مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال دهم، شماره اول، بهار ۱۴۰۲

وفق‌دهی دامنه‌ی بدون نظارت در مسئله‌ی بازشناسایی شخص از طریق یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف

صبا سادات فقیه‌ایمانی[۱]، کاظم فولادی قلعه[۲] و حسین آقابابا[۳]

چکیده

مسئله‌ی بازشناسایی شخص شامل بازیابی تصاویر یک فرد در میان تصاویر جمع‌آوری شده توسط مجموعه‌ای از دوربین‌های غیرهم‌پوشان می‌باشد. باوجود عملکرد موفق‌آمیز مدل‌های عمیق بازشناسایی شخص، هنگام آزمایش مدل روی مجموعه‌داده‌ی بدون برچسب متفاوت با مجموعه‌داده‌ی آموزشی برچسب‌گذاری شده، عملکرد مدل به شدت کاهش می‌یابد. برای حل این مشکل می‌توان از وفق‌دهی دامنه‌ی بدون نظارت استفاده کرد.در این پژوهش مدلی با تعمیم‌پذیری بالا برای وفق‌دهی دامنه‌ی بدون نظارت در مسئله‌ی بازشناسایی شخص ارائه شده است. در این مدل از مجموعه‌داده‌یبرچسب‌گذاری‌شده‌یدامنه‌ی منبع و مجموعه‌داده‌ی بدون برچسب دامنه‌ی هدف برای آموزش مدل استفاده می‌شود و مدل باید در هنگام آزمایش روی دامنه‌ی هدف عملکرد مناسبی داشته باشد. برای این هدف، مدل پیشنهادی توسط سه تابع اتلاف بهینه‌سازی می‌شود. مجموع تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع، تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف و یک تابع اتلاف سه‌گانه به‌منظور یادگیری توأم ویژگی‌های دامنه‌ی منبع و دامنه‌ی هدف، تابع اتلاف نهایی شبکه را تشکیل می‌دهد. مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها در تنظیمات Duke→Market در رتبه‌ی ۱ معیارCMC مقدار ۸۴,۵ درصد و مقدارmAP۶۳ درصد و در تنظیمات Market→Duke در رتبه‌ی ۱ معیارCMC مقدار ۷۰,۱ درصد و مقدار mAP۴۹,۱ درصد را به‌دست آورده است.

کلیدواژه‌ها

بازشناسایی شخص، بازیابی شخص، وفق‌دهی دامنه، یادگیری عمیق

۱ مقدمه

مسئله‌ی بازشناسایی شخص (Person Re-identification)یکی از مسائل پیچیده و پرکاربرد در حوزه‌ی بینایی ماشین است. این مسئله شامل بازیابی تصاویر فرد مورد جستجو از میان مجموعه‌ی تصاویر ثبت شده توسط دوربین‌های غیرهم‌پوشان می‌باشد.

باتوجه به اهمیت حفظ امنیت مکان‌های عمومی و افزایش کاربردهای دوربین‌های امنیتی نظارتی، حجم زیادی از داده‌های تصویری تولید می‌شوند که معمولاً دارای کیفیت و وضوح پایینی می‌باشند. پردازش و تحلیل و بررسی این داده‌ها در هنگام نیاز توسط نیروی انسانی کار بسیار پیچیده و تقریباً غیرممکنی است. به همین دلیل وجود سیستم‌های اتوماتیک و هوشمند برای مسئله‌ی بازشناسایی شخص ضروری می‌باشد.

به دلیل شرایط خاص ثبت تصاویر مجموعه‌داده‌های حوزه‌ی بازشناسایی شخص توسط دوربین‌های امنیتی نظارتی، اغلب این تصاویر چالش‌هایی از قبیل تنوع میزان روشنایی، کیفیت پایین

این مقاله در بهمن‌ماه ۱۴۰۰ دریافت، در فروردین‌ماه ۱۴۰۱ بازنگری و در اردیبهشت‌ماه ۱۴۰۱ پذیرفته شد.

[۱] دانش‌آموخته‌ی کارشناسی ارشد مهندسی فناوری اطلاعات، آزمایشگاه پژوهشی یادگیری عمیق، گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشکدگان فارابی، دانشگاه تهران
رایانامه: saba.sadat.imani@ut.ac.ir

[۲و۳] گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشکدگان فارابی، دانشگاه تهران
رایانامه: {kfouladi,Aghababa}@ut.ac.ir

نویسنده مسئول: کاظم فولادی قلعه

dorl.net/dor/20.1001.1.23831197.1402.10.1.1.9

تصاویر، تنوع زاویه‌دید دوربین‌ها و ... دارند. از طرفی در تصاویر ثبت شده توسط دوربین‌های امنیتی نظارتی، به دلیل کیفیت پایین، چهره‌ی افراد با وضوح قابل مشاهده نیست و همچنین ممکن است افراد چهره‌ی خود را از دوربین‌ها مخفی نگه دارند. بنابراین نمی‌توان از ویژگی‌های چهره‌ی افراد برای بازشناسایی آن‌ها استفاده کرد و معمولاً از ویژگی‌های ظاهری افراد مثل لباس، وضعیت بدن و ... می‌توان بهره برد.

تا کنون مطالعات زیادی در حوزه‌ی تشخیص چهره انجام شده است، اما در کاربردهای عملی، دوربین‌ها همیشه قادر به تهیه‌ی تصویر واضح از صورت افراد نیستند. در نتیجه وجود سیستم‌هایی که قادر به بازشناسایی شخص با استفاده از ویژگی‌های تمام بدن فرد باشند ضروری است. از طرفی دوربین‌ها اغلب محدوده‌های غیرهم‌پوشانی را نظارت می‌کنند. بنابراین بازشناسایی شخص با استفاده از ویژگی‌های کلی بدن افراد و ردیابی فرد در میان دوربین‌های مختلف می‌تواند در کنار تکنولوژی شناسایی چهره در سناریوهای دنیای واقعی بسیار کمک کننده باشد.

باوجود پیچیده بودن مسئله‌ی بازشناسایی شخص، با استفاده از تکنیک‌های یادگیری عمیق نتایج قابل قبولی در این حوزه به‌دست آمده است. اما همچنان چالش‌های مختلفی در مسئله‌ی بازشناسایی شخص وجود دارد. یکی از این چالش‌ها، کمبود داده‌های آموزشی می‌باشد. مدل‌های عمیق اغلب مدل‌های بسیار پیچیده با لایه‌های متعدد هستند. یکی از ملزومات استفاده از مدل‌های عمیق و کارآمدی این مدل‌ها، در دسترس داشتن میزان کافی داده‌های آموزشی است تا مدل دچار بیش‌برازش نشود.

چالش دیگری که در مسئله‌ی بازشناسایی شخص وجود دارد این است که اگر یک مدل روی یک مجموعه‌داده‌ی حوزه‌ی بازشناسایی شخص آموزش داده شود و روی مجموعه‌داده‌ی دیگری از این حوزه آزمایش شود، عملکرد مدل به شدت کاهش می‌یابد. دلیل این افت عملکرد، تفاوت بین دامنه‌های آموزشی و آزمایشی می‌باشد. برای حل این مشکل از تکنیک وفق‌دهی دامنه‌ی بدون نظارت[۱] استفاده می‌شود.

در این پژوهش مدلی با تعمیم‌پذیری بالا برای مسئله‌ی وفق‌دهی دامنه‌ی بدون نظارت در بازشناسایی شخص ارائه شده است. مدل پیشنهادی برای آموزش از یک مجموعه‌داده‌ی برچسب‌گذاری شده‌ی دامنه‌ی منبع[۲] و یک مجموعه‌داده‌ی بدون برچسب دامنه‌ی هدف[۳] استفاده می‌کند. تابع اتلاف نهایی شبکه از مجموع سه تابع اتلاف تشکیل می‌شود. اولین تابع اتلاف مربوط به یادگیری بانظارت ویژگی‌های دامنه‌ی منبع است. تابع اتلاف دوم مربوط به یادگیری بدون نظارت ویژگی‌ها در دامنه‌ی هدف است. مدل پیشنهادی باید هنگام آزمایش روی دامنه‌ی هدف عملکرد خوبی داشته باشد بنابراین اگر بتواند در هنگام آموزش ویژگی‌هایی از دامنه‌ی هدف را یاد بگیرد، در هنگام آزمایش عملکرد بهتری روی این دامنه خواهد داشت. یک روش داده‌افزایی در دامنه‌ی هدف اعمال می‌شود به نحویکه مدل نسبت به تفاوت سبک دوربین‌ها در این دامنه مقاوم‌تر شود. علاوه بر این برای هر تصویر از دامنه‌ی هدف تعدادی همسایه انتخاب می‌شود و در فضای ویژگی هر تصویر از دامنه‌ی هدف به خودش و همسایه‌هایش نزدیک می‌شود. تابع اتلاف سوم یک تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌ی منبع و دامنه‌ی هدف می‌باشد. مدل پیشنهادی علاوه بر عملکرد بسیار خوبی که هنگام آزمایش روی دامنه‌ی هدف دارد، میزان مصرف حافظه در آن کاملاً قابل قبول است.

در بخش (۲) به دو مورد از چالش‌های مهم مسئله‌ی بازشناسایی شخص پرداخته شده است. در بخش (۳) مروری بر کارهای پیشین در این حوزه صورت گرفته است. در بخش (۴) مدل پیشنهادی و جزئیات آن توضیح داده شده و در بخش (۵) تنظیمات مدل و تحلیل نتایج آزمایش‌های مختلف انجام شده روی مدل آورده شده است. همچنین در بخش (۵) عملکرد مدل پیشنهادی با تعدادی از مدل‌های موفق حوزه‌ی وفق‌دهی دامنه‌ی بدون نظارت در بازشناسایی شخص مقایسه شده است. بخش (۶) مربوط به جمع‌بندی و نتیجه‌گیری می‌باشد.

۲ دو چالش مهم در مسئله‌ی بازشناسایی شخص

۲-۱ کمبود تعداد نمونه‌های آموزشی برای هر هویت

تصاویر مجموعه‌داده‌های بازشناسایی شخص چالش‌های متفاوتی دارند [۱]. مدل‌های عمیقی که برای مسئله‌ی بازشناسایی شخص ارائه می‌شوند، مدل‌هایی پیچیده با لایه‌های متعدد می‌باشند. برای اینکه مدل‌ها دچار بیش‌برازش نشوند تعداد زیادی تصویر برچسب‌گذاری شده مورد نیاز است.

اکثر مجموعه‌داده‌های بازشناسایی شخص تعداد تصاویر کمی از هر هویت دارند، برای مثال مجموعه‌داده‌ی VIPeR [۳] به ازای هر هویت، ۲ تصویر دارد. در مجموعه‌داده‌های بزرگ مقیاس این حوزه مانند CUHK03 [۴]، Market1501 [۵] و DukeMTMC-reID [۶] به طور متوسط به ازای هر هویت تعداد ۹/۶، ۱۷/۲ و ۲۳/۵ تصویر وجود دارد. این محدودیت تعداد تصاویر می‌تواند منجر به بیش‌برازش مدل عمیق بازشناسایی شخص شود. برای مقابله با مشکل کمبود تعداد نمونه‌های آموزشی در مسئله‌ی بازشناسایی شخص تاکنون راه‌حل‌های متفاوتی ارائه شده است. یکی از روش‌های مقابله با کمبود تعداد نمونه‌های آموزشی در مسئله‌ی بازشناسایی شخص استفاده از شبکه‌های عصبی سیامی[۴] [۷] است. استفاده از روش‌های مختلف داده‌افزایی

---

[۱] Unsupervised domain adaptation
[۲] Source domain
[۳] Target domain
[۴] Siamese neural network

مثل چرخاندن تصویر، جابه‌جایی تصویر، داده‌افزایی پاک کردن تصادفی[۱] [۸] و ... نیز می‌تواند موثر باشد. همچنین استفاده از شبکه‌های مولد تخاصمی[۲] به منظور تولید داده‌های جدید در مسائل مختلف، یکی از موضوعات جدید و پرطرفدار است.

