

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشکده مهندسی



دانشگاه فردوسی مشهد
دانشکده مهندسی - کروه کامپیوتر
آزمایشگاه بینایی ماشین

پایان نامه کارشناسی ارشد

هوش مصنوعی

تشخیص بی درنگ چهره در محیط های بدون محدودیت

پژوهش و نگارش: سید سجاد اعمی

استاد راهنما: دکتر حمید رضا پور رضا

شهریور ماه ۱۴۰۰



Ferdowsi University of Mashhad
m v l a b . u m . a c . i r

آزمایشگاه بینایی ماشین

تعهدنامه

اینجانب سید سجاد اعمی دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر دانشکده مهندسی دانشگاه فردوسی مشهد نویسنده پایان‌نامه تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت تحت راهنمایی دکتر حمید رضا پور رضامتعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود و یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه فردوسی مشهد می‌باشد و مقالات مستخرج با نام "دانشگاه فردوسی مشهد" و یا "Ferdowsi University of Mashhad" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند در مقالات مستخرج از رساله رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است، اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ
امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه فردوسی مشهد می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

تقدیم به

پدر و مادر عزیزم

و همه کسانی که درست اندیشیدن را به من آموختند.

سپاس‌گزاری

سپاس خداوند یکتای عزتمندی که رحمت و دانش او در سراسر گیتی گسترده شده، آسمان‌ها و زمین همه از آن اوست و علم و دانش حقیقی را برو هر که بخواهد موهبت می‌فرماید. رحمت و لطف او را بی‌نهایت سپاس می‌گوییم چرا که فهم و درک مطالب این پژوهش را بر من ارزانی داشت و مرا به این اصل رساند که علم و ایمان دو بال یک پروازند. توفیق تلاش به من داد و هر بار که خطا کردم فرصتی دوباره، تا با امید، تلاشی تازه را آغاز کنم و به خواست او به نتیجه‌ی مطلوب نائل آیم. به‌راستی که همه چیز از آن اوست و همه‌چیز به خواست اوست.

بسمه تعالیٰ

شناخته: ب/ک/3	<p>صورتجلسه دفاعیه پایان نامه دانشجوی دوره کارشناسی ارشد</p> <p>دانشجوی کارشناسی ارشد</p>	<p>دانشگاه فنی شهرضا مدیریت تحصیلات تکمیلی</p>
		جلسه دفاعیه پایان نامه تحصیلی آقای/خانم: سید سجاد اعمی رشته/گرایش: مهندسی کامپیوتر/هوش مصنوعی تحت عنوان: تشخیص بی درنگ چهره در محیط های بدون محدودیت و تعداد واحد: ۶ در تاریخ ۱۴۰۰/۰۶/۰۶ با حضور اعضای هیأت داوران (به شرح ذیل) تشکیل گردید. پس از ارزیابی توسط هیأت داوران، پایان نامه با نمره به عدد و درجه به حروف مورد تصویب قرار گرفت.
	<u>امضاء</u>	<u>نام و نام خانوادگی</u>
	دکتر	دکتر حمید رضا پور رضا
		استاد/ استادان راهنمای:
		استاد/ استادان مشاور:
	دکتر	متخصص و صاحب نظر داخلی:
		متخصص و صاحب نظر خارجی:
		ناینده تحصیلات تکمیلی دانشگاه (ناظر)
		نام و نام خانوادگی:
		امضاء:

چکیده

در سال‌های اخیر، به دلیل استفاده از یادگیری عمیق، فناوری تشخیص چهره شاهد پیشرفتهای چشمگیری بوده است. با این حال، استراتژی‌های داده محور یک چالش را به همراه می‌آورد: تصاویر ارسال شده به سامانه تشخیص چهره همیشه برای تشخیص مناسب نیستند و ممکن است چهره‌هایی باوضوح کم، چهره‌های تار در حال حرکت، صورت‌های مسدود و حتی مشکلاتی در پس زمینه وجود داشته باشد. متاسفانه، از آنجا که موتور تشخیص چهره قبل‌چنین چهره‌های بی‌کیفیتی را نمیده است، احتمالاً تصمیمات نادرستی در مورد آن‌ها می‌گیرد.

این پایان نامه با محوریت موضوع تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده می‌باشد که از دو بخش اصلی یافتن چهره و شناسایی چهره تشکیل شده است. روش پیشنهادی در بخش شناسایی چهره می‌باشد. هدف نهایی طراحی بخش نرم افزاری یک عینک هوشمند برای افراد نابینا می‌باشد. هنگامی که فرد نابینا از عینک استفاده کرده و در محیط‌های عمومی به راه رفتن پردازد، دوربینی که روی عینک نصب شده است، چهره افرادی که در زاویه دید آن قرار دارند را بررسی کرده و در صورت یافتن یک چهره آشنا، فرد مورد نظر شناسایی شده و نامش از طریق صدا برای فرد نابینا خوانده می‌شود.

در پیاده سازی این سامانه که باید در مکان‌های عمومی، معابر پیاده و محیط‌های کنترل نشده مورد استفاده قرار بگیرد، چند چالش مهم مانند نورپردازی غیریکنواخت، انسداد، تاری خارج از مرکز دوربین و زاویه نامطلوب چهره نسبت به دوربین وجود دارد. از طرفی این سامانه باید به صورت بی‌درنگ رفتار نماید. زیرا فرصت زیادی برای شناسایی فردی که در خیابان در کنار دوربین می‌گذرد، وجود ندارد. از سوی دیگر به دلیل اجرای پردازش‌ها بر روی پردازنگ تلفن همراه، باید محدودیت منابع را نیز در نظر گرفت و الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد. بدین منظور مبنای تحقیق را بر معماری MobileNetV3 قرار دادیم.

در این پایان نامه محوریت اصلی مطالعات بر روی طراحی یک الگوریتم کارا و مناسب برای مرحله شناسایی بی‌درنگ چهره در شرایط بدون محدودیت است.

علاوه بر موارد گفته شده در بالا، ما فرض کردیم که داده‌های محدودی از هر دسته در اختیار داریم. برای مقابله با این چالش از روش‌های few shot learning و one shot learning استفاده می‌نماییم. با بررسی نتایج حاصل از این پژوهش بر روی تصاویر مجموعه داده LFW و YouTube Faces، دقت روش پیشنهادی ما به ترتیب برابر با ۹۶٪ و ۹۴٪ شد که دقت بالاتری نسبت به روش‌های مشابه می‌باشد.

فهرست مطالب

ج

فهرست جداول

ج

فهرست تصاویر

ذ

فهرست نمادها

۱

۱ مقدمه

۲

۱.۱ مقدمه

۲

۲.۱ ویژگی‌های بیومتریک

۲

۱.۲.۱ ارزیابی ویژگی‌های بیومتریک انسان

۳

۳.۱ سامانه تشخیص چهره

۴

۱.۳.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره

۵

۲.۳.۱ نمای کلی یک سامانه تشخیص چهره

۶

۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر

۶

۱.۴.۱ خوشبندی

۷

۲.۴.۱ طبقه‌بندی

۷

۵.۱ مسئله تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده

۸

۱.۵.۱ چالش‌های سامانه تشخیص چهره

۱۱

۲.۵.۱ نوآوری‌های پایان‌نامه

۱۲

۳.۵.۱ ساختار پایان‌نامه

۱۳	مروی بر کارهای گذشته	۲
۱۴	۱.۲ مقدمه	۱.۲
۱۴	۲.۰.۲ یافتن چهره	۲.۰.۲
۱۴	۱.۰.۲.۲ رویکردهای مبتنی بر دانش و تطبیق کلیشه	۱.۰.۲.۲
۱۵	۲.۰.۰.۲ رویکردهای مبتنی بر ویژگی	۲.۰.۰.۲
۱۸	۳.۰.۰.۲ رویکرد مبتنی بر عامل‌های آماری	۳.۰.۰.۲
۲۳	۴.۰.۰.۲ رویکردهای مبتنی بر یادگیری	۴.۰.۰.۲
۲۴	۳.۰.۲ شناسایی چهره	۳.۰.۲
۲۴	۱.۰.۰.۲ رویکردهای سنتی	۱.۰.۰.۲
۲۵	۲.۰.۰.۲ رویکرد مبتنی بر فیلتر گابور	۲.۰.۰.۲
۲۵	۳.۰.۰.۲ رویکردهای سه بعدی	۳.۰.۰.۲
۲۶	۴.۰.۰.۲ رویکردهای مبتنی بر دوربین حرارتی	۴.۰.۰.۲
۲۷	۵.۰.۰.۲ تشخیص چهره مبتنی بر ویدیو	۵.۰.۰.۲
۲۸	۶.۰.۰.۲ رویکرد مبتنی بر چهره ویژه	۶.۰.۰.۲
۳۰	۷.۰.۰.۲ رویکردهای مبتنی بر یادگیری عمیق	۷.۰.۰.۲
۴۰	۴.۰.۲ نتیجه‌گیری	۴.۰.۲
۴۲	۳ مروی بر روش‌های ارائه شده برای مواجهه با شرایط کنترل نشده	۳
۴۳	۱.۰.۳ مقدمه	۱.۰.۳
۴۳	۲.۰.۳ چالش حالت	۲.۰.۳
۵۱	۲.۰.۳ چالش روشنایی	۲.۰.۳
۵۲	۴.۰.۳ چالش انسداد	۴.۰.۳
۵۵	۵.۰.۳ چالش کمبود تصاویر آموزشی	۵.۰.۳
۶۰	۶.۰.۳ چالش منابع محدود	۶.۰.۳

۶۲	توابع ضرر مبتنی بر افزایش حاشیه	۷.۳
۶۶	نتیجه‌گیری	۸.۳
۶۸	روش پیشنهادی	۴
۶۹	مقدمه	۱.۴
۶۹	پیش‌پردازش	۲.۴
۷۰	همسان‌سازی بافت‌نگار	۱۰.۲.۴
۷۰	یافتن چهره	۲۰.۲.۴
۷۱	تراز کردن تصویر	۳۰.۲.۴
۷۱	دسته‌بندی	۴.۴
۷۲	مدل پیشنهادی پایه	۱۰.۳.۴
۷۶	تابع ضرر	۲۰.۳.۴
۷۸	آموزش مدل و استخراج ویژگی	۳۰.۳.۴
۷۹	فناوری‌های استفاده شده	۴.۴
۸۰	ارزیابی روش پیشنهادی	۵
۸۱	مقدمه	۱.۵
۸۱	معیار ارزیابی	۲.۵
۸۲	مجموعه داده	۳.۵
۸۲	مجموعه داده‌های آموزش	۱۰.۳.۵
۸۳	مجموعه داده‌های آزمون	۲۰.۳.۵
۸۵	پیکربندی الگوریتم	۴.۵
۸۵	نتایج آزمون	۵.۵
۸۵	دقیقت دسته‌بندی	۱۰.۵.۵
۸۶	سرعت پاسخ‌دهی	۲۰.۵.۵

۶ نتیجه‌گیری و پیشنهادات

۸۸

۱.۶ مقدمه

۸۹

۲.۶ بحث و نتیجه‌گیری

۹۰

۳.۶ پیشنهادات

۹۱

منابع و مأخذ

فهرست جداول

۱.۲	مقایسه شبکه عصبی موبایل نت با گوگل نت و VGG	۳۸
۱.۳	خلاصه ای از روش های مقابله با شرایط کنترل نشده	۴۳
۱.۴	مقایسه و ارزیابی برخی از معماری شبکه های رایج در زمینه بینایی ماشین	۷۲
۱.۵	مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره	۸۴
۲.۵	دقیق معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره	۸۵
۳.۵	دقیق بر روی مجموعه داده LFW بر حسب تعداد متفاوت دسته ها	۸۶
۴.۵	پیچیدگی شبکه و تعداد پارامترهای معماری های مختلف بر روی مجموعه داده ImageNet	۸۷
۵.۵	نتایج زمانی بر روی Raspberry Pi به ازای هر فریم	۸۷

فهرست تصاویر

۱.۱	نمای کلی یک سامانه تشخیص چهره	۶
۲.۱	خوشبندی سخت و یافتن شاخص نمونه با توزیع گوسی بر روی داده‌ها	۷
۳.۱	یک طبقه بند غیر خطی ساده	۸
۴.۱	نمونه‌ای از عینک دوربین دار برای افراد نایبینا [۱]	۸
۵.۱	مقایسه تفاوت‌های ظاهری ناشی از روشنایی و تفاوت بین چهره افراد [۲]	۹
۶.۱	زاویه شدید چهره نسبت به دوربین باعث کاهش دقت سامانه می‌شود [۳]	۹
۷.۱	تاری خارج از تمرکز به علت عمق کم میدان دوربین [۴]	۱۰
۸.۱	انسداد در اثر پوشیدن ماسک یا عینک	۱۰
۹.۱	نمونه‌ای از یک سامانه تشخیص چهره بی‌درنگ در محیط کنترل نشده [۵]	۱۱
۱۰	اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می‌باشد [۶]	۱۶
۱۱	نمونه‌ای از مستطیل‌های ویژگی‌های Haar [۷]	۱۸
۱۲	مستطیل‌های ویژگی‌های Haar متناسب با بخش‌های چهره می‌باشند [۷]	۱۹
۱۳	هر مستطیل در اندازه‌های مختلف بر روی بخش‌های مختلف تصویر قرار می‌گیرد [۷]	۱۹
۱۴	بخشی از یک تصویر که مراحل آبشاری برای آن محاسبه می‌شود [۷]	۲۰
۱۵	سطح روشنایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می‌شود [۸]	۲۱
۱۶	برای یافتن چهره‌ها، بخش‌هایی از تصویر که به الگوی HOG شبیه‌تر است را پیدا می‌کنیم [۸]	۲۱
۱۷	نتیجه اجرای روش مبتنی بر NPD [۹]	۲۲

- ۹.۲ نمای کلی روش یافتن چهره تک مرحله ای و متراکم RetinaFace بر اساس هرم ویژگی طراحی شده است.
- ۱۰.۲ این رویکرد از تابع ضرر چند کاره استفاده می نماید. [۱۰]
- ۱۱.۲ نمونه از خروجی الگوریتم یافتن چهره retina [۱۰].
- ۱۲.۲ (آ) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۱۱].
- ۱۳.۲ مدل سازی سه بعدی چهره با اشعه فرو سرخ [۱۲].
- ۱۴.۲ نمای کلی یک سامانه تشخیص چهره مبتنی بر یافتن چهره و ردیابی آن [۱۳].
- ۱۵.۲ (آ) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۱۴].
- ۱۶.۲ شبکه عصبی عمیق برای شناسایی چهره [۴۶].
- ۱۷.۲ (آ) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [۱۶].
- ۱۸.۲ معماری کلی شبکه GoogLeNet [۱۷].
- ۱۹.۲ معماری شبکه VGGFace [۱۸].
- ۲۰.۲ نتیجه موقعیت ۱۹۴ شاخص روی چهره [۱۹].
- ۲۱.۲ علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطای همراه می باشد [۱۹].
- ۲۲.۲ شبکه عصبی پیچشی با معماری TCNN [۴۷].
- ۲۳.۲ تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی [۲۴].
- ۲۴.۲ معماری OpenFace [۲۴].
- ۲۵.۲ جریان یادگیری در معماری OpenFace [۲۴].
- ۲۶.۲ کانولوشن DWS که شامل دو کانولوشن point-wise و depth-wise است. [۲۰].
- ۲۷.۲ لایه های MobileNet به همراه مازول های Squeeze-and-Excite [۲۱].
- ۲۸.۲ معماری مازول Shuffle Attention [۲۱].
- ۲۹.۲ مقایسه چگالی دقیقت در معماری های مختلف شبکه عصبی پیچشی [۲۲].
- ۳۰.۲ رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲۵].
- ۳۱.۲ سه گانه تطبیق - عدم تطبیق [۱۵].
- ۳۲.۲ (آ) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [۱۶].
- ۳۳.۲ معماری کلی شبکه GoogLeNet [۱۷].
- ۳۴.۲ نتیجه موقعیت ۱۹۴ شاخص روی چهره [۱۹].
- ۳۵.۲ علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطای همراه می باشد [۱۹].
- ۳۶.۲ شبکه عصبی پیچشی با معماری TCNN [۴۷].
- ۳۷.۲ تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی [۲۴].
- ۳۸.۲ جریان یادگیری در معماری OpenFace [۲۴].
- ۳۹.۲ کانولوشن DWS که شامل دو کانولوشن point-wise و depth-wise است. [۲۰].
- ۴۰.۲ لایه های MobileNet به همراه مازول های Squeeze-and-Excite [۲۱].
- ۴۱.۲ مقایسه چگالی دقیقت در معماری های مختلف شبکه عصبی پیچشی [۲۲].
- ۴۲.۲ رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲۵].

۴۶	۲۳. مقدار دهی اولیه و بهینه سازی AAM [۲۵].
۴۷	۳.۳. نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۲۵].
۴۷	۴.۳. معماری شبکه پیشنهادی VS2VI [۲۶].
۴۸	۵.۳. معماری مدل یادگیری موقعیت چهره [۲۶].
۴۸	۶.۳. معماری مدل یادگیری بازسازی چهره از رو به رو [۲۶].
۴۹	۷.۳. مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۲].
۵۱	۸.۳. اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۲].
۵۳	۹.۳. تصویر انسداد از ترکیب خطی چهره های آموزشی مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۲۸].
۵۴	۱۰.۳. رویکرد کلی الگوریتم GD-HASLR [۲۸].
۵۵	۱۱.۳. ردیابی، یافتن چهره ها و برچسب زنی [۲۹].
۵۶	۱۲.۳. تصاویر با حاشیه قرمز رنگ، به اشتباہ برچسب زنی شده اند [۲۹].
۵۶	۱۳.۳. استفاده از روش خوش بندی گراف و تعیین تصویر شاخص [۲۹].
۵۷	۱۴.۳. تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۳۰].
۵۸	۱۵.۳. تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین. موارد با حاشیه قرمز خروجی اشتباہ هستند [۳۱].
۵۹	۱۶.۳. تصاویر هر ردیف و هر ستون دارای برخی ویژگی های دیداری مشابه هستند [۳۲].
۵۹	۱۷.۳. ساختار شبکه AD GAN - شبکه GN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۳۳].
۶۰	۱۸.۳. معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۳۴].
۶۱	۱۹.۳. نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۳۵].
۶۲	۲۰.۳. چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۳۵].
۶۳	۲۱.۳. تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مزبین دسته ها ایجاد می نماید [۳۶].
۶۵	۲۲.۳. خط لوله ساز و کار تابع ضرر arcface [۳۷].

۶۵	تابع ضرر ArcFace در مقایسه با توابع ضرر مشهور دیگر [۳۷].	۲۳.۳
۶۹	نمای کلی از روش پیشنهادی.	۱.۴
۷۰	نمونه ای از خروجی الگوریتم یافتن چهره	۲.۴
۷۱	به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد.	۳.۴
۷۳	مدل پایه مبتنی بر لایه های کانولوشن و لایه توجه.	۴.۴
۷۴	یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می کند.	۵.۴
۸۶	دقت روش پیشنهادی در مقایسه با روش ArcFace	۱.۵

فهرست نمادها

فصل ١

مقدمة

۱.۱ مقدمه

استفاده روز افرون از سامانه‌های تشخیص هوشمند چهره موجب اهمیت این شاخه از علم هوش مصنوعی شده است. هدف این است که تصویری به کامپیوتر داده شود و سامانه تشخیص دهد که آیا چهره‌ای در تصویر مشاهده می‌کند یا خیر، و در صورت وجود چهره، در صورت امکان آن را شناسایی نماید. عملکرد این سامانه‌ها در شرایط کنترل شده و آزمایشگاهی به حد مناسبی از بلوغ رسیده است. اما تشخیص چهره در شرایط کنترل نشده، موضوعی چالش برانگیز و در حال پیشرفت می‌باشد. در شرایط مختلفی مانند تابش نور غیر یکنواخت، زاویه نامناسب چهره در مقابل دوربین، وضوح پایین حسگر... گاهی چهره‌ای یافت نمی‌شود و یا چهره یافت شده، قابل شناسایی نمی‌باشد. این مشکلات در سامانه‌های تشخیص چهره مبتنی بر ویدیو، به دلیل عدم ثبات شرایط محیطی و انسانی، تاثیر بیشتری داشته و در نتیجه، باعث کاهش دقیق سامانه در تشخیص افراد می‌شود.

در این بخش به منظور آشنایی کلی با موضوع مورد پژوهش، ابتدا توضیح مختصری در مورد ویژگی‌های بیومتریک انسان و ارزیابی آن‌ها آورده شده است. پس از آن به طور خاص بر روی ویژگی بیومتریک چهره متمرکز شده و در مورد کاربردها، انواع، مراحل، بخش‌ها، مزایا و چالش‌های آن به طور کامل بحث شده است. سپس به تعریف یک مستقله خاص در زمینه تشخیص چهره به صورت بی‌درنگ و در محیط کنترل نشده پرداخته و نیازها و ابزارهای مورد نیاز بررسی شده است.

۲.۱ ویژگی‌های بیومتریک

امروزه در زمینه‌های فراوانی نیاز به سامانه‌ای می‌باشد که هویت اشخاص را بر اساس ویژگی‌های بدن آن‌ها شناسایی کند. این زمینه علمی علاقه مندان فراوانی پیدا کرده و استفاده از ویژگی‌های بیومتریک^۱ در سال‌های اخیر به صورت گسترده مورد استفاده قرار گرفته است. این ویژگی‌ها در هر شخص منحصر به فرد است که از آن جمله می‌توان به اثر انگشت، گفتار، نوع راه رفتن و چهره اشاره کرد. ویژگی‌های بیومتریک را نمی‌توان امانت داد یا قرض گرفت. نمی‌توان خرید یا فراموش کرد و جعل آن هم تقریباً غیر ممکن است. یک سامانه بیومتریک در واقع یک سامانه تشخیص الگو است که یک شخص را بر اساس ویژگی‌های خاص فیزیولوژیک بازنگشتنی می‌کند.

۱.۲.۱ ارزیابی ویژگی‌های بیومتریک انسان

معمولًا ویژگی‌های بیومتریک انسان با ۸ عامل مورد ارزیابی قرار می‌گیرند [۳۹] که عبارت اند از:

^۱Biometric

۱. عمومیت: هر شخصی باید دارای آن ویژگی بیومتریک باشد.
 ۲. یکتاپی: آن ویژگی بیومتریک باید برای هر شخصی منحصر بفرد باشد.
 ۳. دوام: معیاری برای سنجش آنکه یک ویژگی بیومتریک چه مدت بدون تغییر باقی می‌ماند.
 ۴. ارزیابی: ویژگی بیومتریک مورد نظر باید سادگی کافی را در استفاده برای ارزیابی نمونه‌های متفاوت داشته باشد.
 ۵. کارایی: استفاده از ویژگی بیومتریک مورد نظر باید دقت، سرعت و پایداری مطلوب داشته باشد.
 ۶. مقبولیت: فناوری استفاده از ویژگی بیومتریک مورد نظر باید در میان جامعه پذیرش شده باشد.
 ۷. تصدیق هویت: ویژگی فرد به سامانه ارسال شود و سامانه پاسخی مثبت یا منفی برای تصدیق هویت فرد ارائه نماید.
 ۸. تشخیص هویت: ویژگی فرد به سامانه ارسال شود و سامانه با جستجو در پایگاه داده، هویت فرد را استخراج نماید.
- ویژگی بیومتریک چهره تمام موارد بالا را شامل می‌شود و یکی از بهترین انتخاب‌ها برای طراحی یک سامانه تشخیص هویت می‌باشد. پس از موفقیت سامانه شناسایی از طریق اثر انگشت در چند سال اخیر، فناوری تشخیص چهره یکی از مهم‌ترین فناوری‌های بیومتریک برای شناسایی افراد محسوب می‌شود.

۳.۱ سامانه تشخیص چهره

تشخیص چهره همواره یکی از موضوعات مورد مطالعه در علوم کامپیوتر و هوش مصنوعی بوده است. این اهمیت و توسعه کاربرد به دو دلیل مهم می‌باشد:

۱. تشخیص چهره برای استفاده در کاربردهای مختلف مانند کاربردهای امنیتی، قابلیت شناسایی خودکار سریع و بدون دخالت شخص را دارد، سرعت پردازش را بالا برد و خطای کاهش داده است.
۲. سامانه تشخیص چهره نسبت به سامانه‌های بیومتریک قابل اعتمادی مانند تشخیص اثر انگشت و عنایی چشم، ارتباط راحت‌تری با کاربر ایجاد کرده و بدون تماس عضوی از بدن با سامانه، عملیات تشخیص انجام می‌گیرد. البته توسعه کاربردهای دوربین‌های دیجیتالی پیشرفته عامل موثری در توسعه و بالا رفتن طرفداران این سامانه بوده است.

سامانه تشخیص چهره بر اساس الگوریتم‌های شناسایی و مقایسه تصاویر کار می‌کند. اساس و پایه این الگوریتم‌ها شناسایی و تجزیه و تحلیل ویژگی‌های مربوط به اندازه، شکل و موقعیت چشم، بینی، گونه‌ها و اعضای چهره می‌باشد. تصاویر رقمی برای سامانه ارسال می‌شود و سامانه به طور خودکار چهره شخص را در تصویر پیدا می‌نماید و ویژگی‌های آن را استخراج و با نمونه‌های دیگر مقایسه می‌کند. نتیجه این پردازش، لیستی از هویت‌ها است که رتبه بندی شده است.

۱.۳.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره

فناوری تشخیص چهره که دارای مزایایی چون دقت بالا و سطح پایین دخالت فرد می‌باشد، در مواردی مانند کنترل دسترسی^۱، امنیت اطلاعات، اجرا و نظارت بر قانون، شناسایی مجرمین، کنترل و ثبت تردد در سامانه‌های حضور و غیاب، کنترل نامحسوس و ایجاد امنیت در بانک، فروشگاه، فرودگاه و... مورد استفاده قرار می‌گیرد و در صنعت و علم مورد توجه قرار گرفته است.

علاوه بر کاربردهای فوق، شناسایی و پردازش چهره کاربردهای دیگری هم دارند که ارتباطی با تشخیص هویت ندارند. دنبال کردن خط دید چشم و تعیین نژاد، جنس، سن و حالت صورت از جمله این کاربردها هستند که بعضی از آن‌ها در ارتباط بین انسان و کامپیوتر مفید هستند. کاربردهای زیادی برای مباحث شناسایی چهره می‌توان متصور شد که محدوده وسیعی از تصاویر متحرک تا تصاویر ثابت و از کاربردهای امنیتی تا کاربردهای تجاری را شامل می‌شود. این کاربردها را بر اساس نوع تصاویری که استفاده می‌کنند، می‌توان به دو گروه تصاویر ثابت و متحرک تقسیم کرد که در مواردی همچون کیفیت تصویر، زمینه تصویر، در دسترس بودن معیار انطباق و... با یکدیگر تفاوت دارند. دو ویژگی مهم یک سامانه تشخیص چهره عبارتند از:

سرعت تشخیص: بدین معنا که یک الگوریتم تشخیص چهره از لحظه قرارگیری فرد در مقابل دوربین، در چه بازه زمانی می‌تواند هویت فرد درون تصویر را تشخیص دهد.

دقت تشخیص: بدین معنا که یک الگوریتم تشخیص چهره با چه ضریب اطمینانی می‌تواند هویت یک فرد درون تصویر را تشخیص دهد. هرچه تعداد افراد مختلفی که در پایگاه داده ثبت نام شده اند، بیشتر باشد، احتمال خطا در سامانه بیشتر می‌شود و به یک الگوریتم دقیق تر نیاز داریم.

بین سرعت تشخیص و دقت تشخیص، بدنه بستان^۲ وجود دارد. یک الگوریتم کارا باید هر دو ویژگی بالا را در نظر بگیرد.

¹Access Control

²Tradeoff

۲.۳.۱ نمای کلی یک سامانه تشخیص چهره

یک سامانه بیومتریک تشخیص چهره شامل بخش‌های مختلفی می‌باشد که در شکل ۱.۱ نشان داده شده است. شش بخش مهم یک سامانه تشخیص چهره عبارتند از:

حسگر دوربین: این بخش وظیفه گرفتن تصویر چهره را بر عهده دارد. دستگاه گیرنده بسته به نیاز و کاربرد می‌تواند یک دوربین سیاه و سفید، رنگی، یک دوربین مخصوص با قابلیت استخراج اطلاعات عمق یا یک دوربین مادون قرمز باشد.

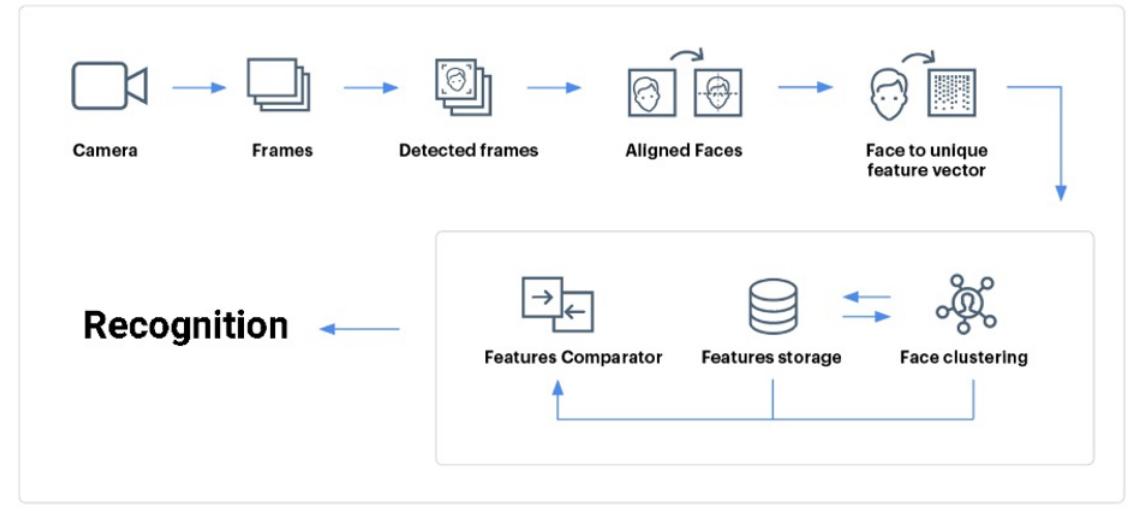
یافتن چهره: تصاویر ورودی به این بخش ابتدا مورد پیش پردازش قرار می‌گیرند. سپس ارزیابی محتوایی شده و داده‌های نامربوط از قبیل پس زمینه، موها و گردن و شانه و... حذف می‌شوند و تنها ناحیه چهره باقی می‌ماند.

تراز کردن چهره: چهره پیدا شده در مرحله قبل، تراز می‌شود. به گونه‌ای که چشم‌ها در راستای افقی قرار بگیرند.

استخراج اطلاعات: در این بخش ویژگی‌های چهره مورد بررسی قرار گرفته و اطلاعات مورد نیاز از تصویر استخراج می‌گردد تا با تصاویر موجود در پایگاه داده مقایسه گردد.

پایگاه داده: این بخش وظیفه ثبت نام، نگهداری و واکنشی ویژگی چهره کاربران را بر عهده دارد. پایگاه داده مجموعه‌ای از تصاویر است که در مرحله طبقه‌بندی مورد استفاده قرار می‌گیرد. در بیشتر روش‌های تشخیص چهره چندین نمای متفاوت از یک شخص در حالتهای مختلف روحی خنده، اخم، عصبانیت، عادی و یا با عینک از کاربر گرفته می‌شود که موجب بالا رفتن ضریب شناسایی سامانه می‌شود.

طبقه‌بندی: در این بخش ویژگی‌های استخراج شده با ویژگی‌های موجود در پایگاه داده تصاویر مقایسه می‌گردد و مشخص می‌شود که آیا چهره مورد نظر در بین چهره‌های موجود می‌باشد یا خیر، و در صورت مثبت بودن جواب، هویت شخص را تایید می‌کند. بر اساس امتیاز بدست آمده از مقایسه که همان درصد تطابق بردار ویژگی گرفته شده با بردارهای ویژگی موجود می‌باشد، چهره مورد نظر مورد تایید قرار گرفته و یا پذیرفته نمی‌شود.



شکل ۱.۱: نمای کلی یک سامانه تشخیص چهره

۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر

۱.۴.۱ خوشبندی

روش خوشبندی^۱ بر اساس محاسبه میزان و معیار شباهت در داده‌ها، آن‌ها را در خوشبندی مختلف قرار می‌دهد. به طور کلی دو

نوع روشنخوشبندی وجود دارد:

خوشبندی سخت^۲: برچسب گذاری به صورت صفر و یک بر روی داده‌ها می‌باشد و مشخص می‌کند داده مورد نظر مربوط به خوش

می‌باشد یا خیر.

خوشبندی نرم^۳: به خوشبندی فازی معروف است و عضویت یک داده به یک خوش به کمک عددی بین صفر و یک به عنوان درجه

عضویت انجام می‌شود و هر داده می‌تواند با توجه به وزن‌های اختصاص داده شده به آن، به چندین خوش تعلق داشته باشد.

داده‌های استفاده شده در فرایند خوشبندی، بدون برچسب بوده و یادگیری بدون ناظر^۴ در فرایند خوشبندی داده‌ها استفاده

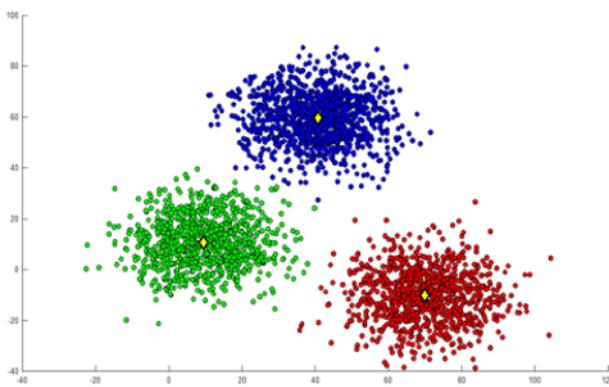
می‌شود. شکل ۲.۱ نمونه‌ای از خوشبندی سخت و یافتن شاخص برای هر خوش را نشان می‌دهد.

