



# تشخیص اسلحه دستی با استفاده از مدل

## شبکه‌های عصبی کانولوشنال سه‌بعدی

سارا معتمد\* و الهام عسکری

استادیار، گروه کامپیوتر، واحد فومن و شفت، دانشگاه آزاد اسلامی، فومن، ایران

### چکیده

از آنجایی که رفتار افراد در ویدئوها به صورت سیگنال‌های سه‌بعدی است و جستجوی یک عمل خاص بسیار دشوار است؛ لذا نیاز به یک روش مناسب جهت تشخیص خودکار دزدان مسلح در ویدئوهای امنیتی در حال ضبط است. در این مقاله روشی سریع و کارا جهت شناسایی موقعیت افراد و سپس تشخیص اسلحه در دست آنها، با استخراج قاب‌های تصاویر برگرفته از ویدئوها و بدون حذف نقاط اصلی، ارائه شده است. در مرحله نخست و به منظور استخراج قاب‌های تصاویر برگرفته از ویدئوها، الگوریتم جداسازی با نرخ قاب مشخص اعمال خواهد شد و تمامی تصاویر در یک پوشه قرار می‌گیرند؛ سپس روی تمامی تصاویر به دست آمده رده‌بندی (HC) Haar Cascade اعمال شده تا نقاط کلیدی یا قاب‌های مربوط به تصاویر کل بدن استخراج شوند و باقی‌مانده‌ها از تصاویر حذف شوند. در انتها، نمونه‌های هر ویدئو در قالب ماتریس چهاربعدی شامل تعداد دنباله قاب‌های هر ویدئو، عرض، ارتفاع و تعداد کانال تصویر به شبکه 3DCNNs ارسال می‌شود تا سلاح در تصاویر شناسایی شوند؛ لذا نوآوری مقاله ترکیب رده‌بندی HC و 3DCNNs به منظور افزایش سرعت و کارایی تشخیص اسلحه است. همچنین به منظور بررسی دقت مدل پیشنهادی، از پارامترهای نرخ مثبت صحیح و مثبت کاذب، مقدار پیش‌بینی مثبت و نرخ تشخیص کاذب استفاده می‌شود.

واژگان کلیدی: شبکه‌های عصبی سه‌بعدی (3DCNN)، رده‌بندی (HC) Haar Cascade، بازشناسی اشیاء و شناسایی کل بدن

## Detection of handgun using 3D convolutional neural network model (3DCNNs)

Sara Motamed\* and Elham Askari

Assistant Professor, Department of Computer Engineering, Fouman and Shaft Branch, Islamic Azad University, Fouman, Iran

### Abstract:

Since the behavior of people in the videos is in the form of three-dimensional signals and it is very difficult to search for a specific action, therefore, there is a need for a suitable technique for the automatic detection of armed robbers in security videos being recorded. In this article, a fast and efficient method is presented to identify the position of people and then recognize the weapon in their hand, by extracting the frames of images taken from the videos and without removing the main points. The database used in this article is called GM Gun Movies and is a standard database derived from videos with 665 images and 640 x 480 size. IMFDB database has also been used to teach the types of weapons in this article. In this article, we have presented an efficient method to prevent and detect armed robbers in images taken from videos. In order to extract the frames of the images and detect the weapon from the video, the separation algorithm with a specific frame rate has been applied to all the samples. Then, the video samples are divided into three categories: training, validation and testing, and

\* Corresponding author

\* نویسنده عهده‌دار مکاتبات



using HC classification, the frames of the whole body images are extracted and the rest of the backgrounds are removed from the images. The reason for choosing this method is that the HC classifier is resistant to rotation of images and also this algorithm has shown a good performance for complex calculations.

Therefore, in our proposed model, we will use this algorithm as a whole body detection. This is done by detecting the ROI region and by cropping the selected regions followed by background subtraction to remove unwanted background. All key points are saved in a folder after selection and extraction. Finally, all images are sent to 3DCNN networks to identify weapons in the images. The reason for choosing 3DCNN networks is that this method leads to more efficient and better results by considering temporal information in dynamic images. As seen in the test results, the highest gun detection rate is related to the 3DCNNs model with a detection rate of 10.96%, followed by the best detection rate of the YOLO V3 model with a detection rate of 60.95%.

**Keywords:** 3D Neural Networks (3DCNNs), Haar-Cascade (HC) Classification, Object Recognition, Full Body Recognition.

ببرند. با بررسی‌های انجام‌شده می‌توان به این نتیجه رسید که برای بازشناسی جرم و جنایات و تشخیص الگوهای جرم، علاوه بر شناسایی سلاح نیاز به بررسی فاکتورهای دیگری مثل ژست انسان، بررسی موقعیت نسبی مجرمان و اندام بدن آن‌ها در تیراندازی بسیار مهم بوده و در تشخیص جرم امری کمک کننده است [۷]؛ اما مشکل زمانی رخ می‌دهد که تصاویر به‌دست‌آمده به‌وسیله دوربین‌های مدار بسته کیفیت بالایی نداشته باشند و با مشکلاتی مثل درجه وضوح پایین، وجود نوفه و یا روشنایی ضعیف همراه باشند. همچنین عوامل دیگری مانند فاصله سلاح و فرد مجرم تا دوربین نیز از مشکلات دیگر این‌گونه تحقیقات هستند [۷]؛ بنابراین در استفاده از مدل‌های داده‌کاوی می‌بایست هوشمندانه عمل کرد تا مشکلاتی نظیر سرریز و نرخ پایین شناسایی وجود نداشته باشد. در این مقاله برای حل چنین مشکلاتی و جهت تشخیص سلاح در دست سارقین و مجرمان از یادگیری عمیق سه‌بعدی استفاده خواهیم کرد. در این مقاله، روشی سریع و کارا جهت تشخیص موقعیت افراد و سپس تشخیص اسلحه در دستشان با استفاده از قاب‌های تصاویر برگرفته از ویدئوها ارائه شده است. هدف انتخاب و استخراج نقاط کلیدی در تصاویر است. در مدل پیشنهادی برای به‌دست‌آوردن نقاط کلیدی، از ترکیب رده‌بند Haar Cascade (HC) و حذف پس‌زمینه‌ها استفاده شده است؛ سپس تمامی تصاویر منتخب به شبکه‌های 3DCNN ارسال شده تا در کوتاه‌ترین زمان ممکن اسلحه بازشناسی شود. دلیل استفاده از مدل پیشنهادی و نوآوری در این مقاله این است که تاکنون در تحقیقات انجام‌شده مدلی کارا و سریع جهت تشخیص موقعیت افراد و بازشناسی اسلحه در دستشان ارائه نشده است؛ همچنین در بیشتر روش‌های تشخیص اشیا بعد رنگی حذف شده که با این کار ویژگی‌های مهمی از تصاویر حذف می‌شوند. از مزایای این شبکه این است که

