

## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش های مبتنی بر

### یادگیری عمیق

سیمین کولایان، دانش آموخته کارشناسی ارشد رشته ی کنترل و علائم، دانشکده راه آهن، دانشگاه علم و صنعت ایران، تهران، ایران  
محمدعلی صنیدزاده (مسئول مکاتبات)، دانشیار دانشکده راه آهن، دانشگاه علم و صنعت ایران، تهران، ایران

**E-mail: sandidzadeh@iust.ac.ir**

پذیرش: ۱۴۰۲/۰۸/۲۵

دریافت: ۱۴۰۲/۰۴/۲۱

### چکیده

راه آهن به عنوان یک شریان ارتباطی موثر جایگاه ویژه ای در رشد و توسعه اقتصادی کشورها دارد. از همین رو افزایش دقت، قابلیت اطمینان و بهینه سازی ظرفیت صنعت حمل و نقل ریلی از اهمیت ویژه ای برخوردار است. بطوری که امروزه صنعت حمل و نقل ریلی در حال حرکت بسمت خودران سازی است. تشخیص موانع ریلی یک بهبود عملیاتی و ایمنی در راه آهن است که حمل و نقل ریلی را کارآمدتر، دقیق تر و قابل اعتمادتر می سازد. امروزه پیشرفت در فناوری های هوش مصنوعی و حسگرها باعث توسعه روش های تشخیص موانع و خودران سازی شده است. در این مقاله با بهره گیری از روش های مبتنی بر هوش مصنوعی و یادگیری عمیق به طراحی ساختاری پرداخته شده است که تشخیص زمان حقیقی موانع ریلی را با توازن بین دقت و بار محاسباتی، ممکن می سازد. در روش پیشنهادی ابتدا به منظور قطعه بندی خطوط ریلی سه مدل یادگیری عمیق با بهره گیری از معماری توجه هرمی و رمزگذارهای شبکه ی رزنت فشرده سازی و تحریک، شبکه ی کارآمد و شبکه بدون نرمال ساز، توسط مجموعه ای از تصاویر رنگی، تحت آموزش و اعتبارسنجی قرار گرفتند. نتایج اعتبارسنجی نشان می دهد که مدل قطعه بندی مبتنی بر شبکه ی بدون نرمال ساز با چند دهم درصد اختلاف از دو مدل دیگر دقیق تر است. این در حالی است که بار محاسباتی مدل مبتنی بر شبکه ی کارآمد حدوداً  $\frac{1}{8}$  و  $\frac{1}{4}$  برابر بار محاسباتی مدل های مبتنی بر شبکه های رزنت سوپر پیچشی و بدون نرمال ساز است. به منظور تشخیص اشیا نسخه های پنجم و هفتم الگوریتم یولو تحت آموزش و اعتبارسنجی قرار گرفتند. نتایج نشان می دهد که دقت نسخه هفتم یولو در تشخیص موانع تعریف شده ۱۷/۶ درصد از دقت نسخه پنجم این مدل بیشتر است. در نهایت پس از پیاده سازی مدل قطعه بندی ریل مبتنی بر شبکه ی کارآمد و نسخه هفتم الگوریتم یولو در ساختار پیشنهادی عملیات تشخیص موانع با بررسی وجود همپوشانی میان ماسک مسیر حرکت قطار و کادرهای محصورکننده ی اشیا انجام شد. در روش پیشنهادی تشخیص حضور موانع در مسیر حرکت قطار به ارسال پیام هشدار می انجامد. از این رو می توان از این روش پیشنهادی به عنوان سیستم کمک راهبر قطار استفاده نمود تا از خطای احتمالی راهبر جلوگیری شود.

واژگان کلیدی: بینایی ماشین، تشخیص موانع، راه آهن، قطعه بندی ریل، یادگیری عمیق

## ۱. مقدمه

ریل یا مسیر حرکت قطار و موانع احتمالی موجود در محدوده ریلی، از داده‌های ثبت شده توسط حسگرها، با استفاده از روش‌های دقیق و زمان حقیقی استخراج شود. لذا در این مسیر سه چالش اساسی وجود دارد:

- ادراک محیط

- تشخیص مسیر حرکت قطار

- تشخیص موانع

هدف بکارگیری دستگاه‌های ادراک محیطی ارائه اطلاعات حیاتی در مورد محیط اطراف سیستم، موانع احتمالی، سرعت، موقعیت مکانی و حتی پیش‌بینی وضعیت آینده آن‌ها است. اطلاعات جمع‌آوری شده توسط دستگاه‌های ادراک محیطی یا حسگرها در واقع ورودی و عنصر اصلی طراحی سیستم‌های تشخیص موانع است. در سیستم‌های تشخیص موانع برای ادراک محیط، معمولاً چندین فناوری سنسجش مانند لیدار، رادار، اولتراسوند و دوربین‌های تصویربرداری را به منظور شناسایی موانع و ارائه موقعیت فیزیکی آن‌ها، ترکیب می‌کنند. ادغام حسگرها به غلبه بر معایب و محدودیت‌های یک حسگر با استفاده از مزایای سایر حسگرها کمک می‌کند. کاربرد اصلی حسگرهای فعال در فاصله سنجی و کاربرد اصلی حسگرهای غیر فعال در ثبت جزئیات، شناسایی و تشخیص است. اگرچه دقت فناوری‌های سنسجش فعال مانند لیدار، رادار و حسگرهای اولتراسوند بالا است، اما این حسگرها بار محاسباتی زیادی مصرف می‌کنند، در اجرا پرهزینه هستند و بدون ترکیب با حسگرهای غیرفعال اطلاعات کافی برای شناسایی محیط با هدف تشخیص موانع را فراهم نمی‌کنند.

در مقاله، [Ristić-Durrant, D., Franke, M., & Michels, K. (2021)] مروری بر ادبیات تشخیص موانع ریلی مبتنی بر دید ارائه شده است. در این مقاله مروری دوربین‌های تصویر برداری به عنوان مهم‌ترین و کارآمدترین حسگرهای ادراک محیطی معرفی شده اند؛ زیرا این حسگرها اطلاعات بصری غنی از محیط اطراف را با تمام جزئیات ممکن،

راه آهن به عنوان یک شریان ارتباطی موثر جایگاه ویژه ای در رشد و توسعه اقتصادی کشورها دارد. قابلیت حمل انبوه، صرفه جویی در مصرف انرژی، سرعت و ایمنی بالا از مزیت‌های حمل‌ونقل ریلی است که باعث تمرکز بسیاری از کشورهای جهان بر توسعه حمل‌ونقل ریلی شده است. یکی از مهم‌ترین پارامترهای حمل و نقل ریلی ایمنی بالای این صنعت و نرخ پایین تصادفات ریلی نسبت به نرخ تصادفات جاده‌ای است. با این حال گزارشات کمیسیون عالی سوانح راه آهن، نشان می‌دهد که ایران از نظر تعداد تصادفات ریلی در بین کشورهای جهان رتبه بالایی دارد لذا باید اقدامات موثری برای افزایش ایمنی حمل‌ونقل ریلی در ایران انجام شود. یکی از عوامل اصلی سوانح ریلی مانند آن چه در سال اخیر در خط مشهد به یزد رخ داد خطای انسانی است. سانحه‌ی قطار مشهد به یزد در تاریخ ۱۸ خرداد ۱۴۰۱ در نزدیکی ایستگاه مزینو در ۵۰ کیلومتری طبس رخ داد. این حادثه به علت خطای انسانی راهبر قطار اتفاق افتاد و قطار به یک بیل مکانیکی برخورد کرد. این قطار ۳۴۸ مسافر داشت. در این سانحه ۱۴ نفر کشته و ۸۶ نفر مصدوم شدند. چنین حوادثی نشان می‌دهد که ایمنی حمل و نقل ریلی در ایران با چالش‌های جدی روبرو است و برای کاهش این چالش‌ها، اقدامات فوری و موثری لازم است.

طراحی یک سیستم تشخیص موانع<sup>۱</sup> ریلی یک بهبود عملیاتی و ایمنی در راه آهن است که خطاهای انسانی را از بین می‌برد و راه آهن را دقیق‌تر و قابل اعتمادتر می‌سازد. هدف این پژوهش طراحی سیستم نظارت تصویری است که وجود مانع در مسیر حرکت قطار را تشخیص داده و راهبر را با تولید پیام هشدار از وجود مانع مطلع سازد. برای دستیابی به هدف تشخیص موانع ریلی نخست به بهره‌گیری از ابزار و حسگرهایی نیاز است که داده‌های محیط اطراف را بصورت دقیق و با جزئیات بالا ثبت کند، سپس باید اطلاعات مفید مورد نیاز؛ مانند محدوده قرارگیری

## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش‌های مبتنی بر یادگیری عمیق

مختلف به‌طور همزمان آموزش داده می‌شوند. در این روش به ماشین (رایانه) مجموعه داده‌ای از تصاویر داده می‌شود که با کلاس‌هایی از اشیاء در هر تصویر مشروح شده‌اند. لایه‌های عمیق در این شبکه‌ها به‌عنوان استخراج‌کننده ویژگی‌های عمومی عمل می‌کنند. این بدین معنی است که یک روش مبتنی بر یادگیری عمیق منجر به مجموعه‌ای از ویژگی‌های تصویر می‌شود که مستقیماً از مشاهدات تصاویر ورودی به‌دست آمده است و این ویژگی‌ها را از تصاویر ورودی آموخته است. این ویژگی‌ها ویژگی‌های غیر دستی<sup>۱۲</sup> نامیده می‌شوند [Nanni, L. et al, 2017]. در سال‌های اخیر، پیشرفت فناوری شبکه‌های عصبی، پیشرفت زیادی را در تشخیص اشیاء به‌وسیله روش‌های مبتنی بر هوش مصنوعی در کاربردهای ترافیک جاده‌ای ممکن کرده است، اما تحقیقات در حوزه‌ی خودران‌سازی حمل و نقل ریلی پیشرفت چشمگیری نداشته است. با این حال، در چند سال اخیر، علاقه فزاینده‌ای به کاربردهای مبتنی بر هوش مصنوعی در حمل و نقل ریلی وجود داشته تا از استفاده ویژگی‌های دست‌ساز مانند مواردی که در روش‌های سنتی بینایی کامپیوتر استفاده می‌شود، اجتناب شود.

