



Computational Intelligence in Electrical Engineering
Vol. 14, No. 1, 2023
Research Paper

Improvement in detection of presence in forbidden locations in video anomaly using optical flow map

Mohammad Rahimpour¹, Mohammad Kazemi², Payman Moallem³, Mehran Safayani⁴

¹ Dept. of Electrical Engineering, University of Isfahan, Isfahan, Iran

² Dept. of Electrical Engineering, University of Isfahan, Isfahan, Iran

³ Dept. of Electrical Engineering, University of Isfahan, Isfahan, Iran

⁴ Dept. of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran

Abstract:

Anomaly detection has been in researchers' scope of study for a long time. The wide variety of anomaly detection use cases ranges from quality control in production lines to providing security in public places. One of the most attractive topics in anomaly detection is in video surveillance systems. In this paper, we propose a method that works based on frame prediction and optical flow to improve anomaly detection in videos. The use of optical flows in normal frames helps the system to better detect the entrance of people or objects to forbidden areas by its information about the amount of movement in different regions of the frames. Based on the optical flow of normal videos and that of current video, the threshold for anomaly decision is adaptively adjusted. This could ultimately lead to a better overall performance of the anomaly detection system compared to the recent similar works. The presented method is general and can be simply incorporated to other video anomaly detection systems to improve the detection accuracy.

Keywords: Anomaly detection, video surveillance, optical flow, deep neural network.



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/ISEE.2022.134125.1571>

بهبود تشخیص حضور در مناطق ممنوعه در تصاویر ناهنجار ویدئویی

با استفاده از نقشه شار نوری

محمد رحیم پور^۱، محمد کاظمی*^۲، پیمان معلم^۳، مهران صفایانی^۴

۱- دانشجوی دکتری گروه مهندسی برق- دانشکده فنی و مهندسی - دانشگاه اصفهان- اصفهان- ایران

mohammadrahimpour@eng.ui.ac.ir

۲- استادیار گروه مهندسی برق- دانشکده فنی و مهندسی - دانشگاه اصفهان- اصفهان- ایران

m.kazemi@eng.ui.ac.ir

۳- استاد گروه مهندسی برق- دانشکده فنی و مهندسی - دانشگاه اصفهان- اصفهان- ایران

p_moallem@eng.ui.ac.ir

۴- دانشیار - دانشکده مهندسی برق و کامپیوتر- دانشگاه صنعتی اصفهان- اصفهان- ایران

safayani@iut.ac.ir

چکیده: پژوهشگران به تشخیص ناهنجاری از دیرباز توجه داشته‌اند. تشخیص ناهنجاری کاربردهای متنوعی شامل کنترل کیفیت خط‌های تولید تا تأمین امنیت معابر و گذرگاههای عمومی می‌شود. یکی از موارد شایان توجه در تشخیص ناهنجاری، تشخیص ناهنجاری در تصاویر دوربین‌های نظارتی است. در این تحقیق، روشی مبتنی بر پیش‌بینی فریم‌های جاری در ویدئو و شار نوری برای بهبود تشخیص ناهنجاری ارائه می‌شود. استفاده از داده‌های مربوط به شار نوری فریم‌های عادی به سیستم کمک می‌کند تا از طریق داشتن اطلاعات درباره میزان حرکت در نواحی مختلف تصویر، ورود احتمالی افراد یا اشیا را به نواحی ممنوعه تصویر بهتر تشخیص دهد و از این طریق، عملکرد کلی سیستم نسبت به کارهای مشابه بهبود می‌یابد. مزیت دیگر این روش، کلی بودن آن است؛ به این معنی که به سادگی به شبکه‌های دیگر برای بهبود دقت در تشخیص ناهنجاری افزوده می‌شود.

واژه‌های کلیدی: تشخیص ناهنجاری، نظارت ویدئویی، شار نوری، شبکه عصبی عمیق.

۱- مقدمه

تولید، تشخیص تصادف به کمک دوربین‌های نظارتی خیابان‌ها، تشخیص اتفاقات ناگوار و غیرمنتظره مانند درگیری در معابر و گذرگاههای عمومی و غیره‌اند. ناهنجاری به معنای وجود هر شی، اتفاق یا وضعیتی است که عادی محسوب نمی‌شود. با توجه به این تعریف، معنای ناهنجاری در مکان‌ها و زمان‌های مختلف متغیر است؛ برای مثال، دمای هوای ۱۰ درجه سانتی‌گراد برای فصل زمستان دمایی عادی محسوب می‌شود؛ اما همین دمای هوا در فصل تابستان مقداری نامتعارف و به عبارت دیگر، ناهنجاری است. بر همین اساس، تعریف ناهنجاری در تصاویر ویدئوها هم متفاوت است. تشخیص ناهنجاری برای انسان کاری ساده محسوب می‌شود؛ اما با توجه به بی‌تعریف بودن

پژوهشگران از دیرباز درباره تشخیص ناهنجاری به‌منزله یک زمینه تحقیقاتی بحث کرده‌اند. کاربردهای بسیار تشخیص ناهنجاری در کنار چالش‌های فراوان این حوزه، آن را به موضوعی همچنان جذاب برای تحقیق تبدیل کرده است. از جمله کاربردهای آن، تشخیص خرابی در خط

^۱ تاریخ ارسال مقاله: ۱۴۰۱/۰۴/۰۱

تاریخ پذیرش مقاله: ۱۴۰۱/۰۵/۰۲

نام نویسنده مسئول: محمد کاظمی

نشانی نویسنده مسئول: ایران - اصفهان - دانشگاه اصفهان - دانشکده فنی و مهندسی، گروه مهندسی برق

ناهنجاری، تشخیص آن برای ماشین امری دشوار و چالش برانگیز است.

