

## تشخیص اشیاء رها شده با استفاده از شبکه عصبی کانولوشنی بهبود یافته

محدثه هنرمند حقیقی<sup>۱</sup> و سارا معتمد<sup>۲</sup>

### چکیده

دوربین‌های نظارت تصویری می‌توانند به عنوان یک ابزار قدرتمند برای خودکارسازی تشخیص موقعیت‌های مختلف و کمک در جهت تصمیم‌گیری‌های مناسب به منظور افزایش سطح امنیتی و حفاظتی بکارگرفته شوند. یکی از مهمترین کاربردهای سیستم‌های نظارت تصویری، تشخیص اشیاء رها شده مانند چمدان‌های رها شده برای جلوگیری از بمب‌گذاری‌های خطرناک و سایر موارد است. در این راستا، در این مقاله یک مدل دومرحله‌ای مبتنی بر یادگیری عمیق جهت تشخیص اشیاء رها شده معرفی شده است. هدف مرحله اول تشخیص همه اشیاء ساکن در صحنه و مرحله دوم دسته‌بندی اشیاء رها شده است. در مرحله اول از مدل مخلوط گاوسی برای مدل‌سازی پس‌زمینه و تشخیص اجسام ساکن استفاده می‌شود. در مرحله دوم نیز برای مشخص کردن اشیاء رها شده از میان کلیه تصاویر استخراج شده از ترکیب شبکه عصبی کانولوشنی و الگوریتم آدابوست استفاده می‌شود. بر اساس نتایج ارزیابی‌ها مدل پیشنهادی از دقت بالاتری در تشخیص اشیاء رها شده نسبت به روش‌های پایه برخوردار است.

### کلید واژه‌ها

اشیاء مشکوک، دوربین‌های نظارتی، تفریق پس‌زمینه، شبکه‌های عصبی کانولوشنی بهبود یافته، الگوریتم آدابوست.

### ۱- مقدمه

دلیل این نوع حملات جامعه بهای گزافی را در رابطه با صدمات اقتصادی و محیطی، تضعیف روحیه ملی و مرگ انسان‌ها متحمل می‌شود. نه تنها روش تهدیدها عوض شده است، بلکه در حوادث اخیر هدف تهدیدها نیز تغییر کرده است. در حال حاضر، هدف حملات، زیربنای اقتصادی و شهروندان بی‌گناه هستند. سازمان‌هایی که مسئول نظارت زیربنای حساس اقتصادی و دارایی‌های ملی مانند فرودگاه‌ها، نیروگاه‌ها، ذخایر آب و مسیرهای حمل‌ونقل عمومی و غیره هستند، این نیاز را احساس می‌کنند که توانایی‌های خود را برای شناسایی اشیاء مشکوک و رفتارهای ناهنجار افزایش دهند و در زمان مناسب در مقابل آن عکس‌العمل نشان دهند [۱، ۲].

امروزه به دلیل افزایش نگرانی‌های امنیتی و هزینه‌های کم سخت‌افزاری، تقاضای رو به رشدی برای دستگاه‌های نظارت خودکار به وجود آمده است. تصاویر گرفته شده توسط دوربین‌های نظارتی عنصر اصلی هر سیستم نظارتی می‌باشند. به همین دلیل، پردازش بینایی کامپیوتری می‌تواند برای استخراج داده‌های مفید از

حوادث اخیری که در دنیا رخ داده است، دولت‌ها و همچنین سازمان‌ها را برآن داشته که درباره دستاوردهای مربوط به مباحث امنیتی تجدیدنظر کنند. بحران‌هایی که امروزه با آن‌ها روبرو هستیم دیگر مربوط به حملات گسترده نظامی دشمنان ما از خارج از مرزها نیست. امروزه ترس ما از این موضوع نشئت می‌گیرد که این امکان وجود دارد که گروهی هرچند کوچک، در داخل مرزها، این توانایی را داشته باشند که موجب بروز صدمات بسیاری شوند. به

این مقاله در بهمن‌ماه ۱۴۰۱ دریافت، در اردیبهشت‌ماه ۱۴۰۲ بازنگری و در خردادماه پذیرفته شد

<sup>۱</sup> دانشجوی کارشناسی ارشد مهندسی کامپیوتر، واحد فومن و شفت، دانشگاه آزاد اسلامی، فومن، ایران

رایانامه: [mohadeseh.h.haghghi@gmail.com](mailto:mohadeseh.h.haghghi@gmail.com)

<sup>۲</sup> گروه کامپیوتر، واحد فومن و شفت، دانشگاه آزاد اسلامی، فومن، ایران  
رایانامه: [motamed.sarah@gmail.com](mailto:motamed.sarah@gmail.com)

نویسنده مسئول: سارا معتمد

dorl.net/dor/20.1001.1.23831197.1402.10.3.6.8

این شناسایی شی مورد علاقه موفق عمل نمی‌کند. مشکلات این روش بیشتر ناشی از ظاهر و اندازه متفاوت اشیاء مختلف است. این روش همچنین به دلیل هزینه بالای محاسباتی برای استفاده در زمان واقعی مناسب نیست. دسته دوم استفاده از روش‌های تفریق پس‌زمینه برای شناسایی ناحیه ایستا و استفاده از آشکارساز شی برای تجزیه و تحلیل بیشتر است. این استراتژی اغلب مفید است و برای تشخیص شیء رها شده و استفاده‌های تجاری به اندازه کافی سریع است. اگرچه معمولاً موفق به تشخیص اشیاء، در صحنه‌های شلوغ و دارای انسداد یا مشکلات روشنی نیست [۱].

در کنار استفاده از این روش‌ها در سال‌های اخیر و با توسعه روش‌های یادگیری عمیق مسیر حرکت پژوهش در این حوزه به سمت استفاده از شبکه‌های عصبی عمیق برای حل این مسئله سوق پیدا کرده‌اند [۷, ۸]. یادگیری عمیق در واقع یادگیری به وسیله شبکه‌های عصبی است که دارای لایه‌های پنهانی زیاد می‌باشند. در یادگیری عمیق برای مثال یک تصویر را به لایه‌های مختلف تقسیم می‌کنند که مغز انسان هم این‌گونه عمل می‌کند و نورون‌های مغز به توده‌ها حساسیت دارند تا بتوانند به کل تصویر حساسیت نشان دهند و آن را پردازش کنند. شبکه‌های عصبی کانولوشنی یکی از این شبکه‌ها است که در کاربردهای مختلفی مانند بازشناسی حروف، اشیاء و غیره در سال‌های اخیر مورد استفاده قرار گرفته است. با اینکه استفاده از روش‌های یادگیری عمیق باعث بهبود قابل توجهی در حوزه بینایی ماشین شده است، تحقیقات انجام شده هنوز در ابتدای مسیر قرار دارند و نیاز به پیشرفت در این حوزه به شدت احساس می‌شود. با توجه به موارد بیان شده، پیشنهاد می‌شود از ظرفیت بالای شبکه‌های عصبی عمیق در یادگیری شناسه‌های پلاک استفاده شود. در این راستا هدف این مقاله نیز استفاده از شبکه‌های یادگیری عمیق برای تشخیص اشیاء رها شده است.

مدل پیشنهادی یک مدل دو مرحله‌ای است. مرحله اول شامل تشخیص اشیاء و مرحله دوم دسته‌بندی شیء رها شده است. برای تشخیص اشیاء ساکن و مدل‌سازی پس‌زمینه در مدل پیشنهادی از مدل مخلوط گاوسی (GMM) و برای دسته‌بندی اشیاء از ترکیب شبکه عصبی کانولوشنی و الگوریتم آدابوست استفاده می‌شود. با اینکه استفاده از شبکه‌های عصبی کانولوشنی در سال‌های اخیر منجر به افزایش قابل توجه دقت در حوزه دسته‌بندی تصاویر شده است و این شبکه‌ها می‌توانند ویژگی‌های محلی را استخراج کنند، اما تمامی ویژگی‌های استخراج شده توسط آن‌ها برای دسته‌بندی مناسب نیستند. برای حل این مشکل بر آن شدیم در مدل پیشنهادی ویژگی‌های حاصل از فیلترها با اندازه‌های مختلف حاصل از شبکه‌های عصبی کانولوشنی مختلف را به لایه‌های پولینگ و دسته‌بندی مجزا بدهیم. پس آن هنگامی که نتایج اولیه توسط هر کدام از دسته‌بندی‌ها به دست آمد، از مکانیسم آدابوست برای تولید نتایج کلی دسته‌بندی استفاده کنیم. مدل پیشنهادی روی سه مجموعه داده آموزشی استاندارد PETS2006، AVSS2007 و

این ویدئوها و تحلیل این اطلاعات با هدف مکانیزه کردن وظایف نظارتی و اعلام هشدار نسبت به رویدادهای پر ریسک و شناسایی اشیاء و رفتارهای مشکوک استفاده شود. بنابراین اگر دوربین‌های نظارت تصویری به جای تنها ضبط فیلم، بتوانند برای شناسایی و اقدام در زمان واقعی مورد استفاده قرار بگیرند، می‌توانند ابزاری به مراتب مفیدتر باشند. در نظارت تصویری، هدف اصلی تشخیص وقایع خاص برای کمک به کارکنان امنیتی است. اکثر سیستم‌های نظارتی که در حال حاضر استفاده می‌شوند، از عملکرد اپراتور انسانی استفاده می‌کنند و انتظار می‌رود که این افراد اغلب به طور هم‌زمان، تعداد زیادی صفحه‌نمایش که توسط دوربین‌های مختلف ضبط شده را تماشا کنند. یکی از وظایف اصلی کارکنان امنیتی، نظارت برای شناسایی رفتار و اشیاء مشکوک جهت ایجاد واکنش مناسب است. اما به مرور که تعداد دنباله‌های ضبط شده توسط دوربین‌های مدار بسته افزایش می‌یابد، وظیفه اپراتور بیشتر، سخت‌تر و خسته‌کننده می‌شود و دقت کاربر کاهش می‌یابد. بنابراین شیوه‌های تجزیه و تحلیل ویدئو که به طور خودکار جریان‌ها از ویدئویی را برای ایجاد هشدار، تجزیه و تحلیل کرده و اشیاء مشکوک و فعالیت‌های ناهنجار در حال وقوع را شناسایی می‌کنند. در هر سیستم نظارت تصویری، تشخیص مؤثر و کارآمد اشیاء رها شده یکی از وظایف مهم برای تأمین امنیت عمومی است. به عنوان مثال چمدان‌های رها شده برای جلوگیری از بمب‌گذاری‌های خطرناک و سایر موارد را می‌توان اشاره کرد [۳, ۴].

