مقاله پژوهشی

مجله ماشین‌بینایی و پردازش‌تصویر
سال دوازدهم، شماره اول، بهار ۱۴۰۴

داده افزایی با استفاده از شبکه‌های مولد تخاصمی جهت بهبود بازتشخیص افراد

هادی کاوه[۱]، محمدشهرام معین[۲] و فربد رزازی[۳]

چکیده

در سال‌های اخیر، پژوهش‌ها در زمینه بازتشخیص افراد به‌طور پیوسته پیشرفت کرده‌اند. در این راستا، شبکه‌های عصبی متخاصم به موفقیت چشمگیری دست یافته‌اند و به عنوان یک رویکرد برجسته در این حوزه شناخته می‌شوند. در این مقاله، با بررسی کاربرد انواع شبکه‌های متخاصم در بازتشخیص افراد، از ترکیب شبکه ATNet و مدل Pix2Pix برای داده‌افزایی استفاده شده است. مدل Pix2Pix که در تبدیل تصویر به تصویر در زمینه‌های مختلف موفقیت‌آمیز بوده، از الگوریتم PatchGAN به عنوان متمایزکننده و U-Net به عنوان تولیدکننده بهره می‌گیرد. روش پیشنهادی برای داده‌افزایی مبتنی بر وجود تصاویر فرد مورد نظر از چهار جهت (روبه‌رو، پشت، سمت چپ و راست) است. پس از دسته‌بندی تصاویر در مجموعه‌داده‌های Market-1501 و CUHK03، تصاویر نماهای ناموجود با استفاده از شبکه ATNet تولید شده‌اند. مقایسه‌های انجام‌شده نشان‌دهنده بهبود عملکرد روش‌های پیشرو در بازتشخیص افراد با داده افزایی پیشنهادی در این مقاله است.

کلید واژه‌ها

بازتشخیص افراد، داده افزایی، شبکه عصبی تخاصمی، سیستم Pix2Pix**، شبکه** ATNet **و مدل** U-Net

۱- مقدمه

با توسعه سریع جوامع مدرن، افزایش تراکم جمعیت در برخی از مکان‌های عمومی به راحتی سبب وقوع حوادث گوناگون می‌گردد. به منظور پیشگیری و مقابله به موقع با چنین حوادثی، تعداد زیادی دوربین نظارتی در مکان‌های عمومی نصب و به کار گرفته می‌شوند.

اما بررسی دقیق و سریع داده‌های نظارتی عظیم توسط انسان امری دشوار است. انسان برای کمک به خود یا حتی جایگزینی خود در برنامه‌های نظارتی، به استفاده از فناوری یادگیری ماشین روی آورده است[۱].

یکی از موضوعات مهم در پردازش تصویر، مسئله بازتشخیص فرد است، به این شکل که پس از این که فرد توسط یک دوربین مشاهده شد، شناسایی دقیق و سریع او در میان تعداد زیادی از افراد به کمک سایر دوربین‌های شبکه نظارتی انجام شود. دستیابی به فناوری بازتشخیص، میزان مشارکت انسان در نظارت تصویری را تا حد زیادی کاهش می‌دهد و همچنین می‌توان به کمک آن مسیر گذر افراد را به طور دقیق تجزیه و تحلیل کرد. بنابراین این فناوری نقش مهمی در پیشگیری از جرم و حفظ امنیت اجتماعی دارد[۲].

در شکل (۱) فرآیند بازشناسی افراد در یک سیستم نظارت تصویری به‌صورت بصری نمایش داده شده است. بازشناسی افراد به فرآیند شناسایی و تطبیق یک فرد خاص در تصاویر یا ویدیوهای مختلف از زاویه‌های گوناگون و شرایط نوری متفاوت اشاره دارد.

این مقاله در مردادماه سال ۱۴۰۳ دریافت شد؛ در اردیبهشت‌ماه بازنگری و سپس پذیرفته گردید.

[۱] دانشجوی دکترای تخصصی رشته برق - مخابرات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
رایانامه: hadi.kaveh۸۶@gmail.com

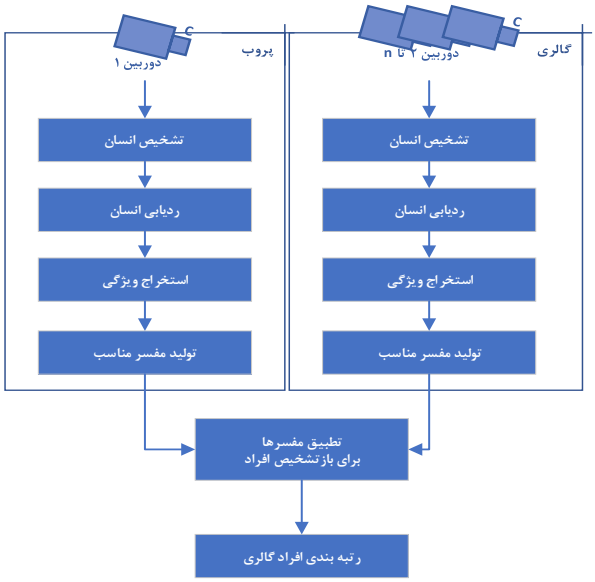
[۲] عضو هیئت علمی پژوهشگاه ارتباطات و فناوری اطلاعات
رایانامه: moin@itrc.ac.ir

[۳] عضو هیئت علمی دانشکده مهندسی برق و کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
رایانامه: razzazi@srbiau.ac.ir

نویسنده مسئول: محمدشهرام معین

این فرآیند به‌ویژه در سیستم‌های نظارتی و امنیتی کاربرد فراوانی دارد و به شناسایی، پیگیری و شبیه‌سازی هویت افراد در محیط‌های مختلف کمک می‌کند. به عبارت دیگر، سیستم بازشناسی افراد تلاش می‌کند تا با استفاده از ویژگی‌های خاص فرد، مانند چهره، رفتار، یا سایر ویژگی‌های فیزیکی، او را در تصاویر و ویدیوهای گرفته‌شده از دوربین‌های مختلف شناسایی کند.

برای مثال، در یک شبکه دوربین‌های نظارتی، ممکن است فردی از میدان دید یک دوربین خارج شده و سپس وارد میدان دید دوربین‌های دیگر شود. در چنین شرایطی، سیستم بازشناسی افراد با تحلیل ویژگی‌های خاص و منحصر به فرد هر فرد (مانند ویژگی‌های چهره یا رفتار حرکتی)، تلاش می‌کند تا همان فرد را مجدداً شناسایی کرده و ردگیری کند. این فرآیند از طریق الگوریتم‌های پیشرفته یادگیری ماشین و بینایی کامپیوتری انجام می‌شود که قادرند تغییرات زاویه دید، شرایط نوری و حتی پوشش‌هایی که ممکن است صورت یا سایر ویژگی‌های فرد را پنهان کنند، شبیه‌سازی کرده و تطبیق دهند.

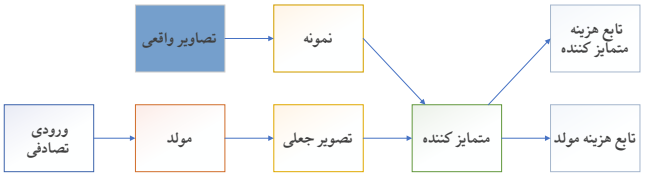


شکل(۱). بازتشخیص افراد در یک سیستم نظارت تصویری [۴]

در حال حاضر، فناوری پردازش تصویر و تشخیص چهره به طور گسترده در بسیاری از کاربردها مورد استفاده قرار می‌گیرد. در سیستم‌های نظارتی مدرن، عوامل مختلف مانند دید دوربین، انسداد ناشی از محیط یا انسداد ناشی از عبور افراد پیاده و همین طور وضوح ناکافی، موجب عدم مشاهده تصویر باکیفیت از چهره افراد می‌شود. این چالش‌ها موجب می‌شود که استفاده از چهره انسان به تنهایی برای تشخیص عابرین از کارآیی کافی برخوردار نباشد؛ بنابراین فناوری بازتشخیص که ویژگی‌های کامل بدن عابرین پیاده را نیز بررسی می‌کند، به یک راه حل جایگزین مهم تبدیل شده است.

پس از ده سال تحقیق، به خصوص در سال‌های اخیر، با توسعه یادگیری عمیق، تحقیقات در مورد بازتشخیص افراد به طور مداوم بهبود پیدا کرده است. در این میان، شبکه‌های عصبی متخاصم به موفقیت بالایی در این زمینه دست یافته اند و کاربرد آنها به عنوان رویکردی ممتاز در این زمینه در نظر گرفته می‌شود[۳].

شبکه‌های عصبی مولد متخاصم یا به اصطلاح GAN ها از یک مدل تولید کننده و یک مدل متمایز کننده تشکیل شده که هر دوی آن‌ها توسط شبکه عصبی کانولوشنی پیاده سازی شده‌اند. از مدل تولید کننده برای تولید داده جدید استفاده می‌شود و به صورت نظارت نشده است، اما از متمایز کننده برای طبقه بندی داده تولید شده و تشخیص داده اصلی از جعلی استفاده می‌شود و به صورت نظارت شده می‌باشد.[۴]. در شکل(۲) ساختار شبکه عصبی متخاصم نشان داده شده است.



شکل(۲). ساختار شبکه عصبی متخاصم

در این مقاله ضمن بررسی انواع شبکه‌های تخاصمی در بازتشخیص افراد، یک معماری شبکه جدید با الهام از شبکه ATNet ارائه شده است. این شبکه جدید به طور هم زمان از مدل Pix2Pix و CycleGAN برای تولید تصاویر بهره می‌برد. بهبود نتایج ناشی از این داده افزایی در بازتشخیص افراد با استفاده از مجموعه داده‌های Market-1501 و CUHK03 و پارامترهای ارزیابی نشان داده شده است.

