

مکان‌یابی و تشخیص (Detection) کاراکترهای کارت ملی در تصاویر با پیچیده‌گی‌های خاص مبتنی بر الگوریتم‌های YOLOv3&5

سامرند مجنونی^۱، طلحه قدوسیان^۲، علی خواجه میرزائی^۳

^۱ دانشجوی کارشناسی ارشد هوش مصنوعی، مجتمع دانشگاهی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران
Samrand.majnooni@gmail.com

^۲ دانشجوی کارشناسی ارشد هوش مصنوعی، مجتمع دانشگاهی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران
T.ghodusiean@gmail.com

^۳ دانشجوی کارشناسی ارشد هوش مصنوعی، مجتمع دانشگاهی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران
alimirzaei420@yahoo.com

چکیده

کارت ملی از جمله اسناد هویتی اصلی و ضروری شهروندان ایرانی است به گونه‌ای که سند مرجع احراز هویت در عملیات و تراکنش‌های بانکی، پروکرایی‌ها و خدمات سازمانی – اداری، سفر و ... به شمار می‌رود همچنین اکثر قریب به اتفاق عملیات و خدمات مزبور نیاز به وارد کردن و استخراج اطلاعات مندرج در کارت ملی دارند امروزه با رواج روز افزون سامانه‌های مکانیزه ارائه خدمات، تشخیص هویت و استخراج هوشمند اطلاعات مندرج در اسناد هویتی از جمله کارت ملی ضرورتی بنیادین است که در جستار پیش‌رو یک رویکرد هوشمند مبتنی بر شبکه‌های عصبی کانولوشنال (CNNs) به منظور تشخیص کاراکترهای کارت ملی در تصاویر با پیچیدگی‌های خاص طبیعی ارائه شده است.

در این راستا به دلیل محرمانگی کارت‌های ملی ابتدا یک الگوریتم^۱ جهت تولید تصاویر کارت ملی و کاراکترهای آن ایجاد شده، و برای افزودن پیچیدگی به تصاویر از ترکیب^۱ سکریپت نرم افزار گرافیکی فتو شاپ و الگوریتم‌های رایج تولید داده پایتون^۱ استفاده شد. سپس از تبدیل هاف^۲ برای اصلاح چرخش تصاویر و الگوریتم یولو^۳ نسخه ۳ و ۵ با هدف مکان‌یابی و تشخیص کاراکترهای کارت ملی بهره گرفته شد نتایج حکایت از عملکرد مطلوب الگوریتم‌های پی‌شده‌ی یولو ۵ با میانگین دقت ۹۸,۵٪ و یولو ۳ با ۹۷٪ در تشخیص کاراکترها کارت ملی در تصاویر پیچیده دارد.

کلمات کلیدی

کارت ملی، شبکه‌های عصبی کانولوشن (CNN)، یولو (YOLO)، تبدیل هاف، تشخیص، پیچیدگی.

^۱ SHakhs Algorithm

^۲ Hough Transform

^۳ YOLO: You Only Look Once

۱- مقدمه

دادگان بومی (کارت ملی) با شبیه سازی پیچیدگی های خاص نظیر تصاویر با پس زمینه گوناگون و کیفیت پایین، تار و کدر بودن، عمق و پرسپکتیو، شرایط نوری گوناگون و ... است. گزارش این بخش از پروژه فقط بر تولید داده، افزودن پیچیدگی و نیز مکان یابی و تشخیص کاراکترها متمرکز است که با تکمیل بخش های دیگر نظیر شناسایی کاراکترها و نیز احراز هویت مبتنی بر تصویر کارت ملی، امید می رود مدل پیچیده ای به یک محصول کاربردی در انجام امورات گوناگون نیازمند به استخراج اطلاعات هویتی تبدیل شود.

۲- چشم انداز پژوهش های مرتبط گذشته

به طوری کلی اغلب پژوهش ها تشخیص متن بر روی اسناد چاپی، دستی، دیجیتالی، تابلوهای راهنما و بیلبردها متمرکز بوده است گزارش های نسبتاً معدودی از تشخیص و استخراج نوشته های اسناد هویتی مثل کارت شناسایی ارائه شده است در ادامه به اختصار نتایج شاخص ترین مدل ها و پروژه های صورت گرفته در زمینه تشخیص و شناسایی کاراکترهای کارت شناسایی ذکر شده اند:

مقاله [۵] ترکیب های الگوریتمی مختلفی را برای تشخیص و شناسایی کاراکترهای کارت شناسایی اندونزیایی در تصاویر پیچیده را مورد ارزیابی قرار داده است از الگوریتم Faster R-CNN و YOLOv5 برای تشخیص متن و ارقام، SVM برای طبقه بندی ارقام و Improved STAR-Net شناسایی نام بهره گرفته است. که بهترین عملکرد مربوط به مدل «Faster R-CNN + YOLOv5 (digit character detection) + SVM + Improved STAR-Net» با میانگین دقت ۷۴٫۲٪ و زمان ۳٫۲S در تمام کاراکترها است که این بین دقت الگوریتم YOLOv5 در مکان یابی و تشخیص ارقام ۹۳٫۶٪ گزارش شده، همچنین کمترین دقت در شناسایی کاراکترها مربوط به نام کامل با ۶۹٫۸٪ است ضمناً لازم به ذکر است پایین بودن دقت مدل پیشنهادی در این پژوهش نسبت دیگر پژوهش های به دلیل پیچیده بودن داده ها و افزودن انواع نویز به آن ها، و نیز حجم نسبتاً کم داده ها در مقایسه با حجم داده مورد نیاز برای یک شبکه عصبی عمیق بوده است.

پروژه [۸] برای شناسایی کاراکترهای کارت شناسایی چینی ترکیبی از الگوریتم های آماری پردازش تصویر و شبکه های عصبی کانولوشن را پیشنهاد کرده است از تبدیل هاف^۲ برای تصحیح چرخش تصویر، از کلاسی فایر AdaBoost برای تشخیص نماد و چهره (به منظور پیدا کردن موقعیت کارت در تصویر کلی)، همچنین تشخیص متن بر اساس پردازش مورفولوژی تصویر صورت گرفته است و در نهایت از سه شبکه عصبی کانولوشنی برای شناسایی کاراکتر بهره برده، که بیشترین دقت ۹۹٫۷٪ مربوط به ResNet-56 است همچنین دقت شبکه پیشنهادی پژوهش مزبور با ۸۰٫۴٪ گزارش شده است. دقت بالا شبکه های مورد استفاده در این پژوهش مربوط به عدم پیچیدگی مجموعه داده ها است.

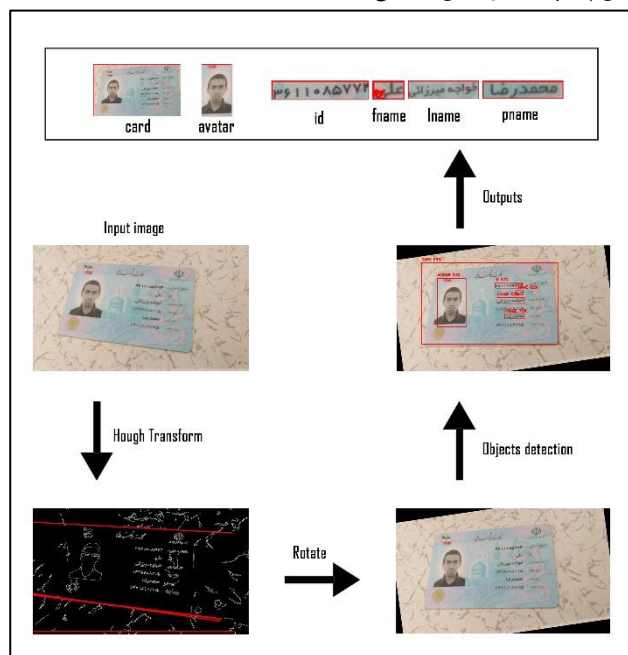
با استفاده از الگوریتم پایه YOLOv5 در مقاله [۱۲] نواحی کاراکترهای پاسپورت روسی تشخیص و سپس شناسایی شدند در این راستا ابتدا تصاویر پیش پردازش و تبدیل به مقیاس خاکستری شدند که دقت مدل پیشنهادی در تشخیص محل و نواحی کاراکترها (متن) ۷۸٫۶٪ و شناسایی آن ها ۹۷٫۴٪ ذکر شده است. برای شناسایی کاراکترهای کارت شناسایی ویتنامی [۱۶] مدلی

