مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال یازدهم، شماره اول، بهار ۱۴۰۳

تفریق پس‌زمینه با استفاده ازشبکه‌های عصبی حافظه کوتاه-مدت بلند عمیق و سازوکار توجه

مهشام کوشکی[۱]، عصمت راشدی[۲]، مریم آموزگار[۳]

چکیده

جداسازی اشیاء متحرک از پس‌زمینه یکی از مسایل مهم و کاربردی در حوزه بینایی ماشین است. در این مقاله یک روش برخط و بدون نظارت مبتنی برای جداسازی پس‌زمینه از پیش‌زمینه در داده‌های ویدیویی ارائه شده است که در آن پس‌زمینه به صورت یک ماتریس کم رتبه L توسط یک شبکه عصبی عمیق استخراج می‌شود و پیش‌زمینه به صورت یک ماتریس تنک با تفریق L از تصویر اصلی به دست می‌آید. در طراحی شبکه عصبی فوق، از یک شبکه حافظه کوتاه-مدت بلند (LSTM) مبتنی بر سازوکار توجه استفاده شده است. یادگیری این روش به صورت بدون نظارت انجام می‌شود و می‌تواند آموزش ببیند به قسمت‌هایی از داده‌ها و تصویر که پیش‌زمینه در آن وجود دارد، وزن بیشتری اختصاص داده و توجه و تمرکز بیشتری داشته باشد. جهت ارزیابی مدل پیشنهادی، پایگاه داده LASIEST انتخاب شده است. کارایی راهکار پیشنهادی به طور کمی با استفاده از معیارهای استاندارد یادآوری، دقت و اندازه‌ی F-measures ارزیابی و با تعدادی از روش‌های معتبر و مطرح مقایسه شده که به ترتیب به میزان ۸%، ۱۰%، و ۵%بهبود داشته است. علاوه بر این از نظر کیفی و شهودی نیز با راهکارهای موجود مورد مقایسه قرار گرفته که موفق به اخذ نتایج مطلوب‌تر شده است.

کلید واژه ها:تفریق پس‌زمینه از پیش‌زمینه، یادگیری عمیق، شبکه عصبیLSTM، سازوکار توجه، یادگیری بدون نظارت، تفریق برخط پس‌زمینه

۱- مقدمه

جداسازی اشیاء متحرک یا پیش‌زمینه از پس‌زمینه در یک دنباله ویدیویی، یکی از مسائل پایه‌ای و مهم در مباحث بینایی ماشین است[۱]. این مسئله در کاربردهایی همچون خودروهای خودران [۲]، نظارت بر جریان ترافیک [۳]، تشخیص ناهنجاری [۴]، ردیابی اشیا[۵] و نظارت ویدیویی [۶] مورد استفاده قرار می‌گیرد مساله تفریق پس‌زمینه با چالش‌های زیادی در سناریوهای دنیای واقعی مانند: تغییرات روشنایی، سایه‌ها، تغییرات دائمی در صحنه‌ها، تغییرات آب و هوا، لرزش دوربین، انسدادجزئی و کلی، انسداد و پیش‌زمینه ثابت روبرو است [۷].

برای مقابله با این چالش‌ها، در چند دهه اخیر روش‌های کلاسیک بسیار زیادی مانند بهبود فرآیند نگهداشت مدل در مسئله تفریق پس‌زمینه[۱]،روش‌های مبتنی بر ردیابی اشیاء [۸] وروش‌های مبتنی بر دسته بندی مسیر [۹]برای تشخیص شیء متحرک معرفی شده است.

در کنار روش‌های کلاسیک، در سال‌های اخیر رویکردهای مبتنی بر شبکه عصبی عمیق به دلیل توانایی آن‌ها در تجزیه و تحلیل حجم زیاد داده و در نظر گرفتن روابط غیرخطی داده‌ها، محبوبیت بسیار زیادی کسب کرده‌اند [۱۰]. این راهکارها به سه دسته آموزش با نظارت، آموزش بدون نظارت و آموزش نیمه نظارتی تقسیم‌بندی می‌شوند. این نکته حائز اهمیت است که اکثر

این مقاله در آذرماه ۱۴۰۲ دریافت شد؛ در خردادماه ۱۴۰۳ بازنگری و سپس پذیرفته گردید.

[۱] کارشناس ارشد از دانشکده مهندسی برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران.
رایانامه: mahshamkushki6724@gmail.com

[۲] دانشیار، دانشکده مهندسي برق و کامپیوتر، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران.
رایانامه: e.rashedi@kgut.ac.ir

[۳] استادیار، گروه کامپیوتر و فناوري اطلاعات، پژوهشگاه علوم و تکنولوژي پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان، ایران.
رایانامه: amoozegar@kgut.ac.ir

نویسنده مسئول: عصمت راشدی

رویکردهای موجود بانظارت و با استفاده از داده‌های برچسب گذاری ارائه شده‌اندمانند[3D-CNN ۱۱]،[۱۲T-CNN]و CNN-[۱۳SFC].با توجه به اینکه روش‌های مطرح شده، متکی بر آموزش با نظارت هستند و فرآیند آموزش در این روش‌ها متکی بر داده‌های برچسب گذاری شده است، قدرت تعمیم‌پذیری آن‌ها کاهش پیدا می‌کند. در واقع شبکه به نمونه‌های آموزشی خود عادت کرده و دچار کاهش کارایی در شناسایی نمونه‌های جدید می‌شود.اکثر الگوریتم‌های تفریق پس‌زمینه در این دسته به صورت برون خط بوده و داده‌ها را چندین بار پردازش می‌کنند[۱۴]، بر خلاف الگوریتم‌های بر خط که به صورت تدریجی-افزایشی عمل کرده و فقط یکبار داده‌ها را پردازش می‌کنند.

شبکه عصبی همگشتی [۱](CNN)یک شبکه کار آمد خاص است کهبه دلیل توانایی یادگیری با استفاده از ویژگی‌های سطح بالا برای تشخیص شیء و تفریق پس‌زمینه استفاده می‌شود[۱۲]. این شبکه دارای دقت و سرعت خوبی در تشخیص اشیاء است و در زمینه‌ی طبقه‌بندی اشیاء و تقسیم‌بندی اشیاء به موفقیت زیادی دست پیدا کرده است. در واقع شبکه‌های همگشتی عمیق مشکل کوتاه بودن حافظه را دارند و اگر یک دنباله به اندازه کافی طولانی باشد، ارسال اطلاعات از مرحله قبلی به مرحله بعدی به مشکل برمی‌خورد و در طول انتشار شبکه دچار محوشدگی گرادیان می‌شوند[۱۵]. محققان مشکل محوشدگی گرادیان را با استفاده ازشبکه‌های عصبی بازگشتی حل کردند [۱۶].

تشخیص ناهنجاری با استفاده از شبکه‌های عصبی بازگشتی در مساله تفریق پس‌زمینه و داده‌های دیگر به نتایج مطلوبی دست پیدا کرده است. به عنوان مثال می توان روش برخط [۲] ارائه شده در [۱۷] مبتنی بر واحدهای بازگشتی گیتی[۳](GRU)، روش بر خط استوار ارائه شده در مرجع [۱۸] و روش تشخیص ناهنجاری متکی بر سری‌های زمانی چندمتغیره در مرجع [۱۹] را ذکر کرد.در سال ۲۰۱۹، فرنوش و همکاران [۲۰]یک مدل پس‌زمینه احتمالی عمیق بدون نظارت به نام[۴]Deep PBM ارائه کردند که بر اساس چارچوب کد گذارهای خودکار متغیر [۲۱]برای تشخیص شیء متحرک کار می‌کند. در این شبکه یک مدل کم بعد از پس‌زمینه تخمین زده می‌شود. برای تفریق پس‌زمینه، با تکیه بر آستانه‌گذاری، تفاوت‌های بین خروجی مدل و قاب ورودی اصلی انجام می‌شود.

