بهبود بازشناسایی شخص با استفاده از یادگیری انتقالی و شبکه‌های سیامی

سجاد عموئی ششکل [۱]، کاظم فولادی قلعه [۲]، حسین آقابابا [۳]

[۱] دانش‌آموخته کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران
sajad.amouei@ut.ac.ir
[۲] استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران؛ سرپرست آزمایشگاه پژوهشی یادگیری عمیق دانشگاه تهران
kfouladi@ut.ac.ir
[۳] دانشیار گروه مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران
aghababa@ut.ac.ir

چکیده

بازشناسایی شخص در جامعه تحقیقاتی محبوبیت بالایی به‌دست آورده و دلیل آن افزایش کاربردها و اهمیت آن در صنعت نظارت است. بازشناسایی شخص به‌دلیل وجود تغییرات درون‌کلاسی و بین‌کلاسی در دوربین‌های مختلف همچنان به‌عنوان یک مسئله‌ی چالشی مورد بررسی قرار می‌گیرد. در این مقاله یک شبکه از نوع سیامی معرفی می‌شود که جفت تصاویر را دریافت کرده و سپس با استفاده از شبکه پیش‌آموزش داده شده، ویژگی‌های تصاویر را استخراج می‌کند و در نهایت خروجی توسط یک تابع اتلاف تصدیق تعیین می‌شود. به‌منظور به‌دست آوردن ویژگی‌های عمیق‌تر از تصاویر عابران پیاده، از شبکه پیش‌آموزش داده شده EfficientNet B0 برای استخراج ویژگی‌ها استفاده کردیم. آزمایش‌ها را روی مجموعه داده CUHK01 برای نشان دادن دقت روش پیشنهادی انجام دادیم. دقت روش پیشنهادی در رتبه ۱، رتبه ۵، رتبه ۱۰، رتبه ۱۵ و رتبه ۲۰ به ترتیب ٪۷۰، ٪۹۵، ٪۹۹، ٪۹۹ و ٪۹۹ می‌باشد. نتایج نشان می‌دهد که روش ارایه شده نسبت به روش‌های به‌روز دارای عملکرد بهتری است.

کلمات کلیدی: بازشناسایی شخص، شبکه سیامی، اتلاف تصدیق، EfficientNet B0.

۱ مقدمه

معمولاً مسئله بازشناسایی شخص به‌عنوان یک مسئله بازیابی تصویر بررسی می‌شود، که هدف آن تطبیق عابرپیاده در چند دوربین است [۱، ۲، ۳]. تصویر عابرپیاده موردنظر به عنوان پرس‌وجو داده می‌شود، و بازشناسایی شخص تعیین می‌کند که این عابرپیاده در کدام یک از دوربین‌های دیگر مشاهده شده است

[۴]. اخیراً در حوزه تحقیقاتی بازشناسایی شخص پیشرفت‌های قابل توجهی انجام شده است که یکی از دلایل آن ایجاد مجموعه داده‌های بزرگتر از تصاویر عابران پیاده است. دلیل دیگر این پیشرفت‌ها مربوط به یادگیری توصیف‌گرهای عابرپیاده توسط شبکه‌های عصبی کانوولوشنال (CNN) است. علیرغم پیشرفت‌های صورت‌گرفته توسط محققان بینایی کامپیوتر، چالش‌های حل‌نشده بسیاری در این حوزه وجود دارد [۱].

مدل‌های بازشناسایی شخص بسیاری توسعه داده شده که از ویژگی‌های سطح پایین مانند رنگ [۵]، بافت و ساختار مکانی [۶] استفاده می‌کردند. این ویژگی‌های بصری در مقابل تغییرات نور، زاویه دید و عدم تراز مقاوم نبودند. درک انسانی، تفاوت افراد را به‌وسیله ویژگی‌های سطح بالا مانند موی بلند، رنگ پیراهن، رنگ کوله‌پشتی و غیره به‌آسانی تشخیص می‌دهد. این خصوصیت می‌تواند بازنمایی از ویژگی‌های معنایی سطح بالا یک شخص را استخراج کند و در برابر تغییرات نور و عدم تراز تصویر نسبت به ویژگی‌های سطح پایین مقاوم‌تر است. برای این منظور نیاز است که داده‌ها برچسب‌گذاری شوند که هزینه بسیار زیادی دارد. در نتیجه دستیابی به مجموعه داده آموزشی کافی دارای برچسب خصوصیت انسانی بسیار دشوار است.