امروزه روش‌های مبتنی بر یادگیری ماشین تلاش می‌کنند تا با یادگیری ویژگی‌های مربوط به رویدادها و شرایط عادی، تعریفی از هنجار به دست بیاورند و سپس با استفاده از این تعریف و مقایسه آن با نمونه‌های مختلف و با اندازه‌گیری میزان شباهت تعریف هنجار و نمونه ناشناس، درباره هنجار یا ناهنجار بودن نمونه تصمیم‌گیری کنند. برای انجام این کار، چالش‌های مختلفی وجود دارند. نخستین چالش، نبود توازن بین نمونه‌های ناهنجار و نمونه‌های هنجار است. در هر محیط یا شرایطی، تعداد دفعات رخداد ناهنجاری بسیار کمتر از تعداد رخداد هنجار است. چالش دوم، تعداد بسیار بالا یا حتی نامتناهی بودن کلاس‌های ناهنجاری است. چالش سوم شباهت بسیار زیاد میان نمونه‌های ناهنجار و هنجار در بعضی موارد است. با توجه به چالش دوم، پژوهشگران معمولاً از روش‌های بدون ناظر یا با نظارت ضعیف برای آموزش استفاده می‌کنند. این روش‌ها در کنار مزایایی که به همراه دارند، ایراداتی را هم به دنبال خواهند داشت. روش بدون ناظر نیازی به داده برچسب دار شده^۱ ندارد. همین موضوع هزینه مالی و زمانی تهیه داده‌های مورد نیاز را برای آموزش بدون ناظر کاهش می‌دهد. استفاده نکردن از داده برچسب‌دار شده معایبی نیز دارد. برخی اوقات تعریف تابع هزینه‌ای که بتواند کل سیستم را به سمت جواب مناسب همگرا کند، کاری پیچیده است. همچنین، جواب نهایی سیستم‌های بدون ناظر معمولاً از کیفیت کمتری نسبت به روش‌های با ناظر برخوردارند.

برای آموزش یک سامانه از ویژگی‌های دست‌ساز یا ویژگی‌های استخراج‌شده به صورت خودکار استفاده می‌شود. با توجه به پیشرفت‌های اخیر در ریاضیات و همچنین افزایش چشمگیر و نمایی قدرت محاسباتی واحدهای پردازنده، روش‌های خودکار مانند شبکه‌های عصبی عمیق از استقبال بیشتری برخوردار بوده‌اند. این روش‌ها به صورت خودکار، ویژگی‌های مناسب‌تر برای حل یک مسئله را استخراج و با آنها مسئله مطرح‌شده را حل می‌کنند.

در ادامه به بررسی برخی روش‌های موجود برای تشخیص ناهنجاری پرداخته می‌شود و در بخش بعد، روش پیشنهادی برای بهبود و ارتقای سیستم‌های تشخیص ناهنجاری در ویدئو معرفی می‌شود. نتایج نهایی استفاده از روش معرفی شده و مقایسه آن با سایر روش‌های موجود نیز در بخش نتایج عملی آورده شده‌اند. در نهایت، جمع‌بندی و نتیجه‌گیری کلی ارائه شده است.

۲- مطالعات پیشین

با توجه به اینکه روش‌های بدون ناظر به دلیل عدم نیاز به داده‌های برچسب‌گذاری شده معمول‌ترند، در اینجا این روش‌ها بررسی می‌شوند؛ روش‌هایی که برای تشخیص ناهنجاری از آنها استفاده می‌شود یا براساس خوشه‌بندی کار می‌کنند یا از خطای بازسازی برای تشخیص ناهنجاری بهره می‌برند. در برخی از موارد نیز از این دو به صورت ترکیبی استفاده می‌شود. روش‌های مبتنی بر خوشه‌بندی سعی می‌کنند داده‌های آموزشی را در خوشه‌هایی با مراکز مجزا قرار دهند. هدف از مرحله آموزش، به دست آوردن خوشه‌هایی است که به خوبی از هم جدا شده‌اند. هر داده‌ای که متعلق به یکی از خوشه‌های آموخته‌شده در مرحله آموزش نباشد، داده ناهنجار در نظر گرفته می‌شود. روش ارائه‌شده در [۱] از یک روش خوشه‌بندی خاصی برای یادگیری زیرمجموعه‌های عادی استفاده می‌کند که می‌تواند به صورت مؤثر زیردسته‌های هنجار را به عنوان مراکز خوشه‌ها فرا گیرد. روش مقاله [۲] یک رویکرد مبتنی بر مسیر و یک رویکرد منطقه‌محور را برای دستیابی به نتایج بهتر ترکیب می‌کند. اگرچه روش معرفی شده در [۳] یک روش مبتنی بر خطای بازسازی برای تشخیص ناهنجاری است، پژوهشگران از خوشه‌بندی برای شناسایی استفاده می‌کنند، تا هم شبکه‌های رمزگذار^۲ مربوط به ظاهر و هم شبکه‌های رمزگذار مربوط به حرکت را مجبور کنند. فاکتورهای رایج متغیر در مجموعه داده را استخراج کنند. روشی مبتنی بر خوشه‌بندی براساس SVM تک‌کلاسه برای

عادی جاده‌ها فراگرفته شود. سبک‌رو و همکارانش در تحقیق انجام‌شده در [۸] دو تشخیص‌دهنده ناهنجاری معرفی کرده‌اند که یکی براساس بازسازی یک دسته فریم ویدئو کار می‌کند و دیگری از نمایش تنک داده‌های ورودی استفاده می‌کند. در هر دوی این سیستم‌ها از خودرمنگاریها استفاده شده است. در تحقیق ارائه‌شده در [۹] از یک ساختار آبشاری استفاده شده که طبقه اول آن یک خودرمنگاری سه‌بعدی است و به کمک آن تشخیص ناهنجاری در دو مرحله انجام خواهد شد. طبقه اول ساختار را یک شبکه با عمق متوسط تشکیل می‌دهد و در طبقه دوم برای تشخیص بهتر، از یک شبکه با عمق بیشتر استفاده شده است. در تحقیق [۱۰] از خودرمنگاریهای متغیر برای یافتن ناهنجاری‌ها استفاده شده که استفاده از آنها در کنار روش فراگیری لغت‌نامه باعث شده است یک نمایش تنک برای داده‌ها ایجاد شود که در نهایت به بهینگی کلی سیستم از منظر کاهش ابعاد داده منجر می‌شود. در روش معرفی شده در [۱۱]، برای یافتن ناهنجاری‌ها از طریق یادگیری لغت‌نامه در فضای ثانویه، یک خودرمنگاری متغیر معرفی شده است. بر اساس روش [۱۲] می‌توان با استفاده از یک خودرمنگاری سه‌بعدی ویژگی‌هایی از ویدئو را استخراج و سپس با استفاده از یک فرهنگ لغت، فضای ویژگی‌های مربوط به داده‌های هنجار را مدل کرد. با استفاده از این مدل، نمونه‌های ناهنجار نیز تشخیص‌پذیرند. مرجع [۱۳] ساختاری تماماً کاتولوشنی معرفی کرده که توانایی فراگیری فضای داده‌های هنجار را داراست. ویژگی اصلی آن استفاده از یادگیری انتقالی برای فراگیری داده‌های هنجار است. در آنجا برای استخراج ویژگی‌های غنی از لایه‌های AlexNet استفاده شده است.

