مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال یازدهم، شماره چهارم، بهار ۱۴۰۴

بهبود بازشناسی پلاک خودرو با استفاده از رویکرد مبتنی بر مدل زبانی

الهام شعبانی نیا[۱]، فاطمه اسدی زیدآبادی[۲] و حسین نظام‌آبادی پور[۲]

چکیده

از گذشته بازشناسی پلاک خودرو به دلیل طیف وسیعی از کاربردها، از جمله کنترل ترافیک، جمع‌آوری عوارض، جستجوهای پلیس و غیره توجه زیادی از جامعه تحقیقاتی را به خود جلب کرده است. پلاک‌ها معمولاً بسته به کشور یا منطقه صدور آن‌ها با الگوهای خاصی طراحی می‌شوند. تفسیر این الگوها برای بهبود دقت بازشناسی حروف به‌ویژه در محیط‌های پیچیده و بدون محدودیت بسیار مهم است. بااین‌وجود روش‌های موجود برای بازشناسی پلاک خودرو معمولاً به شکل دستی این قواعد و الگوها را در نظر می‌گیرند. حال‌آنکه شناخت و درک خودکار «زبان» پلاک‌ها که شامل ترکیبات منحصربه‌فردی از حروف و ارقام می‌شود، می‌تواند عملکرد این سامانه‌ها را به طور قابل‌توجهی بهبود دهد و به آن‌ها اجازه می‌دهد تا به شکل پویا برای مجموعه حروف خاص مورداستفاده در مناطق مختلف طراحی شوند. در این مقاله باتکیه بر موفقیت‌های اخیر رویکردهای مبتنی بر ترنسفورمر در پردازش زبان طبیعی و بینایی رایانه، یک مدل جدید برای بازشناسی پلاک خودرو پیشنهاد شده است که از یک مدل بینایی برای بازشناسی تصاویر پلاک خودرو و یک مدل زبانی برای یادگیری قواعد موجود در پلاک‌ها و تصحیح چندمرحله‌ای نتایج بازشناسی بهره می‌برد. این فرآیند باعث بهبود دقت بازشناسی می‌شود و سازگاری روش را در مواجهه با انواع پلاک‌ها با الگوهای مختلف ارتقا می‌دهد. ارزیابی‌های انجام شده بر روی مجموعه‌داده‌های IR-LPR، UFPR-ALPR و AOLP برتری رویکرد پیشنهادی را در مقایسه با تحقیقات انجام شده نشان می‌دهد. این روش می‌تواند در توسعه سامانه‌های بازشناسی پلاک دقیق و سازگار مفید باشد.

کلیدواژه‌ها

بازشناسی خودکار پلاک خودرو، سامانه‌های حمل‌ونقل هوشمند، یادگیری عمیق، مدل زبانی، ترنسفورمر.

۱- مقدمه

بازشناسی خودکار پلاک خودرو یک فناوری به نسبت قدیمی در بینایی ماشین است که از بازشناسی نوری نویسه‌ها (OCR) برای خواندن و تجزیه‌وتحلیل اطلاعات پلاک خودروها استفاده می‌کند. این سامانه از دوربین‌های دیجیتال برای ضبط تصاویر پلاک خودرو و سپس از نرم‌افزار برای تفسیر تصاویر و بازشناسی حروف الفبایی و عددی روی پلاک استفاده می‌کند [۱, ۲]. بازشناسی پلاک خودرو در زمینه‌های بسیار مختلفی کاربرد دارد. به‌عنوان‌مثال، در سازمان‌های ناظر بر کنترل ترافیک از آن برای اجرای محدودیت‌های زمانی، کاهش ازدحام و بهبود جریان ترافیک استفاده می‌شود. همچنین از این سامانه برای مدیریت پارکینگ‌ها و صدور خودکار بلیط استفاده می‌شود. باجه‌های جمع‌آوری عوارض نیز از بازشناسی پلاک خودرو برای خودکارسازی وصول عوارض استفاده می‌کنند. در مسائل مربوط به اعمال قانون و جستجوی مجرمان نیز برای شناسایی وسایل نقلیه دزدیده شده، شناسایی مجرمان تحت تعقیب و وسایل نقلیه مرتبط با جرایم استفاده می‌شود [۳, ٤].

بازشناسی خودکار پلاک خودرو شامل چند مرحله اصلی است (شکل (۱)) که با استفاده از روش‌های مختلف پردازش تصویر و یادگیری ماشین انجام می‌شوند.

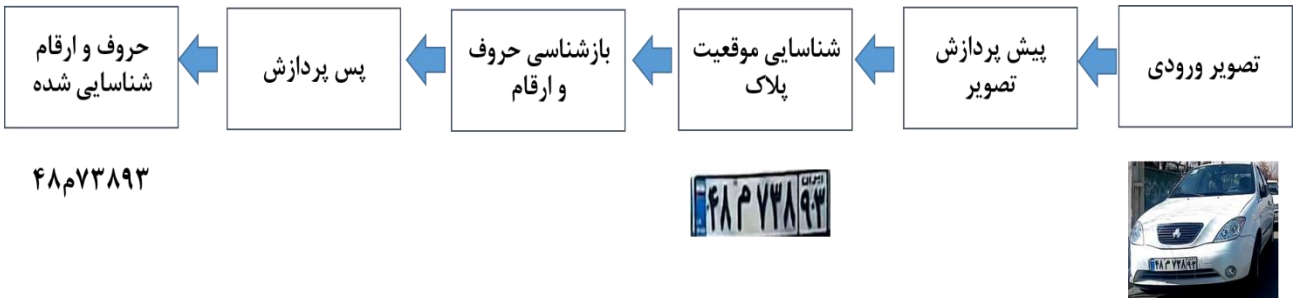
این مقاله در آبان‌ماه ۱٤۰۳ دریافت شد؛ در بهمن‌ماه بازنگری و در اسفندماه پذیرفته گردید. این پژوهش با پشتیبانی مالی اداره کل زندان‌های استان کرمان بر اساس قرارداد شماره ۰۲/٤۳۲/۱۰/۱۷۹۳٦ انجام شده است.

[۱] بخش ریاضی کاربردی، دانشکده علوم و فنون نوین، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران،
رایانامه: e.shabaninia@kgut.ac.ir

[۲] پژوهشگر ارشد در آزمایشگاه پردازش هوشمند داده، دانشگاه شهید باهنر کرمان
رایانامه: fatemeh.asadi.z1998@gmail.com

[۳] گروه مهندسی برق، دانشگاه شهید باهنر کرمان
رایانامه: nezam@uk.ac.ir

نویسنده مسئول: الهام شعبانی‌نیا



شکل (۱) : اجزای مهم یک سامانه بازشناسی خودکار پلاک خودرو.

مرحله پیش‌پردازش، معمولا وظایفی همچون تصحیح نور، حذف نویز، بهبود کیفیت تصویر، مقیاس‌بندی و یا تبدیل رنگ را برعهده می‌گیرد تا پردازش‌های بعدی راحت‌تر و سریع‌تر انجام شوند. در مرحله شناسایی پلاک، ناحیه‌ای که پلاک در آن قرار دارد شناسایی می‌شود. پس از شناسایی ناحیه پلاک، ممکن است نیاز باشد حروف موجود در آن از یکدیگر جدا شوند. در مرحله بازشناسی نیز معمولاً از روش‌های بازشناسی نوری نویسه‌ها برای تفسیر حروف و اعداد موجود بر روی پلاک استفاده می‌شود. در نهایت در مرحله پس‌پردازش، نتایج حاصل از شناسایی حروف بر اساس قالب استاندارد پلاک‌ها در یک کشور یا منطقه، اصلاح و اعتبارسنجی می‌شوند. در این پژوهش، به طور خاص بازشناسی و پس‌پردازش پلاک مورد توجه قرار گرفته است.

بازشناسی پلاک خودرو در چند دهه اخیر توجه زیادی از جامعه تحقیقاتی را به خود معطوف کرده است. در این مدت، روش‌های مختلفی برای حل مشکلات گوناگون مانند شناسایی ناحیه پلاک، جداسازی حروف، بازشناسی تمام حروف و ارقام تحت شرایط مختلف ارائه شده است [۵].