در ساختار پیشنهادی ما از دو مدل یادگیری عمیق به منظور قطعه‌بندی ریل<sup>۱۳</sup> و تشخیص اشیاء استفاده شده است. این مدل‌های یادگیری عمیق توسط مجموعه‌ای از داده‌های برچسب‌دار آموزش داده می‌شوند و با قرارگیری در ساختار شبکه‌ی پیشنهادی عملیات تشخیص موانع را ممکن می‌سازند. خروجی مدل قطعه‌بندی ریل یک ماسک است که بخش‌های مربوط به کلاس‌های مدنظر علامت‌گذاری رنگی شده‌اند. خروجی مدل تشخیص اشیاء مجموعه‌ای از کادرهای محصور کننده است که کلاس‌های مشخص از اشیاء موجود در تصویر را دربر دارد. ساختار شبکه‌ی تشخیص موانع پیشنهادی ما به این صورت است که بعد از دریافت تصاویر رنگی به‌عنوان ورودی، این تصاویر به مدل قطعه‌بندی ریل و تشخیص اشیاء<sup>۱۴</sup> که از پیش آموزش دیده‌اند وارد می‌شود تا اطلاعات مربوط به مسیر

در قالب تصاویر ارائه می‌دهند. در این جا نیز دوربین تصویربرداری رنگی به‌عنوان حسگر ادراک محیطی پایه در نظر گرفته شده است.

برای حل چالش‌های تشخیص ریل و تشخیص موانع از تصاویر رنگی، نیاز به بکارگیری روش‌هایی است که ویژگی‌ها و اطلاعات مفید را با دقت بالا و بصورت زمان حقیقی از تصاویر ثبت شده توسط دوربین استخراج کند و پاسخگوی محیط پیچیده ریلی نیز باشد. با توجه به خط ترمز بلند قطار روش پیشنهادی باید فاصله سنجش و سرعت پردازش مناسبی را فراهم آورد تا مانع در فواصل دور شناسایی و قطار زمان کافی برای کاهش سرعت و توقف را داشته باشد.

بر اساس تکنیک‌های مورد استفاده برای استخراج ویژگی‌های تصویر، همه روش‌های تشخیص موانع بر اساس بینایی را می‌توان به دو دسته‌ی روش‌های مبتنی بر بینایی کامپیوتر<sup>۲</sup> سنتی و روش‌های مبتنی بر هوش مصنوعی<sup>۳</sup> تقسیم کرد. گروه اول شامل روش‌هایی است که از تکنیک‌های بینایی کامپیوتر برای قطعه‌بندی<sup>۴</sup> تصویر مانند تشخیص لبه<sup>۵</sup>، تشخیص گوشه<sup>۶</sup> و قطعه‌بندی آستانه<sup>۷</sup> استفاده می‌کنند. در این روش سنتی، تشخیص اشیاء بر اساس استخراج ویژگی‌های دست‌ساز یا دستی<sup>۸</sup> انجام می‌شود. ویژگی‌های دستی به ویژگی‌های استخراج‌شده از یک تصویر مانند لبه‌ها، گوشه‌ها و توصیف‌کننده‌های جزئی اشاره دارد. این ویژگی‌ها با استفاده از روش‌های مختلف و از اطلاعات موجود در خود تصویر استخراج می‌شوند. روش‌های گروه دوم مبتنی بر یادگیری ماشین<sup>۹</sup> و به‌ویژه یادگیری عمیق<sup>۱۰</sup> است. برخلاف روش‌های سنتی بینایی کامپیوتر که از ویژگی‌های دست‌ساز برای طبقه‌بندی نواحی تصویر استفاده می‌کنند، یادگیری عمیق از مفهوم یادگیری انتها به انتها<sup>۱۱</sup> استفاده می‌کند. در این تکنیک مدل، تمام مراحل بین فاز ورودی اولیه و نتیجه‌ی خروجی نهایی را یاد می‌گیرد. این یک فرآیند یادگیری عمیق است که در آن همه‌ی بخش‌های

حرکت قطار و اشیا موجود در تصویر به طور زمان حقیقی از تصاویر استخراج شود. خروجی مدل اول بصورت ماسکی که تنها در ناحیه مسیر حرکت قطار (محدوده سبز رنگ در شکل) مقدار یک دارد و خروجی مدل تشخیص اشیا به صورت ماسکی که در ناحیه درون کادرهای محصور کننده اشیا مقدار یک و در سایر نقاط مقدار صفر دارد ذخیره می شود. در نهایت با بررسی وجود همپوشانی میان ماسک اشیا و ماسک مسیر حرکت قطار، حضور مانع در مسیر تشخیص داده می شود.

## ۲. ادبیات پژوهش

همان طور که پیشتر اشاره شد مدل های قطعه بندی ریل و تشخیص اشیا بخش های اصلی ساختار پیشنهادی ما را تشکیل می دهند و نقش اساسی در عملکرد مدل نهایی دارند از این رو باید از مدل یادگیری عمیق مناسب برای هر یک از این بخش ها استفاده شود. با توجه به معماری های متنوع مدل های یادگیری عمیق، با هدف دستیابی به معماری مناسب برای حل مسئله تعریف شده و حفظ توازن میان دقت و بار محاسباتی، سه مدل یادگیری عمیق مختلف با بهره گیری از معماری توجه هرمی<sup>۱۵</sup> (PAN) و رمزگذارهای<sup>۱۶</sup>، رزنت فشرده سازی و تحریک (SE\_ResNet)<sup>۱۷</sup> [Hu, J., Shen, L., Sun, G. 2018] شبکه ی کارآمد (EfficientNet)<sup>۱۸</sup> [Tan, M., 2018] و شبکه ی بدون نرمال ساز (NF\_Net)<sup>۱۹</sup> [Le, Q. 2019] به منظور قطعه بندی ریل طراحی و اعتبارسنجی شدند.

### ۲-۱- مروری بر حسگرها و ابزارهای ادراک محیطی

در فن آوری حسگرها برای تشخیص موانع و حفظ ایمنی وسیله حمل و نقل زمینی، درک و آگاهی دقیق از محیط ضروری است، حسگرها از نظر عملکرد و خروجی با یکدیگر متفاوت بوده و از آن ها باید با توجه به قابلیت هایشان استفاده شود. مروری بر حسگرهای آنبورد در ادراک محیط ریلی توسط نویسندگان [Gebauer, O.; Pree, W.; Stadlmann, B, 2018] ارائه شده است.

در این جا این سه مدل قطعه بندی ریل به ترتیب با نام های PAN+SE\_ResNet، PAN+EfficientNet و PAN+NF-Net شناخته می شوند. همچنین نسخه های پنجم و هفتم الگوریتم یولو (YOLOv5 و YOLOv7) به منظور تشخیص کلاس های اشیا مدنظر اعتبارسنجی شده اند. با توجه به نتایج اعتبارسنجی مدل ها مورد بحث که در بخش نتایج به آن پرداخته شده است، نهایتاً مدل قطعه بندی ریل

اگرچه راه آهن یکی از ابزارهای اصلی حمل و نقل زمینی است، تحقیق و توسعه در زمینه ی تشخیص موانع حمل و نقل ریلی تا به امروز بسیار کمتر از حمل و نقل جاده ای بوده است. در نتیجه، تعداد آثار مرتبط منتشر شده به طور قابل توجهی نسبت به حمل و نقل جاده ای محدود است. با این تفاسیر و با در نظر گرفتن برخی شباهت های بین این دو وسیله ی حمل و نقل زمینی، قابل درک است که تحقیقات اخیر در تشخیص موانع ریلی بر پایه تحقیقات در زمینه ی خودرو بنا شده باشد؛ یعنی محققان و توسعه دهندگان از تکنیک هایی که در فناوری های کمک راننده خودرو شناخته اند، برای توسعه سیستم های تشخیص موانع حمل و نقل ریلی استفاده کرده اند.

۲-۱-۱- مروری بر حسگرها و ابزارهای ادراک محیطی

در فن آوری حسگرها برای تشخیص موانع و حفظ ایمنی وسیله حمل و نقل زمینی، درک و آگاهی دقیق از محیط ضروری است، حسگرها از نظر عملکرد و خروجی با یکدیگر متفاوت بوده و از آن ها باید با توجه به قابلیت هایشان استفاده شود. مروری بر حسگرهای آنبورد در ادراک محیط ریلی توسط نویسندگان [Gebauer, O.; Pree, W.; Stadlmann, B, 2018] ارائه شده است.

در این جا این سه مدل قطعه بندی ریل به ترتیب با نام های PAN+SE\_ResNet، PAN+EfficientNet و PAN+NF-Net شناخته می شوند. همچنین نسخه های پنجم و هفتم الگوریتم یولو (YOLOv5 و YOLOv7) به منظور تشخیص کلاس های اشیا مدنظر اعتبارسنجی شده اند. با توجه به نتایج اعتبارسنجی مدل ها مورد بحث که در بخش نتایج به آن پرداخته شده است، نهایتاً مدل قطعه بندی ریل

در این جا این سه مدل قطعه بندی ریل به ترتیب با نام های PAN+SE\_ResNet، PAN+EfficientNet و PAN+NF-Net شناخته می شوند. همچنین نسخه های پنجم و هفتم الگوریتم یولو (YOLOv5 و YOLOv7) به منظور تشخیص کلاس های اشیا مدنظر اعتبارسنجی شده اند. با توجه به نتایج اعتبارسنجی مدل ها مورد بحث که در بخش نتایج به آن پرداخته شده است، نهایتاً مدل قطعه بندی ریل

## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش‌های مبتنی بر یادگیری عمیق

نویسندگان در این روش از مجموعه داده‌ای شامل ۲۹۰۰ تصویر برداشت‌شده توسط یک دوربین با وضوح (۲۱۶۸ × ۴۰۹۶) نصب‌شده بر روی بدنه‌ی قطار استفاده کردند. در این مجموعه داده‌ی دو کلاسه، مسیر حرکت قطار با ماسک سبز و تمام مسیرهای موجود دیگر با ماسک قرمز نشانه‌گذاری شده بودند.