همان‌گونه که گفته شد روش‌های براساس پیش‌بینی فریم‌های بعدی نیز مانند روش‌های براساس بازسازی فریم جاری به دنبال نمونه‌هایی هستند که خطای بازسازی آنها از حد مشخصی بیشتر شود. میزان خطای بازسازی معمولاً با مجذور میانگین خطا یا کمیت‌های معادل آن اندازه‌گیری می‌شود [۱۴]. در تحقیق [۱۵] از یک کدگشای بازسازی و

تشخیص مسیرهای حرکت ناهنجار معرفی شده است که می‌تواند با توجه خاصی که به مسیرهای حرکتی دارد، بدون داشتن دانش قبلی از ناهنجاری‌ها آنها را به‌درستی تشخیص دهد [۴]. می‌توان از یک خودرمنگاری سه‌بعدی برای استخراج ویژگی استفاده کرد و در ادامه بر اساس این ویژگی‌ها یک خوشه‌بندی برای فراگیری فضای داده‌های هنجار انجام داد [۵]. در مرجع [۶]، پژوهشگران از یک خوشه‌بندی دو مرحله‌ای استفاده کرده‌اند تا ناهنجاری‌های مربوط به حرکت وسایل نقلیه را تشخیص دهند؛ به این صورت که ابتدا مسیرهای حرکتی، پیش‌پردازش و از آنها نمونه‌برداری می‌شود و سپس از خوشه‌بندی فضایی استفاده می‌شود تا مسیرهای شبیه به هم در یک خوشه قرار گیرند.

روش‌های مبتنی بر خطای بازسازی در میان پژوهشگران بسیار محبوب‌اند. در این روش‌ها تلاش می‌شود تا ابتدا فضای داده‌های هنجار یا نمایشی از آنها به بهترین نحو ممکن یادگیری شود. در اینجا فرض بر این است که سیستم می‌تواند نمونه‌های فراگرفته‌شده را که همان هنجارها هستند، با کیفیت بالا بازسازی کند؛ اما این امر برای داده‌های ناهنجار میسر نیست؛ به همین دلیل، در این روش‌ها در صورتی که خطای بازسازی نمونه‌ای از آستانه بیشتر شود، آن نمونه به‌عنوان نمونه ناهنجار دسته‌بندی می‌شود. از جمله ابزارهای محبوب برای یادگیری فضای داده‌های هنجار، خودرمنگاری‌ها^۱، خودرمنگاریهای متغیر^۲، رمزنگاری‌های نوین زده^۳ و نمونه‌های تنظیم‌شده و تغییر یافته آنها هستند.

در برخی روش‌ها از خطای بازسازی فریم (های) جاری استفاده می‌شود و در برخی دیگر، سیستم فریم (های) بعدی را بازسازی می‌کند و خطای بین فریم (های) بعدی و بازسازی آن (ها) معیار قرار می‌گیرد. در اینجا هر دو دسته روش، یعنی بازسازی فریم‌های جاری و بازسازی فریم‌های آینده بررسی می‌شوند. سیستم معرفی شده در [۷] از تعقیب اشیا و تشخیص ناهنجاری در کنار هم استفاده می‌کند تا تصادفات جاده‌ای را تشخیص دهد. در این تحقیق از شبکه‌های خودرمنگاری عمیق به‌صورت پشته‌شده^۴ استفاده شده است تا فضای داده‌های هنجار، یعنی ترافیک معمولی و

یک کدگشای پیش‌بین^۸ استفاده شده است. در کنار بهره‌مندی از مزایای کدگشاهای پیش‌بین، استفاده از کدگشای بازساز این امکان را فراهم می‌آورد که بتوان قیودی برای آموزش بهتر اعمال کرد. در روش ارائه‌شده در [۱۶]، ابتدا اختلاف بین فریم پیش‌بینی شده و فریم اصلی به کمترین مقدار ممکن رسانده شده است و سپس فریم پیش‌بینی شده و فریم اصلی به فضای ثانویه برده می‌شوند تا مدل بهینه‌تری از داده‌ها ساخته شود. در مطالعه [۱۷] شبکه‌ای چندمسیره برای پیش‌بینی فریم‌ها معرفی شده است که به ادعای نویسندگان آن، توانایی خوبی در فراگیری مفهومی اشیا و نواحی دارد.

در تحقیق ارائه‌شده در [۲۲] از داده‌های شار نوری برای ساخت ویژگی‌هایی استفاده شده است که قابل استفاده در تشخیص ناهنجاری هستند. در آن تحقیق با استفاده از هیستوگرام دامنه و جهت شار نوری و همچنین، مفهوم انتروپی تلاش شده است تا الگوهایی فراگرفته شوند که به تشخیص ناهنجاری کمک می‌کنند. در تحقیق [۲۳] از یک ساختار خود کدکننده با استفاده از لایه‌های کانولوشنی و CONV-LSTM استفاده شد تا فضای داده هنجار فراگرفته شود. برای آنکه اطلاعات مربوط به سرعت و جهت اشیا موجود در پیش‌زمینه، بهتر استخراج شوند، در مرحله استخراج ویژگی از شار نوری استفاده شده است. می‌توان با استفاده از یک ماسک برای تشخیص پیش‌زمینه و همچنین، انجام نمونه‌برداری در حوزه فضای تصویر، سرعت محاسبات شار نوری را افزایش داد [۲۴]. همچنین، با انتخاب ویژگی‌های مناسب در شار نوری و فیلترکردن این ویژگی‌ها میزان قابلیت اطمینان شار نوری در آن افزایش یافته است. در تحقیق ارائه‌شده در [۲۵] با معرفی ویژگی بصری بافت شار نوری که میزان همگن بودن یک میدان شار نوری را بررسی می‌کند، نمایشی از داده‌ها ارائه شده است که می‌تواند حرکت در تصاویر را به‌منظور تشخیص ناهنجاری به‌طور مؤثری مدل کند.

