

شناسایی رفتارهای ناهنجار در تصاویر ویدئویی با استفاده از شبکه عصبی کانولوشنی

بهنام سبزعلیان^۱، حسین مروی^۲ و علیرضا احمدی‌فرد^۳

چکیده

شناسایی رفتار ناهنجار از لحاظ اهمیت یک ضرورت بصری تبدیل شده است. همچنین این حوزه به عنوان یک چالش در تحقیقات بینایی ماشین بدل شده است. گرچه تلاش‌های بسیاری به منظور رفع این مشکل انجام شده است، اما شناسایی رفتار در یک محیط واقعی و غیرقابل کنترل فاصله معناداری تا به بلوغ رسیدن آن وجود دارد. مشکل اصلی ابهام در تفاوت خصوصیات رفتار غیر نرمال و نرمال است که تعریف آن معمولاً با توجه به زمینه پیشین تصاویر می‌تواند متفاوت باشد. در این مقاله یک سیستم شناسایی و موقعیت‌یابی رفتارهای ناهنجار در سکانس‌های ویدئویی ارائه شده است. جنبه کلیدی این روش درواقع ترکیب شبکه عصبی کانولوشنی زمان-مکانی دو بعدی و سه بعدی به منظور شناسایی رفتار غیر نرمال در فریم‌های متوالی ویدئویی است. همچنین از روش شناساگر (FAST) Features from Accelerated Segment Test به منظور افزایش ضریب اطمینان در شناسایی موقعیت‌های موردنظر در تصاویر ورودی به مدل شبکه عصبی کانولوشنی بهره گرفته شده است. این ویژگی‌ها تنها از حجم پیکسل‌های دارای حرکت استخراج می‌شوند تا بتوانند هزینه محاسبه را کاهش دهند. ساختار مدل شبکه عصبی کانولوشنی به ما اجازه استخراج ویژگی‌های زمان-مکانی که شامل استخراج ویژگی‌هایی با حرکات پیچیده نیز هست را می‌دهد. روش ارائه شده توسط مجموعه داده‌ی متداول که شامل رفتارها و اعمال ناهنجار متفاوت انسانی در موقعیت‌های گوناگون است، مورد آزمایش و ارزیابی قرار گرفته است. نتایج حاصل از آزمایش‌های مربوطه نمایانگر این است که سیستم ارائه شده در مقایسه با بسیاری از روش‌های متداول پیشین، عملکرد بهتری را دارد و کارایی آن در شناسایی رفتار غیر نرمال در مقایسه با روش‌های قبلی بسیار رقابتی است.

کلیدواژه‌ها

بینایی ماشین، شناسایی رفتارهای ناهنجار در تصاویر ویدئویی، شبکه‌های عصبی کانولوشنی، یادگیری ماشین، ویژگی‌های زمان-مکانی (CNN).

۱ مقدمه

این مقاله در مردادماه ۱۳۹۶ دریافت، در خردادماه ۱۳۹۷ بازنگری و در مردادماه همان سال پذیرفته شد.

^۱ دانشجوی کارشناسی ارشد مهندسی رباتیک، دانشکده برق و رباتیک دانشگاه صنعتی شاهروド

رایانامه: behnamsabzalian@gmail.com

^۲ گروه رباتیک، دانشکده برق و رباتیک دانشگاه صنعتی شاهروド ahmadyfard@shahroodut.ac.ir و h.marvi@shahroodut.ac.ir

نویسنده مسئول: بهنام سبزعلیان

Archive of SID

این شبکه با اعمال بر روی توده‌های زمان-مکانی موردنظر، اطلاعات حرکتی مؤثر را در این توده‌ها استخراج می‌کند و مشخص می‌کند کدام به عنوان یک توده نامتعارف شناخته می‌شود. نتایج آزمایش‌های انجام شده بر روی مجموعه داده‌های متدالو نمایان گر این است که روش پیشنهادی، نتایج بسیار رقابتی را در قیاس با روش‌های پیشین ارائه شده کسب کرده است.

این مقاله در ۶ بخش سازماندهی شده است، در بخش ۲ به مرور پژوهش‌های پیشین می‌پردازیم، چهارچوب کلی، معرفی روش پیشنهادی و نحوه استخراج ویژگی‌ها در بخش ۳ پرداخته خواهد شد. در بخش ۴ نتایج آزمایش‌ها و مقایسه روش ارائه شده با روش پیشین مورد بررسی قرار می‌گیرد. بخش‌های ۵ و ۶ به ترتیب به جمع‌بندی و تقدیر و تشکر در این مقاله می‌پردازد.

۲ مروری بر پژوهش‌های پیشین

مرجع [۱] به شناسایی ناهنجاری در حوزه‌های زمان و مکانی می‌پردازد. یکی دیگر از مشکلات در این حوزه نداشتن یک تعریف مشخص از رفتار ناهنجار است. در این تحقیق رفتارهایی که به‌ندرت در تصاویر ویدئویی اتفاق می‌افتد را به عنوان رفتار نامتعارف محسوب می‌کند [۲-۶]. روش‌هایی که برای شناسایی ناهنجاری درگذشته ارائه گردیده است عمدهاً بر مبنای استخراج ویژگی‌ها از مسیرهای شناسایی شده اشیا و یا تغییرات زمان-مکانی حادث در تصاویر ویدئویی است.

برای ذکر نمونه، روش ارائه شده در [۷, ۸] که بر روی مسیرهای شناسایی شده اشیا متمرکز است بر اساس اینکه کدامیک مسیری متعارف در پیش دارد برچسب نرمال و غیر نرمال بودن را به آن اختصاص می‌دهد. این روش‌ها در مواقعي که انسدادی وجود دارد و یا تغییر در روشنایی تصاویر ایجاد می‌شود کارایی خود را به طور محسوسی از دست می‌دهند و همچنین در مواردی که صحنه‌های شلوغی در تصاویر پدید می‌آید پیچیدگی محاسباتی بالایی دارند. بنابراین محققان روش‌هایی را ارائه کردند که از ویژگی‌های سطح پایین مانند شارنوری^۴ و گرادیان استفاده کنند تا ابعاد و ارتباط زمان-مکانی را در این گونه ویژگی‌ها آموختند.

در مقایسه‌ی روش‌های پیشین، این متدانهای می‌توانند بر مبنای مدل [۹, ۵, ۶] یا خوشبندی بر مبنای ویژگی‌های برجسته [۴, ۱۰-۱۱] باشند. در تحقیقات اخیر تنکسازی رویدادها^۵ [۱۲-۱۴] در تصاویر ویدئویی به صورت قابل ملاحظه‌ای مورد توجه قرار گرفته است. نتایج مدل‌های ارائه شده در [۱۷-۱۲] حاکی از این است که کارایی قابل قبولی در شناسایی ناهنجاری داشته‌اند.

با توجه به رشد فزاینده دوربین‌های امنیتی و باره‌زینه‌ای زیاد نظارت انسانی، انگیزه استفاده از سامانه‌های امنیتی خودکار را افزایش می‌دهد. در واقع این سامانه‌ها تلاش دارند کارایی استفاده از دوربین‌های نظارتی را بهبود بخشنند. یکی از مزایای استفاده از سامانه‌های تشخیص رفتار ناهنجار پیشرفت استفاده از دوربین‌های نظارتی در تجارت‌های کوچک و یا خانه‌ها است. سیستم نظارت دوربین‌ها علاوه بر این‌که می‌تواند از وقوع جرم جلوگیری کند و به عنوان یک ابزار آنالیز حوادث مورد استفاده قرار گیرد همچنین می‌تواند به عنوان یک سامانه کمکی در حوادث ناگوار دخالت داشته باشد. هدف اصلی از این تحقیقات یافتن روشی برای شناسایی رفتار ناهنجار است.

یک رفتار ناهنجار، رفتاری است که در تصاویر ویدیویی از یک محیط مورد تحقیق تعریف شده است، اما در آن اتفاق می‌افتد. در حالت کلی این رفتار توسط انسان‌ها واقع می‌گردد. یک رویداد می‌تواند حتی توسط انسان انجام نشود بلکه توسط اشیای دیگر مانند وسایل نقلیه یا حیوانات انجام گیرد. تعریف خاصی که می‌توان از رفتارهای ناهنجار داشت بسیار به محیط موردنظر بستگی دارد. به دلیل اینکه این رفتارها در صحنه‌های بسیاری قابل پذیرش است و به طور معمول اتفاق می‌افتد بنابراین می‌توان گفت که رفتارهای ناهنجار به صورت معمول اتفاق نمی‌افتد و در اذهان عمومی ممکن است قابل پذیرش نباشد. چالش یک سیستم خودکار این است که بین رفتارهای حادث، رفتارهای نرمال از غیر نرمال را مشخص کند. از چالش‌هایی که سامانه‌های شناسایی رفتارهای ناهنجار با آن‌ها مواجه هستند می‌توان به یافتن تعریف دقیقی از رفتار نرمال از غیر نرمال در یک محیط خاص، یافتن اطلاعات تصویری مناسب، انتخاب روش‌های مناسب آموزش و طبقه‌بندی و به دست آوردن نتایجی با نرخ FP^۱ کم و نرخ TP^۲ بالا اشاره داشت. قسمت اعظم تحقیقات در حوزه‌های یادگیری ماشین و بینایی ماشین بر روی بهبود کارایی سامانه‌های نظارت بصری متکرکشده است. اهداف این تحقیقات در ابتدا شناسایی محدودیت‌هایی است که در محیط‌های واقعی اتفاق می‌افتد و بعد از آن آزمایش و ارائه مجموعه روش‌هایی که می‌توانند تا حدودی بر شرایط چالش‌های طبیعی فائق آیند.