شبکه حافظه کوتاه-مدت بلند[۵](LSTM)، یک نمونه شبکه عصبی بازگشتی بهبود یافته است. در [۲۲] یک شبکه رمزگذار و رمزگشا[۶] حافظه‌ی کوتاه مدت بلند برای پیش‌بینیپیکسل‌های ویدیویی ارائه شده است. شبکه‌های LSTM با موفقیت موقعیت‌های مکانی و زمانی را یاد می‌گیرند. در واقع این شبکه رابطه مکانی و زمانی را با پارامترهای کمتری یاد می‌گیرد. آمار و ارقام نشان می‌دهد که شبکه‌های عصبی LSTM عملکرد بهتری نسبت به شبکه‌های GRU دارند.

سازوکار توجه[۷] در حوزه یادگیری ماشین، اساسا مانند توجه و انتخاب بصری انسان است. این کار شبیه به آن چیزی است که افراد در محیط پیرامون خود انجام می‌دهند. این تکنیک می‌تواند در شناسایی افراد، ردیابی آن‌ها، تشخیص چهره وسایر موارد مفید باشد. در واقع، این تکنیک باتوجه به ویژگی‌های مختلف، تشخیص مناطق دارای اهمیت را دقیق‌ترمی‌کند، به این معنا که تمایل دارد به بخش‌های خاصی از اطلاعات که از اهمیت بیشتری برخوردار است (در این مساله بخش‌هایی از تصویر که پیش‌زمینه در آن وجود دارد) وزن بیشتری اختصاص دهد. با توجه به مزایای این تکنیک در تمایز وتمرکز بر ویژگی‌ها، به طور گسترده‌ای در حوزه یادگیری ماشین مورد استفاده قرار گرفته است. از تکنیک توجه می‌توان برای افزایش وزن پیش‌زمینه و بهبود توانایی یادگیری مدل استفاده کرد. این شبکه آموزش می‌بیند که به طور خودکار کدام قسمت از داده‌های آموزشی مرتبط‌تر است و محاسبات خود را روی آن قسمت متمرکز می‌کند [۲۳].

بر این اساس، این مقاله راهکاری برخط برای تفریق پس‌زمینه ارائه می‌کند که متکی بر شبکه‌های عصبی بازگشتی LSTM و مبتنی بر سازوکار توجه است. در واقع شبکه پیشنهادی هر ویدیو را قاب[۸] به قاب در لحظه پردازش می‌کند و لذا برای سناریوهای بلادرنگ مناسب است. نکته حائز اهمیت دیگر اینکه روش پیشنهادی بدون نظارت است و به هیچ برچسبی نیاز ندارد. همچنین سازوکار توجه باعث می‌شود شبکه توجه بیشتری به پیش‌زمینه داشته باشد و اثر پس‌زمینه نامرتبط را کاهش داده و توانایی یادگیری مدل را بهبود بخشد. این شبکه قابلیت رویارویی با چالش‌های ویدیویی مانند تغییرات روشنایی، سایه‌ها، تغییرات دائمی در صحنه‌ها، تغییرات آب و هوا، لرزش دوربین، خود راه اندازی[۹]، انسداد، استتار، انسداد جزئی و کلی را دارد.

در ادامه، در بخش دوم به پیشینه تحقیق در دو قسمت یادگیری کلاسیک ویادگیری عمیق پرداخته شده است. در بخش سوم معماری شبکه پیشنهادی مورد بحث قرارمی‌گیرد.دربخش چهارم پایگاه داده مورد استفاده و در بخش پنجم معیارهای ارزیابی، چگونگی ارزیابی و نتایج کمی و کیفی بروی پایگاه داده ارزیابی شده و مقایسه مدل پیشنهادی با روش‌های دیگر آورده شده است. نهایتا در بخش ششم به جمع‌بندی پرداخته شده است.

---

[۱] Convolution Neural Networks

[۲] Online

[۳] Gated Recurrent Unit

[۴] Deep Probabilistic Background Model

[۵] Long Short-Term Memory

[۶] Encoder and decoder network

[۷] Attention Mechanism

[۸] frame

[۹] Boot Strap

۲-سابقه پژوهش

تفریق پس‌زمینه از پیش‌زمینه در پردازش تصویر، به معنای جداسازی شیء یا اجسام مورد نظر از پس‌زمینه تصویر است. این فرآیند در بسیاری از مسایل بینایی ماشین و پردازش تصویر استفاده می‌شود. در این بخش به انواع روش‌های جداسازی پس‌زمینه از پیش‌زمینه در دو دسته روش‌های کلاسیک آماری و شبکه‌های عصبی عمیق پرداخته شده است. توجه اصلی روی شبکه‌های عصبی عمیق می‌باشد که شامل شبکه‌های عمیق با نظارت[۱]، بدون نظارت[۲] و نیمه‌نظارتی[۳] است.

۲-۱- روش‌های کلاسیک آماری

مطالعات جامعی از روش‌های مبتنی بر رویکرد PCA[۴][۲۴]برای آزمایش و رتبه‌بندی الگوریتم‌های موجود برای تشخیص پیش‌زمینه انجام شده است.این رویکردها عموما به صورت برون خط[۵] کار می‌کنند و با داده‌ها به صورت دسته‌ای[۶] سروکار دارند و برای برنامه‌های بلادرنگ، محدودیت در استفادهدارند. روش دیگری به نام روش مبتنی بر مدل مخلوط گاوسی[۷](MOG)در [۲۵] ارائه شده است. مدل‌های مبتنی بر تحلیل مؤلفه‌های اصلی قوی (RPCA) برای تفکیک پس‌زمینه بازسازی شده به عنوان مؤلفه‌ای با رتبه پایین از پیش‌زمینه به عنوان یک ماتریس تنک ارائه شدند [۲۶]. در برخی موارد جداسازی پس‌زمینه از پیش‌زمینه، دقت پایین و عدم قطعیت وجود دارد که موجب شده برخی از نویسندگان به استفاده از مفاهیم فازی بپردازند [۲۷]. یک روش دیگر مبتنی براستفاده از تخمین چگالی کرنل[۸] برای مدل کردن چگالی احتمال الگوهای محلی[۲۸] می‌باشد. از روش‌های مهم دیگر می‌توان به الگوریتم وایب[۹]، به عنوان یک الگوریتم تفریق پس‌زمینه اشاره کرد [۲۹]. روش پیشنهادی برای هر پیکسل، مجموعه‌ای از مقادیر گذشته در همان مکان یا همسایگی آن را ذخیره می‌کند. سپس این مجموعه را با مقدار پیکسل فعلی مقایسه می‌کند تا تعیین کند آیا آن پیکسل متعلق به پس‌زمینه است یا خیر، و مدل را با انتخاب تصادفی مقادیری از مدل پس‌زمینهبرای جایگزینی، تطبیق می‌دهد.یک روش مدل‌سازی پس‌زمینه مبتنی بر کدگذاری تنک[۱۰] توسط استالیانو و همکاران و معرفی شده است [۳۰]. این روش از تجزیه و تحلیل متغیر فضا استفاده می‌کند.

[۱] supervised
[۲] Unsupervised
[۳] Semi supervised
[۴] Principal component analysis
[۵] Offline
[۶] Batch -Based
[۷] Of Mixture Gaussian
[۸] Kernel density estimation
[۹] ViBe
[۱۰] Sparse

۲-۲-رویکردهای مبتنی بر شبکه عصبی عمیق

در سال‌های گذشته، روش‌های مبتنی بر یادگیری عمیق نشان داده‌اند که قادر به استخراج انتزاعات و داده‌های سطح بالا از مجموعه داده‌های بزرگ هستند.

ساختار شبکه عصبی عمیق شامل لایه‌هایی است که هر لایه به عنوان یک فیلتر برای استخراج ویژگی‌های مختلف از داده‌ها عمل می‌کند. هر لایه از خروجی لایه‌های قبلی خود به عنوان ورودی استفاده می‌کند و خروجی خود را به لایه‌ی بعدی ارسال می‌کند. این شبکه‌ها به کمک لایه‌های پنهان[۱۱]، قابلیت این را دارند که بتوانند ویژگی‌های مختلفی را استخراج کنند. در نتیجه، شبکه عصبی عمیق با پردازش تدریجی و ترکیب ویژگی‌ها، به نتیجه نهایی در خروجی خود می‌رسد.شبکه‌های عصبی عمیق به روش-های یادگیری با نظارت، یادگیری بدون نظارت و یادگیری نیمه‌نظارتی تقسیم می‌شوند. با استفاده از قابلیت این شبکه‌ها، روش‌های مختلفی مبتنی بر یادگیری عمیق جهت تفریق پس‌زمینه ارائه شده که در ادامه مرور شده‌اند.