با توجه به تحقیقات بررسی‌شده، روش کلی برای تشخیص ناهنجاری، فراگیری فضای حالت داده‌های عادی

است و برای انجام این امر می‌توان از روش‌های مبتنی بر خوشه‌بندی یا مبتنی بر خودرمنگارها استفاده کرد. در برخی از روش‌ها برای بهبود عملکرد سیستم، از مسیرهای حرکتی در داده‌های عادی هم استفاده می‌شود؛ اما عیب همه آنها مغفول‌ماندن مکان‌های عاری از هدف متحرک و ممنوعه است که در این مقاله بررسی شده است. در مقالات پیشین مرتبط با شار نوری، با استفاده از شار نوری یا نمونه‌های تغییر یافته آن، نمایشی از داده‌ها ایجاد شود که به دلیل دارا بودن ویژگی‌های حرکتی می‌تواند به‌طور مؤثر برای یادگیری فضای داده معمولی استفاده شود؛ اما در این مطالعه تلاش شده است تا با استفاده از شار نوری و ویژگی‌های استخراج شده از آن، درباره آستانه‌ای تصمیم گرفته شود که برای تشخیص ناهنجاری استفاده می‌شود.

۳- روش پیشنهادی

برای تشخیص ناهنجاری، از یک سیستم مبتنی بر پیش‌بینی فریم بعدی ویدئو استفاده می‌شود. دلیل استفاده از این روش، عملکرد بهتر آن نسبت به روش‌هایی مثل [۹] و [۱۳] است که از بازسازی فریم جاری استفاده می‌کنند. دلیل این امر آن است که در مقایسه با روش‌هایی که از بازسازی فریم جاری استفاده می‌کنند، خطای افزوده‌شده به پیش‌بینی فریم‌های ناهنجاری بیشتر از خطای اضافه‌شده به پیش‌بینی فریم‌های عادی است. برای بهبود کیفیت نتایج نهایی، در فرایند تصمیم‌گیری از یک نقشه شار نوری استفاده شده است که نواحی ممنوعه را به‌طور مؤثری مدل می‌کند.

۳-۱- ساختار روش پیشنهادی

معماری شبکه پیشنهادی برای تشخیص ناهنجاری‌های ویدئو در شکل (۱) نشان داده شده است. با توجه به شکل، این شبکه از سه لایه CONV-LSTM2D تشکیل شده که میان آنها لایه همسان‌ساز^۹ قرار داده شده است. این ساختار در ورودی خود ۹ فریم را دریافت می‌کند و در خروجی خود نیز ۹ فریم ایجاد می‌کند که هرکدام فریم پیش‌بینی شده

موجود در نواحی مختلف تصویر ساخته می‌شود. ایده استفاده از نقشه شار نوری این است که در صورتی که در یک ناحیه از تصویر که در فاز آموزش حرکتی دیده نشده است، حرکتی مشاهده شود، احتمالاً یک ناهنجاری در تصویر وجود دارد و بنابراین، با کاهش آستانه، شرایط برای تشخیص داده شدن فریم به عنوان ناهنجاری تسهیل می‌شود. نکته شایان توجه اینکه نمی‌توان تنها به صرف وجود حرکت در نواحی ممنوعه، فریم را ناهنجاری اعلام کرد؛ دلیل این امر وجود نویز در شار نوری فریم‌ها است؛ اما با تنظیم آستانه می‌توان آن را به درستی استفاده کرد. مراحل مختلف روش پیشنهادی به صورت زیر (الف تا ه) است که به غیر از بند (ه)، بقیه روی داده‌های هنجار اعمال می‌شود.

الف) محاسبه شار نوری تمام فریم‌های هنجار داده‌های آموزشی:

برای محاسبه شار نوری لازم است ابتدا تصاویر با یک فیلتر گوسی مقداری مات شوند تا نویز در شار نوری کاهش یابد. برای این کار از یک فیلتر گوسی با ابعاد 3×3 استفاده می‌شود.

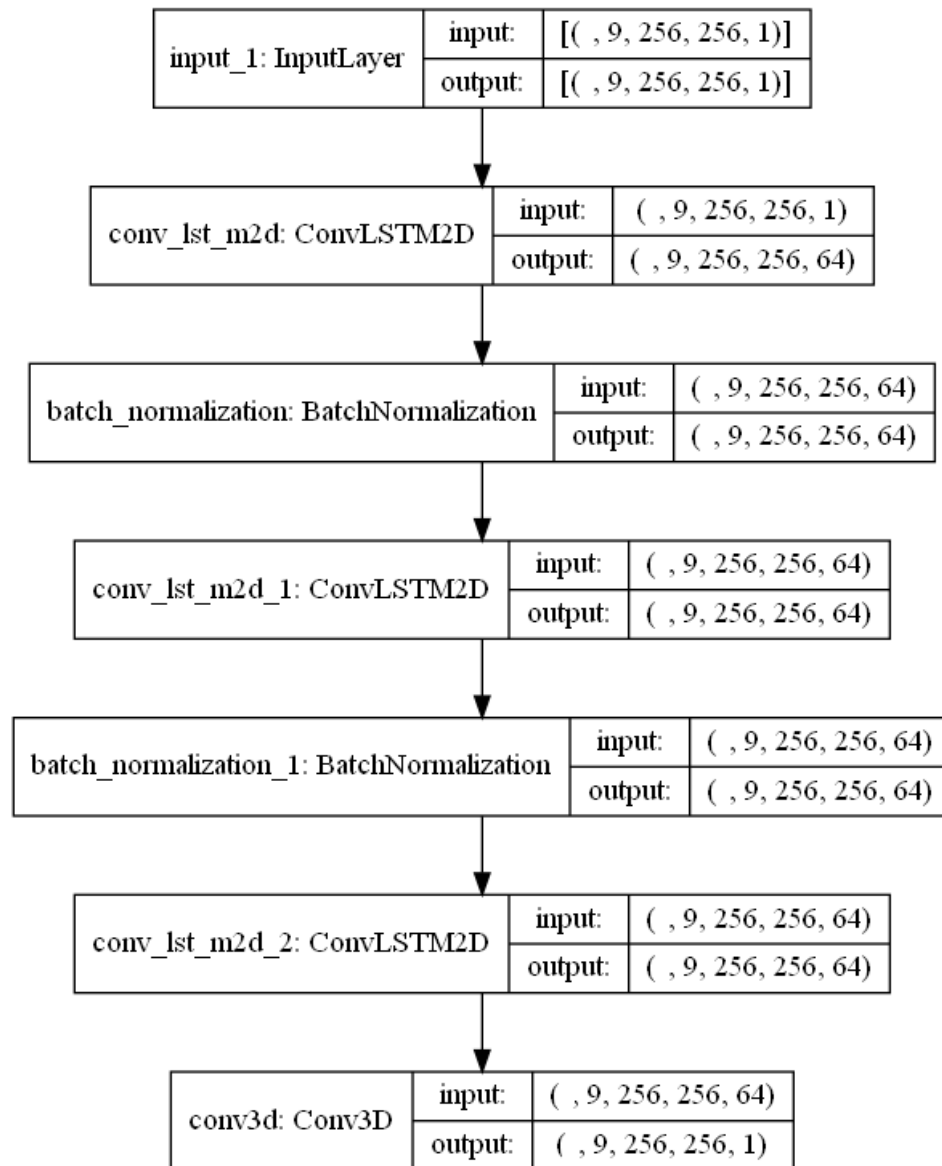
ب) محاسبه نقشه حرکت فریم‌ها:

