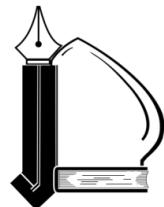


بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشکده مهندسی



دانشگاه فردوسی مشهد
دانشکده مهندسی کروه کامپیوتر
آزمایشگاه بینایی ماشین

دانشگاه فردوسی مشهد

دانشکده مهندسی

گروه مهندسی کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد
هوش مصنوعی

تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت

استاد راهنما: دکتر حمید رضا پور رضا

پژوهش و نگارش: سید سجاد اعمی

۱۳۹۷ مهرماه



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

شناسه: ب/ک/3	صورتجلسه دفاعیه پایان نامه دانشجوی دوره کارشناسی ارشد	 دانشگاه فنی شهرضا مدیریت تحصیلات تکمیلی
دانشجوی کارشناسی ارشد	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 ۱۳۹۷/۰۷/۳۰ با حضور اعضای هیأت داوران (به شرح ذیل) تشکیل گردید.	دانشجوی کارشناسی ارشد
و درجه	پس از ارزیابی توسط هیأت داوران، پایان نامه با نمره به عدد به حروف	مورد تصویب قرار گرفت.
<u>امضاء</u>	<u>نام و نام خانوادگی</u>	<u>عنوان</u>
	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 با وضوح 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
۳	۱۰.۱ ارزیابی ویژگی‌های بیومتریک انسان
۳	۳.۱ سامانه تشخیص چهره
۴	۱۰.۳.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره
۵	۲۰.۳.۱ نمای کلی یک سامانه تشخیص چهره
۶	۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر
۶	۱۰.۴.۱ خوشه‌بندی
۷	۲۰.۴.۱ طبقه‌بندی
۸	۵.۱ مسئله تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده
۸	۱۰.۵.۱ چالش‌های سامانه تشخیص چهره
۱۲	۲ مروری بر کارهای گذشته
۱۳	۱۰.۲ مقدمه

۱۳	یافتن چهره	۲.۲
۱۳	رویکردهای مبتنی بر دانش و تطبیق کلیشه	۱.۲.۲
۱۵	رویکردهای مبتنی بر ویژگی	۲.۲.۲
۱۷	رویکرد مبتنی بر عامل‌های آماری	۳.۲.۲
۲۳	رویکردهای مبتنی بر تصویر	۴.۲.۲
۲۵	شناسایی چهره	۴.۲
۲۵	رویکردهای سنتی	۱.۳.۲
۲۶	رویکرد مبتنی بر فیلتر گابور	۲.۳.۲
۲۶	رویکردهای سه بعدی	۳.۳.۲
۲۷	رویکردهای تجزیه و تحلیل بافت پوست	۴.۳.۲
۲۸	رویکردهای مبتنی بر دوربین حرارتی	۵.۳.۲
۲۸	تشخیص چهره مبتنی بر ویدیو	۶.۳.۲
۳۰	رویکرد مبتنی بر چهره ویژه	۷.۳.۲
۳۱	رویکردهای مبتنی بر شبکه عصبی	۸.۳.۲
۳۶	رویکردهای مبتنی بر نقاط راهنمایی	۹.۳.۲
۴۰	نتیجه‌گیری	۴.۲
۴۲	۳ مروری بر کارهای گذشته در شرایط کنترل نشده	
۴۳	۱.۳ مقدمه	
۴۳	۲.۳ چالش حالت	
۵۳	۲.۳ چالش روشنایی	
۵۵	۴.۳ چالش انسداد	
۵۷	۵.۳ چالش کمبود تصاویر آموزشی	
۶۲	۶.۳ چالش منابع محدود	

۶۸	نتیجه‌گیری	۷.۳
۷۱	روش پیشنهادی	۴
۷۲	مقدمه	۱.۴
۷۲	پیش‌پردازش	۲.۴
۷۳	همسان‌سازی بافت‌نگار	۱.۲.۴
۷۳	یافتن چهره	۲.۲.۴
۷۴	تراز کردن تصویر	۳.۲.۴
۷۵	دسته‌بندی	۳.۴
۷۵	مدل پیشنهادی پایه	۱.۳.۴
۸۰	تابع ضرر	۲.۳.۴
۸۲	آموزش مدل و استخراج ویژگی	۳.۳.۴
۸۳	فناوری‌های استفاده شده	۴.۴
۸۴	ارزیابی روش پیشنهادی	۵
۸۵	مقدمه	۱.۵
۸۵	معیار ارزیابی	۲.۵
۸۶	مجموعه داده	۳.۵
۸۶	مجموعه داده‌های آموزش	۱.۳.۵
۸۷	مجموعه داده‌های آزمون	۲.۳.۵
۸۸	پیکربندی الگوریتم	۴.۵
۸۹	نتایج آزمون	۵.۵
۹۱	نتیجه‌گیری و پیشنهادات	۶
۹۲	مقدمه	۱.۶
۹۲	بحث و نتیجه‌گیری	۲.۶

۳.۶ پیشنهادات

۹۳

منابع و مأخذ

۹۴

فهرست جداول

۱.۲	مقایسه شبکه عصبی موبایل نت با گوگل نت و VGG	۳۸
۱.۴	مقایسه و ارزیابی برخی از معماری شبکه های رایج در زمینه بینایی ماشین	۷۶
۱.۵	مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره	۸۸
۲.۵	دقیق معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره	۸۹
۳.۵	سرعت معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره	۸۹

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 نایینا	۹
۵.۱	مقایسه تفاوت‌های ظاهری ناشی از روش‌نایی و تفاوت بین چهره افراد [؟]	۹
۶.۱	زاویه شدید چهره نسبت به دوربین باعث کاهش دقت سامانه می‌شود [؟]	۹
۷.۱	تاری خارج از مرکز به علت عمق کم میدان دوربین [？]	۱۰
۸.۱	انسداد شدید در اثر نورپردازی [？]	۱۰
۹.۱	نمونه‌ای از یک سامانه تشخیص چهره بی‌درنگ در محیط کنترل نشده [？]	۱۱
۱۰	رویکرد مبتنی بر تطبیق کلیشه [？]	۱۴
۱۱	استفاده از لبه‌های معروف برای استخراج ویژگی‌های مبتنی بر لبه از چهره [۱]	۱۵
۱۲	اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می‌باشد [۱]	۱۶
۱۳	نمونه‌هایی از مستطیل‌های ویژگی‌های Haar [۲]	۱۸
۱۴	مستطیل‌های ویژگی‌های Haar متناسب با بخش‌های چهره می‌باشند [۲]	۱۸
۱۵	هر مستطیل در اندازه‌های مختلف بر روی بخش‌های مختلف تصویر قرار می‌گیرد [۲]	۱۹
۱۶	یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [۲]	۱۹
۱۷	نحوه مقدار دهی به پیکسل‌های تصویر یکپارچه [۲]	۲۰
۱۸	بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [۲]	۲۰

- ۲۱ ۱۰.۲ بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [۲].
- ۲۱ ۱۱.۲ سطح روشنایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می شود [۳].
- ۲۲ ۱۲.۲ برای یافتن چهره ها، بخش هایی از تصویر که به الگوی HOG شبیه تر است را پیدا می کنیم [۳].
- ۲۳ ۱۳.۲ نتیجه اجرای روش مبتنی بر NPD [۴].
- ۲۴ ۱۴.۲ نمای کلی روش یافتن چهره تک مرحله ای و متراکم RetinaFace بر اساس اهرام ویژگی طراحی شده است.
- ۲۵ ۱۵.۲ این رویکرد از تابع ضرر چند کاره استفاده می نماید. [۵].
- ۲۶ ۱۶.۲ ویژگی های هندسی (رنگ سفید) مورد استفاده در آزمایش های تشخیص چهره [۶].
- ۲۶ ۱۷.۲ (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۷].
- ۲۷ ۱۸.۲ مدل سازی سه بعدی چهره با اشعه فرو سرخ [۸].
- ۲۹ ۱۹.۲ نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [۹].
- ۳۰ ۲۰.۲ (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۱۰].
- ۳۱ ۲۱.۲ شبکه عصبی عمیق برای شناسایی چهره.
- ۳۲ ۲۲.۲ سه گانه تطبیق - عدم تطبیق [۱۱].
- ۳۴ ۲۳.۲ (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [۱۲].
- ۳۴ ۲۴.۲ معماری کلی شبکه GoogLeNet [۱۲].
- ۳۵ ۲۵.۲ کادر سبز رنگ یکی از بخش های موازی شبکه رانشان می دهد [۱۳].
- ۳۵ ۲۶.۲ بخش آغازگر شبکه GoogLeNet [۱۳].
- ۳۵ ۲۷.۲ معماری شبکه VGGFace [۱۴].
- ۳۶ ۲۸.۲ نتیجه موقعیت ۱۹۴ شاخص روی چهره [۱۵].
- ۳۶ ۲۹.۲ علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطای همراه می باشد [۱۵].
- ۳۷ ۳۰.۲ (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری TCNN [؟].
- ۳۸ ۳۱.۲ کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن depth-wise و point-wise هست (سمت راست). [۱۶].
- ۳۹ ۳۱.۲ لایه های MobileNet به همراه مازول های Squeeze-and-Excite [۱۷].

٤٠	۳۲.۲ معماری ماژول [۱۷] Attention Shuffle
٤١	۳۳.۲ مقایسه چگالی دقیقت در معماری‌های مختلف شبکه عصبی پیچشی [۱۸]
٤٥	۱.۳ رویکردهای مختلف هم ترازی چهره [۱۹]
٤٦	۲.۳ تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی [۲۰]
٤٧	۳.۳ معماری OpenFace [۲۰]
٤٧	۴.۳ جریان یادگیری در معماری OpenFace [۲۰]
۴۸	۵.۳ رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲۱]
۴۸	۶.۳ مقدار دهی اولیه و بهینه سازی AAM [۲۱]
۴۹	۷.۳ نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۲۱]
۵۰	۸.۳ معماری شبکه پیشنهادی VS2VI [۲۲]
۵۰	۹.۳ (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [۲۲]
۵۱	۱۰.۳ مقایسه روش ارائه شده با سایر روش ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش های دیگر (d) خروجی روش ارائه شده [۲۲]
۵۲	۱۱.۳ ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۲۴]
۵۴	۱۲.۳ اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۲۴]
۵۶	۱۳.۳ تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۲۵]
۵۷	۱۴.۳ رویکرد کلی الگوریتم GD-HASLR [۲۵]
۵۸	۱۵.۳ ردیابی، یافتن چهره ها و برچسب زنی [۲۶]
۵۸	۱۶.۳ تصاویر با حاشیه قرمز رنگ، به اشتباہ برچسب زنی شده اند [۲۶]
۵۹	۱۷.۳ استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [۲۶]
۶۰	۱۸.۳ تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۲۷]

۱۹.۳	تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین در ۴ مجموعه داده مختلف. موارد با حاشیه قرمز خروجی اشتباه هستند [۲۸].	۶۱
۲۰.۳	تصاویر هر ردیف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند [۲۹].	۶۲
۲۱.۳	ساختار شبکه AD GAN - شبکه GN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۳۰].	۶۳
۲۲.۳	معماری MOCHA: دستگاه‌های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می‌کنند [۳۱].	۶۴
۲۳.۳	نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۳۲].	۶۵
۲۴.۳	چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۳۲].	۶۶
۲۵.۳	تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مربوط بین دسته‌ها ایجاد می‌نماید [۳۳].	۶۷
۲۶.۳	رویکردهای مختلف هم ترازی چهره [۳۴].	۶۸
۲۷.۳	رویکردهای مختلف هم ترازی چهره [۳۴].	۶۸
۱.۴	نمای کلی از روش پیشنهادی [۴].	۷۲
۲.۴	نتیجه اعمال یکسان‌سازی بافت‌نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می‌باشند [۳۵].	۷۳
۳.۴	نمونه از خروجی الگوریتم یافتن چهره retina [۵].	۷۴
۴.۴	به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد [۶].	۷۵
۵.۴	مدل پایه مبتنی بر لایه‌های کانولوشن و لایه توجه.	۷۶
۶.۴	یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن‌های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می‌کند.. ۷۸	۷۸
۷.۴	معماری ماژول Attention Shuffle [۱۷].	۷۸
۸.۴	تابع ضرر ArcFace در مقایسه با توابع ضرر مشهور دیگر [۳۴].	۸۲

فهرست نمادها

فصل ١

مقدمة

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 مستله خاص در زمینه تشخیص چهره به صورت بی‌درنگ و در محیط کنترل نشده پرداخته و نیازها و ابزارهای مورد نیاز بررسی شده است.

۲۰۱ 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 بردار ویژگی‌های خاص فیزیولوژیک بازشناسی می‌کند.

^۱Biometric

۱.۲.۱ ارزیابی ویژگی‌های بیومتریک انسان

معمولاً ویژگی‌های بیومتریک انسان با ۸ عامل مورد ارزیابی قرار می‌گیرند که عبارت اند از:

- عمومیت: هر شخصی باید دارای آن ویژگی بیومتریک باشد.
- یکتاپی: آن ویژگی بیومتریک باید برای هر شخصی منحصر به فرد باشد.
- دوام: معیاری برای سنجش آنکه یک ویژگی بیومتریک چه مدت بدون تغییر باقی می‌ماند.
- ارزیابی: ویژگی بیومتریک مورد نظر باید سادگی کافی را در استفاده برای ارزیابی نمونه‌های متفاوت داشته باشد.
- کارایی: استفاده از ویژگی بیومتریک مورد نظر باید دقیق، سرعت و پایداری مطلوب داشته باشد.
- مقبولیت: فناوری استفاده از ویژگی بیومتریک مورد نظر باید در میان جامعه پذیرش شده باشد.
- تصدیق هویت: ویژگی فرد به سامانه ارسال شود و سامانه پاسخی مثبت یا منفی برای تصدیق هویت فرد ارائه نماید.
- 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 عنیبه چشم، ارتباط راحت تری 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

فاویری 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 منصور 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 خطأ در سامانه بیشتر می‌شود و به یک الگوریتم دقیق تر نیاز داریم.

¹Access Control

بین سرعت تشخیص و دقت تشخیص، بدء بستان^۱ وجود دارد. یک الگوریتم کارا باید هر دو ویژگی بالا را در نظر بگیرد.

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. پنج 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.

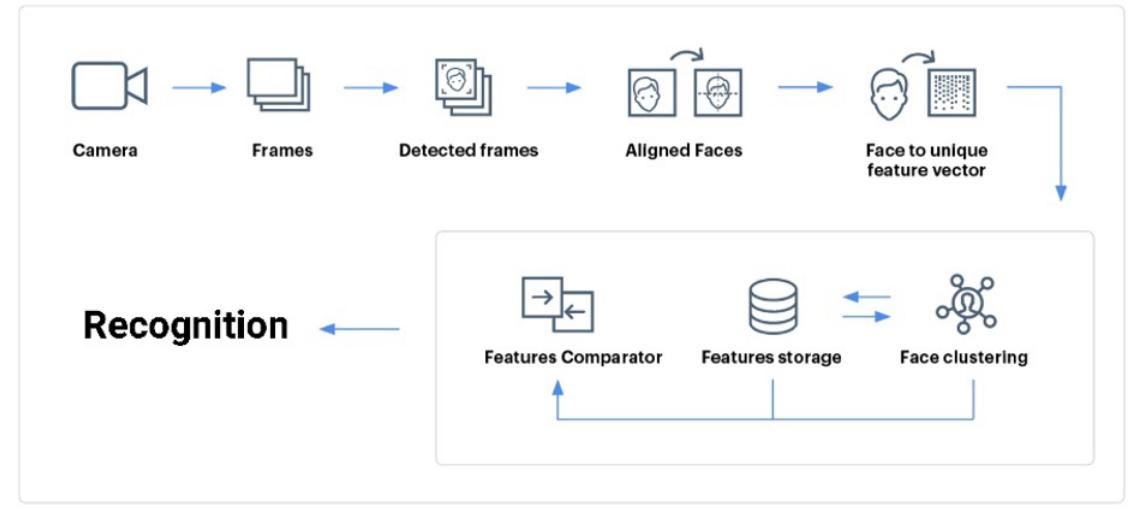
یافتن چهره: تصاویر ورودی به این بخش ابتدا مورد پیش‌پردازش قرار می‌گیرند. سپس ارزیابی محتوایی شده و داده‌های نامربوط از قبیل پس زمینه، موها و گردن و شانه و... حذف می‌شوند و تنها ناحیه چهره باقی می‌ماند.

