

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک

دوربین و روش‌های یادگیری عمیق

زهرا بادامچی شبستری، کارشناس ارشد سنجش از دور، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران،

ایران

علی حسینی نوه (مسئول مکاتبات)، استادیار دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

E-mail: hosseinaveh@kntu.ac.ir

پذیرش: ۱۳۹۹/۰۹/۳۲

دریافت: ۱۳۹۹/۰۷/۳۰

چکیده

در ایران موتورسیکلت‌ها یکی از آسیب‌پذیرترین کاربران جاده‌ها هستند که حجم زیادی از آمار تصادفات را شامل می‌شوند. در این مقاله راهکار مناسبی برای کمک به کاهش تصادفات اتومبیل‌ها با موتورسیکلت‌ها و به ویژه موتورسیکلت‌های مجهز به بادگیرهای مشکی رنگ ارائه شده است که با استفاده از تنها یک دوربین مستقر بر روی آینه بغل سمت کمک راننده، آگاهی راننده نسبت به نقاط کور محدودی بغل و پشت سر افزایش می‌یابد، تا در صورت نزدیک شدن بیش از حد موتورسوار، با اعلام هشدار به راننده از بروز تصادف جلوگیری شود. این عملیات هشدار با توجه به تلفیق اطلاعات بدست آمده از دو مرحله تشخیص موتورسیکلت و سپس برآورد فاصله با استفاده از روش‌های یادگیری عمیق صورت گرفته است. در مرحله تشخیص، مدل‌های مختلفی از الگوریتم‌های YOLO با یکدیگر مقایسه شده‌اند که در میان آنها، مدل بهبود یافته‌ی YOLOV4 با میانگین دقت ۸۰ درصد و سرعت ۳۵ فریم بر ثانیه بهترین عملکرد در شناسایی موتورسیکلت‌های مورد نظر را داشته است. این مدل بر روی پایگاه داده شامل ۲۰۰۰ تصویر اخذ شده از موتورسیکلت‌های شهر تهران آموزش داده شده است. در مرحله‌ی دوم برای برآورد نقشه عمق تک تصویر از آموزش مدل Monodepth2 بر روی ۶۰۰۰ جفت تصویر اخذ شده از خیابان‌های شهر تهران با استفاده از دوربین MYNT-EYE استفاده شده است. با ادغام نتایج بدست آمده از مختصات موتورسیکلت در تصویر و نقشه عمق بدست آمده، الگوریتم پیشنهادی به بهترین عملکرد در تشخیص و برآورد فاصله‌ی دوربین تا موتورسیکلت مورد نظر با میانگین خطای ۳۶ سانتیمتر و سرعت ۲۰ فریم بر ثانیه دست یافته است.

واژه‌های کلیدی: یادگیری عمیق، شناسایی موتورسیکلت، تولید نقشه عمق، YOLOV4، Monodepth2

۱. مقدمه

در ایران طبق گزارش‌های منتشرشده توسط پلیس راهور آمار تلفات موتورسیکلت سواران در سال ۱۳۹۷ حدود ۲۱/۸ درصد از کل تلفات حوادث رانندگی بوده است که از این میان ۳۷۴۵ موتورسوار در حوادث رانندگی جان خود را از دست داده‌اند. بر اساس آمارهای موجود بیشترین آمار تلفات موتورسواران مربوط به استان تهران بوده که طبق بررسی‌های انجام شده بر روی شغل این موتورسواران، ۴۱ درصد این آمار مربوط به موتورسواران با شغل آزاد همچون پیک موتوری بوده است (عصرخودرو، ۲۰۱۹). طبق بررسی‌های صورت گرفته بر روی موتورسیکلت‌ها و همچنین پیک‌های موتوری مورد استفاده در خیابان‌های شهر تهران دیده شده که اکثر این موتورسواران برای راحتی استفاده از موتورسیکلت در طول سال و شرایط مختلف آب و هوایی، همچون شکل ۱ از پوشش‌های بادگیر مشکی رنگی در قسمت جلویی موتورسیکلتشان استفاده می‌کنند (عصرخودرو، ۲۰۲۰). نصب این بادگیرها در جلوی موتورسیکلت باعث می‌شود که تصویری که از روبرو از موتورسیکلت گرفته می‌شود متفاوت از تصویر بدون این پوشش بادگیر مشکی رنگ باشد؛ بنابراین برای پایش این دسته از موتورسیکلت‌ها در تصویر با دقت و سرعت مناسب، می‌بایستی که الگوریتم متناسبی برای شناسایی و تعیین موقعیت آنها ارائه شود.



شکل ۱. تصویر موتورسیکلت با پوشش بادگیر

طبق بررسی‌های انجام شده بیشتر مطالعات صورت گرفته بر روی سیستم‌های هشدار و کمکی راننده، با استفاده از سنجنده‌های گرانتیمتی همچون رادار و لیزر بوده است (Dabbour and Easa, 2014; GONG et al., 2019; Gruyer and Rahal, 2019) و در نتیجه مطالعات کمتری در زمینه‌ی پردازش تصاویر رقومی به ویژه تک تصویر صورت گرفته است (Sivaraman and Trivedi, 2013). توجه به چالش‌های موجود در تشخیص موتورسیکلت‌های در حال حرکت مانند ابعاد کوچک موتورسیکلت‌ها نسبت به اتومبیل‌ها و سخت‌تر بودن تهیه‌ی داده‌های آموزشی و استخراج آن‌ها از تصویر، اکثر روش‌ها با استفاده از یک دوربین انجام شده است که به طور خاص برای تشخیص اتومبیل‌های سواری و برآورد سرعت و مسیر حرکت آنها بوده است (Luvizon et al., 2016; Sochor et al., 2018; Ouyang and Wang, 2019). از طرفی سیستم‌هایی که با هدف تشخیص موتورسیکلت فعالیت داشته‌اند، اکثراً برای شناسایی و رصد موتورسواران بدون پوشش کلاه ایمنی با استفاده از تصاویر اخذ شده از دوربین‌های ثابت مستقر بر روی پل‌ها یا چراغ‌ها و یا سایر نقاط ثابت شهر بوده است (Espinosa et al., 2017b; Vishnu et al., 2017; Espinosa et al., 2018b; Kulkarni et al., 2018). در عین حال در دسته‌ای دیگر از مطالعات دوربین از مسیر جلوی راننده تصویربرداری می‌کند و موتورسیکلت‌های موجود در مسیر حرکت با توجه به پلاک نصب شده در پشت سرشان تشخیص و تعیین موقعیت می‌شوند (Mistry et al., 2017b; Laroca et al., 2018).

با توجه به عدم استفاده‌ی تعداد قابل توجه‌ای از موتورسواران مورد نظر از کلاه ایمنی و همچنین به دلیل نحوه تصویر برداری از روبرو، امکان شناسایی موتورسیکلت از طریق کلاه ایمنی و یا پلاک وجود ندارد. در این مقاله روشی برای تشخیص و شناسایی فاصله موتورسیکلت با پوشش بادگیر در تصاویر با استفاده از یک دوربین

شود (Wang et al., 2015; Jamaluddin and Areni, 2019; Pineda-Deom, 2019). به طور کلی این دسته از سنجنده‌ها به دو دسته‌ی سنجنده‌های فعال^۱ و غیر فعال^۲ تقسیم می‌شوند. روش‌ها و مطالعات بسیاری برای بررسی وضعیت مسیر با استفاده از سنجنده‌های فعالی همچون رادار و لیزر (Dabbour and Easa, 2014; GONG et al., 2019; Gruyer and Rahal, 2019)، سنجنده‌های غیر فعال همچون سنجنده‌های مادون قرمز (Borges and Vidas, 2016; Thakur, 2018) و همچنین دوربین‌های اخذ جفت تصاویر و یا تک تصویر (Sivaraman and Trivedi, 2013; Deigmoeller et al., 2018; Tang et al., 2018; Damian et al., 2019) صورت گرفته است.

سیستم‌های راداری جزو اولین سیستم‌هایی بودند که برای تشخیص مانع بر روی اتومبیل‌ها توسط شرکت‌های خودروسازی استفاده شدند (Van Ratingen et al., 2016). در حوزه‌ی سیستم‌های راداری، برای کمک به افزایش ایمنی سفر در اتومبیل‌های بدون سرنشین از سنجنده‌های راداری برد کوتاه برای تشخیص موانع مختلفی همچون سایر اتومبیل‌ها و موتورسیکلت‌های موجود در مسیر روبرو (Liu et al., 2018) و یا نقاط کور اتومبیل (Fernández et al., 2013; Pineda-Deom, 2019) مورد استفاده قرار گرفتند. در مجموع سیستم‌های راداری به دلیل باریک بودن میدان دید نسبت به عوارض موجود در مسیر خطی جلو و یا پشت‌سر، در تقاطع‌هایی با ترافیک بالا عملکرد خوبی در دنبال کردن وسایل نقلیه موجود در جاده را نداشته‌اند (Sivaraman and Trivedi, 2013). درحالی که مسئله‌ی میدان دید در سنجنده‌های لیزری تا حد قابل قبولی بهبود یافت. از سیستم‌های لیزری به صورت گسترده در تشخیص و هشدار نسبت به هرگونه مانع برای اتومبیل‌های بدون سرنشین با استفاده از نصب تک لیزر (GONG et al., 2019; Gruyer and Rahal, 2019) و یا

نصب شده بر روی آینه‌بغل سمت کمک‌راننده با دید به پشت سر ارائه شده است. بیشتر این تصادفات بر اساس حرکت‌های نامناسب موتورسیکلت‌ها در نقاط کور راننده‌ی اتومبیل اتفاق می‌افتد (کربلایی et al., 2019). این کنترل و نظارت کمک می‌کند تا به طور خاص در مواقعی که موتورسیکلت بیش از حد مجاز به اتومبیل نزدیک شده است و یا راننده تمایل به تغییر مسیر حرکت خود در جاده یا خیابان‌ها را دارد، با هشدار به موقع از وقوع تصادف جلوگیری کند.

در ادامه مقاله، در بخش دوم به بررسی تحقیقات انجام شده قبلی در موضوع مورد نظر پرداخته شده است. در بخش سوم به بیان نوع پایگاه داده‌ی تهیه شده و همچنین روش تحقیق و مدل‌های یادگیری عمیق مورد استفاده در دو مرحله‌ی شناسایی و تخمین فاصله اشاره شده است. در بخش چهارم هم نتایج پیاده سازی و ارزیابی مدل پیشنهادی ارائه شده است. در آخر نیز به نتیجه‌گیری و ارائه‌ی پیشنهادهایی برای ارتقای هدف مورد نظر پرداخته شده است.

۲. پیشینه تحقیق

برای کاهش حجم تصادفات و خسارت‌های مالی و جانی در مسیرهای درون شهری و برون شهری، تحقیقات و مطالعات گسترده‌ای صورت گرفته و در راستای تحقق این امر روش‌های متنوعی برای تقابل با این‌گونه تصادفات طراحی شده است (Dabbour and Easa, 2014; Anaya et al., 2017; Markiewicz et al., 2017; Salahat et al., 2017; Yang et al., 2018; Jamaluddin and Areni, 2019; Pineda-Deom, 2019). برای کمک به کاهش تصادفات سیستم‌های حمل‌ونقل در دهه‌های اخیر سعی شده تا با استفاده از سنجنده‌های مختلف و با طراحی سیستم‌هایی پیشرفته‌ی کمک رسان راننده با نام *Advanced driver assistance system* (ADAS) از بروز حوادث غیرقابل جبران و خطرناک جلوگیری

(et al., 2017). در این مقاله با توجه به استفاده از داده‌های اخذ شده توسط تک دوربین، در ادامه به بررسی عملکرد دوربین‌های رقومی برای شناسایی و برآورد فاصله آنی جهت جلوگیری نسبت به وقوع تصادف اتومبیل‌ها با موتورسیکلت‌های موجود در مسیر پرداخته شده است.

۱-۲ تشخیص و شناسایی نوع وسیله نقلیه با استفاده از

دوربین رقومی

اکثر سیستم‌های هشدار طراحی شده برای پایش و کنترل تصادفات بر مبنای آگاهی نسبت به اتومبیل‌های موجود در اطراف و یا بر مبنای هشدار درباره‌ی کم شدن فاصله‌ی اتومبیل با اتومبیل‌های جلویی (Kim et al., 2016; Sochor et al., 2018) و یا موتورسیکلت‌های روبرو (Damian et al., 2019; Jamaluddin and Areni, 2019) صورت گرفته است. به همین خاطر پژوهش‌های کمتری بر روی استفاده از تک دوربین با دید به عقب، برای کنترل وضعیت وسایل نقلیه در حال نزدیک شدن از پشت سر انجام شده است (Espinosa et al., 2018b; Jiang et al., 2018). تاکنون بیشتر روش‌های تشخیص موتورسیکلت‌های با استفاده از الگوریتم SVM (Chiverton, 2012)، بر اساس استخراج تصاویر چرخ‌ها (Rogers and Papanikolopoulos, 2000) و یا استفاده از الگوریتم‌های طبقه‌بندی کننده‌ی آبخاری (Chen and Liu, 2016) بوده است. طبق نتایج منتشر شده از مدت زمان اجرای این روش‌ها، در اکثر آنها تشخیص ماشین یا موتورسیکلت، به دلیل محاسبات مرحله‌ی مربوط به کلاسه‌بندی به روش SVM زمان‌بر است (Chiverton, 2012) و در نتیجه قابلیت اجرا به عنوان الگوریتم-های آنی را ندارند.

