Arch

6TH INTERNATIONAL CONFERENCE ON APPLIED RESEARCH IN COMPUTER, ELECTRICAL AND INFORMATION TECHNOLOGY
March 6, 2022
Tbilisi - Georgia
CINVU
Kosar University
Avicenna International Community College LLC

مقایسه الگوریتم‌های تشخیص اشیاء YOLO، RCNN و HOG

حسین رعیت پرور[1]، محمدعلی جوادزاده[2]

۱-دانشجوی کارشناسی ارشد دانشگاه جامع امام حسین(ع)

۲-استادیار دانشگاه جامع امام حسین(ع)

چکیده

تشخیص اشیاء یک کار بینایی ماشین مهم است که برای شناسایی نمونه‌هایی از اشیاء بصری کلاس‌های خاص (مثلاً انسان‌ها، حیوانات، اتومبیل‌ها یا ساختمان‌ها) در تصاویر دیجیتالی مانند عکس‌ها یا فریم‌های ویدیو استفاده می‌شود. هدف از تشخیص اشیاء، توسعه مدل‌های محاسباتی است که اساسی‌ترین اطلاعات مورد نیاز برنامه‌های بینایی ماشین را ارائه می‌کند و همچنین افزودن درک تصاویر همانند انسان به رایانه می‌باشد. حال چالش موجود این می‌باشد که کدام یک از الگوریتم‌های تشخیص اشیاء عملکرد قابل قبولی از نظر دقت و زمان پاسخ را دارد. پژوهشگران دیگر نیز این موضوع را برای الگوریتم‌های تشخیص اشیاء که از فناوری شبکه عصبی بهره‌مند می‌باشند مورد بحث و بررسی قرار داده‌اند و به نتایج قابل قبولی رسیده‌اند. حال ما درنظر داریم الگوریتم‌هایی که از فناوری شبکه عصبی استفاده می‌کنند را با الگوریتم‌هایی که از این فناوری استفاده نمی‌کنند را مورد بحث و بررسی قرار دهیم. در این مقاله ما به این نتیجه می‌رسیم که الگوریتم‌هایی که از فناوری شبکه عصبی و یادگیری عمیق را استفاده می‌کنند نیازمند قدرت محاسباتی بالا هستند به تبع زمان پاسخ آن‌ها پایین و از دقت بالایی در تشخیص اشیاء برخوردار هستند.

واژگان کلیدی: تشخیص اشیاء، بینایی ماشین، شبکه عصبی، یادگیری عمیق

www.SID.ir

Arch

6TH INTERNATIONAL CONFERENCE ON APPLIED RESEARCH IN COMPUTER, ELECTRICAL AND INFORMATION TECHNOLOGY
March 6, 2022
Tbilisi - Georgia
CINVU
Kosar University
Avicenna International Community College LLC

۱- مقدمه

یک انسان به راحتی می‌تواند اتفاقات مختلفی را که در اطرافش اتفاق می‌افتد، بسیار سریع تجسم، پیش‌بینی و تحلیل کند. انسان‌ها به وسیله حس بینایی و سیستم عصبی درون مغز وظیفه مشاهده، تجسم و تحلیل مشاهدات را انجام می‌دهند. به همین ترتیب، یک کامپیوتر هم می‌تواند عملکردهای مشابهی را با کمک بینایی ماشین انجام دهد. الگوریتم‌هایی وجود دارند که به رایانه در مورد نحوه پردازش تصاویر و نتیجه‌گیری‌های مختلف آموزش می‌دهند. انسان‌ها به یک تصویر نگاه می‌کنند و بلادرنگ می‌توانند اشیاء، مکان آن اشیاء و چگونگی تعامل یا برقراری ارتباط آن اشیاء درون تصویر را تشخیص و بررسی نمایند. الگوریتم‌های سریع و دقیق برای تشخیص اشیاء به رایانه‌ها اجازه می‌دهد همانند یک انسان یک تصویر را پردازش کنند.