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.

^۱Tradeoff



شکل ۱.۱: نمای کلی یک سامانه تشخیص چهره [۶]

۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر

سامانه شناسایی تصویر یکی از مهم ترین بخش‌های یک سامانه بینایی ماشین به حساب می‌آید. این سامانه با استفاده از یادگیری عمیق^۱، انواع اطلاعات در تصاویر مانند افراد، متون، اشیا و سایر اطلاعات موجود را شناسایی می‌کند. الگوریتم‌های مورد استفاده در فرایند تشخیص تصاویر بر مبنای استفاده از ویژگی‌های بصری تصویر مانند لبه و رنگ و یا ویژگی‌های استخراج شده از شبکه عصبی عمیق^۲ به همراه اعمال الگوریتم‌های مختلف خوشه‌بندی^۳، واپاش^۴ یا طبقه‌بندی^۵ داده‌ها است که سعی در تحلیل، شناسایی و تشخیص تصاویر مختلف دارند. عنصر اصلی این گونه روش‌ها، داده‌های مورد استفاده در فرایند یادگیری می‌باشد. نوع و فراوانی داده‌ها در افزایش دقیق سامانه‌های تشخیص تصاویر مبتنی بر شبکه عصبی بسیار مهم می‌باشد.

۱.۴.۱ خوشه‌بندی

روش خوشه‌بندی بر اساس محاسبه میزان و معیار شباهت در داده‌ها، آن‌ها را در خوشه‌های مختلف قرار می‌دهد. به طور کلی دو نوع روشن خوشه‌بندی وجود دارد:

خوشه‌بندی سخت^۶: برچسب گذاری به صورت صفر و یک بر روی داده‌ها می‌باشد و مشخص می‌کند داده مورد نظر مربوط به خوشه

¹Deep Learning

²Deep Neural Network

³Clustering

⁴Regression

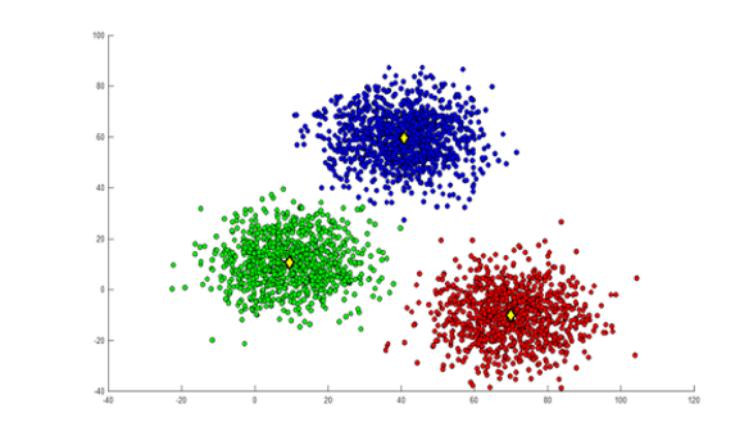
⁵Classification

⁶Hard Clustering

می‌باشد یا خیر.

خوشبندی نرم^۱ : به خوشبندی فازی معروف است و میزان تعلق یک داده به خوشبدها را مشخص می‌کند و هر داده می‌تواند با توجه به وزن‌های اختصاص داده شده به آن، به چندین خوشبندی داشته باشد.

داده‌های استفاده شده در فرایند خوشبندی، بدون برچسب بوده و یادگیری بدون ناظر^۲ در فرایند خوشبندی داده‌ها استفاده می‌شود. شکل ۲.۱ نمونه‌ای از خوشبندی سخت و یافتن شاخص برای هر خوشبندی را نشان می‌دهد.



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

۲.۴.۱ طبقه‌بندی

روش‌های طبقه‌بندی داده‌ها از جمله روش‌های یادگیری با ناظر^۳ 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.

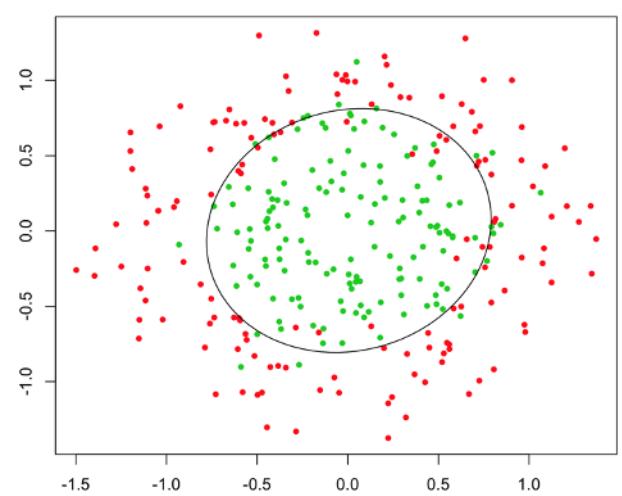
در این نوع از الگوریتم‌ها که از لحاظ تعداد، بار اصلی یادگیری ماشین را بر دوش می‌کشنند، دو نوع داده وجود دارند. نوع اول داده‌های مستقل نامیده می‌شوند که باید بر اساس آن‌ها، یک متغیر دیگر پیش‌بینی شود. نوع دوم داده‌های وابسته یا برچسب هستند که قرار است مقادیر آن‌ها به کمک این الگوریتم‌ها پیش‌بینی شود. برای این منظور باید تابعی ایجاد شود که ورودی‌ها (داده‌های مستقل) را گرفته و خروجی موردنظر (داده‌های وابسته یا هدف) را تولید کنند. فرایند یافتن این تابع، کشف رابطه‌ای بین متغیرهای مستقل و وابسته است که آن را فرآیند آموزش می‌نامند که بر روی داده‌های موجود اعمال می‌شود و تا رسیدن به دقت لازم ادامه می‌یابد.

¹Soft Clustering

²Unsupervised Learning

³Supervised Learning

الگوریتم‌های مختلفی برای طبقه‌بندی داده‌ها وجود دارد که در این میان می‌توان به شبکه عصبی، SVM^۱ و KNN^۲ اشاره کرد.



شکل ۳.۱: یک طبقه بند غیر خطی ساده

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 نابینا می‌باشد. هنگامی که شخص نابینا عینک را بر روی

چشم‌مانش 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 خواهند گرفت.

۱.۵.۱ چالش‌های سامانه تشخیص چهره

سامانه‌های تشخیص چهره به سطح مشخصی از بلوغ رسیده اند، 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

¹Support Vector Machine

²K Nearest Neighbor



شکل ۴.۱: نمونه ای از عینک دوربین دار برای افراد نابینا

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 برشمرده 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

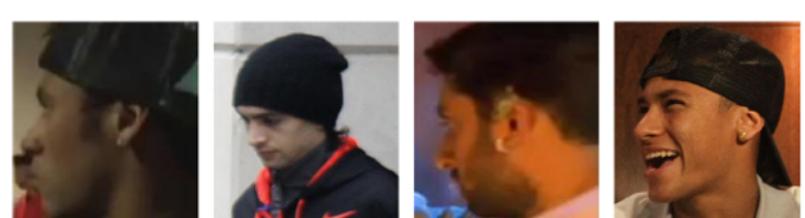
شکل ۵.۱ نشان داده شده است که گاهی تفاوت‌های ظاهری ناشی از روشنایی، بیشتر از تفاوت بین چهره افراد مختلف می‌باشد.



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

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 [؟]

¹Lighting

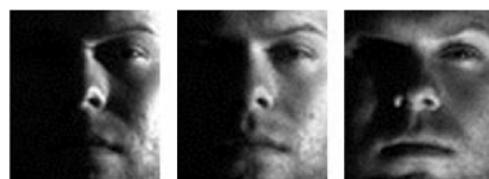
²Pose

تاری خارج از تمرکز^۱: اگر عمق میدان دوربین کم باشد و چهره‌ها با فاصله‌های مختلف از دوربین باشند، مشکل تاری خارج از تمرکز 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 درس ساز شود. نمونه‌ای از این اثر در شکل ۷.۱ قابل مشاهده می‌باشد.



شکل ۷.۱: تاری خارج از تمرکز به علت عمق کم میدان دوربین [۲]

انسداد^۲: اگر در حال تصویر برداری از یک جمع باشیم، احتمال اینکه چهره فردی توسط فرد دیگری مقداری پوشانده شود، بسیار بالاست. همچنین اگر بخشی از چهره فرد توسط موها پوشیده شده باشد، از عینک آفتابی استفاده کند، یا نورپردازی غیر یکنواخت باشد، انسداد چهره رخ خواهد داد. نمونه‌ای از این اثر در شکل ۸.۱ قابل مشاهده می‌باشد.



شکل ۸.۱: انسداد شدید در اثر نورپردازی [۳]

سن^۳: بیشتر روش‌های مرسوم تشخیص چهره تغییرات سن را نادیده می‌کیرند. 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

¹Out Of Focus Blur

²Occlusion

³Aging

می‌باشند. تعداد افراد در محیط کنترل نشده زیاد می‌باشد و به 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 پردازنده پایین‌تر و منبع انرژی محدودتر می‌باشد که باعث می‌شود الگوریتم‌هایی با پیچیدگی محاسباتی بالا بر روی

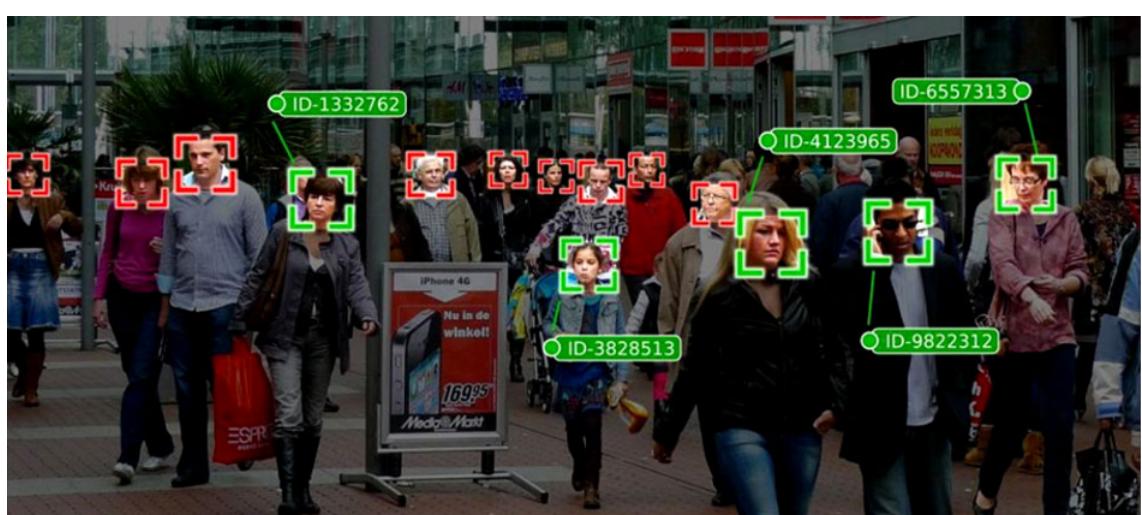
این دستگاهها قابل اجرا نباشد. بنابراین الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد.

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 چند 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 چهره بی‌درنگ در محیط کنترل نشده [۹]

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^۳ و همکاران در [۳۶] یک دسته بندی 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

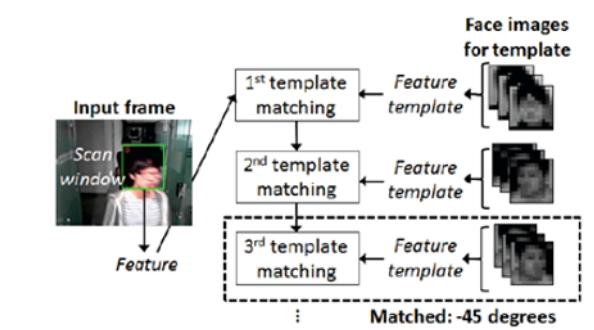
²Face Recognition

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 ۱۵.۳ 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.



شکل ۱۰.۲: رویکرد مبتنی بر تطبیق کلیشه [؟].

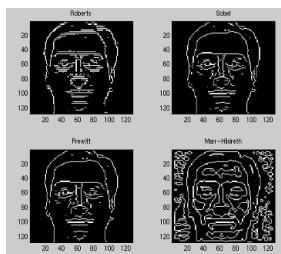
¹Template

۲.۲.۲ رویکردهای مبتنی بر ویژگی

این رویکردها با استخراج ویژگی^۱ 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 [۱] تعدادی از این رویکردها مورد بررسی قرار گرفته است:

۱.۲.۲.۲ رویکرد مبتنی بر لبه

ابتدا به کمک یک الگوریتم لبه یاب^۲، لبه‌های تصویر بدست می‌آید، سپس نازک سازی می‌شوند و شاخه‌های اضافه حذف می‌گردد (شکل ۳.۲). گوشه لبه‌ها تشخیص داده می‌شوند و هر مولفه متصل^۳ به شاخه مرکزی آن کاهش 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.



شکل ۲.۲: استفاده از لبه یاب های معروف برای استخراج ویژگی های مبتنی بر لبه از چهره [۱].

۲.۲.۲.۲ رویکرد مبتنی بر اطلاعات سطح خاکستری

سطوح خاکستری^۴ تصویر چهره شامل اطلاعات مفیدی می‌باشد. برای مثال ابروها، مردمک چشم و لب‌ها معمولاً تاریک تر از سایر نواحی صورت هستند. این ویژگی‌ها می‌توانند به یافتن یک چهره در تصویر کمک نماید. در این رویکرد ابتدا بر روی تصویر ورودی،

¹Feature

²Edge Detection

³Connected Component

⁴Gray level

عملیات بسط تباین^۱ و عملیات مورفولوژی^۲ مبتنی بر سطح خاکستری انجام می‌شود تا تصویر بهبود پیدا کند و یافتن نواحی تیره، 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 ۳.۲).



شکل ۳.۲: اطلاعات سطح خاکستری حتی دروضوح پایین نیز قابل دسترسی می‌باشند [۱].

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^۳ می‌شوند تا اثر نور پردازی از بین برود. گستردگی ترین 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:

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

¹Contrast Stretching

²Morphological Operations

³Normalization

پس می‌توان فقط با داشتن مقدار r و g مقدار b را بدست آورد. با توجه به بافت‌نگار^۱ 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 YCrCb، YIQ، HSV 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.

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^۲ 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^۳ 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 پیشرفته تر و دقیق‌تر نماید، اما انقلاب الگوریتم‌های یافتن چهره در سال ۲۰۰۱ بود. زمانی که Viola و Jones در [۲] 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^۴، برش^۵، تار شدن^۶ و تیزکردن لبه‌های تصویر^۷ انجام می‌شود.

ویژگی‌های Haar: تمام چهره‌های انسانی ویژگی‌های مشترکی دارند. برای مثال ناحیه چشم تاریک‌تر از پیکسل‌های همسایه آن است و ناحیه بینی از چشم روشن‌تر است. ویژگی‌های Haar مستطیل‌هایی هستند که نشان دهنده بخش‌های مختلف صورت می‌باشند.

¹Histogram

²Statistical Parameters

³Bayes Rule

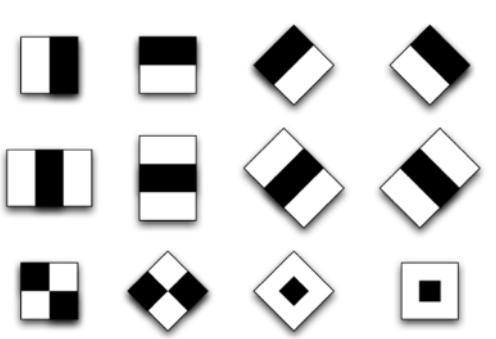
⁴Resizing

⁵Cropping

⁶Blurring

⁷Sharpening

شکل ۴.۲ نمونه‌هایی از مستطیل‌های ویژگی‌های Haar را نشان می‌دهد.