امروزه استفاده از فناوری مرتبط با عامل های هوشمند به موجب قابلیت های ویژه شان در امکان شناسایی تفکر انسانی برای یادگیری ماشین، گسترش و توسعه چشمگیری داشته اند به گونه ای که فناوری های با عامل هوش مصنوعی جایگزین بسیاری از عملیات دستی و تکراری روزمره شده اند، که این امر علاوه بر تسهیل عملیات های فیزیکی، صرفه جویی در زمان و نیز کاهش خطا و نیروی انسانی را به همراه داشته است از جمله پرتکرارترین عملیات روزمره استخراج اطلاعات مدارک هویتی نظیر کارت ملی به منظور وارد کردن و استفاده از آن ها در امورات بانکی، فرآیندهای ثبت نامی، احراز هویت در مراکز و سازمان های مختلف و ... است که برای انجام چنین عملیات و اموراتی به صورت دستی، اغلب با چالش هایی نظیر خطا در وارد کردن ارقام و اطلاعات، هدر رفت زمان و ... مواجه ایم. شناسایی هوشمند کاراکترهای مدارک هویتی از جمله کارت ملی بصورت گسترده در عملیات ثبت نامی و احراز هویت توسعه یافته اند که رایج ترین روش شناسایی هوشمند کاراکتر مدارک هویتی، استفاده از فناوری OCR^۱ ساده است نرخ تشخیص OCR ساده در برخورد با کارت های دارای پس زمینه پیچیده، اعداد نامنظم، کم رنگ شدن ارقام، تصاویر کم کیفیت و ... کاهش پیدا می کند [۱ و ۲] نتایج پژوهش های گوناگون منتشر شده حکایت از عملکرد بسیار مطلوب انواع الگوریتم شبکه های عصبی و یادگیری عمیق نظیر CNN، RNN، ادغام کننده End to End نظیر YOLO و ... در استخراج ویژگی ها تمایزی و شناسایی متون دارد همچنین این الگوریتم ها جایگاه ویژه ای در تشخیص و بازشناسی نوشته ها در تصاویر پیدا کرده اند از دیگر قابلیت الگوریتم های عمیق می توان به تعمیم پذیری مطلوب، مقاومت به نویز بالا، مدلسازی سری های زمانی و ... اشاره نمود [۳ و ۴] در میان شبکه های عصبی تکنیک های بر پایه CNNs از جمله: LeNet، AlexNet، ResNet، YOLO و ... به صورت مجزا و با ترکیبی با دیگر الگوریتم های به دلیل عملکرد و دقت بسیار مطلوب در تشخیص و استخراج کاراکترهای مدارک هویتی بهبود یافته و مورد استفاده قرار گرفته اند که می توان به استخراج کاراکترهای کارت شناسایی اندونزیایی [۵]، چینی [۶ و ۷]، ویتنامی [۸ و ۹]، کره ای [۹]، نیز نویسه های فارسی، اردو و عربی [۱۰] و روسی [۱۱] بر پایه تکنیک های CNNs اشاره نمود. بنابراین با نظر به اهمیت مسئله مزبور، پروژه حاضر بر آن است با استفاده از الگوریتم یولو^۳، یک روش بهینه با دقت مطلوب در مکان یابی و تشخیص کاراکترهای کارت ملی پیشنهاد کند الگوریتم YOLO برخلاف اکثر شبکه عصبی پایه کانولوشنی رایج، یک شبکه End to End است و در تشخیص اشیا در زمان واقعی (بلادرنگ) عملکرد مطلوب و سرعت بالایی دارد همچنین قابلیت تشخیص اشیا در سه مقیاس را دارا است برجسته ترین ضعف YOLO دقت پایین در تشخیص اشیا کوچک نظیر متون است که این چالش با انتخاب داده های مناسب و نیز پیش پردازش آن ها تا حدودی قابل بهبود است و نیز چنین چالش و ضعف هایی در نسخه های ارتقاء یافته تا حد زیادی بر طرف شده و حتی قابلیت ویژه ای در تشخیص متن های چند جهته و نسبتاً پیچیده پیدا کرده اند [۱۳ - ۱۵]. از اهداف اصلی و بنیادین پروژه حاضر کار بر روی بهینه سازی انواع الگوریتم شبکه های عصبی کانولوشن بر روی

^۲ Hough Transform

^۱ Optical Character Recognition

انواع پیچیدگی‌ها اضافه شد. به طور کلی سیر اجرای پروژه پس از تولید داده‌ها بدین صورت است؛ ۱. پیش‌پردازش داده‌ها با اصلاح زوایه چرخش به وسیله تبدیل هاف ۲. انتخاب الگوریتم‌های YOLOv3&5 آموزش داده شده^۴ بر پایه مجموعه دادگان COCO^۵ و تغییر هاپر پارامترهای آن‌ها به صورت تجربی بر مبنای داده مورد نظر ۳. ارزیابی نهایی مدل براساس تصاویر با حالت‌های مختلف و نسبتاً پیچیده در دنیای واقعی، که شماتیکی از سیر اجرایی مدل پیشنهادی در شکل ۱ نشان داده شده است.



شکل (۱): شماتیکی از سیر اجرایی^۶ مدل پیشنهادی با نمونه تصویری از دنیای واقعی

۱-۳- مجموعه دادگان

همان‌گونه که ذکر شد با نظر به محرمانگی پایگاه داده ثبت احوال، دادگان این پروژه به صورت ترکیبی با استفاده از نرم افزار گرافیکی (اسکرپت فتوشاپ) و کتابخانه‌های دارای الگوریتم تولید داده تصویری تولید شدند بدین منظور یک الگوریتم با نام شخص^۷ جهت تولید داده برای پروژه حاضر ایجاد، و سپس با تکنیک‌های مختلف؛ پیچیدگی و داده‌افزایی^۸ بر روی دادگان مورد نظر اعمال شد.

۱-۳-۱- تولید داده

فرایند تولید داده با استفاده از الگوریتم ایجاد شده، شامل دو مرحله دستی و خودکار است ضمناً جزئیات تفصیلی فرایند تولید دیتا و الگوریتم شخص در پیوست ۲ (گزارش تولید دیتا) ذکر شده است.

مرحل اول (دستی): ابتدا چند نوع عکس از کارت‌های ملی مختلف با زوایای و پس زمینه‌های متنوع تهیه شد (پیوست ۱، شکل ۱). در مرحله بعد با استفاده از ابزار های نرم افزار فتوشاپ قسمت‌های مربوط به شماره ملی،

مبتنی بر روش end to end پیشنهاد شده است. که مراحل آن شامل برش کارت شناسایی، تشخیص متن با استفاده از ResNet-50 و RetinaNet به عنوان شبکه پایه (ستون فقرات)، و نیز شناسایی کاراکترهای Inception-v3 و BiLSTM صورت گرفت است که دقت مدل پیشنهادی در کاراکترهای مختلف بین ۶۸ تا ۸۶٪ گزارش شده است. همچنین در پژوهشی دیگر [۱] کاراکترهای کارت شناسایی ویتنامی بر مبنای مجموعه دادگان اسکن شده با الگوریتم CNN-LSTM شناسایی شدند که بیشترین دقت مدل پیشنهادی در شناسایی کاراکترهای گروه قومیتی با ۹۹٪ و کمترین مربوط به محل خواستگاه با ۹۵٪ است.

مدل پیشنهادی [۷] یک شبکه بر پایه یک یادگیری انتقالی به منظور شناسایی نویسه‌های مندرج در کارت‌های شناسایی چینی است که در آن از شبکه GoogLeNet استفاده کرده است و شامل چهار مرحله: ۱. تولید داده و افزودن پیچیدگی ۲. پیش پردازش داده‌ها ۳. آموزش مدل با ۱۳۰۷۰ نمونه ۴. انتقال یادگیری بر روی یک شبکه که با ۱۹ میلیون نویسه چینی آموزش دیده است که دقت نهایی مدل پیشنهادی در شناسایی کاراکترها ۹۹٪ گزارش شده است.

الگوریتم YOLOv3 در مقاله [۱۳] برای تشخیص متن‌های چند جهته در معرض نمایش (متن صفحه) بهبود یافته است که برای این منظور از تابع اتلاف کادر چرخشی (جعبه چرخشی) جدید MD-Closs بر مبنای CIOU و نیز NMS گام به گام بهره گرفته شد که میزان دقت مدل پیشنهادی بر روی داده‌های مورد آزمایش ۸۶٫۲٪ با نرخ فراخوانی ۸۱٫۹٪ و زمان ۲۱٫۳ فریم در ثانیه ذکر شده است. در مقاله [۱۵] نیز الگوریتم YOLOv4 جهت آشکارسازی متن‌های کج در زمان واقعی بر مبنای تابع اتلاف RDIoU-NMS بهبود پیدا کرده است که F ۸۲٫۳٪ در ۶۲٫۵ فریم در ثانیه بر روی یکی از مجموعه داده‌های مورد استفاده گزارش شده است.

