Real-time object detection based on YOLO-v2 for tiny vehicle object

**چکیده:**

در سیستم های رانندگی خودکار (ADS) و سیستم های دستیار راننده (DAS)، شناسایی اشیاء نقشی حیاتی دارد. با این حال، مدل های شناسایی در زمان واقعی موجود برای خودروهای کوچک با مشکلاتی مانند دقت پایین و عملکرد نامطلوب روبرو هستند. برای حل این مشکلات، ما یک مدل جدید برای شناسایی اشیاء در زمان واقعی، بر اساس چارچوب یادگیری عمیق You Only Look Once نسخه ۲ (YOLO-v2)، با نام بهینه سازی شده ی YOLO-v2 یا O-YOLO-v2 برای خودروهای کوچک پیشنهاد می دهیم.

در مدل پیشنهادی، ساختار جدیدی معرفی شده که با افزودن لایه های کانولوشنی در مکان های مختلف، توانایی استخراج ویژگی های شبکه را تقویت می کند. همچنین، با اضافه کردن ماژول های باقی مانده، مشکل ناپدید شدن یا پراکندگی گرادیان که به دلیل افزایش عمق شبکه ایجاد می شود، برطرف شده است. علاوه بر این، برای بهبود دقت شناسایی خودروهای کوچک، ویژگی های سطح پایین و سطح بالای شبکه ترکیب شده اند. نتایج تجربی و تحلیل ها بر روی مجموعه داده KITTI نشان می دهند که این مدل نه تنها دقت شناسایی خودروهای کوچک را افزایش داده، بلکه دقت شناسایی کلی خودروها را نیز بهبود داده است و به دقت ۹۴ درصد دست یافته است، به طوری که سرعت شناسایی کاهش نیافته است.

**مقدمه:**

در سال های اخیر، پیشرفت های بزرگی در شناسایی اشیاء، به ویژه با ترویج شبکه های عصبی کانولوشنی عمیق (CNN)، صورت گرفته است. با توجه به مشکلات ازدحام ترافیکی و ایمنی رانندگی، استفاده از شناسایی خودروها در سیستم های رانندگی خودکار (ADS)، سیستم های دستیار راننده (DAS)، سیستم های حمل و نقل هوشمند (ITS) و موارد مشابه اهمیت زیادی دارد. با این حال، شناسایی اشیاء در صحنه های پیچیده ترافیکی همچنان با چالش های متعددی روبرو است، مانند خودروهای کوچکی که توسط اشیاء دیگر پنهان می شوند و یا شرایط محیطی مختلفی همانند تاری، بارانی بودن و شب را دارند.

در مدل های اولیه شناسایی اشیاء، ویژگی های دستی و ساختارهای طبقه بندی ترتیبی (cascade classifier) بسیار محبوب بودند. این مدل ها ابتدا با استفاده از پنجره متحرک، ناحیه اشیاء را در تصاویر اصلی شناسایی کرده و سپس انواع مختلفی از ویژگی ها را استخراج می کردند و در نهایت از یک طبقه بند برای شناسایی اشیاء مختلف استفاده می شد. برای مثال، Meng و همکاران از ویژگی های Haar-Like برای توصیف ویژگی های لبه، خطی، مرکزی و مورب تصاویر چهره استفاده کرده و سپس از SVM برای شناسایی اشیاء بهره بردند. این مطالعات اولیه به موفقیت های محدودی دست یافتند، اما این مدل ها تنها می توانستند ظاهر و شکل اشیاء را از طریق ویژگی های مختلف توصیف کنند و قادر به بیان ویژگی های عمیق تر نبودند. این امر باعث می شد مدل ها به سختی با مجموعه داده های جدید تطبیق یابند و شناسایی اشیاء در صحنه های متغیر ترافیکی بسیار دشوار بود. با هدایت نظریه یادگیری عمیق، شناسایی اشیاء پیشرفت های زیادی در سرعت و دقت داشته است. به عنوان مثال، Zhang و همکاران مدلی توسعه دادند که ابتدا از الگوریتم دید دوچشمی برای پیش پردازش تصویر استفاده می کند و سپس از Faster R-CNN برای شناسایی خودرو بهره می برد. آزمایش ها نشان داد که این مدل سرعت شناسایی اشیاء را به طور قابل توجهی افزایش داده است. Gao و همکاران نیز یک ساختار شبکه مبتنی بر YOLO بهبود یافته را پیشنهاد کردند تا دقت و سرعت شناسایی اشیاء را افزایش دهند. همه مدل های ذکرشده سعی دارند مدلی را توسعه دهند که نیازهای شناسایی اشیاء در سناریوهای ترافیکی را برآورده کند. اگرچه این مدل ها دقت و سرعت شناسایی اشیاء را افزایش داده اند، اما به دلیل محدودیت های Faster R-CNN، نمی توانند نیازهای شناسایی در زمان واقعی را برآورده کنند. مدل مبتنی بر YOLO سرعت شناسایی را به طور قابل توجهی افزایش داده است، اما به دلیل ساختار شبکه کم عمق، دقت شناسایی در مقایسه با Faster R-CNN کاهش می یابد. همچنین، این مدل در شناسایی اشیاء کوچک عملکرد ضعیفی دارد و نمی تواند خودروهایی که در صحنه های ترافیکی از دید پنهان شده اند را شناسایی کند.

