطالعه‌ی هاتون و همکاران^{۱۲} نشان می‌دهد که سامانه‌های مبتنی بر شبکه‌های عصبی عمیق^۴ می‌توانند با دقت ۹۷٪ نوع خودرو (مانند ون، کامیون، و خودروهای سواری) را حتی در شرایط محیطی متغیر (نور کم، باران، برف) تشخیص دهند. این پژوهش تأکید می‌کند که معماری‌های عمیق مانند ResNet با حذف نیاز به پیش‌پردازش دستی، بر روش‌های سنتی مبتنی بر ویژگی‌های مهندسی شده مانند SIFT + SVM برتری دارند. مطالعه

⁶ MMR: Make and Model Recognition

⁷ Fine-Grained Classification

⁸ LLC: Linear Locally Constrained Coding

⁹ Accuracy

¹⁰ Confidence

¹ Wang et al

² Multi-path DCNN

³ Huttunen et al

⁴ Deep neural networks

⁵ Nazemi et al

روش‌های مبتنی بر HOG و SVM یا قوانین قطعی، وابستگی شدیدی به مکان دقیق پیکسل‌ها دارند و در برابر این تغییرات مکانی کوچک، بسیار شکننده عمل می‌کنند. یک جابجایی جزئی می‌تواند هیستوگرام گرادینان‌ها (HOG) یا اندازه‌گیری‌های هندسی را به‌طور کامل دگرگون ساخته و منجر به خطای فاحش در تشخیص شود. تجربه عملی ما در این پروژه با روش SVM با خطا همراه بود. در مقابل، شبکه عمیق ResNet با یادگیری هزاران ویژگی سلسله‌مراتبی و انتزاعی، به‌طور ذاتی در برابر این تغییرات مکانی مقاومت بالایی دارد. این شبکه بر «مفهوم» هر پلتفرم (مانند شکل کلی، نسبت‌ها و طرح‌های کلی) تمرکز می‌کند، نه بر مکان دقیق لبه‌ها در یک تصویر ثابت. این قابلیت، راهکار ممکن برای دستیابی به دقت ۱۰۰٪ و قابلیت اطمینان در یک محیط پویا و غیرایده‌آل مانند خط تولید بود.

۲-۱- مجموعه داده‌ها

در این پژوهش، برای توسعه و آموزش سامانه تشخیص نوع بدنه خودرو، در مجموع از ۲۵۲ تصویر استفاده شد. تصاویر مورد استفاده به سه کلاس اصلی شامل پلتفرم X200 (شکل ۳) شامل ۱۲۱ تصویر، پلتفرم SP100 (شکل ۴) شامل ۷۵ تصویر و پلتفرم X200 بهینه (شکل ۵) شامل ۵۶ تصویر تعلق دارند. نمونه‌های تصویری این سه کلاس، تنوعی از زوایا، شرایط نوری و پس‌زمینه‌های متفاوت را پوشش می‌دهند.



شکل ۳: پلتفرم X200

مقایسه‌ای و نمودارهای هیستوگرام تحلیل می‌شود و نهایتاً راهکارهایی برای بهبود سیستم در آینده ارائه می‌گردد.

۲- روش

در این پژوهش، انتخاب روش با ارزیابی روش ماشین بردار پشتیبان آغاز شد. همان‌طور که در شکل ۱ نشان داده شده است، در خط تولید، بدنه‌های خودرو بر روی یک سازه فلزی قرار می‌گیرند. قرارگیری بدنه‌ها روی این سازه با تلرانس همراه است. علاوه بر این، سازه فلزی بر روی چرخنده‌هایی قرار دارد که از طریق تسمه به موتور الکتریکی متصل هستند. میزهای انتقال متحرک از سنسورهای القایی برای تشخیص حضور بدنه خودروی واردشونده به میز انتقال استفاده می‌کنند. با تشخیص سنسور، حرکت موتور الکتریکی متوقف می‌شود، اما وجود لقی در ساختار میز، کشیدگی تسمه و سیستم ترمز موتور الکتریکی منجر به توقف بدنه خودرو در نقاط مختلفی از یک بازه مشخص می‌شود. نکته حائز اهمیت این است که مطابق شکل ۲، مکان نصب این سنسورها قابلیت جابجایی دارد و هیچ تضمینی وجود ندارد که بدنه خودرو دقیقاً در یک موقعیت ثابت و مشخص متوقف شود؛ به‌ویژه آنکه در زمان‌های سرویس دوره‌ای، این سنسورها جابجا می‌شوند. این تغییرات جزئی در موقعیت توقف بدنه خودرو (حتی در حد چند سانتیمتر) موجب تغییر در زاویه دید و پرسپکتیو دوربین، جابجایی موقعیت بدنه در کادر تصویر و تغییر در سایه‌ها و انعکاس‌ها می‌گردد.



شکل ۱: بدنه خودرو و سازه فلزی روی میز انتقال



شکل ۴: پلتفرم SP100



شکل ۲: سنسور القایی و محل نصب دارای تنظیم محل نصب

● فاز دوم (نهایی):

در مرحله نهایی توسعه، یک دوربین تحت شبکه از برند هایکوژن^۲ در موقعیتی ثابت پیش از ورودی کابین پاشش رباتیک رنگ رویه نصب شد. با استفاده از پروتکل پخش زنده^۳ [۱۴] و استفاده از FFMPEG^۴ [۱۵]، تصاویر با وضوح ۱۲۸۰×۷۲۰ پیکسل از بدنه خودروهای عبوری از خط تولید جمع‌آوری گردید. داده‌های این فاز به دلیل واقع‌گرایانه بودن شرایط محیطی (نور طبیعی، پس‌زمینه صنعتی و زوایای غیرایده‌آل) جهت آموزش نهایی مدل مورد استفاده قرار گرفتند.

به جز تصاویر مربوط به آموزش و اعتبارسنجی نرم افزار از ۳۹ تصویر X200، ۲۲ تصویر SP100 و ۲۶ تصویر X200 بهینه برای ارزیابی نحوه عملکرد مدل استفاده شد.

۲-۲- روش ارائه شده

سامانه تشخیص نوع بدنه می‌بایست با زیر سیستم‌های دیگر خط رنگ آمیزی رباتیک ارتباط داشته باشد. به منظور تشخیص قرار گیری بدنه خودرو در موقعیت مناسب از دوربین نصب شده قبل از ورودی کابین پاشش رنگ یک بازخورد از سنسور نصب شده در کانوایر خط انتقال قبل از کابین پاشش گرفته می‌شود. این سیگنال بازخورد از طریق سامانه تشخیص بدنه دریافت شده و فرمان ذخیره تصویر جدید را از داده‌های بلادرنگ دوربین اجرا می‌کند. پس از این مرحله فرایند تشخیص نوع بدنه خودرو و طبقه‌بندی کلاس با در اختیار بودن مدل آموزش دیده شده با استفاده از پایگاه داده سامانه انجام می‌شود. در این روش ابتدا یک فرایند پیش پردازش روی تصاویر انجام می‌شود سپس داده‌ها در یک پروسه افزایش داده به عنوان ورودی شبکه عصبی استفاده می‌شوند تا تشخیص سه نوع کلاس بدنه خودرو امکان پذیر شود. پس از طبقه‌بندی و تشخیص نوع بدنه سیگنال مناسب برای کنترل‌کننده ربات‌های پاشش ارسال می‌شود تا بر اساس آن مسیر حرکتی ربات‌های پاشش تعیین گردد.

شکل ۷ نشان‌گر بلوک دیاگرام کلی سامانه می‌باشد. برای سامانه قابلیت نگهداری و به روزرسانی نیز پیش‌بینی شده است. در صورت تغییر اساسی در محیط عملیاتی (مانند نصب سیستم در یک کابین رباتیک جدید با شرایط نوری یا پس‌زمینه کاملاً متفاوت)، اپراتور سیستم می‌تواند به صورت دستی فرایند بازآموزی مدل را آغاز نماید. مسیر غیر عملیاتی بلوک دیاگرام بیانگر این عملکرد می‌باشد. در این فرایند، از داده‌های تصویری که در طول عملیات عادی سیستم در حافظه رایانه ذخیره شده‌اند نیز می‌توان استفاده کرد پس از بروز رسانی پایگاه داده سامانه و آموزش مجدد مدل، در نهایت، مدل جدید به صورت دستی جایگزین مدل قبلی



شکل ۵: پلنفرم X200 بهینه



شکل ۶: نمونه تصاویر استفاده شده در فاز ۱ برای آموزش مدل با نمای روبرو و نور و پس‌زمینه مناسب

مجموعه داده در دو فاز تهیه و مورد استفاده قرار گرفت که نمونه تصویر استفاده شده در فاز ۱ را در شکل ۶ و نمونه تصویر استفاده شده در فاز ۲ را در اشکال ۳ تا ۵ ملاحظه می‌کنید. این فازبندی با در نظر گرفتن ملاحظات عملی و لجستیکی پروژه طراحی شد. با توجه به اینکه نصب و یکپارچه‌سازی دوربین در خط تولید فعال، نیازمند طی مراحل فنی، ایمنی و هماهنگی‌های زمان‌بر بود، فاز اول (آزمایشی) با استفاده از داده‌های جمع‌آوری‌شده توسط تلفن همراه و در شرایط کنترل‌شده آغاز گردید. این رویکرد به موازات فرایند نصب، به ما امکان داد تا زمان پروژه را بهینه کرده و توسعه و عیب‌یابی اولیه الگوریتم‌ها، پیش‌پردازش داده‌ها و ارزیابی اولیه پتانسیل روش را بدون وقفه پیش ببریم. پس از تکمیل نصب دوربین صنعتی، فاز دوم (نهایی) با استفاده از داده‌های واقعی از محیط عملیاتی خط تولید آغاز شد.