اغلب روش‌های سنتی متکی بر استفاده از ویژگی‌های دست-ساز برای پلاک خودرو مانند شکل و اندازه، نوع قلم الفبایی-عددی، اندازه و تعداد ارقام، اطلاعات پس‌زمینه، رنگ متن و حتی خواص بازتابی پلاک‌های استاندارد هستند. متأسفانه این روش‌ها توانایی محدودی در بازشناسی پلاک خودرو دارند [۸-۶].

در مقابل روش‌های مبتنی بر یادگیری عمیق[۱] و به‌ویژه شبکه-های ترنسفورمری[۲] به طور خودکار ویژگی‌های مؤثر را از داده‌ها یاد گرفته و اخیراً به نتایج امیدوارکننده‌ای دست یافته‌اند. رویکردهای یادگیری عمیق در مقایسه با روش‌های سنتی دارای مزایای متعددی مانند دقت بالا، استحکام در برابر تغییرات مختلف و توانایی کار با مجموعه حروف پیچیده و سبک‌های مختلف قلم هستند [۹, ۱۰].

بااین‌وجود، یکی از چالش‌های اصلی در سامانه‌های بازشناسی پلاک خودرو، نبود استاندارد مشخص در طراحی پلاک‌ها است که توسعه سامانه‌های جهانی بازشناسی پلاک را با مشکل مواجه می‌سازد. تنوع در قالب‌ها و زبان‌های مورداستفاده برای پلاک خودرو در کشورهای مختلف، پیچیدگی‌هایی در توسعه سامانه‌های سازگار با تشخیص پلاک خودکار ایجاد می‌کند (شکل (۲)). برای مثال، در ایالات متحده، هر ایالت طراحی، مجموعه حروف و روش شماره‌گذاری خاص خود را دارد. در ایران نیز پلاک‌ها شامل ترکیبی از اعداد، حروف و بخشی کوچک برای نشان دادن شهر مالک خودرو هستند. این تفاوت‌ها سامانه‌های بازشناسی پلاک خودرو را ملزم به شناسایی طیف وسیعی از الگوها و حروف می‌کنند.

برای غلبه بر این چالش‌ها، همان‌طور که پیش‌تر اشاره شد، محققان معمولاً پس از مرحله بازشناسی، از یک مرحله پس‌پردازش استفاده می‌کنند تا با اعمال قوانین و الگوهای دستی تنظیم شده برای هر منطقه یا کشور، خطاهای احتمالی را اصلاح کنند. استفاده از قوانین خاص هر کشور در قالب پلاک‌ها (مانند تعداد حروف مورد انتظار، پیشوندها یا پسوندهای شناخته‌شده) می‌تواند به کاهش خطاهای OCR کمک کند. به‌عنوان‌مثال، می‌توان از الگوریتم‌های تطبیق رشته برای اصلاح اشتباهات رایج بر اساس الگوهای شناخته‌شده پلاک‌های خودرو بهره برد. بااین‌وجود، این روش متکی بر شناخت دقیق قواعد و الگوهای هر کشور یا منطقه است که بایستی به شکل دستی در یک فرآیند که معمولاً زمان‌بر است انجام شود [۱۱].

در سال‌های اخیر، استفاده از مدل‌های زبانی در حوزه پردازش زبان طبیعی به نتایج چشمگیری منجر شده است [۱۲]. اگر فرایند یادگیری الگوها و قواعد موجود در پلاک‌ها را مشابه یادگیری زبان در نظر بگیریم، می‌توان این قواعد را از داده‌های آموزشی به‌طور ضمنی و خودکار استخراج کرد. بنابراین استفاده از مدل زبانی برای یادگیری "زبان پلاک‌ها" می‌تواند به عنوان مرحله پس‌پردازش مورد استفاده قرار گیرد تا خطاهای احتمالی را که به دلایل مختلف بر دقت بازشناسی اثر می‌گذارند، اصلاح کند. این پژوهش به چالش‌های مرتبط با بازشناسی حروف به دلیل نبود استاندارد در طراحی پلاک خودرو و تنوع کیفیت تصاویر می‌پردازد.

---

[۱] Deep learning

[۲] Transformer

شکل (۲): تنوع در الگوهای مورد استفاده برای پلاک خودروها (به دلیل تنوع پلاک در برخی کشورها، پلاک نشان داده شده لزوما نماینده پلاک‌های آن کشور نیست).

در این مطالعه، یک شبکه ترنسفورمری مبتنی بر یادگیری عمیق به نام ABINet [۱۳] مبتنی بر مدل زبانی برای نخستین بار برای بازشناسی پلاک خودرو معرفی می‌گردد. شبکه پیشنهادی پس از آموزش قادر است ساختار خاص پلاک‌ها را به شکل خودکار از روی نمونه‌های آموزشی یادگرفته و با استفاده از مدل زبانی نتایج OCR را اصلاح کند.

این موضوع به‌ویژه در محیط‌های چالش‌برانگیز و بدون محدودیت بسیار حائز اهمیت است. نتایج این تحقیق، نه‌تنها عملکرد سامانه‌های بازشناسی پلاک در ایران را در زمینه‌هایی نظیر مدیریت مرزی و ترددهای بین‌المللی بهبود می‌بخشد، بلکه قابلیت استفاده از آن در سطح بین‌المللی را نیز فراهم می‌سازد.

ازاین‌رو نوآوری‌های کلیدی مقاله عبارت است از:

- حذف نیاز به جداسازی حروف که احتمال خطا در شناسایی را کاهش می‌دهد.
- یادگیری خودکار زبان پلاک‌ها برای بهبود نتایج OCR.
- استفاده از مدل ABINet [۱۳] برای نخستین بار در حل مسئله پلاک‌خوانی.

ازاین‌رو ادامه مقاله به شرح زیر سازماندهی شده است: بخش ۲ رویکردهای مرتبط با بازشناسی خودکار پلاک خودرو را بررسی می‌کند. بخش ۳ روش پیشنهادی برای بازشناسی پلاک خودرو را شرح می‌دهد. بخش ٤ نتایج ارزیابی را ارائه می‌دهد و بخش ٥ مقاله را با جمع بندی به پایان می‌رساند.

۲- پیشینه تحقیق

همان‌طور که پیش‌تر اشاره شد، حوزه بازشناسی پلاک خودرو به دلیل گستردگی کاربردهای آن، فرصت‌های توسعه قابل‌توجهی را تجربه کرده است. این موضوع باعث افزایش تلاش‌ها برای رسیدگی به سناریوهای بدون محدودیت و در نتیجه گسترش کاربردهای سامانه‌های بازشناسی پلاک خودرو شده است.

به همین دلیل، تعداد زیادی از تحقیقات در این زمینه انجام شده است که مرور کامل تمام آن‌ها عملاً بسیار دشوار و خارج از توان این مقاله است. در اینجا، باتوجه‌به تمرکز مقاله بر مرحله بازشناسی، تنها به بررسی تحقیقات اخیر و برجسته در این زمینه پرداخته شده است.

همان‌طور که قبلاً ذکر شد، این مرحله وظیفه بازشناسی حروف الفبایی و عددی محدوده پلاک را بر عهده دارد. ورودی این واحد ناحیه جدا شده از تصویر اصلی است که تنها پلاک خودرو را شامل می‌شود و خروجی، متن موجود در پلاک است. برخی روش‌های موجود مبتنی بر جداسازی هستند که ناحیه مربوط به هر حرف یا عدد را قبل از ارسال به واحد بازشناسی جدا می‌کنند. در مقابل روش‌های مستقل از جداسازی کل ناحیه پلاک را به واحد بازشناسی ارسال می‌کنند. به لطف توسعه فناوری یادگیری عمیق در سال‌های اخیر، روش‌های بدون ناحیه‌بندی به‌تدریج به جریان اصلی تبدیل شده‌اند. البته برخی روش‌ها نیز شناسایی و بازشناسی را توامان مشابه با وظیفه شناسایی شی انجام می‌دهند [۱٤].

۲-۱- روش‌های مبتنی بر جداسازی[۱]

در این روش‌ها، معمولاً تضاد آشکار بین پس‌زمینه و حروف برای جداسازی در نظر گرفته می‌شود. باینری کردن تصویر، این جداسازی را آسان‌تر می‌کند. فناوری‌های جداسازی تصویر موجود شامل جداسازی حروف با استفاده از اتصال پیکسل [۱٥]، جداسازی حروف با استفاده از خطوط پیش‌بینی [۱٦] و جداسازی حروف با استفاده از دانش قبلی [۱۷] است. پس از جداسازی، به طور کلی، سه رویکرد اصلی برای طبقه‌بندی حروف مورد توجه است: استفاده از فناوری تطبیق الگو، بازشناسی با استفاده از ویژگی‌های دستی استخراج شده و یادگیری عمیق.

