Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 8, Number 3, Fall 2019, pp. 40-51
Received: 11/05/2018, Revised: 04/27/2019, Accepted: 07/14/2019

Dominant and rare events detection and localization in video using Generative Adversarial Network

Mohammad khalooei[1], Mohammad Fakhredanesh[2*] and Mohammad Sabokrou[3]
1- Faculty of Electrical and Computer, Malek Ashtar University of Technology, Tehran, Iran.
2*- Faculty of Electrical and Computer, Malek Ashtar University of Technology, Tehran, Iran.
3- School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran.
[1]khalooei@chmail.ir, [2]m-fakhredanesh@mut.ac.ir, and [3]sabokro@ipm.ir

Corresponding author's address: Mohammad Fakhredanesh, Faculty of Electrical and Computer, Malek Ashtar University of Technology, Tehran, Iran.

Abstract- Dominant and rare events detection is one of the most important subjects of image and video analysis field. Due to inaccessibility to all rare events, detecting of them is a challenging task. Today, deep networks are the best tool for video modeling but due to inaccessibility to tagged data of rare data, usual learning of a deep convolutional network is not possible. Due to the success of generative adversarial networks, in this paper an end-to-end deep network based on generative adversarial networks is presented for detecting rare events. This network is competitively trained only by dominant events. To evaluate performance of proposed method, two standard datasets: UCSDped1 and UCSDped2 are utilized. The proposed method can detect rare event with 0.2 and 0.17 equal error rate with the processing speed of 300 frames per second on the mentioned data respectively. In addition to end-to-end structure of the network and its simple train and test phase, this result is comparable to advanced methods results.

Keywords- Detecting rare event, Generative adversarial network, anomaly detection, anomaly localization.

مجله علمی پژوهشی رایانش نرم و فن‌آوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۸، شماره ۳، پاییز ۱۳۹۸، صفحه ۵۱-۴۰
دریافت: ۱۳۹۷/۰۸/۱۴، بازنگری: ۱۳۹۸/۰۲/۰۷، پذیرش: ۱۳۹۸/۰۴/۲۳

تشخیص و مکان‌یابی رویدادهای رایج و نادر در ویدیو با به‌کارگیری شبکه تخاصمی مولد

محمد خالوئی[۱]، محمد فخردانش*[۲]، محمد سبک‌رو[۳]
۱- مجتمع دانشگاهی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران، ایران.
*۲- مجتمع دانشگاهی برق و کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران، ایران.
۳- پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران، ایران.
[1]khalooei@chmail.ir, [2]fakhredanesh@mut.ac.ir, [3]sabokro@ipm.ir

* نشانی نویسنده مسئول: محمد فخردانش، تهران، خیابان شهید شعبانلو، دانشگاه صنعتی مالک‌اشتر، مجتمع دانشگاهی برق و کامپیوتر.

رایج و نادر در ویدیو یکی از مسائل مهم در حوزه تحلیل تصویر و ویدیو است. با توجه به عدم شناخت و در دسترس نبودن رویدادهای نادر، تشخیص آنها یک چالش محسوب می‌شود. امروزه، شبکه‌های ژرف یکی از بهترین ابزارها برای مدل‌سازی ویدیو هستند اما در این مساله با توجه به عدم وجود داده‌های برچسب‌دار از کلاس رویدادهای نادر، آموزش یک شبکه کانولوشنال ژرف به صورت معمول امکان پذیر نیست. با توجه به موفقیت شبکه های عصبی ژرف تخاصمی مولد، در این پژوهش یک شبکه ژرف یکپارچه (انتها به انتها) با الهام از شبکه های ژرف تخاصمی مولد برای تشخیص رویدادهای نادر ارائه شده است. این شبکه فقط با رویدادهای رایج و به صورت تخاصمی آموزش داده شده است. برای نمایش میزان عملکرد معماری پیشنهادی، از مجموعه دادگان استاندارد UCSDped1 و UCSDped2 به‌کارگیری شده است. معماری پیشنهادی روی مجموعه دادگان مذکور دارای نرخ‌خطای‌برابر به ترتیب ۲۰٪ و ۱۷٪ با سرعت پردازش ۳۰۰ فریم بر ثانیه بوده است. این نتیجه علاوه بر ساختار یکپارچه شبکه و سادگی مرحله آموزش و آزمون آن، قابل مقایسه با نتایج روش‌های پیشرفته است.

واژه‌های کلیدی: تشخیص رویداد نادر ، رویداد نادر، رویداد رایج، شبکه تخاصمی مولد، یادگیری ژرف، مکان‌یابی.

۱- مقدمه

تامین امنیت یکی از نیازهای اولیه جوامع است. امروزه، تامین امنیت از طریق نظارت تصویری، جهت کنترل اماکن مختلف رایج شده است. همچنین هزینه و تعداد نیروی انسانی برای کنترل و نظارت محتوای ویدیوهای دریافتی، بالا است. از سویی دیگر در برخی از اماکن، بدلیل شرایط نامناسب محیطی (نظیر مراکز حساس و پرخطر، مکان‌های صعب‌العبور)، امکان حضور نیروی انسانی وجود ندارد. ازاین‌رو نیاز است تا با بهره‌گیری از علم روز، همه اقدامات لازم کنترل و ارزیابی در این سامانه‌های نظارتی به‌صورت خودکار صورت پذیرد. یکی از مواردی که در سامانه‌های نظارتی دنبال می‌شود، تشخیص[۱] رویداد[۲]های نادر[۳] در میان انبوهی از رویدادهای رایج[۴] در مکان‌های مختلف است.

روش‌های مختلفی در سامانه‌های نظارت ویدیویی، در راستای تشخیص رویدادهای رایج و نادر مورد استفاده قرار گرفته است. در این پژوهش سعی شده تا با به‌کارگیری شبکه تخاصمی مولد، ساختاری یکپارچه مبتنی بر یادگیری ژرف[۵] برای تشخیص و

مکان‌یابی[۶] رویدادهای رایج و نادر ارائه گردد. رویدادهای نادر غالبا منجر به بروز حوادث و خسارت در مکان‌های حساس یا عمومی می‌شوند و این سامانه‌ها با تشخیص به موقع جهت ارائه عکس‌العمل مناسب هشدار می‌دهند [۱].

هدف پژوهش حاضر این است که به جای استخدام و نمایش رویدادهای رایج به نگهبانان، سامانه‌ای هوشمند آموزش داده شود و فرآیندهای تشخیص به صورت خودکار اعلان شود. الگوریتم پیشنهادی از شبکه عصبی تخاصمی مولد الهام و پیاده‌سازی شده است. درواقع این الگوریتم همانند یک نگهبان که همه حالات غیررایج را در ذهن خود تداعی می‌کند، بدون نیاز به نمونه‌ی رویدادهای نادر، آن‌ها را تشخیص می‌دهد. برای این منظور، معماری مبتنی بر شبکه‌های کانولوشنال[۷] [۱] و شبکه تخاصمی مولد[۸] [۲] برای تشخیص و مکان‌یابی رویدادهای رایج و نادر در ویدیو ارائه می‌گردد.

این پژوهش در شش بخش سازماندهی شده است. در بخش دوم به مرور پژوهشهای مرتبط و در بخش سوم به تعریف شبکه‌های تخاصمی مولد به عنوان زیرساخت این پژوهش پرداخته خواهد شد. معماری پیشنهادی و آزمایش‌های تجربی نیز به ترتیب در بخش‌های چهارم و پنجم ارائه شده و در بخش ششم نیز نتیجه‌گیری و پیشنهادهای آتی ارائه شده است.

۲- پژوهش‌های مرتبط

در پژوهش‌های ژوهو[۹] و همکارانش [۳] و موسوی و همکارانش [۴]، معانی و دیدگاه‌های مختلفی درمورد رویدادهای نادر مطرح شده است. فضای مسئله تشخیص رویدادهای رایج و نادر، با عدم قطعیت زیادی در شرایط واقعی روبرو می‌شود. به علت این‌که همیشه همه تنوعات موجود در رویدادهای نادر در حال رخ‌دادن نیستند، لذا نمی‌توان مجموعه داده آموزشی مناسبی که حاوی همه نمونه‌های رویداد نادر را در اختیار داشت. در نتیجه حل مسئله با به‌کارگیری رویکردهای بانظارت، موفق نخواهد بود.

یکی از چالش‌های مهم در تشخیص رویدادهای نادر، فقدان وجود تعریف مشخص از رویدادهای نادر است. تعریفی که اکثر پژوهش‌ها به آن اشاره کرده‌اند شامل رویدادهایی است، که نرخ احتمال رخدادشان پایین است. رویدادهایی که در هر محیط به صورت معمول رخ می دهند به عنوان رویداد رایج و بقیه موارد رویداد نادر محسوب می‌شوند [۱۱-۳]. یکی از چالش‌های مطرح دیگر در تشخیص رویدادهای نادر، مکان به وقوع پیوستن رویداد نادر است [۴، ۱۲].