در سالهای اخیر سعی شده تا با ارائه‌ی ایده‌های نوآورانه عملیات تشخیص وسایل نقلیه همچون موتورسیکلت به صورت آنی دنبال

نصب و بکارگیری ۴ لیزر همزمان برای برآورد ابرنقاط اطراف اتومبیل و کلاسه‌بندی موانعی از جمله اتومبیل‌ها، موتورسیکلت‌ها، دوچرخه‌ها و اتوبوس‌ها صورت گرفت (Anaya et al., 2017). داده‌های مربوط به سنجنده‌های راداری و لیزری در اکثر مواقع به دلیل بالا بودن حجم نويز نیازمند اعمال فیلترهای رفع نويز هستند که همین امر اطلاعات مربوط به داده‌ها را دستخوش تغییر می‌کند (Sivaraman and Trivedi, 2013). از طرفی سیستم‌های راداری و یا لیزری جهت طبقه‌بندی عوارض شناسایی شده بر اساس ابعاد و نحوه حرکت آنها عمل می‌کنند (Sivaraman and Trivedi, 2013). به همین خاطر با توجه به محدوده‌ی وسیعی از انواع اشکال و ابعاد متنوع موجود در موتورسیکلت‌های مورد استفاده در سراسر دنیا، برای تشخیص و طبقه‌بندی هر یک از این نوع موتورسیکلت‌ها نیاز به استفاده از الگوریتم‌های مجزا و منحصر به فردی است. تاکنون مطالعه‌ای در زمینه‌ی استفاده از سنجنده‌های راداری یا لیزری برای هشدار نسبت به حضور و فاصله‌ی کم موتورسیکلت‌های با پوشش و ظاهر خاص همچون پوشش مورد بحث با بادگیر مشکی رنگ مورد استفاده در شهرهای ایران صورت نگرفته است.

در سالهای اخیر سنجنده‌های دیگری همچون دوربین‌های رقومی به دلیل قیمت و هزینه‌ی مناسب آنها و همچنین امکان استفاده ساده و راحت از آنها برای اکثر کاربران نسبت به سایر سنجنده‌های همچون رادار و لیزر، به شدت مورد توجه قرار گرفته است (Deigmoeller et al., 2018; Sochor et al., 2018; Tang et al., 2018; Damian et al., 2019). تصاویر بصری به خوبی قابلیت ایجاد تمایز بین پدیده‌های مختلف و کلاسه‌بندی آنها در گروه‌های مربوطه را دارند. این دلایل باعث شده‌اند تا در پروژه‌های مربوط به حوزه‌ی ADAS استفاده از دوربین‌های دیجیتال از اهمیت بسیار بالایی برخوردار شود (Markiewicz

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق

شده‌اند (Girshick et al., 2014; Shine and Jiji, 2020). با وجود دقت مناسب این روش، مسئله محدودیت منابع محاسباتی موجب چالش برانگیز شدن استفاده از مدل‌های CNN در حل مسائل شده است. در کنار R-CNN شبکه‌های عصبی ارتقا یافته‌ی Faster R-CNN (Girshick, 2015) و Fast R-CNN (Ren et al., 2015) در راستای بهبود سرعت و نتایج در شناسایی اشیاء و عوارض موجود در تصویر ارائه شده‌اند. برای کمک به سیستم هوشمند حمل و نقل، از این شبکه‌ها بیشتر برای شناسایی و طبقه‌بندی موتورسیکلت و شناسایی کلاه ایمنی در تصاویر دوربین‌های ثابت (Vishnu et al., 2017; Espinosa et al., 2018b) استفاده شده است.

در روش‌های تک‌مرحله‌ای برای افزایش سرعت و دقت، از همان ابتدا سعی بر این است تا الگوریتم تصویر ورودی را فقط یک بار ببیند و سپس به صورت موازی با استفاده از ترکیب شبکه‌های عمیق پیچشی و تمام متصل ویژگی‌های لازم از تصویر را استخراج و کلاسه‌بندی را انجام دهد. از جمله این دسته از الگوریتم‌ها می‌توان به مدل‌های (YOLO) (You only look once) (Redmon et al., 2016) و (SSD) (Single Shot Detector) (Liu et al., 2016) را نام برد. در این الگوریتم‌ها برخلاف تکنیک پنجره‌های لغزان و ناحیه‌های کاندید، الگوریتم تصویر ورودی را به صورت کلی یا به اصطلاح یک بار دیده و با استفاده از ترکیب شبکه‌های عمیق پیچشی و تمام متصل به صورت موازی ویژگی‌های لازم از تصویر را استخراج می‌کند. این الگوریتم‌ها از لحاظ دقت و سرعت پیشرفت بسیاری داشته‌اند و قابلیت موازی سازی بسیار مناسبی بر روی پردازنده‌هایی مثل GPU^۳ و FPGA^۴ را دارند.

در مجموع روش‌های دو-مرحله‌ای همچون شبکه‌های R-CNN، Fast R-CNN و Faster R-CNN هر یک از اجزای شبکه برای

شوند. در راس آنها مدل‌های شبکه‌های عصبی پیچشی Convolutional Neural Network (CNN) جهت تشخیص و شناسایی موتورسیکلت با استفاده از استخراج ویژگی‌های مربوط به موتورسیکلت (Espinosa et al., 2017a) و یا بر اساس شناسایی کلاه ایمنی مورد استفاده قرار گرفت (Vishnu et al., 2017). شبکه‌های CNN با معماری خاصی که دارند، با کاهش تعداد پارامترها، عملکرد الگوریتم بازگشتی مربوط به فرایند آموزش را ارتقا می‌دهند (Espinosa et al., 2018a). الگوریتم‌های CNN بسیاری در زمینه‌ی تشخیص اشیاء و تعیین موقعیت و کلاسه‌بندی نوع عارضه صورت گرفته‌اند که به طور کلی به دو دسته‌ی روش‌های دو-مرحله‌ای و تک-مرحله‌ای تقسیم می‌شوند (Chen et al., 2019). در روش‌های دو-مرحله‌ای ابتدا منطقه‌ای که امکان حضور شی در آن وجود دارد شناسایی می‌شود و سپس در مرحله‌ی دوم کلاس شی موجود در منطقه‌ی مورد نظر تشخیص و تعیین می‌شود. در روش‌های تک‌مرحله‌ای انجام عملیات تعیین موقعیت و برآورد کلاس عارضه در یک مرحله و به صورت همزمان صورت می‌گیرد. روش‌های تک-مرحله‌ای از سرعت بالاتری نسبت به روش‌های دو-مرحله‌ای برخوردارند.

یکی از این روش‌های دو-مرحله‌ای پرکاربرد در زمینه‌ی سیستم‌های کمکی راننده، مدل Region Based Convolutional Neural Network (R-CNN) است که مبتنی بر ناحیه‌های کاندید هستند. این مدل‌ها یک الگوریتم شناسایی شی هستند که ابتدا تصویر را برای یافتن کادرهای محصورکننده به صورت بالقوه قطعه‌بندی می‌کند و سپس الگوریتم شناسایی را برای یافتن محتمل‌ترین اشیاء در این کادرهای محصورکننده اجرا می‌کند. در مطالعه‌ای خاص بر روی موتورسیکلت‌های کشور هند، عوارض در حال حرکت از پس زمینه ثابت استخراج و موتورسیکلت‌هایی که از کلاه ایمنی استفاده نکرده‌اند با اجرای شبکه‌ی R-CNN تعیین

در راستای کنترل و پایش موتورسواران از طریق دوربین‌های ثابت راهنمایی و رانندگی با استفاده از مدل YOLOV2 (Redmon and Farhadi, 2017) به تشخیص پلاک نصب شده در پشت موتورسیکلت‌ها و بررسی استفاده یا عدم استفاده موتورسوار از کلاه ایمنی پرداخته شده است (Mistry et al., 2017a; Laroca et al., 2018). در سال ۲۰۱۹ از مدل ارتقا یافته‌ی YOLOV3 به عنوان بهترین الگوریتم جهت جداسازی و تمایز بهتر بین شناسایی عابران و موتورسواران در حال حرکت در مسیر تونل‌ها استفاده شد (Mun et al., 2019). در سال ۲۰۲۰ مدل بهبود یافته‌ی جدیدی با نام YOLOV4 (Bochkovski et al., 2020) به شناسایی و تشخیص اتومبیل‌ها، کامیون‌ها و موتورسیکلت‌های موجود در جاده با میانگین دقت ۶۷ درصد و سرعت ۳۸ فریم بر ثانیه در تصویر دست یافت (Mahto et al., 2020). حدود دو ماه پس از معرفی YOLOV4، الگوریتم جدیدی تحت نام YOLOV5 به صورت متن باز منتشر شده که با دقت مشابه با YOLOV4 عملکرد بسیار سریع‌تری در تشخیص اشیا با سرعت حدود ۱۴۰ فریم بر ثانیه را داشته است (Glenn Jocher, 2020). از آنجا که این مدل توسط نویسندگان اصلی نسخه‌های YOLO نوشته نشده و تا کنون نتایج رسمی از این مدل در مقالات معتبر جهانی منتشر نشده است، همچنان مدل YOLOV4 به عنوان آخرین نسخه‌ی رسمی ارائه شده از خانواده‌ی الگوریتم‌های YOLO محسوب می‌شود. به همین خاطر در این مقاله نتایج مدل YOLOV5 مورد بررسی قرار نگرفته است.

به دلیل وجود طیف گسترده‌ای از موتورسیکلت‌ها با ابعاد و ظواهر مختلف، در هنگام تصویر برداری از روبرو لازم است تا الگوریتم، با توجه به پارامترها و ویژگی‌های ظاهری مربوط به نوع موتورسیکلت مورد نظر آموزش داده شده باشد. در بیشتر مطالعات

استخراج ناحیه و محاسبه‌ی ویژگی‌ها باید به صورت جداگانه آموزش داده شوند که به دلیل حجم بالای پردازش‌ها و محاسبات سرعت پایین‌تری دارند و بهینه‌سازی آن‌ها با مشکل بیشتری مواجه است. مقایسه‌ی نتایج بدست آمده از الگوریتم‌های تک_مرحله‌ای با الگوریتم دو_مرحله‌ای R-CNN نشان داده که با وجود عملکرد مشابه این دو گروه در دقت عوارض شناسایی شده، سرعت اجرای الگوریتم‌های تک_مرحله‌ای بالاتر بوده است (Chen et al., 2019). به همین خاطر استفاده از الگوریتم‌های تک_مرحله‌ای انتخاب مناسب‌تری برای شناسایی آبی عوارض در هنگام رانندگی اتومبیل خواهد بود.

مدل‌های SSD مدل‌های بسیار سریعی هستند که با اجرای تنها یک بار از یک شبکه‌ی پیچشی بر روی تصویر ورودی، از آن نقشه‌های ویژگی تولید می‌کنند. در این مدل‌ها با وجود کاهش در عملکرد دقت، سرعت اجرا نسبت به شبکه‌های Faster R-CNN بهتر بوده است (Liu et al., 2016). بررسی مدل‌های SSD و Faster R-CNN در مقایسه با نسخه‌های جدیدی از الگوریتم YOLO در شناسایی عوارض جاده همچون اتومبیل‌ها، موتورسیکلت‌ها و عابران بر روی پایگاه داده‌ی VOC2012 (Everingham et al., 2010) نشان داده که عملکرد YOLOV2 در کنار دقت مشابه، به افزایش سرعت ۶ و ۲ برابر نسبت به مدل‌های Faster R-CNN و SSD دست یافته است (Redmon and Farhadi, 2017). علاوه بر این نتایج آموزش الگوریتم YOLO بر اساس پایگاه داده‌ی Cityscape (Cordts et al., 2016) نشان داده که بدلیل ابعاد کوچک و عدم وجود تصاویر با کیفیت از کلاس موتورسیکلت در پایگاه داده‌ی مورد نظر، تشخیص این دسته از وسایل نقلیه نسبت به سایر ۳ نوع وسیله‌ی دیگر با چالش بیشتری همراه بوده است (Bhujbal and Mane, 2019).

(Mistry et al., 2017a; Vishnu et al., 2017; Laroca et al., 2018) صورت گرفته‌اند. جهت اندازه‌گیری فاصله عارضه در تصویر در روش‌های سنتی با استفاده از حداقل چهار نقطه با مختصات زمینی معلوم در پشت زمینه (YG et al., 2015)، ابعاد مشخص طول و عرض خط‌کشی‌های عابر پیاده (Luvizon et al., 2016)، استفاده از ابعاد معلوم پلاک‌های راهنمایی نصب شده در پشت سر اتومبیل‌ها و موتورسیکلت‌ها (برومندنیا and غلامی، ۲۰۱۹) برای انتقال به سیستم مختصات پرسپکتیو معکوس و اعمال شبکه‌ی مقیاس‌بندی شده بر روی محدوده مورد نظر استفاده شده است (Wang et al., 2015; Wu et al., 2015; Tang et al., 2018).

اندازه‌گیری این دسته از نقاط و ابعاد معلوم نیازمند برداشت‌های زمینی در موقعیت تصویربرداری است. از آنجا که این سیستم‌ها و الگوریتم‌های مورد استفاده برای هشدار آنی به رانندگان می‌بایست بر روی دوربین‌های نصب شده بر روی اتومبیل‌های در حال حرکت طراحی و اجرا شوند، بنابراین در حالتی که اتومبیل‌ها بر روی یک سطح هم تراز نباشند، این عملیات محاسبه‌ی ماتریس هموگرافی و اندازه‌گیری فاصله دچار خطای قابل توجهی خواهد شد.