مقاله [۱۷] شرحی از باز شناسی شماره کارت بانکی بر پایه یادگیری عمیق است که در این پژوهش باز شناسی شماره کارت در سه مرحله؛ مکان‌یابی، قطعه‌بندی و شناسایی کاراکترها صورت گرفته سپس قابلیت الگوریتم‌های رایج در هر مرحله مثل CNN, YOLO, R-CNN, SVM و ... به اختصار ذکر شده‌اند که برای شناسایی شماره کارت از تکنیک LeNet استفاده شد همچنین از بین توابع فعال ساز بکار رفته در آموزش شبکه، Relu پایدارترین و بهترین عملکرد را داشته است، ضمناً YOLOv3 در تشخیص تصاویر کارت‌های با عدم اختلاف پیکسل بالا نسبت به دیگر الگوریتم‌ها موفق‌تر عمل کرده است اما دقت مکان‌یابی پایینی بویژه در تشخیص اشیاء کوچک دارد.

۳- رویکرد پیشنهادی

به دلیل محرمانگی پایگاه دادگان کارت ملی (سازمان ثبت احوال) و عدم دسترسی به آن‌ها، رویکرد پژوهش حاضر ابتدا تولید داده بر مبنای الگوهای کاملاً مشابه با کارت ملی اصلی است. سپس آموزش مدل با نمونه تصاویر تولید شده معمولی، و در ادامه برای بهبود عملکرد شبکه به مجموعه دادگان

^۵ Common Objects in Context

^۶ Pipeline

^۷ SHAKHS

^۸ Data Augmentations

^۳ Scene Text

^۴ <https://github.com/AlexeyAB/darknet#how-to-train-to-detect-your-custom-objects>

نام، نام خانوادگی، نام پدر و تصویر شخص را پاک کرده و زیر ساخت آن باز سازی شد. سپس برای قسمت‌های مذکور لایه‌های جدیدی را با فونت و سایز مناسب با عناوین code, name, family, pedar, avatar تعریف نموده که با فرمت فایل فتوشاپی (psd) ذخیره شدند.

مرحله اول (خودکار): در مرحله اول نیاز به لیستی از متن و تصاویر برای جایگذاری در لایه‌های تولید شده بود که برای این منظور الگوریتمی تحت عنوان شخص ایجاد شد. الگوریتم مزبور شامل تکنیک‌های برای تولید شماره ملی، نام مؤنث، نام مذکر و نام خانوادگی است. همچنین برای لایه آواتار از وبگاه generated Photos^۹، تصاویری از چهره مردانه و زنانه دریافت شد، که به ازای هر فایل psd، ۱۰۰ کارت ملی با یک تصویر مرد و ۱۰۰ کارت ملی با یک تصویر زن در نظر گرفته شد. بدین صورت فرآیند داده تولید صورت گرفت. در مجموع ۱۸۰۰ تصویر کارت ملی مرد و زن تولید شد که به دلیل محدودیت زمانی (بویژه هزینه زمانی برای برچسب زنی کلاس‌ها)، فقط ۹۰۰ تصویر (نسبت زن به مرد ۱:۱) انتخاب شد که با روش‌های داده‌افزایی به منظور افزودن پیچیدگی، مجموع تصاویر با ۱۵۵۰ رسید (جدول ۱).

جدول (۱): جزئیات دادگان پروژه

نوع تصویر	تعداد
پیچیدگی معمولی با ۹ پس زمینه - های متنوع	۹۰۰
پیچیدگی خاص با چرخش (Rotate) در زوایای مختلف با مقادیر تصادفی	۱۵۰
پیچیدگی خاص با کجی عمق‌نمایی (Skew)	۱۰۰
پیچیدگی خاص با فیلتر (Blur)	۱۰۰
پیچیدگی خاص با فیلتر (Brightness)	۱۰۰
پیچیدگی خاص با فیلتر (LensFlare)	۱۰۰
پیچیدگی خاص با فیلتر (Noise)	۱۰۰

۲-۱-۳- افزودن پیچیدگی و داده‌افزایی

همان‌طور که ذکر شده برای بهبود عملکرد مدل پیشنهادی از ترکیب اسکرپت نرم افزار گرافیکی فتو شاپ و الگوریتم‌های رایج تولید داده پایتون، داده‌افزایی و افزودن پیچیدگی به دادگان صورت گرفت. در طی سالیان

اخیر شبکه‌های عصبی عمیق، بویژه با توسعه الگوریتم‌های پایه کانولوشن عملکرد و پیشرفت‌ها قابل توجه و شگرفی در پردازش تصویر، بینایی کامپیوتر و ماشین‌دانش‌اند که می‌توان به پیشرفت و توسعه در زمینه‌هایی همچون طبقه‌بندی و بخش‌بندی تصاویر، تشخیص اشیاء، شناسایی نویسه‌ها و ... اشاره کرد. علیرغم قابلیت‌ها و ظرفیت‌های فراوان شبکه‌های عصبی عمیق؛ بسیار وابسته به مجموعه داده‌های بزرگ هستند به گونه‌ای که محدودیت داده‌ای در این شبکه‌ها موجب بیش‌برازش و عدم تعمیم‌پذیری آن در مواجه با مسائل گوناگونی اخصاً مسائل با داده محدود خواهد شد از جمله رویکردها و راهکارهای مواجه با این چالش، تقویت و داده‌افزایی با روش‌های سنتز داده هم‌مانند تکنیک‌های مبتنی بر شبکه‌های مولد تخصصی^{۱۰} است. در مجموع تکنیک‌های گوناگونی برای تقویت و ساخت مجموعه دادگان بزرگ پیشنهاد شده است که به‌طور کلی می‌توان آن‌ها را به عنوان روش‌های اعوجاج داده^{۱۱} یا بیش‌نمونه‌برداری^{۱۲} تقسیم‌بندی کرد همچنین استفاده از الگوریتم‌های جستجو^{۱۳} که ترکیبی از دو روش مزبور بوده، و با توجه به معماری لایه‌ای شبکه‌های عصبی عمیق که امکان و فرصت‌های زیادی را برای افزایش داده‌ها فراهم می‌کنند اگر چه تکنیک‌های داده‌افزایی را می‌توان در حوزه‌های مختلف گسترش داد اما در زمینه برخی از مجموعه داده‌های کوچک مثل نژاد بسیار خاصی از یک سگ نمی‌تواند بر کمبود و محدودیت داده غلبه کند به گونه‌ای که با هیچ یک از روش‌های داده‌افزایی و تقویت داده از جفت‌سازی نمونه^{۱۴} گرفته تا داده‌افزایی خودکار^{۱۵} و شبکه‌های مولد تخصصی، یک نژاد بسیار خاص از سگ را ایجاد کنند [۱۸]. با نظر به ماهیت پروژه حاضر یعنی شناسایی کاراکتر مدارک هویتی؛ بنابراین هدف اصلی و بنیادین از داده‌افزایی غلبه بر محدودیت مجموعه دادگان نیست بلکه با توجه به ابعاد کارکردی پروژه پیش‌رو، که بستر کاربردی آن در دنیای واقعی عام و گوناگون آن است. فلذا الگوریتم و شبکه پیشنهادی بایستی قابلیت تعمیم‌پذیری با انواع داده‌های ورودی یعنی تصاویر با کیفیت و حالات متفاوت را داشته باشد که در این راستا با تکنیک‌های ترکیبی بر پایه ۲۳ نوع فیلتر و عملگر تغییر تصویر نرم افزار گرافیکی فتو شاپ با انتخاب نرخ تصادفی مقادیر پارامترهای شان، پیچیدگی بر دادگان مورد نظر اعمال شد که قابلیت تغییر مقادیر و نرخ حداکثر و حداقل پارامترهای فقط یک فیلتر و یا عملگر تغییر عکس در یک تصویر می‌تواند بیش از ۱۰۰ داده تصویری دیگر با کیفیت بصری و مورفولوژی مختلف تولید کند به دلیل محدودیت سخت افزاری و نیز هزینه زمانی فقط ۴ نوع پیچیدگی شامل: Blur، LensFlare، Adjust Brightness / Contrast و Noise با فیلترهای گرافیکی و دو نوع پیچیدگی Rotate و Skew با کتابخانه‌های داده‌افزایی^{۱۶} پایتون بر روی دادگان ایجاد شد (شکل ۲). به صورت خیلی مختصر مراحل اصلی افزودن پیچیدگی به تصاویر با استفاده از عملگر و فیلترهای فتو شاپ شامل: ۱. باز کردن نرم افزار فتو شاپ با تکنیک Dispatch ۲. گرفتن آدرس کامل پوشه‌های تصاویر با (walk)، و تعریف حلقه برای باز کردن هر بار تصویری متفاوت از پوشه انتخابی ۳. فعال کردن تمام لایه‌های تصویر باز شده در فتوشاپ با دستور

^{۱۳} Search Algorithms

^{۱۴} Sample Pairing

^{۱۵} Auto Augment

^{۱۶} Data Augmentations

^۹ <https://generated.photos.com>

^{۱۰} Generative Adversarial Networks (GAN)

^{۱۱} warping

^{۱۲} oversampling

برچسب‌گذاری دادگان در ۶ کلاس: کارت ملی (Card)، عکس (Avatar)، شماره ملی (Id)، نام (Fname)، نام خانوادگی (Lname) و نام پدر (Pname) صورت گرفت (شکل ۳).