## ۱- مقدمه

مطالعات جرم‌شناسایی، دزدی و سرقت با بررسی و نظارت ویدئو سال‌هاست که موضوع پژوهش‌ها هستند؛ همچنین سامانه نظارت ویدئوها در همه مکان‌های عمومی و خصوصی قابل بررسی است. برای مثال در فرودگاه‌ها، ایستگاه‌های راه‌آهن، موزه‌ها، بانک‌ها، ساختمان‌های دولتی، نظامی و غیره. داده‌کاوی یکی از مهمترین روش‌هایی است که به‌وسیله آن الگوهای مفید در درون داده‌ها با حداقل دخالت کاربران شناخته می‌شوند و اطلاعاتی را در اختیار کاربران و تحلیل‌گران قرار می‌دهند تا بر اساس آنها تصمیمات مهم و حیاتی در سازمان اتخاذ شود. روش‌ها و الگوریتم‌های داده‌کاوی با ارائه یک الگوریتم یا مدل، داده‌ها را تحلیل و سپس نتایج را تعریف می‌کنند [۱]. چهارچوب داده‌کاوی جرم، روش‌های مختلفی را برای جرم به‌کار گرفته و تحلیل ویژگی انواع جرم می‌تواند به مأموران تحقیق کمک کند تا از روش‌هایی استفاده کنند تا گرایش‌ها و الگوها و نشان‌دهی نواحی را در پیش‌بینی جرم شناسایی کنند. در برخی سازمان‌های پلیسی و قضایی داده‌کاوی را به‌عنوان پیش‌بینی و پیش‌گیری از وقوع جرم استفاده کرده‌اند و بررسی صحنه جرم در بخش داده‌کاوی می‌تواند به کارایی مأموران تحقیق و بررسی صحنه جرم کمک کند [۲]. با استفاده از داده‌کاوی می‌توان به شناسایی شبکه‌های تروریستی، شناسایی تهدیدات درون‌سازمانی و یافتن سرخ‌های بی‌ثباتی در مناطق حساس پرداخت. دلیل اهمیت‌دادن به این موضوعات کم‌کردن جرم و جنایات و کاستن تلفات جانی و مالی است؛ همچنین ایجاد بانک اطلاعات داده‌کاوی به‌عنوان یک نرم‌افزار قدرتمند به‌مراتب کم‌هزینه و کارآمدتر در اختیار سازمان‌های پلیس قرار گرفته است تا با استفاده از داده‌کاوی در آینده بتوانند الگوهای جرم را شناسایی کرده و این‌گونه مشکلات اجتماعی را تا حد امکان کاهش داده یا از بین

پیدا کردن سلاح در سفر مفید هستند، کیسه‌ها یا چمدان‌ها، دامنه این مسائل را بسیار محدود می‌کنند. به‌علاوه این نوع ماشین‌های اسکن بسیار گران‌قیمت هستند. با استفاده از تصاویر رنگی ضبط‌شده به‌وسیله دوربین‌های مدار بسته، سامانه‌های نظارت تصویری برای تشخیص اشیای خطرناک می‌توانند یک گزینه کارا تر و اقتصادی باشند. در همین راستا کارهای متعددی جهت تشخیص سلاح در تصاویر رنگی و روش‌های یادگیری ماشین بیان شده است.

در [۶] یک روش برای تشخیص سلاح در تصاویر رنگی معرفی شد که از بخش‌بندی رنگ و الگوریتم **k-means** برای حذف اشیای غیر مرتبط استفاده می‌شد. در [۷] روش دیگری را برای تشخیص یک تفنگ دستی در یک تصویر و با استفاده از استخراج‌گر تبدیل ویژگی مستقل از مقیاس **SIFT**<sup>1</sup> از مجموعه تصاویر و خوشه‌بندی آن‌ها به‌وسیله الگوریتم **k-means** استفاده کردند. برای کاراتر شدن مدل پیشنهادی روش‌های جدیدی مبتنی بر یادگیری عمیق نیز در مدل‌شان استفاده شد. در این حالت تعداد زیادی از مناطق یا پنجره‌های مختلف در اندازه‌ها و نسبت‌های مختلف در داخل تصویر تولید شده و به رده‌بند شبکه عصبی ارسال شدند. در مطالعات متعددی از این روش جهت تشخیص اسلحه دستی استفاده شده است [۸]. اشکال عمده این نوع سامانه‌ها زمان پردازش بالای مورد نیاز برای رده‌بندی هر یک از این پنجره‌ها و محاسبه اختلاف آن‌ها در زمان واقعی‌شان بود. همچنین روشهای اولیه استفاده‌شده در **CNNs**<sup>2</sup> جهت محاسبه مبتنی بر مناطق<sup>3</sup> استفاده شد. در [۱۱] روشی را بر اساس چارچوب **Faster-RCNN** پیشنهاد کردند که با معماری **VGG-16** به‌عنوان استخراج ویژگی آموزش‌دیده با مجموعه داده‌های **IMFDB** ارائه شد و در [۱۲] برای تشخیص بازو و دست مورد استفاده قرار گرفت. هر دو روش‌های پنجره‌کشویی<sup>4</sup> و **Faster-RCNN** در این مقاله مورد آزمایش قرار گرفتند. **Faster-RCNN** در [۱۳] برای تشخیص تفنگ دستی مورد آزمایش قرار گرفت. تشخیص **Faster-RCNN** پیش آموزش دیده با معماری **VGG-16** بهترین نتایج را در مجموعه داده‌های سفارشی از سه‌هزار تصویر تفنگ یوتیوب داشتند؛ درنهایت، روش‌های دیگری برای تشخیص اشیا بر اساس خانواده یادگیری عمیق و به نام روش **YOLO** نیز ارائه شد [۱۴، ۱۵، ۱۶]. در این معماری‌ها تصاویر به مناطق مختلف ثابت تقسیم

هیچ بعدی از تصاویر حذف نمی‌شود و شبکه‌های **3DCNN** نسبت به تغییر مقیاس و اندازه در تصاویر بسیار مقاوم بوده و حتی در تصاویر نوفه‌دار نیز پاسخ مناسبی را به خروجی ارائه می‌دهند؛ همچنین با بررسی‌های انجام‌شده می‌توان گفت که شبکه‌های **3DCNN** با در نظر گرفتن اطلاعات زمانی در تصاویر پویا منجر به نتایج کارآمدتر و بهتری نسبت به بقیه روش‌های یادگیری عمیق می‌شود.

به‌اختصار سلسله‌مراتب مدل پیشنهادی به‌این‌صورت است که تمامی ویدئوها از پایگاه داده معرفی‌شده در متن خوانده شده و به قاب‌هایی با نرخ ثابت تقسیم‌بندی می‌شود؛ سپس از رده‌بندی **HC** به‌منظور استخراج کل بدن و موقعیت افراد در تصاویر استفاده خواهد شد. تمامی تصاویر کل بدن از هر نمونه ویدئو استخراج شده و پس‌زمینه و اطلاعات اضافی با روش تفریق پس‌زمینه از تصاویر حذف شده و تصاویر به‌دست‌آمده به‌عنوان نقاط کلیدی ذخیره خواهند شد. تمام نقاط کلیدی تصاویر در پوشه‌های مربوط به قاب ویدئو ذخیره می‌شوند. در انتها به‌منظور بازشناسی اسلحه از **3DCNNs** استفاده خواهد شد و به محض بازشناسی اسلحه زنگ خطر به صدا در خواهد آمد.