شکل ۴.۲: نمونه‌هایی از مستطیل‌های ویژگی‌های Haar [۲].

ویژگی‌های 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 می‌شود. مستطیل‌های ویژگی‌های Haar، مناسب با بخش‌های چهره می‌باشند که مثالی از آن در

شکل ۵.۲ نشان داده شده است.

همانطور که در شکل ۶.۲ مشاهده می‌شود، هر مستطیل در بخش‌های مختلف چهره در روندی تکراری با اندازه‌های مختلف قرار

می‌گیرد و نتیجه نهایی از کم کردن مجموع سطح روشنایی پیکسل‌های زیر

بخش‌های سفید به دست می‌آید که یک عدد می‌باشد. از یک پنجره با اندازه 24×24 برای قرار دادن مستطیل‌ها بر روی تصویر استفاده

می‌شود که تعداد زیاد و اندازه‌های مختلف آن‌ها باعث می‌شود برای محاسبه نتیجه نهایی نیاز به انجام بیش از ۱۶۰۰۰۰ محاسبه باشد

که زمان زیادی برای هر تصویر خواهد گرفت.

Ada Boost: همانطور که در شکل ۷.۲ مشاهده می‌شود، تمام ویژگی‌های Haar برای تصویر ورودی مناسب نخواهد بود. بعضی

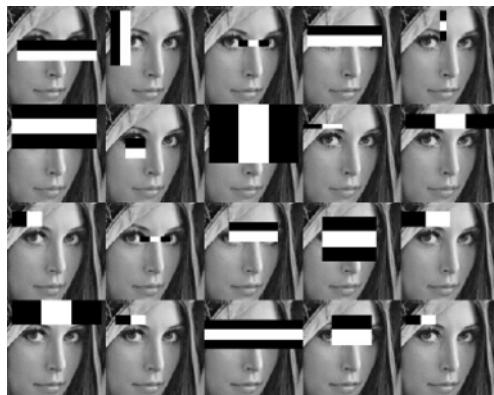
از این ویژگی‌ها باید نادیده گرفته شوند و فقط ویژگی‌های مرتبط انتخاب شوند تا در زمان صرفه جویی شود. این کار به صورت خودکار

به کمک عنصر Ada Boost انجام می‌شود.

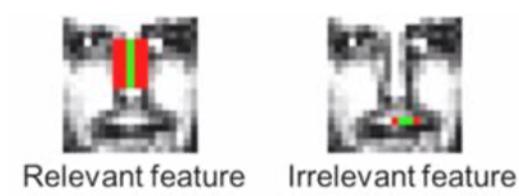
یک الگوریتم مبتنی بر یادگیری ماشین می‌باشد که ویژگی‌های کاربردی را از میان تعداد زیادی ویژگی پیدا می‌کند.



شکل ۵.۲: مستطیل‌های ویژگی‌های Haar مناسب با بخش‌های چهره می‌باشند [۲].



شکل ۶.۲: هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [۲].



شکل ۷.۲: یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [۲].

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 می شود، در صورتی که حداقل بتواند بیش از نیمی از موارد را تشخیص دهد. این ویژگی ها با عنوان طبقه بند های ضعیف^۱ معرفی می شوند. Ada Boost در طبقه بند قوی^۲، تعداد زیادی طبقه بند ضعیف را با هم ترکیب می کند. رابطه کلی آن به صورت زیر می باشد.

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

که در آن F طبقه بند قوی می باشد که از تعدادی f_i که طبقه بند ضعیف می باشد، تشکیل شده است. هر کدام از طبقه بند های ضعیف، یک خروجی صفر یا یک تولید می کنند. α_i وزن مربوط به طبقه بند می باشد. با استفاده از این الگوریتم وزن دادن به ویژگی ها، بیش از ۱۶۰۰۰۰ ویژگی قبلی به کمتر از ۲۵۰۰ ویژگی کاهش پیدا می کند.

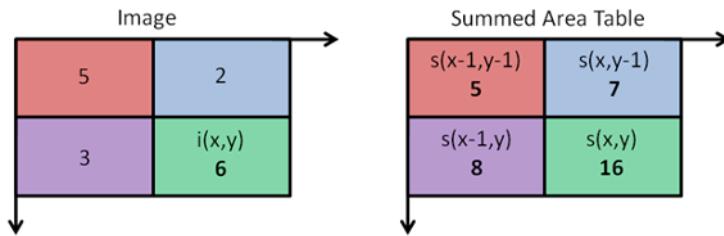
تصویر یکپارچه^۳: تصویر یکپارچه، یا جدول محدوده مجتمع^۴، به منظور ارزیابی سریع تر ویژگی هایی که در بخش اول معرفی شد، استفاده می شود. با توجه به شکل ۸.۲ در یک تصویر یکپارچه مقدار پیکسل در مکان x و y برابر با جمع مقادیر پیکسل های بالا و چپ پیکسل x و y می باشد.

¹Weak Classifier

²Strong Classifier

³Integral Image

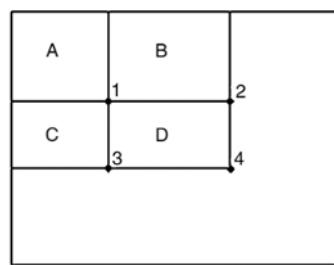
⁴Summed area table



شکل ۸.۲: نحوه مقدار دهی به پیکسل های تصویر یکپارچه [۲].

به عنوان مثال در شکل ۹.۲ مقدار پیکسل ها به صورت زیر محاسبه می شود:

- مقدار پیکسل ۱ در تصویر یکپارچه برابر است با مجموع پیکسل ها در مستطیل A .
- مقدار پیکسل ۲ در تصویر یکپارچه برابر است با مجموع پیکسل ها در مستطیل A و B .
- مقدار پیکسل ۳ در تصویر یکپارچه برابر است با مجموع پیکسل ها در مستطیل A و C .
- مقدار پیکسل ۴ در تصویر یکپارچه برابر است با مجموع پیکسل ها در مستطیل A و B و C و D .
- مجموع پیکسل ها در مستطیل D می تواند به صورت $(۳+۲)-(۱+۴)$ محاسبه شود.



شکل ۹.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [۲].

مراحل آبشاری^۱: اگر تصویر به مربع های 24×24 پیکسلی تقسیم شده و با پردازش هر بخش که 2500 ویژگی دارد، تشخیص داده شود

که چهره ای در تصویر وجود دارد یا خیر، حجم محاسبات بسیار زیاد خواهد بود. مراحل آبشاری این فرایند را سریع تر انجام می دهد.

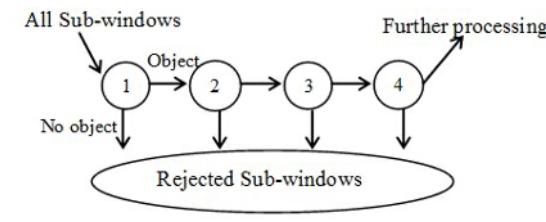
2500 ویژگی هر مربع 24×24 به دسته بندی های مختلف تقسیم می شود. برای مثال 100 ویژگی در دسته اول، 20 ویژگی در دسته

دوم، 100 ویژگی در دسته سوم و ... می توان بعد از پردازش هر دسته، در ارتباط با وجود یا عدم وجود چهره در آن دسته تصمیم گرفت

تا بخش هایی که چهره در آن وجود ندارد زودتر حذف شوند. شکل ۱۰.۲ یک نمای کلی از روند تشخیص آبشاری را نشان می دهد.

¹Cascading

مجموعه‌ای از طبقه‌بندها 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.

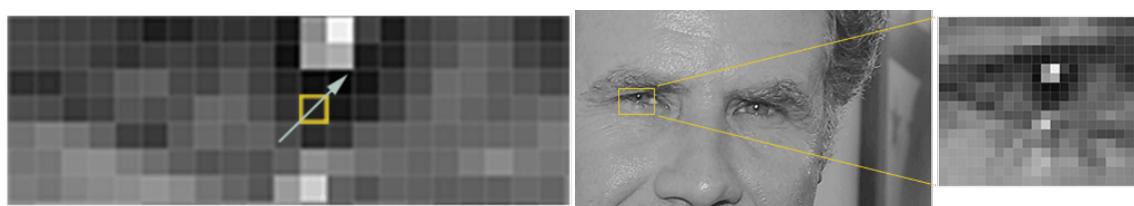


شکل ۱۰.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [۲].

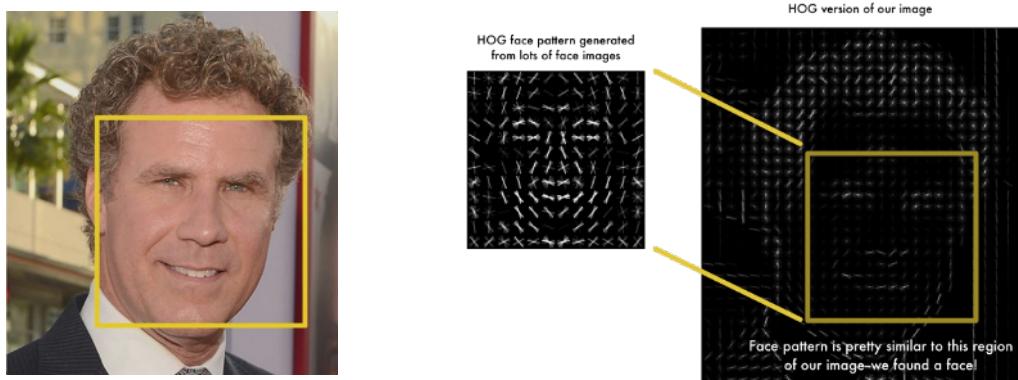
در سال ۲۰۰۵ دلال و همکاران در [۳] روشی به نام بافت نگار شبیه‌های جهت دار^۱ ارائه کردند که به اختصار HOG نامیده می‌شود. در این روش ابتدا تصویر خاکستری می‌شود، زیرا نیازی به رنگ نیست. پیرامون هر پیکسل بررسی می‌شود تا مشخص شود نسبت به پیکسل‌های پیرامونش چقدر تاریک می‌باشد. مطابق شکل ۱۱.۲ جهتی انتخاب می‌شود که به سمت پیکسل‌های تاریکتر باشد. این روند برای همه پیکسل‌های تصویر انجام می‌شود و به ازای هر پیکسل یک جهت ذخیره خواهد شد که روندی از روشنایی به تاریکی را در تصویر نمایش می‌دهند.

دلیل استفاده از جهت‌ها این است که اگر پیکسل‌ها 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 می‌کند. ذخیره جهت‌ها برای تمام پیکسل‌ها باعث افزایش جزئیات می‌شود. لذا روند اصلی روشنایی و تاریکی در سطحی بالاتر در نظر گرفته می‌شود، به طوری که بتوان الگوی اصلی تصویر را دید. تصویر به بخش‌های ۱۶×۱۶ پیکسل تقسیم می‌شود و در هر بخش تعداد جهت‌های به سمت بالا، پایین، چپ و راست شمارش می‌شود. سپس بخش‌های

¹Histograms Of Oriented Gradients



شکل ۱۱.۲: سطح روشنایی پیکسل‌های اطراف های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می‌شود [۳].



شکل ۱۲.۲: برای یافتن چهره‌ها، بخش‌هایی از تصویر که به الگوی HOG شبیه‌تر است را پیدا می‌کنیم [۳].

دروں تصویر با جهت‌هایی که بزرگتر بودند، جایگزین می‌شود. نتیجه نهایی، تبدیل تصویر به یک نمایش ساده شده از ساختار چهره است که در شکل ۱۲.۲ مشاهده می‌شود. برای یافتن چهره‌ها در این الگوریتم، بخش‌هایی از تصویر که به الگوی HOG شبیه‌تر است، مشخص می‌شود. با استفاده از این روش، می‌توان چهره‌ها را در تصویر به سادگی پیدا کرد.

در سال ۲۰۱۵ Shengcai Liao و همکاران در [۴] یک روش دقیق و سریع برای یافتن چهره ارائه دادند که در آن از اختلاف پیکسل هنجارسازی شده یا NDP 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 آمورش نیز لازم نیست و در برابر تغییرات نور، حالت، انسداد، تصاویر باوضوح پایین و... مقاوم 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 $x = y = 0$ 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 $p = s^2$ تعداد پیکسل‌ها باشد، آنگاه تعداد ویژگی‌های استخراج شده برابر $d = p(p - 1)/2$ می‌باشد. سپس علامت^۱ 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^۲ 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 ۱۳.۲ آمده است.

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

¹Sign

²Noise



شکل ۱۳.۲: نتیجه اجرای روش مبتنی بر NPD [۴].

نشده را بهبود داده اند، 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 شده اند. پس نیاز به روش‌هایی هست که بتوانند در شرایط خاصمانه تر مانند تشخیص چهره‌های متعدد در زمینه‌های شلوغ به خوبی عمل کنند.

۴.۲.۲ رویکردهای مبتنی بر تصویر

با توجه به آنچه در [۳۷] 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 ویژگی‌هایی 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. رویکردهای مبتنی بر تصویر نیاز به مجموعه‌ای از تصاویر آموزشی برای پیدا کردن مدل‌های چهره دارند و بر اساس استخراج ویژگی و یادگیری ماشین عمل می‌نمایند. مجموعه‌ای از تصویرهای مختلف چهره طبقه‌بندی می‌شوند و برای تشخیص چهره جدید از این طبقه‌بندی استفاده می‌شود. نمونه‌هایی از چهره و نمونه‌هایی از غیر چهره به طبقه‌بند داده می‌شود تا از روی این تصویرها عمل یادگیری انجام شود. به طور تجربی دقت نتایج رویکردهای مبتنی بر تصویر بهتر از سایر رویکردها می‌باشد. این رویکردها به چند دسته تقسیم می‌شوند که در ادامه شرح داده شده است.

۱.۴.۲.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

ماشین بردار پشتیبان طبقه‌بندی خطی می‌باشد که حاشیه بین ابرصفحه تصمیم و داده‌های آموزش را به حداقل می‌رساند. برای اولین بار در سال ۱۹۹۷ Osuna و همکاران در [۳۸] از این طبقه‌بند برای یافتن چهره استفاده کردند.

۲.۴.۲.۲ رویکرد مبتنی بر شبکه عصبی

یک راه حل غیر خطی برای یافتن چهره، استفاده از شبکه‌های عصبی^۱ است. اولین رویکردهای عصبی در یافتن چهره بر اساس MLP^۲ بود که در مجموعه داده‌های ساده، امیدوار کننده بود. شبکه‌های عصبی با معماری پیمانه ای^۳ امروزی بسیار پیچیده‌تر از MLP ساده هستند. نوع خاصی از شبکه‌های عصبی عمیق برای پردازش تصاویر استفاده می‌شوند که شبکه عصبی پیچشی^۴ نام دارند. ساختار عمیق این شبکه‌ها باعث شد در مجموعه داده‌های بزرگ و دشوار نتایج خوبی بدست آید و ویژگی‌های عمیق به دست آمده به طور گسترده‌ای برای یافتن چهره استفاده می‌شود. یک شبکه عصبی پیچشی از تعدادی تابع و لایه تشکیل شده است.

لایه‌های شبکه عصبی عبارتند از:

- لایه‌های پیچشی^۵ برای لغزاندن یک پنجره بر روی ورودی
- لایه‌های تمام متصل^۶ برای محاسبه مجموع وزن دار تمام واحد های ورودی
- لایه‌های رای گیری^۷ به منظور کاوش حجم داده‌ها با محاسبه مقدار بیشینه، میانگین یا اندازه اقلیدسی هر بخش

در سال ۲۰۱۹ و همکاران در [۵] یک روش مبتنی بر شبکه عصبی پیچشی برای یافتن چهره در تصویر پیشنهاد کردند. در این روش برای آموزش شبکه پیچشی از یک تابع ضرر مبتنی بر یادگیری چندکاره^۸ استفاده شده است که به صورت زیر 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}

¹Neural Network

²Multi Layer Perceptron

³Modular Architecture

⁴Convolutional Neural Network

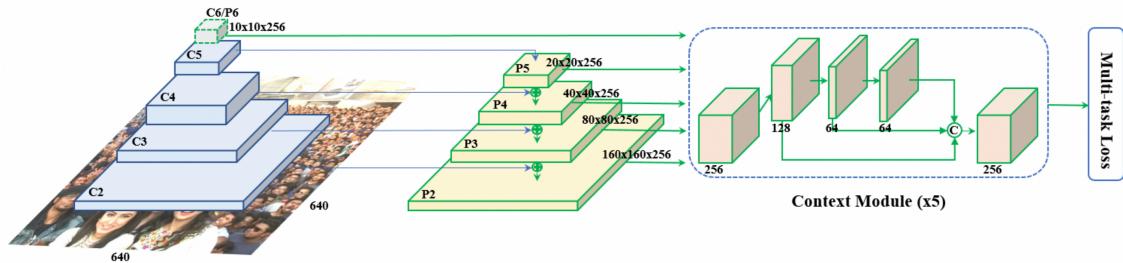
⁵Convolution Layer

⁶Fully Connected Layer

⁷Pooling Layer

⁸Multi Task Learning

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 از روی چهره می‌باشد. استفاده از تابع ضرر مبتنی بر یادگیری چند کاره، 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 ۱۴.۲ مشاهده می‌کنید.