● فاز اول (آزمایشی):

تصاویر اولیه با استفاده از دوربین تلفن همراه اندرویدی با وضوح ۱۲۸۰×۵۷۶ پیکسل در شرایط کنترل‌شده و با پس‌زمینه نسبتاً ساده تهیه شدند. این تصاویر برای طراحی و ارزیابی اولیه مدل مورد استفاده قرار گرفتند. مکان ثبت این تصاویر برای توسعه و آموزش مدل ناحیه اول کابین پاشش رباتیک رنگ رویه موسوم به تک‌رگ^۱ می‌باشد. سامانه تشخیص بدنه قابلیت دریافت این تصاویر از طریق فایل را داراست.

^۱ Tack Rag

^۲ Hikvision

^۳ RTSP:Real-Time Streaming Protocol

^۴ Fast Forward Moving Picture Experts Group

تصادفی در بازه‌ی ۸۰٪ تا ۱۲۰٪ در هر دو محور X و Y است. هدف از این تبدیلات، شبیه‌سازی شرایط مختلف تصویربرداری در محیط‌های صنعتی و افزایش مقاومت مدل در برابر تغییرات جزئی در ورودی است.

یک پایگاه داده‌ی افزوده‌شده برای تصاویر آموزشی ایجاد گردید که در آن داده‌ها پیش از هر بار تغذیه به شبکه، به صورت تصادفی تغییر می‌یابند. این پایگاه داده به عنوان ورودی مرحله‌ی آموزش مورد استفاده قرار گرفت. در مقابل، برای مجموعه‌ی اعتبارسنجی، تصاویر بدون اعمال افزایش داده و تنها با تطبیق اندازه، به مدل ارائه شدند.

۳-۲-۲- آموزش مدل

در مرحله‌ی تنظیم نهایی معماری شبکه عصبی، به منظور انطباق مدل پیش‌آموزش‌دیده‌ی ResNet با مسئله‌ی طبقه‌بندی جدید، لایه‌های خروجی آن متناسب با تعداد کلاس‌های داده‌های مورد نظر بازطراحی شدند. در این راستا، ابتدا گراف لایه‌ی شبکه‌ی پایه استخراج گردید و سپس تعداد کلاس‌ها محاسبه شد.

به‌منظور جایگزینی لایه‌های انتهایی، لایه‌ی Fully Connected موجود با یک لایه‌ی جدید از نوع fullyConnectedLayer با تعداد نوروهای برابر با تعداد کلاس‌ها جایگزین شد. در این لایه‌ی جدید پارامترهای WeightLearnRateFactor و BiasLearnRateFactor هر دو بر روی مقدار ۱۰ تنظیم شدند تا نرخ یادگیری برای وزن‌ها و بایاس‌ها در این لایه افزایش یابد و امکان بهینه‌سازی سریع‌تر و مؤثرتر این لایه‌ها را فراهم شود، در شبکه‌های عمیق، پارامترهای هر لایه در طول فرآیند آموزش با استفاده از الگوریتم به‌روزرسانی وزن‌ها همچون گرادیان نزولی تصادفی یا Adam به‌روزرسانی می‌شوند. روش بهینه‌سازی استفاده شده در فرآیند آموزش شبکه، گرادیان نزولی تصادفی^۴ است. در این روش، وزن‌ها و بایاس‌های شبکه عصبی با استفاده از گرادیان تابع خطا نسبت به پارامترها و بر مبنای داده‌های مینی‌باتچ به‌روزرسانی می‌شوند. معادله به‌روزرسانی کلی برای وزن‌ها در این روش به صورت زیر است:

$$w(t+1) = w(t) - \eta \cdot \nabla wL(w)$$

که در آن:

$w(t)$ وزن در گام t ام است،

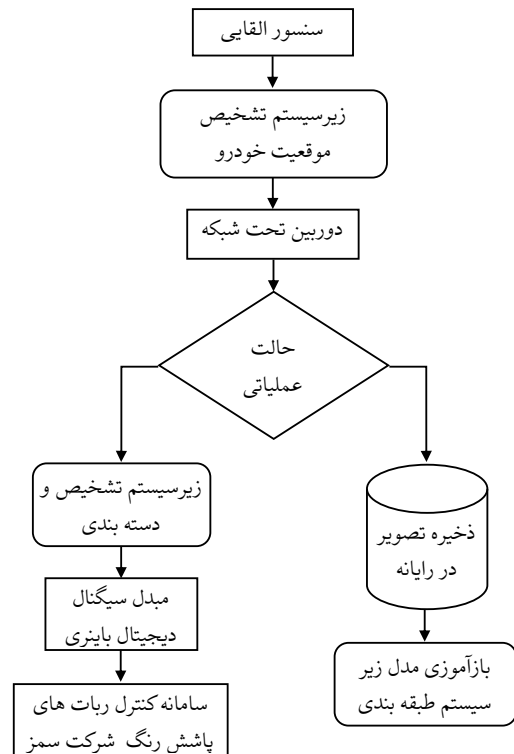
η نرخ یادگیری اولیه^۵ است (در پروژه برابر با ۰/۰۰۱)،

$\nabla wL(w)$ گرادیان تابع زیان^۵ نسبت به وزن‌هاست.

در صورت استفاده از ضریب‌های WeightLearnRateFactor و BiasLearnRateFactor، این رابطه برای وزن‌ها و بایاس‌ها به صورت زیر اصلاح می‌شود:

$$w(t+1) = w(t) - \eta \cdot \alpha w \cdot \nabla wL(w)$$

می‌شود. این قابلیت، انعطاف‌پذیری سامانه را در مواجهه با تغییرات اساسی در محیط عملیاتی به میزان قابل توجهی افزایش می‌دهد.



شکل ۷: بلوک دیاگرام کلی سامانه تشخیص نوع بدنه خودرو

۱-۲-۲- پیش پردازش

برای تضمین پایداری و دقت بالای سامانه شناسایی نوع بدنه در شرایط صنعتی و متنوع، پیش‌پردازش تصاویر ورودی به صورت استانداردسازی ابعاد و فرمت تصاویر انجام می‌شود. همه‌ی تصاویر ابتدا با استفاده از تابع متلب با حفظ نسبت ابعاد به اندازه‌ی ثابت 224×224 پیکسل تغییر اندازه می‌یابند تا با ورودی مورد انتظار شبکه‌های ResNet-18 و ResNet-50 هم‌ساز گردند.

در صورتی که تصویر ورودی به صورت تک‌کاناله سیاه و سفید باشد، هر پیکسل در سه کانال R، G، و B تکرار شده و تصویر به فرمت سه‌کاناله تبدیل می‌شود تا مطابق ساختار لایه‌ی ابتدایی شبکه قرار گیرد.

۲-۲-۲- افزایش داده

در راستای افزایش تنوع داده‌های آموزشی و بهبود توان تعمیم‌دهی مدل شبکه عصبی کانولوشنی^۱، تکنیک افزایش داده‌ها به‌کار گرفته شد. این تبدیلات شامل چرخش تصادفی در بازه‌ی $\pm 10^\circ$ درجه، انتقال افقی و عمودی به‌میزان حداکثر ± 5 پیکسل، و مقیاس‌گذاری

⁴ Initial Learning Rate

⁵ Loss Function

¹ CNN

² Layer Graph

³ SGD: Stochastic Gradient Descent

انتخاب MiniBatchSize کوچک:

- محاسبه گرادینانها بسیار نویزدار می‌شود، چون هر بار فقط با تعداد بسیار کمی از نمونه‌ها انجام می‌شود.
 - ممکن است مدل نوسان زیادی در فرآیند یادگیری داشته باشد و همگرایی ضعیف یا پرنوسان باشد.
 - زمان آموزش بسیار افزایش می‌یابد، چون تعداد دفعات به‌روزرسانی زیاد می‌شود.
- انتخاب MiniBatchSize بزرگ:
- مدل به سمت میانگین داده‌ها گرایش پیدا می‌کند و ممکن بود از یادگیری جزئیات مهم در داده‌ها غافل شود.
 - مصرف حافظه‌ی GPU یا RAM به شدت بالا می‌رود و احتمال خطای حافظه یا کاهش کارایی وجود دارد.
 - همچنین در دیتاست‌های کوچک ممکن است مدل نتواند تفاوت‌های کلاس‌ها را به‌خوبی درک کند.
- انتخاب MiniBatchSize بین ۲۸ یا ۳۲:
- این بازه اندازه‌ی نسبتاً متوسطی دارد که باعث می‌شود نویز گرادینانها کمتر باشد، اما همچنان مدل تنوع خوبی از داده‌ها را در هر به‌روزرسانی می‌بیند.
 - منابع محاسباتی در این بازه به شکل مناسبی استفاده می‌شوند (نه خیلی زیاد که حافظه پر شود، نه خیلی کم که آموزش خیلی کند شود).
 - مناسب برای دیتاست‌هایی با حجم نسبتاً کم.

در صورتی که تعداد کل نمونه‌های آموزشی بر این مقدار بخش‌پذیر نباشد، محیط متلب به‌صورت پیش‌فرض، batch ناقص انتهایی را نادیده می‌گیرد و از آن در روند آموزش استفاده نمی‌کند. [۱۶] این رویکرد به حفظ ثبات در اندازه‌ی ورودی هر به‌روزرسانی و جلوگیری از بروز نوسان در گرادینانها کمک می‌کند. در جدول ۱ مقایسه بین دو مقدار ۲۸ و ۳۲ برای انتخاب minibatchsize مناسب نشان داده شده است.

