فصلنامه مهندسی ترافیک، سال بیست و دوم، شماره ۸۹، پاییز ۱۴۰۱، ص. ۳۷-۲۹

# سامانه خودیادگیر مبتنی بر پردازش تصویر جهت کمک به راننده در نقاط کور با استفاده از شبکه‌های عصبی عمیق

آرش پورحسن‌نژاد، کارشناس ارشد سامانه‌های هوشمند، دانشگاه امیرکبیر، تهران، ایران

مهدی قطعی (مسئول مکاتبات)، عضو هیات علمی گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران

هدیه ساجدی، عضو هیات علمی گروه علوم کامپیوتر، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران

**E-mail: ghatee@aut.ac.ir**

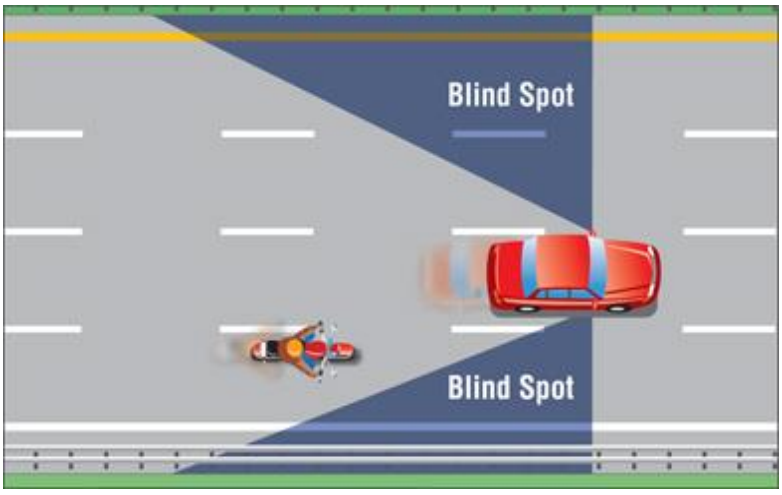
## چکیده

***در این مقاله یک سامانه دستیار راننده مبتنی بر دوربین منفرد برای کسب جزییات در نقاط کور ارائه می‌کنیم که بدون داده‌ی برچسب‌گذاری شده، به‌صورت خودیادگیر آموزش می‌بیند. سامانه پیشنهادی بر اساس شبکه‌های عمیق توسعه یافته است که به‌عنوان ورودی از تصویر خاکستری و جریان نوری [1] استفاده می‌کند. نقطه قوت این مقاله در مقایسه با مقالات مشابه، پردازش اطلاعات دنباله‌ای از تصاویر به‌منظور ارزیابی ریسک بروز تصادف با توجه به اشیای متحرک در نقاط کور است. در این مقاله با استفاده تشخیص و ردگیری اشیا در تصویر، وجود وسایل نقلیه و سرعت نسبی آنها تعیین می‌گردد و از روی آن برای نقاط کور خودرو، ریسک بروز تصادف پیش‌بینی می‌شود. ادغام جریان نوری با تصویر و همچنین ترکیب نتایج با ویژگی‌های استخراج‌شده از یک شبکه عصبی عمیق، باعث توانمندی سامانه پیشنهادی شده است. در کاربرد پیشنهادی مقاله، دوربین بر روی آینه کناری خودرو نصب شده است و با ۹۶ درصد دقت، خطر بروز تصادف هنگام چرخش به طرفین یا تغییر خط تخمین زده شده است.***

**واژه‌های کلیدی**: سامانه دستیار راننده، نقاط کور، پردازش تصویر، یادگیری عمیق، آموزش خودیادگیر، تشخیص شیء در تصویر، ردگیری اشیا

## ۱. مقدمه

با ظهور سامانه‌های هوشمند، خودروهای جدید اغلب به سامانه‌های دستیار راننده مجهز شده‌اند که می‌توانند از بروز تصادفات رانندگی جلوگیری کنند. در این مقاله، یک سامانه دستیار راننده برای نقاط کور پیشنهاد می‌شود که در کاهش تصادفات وسیله نقلیه هنگام گردش به طرفین و یا تغییر خط، می‌تواند تأثیر مهم داشته باشد. ازآنجاکه عدم‌تشخیص وسایل نقلیه در نقاط کور در هنگام چرخش به طرفین یا تغییر خط بسیار خطرناک است و نیز احتمال تصادف در هنگام چرخش، بسیار بالا است، ارائه هشدارهای مناسب به راننده هدف این مقاله قرار داده شده است. در این راستا، سامانه‌های مبتنی بر دوربین، زمینه دید وسیع‌تری و اطلاعات بیشتری نسبت به نمونه‌های دیگر فراهم می‌کنند.