ساختار کلی این مقاله به این صورت است: ابتدا مروری جامع بر روش‌های موجود، مجموعه داده‌های مرجع و معیارهای ارزیابی صورت خواهد گرفت. سپس روش پیشنهادی به همراه معماری شبکه‌های پیشنهادی به تفصیل ارائه می‌شود. در بخش نتایج، پیاده‌سازی روش مورد نظر و مقایسه آن با روش‌های موجود بیان شده‌اند. تحلیل نتایج موضوع بخش بعدی می‌باشد. در پایان، جمع‌بندی و نتیجه‌گیری نهایی ارائه خواهد شد.

۲- مرور روش‌های موجود

این بخش حاوی مرورالگوریتم‌های موجود در زمینه بازتشخیص افراد است. با توجه به نتایج به دست آمده از روش‌های مبتنی‌بر GAN در زمینه یاد شده، این روش‌ها با جزئیات بیشتری مورد بررسی قرار خواهند گرفت.

در سال‌های اخیر با معرفی ترجمه تصویر به تصویر که هدف آن ترجمه تصویر از حوزه منبع به حوزه هدف است، کاربرد GAN ها در تولید تصویر گسترش یافته است. در سال ۲۰۱۶ ایزولا و همکاران یک شبکه تخاصمی شرطی ارائه کردند که قابلیت یادگیری نگاشت تصاویر منبع به تصاویر مورد نظر را داشت. محدودیت اصلی این مدل نیاز آن به نمونه‌های مزدوج است[۵]. برای از بین بردن این محدودیت، لیو و همکاران مدل CoGAN را با هدف یادگیری توزیع مجموع تصاویر دو حوزه بدون نیاز به

مزدوج کردن تصاویر پیشنهاد دادند[۶]. ژو و همکاران مدل CycleGAN را توسعه دادند که مفهوم تابع هزینه مربوط به ثبات چرخه[1] را ارائه کرد تا بتواند دو تابع نگاشت میان حوزه منبع و حوزه هدف را یاد بگیرد. همچنین در این مدل از نگاشت هویت جهت اطمینان از انتقال رنگ مربوط به هویت استفاده شده است[۷]. تابع هزینه مربوط به هویت به صورت رابطه۱ نوشته می‌شود:

(۱)

$$L_{identity} = E_{y\sim P_{data}(y)}[\| G(y)-y \|_1] + E_{x\sim P_{data}(x)}[\| F(x)-x \|_1]$$

که در آن x و y تصاویر ورودی از مجموعه داده X و Y هستند، F تابع نگاشت است که تصاویر را از X به Y تبدیل می‌کند و تابع نگاشت G مخالف آن را انجام می‌دهد. بدون وجود تابع هزینه، بخش تولیدکننده می‌تواند آزادانه رنگ تصاویر ورودی را تغییر دهد زیرا تابع هزینه مربوط به ثبات چرخه فقط رفتار کل نگاشت مزدوج شده را محدود می‌نماید.

در مقاله [۸] نویسندگان از ترکیب رویکردهای یادگیری مبتنی بر ویژگی‌های تصویری و شبکه عصبی اسپایکینگ (SNN) برای استخراج اطلاعات زمانی-مکانی استفاده کرده اند. نتایج نشان می‌دهند که ترکیب ویژگی‌های تصویری به بهبود تشخیص و دقت مدل در سناریوهای مختلف کمک می‌کند. این روش به خصوص برای سیستم‌های نظارت چنددوربینه با شرایط نوری و زاویه‌های مختلف مناسب می‌باشد. روش مبتنی‌بر رویکرد مولد برای باز تشخیص افراد در الگوریتم پیشنهادی [۹] ارائه شده است. در این الگوریتم از مدل‌های جدید تولیدی مانند مدل‌های انتشار (Diffusion Models)بهره گرفته شده است تا تصاویر افراد را به گونه‌ای بازسازی کند که اطلاعات هویتی را از سایر ویژگی‌ها جدا کند. با این روش، امکان ایجاد تغییرات و بررسی تصاویر افراد در شرایط مختلف بدون از دست دادن دقت هویتی فراهم می‌شود. هرچند عملکرد کنونی مدل هنوز به سطح پیشرفته‌ترین روش‌ها نرسیده، اما دارای ظرفیت توسعه در زمینه‌های مختلفی است.

نویسندگان مقاله [۱۰] از مدل‌های توجه و خود-تنظیمی(Self-Supervision)استفاده شده است که برای استخراج ویژگی‌های دقیق از تصاویر افراد مناسب است. این مدل‌ها توانسته‌اند با استفاده از شبکه‌های عصبی پیچشی و مکانیسم‌های توجه، عملکرد قابل قبولی در محیط‌های چالشی و پویای نظارتی داشته باشند. این روش با استفاده از تنظیم و تقویت ویژگی‌های هویتی، کارآیی شناسایی افراد را در تصاویر دوربین‌های مختلف و زوایای گوناگون بهبود بخشیده است. آزمایشات نشان می‌دهند که استفاده از GAN در پردازش داده‌ها به بهبود دقت و کاهش خطا در به باز تشخیص افراد کمک کرده است.

در سال‌های اخیر نسخه‌هایی از شبکه GAN برای بازتشخیص افراد معرفی شده که توانسته موفقیت قابل توجهی به دست آورد. شبکه GAN می‌تواند همزمان به عنوان یک رویکرد تطبیق حوزه‌ها[2] و یک رویکرد داده افزایی[3] عمل کند. مطالعات موجود مبتنی بر GAN در بازتشخیص افراد اغلب برای حل مسئله تطبیق میان دوربین‌ها یا تطبیق بین حوزه‌ها استفاده می‌شود [۱۱] تا [۱۴]. در ادامه به تشریح روشهای ارائه شده برای هر یک از این دو مورد خواهیم پرداخت.

۲-۱- رویکرد تطبیق میان دوربین‌های مختلف مبتنی بر شبکه‌های GAN

تنوع سبک[4]، یک چالش اساسی برای شناسایی مجدد افراد است، چرا که تصاویر افراد گرفته شده توسط دوربین‌های مختلف به دلیل تفاوت در تنظیمات مختلف دوربین و مکان‌های نظارتی متفاوت، باعث ایجاد سبک‌های مختلفی می‌گردد. شبکه GAN برای کاهش شکاف حوزه بین تصاویر گرفته شده توسط دوربین‌های مختلف و همچنین هموارسازی نابرابری‌ها در تصاویر یک دوربین مؤثر است[۱۱، ۱۴ و ۱۵]. ژونگ و همکاران از مدل CamStyle برای تولید تصاویر آموزشی جهت ورود به مدل CycleGAN استفاده نموده اند، زیرا مجموعه داده مربوطه حاوی تصاویر دوربین‌های مختلف بوده و این موضوع موجب تفاوت در سبک تصاویر می‌شود. با فرض وجود N دوربین، جهت کاهش فاصله حوزه میان تصاویری که گرفته شد، تعداد C_N^2 (ترکیب ۲ از N) مدل CycleGAN برای انتقال تصاویر هر دوربین به سبک دوربین‌های دیگر آموزش داده می شود. تصاویری که انتقال سبک شده بودند به عنوان داده پشتیبان به مجموعه داده آموزشی اضافه شده اند.

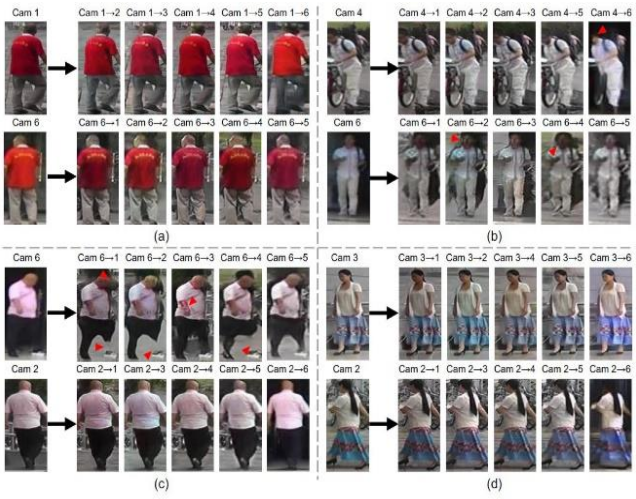
شش دوربین در مجموعه داده Market-1501 وجود دارد. همان‌طور که در شکل(۳) نشان داده شده است، مدل CamStyle هر نمونه آموزشی را به سبک پنج دوربین دیگر منتقل می‌کند و در نتیجه پنج تصویر آموزشی اضافی تولید می‌شود. تصاویر تولید شده از نظر سبک شبیه به تصاویر حوزه هدف هستند. با این حال، به دلیل عدم دقت تابع نگاشت مربوط به CycleGAN، مقداری خطا در تصاویر منتقل شده وجود دارد که موجب ورود نویز به سیستم می‌شود.

آنتروپی[5] بر روی تصاویر واقعی و تابع تنظیم کننده برچسب[6] یا LSR برای تصاویر منتقل شده اعمال می‌شود. تابع LSR به صورت رابطه ۲ محاسبه می‌گردد.

$$L_{LSR} = -(1-\varepsilon)\log p(y) - \frac{\varepsilon}{c}\sum_{c=1}^{c}\log p(c) \qquad (۲)$$

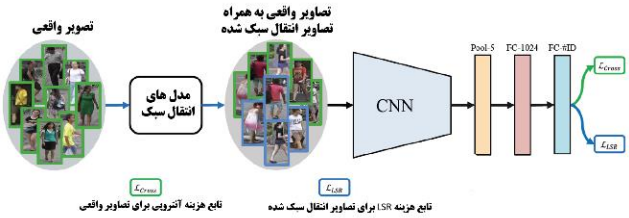
که در آن C تعداد کلاس‌ها است، $p(c)$ احتمال تعلق تصویر ورودی به برچسب c، و $p(y)$ احتمال تعلق هویت y به برچسب y است.