فناوری تطبیق الگو [۱۸] بر این واقعیت استوار است که پلاک‌ها معمولاً قلم‌ها و اندازه حروف ثابتی دارند. با این حال، در یک محیط بدون محدودیت، حروف پلاک ممکن است با چرخش و کشیدگی همراه باشند و یا کج شده باشند. در این حالت بسیاری از الگوهای اضافی نیز برای دستیابی به بازشناسی با دقت بالا بایستی ذخیره شوند. استخراج ویژگی [۱۹]، ویژگی‌های برجسته حروف را از تصویر استخراج می‌کند. این روش هزینه محاسبه را

[۱] Segmentation-Based Methods

کاهش می‌دهد و در برابر چرخش حروف و نویز نیز مقاومت بهتری داراست.

روش‌های یادگیری عمیق [۲۰] می‌توانند مستقیماً ویژگی‌های حروف را از تصویر به دست آورند. در [۲۱]، یک الگوریتم یادگیری با نظارت ضعیف برای شناسایی شماره پلاک خودرو ارائه شده است که پس از جداسازی نویسه‌های پلاک، از یک شبکه عصبی پیچشی[۱](CNN) برای طبقه‌بندی و بازشناسی نویسه‌ها استفاده می‌کند. در [۲۲] نیز پس از مکان‌یابی پلاک خودرو و جداسازی حروف و اعداد در آن، در نهایت از شبکه پیچشی AlexNet برای دسته‌بندی حروف استفاده می‌شود. در [۲۳]، واحد شناسایی پلاک از یک مدل شناسایی شی (YOLOv3) استفاده می‌کند که برای شناسایی پلاک‌ها در تصاویر آموزش دیده است. بازشناسی حروف نیز از یک رویکرد مبتنی بر CNN برای بازشناسی موجودیت‌های منفرد استخراج‌شده از تصویر پلاک خودرو استفاده می‌کند.

به‌طورکلی در روش‌های مبتنی بر جداسازی نتایج طبقه‌بندی مستقیماً تحت‌تأثیر دقت جداسازی حروف قرار می‌گیرند. حال‌آنکه جداسازی تک‌تک نویسه‌ها به دلایلی همچون زوایای دید مختلف، وجود نویز و غیره معمولاً بسیار چالش‌برانگیز است.

۲-۲- روش‌های بدون جداسازی[۲]

به‌منظور اجتناب از عدم قطعیت روش‌های جداسازی، اخیراً بسیاری از محققان از روش بدون جداسازی برای بازشناسی حروف پلاک خودرو استفاده می‌کنند. به‌عنوان‌مثال، شبکه‌های عصبی بازگشتی[۳]به دلیل اتصال قدرتمند، مزایای طبیعی یادگیری توالی را دارا هستند و به طور گسترده‌ای در بازشناسی گفتار، بازشناسی متن، مدل‌سازی زبان و ترجمه ماشینی استفاده می‌شوند.

وانگ و همکارانش [۲٤] از یک شبکه عصبی بازگشتی دو جهته[۴] و ساختار طبقه‌بندی زمانی ارتباطی[۵](CTC) برای شناسایی حروف پلاک خودرو به صورت انتها به انتها استفاده کردند. مرجع [۲۵]، مدل بازشناسی حروف را به دو بخش تقسیم کرد. قسمت اول از شبکه عصبی پیچشی برای استخراج ویژگی های تصویر استفاده می‌کند و قسمت دوم از LSTM و CTC برای به دست آوردن اطلاعات زمینه و سپس رمزگذاری و رمزگشایی آن استفاده می‌کند. مرجع [۲٦] نیز از توجه دو بعدی و یک رمزگذار بر اساس CNN و سپس از مدل LSTM برای رمزگشایی ویژگی‌های تصویر به دست آمده استفاده کرد. نویسندگان در [۵]، روشی بر اساس یادگیری عمیق برای افزایش وضوح‌حروف و خواندن پلاک خودروهای ایرانی ارائه کردند که از شبکه‌های عصبی بازگشتی با ساختار رمزگذار-رمزگشا استفاده می‌کند. شبکه پیشنهادی می‌تواند با یادگیری تصاویر پلاک خودرو در شرایط متنوع، حروف پلاک خودرو را برجسته نماید. پس از آن، حروف پلاک از روی تصویر دودویی شده با استفاده ازشبکه‌های عصبی بازگشتی و بدون نیاز به جداسازی، خوانده می‌شوند. این کار می‌تواند خطای ناشی از جداسازی و ناحیه‌بندی را کاهش دهد.

به‌طورکلی این دسته از روش‌ها به دلیل عدم نیاز به جداسازی نویسه‌ها، با خطای ناشی از ناحیه‌بندی مواجه نیستند، اما در مقابل معمولاً به پردازش محاسباتی بالایی نیاز دارند.

۳- روش پیشنهادی

در اینجا روش استفاده شده برای بازشناسی متن از مرجع [۱۳] گرفته شده است. این روش با استفاده از یک مدل بینایی[۶]مبتنی بر ترنسفورمری (VM) و یک مدل زبانی[۷](LM) وظیفه بازشناسی متون را انجام می‌دهد. در این پژوهش برای اولین بار این شبکه برای بازشناسی پلاک خودرو استفاده شده است. این مدل که به شکل انتها به انتها آموزش داده می‌شود، قادر است مرحله پس-پردازش برای تصحیح متن را به شکل خودکار از طریق یادگیری قواعد موجود در پلاک‌های آموزشی یاد بگیرد. ساختار کلی معماری پیشنهادی در شکل (۳) نشان داده شده است. در ادامه واحدهای اصلی مربوط به این معماری توضیح داده شده است.

۳-۱- مدل بینایی

مدل بینایی از یک شبکه عصبی پیچشی (ResNet45) [۲۷] برای استخراج ویژگی و یک شبکه مبتنی بر ترنسفورمری ViT [۲۸, ۲۹] برای مدل سازی توالی تشکیل شده است. در نتیجه داریم:

$$F_b = R(x) \in R^{H/4 \times W/4 \times C} \qquad (۱)$$

که در آن تصویر x با عرض W و ارتفاع H را به‌عنوان ورودی مدل دریافت کرده و پس از نگاشت با شبکه پیچشی R ویژگی‌های پیچیده و مفهومی تصویر F_b استخراج می‌شوند. در اینجا C نشان‌دهنده تعداد کانال‌های ویژگی استخراج شده است.

پس از آن از یک فرآیند مبتنی بر توجه برای یادگیری دنباله ویژگی-های استخراج شده استفاده می‌شود. این فرآیند بر اساس الگوی پرس‌وجو [۲۸] انجام می‌شود:

$$F_V = softmax\left(\frac{QK^T}{\sqrt{C}}\right)V \qquad (۲)$$

به طور دقیق‌تر، $Q \in R^{T \times C}$ نشان‌دهنده کدگذاری‌های موقعیتی [۲۸] برای ترتیب است و T طول توالی را نشان می‌دهد.

