علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۸، شماره ۲، ۱۳۹۹
صفحات ۳۱- ۳۹
مقاله عادی

شناسایی رویداد در ویدیو با استفاده از شبکه عصبی عمیق بهینه

امیرحسین زنگنه[1*]، مهدی چم‌پور[2]، کامران لایقی[3]

*نویسنده مسئول، دریافت: ۱۴۰۰/۰۲/۲۰، بازنگری: ۱۴۰۰/۰۲/۲۶، پذیرش: ۱۴۰۰/۰۳/۰۵

[1] دانشجوی دکتری، دانشکده مهندسی برق و کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران
[2] استادیار، دانشکده مهندسی کامپیوتر ، دانشگاه صنعتی قوچان، قوچان، ایران
[3] استادیار، دانشکده مهندسی برق و کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران

چکیده

یادگیری عمیق به عنوان یکی از تکنیک‌های یادگیری ماشین، از پیشرفت‌های فناوری واحدهای پردازش گرافیکی استفاده کرده و این امر به نوبه خود استفاده گسترده از آن را فراهم آورده است. تکنیک‌های یادگیری عمیق به نتایج بسیار خوبی در بسیاری از مسائل مهم از جمله شناسایی و تشخیص رویداد در ویدیوی ورزش فوتبال، در مقایسه با روش‌های سنتی دست یافته‌اند. یکی از چالش های عمده استفاده از شبکه‌های عصبی عمیق برای مدیریت و طبقه بندی تصاویر، تعداد بسیار زیاد پارامترهای قابل آموزش در این نوع شبکه‌ها است که منجر به بار محاسباتی بالا و زمان طولانی برای آموزش شبکه عصبی عمیق می‌شود. شبکه عصبی دنس‌نت[1] یکی از آخرین شبکه‌های ارائه شده برای اهداف شناسایی و تشخیص اشیاء هست. ما در این مقاله از شبکه عصبی عمیق دنس‌نت برای شناسایی و تشخیص رویدادهای کارت زرد و قرمز، پنالتی و ضربه آزاد در ویدیوی ورزش فوتبال به عنوان یک معماری پایه استفاده کرده‌ایم. تعداد و اندازه بلوک‌های شبکه دنس‌نت در تعداد پارامترهای قابل آموزش و همچنین دقت شبکه تاثیر گذار است. در این مقاله ما سعی کرده‌ایم با ایجاد تغییر در معماری پایه شبکه عصبی عمیق دنس‌نت با حفظ دقت، تعداد پارامترهای قابل آموزش این شبکه را کاهش دهیم. برای این منظور با بررسی حالت‌های ممکن برای قرار گیری بلوک‌های با سایز مختلف شبکه دنس‌نت اقدام به ارائه یک معماری پیشنهادی برای شبکه عصبی عمیق کرده‌ایم. نتایج ارزیابی‌ها، نشان‌دهنده کاهش تعداد پارامترهای قابل آموزش شبکه عصبی عمیق و در عین حال افزایش دقت معماری پیشنهادی برای شناسایی و تشخیص رویدادهای مهم در ورزش فوتبال است.

کلمات کلیدی: تشخیص رویداد، رویدادهای مهم بازی فوتبال، شبکه عصبی عمیق دنس‌نت، یادگیری عمیق، خلاصه‌سازی ویدیو.

۱- مقدمه

محبوبیت بسیار زیاد وب سایت‌های ویدیویی و شبکه‌های اجتماعی از یک سو، رشد و گسترش تجهیزات و رسانه‌های مختلف ضبط و ذخیره‌سازی ویدیو برای اهداف تجاری، امنیتی و سرگرمی از سویی دیگر سبب تولید گسترده محتوای ویدئویی شده و کابران مختلف در سراسر دنیا با حجم بسیار زیادی از انواع مختلف داده‌های ویدیویی به صورت روزانه در ارتباط هستند.

در میان ویدیوهای مختلف در دسترس کاربران، ویدیوهای ورزشی و به طور خاص با توجه به محبوبیت ورزش فوتبال در سراسر دنیا، ویدیوهای ورزش فوتبال در برابر سایر ویدیوها از نظر حجم و تعداد در صدر تقاضا هستند. ورزش فوتبال با بیش از ۲۶۵ میلیون بازیکن در بیش از ۲۰۰ کشور جهان [۱] و با بیشترین تعداد مخاطب ورزشی در تلویزیون در حال حاضر محبوبترین و پرطرفدارترین ورزش در جهان است [۲].

شکل ۱. تصاویر رویداد مهم بازی فوتبال شامل گل، پنالتی، ضربه آزاد، کارت زرد و قرمز

جذابیت ورزش فوتبال نه تنها هواداران، بلکه محققان زیادی از مناطق مختلف را نیز در سراسر جهان به سوی خود جلب کرده و سبب شده روش‌های مختلفی برای تجزیه و تحلیل اتوماتیک مسابقات فوتبال پیشنهاد شود. از ویژگی‌های بازی فوتبال می‌توان به طولانی بودن زمان آن اشاره کرد که می‌تواند در مواردی مزیت و در مواردی به عنوان یک محدودیت تلقی شود. مساله زمان طولانی بازی فوتبال موجب شده است که علاوه بر حجم زیاد مورد نیاز برای ذخیره‌سازی، در اغلب موارد همه مردم فرصت تماشای ۹۰ دقیقه فوتبال را ندارند ولی از سوی دیگر علاقه‌مند هستند دستکم لحظات مهم و هیجان انگیز بازی را مشاهده کنند. از این رو، اخیرا تحقیقات گسترده ای پیرامون این مسئله آغاز شده است و بطور ویژه می توان به مسائلی مانند خلاصه‌سازی بازی ویدیو متناسب با علایق کاربران [۳]، شناسایی لحظات حساس و رویدادهای مهم [۴]، شناسایی و ردیابی بازیکنان [۵] یا توپ [۶] ، تحلیل آماری بازی و پیش‌بینی نتیجه بازی فوتبال [۷] و مواردی از این قبیل اشاره کرد. اگرچه بسیاری از این برنامه‌های تولید شده قادر به ارائه نتایج قابل قبولی هستند، اما به طور معمول در زمان واقعی کار نمی‌کنند، که این یک نیاز اساسی برای استفاده آنها در پخش زنده است [۸]. علاوه بر این، کارهای انجام شده معمولا استراتژی‌هایی را پیشنهاد می‌کنند که بر حل یا تحلیل جنبه‌های خاصی از ورزش فوتبال تمرکز دارند و نمی‌توان از آن‌ها به عنوان محصول نهایی مناسب برای استفاده توسط شبکه‌های تلویزیونی، کاربران و یا باشگاه‌های فوتبال برای خلاصه‌سازی یا تحلیل و آنالیز ویدیوی فوتبال، استفاده کرد. به همین دلایل، برنامه های جدید مداوم در حال تهیه و ارائه هستند.