لیو و همکاران [۳۱] اولین الگوریتم جهت تفریق پس‌زمینه را بر اساس شبکه‌های همگشتی (CNN)با الهام از قشر بینایی حیوانات ارائه کردند. شبکه‌های مولد متخاصم[۱۲] (GAN)یکی از پیشرفت‌های چشم‌گیر در یادگیری عمیق بوده است. این شبکه اولین بار توسط گودفلوو همکاران ارائه شد. شبکه‌های GANبه عنوان یک چارچوب بسیار قدرتمند برای داده‌های بدون برچسب استفاده می‌شوند. این شبکه از دو شبکه متوالی تشکیل شده است: مولد و ممیز. بخش مولد نگاشت را از ورودی مشاهده و به خروجی می‌آموزد. برای بدست آوردن یک صحنه،از این شبکه برای استخراج پس‌زمینه استفاده شده است [۳۲]و [۳۳].

خودرمزنگار[۱۳] (AE)نوع دیگری از شبکه[۱۴]ANNمی‌باشد که یک شبکه یادگیری بدون نظارت است و برای کاهش داده و حجم داده استفاده می‌شود. این شبکه می‌تواند جهت استخراج پس زمینه مورد بهره برداری قرار بگیرد [۳۴].

یک روش دیگر برای تقسیم‌بندی و پیش‌بینی[۱۵](DCP)پس‌زمینه توسط سلطانا و همکاران ارائه شد [۳۵].در یک معماری شبکه فشرده، رحمان وهمکاران یک شبکه عصبی همگشتی-Motion U Net، را برایمعماری عمیق رمزگذار چند نشانه‌ای توسعه دادند که نشانه‌های حرکت و تغییر را با نشانه‌های ظاهری برای تشخیص اشیاء متحرک،ادغام می‌کند[۳۶]. زنگ و همکاران یک معماری شبکه کاملا پیچیده چند مقیاسی[۱۶] را پیشنهاد کردند که از

[۱۱] Hidden Size
[۱۲] Generative Adversarial Network
[۱۳] Auto encoder
[۱۴] Artificial Neural Networks
[۱۵] Deep Context Prediction
[۱۶] Multiscale Fully convolutional Networks

ویژگی‌های لایه‌های مختلف برای تفریق پس‌زمینه استفاده می‌کند و احتیاجی به استخراج تصاویر پس‌زمینه ندارد [۳۷]. در مطالعه‌ی دیگری، بیان فنگ و همکاران [۳۸] یک روش تشخیص شیء متحرک در صحنه‌های پیچیده با استفاده از معماری ReNet-18 برای دستیابی به طبقه بندی سطح پیکسل را ارائه کردند.

در مرجع [۳۹] یک روش نظارت شده دیگر برای تشخیص اجسام متحرک مبتنی بر شبکه رمزگذار-رمزگشا همگشتی برای استخراج ویژگی‌های سطح پیکسل معرفی شده است که به مدل‌سازی تغییرات پیکسل در طول زمان نیز می‌پردازد. علاوه بر این لایه‌هایی را نیز برای افزایش پایداری مدل در مقابل حرکت پس‌زمینه اضافه کرده است.در سال ۲۰۱۹، فرنوش و همکاران [۲۰] یک مدل پس‌زمینه احتمالاتی عمیق بدون نظارت به نامDeep PBMبر اساس چارچوبی از متغیرها راارائه کردند. در Deep PBM یک مدل با ابعاد کم از پس‌زمینه تخمین زده شده و سپس پس‌زمینه با آستانه‌گذاری مشخص می‌شود. در این شبکه، یککدگذار، نمایش کارآمدی از ویدیوهای ورودی را می‌آموزد و آن را در یک فضای تصادفی با ابعاد پایین‌تر تعیین شده توسط لایه‌های پنهان طراحی می‌کند و بخشبازیابی تلاش می‌کند تا داده‌های اصلی را با توجه به متغیرهای احتمالاتی پنهان رمزگذار بازیابی کند.

دردهه‌های اخیر یادگیری‌های عمیق نیمه‌نظارتی مسیر جدیدی در حوزه یادگیری شبکه‌های عصبی عمیق ایجاد کرده است. در این مدل یادگیری، داده‌های برچسب خورده در کنارحجم زیادی از داده های برچسب نخورده در دسترس است. یادگیری نیمه نظارتی به دلیل حجم بالای داده‌های برچسب نخورده استفاده می‌شود. این روش می‌تواند بسیار مفید باشد چون برچسب گذاری همه داده‌ها، زمان بر و هزینه براست و دارای حجم محاسباتی بالا می‌باشد.

۳-راهکار پیشنهادی

در این بخش، روش پیشنهادی جهت جداسازی پس‌زمینه مبتنی بر شبکه‌های عصبی عمیق بازگشتی و با استفاده از تکنیک توجه معرفی شده است.

در روش‌های کلاسیک، از جمله در مرجع [۴۰] پس‌زمینه و پیش‌زمینه‌ی قاب‌های ویدیویی براساس ماتریس کم رتبه[۱]L، و ماتریس تنک[۲]Sمدل می‌شوند.در واقع تصویر اصلی از دو قسمتLوSبه صورتX = L +Sتشکیل شده است. با الهام از این مدل‌ها، شبکه عصبی عمیق بازگشتی پیشنهادی مجهز به بخش تنک شده است تا ضمن تخمین پس‌زمینه، پیش‌زمینه را نیز استخراج کند. بدین ترتیب قسمت کم رتبه یک تخمین از پس‌زمینه است که به کمک شبکه عصبی عمیق مدل‌سازی می‌شود و قسمت تنک نمایانگر پیش‌زمینه است. ابعاد ماتریس کم رتبه و پیش‌زمینه معادل ابعاد تصویر است. وزن‌های شبکه به صورت برخطبروزرسانی می‌شوند و یادگیری تا آخرین قاب ویدیویی ادامه دارد به طوری که مدل پیشنهادی می‌تواند با تغییرات سازگار شود. یادگیری در این روش، به صورت بدون نظارت است به طوری که در هر گام زمانی،یک قاب ویدیویی پردازش شده و با رسیدن قاب ویدیویی جدید، وزن‌های شبکه بروزرسانی می‌شود. با توجه به اینکه در هر لحظه فقط از یک تصویر برای بروزرسانی شبکه استفاده می‌شود، محاسبات شبکه از پیچیدگی کمتری نسبت به روش‌هایی که از چند قاب استفاده می‌کنند، برخوردار است. در معماری شبکه عصبی عمیق جهت مدل سازی پس‌زمینه، از شبکه بازگشتیLSTMمبتنی بر سازوکار توجه استفاده شده است.

درشبکه‌های عصبی، سازوکارتوجه باعث می‌شود که تمرکز بروی یک سری داده‌های مهم در حین فرآیند آموزش افزایش پیدا کند و اثر بقیه داده‌ها کاهش می‌یابد.در واقع در سازوکارتوجه،در ورودی یک وزن به داده‌ها اختصاص داده می‌شود. این وزن نشان دهنده اهمیت و تاثیر این داده‌ها بر روند آموزش می‌باشد. روش پیشنهادی از سازوکارتوجه بهره برده و از اولین قاب تا آخرین قاب ویدیویی را پردازش می‌کند. معماری شبکه و جزئیات پیاده‌سازی آن در ادامه آورده شده است.