با توجه به مشکلات پژوهش های فوق، ما مدل جدیدی به نام Optimized You Only Look Once Version 2 یا (O-YOLO-v2) را بر اساس YOLO-v2 پیشنهاد می دهیم. نکات برجسته این مدل شامل موارد زیر است: (1) ساختار شبکه را با افزودن لایه های کانولوشنی در مکان های مختلف طراحی کرده ایم تا توانایی استخراج ویژگی های شبکه بهبود یابد. همچنین، ماژول های باقی مانده ای را اضافه کرده ایم تا مشکل ناپدید شدن یا پراکندگی گرادیان که عمدتاً به دلیل لایه های شبکه ایجاد می شود، حل شوند. (2) ویژگی های استخراج شده از سطوح مختلف شبکه را ترکیب می کنیم تا دقت شناسایی اشیاء کوچک افزایش یابد. (3) مدل پیشنهادی، ویژگی های عمیق تری را استخراج می کند و در عین حال سرعت آموزش شبکه را افزایش می دهد، که نیازهای شناسایی اشیاء در صحنه های ترافیکی را برآورده می کند. بخش های بعدی مقاله به صورت زیر سازماندهی شده است: بخش ۲ نظریه YOLO-v2 را توضیح می دهد. بخش ۳ جزئیات فنی مدل پیشنهادی O-YOLO-v2 را معرفی می کند. به طور خاص، نمای کلی مدل پیشنهادی در بخش ۳.۱ نشان داده شده و تابع هزینه مدل در بخش ۳.۲ توصیف شده است. برای تأیید اعتبار مدل پیشنهادی، آزمایش های گسترده ای در بخش ۴ انجام شده اند و بخش ۵ نیز ویژگی های O-YOLO-v2 را خلاصه می کند.

**اصول Yolo-v2:**

YOLO-v2 به طور مستقیم با استفاده از یک شبکه عصبی، bounding box و احتمال دسته بندی را پیش بینی می کند. که تصویر به شبکه ای از سلول ها با ابعاد 𝑆×𝑅 تقسیم می شود، به طوری که هر سلول فقط می تواند یک شیء را پیش بینی کند. نقش هر سلول در این شبکه این است که چیزی را شناسایی کند که مرکز آن در محدوده این سلول قرار گرفته است. لذا هر سلول می تواند bounding box، نمره اطمینان (confidence score) این box ها و احتمالات دسته بندی را پیش بینی کند. هر bounding box شامل ۵ بخش است: (x, y, w, h, confidence C).

