20.1001.1.28212606.1403.7.1.1.9

فصلنامه علمی آماد و فناوری دفاعی، سال ششم، شماره بیست و یکم، بهار ۱۴۰۳

حذف ویژگی‌های مشترک برون کلاسی به منظور بهبود بازشناسی رویداد در تصویر

امیرحسین زنگنه*[۱]، محمود دی‌پیر[۲]، احسان شریفی[۳]

تاریخ دریافت:۱۴۰۲/۱۰/۰۴ تاریخ پذیرش: ۱۴۰۲/۱۲/۲۲

چکیده

امروزه به صورت گسترده برای نظارت و کنترل محیط از سامانه‌های نظارت و پایش تصویری استفاده می‌شود. هدف ما در این مقاله شناسایی و تشخیص رویداد در ویدیو است. ما به منظور شناسایی و تشخیص رویداد در ویدیو، ویدیوی ورزش فوتبال را که دارای پیچیدگی‌ها و چالش‌های فراوان است موردبررسی و آنالیز قرار داده-ایم در میان این چالش‌ها، می‌توان به خلاصه‌سازی، ردیابی، بازشناسی رویدادهای مهم بازی و غیره اشاره کرد، به‌عنوان‌مثال بازشناسی رویدادهایی مانند پنالتی و ضربه آزاد که دارای المان‌های بصری مشترک می‌باشند، دارای چالش بیشتری است. وجود المان‌های مشترک بین دو رویداد سبب استخراج ویژگی‌های مشترک و تفکیک‌ناپذیر در فرآیند بازشناسی این دو رویداد می‌شود. درنتیجه خطای بازشناسی و تفکیک این‌چنین رویدادهایی نسبت به سایر رویدادها بیشتر است. در این مقاله ما یک روش جدید برای حذف ویژگی‌های مشترک بین دو کلاس باهدف همگرا کردن ویژگی‌های درون کلاسی و واگرا نمودن ویژگی‌های برون کلاسی برای افزایش دقت دسته‌بندی و بازشناسی دو رویداد پنالتی و ضربه آزاد ارائه داده‌ایم. نتایج ارزیابی‌های انجام‌شده به وسیله روش پیشنهادی، حاکی از بهبود دقت بازشناسی و تفکیک دو رویداد پنالتی و ضربه آزاد با استفاده از روش پیشنهادی است و دقت شناسایی و تشخیص این دو رویداد به‌طور میانگین نسبت به شبکه عصبی عمیق پایه به میزان ۹,۰۸ درصد افزایش یافته است.

واژگان کلیدی: شبکه عصبی عمیق دنس‌نت، شبکه رزنت، ویژگی‌های مشترک، ویژگی‌های برون کلاسی، یادگیری عمیق

۱ دکتری هوش مصنوعی گروه مهندسی نرم افزار، دانشکده رایانه و سایبر، دانشگاه هوایی شهید ستاری، تهران ، ایران، (نویسنده مسئول) amirhosein@aut.ac.ir .

۲ دانشیار گروه مهندسی نرم افزار، دانشکده رایانه و سایبر، دانشگاه هوایی شهید ستاری، تهران ، ایرانmdeypir@ssau.ac.ir

۳ استادیار گروه مهندسی نرم افزار، دانشکده رایانه و سایبر، دانشگاه هوایی شهید ستاری، تهران ، ایرانsharifi@ssau.ac.ir

۱. مقدمه

با توجه به توسعه سریع فناوری تصویربرداری ویدیویی، امروزه به صورت گسترده برای نظارت و کنترل محیط از سامانه‌های نظارت و پایش تصویری استفاده می‌شود. برای جلوگیری از حوادث ناخواسته و حفاظت از اماکن نظامی و امنیتی، مردم و اموال آن‌ها، سرمایه‌گذاری بسیار زیادی در سامانه‌های نظارت تصویری انجام‌شده و هدف، استفاده حداکثری از تمام دستاوردهای فناورانه موجود در این زمینه برای توسعه سامانه‌های نظارت تصویری است. هدف از شناسایی و تشخیص اشیاء در سامانه‌های نظارت تصویری، دسته‌بندی و برچسب‌گذاری اشیا و تعیین موقعیت دقیق آن‌ها در تصویر یا ویدئو است.

استفاده از این سامانه‌ها موجب مراقبت و کنترل دقیق محیط، کاهش تخلفات، افزایش توانایی در آشکار سازی سریع حوادث و نظم دهی محیط کاری شده است. هدف ما در این مقاله شناسایی و تشخیص رویداد در ویدیو است. ما برای رسیدن به این هدف ویدیوی ورزش فوتبال را که دارای پیچیدگی‌های فراوان است موردبررسی و آنالیز قرار داده‌ایم.

شناسایی و تشخیص رویداد در ویدیوی ورزش فوتبال و نهایتاً خلاصه‌سازی مطلوب یک ویدئوی ورزشی مانند فوتبال فرآیند ساده‌ای نیست و نیازمند پردازش انسانی است چنانکه سراسر یک بازی فوتبال را مورد نظارت قرار داده و بخش های مهم آن را برچسب‌گذاری کند.

در روش‌های خلاصه‌سازی سنتی یک کاربر سراسر یک ویدیو را مورد نظارت قرار داده و بخش‌های مهم آن را برچسب‌گذاری می‌کند که کاری زمان‌گیر و دشوار است. با توجه به مشکلات موجود لزوم خلاصه‌سازی خودکار ویدیوها کاملاً محسوس است. در خلاصه‌سازی خودکار ویدیو با معرفی رویدادهای مهم و حساس، سامانه قادر است پس از دریافت یک ویدیو در ورودی با حذف افزونگی-های بصری و فریم‌های تکراری، یک کلیپ خلاصه‌شده از ویدیو که دربرگیرنده محتوای ویدیوی اصلی است را در اختیار کاربر قرار دهد.