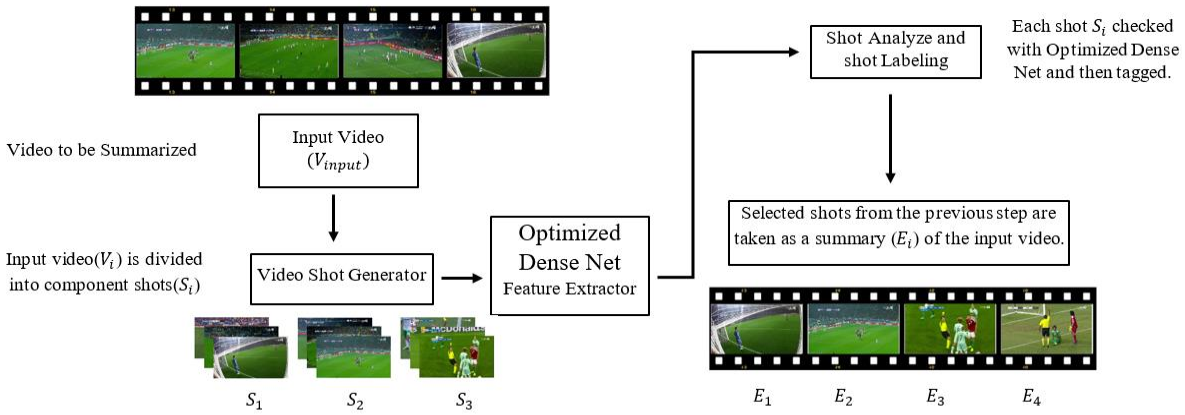
شناسایی رویدادهای یک بازی فوتبال، همچنین انواع خلاصه سازی آنها از مهمترین موضوعات تحقیقات سال‌های اخیر در حوزه تحلیل ویدئوهای ورزشی است. از جمله مهمترین انگیزه‌ها و دلایل خلاصه سازی اتوماتیک رویدادهای ویدیویی صرفه‌جویی در زمان تماشای ویدیو است. از این رو تجزیه و تحلیل خودکار ویدیو فوتبال در تولید محتوای فوتبال بسیار ارزشمند است و می‌تواند به ویراستاران ویدیویی کمک کند تا با صرف زمان کمتری، اقدام به خلاصه‌سازی ویدیو کنند. شرکت‌های مختلفی در سراسر دنیا در حوزه تجزیه و تحلیل ویدیوهای ورزش فوتبال فعال هستند. اما سامانه‌های خودکار تحلیل ویدیو هنوز راه طولانی تا انجام کاملا هوشمند این فرآیند دارند و در اغلب موارد سامانه ها از نیروی انسانی برای تجزیه و تحلیل ویدیو استفاده می‌کنند.

طبقه‌بندی تصاویر برای شناسایی رویدادها، یا به عبارتی تعیین مهمترین وحساس‌ترین رویدادها در بازی فوتبال کاری پیچیده و براساس نظرات کاربران متقاوت است. تصاویر تعدادی از رویداد های مهم بازی فوتبال در شکل ۱ نمایش داده شده است. ما به منظور شناسایی رویدادهای مهم بازی فوتبال و طبقه‌بندی تصاویر موجود بر اساس نظرسنجی میدانی اقدام کردیم و پرسشنامه‌ای شامل ۵ رویداد رایج بازی فوتبال را طراحی و انتخاب مهمترین رویداد را به شرکت‌کنندگان در نظر سنجی واگذار کردیم. فرم نظرسنجی مذکور شامل رویدادهای گل، کارت زرد و قرمز، ضربه آزاد، پنالتی و برخورد توپ با تیرک دروازه بوده است. پرسشنامه را بین تعداد ۲۰۰ نفر در محدوده‌های سنی متفاوت پخش کرده و از مخاطبان درخواست کردیم که به ۵ رویداد مهم مطرح شده در پرسشنامه از عدد یک (کمترین امتیاز) تا عدد ۵(بیشترین امتیاز) یک عدد را اختصاص دهند. نتایج نظرسنجی در جدول ۱ ارائه شده است.

جدول ۱. نتایج نظر سنجی در مورد رویدادهای مهم بازی فوتبال

ردیف	نوع رویداد	میانگین امتیاز دریافتی
۱	گل	۴.۳۲
۲	پنالتی	۴.۰۰
۳	ضربه آزاد	۳.۳۲
۴	برخورد توپ با تیرک دروازه	۳.۲۰
۵	کارت زرد و قرمز	۲.۹۶

رویدادهای اشاره شده در نظرسنجی را به طور کلی می‌توانیم به رویدادهای مبتنی بر یک مشاهده و رویدادهای مرکب مبتنی بر فعالیت تقسیم نمود. در این مقاله هدف ما شناسایی رویدادهای مبتنی بر مشاهده شامل پنالتی، ضربه آزاد و کارت زرد و قرمز است. مطابق شکل ۲ برای خلاصه‌سازی ویدیو، ابتدا ویدیو دریافتی به شات‌های مستقل تقسیم می‌شود، سپس با هدف کاهش تعداد پارامترهای قابل آموزش در شبکه عصبی عمیق دنس‌نت، معماری این شبکه مورد بررسی و اصلاح قرار گرفته و در مرحله بعد هر شات با استفاده از یک شبکه عصبی عمیق بهینه شده مورد بررسی و برچسب‌گذاری قرار می‌گیرد.



شکل ۲. معماری پیشنهادی برای خلاصه‌سازی ویدیو براساس رویدادهای پنالتی، ضربه آزاد، کارت زرد و قرمز

این مقاله در ادامه به شرح زیر سازماندهی شده‌است: در بخش ۲، کارهای انجام شده قبلی در زمینه خلاصه‌سازی ویدیویی فوتبال مورد بررسی قرار می‌گیرد. در بخش ۳، معماری شبکه عصبی عمیق دنس‌نت و نحوه اصلاح معماری شبکه عصبی عمیق دنس‌نت به تفصیل شرح داده می‌شود. در بخش ۴ نتایج تجربی ارائه شده و در نهایت، نتیجه‌گیری در بخش ۵ ذکر شده است.

۲ - کارهای مرتبط

تشخیص خودکار رویداد و تفسیر معنایی صحنه‌ها، یک کار چالش برانگیز در خلاصه‌سازی ویدیو بازی فوتبال است. این کار با استخراج ویژگی‌ها در سطوح معنایی مختلف انجام می‌شود. از ویژگی‌های سطح پایین تصویر مانند رنگ، شکل و بافت، توپ، دهانه دروازه و همچنین از ویژگی‌های ویدیویی سطح بالا از قبیل شناسایی وضعیت ویدیو مانند حالت پخش مجدد بازی و حالت وقفه ایجاد شده در بازی، یا سایر ویژگی‌های موجود مانند صوت، متون اینترنتی مرتبط یا ویدیو و غیره برای شناسایی و تشخیص رویداد در ویدیوی بازی فوتبال استفاده می‌شود.

خلاصه‌سازی ویدئو معمولا به دو روش استاتیک (مبتنی بر فریم‌های کلیدی) و خلاصه‌سازی پویا انجام می‌شود. در روش خلاصه‌سازی استاتیک، مجموعه‌ای از تصاویر مربوط به رویدادهای با اولویت بالا از ویدیو استخراج شده و سپس یک ویدیو شامل فریم‌های استخراج شده‌ی مرحله قبل به عنوان خلاصه استاتیک ویدیوی اصلی تولید می‌شود. در مقابل، در روش خلاصه سازی پویای ویدیو، ابتدا رویدادهای مهم موجود در ویدیو شناسایی شده و سپس شات‌هایی از ویدیو که شامل رویدادهای مهم هستند، به عنوان خلاصه نهایی ویدیو بازی فوتبال تهیه می‌شوند. در روش خلاصه-سازی استاتیک، ویدیوی تولید شده فقط شامل فریم‌های مهم ویدیو بوده و فاقد صدا است در حالیکه در روش خلاصه‌سازی پویا، ویدیوی خلاصه تولید شده شامل همه ویژگی‌های ویدیوی اصلی ازجمله صوت است.

کارهای متنوع و زیادی برای خلاصه‌سازی ویدیو انجام شده که براساس نوع ویژگی‌های استفاده شده در آن‌ها، میتوانیم تحقیقات انجام شده را به سه دسته‌ی: ۱) روش های مبتنی بر منابع خارجی مانند متن‌های اینترنتی مربوط به ویدیو و اطلاعات مرتبط با ویدیو در شبکه‌های اجتماعی ۲) روش‌های خلاصه‌سازی مبتنی بر ویژگی-های مختلف صوتی – تصویری ویدیو و ۳) روش‌های که فقط از ویژگی‌های دیداری و بصری موجود در ویدیو استفاده می‌کنند، تفسیم بندی کنیم.