۲-۲ وفق‌دهی دامنه در بازشناسایی شخص

با وجود عملکرد موفق مدل‌های عمیق بازشناسایی شخص تک‌ـ دامنه‌ای، هنگام آزمایش مدل روی مجموعه‌داده‌ی بدون برچسب دیگری عملکرد مدل به شدت افت پیدا می‌کند. از طرفی فرآیند برچسب‌گذاری تصاویر برای مجموعه‌داده‌های بزرگ مقیاس، بسیار پرهزینه و زمان‌بر است و در سناریوهای دنیای واقعی امکان برچسب‌گذاری همه‌ی مجموعه‌داده‌ها وجود ندارد. بنابراین مدل‌های بانظارت تک دامنه‌ای بازشناسایی شخص برای سناریوهای دنیای واقعی دارای محدودیت هستند.

برای حل این مشکل می‌توان از تکنیک وفق‌دهی دامنه‌ی بدون نظارت بهره برد. وفق‌دهی دامنه بدون نظارت این امکان را می‌دهد که اطلاعات یاد گرفته شده از یک مجموعه‌داده‌ی دارای برچسب را بتوان برای پیش‌بینی روی یک مجموعه‌داده‌ی بدون برچسب مرتبط، استفاده کرد. در حالت کلی وفق‌دهی دامنه از داده‌های دارای برچسب یک و یا بیشتر دامنه‌ی منبع به‌منظور حل وظیفه‌ی دامنه‌ی هدف مرتبط استفاده می‌کند. هرچقدر سطح شباهت و مرتبط بودن بین دامنه‌های منبع و هدف بیشتر باشد، این عملیات موفقیت‌آمیزتر خواهد بود. درواقع در وفق‌دهی دامنه‌ی بدون نظارت، دانش از دامنه‌ی دارای برچسب منبع به دامنه‌ی بدون برچسب هدف منتقل می‌شود.

وفق‌دهی دامنه می‌تواند به سه شکل بانظارت، نیمه نظارتی و بدون نظارت انجام شود. در حالت بانظارت، تصاویر دامنه‌ی هدف دارای برچسب می‌باشند ولی مقدار این داده‌ها برای آموزش یک مدل کامل، کافی نیست. در حالت نیمه نظارتی هم داده‌های دارای برچسب و هم داده‌های بدون برچسب در دامنه‌ی هدف موجود می‌باشد. در حالت بدون نظارت تصاویر دامنه‌ی هدف بدون برچسب می‌باشند. در وفق‌دهی دامنه‌ی بدون نظارت، شبکه روی داده‌های دارای برچسب دامنه‌ی منبع و داده‌های بدون برچسب دامنه‌ی هدف متفاوت ولی مرتبط به دامنه‌ی منبع، آموزش می‌بیند با این هدف که در هنگام آزمایش روی دامنه‌ی هدف عملکرد قابل قبولی داشته باشد.

۳ مرور ادبیات موضوع

امروزه مسئله‌ی بازشناسایی شخص، یکی از مسائل پرطرفدار در حوزه‌ی بینایی ماشین محسوب می‌شود. باتوجه‌به گسترش مجموعه‌داده‌های این حوزه و همچنین به کارگیری تکنیک‌های یادگیری عمیق، نتایج موفقیت‌آمیزی در حوزه‌ی بازشناسایی شخص به‌دست آمده است. مقاله‌های [۹] تا [۱۱] از جمله مقالات مروری هستند که در حوزه‌ی بازشناسایی شخص ارائه شده‌اند.

پژوهش‌ها در حوزه‌ی بازشناسایی شخص با مسئله‌ی ردیابی چند‌ـ دوربینی[۳] در سال ۱۹۹۷ شروع شد [۱۲]. در آن زمان مسئله‌ی بازشناسایی شخص به‌عنوان یک مسئله‌ی مستقل مطرح نبود. در سال ۲۰۰۵، Wojciech Zajdel و همکارانش برای اولین بار در مقاله‌ی خود [۱۳] از عبارت بازشناسایی شخص استفاده کرده و در مسئله‌ی ردیابی چند‌ـ دوربینی به بازشناسایی شخص پرداختند. در سال ۲۰۰۶، Gheissari در مقاله‌ی خود [۱۴] روی مسئله‌ی بازشناسایی شخص به‌صورت مستقل کار کرد. در سال ۲۰۱۰، دو مقاله‌ی [۱۵] و [۱۶] برای بازشناسایی شخص چند‌ـ شاتی ارائه شدند. در سال‌های ۲۰۱۰ و ۲۰۱۱، مقالات [۱۷] و [۱۸] برای مسئله‌ی بازشناسایی شخص از یادگیری معیار استفاده کردند به‌نحوی‌که با بیشینه کردن احتمال کمتر بودن فاصله‌ی زوج‌های منطبق صحیح نسبت به فاصله‌ی زوج‌های منطبق ناصحیح، معیار فاصله‌ی بهینه را یاد بگیرند. در سال ۲۰۱۲، اولین مقاله [۱۹] در حوزه‌ی بازشناسایی شخص مجموعه‌ـ باز[۴] ارائه شد. در بازشناسایی شخص مجموعه‌ـ باز، هویت تصویر ورودی امکان دارد در میان هویت‌های تصاویر موجود در گالری نباشد.

در سال ۲۰۱۴ مقاله‌های [۲۰] و [۲۱] برای اولین بار از یادگیری عمیق برای حل مسئله‌ی بازشناسایی شخص استفاده کردند. این دو مقاله از شبکه‌ی عصبی سیامی برای تعیین احتمال تعلق دو زوج تصویر ورودی به یک هویت واحد، بهره بردند. یکی از دلایل استفاده از شبکه‌ی سیامی برای این هدف، این است که تعداد نمونه‌های آموزشی به ازای هر هویت محدود می‌باشد. با گسترش یادگیری عمیق و استفاده از تکنیک‌های یادگیری عمیق مدل‌های بانظارت و تک‌ـ دامنه‌ای بازشناسایی شخص موفقی ارائه شده‌اند [۲۲] تا [۲۵]. اما این مسئله هنوز چالش‌هایی دارد که در بخش (۲) به دومورد از این چالش‌ها پرداخته شده است.

یکی از راه‌های مقابله با چالش کمبود تعداد نمونه‌های آموزشی در مجموعه‌داده‌های حوزه‌ی بازشناسایی شخص استفاده از استفاده از معماری شبکه عصبی سیامی است. معمولاً هنگامیکه از هر کلاس نمونه‌های محدودی وجود داشته باشد، استفاده از مدل‌های سیامی موفق‌تر از مدل‌های طبقه‌بندی خواهد بود. مدل‌های زیادی برای مسئله‌ی بازشناسایی شخص ارائه شده‌اند که از معماری شبکه سیامی استفاده می‌کنند [۲۶] تا [۲۸].

همچنین داده‌افزایی با استفاده از شبکه‌های مولد تخاصمی یک حوزه‌ی بسیار به‌روز و فعال در بازشناسایی شخص است. در [۶] با استفاده از DCGAN [۲۹] و در [۳۰] با استفاده از WGAN-GP [۳۱] به تولید تصاویر آموزشی جدید برای مسئله‌ی

---

[۱] Random erasing data augmentation

[۲] Generative adversarial networks

[۳] Multi-camera tracking

[۴] Open-set

بازشناسایی شخص پرداخته شده است. در [۲۴] با جابجا کردن کدهای ساختاری و کدهای ظاهری هر دو تصویر آموزشی، تعدادی تصاویر آموزشی جدید تولید می‌شود. در برخی مقالات با تمرکز بر یک چالش خاص، تصاویر آموزشی جدید تولید می‌کنند. مثلاً مقالات [۳۲] و [۳۳] با استفاده از GAN تصاویر آموزشی جدیدی از افراد در حالات قرارگیری مختلف تولید کردند در نتیجه مدل آموزش دیده نسبت به تفاوت حالت قرارگیری افراد مقاوم می‌شود. در مقاله‌ی [۳۴] از هر هویت تصاویری در پس‌زمینه‌های مختلفی که در مجموعه‌داده وجود دارد تولید می‌شود.

یکی از چالش‌های مهمی که در مسئله‌ی بازشناسایی شخص وجود دارد این است که اگر یک مدل روی یک مجموعه‌داده‌ی حوزه‌ی بازشناسایی شخص آموزش داده شود و روی مجموعه‌داده‌ی دیگری از این حوزه آزمایش شود، عملکرد مدل به شدت کاهش می‌یابد. این تفاوت عملکرد به‌دلیل فاصله‌ی دامنه‌های مجموعه‌های داده ناشی از تفاوت رزولوشن تصاویر، تفاوت میزان روشنایی، تفاوت سرعت حرکت افراد، تفاوت پس‌زمینه‌ها و ... می‌باشد. این چالش می‌تواند برای مسئله‌ی بازشناسایی شخص جدی باشد زیرا نمونه‌های آموزشی در دسترس نمی‌توانند برای دامنه‌های آزمایشی جدید کافی و مؤثر باشند. بنابراین توجه به موضوع وفق‌دهی دامنه در مسئله‌ی بازشناسایی شخص گسترش یافته است.

به‌دلیل هزینه‌بر بودن فرآیند برچسب‌گذاری تصاویر و همچنین کاهش قابل توجه کارآیی مدل‌های بازشناسایی شخص هنگام متفاوت بودن مجموعه داده‌های آموزشی و آزمایشی، مقالات زیادی سعی کرده‌اند که روی موضوع وفق‌دهی دامنه در مسئله‌ی بازشناسایی شخص کار کنند [۳۵] تا [۴۴] . به نظر می‌آید که روش‌های بانظارت تک‌ـ دامنه‌ای در مسئله‌ی بازشناسایی شخص، برای سناریوهای واقعی دارای محدودیت می‌باشند. همچنین روش‌های بدون نظارت بازشناسایی شخص نیز عملکرد قابل قبولی ندارند.

در مقاله‌ی [۳۵] برای مقابله با چالش افت عملکرد مدل آموزش دیده در هنگام آزمایش روی مجموعه‌داده‌ی هدف و کاهش فاصله‌ی دامنه‌های منبع و هدف، شبکه‌ی مولد تخاصمی انتقال شخص PTGAN معرفی شده است.

