**بررسی امکان طبقه‌بندی و مکان‌یابی کاروانسراهای ایوان‌دار در تصاویر ماهواره‌ای با بهره‌گیری از تکنیک‌های پردازش تصویر مبتنی بر یادگیری عمیق**

فرناز فرجی۱، محمدحسین سالاری۲\*

۱دانشجوی دکترای رشته مرمت و احیای ابنیه و بافت‌های تاریخی، دانشگاه تهران

\*۲ نویسنده مسئول، دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر، گرایش هوش مصنوعی و رباتیکز، دانشگاه آزاد اسلامی، واحد شیراز

**چکیده:**

با وجود افزایش روز افزورن استفاده از یادگیری عمیق[[1]](#footnote-1) در پردازش تصاویر، استفاده از این تکنولوژی‌ در زمینه‌های مرتبط با میراث فرهنگی بسیار محدود بوده است. علت این استفاده محدود را می‌توان به موانعی مانند نبود پایگاه‌های‌‌داده[[2]](#footnote-2) مرتبط و هزینه‌بر بودن ایجاد پایگاه داده‌ی جدید‌، پیچیدگی کار با شبکه‌های عصبی عمیق، نیاز به منابع سخت‌افزاری گران قیمت برای آموزش شبکه‌های عصبی عمیق و ... مرتبط دانست. جهت بررسی امکان استفاده از تکنیک‌های پردازش تصویر میتنی بر یادگیری عمیق در مکان‌یابی و طبقه‌بندی مسائل مرتبط با میراث فرهنگی، در این مقاله ابتدا یک پایگاه داده تصاویر ماهواه‌ای از کاروانسراها (شامل ۳۱۹ تصویر از کاروا‌سرا، ۶۳۶ تصویر از محیط پیرامون هر کاروانسرا و ۱۱۳‌ تصویر از سایر بناها) با استفاده از منابع آنلاین رایگان ایجاد شده است، سپس با استفاده از تکنیک یادگیری انتقالی[[3]](#footnote-3) شبکه‌های عصبی ResNet18، ResNet34 و ResNet50 برای طبقه‌بندی تصاویر و وزن‌های سبک[[4]](#footnote-4) معماری YOLOv5برای تشخیص محل کاروانسرا بر روی این پایگاه داده آموزش داده شده‌اند. در بررسی‌های انجام شده بالاترین دقت بدست آمده برای طبقه بندی تصاویر کاراونسراها برابر با 85.83% بوده دقت الگوریتم YOLOv5 در یافتن محل کاروانسرا برابر با 82.47% mAP\_0.5 بوده است. با توجه به نتایج به دست آمده می‌توان گفت که وزن‌های سبک معماری YOLOv5قابلت آن را دارند تا به عنوان ابزاری موثر در تشخیص محل کاروانسراهای ایوان‌دار در تصاویر ماهواره‌ای استفاده شوند، همچنین یافته‌های این مقاله باینگر این موضوع می‌باشد که شبکه‌های عصبی عمیق می‌توان برای طبقه‌بندی تصاویر ماهواره‌ای شامل کاروانسرا بهره برد.

**واژگان کلیدی:** بینایی ماشین (Computer Vision)، یادگیری عمیق (Deep Learning)، کاروانسراهای ایوان‌دار، تصاویر ماهواره‌ای، شبکه‌های عصبی پیچیده (CNNs)، الگوریتم YOLOv5

**۱- مقدمه**

میراث فرهنگی و به‌طور ویژه میراث معماری و شهری به‌عنوان یکی از قدرتمند‌ترین ابزار انتقال و حفاظت از هویت جامعه در طول تاریخ بشر شناخته شده ‌است. ارزش والا و بی‌بدیل این آثار ضرورت اقدامات مراقبتی و حفاظتی از این میراث را دوچندان می‌سازد. عملیات مستندنگاری، ثبت و برداشت از سازه‌ها، ساختارها و ابنیه تاریخی و فرهنگی به‌عنوان یکی از اقدامات اساسی در حفاظت از میراث فرهنگی و دانش تاریخی جامعه شناخته می‌شود که عمل به آن می‌تواند به تصمیم‌گیری‌های بهینه در فرایند حفاظت از میراث یاری رساند. در سال‌های اخیر، استفاده از ابزارها و تکنولوژی‌های نوین به انتخابی کارآمد، کم‌هزینه و کم ریسک در مراقبت، حفاظت، مستندنگاری و مدیریت میراث فرهنگی بدل شده ‌است که بقای طولانی مدت این آثار را تضمین می‌کند (Belhi et al, 2019). از میان تکنولوژی‌ها و فناوری‌های معاصر ابزارهایی نظیر فتوگرامتری و برداشت از راه دور، استفاده از تکنولوژی‌های واقعیت مجازی و تکنیک‌های هوش مصنوعی در شاخه‌های مختلف به‌عنوان بخشی از راهکارهای دیجیتال‌سازی میراث فرهنگی شناخته شده‌اند. از میان ابزارهای ذکر شده، تکنیک‌های یادگیری عمیق، طبقه‌بندی و شناسایی آثار به کمک پردازش تصاویر و استفاده از تکنیک‌های شبکه‌های عصبی بیشترین کاربرد را در علوم حفاظت داشته‌اند (Kersten & Lindstaedt, 2012:2). از جمله مزایای استفاده از یادگیری ماشین و یادگیری عمیق در حوزه‌ی حفاظت از میراث فرهنگی می‌توان به امکان گسترش دسترسی کاربران به مجموعه داده‌های عظیم به منظور طبقه‌بندی و سازمان‌دهی اطلاعات موجود، تقویت فرصت به‌روزرسانی و بازگردانی آرشیو داده‌های تاریخی برای استفاده در عملیات حفاظت و مدیریت میراث و دقت بسیار بالای تکنیک‌های زیرمجموعه یادگیری عمیق در تشخیص و جستجوی تصاویر و اشیا نام برد. استفاده از یادگیری ماشینی باعث کاهش زمان و هزینه‌ی استفاده از نیروی انسانی متخصص در مدیریت پروژه‌های میراث فرهنگی و تسهیل فعالیت‌های متخصصان حفاظت و مرمت میراث را می‌شود (Condorelli et al, 2020).

پژوهش حاضر بر آن است تا به بررسی امکان استفاده از تکنیک‌های یادگیری عمیق و پردازش تصاویر ماهواره‌ای را در طبقه‌بندی و مکان‌یابی مجموعه کاروانسراهای باقی‌مانده در کشور بپردازد.کاروانسراها مجموعه‌ای از ابنیه تاریخی درون و برون شهری در پهنه ایران‌اند که سابقه احداث آن‌ها به گذشته‌های دور باز می‌گردد و به گواه منابع مختلف کشور ایران در احداث کاروانسراها مبتکر و متقدم بود ‌است (کیانی، ۱۳۹۲). متاسفانه بسیاری از کاروانسراهای موجود در ایران به علت متروک ماندن و عدم رسیدگی و مرمت مستمر تخریب شده و یا تحول و تغییر یافته‌اند و تنها تعداد اندکی از آن‌ها در وضعیتی مطلوب حفاظت شده‌اند. اهمیت طبقه‌بندی و مکان‌یابی اتوماتیک این بناها در آن است که نخست، تاکنون بشکل مدون و دیجیتال هیچ‌گونه پایگاه از موقعیت جغرافیایی و نقشه‌ی پراکندگی این ابنیه تاریخی ارزشمند ایران به شکل رایگان و در بستر اینترنت در دسترس قرار نگرفته است و دوم آنکه، طبقه‌بندی و مکان‌یابی این ابنیه علاوه‌بر کمک به مستندنگاری کاروانسراهای فراموش شده به شناسایی مسیرهای تاریخی کاروان‌رو و مهم در سرزمین ایران یاری می‌رساند. این امیر می‌تواند نقش به‌سزایی در مطالعات کارشناسان مطالعات تاریخ معماری ایران و حفاظت‌گران و مرمتگران میراث معماری و شهری داشته ‌باشد.