[1] Cycle Consistency Loss
[2] Domain Adaptation
[3] Data Augmentation
[4] Style Variation
[5] Cross-Entropy Loss
[6] Label Smooth Regularization



شکل(۳). نمونه‌هایی از تصاویر منتقل شده در Market-1501 [۱۱]

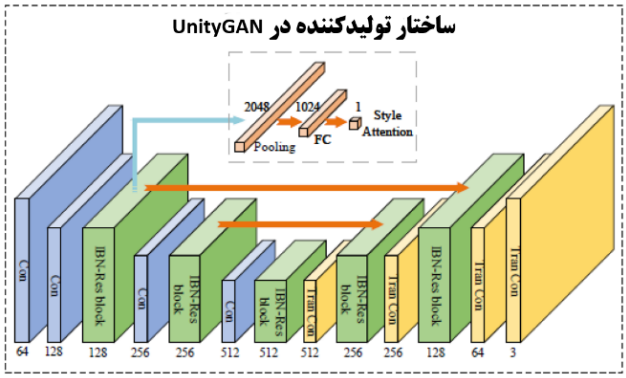
مدل CamStyle با معرفی تصاویر منتقل شده به مجموعه آموزشی، نه تنها بایاس دامنه بین تصاویر گرفته شده توسط دوربین‌های مختلف را محدود می‌کند، بلکه مجموعه داده‌های آموزشی را نیز گسترش می‌دهد تا نمونه‌های آموزشی بیشتری برای بازتشخیص فرد ارائه دهد. مدل CamStyle در شکل(۴) نشان داده شده است.



شکل(۴). معماری مدل CamStyle [۱۱]

مدل CamStyle بایاس دامنه را در بین دوربین‌های مختلف کاهش می‌دهد، اما باید این را در نظر داشت که به دلیل تفاوت زمان ضبط و تفاوت فاصله عابران پیاده با عدسی، تفاوت سبک در تصاویر گرفته‌شده توسط یک دوربین نیز وجود دارد. لیو و همکاران [۷] با توجه به این مشکل، مدل UnityStyle را جهت کاهش بایاس دامنه بین تصاویر مختلف یک دوربین (تعریف سبک) پیشنهاد دادند.

مدل UnityStyle برای کاهش فاصله حوزه‌های تصاویر گرفته شده توسط یک دوربین و همچنین دوربین‌های مختلف، از UnityGAN برای تولید تصاویر استفاده می‌نماید. برخلاف CycleGAN که هدف آن یادگیری توابع نگاشت بین هر دو سبک است، هدف UnityGAN ایجاد یک سبک یکنواخت برای همه تصاویر است. UnityGAN از DiscoGAN و CycleGAN استفاده می‌کند، و شبکه حاصل از ترکیب این دو مدل را با معرفی بلوک‌های باقیمانده[1] و در نظر نگرفتن اتصالات بهبود می‌دهد. علاوه بر این، مدل IBN-Res جهت تقویت نتایج مربوط به تغییر سبک اتخاذ می شود. معماری بخش تولیدکننده UnityGAN در شکل(۵) نشان داده شده است.



شکل(۵). معماری مربوط به بخش تولید کننده UnityGAN [۷]

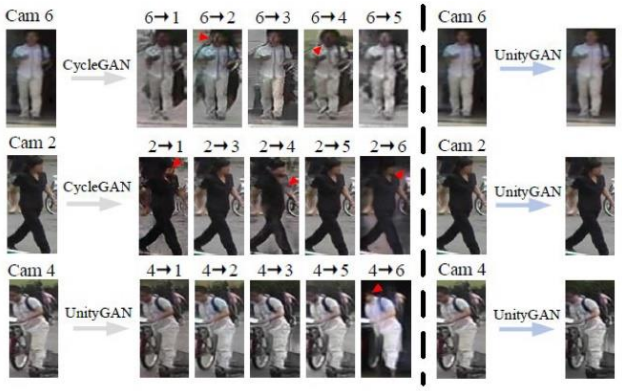
تابع هزینه مربوط به هویت در UnityStyle به صورت رابطه ۳ می‌باشد.

$$L_{ID} = E_{x\sim l_x}(\| F(x)-x \|_1) + E_{y\sim l_y}(\| G(y)-y \|_1) \quad (۳)$$

که در آن F تابع نگاشت برداری است که برای انتقال تصاویر از حوزه X به Y استفاده می‌شود و تابع نگاشت G برعکس این عمل را انجام می‌دهد. علاوه بر این، برای تضمین اینکه UnityGAN می‌تواند تصاویر UnityStyle تولید کند، ماژول Style Attention به بخش تولیدکننده UnityGAN اضافه شده است که به صورت رابطه زیر تعریف می‌گردد:

$$A(x) = Sigmoid(A_{style}(G_1(x))) \quad (۴)$$

در این رابطه G_1 اولین بلوک خروجی IBN-Res است[۱۲]. همان‌طور که در شکل(۶) نشان داده شده است، چندین خطا در تصاویر تولید شده توسط CycleGAN وجود دارد. اما در مقایسه با روش قبل، UnityGAN می‌تواند تصاویری با ساختار پایدارتر تولید کند.



شکل(۶). مقایسه تصاویر تولید شده توسط CycleGAN و UnityGAN [۷]

برای آموزش جهت بازتشخیص فرد، تصاویر واقعی و تصاویر UnityStyle به عنوان یک مجموعه آموزشی برای آموزش تحت نظارت ترکیب می‌شوند. با توجه به کیفیت بالای تصاویر تولید شده، LSR اتخاذ شده در CamStyle دیگر در UnityStyle مورد نیاز نیست، تصاویر تولید شده را می‌توان مانند تصاویر اصلی در نظر گرفت. در بخش آزمایش، به طور مستقیم تصاویر UnityStyle به جای تصاویر اصلی استفاده می‌شود تا اطمینان حاصل شود که

[1] Residual Blocks

تصاویر پرس و جو[1] و تصاویر گالری مربوط به یک نفر هستند (دو مجموعه تصاویر پرس و جو و گالری دو مجموعه با اشتراک تهی هستند و از آن‌ها جهت آزمایش مدل استفاده می‌شود. بر این اساس مدل باید هویت مربوط به تصویر پرس و جو شده را با ارائه تصویری از همان فرد از میان تصاویر گالری تعیین نماید).

۲-۲- رویکردهای انطباق بین حوزه‌ای براساس شبکه GAN (تطبیق میان مجموعه داده‌های مختلف)

اگرچه رویکردهای موجود در حین آموزش و آزمایش مدل‌ها بر روی یک مجموعه داده به موفقیت قابل‌توجهی دست یافته‌اند، اما عملکرد آن‌ها در حین آموزش و آزمایش بر روی مجموعه داده‌های مختلف به طور قابل‌توجهی کاهش می‌یابد. همانطور که در شکل(۷) نشان داده شده است، تصاویر افراد از مجموعه داده‌های مختلف سبک‌های متفاوتی را ارائه می‌دهند، سبک‌های مختلف را می‌توان به عنوان حوزه‌های مختلف مشاهده کرد و بازتشخیص فرد به راحتی تحت تأثیر وجود بایاس دامنه قرار می‌گیرد. درکاربردهای واقعی، به دلیل هزینه گران برای نشانه گذاری نمونه‌ها، به طور معمول جمع آوری داده‌های آموزشی کافی غیرعملی است. برای حل مشکل کمبود نمونه‌های آموزشی، برخی از مطالعات از GAN برای انتقال داده‌های نشانه گذاری شده موجود به سبک مجموعه داده‌های هدف استفاده می‌کنند. تصاویر منتقل شده به عنوان نمونه‌های آموزشی تکمیلی استفاده می‌شوند.

شکل(۷). سبک تصاویر مربوط به افراد در مجموعه داده‌های مختلف

وی و همکاران یک رویکرد داده افزایی به نام GAN Transfer Person (PTGAN) را برای انتقال تصاویر نشانه گذاری شده از مجموعه داده A به مجموعه داده B و به سبک آن پیشنهاد کرده‌اند[۱۲].

از آنجایی که هیچ نمونه تصویر مزدوج در مجموعه داده‌ها وجود ندارد، انتقال بین مجموعه داده‌ها را می‌توان ترجمه تصویر به تصویر غیر مزدوجی در نظر گرفت. با توجه به اثربخشی CycleGAN در کار ترجمه بدون مزدوج تصویر به تصویر، PTGAN از CycleGAN به عنوان معماری شبکه استفاده می‌کند. علاوه بر این، یک محدودیت اضافی برای حفظ ثبات رنگ در طول انتقال اتخاذ می‌شود که به صورت رابطه ۵ نوشته می‌شود.