بعد از اینکه شار نوری برای همه داده‌ها محاسبه شد، یک نقشه حرکت از طریق میانگین‌گیری روی تمامی شارهای نوری فریم‌ها ساخته می‌شود. تصویر حاصل شده، تصویری است که در آن شدت پیکسل‌ها در جاهایی که حرکت به صورت میانگین بیشتر بوده بالاتر است و هرچه جایی از تصویر تاریک‌تر باشد، در آن ناحیه حرکت کمتری وجود داشته است. برای ساخت نقشه حرکت فریم‌ها ابتدا تصویر حاصل در مرحله قبل بین صفر و یک همسان‌سازی می‌شود. سه نمونه از این تصاویر همسان‌شده در شکل (۲) آورده شده‌اند. مجدداً برای کاهش نویز، عمل مورفولوژیک سایش^۱ با پنجره مربعی 5×5 به آن اعمال می‌شود. نتیجه نهایی پس از پردازش با عمل سایش و میانگین‌گیری روی فریم‌های آموزشی در شکل (۳) نشان داده شده است.

بعدی برای فریم متناظر خود در ورودی است. به عبارت دیگر، این شبکه ۹ فریم دریافت می‌کند و برای هر کدام از آنها یک پیش‌بینی برای فریم بعدی ایجاد می‌کند. دلیل استفاده از لایه‌های CONV LSTM2D توانایی بالای آنها در فراگیری داده‌های مکانی - زمانی موجود در ویدئوها است. طراحی این لایه‌ها به صورتی انجام شده است تا هم توانایی لایه‌های کانولوشنی در فراگیری داده‌های دوبعدی را دارا باشند و هم از قابلیت‌های لایه‌های LSTM در فراگیری داده‌های زمانی بهره ببرند. با توجه به ماهیت ویدئو، استفاده از این لایه‌ها منطقی به نظر می‌رسد. برای آموزش این شبکه از روش بدون ناظر به این صورت استفاده می‌شود که در ورودی ۹ فریم از ویدئوهای هنجار به عنوان ورودی داده می‌شود و در خروجی از شبکه انتظار می‌رود ۹ پیش‌بینی برای ورودی‌ها ایجاد کند. با توجه به در دسترس بودن فریم‌های بعدی، می‌توان فریم‌های بعدی را خروجی مورد انتظار در نظر گرفت. از بهینه‌ساز ADAM برای کمینه کردن خطای بازسازی استفاده می‌شود و میزان خطای بازسازی با میانگین مجذور خطا بین فریم‌های بعدی واقعی و پیش‌بینی انجام شده محاسبه می‌شود. در صورتی که فرایند آموزش به درستی انجام شود، انتظار می‌رود شبکه به خوبی فریم‌های هنجار را بازسازی کند؛ اما از بازسازی فریم‌های ناهنجاری با کیفیت خوب ناتوان باشد. با توجه به مطالب گفته شده، در صورتی که خطای بازسازی در فاز تست از آستانه از پیش تعیین شده فراتر برود، آن فریم ناهنجار تشخیص داده می‌شود.

۳-۲- نقشه شار نوری

در بخشی از ناهنجاری‌های موجود در ویدئوها، حضور افراد، حیوانات یا وسایل نقلیه در برخی نواحی ممنوع است؛ برای مثال، حضور افراد روی چمن یا ورود افراد به نواحی ممنوعه، ناهنجاری تلقی می‌شود. برای تشخیص این نوع ناهنجاری‌ها، استفاده از نقشه شار نوری پیشنهاد می‌شود. نقشه شار نوری نقشه‌ای است که از روی میزان حرکت



شکل (۱): ساختار شبکه پیش‌بینی‌کننده فریم



شکل (۲): سه نمونه از تصاویر شار نوری همسان‌سازی شده بین صفر و یک (ردیف بالا) و فریم مربوطه به آنها (ردیف پایین)

که در آن OFM و $GOFM$ به ترتیب نقشه شار نوری و حالت شبکه‌بندی شده آن است. هدف این است که تصویر ایجاد شده به ۱۲ بلوک مساوی تقسیم شود و در نهایت به جای یک تصویر، یک ماتریس 4×3 به وجود آید که هر عنصر آن برابر با میانگین شار نوری در بلوک متناظر آن در نقشه نوری است.