در تحقیقات مختلف انجام‌شده برای خلاصه‌سازی ویدیو از ویژگی‌های مختلف صوتی و تصویری سطح پایین و سطح بالا استفاده شده است. روش‌های مبتنی بر ویژگی‌های بصری سطح پایین نمی-

توانند به خوبی مفاهیم موجود در ویدیو را مشخص و آشکار کنند و روش‌های چندمنظوره[1] که از ویژگی‌های صوتی، تصویری و رفتاری بیننده به‌طور هم‌زمان بهره می‌گیرند، مانند روش‌های مبتنی بر ویژگی‌های سطح بالا سعی در استخراج مفاهیم ویدیو با استفاده از الگوریتم‌های پیچیده و زمان‌بر دارند و دارای محدودیت‌هایی مانند نیاز به حسگرها و تجهیزات سخت‌افزاری اضافی برای ضبط صدا، محدودیت در فاصله ضبط داده‌ها، نیاز به حذف نویز و صداهای اضافی موجود در ویدیو و یا پردازش اولیه ویدیو توسط یک کاربر انسانی و غیره می‌باشند.

خلاصه‌سازی ویدئوی مبتنی بر رویداد که با توجه به ویژگی‌های بصری موجود در ویدیو است می‌تواند مشکلات یادشده در سایر روش‌های خلاصه‌سازی را از طریق پیش تعریف رویدادهایی که برای تشخیص رخداد در یک ویدیو مورداستفاده قرار می‌گیرند، حل کند. بر اساس شکل شماره ۱ در این روش ابتدا ویدیوی ورودی به شات‌های تشکیل‌دهنده‌ی خود تجزیه می‌شود و سپس در مرحله بعد هر شات بر اساس نوع رویدادی که شامل آن است توسط یک سامانه آموزش‌دیده شده دسته‌بندی می‌شود. این روش می‌تواند یک خلاصه ویدئویی مرتبط با رویداد را تولید کند.

شکل ۱. مراحل خلاصه‌سازی ویدیو توسط یک سامانه خودکار

شناسایی و تشخیص رویداد در ویدئوی ورزش فوتبال با چالش‌های متفاوتی روبرو است و این چالش‌ها بین رویدادهایی مانند پنالتی و ضربه آزاد که دارای المان‌های بصری مشترک می‌باشند، بیشتر

[1] Multimodal

است. وجود المان‌های مشترک بین دو رویداد سبب استخراج ویژگی‌های مشترک و غیر مؤثر در فرآیند شناسایی و تشخیص این دو رویداد می‌شود. درنتیجه خطای شناسایی و تفکیک این دو رویداد نسبت به سایر رویدادها بیشتر است. برای کاهش خطای شناسایی و تشخیص دو رویداد پنالتی و ضربه آزاد ما در این مقاله روشی را برای حذف ویژگی‌های مشترک بین دو کلاس باهدف نزدیک کردن ویژگی‌های درون کلاسی و دور کردن ویژگی‌های برون کلاسی دو رویداد یادشده ارائه کرده‌ایم.

در کاربردهای نظامی دفاعی و آفندی نیز برای افزایش دقت شناسایی و تشخیص اهدافی که باهم از نظر بصری و دیداری به هم نزدیک هستند، باهم شباهت‌های زیادی دارند، می‌توان از این روش استفاده کرد. همان‌طور که در شکل شماره ۲ نشان داده شده است، هواپیمای مسافربری بوئینگ ۷۴۷ (الف) با هواپیمای ترابری نظامی **C17** (ب) از نظر بصری و دیداری دارای شباهت‌های زیادی هستند و در سامانه‌های شناسایی و تشخیص مبتنی بر تصویر در صورتی که ما از همه ویژگی‌ها برای دسته‌بندی و تشخیص این دو هواپیما استفاده کنیم، با توجه به اشتراک بسیاری از ویژگی‌ها، با خطای زیادی روبرو خواهیم شد، درنتیجه برای کاهش خطا باید تا حد امکان ویژگی‌های مشترک بین این دو کلاس را حذف و از ویژگی‌های غیر مشترک مانند ویژگی‌های استخراج‌شده از دم و غیره هنگام شناسایی و تشخیص این دو نوع هواپیما استفاده کنیم.

امروزه با توجه به استفاده از تکنیک‌های پیشرفته جنگ الکترونیک امکان قطع ارتباط سامانه‌های دفاعی و آفندی با مرکز فرماندهی کنترل امکان‌پذیر است، درنتیجه سامانه‌های مورداستفاده در نیروهای مسلح باید قابلیت شناسایی و تشخیص اهداف با دقت بالا را به صورت خودکار و بدون نیاز به ارتباط با مرکز فرماندهی داشته باشند و در صورت وجود تشابه بین اهداف نظامی و غیرنظامی بتوانند با دقت بالا هدف نظامی را تشخیص دهند. یکی از روش‌های شناسایی و تشخیص اهداف استفاده از سامانه-های نظارت ویدیویی است که با توجه به ویژگی‌های بصری موجود در ویدیو، اقدام به شناسایی و تشخیص هدف می‌کنند.

شکل ۲. مقایسه تصویر هواپیمای مسافربری بوئینگ ۷۴۷ (الف) با هواپیمای ترابری نظامی C17 (ب).

ادامه این مقاله به شرح زیر سازمان‌دهی شده است: در بخش ۲، کارهای انجام‌شده در زمینه خلاصه سازی ویدیویی فوتبال موردبررسی قرار می‌گیرد، سپس در بخش ۳، روش پیشنهادی به‌تفصیل شرح داده می‌شود؛ در بخش ۴ نتایج تجربی ارائه شده است و درنهایت، نتیجه‌گیری در بخش ۵ ذکر شده است.

۲. پیشینه تحقیق

تشخیص خودکار رخدادها و تفسیر معنایی صحنه‌ها، یک کار چالش‌برانگیز در خلاصه‌سازی ویدیو بازی فوتبال است. به‌طورکلی روش‌های خودکار شناسایی و تشخیص خودکار رویداد در ویدیوی ورزش فوتبال را بر اساس ویژگی‌های مورداستفاده در آن‌ها می‌توان به دو دسته کلی تقسیم نمود. در دسته اول از ویژگی‌های مختلف مانند ویژگی صوتی، متون اینترنتی مرتبط با ویدیو یا ویژگی‌های بصری اضافه‌شده به ویدیو مانند صحنه‌های تکرار آهسته [۱] و تشخیص لوگو [۲] استفاده می‌شود، اما در دسته دوم از ویژگی‌های بصری موجود در ویدیو استفاده می‌شود [۳] که در ادامه مزایا و معایب هر روش را به‌تفصیل شرح خواهیم داد.