جدول ۱: مقایسه minibatchsize

معیار	اندازه گروه داده=۳۲	اندازه گروه داده=۲۸
تصاویر استفاده‌شده	۱۹۲ (۹۵/۵٪≈)	۱۹۶ (۹۷/۵٪≈)
Iteration/Epoch	۶	۷

با توجه به این که با minibatchsize با مقدار ۲۸، داده آموزشی بیشتری وارد شبکه عصبی می‌شود این مقدار را انتخاب می‌کنیم. تعداد به‌روزرسانی^۳ در هر epoch نیز به صورت زیر محاسبه می‌شود:

$$\text{IterationsperEpoch} = \frac{N}{\text{minibatchsize}} = \frac{202}{28} \approx 7$$

در فرآیند آزمایشی انتخاب تعداد epochs، مشاهده شد که استفاده از مقادیر کمتر از ۱۰ منجر به دقت نامطلوب مدل و ناکامل بودن فرآیند یادگیری می‌شود؛ به‌طوری‌که مدل هنوز به سطح

$$b(t+1) = b(t) - \eta \cdot ab \cdot \nabla bL(b)$$

که در آن:

ab و aw به ترتیب ضریب‌های یادگیری وزن و بایاس هستند (در این پروژ به برابر با ۱۰)، b نمایانگر بایاس‌ها است.

در فرآیند آموزش مدل یادگیری عمیق، تفکیک داده‌ها به مجموعه‌های آموزش و اعتبارسنجی نقش کلیدی در جلوگیری از بیش‌برازش و ارزیابی عملکرد تعمیم‌پذیری مدل دارد. در این پروژه، تقسیم تصادفی داده‌های تصویر به دو مجموعه‌ی آموزش و اعتبارسنجی انجام شده است. به طور خاص، ۸۰٪ از داده‌ها به عنوان مجموعه‌ی آموزش^۱ و ۲۰٪ باقی‌مانده به عنوان مجموعه‌ی اعتبارسنجی^۲ تخصیص یافته‌اند. این تفکیک با حفظ نسبت تعداد نمونه‌ها در هر کلاس انجام می‌شود تا تعادل آماری کلاس‌ها در هر دو زیرمجموعه حفظ گردد. در پیاده‌سازی، تقسیم داده‌ها به صورت تصادفی و غیرقطعی پیاده‌سازی شده است تا در افزایش اعتبار ارزیابی مدل مؤثر باشد. این رویکرد مطابق با رویه‌های استاندارد در آموزش شبکه‌های عمیق مبتنی بر پردازش تصویر می‌باشد. با توجه به توضیحات فوق تعداد تصاویر استفاده شده برای آموزش برابر است با:

$$252 \times 0.8 = 201.6 \approx 202$$

استفاده از گرادینان نزولی تصادفی در ترکیب با داده‌های افزایش یافته و نرخ‌های یادگیری تقویتی برای لایه‌های جدید، منجر به همگرایی سریع‌تر و دقیق‌تر مدل در مواجهه با مجموعه داده‌های تخصصی‌تر مانند تصاویر خودرو در محیط‌های صنعتی شده است. در فرآیند آموزش شبکه‌های عصبی عمیق، پارامترهایی مانند MaxEpochs، MiniBatchSize، و ValidationFrequency نقش مهمی در پایداری و سرعت همگرایی مدل دارند. در به‌روزرسانی وزن‌ها با الگوریتم گرادینان نزولی تصادفی، استفاده از مینی‌بچ باعث کاهش نوسانات ناخواسته در گرادینانها و بهبود پایداری آموزش می‌شود. به‌جای محاسبه گرادینان کل تابع زیان $L(w)$ بر روی تمام داده‌ها، گرادینان به‌صورت تقریبی روی یک دسته کوچک از داده‌ها $B \subset B$ محاسبه می‌شود:

$$\nabla L(w) \approx \frac{1}{|B|} \sum_{i \in B} \nabla \ell(w; x_i, y_i)$$

که در آن:

$$\nabla L(w) \text{ گرادینان کل تابع زیان،}$$

B مجموعه‌ای از داده‌های انتخاب‌شده به‌صورت تصادفی

در هر mini batch

$$\nabla l(w; x_i, y_i) \text{ گرادینان زیان برای نمونه } (x_i, y_i)$$

$|B|$ تعداد نمونه‌ها در mini batch

انتخاب MiniBatchSize، تعادلی میان دقت گرادینان و

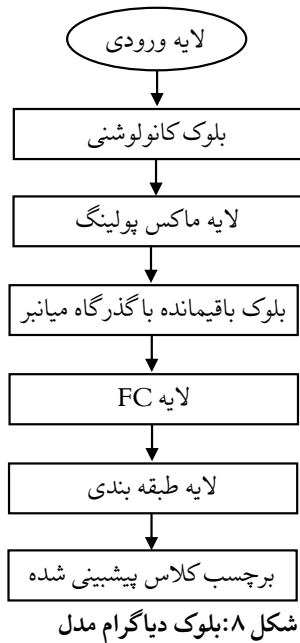
سرعت محاسباتی ایجاد می‌کند.

³ Iteration

¹ Training Set

² Validation Set

نرمال‌سازی دسته‌ای، یک تابع فعال‌ساز ReLU و یک لایه ادغام حداکثری قرار دارد. سپس، داده‌ها از طریق چهار مرحله متوالی از بلوک‌های باقیمانده عبور می‌کنند که هسته اصلی این معماری را تشکیل می‌دهند. هر بلوک از چندین لایه کانولوشنی (با هسته‌های 1×1 ، 3×3 و 1×1)، لایه‌های نرمال‌سازی دسته‌ای و توابع فعال‌ساز ReLU تشکیل شده است. مؤلفه حیاتی این بلوک‌ها، اتصال پرشی است که ورودی بلوک را مستقیماً به خروجی آن می‌افزاید. این مکانیزم، جریان گرادینان را در طول فرآیند آموزش با کاهش مشکل محو شدن گرادینان، پایدار ساخته و آموزش شبکه‌های بسیار عمیق را ممکن می‌سازد. پس از استخراج ویژگی‌ها، از یک لایه ادغام سراسری میانگین برای فشرده‌سازی نقشه‌های ویژگی به یک بردار ویژگی یک‌بعدی استفاده می‌شود. این بردار سپس به یک لایه تمام‌متصل جدید تغذیه می‌گردد که برای تطبیق با تعداد کلاس‌های مسئله ما یعنی سه پلتفرم بدنه خودرو X200، SP100 و X200 بهینه بازطراحی شده است. در نهایت، یک لایه طبقه‌بندی با تابع فعال‌ساز Softmax، احتمال تعلق تصویر ورودی به هر یک از کلاس‌ها را محاسبه کرده و کلاس نهایی با بالاترین میزان اطمینان را به عنوان خروجی پیش‌بینی می‌کند.



۳- نتایج و بحث

در این تحقیق، کلیه مراحل آموزش و ارزیابی مدل‌ها با استفاده از نرم‌افزار متلب نسخه ۲۳/۲۰b و بر روی سامانه دارای سخت‌افزار شامل پردازنده Intel® Core™ i5-13400 (13th Gen) با فرکانس پایه ۲/۵۰ GHz و حافظه رم ۳۲ گیگابایت انجام شده است. فرآیند آموزش به‌طور کامل بر پایه CPU و در محیط سیستم عامل Windows 64-bit اجرا شده و از شتاب‌دهنده

همگرایی مناسب نرسیده بود. در عین حال، افزایش تعداد epochs به بیش از ۱۰ نیز بهبود قابل توجهی در عملکرد مدل ایجاد نکرد و با توجه به پیچیدگی محاسباتی و زمان آموزش، مقدار ۱۰ به عنوان انتخاب بهینه تعیین شد، چراکه هم از لحاظ دقت خروجی قابل قبول بود و هم از لحاظ بهره‌وری محاسباتی. بنابراین، این مقدار با هدف تعادل میان دقت و زمان آموزش انتخاب گردید. به منظور ارزیابی عملکرد مدل در فرآیند طبقه‌بندی، از معیار دقت و اطمینان استفاده گردید. معیار دقت، درصد نمونه‌هایی را که مدل به درستی طبقه‌بندی کرده است، نسبت به کل نمونه‌ها محاسبه می‌نماید و به صورت زیر تعریف می‌شود:

$$\text{معیار اطمینان} = \frac{\text{تعداد پیش‌بینی‌های صحیح}}{\text{کل پیش‌بینی‌ها}} \times 100 = \text{دقت}$$

معیار اطمینان به صورت زیر محاسبه می‌شود:

$$\text{اطمینان} = \max(\text{score}) * 100$$

Score یک بردار از احتمالات است که توسط شبکه عصبی پیش‌بینی می‌شود. هر عضو این بردار نشان‌دهنده احتمال تعلق تصویر به یکی از کلاس‌ها است.