¹Clustering

²Hard Clustering

³Soft Clustering

⁴Unsupervised Learning



شکل ۲.۱: خوشبندی سخت و یافتن شاخص نمونه با توزیع گوسی بر روی داده‌ها

۲.۴.۱ طبقه‌بندی

روش‌های طبقه‌بندی^۱ داده‌ها از جمله روش‌های یادگیری با ناظر^۲ هستند که با توجه به یادگیری داده‌های آموزشی به همراه برچسب آن‌ها، داده‌های آزمایشی را نیز برچسب زنی می‌کنند. در طبقه‌بندی داده‌ها می‌توان از توابع سنجش مختلفی برای سنجش میزان تعلق یک داده به هر دسته استفاده کرد. شکل ۳.۱ یک طبقه‌بند غیر خطی ساده را نشان می‌دهد.

در روش‌های طبقه‌بندی که از لحاظ تعداد، بار اصلی الگوریتم‌های یادگیری ماشین را بر دوش می‌کشند، داده‌ها به دو بخش تقسیم می‌شوند. اول داده‌های ورودی (X) نامیده می‌شوند که باید بر اساس آن‌ها، فرآیند پیش‌بینی انجام پذیرد. دوم داده‌های برچسب (Y) هستند که باید مقادیر آن‌ها پیش‌بینی شود. برای این منظور باید تابعی ایجاد شود که ورودی‌ها را گرفته و خروجی موردنظر را تولید کند. فرآیند یافتن این تابع، کشف رابطه‌ای بین مقادیر ورودی (X) و خروجی (Y) است که آن را فرآیند آموزش می‌نامند و تا رسیدن به دقت لازم ادامه می‌باید. الگوریتم‌های مختلفی برای طبقه‌بندی داده‌ها وجود دارد که در این میان می‌توان به شبکه عصبی،^۳ SVM و KNN^۴ اشاره کرد.

۵.۱ مسئله تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده

این پایان نامه با محوریت موضوع تشخیص بی‌درنگ چهره افراد در محیط‌های کنترل نشده با در نظر گرفتن شرایط سخت می‌باشد.

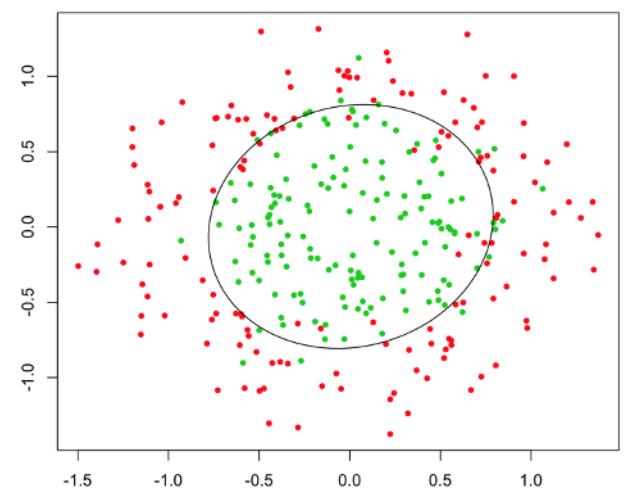
هدف ما، طراحی و ساخت یک عینک هوشمند مانند شکل ۴.۱ برای افراد نایینا می‌باشد. هنگامی که شخص نایینا عینک را بر روی

¹Classification

²Supervised Learning

³Support Vector Machine

⁴K Nearest Neighbor



شکل ۳.۱: یک طبقه بند غیر خطی ساده

چشمانش قرار داده و در محیط‌های عمومی راه می‌رود، دوربینی که بر روی عینک نصب شده است، شروع به بررسی چهره افرادی می‌کند که در زاویه دید آن قرار دارند. در صورت یافتن یک چهره آشنا، فرد مورد نظر شناسایی شده و نام فرد برای شخص نابینا خوانده می‌شود. این مسئله شامل دو بخش اصلی یافتن چهره و شناسایی چهره می‌شود. هریک از این بخش‌ها وزیر بخش‌های آن‌ها در فصل دوم مورد بررسی قرار خواهد گرفت.



شکل ۴.۱: نمونه‌ای از عینک دوربین دار برای افراد نابینا [۱]

۱.۵.۱ چالش‌های سامانه تشخیص چهره

سامانه‌های تشخیص چهره به سطح مشخصی از بلوغ رسیده‌اند، اما توسعه آن‌ها در شرایط کنترل نشده و کاربردهای واقعی هنوز مسیر طولانی در پیش دارد. برای مثال، تشخیص چهره در ویدیو در محیطی با تغییرات شدید نورپردازی و حالت چهره، انسداد صورت، وضوح پایین تصویر و...، و ردیابی آن در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی و... مشکل می‌باشد. دلیل اصلی به وجود

آمدن چالش‌ها این است که چهره انسان یک شی صلب نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. در ادامه مهم ترین چالش‌ها بر شمرده شده است.

نورپردازی^۱: روشنایی محیط در شب و روز و یا در محیط داخلی و خارجی به شدت تغییر می‌کند. با توجه به ساختار سه بعدی چهره، یک منبع نور مستقیم می‌تواند سایه‌های قوی بر روی چهره ایجاد کند که برخی ویژگی‌های چهره را تغییر می‌دهد. همانطور که در

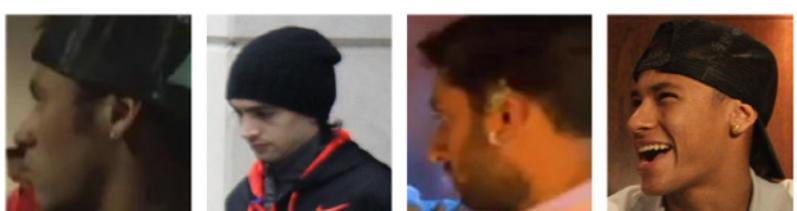
شکل ۵.۱ نشان داده شده است که گاهی تفاوت‌های ظاهری ناشی از روشنایی، بیشتر از تفاوت بین چهره افراد مختلف می‌باشد.



شکل ۵.۱: مقایسه تفاوت‌های ظاهری ناشی از روشنایی و تفاوت بین چهره افراد [۲]

حالت^۲: زاویه چهره نسبت به حسگر دوربین، می‌تواند سامانه را با چالش مواجه نماید. چهره با زاویه تند و چهره‌های نیم رخ مانند

شکل ۶.۱ باید برای الگوریتم تشخیص چهره، قابل شناسایی باشند. بنابراین ویژگی‌های استخراج شده توسط الگوریتم باید به گونه‌ای باشند که در هر زاویه‌ای امکان استخراج آن‌ها وجود داشته باشد.



شکل ۶.۱: زاویه شدید چهره نسبت به دوربین باعث کاهش دقت سامانه می‌شود [۳]

تاری خارج از مرکز^۳: اگر عمق میدان دوربین کم باشد و چهره‌ها با فاصله‌های مختلف از دوربین باشند، مشکل تاری خارج از مرکز

رخ خواهد داد. اگر دوربین طوری تنظیم شود که چهره نزدیک‌تر، واضح‌تر دیده شود، در مقابل باعث می‌شود که چهره دورتر، مقداری

مات شود و برعکس. این مسئله می‌تواند برای سامانه تشخیص چهره در دسر ساز شود. نمونه‌ای از این اثر در شکل ۷.۱ قابل مشاهده

می‌باشد.

انسداد^۴: اگر در حال تصویر برداری از یک جمع باشیم، احتمال اینکه چهره فردی توسط فرد دیگری مقداری پوشانده شود، بسیار

¹Lighting

²Pose

³Out Of Focus Blur

⁴Occlusion



شکل ۷.۱: تاری خارج از مرکز به علت عمق کم میدان دوربین [۴]



شکل ۸.۱: انسداد در اثر پوشیدن ماسک یا عینک

بالاست. همچنین اگر بخشی از چهره فرد توسط موها پوشیده شده باشد، از عینک آفتابی یا ماسک استفاده کند، انسداد چهره رخ خواهد داد. نمونه ای از این اثر در شکل ۸.۱ قابل مشاهده می‌باشد.

سن^۱ : بیشتر روش‌های مرسوم تشخیص چهره تغییرات سن را نادیده می‌گیرند. بعضی از رویکردها به طور منظم پایگاه داده تصاویر را به روز رسانی کرده و بازآموزی سامانه را انجام می‌دهند. این راه حل فقط برای سامانه‌هایی مناسب است که اغلب وظیفه احراز هویت کارمندان را انجام می‌دهد. در شرایط دیگر سن موضوع را باید جدی گرفت و تلاش کرد تا سامانه نسبت به این نوع تغییرات قوی‌تر شود.

کمبود داده‌های آموزشی: سامانه‌های تشخیص چهره در کاربردهای واقعی دارای مشکل کمبود داده‌های آموزشی برای آموزش سامانه می‌باشند. تعداد افراد در محیط کنترل نشده زیاد می‌باشد و قادر به در اختیار داشتن حجم بالایی از داده‌های آموزشی برای هر کدام از افراد نیستیم. از طرفی کاهش تعداد داده‌های آموزشی می‌تواند دقیق سامانه را به شدت کاهش دهد.

منابع محدود: در صورت اجرای پردازش‌های سامانه توسط تلفن همراه، باید محدودیت منابع را در نظر گرفت. تلفن همراه نسبت به رایانه، دارای قدرت پردازندۀ و حافظه پایین‌تر و منبع انرژی محدودتر می‌باشد که باعث می‌شود الگوریتم‌هایی با پیچیدگی محاسباتی بالا بر روی این دستگاه‌ها قابل اجرا نباشد. بنابراین الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد.

¹Aging

زمان: یکی از چالش‌های موجود، فضایی پر از چهره‌های مختلف در مکان‌های عمومی و در مقابل، نیاز به واکنش سریع توسط سامانه است. مشابه شکل ۹.۱ در فضاهای عمومی و معابر پیاده مردم با سرعت از کنار دوربین عبور می‌کنند و سامانه باید قابلیت تشخیص چهره آن‌ها در کسری از ثانیه را داشته باشد. اگر کاربر سامانه با افراد جدید دیدار داشته باشد، سامانه باید به سرعت یاد بگیرد که چهره افراد جدید را تشخیص دهد.



شکل ۹.۱: نمونه‌ای از یک سامانه تشخیص چهره‌بی‌درنگ در محیط کنترل نشده [۵]

۲.۵.۱ نوآوری‌های پایان‌نامه

همان‌طور که در بالا شرح داده شد، شناسایی بی‌درنگ چهره در محیط بدون محدودیت و با دقت بالا با چالش‌های بسیاری همراه است. بنابراین در این پژوهش تلاش می‌کنیم تا روشی برای تشخیص دقیق‌تر و بی‌درنگ چهره توسط شبکه عصبی عمیق در محیط بدون محدودیت پیشنهاد دهیم. در این پژوهش دو نوآوری ارائه می‌شود. اول یک معماری شبکه عصبی پیچشی به نام MobileNetV3 معرفی می‌شود که از ترکیب معماری شبکه عصبی پیچشی MobileNetV3 و یک واحد توجه به نام MobileNetV3 می‌باشد. در ادامه تابع ضرر MobileNetV3 اصلاح شده را معرفی می‌کنیم. با استفاده از ترکیب این دو نوآوری، دقت و سرعت بالاتری نسبت به سایر روش‌های مشابه بدست می‌اوریم.

۳.۵.۱ ساختار پایان نامه

در فصل دوم به مرور رویکردهای مختلف در زمینه تشخیص چهره خواهیم پرداخت. در فصل سوم مروری بر روش‌های ارائه شده برای مواجهه با شرایط کنترل نشده خواهیم داشت. در فصل چهارم روش پیشنهادی را ارائه می‌نماییم. در فصل پنجم به ارزیابی روش پیشنهادی و مقایسه آن با سایر روش‌ها می‌پردازیم و در پایان در فصل ششم به جمع بندی و نتیجه‌گیری خواهیم پرداخت.

۲ فصل

مروری بر کارهای گذشته

۱.۲ مقدمه

همانطور که در فصل ۱ بیان شد، دو موضوع یافتن چهره^۱ در تصویر و شناسایی چهره^۲، بخش‌های اصلی سامانه تشخیص چهره می‌باشند. اگرچه این دو بخش برای انسان کار ساده‌ای به نظر می‌رسد، اما برای کامپیوترها همیشه با چالش همراه بوده است. دلیل این سختی می‌تواند تفاوت تصویرها در مقیاس، حالت چهره، پس زمینه، تابش نور، انسداد و... باشد. در ادامه به بررسی روش‌هایی برای یافتن و شناسایی چهره در دو بخش مجزا پرداخته شده است.

۲.۱ یافتن چهره

یافتن مکان چهره در تصویر، اولین گام در فرایند تشخیص چهره می‌باشد که نقشی کلیدی در سامانه ایفا می‌نماید. هدف اصلی الگوریتم یافتن چهره این است که تعیین کند آیا چهره‌ای در تصویر وجود دارد یا خیر و در صورت وجود چهره، مکان آن را بیابد. یافتن چهره در تصویر، امری پیچیده است. زیرا چهره انسان همواره دستخوش تغییراتی مانند شرایط روشنایی ووضوح تصویر، حالت چهره، رنگ پوست، حضور عینک یا موی صورت و... می‌شود. در سال ۲۰۰۲ Yang و همکاران^[۴۰] یک دسته‌بندی برای روش‌های یافتن چهره ارائه کردند که در ادامه شرح داده می‌شود.

۱.۲.۲ رویکردهای مبتنی بر دانش و تطبیق کلیشه

این رویکردها به مجموعه‌ای از قوانین بستگی دارند و بر اساس دانش انسان در مکان قرار گرفتن اجزای چهره و ویژگی‌های خاصی که در چهره انسان وجود دارد، عمل می‌کنند. یافتن یک جفت چشم در تصویر و سپس جستجو اطراف آن برای یافتن چهره، مثالی از این روش می‌باشد. ابتدا مکان چشم‌ها و بالاترین نقطه سر پیدا می‌شود. سپس فاصله بین چشم تا بالاترین نقطه محاسبه شده و به عنوان یک مرجع برای یافتن نواحی دیگر مانند بینی و دهان مورد استفاده قرار می‌گیرد. این روش زمانی که مو بر روی پیشانی ریخته باشد یا در حضور عینک، به درستی عمل نمی‌کند. یا به عنوان مثالی دیگر، اگر یک توالی از چند فریم در اختیار باشد، می‌توان به کمک اطلاعات حرکت^۳، یافتن چهره را انجام داد. برای این کار به کمک تفاضل فریم‌ها، قسمت متحرک نسبت به پس زمینه شناسایی شده و بخش بالای آن جدا می‌شود. بدین ترتیب می‌توان با احتمال بالا، مکان چهره یک فرد را در یک تصویر پیدا نمود. این رویکرد در مواجه

¹Face Detection

²Face Recognition

³Motion

با اجسام متحرک دیگری مانند اتومبیل دچار اشتباه می‌شود. نمونه‌های دیگری از این قوانین عبارتند از:

• هر چهره دارای دو فرو رفتگی برای چشم‌ها است و چیزی شبیه به ابرو روی این فرو رفتگی‌ها قرار دارد.

• صورت شامل بینی، چشم‌ها و دهان در فاصله‌ها و موقعیت‌های خاصی با یکدیگر می‌باشد.

• چهره مانند یک ناحیه کوچکتر است که بر روی یک ناحیه بزرگتر مانند شانه‌ها قرار گرفته است.

• چهره انسان متقارن است.

این رویکردها با استفاده از قالب^۱ های از پیش تعیین شده برای یافتن چهره‌ها با همبستگی بین الگوها و تصاویر ورودی استفاده می‌نمایند. چهره انسان را می‌توان به چشم، صورت، بینی و دهان تقسیم کرد. همچنین، با استفاده از روش تشخیص لبه، یک مدل صورت می‌تواند توسط لبه‌ها ساخته شود. شکل ۱۱.۳ یک نمای کلی از رویکرد مبتنی بر کلیشه را نشان می‌دهد. این رویکرد ساده است، اما برای تشخیص چهره ناکافی است. با این حال، قالب‌های انعطاف‌پذیر برای مقابله با این مشکلات پیشنهاد شده‌اند. مشکل بزرگ این رویکردها، ساختن یک مجموعه مناسب از قوانین است. اگر قوانین خیلی ساده یا خیلی دقیق باشند، الگوریتم همیشه به درستی عمل نمی‌کند. این رویکرد به تنها بی کافی نیست و موفق به یافتن چهره‌ها در شرایط کنترل نشده با تعداد زیادی چهره نمی‌باشد.

۲۰.۲ رویکردهای مبتنی بر ویژگی

این رویکردها با استخراج ویژگی^۲ های ساختاری چهره، چهره‌ها را پیدا می‌نمایند. ابتدا به عنوان یک طبقه‌بند، آموزش دیده و سپس برای تمایز میان نواحی شامل چهره و بدون چهره در تصویر استفاده می‌شوند. ایده این رویکرد، غلبه بر محدودیت دانش ما از چهره‌ها می‌باشد. در [۶] تعدادی از این رویکردها مورد بررسی قرار گرفته است:

¹Template

²Feature

۱.۲.۲.۲ رویکرد مبتنی بر لبه

ابتدا به کمک یک الگوریتم لبه یاب^۱، لبه‌های تصویر بدست می‌آید، سپس نازک سازی می‌شوند و شاخدهای اضافه حذف می‌گردد. گوشه لبه‌ها تشخیص داده می‌شوند و هر مولفه متصل^۲ به شاخه مرکزی آن کاوش می‌یابد. اجزایی که ویژگی چهره در آن‌ها نیست حذف می‌شوند و اجزای نهایی به عنوان سمت چپ چهره، خط مو، یا سمت راست چهره برچسب گذاری می‌شوند. در یک آزمایش که ۶۰ تصویر دارای پس زمینه پیچیده حاوی ۹۰ چهره به این سامانه داده شده است، سامانه قادر به یافتن ۷۶٪ چهره‌ها بوده است.

۲.۲.۲.۲ رویکرد مبتنی بر اطلاعات سطح خاکستری

سطوح خاکستری^۳ تصویر چهره شامل اطلاعات مفیدی می‌باشد. برای مثال ابروها، مردمک چشم و لب‌ها معمولاً تاریک‌تر از سایر نواحی صورت هستند. این ویژگی‌ها می‌توانند به یافتن یک چهره در تصویر کمک نماید. در این رویکرد ابتدا بر روی تصویر ورودی، عملیات بسط تباین^۴ و عملیات مورفولوژی^۵ مبتنی بر سطح خاکستری انجام می‌شود تا تصویر بهبود پیدا کند و یافتن نواحی تیره تر، راحت شود. سپس تصویر به چندین بخش تقسیم می‌شود و سطوح خاکستری بخش‌ها مورد بررسی قرار می‌گیرد. مزیت این رویکرد، کارایی در تصاویر باوضوح پایین می‌باشد (شکل ۱.۲).



شکل ۱.۲: اطلاعات سطح خاکستری حتی دروضوح پایین نیز قابل دسترسی می‌باشد [۶].

¹Edge Detection

²Connected Component

³Gray level

⁴Contrast Stretching

⁵Morphological Operations

۳.۲.۲.۲ رویکرد مبتنی بر اطلاعات رنگی

رنگ در تصویر اطلاعات با ارزشی به ما می‌دهد و می‌توان از رنگ پوست انسان برای یافتن چهره در تصویر استفاده کرد. برای این کار

ابتدا رنگ‌ها هنجار سازی^۱ می‌شوند تا اثر نور پردازی از بین برود. گسترده‌ترین فضای رنگ مورد استفاده RGB می‌باشد.

$$r = \frac{R}{R + G + B} \quad (1.2)$$

$$g = \frac{G}{R + G + B} \quad (2.2)$$

$$b = \frac{B}{R + G + B} \quad (3.2)$$

در روابط بالا R میزان سطح رنگ قرمز، G میزان سطح رنگ سبز و B میزان سطح رنگ آبی در هر پیکسل از تصویر می‌باشد.

همچنین r ، g و b به ترتیب مقدار هنجار سازی شده برای رنگ قرمز، سبز و آبی می‌باشد. واضح است که:

$$r + g + b = 1 \quad (4.2)$$

پس می‌توان فقط با داشتن مقدار r و g مقدار b را بدست آورد. با توجه به بافت‌نگار^۲ رنگ سبز و قرمز تصویر، رنگ پوست انسان، بخش کوچکی از بافت‌نگار را اشغال می‌کند. بنابراین با بررسی پیکسل‌های تصویر، می‌توان با دقت بالایی احتمال حضور چهره در تصویر را تشخیص داد و رنگ پوست انسان را می‌توان به راحتی با یکتابع گوسی تخمین زد. فضاهای رنگی دیگری نیز در این زمینه مورد استفاده قرار گرفته است. مانند HSV، YIQ، YCrCb و Lab. یک ایده خوب این است که وقتی شخص از دوربین فاصله زیادی دارد از رنگ پوست استفاده کنیم و وقتی شخص نزدیک به دوربین می‌باشد از ویژگی‌های قدرتمندتر چهره استفاده نماییم.

¹Normalization

²Histogram

۳.۲.۲ رویکرد مبتنی بر عامل‌های آماری

رویکردهای بخش‌های قبل بر روی اطلاعات استخراج شده از تصاویر چهره در شرایط آزمایشگاهی تکیه می‌کنند و اگر یک تصویر

چهره در شرایط کنترل نشده و پس زمینه پیچیده باشد، بسیاری از این رویکردها شکست می‌خورند. استفاده از عامل‌های آماری^۱

برای ویژگی‌ها و ارائه یک مدل احتمالی برای چهره، باعث انعطاف پذیری بیشتر سامانه می‌شود. این رویکرد قادر است در مواجهه با جا

به جایی، چرخش و تغییر مقیاس با دقت بیشتری عمل نماید. در یک رویکرد دقیق‌تر از شبکه‌های بیز^۲ برای یافتن احتمالاتی چهره

بهره گرفته شده است. یافتن چهره در حضور عینک و برخی ویژگی‌های از دست رفته نیز توسط این رویکرد انجام می‌شود.

صدها رویکرد برای یافتن چهره ارائه شده است تا آن را پیشرفت‌تر و دقیق‌تر نماید، اما انقلاب الگوریتم‌های یافتن چهره در سال ۲۰۰۱

بود. زمانی که Viola و Jones در [۷] یک الگوریتم چهره یاب بی‌درنگ معرفی کردند که قادر به یافتن چهره با دقت بالا بود. در

ادامه به شرح این الگوریتم می‌پردازیم.

پیش‌پردازش: ابتدا تصویر از فضای RGB به تصویر سطح خاکستری تبدیل می‌شود. زیرا تشخیص چهره‌ها در تصویر خاکستری

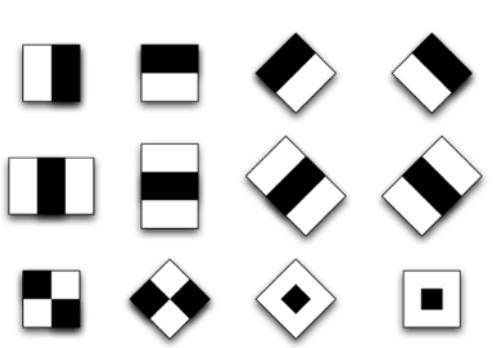
برای سامانه آسان است. سپس در صورت نیاز، پیش‌پردازش‌هایی مانند تغییر اندازه^۳، برش^۴، تار شدن^۵ و تیزکردن لبه‌های تصویر

^۶ انجام می‌شود.

ویژگی‌های Haar: تمام چهره‌های انسانی ویژگی‌های مشترکی دارند. برای مثال ناحیه چشم تاریک‌تر از پیکسل‌های همسایه آن است

وناحیه بینی از چشم روشن‌تر است. ویژگی‌های Haar مستطیل‌هایی هستند که نشان دهنده بخش‌های مختلف صورت می‌باشند.

شکل ۲.۲ نمونه‌هایی از مستطیل‌های ویژگی‌های Haar را نشان می‌دهد.



شکل ۲.۲: نمونه‌هایی از مستطیل‌های ویژگی‌های Haar.

¹Statistical Parameters

²Bayes Rule

³Resizing

⁴Cropping

⁵Blurring

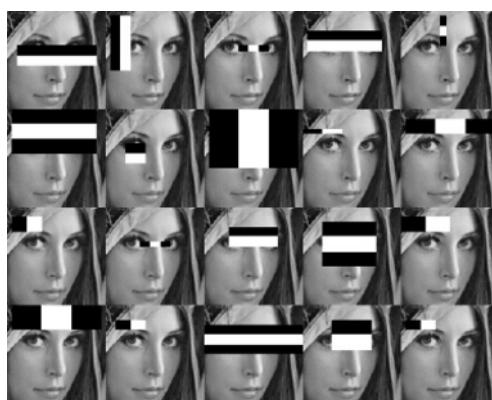
⁶Sharpening



شکل ۳.۲: مستطیل های ویژگی های Haar مناسب با بخش های چهره می باشند [۷].

ویژگی های Haar برای تشخیص چشم، بینی، دهان و... با کمک تشخیص لبه، تشخیص خط و تشخیص مرکز در تصویر و استخراج ویژگی برای یافتن چهره استفاده می شود. مستطیل های ویژگی های Haar، مناسب با بخش های چهره می باشند که مثالی از آن در شکل ۳.۲ نشان داده شده است.

همانطور که در شکل ۴.۲ مشاهده می شود، هر مستطیل در بخش های مختلف چهره در روندی تکراری با اندازه های مختلف قرار می گیرد و نتیجه نهایی از کم کردن مجموع سطح روشنایی پیکسل های زیر بخش های سفید به دست می آید که یک عدد می باشد. از یک پنجره با اندازه 24×24 برای قرار دادن مستطیل ها بر روی تصویر استفاده می شود که تعداد زیاد و اندازه های مختلف آن ها باعث می شود برای محاسبه نتیجه نهایی نیاز به انجام بیش از ۱۶۰۰۰۰ محاسبه باشد که زمان زیادی برای هر تصویر خواهد گرفت.



شکل ۴.۲: هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [۷].

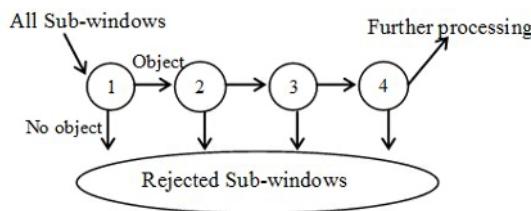
شکل ۴.۲: تمام ویژگی های Haar برای تصویر ورودی مناسب نخواهد بود. بعضی از این ویژگی ها باید نادیده گرفته شوند و فقط ویژگی های مرتبط انتخاب شوند تا در زمان صرفه جویی شود. این کار به صورت خودکار به کمک عنصر Ada Boost انجام می شود. Ada Boost یک الگوریتم مبتنی بر یادگیری ماشین می باشد که ویژگی های کاربردی را از میان تعداد زیادی ویژگی پیدا می کند. بعد از شناسایی ویژگی های مختلف، مشخص می گردد که هر یک از پنجره ها برای بخشی از چهره مناسب می باشد یا خیر. هر کدام از

ضرایب انتخاب شده مثبت در نظر گرفته می‌شود، در صورتی که حداقل بتواند بیش از نیمی از موارد را تشخیص دهد. این ویژگی‌ها با عنوان طبقه‌بندی‌های ضعیف^۱ معرفی می‌شوند. Ada Boost در طبقه‌بند قوی^۲، تعداد زیادی طبقه‌بند ضعیف را با هم ترکیب می‌کند. رابطه کلی آن به صورت زیر می‌باشد.

$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \dots \quad (5.2)$$

که در آن F طبقه‌بند قوی می‌باشد که از تعدادی f_i که طبقه‌بند ضعیف می‌باشد، تشکیل شده است. هر کدام از طبقه‌بندی‌های ضعیف، یک خروجی صفر یا یک تولید می‌کنند. α_i وزن مربوط به طبقه‌بند می‌باشد. با استفاده از این الگوریتم وزن دادن به ویژگی‌ها، بیش از ۱۶۰۰۰ ویژگی قبلی به کمتر از ۲۵۰۰ ویژگی کاهش پیدا می‌کند.

مراحل آشنازی^۳: اگر تصویر به مربع‌های 24×24 پیکسلی تقسیم شده و با پردازش هر بخش که ۲۵۰۰ ویژگی دارد، تشخیص داده شود که چهره‌ای در تصویر وجود دارد یا خیر، حجم محاسبات بسیار زیاد خواهد بود. مراحل آشنازی این فرایند را سریع‌تر انجام می‌دهد. ۲۵۰۰ ویژگی هر مربع 24×24 به دسته‌بندی‌های مختلف تقسیم می‌شود. برای مثال ۱۰ ویژگی در دسته اول، ۲۰ ویژگی در دسته دوم، ۱۰۰ ویژگی در دسته سوم و... می‌توان بعد از پردازش هر دسته، در ارتباط با وجود یا عدم وجود چهره در آن دسته تصمیم گرفت تا بخش‌هایی که چهره در آن وجود ندارد زودتر حذف شوند. شکل ۵.۲ یک نمای کلی از روند تشخیص آشنازی را نشان می‌دهد. مجموعه‌ای از طبقه‌بندی‌ها به هر زیرپنجره اعمال می‌شود. طبقه‌بند اولیه تعداد زیادی از نمونه‌های منفی را حذف می‌کند و پردازش کمی دارد. لایه‌های بعد، منفی‌های اضافی را حذف می‌کنند که نیاز به محاسبات بیشتری دارند. این الگوریتم عملکرد بسیار خوبی در برنامه‌های کاربردی بی‌درنگ و در حضور پس زمینه‌های شلوغ نشان داده است. اما هنوز در برای چهره‌هایی که رو به روی دوربین نیستند، تغییرات شدید نور، انسداد و... دارای چالش می‌باشد.

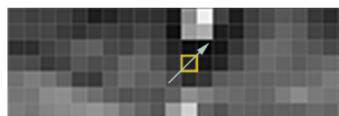


شکل ۵.۲: بخشی از یک تصویر که مراحل آشنازی برای آن محاسبه می‌شود [۷].

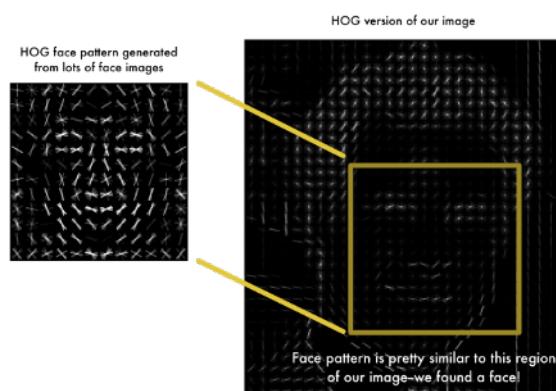
¹Weak Classifier

²Strong Classifier

³Cascading



شکل ۶.۲: سطح روشنایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می شود [۸].



شکل ۷.۲: برای یافتن چهره ها، بخش هایی از تصویر که به الگوی HOG شبیه تر است را پیدا می کنیم [۸].

در سال ۲۰۰۵ دلال و همکاران [۸] روشی به نام بافت نگار شبیه های جهت دار^۱ ارائه کردند که به اختصار HOG نامیده می شود.

در این روش ابتدا تصویر خاکستری می شود، زیرا نیازی به رنگ نیست. پیرامون هر پیکسل بررسی می شود تا مشخص شود نسبت به پیکسل های پیرامونش چقدر تاریک می باشد. مطابق شکل ۶.۲ جهتی انتخاب می شود که به سمت پیکسل های تاریکتر باشد. این روند برای همه پیکسل های تصویر انجام می شود و به ازای هر پیکسل یک جهت ذخیره خواهد شد که روندی از روشنایی به تاریکی را در تصویر نمایش می دهند.

دلیل استفاده از جهت ها این است که اگر پیکسل ها به طور مستقیم استفاده شوند، تصویر تاریک و تصویر روشن از یک چهره مشخص، دارای سطح روشنایی متفاوتی خواهند بود. اما با در نظر گرفتن جهتی تغییر روشنایی، هم تصویر تیره و هم تصویر روشن، نمایش یکسانی خواهند داشت که حل مسئله را آسان تر می کند. ذخیره جهت ها برای تمام پیکسل ها باعث افزایش جزئیات می شود. لذا روند اصلی روشنایی و تاریکی در سطحی بالاتر در نظر گرفته می شود، به طوری که بتوان الگوی اصلی تصویر را دید. تصویر به بخش های ۱۶×۱۶ پیکسل تقسیم می شود و در هر بخش تعداد جهت های به سمت بالا، پایین، چپ و راست شمارش می شود. سپس بخش های درون تصویر با جهت هایی که بزرگتر بودند، جایگزین می شود. نتیجه نهایی، تبدیل تصویر به یک نمایش ساده شده از ساختار چهره است که در شکل ۷.۲ مشاهده می شود. برای یافتن چهره ها در این الگوریتم، بخش هایی از تصویر که به الگوی HOG شبیه تر است، مشخص می شود. با استفاده از این روش، می توان چهره ها را در تصویر به سادگی پیدا کرد.