محققان از دیدگاه‌های گوناگون به مسئله تشخیص رویدادهای رایج و نادر در ویدیو پرداخته‌اند [۸، ۹، ۱۷-۱۳]. هر شرایطی امکان دارد متناسب با محیط فیلم‌برداری شده، نادر یا رایج باشد. برای مثال در پیاده‌رو، عبور و مرور مردم امری طبیعی و رایج است اما تردد وسایل نقلیه، رویدادی نادر در آن محیط به شمار می‌رود. مسئله تشخیص رویداد نادر، یک مسئله چندمرحله‌ای است و در اغلب موارد، رسیدن به دقت مطلوب نیازمند پیش‌پردازش‌هایی نظیر جداکردن پس‌زمینه، رده‌بندی اشیا و بخش‌بندی[۱۰] محیط است [۱۱]. هر یک از این عملیات نیز به تنهایی یک مسئله و چالش محسوب می‌شوند.

پارامترهای لازم متناسب با مسئله مورد نظر و همچنین محل نصب دوربین‌ها در محیط‌های مختلف از جمله راهرو، پارکینگ و اتاق متفاوت هستند. ویدیوهایی که توسط دوربین‌های نظارتی ایستان تهیه شده‌اند، جزو ویدیوهای سوم شخص محسوب می‌شوند. دسته دیگری با عنوان ویدیو اول‌شخص وجود دارد که در آنها دوربین بر روی لباس، بدن یا کلاه فرد نصب می شود. این دسته ویدیو در این پژوهش بررسی نمی‌شود [۴].

دیدگاه‌ها و دسته‌بندی‌های مختلفی در بررسی و تحلیل ورودی مسئله تشخیص رویدادهای رایج و نادر وجود دارد. دسته‌بندی ویدیوها به ویدیوهای محیط‌های خلوت و شلوغ[۱۱] از دسته‌بندی‌های معمول این حوزه است. همچنین موارد دیگری از قبیل ویدیوهایی که برای یک محیط، اما از چند نقطه، فیلم‌برداری شده‌اند [۱۰] و ویدیوهایی که مربوط به محیط‌های بسته یا باز هستند [۱۵] نیز از دسته‌بندی‌های مختلف دیگر به شمار می‌روند. پژوهش حاضر مربوط به تشخیص رویدادهای نادر در دسته اول (محیط‌های خلوت و شلوغ) قرار می‌گیرد.

به منظور شناسایی رویدادهای نادر، دقت و عملکرد الگوریتم‌ها، در محیط‌های خلوت، بهتر از محیط‌های شلوغ است. اکثر روش‌های ساده و سنتی با دقت‌های نسبتا مناسبی تشخیص‌های مطلوب را ارائه می‌کنند [۱۸]. از مهم‌ترین چالش‌های تشخیص رویدادهای نادر در محیط‌های شلوغ، مشکل همپوشانی افراد و اشیای محیط است. هزینه محاسباتی روش‌های دستی و یکنواخت نبودن میزان شلوغی ویدیو، از مشکلات دیگر محیط‌های شلوغ محسوب می‌شود. فعالیت‌ها و پژوهش‌هایی همانند تشخیص حرکت و عبور قایق در محوطه‌های ممنوعه، پارک، پیاده‌رو، قطارشهری از نمونه‌های بارز پژوهش‌های انجام شده در محیط‌های خلوت هستند [۵، ۶، ۸، ۱۴، ۱۹، ۲۰].

برای تشخیص رویداد رایج و نادر، ویژگی مکانی، زمانی و مکانی-زمانی قابل تعریف است. ویژگی مکانی، بازنمایی شکل و بافت

اشیا موجود در یک فریم از ویدیو را شامل می‌شود. در این نوع ویژگی، رابطه بین فریم‌های متوالی یک ویدیو در نظر گرفته نمی‌شود و تمرکز اصلی روی ویژگی نواحی کنار هم در داده ورودی است [۱۸]. ویژگی زمانی، مرتبط با رابطه میان نواحی صحنه‌ها در طول زمان (فریم‌های متوالی) می‌باشد [۵، ۶، ۸، ۱۴، ۱۹، ۲۰]. ویژگی زمانی به دو دسته کلی ویژگی زمانی بلندمدت و ویژگی زمانی کوتاه‌مدت تقسیم‌بندی می‌شود [۱۸]. ویژگی مکانی-زمانی، حالت ترکیبی از دو ویژگی بیان‌شده می‌باشد. در فضای پژوهش‌های مرتبط با حوزه تحلیل ویدیو، به‌کارگیری این ویژگی در راستای شناسایی روند تغییرات هرناحیه، کمک دوچندانی به حل مسائل مرتبط با تشخیص رفتار می‌کند [۵، ۸، ۱۱، ۱۴، ۱۹، ۲۱].

با توجه به محدودیت‌های موجود در تشخیص رویدادهای نادر، تمرکز بر یادگیری رویدادهای رایج است. درواقع پس از یادگیری رویدادهای رایج توسط الگوریتم، متناسب با شرایط محیطی رویدادهایی که خارج از حد آستانه مناسب قرار گیرند، به عنوان رویداد نادر شناخته می‌شوند [۲۲]. منظور از حد آستانه، میزان هم‌خوانی ویژگی‌های استخراج شده توسط شبکه، نسبت به ویژگی‌های فراگیری شده در زمان آموزش است [۲۳]. برخی پژوهش‌ها نظیر دنگ و همکارانش [۲۴]، تمرکز اصلی را بر شناسایی حدآستانه مطلوب گذاشته‌اند. آلبوساک[12] و همکارانش [۲۵]، با به‌کارگیری تولید سه قانون برای هر دوربین، رفتار غیررایج اشیاء را با کمک سه مولفه کلاس شیء، موقعیت شیء و سرعت شیء تشخیص می‌دهند. برای مقابله با شرایط عدم قطعیتی که هر دوربین، منطق فازی در تصمیم‌گیری‌ها به‌کارگیری شده است.

روش‌های اولیه تشخیص رویدادهای رایج و نادر، اغلب مبتنی بر مدل‌سازی مسیرحرکت اشیاء بنا نهاده شده است. در رویکرد مذکور، تشخیص رویداد نادر از طریق شناسایی داده‌های پرت موجود در رویدادهای رایج صورت می‌پذیرد. درواقع در صورتی که شیء مورد نظر، مسیر حرکت رایجی را طی نکرده باشد، به عنوان حرکت غیررایج آن شیء تلقی می‌گردد. جیانگ[13] و همکارانش [۱۹]، ایده به‌کارگیری سه سطح مفاهیم مکانی-زمانی و مسیر را در راستای تشخیص رویدادهای نادر ارائه کردند. شلندونگ‌وو[14] و همکارانش [۲۶]، با خوشه‌بند K-means ناهنجاری‌های موجود را تشخیص دادند. کارتز و نیشینو [۲۷]، موقعیت الگوهای حرکتی را از طریق توزیع سه‌بعدی گوسی گرادیان مکانی-زمانی شناسایی کردند. در رویکرد مذکور، رویدادهای نادر توسط مدل مخفی مارکوف[15] تشخیص داده می‌شوند.

اغلب پژوهش‌های ذکر شده، مبتنی بر ویژگی‌های دستی پیچیده، جهت بازنمایی صحنه و حرکات درون آن درنظر گرفته شده‌اند. لیکن اخیرا روند فعالیت‌های پژوهشی این حوزه، مبتنی بر یادگیری ژرف صورت پذیرفته است [۷، ۹، ۱۴، ۱۶، ۱۹]. همچنین رویکرد استخراج تکه‌های مکعبی، توسط رده‌بند آبشاری[16] جهت شناسایی و بررسی نوع تکه استخراجی توسط سبکرو و همکارانش [۵] برای یادگیری ویژگی خودکار به‌کارگیری شده است. به‌کارگیری ساختار تمام متصل در ارائه بازنمایی ویژگی مرتبط با رویداد نادر نیز در پژوهش سبک‌رو و همکاران با عنوان deep-anomaly از آن یاد شده است [۵]. با توجه به تغییر رویکردی که در سایر مسائل بینایی و یادگیری ماشین نیز مشهود است، این پژوهش نیز از رویکردهای یادگیری ژرف و شبکه تخاصمی مولد برای حل مساله تشخیص و مکان‌یابی رویدادهای رایج و نادر در ویدیو استفاده شده که در ادامه بیان خواهد شد.

۳- شبکه تخاصمی مولد

شبکه تخاصمی مولد، که اختصاراً GAN نامیده می‌شود به عنوان جرقه‌ای برای یادگیری ژرف خطاب شده است [۲۸]. در ادامه جزئیات بیشتری از این شبکه ارائه خواهد شد.

۳-۱- معرفی شبکه تخاصمی مولد

این شبکه توسط گودفلو و همکارانش [۲] در سال ۲۰۱۴ ارائه شده است. پژوهش مذکور، با به‌کارگیری پیاده‌سازی ایده یادگیری دو شبکه عصبی در فضای یادگیری تخاصمی، شبکه تخاصمی مولد را در حوزه یادگیری ماشین و شبکه‌های ژرف ایجاد نموده‌است. GAN از دو زیر شبکه تشکیل شده که به نام‌های مولد[17] و ممیز[18] (متمایزگر) شناخته می‌شوند. هرکدام از این شبکه‌ها ساختار معماری اختصاصی دارند اما خروجی مولد به‌عنوان ورودی ممیز به‌کار گرفته می‌شود.