به منظور نیل به اهداف ذکر شده در این مقاله ابتدا با استفاده از منابع مختلف رایگان دو پایگاه داده از تصاویر ماهواره‌ای و یک پایگاه داده از موقعیت جغرافیایی کاروانسراهای ایوان‌دار ایران تهیه شده است. سپس با استفاده از این پایگاه‌داده‌ها الگوریتم‌های یادگیری عمیق خانواده ResNet برای بررسی توانایی یادگیری عمیق در طبقه‌بندی تصاویر کاراون‌سراها و الگوریتم YOLOv5 مکان‌یابی کاراونسارها در تصاویر ماهواره‌ای آموزش داده شده‌اند.

در ادامه‌ای این مقاله ابتدا بررسی پیشینه پژوهش مرتبط در بخش ۲ آورده شده است، سپس مراحل ایجاد پایگاه‌های داده آورده شده است. بخش ۳ به شرح انواع کاران‌سرا پرداخته و در بخش ۴ شبکه‌های عصبی عمیق مورد استفاده در این تحقیق و معیارهای دقت معرفی شده‌اند. در بخش ۶ مراحل ایجاد پایگاه‌داده‌های مورد نیاز آورده شده است و در نهایت بخش ۷ این مقاله به تشریح نتایج حاصل از آزمایش‌های انجام شده اختصاص داده شده است.

**۲- پیشینه پژوهش:**

بررسی پیشینه پژوهش مرتبط با موضوع نشان می‌دهد که علی‌رغم انجام مطالعات پژوهشی گسترده در ادبیات بین‌المللی بر موضوع بهره‌گیری از تکنیک یادگیری ماشینی و یادگیری عمیق در مستندنگاری و حفاظت از میراث فرهنگی، این حوزه در مطالعات داخلی کشور تاکنون مورد بررسی و پژوهش قرار نگرفته است.

در میان پژوهش‌های بین المللی انجام شده بلهی و همکارن (۲۰۱۹) به موضوع کاربرد علوم هوش مصنوعی در حفاظت از میراث فرهنگی پرداخته‌اند. در این پژوهش آن‌ها برآن بوده‌اند تا به کمک پروژه موردی CEPROQHA کاربرد روش‌ها و حوزه‌های مختلف یادگیری عمیق را در حفاظت از میراث مورد واکاوی قرار دهند. در انتهای آن پژوهش اثبات شده است که مجموعه رویکردهای یادگیری عمیقی که توسط آن‌ها توسعه داده شده ‌اند قابلیت کاربرد در طبقه‌بندی و جهت‌دهی به داده‌های فرهنگی را دارند(Belhi et al, 2019). در پژوهشی خوزه لاماس و همکاران (۲۰۲۰) به طبقه‌بندی و دسته‌بندی سبک‌های مختلف ابنیه معماری شناخته شده به‌عنوان میراث معماری با بهره‌گیری از تکنیک‌های یادگیری عمیق پرداخته‌اند. در این پژوهش، نویسندگان دو هدف انتشار پایگاه داده برای دسترسی عموم و بررسی امکان استفاده از شبکه‌های در مستندسازی میراث معماری دارای ارزش را دنبال کرده اند. پایگاه داده مورد استفاده در این پژوهش شامل پایگاه داده اجزای میراث معماری[[5]](#footnote-5) است که برای آموزش و اعتبارسنجی مورد استفاده قرار گرفته‌است. این پژوهش به میانگین دقت ۹۴٪ دست یافته است (Llamas et al, 2020). در پژوهشی دیگر کوندرلی و همکاران (۲۰۲۰) به تشخیص میراث گمشده در ویدیو‌های تاریخی با کمک شبکه‌های عصبی پرداخته‌اند. این پژوهشگران معتقدند که یکی از روش‌های بازگردانی اطلاعات در خصوص میراث مفقود، رجوع به تصاویر تاریخی آرشیو‌ها است. سلسله مراتب پژوهشی در این تحقیق بدین صورت است که پس از شناسایی میراث مفقود به کمک تکنیک یادگیری عمیق، بازسازی سه‌بعدی مجازی از آن‌ها با روش فتوگرامتری انجام شده‌است. این روش برروی دو نمونه موردی کلیسای تور سنت جکویس[[6]](#footnote-6) و ساختمان لس هالِس[[7]](#footnote-7) و بر مبنای وضعیت معماری این دو بنا پیش از تخریب انجام شده است. مهم‌ترین مشکل در این مطالعه کیفیت پایین تصاویر تاریخی موجود از این دو بنا و تعداد اندک تصاویر است که عملکرد شبکه عصبی را محدود ساخته‌است. یافته‌های این تحقیق نشان از اهمیت بالای نواوری‌های موجود در زمینه مستندنگاری یادمان‌های تخریب‌شده دارد و شیوه‌های جدیدی را برای بازگردانی اطلاعات گذشته در اختیار قرار می‌دهد (Condorelli et al, 2020). اسس و همکاران (۲۰۱۴) به مطالعه تشخیص و طبقه‌بندی آسیب‌های روی مصالح بنایی دارای ارزش میراث فرهنگی به کمک داده‌های تصویری پرداخته‌اند. هدف از این مطالعه تعیین و تدقیق نحوه استفاده از فناوری‌های هوش مصنوعی و یادگیری ماشینی در تصمیم‌گیری بر سر میزان عملیات حفاظتی میراث معماری است. در این پژوهش پایگاه‌داده‌ای تصویری برای تشخیص آسیب‌های مصالح سازه‌ای (سنگ) توسعه داده‌شده است. بخش اعظمی از این پژوهش با مشاهدات بصری از طریق تصاویر انجام شده‌است. نتیجه بدست آمده در قالب چارچوبی برای ارزیابی و تشخیص نوع آسیب مصالح سازه‌ای پیشنهاد شده‌است که به طبقه‌بندی نوع آسیب‌های موجود می‌پردازد. اهمیت این پژوهش در آن است که نشان می‌دهد می‌توان از یادگیری عمیق و یادگیری ماشینی برای آسیب‌شناسی پایش میراث فرهنگی استفاده نمود (Oses et al, 2014).