---

[۱] Convolutional Neural Network

[۲] Segmentation-Free Methods

[۳] Recurrent Neural Network (RNN)

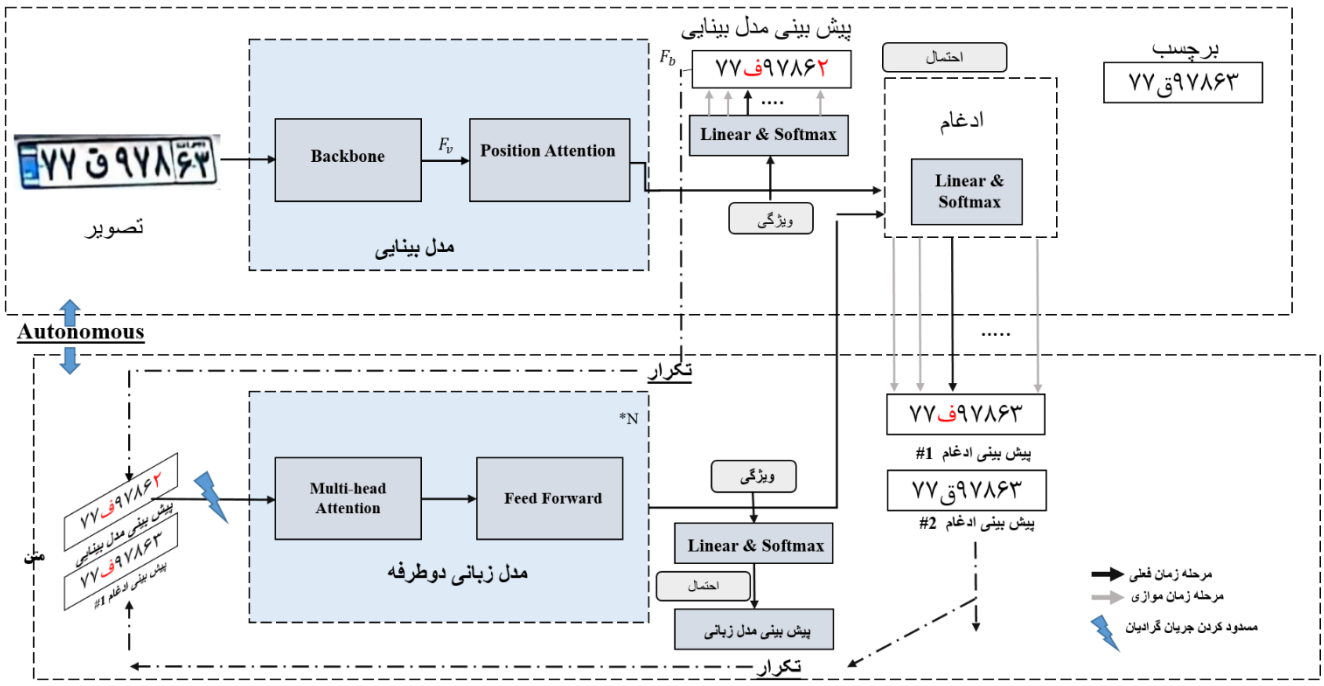
[۴] Bidirectional recurrent neural networks (BRNN)

[۵] Connectionist Temporal Classification

[۶] Vision Model

[۷] Language Model

[۸] Positional Encodings



شکل (۳): ساختار کلی معماری پیشنهادی برای بازشناسی و تصحیح پلاک خودرو. این معماری از دو بخش بینایی-پیچشی و مدل زبانی تشکیل شده است [۱۳].

در این مدل، $K = g(F_b) \in R^{\frac{HW}{16} \times C}$ نشان‌دهنده‌ی ویژگی‌هایی است که از طریق تابع g(.) به‌دست می‌آیند و این تابع با استفاده از یک شبکه U-Net [۳۰] پیاده‌سازی شده است، تا بتواند ویژگی‌های مهم تصویر را استخراج کند.

از سوی دیگر، $V = H(F_b) \in R^{\frac{HW}{16} \times C}$ نشان‌دهنده‌ی ویژگی‌هایی است که از طریق تابع H(.) به‌دست می‌آیند؛ این تابع یک نگاشت همانی است که بدون تغییر، ویژگی‌های[۱] اولیه را به‌عنوان ورودی به بخش‌های بعدی مدل انتقال می‌دهد. این سازوکار به مدل اجازه می‌دهد تا روابط بین توالی حروف و ویژگی‌های بصری را به‌خوبی درک کند و در نتیجه عملکرد بهتری در پردازش داده‌های تصویری و متنی ارائه دهد.

۳-۲-مدل زبانی

در [۱۳]، یک شبکه BCN به‌عنوان مدل زبانی به کار گرفته شده است که در واقع نسخه‌ای اصلاح‌شده از رمزگذار ترنسفورمر با L لایه می‌باشد. هر لایه در BCN شامل مکانیزم‌های توجه چندسر ضربدری و شبکه‌های تغذیه پیشرو است که به همراه اتصالات باقیمانده و نرمال‌سازی لایه عمل می‌کنند.

برخلاف ترنسفورمر استاندارد، در BCN بردارهای یک توالی به طور مستقیم به بلوک‌های توجه چندسر وارد می‌شوند و از لایه اول شبکه عبور داده می‌شوند. همچنین، ماسک توجه در این مکانیزم به‌گونه‌ای طراحی شده است که مدل نتواند به همان گام زمانی دسترسی پیدا کند که از ارجاع به خود جلوگیری می‌کند.

BCN همچنین توجه به خود را حذف می‌کند تا از نشت اطلاعات بین گام‌های زمانی مختلف جلوگیری شود. در نتیجه عملکرد بهتری در پردازش زبان طبیعی و تحلیل متن ارائه دهد. این راهبرد به طور مؤثر مدل زبانی را به‌عنوان یک سامانه مستقل برای تصحیح املاء در نظر می‌گیرد، جایی که بردارهای احتمالات حرف را پردازش کرده و توزیع‌های احتمالی حروف را تولید می‌کند.

این مدل لایه‌های رمزگذار ترنسفورمر را ادغام می‌کند و به‌گونه‌ای طراحی شده است که بردارهای حرف از مرحله قبل، مستقیماً به بلوک‌های توجه چندسر وارد شوند. مکانیزم توجه در BCN به‌صورت زیر تعریف می‌شود:

$$M_{ij} = \begin{cases} 0, & i \neq j \\ -\infty, & i = j \end{cases}$$
$$K_i = V_i = P(y_i)W_L \qquad (۳)$$
$$F_{mha} = Softmax\left(\frac{QK^T}{\sqrt{C}} + M\right)V$$

در اینجا، Q نشان‌دهنده کدگذاری‌های مکانی یا خروجی‌های لایه آخر است. K و V از احتمال کاراکتر $P(y_i)$ برای رشته متنی $y = (y_1, \ldots, y_n)$ به‌دست می‌آید. ماتریس W_L ماتریس نگاشت خطی است که احتمال حروف را به فضای برداری C منتقل می‌کند و M ماتریس ماسک توجه را نشان می‌دهد.

معماری BCN که مشابه ترنسفورمر است، این امکان را فراهم می‌آورد که محاسبات به‌صورت مستقل و موازی انجام شود و در مقایسه با مدل‌های تجمیع‌شده، پارامترها و محاسبات کمتری را نیاز دارد. این ویژگی‌ها باعث می‌شود که BCN نه‌تنها کارآمد باشد، بلکه توانایی پردازش داده‌ها را نیز افزایش دهد.

[۱] Identity Mapping
[۲] Bidirectional Contextual Network

برای مدیریت ورودی‌های نویزی، مدل زبانی به مدت M تکرار اجرا می‌شود. در تکرار اولیه، خروجی معادل پیش‌بینی احتمال از مدل بینایی است. در تکرارهای بعدی، این خروجی منعکس‌کننده پیش‌بینی احتمال از مدل ادغام شده است که از تکرار قبلی به‌دست‌آمده است.

این فرآیند تکراری به مدل زبانی اجازه می‌دهد تا پیش‌بینی‌های مدل بینایی را به طور تدریجی تصحیح و بهبود بخشد. با هر تکرار، مدل می‌تواند به طور مؤثرتری به نویزهای موجود در ورودی پاسخ دهد و دقت پیش‌بینی‌های خود را افزایش دهد. این رویکرد نه‌تنها به بهبود کیفیت خروجی کمک می‌کند، بلکه به مدل اجازه می‌دهد تا از تجربیات قبلی خود یاد بگیرد و در برابر ورودی‌های متغیر و پیچیده واکنش مناسب‌تری نشان دهد.

این معماری به‌گونه‌ای طراحی شده است که گرادیان‌ها میان مدل زبانی و مدل بینایی مسدود می‌شوند. این امر تضمین می‌کند که مدل زبانی قواعد زبانی را به‌صورت مستقل و بدون وابستگی به ویژگی‌های بصری یاد بگیرد. درعین‌حال، مدل بینایی همچنان از طریق نظارت مستقیم و گرادیان‌های حاصل از مدل ادغام‌شده (Fusion) آموزش می‌بیند تا فضای ویژگی‌های خود را با مدل زبانی هم‌راستا کند. این رویکرد نه‌تنها به بهبود عملکرد هر دو مدل کمک می‌کند، بلکه از نشت اطلاعات نامرتبط در فرایند یادگیری جلوگیری می‌کند.