در [۳۶] با استفاده از شبکه‌ی مولد تخاصمی تصاویر دارای برچسب دامنه‌ی منبع به دامنه‌ی هدف منتقل می‌شوند به‌نحوی‌که تصاویر منتقل شده سبک مشابه با تصاویر دامنه‌ی هدف دارند. در گام بعدی تصاویر تغییر سبک داده شده به همراه برچسب‌های متناظر خود برای یادگیری بانظارت در دامنه‌ی هدف مورداستفاده قرار می‌گیرند. در هر تصویر، هویت آن باید پس از ترجمه ثابت بماند همچنین دو دامنه هویت‌های کاملاً متفاوتی دارند بنابراین هویت تصویر ترجمه شده باید با تمامی هویت‌های تصاویر دامنه‌ی هدف متفاوت باشد. برای رسیدن به این هدف دو نوع شباهت بدون نظارت تعریف می‌شود. مورد اول شباهت یک تصویر با خودش قبل و بعد از فرآیند ترجمه است و مورد دوم عدم شباهت تصویر ترجمه شده‌ی دامنه‌ی منبع و تصاویر دامنه‌ی هدف می‌باشد. هر دو قید توسط شبکه‌ای با عنوان SPGAN پیاده‌سازی می‌شوند.

در مقاله‌ی [۳۷] برای انجام وظیفه‌ی بازشناسایی شخص کنار ـ دامنه‌ای به‌طور مؤثر، یک شبکه‌ی انتقال وفقی ATNet معرفی شده است. شبکه‌ی ATNet به دلایل اصلی ایجاد فاصله در دامنه‌های هدف و منبع نگاه می‌کند و از طریق اصل «تقسیم و غلبه» به این مسئله می‌پردازد. این شبکه انتقال کنار ـ دامنه‌ای را به مجموعه‌ای از زیرانتقال‌ها که هرکدام از آن‌ها به تغییر سبک با یک فاکتور خاص مثل روشنایی، رزولوشن، زاویه دید دوربین و ... تمرکز دارند، تجزیه می‌کند. سپس یک استراتژی ترکیبی وفق‌دهی ارائه می‌کند که در آن زیرانتقال‌ها باتوجه‌به اهمیت تأثیر فاکتورهای مختلف روی تصاویر، ادغام می‌شوند.

مقاله‌ی [۳۸] به منظور حذف کردن تأثیر سبک و حالت قرارگیری افراد در بازشناسایی شخص کنار ـ دامنه‌ای، یک الگوریتم یادگیری دیکشنری[۱] بر مبنای تجزیه‌ی ماتریسی[۲] ارائه کرده است.

در اکثر مقالات وفق‌دهی دامنه برای بازشناسایی شخص سعی شده است که فاصله‌ی بین دامنه‌ی منبع و دامنه‌ی هدف را از لحاظ سطح تصویر [۳۶و۳۵] و یا سطح ویژگی [۴۳] کاهش دهند. باوجود مؤثر بودن این روش‌ها اغلب تفاوت‌های درون ـ دامنه‌ای در دامنه‌ی هدف درنظر گرفته نمی‌شود. اما در [۳۹] به‌طور صریح تفاوت درون ـ دامنه‌ای تنوع تصاویر ناشی از دوربین‌های مختلف در دامنه‌ی هدف درنظر گرفته می‌شود. در [۴۰] , [۴۱], [۴۲] علاوه‌بر یادگیری تغییرناپذیری نسبت به دوربین‌ها، تغییرناپذیری نسبت به نمونه‌ها و تغییرپذیری نسبت به همسایه‌ها نیز در دامنه‌ی هدف بررسی می‌شوند. تفاوت عمده‌ی این سه مقاله در روش یادگیری تغییرناپذیری نسبت به همسایه‌ها می‌باشد.

در مقاله‌ی [۴۴] مدلی برای وفق‌دهی دامنه در بازشناسایی شخص ارائه شده است که با استفاده از شبکه‌ی کانوولوشنال گرافی به موضوع همبستگی ویژگی‌های معنایی درون ـ دامنه‌ای و میان ـ دامنه‌ای پرداخته است.

۴ مدل پیشنهادی

به دلیل کاهش قابل توجه عملکرد مدل آموزش‌دیده ی بازشناسایی شخص به هنگام آزمایش روی مجموعه‌داده‌ی متفاوت و همچنین هزینه بر بودن فرآیند برچسب گذاری تصاویر آموزشی، ارائه‌ی مدلی که تعمیم‌پذیری بالایی هنگام آزمایش روی مجموعه‌داده‌ی بدون برچسب داشته باشد بسیار با ارزش است [۲]. در این پژوهش مدلی برای وفق‌دهی دامنه‌ی بدون نظارت برای بازشناسایی شخص ارائه شده‌است. این مدل از یک مجموعه‌داده‌ی برچسب‌گذاری شده

[۱] Dictionary learning

[۲] Matrix factorization

به عنوان مجموعه‌داده‌ی منبع $\{x_{s,i}, y_{s,i}\}_{i=1}^{N_s}$ و از یک مجموعه‌داده‌ی بدون برچسب به عنوان مجموعه‌داده‌ی هدف $\{x_{t,i}\}_{i=1}^{N_t}$ استفاده می‌کند. مدل پیشنهادی به هنگام آزمایش روی دامنه‌ی هدف عملکرد بسیار خوبی دارد. برای تحقق این هدف تابع اتلاف نهایی شبکه از مجموع سه تابع اتلاف مختلف تشکیل می‌شود:۱)تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع ۲) تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف ۳) تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف. ساختار کلی مدل پیشنهادی در شکل (۱) نمایش داده شده است.

۴-۱ یادگیری بانظارت ویژگی‌های دامنه‌ی منبع

مجموعه‌داده‌ی منبع یک مجموعه‌داده‌ی برچسب گذاری شده است. تصاویر $x_{s,i}$ برچسب‌های $y_{s,i}$ را دارند. به منظور یادگیری ویژگی‌های دامنه‌ی منبع، به مسئله‌ی بازشناسایی شخص به شکل یک مسئله‌ی طبقه‌بندی تصاویر نگاه می‌شود. براساس ویژگی‌هایی که از تصویر استخراج می‌شود باید کلاس مربوط به تصویر توسط شبکه پیش‌بینی شود. احتمال پیش‌بینی برچسب $y_{s,i}$ برای تصویر $x_{s,i}$توسط تابع softmax کدگذاری می‌شود. سپس تابع اتلاف شناسایی از طریق آنتروپی متقاطع[1] تعریف می‌شود و فرآیند آموزش برای دامنه‌ی منبع با کمینه کردن تابع اتلاف زیر به دست می‌آید.

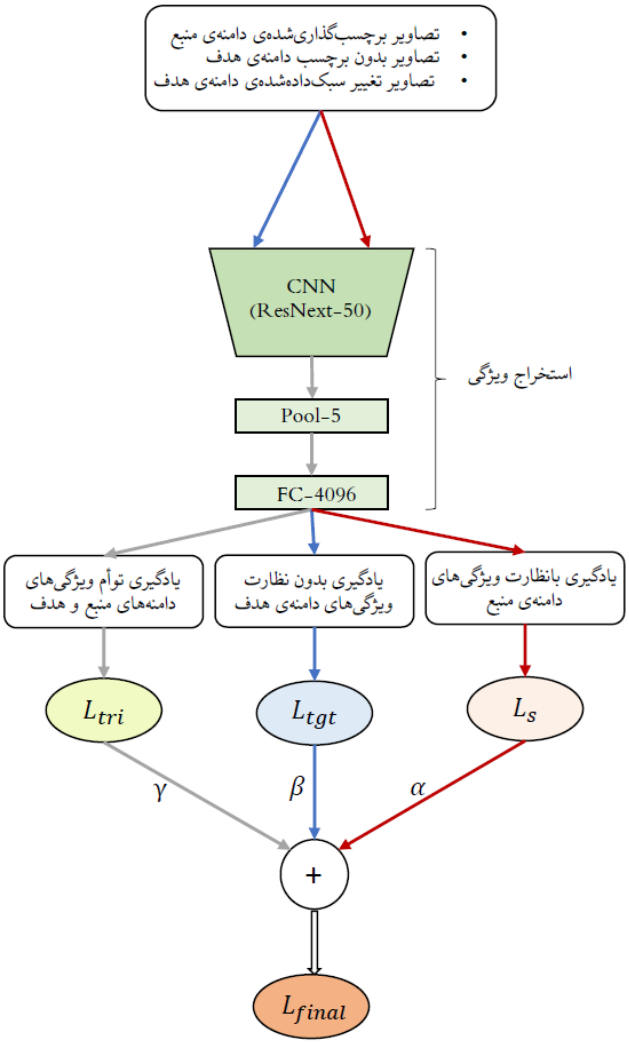
$$L_s = -\frac{1}{N_s}\sum_{i=1}^{N_s} \log p(y_{s,i}|x_{s,i}) \quad (۱)$$

۴-۲ یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف

۴-۲-۱- کمبود تعداد نمونه‌ها و تفاوت سبک دوربین‌ها در دامنه‌ی هدف

کمبود تعداد نمونه‌های آموزشی به ازای هر هویت در مجموعه‌داده‌های بازشناسایی شخص یکی از چالش‌های مهم آموزش مدل‌های عمیق بازشناسایی شخص می‌باشد. مدل‌های عمیق بازشناسایی شخص، معمولاً مدل‌هایی با لایه‌های متعدد و پیچیدگی زیادی هستند. برای آموزش چنین مدل‌هایی حجم زیادی از داده‌های آموزشی نیاز است تا مدل آموزش دیده بتواند قدرت تمییزدهندگی مناسبی داشته باشد. کمبود تعداد نمونه‌های آموزشی باعث بیش‌برازش شده و مانع عملکرد مناسب مدل می‌شود.

ازطرفی در مسئله‌ی بازشناسایی شخص تصاویر افراد توسط دوربین‌های مختلف ثبت می‌شود. بنابراین این تصاویر تفاوت‌هایی در کیفیت و رزولوشن ، میزان روشنایی ، پس‌زمینه و ... دارند.



شکل (۱): ساختار کلی مدل پیشنهادی. در فرآیند آموزش، داده‌های برچسب‌گذاری‌شده‌ی دامنه‌ی منبع و داده‌های بدون برچسب دامنه‌ی هدف به منظور استخراج ویژگی وارد شبکه‌ی عمیق می‌شوند. سپس شبکه با سه تابع اتلاف بهینه‌سازی می‌شود. L_s : تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع، L_{tgt} : تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف و L_{tri}: تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف. خطوط قرمز بیانگر جریان داده‌ی دامنه‌ی منبع، خطوط آبی بیانگر جریان داده‌ی دامنه‌ی هدف و خطوط خاکستری بیانگر جریان داده‌ی دامنه‌های منبع و هدف هستند.

این تفاوت‌ها مربوط به تفاوت سبک دوربین‌ها است. تفاوت سبک دوربین‌ها در تصاویر مجموعه‌داده‌های بازشناسایی شخص یکی از چالش‌ها است.

در مدل پیشنهادی برای مقابله با کمبود تعداد نمونه‌های آموزشی و همچنین یادگیری تفاوت سبک دوربین‌ها در دامنه‌ی هدف، از روش داده‌افزایی CamStyle [۴۵] استفاده شده است.

با آموزش یک مدل CamStyle با استفاده از CycleGAN [۴۶] روی دامنه‌ی هدف، تصاویر جدیدی با سبک دوربین‌های

[1] Cross entropy

مختلف در دامنه‌ی هدف تولید می‌شـــود. اگر تعداد N_c دوربین در دامنه‌ی هدف موجود باشـــد، به ازای هر تصـــویر تعداد $N_c - 1$ تصـــویر جدید به ســـبک ســـایر دوربین‌ها در دامنه‌ی هدف تولید می‌شـــود. این روش داده‌افزایی علاوه بر افزایش تعداد نمونه‌های دامنه‌ی هدف، مدل را نسبت به تفاوت سبک دوربین‌ها در دامنه‌ی هدف مقاوم کرده و باعث می‌شـــود که مدل آموزش دیده هنگام آزمایش روی دامنه‌ی هدف عملکرد بهتری داشته باشد.

