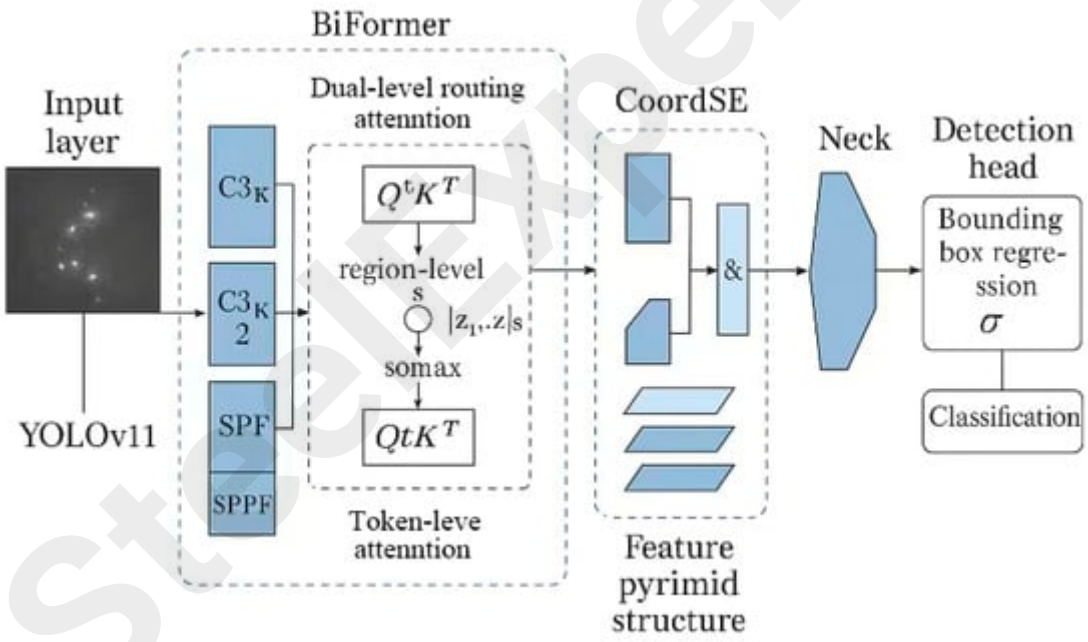
مرکز دانش شبکه خبرگان صنعت فولاد

بهبود YOLOv11 با BiFormer و CoordSE برای تشخیص بلادرنگ ذرات سرباره در ریخته‌گری پیوسته فولاد

مقالات تخصصی | علی ارجمندی اصل | ۱۴۰۵/۰۸/۰۵

مدل بهبودیافته YOLOv11 با BiFormer و CoordSE برای تشخیص ذرات سرباره در ریخته‌گری پیوسته: نتایج ۲۸۴۷ تصویر صنعتی، Precision برابر ۸۲٫۵٪، Recall برابر ۶۹٫۱٪ و mAP@0.5 برابر ۸۰٫۶٪، همراه محدودیت‌های استقرار.



ترجمه و تنظیم فارسی شبکه خبرگان صنعت فولاد | نویسندگان: Binhong Li، Pengfei Cheng و Xichen Liu | *Applied Sciences*، ۲۰۲۶، جلد ۱۶، مقاله ۳۹۶۵ | منبع و DOI | مجوز CC BY 4.0.

نکته اجرایی: سرعت ۴۵٫۲ فریم‌برثانیه فقط روی RTX 4090 ثبت شده است. نرخ بازیابی ۶۹٫۱ درصد نشان می‌دهد برخی ذرات تشخیص داده نمی‌شوند. اعتبارسنجی لبه‌ای و صنعتی هنوز انجام نشده است.

چکیده

پایش دقیق افزودن سرباره در فرایند ریخته‌گری پیوسته فولاد اهمیت زیادی دارد، اما شرایط خشن کارخانه تشخیص ذرات کوچک را دشوار می‌کند. در این پژوهش، چارچوبی مبتنی بر یادگیری عمیق پیشنهاد شده است که دو سازوکار توجه BiFormer و CoordSE را به معماری YOLOv11 اضافه می‌کند. BiFormer با مسیریابی توجه در دو سطح ناحیه و توکن، ویژگی‌های ذرات کوچک سرباره را از پس‌زمینه پیچیده فولاد مذاب جدا می‌کند؛ در مقابل، CoordSE با ترکیب تجمیع جهت‌دار و لایه‌های تمام‌متصل چندشاخه، وابستگی‌های مکانی و کانالی را حفظ می‌کند. مدل نهایی روی مجموعه‌ای اختصاصی شامل ۲۸۴۷ تصویر صنعتی با وضوح بالا آزمایش شد و به دقت پیش‌بینی ۸۲٫۵ ± ۰٫۲ درصد، بازیابی ۶۹٫۱ ± ۰٫۳ درصد و mAP@0.5 برابر ۸۰٫۶ ± ۰٫۲ درصد رسید. در آزمایش حذف مؤلفه‌ها، BiFormer به‌تنهایی mAP@0.5 را نسبت به مدل پایه ۲۳٫۴ واحد درصد بهبود داد و ترکیب کامل نسبت به نسخه صرفاً BiFormer بهبود دیگری معادل ۱۲٫۳ واحد درصد ایجاد کرد. سرعت استنتاج روی پردازنده گرافیکی RTX 4090 حدود ۴۵٫۲ فریم‌برثانیه بود. بااین‌حال، نرخ بازیابی ۶۹٫۱ درصد و نبود آزمون عملی روی رایانه‌های لبه‌ای کم‌مصرف، محدودیت‌هایی مهم پیش از استقرار صنعتی گسترده به‌شمار می‌روند.

۱. مقدمه

صنعت فولاد از ارکان اصلی تولید صنعتی است و ریخته‌گری پیوسته، به‌عنوان روش غالب تبدیل فولاد مذاب به محصولات نیمه‌ساخته، با حذف برخی مراحل میانی، بهره‌وری تولید را افزایش داده است. بااین‌حال، پیچیدگی عملیات ریخته‌گری، به‌ویژه مرحله افزودن سرباره، نظارت بر فرایند و کنترل کیفیت محصول را دشوار می‌کند. [1]

پایش قابل‌اعتماد در فولادسازی می‌تواند از روش‌هایی مانند تحلیل ارتعاش مبتنی بر شتاب‌سنج نیز بهره بگیرد. سرباره افزوده‌شده به فرایند ریخته‌گری پیوسته، نقش‌هایی همچون عایق‌کاری حرارتی فولاد مذاب، جذب ناخالصی‌های غیرفلزی و جلوگیری از اکسیداسیون دارد. توزیع و ضخامت لایه سرباره بر کیفیت سطح، پاکیزگی داخلی و خواص مکانیکی فولاد اثر می‌گذارد. روش‌های قدیمی بینایی ماشین، مانند تحلیل لکه، آستانه‌گذاری ریخت‌شناختی و بازسازی عناصر ساختاری، در محیط‌های تصویری پایدار مفیدند؛ اما تغییر روشنایی، دمای بسیار بالا و شکل نامنظم اهداف در ریخته‌گری پیوسته، توان این روش‌ها را محدود می‌کند. به همین علت، روش‌های یادگیری عمیق مورد توجه قرار گرفته‌اند. [2]، [3]، [4]، [5]، [6]

یادگیری عمیق در پایش تصویری فرایندهای متالورژیکی پیشرفت قابل‌توجهی ایجاد کرده است. پژوهشگران با سازوکارهای توجه ویژه، استخراج ویژگی از بافت‌های پیچیده سطح فولاد را تقویت کرده‌اند؛ برای نمونه، ماژول «برانگیختگی تجمیع‌شده فشرده» برای تشخیص عیوب سطحی طراحی شده است. بااین‌حال، تابش حرارتی

شدید ممکن است اطلاعات مختصات مکانی ذرات بسیار کوچک سرباره را مخدوش کند. بنابراین، به سازوکاری نیاز است که موقعیت هدف را دقیق‌تر حفظ کند. [7]

محیط عملیاتی ماشین ریخته‌گری پیوسته ویژگی‌های دشواری دارد: دمایی بالاتر از ۱۶۰۰ درجه سلسیوس، تداخل الکترومغناطیسی شدید، گردوغبار و بخار. چنین عواملی کارایی حسگرهای متعارف را کاهش می‌دهند. حرکت پیوسته سطح مذاب و اندازه کوچک ذرات سرباره نیز تشخیص مطمئن آن‌ها را مشکل‌تر می‌کند. [3]، [8]

پیشرفت بینایی ماشین و یادگیری عمیق امکان پایش بلادرنگ فرایندهای صنعتی را افزایش داده است. خانواده YOLO که نام آن از عبارت «فقط یک‌بار نگاه می‌کنی» گرفته شده، به سبب اجرای سریع تشخیص اشیا شناخته می‌شود. نسخه‌های پیاپی آن، از جمله YOLOv7 و YOLOv11، با تغییر در استخراج ویژگی و تشخیص اجسام کوچک، بهبودهایی در کارایی و دقت ارائه کرده‌اند. [9]، [10]، [11]، [12]، [13]

به‌کارگیری مستقیم آشکارسازهای عمومی در عملیات متالورژیکی با محدودیت مواجه است. بازتاب‌های پیچیده سطح فولاد مذاب، اندازه کم و ظاهر متغیر ذرات سرباره، استخراج ویژگی در شبکه‌های کانولوشنی متداول را دشوار می‌کنند. علاوه‌براین، الزام به تصمیم‌گیری در زمان واقعی اجازه نمی‌دهد دقت را بدون توجه به هزینه محاسباتی افزایش دهیم. [14]، [15]

سازوکارهای توجه به شبکه عصبی کمک می‌کنند بخش‌های مرتبط تصویر را برجسته و عوامل مزاحم پس‌زمینه را تضعیف کند. BiFormer از ساختار توجه با مسیریابی دوسطحی استفاده می‌کند: ابتدا نواحی مهم‌تر انتخاب می‌شوند و سپس ارتباطات دقیق‌تر در سطح اجزای تصویری یا توکن‌ها بررسی می‌گردد. نتیجه، کاهش پردازش غیرضروری همراه با حفظ جزئیات ظریف تصویر است. [16]، [17]، [18]

ترکیب توجه مکانی و توجه کانالی برای کاربردهای صنعتی ارزش ویژه‌ای دارد؛ زیرا تشخیص درست هم به مکان دقیق هدف وابسته است و هم به توان جداسازی ویژگی‌های آن. توجه مختصاتی، اطلاعات موقعیت را که در تجمیع سراسری معمولاً از دست می‌رود حفظ می‌کند و شبکه‌های Squeeze-and-Excitation وزن کانال‌های ویژگی را دوباره تنظیم می‌کنند. ترکیب این دو رویکرد می‌تواند در صحنه‌های پیچیده دقت را بالا ببرد. [19]، [20]، [21]، [22]، [23]، [24]

پژوهش حاضر چهار دستاورد را دنبال می‌کند: نخست، افزودن توجه دوسطحی BiFormer به YOLOv11 برای استخراج بهتر ویژگی اجسام کوچک؛ دوم، معرفی ماژول CoordSE با تلفیق توجه مختصاتی و SENetV2 تقویت‌شده؛ سوم، ارزیابی تجربی مدل در مقایسه با الگوریتم‌های شناخته‌شده؛ و چهارم، گردآوری یک مجموعه‌داده تخصصی برای پایش افزودن سرباره در ریخته‌گری پیوسته.