در روش‌های ماشین بینایی و شبکه‌های عصبی درک شکل و وضعیت یک صحنه از طریق تک تصویر همواره یک چالش اساسی محسوب می‌شود (Godard et al., 2017). تاکنون از این روش‌ها جهت تولید نقشه‌ی عمق در کاربردها و فعالیت‌های متنوعی همچون علم رباتیک (Lenz et al., 2015)، تعیین موقعیت و تقلید از نحوه‌ی حرکت اعضای بدن انسان (Shotton et al., 2018; Zhou and Tuzel, 2011) و تبدیل خودکار تصاویر دوبعدی به سه بعدی فیلم‌ها (Xie et al., 2016) استفاده شده است. در زمینه‌ی اتومبیل‌های بدون سرنشین نیز از آموزش شبکه بر اساس ویدیوهای پیوسته‌ی اخذ شده از نحوه حرکت صحنه‌های

صورت گرفته شناسایی موتورسیکلت در اجرا در دو فاز پشت سر هم (سرعت اجرای پایین‌تر) انتخاب نوع کلاس «انسان» و همچنین «کلاه ایمنی» صورت گرفته است (Mistry et al., 2017b; Vishnu et al., 2017). تاکنون مطالعه متمرکزی برای شناسایی موتورسیکلت‌های با پوشش ظاهری متفاوت همچون نصب بادگیرهای مشکی موردنظر بدون نیاز به تشخیص از روی پلاک، کلاه ایمنی و یا راننده‌ی موتورسیکلت انجام نشده است. در صورتی که طبق اعلام پلیس راهور تهران، امروزه حجم زیادی از موتورسیکلت‌های در حال تردد از این پوشش مخصوص در طول سال استفاده می‌کنند (عصرخودرو، ۲۰۲۰).

استفاده از شبکه‌های تک_مرحله‌ای همچون الگوریتم YOLO به دلیل دقت قابل قبول، حجم کمتر محاسبات و قابلیت اجرا به صورت آنی، پاسخگوی مناسبی برای شناسایی و تشخیص آنی موتورسیکلت‌های مورد نظر در این مقاله خواهد بود. در این مقاله در راستای هدف مورد نظر، بررسی بر روی عملکرد سرعت و دقت دو نسخه‌ی آخر رسمی منتشر شده از این خانواده YOLO صورت گرفته تا بررسی شود کدام یک از این الگوریتم‌ها می‌تواند به عنوان بهترین روش تشخیص موتورسیکلت‌های مورد نظر استفاده قرار گیرند.

۲-۲ محاسبه‌ی و اندازه‌گیری فاصله با استفاده از

دوربین رقومی

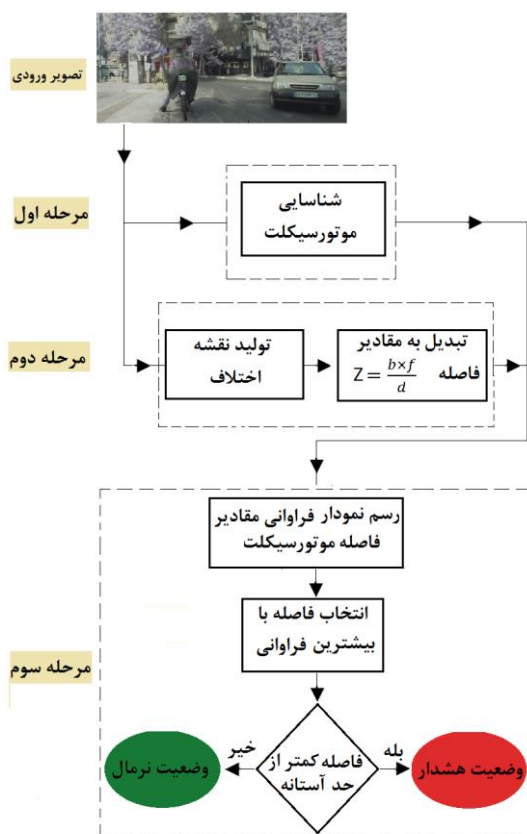
همانطور که گفته شد مبحث بررسی وسایل در حال سبقت در مسیر شامل دو مرحله‌ی اصلی تشخیص وسیله و سپس برآورد فاصله‌ی وسیله است. بیشتر این مطالعات بر روی پایش نحوه‌ی رانندگی وسایل نقلیه به ویژه اتومبیل‌های سواری (Wang et al., 2016; Chabot et al., 2017) و کنترل و تعیین موقعیت این دسته از وسایل، از تصاویر دوربین‌های ثابت نصب شده بر روی نقاط مختلف شهر همچون پل‌های هوایی و یا چراغ‌های راهنمایی

در سال ۲۰۱۹ از این مدل برای تخمین فاصله‌ی نسبی عابران پیاده و اتومبیل‌های موجود در تصویر استفاده شده است (Chen et al., 2019).

در بخش دوم روش پیشنهادی با توجه به هدف مقاله به بررسی این موضوع پرداخته شد که با چه دقت و سرعتی می‌توان از الگوریتم Monodepth برای تخمین نقشه عمق و فاصله پیکسل‌های تصویر تا دوربین استفاده کرد.

۳. روش پیشنهادی

روش ارائه شده در این مقاله برای هشدار به راننده در هنگام نزدیک شدن موتورسیکلت از پشت سر به اتومبیل بر پایه‌ی مدل‌های شبکه‌های عصبی و یادگیری عمیق بوده است که در سه مرحله‌ی اصلی طبق فلوچارت نمایش داده شده در شکل ۲ انجام می‌شود.



شکل ۲. فلوچارت عملکرد الگوریتم هشدار پیشنهادی

اطراف اتومبیل جهت برآورد نقشه عمق تک تصویر رنگی استفاده شده است (Jiang et al., 2018). مطالعه بر روی مدل‌های یادگیری عمیق نشان داده است که در صورت آموزش این مدل‌ها با استفاده از پایگاه داده‌ای غنی با استفاده از دوربین نصب شده بر روی جلوی اتومبیل، می‌توان به تولید نقشه‌ی عمق رسید و موقعیت نسبی اشیاء در تصویر را تعیین کرد (Dijk and Croon, 2019). در سالهای اخیر تلاش‌های بسیار زیادی در زمینه‌ی تولید نقشه‌ی عمق از تک تصویر با استفاده از شبکه‌های عصبی عمیق صورت گرفته (Eigen et al., 2014; Liu et al., 2015; Godard et al., 2017). روش‌های یادگیری عمیقی که به استخراج عمق بر اساس تک تصویر می‌پردازند، به شکل کلی به دو گروه روش-های یادگیری نظارت شده^۵ و نظارت نشده^۶ تقسیم شده‌اند (Godard et al., 2017).

در روش‌های بر مبنای آموزش نظارت شده (Eigen et al., 2014)، شبکه بر اساس پایگاه داده‌ی بزرگی از مجموعه تصاویر به همراه داده‌های معلوم از اندازه‌گیری‌های واقعی زمینی عمق آموزش داده می‌شود. با استفاده از پارامترهای وزن آموزش دیده به تخمین عمق هر پیکسل از تصویر ورودی پرداخته می‌شود. در حالی که این مدل‌ها دقت قابل قبولی در برآورد عمق دارند، اما مسئله‌ی تهیه‌ی یک پایگاه داده‌ی غنی به همراه مقادیر فاصله یا نقشه‌های عمق معلوم به ازای هر تصویر چالش بسیار بزرگی است. در مقابل، روش‌های یادگیری نظارت نشده برای برآورد عمق، در مرحله آموزش نیازی به داده‌های زمینی عمق ندارند (Garg et al., 2016). به تازگی مدلی با نام Monodepth در زمینه‌ی تولید نقشه‌ی عمق از تک تصویر بدون استفاده از اندازه‌گیری‌های زمینی عمق ارائه شده است (Godard et al., 2017) که با استفاده از جفت تصاویر آموزش داده می‌شود و سپس با اجرا روی تک تصویر نقشه‌ی اختلاف^۷ آن تصویر به عنوان خروجی نمایش داده می‌شود.

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق

دو برابر نسبت به دیگر سیستم‌های تشخیص اشیای دست یافته است (Redmon et al., 2016). در راستای بهبود این الگوریتم، نسخه‌های دیگری تحت عنوان YOLOV2، YOLOV3 و YOLOV4 برای تشخیص آنی اشیاء در تصویر به صورت رسمی ارائه شده است.

۳-۱-۱ ساختار YOLOV2 و YOLOV3

مدل YOLOV2 با نام دیگر Darknet-19 نیز معرفی شده است که عدد ۱۹ در آن بیانگر ۱۹ لایه‌ی پیش‌بینی موجود در ساختار شبکه است (Redmon and Farhadi, 2017). در این مدل تغییراتی برای بهبود عملکرد و سرعت اجرا صورت گرفته که ۱۷ درصد معیار دقت نسبت به نسخه قبلی افزایش یافته است. از جمله‌ی این تغییرات اعمال شده شامل بالا بردن توان تفکیک، نرمالسازی داده‌های ورودی هر لایه و استفاده از کادرهای محصور کننده با ابعاد مختلف است. در مدل YOLOV3 از ۵۳ لایه‌ی پیش‌بینی برای استخراج ویژگی در شبکه استفاده شده است (Redmon and Farhadi, 2018). در کنار این افزایش عمق، برخی تغییرات نیز در ساختار شبکه اعمال شده است که نسبت به نسخه قبلی موجب افزایش ۲۰ درصد میانگین دقت این مدل بر روی پایگاه داده‌ی COCO (Lin et al., 2014) شده است (Redmon and Farhadi, 2018). این تغییرات شامل در نظر گرفتن و اضافه کردن پارامترهایی همچون مقدار احتمال حضور شی در کادر پیش‌بینی شده از شی، اتصالات جدید به لایه‌های شبکه و انجام پیش‌بینی در سه مرحله جداگانه برای ارتقای الگوریتم در راستای شناسایی عوارض کوچکتر در تصویر بوده است.

۳-۱-۲ ساختار YOLOV4

YOLOV4 آخرین نسخه رسمی منتشر شده از سری مدل‌های YOLO است که در سال ۲۰۲۰ در محیط Darknet نوشته شده است (Bochkovski et al., 2020). در راستای بهبود

در مرحله‌ی اول با توجه به تک تصویر رنگی اخذ شده توسط دوربین نصب شده بر روی آینه‌ی سمت کمک راننده، موتورسیکلت موجود در نقطه‌ی کور پشت سر تشخیص داده می‌شود و مختصات موقعیت آن در تصویر تعیین می‌شود. در مرحله‌ی دوم نقشه اختلاف تصویر تولید می‌شود و فاصله‌ی پیکسل‌های تصویر تا دوربین محاسبه می‌شوند. در مرحله سوم با تلفیق نتایج حاصل از دو مرحله‌ی قبل با توجه به فراوانی مقادیر فاصله‌ی بدست آمده از پیکسل‌های مربوط به موتورسیکلت، فاصله‌ی موتورسیکلت تا دوربین مشخص می‌شود. در آخر در صورتی که این فاصله کمتر از حد مجاز انتخابی باشد فرایند هشدار به راننده صورت خواهد گرفت.

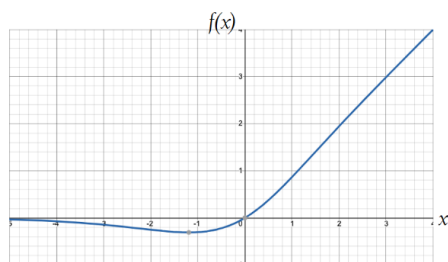
۳-۱-۳ مرحله اول: تشخیص موتورسیکلت

در مرحله‌ی اول، با مقایسه‌ی مدل‌های مختلف موجود از الگوریتم‌های YOLO، مناسب‌ترین مدل برای هدف مورد نظر انتخاب و آموزش داده شده است. هدف از آموزش مدل مورد نظر این است که می‌توان پس از اخذ تصویر و اعمال آن به عنوان ورودی شبکه‌ی پیش‌بینی، موقعیت و کلاس اشیای خاص مورد نظر در تصویر را با دقت بهتر مشخص کرد.