در سال ۲۰۱۵ شنگ‌کای لیاوه [۹] یک روش دقیق و سریع برای یافتن چهره ارائه دادند که در آن از اختلاف پیکسل

^۱Histograms Of Oriented Gradients

هنجارسازی شده یا NPD^۱ برای یافتن چهره استفاده می‌شود. ارزیابی ویژگی NPD بسیار سریع است و دسترسی به حافظه تنها با استفاده از یک جدول جستجو می‌باشد. در این روش نشانه گذاری یا خوشه بندی در مرحله آموزش نیز لازم نیست و در برابر تغییرات نور، حالت، انسداد، تصاویر باوضوح پایین و... مقاوم است. NPD بین دو پیکسل در یک تصویر به صورت زیر تعریف شده است:

$$f(x, y) = \frac{x - y}{x + y} \quad (6.2)$$

که در آن x و y بزرگتر از صفر هستند و $f(0, 0) = 0$ برای حالتی که $x = y$ باشد، برابر صفر است. این عمل بر روی هر جفت پیکسل از تصویر اجرا می‌شود. اگر تصویر ورودی مربعی با ابعاد $s \times s$ باشد و p تعداد پیکسل‌ها باشد، آنگاه تعداد ویژگی‌های استخراج شده برابر $d = p(p - 1)/2$ می‌باشد. سپس علامت^۲ ویژگی‌های استخراج شده مورد استفاده قرار می‌گیرد که وابسته به اندازه سطح روشنایی پیکسل‌ها نمی‌باشد. بلکه تنها نشان می‌دهد کدام ناحیه روشن‌تر و کدام ناحیه تیره‌تر می‌باشد. همچنین این ویژگی‌ها به خدشه^۳ حساس نمی‌باشند. در نهایت ویژگی‌های استخراج شده به عنوان ورودی به یک سامانه یادگیری داده می‌شود.

نمونه‌ای از نتیجه اجرای این الگوریتم بر روی مجموعه داده Fddb در شکل ۸.۲ آمده است.



شکل ۸.۲: نتیجه اجرای روش مبتنی بر NPD.

تمام رویکردهای ارائه شده که یافتن چهره را با مدل سازی صریح از ویژگی‌های صورت انجام می‌دهند، در برابر تغییرات غیر قابل پیش‌بینی چهره و شرایط محیطی دچار مشکل می‌شوند. اگرچه بعضی از تلاش‌های اخیر مبتنی بر ویژگی، توانایی مقابله با شرایط کنترل نشده را بهبود داده‌اند، اما بیشتر آن‌ها هنوز به چهره‌های رو به رو و شرایط کنترل شده محدود می‌شوند، و به عنوان یکی از روش‌های یک سامانه ترکیبی در نظر گرفته شده‌اند. پس نیاز به روش‌هایی هست که بتوانند در شرایط خصم‌مانه تر مانند تشخیص چهره‌های متعدد در زمینه‌های شلوغ به خوبی عمل کنند.

¹Normalized Pixel Difference

²Sign

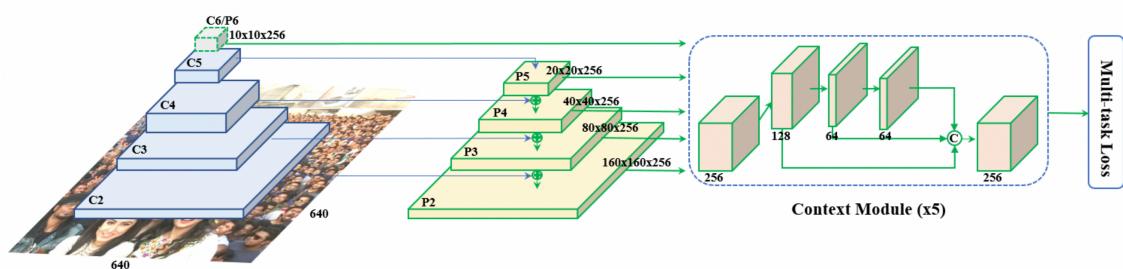
³Noise

۴.۲.۲ رویکردهای مبتنی بر یادگیری

با توجه به آنچه در [۴۱] آمده است، چالش اصلی در یافتن چهره این است که ویژگی‌هایی مانند Haar و HOG اطلاعات برجسته چهره را در شرایط مختلف نما، نورپردازی، رنگ پوست، انسداد، استفاده از لوازم آرایشی و... استخراج نمی‌کنند. این محدودیت بیشتر به دلیل ویژگی‌های استفاده شده در طبقه‌بندها است. با پیشرفت‌های اخیر در رویکردهای یادگیری عمیق و در دسترس بودن پردازنده‌های گرافیکی، استفاده از شبکه‌های عصبی پیچشی عمیق برای استخراج ویژگی امکان پذیر شده است. در سال ۲۰۱۹ و همکاران [۱۰] یک روش مبتنی بر شبکه عصبی پیچشی برای یافتن چهره در تصویر پیشنهاد کردند. در این روش برای آموزش شبکه پیچشی از یکتابع ضرر مبتنی بر یادگیری چندکاره^۱ استفاده شده است که به صورت رابطه ۷.۲ می‌باشد.

$$L = L_{cls} + L_{box} + L_{pts} + L_{pixel} \quad (7.2)$$

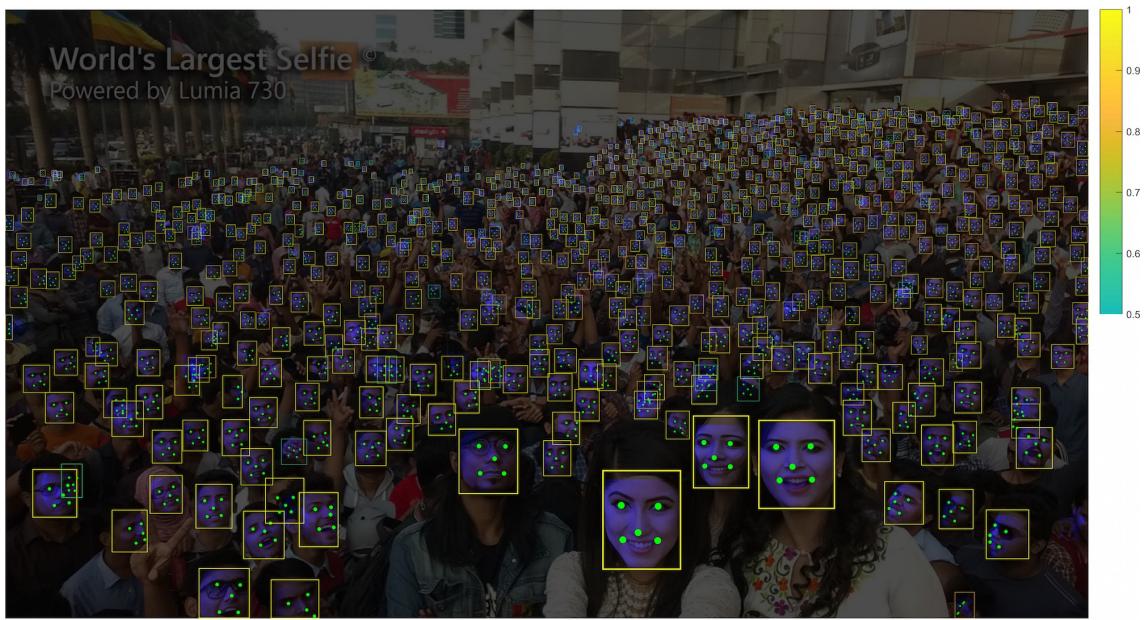
که در آن L_{cls} تابع ضرر مربوط به یافتن یا عدم یافتن چهره می‌باشد. L_{box} تابع ضرر مربوط به مکان چهره می‌باشد. همچنین L_{pts} تابع ضرر مربوط به یافتن نقاط ویژه روی اجزای چهره می‌باشد، و L_{pixel} تابع ضرر مربوط به یافتن یک مدل سه بعدی مبتنی بر مشاهده می‌باشد. استفاده از تابع ضرر مبتنی بر یادگیری چندکاره، کمک می‌نماید تا فضای مسئله محدود تر شود و الگوریتم از روی چهره می‌باشد. استفاده از تابع ضرر مبتنی بر یادگیری چندکاره، کمک می‌نماید تا فضای مسئله محدود تر شود و الگوریتم بهینه سازی مورد نظر زودتر به سمت نقطه بهینه همگرا شود. شمای کلی این روش را در شکل ۹.۲ مشاهده می‌کنید.



شکل ۹.۲: نمای کلی روش یافتن چهره تک مرحله‌ای و متراکم. RetinaFace بر اساس هرم ویژگی طراحی شده است. این رویکرد از تابع ضرر چندکاره استفاده می‌نماید. [۱۰].

نمونه‌ای از خروجی این روش را در شکل ۱۰.۲ مشاهده می‌کنید.

¹Multi Task Learning



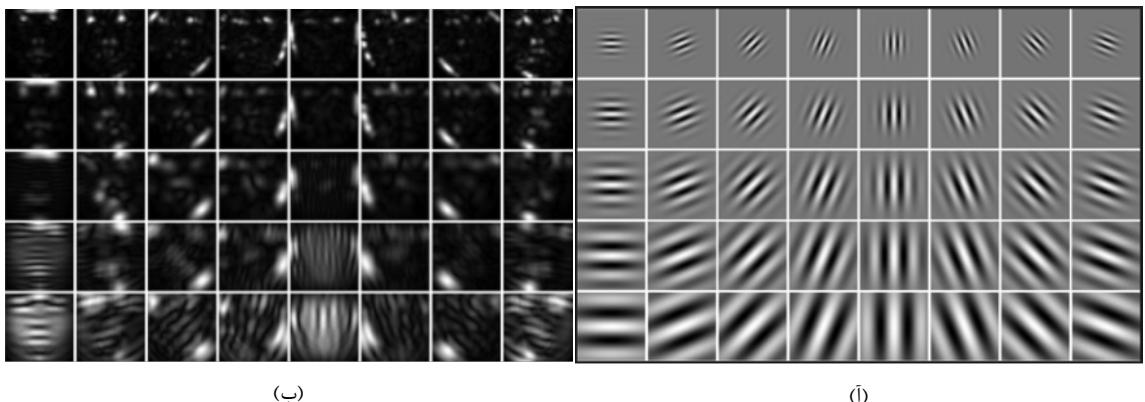
شکل ۱۰.۲: نمونه از خروجی الگوریتم یافتن چهره retina [۱۰]

۳.۲ شناسایی چهره

شناسایی چهره در دو مرحله انجام می‌شود. مرحله اول استخراج ویژگی و مرحله دوم، طبقه‌بندی است. الگوریتم‌های شناسایی چهره را می‌توان به دو دسته اصلی تقسیم کرد. الگوریتم‌های هندسی که بر مبنای استخراج ویژگی‌های متمایز چهره‌ها کار می‌کنند، و الگوریتم‌های تصویری که تصویر را تبدیل به یک الگو می‌نماید و الگوها را مقایسه می‌نماید. رویکردهای مختلفی برای شناسایی چهره طراحی شده است که در ادامه مهم‌ترین آن‌ها آمده است.

۱.۳.۲ رویکردهای سنتی

این رویکردها ویژگی‌های چهره را با علامت‌ها و اندازه‌ها از تصویر استخراج می‌کنند. برای مثال موقعیت نسبی، اندازه و یا شکل چشم‌ها، بینی، گونه‌ها و فک را محاسبه کرده و تجزیه و تحلیل می‌کنند. سپس از این ویژگی‌ها برای جستجوی تصاویر دیگر در پایگاه داده استفاده می‌کنند. در سال ۱۹۹۳ Roberto Brunelli و همکاران [۴۲] یکی از اولین الگوریتم‌ها در این زمینه را ارائه دادند که رویکرد موفقی مبتنی بر روش‌های تطبیق الگو داشت. این رویکرد در شرایط کنترل شده به دقت ۹۰٪ رسید.



شکل ۱۱.۲: (آ) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۱۱].

۲.۳.۲ رویکرد مبتنی بر فیلتر گابور

در سال ۲۰۰۷ F. Abate و همکاران [۱۱] یک رویکرد مبتنی بر فیلتر گابور ارائه دادند. در این رویکرد ابتدا تصویر را بخش بندی^۱ کرده، سپس بر روی بخش‌های مختلف آن، فیلتر گابور^۲ اعمال می‌شود و نتیجه بدست آمده با یک طرح از پیش آماده شده، با یک آستانه گذاری مطابقت داده می‌شود. شکل ۱۱.۲ فیلترهای چندگانه گابور و تاثیر این فیلترها بر روی تصویر چهره انسان را نشان می‌دهد. دلیل استفاده از فیلتر گابور این است که عملکرد این فیلتر به سامانه بصری انسان بسیار شباهت دارد.

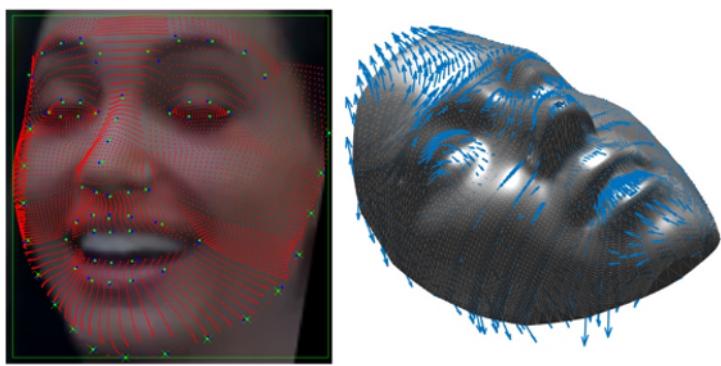
۳.۳.۲ رویکردهای سه بعدی

در سال ۲۰۰۹ JafriRabia و همکاران [۱۲] نشان داده‌اند که داده‌های سه بعدی دقیق ت تشخیص چهره را به شدت بهبود می‌بخشد. روش تشخیص سه بعدی چهره از یک منتشر کننده نور فرو سرخ و یک حسگر به عنوان دریافت کننده استفاده می‌کند. شبکه‌ای از نورهای فرو سرخ که برای انسان قابل رویت نیست، روی چهره تابانده می‌شود. سپس یک حسگر ویژه، پرتوهای بازتاب را دریافت کرده و اطلاعات عمق تصویر پردازش می‌شود. این دسته از الگوریتم‌ها برای شناسایی دقیق اشخاص، برای هر نفر برداری‌های سه بعدی می‌سازند. عیب رویکردهای سه بعدی، نیاز به تجهیزات پیشرفته و غیر قابل استفاده بودن در شرایط کنترل نشده مانند خیابان و معابر پیاده می‌باشد. شکل ۱۲.۲ یک مدل سازی سه بعدی چهره با اشعه فروسرخ را نشان می‌دهد.

در سال ۲۰۱۷ Guosheng Hu و همکاران [۴۳] یک روش سه بعدی برای توصیف ویژگی‌های چهره ارائه کردند که در آن از مدل سازی سه بعدی چهره به همراه تجزیه و تحلیل بافت پوست استفاده شده است. یکی دیگر از رویکردهای در حال ظهور، استفاده از بافت

¹Grid

²Gabor



شکل ۱۲.۲: مدل سازی سه بعدی چهره با اشعه فرو سرخ [۱۲].

پوست برای شناسایی چهره می‌باشد که خطوط، الگوها و لکه‌های پوست را به یک فضای ریاضی تبدیل می‌کند. تجزیه و تحلیل بافت بسیار شبیه روش شناسایی چهره است. تصویری از پوست گرفته می‌شود و به بخش‌های کوچکتر تقسیم می‌شود. سپس هر بخش به یک فضای ریاضی قابل اندازه‌گیری تبدیل می‌شود و خطوط، منافذ و بافت پوست تشخیص داده می‌شود. این رویکرد می‌تواند تفاوت بین دوقلوهای یکسان را شناسایی کند که با استفاده از تشخیص چهره به تنهایی امکان پذیر نیست. آزمایش‌ها نشان دادند که با افزودن تحلیل بافت پوست، عملکرد سامانه تشخیص چهره می‌تواند ۲۰ تا ۲۵ درصد افزایش یابد. این سامانه شایستگی استفاده در کاربردهای مختلف امنیتی و نظامی با شناسایی خودکار سریع و بدون دخالت شخص را دارد و سرعت پردازش را بالا و خطرا کاهش داده است. برتری روش سه بعدی در عدم وابستگی به حرکت و جا به جایی صورت است. انتقال و نصب سامانه تصویربرداری بسیار ساده است. زاویه دید حسگر چندان مهم نیست. همچنین نورپردازی نامناسب تاثیری در این شیوه ندارد و عملیات آن ساده است. برخلاف روش تشخیص دو بعدی، روش سه بعدی و تجزیه و تحلیل بافت پوست به تجهیزات بسیار پیچیده تری نیاز دارد، و با توجه به آنکه تمرکز ما بر روی تشخیص چهره به صورت بی‌درنگ در شرایط کنترل نشده مانند معابر پیاده و خیابان می‌باشد، به توضیح مختصر رویکردهای سه بعدی و تجزیه و تحلیل بافت پوست بسنده می‌کنیم.

۴.۳.۲ رویکردهای مبتنی بر دوربین حرارتی

در این رویکرد، دوربین حرارتی شکل صورت را تشخیص می‌دهد و از لوازم جانبی مانند عینک، کلاه یا آرایش چشم پوشی می‌کند. برخلاف دوربین‌های معمولی، دوربین‌های حرارتی می‌توانند تصاویر را حتی در شرایط کم نور مانند شب، بدون استفاده از فلاش و قرار گرفتن در معرض مستقیم دوربین ضبط کنند. با این حال، یکی از مشکلهای استفاده از تصاویر حرارتی برای تشخیص چهره این است که مجموعه داده‌های آن برای شناسایی چهره محدود است.

در سال ۲۰۰۳ Diego Socolinsky و همکاران [۴۴] از شناسایی چهره مبتنی بر دوربین حرارتی در کاربردهای واقعی بهره برداری کردند و یک مجموعه داده جدید از تصاویر حرارتی چهره ایجاد کردند. آنها از حسگرهای الکتریکی فروسرخ با حساسیت کم و با توانایی جذب حرارت طولانی مدت یا LWIR^۱ استفاده کردند. نتایج نشان می‌دهد که تلفیق LWIR و دوربین‌های معمولی، نتایج بهتری در شرایط کنترل نشده دارد. در این مطالعه ۲۴۰ چهره مجزا در طی ۱۰ هفته برای ایجاد پایگاه داده جدید استفاده شده است. داده‌ها در روزهای آفتابی، بارانی و ابری جمع آوری شد. در شرایط کنترل شده دوربین معمولی دقیق ۹۷٪ درصد دارد، در حالی که روش LWIR دارای دقیق ۹۳٪ می‌باشد و ترکیب این دو دارای دقیق ۹۸٪ است. در شرایط کنترل نشده دوربین معمولی دقیق ۶۷٪، دوربین LWIR دقیق ۸۳٪ و ترکیب این دو دارای دقیق ۸۹٪ است.

۵.۳.۲ تشخیص چهره مبتنی بر ویدیو

در سال ۲۰۰۹ Wang و همکاران [۱۳] یک برآورد کلی از رویکردهای مبتنی بر ویدیو ارائه دادند. تشخیص چهره در ویدیو در طی چند سال گذشته مورد توجه قرار گرفته و طیف گسترده‌ای از برنامه‌های کاربردی تجاری و اجرای قانون را در بر گرفته است. فیلم‌ها قادر به ارائه اطلاعات بیشتر نسبت به تصاویر ثابت هستند. مزایای عمدۀ استفاده از ویدیو عبارتند از:

۱. امکان استفاده از افزونگی موجود در توالی ویدیو برای بهبود عملکرد تشخیص نسبت به تصاویر ثابت وجود دارد. تشخیص

چهره و پیگیری آن در طول زمان، موجب انتخاب فریم‌های خوب می‌شود که حاوی چهره‌های رو به رو یا نشانه‌های ارزشمند است که شرایط نور، انسداد، حالت چهره و... در آن رضایت بخش می‌باشد.

۲. نمایش‌های موثرتر مانند مدل چهره سه بعدی یا تصاویر super resolution می‌توانند از اطلاعات فریم‌های ویدیو گرفته

شده برای بهبود شناخت استفاده کنند.

۳. یادگیری و به روز رسانی مدل در طول زمان امکان پذیر می‌باشد.

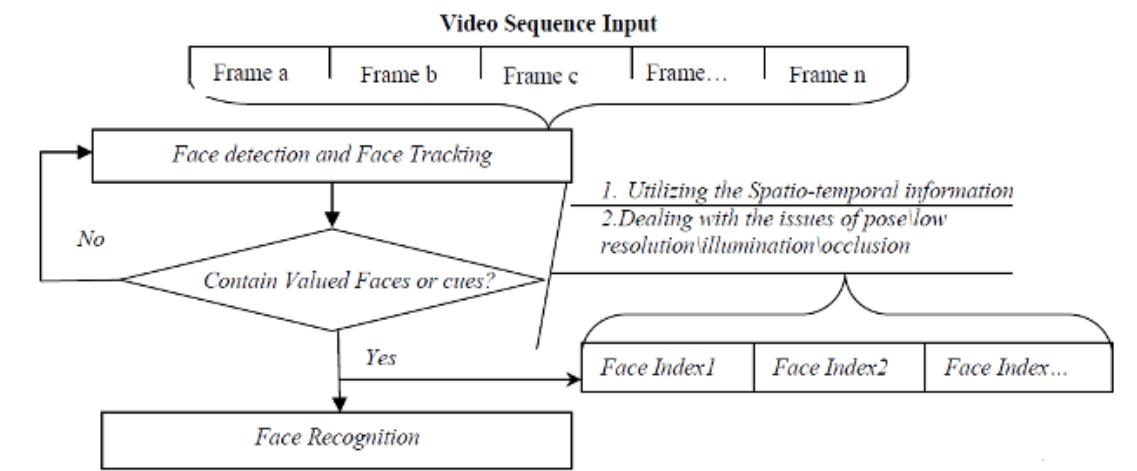
برای تشخیص چهره در تصاویر ویدیویی، دو رویکرد کلی وجود دارد:

مبتنی بر قاب^۲: در این رویکرد برای شناسایی چهره، هر قاب به صورت جداگانه مورد پردازش قرار می‌گیرد که عیب آن نادیده گرفتن اطلاعات زمانی ارائه شده توسط توالی ویدیویی می‌باشد.

¹Longwave Infrared

²Frame

یافتن و ردیابی^۱ : یافتن چهره در اولین قاب و سپس ردیابی آن از طریق توالی قاب‌ها. شکل ۱۳.۲ نمای کلی این رویکرد را نشان می‌دهد.



شکل ۱۳.۲: نمای کلی یک سامانه تشخیص چهره مبتنی بر یافتن چهره و ردیابی آن [۱۳].

۶.۳.۲ رویکرد مبتنی بر چهره ویژه

با افزایش حجم داده‌های موجود، نیاز به کاهش ابعاد داده‌ها می‌باشد. تحلیل مؤلفه‌های اساسی یا PCA^۲ یک روش کاهش ابعاد داده‌ها است که در برخی مسئله‌ها مانند پردازش تصویر به خوبی و با سرعت بالا عمل می‌کند. استفاده از این روش در پردازش سریع تر داده‌ها کمک می‌کند و از رخدادن مشکل بیش‌برازش^۳ جلوگیری می‌نماید. اگر یک پایگاه داده عظیم از تصاویر چهره اشخاص با وضوح بالا موجود باشد که هر کدام دارای تعدادی زیاد ویژگی هستند، و بخواهیم یک تصویر آزمایش را با این پایگاه داده مقایسه کرده و شخص شبیه به آن را پیدا کنیم، مقایسه تصاویر بسیار زمان برو در مواردی غیر ممکن خواهد بود. PCA در این مسئله به خوبی عمل می‌کند. با اعمال تکنیک کاهش بعد به تصویر و با به دست آوردن تصویر ویژه چهره‌ها^۴ می‌توان ویژگی‌ها را کاهش داد و نتیجه مطلوب را در زمان بسیار کم گرفت. شکل ۱۴.۲ تعدادی چهره و چهره‌های ویژه متناظر با آن‌ها را نشان می‌دهد.

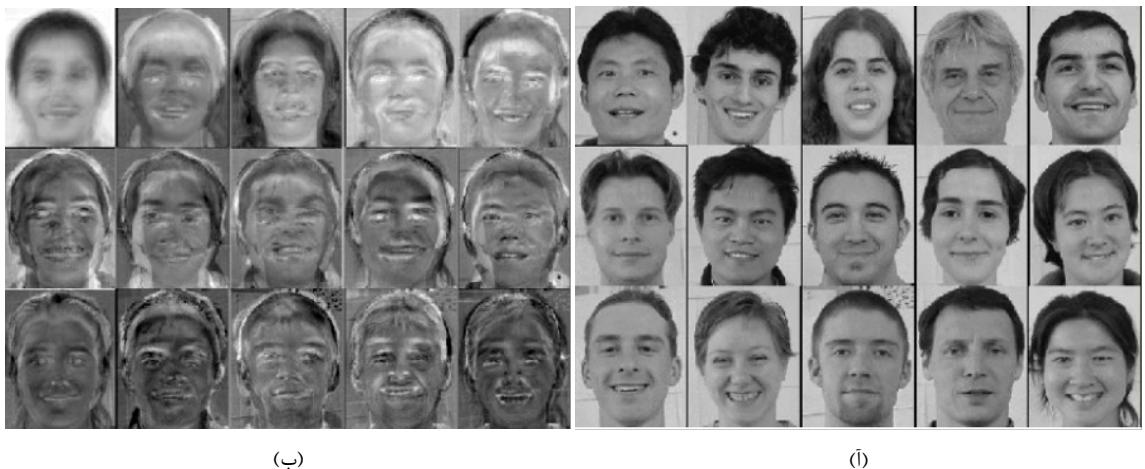
در سال ۲۰۱۴ Xiao Luan^۵ و همکاران [۱۴] یک روش تشخیص چهره مبتنی بر PCA ارائه دادند که تا حدی در برابر تغییرات نورپردازی و انسداد مقاوم می‌باشد. PCA راستای بیشترین تغییرات را با توجه به تعداد ویژگی‌ها و نوع آن‌ها به ما می‌دهد. به همین دلیل در برخی موارد که تنها محوری که بیشترین تغییرات یا پراکندگی را دارد برای ما مهم است، راه حل مناسبی خواهد بود.

¹Detection And Tracking

²Principle Component Analysis

³Overfitting

⁴Eigenface



(ب)

(ا)

شکل ۱۴.۲: (ا) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۱۴].

در سال ۲۰۱۶ K. R. Sreelakshmi و همکاران [۴۵] یک روش شناسایی چهره مبتنی بر چهره های ویژه ارائه دادند که در آن ابتدا تصویر ورودی با استفاده از ماتریس بردارهای ویژه، به فضای دیگری منتقل می شود، سپس در فضای کاهش بعد یافته با داده های موجود مقایسه شده و شبیه ترین تصویر به آن انتخاب می شود. برای مقایسه از معیارهایی مانند معیار اقلیدسی و منهتن می توان استفاده کرد. از مزایای این روش می توان به سهولت پیاده سازی و استفاده، کاهش حجم داده ها و سرعت بالا اشاره کرد. در نظر نگرفتن پراکندگی درون کلاسی و بین کلاسی داده ها، عدم توجه به برچسب تصویر برای شناسایی، تمایز قابل نشدن بین تصاویر مختلف یک شخص در پایگاه و نیاز به بروز رسانی تمامی اطلاعات موجود با ورود یک تصویر جدید به پایگاه داده از معایب این روش است. محاسبات ریاضی و مراحل انجام آن ها:

۱. تبدیل ماتریس تصاویر به بردار و کنار هم قرار دادن آن ها برای تشکیل ماتریس داده ها

۲. محاسبه میانگین ماتریس بدست آمده و انتقال داده ها به مرکزیت صفر

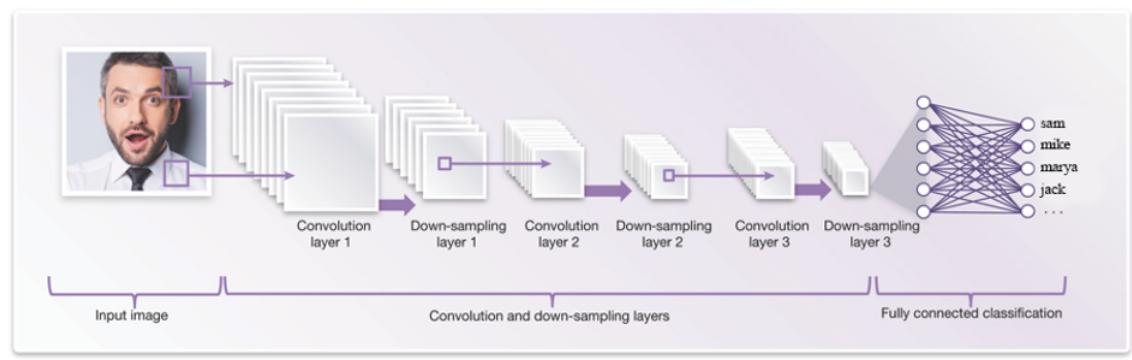
۳. محاسبه ماتریس کوواریانس بردارها و مقادیر ویژه آن

۴. انتقال ماتریس داده ها به زیرفضای جدید با استفاده از ماتریس بردارهای ویژه

۵. بررسی شباهت بین بردار منتقل شده و بردارهای موجود و انتخاب شبیه ترین بردار

۷.۳.۲ رویکردهای مبتنی بر یادگیری عمیق

یک راه حل غیر خطی برای شناسایی چهره، استفاده از شبکه عصبی پیچشی است که به طور شگفت انگیزی در طبقه‌بندی تصاویر چهره خوب کار می‌کند و ویژگی‌های ارزشمندی را از تصویر چهره استخراج می‌کند. بنابراین می‌توان از آن در حل مسئله شناسایی چهره و تأیید هویت استفاده کرد. شکل ۱۵.۲ ساختار کلی یک شبکه عصبی عمیق برای شناسایی چهره را نشان می‌دهد.



شکل ۱۵.۲: شبکه عصبی عمیق برای شناسایی چهره [۴۶].

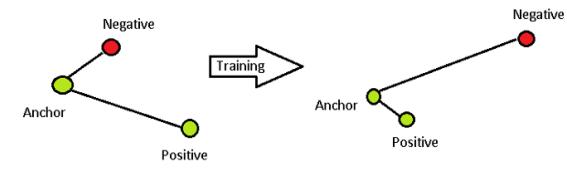
معمولاً به عنوان تابع فعالیت از توابع غیرخطی مانند ReLU^۱ استفاده می‌شود. عملیات بهینه سازی به روش پس انتشار خطا انجام می‌گردد. و در لایه خروجی از تابع SoftMax برای طبقه‌بندی استفاده می‌شود که خروجی‌های لایه آخر را هنجار می‌کند. شبکه‌های عصبی پیچشی ویژگی‌های یک چهره را استخراج می‌کنند که می‌توان به عنوان یک شناسه برای یک فرد خاص در نظر گرفت. هنگامی که دو تصویر مختلف از چهره یک شخص به عنوان ورودی داده می‌شود، شبکه باید خروجی‌های مشابه (ویژگی‌های نزدیک تر) را برای هر دو تصویر تولید نماید، در حالی که برای چهره دو شخص مختلف، شبکه باید خروجی‌های بسیار متفاوت برای دو تصویر تولید نماید. شبکه عصبی نیاز به آموزش دارد تا به طور خودکار ویژگی‌های مختلف چهره‌ها را شناسایی کند و بر اساس آن محاسبات را انجام دهد. در ادامه چند شبکه عصبی پیچشی معروف مورد بررسی قرار گرفته است.

۱.۷.۳.۲ FaceNet شبکه

در سال ۲۰۱۵ Florian Schroff و همکاران^[۱۵] یک شبکه عصبی عمیق به نام FaceNet ارائه دادند. FaceNet یک مدل شبکه است که می‌آموزد چگونه تصاویر چهره را به یک فضای اقلیدسی فشرده نگاشت دهد تا فاصله تصاویر به طور مستقیم با میزان شباهت چهره‌ها مرتبط باشد. هنگامی که این فضا تولید شود، شناسایی چهره، تأیید هویت و خوشه‌بندی می‌تواند به راحتی با استفاده

^۱Rectified Linear Unit

از روش‌های استاندارد توسط FaceNet انجام شود. این شبکه برای آموزش از سه گانه تطبیق - عدم تطبیق استفاده می‌نماید. با توجه به شکل ۱۶.۲، سه گانه تطبیق - عدم تطبیق یک مجموعه از سه تصویر شامل یک تصویر لنگر^۱، یک تصویر منطبق بر تصویر لنگر (مثبت^۲) و یک تصویر غیر منطبق بر تصویر لنگر (منفی^۳) است که باید فاصله بین تصویر لنگر و تصویر مثبت را به حداقل برساند، زیرا هر دو دارای هویت مشابه هستند و فاصله بین تصویر لنگر و تصویر منفی را به حداقل برساند، زیرا این تصاویر دارای هویت متفاوت می‌باشند.



شکل ۱۶.۲: سه گانه تطبیق - عدم تطبیق [۱۵].