### ۲-۳ مروری بر روش‌های تشخیص موانع

مطالعه پژوهش‌های مرتبط نشان می‌دهد که نویسندگان برای تشخیص موانع ریلی عمدتاً از روش‌ها یا شبکه‌های مبتنی بر یادگیری عمیق برای تشخیص اشیاء استفاده می‌کنند که پیش‌تر در برنامه‌های کاربردی دیگر استفاده شده بودند.

در [Ye, T. et al, 2020] یک شبکه تشخیص شیء با معماری شبکه کاملاً پیچشی به نام DFF-Net پیشنهاد شده است که می‌تواند مسیرهای ریلی و چند نوع مانع مختلف که معمولاً در محیط راه‌آهن وجود دارند را شناسایی کند. یک روش مبتنی بر یادگیری عمیق برای تشخیص اجسام کوچک در محیط راه‌آهن در [Gleichauf, J. et al, 2020] پیشنهاد شده است، روش پیشنهادی مبتنی بر FE-SSD<sup>۲۵</sup> است. نتایج ارزیابی نشان داده است که روش پیشنهادی تعادل خوبی بین دقت و عملکرد زمان حقیقی را برای تشخیص موانع فراهم می‌کند و به‌طور قابل توجهی عملکرد تشخیص اشیاء کوچک را نسبت به چندین روش دیگر بهبود می‌بخشد [Li, J.; Zhou, F.; Ye, T., 2018].

نتایج ارائه‌شده در [Haseeb, M.A. et al, 2018, Ristić, Durrant, et al, 2020] با استفاده از یکی از پیشرفته‌ترین روش‌های مبتنی بر یادگیری عمیق به نام الگوریتم YOLO به دست آمد [Redmon, J.; Farhadi, A., I hsj2021]. مزیت اصلی YOLO سرعت این شبکه است که باعث می‌شود برای تشخیص زمان حقیقی در محیط ریلی مناسب باشد. الگوریتم YOLO مورد استفاده در این پژوهش برای اولین بار با مجموعه داده‌ی COCO آموزش داده شد [Lin, T, et al, 2020]. به‌منظور فعال کردن الگوریتم YOLO برای تشخیص

شده است. [Wang, Z. et al, 2018, Xu, Y. et al, 2019, Ye, T. et al, 2018] مجموعه داده ارزیابی در این پژوهش‌ها ویدیوی است که توسط یک دوربین نصب شده روی قطار ضبط شده است.

### ۲-۲ مروری بر روش‌های تشخیص ریل

به‌منظور تشخیص ریل به‌وسیله روش‌ها مبتنی بر یادگیری عمیق معمولاً از شبکه‌های قطعه‌بندی<sup>۲۰</sup> استفاده می‌شود بدین معنی که مدل هوش مصنوعی با استفاده از یکسری داده که در آن خطوط ریلی نشانه‌گذاری رنگی شده‌اند آموزش می‌بیند. پژوهشگران در [Wang, H. et al, 2018] روشی مبتنی بر شبکه‌ی عصبی پیچشی<sup>۲۱</sup> به‌منظور تشخیص خطوط راه‌آهن پیشنهاد داده‌اند. در این روش آموزش مدل قطعه‌بندی توسط یک مجموعه داده سفارشی با وضوح تصویر ۱۲۸۰ × ۷۲۰ که در آن تمام مسیرهای ریلی برجسب‌گذاری سبز شده بودند انجام شد. در این پژوهش [Wang, H. et al, 2018] یک مدل قطعه‌بندی ریل به نام RailNet ارائه شد. یک الگوریتم قطعه‌بندی مسیر ریلی مبتنی بر یادگیری عمیق است که از یک شبکه استخراج ویژگی و یک شبکه قطعه‌بندی تشکیل شده است. شبکه استخراج ویژگی از یک ساختار هرمی و ستون فقرات ResNet50 برای انتشار ویژگی‌ها و به‌دست آوردن یک بردار ویژگی ترکیبی استفاده می‌کند. برای آموزش و آزمایش شبکه‌ی RailNet، نویسندگان مجموعه داده‌ای را ایجاد کردند که همه تصاویر از محیط ریلی و در دنیای واقعی گرفته شده بود. در برجسب‌گذاری مجموعه داده‌های راه‌آهن، مسیر ریلی به‌عنوان تمام پیکسل‌های بین دو ریل تعریف شد.

در [S., Belyaev, I, Popov, 2020] نویسندگان شبکه‌ی اصلاح‌شده‌ای به نام RailCNN ارائه دادند که معماری شبکه قطعه‌بندی پیشنهادی شامل بلوک‌های توجه<sup>۲۲</sup> و ادغام هرم فضایی<sup>۲۳</sup> با بهره‌گیری از ستون فقرات<sup>۲۴</sup> ResNet34 بود که استخراج و حفظ ویژگی‌های مفید تصویر را میسر می‌کرد

مسیرهای ممکن با دقت ۹۴٪ و سرعت پردازش ۱۷ فریم بر ثانیه قطعه‌بندی می‌کند.

در بخش شناسایی اشیا قدرتمندترین شبکه‌های ارائه شده از خانواده R\_CNN ها و Faster\_RCNN ها هستند اما با روی کار آمدن الگوریتم‌های YOLO با توجه سرعت بالایی که این الگوریتم‌ها در پردازش زمان حقیقی ارائه دادند مبنای کار قرار گرفتند. در [Ye, T; Zhang, Z, 2020] روشی بر مبنای YOLOV3 ارائه شده است که کلاس‌های اشیا تعریف شده را با دقت ۸۹/۵ درصد و سرعت پردازش ۳۸ فریم بر ثانیه شناسایی می‌کند. (در موارد بالا منظور از دقت شاخص IOU و پردازنده گرافیکی مورد استفاده NVIDIA GTX1080Ti است.)

در روش پیشنهادی ما، بدون استفاده از حسگرهای متعدد و تنها با استفاده از داده دوربین تصویربرداری رنگی، بعد از بررسی الگوریتم‌های مختلف ساختاری طراحی و پیشنهاد شده است که با رعایت توازن میان دقت و بار محاسباتی تشخیص مسیرهای ریلی، تفکیک مسیر حرکت قطار از سایر مسیرهای موجود، تشخیص اشیا و تعیین حضور موانع در مسیر حرکت قطار را با دقت قابل قبولی ممکن می‌سازد. در روش پیشنهادی در صورت تشخیص حضور موانع پیام هشدار صادر می‌شود. از این رو می‌توان از این سیستم به‌عنوان سیستم کمک راهبر قطار استفاده نمود. این مطالعه شامل بخش‌های روش، نتایج و جمع‌بندی است.

### ۳. روش پژوهش

در روش پیشنهادی ما از ترکیب چند شبکه‌ی عصبی پیچشی برای پردازش تصاویر ورودی و شناسایی ریل و موانع موجود استفاده شده است. شبکه عصبی پیچشی نوعی شبکه عصبی عمیق است و معمولاً برای طبقه‌بندی و قطعه‌بندی تصاویر استفاده می‌شود. یک شبکه عصبی پیچشی از یک لایه‌ی ورودی، لایه‌ی خروجی و همچنین چندین لایه‌ی پنهان تشکیل شده است. لایه‌های پنهان، معمولاً از لایه‌های پیچشی، لایه‌های ادغام<sup>۲۶</sup> و لایه‌های تماماً متصل<sup>۲۷</sup> تشکیل شده‌اند. وظیفه اصلی

فصلنامه مهندسی حمل و نقل / سال شانزدهم / شماره دوم (۶۳) / زمستان ۱۴۰۳

اشیاء در محیط ریلی، با تمرکز ویژه بر اجسام دوربرد، الگوریتم YOLO بر روی مجموعه داده‌های سفارشی راه‌آهن دوربرد توسعه یافته به‌عنوان بخشی از پروژه SMART بازآموزی شد [Ristic-Durrant, et al, 2020].

پس از مروری بر ادبیات موضوع میتوان دریافت که تشخیص موانع ریلی امر پیچیده‌ای است که با چالش‌های متعددی مانند تشخیص دوربرد اشیا، مقابله با عوامل محیطی، تشخیص ریل، تفکیک مسیر حرکت قطار از سایر مسیرهای موجود، بار محاسباتی پایین، پردازش زمان حقیقی و دقت بالا روبرو است. همچنین اکثر مطالعات موجود در این زمینه راه‌حل‌های جداگانه‌ای برای تشخیص موانع و تشخیص ریل ارائه داده‌اند تعداد محدودی از پژوهش‌های موجود به تشخیص همزمان مسیر حرکت قطار و موانع پرداخته‌اند و در اکثر پژوهش‌ها به صورت اختصاصی به یکی از این دو چالش پرداخته شده است [Ristić-Durrant, D., Franke, M., & Michels, K, 2021]. پژوهش انجام شده در [Ye, T. et al, 2020] بر مبنای تصاویر برداشت شده با یک دوربین رنگی و یک رادار موج میلی‌متری، به‌صورت همزمان به تشخیص ریل و تشخیص موانع پرداخته است. در این پژوهش راهکاری برای تفکیک مسیر حرکت قطار از سایر مسیرهای موجود در تصویر و تعیین حضور یا عدم حضور اشیا شناسایی شده، در مسیر حرکت قطار ارائه نشده است و شبکه پیشنهادی کلاس‌های اشیا و مسیرهای ریلی را با دقت میانگین ۷۹/۸۹ درصد با سرعت پردازش ۵۴ فریم بر ثانیه شناسایی می‌کند.