به منظور مقاوم شـــدن مدل آموزش دیده نســـبت به تفاوت ســـبک دوربین‌ها در دامنه‌ی هدف، باید یک تصـــویر و تصـــاویر جدید تولیدشـــده‌ی متناظرش متعلق به یک هویت باشـــند و در فضـــای ویژگی به هم نزدیک شوند.

۲-۲-۴- مقاوم شدن مدل نسبت به تغییرات درون‌دامنه‌ای در دامنه‌ی هدف

تصـاویر در دامنه‌ی هدف برچسـب‌گذاری نشـده‌اند بنابراین هویت تصـاویر مشـخص نیسـت. اما از هر هویت تعدادی تصـویر موجود اسـت. اگر تصـاویری که متعلق به یک هویت هسـتند در فرآیند آموزش قابل استخراج باشند، می‌توان مدل بازشناسایی شخص با قابلیت تعمیم‌پذیری بیشـتری روی دامنه‌ی هدف داشـت. در فرآیند آموزش در دامنه‌ی هدف برای هر تصـویر تعدادی همسـایه تعیین می‌شـود و هر تصـویر و همسـایه‌هایش در فضـای ویژگی به هم نزدیک می‌شـوند و از سـایر نمونه‌ها دور می‌شـوند. از طرفی هر تصـویر ویژگی‌های ظاهری منحصـربفردی دارد. حتی تصـاویری که در اصل متعلق به یک هویت هستند از آنجاییکه توسط دوربین‌های مختلف و در شرایط مختلف ثبت شده‌اند باهم تفاوت‌هایی دارند. برای یادگیری ویژگی‌های تصـاویر در دامنه‌ی هدف، در فضـای ویژگی هر تصویر به خودش و همسایه‌هایش نزدیک شده و از سایر تصاویر دور می‌شود.

برای رسـیدن به این مقصـود ویژگی‌های به روز تمامی نمونه‌های دامنه‌ی هدف در یک سـاختار حافظه‌ای [۴۰] ذخیره می‌شـود. برای انتخاب همسـایه‌های نمونه‌های دامنه‌ی هدف، دو اسـتراتژی مورد آزمایش قرار گرفته است.

استراتژی اول انتخاب همسایه برای نمونه‌های دامنه‌ی هدف

در اســـتراتژی اول برای هر تصـــویر در دامنـه‌ی هـدف، پس از محاسبه‌ی شباهت ویژگی آن تصویر با ویژگی‌های ذخیره شده در حافظه، تعداد k نزدیک‌ترین همســـایه از میان نمونه‌های موجود در حافظه، به عنوان همسایه‌های آن تصویر انتخاب می‌شوند. در نتیجه برای تمامی تصاویر دامنه‌ی هدف به تعداد ثابت k همسایه انتخاب خواهد شد [۴۰]. فرمول $w_{i,j}$ مشابه زیر تعریف می‌شود.

$$w_{i,j} = \begin{cases} 1 & i = j \\ \frac{1}{k} & x_{t,j} \; is \; a \; neighbor \; of \; x_{t,i} \\ 0 & o.w. \end{cases} \quad (۲)$$

با کمینه کردن تابع اتلاف زیر هر تصـــویر از دامنه‌ی هدف در فضـای ویژگی به خودش و همسـایه‌هایش نزدیک شـده و از سـایر نمونه‌ها دور می‌شود.

$$L_{tgt} = -\frac{1}{N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} w_{i,j} \log(j|\tilde{x}_{t,i}) \quad (۳)$$

تصـــویر $\tilde{x}_{t,i}$ از اجتماع مجموعه تصـــویر $x_{t,i}$ و تصـــاویر تغییرسبک داده‌شده‌ی متناظرش انتخاب می‌شود. بدین ترتیب داده افزایی سـبک دوربین‌ها نیز به مدل اعمال می‌شـود. در فرمول (۳) اگر $i = j$ باشـد، $w_{i,j} = 1$ خواهد بود و نمونه‌ی $x_{t,i}$ در فضـای ویژگی به خودش نزدیک می‌شود. اگر i≠j و $x_{t,j}$ یکی از همسایه‌های $x_{t,i}$ باشد نیز $w_{i,j} \neq 0$ می‌باشد و نمونه‌های $x_{t,i}$ و $x_{t,j}$ در فضای ویژگی به هم نزدیک می‌شوند.

استراتژی دوم انتخاب همسایه برای نمونه‌های دامنه‌ی هدف

در اسـتراتژی دوم، برای انتخاب همسـایه‌ها میزان شـباهت ویژگی یک تصـویر دامنه‌ی هدف با ویژگی‌های سـایر نمونه‌های دامنه‌ی هدف محاسبه می‌شود. تصاویری که میزان شباهت ویژگی‌های آن‌ها به ویژگی‌های تصـویر مدنظر از یک مقدار آسـتانه بیشـتر باشـد به عنوان همسـایه‌های آن تصـویر انتخاب می‌شـوند. فرمول $w_{i,j}$ به شکل زیر تعریف می‌شود.

$$w_{i,j} = \begin{cases} 1 & i = j \\ 1 & x_{t,j} \; is \; a \; neighbor \; of \; x_{t,i} \\ 0 & o.w. \end{cases} \quad (٤)$$

با کمینه کردن تابع اتلاف زیر هر تصـــویر از دامنه‌ی هدف در فضـای ویژگی به خودش و همسـایه‌هایش نزدیک شـده و از سـایر نمونه‌ها دور می‌شود.

$$L_{tgt} = -\frac{1}{N_t}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} w_{i,j} \log(j|\tilde{x}_{t,i}) \quad (۵)$$

تصـــویر $\tilde{x}_{t,i}$ از اجتماع مجموعه تصـــویر $x_{t,i}$ و تصـــاویر تغییرسبک داده‌شده‌ی متناظرش انتخاب می‌شود. بدین ترتیب داده افزایی سـبک دوربین‌ها نیز به مدل اعمال می‌شـود. در فرمول (۵) اگر i=j باشد، $w_{i,j} = 1$ خواهد بود و نمونه‌ی $x_{t,i}$ در فضای ویژگی به خودش نزدیک می‌شـود. اگر i≠j و $x_{t,j}$ یکی از همسـایه‌های $x_{t,i}$ باشـد نیز $w_{i,j} = 1$ می‌باشـد و نمونه‌های $x_{t,i}$ و $x_{t,j}$ در فضـای ویژگی به هم نزدیک می‌شوند.

از آنجاییکه در اسـتراتژی دوم تعداد همسـایه‌های انتخاب شـده برای هر تصـویر در دامنه‌ی هدف متفاوت اسـت. مشـکلی که وجود دارد این اسـت که اگر تعداد همسـایه‌های یک تصـویر بسـیار زیاد باشـد، مجموع اتلاف بین تصـویر و همسـایه‌هایش بسـیار زیاد می‌شـود. اگر تعداد همسـایه‌های یک تصـویر بسـیار کم باشـد،

مجموع اتلاف بین تصویر و همسایه‌هایش بسیار کم می‌شود. این عدم توازن می‌تواند مدل را با مشکل مواجه کند. برای حل این مشکل از یک جریمه در تابع اتلاف استفاده می‌شود [۴۱]. اگر تعداد همسایه‌های تصویر $x_{t,i}$ برابر با N_{n_i} باشد عبارت جریمه $\frac{1}{N_{n_i}\log(N_{n_i})}$ خواهد بود. بدین ترتیب فرمول تابع اتلاف دامنه‌ی هدف به شکل زیر تعریف می‌شود.

$$L_{tgt} = -\frac{1}{N_t}\sum_{i=1,N_{n_i}>1}^{N_t}\frac{1}{N_{n_i}\log(N_{n_i})}\sum_{j=1}^{N_t} v_{i,j}\log p\left(j|\tilde{x}_{t,i}\right) \quad (۶)$$

اگر تصویری تعداد همسایه‌های زیادی داشته باشد،با استفاده از جریمه‌ی $\frac{1}{N_{n_i}\log(N_{n_i})}$ اتلاف بین آن تصویر و همسایه‌هایش کاهش می‌یابد تعداد همسایه‌ها متوازن می‌شود.

نمونه‌هایی که غیر از خودشان حداقل یک همسایه‌ی دیگر داشته باشند به عنوان نمونه‌های آموزشی استفاده می‌شوند بنابراین باید $N_{n_i} > 1$ باشد. در فرمول (۶) نیز برای اعمال داده‌افزایی سبک دوربین‌ها نمونه‌ی $\tilde{x}_{t,i}$ از اجتماع مجموعه تصویر $x_{t,i}$ و تصاویر تغییرسبک داده‌شده‌ی متناظرش انتخاب می‌شود. بدین ترتیب تابع اتلاف دامنه‌ی هدف شکل می‌گیرد. این تابع اتلاف علاوه بر مقاوم کردن مدل نسبت به تفاوت سبک دوربین‌ها در دامنه‌ی هدف، باعث می‌شود هر نمونه از دامنه‌ی هدف در فضای ویژگی به خودش و همسایه‌هایش نزدیک شده و از سایر نمونه‌ها دور شده و ویژگی‌های درون‌ـ دامنه‌ای در دامنه‌ی هدف توسط مدل یاد گرفته شود.

مقایسه‌ی استراتژی اول و استراتژی دوم انتخاب همسایه برای نمونه‌های دامنه‌ی هدف

در استراتژی اول انتخاب همسایه‌ها، برای هر نمونه از دامنه‌ی هدف تعداد k نزدیکترین نمونه به آن تصویر، به عنوان همسایه‌هایش انتخاب می‌شوند. در این استراتژی برای تمامی نمونه‌های دامنه‌ی هدف تعداد یکسانی همسایه انتخاب می‌شود. این مسئله می‌تواند نقطه‌ی ضعف این استراتژی باشد چرا که می‌دانیم در مجموعه‌داده‌های بازشناسایی شخص به ازای هر هویت لزوماً تعداد تصاویر یکسانی وجود ندارد. برای حل این مشکل، استراتژی دوم انتخاب همسایه‌ها پیشنهاد می‌شود.

در استراتژی دوم انتخاب همسایه‌ها، یک مقدار آستانه مشخص می‌شود و تصاویری که میزان شباهت آن‌ها با یک نمونه از دامنه‌ی هدف از میزان آستانه بیشتر باشد به عنوان همسایه‌های آن نمونه انتخاب می‌شوند. در این روش تعداد همسایه‌های انتخاب شده برای تصاویر مختلف از دامنه‌ی هدف متفاوت می‌باشد. اما اگر برای نمونه‌های مختلف در دامنه‌ی هدف تعداد تصاویر بسیار متفاوتی انتخاب شود، این عدم توازن تعداد همسایه‌ها باعث می‌شود فرآیند آموزش با مشکل مواجه شود. اگر تعداد همسایه‌های یک تصویر بسیار زیاد باشد، در فرمول (۵) مجموع اتلاف بین تصویر و همسایه‌هایش بسیار زیاد می‌شود و اگر تعداد همسایه‌های یک تصویر بسیار کم باشد، در فرمول (۵) مجموع اتلاف بین تصویر و همسایه‌هایش بسیار کم می‌شود. برای حل این مشکل از یک عبارت جریمه در تابع اتلاف استفاده می‌شود و باعث ایجاد توازن در تعداد همسایه‌ها برای نمونه‌های دامنه‌ی هدف می‌شود.با اعمال عبارت جریمه، تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف به صورت فرمول (۶) تعریف می‌شود.