دو مطالعه زیر تاثیر شایانی در انتخاب روش و پیش‌برد پژوهش حاضر داشته‌اند. در پژوهش نخست، گرونر و همکاران (۲۰۲۰)، به مقایسه مدل‌های مورد استفاده در یادگیری عمیق برای شناسایی اشیا به کمک تصاویر ماهواره‌ای پرداخته‌اند. از یافته‌های آن تحقیق می توان به اشثبات امکان استفاده از مدل پنجره‌ لغزان[[8]](#footnote-8) در شناسایی اشیای مورد نظر در تصاویر ماهواره‌ای نام برد. (Groener et al, 2020) در تکنیک پنجره‌ی لغزان ناحیه‌ی مورد مطالعه به تصاویر کوچک‌تر تقسیم شده و سپس این تصاویر به عنوان ورودی به یک الگوریتم طبقه‌بندی داده می‌شوند. پس می‌توان نتیجه گرفت که اگر امکان طبقه‌بندی تصاویر با دقت قابل قبول وجود داشته باشد بدون سربار اضافی و تنها با استفاده از تکنیک پنجره لغزان می‌توان به مکان‌یابی اشیا در تصاویر ماهواره‌ی دست یافت. در پژوهش دیگر بالانویک و همکاران (۲۰۲۰) موقعیت سد‌های احداث شده یا بقایای سدهای موجود را با بهره‌گیری از تکنیک یادگیری عمیق توسط تصاویر ماهواره‌ای شناسایی می‌کند. اهمیت این پژوهش در آن است که پایگاه داده‌های پژوهش حاضر نیز همانند پژوهش بالانویک و همکاران مجموعه‌ای متشکل از تصاویر ماهواره‌ای سه کاناله RGB است که به کمک مدل شبکه‌های مشابه مورد استفاده قرار گرفته است. یافته‌های پژوهش بالانویک نشان می‌دهد که استفاده از تکنیک یادگیری عمیق هزینه‌های شناسایی و موقعیت‌سنجی عوارض زمینی مشابه را می‌توان به حداقل رساند و می‌توان از این تکنیک‌ها برای شناسایی سدها و کاروانسراها به ‌عنوان عوارض ساخته شود توسط انسان بهره‌ گرفت (Balaniuk et al, 2020).

**۳- کاروانسرا و گونه‌های آن:**

از دیرباز آدمی همواره در پی ایجاد ارتباط با هم نوع خویش بوده و این ارتباط به شیوه‌ها و با اهداف مختلفی صورت پذیرفته‌است. این ارتباط تنها محدود به انسان‌های یک جامعه نشده، بلکه از مرزهای جغرافیایی فراتر رفته‌است که نشانگر نیاز انسان به تعامل با سایر هم‌نوعان خود در سراسر کره خاکی است. در گذشته‌های دور و قبل از اختراع هرگونه وسیله ارتباطی، پیام‌ها از نقطه‌ای به نقطة دیگر در یک سرزمین، از طریق طی مسافت توسط انسان‌های سوار بر چهارپایان صورت می‌گرفت. از آنجایی که مسافت بین مبدأ و مقصد اغلب طولانی بود و پیام‌رسان و چهارپایش نیاز به استراحت داشتند، بناهایی در میان مسیر‌ها ساخته می‌شد تا مأمنی برای نامه‌رسان باشد. علاوه‌بر این، همواره انسان‌هایی بودند که بنا به هر دلیلی قصد سفر از دیار خود داشتند. به دلیل امکانات محدود و عدمِ وجودِ امنیتِ کافی در مسیر، مسافران با مقصد مشترک گرد هم می‌آمدند تا به صورت گروهی سفر کنند. کاروان نامی بود که به این گروه از مسافران اطلاق می‌شد. جهت تأمینِ آسایش و رفاهِ کاروانیان همانند نامه‌رسان‌ها، بناهایی در میان راه‌ها و در فواصل مشخص از هم که اغلب طی آن به مدتِ نصفِ روز به طول می‌انجامید، احداث می‌شدند که کاروانسرا نام داشتند. کاروانسراها محلی برای استراحت و یا سپری کردن شب برای کاروانیان بود و در برخی کاروانسراها دادو ستد نیز صورت می‌گرفت (کیانی، ۱۳۹۲: ۲۶۵-۲۶۶). این بناها علاوه‌بر نقشِ خود به‌عنوان مأمن، نوعی رسانه قدرتمند بودند که انسان‌هایی از جوامع و فرهنگ‌های مختلف را با یکدیگر آشنا و مرتبط می‌ساختند. گاهی این رسانه‌ها پیامِ اعتبار و قدرتِ بانیان خود را به مسافران، به ویژه مسافران از کشورهای دیگر، انتقال می‌دادند. بدین‌جهت ساختار کاروانسراها بسته به شرایط و کاربرد آن‌ها از تنوع چشمگیری برخوردار بوده‌است.

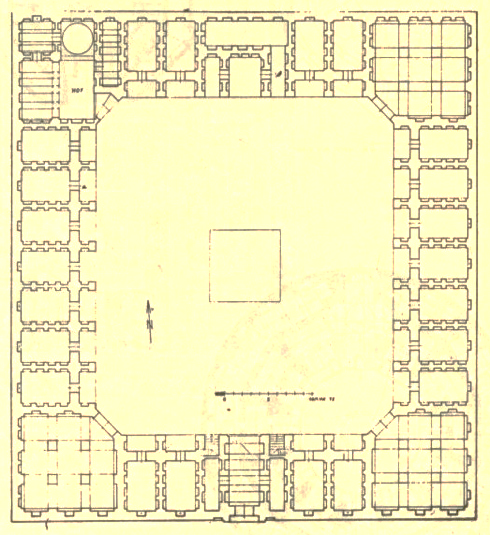
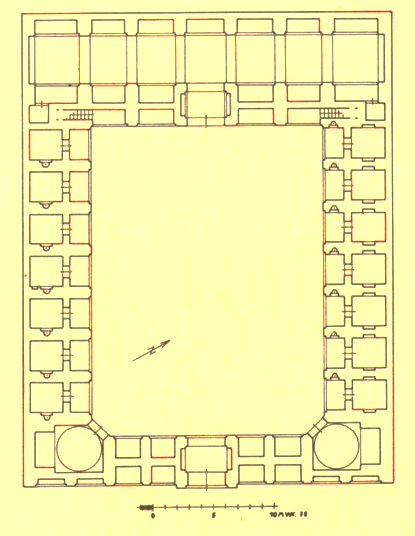
یکی از مهم‌ترین دلایل احداث و حضور کاروانسراها در سرتاسر پهنه ایران، قرارگیری کشور در میانه راه جاده‌های تجاری مهم نظیر جاده ابریشم و ادویه است که در طول تاریخ نقش مهمی را در مبادلات تجاری ایفا کرده و در تبادل اندیشه‌ها، فنون و باور‌های دینی بسیار تأثیرگذار بوده‌اند. به‌عنوان نمونه، مسیر جاده ابریشم از کشور چین شروع شده، از فلات ایران می‌گذشت و پس از گذر از بین النهرین و شامات، به انطاکیه در کنار دریای مغرب (مدیترانه) می‌پیوست. در ایران، جادة ابریشم از بلخ به طرف مرو و ناحیه جنوب دریای خزر می‌آمد، و از شهرهایی نزدیک دامغان امروزی عبور کرده و به ری می‌رسید و در ادامه به سمت همدان با گذر از زاگرس به سمت بین‌النهرین و انطاکیه کشیده شده‌بود. شاخه‌ای از جادة ابریشم در غرب ایران به سمت جنوب و سرزمین‌های پست بین النهرین از جمله بغداد امتداد یافته و شاخه دیگر در قزوين به شعباتی تقسیم می‌شد که بخشی از آن وارد زنجان شده و از شهر های میانه، بستان آباد، تبریز، مرند عبور کرده و از شهر ماکو در شمال غرب ایران کنونی به مقصد اروپا خارج می‌شد و سپس به آذربايجان و آران و ياطرابوزان مي‌پيوست (ابن حوقل : 91؛ دورانت، ويل و اريل : 1366 ،267 ). گذر این راه از استان آذربایجان و اتصال آن با کشور های همجوار و در نتیجه مبادلات فراوان و عبور کاروان ها از این مسیر، منجر به احداث کاروانسرا های متعدد و متنوع در این ناحیه شده است.