در روش‌های مبتنی بر استخراج ویژگی‌های ویدیو و اطلاعات خارجی مرتبط با ویدیو (روش‌های مالتی‌مُدال) از اطلاعات و داده‌های صوتی-تصویری موجود در ویدیو و همچنین اطلاعات و داده‌های مرتبط با ویدیو مانند متون اینترنتی یا اطلاعات موجود در شبکه‌های اجتماعی که مربوط به یک

ویدیوی مسابقه فوتبال هستند، برای شناسایی رویداد و نهایتاً خلاصه‌سازی ویدیو فوتبال بهره می-گیرند[۴].

در [۵] ویژگی‌های صوتی شامل تشویق تماشاگران و هیجان مفسران ورزشی را استخراج کرده‌اند، و هم‌زمان ویژگی‌های بصری را تشخیص دادند. بعد از استخراج مفهوم معنایی و توجه به توالی معنایی رویدادهایی که باهم در ارتباط هستند، مانند ورود توپ به دروازه و هلهله تماشاچیان، قوانین موجود برای شناسایی رویداد مورداستفاده قرار می‌گیرند. در کاری مشابه در [۶] نیز برای تجزیه و تحلیل محتوی ویدیو اقدام به استخراج ویژگی‌های سطح پایین و سطح میانی از کانال‌های صدا / تصویری کرده‌اند.

در [۷] با استفاده از یک شبکه بیزی روشی برای آنالیز معنایی ویدیو و خلاصه‌سازی ویدیو با شناسایی مفاهیم معرفی کردند که در آن، رویدادهای مهم بازی با استفاده از ویژگی‌های صوتی و استفاده از قوانین تولیدشده و دانش این حوزه، شناسایی می‌شوند. سپس مجموعه‌ای از کلیپ‌های برجسته که شامل رویدادهای حساس بازی هستند، برچسب‌گذاری شده و در یک خلاصه ویدئویی برای کاربردهای مختلف مانند مرور رویدادهای مهم، بازیابی و شاخص گذاری ویدیو بکار برده می‌شوند.

در [۸] برای شناسایی وقفه‌های ایجادشده در بازی، ویژگی‌های صوتی (صدای سوت داور) و تصویری ویدیو را استخراج کردند. در بازی فوتبال زمانی که سوت داور شنیده می‌شود به این معنی است که یک خطا اتفاق افتاده یا توپ از میدان بازی خارج بوده و درنتیجه یک وقفه در بازی رخ داده است. ازجمله مزایای روش ارائه‌شده، عمومی بودن و کاربردی بودن آن برای همه بازی‌های دارای ساختار بازی/ وقفه است.

روش‌هایی شناسایی رویداد مبتنی بر ویژگی‌های سمعی- بصری با محدودیت‌های ازجمله: ۱- افزایش تعداد حسگرها و تجهیزات سخت‌افزاری، ۲- محدودیت در فاصله ضبط داده‌ها، ۳- حذف نویز و صداهای اضافی موجود در ویدیو که می‌تواند موجب خطا در عملکرد سامانه شود، مواجه هستند.

در [۹] و [۱۰]برای یافتن کلمات کلیدی مرتبط با رویدادهای مهم بازی فوتبال مانند گل، ضربه آزاد، کارت و پنالتی محتوای متون اینترنتی را مورد جستجو قرار می‌گیرند، سپس اطلاعات مربوط به زمان و افراد مرتبط با رویداد را استخراج می‌کنند. با توجه به زمان ثبت رویداد، اقدام به تهیه یک خلاصه ویدیویی شامل زمان ثبت شده می‌کنند.

در [۱۱] برای شناسایی رویدادهای مهم و سپس خلاصه‌سازی ویدیوی ورزش فوتبال از محتوی موجود در شبکه‌های اجتماعی مانند توییتر استفاده کرده‌اند. با استخراج توییت‌های مرتبط با ویدیوهای ورزش فوتبال و سپس تحلیل و بررسی آن‌ها، اقدام به تشخیص و شناسایی رویدادهای مهم ویدیوی بازی فوتبال کرده‌اند.

استفاده از منابع اطلاعاتی اضافی مانند متون اینترنتی و اطلاعات موجود در شبکه‌های اجتماعی می‌تواند منجر به افزایش دقت در تشخیص و شناسایی رویدادهای مهم در ویدیو ورزش فوتبال شود، اما ازجمله اشکالات مهم این روش‌ها این است که اطلاعات مربوط به رویدادهای ورزشی با تأخیر زمانی توسط بینندگان و منتقدان در شبکه‌های اجتماعی منتشر می‌شوند و عملاً این روش‌های خلاصه‌سازی، تا زمان دریافت این اطلاعات با تأخیر روبرو می‌شوند. اما در روش‌های مبتنی بر ویژگی‌های دیداری منحصراً از ویژگی‌های دیداری موجود در فریم‌های ویدیو برای شناسایی رویداد، استفاده می‌شود.

هنگام خلاصه‌سازی ویدئوی ورزش فوتبال با چالش‌های مختلفی مواجه می‌شویم و این چالش‌ها و دشواری‌ها در مورد رویدادهایی که دارای عناصر بصری مشترک می‌باشند، بیشتر است[۱۲]. وجود المان‌های بصری مشترک بین دو رویداد موجب استخراج ویژگی‌های مشترک بین دو کلاس شده و درنتیجه فرآیند بازشناسی و تشخیص این دو رویداد با دقت پایینی انجام می‌شود. برای کاهش خطای بازشناسی و تشخیص دو رویداد پنالتی و ضربه آزاد ما در این مقاله با ارائه یک روش تحلیل داده‌های ورودی به شبکه عصبی، ویژگی‌های درون کلاسی برای دو کلاس ضربه آزاد و پنالتی را به هم نزدیک و ویژگی‌های برون کلاسی دو رویداد یادشده را از هم دور می‌کنیم که سبب بهبود بازشناسی تفکیک داده‌های دو کلاس می‌شود.