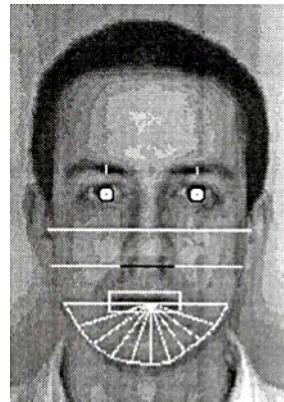
شکل ۱۴.۲: نمای کلی روش یافتن چهره تک مرحله ای و متراکم. RetinaFace بر اساس اهرام ویژگی طراحی شده است. این رویکرد از تابع ضرر چند کاره استفاده می‌نماید. [۵].

۳.۲ شناسایی چهره

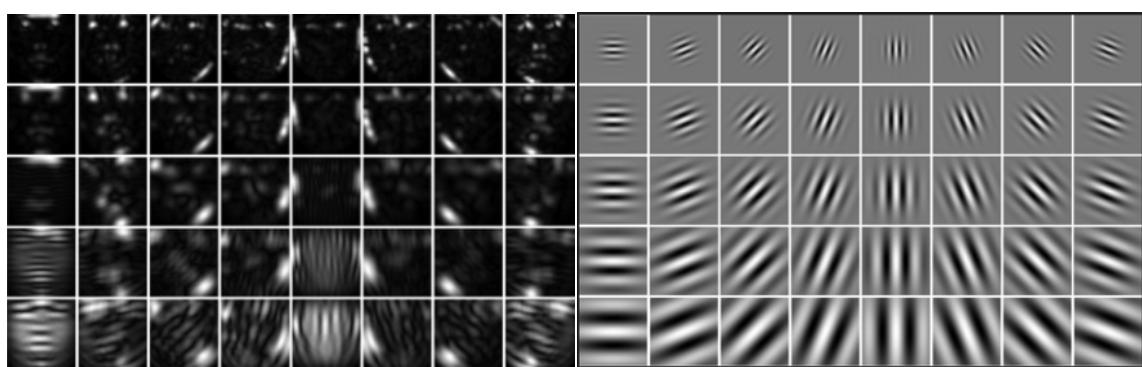
شناسایی چهره در دو مرحله انجام می‌شود. مرحله اول استخراج ویژگی و مرحله دوم، طبقه‌بندی است. الگوریتم‌های شناسایی چهره را می‌توان به دو دسته اصلی تقسیم کرد. الگوریتم‌های هندسی که بر مبنای استخراج ویژگی‌های متمایز چهره‌ها کار می‌کنند، و الگوریتم‌های تصویری که تصویر را تبدیل به یک الگو می‌نمایند و الگوها 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 در [۶] یکی از اولین الگوریتم‌ها در این زمینه را ارائه دادند که رویکرد موفقی مبتنی بر روش‌های تطبیق الگو داشت (شکل ۱۵.۲). این رویکرد در شرایط کنترل شده به دقت ۹۰٪ رسید.



شکل ۱۵.۲: ویزگی های هندسی (رنگ سفید) مورد استفاده در آزمایش های تشخیص چهره [۶].



شکل ۱۶.۲: (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۷].

۲.۳.۲ رویکرد مبتنی بر فیلتر گابور

همانطور که در [۷] آمده است، در این رویکرد ابتدا تصویر را بخش بندی^۱ کرده، سپس بر روی بخش های مختلف آن، فیلتر گابور^۲ اعمال می شود و نتیجه بدست آمده با یک طرح از پیش آمده شده، با یک آستانه گذاری مطابقت داده می شود. شکل ۱۶.۲ فیلترهای چندگانه گابور و تاثیر این فیلترها بر روی تصویر چهره انسان را نشان می دهد. دلیل استفاده از فیلتر گابور این است که عملکرد این فیلتر به سامانه بصری انسان بسیار شباهت دارد.

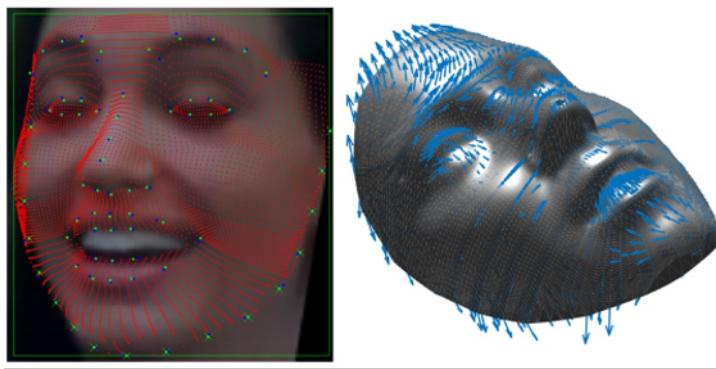
۳.۳.۲ رویکردهای سه بعدی

همانطور که در [۸] آمده است، داده های سه بعدی دقت تشخیص چهره را به شدت بهبود می بخشد، اختلاف داده های ورودی با داده های ذخیره شده زیادتر است و سامانه با دقت بیشتری عمل می کند. روش تشخیص سه بعدی چهره از یک منتشر کننده نور

¹Grid

²Gabor

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 ۱۷.۲ یک مدل سازی سه بعدی چهره با اشعه فروسرخ را نشان می‌دهد.



شکل ۱۷.۲: مدل سازی سه بعدی چهره با اشعه فروسرخ [۸].

۴.۳.۲ رویکردهای تجزیه و تحلیل بافت پوست

یکی دیگر از رویکردهای در حال ظهور، استفاده از بافت 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 تنها یک امکان پذیر نیست. آزمایش‌ها نشان دادند که با افزودن تحلیل بافت پوست، عملکرد سامانه تشخیص چهره می‌تواند ۲۰ تا ۲۵ درصد افزایش یابد.

در سال ۲۰۱۷ Guosheng Hu و همکاران در [۳۹] 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. 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 خطأ را کاهش داده است. برتری روش سه بعدی در عدم وابستگی به حرکت و جا به جایی صورت است. انتقال و نصب سامانه تصویر برداری بسیار ساده است.

زاویه دید حسگر چندان مهم نیست. همچنین نورپردازی نامناسب تاثیری در این شیوه ندارد و عملیات آن ساده است. بر خلاف 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 بی درنگ در شرایط کنترل نشده مانند معابر پیاده و خیابان می باشد، به توضیح مختصر رویکردهای سه بعدی و تجزیه و تحلیل بافت پوست بسنده می کنیم.

۵.۳.۲ رویکردهای مبتنی بر دوربین حرارتی

در این رویکرد، دوربین حرارتی شکل صورت را تشخیص می دهد و از لوازم جانبی مانند عینک، کلاه یا آرایش چشم پوشی می کند. برخلاف دوربین های معمولی، دوربین های حرارتی می توانند 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 مشکل های استفاده از تصاویر حرارتی برای تشخیص چهره این است که مجموعه داده های آن برای شناسایی چهره محدود است.

در سال ۲۰۰۳ Diego Socolinsky و همکاران در [۴۰] 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 می دهد که تلفیق LWIR و دوربین های 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. داده ها در روزهای آفتابی، بارانی و ابری جمع آوری شد. در شرایط کنترل شده دوربین معمولی دقیق ۹۷٪.۵٪ دارد، در حالی که روش LWIR دارای دقیق ۹۳٪.۹٪ می باشد و ترکیب این دو دارای دقیق ۹۸٪.۴٪ است. در شرایط کنترل نشده دوربین معمولی دقیق ۸۶٪.۰٪، دوربین LWIR دقیق ۸۳٪.۰٪ و ترکیب این دو دارای دقیق ۸۹٪.۰٪ است.

۶.۳.۲ تشخیص چهره مبتنی بر ویدیو

در سال ۲۰۰۹ Wang Huafeng و همکاران در [۹] 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 گسترده ای از برنامه های کاربردی تجاری و اجرای قانون را در بر گرفته است. فیلم ها قادر به ارائه اطلاعات بیشتر نسبت به تصاویر ثابت هستند. مزایای عمدی استفاده از ویدیو عبارتند از:

^۱Longwave Infrared

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 رضایت‌بخش می‌باشد.

۲. مطالعات روان‌پژوهی نشان داده است که اطلاعات پویا در فرایند تشخیص فرد بسیار حائز اهمیت است.

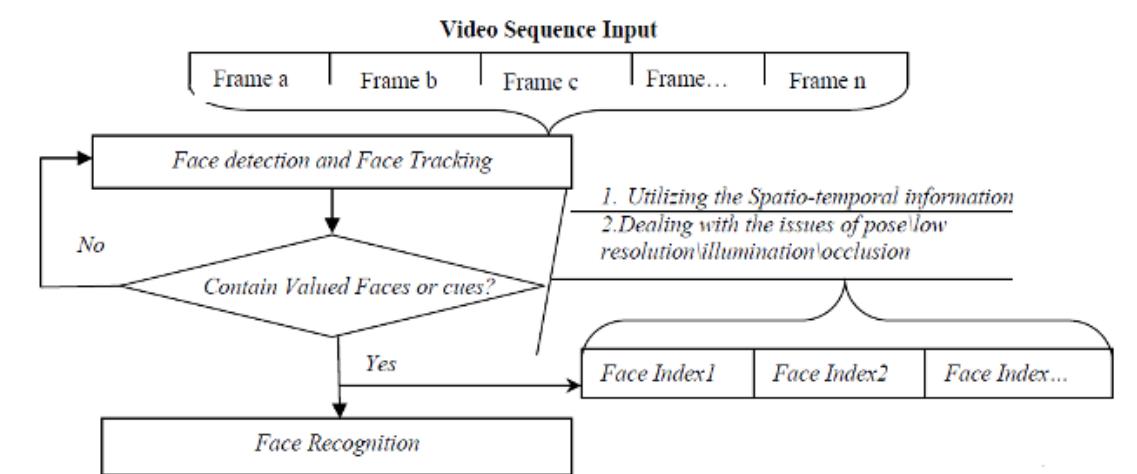
۳. نمایش‌های موثرتر مانند مدل چهره سه بعدی یا تصاویر super resolution می‌توانند از اطلاعات فریم‌های ویدیو گرفته شده برای بهبود شناخت استفاده کنند.

۴. 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^۱: 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.

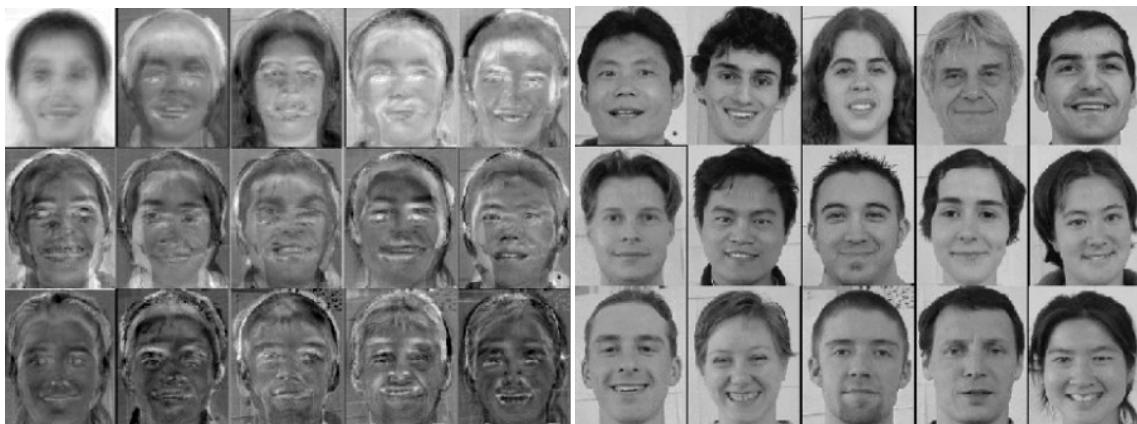
یافتن و ردیابی^۲: یافتن چهره در اولین قاب و سپس ردیابی آن از طریق توالی قاب‌ها. شکل ۱۸.۲ نمای کلی این رویکرد را نشان می‌دهد.



شکل ۱۸.۲: نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [۹].

¹Frame

²Detection And Tracking



شکل ۱۹.۲: (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۱۰].

۷.۳.۲ رویکرد مبتنی بر چهره ویژه

با افزایش حجم داده های موجود، نیاز به کاهش ابعاد داده ها می باشد. تحلیل مؤلفه های اساسی یا PCA^۱ یک روش کاهش ابعاد داده ها است که در برخی مسئله ها مانند پردازش تصویر به خوبی و با سرعت بالا عمل می کند. استفاده از این روش در پردازش سریع تر داده ها کمک می کند و از رخداد مشکل بیش برآزندن^۲ جلوگیری می نماید. 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 بر و در مواردی غیر ممکن خواهد بود. PCA در این مسئله به خوبی عمل می کند. با اعمال تکنیک کاهش بعد به تصویر و با به دست آوردن تصویر ویژه چهره ها^۳ می توان ویژگی ها را کاهش داد و نتیجه مطلوب را در زمان بسیار کم گرفت. شکل ۱۹.۲ تعدادی چهره و چهره های ویژه متناظر با آن ها را نشان می دهد.

در سال ۲۰۱۴ Xiao Luan و همکاران در [۱۰] یک روش تشخیص چهره مبتنی بر PCA ارائه دادند که تا حدی در برابر تغییرات نورپردازی و انسداد مقاوم می باشد. PCA راستای بیشترین تغییرات را با توجه به تعداد ویژگی ها و نوع آن ها به ما می دهد. 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.

در سال ۲۰۱۶ K. R. Sreelakshmi و همکاران در [۴۱] 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 شبیهترین تصویر به آن انتخاب می شود. برای مقایسه از معیارهایی مانند معیار اقلیدسی و منتهن

¹Principle Component Analysis

²Overfitting

³Eigenface

میتوان استفاده کرد. از مزایای این روش میتوان به سهولت پیادهسازی و استفاده، کاهش حجم دادهها و سرعت بالا اشاره کرد. در نظر نگرفتن پراکندگی درون کلاسی و بین کلاسی دادهها و عدم توجه به برچسب تصویر برای شناسایی و تمایز قابل نشنیدن 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 آن‌ها:

۱. تبدیل ماتریس تصاویر به بردار و کنار هم قرار دادن آن‌ها برای تشکیل ماتریس دادهها

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

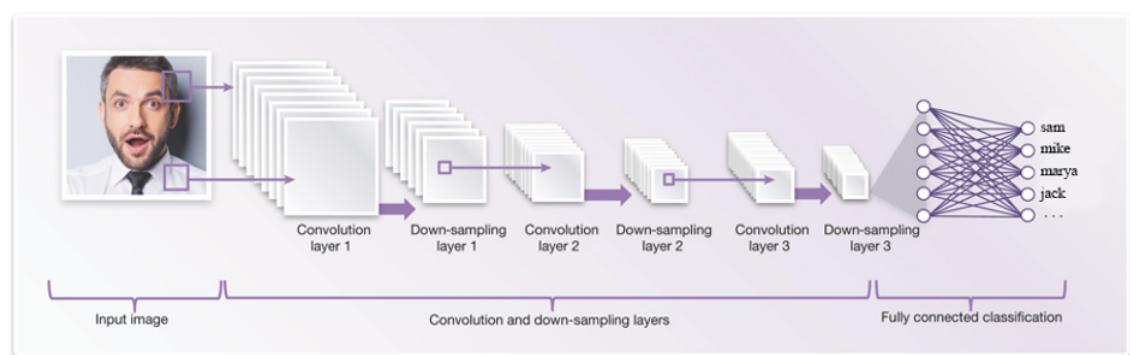
۳. محاسبه ماتریس کوواریانس بردارها و مقادیر ویژه آن

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

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

۸.۳.۲ رویکردهای مبتنی بر شبکه عصبی

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 ۲۰.۲ ساختار کلی یک شبکه عصبی عمیق برای شناسایی چهره را نشان می‌دهد.