پژوهش‌های انجام‌شده نشان می‌دهد که عواملی نظیر اقلیمِ منطقة ساختِ بنا، نیازهای زمانة ساخت، اقتصادِ دورة ساخت و ویژگی‌های بسترِ ساخت در ساختار کاروانسراها و در نتیجه طبقه‌بندی معماری این نوع از ابنیه تاثیر گذاشته‌است. در این زمینه، سیرو در مطالعة کاروانسراهای میان‌راهی و درون شهری ایران براساس عاملیت اقلیم، این نوع از ابنیه و ساختارهای تاریخی را به زیر مجموعه‌های کاروانسراهای کاملا پوشیدة منطقة کوهستانی، کاروانسراهای کرانه خلیج فارس، کاروانسراهای سواحل جنوبی دریای خزر، کاروانسراهای حیاط‌دار مناطق مرکزی ایران و کاروانسراهای چند ضلعی دسته بندی می‌کند (کیانی، ۱۳۸۹ به نقل از سیرو، ۱۳۳۹). لازم به‌ذکر است که یکی از تفاوت‌های بنیانی میان کاروانسراهای درون شهری و برون شهری مطابق با دسته‌بندی انجام شده نوع و نحوه عملکرد هرکدام از این کاروانسراهاست. به‌طور کلی کاروانسراهای درون‌شهری علاوه بر ماهیت اقامتی، مرکزی برای خرید و فروش بارهای کاروان‌ها و محلی برای داد ‌و ستد میان کاروان‌های تجاری مختلف به‌‌شمار می‌آمدند، در حالی‌که کاروانسراهای میان‌راهی عمدتا عملکردی اقامتی داشتند و گاه خرید و فروش جزئی میان مسافران یا دریافت خراج مسیر کاروان‌ها صورت می‌گرفت.

در ادامه گونه‌بندی معرفی‌شده از کاروانسراهای درون و برون شهری، پیرنیا تحول ساختاری کاروانسراها را بنا به دوره‌های تاریخی مختلف طبقه‌بندی کرده و تداوم و تکامل ساختاری آن‌ها را از گونه کاروانسراهای حیاط‌ دار به کاروانسراهای چهارایوانی که از دوران ایلخانی تا اواخر دوران قاجار رواج یافته‌اند، تقسیم‌بندی کرده است (معماریان،۱۳۷۶:۴۶۶). کیانی نیز ساختمان کاروانسراهای ایران را در پنج دستة کاروانسراهای دو ایوانی، کاروانسراهای چهار ایوانی، کاروانسراها با تالار ستون‌دار، کاروانسراهای با پلان متفرقه و کاروانسراهای چندضلعی معرفی کرده است (کیانی، ۱۳۷۹:۲۷۳). کاروانسراهای مورد هدف در پژوهش حاضر اغلب در دسته کاروانسراهای برون شهری و ایوان‌دار (دو ایوانی و چهار ایوانی) قرار دارند.

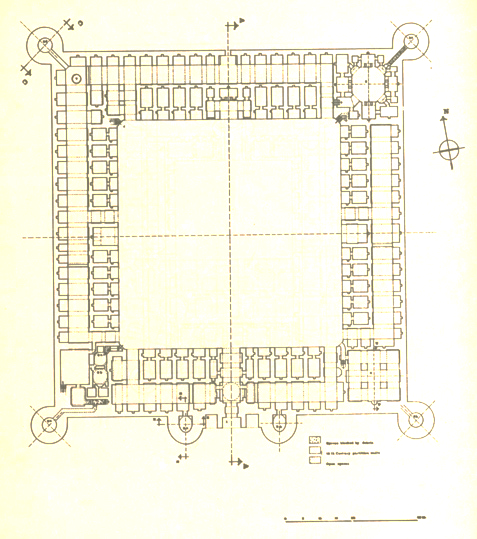
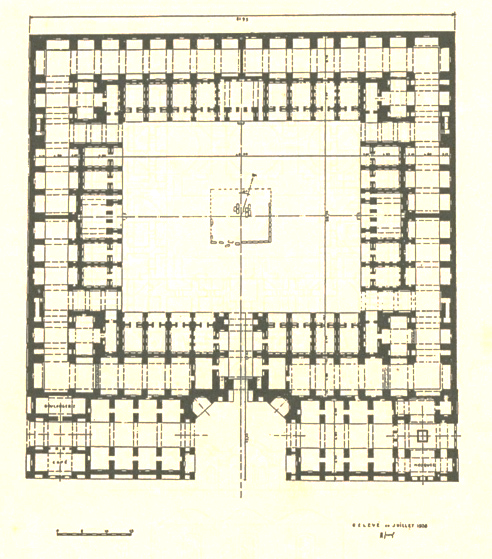
**کاروانسراهای ایوان‌دار** پیشینه ساخت کاروانسراهای ایوان‌دار، شامل کاروانسراهای دو ایوانی، سه‌ایوانی یا چهار ایوانی به دوران ساسانی باز می‌گردد. در این دوره به‌علت توسعه اقتصادی سرزمین ایران ایجاد راه‌ها و ساختمان‌های وابسته به آن اهمیتی دوچندان یافت و در نتیجه کاروانسراهای بسیاری در طی مسیر جاده‌ها و راه‌ها ساخته شدند که نقشه عمومی این بناها به‌صورت چهار ایوانی است. پس از این دوران، استفاده از طرح چهار ایوانی و دو ایوانی بیشتر رواج یافت. به‌طوری که در طی دوران سلجوقی استفاده از این طرح‌مایه در ساخت اغلب ابنیه تاریخی به‌طور ویژه کاروانسراها به طرح غالب تبدیل شد. تا شروع دوره صفویه، اغلب کاروانسراهای میان راهی در پهنه ایران به‌این صورت ساخته شدند.

**کاروانسراهای دو ایوانی** همانند مدارس و مساجد، اغلب به فرم مربع یا مستطیل ساخته‌شده‌اند. عموما ایوان‌های این کاروانسراها یکی در مدخل ورودی و دیگری روبه‌روی آن قرار دارد. فضاهای دیگر نیز دورتادور میان‌سرای مربع شکل مرکزی جای گرفته‌اند. نمونه این طرح‌مایه، کاروانسراهای چاه خوشاب و کاروانسرای دوکوهک است (کیانی، ۱۳۹۲).



تصویر ۳-۱: (سمت راست)کاروانسرای چاه‌خوشاب، نزدیک فردوس منبع: کیانی و کلاویس، ۱۳۶۴:۶۱۲ و (سمت چپ) کاروانسرای دو دهک، منبع: کیانی و کلاویس، ۱۳۶۴:۵۳