که در آن a~pdata (a) توزیع داده مجموعه داده A است و b~pdata (b) توزیع داده مجموعه داده B می‌باشد. G نشان دهنده تابع تبدیل از مجموعه داده A به مجموعه داده B بوده و $\bar{G}$ نشان دهنده تابع تبدیل از مجموعه داده B به A است.

$$L_{ID} = E_{a\sim Pdata(a)}[\| (G(a)-a)\odot M(a)\|_2] + E_{b\sim Pdata(b)}[\| (\bar{G}(b)-b)\odot M(b)\|_2] \quad (۵)$$

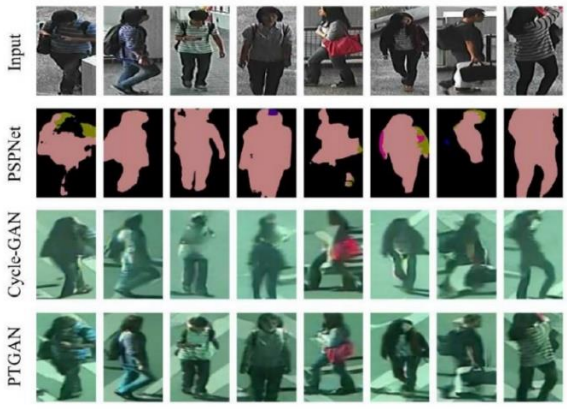
M(a) و M(b) نشان دهنده انسداد تصاویر ورودی a و b هستند. توجه داشته باشید که محاسبه L_{ID} با تابع هزینه هویت که در بخش‌های قبل ذکر شد، متفاوت است. در تابعL_{ID} با استفاده از تقسیم بندی بدن انسان، میزان انسداد به دست می‌آید تا GAN رنگ را در نواحی مختلف حفظ نماید، در حالی که حفظ رنگ پشت زمینه اهمیتی ندارد. تابع هزینه کلی PTGAN به صورت رابطه ۶ می‌باشد.

$$L = L_{GAN}(G, D_B, A, B) + L_{GAN}(\bar{G}, D_A, B, A) + \lambda_1 L_{cyc} + \lambda_2 L_{1D} \quad (۶)$$

که در آن D_A و D_B بخش متمایز کننده مدل‌های مربوط به مجموعه داده A و B است.

همان‌طور که در شکل(۸) نشان داده شده است، کیفیت تصاویر منتقل شده توسط PTGAN به طور قابل توجهی بالاتر از CycleGAN است.

در حالی که PTGAN وظیفه انتقال بین حوزه‌ای را یک جا انجام می‌دهد، لیو و همکاران [۱۳] بر این نکته تأکید کردند که سه عامل اصلی شامل روشنایی، وضوح و زاویه دید بر کیفیت کلی انتقال تأثیر می‌گذارد. ایشان پیشنهاد دادند که وظیفه انتقال میان حوزه‌ها به سه وظیفه فرعی تقسیم شوند، و در نتیجه برای هر وظیفه فرعی یک شبکه GAN به طور جداگانه آموزش می‌بینند. سپس، برای بازتشخیص فرد یک راهبرد جهت ترکیب ویژگی‌های به دست آمده از سه شبکه اتخاذ می شود.



شکل(۸). تصاویر منتقل شده از CHUHK03 به PRID-CAM1

همان‌طور که در شکل(۹) مشخص شده است، این شبکه با نام ATNet[2] از چهار زیرشبکه شامل سه شبکه GAN برای روشنایی،

[1] Query

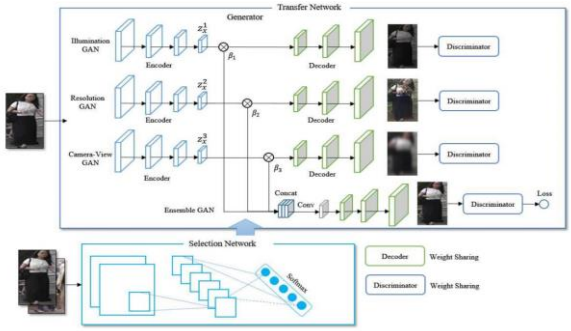
[2] Adaptive transfer Network

وضوح و زاویه دید، یک شبکه گزینش و یک شبکه برای ترکیب نتایج تشکیل شده است. این شبکه برای اولین بار در مقاله [۱۳] ارائه شده است که یک شبکه مولد برای تبدیل تصویر به تصویر (Image-to-Image Translation) است و به‌طور خاص از *Transformers* برای پردازش ویژگی‌های تصویری استفاده می‌کند. برخلاف شبکه‌های کانولوشنی معمولی (CNN) که در بسیاری از مدل‌های تبدیل تصویر به تصویر استفاده می‌شوند، ATNet از قدرت *self-attention* و *cross-attention* در Transformers بهره می‌برد تا توانایی بهتری برای یادگیری ویژگی‌های پیچیده و غیرخطی از داده‌های تصویری مختلف داشته باشد.

ATNetبه‌ویژه در تبدیل تصاویر از یک دامنه به دامنه دیگر، مانند تبدیل تصاویر سبک‌های هنری به عکس‌های واقعی یا ایجاد تغییرات پیچیده در تصاویر، به خوبی عمل می‌کند. این معماری از یک شبکه مولد و یک شبکه تمایزدهنده استفاده می‌کند که قادر به یادگیری ویژگی‌های پیچیده از تصاویر در مقیاس‌های مختلف است. یکی از ویژگی‌های برجسته ATNet این است که مدل به‌طور پویا و سازگار با داده‌ها عمل می‌کند و می‌تواند بر اساس نیازهای خاص هر وظیفه، به‌صورت خودکار توجه را به بخش‌های مختلف تصویر معطوف کند. این امر موجب افزایش دقت و کارایی مدل در مسائل پیچیده‌تری همچون تبدیل‌های غیررسمی تصویر می‌شود.

سه شبکه انتقال و شبکه مربوط به یادگیری گروهی[1] مبتنی بر CycleGAN هستند و تابع هزینه با استفاده از رابطه ۷ محاسبه می‌گردد[۱۶ و ۱۷].

$$L_{gan} = L_{adv} + \lambda_1 L_{cyc} + \lambda_2 L_{idc} \qquad (۷)$$



شکل(۹). معماری شبکه ATNet معرفی شده در مرجع [۱۳]

۳- مجموعه داده‌های مرجع و معیارهای ارزیابی

در این مقاله از مجموعه داده‌های Market-1501 و CUHK03 برای آموزش شبکه پیشنهادی و ارزیابی عملکرد آن در بازتشخیص افراد استفاده شده است. مجموعه داده Market-1501 یک مجموعه داده معیار برای بازتشخیص افراد است. این مجموعه شامل ۱۵۰۱ فرد است که تصاویر آنها توسط شش دوربین مختلف اخذ شده است. هر فرد به طور متوسط ۳/۶ تصویر دارد. پر واضح هست که هر فرد در هر نما تصویر ندارد[۱۸].

مجموعه داده CUHK03 نیز متشکل از ۱۴۰۹۷ تصویر از ۱۴۶۷ فرد مختلف است که در آن نیز از شش دوربین برای جمع آوری تصاویر استفاده شده است و هر فرد توسط ۲ دوربین در محوطه دانشگاه ثبت شده است[۱۸]. تصاویر نمونه از افراد مختلف در این مجموعه داده‌ها در شکل(۱۰) نشان داده شده است.

الف) ب)

شکل(۱۰). تصاویر نمونه از مجموعه داده‌های معیار
الف) CUHK03 و ب) Market-1501

دو معیار رایج جهت ارزیابی الگوریتم‌های بازتشخیص افراد «تطبیق تجمیعی»[2] و «میانگین دقت متوسط»[3] است. روابط ۸ و ۹ به ترتیب نشان‌دهنده نحوه محاسبه این معیارها می‌باشند.

$$CMC(K) = \frac{1}{N}\sum_{i=1}^{N} 1.(rank(i) \le k) \qquad (۸)$$

$$mAP = \frac{1}{Q}\sum_{i=1}^{Q} AP(i), AP = \frac{1}{m}\sum_{k=1}^{n} p(k).1 \qquad (۹)$$

در رابطه ۸ N تعداد افراد در مجموعه گالری هستند و n رتبه اولین تطبیق صحیح برای نمونه i ام می‌باشد. در واقع معیار تطبیق تجمیعی درصد تطبیق در رتبه‌های بالای لیست می‌باشد. در رابطه ۹ نیز mAP میانگین دقت‌های محاسبه شده برای هر پرس و جو بوده و در آن n تعداد کل نتایج ارزیابی، m تعداد تطبیق‌های صحیح و $p(k)$ دقت در رتبه k می‌باشد.

معیار تطبیق تجمیعی مرتبه[4] n نشان دهنده این احتمال است که در n تشخیص اولیه، هویت پیش بینی شده درست وجود داشته باشد. این روش زمانی دقیق است که برای هر شخص مورد نظر تنها یک جواب درست موجود باشد. هر چند که در اکثر شبکه‌های بزرگ از دوربین‌ها معمولاً این فرض درست نیست و روش تطبیق تجمیعی نمی‌تواند به طور کامل نشان دهنده تمایزپذیری یک مدل در بین چندین دوربین باشد.

روش میانگین دقت متوسط در واقع میانگینی از عملکرد بازیابی تصاویر در هنگام وجود چند گزینه درست است. این روش برای مقایسه دو مدل که در بازیابی گزینه درستِ اول، یکسان هستند، اما در بازیابی گزینه دوم تفاوت دارند، کارآیی مناسبی دارد.

[1] Ensemble Learning

[2] Cumulative Matching Characteristics (CMC)
[3] Mean Average Precision (mAP)
[4] Rank

۴- روش پیشنهادی

در داده افزایی جهت بازتشخیص افراد معمولاً از روش‌هایی مانند تغییر تصادفی مبتنی‌بر اندازه، بریدن تصویر و چرخش افقی و عمودی و ... استفاده می‌شود. همچنین جهت ایجاد تنوع در مجموعه داده آموزشی از نمونه‌های ماسک شده نیز بهره برده می‌شود. همچنین برای اضافه نمودن نویز به تصاویر ورودی، راهبرد حذف تصادفی نیز کاربرد دارد. این روش‌ها موجب بهبود الگوریتم نظارتی و تعمیم بهتر آن بر روی داده‌های تست می‌گردد.