در این مقاله به منظور شناسایی و موقعیت‌یابی رفتارهای ناهنجار در تصاویر ویدئویی به معرفی یک مدل شبکه عصبی کانولوشنی زمان-مکانی می‌پردازیم که علاوه بر اینکه به اطلاعات بصری موجود در تصاویر ایستادسترسی دارد، همچنین می‌تواند اطلاعات حرکتی پیچیده را نیز از فریم‌های متوالی استخراج کند. برای شناسایی رفتارهای نامتعارف در قسمت کوچکی از یک فریم، مدل زمان-مکانی شبکه عصبی کانولوشنی بر روی توده‌های زمان-مکانی^۳ از تصاویر اعمال می‌شود.

¹ False-Positive

² True-Positive

³ Spatial-Temporal Patches

Archive of SID

از یک Auto-Encoder برای حذف تپیر بهمنظر بارسازی تصاویر ورودی و آموزش ویژگی‌های موردنظر در یک شبکه عمیق استفاده شده است. گرچه در این مقاله از یک طبقه بند SVM یک کلاسه در لایه‌ی بالایی مدل استفاده کرده است. اما این طبقه بند به صورت بهینه با ویژگی‌های شبکه عمیق طراحی شده منطبق نشده است. روشی که در [۲۸] ارائه شده است به معنی یک شبکه عصبی کانولوشنی بهبودیافته بهمنظر شناسایی رفتار انسان در تصاویر ویدئویی سه‌بعدی RGB-D می‌پردازد. روش ارائه شده در [۲۹] با بهره‌گیری از یک شبکه PCANet بهمنظر استخراج ویژگی‌های حرکتی از گریدیان سه‌بعدی به ساخت یک مدل ترکیبی گوسی عمیق برای مدل کردن الگوی رویدادهای نرمال پرداخته است. در مقاله [۳۰] از یک شبکه عصبی کانولوشنی سه‌بعدی برای شناسایی محل فرود یک بالگرد خودکار استفاده می‌کند.

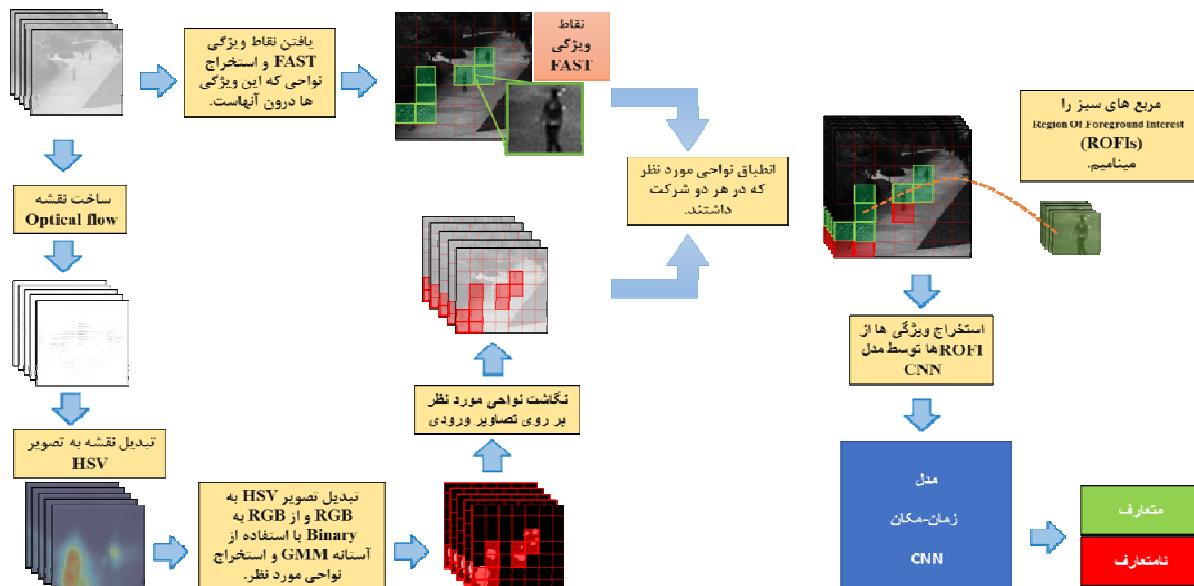
در مقاله پیش رو از ترکیب لایه‌های کانولوشن در ابعاد دو‌بعدی و سه‌بعدی، بهمنظر استخراج ویژگی‌های زمان-مکانی مقاوم در برابر برخی از مشکلات نظیر تغییر روش‌نمایی شدید، لرزش دوربین و ...، استفاده شده و به طراحی یک شبکه عصبی کانولوشنی نوین انجامیده است. همچنین از روش شناساگر FAST بهمنظر افزایش اطمینان در شناسایی موقعیت‌های موردنظر در تصاویر ورودی به مدل شبکه عصبی کانولوشنی بهره گرفته شده است. این ویژگی‌ها تنها از پیکسل‌های دارای حرکت استخراج می‌شوند تا بتوانند هزینه محاسبه را کاهش دهند.

اغلب روش‌های ارائه شده در این زمینه از یک چهارچوب کلی برای شناسایی الگوی موردنظر پیروی می‌کنند. این چهارچوب کلی شامل دو مرحله است که مرحله اول آن محاسبه ویژگی‌های دستی از فریم‌های خام ویدئویی است و دومین مرحله شامل آموزش طبقه بندها بر مبنای ویژگی‌های استخراج شده است.

در حالت واقعی دانستن این مسئله که کدام ویژگی در شناسایی رفتار غیر نرمال نقش مؤثر دارد بسیار مشکل است. در حالت کلی ویژگی‌هایی اعم از ویژگی‌های زمان-مکان، گرادیان، اطلاعات شارنوری بهمنظر شناسایی موقعیتی که رویداد نامتعارفی در آن رخداده است، مورد استفاده قرار می‌گیرد.

اخیراً کارهای بسیاری ارائه شده است که توانایی شبکه‌های عصبی کانولوشنی [۱۸] را در حوزه‌های مختلف بینایی ماشین مانند شناسایی متن [۱۹] و شناسایی طبقه‌بندی اشیا [۲۰، ۲۱]، شناسایی لبه [۲۲] و شناسایی صورت انسان [۲۳] نشان می‌دهد.

در حوزه طبقه‌بندی ویدیو نیز شبکه‌های عصبی کانولوشنی نقش مؤثری را بازی می‌کنند. شبکه عصبی کانولوشنی سه‌بعدی در [۲۴] برای شناسایی اعمال و رفتار انسان در تصاویر ویدئویی ارائه شده است. روش ارائه شده در [۲۵] مبتنی بر توده‌های تصویری که با بهره‌گیری از یک شبکه عصبی عمیق سه‌بعدی به صورت آ بشاری رویدادهای ناهمجارت در صحنه‌های ویدئویی را شناسایی کند. اخیراً روشی در [۲۶] ارائه گردیده است که با استفاده از یک شبکه عصبی کانولوشنی سه‌بعدی بر مبنای Auto-Encoder سعی دارد از طریق ویژگی‌های بصری و حرکتی رفتارهای ناهمجارت را شناسایی کند، همچنین در روشی دیگر [۲۷]



شکل ۱ شماتیکی روش ارائه شده در این مقاله

Archive of SID

می‌کند و نواحی موردنظر را در آن‌ها می‌یابد. این مکعب‌ها در واقع نواحی محلی مکان-زمانی موردنظر را نیز توصیف می‌کنند. نواحی مربوط به اعمال نرمال معمولاً ارتباطات مشابهی بین همسایگان خود دارند و همچنین احتمال وقوع آن‌ها بالاتر است. اما در مورد اعمال غیر نرمال تشابه بین این نواحی و نواحی در همسایگی آن‌ها یک الگوی واحد را نتیجه نمی‌دهد. بدیهی است که احتمال وقوع نواحی غیر نرمال به مراتب کمتر از نواحی معمول است.