زمانی که تعداد داده‌های آموزشی به مقدار زیادی موجود است، انتقال بازنمایی یادگرفته شده از مجموعه داده بزرگ اهمیت ویژه‌ای پیدا می‌کند. اکنون از یادگیری انتقالی در اکثر کارهای مربوط به بازشناسایی شخص استفاده شده است. در مسئله بازشناسایی شخص تعداد تصاویر برچسب‌گذاری شده به چند صد عدد می‌رسد و روش‌های موجود عموماً از مدل‌های پیش آموزش داده شده روی مجموعه داده‌های بزرگتر استفاده می‌کنند و سپس مدل را روی مجموعه داده هدف تنظیم می‌کنند. زمانی که از مجموعه داده بازشناسایی شخص بزرگتر برای یادگیری انتقالی استفاده می‌کنیم، ممکن است تفاوت زیادی در زاویه دوربین و شرایط تصویربرداری وجود داشته باشد. در نتیجه مدل‌هایی که از یادگیری انتقالی استفاده می‌کنند بهبود عملکرد کم یا دارای عملکرد منفی می‌شوند [۱].

روش ارایه شده برای استخراج ویژگی‌های سطح بالا از CNN استفاده می‌کند. در سال‌های اخیر معماری‌های CNN بهبود عملکرد چشمگیری در حل وظایف بینایی ماشین از خود نشان دادند. همچنین مطالعاتی در زمینه ویژگی‌های به‌دست آمده توسط CNN انجام شد. ویژگی‌هایی که توسط CNNها بدست می‌آید دارای ساختار سلسله مراتبی است. ویژگی‌های به‌دست آمده توسط لایه‌های پایینی CNN مشابه ویژگی‌های استخراج شده سطح پایین مانند فیلترهای رنگ و لبه است. لایه‌های بالاتر CNN ویژگی‌های بسیار متفاوت وسطح بالایی را استخراج می‌کنند که شبکه می‌تواند با این ویژگی‌ها کلاس مورد نظر را تشخیص دهد [۷]. شبکه پیشنهادی ما می‌تواند ویژگی‌های سطح بالا توسط لایه‌های بالایی CNN استخراج کند. شبکه پیشنهادی از مدل پیش‌آموزش داده شده EfficientNet B0 [۸] بهره می‌برد که روی مجموعه داده ImageNet آموزش داده شده است.

اگر تابع تطبیق مناسبی برای محاسبه فاصله بین ویژگی‌ها تعیین شود، استخراج کننده ویژگی می‌تواند ویژگی‌های متمایز کننده‌ای از بازنمایی شخص را آموزش ببیند. تعدادی از روش‌های موجود توزیع مشابهت روی جفت تصاویر پرس و جو و گالری براساس نقشه ویژگی[1] آنها می‌سازند. در روش ما شبکه محدود به

[1] Feature map

شکل ۱: نمونه‌ای از پرس‌وجو و تصاویر بازیابی شده روی مجموعه داده CUHK01

مقایسه نقشه ویژگی‌ها متناظر دو تصویر نیست و با مقایسه ویژگی‌های سطح بالا و ترکیب آنها دارای استراتژی جدیدی است. روش ارایه شده از چندین نقشه ویژگی برای مقایسه دو تصویر استفاده می‌کند. علاوه بر این، ارتباط بین این ویژگی‌ها را به‌صورت داده‌محور بررسی می‌کند.