بخش‌های این مقاله به‌این‌صورت سازماندهی شده است که بخش ۲ مرتبط با کارهای پیشین جهت کشف سلاح، تخمین ژست و بررسی مدل‌های تشخیص است. در بخش ۳ روش پیشنهادی ارائه خواهد شد. نتایج آزمایش‌ها و نتیجه‌گیری در بخش‌های ۴ و ۵ مطرح می‌شوند.

## ۲ - کارهای مرتبط

### ۲-۱- کشف سلاح به‌صورت خودکار

دستگاه‌های اسکن اشعه ایکس در فضاهای عمومی مانند فرودگاه‌ها، ایستگاه‌های قطار و یا موزه‌ها با هدف پیدا کردن سلاح‌های پنهان در چمدان مورد استفاده قرار می‌گیرند. تصاویر تولید شده با اشعه ایکس به صورت دستی توسط یک اپراتور امنیتی تجزیه و تحلیل می‌شوند. در [۳] یک سامانه تشخیص براساس سگمنت‌کردن تصویر و بردارهای ویژگی مبتنی بر لبه استفاده شده است. در [۴] یک روش مبتنی بر ویژگی‌های مشابه **Haar** و رده‌بندی جهت تشخیص سلاح دستی در تصاویر ارائه شده‌است؛ همچنین توصیف‌کننده‌های نقطه سه‌بعدی جهت رده‌بندی شیء در چمدان مورد مطالعه قرار گرفته است [۵]. از آنجایی که سامانه‌های مبتنی بر تصویر برداری اشعه ایکس برای

<sup>1</sup> Scale Invariant Feature Transform

<sup>2</sup> Convolutional Neural Networks

<sup>3</sup> Region-based

<sup>4</sup> sliding window

افراد حاضر در یک تصویر اعمال و سپس یک برآورد ژست تک نفره جهت پیدا کردن ژست دوبعدی از بدن استفاده می‌شود [۲۲، ۲۳]. در این خانواده از روش‌ها، هزینه محاسباتی به شدت بستگی به تعداد افراد و ظاهر شدن آن‌ها در تصویر دارد. از سوی دیگر مدل‌های پایین‌به‌بالا اول نیاز به درخواست یک آشکارساز فرد ندارد، اما نقاط کلیدی دوبعدی بدن برای کل تصویر حائز اهمیت بوده و نیاز به پیش‌بینی دارد. **OpenPose** نمونه‌ای از این نوع رویکردها است [۲۴] و یک برآوردکننده ژست چندنفره است که یک منبع باز است، که مجموعه‌ای از بردارهای دوبعدی را جهت اندازه‌گیری اندام در دامنه تصویر کدگذاری می‌کنند.

برآورد ژست انسان به‌طور گسترده‌ای برای وظیفه تشخیص عمل<sup>۷</sup> مورد استفاده قرار گرفته است. در [۲۵] یک روش برای تشخیص عمل انسان مطرح شده است. در این پژوهش از هیستوگرام توصیفگر گرادین‌گرا (**HOG**<sup>۸</sup>) برای تشخیص اولیه ژست استفاده شده، سپس اعمال فرد با مقایسه هیستوگرام تولیدشده پیش‌بینی می‌شوند. در مقاله [۲۶]، ترکیب رده‌بندی مبتنی بر سیگنال با ویژگی‌های مبتنی بر مدل ارائه شده است. در این مدل جهت تشخیص بالاتنه ژست از مدل‌های یادگیری عمیق استفاده شده است؛ سپس به‌منظور رده‌بندی حرکت یا عمل از (**LSTM**<sup>۹</sup>) استفاده کرده‌اند [۲۷]. ترکیب بین ظاهر بصری در یک تصویر و ژست انسان برای تشخیص عمل در [۲۸] مورد مطالعه قرار گرفته است. در [۲۹] معماری مبتنی بر **CNN** پیشنهاد شده و به‌طور مشترک جهت برآورد و تشخیص ظاهر و ترکیب آن‌ها به‌منظور پیش‌بینی در امر بازشناسی استفاده شده است. در بیش‌تر این روش‌ها ابتدا طبقه‌بندی را آموزش داده و سپس در طول فرایند تشخیص آن را بر روی تعدادی از مناطق از تصویر ورودی اجرا و با استفاده از هر دو رویکرد پنجره کشویی و یا رویکرد پیشنهادی منطقه اعمال می‌کنند. در [۱] با استفاده از (**HOG**) به استخراج ویژگی و پیش‌بینی رده‌ی شیء در هر پنجره پرداخته شده است. نخستین مدل تشخیص در [۱] **CNN** مبتنی بر منطقه (**RCNN**) است [۵] که حدود دوهزار جعبه محدودکننده با استفاده از روش جستجوی انتخابی تولید می‌کرد [۲۴]. امتیازات جعبه‌ها با استفاده از ماشین بردار پشتیبان (**SVM**<sup>۱۰</sup>)، با جعبه‌های محدودشده و با استفاده از یک مدل خطی تنظیم‌شده محاسبه می‌شدند.

شده و احتمالات و جعبه‌های محدودکننده<sup>۱</sup> برای هر منطقه پیش‌بینی شده‌اند [۱۷].

در مقاله [۱۸] به اهمیت سامانه‌های نظارتی که قادر به ردیابی مسیرهای افراد است، پرداخته شده است؛ اما مشکل اصلی تشخیص اشیایی است که توسط افراد حمل خواهد شد. با توجه به مسائل امنیتی و اهمیت این موضوع، در این مقاله روش پیشرفته مبتنی بر شبکه‌های عصبی ساختاری بازگشتی (**SRNN**<sup>۲</sup>) ارائه شده است. نرخ تشخیص اسلحه در دست افراد، با استفاده از روش پیشنهادی‌شان به‌طور میانگین ۸۷.۳ درصد بوده است. در [۱۹] روشی جدید جهت تشخیص اسلحه معرفی نشد، اما سامانه‌ای کارا جهت استفاده از الگوریتم‌های موجود بهینه ارائه شد تا مورد استفاده در سرورهای ابری<sup>۳</sup> و دستگاه لبه<sup>۴</sup> باشد.

همچنین در مقاله [۲۰] از مدل تشخیص شیء **YOLO V3** جهت شناسایی اسلحه استفاده شده و با مدل‌های **YOLO V2** و شبکه عصبی کانولوشنال مقایسه شد. طبق بررسی‌های انجام‌شده در این مقاله به این نتیجه رسیدند که **YOLO V3** جواب بهتری را نسبت به مدل‌های رقیب ارائه داده است. از طرفی دیگر در مقاله [۲۱] از متدهای یادگیری عمیق و با در نظر گرفتن اطلاعات وضعیت و موقعیت افراد و تشخیص تفنگ دستی استفاده شده است. شبکه پیشنهادی در آن مقاله نتایج موفقیت‌آمیزی، با بیشینه بهبود ۱۷.۵٪ را جهت بررسی موقعیت افراد و سپس تشخیص تفنگ دستی با استفاده از روش دقت متوسط<sup>۵</sup> ارائه داده است.