لایه‌ی طبقه‌بندی نهایی شبکه که به‌طور پیش‌فرض برای ۱۰۰۰ کلاس تعریف شده بود (مطابق با ساختار ImageNet)، با یک لایه طبقه‌بندی جدید جایگزین گردید. این لایه جدید در ساختار گراف لایه‌ای شبکه اعمال شد. اگر این لایه متناسب با مسئله‌ی جدید طراحی نشود، شبکه قادر به یادگیری مناسب روی کلاس‌های جدید نخواهد بود و فرآیند تنظیم دقیق^۱ به درستی انجام نخواهد شد. زمانی که این لایه با پارامترهای پیش‌فرض در ساختار شبکه قرار می‌گیرد، به‌صورت پیش‌فرض از تابع آنتروپی متقاطع^۲ برای محاسبه خطا استفاده می‌شود. این تابع با اندازه‌گیری اختلاف بین بردار پیش‌بینی شده و بردار برچسب واقعی به صورت one-hot، مقدار loss را برای هر نمونه محاسبه می‌کند. ساختار کلی این تابع به صورت زیر تعریف می‌شود:

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i)$$

که در آن:

L: مقدار تابع زیان

C: تعداد کلاس‌ها

y_i : مقدار واقعی برای کلاس i به صورت one-hot

\hat{y}_i : احتمال پیش‌بینی شده توسط مدل برای کلاس i

شکل ۸ بلوک دیاگرام مدل با تغییرات اعمال شده را نشان می‌دهد. شکل ۸ معماری شبکه عصبی عمیق پیشنهادی مبتنی بر ResNet-50 را نشان می‌دهد. در گام نخست، تصویر ورودی با ابعاد 224×224 پیکسل وارد شبکه می‌شود. پردازش با یک لایه کانولوشنی با ۶۴ فیلتر آغاز می‌گردد که به دنبال آن یک لایه

² Cross Entropy Loss

¹ Fine-tuning

نیز در صورت فراهم بودن شرایط مناسب تصویربرداری، توانایی بالایی در تفکیک دقیق کلاس‌های بدنه را داراست. در مراحل بعدی پیاده‌سازی، با ملاحظات ایمنی ناشی از وجود مواد قابل اشتعال در فضای داخل کابین پاشش رنگ، استقرار دوربین در داخل کابین امکان‌پذیر نبود. از سوی دیگر، مسیر حرکتی بدنه‌ها بر روی کانوایر به‌گونه‌ای بود که در صورت نصب دوربین درون کابین، برخورد فیزیکی با خودرو اجتناب‌ناپذیر می‌بود. به همین دلیل، دوربین تحت شبکه با رزولوشن 1280×720 در موقعیت ثابت و خارج از کابین نصب گردید؛ به‌نحوی که زاویه دید آن به‌صورت مایل نسبت به سطح جلویی بدنه خودرو تنظیم شد. در این حالت، تصاویر دریافتی شامل پس‌زمینه‌های متغیر و شرایط نوری ناپایدار بودند که موجب پیچیدگی قابل توجهی در ویژگی‌های تصویری ورودی شد.

ارزیابی مدل ResNet-18 در شرایط جدید نشان داد که دقت آن در تشخیص نوع بدنه کاهش یافته و مدل در برخی موارد دچار اشتباه در طبقه‌بندی گردید در حالی که سیستم امکان تحمل خطا در تشخیص را ندارد و خطا ممکن است باعث برخورد ربات‌های پاشش رنگ به بدنه خودرو گردد. این افت عملکرد را می‌توان ناشی از افزایش نویز پس‌زمینه، تغییر زاویه دید و ویژگی‌های نوری ناهمگن دانست که مدل اولیه، با ساختار کم‌عمق خود، قادر به استخراج مؤثر ویژگی‌های تمایزبخش در آن شرایط نبود. این نتایج ضرورت بهره‌گیری از معماری‌های عمیق‌تر با ظرفیت یادگیری بالاتر، مانند ResNet-50، را در محیط‌های صنعتی متغیر آشکار ساختند.

جدول ۲ نمونه تست یک پوشه شامل ۲۲ تصویر بدنه خودرو SP100 مستقل از داده‌های آموزش را نشان می‌دهد که سیستم در آن دچار خطا شده است و دو نمونه را به اشتباه به عنوان بدنه خودرو X200 بهینه تشخیص داده است. مطابق شکل ۱۲ میزان اطمینان مدل به پیش‌بینی‌ها نسبت به زمانی که تصاویر از داخل کابین به صورت روبرو و با پس‌زمینه ساده بودند بشدت تغییر کرده است. این مشاهده نشان می‌دهد که چالش اصلی در این مسئله، صرفاً کمبود داده نیست، بلکه ماهیت و درجه دشواری ذاتی هر کلاس برای تشخیص است. به نظر می‌رسد ویژگی‌های بصری کلاس SP100 (مانند تفاوت‌های بسیار ظریف با سایر پلتفرم‌ها) حتی با وجود داده بیشتر، برای یک مدل کم‌ظرفیت‌تر (ResNet18) چالش‌برانگیزتر بوده است. این نتیجه بر برتری ذاتی معماری‌های عمیق‌تر مانند ResNet50 در استخراج و یادگیری این ویژگی‌های پیچیده و ظریف، حتی در حضور داده نسبتاً کمتر، تأکید می‌کند.

جدول ۲: تست یک پوشه شامل ۲۲ تصویر SP100

نوع بدنه	X200 بهینه	SP100	X200
تعداد تشخیص	۲	۲۰	۰
اطمینان	۵۱/۲۴ - ۵۶/۳۷	۵۲/۹۹ - ۱۰۰	-

گرافیکی (GPU) استفاده نگردیده است. زمان‌های گزارش شده برای آموزش مدل‌ها، آزمون دسته‌ای تصاویر، و تحلیل‌های عملکردی همگی در همین بستر محاسباتی اندازه‌گیری شده‌اند.

با هدف تعیین مقدار بهینه برای پارامتر MaxEpochs به منظور آموزش شبکه ResNet18، سه مقدار مختلف شامل ۲، ۱۰ و ۵۰ مورد بررسی قرار گرفتند. که به ترتیب در شکل‌های ۹ تا ۱۱ نشان داده شده‌اند. نتایج نشان دادند که در مقدار ۲، مدل ResNet تنها طی ۸ ثانیه آموزش، به دقت ۹۰٪ بر روی مجموعه اعتبارسنجی رسید؛ با این حال منحنی همگرایی آن کامل نبود. در مقدار ۱۰ که زمان آموزش برابر با ۳۴ ثانیه بود، دقت مدل به ۱۰۰٪ افزایش یافت و مقادیر اطمینان نیز به‌طور کامل تثبیت شدند. در مقدار ۵۰، اگرچه دقت مدل همچنان در سطح ۱۰۰٪ باقی ماند، اما زمان آموزش به ۱۶۹ ثانیه افزایش یافت و در برخی نمونه‌ها نوسان جزئی در مقادیر loss مشاهده گردید که می‌تواند نشانه شروع بیش‌برازش باشد. بنابراین، مقدار ۱۰ برای epoch به عنوان نقطه تعادل بین دقت، پایداری و کارایی محاسباتی انتخاب شد. این تحلیل نشان می‌دهد که افزایش بیش‌ازحد تعداد epochs نه تنها بهبود قابل توجهی در عملکرد مدل ایجاد نمی‌کند، بلکه می‌تواند موجب اتلاف منابع و کاهش تعمیم‌پذیری مدل شود.

در راستای سنجش عملکرد مدل در محیط‌های واقعی و مستقل از داده‌های آموزشی، قابلیت پردازش دسته‌ای به نرم‌افزار افزوده شد تا تصاویر موجود در یک پوشه به‌صورت خودکار شناسایی و طبقه‌بندی شوند. در این آزمون، مجموعه‌ای شامل ۳۹ تصویر از بدنه خودروی X200 و ۳۲ تصویر از بدنه خودروی SP100 که در شرایط یکنواخت نوری و زاویه ثابت، در داخل کابین پاشش رنگ و با دوربین اندرویدی ثبت شده بودند، به صورت فایل به سامانه معرفی گردیدند.

نکته حائز اهمیت، عملکرد قوی سامانه پیشنهادی در شرایط واقعی و پرچالش صنعتی است که در آزمون‌های خارج از کابین (با نور طبیعی و پس‌زمینه متغیر) مورد ارزیابی قرار گرفت. همانطور که در نمودارهای هیستوگرام اطمینان (شکل‌های ۱۵، ۱۶ و ۱۷) مشهود است، مدل ResNet-50 در مواجهه با این داده‌های پیچیده، در اکثر قریب به اتفاق موارد با سطح اطمینانی بالاتر از ۸۰٪ اقدام به طبقه‌بندی صحیح نمود. این امر به وضوح نشان‌دهنده قابلیت اطمینان بالای معماری یادگیری عمیق در استخراج ویژگی‌های معنادار حتی در حضور نویز، تغییرات نوری و زاویه‌ای است؛ ویژگی‌ای که روش‌های مبتنی بر قانون یا مدل‌های کلاسیک فاقد آن هستند و همین امر انتخاب رویکرد پیشنهادی را برای کاربرد در خط تولید توجیه می‌کند.