رویکردهای مختلفی جهت تشخیص خودکار شیء رها شده توسط دوربین‌های نظارتی برای استفاده‌های خاص نظیر امنیت عمومی، نظارت بر ترافیک، نظارت بر صحنه‌های شلوغ مثل فرودگاه و ایستگاه مترو و غیره ارائه شده است. اشیاء رها شده مثالی از رفتار غیر معمول در ناحیه نظارت هستند، یعنی جسمی که برای مدت معینی رها می‌شود و در حالت سکون باقی می‌ماند. تشخیص اشیاء رها شده بسیار مهم است زیرا ممکن است باعث مشکل امنیتی شود. تشخیص اشیاء رها شده یک کار چالش برانگیز در سیستم‌های نظارت تصویری است و تحقیقات زیادی برای بهبود و خودکارسازی نظارت انجام شده است.

اجسام رها شده را می‌توان به دو دسته تقسیم کرد: اشیاء خطرناک و اشیایی که خطرناک نیستند. اشیایی که خطرناک نیستند معمولاً برای مدتی رها می‌شوند و سپس برداشته می‌شوند، مثل یک چمدان رها شده که توسط صاحبش تحت نظارت است و برای برداشتن آن برمی‌گردد [۵, ۶]. تشخیص اشیاء رها شده یک رویداد مهم است که بسیاری از محققان در تلاش برای حل آن هستند و در صحنه‌های شلوغ چالش‌های متعددی نظیر سرعت و دقت تشخیص شیء رها شده مشکوک و همچنین انسداد وجود دارد. روش‌های تشخیص شیء رها شده را غالباً می‌توان به دو دسته تقسیم کرد. دسته اول استفاده از آشکارساز جسم و تکنیک‌های استخراج ویژگی برای مکان‌یابی و پیدا کردن اشیاء مورد علاقه در صحنه است. اگرچه این روش دقت خوبی دارد اما در برخی مواقع

به صورت اتوماتای حالت محدود<sup>۹</sup> مدل سازی کردند، حال آنکه گروهی دیگر از استنتاج زمانی مبتنی بر منطق به عنوان راه حل جایگزین استفاده نمودند. کونگ و همکاران [۱۰]، متفاوت از همه پژوهش های دیگر، تلاش کردند تا اشیاء رها شده در جاده را به کمک یک دوربین متحرک تشخیص دهند. فریمن و دیگران [۱۱] یک الگوریتم ارزیابی تهدید ارائه کردند تا مفهوم مالکیت<sup>۱۰</sup> را با فهم خودکار ارتباطات اجتماعی ترکیب کرده و رهاشدگی اشیاء را نتیجه گیری کنند. در ادامه، سیزوچ [۱۲] الگوریتمی را برای تشخیص نواحی ساکن به کمک مقایسه این نواحی با کانتورهای<sup>۱۱</sup> اشیاء در حال حرکت، مطرح نمود. داهی و همکاران [۱۳] روشی را بر اساس روش های تشخیص لبه ساکن و طبقه بندی، ارائه کردند. فام و دیگران [۱۴] یک روش دو مرحله ای برای تشخیص اشیاء ای که مورد اعتنا قرار نمی گیرند، پیشنهاد نمودند. مدل پیشنهادی آن ها در قدم اول تلاش می کرد تا تمامی اشیاء رها شده ممکن، تشخیص داده شوند تا از مثبت کاذب ها جلوگیری شود. در گام دوم، روش آن ها هشدارهای کاذب را کاهش می دهد. این کار با تطبیق مشابهت ها<sup>۱۲</sup> بین نتایج مرحله اول و مدل پیش زمینه انجام می شود.

فن و همکاران [۱۵] نرخ هشدار کاذب را با الگوبرداری از یک شیء موقت ایستا، مانند خودرویی که مدتی متوقف می شود و سپس دوباره حرکت می کند، با استفاده از یک مدل واحد پس زمینه و ماشین حالت نامحدود (FSM) کاهش دادند. مطالعه آن ها همچنین شیء نامشخص را را که قبلاً در پس زمینه جذب شده بود در نظر گرفت، اما مسئله تغییر روشنایی را مورد بررسی قرار نداد. از سوی دیگر، تیان و همکاران [۱۶] از یک مدل پس زمینه متشکل از سه مدل مخلوط گوسی استفاده کردند. در مطالعه آن ها اطلاعات شدت و بافت برای حذف سایه ها و مقاوم در برابر تغییرات روشنایی یکپارچه شده است و در آن از پس زمینه برای تشخیص وقوع تغییرات روشنایی استفاده می شود. لازم به ذکر است که تاریخچه پس زمینه به گونه ای ذخیره می شود که الگوریتم تطبیق در برابر تغییرات قوی کار کند. با بازگرداندن پیش زمینه قبلی، قبل و بعد از تغییر نور به طور موقت تغییر می کند، اما آن ها نمی توانند تغییرات مداوم را کنترل کنند. در مدل پیشنهادی آن ها اگر تغییرات پس از رها شدن شیء و قبل از وقوع زنگ رویداد رخ دهد، ردیابی می تواند به دلیل از دست دادن اطلاعات پیش زمینه موجود، شکست بخورد. مدل پس زمینه دوگانه شامل یک مدل کوتاه مدت با سرعت یادگیری سریع و یک مدل بلند مدت با سرعت یادگیری کند است [۱۷]. اجسام ثابت از تفاوت بین این دو مدل استخراج می شوند. یکی از معایب نرخ مثبت کاذب بالا ناشی از ناقص بودن مدل پس زمینه است.

ABODA مورد ارزیابی قرار گرفت و بر اساس نتایج آزمایشات، مدل پیشنهادی از دقت بالاتری نسبت به سایر روش های پایه برخوردار است.

ادامه این مقاله به صورت زیر سازماندهی شده است. بخش دوم از این مقاله به کارهای پیشین اختصاص دارد. در بخش سوم روش پیشنهادی تشریح می گردد. آزمایشات و نتایج بررسی ها در بخش چهارم به تفصیل بیان شده است. بخش پنجم نیز شامل نتیجه گیری و کارهای آینده است.

## ۲- پیشینه پژوهش

تشخیص اشیاء یکی از الگوریتم های پردازش تصویر است که می تواند برای پیدا کردن اشیاء ساکن در تصویر مورد استفاده قرار گیرد. برای اینکه این الگوریتم بتواند شیء را در تصویر پیدا کند، نیاز است که یک مدل از آن شیء را بسازد. در مرحله بعد الگوریتم شروع به جستجو در عکس می نماید. الگوریتم شناسایی اشیاء عکس را جستجو می کند تا شیء مورد نظر را پیدا کند. پیدا کردن یک شیء در پردازش تصویر می تواند با اهداف متفاوتی مانند اعلام مختصات قطعه به تجهیزاتی دیگر (مثل ربات و...)، شناسایی موقعیت یک شیء در سیستم های ردیابی موقعیت، شناسایی اشیاء رها شده و ده ها کاربرد دیگر به کار رود. با توجه به اینکه تأکید این پژوهش بر روی شناسایی اشیاء رها شده است، مطالعات انجام شده در این حوزه به صورت خلاصه مورد بررسی قرار گرفته است.