اهمیت عملی این کار فقط در طراحی یک معماری یادگیری عمیق نیست. اگر آشکارسازی در شرایط واقعی قابل

اعتماد شود، می‌تواند ورودی سامانه‌های پایش و کنترل کیفیت ریخته‌گری باشد و به تصمیم‌گیری سریع‌تر درباره وضعیت سرباره کمک کند. البته تبدیل این نتیجه آزمایشگاهی به یک سامانه عملیاتی، به اعتبارسنجی تکمیلی نیاز دارد. [25]، [26]

۲. پیشینه و پژوهش‌های مرتبط

۲.۱. تشخیص اشیا در کاربردهای صنعتی

استفاده از بینایی ماشین برای پایش صنعتی همراه با پیشرفت پردازنده‌ها و الگوریتم‌های یادگیری عمیق گسترش یافته است. اما سامانه‌ای که در محیط کارخانه کار می‌کند با محدودیت‌هایی فراتر از کاربردهای عمومی روبه‌رو است؛ از جمله آلودگی تصویری، نیاز به دوربین و فیلتر تخصصی و محدودیت سخت‌گیرانه زمان پاسخ. [4]، [14]، [8]

نسل نخست سامانه‌های بینایی صنعتی بیشتر به تشخیص لبه، عملیات مورفولوژیکی و تطبیق الگو متکی بود. این روش‌ها برای اجسام با شکل و روشنایی تقریباً ثابت مناسب بودند، ولی در مواجهه با تغییر نور، پس‌زمینه متحرک و اهداف نامنظم، به‌خصوص در محیط‌های فولادسازی، دقت کافی نداشتند. [6]، [27]

شبکه‌های عصبی کانولوشنی یا CNN تحولی مهم به وجود آوردند: به‌جای طراحی دستی ویژگی‌ها، خود شبکه الگوهای لازم را از داده‌ها فرا می‌گیرد. بااین‌حال، انتقال مستقیم معماری‌های عمومی CNN به کارخانه همواره موفق نیست؛ زیرا برخی ویژگی‌های تصویری ویژه و محدودیت زمان پردازش در فرایند صنعتی در طراحی اولیه لحاظ نشده‌اند. [9]، [8]

پژوهش‌های جدید بر معماری‌های تخصصی و شیوه‌های آموزشی متناسب با صنعت متمرکز شده‌اند. کمبود نمونه‌های آموزشی، اختلاف میان تصاویر آزمایشگاه و خط واقعی تولید و نیاز به اتصال خروجی بینایی ماشین به زیرساخت کنترل صنعتی از چالش‌های مهم این مسیر هستند. [14]

مدل‌های استاندارد YOLOv5 تا YOLOv10 در محیط پویای متالورژیکی ممکن است در شناسایی لبه‌های ذرات ریز سرباره دچار افت ویژگی شوند؛ زیرا تابش دمای بالا و تغییرات پس‌زمینه، علامت تصویری هدف را کم‌رنگ می‌کند. [14]

۲.۲. سیر تکامل YOLO و کاربردهای صنعتی

خانواده YOLO یکی از رویکردهای اصلی تشخیص بلادرنگ اشیاست. ایده محوری نسخه نخست آن، تشخیص تک‌مرحله‌ای بود: پیشنهاد نواحی و طبقه‌بندی اشیا در دو فرایند جداگانه انجام نمی‌شود، بلکه شبکه در یک گذر، مکان و نوع اهداف را تخمین می‌زند. نسخه‌های بعدی دقت و سرعت را با اصلاح معماری افزایش دادند. [10]، [11]، [28]، [9]

در YOLOv5 از شبکه‌های CSP و تجمیع هرمی فضایی SPP برای استخراج بهتر ویژگی بدون افزایش شدید محاسبات استفاده شد. YOLOv8 نیز با سرهای آشکارسازی بدون لنگر و روش‌های تقویت داده، کیفیت مدل را ارتقا داد. [10]، [11]

در YOLOv11 بلوک‌های C3K2 به استخراج ویژگی کمک می‌کنند، بخش SPPF ویژگی‌های چندمقیاسی را با سرعت مناسب ترکیب می‌کند و ماژول C2PSA با توجه فضایی موازی، اهمیت بخش‌های مختلف تصویر را تنظیم و نویز پس‌زمینه را کاهش می‌دهد. این قابلیت‌ها آن را برای اصلاح در کاربردهای صنعتی مناسب می‌کنند. [12]، [13]، [29]

YOLO در کنترل کیفیت تولید، ناوبری خودکار و پایش فرایند نتایج کاربردی داشته است. بااین‌وجود، ویژگی‌های ریخته‌گری پیوسته فولاد به حدی خاص هستند که برای حفظ عملکرد مدل، به تغییراتی تخصصی در معماری و آموزش نیاز است. [14]، [28]، [5]، [8]

۲.۳. سازوکارهای توجه در بینایی ماشین

سازوکار توجه به شبکه امکان می‌دهد هنگام تحلیل داده‌های پیچیده، روی اطلاعات مؤثرتر تمرکز کند و سهم اطلاعات نامرتبط را کاهش دهد. این مفهوم پس از موفقیت در پردازش زبان طبیعی، به حوزه بینایی ماشین منتقل شد و در بسیاری از مسائل تشخیص تصویر کاربرد پیدا کرد. [16]، [30]

ماژول Squeeze-and-Excitation یا SE با توجه کانالی کار می‌کند: رابطه میان کانال‌های نقشه ویژگی را می‌آموزد و اهمیت آن‌ها را بازتنظیم می‌کند. SENetV2 این ایده را با معماری چندشاخه توسعه می‌دهد تا ارزیابی وابستگی‌های کانالی انعطاف بیشتری داشته باشد. [21]، [22]، [31]

توجه فضایی می‌کوشد محل مهم‌ترین نقاط تصویر را مشخص کند. توجه مختصاتی یا Coordinate Attention، برخلاف تجمیع سراسری که اطلاعات مکان را حذف می‌کند، مکان تقریبی ویژگی را در محورهای افقی و عمودی نگه می‌دارد. این خصوصیت برای یافتن اجسام کوچک اهمیت زیادی دارد. [19]، [20]، [15]، [32]

BiFormer سازوکاری سلسله‌مراتبی دارد: در سطح نخست، نواحی تصویری با احتمال ارتباط بیشتر انتخاب می‌شوند؛ در مرحله بعد، توجه در سطح توکن، روابط ظریف ویژگی‌ها را در همان نواحی بررسی می‌کند. چنین الگویی می‌تواند هم وابستگی‌های گسترده تصویر و هم جزئیات موضعی را با هزینه مناسب مدل کند. [17]، [18]

ترکیب چند نوع توجه در یک شبکه، راهی برای استفاده هم‌زمان از مزیت‌های آن‌هاست. پژوهش‌های پیشین نشان داده‌اند که این رویکردهای ترکیبی به‌خصوص در تشخیص اشیای کوچک و پس‌زمینه‌های پیچیده می‌توانند از یک سازوکار توجه منفرد موفق‌تر باشند. [23]، [24]، [31]، [32]

۲.۴. پایش فرایند ریخته‌گری پیوسته فولاد

در ریخته‌گری پیوسته، فولاد مذاب از مراحل مختلفی مانند عملیات پاتیل، توزیع در تاندیش، انجماد در قالب و خنک‌کاری ثانویه عبور می‌کند. کیفیت محصول وابسته به کنترل دقیق این مراحل است؛ بنابراین پایش شرایط جریان مذاب و مواد همراه اهمیت زیادی دارد. [1]، [2]، [4]

افزودن سرباره در این فرایند برای حفظ گرما، جذب آخال‌ها و محافظت در برابر اکسایش اهمیت دارد. توزیع و ضخامت لایه آن می‌تواند بر کیفیت سطح و ساختار داخلی شمش یا تختال و حتی ویژگی‌های مکانیکی محصول نهایی اثرگذار باشد. [4]، [5]

روش‌های متعارف نظارت بر وضعیت سرباره معمولاً به مشاهده اپراتور و تصمیم‌گیری بر اساس تجربه متکی هستند. این روش‌ها هم با تغییر شیفت و فرد، نوسان دارند و هم در محیط داغ، پرگردوغبار و دارای تداخل الکترومغناطیسی، مستعد خطا هستند. [3]، [8]

برای نظارت بر ریخته‌گری پیوسته، روش‌هایی مانند تصویربرداری حرارتی، حسگرهای فراصوت و شتاب‌سنج، و حسگرهای الکترومغناطیسی بررسی شده‌اند. بااین‌حال، همه آن‌ها در محیط سخت کارخانه به یک اندازه پایدار نیستند و بعضی نیز تفکیک‌پذیری مکانی لازم برای مشاهده ذرات کوچک سرباره را ندارند. [3]، [25]

سامانه‌های مبتنی بر تصویر، امکان پایش غیرتماسی و با جزئیات بالا را فراهم می‌کنند. مانع اصلی آن‌ها تنوع زیاد بافت سطح مذاب، روشنایی متغیر و کوچک‌بودن ویژگی‌های هدف در مقایسه با کل میدان دید است؛ مشکلی که در برخی حالات به ابعاد نزدیک یا حتی کوچک‌تر از یک پیکسل مؤثر می‌رسد. [5]، [8]، [14]

۲.۵. چالش‌های تشخیص اجسام کوچک

تشخیص اشیای کوچک از دشوارترین مسائل بینایی ماشین است. هدف کوچک تنها تعداد محدودی پیکسل

دارد، اطلاعات شکل و بافت آن ناکافی است و نویز پس‌زمینه می‌تواند آن را پنهان کند. در نتیجه، پردازش تصویر با وضوح مناسب و استخراج دقیق ویژگی اهمیت زیادی پیدا می‌کند. [33]، [34]، [35]

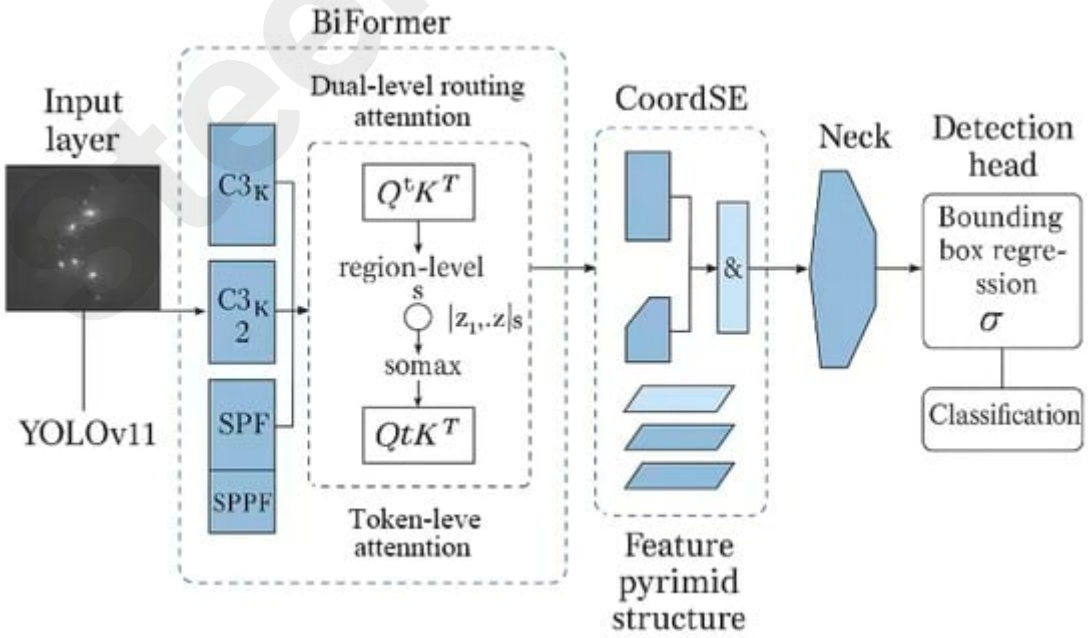
این دشواری در کارخانه شدیدتر است: محدودیت زمان پاسخ، محیط تصویری نامطمئن و نیاز به خطای کم، مسئله را سخت‌تر می‌کنند. ذرات سرباره در ریخته‌گری پیوسته نمونه روشنی از چنین اهدافی هستند؛ زیرا نسبت به تصویر کوچک، از نظر ظاهری متغیر و در زمینه‌ای پیچیده قرار دارند. [8]، [14]، [5]