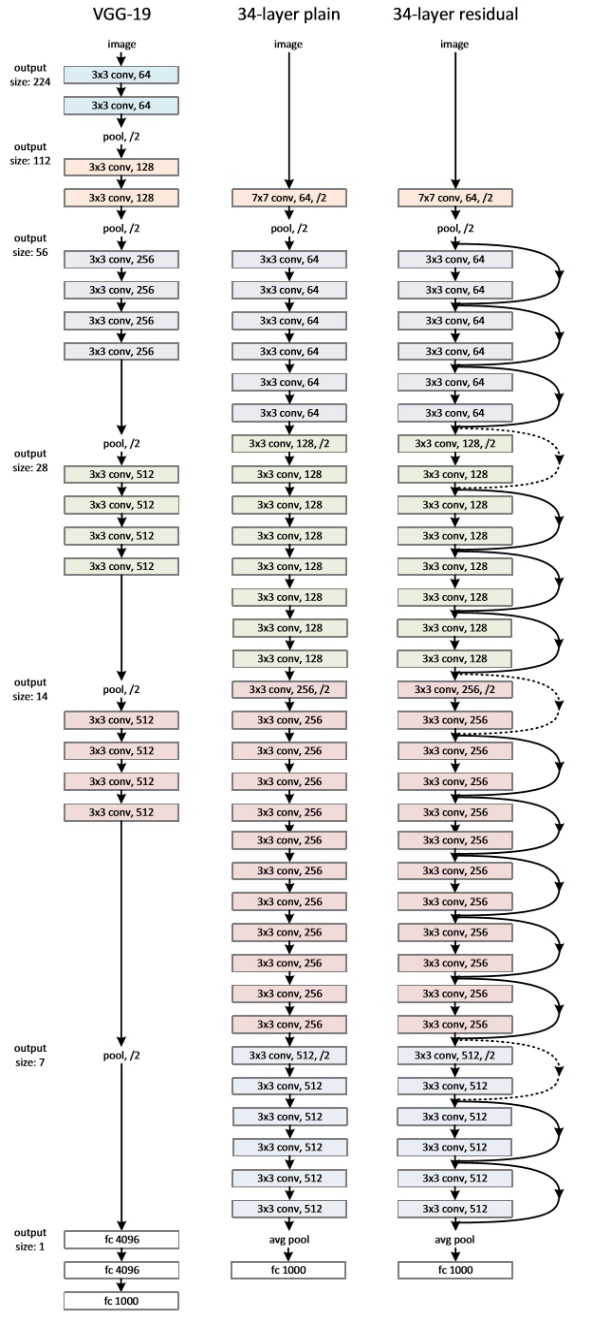
**ساخت کاروانسراهای چهار ایوانی** با جیاط مرکزی باز سابقه‌ای طولانی در تاریخ معماری ایران دارد. اغلب این کاروانسراها با طرح ثابت مربع یا مستطیل و با ایوان‌هایی در مرکز هر جبهه ساخته‌شده‌اند؛ لیکن در جزئیات شکل داخل و خارج، دروازه ورودی، برج‌ها و سایر فضاها دارای ویژگی‌های متفاوتی هستند. نمونه این نوع کاروانسراها، کاروانسرای دیرگچین و کاروانسرای مهیار است (همان).

****

تصویر ۳-۲: (سمت راست) کاروانسرای دیرگچین، نزدیک شهر ری، منبع: کیانی و کلاویس، ۱۳۶۴:۶۴۹ و (سمت چپ) کاروانسرای مهیار، منبع: کیانی و کلاویس، ۱۳۶۴:۵۴۱

**۴- معرفی معماری شبکه‌های عصبی عمیق استفاده شده و معیارهای خطا:**

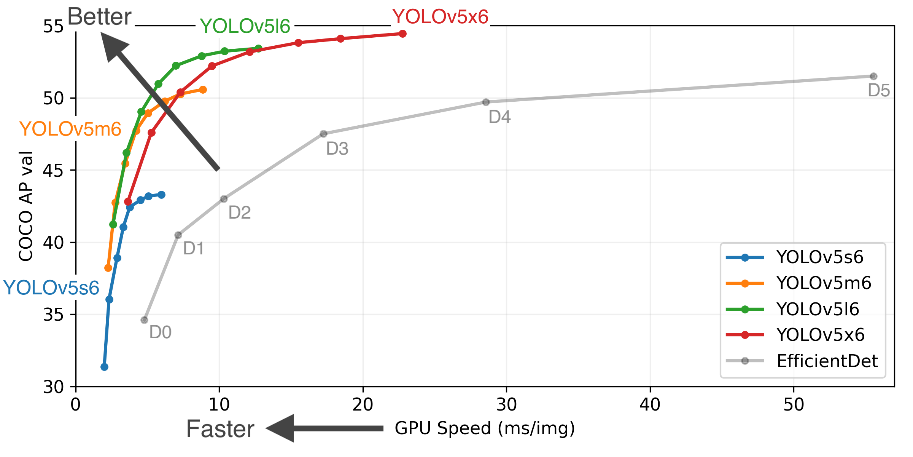
مجموعه‌ی شبکه‌های عصبی خانواده‌ی ResNet توسط He و همکاران (۲۰۱۵) و به منظور طبقه‌بندی تصاویر مسابقه‌ی ImageNet طراحی شده است. این خانواده از معماری‌های شبکه‌ی عصبی عمیق به دلیل پیاده سازی ساده و قدرت بسیار بالا در طبقه‌بندی تصاویر به یکی از رایج‌ترین روش‌های مورد استفاده تبدیل شده است. ویژگی اصلی این شبکه استفاده از پرش از روی اتصالات[[9]](#footnote-9) در بین لایه‌های مختلف است که این تکنیک مشکل ناپدید شدن گرادیان[[10]](#footnote-10) را به صورت قابل توجهی کاهش می‌دهد و باعث می‌شود تا بتوان شبکه‌های عصبی عمیق با تعداد لایه‌ی زیاد را آموزش داد. معماری ResNet در تصویر ۴-۱ نشان داده شده است.



تصویر ۴-۱: معماری شبکه‌ی ResNet

در این مقاله به علت تعداد نسبتا محدود داده‌های آموزشی از شبکه‌های ResNet18، ResNet34 و ResNet50 استفاده شده است و معماری‌های با عمق بیشتر شبکه‌ی ResNet مورد مطالعه قرار نگرفته‌اند. لازم به ذکر است که اعداد ۱۸، ۳۴ و ۵۰ آمده در نام شبکه‌های ذکر شده در بالا بیانگر تعداد لایه‌های هر شبکه می‌باشد.

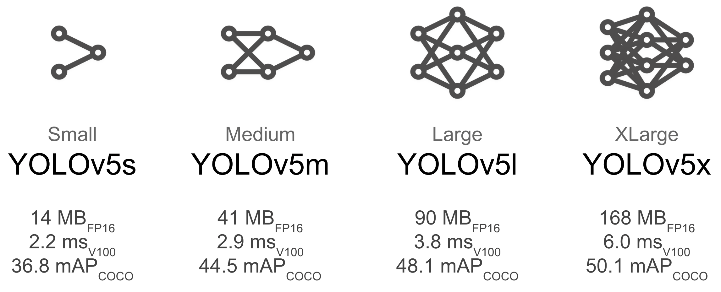
الگوریتم YOLOv5 یکی از جدیدترین شبکه‌های عصبی موجود برای تشخیص شی است. YOLOv5 دارای دقت بالا و بسیار سریع است. این معماری به قدرت پردازشی کمتری به نسبت سایر روش‌های تشخیص شی مبتنی بر شبکه عصبی نیاز داشته و در عین حال از نظر دقت با آن‌ها برابری می‌کند [۱۴و ۱۵]. نمودار ۴-۱ سرعت و دقت YOLOv5 را در مقایسه با روش‌ EfficientDet که یکی دیگر از معروف‌ترین روش‌های تشخیص شی است نشان می‌دهد.



نمودار۴-۱: مقایسه دقت معماری YOLOv5 با EfficientDet [16]

علت انتخاب شبکه‌ی YOLOv5 در این مقاله سرعت بالای این شبکه است. با توجه به مساحت بسیار زیاد کشور استفاده از الگوریتم‌های کندتر (هرچند با دقت بالاتر) به علت زمانبر و هزینه‌بر بودن قابلیت استفاده عملیاتی نخواهند داشت. در صورتی که یک الگوریتم تشخیص شی سریع (همانند الگوریتم YOLOv5) پس از اتمام آموزش بر روی پایگاه داده‌ی آزمایشگاهی می‌تواند در زمان قابل قبولی به پایش تصاویر ماهواره‌ای از کل سرزمین ایران پرداخته و کاروانسراهای ایوان‌دار را با هزینه‌ی پایین در پهنه‌ی کشور شناسایی کند.