تشخیص اشیاء وظیفه شناسایی نمونه‌هایی از اشیاء یک کلاس خاص در یک تصویر را فراهم می‌کند. روش‌های تشخیص اشیاء را می‌توان به دو دسته اصلی تقسیم کرد: رویکردهایی که از فناوری شبکه عصبی بهره می‌برند و رویکردهای غیرشبکه عصبی. رویکردهایی که از فناوری شبکه عصبی استفاده می‌کنند نیز خود به دو دسته روش‌های تک مرحله‌ای و روش‌های دو مرحله‌ای تقسیم می‌شوند. روش‌های تک مرحله‌ای سرعت پردازش را در اولویت قرار می‌دهند یعنی این الگوریتم‌ها در کمترین زمان پاسخ می‌دهند و مدل‌های نمونه شامل YOLO، SSD و RetinaNet می‌شوند. روش‌های دو مرحله‌ای دقت تشخیص را اولویت‌بندی می‌کنند و از مدل‌های نمونه این دسته می‌توان Faster RCNN، Mask RCNN وRCNN می‌توان نام برد. رویکردهایی که از فناوری شبکه عصبی استفاده نمی‌کنند در آن‌ها باید ویژگی‌ها را تعیین کرده و سپس توسط یک طبقه‌بند آن‌ها را دسته‌بندی کرد. روش‌های تعیین ویژگی مانند Scale-invariant feature transform (SIFT) و Histogram of oriented gradients (HOG) هستند. مسئله ما این می‌باشد که از بین تمامی دسته‌های نام برده شده کدام یک بهترین عملکرد را از منظر دقت و زمان پاسخ دارند. اهمیت این موضوع زمانی مطرح می‌شود که بحث سیستم‌های بلادرنگ می‌شود، ما نیاز داریم یک سیستم بلادرنگ در کمترین زمان ممکن بتواند پاسخی با دقت بالا ارائه کند. پس تشخیص اشیاء در پهبادها، ماشین‌های خودران و همچنین در دوربین‌های کنترل سرعت از اهمیت خیلی بالایی برخوردار است. همچنین نمی‌توان از کاربردهای تشخیص اشیاء در زمینه‌های تشخیص عابر پیاده، تشخیص چهره، تشخیص متن و غیره چشم‌پوشی کرد.(Z.Zou et al, 2019)

۱-۱- پیشینه تحقیق

از پیشینه تشخیص اشیاء می‌توان به دو دوره اساسی یاد برد. درسال ۲۰۱۲ با معرفی فناوری یادگیری عمیق و شبکه‌های عصبی باعث تحولی در عرصه تشخیص اشیاء شد که تشخیص اشیاء به دوره قبل ۲۰۱۴ مبتنی بر رویکردهای سنتی و بعدازظهر ۲۰۱۴ که مبتنی بر فناوری‌های نوظهور یادگیری عمیق وشبکه عصبی می‌باشد. تشخیص اشیاء مبتنی بر رویکردهای سنتی با استفاده از

www.SID.ir

تعیین ویژگی و استخراج آن از تصاویر، اشیاء را تشخیص می‌دادند؛ این نوع اصلا نیازمند محاسبات پیشرفته‌ای نبود. بعداز ظهور شبکه‌های عصبی کانولوشنال و یادگیری عمیق در سال ۲۰۱۲ تشخیص اشیاء دچار پیشرفت قابل توجهی شد. از آنجایی که یک شبکه کانولوشن عمیق قادر به یادگیری نمایش ویژگی‌های قوی و سطح بالا از یک تصویر است دقت بالاتری را از رویکردهای سنتی مبتنی بر استخراج ویژگی ارائه می‌دهد. (Z.Zou et al, 2019)