٤- آزمایش‌ها و نتایج

٤-١-پایگاه‌های داده

آزمایش‌ها با استفاده از سه پایگاه‌داده عمومی و در دسترس انجام می‌شوند: IR-LPR [۳۱]، AOLP [۳۲] و UFPR-ALPR [۳۳]. یک نمای کلی از این پایگاه‌های داده در جدول (۱) ارائه شده است.

جدول (۱):مروری بر پایگاه‌های داده مورد استفاده در آزمایش‌ها.

پایگاه‌داده	سال	تصاویر	پروتکل ارزیابی	طرح‌بندی LP
IR-LPR	۲۰۲۲	٤٨٧١٢	بله	ایرانی
UFPR-ALPR	۲۰۱۸	٤٥٠٠	بله	برزیلی
AOLP	۲۰۱۳	۲۰٤۹	خیر	تایوانی

تصاویر نمونه از این پایگاه‌داده‌ها نیز در جدول (۲) قابل مشاهده است.

۱) پایگاه‌داده *IR-LPR*

پایگاه‌داده IR-LPR برای اهداف مختلفی از جمله مکان‌یابی پلاک خودرو و بازشناسی حروف ایجاد شده است.

این پایگاه‌داده شامل سه مجموعه از تصاویر است. مجموعه اول ۲۰۹٦۷ تصویر از تصاویر صحنه که در آن‌ها یک یا چند خودرو در حال حرکت هستند را شامل می‌شود. تصاویر این مجموعه از محیط‌های متنوعی مانند چندین وب‌سایت فروش خودرو، پارکینگ‌ها و غیره در زمان‌های مختلف روز و تحت شرایط نوری گوناگون گردآوری شده‌اند. فاصله‌های ثبت تصاویر از ۱ متر تا بیش از ۱۰ متر متغیر بوده و اغلب تصاویر با استفاده از دوربین‌های تلفن همراه با کیفیت وضوح و فرمت‌های مختلف ضبط شده‌اند.

مجموعه دوم و سوم تنها تصاویر پلاک خودرو را شامل می‌شوند که در مجموع مشتمل بر ٤٨,٧١٢ تصویر هستند. تفاوت این دو مجموعه در آن است که تصاویر مجموعه سوم واقعی نیستند. بلکه پلاک‌ها به شکل مصنوعی از طریق قرار دادن اعداد و حروف روی پس‌زمینه‌های پلاک ایجاد شده‌اند. از آن‌جا که در این پژوهش روی مرحله بازشناسی تمرکز شده است، تنها از تصاویر این دو مجموعه آخر برای آموزش مدل پیشنهادی برای وظیفه بازشناسی استفاده شده است.

این پایگاه‌داده به شرایط مختلف جوی و نوری، فاصله دوربین تا خودرو، زاویه دید پلاک‌ها و استفاده از دوربین‌هایی با کیفیت‌های متفاوت توجه ویژه‌ای دارد. بااین‌حال متأسفانه برچسب پلاک‌ها با خطای زیادی همراه است که در این پژوهش کلیه پلاک‌ها مجدد مورد بازبینی قرار گرفته است. علاوه بر این تعداد ۳,۷۷۰تصویر از تصاویر موجود به دلیل کیفیت پایین حذف شده است.

۲) پایگاه‌داده *AOLP*

این پایگاه‌داده شامل ۲,۰٤۹ تصویر از خودروها است که برای ارزیابی سامانه‌های شناسایی پلاک طراحی شده است. این پایگاه داده به سه زیرمجموعه اصلی تقسیم می‌شود:

زیر مجموعه (AC)[۱]: شامل ٦۸۱ تصویر از خودروها است که در شرایط کنترل‌شده مانند ورودی‌ها و خروجی‌های پارکینگ‌ها گرفته شده‌اند.

زیرمجموعه (LE)[۲]: شامل ۷۵۷ تصویر از خودروهای در حال حرکت در جاده‌ها و بزرگراه‌ها است که برای کاربردهای نظارتی و اجرای قانون استفاده می‌شود.

زیرمجموعه (RP)[۳]: شامل ٦۱۱ تصویر از خودروها است که برای شبیه‌سازی شرایط گشت جاده‌ای جمع‌آوری شده‌اند.

هر زیرمجموعه شامل تصاویر متنوعی از نظر زاویه دید، فاصله دوربین تا خودرو، شرایط جوی و نوری است. تمام تصاویر به‌دقت

[۱] Access Control
[۲] Traffic Law Enforcement
[۳] Road Patrol

حاشیه‌نویسی شده‌اند و موقعیت و محتوای پلاک‌ها در آن‌ها مشخص شده است که این مجموعه را برای توسعه و ارزیابی الگوریتم‌های شناسایی پلاک و بازشناسی حروف مناسب می‌سازد.

۳) پایگاه‌داده *UFPR-ALPR*

یک پایگاه‌داده برزیلی توسط لاروکا و همکاران گردآوری شده است. این پایگاه‌داده شامل تصاویر ۳۰۰ وسیله نقلیه است. هر یک از تصاویر، با وضوح ۱۰۸۰ × ۱۹۲۰ پیکسل، چندین عکس از یک وسیله نقلیه در حال حرکت را نشان می‌دهند.

این پایگاه‌داده، شامل ٤٥۰۰ تصویر است که به‌ویژه برای توسعه سامانه‌های بازشناسی پلاک خودرو که برای استقرار در وسایل نقلیه پلیس در نظر گرفته می‌شوند، ارزشمند است. پایگاه داده UFPR-ALPR به دلیل استفاده از سه دوربین غیر استاتیک مختلف که تصاویر انواع وسایل نقلیه مختلف (ماشین‌ها، موتورسیکلت‌ها، اتوبوس‌ها و کامیون‌ها) را در پس‌زمینه‌های پیچیده و در شرایط نوری متنوع ثبت می‌کند، چالش‌برانگیز است.

علاوه بر این، تصاویر این پایگاه، قاب‌های ویدیویی از تعداد محدودی خودرو را شامل می‌شوند.

٤-۲-معیارهای ارزیابی

در سامانه‌های بازشناسی، ارزیابی یک مرحله حیاتی است؛ زیرا عملکرد و دقت سامانه را تعیین می‌کند. در این مقاله، برای ارزیابی روش پیشنهادی از معیار دقت به‌عنوان یکی از کلیدی‌ترین شاخص‌های ارزیابی عملکرد سامانه استفاده شده است. این معیار بیانگر درصدی از کل پلاک‌های شناسایی شده است که به‌درستی بازشناسی شده‌اند. این دقت تحت‌تأثیر عواملی همچون کیفیت تصویر قرار دارد. این معیار به‌صورت رابطه (٤) قابل‌بیان است [٣٤].

$$\text{دقت} = \frac{\text{تعداد پلاک های بازشناسی شده درست}}{\text{تعداد کل پلاک ها}} * 100 \qquad (٤)$$

جدول (۲): نمونه تصاویر از پایگاه داده‌های مورد استفاده

IR-LPR			
UFPR-ALPR			
AOLP			

٤-٣-نتایج بازشناسی پلاک خودرو

برای آموزش مدل، تصاویر موجود در پایگاه‌داده باید به مجموعه‌های آموزشی، ارزیابی و آزمون تقسیم شوند. جدول (۳) ساختار مجموعه‌های آموزش، ارزیابی و آزمون را در سه پایگاه داده نشان می‌دهد. همانطور که در جدول (۱) ذکر شد، برای دو پایگاه IR-LPR و UFPR-ALPR پروتکل ارزیابی وجود دارد. این بدان معنی است که در مراجع اصلی این دو پایگاه تصاویر آموزش، ارزیابی و آزمون تفکیک شده است و آزمایش‌ها با تنظیمات مقالات مرجع انجام شده است. برای پایگاه AOLP چنین تفکیکی وجود ندارد، با این حال مشابه کارهای دیگران در این پژوهش نیز مطابق با جدول (۳) تصاویر به شکل تصادفی تفکیک شده است.