۳. روش پیشنهادی

شناسایی و تشخیص رویداد در ویدئوی ورزش فوتبال با چالش‌های متفاوتی روبرو است و این چالش‌ها بین رویدادهای که دارای المان‌های بصری مشترک می‌باشند، مانند رویداد پنالتی و ضربه آزاد، بیشتر است و درنتیجه خطای شناسایی و تفکیک این دو رویداد نسبت به سایر رویدادها بیشتر است. ما در بخش بعد روش حذف المان‌های بصری مشترک و نهایتاً ویژگی‌های مشترک بین دو رویداد پنالتی و ضربه آزاد را باهدف افزایش دقت شناسایی و تشخیص این دو رویداد ارائه می‌دهیم.

۳,۱ معماری مدل پایه

در این بخش دو شبکه عصبی عمیق دنس‌نت-۱۲۱ و معماری شبکه عصبی عمیق رزنت که به عنوان شبکه پایه برای بررسی روش پیشنهادی در بازشناسی دو رویداد پنالتی و ضربه آزاد مورداستفاده قرار داده‌ایم، را شرح می‌دهیم

۱-۲-۲-۴-۴- شبکه عصبی عمیق دنس‌نت

این شبکه عصبی عمیق سال ۲۰۱۷ توسط هانگ و همکاران در[۱۳] معرفی گردید. شبکه عصبی عمیق دنس‌نت، یکی از آخرین شبکه‌های عصبی ارائه‌شده برای اهداف شناسایی و تشخیص اشیاء است. از نظر معماری، این شبکه دارای معماری مشابه شبکه عصبی عمیق رزنت[1]بوده، اما دارای چند تفاوت اساسی است. این معماری نسبت به معماری‌های قبلی بر روی دیتابیس‌های CIFAR [۱۴] و SVHN [۱۵] دارای نرخ خطای کمتری است. همچنین این معماری نسبت به معماری رزنت روی به دیتابیس Imagenet تعداد پارامتر کمتری برای شناسایی اشیاء نیاز دارد درحالی‌که دقت دو روش تقریباً مشابه است[۱۶].

لایه‌های ابتدایی در شبکه‌های عصبی کانولوشن ویژگی‌های سطح پایین مانند لبه‌ها را استخراج و ویژگی‌های سطح بالا مانند بافت‌ها، و اشکال پیچیده و غیره توسط لایه‌هایی که در انتهای این زنجیره قرار دارند، استخراج می‌شوند. ویژگی‌های سطح پایین استخراج‌شده در عملیات طبقه‌بندی یک کلاس در مواردی ممکن است، از ویژگی‌های سطح بالای استخراج‌شده مهم‌تر و مؤثرتر باشند، درنتیجه با توجه به اتصال لایه‌هایی ابتدایی به لایه‌های انتهایی در معماری شبکه عصبی عمیق دنس‌نت، این شبکه می‌تواند یاد بگیرد که برای کلاس موردنظر فقط از ویژگی‌های سطح پایین، ویژگی‌های سطح بالا یا از ترکیب ویژگی‌های سطح پایین مانند لبه‌ها و ویژگی‌های سطح بالا مانند بافت‌ها استفاده کند.

۲-۲-۲-۴-۴- شبکه عصبی عمیق رزنت

شبکه عصبی دیگری که برای ارزیابی روش پیشنهادی، مورداستفاده و بررسی قرارگرفته شبکه رِزنت است که برای شناسایی و تشخیص دو رویداد پنالتی و ضربه آزاد مورداستفاده قرار گرفته است[۱۷]. معماری شبکه رزنت استفاده‌شده، دارای ۱۸ لایه است که شامل چهار بلوک رزیژوال با ساختار یکسان

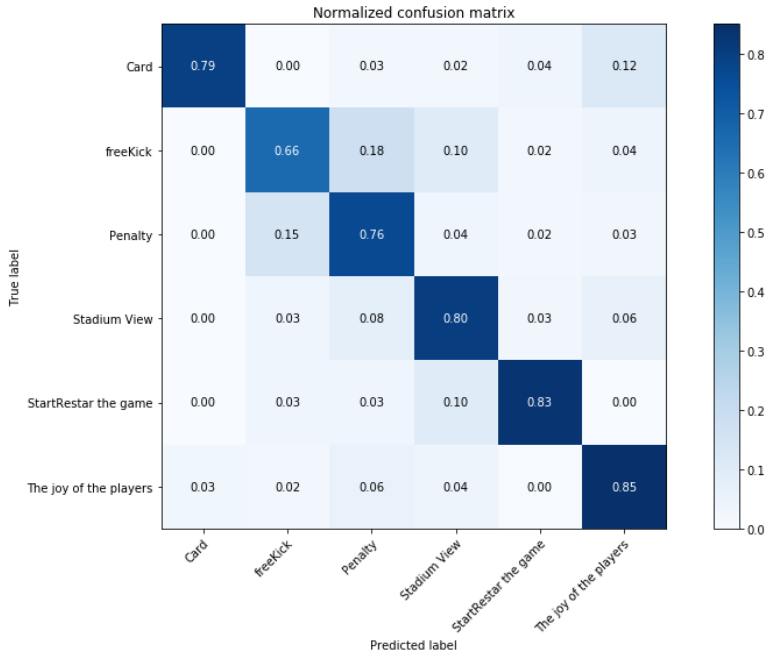
[1] ResNet

است. در مدل ارائه‌شده، لایه عادی‌سازی دسته‌ای[1] و تابع فعال‌سازی[2] ریلو قبل از لایه‌های کانولوشن بکار برده شده‌اند. همچنین در هر بلوک رزیژوال، تعداد مشابهی فیلتر بکار رفته است، به استثنای تعداد فیلترهای لایه کانولوشن آخر که برای حفظ پیچیدگی محاسباتی هر لایه، این تعداد دو برابر شده است. بدین ترتیب، تعداد فیلترها در بلوک‌های رزیژوال به ترتیب با عمیق‌تر شدن شبکه برابر با ۶۴، ۱۲۸، ۲۵۶ و ۵۱۲ است.