هدف از بیان این مسئله و پژوهش روی آن، این می‌باشد که ما بتوانیم الگوریتم‌های تشخیص اشیاء را با توجه به عمکردشان شناسایی کنیم و در مواقع نیاز از آن‌ها استفاده کنیم. باتوجه به اینکه الگوریتم‌های تشخیص اشیاء پیشرفت‌های بی‌نظیری داشته‌اند و با توجه به پیشرفت تکنولوژی و بروزرسانی فناوری‌ها، این مسئله پیش می‌آید که از الگوریتم‌ها و رویکردهای کنونی کدام یک عملکرد قابل قبولی ارائه می‌دهند. حال ما در نظر داریم با مقایسه الگوریتم‌هایی که از فناوری شبکه عصبی در تک مرحله و دو مرحله استفاده می‌کنند و رویکردهای سنتی مبتنی بر استخراج ویژگی را از دو منظر معیار دقت و زمان پاسخ مقایسه و به بحث و بررسی بگذاریم. الگوریتم‌هایی که ما به منظور پیشبرد پژوهش استفاده می‌کنیم شامل RCNN، YOLO و HOG می‌باشد.

۲- ادبیات موضوعی

۲-۱- الگوریتم YOLO

در این الگوریتم شما یک بار به تصویر نگاه انداخته و اشیاء داخل آن را تشخیص می‌دهیم. YOLO بسیار ساده است، یک شبکه کانولوشنال منفرد که به طور همزمان چندین جعبه مرزی و احتمالات کلاس را برای آن جعبه‌ها پیش‌بینی می‌کند. YOLO روی تصاویر کامل تمرین می‌کند و به طور مستقیم عملکرد تشخیص را بهینه می‌کند. این مدل یکپارچه مزایای متعددی نسبت به روش‌های سنتی تشخیص اشیاء دارد، به دلیل اینکه در این الگوریتم از یک برنامه رگرسیون استفاده می‌کند پس نیازی به استفاده از یک خط لوله پیچیده ندارد و به سادگی تمام شبکه عصبی برای تشخیص اشیاء در تصاویر مورد استفاده واقع می‌شود. (J.Redmon et al, 2016)

YOLO هنگام پیش‌بینی تصویر را به صورت سراسری و کلی بررسی می‌کند، برخلاف روش‌های پیشنهادی مبتنی بر کشف منطقه، YOLO کل تصویر را در طول زمان آموزش می‌بیند و تست می‌کند، بنابراین به طور ضمنی اطلاعات متنی در مورد کلاس‌ها و اشیاء درون تصویر و همچنین ظاهر آن‌ها را رمزگذاری می‌کند و در یک بار عبور از شبکه عصبی می‌تواند باتوجه به وزن‌دهی‌هایی که انجام داده است، اشیاء را مورد تشخیص قرار دهد. (J.Redmon et al, 2016)

YOLO قابلیت تعمیم اشیاء مشابه به یک کلاس را در طول مسیر تشخیص خود دارد. زمانی که YOLO بر روی مجموعه تصاویر آموزشی، آموزش دید و روی مجموعه تست آزمایش شد، از روش‌های تشخیص دیگر با اختلاف بسیار بهتر عمل می‌کند. از

www.SID.ir

آنجایی که YOLO بسیار قابل تعمیم است، در صورت اعمال دامنه‌های جدید یا ورودی‌های غیرمنتظره، احتمال خراب شدن و نتیجه ضعیف آن کمتر است. در شبکه عصبی الگوریتم، از ویژگی‌های کل تصویر برای پیش‌بینی هر کادر محدود استفاده می‌شود. همچنین تمام کادرهای محدود کننده در همه کلاس‌ها را برای یک تصویر به طور همزمان پیش‌بینی می‌کند. این بدان معناست که شبکه عصبی به طور سراسری در مورد اشیاء موجود در تصویر نتیجه‌گیری می‌کند. طراحی YOLO با حفظ میانگین دقت بالا، آموزش سرتاسر و سرعت بی‌درنگ را امکان پذیر می‌کند.(J.Redmon et al, 2016)(PeiyuanJiang et al, 2022)