۲-۳ نواحی پیش‌زمینه و شارنوی

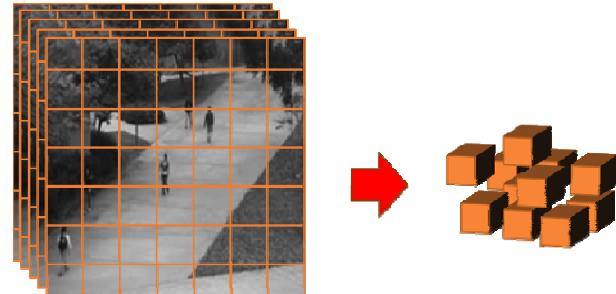
اطلاعات مربوط به پیش‌زمینه تصاویر برای شناسایی رفتار غیرمعمول در ویدیو به منظور استخراج نواحی پیش‌زمینه موردنظر Region Of Foreground Interests (ROFIs) بسیار اهمیت دارد. بنابراین نقشه شارنوی که مزایای آن در شناسایی حرکات و ردگیری اشیا در بسیاری از حوزه‌های بینایی ماشین به اثبات رسیده است، محاسبه می‌شود. در ادامه توسط روش حذف پشت زمینه در این نقشه که با استفاده از روش مدل‌های مخلوط گوسی (GMM) Gaussian Mixture Model انجام می‌شود، این نقشه را به تصویر بازنگری تبدیل می‌کند. در این مرحله نواحی کلیدی در تصویر بازنگری به دست می‌آید که این نواحی در تصویر ورودی نگاشت می‌شوند. شکل ۳ چگونگی انتخاب نواحی موردنظر را در تصویر بازنگری نشان می‌دهد.

۳ معرفی روش پیشنهادی

شمای کلی از روش پیشنهادی در این مقاله، در شکل ۱ نشان داده شده است. جزئیات مربوط به این روش در ادامه شرح داده خواهد شد.

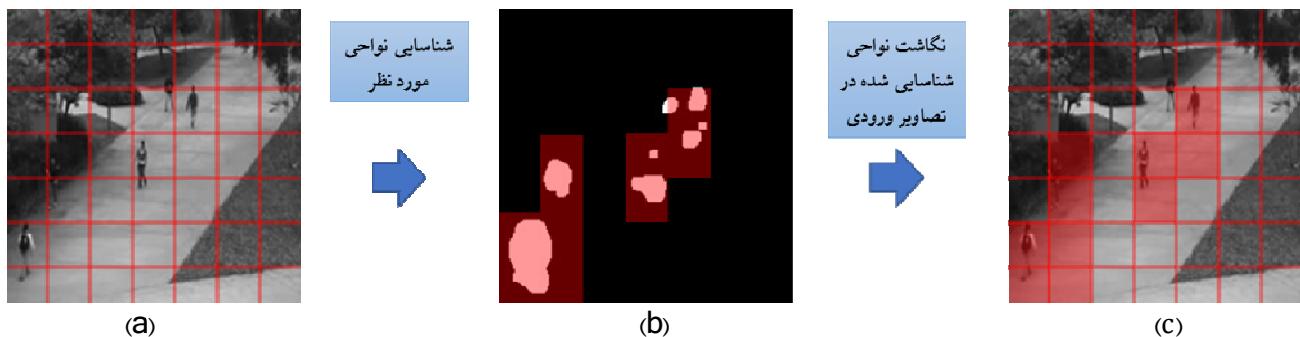
۱-۳ نحوه نمایش ویدیو

به منظور آماده‌سازی ویدیو ابتدا می‌باشد هر یک از فریم‌های ویدئویی را به تعدادی سلول‌های مکعبی غیر همپوشان تبدیل کنیم که در شکل ۲ نشان داده شده است.



شکل ۲ نحوه نمایش ویدیو. هر یک از فریم‌های ویدیو به صورت تعدادی سلول‌های مکعبی غیر همپوشان تبدیل می‌شوند.

اندازه این نواحی غیر همپوشان برابر هستند و با توجه به نواحی به دست آمده تعداد هفت فریم متوالی با یکدیگر تشکیل یک مکعب زمان-مکانی را می‌دهد. روش ارائه شده به منظور شناسایی رویدادهای غیرمعمول از همین سلول‌های غیر همپوشان استفاده



شکل ۳ شماتیک از استخراج نواحی موردنظر. a) فریم ورودی و سلول‌های غیر همپوشان b) تصویر بازنگری شارنوی و شناسایی نواحی موردنظر c) نگاشت این نواحی در فریم ورودی.

اندازه اشیای یافته شده و همچنین حضور این اشیا در فریم‌های مختلف را می‌تواند به‌آسانی با شمارش تعداد پیکسل‌های پیش‌زمینه در هر سلول مکعبی مانند B_t به دست آورد.

$$\sum_{d=1}^{m_t} \sum_{j=1}^{m_y} \sum_{x=1}^{m_x} u(i, j, d) \quad (1)$$

تنها سلول‌هایی که دارای حداقل آستانه اشغال پیش‌زمینه $F(u)$ هستند به عنوان سلول‌های فعال به شمار می‌روند. به طور معمول اگر سلولی حداقل ۱۰٪ پیکسل‌هایش مربوط به پیش‌زمینه باشد به عنوان سلول فعال شناخته می‌شود.

اگر هر کدام از نواحی شناسایی شده در تصویر بازنگری را به عنوان B_t در فریم t بنامیم، در این صورت سلول مکعبی را به عنوان $u \in \mathbb{R}^3$ در B_t تعریف می‌کنیم. سلول مکعبی u دارای ابعاد m_x, m_y, m_z است که ابعاد m_x و m_y توسط ابعاد افقی و عمودی هر سلولها تعیین می‌شود و m_z نمایانگر تعداد فریم‌های متوالی در نظر گرفته شده است که در تمامی سلول‌های ویدئویی مکعبی مقداری ثابت دارد.

Archive of SID

[۲۹] و [۳۰]. این روش‌ها اغلب تعداد زیادی اوتونکه‌های غیرهمپوشان و یا کل فریم را به عنوان یک مجموعه فارغ از اینکه این تکه‌ها شامل اطلاعات حرکتی باشد یا خیر، مورد پردازش قرار می‌دهند. بنابراین این‌گونه متدها هزینه محاسباتی بسیار و همچنین خطای بالای شناسایی را متحمل می‌گردند. به منظور به دست آوردن دقت بالاتر و همچنین هزینه محاسباتی پایین می‌تواند از تکه سلول‌هایی که به صورت غیرفعال هستند صرف نظر کرد.

در روش پیشنهادی حاضر از شبکه کانولوشنی CNN زمان-مکانی که ارائه شده است به منظور شناسایی رفتار غیرمعمول در تصاویر ویدئویی استفاده شده است. در ادامه به بررسی مدل CNN و ساختار این شبکه پرداخته می‌شود. به منظور شناسایی CNN رفتار غیرمعمول در هر دو حوزه مکان و زمان، مدل CNN به صورت زمان-مکان ارائه شده است که ویژگی‌های حرکتی و بصری این رفتار ناهمجارت را استخراج می‌کند.

۱-۴-۳ استخراج ویژگی‌های زمان-مکان

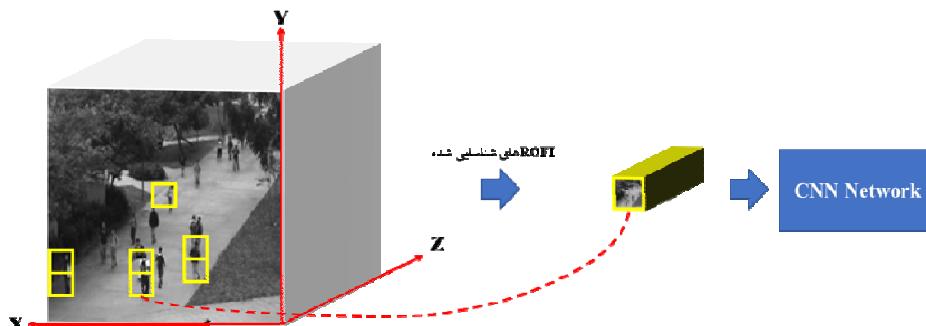
مدل CNN زمان-مکانی ارائه شده به جای استفاده از کل فریم از اطلاعات موجود در سلول‌های فعل استفاده می‌کند. بنابراین تنها از پیکسل‌هایی که اطلاعات غنی‌تری از رویداد وقوع یافته دارند بهره می‌برد. اندازه مکانی هر یک از تکه سلول‌ها قبل از وارد شدن به مدل CNN می‌باشد طبق ورودی مدل CNN تنظیم و اندازه آن‌ها تغییر پیدا کند. در ادامه مدل CNN زمان-مکانی با اعمال کانولوشن زمان-مکانی بر روی سلول‌ها، ویژگی‌های سطح بالای زمان-مکان که در شناسایی رفتار غیرمعمول مؤثر است، استخراج می‌شوند. این فرایند در شکل ۴ نشان داده شده است.