شبکه‌های RailNet و RailCNN از بهترین شبکه‌هایی هستند که تاکنون به هدف قطعه‌بندی ریل بر مبنای تصاویر رنگی ارائه شده‌اند. شبکه RailNet تمام مسیرهای ریلی موجود در تصویر را در یک کلاس قرمز رنگ با دقت ۸۹/۸ درصد و سرعت پردازش ۲۰ فریم بر ثانیه قطعه‌بندی می‌کند. شبکه RailCNN مسیرهای ریلی را در دو کلاس تمام مسیرهای موجود و

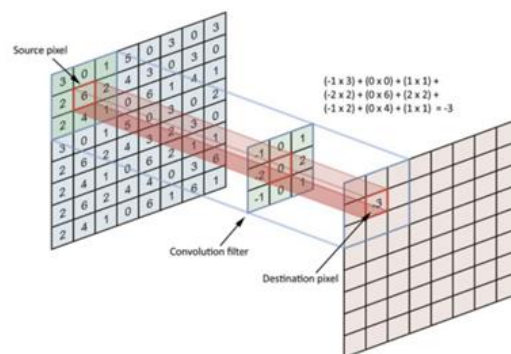
## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش‌های مبتنی بر یادگیری عمیق

می‌کند. یکی از سخت‌ترین چالش‌های شبکه‌های عصبی پیچشی در بینایی کامپیوتر، تشخیص اشیا و قطعه‌بندی تصویر است. تشخیص اشیا و وظیفه‌ای برای طبقه‌بندی اشیا در یک تصویر به برجسب‌های اشیا شناخته شده و قطعه‌بندی وظیفه‌ای برای تقسیم تصویر به مناطق مد نظر یا اشیا است. تاکنون معماری‌های متفاوتی مانند VGGNet، GoogLeNet و ResNet به منظور تشخیص اشیا و معماری‌هایی نظیر PSPNet، Unet و PAN به منظور قطعه‌بندی تصویر معرفی شده‌اند. ما برای شناسایی موانع ریلی در تصاویر به معماری و شبکه‌هایی نیاز داریم که علاوه بر دقت بالا سرعت پردازش بالایی فراهم آورد و مناسب مصرف زمان حقیقی باشد.

### ۳-۱ مدل قطعه‌بندی ریل

وظیفه مدل قطعه‌بندی ریل استخراج نقشه‌ی ویژگی‌های مربوط به خطوط ریلی از تصویر ورودی و تولید ماسک خروجی است. در این جا برای پیاده‌سازی شبکه قطعه‌بندی ریل از معماری توجه هرمی استفاده شده است. شبکه توجه هرمی یک مدل بینایی کامپیوتری است که به‌طور خاص برای کمک به درک بهتر تصاویر طراحی شده است و از یک ساختار «هرمی» برای تجزیه و تحلیل تصاویر در سطوح مختلف جزئیات استفاده می‌کند. همچنین دارای مکانیزم توجه است که به مدل کمک می‌کند تا بر مهم‌ترین قسمت‌های تصویر تمرکز کند. این باعث می‌شود شبکه توجه هرمی یک ابزار قدرتمند برای رویکردهای مانند طبقه‌بندی تصویر، تشخیص اشیا و قطعه‌بندی معنایی باشد. معماری این شبکه از سه جزء اصلی شامل رمزگذار، رمزگشا<sup>۳۰</sup> و سر قطعه‌بندی<sup>۳۱</sup> تشکیل شده است. رمزگذار مسئول گرفتن تصویر ورودی و تولید مجموعه‌ای از نقشه‌های ویژگی است. رمزگذار با سه کانال ورودی قرمز، سبز و آبی تصاویر را می‌گیرد و از ویژگی‌های آموخته‌شده برای ایجاد مجموعه‌ای از نقشه‌های ویژگی استفاده می‌کند. رمزگشا مسئول گرفتن نقشه‌های ویژگی تولیدشده توسط رمزگذار و تولید یک ماسک قطعه‌بندی است.

لایه‌ی پیچشی استخراج ویژگی‌ها از تصویر ورودی یا لایه‌ی قبلی، با استفاده از یک هسته پیچشی است [Liu, Y. H., 2018]. هر لایه‌ی پیچشی از تعداد زیادی نقشه‌های ویژگی تشکیل شده است. هر نقشه‌ی ویژگی خروجی، پیچش بین هسته‌ی فعلی و نقشه‌ی ویژگی لایه‌ی قبلی است.



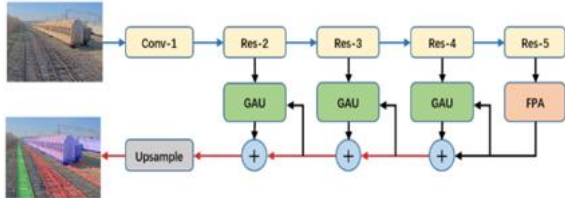
شکل ۱. نمونه‌ای از عملیات پیچشی

محاسبه‌ی لایه‌ی پیچشی را می‌توان به‌صورت فرمول زیر نشان داد.

$$y_j^l = f\left(\sum_{i \in M_j} y_i^{l-1} \otimes k_{ij}^l + b_j^l\right) \quad (1)$$

در این رابطه  $k_{ij}^l$  هسته‌ی پیچشی بین  $i$  امین نقشه‌ی ویژگی از  $l$  امین لایه و  $j$  امین نقشه‌ی ویژگی از  $l-1$  امین لایه است،  $M_j$  یک نمونه از نقشه‌ی ویژگی ورودی است. پارامتر  $b_j^l$  بایاس  $j$  امین نقشه‌ی ویژگی از  $l$  امین لایه است و  $f$  تابع فعال‌ساز شامل توابع Sigmoid، Tanh، ReLU و غیره است. لایه‌ی ادغام ابعاد هر نقشه‌ی ویژگی را کاهش می‌دهد و درعین‌حال اطلاعات ارزشمند را حفظ می‌کند [۵۷]. به‌طور عمده دو نوع لایه‌ی ادغام وجود دارد، لایه‌ی ادغام حداکثر<sup>۲۸</sup> و لایه‌ی ادغام میانگین<sup>۲۹</sup>. لایه‌ی ادغام حداکثر، با ارزش‌ترین عدد را از هر نقشه و لایه‌ی ادغام میانگین عدد متوسط هر نقشه را انتخاب می‌کند. پس از آن، لایه‌ی تماماً متصل، ماتریس نقشه‌ی ویژگی لایه‌ی قبلی را به شکل برداری تبدیل می‌کند. درنهایت، لایه‌ی خروجی، کلاس را با استفاده از یک تابع فعال‌سازی که شامل رگرسیون softmax یا رگرسیون Sigmoid است، تولید و تصاویر را طبقه‌بندی

نهایی مدل و مسئول تولید ماسک قطعه‌بندی نهایی است. این بخش از مجموعه‌ای از لایه‌های پیچشی تشکیل شده است که نقشه‌های ویژگی ارتقاء‌یافته‌ی تولیدشده توسط رمزگشا را می‌گیرند و ماسک قطعه‌بندی نهایی را خروجی می‌دهند.

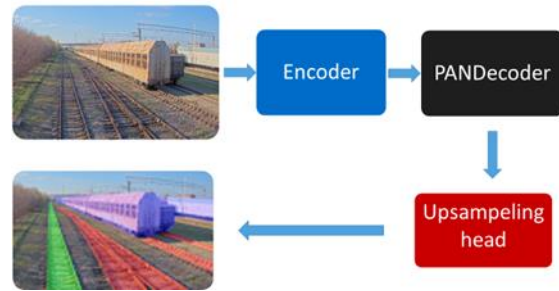


شکل ۳. نمای کلی شبکه‌ی PAN. با ستون فقرات ResNet

خطوط آبی و قرمز به ترتیب نشان دهنده عملگرهای نمونه‌برداری کاهشی و نمونه‌برداری افزایشی هستند.

هر سه مدل قطعه‌بندی PAN+EfficientNet، PAN+NF\_Net و PAN+SE\_ResNet، مدل‌های مبتنی شبکه عصبی پیچشی هستند که برای طبقه‌بندی تصویر طراحی شده‌اند. با این حال، از نظر معماری و عملکرد با یکدیگر تفاوت دارند. در بخش رمزگزار شبکه‌ی PAN+EfficientNets ایده مقیاس‌بندی ترکیبی استفاده شده است تا دقت و کارایی بهبود یابد. مقیاس‌بندی ترکیبی شامل مقیاس‌بندی هر سه بعد عمق، عرض و وضوح است. عمق، عرض و وضوح به ترتیب به تعداد لایه‌های پیچشی شبکه، تعداد فیلترها در هر لایه و اندازه تصویر ورودی اشاره دارد. شبکه‌ی SE\_ResNet از بلوک‌های فشرده‌سازی و تحریک (SE) استفاده می‌کند. بلوک‌های SE به بهبود عملکرد مدل با تنظیم اهمیت کانال‌های مختلف کمک می‌کنند و باعث حفظ و تمرکز مدل بر ویژگی‌های مهم استخراج شده می‌شود. بلوک SE از دو بخش فشرده‌سازی و برانگیختگی تشکیل شده است. مرحله فشرده‌سازی ویژگی‌های هر کانال را به یک مقدار خلاصه می‌کند و مرحله برانگیختگی از مقدار خلاصه شده برای تنظیم اهمیت هر کانال استفاده می‌کند.