۲-۱- روش های مبتنی بر منابع خارجی

این روش‌ها از اطلاعات و داده‌های غیر صوتی - تصویری موجود در متون اینترنتی یا اطلاعات موجود در شبکه‌های اجتماعی که مربوط به یک ویدیوی مسابقه فوتبال هستند، برای شناسایی رویداد و نهایتا خلاصه‌سازی ویدیو فوتبال بهره می‌گیرند[۸].

ونگ و همکاران در [۹] محتوای متون اینترنتی را برای یافتن کلمات کلیدی مرتبط با رویدادها مانند گل، ضربه آزاد، کارت، پنالتی و ... مورد جستجو قرار می‌دهند و اطلاعات مربوط به زمان و افراد مرتبط با رویداد را استخراج می‌کنند. سپس، با استفاده از زمان ثبت رویداد، اقدام به تهیه یک ویدیو خلاصه شامل زمان ثبت شده می‌کنند. تانگ و همکاران در[۱۰] نیز در کاری مشابه برای شناسایی رویدادهای مهم بازی فوتبال از متون اینترنتی مرتبط با مسابقات فوتبال به عنوان اطلاعات کمکی استفاده کردند. آندلوسی و همکاران در [۱۱] با استفاده از محتوی موجود در شبکه های اجتماعی مانند توییتر اقدام به شناسایی رویداد و سپس خلاصه‌سازی ویدیوی ورزش فوتبال کرده‌اند. آن‌ها با استخراج توییت‌های مربوط به ویدیوهای ورزش فوتبال و سپس تحلیل و بررسی آن‌ها اقدام به شناسایی رویدادهای مهم بازی و نهایتا اقدام به خلاصه‌سازی ویدیو ورزش فوتبال کرده‌اند.

استفاده از منابع اطلاعاتی اضافی مانند متون اینترنتی و اطلاعات موجود در شبکه های اجتماعی می‌تواند به افزایش دقت در شناسایی رویدادهای مهم در ویدیو ورزش فوتبال کمک کند اما از اشکالات مهم این روش‌ها این است که برای شناسایی رویداد در ویدیو باید بین اطلاعات دریافتی از منابع خارجی با سیگنال سمعی بصری همگام سازی ایجاد نمود. علاوه بر این، باید به این نکته نیز توجه نمود که اطاعات مربوط به رویدادهای ورزش با یک تاخیر زمانی توسط بینندگان، منقدان و غیره در شبکه‌های اجتماعی ثبت می‌شود و عملا روش‌های خلاصه‌سازی ویدیو مبتنی بر منابع خارجی تا زمان دریافت این اطلاعات با تاخیر روبرو می‌شوند. البته در مواردی این تاخیرها می‌تواند قابل توجه باشد.

۲-۲- روش‌هایی مبتنی بر ویژگی‌های سمعی- بصری (روش‌های مالتی‌مُدال)

در این روش‌ها از ویژگی‌های صوتی موجود در ویدیوهای ورزشی به عنوان یک عامل مهم در کنار ویژگی‌های بصری موجود در ویدیو برای شناسایی رویدادهای مهم و سپس خلاصه‌سازی ویدیوهای ورزش فوتبال استفاده می‌شود.

شی و همکاران در [۱۲] ویژگی‌های صوتی شامل تشویق تماشاگران و هیجان مفسران ورزشی را استخراج کرده‌اند، و همزمان نشانه‌های (ویژگی‌های) بصری را تشخیص دادند. بعد از استخراج مفهوم معنایی و توجه به توالی معنایی رویدادهایی که با هم مرتبط هستند، مانند ورود توپ به دروازه و هلهله تماشاچیان، قوانین موجود برای شناسایی رویداد به کار گرفته می‌شوند. در کاری مشابه شوو و همکاران در [۱۳] برای تجزیه و تحلیل محتوی ویدیو اقدام به استخراج ویژگی‌های سطح پایین و سطح میانی از کانال‌های صدا / تصویری کردند.

کوول‌کر و همکاران در [۱۴] روشی برای آنالیز معنایی ویدیو و خلاصه‌سازی ویدیو با شناسایی مفاهیم با استفاده از یک شبکه بیزی معرفی کردند که در آن، رویدادهای برجسته بازی با استفاده از ویژگی‌های صوتی با استفاده از قوانین تولید شده و دانش این حوزه از کلیپ‌های ویدیو، شناسایی می‌شوند. مجموعه‌ای از کلیپ‌های برجسته که شامل رویدادهای حساس بازی هستند، برچسب‌گذاری شده و در یک چکیده ویدئویی برای کاربردهای مختلف مانند مرور رویدادهای مهم، شاخص‌گذاری و بازیابی ویدیو بکار برده می‌شوند.

جاندنگرو و همکاران در [۱۵] با استخراج ویژگی‌های صوتی (صدای سوت داور) و تصویری ویدیو اقدام به شناسایی وقفه‌های ایجاد شده در بازی کردند. برای مثال در بازی فوتبال زمانی که سوت داور شنیده می‌شود به این معنی است که یک خطا اتفاق افتاده یا توپ خارج از میدان بوده و در نتیجه یک وقفه در بازی رخ داده است. از جمله مزایای کار یاد شده، عمومی بودن و کاربردی بودن آن برای همه بازی‌هایی است که دارای ساختار بازی/ وقفه هستند.

روش‌هایی مبتنی بر ویژگی‌های سمعی- بصری با محدیدویت‌های از جمله: ۱- افزایش تعداد سنسورها و تجهیزات سخت افزاری به منظور ضبط صوت، ۲- محدودیت در فاصله ضبط داده‌ها، با استفاده از دوربین می‌توان رویدادها را از فاصله دور ثبت و ضبط نمود در حالیکه اگر بخواهیم همان ویدیو را با صدا تهیه کنیم با محدودیت فاصله روبرو خواهیم بود. ۳- حذف نویز و صداهای اضافی موجود در ویدیو که توسط تماشاچیان تولید می‌شود و می‌تواند موجب خطا در عملکرد سیستم شود. به عنوان مثال روش‌هایی که با شناسایی صدای سوت داور اقدام به شناسایی رویداد می‌کنند در مواردی که تماشاچیان اقدام به سوت زدن در حین بازی می‌کنند با خطا روبرو می‌شوند. ۴- در ویدیوهایی که در ورزشگاه‌های سرپوشیده تهیه می‌شوند، صدای تماشاچیان صدای غالب بوده و عملا صدای سوت داور و بازیکنان توسط روش‌های مالتی‌مُدال قابل استفاده نیستند. ۵- روش‌های مالتی‌مدال فقط روی ویدیوهایی که تحت شرایط خاصی تهیه شده‌اند، قابلیت استفاده را دارا بوده و عمومی نیستند.

۲-۳- روش‌های مبتنی بر ویژگی‌های دیداری

شرکت‌های پخش ویدیویی از تکرارهای[۲] صحنه‌های هیجان انگیز و مهم استفاده می‌کنند تا روی رویدادهای خاص بازی با جزئیات کامل تأکید کرده و آنها را برای بینندگان خود نمایش دهند. صحنه تکرار به طور عمده شامل نمایش حرکت آهسته یک رویداد جالب و گاهی اوقات لوگو بازی (علامت ویژه مسابقه یا علامت تجاری اسپانسر برای برخی از فریم‌ها) است، که در آغاز و پایان صحنه تکرار استفاده می‌شود. استفاده از ویژگی تکرار رویدادهای حساس بازی نیز در برخی از کارهای مشابه برای خلاصه‌سازی ویدیو مورد استفاده قرار می‌گیرد.