**شکل ۱. نقاط کور خودرو که تمام وسایل نقلیه در آن نقاط با آینه‌های بغل قابل‌رؤیت نیست.**

بر اساس آمارهای اداره ملی ایمنی ترافیک بزرگراه‌های آمریکا، تقریباً ۸٤۰ هزار تصادف از نقطه کور در ایالات‌متحده هرسال اتفاق می‌افتد که منجر به ۳۰۰ مرگ به‌صورت تقریبی می‌شود. این ارقام در سایر کشورها مانند ایران نیز قابل‌توجه است. طبق برخی آمار، سالانه حدود ۸۰۰ هزار تصادف ترافیکی در ایران رخ می‌دهد و هرساله حدود ۱٦ هزار نفر جان خود را از دست می‌دهند؛ بنابراین توسعه سامانه‌های دستیار راننده می‌تواند نقش مهمی در کاهش نرخ مرگ‌ومیر داشته باشد.

در مقالات مشابه، دوربینی برای دریافت تصاویر از نقاط کور بر روی خودرو تعبیه شده است. در تمامی مقالات ارائه‌شده تشخیص وسایل نقلیه در نقاط کور هدف‌گذاری شده است. در مقالات [۱،۲،۳،٤]، تحلیل با استفاده از یک قاب تصویر بوده است و در [۵] با کمک ویژگی‌های وابسته به جریان نوری تشخیص وسیله نقلیه صورت گرفته است. در تمامی این مقالات، برای هر تصویر وجود یا عدم وجود وسایل متحرک بدون توجه به سرعت نسبی آن‌ها موردتوجه قرار گرفته است. با توجه به جدول ۱ مشاهده می‌شود که تشخیص انواع اشیای متحرک نیز در مقالات یادشده مورد غفلت قرار گرفته است. در این مقاله برای حل این دو چالش از ردگیری اشیای متحرک در دنباله تصاویر استفاده شده است که می‌تواند یک نوآوری در سامانه‌های حمل‌ونقل هوشمند مبتنی بر پردازش تصویر ایجاد نماید.

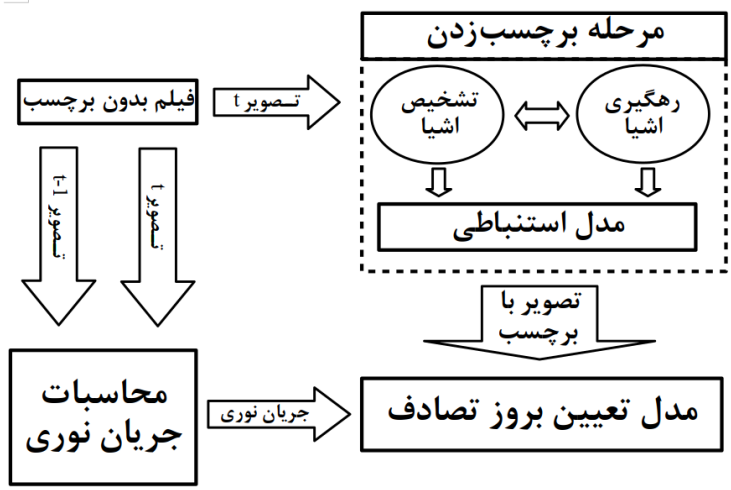
جدول ۱. ارزیابی مقالات مرتبط و مقایسه نقاط قوت و ضعف

| شماره مقاله | دقت گزارش‌شده | مجموعه داده | تشخیص موتور | تشخیص ماشین |
|---|---|---|---|---|
| ۱ | ۹۸ | تهیه‌شده صرفاً برای مقاله متناظر | - | * |
| ۲ | ۹۶ | تهیه‌شده صرفاً برای مقاله متناظر | * | - |
| ۳ | ۹۹ | تهیه‌شده صرفاً برای مقاله متناظر | - | * |
| ٤ | ۹۸ | تهیه‌شده صرفاً برای مقاله متناظر | - | * |
| ٥ | ۸۷ | تهیه‌شده صرفاً برای مقاله متناظر | * | * |