آزمایش‌های انجام شده روی مدل پیشنهادی نشان می‌دهد که استراتژی دوم انتخاب همسایه برای نمونه‌های دامنه‌ی هدف عملکرد بهتری نسبت به استراتژی اول انتخاب همسایه برای نمونه‌های دامنه‌ی هدف دارد.

۳-۴ یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف

در مدل پیشنهادی به منظور یادگیری توأم ویژگی‌های دامنه‌ی منبع و دامنه‌ی هدف از یک تابع اتلاف سه‌گانه استفاده شده است. تابع اتلاف سه‌گانه اولین بار توسط گوگل برای وظیفه‌ی تشخیص چهره ارائه شد [۴۷]. هدف تابع اتلاف سه‌گانه بیشینه کردن تفاوت های بیرون‌ـ کلاسی و کمینه کردن تفاوت های درون‌ـ کلاسی است. تابع اتلاف سه‌گانه نیاز به سه‌تایی‌هایی دارد که تشکیل شده اند از یک تصویر لنگر $x_{1,i}$، یک نمونه ی مثبت $x_{2,i}$ که متعلق به کلاس تصویر لنگر است و یک نمونه‌ی منفی $x_{3,i}$ که متعلق به کلاسی متفاوت با کلاس تصویر لنگر است. ایده ی اصلی تابع اتلاف سه‌گانه این است که فاصله بین زوج‌های مثبت باید کمتر از فاصله بین زوج‌های منفی باشد.

می‌توان تابع اتلاف سه‌گانه را به طور کلی به صورت زیر تعریف کرد.

$$L_{triplet}(X) = \sum_{i=1}^{N} max\left\{\left\|f(x_{1,i}) - f(x_{2,i})\right\|^2 - \left\|f(x_{1,i}) - f(x_{3,i})\right\|^2 + \rho, 0\right\} \quad (۷)$$

که در آن X مجموعه‌ای از سه‌تایی ها است به طوری‌که $X_i =< x_{1,i}, x_{2,i}, x_{3,i} >,\ 1 \le i \le N$. پارامتر ρ ثابت حاشیه نام دارد.

در مسئله‌ی بازشناسایی شخص، هویت‌های موجود در دو مجموعه‌داده‌ی منبع و هدف کاملاً متفاوت هستند. در واقع هر نمونه از دامنه‌ی منبع $(x_{s,i})$ با هر نمونه از دامنه‌ی هدف $(x_{t,i})$ متعلق به کلاس‌های مختلفی می‌باشند و یک زوج منفی را تشکیل می‌دهند. از طرفی، تصاویر تولیدشده با سبک دوربین‌های مختلف از یک تصویر دامنه‌ی هدف $(x^*_{t,i})$، هویتی مشابه به تصویر اصلی $(x_{t,i})$ دارند. بنابراین یک تصویر از دامنه‌ی هدف با یکی از تصاویر تغییرسبک‌داده‌شده‌ی متناظرش، یک زوج مثبت را می‌سازند. دو نمونه از سه‌تایی‌های مد نظر در شکل (۲) نمایش داده شدند. ایده‌ی تابع اتلاف سه‌گانه‌ی پیشنهادشده از مقاله‌ی [۳۹] برگرفته

شده است. بنابراین با داشتن سه‌تایی‌هایی تشکیل‌شده از یک نمونه از دامنه‌ی منبع، یک نمونه از دامنه‌ی هدف و یک نمونه از تصاویر تولیدشده‌ی متناظر با نمونه‌ی دامنه‌ی هدف، می‌توان تابع اتلاف جدیدی را تعریف نمود.

$$L_{tri}(X) = \sum_{i=1}^{N} max\left\{\left\|f(x_{t,i}) - f(x^*_{t,i})\right\|^2 - \left\|f(x_{t,i}) - f(x_{s,i})\right\|^2 + \rho, 0\right\} \quad (۸)$$

استفاده از تابع اتلاف سه‌گانه‌ی فرمول (۸)، باعث می‌شود که نمونه‌های دامنه‌ی منبع از نمونه‌های دامنه‌ی هدف دور شوند و همچنین نمونه‌های دامنه‌ی هدف به نمونه‌های تولیدشده‌ی متناظر خود نزدیک شوند. بدین ترتیب مدل آموزش دیده نسبت به تغییرات درون‌ـ دامنه‌ای در دامنه‌ی هدف و نسبت به تغییرات بین دامنه‌ای میان دامنه‌های منبع و هدف مقاوم می‌شود.

۴-۴ تابع اتلاف نهایی مدل

تابع اتلاف کلی شبکه از مجموع توابع اتلاف فرمول (۱) و فرمول (۶) و فرمول(۸) تشکیل می‌شود. در تابع اتلاف نهایی شبکه، L_s باعث می‌شود که مدل ویژگی‌های درون‌ـ دامنه‌ای در دامنه‌ی منبع را یاد بگیرد و و نسبت به تغییرات درون‌ـ دامنه‌ای در دامنه‌ی هدف که مشابه با تغییرات درون‌ـ دامنه‌ای در دامنه‌ی منبع می‌باشند، مقاوم‌تر شود. L_{tgt} باعث می‌شود که مدل نسبت به تغییرات سبک دوربین‌ها در دامنه‌ی هدف مقاوم شود و همچنین بازنمایی تصاویر

شکل (۲): یک تصویر از دامنه‌ی هدف ($x_{t,i}$) با تصویر تغییر سبک‌داده‌ی خودش ($x^*_{t,i}$) زوج مثبت و با یک تصویر از دامنه‌ی منبع ($x_{s,i}$) یک زوج منفی را می‌سازد.

بدون برچسب و تغییرات درون‌ـ دامنه‌ای در دامنه‌ی هدف را یادبگیرد. L_{tri} با استفاده از تابع اتلاف سه‌گانه، علاوه بر تغییرات در دامنه‌ی هدف، تفاوت‌های بین دامنه‌ی منبع و دامنه‌ی هدف را نیز یاد می‌گیرد. تابع اتلاف نهایی شبکه به شکل زیر تعریف می‌شود.

$$L_{final} = \alpha L_s + \beta L_{tgt} + \gamma L_{tri} \quad (۹)$$

پارامترهای α و β و γ اعداد ثابتی هستند که میزان تأثیر توابع اتلاف L_s ، L_{tgt} و L_{tri} را در تابع اتلاف نهایی کنترل می‌کنند.

۴-۵ معماری ResNext-50

در مدل پیشنهادی به منظور استخراج ویژگی نمونه‌های آموزشی از شبکه‌ی کانوولوشنال ResNsxt-50 استفاده شده است. ResNext [۴۸] یکی از گسترش‌های ResNet [۴۹] است که در سال ۲۰۱۷ معرفی شد.تفاوت ResNext با ResNet در وجود تعدادی انشعاب یا مسیر موازی داخل بلاک باقی‌مانده[۱] می‌باشد. درواقع در ResNext به‌جای اعمال کانوولوشنال به تمام نقشه‌ی ویژگی ورودی، ورودی یک بلاک به دنباله‌هایی با ابعاد (تعداد کانال) بازنمایی کمتر تبدیل شده و چند فیلتر کانوولوشنال قبل از ادغام نتیجه به‌طور جداگانه روی آن‌ها اعمال می‌شود. همچنین تعداد پارامترها در ResNext کمتر از تعداد پارامترها در ResNet می‌باشد.

۵ آزمایش‌ها

۵-۱ مجموعه‌های داده

مدل پیشنهادی روی دو مجموعه‌داده‌ی بزرگ‌ـ مقیاس DukeMTMC-reID [۶] و Market1501 [۵] ارزیابی شده است.

مجموعه‌داده‌ی Market1501 : مجموعه‌داده‌ی Market1501 در سال ۲۰۱۵ برای مسئله‌ی بازشناسایی شخص ارائه شده است. تصاویر این مجموعه‌داده در مقابل فروشگاهی در دانشگاه Tisinghua ثبت شده است. در مجموع ۶ دوربین تصاویر را ثبت کردند. ۵ دوربین دارای رزولوشن بالا و یک دوربین دارای رزولوشن پایین هستند. تصاویر هویت‌های این مجموعه‌داده حداقل توسط ۲ دوربین و حداکثر توسط ۶ دوربین ثبت شده‌اند. مجموعه‌داده‌ی Market1501 دارای ۳۲۶۶۸ تصویر از ۱۵۰۱ هویت می‌باشد. معمولاً ۷۵۱ هویت برای آموزش مورد استفاده قرار می‌گیرند و ۷۵۰ هویت نیز به منظور آزمایش مدل استفاده می‌شوند. تصاویر این مجموعه‌داده به سه دسته تقسیم می‌شوند، ۱۲۹۳۶ تصویر از ۷۵۱ هویت برای آموزش، ۱۹۷۳۲ تصویر از ۷۵۰ هویت در گالری و ۳۳۶۸ تصویر از همان ۷۵۰ هویت موجود در گالری، به عنوان کوئری موجود هستند.

[۱] Residual block

مجموعه‌داده‌ی DukeMTMC-reID : این مجموعه‌داده زیرمجموعه‌ای از مجموعه‌داده‌ی DukeMTMC [۵۰] است که در سال ۲۰۱۷ برای وظیفه‌ی بازشناسایی شخص برمبنای تصویر، ارائه شده است. مجموعه‌داده‌ی DukeMTMC شامل ۵۸ دقیقه ویدیوی با رزولوشن بالا از ۸ دوربین مختلف می‌باشد. در DukeMTMC-reID تصاویر عابرین پیاده از ویدیوها در هر ۱۲۰ فریم، برش داده شده‌اند. در مجموع ۳۶۴۱۱ تصویر محدودشده‌ی با برچسب در این مجموعه‌داده وجود دارد. ۷۰۲ هویت برای آموزش مورداستفاده قرار می‌گیرند و ۷۰۲ هویت دیگر برای آزمایش مدل استفاده می‌شوند. تصاویر موجود در این مجموعه‌داده نیز همانند مجموعه‌داده‌ی Market1501 به سه دسته تقسیم می‌شوند، ۱۶۵۲۲ تصویر از ۷۰۲ هویت برای آموزش، ۱۷۶۶۱ تصویر از ۷۰۲ هویت دیگر در گالری و ۲۲۲۸ تصویر از همان ۷۰۲ هویت موجود در گالری، به عنوان کوئری موجود هستند.

در شکل ۳ نمونه‌هایی از تصاویر مجموعه‌داده‌های Market1501 و DukeMTMC-reID نمایش داده شده است.

شکل (۳): نمونه‌هایی از تصاویر مجموعه‌داده‌های Market1501 و DukeMTMC-reID. در هر سمت تصاویری که در ردیف قرار دارند متعلق به یک هویت می‌باشند و توسط دوربین‌های مختلف ثبت شده‌اند.