۳,۲ روش پیشنهادی برای حذف ویژگی‌های مشترک بین دو کلاس

با توجه به مشترک بودن برخی از المان‌های موجود در تصاویر ضربه آزاد و پنالتی، هنگامی‌که ما اقدام به استخراج ویژگی می‌کنیم، برخی از ویژگی‌های استخراج‌شده در هر دو کلاس پنالتی و ضربه آزاد مشترک است. ما برای نمایش توزیع ویژگی‌های استخراج‌شده برای دو کلاس پنالتی و ضربه آزاد از توابع چگالی احتمال استفاده کرده‌ایم. برای بررسی میزان تداخل شناسایی و تشخیص رویدادهای بازی فوتبال، ما مسئله شناسایی و تشخیص ۶ رویداد کارت، ضربه آزاد، پنالتی، نمای ورزشگاه، شروع بازی و شادی بازیکنان را به عنوان یک مسئله ۶ کلاسه بررسی کردیم و نتایج حاصل را در شکل شماره ۳ ارائه کرده‌ایم، همان‌طور که قابل مشاهده است دو رویداد ضربه آزاد و پنالتی دارای بیش‌ترین تداخل هستند و درنتیجه توزیع ویژگی‌های استخراج‌شده برای کلاس ضربه آزاد با توزیع ویژگی‌های استخراج‌شده برای کلاس پنالتی هم‌پوشانی داشته و ناحیه مشترک بین دو نمودار شامل ویژگی‌های هستند که در هر دو کلاس مشترک هستند که همپوشانی دو تابع بیانگر عدم توانایی ویژگی‌های استخراج‌شده در دسته‌بندی مطلوب تصاویر دو کلاس پنالتی و ضربه آزاد است.

[1] Batch normalization
[2] Activation function



شکل ۳. ماتریس اعوجاج رویدادهای ورزش فوتبال به ازای ۶ رویداد مهم کارت، ضربه آزاد، پنالتی، نمای ورزشگاه، شرع بازی و شادی بازیکنان.

ما به منظور محاسبه میزان همپوشانی توابع چگالی احتمال، روابط (۱) تا (۶) را بسط داده‌ایم. حاصل رابطه (۶) مقداری عددی در بازه صفر تا یک خواهد بود که هر چه مقدار آن کمتر باشد میزان همپوشانی دو تابع چگالی احتمال کمتر خواهد بود و نشان‌دهنده‌ی این است که مجموعه ویژگی‌های استخراج‌شده به ازای دو کلاس پنالتی و ضربه آزاد از هم دور شده‌اند که هدف نهایی ما است. بنابراین بر اساس توزیع نرمال دو تابع چگالی احتمال داریم:

$$\mu_{\phi(X_i)} = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i) \qquad (۱)$$

$$\mu_{\phi(Y_j)} = \frac{1}{n} \sum_{j=1}^{m} \phi(Y_j) \qquad (۲)$$

که در آن $\mathbf{X_i}$ مجموعه ویژگی‌های استخراج‌شده برای کلاس پنالتی، $\mathbf{Y_j}$ مجموعه ویژگی‌های استخراج‌شده برای کلاس ضربه آزاد، $\boldsymbol{\phi(X_i)}$ تابع چگالی احتمال ویژگی‌های استخراج‌شده برای کلاس پنالتی، $\boldsymbol{\phi(Y_j)}$ تابع چگالی احتمال ویژگی‌های استخراج‌شده برای کلاس ضربه آزاد، $\mu_{\phi(X_i)}$ میانگین تابع چگالی احتمال مجموعه ویژگی‌های استخراج‌شده کلاس پنالتی و $\mu_{\phi(Y_j)}$ تابع چگالی احتمال مجموعه ویژگی‌های استخراج‌شده کلاس ضربه آزاد می‌باشند. برای محاسبه فاصله ۲ تابع چگالی احتمال داریم:

$$D = |P - Q|^2 \quad (۳)$$

$$= \left|\mu_{\phi(X_i)} - \mu_{\phi(Y_j)}\right|^2 \quad (۴)$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n}\phi(X_i) - \frac{1}{m}\sum_{j=1}^{m}\phi(Y_j)\right|^2 \quad (۵)$$

که با بسط آن خواهیم داشت:

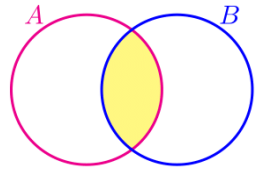
$$= \left(\frac{1}{n}\sum_{i=1}^{n}\phi(X_i) - \frac{1}{m}\sum_{j=1}^{m}\phi(Y_j)\right)^T \times \left(\frac{1}{n}\sum_{i=1}^{n}\phi(X_i) - \frac{1}{m}\sum_{j=1}^{m}\phi(Y_j)\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(X_i)^T\,\phi(X_j) + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\phi(Y_i)^T\,\phi(Y_j) - \frac{2}{n\,m}\sum_{i=1}^{n}\sum_{j=1}^{m}\phi(X_i)^T\,\phi(Y_j) \quad (۶)$$

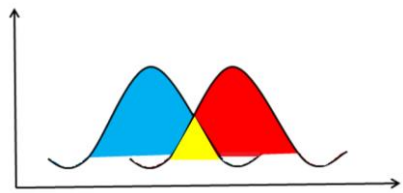
حاصل رابطه (۶) مقداری عددی در بازه صفرتا یک خواهد بود و بدیهی است هر چه مقدار همپوشانی ویژگی‌های استخراج‌شده به ازای دو کلاس پنالتی و ضربه آزاد، که از رابطه (۶) به دست می‌آید کمتر باشد یا به عبارتی به عدد ۰ نزدیک‌تر باشد، ویژگی‌های استخراج‌شده به ازای دو کلاس پنالتی و ضربه آزاد از هم دورتر و قابلیت بهتری در تفکیک دو کلاس پنالتی و ضربه آزاد خواهند داشت.