در این مقاله از الگوریتم‌های پیشنهادی بر مبنای شبکه‌های GAN به منظور داده افزایی و بهبود نتایج بازتشخیص استفاده شده و جهت صحت سنجی، آزمایش‌ها برروی دو مجموعه داده معیار در بازتشخیص افراد صورت گرفته است.

فرض بر این است که چهار تصویر از کل بدن یک فرد، شامل چهار نما از روبه رو (F)، پشت سر (B) و طرفین (R, L)، بیشترین اطلاعات را از فرد ارائه می‌کند و اگر تصاویر هر چهار نما موجود باشد، بازتشخیص یک فرد با دقت بالایی انجام می‌شود.

در بسیاری از مواقع، همه این چهار نما از یک فرد در دسترس نیستند. در این پروژه با الهام از روش یادگیری گروهی مورد استفاده در ساختار ATNet و استفاده هم زمان از دو روش Pix2Pix و CycleGAN جهت تولید تصاویر غیر موجود، مجموعه کامل این چهار تصویر برای هر فرد تولید شده و سپس با افزودن آن‌ها به داده آموزشی، دقت شبکه عصبی افزایش می‌یابد.

مدل Pix2Pix معرفی شده در در مقاله [۱۹]، که از شبکه‌های عصبی کانولوشنی (CNN) و شبکه‌های مولد رقابتی (GAN) برای تبدیل تصویر به تصویر استفاده می‌کند .Pix2Pix از یک معماری GANشرطی استفاده می‌کند که در آن ورودی به‌عنوان یک تصویر (مثلاً عکس یک خانه) به شبکه داده می‌شود و هدف آن تولید تصویری مشابه از تصاویر هدف (مثلاً تصویر خانه‌ای که در شب است) است. در این مدل، از یک شبکه تولیدکننده و یک شبکه تمایزدهنده برای آموزش مشترک استفاده می‌شود.

مدل CycleGAN نیز در مقاله [۲۰] معرفی شده است که یکی از مدل‌های محبوب در زمینه تبدیل تصاویر بدون نیاز به داده‌های جفت‌شده است. برخلاف Pix2Pix که به‌داده‌های جفت‌شده نیاز دارد (یعنی هر تصویر ورودی باید یک تصویر هدف داشته باشد)، CycleGANبرای آموزش به تصاویر بدون جفت‌شده نیز قادر است. این مدل از دو شبکه مولد و تمایزدهنده و یک قیود چرخه‌ای (cycle consistency)استفاده می‌کند که موجب می‌شود مدل‌ها قادر به نگه‌داشتن ویژگی‌های اصلی تصاویر حتی در زمان عدم حضور تصویر جفت گردند.

پس از بررسی داده‌های مربوط به هر فرد در مجموعه داده‌های مورد استفاده شامل Market-1501 و CUHK03، ابتدا تصاویری که با وضوح بیشتر تصویر فرد را از رو به رو، پشت سر، طرف چپ و طرف راست نشان می‌دهد انتخاب می‌شوند (در برخی تصاویر هر چهار نما در دسترس نیستند و بنابراین به عنوان داده آموزشی مورد استفاده قرار نگرفته‌اند. بدیهی است بعد از مرحله آموزش، نماهای ناموجود توسط شبکه پیشنهادی تولید گردیدند). در شکل (۱۱) مثال‌هایی جهت تبیین روش انتخاب تصویر برای هر فرد جهت تولید مجموعه داده آموزشی شخصی سازی شده نشان داده شده است.

شکل(۱۱). نمونه‌هایی از انتخاب تصاویر یک فرد در چهار نما

۴-۱- معماری شبکه‌های پیشنهادی

تفاوت اصلی معماری پیشنهادی با ATNet در استفاده هم زمان از شبکه Pix2Pix و CycleGAN برای تولید تصاویر است. در معماری پیشنهادی مانند ATNet از یادگیری گروهی استفاده شده است.

در این معماری، نمایی که وجود ندارد به عنوان خروجی در نظر گرفته می‌شود. همان‌طور که گفته شد، هدف تولید این نما با استفاده از تصاویر موجود نماهای دیگر می‌باشد. با توجه به پیشنهاد ساختار Pix2Pix برای تولید نمای چهارم، به سه شبکه با ورودی ترکیبی از دو نما از سه نمای در دسترس، نیاز خواهد بود. به عنوان مثال در شکل(۱۲) هدف تولید نمای پشت سر است. در واقع این شکل نشان دهنده معماری شبکه پیشنهادی برای داده افزایی به منظور بازتشخیص افراد می‌باشد. همان‌طور که گفته شد، تمامی افراد دارای تصویر از تمامی نماهای مختلف نیستند و خروجی برخی از زیر شبکه‌های Pix2Pix موجود در این شکل قابل تولید نمی باشد.

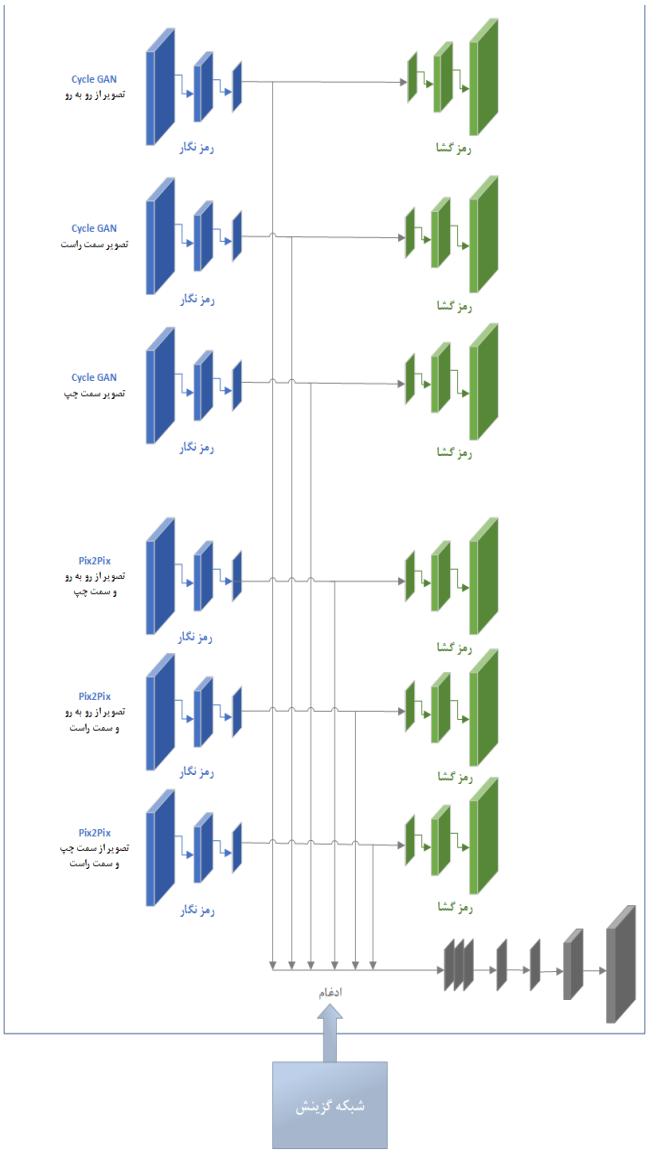
در این صورت هر نمای قابل تولید در مرحله ادغام مورد استفاده قرار می‌گیرد. پر واضح است که اضافه شدن نمای جدید منجر به بالا رفتن دقت بازتشخیص خواهد شد. در جدول۱ تعداد تصاویر مختلف با نماهای موجود در مجموعه داده‌های معیار نشان داده شده است.

جدول۱. تعداد تصاویر با نماهای موجود در مجموعه داده‌های معیار Market-1501 و CUHK03

مجموعه داده	دارای چهار نما	دارای سه نما	دارای دو نما	دارای یک نما
Market-1501	۱۴۳	۴۵۳	۱۲۶۴	۱۳۹۸
CUHK03	۰	۰	۱۴۶۷	۰

در این الگوریتم از سه شبکه CycleGAN مجزا با ورودی‌هایی شامل سه نما در دسترس برای تولید نمای ناموجود بهره گرفته شده است. بنابراین معماری شبکه الگوریتم پیشنهادی شامل شش شبکه مولد می‌باشد.

همان‌گونه که در این جدول مشاهده می‌گردد، به دلیل ساختار مجموعه داده CUHK03 و گزارش هر فرد توسط دو دوربین، در عمل فقط دو نما از فرد وجود دارد. این دو نما به صورت مزدوج چپ و پشت سر و مزدوج راست و روبرو می‌باشند.



شکل(۱۲). معماری شبکه پیشنهادی برای داده افزایی

به منظور بررسی نتایج حاصل از این شش شبکه و محاسبه وزن هر کدام در پیش بینی بهتر نمای چهارم از یک شبکه گزینش استفاده می‌شود. در شبکه گزینش ابتدا شبکه ResNet به منظور استخراج ویژگی تصاویر تولید شده به کار برده می‌شود. در ادامه از معیار فاصله اقلیدسی برای شباهت سنجی بردارهای ویژگی استفاده می‌گردد. دلیل این کار این است که تولید تصاویر به صورت شرطی تولید می‌گردند. سپس فواصل به دست آمده با بازه [۰ ,۱] نرمالیزه می‌شود. پس از آموزش شبکه گزینش، از وزن‌های مؤثر هر یک از شش شبکه بر اساس سهم آن‌ها در باز تولید نمای ناموجود میانگین‌گیری می‌گردد. محاسبه میانگین وزن شامل میانگین گرفتن از وزن‌های به دست آمده از شبکه گزینش برای هر شبکه و محاسبه آن برروی تمام نمونه‌های مجموعه آموزشی است. این فرآیند میانگین‌گیری امکان استفاده از هر شش شبکه را برای تولید نمای ناموجود فراهم می‌نماید.