شکل (۳): نمونه تصویر کلاس‌های برچسب‌گذاری شده با labelImg-master

۳-۲-۲- اصلاح چرخش تصاویر

اصلاح چرخش تصاویر در این پژوهش بر اساس تبدیل هاف^{۱۷} صورت گرفت. تبدیل هاف یک روش بسیار موثر برای تصحیح افقی تصاویر کارت‌های شناسایی است. تبدیل هاف روشی است که پیکسل‌های تصویر را به یک فضای پارامتریک نگاشت می‌کند و در تشخیص اشکال منظم مانند خطوط مستقیم و دایره در یک تصویر بسیار موثر است. بزرگترین مزیت تبدیل هاف تحمل فاصله در توضیحات لبه ویژگی است که تحت تأثیر نویز تصویر قرار نمی‌گیرد (پیوست ۱).

استراتژی اصلی تبدیل هاف محاسبه مسیرهای ممکن نقاط مرجع در فضای پارامتریک با استفاده از مختصات فضای تصویر پیکسل هدف و محاسبه تعداد نقاط مرجع در انباشته‌کننده^{۱۸} است. مختصات قطبی فضای پارامتر (ρ, θ) است، که فرمول تبدیل مختصات مربوطه به صورت زیر است:

$$\rho = x * \cos(\theta) + y * \sin(\theta) \quad (۱)$$

در معادله بالا x و y مختصات دکارتی پیکسل مربوطه هستند. که پس از اعمال تبدیل هاف به تصویر مقیاس خاکستری، خط مستقیم شناسایی شده و زاویه شیب را می‌توان به دست آورد و با چرخش زاویه مربوطه به تصویر اصلی، تصویر اصلاح شده افقی را می‌توان به دست آورد.

تبدیل هاف، تمایل به گرفتن خطوط متعدد در یک تصویر دارد با توجه به اینکه کارت شناسایی ملی، یک ساختار مستطیل شکل دارد که اضلاع مجاور بر یکدیگر عمود هستند. انتخاب اینکه کدام خط به عنوان مبنای چرخش استفاده شود زاویه بین اضلاع مجاور مستطیل بوده که ۹۰ درجه است. بنابراین، در شرایطی که زاویه شیب تصویر کمتر از ۴۵ درجه باشد، فقط باید سمتی را که زاویه شیب کمتر از ۴۵ درجه دارد به عنوان خط مرجع پیدا کرد. هنگامی که زاویه چرخش تصویر بیشتر از ۴۵ درجه باشد، تشخیص اینکه کدام لبه یک لبه افقی یا یک لبه عمودی است یا پیدا کردن لبه در تصاویر بسیار پیچیده دشوار است [۸]. روش اصلاح چرخش تبدیل هاف در این پژوهش قادر به اصلاح زوایای چرخش کمتر از ۴۵ درجه است همچنین با تبدیل مقیاس

app.Application.ActiveDocument، همچنین برای اعمال تغییرات با دستور ArtLayers() لایه مادر به active_layer نسبت داده می‌شود. ۴. انتخاب فیلترهای مورد نظر و تنظیم پارامترهای شان. ضمناً لازم به ذکر است که گزارش تفصیلی افزودن پیچیدگی به دادگان در پیوست ۳ به اختصار تشریح شده است.



شکل (۲): نمونه‌ای از پیچیدگی‌های افزوده شده به دادگان (پیچیدگی‌ها شامل: انواع تاری، نویزهای تیره و روشن، چرخش، پرسپکتیو و ... هستند که با درجه‌های مختلف به صورت رندم بر روی هر مجموعه تصاویر اعمال شده‌اند)

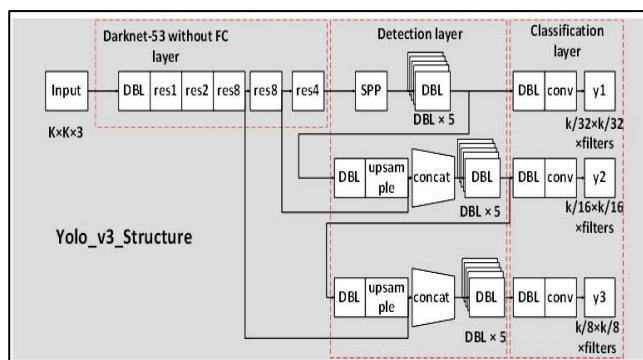
۳-۲- پیش‌پردازش و آماده‌سازی دادگان

۳-۲-۱- برچسب‌گذاری

داده‌های تصویری کارت ملی قبل از ورود به مدل برای آموزش شبکه، نیازمند برچسب‌گذاری هستند بدین منظور با کمک نرم‌افزار labelImg-master

^{۱۸} Accumulator

^{۱۷} Hough Transform



شکل (۳): معماری الگوریتم YOLOv3

تابع اتلاف موقعیت^{۲۲}: الگوریتم یولو یک تصویر را به یک سیستم شبکه ای $S \times S$ تقسیم می کند. هر شبکه روی تصویر ورودی مسئول تشخیص یک شی مرتبط است. در حال حاضر، سلول شبکه تعداد جعبه های محدود کننده در هر شی را پیش بینی می کند. هر جعبه کراندار شامل ۵ عدد عنصر (x, y, w, h) امتیاز اطمینان (Pr) است. در جایی که «x» و «y» مختصات مورد داخل تصویر ورودی هستند، «w» و «h» به ترتیب عرض و ارتفاع شی هستند. امتیاز اطمینان احتمالی است که جعبه با یک شی مرتبط است و میزان دقیق آن همان جعبه مرزی است.

اطمینان = احتمال (شیء) * $IoU^{۲۳}$

IoU برای اندازه گیری دقت موقعیت استفاده می شود. برای تابع لاس یا زیان در یولو ۳، به جای میانگین مربعات خطا^{۲۴} (MSE) از تابع آنتروپی متقابل (cross-entropy function) استفاده شده است. تابع لاس تابع آنتروپی متقابل به شرح زیر است:

$$\sum_{c=1}^m \int \log(P(x \in c)) \quad (2)$$

که در آن m تعداد کلاس ها و c نشان دهنده شاخص کلاس، X مشاهده است و $\log(P(x \in C))$ لگاریتم طبیعی است که احتمال مشاهده X متعلق به کلاس C را پیش بینی می کند. پس از تشخیص در تصاویر با استفاده از یولو ۳ بهبود یافته، کادرهای محدود کننده شی شناسایی شده را بدست می آوریم. سپس، عملیات پس از پردازش باید انجام شود زیرا چندین جعبه محدود کننده برای یک شی ظاهر شده است. برای حل این مشکل، از نگاشت جعبه مرزی یا محدود کننده^{۲۵} و جداکننده غیر حداکثری^{۲۶} (NMS) استفاده کنید. NMS تمام کادرهای محدود کننده همپوشانی را حذف می کند و کادر محدود کننده سمت راست را برمی گرداند. پیش بینی جعبه های مرزی مانند یولو ۲ است. بنابراین، شبکه چهار مختصات را برای هر جعبه مرزی پیش بینی می کند که عبارتند از tx, th, ty و ty . تصویر جعبه مرزی در شکل ۴ ارائه شده است.

RGB تصاویر به مقیاس خاکستری، تا حدودی می توان بر چالش های پیچیدگی تصاویر و پیدا کردن لبه ها غلبه کرد.

۳-۳-۳ مدل پیشنهادی

داده های پس از پیش پردازش و آماده سازی به مجموعه آموزشی^{۱۹} و آزمایشی^{۲۰} با نسبت ۷۰ به ۳۰ درصد تقسیم شدند. از آنجایی که تنوع دادگان مورد استفاده این در پروژه نسبتاً بالا، و حجم آن های برای آموزش یک شبکه عصبی بسیار کم و محدود (۱۵۵۰ تصویر کارت ملی) است بنابراین دادهای برای مجموعه ارزیابی^{۲۱} در نظر گرفته نشد. شبکه های مورد استفاده در این پروژه برای آموزش مدل، نسخه های ۳ و ۵ الگوریتم یولو (YOLOv3&5) هستند لازم به ذکر است الگوریتم مورد تاکید یولو ۳ بوده، که از نسخه ۵ آن برای مقایسه و ارزیابی نتایج یولو ۳ بهره برده شد در ادامه شرح مختصری از معماری الگوریتم یولو با تاکید بر نسخه ۳ ذکر شده است.