الدیب و همکاران در [۱۶] برای شناسایی رویدادهای حساس بازی اقدام به شناسایی لوگو بازی کرده‌اند. آن‌ها هنگامی رویداد گل شناسایی می‌کنند که یک وقفه در مسابقه تشخیص داده می‌شود یا برخی علایم از تشویق بازیکنان مشاهده می‌شوند و یا پخش مجدد بازی از زوایای مختلف که توسط دوربین های مختلف بدست آمده‌اند، نمایش داده می‌شوند. هنگامی که لوگوی مسابقات در ویدیو پخش می‌شود اقدام به تشخیص صحنه تکرار می‌کنند و سپس برای خلاصه سازی ویدیو با استفاده از شناسایی صحنه‌ی تکرار، شناسایی مبتنی بر قاعده گل و تشخیص حمله، اقدام می‌کنند. این تشخیص از طریق تشخیص مرز براساس دهانه‌ی دروازه، طبقه‌بندی عکس، تشخیص صحنه‌ی تکرار، و تشخیص بورد ثبت امتیازات امکان‌پذیر است.

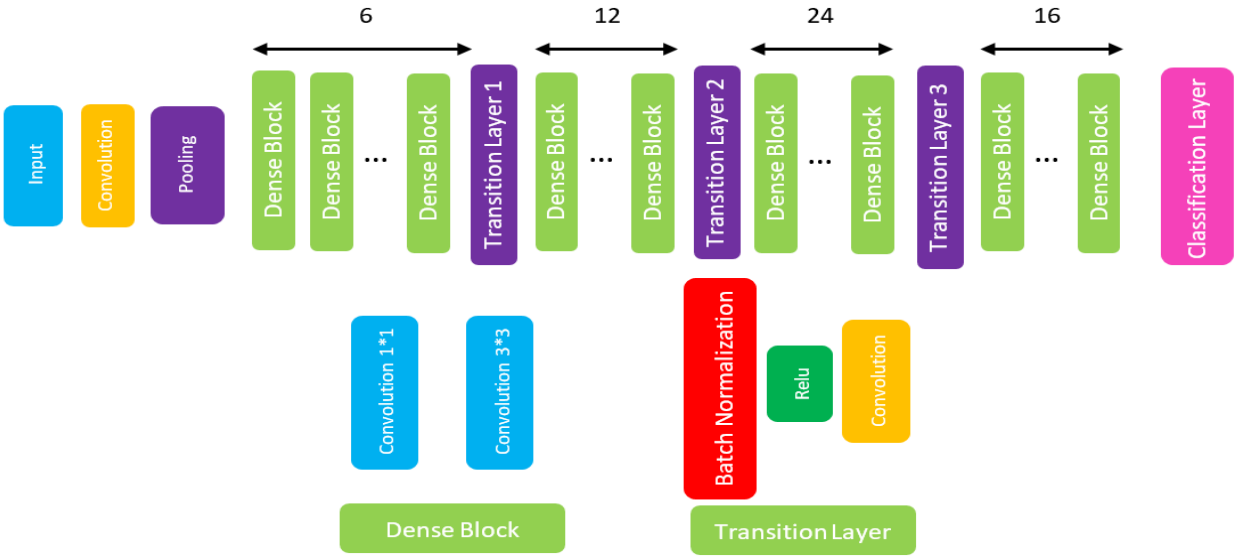
فخار و همکاران در [۱۷] برای شناسایی رویدادهای حساس بازی فوتبال اقدام به تشخیص صحنه‌های پخش مجدد در ویدیو کردند. در روش آن‌ها صحنه‌های پخش مجدد، شامل رویدادهای مهم بازی هستند. برای شناسایی صحنه‌های پخش مجدد نیز اقدام به شناسایی لوگوی بازی در فریم‌های ویدیو کرده‌اند. آن‌ها به محض تشخیص لوگوی مسابقات در یک فریم، به فریم‌های قبلی برگشته و این کار را تا رسیدن به فریمی که حاوی یک تصویر از نمای دور[۳]، است ادامه می‌دهند. مجموعه فریم‌های بین تصویر نمای دور و لوگوی مسابقات به عنوان رویداد مهم بازی خلاصه می‌شوند.

یو و همکاران در [۴]، نیز برای شناسایی رویدادهای حساس بازی اقدام به تشخیص صحنه‌های پخش مجدد ویدیو کرده‌اند. به عقیده آن‌ها لوگوی مسابقات قبل و بعد صحنه‌های پخش مجدد ویدیو قرار دارند. آن‌ها برای تشخیص وجود یا عدم وجود لوگوی مسابقات در فریم‌های ویدیو از شبکه عصبی کانولوشن استفاده کرده و براساس رویدادهای حساس شناسایی شده اقدام به خلاصه سازی ویدیو کرده‌اند.

در روش‌های مبتنی بر شناسایی لوگوی [۲۰]–[۱۸] مسابقات ۱- باید لوگوی مسابقات برای سیستم تعریف شود ۲- سیستم فقط به تصویر لوگو مسابقه حساس بوده و هیچ دانشی در مورد نوع رویداد اتفاق افتاده نداشته و در نتیجه امکان خلاصه سازی ویدیو براساس نوع رویداد در این روش وجود ندارد و ۳- این روش عمومی نبوده و فقط برای ویدیوهایی طراحی شده که توسط یک کاربر انسانی از قبل مورد بررسی قرار گرفته باشد که سلیقه و انتخاب کاربر شرکت‌های پخش ویدیویی در آن دخیل است. با توجه به محدودیت‌های موجود در روش‌های خلاصه‌سازی ویدیو که به آن اشاره شد، هدف ما در این مقاله خلاصه‌سازی ویدیوی ورزش فوتبال باتوجه به ویژگی‌های دیداری موجود در ویدیو است. روش ما قادر است به صورت آنلاین و بدون نیاز به هیچ گونه پیش‌پردازشی یا حتی دخالت انسان اقدام به خلاصه‌سازی رویدادهای مهم ورزش فوتبال نماید.

۳- روش پیشنهادی

در این بخش با توجه به ضرورت توسعه روش‌های خودکار و موثر برای شناسایی رویداد در ویدیوی ورزش فوتبال و سپس خلاصه‌سازی رویدادهای مهم، به معرفی روش پیشنهادی می‌پردازیم. ما ابتدا از شبکه عصبی عمیق دنس‌نت-۱۲۱ لایه به عنوان یک مدل پایه یادگیری عمیق برای استخراج ویژگی‌ها استفاده می‌کنیم، سپس با هدف کاهش بار محاسباتی و زمان اجرا، اقدام به بهبود معماری شبکه عصبی پایه می‌نماییم.



شکل ۳. معماری شبکه عصبی عمیق *Dense net-121* [۲۳]

در ادامه ابتدا معماری شبکه عصبی عمیق دنس‌نت-۱۲۱ که در این مقاله مورد استفاده قرار گرفته شده را شرح می‌دهیم، سپس در زیربخش بعدی نحوه بهبود و یافتن معماری بهینه برای شبکه عصبی عمیق پایه را ارائه و تشریح می‌کنیم.

۳-۱- معماری مدل پایه

ما در این مقاله از مدل پایه شبکه عصبی عمیق دنس‌نت-۱۲۱ برای شناسایی رویدادهای مبتنی بر مشاهده شامل پنالتی، ضربه آزاد و کارت زرد و قرمز، استفاده می‌کنیم. شبکه عصبی عمیق دنس‌نت سال ۲۰۱۷ توسط هانگ و همکاران در [۲۱] معرفی گردید. این شبکه، یکی از آخرین شبکه‌های ارائه شده برای اهداف شناسایی و تشخیص اشیاء است. از نظر معماری، این شبکه دارای معماری مشابه شبکه عصبی عمیق ResNet بوده، اما دارای چند تفاوت اساسی است. این معماری نسبت به سایر معماری های قبلی نرخ خطای کمتری بر روی دیتابیس های CIFAR و SVHN دارد. همچنین براساس نتایج بدست آمده برای شناسایی اشیاء، این معماری برای دیتابیس ImageNet نسبت به معماری ResNet به تعداد پارامتر کمتری نیاز دارد در حالی که دقت دو روش تقریبا مشابه است.