شبکه مولد به منظور یادگیری تولید نمونه‌های مشابه مجموعه دادگان اصلی مسئله مورد استفاده قرار می‌گیرد. این شبکه در نسخه اولیه در ورودی خود، بردار نمونه‌برداری شده از توزیع مشخص (مثلا نرمال) را داراست. این بردار طی عبور از لایه‌ها به ساختاری متناسب با ساختار داده مجموعه دادگان اصلی در لایه خروجی میرسد. درواقع در شبکه مولد ساده، نگاشت از بردار ورودی به ساختار داده مجموعه داده اصلی معمول است. درواقع داده‌ای تولید می‌شود که در ابعاد داده مجموعه داده اصلی مسئله

است. لازم به ذکر است که منظور از داده واقعی، همان مجموعه دادگان آموزشی مسئله و منظور از داده جعلی، داده تولید شده توسط شبکه مولد مدنظر است.

در پژوهش گودفلو و همکارانش [۲]، شبکه مولد به «تلاش گروهی جاعل[۱۹]، جهت تولید پول جعلی[۲۰]» و شبکه ممیز به «پلیسی که سعی در تشخیص پول‌های جعل شده دارد»، تشبیه شده‌است. جاعل(شبکه مولد) سعی می‌کند تا پلیس(شبکه ممیز) را فریب[۲۱] دهد اما در مقابل، پلیس سعی می‌کند تا فریب نخورد. شبکه مولد با تابع G[۲۲] و شبکه ممیز با تابع D[۲۳] شناخته می‌شوند. این دو شبکه همانند بازیکن، با قاعده min-max و قانون جمع‌صفر با یک‌دیگر در حال رقابت هستند. تابع هزینه کلی $V(G,D)$ به صورت زیر تعریف می‌شود [2]:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right] \quad (۱)$$

شبکه ممیز از مجموعه دادگان آموزشی تغذیه می‌شود و سعی در فراگیری توزیع دادگان اصلی دارد. این شبکه در تلاش است تا احتمال واقعی بودن دادگان آموزشی (داده واقعی) را به سمت یک و احتمال واقعی بودن همه خروجی‌های شبکه مولد (داده جعلی) را به سمت صفر سوق دهد. در رابطه فوق، میزان احتمال شبکه ممیز به ازای داده ورودی x با $D(x)$ نشان داده شده است. در روند آموزش شبکه ممیز، وقتی که x از توزیع دادگان آموزشی $p_{data}(x)$ نمونه‌برداری شده باشد، احتمال این نمونه به سمت یک سوق پیدا می‌کند. بردار z با نمونه‌برداری از توزیع مشخص p_z تولید شده است. شبکه مولد با دریافت نمونه‌های z درخروجی خود $G(z)$ را تولید می‌کند (این خروجی همتا با ساختار داده اصلی مسئله است). شبکه ممیز می‌بایست با دریافت داده مصنوعی (جعلی) تولید شده $G(z)$، احتمال صفر را به‌عنوان مقدار $D(G(z))$ در خروجی تولید کند. این درحالی است که شبکه مولد، در تلاش است تا داده تولیدی $G(z)$ شبیه به دادگان آموزشی مسئله باشد و در نهایت $D(G(z))$ برابر یک گردد. این روند رقابت بین سه تابع هزینه بیان شده تا زمان فراگیری نسبی درک دو شبکه پیش می‌رود.

شبکه‌های D و G، دارای پارامترها و معماری‌های مختص به خود هستند. اگر پارامترهای شبکه G با $\theta^{(g)}$ و پارامترهای شبکه D با $\theta^{(D)}$ درنظر گرفته شوند، تابع هزینه با $V\left(\theta^{(g)}.\theta^{(d)}\right)$، خروجی شبکه مولد با $\hat{x} = G(z;\theta^{(g)})$ و شبکه ممیز با $D(x;\theta^{(d)})$ نشان داده می‌شود.

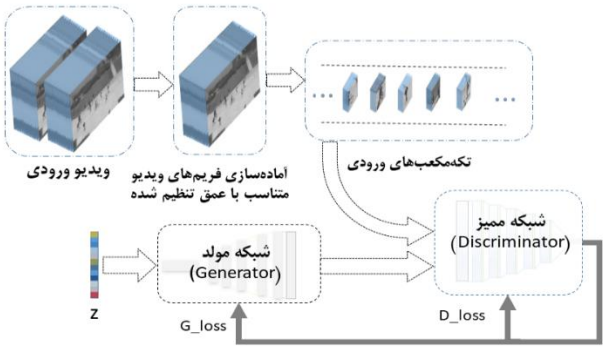
۳-۲- توسعه و کاربرد شبکه تخاصمی مولد

طبق ایده ارائه شده در پژوهش میرزا و اوسیندرو [۲۹]، با تزریق اطلاعات مجزا، به شبکه تخاصمی مولد شرطی تبدیل می‌شود. درواقع اطلاعات شرطی در قالب اطلاعات اضافه به مجموعه شبکه‌ها تزریق می‌شود. در مثال ارقام دست‌نویس MNIST [۳۰]، شرط مذکور روی برچسب کلاس در نظر گرفته شده است. از این ایده برای بیان رویکردهای یادگیری مدل‌های چندحالته در مجموعه دادگان MIRFlicker-25000 [۳۱]، نیز استفاده شده است. به‌کارگیری لایه‌های کانولوشنال در زیرساخت شبکه تخاصمی مولد پژوهش رادفورد و همکارانش [۳۲] به‌کارگیری شده است. به طور معمول شبکه‌های کانولوشنال از یادگیری بانظارت استفاده می‌کنند. به این ترتیب با توجه به نوع یادگیری کلی بدون‌نظارت در GAN، سعی شد تا با به‌کارگیری لایه‌های کانولوشنال، مزیت‌های یادگیری بانظارت و یادگیری بدون‌نظارت در کنار هم استفاده شوند [۳۲].

هدف تقطیع مفهومی[۲۴]، انتساب برچسب به هرکدام از پیکسل‌های تصویر است. برای انتساب، نیاز به تعداد مشخصی از داده برچسب خورده سطح پیکسلی[۲۵] است و اغلب در دسترس نیست. برای نشان‌دادن این فقدان اطلاعاتی، در پژوهش سولی[۲۶] و همکارانش [۳۳]، سعی شده تا اولاً در حجم وسیعی از اطلاعات بدون برچسب یا نیمه برچسب خورده و ثانیاً روی تصاویر غیرواقعی تولید شده توسط شبکه تخاصمی مولد، تمرکز شود. معماری پایه پژوهش مذکور، به صورت ساختار نیمه‌نظارت شده مبتنی بر شبکه تخاصمی مولد، ارائه شده است. شلگ[۲۷] و همکارانش [۳۴]، از روش‌های بدون‌نظارت، برای شناسایی ناهنجاری در دادگان فیلم‌برداری شده، همراه با نشانه‌های کاندیدشده استفاده کرده‌اند. روش AnoGAN به منظور فراگیری تغییرات رویه نرمال آناتومیک[۲۸]، همراه با شمای امتیازدهی ناهنجاری مبتنی بر نگاشت فضای‌تصویر به فضای‌پنهان[۲۹] است. در واقع هدف، اعمال‌کردن برچسب ناهنجاری به دادگان جدید و امتیازدهی به تکه‌های[۳۰] تصاویر در جهت برازش بر توزیع یادگرفته شده است [۳۴].

٤- معماری پیشنهادی

در این بخش، جنبه‌های مختلف معماری پیشنهادی اعم از دادگان مورد استفاده، معماری و الگوریتم تشریح می‌شود. این معماری رویدادهای نادر موجود در یک ویدیو را شناسایی ومکان‌یابی می‌نماید و طبیعتا سایر رویدادها رایج درنظر گرفته می‌شود.



شکل ۱: شمای کلی از معماری پیشنهادی

روال کار این معماری به این صورت است که در ابتدا، فریم‌هایی از ویدیوی دریافتی، طبق تعداد فریم تعریف شده، استخراج می‌شوند. در گام بعدی، همه فریم‌های موجود در یک دسته[۳۱]، متناسب با پارامتر تنظیم اندازه ناحیه، تکه‌تکه می‌شوند. در انتهای این گام، مکعب‌های فریمی محیا می‌شوند. اگر تعداد فریم تنظیم شده واحد در نظر گرفته شود، این مکعب به فریم واحد تبدیل می‌شود. فرآیند آماده‌سازی دادگان در مراحل آموزش و آزمون به همین منوال است.

معماری الگوریتم پیشنهادی همانند شبکه تخاصمی مولد، متشکل از دو شبکه مولد و ممیز بر مبنای دو فاز آموزش و آزمون است. در فاز آموزش، ساختار شماتیک شکل ۱ به‌کارگیری می‌شود. در فاز آزمون، فقط از شبکه ممیز به‌کارگیری می‌شود. در واقع در فاز آزمون، داده‌های ورودی مطابق مراحل مذکور آماده می‌شوند و با ورود این دادگان به شبکه ممیز، احتمالی مبنی بر رایج بودن داده مورد نظر داده می‌شود. سپس در مرحله پس‌پردازش، متناسب با شرایط محیطی و پارامتر حدآستانه، رویداد رایج و نادر مشخص می‌گردد. در ادامه بخش‌های مختلف معماری تشریح می‌شوند.