از آنجایی که روش یادگیری عمیق روی روش‌های بازشناسی اشیا بسیار کارا عمل کرده و جزئیات ریز تصاویر با کمک این روش‌ها قابل شناسایی است، در این مقاله نیز یکی از روش‌های یادگیری عمیق را با عنوان شبکه عصبی کانولوشنال سه‌بعدی به‌منظور تشخیص اسلحه در دست سارقان و مجرمان در تصاویر استخراج‌شده از ویدئوهای ضبط‌شده امنیتی ارائه خواهیم کرد.

## ۲-۲- تخمین ژست<sup>۶</sup> انسان

در ادبیات، روش‌های تخمین ژست انسان به دو گروه اصلی تقسیم می‌شوند. مجموعه نخست برای برآورد ژست چندنفره به‌عنوان روش‌های بالابه‌پایین شناخته می‌شوند. در این حالت یک آشکارساز شخص برای تشخیص تمام

<sup>1</sup> bounding boxes

<sup>2</sup> Structural Recurrent Neural Network

<sup>3</sup> Cloud Server

<sup>4</sup> Edge Device

<sup>5</sup> Average Precision

<sup>6</sup> pose

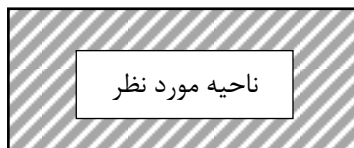
<sup>7</sup> action recognition

<sup>8</sup>The Histogram of Oriented Gradients

<sup>9</sup> Long Short Term Memory

<sup>10</sup> Support Vector Machine

$(x1, y1)$



$(x1 + w, y1 + h)$

(شکل-۲): موقعیت بدن افراد به عنوان ناحیه مورد نظر  
(Fig.2): Position of the body as the desired area

برای پیدا کردن مختصات ROI و برش تصویر کل بدن افراد در تصاویر گرفته شده، شکل (۲) به تصویر کشیده شده است. الگوریتم HC قادر به شناسایی یک یا چند منطقه ROI است. این کار از طریق تجزیه و تحلیل اختلاف فاصله برای هر ROI تولید شده از قبل و مطلوب-ترین ROI محاسبه می‌شود. مرحله بعدی بریدن یا جدا کردن مناطق کلیدی از مناطق بی‌اهمیت یا پس‌زمینه است. در انتها تمامی تصاویر با جزئیات مهم به شبکه عصبی کانولوشنال سه بعدی ارسال شده تا اسلحه در این تصاویر تشخیص داده و بازشناسی شوند [۳۰].

بر اساس شکل (۲)،  $w$  و  $h$  واحدهای گسترش داده از مختصات مناطق مورد نظر هستند که به‌طور خودکار برای هر فرد واقع در ناحیه ROI شناسایی و بریده می‌شود [۳۰].

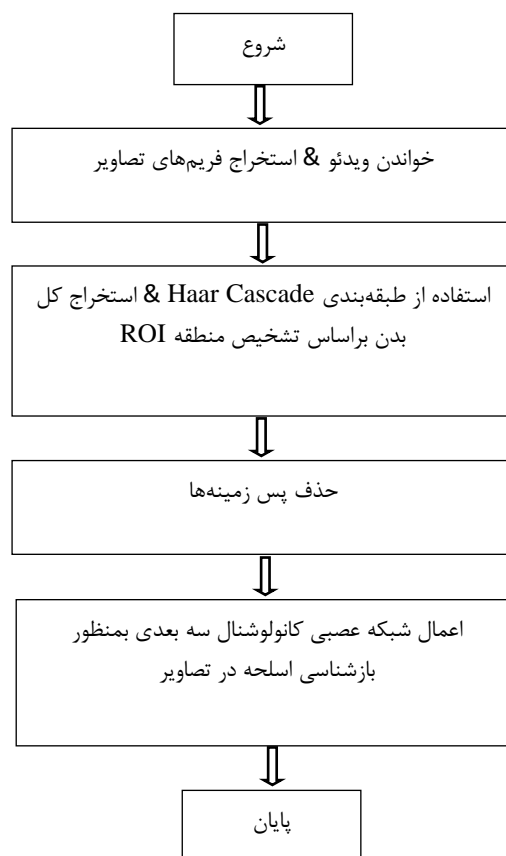
### ۳-۱ - پایگاه داده

بیشتر پایگاه داده‌های مربوط به تصاویر دوربین مداربسته در مجموعه داده‌های بازشناسی سلاح یا بررسی سرقت، جرم و جنایات به اندازه کافی واقع بینانه نیستند و نشان دادن ژست‌های غیر طبیعی برای حملات تفنگ دستی یا تیراندازی دسته‌جمعی به‌طور کامل ساختگی است؛ اما مطلب مهم در انتخاب پایگاه داده، این است که تصاویر برگرفته از ویدئوهای ضبط شده دارای کیفیت و وضوح مناسب باشند و همچنین ویدئوهای افراد از زمان ورود به یک مکان و تا زمان خروج از مکان مورد نظر، در دسترس باشند. پایگاه داده‌ای که در این مقاله مورد استفاده قرار گرفته، به نام فیلم‌های تفنگ (GM<sup>3</sup>) است [۳۱]. GM یک پایگاه داده استاندارد و برگرفته از ویدئوهایی با ۶۶۵ تصویر و با اندازه  $۶۴۰ \times ۴۸۰$  است. در این کلیپ‌ها یک مرد با ننگ داشتن یک تفنگ دستی در چند ژست تیراندازی و در یک اتاق سرپوشیده حرکت

<sup>2</sup> CCTV  
<sup>3</sup> Gun Movies

### ۳ - روش پیشنهادی

همانطور که در شکل (۱) نشان داده می‌شود، موقعیت کلیدی یا همان شناسایی سارقان در تصاویر به وسیله رده‌بندی کننده HC استخراج می‌شود. رده‌بندی کننده HC با مکان‌یابی مختصات  $x$  و  $y$ ، منطقه مورد علاقه (ROI<sup>1</sup>) را که در این مقاله کل بدن افراد یا فرد در تصاویر است و جزء مناطق کلیدی در تصاویر است، استخراج می‌کند. در اینجا، روی تصاویر مثبت و منفی، رده‌بندی کننده HC آموزش داده می‌شود. تصاویر مثبت شامل افراد در هنگام ورود و داشتن شیء مشکوک در دست‌شان است و تصاویر منفی شامل تصاویری از مکان مورد نظر (بدون انسان) و پس‌زمینه است. رده‌بندی کننده HC باید روی تصاویری با تعداد مناسب آموزش داده شود. تعداد تصاویر گرفته شده از پایگاه داده معرفی شده در متن ۶۶۵ است که جهت آموزش به رده‌بندی کننده HC ارسال می‌شود. تصاویر منفی به عنوان نمونه‌های پس‌زمینه است که باید با عمل تفریق از بخش تصاویر مثبت حذف شوند. تمامی تصاویر در یک پوشه ذخیره شده و برای تشخیص اسلحه به شبکه عصبی سه بعدی ارسال می‌شوند.



(شکل-۱): سلسله مراتب مدل پیشنهادی  
(Fig-1): Hierarchy of proposed model

<sup>1</sup> Region of Interest

در فاز آزمایش، ویدئوها و نمونه‌های یادگیری آماده شده بدین صورت است که یک مرد از جلوی دوربین قدم زده و با خود اسلحه حمل می‌کند. این اهداف از زوایای مختلف و فواصل متفاوت از دوربین گرفته شده است.