۳-۳ شناساگر ویژگی FAST و شناسایی سلول‌های فعال

از شناساگر ویژگی FAST به منظور اطمینان از نواحی که در بخش پیشین حاصل شد، بهره گرفته می‌شود. FAST یک متدهاینرا است که نقاط موردنظر را با مقایسه شدت یک پیکسل به نام P با همسایه‌هایش، شناسایی می‌کند. اگر تمام شدت پیکسل‌های همسایگان بیشتر یا کمتر از شدت پیکسل P باشد در این صورت P را به عنوان یک نقطه موردنظر در نظر می‌گیرد. این شناساگر باینری دارای مزایایی از جمله سرعت در شناسایی است که در این‌گونه مسائل می‌تواند پراهمیت باشد [۳۱، ۳۲]. حال شناساگر FAST را بر روی تصاویر ویدئویی اعمال می‌شود. برای هر نقطه با موقعیت مکانی (x_p, y_p) در بعد زمانی سلول فعالی شناسایی می‌شود. بعد از انتخاب نقطه موردنظر یک سلول ویدئویی $V \in R^3$ با سایز m_{x_p, y_p, t_p} به مرکزیت (x_p, y_p, t_p) ساخته می‌شود. سایز m_x, m_y توسط سایز سلول فعل موردنظر تعیین می‌گردد و m_t در تمام حجم‌های ویدئویی ثابت است. مجموعه سلول‌های به دست آمده توسط این روش را با مجموعه سلول‌هایی که در بخش پیشین حاصل آمد، در نظر گرفته می‌شود و سلول‌هایی که در هر دو مجموعه از نظر بعد مکانی با یکدیگر به میزان حداقل ۱۰٪ اشتراک دارند به عنوان یک سلول فعل و یا یک سلول شناخته خواهند شد.

۴-۳ استخراج ویژگی CNN

برای شناسایی و موقعیت یابی رفتار غیرمعمول که در نواحی محلی روی می‌دهد روش‌های متفاوتی در گذشته ارائه شده‌اند [۱۵] و



شکل ۴ شمایی از نواحی موردنظر که به عنوان ورودی به مدل CNN ارائه شده داده می‌شود.

درواقع شبکه فیلترهایی را که ویژگی‌های خاصی را در برخی موقعیت‌های مکانی ورودی شناسایی می‌کند فعال و آموزش می‌دهد. موضوع دیگری که در مدل‌های CNN از اهمیت خاصی برخوردار است لایه‌های ادغام و با pooling عمل می‌کنند. لایه pooling این‌ها به فرم یک نمونه بردار غیرخطی عمل می‌کنند. لایه pooling در واقع باعث کاهش سایز، همچنین باعث کاهش تعداد پارامترها، کاهش حجم محاسبات می‌شود که درنتیجه باعث می‌شود پدیده

۴-۴-۳ کانولوشن زمان-مکان

لایه کانولوشن قسمت اصلی مدل CNN است. پارامترهای لایه شامل مجموعه‌ای از فیلترها و یا هسته‌های باقابلیت آموزش هستند که در حوزه‌های کوچکی باهم در ارتباط‌اند اما در تمام عمق یک سلول مکعبی گسترش پیدا کرده است.

Archive of SID

می‌شود. با این ساختار نقشه‌های ویژگی در لایه کانولوشنی، به چندین فریم متوالی در لایه قبل متصل می‌شود. بنابراین اطلاعات حرکتی را می‌تواند استخراج کرد.

عملگر کانولوشن زمان-مکانی بین هسته سه بعدی W_n و سلول مکعبی زمان-مکان $a_{(i-1)n}$ را می‌توان به صورت زیر تعریف کرد.

$$[W_n * a_{(i-1)n}](x, y, z) \quad (3)$$

که

$$\sum_n \sum_{u=0}^{U_i-1} \sum_{v=0}^{V_i-1} \sum_{r=0}^{R_i-1} W_n^{uvr} * a_{(i-1)n}^{(x+u)(y+v)(t+r)} \quad (4)$$

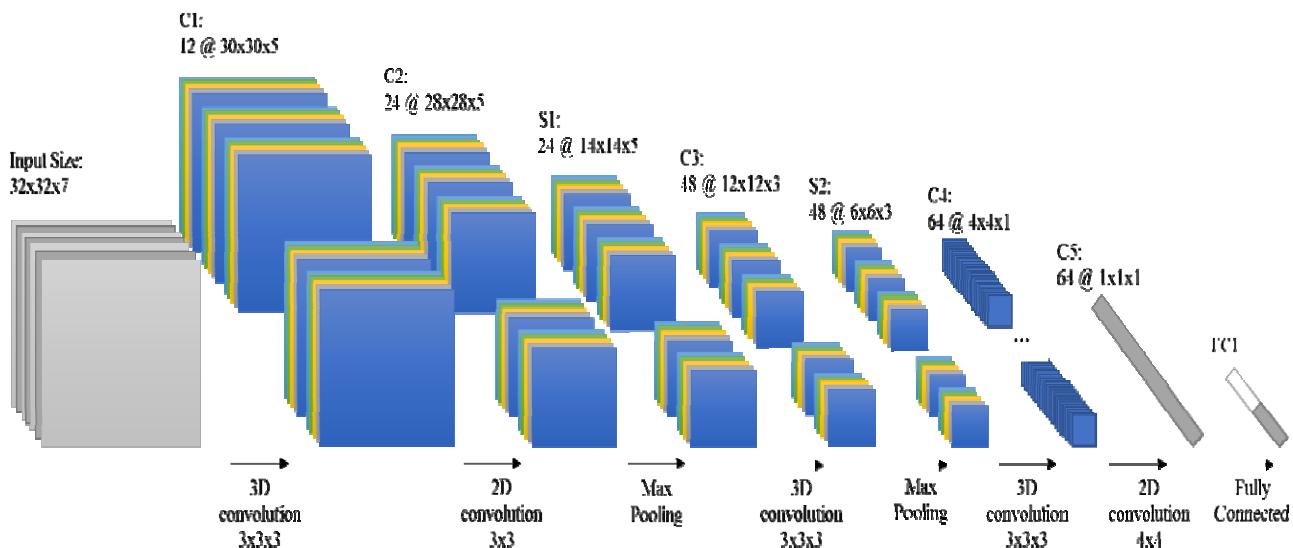
که U_i و V_i و R_i به ترتیب طول، عرض و طول زمانی هسته سه بعدی است و $x \times y \times t$ سایز زمان-مکان مربوط به سلول مکعبی $a_{(i-1)n}$ است.

برازش را کنترل کند. در ز امین نقشه ویژگی a_{ij} را می‌توان در ز امین لایه به صورت زیر تعریف کرد.

$$a_i = f(W_n * a_{(i-1)n} + b_{ij}) \quad (2)$$

که $f(\cdot)$ یک تابع فعال‌ساز است که به صورت ReLU گفته می‌شود. W هسته تعریف می‌شود، به آن لایه ReLU نمایه مجموعه نقشه ویژگی است که به نقشه ویژگی در $(i-1)$ امین لایه متصل شده است. * درواقع همان عملگر کانولوشن است و b_{ij} مقدار بایاس نقشه ویژگی کنونی است. W و b می‌بایست آموزش بینند که بتوانند ویژگی‌های بهتری را استخراج کنند.

برای استخراج اطلاعات زمان-مکانی، عملیات کانولوشن زمان-مکانی در لایه‌های کانولوشنی CNN اعمال می‌شود [۲۴]. کانولوشن زمان-مکانی توسط کانولوپ یک هسته سه بعدی روی سلول‌های مکعبی زمان-مکانی به دست می‌آید. درواقع این کانولوشن سه بعدی توسط یک هسته سه بعدی بر روی سلول‌های مکعبی که از انباره سازی فریم‌های متوالی به دست می‌آید اعمال



شکل ۵ ساختار مدل CNN برای شناسایی رفتار غیر نرمال. این ساختار شامل سه لایه کانولوشنی سه بعدی، دو لایه کانولوشنی دو بعدی و یک لایه کاملاً متصل است.

ساخته شده است. نواحی ROFI که در قسمت‌های قبلی به آن اشاره شد می‌بایست قبل از اینکه مورداستفاده مدل CNN قرار گیرد به ابعاد اشاره شده $7 \times 32 \times 32$ تغییر سایز پیدا کند.