شکل ۲: نحوه آماده‌سازی جهت ارسال به شبکه تخاصمی مولد

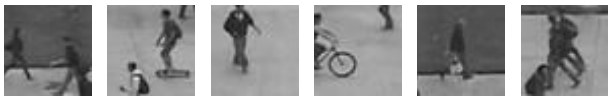
٤-١- آماده‌سازی داده

ورودی معماری پیشنهادی تعدادی فریم ویدیویی خواهد بود. هر فریم با توجه به نقاط مکانی و زمانی، در راستای دقتی که موردنیاز است ناحیه‌بندی می‌گردد. سپس به‌صورت موازی، تکه‌ها ایجاد و جهت پیش‌پردازش آماده‌سازی می‌گردند (شکل ۲). در وهله بعد، داده‌ها به ساختار آموزش یا آزمون شبکه تخاصمی مولد داده می‌شود (شکل ۱).

نکته مهمی که در ساختار اغلب پژوهش‌های مبتنی بر شبکه‌های کانولوشنال وجود دارد، عدم نیاز به پیش‌پردازش‌های حرفه‌ای است [۳، ۷، ۸، ۱۸، ۲۴، ۳۸-۳۴].

٤-٢- ورود اطلاعات به شبکه

در هنگام تغذیه شبکه ممیز با نمونه‌های مجموعه دادگان اصلی (شکل ۳)، این شبکه سعی می‌کند احتمال نمونه‌های مجموعه داده اصلی یا همان $P_{real\_data}(x)$ را به بیشینه مقدار واقعی(یعنی یک) نزدیک کند. در شبکه مولد، شرایط به خاطر ماهیت مولدی بودن آن، کمی متفاوت است. درواقع شبکه مولد با شروع از یک بردار نمونه‌برداری شده از توزیع مشخص، مقادیر را به‌گونه‌ای مدیریت می‌کند تا در انتها به ساختاری هم‌اندازه و شبیه به مجموعه دادگان ورودی شبکه ممیز برسد. سپس داده تولید شده جهت بررسی تحویل ممیز داده می‌شود.



شکل ۳: نمونه‌ای از ناحیه دادگان آماده شده جهت ارسال به شبکه رقابتی

پژوهش حاضر مبتنی بر زیرساخت اولیه پژوهش رادفورد و همکاران[۳۲] است. در ساختار پژوهش جاری، ایده افزایش سرعت نرمال‌سازی دسته‌ای بیان شده در پژوهش لاف[۳۲] و همکارانش [۳۵] نیز به‌کارگیری شده است. همچنین توابع فعال‌ساز مختلف از جمله ReLU و PLReLU [40] در معماری به‌کارگیری شده است. چارچوب یادگیری شبکه‌های خصمانه مولد به‌گونه‌ای است که دید خوبی از توزیع دادگان اصلی، ارائه می‌کند. همچنین با توجه به بدون‌نظارت بودن تشخیص رویدادهای نادر، می‌توان با استفاده از این چارچوب یادگیری، به فراگیری مجموعه دادگان رایج محیط پرداخته شود و محدودیت‌های قبل را پوشش داد. روال تشخیص رویداد نادر نیز هم‌راستا با پاسخ شبکه ممیز است. در معماری شبکه مولد به عنوان تولیدکننده شرایط غیررایج به‌کار می‌رود. این شبکه در در فعالیت‌هایی نظیر افزایش‌داده[۳۳] نیز قابل استفاده است. نکته حائز اهمیت آن است که الزاما همه موارد تولید شده شبکه مولد، جزو شرایط نادر قرار نمی‌گیرند، لذا در این پژوهش آن‌ها را موارد غیر رایج می‌نامیم. به عبارت دیگر ممکن است مواردی غیررایج وجود داشته باشد که اصلا امکان تحقق در محیط واقعی به عنوان نادر وجود نداشته باشند. لذا این رویدادهای غیررایج در هیچ دسته‌ای

به جهت غیرواقعی بودن قرار نگرفته‌اند (بنابراین نه نادر و نه رایج هستند).

در این رویکرد، تشخیص رویداد رایج و نادر، مبتنی بر مقدار احتمال شبکه ممیز است. از آنجا که هر ویدیو، از تعدادی فریم تشکیل‌شده، اطلاعات ورودی، مطابق شرایط مکانی هر بخش از فریم، به شبکه وارد می‌شود. درواقع، نواحی فریم‌های رویدادهای رایج، به عنوان مجموعه‌دادگان آموزشی جهت آموزش در نظر گرفته می‌شوند. پس از مرحله آموزش، مدل شبکه ممیز قابلیت تشخیص رویدادهای رایج و نادر را دارد.

٤-٣- معماری و آموزش شبکه تخاصمی مولد

در این ساختار، شبکه ممیز به ازای هر داده آموزشی ویدیوی رایج، یک مقدار احتمال $P_{normal\_data}(x)$ در نظر می‌گیرد. در حین روال آموزش، شبکه ممیز این احتمال را برای دادگان آموزشی به سمت یک و برای نمونه‌های تولیدی شبکه مولد (جعلی)، به سمت صفر سوق دهد. در انتهای یادگیری دوشبکه، رابطه (۲) برقرار می‌شود [2].

$$D^*(x) = \frac{P_{normal\_data}(x)}{P_{normal\_data}(x) + P_{generated\_data}(x)} \quad (۲)$$

مطابق رابطه (۲)، در انتهای شرایط بهینه آموزشی، داده تولیدشده توسط شبکه مولد، بسیار شبیه به مجموعه دادگان اصلی است، لذا شبکه ممیز به آن احتمال یک (۱۰۰٪ واقعی) تخصیص می‌دهد. حال طبق رابطه مذکور، مدل شبکه ممیز از این پس مجبور است با احتمال ۵۰٪ به هر داده، واقعی بودن را اعلام کند. در این صورت، شبکه مولد آنچنان خوب عمل کرده که شبکه ممیز قادر به تفکیک داده اصلی و داده جعلی تولید شده نیست. شبکه مولد در حین روال آموزش در تلاش است تا خروجی تولیدی به گونه‌ای باشد که میزان احتمالی که شبکه ممیز به آن تخصیص می‌دهد به یک نزدیک باشد. این روال، به صورت دوره‌ای در حین آموزش به تعداد مشخصی تکرار می‌شود. سپس شبکه مولد مبتنی بر هدف خود، وزن‌ها را بروزرسانی می‌کند.

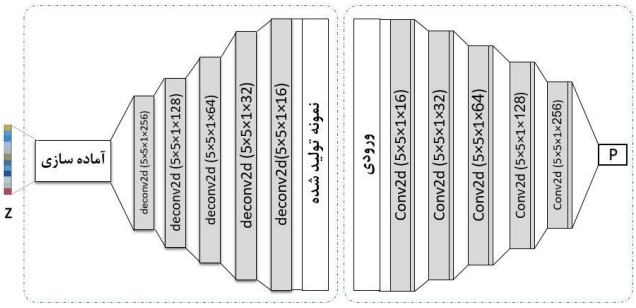
یکی از چالش‌های چارچوب یادگیری شبکه تخاصمی مولد، روند توقف آموزش است. همواره ساختار آموزش به نحوی پیش می‌رود که توزیع فراگیری شده در راستای توزیع دادگان آموزشی مسئله (فریم‌های ویدیوهای رایج) باشد. همچنین مقادیر خروجی شبکه ممیز به ازای دادگان غیررایج با مقادیر نزدیک به صفر ارائه شود. خروجی شبکه ممیز در این پژوهش، به عنوان عامل تشخیص‌دهنده و معیار نمره‌دهی به هر ناحیه از داده ورودی در نظر گرفته شده است. در رابطه (۳) مسئله بهینه‌سازی روند یادگیری چارچوب الهام‌گرفته پیشنهادی درقالب مسئله min-max بیان شده است [2]:

$$\min_G \max_D V(D|G) = \mathbb{E}_{ic\sim P_{normal\_dc}}[logD^*(ic)] + \mathbb{E}_{gc\sim P_g}[\log(1-D^*(gc))] \quad (۳)$$

منظور از ic مکعب داده ورودی، $normal\_ic$ داده رایج موجود در مسئله مورد نظر، P_{normal_ic} توزیع مجموعه دادگان رایج، P_g توزیع یادگرفته شده شبکه مولد، gc مکعب تولیدشده توسط شبکه مولد است. از نماد $\mathbb{E}$ برای نمونه‌برداری مجموعه داده اصلی و نیز مجموعه داده تولیدی شبکه مولد متناسب با پایین‌نویس مرتبط استفاده می‌شود. در این رابطه شبکه ممیز (D)، در تلاش است تا به ازای ورودی‌های اصلی x، مقدار احتمال یک را در خروجی نشان دهد. همچنین در تلاش است مقدار صفر به ازای داده تولیدشده شبکه مولد در خروجی نمایان گردد. هدف شبکه مولد، برعکس هدف شبکه ممیز است. در واقع شبکه مولد در تلاش است تا داده تولیدی به گونه‌ای تولید شود تا میزان احتمال $D^*(gc)$ به یک (میزان واقعی بودن داده) نزدیک شود.