برای انتخاب اندازه بلوک‌بندی (4×3) ، با توجه به شکل کلی تصاویر که مستطیل شکل است، شبکه‌بندی فریم‌ها به بلوک‌های مستطیلی $M \times N$ منطقی است. برای انتخاب دقیق مقادیر M و N حالات 5×4 و 3×2 نیز تست شدند. با بررسی بصری نتایج حاصل شده از این سه نوع دسته‌بندی نتیجه‌گیری شد که در حالت 3×2 به دلیل بیش از اندازه بزرگ بودن بلوک‌ها، دقت کافی به دست نمی‌آید. در حالت 5×4 دقت کافی است؛ اما مزیت خاصی نسبت به حالت 4×3 کسب نمی‌شود و از منظر کاهش بار محاسباتی، حالت 4×3 بهینه‌تر است.

ه) استفاده از نقشه شار نوری شبکه‌بندی شده در تشخیص ناهنجاری:

با انجام مراحل قبلی نقشه شار نوری شبکه‌بندی شده محاسبه شده است. حال در صورتی که در تعدادی از بلوک‌های تصویر بررسی شده، مقدار شار نوری شبکه‌بندی شده محاسبه شده برای یک فریم بیشتر از شار نوری شبکه‌بندی شده برای داده‌های هنجار بود، آستانه خطای بازسازی کاهش داده شود تا از این طریق احتمال ناهنجاری فریم افزایش یابد.

برای استخراج میزان کاهش آستانه، با استفاده از حلقه For حالات مختلفی بررسی شدند و سیستم با مقادیر مختلف کاهش آستانه تست شد. برای این کار، تمام مقادیر ممکن از کاهش ۳ درصدی تا کاهش ۵۰ درصدی با دقت ۱ درصد تست شدند و بهترین نتایج با دقت یکسان برای بازه ۱۴ درصد تا ۱۷ درصد حاصل شد که برای مقاله، مقدار ۱۵ درصد استفاده شد و با آزمایشی مشابه تعداد بلوک، حداقل دو به دست آمد.

به طور خلاصه، فریم‌های هنجار در هر قسمت از تصویر، شار نوری متوسط مشخصی دارند. اگر فریمی دارای نقشه شار نوری متفاوت از متوسط هنجارها داشت (در دو



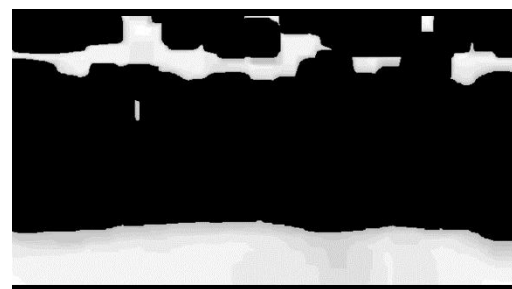
شکل (۳): نقشه حرکت فریم‌ها بعد از پردازش با عمل مورفولوژیک سایز

ج) ساخت نقشه شار نوری با استفاده از نقشه حرکت: در این مرحله، ابتدا نتیجه حاصل از تصویری تماماً یک و هم‌ابعاد با تصویر اصلی کم می‌شود، سپس مجدداً با عمل سایز، پردازش و در تصویر حاصل پیکسل‌های کمتر از ۰,۵ طبق رابطه (۱) صفر می‌شود؛ چون نیاز است این تصویر نواحی با میزان حرکت کمتر از متوسط را برای ما آشکار کند.

$$OFM(x, y) = \text{Erosion}(\text{ones}(M, N) - \text{Motion map}(x, y)) \quad (1)$$

$$> 0.5$$

که در آن M و N به ترتیب عرض و ارتفاع تصویرند و Motion map خروجی مرحله (ب) است. نقشه شار نوری حاصل در شکل (۴) نشان داده شده است.



شکل (۴): نقشه شار نوری پایگاه داده Avenue

د) ساخت نقشه نوری شبکه‌بندی شده:

برای ساخت نقشه شار نوری شبکه‌بندی شده کافی است از رابطه (۲) استفاده شود.

$$GOFM(x, y) = \frac{1}{\frac{M}{4} \times \frac{N}{3}} \sum_{i=M \times x/4}^{M \times (x+1)/4} \sum_{j=N \times y/3}^{N \times (y+1)/3} OFM(i, j) \quad (2)$$

به نواحی ممنوعه نیست، استفاده از این پایگاه داده برای بررسی روش پیشنهادی چندان مؤثر به نظر نمی‌رسد و به همین دلیل از این پایگاه داده استفاده نشد. در این پایگاه، اگر به‌طور مثال، دوچرخه در قسمتی از پیاده‌رو مجاز به حضور بود و در قسمتی دیگر نبود و وارد آن قسمت می‌شد، می‌توان برای تست الگوریتم خود از آن استفاده کرد. بنابراین، شبکه روی داده‌های آموزشی پایگاه داده Avenue به تعداد ۱۰ اپوک آموزش داده شده و از داده‌های تست همان پایگاه داده برای سنجش کیفیت عملکرد کلی سیستم استفاده شده است. معیار سنجش عملکرد روش‌های تشخیص ناهنجاری به دلیل نابرابری تعداد نمونه‌های هنجار و ناهنجار معمولاً سطح زیر منحنی^{۱۱} (AUC) نرخ نمونه مثبت صحیح^{۱۲} بر حسب نرخ نمونه مثبت اشتباه^{۱۳} است. برای رسم این نمودار لازم است مقادیر یادشده برای آستانه‌های مختلف، محاسبه و نسبت به هم رسم شوند. معیار AUC مساحت زیر این نمودار است. برای سیستم ایدئال، این مقدار یک خواهد بود و هرچه این مقدار به صفر نزدیک شود، سیستم عملکرد ضعیف‌تری خواهد داشت.

روش پیشنهادی با روش‌های ارائه شده در [۱۸]، [۱۹] و [۲۰] مقایسه شده است. در روش [۱۸]، از لایه‌های CONV LSTM2D برای ساخت یک خودرمنگار استفاده شده است تا با بهره‌گیری از آنها ویژگی‌های مکانی - زمانی به‌خوبی فرا گرفته شود. در تحقیق [۱۹] از خودرمنگار نویززا استفاده شده است که نسبت به خودرمنگار عادی عملکرد بهتری دارد. در روش [۲۰] از یک خودرمنگار ساخته شده با لایه‌های کانولوشن دوبعدی استفاده شده است. این روش‌ها و روش پیشنهادی، آزمایش و نتایج در جدول (۱) آورده شدند.