مزیت شبکه‌ی PAN+NFNet این است که بدون استفاده از لایه‌های نرمال‌سازی، مانند نرمال‌ساز دسته‌ای، آموزش داده می‌شود این فرآیند منجر به آموزش سریع‌تر و دقت بهتر



شکل ۲. ساختار مدل قطعه‌بندی ریل

در این جا سه شبکه‌ی پیچشی مختلف در نسخه‌های مشخص زیر به عنوان رمزگذار استفاده شده است [Pavel Iakubovskii 2023].

tf\_efficientnet\_b4 -  
eca\_nfnet\_l2 -  
seresnet152 -

این شبکه‌ها با هدف دستیابی به توازن میان دقت و بار محاسباتی در معماری مدل قطعه‌بندی تعریف شده، پیاده‌سازی شده و سه مدل قطعه‌بندی ریل PAN+EfficientNet، PAN+NF\_Net و PAN+SE\_ResNet را تولید می‌کنند. رمزگشا در مدل‌های قطعه‌بندی ذکر شده یک PANDecoder [Pavel Iakubovskii 2023] است که از بلوک‌های

هرمی<sup>۳۲</sup>، بلوک‌های توجه<sup>۳۳</sup> و لایه‌های نمونه‌برداری<sup>۳۴</sup> تشکیل شده است. این بلوک‌ها و لایه‌ها نقشه‌های ویژگی را اصلاح می‌کنند و یک ماسک قطعه‌بندی با وضوح بالا ایجاد می‌کنند. بلوک‌های هرمی استخراج ویژگی‌های میانی را بر روی نقشه‌های ویژگی تولیدشده توسط رمزگذار انجام می‌دهند. هر بلوک هرمی از یک سری عملیات پیچشی و نمونه‌برداری کوچک تشکیل شده است. بلوک‌های توجه به مدل کمک می‌کنند تا روی ویژگی‌های مهم تمرکز کند و موارد نامربوط را نادیده بگیرد. هر بلوک توجه، خروجی یک بلوک هرمی را می‌گیرد و از آن برای محاسبه نقشه‌ی توجه<sup>۳۵</sup> استفاده می‌کند که روی نقشه‌های ویژگی ورودی اعمال می‌شود. لایه‌های نمونه‌برداری نقشه‌های ویژگی تولیدشده توسط هرم و بلوک‌های توجه را به اندازه‌ی اصلی تصویر ورودی، نمونه‌برداری می‌کنند. سر قطعه‌بندی جزء

## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش‌های مبتنی بر یادگیری عمیق

نسبت به شبکه‌های دارای نرمال‌ساز دسته‌ای می‌شود. شبکه‌ی PAN+NFNet در آموزش سریع‌تر بود و می‌توانست اندازه دسته‌ای تا ۴۰۹۶ داشته باشد. نرمال‌سازی دسته‌ای تکنیکی است که در یادگیری عمیق برای بهبود عملکرد شبکه‌های عصبی استفاده می‌شود. این تکنیک شامل نرمال‌سازی ورودی‌های هر لایه در شبکه با کم کردن میانگین دسته‌ای و تقسیم بر انحراف از معیار دسته‌ای است. این روش به جلوگیری از مشکل ناپدید شدن یا انفجار گرادیان کمک می‌کند، مشکل گرادیان زمانی رخ می‌دهد که ورودی‌های شبکه خیلی بزرگ یا خیلی کوچک هستند.

اعتبارسنجی سه شبکه‌ی قطعه‌بندی مورد بحث به وسیله‌ی شاخص نسبت اشتراک به اجتماع (IOU)<sup>۳۷</sup> و تابع زیان جاکارد<sup>۳۸</sup> انجام شد.

$$\text{Jaccard Loss} = 1 - \text{IoU} \quad (۲)$$

$$\text{IoU} = (\text{Area of Intersection}) / (\text{Area of Union}) \quad (۳)$$

این معیار یکی از رایج‌ترین و اساسی‌ترین مفاهیم برای اندازه‌گیری عملکرد شبکه‌های عصبی کانونوشال است. این روش برای ارزیابی دقیق بودن ماسک‌های پیش‌بینی شده، شباهت بین کادرهای حقیقت مبنای<sup>۳۹</sup> و کادرهای پیش‌بینی شده را کمی‌سازی<sup>۴۰</sup> می‌کند. امتیاز IoU بین ۰ تا ۱ متغیر است، هر چه این دو کادر به هم شبیه‌تر باشند، امتیاز IoU بالاتر است. به منظور به‌روزرسانی پارامترهای مدل در طول آموزش از تابع بهینه‌ساز AdamW استفاده شد. AdamW یک نسخه اصلاح‌شده از الگوریتم Adam است که برای پایداری بیشتر در آموزش مدل‌های بزرگ طراحی شده است.

### ۲-۳ مدل تشخیص شیء

در این پژوهش مدل‌های YOLOv5 [Ultralytics, 2024] و YOLOv7 [WongKinYiu, 2024] به منظور انجام عملیات تشخیص اشیا مقایسه و اعتبارسنجی شده‌اند. الگوریتم YOLO یک شبکه عصبی پیچشی است که عملیات تشخیص شیء را به‌عنوان یک مسئله رگرسیون در نظر می‌گیرد و کادرهای محصورکننده و احتمالات کلاس مرتبط با آن‌ها را از هم جدا می‌کند. در الگوریتم YOLO از شبکه عصبی منفرد<sup>۴۱</sup> برای

نسبت به شبکه‌های دارای نرمال‌ساز دسته‌ای می‌شود. شبکه‌ی PAN+NFNet در آموزش سریع‌تر بود و می‌توانست اندازه دسته‌ای تا ۴۰۹۶ داشته باشد. نرمال‌سازی دسته‌ای تکنیکی است که در یادگیری عمیق برای بهبود عملکرد شبکه‌های عصبی استفاده می‌شود. این تکنیک شامل نرمال‌سازی ورودی‌های هر لایه در شبکه با کم کردن میانگین دسته‌ای و تقسیم بر انحراف از معیار دسته‌ای است. این روش به جلوگیری از مشکل ناپدید شدن یا انفجار گرادیان کمک می‌کند، مشکل گرادیان زمانی رخ می‌دهد که ورودی‌های شبکه خیلی بزرگ یا خیلی کوچک هستند.

### ۳-۱-۱ آموزش و اعتبارسنجی شبکه‌های قطعه‌بندی ریل

در اینجا برای آموزش مدل قطعه‌بندی ریل، از یک مجموعه داده‌ی ویژه استفاده شده است. این مجموعه داده شامل ۸۲۰۳ تصویر رنگی برچسب‌گذاری شده. با وضوح (۱۵۱۲ × ۱۹۲۰ × ۱۰۸۰) است که توسط یک دوربین استاندارد از خطوط ریلی کشور روسیه برداشت شده است [dazzle-me jul 2022]. این مجموعه داده شامل سه دسته برچسب است. برچسب سبز برای قطعه‌بندی مسیری که قطار روی آن در حال حرکت است، برچسب قرمز برای تمام خطوط اطراف و برچسب آبی برای قطارهای موجود در تصویر می‌باشد.



شکل ۴. نمونه داده قطعه‌بندی ریلی

به‌منظور عملیات تشخیص ریل، سه مدل پیشنهادی روی مجموعه داده‌ی مذکور آموزش داده شدند. ۸۵ درصد از مجموعه داده برای آموزش و ۱۰ درصد مجموعه داده برای اعتبارسنجی و ۵ درصد برای تست استفاده شده است. همچنین به‌منظور

(وسیله نقلیه، عابر پیاده، دوچرخه، اتوبوس و علائم ساخت و ساز) است.



شکل ۵. نمونه داده تشخیص موانع برجسب گذاری شده

۷۰ درصد از مجموعه داده برای آموزش، ۲۰ درصد برای اعتبارسنجی و ۱۰ درصد برای آزمایش استفاده شده است. برای اعتبارسنجی شبکه‌های YOLO از شاخص‌های مختلفی استفاده می‌شود. یکی از این شاخص‌ها، نسبت اشتراک به اجتماع<sup>۴۹</sup> (IOU) است. در هر تشخیص، محاسبه امتیاز IoU به این صورت است که مقادیر IoU که بالاتر از آستانه تعریف شده‌اند، پیش‌بینی مثبت و موارد زیر آستانه به‌عنوان پیش‌بینی نادرست در نظر گرفته می‌شوند. به‌طور دقیق‌تر، پیش‌بینی‌ها به‌صورت مثبت واقعی (TP) ۵۰، منفی واقعی (TN) ۵۱، منفی کاذب<sup>۵۲</sup> (FN) و مثبت کاذب (FP)<sup>۵۳</sup> طبقه‌بندی می‌شوند [M. E. Aidouni, 2019]. به‌طور کلی، زمانی که کلاس پیش‌بینی شده با کلاس حقیقت مبنای مطابقت داشته باشد و IoU از یک آستانه معین بیشتر باشد، این پیش‌بینی به‌عنوان مثبت واقعی در نظر گرفته می‌شود. در غیر این صورت، زمانی که کلاس پیش‌بینی با کلاس حقیقی مطابقت نداشته باشد یا IoU کمتر از آستانه باشد، آنگاه به‌عنوان مثبت کاذب در نظر گرفته می‌شود. منفی کاذب به این اشاره دارد که داده حقیقی قابل تشخیص نیست و منفی واقعی نشان‌دهنده یک تشخیص با عنوان صحیح که در حقیقت اشتباه است می‌باشد.