در این مقاله هدف ما ایجاد یک شبکه انتها به انتها است که به هر دو تصویر عابرپیاده ورودی به شبکه، یک امتیاز مشابهت تعیین کند. مثالی از پیش‌بینی شبکه در شکل ۱ نشان داده شده است. روش ارایه شده از شبکه پیش‌آموزش داده شده EfficientNet B0 به‌عنوان استخراج‌کننده ویژگی‌های سطح بالا بهره می‌برد که می‌تواند با استفاده از آن رابطه بین‌کلاسی[۲] و درون‌کلاسی[۳] در ویژگی‌های عمیق سطح بالا را پیدا کند. براساس مشاهدات، شبکه پیش‌آموزش داده EfficientNet B0 به‌وسیله‌ی متعادل کردن عمق، عرض و وضوح به کارآیی بهتری دست می‌یابد. EfficientNet B0 از یک روش مقیاس‌گذاری ترکیبی برای ثابت نگه داشتن نسبت‌ها در سه بُعد بهره می‌برد که منجربه تقویت محاسبات و همچنین دقت می‌شود [۸]. امتیاز مشابهت به‌وسیله‌ی تجزیه و تحلیل رابطه بین ویژگی‌های استخراج شده محاسبه می‌شود. مقدار اولیه پارامترهای شبکه پیش‌آموزش داده شده براساس مجموعه داده ImageNet مقداردهی شده است. برای تنظیم پارامترها باید آن را روی مجموعه داده آموزشی بازشناسایی شخص مجدداً آموزش دهیم که باعث می‌شود پارامترهای شبکه طبق مسئله به‌روزرسانی شود و بتواند ویژگی‌های متمایزکننده از تصاویر عابران پیاده را به‌خوبی استخراج کند.

[2] Inter-Class
[3] Intra-Class

۲ کارهای گذشته

به‌طور کلی فرآیند بازشناسایی شخص شامل دو بخش است: یک بخش مربوط به استخراج ویژگی‌ها از تصاویر ورودی و بخش دیگر معیار مشابهت برای مقایسه ویژگی‌های سراسر تصویر. هدف اصلی جست‌وجو برای یافتن بازنمایی ویژگی‌ها، یافتن ویژگی متمایزکننده است که در مقابل تغییرات شرایط نور، حالت شخص و زاویه دوربین مقاومت بیشتری داشته باشد. روش‌های اولیه به ویژگی‌های دستی طراحی شده مانند HSV هیستوگرام رنگ [۶]، LBP و ویژگی‌های Garbo [۹]، SIFT [۲] و غیره متکی بودند. در کنار این ویژگی‌ها، از معیارهای مشابهت مانند یادگیری معیار Mahalanobis [۱۰]، LADF [۱۱] و مسافت‌های وزنی برابری [۱۲] و غیره استفاده می‌شد.

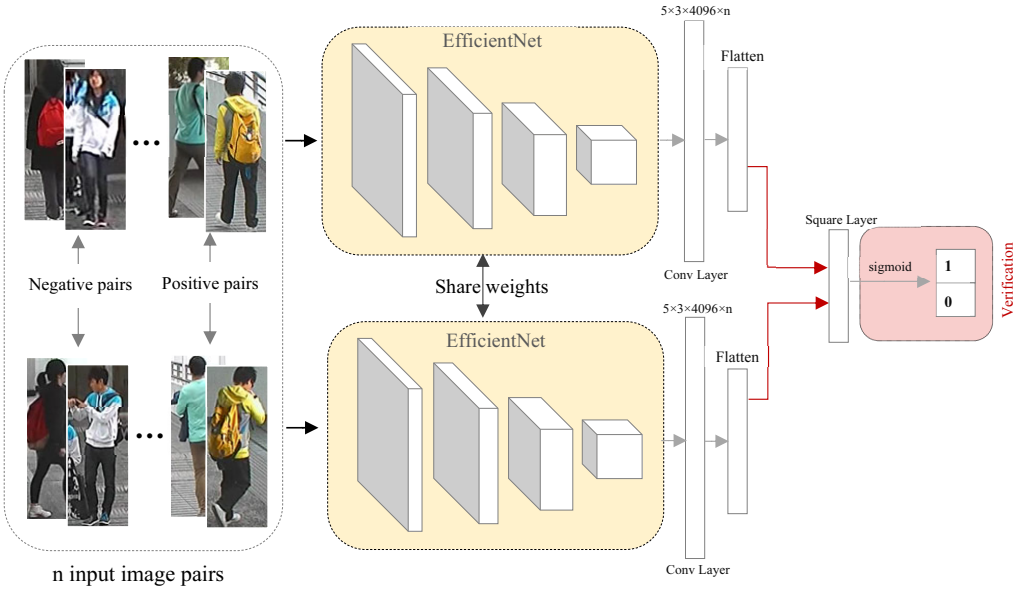
در سال‌های اخیر، عملکرد خوب یادگیری عمیق دلیلی شد تا بسیاری از محققان از آن برای به‌دست آوردن ویژگی‌های ظاهری و معیارهای فاصله برای بازشناسایی شخص استفاده کنند [۴]، [۱۳]، [۱۴]. روش یادگیری معیار عمیق [۱۵]، تصویر ورودی را به سه قسمت افقی بخش‌بندی می‌کند و این بخش‌ها از دولایه کانوولوشنال و یک لایه تمام متصل عبور می‌کند و در خروجی یک بردار برای تصویر به‌دست می‌آید. مشابهت دو بردار خروجی توسط فاصله کسینوسی محاسبه می‌شود. معماری FPNN [۴] با داشتن یک لایه تطبیق تکه نسبت به معماری قبل متفاوت است، که پاسخ‌های کانوولوشنال دو تصویر در راه‌راه‌های افقی متفاوت ضرب می‌کند. روش ImprovedReID [۱۳] مدل FPNN به‌وسیله محاسبه ویژگی‌های اختلاف همسایگی متقابل ورودی بهبود داده است، که ویژگی‌های تصویر ورودی با ویژگی‌های محلی همسایه تصویر دیگر مقایسه می‌کند. با این حال ممکن است استراتژی‌های تطبیق با کارآیی محاسباتی کم یا محدودیت اطلاعات ساختار مکانی مواجه شوند.