اکثر مطالعات موجود برای تشخیص اجسام رها شده، اطلاعات پیش زمینه تولید شده از یک یا چند مدل پس زمینه را برای تمایز بین اجسام ثابت و متحرک تجزیه و تحلیل کرده و اجسام ساکن را برای مدت معینی ردیابی می کند تا مشخص شود که آنها رها شده اند یا خیر. در اکثر پژوهش ها در زمینه تشخیص اشیاء رها شده، از روش تفریق پس زمینه به عنوان یک گام مقدماتی سطح-پایین استفاده می شود تا اشیاء یا نواحی پیش زمینه تشخیص داده شوند. البته اسمیت و همکاران [۹]، مستقیماً با دنبال کردن<sup>۱</sup> چندین شیء در صحنه به کمک زنجیره مارکوف مونت کارلو<sup>۲</sup> (MCMC) فرا بعدی<sup>۳</sup> کار را آغاز کردند. پس از تفریق پس زمینه، برخی از پژوهش ها تلاش کردند تا تشخیص های مثبت کاذب<sup>۴</sup> را با روش های دنبال کردن اشیاء<sup>۵</sup> و طبقه بندی<sup>۶</sup> کاهش دهند در حالی که برخی دیگر چارچوب های تشخیص لبه<sup>۷</sup> یا مدل های مولد<sup>۸</sup> را به کار گرفتند. بعضی چارچوب ها مسئله تشخیص اشیاء رها شده را

<sup>1</sup> Tracking

<sup>2</sup> Markov Chain Monte Carlo

<sup>3</sup> Trans-dimensional

<sup>4</sup> False positive

<sup>5</sup> object tracking

<sup>6</sup> Classification

<sup>7</sup> Edge detection frameworks

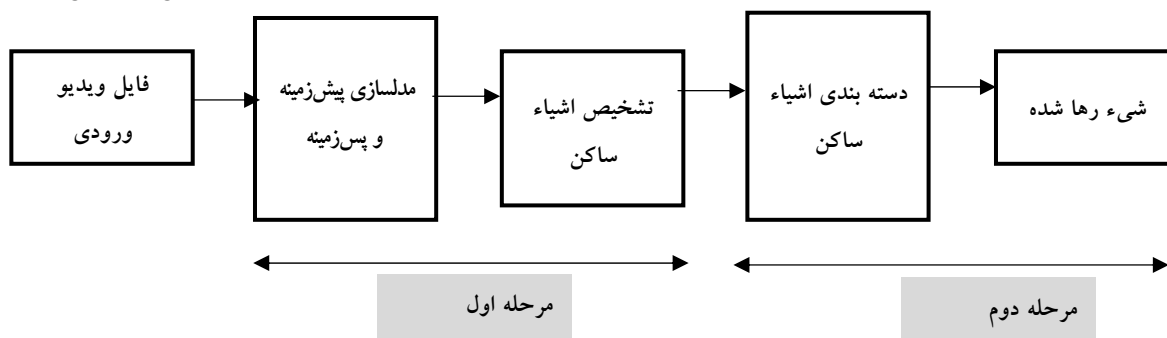
<sup>8</sup> Generative models

<sup>9</sup> Finite state automata

<sup>10</sup> Ownership

<sup>11</sup> Contours

<sup>12</sup> Similarity matching



شکل (۱): مراحل کلی مدل پیشنهادی

زمینه جذب نمی‌کند. این مدل می‌تواند یک شیء ثابت را برای مدت طولانی ردیابی کند و انسداد نیز به راحتی قابل کنترل است. با این حال، مدل نمی‌تواند تغییرات روشنایی را به طور مناسب جبران کند. اگر تغییرات اتفاق بیفتند، تصاویر پیش زمینه جدید تأثیرگذار در مدل بلند مدت بدون جذب باقی می‌مانند و تعداد زیادی اجسام ثابت اشتباه تولید می‌کنند.

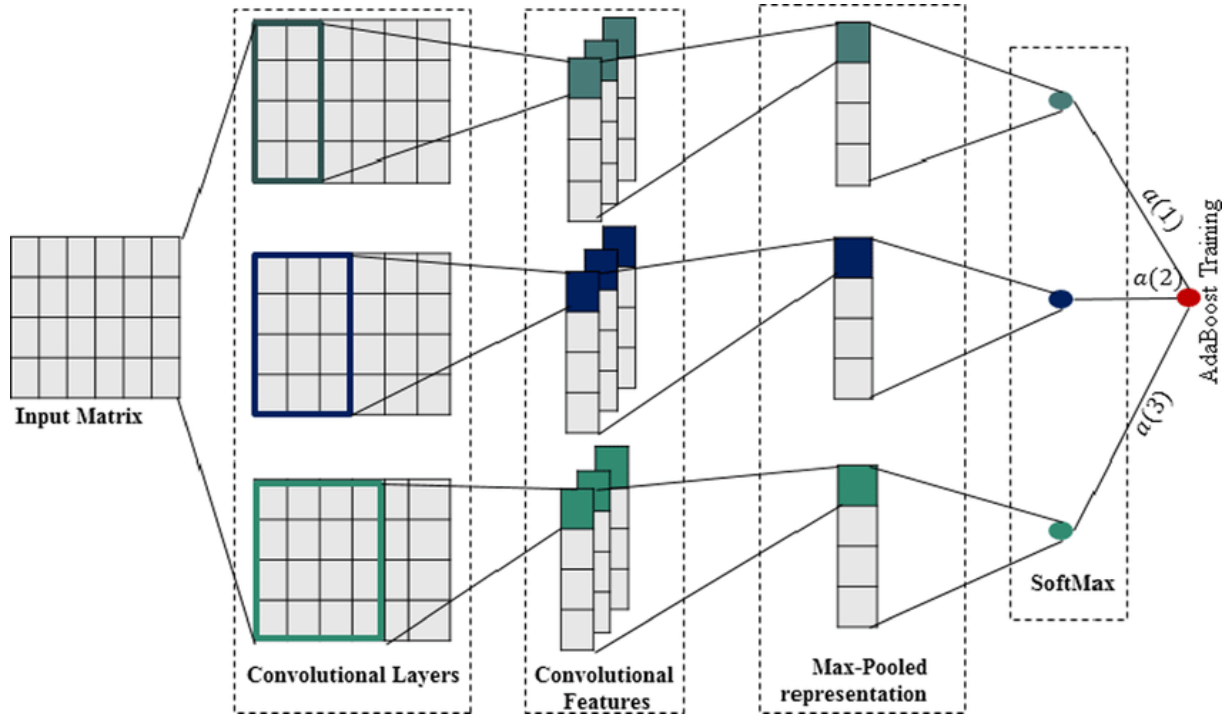
با توسعه مدل‌های یادگیری عمیق، استفاده از این شبکه‌ها در حوزه تشخیص اشیاء رها شده نیز در مرکز توجه گرفت. سومران و همکاران [۷] از یک رویکرد دو مرحله‌ای استفاده کردند که همه اجسام ثابت را در مرحله اول تشخیص داده و در مرحله دوم موارد رها شده را از اشیاء شناسایی شده انتخاب می‌کرد. آنها با آموزش تصویر فرد دارای یک شیء به عنوان یک تصویر منفی، همه اشیاء نزدیک فرد را مورد توجه قرار می‌دهند. آنها همچنین تصاویر ترکیبی از تصاویر مختلف چمدان را که در مکان‌های تصادفی در پس زمینه اندازه‌گیری شده قرار می‌گیرند، ایجاد می‌کنند و با برچسب‌گذاری به عنوان یک تصویر مثبت، مدل تشخیص را آموزش می‌دهند. این روش به بیش از حد از لحاظ تکنیک‌های یادگیری عمیق توجه دارد و در سناریوها، به نظر نمی‌رسد که تشخیص و تغییر روشنایی مالک در نظر گرفته شود. سیداکین و همکاران [۲۲] نیز مدل تشخیص را با استفاده از تصاویر مختلف چمدان آموزش می‌دهند و در عین حال، تصاویر دقیق اشیاء در زمینه‌ای خاص که به عنوان تصاویر پس زمینه برچسب‌گذاری شده‌اند را مشخص می‌کنند.

با اینکه استفاده از یادگیری عمیق باعث پیشرفت قابل توجهی در این حوزه شده است، اما یکی از مشکلات رایج در روش‌های یادگیری عمیق در تشخیص اشیاء، مشکل تحلیل دقیق و قابل توجه برای شناسایی ویژگی‌های مهم و مفید است. به عنوان مثال، در صورتی که یک سیستم تشخیص اشیاء تنها با تصاویری آموزش دیده باشد که در شرایط خاصی گرفته شده‌اند، ممکن است در تشخیص اشیاء در شرایط دیگر ناموفق باشد. این مسئله به دلیل عدم تعمیم‌پذیری و جزئیات بیش از حد در داده‌های آموزشی رخ می‌دهد. علاوه بر این، مسئله تعداد کم داده‌ها، نیز یکی دیگر از مشکلات این روش‌ها است. در بسیاری از موارد، تعداد داده‌های موجود برای هر دسته از اشیاء کافی نیست و باعث می‌شود که دقت سیستم در تشخیص و دسته بندی اشیاء کاهش یابد. به طور کلی، مشکلات روش‌های یادگیری عمیق در تشخیص اشیاء شامل

در راستای حل مشکل عدم توجه به تداوم تصاویر مطالعات مختلفی انجام شد. در این راستا، لین و همکاران [۱۸] مجموعه داده شیء رها شده (ABODA) را ایجاد کرد. مطالعه آن‌ها تلاش کرد با تشخیص وضعیت پیکسل تصاویر پی در پی بلند مدت و کوتاه مدت پیوسته برای مدت زمان کافی با استفاده از دستگاه حالت پیکسل (PFSM)، عملکرد تشخیص شیء را بهبود بخشد. روش آنها شامل پیگیری وضعیت پیکسل تصاویر پیش زمینه تا پایان زمان سنج و تعیین این که آیا گروه پیکسل‌هایی که تا این زمان حفظ شده‌اند یک شیء رها شده هستند یا خیر بود. یکی از مشکلات این شیوه این است که ردیابی زمانیکه جسم دیگری شیء مورد نظر را مسدود کند، پیچیده می‌شود. مشکل دیگر این است که مدل پس زمینه به تدریج پیش‌زمینه را به پس‌زمینه جذب می‌کند و پس از پایان جذب، ردیابی غیرممکن خواهد بود که منجر به کاهش دقت ردیابی می‌شود.