۱-۳- معماری شبکه

در روش پیشنهادی، از شبکه عصبی بازگشتی LSTM استفاده شده است. چراکه درپردازش ویدیو، با داده‌های سری و متوالی سروکار داریم که از هم مستقل نیستند و هر داده اطلاعاتی مشابه داده‌های قبلی را در خود دارد. همچنین از سازوکارتوجه جهت بهبود کارایی شبکه استفاده شده است و وزن‌ها با توجه به اهمیت داده‌ها بروزرسانی و با تغییرات سازگار می‌شوند. در روش پیشنهاد شده، تمرکز روی برخی از قسمت‌های داده افزایش پیدا می‌کند. شبکه پیشنهادی به اختصار[۳] (ALSTM)نامیده شده است. شبکه با استفاده از لایه‌های تکراری پنهان قادر به یادگیری الگوهای متوالی و تخمین داده‌هایی با ابعاد کم می‌باشد. روش پیشنهادی به صورت برخط و به ازای هر فریم و بدون نظارت آموزش می‌بیند و لذا از پیچیدگی و محاسبات کمتری برخوردار است.

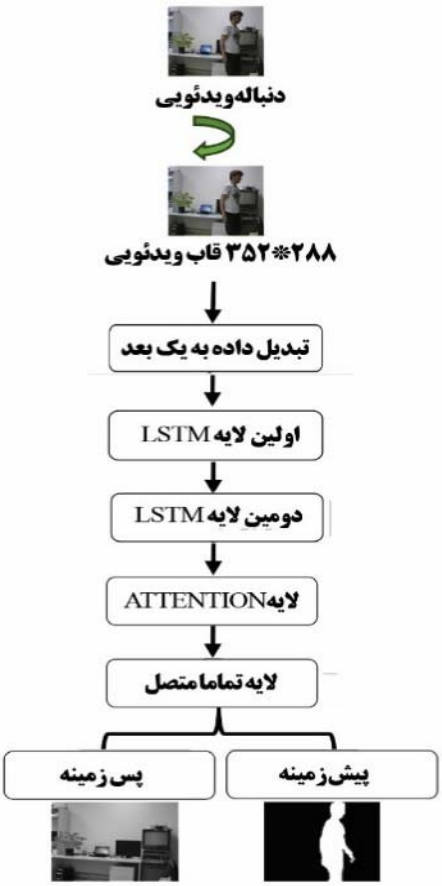
نمای کلی روش پیشنهادی در شکل۱ ارائه شده است. در این روش، اثر آموزشی هر قاب ویدیویی حتی پس از گذشت زمان در شبکه باقی خواهد ماند و با گذشت زمان عملکرد مدل بهبود می‌یابد.همانطورکه در شکل۱ نمایش داده شده است، این شبکه از دو لایهLSTM، یک لایه توجه و یک لایه تماما متصل تشکیل شده است. لازم به ذکر است خروجی لایه دومLSTMاز یک لایه تمامامتصل با ابعاد تصویر ورودی گذارنده می‌شود. در این رویکرد، معماری شبکه از کمترین پیچیدگی برخوردار است. در

[۱] Low – Rank Matrix
[۲] Sparse Matrix
[۳] Attention Long Short-Term Memory Network

روش پیشنهادی، از مزایای شبکه‌های عصبی بازگشتی برای پردازش داده‌های متوالی، بدون نیاز به داده‌های برچسب گذاری شده و آموزش بدون نظارت استفاده می‌شود.در هرگام زمانی، یک قاب ویدیویی وارد مدل می‌شود و با استفاده از شبکه عصبی عمیق، یک ماتریس کم رتبه بازسازی می‌شود.ماتریس تنکا تفریق ماتریس کم رتبه از ماتریس قاب ویدیویی حاصل می‌شود. همانطور که در ابتدا گفته شده، ماتریس کم رتبه Lحاوی صحنه پس‌زمینه و ماتریس پراکنده Sحاوی صحنه‌ی پیش‌زمینه است. از صحنه‌ی پس‌زمینه بازسازی شده برای محاسبه تابع خطا و آموزش شبکه با الگوریتم پس انتشار خطا استفاده می‌شود.

۲-۳- پیش پردازش داده‌ها

همانطور که گفته شد، ویدیوهای انتخاب شده از پایگاه[۱]LASIESTAهستند. این ویدیوها به فضای خاکستری تبدیل می‌شوند و با ابعاد اندازه اصلی خودشان ۲۸۸×۳۵۲ به صورت داده های یک بعدی در می‌آیند و وارد شبکه می‌شوند.



شکل(۱):معماری شبکه پیشنهادی ALSTM

۳-۳- آموزش شبکه پیشنهادی

توالی ویدیویی $X_t \in X=\{X_1, X_2, ..., X_t, ..., X_N\}$ را در نظر بگیرید که در آن X_tقاب ویدیویی ورودی در زمان t وN تعداد قاب‌های ویدیویی یا همان طول دنباله‌ی ویدیویی است. قاب‌های ویدیویی به فضای خاکستری تبدیل می‌شوند. فرآیند آموزش روش پیشنهادی با استفاده از الگوریتم پس انتشار انجام می‌شود. به این ترتیب، حل مسئله بهینه سازی با تنظیم تابع تلفات به صورت تابع مربعات خطا[۲](RMSE) می‌تواند مدل را با بروزرسانی وزن‌ها آموزش دهد. به منظور دستیابی به یک مدل تفریق پس‌زمینه، هر مرحله شامل بروزرسانی وزن‌های شبکهALSTMبر اساس قاب‌های ویدیویی ورودی است. پیاده سازی روش پیشنهادی بر اساس توضیحات ذکر شده در الگوریتم ۱ ارائه شده است که در بخش‌های بعدی اجزا و عملکرد آن تشریح می‌شود.

الگوریتم ۱: شبکه عصبی بازگشتی (ALSTM) جهت تفریق پس‌زمینه

ورودی: قاب ویدیویی $X_t \in X=\{X_1, X_2, ..., X_t, ..., X_N\}$

خروجی: صحنه ی پس‌زمینه و پیش‌زمینه‌ی X_t.

- مقداردهی اولیه:

ماتریس کم رتبه برابر با صفر در نظر گرفته می‌شود$(L=0)$.

وزن‌های شبکه به صورت تصادفی در نظر گرفته می‌شوند.

- به ازای هر قاب ویدیویی مراحل زیر انجام می‌شود:

۱. تخمین مؤلفه‌ی کم رتبهدر خروجی شبکه.

۲. محاسبه‌ی ماتریس تنکا استفاده از تفریق پیش‌زمینه از قاب ویدیویی جاری:

$$S = X_t - L$$

۳. آموزش وزن های شبکه.

پایان توالی ویدیویی

۴-۳-مدل‌سازی پس‌زمینه

روش پیشنهادی بر اساس نگهداشت و بروزرسانی مدل پس‌زمینه کار می‌کند و با گذشت زمان، مدل الگوی پس‌زمینه را یاد می‌گیرد و از شبکهLSTMبرای نگهداری اطلاعات و تاثیر آن بر آموزش شبکه و همچنین حذف اطلاعات اضافی استفاده می‌کند. این شبکه به گونه‌ای بروزرسانی و تنظیم می‌شود که صحنه پس‌زمینه را که یک الگوی ثابت با کمترین تغییرات استتخمین بزند. خروجی آخرین لایه LSTM به یک لایه تماما متصل منتقل می‌شود تا خروجی نهایی یا همان پس‌زمینه تصویر را تولید کند. اجسام متحرک به عنوان پیش‌زمینه در نظر گرفته می‌شوند و با تفریق پس‌زمینه از قاب جاری استخراج می‌شوند.