مدل آموزش دیده‌ی ResNet-18 موفق شد تمامی نمونه‌ها را به‌درستی طبقه‌بندی کند که معادل دقت ۱۰۰٪ در این ارزیابی است. مقادیر اطمینان خروجی در کلاس X200 بین ۹۹/۶۹٪ تا ۱۰۰٪ و در کلاس SP100 بین ۹۸/۴۵٪ تا ۱۰۰٪ گزارش شد. این نتایج نشان می‌دهند که حتی مدل نسبتاً کم‌عمق ResNet-18

در جدول ۳ مقایسه عملکرد بین دو شبکه عصبی Resnet18 و Resnet50 برای تشخیص نوع خودرو در سه کلاس تعریف شده را می توان مشاهده کرد. به منظور ارزیابی کارایی زمانی سامانه در شرایط عملیاتی، زمان اجرای تابع طبقه بندی برای اجرای تک تصویر در محیط رابط کاربری گرافیکی اندازه گیری شد. در این آزمایش، برای هر یک از کلاس ها، به ترتیب ۳۶ تصویر از X200، ۳۲ تصویر از SP100 و ۲۴ تصویر X200 بهینه مستقل از مجموعه داده آموزشی انتخاب گردید و فرآیند طبقه بندی به صورت جداگانه برای هر تصویر اجرا شد. مدت زمان صرف شده برای اجرای طبقه بندی در هر بار فراخوانی تابع محاسبه گردید. میانگین زمان ها و بازه مقادیر مشاهده شده در جدول ۴ ارائه شده است:

جدول ۴: میزان زمان تشخیص و طبقه بندی نمونه

بازه زمانی مشاهده شده (ثانیه)	میانگین زمان اجرای classify (ثانیه)	تعداد نمونه آزمایشی	کلاس بدنه خودرو
۰/۱۶ - ۰/۲۹	۰/۲۱	۳۲	SP100
۰/۱ - ۰/۲۶	۰/۲۲	۲۴	X200 بهینه
۰/۱۶ - ۰/۲۴	۰/۲	۳۶	X200

با توجه به نتایج حاصل از ارزیابی زمان اجرای مدل در محیط نرم افزاری، میانگین زمان inference برای هر تصویر در حدود ۲۰۰ میلی ثانیه برآورد شده است. با لحاظ این مقدار و در نظر گرفتن سرعت خط تولید خودرو (کانوایر) حداکثر ۴/۶ متر بر دقیقه در کابین پاشش، تشخیص نوع بدنه با اطمینان کامل پیش از ورود به بخش رباتیک سامانه پاشش رنگ انجام شده و فرصت کافی برای تنظیم پارامترهای ربات فراهم می شود. این ویژگی حاکی از قابلیت عملیاتی سامانه در شرایط واقعی کارخانه بوده و مشخص می کند که مدل پیشنهادی از نظر زمان پاسخگویی کاملاً با نیازهای زمان بندی خط رنگ آمیزی خودرو سازگار است.

۴- نتیجه گیری

در این مطالعه، عملکرد دو مدل ResNet18 و ResNet50 در تشخیص خودروهای X200، SP100 و X200 بهینه با استفاده از پردازش تصویر مورد بررسی قرار گرفت. نتایج نشان داد که ResNet50 در تمامی کلاس ها عملکرد بهتری نسبت به ResNet18 دارد. به طور خاص، میانگین اطمینان برای ResNet50 در تشخیص X200، SP100 و X200 بهینه به ترتیب ۹۹/۵۶٪، ۹۸/۴۵٪ و ۹۹/۸۵٪ بود، در حالی که این مقادیر برای ResNet18 به ترتیب ۹۷/۴۵٪، ۷۶/۴۵٪ و ۹۹/۸۷٪ گزارش شد. علاوه بر این، ResNet50 در تمامی کلاس ها به دقت ۱۰۰٪ دست یافت، در حالی که ResNet18 در تشخیص SP100 و X200 بهینه به ترتیب دقت ۸۵/۷۱٪ و ۹۵/۸۳٪ را نشان داد. این بهبود عملکرد ResNet50 به دلیل عمق بیشتر شبکه و توانایی بهتر آن در یادگیری ویژگی های پیچیده است. همچنین، ResNet50 توانست خطاهای تشخیص ResNet18 را به طور کامل برطرف کند و هیچ تشخیص اشتباهی در هیچ یک از کلاس ها گزارش نشد. با توجه به این نتایج،

مقدار اطمینان مدل که در شرایط کنترل شده در نزدیکی ۱۰۰٪ بود، اکنون در بازه های بین ۵۵٪ تا ۱۰۰٪ توزیع شد. این کاهش اطمینان و افزایش نرخ خطا در خروجی مدل، به وضوح ضعف ResNet-18 را در استخراج ویژگی ها با مقاومت در برابر تغییرات محیطی را نمایان ساخت. با توجه به کاهش دقت و اطمینان مدل ResNet-18 در شرایط واقعی محیط صنعتی، ساختار شبکه به معماری عمیق تر ResNet-50 ارتقاء یافت. این شبکه با ۵۰ لایه و بیش از ۲۵ میلیون پارامتر، ظرفیت بالاتری برای یادگیری ویژگی های پیچیده تر در داده های تصویری دارد. فرآیند آموزش مدل ResNet-50 با استفاده از همان مجموعه داده و پارامترهای مشابه نسخه ResNet-18 انجام شد (شامل نرخ یادگیری، تعداد epoch، و اندازه batch) با وجود عدم تغییر در این پارامترها، زمان آموزش مدل از ۳۴ ثانیه در نسخه ResNet-18 به ۵۲ ثانیه در نسخه ResNet-50 افزایش یافت. افزایش زمان آموزش ناشی از پیچیدگی بیشتر مدل و افزایش تعداد پارامترها بود، اما در مقابل، دقت مدل و اطمینان آن در شرایط غیرایزوله صنعتی به طور محسوسه بهبود یافت. این تحلیل نشان می دهد که با وجود افزایش نسبی هزینه محاسباتی، استفاده از معماری های عمیق تر در مسائل صنعتی که با نویز و تغییرات محیطی مواجه اند، مزیت قابل توجهی در پایداری و دقت عملکرد نهایی سامانه ایجاد می کنند. فرآیند آموزش شبکه و زمان سپری شده برای این کار در شکل ۱۱ دیده می شود.

برای بررسی کیفیت پیش بینی های مدل ResNet-50 در شرایط صنعتی، توزیع مقادیر اطمینان مربوط به خروجی طبقه بندی برای هر یک از سه کلاس بدنه شامل X200، SP100 و X200 بهینه مورد تحلیل قرار گرفت. در این ارزیابی، به ترتیب ۳۶ تصویر از X200، ۳۲ تصویر از SP100 و ۲۴ تصویر از X200 بهینه توسط سامانه پردازش و طبقه بندی شدند. هیستوگرام های حاصل در شکل های ۱۵ تا ۱۷ نشان می دهند که مقادیر اطمینان در تمامی کلاس ها عمدتاً در بازه بالای ۸۵٪ تا ۱۰۰٪ قرار داشته اند، و توزیع این مقادیر به وضوح به سمت مقادیر نزدیک به ۱۰۰٪ گرایش دارد. این تمرکز بالا در ناحیه اطمینان کامل، نه تنها بیانگر دقت بالا در طبقه بندی نمونه هاست، بلکه حاکی از پایداری عملکرد مدل ResNet-50 در شرایط تصویربرداری صنعتی (با زاویه دید مایل، پس زمینه متغیر، و نور نامنظم) نیز می باشد. به ویژه، عدم حضور مقادیر پایین اطمینان نشان می دهد که حتی در موارد دشوارتر نیز مدل با سطح اطمینان قابل قبولی تصمیم گیری کرده است. این در حالی است که سامانه نصب شده توسط شرکت سمز فرانسه که از سنسور های پرده نوری برای سنسجش طول بدنه در مقاطع مختلف عرضی خودرو، مطابق شکل ۱۴ برای تشخیص نوع بدنه خودرو بهره می گیرد. توانایی تشخیص تفاوت بین بدنه X200 و X200 بهینه که دارای نمای جانبی یکسان و نمای روبروی متفاوت هستند را ندارد.

Residual-Inception Networks for Real-Time Object Detection. 2017 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2017.

[11] Wang, H., Peng, J., Zhao, Y., Fu, X. *Multi-Path Deep CNNs for Fine-Grained Car Recognition*. IEEE Transactions on Vehicular Technology, 69(10), 10484–10493, 2020.

[12] Huttunen, H., Yancheshmeh, F. S., Chen, K. *Car Type Recognition with Deep Neural Networks*. 2016 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2016.

[13] Nazemi, A., Shafiee, M. J., Azimifar, Z., Wong, A. *Unsupervised Feature Learning Toward a Real-Time Vehicle Make and Model Recognition*. arXiv preprint arXiv:1806.03028, 2018.

[14] Schulzrinne, H., Rao, A., Lanphier, R., Westerlund, M., Stiemerling, M. *Real-Time Streaming Protocol (RTSP)*. RFC 2326, April 1998. [Online]. Available: <https://tools.ietf.org/html/rfc2326>.

[15] FFmpeg Documentation. *FFmpeg, 2025*. [Online]. Available: <https://ffmpeg.org/ffmpeg-all.html>.

[16] MathWorks. *Training Options Documentation*. 2019. [Online]. Available: <https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html>. Accessed: May 14, 2025.



میرمسعود حسینی کوکمری مدرک کارشناسی و کارشناسی ارشد خود را به ترتیب در رشته مهندسی مهندسی برق از دانشگاه صنعتی مالک اشتر و دانشگاه علم و صنعت ایران در سال‌های ۱۳۹۰ و ۱۳۹۳ دریافت نمود. وی تاکنون در شرکت سایپا به عنوان مهندس ارشد برق و اتوماسیون مشغول به فعالیت بوده و در زمینه طراحی سخت‌افزار،

سیستم‌های کنترل صنعتی و هوشمندسازی خطوط رنگ و رباتیک تجربه دارد زمینه‌های پژوهشی مورد علاقه ایشان شامل بینایی ماشین، یادگیری عمیق، و سامانه‌های هوشمند مبتنی بر پردازش تصویر می‌باشد.



رضا آقائیان زاده ظروفی مدرک دکتری خود را در زمینه آنالیز تصاویر پزشکی از دانشگاه ازاکای ژاپن در سال ۱۳۷۵ دریافت نمود. وی همچنین از سال ۱۳۷۵ لغایت ۱۳۷۸ به‌عنوان محقق دکتری و فوق دکتری در وزارت صنایع ژاپن، مرکز تحقیقات قلب و عروق ژاپن و دانشکده پزشکی از دانشگاه ازاکای ژاپن همکاری نمود. وی به عنوان

استادیار در اسفند سال ۱۳۷۸ به دانشگاه تهران پیوست و در حال حاضر استاد دانشکده مهندسی برق و کامپیوتر این دانشگاه می‌باشد. او همچنین از سال ۱۳۷۸ تاکنون به عنوان محقق و استاد مدعو با دانشگاه های ژاپن همکاری علمی مستمر داشته است. زمینه کاری تخصصی ایشان پردازش، مهندسی و مدیریت جامع تصاویر پزشکی می‌باشد.