مختصات (x, y) نسبت به موقعیت سلول شبکه به مرکز box اشاره دارند. از نظر دامنه مختصات، این مختصات را به بازه ۰ تا ۱ نرمال سازی می کنیم. عرض و ارتفاع bounding box نیز با ابعاد (w, h) نسبت به اندازه تصویر تنظیم می شوند که بازه آن ها هم ۰ تا ۱ است. از طرفی confidence score پیش بینی شده دو مفهوم را نشان می دهند: اول، اطمینان مدل از وجود شیء در box و دوم، میزان دقت پیش بینی مدل از این box.

confidence score صفر به این معنی است که سلول حاوی هیچ شیء ای نیست. در غیر این صورت، انتظار داریم confidence score برابر با تلاقی بر اتحاد (IOU) بین box پیش بینی شده و واقعیت موجود باشد. در برخی موارد، یک سلول شبکه ممکن است حاوی چندین شیء باشد. شکل ۱ فرآیند محاسباتی شبکه و bounding box را نشان می دهد و شکل ۲ نمودار شناسایی در الگوریتم YOLO-v2 را نمایش می دهد. ساختار شبکه YOLO-v2 در جدول ۱ ارائه شده است.

|  |
| --- |
|  |
|  |

مقدار confidence score مطابق با معادله (1) محاسبه می شود.

|  |
| --- |
|  |

در اینجا، IOU به معنی تداخل مشترک بین واقعیت موجود و جعبه پیش بینی شده است. Pr(project) احتمال پیش بینی است که آیا boundary object حاوی شیء خودرو است یا خیر. اگر شیء ای وجود داشته باشد، Pr(project)=1 و در غیر این صورت، Pr(project)=0. همچنین احتمال دسته بندی p طبق معادله (2) محاسبه می شود:

|  |
| --- |
|  |
|  |

**مدل شناسایی پیشنهادی:**

اگرچه YOLO-v2 مزایایی دارد از جمله توانایی پیش بینی مستقیم موقعیت، اندازه و احتمال دسته بندی شیء از تصویر ورودی و همچنین دستیابی به نتایج خوب در شناسایی و سرعت بالای آن، اما در صحنه های ترافیکی واقعی، ما غالباً با اشیاء خودروهای کوچک مواجه هستیم که بیشتر به ویژگی های سطح پایین وابسته اند. به دلیل ساختار شبکه کم عمق YOLO-v2، ویژگی های سطح پایین نمی توانند به طور کامل از طریق مدل YOLO-v2 استخراج شوند و اطلاعات ویژگی ها در فرآیند انتقال بین لایه های مختلف از دست می روند. بنابراین، اطلاعات ویژگی های چندلایه به طور کامل در فرآیند پیش بینی مورد استفاده قرار نمی گیرد که این امر دقت مکان یابی اشیاء را کاهش می دهد.

* **مدل شناسایی پیشنهادی**

به منظور افزایش دقت شناسایی اشیاء و بهبود عملکرد شناسایی اشیاء کوچک، این مقاله به توسعه یک مدل جدید می پردازد که O-YOLO-v2 نامیده می شود و بر اساس YOLO-v2 طراحی شده است. هدف اصلی این معماری جدید، ارتقای عملکرد استخراج ویژگی ها است. به علاوه، ماژول باقی مانده ای برای حل مشکل ناپدید شدن یا پراکندگی گرادیان که ناشی از افزایش عمق شبکه است، پیشنهاد شده است. علاوه بر این، برای استفاده کامل از ویژگی ها در نقشه های ویژگی سطح کم و بهبود دقت شناسایی خودروهای کوچک، مدل ویژگی های سطح پایین و سطح بالا را ترکیب می کند. همچنین، با توجه به اینکه خطای موقعیت ایجاد شده توسط O-YOLO-v2 در حین آموزش بیشتر از خطای طبقه بندی است، تنظیم پارامترهای تابع هزینه انجام می شود تا عملکرد مدل بیشتر ارتقا یابد. معماری پیشنهادی در شکل ۳ نشان داده شده است که در آن Conv3-X به معنای لایه کانولوشنی است؛ اندازه هسته کانولوشن 3×3 است؛ تعداد کانال X است؛ اندازه گام پیش فرض برابر ۱ است؛ Pool2 به معنای لایه حداکثر pooling است؛ فیلتر 2×2 است و اندازه گام پیش فرض برابر ۲ است؛ Residual به معنای ساختار بلوک باقی مانده است. این معماری شامل ۳۱ لایه کانولوشنی، ۶ لایه pooling، ۱ ساختار سازماندهی ویژگی و ۱۱ ماژول باقی مانده است. ورودی به شبکه 416×1248×3 و خروجی 13×39×30 است.