۱-۳-۳-۳ معماری یولو

الگوریتم یولو یک روش شناسایی شیء end to end است که مستقیماً موقعیت و دسته شی را در خروجی رگرسیون کرده و یک تشخیص سریع انجام می شود. ساختار شبکه یولو ۳ عمدتاً شامل دو لایه است. لایه اول استخراج ویژگی است که از شبکه Darknet-53 برای بدست آوردن نقشه ویژگی استفاده می کند و سلول های شبکه را به اندازه نقشه ویژگی تنظیم می کند. برای هر سلول شبکه، سه کادر محدود کننده پیش بینی می شود. علاوه بر این، در شبکه یولو ۳، مقیاس نقشه های ویژگی مورد استفاده برای تشخیص شی به سه نسبت، یعنی $۱۳ * ۲۶ * ۵۲$ و $۵۲ * ۵۲ * ۵۲$ تقسیم می شود. لایه دوم فرآیند خروجی است که اطلاعات مکان را تولید می کند (یعنی $x: y: w: h: pr$). سپس، کادر محدود کننده بدون کیفیت توسط الگوریتم Itering حذف می شود و از الگوریتم ساخت خط متن برای تولید یک خط متن استفاده شده که برای تشخیص متن به مرحله بعدی انتقال داده می شود. یولو ۳ از یک روش ساخت شبیه به هرم ویژگی استفاده می کند تا سه نقشه ویژگی را از طریق دو لایه upsampling بدست آورد. که امر امکان شناسایی اشیاء با اندازه های مختلف را بویژه در زمان واقعی فراهم می کند [۱۴]. معماری پایه شبکه یولو ۳ در شکل ۳ نشان داده شده است. شامل سه لایه اصلی شبکه استخراج ویژگی (Darknet-53)، لایه تشخیص و لایه طبقه بندی است که DBL جزء اصلی یولو ۳ است که از لایه Convolutional، لایه BN و لایه Leaky ReLU تشکیل شده است [۱۳].

^{۲۳} Intersection over Union (IoU)

^{۲۴} Mean Squared Error

^{۲۵} Bonding box

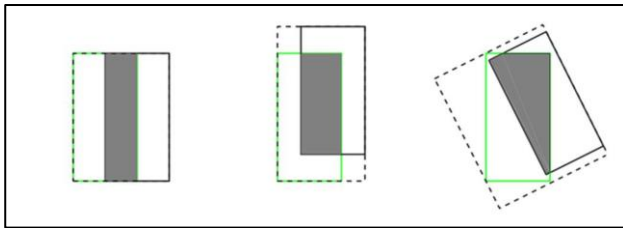
^{۲۶} Non-Maximum Suppression

^{۱۹} Train Set

^{۲۰} Test Set

^{۲۱} Dev Set

^{۲۲} Location Loss Function



شکل (۵) سه روش متقاطع مختلف با یک مقدار IOU (۰,۳۳) مطابقت دارند.

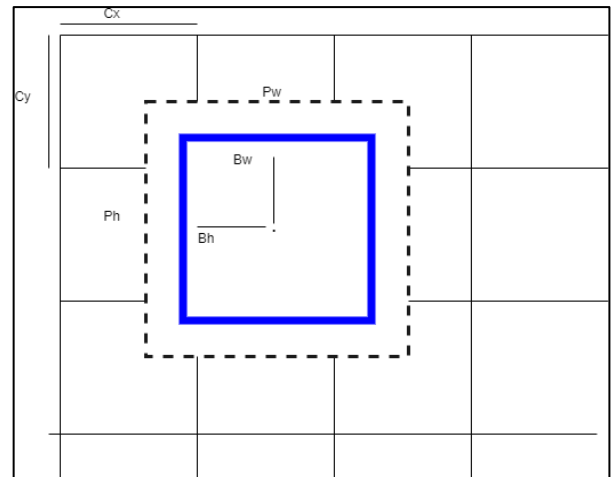
برای حذف موارد تکراری از شئی مشابه، یولو جداکننده غیر حداکثری (NMS) را انجام می‌دهد. اگر بین هر یک از پیش‌بینی‌های تصویر $\leq \text{IoU}$ آستانه داشته باشیم، سرکوب غیر حداکثری، پیش‌بینی با کمترین امتیاز اطمینان را حذف می‌کند. یولو ۳ یک کادر محصورکننده^{۲۸} برای هر شئی اختصاص می‌دهد. در یولو ۳ از شبکه‌های هرمی ویژگی برای پیش‌بینی جعبه‌ها در ۳ مقیاس مختلف استفاده می‌شود [۱۴].

۲-۳-۳- آموزش مدل

الگوریتم‌های مورد استفاده برای آموزش مدل در پروژه حاضر، از دو نمونه شبکه که از پیش آموزش دیده یولو ۲۹۳ و یولو ۳۰۵ بر روی مجموعه داده‌گان COCO^{۳۱} انتخاب شدند که هاپیرپارامترهای آن به صورت تجربی در طی چندین مرحله آزمون-خطا در فرآیند آموزش مدل بر اساس مجموعه داده‌گان مورد نظر تنظیم شد (جدول ۲) و نهایتاً آموزش مدل با دو الگوریتم یولو ۳ و یولو ۵ در محیط Google Colab با زبان برنامه نویسی پایتون نسخه 3.8.10 صورت گرفت. ضمناً لازم به ذکر است که جزئیات تفصیلی فرآیند آموزش مدل و گزارش کد الگوریتم‌ها در پیوست ۳ به اختصار تشریح شده‌اند.

جدول (۲): هاپیرپارامتر الگوریتم‌های پیشنهادی

یولو ۵	یولو ۳	
۶۴۰*۶۴۰	۴۱۶*۴۱۶	سایز تصاویر
۶۴	۶۴	اندازه بچ
۱۰۰	۳۵۴	دوره (Epoch)
۰,۰۱	۰,۰۰۱	نرخ یادگیری
۰,۹۳۷	۰,۹	ممنت
Adam	Leaky ReLU	تابع فعال‌ساز
-	۰,۳	NMS



شکل (۴) جعبه قبلی (نقطه سیاه)، جعبه پیش‌بینی شده (آبی)، پیش

بینی جعبه مرزی

برای محاسبه مختصات جعبه محصورکننده از فرمول‌های زیر استفاده کنید:

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \quad (۳)$$

محاسبه IoU به محاسبه میانگین دقت متوسط^{۲۷} (MAP) کمک می‌کند. IoU به ما کمک می‌کند تا تعیین کنیم که آیا جعبه پیش‌بینی شده مثبت کاذب، مثبت درست یا منفی کاذب است. در اینجا منفی واقعی وجود ندارد زیرا فرض می‌کند که جعبه مرزی ممکن است چیزی در داخل خود داشته باشد، که به نوبه خود به عنوان جعبه مرزی خالی نیست. ما یک مقدار آستانه را از قبل برای IoU تا ۰,۵ تعریف می‌کنیم که معمولاً استفاده می‌شود.

• اگر $\text{IoU} > 0.5$ باشد، می‌توان گفت که مثبت واقعی است.

• اگر $\text{IoU} < 0.5$ باشد، مثبت کاذب است.

• اگر $\text{IoU} > 0.5$ باشد، اما اگر شبیهی به اشتباه طبقه بندی شده باشد، منفی کاذب است.

در کد به روز شده الگوریتم یولو ۳، از دست دادن موقعیت از تابع اتلاف IOU برای رگرسیون استفاده می‌کند، اما فقط برای کادرهای مستطیلی افقی مناسب است و نمی‌تواند با الگوریتم شناسایی هدف در حال چرخش سازگار شود، و از دست دادن IOU دارای دو نقص زیر است:

(۱) اگر دو هدف با هم همپوشانی نداشته باشند، IOU خواهد بود. در این زمان، صرف نظر از فاصله بین دو جعبه، IOU نمی‌تواند منعکس شود. اگر IOU به عنوان تابع ضرر استفاده شود، گرادینت صفر است و نمی‌توان آن را بهینه کرد.

(۲) IOU نمی‌تواند بین ترازهای مختلف بین دو شی تمایز قائل شود. به طور دقیق‌تر، IOU دو جسم روی هم‌پوشانی با سطح تقاطع یکسان در جهات مختلف دقیقاً یکسان خواهد بود، همانطور که در شکل ۵ نشان داده شده است.

^{۲۷} Mean Average Precision (MAP)

^{۲۸} Anchor Box

^{۲۹} <https://github.com/AlexeyAB/darknet#how-to-train-to-detect-your-custom-objects>

^{۳۰} <https://github.com/ultralytics/yolov5>

^{۳۱} Common Objects in Context

شبکه مورد نظر ابتدا با داده‌های فاقد پیچیدگی به وسیله الگوریتم یولو ۳ آموزش داده شد و نتایج آن حکایت از میانگین دقت ۹۹ درصدی شبکه در تشخیص کاراکترهای کارت ملی هم در فاز آموزشی و هم در فاز آزمایشی دارد که این میزان دقت فقط براساس وزن‌های حاصله از آموزش مدل با دادگان ساده بوده (قبل از افزودن پیچیدگی‌ها)، به گونه‌ای که با چرخش بیش از ۱۰ درجه زاویه تصاویر، شبکه قادر به تشخیص صحیح کاراکترها نیست. همچنین لازم به ذکر است در فرآیند آموزش شبکه با داده‌های ساده، بیش‌برازش الگوریتم بر روی کلاس کارت ملی (Card) رخ داد و به نوعی الگوریتم در هر تصویر (حتی تصاویر فاقد کارت ملی)، وجود کارت ملی تشخیص را داده، که این چالش با تغییر مقادیر هاپیرپارامترها و افزایش تعداد اپوک‌های تا حدودی برطرف شد (جدول ۳ و شکل ۱).