استفاده از مدل ResNet50 به‌عنوان مدل پیشنهادی برای تشخیص خودروهای X200 و SP100، بهینه توصیه می‌شود. در آینده، می‌توان با افزایش حجم داده‌های آموزشی و استفاده از تکنیک‌های پیشرفته‌تر یا مدل‌های ترکیبی، عملکرد مدل را حتی بیشتر بهبود بخشید.

۵- سپاسگزاری

نویسندگان مقاله بدین‌وسیله مراتب قدردانی و سپاس خود را از آقای رضا ارش، مدیر نگهداری و تعمیرات شرکت سایپا و آقای پیمان حسینی، رئیس اداره نگهداری و تعمیرات سالن رنگ ۲ شرکت سایپا، به‌منظور حمایت‌ها و همکاری‌های صمیمانه ایشان در فرآیند اجرایی این پژوهش ابراز می‌دارند. راهنمایی‌ها و تسهیلات فراهم‌شده از سوی ایشان نقش مؤثری در پیشبرد مراحل عملیاتی پروژه داشته است.

مراجع

[1] Pendar, M.-R., Rodrigues, F., Páscoa, J. C., Lima, R. *Review of Coating and Curing Processes: Evaluation in Automotive Industry*. Physics of Fluids, 34(10), 2022.

[2] Culda, L. I. *RFID Technology in Production and Post Production for the Automotive Industry*. Robotica & Management, 28(1), 2023.

[3] Kuş, A. *Implementation of 3D Optical Scanning Technology for Automotive Applications*. Sensors, 9(3), 1967–1979, 2009.

[4] O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., et al. *Deep Learning vs. Traditional Computer Vision*. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*, Volume 1, Springer, 2020.

[5] Jung, H., Rhee, J. *Application of YOLO and ResNet in Heat Staking Process Inspection*. Sustainability, 14(23), 15892, 2022.

[6] Li, Z., Tian, X., Liu, X., Liu, Y., Shi, X. *A Two-Stage Industrial Defect Detection Framework Based on Improved-YOLOv5 and Optimized-Inception-ResNetv2 Models*. Applied Sciences, 12(2), 834, 2022.

[7] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. *YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications*. arXiv preprint arXiv:2209.02976, 2022.

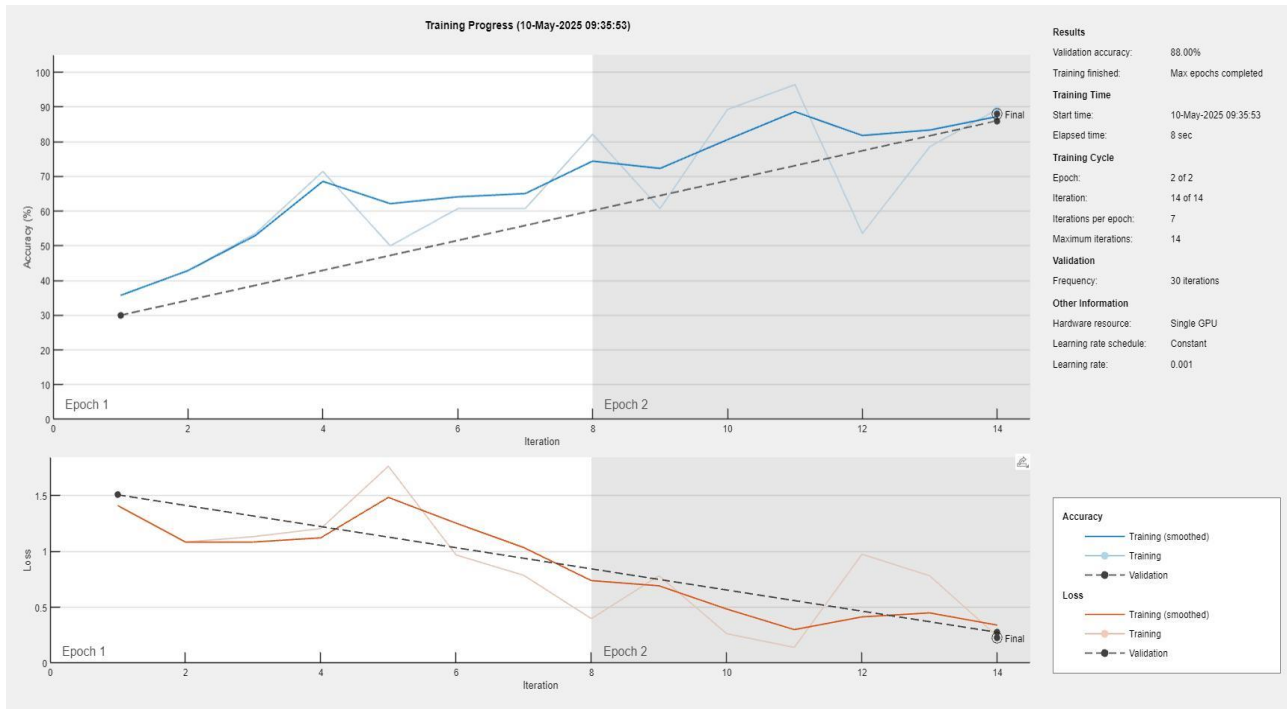
[8] Puttemans, S., Callemeyn, T., Goedemé, T. *Building Robust Industrial Applicable Object Detection Models Using Transfer Learning and Single Pass Deep Learning Architectures*. arXiv preprint arXiv:2007.04666, 2020.

[9] Chen, Z., He, Z., Chao, B., Guo, H. *Visual Detection Application of Lightweight Convolution and Deep Residual Networks in Wood Production*. Wireless Communications and Mobile Computing, 2022(1), 9465433, 2022.

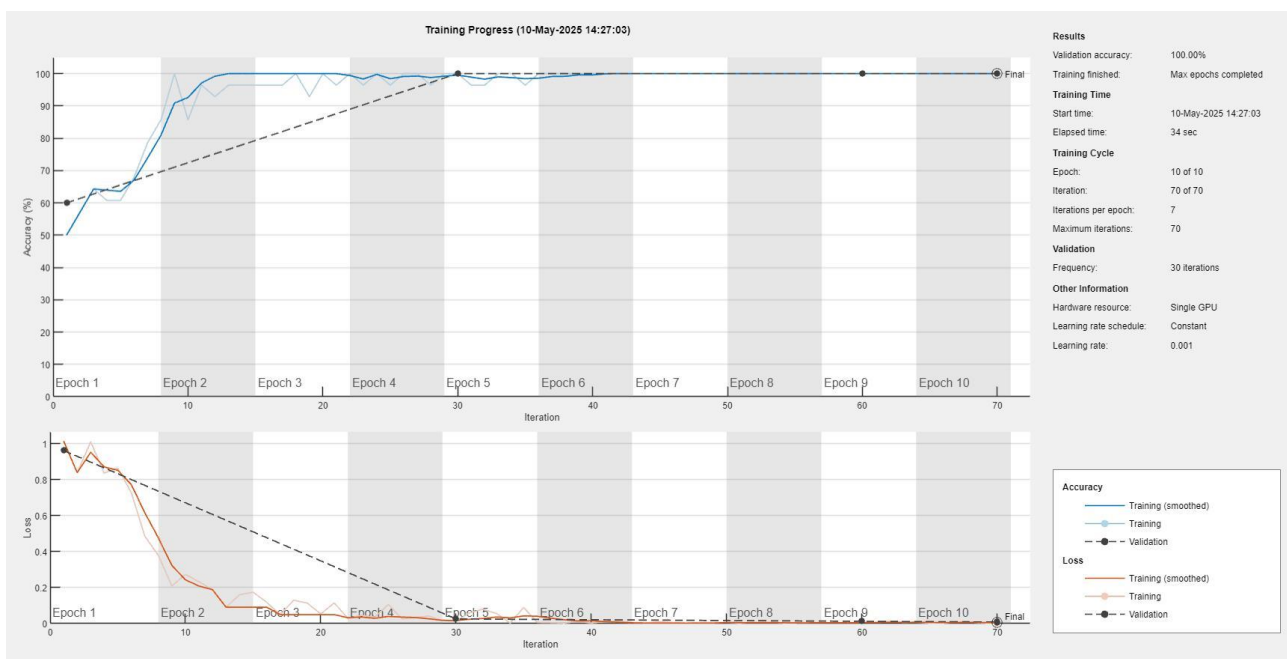
[10] Lee, Y., Kim, H., Park, E., Cui, X., Kim, H. *Wide-*