از دیدگاه ابزاری، روش‌های گوناگونی برای تشخیص اشیا در تصاویر و تحلیل پارامترهای مهم در حمل‌ونقل هوشمند توسعه داده شده است. به‌عنوان نمونه در [٦] از یک روش یادگیری عمیق ترکیبی با ماشین یادگیری تشدیدی برای شناسایی خصوصیات تصادفات بر اساس پردازش تصاویر پیشنهاد شده است. در این خصوص توجه به یک نکته بسیار اساسی است. در اکثر مقالات مرتبط که در جدول ۱ نشان داده شده‌اند، مجموعه داده صرفاً برای همان مقاله تولید شده است. لذا مقایسه روش‌های مزبور چندان منصفانه نیست. به‌علاوه احتمال بایاس شدن نتایج به سمت داده‌های جمع‌آوری‌شده، وجود دارد. برای رفع این مشکل، به‌تازگی مجموعه دادگان بزرگ همچون COCO به‌عنوان یک چارچوب یکسان برای توسعه روش‌های یادگیری و انجام مقایسه، توسعه داده شده است. این مجموعه داده شامل ۸۰ نوع از اشیا معمول برچسب‌گذاری شده در تصویر است که چالش‌های بزرگی در راه توسعه الگوریتم‌های پردازش تصویر ایجاد می‌کند. در این صورت می‌توان امیدوار بود که الگوریتمی که روی COCO نتایج مناسبی داشته باشد، در جهان واقع نیز جواب‌های مورد قبولی ایجاد می‌کند. به‌علاوه از این مجموعه دادگان، می‌توان استخراج دانش از تصویر را به‌خوبی انجام داد. در همین راستا، یادگیری عمیق در قالب پروژه‌های بسیار بزرگ و مقالات بسیار گوناگون روی این دادگان پیاده-سازی شده است. در مقاله حاضر نیز مجموعه داده COCO برای تشخیص وسایل نقلیه استفاده شده است. در این راستا از سامانه‌ها و الگوریتم‌های تشخیص اشیا در تصاویر همانند YOLO و FRCNN استفاده شده است.

## ۲. روش پژوهش

سامانه پیشنهادی همان‌گونه که در شکل ۲ نشان داده شده است، دارای سه بخش اصلی است. در بخش اول به توسعه یک مدل برای تشخیص انواع وسایل نقلیه اعم از خودروهای سواری، دوچرخه، موتور، کامیون، اتوبوس و حتی انسان می‌پردازد. تشخیص انسان برای تشخیص عبور انواع اسکیت سواران یا حتی دویدن عابران پیاده ضروری است.




**شکل ۲. نحوی برچسب زدن داده‌های بدون برچسب برای مدل پیشنهادی برای تعیین بروز تصادف**

در بخش دیگر با استفاده از خروجی مدل تشخیص وسایل نقلیه در یک تصویر و به کمک الگوریتم ردگیر در دنباله‌ای از تصاویر، سرعت و رفتار هر یک از اشیای تشخیص داده‌شده را بررسی می‌کنیم و با استفاده از مدلی استنباطی و مبتنی بر قانون، برای هر قاب تصویر، خطرناک بودن گردش به طرفین، به‌عنوان برچسب آن قاب تصویر مشخص می‌شود.

در بخش انتهایی با استفاده از برچسب آن قاب تصویر، جریان نوری و خود قاب تصویر، یک شبکه عصبی عمیق آموزش داده می‌شود که بسیار کوچک بوده و به‌منظور پردازش به‌صورت بلادرنگ بر روی CPU نیز مناسب است. این شبکه‌ی کوچک خطرناک بودن گردش به طرفین را مشخص می‌کند و در صورت نیاز می‌توان هشدار لازم را برای راننده ارسال کرد.