در ابتدا کانولوشن زمان-مکان سه بعدی با هسته با سایز $3 \times 3 \times 3$ را بر روی داده‌های ورودی اعمال می‌کنیم (3×3 بعد مکانی و 3 بعد زمانی هسته است). می‌بایست به این نکته توجه داشت که یک هسته سه بعدی می‌تواند تنها یک نوع از ویژگی را از ROFI به دست آورد، بنابراین بهمنظور اینکه انواع ویژگی متفاوت را استخراج کنیم از 12 هسته سه بعدی متفاوت استفاده می‌کنیم و آن‌ها را در ورودی اعمال می‌کنیم که این امر باعث تولید 12 نقشه ویژگی در لایه C1 می‌شود. سایز هر کدام از نقشه‌های ویژگی $5 \times 30 \times 30$ است.

۳-۴-۳ ساختار مدل CNN زمان-مکان

با توجه به توضیحاتی که در قسمت قبل در مورد کانولوشن سه بعدی داده شد، ساختارهای بسیار متفاوتی از مدل CNN را می‌توان به وجود آورد. در ادامه به بررسی ساختار مدل CNN که برای شناسایی رفتار انسان توسعه داده‌ایم، پرداخته می‌شود.

در این شبکه تعداد هفت فریم متوالی که ابعاد هر کدام 32×32 است را به عنوان ورودی مدل CNN سه بعدی در نظر می‌گیریم. این مدل را می‌توان در شکل ۵ مشاهده کرد.

ساختار این شبکه زمان-مکانی از هشت لایه تشکیل شده است. سایز ورودی در ساختار این شبکه به صورت $7 \times 32 \times 32$ است، همان‌طور که اشاره شد، از 7 فریم متوالی به ابعاد 32×32

Archive of SID

شده است. منحنی ROC یک نمودار پراکندگی از حساسیت برای یک سیستم طبقه‌بندی کننده بایزی است که آستانه‌ی آن متغیر است.

بهمنظور تعیین فریم‌های غیر نرمال از دو سطح فریم^۲ و سطح پیکسل^۳ بهره گرفته شده است. این دو مقدار اندازه‌گیری به صورت زیر تعریف می‌شوند.

الف) اندازه‌گیری در سطح فریم: اگر یک پیکسل از هر فریم به عنوان غیرمعمول شناخته شود، آن فریم به عنوان یک فریم غیر نرمال محسوب می‌شود.

ب) اندازه‌گیری در سطح پیکسل: اگر حداقل ۴۰٪ از پیکسل‌های زمینه درست^۴ توسط پیکسل‌هایی که الگوریتم به عنوان غیرمعمول شناخته است، پوشش داده شود، فریم مربوطه را به عنوان فریم غیر نرمال محسوب می‌کنیم.

۱-۴ مجموعه داده^۵

بهمنظور آموزش مدل CNN و ارزیابی روش ارائه شده، از مجموعه داده UCSD که تقریباً در تمامی مقالات ارائه شده در این حوزه استفاده شده است، بهره گرفته‌ایم.

این مجموعه داده شامل دو زیرمجموعه با نام‌های Peds1 و Peds2 است که از دو فضای باز متفاوت گرفته شده است. هر دو زیرمجموعه توسط یک دوربین ایستا در ۱۰ فریم بر ثانیه به ترتیب با رزولوشن ۱۵۸×۲۳۴ و ۳۶۰×۲۴۰ ضبط شده است. فایل‌های زمینه درست در آن اجزا ارزیابی سطوح فریم و پیکسل را می‌دهد. در زیرمجموعه 1 Peds از ۳۶ ویدیو برای آزمایش و ۳۴ ویدیو برای آموزش و در زیرمجموعه 2 Peds از ۱۶ ویدیو برای آزمایش و ۱۲ ویدیو برای آموزش استفاده شده است. در شکل ۶ نمونه فریم‌هایی از این مجموعه داده به نمایش گذشته شده است.

در این مجموعه داده هر یک از ROFI‌ها به عنوان یک فضای مستطیلی به ابعاد ۱۵×۱۵ انتخاب شده‌اند که هم برای به دست آوردن یک موقعیت غیر نرمال به اندازه کافی کوچک و هم برای استخراج جزئیات تصویری به اندازه کافی بزرگ است. طول زمانی ROFIs به اندازه ۷ فریم است که درواقع یک توازن بین توانایی شناسایی رفتارهای ناهنجار و فضای حافظه موردنیاز برای این رفتارهاست.

درنهایت مکعب‌های ROFIs در اندازه ۱۵×۱۵×۷ انتخاب می‌شوند. سایز انتخابی برای این مکعب‌ها بستگی به نوع محیط و نحوه قرارگیری دوربین‌ها دارد که بعد از استخراج به اندازه ثابت موردنیاز در شبکه تغییر می‌کند و به شبکه کانولوشنی داده می‌شود.

بعد از آن یک هسته دو بعدی با سایز ۳×۳ برای لایه کانولوشنی بعدی اعمال می‌شود. سپس یک عملگر ادغام بر روی نتایج هر کانولوشن زمان-مکانی اعمال می‌شود. در لایه Pooling با نام S1 در هر یک از نقشه ویژگی در لایه C2 عملیات نمونه‌برداری با یک فاکتور مکانی ۲×۲ انجام می‌پذیرد که باعث کاهش ابعاد مکانی با همان تعداد نقشه ویژگی می‌شود و باعث به وجود آمدن نقشه‌های ویژگی می‌شود که در برابر اعوجاج‌های کوچک مستحکم‌تر است. بهمنظور تولید مجموعه دیگری از نقشه‌های ویژگی کانولوشن‌های زمان-مکانی دیگری را در لایه‌های عمیق‌تر بر روی نقشه‌های ویژگی اعمال می‌کنیم.

لایه کانولوشنی C3 با اعمال کانولوشن با یک هسته سه بعدی با سایز ۳×۳×۳ بر روی نقشه ویژگی‌های S1 به دست می‌آید. لایه C4 همانند لایه S1 عملیات مشابهی را انجام می‌دهد. لایه C2 کانولوشن ۳×۳×۳ را اعمال می‌کند که ۶۴ نقشه ویژگی را نتیجه می‌دهد. بعد از اعمال سه لایه کانولوشن زمان-مکانی، بعد زمانی نقشه‌های ویژگی به دست آمده به یک کاهش پیدا می‌کند. به دنبال لایه C4 یک لایه کاملاً متصل C5 قرار دارد. در لایه C5 یک کانولوشن دو بعدی بهمنظور به دست آوردن ویژگی‌های پیچیده سطح بالا اعمال می‌شود. سایز هسته کانولوشن در این لایه ۴×۴ است، بنابراین سایز نقشه ویژگی‌های خروجی این لایه به ۱×۱ کاهش پیدا می‌کند که هر کدام از آن‌ها به تمامی ۶۴ نقشه‌های ویژگی در لایه C4 متصل می‌شود. نقشه‌های ویژگی خروجی لایه C5 با یکدیگر ادغام می‌شود که به هر واحد بردار ویژگی کاملاً متصل^۶ گفته می‌شود. خروجی این لایه ۲ واحد است که مطابق با تعداد انواع رفتارها در ویدیو (Riftar نرمال و غیر نرمال) است و هریک از این واحدها احتمالی از یک Riftar را نمایش می‌دهد. این احتمالات توسط روابط زیر به دست می‌آید.

$$p_1 = 1 / (1 + \exp(u_2 - u_1)) \quad (5)$$

$$p_1 = 1 / (1 + \exp(u_1 - u_2)) \quad (6)$$

که $i=1,2$ ، u_i خروجی واحد i ام در لایه FC1 است و بهمنظور نرمال‌سازی احتمالات از روش رگرسیون منطقی بهره گرفته شده است. در ادامه این شبکه با توجه به احتمالات موردمحاسبه در مورد اینکه کدام توده از سلول‌های فعل شامل Riftar نامتعارف می‌شود تصمیم‌گیری می‌کند.

۴ ارزیابی و نتیجه‌گیری

روش ارائه شده در این مقاله با روش‌های متداول پیشین بر روی مجموعه داده UCSD [۳۲] که در ادامه در مورد آن توضیحاتی داده خواهد شد، مورد مقایسه قرار گرفته است. این مقایسه با Receiver Operation Characteristic استفاده از معیارهای منحنی ROC (ROC) و نرخ Equal Error Rate (EER) ارزیابی و نتیجه‌گیری

² Sensitivity

³ Frame level

⁴ Pixel level

⁵ Ground Truth

⁶ Dataset

¹ Fully-Connected

Archives of SID



شکل ۶ نمونه تصاویر مربوط به مجموعه داده‌ی مورد استفاده جهت ارزیابی روش ارائه شده. ردیف اول و ردیف دوم به ترتیب نمونه تصاویر مربوط به Peds1 و Peds2 از مجموعه داده UCSD است.