۳ روش ارایه شده

۱.۳ نمای کلی

معماری شبکه بازشناسایی شخص ارایه شده در شکل ۲ به تصویرکشیده شده است. این شبکه برپایه یک مدل کانوولوشنال سیامی است که از اتلاف تصدیق بهره می‌برد. هدف از مدل ارایه شده، یادگیری بازنمایی ویژگی‌های محلی جفت تصاویر ورودی است که با استفاده از آنها امتیاز مشابهت تعیین می‌شود یا ویژگی‌های متمایزکننده برای طبقه‌بندی تصاویر ورودی مربوط به کلاس مختلف را یاد می‌گیرد. این شبکه دارای اتلاف تصدیق است. ابتدا تصاویر برای استخراج ویژگی‌ها وارد شبکه می‌شوند. بعد از آخرین لایه مدل پیش‌آموزش داده شده، یک توصیف‌کننده ویژگی N بُعدی قرار دارد. سپس ویژگی‌های استخراج شده سطح بالا برای مقایسه وارد یک لایه مربع[۴] بدون پارامتر می‌شوند. این لایه ماتریس را به‌عنوان ورودی گرفته و بعد از تفریق و مربع کردن، یک ماتریس به عنوان خروجی می‌دهد. لایه مربع به صورت $f_s = (f_1 - f_2)^2$ نوشته می‌شود. جایی که f_1 و f_2 ویژگی‌های سطح بالا استخراج شده از جفت تصاویر ورودی می‌باشند و f_s ماتریس خروجی

[4] Square layer



شکل ۲: معماری روش پیشنهادی

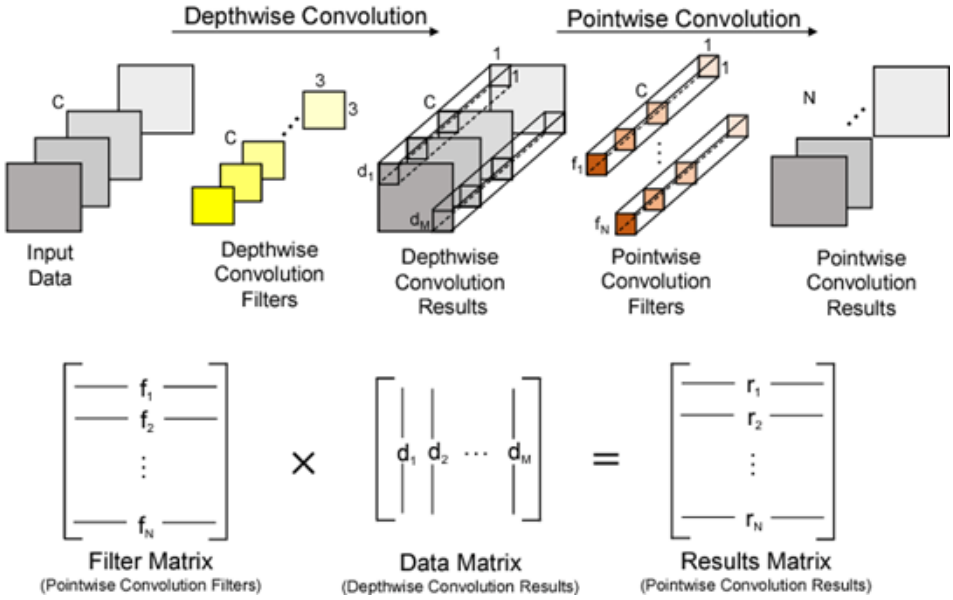
است. در ادامه یک لایه حذف تصادفی (برون‌اندازی)[5] برای جلوگیری از بیش‌برازش شبکه قرار دارد. در انتهای شبکه تابع سیگموید قرار دارد که تشابه یا عدم تشابه جفت تصویر ورودی را تعیین می‌کند.