در سال‌های اخیر، ادغام ویژگی‌های چندمقیاسی، شبکه‌های هرمی ویژگی و سازوکارهای توجه برای تشخیص اشیای کوچک توسعه یافته‌اند. در این روش‌ها، توجه محاسبات را به بخش‌هایی می‌برد که اطلاعات بیشتری برای آشکارسازی دارند و از اتلاف منابع پردازشی روی پس‌زمینه می‌کاهد. [33]، [34]، [35]، [32]

۳. روش‌شناسی

۳.۱. طراحی کلی معماری

معماری پیشنهادی BiFormer-CoordSEBlock-YOLOv11 ترکیبی از آشکارساز تک‌مرحله‌ای YOLOv11 و دو ماژول توجه تخصصی است. هدف آن حفظ سرعت موردنیاز پایش صنعتی همراه با افزایش توان تشخیص ذرات ریز سرباره در پس‌زمینه پیچیده فولاد مذاب است. نمای کلی معماری در شکل ۱ آمده است.



شکل 1. معماری BiFormer-CoordSEBlock-YOLOv11؛ برگرفته از مقاله اصلی با مجوز CC BY 4.0.

طراحی شبکه بر توجه سلسله‌مراتبی استوار است. در بخش Backbone یا استخراج‌کننده ویژگی، BiFormer امکان تحلیل مؤثر ویژگی‌های تصویر را فراهم می‌کند. در بخش Neck که وظیفه ترکیب ویژگی‌ها را بر عهده دارد، CoordSE اطلاعات مکان و ارتباط میان کانال‌های ویژگی را هنگام ادغام مقیاس‌های مختلف حفظ و تقویت می‌کند.

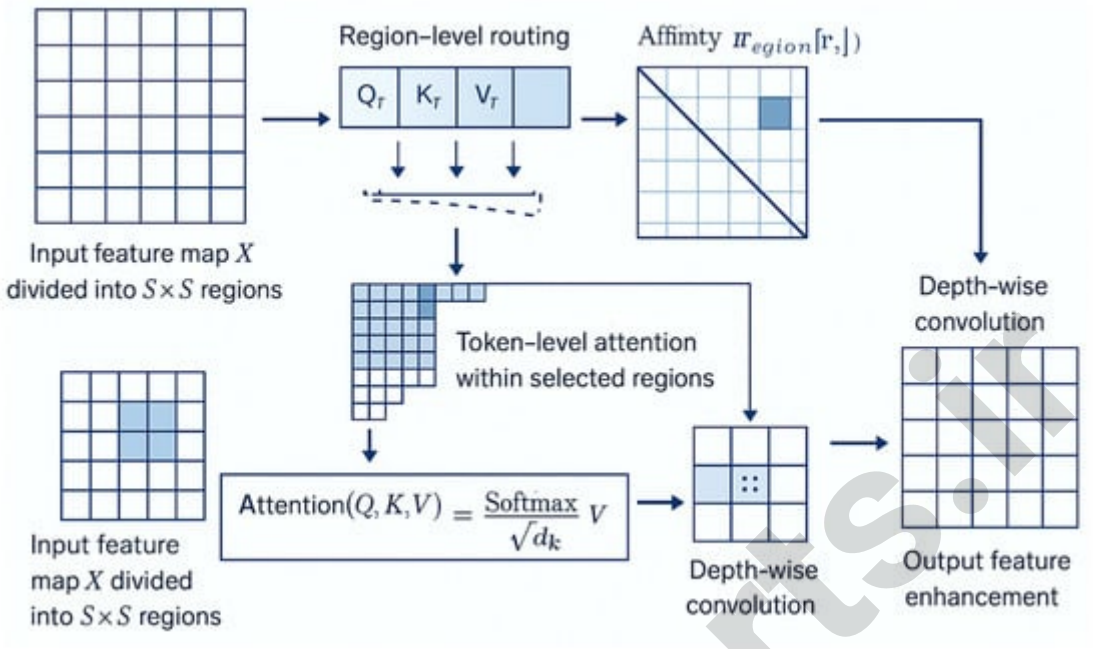
هدف از این ترکیب، رسیدن به تعادلی میان هزینه محاسبات و کیفیت تشخیص است. طراحی ماژولار می‌تواند در کاربردهای صنعتی دیگر نیز به‌کار رود؛ به شرط آنکه با نیازهای پردازشی، ویژگی‌های هدف و سخت‌افزار هر کاربرد دوباره ارزیابی شود.

۳.۲. ادغام توجه مسیریابی دوسطحی BiFormer

BiFormer با تحلیل توجه در دو سطح ناحیه و توکن، یکی از بهبودهای مهم در طراحی سازوکارهای توجه به‌شمار می‌رود. افزودن این ماژول به بخش استخراج ویژگی YOLOv11 با هدف شناسایی ذرات کوچک سرباره روی پس‌زمینه متغیر و درخشان سطح فولاد مذاب انجام شده است. [17]

در تصاویر صنعتی با وضوح بالا، توجه سراسری معمولی باید روابط شمار زیادی از نقاط را بررسی کند و هزینه زیادی تحمیل می‌شود. مشاهده خط تولید نشان می‌دهد که ذرات سرباره بخش بسیار کوچکی از تصویر را اشغال می‌کنند. BiFormer ابتدا با مسیریابی ناحیه‌ای، قسمت‌های وسیع و نامرتبط سطح مذاب را کنار می‌گذارد؛ سپس توجه در سطح توکن، جزئیات ذرات کوچک موجود در نواحی انتخاب‌شده را پالایش می‌کند. این راهبرد از محاسبات غیرضروری می‌کاهد و در عین حال به تشخیص جزئیات هدف کمک می‌کند. شکل ۲ سازوکار آن را نمایش می‌دهد.

SteelExperts.ir



شکل 2. سازوکار توجه مسیریابی دوسطحی BiFormer؛ برگرفته از مقاله اصلی با مجوز CC BY 4.0.

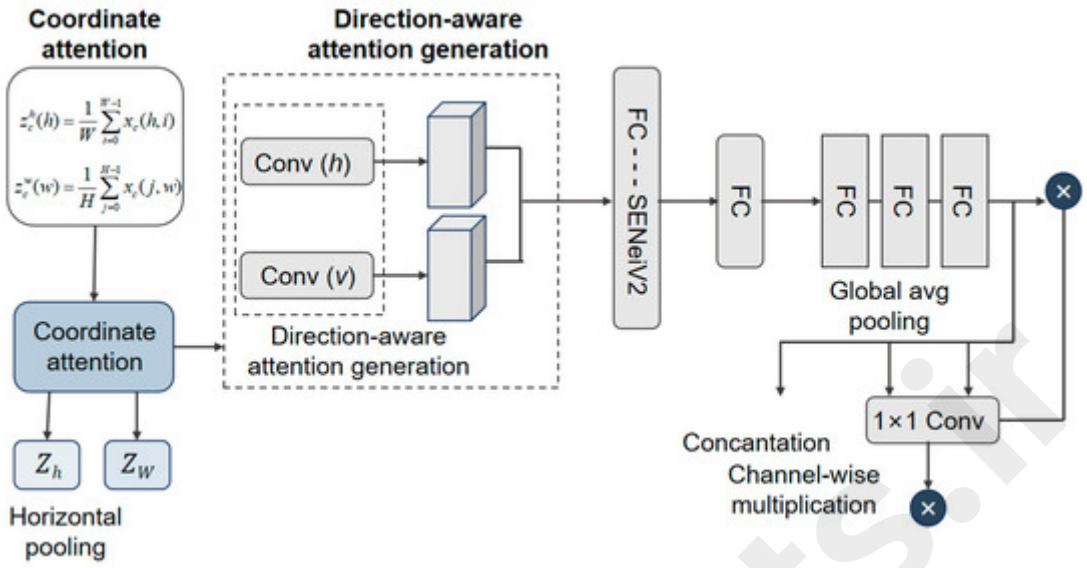
در نمایش ریاضی سازوکار، A_region ماتریس شباهت ناحیه‌هاست؛ I_r شاخص‌های k ناحیه مرتبط‌تر برای ناحیه r و d_k بعد بردارهای کلید را نشان می‌دهد. به این ترتیب، سهم اصلی توان پردازشی به بخش‌های اطلاع‌رسان‌تر تصویر اختصاص می‌یابد.

پس از انتخاب نواحی مهم، توجه در سطح توکن، با عملیات توجه چندسری، ویژگی‌های آن بخش‌ها را دقیق‌تر می‌کند. مزیت رویکرد دومرحله‌ای آن است که روابط فضایی دوربرد و وابستگی‌های محلی ظریف را با هم مدل می‌کند؛ امری که برای تشخیص هدف‌های کم‌پیکسل اهمیت دارد.

پژوهشگران BiFormer را در جایگاه‌های انتخابی بلوک C3K2 در Backbone قرار داده‌اند تا استخراج ویژگی ارتقا یابد و جریان کلی شبکه مختل نشود. پس از عملیات توجه، یک کانولوشن عمقی با کرنل ۵ نیز برای تکمیل تحلیل مکانی به‌کار رفته است؛ روشی که افزایش توان استخراج ویژگی را با هزینه محاسباتی کنترل‌شده دنبال می‌کند.

۳.۳. طراحی و پیاده‌سازی ماژول CoordSE

CoordSE حاصل ادغام توجه مختصاتی با نسخه تقویت‌شده SENetV2 است. این ماژول برای حل دو مسئله هم‌زمان طراحی شده: حفظ موقعیت دقیق هدف و جداسازی کانال‌های ویژگی مفید از کانال‌های کم‌اهمیت. شکل ۳ اجزای اصلی ساختار پیشنهادی را نشان می‌دهد.



شکل 3. ساختار CoordSE با ترکیب توجه مختصاتی، SENetV2 تقویت‌شده و SENetV؛ برگرفته از مقاله اصلی با مجوز CC BY 4.0.

تجمیع میانگین سراسری، در عین فشرده‌سازی اطلاعات، مکان دقیق ویژگی‌ها را از بین می‌برد. توجه مختصاتی به‌جای حذف یک‌باره هر دو بعد تصویر، از تجمیع جهت‌دار در محور افقی و عمودی استفاده می‌کند تا اطلاعات مربوط به موقعیت هدف همچنان در نقشه ویژگی باقی بماند.