الگرویتم YOLOv5 در چهار وزن مختلف منتشر شده است. در این مقاله به علت تعداد نسبتا کم داده‌های آموزشی و جهت جلوگیری از Under fitting از وزن‌های سبک YOLOv5s استفاده شده است. پرواضح است که معماری YOLOv5s به علت پارامترهای کمتر نیاز به دیتای کمتری برای آموزش دارد.



تصویر ۴-۲: وزن‌های مختلف معماری YOLOv5

معماری YOLOv5 از بسیاری از بخش‌های معماری YOLOv4 [۱۸] استفاده می‌کند. انکدر[[11]](#footnote-11) استفاده شده در YOLOv5 همان CSPDarknet [۱۸] می‌باشد، انکدر ذکر شده در کنار Path Aggregation Network (PANet) [17] تمامی بخش‌های شبکه‌ی YOLOv5 را شامل. در مقایسه با YOLOv4 در YOLOv5 توابع فعال ساز تغییریافته اند (توابع فعال ساز Leaky ReLU و Hardswish با SiLU [19] جایگزین شده‌اند.

تابع خطا[[12]](#footnote-12) معماری YOLOv5 از جمع سه تابع خطای زیر بدست می‌آید:

● Bounding box regression Loss - penalty for wrong anchor box detection, Mean Squared  
Error calculated based on predicted box location (x, y, h, w),  
● Classification Loss - Cross Entropy calculated for object classification,  
● Objectness Loss - Mean Squared Error calculated for Objectness-Confidence Score  
(estimation if the anchor box contains an object).

معیار precision تقسیم تعداد مواردی که توسط مدل درست تشخیص داده شده‌است بر تعداد مواردی که واقعاً درست هستند.

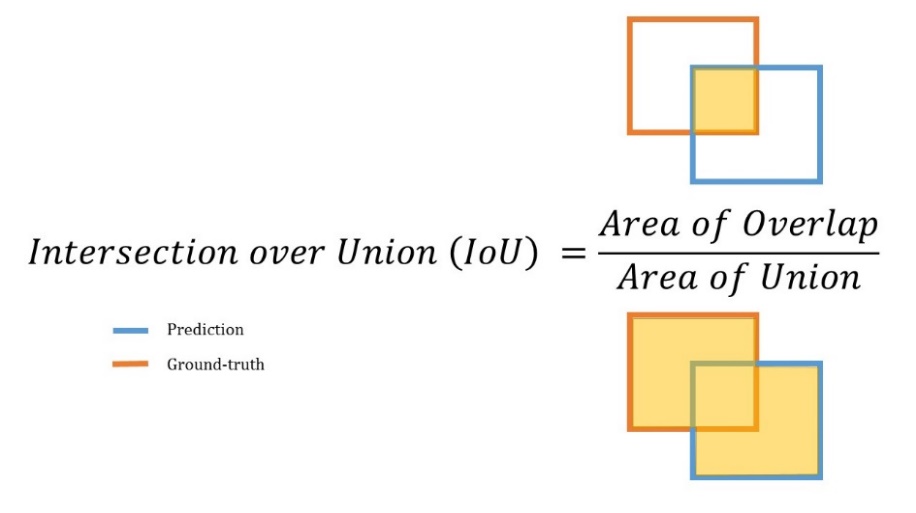
فرمول ۴-۱:

در الگوریتم‌های طبقه‌بندی دقت[[13]](#footnote-13) برابر است با نسبت نمونه‌هایی که برچسب درست خورده‌اند به کل نمونه‌ها (داده‌های آزمایش) که با استفاده از رابطه محاسبه می‌شود:

فرمول ۴-۲:

معیار IoU[[14]](#footnote-14) به منظور اندازه‌گیری کیفیت خروجی الگوریتم‌های تشخیص شی استفاده می‌شود. این معیار مقداری بین ۰ تا ۱ دارد که مقدار ۱ نشانگرد همپوشانبی بیشتر بین خروجی الگوریتم و مقادیر مرجع و در نتیجه بهتر بودن خروجی الگوریتم می‌باشد.

فرمول ۴-۳:



معیار mAP[[15]](#footnote-15) مقادیر ground-truth bounding box را با bounding boxهای تشخیص داده شده توسط الگوریتم مقایسه می‌کند و یک امتیاز[[16]](#footnote-16) را بر می‌گرداند. هرچه مقدار این امتیاز بیشتر باشد بیانگر عمکرد بهتر الگوریتم می‌‌باشد.

برای محاسبه‌ی mAP ابتدا باید average precision (AP) محاسبه گردد. AP راهی‌است برای خلاصه سازی نمودار precision-recall به یک مقدار عددی که بیانگر متوسط همه‌ی precisionها است. فرمول محاسبه‌ی mAP در زیر آمده است.

فرمول ۴-۴:

در کاربراهای تشخیص شی یک مقدار آستانه برای محاسبه‌ی مثبت یا منفی بودن خروجی الگوریتم انتخاب می‌گردد و اگر IoU خروجی الگوریتم از این مقدار بیشتر باشد آن خروجی به عنوان نمونه مثبت و اگر کمتر باشد به عنوان نمونه منفی طبقه بندی می‌گردد.

فرمول ۴-۵:

## برای محاسبه‌ی mAP ابتدا مقدار AP برای هی یک از کلاس‌‌های داده محاسبه می‌شود و در نهایت میانگین این مقادیر به همراه آستانه‌ی انتخاب شده به عنوان mAP گزارش می‌گردد. به شکل معمول مدل‌های تشخیص شی با چند مقدارآستانه‌ متفاوت ارزیابی می‌شوند، به عنوان مثال mAP@.5 به معنی متوسط AP کلاس‌های مختلف برای IoU با آستانه‌ی ۰.۵ می‌باشد.

## فرمول ۴-۶:

**۵-آماده سازی پایگاه داده:**

**۱-۵ ایجاد پایگاه‌ها داده:**

جهت ایجاد پایگاه داده تصاویر ماهواره‌ای از کاروانسراها ابتدا نیاز بود تا موقعیت جغرافیایی کاروانسراها استخراج شود. با توجه به اینکه سابق بر این هیچ داده‌ی دیجیتال و دقیقی از مختصات جغرافیایی کاراوانسراها ایجاد نشده است، ابتدا لیستی از کاروانسراها از کتاب فهرست کاروانسراهای ایران و جست و جوی عبارت «کاروانسرا» در نقشه‌ی گوگل ایجاد شد. سپس تمامی مختصات دقیق این کاروان‌سراها به صورت دستی با کمک سرویس نقشه‌ی گوگل[[17]](#footnote-17) و سرویس نقشه‌ی بینگ[[18]](#footnote-18) بدست آمده و پس از راستی آزمایی و حذف موارد تکرای فهرست نهایی موقعیت جغرافیایی کاراونسراهای ایوان‌دار ایران ایجاد شد. نمونه‌ای از این داده‌ها را می توان در جدول ۵-۱ مشاهده نمود.