این شیوه‌ها در مورد تغییرات روشنایی سناریوها موثر نبودند و در این راستا واهیونگتال [۱۹] و فیلونگتال [۲۰] به جای استفاده از مدل‌های کوتاه مدت و بلند مدت، از مدل پشت زمینه برای جدا سازی پس زمینه مدلی فعلی استفاده کردند. در مطالعه آن‌ها از تکنیک مدل سازی پس زمینه سازگار برای مقاومت در برابر تغییرات موقت روشنایی استفاده شد، اما در سناریوهای بزرگ روشنایی مداوم ABODA شکست خورد. شامتال و همکاران استخراج اشیاء ثابت از مدل پس‌زمینه دوگانه را با استفاده از روش مدل‌سازی پس‌زمینه sViBe که ترکیب ViBe مبتنی بر رنگ و SILTP مبتنی بر بافت بود، انجام دادند. مدل پیشنهادی آن‌ها عملکرد فوق‌العاده‌ای را در سناریوهای تغییر روشنایی ABODA نشان داد. با این حال، الگوریتم‌های ویروسی نیز بستگی زیادی به اطلاعات پیش زمینه دارند و مانند مدل دیگر، از مدل پس زمینه تولید می‌شوند. بنابراین، به نظر می‌رسد بعید است که آنها در برابر تغییرات روشنایی که پس از بی‌توجهی یک شیء و قبل از فعال شدن زنگ هشدار رخ می‌دهد، قوی عمل کنند.

در ادامه، کواس و همکاران [۲۱] یک مدل پس زمینه سه گانه متشکل از سه مدل کوتاه مدت، میان مدت و بلند مدت را پیشنهاد کردند تا امکان ردیابی یک شیء برای یک دوره طولانی فراهم شود. در این مدل، مدل کوتاه مدت پیش زمینه را سریعتر از مدل میان مدت جذب می‌کند. مدل بلندمدت هرگز اجسام ثابت پیش‌زمینه را بدون توجه به مدت زمانی که جابجا نشده‌اند، به پس



شکل (۲): شبکه عصبی کانولوشنی تلفیق شده با الگوریتم آدا بوست برای دسته بندی تصاویر

### ۱-۳- مرحله اول: تشخیص اشیاء

در این مرحله از مدل مخلوط گاوسی (GMM) برای مدل سازی پس زمینه و تشخیص اجسام ساکن استفاده می شود. به طور کلی، چون امکان تخمین یک شیء ساکن در پس زمینه در زمان واقعی وجود دارد، روش مدل سازی پس زمینه برای تشخیص شیء مناسب است. آن ها همچنین در برابر نسبت نویز دوربین، تغییرات آهسته در شرایط نور و حرکت دوره ای پس زمینه های درهم نیز مقاوم هستند.

بر اساس مدل مخلوط گاوسی، احتمال اینکه یک پیکسل خاص در فریم در زمان  $t$  مقدار  $X_t$  را به عنوان مجموع  $K$  وزن گاوسی به خود بگیرد برابر است با:

$$P(X_t) = \sum_{k=1}^K \omega_{k,t} \eta_{k,t}(X_t, \mu_{k,t}, \Sigma_{k,t}) \quad (1)$$

$$\eta_{k,t} = N(\mu_{k,t}, \Sigma_{k,t}), \Sigma_{k,t} = \sigma_{k,t}^2 \cdot I \quad (2)$$

در اینجا  $\omega_{k,t}$  وزن  $k$ -امین گوسین در مخلوط،  $\eta_{k,t}(X_t, \mu_{k,t}, \Sigma_{k,t})$  توزیع نرمال، مقدار میانگین و  $\Sigma_{k,t}$  ماتریس کوواریانس وابسته به کانال رنگی تصاویر و  $\sigma_{k,t}^2$  نشان دهنده نوع تصویر و  $I$  ماتریس واحد است. گوسین در مخلوط دنباله ای از مقادیر  $\frac{\omega_{k,t}}{\sigma_{k,t}}$  است. در رویکرد دیگر مدل پس زمینه گاوسی، ابتدا ویژگی های پیش زمینه یا اجسام متحرک طبقه بندی می شوند.

مسائلی مانند عدم تعمیم پذیری، کمبود داده ها، نویز و ابهام در داده ها و همچنین مشکلات مربوط به ساختار و پیچیدگی شبکه های عصبی است. با این حال، این روش ها همچنان به صورت فزاینده ای بهبود یافته و جایگاه خود را در زمینه تشخیص اشیاء پررنگ تر کرده اند.

### ۳- مدل پیشنهادی

شیء رها شده مصداق جسمی که برای مدت معینی رها می شود و در حالت سکون باقی می ماند. تشخیص اشیاء رها شده یک کار چالش برانگیز در سیستم های نظارت تصویری است و در سال های اخیر تحقیقات زیادی برای بهبود و خود کار سازی فرآیند تشخیص اشیاء انجام شده است. با اینکه مطالعات مختلفی در این حوزه صورت گرفته و بکارگیری شبکه های یادگیری عمیق نیز باعث پیشرفت قابل توجهی در این زمینه شده است، هنوز در ابتدای مسیر پیشرفت بوده و نیاز به پژوهش در این حوزه همچنان احساس می شود.

در این راستا، در این مقاله یک مدل دومرحله ای برای تشخیص اشیاء رها شده معرفی شده است. هدف مرحله اول تشخیص همه اشیاء ساکن در صحنه است. با توجه به اینکه از میان این اشیاء، تعداد زیادی تشخیص مثبت کاذب وجود خواهد داشت که در واقع اشیایی به غیر از چمدان رها شده هستند. مثلاً افرادی که به طور ثابت ایستاده اند یا حتی چمدان های رها نشده، مرحله دوم مدل پیشنهادی تشخیص بین اشیاء رها شده از اشیاء دیگر ضروری است. مراحل کلی مدل پیشنهادی در شکل ۱ نشان داده شده است و جزئیات آن در ادامه بیان شده است.

نتایج حاصل از دسته‌بندی به کمک الگوریتم آدابوست با هم ادغام شوند تا بهترین دسته‌بند حاصل شود.

ساختار شبکه عصبی کانولوشنی تلفیق شده با الگوریتم آدابوست برای دسته بندی شیء رها شده در شکل ۲ نشان داده شده است. همان‌طور که مشخص است یک لایه کانولوشنی از دو مرحله تشکیل شده است:

در مرحله اول برای انجام عملیات کانولوشن یک فیلتر  $F \in R^{d \times h}$  روی مجموعه ورودی اعمال می‌شود. در صورتی که فرض شود  $A[i:f]$  پنجره‌ای شامل بردار ویژگی‌هایی باشد که میان  $i$  امین و  $f$  امین سطر ماتریس  $A$  قرار دارند، ویژگی  $\bar{c}_i$  بر اساس پنجره‌ای از ویژگی‌ها با اندازه  $h$  بر اساس رابطه (۴) محاسبه می‌شود [۲۳].

$$\bar{c}_i = g(w \circ F[i:t+c-1] + b) \quad (4)$$

در این رابطه متغیر  $i$  به صورت  $i = 1 \dots n - c + 1$  مقدار می‌گیرد و  $F \in R^{d \times h}$  و  $b \in R$  نیز به ترتیب نشان‌دهنده ماتریس فیلتر کانولوشن و عملیات بایاس هستند.  $g$  و  $\circ$  نیز به ترتیب توابع فعال‌سازی (مانند Tanh و Relu) و عملگر کانولوشن هستند. با اعمال فیلتر  $F$  به تمامی پنجره‌های ممکن از تصاویر ورودی و قرار دادن ویژگی‌های حاصل از آن‌ها در یک بردار یک نداشت ویژگی به صورت  $C = [\bar{c}_1, \bar{c}_2, \dots, \bar{c}_{n-c+1}]$  تشکیل خواهد شد که در آن  $C \in R^{n-c+1}$  است.

سپس ویژگی‌های استخراج شده به لایه ادغام داده می‌شوند. هدف این لایه حفظ بهترین ویژگی‌ها به نحوی است که ارتباط پیوسته ویژگی‌ها حفظ شود. به طور کلی در لایه ادغام، بلوک مستطیل‌های کوچکی که از لایه کانولوشنی دریافت می‌شوند، نمونه‌برداری شده و یک خروجی منفرد از آن بلوک ایجاد می‌شود. روش‌های مختلفی برای انجام عملیات ادغام وجود دارد مانند میانگین‌گیری، بیشینه‌گیری و یا یک ترکیب خطی از نورون‌های داخل بلوک. در روش پیشنهادی ما از ادغام به روش بیشینه‌گیری استفاده کرده و بیشترین مقدار را بین ویژگی‌ها انتخاب می‌کنیم (رابطه ۵).

$$c_{max} = \max\{C\} = \max\{c_1, \dots, c_{n-h+1}\} \quad (5)$$

لازم به ذکر است که لایه ادغام منجر به کاهش اندازه ویژگی‌ها و پیرو آن محاسبات آینده می‌شود. همچنین لایه ادغام انتقالات را بدون تغییر بدست می‌آورد و این اطمینان را می‌دهد که شبکه در مقابل موقعیت دارای مقاومت است. در ادامه ویژگی‌های ادغام شده توسط یک تابع غیر خطی قبل از اینکه وارد دسته‌بند شوند، مورد پردازش قرار می‌گیرند.

برای مقابله با مشکل بیش‌برازش که تمامی شبکه‌های عصبی به عنوان یک نقطه ضعف به حساب می‌آید، از روش از قلم

$$B = \underset{b}{\operatorname{argmin}} (\sum_{k=1}^b \omega_{k,t} > T) \quad (3)$$

در اینجا  $T$  آستانه حساسیت است که به صورت دستی انتخاب می‌شود. هنگامی که یک قاب جدید ظاهر شود، شدت مقادیر پیکسل انتخاب شده در مقایسه با مخلوط گاوسی و پارامترهای مخلوط گاوسی به روزسانی می‌شود.