در شبکه های عصبی کانولوشن لایه‌های ابتدایی ویژگی‌های سطح پایین مانند لبه‌ها را استخراج و لایه‌هایی که در انتهای این زنجیره قرار گرفتند ویژگی‌های سطح بالا مثل بافت‌ها، و اشکال پیچیده و غیره را استخراج می‌کنند. در برخی موارد ممکن است ویژگی‌های سطح پایین استخراج شده در عملیات طبقه‌بندی یک کلاس از ویژگی‌های سطح بالای استخراج شده کاراتر و موثرتر باشند، در نتیجه با توجه به متصل بودن لایه‌هایی ابتدایی به لایه‌های انتهایی در معماری شبکه عصبی عمیق دنس‌نت، این شبکه می‌تواند یاد بگیرد که برای کلاس مورد نظر فقط از ویژگی های سطح پایین، ویژگی‌های سطح بالا یا از ترکیب ویژگی های سطح پایین و بالا استفاده کند.

۳-۲- یافتن معماری بهینه شبکه عصبی عمیق

استفاده از شبکه‌های عصبی عمیق برای طبقه‌بندی تصاویر بطور کلی با دو چالش ۱) چگونه می توان دقت شبکه‌های عصبی عمیق را بیشتر بهبود داد؟ و ۲) چگونه می‌توان شبکه را سبکتر کرد یا به عبارتی تعداد پارامترهای قابل آموزش را در شبکه‌های عصبی عمیق کاهش داد؟ مواجه هستند.

تعداد و اندازه بلوک‌های شبکه دنس‌نت در تعداد پارامترهای قابل آموزش و همچنین دقت شبکه موثر است. در شبکه عصبی عمیق دنس‌نت-۱۲۱، لایه‌های کانولوشن، مکس‌پولینگ، بتچ نرمالیزیشن و ریلو در بلوک‌های متوالی، ۱۲۱ لایه پردازشی را بوجود می‌آورند. لایه‌های شبکه دنس‌نت در ۴ بلوک پی‌درپی به ترتیب ۶، ۱۲، ۲۴ و ۱۶ قرار گرفته‌اند که در نهایت مجموع تعداد ساب‌پروسس‌های کانولوشنی برابر است با ۵۸ (۵۸ = ۱۶ + ۲۴ + ۱۲ + ۶) و تعداد کل لایه‌های کانولوشن در شبکه دنس نت برابر است با ۱۱۶ لایه (۱۱۶ = ۲ * ۵۸) است.

در شبکه عمیق دنس‌نت به منظور بررسی امکان کاهش تعداد پارامترهای قابل آموزش و همچنین افزایش دقت در شبکه عمیق دنس‌نت با ثابت نگه داشتن تعداد کل ساب‌پروسس‌های کانولوشنی، اقدام به تغییر سایز بلوک‌های شبکه و همچنین بررسی حالت های مختلف قرارگیری بلوک‌های شبکه دنس‌نت در کنار هم کرده‌ایم. برای این منظور تعداد کل حالت‌هایی که بلوک‌های با سایزهای مختلف می‌توانند در شبکه عمیق در کنار هم قرار بگیرند و مجموع تعداد ساب‌پروسس‌ها بدون تغییر باقی بماند، را بررسی کرده‌ایم.

در نتیجه مطابق با رابطه شماره ۱ مساله پیدا کردن تعداد کل ترکیب‌های ممکن از مجموعه $S = \{S_1 . S_2 . \ldots . S_k\}$ است که مجموع هر ترکیب برابر با عدد صحیح N شود:

$$N = \sum_{k=1\ldots m} x_k \, S_k \quad where \; x_k \geq 0 \, . k \in \{1. 2. \ldots . m\} \qquad (۱)$$

براساس رابطه ۱، مساله پیدا کردن همه ترکیب‌های مختلف از S_k ها بطوری که مجموع همه S_k برابر با N= 58 شود، مشابه مساله خرد کردن پول[۴] است. برای حل این مساله از تکنیک برنامه‌ریزی پویا استفاده نموده‌ایم. در نتیجه برای حل مساله یک اندازه حداقلی(۶، حداقل سایز بلوک شبکه دنس‌نت پایه) و یک مقدار حداکثر(۲۴، حداکثر سایز بلوک شبکه دنس‌نت پایه) برای سایز بلوک‌ها در نظر گرفته‌ایم. سایز بلوک‌ها هم امکان افزایش ۲ واحدی دارند. با توجه به تعداد کل ساب‌پروسس‌های کانولوشن (N = ۵۸) و سایزهای ممکن برای بلوک‌های شبکه عصبی عمیق

($S = \{6.8.10.12.14.16.18.20.22.24\}$) ، تعداد کل حالت‌هایی که می‌توان اعداد مجموعه S را باهم ترکیب کرد تا مجموع برابر با N = ۵۸ شود، را با هدف کاهش تعداد پارامترهای قابل آموزش شبکه عصبی عمیق دنس‌نت و در عین حال افزایش دقت شبکه مورد بررسی و ارزیابی قرار داده‌ایم. برای کوچک کردن فضای مساله فقط ترکیب‌های صعودی از S_k ها را مورد بررسی قرار داده‌ایم. در این حالت تعداد کل حالت‌های مورد بررسی برابر با ۲۵ ترکیب مختلف می‌شود. برای ساختارهای مختلف شبکه عصبی جدید ارائه شده، معیار دقت و تعداد پارامترهای قابل آموزش را محاسبه و در نهایت با استفاده از رابطه ۲ بهترین معماری را برای شبکه عصبی عمیق انتخاب کرده‌ایم.

$$Structure\ Gain = \frac{Structure\ Accuracy}{Trainable\ Parameters} \qquad (۲)$$

۴- نتایج تجربی و آزمایش‌ها

در این بخش به ارزیابی و تحلیل روش پیشنهادی می‌پردازیم. در ابتدا مشخصات پایگاه داده تصاویر مورد استفاده را معرفی کرده و سپس روش پایه و ساختار معماری پیشنهادی را مورد ارزیابی قرار می‌دهیم. همچنین در انتهای این بخش مقایسه‌ای بین نتایج معماری پایه و معماری پیشنهادی انجام پذیرفته است. در این راستا، ما یک استراتژی یکسان در همه آزمایش‌ها داریم. هنگامی که تصاویر مطلوب (تصاویر کلاس کارت، پنالتی و ضربه آزاد) را انتخاب می‌کنیم، آن‌ها را به سه بخش آموزش، اعتبار سنجی، و زیرمجموعه‌های تست با نسبت ۵۰ ٪، ۲۵ درصد و ۲۵ درصد، تقسیم می‌کنیم. پارامترهای مربوط به شبیه‌سازی نیز در جدول شماره ۲ ارائه شده است.

جدول ۲. پارامترهای مربوط به شبیه سازی

Parameter	Value
Optimizer	Adam
Loss function	Binary cross-entropy
Performance metric	Accuracy
Total Classes	2 (Card, Penalty and Free Kick)
Batch Size	32

۴-۱- پایگاه‌داده تصاویر

با توجه به تحقیقات انجام شده، در حال حاضر تنها مجموعه داده اختصاصی جامع برای تحلیل اشیاء موجود در زمین فوتبال و تشخیص رویداد که بصورت دسترسی رایگان برای امور تحقیقاتی وجود دارد، پایگاه داده پایگاه داده IAU-SD [۵] است. سایر پایگاه داده‌های موجود یا اکثرا فقط شامل ویدئو هستند یا تعداد تصاویر در آن‌ها خیلی کم است. همچنین عدم تنوع در شرایط مختلف روشنایی، آب و هوا و غیره سبب کاهش جامعیت دیتاست‌های موجود شده است [۴].