شکل ۲۰.۲: شبکه عصبی عمیق برای شناسایی چهره.

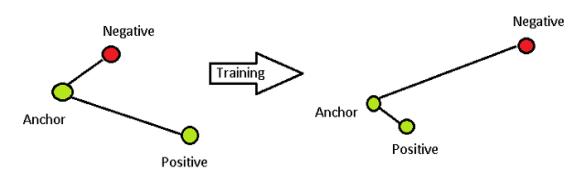
معمولًا به عنوان تابع فعالیت از توابع غیرخطی مانند ReLU ^۱ استفاده می‌شود. و عملیات بهینه سازی به روش پس انتشار خطا

^۱Rectified Linear Unit

انجام می‌گردد. و در لایه خروجی ازتابع 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.

FaceNet شبکه ۱.۸.۳.۲

در سال ۲۰۱۵ Florian Schroff و همکاران در [۱۱] یک شبکه عصبی عمیق به نام FaceNet ارائه دادند. یک مدل یکپارچه 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 فضای تولید شود، شناسایی چهره، تایید هویت و خوشبندی می‌تواند به راحتی با استفاده از روش‌های استاندارد توسط FaceNet انجام شود. این شبکه برای آموزش از سه گانه تطبیق - عدم تطبیق استفاده می‌نماید. با توجه به شکل ۲۱.۲، سه گانه تطبیق - عدم تطبیق یک مجموعه از سه تصویر شامل یک تصویر مرجع، یک تصویر منطبق بر تصویر مرجع و یک تصویر غیر منطبق بر تصویر مرجع است که باید فاصله بین تصویر مرجع و تصویر منطبق را به حداقل برساند، زیرا هر دو دارای هویت مشابه هستند و فاصله بین تصویر مرجع و تصویر غیر منطبق را به حداقل برساند، زیرا این تصاویر دارای هویت متفاوت می‌باشند.



شکل ۲۱.۲: سه گانه تطبیق - عدم تطبیق [۱۱].

برای هر داده آموزشی A مجموعه‌ای از داده‌های مشابه Positive و مجموعه‌ای از داده‌های نامرتبط Negative در نظر گرفته می‌شود. سپس داده‌ها با تابع ضرر سه گانه طوری آموزش می‌بینند که رابطه ۸.۲ برای هر کدام از داده‌های آموزشی بقرار 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 مشابه و غیر مشابه را به راحتی تفکیک نماید. تابع هزینه در این مدل برای هر نمونه آموزشی x به صورت زیر تعریف می‌شود:

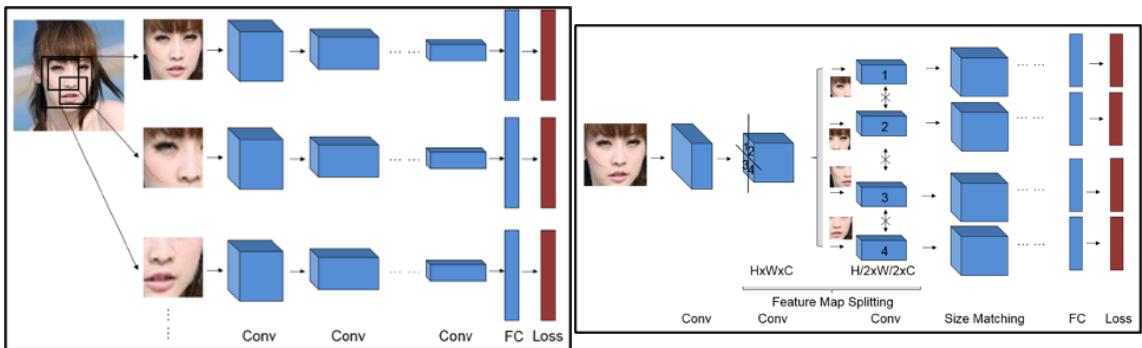
$$L = \sum_{i=1}^n f(x_i^A) - f(x_i^P)^2 - f(x_i^A) - f(x_i^N)^2 + \alpha \quad (9.2)$$

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 ۶۳٪.۹۹ رسیده است و در مجموعه داده YouTube Faces DB دقت آن به ۱۲٪.۹۵ رسیده است.

شبکه SplitNet ۲.۸.۳.۲

در سال ۲۰۱۸ و همکاران در [۱۲] 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 ۲۲.۲ 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.

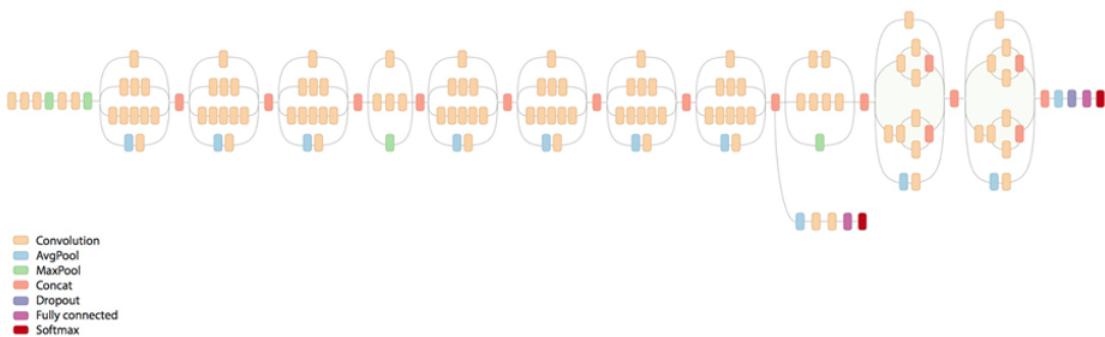
^۱Feed Forward



شکل ۲۲.۲: (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [۱۲].

GoogLeNet شبکه ۳.۸.۳.۲

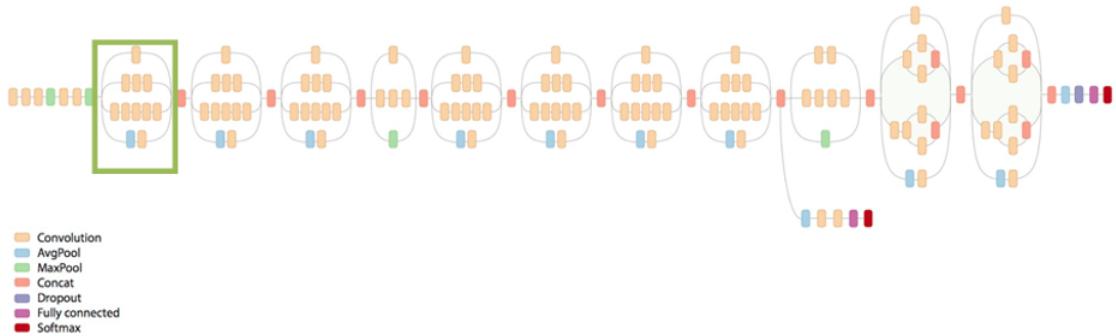
در سال ۲۰۱۵ و همکاران در [۱۳] یک شبکه عصبی عمیق به نام GoogLeNet ارائه دادند. همانطور که در شکل ۲۳.۲ مشاهده می‌شود، GoogLeNet یک شبکه عصبی پیچشی با ۲۲ لایه است که یکی از اولین معماری‌های شبکه عصبی پیچشی بود که از رویکرد کلی قرار دادن تعداد زیادی از لایه‌های پیچشی و رای‌گیری در کنار هم در یک ساختار متواലی بدست آمد. نویسنده‌گان این مقاله همچنین تأکید کردند که این مدل جدید، توجه قابل ملاحظه‌ای به مصرف حافظه و مصرف انرژی دارد، زیرا کنار هم چیدن تعداد زیادی لایه و فیلتر دارای هزینه محاسباتی و حافظه است که احتمال بیش‌برازاندن را افزایش می‌دهد.



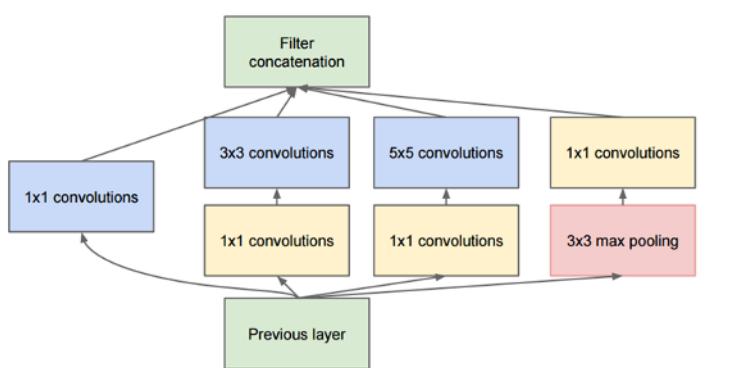
شکل ۲۳.۲: معماری کلی شبکه GoogLeNet.[۱۳]

در تمام محاسبات به طور متواالی اتفاق نمی‌افتد، بلکه هر بخش شبکه روندی موازی دارد. کادر سبز رنگ در شکل ۲۴.۲ بخش آغازگر نامیده می‌شود. در ادامه نگاهی دقیق‌تر به این بخش خواهیم داشت.

کادر سبز پایین در شکل ۲۵.۲ ورودی این بخش و کادر بالایی خروجی می‌باشد. در هر لایه شبکه‌های پیچشی معمولی، باید بین یک لایه پیچشی یا رای‌گیر، یکی را انتخاب نمود. در حالی که اینجا می‌توان تمام این عملیات را به صورت موازی انجام داد. این همان ایده ساده‌ای بود که نویسنده‌گان مقاله روی آن تمرکز کردند.



شکل ۲۴.۲: کادر سبز رنگ یکی از بخش های موازی شبکه را نشان می دهد [۱۳].



شکل ۲۵.۲: بخش آغازگر شبکه [۱۳] GoogLeNet.

۴.۸.۳.۲ شبکه VGGFace

در سال ۲۰۱۵ Omkar M. Parkhi و همکاران در [۱۴] شبکه عمیق VGGFace را ارائه کردند که شامل یک توالی طولانی از لایه های پیچشی می باشد. با توجه به شکل ۲۶.۲ این شبکه که در لایه آخر به عنوان یک طبقه بند عمل می نماید، هر تصویر آموزشی چهره را توسط لایه تماما متصل وتابع ضرر softmax log-loss به یک بردار تبدیل می نماید 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 می نماید.



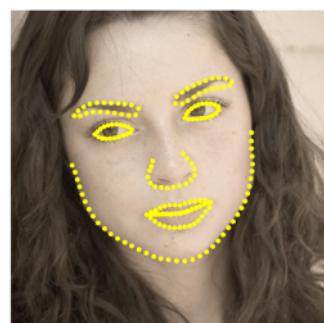
شکل ۲۶.۲: معماری شبکه [۱۴] VGGFace.

^۱Triplet Loss Function

۹.۳.۲ رویکردهای مبتنی بر نقاط راهنما

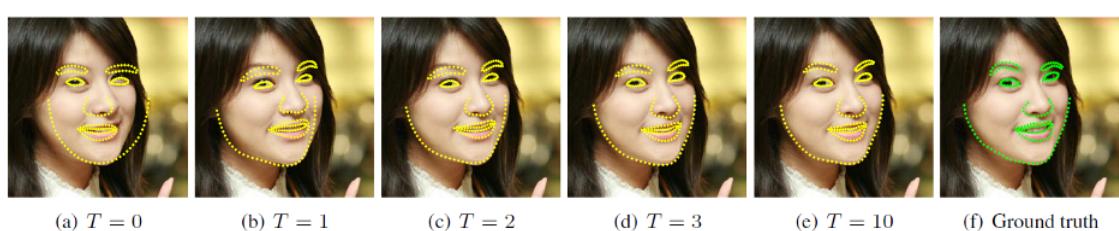
چهره‌هایی که در جهت‌های مختلفی هستند، برای سامانه تشخیص چهره، متفاوت به نظر می‌رسند. برای غلبه بر این چالش در رویکردهای مبتنی بر نقاط راهنما سعی می‌شود تصویر را چرخانده و جایه جا نمود، بطوریکه چشمها و لب‌ها در یک موقعیت خاص در تصویر قرار بگیرند. بدین ترتیب مقایسه چهره‌ها در مرحله بعد بسیار ساده‌تر خواهد شد.

در سال ۲۰۱۴ وحید کاظمی و جوزفین سالیوان در [۱۵] یک الگوریتم برای یافتن نقاط راهنما^۱ بر روی چهره ارائه دادند که از ۱۹۴ نقطه خاص که در هر چهره‌ای وجود دارد استفاده می‌نماید. شکل ۲۷.۲ مکان این نقاط را بر روی گونه، لبه‌های بیرونی چشم، کناره ابرو و... نشان می‌دهد. سپس این ۱۹۴ نقطه به سامانه آموزش داده می‌شود تا در هر چهره‌ای آن‌ها را تشخیص دهد.



شکل ۲۷.۲: نتیجه موقعیت ۱۹۴ شاخص روی چهره [۱۵].

پس از این که دانستیم چشم‌ها، دهان و... کجاست، به راحتی می‌توانیم تناسب تصویر را تغییر داده و آن را چرخانده یا برش 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 ۲۸.۲ علامت گذاری تکراری خطوط راهنما بر روی چهره را نشان می‌دهد.



شکل ۲۸.۲: علامت گذاری خطوط راهنما بر روی چهره که در هر تکرار با کاهش خطأ همراه می‌باشد [۱۵].

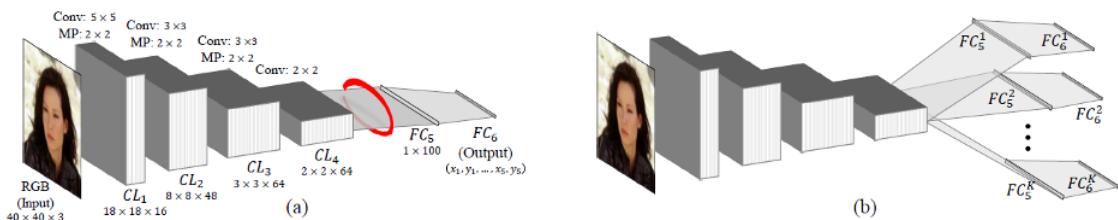
در سال ۲۰۱۶ Yue Wu و همکاران در [۴۲] رویکردی برای یافتن نقاط راهنما بر روی چهره مبتنی بر شبکه عصبی پیچشی ارائه

¹Landmark

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 بروی چهره مشخص می‌کند. در این مقاله m برابر با ۵ در نظر گرفته شده است.تابع فعالیت برای هریک از لایه‌های پیچشی $f(x) = \tanh(x)$ و تابع فعالیت برای لایه تمام متصل $f(x) = |\tanh(x)|$ به عنوان تابع ضرر معرفی شده است.

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

که در آن P_i یک بدار $m * 2$ برای مختصات پیش‌بینی شده تصویر I_i و P_i و مختصات محل دقیق آن نقاط می‌باشد. $P_{i,1}$ و $P_{i,2}$ مختصات چشم‌ها در تصویر مرجع می‌باشند. 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 ۲۹.۲ قسمت b قابل مشاهده می‌باشد. این شبکه برای آموزش از مجموعه داده LFW استفاده کرده است.



شکل ۲: (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری TCNN.

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.