برای هر داده آموزشی A مجموعه‌ای از داده‌های مثبت و مجموعه‌ای از داده‌های منفی در نظر گرفته می‌شود. سپس داده‌ها با تابع ضرر سه گانه طوری آموزش می‌بینند که رابطه ۸.۲ برای هر کدام از داده‌های آموزشی برقرار باشد.

$$f(A) - f(P)^2 f(A) - f(N)^2 \quad (8.2)$$

که در آن تابع f ویژگی‌های استخراج شده از تصاویر است. شبکه در مرحله آموزش با استفاده از داده‌های برچسب‌گذاری شده، می‌آموزد فاصله بین ویژگی‌های شبیه به هم، کمتر از فاصله بین ویژگی‌های دور باشد و به این ترتیب در مرحله آزمایش می‌تواند داده‌های مثبت و منفی را به راحتی تفکیک نماید. تابع هزینه در این مدل برای هر نمونه آموزشی x_i به صورت زیر تعریف می‌شود:

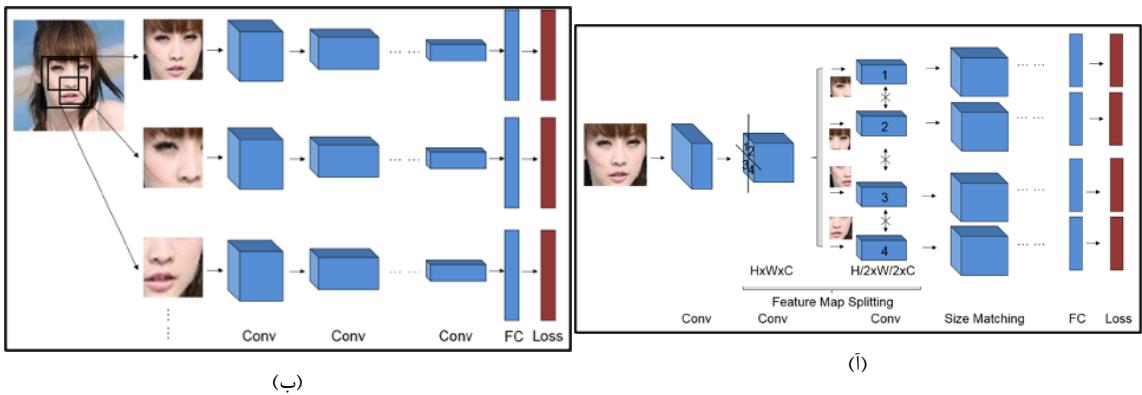
$$L = \sum_{i=1}^n f(x_i^A) - f(x_i^P)^2 - f(x_i^A) - f(x_i^N)^2 + \alpha \quad (9.2)$$

که در آن x_i^P و x_i^N نمونه‌های مثبت و منفی برای نمونه آموزشی x_i^A می‌باشند و α حاشیه بین داده‌های مثبت و منفی را برای هر داده آموزشی مشخص می‌کند. این شبکه در مجموعه داده برچسب دار LFW به دقت جدید ۹۹,۶۳٪ رسیده است و در مجموعه داده YouTube Faces DB دقت آن به ۹۵٪ رسیده است.

¹Anchor

²Positive

³Negative



شکل ۱۷.۲: (آ) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [۱۶].

۲.۷.۳.۲ شبکه SplitNet

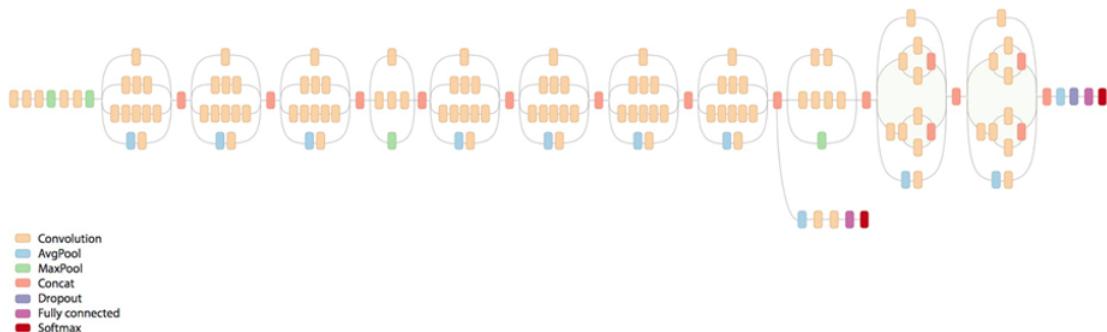
در سال ۲۰۱۸ و همکاران [۱۶] یک شبکه عمیق به نام SplitNet برای شناسایی چهره ارائه دادند. با توجه به ساختار معنایی چهره، یک بخش محلی از تصویر چهره همانند تصویر کلی چهره حاوی ویژگی‌ها و اطلاعات مفیدی برای یادگیری عمیق است. به منظور استفاده همزمان از اطلاعات سراسری و محلی، روش‌های یادگیری عمیق موجود برای شناسایی چهره، چندین شبکه CNN را آموزش می‌دهند و ویژگی‌های مختلف را بر اساس مکان تصاویر محلی ترکیب می‌کنند که نیاز به عملیات متعدد و محاسبات بسیار بیشتری برای هر تصویر دارد. هدف این مقاله بهبود تشخیص چهره تنها با یک عملیات پیشخور^۱ است که به طور همزمان از اطلاعات سراسری و محلی در یک مدل استفاده می‌کند. آن‌ها یک چارچوب یکپارچه به نام SplitNet ارائه دادند که به جای آن که تصویر اصلی را برش دهد، ویژگی‌های میانی را به چندین شاخه تقسیم می‌کند. شکل ۱۷.۲ شبکه عصبی پیچشی SplitNet را در مقابل شبکه عصبی پیچشی معمولی نشان می‌دهد. نتایج تجربی نشان می‌دهد که این رویکرد می‌تواند به طور موثر دقت تشخیص چهره را با محاسبات کمتر افزایش دهد.

۲.۷.۳.۲ شبکه GoogLeNet

در سال ۲۰۱۵ و همکاران [۱۷] یک شبکه عصبی عمیق به نام GoogLeNet ارائه دادند. همانطور که در شکل ۱۸.۲ مشاهده می‌شود، یک شبکه عصبی پیچشی با ۲۲ لایه است که یکی از اولین معماری‌های شبکه عصبی پیچشی بود که از رویکرد کلی قرار دادن تعداد زیادی از لایه‌های پیچشی و رای‌گیری در کنار هم در یک ساختار متوالی بدست آمد. نویسنده‌گان این مقاله همچنین تأکید کردند که این مدل جدید، توجه قابل ملاحظه‌ای به مصرف حافظه و مصرف انرژی دارد.

^۱Feed Forward

زیرا کنار هم چیدن تعداد زیادی لایه و فیلتر دارای هزینه محاسباتی و حافظه است که احتمال بیش برازش را افزایش می‌دهد.



شکل ۱۸.۲: معماری کلی شبکه [۱۷] GoogLeNet.

۴.۷.۳.۲ شبکه VGGFace

در سال ۲۰۱۵ Omkar M. Parkhi و همکاران [۱۸] شبکه عمیق VGGFace را ارائه کردند که شامل یک توالی طولانی از لایه‌های پیچشی می‌باشد. با توجه به شکل ۱۹.۲ این شبکه که در لایه آخر به عنوان یک طبقه‌بند عمل می‌نماید، هر تصویر آموزشی چهره را توسط لایه تماماً متصل وتابع ضرر softmax log-loss به یک بردار تبدیل می‌نماید که هر مقدار در این بردار، نشان دهنده احتمال برای یک هویت فردی است. VGGFace مشابه FaceNet از یک تابع ضرر سه گانه^۱ در آموزش برای بهبود عملکرد کلی استفاده می‌نماید.



شکل ۱۹.۲: معماری شبکه [۱۸] VGGFace.

۵.۷.۳.۲ نقاط راهنمای

چهره‌هایی که در جهت‌های مختلفی هستند، برای سامانه تشخیص چهره، متفاوت به نظر می‌رسند. برای غلبه بر این چالش در رویکردهای مبتنی بر نقاط راهنمای سعی می‌شود تصویر را چرخانده و جایه جا نمود، بطوریکه چشم‌ها و لب‌ها در یک موقعیت خاص در

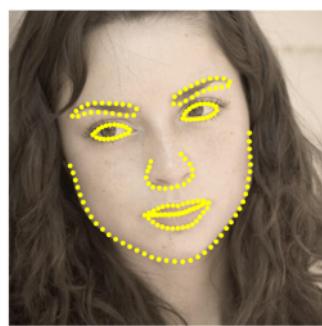
^۱Triplet Loss Function

تصویر قرار بگیرند. بدین ترتیب مقایسه چهره‌ها در مرحله بعد بسیار ساده‌تر خواهد شد.

در سال ۲۰۱۴ وحید کاظمی و جوزفین سالیوان در [۱۹] یک الگوریتم برای یافتن نقاط راهنما^۱ بر روی چهره ارائه دادند که از

نقطه خاص که در هر چهره ای وجود دارد استفاده می‌نماید. شکل ۲.۲۰ مکان این نقاط را بر روی گونه، لبه‌های بیرونی چشم، کناره

ابرو و... نشان می‌دهد. سپس این نقطه به سامانه آموزش داده می‌شود تا در هر چهره‌ای آن‌ها را تشخیص دهد.



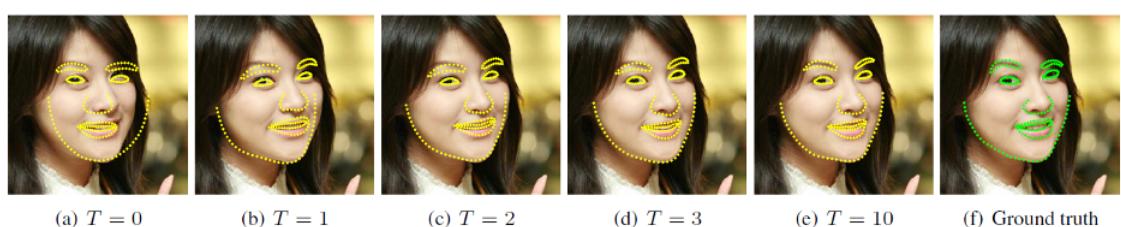
شکل ۲۰.۲: نتیجه موقعیت ۱۹۴ شاخص روی چهره [۱۹].

پس از این که دانستیم چشم‌ها، دهان و... کجاست، به راحتی می‌توانیم تناسب تصویر را تغییر داده و آن را چرخانده یا پرش بزنیم. به

طوری که چشم‌ها و دهان در بهترین حالت ممکن در مرکز قرار گیرد. با استفاده از تغییرات اساسی و اصلی تصویر، مانند تغییر اندازه،

چرخش، خطوط موازی را حفظ می‌کنیم که در ریاضی به آن تغییرات نسبت یا افاین می‌گویند. شکل ۲۱.۲ علامت گذاری تکراری

خطوط راهنما بر روی چهره را نشان می‌دهد.



شکل ۲۱.۲: علامت گذاری خطوط راهنما بر روی چهره که در هر تکرار با کاهش خط‌ها همراه می‌باشد [۱۹].

در سال ۲۰۱۶ Yue Wu و همکاران [۴۷] رویکردی برای یافتن نقاط راهنما بر روی چهره مبتنی بر شبکه عصبی پیچشی ارائه

دادند. در این مقاله یک معماری جدید برای شبکه عصبی پیچشی به نام Tweaked CNN پیشنهاد شده است که به اختصار

TCNN نامیده می‌شود. این شبکه عصبی عمیق از ۴ لایه پیچشی ($CL_1 \dots CL_4$) با لایه‌های رای‌گیری در میان آن‌ها تشکیل

شده است و در انتهای یک لایه تمام متصل FC_5 و پس از آن یک لایه خروجی با اندازه $m * 2$ آمده است که مختصات m نقطه

ویژه را بر روی چهره مشخص می‌کند. در این مقاله m برابر با ۵ در نظر گرفته شده است.تابع فعالیت برای هریک از لایه‌های پیچشی

¹Landmark

۱۰.۲ تابع فعالیت برای لایه تمام متصل ($f(x) = \tanh(x)$) در نظر گرفته شده است. و در نهایت رابطه

به عنوان تابع ضرر معرفی شده است.

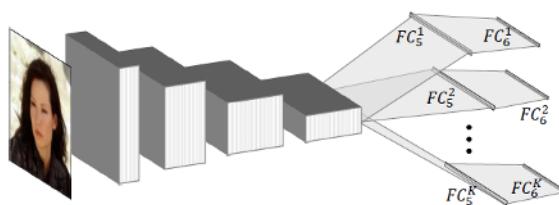
$$L(P_i, \hat{P}_i) = \frac{(P_i - P_{i,2}^2)}{(P_{(i,1)} - P_{(i,2)}^2)^2} \quad (10.2)$$

که در آن P_i یک بردار $m * 2$ برای مختصات پیش بینی شده تصویر I_i و P_i و مختصات محل دقیق آن نقاط می باشد.

مختصات چشم ها در تصویر مرجع می باشند. در نهایت خروجی لایه تمام متصل توسط الگوریتم GMM به ۶۴ خوشه تقسیم شده

و هریک به صورت جداگانه بررسی شده است. معماری TCNN در شکل ۲۲.۲ قابل مشاهده می باشد. این شبکه برای آموزش از

مجموعه داده LFW استفاده کرده است.



شکل ۲۲.۲: شبکه عصبی پیچشی با معماری TCNN.^[۴۷]

OpenFace ۶.۷.۳.۲

در سال ۲۰۱۶، Amos و همکاران^[۲۴] یک روش شناسایی چهره به نام OpenFace ارائه دادند که ویژگی اصلی آن، آموزش شبکه

عصبی عمیق در کمترین زمان و قابلیت اجرا بر روی دستگاه های قابل حمل مانند تلفن همراه با در نظر گرفتن منابع محدود می باشد.

یک تصویر شامل تعدادی چهره به الگوریتم داده می شود. پس از یافتن چهره ها و مجزا کردن^۱ آن ها از یکدیگر، هر چهره به طور

جداگانه مورد پیش پردازش^۲ قرار می گیرد و حجم آن کاهش می باید. کاهش حجم تصویر برای عملکرد مناسب یک طبقه بندی

بهینه بسیار مهم می باشد. تصاویر چهره ها باید هنجارسازی شده و ابعاد آن ها ثابت گردد تا به بخش شناسایی چهره راه یابند. هر

تصویر چهره باید مورد تبدیل قرار بگیرد تا چشم ها، بینی و دهان، در مکان مشخصی قرار گیرند. بدین منظور از یک تبدیل هم نسبی^۳

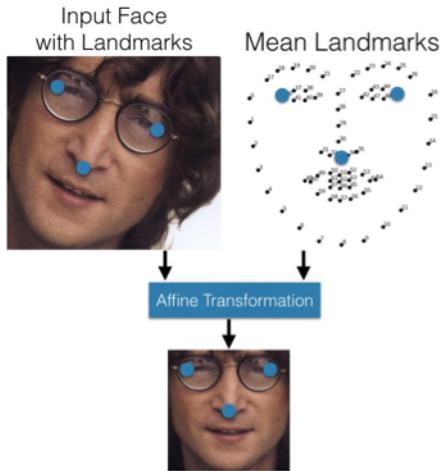
دوبعدی ساده استفاده می گردد. ابتدا باید چهره توسط ۶۸ نقطه ویژه، نشانه گذاری شود. سپس نشانه های اطراف چشم ها و بینی

¹Isolate

²Preprocessing

³Affine Transformation

(شکل ۲۳.۲) برای محاسبه عامل های تبدیل هم نسبی استفاده می شوند. پس از انجام تبدیل هم نسبی، تصاویر چهره برش زده شده و اندازه آن ها 96×96 پیکسل می شود.



شکل ۲۳.۲: تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی [۲۴].

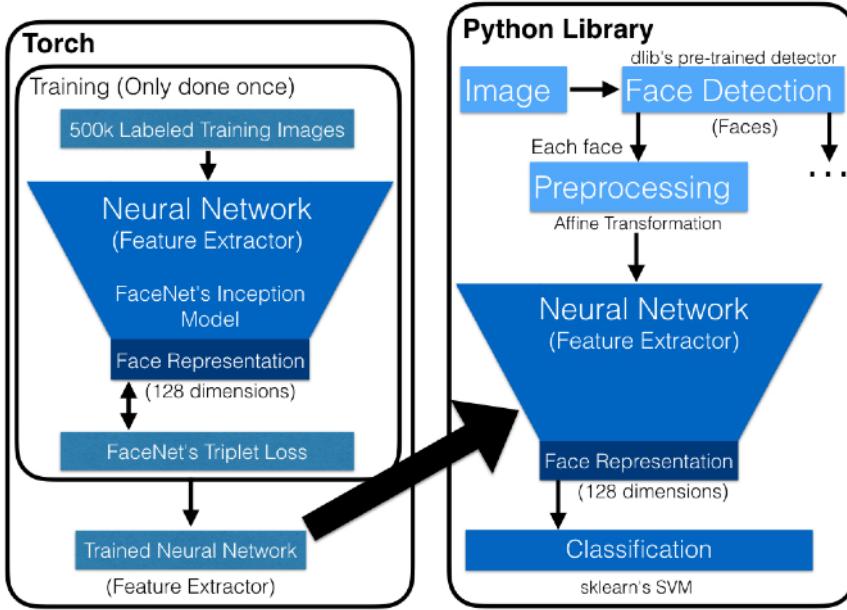
پس از پیش پردازش، تصاویر چهره ها به عنوان ورودی به یک شبکه عصبی پیچشی داده می شوند (شکل ۲۴.۲). این الگوریتم برای تعلیم شبکه از مجموعه داده کوچکی با 500 هزار تصویر چهره استفاده می کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری شده به نام FaceScrub و CASIA-WebFace یک نسخه اصلاح شده از شبکه nn4 الگوریتم FaceNet می باشد. شبکه nn4 مبتنی بر معماری GoogLeNet می باشد. برای تعیین میزان شباهت نتیجه، از فاصله اقلیدسی استفاده شده است.

هر تصویر از یک شبکه یکتا به یک سه گانه نگاشت داده می شود. گرادیان خطای سه گانه برای هر تصویر محاسبه شده و پس انتشار خطا صورت می پذیرد. در هر دسته کوچک^۱، تصویر برای هر نفر از Q نفر، در مجموعه داده انتخاب می شود. سپس $M \approx PQ$ تصویر به شبکه داده می شود تا عملیات رو به جلو^۲ انجام پذیرد. در این مقاله از $P = 20$ و $Q = 15$ استفاده شده است. تمام جفت های لنگر-مشبт برای بدست آوردن سه گانه های $N = Q \binom{P}{2}$ مورد استفاده قرار می گیرند. خطای سه گانه محاسبه شده و مشتق آن برای پس انتشار خطا استفاده می شود. شکل ۲۵.۲ چگونگی آموزش شبکه را نشان می دهد.

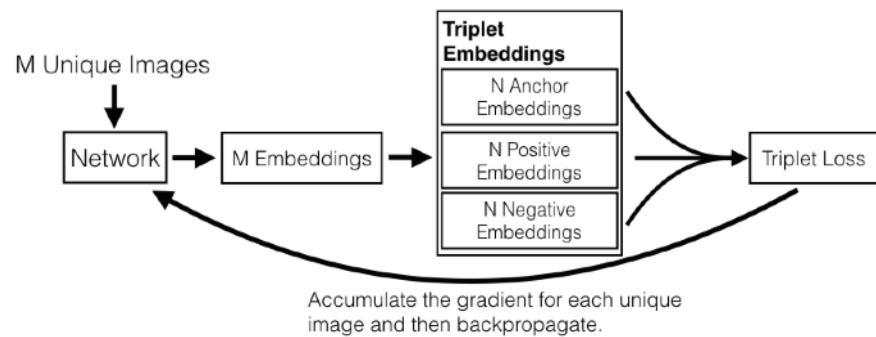
مجموعه داده LFW یک معیار استاندارد برای سنجیدن میزان دقیقیت الگوریتم های تشخیص چهره می باشد. الگوریتم OpenFace بر روی این مجموعه داده مورد سنجش قرار گرفت که به دقیقیت $0.9292 \pm 0.0134\%$ رسید.

¹Mini Batch

²forward



.[۲۴] OpenFace معماری



.[۲۵] OpenFace جریان یادگیری در معماری

۷.۷.۳.۲ شبکه MobileNet

در سال‌های بعد شبکه‌های عمیق‌تر، پهن‌تر و البته پیچیده‌تر مانند ResNet، ResNext و GoogleNet برای رسیدن به دقت بالاتر مطرح شد. با وجود پیچیدگی در طراحی این شبکه‌ها، تمرکز اصلی بر روی دقت بود و جای خالی شبکه‌های با سایز کوچک و سرعت بالا با قابلیت استفاده در رباتیک، بردهای مینی کامپیوتری و البته موبایل‌ها احساس می‌شد که ایده دسته جدیدی از شبکه‌های کانولوشنی سبک با پارامترهای کم شکل گرفت.

یکی از شاخص‌ترین شبکه‌های سبک، شبکه عصبی MobileNet نام دارد که توسط محققان گوگل [۲۰] با هدف طراحی شبکه‌های کارآمد، سبک، سریع و با دقت قابل قبول مطرح شده است. در این مقاله یک نوع کانولوشن جدید به نام کانولوشن^۱ DWS معرفی

^۱Depth-Wise Separable convolution

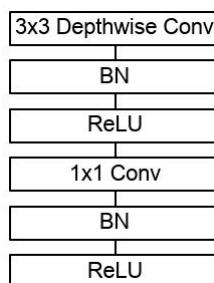
جدول ۱.۲: مقایسه شبکه عصبی موبایل نت با گوگل نت و VGG

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

شد که قلب تپنده شبکه موبایل نت است. در کانولوشن DWS ابتدا کانولوشن عمقی اعمال می‌شود و سپس کانولوشن نقطه‌ای که

به ترتیب نقش مراحل فیلتر و ادغام در کانولوشن استاندارد را دارند.

در کانولوشن استاندارد M کرنل $k \times k$ داشتیم. اما در اینجا تنها یک کرنل $k \times k$ داریم. با این کرنل، مرحله اول کانولوشن را انجام می‌دهیم. با انجام عمل فیلتر، هر صفحه از نقشه ویژگی ورودی F کانوالومی شود. به این مرحله کانولوشن عمقی گفته می‌شود. مرحله دوم، کانولوشن نقطه‌ای^۱ است. این مرحله معادل با مرحله ادغام در کانولوشن استاندارد است. اما یک تفاوت اساسی بین مرحله ادغام در کانولوشن استاندارد و کانولوشن DWS این است که مرحله ادغام در کانولوشن استاندارد، یک جمع ساده است، اما مرحله ادغام در کانولوشن DWS شامل یک کانولوشن 1×1 است. کانولوشن نقطه‌ای همان کانولوشن استاندارد یا رایجی هست که می‌شناسیم و در بسیاری از شبکه‌های کانولوشنی استفاده می‌شود. این کانولوشن 1×1 وظیفه مهمی دارد و خروجی‌های کانولوشن عمقی (مرحله اول) را با هم ادغام می‌کند. در مرحله قبل بجای تعریف M کرنل، تنها یک کرنل تعریف کردیم. اما در این مرحله، M کرنل 1×1 تعریف می‌کنیم.



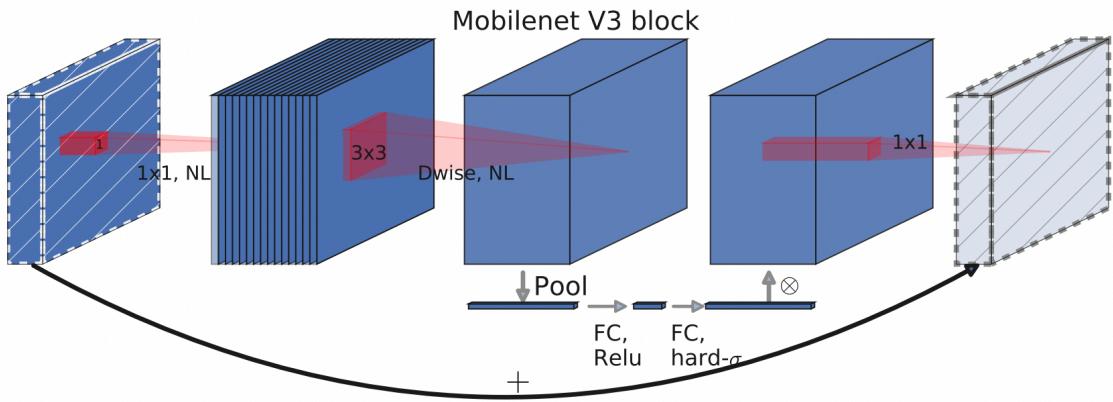
شکل ۲۶.۲: کانولوشن DWS که شامل دو کانولوشن point-wise و depth-wise است. [۲۰]

شبکه عصبی موبایل نت ۲۰.۴ میلیون پارامتر دارد. وقتی تعداد پارامترهای این شبکه را با شبکه محبوب ResNet-۱۸ با ۱۱ میلیون پارامتر مقایسه کنیم، متوجه می‌شویم که چقدر میزان پارامترها کمتر است. این مقایسه در جدول ۱.۲ قابل مشاهده می‌باشد.

در سال ۲۰۱۹ Andrew Howard و همکاران [۴۸] معماری MobileNet نسخه ۳ را ارائه دادند. در این معماری مسیر استخراج

¹point-wise

ویژگی از ۴ لایه کانولوشن، ۱ لایه رای گیری، ۱۵ لایه تنگنا^۱ و ۸ مازول توجه^۲ تشکیل شده است، تصویر ورودی به بخش استخراج ویژگی داده می‌شود و مدل در این مسیر به طور خودکار یک سلسله مراتب ویژگی را از تصاویر ورودی آموزش خواهد دید و در نهایت این ویژگی‌های استخراج شده به عنوان ورودی دسته بند مورد استفاده قرار می‌گیرد. ایده اصلی این مقاله استفاده از بلاک‌های Squeeze-and-Excite می‌باشد. عماری MobileNet نسخه در شکل ۲۷.۲ قابل مشاهده می‌باشد.



[۲۱] شکل ۲۷.۲: لایه های MobileNet به همراه مازول های Squeeze-and-Excite

۸.۷.۳.۲ لایه توجه

در سال ۲۰۲۱ yang و همکاران [۲۱] یک مازول مبتنی بر لایه توجه ارائه دادند. لایه‌های توجه که یک شبکه عصبی را قادر می‌سازد تا دقیقاً بر روی عناصر مهم مربوط به ورودی مرکز شود، به یک جز اساسی برای بهبود عملکرد شبکه‌های عصبی عمیق تبدیل شده است. عمدتاً دو مکانیسم توجه به طور گسترشده در بینایی رایانه مورد استفاده قرار می‌گیرد: لایه توجه وابسته به کanal^۳ و لایه توجه وابسته به موقعیت^۴ که به ترتیب به منظور توجه به رابطه دو به دو در سطح کanal و در سطح پیکسل هستند. اگرچه تلفیق آن‌ها ممکن است عملکرد بهتری نسبت به پیاده سازی‌های منفرد آنها به دست آورد، اما سربار محاسباتی را افزایش می‌دهد.

در این مقاله، یک مازول Shuffle Attention کارآمد برای کاهش بار محاسباتی پیشنهاد شده است که آن را به اختصار مازول SA می‌نامیم. ابتدا ابعاد کanal را به چندین ویژگی فرعی قبل از پردازش موازی آنها تقسیم می‌کند. سپس، برای هر زیرویژگی از یک واحد Shuffle برای به تصویر کشیدن وابستگی‌های ویژگی در هر دو بعد مکانی و کanal استفاده می‌کند. پس از آن، همه زیرویژگی‌ها جمع

¹Bottleneck

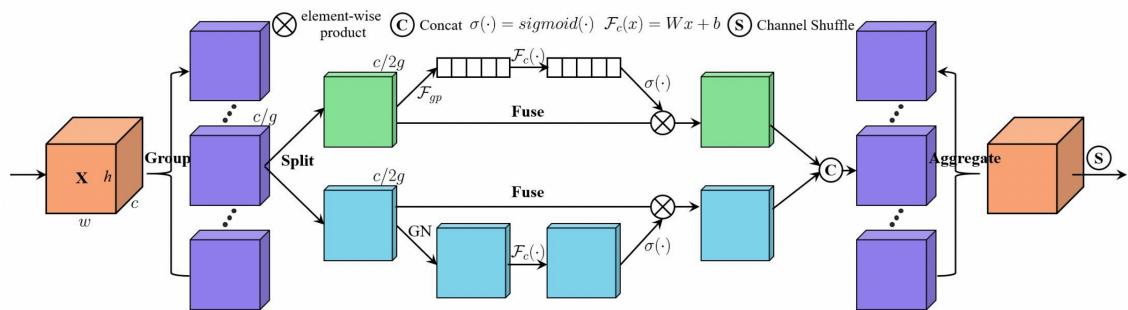
²Attention

³Channel Attention Module

⁴Spatial Attention Module

می‌شوند و یک عملگر تغییر کانال برای امكان برقراری ارتباط اطلاعاتی بین ویژگی‌های فرعی مختلف به کار گرفته می‌شود. معماری

این مازول در شکل ۲۸.۲ آمده است.



شکل ۲۱: معماری مازول [21].

پارامترها و محاسبات SA در شبکه ResNet50 به ترتیب ۳۰۰ میلیون و ۵۶.۲۵ میلیون است، اما افزایش عملکرد بیش از ۳۴٪ را

به ارمغان می‌آورد. نتایج تجربی نشان می‌دهد که SA برای دستیابی به دقت بالاتر مناسب است، در حالی که دارای پیچیدگی مدل

کمتری است و از روش‌های جدید^۱ کنونی به مرتب بهتر عمل می‌کند. پس مازول SA بهینه و در عین حال کارآمد است.

۴.۲ نتیجه‌گیری

در این فصل مفاهیم پایه در مبحث یافتن و تشخیص چهره در تصاویر و انواع الگوریتم‌های دسته‌بندی از روی تصاویر رنگی بررسی

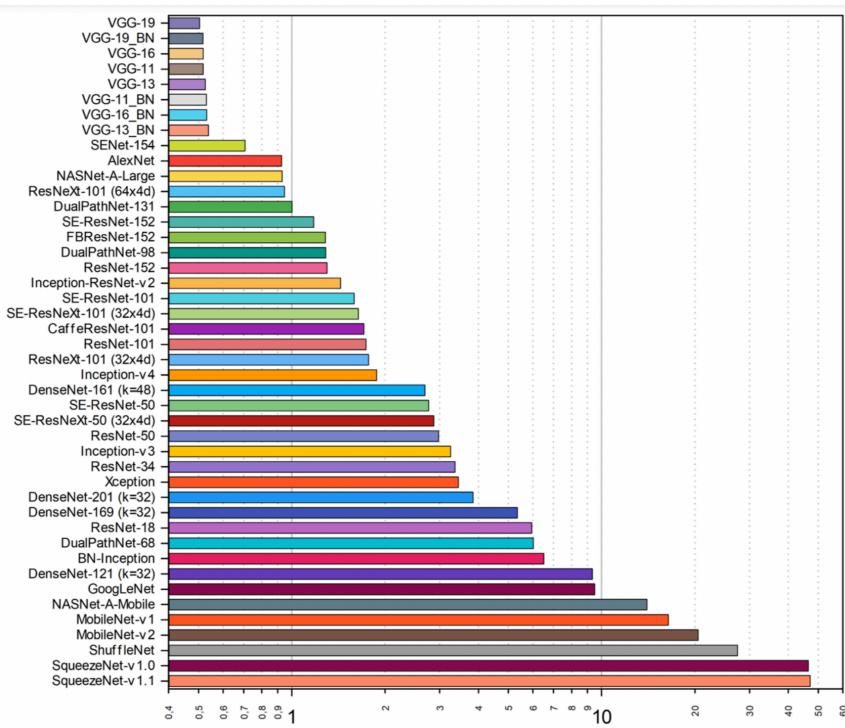
شد. با بررسی شبکه‌های بهروز از جمله MobileNet ، SqueezeNet ، VGG19 ، ResNet-50 ، EfficientNetB0 و

MobileNetV2 و مقایسه دقت و زمان پاسخگویی آن‌ها، به این نتیجه می‌رسیم که شبکه‌های SqueezeNet و MobileNetV2

دارای چگالی دقت بالاتری می‌باشند. بدین معنا که نسبت دقت دسته‌بندی به تعداد پارامترهای شبکه در آن‌ها بیشتر می‌باشد. بنابرین

می‌توان سرعت اجرای مناسب و همچنین دقت مناسب را از آن‌ها انتظار داشت. نتایج بررسی در شکل ۲۹.۲ آمده‌اند.

¹State of the Art



شکل ۲۹.۲: مقایسه چگالی دقت در معماری‌های مختلف شبکه عصبی پیچشی [۲۲].

فصل ۳

مروری بر روش‌های ارائه شده برای مواجهه با شرایط کنترل نشده

در سال های اخیر روش های تشخیص چهره بسیار زیادی به منظور یافتن و شناسایی چهره افراد در تصویر پیشنهاد شده است که توانایی مقاومت در برابر مشکلات و چالش های رایج مانند تغییرات شدید روشنایی، تغییر حالت و زاویه چهره، انسداد، تاری خارج از تمکر، سالخوردگی و... را ندارند و در کاربردهایی نظری شرایط کنترل نشده قابل استفاده نیستند. در بخش مقدمه در مورد چالش های موجود در فرایند تشخیص چهره صحبت شد. برای رفع این چالش ها و بهبود طبقه بندی، راه حل هایی پیشنهاد شده است که در این بخش مورد بررسی قرار گرفته اند. خلاصه ای از روش های مقابله با شرایط کنترل نشده در جدول ۱.۳ آمده است.

جدول ۱.۳: خلاصه ای از روش های مقابله با شرایط کنترل نشده

مقاله ها	چالش مورد نظر	رویکرد	مزیت ها	مشکل ها
[۴۹، ۲۵، ۲۴، ۲]	حالات چهره	تبدیل دو بعدی	پیچیدگی محاسباتی قابل قبول	استخراج نقاط ویژه باید دقیق تر باشد
[۵۲، ۵۰، ۴۷، ۲۶]	حالات چهره	استفاده از شبکه عصبی عمیق	دققت بالا در شرایط کنترل نشده	پیچیدگی محاسباتی، وابستگی به داده های آموزش
[۵۶، ۵۲، ۴۳]	حالات چهره	تبدیل مدل دو بعدی به سه بعدی	دققت بالا پیچیدگی محاسباتی	تبدیل مدل سه بعدی به سه بعدی
[۵۷]	حالات چهره	تبدیل مدل سه بعدی به دو بعدی	دققت بالا پیچیدگی محاسباتی	همسان سازی بافت-نگار
[۵۸، ۰۲]	روشنایی	پیچیدگی محاسباتی قابل قبول	قابل استفاده در تصاویر خاکستری	استفاده از روش های شناسایی الگو و واپیش
[۵۹، ۲۸]	انسداد	دققت بالا در انسداد شدید	انسداد	
[۴۹]	محدودیت داده	تهیه مجموعه داده با دیدایی چهره در ویدیو	نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو	
[۳۵، ۳۴]	محدودیت منابع	استفاده از رایانش ابری	سرعت بالا	تجهیزات پیشرفته و زمان تاخیر ناهمگن

در ادامه یه شرح بیشتر این مشکلات و روش های ارایه شده مواجهه با آن می پردازیم.