می‌توان گفت، با میانگین‌گیری وزن‌ها، مجموعه‌ای به دست می‌آید که نشان دهنده اهمیت کلی هر شبکه و هر یک از وزن‌ها است. سپس آخرین شبکه GAN، نتیجه حاصل از این شش شبکه را با هم ادغام[1] می‌نماید.

لازم به ذکر است، مجموعه داده و کدهای مربوط به معماری پیشنهادی در آدرس اینترنتی[2] قابل مشاهده است. در ادامه مقاله توضیحات بخش‌های مختلف شبکه با جزئیات بیشتری مورد بحث قرار گرفته است.

۴-۱-۱- شبکه Pix2Pix به کار رفته

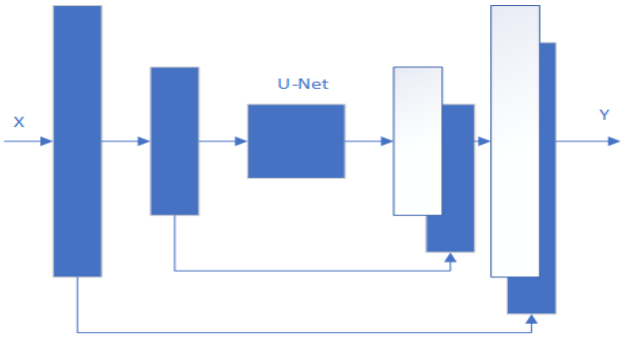
شبکه Pix2Pix یک نوع شرطی از معماری GAN است و ورودی آن شامل دو تصویری است که از دو حوزه مختلف در نظر گرفته شده است. در این شبکه، تولیدکننده و متمایزکننده براساس تصویر ورودی یا اطلاعات کمکی شرطی می‌شوند تا موجب تولید تصاویری طبق حوزه هدف شوند. در مدل پیشنهادی از الگوریتم PatchGAN [۱۹] به عنوان متمایزکننده و از U-Net به عنوان تولیدکننده استفاده شده است.

شبکه PatchGAN به‌عنوان یک معماری برای شبکه تمایزدهنده در GANها معرفی شد که به‌جای ارزیابی تصویر به‌طور کامل، آن را به بخش‌های کوچک (پچ‌ها) تقسیم کرده و در سطح بخش‌ها تمایز می‌دهد. این شبکه در Pix2Pix به‌طور خاص به‌عنوان تمایزدهنده استفاده می‌شود و از آنجا که می‌تواند ویژگی‌های کوچک‌تر تصاویر را شناسایی کند، برای انجام تبدیل‌های پیچیده‌تر تصویر به تصویر بسیار مؤثر است.

U-Net شامل یک شبکه کدگذار[3] و یک شبکه کدگشا[4] به طور سلسله مراتبی است. شکل(۱۳) نشان‌دهنده مدل U-Net می‌باشد.

---

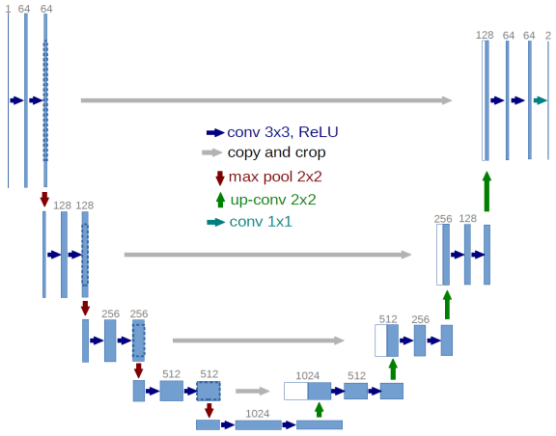
[1] Concatenation
[2] https://github.com/hdkvh/Data_Augmentation4Person_Reidentification
[3] Encoder
[4] Decoder



شکل(۱۳). معماری شبکه U-Net

- نیمه اول معماری شبکه کدگذار است که از نوع شبکه طبقه بند ResNet است.
- نیمه دوم معماری شبکه کدگشا است. هدف، نگاشت مشخصه‌های متمایزکننده (که توسط کدگذار یادگیری شده است) به فضای پیکسل است (رزولوشن بالاتر) تا مقادیر لایه کامل متصل به دست آید. شبکه کدگشا شامل نمونه برداری افزایشی[1]، ادغام و در نهایت عملیات کانولوشن می‌باشد.

معماری زیرشبکه‌های کدگذار و کدگشا در شکل(۱۴) نشان داده شده‌اند.



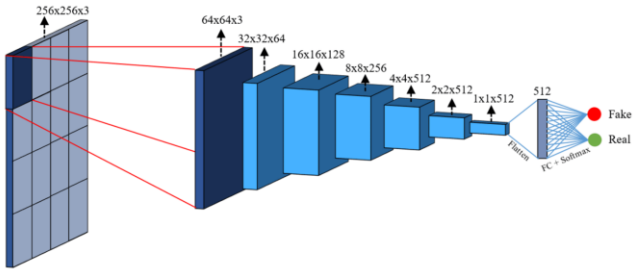
شکل(۱۴). معماری زیر شبکه‌های کدگذار و کدگشا در شبکه U-Net پیشنهادی

آخرین لایه شبکه کدگذار که به لایه گلوگاه[2] معروف است، لایه نرمال‌سازی دسته‌ای است که تابع فعال‌ساز آن ReLU می‌باشد. در نهایت، خروجی مدل تولید کننده یک لایه کانولوشنی با تابع فعال‌ساز tanh است.

تابع هزینه شامل هر دو تابع آنتروپی و میانگین خطای مطلق می‌باشد. جهت دریافت کوچکترین اطلاعات میان ورودی و خروجی، بین هر لایه i و لایه n-i اتصالات پرش[3] اضافه می‌شود. هر اتصال پرش همه کانال‌های موجود در لایه i را به کانال‌های لایه n-i متصل می‌کند.

از شبکه PatchGAN به عنوان یک مدل متمایزکننده یا تشخیص دهنده جعلی بودن تصویر تولیدی مورد استفاده قرار گرفته است. بنابراین جزئیات با فرکانس بالاتر محدود و تعداد پارامترها کمتر و در نتیجه طبقه‌بندی سریعتر انجام می‌شود. این شبکه دو جفت تصویر شامل ورودی و هدف و ورودی و تصویر تولید شده را به عنوان ورودی می‌پذیرد. این تصاویر با هم ادغام می‌شوند.

پنجره مورد استفاده در این مدل ۶۴ در ۶۴ می‌باشد که در نتیجه خروجی مدل به یک تصویر ۶۴ در ۶۴ از تصاویر ورودی نسبت داده می‌شود. به بیان دیگر یک شبکه PatchGAN با اندازه ۶۴ در ۶۴ می‌تواند تکه‌های ۶۴ در ۶۴ از تصویر ورودی را به عنوان تصاویر واقعی یا جعلی طبقه بندی کند. پیکربندی شبکه PatchGANپیشنهادی به صورت شکل(۱۵) می‌باشد.

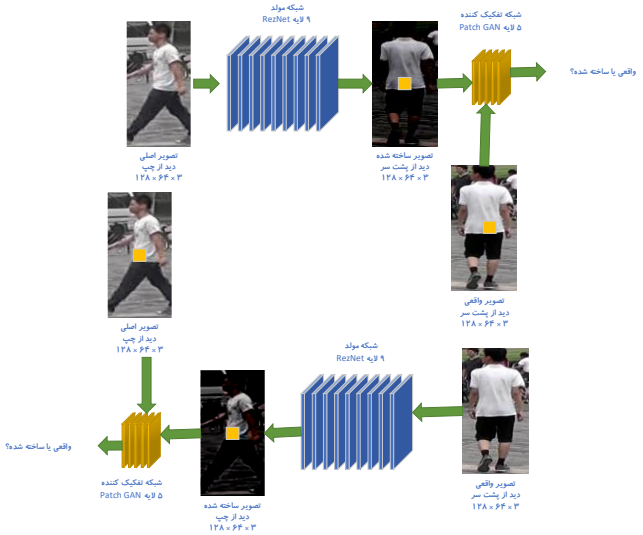


شکل(۱۵). پیکر بندی شبکه PatchGAN پیشنهادی

۴-۱-۲- شبکه CycleGAN

همان‌گونه که در بخش‌های قبلی نیز اشاره شد، شبکه CycleGAN ابزار قدرتمندی برای تولید داده‌های متنوع مصنوعی برای غنی سازی مجموعه داده‌ها می‌باشد.

این شبکه با استفاده از معماری شبکه‌های متخاصم نه تنها به افزایش داده‌های مجموعه داده کمک می‌نماید به طور کیفی نیز با حفظ ویژگی‌های کلیدی داده‌ها در بهبود عملکرد الگوریتم‌های اجرا شده بر روی مجموعه داده‌ها تأثیرگذار است. در شکل(۱۶) معماری شبکه CycleGAN پیشنهادی نشان داده شده است.



شکل(۱۶). معماری شبکه CycleGAN پیشنهادی

[1] Upsampling
[2] Bottleneck Layer
[3] Skip Connection

در جدول۲ به صورت کلی ساختار و معماری شبکه پیشنهادی ارائه شده است. این جدول شامل اطلاعات مربوط به معماری، تعداد لایه‌ها و مشصات هر یک، ابعاد و همچنین دلایل استفاده از هر معماری به همراه جزئیات استفاده شده می‌باشد.