برای مجموعه تست<sup>۵</sup> و آزمایش<sup>۶</sup> نمونه‌های مثبت (قاب‌هایی با شخص مسلح) و نمونه‌های منفی (قاب‌هایی بدون شخص مسلح) انتخاب شده است.

در ادامه مرحله پیش‌پردازش، تمامی تصاویر به وضوح  $3 \times 240 \times 240$  تغییر کرده و به الگوریتم HC ارسال می‌شوند.

### ۳-۳- الگوریتم Haar Cascade (HC)

HC یکی از موثرترین الگوریتم‌های تشخیص شی مبتنی بر ویژگی است. از این روش بیشتر برای تشخیص چهره، تشخیص راه رفتن و تشخیص حرکت انسان استفاده می‌شود. در [۳۱] جهت تشخیص چهره در زمان واقعی<sup>۷</sup> از رده‌بند AdaBoost با HC و به‌منظور استخراج ویژگی از تجزیه مؤلفه اصلی (PCA<sup>۸</sup>) و آنالیز تشخیصی خطی (LDA<sup>۹</sup>) استفاده شده است. در [۳۳] به‌منظور بازشناسی صورت سلسله‌مراتبی<sup>۱۰</sup> و تشخیص چشم از HC به‌عنوان رده‌بندی‌کننده استفاده شده است.

طبق پژوهش‌های انجام‌شده HC نسبت به چرخش تصاویر مقاوم بوده و همچنین این الگوریتم نسبت به محاسبات پیچیده عملکرد خوبی را نشان داده است [۳۴]. علاوه بر این، الگوریتم‌هایی که از HC استفاده کرده‌اند موجب بهبود عملکرد روش‌های یادگیری ماشین هم در زمینه تشخیص شیء بصری و هم در زمینه بازشناسی شده‌اند. همان‌طور که در [۳۵] توضیح داده شده HC برای پایگاه داده‌های بزرگ از تصاویر مناسب است؛ بنابراین در مدل پیشنهادی ارائه‌شده در این مقاله از این الگوریتم به‌عنوان تشخیص کل بدن استفاده خواهیم کرد. نکته‌ای که لازم است در این بخش توضیح داده شود، این است که در بین کلیپ‌های ضبط‌شده یک شخص با نگاه داشتن یک تفنگ دستی در چند ژست تیراندازی و در یک اتاق سرپوشیده حرکت می‌کند. در این مقاله از الگوریتم HC بالاتنه و پایین‌تنه استفاده شده و این الگوریتم قادر به تشخیص بیش از یک نفر در تصاویر است.

می‌کند. فاصله دوربین، وضوح تصویر، شرایط روشنایی و سناریوهای دوربین مداربسته خوب نمایش داده شده است. در شکل (۳) دو مثال از تصاویر پایگاه داده GM نشان داده شده است [۳۱]. همچنین برای آموزش انواع سلاح در این مقاله، از پایگاه داده IMFDB نیز استفاده شده است. در این پایگاه داده تصاویر و ویدئوهای زیادی مربوط به موضوع پیشنهادی مان وجود دارد. تصاویر منفی نیز در این پایگاه داده موجود است که اسلحه نبوده و شامل سشوار، مته و غیره است [۳۲]. IMFDB شامل چهل هزار تصاویر تفنگ<sup>۱</sup> و هفت تیر<sup>۲</sup>، تفنگ ساچمه‌ای<sup>۳</sup> و غیره است. در این مقاله تنها از پانصد تصویر هفت تیر و تفنگ و تصاویر منفی استفاده شده است؛ همچنین برای تشخیص تفنگ‌های فلزی<sup>۴</sup> از بردارهای ویژگی مبتنی بر لبه و بخش‌بندی استفاده شده است.



(شکل-۳): نمونه‌ای از تصاویر پایگاه داده GM  
(Fig-3): Samples from GM dataset images

### ۳-۲- پیش پردازش

برای بازشناسی سلاح در دست، در مرحله نخست، آماده‌سازی پایگاه داده و عملیات پیش‌پردازش انجام می‌شوند. این مراحل به صورت زیر دسته‌بندی شده است: در مرحله آماده‌سازی پایگاه داده ابتدا، مجموعه پایگاه داده به سه بخش آموزش، آزمایش و اعتبارسنجی تقسیم می‌شوند. در بخش آموزش به تصاویر منفی (شامل سشوار، مته و غیره) برچسب صفر، به تصاویر تفنگ برچسب ۱ و به تصاویر مربوط به هفت تیر برچسب ۲ داده می‌شود.

در مرحله بعد، برای استخراج قاب‌های تصاویر از ویدئو الگوریتم جداسازی با نرخ قاب مشخص بر روی همه نمونه‌ها اعمال خواهد شد. در این آزمایش‌ها نرخ قاب تصاویر ۱۰ (یعنی در هر ثانیه ۱۰ قاب را استخراج می‌کند) است.

<sup>5</sup> Test

<sup>6</sup> Train

<sup>7</sup> Real Time

<sup>8</sup> Principal Component Analysis

<sup>9</sup> Linear Discriminant Analysis

<sup>10</sup> Hierarchical Face

<sup>1</sup> Rifles

<sup>2</sup> Revolves

<sup>3</sup> Shotgun

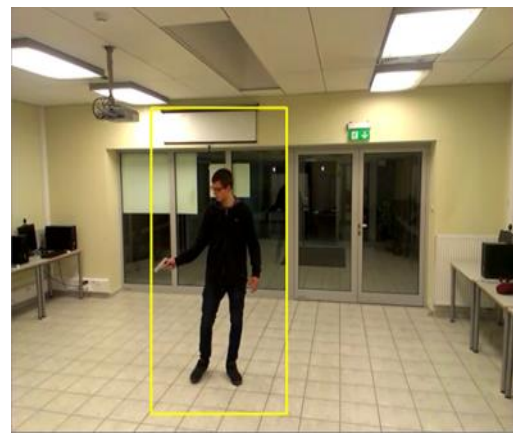
<sup>4</sup> Metallic Pistols

اطلاعات زمانی را از دست می‌دهد. اما در تصاویر وسط و سمت راست، کانولوشن سه بعدی نمایش داده شده است. همان‌طور که مشاهده می‌شود، اطلاعات زمانی در هر دو شکل حفظ می‌شوند و خروجی آنها به صورت سه بعدی است [۳۶].

این شبکه یک مدل از یادگیری باناظر است و معماری آن شامل دو بخش است، یک شبکه رمزگشایی سه بعدی و به دنبال آن بلوک رده بندی لایه‌های کانولوشنال **Max Pooling** است. رمزگشایی اجازه می‌دهد تا یادگیری با چند مقیاس پیچیده و با حجم سه بعدی صورت گیرد. همچنین، اتصالات پرش اجازه می‌دهد تا یادگیری با ویژگی‌هایی با وضوح بالا و پاسخی با وضوح بالاتر و تأثیرات چند مقیاسی انجام شود. همچنین خروجی ترکیب لایه‌های مخلوط از دو بلوک و به کمینه‌رساندن تابع هدف چندمقیاسی است.