۲.۳ چالش حالت

چالش حالت زمانی پیش می آید که چهره فرد کاملا رو به روی دوربین قرار نگیرد و دارای زاویه زیادی باشد. در این شرایط با توجه به ساختار سه بعدی چهره، ممکن است سامانه نتواند ویژگی های درستی از چهره استخراج نماید و در تشخیص هویت دچار اشتباه شود. گرچه شبکه عصبی پیچشی توانایی مقابله با این چالش را از طریق استفاده از مجموعه داده های بزرگ و آموزش تصاویر مختلف از حالات چهره دارد، اما این کار باعث بزرگ شدن پایگاه داده و کند شدن سامانه می شود. استفاده از روش های ابتکاری^۱ به منظور کاهش حجم داده های آموزش می تواند نتایج بهتری به دنبال داشته باشد. یکی از راه حل های مقابله با این چالش، هنجار سازی، رو به رو سازی^۲ و هم ترازی^۳ چهره می باشد. در اینجا برخی رویکردهای رو به رو سازی و هم ترازی چهره در شرایط کنترل نشده را به صورت خلاصه دسته بندی می کنیم. سپس در ادامه شرح داده خواهند شد.

¹Heuristic

²Frontalization

³Alignment

۱. رویکردهای دو بعدی با پیچیدگی محاسباتی قابل قبول (بیشتر ایده های مبتنی بر نشانه گذاری^۱ قدیمی) مانند [۲۴، ۲]،

[۴۹، ۲۵]

مشکل: در محیط های بدون محدودیت مبتنی آنچه که در این پروژه داریم، استخراج دقیق مکان نشانه های صورت از تصاویر دو بعدی نیاز به توجه بیشتری دارد. پیشرفت های اخیر مانند [۲۵] است.

مزیت: این الگوریتم ها از نظر پیچیدگی محاسباتی^۲ قابل قبول هستند و کاملا برای شرایط این پروژه مناسب می باشند.

۲. رویکردهای مبتنی بر شبکه عصبی برای تخمین و اصلاح موقعیت چهره (آموزش و آزمایش با تصاویر دو بعدی) مانند [۲۶]،

[۴۷، ۵۰-۵۲]

مشکل: این الگوریتم ها، به طور متوسط، کندر از دسته پیشین می باشند. اما بستگی به این دارد که عمق شبکه عصبی چه مقدار باشد. وابستگی آن ها به داده های آموزش می باشد و مراحل مجازی رو به رو سازی و هم ترازی چهره ندارند.

مزیت: بدون نیاز به تصمیم گیری در مورد مجموعه بهینه ای از نشانه های چهره و دارای دقت بیشتر در شرایط کنترل نشده با انسداد و... . ایده هایی مانند [۶۰] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند.

۳. رویکردهای سیک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۴۳، ۵۳-۵۶].

مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده ترین این الگوریتم ها در مورد پیچیدگی محاسباتی، [۵۳] است که در شرایط بدون محدودیت آموزش دیده و آزمایش شده است. شامل مراحل مجازی رو به رو سازی و تراز بندی چهره می باشند. اما برای محدودیت های این پروژه قابل استفاده نمی باشند.

مزیت: با استفاده از اطلاعات سه بعدی، این روش ها به بالاترین دقت تصمیم گیری در میان سه نفر رسید.

۴. رویکردهای تبدیل مدل سه بعدی چهره به مدل دو بعدی چهره (روش های مبتنی بر پنجره^۳ بر اساس چند نمایش دو بعدی

مختلف از چهره) مانند [۵۷] مشکل: زمان محاسباتی (نه به اندازه الگوریتم های دسته سوم).

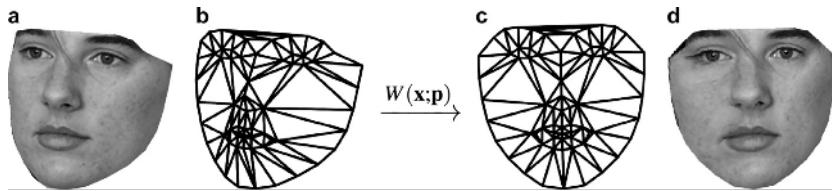
مزیت: عملکرد بهتر در رو به رو سازی چهره در شرایط کنترل نشده نسبت به الگوریتم های دسته اول. ممکن است برای شرایط این پروژه مناسب باشند.

¹Landmark

²Computational Complexity

³Patch-Based

در سال ۲۰۱۶ محمد حقیقت و همکاران [۲۵] یک روش برای هنجارسازی حالت چهره بر اساس تنظیم کردن مدل ظاهری فعال^۱ یا AAM ارائه دادند. یک مدل پارامتری است که برای ارائه یک شکل مانند چهره انسان استفاده می‌شود. در این الگوریتم ابتدا یک AAM بر روی تصویر چهره قرار گرفته، با روندی تکراری و به صورت بهینه شونده، بر روی چهره تنظیم می‌شود. سپس با استفاده از یک تبدیل هم نسبی، مرحله رو به رو سازی بر روی چهره انجام می‌پذیرد. شکل ۱.۳ رویکرد کلی این الگوریتم را نشان می‌دهد.



شکل ۱.۳: رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲۵].

در این مدل یک تصویر چهره با مجموعه ای از نقاط ویژه هنجارسازی شده مدل می‌شود که به صورت $[x_i, y_i]$ تعریف می‌شود که در آن $i = 1, 2, \dots, n$. برای انجام این کار یک مرحله یادگیری نیاز است. سپس الگوریتم PCA اعمال می‌شود تا کاهش میزان وابستگی میان نقاط ویژه در هر مجموعه انجام می‌شود و نتیجه یک مدل خطی است که یک مدل شکل نمونه را به صورت رابطه ۱.۳ نمایش می‌دهد.

$$S = s_0 + \sum_{i=1}^n p_i s_i \quad (1.3)$$

که در آن s_0 شکل پایه، s_i نشان دهنده i امین شکل پایه و $[p_1, p_2, \dots, p_n]$ عامل‌های شکل می‌باشند. ظاهر^۲ مدل AAM یک تصویر (x) می‌باشد که در آن X مجموعه پیکسل‌های داخل شکل پایه s_0 می‌باشد. مدل ظاهر یک چهره خاص از یک ظاهر پایه a_0 و ترکیب خطی از بردارهای ویژه a_i ، $i = 1, 2, \dots, m$ تشکیل می‌شود که به صورت رابطه ۲.۳ تعریف می‌گردد.

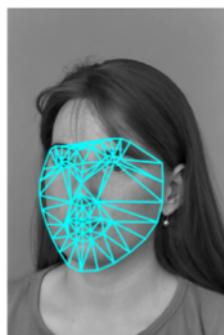
$$A(x) = a_0(x) + \sum_{i=1}^m q_i a_i(x) \quad (2.3)$$

که در آن $[q_1, q_2, \dots, q_m]$ عامل‌های ظاهر می‌باشند. عامل‌های شکل و ظاهر برای هر تصویر در فرایند AAM بدمست می‌آید.

¹Active Appearance Model

²Appearance

الگوریتم های POIC^۱ و SIC^۲ دو الگوریتم شناخته شده برای این منظور می باشند. رویکرد SIC نسبت به POIC در شرایطی که تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می باشد. در این مقاله از یک روش SIC سریع برای حل مسئله بهینه سازی با 100×100 تکرار استفاده شده است. اگر $p = [p_1, p_2, \dots, p_n]$ در این مجموعه عامل های بدست آمده باشد، یک تبدیل هم نسبی قطعه ای^۳ $W(x; p)$ برای روبه رو سازی چهره مورد استفاده قرار می گیرد که در آن هریک از مثلث های روی توری، به صورت جداگانه به تصویر نتیجه با استفاده از درونیابی نزدیک ترین همسایه^۴ نگاشت پیدا می نمایند. برای مقداردهی اولیه از یک مدل پایه s_0 استفاده می شود که مقدار p در آن صفر می باشد. شکل ۲.۳ نمایی از این روش را نشان می دهد.



شکل ۲.۳: مقداردهی اولیه و بهینه سازی AAM.^[۲۵]

پس از تنظیم کامل مدل بر روی چهره، یک تبدیل هم نسبی با پارامترهای بدست آمده توسط الگوریتم یادگیری یاد شده، می تواند حالت چهره را هنجارسازی نماید. در بخش شناسایی چهره، ابتدا بخش چانه از تصویر حذف می شود زیرا چانه تقریباً تاثیری در شناسایی یک چهره ندارد. سپس تصویر چهره به اندازه 64×64 پیکسل تبدیل می شود و به 64×8 بخش غیر هم پوشان با اندازه تقسیم می شود. سپس در هر بخش تبدیل DCT^۵ انجام می شود. ضرایب خروجی تبدیل DCT بر حسب یک پویش زیگزاگی مرتب می شوند. اولین ضریب در نظر گرفته نمی شود. زیرا نشان دهنده میانگین سطح خاکستری پیکسل های بخش می باشد. 10×8 ضریب بعدی که ضرایب فرکانس پایین می باشند، برای ایجاد بردار ویژگی چهره استفاده می شوند. برای آموزش و آزمایش از مجموعه داده FERET و LFW استفاده شده است که در آن تصاویر چهره با زوایایی چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این مقاله موفق به دستیابی به شناسایی چهره با دقیقیت 87.8% شده است. در شکل ۳ نتیجه آزمایش بر روی مجموعه داده FERET در زوایه های متفاوت قابل مشاهده می باشد.

¹Project-Out Inverse Compositional

²Simultaneous Inverse Compositional

³Piecewise Affine Transformation

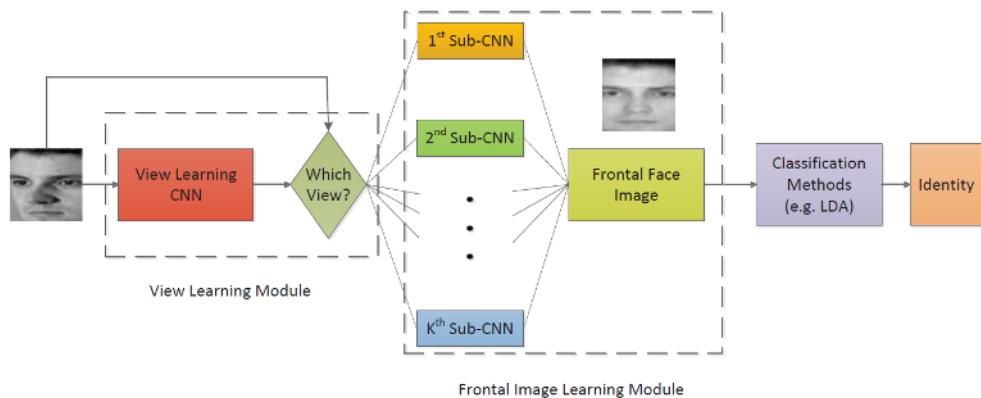
⁴Nearest Neighbor Interpolation

⁵Discrete Cosine Transform



شکل ۳.۳: نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۲۵].

در سال ۲۰۱۶ و همکاران [۲۶] یک روش رو به رو سازی چهره ارائه دادند که شناسایی چهره را مستقل از نمای چهره^۱ انجام می دهد. این الگوریتم یادگیری عمیق که VS2VI نامیده می شود، از دو بخش اصلی تشکیل شده است. بخش اول یک شبکه عصبی پیچشی برای یادگیری نما و زاویه چهره می باشد و بخش دوم از تعدادی شبکه عصبی پیچشی تشکیل شده است که هر کدام برای یادگیری تناظر^۲ بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص می باشد (شکل ۴.۳). این الگوریتم که می تواند با تعداد کمی داده نمونه، به خوبی آموزش ببیند، دو بخش تشکیل شده از شبکه عصبی پیچشی را به هم متصل می نماید تا مشکل نمای چهره در سامانه شناسایی چهره را برطرف نماید. در این معماری برای بازسازی چهره از زاویه رو به رو از لایه های واپیچشی^۳ به جای لایه های تمام متصل استفاده شده است.



شکل ۴.۳: معماری شبکه پیشنهادی VS2VI [۲۶].

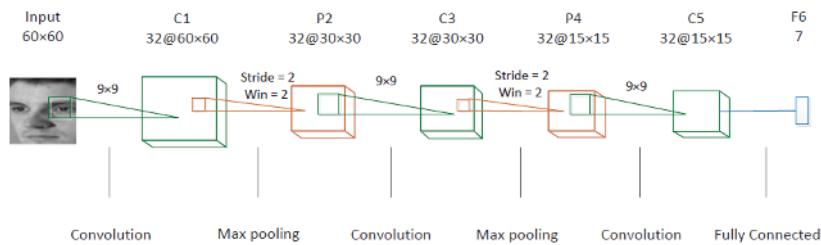
مدل VS2VI از دو بخش اصلی تشکیل شده است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با ابعاد 60×60 دریافت می کند و آن را با توجه به نمای چهره طبقه بندی می کند. سپس تصویر وارد بخش دوم می شود که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص

¹Facial View

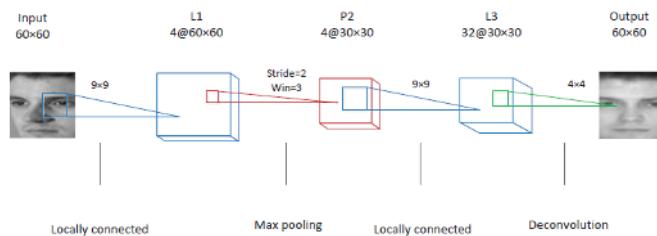
²correspondence

³deconvolutional

می باشد، تشکیل شده است. در این بخش چهره با نمای رو به رو بدست آمده و بررسی می شود تا هویت فرد مشخص شود. برای این منظور نیز از الگوریتم LDA^۱ برای طبقه بندی استفاده شده است. الگوریتم LDA برای یادگیری موقعیت چهره استفاده نمی شود و فقط برای دسته بندی نهایی مورد استفاده قرار می گیرد. معماری این مدل به ترتیب در شکل ۵.۳ و ۶.۳ قابل مشاهده می باشد.



شکل ۵.۳: معماری مدل یادگیری موقعیت چهره [۲۶].



شکل ۶.۳: معماری مدل یادگیری بازسازی چهره از رو به رو [۲۶].

بخش اول از یک شبکه عصبی پیچشی تشکیل شده است که شامل سه لایه پیچشی، دو لایه رای گیری و یک لایه تمام متصل می باشد. ورودی آن یک تصویر با هر موقعیت و زاویه دلخواه و خروجی آن احتمال قرار داشتن تصویر ورودی در هر دسته از دسته های مربوط به نماهای مختلف می باشد. برای لایه های پیچشی از تابع فعالیت ReLU استفاده شده است. و لایه تمام متصل از softmax به عنوان تابع هزینه استفاده کرده است.

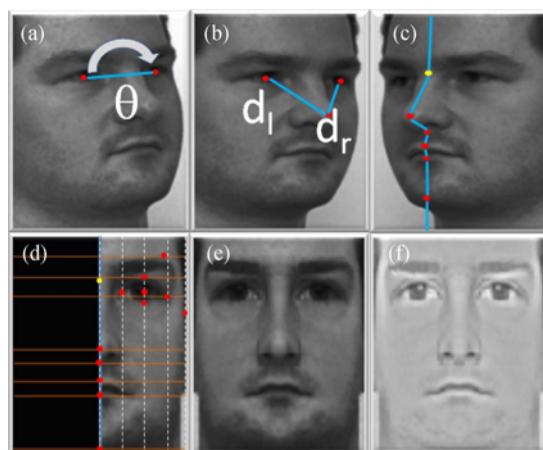
بخش دوم از تعدادی زیر شبکه پیچشی که هر کدام برای یادگیری تناظر بین چهره از رو به رو با یک چهره از یک نمای خاص می باشد، تشکیل شده است. هر یک از این زیر شبکه ها شامل دو لایه با اتصال محلی، یک لایه رای گیری و یک لایه واپیچشی می باشند. سه لایه اول برای استخراج ویژگی ها و لایه آخر برای بازیابی چهره از رو به رو می باشند. ورودی و خروجی این لایه ها تصویر چهره می باشد. لایه آخر به جای لایه تمام متصل از لایه واپیچشی استفاده شده است. این کار حجم محاسبات را به طور قابل توجهی کاهش می دهد. یک لایه تماماً متصل به 10^3 میلیون پارامتر نیاز دارد، در حالی که لایه واپیچشی به ۴۶۰ هزار پارامتر نیاز دارد. لایه اول که اتصال محلی دارد، از تابع PreLU به عنوان تابع فعالیت استفاده کرده است. لایه واپیچشی برای نمونه افزایی از درون یابی دو خطی استفاده کرده

¹linear discriminant analysis

است. برای یادگیری شبکه از الگوریتم پس انتشار خط‌الکتری^۱ استفاده شده است. الگوریتم VS2VI به دقت ۹۵٪/۶٪ در تشخیص چهره با زاویه ۴۵ درجه رسیده است.

در سال ۲۰۱۸ Andrey V.Savchenko و همکاران [۲۷] یک روش مبتنی بر ML^۲ برای شناسایی چهره در محیط‌های بدون محدودیت با تعداد کم نمونه‌ها بر اساس محاسبه فاصله بین ویژگی‌های با ابعاد بالا که توسط شبکه عصبی پیچشی عمیق مانند VGG، ResNet و SENet استخراج شده است ارائه دادند. این روش جدید شناسایی آماری، احتمال فاصله‌ها را نسبت به تمام تصاویر Kullback–Leibler مجموعه داده‌ها با استفاده از قانون بیز به حداقل می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی بین ویژگی‌های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده‌های IJB-A و YTF، LFW و ایجاد آزمایش قرار گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقت ۳٪ تا ۵٪ درصد در مقایسه با روش‌های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند.

در سال ۲۰۱۳ Marsico و همکاران [۲] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم STASM^۳ برای به دست آوردن ۶۸ نقطه ویژه بر روی چهره استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه (SP)^۴ محاسبه می‌شود و در صورتی که مقدار آن کمتر از یک آستانه باشد، تصویر مردود شده و در غیر این صورت به مرحله بعد برای هنجارسازی حالت فرستاده می‌شود. هرچه مقدار شاخص SP بالاتر باشد، تصویر چهره به حالت تمام رخ نزدیکتر است و اصلاح زاویه کمتری نیاز دارد. شکل ۷.۳ قسمت a تا c معیارهای مورد نیاز برای محاسبه شاخص SP را نشان می‌دهد.



شکل ۷.۳: ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۲].

چرخش: چرخش سر در جهت عقربه‌های ساعت یا عکس آن می‌باشد. و طبق رابطه ۳.۳ به صورت زاویه θ تعریف می‌شود که زاویه

¹Backpropagation

²Maximum Likelihood

³Extended Active Shape Model

⁴Simple Pose

بین خط عبوری از مرکز چشم‌ها و محور افقی x می‌باشد.

$$roll = \min\left(\left|\frac{2\theta}{\pi}\right|, 1\right) \quad (4.3)$$

انحراف: چرخش در راستای محور افقی است و طبق رابطه ۴.۳ مقادیر r و θ فاصله مرکز چشم چپ و راست از نوک بینی می‌باشد.

اندازه گیری این فاصله‌ها در صورت برابر بودن، برای تشخیص تمام رخ بودن تصویر چهره مورد استفاده قرار می‌گیرد.

$$yaw = \frac{\max(d_l, d_r) - \min(d_l, d_r)}{\max(d_l, d_r)} \quad (4.3)$$

شیب: بر حسب رابطه ۵.۳ چرخش سر در راستای محور عمودی را اندازه گیری می‌کند.

$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)} \quad (5.3)$$

با محاسبه ۳ شاخص فوق، شاخص SP مطابق رابطه ۶.۳ محاسبه می‌شود:

$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch) \quad (6.3)$$

که در آن $1 = \alpha + \beta + \gamma$ می‌باشد که مقادیر این ضرایب از طریق آزمون و خطا به دست می‌آیند. سپس در مرحله تمام

رخ کردن تصویر چهره، بین دو فاصله dr و dl هر کدام بزرگتر باشند، نشان می‌دهد آن سمت از چهره بیشتر در دید دوربین است.

اگر نیمه سمت راست صورت به طرف دوربین باشد ($dl \geq dr$)، تصویر بدون تغییر باقی می‌ماند. در غیر این صورت، تصویر حول

محور عمودی برعکس می‌شود که باعث می‌شود همیشه نیمه سمت راست تصویر پردازش شود. سپس برای ثابت کردن طول سطرها،

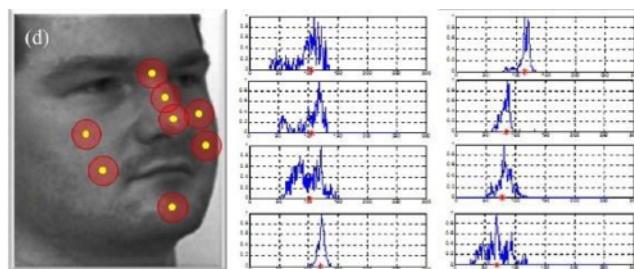
سطرها بسط داده می‌شوند. مطابق شکل ۷.۳ نیمه سمت چپ تصویر حذف شده و از روی تصویر نیمی از چهره، نیمه دیگر نیز ساخته

می‌شود و تصویر تمام رخ چهره به دست می‌آید.

۳.۳ چالش روشنایی

متعادل سازی بافت نگار یکی از الگوریتم های مهم در پردازش تصویر است که هدف آن افزایش وضوح تصویر با یکنواخت سازی بافت نگار تصویر است، به گونه ای که علت روشنایی کم یا زیاد، پنهان می باشد، قابل مشاهده شوند. متعادل سازی بافت نگار قدرتمندترین و رایج ترین روش برای اصلاح روشنایی تصاویر است. اما ضعف این روش، سراسری بودن آن می باشد. برای رفع این مشکل باید از الگوریتم های محلی استفاده کرد.

در سال ۲۰۱۳ Marsico و همکاران [۲] یک روش هنجار سازی نورپردازی برای تصاویر چهره ارائه دادند و از این روش برای محاسبه شاخص روشنایی نمونه (SI)^۱ استفاده کردند. زمانی که تصویر روشنایی یکنواخت دارد، بیشتر بخش های چهره توزیع یکنواخت سطح خاکستری دارند. اما وقتی روشنایی یکنواخت نباشد، برخی از نواحی خاص چهره، توزیع یکنواخت سطح خاکستری ندارند. برای مثال جلوی بینی، گونه ها و چانه معمولاً نور را منعکس می کنند. ۸ ناحیه در شکل ۸.۳ با توجه به چنین اصلی انتخاب شده اند. ۸ بابت نگار با رنگ آبی و مرکز آن ها با رنگ قرمز مشاهده می شود.



شکل ۸.۳: اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۲].

۸ بابت نگار فوق به یک توزیع نرمال با انحراف معیار کم در همسایگی از مرکز حجم بافت نگار اشاره دارد. بافت نگار هر یک از ناحیه ها بدست آمده و مرکز ثقل آن مطابق رابطه ۷.۳ محاسبه می شود:

$$mc(w) = \frac{\sum_{i=0}^{255} i \times h_w(i)}{\sum_{i=0}^{255} h_w(i)} \quad (7.3)$$

که در آن w نشان دهنده یکی از نواحی ۸ گانه می باشد. ۸ مرکز جرم محاسبه شده، بردار mc را تشکیل می دهند. با توجه به فرض تشابه ذکر شده در میان نواحی صورت در نظر گرفته شده، انتظار می رود هیچ تنوع قابل توجهی در میان عناصر بردار وجود نداشته باشد و توزیع های یکسانی از سطوح خاکستری را نمایش دهند. برای دستیابی به این منظور پراکندگی مراکز حجم ها از ۸ نمودار

¹Sample Illumination

بافت نگار محاسبه شده است. سپس عناصر بردار mc توسطتابع سیگموید F در بازه $[0, 1]$ هنجارسازی می‌شوند و شاخص کیفیت

روشنایی مطابق رابطه ۸.۳ محاسبه می‌شود که یک عدد می‌باشد:

$$SI = 1 - F(std(mc)) \quad (8.3)$$

هرچه مقدار SI بیشتر باشد، یعنی تصویر روشنایی یکنواخت تری دارد. اگر این شاخص به اندازه کافی رضایت‌بخش نباشد، تصویر رد می‌شود. در غیر این صورت برای هنجارسازی روشنایی وارد بخش بعدی خواهد شد. در صورت رد شدن تصویر، سیاست‌های جایگزین برای رسیدگی به این موضوع در دسترس هستند. برای مثال ممکن است یک نمونه جدید درخواست شود که در شرایط برون خط امکان پذیر نیست. یا مداخله انسانی می‌تواند به صورت دستی نمونه را طبقه‌بندی کند. اگر تصویر به مرحله بعد وارد شد، با استفاده از الگوریتم SQI توسط یک ماسک مربعی با اندازه 8×8 مقدار هر پیکسل بر مقدار میانگین همسایگانش تقسیم می‌شود و نتیجه نهایی حاصل می‌شود. نتیجه به صورت شکل ۷.۳ می‌باشد.

۴.۳ چالش انسداد

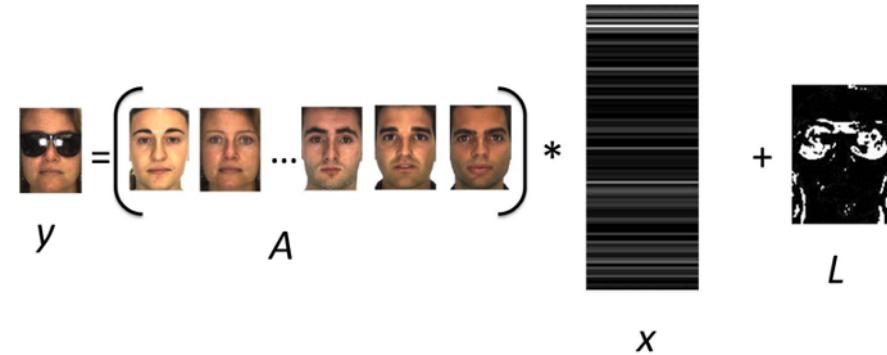
در سال ۲۰۱۸ Cho Ying Wu و همکاران^۱ یک رویکرد مبتنی بر واپاش با جهت گرادیان برای شناسایی چهره‌های در معرض انسداد به نام $GD\text{-HASLR}$ ^۱ ارائه دادند. در کاربردهای واقعی، تعداد داده‌های آموزش بسیار کم می‌باشد (شاید یک تصویر به ازای هر شخص). این رویکرد توانایی برخورد با این شرایط را دارد و در مقابل تصاویری که نزدیک به 80 درصد از چهره در شرایط انسداد قرار دارد، به خوبی عمل می‌کند. نتایج نشان می‌دهد که با تعداد بسیار کمی از تصاویر آموزشی، مدل پیشنهاد شده $GD\text{-HASLR}$ بهترین عملکرد را در مقایسه با سایر روش‌های پیشرفته، از جمله روش‌های مبتنی بر شبکه عصبی پیچشی دارد. مجموعه داده تصاویر می‌باشد. داده‌های آموزشی چهره‌های طبیعی و بدون انسداد می‌باشند. $L = \mathbb{R}^d$ می‌توان از یک ترکیب خطی داده‌های آموزش برای تخمین زدن داده آزمایش استفاده کرد که شامل یک عبارت خطای $A = \mathbb{R}^{d \times n}$ در نظر گرفته شده که در آن n تعداد داده‌های آموزشی و d حاصل ضرب تعداد پیکسل‌های طول و عرض آموزشی می‌باشد. مطابق رابطه

$$y = \mathbb{R}^d \quad \text{می‌توان از یک ترکیب خطی داده‌های آموزش برای تخمین زدن داده آزمایش استفاده کرد که شامل یک عبارت خطای} \\ L = \mathbb{R}^d \quad \text{می‌باشد.}$$

¹gradient direction-based hierarchical adaptive sparse and low-rank

$$y = Ax + L \quad (9.3)$$

که در آن x بردار ضرایب با n بعد می‌باشد. (شکل ۹.۳)



شکل ۹.۳: تصویر انسداد از ترکیب خطی چهره‌های آموزشی مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۲۸].

برای آنکه شرط تنک بودن به رابطه بالا اضافه شود، مسئله به صورت رابطه ۱۰.۳ نوشته می‌شود:

$$\operatorname{argmin} \|x\|_1 \text{ s.t. } \|y - Ax\| \leq \epsilon \quad (10.3)$$

که در آن ϵ یک آستانه خطا می‌باشد. برای تصویر ورودی و تصاویر مجموعه داده آموزش، گرادیان مرتبه اول، دوم و سوم محاسبه شده

و به عنوان ویژگی هر تصویر در نظر گرفته می‌شود. در ادامه شرط کم رتبه بودن ماتریس ویژگی‌ها نیز به این رابطه اضافه می‌شود. با

استفاده از روش ضرایب لاغرانژ، رابطه بالا را می‌توان به صورت یک مسئله بهینه سازی بدون محدودیت نوشت و حل نمود.

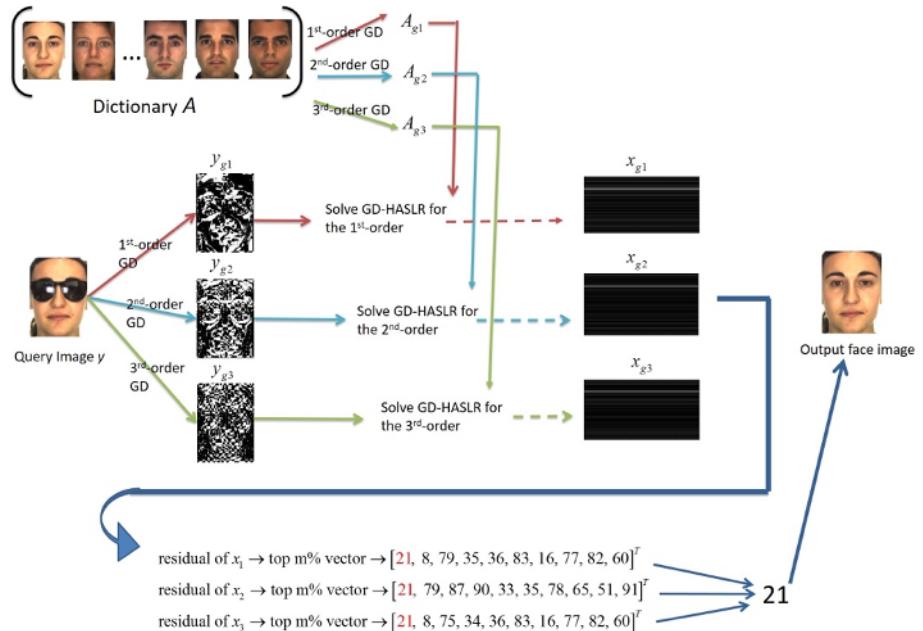
$$\mathcal{L}(x, L, z) = \alpha \|L_M\| + \sum \pi_\lambda(|x_i|) + z^T(y - Ax - L) + \frac{\beta}{2} \|y - Ax - L\|_2^2 \quad (11.3)$$

که در آن z ضریب لاغرانژ و β عامل مجازات می‌باشد. پس از بدست آوردن بردار تنک x می‌توان باقیمانده دسته i ام را به صورت

رابطه ۱۲.۳ محاسبه نمود:

$$r_i = y - Ai(x)2 \quad (12.3)$$

که در آن (x_i) نشان دهنده i امین انتخاب کننده دسته می‌باشد که فقط ورودی‌های مربوط به دسته i را حفظ می‌کند و در سایر قسمت‌ها برابر با صفر می‌باشد. در نهایت دسته‌ای که کمترین باقیمانده را داشته باشد، انتخاب می‌شود. رویکرد کلی الگوریتم در شکل ۱۰.۳ آمده است.



شکل ۱۰.۳: رویکرد کلی الگوریتم [۲۸].

در سال ۲۰۱۴ J. Li و همکاران [۵۹] یک روش تشخیص چهره پوشیده شده در پس زمینه پیچیده ارائه کردند. این الگوریتم از دو مرحله تشکیل شده است. در مرحله اول تعیین می‌کنند که آیا شی یک شخص می‌باشد یا خیر و در مرحله دوم بررسی می‌شود که آیا چهره پوشیده شده می‌باشد یا خیر و در صورت پوشش چهره، نوع پوشش و اینکه پوشیدگی با ماسک، کلاه، عینک یا ... است را مشخص می‌کند.

در مرحله اول یک رویکرد تشخیص شی در پیش زمینه در حالت پویا و ایستا پیشنهاد شده است. برای تشخیص هدف ایستا از تشخیص مبتنی بر ویژگی HOG استفاده شده است. از آنجا که سرعت HOG نسبتاً پایین است، از LBP^۱ به همراه آن نیز استفاده کرده اند.

در مرحله دوم از طبقه بند Adaboost برای طبقه بندی چهره های پوشیده شده است که برای انواع پوشیدگی آموزش داده شده است.

^۱Local Binary Pattern

۵.۳ چالش کمبود تصاویر آموزشی

دلیل اصلی به وجود آمدن چالش این است که چهره انسان یک شی چسب نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده‌های آموزشی بسیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویربرداری شده باشد. در مقابل فرض بر این است که داده‌های آموزش بسیار کم هستند. از این رو مسئله تشخیص چهره باید در شرایطی حل شود که داده‌های آموزشی کافی در اختیار نمی‌باشد. بنابراین به الگوریتمی نیاز داریم که به ما کمک کند با تولید داده‌های غیر واقعی، مشکل کمبود داده‌های آموزشی را حل نماییم. از سویی دیگر محدودیت منابع برای اجرای پردازش‌ها بر روی تلفن همراه وجود دارد و الگوریتم ارائه شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد.