جدول ۳: مقایسه عملکرد شبکه ResNet50 و ResNet18

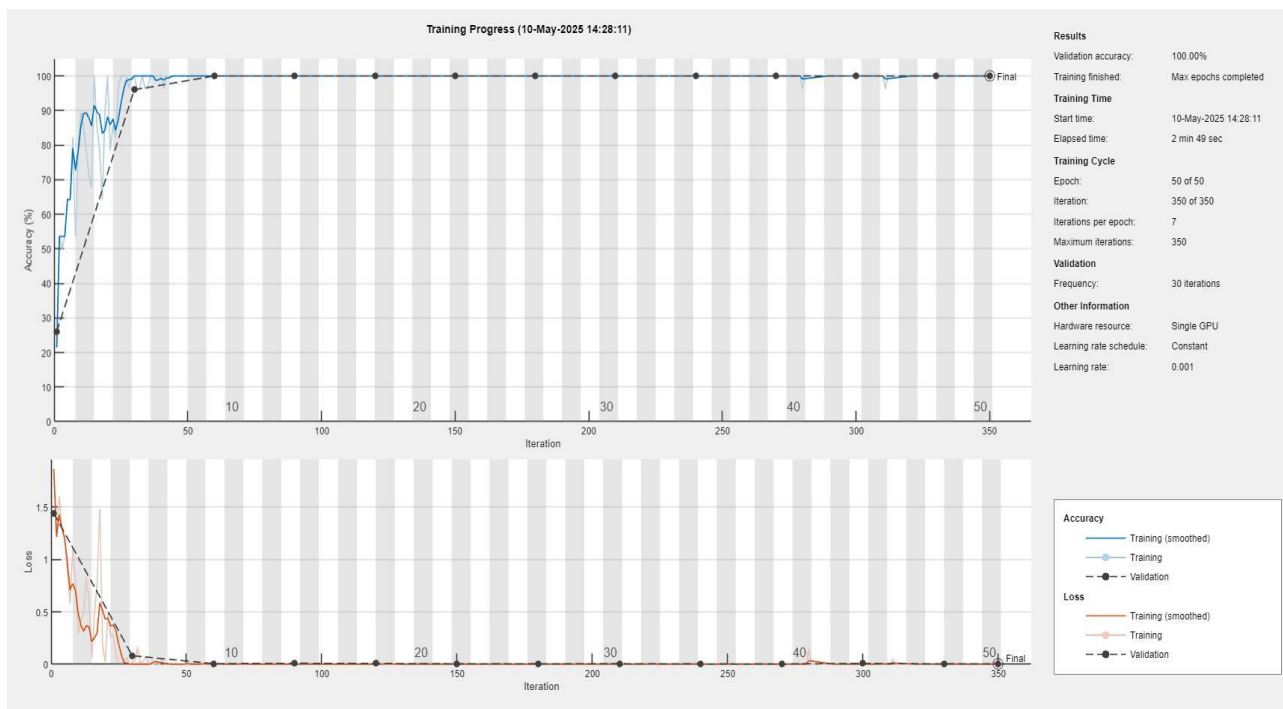
نوع بدنه	مدل	تعداد نمونه تست	حداقل درصد اطمینان	حداکثر درصد اطمینان	میانگین درصد اطمینان	تعداد تصاویر با اطمینان زیر ۹۰٪	تعداد تشخیص اشتباه	دقت مدل (%)
X200	ResNet18	۳۶	۹۳/۷	۱۰۰	۹۹/۸	۰	۰	۱۰۰
X200	ResNet50	۳۶	۷۰/۱	۱۰۰	۹۹/۵	۱	۰	۱۰۰
SP100	ResNet18	۳۲	۵۱/۲	۹۹/۱	۷۶/۴	۱۲	۲	۸۵/۷
SP100	ResNet50	۳۲	۸۵/۹	۹۹/۹	۹۸/۴	۱	۰	۱۰۰
X200 بهینه	ResNet18	۲۴	۵۶/۷	۹۹/۹	۹۷/۴	۲	۱	۹۵/۸
X200 بهینه	ResNet50	۲۴	۸۱/۸	۱۰۰	۹۹/۸	۱	۰	۱۰۰



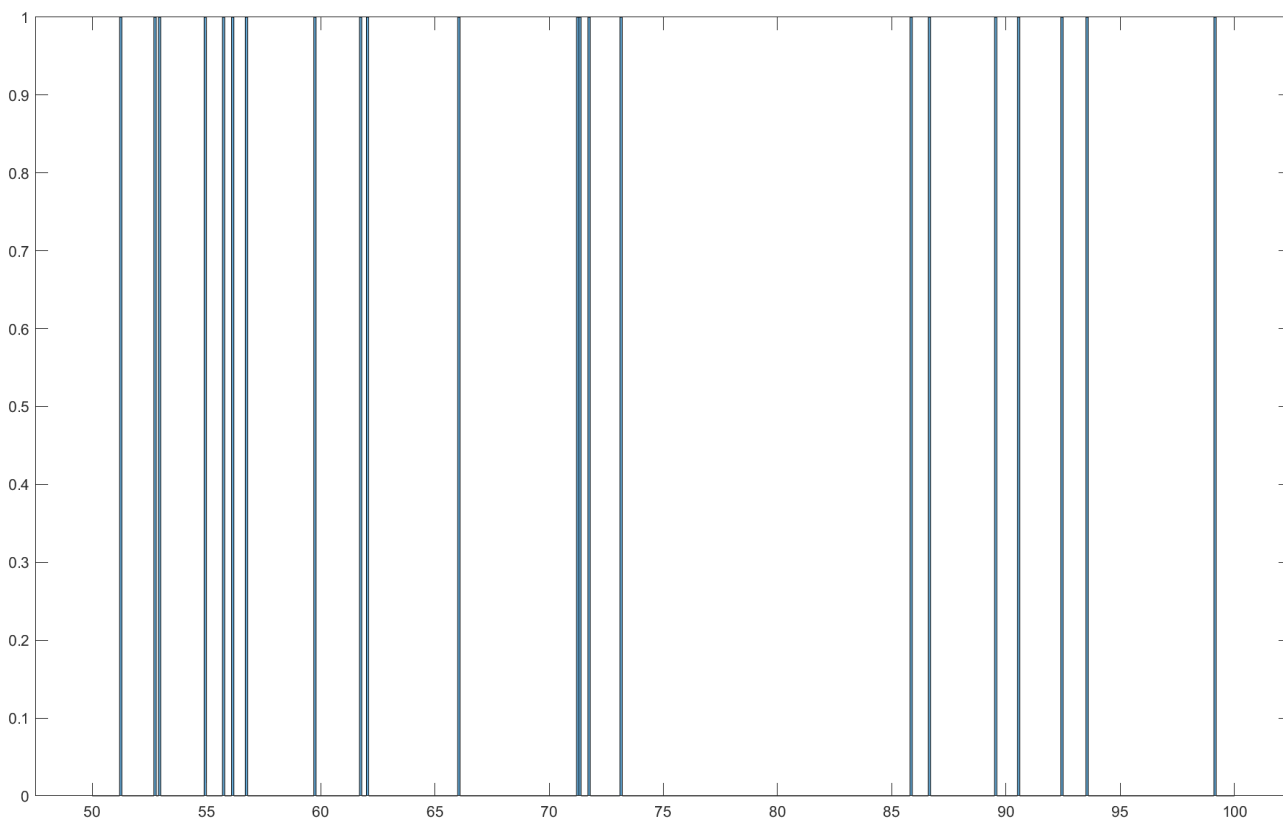
شکل ۹: آموزش شبکه ResNet18 با Maxepoch=2



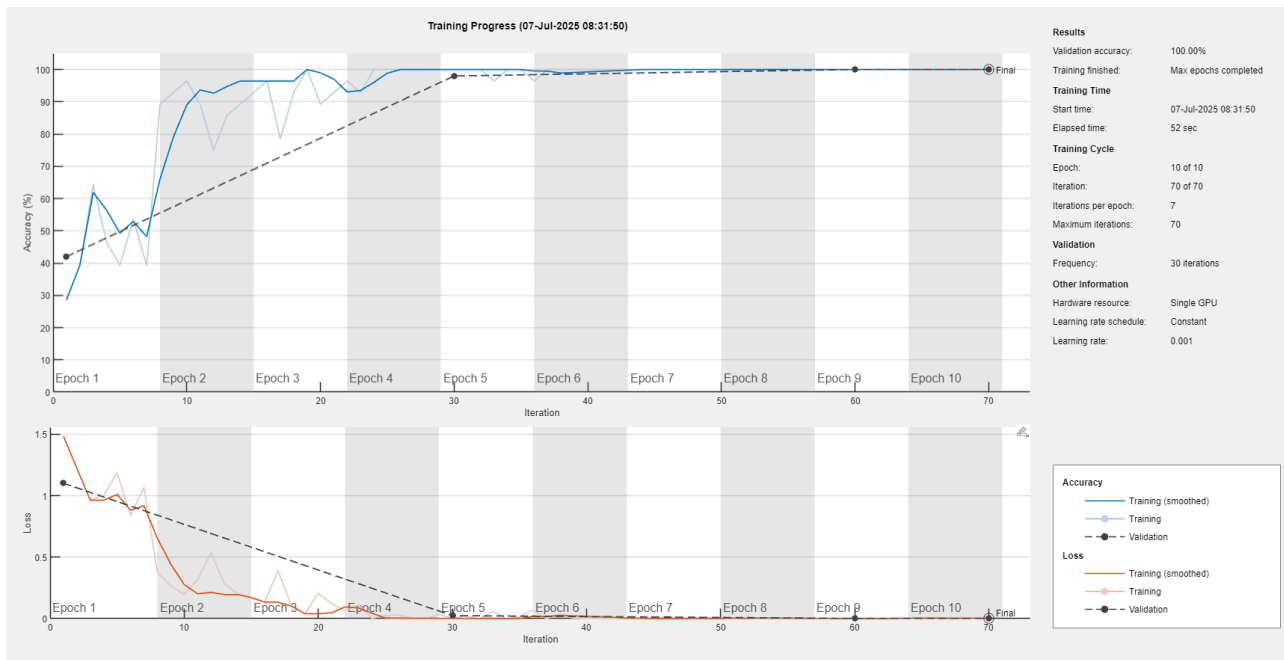
شکل ۱۰: آموزش شبکه ResNet18 با Maxepoch=10