پایگاه داده استفاده شده شامل ۱۰۰.۰۰۰ تصویر از تعداد ۳۳ مسابقه فوتبال شامل مسابقات ملی و باشگاهی از سراسر جهان هست که در آن تصاویر در ۱۰ کلاس دروازه، حالت شروع بازی/شروع مجدد، شادی بازیکنان، کارت زرد و قرمز، توپ فوتبال، تصویر ورزشگاه، تصویر داور، پنالتی و ضربه آزاد دسته‌بندی شده‌اند.

۴-۲- معیارهای ارزیابی روش پیشنهادی

ما به منظور ارزیابی عملکرد روش پیشنهادی از ۴ معیار ارزیابی شامل بازیابی [۶] (رابطه ۳)، وضوح [۷] (رابطه ۴)، معیار-اف [۸] (رابطه ۵) و دقت [۹] (رابطه ۶) استفاده کرده‌ایم. در این ارزیابی‌ها هدف پیدا کردن فریم‌های شامل رویدادهای مهم کارت زرد و قرمز، پنالتی و ضربه آزاد، به منظور خلاصه‌سازی ویدیو بازی فوتبال است.

$$Recall = \frac{TP}{TP+FP} \qquad (۳)$$

$$Precision = \frac{TP}{TP+FN} \qquad (۴)$$

که در آن TP [۱۰] تعداد نمونه های مثبتی است که به درستی مثبت شناسایی شده‌اند، TN [۱۱] تعداد نمونه‌های منفی که به درستی منفی شناسایی شده‌اند، FP [۱۲] تعداد شناسایی‌های مثبت کاذب و FN [۱۳] تعداد شناسایی‌های منفی کاذب است. سپس مقدار معیار-اف f- و دقت به شرح زیر تعریف می‌شوند:

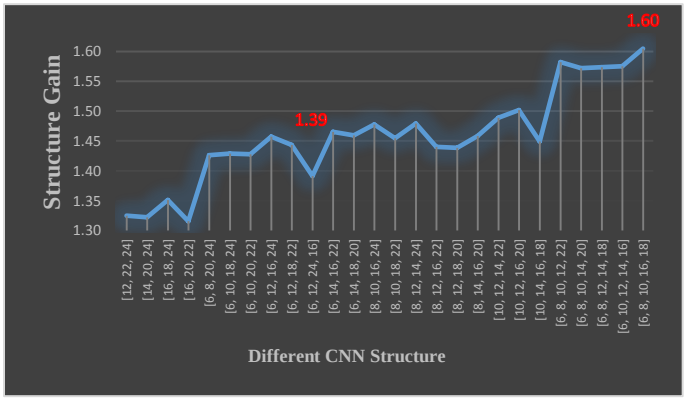
$$\mathrm{f} - measure = \frac{2*Precision*Recall}{Precision+Recall} \qquad (۵)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۶)$$

۴-۳- انتخاب ساختار بهینه شبکه عصبی عمیق

در این بخش با توجه به بخش ۳-۲ معماری‌های مختلف برای شبکه عصبی عمیق با تعداد N = ۵۸ ساب‌پروسس کانولوشنی را با هدف ایجاد یک موازنه بین دقت - کاهش تعداد پارامترهای قابل آموزش شبکه مورد بررسی قرار داده‌ایم. با توجه به رابطه ۱ تعداد کل حالت‌های مختلف برای ایجاد معماری جدید برای شبکه عصبی عمیق برابر با ۲۵ هست، که نتایج برای تشخیص رویداد ضربه آزاد در جدول شماره ۳ ارائه شده‌اند.

با توجه به نتایج ارائه شده در جدول شماره ۳ و همانطور که در شکل شماره ۴ مشهود است معماری ارائه شده با ساختار ([۶, ۸, ۱۰, ۱۶, ۱۸]) نسبت به معماری پایه ارائه شده ([۶, ۱۲, ۲۴, ۱۶]) دارای دقت بالاتر و همچنین تعداد پارامترهای کمتری است، در نتیجه این معماری به عنوان معماری پیشنهادی انتخاب می‌شود. در جدول شماره ۴ نیز جزئیات بیشتری از شبکه پیشنهادی و شبکه دنس‌نت ارائه شده است. با توجه به نتایج ارائه شده، ساختار پیشنهادی دارای پارامترهای قابل آموزش کمتر و دقت بالاتری نسبت به ساختار معرفی شده توسط هانگ و همکاران است.



شکل ۴. نمودار مربوط به بهره معماری‌های مختلف شبکه دنس‌نت.

جدول ۳. مشخصات معماری‌های مختلف شبکه دنس نت

Dense net Structure	Total parameter	Accuracy	Structure Gain
[12, 22, 24]	7,367,105	0/9758	1/32
[14, 20, 24]	7,292,737	0/9641	1/32
[16, 18, 24]	7,240,385	0/9784	1/35
[16, 20, 22]	7,319,105	0/9628	1/32
[6, 8, 20, 24]	6,807,617	0/9708	1/43
[6, 10, 18, 24]	6,732,225	0/9618	1/43
[6, 10, 20, 22]	6,798,401	0/9708	1/43
[6, 12, 16, 24]	6,678,849	0/9732	1/46
[6, 12, 18, 22]	6,738,753	0/9727	1/44
[6, 12, 24, 16]	7,044,417	0/9799	1/39
[6, 14, 16, 22]	6,701,121	0/982	1/47
[6, 14, 18, 20]	6,775,745	0/9887	1/46
[8, 10, 16, 24]	6,601,729	0/9754	1/48
[8, 10, 18, 22]	6,658,497	0/9685	1/45
[8, 12, 14, 24]	6,564,609	0/9712	1/48
[8, 12, 16, 22]	6,615,105	0/9527	1/44
[8, 12, 18, 20]	6,686,593	0/962	1/44
[8, 14, 16, 20]	6,658,945	0/9711	1/46
[10, 12, 14, 22]	6,524,225	0/9715	1/49
[10, 12, 16, 20]	6,586,305	0/9891	1/50
[10, 14, 16, 18]	6,651,713	0/9638	1/45
[6, 8, 10, 12, 22]	6,038,337	0/9552	1/58
[6, 8, 10, 14, 20]	6,081,601	0/9561	1/57
[6, 8, 12, 14, 18]	6,127,169	0/9641	1/57
[6, 10, 12, 14, 16]	6,151,297	0/969	1/58
[6, 8, 10, 16, 18]	6,145,857	0/9863	1/60

جدول ۴. مقایسه معماری پایه شبکه دنس نت و معماری پیشنهادی برای شبکه عصبی عمیق

Structure Name	Trainable parameter	Total parameter	Accuracy	Structure Gain
Base Dense Net Structure	6,774,017	7,044,417	0/9799	**1/39**
Proposed Structure	6,075,393	6,145,857	0/9863	**1/60**

۴-۴- ارزیابی روش ارائه شده

در این بخش، مساله تشخیص هر مشاهده(کارت زرد و قرمز، پنالتی و ضربه آزاد) در مقابل مشاهدات دیگر به عنوان یک مساله دو کلاسه مورد بررسی قرار گرفته‌است. بنابراین، در این آزمایش، ما تصاویری را با تنها یک برچسب مطلوب (به عنوان مثال، تنها شامل ضربه آزاد) انتخاب کردیم. نتایج ارزیابی مدل‌های پایه شبکه دنس‌نت و معماری پیشنهادی با استفاده از داده‌های پایگاه داده IAU-SD ارائه شده است. نتایج حاصل از مقایسه دو مدل پایه و معماری پیشنهادی براساس پارامتر دقت در جدول شماره ۵ ارائه شده است. نتایج حاصل از شناسایی اشیاء به ترتیب ۱- کارت زرد و قرمز، ۲- پنالتی و ۳- ضربه آزاد با استفاده از مدل پایه شبکه عصبی عمیق دنس‌نت و معماری پیشنهادی برای شبکه عصبی عمیق به ترتیب در جدول‌های شماره ۶ و ۷ ارائه شده‌اند.