۲-۲- الگوریتم RCNN

تشخیص اشیاء شامل دو وظیفه از هم جدا می‌باشد که شامل طبقه بندی و محلی سازی است. RCNN مخفف شبکه عصبی کانولوشنال مبتنی بر منطقه است. مفهوم کلیدی RCNN پیشنهادات منطقه‌ای است. پیشنهادات منطقه‌ای برای بومی سازی اشیاء در یک تصویر استفاده می‌شود. شبکه پیشنهادات منطقه‌ای را تولید می‌کند که در آن شی مورد تجزیه و تحلیل قرار می‌گیرد و منطقه شبکه از این پیشنهادها برای شناسایی اشیاء در یک تصویر استفاده می‌کند. در این الگوریتم از مفهوم Region proposals استفاده شده که به معنی بخش‌های کوچکی از تصویر هستند که می‌توانند شامل اشیایی باشند که دنبال آن‌ها هستیم. ما نیازمند تعدادی منطقه پیشنهادی هستیم، پس برای ایجاد تعدادی منطقه پیشنهادی در تصویر از الگوریتم جستجوی انتخابی استفاده می‌شود؛ نواحی انتخاب شده به مدل شبکه عصبی کانولوشنال داده می‌شود و یک بردار ویژگی ازاین نواحی استخراج می‌شود. برای بخش آموزش مدل شبکه هریک از بخش‌های انتخابی به صورت مستقل و جداگانه‌ای آموزش می‌بینند؛ بعداز استخراج بردار ویژگی هر بخش انتخابی، مرحله طبقه‌بندی بردارها می‌باشد. طبقه مورد استفاده در این الگوریتم دسته‌بند ماشین بردار پشتیبان می‌باشد. خروجی‌های دسته‌بندی بخش‌ها را باید در یک تصویر یکتا مشخص کرد که از روش gready non-maximum suppression استفاده می‌کند؛ همچنین از معیاری به نام intersection-over-union(IoU) برای انتخاب ناحیه‌هایی که با یکدیگر هم پوشانی دارند، استفاده می‌کند.(Ross Girshick et al, 2014)

۲-۳- الگوریتم HOG

هیستوگرام گرادیان‌های جهت‌دار، همچنین به عنوان HOG شناخته می‌شود، یک توصیف کننده ویژگی یا تعیین کننده ویژگی است، در بینایی ماشین و پردازش تصویر به منظور تشخیص اشیاء استفاده می‌شود. این تکنیک، جهت گیری گرادیان را در قسمت محلی یک تصویر شمارش می‌کند. این روش کاملاً شبیه به هیستوگرام‌های جهت گیری لبه و تبدیل ویژگی ثابت مقیاس است. توصیف‌کننده HOG بر ساختار یا شکل یک شی تمرکز می‌کند. از هر توصیف کننده لبه‌ای بهتر است زیرا از بزرگی و همچنین زاویه گرادیان برای محاسبه ویژگی‌ها استفاده می‌کند. برای مناطق تصویر، هیستوگرام‌هایی را با استفاده از بزرگی و جهت گیری گرادیان تولید می‌کند.