۵-۳- سازوکار توجه

همانطور که اشاره شد شبکه LSTM می‌تواند مؤلفه رتبه پایین را تخمین بزند و پس از تخمین مؤلفه با رتبه پایین، ماتریس تنک Sقابل استخراج است. لازم به ذکر است خروجی لایه‌ی LSTM

---

[۱] http://www.gti.ssr.upm.es/data/LASIESTA

[۲] Root Mean Square Error (RMSE)

دوم از یک لایه‌ی تمام‌متصل با ابعاد تصویر ورودی گذارنده می‌شود. سازوکارتوجه امکان مدلسازی و انعکاس وابستگی‌هارا بدون توجه به فاصله آن‌ها در توالی ورودی فراهم می‌کند [۴۱]. یک تابع توجه را می‌توان به عنوان نگاشت یک پرس و جو و مجموعه‌ای از جفت‌های کلید-مقدار به یک خروجی توصیف کرد که در آن پرس‌وجو، کلیدها و خروجی همه بردار هستند.خروجی به عنوان مجموع وزنی از مقادیر ورودی محاسبه می‌شود، که در آن وزن اختصاص داده شده به هر مقدار توسط تابع سازگاری پرس و جو با کلید مربوطه محاسبه می‌شود[۴۱].

استفاده از سازوکار توجه درشبکه‌های عصبی عمیق باعث بهبود و کیفیت شبکه‌های عصبی عمیق است و این امکان را به داده‌ها می‌دهد تا تمرکز را روی داده‌های ورودی افزایش دهد. این کار با وزن دادن به داده‌های ورودی انجام می‌شودتا برخی از ویژگی‌های ورودی را نسبت به سایرین اولویت دهد.این وزن‌ها نشان‌دهنده این است که هر قسمت از داده‌های ورودی چقدر اهمیت دارد و به چه میزان باید بر روی آنها تمرکز شود این وزن‌ها به طور خودکار تنظیم می‌شوند. ساز و کار توجه این امکان را فراهم می‌کند که در حین پردازش قاب به قاب تصویر بر روی اساسی‌ترین بخش داده‌ها تمرکز شود.

در مدل پیشنهادی از یک توجه چند سر[۱] استفاده شده است که به موقعیت‌های مختلف و موقعیت‌های متفاوت توجه می‌کند. این تکنیک کاربردهای وسیعی دارد. سازوکار توجه مدل را قادرمی‌سازد تا به طور انتخابی بر بخش‌های مختلف ورودی تمرکز کند.

۴-پایگاه داده

مجموعه داده استفاده شده LASIESTA نام دارد که شامل ۱۸۴۲۵ قاب ویدیو در اندازه ۳۵۲×۲۸۸ است که در ۴۸ ویدیو و ۱۴ دسته مختلف تقسیم‌بندی شده‌اند و به صورت رایگان در دسترس عموم است[۷]. این پایگاه داده یک پایگاه داده جامع و فشرده و مناسب برای تشخیص شیء متحرک می‌باشد. این پایگاه داده شامل مجموعه‌ای از داده‌ها است که شامل سناریوهای داخلی و خارجی است. قاب‌های واقعی بیرونی و داخلی زیادی در این پایگاه داده وجود دارد که در دسته بندی‌های مختلف سازمان‌دهی شده‌اند و هر دسته بندی چالش خاصی را پوشش را می‌دهد مانند تغییرات شدید روشنایی، خود راه اندازی، استتار، انسداد، پس‌زمینه اصلاح شده، صحنه های ساده، تغییرات مدام پس زمینه. یکی دیگر از ویژگی‌های برجسته آن داشتن برچسب[۲] برای تمام قاب‌های ویدیویی می‌باشد. همان‌طورکه که قبلا اشاره شد،این پایگاه داده دارای صحنه‌های داخلی و خارجی است. تمام ویدئوهای داخلی با چالش‌های مختلف در ۷ دسته و ویدیوهای خارجی در ۵ دسته تقسیم بندی شده‌اند.

۵- ارزیابی

اغلب الگوریتم‌های تفریق پس‌زمینه در استخراج پیش‌زمینه در شرایط ایده آل عملکرد خوبی دارند [۱]. اما گاهی اوقات توقف ناگهانی یک شیء متحرک، حرکت شاخه‌های درخت، استتار جسم متحرک، تغییرات مداوم پس‌زمینه و شرایط بد آب و هوایی تشخیص پیش‌زمینه را دچار مشکل می‌کنند. الگوریتم پیشنهادی بر روی یک مجموعه داده جامع و چالش برانگیز ارزیابی شده است تا قابلیت تعمیم آن را نشان دهد. به منظور مقایسه راهکار پیشنهادی جدیدترین و متناسب‌ترین راهکارهای موجود انتخاب شدند.یکی از این راهکارها، روش Deep PBM که یک روش پیشرفته و بدون نظارت است. کد این مقاله توسط نویسندگان به اشتراک گذاشته شده است [۴۲]. همچنین روش GRU[۱۵]نیز روی پایگاه داده فوق مقایسه و ازیابی شده است. مدل پیشنهادی با استفاده از زبان پایتون و در کتابخانه pytorch نوشته شده است. برای به‌روزرسانی وزن‌های شبکه پیشنهادی از روش بهینه‌سازی Adam استفاده شده است. این الگوریتم از ترکیب دو روش بهینه‌سازیRMSprop و Momentum برای بهبود سرعت و کارایی شبکه استفاده می‌کند.همچنین نرخ یادگیری برابر با ۲×۱۰^{-۴} تنظیم شده است.

۵-۱-معیارهای ارزیابی

باآزمایش بر روی مجموعه داده‌ی ذکر شده، یک ماتریس درهم‌ریختگی[۳] حاصل از مقایسه‌ی ماسک‌های پیش‌زمینه باینری با ماسک‌های برچسب گذاری شده مجموعه داده، بدست می‌آید که در آن (TP)[۴]، (TN)[۵]، (FP)[۶]، (FN)[۷] به ترتیب تعداد پیکسل‌های مثبت واقعی، منفی واقعی، مثبت کاذب و منفی کاذب را نشان می‌دهند. در جدول ۱چگونگی تعریف این پارامترها مشخص شده است.

برای ارزیابی عملکرد روش پیشنهادی، از معیارهای یادآوری، دقت و اندازه‌ی F که به ترتیب در روابط ۱ تا ۳ تعریف شده‌اند، استفاده شده است.

$$RECALL = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{T}{TN}\right) \quad (۱)$$

$$Precision = \frac{1}{2}\left(\frac{TP}{TP+FP} + \frac{T}{TN}\right) \quad (۲)$$

$$F\text{-}Measure = \frac{1}{n}\sum_{i=1}^{n}\frac{Precision \times RE}{Precision + RE} \quad (۳)$$

---

[۱] Multi-Head Attention

[۲] Ground truth

[۳] Confusion matrix

[۴] True Positive

[۵] True Negative

[۶] False Positive

[۷] False Negative

جدول (۱):تعیین پارامترهای ماتریس درهم ریختگی.

	ماتریس کم رتبه (صحنه پس‌زمینه)	ماتریس تنک (صحنه پیش‌زمینه)
صحنه پس‌زمینه (طبق‌ماسک‌برچسب شده در پایگاه داده)	منفی واقعی(TN)	مثبت کاذب(FP)
صحنه پیش‌زمینه(طبق‌ماسک‌برچسب شده در پایگاه داده)	منفی کاذب (FN)	مثبت واقعی(TP)

یکی از رایج‌ترین معیارها، میانگین اندازه‌ی F است. اندازه‌ی F با ترکیب دو معیار مهم ارزیابی شامل یادآوری و دقت پس‌زمینه محاسبه‌می‌شود. بنابراین عملکرد کلی یک الگوریتم تفریق پس‌زمینه توسط مقدار اندازه‌ی Fو معیار دقت آن قابل بررسی است و برای ارزیابی و تعمیم‌پذیری آن در موقعیت‌های مختلف مناسب است. بنابراین در مقایسه‌ها، بیشترین توجه به معیار اندازه‌ی Fو دقت معطوف شده است.

۲-۵- نتایج کمی

جهت ارزیابی، روش پیشنهادی روی پایگاه داده پیاده سازی و اجرا شده است. نتایج با روش Deep PBM[۴۲]و GRU[۱۵]مقایسه شده است.

در جدول ۲ معیارهای ارزیابی محاسبه و برای صحنه‌های داخلی، مجموعه داده انتخابی نمایش داده شده است. قابل توجه است که در هر دسته چالش، دو ویدیو وجود دارد به عنوان مثال برای چالش صحنه‌های ساده، دو ویدیوی I_SI_01 و I_SI_02ارائه شده است.