بر اساس شکل ۴، توزیع ویژگی‌های استخراج‌شده برای کلاس ضربه آزاد با توزیع ویژگی‌های استخراج‌شده برای کلاس پنالتی هم‌پوشانی داشته و ناحیه مشترک بین دو نمودار شامل ویژگی‌های

هستند که در هر دو کلاس مشترک هستند درنتیجه بر اساس رابطه ۶ مقدار عددی حاصل عددی نزدیک به یک خواهد بود ($D \rightarrow 1$) و در صورتی که از این ویژگی‌های برای دسته‌بندی تصاویر کلاس ضربه آزاد و پنالتی استفاده شود، موجب افزایش خطای دسته‌بندی می‌شوند.



الف



ب

شکل ۴. الف - فضای ویژگی‌های مربوط به دو کلاس پنالتی (A) و کلاس ضربه آزاد (B) و ویژگی‌های مشترک بین هر دو کلاس (رنگ زرد)

ب- توزیع نرمال ویژگی‌های مربوط به دو کلاس پنالتی و کلاس ضربه آزاد.

به عبارتی هدف ما حذف ویژگی‌های مشترک بین دو مجموعه باهدف افزایش دقت دسته‌بندی تصاویر است. برای این منظور باید دو مجموعه A و B از هم جدا یا به عبارتی باهم ناسازگار باشند. دو مجموعه از هم جدا یا باهم ناسازگار هستند اگر اشتراک بین دو مجموعه تهی باشد. درنتیجه:

$$A = (f_1, f_2, f_3, \dots, f_n) = \sum_{i=1}^{n} \odot P(A_i)$$

$$B = (f_1, f_2, f_3, \dots, f_n) \sum_{i=1}^{n} \odot P(B_i)$$

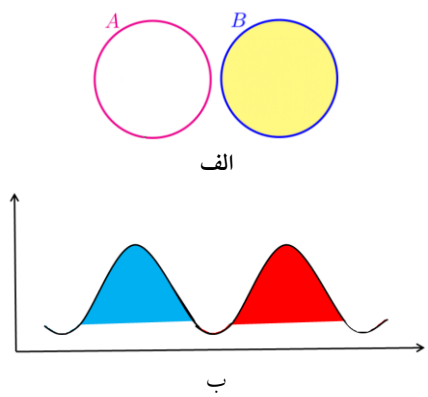
$$A \cap B = \{f \mid f \epsilon A \text{ and } f \epsilon B\}$$

در روابط بالا n تعداد پیکسل‌های هر تصویر است. A و B مجموعه ویژگی‌های استخراج‌شده برای هر کلاس و ناحیه مشترک بین دو مجموعه A و B شامل ویژگی‌هایی است که در هر دو مجموعه مشترک هستند و در فرآیند دسته‌بندی تأثیرگذاری مثبتی ندارند. درنتیجه برای حذف اشتراک بین دو مجموعه داریم[۱۸]:

$$\left.\begin{matrix}A\\B\end{matrix}\right\} \rightarrow A \cap B = \emptyset \leftrightarrow A \cup B = (A - B) \cup (B - A) \quad (۷)$$

با توجه به اینکه ویژگی‌های مشترک استخراج‌شده موجب افزایش خطای دسته‌بندی دو کلاس پنالتی و ضربه آزاد می‌شوند ما اقدام به حذف این ویژگی‌های مشترک کرده‌ایم، و در عمل با دور کردن ویژگی-های هر دو کلاس از یکدیگر، دقت دسته‌بندی تصاویر این دو کلاس را افزایش داده‌ایم.

همان‌طور که در شکل ۵، قابل‌مشاهده است توزیع ویژگی‌های استخراج‌شده برای کلاس ضربه آزاد از توزیع ویژگی‌های استخراج‌شده برای کلاس پنالتی تفکیک‌شده و نهایتاً با حذف ویژگی‌های مشترک بین دو کلاس، نمودار توزیع ویژگی‌های دو کلاس با یکدیگر هم‌پوشانی ندارند. درروش پیشنهادی آزاد برای شناسایی، تشخیص و تفکیک رویداد پنالتی و ضربه آزاد ما اقدام به حذف ویژگی‌های مشترک بین دو کلاس کرده‌ایم.



شکل ۵. الف - فضای ویژگی‌های مربوط به دو کلاس پنالتی (A) و کلاس ضربه آزاد (B)

ب- توزیع نرمال ویژگی‌های مربوط به دو کلاس پنالتی و کلاس ضربه آزاد.

ما برای استخراج ویژگی‌های کلاس پنالتی و ضربه آزاد از دو نقشه راهنما استفاده می‌کنیم. با توجه به وجود اشتراکات زیاد بین تصاویر پنالتی و ضربه آزاد، مانند دروازه، شناسایی و تشخیص این دو رویداد با خطای زیادی روبرو است. برای این منظور ما نواحی مشترک بین نقشه‌های راهنمای

مورداستفاده برای استخراج ویژگی را حذف کرده‌ایم و سپس از نقشه‌های راهنمای نهایی که دارای نواحی مشترک با کلاس دیگر نیستند، برای استخراج ویژگی‌ها استفاده کرده‌ایم. سپس در مرحله بعد نقشه راهنمای هر کلاس را در کلیه تصاویر آموزشی و آزمون آن کلاس ضرب نموده‌ایم و با این روش توانسته‌ایم نواحی ویژگی‌های مشترک حاصل از نواحی مشترک بین تصاویر پنالتی و ضربه آزاد را حذف کنیم.

جدول ۱. پارامترهای مربوط به پیاده‌سازی شبکه عمیق پیشنهادی

Parameter	Value
Optimizer	**Adam**
Loss function	**Binary cross-entropy**
Performance metric	**Accuracy**
Total Classes	**2 (Penalty and Free Kick)**
Batch Size	**32**
Epoch	**50**

۴. نتایج تجربی و آزمایش‌ها

در این بخش به ارزیابی و تحلیل روش پیشنهادی می‌پردازیم. در ابتدا مشخصات پایگاه داده تصاویر مورداستفاده را معرفی کرده و سپس روش پایه و روش پیشنهادی را مورد ارزیابی قرار می‌دهیم. همچنین در انتهای این بخش مقایسه‌ای با سایر کاره‌ای مشابه انجام پذیرفته است. پلتفرم مورداستفاده در پیاده‌سازی این تحقیق پایتون بوده و پارامترهای مربوط به شبکه عمیق پیشنهادی نیز در جدول شماره ۱ ارائه شده است.