۵-۲ تنظیمات آزمایش

در ابتدا داده‌های آموزشی و آزمایشی آماده‌سازی می‌شوند. داده‌های آموزشی از دامنه‌های منبع و هدف نرمال‌سازی شده و داده‌افزایی روی آن‌ها اعمال شده و اندازه‌ی تصاویر ورودی به ۱۲۸*۲۵۶ تبدیل می‌شود. روی داده‌های دامنه‌ی هدف داده‌افزایی CamStyle [۴۵] اعمال شده‌است. پس از آماده‌سازی داده‌ها فرآیند آموزش مدل آغاز می‌شود. برای آموزش مدل، از مدل پیش‌آموزش دیده‌ی ResNet-50 [۴۸] که پارامترهای آن روی مجموعه‌داده‌ی ImageNet [۴۹] آموزش دیده‌اند، استفاده شده است. به‌منظور صرفه‌جویی در حافظه‌ی GPU، دو لایه‌ی باقی‌مانده‌ی ابتدایی ثابت می‌شوند. پس از لایه‌ی Pool-5 در ResNet-50 یک لایه‌ی تماماً متصل با اندازه‌ی ۴۰۹۶ اضافه شده و نرخ برون‌اندازی[۱] با احتمال ۰/۵، نرمال‌سازی دسته‌ای[۲] و ReLU پس از آن لایه اعمال می‌شوند.

پس از مرحله‌ی استخراج ویژگی، معماری مدل دارای سه شاخه‌ی مختلف می‌شود. شاخه‌ی اول مربوط به طبقه‌بندی داده‌های دارای برچسب دامنه‌ی منبع می‌باشد. برای این هدف یک لایه‌ی تماماً متصل با اندازه‌ی تعداد هویت‌های موجود در دامنه‌ی منبع، به مدل اضافه می‌شود. تابع اتلاف مربوط به شاخه‌ی اول همان L_s است. میزان تأثیر این تابع اتلاف در تابع اتلاف نهایی شبکه، با پارامتر α کنترل می‌شود که مقدار آن در مدل با استراتژی اول انتخاب همسایه برابر با ۰/۶ و در مدل با استراتژی دوم انتخاب همسایه در تنظیمات Duke→Market برابر با ۰/۴ و در تنظیمات Market→Duke برابر با ۰/۶ درنظر گرفته شده است.

شاخه‌ی دوم مربوط به یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف می‌باشد. در استراتژی اول انتخاب همسایه‌ها، مقدار k (تعداد نمونه‌های مثبت همسایگی) برابر با ۶ درنظر گرفته شده است. در استراتژی دوم انتخاب همسایه‌ها، مقدار آستانه‌ی شباهت یک تصویر با همسایه‌هایش برابر با ۰/۵۵ درنظر گرفته شده است. تابع اتلاف مربوط به شاخه‌ی دوم همان L_{tgt} است. میزان تأثیر این تابع اتلاف در تابع اتلاف نهایی شبکه با پارامتر β کنترل می‌شود. که مقدار آن در مدل با استراتژی اول انتخاب همسایه برابر با ۰/۴ و در مدل با استراتژی دوم انتخاب همسایه در تنظیمات Duke→Market برابر با ۰/۶ و در تنظیمات Market→Duke برابر با ۰/۴ درنظر گرفته شده است.

شاخه‌ی سوم مربوط به تابع اتلاف سه‌گانه است. در این شاخه یک لایه‌ی تماماً متصل با اندازه‌ی ۲۵۶ اضافه می‌شود. خروجی این لایه به‌عنوان ویژگی مورداستفاده در محاسبه‌ی تابع اتلاف سه‌گانه کاربرد دارد. مقدار پارامتر حاشیه‌ای ρ در تابع اتلاف سه‌گانه برابر ۰/۳ درنظر گرفته شده است. میزان تأثیر این تابع اتلاف سه‌گانه در تابع اتلاف نهایی شبکه با پارامتر δ کنترل می‌شود و مقدار این پارامتر در مدل پیشنهادی با استراتژی اول انتخاب همسایه و مدل پیشنهادی با استراتژی دوم انتخاب همسایه برابر با ۰/۵ است.

برای آموزش مدل از بهینه‌ساز SGD با مومنتوم ۰/۹ و کاهش وزنی[۳] برابر با $4 * 10^{-5}$ استفاده شده است. در ۴۰ دوره‌ی[۴] اول نرخ یادگیری برای لایه‌های پایه در ResNet-50 برابر با ۰/۰۱ و برای سایر لایه‌ها برابر با ۰/۱ می‌باشد. پس از ۴۰ دوره نرخ یادگیری بر ۱۰ تقسیم می‌شود. اندازه‌ی دسته برای داده‌های منبع و هدف برابر ۱۲۸ و برای داده‌های سه‌گانه برابر ۶۴ درنظر گرفته می‌شود.

---

[۱] Dropout

[۲] Batch normalization

[۳] Weight decay

[۴] Epoch

۵-۳ بررسی عملکرد مدل با معماری‌های مختلف CNN

به منظور بررسی تأثیر معماری CNN در عملکرد مدل نتایج اجرای آزمایشات برای معماری‌های ResNet-50 [۴۹]، ResNext-50 [۴۸] و WideResNet-50 [۵۲] در جدول (۱) و جدول (۲) آورده شده است. به‌منظور صرفه‌جویی در حافظه‌ی GPU، دو لایه‌ی باقی‌مانده‌ی ابتدایی در هر سه معماری ثابت شده اند. همانطور که در جدول (۱) و جدول (۲) مشخص است بهترین نتایج مربوط به زمانی است که از ResNext-50 برای استخراج ویژگی استفاده شده است. علاوه بر این ResNext-50 نسبت به ResNet-50 و WideResNet-50 تعداد پارامترهای کمتری دارد و این موضوع یک مزیت به حساب می‌آید. بنابراین در مدل پیشنهادی برای استخراج ویژگی از شبکه‌ی پیش‌آموزش‌دیده‌ی ResNext-50 استفاده می‌کنیم.

جدول (۱): تأثیر معماری CNN بر عملکرد مدل با استراتژی اول انتخاب همسایه‌ها در تنظیمات Market→Duke و Duke→Market.

CNN Architecture	#Trainable Parameters	Duke→Market		Market→Duke	
		R-1	mAP	R-1	mAP
ResNet-50	≈ 36 million	76.7	45.5	63.4	40.8
ResNext-50	≈ 35 million	**77.4**	**46**	**64**	**41.1**
WideResNet-50	≈ 77 million	76	43.2	63.8	41

جدول (۲): تأثیر معماری CNN بر عملکرد مدل با استراتژی دوم انتخاب همسایه‌ها در تنظیمات Market→Duke و Duke→Market.

CNN Architecture	#Trainable Parameters	Duke→Market		Market→Duke	
		R-1	mAP	R-1	mAP
ResNet-50	≈ 36 million	83.3	60.5	67.7	47.3
ResNext-50	≈ 35 million	**84.5**	**63**	**70.1**	**49.1**
WideResNet-50	≈ 77 million	83.8	62.3	69.9	48.7

۵-۴ بررسی تأثیر اجزای مختلف تابع اتلاف نهایی شبکه روی عملکرد مدل

تابع اتلاف نهایی شبکه که در بخش ۴ـ۴ توضیح داده شده است، از مجموع تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع (L_s)، تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف (L_{tgt}) و یک تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف (L_{tri}) تشکیل شده است. به‌منظور بررسی تأثیر هرکدام از مؤلفه‌های تابع اتلاف نهایی روی عملکرد مدل آزمایش‌های متعددی صورت گرفته و نتایج این آزمایش‌ها روی مجموعه‌داده‌های Market1501 و DukeMTMC-reID در جدول (۳) و جدول (۴) ثبت شده است. همچنین در جدول (۳) و جدول (۴) عملکرد مدل به شکل بانظارت (شرایطی که مجموعه‌داده‌ی هدف همان مجموعه‌داده‌ی منبع می‌باشد) نیز آورده شده است.

همانطورکه از نتایج ثبت شده در جدول مشخص است، عملکرد مدل در حالت بانظارت قابل قبول می‌باشد. اما در شرایطی که مجموعه‌داده‌های منبع و هدف متفاوت هستند، استفاده از تابع اتلاف L_s به تنهایی، عملکرد ضعیفی دارد. در این حالت فقط مجموعه‌داده‌ی منبع در فرآیند آموزش شرکت دارد و آزمایش روی دامنه‌ی هدف صورت می‌گیرد. بنابراین مدل قابلیت تعمیم پذیری مناسبی روی دامنه‌ی هدف ندارد. توابع اتلاف L_{tgt} و L_{tri} نیز هر دو تأثیر مثبتی در عملکرد مدل دارند.

جدول (۳): بررسی عملکرد مدل پیشنهادی به شکل بانظارت و تأثیر مؤلفه‌های تابع اتلاف نهایی شبکه بر عملکرد مدل پیشنهادی در هنگام آزمایش روی Market1501. L_s : تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع، L_{tgt} تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف و L_{tri}: تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف.

Method	Market1501		
	Src	R-1	mAP
Supervised Learning	Market	86.6	70.9
Ours (L_s)	Duke	43.1	18.1
Ours ($L_s + L_{tgt}$)	Duke	82.5	60.6
Ours ($L_s + L_{tri}$)	Duke	67.1	34.3
Ours ($L_s + L_{tgt} + L_{tri}$)	Duke	84.5	63

جدول (۴): بررسی عملکرد مدل پیشنهادی به شکل بانظارت و تأثیر مؤلفه‌های تابع اتلاف نهایی شبکه بر عملکرد مدل پیشنهادی در هنگام آزمایش روی DukeMTMC-reID. L_s: تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع، L_{tgt} تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف و L_{tri}: تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف.

Method	DukeMTMC-reID		
	Src	R-1	mAP
Supervised Learning	Duke	75.6	57.8
Ours (L_s)	Market	29	14.8
Ours ($L_s + L_{tgt}$)	Market	68.8	47.5
Ours ($L_s + L_{tri}$)	Market	48.2	29
Ours ($L_s + L_{tgt} + L_{tri}$)	Market	70.1	49.1

۵-۵ بررسی میزان مصرف حافظه‌ی GPU

مدل پیشنهادی و مدل‌های ECN [۴۰]، AE [۴۱] و GPP [۴۲] از لحاظ یادگیری بدون نظارت ویژگی‌های درون‌ـ دامنه‌ای در دامنه‌ی هدف رویکردهای تقریباً مشابهی دارند. تفاوت عمده‌ی آن‌ها در استراتژی انتخاب همسایه برای نمونه‌های دامنه‌ی هدف می‌باشد. مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها عملکرد بهتری از مدل‌های ECN و AE دارد. اما مدل GPP نسبت به مدل پیشنهادی، به ویژه در هنگام آزمایش روی مجموعه‌داده‌ی DukeMTMC-reID، اندکی دقت بالاتری دارد. مدل GPP از ساختار گرافی برای انتخاب همسایه‌ها در دامنه‌ی هدف بهره می‌برد. ساختار گراف قادر به استخراج روابط پیچیده در داده‌ها می‌باشد. به همین دلیل عملکرد این مدل هنگام آزمایش روی مجموعه‌داده‌ی DukeMTMC-reID، که مجموعه‌داده‌ای چالشی‌تر نسبت به مجموعه‌داده‌ی Market1501 است، بهبود قابل توجهی دارد. اما مدل GPP از لحاظ پیچیدگی و میزان مصرف حافظه‌ی GPU تفاوت قابل ملاحظه‌ای با مدل پیشنهادی دارد.