صحت (Precision): صحت میزان درک و دقت یک مدل در تشخیص شیء مربوطه است که به‌عنوان تعداد مثبت‌های واقعی

شناسایی اشیاء درون تصویر و پیش‌بینی کادرهای محصورکننده آن‌ها با احتمالات کلاس مرتبط استفاده می‌شود [Redmon, J., et al. 2016]. این الگوریتم، از تکنیک پنجره کشویی برای عبور منطقه به منطقه<sup>۴۲</sup> تصویر استفاده نمی‌کند. در عوض، یک بار در طول آموزش و آزمایش به کل تصویر نگاه می‌کند تا اطلاعات زمینه‌ای<sup>۴۳</sup> در مورد ظاهر شیء و کلاس را فرموله کند. این باعث می‌شود YOLO بسیار سریع‌تر از سایر الگوریتم‌های تشخیص باشد.

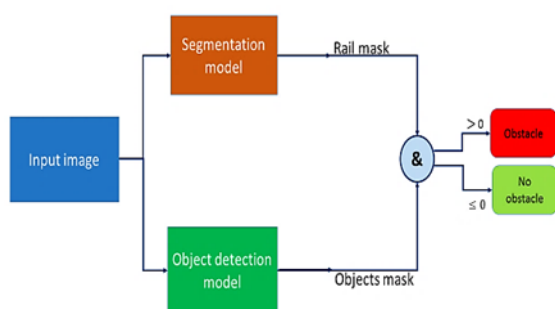
مدل YOLOv5 در پایتورچ پیاده‌سازی شده است. در معماری شبکه‌ی این مدل، قسمت ستون فقرات از نوع CSP<sup>۴۴</sup> و قسمت گردن<sup>۴۵</sup> از نوع PA-NET است. ویژگی‌های اصلی این نسخه، شامل داده افزایشی موزاییکی<sup>۴۶</sup> و یادگیری خودکار کادرهای محصورکننده است. همچنین در این نسخه لایه‌ی توجه به الگوریتم افزوده شده است که به کاهش تعداد پارامترها، تعداد عملیات نقطه شناور (FLOPS)<sup>۴۷</sup> و بهبود حافظه کمک کرده است.

مدل YOLOv7 از نسخه‌ی اصلاح‌شده‌ی معماری نسخه‌ی پنجم استفاده می‌کند که شامل یک شبکه ستون فقرات برای استخراج ویژگی، یک شبکه‌ی گردن برای ترکیب ویژگی‌ها از مقیاس‌های مختلف و یک شبکه سر برای تشخیص و طبقه‌بندی اشیاء است. شبکه‌ی ستون فقرات از معماری CSPResNet50 اصلاح‌شده به‌وسیله‌ی مازولی با نام ادغام هرم فضایی استفاده می‌کند و شبکه اصلی از شبکه‌ی سر مدل نسخه‌ی پنجم YOLO با معماری تغییر یافته به همراه تشخیص‌های بدون لنگر<sup>۴۸</sup> استفاده می‌کند.

### ۳-۲-۱ آموزش و اعتبارسنجی مدل تشخیص شیء

دو نسخه‌ی مختلف شبکه YOLO توسط مجموعه داده‌ای منتشر شده در سایت Roboflow [CV, 2022] تحت آموزش و اعتبارسنجی قرار گرفتند. این مجموعه داده شامل ۹۹۹۵ تصاویر برداشت‌شده با دوربین رنگی در ۵ کلاس متفاوت

## نظارت تصویری بر موانع حمل و نقل ریلی با استفاده از روش‌های مبتنی بر یادگیری عمیق



شکل ۶. ساختار شبکه تشخیص موانع

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

### ۴. تحلیل داده‌ها

در این بخش به بررسی نتایج اعتبارسنجی مدل‌های قطعه‌بندی ریل PAN+NF\_Net، PAN+Efficient\_Net و PAN+SE\_ResNet در قطعه‌بندی ریل و بررسی نتایج اعتبارسنجی مدل‌های YOLOv7 و YOLOv5 در تشخیص اشیاء و آزمایش مدل نهایی پرداخته شده است.

#### ۴-۱ نتیجه اعتبارسنجی مدل‌های قطعه‌بندی ریل

نتایج آموزش مدل‌های قطعه‌بندی ریل روی مجموعه داده‌ی مربوطه در جدول ۱ گزارش شده است. پردازنده مورد استفاده در این پروژه NVIDIA Tesla P100 می‌باشد. با توجه به جدول ۱ مشخص می‌شود که بار محاسباتی مدل‌های قطعه‌بندی مورد بررسی، اختلافات قابل‌توجهی دارند. به طوری که تعداد عملیات نقطه‌شناور (FLOP) مدل‌های PAN+SE\_resNet و PAN+NF\_Net دو و شش برابر تعداد FLOP مدل PAN+EfficientNet است. این در حالی است که، دقت ارائه شده توسط این مدل‌ها مقداری نزدیک به هم دارند؛ در نتیجه با توجه به موارد ذکر شده و همچنین سرعت پردازش بالاتری که مدل PAN+EfficientNet ارائه می‌دهد می‌توان به این نتیجه رسید که مدل قطعه‌بندی PAN+EfficientNet برای عملیات مدنظر در این پروژه مناسب‌تر است.

تقسیم بر مجموع مثبت‌های واقعی و مثبت‌های کاذب تعریف می‌شود.

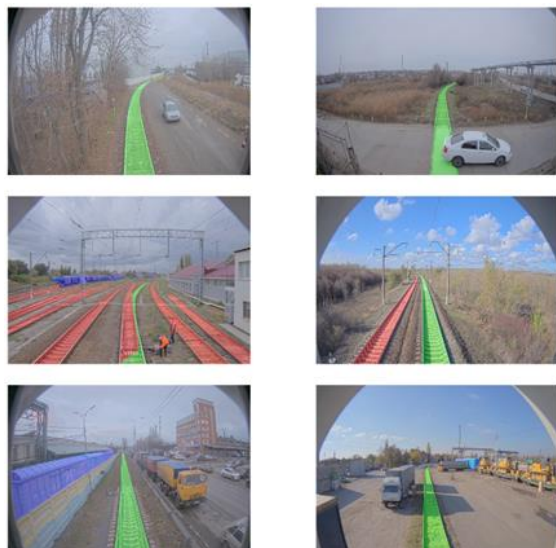
**حساسیت (Recall):** حساسیت توانایی یک مدل برای شناسایی تمام اشیاء مربوطه است که به عنوان تعداد مثبت‌های واقعی تقسیم بر مجموع مثبت‌های واقعی و منفی‌های کاذب (همه حقایق پایه) تعریف می‌شود.

با تعیین آستانه امتیاز اطمینان در سطوح مختلف، می‌توان جفت‌های مختلف دقت و حساسیت را به دست آورد و سپس منحنی دقت - حساسیت (P-R) را ترسیم کرد، که نشان‌دهنده‌ی رابطه بین این دو معیار است و به قضاوت در مورد عملکرد یک شاخص کمک می‌کند. همیشه مقایسه مستقیم بین آشکارسازهای مختلف زمانی که منحنی‌ها با یکدیگر تلاقی می‌کنند آسان نیست. بنابراین، دقت متوسط (AP)<sup>۵۴</sup> یک متریک عددی، که مساحت زیر منحنی P-R را محاسبه می‌کند، اتخاذ می‌شود.

### ۳-۳ مدل تشخیص موانع

در روش پیشنهادی ما موانع با بررسی وجود همپوشانی میان ماسک مسیر حرکت قطار و کادرهای محصورکننده اشیاء صورت می‌گیرد. به منظور بررسی وجود همپوشانی ابتدا ماسک قرمز و آبی از ماسک پیش‌بینی شده توسط مدل قطعه‌بندی ریل باید حذف شوند تا ماسکی که تنها شامل کلاس سبز یعنی مسیر حرکت قطار است تولید شود. ماسک حاصل در مقیاس خاکستری ذخیره می‌شود. همچنین ماسک کادرهای محصورکننده اشیاء با اختصاص دادن مقدار یک به پیکسل‌های درون کادر کلاس‌های معین و اختصاص دادن مقدار صفر به سایر پیکسل‌ها تولید می‌شود. در نهایت با ضرب پیکسل به پیکسل ماسک مسیر حرکت قطار در ماسک اشیاء و بررسی مقدار خروجی وجود همپوشانی و حضور مانع در مسیر حرکت قطار بررسی می‌شود. خروجی بزرگ‌تر از صفر به معنای همپوشانی ماسک‌ها یا به عبارتی حضور مانع در مسیر حرکت قطار است و خروجی کوچک‌تر از صفر به معنای عدم همپوشانی یا عدم حضور مانع در مسیر حرکت قطار است.

ویژگی‌های تصویر در نظر گرفته شده باشد و مجموعه داده مناسب برای آموزش مدل فراهم شود. مجموعه داده با تعداد برچسب متناسب برای هر کلاس نقش اساسی در دقت نهایی آموزش دارد.



شکل ۷. نمونه عملکرد مدل PAN+EfficientNet روی مجموعه

داده آزمایش

جدول ۱. نتایج اعتبارسنجی مدل‌های قطعه‌بندی ریل

PAN+SE\_Resnet, PAN+EfficientNet, PAN+NfNet

| مدل          | زمان آموزش (ساعت) | MIOU (%) | FLOP (billion) | fps  |
|--------------|-------------------|----------|----------------|------|
| +PAN         |                   |          |                |      |
| SE_Resnet    | ۱۳/۵              | ۸۵/۸     | ۳۳/۵۲          | ۷/۳۵ |
| EfficientNet | ۹                 | ۸۵/۷     | ۱۷/۸۴          | ۸/۳۳ |
| NF_Net       | ۱۸/۷۵             | ۸۶/۵     | ۱۰۲/۲۲         | ۷/۱  |

۲-۴ نتیجه اعتبارسنجی مدل تشخیص شی

نتایج ثبت شده در جدول ۲ نشان می‌دهند که عملکرد YOLOv7 روی مجموعه داده تشخیص شی به میزان قابل توجهی نسبت به YOLOv5 دقیق‌تر است. معماری و وضوح بالایی که YOLOv7 فراهم می‌کند باعث شده این مدل قادر باشد اشیاء کوچک‌تر را شناسایی کند و در کل دقت بالاتری نسبت به YOLOv5 داشته باشد.