YOLO به معنای «شما فقط یک بار به تصویر نگاه می‌کنید» است. در واقع، این عبارت به همان قابلیت سیستم بینایی انسان اشاره دارد که با یک نگاه عمل تشخیص اشیاء را انجام می‌دهد. این الگوریتم به مساله تشخیص اشیاء به صورت پیاده‌سازی همزمان استخراج نقشه‌های ویژگی توسط شبکه پیش‌بینی و اعمال رگرسیون برای تشخیص عوارض می‌نگرد که مستقیم از پیکسل‌های تصویر به مختصات کادر شی و احتمال کلاس آن می‌رسد (Redmon et al., 2016). YOLO اولین شبکه‌ی تشخیص عارضه است که دو چالش رسم کادر شی^۱ و انتخاب نوع کلاس عارضه را در یک شبکه ادغام کرده و به Mean Average Precision (mAP)

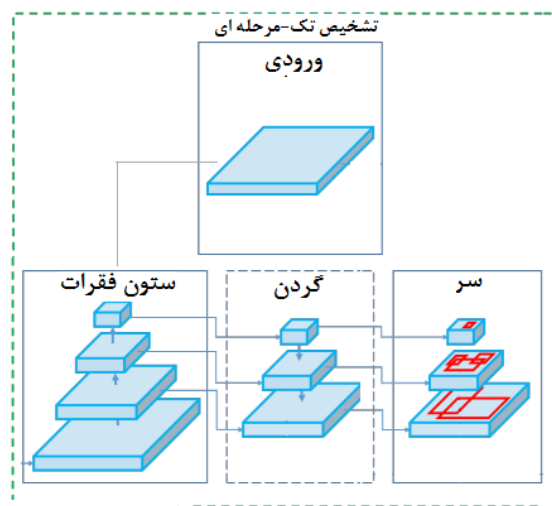
فصلنامه مهندسی حمل و نقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱

Complete union (GIoU) به عنوان تابع هزینه، از پارامتر $\text{Bochkovski et al., 2020}$ (CIoU) استفاده شده است. در این تابع علاوه بر در نظر گرفتن میزان اشتراک، به فاصله‌ی بین مراکز و نسبت ابعاد کادر محاسبه شده با مقدار کادر واقعی نیز اهمیت داده شده است. علاوه بر این در این مدل برای نرمالسازی داده‌ها از مدل بهبود یافته $\text{Cross mini-Batch Normalization (CmBN)}$ استفاده شده (Bochkovski et al., 2020) تا مدل قابلیت اجرا بر روی سیستم‌های تک GPU را نیز داشته باشد. با توجه به این تصحیحات و بهبودهای صورت گرفته در این مدل نتایج بدست آمده از ارزیابی بر روی پایگاه داده‌ی COCO نشان از افزایش ۱۰ درصد دقت و ۱۲ درصد سرعت اجرای مدل YOLOV4 نسبت به مدل قبلی YOLOV3 را داشته است (Bochkovski et al., 2020). علاوه بر تغییراتی که در ساختار شبکه و معماری شبکه رخ داده است، برای بهبود عملکرد و بهینه‌سازی نتایج از تابع فعالسازی Mish (Misra, 2019) به جای تابع Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) در مرحله‌ی آموزش شبکه استفاده شده است. تابع فعالسازی Mish در سال ۲۰۱۹ طبق رابطه‌ی ۱ به عنوان تابع فعالسازی بهینه در شبکه‌های یادگیری عمیق ارائه شد. با توجه به نمودار این معادله که در شکل ۴ رسم شده، این تابع فعالسازی یک تابع پیوسته با شیب ملایم است که در آن به روش منظم غیر یکنواختی، شبکه را به حالت بهینه‌ی خود نزدیک می‌کند.



شکل ۴. نمودار تابع Mish

دقت شبکه‌های عصبی پیچشی روش‌های زیادی وجود دارد که در نسخه‌ی چهارم YOLOV4 از بسیاری از آن‌ها استفاده شده است. به گونه‌ای که اجرای الگوریتم بر روی داده‌های COCO (Lin et al., 2014) حاکی از میانگین دقت ۶۵/۷ درصد و سرعت ۶۵ فریم بر ثانیه بوده است (Bochkovski et al., 2020). به صورت کلی، ساختار این شبکه که به شکل تک-مرحله‌ای به تشخیص عارضه‌ی مورد نظر در تصویر می‌پردازد همچون شکل ۳ است. به ترتیب ابتدا تصویر مورد نظر به عنوان ورودی وارد بخش ستون فقرات شبکه می‌شود تا نقشه‌های ویژگی مربوط به تصویر ورودی از آن استخراج شود. خود این قسمت ستون فقرات شامل دو بخش گردن و سر است. بخش گردن برای کمک به استخراج نقشه‌های ویژگی متمایز و بهتر از تصاویر در سطوح مختلف است که از بلوک‌های متراکم شامل مجموعه‌ای از لایه‌های پیچشی تشکیل شده‌اند که در بین آنها لایه‌های میانی نیز وجود دارند که در آن عملیات نرمالسازی و تجمیع‌سازی^۹ انجام می‌شود (Huang et al., 2017). بخش سر شبکه نیز پیش‌بینی نهایی را انجام می‌دهد.



شکل ۳. معماری شبکه‌ی تک مرحله‌ای YOLOV4

در راستای افزایش دقت و سرعت تشخیص کادر محصور نیز به جای محاسبه‌ی پارامتر Generalized Intersection over

زمینه‌ی برآورد عمق در تک تصویر داشته‌اند (Godard et al., 2017). اکثر این روش‌ها در برخورد با این مسئله به صورت رگرسیون نظارت شده برخورد می‌کنند که برای آموزش شبکه به مقادیر وسیعی از تصاویر به همراه داده‌های زمین مربوط به اطلاعات عمق واقعی آنها نیاز است (Eigen et al., 2014). ضبط داده‌های عمق در طیف وسیعی از محیط‌ها چالش بسیار بزرگ و سختی است. در سال ۲۰۱۷ روشی تحت عنوان Monodepth (Godard et al., 2017) برای تولید نقشه عمق از تک تصویر ارائه شد که در مرحله‌ی آموزش شبکه به جای استفاده از داده‌های معلوم عمق زمینی، از جفت تصاویر اخذ شده به عنوان داده‌های آموزشی استفاده می‌شود. این روش یک شبکه‌ی پیچشی ۷ لایه‌ای را ارائه داده است که بدون در نظر داشتن هندسه صحنه و نوع عوارض موجود، به برآورد نقشه عمق بر روی تک تصویر می‌پردازد.

نوآوری این روش در استفاده از قید مربوط به ارتباط بین نقشه‌های اختلاف (اختلاف منظر^۱ در راستای افق) تولید شده از جفت تصاویر چپ و راست است. نحوه عملکرد این الگوریتم به این صورت است که طبق شکل ۵ با استفاده از جفت تصویر، ابتدا نقشه‌های اختلاف تصاویر سمت چپ d^l و راست d^r تولید می‌شوند و سپس با استفاده از اطلاعات موجود از تصویر چپ I^l و نقشه اختلاف چپ به بازسازی تصویر راست \hat{I}^r پرداخته می‌شود و همچنین با استفاده از اطلاعات موجود از تصویر راست I^r و نقشه اختلاف راست، تصویر سمت چپ \hat{I}^l بازسازی می‌شود. با توجه به توابع هزینه‌ی تعریف شده در این شبکه، پارامترهای وزن شبکه در طول روند آموزش به گونه‌ای محاسبه می‌شوند که تصاویر چپ و راست بازسازی شده بیشترین شباهت و نزدیکی را به جفت تصاویر چپ و راست اخذ شده داشته باشد.

در نمودار این تابع مقادیر کوچکی از اعداد منفی در قسمت غیرخطی نمودار تعریف شده‌اند که باعث شده این تابع بر خلاف توابعی همچون ReLU و Leaky-ReLU در مقادیر منفی نیز مشتق‌پذیر باقی بماند و در نتیجه بروزرسانی بر روی اکثر نرون‌های شبکه امکان‌پذیر شود. با توجه به عملکرد این تابع، استفاده از تابع Mish به جای ReLU باعث افزایش دقت و سرعت الگوریتم خواهد شد. نتایج بدست آمده از دقت نهایی بدست آمده از تابع فعالسازی جدید Mish بر روی شبکه‌ی ResNet-50 (He et al., 2016) افزایش دقت حدود ۱ درصد نسبت به ReLU و ۰/۵ درصد نسبت به Swish (Ramachandran et al., 2017) داشته است (Misra, 2019).

$$f(x) = x \tanh(\ln(1 + e^x)) \quad (1)$$

در این مقاله با توجه به مزایا و ویژگی‌های مطرح شده در استفاده از الگوریتم‌های YOLO در سیستم‌های آنی، مقایسه‌ای بر روی عملکرد مدل‌های آموزش دیده و بهینه‌سازی شده‌ی YOLOV3 و YOLOV4 برای تشخیص موتورسیکلت‌های مورد نظر صورت گرفته است. در نهایت نیز مناسب‌ترین مدل از نظر سرعت و دقت در برآورد هدف مورد نظر در شناسایی موتورسیکلت، انتخاب و به عنوان مدل مورد استفاده در مرحله‌ی اول روش پیشنهادی استفاده شده است.

۳-۲ مرحله دوم: تولید نقشه عمق از تک تصویر

در اکثر روش‌های سنتی مطرح شده برای برآورد مقیاس، از ابعاد واقعی یک عارضه‌ی خاص (به عنوان مثال ابعاد پلاک اتومبیل) در تصویر استفاده شده است (Wang et al., 2015; Wu et al., 2015; Tang et al., 2018). در این دسته از تصاویر اخذ شده در این مقاله، تصویری از پلاک موتورسیکلت و یا اطلاعاتی از ابعاد معلوم عوارضی همچون خط‌کشی خیابان‌ها وجود ندارد. در این میان روش‌های مبتنی بر یادگیری عمیق نتایج بسیار مناسبی در

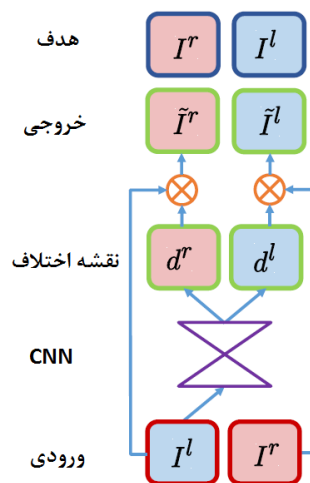
پارامترهای کالیبراسون دوربین آموزش می‌بینند، بنابراین برای رسیدن به نقشه‌ی عمق دقیق و درست بر روی تک تصویر، لازم است تا دوربین همان دوربینی باشد که با آن پایگاه داده‌ی مربوط به تصاویر آموزشی ساخته شده است. از آنجایی که پایگاه داده‌ی KITTI به نسبت شباهت زیادی با صحنه‌های تصویر برداری مورد نظر این مقاله دارد، بنابراین از مقادیر از قبل آموزش دیده‌ی این مدل (Godard et al., 2019) به عنوان پارامترهای وزن اولیه در آموزش شبکه بر روی پایگاه داده‌ی مخصوص تهیه شده استفاده شده است.

۳-۳ مرحله سوم: ادغام نتایج مراحل قبل

پس از اجرای دو مدل آموزش دیده در مراحل قبل، موقعیت موتورسیکلت و فاصله‌ی واقعی پیکسل‌ها در تصویر تعیین و محاسبه می‌شوند. در مرحله‌ی سوم لازم است تا با ادغام نتایج بدست آمده از دو مرحله‌ی قبل مقدار واحدی به عنوان طول فاصله‌ی موتورسیکلت تا دوربین تعیین شود تا در صورتی که این فاصله کمتر از حدآستانه‌ی انتخابی باشد به راننده در رابطه با وجود موتورسیکلت نزدیک شده هشدار داده شود.

هدف اصلی مقاله تنها مقادیر پیکسل‌های محدوده‌ی مربوط به کادر محصور در اطراف موتورسیکلت که ابعاد و مختصات آن توسط الگوریتم YOLOV4 صورت گرفته. به همین خاطر ادغام و تلفیقی بر روی نتایج بدست آمده از دو مرحله‌ی قبل انجام شده است. با توجه به این که نقاط تصویر برداری شده از جلوی موتورسیکلت شامل مجموعه نقاطی از چرخ تا قسمت شیشه‌ی بالایی بادگیر است، بنابراین نمی‌توان این نقاط ثبت شده از موتورسیکلت را در یک صفحه واحد و مسطح تعریف کرد. مقادیر متنوعی از مقدار فواصل بدست آمده بر اساس پیکسل‌های موجود در محدوده‌ی محصور موتورسیکلت وجود دارند. در سال ۲۰۱۹ برای برآورد مقدار عمق پیکسل‌های منتصب شده به کلاس عابران

فصلنامه مهندسی حمل و نقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱



شکل ۵. نمونه‌ی عملکرد آموزش شبکه‌ی Monodepth (Godard et al., 2017)

در این شبکه مقادیر نقشه اختلاف منظر تولید شده مقادیر نسبی بوده‌اند. برای رسیدن به مقادیر فاصله بر حسب واحد مشخص لازم است تا با استفاده از اطلاعات کالیبراسیون دوربین، مقدار نقشه اختلاف تصویر چپ طبق فرمول ۲ به نقشه عمق تبدیل شود (Hirschmuller, 2005). در این رابطه D فاصله تا دوربین، b طول بین دو دوربین اخذ جفت تصویر^{۱۱}، f فاصله کانونی و d مقدار اختلاف منظر است.

$$D = \frac{b \times f}{d} \quad (2)$$

در راستای افزایش دقت و سرعت آموزش این الگوریتم، نسخه جدیدی از مدل اشاره شده تحت عنوان Monodepth2 نیز در سال ۲۰۱۹ ارائه شده است (Godard et al., 2019). از جمله پارامترهای بهبود داده شده در این مدل می‌توان به تلاش به حفظ بهتر لبه‌ها در تصاویر عمق اشاره کرد. تا کنون این شبکه بر روی جفت تصاویر حاصل از پایگاه داده‌های متنوعی از جمله CityScape (Cordts et al., 2016)، ImageNet (Deng et al., 2009) و KITTI (Geiger et al., 2013) آموزش داده شده. با توجه به اینکه پارامترهای وزن موجود در شبکه طبق

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق

آستانه تعیین شده باشد، الگوریتم نسبت به حضور موتورسیکلت در نزدیکی اتومبیل هشدار داده می‌شود. در این مقاله طبق بررسی‌های تجربی مقدار فاصله‌ی ۳ متر به عنوان حد آستانه دوربین تا موتورسیکلت انتخاب شده است که با توجه به تغییر شرایط و نوع وسیله این مقدار انتخابی قابل تغییر است.