جدول۲ . ساختار و معماری شبکه پیشنهادی

شبکه	معماری	تعداد لایه‌ها	مشخصات لایه‌ها	ابعاد ورودی	ابعاد خروجی	دلیل استفاده	سایر جزئیات
CycleGAN	مولد ResNet تفکیک کننده PatchGAN	۹ لایه مولد ۵ لایه تفکیک کننده	کانولوشن ReLU Tanh	۲۵۶×۱۲۸×۳	۲۵۶×۱۲۸×۳	یادگیری تبدیل بدون نظارت، امکان تولید نماهای ناموجود بدون جفت داده	استفاده از دو شبکه برای تبدیل بین نماها
Pix2Pix	مولد U-Net تفکیک کننده PatchGAN	۸ لایه مولد ۵ لایه تفکیک کننده	کانولوشن BatchNorm ReLU Tanh	۲۵۶×۱۲۸×۳	۲۵۶×۱۲۸×۳	یادگیری تحت نظارت، بهبود کیفیت تصاویر با اطلاعات اضافی، تولید دقیق‌تر تصاویر	یادگیری تحت نظارت برای تولید تصاویر نماهای ناموجود
ResNet	ResNet-50	۵۰ لایه	کانولوشن MacPool ResBlock(50) Global Pooling	۲۵۶×۱۲۸×۳	۲۰۴۸ ویژگی	استخراج ویژگی‌های عمیق، افزایش دقت بازشناسی، توانایی یادگیری ویژگی‌های مهم	استخراج ویژگی‌های عمیق از افراد
PatchGAN	کانولوشن	۷ لایه	کانولوشن LeekyReLU Sigmod	۶۴×۶۴×۳	۱ تصمیم واقعی از جعلی	تشخیص محلی واقعی/جعلی بودن تصویر، بهبود کیفیت خروجی با توجه به جزییات منطقه‌ای	استفاده در Pix2Pixو CycleGANبرای تفکیک واقعیت از جعل
U-Net	کدگذار کدگشا	۸ لایه	کدگذار: کانولوشن BatchNorm ReLU MaxPool کدگشا: کانولوشن BatchNorm ReLU Skip Conn	۲۵۶×۱۲۸×۳	۲۵۶×۱۲۸×۳	استفاده برای بازسازی دقیق‌تر، حفظ جزئیات تصویر با اتصال‌های میان‌بُر	استفاده در Pix2Pixبرای بازسازی تصاویر

۵- نتایج پیاده سازی و مقایسه با روش‌های موجود

در ادامه این بخش به بررسی نتایج به دست آمده پرداخته شده است.

۵-۱- بررسی تصاویر تولید شده

بعد از آموزش شبکه پیشنهادی نشان داده شده در شکل(۱۰)، نمای ناموجود افراد تولید شدند و به مجموعه داده‌های معیار اضافه گردیدند. شکل(۱۷) نشان دهنده تعدادی از نماهای تولید شده افراد مختلف توسط الگوریتم پیشنهادی می‌باشد.

۵-۱- مقایسه با الگوریتم‌های موجود

همانگونه که بخش‌های قبل گفته شد هدف از الگوریتم پیشنهادی، داده افزایی و تولید تصاویر نماهای ناموجود از افراد مختلف در مجموعه گالری می‌باشد. بعد از افزایش مجموعه داده، از الگوریتم‌های موجود برای بازتشخیص افراد بهره گرفته شده است.

در جدول۲ مقایسه عملکرد الگوریتم‌های موجود که راهکار آن‌ها مبتنی بر شبکه‌های GAN می‌باشند، بیان شده اند.

بعد از انجام آموزش دوباره، ارزیابی به همراه محاسبه معیار تطبیق جمعی و میانگین دقت متوسط صورت گرفته است.

نتایج ارزیابی تأثیر الگوریتم پیشنهادی در جدول فوق برای مدل‌های برتر بیان شده‌اند. این مدل‌های برتر شامل UnityStyle، FD-GAN و DG-Net است.

در بخش ۲-۱ به تفصیل در مورد UnityStyle و کارکرد آن در یکنواخت ساختن سبک همه تصاویر بحث شد.

در جدول۳ پارامترهای مربوط با روش‌های برتر به صورت تیره‌تر نشان داده شده است. روش کار به این صورت می‌باشد که از مجموعه داده جدید تولید شده، برای آموزش دوباره شبکه‌های الگوریتم‌های پیشنهادی این مقالات استفاده شده است. در نهایت

در FD-GAN با توجه به تعدد ژست‌های مختلف افراد در تصویر سعی شده که با شناسایی ویژگی‌های مربوط به هویت افراد و حذف ویژگی‌های مربوط به ژست بتوان تمایز میان افراد را در تصاویر مختلف بهتر انجام داد.

شکل(۱۷). نمونه‌هایی از نماهای تولید شده افراد مختلف موجود در داده‌های معیار با استفاده از الگوریتم پیشنهادی به منظور داده افزایی
الف و ب) تصاویر تولید شده در مرحله آموزش و مقایسه با تصویر هدف بعد از Epoch به ترتیب ۵۰ و ۱۰۰ (ردیف اول نمای ورودی، ردیف دوم نمای تولید شده و ردیف سوم نمای هدف برای تولید)
ج) تصاویر تولید شده در مرحله پیش بینی (ردیف اول نمای ورودی و ردیف دوم نمای تولید شده)

مدل DG-Net شامل یک بخش تولیدکننده است که تصویر هر فرد را براساس شکل ظاهری (رنگ لباس و کفاش، سبک و بافت تصویر) و ساختار تصویر (اندازه بدن، مو، زمینه، مکان، ژست، منظر) به طور جدا کدگذاری می کند. همچنین در این معماری بخش کدگذاز شکل ظاهری در دو مدل متمایزکننده و تولیدکننده مشترک است. با تغییر کد کننده‌های مربوط به شکل ظاهری و ساختار، تصاویر گوناگونی از هویت‌های موجود ساخته می‌شود. فیدبک این تصاویر ساخته شده به کدگذار شکل ظاهری وارد می‌گردد و موجب بهبود متمایزکننده می‌شود. این معماری بدون استفاده از داده افزایی حاصل از مدل تولیدکننده، نتایج را به طور چشمگیر بهبود داده است.

جدول۳. مقایسه روش‌های موجود در دو مجموعه داده معیار [۲۱] و ارزیابی تاثیر الگوریتم پیشنهادی با روش‌های موجود

روش‌ها	دیتاست Market-1501		دیتاست CUHK03	
	mAP	Rank-1	mAP	Rank-1
CamStyle [11]	۶۸/۷	۸۸/۱	-----	-----
UnityStyle[14]	**۸۹/۳**	۹۳/۲	-----	-----
LSRO[22]	۶۶/۱	۸۴/۰	۸۷/۴	۸۴/۶
PNGAN[23]	۷۲/۶	۸۹/۴	-----	۷۹/۸
PT [24]	۶۸/۹	۸۷/۶	۴۲/۰	۴۵/۱
MpRL[25]	۶۷/۵	۸۵/۸	۸۷/۵	۸۵/۴
DG-Net[26]	۸۶/۰	**۹۴/۸**	-----	-----
FD-GAN[27]	۷۷/۷	۹۰/۵	**۹۱/۳**	**۹۲/۶**
افزایش دادگان + UnityStyle[14]	**۹۱/۱**	۹۵/۹	-----	-----
افزایش دادگان + DG-Net[26]	۸۷/۱	**۹۷/۱**	-----	-----
افزایش دادگان + FD-GAN[27]	۷۹	۹۳/۳	**۹۳/۹**	**۹۴/۲**

۶- تحلیل نتایج

الگوریتم‌های مورد استفاده در مقالات [۱۱، ۲۳ و ۲۴] از ابتدا[1] و بدون هیچ گونه تنظیم مجدد[2] با استفاده از دادگان پیشین به علاوه داده افزوده شده توسط این مقاله تحت آموزش قرار گرفتند تا بتوان اثر داده افزوده شده به روش کنونی بر میزان بهبود نتایج را مشخص نمود.

همان‌طور که در جدول۳ نشان داده شده است، استفاده از مجموعه داده پیشنهادی سبب بهبود عملکرد روش‌های موجود برای بازتشخیص افراد در هر سه مدل برتر شده است. هر یک از این ۳ روش از مشخصه های متفاوتی در جهت بهبود تمایز میان تصاویر استفاده کرده اند، و همچنین داده‌های افزوده شده موجب بهبود هر سه روش شده است. لذا می توان نتیجه گرفت که تصاویر تولید شده به روش پیشنهادی توانسته است مشخصه‌های سبک، ژست و شکل ظاهری را به خوبی بازتولید کند طوری که با افزودن این داده‌ها، صحت عملکرد روش‌های اولیه بهتر شده است.

[1] From Scratch
[2] Fine Tuning

باید توجه داشت که برچسب گذاری جدید افراد حاضر در مجموعه داده‌های معیار به صورت نسبی و با توجه به نمای هر فرد انجام شده است. به طور مثال ممکن است نمای بدن فرد پشت به دوربین و زاویه دید چهره فرد به راست باشد که در بر چسب گذاری جدید در گروه نمای پشت به دوربین طبقه بندی شده است. در بخش بعد، آنالیز حساسیت ارائه گردیده است.

۶-۱- آنالیز حساسیت

در این قسمت برای تکمیل تحلیل نتایج، به بررسی تاثیر بخش‌های مختلف شبکه پرداخته شده است. برای انجام یک آنالیز حساسیت (Ablation Study) می‌توان به موارد زیر اشاره کرد.

۱. تعریف بلوک‌ها و اجزای مدل

ما دو بلوک اصلی برای آزمایش داریم:

- **بلوک افزایش داده‌ها (Data Augmentation)**: این بلوک شامل تکنیک‌های مختلفی است که برای افزایش تعداد نمونه‌ها از داده‌هایاولیه استفاده می‌کنیم تا مدل بتواند ویژگی‌های بیشتری را بیاموزد.
- **بلوک‌های شبکه عصبی (CNN Layers, Attention Mechanisms, etc.)**: مدل شامل لایه‌های مختلف شبکه عصبی است که برای استخراج ویژگی‌ها از تصاویر استفاده می‌شود.