می‌توان دریافت که ALSTM به اندازه‌ی F بالاتری نسبت به روش Deep PBM که در بین روش‌های بدون نظارت، پیشرفته‌تر است، دست پیدا کرده است. همچنین نسبت به روش GRU به مقادیر بهتری دست یافته است. این معیار برای دسته‌های چالش برانگیزی مانند انسداد، تغییرات روشنایی، و استتار نیز به خوبی عمل کرده است.

یادآوری می‌شود ALSTM از اولین قاب یک ویدیو از مجموعه داده LASIEST، بدون در نظر گرفتن مرحله آموزشی، ارزیابی شده است و قاب‌ها را یک به یک دریافت می‌کند. همچنین این نکته حائز اهمیت است که فرصت تصمیم گیری در مورد هر قاب تنها یک بار است.

همچنین در این جدول معیار دقت در روش پیشنهادی نسبت به روش DeepPBM و GRU به مقادیر بهتری دست یافته است. معیار بازخوانی در بعضی از چالش‌های این پایگاه داده هم اندازه بوده و در بسیاری از موارد نسبت به دو روش دیگر به مقدار بهتری داشته است. لازم به ذکر است که در این پایگاه داده، هر چالش شامل دو ویدیو می‌باشد که به دلیل نیاز به‌توان محاسباتی بالا برای روشDeepPBMو محدویت منابع، این روش تنها بر ویدیوهای شماره ۱ این پایگاه داده ارزیابی و اجرا شده است.

جدول (۲) :نتایج کمی صحنه‌های داخلی برروی پایگاه دادهLASIESTA(برحسب درصد).

طبقه بندی	Method	Recall	Precision	F-Measure
صحنه ساده (I_SI_01)	ALSTM	۵۸	۸۴	۷۱
	GRU	۵۶	۸۰	۶۸
	DeepPBM	۵۷	۶۳	۶۳
صحنه ساده (I_SI_02)	ALSTM	۵۱	۷۰	۶۰
	GRU	۴۹	۶۸	۵۸
صحنه استتار (I_CA_01)	ALSTM	۵۱	۷۱	۶۱
	GRU	۴۹	۷۰	۵۴
	DeepPBM	۶۲	۶۱	۶۱
صحنه استتار (I_CA_02)	ALSTM	۵۳	۷۶	۶۴
	GRU	۵۲	۷۲	۶۲
صحنه انسداد (I_OC-01)	ALSTM	۵۱	۵۶	۵۴
	GRU	۵۰	۵۶	۵۱
	DeepPBM	۵۱	۵۱	۵۳
صحنه انسداد (I_OC-02)	ALSTM	۴	۵۹	۵۲
	GRU	۴۵	۵۶	۵۱
صحنه تغییرات روشنایی (I-IL-01)	ALSTM	۵۳	۶۲	۵۸
	GRU	۵۱	۶۰	۵۹
	DeepPBM	۴۸	۵۲	۵۰
صحنه تغییرات روشنایی (I-IL-02)	ALSTM	۶۰	۵۹	۵۵
	GRU	۴۹	۵۳	۵۱
صحنه پس‌زمینه اصلاح شده (I-MB-01)	ALSTM	۵۱	۶۳	۵۶
	GRU	۴۹	۵۷	۵۲
	DeepPBM	۵۵	۶۲	۶۱
صحنه پس‌زمینه اصلاح شده (I-MB-02)	ALSTM	۵۳	۷۲	۶۴
	GRU	۵۱	۶۶	۶۱
صحنه خود راه اندازی (I-BS-01)	ALSTM	۵۲	۶۸	۶۲
	GRU	۵۳	۶۵	۶۰
	DeepPBM	۵۵	۶۰	۶۰
صحنه خود راه اندازی (I-BS-02)	ALSTM	۵۱	۵۹	۵۴
	GRU	۵۰	۵۳	۵۱

جدول (۳) نتایج کمی صحنه های خارجی برروی پایگاه دادهLASIESTA(برحسب درصد).

طبقه بندی	Method	Recall	Precision	F-Measure
صحنه‌های ابری(O_CL_01)	ALSTM	۵۳	۶۵	۶۲
	GRU	۵۱	۶۱	۵۸
	DeepPBM	۵۴	۶۱	۶۱
صحنه‌های ابری (O_CL_02)	ALSTM	۵۳	۹۰	۶۸
	GRU	۵۲	۷۸	۶۳
صحنه‌های شرایط بارانی(O_RA_01)	ALSTM	۵۳	۶۵	۶۵
	GRU	۵۲	۶۱	۵۸
	DeepPBM	۵۴	۶۰	۶۰
صحنه‌های شرایط بارانی(O_RA_02)	ALSTM	۵۲	۷۰	۶۸
	GRU	۵۰	۶۵	۶۵
صحنه‌های شرایط برفی(O_SN_01)	ALSTM	۶۷	۶۰	۶۹
	GRU	۶۳	۵۸	۶۵
	DeepPBM	۶۱	۵۶	۶۵
صحنه‌های شرایط برفی(O_SN_02)	ALSTM	۶۵	۶۳	۷۲
	GRU	۶۴	۵۴	۷۰
صحنه‌های شرایط آفتابی(O_SU_01)	ALSTM	۶۳	۸۶	۷۵
	GRU	۶۲	۸۹	۷۳
	DeepPBM	۶۱	۶۰	۶۴
صحنه‌های شرایط آفتابی(O_SU_02)	ALSTM	۵۱	۸۶	۶۵
	GRU	۵۰	۸۲	۶۳

نتایج نشان می‌دهد که روش پیشنهادی به عملکرد مناسبی در پایگاه داده LASIESTA دست‌یافته است و قادر است روی داده های کاملا ناآشنا و دیده نشده پیاده‌سازی شود.

این نکته حائز اهمیت است که این پایگاه داده از دو مجموعه ویدیوی داخلی و خارجی تشکیل شده است. جدول شماره ۳ تمام معیارهایی که برای فضای داخلی محاسبه شده بودند را برای صحنه‌های خارجی نیز نمایش می‌دهد.

تحلیل نتایج جدول ۳ برتری روش پیشنهادی را در صحنه‌های خارجی نیز تایید می‌کند. بر این اساس می‌توان گفت که روش پیشنهادی در مجموعه داده LASIESTA توانسته بهتر از روش‌های دیگر عمل کند.

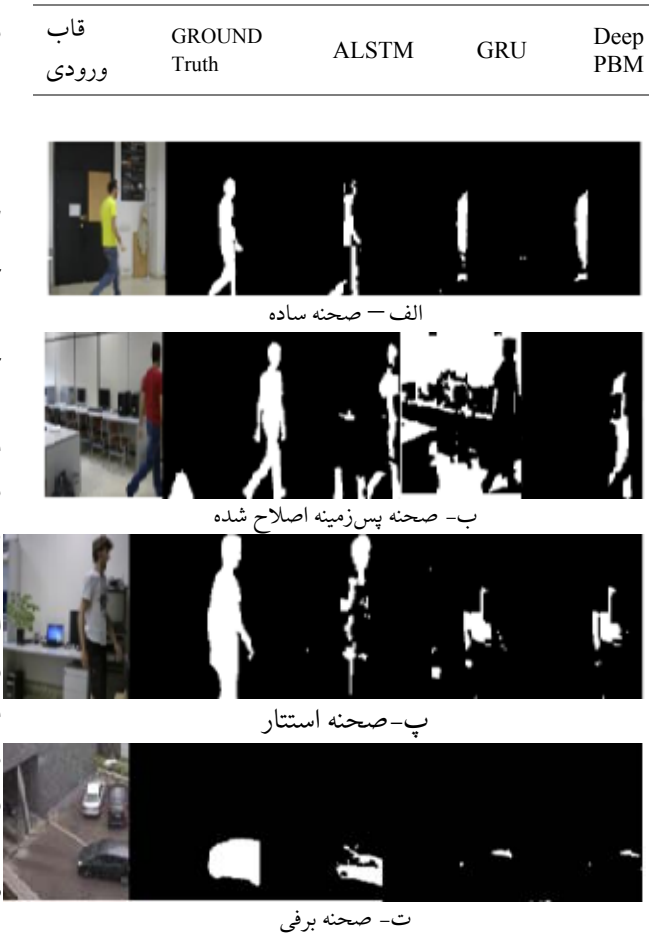
روش پیشنهادی یک رویکرد بدون نظارت و برخط است که هر قاب را فقط یک بار پردازش می‌کند. در مقابل، روش‌های یادگیری عمیق مورد مقایسه برون خط هستند و اغلب در چندین تکرار و تحت نظارت فرآیند آموزش را انجام می‌دهند. در ادامه برای ارائه ارزیابی کامل‌تر، روش پیشنهادی با تعدادی از روش‌های مبتنی بر یادگیری عمیق تحت نظارت، مقایسه شد و نتایج به شرح زیر در جدول (٤) ارائه شده است. همانطور که نتایج نشان می‌دهد روش پیشنهادی نتایج قابل قبولی به دست آورده ضمن اینکه یک روش بدون نظارت و برخط است. بقیه روش‌ها در چندین تکرار قاب‌های ویدیو را چندین بار پردازش می‌کنند و از اطلاعات برچسب نیز بهره می‌برند.