در آموزش مدل یادگیری عمیق دقت مؤثر زمانی صورت می‌گیرد که تعداد لایه‌های پیش‌بینی و معماری مناسب برای استخراج

جدول ۲. نتایج اعتبارسنجی شبکه‌های YOLOV5 و YOLOV7

| YOLOV5.s                                                                   |        |        |       |       |         |        |                 |
|----------------------------------------------------------------------------|--------|--------|-------|-------|---------|--------|-----------------|
| custom_YOLOv5s summary: 232 layers, 7257306 parameters, 16.8 GFLOPs        |        |        |       |       |         |        |                 |
| anchors/target = 3.90, Best Possible Recall (BPR) = 0.985 conf_thres=0.4   |        |        |       |       |         |        |                 |
| Class                                                                      | Images | Labels | P     | R     | F1score | mAP@.5 | mAP@.5:.95:100% |
| all                                                                        | 1999   | 24833  | 0.744 | 0.588 | 0.656   | 0.656  | 0.369           |
| bus                                                                        | 1999   | 281    | 0.766 | 0.683 | 0.722   | 0.756  | 0.509           |
| construction                                                               | 1999   | 2637   | 0.722 | 0.438 | 0.545   | 0.5    | 0.222           |
| cyclist                                                                    | 1999   | 221    | 0.611 | 0.529 | 0.567   | 0.558  | 0.312           |
| pedestrian                                                                 | 1999   | 3304   | 0.776 | 0.526 | 0.627   | 0.628  | 0.275           |
| vehicle                                                                    | 1999   | 18390  | 0.847 | 0.764 | 0.795   | 0.838  | 0.529           |
| YOLOV7 v0.1                                                                |        |        |       |       |         |        |                 |
| Model Summary: 415 layers, 37218132 parameter, 105.2 GFLOPs                |        |        |       |       |         |        |                 |
| anchors/target = 4.37, Best Possible Recall (BPR) = 0.9951, conf_thres=0.1 |        |        |       |       |         |        |                 |
| Class                                                                      | Images | Labels | P     | R     | F1score | mAP@.5 | mAP@.5:.95:100% |
| all                                                                        | 1999   | 24833  | 0.837 | 0.709 | 0.7686  | 0.832  | 0.48            |
| bus                                                                        | 1999   | 281    | 0.851 | 0.819 | 0.8346  | 0.87   | 0.658           |
| construction                                                               | 1999   | 2637   | 0.798 | 0.55  | 0.6511  | 0.671  | 0.299           |
| cyclist                                                                    | 1999   | 221    | 0.775 | 0.683 | 0.726   | 0.735  | 0.452           |
| pedestrian                                                                 | 1999   | 3304   | 0.855 | 0.653 | 0.74    | 0.8    | 0.387           |
| vehicle                                                                    | 1999   | 18390  | 0.904 | 0.842 | 0.872   | 0.923  | 0.605           |

#### ۳-۴ تشریح نتایج روش مدل تشخیص موانع

در روش تشخیص موانع پیشنهادی موانع احتمالی به وسیله بررسی وجود همپوشانی میان ماسک مسیر حرکت قطار و ماسک کادرهای پیش بینی شده صورت می گیرد. برای بررسی همپوشانی از عملگر ضرب پیکسل به پیکسل استفاده شد و دقت شبکه نهایی تشخیص موانع به طور کامل وابسته به دقت مدل های قطعه بندی ریل و تشخیص اشیاء است که با چه دقتی ریل و کلاس های اشیاء را شناسایی می کنند.

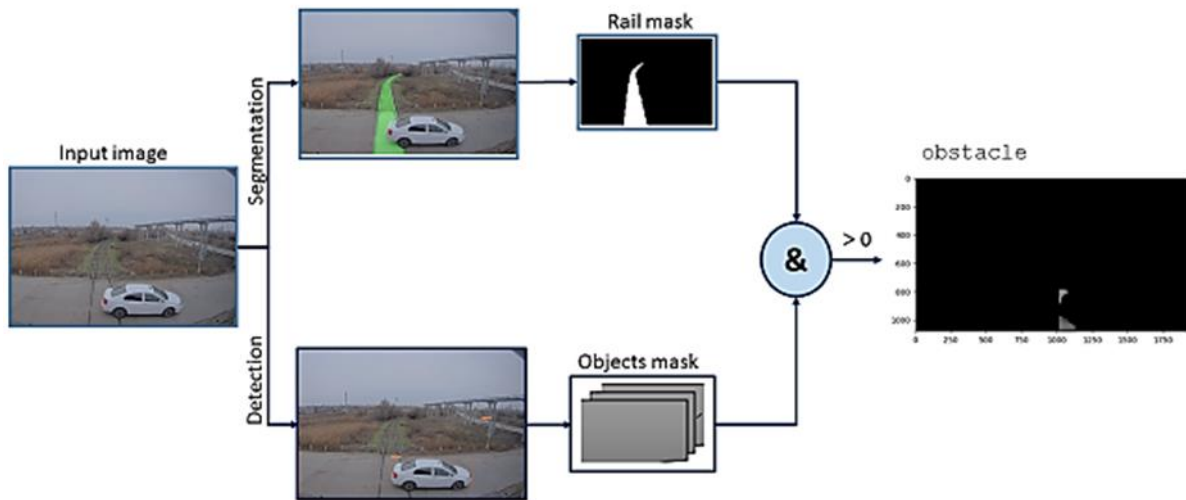
مدل تشخیص ریل PAN+EfficientNet ریل را با دقت ۸۵/۷ درصد و مدل YOLOv7 کلاس های اشیاء را با دقت میانگین ۸۳/۲ درصد تشخیص می دهد. سرعت پردازش هر تصویر توسط مدل PAN+EfficientNet ۱۲۰ میلی ثانیه و سرعت پردازش هر فریم تصویر توسط مدل YOLOv7 ۱۳ میلی ثانیه است. در ادامه نمونه ای از فرایند عملکرد شبکه پیشنهادی در تشخیص موانع آورده شده است.

مدل طراحی شده با موفقیت موانع موجود در مسیر حرکت قطار را تشخیص می دهد. زمان پردازش هر فریم تصویر ورودی توسط شبکه با استفاده از روش پردازش موازی روی پردازنده NVIDIA Tesla p100، ۱۲۵ میلی ثانیه یعنی ۰,۷۸ درصد زمان توقف اضطراری یک قطار معمولی با سرعت ۷۰ کیلومتر بر ساعت می باشد.

به عنوان مثال با مقایسه نتایج ثبت شده در جدول می توان مشاهده کرد که، الگوریتم YOLO کلاس هایی که از پیچیدگی ظاهری کمتری برخوردارند را می تواند با تعداد کمتری برچسب بیاموزد، مانند کلاس اتوبوس. اما همین مدل برای تشخیص انسان به عنوان یک شیء کوچک با پیچیدگی ظاهری فراوان به برچسب های بیشتر و متنوع تری نیاز دارد. به این معنا که مجموعه داده آموزشی برای هر کلاس خاص باید شامل تصاویری، با مقیاس، زاویه چرخش، نورپردازی، پس زمینه و جهات متفاوت باشد. به طور کلی برای افزایش دقت یادگیری باید برای هر کلاس با توجه به پیچیدگی آن کلاس، تعداد مناسبی برچسب در نظر گرفته شود. در مجموعه داده ی تشخیص شیء استفاده شده در این پروژه تعداد و تنوع برچسب های در نظر گرفته شده برای تشخیص کلاس علائم ساخت و ساز، دوچرخه سوار مناسب نیست و همین امر باعث کاهش دقت مدل در یادگیری این کلاس ها نسبت به سه کلاس دیگر و کاهش دقت کلی شده است. قابل ذکر است که اگرچه دقت مدل در تشخیص این دو کلاس بالا نیست اما برچسب داشتن این کلاس ها باعث شده دقت مدل در تشخیص کلاس های دیگر بالاتر برود. مدل آموزش دیده درجه اطمینان مناسبی را برای تشخیص کلاس های وسایل نقلیه عابر پیاده و اتوبوس فراهم می کند.



شکل ۸. نمونه عملکرد مدل YOLOv7 روی داده آزمایشی



شکل ۹. نمونه خروجی شبکه تشخیص موانع روی داده آزمایش.

استفاده از یک مجموعه داده، آموزش دیده و اعتبارسنجی شدند. نتایج نشان داد که مدل YOLOv7 نسبت به مدل YOLOv5 دقیق‌تر است و YOLOv5 در تشخیص اشیاء کوچک با پیچیدگی ظاهری فراوان دچار خطا می‌شود و دقت مناسبی ندارد. این در حالی است که معماری و وضوح بالایی که مدل YOLOv7 فراهم می‌کند باعث شده این مدل قادر باشد اشیاء کوچک‌تر را شناسایی کند و دقت بالاتری نسبت به YOLOv5 ارائه دهد.

در نهایت مدل تشخیص موانع پیشنهادی با بهره‌گیری از مدل قطعه‌بندی ریل PAN+EfficientNet و مدل تشخیص شیء YOLOv7 طراحی و پیاده‌سازی شد. در روش پیشنهادی تشخیص موانع ریلی با بررسی وجود همپوشانی میان ماسک مسیر حرکت قطار و کادرهای محصورکننده اشیاء موجود در تصویر، موانع موجود روی مسیر حرکت قطار تشخیص داده می‌شوند. خروجی روش پیشنهادی به دو کلاس بدون مانع و با مانع تقسیم می‌شود.