پیکربندی مدل ABINet برای بازشناسی پلاک شامل تنظیم چندین ابر پارامتر است که تنظیمات آموزش مدل و عملکرد آن را تعیین می‌کند. در این مقاله، ابر پارامترهای کلیدی زیر بر روی مدل ABINet تنظیم شده است:

- اندازه تصویر: وضوح ورودی مدل ABINet را تعیین می‌کند.
- نرخ یادگیری: سرعتی را که مدل ABINet وزن خود را در طول آموزش تنظیم می‌کند، کنترل می‌کند. مقدار نرخ یادگیری بهینه به پایگاه‌داده و پیچیدگی کار بستگی دارد.
- اندازه دسته: تعداد تصاویری را که به طور هم‌زمان در طول هر تکرار آموزشی پردازش می‌شوند را تعیین می‌کند. اندازه دسته بهینه به پیچیدگی مدل بستگی دارد.
- دوره: تعداد دفعاتی را که مدل ABINet از کل پایگاه‌داده آموزشی عبور می‌کند، تعیین می‌کند. تعداد بهینه دوره‌ها ممکن است به‌اندازه پایگاه‌داده، پیچیدگی و سرعت همگرایی مدل بستگی داشته باشد.

جدول (۳): تعداد تصاویر آموزش، ارزیابی و آزمون در پایگاه‌های داده.

	پایگاه‌داده AOLP	پایگاه‌داده UFPR-ALPR	پایگاه‌داده IR-LPR
نمونه‌های آموزش	۱۱۲٦	۱۸۰۰	۳٤۰۹۸
نمونه‌های اعتبارسنجی	۲۲٥	۹۰۰	۷۳۰۸
نمونه‌های آزمون	٦۹۸	۱۸۰۰	۷۳۰٦
مجموع	۲۰٤۹	٤٥۰۰	٤۸۷۱۲

ابرپارامترهای مورداستفاده در جدول (٤) به دست آمده‌اند. این پارامترها به شکل تجربی برای هر پایگاه‌داده حاصل شده‌اند.

در طول آزمایش، رایانه با پردازنده Intel(R) Core(TM) i9-۱۰۹۰۰k @ ۳/۷۰GHz، ۳۲ گیگابایت رم، ۱۰۲٤ گیگابایت SSD و یک پردازنده گرافیکی اختصاصی NVIDIA Geforce RTX ۳۰۷۰ با حافظه ۸ گیگابایتی GDDR6 استفاده شده است.

جدول (٤): ابرپارامترهای تنظیم شده و مقدار آنها.

ابر پارامتر	پایگاه‌داده AOLP	پایگاه‌داده UFPR-ALPR	پایگاه‌داده IR-LPR
اندازه تصویر	۳۲×۱۲۸	۳۲×۱۲۸	۳۲×۱۲۸
اندازه دسته	۳۲	۳۲	٦٤
دوره	۲۰۰۰	۱۰۰۰	٥۰۰
نرخ یادگیری	۰/۰۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱

جدول (٥) نتایج مدل پیشنهادی را در مراحل آموزش، ارزیابی و آزمایش برای هر سه پایگاه داده IR-LPR، AOLP و UFPR-ALPR نشان می‌دهد. همانطور که از جدول پیداست، آموزش، ارزیابی و آزمون با دقت خوبی روی هر سه پایگاه داده انجام شده است.

جدول (٥) : نتایج دقت (%) روش پیشنهادی.

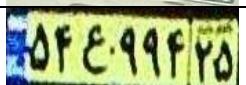
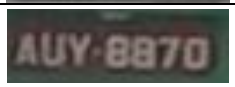
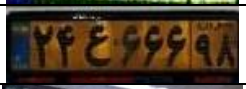
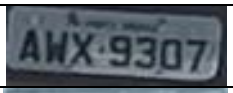
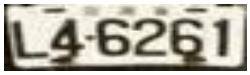
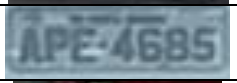
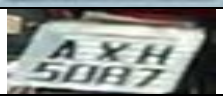
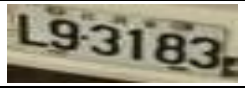
	پایگاه‌داده IR-LPR	پایگاه‌داده UFPR-ALPR	پایگاه‌داده AOLP
آموزش	۹۹/۹۷	۹۹/۹۹	۱۰۰
اعتبارسنجی	۹۷/۰۳	-	۹۹/۹۹
آزمون	۹۷/۱۲	۹۹/۹۳	۹۹/٤

جدول (٦) نیز نتایج کیفی را برای تعدادی از تصاویر این سه پایگاه داده به همراه نتایج بازشناسی مربوطه تولید شده توسط مدل پیشنهادی نشان می‌دهد. این نتایج نشان می‌دهند مدل پیشنهادی قادر به بازشناسی بسیاری از پلاک‌ها با دقت بسیار بالایی است. با این وجود برای برخی از پلاک‌ها نیز حتی استفاده از مدل زبانی نتوانسته است به بهبود نتایج کمک کند. به عنوان مثال در این شکل سطر آخر پلاک مربوط به پایگاه‌داده AOLP را نشان می‌دهد که در آن عدد ۸ با حرف B اشتباه شده است. متاسفانه چون در این موقعیت در این پایگاه داده هم حرف و هم عدد ممکن است ظاهر شوند و قاعده مشخصی در این خصوص ظاهرا وجود ندارد، مدل نیز قادر به بازشناسی صحیح نبوده است.

٤-٤-ارزیابی کارایی محاسباتی

مدل ABINet یک مدل مبتنی بر ترنسفورمر است. این روش در مقایسه با روش‌های سنتی، با افزایش نیازهای محاسباتی و منابع همراه است. به طور خلاصه، انتخاب بین روش‌های OCR سنتی و مبتنی بر ترنسفورمر به نیازهای کاربردی و به تعادل بین دقت، منابع و محیط‌های استقرار بستگی دارد.

جدول (٦) : نتایج کیفی خروجی تعدادی از تصاویر آزمون با استفاده از معماری پیشنهادی.

پایگاه‌داده IR-LPR		پایگاه‌داده UFPR-ALPR		پایگاه‌داده AOLP	
متن خروجی	تصویر ورودی	متن خروجی	تصویر ورودی	متن خروجی	تصویر ورودی
۹۵ت ۳۸٦۳۳		MLS5511		8D9186	
۹۳ث۷٦۸۱۱		AQY6388		EZ8402	
٥٤ع٩٩٤٢٥		AUY8870		0329HB	
٢٤ع٦٦٦٩٨		AVI2214		LE7857	
٣٤ی٨١٣٢٦		AWX9307		L46261	
٤٤س٢١٢٣١		APE4685		T26406	
٢٥د٨٥٦١٤		AXN5087		L931B3	

جدول (۷) تعداد پارامترها و الزامات محاسباتی روش ABINet و سرعت متوسط استنتاج برای هر تصویر را نشان می‌دهد. در اینجا سرعت بازشناسی مدل پیشنهادی با استفاده از سخت افزار فوق، به طور متوسط برای هر تصویر ۳۲/۱٦٥ میلی‌ثانیه بدست آمد. همان‌طور که از جدول پیداست، علی‌رغم بالابودن تعداد پارامترها، سرعت استنتاج برای کاربردهای بی‌درنگ مناسب به نظر می‌رسد.

٤-٥-ارزیابی نتایج در مقایسه با سایر کارها

جدول (۸) نتایج مدل پیشنهادی را در مقایسه با سایر روش‌ها نشان می‌دهد. طبق نتایج موجود در جدول (۸)، مدل پیشنهادی در هر سه پایگاه داده AOLP، UFPR-ALPR و IR-LPR دقت بازشناسی بالاتری نسبت به سایر روش‌ها نشان می‌دهد. در پایگاه‌داده AOLP، دقت بازشناسی مدل پیشنهادی به ۹۹/۹۹ درصد رسید که از بهترین نتیجه به‌دست‌آمده توسط روش‌های قبلی (۹۹/۷ درصد) به میزان ۰/۲ درصد بیشتر است. همچنین در پایگاه‌داده UFPR-ALPR نیز مدل پیشنهادی با دقت ۹۹/۹۳ درصد توانسته است از سایر روش‌ها پیشی بگیرد که تفاوت معناداری با سایر روش‌ها دارد.