۲.۳ مدل پیش‌آموزش داده EfficientNet B0

طبق مشاهدات با اعمال تعادل بین همه بُعدهای شبکه، در کارآیی و دقت آن بهبود حاصل می‌شود. در EfficientNet B0 برای بهبود عملکرد CNNها، در سه بُعد عرض، عمق و وضوح از مجموعه ثابت استفاده می‌شود که این ضرایب مقیاس‌گذاری ثابت برخی از محدودیت‌های خاص را برآورده می‌کند. در شکل ۳ بازنمایی از کانوولوشنال عمقی و نقطه‌ای استفاده شده در EfficientNet B0 ترسیم شده است. این شبکه پیش‌آموزش داده در مجموع دارای ۱۸ لایه کانوولوشنال است که $D = ۱۸$ و هر لایه دارای هسته‌های $k(۳,۳)$ یا $k(۵,۵)$ است. تصاویر ورودی شامل سه کانال رنگ R, G, B است. این شبکه روی تصاویر با ابعاد ۲۲۴ × ۲۲۴ آموزش دیده است. ابعاد مجموعه تصاویر بازشناسایی شخص ۸۰ × ۱۶۰ می‌باشد که شبکه EfficientNet B0 را در این ابعاد تصاویر تنظیم می‌کنیم تا یادگیری انجام شود. لایه‌های بعدی برای کوچک کردن اندازه نقشه ویژگی، مقیاس وضوح را کاهش می‌دهند ولی برای افزایش دقت، مقیاس عرض را افزایش می‌دهند. برای نمونه لایه کانوولوشنال دوم شامل $w = ۱۶$ فیلتر است، و تعداد فیلترهای لایه بعدی $w = ۲۴$ است.

روش متداول، استفاده از هسته‌های $k(۳,۳)$، $k(۵,۵)$ یا $k(۷,۷)$ است. [۱۶]. با این حال هسته‌های

[5] Dropout layer



شکل ۳: بازنمایی از کانوولوشن عمقی و نقطه‌ای [۱۸]

بزرگ می‌توانند باعث بهبود دقت و کارآیی مدل شوند. علاوه بر این، هسته‌های بزرگ به ضبط الگوهای با وضوح بالا کمک می‌کنند، در حالی که هسته‌های کوچک باعث بهتر استخراج شدن الگوهای با وضوح پایین می‌شوند [۱۷].

۳.۳ یادگیری انتقالی

برای آموزش شبکه‌های عصبی نیاز به جمع‌آوری مقدار کافی داده است. دستیابی به مقدار زیادی از داده‌ها دشوار می‌باشد، زیرا جمع‌آوری داده‌های برچسب‌دار نیاز به زمان و هزینه زیادی دارند و احتمال خطا در آنها بالا است [۱۹]. برای این منظور، یادگیری انتقالی به عنوان روشی مؤثر برای انتقال دانش استخراج شده از یک منبع به دامنه هدف در نظر گرفته می‌شود [۱۹]. در این روش، یادگیری انتقالی به ما این امکان را می‌دهد که از پارامترهای موجود و وزن‌های لایه کانوولوشن از یک مدل یادگیری شده روی مجموعه داده بزرگ برای مدل جدید خود با مجموعه داده کوچک استفاده کنیم. ما از وزن‌های مدل پیش آموزش داده شده روی مجموعه داده ImageNet استفاده کردیم که به دلیل داشتن تصاویر عمومی و محیط می‌تواند در مسئله بازشناسایی شخص مورد استفاده قرار بگیرد و نتایج خوبی بدهد.