|  |  |  |
| --- | --- | --- |
| موقعیت جغرافیایی | نام کاروانسرا | شماره |
| 38.344328, 45.834966 | کاروانسرای یام | 001 |
| 38.977452, 45.577038 | کاروانسرای خواجه نظر | 002 |
| 37.271923, 47.843075 | کاروانسرای جمال آباد | 003 |
| 36.266490, 60.655253 | کاروانسرای رباط شرق | 004 |
| … | … | … |
| 33.281254, 58.910111 | کاروانسرای رباط ده چاهک | 160 |

جدول ۵-۱: نمونه‌ی دیتای موقعیت جغیرافیایی کاروانسراهای ایوان‌دار ایران

برای تهیه‌ی تصاویر ماهواره‌ای موقعیت‌های جغرافیایی استخراج شده در مرحله‌ی قبل به عنوان مرکز تصویر انتخاب شده و با استفاده از سرویس‌های نقشه‌ی گوگل و بینگ و به صورت نیمه اتوماتیک یک تصویر از هر کاروانسرا در سطح نزدیکی به زمین[[19]](#footnote-19) ۱۰۰متر تهیه شده است. همچنین از محیط اطراف هر کاروانسرا ۲ تصویر و از ۲۰ مکان تاریخی (مانند حافظیه شیراز، میدان آزادی و ...) در همان سطح نزدیکی به زمین تصاویر نمونه منفی[[20]](#footnote-20) تهیه شده‌اند. تصاویر نمونه‌ی منفی بگونه‌ای انتخاب شده‌اند که بافتی مشابه با کاروانسراها را داشته باشند تا شبکه‌ی عصبی عمیق آموزش داده شده در مراحل بعد دارای تعمیم‌پذیری بالایی باشد.

تصاویر تهیه شده در مرحله‌ی قبل به ابعاد ۶۰۰×۶۰۰ پیکسل کاهش اندازه داده شده و دیتاست طبقه‌بندی با نسبت ۲۰٪ برای اعتبارسنجی[[21]](#footnote-21) و ۸۰٪ دیتای آموزش[[22]](#footnote-22) ایجاد شده است. نمونه‌ای از تصاویر ایجاد شده در این مرحله در تصویر ۵-۱ آمده است.



تصویر ۵-۱: نمونه تصاویر ایجاد شده در پایگاه داده طبقه‌بندی

لازم به ذکر است که با توجه به اینکه از هر کاروانسرا ۲ تصویر (یک تصویر از سرویس نقشه گوگل و تصویر دیگر از سرویس نقشه بینگ) در پایگاه داده وجود دارد، زوج تصویر کاروانسرا تنها در یک گروه از تصاویر اعتبارسنجی یا آموزش قرار داده شده‌اند.

در جدول ۵-۲ تعداد دقیق تصاویر پایگاه داده‌ی طبقه‌بندی آورده شده است.

|  |  |  |
| --- | --- | --- |
| تعداد داده راستی آزمایی | تعداد داده آموزش | شی |
| 61 | 228 | کاروانسرا | |
| 140 | 605 | سایر | |
| 636 | 319 | کل | |

جدول ۵-۲: توزیع تعداد تصاویر پایگاه داده طبقه‌بندی

برای ایجاد پایگاه داده تشخیص شی با استفاده از ابزار LabelImg[[23]](#footnote-23) اطراف هر کاروانسرا یک جعبه[[24]](#footnote-24) ترسیم شده است. همچنین تصاویری که در آن‌ها هیچ کاروانسرایی وجود نداشته‌اند به عنوان نمونه منفی در پایگاه داده تشخیص شی قرار داده شده است.



تصویر ۵-۲: نمونه تصاویر پایگاه داده تشخیص شی

لازم به ذکر است که خروجی نرم‌افزار LabelImg با فرمت داده‌های yolo و در پوشه‌ای با نام labels در تصاویر پایگاه داده ایجاد شده جهت تشخیص شی ذخیره شده‌اند.

**۲-۵ پیش پردازش داده‌ها:**

تمامی تصاویر مورد استفاده در الگوریتم طبقه‌بندی در اندازه‌ی ۱۲۸×۱۲۸ پیکسل کوچک شده‌اند و جهت داده افزایی عملیات‌های زیر بر روی آن‌ها انجام شده است:

* چرخص تصادفی تصویر تا حداکثر ۷۵ درجه
* تغییر شد روشنایی تصویر
* تغییر در کشیدی تصویر
* سر وته عمودی تصویر

نمونه‌ای داده افزایی‌های انجام شده بر روی یک عکس در تصویر ۵-۳ آمده‌است.

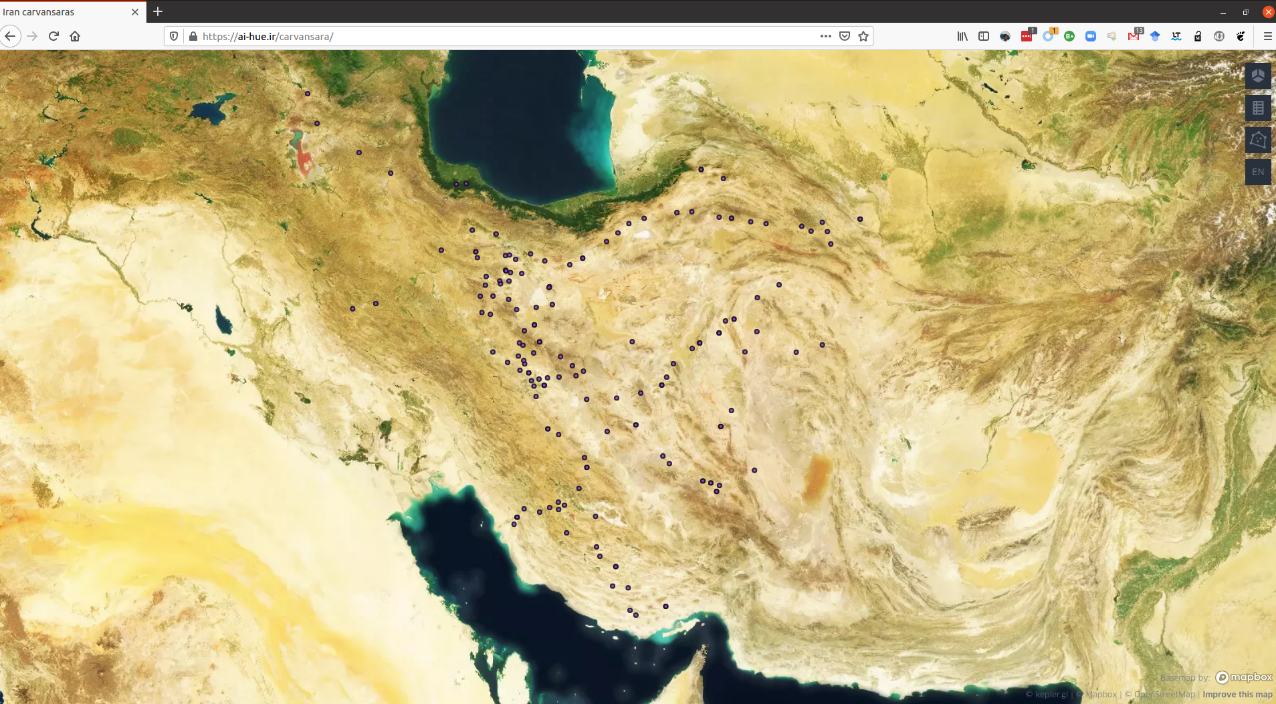
****

تصویر ۵-۳ نمونه‌ای از تغییرات ایجاد شده بر روی تصویر یک کاروانسرا جهت داده افزایی

هنگام آموزش الگوریتم YOLOv5 از داده‌افزایی پیشفرض مقاله‌ی اصلی استفاده شده است.