جدول (۴): نتایج اندازه گیریF-measuresدر روش‌های نظارت شده بر روی پایگاه دادهLASIESTA (برحسب درصد).

Method	I_SI_02	I_CA_02	I_OC-02	I-IL-02	I-MB-02	I-MS-02	O-CL-02	O-RA-02	O-SN-02	O-SU-02	AVG
BSUV-Net [۴۳]	۸۹	۶۰	۹۵	۸۹	۷۶	۶۹	۸۹	۹۳	۷۰	۹۱	۸۲
3DCD [۴۴]	۸۶	۴۹	۹۳	۸۵	۷۹	۸۷	۸۷	۸۷	۴۹	۸۳	۷۹
FgSegNet-s [۴۵]	۲۰	۶۰	۵۳	۲۵	۶۰	۲۸	۱۹	۱۶	۵	۱۸	۳۰
FgSegNet-M [۴۵]	۵۶	۵۵	۶۵	۴۲	۵۶	۱۹	۲۸	۱۸	۱	۳۳	۳۷
MSFS [۴۶]	۵۳	۵۸	۲۵	۴۱	۶۳	۲۵	۵۴	۵۴	۵	۲۹	۴۱
ChangeDet [۴۷]	۸۳	۶۶	۸۹	۸۷	۷۷	۸۲	۹۰	۸۵	۵۱	۸۱	۷۹
ALSTM	۶۰	۶۴	۵۲	۵۵	۶۴	۵۴	۶۸	۶۸	۷۲	۷۵	۶۳

۳-۵-نتایج کیفی

ویدیویی‌ها که مورد مطالعه قرار گرفته‌اند شامل اکثر چالش‌هایی هستند که در تفریق پس‌زمینه از پیش‌زمینه وجود دارند؛ مانند: شرایط انسداد، استتار، سایه‌ها، تغییرات نور و روشنایی و پس‌زمینه پویا.



شکل (۲): .مقایسه کیفی نتایج. تصاویر از چپ به ترتیب تصویر اصلی، جواب صحیح، و نتایج روش‌های GRU، DEEP PBM و ALSTMهستند.

شکل ۲ نتایج سه روش ذکر شده را نشان می‌دهد. از نظر بصری می‌توان دید که نتایج روش ALSTM بهتر از روشGRU و Deep PBM است. نتایج تشخیص پیش‌زمینه ALSTM دقیق‌تراست. همچنین در صحنه‌های ساده بدون استتار، بدون انسداد، تغییر نور، تغییرات مداوم پس‌زمینه و خود راه اندازی، جداسازی پس‌زمینه از پیش‌زمینه به راحتی قابل مشاهده است. در شرایط ابری، بارانی، برفی و آفتابی که به دلیل حرکت شاخه‌های درخت و سایه‌ها پس‌زمینه‌های پویا هستند، روش پیشنهادی می‌تواند اشیاء پیش‌زمینه را استخراج کند

۶-نتیجه گیری

جداسازی پس‌زمینه از پیش‌زمینه یکی از مسائل مهم در پردازش تصویر و شبکه‌های عصبی عمیق است. در این تحقیق روشی برای حل مسئله‌ی تفریق پس‌زمینه با تعمیم چارچوب قوی X=L+S با استفاده از شبکه‌های عصبی عمیق با حافظه طولانی-کوتاه مدت مبتنی بر تکنیک توجه پیشنهاد شده است. این روش، یک رویکرد بدون نظارت و برخط است که در هر تکرار آموزش فقط از یک قاب استفاده می‌کند.

برای ارزیابی روش پیشنهادی از پایگاه داده LASIESTA استفاده شده که یک پایگاه داده جامع و کامل بوده، تمام چالش‌ها را در بر می‌گیرد و امکان مقایسه و ارزیابی با مدل‌های دیگر را می‌دهد. مدل پیشنهادی با روش DeepPBMو GRUکه از روش-های به‌روز و بدون نظارت می‌باشند، مقایسه شده است. این روش را می‌توان برای صحنه‌های دیده نشده که حاوی تغییرات روشنایی، استتار، انسداد، سایه‌ها، خودراه اندازی، تغییرات مداوم پس‌زمینه و شرایط بد آب و هوایی هستند، اعمال کرد. بنابراین روش پیشنهادی برای توالی‌های ویدیویی که دارای پیچیدگی بالا و دنباله‌های طولانی در کاربردهای نظارت بر فعالیت‌های انسانی، نظارت بر ترافیک، نظارت بر فرودگاه‌ها و دریاها هستند کاربرد دارد. توسعه این روش به ویدیوهای با دوربین متحرک می‌تواند در تحقیقات آتی دنبال شود.

مراجع

[۱] مریم. اموزگار, معصومه. اکبری زاده ,"بهبود فرآیند مدل در مسئله تفریق پس‌زمینه" نشریه علوم رایانش و فناوری اطلاعات ۱۴۰۰

[2] W. Farag and Z. Saleh "An advanced vehicle detection and tracking scheme for self-driving cars," 2nd Smart Cities Symposium (SCS 2019),2019, pp.1-6.

[3] P. Zhang, W. Liu, Y. Lei, H. Wang, and H. Lu, “RAPNet: Residual atrous pyramid network for importance-aware street scene parsing,” IEEE Trans. Image Process., vol. 29, pp. 5010–5021, 2020.

[4] K. M. Biradar, A. Gupta, M. Mandal, and S. K. Vipparthi, “Challenges in time-stamp aware anomaly detection in traffic videos,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2019, pp. 13–20.

[5] X. Dong, J. Shen, D. Wu, K. Guo, X. Jin, and F. Porikli, “Quadruplet network with one-shot learning for fast visual object tracking,” IEEE Trans. Image Process., vol. 28, no. 7, pp. 3516–3527, Jul. 2019.

[6] Z. Fu, Y. Chen, H. Yong, R. Jiang, L. Zhang, and X.-S. Hua, “Foreground gating and background refining network for surveillance object detection,” IEEE Trans. Image Process., vol. 28, no. 12, pp. 6077–6090, Dec. 2019.

[7] C. Cuevas, E. M. Yáñez, and N. García, "Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA," *Computer Visionand Image Understanding,* vol. 152, pp. 103-117, 2016.

[8] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 25, no. 5, pp. 564-577, 2003.

[9] X. Yin, B. Wang, W. Li, Y. Liu, and M. Zhang, "Background Subtraction for Moving Cameras based on trajectory-controlled segmentation and Label Inference," *Ksii Transactions on Internet & Information Systems,* vol. 9, no. 10, 2015.

[10] T.Bouwmans.Ja,Sved,M.Sultana, and S.k.Jung,"Deep neural network concept for background subtraction :A systematic review and comparative evaluation ,"Neural Network , vol.117,pp8-66,2019

[11] C. Rao and Y. Liu, "Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization," *Computational Materials Science,* vol. 184, p. 109850, 2020.