بخش دوم مدل پیشنهادی بلوک رده بندی، از سه واحد **Convolution Pooling** سه بعدی، شش لایه کانولوشنال و سه لایه **Max Pooling**، یک لایه پهن شده<sup>۱</sup>، و دو لایه به طور کامل لایه‌های متصل تشکیل شده است. لایه‌های خروجی مرحله‌ی قبل به شکل بلوک در آمده و این بلوک اجازه می‌دهد تا یک تراکنشی از نمایش سه بعدی با پهن کردن داده‌ها<sup>۲</sup> ارائه شود. همین امر موجب می‌شود تا یادگیری مدل از لایه‌های چندمقیاسی و با استفاده از معماری پیچیده انجام شود [۳۷]. برای شش لایه کانولوشنال آخر، ما از یک پارامتر اصلاح واحد خطی (**PreLU**) به عنوان تابع فعال-ساز، که از تابع **ReLU** تولید شده است، استفاده می‌کنیم [۳۷]. این تابع فعال‌ساز به طور تطبیقی پارامترهای یک اصلاح کننده را یاد می‌گیرد و دقت را افزایش می‌دهد [۳۷].

ورودی **ConvNets** اندازه ثابت ۲۴۰ در ۲۴۰ رنگی در طول آموزش است. همچنین در بخش آموزش شبکه و برای تعیین ابر پارامترهای شبکه از روش سعی و خطا و تابع کراس آن‌تروپی و بهینه کننده آدام با نرخ یادگیری ۰.۰۰۱ استفاده شده است. تعداد کل اپک‌های<sup>۴</sup> به کار رفته در مدل پیشنهادی ۱۵۰ بوده و از روش **Fold** ۱۰ برای اعتبارسنجی متقابل<sup>۵</sup> داده‌ها استفاده شده است (جدول ۱).



(شکل ۴): تصاویر استخراج شده از کل بدن

به وسیله منطقه ROI

(Fig-4): Images extracted from the whole body by ROI area

یکی از مهمترین بخشهای مدل پیشنهادی یافتن نقاط کلیدی و حذف پس زمینه از بخشهای انتخاب شده است؛ سپس روی بخش‌های انتخاب شده از تصاویر شبکه‌های عصبی سه بعدی اعمال می‌شود. دلیل انتخاب رده بندی **3DCNN** این است که این رده بند قادر به ذخیره اطلاعات در کل شبکه است و توانایی کار با دانش ناکامل را داشته و همچنین توانایی تحمل خطا بالایی را دارد؛ لذا انتظار می‌رود که این رده بند نتایج مطلوبی را نمایش دهد.

### ۳-۴- شبکه‌های عصبی کانولوشن سه بعدی (3DCNN)

شبکه عصبی کانولوشن سه بعدی دارای فیلتر سه بعدی است که به تصویر اعمال می‌شود و آن را **3DC** یا **3DCNNs** می‌نامند. در شبکه عصبی کانولوشن دو بعدی، عملیات کانال و ادغام بر روی تصاویر دو بعدی به صورت مکانی اعمال می‌شود در حالیکه در شبکه عصبی کانولوشن سه بعدی این عملیات با در نظر گرفتن زمان و مکان در تصاویر پویا اعمال می‌شوند.

از این رو در شبکه‌های **3DCNN** با در نظر گرفتن اطلاعات زمانی در تصاویر پویا منجر به نتایج کارآمدتر و بهتری می‌شوند. می‌شوند. شکل (۵)، فیلترگذاری روی شبکه‌های دو بعدی و سه بعدی را نمایش می‌دهد. با توجه به تصویر سمت چپ، کانولوشن دو بعدی بر روی چندین تصویر (بطور مثال بر روی کانال‌های مختلف تصویر در تصاویر رنگی) اعمال می‌شود. در نتیجه خروجی آنها نیز یک تصویر دو بعدی است. لذا کانولوشن دو بعدی

1 Flatten Layer  
2 Fully Connected  
3 Flatten Data  
4 Epochs  
5 Cross Validation

(جدول - ۱): پارامترهای استفاده شده در 3DCNN

(Table - 1): Parameters Used in 3DCNN

Method	Layers	Optimizer	LR	Batch size
3DCNN	35	Adam	0.001	64

(جدول - ۲): نتایج تشخیص اسلحه با استفاده از HC و رده‌بندی‌کننده‌های مختلف

(Table -2): Results of gun detection using HC and different Classifiers

Algorithm	Time (S)	Accuracy (%)	TRP	FPR	PRV	FDR
3DCNNs	~ 40 sec	96.1	93	6	99	1
YOLO V3	~ 30 sec	95.6	91	4	99	1
RCNN	~ 48 sec	94.8	90	7	99	1
CNN	~ 59 sec	93.5	89	7	98	2
Linear SVM	~ 66 sec	90	88	8	94	6
KNN	~ 71 sec	86.2	89	17	87	13

تشخیص تفنگ در تصاویر استفاده کرده‌ایم. نتایج آزمایش‌ها در جداول (۲) و (۳) نشان داده شده است. همان‌طور که در جدول (۲) مشاهده می‌شود و نتایج این جدول مربوط به انتخاب ROI مناسب با استفاده از الگوریتم HC و سپس استفاده از پنج رده‌بندی‌کننده‌های منتخب در این مقاله است، بالاترین نرخ تشخیص اسلحه مربوط به شبکه 3DCNNs و با نرخ تشخیص ۹۶٪.۱۰ و زمان اجرای چهار ثانیه است. از طرفی کم‌ترین زمان اجرای تشخیص اسلحه مربوط به روش YOLO V3 است و دقت تشخیص این روش به‌طور تقریبی مشابه روش پیشنهادی است؛ بنابراین می‌توان نتیجه گرفت که YOLO V3 سرعت بالایی در تشخیص دارد، اما در تعیین دقیق اشیای کوچک بهتر از مدل پیشنهادی عمل نمی‌کند.

همچنین زمان اجرای مدل پیشنهادی نسبت به مدل‌های رقیب CNN، SVM خطی و KNN کمتر بوده و دلیل آن نیز انتخاب مؤثرترین ویژگی‌ها از تصاویر برگرفته از ویدئوها و عملکرد شبکه 3DCNN است.

در سری دوم آزمایش‌ها، به‌منظور بررسی عملکرد HC از این رده‌بند استفاده نمی‌شود؛ بنابراین بعد از پیش‌پردازش و تغییر اندازه تصاویر، تمام تصاویر

#### ۴ - تشریح و بحث نتایج

همان‌طور که در بخش قبل توضیح داده شد، پس از خواندن ویدئو و تقسیم‌بندی تصاویر به دسته‌های آموزش و تست با استفاده از رده‌بندی HC موقعیت افراد در تصاویر به دست می‌آید و پس‌زمینه از بخش ROI حذف خواهد شد. در انتها بر روی تمامی بخش‌های مورد نظر و به‌منظور تشخیص اسلحه، 3DCNN اجرا می‌شود.

عملکرد سامانه از نظر دقت و با پارامترهای نرخ مثبت صحیح ( $TPR^1$ )، نرخ مثبت کاذب ( $FPR^2$ )، مقدار پیش‌بینی مثبت ( $PPV^3$ ) و نرخ تشخیص کاذب ( $FDR^4$ ) ارزیابی می‌شود.