ماژول SE استاندارد با استفاده از میانگین‌گیری سراسری، اطلاعات مختصات را از دست می‌دهد؛ در حالی که همین اطلاعات برای تشخیص ریزذرات سرباره ضروری است. CoordSE دو مسیر مکمل دارد: مسیر توجه مختصاتی، ویژگی‌های افقی و عمودی را استخراج می‌کند و خروجی آن به لایه‌های تمام‌متصل چندشاخه SENetV2 انتقال می‌یابد. بدین ترتیب، وزن‌دهی کانال‌ها تحت هدایت اطلاعات مکانی انجام می‌شود و مدل در برابر تداخل بخار متراکم، موقعیت‌یابی و انتخاب ویژگی مناسب‌تری دارد.

ترکیب دو سازوکار توجه باعث می‌شود بهبود مکان‌یابی و بهبود تمایز کانالی یکدیگر را تقویت کنند. اهمیت این هم‌افزایی در محیط‌هایی است که ظاهر ذره و زمینه بسیار شبیه‌اند و خطا در هر یک از دو مرحله، تشخیص نهایی را تضعیف می‌کند.

۳.۴. اصلاحات معماری شبکه

قرارگیری BiFormer و CoordSE در شبکه باید به‌گونه‌ای باشد که ارتباط میان لایه‌ها، مسیر جریان گرادیان و هزینه اجرای مدل مختل نشود. اصلاحات انجام‌شده بر ارتقای توان استخراج ویژگی و توجه متمرکز است، نه بر کنارگذاشتن ساختار کارآمد تک‌مرحله‌ای YOLO.

در بخش Neck، ماژول‌های CoordSE در فرایند ادغام ویژگی‌ها حضور دارند و ارتباطات مکانی و کانالی را در

نقشه‌هایی با وضوح متفاوت حفظ می‌کنند. این جانمایی باید به‌اندازه کافی وضوح تصویر را نگه دارد تا جزئیات ریز هدف، هنگام تجمیع ویژگی‌ها از بین نرود.

ادغام ویژگی‌ها در مقیاس‌های گوناگون برای آشکارسازی ذرات کوچک ضروری است. CoordSE در Neck کمک می‌کند ویژگی‌های به‌دست‌آمده در بخش‌های مختلف شبکه با حفظ نشانه‌های مکانی ترکیب شوند و خروجی برای تعیین محل و نوع هدف مناسب‌تر باشد.

معماری اصلاح‌شده همچنان از سه مقیاس آشکارسازی YOLOv11 بهره می‌گیرد؛ یعنی سرهای تشخیص با وضوح فضایی متفاوت به اجسامی با اندازه‌های متفاوت پاسخ می‌دهند. BiFormer و CoordSE با بهبود کیفیت نقشه‌های ویژگی تغذیه‌کننده این سرها، تلاش می‌کنند عملکرد تشخیص را در هر سه مقیاس تقویت کنند.

در طراحی اتصالات میان‌بر و مسیرهای تجمیع، سازگاری با SPPF و سرهای تشخیص حفظ شده است. تداوم جریان گرادیان برای آموزش مؤثر شبکه ضروری است و جانمایی سازوکارهای توجه نباید مانع یادگیری بخش‌های موجود شود.

۳.۵. تابع زیان و راهبرد آموزش

راهبرد آموزش شبکه پیشنهادی برای تشخیص اجسام کوچک در محیط صنعتی تنظیم شده است. تابع زیان چند بخش دارد تا مکان‌یابی کادر هدف، طبقه‌بندی ذره و یادگیری سازوکارهای توجه به‌صورت هماهنگ بهینه شوند.

در تنظیم تابع زیان مطابق YOLOv11، به‌جای تکیه بر زیان استاندارد Objectness، از زیان کانونی توزیعی یا DFL استفاده شده است. تابع زیان کل از سه مؤلفه تشکیل می‌شود: زیان کادر محصورکننده L_box، زیان طبقه‌بندی L_cls و زیان توزیعی L_dfl. رابطه مورد استفاده چنین است: L_total = 7.5 × L_box + 0.5 × L_cls + 1.5 × L_dfl. وزن بیشتر زیان کادر، نشان می‌دهد که تعیین مرز ذره کوچک برای این کاربرد دشوارتر و مهم‌تر از صرف تشخیص دسته آن تلقی شده است.

$$L_{total} = 7.5\, L_{box} + 0.5\, L_{cls} + 1.5\, L_{dfl}$$

برای طبقه‌بندی، از آنتروپی متقاطع دودویی و روش‌های متعادل‌سازی کلاس‌ها استفاده شده تا اختلاف فراوانی نمونه‌های سرباره و زمینه بر آموزش اثر نامطلوب نگذارد. برای تنظیم کادر محصورکننده نیز زیان Complete IoU یا CIoU به‌کار رفته است تا علاوه بر همپوشانی، کیفیت مکان‌یابی بهبود یابد.

در فرایند آموزش، پژوهشگران ورود تدریجی اثر سازوکارهای توجه را برای پایداری همگرایی در نظر گرفته‌اند.

هدف این کار جلوگیری از غالب‌شدن ناگهانی لایه‌های توجه در مراحل اولیه و فراهم‌کردن فرصت برای یادگیری پایدار شبکه پایه است.

تقویت داده برای بازنمایی شرایط صنعتی شامل تغییر روشنایی و رنگ، تغییر هندسی و زاویه مشاهده، و تزریق نویز مصنوعی است. این تغییرات به مدل کمک می‌کنند در برابر تغییرات نور، حالت سطح مذاب و تداخل‌هایی که در کارخانه رخ می‌دهد مقاوم‌تر شود.

راهبرد بهینه‌سازی همچنین شامل زمان‌بندی تطبیقی نرخ یادگیری، دوره گرم‌کردن اولیه، محدودکردن گرادیان و منظم‌سازی Weight Decay است. این تنظیمات برای پیشگیری از بیش‌برازش و حفظ پایداری آموزش شبکه دارای سازوکار توجه پیچیده استفاده می‌شوند.

۴. محیط آزمایش و مجموعه‌داده

۴.۱. ساخت و ویژگی‌های مجموعه‌داده

ازآنجاکه مجموعه‌داده عمومی مناسبی برای آشکارسازی ذرات سرباره در هنگام افزودن به ریخته‌گری پیوسته وجود نداشت، پژوهشگران یک مجموعه اختصاصی تهیه کردند. تصاویر با همکاری واحدهای تولید فولاد و در شرایطی گردآوری شدند که تا حد ممکن ویژگی‌های محیط واقعی خط را بازنمایی کنند.

گردآوری داده در چند مجموعه فولادسازی و طی شش ماه انجام شد تا تفاوت شرایط کار، چیدمان تجهیزات و شیوه افزودن سرباره در داده‌ها حضور داشته باشد. دوربین‌های صنعتی با وضوح بالا و فیلترهای ویژه برای ثبت تصویر در نزدیکی دمای بالا، تداخل الکترومغناطیسی و روشنایی متغیر به‌کار رفتند.

مجموعه نهایی شامل ۲۸۴۷ تصویر با ابعاد ۱۹۲۰ × ۱۰۸۰ پیکسل است. در هر تصویر بین یک تا ۱۵ ذره سرباره دیده می‌شود و میانگین تعداد ذرات ۴٫۲ است. اندازه، شکل و ظاهر ذرات تغییر می‌کند و این تنوع، شرایط طبیعی فرایند صنعتی را منعکس می‌کند.

برچسب‌گذاری تصویرها به کمک کارشناسان آشنا با ریخته‌گری پیوسته انجام شد. هر برچسب، مختصات دقیق کادر محصورکننده، دسته‌بندی ذره بر اساس ویژگی‌های ظاهری و اندازه، و اطلاعات زمینه‌ای مربوط به شرایط ریخته‌گری هنگام تصویربرداری را در بر می‌گیرد.

اندازه اهداف از حدود ۱۵ × ۱۵ تا ۱۲۰ × ۸۰ پیکسل متغیر است. ۶۸ درصد آن‌ها در گروه اشیای کوچک‌تر از ۳۲ × ۳۲ پیکسل جای می‌گیرند. بافت متحرک سطح فولاد مذاب و بازتاب‌های متفاوت آن، زمینه‌ای پیچیده برای

سنجش عملکرد الگوریتم ایجاد می‌کند.

در این مجموعه‌داده عدم‌تعادل شدیدی میان هدف و پس‌زمینه وجود دارد: ذرات سرباره به‌طور میانگین فقط ۲٫۳ درصد مساحت تصویر را اشغال می‌کنند. این نسبت کوچک شبیه بسیاری از کاربردهای واقعی نظارت صنعتی است. تصویرها در روشنایی طبیعی و مصنوعی جمع‌آوری شده‌اند تا امکان بررسی مقاومت مدل در سناریوهای مختلف فراهم شود.

نسخه‌ای نمونه از مجموعه Add_Slag2025 همراه با برچسب‌ها و تقسیم‌بندی آزمون، بنا به درخواست از نویسندگان قابل دریافت است؛ اما کل داده‌ها به علت محرمانگی اطلاعات خطوط تولید کارخانه‌ها به‌صورت عمومی منتشر نشده است. این موضوع امکان تکرار مستقل همه آزمایش‌ها را محدود می‌کند.

۴.۲. پیکربندی آزمایش و سخت‌افزار

داده‌ها با نسبت ۷۰ درصد آموزش، ۱۵ درصد اعتبارسنجی و ۱۵ درصد آزمون تقسیم شدند؛ یعنی ۱۹۹۳ تصویر آموزشی، ۴۲۷ تصویر اعتبارسنجی و ۴۲۷ تصویر آزمون. آموزش مدل پیشنهادی و مدل‌های مقایسه‌ای با سه بذر تصادفی مختلف تکرار شد. اعداد گزارش‌شده میانگین و انحراف معیار این سه اجرا هستند.

آموزش و ارزیابی با استفاده از پردازنده‌های گرافیکی NVIDIA RTX 4090 انجام شد. این سخت‌افزار توان محاسباتی لازم برای بررسی مدل‌های دارای سازوکارهای توجه و تحلیل دقیق زمان استنتاج را فراهم کرده است. این نتایج را نباید بدون آزمون، به سخت‌افزارهای ضعیف‌تر تعمیم داد.

به دلیل مصرف حافظه تصاویر صنعتی با وضوح بالا، اندازه دسته آموزشی ۱۶ انتخاب شد. بهینه‌ساز AdamW با نرخ یادگیری اولیه ۰٫۰۱ در ۱۰۰ دوره آموزشی استفاده شد. مقدار Weight Decay نیز ۰٫۰۰۰۵ قرار گرفت تا خطر بیش‌برازش کاهش یابد. جدول ۱ تنظیمات اصلی آموزش را نشان می‌دهد.

ابرپارامتر	مقدار
نرخ یادگیری اولیه	0.01
بهینه‌ساز	AdamW
تعداد دوره‌های آموزش	100
اندازه دسته (Batch Size)	16
ضریب کاهش وزن (Weight Decay)	0.0005

تقویت داده شامل وارون‌سازی افقی با احتمال ۰٫۵، مقیاس‌دهی تصادفی بین ۰٫۸ تا ۱٫۲، تغییر روشنایی و

کنتراست هر یک به میزان ±۰٫۲ و تزریق نویز گاوسی با انحراف معیار ۰٫۰۱ است. نویز افزوده‌شده برای شبیه‌سازی بخشی از اغتشاش‌های تصویری ناشی از محیط الکترومغناطیسی صنعتی در نظر گرفته شده است.