جدول ۵. مقایسه دقت معماری‌های شبکه عصبی عمیق دنس‌نت پایه و معماری پیشنهادی در تشخیص رویدادهای مهم ویدیوی بازی فوتبال.

Method	Card Detection	Free Kick Detection	Penalty Detection
Initial Dense Net	0/93	0/9799	0/9696
Proposed Structure	0/9883	0/9744	0/9894

جدول۶. نتایج معماری پایه در شناسایی رویدادهای مهم ویدیوی بازی فوتبال.

Method Dense net	Yellow Card Detection	Free Kick Detection	Penalty Detection
Recall	0.9257	0.9738	0.9619
Precision	0.935	0.9864	0.978
f-measure	0.9303	0.98	0.9699

جدول ۷ . نتایج معماری پیشنهادی شبکه دنس‌نت در شناسایی رویدادهای مهم بازی فوتبال.

Method Dense net	Yellow Card Detection	Free Kick Detection	Penalty Detection
Recall	0.9837	0.9661	0.9771
Precision	0.9931	0.9832	0.992
f-measure	0,9884	0.9746	0.9845

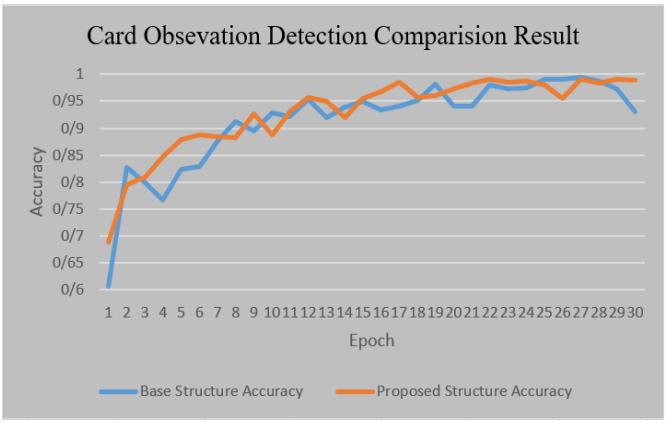
نمودارهای مربوط به مقایسه نتایج حاصل از معماری پیشنهادی برای شبکه عصبی عمیق در مقایسه با نتایج شبکه عصبی عمیق دنس‌نت پایه در شناسایی رویدادهای کارت زرد و قرمز، پنالتی و ضربه آزاد در شکل ۵ نشان داده شده است. براساس نتایج ارائه شده، معماری پیشنهادی برای شبکه عصبی عمیق با اختلافی اندک بهتر از نتایج معماری پایه دنس‌نت است، با این تفاوت که تعداد پارامترهای قابل آموزش در معماری پیشنهادی برای شبکه عصبی عمیق به میزان ۱۰.۳۱ درصد کاهش داشته است.

۵- نتیجه‌گیری

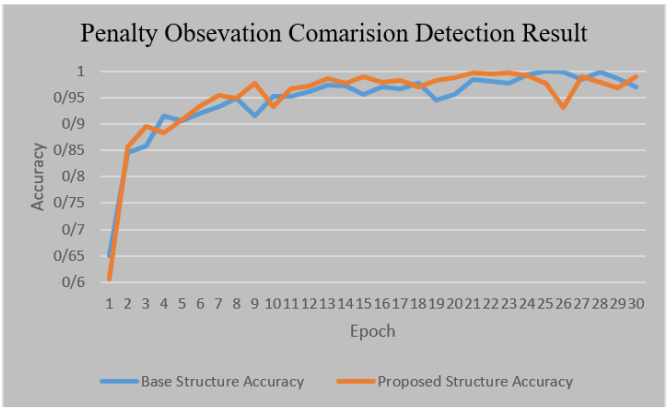
حجم بسیار فراوانی از ویدئوهای مختلف در دسترس کاربران در سراسر جهان قرار دارد. برخی از این ویدیوها مربوط به حوزه سرگرمی و برخی دیگر نیز مرتبط با حوزه‌های نظارتی و امنیتی هستند. از جمله این ویدئوها، ویدیوهای ورزشی و خصوصا ویدئوهای ورزش فوتبال است که به دلیل علاقه‌مندی طیف گسترده‌ای از مردم جهان به این ورزش دارای اهمیت بالایی است. علاوه بر موضوع علاقه‌مندی فوتبال دوستان، مسئله زمان طولانی بازی فوتبال است که در اغلب موارد همه مردم فرصت تماشای ۹۰ دقیقه بازی فوتبال را ندارد و البته علاقه‌مند هستند دستکم لحظات مهم و هیجان انگیز بازی را مشاهده کنند. به همین علت اخیرا برخی از فراهم‌کنندگان خدمات ارائه ویدئوهای ورزشی به خلاصه‌سازی بازی فوتبال پرداخته‌اند که با استقبال کاربران‌شان مواجه شده است.

تشخیص خودکار رخدادها و تفسیر معنایی صحنه‌ها، یک فرآیند چالش برانگیز در خلاصه‌سازی ویدیو بازی فوتبال است که نیازمند تحلیل هوشمند است. امروزه تکنیک‌های یادگیری عمیق بطور گسترده مورد استفاده قرار می‌گیرند و در بسیاری از مسائل مهم از جمله شناسایی و تشخیص رویداد در ویدیوی ورزش فوتبال، به نتایج خوب و قابل قبولی دست‌یافته‌اند. اما یکی از چالش‌های عمده استفاده از شبکه‌های

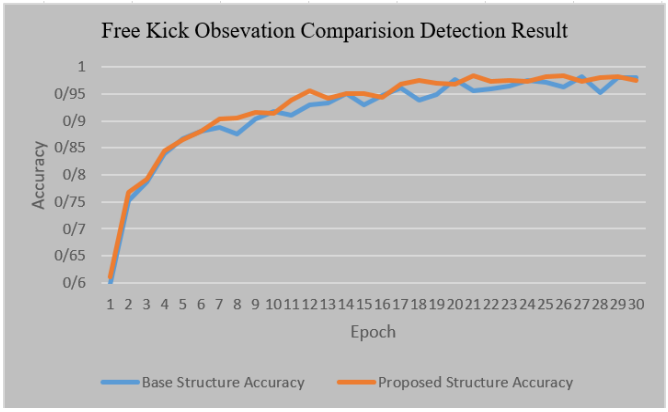
عصبی عمیق برای مدیریت و طبقه‌بندی تصاویر تعداد بسیار زیاد پارامترهای قابل آموزش در شبکه‌های عصبی عمیق است. در این مقاله ما معماری شبکه عصبی عمیق دنس‌نت را مورد بررسی قرار داده و با تغییر در معماری آن توانسته‌ایم هم دقت شبکه را افزایش و هم تعداد پارامترهای قابل آموزش شبکه را با هدف کاهش بار محاسباتی و زمان اجرا، به میزان قابل توجه‌ای کاهش دهیم.



(الف)



(ب)



(پ)

شکل ۵. مقایسه نتایج روش‌های معماری پیشنهادی برای شبکه عصبی عمیق و شبکه عصبی عمیق دنس‌نت ۱۲۱ لایه در تشخیص رویدادهای کارت زرد و قرمز (الف)، پنالتی (ب) و ضربه آزاد (پ).

۶- مراجع

[1] R. da Silva and S. R. Dahmen, "Universality in the distance between two teams in a football tournament," *Physica A: Statistical Mechanics and its Applications*, vol. 398, pp. 56–64, 2014.