### ۲-۱ تشخیص اشیای تصویر

برای تشخیص اشیای موردنیاز در داخل تصویر، نوعی خاصی از YOLO با نام YOLOv3 استفاده شده است. این شبکه بسیار عمیق است و قابلیت پردازش ۲۰ تصویر در یک ثانیه را تنها با استفاده از کارت‌های گرافیک بسیار قدرتمند دارد. استفاده از این شبکه با این حجم محاسباتی، برای حل مسئله تشخیص خودرو در نقاط کور مقرون‌به‌صرفه نیست و لذا تنها در مرحله آموزش از آن استفاده شده است که بلادرنگ بودن شبکه ضروری نیست و زمان مصرفی برای یک قالب تصویر بر روی CPU تقریباً ۷.۰ ثانیه است.

در مقاله اصلی، این شبکه بر روی مجموعه داده COCO آموزش داده شده است که شامل ۸۰ کلاس است. همان‌طور که ذکر شد، اشیای موردنیاز برای این مسئله کامیون، خودرو، اتوبوس، موتور، دوچرخه و انسان است پس تمامی ۸۰ کلاس COCO موردنیاز نیست. برای استفاده از YOLO، داده‌های ٦ کلاس مرتبط با اهداف مقاله از COCO جدا و بقیه کلاس‌ها از این مجموعه داده حذف شده است. برای آموزش این شبکه بر روی این ٦ کلاس تغییرات کوچکی در ساختار مدل داده شد تا تشخیص اشیا برای تنها آن ٦ کلاس انجام پذیرد. خروجی مدل تشخیص اشیای موردنیاز به‌صورت لیستی از نام تمامی اشیای داخل تصویر، جعبه دربردارنده هر یک از اشیا و امتیاز تشخیص آن شیء است.

### ۲-۲ تحلیل اشیای داخل تصویر

همان‌طور که در قسمت قبل گفته شد ورودی الگوریتم تشخیص اشیا داخل تصویر تنها یک قاب تأثیر است؛ و برای بررسی اشیا در دنباله‌ای از تصاویر نیازمند یک الگوریتم ردگیر هستیم. دانلجان و همکاران یکی از بهترین الگوریتم‌های ردگیری اشیا را ارائه داده‌اند. در سامانه پیشنهادی مقاله نیز از این الگوریتم استفاده شده است. این الگوریتم به ما این امکان را می‌دهد که یک شیء

را در دنباله‌ای از تصاویر تعقیب کنیم و با استفاده از اندازه جعبه دربردارنده هرشی، می‌توان تخمین زد که آن شیء در حال نزدیک شدن یا دور شدن به دوربین است.

با استفاده از اندازه سرعت نسبی اشیای ردگیری‌شده و همچنین خروجی مدل تشخیص اشیا که شامل جعبه دربردارنده و نوع هر شیء است، مدل استنباطی مبتنی بر قانون تنظیم شده است که بر اساس یک سری قوانین اگر-آنگاه درباره بروز تصادف در هنگام گردش به طرفین و یا تغییر خط خودرو تصمیم می‌گیرد. تمامی قوانین این مدل استنباطی با نظارت فرد خبره تنظیم شده است.

### ۲-۳ مدل پیشنهادی برای نقاط کور

در این بخش به بررسی مدل کلی برای تعیین بروز تصادف در هنگام گردش به طرفین و یا تغییر خط خودرو می‌پردازیم. جزییات معماری این مدل در جدول ۲ ارائه شده است.