٥- پیاده‌سازی

معماری ارائه شده در پژوهش رادفورد و همکاران [۳۲]، به عنوان معماری پایه مورد استفاده این پژوهش قرار گرفته است. در رویکرد تک‌فریم، همه لایه‌های کانولوشنال با کرنل‌های ۱×۵×۵ انجام می‌شود. این کرنل‌ها با گام حرکت ۲ در هر راستا عبور می‌کنند. درواقع روی داده ورودی لایه کانولوشنال، کرنل ۵×۵ با عمق یک اعمال می‌گردد. در سمت چپ شکل ۴ معماری شبکه مولد و در سمت راست آن، معماری شبکه ممیز ترسیم شده است. شبکه‌های ممیز و مولد از ۵ لایه کانولوشنال مطابق با معماری الهام گرفته از پژوهش آقای رادفورد[۳۴] و همکارانش [۳۲] طراحی شده است.



شکل ۴ شبکه‌های مولد و ممیز در شبکه تخاصمی مولد کانولوشنی

در پیکربندی پیاده‌سازی شده، میزان فضای‌ویژگی[۳۵] مرتبط با هر لایه کانولوشنال، متناسب با تجهیزات پردازشی تنظیم می‌گردد. از آن‌جا که مبنای معماری این پژوهش، پژوهش آقای رادفورد و

همکارانش [۳۲] است، رشد فضای‌ویژگی در لایه‌بندی معماری، ضریبی از توان‌های ۲ در نظر گرفته شده است. لازم به ذکر است که به‌کارگیری معماری پیشنهادی، الهام‌گرفته از معماری پژوهش رادفورد و همکاران [۳۲] و پژوهش اله و پتروسینو [۴۱] است. در پژوهش اله و پتروسینو [۴۱] ویژگی‌های هرمی به تفصیل بررسی شده است. همچنین در آزمایش‌های مختلف، بردار ورودی شبکه مولد در اندازه ۱۰۰، ۲۰۰، ۴۰۰ و ۵۰۰ به صورت جداگانه در نظر گرفته شده تا تاثیر اندازه ورودی بردار z نیز در خروجی شبکه مولد و نیز آموزش شبکه ممیز بررسی گردد.

در معماری ارائه شده پژوهش جاری، خروجی هر لایه کانولوشنال، به لایه نرمال‌سازی دسته‌ای مطابق پژوهش لاف و همکارانش [۳۵]، با مقادیر پارامترهای جدول ۱ داده شده است. لازم به ذکر است که در این پژوهش باتوجه به پیاده‌سازی‌های مختلف از تابع نرمال‌سازی دسته‌ای، جهت اعمال این لایه ، پیاده‌سازی پژوهش لاف و همکاران [۳۵] در تابع اختصاصی کتابخانه تنسورفلو[۳۶] به نام tf.contrib.layers.batch_norm [۴۲] به‌کارگیری شده است. خروجی مرحله نرمال‌سازی دسته‌ای، از آشکارساز (تابع فعالیت) نیز عبورکرده است.

جدول ۱: آرگومان‌های لایه نرمال‌سازی دسته‌ای

نام پارامتر	مقادیر
Epsilon	1E-5
Momentum	0.9

لازم به ذکر است که در لایه کانولوشنال، برای مقداردهی اولیه وزن‌ها، توزیع‌نرمال بریده‌شده[۳۷] [۳۶] با انحراف‌معیار ۰.۰۲ و مقدار بایاس صفر به‌کارگیری شده است. همچنین از مقداردهی اولیه ژاویر[۳۸] [۴۳] نیز در برخی آزمایش‌های معماری پیشنهادی برای مقداردهی اولیه استفاده شده است. همچنین میزان ضریب نشت‌اطلاعاتی در LReLU به مقدار ۰.۲ در نظر گرفته شده است. در این پژوهش نسخه ۱.۲ تنسورفلو به‌کارگیری شده است. جهت پیاده‌سازی عملیات دیکانولوشنال، ترانهاده کانولوشنال پیاده شده در کتابخانه تنسورفلو با نام tf.nn.conv2d_transpose [۴۴]، به‌کارگیری شده است. در صورتی که از نسخه‌های پایین‌تر جهت پیاده‌سازی به‌کارگیری شود، محدودیت‌هایی وجود دارد که درنظر گرفته شده است. برای مثال در نسخه‌های پایین‌تر از ۰.۷ تنسورفلو، برای پیاده‌سازی عملیات دیکانولوشنال، tf.nn.deconv2d [۴۴] به‌کارگیری می‌شود.

۶- آزمایشهای تجربی

این بخش به بیان معیارهای ارزیابی و مجموعه دادگان و سپس تحلیل آزمایش‌های مختلف روی روش پیشنهادی می‌پردازد.

۶-۱- معیارهای ارزیابی

باتوجه به هدف اصلی در تشخیص رویدادهای رایج و نادر، این مسئله به صورت یک مسئله دو کلاسه مدل‌سازی می‌شود. از روش‌های معمول در بررسی و ارزیابی مسائل دوکلاسه، نمایش عملکرد توسط منحنی مشخصه عملکرد (ROC[۳۹]) و مساحت سطح زیر منحنی(AUC[۴۰]) می‌باشد. این معیارها مبتنی بر مشخصه‌هایی نظیر نرخ مثبت واقعی و نرخ مثبت کاذب که در ادامه آورده‌شده سنجیده می‌شوند:

۶-۱-۱- نرخ مثبت واقعی(TPR[41])

نرخ میزان صحت دادگان پیشنهادی از کل دادگان صحیح، با نرخ مثبت واقعی یا حساسیت یاد می‌شود که رابطه آن به صورت رابطه ۴ است. منظور از TP[42] و FP[43] به ترتیب مثبت واقعی و مثبت کاذب می‌باشد.

$$TPR = \frac{TP}{TP + FP} \qquad (۴)$$

۶-۱-۲- نرخ مثبت کاذب(FPR[44])

نرخ میزان اشتباه عملکردی معماری پیشنهادی در تشخیص نسبت به کل میزان تشخیص‌های موجود، به عنوان نرخ مثبت کاذب شناخته می‌شود که رابطه آن مطابق رابطه ۵ است.

$$FPR = \frac{FP}{TN + FP} \qquad (۵)$$

۶-۱-۳- مشخصه FPS[۴۵]

جهت ارزیابی بازدهی زمانی سامانه‌های تحلیل ویدیو، از مشخصه FPS استفاده می‌شود. این مشخصه نشان دهنده تعداد فریم محاسبه شده در واحد زمان (ثانیه) است.

۶-۲- دادگان مورد استفاده

در این پژوهش از دو مجموعه داده UCSD[۴۶] با نام‌های UCSDped1 و UCSDped2 به منظور آموزش و آزمون معماری پیشنهادی به‌کارگیری شده است [۴۵]. این دو مجموعه داده شامل تصاویر ویدیویی دوربین‌های نظارتی تردد مردم در پیاده‌رو و خیابان فیلم‌برداری شده است. در این مجموعه‌داده، عابرین پیاده، افراد اسکیت‌سوار، دوچرخه سوار و وسایل نقلیه عبور می‌کنند. مجموعه داده UCSDped1، شامل ۳۴ نمونه ویدیو آموزشی و ۳۶ نمونه ویدیو آزمون است. مجموعه داده UCSDped2، متشکل از ۱۶ نمونه ویدیو آموزشی به همراه ۱۲ نمونه ویدیو آزمون است. در این مجموعه داده، حرکت افراد درجهت افقی صورت گرفته است. اغلب نمونه‌های آموزشی شامل

۱۲۰ فریم می‌باشند. نمونه‌هایی نظیر حرکت اسکیت‌سوار، عبور وسیله‌نقلیه، حمل کوله‌پشتی از نمونه تکه‌داده‌های نادر و بقیه قسمت‌ها (همانند راه‌رفتن عابرین، طبیعت محیط) از نمونه تکه‌های رایج می‌باشند. در این آزمایش، معماری شبکه روی اندازه‌های مختلف تکه‌فریم ورودی از ۳۵×۳۵ الی ۶۰×۶۰ مورد بررسی قرار گرفته و ۴۵×۴۵ جزو بهترین حالاتی بود که متناسب با موقعیت دوربین و اشیای مورد انتظار رایج عبوری و معیارهای بررسی مسئله جهت پیشروی پژوهش انتخاب شد.