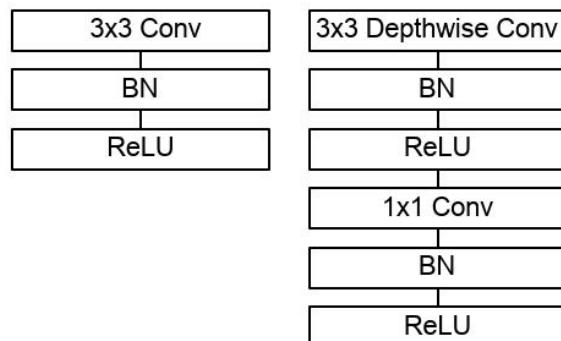
یکی از شاخص‌ترین شبکه‌های سبک، شبکه عصبی MobileNet نام دارد که توسط محققان گوگل در [۱۶] با هدف طراحی شبکه‌های کارآمد، سبک، سریع و با دقت قابل قبول مطرح شده است. در این مقاله یک نوع کانولوشن جدید به نام depth-wise separable convolution معرفی شد که قلب تپنده شبکه موبایل نت است. در کانولوشن dws ابتدا کانولوشن عمیق اعمال می‌شود و سپس

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

کانولوشن نقطه‌ای که به ترتیب نقش مراحل فیلتر و ادغام در کانولوشن استاندارد را دارد.

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 کرنل در یک صفحه ویژگی ورودی F کانولوشنی 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.



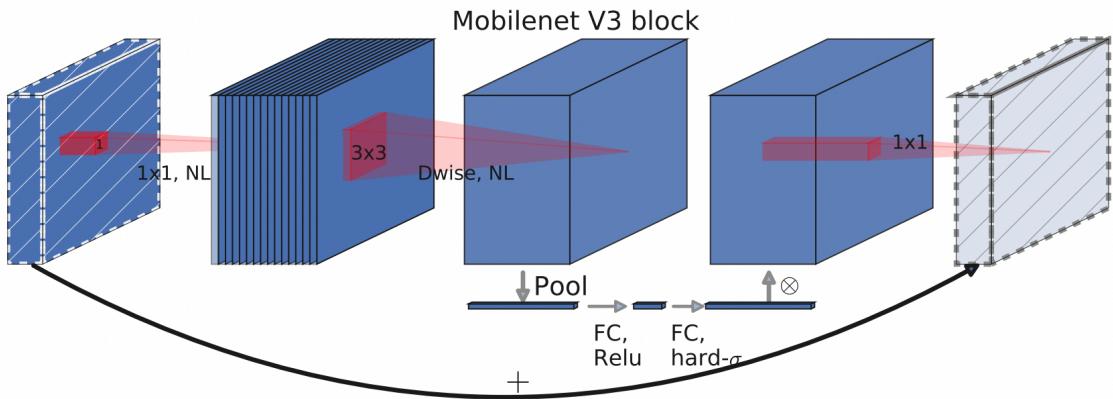
شکل ۳۰.۲: کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن depth-wise و point-wise هست (سمت راست). [۱۶].

شبکه عصبی موبایل نت ۲۰۴ 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 در [۴۳] معماری MobileNet نسخه ۳ را ارائه دادند. در این معماری مسیر

¹point-wise

استخراج 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 که در شکل قابل مشاهده می‌باشد.



[۱۷] شکل ۳۱.۲: لایه‌های MobileNet به همراه مازول‌های Squeeze-and-Excite

در سال ۲۰۲۱ yang و همکاران در [۱۷] یک مازول مبتنی بر لایه توجه ارائه دادند. لایه‌های توجه که یک شبکه عصبی را قادر می‌سازد تا دقیقاً بر روی تمام عناصر مربوط به ورودی مرکز شود، به یک جز اساسی برای بهبود عملکرد شبکه‌های عصبی عمیق تبدیل شده است. عمدتاً دو مکانیسم توجه به طور گسترده 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 SA کارآمد برای پرداختن به این مسئله پیشنهاد شده است که آن را به اختصار SA مازول^۵ نامیم، ابتدا ابعاد کanal 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 کanal برای امکان برقراری ارتباط اطلاعاتی بین ویژگی‌های فرعی مختلف به کار گرفته می‌شود. معماری این مازول در شکل ۳۲.۲ آمده است.

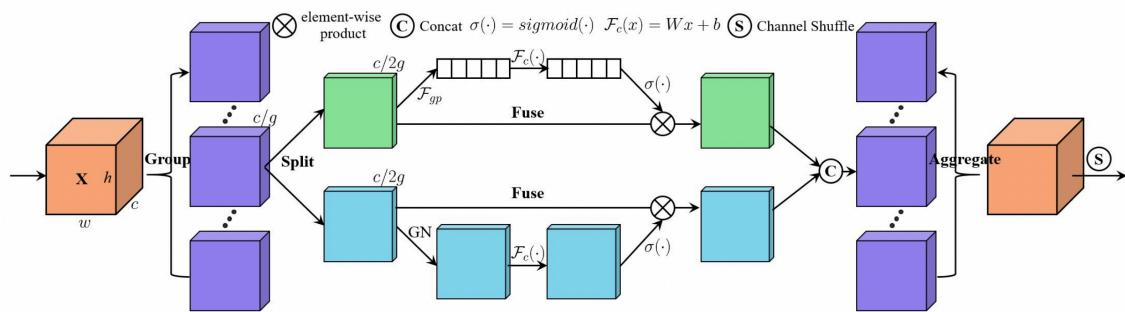
مازول SA بهینه و در عین حال کارآمد است، به عنوان مثال، پارامترها و محاسبات SA در شبکه ResNet^۶ ۳۰۰ در مقابل

¹Bottleneck

²Attention

³Channel Attention Module

⁴Spatial Attention Module



. [۱۷] Attention Shuffle معماری مازول

۵۶.۲۵ میلیون است، اما افزایش عملکرد بیش از ۳۴٪.۱ را به ارمغان می‌آورد. نتایج تجربی نشان می‌دهد که SA برای دستیابی به دقث بالاتر مناسب است، در حالی که دارای پیچیدگی مدل کمتری است و از روش‌های SOTA^۱ کنونی به مراتب بهتر عمل می‌کند.

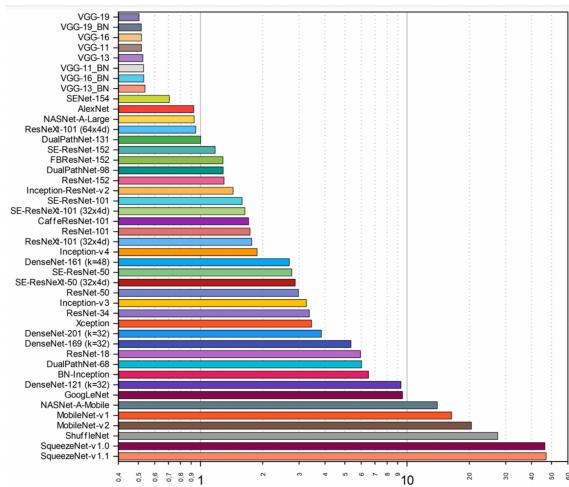
۴.۲ نتیجه گیری

در این فصل مفاهیم پایه در مبحث یافتن و تشخیص چهره در تصاویر و انواع الگوریتم‌های دسته بندی از روی تصاویر رنگی بررسی شد. همانطور که در قبل نیز بیان شد، برای حل مساله دسته بندی چهره دور روش کلی، مبتنی بر تصویر و روش‌های مبتنی بر استخراج ویژگی وجود دارد. همچنین 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: همچنین روش‌های مبتنی بر استخراج ویژگی که در سال‌های اخیر بسیار مورد توجه قرار گرفته‌اند شامل رویکردهای مبتنی بر شبکه عصبی، بردار پشتیبان و... می‌باشند.

با بررسی شبکه‌های به روز از جمله

MobileNetV2، MobileNet، NASNetMobile، SqueezeNet، VGG19، ResNet-50، EfficientNetB0 و مقایسه دقث وزمان پاسخگویی آن‌ها به کمک یادگیری انتقال، به این نتیجه می‌رسیم که شبکه‌های SqueezeNet و MobileNetV2 دارای چگالی دقث بالاتری می‌باشند و نسبت دقث دسته بندی به تعداد پارامترهای شبکه در آن‌ها بیشتر می‌باشد. بنابرین می‌توان سرعت اجرای مناسب و همچنین دقث مناسب را از آن‌ها انتظار داشت. نتایج بررسی در شکل ۳۳.۲ آمده‌اند.

¹State of the Art



شکل ۳۳.۲: مقایسه چگالی دقت در معماری‌های مختلف شبکه عصبی پیچشی [۱۸].

٣ فصل

مروری بر کارهای گذشته در 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 تمرکز، سالخوردگی و... را ندارند و در کاربردهایی نظیر 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. جدول^۱ خلاصه ای از روش های مقابله با شرایط کنترل نشده

مقاله ها	چالش مورد نظر	رویکرد	مزیت ها	مشکل ها
[۴۴،۲۴،۲۱،۲۰]	حالات چهره	تبديل دو بعدی	پیچیدگی محاسباتی قابل قبول	استخراج نقاط ویژه باید دقیق تر باشد
[۴۷،۲۵،۴۲،۲۲]	حالات چهره	استفاده از شبکه عصبی عمیق	دقت بالا در شرایط کنترل نشده	پیچیدگی محاسباتی، وابستگی به داده های آموزش
[۵۱،۴۸،۴۹]	حالات چهره	تبديل مدل دو بعدی به سه بعدی	دقت بالا پیچیدگی محاسباتی	
[۵۲]	حالات چهره	تبديل مدل سه بعدی به دو بعدی	دقت بالا پیچیدگی محاسباتی	
[۵۳،۰۴]	روشنایی	همسان سازی بافت-نگار	پیچیدگی محاسباتی قابل قبول	قابل استفاده در تصاویر خاکستری
[۵۴،۰۵]	انسداد	استفاده از روش های شناسایی الگو و واپاش	دقت بالا در انسداد شدید	
[۷۶]	محدودیت داده	تهییه مجموعه داده با دقت بالا	تهییه مجموعه داده با دقت بالا	نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو
[۳۲،۰۱]	محدودیت منابع	استفاده از رایانش ابری	سرعت بالا	تجهیزات پیشرفته و زمان تاخیر ناهمگن

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 یک پی برنده^۲ 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^۳ و هم ترازی^۴ چهره می باشد. در ادامه برخی رویکردهای رو به رو سازی و هم ترازی چهره در شرایط کنترل نشده را دسته بندی می کنیم.

۱. رویکرد های دو بعدی با پیچیدگی محاسباتی قابل قبول (بیشتر ایده های مبتنی بر نشانه گذاری^۴ قدیمی) مانند [۲۰، ۲۱، ۰۲]،

¹Heuristic

²Frontalization

³Alignment

⁴Landmark

مشکل: در محیط های بدون محدودیت مانند آنچه که در این پروژه داریم، استخراج دقیق مکان نشانه های صورت از تصاویر دو بعدی نیاز به توجه بیشتری دارد. پیشرفتهای اخیر مانند [۲۱] است.

مزیت: این الگوریتم ها از نظر پیچیدگی محاسباتی^۱ 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 [۲۲، ۴۲].

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 [۵۵] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند.

۳. رویکردهای سبک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۳۹، ۴۸-۵۱].
مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده ترین این الگوریتم ها در مورد پیچیدگی محاسباتی، [۴۸] 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^۲ بر اساس چند نمایش دو بعدی مختلف از چهره) مانند [۵۲] 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.

مرجع [۱۹] در سال ۲۰۱۸ خلاصه ای از رویکردهای مختلف برای حل مسئله هم ترازی را در شکل ۱.۳ نشان داده است.

تصویر سمت چپ، چهره ورودی می باشد. (a) هم ترازی با استفاده از تبدیلات دو بعدی ساده می باشد. (b) داده افزایی^۳ با تغییر

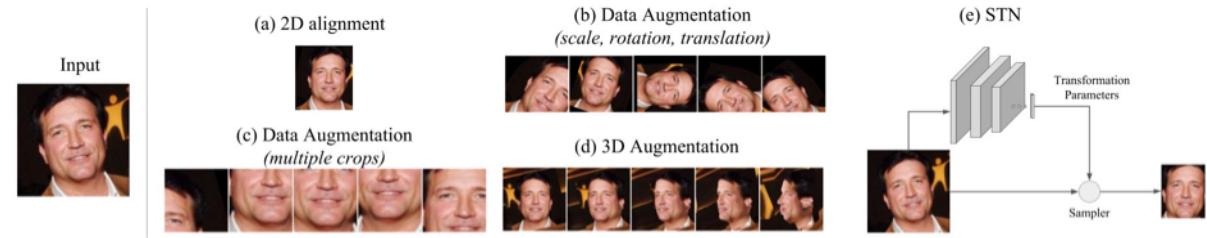
¹Computational Complexity

²Patch-Based

³Data Augmentation

مقیاس، تغییر زاویه و جا به جایی می‌باشد. (c) برش‌های چندگانه می‌باشد. (d) داده افزایی مبتنی بر روش‌های سه بعدی می‌باشد.

(e) از هیچ ابزاری برای هم ترازی مستقیم استفاده نمی‌نماید. اما یک شبکه را آموزش می‌دهد تا عامل‌های مورد نیاز برای تبدیل هم ترازی را بدست آورد.



شکل ۱۰.۳: رویکردهای مختلف هم ترازی چهره [۱۹].

در سال ۲۰۱۶ Brandon Amos و همکاران در [۲۰] 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 دستگاه‌های قابل حمل مانند تلفن همراه با در نظر گرفتن متابع

محدود می‌باشد. یک تصویر شامل تعدادی چهره به الگوریتم داده می‌شود. پس از یافتن چهره‌ها و مجزا کردن^۱ آن‌ها از یکدیگر، هر

چهره به طور جداگانه مورد پیش پردازش^۲ قرار می‌گیرد و حجم آن کاهش می‌باید. کاهش حجم تصویر برای عملکرد مناسب یک

طبقه‌بندی بهینه بسیار مهم می‌باشد. تصاویر چهره‌ها باید هنجارسازی شده و ابعاد آن‌ها ثابت گردد تا به بخش شناسایی چهره راه

یابند. هر تصویر چهره باید مورد تبدیل قرار بگیرد تا چشم‌ها، بینی و دهان، در مکان مشخصی قرار گیرند. بدین منظور از یک تبدیل

هم نسبی^۳ دو بعدی ساده استفاده می‌گردد. ابتدا باید چهره توسط ۶۸ نقطه ویژه، نشانه گذاری شود. سپس نشانه‌های اطراف چشم

ها و بینی (شکل ۲.۳) برای محاسبه عامل‌های تبدیل هم نسبی استفاده می‌شوند. پس از انجام تبدیل هم نسبی، تصاویر چهره برش

زده شده و اندازه آن‌ها 96×96 پیکسل می‌شود.

پس از پیش پردازش، تصاویر چهره‌ها به عنوان ورودی به یک شبکه عصبی پیچشی داده می‌شوند (شکل ۳.۳). این الگوریتم برای

تعلیم شبکه از مجموعه داده کوچکی با ۵۰۰ هزار تصویر چهره استفاده می‌کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری

شده به نام FaceScrub و CASIA-WebFace بدست آمده است. شبکه مورد استفاده در این الگوریتم یک نسخه اصلاح شده

از شبکه nn4 الگوریتم FaceNet می‌باشد. شبکه nn4 مبتنی بر معماری GoogLeNet می‌باشد. برای تعیین میزان شباهت

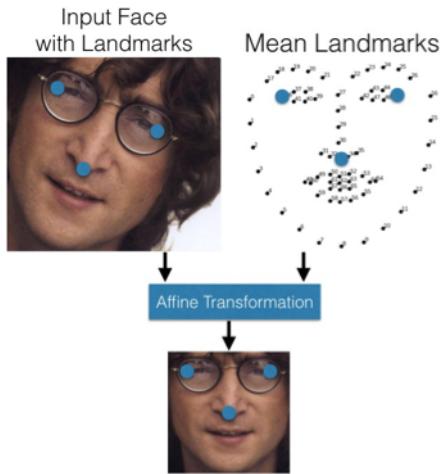
نتیجه، از فاصله اقلیدسی استفاده شده است.

هر تصویر از یک شبکه یکتا به یک سه گانه نگاشت داده می‌شود. گرادیان خطای سه گانه برای هر تصویر محاسبه شده و به عقب انتشار

¹Isolate

²Preprocessing

³Affine Transformation



شکل ۲.۳: تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی [۲۰].