در جدول (۵) مدل‌های ECN، AE، GPP و مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها از نظر عملکرد و میزان مصرف حافظه‌ی GPU با هم مقایسه شده‌اند. از نتایج ثبت شده در جدول (۵) می‌توان نتیجه گرفت که مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها با وجود عملکرد خوبی که دارد میزان مصرف حافظه‌ی آن نیز قابل قبول است. همچنین استفاده از تابع اتلاف سه‌گانه در مدل پیشنهادی میزان مصرف حافظه‌ی GPU را به میزان اندکی افزایش داده ولی نتیجه را بهبود داده است.

جدول (۵): بررسی میزان مصرف حافظه‌ی GPU در زمان آموزش مدل و عملکرد مدل پیشنهادی و سه مدل دیگر در تنظیمات Market→Duke و Duke→Market. L_s: تابع اتلاف یادگیری بانظارت ویژگی‌های دامنه‌ی منبع، L_{tgt} : تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف و L_{tri}: تابع اتلاف سه‌گانه برای یادگیری توأم ویژگی‌های دامنه‌های منبع و هدف.

Method	GPU (MB)	Duke→Market		Market→Duke	
		R-1	mAP	R-1	mAP
ECN [۴۰]	≈7200	75.1	43	63.3	40.4
AE [۴۱]	≈7000	81.6	58	67.9	46.7
GPP [۴۲]	≈9800	84.1	63.8	74	54.4
Ours ($L_s + L_{tgt}$)	≈7000	82.5	60.6	68.8	47.5
Ours ($L_s + L_{tgt} + L_{tri}$)	≈7100	84.5	63	70.1	49.1

۶-۵ مقایسه‌ی مدل پیشنهادی با سایر مدل‌ها

در جدول (۶) عملکرد مدل پیشنهادی با ۱۱ مدل وفق‌دهی دامنه‌ی بدون نظارت برای بازشناسایی شخص مقایسه شده است. همانطور که از نتایج جدول مشخص است عملکرد مدل پیشنهادی با استراتژی اول انتخاب همسایه‌ها از تمامی روش‌های موجود در جدول به‌غیر از AE [۴۱] و GPP [۴۲] بهتر می‌باشد. عملکرد مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها نیز از تمامی روش‌های موجود در جدول به‌غیر از GPP بهتر می‌باشد.

یکی از دلایل اصلی بهبود عملکرد در روش پیشنهادی، درنظرگرفتن ویژگی‌های درون‌ـ دامنه‌ای در دامنه‌ی هدف، طی فرآیند آموزش می‌باشد.علاوه‌بر تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف، در روش پیشنهادی از یک تابع اتلاف سه‌گانه نیز بهره برده شده است. استفاده از تابع اتلاف سه‌گانه سبب می‌شود که علاوه‌بر تغییرات درون‌ـ دامنه‌ای در دامنه‌ی هدف، تفاوت‌های بین دامنه‌های منبع و هدف نیز یاد گرفته شود. بدین‌ترتیب مدل می‌تواند دانش یادگرفته شده از دامنه‌ی برچسب‌گذاری‌شده‌ی منبع را به دامنه‌ی بدون برچسب هدف منتقل کند. یکی دیگر از دلایل بهبود عملکرد مدل پیشنهادی، استفاده از شبکه‌ی پیش‌آموزش دیده‌ی ResNext-50 به‌منظور استخراج ویژگی است. استفاده از این شبکه نسبت به شبکه‌ی ResNet-50 که معماری متداول در مدل‌های عمیق بازشناسایی شخص می‌باشد، باعث بهبود عملکرد مدل شده است و تعداد پارامترهای کمتری نیز دارد.

همانطور که از نتایج جدول (۶) مشخص است، عملکرد مدل در تنظیمات Duke→Market بهتر از عملکرد مدل در تنظیمات Market→Duke می‌باشد. مجموعه‌داده‌ی DukeMTMC-reID نسبت به مجموعه‌داده‌ی Market1501 تصاویر چالشی‌تری دارد. بنابراین هنگامی‌که مجموعه‌داده‌ی DukeMTMC-reID به عنوان مجموعه‌داده‌ی برچسب‌گذاری‌شده‌ی دامنه‌ی منبع و مجموعه‌داده‌ی Market1501 به عنوان دامنه‌ی هدف برای آموزش مدل استفاده شوند، مدل عملکرد بهتری نسبت شرایطی دارد که مجموعه‌داده‌ی Market1501 به عنوان دامنه‌ی منبع و مجموعه‌داده‌ی DukeMTMC-reID به عنوان دامنه‌ی هدف استفاده شدند. به طور کلی هرچه تصاویر دامنه‌ی منبع پیچیده‌تر و چالشی‌تر باشند و تصاویر دامنه‌ی هدف پیچیدگی کمتری داشته باشند عملکرد مدل بهتر است.

مدل‌های ECN، AE و GPP علاوه بر تلاش برای کم کردن فاصله‌ی بین دامنه‌های منبع و هدف، سعی می‌کنند ویژگی‌های دامنه‌ی هدف را نیز طی فرآیند آموزش یاد بگیرند. این سه روش، استراتژی‌های انتخاب همسایه‌ی متفاوتی دارند. مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها از روش‌های ECN و AE روی هر دو مجموعه‌داده‌ی Maket1501 و DukeMTMC-reID عملکرد بهتری دارد. مقدار معیار CMC در تنظیمات Duke→Market در مدل پیشنهادی از تمامی روش‌های موجود در جدول بهتر است. اما

در این تنظیمات معیار mAP در مدل GPP به میزان ۰/۸ درصد نسبت به مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها بالاتر است. همچنین در تنظیمات Market→Duke روش GPP عملکرد بهتری نسبت به مدل پیشنهادی دارد. دلیل عملکرد بهتر مدل GPP نسبت به مدل پیشنهادی، استفاده از ساختار گرافی در استراتژی انتخاب همسایه‌ها در این مدل می‌باشد. اما از لحاظ میزان پیچیدگی مدل و میزان مصرف حافظه‌ی GPU تفاوت قابل ملاحظه‌ای با مدل پیشنهادی ما دارد. این روش نسبت به مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها مقدار MB ۲۷۰۰ حافظه‌ی GPU بیشتری مصرف می‌کند. بنابراین مدل پیشنهادی علاوه‌بر عملکرد مناسب، میزان مصرف حافظه‌ی معقولی نیز دارد.

جدول (۶): مقایسه‌ی عملکرد مدل پیشنهادی با استراتژی اول و استراتژی دوم انتخاب همسایه‌ها در دامنه‌ی هدف با سایر مدل‌های وفق‌دهی دامنه‌ی بدون نظارت در بازشناسایی شخص در تنظیمات Market→Duke و Duke→Market.

Method	Duke→Market (%)				Market→Duke (%)			
	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
PTGAN [۳۵]	38.6	-	66.1	-	27.4	-	50.7	-
SPGAN [۳۶]	51.5	70.1	76.8	22.8	41.1	56.6	63	22.3
TJ-AIDL [۴۳]	58.2	74.8	81.1	26.5	44.3	59.6	65	23
CamStyle [۵۳]	58.8	78.2	84.3	27.4	48.4	62.5	68.9	25.1
HHL [۳۹]	62.2	78.8	84	31.4	46.9	61	66.7	27.2
ARN [۵۴]	70.3	80.4	86.3	39.4	60.2	73.9	79.5	33.4
ATNet [۳۷]	55.7	73.2	79.4	25.6	45.1	59.5	64.2	24.9
DAL [۵۵]	64.3	-	-	34.5	55.4	-	-	36.7
ECN [۴۰]	75.1	87.6	91.6	43	63.3	75.8	80.4	40.4
AE [۴۱]	81.6	91.9	94.6	58	67.9	79.2	83.6	46.7
GPP [۴۲]	84.1	92.8	95.4	**63.8**	**74**	**83.7**	**87.4**	**54.4**
Ours (strategy1)	77.4	89.1	92.7	46	64	76.8	80.9	41.1
Ours (strategy2)	**84.5**	**93.1**	**95.7**	63	70.1	80.8	84.1	49.1

۶ جمع‌بندی

در این پژوهش مدلی برای وفق‌دهی دامنه‌ی بدون نظارت در حوزه‌ی بازشناسایی شخص ارائه شده است. در این مدل، داده‌های دارای برچسب دامنه‌ی منبع و داده‌های بدون برچسب دامنه‌ی هدف، برای آموزش مدل استفاده می‌شوند و مدل در هنگام آزمایش روی دامنه‌ی هدف، عملکرد بسیار خوبی دارد. به‌منظور استخراج ویژگی داده‌های دامنه‌های منبع و هدف، از شبکه‌ی پیش‌آموزش دیده‌ی ResNext-50 استفاده شده است که نسبت به شبکه‌ی ResNet-50 تعداد پارامترهای قابل آموزش کمتر و عملکرد بهتری دارد.

در مدل پیشنهادی، از تابع اتلاف یادگیری بانظارت ویژگی‌ها در داده‌های منبع، تابع اتلاف یادگیری بدون نظارت ویژگی‌ها در دامنه‌ی هدف و تابع اتلاف سه‌گانه استفاده شده است. در تابع اتلاف یادگیری بدون نظارت ویژگی‌های دامنه‌ی هدف، برای نحوه‌ی انتخاب همسایه‌ها، دو استراتژی مورد آزمایش قرار گرفته‌اند. مدل پیشنهادی با استراتژی اول انتخاب همسایه‌ها توانسته است در تنظیمات Duke→Market در رتبه‌ی ۱ معیار CMC مقدار ۷۷/۴ درصد و مقدار mAP ۴۶ درصد را به‌دست آورد. همچنین مدل پیشنهادی با استراتژی اول انتخاب همسایه‌ها در تنظیمات Market→Duke در رتبه‌ی ۱ معیار CMC مقدار ۶۴ درصد و مقدار mAP ۴۱/۱ درصد را به‌دست آورده است.

مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها توانسته در تنظیمات Duke→Market در رتبه‌ی ۱ معیار CMC مقدار ۸۴/۵ درصد و مقدار mAP ۶۳ درصد را به‌دست آورد. همچنین مدل پیشنهادی با استراتژی دوم انتخاب همسایه‌ها در تنظیمات Market→Duke در رتبه‌ی ۱ معیار CMC مقدار ۷۰/۱ درصد و مقدار mAP ۴۹/۱ درصد را به‌دست آورده است. این مقادیر به‌دست آمده نتایج بسیار خوبی در این حوزه به‌شمار می‌روند.

مراجع

[۱] فقیه ایمانی، صبا سادات و فولادی قلعه، کاظم، «**مروری بر روش‌های ارائه شده برای مسئله‌ی بازشناسایی شخص با رویکرد یادگیری عمیق و چالش‌های آن**»، یازدهمین کنفرانس ملی و اولین کنفرانس بین‌المللی بینایی ماشین و پردازش تصویر ایران، ۱۳۹۸.

[۲] فقیه ایمانی، صبا سادات، فولادی قلعه، کاظم و آقابابا، حسین، «**مدلی تعمیم‌پذیر برای وفق‌دهی دامنه‌ی بدون نظارت در مسئله‌ی بازشناسایی شخص**»، پنجمین کنفرانس بین‌المللی بازشناسی الگو و تحلیل تصویر، ایران، کاشان، ۱۴۰۰.