۴. پیاده‌سازی و ارزیابی نتایج

در این بخش بعد از بیان تصاویر و داده‌های مورد نیاز برای آموزش الگوریتم‌های مورد استفاده، نتایج پیاده‌سازی و ارزیابی صورت گرفته ارائه شده است. تمامی عملیات آموزش و آزمون بر روی سیستم عامل Ubuntu 18.04 با کارت گرافیکی GTX-1060 با حافظه دسترسی تصادفی ۱۶ گیگابایت و حافظه گرافیکی ۶ گیگابایت صورت گرفته است.

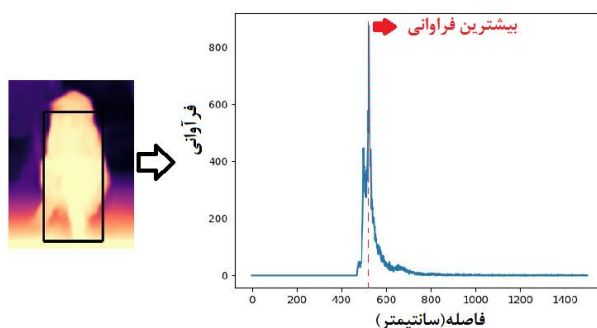
۴-۱ پایگاه‌های داده‌ی تهیه شده

از آنجایی که روش مورد استفاده در این مقاله بر اساس دو مدل مجزا در دو مرحله‌ی شناسایی و برآورد فاصله انجام می‌شود، به همین خاطر برای آموزش هر یک از مدل‌ها نیاز به یک پایگاه داده مجزا و متناسب است.

۴-۱-۱ پایگاه داده‌ی مربوط به مدل شناسایی موتورسیکلت

برای آموزش مناسب مدل YOLO مورد نظر از مجموع دو هزار تصویر اخذ شده از کلاس موتورسیکلت‌های مورد نظر استفاده شد. همه‌ی این تصاویر به صورت دستی و با دوربین آیفون ۸ از موتورسیکلت‌های در حال حرکت در خیابان‌های شهر تهران بدون نیاز به پیش پردازش در زوایا و فواصل مختلف در طول روز و شرایط آب و هوایی آفتابی و ابری اخذ شده است. در مجموع این تصاویر با نسبت ۱ به ۴ حدود برای مرحله‌ی آزمون و آموزش شبکه مورد استفاده قرار گرفتند. برای بهبود عملکرد مدل، سعی شد تا با استفاده از روش‌هایی همچون تغییر ابعاد و برش تصویر

پیاده در تصویر از روش نمودار فراوانی به عنوان بهینه‌ترین روش استفاده شد (Chen et al., 2019). در روش پیشنهادی این مقاله نیز برای انتخاب مقدار واحدی به عنوان فاصله‌ی موتورسیکلت از روش برآورد نمودار فراوانی مقادیر پیکسل‌های نقشه عمق موجود در کادر محصور موتورسیکلت و انتخاب فاصله‌ی با بیشترین فراوانی استفاده شده است. در این روش به ازای هر موتورسیکلت شناسایی شده طبق شکل ۶ نمودار هیستوگرام مربوط به مقادیر طول فواصل برآورد شده در کادر محدوده‌ی موتورسیکلت محاسبه می‌شود.



شکل ۶. نمایش نمودار فراوانی مربوط به مقادیر پیکسل‌های نقشه عمق مربوط به محدوده‌ی موتورسیکلت شناسایی شده در تصویر توسط روش پیشنهادی

فاصله‌ی با بیشترین مقدار فراوانی به عنوان طول فاصله‌ی موتورسیکلت تا دوربین انتخاب می‌شود. در این هیستوگرام مقادیر محور افقی بیانگر طول فواصل محاسبه شده بر حسب سانتی‌متر و محور عمودی بیانگر میزان فراوانی مقادیر است. در نهایت با استفاده از این نمودار می‌توان فاصله‌ی محاسبه شده با بیشترین میزان فراوانی در محدوده‌ی کادر موتورسیکلت را به عنوان مقدار بهینه‌ی طول فاصله موتورسیکلت تا دوربین انتخاب کرد.

با تلفیق نتایج مرحله‌ی اول و دوم الگوریتم و محاسبه‌ی طول فاصله‌ی موتورسیکلت تا دوربین، در صورتی که طبق این روند محاسبات، این فاصله محاسبه شده از موتورسیکلت کمتر از حد

۷ موقعیت مرکز کادر موتورسیکلت در تصویر، عرض تصویر و ارتفاع تصویر بوده است.

۴-۱-۲ پایگاه داده‌ی مربوط به مدل برآورد فاصله

موتورسیکلت

مدل **Monodepth2** به عنوان یک شبکه‌ی نظارت نشده با استفاده از مجموع جفت تصاویر موجود در پایگاه داده به همراه اطلاعات فایل کالیبراسیون دوربین مربوط به مقادیر طول پایه بین جفت دوربین، فاصله کانونی و مختصات مراکز تصویر آموزش داده می‌شود. این مدل یاد می‌گیرد که چگونه از تک تصویر رنگی ورودی نقشی اختلاف منظر آن را تولید کند. با توجه به مقادیر پارامترهای وزنی که در مرحله‌ی آموزش تعریف می‌شوند، لازم است تا برای رسیدن به برآورد دقیق از مقادیر فاصله، در مرحله آزمون نیز از همان دوربینی که در مرحله‌ی تهیه پایگاه داده‌ی آموزشی استفاده شده بود، استفاده شود. در این مقاله از دوربین **MYNT-EYE** که این یک دوربین با قابلیت تصویر برداری به صورت جفت تصاویر و همچنین تک تصویر است استفاده شده است. مشخصات فنی و برخی پارامترهای کالیبراسیون این دوربین در جدول ۱ آورده شده است. برای آموزش این شبکه، حدود شش هزار جفت تصویر ترمیم شده^{۱۴} از سطح خیابان‌های شهر تهران در حالت‌ها و موقعیت‌های مختلفی از حضور اتومبیل‌ها و موتورسیکلت‌ها در طول روز و شرایط نوردهی مناسب به صورت دستی توسط این دوربین اخذ شد.

جدول ۱. ویژگی و مشخصات دوربین **MYNT-EYE** مورد استفاده

مدل دوربین	مشخصات فنی	مولفه
	توان تفکیک	۴۸۰ × ۶۴۰
	اندازه پیکسل	۳/۷۵ × ۳/۷۵ میکرون
	فاصله Baseline	۱۲۰ میلی‌متر
	فاصله کانونی	۲/۴۵ میلی‌متر

مجموعه‌ای از تصاویر با ابعاد متنوع در پایگاه داده‌ی مورد نظر وجود داشته باشد. از طرفی در این مقاله از تکنیک الگوریتم **Self-Adversarial Training (SAT)** و موزاییک کردن تصاویر نیز برای تولید تصاویر جدید استفاده شد (Bochkovskiy et al., 2020). طبق شکل ۷ با ترکیب هر چهار تصویر موجود در پایگاه داده به ساخت تصاویر ترکیبی جدید برای آموزش بهتر شبکه‌های **YOLOV3** و **YOLOV4** استفاده شده است.



شکل ۷. نمونه‌ای از تصویر تولید شده از روش موزاییک کردن در

پایگاه داده ایجاد شده برای تشخیص موتورسیکلت

پس از اخذ و جمع‌آوری تصاویر، با استفاده از نرم‌افزار **LabelImg** در محیط **Ubuntu** به صورت دستی موقعیت موتورسیکلت‌ها در هر یک از تصاویر به صورت جداگانه مشخص شد. یک فایل با فرمت ***.txt** نیز به ازای هر تصویر برای مشخص شدن اطلاعات مورد نیاز برای آموزش شبکه ساخته شد. هر سطر این فایل به ترتیب بیانگر پنج پارامتر از جمله: شماره‌ی کلاس، مختصات **X** موقعیت مرکز کادر موتورسیکلت در تصویر، مختصات

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق

فواصل زیر ۷ متر مربوط به محدوده‌ی پنجره‌ی شامل ۱۴ پیکسل انتخابی بر روی تصویر با دقت ۲ سانتی‌متر برآورد و ثبت می‌کند.

۴-۲ نتایج شناسایی موتورسیکلت

مدل‌های اصلی و اولیه ارائه شده از الگوریتم‌های YOLO بر اساس پایگاه داده‌های متنوعی آموزش داده شدند و سپس پارامترهای وزن محاسبه شده آنها در دسترس علاقه‌مندان قرار گرفته است (Redmon, 2018). در این مقاله هریک از مدل‌های از قبل آموزش دیده‌ی YOLOV3 و YOLOV4 مجدداً بر روی پایگاه داده‌ی تهیه شده، مورد آموزش قرار گرفتند و مقادیر وزن آنها متناسب با پایگاه داده‌ی موردنظر و بر اساس هدف مورد نظر اصلاح و بهینه‌سازی شد. همچنین برای بررسی مناسب‌ترین روش برای شناسایی موتورسیکلت‌های مورد نظر، مقایسه‌ای بر روی نتایج مدل‌های بهینه‌سازی و آموزش دیده‌ی YOLOV3 و YOLOV4 صورت گرفت.

دو شبکه‌ی YOLOV3 و YOLOV4 هر دو در قالب اصلی زبان برنامه نویسی C نوشته شده‌اند. در این مقاله برای راحتی استفاده و امکان ایجاد تغییر و بهبود در بخش‌های مختلف برنامه، از تبدیل زبان C به زبان ساده‌تر Python و از کتابخانه‌های پرکاربرد و رایجی همچون Tensorflow و Pytorch استفاده شده است. هر کدام از این دو شبکه در آخرین لایه آنها که مربوط به انتخاب کلاس عارضه‌ی شناسایی شده است، به جای ۸۰ کلاس اولیه معرفی شده در طراحی اصلی شبکه‌های YOLO، تنها یک کلاس موتورسیکلت معرفی شده است. به طور کلی سرعت عملکرد و اجرای الگوریتم با توجه به کمیت و تعداد پارامترهای تعریف شده-ی مورد استفاده در کل شبکه عصبی تعریف می‌شود (Chen et al., 2019). از آنجا که کاهش در تعداد کلاس‌های معرفی شده باعث کاهش در تعداد پارامترهای موجود در تنها لایه‌های تمام اتصال (لایه آخر) خواهد شد. کاهش کلاس‌های تعریف شده از

مدل دوربین	مشخصات فنی	مولفه
MYNT-EYE D1000-IR- 120/Color	زوایای دید	D:۱۲۱° H:۱۰۵° V:۵۸°
	تصحیح اعوجاجات شعاعی	k1=-۰/۳۰۶۶ k2=۰/۰۸۶۱
	تصحیح اعوجاجات غیر شعاعی	p1=-۰/۰۰۰۳ p2=۰/۰۰۱۵

در راستای پیش پردازش‌های مورد نیاز طبق برنامه طراحی شده بر روی دوربین در لحظه‌ی اخذ تصویر، عملیات تصحیح اعوجاجات شعاعی و تصحیحات غیر شعاعی انجام شده است. از طرفی عملیات ترمیم بر روی جفت تصاویر نیز صورت گرفته است تا تصاویر اپی‌پولار مربوط به هر جفت تصویر تولید شود. در واقع تصاویر اپی‌پولار تصاویر ترمیم شده‌ای هستند که در آن‌ها هر سطر در تصویر چپ با همان سطر در تصویر راست‌اش متناظر است. شکل ۸ نمونه‌ای از یک جفت از این تصاویر ترمیم شده است که توسط دوربین MYNT-EYE از خیابان شهر تهران گرفته شده و برای آموزش شبکه استفاده شده است.



شکل ۸. نمونه‌ای از جفت تصویر ترمیم شده اخذ شده توسط دوربین

MYNT-EYE برای استفاده در پایگاه داده‌ی موردنظر

علاوه بر پایگاه داده‌ی مربوط به آموزش شبکه، جهت کنترل و ارزیابی روش کلی مقاله، تعداد دویست عدد تک تصویر از موتورسیکلت‌های مختلف در فواصل معلوم و مشخص بین ۰/۳ تا ۷ متر توسط دوربین MYNT-EYE تهیه شده است. فرآیند اندازه‌گیری این فواصل معلوم موتورسیکلت در هر تصویر بر اساس الگوریتم طراحی شده در داخل دوربین است. به ازای هر تصویر فصلنامه مهندسی حمل‌ونقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱



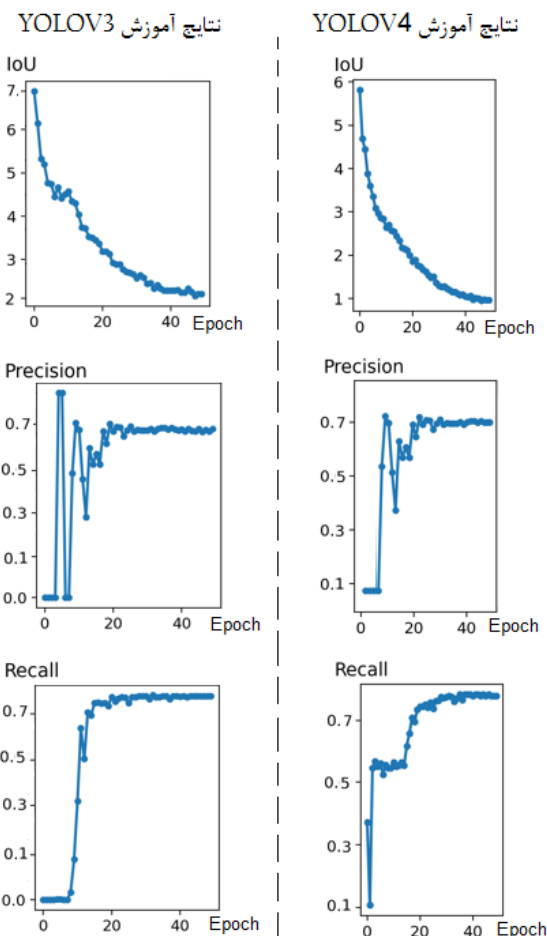
شکل ۹. الف) حضور همزمان دو موتورسیکلت در تصویر، ب) تداخل و پوشیده شدن تصویر موتورسیکلت با سایر عوارض مسیر