## ۲-۳- مرحله دوم: دسته‌بندی شیء رها شده

در این مرحله برای مشخص کردن اشیاء رها شده از میان کلیه تصاویر استخراج شده از ترکیب شبکه عصبی کانولوشنی و الگوریتم آدابوست استفاده می‌شود. هدف استفاده از شبکه عصبی کانولوشنی استخراج ویژگی‌های سطح پایین تصویر است. به طور کلی، شبکه عصبی کانولوشنی، یک نوع از شبکه‌های عصبی عمیق است که برای پردازش تصاویر و شناسایی الگوهای مختلف در آن‌ها استفاده می‌شود. ساختار این شبکه‌ها شامل لایه‌های مختلفی است که هر کدام وظیفه خاص خود را دارند. در ورودی این شبکه‌ها، تصویر به صورت ماتریسی ارائه می‌شود و با استفاده از فیلترهای کانولوشنی، ویژگی‌هایی مانند لبه‌ها و نقاط روشن و تاریک را استخراج می‌کند. سپس با استفاده از لایه‌های مختلف، این ویژگی‌ها با هم ترکیب و به عنوان ویژگی‌های بالاتر به لایه‌های بعدی منتقل شده و در نهایت با استفاده از لایه‌های کاملاً متصل، دسته‌بندی تصاویر انجام می‌شود. در واقع، شبکه‌های عصبی کانولوشنی از لایه‌هایی به همراه فیلترهای کانولوشنی استفاده می‌کنند که این فیلترها به ویژگی‌های محلی اعمال می‌شوند. در واقع، در این نوع شبکه‌ها به جای اینکه تمامی نورون‌های یک لایه به لایه بعد متصل باشند، تنها قسمتی از نورون‌های یک لایه به لایه بعد متصل هستند. در نتیجه این نوع شبکه‌های عصبی قابلیت این را دارند که بتوانند ویژگی‌های جدیدی با استفاده از ویژگی‌های موجود استخراج کنند که اصطلاحاً به آن مهندسی ویژگی‌ها گفته می‌شود که در رویکردهای یادگیری ماشین به صورت دستی انجام می‌شود. در حوزه پردازش تصویر با استفاده از این نوع شبکه‌ها دیگر لازم نیست که ویژگی‌های جدید دستی استخراج شوند و به کمک این شبکه‌ها می‌توان ویژگی‌های سطح پایین مانند لبه‌ها را بطور خودکار استخراج کرد. لازم به ذکر است که استفاده از فیلترها با اندازه‌های مختلف روی شبکه‌های کانولوشنی منجر به ایجاد ویژگی‌های مختلفی می‌شوند که هر کدام از این ویژگی‌ها ممکن است تاثیر متفاوتی روی دسته‌بندی نهایی داشته باشند. در روش کلاسیک شبکه‌های عصبی کانولوشنی این ویژگی‌ها پس از اعمال عملیات پولینگ با هم ادغام شده و دسته‌بندی روی ویژگی‌های ادغام شده انجام می‌شود. این مساله می‌تواند تاثیر برخی از ویژگی‌ها که دارای اطلاعات مناسب‌تری هستند را از بین ببرد. در همین راستا ما تصمیم گرفتیم که ویژگی‌های حاصل از هر فیلتر، لایه پولینگ و دسته‌بندی جداگانه‌ای داشته باشند. پس از آن

شبکه مرتبط با آدابوست به کمک بازگشت به عقب<sup>۴</sup> آموزش می‌بیند می‌بیند و فرآیند آموزش آدابوست به شرح زیر است [۲۵]:

(۱) اختصاص توزیع  $D^1$  برای نمونه‌های آموزشی به صورت یکسان،  $D^t$  نشان دهنده  $t$ مین توزیع نمونه‌های آموزشی است. برای هر شاخص  $i$  داریم  $D_i^1 = \frac{1}{\#training\_samples}$

(۲) در یک دوره<sup>۵</sup> آموزشی  $t$ : سه دسته‌بند شبکه به طور پیوسته به کمک بازگشت به عقب آموزش می‌بینند، در نتیجه فرآیندهایی که در ادامه به آن‌ها اشاره شده است روی تمامی دسته‌بندها در هر دوره اعمال می‌شود.

(الف) تخمین آمار دسته‌بند ضعیف: پس از آن‌ها دسته‌بندها آموزش دیده و برچسب جملات پیش‌بینی شد، معماری پیشنهادی آمار مربوط به دسته‌بندی را در نمونه‌های آموزشی ذخیره کرده و خطای دسته‌بند ضعیف  $G_m(x)$  را بصورت رابطه (۷) محاسبه می‌کند.

$$e_m^t = \sum_i D_i^t 1(G_m(x) \neq y(x)) \quad (7)$$

(ب) تنظیم وزن: هر بار که یک دسته‌بند ضعیف آموزش می‌بیند، معماری پیشنهادی از خطای دسته‌بندی برای تغییر توزیع روی مجموعه آموزشی استفاده کرده و شاخص خطا و وزن دسته‌بند ضعیف را محاسبه می‌کند. محاسبه وزن دسته‌بند (رابطه ۸):

$$a(m) = \frac{1}{2} \ln \frac{1 - e_m^t}{e_m^t} \quad (8)$$

تنظیم توزیع (رابطه ۹)

$$D_i^{t+1} = \frac{D_i^t \exp(-a(m)y(x)G_m(x))}{D^t} \quad (9)$$

(ج) ارزیابی عملیات انجام شده توسط آدابوست: پس از آموزش دسته‌بند ما ضرب بین عناصر خروجی ها و وزن‌ها را اعمال کرده و دسته نهایی که تصویر به آن تعلق دارد را پیش‌بینی می‌کنیم. ما از وزن آموزش دیده  $a$  برای ارزیابی آدابوست به کمک رابطه (۱۰) استفاده می‌کنیم.  $t$  نشان دهنده شاخص دسته‌بند،  $a$  نشان دهنده مجموعه‌ای از وزن‌های دسته‌بندها و  $l$  نشان دهنده برچسب‌های خروجی دسته‌بندها است.

$$L(s) = \sum_i a(i) * l(i) \quad (10)$$

الگوریتم مدل پیشنهادی در ادامه بیان شده است.

انداختن<sup>۱</sup> استفاده خواهد شد. در این روش به جای مقادیر تعدادی از ویژگی‌های لایه مخفی، مقدار صفر قرار داده می‌شود. یعنی برای ویژگی‌های ادغام شده در لایه قبل،  $C_{max} = \{c_{max}^1, c_{max}^2, \dots, c_{max}^m\}$  (با فرض اینکه تعداد فیلترها در لایه کانولوشنی برابر  $m$  باشد)، ما بخشی از آن‌ها را قبل از ورود به سافت مکس برابر صفر قرار می‌دهیم و به این ترتیب تنها واحدهای بدون تاثیر برای محاسبه گرادیان مورد استفاده قرار می‌گیرند. مقدار واحدهای از قلم افتاده یک ابرپارامتر<sup>۲</sup> است که مقدار آن در طول آموزش مدل تعیین می‌شود. نتیجه دسته‌بندی خروجی لایه سافت مکس پس از ادغام است. این لایه با استفاده از ویژگی‌های تنظیم شده، توزیع احتمالی ورودی را برحسب تمام برچسب‌ها محاسبه می‌کند. هدف اصلی این لایه مشخص کردن دسته تصاویر است. اساس لایه دسته‌بندی، دسته‌بند رگرسیون لجستیک<sup>۳</sup> است که با داشتن ورودی با ابعاد مشخص از لایه‌های قبلی عملیات طبقه‌بندی را به کمک تابع فعال‌ساز سافت‌مکس (رابطه (۶)) انجام می‌دهد.

$$P(y = j|x) = \text{softmax}_j(x^T w + b) = \frac{e^{x^T w_j + b_j}}{\sum_{k=1}^K e^{x^T w_k + b_k}} \quad (6)$$

در این رابطه  $w_k$  وزن ورودی‌ها،  $b_k$  بایاس و  $P_i$  کلاس خروجی و  $k$  نشان‌دهنده تعداد دسته‌ها است [۲۴]. در ادامه ویژگی‌های بدست آمده به الگوریتم آدابوست داده می‌شوند. آدابوست یک الگوریتم است که می‌تواند دسته‌بندهای ضعیف را در یک دسته‌بند قوی ادغام کند. به طور کلی، در روش آدابوست یک نمونه توسط چندین دسته‌بند، دسته‌بندی شده و نتایج دسته‌بندی‌ها به شکل هوشمندانه‌ای با هم ترکیب شده و نتیجه نهایی برای آن نمونه خاص تعیین می‌گردد. معمولاً استفاده از روش آدابوست باعث افزایش کارایی نسبت به دسته‌بندهای تکی می‌شود. این الگوریتم منجر به کاهش خطا و واریانس داده‌های آموزشی نیز می‌شود. در الگوریتم آدابوست، هر دسته‌بند با یک زیر مجموعه تصادفی و منتخب از کل نمونه‌ها آموزش داده می‌شود و با شکل گرفتن چندین دسته‌بند متفاوت، دسته‌بند نهایی که نتیجه نگاه جمعی است و دارای کارایی بالاتری است؛ ایجاد خواهد شد.