www.SID.ir

Arch

تصور اساسی در پشت هیستوگرام توصیف کننده گرادیان‌های جهت‌دار این است که ظاهر و شکل شی در یک تصویر را می‌توان با توزیع گرادیان‌های شدت یا جهت‌های لبه توصیف کرد. تصویر به مناطق کوچک متصل به نام سلول تقسیم می‌شود و برای پیکسل‌های درون هر سلول، یک هیستوگرام از جهت‌های گرادیان محاسبه می‌شود. توصیف کننده وصل کردن و بررسی این هیستوگرام‌ها است. برای دقت بهتر، هیستوگرام‌های محلی را می‌توان با محاسبه اندازه‌گیری شدت در ناحیه بزرگ‌تری از تصویر، به نام بلوک و سپس با استفاده از این مقدار برای عادی‌سازی تمام سلول‌های داخل بلوک، با تضادهای سفید و مشکی نرمال کرد. این نرمال سازی باعث یک همگام سازی یا قابلیت تغییر نکردن نسبت به تغییراتی که در مناطق روشن و سایه تصویر می‌شود. توصیف‌کننده HOG چند مزیت کلیدی نسبت به سایر توصیف‌کننده‌ها دارد. از آنجایی که بر روی سلول‌های محلی کار می‌کند، به جز برای شی گرایی، نسبت به تبدیل‌های هندسی و فتومتریک ثابت است. چنین تغییراتی فقط در مناطق فضایی بزرگتر ظاهر می‌شود. علاوه بر این، نمونه‌برداری فضایی بزرگ، نمونه‌گیری جهت‌گیری دقیق و نرمال‌سازی فتومتریک محلی قوی اجازه می‌دهد تا زمانی که افراد درون تصاویر وضعیت تقریباً عمودی خود را حفظ کنند، حرکات بدن فردی نادیده گرفته شود. بنابراین توصیف‌کننده HOG به ویژه برای تشخیص انسان در تصاویر مناسب است.(Kosuke Mizuno et al, 2012)

۳- روش تحقیق

روش پژوهشی این مقاله به این صورت می‌باشد که یک مجموعه داده که تعداد زیادی تصویر را دارا است که هر تصویر شامل کلاس‌ها یا ویژگی‌های مختلفی شامل انسان، خودرو، هواپیما، اتوبوس و غیره می‌باشد؛ این مجموعه تصاویر به دو بخش داده آموزشی برای مدل‌هایمان و داده تست و ارزیابی که هفتاد درصد داده آموزشی و سی درصد داده تست می‌باشد، تقسیم می‌کنیم. معیار ارزیابی عملکرد الگوریتم‌های تشخیص اشیاء که از دو منظر دقت و زمان پاسخ می‌باشد به این صورت است که برای زمان پاسخ زمانی که تصویر آماده وارد شدن به مدل چه در رویکردهایی که از فناوری شبکه عصبی بهره می‌برند و چه رویکردهای مبتنی بر تعیین ویژگی به طور خودکار زمان سنج فعال می‌شود و زمان خروجی نهایی که شامل یک تصویر که اشیاء داخل آن مشخص شده است زمان سنج خودکار متوقف شده است و زمان خروجی را نمایش می‌دهد و برای سه مدل ذکر شده این روند ادامه دارد تا کمترین زمان پاسخ را بتوان به دست آورد؛ همچنین تصویر وارد شده به مدل‌هایمان یک تصوی ثابت می‌باشد. برای به دست آوردن دقت الگوریتم‌ها از معیارهای Precision و Recall استفاده می‌کنیم(فرمول‌های ۱ و ۲). بعداز به دست آوردن نتایج معیارها آن‌ها را بایکدیگر مقایسه می‌کنیم.