برای مقایسه و ارزیابی مدل‌های تشخیص و شناسایی عارضه از برخی شاخص‌های آماری متداول از جمله: Recall, Precision و Average IOU استفاده می‌شود (Godard et al., 2017). Precision یا دقت بیانگر این است که چه مقداری از کل مقادیر پیش‌بینی شده در کلاس مورد نظر به درستی تخمین زده شده است. Recall یا صحت بیانگر این است چه مقدار از مقادیر واقعی کلاس مورد نظر به درستی پیش‌بینی شده است. IOU بیانگر نسبت اشتراک بین کادر محصور محاسبه شده و واقعی به اجتماع بین این دو کادر است و مقدار Average IOU میانگین تمام مقادیر IOU مربوط به تصاویر موجود در پایگاه داده است. هر یک از این مقادیر طبق فرمول‌های ۳ تا ۵ محاسبه می‌شوند. در این روابط TP بیانگر مقادیری است که به درستی به عنوان کلاس مورد نظر پیش‌بینی شده‌اند. FP بیانگر مقادیری است که به غلط به عنوان کلاس مورد نظر پیش‌بینی شده‌اند. FN بیانگر مقادیری از کلاس مورد نظر است که اشتباه پیش‌بینی شده‌اند. A مقدار مساحت کادر محصور پیش‌بینی شده و B مقدار مساحت کادر محصور واقعی می‌باشد. با توجه به روابط و نوع مقادیر نسبت‌ها، هر سه پارمتر Precision, Recall و IOU به نحوی بیانگر مقادیر دقت عددی (بدون واحد) الگوریتم هستند.

۸۰ کلاس به ۱ کلاس موجب کاهش در تعداد پارامترهای لایه‌ی آخر این شبکه‌ها شده است. با مقایسه و بررسی نتایج بدست آمده، افزایش سرعت حدود ۲/۴ فریم بر ثانیه در اجرای الگوریتم تشخیص صورت گرفته است. نتایج آموزش مجدد هر دو شبکه بر روی پایگاه داده‌ی خاص تهیه شده از موتورسیکلت‌ها نشان می‌دهد که در صورتی که در مرحله آموزش از پایگاه داده‌ی غنی و مشابه با هدف نهایی استفاده شود، آنگاه دقت یا Precision نتایج بدست آمده به میزان ۱۵ درصد نسبت به مدل‌های اولیه‌ی آموزش دیده بر روی پایگاه داده‌ی COCO رشد داشته است.

در این الگوریتم در صورت شناسایی همزمان چند موتورسیکلت و یا تداخل تصویر دو موتورسیکلت با یکدیگر در تصویر، موتورسیکلتی مورد تمرکز و بررسی قرار می‌گیرد که با احتمال و امتیاز بالاتر شناسایی شده باشد. این به این معنی است که همچون شکل ۹ الف در وضعیت تداخل و یا حضور همزمان، موتورسیکلت نزدیکتر که تداخل تصویری ندارد و با وضوح کامل‌تر در تصویر قرار گرفته با امتیاز بالاتر شناسایی می‌شود و در اولویت بررسی ادامه الگوریتم قرار می‌گیرد. از طرفی همچون شکل ۹ ب در صورت پوشیده شدن و تداخل بیش از نیمی از تصویر موتورسیکلت با سایر عوارض (مثلاً سایر اتومبیل‌های پشت سر)، الگوریتم توانایی شناسایی و تشخیص موتورسیکلت را نخواهد داشت. این وضعیت هیچ خلل و مشکلی برای هدف اصلی مقاله ایجاد نخواهد کرد زیرا در این وضعیت عارضه یا وسیله‌ی قرار گرفته در جلوی موتورسیکلت همچون مانعی بین موتورسیکلت و اتومبیل مورد نظر خواهد شد و از بروز وضعیت خطر و اضطراب جلوگیری خواهد کرد.

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق



شکل ۱۰. نمودار نتایج شاخص‌های آماری محاسبه شده بر روی مدل‌های YOLOV3 و YOLOV4 در طول مراحل آموزشی

بر اساس شاخص‌های آماری تعریف شده، عملیات شناسایی موتورسیکلت‌های مورد نظر بر روی مدل‌های بهینه سازی شده و آموزش یافته‌ی YOLOV3 و YOLOV4 با توجه به پایگاه داده‌ی آزمون مشترک صورت گرفته است. نتایج بدست آمده از ارزیابی و مقایسه مقادیر دقت و سرعت این دو الگوریتم در جدول ۲ مشاهده می‌شود. پس از آموزش الگوریتم‌های YOLOV3 و YOLOV4 در ۴۰ مرحله، به ترتیب دقت شناسایی موتورسیکلت‌های مورد نظر ۶۹ درصد و ۷۵ درصد محاسبه شده است. با توجه به نتایج بدست آمده عملکرد مدل YOLOV4 در

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (۳)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (۴)$$

$$\text{Average IOU} = (\text{A} \cap \text{B})/(\text{A} \cup \text{B}) \quad (۵)$$

در طول روند آموزش شبکه‌ها مقادیر بدست آمده از پارامترهای Precision, IOU و Recall به ازای هر مرحله^{۱۵} آموزشی شبکه ثبت و نمودار هر یک از این پارامترها طبق نمودارهای شکل ۱۰ محاسبه شده است. در این نمودارها محورهای افقی بیانگر شماره‌ی مرحله‌ی آموزشی و محورهای عمود بر بیانگر مقدار شاخص دقت مورد نظر است. طبق این نمودارها روند همگرایی هر دو مدل در طول مراحل آموزشی به خوبی قابل مشاهده است. مقادیر دقت بدست آمده از پارامتر Precision در هر دو مدل بعد از مرحله‌ی ۴۰ ام روند همگرایی پیدا کرده و تقریباً ثابت شده. به همین خاطر پارامترهای وزن ثبت شده در مرحله‌ی ۴۰ ام به عنوان نتایج مرحله-ی بهینه و پایانی ثبت شده است. با توجه به آموزش هر دو مدل در طول ۴۰ مرحله، پارامترهای دقت بدست آمده از YOLOV4 در مدت زمان کوتاهتری به سمت مقادیر بهینه‌تر حرکت داشته است. از طرفی با توجه به فرمول ۵ هر چه مقدار IOU محاسبه شده به یک نزدیک‌تر باشد، شبکه به صورت بهینه‌تری آموزش دیده شده است. طبق نمودار مقادیر IOU مربوط به مدل آموزش دیده YOLOV4 در مرحله مشابه با YOLOV3 مقادیر IOU بهینه-تر و مناسب‌تری را نشان داده است.

شناسایی این دسته از موتورسیکلت‌ها در پارامتر دقت حدود ۸ درصد بالاتر از نتایج مدل YOLOV3 در هنگام اجرا بوده است. از طرفی نیز مدل YOLOV4 توانسته موتورسیکلت‌های موجود در تصاویر را با سرعت ۳۵ فریم بر ثانیه (هر تصویر را در مدت ۰/۲۸ ثانیه) پیش‌بینی و در تصویر تعیین موقعیت کند. این زمان حدود ۱۰ درصد کوتاه‌تر از زمان شناسایی الگوریتم آموزش دیده‌ی YOLOV3 بود است.

تصویر است. همواره برآورد این فاصله با استفاده از یک دوربین در حال حرکت چالش بزرگی بوده است. در صورت برآورد این فاصله با استفاده از پارامترها و ابعاد از قبل اندازه‌گیری شده در صحنه‌ی تصویر برداری و همچنین پارامترهای توجیه خارجی دوربین، خطا به صورت تصاعدی در محاسبه‌ی فاصله بروز پیدا می‌کند (Cao et al., 2019). به همین خاطر در این مقاله از روش‌های یادگیری عمیق و مدل Monodepth2 استفاده شده است. در این مدل نیازی به تهیه‌ی پایگاه داده‌ای بزرگ و وسیع از تصاویر با مقادیر عمق زمینی معلوم آنها نیست و با آموزش شبکه تنها با استفاده از مجموعه‌ای از جفت تصاویر، شبکه قابلیت محاسبه و تولید نقشه عمق از تک تصویر ورودی را خواهد داشت (Godard et al., 2019).

جدول ۲. ارزیابی سرعت و دقت مدل‌های آموزش دیده برای تشخیص موتورسیکلت‌های مورد نظر

نام مدل آموزش دیده	YOLOV3	YOLOV4
Precision شناسایی موتورسیکلت	۶۹ درصد	۷۵ درصد
Recall شناسایی موتورسیکلت	۷۵ درصد	۷۹ درصد
IOU شناسایی موتورسیکلت	۱/۹۶	۰/۹۷
مدت زمان شناسایی موتورسیکلت	۳۲ فریم بر ثانیه	۳۵ فریم بر ثانیه

مدل Monodepth2 بر روی پایگاه داده‌های مختلفی از جمله ImageNet (Deng et al., 2009)، KITTI (Geiger et al., 2013) و CityScape (Cordts et al., 2016) آموزش و اجرا شده است و مقادیر پارامترهای وزن مدل‌های آموزش داده شده در اختیار کاربران علاقه‌مند به صورت رایگان قرار گرفته است (Clément Godard, 2019). فضاها و تصاویر موجود در پایگاه داده‌ی KITTI بیشترین نزدیکی و شباهت را با فضای شهری مورد نظر مقاله (خیابان‌های شهر تهران) دارند. به همین خاطر برای رسیدن به نتایج بهتر مدل آموزش دیده بر روی KITTI به عنوان مقادیر وزن اولیه شبکه مورد استفاده قرار گرفته است. از آنجا که برای رسیدن به دقت بهینه نیاز به ثابت ماندن پارامترهای دوربین در مرحله‌ی آموزش و آزمون است، مدل Monodepth2 بر اساس پایگاه داده‌ی حاصل از جفت تصاویر دوربین MYNT-EYE از سطح خیابان‌های شهر تهران مجدداً آموزش دیده و پارامترهای وزن شبکه تنظیم و بهبود یافته است.

از آنجا که هدف مورد نظر در این مقاله مشخص کردن موقعیت دقیق موتورسیکلت‌ها در تصاویر با استفاده از مناسب‌ترین روش موجود است، بنابراین با توجه به دقت قابل قبول ۷۵ درصدی و صحت ۷۹ درصدی نتایج آموزش شبکه‌ی YOLOV4 برای تشخیص کلاس موتورسیکلت‌های دارای بادگیر مشکی رنگ، از این مدل در مرحله‌ی اول این الگوریتم پیشنهادی استفاده شده است.

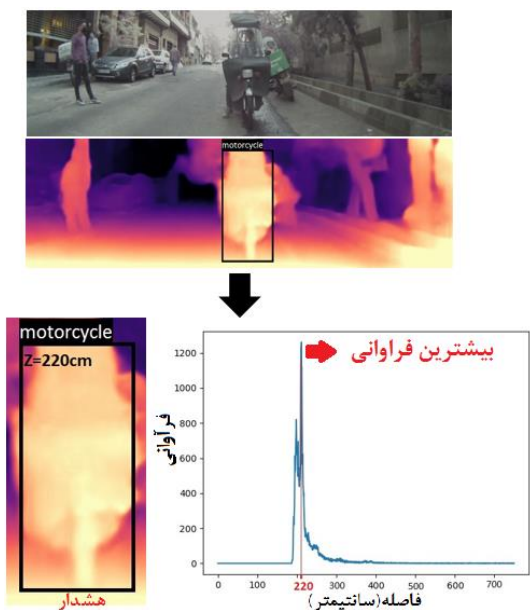
۳-۴ نتایج تولید نقشه عمق

شناسایی و تشخیص موتورسیکلت در تصویر به تنهایی دلیل موجهی بر لزوم هشدار به راننده نیست. پس از آگاهی الگوریتم نسبت به حضور این دسته از وسایل نقلیه، نیازمند اجرای مرحله‌ی دوم الگوریتم تحت عنوان محاسبه‌ی فاصله موتورسیکلت در

ارائه‌ی الگوریتم هشدار نسبت به موتورسیکلت‌های پشت‌سر با استفاده از تک دوربین و روش‌های یادگیری عمیق

۴-۴ ادغام نتایج بدست آمده از شبکه‌های آموزش دیده

هدف از تلفیق نتایج بدست آمده از شناسایی موقعیت کادر موتورسیکلت مورد نظر و محاسبه‌ی مقادیر فاصله‌ی پیکسل‌های تصویر، در راستای کنترل شرایط و هشدار نسبت به وضعیت خطر است. نمونه‌ای از اجرای کامل مدل تلفیقی استفاده از دو شبکه‌ی آموزش داده شده‌ی YOLOV4 و Monodepth2 بر روی یکی از تک تصاویر اخذ شده از موتورسیکلت در حال حرکت و مجهز به بادگیر مشکی رنگ در شکل ۱۲ قابل مشاهده است.