از این سازوکار برای پیدا کردن مناسب‌ترین وزن برای دسته‌بندها استفاده می‌شود به طوری که وزن انتخابی برای  $n$ -گرام‌ها با اندازه‌های مختلف مناسب باشد. دسته‌بندها ویژگی‌ها را در محدوده  $\{-1, +1\}$  نگاشت می‌کنند به طوریکه  $+1$  نشان‌دهنده دسته شیء رها شده و  $-1$  نشان‌دهنده شیء ثابت عادی است. باتوجه به کاربرد آدابوست ما باید آمار مربوط به نتایج دسته‌بندی ضعیف را در نمونه‌های آموزشی یافته و وزن نمونه‌های آموزشی و دسته‌بند را برای رسیدن به دسته‌بندهای قوی تنظیم کنیم.

<sup>1</sup> Dropout

<sup>2</sup> Hyper parameter

<sup>3</sup> Logistic Regression

<sup>4</sup> Back propagation

<sup>5</sup> Epoch

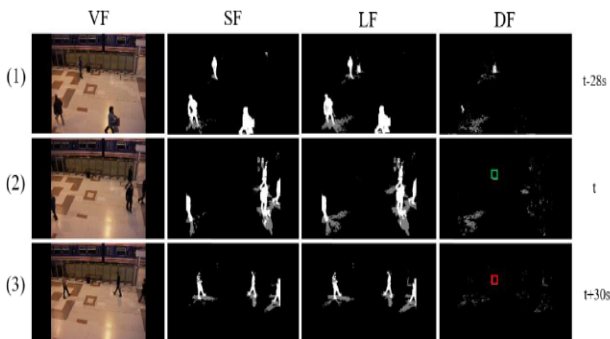
تشخیص اشیاء رها شده با استفاده از شبکه عصبی کانولوشنی بهبود یافته

موقعیت‌هایی که فرد چمدان را می‌اندازد و می‌رود، ردیابی یک شیء مورد بازدید برای مدت طولانی و دسترسی افراد غیر مالک می‌پردازد [۲۶]. شکل ۳ روند تشخیص یک شیء رها شده در سناریوی هفتم این مجموعه داده را نشان می‌دهد، که چالش‌انگیزترین مساله در مجموعه داده PETS2006 است.

شکل (۳): تشخیص شیء رها شده در سناریوی هفتم مجموعه داده PETS2006 (VF, SF, LF, DF) به ترتیب نمایانگر قاب ویدئو، پیش زمینه کوتاه مدت، پیش زمینه بلند مدت و پیش زمینه تفاوت)

در این سناریو، مالک مدت طولانی در اطراف یک شیء می‌چرخد و سپس ناپدید می‌شود. سپس چند گروه از افراد از نزدیک شیء عبور می‌کنند. یکی از موضوعات مهم در این سناریو ردیابی طولانی مدت شیء متغیرها به دلیل پیش فرض بودن موضوع به طور تدریجی در پس زمینه است. مسئله دیگر توانایی تشخیص مالک اصلی است. ما می‌توانیم ببینیم که پیش زمینه شیء مورد نظر قبلاً ناپدید شده است و صاحب آن شیء را در (۲) رها می‌کند. اما الگوریتم هنوز می‌تواند شیء را ردیابی کند و در نهایت با وجود احراز هویت در (۳)، شیء را به عنوان یک شیء رها شده تعیین می‌کند.

#### • مجموعه داده ABODA



ABODA مجموعه داده‌ای است که توسط لین و همکاران ایجاد شده است. در این مجموعه داده یازده سناریو ارائه شده است که شامل محیط‌های داخلی، خارجی، شب، تغییرات روشنایی و صحنه‌های شلوغ است [۲۸]. از آنجایی که ABODA از استاندارد زمانی PETS2006 که ۳۰ ثانیه است پیروی نمی‌کند، ما به طور خودسرانه زمان تعیین ترک برای آزمایش را کاهش دادیم. در این مجموعه داده فریم ۱۱ را حذف نمی‌شود تا که قوانین PETS2006 از نظر زمانی و مکانی حفظ شود. شکل ۴ روند تشخیص یک شیء رها شده در این مجموعه داده نشان می‌دهد.

#### Algorithm 1: Training and testing steps of our model provided

Dataset division: Training, Validation, and Testing  
Output: Class labels

#### Training :

1. Model parameters
  - Filter size: 3,4,5
  - Number of filters :150
  - Activation function: ReLU
  - Optimizer: ADADELTA
  - Learning rate: 0.01
  - Dropout rate:0.05
  - Mini-batch size: 32
  - N: Number of samples
  - Number of epochs: 40
2. Set the number of mini-batches as:  $N_b = Nb2$ .
3. For iteration = 1: number of epochs
  - 3.1 For batch =1 number of mini-batches
    - The obtained training set is fed to CNN+AdaBoost
    - The augmented images batch is fed to CNN+AdaBoost
    - The classification token is fed to the token classifier
    - Calculated the loss function
    - Loss backpropagation
    - Updating the model parameters

#### Model testing:

1. Feed the input images to the model
2. Calculate the prediction label using output label Y

## ۴- پیاده‌سازی و نتایج آزمایشات

### ۴-۱- مجموعه داده

برای پیاده‌سازی مدل پیشنهادی این مقاله از سه مجموعه داده آموزشی استاندارد [۲۶] PETS2006، [۲۷] AVSS2007 و [۲۸] ABODA استفاده شده است. که توضیحات آن‌ها در ادامه آمده است.

#### • مجموعه داده PETS2006

مجموعه داده PETS2006 شامل هفت سناریو در محیط داخلی است. هر سناریو به مسائل مربوط به تشخیص اجسام رها شده در شرایط عادی، مانند افراد ثابت، کیسه‌هایی به بزرگی افراد،



## ۲-۴- روش ارزیابی

به منظور ارزیابی مدل پیشنهادی از معیار ارزیابی دقت، صحت و فراخوانی و امتیاز F استفاده می‌شود و به ترتیب مطابق رابطه (۱۰)، (۱۱)، (۱۲) و (۱۳) محاسبه می‌شوند:

$$Accuracy = (TP + TN) / N \quad (10)$$

$$precision = TP / (TP + FP) \quad (11)$$

$$recall = TP / (TP + FN) \quad (12)$$

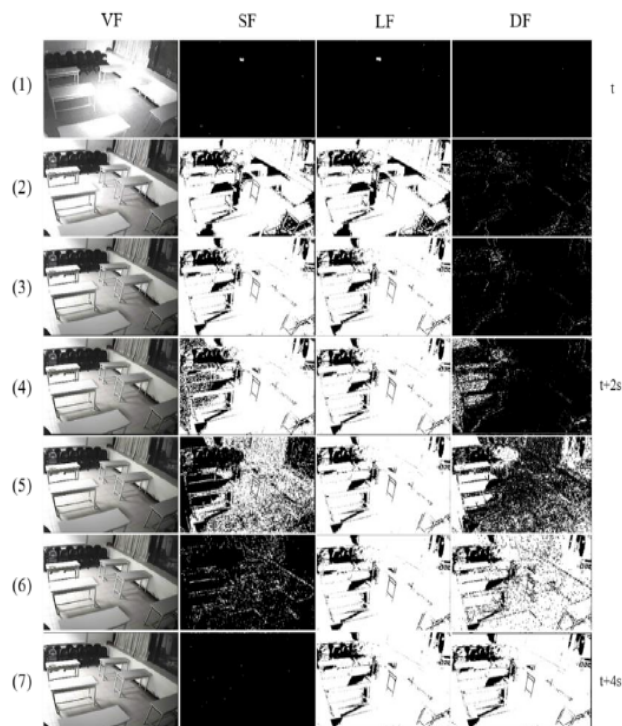
$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

در روابط فوق TP و TN به ترتیب نمونه‌های مثبت و منفی هستند که درست طبقه‌بندی شده‌اند. FN و FP به ترتیب نمونه‌های مثبت و منفی هستند که نادرست طبقه‌بندی شده‌اند و N نیز برابر تعداد کل نمونه‌ها است.

## ۳-۴- جزئیات پیاده‌سازی

از آنجایی که اجرای برنامه بر پایه یادگیری عمیق به جهت محاسبات و پردازش اطلاعات از میان میلیون‌ها داده مختلف صورت می‌گیرد، از یک پردازنده معمولی نمی‌توان انتظار داشت این عملیات را انجام دهد. لذا ضرورت تهیه سخت‌افزاری با سرعت بالاتر و قوی‌تر از نیازهای مهم فرایند فوق می‌باشد. برای پیاده‌سازی روش پیشنهادی از زبان برنامه‌نویسی پایتون استفاده شده است که باعث تسهیل طراحی و پیاده‌سازی الگوریتم‌های یادگیری ماشین و یادگیری عمیق شده است. برای استفاده از محیط برنامه‌نویسی پایتون، از آناکوندا استفاده شده است. به طور خلاصه، کلیه پیاده‌سازی‌های این مقاله به کمک پایتون ۳ و کتابخانه تانسورفلو ۰٫۱٫۲، روی سیستم با پردازنده Intel GH Xeon 2 E5-2620 2.0 و ۸ گیگابایت رم در محیط لینوکس انجام شده است.