۴.۳ اتلاف تصدیق

همان‌طور که در شکل ۲ مشاهده می‌شود، از لایه مربع برای مقایسه ویژگی‌ها استفاده شده است. در این شبکه، لایه مربع f_1 و f_2 را به عنوان ورودی دریافت کرده و f_s به عنوان خروجی لایه مربع می‌دهد. لایه

مربع به صورت زیر نمایش داده می‌شود:

$$f_s = (f_1 - f_2)^2 \quad (1)$$

برای حل مسئله تصدیق عابرپیاده همانند مسئله طبقه‌بندی دودویی رفتار می‌کنیم و برای احتمال پیش‌بینی شده از اتلاف آنتروپی متقاطع به صورت زیر استفاده می‌کنیم:

$$\hat{q} = \text{softmax}(\theta_s^* f_s) \quad (2)$$

$$\text{verify}(f_1, f_2, s, \theta_s) = \sum_{i=1}^{2} -q_i \log(\hat{q}_i) \quad (3)$$

جایی که ابعاد f_1 و f_1 برابر ۴۰۹۶ × ۱ × ۱، s کلاس هدف (یکسان/ متفاوت)، θ_s پارمترهای لایه کانوولوشنال و $\hat{q}$ احتمال پیش بینی شده است. اگر جفت تصویر ورودی مربوط به یک شخص باشد، $q_1 = 0, q_2 = 1$ در غیر اینصورت برابر با $q_1 = 0, q_2 = 1$ است.

۴ نتایج پیاده‌سازی

۱.۴ مجموعه داده

آزمایش‌های ما روی مجموعه داده CUHK01 انجام شده است. این مجموعه داده در دو زاویه دوربین، ضبط و جمع‌آوری شده است. مجموعه داده شامل تصاویر گرفته شده از ۹۷۱ شخص است و هر شخص دو تصویر از دوربین A و دو تصویر از دوربین B دارد. دوربین A از زاویه روبه‌روی شخص و دوربین B از زاویه نیمرخ شخص تصویر ضبط می‌کند.

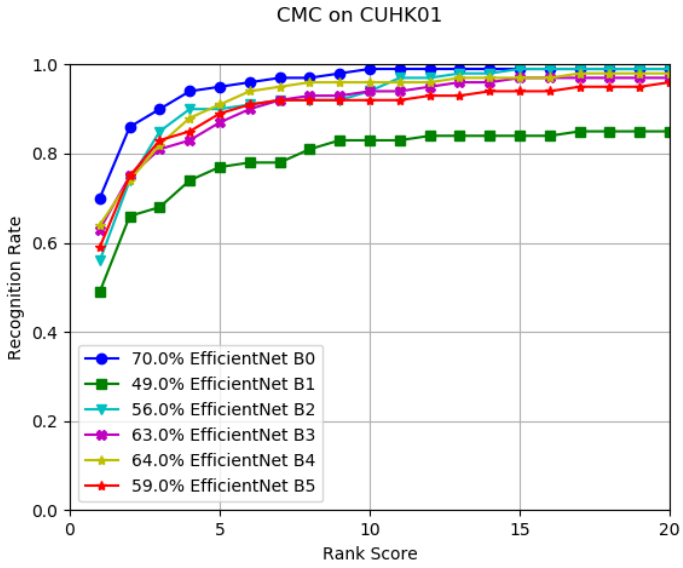
۲.۴ تنظیمات یادگیری

در فرآیند آموزش جفت تصاویر به دسته‌های ۴۸تایی به شبکه داده می‌شوند. از کاهش گرادیانی به‌عنوان روش بهینه‌سازی برای حداقل کردن خطای آنتروپی متقاطع استفاده می‌شود. نرخ یادگیری برابر با ۰/۰۰۱ است.