با بررسی آزمایش‌های متعدد دریافتیم که روش 3DCNN برای شرایط مختلف مانند پس‌زمینه‌های مختلف به‌خوبی کار می‌کند. ما از پنج رده‌بندی‌کننده مختلف دیگر به نام‌های CNN، RCNN، YOLO V3، SVM خطی و K نزدیک‌ترین همسایگی ( $KNN^5$ )، به‌عنوان مدل‌های رقیب جهت رده‌بندی و بررسی نرخ

<sup>1</sup> True Positive Rate

<sup>2</sup> False Positive Rate

<sup>3</sup> Positive Prediction Value

<sup>4</sup> False Detection Rate

<sup>5</sup> K-Nearest Neighbors



نرمال‌سازی شده PCA اعمال شده تا با عمل کاهش ابعاد در بردارها، ویژگی‌های خوب تصاویر استخراج شده و روند اجرا سریعتر صورت گیرد. در انتها روی تمامی بردارهای ویژگی استخراج شده رده‌بندهای CNN، RCNN، SVM و KNN اعمال شده است. در این آزمایش بهترین نرخ تشخیص مربوط است به روش PCA و RCNN و با نرخ تشخیص ۹۰.۸۰٪ است. نتایج این آزمایشات در جدول (۴) نمایش داده شده است.

### ۵- نتیجه‌گیری

در این مقاله ما یک روش کارا جهت پیش‌گیری و تشخیص دزدان مسلح در تصاویر گرفته‌شده از ویدئوها ارائه داده‌ایم. برای استخراج قاب‌های تصاویر و تشخیص سلاح از ویدئو، الگوریتم جداسازی با نرخ قاب مشخص بر روی همه نمونه‌ها اعمال شده است؛ سپس نمونه‌های ویدئو به سه دسته آموزش، اعتبارسنجی و آزمایش تقسیم می‌شوند و با استفاده از رده‌بندی HC قاب‌های تصاویر کل بدن استخراج شده و باقی‌پس‌زمینه‌ها از تصاویر حذف می‌شوند. دلیل انتخاب این روش این است که رده‌بند HC نسبت به چرخش تصاویر مقاوم بوده و همچنین این الگوریتم نسبت به محاسبات پیچیده عملکرد خوبی را نشان داده است؛ بنابراین در مدل پیشنهادی‌مان از این الگوریتم بعنوان تشخیص کل بدن استفاده خواهیم کرد. این کار با تشخیص منطقه ROI و با بریدن نواحی انتخاب شده و به‌دنبال آن تفریق زمینه برای از بین بردن پس‌زمینه ناخواسته انجام می‌شود. تمامی نقاط کلیدی پس از انتخاب و استخراج در داخل یک پوشه ذخیره می‌شوند. در انتها، تمامی تصاویر به شبکه‌های 3DCNN ارسال می‌شوند تا سلاح در تصاویر شناسایی شوند. دلیل انتخاب شبکه‌های 3DCNN این است که این روش با در نظر گرفتن اطلاعات زمانی در تصاویر پویا منجر به نتایج کارآمدتر و بهتری می‌شود. همان‌طور که در نتایج آزمایش‌ها دیده می‌شود، بالاترین نرخ تشخیص اسلحه مربوط به مدل 3DCNNs و با نرخ تشخیص ۹۶.۱۰٪ است و بعد از آن بهترین نرخ مدل تشخیص مربوط به YOLO V3 و با نرخ تشخیص ۹۵.۶۰٪ است.

نرمال‌سازی شده یک‌بار به 3DCNN و بار دیگر به RCNN، YOLO V3 و CNN ارسال می‌شوند. همان‌طور که در جدول (۳) نشان داده می‌شود، نرخ تشخیص اسلحه در مدل پیشنهادی برابر با ۹۳.۱۰٪ است و نسبت به سایر روش‌ها بهتر جواب داده است.

با مشاهده نتایج جدول (۳) به این نتیجه می‌رسیم که سلسله‌مراتب مدل پیشنهادی جهت پیدا کردن موقعیت افراد و استخراج نقاط کلیدی و سپس تشخیص اسلحه، عملکرد بهتری را نسبت به سایر روش‌های رقیب دارد؛ همچنین نرخ تشخیص با در نظر گرفتن نقاط کلیدی بدون در نظر گرفتن نقاط کلیدی نیز مشخص می‌کند که سلسله‌مراتب مدل پیشنهادی موجب بهبود نرخ تشخیص می‌شود. به همین دلیل می‌توان ادعا کرد که مدل پیشنهادی موجب بهبود عملکرد تشخیص اسلحه در تصاویر شده است.

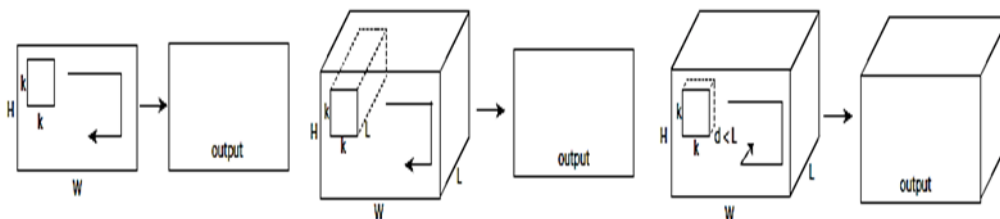
(جدول ۳): نتایج تشخیص اسلحه بدون استفاده از HC  
(Table - 3): Gun Detection Results without Using HC

Training	Accuracy (%)
3DCNN (RGB images)	87.70
3DCNN (Background Subtraction in RGB images)	93.10
YOLO V3	91.70
RCNN	89.30
CNN	87.50

(جدول ۴): نتایج تشخیص اسلحه با استفاده از استخراج ویژگی PCA رده‌بندی کننده‌های (RCNN, CNN, SVM, KNN)  
(Table - 4): Gun Detection Results Using PCA Feature Extraction and Classifiers (CNN, SVM, KNN, RCNN)

Training	Accuracy (%)
PCA + RCNN	90.80
PCA + CNN	89.01
PCA+SVM	88.25
PCA+KNN	84.64

در آزمایش چهارم، بمنظور بررسی مدل پیشنهادی پس از پیش‌پردازش، روی تمامی تصاویر



(شکل-۵): نمایش فیلتر شبکه عصبی کانولوشن دو بعدی (چپ)، شبکه عصبی کانولوشن سه بعدی تصاویر رنگی (وسط) و شبکه عصبی سه بعدی برای ویدئو (راست) [۳۶]

(Fig-5): 2D convolutional neural network filter display (left), 3D convolutional neural network of color images (middle) and 3D neural network for video (right)

[9] Girshick, R. "Fast R-CNN", Proceedings of the IEEE international conference on computer vision. pp. 1440 - 1448. 2015.

[10] Ren, S., He, K., Girshick, R., Sun, J. "Faster R-CNN: Towards real-time object detection with region proposal networks", Advances in Neural Information Processing Systems. pp. 91 - 99. 2015.

[11] Verma, G. K., Dhillon, A. "A handheld gun detection using faster R-CNN deep learning", Proceedings of the 7th International Conference on Computer and Communication Technology. pp. 84 - 88. 2017.