**جدول ۲. معماری پیشنهادی برای مدل تعیین بروز تصادف**

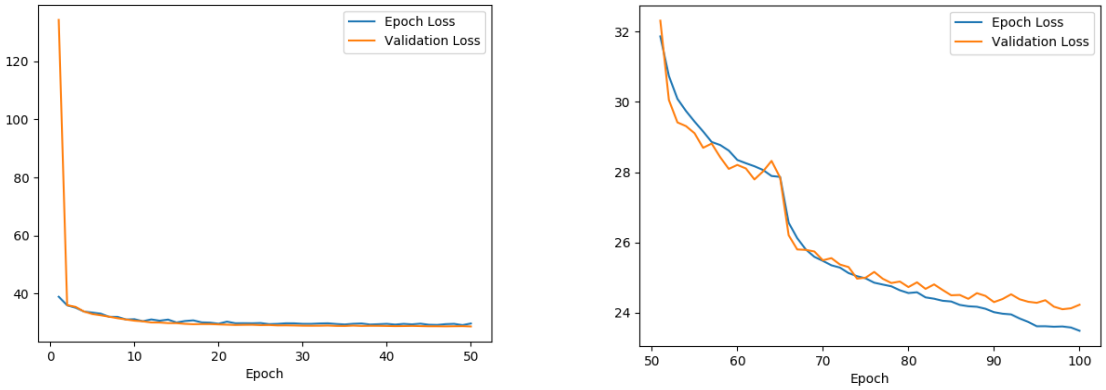
| شماره | نوع لایه | تعداد فیلترها | اندازه |
|---|---|---|---|
| ۱ | Conv2D | ۱٦ | (۳ ,۳) |
| ۲ | Conv2D | ۱٦ | (۳ ,۳) |
| ۳ | MaxPooling2D | | (۲ ,۲) |
| ٤ | Conv2D | ۳۲ | (۳ ,۳) |
| ٥ | Conv2D | ۳۲ | (۳ ,۳) |
| ٦ | MaxPooling2D | | (۲ ,۲) |
| ۷ | Conv2D | ٦٤ | (۳ ,۳) |
| ۸ | Conv2D | ٦٤ | (۳ ,۳) |
| ۹ | MaxPooling2D | | (۲ ,۲) |
| ۱۰ | Dense | | (۱ ,۳۲) |
| ۱۱ | Dense | | (۱ ,۱٦) |
| ۱۲ | Dense | | (۱ ,۱٦) |
| ۱۳ | Dense | | (۱ ,۱) |

مدل مذکور یک تنسور با شکل (۳، ۱۲۸، ۱۲۸) به‌عنوان ورودی دریافت می‌کند. این تنسور از ترکیب تصویر خاکستری و تک کانال با خروجی تنسور جریان نوری متراکم ساخته می‌شود؛ بنابراین می‌توان هر یک از اجزای ۱۲۸*۱۲۸ از تنسور ورودی را به شکل برداری که نشان‌دهنده شدت نور، سرعت جابه‌جایی در راستای طولی و سرعت جابه‌جایی در راستای عرضی برای هر یک از پیکسل‌ها است، در نظر گرفت. الحاق تصویر با تنسور جریان نوری متراکم باعث می‌شود که پیکسل‌های پس‌زمینه از اشیا متحرک داخل تصویر به‌راحتی متمایز شوند.

برای آموزش این شبکه همان‌طور که اشاره شد، به داده برچسب‌گذاری شده از قبل، نیاز نیست و با استفاده از چند ساعت فیلم ضبط‌شده و برچسب تعیین‌شده به کمک مدل استنباطی، این آموزش صورت می‌گیرد. فرایند آموزش در شکل ۲ ارائه شده است.