می یابد. در هر دسته کوچک^۱, P 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 forward انجام پذیرد. در این مقاله از $Q = 15$ و $P = 20$ استفاده شده است. تمام جفت های anchor-positive برای بدست آوردن سه گانه های $N = Q^P / 2$ مورد استفاده قرار می گیرند. خطای سه گانه محاسبه شده و مشتق آن برای پس انتشار خطا استفاده می شود. شکل ۴.۳ چگونگی آموزش شبکه را نشان می دهد.

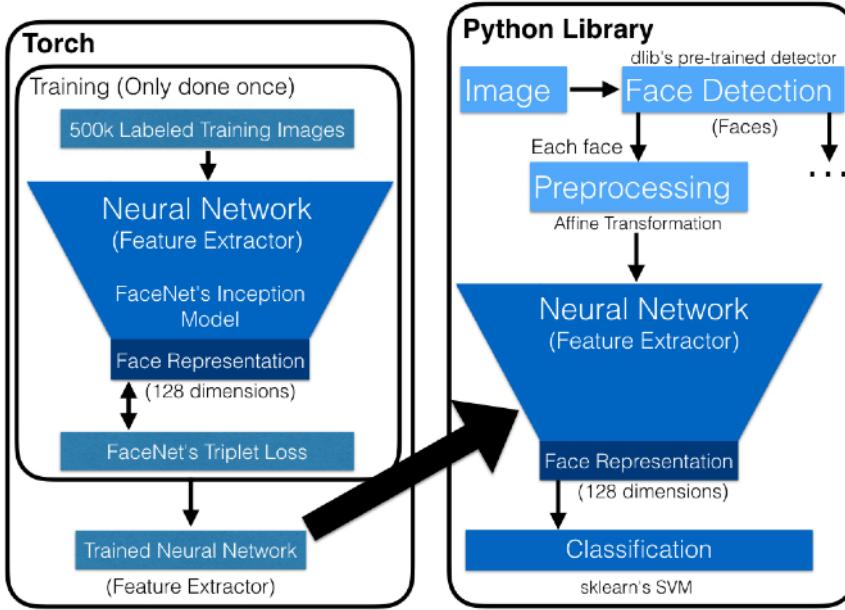
مجموعه داده LFW یک معیار استاندارد برای سنجیدن میزان دقیقیت الگوریتم های تشخیص چهره می باشد. الگوریتم OpenFace بر روی این مجموعه داده مورد سنجش قرار گرفت که به دقت $0.9292 \pm 0.0134\%$ رسید.

در سال ۲۰۱۶ Mohammad Haghhighat و همکاران در [۲۱] یک روش برای هنجارسازی حالت چهره بر اساس تنظیم کردن مدل ظاهری فعال^۲ یا AAM ارائه دادند. AAM یک مدل پارامتری است که برای ارائه یک شکل مانند چهره انسان استفاده می شود. در این الگوریتم ابتدا یک AAM بر روی تصویر چهره قرار گرفته، با روندی تکراری و به صورت بهینه شونده، بر روی چهره تنظیم می شود. سپس با استفاده از یک تبدیل هم نسبی، مرحله رو به رو سازی بر روی چهره انجام می پذیرد. شکل ۵.۳ رویکرد کلی این الگوریتم را نشان می دهد.

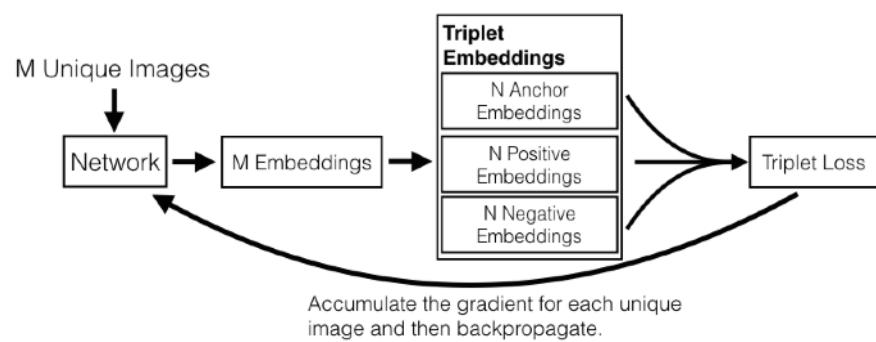
در این مدل یک تصویر چهره با مجموعه ای از نقاط ویژه هنجارسازی شده مدل می شود که به صورت $[x_i, y_i]$ تعریف می شود که در آن $n \dots 1 \dots 2 \dots i$. برای انجام این کار یک مرحله یادگیری نیاز است. سپس الگوریتم PCA اعمال می شود تا کاهش میزان وابستگی میان نقاط ویژه در هر مجموعه انجام می شود و نتیجه یک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

¹Mini Batch

²Active Appearance Model



. [۲۰] OpenFace معماری .

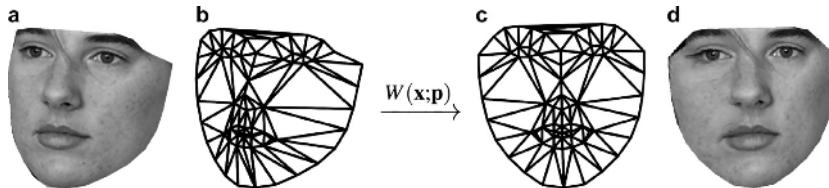


. [۲۰] OpenFace در معماری جریان یادگیری

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

که در آن s_0 شکل پایه، s_i نشان دهنده i امین شکل پایه و $[p_n, p_2, \dots, p_1]$ عامل های شکل می باشند. ظاهر^۱ مدل AAM یک تصویر (x) می باشد که در آن x مجموعه پیکسل های داخل شکل پایه s_0 می باشد. مدل ظاهر یک چهره خاص از یک ظاهر پایه تشکیل می شود که به صورت رابطه ۱.۴ تعریف می گردد.

¹ Appearance



شکل ۵.۳: رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲۱].

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

که در آن $[q_1, q_2, \dots, q_m]$ عامل های ظاهر می باشند. عامل های شکل و ظاهر برای هر تصویر در فرایند AAM بدست می آید.

الگوریتم های POIC^۱ و SIC^۲ دو الگوریتم شناخته شده برای این منظور می باشند. رویکرد SIC نسبت به POIC در شرایطی که

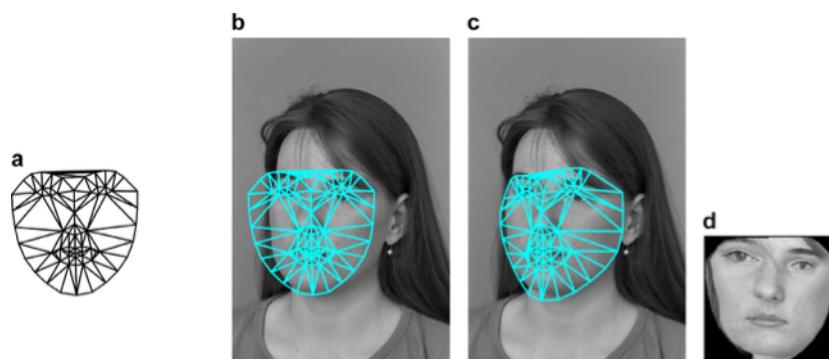
تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می باشد.

در این مقاله از یک روش SIC سریع برای حل مسئله بهینه سازی با ۱۰۰ تکرار استفاده شده است. اگر $p = [p_1, p_2, \dots, p_n]$

مجموعه عامل های بدست آمده باشد، یک تبدیل هم نسبی قطعه ای^۳ $W(x; p)$ برای رو به رو سازی چهره مورد استفاده قرار می

گیرد که در آن هر یک 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 ۶.۳ قسمت a).



شکل ۶.۳: مقدار دهی اولیه و بهینه سازی AAM [۲۱].

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

¹Project-Out Inverse Compositional

²Simultaneous Inverse Compositional

³Piecewise Affine Transformation

⁴Nearest Neighbor Interpolation

شناسایی یک چهره ندارد. سپس تصویر چهره به اندازه 64×64 پیکسل تبدیل می‌شود و به 64×8 بخش غیر هم پوشان با اندازه 8×8 تقسیم می‌شود. سپس در هر بخش تبدیل DCT^۱ انجام می‌شود. ضرایب خروجی تبدیل DCT بر حسب یک پویش زیگزاگی مرتب می‌شوند. اولین ضریب در نظر گرفته نمی‌شود. زیرا نشان دهنده میانگین سطح خاکستری پیکسل های بخش می‌باشد. 10 ضریب بعدی که ضرایب فرکانس پایین می‌باشند، برای ایجاد بردار ویژگی چهره استفاده می‌شوند. برای آموزش و آزمایش از مجموعه داده FERET و LFW استفاده شده است که در آن تصاویر چهره با زوایای چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این مقاله موفق به دستیابی به شناسایی چهره با دقت 87.3% شده است. در شکل ۷.۳ نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت قابل مشاهده می‌باشد.



شکل ۷.۳: نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۲۱].

در سال ۲۰۱۶ Zhang و همکاران در [۲۲] یک روش رو به رو سازی چهره ارائه دادند که شناسایی چهره را مستقل از نمای چهره^۲ انجام می‌دهد. این الگوریتم یادگیری عمیق که VS2VI نامیده می‌شود، از دو بخش اصلی تشکیل شده است. بخش اول یک شبکه عصبی پیچشی برای یادگیری نما و زاویه چهره می‌باشد و بخش دوم از تعدادی شبکه عصبی پیچشی تشکیل شده است که هر کدام برای یادگیری تناظر^۳ بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص می‌باشد (شکل ۸.۳). این الگوریتم که می‌تواند با تعداد کمی داده نمونه، به خوبی آموزش بینند، دو بخش تشکیل شده از شبکه عصبی پیچشی را به هم متصل می‌نماید تا مشکل نمای چهره در سامانه شناسایی چهره را بطرف نماید. در این معماری برای بازسازی چهره از زاویه رو به رو از لایه های واپیچشی^۴ به جای لایه های تمام متصل استفاده شده است.

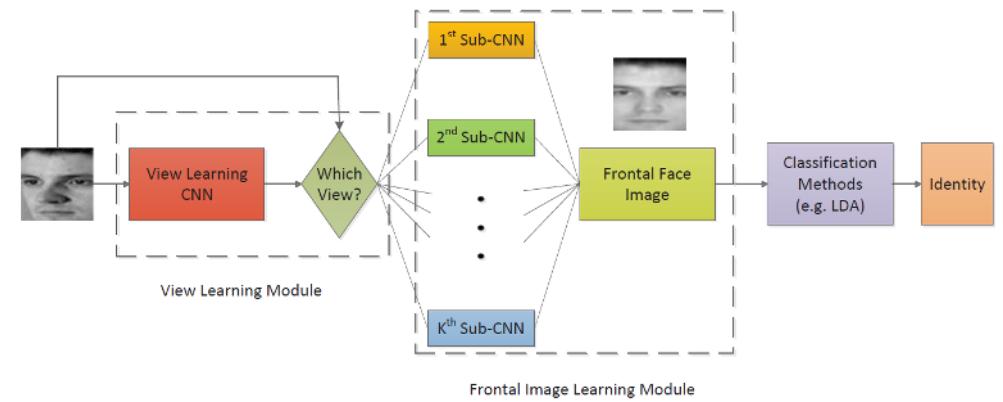
مدل VS2VI از دو بخش اصلی تشکیل شده است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با ابعاد 60×60 دریافت می‌کند و آن را با توجه به نمای چهره طبقه بندی می‌کند. سپس تصویر وارد بخش دوم می‌شود که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از رو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

¹Discrete Cosine Transform

²Facial View

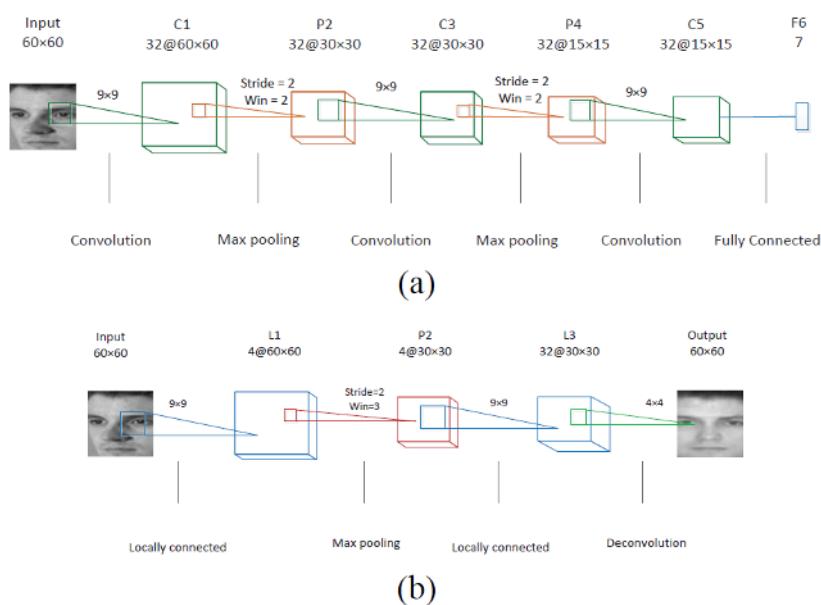
³correspondence

⁴deconvolutional



شکل ۸.۳: معماری شبکه پیشنهادی [۲۲].

می باشد، تشکیل شده است. در این بخش چهره با نمای رو به رو بدست می آید و را مورد شناسایی قرار می دهیم 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 در شکل ۹.۳ قابل مشاهده می باشد.



شکل ۹.۳: (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [۲۲].

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

¹linear discriminant analysis

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 می‌باشد، تشکیل شده است. هر یک از این زیر شبکه‌ها شامل دو لایه با اتصال محلی، یک لایه رای گیری و یک لایه واپیچشی می‌باشند. سه لایه اول برای استخراج ویژگی‌ها و لایه آخر برای بازیابی چهره از رو به رو می‌باشند. ورودی و خروجی این لایه‌ها تصویر چهره می‌باشد. لایه آخر به جای لایه تمام متصل از لایه واپیچشی استفاده شده است. زیرا حجم محاسبات را به طور قابل توجهی کاهش می‌دهد. یک لایه تماماً متصل به 10^3 میلیون پارامتر نیاز دارد، در حالی که لایه واپیچشی به 46^0 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 و تابع هزینه آن $loss = \ell_2$ می‌باشد. برای یادگیری شبکه از الگوریتم پس انتشار خط ^۱ استفاده شده است. الگوریتم VS2VI به دقت ۹۵٪ در تشخیص چهره با زاویه ۴۵ درجه رسیده است.

در سال ۲۰۱۸ Andrey V.Savchenko و همکاران در [۲۳] 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 مجموعه داده‌ها با استفاده از قانون بیز به حداقل می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی SENet و ResNet استخراج شده است ارائه دادند. این روش جدید شناسایی آماری، احتمال فاصله‌ها را نسبت به تمام تصاویر Kullback–Leibler می‌تواند با استفاده از قانون بیز به حداقل می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی IJB-A و YTF، LFW و قرار مجموعه داده‌های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده‌های گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقت ۳.۰ تا ۵.۵ درصد در مقایسه با روش‌های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند. مقایسه روش ارائه شده با سایر روش‌ها در شکل ۱۰.۳ مشاهده می‌شود.



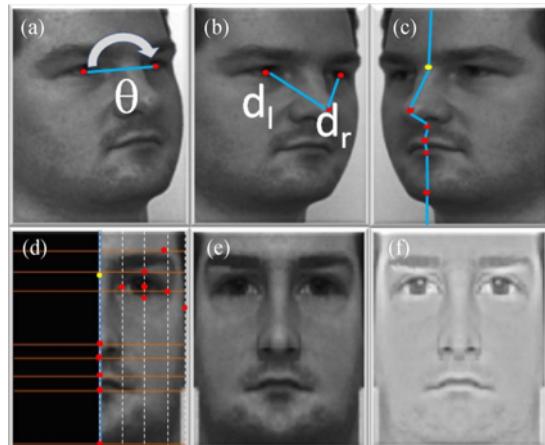
شکل ۱۰.۳: مقایسه روش ارائه شده با سایر روش‌ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش‌های دیگر (d) خروجی روش ارائه شده [۲۳].