همان طور که در شکل ۳ نشان داده شده است، دو ماژول باقی مانده از طریق لایه های کانولوشنی ششم، هفتم، دهم و یازدهم ساخته می شوند (لایه های ششم و هفتم برای ماژول باقی مانده اول؛ لایه های دهم و یازدهم برای ماژول باقی مانده دوم). پس از لایه یازدهم، یک ماژول باقی مانده جدید اضافه می شود که شامل دو لایه کانولوشن است. پس از آن، دو ماژول باقی مانده دیگر از طریق لایه های کانولوشنی شانزدهم، هفدهم، هجدهم و نوزدهم ایجاد می شوند (لایه های شانزدهم و هفدهم برای ماژول باقی مانده اول؛ لایه های هجدهم و نوزدهم برای ماژول باقی مانده دوم).

پس از لایه نوزدهم، دو ماژول باقی مانده جدید اضافه می شوند. سپس، دو ماژول باقی مانده دیگر از طریق لایه های کانولوشنی بیست و ششم، بیست و هفتم، بیست و هشتم و بیست و نهم ساخته می شوند (لایه های بیست و ششم و بیست و هفتم برای ماژول باقی مانده اول؛ لایه های بیست و هشتم و بیست و نهم برای ماژول باقی مانده دوم).

پس از لایه بیست و نهم، دو لایه کانولوشن حذف شده و دو ماژول باقی مانده جدید اضافه می شوند. در نهایت، نقشه های ویژگی خروجی از لایه های سیزدهم، بیست و سوم و سی و سوم ترکیب می شوند و همان طور که در شکل ۳ نشان داده شده است، feature map (52×156×256) خروجی از لایه سیزدهم، feature map (26×78×512) خروجی از لایه بیست و سوم و feature map (13×39×1024) خروجی از لایه سی و سوم با هم ترکیب می شوند. feature map (52×156×256) از طریق یک میانگین pooling، نقشه ویژگی (13×39×256) را تولید می کند. feature map (13×39×256) پس از سازماندهی ویژگی و یک لایه کانولوشن از feature map (26×78×512) تولید می شود. پس از افزودن feature map لایه های سیزدهم، بیست و سوم و سی و سوم، feature map (13×39×1536) تولید شده و سپس به لایه بعدی ورودی داده می شود.

زیرا ویژگی های اشیاء کوچک عمدتاً در لایه های سطح کم شبکه قرار دارند و ویژگی های اشیاء بزرگ در لایه های عمیق تر شبکه موجودند، ما ویژگی های چندلایه ای از ویژگی های سطح پایین و سطح بالا را ترکیب می کنیم تا توانایی بیانfeature های بهبود یافته برای اشیاء کوچک را به دست آوریم و دقت شناسایی اشیاء کوچک را افزایش دهیم.

|  |
| --- |
|  |

برای حل مشکل ناپدید شدن یا پراکندگی گرادیان که ناشی از افزایش عمق شبکه است، این مقاله پیشنهاد می کند که ماژول های باقی مانده به مدل اضافه شوند. علاوه بر این، ماژول های باقی مانده افزوده شده می توانند توانایی استخراج ویژگی های مدل را نیز افزایش دهند. ماژول باقی مانده پیشنهادی در شکل ۴ نشان داده شده است که شامل یک لایه کانولوشنی 1×1، یک لایه کانولوشنی 3×3 و یک نگاشت هویتی (identity mapping) است. لایه کانولوشنی 1×1 و لایه کانولوشنی 3×3 با حالت پرکننده SAME پر شده اند. در شرایطی که از حالت پرکننده SAME استفاده می شود، اندازه feature map تغییر نمی کند و معادل اندازه نگاشت هویتی است، بنابراین می توان آن ها را به طور مستقیم جمع کرد.