در پایگاه‌داده IR-LPR مدل پیشنهادی با دقت ۹۷/۱۲ درصد به عملکرد خیلی بهتری نسبت به سایر مدل‌ها دست یافته است. باید توجه داشت که دو پایگاه داده نخست متاسفانه تعداد نمونه‌های محدودی دارند که قدرت مدل پیشنهادی را در مقایسه با سایر روش‌ها کمتر می‌توانند نشان دهند چرا که با وجود برتری مدل پیشنهادی، برای این تعداد محدود نمونه‌ها، سایر روش‌ها نیز با دقت قابل قبولی عمل می‌کنند. در حالی که برای پایگاه داده IR-LPR که تنوع و تعداد بسیار بیشتری را شامل می‌شود، تفاوت با سایرین چشم‌گیر است.

جدول (۷): پارامترها و الزامات محاسباتی روش پیشنهادی.

پارامترها ($*10^6$)	FLOPs ($*10^6$)	سرعت استنتاج (ms/ تصویر)
۳۷	۷/۲	۳۲/۱٦٥

مقایسه روی این پایگاه داده با دو روش [۳۱] و [۳٥] که هر دو مبتنی بر یک مدل بینایی هستند، نشان می‌دهد استفاده از مدل زبانی در کنار مدل بینایی توانسته است به بهبود دقت بازشناسی کمک کند. با این حال هر دو این روش‌ها از شبکه‌های عصبی پیچشی برای مدل بینایی خود استفاده می‌کنند. برای مقایسه بهتر، روش پیشنهادی روی این پایگاه داده با حالتی که تنها از مدل بینایی مبتنی بر شبکه ترنسفورمری استفاده شده است [۳٤] نیز مقایسه شده است که پیاده سازی این مدل نیز به دقت بسیار پایین‌تر یعنی ۹٤/۱۸ درصد دست یافت.

این برتری عملکرد می‌تواند به دلیل استفاده توأمان از مدل‌های بینایی و مبتنی بر زبان در مرحله پردازش داده‌ها و استخراج ویژگی‌ها به‌ویژه با استفاده از مدل‌های زبانی پیشرفته حاصل شود. این روش باعث می‌شود که سامانه قادر به شبیه‌سازی بهتر ساختارهای پیچیده اطلاعاتی در پلاک‌ها شود و علاوه بر شناسایی بهتر، از دقت بالاتری در شبیه‌سازی کلمات و اعداد در محیط‌های پیچیده بهره‌برداری کند.

در مجموع، مدل پیشنهادی توانسته است دقت بازشناسی را در همه پایگاه‌داده‌های مورداستفاده بهبود بخشد. این نتایج گویای عملکرد قوی‌تر مدل پیشنهادی در مقایسه با سایر روش‌های مشابه و رقبا است، به‌ویژه در شرایط غیر ایده‌آل و تصاویر باکیفیت پایین که معمولاً باعث افت دقت در سایر مدل‌ها می‌شود.

متأسفانه، مقایسه روش پیشنهادی از نظر سرعت و نیازهای محاسباتی با سایر روش‌ها به دلیل عدم وجود اطلاعات کافی در مقالات مرجع مربوطه امکان‌پذیر نیست. بااین‌حال در جدول (۸) یک ستون با عنوان "معماری" در نظر گرفته شده است. این ستون کمک می‌کند تا بتوان با مقایسه نوع معماری به‌کاررفته در هر یک از روش‌ها، درکی از پیچیدگی محاسباتی روش‌ها بدست آورد. همانطورکه از جدول پیداست، غالب روش‌های مورد استفاده برای پلاک خوانی از شبکه‌های عصبی پیچشی استفاده می کنند که معمولا تعداد پارامتر کمتری نسبت به شبکه‌های ترنسفرمری دارند. تنها یکی از مراجع [۳۷] وجود دارد که در کنار مدل بینایی از مدل زبانی هم با استفاده از یک مدل زبانی بزرگ چندوجهی[1] به نام QwenLM با ۷ میلیون پارامتر استفاده می کند. این مدل علی‌رغم داشتن تعداد پارامتر بیشتر نسبت به مدل پیشنهادی (۳/۷ میلیون پارامتر) از آنجا که برای پلاک خوانی از هیچ آموزش یا تنظیم[2] استفاده نمی کند، دقت پایینتری نسبت به مدل پیشنهادی دارد.

۵- جمع‌بندی

در این مقاله، استفاده از یک مدل زبان برای تصحیح متن، همراه با یک مدل بینایی برای افزایش دقت بازشناسی پلاک خودرو مورد استفاده قرار گرفت. مدل بینایی دارای یک معماری پیچشی-ترنسفورمری است که از ResNet45 برای استخراج ویژگی و از ترنسفورمر برای مدل‌سازی دنباله استفاده می‌کند. همچنین از مدل زبانی برای اصلاح و بهبود دقت متن به‌دست‌آمده استفاده می‌شود. نتایج تجربی بر روی سه پایگاه‌داده نشان‌دهنده اثربخشی برتر در تقابل با سایر روش‌ها با دستیابی به نرخ بازشناسی ۹۷/۱۲ درصد، ۹۹/۹۳ درصد و ۹۹/٤ در مجموعه‌داده‌های IR-LPR، UFPR-ALPR و AOLP است. استفاده از مدل زبان در کنار مدل بینایی پیشرفته به بازشناسی کاراکترهای الفبایی و عددی موجود در ساختار پلاک‌ها کمک می‌کند و کارآمدی رویکرد پیشنهادی را در انجام وظایف مختلف و چالش‌برانگیز بازشناسی پلاک خودرو نشان می‌دهد.

جدول (۸) : مقایسه دقت روش پیشنهادی با سایر روش‌ها.

پایگاه‌داده	مرجع	معماری	سال	دقت
IR-LPR	[۳۱]	مدل بینایی (CNN)	۲۰۲۲	۹۳/۳
	[۳۵]	مدل بینایی (CNN)	۲۰۲٤	۹٤/۹
	پیشنهادی	مدل بینایی (ترنسفورمری)	۲۰۲٤	۹٤/۱۸
	پیشنهادی	مدل بینایی + مدل زبانی (ترنسفورمری)	**۲۰۲٤**	**۹۷/۱۲**
UFPR-ALPR	[۳]	مدل بینایی (CNN)	۲۰۱۹	۹۷/۵۷
	[۳٦]	مدل بینایی (CNN)	۲۰۲۰	۹٦/۸۷
	[۳۷]	مدل بینایی + مدل زبانی (ترنسفورمری)	۲۰۲٤	۹۰
	پیشنهادی	مدل بینایی + مدل زبانی (ترنسفورمری)	**۲۰۲٤**	**۹۹/۹۳**
AOLP	[۳۸]	مدل بینایی (CNN)	۲۰۲۱	۹۹/۲
	[۳۹]	مدل بینایی (CNN)	۲۰۱۸	۹۸/۳٦
	پیشنهادی	مدل بینایی + مدل زبانی (ترنسفورمری)	**۲۰۲٤**	**۹۹/٤**

مراجع

[1] Khan, M.M., et al., License plate recognition methods employing neural networks. IEEE Access, 2023. 11: p. ۷۳٦۱۳-۷۳٦٤٦.

[2] Tao, L., et al., A Real-Time License Plate Detection and Recognition Model in Unconstrained Scenarios. Sensors, 2024. 24(9): p. 2.۷۹۱

[3] Rakhshani, S., E. Rashedi, and H. Nezamabadi-pour, License plate recognition using deep learning. Journal of

[1] multi modal large language model

[2] fine-tunning

Machine Vision and Image Processing, 2019. 6(1): p. ۳۱-٤٦.

[4] Laroca, R., et al., An efficient and layout - independent automatic license plate recognition system based on the YOLO detector. IET Intelligent Transport Systems, ۲۰۲۱. ۱٥(٤): p. ٤٨۳-٥۰۳.

[٥] ساجد, ر.ا., ر. عصمت, and ن.آ.پ. حسین, بازشناسی پلاک خودرو با استفاده از یادگیری ژرف. ۲۰۱۹.

[6] Abolghasemi, V. and A. Ahmadyfard, An edge-based color-aided method for license plate detection. Image and Vision Computing, 2009. 27(8): p. 1134-1142.

[7] Lalimi, M.A., S. Ghofrani, and D. McLernon, A vehicle license plate detection method using region and edge based methods. Computers & Electrical Engineering, ۲۰۱۳. ۳۹(۳): p. ۸۳٤-۸٤٥.