۳.۴ متعادل کردن زوج های مثبت و منفی

هر شخص در مجموعه داده دارای چهار جفت تصویر مثبت و تعداد زیادی جفت تصویر منفی است. به دلیل کم بودن جفت‌های مثبت نسبت به جفت‌های منفی ممکن است مشکل بیش‌برازش[۶] رخ دهد. برای جلوگیری از این مشکل و متعادل کردن جفت‌های مثبت و جفت‌های منفی از روش‌های افزایش داده استفاده

[6] Overfitting



شکل ۴: نمودار CMC مربوط به روش‌های مختلف روی مجموعه داده CUHK01

می‌کنیم. تصاویر هر شخص را با استفاده از تکنیک های افزایش داده مانند آیینه، نزدیک‌نمایی و جابه‌جایی تصویر افزایش دادیم. سپس تعداد جفت‌های مثبت افزایش داشته و تعادل آن نسبت به جفت‌های منفی برقرار شد.

در این بخش کارآیی مدل پیشنهادی با دیگر روش‌های توسعه یافته در سال‌های گذشته مانند LMNN، KISSME، ImprovedDeep، Quadruplet و غیره مقایسه می‌شود. از منحنی تطبیق CMC برای ارزیابی کمی روش‌ها استفاده می‌کنیم. در جدول شماره ۱ مقایسه روش ارایه شده با دیگر روش‌ها قرار گرفته است. در شکل ۴ نمودار CMC رسم شده است و مشاهده می‌شود روش پیشنهادی نسبت به روش‌های دیگر دارای بهبود است.

۵ نتیجه‌گیری

بازشناسایی شخص به عنوان یک زیرمسئله از مسئله بازیابی تصاویر درنظر گرفته می‌شود که می‌توان آن را با استفاده از روش‌های سنتی تجزیه و تحلیل تصاویر حل کرد. در این مقاله ما از یک روش انتها به انتها یادگیری عمیق برای حل مسئله بازشناسایی شخص استفاده کردیم. از لایه‌های کانوولوشنال و معماری سیامی در کنار شبکه پیش‌آموزش داده EfficientNet برای استخراج ویژگی‌ها استفاده کردیم. برخلاف روش‌های دیگر بازشناسایی شخص، روش ارائه شده توانست ویژگی‌های سطح بالا و متمایز کننده را به خوبی از تصاویر استخراج کند. به‌دلیل تعداد کم پارامترهای EfficientNet، سرعت یادگیری افزایش یافته و نیاز به تعداد کمتری از مجموعه داده است و در نتیجه کارآیی و دقت شبکه افزایش قابل توجهی داشته است. معیار

جدول ۱: مقایسه روش‌های بازشناسایی شخص

Paper	R-1	R-5	R-10	R-15	R-20	Year	Source Title
Pedestrian recognition with a learned metric (LMNN)	13.5	31.3	42.3	-	54.1	2010	Asian conference on Computer vision
Large scale metric learning from equivalence constraints (KISSME)	29.4	57.7	72.4	-	86.1	2012	CVPR
Person re-identification by local maximal occurrence representation and metric learning (LOMO+XQDA)	63.2	83.9	90.1	-	94.2	2015	CVPR
DeepReID: Deep filter pairing neural network for person re-identification (DeepReID)	27.9	58.2	73.5	-	86.3	2014	CVPR
An improved deep learning architecture for person reidentification (Improved-Deep)	47.5	72.3	80.1	-	83.9	2015	CVPR
Sample-specific SVM learning for person re-identification (LSSCDK)	66	-	90	93.3	95	2016	CVPR
Beyond triplet loss: a deep quadruplet network for person re-identification (Quadruplet)	62.6	-	86	88.9	89.8	2017	CVPR
Deepreid: Deep filter pairing neural network for person re-identification	27.87	-	-	-	-	2014	CVPR
Person re-identification using CNN features learned from combination of attributes	46.8	71.8	80.5	-	-	2016	ICPR
Learning deep feature representations with domain guided dropout for person re-identification	71.7	88.6	92.6	-	-	2016	CVPR
Person re-identification by multi-channel parts-based CNN with improved triplet loss function	53.7	84.3	91	-	-	2016	CVPR
Joint learning of single-image and cross-image representations for person re-identification	71.8	-	-	-	-	2016	CVPR
Deep ranking for person re-identification via joint representation learning	70.94	92.3	96.9	-	-	2016	IEEE Transactions on Image Processing
An improved deep learning architecture for person re-identification	65	89.5	93	-	-	2015	CVPR
Personnet: Person re-identification with deep convolutional neural networks	71.14	90	95	-	-	2016	arxiv
Embedding Deep Metric for Person Re-identification: A Study Against Large Variations	69.38	-	-	-	-	2016	European conference on computer vision
Beyond triplet loss: a deep quadruplet network for person re-identification	62.55	83.44	89.71	-	-	2017	CVPR
A new patch selection method based on parsing and saliency detection for person re-identification	83.2	-	97.1	98.4	98.8	2020	Neurocomputing
A Discriminatively Learned CNN Embedding for Person Re-identification	41	72	87	91	93	2017	ACM Transactions (TOMM)
Proposed	**70**	**95**	**99**	**99**	**99**		