جدول (۱): مقایسه عملکرد روش پیشنهادی با روش‌های

موجود

روش	(AUC) مساحت زیر منحنی
ارائه شده در [۱۸]	۷۷
ارائه شده در [۱۹]	۸۱٫۷
ارائه شده در [۲۰]	۷۴٫۵
روش پیشنهادی	۸۱٫۷

بلوک یا بیشتر)، آن فریم با سهولت بیشتری (کاهش سطح آستانه) ناهنجار اعلام می‌شود. درخور ذکر است میزان اثرگذاری این روش به میزان وجود ناهنجاری از جنس حضور شیء یا شخص در مناطق ممنوعه بستگی دارد. شبه‌کد الگوریتم در شکل (۵) آورده شده است.

```
for frames in video
  Threshold = Initial_Threshold
  Suspicious_blocks = 0
  for all sub-blocks in the frame
    if GOFM_Frame > GOFM_Normal
      Suspicious_blocks = Suspicious_blocks + 1
  if Suspicious_blocks >= 2
    Threshold = Threshold * 0.85
  if Reconstruction_error_frame > Threshold
    frame_label = 'Anomaly'
  else
    frame_label = 'Normal'
```

شکل (۵): شبه‌کد الگوریتم پیشنهادی

۴- نتایج عملی

برای پیاده‌سازی شبکه پیشنهادی و همچنین، ایجاد نقشه شار نوری شبکه‌بندی شده از پایتون نسخه ۳٫۸ و کتابخانه‌های keras و opencv استفاده شده است. برای انجام فرایند آموزش از کامپیوتری مجهز به دو کارت گرافیک Nvidia 3060 ti، 64GB حافظه رم و یک CPU نسل ۱۰ هفت هسته‌ای استفاده شده است.

برای بررسی تشخیص ناهنجاری در تصاویر ویدئویی دو پایگاه داده بسیار محبوب Avenue [۲۱] و UCSD [۲۶] در بین پژوهشگران وجود دارد. پایگاه داده Avenue شامل تصاویر یک پیاده‌رو و رفت‌وآمد عابران پیاده به‌عنوان داده معمولی است. برای داده تست از تعدادی ویدئو استفاده می‌شود که در آنها عابران اعمالی را انجام می‌دهند که در داده‌های معمولی دیده نشده است. این اعمال شامل دویدن در پیاده‌رو، عبور از یک ناحیه‌ای که عابران پیاده از آن نواحی در داده‌های معمولی عبور نکرده‌اند، پرتاب اشیاء به سمت بالا و غیره می‌شود. تصاویر پایگاه داده UCSD شامل تصاویر نواحی دو پیاده‌رو از دو زاویه مختلف با عناوین UCSD Ped1 و UCSD Ped2 است. ناهنجاری‌های موجود در این پایگاه داده شامل عبور دوچرخه، اسکیت سوار، ویلچیر و خودرو از پیاده‌رو است. با توجه به اینکه ذات ناهنجاری‌های موجود در این پایگاه داده از نوع ورود

دلیل اثرگذاری شار نوری، کمک در تشخیص ناهنجاری‌هایی مانند شکل (۶) است. در این فریم، شخص عابر وارد فضایی شده است که در داده‌های آموزشی شخصی از آن عبور نمی‌کند. مزیت روش ما در داده‌هایی دیده می‌شود که ناهنجاری به صورت ورود فرد یا شیء به منطقه‌ای ممنوعه است. روش‌های دیگر به دلیل آنکه به این مهم توجه نداشته‌اند، نمونه را نمونه نهنجار در نظر خواهند گرفت؛ حال آنکه در روش ما به دلیل کاهش آستانه ناشی از *OFM* کم، تشخیص ناهنجاری به درستی صورت می‌گیرد.

۵- نتیجه‌گیری

در این تحقیق، با استفاده از شبکه پیش‌بین فریم بعدی ویدئو و یک نقشه شار نوری شبکه‌بندی‌شده، روشی برای بهبود نتایج تشخیص ناهنجاری ارائه و ارزیابی شده است. نتایج عملی حاصل‌شده از تست شبکه روی پایگاه داده Avenue حاکی از آن است که شبکه، نحوه بازسازی داده‌های هنجار را به درستی آموزش دیده است و داده‌های ناهنجار را با مقایسه خطای بازسازی با یک آستانه تشخیص می‌دهد. میزان آستانه با توجه به محتوای نقشه شار نوری شبکه‌بندی‌شده تنظیم می‌شود. محتوای نقشه شار نوری بررسی ورود به نواحی ممنوعه در فریم‌ها استفاده می‌شود. مدل‌سازی نواحی ممنوعه در این کاربرد برای نخستین بار مطرح و بررسی شدند. دقت روش نسبت به غالب روش‌های مقایسه‌شده بیشتر است؛ اما با یک روش مساوی است که علت آن محدودیت پایگاه‌های داده موجود است. علاوه بر این، رویکرد معرفی‌شده برای مدل‌کردن نواحی ممنوعه کلی است و می‌تواند به خوبی به سایر سیستم‌های تشخیص ناهنجاری ویدئویی اضافه شود.

مراجع

- [1] Z. Wang, Y. Zou, and Z. Zhang, "Cluster Attention Contrast for Video Anomaly Detection". Proceedings of the 28th ACM International Conference on Multimedia, 2020.
- [2] L. Hao, A. Achim, and D. Bull, "Unsupervised video anomaly detection using feature clustering", IET signal processing, Vol. 6, No. 5, pp. 521-533, 2012
- [3] Y. Chang, Z. Tu, W. Xie, J. Yuan, "Clustering driven deep autoencoder for video anomaly

همان‌گونه که پیداست روش ما توانست نتایج بهتری در مقایسه با غالب روش‌ها به دست آورد و در واقع در رتبه اول مشترک با روش [۱۹] قرار دارد؛ اما ضعف روش [۱۹] این است که خطای تمامی نواحی به صورت یکسان نگاه می‌شوند و برخلاف روش ما برای بررسی نواحی ممنوعه به طور دقیق راهکاری ندارد. روی این موضوع، با وجود کاربردی و مهم بودن، در پایگاه‌های داده موجود تأکید نشده است. به عبارت دیگر، اگر پایگاه داده دیگری با پوشش بیشتر نواحی ممنوعه وجود داشت، روش پیشنهادی ما بهتر بود.