۴,۱ پایگاه داده تصاویر

بر اساس تحقیقات انجام‌شده توسط نویسندگان مقاله، در حال حاضر تنها پایگاه داده جامع و کامل برای تحلیل اشیاء موجود در زمین فوتبال و تشخیص رویداد در ویدیوی ورزش فوتبال، که به صورت دسترسی رایگان برای امور تحقیقاتی وجود دارد، پایگاه داده IAUFD[1] است [۱۹]. سایر پایگاه داده‌های موجود اکثراً فقط شامل ویدئو هستند یا تعداد تصاویر در آن‌ها خیلی کم است. همچنین عدم

[1] Islamic Azad University Football Dataset

تنوع در تصاویر موجود در پایگاه داده‌ها، شرایط مختلف روشنایی، آب و هوا و غیره سبب کاهش جامعیت دیتا ست‌های موجود شده است [۲۰]. یک مقایسه اجمالی بین پایگاه داده‌های موجود و پایگاه داده مورداستفاده در جدول شماره ۲ ارائه شده است.

جدول ۲. مقایسه بین پایگاه داده‌های موجود به منظور تحلیل ویدیو ورزش فوتبال

ردیف	نام پایگاه داده	نوع داده	تعدادتصاویر	تعدادکلاس	قابلیت افزایش تعداد کلاس
۱	سوکرنت [۲۱]	ویدیو	–	۴	ندارد
۲	سوکر دی یی [۲۲]	ویدیو/تصویر	۵۵,۲۹۰	۳	ندارد
۳	سوکرنت ورژن ۲ [۲۳]	ویدیو	–	۱۷	ندارد
۴	پایگاه داده مورداستفاده (100k Soccer Images)	ویدیو / تصویر	۱۰۰,۰۰۰	۱۰	دارد

۴,۲ معیارهای ارزیابی روش پیشنهادی

ما به منظور ارزیابی عملکرد روش پیشنهادی از ۴ معیار ارزیابی شامل بازیابی[1](رابطه ۸)، وضوح[2] (رابطه ۹)، معیار-اف[3](رابطه ۱۰) و دقت[4](رابطه ۱۱) استفاده کرده‌ایم[۲۴]. همچنین بر اساس این پارامترها مشخصه عملکرد سامانه[5]را نیز محاسبه می‌کنیم.

$$Recall = \frac{TP}{TP+FP} \quad (۸)$$

[1] Recall
[2] Precision
[3] F-measure
[4] Accuracy
[5] ROC

$$Precision = \frac{TP}{TP+FN} \quad (۹)$$

که در آن TP[1] تعداد نمونه‌های مثبتی است که به درستی مثبت شناسایی‌شده‌اند، TN[2] تعداد نمونه‌های منفی که به درستی منفی شناسایی‌شده‌اند، FP[3] تعداد شناسایی‌های مثبت کاذب و FN[4] تعداد شناسایی‌های منفی کاذب می‌باشند. سپس مقدار معیار-اف f- و دقت به شرح زیر تعریف می‌شوند:

$$f-measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (۱۰)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۱۱)$$

۴،۳ ارزیابی و مقایسه روش ارائه‌شده

در این بخش، نتایج ارزیابی روش پیشنهادی در پیش‌پردازش داده‌های ورودی به شبکه عصبی عمیق باهدف افزایش دقت شناسایی و تشخیص دو رویداد پنالتی و ضربه آزاد، ارائه شده است. طی این ارزیابی از دو شبکه عصبی عمیق دنس‌نت و رزنت برای ارزیابی روش پیشنهادی استفاده کرده‌ایم. نتایج این ارزیابی که هدف آن شناسایی و تشخیص دو رویداد پنالتی و ضربه آزاد در تصاویر است، در جدول شماره ۳ ارائه شده است.

جدول ۳. دقت شناسایی رویدادهای پنالتی و ضربه آزاد با روش پیشنهادی

Method	Acuracy
ResNet-18	**62.79**
DenseNet-121	**59.50**

[1] True Positive
[2] True Negative
[3] False Positive
[4] False Negative

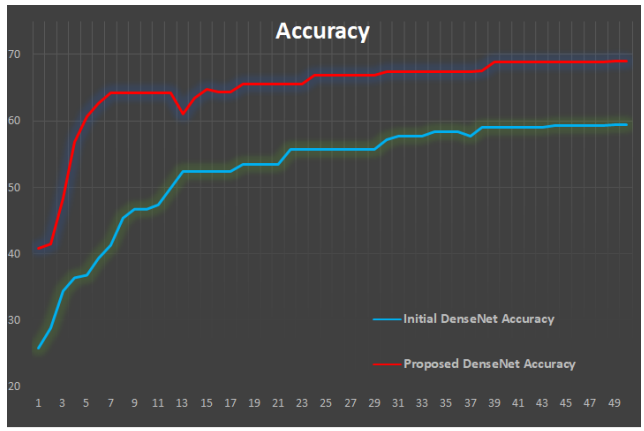
ResNet-18 with Proposed Method	**71.36**
DenseNet with Proposed Method	**69.08**

چنانکه مشاهده می‌شود دقت روش پیشنهادی نسبت به روش‌های پایه برای شناسایی رویداد پنالتی و ضربه آزاد بهبود یافته است. نتایج ارزیابی روش پیشنهادی روی شبکه‌های عصبی عمیق پایه رزنت و دنس نت بر اساس معیارهای ارائه‌شده در بخش ۴-۲ در جدول شماره ۴ ارائه شده است.