شباهت این شبکه زیر نظر اتلاف تصدیق یادگیری را انجام می‌دهد. این شبکه روی مجموعه داده CUHK01 که یکی از چالشی‌ترین مجموعه داده‌های بازشناسایی شخص می‌باشد آزمایش شده است. تعداد تصاویر این مجموعه داده کم و تصاویر آن دارای وضوح و چالش‌های فراوانی هستند. آزمایش‌های ما روی این مجموعه داده کارآیی روش ارایه شده را نشان می‌دهد.

مراجع

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person re-identification: Past, present and future,” arXiv Prepr. arXiv1610.02984, 2016.

[2] R. Zhao, W. Ouyang, and X. Wang, “Unsupervised salience learning for person re-identification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586–3593.

[3] Z. Wang et al., “Zero-shot person re-identification via cross-view consistency,” IEEE Trans. Multimed., vol. 18, no. 2, pp. 260–272, 2015.

[4] W. Li, R. Zhao, T. Xiao, and X. Wang, “Deepreid: Deep filter pairing neural network for person re-identification,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 152–159.

[5] C. Madden, E. D. Cheng, and M. Piccardi, “Tracking people across disjoint camera views by an illumination-tolerant appearance representation,” Mach. Vis. Appl., vol. 18, no. 3–4, pp. 233–247, 2007.

[6] L. Bazzani, M. Cristani, and V. Murino, “Symmetry-driven accumulation of local features for human characterization and re-identification,” Comput. Vis. Image Underst., vol. 117, no. 2, pp. 130–144, 2013.

[7] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in European conference on computer vision, 2014, pp. 818–833.

[8] M. Tan and Q. V Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” arXiv Prepr. arXiv1905.11946, 2019.

[9] W. Li and X. Wang, “Locally aligned feature transforms across views,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3594–3601.

[10] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in 2012 IEEE conference on computer vision and pattern recognition, 2012, pp. 2288–2295.

[11] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, “Learning locally-adaptive decision functions for person verification,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3610–3617.

[12] N. Martinel, C. Micheloni, and G. L. Foresti, “Saliency weighted features for person re-identification,” in European Conference on Computer Vision, 2014, pp. 191–208.

[13] E. Ahmed, M. Jones, and T. K. Marks, “An improved deep learning architecture for person re-identification,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3908–3916.

[14] S. Paisitkriangkrai, C. Shen, and A. Van Den Hengel, “Learning to rank in person re-identification with metric ensembles,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1846–1855.

[15] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Deep metric learning for person re-identification,” in 2014 22nd International Conference on Pattern Recognition, 2014, pp. 34–39.

[16] H. Cai, L. Zhu, and S. Han, “Proxylessnas: Direct neural architecture search on target task and hardware,” arXiv Prepr. arXiv1812.00332, 2018.

[17] A. G. Howard et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv Prepr. arXiv1704.04861, 2017.

[18] H. T. Kung, B. McDanel, and S. Q. Zhang, “Adaptive tiling: Applying fixed-size systolic arrays to sparse convolutional neural networks,” in 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 1006–1011.

[19] B. Lavi, I. Ullah, M. Fatan, and A. Rocha, “Survey on Reliable Deep Learning-Based Person Re-Identification Models: Are We There Yet?,” arXiv Prepr. arXiv2005.00355, 2020.