در سال ۲۰۱۷ Ya Wang و همکاران [۲۹] روشی برای تشخیص چهره در دوربین‌های نظارتی در محیط بدون محدودیت به وسیله شبکه عصبی پیچشی عمیق ارائه دادند. از آنجایی که داده‌های آموزشی ورودی به مدل از اهمیت بالایی برای تشخیص برخوردار هستند و همچنین به تعداد زیادی از داده‌های هر دسته برای بهبود عملکرد سامانه نیاز است، نوآوری این رویکرد، ساختن یک مجموعه داده استاندارد برای شبکه عصبی از روی دوربین‌های نظارتی در محیط است که در چهار مرحله به صورت زیر ساخته می‌شوند. با توجه به اینکه تصاویر مورد نظر برای هر فرد در مجموعه فریم‌های پشت سر هم از یک دوربین موجود است، می‌توان مجموعه تصاویر یک فرد را بوسیله ترکیب الگوریتم تشخیص چهره و ردیابی چهره جمع آوری کرد. پس از شناسایی یک چهره، با ردیابی آن، مجموعه تصاویری از آن به عنوان یک دسته طبقه‌بندی می‌شود.



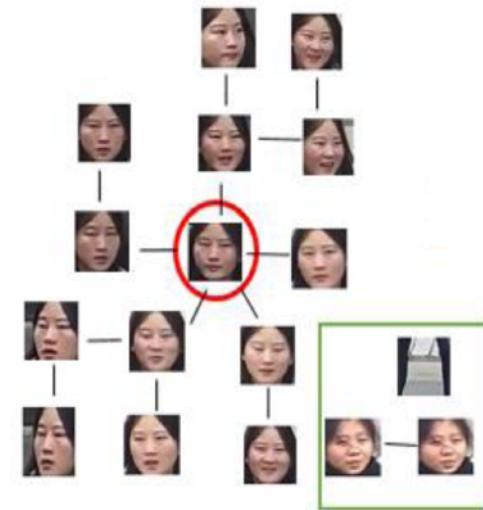
شکل ۱۱.۳: ردیابی، یافتن چهره‌ها و برچسب زنی [۲۹].

برخی تصاویر در هر دسته به اشتباه در مرحله اول به عنوان تصویر یک فرد در نظر گرفته شده اند (شکل ۱۲.۳).



شکل ۱۲.۳: تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند [۲۹].

با استفاده از روش خوشه بندی گراف^۱ روی ویژگی های استخراج شده از شبکه VGG-Face، تشخیص و پاک سازی تصاویر اشتباه انجام می شود. فاصله کسینوسی بین ویژگی های تصاویر چهره محاسبه می شود و اگر این فاصله برای هر دو تصویر کمتر از یک مقدار آستانه باشد، این تصاویر متعلق به یک فرد هستند. با توجه به شکل ۱۳.۳ تصویری که بیشترین شباهت را به تصاویر دیگر دارد، به عنوان شاخص برای آن شخص انتخاب می شود.



شکل ۱۳.۳: استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [۲۹].

با استفاده از محاسبه فاصله بین هر داده با داده مرکزی و در نظر گرفتن یک آستانه، داده های تکراری در هر دسته مشخص شده و حذف می شوند. با توجه به مقدار داده های درون هر دسته، تصفیه بین دسته ای انجام می شود. اگر مجموعه داده های درون هر دسته کمتر از ۱۰۰ تصویر باشد، آن دسته از مجموعه داده حذف می شود. دقت خوشه بندی و جمع آوری مجموعه داده ۹۹٪ شده است. در نهایت از یک مدل پیش آموزش دیده شده شبکه VGG-Face با Fine-tuning همراه با طبقه بندی تصاویر آزمایشی استفاده شده است که به دقت ۹۲٪ رسیده است.

در سال ۲۰۱۶ رادفورد و همکاران [۳۰] از شبکه های مولد تخصصی برای تولید داده ها استفاده کردند که به اختصار GAN نامیده می شوند. GAN از دو شبکه مستقل تولید کننده و تمیز دهنده تشکیل شده است. شبکه تولید کننده از روی بردار Z که می تواند

¹Graph Clustering

یک نویز تصادفی باشد، یک تصویر تولید می‌کند و شبکه تمیز دهنده وظیفه دارد تصاویر واقعی را از تصاویر تولید شده توسط شبکه تولید کننده تشخیص دهد. بنابراین هر تصویر با یک بردار Z معرفی می‌شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می‌شود. به عنوان مثال بردار Z برای تصویر خانمی که عینک آفتابی نزدیک است از بردار Z برای تصویر خانمی که عینک آفتابی نزدیک است، کم می‌شود و حاصل آن، بردار مربوط به یک عینک آفتابی می‌باشد. سپس این بردار با بردار تصویر آقایی که عینک نزدیک است جمع می‌شود. نتیجه نهایی تصویر همان آقا با عینک آفتابی می‌باشد. به عنوان مثالی دیگر، با میانگین گیری بردارهای مربوط به دو تصویر از چهره شخصی که به سمت راست و چپ متمایل است، توانسته چهره رو به روی شخص را بازسازی نماید. اما کیفیت کار هنوز تا حالت مطلوب فاصله دارد. نمونه‌ای از خروجی این روش در شکل ۱۴.۳ قابل مشاهده می‌باشد.



شکل ۱۴.۳: تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۳۰].

در سال ۲۰۱۸ بانرجی و همکاران [۳۱] یک شبکه عمیق مبتنی بر GAN را با نام LR-GAN پیشنهاد دادند، که تصاویر واقع گرایانه با وضوح بالا از روی تصاویر با وضوح پایین بازسازی می‌کند. این تصاویر چهره غیر واقعی اما واقع گرایانه و با کیفیت، باعث عملکرد بهتر سامانه شناسایی چهره برای مقایسه تصاویر می‌شود. رویکرد اصلی مقاله در روش آموزش تخصصی LR-GAN بهینه سازی تابع ضرر بازسازی چند مقیاسی است، بر اساس شاخص‌های مانند: شاخص شباهت ساختاری چند مقیاسی (SSIM)^۱، میانگین مربعات خطای هر قسمت (PMSE)^۲، واگرایی جنسن شانون اصلاح شده (JSD)^۳ و تنوع متقابل در اطلاعات (MVI)^۴.

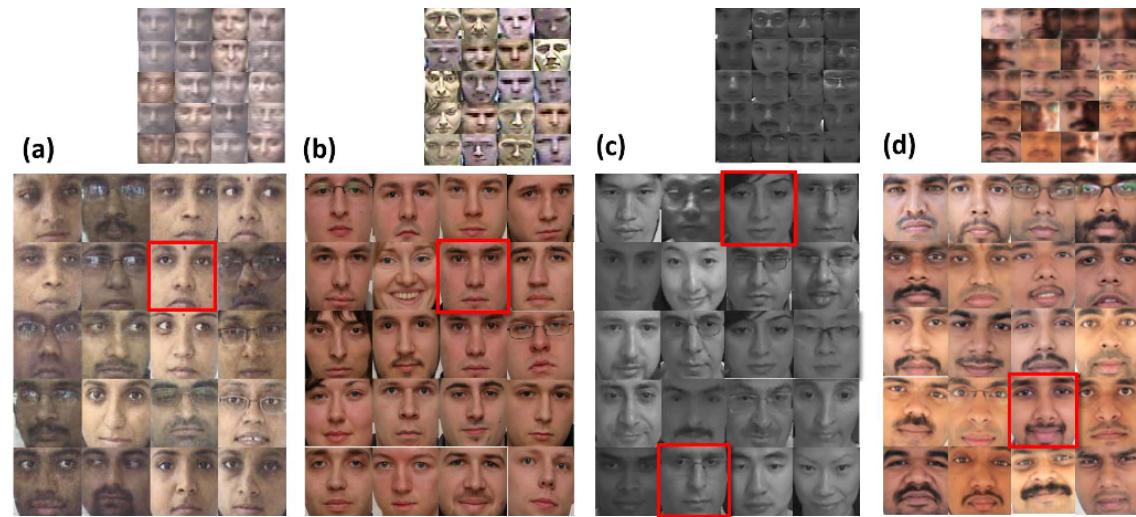
^۱Structural Similarity Index

^۲patch-wise mean-squared error

^۳Jensen–Shannon Divergence

^۴Mutual Variation in Information

شبکه تمیز دهنده در LR-GAN، بر اساس اطلاعات طبقه بندی که به طور ضمنی در طول آموزش آموخته می‌شود، هویت هر شخص را حفظ می‌کند. این رویکرد سریعتر از شبکه‌های مبتنی بر GAN اخیر به یک همگرایی می‌رسد. نمونه‌ای از خروجی این روش در شکل ۱۵.۳ قابل مشاهده می‌باشد.



شکل ۱۵.۳: تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین. موارد با حاشیه قرمز خروجی اشتباه هستند [۳۱].

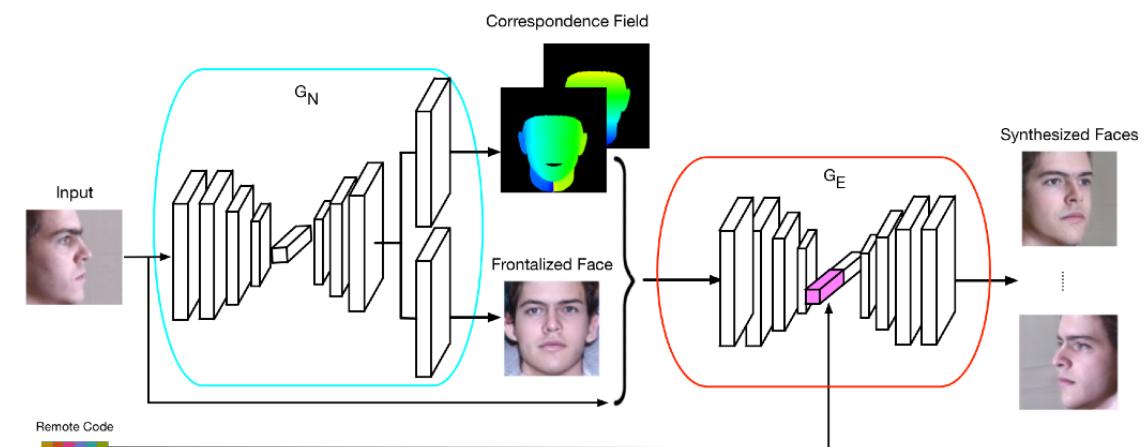
شبکه‌های GAN یاد می‌گیرند تصاویر جدیدی تولید کنند که شبیه به تصاویر واقعی باشند. اما این شبکه‌ها عموماً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند. در سال ۲۰۱۹ کاراس و همکاران [۳۲] یک شبکه GAN جدید پیشنهاد دادند که بخش تولید کننده آن به طور خودکار یاد می‌گیرد بدون هیچ ناظر انسانی ویژگی‌های بصری را از یکدیگر جدا نماید. پس از اتمام مرحله یادگیری، ما می‌توانیم این ویژگی‌های بصری را به دلخواه خود ترکیب نماییم. برای مثال ویژگی‌های اساسی مانند جنسیت، سن، طول مو، وجود عینک و زاویه چهره را از تصویر ۱ با ویژگی‌های دیگری از تصویر ۲ ترکیب کرد و یک چهره جدید تولید نمود. نگاه این شبکه تولید کننده به هر تصویر، مجموعه ای از ویژگی‌های بصری می‌باشد. هر ویژگی بصری با اندازه مشخص، جلوه‌های تصویر را کنترل می‌کند. ویژگی‌های بصری غالب مانند زاویه چهره، مو، شکل صورت؛ ویژگی‌های بصری میانی مانند فرم لب و چشم‌ها و ویژگی‌های ساده‌تر مانند رنگ. ما می‌توانیم این ویژگی‌های بصری را با ضرایب دلخواه خود ترکیب نماییم. نمونه‌ای از خروجی این روش در شکل ۱۶.۳ قابل مشاهده می‌باشد.

در سال ۲۰۱۹ کافو و همکاران [۳۳] به موضوع تولید چهره در زوایای دلخواه پرداختند. در این مقاله از دو شبکه GAN استفاده شده است که در شبکه اول از روی چهره زاویه‌دار، چهره رویه‌رو تولید شده است. سپس با استفاده از شبکه GAN دوم از روی تصویر چهره رویه‌رو، تصویر با زاویه دلخواه با استفاده از یک پارامتر کنترلی تولید می‌شود. چالشی که در این مقاله به آن اشاره شده است،



شکل ۱۶.۳: تصاویر هر ردیف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند [۳۲].

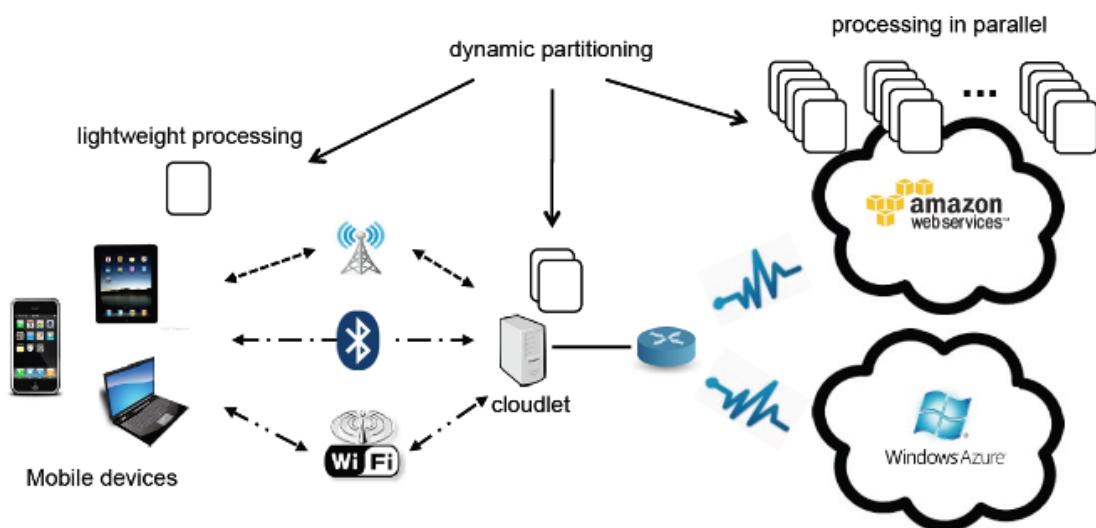
مسئله عدم توازن داده‌ها در وجود برخی ویژگی‌ها در تصاویر می‌باشد. این مقاله در برخی تصاویر چهره زاویه‌دار به مشکل برخورد می‌کرد. به عنوان مثال چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره روبه‌رو بدون عارضه تولید شده است. این چالش از جایی نشات می‌گیرد که تصاویر با عارضه پوستی در مجموعه داده بسیار کم می‌باشند و شبکه در مواجهه با این مسئله ایده‌ای برای آن ندارد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند. نمونه‌ای از خروجی این روش در شکل ۱۷.۳ قابل مشاهده می‌باشد.



شکل ۱۷.۳: ساختار شبکه AD GAN - شبکه G_N برای رو به رو سازی چهره و شبکه G_E برای تولید چهره از زوایای مختلف [۳۳].

۶.۳ چالش منابع محدود

در سال ۲۰۱۲ و همکاران [۳۴] یک روش تشخیص چهره بی درنگ مبتنی بر بینایی ابری^۱ با استفاده از معماری MOCHA^۲ ارائه کردند (شکل ۱۸.۳). با فراگیر شدن تلفن همراه هوشمند در میان شهروندان، سامانه تشخیص چهره می‌تواند از همکاری مشترک محاسبات تلفن همراه و رایانش ابری استفاده کند. چالش این سامانه، چگونگی تجزیه انجام وظیفه بین تلفن همراه و فضای ابری، توزیع بار محاسبه در میان سرورهای ابر برای به حداقل رساندن زمان پاسخ با توجه به تأخیر ارتباطات مختلف و قدرت محاسبه سرور می‌باشد. نتایج نشان می‌دهد که الگوریتم‌های بخش بندی پردازش بین تلفن همراه و فضای ابری با توجه به زمان تأخیر ناهمگن، توانایی محاسبه را به طور قابل توجهی افزایش می‌دهند.



شکل ۱۸.۳: معماری MOCHA: دستگاه‌های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می‌کنند [۳۴].

این سامانه از لحاظ ساختار به سه بخش تقسیم می‌شود:

دستگاه همراه: تلفن‌های همراه و iPad ها نقش تهیه و ارسال تصاویر را دارند. تصاویر با فرمت RAW فرستاده می‌شوند تا قابلیت پیش‌پردازش بهتری داشته باشند. اگر سرور ابری به دستگاه همراه نزدیک باشد و ارتباط با سرعت بالا امکان پذیر باشد، تصاویر پیش‌پردازش به سرور فرستاده می‌شوند. در غیر این صورت مرحله پیش‌پردازش در دستگاه همراه انجام می‌شود، ویژگی‌های Haar به سرور فرستاده می‌شوند و محاسبه ویژگی‌های هار و طبقه‌بندی در سرور انجام می‌شوند. پس از اتمام فرایند تشخیص چهره، نتیجه نهایی برای تلفن همراه فرستاده می‌شود.

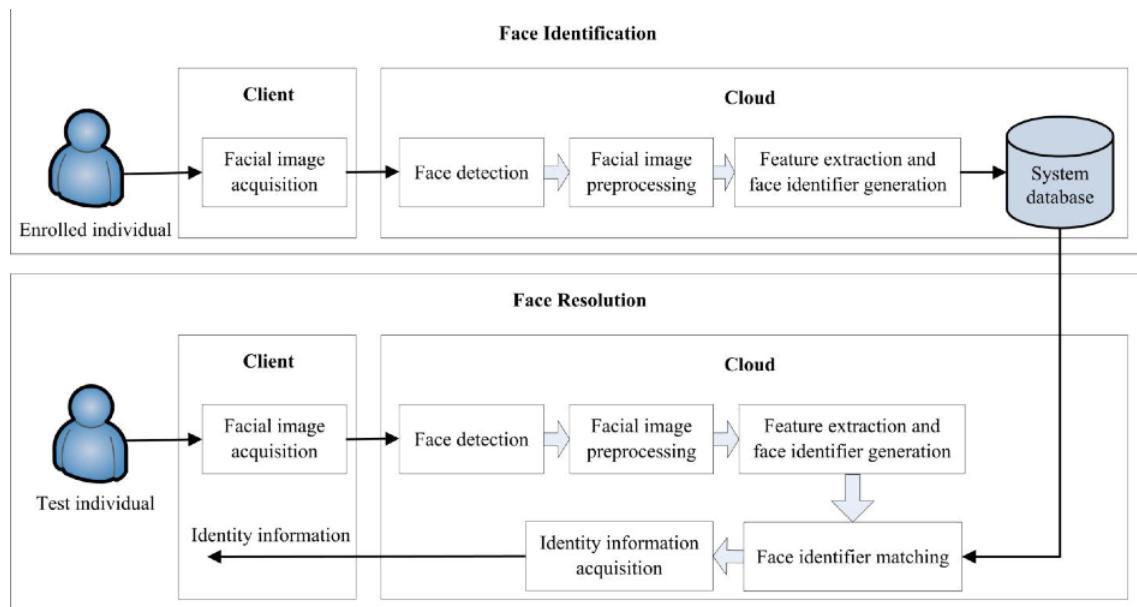
¹Cloud Vision

²mobile-cloudlet-cloud architecture

ابر کوچک : سرورها و رایانه هایی که توانایی پردازشی متوسطی دارند، ابر کوچک یا cloudlet نامیده می شوند. این دستگاه ها که به عنوان میان دستگاه های همراه و سرورهای ابری اصلی قرار دارند، مجهز به GPU می باشند تا بتوانند پردازش موازی را در زمان مطلوبی انجام دهند.

ابر: سرورهای ابر دارای توان پردازشی و پاسخگویی بسیار بالا می باشند که با محاسبات سنگین سامانه را به دوش می کشند و تصمیم گیری نهایی بر روی آن انجام می پذیرد.

در سال ۲۰۱۸ Pengfei Hu و همکاران [۳۵] یک رویکرد تشخیص چهره مبتنی بر رایانش ابری ارائه کردند. افزایش برنامه های کاربردی در زمینه کلان داده ها^۱ باعث افزایش تقاضای سامانه های شناسایی چهره برای محاسبات قدرتمند و ظرفیت ذخیره سازی بالا می شود. این سامانه به طور کامل از مزایای محاسبات ابری بهره می برد تا به طور موثر توانایی محاسبات و ظرفیت ذخیره سازی را بهبود بخشد. نتایج تجربی نشان می دهد که طرح پیشنهادی عملکار پذیر است و می تواند سرویس شناسایی مؤثر چهره را فراهم کند. همانطور که در شکل ۱۹.۳ مشاهده می شود، تنها تهیه تصویر بر عهده دستگاه سرویس گیرنده می باشد و تمام محاسبات یافتن و شناسایی چهره بر روی ابر انجام می شود.



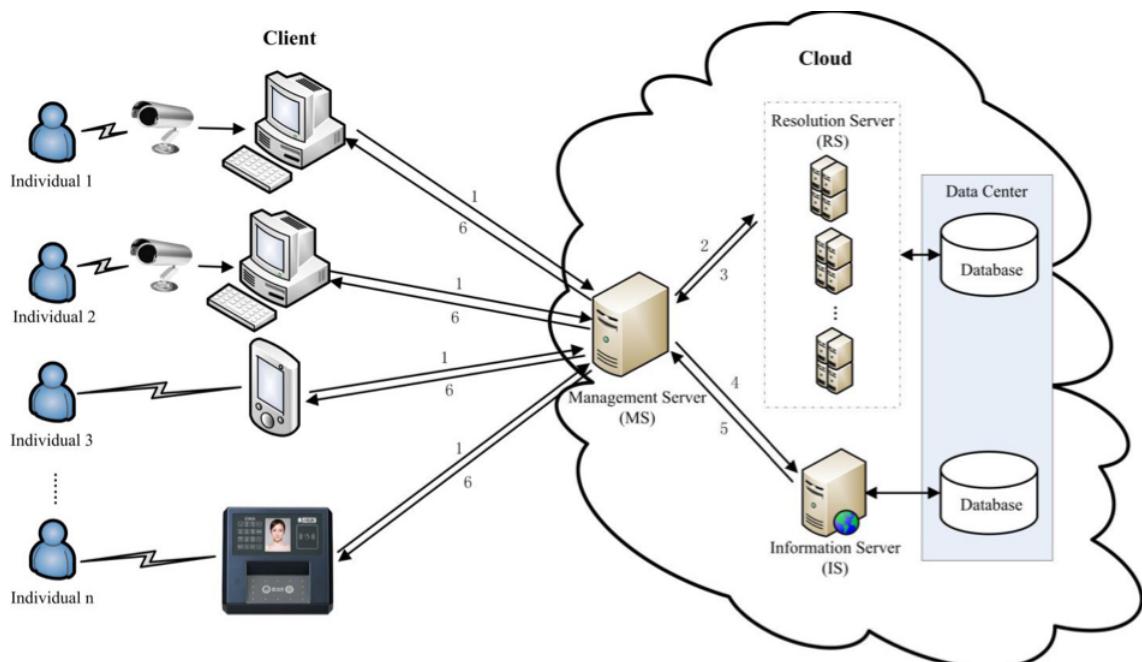
شکل ۱۹.۳: نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۳۵].

در این سامانه تصویر با فرمت RAW برای ارسال شده و برای یافتن چهره از ویژگی های Haar استفاده شده است. سپس عملیات همسان سازی بافت نگار بر روی تصویر چهره اعمال می شود تا بهبود جزئی حاصل شود. سپس از الگوریتم LBP^۲ برای استخراج

¹Big Data

²Local Binary Patterns

ویژگی های چهره استفاده شده، برای هر تصویر یک شناسه تولید می گردد و در نهایت با استفاده از فاصله اقلیدسی با شناسه تصاویر موجود در پایگاه داده مطابقت داده می شود. همانطور که در شکل ۲۰.۳ مشاهده می شود این سامانه ابری از بخش های سرور مدیریت^۱، سرور اطلاعات (IS)^۲، سرور شناسایی (RS)^۳ و پایگاه داده تشکیل شده است. به علت قدرت بالای پردازش در سرور ابری، امکان پردازش موازی نیز در این سامانه وجود دارد که باعث افزایش سرعت محاسبات و کاهش زمان پاسخ دهی سامانه می گردد.



شکل ۲۰.۳: چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۳۵].

۷.۳ توابع ضرر مبتنی بر افزایش حاشیه

علاوه بر رویکردهای بالا که هر یک بر روی حل یک مسئله خاص تمرکز کرده بودند، برخی روش هایی که اخیراً معرفی شده اند، سعی بر این داشته اند که یک راه حل نسبتاً همه جانبه در مورد مسئله تشخیص چهره و مشکلات آن ارائه دهند. یکی از این رویکردها، استفاده از تابع ضرر CosFace می باشد که در سال ۲۰۱۸ و همکاران [۳۶] ارائه دادند. این تابع ضرر کسینتوسی با حاشیه زیاد^۴ شباهت بسیار زیادی به تابع ضرر softmax دارد. با این تفاوت که به جای ضرب ماتریس ضرب های W در بردار ویژگی x حاصل ضرب مقادیر موجود در ویژگی های استخراج شده و آخرین لایه کامل متصل را به صورت (j)

$$W_j^T x_i = ||W_j|| |x_i| \cos(j)$$

¹Management Server

²Information Server

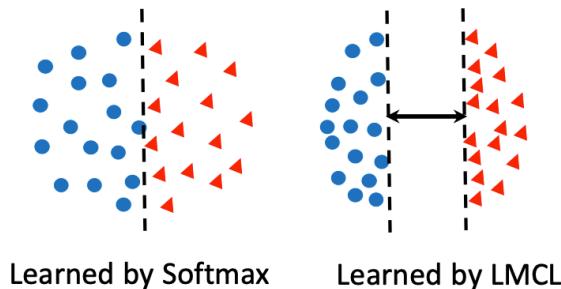
³Resolution Server

⁴Large Margin Cosine Loss

تبدیل می کند، که θ_j زاویه بین وزن W_j و ویژگی x_i است.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i}-m))}}{e^{s(\cos(\theta_{y_i}-m))} + \sum_{j=1}^n e^{s(\cos(\theta_{j,i}))}} \quad (13.3)$$

همانطور که در شکل؟ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می کند اما در مزهای تصمیم گیری ابهام قابل توجهی به وجود می آید، در حالی که تابع ضرر معروفی شده می تواند فاصله بیشتری را بین دسته های نزدیک اعمال کند.



شکل ۲۱.۳: تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مز بین دسته ها ایجاد می نماید [۳۶].

در سال ۲۰۱۹ و همکاران [۳۷] یک تابع ضرر پیشنهاد دادند که کمک می کند ویژگی های استخراج شده که متعلق به دو دسته متفاوت هستند، فاصله بیشتری از هم داشته باشند و در مقابل ویژگی های استخراج شده برای دو تصویر از چهره یک فرد یکسان، فاصله کمتری از هم داشته باشند. این روش که به دنبال مقاله [۳۶] آمده است، سعی در بهبود روش پیشین و افزایش دقت کرده است. برای آموزش شبکه های عصبی پیچشی عمیق برای تشخیص چهره، دو رویکرد اصلی وجود دارد. روش اول دسته بندی را آموزش می دهند که می تواند هویت های مختلف را در مجموعه آموزش از هم جدا کند، مانند استفاده از طبقه بندی softmax و رویکرد دوم که مستقیماً تولید یک بردار تعییه^۱ را یاد می گیرند، مانند loss triplet. بر اساس داده های آموزش در مقیاس بزرگ، هر دو روش می توانند عملکرد بسیار خوبی در تشخیص چهره داشته باشند. با این حال، هم رویکرد softmax و هم رویکرد triplet loss softmax اشکالاتی دارند. برای

۱. اندازه ماتریس تبدیل W به طور خطی با افزایش تعداد دسته ها (n) افزایش می یابد.
۲. ویژگی های آموخته شده برای مسئله های طبقه بندی با مجموعه بسته قابل تفکیک هستند اما به اندازه کافی برای مسئله تشخیص چهره که یک مسئله باز می باشد، مناسب نیستند.

¹Embedding

:triplet loss برای

۱. برای مجموعه داده‌های مقیاس بزرگ، رشد شدید در تعداد ترکیب‌های تعداد تصاویر سه گانه وجود دارد که منجر به افزایش

قابل توجه تعداد مراحل تکرار می‌شود.

۲. استخراج مجموعه تصاویر سه گانه یک مسئله دشوار برای آموزش موثر می‌باشد.

برای افزایش بیشتر قدرت تمایز مدل تشخیص چهره و ایجاد ثبات در روند آموزش، تابع ضرر مبتنی بر توابع مثلثاتی پیشنهاد شده

است. حاصل ضرب نقطه‌ای مقادیر موجود در ویژگی‌های استخراج شده و آخرين لایه کاملاً متصل، برابر با ضرب کسینوسی آن‌ها

پس از نرمال‌سازی می‌باشد. از تابع مثلثاتی کسینوسی برای محاسبه زاویه بین ویژگی فعلی و وزن هدف استفاده شده است. سپس

یک حاشیه زاویه‌ای به زاویه هدف اضافه شده، در انتهای دوباره مقادیر به فضای خطی برگردانده شده است. مراحل بعدی دقیقاً مانند

softmax هستند. رابطه ریاضی softmax معروف ترین تابع ضرر طبقه‌بندی که به طور گسترده استفاده می‌شود، به شرح زیر

است:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (14.3)$$

که در آن x_i نشان دهنده ویژگی عمیق نمونه i از دسته y است. W_j ستون j ام از وزن W می‌باشد و b_j بایاس است. مقدار N اندازه

دسته و n تعداد دسته‌ها است. از آنجا که تابع softmax ذات زاویه‌ای و قطبی دارد، بنابرین می‌توان آن را به صورت رابطه

بازنویسی کرد:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\|W_{y_i}\| \|x_i\| \cos(\theta_{y_i})}}{\sum_{j=1}^n e^{\|W_j\| \|x_i\| \cos(\theta_j)}} \quad (15.3)$$

این تابع مستقیماً ویژگی استخراج شده را برای اعمال شباهت بالاتر برای نمونه‌های درون کلاس و فاصله بیشتر برای نمونه‌های بین

کلاسی بینه نمی‌کند، که منجر به ایجاد مشکل در عملکرد آن برای تشخیص چهره عمیق تحت تغییرات ظاهری بزرگ درون کلاس

می‌شود (به عنوان مثال تغییرات زاویه چهره و تغییرات سنی). رابطه فوق را مبنای محاسبات قرار داده‌اند و برای سادگی مقدار بایاس را

$W_j^T x_i = 0$ صفر در نظر گرفتند. سپس حاصل ضرب مقادیر موجود در ویژگی‌های استخراج شده و آخرين لایه کاملاً متصل را به صورت

(θ_j زاویه بین وزن j ویژگی x_i است. نرمال‌سازی ویژگی‌ها و وزن‌ها باعث می‌شود که

خروجی فقط به زاویه بین ویژگی و وزن بستگی داشته باشد. به کمک نرم‌السازی مقادیر وزن $\|W_j\|$ را برابر ۱ در نظر می‌گیریم.

همچنین ویژگی استخراج شده $\|x_i\|$ را نرمال کرده و یک ضریب ثابت s برای آن در نظر می‌گیریم. بنابرین ویژگی‌های استخراج شده

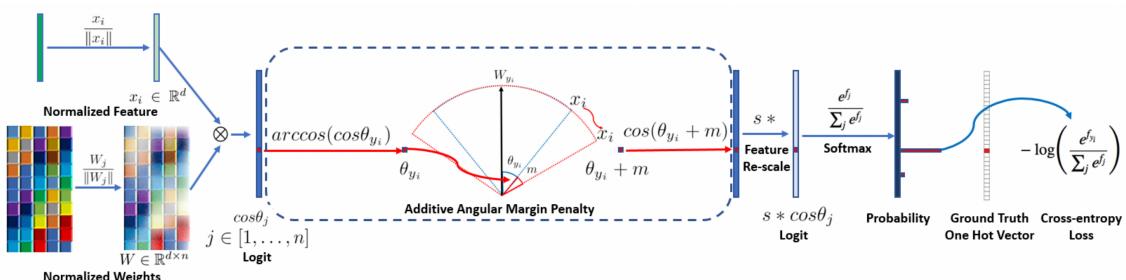
در یک ابر کره با شعاع s توزیع می‌شوند. برای افزایش حاشیه بین x_i و W_j یک مقدار m اضافه می‌کنیم تا به طور همزمان فشرده

سازی درون کلاسی و اختلاف بین کلاسی را افزایش دهیم. رابطه نهاییتابع ضرر ArcFace به صورت ۱۶.۳ می‌باشد.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s(\cos(\theta_j))}} \quad (16.3)$$

این روش دقت ۹۹,۵۳٪ را بر روی مجموعه داده LFW با معماری ResNet50 بدست آورده است. خلاصه این روش در شکل ۲۲.۳

آمده است.

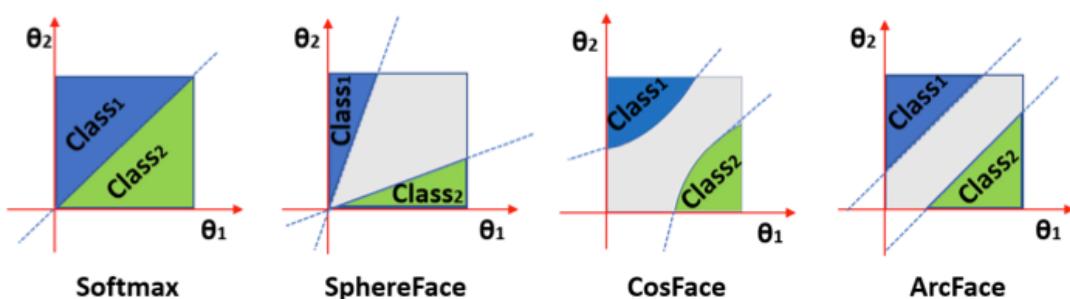


شکل ۲۲.۳: خط لوله ساز و کارتای ضرر Arcface [۳۷].