شکل ۱۲. نتیجه‌ی اجرای مدل پیشنهادی بر روی نمونه‌ی تک تصویر اخذ شده از موتورسیکلت موردنظر

در این نمونه پس از برآورد نمودار فراوانی پیکسل‌های عمق مربوط به محدوده‌ی موتورسیکلت، بیشترین تعداد این پیکسل‌ها فاصله‌ی ۲۲۰ سانتی متر را داشته‌اند. به همین خاطر مقدار فاصله‌ی ۲۲۰ سانتی متر به عنوان طول فاصله‌ی موتورسیکلت تا دوربین منتسب شده است. از آنجا که این مقدار فاصله کمتر از مقدار حد آستانه‌ی انتخابی این مقاله است (در این مقاله حد آستانه ۳ متر انتخاب شده است)، بنابراین به راننده اتومبیل در رابطه با حضور نزدیک

بازیابی و آموزش مجدد پارامترهای مدل Monodepth2 با استفاده از زبان برنامه نویسی Python و با استفاده از کتابخانه‌های Opencv و Pythorch در مدت زمان حدود ۳ ساعت با نرخ آموزشی ۱۰^{-۳} صورت گرفته است. در مجموع تصاویر در غالب ۲۰ سکانس مختلف تهیه شد که آموزش شبکه بر روی ۱۶ سکانس از این تصاویر انجام شد و سپس از تصاویر مربوط به ۴ سکانس دیگر به عنوان داده‌های ارزیابی استفاده شد. با توجه به داده‌های آموزش و ارزیابی، این شبکه پس از ۲۰ مرحله آموزش، به وضعیت بهینه با مجموع مربعات خطای ۰/۲۸ دست یافت.

مقادیر نقشه اختلاف بدست آمده از الگوریتم شاخص نسبی و غیر متریکی برای هدف مقاله است. برای برآورد فاصله‌ی موتورسیکلت بر حسب متر، طبق فرمول ۲ (که در قسمت روش تحقیق به آن اشاره شد) مقادیر اختلاف منظر مربوط به پیکسل‌ها تبدیل به مقادیر فاصله بر حسب واحد متر شوند. نمونه‌ای از نتیجه بدست آمده از این مدل در برآورد فاصله‌ی پیکسل‌های تصویر اخذ شده از موتورسیکلت در شکل ۱۱ نشان داده شده است. در تصویر عمق تولید شده در این مدل، پیکسل‌های مربوط به عوارض نزدیک‌تر با رنگ روشن‌تر و پیکسل‌های مربوط به عوارض دورتر با رنگ تیره‌تر نمایش داده شده است.



شکل ۱۱. تصویر عمق تولید شده توسط پارامترهای بهبود یافته‌ی مدل Monodepth2 بر روی تک تصویر اخذ شده از موتورسیکلت

به محیط و پشت زمینه، الگوریتم به خوبی قادر به برآورد نقشه عمق نخواهد شد. به شکلی که این میانگین خطای نسبی در محاسبه‌ی این دسته از فواصل زیر ۳ متر گاه تا ۵۰ سانتی متر هم پیش می‌رود. از آنجا که تمرکز این مقاله اندازه‌گیری و کنترل وضعیت قبل از بروز هشدار و رسیدن به فاصله‌ی ۳ متر است، لذا توجه اصلی در برآورد دقیق‌تر فواصل محدوده‌ی ۳ تا ۷ متر خواهد بود. در این الگوریتم این محدوده فواصل با دقت بیشتر محاسبه شده‌اند.

۵. نتیجه‌گیری و پیشنهادات

با توجه به لزوم پایش موتورسیکلت‌های مورد استفاده در شهرهای پرتددی همچون کلان شهر تهران، سعی شده تا با تمرکز بر روی موتورسیکلت‌ها به ویژه موتورسیکلت‌های با پوشش خاص بادگیر مشکی رنگ تا حدی بتوان میزان تصادفات بین این دسته از موتورسیکلت‌ها و اتومبیل‌ها را کاهش داد. در راستای این امر در این مقاله از الگوریتم‌ها و مدل‌های یادگیری عمیق استفاده شد که مشکلاتی همچون سرعت نسبتاً پایین روش‌های سنتی ماشین بینایی (همچون الگوریتم‌های SVM و طبقه‌بندی کننده‌های آبخاری) در شناسایی موتورسیکلت و همچنین دقت پایین در برآورد فاصله با توجه به برخی ابعاد معلوم تصویر و توجیه پارامترهای ماتریس هموگرافی را نداشته است. این روش پیشنهادی بر اساس اجرای سه مرحله‌ی تشخیص این دسته از موتورسیکلت‌ها توسط الگوریتم آموزش داده شده‌ی YOLOV4 و در مرحله‌ی دوم برآورد نقشه عمق بر اساس مدل آموزش داده شده‌ی Monodepth2 بر روی پایگاه داده‌ی تهیه شده از دوربین MYNT-EYE صورت گرفت. در مرحله سوم برای محاسبه‌ی فاصله هر یک از این موتورسیکلت‌های مشخص شده در تصویر از تلفیق نتایج بدست آمده از موقعیت موتورسیکلت و مقدار فواصل عمق اندازه‌گیری شده در مرحله‌ی قبل استفاده شده است. از بین مقادیر عمق پیکسل‌های منتسب به فصلنامه مهندسی حمل و نقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱

موتورسیکلت هشدار داده خواهد شد. این مدل بر روی مجموعه داده‌های مربوط به تصاویر موتورسیکلت‌ها پیاده سازی شده است و مقایسه و ارزیابی بر روی نتایج بدست آمده با فواصل واقعی زمینی صورت گرفته است. این مدل پیشنهادی با توجه به شاخص‌های آماری متداول همچون میانگین مربعات خطا^{۱۶} و میانگین خطای نسبی^{۱۷} ارزیابی شده است. با توجه به تعریف نحوه‌ی محاسبه‌ی این شاخص‌ها در فرمول‌های ۶ و ۷ (Li et al., 2017)، هرچه مقدار این دو شاخص کمتر باشد، نشان دهنده‌ی نتایج دقیق‌تر مدل در برآورد فاصله عارضه مورد نظر است. در این رابطه d_i^{gt} بیانگر مقدار فاصله‌ی واقعی موتورسیکلت i ام تا دوربین، d_i بیانگر مقدار فاصله‌ی محاسبه شده‌ی موتورسیکلت i ام تا دوربین، T تعداد کل موتورسیکلت‌های موجود در ارزیابی است.

$$RMSE = \sqrt{\frac{1}{T} \sum_i (d_i^{gt} - d_i)^2} \quad (6)$$

$$MRE = \frac{1}{T} \sum_i \frac{|d_i^{gt} - d_i|}{d_i^{gt}} \quad (7)$$

ارزیابی‌های صورت گرفته در اجرای مدل پیشنهادی بر روی تصاویر اخذ شده از موتورسیکلت‌های در فاصله‌ی زیر ۷ متر، نشان داده است که به طور میانگین، در هر تک تصویر موقعیت و فاصله موتورسیکلت در مدت زمان ۰/۰۲۲ ثانیه (معادل ۴۵ فریم بر ثانیه) با میانگین خطای مربعات ۳۶ سانتی متر قابل اندازه‌گیری و محاسبه است. طبق بررسی‌های صورت گرفته در برآورد طول‌های محاسبه شده توسط الگوریتم در فواصل مختلف، مقادیر خطاهای متفاوتی بدست آمده است. در این میان میانگین مجموع مربعات خطا در برآورد فواصل کمتر از ۳ متر به نسبت فواصل ۳ تا ۷ متر بیشتر محاسبه شده است در صورتی که این فاصله کمتر از ۳ متر باشد، پارالاکس مربوط به موتورسیکلت افزایش پیدا می‌کند و باعث افزایش خطای الگوریتم می‌شود. علاوه بر این بخش عمده‌ای از تصویر را نیز موتورسیکلت در بر می‌گیرد و به دلیل عدم دید کافی

9. Pooling
10. Parallax
11. Camera Baseline
12. Random-access memory (RAM)
13. Video Random Access Memory (VRAM)
14. Rectified images
15. Epoch
16. Root mean squared error (RMSE)
17. Mean relative error (MRE)

۷. منابع

- برومندینیا، ع. و غلامی، م. (۲۰۱۹). "تشخیص کاراکتر فارسی پلاک خودرو، مستقل از چرخش و اندازه با استفاده از ممان‌های متعامد". فصلنامه مهندسی حمل و نقل، شماره ۴، ص. ۹۶۱-۹۸۵.
- عصرخودرو. (۲۰۱۹). "آمار تصادفات موتورسیکلتها در سال ۹۷" <http://www.asrekhodro.com/News/163208/>
- عصرخودرو. (۲۰۲۰). "استفاده از بادگیر برای موتورسیکلت سواران" <http://www.asrekhodro.com/News/192517>.
- کربلایی، م.، آیتی، ا. و صادقی، ع. (۲۰۱۹). "بررسی و مدل سازی تصادفات جرحی و حرکت های نامناسب موتورسیکلت سواران در میدان های شهری". فصلنامه مهندسی حمل و نقل.

- Anaya, J. J., Ponz, A., García, F., & Talavera, E. (2017). "Motorcycle detection for ADAS through camera and V2V Communication, a comparative analysis of two modern technologies". Expert Systems with Applications, Vol. 77, pp. 148-159.
- Bhujbal, A., & Mane, D. (2019). "Vehicle Type Classification Using Deep Learning". Paper presented at the International Conference on Soft Computing and Signal Processing, pp. 279-290.

موتورسیکلت، مقدار فاصله‌ی موتورسیکلت تا دوربین بر اساس رسم نمودار فراوانی و انتخاب فاصله با بیشترین فراوانی صورت گرفته است. طبق نتایج بدست آمده از ارزیابی این مدل بر روی فواصل زیر ۷ متر، میانگین خطا ۳۶ سانتی‌متر در برآورد فاصله‌ی این دسته از موتورسیکلت‌ها محاسبه شده است. از آنجا که آموزش شبکه‌ها بر روی تصاویر اخذ شده در طول روز و آب و هوای ابری و آفتابی انجام شده، به همین خاطر مدل مورد استفاده در برآورد نقشه‌ی عمق در وضعیت بارندگی و کاهش دید محیط، توانایی عملکرد خوبی نداشته و با خطا همراه بوده است. با توجه به اینکه اجرای این سه مرحله در مجموع حدود ۰/۰۵ ثانیه طول کشیده است. این مقدار در واقع معادل پردازش حدود ۲۰ فریم در هر ثانیه است. به همین خاطر می‌توان ادعا داشت که با استفاده از این دو روش پیشنهادی می‌توان به بهترین عملکرد برای اجرای سیستم هشدار به صورت آنی رسید.

در انتها پیشنهاد می‌شود که می‌توان با اضافه کردن کلاس‌های بیشتر همچون دوچرخه، اتومبیل، اتوبوس و کامیون این هشدار را نسبت به سایر دیگر کاربران جاده نیز اعمال کرد و همچنین در ادامه می‌توان با بررسی مداوم و پایش پیوسته‌ی نحوه‌ی حرکت وسیله‌ی مورد نظر با استفاده از الگوریتم‌های یادگیری عمیق به پیش‌بینی چگونگی حرکت و انتخاب مسیرهای آتی وسیله دست یافت. می‌شود.

۶. پی‌نوشت‌ها

1. Active sensor
2. Passive sensor
3. Graphics processing unit
4. Field Programmable Gate Arrays
5. Supervised
6. Unsupervised
7. Disparity map
8. Bounding box

فصلنامه مهندسی حمل و نقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱

- Clément Godard, O. M. A., Michael Firman and Gabriel J. Brostow. (2019). "monodepth2". www.github.com/nianticlabs/monodepth2
- Cordts, M., & Omran, M. (2016). "The cityscapes dataset for semantic urban scene understanding". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213-3223.
- Dabbour, E., & Easa, S. (2014). "Proposed collision warning system for right-turning vehicles at two-way stop-controlled rural intersections". Transportation research part C: emerging technologies, Vol. 42, pp. 121-131.
- Damian, C., Grigorescu, I., & Robu, M. (2019). "Using mono and stereo camera system for static and moving objects detection". Paper presented at the 2019 International Conference on Electromechanical and Energy Systems (SIELMEN), pp. 1-5.
- Deigmoeller, J., Einecke, N., Fuchs, O., & Janssen, H. (2018). "Road Surface Scanning using Stereo Cameras for Motorcycles". Paper presented at the VISIGRAPP (5: VISAPP), pp. 549-554.
- Deng, J., Dong, W., Socher, R. (2009). "Imagenet: A large-scale hierarchical image database". Paper presented at the 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255.
- Dijk, T. v., & Croon, G. d. (2019). "How do neural networks see depth in single images?". Paper presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 2183-2191.
- Eigen, D., & Puhrsch, C. (2014). "Depth map prediction from a single image using a multi-scale
فصلنامه مهندسی حمل و نقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱
- Bochkovskiy, A., & Liao, H.-Y. M. (2020). "YOLOv4: Optimal Speed and Accuracy of Object Detection". arXiv preprint arXiv:2004.10934.
- Borges, P. V. K., & Vidas, S. (2016). "Practical infrared visual odometry". IEEE Transactions on Intelligent Transportation Systems, Vol. 17, No. 8, pp. 2205-2213.
- Cao, Z., Yang, D., Jiang, K., Xu, S., Wang, S., Zhu, M., & Xiao, Z. (2019). "A geometry-driven car-following distance estimation algorithm robust to road slopes". Transportation research part C: emerging technologies, Vol. 102, pp. 274-288.
- Chabot, F., & Chaouch, M. (2017). "Deep manta: A coarse-to-fine many-task network for joint 2d and 3d vehicle analysis from monocular image". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2040-2049.
- Chen, Y., & Liu, H.-L. (2016). "Overview of landmarks for autonomous, vision-based landing of unmanned helicopters". IEEE Aerospace and Electronic Systems Magazine, Vol. 31, No. 5, pp. 14-27.
- Chen, Z., & Khemmar, R. (2019). "Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility". Paper presented at the 2019 Eighth International Conference on Emerging Security Technologies (EST), pp. 1-6.
- Chiverton, J. (2012). "Helmet presence classification with motorcycle detection and tracking". IET Intelligent Transport Systems, Vol. 6, No. 3, pp. 259-269.