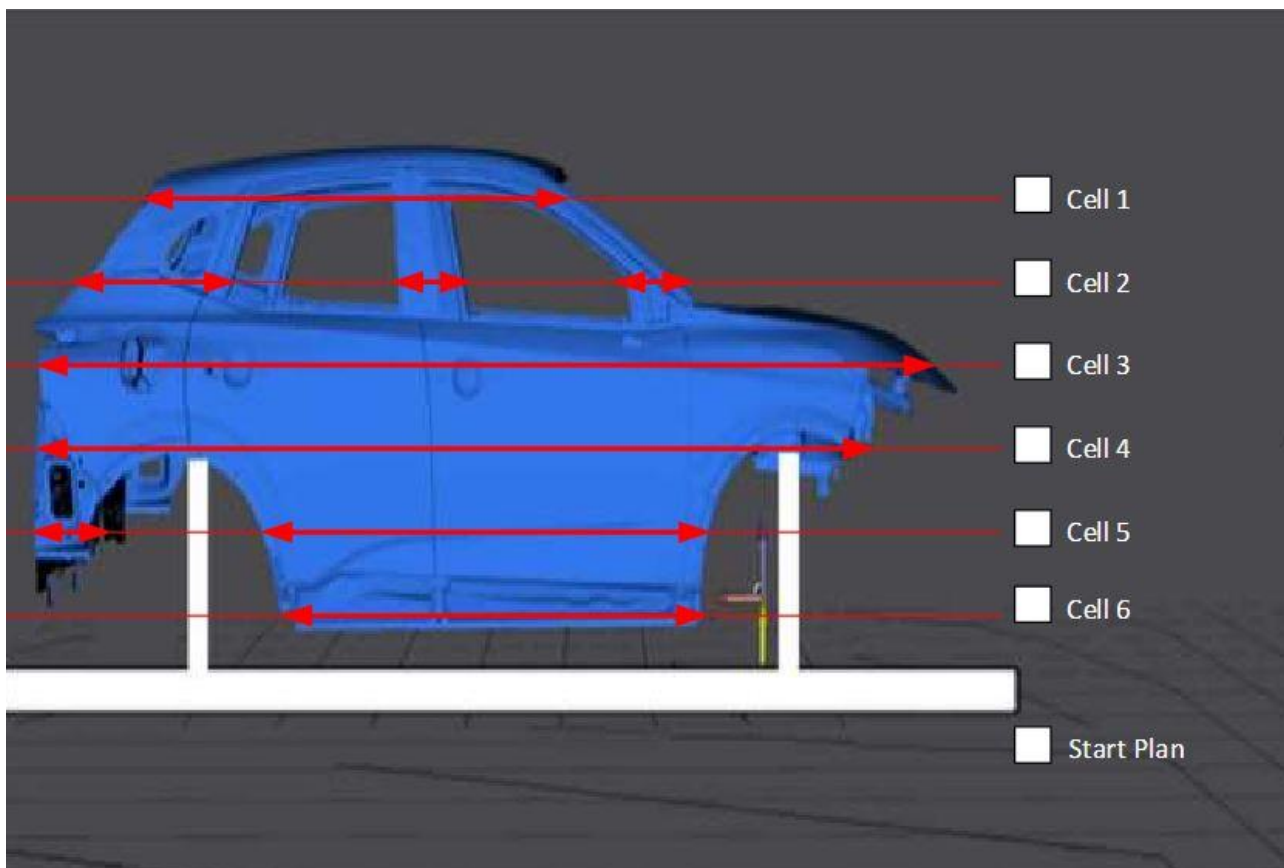
شکل ۱۱: آموزش شبکه ResNet18 با Maxepoch=50



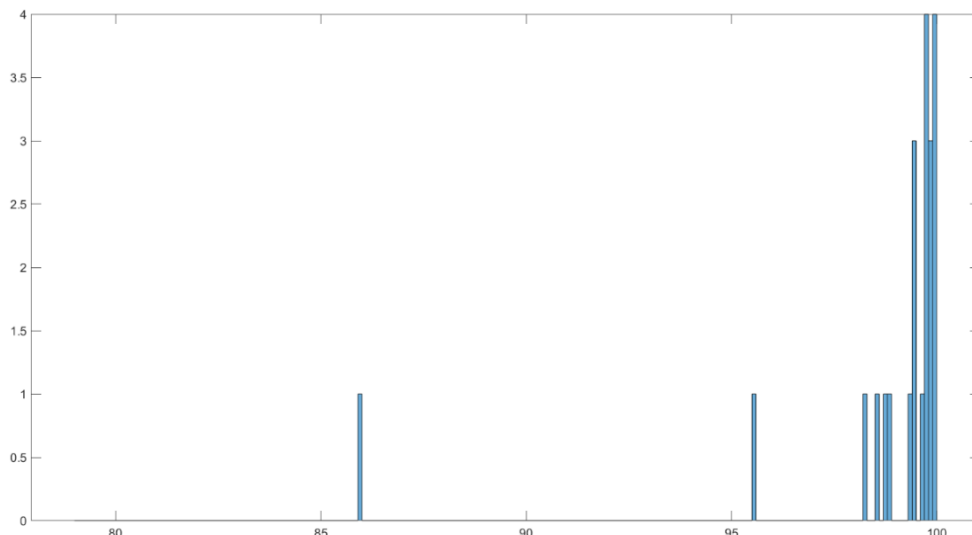
شکل ۱۲: هیستوگرام اطمینان برای تست تصاویر SP100 خارج کابین پاشش رنگ با مدل resnet18



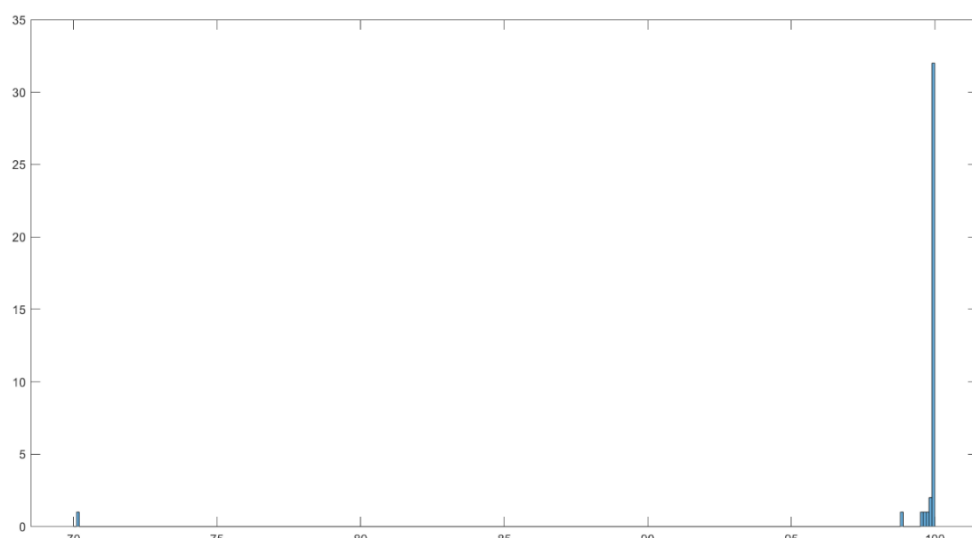
شکل ۱۳: آموزش شبکه ResNet50



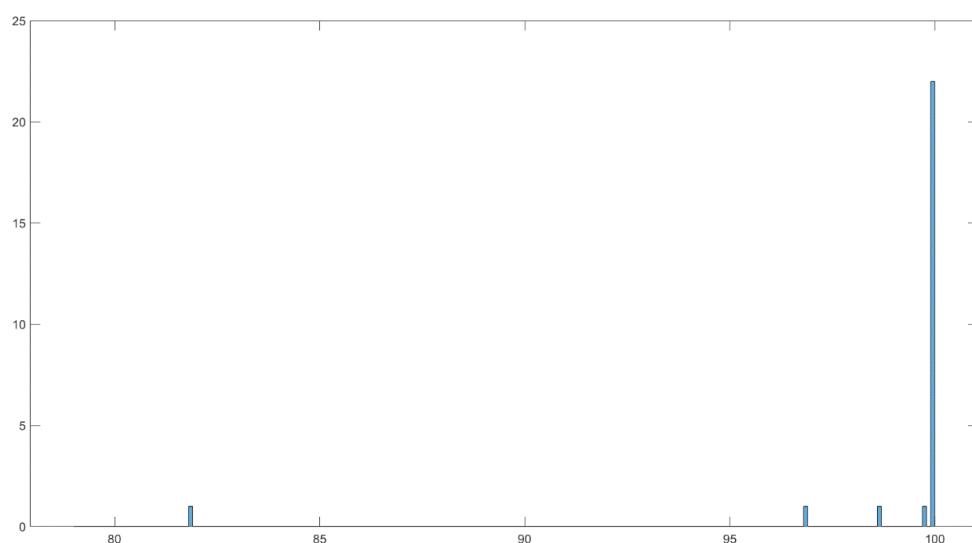
شکل ۱۴: مقاطع قابل اندازه گیری سیستم تشخیص بدنه شرکت sames



شکل ۱۵: هیستوگرام اطمینان برای تست تصاویر SP100 خارج کابین پاشش رنگ با مدل resnet50



شکل ۱۶: هیستوگرام اطمینان برای تست تصاویر X200 خارج کابین پاشش رنگ با مدل resnet50



شکل ۱۷: هیستوگرام اطمینان برای تست تصاویر X200 بهینه خارج کابین پاشش رنگ با مدل resnet50