همانطور که در شکل ۲۳.۳ نشان داده شده است، softmax ویژگی‌های تقریباً قابل تفکیکی ایجاد می‌کند اما در مرزهای تصمیم

گیری ابهام قابل توجهی به وجود می‌آید، در حالی که تابع ضرر ArcFace می‌تواند فاصله بیشتری را بین دسته‌های نزدیک اعمال

کند.



شکل ۲۳.۳: تابع ضرر ArcFace در مقایسه با توابع ضرر مشهور دیگر [۳۷].

۸.۳ نتیجه‌گیری

بیشتر سامانه‌های تشخیص چهره عملکردهای قابل قبولی را در محیط‌های کنترل شده ارائه می‌دهند، اما در محیط‌های بدون محدودیت و در معرض تخریب شدید تصاویر چهره، عملکرد خوبی ندارند و در کاربردهای واقعی هنوز مسیری طولانی برای بهبود در پیش دارند. از جمله چالش‌های مهم، اساسی و عمومی در سامانه‌های تشخیص چهره می‌توان به موارد زیر اشاره نمود:

- تشخیص چهره در محیطی با تغییرات شدید نورپردازی مانند روز و شب (illumination)
- تغییر زاویه و حالت چهره نسبت به دوربین (pose)
- انسداد صورت توسط اشیایی مانند عینک آفتابی و شال گردن (occlusion)
- تغییرات اساسی در چهره با گذر زمان، مانند رشد موها و ریش‌ها و یا بالا رفتن سن مانند سفید شدن موها (aging)
- تاری خارج از تمرکز دوربین (blurring)
- وضوح پایین تصویر (low resolution)
- ردیابی چهره در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی (face tracking)

دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چلپ نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده‌های آموزشی بسیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویربرداری شده باشد.

مقاله [۲۵] روشنی برای رو به رو سازی تصویر چهره پیشنهاد کرده بود که در برخی موارد، چهره را به خوبی می‌چرخاند، اما در نیمی از موقع نیز نتیجه خروجی الگوریتم، تصویر چهره را دچار اعوجاج‌هایی می‌نماید که روند تشخیص چهره را با مشکل بیشتری مواجه می‌سازد. از این رو فرایند رو به رو سازی به طور میانگین کمک شایانی به بالا رفتن دقت تشخیص چهره نمی‌نماید.

مقاله [۳۰] روشنی مبتنی بر GAN برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی مواقع به تصویر چهره لطمه وارد می‌نماید به طوری که شخص مورد نظر قابل شناسایی توسط سامانه یادگیری نمی‌باشد.

مقاله [۳۱] در تولید تصاویر با وضوح بالا بسیار موفق عمل کرده است. اما سایر موارد چالش برانگیز را مورد توجه قرار نداده است. برای مثال اصلاح نورپردازی و زاویه چهره را نادیده گرفته است.

در مقاله [۳۳] چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره روبرو بدون عارضه تولید شده است. این چالش به خاطر کمود تصاویر با عارضه پوستی در مجموعه داده می‌باشد و شبکه در مواجهه با این مسئله راه کاری ارائه نمی‌دهد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند. به تازگی یادگیری عمیق در تشخیص چهره و بسیاری از زمینه‌های هوش مصنوعی به راه حل غالب تبدیل شده است. ما یک سوال مطرح می‌کنیم: آیا یادگیری عمیق واقعاً مسئله تشخیص چهره را حل می‌کند؟ چالش روش‌های یادگیری عمیق در تشخیص چهره چیست؟ در مقایسه با تشخیص شیء عمومی، تشخیص چهره به دلیل طیف گسترده‌ای از تغییرات در ظاهر چهره‌ها چالش برانگیز است. نورپردازی کنترل نشده، انسداد ناشی از عینک، مو، ریش، کلاه و...، تاری خارج از تمرکز دوربین، کیفیت پایین تصویر، بالا رفتن سن افراد و کمبود داده‌های آموزشی از مواردی می‌باشد که می‌توانند سامانه تشخیص چهره را با مشکل روبرو نمایند. از طرفی اکثر مجموعه داده‌ها تنها شامل چند هزار عکس می‌باشد. یک مجموعه داده حاوی اطلاعات بدون محدودیت و مقیاس بزرگ، سامانه چارچوب چهره را به چالش‌هایی همچون گرایش‌های شدید، نور کم و تصاویر کوچک و تاریک چهره تبدیل می‌کند. محققان فرض کرده‌اند که لایه‌های عمیق CNN‌ها می‌توانند اطلاعات انتزاعی بیشتری مانند هویت، ظاهر و ویژگی‌ها را رمزگذاری کنند؛ با این حال هنوز هنوز کاملاً مطالعه نشده است که لایه‌ها دقیقاً با ویژگی‌های محلی برای تشخیص مطابقت دارند.

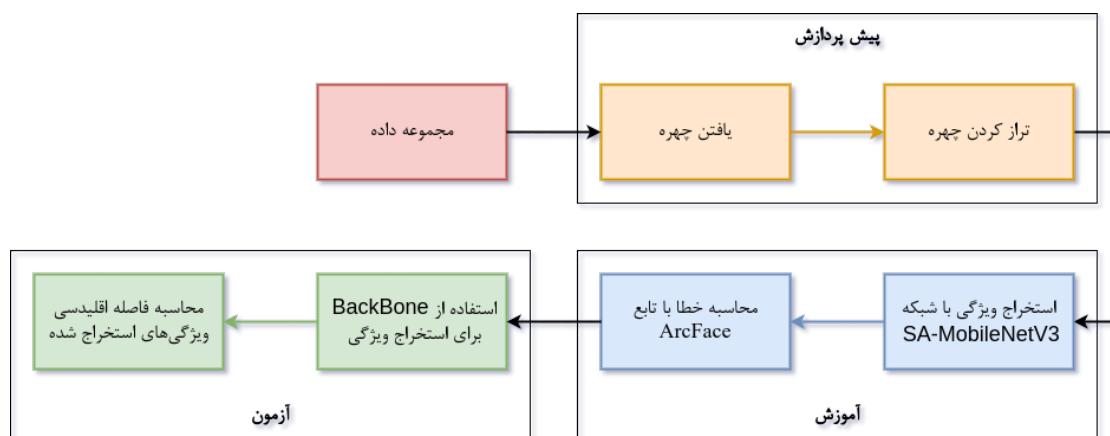
برای شناسایی چهره، عملکرد یادگیری را می‌توان با یادگیری یک معیار اندازه‌گیری فاصله متمایز کننده بهبود داد. با این حال، با توجه به محدودیت‌های حافظه کارت گرافیک‌ها، نحوه انتخاب جفت‌ها یا سه گانه‌های اطلاعاتی و روش‌های آموزش آنلاین (به عنوان مثال، گرادیان نزولی) در مجموعه داده‌های بزرگ، هنوز یک مشکل باز است. یکی دیگر از مشکلات چالش برانگیز این است که پردازش ویدیو در شبکه‌های عمیق را برای استفاده از تجزیه و تحلیل چهره مبتنی بر ویدئو ترکیب کند.

فصل ۴

روش پیشنهادی

۱.۴ مقدمه

شناسایی بی‌درنگ چهره در محیط بدون محدودیت و با دقت بالا با چالش‌های بسیاری همراه است. همچنین دقت بالا و زمان پردازش سریع باهم در تقابل هستند. علاوه بر این‌ها، فرض کمبود داده آموزشی نیز چالش بزرگی محسوب می‌شود. بنابراین در این فصل تلاش می‌کنیم تا روشی برای تشخیص دقیق‌تر و بی‌درنگ چهره توسط شبکه عصبی عمیق در تصاویر بدون محدودیت پیشنهاد دهیم. در ابتدا مرحله پیش‌پردازش شرح داده می‌شود. در قسمت بعد، رویکرد استفاده شده مبتنی بر شبکه‌های پیچشی به منظور استخراج ویژگی از تصاویر چهره شرح داده خواهد شد. نمای کلی روش ارائه شده در شکل ۱.۱ خلاصه شده است که در ادامه هر یک تشریح خواهد شد.



شکل ۱.۴: نمای کلی از روش پیشنهادی.

۲.۴ پیش‌پردازش

بیشتر الگوریتم‌های شناسایی چهره نیاز به اعمال پیش‌پردازش‌هایی بر روی تصویر ورودی دارند. در این روش، پیش‌پردازش شامل همسان‌سازی بافت‌نگار به منظور افزایش تباين، یافتن چهره و تراز کردن تصویر می‌باشد. در ادامه به شرح مراحل پیش‌پردازش می‌پردازیم.

۱.۲.۴ همسان سازی بافت‌نگار

یکی از روش‌های بهبود تصویر، افزایش تباين تصویر است. برای بهبود تباين، می‌توان از تکنیک یکنواخت‌سازی بافت‌نگار تصویر استفاده کرد. این عمل باعث افزایش تباين تصویر می‌شود که به معنای بهبود کیفیت تصویر و افزایش دقت پردازش‌های بعدی است. در سال ۲۰۱۸ چو و همکاران [۳۸] یک الگوریتم بهبود تباين مبتنی بر یکنواخت‌سازی بافت‌نگار ارائه دادند که در این بخش از آن استفاده می‌کنیم.

۲.۲.۴ یافتن چهره

برای یافتن چهره‌ها در تصویر از الگوریتم RetinaFace که توسط Deng و همکاران [۱۰] در سال ۲۰۱۹ معرفی شده است استفاده می‌کنیم. ما برای رسیدن به خروجی بی‌درنگ، این روش را بر روی معماری MobileNetV3 پیاده سازی کرده و آموزش دادیم. نمونه‌ای از خروجی این روش را در شکل ۲.۴ مشاهده می‌کنید.



شکل ۲.۴: نمونه‌ای از خروجی الگوریتم یافتن چهره

۳.۲.۴ تراز کردن تصویر

در ادامه روند پیش‌پردازش، نوبت به تراز کردن تصویر چهره^۱ می‌رسد. پس از یافتن چهره، به تصویر ورودی مناسب شبکه نزدیک‌تر می‌شویم، اما پس از تراز کردن تصویر جهت آموزش شبکه، بهبود دقت نهایی مشهود است. بدین منظور با استفاده از یک تبدیل غیرخطی، تصویر چهره را به گونه‌ای می‌چرخانیم که چشم‌ها در راستای خط افقی قرار بگیرند. روش‌های کلی برای تشخیص چهره از زاویه‌ی رو به رو به خوبی عمل می‌کنند اما مقاومت این روش‌ها در مقابل تغییرات زاویه مناسب نیست، به این علت که ویژگی‌های ظاهری با تغییرات زاویه بسیار تغییر پذیر هستند. با تراز کردن تصویر چهره پیش از اعمال طبقه‌بندی می‌توان این مشکل را بهبود داد. در طول تراز کردن تصویر، نقاط خاصی از تصویر (مانند نقطه وسط دو چشم و نقاط دو طرف دهان) در نظر گرفته می‌شود و به مختصات مشخصی منتقل می‌شوند. برای این منظور از ۵ نقطه ویژه استخراج شده در مرحله یافتن چهره توسط الگوریتم retina استفاده می‌کنیم. نتیجه اعمال این فرآیند را در شکل ۳.۴ مشاهده می‌کنید. در نهایت تصاویر چهره با اندازه 112×112 پیکسل ذخیره می‌گردند تا در مرحله آموزش شبکه، مورد استفاده قرار گیرند.



شکل ۳.۴: به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد.

۳.۴ دسته بندی

روش پیشنهادی برای تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت، یک الگوریتم مبتنی بر شبکه‌های پیچشی می‌باشد. این فرآیند را می‌توان به صورت یک مسئله بهینه سازی فرموله کرد که در ادامه به بخش‌های مختلف آن می‌پردازیم.

¹Face Alignment

در این بخش ابتدا باید معماری مناسب شبکه پایه برای مسئله را به دست آوریم. با بررسی شبکه‌های متداول و مقایسه دقیق و زمان پاسخگویی آن‌ها به کمک یادگیری انتقال، به این نتیجه می‌رسیم که شبکه MobileNetV3 دارای چگالی دقیق بالاتری در مقایسه با شبکه‌های دیگر می‌باشد و نسبت دقیق دسته‌بندی به تعداد پارامترهای شبکه در آن بیشتر می‌باشد. بنابرین می‌توان سرعت اجرای مناسب و همچنین دقیق مناسب را از این شبکه انتظار داشت. از نتایج در می‌یابیم که بهترین معماری شبکه برای مسئله ما معماری MobileNetV3 است. نتایج آزمایش در جدول^۱ آمده است. آزمایش‌هایی بر روی معماری‌های مطرح دیگر نیز انجام شد که به علت ضعیف بودن نتایج یا بالا بودن زمان پاسخ دهنده در جدول درج نشده‌اند.

جدول ۱.۴: مقایسه و ارزیابی برخی از معماری‌های رایج در زمینه بینایی ماشین

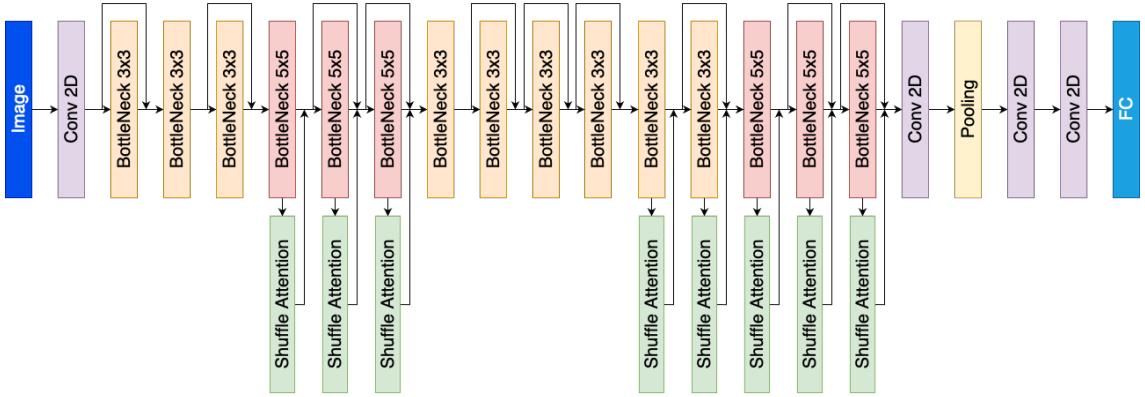
نام شبکه	تعداد پارامترها	دقیق بر روی مجموعه داده LFW
MobileNetV2	3.53M	93.50
MobileNetV3	2.5M	95.8
SqueezeNet	1.25M	89.2
NASNetMobile	5.32M	90.60
EfficientNetB0	5.3M	85.50

تشخیص چهره به دلیل وجود پیکسل‌های مشابه از نظر شدت روشنایی در تصویر بسیار چالش برانگیز بوده است. از آنجا که عملیات کانولوشن به پنجره محلی از شدت روشنایی پیکسل‌های تصویر هدایت می‌شود، بنابراین، این امکان وجود دارد که ویژگی‌های مربوط به پیکسل‌های تصاویر دارای برچسب یکسان، تفاوت‌هایی داشته باشند و یا ویژگی‌های مربوط به پیکسل‌های تصاویر دارای برچسب متفاوت، یکسان باشند. این اختلالات باعث کاهش جدایی‌بردارهای ویژگی خروجی می‌شوند. برای حل این مشکل، از اطلاعات کلی تصاویر به وسیله لایه‌های توجه استفاده می‌کنیم. لایه توجه استفاده شده در این پژوهش، شامل لایه توجه وابسته به کanal و لایه توجه وابسته به موقعیت می‌باشد که به MobileNetV3 اضافه شده است. معماری شبکه پیشنهادی در شکل ۴.۴ آمده است.

در این معماری مسیر استخراج ویژگی از ۴ لایه کانولوشن، ۱ لایه رای گیری، ۱۵ لایه تنگنا^۱ و ۸ مازول توجه^۲ تشکیل شده است که در ادامه توضیح داده خواهد شد. تصویر ورودی با ابعاد 224×224 به بخش استخراج ویژگی داده می‌شود و مدل در این مسیر به طور خودکار یک سلسله مراتب ویژگی را از تصاویر ورودی آموزش خواهد دید و در نهایت این ویژگی‌های استخراج شده به عنوان ورودی

¹Bottleneck

²Attention



شکل ۴: مدل پایه مبتنی بر لایه‌های کانولوشن و لایه توجه.

دسته بند مورد استفاده قرار می‌گیرد.

۱.۱.۳.۴ لایه کانولوشن

۴ لایه کانولوشن دو بعدی همراه با گام^۱ یک یا دو استفاده شده است. در اولین لایه کانولوشن اندازه پنجره فیلترها 3×3 و در لایه‌های کانولوشن بعدی اندازه پنجره فیلترها 1×1 می‌باشد. دلیل انتخاب سایز کوچک پنجره فیلترها کاهش پیچیدگی محاسباتی و همچنین عملکرد خوب آن‌ها در استخراج ویژگی می‌باشد. چگونگی عملکرد یک لایه کانولوشن از رابطه ۱.۴ بدست می‌آید.

$$x_j^{(l)} = \sum_{i=0}^c w_{c_{ij}}^{(l)} * x_i^{(l-1)} + b_j^{(l)} \quad (1.4)$$

که در آن l نشان دهنده شماره لایه کانولوشن، b و w پارامترهای مدل، x خروجی هر لایه $j \in [1, n]$ بیانگر شماره فیلتر موجود در لایه l و همچنین n بیانگر تعداد کل فیلترها در لایه l و نشان دهنده عملکرد کانولوشن می‌باشد.

۲.۱.۳.۴ لایه تنگنا و واحد باقیمانده

به عنوان مثال به جای پردازش یک نگاشت ویژگی عظیم با ۲۵۶ عمق، ابتدا همه این اطلاعات را در نگاشتهای ویژگی ۶۴ بعدی فشرده می‌کنیم. زمانی که این فشردگی انجام شد، از کانولوشن 3×3 استفاده می‌کنیم که وقتی روی ۶۴ نگاشت ویژگی به جای نگاشت اعمال شود، بسیار سریع‌تر است و چنین پردازشی می‌تواند همان نتایج یا نتایج بهتری نسبت به پشتنهای معمولی 3×3 ارائه

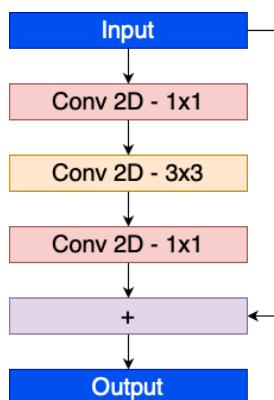
¹Stride

کند. در نهایت با استفاده از 1×1 مجدداً به نگاشت اصلی ۲۵۶ خود بازمی‌گردیم.

از سوی دیگر شبکه‌های عمیق با واحدهای باقیمانده^۱ بر روی پایگاه داده‌های مختلف مانند ImageNet و COCO دقت و همگرایی خوبی را از خود نشان داده‌اند. با استفاده از مسیرهای پرش^۲، واحدهای باقیمانده می‌توانند به سیگنال‌ها اجازه دهنده که مستقیماً از یک بلوک به بلوک‌های دیگر منتقل شوند. به طور کلی، واحدهای باقیمانده را می‌توان به صورت رابطه ۲.۴ بیان کرد.

$$x_{l+1} = x_l + R(x_l, W_l) \quad (2.4)$$

در اینجا R نشان دهنده تابع واحد باقیمانده است، x_l ویژگی ورودی به واحد باقیمانده l ام و W_l مجموعه‌ای از پارامترهای مربوط به واحد باقیمانده l ام می‌باشد. ایده اصلی شبکه‌های باقیمانده، عمیق‌تر کردن یک شبکه به منظور افزایش دقت شبکه مورد نظر می‌باشد. بنابراین با این عملیات در واقع به شبکه اجازه داده می‌شود که در صورت نیاز، ویژگی‌های لایه قبل بدون تغییر و به صورت مستقیم به لایه بعد منتقل شود. در شکل ۵.۴ لایه تنگنا با واحد باقیمانده طراحی شده در این مدل به نمایش گذاشته شده است.



شکل ۵.۴: یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن‌های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می‌کند.

۳.۱.۳.۴ واحد توجه

لایه توجه استفاده شده در این پژوهش یک واحد توجه SA^۳ است که در مقاله [۲۱] معرفی شده است. این لایه ابتدا ابعاد کانال را به چندین ویژگی فرعی قبل از پردازش موازی آنها تقسیم می‌کند. سپس، برای هر زیر ویژگی از یک واحد Shuffle برای به تصویر کشیدن وابستگی‌های ویژگی در هر دو بعد مکانی و کanal استفاده می‌کند. پس از آن، همه زیر ویژگی‌ها جمع می‌شوند و یک عملگر تغییر

¹Deep Residual Network

²Shortcut Pathway

³Shuffle Attention

کanal برای امکان برقراری ارتباط اطلاعاتی بین ویژگی های فرعی مختلف به کار گرفته می شود. معماری این ماژول در شکل ۲۸.۲ آمده است. این لایه شامل دو بخش اصلی وابسته به کanal^۱ و وابسته به موقعیت^۲ می باشد.

یک نقشه ویژگی^۳ به نام X با ابعاد $C \times H \times W$ که در آن C و H و W به ترتیب عمق کanal، ارتفاع و عرض هستند، به عنوان ورودی واحد توجه در نظر گرفته می شود. ابتدا X به گروه در طول کanal تقسیم می شود. رابطه ۳.۴ این موضوع را نشان می دهد. سپس هر X_k در طول کanal به دو نیم تقسیم شده و X_{k1} و X_{k2} را تشکیل می دهند. یکی از این دو زیر نقشه ویژگی^۴ برای توجه وابسته به کanal و دیگری برای توجه وابسته به موقعیت مورد استفاده قرار می گیرد. این عملیات توجه به مفهوم چه چیزی؟ و کجا؟ معنا می بخشد.

$$X = [X_1, \dots, X_G], X_k \in \mathbb{R}^{C/G \times H \times W} \quad (3.4)$$

برای اتصال صحیح ماژول واحد توجه به معماری MobileNetV3 مقدار G را برابر^۴ در نظر گرفتیم.

توجه وابسته به کanal: با بهره گیری از وابستگی های بین کanal ها، می توان بر ویژگی های وابسته، تأکید کرده و استخراج ویژگی های معانی خاص را بهبود بخشد. بنابراین در این پژوهش یک ماژول توجه وابسته به کanal ایجاد شده تا به طور واضح وابستگی بین کanal ها را مدل کند. ابتدا نقشه ویژگی^۵ X_{k1} که از مرحله قبل بدست آمده است، توسط یک لایه پولینگ میانگین گیر سراسری^۵ به یک نقشه ویژگی s با ابعاد $1 \times 1 \times C/2G$ تبدیل می شود. رابطه ۴.۴ نشان دهنده چگونگی ایجاد بردار نقشه توجه وابسته به کanal می باشد.

$$s = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j) \quad (4.4)$$

با انجام این عملیات، اطلاعات مکانی کلی برای هر کanal به صورت جداگانه محاسبه می شود و در s قرار می گیرد. سپس خروجی لایه توجه وابسته به کanal مطابق رابطه ۵.۴ بدست می آید.

$$X'_{k1} = \sigma(W_1 s + b_1).X_{k1} \quad (5.4)$$

¹Channel Attention Module

²Spacial Attention Module

³feature map

⁴sub feature map

⁵Global Averaging Pooling

که در آن $b_1 \in R^{C/2G \times 1 \times 1}$ و $W_1 \in R^{C/2G \times 1 \times 1}$ به ترتیب نشان دهنده وزن و بایاس می‌باشند. سپس نقشه توجه وابسته

به کanal بدست آمده در نقشه ویژگی‌های ورودی ضرب شده و خروجی حاصل را ارائه می‌دهد.

توجه وابسته به موقعیت: وجود ویژگی‌های متمایز برای درک تصویر ورودی ضروری می‌باشد، که این ویژگی‌ها می‌توانند در

باذه بزرگی از تنوع قرار بگیرند. به منظور مدل سازی روابط مبتنی بر روی ویژگی‌های محلی، مازول توجه وابسته به موقعیت استفاده

می‌شود. ابتدا ورودی $X_{k,2}^1$ توسط یک عملگر GN^۱ نرمال سازی می‌شود و خروجی این مرحله طبق رابطه ۶.۴ بدست می‌آید.

$$X'_{k,2} = \sigma(W_2GN(X_{k,2}) + b_2) \cdot X_{k,2} \quad (6.4)$$

که در آن $b_2 \in R^{C/2G \times 1 \times 1}$ و $W_2 \in R^{C/2G \times 1 \times 1}$ به ترتیب نشان دهنده وزن و بایاس می‌باشند. سپس نقشه توجه وابسته به

کanal بدست آمده در نقشه ویژگی‌های ورودی ضرب شده و خروجی حاصل را ارائه می‌دهد. در نهایت نقشه ویژگی‌های $X'_{k,1}$ و $X'_{k,2}$

در کنار هم قرار گرفته و نقشه ویژگی $X'_k \in \mathbb{R}^{C/G \times H \times W}$ بدست می‌آید.

تجمع: پس از مراحل بالا، تمام زیر ویژگی‌های X_k جمع می‌شوند و سرانجام یک عملگر مخلوط کننده کanal^۲ را برای ادغام

اطلاعات گروه‌ها در طول بعد کanal به کار می‌بریم. خروجی نهایی واحد توجه به همان اندازه X است که یکپارچه سازی SA با

معماری‌های مدرن را آسان می‌کند.^۳ [۲۱].

۲.۳.۴ تابع ضرر

یکی از چالش‌های اصلی در یادگیری ویژگی‌ها با استفاده از شبکه‌های عصبی پیچشی عمیق (DCNN)^۳ برای شناسایی چهره در

مقیاس بزرگ، طراحی تابع ضرر مناسب است که قدرت تفکیک را افزایش دهد. از آن جایی که شبکه MobileNetV3 معماری بسیار

سبک تری نسبت به معماری‌های شناخته شده دیگر که در فصل ۲ معرفی شدند، دارد؛ بنابراین استخراج ویژگی‌ها از چهره و دسته

بندی تصاویر چهره با دقت بالا برای این شبکه بسیار دشوار است. همچنین در مواردی شباهت چهره افراد به یکدیگر کار را از آن چه

هست سخت‌تر خواهد کرد. تابع ضرر ArcFace^{۳۷} [۳۷] کمک می‌کند ویژگی‌های استخراج شده که متعلق به دو دسته متفاوت هستند،

فاصله بیشتری از هم داشته باشند و در مقابل ویژگی‌های استخراج شده برای دو تصویر از چهره یک فرد یکسان، فاصله کمتری از هم

داشته باشند؛ این تابع ضرر که می‌توان آن را به راحتی با هزینه‌های محاسباتی ناچیز پیاده سازی کرد، به کاهش مشکلات ذکر شده

¹Group Norm

²channel shuffle

³Deep convolutional neural network

کمک می‌نماید. رابطه نهایی تابع ضرر ArcFace به صورت ۱۶.۳ می‌باشد.

در این تابع مقدار حاشیه m فقط برای زاویه مقدار مربوط به برچسب صحیح اعمال می‌شود و باقی مقادیر بردار دست نخورده باقی

می‌مانند. از آنچا که در فاصله بین 0 تا π هرچه زاویه θ افزایش پیدا کند، باعث می‌شود $\cos(\theta)$ کاهش پیدا کند، بنابرین همگرایی

شبکه با کندی موافق خواهد شد. ما این مشکل را اصلاح کردیم و مقدار حاشیه m را به تمام مقادیر بردار خروجی اضافه می‌نماییم.

تابع نهایی به صورت رابطه ۷.۴ نوشته می‌شود.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{\sum_{j=1}^n e^{s(\cos(\theta_j + m))}} \quad (7.4)$$

حاصل ضرب وزن‌ها در ویژگی‌های استخراج شده محاسبه می‌گردد، که برابر $\arccos(\theta_j)$ می‌شود. سپس آن محاسبه شده

که مقدار θ_j را به ما می‌دهد. سپس برای افزایش حاشیه بین x_i و W_j یک مقدار حاشیه زاویه‌ای m به تمام زاویه‌ها اضافه می‌کنیم

تا به طور همزمان فشرده سازی درون کلاسی و اختلاف بین کلاسی را افزایش دهیم. در انتهای دوباره $\cos(\theta_j)$ محاسبه شده و در

ثابت s ضرب می‌شود. مراحل بعدی دقیقاً مانند ArcFace هستند. در این روش بنا به پیشنهاد مقاله [۳۷] مقدار $m = 0.5$ و مقدار

$s = 64$ در نظر گرفته شده است. مزایای این روش پیشنهادی را می‌توان به شرح زیر خلاصه کرد:

- در مجموعه داده‌های تصویر و فیلم در مقیاس بزرگ، به عملکرد مناسبی دست می‌یابد. زیرا ویژگی‌هایی که استخراج می‌شود،

علاوه بر اینکه قابل تفکیک^۱ هستند، متمایز کننده^۲ نیز هستند، و کمک می‌کند ویژگی‌های استخراج شده که متعلق به دو

دسته متفاوت هستند، فاصله بیشتری از هم داشته باشند و در مقابل ویژگی‌های استخراج شده برای دو تصویر از چهره یک

فرد یکسان، فاصله کمتری از هم داشته باشند.

- فقط به چندین خط کد نیاز دارد و اجرای آن در چارچوب‌های^۳ یادگیری عمیق مبتنی بر Pytorch و Tensorflow آسان

است. برای داشتن عملکرد پایدار نیازی به ترکیب با سایر توابع ضرر ندارد و به راحتی همگرا می‌شود.

- هنگام آموزش فقط پیچیدگی محاسباتی ناچیز را اضافه می‌کند. پردازندۀ های گرافیکی کنونی می‌توانند به راحتی از هزاران

دسته مختلف برای آموزش پشتیبانی کنند و مدل به راحتی می‌تواند هویت‌های بیشتری را پشتیبانی کند.

¹Separable

²Discriminative

³FrameWork

۳.۳.۴ آموزش مدل و استخراج ویژگی

استفاده از روش پیشنهادی ما کمک می‌کند تا تعداد دسته‌ها پس از آموزش قابل تغییر باشد و همچنین برای یادگیری تک تصویری^۱ بسیار مناسب است. در این بخش به آموزش روش پیشنهادی می‌پردازیم. مجموعه داده خود را پس از پیش‌پردازش و افزایش داده‌ها^۲، آمده می‌کنیم. برای افزایش داده‌ها از چرخش افقی^۳ استفاده می‌کنیم. داده‌های آموزش متفاوت را با رگزاري کردیم و آموزش را در ۳۰ دوره^۴ انجام داده‌ایم. در ابتدای روند آموزش، لازم است پارامترهای مدل مقدار دهی اولیه شوند. انتخاب پارامترهای اولیه میتواند تأثیر زیادی در مدل آموزش یافته داشته باشد. در این پژوهش به منظور مقدار دهی اولیه پارامترها ازتابع توزیع یکنواخت استفاده شده است.

از دیگر چالش‌های اساسی برای روش‌های بهینه سازی مبتنی بر گرادیان، انتخاب میزان نرخ یادگیری مناسب است. روش‌های کلاسیک گرادیان تصادفی از نرخ یادگیری ثابت یا کاهشی استفاده میکنند، که برای همه پارامترهای مدل یکسان است. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف می‌توانند از نظر مقدار متفاوت باشند، که می‌تواند به نرخ یادگیری مختلفی نیاز داشته باشد. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف می‌توانند از نظر مقدار تفاوت قابل توجهی داشته باشند، که می‌تواند به نرخ یادگیری مختلفی نیاز داشته باشد. در سال‌های اخیر، تمایل به توسعه روش‌هایی برای انتخاب خودکار نرخ یادگیری مستقل افزایش یافته است. اکثر روش‌ها به عنوان مثال، AdamGrad AdaDelta، RMSprop و Adam آمارهای مختلف مشتقات جزئی را در چندین تکرار جمع آوری می‌کنند و از این اطلاعات برای تعیین میزان یادگیری سازگار برای هر پارامتر استفاده می‌کنند. این امر به ویژه برای آموزش شبکه‌های عمیق بسیار مهم است، جایی که نرخ یادگیری مطلوب اغلب برای هر لایه بسیار متفاوت است. در این پژوهش در آزمایشات انجام شده از همه روش‌های نام برده استفاده شد ولی روش Adam عملکرد بهتری ارائه داده است.

۱.۳.۳.۴ دسته‌بندی

در مرحله آزمون به منظور تشخیص هویت یک تصویر چهره، پس از پیش‌پردازش تصویر را مطابق با ورودی شبکه تغییر اندازه می‌دهیم و جهت استخراج ویژگی به آن شبکه می‌دهیم. پس از استخراج ویژگی‌ها توسط شبکه، بردار ویژگی^۵ ۵۱۲ تایی بدست آمده را با بردارهای مربوط به چهره‌های بانک اطلاعاتی مقایسه کرده و با محاسبه فاصله اقلیدسی بردارها، نزدیک‌ترین شخص مورد نظر انتخاب شده

¹one shot learning

²Data Augmentation

³Horizontal Flip

⁴Epoch

و در صورتی که فاصله میان بردار ویژگی آن‌ها از حد آستانه کمتر باشد، عمل دسته بندی انجام شده و هویت چهره مورد نظر تعیین می‌شود. در غیر این صورت اعلام می‌داریم که شخص مورد نظر قابل شناسایی نمی‌باشد.