۶-۳- جزئیات سخت‌افزاری و نرم افزاری مورد استفاده پژوهش

بستر پردازشی پژوهش، دارای کارت گرافیک با مشخصات NVIDIA Geforce GTX TITAN، با فضای ۳۲ گیگابایت حافظه‌اصلی[47] است. البته متناسب با شرایط با چند بستر پردازشی دیگر نیز ارزیابی گردیده که جزئیات آن‌ها نیز مطرح شده است. همچنین در بحث نرم‌افزاری، از سیستم عامل لینوکس و توزیع اوبونتو[48] و نسخه ۱۶.۰۴ استفاده شده است. لازم به ذکر است که جهت پیاده‌سازی از نسخه ۳.۵ پایتون و ۱.۲ تنسورفلو به‌کارگیری شده است (جدول ۲).

جدول ۲: جزئیات سخت‌افزاری/نرم‌افزاری مبنای پژوهش جاری

سخت‌افزار/ نرم‌افزار	جزئیات
Processor	Intel Core i7-3770 @3.40GHz ×8
RAM	32.0 GB
System type	64-bit
GPU	NVIDIA Geforce GTX TITAN
OS[49]	Ubuntu 16.04

۶-۴- نتایج بررسی روی UCSDped1و UCSDped2

مجموعه UCSDped1 به علت حرکت عمقی که افراد دارند (از دوربین دور می‌شوند و اشیاء موجود به مراتب از بزرگ به کوچک به علت موقعیت دوربین تغییر موقعیت پیدا می‌کنند) جزو مجموعه‌های سخت محسوب می‌شود. در ادامه نتایج منحنی مشخصه عملکرد روی این مجموعه‌دادگان آورده شده است. نرخ مثبت کاذب در این مجموعه‌دادگان به نسبت دیگر مجموعه‌دادگان مشابه به علت سخت بودن شرایط تشخیص، بالاست. در نتیجه میزان نرخ خطای برابر نیز نسبت به UCSDped2 بالاتر بوده است.

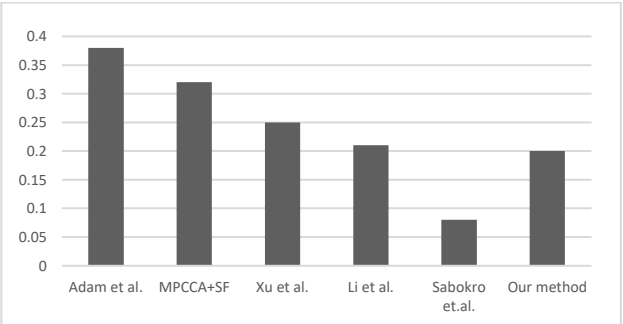
نتایج حاصل‌شده از مجموعه داده UCSDped1 و UCSDped2 به ترتیب در جدول‌های ۳ و ۴ آورده شده است. همچنین نمودار مقایسه‌ای آن‌ها در شکل‌های ۵ و ۶ آورده شده است. نتایج این پژوهش نسبت به میانگین پژوهش های اخیر انجام شده دقت خوبی داشته و این میزان نسبت به سرعت پردازشی آن در مقایسه با سایر پژوهش‌ها در رده‌بندی مناسب بی قرار دارد.

همچنین راحتی آموزش و کوتاه بودن مراحل آموزش آن نیز از دیگر مزایای این پژوهش است. درواقع بهره‌گیری از رویکرد تخاصمی‌مولد پیشنهادی این پژوهش جهت حل مسئله، نسبت به روش‌های سنتی و اخیر مورد استفاده در این حوزه، رشد مطلوب ۱۸ درصدی نسبت به رویکرد پایه داشته است. البته نسبت به برخی پژوهش‌ها نظیر سبکرو و همکاران [۳، ۵] نرخ‌خطای‌برابر بیشتری کسب کرده ولی مزیت این پژوهش نسبت به روش‌های پیشین، پیاده‌سازی زیرساخت یکپارچه (انتها به انتها) و سهولت زیاد در روال آموزش و آزمون است. همچنین بهره‌مندی از خاصیت‌های جانبی شبکه مولد در تولید و افزایش داده در کنار هدف اصلی آن نیز می‌توان اشاره کرد. لازم به ذکر است که مکان‌یابی نیز به صورت پاسخ واحد، توسط تشخیص شبکه ممیز انجام می‌شود و نیاز به مراحل خاص دیگری ندارد.

جدول ۳: نتایج حاصل‌شده از مجموعه UCSDped1

روش	EER (%)	AUC (%)
Adam et al. [۴۶]	۳۸	۶۵
MPCCA+SF [۲۰]	۳۲	۵۹
Xu et al. [۴۷]	۲۵	۸۲
Li et al. [۸]	۲۱	۸۷.۹
Sabokro et al. [۳]	۰۸	۹۳.۲
معماری پیشنهادی	۲۰	۶۴

در شکل ۵، نتایج نرخ‌خطای‌برابر، روش‌های مختلف برای مجموعه دادگان UCSDped1 مصور شده است. رویکرد پژوهش جاری توانسته نسبت به روش‌های پایه و متناسب با یادگیری تخاصمی، به میزان بسیار خوبی نقش ایفا کند. مزیت آموزش یکپارچه (انتها به انتها) پژوهش نسبت به رویکرد آموزش مقطعی و پیچیده سبکرو و همکاران [۳] از مزایای بهینه این روش جهت پیشبرد هدف در زمان آموزش و آزمون است.



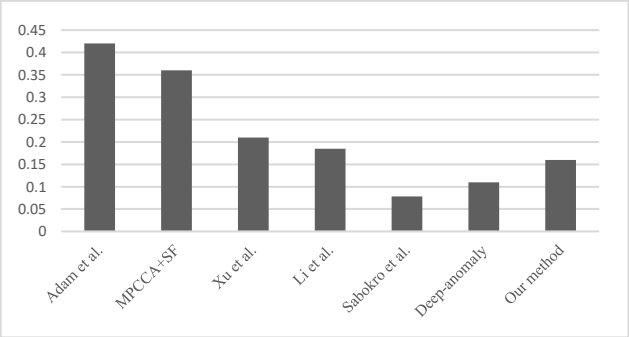
شکل ۵ نرخ‌خطای‌برابر مجموعه دادگان UCSDped1 در روش‌های مختلف

نتایج بررسی مدل پیشنهادی در تشخیص شرایط نادر مجموعه داده UCSDped2 نیز در شکل ۶ آورده شده است. همان‌طور که مشخص است؛ رویکرد پیشنهادی نسبت به روش‌های پایه

توانسته نتایج بهتری داشته باشد. همچنین با توجه به سادگی مجموعه داده UCSDped2 نسبت به UCSDped1، پژوهش Deep-anomaly [۵] نیز صرفا به بیان نتایج UCSDped2 بسنده کرده است. اما نتایج پژوهش جاری روی هر دو دادگان بوده، که نتایج حاکی از عملکرد بهتر آن روی مجموعه داده UCSDped1 بوده است. همان‌طور که در بخش پیشنهادهای آتی آورده شده، رویکرد تخاصمی‌مولد به عنوان یک مسیر جدید در عرصه تشخیص و مکان‌یابی رویدادهای رایج و نادر خواهد بود.

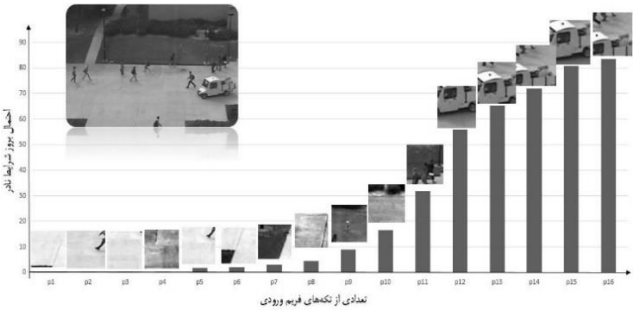
جدول ۴: نتایج به‌دست‌آمده از مجموعه UCSDped2

روش	EER (%)	AUC (%)
Adam et al. [۴۶]	۴۲	۶۳
MPCCA+SF [۲۰]	۳۶	۶۱.۲
Xu et al. [۴۷]	۲۱	۸۸.۲
Li et al. [۸]	۱۸.۵	--
Sabokro et al. [۳]	۷.۵	۹۳.۹
Deep-anomaly [۵]	۱۱	--
معماری پیشنهادی	۱۷	۸۰



شکل ۶: نرخ‌خطای‌برابر مجموعه دادگان UCSDped2 در روش‌های مختلف

در شکل ۷، امتیاز رخداد رویدادهای نادر در تکه‌تکه‌های یکی از فریم‌های ویدیو UCSDped2 در قالب نمودار آماری با نگرش بصری‌سازی نتایج آورده شده است.