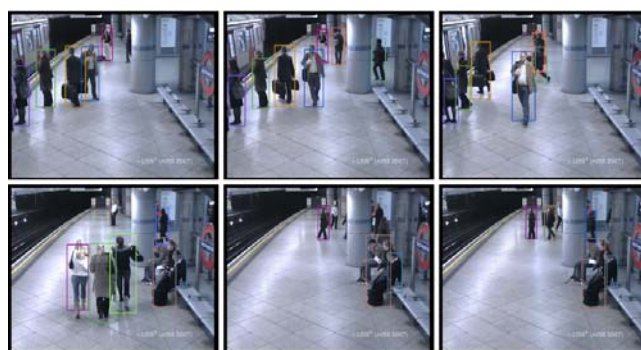
در پیاده‌سازی‌های این مقاله از روش انتخاب مدل Grid Search برای جستجوی بهترین مدل استفاده شده است. در این راستا، ما از سه شبکه کانولوشنی مجزا استفاده کردیم که در هر کدام از آن‌ها اندازه فیلتر یکسان بود. به طور کلی، برای پیاده‌سازی شبکه‌های کانولوشنی، اندازه فیلترها برابر ۳، ۴ و ۵ و تعداد فیلترها برابر ۱۵۰ بود. تابع غیر خطی ReLU نیز به عنوان تابع فعال‌ساز در این شبکه استفاده شده است. از تابع سافت‌مکس نیز برای تعیین دسته‌بندی نهایی استفاده شد. قانون به روز رسانی وزن ADADELTA با نرخ یادگیری ۰٫۰۱ و نرخ از قلم انداختن ۰٫۰۵ برای آموزش مدل مورد استفاده قرار گرفت. لازم به ذکر است که مدل پیشنهادی در ۶۰ اپیک آموزش دید که پس از نزدیک به ۴۰ دور آموزش روی هر سه مجموعه داده همگرا شده است.



شکل (۴): تشخیص شیء رها شده در مجموعه داده ABODA (VF، SF، LF، DF به ترتیب نمایانگر قاب ویدئو، پیش زمینه کوتاه مدت، پیش زمینه بلند مدت و پیش زمینه تفاوت)

### • مجموعه داده AVSS2007

این مجموعه داده با دو هدف تشخیص اشیاء رها شده و ماشین‌های پارک شده ایجاد شده است. با توجه به اینکه هدف این مقاله تشخیص اشیاء رها شده است، بخش اول این مجموعه داده در آزمایشات مورد استفاده قرار گرفته است [۲۷]. در این مجموعه داده در صورتیکه در طول کلیپ، شخصی شیء‌ای را که در اختیار او بوده در منطقه تشخیص قرار داده و بدون شیء از منطقه تشخیص خارج شود و پس از گذشت شصت ثانیه از منطقه تشخیص، به آن شیء بازنگردد و جسم در ناحیه تشخیص باقی بماند، شیء به عنوان شیء رها شده تشخیص داده می‌شود. فرآیند پیگیری کردن به منظور تشخیص شیء رها شده در این مجموعه داده در شکل ۵ نشان داده شده است.



شکل (۵): تشخیص شیء رها شده در مجموعه داده AVSS2007

## ۴-۴- نتایج آزمایشات

پیشنهادی را در مجموعه داده‌های آموزشی مورد بررسی نشان می‌دهد.

جدول ۱ نتایج حاصل از ارزیابی مدل پیشنهادی را در مقایسه با مدل پایه CNN، تشخیص شیء ایستا و شبکه عصبی کانولوشنی (SOD+CNN) و مدل‌های ارائه شده توسط سومرنا ویونسکو که از شبکه عصبی کانولوشنی آبخاری (CCNN) و شبکه عصبی کانولوشنی آبخاری و نمونه‌های تولیدی (CCNN + Generated Samples) برای شناسایی اشیاء رها شده بهره بردند، نشان داده شده است. لازم به ذکر است که نتایج مدل‌های پایه روی مجموعه داده‌های AVSS2007 و PETS2006 از مقالات مرتبط [۷] استخراج شده و روی مجموعه داده ABODA پیاده‌سازی صورت گرفته است.



مجموعه داده AVSS2007



مجموعه داده PETS2006

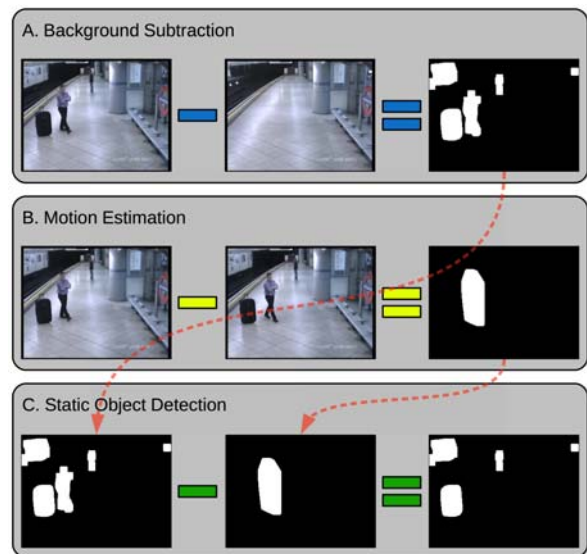


مجموعه داده ABODA

شکل (۷): نمونه‌ای از اشیاء رها شده شناسایی شده توسط مدل پیشنهادی

بر اساس نتایج حاصل از آزمایشات، کاملاً واضح است که مدل پیشنهادی نسبت به سایر مدل‌های پایه روی هر سه مجموعه داده از عملکرد بهتری برای دسته‌بندی بر اساس هر سه معیار ارزیابی هم در سطح فریم و هم در سطح پیکسل برخوردار برخوردار است. بر این اساس می‌توان نتیجه گرفت که استفاده از الگوریتم آدابوست که بر تقویت دسته‌بندی‌های ضعیف استوار است منجر به افزایش دقت دسته‌بندی شده است زیرا این الگوریتم در هر تکرار نقاط داده‌های طبقه‌بندی نشده را مشخص می‌کند و وزن آن‌ها را افزایش می‌دهد (وزن نقاط صحیح را کاهش می‌دهد)، به طوری که طبقه‌بندی کننده بعدی توجه بیشتری به آن‌ها می‌کند که این فرآیند در نهایت منجر به افزایش دقت دسته‌بندی توسط شبکه عصبی کانولوشنی می‌شود.

پس از تأمین بسترهای سخت افزاری و نرم‌افزاری، فرآیند پیش‌پردازش و استخراج ویژگی و مدل‌سازی روی داده‌ها اعمال می‌گردد. در این بخش مقایسه‌ای با هدف ارزیابی کارایی مدل پیشنهادی در مقایسه با سایر مدل‌های سنتی صورت می‌گیرد. همان‌طور که پیش از این اشاره شد، مدل پیشنهادی از دو مرحله مجزا تقسیم شده است. مرحله اول تشخیص شیء و مرحله دوم دسته‌بندی شیء رها شده است. شکل ۶ فرآیند مرحله اول انجام شده در مدل پیشنهادی را نشان می‌دهد که شامل سه بخش است. بخش اول تخمین پیش‌زمینه بر اساس تفریق پس‌زمینه. بخش دوم حرکت بر اساس تفریق فریم‌های ویدئویی به صورت موقت و بخش سوم تشخیص شیء ایستا بر اساس تفریق ماسک حرکت از ماسک پیش‌زمینه. در این مثال خاص، اشیاء ساکن یک قطار مترو، یک چمدان و دو نفر هستند.



شکل (۶): نمونه‌ای از اجرای بخش اول مدل پیشنهادی روی مجموعه داده AVSS2007

مرحله دوم دسته‌بندی شیء رها شده است. در دسته‌بندی مدل پیشنهادی از ترکیب شبکه CNN و آدابوست استفاده شده است. در این راستا ویژگی‌های حاصل از فیلترها با اندازه‌های مختلف حاصل از شبکه‌های عصبی کانولوشنی مختلف را به لایه‌های پولینگ و دسته‌بندی مجزا داده شده و پس آن هنگامی که نتایج اولیه توسط هرکدام از دسته‌بندی بدست آمد، از مکانیزم آدابوست برای تولید نتایج کلی دسته‌بندی استفاده می‌شود. هدف الگوریتم آدابوست افزایش میزان یادگیری دسته‌ها است. این الگوریتم با ترکیب چند دسته ضعیف یک رمز مناسب جهت تفکیک داده‌های بین دو دسته را بدست می‌آورد و به کمک آن می‌توان دسته‌بندی‌هایی که به صورت نادرست دسته‌بندی شده‌اند را اصلاح کرد. الگوریتم آدابوست نسبت به داده‌های نویزی و پرت حساس است؛ ولی نسبت به مشکل بیش برآزش از بیشتر الگوریتم‌های یادگیری برتری دارد. شکل ۷ نمونه‌ای از شناسایی اشیاء رها شده توسط مدل

## ۴- نتیجه‌گیری

ترکیب ده‌ها یا صدها تصمیم ساده را با هدف تولید یک مدل پیچیده‌تر و با دقت بالاتر به کار می‌برد. برخلاف روش‌های سنتی دسته‌بندی، این الگوریتم ضعف جنبش در اطراف نقاط داده‌ای را خطا می‌گیرد و به جای آن به دنبال نقاط داده‌ای است که سخت‌ترین تصمیم‌گیری را دارند.