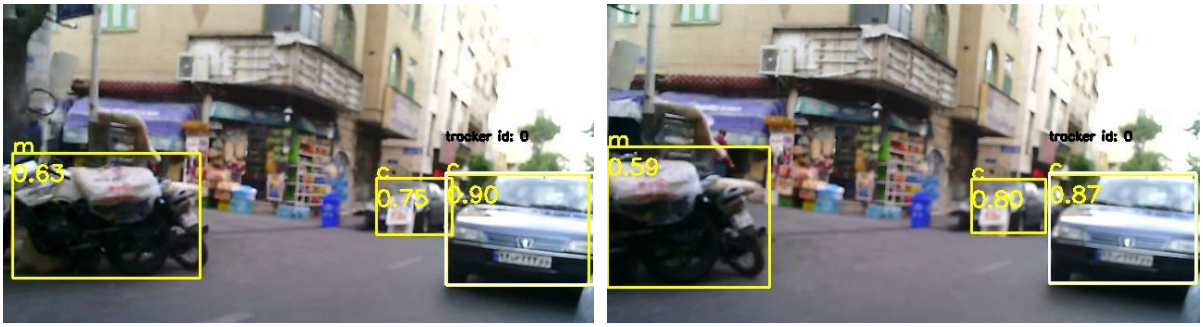
## ۳. تحلیل داده‌ها

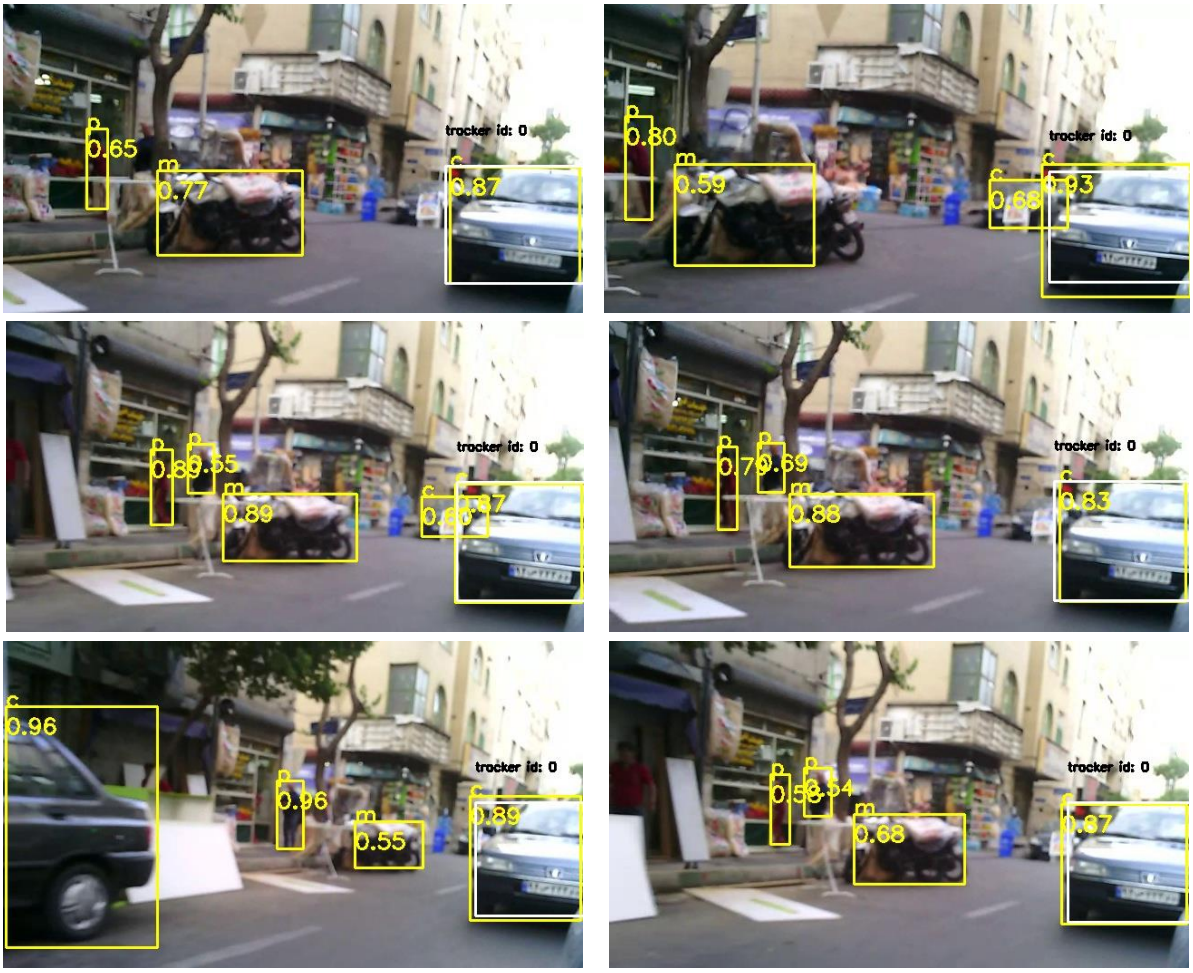
برای آموزش تشخیص اشیا در تصویر از دو کارت گرافیک GTX 1080 Ti استفاده شده است و فرایند آموزش به مدت دو ماه به طول انجامید. برای جلوگیری از بیش‌برازش و نابودی گرادیان، ابتدا لایه‌های ثانویه آموزش دیدند. سپس بعد از ۵۰ مرحله آموزش و هنگامی‌که نرخ خطا به ثبات رسید، یادگیری کل شبکه به همراه لایه‌های ابتدایی که وظیفه استخراج ویژگی را بر عهده دارند، با نرخ یادگیری اندک و به همراه برنامه‌ریزی نرخ یادگیر صورت پذیرفت. نمودار نرخ خطا در شکل ۳ ارائه شده است.