را چنانچه شامل حداقل یک ناحیه غیرعادی باشد، فریم غیرنرمال می‌داند و به جایگاه وقوع رفتار غیرنرمال هیچ حساسیتی ندارد.

در سویی دیگر معیار سطح پیکسل معیاری است که موقعیت زمان-مکانی در فریم را مشخص می‌کند. همان‌طور که اشاره شد اگر حداقل ۴۰٪ از پیکسل‌های زمینه درست توسط پیکسل‌هایی که الگوریتم به عنوان رفتار غیرمعمول شناخته است، پوشش داده شود آن فریم را به عنوان فریم غیرمعمول شناسایی می‌کند.

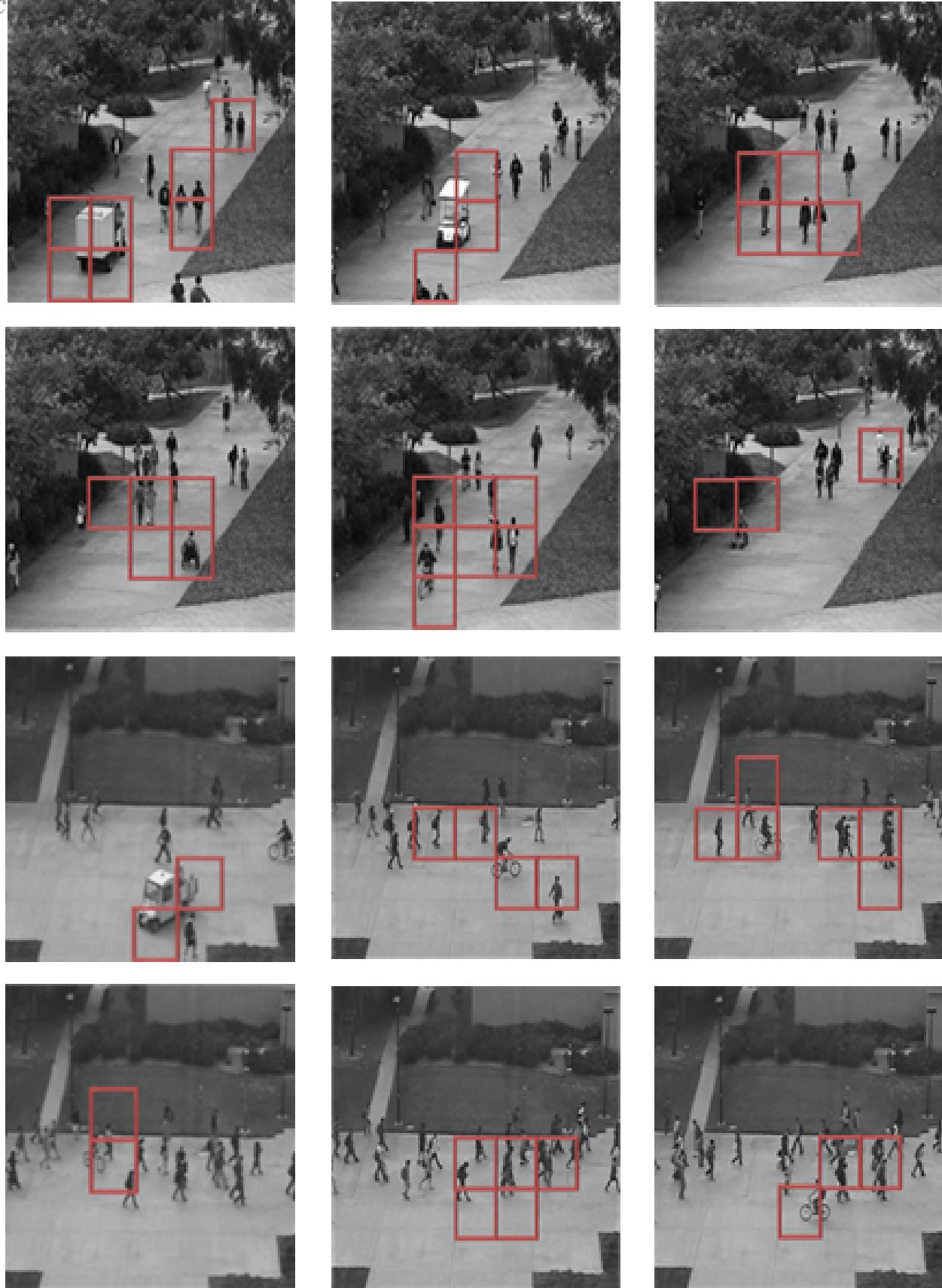
در ادامه با محاسبه نرخ‌های TP و FP می‌توان معیار ROC برای ارزیابی کارایی الگوریتم را به دست آورد.

۴-۴ نحوه ارزیابی

روش ارائه شده در این مقاله با برخی از روش‌های متداول که تاکنون ارائه شده مورد ارزیابی قرارگرفته است [۱۳, ۸, ۲, ۱۵, ۳۱, ۳۴-۳۸]. به منظور ارزیابی روش ارائه شده در مجموعه داده UCSD دو معیار برای ارزیابی میزان دقیقت شناسایی رفتارهای غیرمعمول، معیار سطح پیکسل و سطح فریم، مورد استفاده قرارگرفته است.

معیار سطح فریم تنها بر روی تغییراتی متمرکز است که پیش‌بینی می‌کند کدام فریم شامل رفتار غیرمعمول است بدون اینکه محل وقوع آن را مشخص کند. معیار سطح فریم، یک فریم

Architectural SID



شکل ۷ نمونه‌هایی از شناسایی نواحی وقوع رفتار غیرعادی توسط روش ارائه شده.

Archive of SID

جدول ۱ نرخ ERR در زیرمجموعه Peds1 از مجموعه داده UCSD

نرخ ERR در سطح پیکسل	نرخ ERR در سطح فریم	نام نویسنده مقالات
۳۸/۸	۱۹/۹	[۳۱] Cheng et al.
۵۱/۲	۲۳	[۱۳] Cong et al.
۳۸/۹	۲۲/۶	[۲] Adam
۳۹/۶	۱۹/۶	[۹] Kim
۲۷	۲۱/۱	[۳۵] Kaltsa
-	۲۷/۹	[۲۶] Conv-AE
۲۸	۲۱/۳	روش ارائه شده

جدول ۲ نرخ ERR در زیرمجموعه Peds2 از مجموعه UCSD

نرخ ERR در سطح پیکسل	نرخ ERR در سطح فریم	نام نویسنده مقالات
۴۳/۸	۲۲/۴	[۲] Adam
۳۱/۱	۲۲/۴	[۹] Kim
۲۶/۹	۲۵/۱	[۳۵] Kaltsa
-	۲۱/۷	[۲۶] Conv-AE
۲۷/۶	۱۹/۲	روش ارائه شده

نتایجی که از سایر روش‌ها آورده شده، از مراجع مربوطه که این روش‌ها در آن‌ها معرفی شده‌اند، اقتباس شده است. در شکل ۷ نتایج شناسایی رفتارهای غیرعادی آورده شده است. منحنی‌های مربوط به ROC در مجموعه داده UCSD در شکل ۸ گزارش داده شده است. با توجه به تصویر زیر می‌توان مشاهده کرد که متدهای ارائه شده کارایی قابل رقابتی را در مقایسه با روش‌های پیشین ارائه داده است.