[12] IMFDB: Internet Movie Firearms Database, [http://www.imfdb.org/wiki/Main\\_Page](http://www.imfdb.org/wiki/Main_Page), 2020.

[13] Olmos, R., Tabik, S., Herrera, F. "Automatic handgun detection alarm in videos using deep learning", Neurocomputing. Vol. 275, pp. 66 - 72. 2018.

[14] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. "You only look once: Unified, real-time object detection", Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779 - 788. 2016.

[15] Redmon, J., Farhadi, A. "Yolo9000: better, faster, stronger", Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7263 - 7271. 2017.

[16] Farhadi, A., Redmon, J. "Yolov3: An incremental improvement", Computer Vision and Pattern Recognition. 2018.

[17] de Azevedo Kanehisa, R. F., de Almeida Neto, A. "Firearm detection using convolutional neural networks", ICAART. Vol. 2, pp. 707 - 714. 2019.

[18] Susarla, P., Agrawal, U., Jayagopi, D. B. "Human weapon-Activity recognition in surveillance videos using structural-RNN," MedPRAI '18, Rabat, Morocco, pp.101-108. 2018.

[19] Qi., D, Tan., W, Liu., Z, Yao., Q, Liu., J, "A dataset and system for real-time gun detection in surveillance video using deep learning," IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2021.

## 6- References

## ۶- مراجع

[1] Enrquez, F., Soria, L. M., Alvarez-Garca, J. A., Caparrini, F. S., Velasco, F., Deniz, O., Valez, N. "Vision and crowdsensing technology for an optimal response in physical-security", International Conference on Computational Science, Springer, pp. 15 - 26, 2019.

[2] Tessler, R. A., Mooney, S. J., Witt, C. E., O'Connell, K., Jenness, J., Vavilala, M. S., Rivara, F. P. "Use of rearms in terrorist attacks: differences between the United States, Canada, Europe, Australia, and New Zealand", JAMA internal medicine. Vol. 177, pp.1865 - 1868. 2017.

[3] Nercessian, S., Panetta, K., Agaian, S. "Automatic detection of potential threat objects in x-ray luggage scan images", IEEE Conference on Technologies for Homeland Security, IEEE. pp. 504 - 509. 2008.

[4] Xiao, Z., Lu, X., Yan, J., Wu, L., Ren, L. "Automatic detection of concealed pistols using passive millimeter wave imaging", IEEE International Conference on Imaging Systems and Techniques (IST), IEEE. pp. 1 - 4. 2015.

[5] Flitton, G., Breckon, T. P., Megherbi, N. "A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery", Pattern Recognition. 2420 - 2436. 2013.

[6] Tiwari, R. K., Verma, G. K. "A computer vision based framework for visual gun detection using Harris interest point detector", Procedia Computer Science. Vol. 54, pp. 703 - 712. 2015.

[7] Halima, N. B., Hosam, O. "Bag of words based surveillance system using support vector machines", Int. J. Secur. Appl. Vol. 10, pp. 331 - 346. 2016.

[8] Gelana, F., Yadav, A. "Firearm detection from surveillance cameras using image processing and machine learning techniques", Smart Innovations in Communication and Computational Sciences, Springer. pp. 25- 34. 2019.

- [34] Lienhart, R., Kuranov, A., Pisarevsky, V., Report, M. R. L. T. "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection." in Joint Pattern Recognition Symposium, pp. 297–304. 2003.
- [35] Paulo Menezes, J., Carlos Barreto, J. "Face Tracking Based On Haar-Like Features And Eigenfaces," in IFAC/EURON Symposium on Intelligent Autonomous Vehicles , pp. 1–6, 2004.
- [36] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., Zitnick, C. L. "Microsoft COCO: Common Objects in Context", European conference on computer vision, Springer. pp. 740 - 755. 2014.
- [37] Almaadeed, N., Elharrouss, O., Q'AlMaadeed, S., Bouridane, A., Beghdadi, A. "A Novel Approach for Robust Multi Human Action Recognition and Summarization based on 3D Convolutional Neural Networks", Computer Vision and Pattern Recognition, pp. 1-12. 2021.
- [20] Narejo, S., Pandey, B., Esenarro vargas, D., Rodriguez, R., Anjum, M.R. "Weapon detection using YOLOV3 for smart surveillance system," Mathematical Problems in Engineering, pp. 1-9. 2021.
- [21] Velasco-Mata., A, Ruiz-Santaquiteria., J, Valez., N, Deniz., O, "Using human pose information for handgun detection," Neural Computing and Applications, 33: pp.17273–17286. 2021.
- [22] Pishchulin, L., Jain, A., Andriluka, M., Thormahlen, T., Schiele, B. "Articulated people detection and pose estimation: Reshaping the future", IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 3178 - 3185. 2012.
- [23] Gkioxari, G., Hariharan, B., Girshick, R., Malik, J. "Using k-pose lets for detecting people and localizing their key points", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3582 - 3589. 2014.
- [24] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., Sheikh, Y. A. "OpenPose: real time multi-person 2D pose estimation using Part Anntity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 43, pp.172 - 186. 2019.
- [25] Thureau, C., Hlavac, V. "Pose primitive based human action recognition in videos or still images", in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 1 - 8. 2008.
- [26] Reiss, A., Hendeby, G., Bleser, G., Stricker, D. "Activity recognition using biomechanical model based pose estimation", European Conference on Smart Sensing and Context, Springer. pp. 42 - 55. 2010.
- [27] Eiert, S. "Activity Recognition from 2D pose using an LSTM RNN", 2020.
- [28] Luvizon, D. C., Picard, D., Tabia, H. "2D/3D pose estimation and action recognition using multitask deep learning", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5137 - 5146. 2018.
- [29] Velasco-Mata, A. "Human pose information as an improvement factor for handgun detection", Master's thesis, Escuela Superior de Informatica (UCLM), 2020.
- [30] Puad, A., Tahir2, M. D. "Human Gait Silhouettes Extraction Using Haar Cascade Classifier on OpenCV", International Conference on Modelling & Simulation. Vol. 25, pp. 105 – 111. 2017.
- [31] Grega, M., Matiolański, A., Guzik, P., Leszczuk, M. "Automated detection of firearms and knives in a CCTV image", Sensors. Vol. 16. 2017.
- [32] [http://www.imfdb.org/wiki/Main\\_Page](http://www.imfdb.org/wiki/Main_Page).
- [33] Schmidt, A., Kasiński, A. "The performance of the haar cascade classifiers applied to the face and eyes detection," in Computer Recognition Systems 2, vol. 45, pp. 816–823, 2007.



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

[motamed.sarah@gmail.com](mailto:motamed.sarah@gmail.com)



**الهام عسکری** دکترای خود را در رشته مهندسی رایانه گرایش هوش مصنوعی در سال ۱۳۹۷ دریافت کرده است. ایشان هم اکنون استادیار بخش مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد فومن و شفت است. زمینه‌های پژوهشی وی منطق، سامانه‌های فازی و کاربردهای آنها، هوش مصنوعی و یادگیری ماشین است. نشانی رایانامه ایشان عبارت است از:

[askary.elham@gmail.com](mailto:askary.elham@gmail.com)