[8] Wu, Y., S. Liu, and X. Wang. License plate location method based on texture and color. in 2013 IEEE 4th International Conference on Software Engineering and Service Science. 2013. IEEE.

[9] Asadi-Zeydabadi, F., et al .Farsi Optical Character Recognition Using a Transformer-based Model. in 2023 13th International Conference on Computer and Knowledge Engineering (ICCKE). 2023. IEEE.

[10] Afkari-Fahandari, A., et al., Farsi Handwritten Text Recognition via a Lightweight Attention-driven Sequence Recognition Network.

[11] Del Castillo Velarde, M. and G. Velarde, Benchmarking Algorithms for Automatic License Plate Recognition. arXiv e-prints, 2022: p. arXiv: 2203.14298.

[12] Afkari-Fahandari, A., et al. Enhancing Farsi Text Recognition via Iteratively Using a Language Model. in 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP). 2024. IEEE.

[13] Fang, S., et al. Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[14] Jiang, Y., et al., An efficient and unified recognition method for multiple license plates in unconstrained scenarios. IEEE Transactions on Intelligent Transportation Systems, 2023. 24(5): p. 5376-5389.

[15] Wu, B.-F., S.-P. Lin, and C.-C. Chiu, Extracting characters from real vehicle licence plates out-of-doors. IET Computer Vision, 2007. 1(1): p. 2-10.

[16] Du, S., et al., Automatic license plate recognition (ALPR): A state-of-the-art review. IEEE Transactions on circuits and systems for video technology, 2012. ۲۳(۲): p. ۳۱۱-۳۲٥.

[17] Paliy, I., et al. Approach to recognition of license plate numbers using neural networks. in 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). 2004. IEEE.

[18] Goel, S. and S. Dabas. Vehicle registration plate recognition system using template matching. in 2013 International Conference on Signal Processing and Communication (ICSC). 2013. IEEE.

[19] Pan, M.-S., J.-B. Yan, and Z.-H. Xiao, Vehicle license plate character segmentation. International Journal of Automation and Computing, 2008. 5(4): p. 425-432.

[20] Chen, R.-C., Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. Image and Vision Computing, 2019. 87: p. 47-56.

[۲۱] محمدی, شناسایی شماره پلاک خودرو بر اساس یادگیری عمیق با نظارت ضعیف. مجله بینایی ماشین و پردازش تصویر, ۲۰۲۱. ۷(۲): p. 25-34.

[۲۲] گیوکی, et al., معرفی دادگان ایرپلاک برای بازشناسی پلاک خودروهای ایرانی. مجله بینایی ماشین و پردازش تصویر, ۲۰۲۲. ۹(۲): p. 81-95.

[23] Altyar, S.S., S.S. Hussein, and L.A. Tawfeeq, Accurate license plate recognition system for different styles of Iraqi license plates. Bulletin of Electrical Engineering and Informatics, 2023. 12(2): p. 1092-1102.

[24] Wang, J., et al., Sequence recognition of Chinese license plates. Neurocomputing, 2018. 317: p. 149-158.

[25] Wang, H., et al., Robust Korean license plate recognition based on deep neural networks .Sensors, ۲۰۲۱. ۲۱(۱۲): p. ٤۱٤۰.

[26] Zhang, L., et al., A robust attentional framework for license plate recognition in the wild. IEEE Transactions on Intelligent Transportation Systems, 2020. 22(11): p. ٦۹٦۷-٦۹۷٦.

[27] Fang, S., et al., Abinet++: Autonomous, bidirectional and iterative language modeling for scene text spotting. IEEE transactions on pattern analysis and machine intelligence, 2022. 45(6): p. 7123-7141.

[28] Vaswani, A., Attention is all you need. Advances in Neural Information Processing Systems, 2017.

[29] Wang, T., et al. Decoupled attention network for text recognition. in Proceedings of the AAAI conference on artificial intelligence. 2020.

[30] Ronneberger, O., P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October ٥-۹, ۲۰۱٥, proceedings, part III ۱۸. ۲۰۱٥. Springer.

[31] Rahmani, M., et al. IR-LPR: A Large Scale Iranian License Plate Recognition Dataset. in 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE). 2022. IEEE.

[32] Hsu, G.-S., J.-C. Chen, and Y.-Z. Chung, Application-oriented license plate recognition. IEEE

transactions on vehicular technology, 2012. 62(2): p. ۵۵۲-۵۶۱.

[33] Laroca, R., et al. A robust real-time automatic license plate recognition based on the YOLO detector. in 2018 international joint conference on neural networks (ijcnn). 2018. IEEE.

[34] Asadi-zeydabadi, F., et al., IDPL-PFOD2: A New Large-Scale Dataset for Printed Farsi Optical Character Recognition. arXiv preprint arXiv:2312.01177, 2023.

[35] Wang, H., L. Wu, and H. Liang, A convolutional Neural Network Approach for Character Recognition in Non-English License Plates. Journal of Optics, 2024. 53(2): p. ۱۲۵۶-۱۲۶۳.

[36] Fernandes, L.S., et al. A robust automatic license plate recognition system for embedded devices. in Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–۲۳, ۲۰۲۰, Proceedings, Part I 9. 2020. Springer.

[37] Ding, H., et al., A Training-Free Framework for Video License Plate Tracking and Recognition with Only One-Shot. arXiv preprint arXiv:2408.05729, 2024.

[38] Silva, S.M. and C.R. Jung. License plate detection and recognition in unconstrained scenarios. in Proceedings of the European conference on computer vision (ECCV). ۲۰۱۸.

[39] Pham, T.-A., Effective deep neural networks for license plate detection and recognition. The Visual Computer, ۲۰۲۳. ۳۹(۳): p. ۹۲۷-۹۴۱.

الهام شعبانی‌نیا دارای مدرک کارشناسی ارشد مهندسی کامپیوتر از دانشگاه صنعتی شریف (سال ۱۳۸۸) و دکتری هوش مصنوعی از دانشگاه اصفهان (سال ۱۳۹۷) است. پس از گذراندن دوره پسادکتری در دانشگاه شهید باهنر کرمان، از سال ۱۴۰۲ به عنوان استادیار در گروه ریاضی کاربردی دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته مشغول به فعالیت است. زمینه تحقیقاتی او شامل یادگیری ماشین و هوش مصنوعی است و تحقیقات متعددی در حوزه‌های مختلفی مانند ردگیری، بازشناسی کنش، OCR و پردازش هوشمند اسناد انجام داده است. نتیجه این تلاش‌ها انتشار مقالات علمی متعدد در این زمینه بوده است.

فاطمه اسدی زیدآبادی مدرک کارشناسی، کارشناسی ارشد خود را به ترتیب در سالهای ۱۳۹۹ و ۱۴۰۲در رشته مهندسی برق-الکترونیک از دانشگاه صنعتی سیرجان و دانشگاه شهید باهنر کرمان دریافت کرد. وی هم اکنون به عنوان پژوهشگر ارشد در آزمایشگاه پردازش هوشمند داده در دانشگاه شهید باهنر کرمان مشغول به فعالیت هستند. ایشان نویسنده و هم-نویسنده چندین مقاله در مجلات و کنفرانسهای علمی است. زمینه های علاقه مندی ایشان، پردازش متن، داده کاوی، پردازش تصویر، یادگیری ماشینی و بازشناسی نوری حروف است.

حسین نظام‌آبادی‌پور در سال ۱۳۷۷ مدرک کارشناسی خود را در رشته مهندسی برق-الکترونیک از دانشگاه شهید باهنر کرمان دریافت کرد. سپس، مدرک کارشناسی ارشد و دکتری را در رشته مهندسی برق-الکترونیک از دانشگاه تربیت مدرس به ترتیب در سال های ۱۳۷۹ و ۱۳۸۳ دریافت کرد. ایشان در سال ۱۳۸۳ به عنوان استادیار به بخش مهندسی برق دانشگاه شهید باهنر کرمان پیوست و در سال ۱۳۹۱ به درجه استادی ارتقا یافت. وی نویسنده و هم-نویسنده بیش از ۴۵۰ مقاله در مجلات و کنفرانس‌های معتبر علمی بوده است. زمینه‌های علاقه‌مندی ایشان، شامل پردازش تصویر، بازشناسی الگو، رایانش نرم و الگوریتم های فرا ابتکاری است.