شکل (۶) عملکرد الگوریتم یولو ۳ بر اساس وزن‌های آموزش دیده با داده‌های فاقد پیچیدگی در نمونه تصاویر دنیای واقعی

با افزودن انواع داده‌های پیچیده به مدل، میانگین دقت الگوریتم یولو ۳ با یک و نیم درصد کاهش به ۹۷٫۵ درصد رسید با این وجود الگوریتم قادر به تشخیص کاراکترهای کارت ملی در تصاویر با انواع پیچیدگی‌های خاص نظیر چرخش زاویه، نویز، تاری و ... بوده است بهترین عملکرد الگوریتم یولو ۳ در تشخیص کاراکترها مربوط به کلاس خود کارت ملی (Card) و عکس (Avatar) با میانگین دقت ۹۸٪، و ضعیف‌ترین عملکرد مربوط به تشخیص کلاس‌های نام و نام خانوادگی با میانگین دقت ۹۶٪ است که این اختلاف با توجه به عملکرد بهتر یولو در تشخیص اشیاء با مقیاس بزرگتر نسبت به اشیاء با مقیاس کوچک‌تر، پدیده معمولی به شمار می‌رود (جدول ۳ و شکل ۷).

جدول (۳): میانگین دقت متوسط عملکرد الگوریتم یولو ۳ در تشخیص کاراکترهای کارت ملی، قبل و بعد از افزودن پیچیدگی به داده‌ها

۴- نتایج و بحث

برای ارزیابی عملکرد دو الگوریتم مورد استفاده و مقایسه نتایج آن‌ها از روابط زیر بهره گرفته شد:

$$P = \frac{T_C}{T_D}, \quad (4)$$

$$R = \frac{T_C}{T_G}, \quad (5)$$

$$F = \frac{2(R \times P)}{R + P}, \quad (6)$$

TC تعداد کل کاراکترهایی را که می‌توان به درستی تشخیص داد (تعداد کاراکترهای صحیح شناسایی شده)، TD تعداد کل کاراکترهایی شناسایی شده توسط الگوریتم را نشان می‌دهد (کاراکترهای صحیح و غلط شناسایی شده) و TG نشان دهنده تعداد کل کاراکترهای شناسایی شده موجود در مجموعه داده‌ها آزمایشی یا در واقع تعداد کاراکترهای صحیح شناسایی شده + تعداد کاراکترهای غیر قابل شناسایی است.

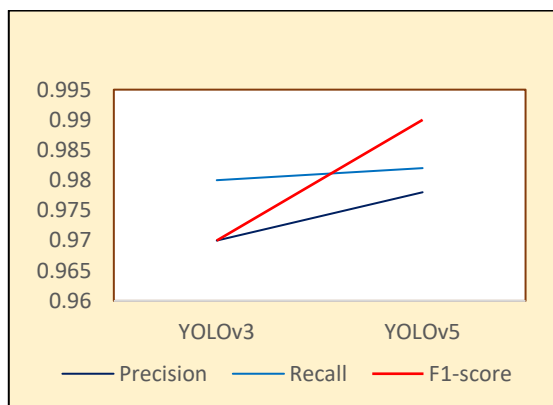
از رابطه‌های بالا می‌توان دریافت که میزان دقت^{۳۲} (P) در واقع نرخ صحیح تشخیص کاراکترهایی است که باید شناسایی شود، و نرخ فراخوانی^{۳۳} (R) نسبت به میزان دقت P می‌تواند نرخ تشخیص آنچه که از دست رفته یا قابل شناسایی نیست را منعکس کند. هنگامی که کاراکترهایی از دست رفته یا شناسایی نشده بیشتری در مجموعه داده وجود داشته باشد مقدار R مربوطه کمتر است. هنگامی که کاراکترهای اشتباه شناسایی شده بیشتری در مجموعه داده وجود داشته باشد مقدار P مربوطه کمتر است. میانگین عملکرد^{۳۴} F، وزن و میانگین بین میزان دقت P و نرخ فراخوان R را منعکس می‌کند این رابطه نه تنها می‌تواند به طور جامع نتایج میزان دقت و نرخ فراخوانی را منعکس کند بلکه نشان دهنده عملکرد جامع الگوریتم نیز است [۱۳].

^{۳۲} Precision

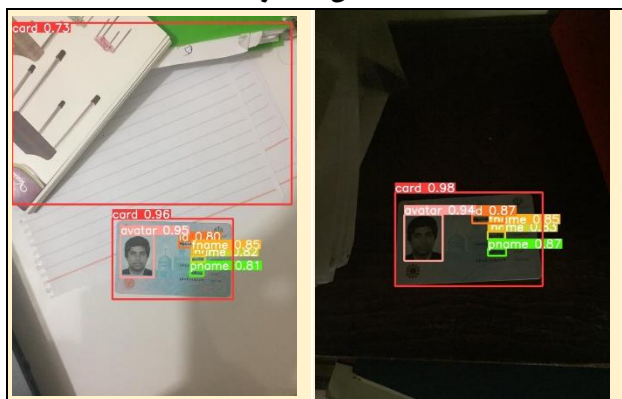
^{۳۳} Recall

^{۳۴} F1-score

نام پدر (Pname)	۹۷,۹۲	۹۸,۴
Precision	۰,۹۷	۹۷,۸
Recall	۰,۹۸	۹۸,۲
F1-score	۰,۹۷	۹۸,۱
سرعت تشخیص (زمان اجرا)	۷,۰۰۸۶	۵,۹۴۲۵

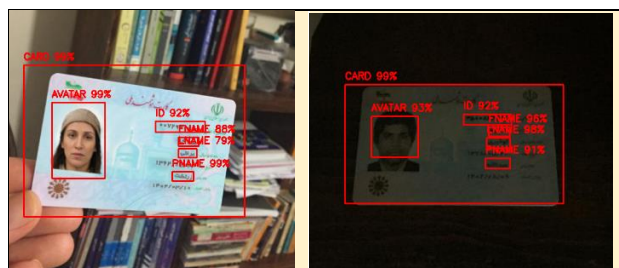


شکل (۸) دیاگرام مقایسه عملکرد الگوریتم یولو ۳ و ۵ در سه شاخص P, R و F



شکل (۹) عملکرد الگوریتم یولو ۵ در نمونه تصاویر دنیای واقعی، تشخیص اشتباه بخشی از فضای تصویر به عنوان کلاس کارت

کلاس‌ها	قبل از افزودن پیچیدگی به داده‌ها					بعد از افزودن پیچیدگی به داده‌ها				
	MAP%	MAP%	MAP%	MAP%	MAP%	MAP%	MAP%	MAP%	MAP%	MAP%
کارت (Card)	۹۹,۴۱	۹۹,۴۲	۹۸,۳۳	۹۷,۰۰	۹۸,۳۳	۹۹,۴۱	۹۹,۴۲	۹۸,۳۳	۹۷,۰۰	۹۸,۳۳
عکس (Avatar)	۹۹,۴۱	۹۹,۴۱	۹۸,۰۵	۹۶,۴۵	۹۸,۰۵	۹۹,۴۱	۹۹,۴۱	۹۸,۰۵	۹۶,۴۵	۹۸,۰۵
شماره ملی (Id)	۹۹,۰۲	۹۹,۰۳	۹۷,۵۳	۹۵,۰۷	۹۷,۵۳	۹۹,۰۳	۹۹,۰۳	۹۷,۵۳	۹۵,۰۷	۹۷,۵۳
نام (Fname)	۹۹,۴۴	۹۹,۴۶	۹۶,۲۷	۹۲,۵۴	۹۶,۲۷	۹۹,۴۴	۹۹,۴۶	۹۶,۲۷	۹۲,۵۴	۹۶,۲۷
نام خانوادگی (Lname)	۹۹,۹۹	۹۹,۹۹	۹۶,۶۲	۹۱,۵۱	۹۶,۶۲	۹۹,۹۹	۹۹,۹۹	۹۶,۶۲	۹۱,۵۱	۹۶,۶۲
نام پدر (Pname)	۹۹,۵۵	۹۹,۵۵	۹۷,۹۲	۹۳,۸۶	۹۷,۹۲	۹۹,۵۵	۹۹,۵۵	۹۷,۹۲	۹۳,۸۶	۹۷,۹۲
میانگین	۹۹,۱۴	۹۹,۱۴	۹۷,۴۵	۹۴,۴۱	۹۷,۴۵	۹۹,۱۴	۹۹,۱۴	۹۷,۴۵	۹۴,۴۱	۹۷,۴۵