**شکل ۳. نمودار نرخ خطا در طول مراحل آموزش برای شبکه YOLOv3**

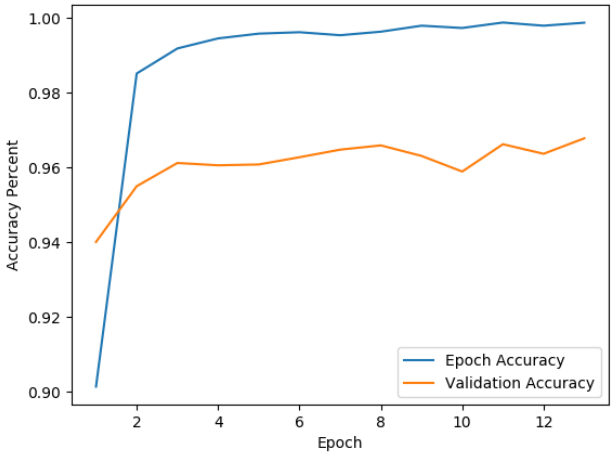
بعد از آموزش مدل تشخیص اشیا در تصویر، چندین فیلم خام و بدون برچسب از رانندگی در محیط‌های شهری و بزرگراه با کیفیت ۷۲۰ با دوربینی که بر روی آینه کناری خودرو نصب شده است، جمع‌آوری شد. با استفاده از مدل ردگیر و مدل استنباطی، این فیلم برچسب زده شده است. شکل ٤ چندین تصویر از فیلم جمع‌آوری‌شده را نشان می‌دهد که در آن جعبه‌های در بردارنده اشیایی که دارای امتیاز تشخیص بیشتر از ۵.۰ هستند با رنگ زرد نشان داده شده است. همچنین امتیاز تشخیص و نوع هر شیء نیز مشخص شده است. خروجی مدل ردگیر نیز با استفاده از رنگ سفید جعبه دربردارنده شیء مشخص شده است و شماره شناسایی ردگیری شیء که برای مدل استنباطی از اهمیت برخوردار است، نیز با رنگ سیاه مشخص شده است.






**شکل ٤. چند فریم متوالی از داده‌های ضبط‌شده. ماشین، موتور و انسان به ترتیب با حروف c، m و p مشخص شده است.**

برای آموزش شبکه عصبی عمیق جهت تعیین بروز تصادف نیز از روش‌های متعددی مانند انتخاب وزن‌های اولیه مناسب، برنامه‌ریز نرخ یادگیر و نرمال‌سازی ورودی استفاده شده است. بعد از برچسب زدن قاب‌های تصویر فیلم‌های خام ضبط‌شده، نیمی از تصاویر برای ارزیابی و نیمی دیگر برای آموزش به‌صورت تصادفی انتخاب شده‌اند. نمودار دقت مدل بر روی داده‌های یادگیری و ارزیابی در شکل ۵ آمده است و دقت ۹۶ درصد را به‌طور تقریبی نشان می‌دهد.




**شکل ۵. نرخ دقت آموزش و ارزیابی مدل تعیین بروز تصادف**

قابل‌توجه است که شبکه عصبی عمیق جهت تعیین احتمال بروز تصادف بسیار کوچک و با دقت بسیار بالا، هزینه پردازشی مقرون‌به‌صرفه‌ای دارد و تنها ۹ میلیون فلاپس هزینه پردازشی، بر روی ضعیف‌ترین پردازنده‌های CPU نیز می‌تواند ٤٠ تصویر در ثانیه را پردازش کند.