[3] Gray, Douglas, and Hai Tao. "Viewpoint invariant pedestrian recognition with an ensemble of localized

features." In European conference on computer vision, pp. 262-275. Springer, Berlin, Heidelberg, 2008.

[4] Li, Wei, Rui Zhao, Tong Xiao, and Xiaogang Wang. "Deepreid: Deep filter pairing neural network for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 152-159. 2014.

[5] Zheng, Liang, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. "Scalable person re-identification: A benchmark." In Proceedings of the IEEE international conference on computer vision, pp. 1116-1124. 2015.

[6] Zheng, Zhedong, Liang Zheng, and Yi Yang. "Unlabeled samples generated by gan improve the person re-identification baseline in vitro." In Proceedings of the IEEE international conference on computer vision, pp. 3754-3762. 2017.

[7] Bromley, Jane, James W. Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. "Signature verification using a "siamese" time delay neural network." International Journal of Pattern Recognition and Artificial Intelligence 7, no. 04 (1993): 669-688.

[8] Zhong, Zhun, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. "Random erasing data augmentation." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 07, pp. 13001-13008. 2020.

[9] Ye, Mang, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. "Deep learning for person re-identification: A survey and outlook." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021.

[10] Wu, Di, Si-Jia Zheng, Xiao-Ping Zhang, Chang-An Yuan, Fei Cheng, Yang Zhao, Yong-Jun Lin, Zhong-Qiu Zhao, Yong-Li Jiang, and De-Shuang Huang. "Deep learning-based methods for person re-identification: A comprehensive review." Neurocomputing 337, 354-371. 2019.

[11] Yaghoubi, Ehsan, Aruna Kumar, and Hugo Proença. "Sss-pr: A short survey of surveys in person re-identification." Pattern Recognition Letters 143, 50-57. 2021.

[12] Huang, Timothy, and Stuart Russell. "Object identification in a bayesian context." In IJCAI, vol. 97, pp. 1276-1282. 1997.

[13] Zajdel, Wojciech, Zoran Zivkovic, and Ben JA Krose. "Keeping track of humans: Have I seen this person before?". In Proceedings of the 2005 IEEE international conference on robotics and automation, pp. 2081-2086. IEEE, 2005.

[14] Gheissari, Niloofar, Thomas B. Sebastian, and Richard Hartley. "Person reidentification using spatiotemporal appearance." In 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), vol. 2, pp. 1528-1535. IEEE, 2006.

[15] Bazzani, Loris, Marco Cristani, Alessandro Perina, Michela Farenzena, and Vittorio Murino. "Multiple-shot person re-identification by hpe signature." In 2010 20th International Conference on Pattern Recognition, pp. 1413-1416. IEEE, 2010.

[16] Farenzena, Michela, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. "Person re-identification by symmetry-driven accumulation of local features." In 2010 IEEE computer society conference on computer vision and pattern recognition, pp. 2360-2367. IEEE, 2010.

[17] Dikmen, Mert, Emre Akbas, Thomas S. Huang, and Narendra Ahuja. "Pedestrian recognition with a learned metric." In Asian conference on Computer vision, pp. 501-512. Springer, Berlin, Heidelberg, 2010.

[18] Zheng, Wei-Shi, Shaogang Gong, and Tao Xiang. "Person re-identification by probabilistic relative distance comparison." In CVPR 2011, pp. 649-656. IEEE, 2011.

[19] Zheng, Wei-Shi, Shaogang Gong, and Tao Xiang. "Transfer re-identification: From person to set-based verification." In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2650-2657. IEEE, 2012.

[20] Yi, Dong, Zhen Lei, Shengcai Liao, and Stan Z. Li. "Deep metric learning for person re-identification." In 2014 22nd International Conference on Pattern Recognition, pp. 34-39. IEEE, 2014.

[21] Li, Wei, Rui Zhao, Tong Xiao, and Xiaogang Wang. "Deepreid: Deep filter pairing neural network for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 152-159. 2014.

[22] Quan, Ruijie, Xuanyi Dong, Yu Wu, Linchao Zhu, and Yi Yang. "Auto-reid: Searching for a part-aware convnet for person re-identification." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3750-3759. 2019.

[23] Wang, Guangcong, Jianhuang Lai, Peigen Huang, and Xiaohua Xie. "Spatial-temporal person re-identification." In Proceedings of the AAAI conference on artificial intelligence, vol. 33, no. 01, pp. 8933-8940. 2019.

[24] Zheng, Zhedong, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. "Joint discriminative and generative learning for person re-identification." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2138-2147. 2019.

[25] Zhou, Kaiyang, Yongxin Yang, Andrea Cavallaro, and Tao Xiang. "Omni-scale feature learning for person re-identification." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3702-3712. 2019.

[26] Varior, Rahul Rama, Mrinal Haloi, and Gang Wang. "Gated siamese convolutional neural network architecture

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7202-7211. 2019.

[38] Yan, Shuanglin, Yafei Zhang, Minghong Xie, Dacheng Zhang, and Zhengtao Yu. "Cross-domain person re-identification with pose-invariant feature decomposition and hypergraph structure alignment." Neurocomputing 467, 229-241. 2022.

[39] Zhong, Zhun, Liang Zheng, Shaozi Li, and Yi Yang. "Generalizing a person retrieval model hetero-and homogeneously." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 172-188. 2018.

[40] Zhong, Zhun, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. "Invariance matters: Exemplar memory for domain adaptive person re-identification." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 598-607. 2019.

[41] Ding, Yuhang, Hehe Fan, Mingliang Xu, and Yi Yang. "Adaptive exploration for unsupervised person re-identification." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16, no. 1 (2020): 1-19.

[42] Zhong, Zhun, Liang Zheng, Zhiming Luo, Shaozi Li, and Yi Yang. "Learning to adapt invariance in memory for person re-identification." IEEE transactions on pattern analysis and machine intelligence (2020).

[43] Wang, Jingya, Xiatian Zhu, Shaogang Gong, and Wei Li. "Transferable joint attribute-identity deep learning for unsupervised person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2275-2284. 2018.

[44] Wang, Jun. "Graph-based local feature adaptation for cross-domain person re-identification." IEEE Access 2022.

[45] Zhong, Zhun, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. "Camera style adaptation for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5157-5166. 2018.

[46] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.

[47] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.

[48] Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492-1500. 2017.

for human re-identification." In European conference on computer vision, pp. 791-808. Springer, Cham, 2016.

[27] Chung, Dahjung, Khalid Tahboub, and Edward J. Delp. "A two stream siamese convolutional neural network for person re-identification." In Proceedings of the IEEE international conference on computer vision, pp. 1983-1991. 2017.

[28] Zheng, Meng, Srikrishna Karanam, Ziyan Wu, and Richard J. Radke. "Re-identification with consistent attentive siamese networks." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5735-5744. 2019.

[29] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[30] Huang, Yan, Jingsong Xu, Qiang Wu, Zhedong Zheng, Zhaoxiang Zhang, and Jian Zhang. "Multi-pseudo regularized label for generated data in person re-identification." IEEE Transactions on Image Processing 28, no. 3 (2018): 1391-1403.

[31] Gulrajani, Ishaan, Ahmed, Faruk, Arjovsky, Martin, Dumoulin, Vincent, and Courville, Aaron C. "Improved training of wasserstein gans." In Advances in neural information processing systems, pages 5767–5777, 2017.

[32] Ge, Yixiao, Li, Zhuowan, Zhao, Haiyu, Yin, Guojun, Yi, Shuai, Wang, Xiaogang, et al. "Fd-gan: Pose-guided feature distilling gan for robust person re-identification." In Advances in neural information processing systems, pages 1222–1233, 2018.

[33] Ma, Liqian, Jia, Xu, Sun, Qianru, Schiele, Bernt, Tuytelaars, Tinne, and Van Gool, Luc." Pose guided person image generation." In Advances in Neural Information Processing Systems, pages 406–416, 2017.

[34] Yaghoubi, Ehsan, Diana Borza, SV Aruna Kumar, and Hugo Proença. "Person re-identification: Implicitly defining the receptive fields of deep learning classification frameworks." Pattern Recognition Letters 145, 23-29. 2021.

[35] Wei, Longhui, Shiliang Zhang, Wen Gao, and Qi Tian. "Person transfer gan to bridge domain gap for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 79-88. 2018.

[36] Deng, Weijian, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 994-1003. 2018.

[37] Liu, Jiawei, Zheng-Jun Zha, Di Chen, Richang Hong, and Meng Wang. "Adaptive transfer network for cross-domain person re-identification." In Proceedings of the

[49] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[50] Ristani, Ergys, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. "Performance measures and a data set for multi-target, multi-camera tracking." In European conference on computer vision, pp. 17-35. Springer, Cham, 2016.

[51] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.

[52] Zagoruyko, Sergey, and Nikos Komodakis. "Wide residual networks." arXiv preprint arXiv:1605.07146 (2016).

[53] Zhong, Zhun, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. "Camstyle: A novel data augmentation method for person re-identification." IEEE Transactions on Image Processing 28, no. 3 (2018): 1176-1190.

[54] Li, Yu-Jhe, Fu-En Yang, Yen-Cheng Liu, Yu-Ying Yeh, Xiaofei Du, and Yu-Chiang Frank Wang. "Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 172-178. 2018.

[55] Qi, Lei, Lei Wang, Jing Huo, Luping Zhou, Yinghuan Shi, and Yang Gao. "A novel unsupervised camera-aware domain adaptation framework for person re-identification." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8080-8089. 2019.

صباسادات فقیه‌ایمانی مدرک کارشناسی مهندسی کامپیوتر-نرم‌افزار را در سال ۱۳۹۷ و مدرک کارشناسی ارشد مهندسی فناوری اطلاعات را در سال ۱۳۹۹ از دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران اخذ کرد. زمینه‌های پژوهشی موردعلاقه ایشان پردازش تصویر و یادگیری عمیق می‌باشد.

کاظم فولادی قلعه مدرک دکتری خود را در رشته‌ی مهندسی برق و کامپیوتر در گرایش هوش مصنوعی و رباتیک از دانشکده مهندسی برق و کامپیوتر دانشکدگان فنی دانشگاه تهران اخذ نمود. ایشان از سال ۱۳۹۵ تا کنون عضو هیئت علمی دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران و در حال حاضر سرپرست آزمایشگاه پژوهشی فضای سایبر و آزمایشگاه پژوهشی یادگیری عمیق می‌باشند. حوزه‌های پژوهشی مورد علاقه ایشان شامل هوش مصنوعی، پردازش تصویر و بینایی کامپیوتری، یادگیری عمیق، سایبرنتیک و مطالعات فضای سایبر می‌باشد.

حسین آقابابا مدرک دکتری خود را در رشته‌ی مهندسی برق در گرایش الکترونیک از دانشکده مهندسی برق و کامپیوتر دانشکدگان فنی دانشگاه تهران اخذ نمود. ایشان از سال ۱۳۹۴ تا کنون عضو هیئت علمی دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران و در حال حاضر سرپرست آزمایشگاه پژوهشی محاسبات و ارتباطات کوانتومی می‌باشند. حوزه‌های پژوهشی مورد علاقه ایشان شامل محاسبات کوانتومی و یادگیری ماشین کوانتومی می‌باشد.