|  |
| --- |
|  |

* **تابع هزینه یا Loss Function**

با توجه به اینکه خطای موقعیت ایجاد شده توسط O-YOLO-v2 پیشنهادی در حین آموزش بیشتر از خطای طبقه بندی است، ما پارامترهای تابع هزینه را تنظیم می کنیم. وزن خطای موقعیت را به ۵.۵ افزایش می دهیم (وزن خطای موقعیت اولیه ۵ بوده است) و وزن خطای طبقه بندی را به ۰.۵ کاهش می دهیم (وزن اطمینان اولیه ۱ بوده است). علاوه بر این، برای محاسبه خطای طبقه بندی از تابع هزینه cross-entropy استفاده می کنیم که فرآیند همگرایی را در مقایسه با تابع هزینه مربعی اولیه تسریع می کند. تابع هزینه استفاده شده در این مقاله طبق معادله های (3)، (4) و (5) محاسبه می شود:

|  |
| --- |
|  |

در معادله (3)، 𝐿object​ نشان دهنده ی هزینه مدل برای اشیاء واقع در bounding box است؛ در معادله (4)، 𝐿no-object هزینه مدل برای اشیاء واقع نشده در bounding box را نشان می دهد؛ و در معادله (5)، 𝐿total نشان دهنده ی کل هزینه مدل است. xi, yi, wi, hi, ci, pi مقادیر برچسب واقعی و ~~x~~i, ~~y~~i, ~~w~~i, ~~h~~i, ~~c~~i, ~~p~~i مقادیر پیش بینی شده مدل هستند.

**آزمایش ها:**

به منظور تأیید عملکرد O-YOLO-v2 پیشنهادی، این مدل را با YOLO-v2 و YOLO-v3 مقایسه می کنیم.

* **تنظیمات آزمایش**

ما از فریم ورک TensorFlow و Keras استفاده می کنیم و از زبان برنامه نویسی Python 3 برای آموزش مدل یادگیری عمیق خود بهره می بریم. مدل آموزشی با استفاده از Nvidia Tesla K40 با 11G Memory، سیستم عامل CentOS 7.0 و ۲۴ گیگابایت حافظه انجام شد. مدل تست نیز با استفاده از Nvidia GeForce 1050 با 4G Memory، CPU Intel(R) Core(TM) i7-7700HQ با فرکانس ۲.۸ گیگاهرتز، ۱۶ گیگابایت حافظه و سیستم عامل Windows 10 صورت گرفت.

* **مجموعه داده ها یا** **Datasets**

ما از مجموعه داده KITTI برای آموزش و آزمایش مدل استفاده می کنیم. KITTI یک Dataset عمومی است که برای تأیید مدل های شناسایی شامل شناسایی خودرو، ردیابی خودرو و تقسیم بندی معنایی به کار می رود. ما تصاویر اشیاء ۲ بعدی در KITTI را به عنوان Dataset خود انتخاب می کنیم. در مجموع، ۷۴۸۱ تصویر از صحنه های ترافیکی واقعی در KITTI وجود دارد که از این تعداد، ۶۱۷۸ تصویر به صورت تصادفی به عنوان مجموعه آموزشی و ۱۳۰۳ تصویر باقی مانده به عنوان مجموعه اعتبارسنجی یا تستی انتخاب شده اند. به منظور متناسب سازی تصاویر با ساختار شبکه ما و جلوگیری از، از دست دادن اطلاعات در فرآیند کانولوشن، از روش درونیابی (interpolation) به منظور interpolation در قسمت های راست و زیر تصاویر استفاده می شود و اندازه تصویر به 416×1248 تغییر می کند. در مجموعه داده KITTI، هشت دسته وجود دارد. ما برچسب های دسته Van، Truck و Tram را به دسته Car تغییر می دهیم و دسته خودرو اصلی بدون تغییر باقی می ماند. سایر دسته ها به عنوان پس زمینه نادیده گرفته می شوند. طبق ساختار دایرکتوری مجموعه داده VOC، فایل های حاشیه نویسی Dataset به فرمت XML تبدیل می شوند. سمت چپ شکل ۵ نمونه ای از تصاویر اصلی و سمت راست شکل ۵ نمونه ای از تصاویر پردازش شده را نشان می دهد.