شکل (۷) عملکرد الگوریتم یولو ۳ بر اساس وزن‌های آموزش دیده با داده‌های دارای پیچیدگی ویژه در نمونه تصاویر دنیای واقعی

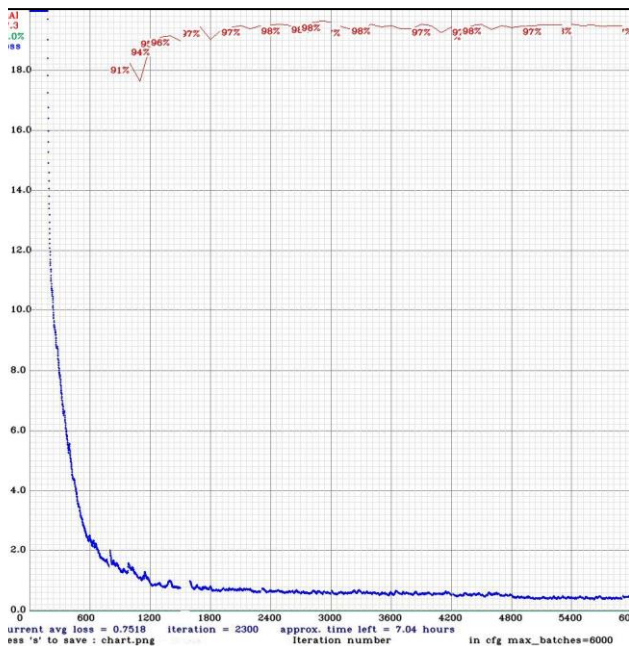
نتایج آموزش مدل با الگوریتم یولو ۵ نیز حکایت از بهبود جزئی عملکرد آن نسبت به یولو ۳ در تشخیص انواع کاراکترها دارد. میانگین دقت متوسط الگوریتم یولو ۵ در تشخیص کاراکترها ۹۸٪ گزارش شد همچنین لازم به ذکر است این مقدار از دقت در یولو ۳ با صرف مدت زمان آموزش بسیار کمتر در ۱۰۰ اپوک، نسبت به یولو ۳ با ۳۵۴ اپوک به دست آمده است. عدم بروزسانی کامل وزن‌ها در الگوریتم یولو ۵ (به دلیل محدودیت زمانی و سخت افزاری) موجب تشخیص اشتباه یک شئی یا قسمتی از فضا در یک تصویر به عنوان کلاس کارت ملی شده است سرعت تشخیص یک تصویر در الگوریتم یولو پنج ۵,۹ فرم بر ثانیه بوده که در مقایسه با سرعت تشخیص ۷ فرم بر ثانیه یولو سه انداکی بیشتر است (جدول ۴ و شکل ۱۱-۸).

جدول (۴) : عملکرد یولو ۳ و ۵ در شاخص‌های مختلف

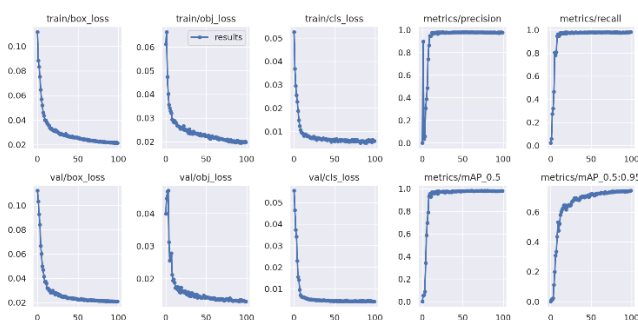
YOLOv3	YOLOv5	
۹۸,۹	۹۸,۳۳	کارت (Card)
۹۸,۷	۹۸,۰۵	عکس (Avatar)
۹۸,۱	۹۷,۵۳	شماره ملی (Id)
۹۷,۹	۹۶,۲۷	نام (Fname)
۹۶,۵	۹۶,۶۲	نام خانوادگی (Lname)

۵- نتیجه گیری

امروزه با رواج روز افزون سامانه‌های مکانیزه ارائه خدمات، تشخیص هویت و استخراج هوشمند اطلاعات مندرج در اسناد هویتی از جمله کارت ملی ضرورتی بنیادین است که در این پروژه رویکردی هوشمند مبتنی بر الگوریتم‌های یولو ۳ و ۵ به منظور تشخیص کاراکترهای کارت ملی در تصاویر با پیچیدگی‌های خاص طبیعی ارائه شد. در این راستا به دلیل محرمانگی پایگاه داده کارت ملی (سازمان ثبت احوال) ابتدا تولید داده تصویری کارت ملی با الگوریتم ایجاد شده شخص و اسکرپت نرم افزار گرافیکی صورت گرفت سپس با نظر به ماهیت پروژه حاضر یعنی شناسایی کاراکتر مدارک هویتی؛ بنابراین الگوریتم و شبکه پیشنهادی بایستی قابلیت تعمیم‌پذیری با انواع داده‌های ورودی یعنی تصاویر با کیفیت و پیچیدگی‌های گوناگون را داشته باشد که در این خصوص با تکنیک‌های ترکیبی برپایه ۲۳ نوع فیلتر و عملگر تغییر تصویر نرم افزار گرافیکی فتوشاپ با انتخاب نرخ تصادفی مقادیر پارامترهای هر فیلتر، پیچیدگی بر دادگان مورد نظر اعمال شد که قابلیت تغییر مقادیر و نرخ حداکثر و حداقل پارامترهای فقط یک فیلتر و یا عملگر تغییر عکس در یک تصویر می‌تواند بیش از ۱۰۰ داده تصویری دیگر با کیفیت بصری و مورفولوژی مختلف تولید کند. نتایج عملکرد الگوریتم‌های پیشنهادی حکایت از دقت ۹۷ درصدی الگوریتم یولو ۳ و دقت ۹۸٫۵ درصدی یولو ۵ در تشخیص کاراکترهای کارت ملی در تصاویر با پیچیدگی‌های خاص دارد. پروژه پیش‌رو از مراحل مقدماتی استخراج و شناسایی کاراکترهای کارت ملی به شمار می‌رود که با تکمیل مراحل و بخش‌های دیگر، نظیر شناسایی کاراکترها و نیز احراز هویت مبتنی بر تصویر کارت ملی امید می‌رود مدل پیشنهادی نهایی به یک محصول کاربردی در انجام امورات گوناگون نیازمند به استخراج اطلاعات هویتی تبدیل شود.



شکل (۱۰) دیاگرام دقت و Loss الگوریتم یولو ۳ در آموزش مدل با دادگان پیچیده



شکل (۱۱) عملکرد الگوریتم یولو ۵ در آموزش مدل با دادگان پیچیده بر مبنای متریک‌های مختلف

میانگین دقت متوسط الگوریتم‌های یولو ۳ و ۵ پیشنهادی پروژه حاضر در تشخیص کاراکترهای کارت شناسایی با تصاویر دارای پیچیدگی‌های خاص نسبتاً مطلوب ارزیابی می‌شود به گونه‌ای که مدل پیشنهادی پروژه پیش‌رو با دقت تشخیص ۹۸٪ در مقایسه با؛ دقت ۹۳٪ الگوریتم یولو ۵ در تشخیص کاراکترهای کارت‌های شناسایی اندونزیایی [۵]، دقت الگوریتم یولو ۵ مورد استفاده در مکان‌یابی و تشخیص کاراکترهای پاسپورت روسی با میانگین ۷۸٪ [۱۱]، و همچنین دقت تشخیص ۸۶٪ متن‌های چند جهته در صحنه و زمان واقعی با الگوریتم یولو ۳ [۱۳] عملکرد بهتری داشته است البته لازم به ذکر است که مقایسه عملکرد الگوریتم‌های مزبور با مدل پیشنهادی به صورت مطلق با نظر به نوع و حجم دادگان و قدرت شبکه‌های مورد استفاده قابل استناد نیست بنابراین مقصود از مقایسه دقت نتایج الگوریتم‌های پیشنهادی با نمونه‌های نام‌برده، ارزیابی عملکرد نسبی آن جهت رسیدن به یک مدل با نوع و مقادیر پارامتر و هاپر پارامترهای تعمیم‌پذیر با عملکرد قابل استناد در مراحل شناسایی و استخراج کاراکترهای کارت شناسایی ملی است.

سپاسگزاری

ابتدا به رسم قدردانی از زحمات ارزشمند استاد درس یادگیری ماشین عالیجناب کیوان‌راد خاضعانه سپاسگزاری می‌نمایم. همچنین از دستیاران محترم درس مهندسان رحمانی، فیضی و رجبی به پاس رهنمودهای ارزشمندشان کمال تشکر را داریم.