## ۶. پیشنهادات

برای توسعه روش تشخیص موانع پیشنهادی موارد زیر پیشنهاد می‌شوند:

## ۵. نتیجه‌گیری

در این مقاله، به طراحی یک الگوریتم نظارت تصویری بر موانع حمل‌ونقل ریلی با استفاده از روش‌های مبتنی بر هوش مصنوعی و یادگیری عمیق پرداخته شده است. پس از مطالعه‌ی پژوهش‌های پیشین و انتخاب دوربین تصویربرداری به‌عنوان حسگر پایه برای حل مسئله تشخیص موانع ریلی، طراحی الگوریتمی جهت استخراج اطلاعات مفید از تصویر نظیر قطعه‌بندی مسیر حرکت قطار و تشخیص اشیاء مبنای کار قرار گرفت. به این منظور سه مدل PAN+Efficient\_Net، PAN+SE\_ResNet و PAN+NF\_Net به منظور قطعه‌بندی ریل و دو مدل YOLOv7 و YOLOv5 به منظور تشخیص اشیاء از نظر دقت و بار محاسباتی برای مصارف زمان حقیقی در محیط ریلی مقایسه و اعتبارسنجی شدند.

نتایج آموزش و اعتبارسنجی مدل‌های قطعه‌بندی ریل نشان داد که مدل قطعه‌بندی PAN+NF-Net با چند دهم درصد اختلاف از دو مدل دیگر دقیق‌تر است. این در حالی است که بار

محاسباتی مدل PAN+EfficientNet حدوداً  $\frac{1}{5}$  و  $\frac{1}{4}$  برابر بار محاسباتی دو مدل دیگر است.

در بخش تشخیص اشیاء مدل‌های YOLOv7 و YOLOv5

36. Data augmentation
37. Intersection over Union
38. Jaccard index
39. Ground truth
40. Quantifying
41. Single neural network
42. Region by region
43. Contextual
44. Cross Stage Partial Network
45. Neck
46. Mosaic data augmentation
47. Total number of floating point operations
48. Anchor-free detection
49. Intersection Over Union
50. True positive
51. True Negative
52. False negative
53. Fuls positive
54. Average precision

- استفاده از روش‌های ادغام حسگرها برای دستیابی به داده‌ی عمق و فاصله سنجی.
- جمع آوری مجموعه داده از خطوط ریلی کشور برای آموزش مدل‌های یادگیری عمیق.
- استفاده از دوربین استریو و تولید نقشه‌ی سه‌بعدی خطوط ریلی

## ۷. پی‌نوشت‌ها

1. Obstacle Detection
2. Computer vision
3. Artificial intelligence
4. Segmentation
5. Edge detection
6. Corner detection
7. Threshold segmentation
8. Hand-crafted
9. Machine Learning
10. Deep Learning
11. End to End
12. Non hand-crafted
13. Rail Segmentation
14. Object Detection
15. Pyramid Attention Network
16. Encoder
17. Squeeze-and-Excitation Residual Network
18. Efficient Network
19. Normalizer free network
20. Segmentation
21. Convolutional Neural Network
22. Attention
23. Spatial pyramid pooling
24. Backbone
25. Feature Enhanced Single Shot Detector
26. Pooling layers
27. Fully connected layers
28. Max-pooling layer
29. Average-pooling layer
30. Decoder
31. Segmentation head
32. Pyramid blocks
33. Attention blocks
34. Upsampling layers
35. Attention map

## ۸. مراجع

- Brock, A., et al. (2021). High performance large-scale image recognition without normalization. International Conference on Machine Learning. PMLR.
- Belyaev, S., Popov, I., Shubnikov, V., Popov, P., Boltenkova, E., & Savchuk, D. (2020, September). Railroad semantic segmentation on high-resolution images. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1-6). IEEE.
- Dazzle-me jul 2022 [Online]. Available: <https://github.com/dazzle-me/digital-breakthrough-hack#download-data>.
- Gebauer, W. Pree, and B. Stadlmann. (2018). Autonomously Driving Trains on Open Tracks—Concepts, System Architecture and Implementation Aspects. IEEE Intelligent Transportation Systems Magazine, 10(1), 58-70.
- Gleichauf, J., Vollet, J., Pfitzner, C., Koch, P., & May, S. (2020). Sensor Fusion Approach for

- recognition, 71, 158-172.
- Pavel Iakubovskii 2023. [Online]. Available:<https://smp.readthedocs.io/en/latest/index.html>
  - Ristić-Durrant, D., Haseeb, M.A., Franke, M., Banić, M., Simonović, M., & Stamenković, D. (2020). Artificial Intelligence for Obstacle Detection in Railways: Project SMART and Beyond. In S. Bernardi, F. Flammini, S. Marrone, D. Schneider, N. Nostro, A. Di Salle, V. Vittorini, R. Nardone, R. Adler, P. Schleiβ, et al. (Eds.), *Dependable Computing – EDCC 2020* (pp. 44–55). Springer.
  - Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
  - Redmon, J., & Farhadi, A. (2019). YOLOv3: An Incremental Improvement. arXiv. Retrieved from <https://arxiv.org/pdf/1804.02767.pdf>
  - Ristić-Durrant, D., Franke, M., & Michels, K. (2021). A review of vision-based on-board obstacle detection and distance estimation in railways. *Sensors*, 21(10), 3452.
  - Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning*. PMLR.
  - Wang, H., Zhang, X., Damiani, L., Giribone, P., Revetria, R., & Ronchetti, G. (2018). Transportation Safety Improvements Through Video Analysis: An Application of Obstacles and Collision Detection Applied to Railways and Roads. In S.I. Ao, H. Kim, O. Castillo, A.S. Chan, & H. Katagiri (Eds.), *Transactions on Engineering Technologies* (pp. 1–15). Springer.
  - an Autonomous Shunting Locomotive. In O. Gusikhin & K. Madani (Eds.), *Informatics in Control, Automation and Robotics* (pp. 603-624). Cham, Switzerland: Publishing House.
  - Haseeb, M.A.; Guan, J.; Ristić-Durrant, D.; Gräser, A. A Novel Method for Distance Estimation from Monocular Camera. In *Proceedings of the 10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18), IROS, Madrid, Spain, 1 October 2018; Volume 10.*
  - Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA* (pp. 7132-7141). doi: 10.1109/CVPR.2018.00745.
  - Li, J.; Zhou, F.; Ye, T. Real-world railway traffic detection based on faster better network. *IEEE Access* 2018, 6, 68730–68739.
  - Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects - in ContextN. In *Computer Vision—ECCV 2014, Part V; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2020; pp. 740–755. [CrossRef]*
  - Liu, Y. H. (2018). Feature extraction and image recognition with convolutional neural networks. In *Journal of Physics: Conference Series* (Vol. 1087, p. 062032). IOP Publishing.
  - M. E. Aidouni, October 2019.[Online]. Available:<https://manalaidouni.github.io/manalaidouni.github.io/Evaluating-Object-Detection-Models-Guide-to-Performance-Metrics.html>.
  - Nanni, L., Ghidoni, S., & Brahmam, S. (2017). Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern*

- Wang, Y., Wang, L., Hu, Y.H., & Qiu, J. (2019). RailNet: A Segmentation Network for Railroad Detection. *IEEE Access*, 7, 143772–143779.

- Wang, Z., Wu, X., Yu, G., & Li, M. (2018). Efficient Rail Area Detection Using Convolutional Neural Network. *IEEE Access*, 6, 77656–77664.

- Xu, Y., Gao, C., Yuan, L., Tang, S., & Wei, G. (2019). Real-time Obstacle Detection Over Rails Using Deep Convolutional Neural Network. In *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 27–30 October 2019 (pp. 1007–1012).

- Ye, T., Wang, B., Song, P., & Li, J. (2018). Automatic Railway Traffic Object Detection System Using Feature Fusion Refine Neural Network under Shunting Mode. *Sensors*, 18, 1916.

- Ye, T., Wang, B., Song, P., & Li, J. (2020). Railway Traffic Object Detection Using Differential Feature Fusion Convolution Neural Network. *IEEE Transactions on Intelligent Transportation Systems*.

- Ye, T., Zhang, Z., Zhang, X., & Zhou, F. (2020). Autonomous railway traffic object detection using feature-enhanced single-shot detector. *IEEE Access*, 8, 145182-145193.

```
- @misc{ tesi_dataset,title = { tesi Dataset
},type = { Open Source
- Dataset },author = { CV 2022 },howpublished
= { \url{ https://universe.roboflow.com/cv-
2022-kyjj6/tesi } },url = {
https://universe.roboflow.com/cv-2022-
kyjj6/tesi },journal = { Roboflow Universe
},publisher = { Roboflow },year = { 2022
},month = { nov }}
```

## سیمین کولائیان، محمدعلی صندیدزاده

سیمین کولائیان، درجه‌ی کارشناسی را در رشته‌ی مهندسی برق کنترل در سال ۱۳۹۹ از دانشگاه صنعتی نوشیروانی بابل و درجه‌ی کارشناسی ارشد را در رشته‌ی مهندسی کنترل و علائم در سال ۱۴۰۲ از دانشگاه علم و صنعت ایران اخذ نمود. زمینه‌های پژوهشی مورد علاقه‌ی ایشان کنترل و سیگنالینگ راه‌آهن و خودران‌سازی قطار است.



محمدعلی صندیدزاده، درجه‌ی کارشناسی را در رشته‌ی مهندسی برق الکترونیک در سال ۱۳۷۰ از دانشگاه تهران و درجه‌ی کارشناسی ارشد را در رشته‌ی مهندسی برق کنترل در سال ۱۳۷۲ از همان دانشگاه اخذ نمود. در سال ۱۳۷۹ موفق به اخذ درجه‌ی دکتری در رشته‌ی مهندسی برق کنترل، از دانشگاه صنعتی امیرکبیر گردید. زمینه‌های مورد علاقه‌ی ایشان کنترل و سیگنالینگ راه‌آهن و قطار بوده و در حال حاضر عضو هیات علمی با رتبه دانشیار در دانشگاه علم و صنعت ایران است.