شکل ۷: بصری‌سازی نمودار امتیاز نواحی در تکه‌فریم ویدیویی توسط شبکه ممیز به ازای داده UCSDped2

۶-۵- تاثیر اندازه ورودی شبکه مولد (z)

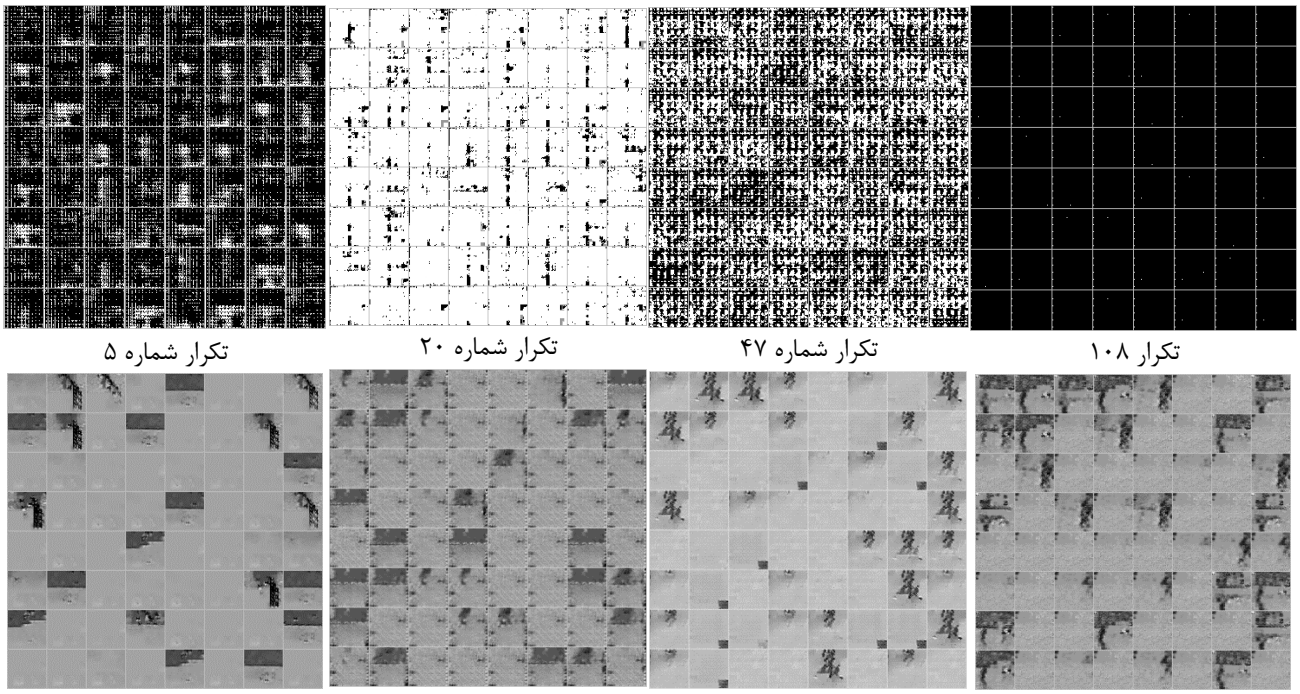
در این ارزیابی، به ترتیب اندازه‌های بردار ورودی z در ابعاد ۱۰۰، ۲۰۰، ۴۰۰ و ۵۰۰ جهت بررسی میزان بهبود آموزش در شبکه مولد در نظر گرفته شده است. تغییر چنین پارامترهایی نیازمند تغییر پارامترهای مختلف دیگر برای همگرایی آموزش است. نتایج بررسی این آزمایش نشان داده که پیچیده‌تر شدن شبکه مولد با انتخاب مقادیر بزرگ نسبت به انتخاب مقادیر کوچک امری بدیهی است. پیچیدگی شبکه مولد، حساسیت و نیازمندی به بهینه انتخاب کردن سایر پارامترها را در آموزش سخت‌تر می‌کند. ارزیابی‌ها مقدار ۲۰۰ را به عنوان میزان همگرایی بهتر در شبکه مولد جهت تولید داده مطابق شکل ۸ در حالت به‌کارگیری نرمال‌سازی دسته‌ای نشان داده‌اند.

۶-۶- تاثیر نرمال‌سازی دسته‌بندی

در این آزمایش، معماری شبکه پیشنهادی، با پارامترها و ساختار یکسان، در دو شرایط مختلف، پیاده‌سازی و اجرا شده است. به‌کارگیری و عدم به‌کارگیری نرمال‌سازی دسته‌ای در چینش لایه‌ای از شرایط این آزمایش در نظر گرفته شده است. نتایجی که در شکل ۸ آورده شده، نشان‌دهنده تاثیر بسیار زیاد نرمال‌سازی دسته‌ای در روند یادگیری شبکه پیشنهادی و نیز شبکه‌های مبتنی بر شبکه تخاصمی مولد است. درواقع به‌کارگیری نرمال‌سازی دسته‌ای با کاهش شیفت کواریانس داخلی، تمرکز زیاد بر مقداردهی پارامترها را کم و این روال را در ورودی لایه خود کنترل می‌کند [۳۶].

همان‌طور که در شکل ۸ نمایی از خروجی معماری پیشنهادی بدون به‌کارگیری نرمال‌سازی دسته‌ای(ردیف اول) و با به‌کارگیری نرمال‌سازی دسته‌ای (ردیف دوم) مشاهده می‌شود، نتایج بدست آمده با به‌کارگیری نرمال‌سازی دسته‌ای بهتر از عدم به‌کارگیری آن است.

در حالت بدون به‌کارگیری نرمال‌سازی دسته‌ای (سطر اول شکل ۸)، از تکرارهای ۹۸ به بعد، فضای یادگیری شبکه مولد به کلی از بین رفته و نیاز است تا پارامترهای مختلف شبکه جهت همگرایی بهتر، تغییراتی داشته باشند. درحالت به‌کارگیری نرمال‌سازی دسته‌ای (سطر دوم شکل۸)، شبکه همگرایی خوبی در فضای یادگیری مجموعه‌دادگان آموزشی داشته است. در مقایسه با حالت قبل، از تکرارهای ۹۸ به بعد، کلیات توزیع را فراگرفته و تمرکز خود را روی جزئیات توزیع دادگان متمرکز کرده است. لذا با به‌کارگیری نرمال‌سازی دسته‌ای، آموزش شبکه بهتر صورت می‌پذیرد.



شکل ۸: نمایی از خروجی معماری پیشنهادی بدون به‌کارگیری نرمال‌سازی دسته‌ای(ردیف اول) و با به‌کارگیری نرمال‌سازی دسته‌ای (ردیف دوم)

۷- نتیجه‌گیری و پیشنهادهای آتی

در این پژوهش، معماری و نگرش جدیدی مبتنی بر زیرساخت شبکه تخاصمی مولد، برای تشخیص رویدادهای رایج و نادر ارائه گردید. این معماری بر پایه استخراج و استفاده خودکار از ویژگی‌های داده ورودی ویدیویی است. این پژوهش جهت تشخیص رویدادهای رایج و نادر با تمرکز بر شبکه ممیز پیش می‌رود. همچنین شیوه آموزش با به‌کارگیری مجموعه دادگان آموزشی (شرایط رایج) به عنوان داده آموزشی پیش‌می‌رود. ایده اصلی پژوهش، به‌کارگیری مدل یادگرفته شده شبکه ممیز در مرحله آزمون به منظور تشخیص می‌باشد. این تشخیص بدین صورت است که به هر ناحیه امتیازی تعلق می‌گیرد و مبتنی بر امتیاز بدست آمده، با حد آستانه مقایسه و سپس نادر یا رایج بودن آن مشخص می‌شود. به‌کارگیری نرمال‌سازی دسته‌ای نیز به عنوان یک عامل موثر در بهبود سرعت و روند همگرایی شبکه تخاصمی مولد در معماری از دیگر نتایج پژوهش جاری رقم می‌خورد. اندازه مطلوب تکه‌های مجموعه دادگان اصلی با اندازه ۴۵×۴۵ انتخاب شد. همچنین بردار ورودی شبکه مولد در ابعاد مختلف مورد ارزیابی قرار گرفت. نتایج نرخ‌خطای‌برابر در مجموعه‌دادگان UCSDped1 و UCSDped2 به ترتیب ۰.۲ و ۰.۱۷ در منحنی مشخصه عملکرد حاصل گردید.

لازم به ذکر است که ارزیابی در شرایط شدید نادر (برای مثال تردد اشیاء دور از دوربین و هم‌رنگ با محیط) با میزان خطای برابر ۰.۲ و در شرایط متوسط نادر (برای مثال تردد وسیله نقلیه در پیاده‌رو) با میزان نرخ‌خطای‌برابر ۰.۰۴ در منحنی مشخصه عملکرد حاصل شده است. همچنین تشخیص در زمان ۰.۰۲ ثانیه و معادل ۳۰۰ فریم برثانیه برای یک مجموعه فریم ورودی، نیز از مزایای دیگر این معماری پیشنهادی در مقایسه با معماری‌های مشابه بوده که قابلیت استفاده در فضای بلادرنگ را بهتر نشان می‌دهد. همچنین استفاده از پیاده‌سازی موازی جهت تشخیص هر ناحیه از تصویر باعث می‌شود تا روند امتیازدهی به نواحی مختلف سریع انجام پذیرد و در مدت زمان بسیار اندک، امتیاز ناحیه مشخص گردد. درپایان نتایج به‌دست‌آمده در کنار سایر معماری‌هایی که نتایج مطلوبی در این حوزه داشته‌اند، مورد بررسی قرار گرفت. باتوجه به به‌کارگیری رویکرد یکپارچه(انتها به انتها) این پژوهش، سهولت در مرحله یادگیری (آموزش) نیز از مزایای آن نسبت به پژوهش‌های پیشین با نرخ‌خطای‌برابر مطلوب بوده و نتایج قابل مقایسه خوبی ارائه شده است.