پروتکل ارزیابی شامل Precision، Recall، میانگین دقت متوسط یا mAP با آستانه همپوشانی ۰٫۵ و بازه ۰٫۵ تا ۰٫۹۵ و نیز سرعت استنتاج است. بررسی روی داده‌های اعتبارسنجی و آزمون انجام شده تا عملکرد فقط بر اساس داده‌های استفاده‌شده در تنظیم مدل گزارش نشود.

۴.۳. روش‌های مرجع و چارچوب مقایسه

برای سنجش تأثیر اصلاحات، مدل پیشنهادی با مجموعه‌ای از آشکارسازهای شناخته‌شده که سبک‌های معماری متفاوتی دارند مقایسه شد. هدف این مقایسه، تعیین مزیت واقعی دو ماژول توجه در مقابل بهبودهایی است که صرفاً از انتخاب یک نسل جدیدتر الگوریتم ناشی می‌شود.

نسخه‌های YOLOv9t، YOLOv8n، YOLOv5s و YOLOv10n به‌عنوان نمونه‌های مقایسه‌ای انتخاب شدند. همه مدل‌های پایه، از جمله معماری‌های کلاسیک اضافه‌شده به آزمایش، دوباره از ابتدا آموزش داده شدند. تنظیمات تقویت داده، وضوح ورودی و بودجه جست‌وجوی ابرپارامترها تا حد امکان یکسان نگه داشته شدند تا مقایسه منصفانه باشد.

مدل RT-DETR نیز برای نمایش رویکرد آشکارسازی مبتنی بر ترنسفورمر در مقایسه گنجانده شد. تعداد پارامترهای آن در سطحی قابل‌مقایسه تنظیم شد تا تفاوت نتایج تا حد امکان به نوع معماری و سازوکار توجه مرتبط باشد.

برای پایش ریخته‌گری، سرعت بلادرنگ اهمیت دارد؛ بنابراین تمرکز اصلی بر آشکارسازهای تک‌مرحله‌ای و ترنسفورمرهای سریع بود. نویسندگان توضیح می‌دهند که بسیاری از مدل‌های دومرحله‌ای متعارف دستیابی به هدف ۳۰ فریم‌برثانیه را دشوار می‌کنند و اتکای EfficientDet به لنگرهای نسبتاً ثابت نیز با شکل متغیر سرباره مذاب سازگار نیست.

بااین‌حال، برای کامل‌تر شدن مقایسه، معماری‌های Faster R-CNN و EfficientDet نیز ارزیابی شدند. همچنین نسخه YOLOv11 همراه با توجه SE استاندارد به آزمایش اضافه شد تا ارزش CoordSE جدید نسبت به سازوکار توجه کانالی قدیمی‌تر جداگانه مشخص شود.

مقایسه فقط بر یک معیار متکی نیست؛ بلکه معیارهای دقت پیش‌بینی، بازیابی، mAP، سرعت استنتاج، تعداد پارامترها و حافظه مصرفی را در بر می‌گیرد. علاوه بر این، با مشاهده کیفی نتایج، نقشه‌های توجه و موارد شکست، رفتار مدل در شرایط دشوار تحلیل شده است.

۴.۴. طراحی آزمایش حذف مؤلفه‌ها

در آزمایش حذف مؤلفه‌ها یا Ablation Study، اجزای مدل به‌صورت کنترل‌شده حذف یا اضافه می‌شوند تا سهم واقعی هر یک در نتیجه نهایی مشخص شود. این روش بررسی می‌کند که بهبود عملکرد از کدام مؤلفه ناشی شده و آیا ترکیب آن‌ها اثر افزایشی دارد یا خیر.

چهار پیکربندی اصلی عبارت‌اند از YOLOv11 پایه، مدل پایه همراه فقط BiFormer، مدل پایه همراه فقط CoordSE و معماری کامل دارای هر دو ماژول. این چهار حالت امکان بررسی مستقل هر مؤلفه و اثر ترکیبی آن‌ها را فراهم می‌کنند.

آزمایش‌های تکمیلی به انتخاب‌های جزئی‌تر معماری نیز می‌پردازند؛ مانند ابعاد تقسیم‌بندی ناحیه‌ها در BiFormer برابر ۴ × ۴، ۷ × ۷ و ۱۴ × ۱۴، تعداد شاخه‌های SENetV2 تقویت‌شده برابر ۲، ۴ و ۸، و محل قرارگیری ماژول‌های توجه در شبکه.

در ارزیابی حذف مؤلفه‌ها، هزینه محاسباتی هر ترکیب نیز در نظر گرفته شده است. این نکته برای کاربرد صنعتی مهم است؛ زیرا تنها بهبود دقت کافی نیست و نتیجه باید در محدودیت زمان پاسخ و توان سخت‌افزار کارخانه قابل استفاده باشد.

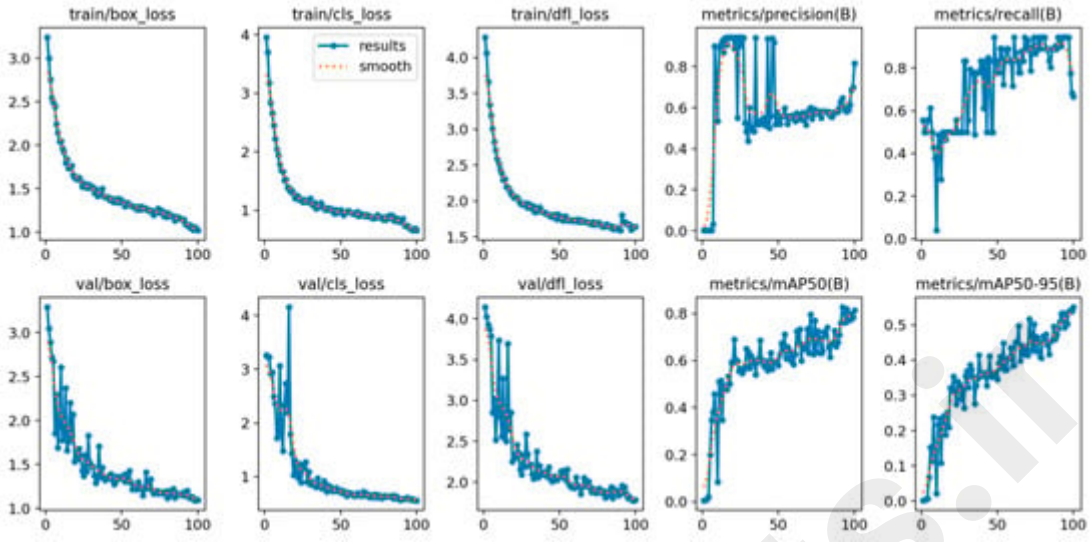
۵. نتایج و تحلیل

۵.۱. مقایسه عملکرد کلی

نتایج آزمایشگاهی نشان می‌دهد که ترکیب BiFormer و CoordSE در چارچوب YOLOv11 در مجموع عملکرد رقابتی دارد و در معیار mAP@0.5 از نمونه‌های مقایسه‌ای انتخاب‌شده بهتر عمل می‌کند. بااین‌حال، هنگام ارزیابی کاربرد صنعتی باید به هر سه شاخص دقت، بازیابی و سرعت توجه کرد؛ زیرا افزایش یک معیار ممکن است با افت معیار دیگر همراه باشد.

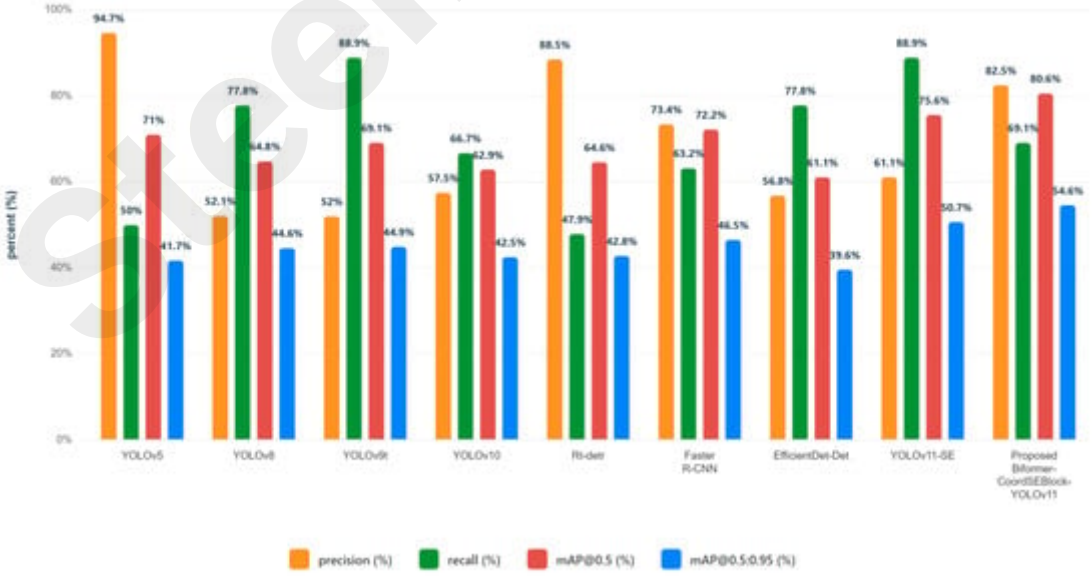
شکل ۴ روند همگرایی مدل طی ۱۰۰ دوره آموزش را نمایش می‌دهد: منحنی‌های زیان آموزش و اعتبارسنجی و معیارهای اصلی آشکارسازی در آن مشخص‌اند. طبق گزارش نویسندگان، زیان‌های مکان‌یابی، طبقه‌بندی و DFL به‌طور نسبتاً هموار کاهش یافته‌اند و نشانه برجسته‌ای از ناپایداری آموزش مشاهده نشده است. افزایش اولیه Precision و Recall نیز می‌تواند نشانه آن باشد که BiFormer به شبکه کمک می‌کند در همان دوره‌های نخست، ویژگی‌های مرتبط با ریزذرات سرباره را پیدا کند.

SteelExperts.ir



شکل 4. نمودار همگرایی شبکه نهایی در طول آموزش؛ برگرفته از مقاله اصلی با مجوز CC BY 4.0.

شکل ۵ معیارهای مدل‌های مختلف را مقایسه می‌کند و نشان می‌دهد که Precision بالا الزاماً به معنای تشخیص همه ذرات نیست. برای نمونه، RT-DETR با دقت پیش‌بینی ۸۸٫۵ ± ۰٫۵ درصد و YOLOv5 با ۹۴٫۷ ± ۰٫۳ درصد، بازیابی‌هایی فقط برابر ۴۷٫۹ ± ۱٫۱ و ۵۰٫۰ ± ۰٫۸ درصد داشته‌اند. نسخه YOLOv11-SE با توجه SE استاندارد به mAP@0.5 برابر ۷۵٫۶ ± ۰٫۴ درصد رسیده است. مدل کامل پیشنهادی در این مقایسه به mAP@0.5 بالاتر دست یافته، اما همچنان بخشی از ذرات را از دست می‌دهد؛ موضوعی که برای کنترل فرایند باید جداگانه بررسی شود.