۳-۴ نتایج

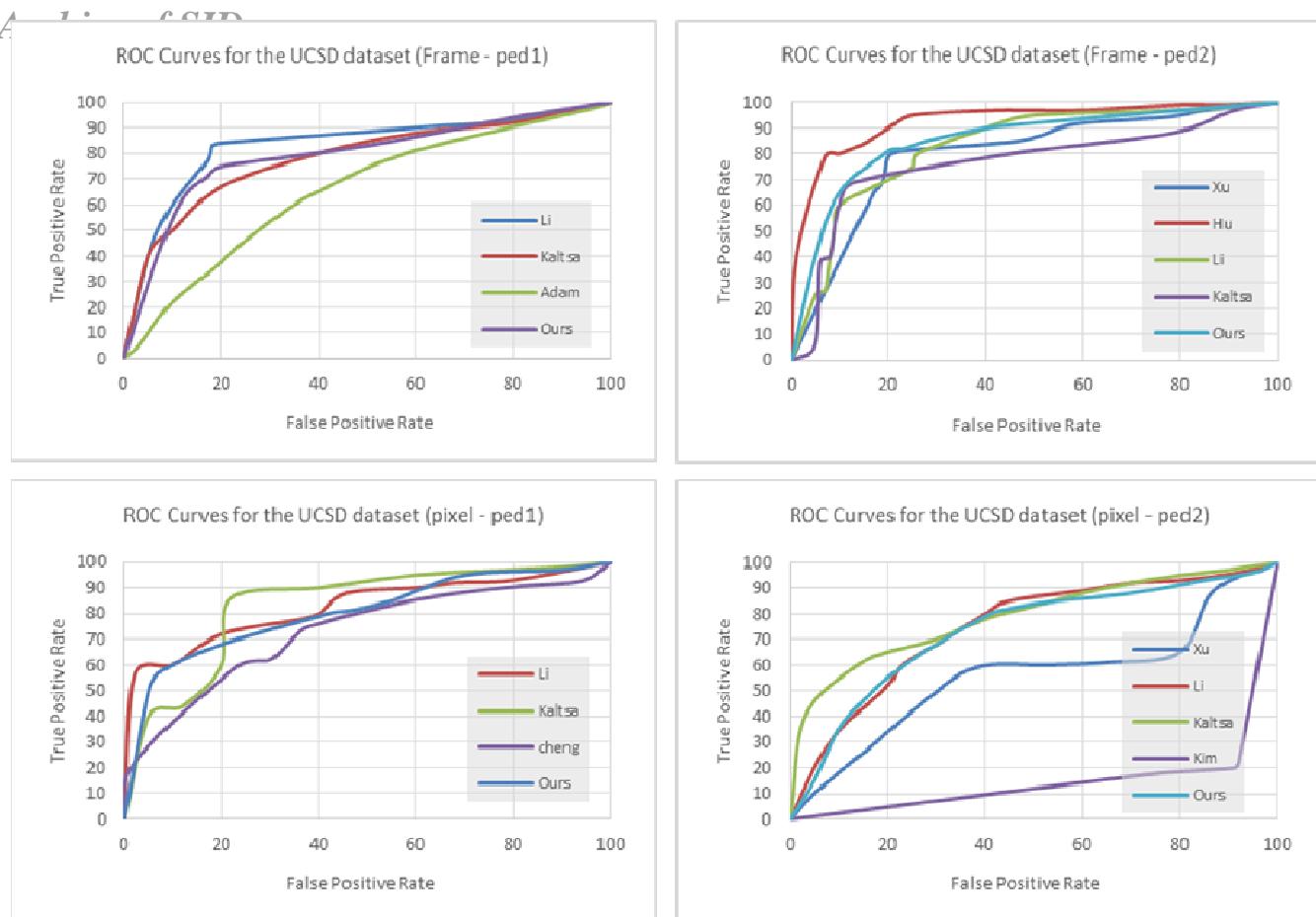
همان‌طور که در بخش ۱-۴ بیان شد، هر یک از مکعب‌های ROFIs در سایزی معادل $15 \times 15 \times 7$ انتخاب می‌شوند. با توجه به اینکه هیچ‌یک از کلیپ‌های آموزشی مجموعه‌های Ped1 و Ped2 شامل ناهنجاری نیست، به صورت تصادفی نصف کلیپ‌های آزمایشی مربوط به Ped1 و Ped2 برای استفاده در مدل آموزشی و مابقی کلیپ‌ها برای آزمایش نمونه‌ها مورد استفاده قرار می‌گیرند.

هر یک از مجموعه‌های Ped1 و Ped2 به طور جداگانه آموزش داده می‌شوند. تعداد ۱۴۰۲۴۸ نمونه نرمال و ۳۵۲۱۵ نمونه غیر نرمال از مجموعه Ped1 و تعداد ۶۳۵۷۹ نمونه نرمال و ۲۰۶۳۸ نمونه غیر نرمال از مجموعه Ped2 استخراج شده است.

با توجه به اینکه تعداد نمونه‌های غیر نرمال بسیار کمتر از نمونه‌های نرمال است، این امکان وجود دارد که مشکل عدم تعادل در کلاس به وجود آید. برای حل این مشکل با نمونه‌برداری دوباره، سعی شد تعداد نمونه‌های نرمال و غیر نرمال در یک تعادل قرار گیرند.

روش ارائه شده در این مقاله توسط یک رابط نرم‌افزاری تحت عنوان [۳۹] Keras که از شبکه‌های عصبی و کانولوشنی با بهره‌گیری از کتابخانه نرم‌افزارهای [۴۱، ۴۰] Theano و [۴۲] Tensor Flow پشتیبانی می‌کند، پیاده‌سازی شده است. این رابط نرم‌افزاری به زبان Python نوشته شده است و قابلیت اجرا بر روی CPU و GPU را دارد. روش ارائه شده بر روی یک کامپیوتر با ۸G RAM ۲.۴ GHz Intel(R) Core(TM) i5 CPU و ۲G NVIDIA GT620 کارت گرافیکی مورد استفاده در این روش می‌باشد.

نتایج مربوط به روش ارائه شده و همچنین تحقیقات پیشین، در جدول ۱ و جدول ۲ از مجموعه داده UCSD در Peds1 و Peds2 نمایش داده شده است.



شکل ۸ منحنیهای ROC در مجموعه داده UCSD در سطوح پیکسل و فریم

به عنوان ورودی به مدل CNN داده می‌شود. مدل زمان-مکانی CNN بهمنظور ساخت ویژگی‌های مقاوم در دو بعد زمان و مکان با اعمال کانولوشن سه‌بعدی طراحی شده است، بنابراین ویژگی‌های ظاهری به خوبی اطلاعات حرکتی موجود در فریم‌های متوالی استخراج می‌شوند.

برای به دست آوردن مکعب‌های غیر همپوشان که حاوی اطلاعات حرکتی مؤثر هستند ابتدا نقشه شارنوری مربوطه را به دست می‌آوریم. این نقشه حاوی اطلاعات حرکتی پیکسل‌های موجود در تصاویر ویدئویی است و بهمنظور استخراج نواحی که اطلاعات حرکتی در آن پررنگ‌تر است ابتدا این نقشه را به فضای تصویر HSV تبدیل می‌کنیم و با استفاده از متند GMM توسط دو گوسی آستانه مربوط به باینری کردن تصویر را به دست می‌آوریم. توسط این آستانه تصویر مربوطه را به یک تصویر باینری بدل می‌کنیم. حال می‌توان نواحی که اطلاعات حرکتی مؤثری را دارا هستند، مشاهده نمود. این نواحی به دست آمده نیز توسط مکعب‌های غیر همپوشان تقسیم می‌شوند.

با توجه به این‌که نواحی به دست آمده می‌توانند تحت تأثیر عواملی دارای خطای باشند، از روش شناساگر FAST کمک گرفته شده است تا میزان خطای شناسایی این نواحی به طرز محسوسی کاهاش یابد.

به خصوص نتایج حاصله در سطح فریم در روش ارائه شده بسیار شبیه به نتایج بهترین متدهایی است که در این زمینه معرفی شده‌اند. در سطح پیکسل نیز روش ارائه شده کارایی قابل رقابتی را در مقایسه با دیگر روش‌ها دارد. در حالت کلی می‌توان گفت که روش ارائه شده، کارایی بسیار رقابتی در مجموعه داده UCSD با نتایج متدهای دیگر را دارد.

۵ جمع‌بندی

در این مقاله روشی برای شناسایی رفتارهای ناهمچار در تصاویر ویدئویی بخصوص در صفحه‌های پیچیده و شلوغ ارائه شده است. این روش ارائه شده مجموعه ویژگی‌هایی را بر مبنای اطلاعات شارنوری، نقاط ویژگی الگوریتم FAST و شبکه عمیق CNN استخراج می‌کند.

این روش از یک مدل CNN زمان-مکانی نوین برای شناسایی و موقعیت‌یابی رفتارهای ناهمچار در صفحه‌های مختلف ویدئویی بهره برده است. مدل زمان-مکانی CNN بهمنظور تولید ویژگی‌ها در دو بعد زمانی و مکانی توسط کانولوشن‌های زمان-مکانی طراحی شده است. در این روش تصاویر ویدئویی به مکعب‌های غیر همپوشان تقسیم می‌شود و هر یک از این مکعب‌های زمان-مکانی که دارای اطلاعات حرکتی غنی‌تری باشد

Archive of SID

- IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07. 2007.
- [6] Li, J., S. Gong, and T. Xiang. Global Behaviour Inference using Probabilistic Latent Semantic Analysis. in BMVC. 2008.
 - [7] Jiang, F., et al., Anomalous video event detection using spatiotemporal context. Computer Vision and Image Understanding, 2011. 115(3): p. 323–333.
 - [8] Wu, S., B.E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.
 - [9] Kim, J. and K. Grauman. Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). . 2009.
 - [10] Zhong, H., J. Shi, and M. Visontai. Detecting unusual activity in video. in IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR. . 2004.
 - [11] Hamid, R., et al. Detection and explanation of anomalous activities: Representing activities as bags of event n-grams. in IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR. . 2005.
 - [12] Lu, C., J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. in Proceedings of the IEEE International Conference on Computer Vision. 2013.
 - [13] Cong, Y., J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011.
 - [14] Cong, Y., J. Yuan, and Y. Tang, Video anomaly search in crowded scenes via spatio-temporal motion context. IEEE Transactions on Information Forensics and Security, 2013. 8(10): p. 1590–1599.
 - [15] Li, W., V. Mahadevan, and N. Vasconcelos, Anomaly detection and localization in crowded scenes. IEEE transactions on pattern analysis and machine intelligence, 2014. 36(1): p. 18–32.
 - [16] Mehran, R., A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. in IEEE Conference on Computer Vision and Pattern Recognition. CVPR. . 2009.
 - [17] Roshtkhari, M.J. and M.D. Levine, An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. Computer vision and image understanding, 2013. 117(10): p. 1436–1452.
 - [18] LeCun, Y., et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998. 86(11): p. 2278–2324.