|  |
| --- |
|  |

* **شاخص های ارزیابی**

IOU، Recall، Precision و دقت متوسط (Average Precision یا AP) برای ارزیابی O-YOLO-v2 استفاده می شوند. IOU مخفف Intersection over Union بین واقعیت موجود و bounding box های پیش بینی شده توسط مدل است که در معادله (6) نشان داده شده است. در این معادله، 𝐵pb​ معادل با bounding box پیش بینی شده و 𝐵gt​ نیز bounding box واقعی (ground truth) است؛ area() معادل با union shape or intersection shape از 𝐵pb و 𝐵gt را نشان می دهد. وقتی که شرط IOU بیشتر از مقدار آستانه 𝑇iou باشد، bounding box به عنوان نتیجه صحیح در نظر گرفته می شود. رابطه بین واقعیت موجود و مقدار پیش بینی بر اساس این اصل اندازه گیری می شود که همبستگی و مقدار تابعی از یک رابطه تناسبی پیروی می کنند.

|  |
| --- |
|  |

دقت متوسط (AP) مدل شناسایی ما با استفاده از IOU محاسبه می شود. bounding box پیش بینی شده 𝐵pb​ از طریق مدل و بر اساس تصویر ورودی پیش بینی می شود. اگر مقدار IOU بین 𝐵pb​ و 𝐵gt از مقدار آستانه 𝑇iou فراتر رود و متغیرها از قانون معادله (6) پیروی کنند، پیش بینی به عنوان درست در نظر گرفته می شود. شکل ۶ نمونه ای از شناسایی خودرو در یک تصویر را نشان می دهد؛ جعبه قرمز یا آبی نمایانگر bounding box پیش بینی شده است و جعبه سبز نشان دهنده bounding box واقعیت موجود (ground truth) است. هدف ما محاسبه IOU بین bounding box های پیش بینی شده و واقعیت موجود است. وقتی IOU بیش از آستانه ۵۰٪ باشد، نتیجه آزمایش به عنوان پیش بینی مثبت (TP) تعریف می شود و مقدار کمتر از آستانه به عنوان پیش بینی مثبت نادرست (FP) در نظر گرفته می شود. FN نشان می دهد که در تصویر هیچ خودرویی وجود ندارد، اما در واقع، خودرویی در تصویر موجود است.

|  |
| --- |
|  |

دقت آشکارساز شیء معمولاً با استفاده از IOU ارزیابی می شود. اهمیت IOU را می توان به این صورت توضیح داد: تخصیص anchor boxes هنگام آماده سازی مجموعه داده های آموزشی و همچنین مفید بودن استفاده از الگوریتم non-max suppression برای حذف box های اضافی که مدل برای یک شیء پیش بینی کرده است. مقدار آستانه 𝑇IOU روی ۰.۵ تنظیم شده است، به این معنی که همان ناحیه باید حداقل توسط نیمی از واقعیت موجود و جعبه پیش بینی شده پوشش داده شود. وقتی IOU از آستانه ۵۰٪ عبور کند، خودرو با موفقیت شناسایی شده است. دقت به صورت تقسیم تعداد پیش بینی های مثبت درست (True Positive) بر تعداد کل تصاویر شناسایی شده تعریف می شود که با معادله (7) محاسبه می شود. در این معادله، 𝑇𝑃 نشان دهنده پیش بینی مثبت درست و 𝐹𝑃 نشان دهنده پیش بینی مثبت نادرست است.

|  |
| --- |
|  |

دقت بازیابی (Recall) به عنوان نسبت اشیاء مرتبط بازیابی شده به کل تعداد اشیاء مرتبط در تصاویر تعریف می شود که با معادله (8) محاسبه می شود. در این معادله، 𝐹𝑁 نشان دهنده پیش بینی منفی نادرست (False Negative) است.