شکل 5. مقایسه معیارهای عملکرد در پایش افزودن سرباره در ریخته‌گری پیوسته فولاد؛ برگرفته از مقاله اصلی با مجوز CC BY 4.0.

۵.۲. نتایج آزمایش حذف مؤلفه‌ها

آزمایش حذف مؤلفه‌ها نشان می‌دهد هر دو ماژول BiFormer و CoordSE در افزایش عملکرد نسبت به مدل پایه سهم دارند. ترکیب آن‌ها نیز از افزودن مستقل هرکدام بهتر بوده است؛ بااین‌حال، میزان بهبود باید در کنار تغییرات بازیابی خوانده شود.

در مدل پایه YOLOv11، دقت پیش‌بینی ۵۴٫۰ ± ۰٫۷ درصد، بازیابی ۸۳٫۳ ± ۰٫۵ درصد و mAP@0.5 برابر ۴۴٫۹ ± ۰٫۵ درصد ثبت شد. افزودن فقط BiFormer این مقادیر را به‌ترتیب به ۷۱٫۲ ± ۰٫۴، ۷۸٫۹ ± ۰٫۶ و ۶۸٫۳ ± ۰٫۴ درصد رساند. بنابراین Precision به اندازه ۱۷٫۲ واحد درصد و mAP@0.5 به میزان ۲۳٫۴ واحد درصد افزایش یافت، اما Recall نسبت به پایه کاهش پیدا کرد.

وقتی فقط CoordSE به YOLOv11 افزوده شد، Precision به ۶۸٫۵ ± ۰٫۵ درصد، Recall به ۷۵٫۲ ± ۰٫۴ درصد و mAP@0.5 به ۶۵٫۷ ± ۰٫۵ درصد رسید. این اعداد در مقایسه با مدل پایه بهبود ۱۴٫۵ واحد درصد در Precision و ۲۰٫۸ واحد درصد در mAP@0.5 را نشان می‌دهند. نقش CoordSE در حفظ اطلاعات مکان و تمایز کانال‌های تصویری از این نتایج قابل برداشت است.

معماری کامل دارای هر دو ماژول، Precision برابر ۸۲٫۵ ± ۰٫۲ درصد، Recall برابر ۶۹٫۱ ± ۰٫۳ درصد و mAP@0.5 برابر ۸۰٫۶ ± ۰٫۲ درصد به‌دست آورد. نسبت به مدل فقط BiFormer، دقت پیش‌بینی طبق گزارش مقاله ۱۱٫۴ واحد درصد و mAP@0.5 برابر ۱۲٫۳ واحد درصد افزایش پیدا کرد. در مقابل، نرخ بازیابی از ۷۸٫۹ به ۶۹٫۱ درصد کاهش یافت. این مبادله میان هشدار اشتباه و ازدست‌دادن هدف، در کاربردهای حساس پایش فولاد باید آگاهانه ارزیابی شود.

پیکربندی	Precision	Recall	mAP@0.5
YOLOv11 پایه	0.7 ± 54.0	0.5 ± 83.3	0.5 ± 44.9
YOLOv11 + BiFormer	0.4 ± 71.2	0.6 ± 78.9	0.4 ± 68.3
YOLOv11 + CoordSE	0.5 ± 68.5	0.4 ± 75.2	0.5 ± 65.7
مدل کامل BiFormer + CoordSE	0.2 ± 82.5	0.3 ± 69.1	0.2 ± 80.6

BiFormer عمدتاً با متمرکزکردن توجه بر نواحی مهم و حذف زمینه نامرتبط، تعداد تشخیص‌های مثبت اشتباه را کاهش می‌دهد. CoordSE به مکان‌یابی و تشخیص ویژگی‌های ظریف کمک می‌کند. نتایج ترکیبی نشان‌دهنده اثر مکمل آن‌هاست، ولی عملکرد نهایی از نظر همه معیارها به‌طور هم‌زمان بهترین نیست.

۵.۳. تحلیل کارایی محاسباتی

تحلیل کارایی محاسباتی نشان می‌دهد که افزودن ماژول‌های توجه، با وجود افزایش دقت تشخیص، هنوز اجازه اجرای مدل در نرخ فریم بلادرنگ را روی سخت‌افزار آزمایشگاهی می‌دهد. برای کاربرد صنعتی، این توازن میان کیفیت تشخیص و هزینه اجرا باید با الزامات خط تولید مقایسه شود.

مدل پیشنهادی روی NVIDIA RTX 4090 با سرعت ۴۵٫۲ ± ۰٫۳ فریم‌برثانیه اجرا شده است؛ درحالی‌که مدل پایه YOLOv11 به ۵۲٫۱ ± ۰٫۲ فریم‌برثانیه می‌رسد. این اختلاف معادل کاهش حدود ۱۳٫۲ درصدی سرعت استنتاج است. نویسندگان این افت را در برابر افزایش mAP قابل قبول ارزیابی کرده‌اند.

تعداد پارامترهای مدل از حدود ۲٫۶ میلیون در نسخه پایه به ۳٫۱ میلیون در نسخه پیشنهادی رسیده؛ یعنی افزایشی در حدود ۱۸٫۷ درصد. لایه‌های تمام‌متصل اضافی SENetV2 تقویت‌شده و محاسبات توجه BiFormer از عوامل اصلی این افزایش هستند.

در هنگام استنتاج، مصرف حافظه پردازنده گرافیکی مدل پیشنهادی نزدیک به ۱۵ درصد بیشتر از مدل پایه گزارش شده است. این سربار عمدتاً به محاسبه میانی و نگهداری نقشه‌های توجه مربوط است؛ ازاین‌رو هنگام انتخاب سخت‌افزار و تعداد دوربین‌های هم‌زمان باید لحاظ شود.

نویسندگان گزارش می‌کنند که توجه دومرحله‌ای BiFormer از نظر تعداد توکن‌ها پیچیدگی تقریباً خطی دارد؛ در مقایسه با خودتوجهی استاندارد که پیچیدگی آن به‌صورت مربعی رشد می‌کند. این تفاوت در پردازش تصاویر بزرگ صنعتی مهم است؛ زیرا شمار نقاط و روابط بالقوه با افزایش وضوح به‌سرعت بالا می‌رود.

۵.۴. بررسی کیفی و نمایش نتایج

بررسی کیفی به روشن‌شدن رفتار شبکه کمک می‌کند: معیارهای عددی به‌تنهایی نشان نمی‌دهند مدل به کدام بخش تصویر توجه کرده، در چه وضعیتی اشتباه کرده و چرا برخی ذرات را ندیده است.

نقشه‌های توجه BiFormer نشان می‌دهند شبکه بیشتر روی نواحی دارای احتمال حضور سرباره تمرکز می‌کند و سهم بخش‌های نامرتبط سطح مذاب را کاهش می‌دهد. مرحله ناحیه‌ای موقعیت‌های مهم را برمی‌گزیند و مرحله توکن جزئیات همان مناطق را برای تعیین مرز ذره پالایش می‌کند.

در نقشه‌های توجه CoordSE، بخش مختصاتی به حفظ نشانه‌های مکان ذرات کوچک کمک می‌کند و SENetV2 تقویت‌شده کانال‌هایی را برجسته می‌کند که برای جداسازی سرباره از زمینه مناسب‌ترند. این دو کارکرد به شکل مکمل در جریان استخراج ویژگی حضور دارند.

تحلیل خطا نشان می‌دهد که برخی بازتاب‌های پیچیده یا ناهمواری‌های سطح مذاب، تشخیص مثبت اشتباه ایجاد می‌کنند. همچنین ذرات بسیار کوچک یا تاحدی پوشیده‌شده ممکن است اصلاً تشخیص داده نشوند. این موارد محدودیت‌هایی واقعی‌اند و لزوماً با افزایش ساده تعداد لایه‌های شبکه برطرف نمی‌شوند.

مقایسه در نورها و وضعیت‌های ظاهری مختلف، طبق ارزیابی نویسندگان، پایداری نسبی روش پیشنهادی را نشان می‌دهد. سازوکارهای توجه کمک می‌کنند تغییر اندازه ذرات، تغییر بافت زمینه و روشنایی گوناگون کمتر به نتیجه لطمه بزند. بااین‌حال، میزان این پایداری فقط در دامنه داده‌های گردآوری‌شده سنجیده شده است.

۵.۵. ملاحظات استقرار در صنعت

بررسی ملاحظات استقرار نشان می‌دهد که ترکیب دقت بهتر، سرعت پردازش مناسب و سربار محاسباتی کنترل‌شده می‌تواند مبنایی برای اتصال این سامانه به زیرساخت بینایی صنعتی ریخته‌گری پیوسته باشد. ولی امکان اتصال فنی به‌تنهایی به معنای اثبات قابلیت اعتماد عملیاتی نیست.

سرعت ۴۵٫۲ فریم‌برثانیه روی سخت‌افزار مورد آزمایش، از حد هدف ۳۰ فریم‌برثانیه بالاتر است و برای مشاهده تغییرات سریع در افزودن سرباره وضوح زمانی مناسبی فراهم می‌کند. بااین‌وجود، سرعت نهایی سامانه کارخانه به مراحل دریافت تصویر، پیش‌پردازش، انتقال داده و نمایش یا فرمان کنترلی نیز وابسته خواهد بود.

ارزیابی در شرایط شبیه‌سازی‌شده صنعتی نشان داده که مدل می‌تواند در برابر تغییر نور، برخی آثار تداخل الکترومغناطیسی و نوسان شرایط محیطی عملکرد خود را حفظ کند. بخش مهمی از این ویژگی به شیوه تقویت داده هنگام آموزش برمی‌گردد؛ نه به آزمون کامل همه تنش‌های محیط واقعی کارخانه.

طبق تحلیل نویسندگان، ورودی و خروجی استاندارد الگوریتم و طراحی ماژولار آن، اتصال به سامانه‌های بینایی صنعتی موجود را تسهیل می‌کند. بااین‌حال، سازگاری با دوربین، شبکه و سامانه کنترل هر واحد باید در فرایند مهندسی و آزمون همان سایت بررسی شود.

تحلیل هزینه–فایده مقاله بر این فرض استوار است که تشخیص دقیق‌تر سرباره می‌تواند به کاهش اتلاف مواد، کیفیت بهتر محصول و کنترل مناسب‌تر فرایند کمک کند. این مزیت‌ها ممکن است هزینه افزایش محاسبات و پیچیدگی مدل را توجیه کنند؛ اما متن مقاله داده‌ای از بازگشت سرمایه واقعی یک پروژه صنعتی ارائه نمی‌کند.

بخش مهمی از ارزیابی استقرار به محدودیت سخت‌افزار مربوط است. ۴۵٫۲ فریم‌برثانیه روی RTX 4090 اندازه‌گیری شده، درحالی‌که بسیاری از کارخانه‌ها برای رایانه‌های لبه‌ای محدودیت توان و فضا دارند. نویسندگان درباره اجرای مدل روی Jetson AGX Orin تنها برآورد نظری ارائه کرده‌اند و احتمال داده‌اند به کمک TensorRT

و محاسبه نیم‌دقت FP16 بتوان به نزدیکی ۳۰ فریم‌برثانیه رسید. **این عدد روی Jetson اندازه‌گیری نشده است** و باید در آزمایش واقعی تأیید شود.