در این پژوهش تلاش شد تا با ارائه نگرشی جدید نسبت به عملکرد معماری‌های مبتنی بر شبکه‌های ژرف خصمانه مولد، مسیری برای حل مسائل مختلف بدون نظارت با رویکرد نیمه نظارتی و زیرساخت تخاصمی مولد ارائه گردد. در ادامه می‌توان به به‌کارگیری زیرساخت‌های مشابه معماری پیشنهادی استفاده کرد. همچنین تمرکز در بهبود رویکرد بهینه‌سازی روند یادگیری تخاصمی این پژوهش جهت حل مسئله نیز مفید خواهد بود. به‌کارگیری روش‌های بهینه جهت بررسی همه حالات پارامترهای ورودی

همانند جستجوی فراگیر[۵۰]، قابلیت ارائه محدوده مقادیر هر پارامتر، جهت اعمال در پارامترهای ورودی از دیگر مواردی اســـت که به بهبود این پژوهش کمک شایانی خواهد کرد.

سپاسگزاری

در اینجا لازم است از همه کسانی که در این پژوهش همکاری نمودند، به ویژه پژوهشگاه دانش‌های بنیادی و آزمایشگاه پردازش‌های هوشمند چندرسانه‌ای LIMP[۵۱] دانشگاه صنعتی امیرکبیر که علاوه بر دانشگاه صنعتی مالک اشتر بسترهای پردازشی را در اختیار قرار دادند، تشکر نماییم.

مراجع

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[2] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." In Advances in neural information processing systems, pp. 2672-2680. 2014.

[3] "Multi-scale and real-time non-parametric approach for anomaly detection and localization - ScienceDirect." [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314211002104. [Accessed: 17-Jun-2018].

[4] "Analyzing Tracklets for the Detection of Abnormal Crowd Behavior - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/7045881/. [Accessed: 17-Jun-2018].

[5] 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, Apr. 2017.

[6] S. Biswas and R. Venkatesh Babu, "Anomaly detection via short local trajectories," *Neurocomputing*, vol. 242, pp. 63–72, Jun. 2017.

[7] *Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes - ScienceDirect*. .

[8] Y. S. Chong and Y. H. Tay, "Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder," in *Advances in Neural Networks - ISNN 2017*, 2017, pp. 189–196.

[9] N. Anjum and A. Cavallaro, "Trajectory Association and Fusion across Partially Overlapping Cameras," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2009, pp. 201–206.

[10] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly Detection and Localization in Crowded Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[11] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang, "Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes," *Signal Processing: Image Communication*, vol. 47, pp. 358–368, Sep. 2016.

[12] T. Xiang and S. Gong, "Video behaviour profiling and abnormality detection without manual labelling," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, vol. 2, pp. 1238-1245 Vol. 2.

[13] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A Review of Anomaly Detection in Automated Surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1257–1272, Nov. 2012.

[14] "Learning Spatiotemporal Features with 3D Convolutional Networks." [Online]. Available: https://dl.acm.org/citation.cfm?id=2919929. [Accessed: 17-Jun-2018].

[15] "Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing." [Online]. Available: https://dl.acm.org/citation.cfm?id=1886106. [Accessed: 17-Jun-2018].

[16] J. R. Medel and A. Savakis, "Anomaly Detection in Video Using Predictive Convolutional Long Short-Term Memory Networks," *arXiv:1612.00390 [cs]*, Dec. 2016.

[17] T. Xiang and S. Gong, "Incremental and adaptive abnormal behaviour detection," *Computer Vision and Image Understanding*, vol. 111, no. 1, pp. 59–73, Jul. 2008.

[18] L. Xia, I. Gori, J. K. Aggarwal, and M. S. Ryoo, "Robot-centric Activity Recognition from First-Person RGB-D Videos," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 357–364.

[19] V. Saligrama, E. Arias-Castro, R. Chellappa, A. O. Hero, R. Nowak, and V. V. Veeravalli, "Introduction to the issue on anomalous pattern discovery for spatial, temporal, networked, and high-dimensional signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 1–3, Feb. 2013.

[20] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1975–1981.

[۲۱] ع. سردار و ر. هاونگی، "بهبود عملکرد الگوریتم خوشه‌یابی خودکار تصاویر رنگی به کمک پیش‌پردازش با شبکه عصبی خودسامانده (SOM)"، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۷، شماره سوم، صفحه ۱۰۷۳-۱۰۸۲، سال ۲۰۱۷

[22] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Holistic features for real-time crowd behaviour anomaly detection," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 918–922.

[23] R. J. Morris and D. C. Hogg, "Statistical Models of Object Interaction," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 209–215, Jun. 2000.

[24] Q. Dong, Y. Wu, and Z. Hu, "Pointwise Motion Image (PMI): A Novel Motion Representation and Its Applications to Abnormality Detection and Behavior Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 407–416, Mar. 2009.

[25] J. Albusac, J. J. Castro-Schez, L. M. Lopez-Lopez, D. Vallejo, and L. Jimenez-Linares, "A supervised learning approach to automate the acquisition of knowledge in surveillance systems," *Signal Processing*, vol. 89, no. 12, pp. 2400–2414, Dec. 2009.

[26] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2054–2060.

[27] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1446–1453.

[28] "What are some recent and potentially upcoming breakthroughs in deep learning? - Quora." [Online]. Available: https://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning. [Accessed: 17-Jun-2018].

[29] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv:1411.1784 [cs, stat]*, Nov. 2014.

[30] "MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges." [Online]. Available: http://yann.lecun.com/exdb/mnist/. [Accessed: 17-Jun-2018].

[31] "The MIRFLICKR Retrieval Evaluation." [Online]. Available: https://press.liacs.nl/mirflickr/. [Accessed: 17-Jun-2018].

[32] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv:1511.06434 [cs]*, Nov. 2015.

[33] N. Souly, C. Spampinato, and M. Shah, "Semi and Weakly Supervised Semantic Segmentation Using Generative Adversarial Network," *arXiv:1703.09695 [cs]*, Mar. 2017.

[34] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery," in *Information Processing in Medical Imaging*, 2017, pp. 146–157.

[35] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," p. 9.
[36] P. Warden, *tf.truncated_normal_initializer | TensorFlow*. 2017.
[37] P. M. Jodoin, J. Konrad, and V. Saligrama, "Modeling background activity for behavior subtraction," in *2008 Second ACM/IEEE International Conference on Distributed Smart Cameras*, 2008, pp. 1–10.
[38] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, no. 3, pp. 323–333, Mar. 2011.
[39] Y. Ganin *et al.*, "Domain-Adversarial Training of Neural Networks," in *Domain Adaptation in Computer Vision Applications*, Springer, Cham, 2017, pp. 189–209.
[40] "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/7410480. [Accessed: 12-Apr-2019].
[41] I. Ullah and A. Petrosino, "A Strict Pyramidal Deep Neural Network for Action Recognition," in *Image Analysis and Processing — ICIAP 2015*, 2015, pp. 236–245.
[42] "tf.contrib.layers.batch_norm," *TensorFlow*. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/contrib/layers/batch_norm. [Accessed: 17-Jun-2018].
[43] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.
[44] "tf.nn.conv2d_transpose," *TensorFlow*. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/nn/conv2d_transpose. [Accessed: 17-Jun-2018].
[45] "UCSD Anomaly Detection Dataset." [Online]. Available: http://www.svcl.ucsd.edu/projects/anomaly/dataset.html. [Accessed: 17-Jun-2018].
[46] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, Mar. 2008.
[47] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning Deep Representations of Appearance and Motion for Anomalous Event Detection," *arXiv:1510.01553 [cs]*, Oct. 2015.

پاورقی‌ها:

[1] Detection
[2] Event
[3] Rare
[4] Common
[5] Deep learning
[6] Localization
[7] Convolutional Neural Networks
[8] Generative Adversarial Network
[9] Shifub Zhou
[10] Segmentation
[11] Crawd
[12] Albusac
[13] Fan Jiang
[14] Shandong Wo
[15] Hidden Markov Model
[16] Cascade classifier
[17] Generator
[18] Discriminator
[19] Counterfeit
[20] Fake currency
[21] Fool
[22] Generator
[23] Discriminator
[24] Semantic segmentation
[25] Pixel-level
[26] Nasim Souly
[27] Thomas Schlegl
[28] Manifold of normal anatomical variability
[29] Latent space
[30] Patches
[31] Batch
[32] Surgey Loff
[33] Data augmentation
[34] Radford
[35] Feature map
[36] TensorFlow
[37] Truncated normal distribution
[38] Xavier
[39] Receiver Operating Characteristic
[40] Area Under the Curve
[41] True Positive Rate
[42] True Positive
[43] False Positive
[44] False Positive Rate
[45] Frames Per Second
[46] University of California San Diego
[47] Read Only Memory
[48] Ubuntu distribution
[49] Operation System
[50] brute-force
[51] LIMP (Laboratory of Intelligent and Multimedia Processing)