نکته مهم‌تر اینکه، روش ارائه‌شده ما می‌تواند به رویکرد و شبکه مقاله [۱۹] یا سایر مقالات، افزوده شود و به بهبود نتایج کمک کند. ساختار پیشنهادی، بدون ایجاد تغییر در معماری شبکه‌های عصبی و صرفاً با بررسی اطلاعات مربوط به شار نوری که از فریم‌ها به دست می‌آید با تنظیم آستانه تصمیم‌گیری به بهبود تشخیص ناهنجاری‌ها کمک می‌کند. برای تست این روش روی شبکه‌های دیگر نیازی به تغییر ساختار و ایجاد تغییرات اساسی در آنها نیست و صرفاً کافی است آستانه روش‌های دیگر با آستانه پیشنهادی توسط این سیستم جایگزین شود؛ اما برای انجام مقایسه دقیق و منصفانه با دیگر کارها، لازم است شبکه آموزش داده شده با وزن‌های دقیق در دسترس باشد. با توجه به اینکه پژوهشگران، پارامترهای آموزش شبکه‌ها را در بسیاری از موارد از جمله تعداد اپوک، نوع بهینه‌ساز، وزن‌های اولیه استفاده‌شده در شبکه و غیره گزارش ندادند، امکان پیاده‌سازی دقیق شبکه‌های پیشنهادی ممکن نیست و این امر امکان مقایسه صحیح با سایر کارها را ناممکن می‌سازد.



شکل (۶): نمونه‌ای از فریم ناهنجاری تشخیص داده شده

به وسیله نقشه شار نوری

- [15] Y. Lai, R. Liu, and Y. Han, "Video anomaly detection via predictive autoencoder with gradient-based attention", IEEE International Conference on Multimedia and Expo, 2020
- [16] Y. Zhang, X. Nie, R. He, M. Chen, and Y. Yin, "Normality learning in multispace for video anomaly detection," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 31, No. 9, pp. 3694-3706, 2020
- [17] X. Wang, Z. Che, B. Jiang, N. Xiao, K. Yang, J. Tang, J. Ye, J. Wang, Q. Qi, "Robust unsupervised video anomaly detection by multipath frame prediction", IEEE Transactions on Neural Networks and Learning Systems, Vol. 23, No. 6, pp. 2301-2312, 2022.
- [18] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional lstm for anomaly detection", In 2017 IEEE International Conference on Multimedia and Expo, pp. 439-444, 2017
- [19] P. Vincent, H. Larochelle, Y. Bengio, and P.A. Manzagol, "Extracting and composing robust features with denoising autoencoders", In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103, 2008.
- [20] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733-742, 2016.
- [21] http://www.cse.cuhk.edu.hk/leojia/projects/detect_abnormal/dataset.html
- [22] R.V.H.M. Colque, C. Caetano, M. T. L. de Andrade and W. R. Schwartz, "Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 27, No. 3, pp. 673-682, 2016.
- [23] E. Duman, and Q. A. Erdem, "Anomaly detection in videos using optical flow and convolutional autoencoder", IEEE Access, Vol. 7, pp.183914-183923, 2019.
- [24] H. Tan, Y. Zhai, Y. Liu, and M. Zhang, "March. Fast anomaly detection in traffic surveillance video based on robust sparse optical flow", In IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 1976-1980, 2016
- [25] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "August textures of optical flow for real-time anomaly detection in crowds. In IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 230-235, 2011.
- [26] <http://www.svcl.ucsd.edu/projects/anomaly/dataset.html>
- [4] C. Piciarelli, C. Micheloni, and G.L. Foresti, "Trajectory-based anomalous event detection", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 11, pp. 1544-1554, 2008
- [5] M. Asad, J. Yang, E. Tu, L. Chen, and X. He, "Anomaly3D: Video anomaly detection based on 3D-normality clusters", Journal of Visual Communication and Image Representation, Vol. 75, 2021
- [6] Z. Fu, W. Hu, T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection", In IEEE International Conference on Image Processing, pp. II-602, 2005
- [7] D. Singh, C. K. Mohan, "Deep spatio-temporal representation for detection of road accidents using stacked autoencoder", IEEE Transactions on Intelligent Transportation Systems, Vol 20, No. 3, pp. 879-887, 2018
- [8] M. Sabokrou, M. Fathy, M. Hoseini, "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder", Electronics Letters, Vol. 52, No. 13, pp. 1122-1124, 2016
- [9] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes", IEEE Transactions on Image Processing, Vol. 26, No. 4, pp. 1992-2004, 2017.
- [10] T. Wang, M. Qiao, Z. Lin, C. Li, H. Snoussi, Z. Liu, and C. Choi, "Generative neural networks for anomaly detection in crowded scenes", IEEE Transactions on Information Forensics and Security, Vol. 14, No. 5, pp. 1390-1399, 2018
- [11] J. Sun, X. Wang, N. Xiong and J. Shao, "Learning sparse representation with variational auto-encoder for anomaly detection", IEEE Access, Vol. 6, pp. 33353-33361, 2018.
- [12] W. Chu, H. Xue, C. Yao, D. Cai, "Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos", IEEE Transactions on Multimedia, Vol. 21, No. 1, pp. 246-255, 2018.
- [13] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes", Computer Vision and Image Understanding, Vol. pp. 172, 88-97, 2018
- [14] M. Ye, X. Peng, W. Gan, W. Wu, and Y. Qiao, "Anopcn: Video anomaly detection via deep predictive coding network", In Proceedings of the 27th ACM International Conference on Multimedia, pp. 1805-1813, 2019.

¹ Labeled data

- ² Encoder
- ³ Autoencoder
- ⁴ Variational Autoencoder
- ⁵ Denoising Autoencoder
- ⁶ Stacked deep Autoencoder
- ⁷ Reconstructing decoder
- ⁸ Predictive decoder
- ⁹ Batch normalization
- ¹⁰ Erosion
- ¹¹ Area Under Curve
- ¹² True Positive Rate
- ¹³ False Positive Rate