۶. بحث

۶.۱. دستاوردهای نظری

این پژوهش از نظر طراحی مدل نشان می‌دهد که ترکیب توجه دوسطحی و توجه مختصاتی–کانالی می‌تواند برای تشخیص ذرات کوچک در محیط صنعتی مفید باشد. نوآوری اصلی، استفاده هم‌زمان از دو سازوکار مکمل در معماری YOLOv11 برای مسئله مشخص سرباره است.

مبنای نظری این رویکرد آن است که تشخیص هدف کوچک هم به آگاهی از ساختار کلی تصویر و هم به پالایش جزئیات محلی نیاز دارد. BiFormer بیشتر وظیفه انتخاب نواحی مرتبط و مدل‌کردن روابط آن‌ها را انجام می‌دهد و CoordSE موقعیت و کانال‌های ویژگی را هنگام ادغام حفظ می‌کند.

مسیریابی توجه در BiFormer با محدودکردن محاسبات به نواحی منتخب می‌تواند هزینه تحلیل روابط گسترده تصویر را کاهش دهد. استفاده از این الگو در آشکارسازی سرباره نشان می‌دهد که کاربرد آن می‌تواند فراتر از مسائل عمومی بینایی ماشین باشد.

تقویت توجه کانالی با داده‌های مختصاتی، محدودیت تجمیع سراسری در مدل‌های SE معمولی را کاهش می‌دهد. حفظ نشانه‌های مکانی به‌ویژه برای اجسامی مهم است که بخش بسیار کوچکی از میدان دید را اشغال می‌کنند و با اندکی جابه‌جایی یا محوشدگی ممکن است گم شوند.

بررسی نتایج ترکیبی نشان می‌دهد که طراحی دقیق دو سازوکار توجه می‌تواند مزایایی فراتر از اعمال هرکدام به‌تنهایی داشته باشد. بااین‌حال، برای ادعای هم‌افزایی در تمام معیارها باید افت Recall را نیز در نظر گرفت؛ آنچه بهبود یافته، در این مطالعه بیش از همه Precision و mAP بوده است.

۶.۲. کاربردهای عملی در صنعت

کاربرد احتمالی این روش محدود به افزودن سرباره در ریخته‌گری پیوسته نیست. هر مسئله صنعتی که در آن هدف کوچک، زمینه پیچیده و محدودیت زمانی وجود دارد، ممکن است از ایده انتخاب ناحیه و حفظ مختصات بهره بگیرد؛ مانند کنترل کیفیت تصویری، تشخیص ذرات و پایش اجزای کوچک تجهیزات.

سامانه‌های بینایی که بتوانند ضمن حفظ نرخ فریم، تصویر را با دقت مناسب تحلیل کنند، برای خودکارسازی کنترل فرایند ارزشمندند. چنین سامانه‌هایی امکان تشخیص به‌موقع تغییرات و فراهم‌کردن داده برای تصمیم‌گیری اپراتور یا سامانه کنترلی را ایجاد می‌کنند.

از نظر نظری و داده‌های آزمایشی، روش پیشنهادی با هدف مقاومت در برابر تغییر روشنایی و آشفتگی تصویری طراحی شده است. پایداری در برابر دمای بالا و تداخل الکترومغناطیسی باید در سطح یک سامانه کامل، شامل دوربین و تجهیزات جانبی، آزموده شود؛ زیرا مقاومت نرم‌افزار به‌تنهایی از تجهیزات در برابر آسیب فیزیکی محافظت نمی‌کند.

ساختار ماژولار، امکان انتخاب BiFormer یا CoordSE بر اساس محدودیت محاسبات و نیاز دقت را فراهم می‌کند. این انعطاف در کارخانه‌هایی با سخت‌افزار ضعیف‌تر یا الگوی متفاوت خطا می‌تواند برای طراحی نسخه سبک‌تر شبکه مفید باشد.

مجموعه‌داده تخصصی Add_Slag2025 نیز می‌تواند مبنایی برای توسعه روش‌های بعدی باشد؛ زیرا نمونه‌های تصویری از محیط واقعی با برچسب‌های کارشناسی را دربر می‌گیرد. بااین‌حال، انتشار محدود داده‌ها به درخواست نویسندگان، دسترسی عمومی به یک معیار آزمون کاملاً باز را فراهم نکرده است.

۶.۳. محدودیت‌ها و مسیرهای پژوهش آینده

پژوهشگران در کنار بهبودهای گزارش‌شده، محدودیت‌های مشخصی را نیز بیان می‌کنند. شناخت این محدودیت‌ها برای تبدیل یک مدل تحقیقاتی به سامانه پایش قابل‌اتکا در کارخانه ضروری است.

افزایش مصرف پردازنده و حافظه ناشی از لایه‌های توجه ممکن است اجرای روش را روی سخت‌افزار لبه‌ای محدود کند. پژوهش‌های بعدی می‌توانند از سازوکارهای توجه سبک‌تر یا توجه تطبیقی بهره بگیرند تا هزینه اجرا بر اساس پیچیدگی صحنه تنظیم شود.

برای ذرات بسیار ریز، به‌خصوص کوچک‌تر از ۱۶ × ۱۶ پیکسل، افت کیفیت تشخیص همچنان مسئله‌ساز است. استفاده از روش‌های افزایش تفکیک‌پذیری تصویر و آموزش چندمقیاسی از گزینه‌هایی است که برای بررسی آینده پیشنهاد شده‌اند.

گسترش نتایج به دیگر کاربردهای صنعتی نیازمند شواهد تجربی مستقل است. گرچه اصول معماری می‌تواند در محیط‌های دیگر هم مفید باشد، تفاوت جنس هدف، دوربین، روشنایی و پس‌زمینه ممکن است نتیجه را دگرگون کند.

مدل فعلی هر فریم را مستقل پردازش می‌کند. استفاده از اطلاعات زمانی و حرکت ذرات در ویدئو می‌تواند مسیر مهمی برای بهبود آشکارسازی باشد؛ زیرا ذره‌ای که در یک فریم پنهان است، ممکن است در فریم‌های قبل یا بعد دیده شود.

توجه تطبیقی که میزان و نوع پردازش خود را با شرایط نور، نوع هدف و تغییرات صحنه تنظیم کند، از جهت افزایش بهره‌وری و حفظ دقت گزینه‌ای قابل بررسی است. چنین رویکردی می‌تواند هزینه پردازش را هنگام ساده‌بودن صحنه کاهش دهد.

نرخ بازیابی ۶۹٫۱ درصد برای یک وظیفه پایش حساس، محدودیت جدی به‌شمار می‌رود. برخی ریزذرات در روشنایی شدید سطح مذاب با پس‌زمینه یکی می‌شوند و مدل در فریم منفرد آن‌ها را نمی‌بیند. پژوهشگران فرض کرده‌اند که پیوستگی زمانی و رهگیری ذرات در جریان ویدئو شاید این مشکل را کاهش دهد؛ اما در مقاله آزمایش رهگیری زمانی ارائه نشده و این احتمال هنوز صرفاً فرضیه است.

۶.۴. تأثیر گسترده‌تر بر خودکارسازی صنعتی

پیشرفت آشکارسازهای بلادرنگ در محیط‌های دشوار می‌تواند به گسترش کاربرد بینایی ماشین در کارخانه و تحقق برخی مؤلفه‌های صنعت ۴٫۰ کمک کند. این پژوهش نمونه‌ای از به‌کارگیری معماری‌های توجه در مسئله واقعی متالورژی است.

ترکیب هوش مصنوعی با فرایندهای سنتی تولید، یکی از مسیرهای توسعه سامانه‌های تولید هوشمند است. نتیجه این مطالعه نشان می‌دهد که برای مسئله مشخصی مانند آشکارسازی ذرات کوچک می‌توان از یادگیری عمیق راه‌حل تخصصی طراحی کرد؛ اما برای پیاده‌سازی در مسائل دیگر ارزیابی مجدد لازم است.

تشخیص بهتر ممکن است در نهایت به کنترل پایدارتر، کاهش اتلاف و کیفیت مطلوب‌تر محصول منجر شود. این پیامدهای اقتصادی و زیست‌محیطی در مقاله به‌عنوان فرصت و ظرفیت فناوری مطرح شده‌اند و جایگزین اندازه‌گیری واقعی شاخص‌های تولید نیستند.

در صورت اعتبارسنجی مناسب، پردازش بلادرنگ می‌تواند داده لازم برای سامانه کنترل حلقه‌بسته فراهم کند تا تغییر شرایط فرایند سریع‌تر شناسایی شود. طراحی چنین حلقه‌ای نیازمند تعیین حد اطمینان، تأیید اپراتور و راهبرد ایمنی مستقل از مدل است.

دسترسی به بخشی از داده‌های پژوهش بنا به درخواست و شرح جزئیات معماری، امکان مطالعه و بازتولید محدود نتایج را برای پژوهشگران بعدی فراهم می‌کند. توسعه و آزمون در سایت‌های صنعتی متفاوت می‌تواند این پایه پژوهشی را تکمیل کند.

۷. نتیجه‌گیری

این مقاله معماری جدیدی برای پایش بلادرنگ افزودن سرباره در ریخته‌گری پیوسته فولاد معرفی کرده که با افزودن سازوکار توجه دومرحله‌ای BiFormer و سازوکار مختصاتی–کانالی CoordSE به YOLOv11 ساخته شده است. آزمایش‌ها نشان داده‌اند که می‌توان هم‌زمان با حفظ نرخ پردازش مناسب روی سخت‌افزار قوی، دقت معیار mAP را نسبت به نسخه پایه افزایش داد.

مدل BiFormer-CoordSEBlock-YOLOv11 به Precision برابر ۸۲٫۵ ± ۰٫۲ درصد، Recall برابر ۶۹٫۱ ± ۰٫۳ درصد و mAP@0.5 برابر ۸۰٫۶ ± ۰٫۲ درصد رسیده است. در مقایسه با YOLOv11 پایه، افزودن BiFormer به‌تنهایی ۲۳٫۴ واحد درصد در mAP بهبود ایجاد کرده و ترکیب کامل نسبت به مدل فقط BiFormer، ۱۲٫۳ واحد درصد بهبود دیگر داشته است. باید توجه داشت که این افزایش‌ها بر حسب «واحد درصد» هستند، نه رشد نسبی درصدی.

یافته نظری اصلی مطالعه آن است که ترکیب توجه سلسله‌مراتبی و توجه آگاه از مختصات می‌تواند برای آشکارسازی اجسام کوچک در تصاویر دشوار صنعتی مؤثر باشد. این نتیجه برای طراحی نسل بعدی سامانه‌های بینایی ماشین متالورژیکی نیز قابل‌استفاده است، به شرط آنکه در مسئله جدید دوباره آزمایش شود.

مدل نهایی روی RTX 4090 سرعت ۴۵٫۲ فریم‌برثانیه داشته است؛ عددی که امکان تحلیل بلادرنگ تصویر را در محیط آزمایش نشان می‌دهد. راهبرد تقویت داده و طراحی ماژول‌ها برای مقابله با نور متغیر و نویز تصویری مفید گزارش شده، اما این نتیجه نباید با تضمین عملکرد یک سامانه کامل در کارخانه اشتباه گرفته شود.