International Journal of Robotics Research, Vol. 32, No. 11, pp. 1231-1237.

- Girshick, R. (2015). "Fast r-cnn". Paper presented at the Proceedings of the IEEE international conference on computer vision, pp. 1440-1448.

- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587.

- Glenn Jocher, L. C., & Adam Hogan. (2020). ultralytics/yolov5: Initial Release (Version v1.0): Zenodo. <https://doi.org/10.5281/zenodo.3908560>

- Godard, C., Mac Aodha, O., & Brostow, G. J. (2017). "Unsupervised monocular depth estimation with left-right consistency". Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 270-279.

- Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). "Digging into self-supervised monocular depth estimation". Paper presented at the Proceedings of the IEEE international conference on computer vision, pp. 3828-3838.

- GONG, D.-W., Dai, X., Chen, Y., & WANG, S.-F. (2019). "Single-layer Laser Scanner-based Approach for a Transportation Participants Recognition Task". Lasers in Engineering (Old City Publishing), Vol. 43, pp. 10-12.

- Gruyer, D., & Rahal, M.-C. (2019). "Multi-Layer Laser Scanner Strategy for Obstacle Detection and Tracking". Paper presented at the 2019 IEEE/ACS 16th International Conference on

deep network". Paper presented at the Advances in neural information processing systems, pp. 2366-2374.

- Espinosa, J. E., & Velastin, S. A. (2017a). "Motorcycle classification in urban scenarios using convolutional neural networks for feature extraction".

- Espinosa, J. E., Velastin, S. A., & Branch, J. W. (2017b). "Motorcycle classification in urban scenarios using convolutional neural networks for feature extraction". pp. 26-26.

- Espinosa, J. E., Velastin, S. A., & Branch, J. W. (2018a). "Motorcycle detection and classification in urban Scenarios using a model based on Faster R-CNN". pp. 16-16.

- Espinosa, J. E., Velastin, S. A., & Branch, J. W. (2018b). "Motorcycle detection and classification in urban Scenarios using a model based on Faster R-CNN". arXiv preprint arXiv:1808.02299.

- Everingham, M., & Van Gool, L. (2010). "The pascal visual object classes (voc) challenge". International journal of computer vision, Vol. 88, No. 2, pp. 303-338.

- Fernández, C., & Llorca. (2013). "Real-time vision-based blind spot warning system: Experiments with motorcycles in daytime/nighttime conditions". International Journal of Automotive Technology, Vol. 14, No. 1, pp. 113-122.

- Garg, R., Bg, & V. K. (2016). "Unsupervised cnn for single view depth estimation: Geometry to the rescue". Paper presented at the European conference on computer vision, pp. 740-756.

- Geiger, A., Lenz, P., Stiller, & C. (2013). "Vision meets robotics: The kitti dataset". The فصلنامه مهندسی حمل‌ونقل / سال سیزدهم / شماره سوم (۵۲) / بهار ۱۴۰۱

- Kulkarni, Y., Bodkhe, S., Kamthe, A., & Patil, A.(2018). "Automatic number plate recognition for motorcyclists riding without helmet". Paper presented at the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), pp. 1-6.
- Laroca, R., & Severo, E.(2018). "A robust real-time automatic license plate recognition based on the YOLO detector". Paper presented at the 2018 international joint conference on neural networks (ijcnn), pp. 1-10.
- Lenz, I., Lee, H., & Saxena, A. (2015). "Deep learning for detecting robotic grasps". The International Journal of Robotics Research, Vol. 34, No. 4-5, pp. 705-724.
- Li, J., Klein, R., & Yao, A.(2017). "A two-streamed network for estimating fine-scaled depth maps from single rgb images". Paper presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 3372-3380.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L.(2014). "Microsoft coco: Common objects in context". Paper presented at the European conference on computer vision, pp. 740-755.
- Liu, F., Shen, C., Lin, G., & Reid, I. (2015). "Learning depth from single monocular images using deep convolutional neural fields". IEEE transactions on pattern analysis and machine intelligence, Vol. 38, No. 10, pp. 2024-2039.
- Liu, J., Sun, Q., Fan, Z., & Jia, Y.(2018). "TOF lidar development in autonomous vehicle". Paper presented at the 2018 IEEE 3rd Optoelectronics Global Conference (OGC), pp. 185-190.
- Computer Systems and Applications (AICCSA), pp. 1-8.
- He, K., Zhang, X., Ren, S., & Sun, J.(2016). "Deep residual learning for image recognition". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778.
- Hirschmuller, H.(2005). "Accurate and efficient stereo processing by semi-global matching and mutual information". Paper presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, pp. 807-814.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q.(2017). "Densely connected convolutional networks". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708.
- Jamaluddin, N. L., & Areni, I. S.(2019). "Detection and Distance Estimation against Motorcycles as Navigation Aids for Visually-impaired People". Paper presented at the 2019 12th International Conference on Information & Communication Technology and System (ICTS), pp. 224-228.
- Jiang, H., & Larsson, G.(2018). "Self-supervised relative depth learning for urban scene understanding". Paper presented at the Proceedings of the European Conference on Computer Vision (ECCV), pp. 19-35.
- Kim, M., Liu, Z., & Kang, D. (2016). "On road vehicle detection by learning hard samples and filtering false alarms from shadow features". Journal of Mechanical Science and Technology, Vol. 30, No. 6, pp. 2783-2791.

Paper presented at the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1-6.

- Mun, S., Nguyen, M. D., Kweon, S., & Bae, Y. H. (2019). "Deep Learning Object Detection to Clearly Differentiate Between Pedestrians and Motorcycles in Tunnel Environment Using YOLOv3 and Kernelized Correlation Filters", Vol. 24, No. 7, pp. 1266-1275.

- Nair, V., & Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines". Paper presented at the ICML.

- Ouyang, L., & Wang, H. (2019). "Vehicle target detection in complex scenes based on YOLOv3 algorithm". Paper presented at the IOP Conference Series: Materials Science and Engineering, Vol. 569, No. 5, pp. 052018.

- Pineda-Deom, D. (2019). "Motorcycle blind spot detection system and rear collision alert using mechanically aligned radar." Google Patents.

- Ramachandran, P., Zoph, B., & Le, Q. V. (2017). "Searching for activation functions". arXiv preprint arXiv:1710.05941.

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). "You only look once: Unified, real-time object detection". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788.

- Redmon, J., & Farhadi, A. (2017). "YOLO9000: better, faster, stronger". Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271.

- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). "Ssd: Single shot multibox detector". Paper presented at the European conference on computer vision, pp. 21-37.

- Luvizon, D. C., Nassu, B. T., & Minetto, R. (2016). "A video-based system for vehicle speed measurement in urban roadways". IEEE Transactions on Intelligent Transportation Systems, Vol. 18, No. 6, pp. 1393-1404.

- Mahto, P., Garg, P., Seth, P., & Panda, J. (2020). "Refining Yolov4 for Vehicle Detection". International Journal of Advanced Research in Engineering and Technology (IJARET), Vol. 11, No. 5, pp. 409-419.

- Markiewicz, P., Długosz, M., & Skruch, P. (2017). "Review of tracking and object detection systems for advanced driver assistance and autonomous driving applications with focus on vulnerable road users sensing". Paper presented at the Polish Control Conference, pp. 224-237.

- Misra, D. (2019). "Mish: A self regularized non-monotonic neural activation function". arXiv preprint arXiv:1908.08681.

- Mistry, J., Misraa, A. K., & Agarwal. (2017a). "An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network". Paper presented at the Image Processing Theory, Tools and Applications (IPTA), 2017 Seventh International Conference on, pp. 1-6.

- Mistry, J., Misraa, & A. K. (2017b). "An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network".

- Sochor, J., Juránek, R., & Špaňhel, J. (2018). "Comprehensive data set for automatic single camera visual speed measurement". IEEE Transactions on Intelligent Transportation Systems, Vol. 20, No. 5, pp. 1633-1643.
- Tang, Z., Wang, G., Xiao, H., Zheng, A., & Hwang, J.-N.(2018). "Single-camera and inter-camera vehicle tracking and 3D speed estimation based on fusion of visual and semantic features". Paper presented at the CVPR Workshop (CVPRW) on the AI City Challenge,
- Thakur, R. (2018). "Infrared Sensors for Autonomous Vehicles". Recent Development in Optoelectronic Devices, pp. 81.
- Van Ratingen, M., & Williams, A. (2016). "The European new car assessment programme: a historical review". Chinese journal of traumatology, Vol. 19, No. 2, pp. 63-69.
- Vishnu, C., Singh, D., & Mohan, C. (2017). "Detection of motorcyclists without helmet in videos using convolutional neural network". Paper presented at the 2017 International Joint Conference on Neural Networks (IJCNN), pp. 3036-3041.
- Wang, S., Liu, F., Gan, Z., & Cui, Z.(2016). "Vehicle type classification via adaptive feature clustering for traffic surveillance video". Paper presented at the 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP), pp. 1-5.
- Wang, W., Yang, S., Li, Y., & Ding, W.(2015). "A rough vehicle distance measurement method using monocular vision and license plate". Paper presented at the 2015 IEEE International Conference on Cyber Technology in Automation,
- Redmon, J., & Farhadi, A. (2018). "Yolov3: An incremental improvement". arXiv preprint arXiv:1804.02767.
- Redmon, J. a. F., Ali. (2018). "yolo". <https://pjreddie.com/darknet/yolo/>
- Ren, S., He, K., Girshick, R., & Sun, J.(2015). "Faster r-cnn: Towards real-time object detection with region proposal networks". Paper presented at the Advances in neural information processing systems, pp. 91-99.
- Rogers, S., & Papanikolopoulos, N. P.(2000). "Counting bicycles using computer vision". Paper presented at the Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE, pp. 33-38.
- Salahat, S., Al-Janahi, & A.(2017). "Speed estimation from smart phone in-motion camera for the next generation of self-driven intelligent vehicles". Paper presented at the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), pp. 1-5.
- Shine, L., & Jiji, C. V. (2020). "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN". Multimedia Tools and Applications, pp. 1-21.
- Shotton, J., Fitzgibbon, A., Cook, M., & Sharp, T.(2011). "Real-time human pose recognition in parts from single depth images". Paper presented at the CVPR 2011, pp. 1297-1304.
- Sivaraman, S., & Trivedi, M. M. (2013). "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis". IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 4, pp. 1773-1795.

Control, and Intelligent Systems (CYBER), pp. 426-430.

- Wu, W., Kozitsky, V., Hoover, M. E., Loce, R., & Jackson, D. T.(2015). "Vehicle speed estimation using a monocular camera". Paper presented at the Video Surveillance and Transportation Imaging Applications 2015, Vol. 9407, pp. 940704.

- Xie, J., Girshick, R., & Farhadi, A.(2016). "Deep3d: Fully automatic 2d-to-3d video conversion with deep convolutional neural networks". Paper presented at the European Conference on Computer Vision, pp. 842-857.

- Yang, D., Jiang, K., Zhao, D., Yu, C., Cao, Z., Xie, S., Zhang, K. (2018). "Intelligent and connected vehicles: Current status and future perspectives". Science China Technological Sciences, Vol. 61, No. 10, pp. 1446-1471.

- YG, A. R., Kumar, S., Amaresh, H., & Chirag, H.(2015). "Real-time speed estimation of vehicles from uncalibrated view-independent traffic cameras". Paper presented at the TENCON 2015-2015 IEEE Region 10 Conference, pp. 1-6.

- Zhou, Y., & Tuzel, O.(2018). "Voxelnet: End-to-end learning for point cloud based 3d object detection". Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4490-4499.

زهرا بادامچی شبستری، علی حسینی نوه

زهرا بادامچی شبستری، درجه کارشناسی در رشته مهندسی عمران-نقشه‌برداری را در سال ۱۳۹۶ از دانشگاه صنعتی خواجه نصیرالدین طوسی و درجه کارشناسی ارشد در رشته‌ی مهندسی سنجش از دور را در سال ۱۳۹۹ از دانشگاه صنعتی خواجه نصیرالدین طوسی اخذ نموده است. زمینه‌های پژوهشی مورد علاقه ایشان ماشین بینایی، پردازش تصویر، شبکه‌های عصبی و یادگیری عمیق بوده است.



علی حسینی نوه، درجه کارشناسی در رشته عمران-نقشه‌برداری را در سال ۱۳۸۳ از دانشگاه آزاد اسلامی واحد میبد و درجه کارشناسی ارشد در رشته مهندسی فوتوگرامتری را در سال ۱۳۸۷ از دانشگاه خواجه نصیرالدین طوسی اخذ نمود. در سال ۱۳۹۲ موفق به کسب درجه دکتری در رشته فوتوگرامتری از دانشگاه University College London گردید. زمینه‌های پژوهشی مورد علاقه ایشان مدل‌سازی سه بعدی، بینایی ماشین، رباتیک و فوتوگرامتری بوده و در حال حاضر عضو هیات علمی با مرتبه استادیاری در دانشگاه خواجه نصیرالدین طوسی گروه فوتوگرامتری و سنجش از دور است.