۴.۴ فناوری‌های استفاده شده

پیاده‌سازی این الگوریتم به کمک زبان برنامه نویسی پایتون و کتابخانه OpenCV و PyTorch انجام شده است. از کتابخانه‌های مهم مورد استفاده دیگر در این کار می‌توان به NumPy برای انجام محاسبات ماتریسی و SciPy و Scikit Learn اشاره کرد. برای آموزش شبکه عصبی مربوط به دسته بندی ۱۲ گیگابایت حافظه اصلی و ۱۲ گیگابایت حافظه گرافیکی در اختیار گرفتیم.

فصل ۵

ارزیابی روش پیشنهادی

۱.۵ مقدمه

در این کار سعی بر این داشته‌ایم تا به کمک روش‌های یادگیری ژرف در راستای شناسایی چهره در ویدیوهای بدون محدودیت قدمی برداریم و در این حوزه عملکرد هوش مصنوعی و یادگیری ژرف را بهبود دهیم. بنابراین با تحقیق و آزمایش روشی برای دسته بندی دقیق‌تر چهره افراد در تصاویر ویدیویی پیشنهاد داده‌ایم که شرح آن در فصل قبل انجام شد و نوبت آن است که الگوریتم پیشنهادی را ارزیابی کرده و با بیان نتایج به مقایسه با کارهای دیگر می‌پردازیم.

۲.۵ معیار ارزیابی

یکی از معیارهایی که بسیار در زمینه‌های تحقیقاتی و مسائل دسته بندی حائز اهمیت است، معیار دقت می‌باشد. در این کار ما با تصاویر چهره دارای هویت‌های مختلف سروکار داریم. بنابراین معیار دقت در این کار به معنی درصد نمونه‌هایی است که هویت آن‌ها به درستی تشخیص داده شده است. فرمول این معیار در رابطه ۱.۵ بیان شده است.

$$Accuracy = \frac{TP + TN}{P + N} \quad (1.5)$$

که در آن مثبت‌های صحیح^۱ یا TP تعداد نمونه‌هایی است که به درستی تشخیص داده شده‌اند. منفی‌های صحیح^۲ یا TN تعداد نمونه‌هایی است که به درستی تشخیص داده نشده‌اند. مقدار P تعداد کل تصاویر مثبت و N تعداد کل تصاویر منفی را نشان می‌دهد. همانطور که می‌دانیم، به طور معمول معماری‌های شبکه عصبی عمیق‌تر، دارای دقت بالاتر و همچنین زمان پردازش بیشتری می‌باشند. از طرفی هرچه معماری شبکه عصبی، سبک‌تر و تعداد پارامترهای آن کمتر باشد، می‌توان انتظار داشت که سرعت بالاتری در زمان اجرا داشته باشد، اما دقت آن با کاهش همراه است. بر همین اساس در مقابل معیار دقت، معیار دیگری به نام تراکم دقت یا چگالی دقت وجود دارد که از تقسیم مقدار دقت بر تعداد پارامترهای معماری شبکه بدست می‌آید و می‌تواند معیار خوبی برای ارزیابی دقت معماری‌های مختلف نسبت به سرعت اجرای آن‌ها باشد.

$$AccuracyDensity = \frac{Accuracy}{Mparams} \quad (2.5)$$

که در آن $Mparams$ تعداد پارامترهای شبکه بر حسب میلیون می‌باشد.

¹True Positive

²True Negative

۳.۵ مجموعه داده

در زمینه تشخیص چهره، مجموعه داده های بسیار زیادی وجود دارند. تعداد از این مجموعه داده ها، حاوی تصاویر چهره در محیط های آزمایشگاهی و کنترل شده می باشند که در بحث ما گنجانده نمی شوند. با رشد الگوریتم های یادگیری عمیق و بدست آمدن نتایج مناسب در زمینه تشخیص چهره، مجموعه داده های جدیدی منتشر شدند که دارای تصاویر چهره در محیط های بدون محدودیت هستند. برخی از این مجموعه داده ها را برای آموزش و برخی را برای آزمون استفاده کرده ایم که به شرح آن ها می پردازیم.

۱.۳.۵ مجموعه داده های آموزش

۱.۱.۳.۵ CASIA Web-Face مجموعه داده

مجموعه داده Web-Face CASIA که شامل ۴۹۴۴۱۴ تصویر چهره متفاوت از ۱۰۵۷۵ فرد است، پاک، بدون نویز و بدون اشتباه است. این مجموعه داده در مسائل تایید چهره و تشخیص چهره کاربرد دارد. [۶۱]

۲.۱.۳.۵ MS-Celeb-1M مجموعه داده

مجموعه داده MS-Celeb-1M^۱ که به مراتب مجموعه داده بزرگتری محسوب می شود، شامل بیش از ۸ میلیون تصویر چهره متفاوت از ۱۰۰ هزار فرد است. برخلاف مجموعه داده CASIA Web-Face، این مجموعه داده با نویز و برچسب اشتباه همراه است. این مجموعه داده را شرکت مایکروسافت ایجاد کرده است. [۶۲]

۳.۱.۳.۵ VGGFace2 مجموعه داده

محققان دانشگاه آکسفورد نسخه ۲ مجموعه داده VGGFace را با ۳۱۰.۳ میلیون تصویر از ۹۱۳۱ فرد مختلف ارائه کردند. این تصاویر با کمک جستجوی تصویر گوگل جمع آوری شده و شامل تغییرات مختلف برای هر فرد نظیر سن، جهت، نور و ... هستند. این مجموعه داده شامل افراد مختلفی نظیر سیاستمداران، وزشکاران، بازیگران و ... است و به طور تقریبی از هر فرد ۳۶۲ تصویر مختلف موجود است. [۶۳]

^۱Microsoft celebrities

۲.۳.۵ مجموعه داده‌های آزمون

۱.۲.۳.۵ LFW مجموعه داده

مجموعه داده LFW^۱ که شامل ۱۳۲۳۳ تصویر چهره متفاوت از ۵۷۴۹ فرد مختلف است، از اولین مجموعه داده‌های منتشر شده برای مسائل تشخیص چهره بدون محدودیت می‌باشد. [۶۴]

۲.۲.۳.۵ PubFig مجموعه داده

مجموعه داده PubFig^۲ که شامل ۵۸۷۹۷ تصویر چهره متفاوت از ۲۰۰ فرد است، از اینترنت جمع آوری شده است. این تصاویر شامل تغییرات مختلف مانند جهت، نور، انسداد و ... هستند. [۶۵]

۳.۲.۳.۵ YouTube Faces مجموعه داده

این مجموعه داده برای استفاده در کارهای تشخیص چهره در تصاویر ویدیویی ایجاد شده است. این مجموعه داده شامل ۳۴۲۵ ویدیو از ۱۵۹۵ فرد مختلف است و تمام ویدیوها از سایت یوتیوب دانلود شده‌اند. به طور میانگین ۱۵.۲ فیلم برای هر شخص در دسترس است. کوتاه‌ترین مدت ویدیو ۴۸ فریم، طولانی‌ترین ویدیو ۶۰۷۰ فریم و متوسط طول یک ویدیو ۳۰۱۸۱ فریم است. [۶۳]

۴.۲.۳.۵ CFP مجموعه داده

مجموعه داده CFP^۳ که شامل ۷۰۰۰ تصویر چهره متفاوت از ۵۰۰ فرد است، تصاویر افراد مشهور را در حالت‌های تمام رخ و نیم رخ جمع آوری کرده است. این مجموعه داده می‌تواند ابزار ارزیابی بسیار خوبی برای چالش زاویه چهره باشد. [۶۴]

¹Labeled Faces in the Wild

²Public Figures

³Celebrities in Frontal-Profile in the Wild

۵.۲.۳.۵ مجموعه داده CACD

مجموعه داده CACD^۱ شامل ۱۶۳۴۴۶ تصویر از ۲۰۰۰ فرد مشهور است که از اینترنت جمع آوری شده است. تصاویر از موتورهای جستجو با استفاده از نام افراد مشهور و سال (۲۰۰۴-۲۰۱۳) به عنوان کلمات کلیدی جمع آوری شده است. بنابراین، می‌توان با تفربیق سال تولد افراد از سال عکس گرفته شده، به سادگی سن افراد در تصاویر را تخمین زد. این مجموعه داده می‌تواند ابزار ارزیابی بسیار خوبی برای چالش تغییرات سن باشد. [۶۶]

۶.۲.۳.۵ مجموعه داده MegaFace

محققان دانشگاه واشنگتن مجموعه داده MegaFace^۲ عظیم را ارائه کرده اند که هم در مسائل تایید چهره و هم در مسائل تشخیص چهره کاربرد دارد. این مجموعه داده شامل یک زیر مجموعه آزمایش است که خود از دو مجموعه FaceScrub شامل تصاویر افراد مشهور و FGNet شامل تصاویر مخصوص چالش سن، تشکیل شده است. اگر الگوریتم مورد نظر بر روی مجموعه FGNet دقت بالا بدست آورد، نشان دهنده قوت الگوریتم در تصاویر با تفاوت سن‌های بالا می‌باشد [۶۷]. اطلاعات تکمیلی درباره مجموعه داده‌های آزمایش را در جدول ؟؟ مشاهده می‌نمایید.

جدول ۱.۵: مجموعه‌داده‌های ارزیابی رایج در زمینه بازناسی چهره

نام	تعداد تصاویر چهره	تعداد افراد	نوع کاربرد
LFW	۱۳۲۲۳	۵۷۴۹	تشخیص چهره
PubFig	۵۸۷۹۷	۲۰۰	تشخیص چهره
YouTube Faces	۳۴۲۵	۱۵۹۵	تشخیص چهره
CFP	۷۰۰۰	۵۰۰	تشخیص چهره
CACD	۱۶۳۴۴۶	۲۰۰۰	تشخیص چهره
MegaFace	۹۷۵ + ۱۴۱۰۰۰ + ۱۰۰۰۰۰	۸۲ + ۶۹۵ + ۶۹۰۰۰	تشخیص چهره و تایید چهره

¹Cross-Age Celebrity Dataset

²Million-Scale Face Recognition Dataset

۴.۵ پیکربندی الگوریتم

به منظور آموزش شبکه، در هر دوره 20% تعداد داده‌های آموزشی را به عنوان داده‌های ارزیابی^۱ در نظر می‌گیریم تا روند آموزش شبکه را بر اساس عامل‌های دیگر مورد بررسی قرار دهیم. همچنین در هر تکرار، اندازه دسته‌هایی که به شبکه برای آموزش داده می‌شود را برابر 64 قرار دادیم. توصیه می‌شود این مقدار توانی از 2 باشد که مقدار 64 با تجربه و توجه به ظرفیت حافظه پردازنده گرافیکی بدست آمده است. از بهینه‌ساز Adam جهت آموزش استفاده کرده‌ایم. این بهینه‌ساز نیز نرخ آموزش را بر اساس خطا به صورت تطبیقی کم یا زیاد می‌کند. همچنین میانگین کاهش گرادیان‌های تکرارهای قبل را نگهداری می‌کند تا بر اساس آن‌ها جهت گرادیان تکرار جدید را محاسبه کند. برای تابع ضرر از معادله 7.4 استفاده کرده‌ایم. این تابع میزان ضرر را به صورتی که در فصل قبل صحبت شد، محاسبه می‌کند.

همچنین برای به دست آوردن بهترین نتیجه از آموزش، از روش ارزیابی تقاطعی استفاده کرده‌ایم. به این صورت که در هر مرتبه آموزش، تعداد داده‌های آموزش را به پنج دسته تقسیم می‌کنیم. چهار قسمت را برای آموزش و یک قسمت را برای ارزیابی در نظر می‌گیریم. به این ترتیب مدل شبکه را 5 بار آموزش دادیم و بهترین نتیجه را برای آزمون بر روی مجموعه داده‌های آزمون انتخاب می‌کنیم.

۵.۵ نتایج آزمون

۱.۵.۵ دقیقت دسته بندی

با توجه به پیکربندی بیان شده، مدل را آموزش داده‌ایم و به سراغ مجموعه داده‌های آزمون می‌رویم. مقدار دقیقت را طبق رابطه 1.5 برای معماری‌های مرتبط و مشهور و همچنین روش پیشنهادی محاسبه کردیم. در جدول 2 ؟ نتایج حاصل از محاسبه دقیقت را در معماری‌های مختلف بر روی مجموعه داده‌های مشهور مشاهده می‌کنید.

جدول ۲: دقیقت معماری‌های مختلف بر روی مجموعه داده‌های ارزیابی رایج در زمینه بازنگاری چهره

ImageNet	Cifar100	Mnist	MegaFace	AgeDB-30	LFW	Custom Dataset	Model
			93.09	95.91	99.50	99.70	ArcFace
75.2	80.01	99.7	90.59	93.05	98.10	92.50	MobileNetV3
79.8	82.47	99.9	93.50	96.02	99.65	99.80	SA-MobileNetV3

ایده پایه در این پژوهش، طراحی یک عینک هوشمند با قابلیت شناسایی چهره برای افراد نایبینا بود. در یک کاربرد واقعی، تعداد افراد

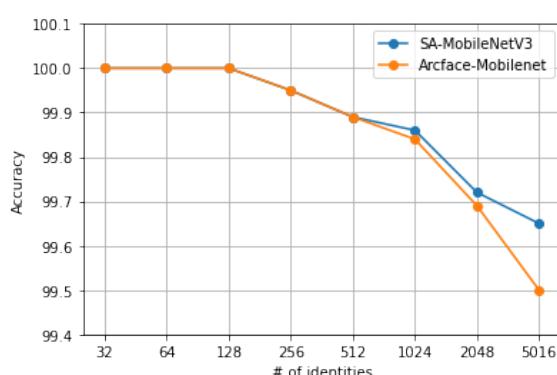
¹Validation

آشنا برای یک شخص، محدود می‌باشد. بر همین اساس با در آزمونی دیگر با تعداد محدود داده‌های آموزشی، روش پیشنهادی را با روش دنگ و همکاران [۳۷] معروف به ArcFace مقایسه کردیم. در این ارزیابی که بر روی مجموعه داده LFW انجام شد، تعداد افراد آشنا را به ترتیب برابر با ۳۲، ۶۴، ۱۲۸ و ... در نظر گرفتیم و تعداد افراد ناشناس (افراد منفی) را برابر با تعداد افرادی که فقط یک تصویر از آن‌ها موجود می‌باشد، یعنی ۴۰۶۹ در نظر گرفتیم. نتایج ارزیابی در جدول؟ مشاهده می‌شود.

جدول ۳.۵: دقت بر روی مجموعه داده LFW بر حسب تعداد متفاوت دسته‌ها

Accuracy (Arcface-MobileNet)	Accuracy (SA-MobileNetV3)	Number of identites
100	100	32
100	100	64
99.97	100	128
99.94	99.95	256
99.89	99.89	512
99.86	99.86	1024
99.69	99.72	2048
99.50	99.65	5016

مقادیر بدست آمده در جدول؟ را به صورت منحنی بر روی نمودار در شکل؟ رسم نمودیم. همانطور که مشاهده می‌شود، روش پیشنهادی در مقایسه با روش ArcFace از بهبود نسبی برخوردار می‌باشد.



شکل ۱.۵: دقت روش پیشنهادی در مقایسه با روش ArcFace

۲.۵.۵ سرعت پاسخ‌دهی

با توجه به اینکه هدف اصلی روش پیشنهادی، اجرا بر روی تلفن همراه هوشمند می‌باشد، علاوه بر ارزیابی دقت دسته بندی، از لحاظ

پیچیدگی شبکه، تعداد پارامترها، سرعت پاسخ‌دهی و زمان اجرای عملیات پیشرو^۱ نیز روش پیشنهادی را در مقایسه با سایر روش‌ها

^۱Forward

سنجدیده ایم. در جدول ؟؟ نتایج حاصل از محاسبه پیچیدگی شبکه و تعداد پارامترها را در معماری های مختلف بر روی مجموعه داده ImageNet مشاهده می کنید.

جدول ٤.٥: پیچیدگی شبکه و تعداد پارامترهای معماری های مختلف بر روی مجموعه داده ImageNet

Madds	Accuracy density	Number of Parameters	Accuracy (imageNet)	Model
8220 M	3.32	25 M	83.0	ResNet50
627.69 M	20.73	3.5 M	72.56	MobileNetV2
448.69 M	13.92	5.4 M	75.2	MobileNetV3
445.68 M	21.06	3.8 M	79.8	SA-MobileNetV3

برای مقایسه زمان اجرای الگوریتم روش پیشنهادی با روش ArcFace آزمایش سرعت را بر روی دستگاه Raspberry Pi اجرا کردیم.

مشخصات سخت افزاری این دستگاه به شرح زیر می باشد:

Model: Raspberry Pi 4 Model B •

CPU: Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz •

Memory: 4GB •

Operating System: Raspberry Pi OS, Kernel version: 5.10 •

نتایج ارزیابی را به دو صورت ارائه کردیم. در حالت اول فقط زمان اجرای یک بار عملیات پیشرو در شبکه تشخیص چهره را گزارش دادیم و در حالت دوم زمان اجرای کامل خط لوله را گزارش دادیم. یک خط لوله کامل به معنای ضبط فریم، پیش پردازش، یافتن چهره، تراز کردن چهره و تشخیص چهره می باشد. در جدول ؟؟ نتایج حاصل از محاسبه سرعت آمده است.

جدول ٥.٥: نتایج زمانی بر روی Raspberry Pi به ازای هر فریم

recognition	فقط زمان اجرای شبکه	زمان اجرای کل خط لوله	Model
25 fps		6 fps	SA-MobileNetV3
20 fps		5 fps	Arcface-Mobilenet

با توجه به نتایجی که در جدول ها و شکل ها ارائه شد، متوجه می شویم که روش پیشنهادی ما از لحاظ دقیق دسته بندی و سرعت پردازش نسبت به سایر روش های جدید در این زمینه بهتر عمل می نماید. ما محدودیت منابع تلفن همراه را در نظر گرفتیم و بر مبنای آن روشی ارائه دادیم که علاوه بر پیچیدگی کم و سرعت مناسب، دارای دقیق دسته بندی بسیار بالایی باشد. معماری شبکه عصبی ارائه شده در این پژوهش، علاوه بر مسئله تشخیص چهره، می تواند در سایر مسائل و مجموعه داده های عمومی نیز کارآمد باشد.

فصل ٦

نتیجه‌گیری و پیشنهادات

در این بخش جمع‌بندی و نتیجه‌گیری این پژوهش بیان می‌شود. هدف کار ارائه شده در این پژوهش ارائه و بهبود روش‌های خودکار مبتنی بر هوش مصنوعی و یادگیری ژرف به منظور تشخیص چهره به صورت بی‌درنگ در شرایط بدون محدودیت در تصاویر ویدیویی است.

۲.۶ بحث و نتیجه‌گیری

همانطور که در فصل‌های قبل نیز بیان شد، روش‌های مختلفی برای یافتن چهره و شناسایی چهره ارائه شده‌اند. برای حل مساله دسته‌بندی چهره دو روش کلی، مبتنی بر تصویر و روش‌های مبتنی بر استخراج ویژگی وجود دارد. روش‌های مبتنی بر تصویر خود دارای رویکردهای مختلفی از جمله روش‌های مبتنی بر رنگ بندی، روش‌های مبتنی بر شکل و روش‌های مبتنی بر گرادیان می‌باشد. دسته‌ای از این روش‌ها سعی کرده‌اند تا به کمک روش‌های غیر ژرف و استخراج ویژگی‌های معنادار از پیکسل‌های چهره به این هدف دست یابند و در نهایت یک دسته‌بند را با ویژگی‌های ارائه شده آموزش دهند و از این دسته‌بند برای آزمون استفاده کنند.

همچنانی روش‌های مبتنی بر استخراج ویژگی که در سال‌های اخیر بسیار مورد توجه قرار گرفته اند شامل رویکردهای مبتنی بر شبکه عصبی، بردار پشتیبان و... می‌باشند. این روش‌ها با تکیه بر ویژگی‌های ژرف استخراج شده از شبکه‌های عصبی پیچشی سعی دارند چهره‌ها را در تصاویر و ویدیو شناسایی کنند. به این صورت که تصاویر آموزشی را به شبکه می‌دهند و شبکه در مرحله آموزش سعی می‌کند ویژگی‌های لازم و تفکیک‌پذیر بین چهره افراد مختلف را شناسایی کند. واضح است که ویژگی‌های استخراج شده توسط شبکه با ویژگی‌های معنادار استخراج شده توسط انسان همیشه برای نیستند و حتی ممکن است ویژگی‌های شبکه اصلاً قابل تفسیر نباشند. در سال‌های گذشته، انواع شبکه‌های عمیق در زمینه دسته‌بندی تصاویر چهره به شدت مورد استفاده قرار گرفته است. مدل‌های مختلف ارائه شده با گرفتن تصاویر مختلف چهره به عنوان ورودی، هویت شخص را به عنوان خروجی تولید می‌کنند. اما روش‌های ارائه شده، دارای محدودیت‌هایی نظیر پایین بودن دقت، تعداد زیاد پارامترهای شبکه‌های ارائه شده، سرعت کم پردازش و همچنانی تعداد زیاد تصاویر ورودی برای آموزش می‌باشند. ما با توجه به شرایط بی‌درنگ و محدودیت‌های دیگر مسئله، در میان روش‌های استخراج ویژگی توسط یادگیری ژرف به دنبال معماری سبک‌تر و سریع‌تر بودیم که موفق شدیم با تحلیل و آزمایش به معماری مورد نظر دست پیدا کنیم که علاوه بر پردازش‌های سبک و سرعت بالا، دارای دقت لازم و کافی در دسته‌بندی نیز باشد.

۳.۶ پیشنهادات

به نظر می‌رسد افزودن لایه‌های خاص منظوره به معماری شبکه‌های عصبی معروف، باعث افزایش دقت الگوریتم در کاربردهای خاص می‌شود. به منظور بھبود در نتایج باید در تغییر لایه‌های معماری شبکه دقت بیشتری داشت. بنابراین برای بھبود ویژگی‌های ژرف، استفاده از ویژگی‌های چند مدل شبکه در کنار هم می‌تواند کمک کننده باشد. به این صورت که از هر مدل، لایه‌های سودمندتر را انتخاب کرده و از کنار هم قرار دادن آن‌ها، بردار ویژگی ما غنی‌تر خواهد شد.

همچنین می‌توان تابع ضرر را بروز رسانی و بهینه کرد و به این ترتیب بردار ویژگی‌هایی به دست خواهد آمد که از فاصله درون دسته‌ای کمتر و فاصله برون دسته‌ای بیشتری برخوردار باشند. ما در طراحی تابع ضرر خود، مقدار پارامتر حاشیه m را ثابت در نظر گرفتیم. می‌توان مقدار بهینه m را به صورت وفقی^۱ طی فرایند آموزش بدست آورد. از طرفی در این میان ممکن است ویژگی‌هایی یکسان از شبکه‌ها استخراج شود، و یا این که ویژگی‌هایی وجود داشته باشند که در تشخیص هویت چهره مورد نظر زیاد نقش موثری نداشته باشند. به این ترتیب استفاده از روش‌های کاهش بردار ویژگی برای استخراج ویژگی‌های اصلی برای ایجاد مرز بهتر بین دسته‌ها می‌تواند مفید باشد. در مقابل با تحقیق می‌توان ویژگی‌های متمایز کننده معنادار دیگری برای چهره به دست آورد تا به بردار ویژگی اضافه شوند. روش‌های مختلفی برای پیش‌پردازش تصاویر چهره وجود دارند. استفاده از روش‌های جدید پیش‌پردازش به منظور کاهش بیشتر نویز، یکسان سازی روش‌نایاب تصاویر بافت‌نگار و ... می‌تواند به استخراج ویژگی‌های مفیدتر کمک کند تا بتوان نتایج بهتر برای این منظور به دست آورد.

در بحث استخراج ویژگی می‌توان با استفاده از لایه‌های بیشتر و همچنین استفاده از مدل‌های دیگری از لایه توجه، ویژگی‌های بهتری استخراج کرد و در نهایت ویژگی‌های استخراج شده منجر به دسته بندی بهتر تصاویر خواهد شد. همچنین به دلیل وجود دسته‌های زیاد در پایگاه داده، می‌توان از تابع ضرری که مبتنی بر تعداد زیاد دسته‌ها باشد استفاده کرد که منجر به آموزش بهتر مدل و همچنین کم شدن اشتباهات دسته بندی می‌شود.

یکی دیگر از مشکلات موجود در حوزه تشخیص چهره، نبود داده کافی به منظور آموزش مدل می‌باشد. به منظور حل این مشکل می‌توان مدلی به منظور تولید تصاویر مصنوعی ارائه داد تا با مشکل نبود داده کافی مقابله کرد و در نهایت مدلی کارآمدتر و بهتر برای تشخیص ارائه داد. و همچنان راه حل‌ها و ایده‌های دیگری می‌توانند در جهت افزایش دقت این سامانه کمک کننده باشند

¹Adaptive

منابع و مأخذ

- [1] "Smart glasses led the blind," <https://www.neotaos365.top/ProductDetail.aspx?iid=246782895&pr=58.99>. Accessed: 2021-04-24.
- [2] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition for uncontrolled pose and illumination changes," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.43, no.1, pp.149–163, 2013.
- [3] B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pp.186–192, 2013.
- [4] M. J. Wilber, V. Shmatikov, and S. Belongie, "Can we still avoid automatic face detection?," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp.1–9, 2016.
- [5] S. Chaudhry and R. Chandra, "Face detection and recognition in an unconstrained environment for mobile visual assistive system," *Applied Soft Computing*, vol.53, pp.168–180, 2017.
- [6] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol.83, no.3, pp.236–274, 2001.
- [7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol.1, pp.I–I, 2001.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol.1, pp.886–893 vol. 1, 2005.
- [9] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.211–223, 2016.
- [10] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-stage dense face localisation in the wild," 2019.
- [11] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," *Pattern Recognition Letters*, vol.28, no.14, pp.1885–1906, 2007. Image: Information and Control.
- [12] JafriRabia and A. R., "A survey of face recognition techniques," *Journal of Information Processing Systems*, vol.5, no.2, pp.41–68, 2009.

- [13] H. Wang, Y. Wang, and Y. Cao, “Video-based face recognition: A survey,” *World Academy of Science, Engineering and Technology*, vol.60, pp.293–302, 2009.
- [14] X. Luan, B. Fang, L. Liu, W. Yang, and J. Qian, “Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion,” *Pattern Recognition*, vol.47, no.2, pp.495–508, 2014.
- [15] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.815–823, 2015.
- [16] G. Wen, Y. Mao, D. Cai, and X. He, “Split-net: Improving face recognition in one forwarding operation,” *Neurocomputing*, vol.314, pp.94–100, 2018.
- [17] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1–9, 2015.
- [18] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [19] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1867–1874, 2014.
- [20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilennets: Efficient convolutional neural networks for mobile vision applications,” 2017.
- [21] Q.-L. Z. Y.-B. Yang, “Sa-net: Shuffle attention for deep convolutional neural networks,” 2021.
- [22] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” *IEEE Access*, vol.6, p.64270–64277, 2018.
- [23] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, “Deep face recognition: A survey,” *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp.471–478, 2018.
- [24] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, “Openface: A general-purpose face recognition library with mobile applications,” tech. rep., CMU-CS-16-118, CMU School of Computer Science, 2016.
- [25] M. Haghigiat, M. Abdel-Mottaleb, and W. Alhalabi, “Fully automatic face normalization and single sample face recognition in unconstrained environments,” *Expert Systems with Applications*, vol.47, pp.23–34, 2016.
- [26] T. Zhang, Q. Dong, and Z. Hu, “Pursuing face identity from view-specific representation to view-invariant representation,” in *2016 IEEE International Conference on Image Processing (ICIP)*, pp.3244–3248, 2016.
- [27] A. V. Savchenko and N. S. Belova, “Unconstrained face identification using maximum likelihood of distances between deep off-the-shelf features,” *Expert Systems with Applications*, vol.108, pp.170–182, 2018.

- [28] C. Y. Wu and J. J. Ding, “Occluded face recognition using low-rank regression with generalized gradient direction,” *Pattern Recognition*, vol.80, pp.256–268, 2018.
- [29] Ya Wang, Tianlong Bao, Chunhui Ding, and Ming Zhu, “Face recognition in real-world surveillance videos with deep learning method,” in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pp.239–243, 2017.
- [30] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2016.
- [31] S. Banerjee and S. Das, “Lr-gan for degraded face recognition,” *Pattern Recognition Letters*, vol.116, pp.246–253, 2018.
- [32] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” 2019.
- [33] J. Cao, Y. Hu, B. Yu, R. He, and Z. Sun, “3d aided duet gans for multi-view face image synthesis,” *IEEE Transactions on Information Forensics and Security*, vol.14, no.8, pp.2028–2042, 2019.
- [34] T. Soyata, R. Muraleedharan, C. Funai, M. Kwon, and W. Heinzelman, “Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture,” in *2012 IEEE Symposium on Computers and Communications (ISCC)*, pp.000059–000066, 2012.
- [35] P. Hu, H. Ning, T. Qiu, Y. Xu, X. Luo, and A. K. Sangaiah, “A unified face identification and resolution scheme using cloud computing in internet of things,” *Future Generation Computer Systems*, vol.81, pp.582–592, 2018.
- [36] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “Cosface: Large margin cosine loss for deep face recognition,” 2018.
- [37] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” 2019.
- [38] S. W. Cho, N. R. Baek, M. C. Kim, J. H. Koo, J. H. Kim, and K. R. Park, “Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network,” *Sensors*, vol.18, no.9, 2018.
- [39] “Evaluation of human biometric features,” <https://en.wikipedia.org/wiki/Biometrics>. Accessed: 2021-04-24.
- [40] Ming-Hsuan Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: a survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.1, pp.34–58, 2002.
- [41] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, “Deep learning for understanding faces: Machines may be just as good, or better, than humans,” *IEEE Signal Processing Magazine*, vol.35, no.1, pp.66–83, 2018.
- [42] R. Brunelli and T. Poggio, “Face recognition: features versus templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.15, no.10, pp.1042–1052, 1993.

- [43] G. Hu, F. Yan, J. Kittler, W. Christmas, C. H. Chan, Z. Feng, and P. Huber, “Efficient 3d morphable face model fitting,” *Pattern Recognition*, vol.67, pp.366–379, 2017.
- [44] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, “Face recognition with visible and thermal infrared imagery,” *Computer Vision and Image Understanding*, vol.91, no.1, pp.72–114, 2003. Special Issue on Face Recognition.
- [45] K. R. Sreelakshmi, R. Anitha, and K. R. Rebitha, “Multiple media based face recognition in unconstrained environments using eigenfaces,” in *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, pp.1–6, 2016.
- [46] “Facial expression recognition,” <http://www.sfu.ca/~llockhar/projects/facial-expression-recognition>. Accessed: 2021-04-24.
- [47] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, “Facial landmark detection with tweaked convolutional neural networks,” 2016.
- [48] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” 2019.
- [49] J.-J. Lv, C. Cheng, G.-D. Tian, X.-D. Zhou, and X. Zhou, “Landmark perturbation-based data augmentation for unconstrained face recognition,” *Signal Processing: Image Communication*, vol.47, pp.465–475, 2016.
- [50] W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, R. Nevatia, and G. Medioni, “Face recognition using deep multi-pose representations,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp.1–9, 2016.
- [51] I. Masi, S. Rawls, G. Medioni, and P. Natarajan, “Pose-aware face recognition in the wild,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4838–4846, 2016.
- [52] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim, “Rotating your face using multi-task deep neural network,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.676–684, 2015.
- [53] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” 2014.
- [54] Xiangyu Zhu, Z. Lei, Junjie Yan, D. Yi, and S. Z. Li, “High-fidelity pose and expression normalization for face recognition in the wild,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.787–796, 2015.
- [55] C. Ding, C. Xu, and D. Tao, “Multi-task pose-invariant face recognition,” *IEEE Transactions on Image Processing*, vol.24, no.3, pp.980–993, 2015.
- [56] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, “Unconstrained face recognition: Identifying a person of interest from a media collection,” *IEEE Transactions on Information Forensics and Security*, vol.9, no.12, pp.2144–2157, 2014.
- [57] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” *Pattern Recognition*, vol.66, pp.144–152, 2017.

- [58] J. Hussain Shah, M. Sharif, M. Raza, M. Murtaza, and Saeed-Ur-Rehman, “Robust face recognition technique under varying illumination,” *Journal of Applied Research and Technology*, vol.13, no.1, pp.97–105, 2015.
- [59] J. Li, B. Li, Y. Xu, K. Lu, K. Yan, and L. Fei, “Disguised face detection and recognition under the complex background,” in *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pp.87–93, 2014.
- [60] L. Di Martino, J. Preciozzi, F. Lecumberry, and A. Fernández, “Face matching with an a contrario false detection control,” *Neurocomputing*, vol.173, 08 2015.
- [61] “Casia web-face dataset,” <https://paperswithcode.com/dataset/casia-webface>. Accessed: 2021-04-24.
- [62] “Ms-celeb-1m,” <https://www.microsoft.com/en-us/research/project/ms-celeb-1m-challenge-recognizing-one-million-celebrities-real-world>. Accessed: 2021-04-24.
- [63] “Vggface2 dataset,” https://www.robots.ox.ac.uk/~vgg/data/vgg_face/. Accessed: 2021-04-24.
- [64] “Labeled faces in the wild dataset,” <http://vis-www.cs.umass.edu/lfw/>. Accessed: 2021-04-24.
- [65] “Public figures face database,” <https://www.cs.columbia.edu/CAVE/databases/pubfig/>. Accessed: 2021-04-24.
- [66] “Cross-age celebrity dataset,” <https://paperswithcode.com/dataset/cacd>. Accessed: 2021-04-24.
- [67] “Megaface: Million-scale face recognition dataset,” <http://megaface.cs.washington.edu>. Accessed: 2021-04-24.

Abstract

Ferdowsi University Mashhad (FUM)
Department of Computer

Thesis submitted
for the degree of M.Sc.

Title:

Realtime Face Recognition in Unconstraint Environments

Supervisor: DR. Hamid Reza Pour Reza

By: Sajjad Aemmi

June 2021