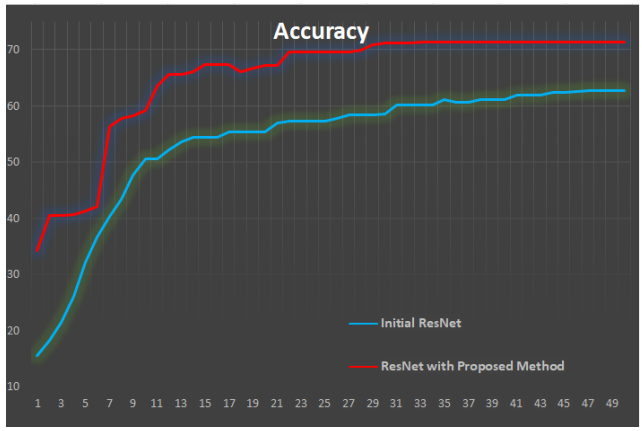
جدول ۴. نتایج روش پیشنهادی بر اساس معیارهای ارزیابی روی معماری‌های پایه دنس‌نت و رزنت.

Method	**ResNet**	**DenseNet**
TP	72/56	**70/45**
TN	70/02	**67/72**
FN	27/44	**29/55**
FP	29/98	**32/28**
Recall	0/7076	**0/6857**
Precision	0/7256	**0/7045**
F-Measure	0/7165	**0/695**

در شکل ۶ و ۷ نیز نمودار دقت روش پیشنهادی به ترتیب برای معماری‌های دنس‌نت و رزنت ارائه شده است. چنانکه در این تصاویر مشاهده می‌شود پس از مراحل ابتدایی آموزش و بعد از اپیک هفتم، نتایج روش پیشنهادی در شناسایی و تشخیص هردو رویداد پنالتی و ضربه آزاد نسبت به معماری‌های پایه مورداستفاده بهتر است و روش پیشنهادی توانسته به‌طور میانگین دقت شناسایی و تشخیص دو رویداد را به میزان ۹,۰۸ درصد افزایش دهد.



شکل ۶. مقایسه دقت روش پیشنهادی و معماری پایه دنس‌نت.



شکل ۷. مقایسه دقت روش پیشنهادی و معماری پایه رزنت.

نتیجه‌گیری

امروزه حجم بسیار فراوانی از ویدئوهای مختلف در اختیار کاربران در سراسر جهان قرار دارد. برخی از این ویدیوها مربوط به حوزه سرگرمی و برخی دیگر نیز مرتبط با حوزه‌های نظارتی و امنیتی می‌باشند. ازجمله این ویدئوها، ویدیوهای ورزشی و خصوصاً ویدئوهای ورزش فوتبال است که به دلیل علاقه‌مندی طیف گسترده‌ای از مردم جهان به این ورزش دارای اهمیت بالایی است.

شناسایی و تشخیص رویداد در ویدیوی ورزش فوتبال با چالش‌های بسیار زیادی روبرو است و این چالش‌ها هنگام شناسایی و تشخیص رویدادهایی مانند پنالتی و ضربه آزاد که دارای المان‌های بصری مشترک می‌باشند، بیشتر است. وجود المان‌های مشترک بین دو رویداد سبب استخراج ویژگی‌های

مشترک و غیر مؤثر در فرآیند شناسایی و تشخیص این دو رویداد می‌شود. درنتیجه خطای شناسایی و تفکیک این دو رویداد نسبت به سایر رویدادها بیشتر است.

ما در این مقاله برای کاهش خطای شناسایی و تشخیص دو رویداد پنالتی و ضربه آزاد، روشی را برای حذف المان‌های بصری مشترک و نهایتاً حذف ویژگی‌های مشترک بین دو کلاس ضربه آزاد و پنالتی ارائه داده‌ایم که بر اساس آن ویژگی‌های درون کلاسی برای دو کلاس ضربه آزاد و پنالتی به هم نزدیک و ویژگی‌های برون کلاسی دو رویداد یادشده از هم دور می‌شود. بر اساس نتایج ارائه‌شده، با استفاده از روش پیشنهادی دقت شناسایی و تشخیص این دو رویداد به‌طور میانگین نسبت به شبکه عصبی عمیق پایه به میزان ۹,۰۸ درصد افزایش یافته است.

مراجع

[1] S. Sarkar, S. Ali, and A. Chakrabarti, "Shot classification and replay detection in broadcast soccer video," *Advanced Computing and Systems for Security: Volume Twelve*, pp. 57–66, 2020.

[2] J. O. Valand *et al.*, "Automated clipping of soccer events using machine learning," in *2021 IEEE International Symposium on Multimedia (ISM)*, IEEE, 2021, pp. 210–214.

[3] L. F. K. Tani, A. Ghomari, and M. Y. K. Tani, "A semi-automatic system of web videos annotation and retrieval: application to events detection in soccer domain," *International Journal of Computer Aided Engineering and Technology*, vol. 16, no. 4, pp. 512–533, 2022.

[4] 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.

[5] 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*, IEEE, 2009, pp. 1–4.

[6] 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.

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

[8] 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.

[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)*, IEEE, 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*, IEEE, 2014, pp. 1–7.

[12] A. Zanganeh, E. Sharifi, and M. Jampour, "Converge intra-class and Diverge inter-class features for CNN-based Event Detection in football videos," in *2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, IEEE, 2023, pp. 1–6.

[13] 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.

[14] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.

[15] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, Springer, 2016, pp. 630–645.

[18] R. P. Grimaldi, *Discrete and combinatorial mathematics, 5/e*. Pearson Education India, 2006.

[19] A. Zanganeh, M. Jampour, and K. Layeghi, "IAUFD: A 100k images dataset for automatic football image/video analysis," *IET Image Processing*, vol. 16, no. 12, pp. 3133-3142 (2022)., doi: 10.1049/ipr2.12543.

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

[21] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "Soccernet: A scalable dataset for action spotting in soccer videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1711–1721.

[22] C. Wang and H. Liu, "Comprehensive Soccer Video Understanding: Towards Human-comparable Video Understanding System in Constrained Environment," *arXiv preprint arXiv:1912.04465*, 2019.

[23] A. Deliege *et al.*, "Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4508–4519.

[24] A. Zanganeh and M. Jampour, "Automatic Weak Learners Selection for Pattern Recognition and its application in Soccer Goal Recognition," in *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, IEEE, 2019, pp. 240–245.