**۶- تحلیل یافته‌ها**

به منظور استفاده‌ی بهینه از پایگاه داده‌ی موقعیت‌هایی جغرافیایی ایجاد شده در این مقاله یک نقشه‌ی تعاملی[[25]](#footnote-25) با استفاده از کتابخانه‌ی keplergl ایجاد شده است. این نقشه در آدرس <https://ai-hue.ir/caravansary/map.html> در دسترس است.



در تصویر ۶-۱: نقشه‌ی پراکندگی موقعیت جغرافیایی کاروانسراهای ایوان‌دار ایران

همانطوری که در تصویر بالا مشخص است اکثر کاراوانسرهای ایوان‌دار باقی مانده در ایران در مناطق با میزان بارش باران نسبتا کم قرار گرفته‌اند. مسیر این کارانسراها از دو بندر مهم جنوب کشور، بندر عباس و بندر بوشهر، شروع شده و به سمت مرکز و شمال شرق کشور ادامه پیدا می‌کند.

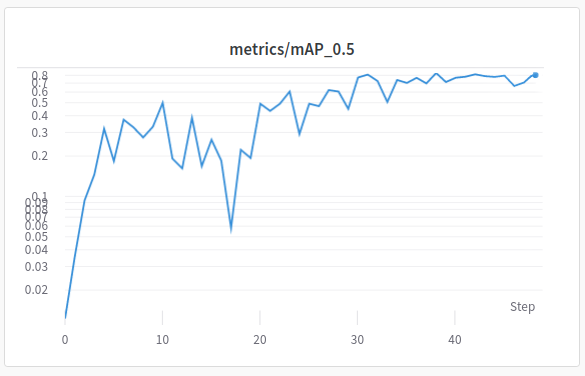
شبکه‌های عصبی ResNet18، ResNet34 و ResNet50 با استفاده از کتابخانه‌ی FastAI و با استفاده از تکنیک یادگیری انتقالی بر روی پایگاه‌ داده‌ی طبقه‌بندی ایجاد شده آموزش داده شده‌اند. نتایج حاصل از این آزمایش در جدول ۶-۱ آمده است.

|  |  |
| --- | --- |
| دقت | معماری |
| 79.2% | ResNet18 |
| 85.83% | ResNet34 |
| 82.9% | ResNet50 |

جدول ۶-۱ نتایج طبقه بندی بر روی پایگاه داده کاروانسراها

همانطور که در جدول ۶-۱ مشخص است بهترین دقت برای شبکه‌ی ResNet34 بدست آمده‌ است. علت این مساله‌ را می توان کوچک‌ بودن پایگاه داده‌ی مورد مطالعه در این مقاله دانست؛ به گونه‌‌ای که تعداد داده‌ی موجود برای آموزش وزن‌های شبکه‌ی ResNet50 کافی نبوده و Under fitting در هنگام آموزش این شبکه رخ داده است، همچنین شبکه‌ی ResNet18‌ ظرفیت کافی را برای طبقه‌بندی بهینه ی این پایگاه داده نداشته است.

الگوریتم YOLOv5 با استفاده از فضای ابری Google Colab بر روی پایگاه داده‌ی شناسایی شی ایجاد شده در این مقاله آموزش داده شده است. نمودار تغییرات metrics/mAP\_0.5 در گام‌های مختلف آموزش معماری YOLOv5 در نمودار ۶-۱ آمده است.

****

نمودار ۶-۱: نمودار تغییرات metrics/mAP\_0.5

همانطور که در تصویر بالا مشخص است دقت نهایی این معماری در یافتن کاروانسراها 82.47% mAP\_0.5 است که عدد بسیار خوب و قابل قبولی می‌باشد. نمونه‌ای از خروجی الگوریتم YOLOv5 بر روی تصاویر راستی آزمایی پایگاه داده در تصویر ۶-۲ نشان داده شده است.

****

تصویر ۶-۲ خروجی الگوریتم YOLOv5 بر روی تصاویر راستی آزمایی

**۷- نتیجه‌گیری**

دقت ۸۵.۸۳٪ بدست آمده توسط شبکه‌ی عصبی ResNet50 برای طبقه‌بندی کاروانسراها از سایر بناها و دقت 82.47% mAP\_0.5 الگوریتم YOLOv5 در مکان‌یابی کاروانسرها با توجه به محدود بودن تعداد داده‌های در دسترس بسیار امیدوار کننده است. نتایج حاصل از آزمایش‌های انجام شده بیانگر این موضوع می‌باشد که می‌توان از تکنیک‌های پردازش تصویر مبتنی بر یادگیری عمیق با وجود تعداد دیتای کم در موارد مرتبط به میراث فرهنگی استفاده کرد. همچنین پایگاه داده‌ی موقعیت‌های جغرافیایی و دو پایگاه‌داده‌ی تصاویر ماهوراه‌ای کاروانسرهای ایوان‌دار که در این تحقیق ایجاد شده‌اند به شکل منبع باز[[26]](#footnote-26) در آدرس <https://ai-hue.ir/caravansary> برای دانلود در دسترس محقین قرار داده شده‌اند و این پایگاه‌های داده‌ها می‌توانند در سایر تحقیقات حوزه‌ی میراث فرهنگی مورد استفاده قرار گیرند.

**منابع:**

[1] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” in CVPR, 2016.

14. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149  
(2017).

15. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conf. Artif. Intell. AAAI 2017*4278–4284 (2017).

16. YOLOv5 official release note: https://github.com/ultralytics/yolov5,  
Accessed 11 May 2021

17. Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path Aggregation Network for Instance Segmentation. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 8759–8768 (IEEE, 2018). doi:10.1109/CVPR.2018.00913.

18. Bochkovskiy, A., Wang, C. Y. & Liao, H. Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* (2020).

19. Elfwing, S., Uchibe, E. & Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* **107**, 3–11 (2018).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

1. Deep Learning [↑](#footnote-ref-1)
2. Dataset [↑](#footnote-ref-2)
3. Transfer learning [↑](#footnote-ref-3)
4. light-weight [↑](#footnote-ref-4)
5. Architectural Heritage Elements Dataset [↑](#footnote-ref-5)
6. Tour Saint Jacques [↑](#footnote-ref-6)
7. Les Halles [↑](#footnote-ref-7)
8. Sliding Window Model [↑](#footnote-ref-8)
9. skip connection [↑](#footnote-ref-9)
10. vanishing gradient [↑](#footnote-ref-10)
11. encoder [↑](#footnote-ref-11)
12. loss function [↑](#footnote-ref-12)
13. Accuracy [↑](#footnote-ref-13)
14. Intersection over Union [↑](#footnote-ref-14)
15. mean average precision [↑](#footnote-ref-15)
16. score [↑](#footnote-ref-16)
17. https://www.google.com/maps [↑](#footnote-ref-17)
18. https://www.bing.com/maps [↑](#footnote-ref-18)
19. Zoom [↑](#footnote-ref-19)
20. Negative sample [↑](#footnote-ref-20)
21. Validation‌ [↑](#footnote-ref-21)
22. Train [↑](#footnote-ref-22)
23. https://github.com/tzutalin/labelImg [↑](#footnote-ref-23)
24. Bonding box [↑](#footnote-ref-24)
25. interactive [↑](#footnote-ref-25)
26. Open Source [↑](#footnote-ref-26)