لازم به ذکر است در شرایطی که بتوان مدل‌های یادگیری عمیق را با حجم عظیمی از داده‌ها آموزش داد، این مدل‌ها از کارایی خوبی برخوردار خواهند بود؛ اما همواره در کاربردهای مختلف مجموعه داده به اندازه کافی بزرگ، که بتوان این مدل‌ها را با آن آموزش داد، وجود ندارد. به همین دلیل پیشنهاد می‌گردد از روش‌های انطباق دامنه و انتقال یادگیری در این کاربرد استفاده شود. به این ترتیب می‌توان از دانش به‌دست آمده از سایر مجموعه داده‌های موجود، که ممکن است دامنه متفاوتی با کاربرد مورد نظر داشته باشند، در یادگیری مدل‌ها استفاده کرد.

در سطح مدل یادگیری و در همین مدل پیشنهادی نیز می‌توان تعداد لایه‌های کانولوشنی افزایش داد و میزان تاثیر آن را مشاهده نمود. همچنین با توجه به اینکه هر کدام از مدل‌های یادگیری عمیق دارای مزایا و معایب خود هستند، می‌توان مدل‌های مختلف یادگیری عمیق را با هم ترکیب کرد.

یکی از وظایف اصلی پرسنل امنیتی، تماشای تصاویر دوربین‌های مختلف نظارتی، برای شناسایی رفتار و اشیاء مشکوک و واکنش مناسب است. اما با زیاد شدن تعداد دوربین‌های امنیتی و نظارت هم زمان یک مأمور امنیتی بر تصاویر چندین دوربین، وظیفه اپراتور بیش‌تر، سخت‌تر و خسته‌کننده می‌شود و دقت اپراتور کاهش می‌یابد. بنابراین نیاز به تکنیک‌هایی است که جریان‌های ویدئویی را به صورت خودکار تجزیه تحلیل کند. وقایع مشکوک را در زمان واقعی شناسایی کند و در مواقع لزوم هشدارهای مورد نیاز را برای کمک به مأمورین امنیتی صادر کند. ارائه و بهبود این تکنیک‌ها در سال‌های اخیر، بسیار مورد توجه جوامع علمی قرار گرفته است. مدل پیشنهادی روی سه مجموعه داده آموزشی استاندارد PETS2006، AVSS2007 و ABODA مورد ارزیابی قرار گرفت. بر اساس نتایج حاصل از آزمایشات، مدل پیشنهادی نسبت به روش‌های پایه روی هر سه مجموعه داده از دقت بالاتری برای دسته‌بندی برخوردار است. براساس نتایج آزمایشات می‌توان نتیجه گرفت که استفاده از الگوریتم آدابوست منجر به افزایش دقت دسته‌بندی داده‌های سخت، کاهش خطا در مدل‌های پیچیده، قابلیت استفاده از چندین الگوریتم یادگیری به طور همزمان فراهم می‌آورد. به عبارت دیگر، آدابوست به صورت

جدول (۱): مقایسه نتایج حاصل از پیاده‌سازی مدل پیشنهادی و مدل‌های پایه

Dataset	Method	Frame level (%)			Pixel level (%)		
		Precision	Recall	F-score	Precision	Recall	F-score
AVSS2007	SOD+CNN	60.17	60.87	60.52	41.19	53.03	46.37
	SOD+CCNN	97.77	51.89	67.80	97.77	51.89	67.80
	SOD + CCNN + Generated Samples	97.48	66.59	79.13	97.47	65.70	78.49
	SOD+CNN+AdaBoost	<b>98.12</b>	<b>66.88</b>	<b>80.12</b>	<b>98.63</b>	<b>66.75</b>	<b>79.64</b>
PETS2006	SOD+CNN	68.01	60.54	68.77	68.00	69.54	68.76
	SOD+CCNN	83.25	69.54	75.78	83.25	70.12	75.78
	SOD + CCNN + Generated Samples	95.67	83.74	89.31	95.67	83.47	78.49
	SOD+CNN+AdaBoost	<b>96.72</b>	<b>84.25</b>	<b>90.31</b>	<b>96.24</b>	<b>83.11</b>	<b>90.12</b>
ABODA	SOD+CNN	70.32	76.33	73.20	66.22	74.65	71.18
	SOD+CCNN	86.54	74.41	80.12	86.52	74.40	80.05
	SOD + CCNN + Generated Samples	95.64	75.83	82.12	87.35	76.45	85.12
	SOD+CNN+AdaBoost	<b>96.74</b>	<b>85.12</b>	<b>91.12</b>	<b>96.73</b>	<b>85.46</b>	<b>90.12</b>

- [15] Fan, Q. and S. Pankanti. *Modeling of temporarily static objects for robust abandoned object detection in urban surveillance*. in *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2011. IEEE.
- [16] Tian, Y., et al., *Robust detection of abandoned and removed objects in complex surveillance videos*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2010. **41**(5): p. 565–57.۶
- [17] Porikli, F., Y. Ivanov, and T. Haga, *Robust abandoned object detection using dual foregrounds*. EURASIP Journal on Advances in Signal Processing, 2007. **2008**: p. 1–11.
- [18] Lin, K., et al. *Left-luggage detection from finite-state-machine analysis in static-camera videos*. in *2014 22nd International Conference on Pattern Recognition*. 2014. IEEE.
- [19] Filonenko, A. and K.-H. Jo, *Unattended object identification for intelligent surveillance systems using sequence of dual background difference*. IEEE Transactions on Industrial Informatics, 2016. **12**(6): p. 2247–2255.
- [20] Filonenko, A. and K.-H. Jo. *Detecting abandoned objects in crowded scenes of surveillance videos using adaptive dual background model*. in *2015 8th International Conference on Human System Interaction (HSI)*. 2015. IEEE.
- [21] Cuevas, C., et al., *Detection of stationary foreground objects using multiple nonparametric background-foreground models on a finite state machine*. IEEE Transactions on image processing, 2016. **26**(3): p. 1127–1142.
- [22] Sidyakin, S. and B. Vishnyakov. *Real-time detection of abandoned bags using CNN*. in *Automated Visual Inspection and Machine Vision II*. 2017. SPIE.
- [23] Irsoy, O. and C. Cardie. *Deep recursive neural networks for compositionality in language*. in *Advances in neural information processing systems*. 2014.
- [24] LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. Nature, vol. 521, no. 7553, pp. 436–444, May, 2015.
- [25] Schapire, R.E., *Explaining adaboost*. Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, 2003 :p. 37–52.
- [26] Thirde, D., L. Li, and F. Ferryman. *Overview of the PETS2006 challenge*. in *Proc. 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006)*. 2006.
- [27] Branch, H.O.S.D. *Imagery library for intelligent detection systems (i-lids)*. in *2006 IET conference on crime and security*. 2006. IET.
- [28] Lin, K., et al., *Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance*. IEEE Transactions on Information Forensics and Security, 2015. **10**(7): p. 1359–1370.
- [1] Luna, E., et al., *Abandoned object detection in video-surveillance: survey and comparison*. Sensors, 2018. **18**(12): p. 4290.
- [2] Shyam, D., A. Kot, and C. Athalye. *Abandoned object detection using pixel-based finite state machine and single shot multibox detector*. in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. 2018. IEEE.
- [3] Savaliya, V. and P. Thakar, *Abandoned object detection system—a review*. Int. Res. J. Eng. Technol, 2018. **5**(10): p. 1668–1677.
- [4] Patel, A.S., et al. *Abandoned Luggage Detection: An Ontology-based approach for unusual activity recognition*. in *2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. 2022. IEEE.
- [5] Din, M., et al., *Abandoned object detection using frame differencing and background subtraction*. International Journal of Advanced Computer Science and Applications, 2020. **11**(7).
- [6] Teja, Y., *Static object detection for video surveillance*. Multimedia Tools and Applications, 2023: p. 1–13.
- [7] Smeureanu, S. and R.T. Ionescu. *Real-time deep learning method for abandoned luggage detection in video*. in *2018 26th European Signal Processing Conference (EUSIPCO)*. 2018. IEEE.
- [8] Pathak, A.R., et al., *Assessment of object detection using deep convolutional neural networks*, in *Intelligent Computing and Information and Communication*. 2018, Springer. p. 457–466.
- [9] Smith, K.C., P. Quelhas, and D. Gatica-Perez, *Detecting abandoned luggage items in a public space*. 2006, IDIAP.
- [10] Kong, H., J.-Y. Audibert, and J. Ponce, *Detecting abandoned objects with a moving camera*. IEEE Transactions on Image Processing, 2010. **19**(8): p. 2201–2210.
- [11] Ferryman, J., et al., *Robust abandoned object detection integrating wide area visual surveillance and social context*. Pattern Recognition Letters, 2013. **34**(7): p. 789–798.
- [12] Szwoch, G., *Extraction of stable foreground image regions for unattended luggage detection*. Multimedia Tools and Applications, 2016. **75**(2): p. 761–786.
- [13] Ilias, D., et al., *An edge-based method for effective abandoned luggage detection in complex surveillance videos*. Computer Vision and Image Understanding, 2017. **158**: p. 141–151.
- [14] Pham, N.T., et al. *Two-stage unattended object detection method with proposals*. in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*. 2017. IEEE.



محدثه هنرمند حقیقی کارشناسی خود را در رشته مهندسی رایانه گرایش نرم افزار در سال ۱۳۸۷ دریافت کرده است. ایشان دانشجوی کارشناسی ارشد مهندسی رایانه گرایش نرم افزار در دانشگاه فومن- شفت می باشد.



سارا معتمد دکترای خود را در رشته مهندسی رایانه گرایش هوش مصنوعی در سال ۱۳۹۶ دریافت کرده است. ایشان هم اکنون استادیار بخش مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد فومن و شفت است. زمینه‌های پژوهشی وی هوش مصنوعی، پردازش تصاویر، بینایی ماشین و علم داده است.