۲. طراحی آزمایش‌ها

سه مدل مختلف را پیاده‌سازی شده است:

- **مدل پایه (Base Model)**: این مدل بدون استفاده از هیچ‌گونه تکنیک افزایش داده یا ویژگی‌های اضافی است.
- **مدل با افزایش داده (Augmentation Model)**: این مدل از تکنیک‌های افزایش داده مانند چرخش، تغییر اندازه، تغییرات نور و کنتراست، و غیره استفاده می‌کند.
- **مدل با تمام ویژگی‌ها (Full Model)**: این مدل از تمام تکنیک‌های افزایش داده و همچنین از ماژول‌های شبکه عصبی پیشرفته استفاده می‌کند.

۳. آموزش و ارزیابی مدل‌ها

برای آموزش، از مجموعه داده‌های معروف در زمینه بازشناسی افراد مانند **Market-1501** و **CUHK03** استفاده شده است.

- **معیارهای ارزیابی**: از معیارهای استاندارد بازشناسی افراد مانند دقت (Accuracy) و میانگین دقت (mAP) برای مقایسه مدل‌ها استفاده می‌شود.

۴. نتایج آزمایش‌ها

پس از آموزش مدل‌ها و ارزیابی آن‌ها، نتایج در جدول(۳) زیر ارائه شده است.

۷- جمع‌بندی و نتیجه‌گیری

در این مقاله از شبکه تخاصمی ATNet در ترکیب با مدل Pix2Pix که در ترجمه تصویر به تصویر در زمینه‌های مختلف موفق بوده است، جهت داده افزایی در حوزه بازتشخیص افراد استفاده شده است. روش پیشنهادی مبتنی بر وجود تصاویر فرد از چهار جهت رو به رو، پشت سر، سمت چپ و سمت راست است. با شبکه پیشنهادی، تصاویر نماهای ناموجود مجموعه داده‌های معیار تولید و به مجموعه داده‌های اولیه اضافه شده اند.

شبکه‌های روش‌های موجود برتر در زمینه بازتشخیص افراد شامل UnityStyle، FD-GAN و DG-Net با استفاده از مجموعه داده جدید مجدداً آموزش داده شدند و عمکرد آن‌ها مورد ارزیابی قرار گرفت. نتایج حاصل نشان دهنده بهبود عملکرد همه این روش‌ها است. بنابراین می‌توان نتیجه گرفت که داده افزایی توسط الگوریتم پیشنهادی، سبب بهبود عملکرد روش‌های موجود برای بازتشخیص افراد در چالش تغییر نما و زاویه دید افراد می‌گردد.

در برچسب گذاری جدید تنها چهار نمای روبرو، پشت سر، راست و چپ در نظر گرفته شدند. با بالابردن دقت زاویه‌ای و برچسب گذاری داده با دقت بالاتر می‌توان به بهبود الگوریتم‌ها کمک کرد. پیشنهاد می‌گردد برای تشخیص نمای افراد از الگوریتم‌های موجود در این زمینه استفاده گردد.

مراجع

[1] *Bhuiyan, Md Roman, et al. "Video analytics using deep learning for crowd analysis: a review." Multimedia Tools and Applications 81.19 (2022): 27895-27922.*

[2] *Zheng, Zhedong, Liang Zheng, and Yi Yang. "A discriminatively learned cnn embedding for person reidentification." ACM transactions on multimedia computing, communications, and applications (TOMM) 14.1 (2017): 1-20.*

[3] *Alqahtani, Hamed, Manolya Kavakli-Thorne, and Gulshan Kumar. "Applications of generative adversarial networks (gans): An updated review." Archives of Computational Methods in Engineering 28 (2021): 525-552*

[4] *Creswell, Antonia, et al. "Generative adversarial networks: An overview." IEEE signal processing magazine 35.1 (2018): 53-65.*

[5] *Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*

[6] *Liu, Ming-Yu, Oncel Tuzel. "Coupled generative adversarial networks." Advances in neural information processing systems29, 2016*

[7] *Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.*

[8] *Cao, Chengzhi, et al. "Event-guided person re-identification via sparse-dense complementary learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.*

[9] *Asperti, Andrea, Salvatore Fiorilla, and Lorenzo Orsini. "A generative approach to person reidentification." Sensors 24.4 2024.*

[10] *Chen, Weihua, et al. "Beyond appearance: a semantic controllable self-supervised learning framework for human-centric visual tasks." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.*

[11] *Zhong, Zhun, et al. "Camera style adaptation for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.*

[12] *Wei, Longhui, et al. "Person transfer gan to bridge domain gap for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.*

[13] *Liu, Jiawei, et al. "Adaptive transfer network for cross-domain person re-identification." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.*

[14] *Liu, Chong, Xiaojun Chang, and Yi-Dong Shen. "Unity style transfer for person re-identification." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.*

[15] *Pan, Xingang, et al. "Two at once: Enhancing learning and generalization capacities via ibn-net." Proceedings of the European Conference on Computer Vision (ECCV). 2018.*

[16] *Zhang, Taiping, et al. "Face recognition under varying illumination using gradientfaces." IEEE Transactions on image processing 18.11 (2009): 2599-2606.*

[17] *Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition 2017*

[18] *Zheng, Liang, et al. "Scalable person re-identification: A benchmark." Proceedings of the IEEE international conference on computer vision. 2015.*

[19] *Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*

[20] *Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.*

[21] *Ming, Zhangqiang, et al. "Deep learning-based person re-identification methods: A survey and outlook of recent works." Image and Vision Computing 119 (2022): 104394*

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

[23] *Qian, Xuelin, et al. "Pose-normalized image generation for person re-identification." Proceedings of the European conference on computer vision (ECCV). 2018.*

[24] *Liu, Jinxian, et al. "Pose transferrable person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.*

[25] *Huang, Yan, et al. "Multi-pseudo regularized label for generated data in person re-identification." IEEE Transactions on Image Processing 28.3 (2018): 1391-1403.*

[26] *Zheng, Zhedong, et al. "Joint discriminative and generative learning for person re-identification." proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.*

[27] *Ge, Yixiao, et al. "Fd-gan: Pose-guided feature distilling gan for robust person re-identification." Advances in neural information processing systems 31 (2018).*

هادی کاوه، پژوهشگر و فعال حوزه هوش مصنوعی، در زمینه طراحی و توسعه سیستم‌های هوشمند مبتنی بر پردازش سه‌بعدی و اسکنرهای لیزری فعالیت می‌کند. وی در سال‌های اخیر بر روی به‌کارگیری الگوریتم‌های یادگیری ماشین و بینایی ماشین در کاربردهای صنعتی، از جمله هوشمندسازی معادن، دیجیتال‌سازی محیط‌های فیزیکی، و تحلیل داده‌های حجیم سه‌بعدی متمرکز بوده است. حوزه‌های علاقه‌مندی او شامل تلفیق سخت‌افزار و نرم‌افزار در سامانه‌های هوشمند، پردازش داده‌های لیزری، و توسعه راهکارهای نوآورانه برای صنایع مبتنی بر فناوری‌های نو است.

محمدشهرام معین مدارک کارشناسی مهندسی الکترونیک، کارشناسی ارشد مهندسی الکترونیک و دکترای مهندسی برق را به ترتیب از دانشگاه صنعتی امیرکبیر در سال ۱۳۶۷ ، دانشکده فنی دانشگاه تهران در سال ۱۳۶۹ و پلی تکنیک مونترال کانادا در سال ۱۳۷۹ اخذ نموده است. ایشان با مرتبه دانشیاری به عنوان مشاور رئیس پژوهشگاه ارتباطات و فناوری اطلاعات در حوزه هوش مصنوعی در این پژوهشگاه مشغول به کار و مجری پروژه های متعددی در حوزه های هوش مصنوعی، بیومتریک، چندرسانه ای و کلان داده ها بوده است. دکتر معین سابقه تدریس دروس یادگیری عمیق، بازشناسی الگو، فشرده سازی اطلاعات، داده کاوی، پردازش سیگنالهای دیجیتال و فرآیندهای تصادفی در مقاطع تحصیلات تکمیلی را دارا می باشد. ایشان در انتشار دو کتاب در حوزه هوش مصنوعی، ۴۹ مقاله ژورنال، ۷ فصل کتاب و ۷۹ مقاله کنفرانس مشارکت داشته اند. زمینه های تحقیقاتی ایشان هوش مصنوعی، بازشناسی الگو، پردازش تصویر، تحلیل داده ها و بیومتریک است. دکتر معین رئیس انجمن سیستم‌های هوشمند ایران و سردبیر نشریه علمی فناوری اطلاعات و ارتباطات ایران می باشد.

فربد رزازی تحصیلات خود را در مقطع کارشناسی و کارشناسی ارشد به ترتیب در سالهای ۱۳۷۳ و ۱۳۷۶ رشته مهندسی برق با گرایش مخابرات سیستم از دانشگاه صنعتی شریف اخذ نموده است. سپس وی دکترای خود را در سال ۱۳۸۲ در رشته مهندسی برق با گرایش مخابرات سیستم از دانشگاه صنعتی امیرکبیر دریافت کرده است. از سال ۱۳۷۷ وی به عضویت هیات علمی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران درآمد که در حال حاضر با رتبه دانشیار در این دانشگاه مشغول به کار است. زمینه های تحقیقاتی ایشان در حال حاضر، سیستمهای شناسایی الگو و پردازش سیگنال در کاربردهای نظارتی، احراز هویت، جرم‌کاوی و حفظ حریم شخصی است.