تدوین داده‌های تصویری ویژه ریخته‌گری پیوسته، از دیگر دستاوردهای مطالعه است. اندازه کوچک ذرات، زمینه پرنوسان و شرایط متغیر نور، مجموعه را برای سنجش روش‌های تشخیص صنعتی چالش‌برانگیز کرده‌اند. کل مجموعه به دلیل الزامات محرمانگی کارخانه‌ها به‌صورت آزاد منتشر نشده است.

از منظر صنعتی، این چارچوب بخشی از روند استفاده از هوش مصنوعی برای بهبود پایش و خودکارسازی فرایندهاست. اگر محدودیت‌های فعلی برطرف شود و آزمون عملی موفقیت‌آمیز باشد، ممکن است به کاهش ضایعات و افزایش پایداری کیفیت کمک کند.

گام‌های بعدی پیشنهادی شامل سبک‌سازی ماژول‌های توجه، استفاده از پیوستگی زمانی و اطلاعات ویدئو، و طراحی توجه تطبیقی با شرایط واقعی تولید است. این مسیرها می‌توانند هم نرخ بازیابی را افزایش دهند و هم اجرای مدل روی تجهیزات محدودتر را امکان‌پذیر کنند.

افزودن ماژول‌های توجه پیشرفته به معماری آشکارسازی شناخته‌شده، رویکردی قابل تعمیم برای توسعه بینایی صنعتی است. این پژوهش نشان می‌دهد که اصلاح هدفمند مدل برای یک مسئله واقعی، ممکن است از استفاده بدون تغییر از مدل عمومی موفق‌تر باشد.

نویسندگان صریحاً محدودیت‌های کنونی را می‌پذیرند: بازیابی ۶۹٫۱ درصد به این معناست که در شرایط آشفته ریخته‌گری، برخی ذرات سرباره بسیار کوچک یا پوشیده‌شده از دید مدل پنهان می‌مانند. سرعت ۴۵٫۲ فریم‌برثانیه نیز فقط روی RTX 4090 آزموده شده است. پیش‌بینی امکان استفاده از Jetson AGX Orin، بدون اندازه‌گیری عملی و صرفاً بر مبنای تحلیل نظری است. پژوهش‌های آینده باید کوانتیزه‌سازی مدل و رهگیری زمانی را بررسی کنند تا هم محدودیت سخت‌افزار و هم موارد تشخیص‌نداده‌شده کاهش یابد.

دریافت نمونه‌ای از Add_Slag2025 از نویسندگان و مستندات روش، زمینه‌ای برای پژوهش‌های بعدی فراهم می‌کند. برای رسیدن به نتیجه صنعتی معتبر، آزمون مستقل روی داده‌های واقعی بیشتر و سناریوهای عملیاتی متفاوت لازم خواهد بود.

در نهایت، مقاله نمونه‌ای از پیوند فناوری‌های یادگیری عمیق با فرایندهای سنتی فولادسازی است. دستاورد آن طراحی و ارزیابی یک آشکارساز تخصصی است، نه اثبات راه‌اندازی یک سامانه کنترل صنعتی بدون نظارت. تحقق کاربرد عملی به همکاری متخصصان فرایند، اتوماسیون و بینایی ماشین نیاز دارد.

نقش نویسندگان

ایده‌پردازی، روش تحقیق، توسعه نرم‌افزار، اعتبارسنجی، تحلیل، گردآوری منابع و داده‌ها و تهیه پیش‌نویس اولیه توسط B. Li انجام شده است. بازبینی و ویرایش مقاله با مشارکت P. Cheng و X. Liu بوده؛ تهیه تصاویر توسط X. Liu و سرپرستی و مدیریت پروژه توسط P. Cheng و X. Liu انجام شده است. همه نویسندگان نسخه نهایی مقاله را مطالعه و تأیید کرده‌اند.

تأمین مالی

این پژوهش حمایت مالی خارجی دریافت نکرده است.

بیانیه کمیته اخلاق

بیانیه کمیته اخلاق پژوهش: موضوعیت ندارد.

بیانیه رضایت آگاهانه

بیانیه رضایت آگاهانه: موضوعیت ندارد.

دسترسی به داده‌ها

داده‌های ارائه‌شده در این مطالعه بنا به درخواست از نویسندگان قابل دریافت‌اند.

قدردانی

نویسندگان از Pengfei Cheng به‌دلیل راهنمایی و از خانواده‌های خود به‌سبب حمایت قدردانی کرده‌اند.

تعارض منافع

نویسندگان اعلام کرده‌اند که هیچ تعارض منافعی ندارند.

منابع

1. Louhenkilpi, S. Chapter 1.8—Continuous Casting of Steel. In Treatise on Process Metallurgy ; Seetharaman, S., Ed.; Elsevier: Boston, MA, USA, 2014; pp. 373–434.
2. Sehri, M.; Ayres, L.M.; Boldt, F.d.A.; Dumond, P.; Cuadros, M.A.d.S.L. An Accelerometer-Based Dataset for Monitoring Slag in Steel Manufacturing. BMC Res. Notes 2025 , 18 , 420. DOI
3. Chaubal, P.C. Slag-Metal Reactions during the Continuous Casting of Steel. Iron

.Steelmak. 1997 , 24 , 43–52

.4 Botnikov, S.A.; Khlybov, O.S.; Kostychev, A.N. Development of the Metal Temperature Prediction Model for Steel-Pouring and Tundish Ladles Used at the Casting and Rolling Complex. Metallurgist 2019 , 63 , 792–803. DOI

.5 Lu, Z.; Ren, N.; Xu, X.; Li, J.; Panwisawas, C.; Xia, M.; Dong, H.; Tsang, E.; Li, J. Real-Time Prediction and Adaptive Adjustment of Continuous Casting Based on Deep Learning. Commun. Eng. 2023 , 2 , 34. DOI

.6 Ling, Q.; Isa, N.A.M. Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. IEEE Access 2023 , 11 , 15921–15944. DOI

.7 Li, C.; Huang, H.; Mo, H.; Wen, Z.; Zhou, S.; Zhu, Z. Novel Efficient Steel Surface Defect Detection Model Based on ConvNeXt v2 and Squeeze Aggregated Excitation Attention. IEEE Access 2025 , 13 , 107512–107531. DOI

.8 Tapia, G.; Elwany, A. A Review on Process Monitoring and Control in Metal-Based Additive Manufacturing. J. Manuf. Sci. Eng. 2014 , 136 , 060801. DOI

.9 Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. DOI

.10 Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), .Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475

.11 Jocher, G.; Chaurasia, A.; Qiu, J. YOLOv8: A new state-of-the-art for object detection. arXiv 2023 . DOI

.12 Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. arXiv 2024 , arXiv:2410.17725. DOI

.13 Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. YOLOv9: Learning What You Want to Learn

Using Programmable Gradient Information. arXiv 2024 , arXiv:2402.13616. DOI

14. Wang, Q.; Wang, M.; Sun, J.; Chen, D.; Shi, P. Review of Surface-Defect Detection Methods for Industrial Products Based on Machine Vision. IEEE Access 2025 , 13 , 90668–90697. DOI

15. Messai, O.; Zein-Eddine, A.; Bentamou, A.; Picq, M.; Duquesne, N.; Puydarrieux, S.; Gavet, Y. Small object detection in industrial recycling: A new dataset and YOLO performance evaluation. J. Electron. Imaging 2026 , 35 , 031203. DOI

16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.

17. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R. BiFormer: Vision Transformer with Bi-Level Routing Attention. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 10323–10333.

18. BaoLong, N.; Zhang, C.; Shi, Y.; Hirakawa, T.; Yamashita, T.; Matsui, T.; Fujiyoshi, H. DeBiFormer: Vision Transformer with Deformable Agent Bi-level Routing Attention. In Computer Vision—ACCV 2024 ; Springer: Singapore, 2025; pp. 445–462.

19. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.

20. Peng, C.; Li, B.; Zou, K.; Zhang, B.; Dai, G.; Tsoi, A.C. An Innovative Neighbor Attention Mechanism Based on Coordinates for the Recognition of Facial Expressions. Sensors 2024 , 24 , 7404. DOI

21. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

SteelExperts.ir

22. Zhang, M.; Hu, Y.; Xu, B.; Luo, L.; Wang, S. DSF-YOLO for weld defect detection in X-ray images with dynamic staged fusion. Sci. Rep. 2025 , 15 , 23305. DOI

23. Jobaer, S.; Tang, X.-s.; Zhang, Y.; Li, G.; Ahmed, F. A novel knowledge distillation framework for enhancing small object detection in blurry environments with unmanned aerial vehicle-assisted images. Complex Intell. Syst. 2024 , 11 , 63. DOI

24. Wang, S.; Li, Y.; Qiao, S. ALF-YOLO: Enhanced YOLOv8 based on multiscale attention feature fusion for ship detection. Ocean Eng. 2024 , 308 , 118233. DOI

25. Nikula, R.-P.; Remes, A.; Kaartinen, J.; Kortelainen, J.; Loponen, T.; Ruuska, J.; Ruusunen, M. Autonomous residual monitoring of metallurgical digital twins. Miner. Eng. 2025 , 220 , 109107. DOI

26. Kusiak, A. Smart Manufacturing. Int. J. Prod. Res. 2018 , 56 , 508–517. DOI

27. Tariq, M.F.; Javed, M.A. Small Object Detection with YOLO: A Performance Analysis Across Model Versions and Hardware. arXiv 2025 , arXiv:2504.09900. DOI

28. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv 2022 , arXiv:2209.02976. DOI

29. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2015 , 37 , 1904–1916. DOI

30. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv 2020 , .arXiv:2010.11929

31. Wang, J.; Zhao, H. Improved YOLOv8 Algorithm for Water Surface Object Detection. Sensors 2024 , 24 , 5059. DOI

32. Shi, W.; Zheng, X.; Zhang, L.; Ji, C.; Zhang, Y.; Bian, J. Multi-Object Tracking based on Optimal Transport and Coordinate Attention Mechanism. Signal Process. 2025 ,

236 , 110058. DOI

33. Zhang, X.; Xie, Q.; Sun, W.; Wang, T. Fall detection method based on spatio-temporal coordinate attention for high-resolution networks. Complex Intell. Syst. 2024 , 11 , 1. DOI

34. Lu, L.; He, D.; Deng, Z.; Liu, C.; Shen, Z. MASF-YOLO: An Improved YOLOv11 Network for Small Object Detection on Drone View. arXiv 2025 . DOI

35. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ; IEEE: Honolulu, HI, USA, 2017; pp. 936–944.

یادداشت ناشر: مسئولیت دیدگاه‌ها، داده‌ها و مطالب منتشرشده بر عهده نویسندگان است و ناشر MDPI یا ویراستاران مسئول پیامدهای استفاده از ایده‌ها، روش‌ها و محصولات ذکرشده در مقاله نیستند.

مشخصات منبع: Li, B.; Cheng, P.; Liu, X. (2026). *Enhanced YOLOv11 with BiFormer and CoordSE for Real-Time Steel Continuous Casting Slag Object Detection.* Applied Sciences, 16(8), 3965. https://doi.org/10.3390/app16083965. ترجمه و تنظیم فارسی؛ نمودارها و داده‌ها متعلق به منبع اصلی‌اند. مجوز CC BY 4.0.

نسخه آنلاین مقاله: https://landing.steelexperts.ir/2026/10/27/sx-r045-yolov11-biformer-coordse-slag-detection/