$$Precision = tp / tp + fp \quad (1)$$

$$Recall = tp / tp + fn \quad (2)$$

www.SID.ir

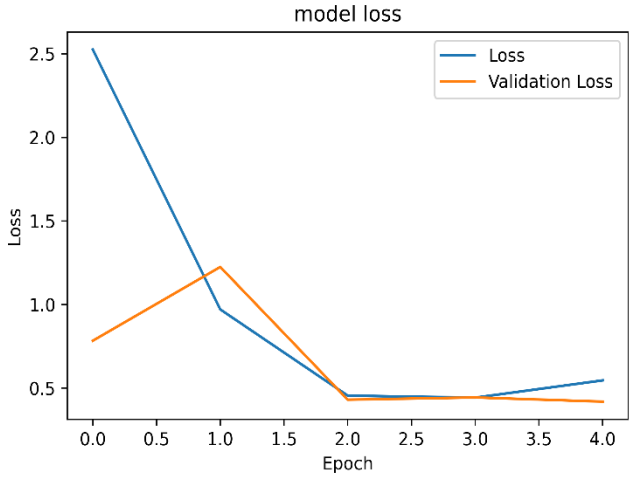
Arch

٤- یافته‌ها

باتوجه به مجموعه داده‌هایمان، الگوریتم YOLO در مقایسه با الگوریتم‌های دیگری که از فناوری شبکه عصبی بهره می‌بردند، بسیار سریع‌تر است. این الگوریتم نمایش شیء تعمیم‌یافته را درک می‌کند که به آن اجازه می‌دهد شبکه را بر روی تصاویر زمان واقعی آموزش دهد و اشیاء را کاملاً دقیق تشخیص دهد. سرعت از حدود ۴۵ تا ۱۵۵ فریم در ثانیه متغیر است. این الگوریتم برخلاف دیگر الگوریتم‌ها که تشخیص را در مناطق مختلف انجام می‌دهند و در نتیجه الگوریتم برای پیش‌بینی، چندین بار تصویر را برای پیدا کردن مناطق مختلف یک تصویر بررسی می‌کند، در الگوریتم YOLO یک شبکه عصبی کاملاً پیچیده است که تصویر را تنها یک بار عبور می‌دهد و خروجی به عنوان مرحله بعدی پیش بینی می‌شود. تصاویری که آموزش داده می‌شوند تصویر کامل هستند و این عملکرد الگوریتم را برای تشخیص اشیاء در یک تصویر بهینه می‌کند. اطلاعات تصویری که در کل مجموعه داده آموزش داده می‌شود و جزئیات تصاویر به صورت سراسری و کلی تحلیل می‌شوند. الگوریتم YOLO مرسوم ترین الگوریتمی است که برای تشخیص تصاویر طبیعی استفاده می‌شود، زیرا این الگوریتم کارآمدترین الگوریتم برای دامنه‌های جدید و ورودی‌هایی است که غیرمنتظره هستند. هر شی از روی جزئیات کل تصویر پیش‌بینی می‌شود و این امکان را فراهم می‌کند که آموزش کل تصویر دارای سرعت واقعی بالا و دقت بالا باشد. YOLO می‌تواند بسیاری از جعبه‌های مرزی را در یک سلول یا نرون شبکه عصبی تحلیل و پیش‌بینی کند و در طول آموزش مدل‌ها، تنها یک پیش‌بینی‌کننده از جعبه‌های مرزی وجود دارد که هر شی درون تصاویر را پیش‌بینی می‌کند. پیش‌بینی‌کننده‌های یک شی منفرد در تصویر برای پیش‌بینی اشیاء مختلف دیگری در یک تصویر اختصاص داده می‌شود. آموزش مدل‌ها بر این اساس انجام می‌شود و این به الگوریتم کمک می‌کند تا اشیاء موجود در یک تصویر را به طور موثر تعریف و پیش‌بینی کند. محدودیت مکانی و فضایی یکی از اشکالات اصلی الگوریتم است زیرا به عنوان یک شبکه می‌تواند تنها تا دو بلوک را تجزیه و تحلیل کند و این دو بلوک می‌توانند فقط یک کلاس را شامل شوند. این باعث کاهش در تشخیص اجسام نزدیک می‌شود. تشخیص تصاویر کوچک در یک تصویر گروهی با کمک این الگوریتم بسیار دشوار است.