|  |
| --- |
|  |

مخرج کسر دقت بازیابی (Recall) برابر است با مجموع پیش بینی های مثبت درست (True Positive) و منفی نادرست (False Negative). در معادله (8)، مجموع این دو مقدار برابر با تعداد کل خودروهای واقعیت موجود است. آخرین روش ارزیابی، دقت متوسط (AP)، به صورت مساحت زیر منحنی های دقت (Precision) و دقت بازیابی (Recall) نمایش داده می شود. آزمایش ها نشان می دهند که مقدار AP مستقیماً با دقت شناسایی متناسب است. با محاسبه مقادیر TP، FP و FN برای هر نمونه، مقادیر دقت بازیابی مربوط به هر نمونه به دست می آیند و سپس منحنی های دقت بازیابی رسم می شوند. مقدار AP که بین ۰ تا ۱ قرار می گیرد، طبق معادله (9) محاسبه می شود و در آن 𝑓𝑃𝑅 نمایانگر منحنی های دقت و دقت بازیابی است.

|  |
| --- |
|  |

* **آموزش**

ما مدل های O-YOLO-v2، YOLO-v2 و YOLO-v3 را بر روی مجموعه داده KITTI آموزش می دهیم تا عملکرد آن ها را مقایسه کنیم. به منظور افزایش سرعت آموزش مدل، از بهینه ساز Adam استفاده کردیم. در این آزمایش، impulse constant برابر با ۰.۹ و ضریب کاهش وزن برابر با ۰ در نظر گرفته شده است. همچنین نرخ یادگیری را به ۰.۰۰۰۰۵ تنظیم کردیم. اندازه دسته (batch size) برابر با ۴ تعیین شده، به این معنی که در هر تکرار، ۴ نمونه از مجموعه آموزشی پردازش می شوند و ۲۴۰ دوره (epoch) انجام می گیرد. تعداد کل تکرارهای آموزش ۳۱۲,۷۲۰ است. با استفاده از Dataset پردازش شده KITTI، به دست آوردن وزن های نهایی مدل شناسایی خودرو تقریباً ۴ روز بر روی GPU زمان برد. شکل ۷ منحنی کاهش هزینه در طول آموزش مدل را نشان می دهد، که در آن مقدار هزینه O-YOLO-v2 به ۰.۰۲۷، هزینه YOLO-v2 به ۰.۰۳۲ و هزینه YOLO-v3 به ۰.۰۳۰ کاهش یافته است.

|  |
| --- |
|  |

* **آزمایش و تحلیل ها**

همان طور که در جدول ۲ نشان داده شده است، نتایج مختلف آزمایش عملکرد مدل پیشنهادی O-YOLO-v2، مدل YOLO-v2 و مدل YOLO-v3 را بر روی مجموعه آزمایش نمایش می دهند. می توان نتیجه گرفت که مدل O-YOLO-v2 پیشنهادی، همان سرعت شناسایی YOLO-v2 را حفظ کرده است. با این حال، دقت بازیابی (Recall) به میزان ۶٪، دقت (Precision) به میزان ۳٪ و دقت متوسط (AP) به میزان ۷٪ افزایش یافته است. در مقایسه با مدل YOLO-v3، دقت مدل O-YOLO-v2 پیشنهادی به میزان ۱٪ افزایش یافته است. در حالی که دقت بازیابی و دقت متوسط مدل O-YOLO-v2 پیشنهادی به ترتیب ۲٪ و ۱٪ کاهش یافته اند، اما سرعت شناسایی O-YOLO-v2 پیشرفت چشمگیری داشته و ۵۳٪ افزایش یافته است. به طور خلاصه، مدل O-YOLO-v2 پیشنهادی دقت شناسایی را نسبت به YOLO-v2 افزایش داده و سرعت شناسایی را نیز حفظ کرده است. در مقایسه با YOLO-v3، اگرچه مدل O-YOLO-v2 پیشنهادی تا حدی از دقت شناسایی صرف نظر کرده، اما بهبود چشمگیری در سرعت شناسایی ایجاد کرده است.