## ٤. نتیجه‌گیری

در راستای ارتقای ایمنی، در این مقاله یک سامانه دستیار راننده مبتنی بر دوربین منفرد منصوب روی آینه بغل خودرو پیشنهاد شده است. این سامانه جزییات در نقاط کور را ارائه می‌دهد. در این سامانه، از شبکه‌های عمیق توسعه‌یافته استفاده شده است. ورودی این شبکه، تصویر خاکستری و جریان نوری است. همچنین، اطلاعات دنباله‌ای از تصاویر به‌منظور ارزیابی ریسک بروز تصادف با توجه به اشیای متحرک در نقاط کور پردازش می‌شوند. در کاربرد، این سامانه با استفاده از تشخیص و ردگیری اشیا در تصویر، وجود وسایل نقلیه و سرعت نسبی آنها را تعیین کرده برای نقاط کور خودرو، ریسک بروز تصادف پیش‌بینی می‌شود. ادغام جریان نوری با تصویر و همچنین ترکیب نتایج با ویژگی‌های استخراج‌شده از یک شبکه عصبی عمیق، باعث توانمندی سامانه پیشنهادی شده است. آزمایش‌های انجام‌شده ثابت کرده است که دقت سامانه پیشنهادی برای تخمین تصادف با توجه به تصاویر دوربین بر روی آینه کناری خودرو، ۹٦ درصد بوده است.

## ٥. پی‌نوشت‌ها

1. Optical Flow

## ٦. مراجع


– Zhao, Y., Bai, L., Lyu, Y., & Huang, X. (2019). Camera-Based Blind Spot Detection with a General Purpose Lightweight Neural Network. Electronics, 8(2), 233.

– Baek, J. W., Han, B. G., Kang, H., & Chung, Y. (2017, October). Fast and reliable two-wheeler detection algorithm for blind spot detection systems. In 2017 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 513-516). IEEE.

– Sotelo, M. Á., & Barriga, J. (2008). Blind spot detection using vision for automotive applications. Journal of Zhejiang University-Science A, 9(10), 1369-1372

– Jung, K. H., & Yi, K. (2018). Vision-based blind spot monitoring using rear-view camera and its real-time implementation in an embedded system. Journal of Computing Science and Engineering, 12(3), 127-138

– Song, K. T., & Chen, H. Y. (2007, June). Lateral driving assistance using optical flow and scene analysis. In 2007 IEEE Intelligent Vehicles Symposium (pp. 624-629). IEEE.

– Pashaei, A., Ghatee, M., and Sajedi, H. (2019). Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images. Journal of Real-Time Image Processing: 1-16.

– ILin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

– Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

– Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

– Danelljan, M., Häger, G., Khan, F., & Felsberg, M. (2014). Accurate scale estimation for robust visual tracking. In British Machine Vision Conference, Nottingham, September 1-5, 2014. BMVA Press.

# Self-Supervised Blind Spot Detection System based on Image Processing by Deep Neural Network

Arash Pourhasan Nezhad, MSc from Department of Computer Science, Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran
Mehdi Ghatee*, Associate Professor with Department of Computer Science, Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran
Hedieh Sajedi, Associate Professor with Department of Computer Science, Faculty of Mathematics, Statistics and Computer Science, University of Tehran, Tehran, Iran

**E-mail: ghatee@aut.ac.ir**

## Abstract

In this paper, we propose a single camera-based driver assistance system for blind spots that is self-supervised and learn without labeled data. The proposed system is developed based on deep neural network and uses gray scale image and optical flow as input. The strength of this paper compared to similar articles is the processing of image sequence information to assess the risk of accident due to moving objects in blind spots.

In this paper, based on detection and tracking of the required objects, the existence of vehicles and its relative speed are estimated and the accident can be predicted for the blind spots of the car. The fusion of image with optical flow and features extracted using a deep neural network has increased robustness of the proposed system. In the proposed application of the article, the camera is mounted on the side mirror of the vehicle and is estimated to have a 60% accuracy, the risk of accidentally turning on the sides or changing lanes.

**Keywords:** Traffic light planning, Traffic Assignment, Data mining, Meta-heuristic Algorithms, Cooperative management, Master-slave