طبق شواهد نشان می‌دهد که الگوریتم‌های RCNN و YOLO هردو از دقت بالایی برخوردارند. به طور مثال در مجموعه داده احتمال اینکه یک شی را به عنوان خودرو در نظر بگیرند بسیار زیاد حدود ۹۰ درصد می‌باشد؛ این عدد نتیجه فرمول‌های یک و دو می‌باشد که در مقایسه با یکدیگر قرار گرفته‌اند. الگوریتم HOG به دلیل اینکه بر ساختار یک شی متمرکز است و سعی می‌کند یک توصیف‌کننده ویژگی و زوایای شی را براساس گرادیان محاسبه کند دقت آنچنان بالایی نسبت به الگوریتم‌هایی که از شبکه عصبی استفاده می‌کنند ندارد زیرا احتمال اینکه زوایا و لبه‌های یک شی در تصویر واضح نباشد وجود دارد و مقدار نتایج معیارهایمان نتیجه‌ای ۵۰ درصدی را ارائه کرد.

www.SID.ir



شکل ۱- نمودار نمایش خطا در RCNN

در هر دوره محاسبات انجام شده در مدل الگوریتم RCNN مقدار خطای آموزش و خطای تست و ارزیابی مقدار زیادی کاهش پیدا کرده‌اند و در نهایت مقدار ۹۰ درصد دقت را ارائه می‌کند.

از نظر زمان پاسخ و اجرا الگوریتم HOG و YOLO سریعتر عمل کردند اما در مقایسه با یکدیگر الگوریتم HOG سریعتر بود اما الگوریتم YOLO به قدرت محاسباتی بیشتری نیاز دارد زیرا از شبکه عصبی استفاده می‌کند. از زمان ظهور شبکه‌های عصبی و یادگیری عمیق کاربرد تشخیص اشیاء هیستوگرام گرادیان کم شد و باتوجه به دقت بالای رویکردهای بهره‌مند از شبکه عصبی، رویکردهای مبتنی بر تعیین ویژگی بیشتر برای موارد خاص مانند تشخیص انسان در تصاویر مورد استفاده قرار گرفته‌اند. معیارهای دقت در الگوریتم RCNN نشان داد که برای بررسی یک تصویر از مجموعه داده زمانی حدود ۲۷.۶ ثانیه نیازمند است که این زمان نسبت به دیگر الگوریتم‌ها زمان زیادی می‌باشد و به این دلیل که در این الگوریتم بررسی برای تشخیص یک شی در تصویر به یک بار بسنده نمی‌کند بلکه یک تصویر را چند بار مورد بررسی قرار می‌دهد.

جدول ۱- جدول زمان پاسخ الگوریتم‌ها

الگوریتم	زمان پاسخ
HOG	۲ ثانیه
YOLO	۲.۵ ثانیه
RCNN	۲۷.۶ ثانیه

www.SID.ir

Arch

۵- بحث و نتیجه‌گیری

مقایسه با سه الگوریتم مختلف انجام شد و نتایج به دست آمده مورد بررسی قرار گرفت. با مقایسه این سه الگوریتم به این نتیجه رسیدیم که YOLO از نظر دقت چندان تفاوتی با RCNN ندارد و بهتر از HOG می‌باشد. از لحاظ زمان پاسخ یا پردازش الگوریتم HOG به دلیل اینکه شبکه عصبی و یادگیری عمیق در آن استفاده نشده است عملکرد بهتری دارد. الگوریتم RCNN از نظر زمان پاسخ ضعیف‌ترین عملکرد ممکن را داشت زیرا چند مرحله تصویر را مورد بررسی قرار می‌دهد؛ همچنین می‌توان به نیازمند بودن سیستم به قدرت محاسباتی اشاره‌ای داشت که الگوریتم HOG کمترین قدرت محاسباتی را نیازمند است.

٦- منابع

Z. Zou, Z. Shi, Y. Guo, J. Ye, and S. Member, “Object Detection in 20 Years : A Survey,” pp. 1–39, 2019.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo Algorithm Developments,” *Procedia Comput. Sci.*, vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.

R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

Mizuno, K., Terachi, Y., Takagi, K., Izumi, S., Kawaguchi, H. & Yoshimoto, M. (2012). Architectural Study of HOG Feature Extraction Processor for Real-Time Object Detection.. *SiPS* (p./pp. 197-202), : IEEE. ISBN: 978-1-4673-2986-6

www.SID.ir