[12] D. Alapatt *et al.*, "Temporally Constrained Neural Networks (TCNN): A framework for semi-supervised video semantic segmentation," *arXiv preprint arXiv:2112.13815,* 2021.

[13] D. Zeng, M. Zhu, and A. Kuijper, "Combining background subtraction algorithms with convolutional neural network," *Journal of Electronic Imaging,* vol. 28, no. 1, pp. 013011-013011, 2019.

[14] H. Ahn and M. Kang, "Dynamic background subtraction with masked RPCA," *Signal, Image and Video Processing,* vol. 15, pp. 467-474, 2021.

[15] A. Sedghi, E. Rashedi, M. Amoozegar, and F. Afsari, "Online vehicle detection using gated recurrent units," in international conference on Artificial intelligence and smart Vehicle, 2023.

[16] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259,* 2014.

[17] A. Sedghi, M. Amoozger , E. Rashedi, and F.Afsari, "ORGU: Online Robust Gated Recurrent Units for Real _Time Background Subtraction in video sequences,"10.21203/rs.3.rs-3138445/v1, 2023.

[۱۸] مریم ,اموزگار. مرتضی, فائزی نیا , بهروز, مینایی بیدگلی "تشخیص برخط ناهنجاری با استفاده از شبکه عصبی بازگشتی "نشریه مدلسازی در مهندسی ۱۴۰۲

[19] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion,* vol. 91, pp. 93-102, 2023.

[20] A. Farnoosh, B. Rezaei, and S. Ostadabbas, "DeepPBM: deep probabilistic background model estimation from video sequences," in International Conference on pattern Recognition (ICPR), 2021,pp.608-621

[21] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114,* 2013.

[22] S.choo,w.seo,Dudhance ,and N.I.Cho.”Multi-scale recurrent encoder-decoder network for dense temporal classification ‘. In 2018 24 th International Conference on Pattern Recognition (ICPR) ,2018,pp.103_108

[23] P. Liu, J. Feng, J. Sang, and Y. K. Kim, "Fusion Attention Mechanism for Foreground Detection Based on Multiscale U-Net Architecture," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.

[24] H. Ahn and M. Kang,"Daynamic background subtraction with masked RPCA" signalImage Video Proscess.,vol.15.no.3 pp.467-474,2021

[25] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent patents on computer science,* vol. 1, no. 3, pp. 219-237, 2008.

[26] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Background–foreground modeling based on spatiotemporal sparse subspace clustering," *IEEE Transactions on Image Processing,* vol. 26, no. 12, pp. 5840-5854, 2017.

[27] T. Bouwmans, "Background subtraction for visual surveillance: A fuzzy approach," *Handbook on soft computing for video surveillance,* vol. 5, pp. 103-138, 2012.

[28] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li, "Modeling pixel process with scale invariant local

patterns for background subtraction in complex scenes," in *2010 IEEE computer society conference on computer vision and pattern recognition*, 2010: IEEE, pp. 1301-1306.

[29] O.Barnich and M.VanDroogcbroeck ,"ViBe:Auniversal background subtraction algorithm for video sequence",IEEE Trans. Image process. Vol. 20. no .6. pp1709-1724,2010.

[30] A. Stagliano,N ,Noceti,A.Verri, and F.Odone." Online space-variantbackgroundmodelingwithsparsecoding,"IEEE.Trans .Image process. Vol.24. no. 8, pp.

[31] Y. Zhang. X. Li,Z.Zhang,F,Wu, and L.Zhao,"Deep learning driven bhockwise moving object detection with binary scene modeling " Neurocomputing, vol 168,pp.454-463.2015

[32] M. C. Bakkay, H. A. Rashwan, H. Salmane, L. Khoudour, D. Puig, and Y. Ruichek, "BSCGAN: Deep background subtraction with conditional generative adversarial networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018: IEEE, pp. 4018-4022.

[33] W. Zheng, K. Wang, F.Y. Wang, "A novel background subtraction algorithm based on parallel vision and Bayesian GANs", Neurocomputing, Volume 394, , pp. 178-200, 2020.

[34] B. Sauvalle, A. Fortelle, "Autoencoder-based background reconstruction and foreground segmentation with background noise estimation", cvf open access, 2023.

[35] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung, "Unsupervised deep context prediction for background estimation and foreground segmentation," *Machine Vision and Applications,* vol. 30, pp. 375-395, 2019.

[36] G. Rahmon, F. Bunyak, G. Seetharaman, and K. Palaniappan, "Motion U-Net: Multi-cue encoder-decoder network for motion segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021: IEEE, pp. 8125-8132

[37] D. Zeng and M Zhu"Back ground subtraction using multiscale fully convolutional network" IEEE Access,vol . 6. pp.16010-16021.2018.

[38] X. Bian, S. N. Lim. and N.Zhou,"Multiscale Fully convolutional network with application to industrial inspection,"in 2016 IEEEwinter conference on applications of computer vision (WACV).2016.

[39] Y. Chen, J. Wang, B. Zhu, M. Tang, and H. Lu, "Pixelwise deep sequence learning for moving object detection," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 29, no. 9, pp. 2567-2579, 2017.

[40] C. Zhou and R.C. Paffenroth ,"Anomaly detection with robust deep autoencoders ," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and data Mining , Aug .2017 , Vol . Part F1 29685, pp.665-674. doi:10.1145/3097983.3098052.

[41] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.

[42] "Deep PBM".https://github . com / ostadabbas/DeepPBM.

[43] M. O. Tezcan .p .Ishwar, and.JKonraud, "BSUV-Net 2.0: Spatio-temporal data augmentation for video-agnostic supervised background subtraction "IEEE Access. Vol. 9.pp.53849-53860.2021

[44] M. Mandal, V,Dhar, A, Mishra,S.K .Vipparthi , and . M.Abodel-Mottaleb,"3DCD:Scene independent end- to -end spatiotemporal feature learning framework for change detection unseen videos,"IEEE Trans. Image process. Vol. 30, pp.546-558.2020

[45] L .A .Lim .H. Y. Keles,"Foreground segmentation using a triplet convolution neural network for multiscale feature encoding arXiv 2018" arXivpreprarXiv 1801.02225

[46] L .A .Lim .H. Y. Keles "Learning multi-scale features for foreground segmentation " pattern Anal, Appl, vol 23. no 3, .pp, 1369-1380,2020

[47] M. Mandal and S.K.Vipparthi"Scene independency matters: An empirical study of scene dependent and scenes independent evaluation for CNN-based change detection "IEEE Trans . Intell. Transp. Syst, 2020

مهشام کوشکی مدرک کارشناسی را سال ۱۳۹۸ و کارشناسی ارشد خود را در سال ۱۴۰۲ در رشته مخابرات سیستم در دانشگاه تکمیلی صنعتی و فناوری‌های پیشرفته دریافت کرد. هم اکنون دانشجوی دکتری مخابرات سیستم در دانشگاه تکمیلی صنعتی و فناوری‌های پیشرفته می‌باشد. زمینه علاقمندی ایشان پردازش تصویر،یادگیری ماشین وهوش مصنوعی است.

عصمت راشدی دارای دکترا در رشته مهندسی برق گرایش مخابرات سیستم می‌باشد. وی هم اکنون دانشیار گروه مهندسی برق دانشگاه تحصیلات تکمیلی صنعتی و فناوری های پیشرفته در شهر کرمان است. زمینه های پژوهشی ایشان پردازش تصویر و ویدیو، بازشناسی الگو، الگوریتم های بهینه سازی ابتکاری و شبکه های عصبی عمیق است.

مریم آموزگار مدرک دکتری خود در رشته مهندسی نرم افزار را از دانشگاه علم و صنعت ایران دریافت کرده است. وی اکنون استادیار پژوهشکده کامپیوتر و فناوری اطلاعات در پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی در دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته است. علایق تحقیقاتی او تشخیص ناهنجاری، پردازش داده های جریانی، یادگیری ماشین و الگوریتمهای فراابتکاری می باشد.