|  |
| --- |
|  |

شکل ۸ منحنی های دقت بازیابی (Recall) و دقت (Precision) مدل های O-YOLO-v2، YOLO-v2 و YOLO-v3 را بر روی مجموعه آزمایش نشان می دهد. همان طور که در شکل ۸ مشاهده می شود، منحنی آبی مربوط به دقت بازیابی و دقت مدل O-YOLO-v2 بر روی مجموعه آزمایش است. منحنی های سبز و قرمز به ترتیب نمایانگر منحنی های دقت بازیابی و دقت مدل های YOLO-v2 و YOLO-v3 بر روی مجموعه آزمایش هستند. می توان مشاهده کرد که منحنی قرمز به نقطه مختصات (1,1) نزدیک تر است و ناحیه تشکیل شده بین محورهای 𝑥 و 𝑦 و منحنی های دقت بازیابی و قرمز بزرگ تر است، که نشان دهنده ی مقدار AP بالاتری است. در مقایسه با منحنی قرمز مدل YOLO-v3، مقدار AP مدل O-YOLO-v2 بر روی مجموعه آزمایش ۱٪ کاهش یافته است.

|  |
| --- |
|  |

شکل ۹ نتایج مقایسه مدل های YOLO-v2 و O-YOLO-v2 را بر روی مجموعه آزمایش نشان می دهد. از شکل ۹ می توان مشاهده کرد که مدل O-YOLO-v2 پیشنهادی قادر است تعداد بیشتری از خودروها را شناسایی کند و همچنین نتایج خوبی برای شناسایی خودروهای کوچک و دور از دسترس دارد.

|  |
| --- |
|  |

همان طور که در بالا نشان داده شده است، با حفظ سرعت شناسایی، مدل پیشنهادی دقت بازیابی (Recall) مدل را در شناسایی اشیاء بهبود می بخشد. در مقایسه با دو مدل دیگر، مدل O-YOLO-v2 پیشنهادی می تواند برای شناسایی اشیاء کوچک خودرو در سناریوهای دنیای واقعی کارآمدتر باشد.

**نتیجه گیری:**

در این مقاله، ما یک مدل شناسایی اشیاء جدید به نام O-YOLO-v2 را بر اساس YOLO-v2 معروف معرفی می کنیم. در این مدل، با استفاده از یک معماری شبکه جدید، عملکرد شناسایی برای اشیاء کوچک بهتر از YOLO-v2 است. معماری معرفی شده توانایی استخراج ویژگی های شبکه را بهبود می بخشد و در عین حال، ماژول های باقی مانده برای حل مشکل ناپدید شدن یا پراکندگی گرادیان که عمدتاً به دلیل انباشت لایه های شبکه ایجاد می شود، اضافه شده اند. علاوه بر این، با توجه به معرفی اختلال در ویژگی های چند بعدی، مدل پیشنهادی دقت شناسایی اشیاء کوچک را افزایش داده است. نتایج آزمایش نشان می دهند که مدل O-YOLO-v2 دقت شناسایی بالاتری را به دست آورده و به طور کلی نیازهای شناسایی ویدئویی را برآورده می کند. این مدل می تواند نه تنها برای شناسایی خودرو در سناریوهای ترافیکی واقعی بلکه برای شناسایی اشیاء در سایر حوزه ها نیز مورد استفاده قرار گیرد.

اگرچه مدل پیشنهادی در این مقاله نتایج شناسایی بهتری به دست آورده، هنوز فضای بهبودی در دقت مدل وجود دارد. در عین حال، تصاویری که در این مقاله استفاده شده، همگی در شرایط نوری مناسب بوده اند. بنابراین، نیاز است که مدل شبکه برای صحنه هایی با شرایط نوری نامناسب بیشتر مورد تأیید قرار گیرد، که این موضوع در آینده بیشتر مورد مطالعه قرار خواهد گرفت.