[2] K. Bandyopadhyay, *Legacies of Great Men in World Soccer: Heroes, Icons, Legends*. Routledge, 2017.

[3] M. Fei, W. Jiang, and W. Mao, "Creating personalized video summaries via semantic event detection," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2018.

[4] J. Yu, A. Lei, and Y. Hu, "Soccer Video Event Detection Based on Deep Learning," in *International Conference on Multimedia Modeling*, 2019, pp. 377–389.

[5] M. N. Ali, M. Abdullah-Al-Wadud, and S.-L. Lee, "An efficient algorithm for detection of soccer ball and players," *Proc. 16th ASTL Control and Networking*, vol. 16, 2012.

[6] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews, "Large-scale analysis of soccer matches using spatiotemporal tracking data," in *2014 IEEE International Conference on Data Mining*, 2014, pp. 725–730.

[7] W. Dubitzky, P. Lopes, J. Davis, and D. Berrar, "The open international soccer database for machine learning," *Machine Learning*, vol. 108, no. 1, pp. 9–28, 2019.

[8] C. Cuevas, D. Quilón, and N. García, "Techniques and applications for soccer video analysis: A survey," *Multimedia Tools and Applications*, vol. 79, no. 39, pp. 29685–29721, 2020.

[9] Z. Wang, J. Yu, and Y. He, "Soccer video event annotation by synchronization of attack–defense clips and match reports with coarse-grained time information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 5, pp. 1104–1117, 2016.

[10] K. Tang, Y. Bao, Z. Zhao, L. Zhu, Y. Lin, and Y. Peng, "AutoHighlight: Automatic Highlights Detection and Segmentation in Soccer Matches," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 4619–4624.

[11] S. Jai-Andaloussi, A. Mohamed, N. Madrane, and A. Sekkaki, "Soccer video summarization using video content analysis and social media streams," in *2014 IEEE/ACM International Symposium on Big Data Computing*, 2014, pp. 1–7.

[12] P. Shi and X. Yu, "Goal event detection in soccer videos using multi-clues detection rules," in *Management and Service Science, 2009. MASS'09. International Conference on*, 2009, pp. 1–4.

[13] M.-L. Shyu, Z. Xie, M. Chen, and S.-C. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 252–259, 2008.

[14] M. H. Kolekar, "Bayesian belief network based broadcast sports video indexing," *Multimedia Tools and Applications*, vol. 54, no. 1, pp. 27–54, 2011.

[15] D. W. Tjondronegoro and Y.-P. P. Chen, "Knowledge-discounted event detection in sports video," *Ieee transactions on systems, man, and cybernetics-part a: Systems and humans*, vol. 40, no. 5, pp. 1009–1024, 2010.

[16] M. Y. Eldib, B. S. A. Zaid, H. M. Zawbaa, M. El-Zahar, and M. El-Saban, "Soccer video summarization using enhanced logo detection," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 4345–4348.

[17] B. Fakhar, H. R. Kanan, and A. Behrad, "Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model," *Multimedia Tools and Applications*, pp. 1–31, 2019.

[18] Z. Dang, J. Du, Q. Huang, and S. Jiang, "Replay detection based on semi-automatic logo template sequence extraction in sports video," in *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, 2007, pp. 839–844.

[19] H. Pan, P. Van Beek, and M. I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation," in *icassp*, 2001, pp. 1649–1652.

[20] H. M. Zawbaa, N. El-Bendary, A. E. Hassanien, and T. Kim, "Event detection based approach for soccer video summarization using machine learning," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 7, no. 2, pp. 63–80, 2012.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[1] Dense net
[2] Replay
[3] Long View Shot
[4] Coin change
[5] Islamic Azad University Soccer Dataset - https://sites.google.com/view/image-and-video-analysis/home
[6] Recall
[7] Precision
[8] F-measure
[9] Accuracy
[10] True Positive
[11] True Negative
[12] False Positive
[13] False Negative

امیرحسین زنگنه مدرک کارشناسی مهندسی کامپیوتر(مهندسی نرم افزار) را از دانشگاه هوایی شهید ستاری، تهران، ایران در سال ۱۳۸۸ دریافت کرد. مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه صنعتی امیرکبیر، تهران، ایران در سال ۱۳۹۰ دریافت نمود. وی هم اکنون در حال پیگیری مقطع دکتری در دانشگاه آزاد اسلامی – واحد تهران شمال، تهران، ایران می-باشد. زمینه های مورد علاقه وی شامل پردازش تصویر و فیلم و بینایی ماشین است. آدرس پست الکترونیکی ایشان عبارت است از:
amirhosein@aut.ac.ir

مهدی چم‌پور مدرک کارشناسی مهندسی کامپیوتر(مهندسی نرم افزار) را از دانشگاه دانشگاه شهید باهنر، کرمان، ایران در سال ۱۳۸۵ دریافت کرد. او مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه آزاد اسلامی، مشهد، ایران در سال ۱۳۸۸ دریافت نمود. ایشان مدرک دکتری را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه فنی گراتس، اتریش، ۲۰۱۶ در یافت نمود. ایشان مدرک پسادکتری را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران در سال ۱۳۹۶ در یافت نمود. زمینه های مورد علاقه ایشان شامل پردازش تصویر و بینایی ماشین است. آدرس پست الکترونیکی ایشان عبارت است از:
jampour@icg.tugraz.at

کامران لایقی مدرک کارشناسی مهندسی کامپیوتر (نرم افزار) را از دانشگاه بیرمنگام سیتی، بیرمنگام، انگلستان در سال ۱۹۸۰ دریافت کرد. ایشان مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه استون، بیرمنگام، انگلستان در سال ۱۹۸۲ دریافت نمود. ایشان مدرک دکتری را در رشته مهندسی کامپیوتر(هوش مصنوعی) از دانشگاه کیلی، نیوکاسل، انگلستان در سال ۱۹۹۳ دریافت نمود. زمینه های مورد علاقه ایشان شامل علوم شناختی و بینایی ماشین است. آدرس پست الکترونیکی ایشان عبارت است از:
K_layeghi@iau-tnb.ac.ir

The CSI Journal on Computing Science and Information Technology
Vol. 18, No.2, 2021

Event detection in video using optimal deep neural network

Amirhosein Zanganeh[1], Mahdi Jampour[2] , Kamran Layghi [3]

[1,3] Faculty of Computer and Electrical Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.
[2] Faculty of Computer and Electrical Engineering, Quchan University of Technology, Quchan, Iran.

Abstract

As one of the machine learning techniques, deep learning has utilized the technological improvements of graphics processing units (GPUs) and this in turn is the reason of its extensive use. Comparing to traditional methods, deep learning techniques obtained excellent results on many important problems such as event recognition in videos related to football. Using deep neural networks, one of the major challenges is the large number of parameters trained in this type of network when managing and classifying images, which leads to a high computational load and long time for training the deep neural network. The neural network Dense Net is considered as one of the latest networks presented for the object identification and recognition objectives. In this paper, the Dense Net deep neural network is used to identify and recognize the events of yellow and red cards, penalties and free kick as a basic architecture in videos related to football. The number and size of Dense Net network blocks affect the number of trainable parameters as well as the network accuracy. In this paper, it is aimed to reduce the number of trainable network parameters by changing the basic architecture of the Dense Net deep neural network while maintaining the accuracy. Therefore, it has been tried to propose an architecture for deep neural network by examining the possible states for deploying the blocks with different sizes in the Dense Net network. The evaluation results represent a significant reduction in the number of trainable parameters for the deep neural network, while increasing the accuracy of proposed architecture to identify and recognize the significant events in football.

Keywords: Event Detection, Yellow and Red Card Event, Penalty Event, Free Kick Event, Dense Deep Neural Network, Deep Learning, Video Summary.