یک شناساگر باینری است که نقاط موردنظر را با مقایسه شدت یک پیکسل با همسایه‌هایش شناسایی می‌کند که اگر شدت تمامی پیکسل‌های همسایه بیشتر یا کمتر از پیکسل موردنظر باشد، آن را به عنوان یک نقطه موردنظر در نظر می‌گیرد. شناساگر همچنین در شناسایی نقاط از سرعت عمل بالایی برخوردار است.

شناساگر FAST بر روی تصاویر ورودی اعمال می‌شود. در هر نقطه‌ای که به دست می‌آید یک سلول مکعبی به مرکزیت مختصات نقطه به دست آمده در نظر گرفته می‌شود. حال می‌توان با اشتراک سلول‌های مکعبی که از باینری کردن نقشه شارنوری به FAST دست آمد و همچنین سلول‌هایی که توسط نقاط شناساگر حاصل شد، مجموعه‌ای تحت عنوان سلول‌های فعال را به دست آورد که در این مقاله به این سلول‌ها ROFIs گفته می‌شود. این سلول‌های ROFIs درواقع همان سلول‌هایی هستند که می‌بایست برای شناسایی رفتار نا亨جار مورد بررسی قرار گیرند. با توجه به آنچه ذکر شد، هر یک از این سلول‌ها به عنوان ورودی مدل CNN برای شناسایی رفتار نا亨جار مورد استفاده قرار می‌گیرند. مدل CNN پیشنهادی با توجه به محاسبه احتمال متعارف و یا نامتعارف بودن این سلول‌ها تصمیم می‌گیرد که کدام سلول حاوی اطلاعات نا亨جار است و آن را به عنوان یک سلول نا亨جار معرفی می‌کند.

سپاسگزاری

در اینجا لازم است از تمامی اعضای پژوهشگاه پردازش سیگنال و تصویر دانشگاه صنعتی شاهroud و همچنین پژوهشگاه دانش‌های بنیادی به منظور حمایت از این پژوهه تشکر و قدردانی به عمل آورده شود.

مراجع

- [1] Cong, Y., J. Yuan, and J. Liu, Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition, 2013. 46(7): p. 1851–1864.
- [2] Adam, A., et al., Robust real-time unusual event detection using multiple fixed-location monitors. IEEE transactions on pattern analysis and machine intelligence, 2008. 30(3): p. 555–560.
- [3] Basharat, A., A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. in IEEE Conference on Computer Vision and Pattern Recognition. CVPR. . 2008.
- [4] Boiman, O. and M. Irani, Detecting irregularities in images and in video. International journal of computer vision, 2007. 74(1): p. 17–31.
- [5] Wang, X., X. Ma, and E. Grimson. Unsupervised activity perception by hierarchical bayesian models. in www.SID.ir

Archive of SID

- [33] UCSD Dataset. Available from: www.svcl.ucsd.edu/projects/anomaly/dataset.html.
- [34] Hu, Y., Y. Zhang, and L. Davis. Unsupervised abnormal crowd activity detection using semiparametric scan statistic. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013.
- [35] Kaltsas, V., et al., Swarm intelligence for detecting interesting events in crowded environments. *IEEE transactions on image processing*, 2015. 24(7): p. 2153–2166.
- [36] Xu, D., et al., Video anomaly detection based on a hierarchical activity discovery within spatio-temporal contexts. *Neurocomputing*, 2014. 143: p. 144–152.
- [37] Zhu, X., et al., Sparse representation for robust abnormality detection in crowded scenes. *Pattern Recognition*, 2014. 47(5): p. 1791–1799.
- [38] Saligrama, V. and Z. Chen. Video anomaly detection based on local statistical aggregates. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012..
- [39] Chollet, F. keras,. 2015; Available from: <https://github.com/fchollet/keras>.
- [40] James, B., et al. Theano: a CPU and GPU math expression compiler. in *Proceedings of the Python for Scientific Computing Conference (SciPy)*.
- [41] Al-Rfou, R., et al., Theano: A Python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, 2016.
- [42] Abadi, M., et al. TensorFlow: A system for large-scale machine learning. in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. Savannah, Georgia, USA. 2016.
- [19] Goodfellow, I.J., et al., Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*, 2013.
- [20] Krizhevsky, A., I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in *Advances in neural information processing systems*. 2012.
- [21] Sermanet, P., et al., Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.
- [22] Shen, W., et al. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
- [23] Taigman, Y., et al. Deepface: Closing the gap to human-level performance in face verification. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
- [24] Ji, S., et al., 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2013. 35(1): p. 221–231.
- [25] Sabokrou, M., et al., Deep-cascade: cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. *IEEE Transactions on Image Processing*, 2017. 26(4): p. 1992–2004.
- [26] Hasan, M., et al. Learning temporal regularity in video sequences. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [27] Xu, D., et al., Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision and Image Understanding*, 2017. 156: p. 117–127.
- [28] Wang, K., et al. 3D human activity recognition with reconfigurable convolutional neural networks. in *Proceedings of the 22nd ACM international conference on Multimedia*. 2014. ACM.
- [29] Feng, Y., Y. Yuan, and X. Lu, Learning deep event models for crowd anomaly detection. *Neurocomputing*, 2017. 219: p. 548–556.
- [30] Maturana, D. and S. Scherer. 3d convolutional neural networks for landing zone detection from lidar. in *IEEE International Conference on Robotics and Automation (ICRA)*. 2015. IEEE.
- [31] Cheng, K.-W., Y.-T. Chen, and W.-H. Fang, Gaussian Process Regression-Based Video Anomaly Detection and Localization With Hierarchical Feature Representation. *IEEE Transactions on Image Processing*, 2015. 24(12): p. 5288–5301.
- [32] Laptev, I. and T. Lindeberg. Space-time interest points. in *IEEE conference proceedings 9th International Conference on Computer Vision*, Nice, France. 2003..

Archive of SID



بهنام سبزعلیان مدرک کارشناسی خود را در رشته نرم افزار از دانشگاه سمنان اخذ نموده است و مقطع کارشناسی ارشد را در رشته مهندسی رباتیک در دانشگاه صنعتی شاهروд به پایان رسانده است. زمینه های مورد علاقه ایشان پردازش تصویر و ویدیو، بینایی ماشین و یادگیری عمیق است.



حسین مریم مدرک کارشناسی خود را در رشته برق الکترونیک از دانشگاه فردوسی مشهد اخذ نمودند. ایشان مقطع کارشناسی ارشد را در دانشگاه شیراز در رشته برق گرایش مخابرات به پایان رساندند. مدرک دکتری تخصصی خود را از کشور انگلستان دانشگاه Surrey مرکز CVSSP و در زمینه پردازش گفتار اخذ نمودند. نامبرده هم اکنون به عنوان هیات علمی با مرتبه دانشیاری در دانشکده مهندسی برق و رباتیک دانشگاه صنعتی شاهرود مشغول فعالیت می باشند. زمینه های مورد علاقه ایشان پردازش سیگنال ها، زمینه های مختلف پردازش گفتار و پردازش سیگنال های حیاتی می باشد.



علیرضا احمدی فرد مدرک کارشناسی را از دانشگاه صنعتی اصفهان در مهندسی الکترونیک و کارشناسی ارشد را از دانشگاه صنعتی امیرکبیر در مهندسی مخابرات اخذ نمودند. ایشان مدرک دکتری تخصصی را در زمینه پردازش تصویر و ماشین بینایی از مرکز دانشگاه CVSSP Surrey در سال ۲۰۰۲ اخذ نمودند. زمینه های تحقیقاتی مورد علاقه ایشان پردازش سیگنال، پردازش تصاویر و شناسایی الگو می باشد.