ضمایم

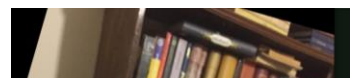
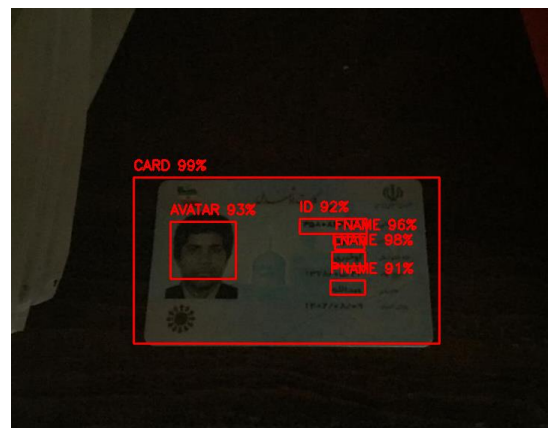
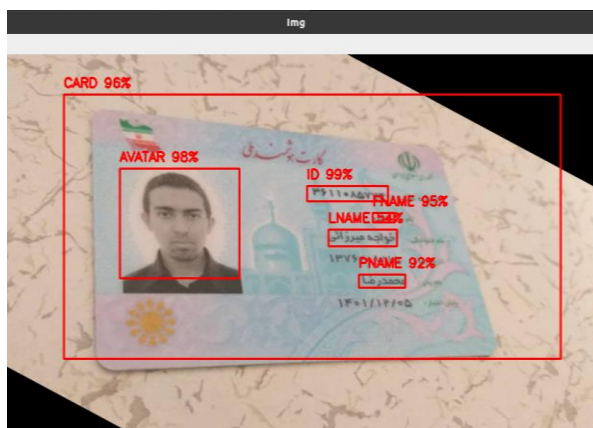
1. <https://github.com/AlexeyAB/darknet#how-to-train-to-detect-your-custom-objects> (لینک شبکه یولو ۳ مورد استفاده برای آموزش مدل)
2. <https://github.com/ultralytics/yolov5> (لینک شبکه یولو ۵ مورد استفاده برای آموزش مدل)
3. <https://generated.photos.com> (لینک تصاویر مورد استفاده در تولید آواتار کارت ملی)

مراجع

- [1] N. E. Thanh, L. H. Lam and N. H. Nam (2020) "An Efficient Method for Automatic Recognizing Text Fields on Identification

- [16] H. D. Liem and *et al.*, (2019 ") **FVI: An End-to-end Vietnamese Identification Card Detection and Recognition in Images**", *NICS 2018 Proceedings of 201 5th NAFOSTED Conference on Information and Computer Science*, pp. 338–340, doi: 10.1109/NICS.2018.8606831.
- [17] Y. Gao, Ch. Xu, Z. Shi and H. Zhang (2019) "**Bank Card Number Recognition System based on Deep Learning**" Association for Computing Machinery: <https://doi.org/10.1145/3358331.3358383>.
- [18] C. Shorten and T. M., Khoshgoftaar (2019) " **Asurvey on Image Data Augmentation for Deep Learning**" *Journal of Big Data*, 6.60, pp:1-48.Springer, <https://doi.org/10.1186/s40537-019-0197-0>.
- Card" *VNU Journal of Science: Mathematics – Physics*, 36.1, pp: 64-70.
- [2] J. Zhu, H. Ma, J. Feng and L. Dai (2018) "**ID Card Number Detection Algorithm based on Convolutional Neural Network**" AIP Conference Proceedings, 040124. <https://doi.org/10.1063/1.5033788>.
- [3] Sh. Long, X. He and C. Yao (2020) "**Scene Text Detection and Recognition: The Deep Learning Era**" *International Journal of Computer Vision*: 10.1007, Springer Nature.
- [4] X. Chen, L. Jin, Y. Zhu, C. Luo, and T. Wang, (2020) "**Text Recognition in the Wild: A Survey**" *arXiv preprint arXiv: 2005.03492*.
- [5] A. Chandra and R. Stefanus (2021) "**An End-to-End Optical Character Recognition Pipeline for Indonesian Identity Card**" 9th International Conference on Information and Communication Technology (ICoICT): UTC from IEEE Xplore.
- [6] J. Wang, R. Wu, and Sh. Zhang (2021) "**Robust Recognition of Chinese Text from Cellphone-acquired Low-quality Identity Card Images Using Convolutional Recurrent Neural Network**" *Sensors and Materials*, 33.4, pp: 1187–1198 1187.
- [7] Y. Li, H. Chang and D. Lin (2022) "**Large-Scale Printed Chinese Character Recognition for ID Cards Using Deep Learning and Few Samples Transfer Learning**" MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations: <https://doi.org/10.3390/app12020907>.
- [8] J. Xu, and X. Wu, (2018) "**A System to Localize and Recognize Texts in Oriented ID Card Images**" Authorized licensed use limited to: ULAKBIM UASL – CUKUROVA UNIVERSITESI: UTC from IEEE Xplore.
- [9] S. Lee and *et.al* (2018) "**Variations of AlexNet and GoogLeNet to Improve Korean Character Recognition Performance**" *J Inf Process Syst*, 14.1, pp: 205-217.
- [۱۰] کاشف، شیما، نظام‌آبادی‌پور، حسین و الهام شعبانی‌نیا، "مروری بر روش های یادگیری ژرف در بازشناسی نوری نویسه ها با تاکید بر رسم الخط های فارسی، عربی و اردو" *مجله ماشین بینایی و پردازش تصویر*، ۸، ۴: ۵۱-۸۵، ۱۴۰۰.
- [11] V. Mustafina and S. Ivanov (2021) " **Identity Document Recognition: Neural Network Approach** " *International Russian Automation Conference (RusAutoCon)*, 978-1-6654-1523-1/21/\$31.00 ©2021 IEEE, pp: 806-810.
- [12] P. Raka and S. P. Agrawal (2019) "**OCR to read embossed text from Credit/Debit card**" *International Journal of Advance Research, Ideas and Innovations in Technology*. 5: 2.
- [13] L. Xiao and *et al.*, (2021) "**Multi-Directional Scene Text Detection Based on Improved YOLOv3**" MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations, 21. 4870: <https://doi.org/10.3390/s21144870>.
- [14] L. Cao and *et al.*, (2020) " **A Text Detection Algorithm for Image of Student Exercises Based on CTPN and Enhanced YOLOv3**" *IEEE ACCESS*, 8, PP: 176924-176934, <https://creativecommons.org/licenses/by/4.0/> .
- [15] X. Wang and *et al.*, (2021) " **R-YOLO: A Real-Time Text Detector for Natural Scenes with Arbitrary Rotation**", *Sensors*, 21, 888, pp: 1-20. <https://doi.org/10.3390/s21030888>.

پیوست ۱ (نمونه تصاویر خروجی نهایی مدل در شناسایی کاراکترها تصاویر کارت ملی در دنیای واقعی)
YOLO v3



Localization and detection of ID card characters in images with special complexities based on the YOLOv3&5 algorithm

Samrand Majnooni¹, Talhe Ghodousiyan², Ali Khajeh mirzaei³

¹Master student of artificial intelligence, Faculty of Electrical and Computer Engineering, Malek Ashtar University of Technology, Tehran

Samrand.majnooni@gmail.com

²Master student of artificial intelligence, Faculty of Electrical and Computer Engineering, Malek Ashtar University of Technology, Tehran

T.ghodusiean@gmail.com

³Master student of artificial intelligence, Faculty of Electrical and Computer Engineering, Malek Ashtar University of Technology, Tehran

alimirzaei420@yahoo.com



Abstract

The ID card is one of the essential identity documents of Iranian citizens, as it is an identity reference document in banking operations, bureaucracies and organizational services. Also, most of these services need to enter and extract information contained in the national card. Nowadays, with the increasing prevalence of mechanized systems for providing services, identification and intelligent extraction of information contained in identity documents, including the national card, it is a fundamental necessity that in this paper, an intelligent approach based on Convolution neural networks (CNNs) to detection of ID card characters Presented with special natural complexities.

In this study, due to the confidentiality of ID cards, first an algorithm was written to produce images of the ID card and its characters, and then a combination of Photoshop graphic software script and common Python data generation algorithms was used to add complexity to the images. Hough Transformation was then used to modify the rotation of the images and the Yolov3&5 algorithms were used to Localization and detection of the ID card characters. The results indicate the optimal performance of the proposed algorithms; YOLO 5 has 98.5% accuracy and YOLO 3 has 97% accuracy in detection of the ID card characters in complex images.

Keywords

ID card, convolutional neural networks (CNN), YOLO, Hough Transform, detection, complexity.