^۱Backpropagation

^۲Maximum Likelihood

در سال ۲۰۱۳ میلادی Marsico و همکاران در [۲۴] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم STASM^۱ برای به دست آوردن ۶۸ نقطه ویژه بر روی چهره استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه (SP)^۲ 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 ۱۱.۳ قسمت a تا c معیارهای مورد نیاز برای محاسبه شاخص SP را نشان می‌دهد.



شکل ۱۱.۳: ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۲۴].

چرخش: چرخش سر در جهت عقربه‌های ساعت یا عکس آن می‌باشد. و طبق رابطه ۳.۳ به صورت زاویه θ تعریف می‌شود که زاویه بین خط عبوری از مرکز چشم‌ها و محور افقی x می‌باشد.

$$roll = \min\left(\left|\frac{2\theta}{\pi}\right|, 1\right) \quad (3.3)$$

انحراف: چرخش در راستای محور افقی است و طبق رابطه ۴.۳ مقادیر r و l 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.

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

¹Extended Active Shape Model

²Simple Pose

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

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

با محاسبه ۳ شاخص فوق، شاخص SP مطابق رابطه ۶.۳ محاسبه می شود:

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

که در آن $\alpha + \beta + \gamma = 1$ می باشد که مقادیر این ضرایب از طریق آزمون و خطای دست به دست می آیند. سپس در مرحله تمام رخ کردن تصویر چهره، بین دو فاصله dr و dl هر کدام بزرگتر باشند، نشان می دهد 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 می ماند. در غیر این صورت، تصویر حول محور عمودی برعکس می شود که باعث می شود همیشه نیمه سمت راست تصویر پردازش شود. سپس برای ثابت کردن طول سطرها، سطرها بسط داده می شوند. مطابق شکل ۱۱.۳ قسمت d و e نیمه سمت چپ تصویر حذف شده و از روی تصویر نیمی از چهره، نیمه دیگر نیز ساخته می شود و تصویر تمام رخ چهره به دست می آید.

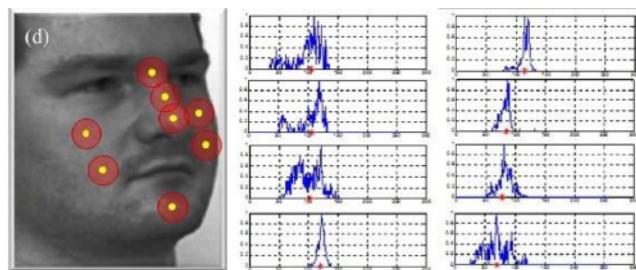
۳.۳ چالش روشنایی

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 بخش های از تصویر که به علت روشنایی کم یا زیاد، پنهان می باشند، 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 در [۲۴] 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

¹Sample Illumination

یکنواخت سطح خاکستری دارند. اما وقتی روشنایی یکنواخت نباشد، برخی از نواحی خاص چهره، توزیع یکنواخت سطح خاکستری ندارند. برای مثال جلوی بینی، گونه‌ها و چانه معمولاً نور را منعکس می‌کنند. ۸ ناحیه در شکل ۱۲.۳ با توجه به چنین اصلی انتخاب شده‌اند. ۸ بافت نگار با رنگ آبی و مرکز آن‌ها با رنگ قرمز مشاهده می‌شود.



شکل ۱۲.۳: اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۲۴].

۸ بافت نگار فوق به یک توزیع یکنواخت با انحراف معیار کم در همسایگی از مرکز حجم بافت نگار اشاره دارد. بافت نگار هر یک از ناحیه‌ها بدست آمده و مرکز ثقل آن مطابق رابطه ۷.۳ محاسبه می‌شود:

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

که در آن w نشان دهنده یکی از نواحی ۸ گانه می‌باشد. ۸ مرکز جرم محاسبه شده، بردار mc را تشکیل می‌دهند. با توجه به فرض تشابه ذکر شده در میان نواحی صورت در نظر گرفته شده، انتظار می‌رود هیچ تنوع قابل توجهی در میان عناصر بردار وجود نداشته باشد و توزیع‌های یکسانی از سطوح خاکستری را نمایش دهند. برای دستیابی به این منظور پراکندگی مراکز حجم‌ها از ۸ نمودار بافت نگار محاسبه شده است. سپس عناصر بردار mc توسط تابع سیگموید F در بازه $[0, 1]$ هنجارسازی می‌شوند و شاخص کیفیت روشناختی 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 طبقه-بندی کند. در هر صورت بیشتر بار طبقه-بندی بر دوش سامانه

خواهد بود. 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 قسمت f در شکل ۱۱.۳ می‌باشد.

در سال ۲۰۱۵ Jamal Hussain Shah و همکاران در [۵۳] رویکردی برای تشخیص چهره در تغییرات شدید روشنایی پیشنهاد دادند کرده اند که به سه مرحله تقسیم شده است:

۱. برای اصلاح روشنایی غیر یکنواخت، همسان سازی بافت نگار براساس بر اساس ناحیه استفاده می‌شود.

۲. ویژگی‌های مبتنی بر LDA از تصویر چهره استخراج می‌شود.

۳. فرایند طبقه‌بندی بر اساس مدل OPPM انجام می‌شود.

۴.۳ چالش انسداد

در سال ۲۰۱۸ Cho Ying Wu و همکاران در [۲۵] یک رویکرد مبتنی بر واپاش با جهت گرادیان برای شناسایی چهره‌های در معرض انسداد 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-HASLR بهترین عملکرد را

در مقایسه با سایر روش‌های پیشرفتی، از جمله روش‌های مبتنی بر شبکه عصبی پیچشی دارد. مجموعه داده آموزشی $A = \mathbb{R}^{d \times n}$ در نظر گرفته شده که در آن n تعداد داده‌های آموزشی و d حاصل ضرب تعداد پیکسل‌های طول و عرض تصاویر می‌باشد. داده‌های آموزشی چهره‌های طبیعی و بدون انسداد می‌باشند. $y = \mathbb{R}^d$ یک داده آزمایشی می‌باشد. مطابق رابطه 9.3 می‌توان از یک ترکیب خطی داده‌های آموزش برای تخمین زدن داده آزمایش استفاده کرد که شامل یک عبارت خطی $L = \mathbb{R}^d$ نیز می‌باشد.

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

که در آن x بردار ضرایب با n بعد می‌باشد. (شکل ۱۳.۳)

برای آنکه شرط تنک بودن به رابطه بالا اضافه شود، مسئله به صورت رابطه 10.3 نوشته می‌شود:

شکل ۱۳.۳: تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۲۵].

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

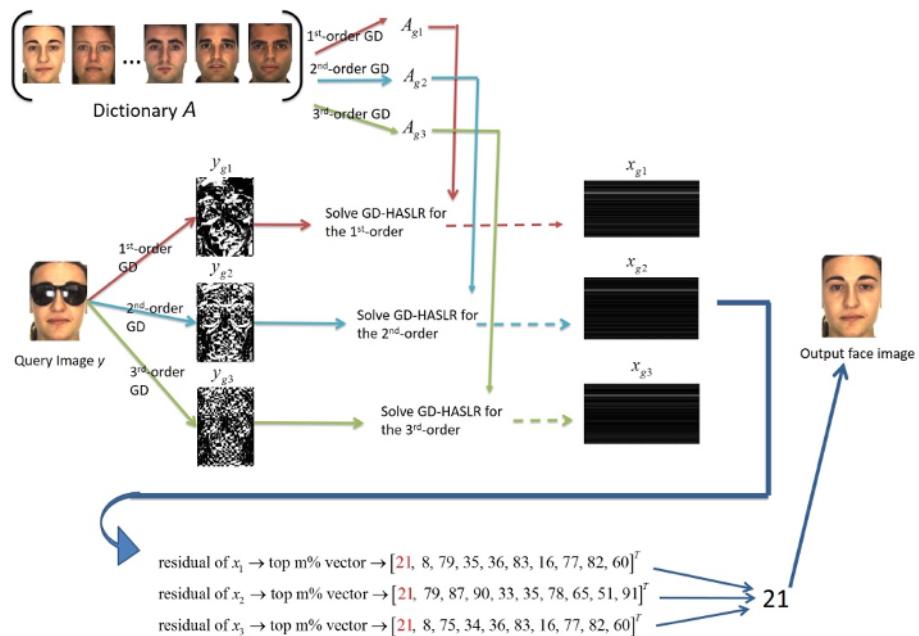
که در آن ϵ یک آستانه خطای باشد. برای تصویر ورودی و تصاویر مجموعه داده آموزش، گرادیان مرتبه اول، دوم و سوم محاسبه شده و به عنوان ویژگی هر تصویر در نظر گرفته می شود. در ادامه شرط کم رتبه بودن ماتریس ویژگی ها نیز به این رابطه اضافه می شود. با استفاده از روش ضرایب لاغرانژ، رابطه بالا را می توان به صورت یک 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 می باشد. پس از بدست آوردن بردار تنک x می توان باقیمانده دسته i ام را به صورت رابطه ۱۲.۳ محاسبه نمود:

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

که در آن $(x)_i \delta$ نشان دهنده i امین انتخاب کننده دسته می باشد که فقط ورودی های مربوط به دسته i ام را حفظ می کند و در سایر قسمت ها برابر با صفر می باشد. در نهایت دسته ای که کمترین باقیمانده را داشته باشد، انتخاب می شود. رویکرد کلی الگوریتم در شکل ۱۴.۳ آمده است.



شکل ۱۴.۳: رویکرد کلی الگوریتم [۲۵].

در سال ۲۰۱۴ J. Li و همکاران در [۵۴] 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.

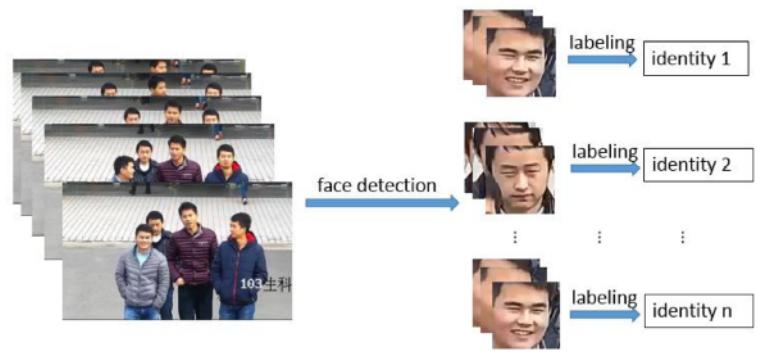
در مرحله اول یک رویکرد تشخیص شی در پیش زمینه در حالت پویا و ایستا پیشنهاد شده است. برای تشخیص هدف ایستا از تشخیص مبتنی بر ویژگی HOG استفاده شده است. از آنجا که سرعت HOG نسبتاً پایین است، از LBP به همراه آن نیز استفاده کرده اند. در مرحله دوم از طبقه بند Adaboost برای طبقه بندی چهره های پوشیده شده است که برای انواع پوشیدگی آموزش داده شده است.

۵.۳ چالش کمبود تصاویر آموزشی

دلیل اصلی به وجود آمدن چالش 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. 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 در [۲۶] روشی برای تشخیص چهره در دوربین‌های ناظارتی 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 بیرون عملکرد سامانه نیاز است، نوآوری این رویکرد، ساختن یک مجموعه داده استاندارد برای شبکه عصبی از روی دوربین‌های ناظارتی 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 به وسیله الگوریتم KCF، مجموعه تصاویری از آن به عنوان یک دسته طبقه بندی می‌شود.



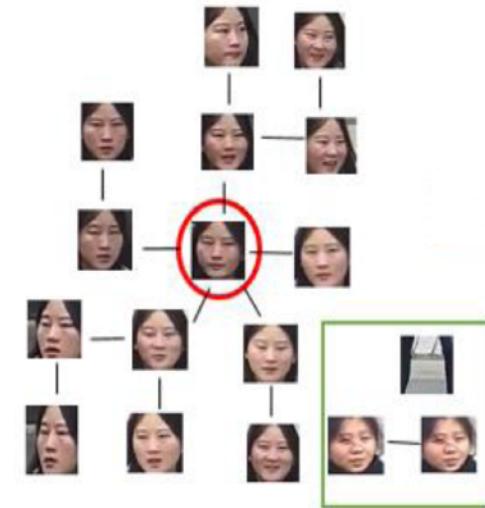
شکل ۱۵.۳: ردیابی، یافتن چهره‌ها و برچسب زنی [۲۶].

برخی تصاویر در هر دسته به اشتباه در مرحله اول به عنوان تصویر یک فرد در نظر گرفته شده اند (شکل ۱۶.۳).



شکل ۱۶.۳: تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند [۲۶].

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 ۱۷.۳ تصویری که بیشترین شباهت را به تصاویر دیگر دارد، به عنوان شاخص برای آن شخص انتخاب می شود.



شکل ۱۷.۳: استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [۲۶].

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٪.۹٪ شده است. در نهایت از یک مدل پیش آموزش دیده شده شبکه VGG-Face با Fine-tuning همراه با شبکه طبقه بندی تصاویر آزمایشی استفاده شده است که به دقت ۹۲٪ رسیده است.

مقاله [۲۷] از شبکه های مولد تخصصی برای تولید داده ها استفاده کرده است که به اختصار GAN نامیده می شوند. شبکه مستقل تولید کننده و تمیز دهنده استفاده تشكیل شده است. شبکه تولید کننده از روی بردار Z که می تواند یک نویز تصادفی باشد، یک تصویر تولید می کند 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 می شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می شود. به عنوان مثال بردار Z برای تصویر خانمی که عینک آفتابی نزدی است از بردار Z برای تصویر خانمی که عینک آفتابی زده است، کم می شود و حاصل آن، بردار مربوط به یک عینک آفتابی می باشد. سپس این بردار با بردار تصویر آقایی

¹Graph Clustering

که عینک نزد 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 ۱۸.۳ قابل مشاهده می‌باشد.



شکل ۱۸.۳: تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۲۷].

مقاله [۲۸] یک شبکه عمیق مبتنی بر GAN را با نام LR-GAN پیشنهاد می‌دهد، 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)، میانگین مربعات خطأ برای هر قسمت (PMSE)، واگرایی جنسن شanon اصلاح شده (JSD) و تنوع متقابل در اطلاعات (MVI). شبکه تمیز دهنده در GAN، بر اساس اطلاعات طبقه‌بندی 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. این مدل که به دقت بالای ۹۰٪ رسیده است، رتبه اول را در ۴ مجموعه داده شرایط بدون محدودیت کسب کرده است. نمونه‌ای از خروجی این روش در شکل ۱۹.۳ قابل مشاهده می‌باشد.

شبکه‌های GAN یاد می‌گیرند تصاویر جدیدی تولید کنند که شبیه به تصاویر واقعی باشند. اما این شبکه‌ها معمولاً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند. مقاله [۲۹] یک شبکه GAN جدید پیشنهاد می‌دهد که بخش تولید کننده آن به طور خودکار یاد می‌گیرد بدون هیچ ناظر انسانی ویژگی‌های بصری متفاوت تصاویر را از یکدیگر جدا نماید. پس از اتمام مرحله یادگیری، ما



شکل ۱۹.۳: تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین در ۴ مجموعه داده مختلف. موارد با حاشیه قرمز خروجی اشتباه هستند [۲۸].

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، مجموعه‌ای 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 سبک تر مانند رنگ. ما می‌توانیم این ویژگی‌های بصری را با ضرایب دلخواه خود ترکیب نماییم. نمونه‌ای از خروجی این روش در شکل ۲۰.۳ قابل مشاهده می‌باشد.

در مقاله [۳۰] به موضوع تولید چهره در زوایای دلخواه پرداخته شده است. در این مقاله از دو شبکه GAN استفاده شده است که در شبکه اول از روی چهره زاویه‌دار، چهره روبه‌رو تولید شده است. سپس با استفاده از شبکه GAN دوم از روی تصویر چهره روبه‌رو، تصویر با زاویه دلخواه با استفاده از یک پارامتر کنترلی تولید می‌شود. چالشی که در این مقاله به آن اشاره شده است، 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 ۲۰.۳ قابل مشاهده می‌باشد.



شکل ۲۰.۳: تصاویر هر ردیف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند [۲۹].

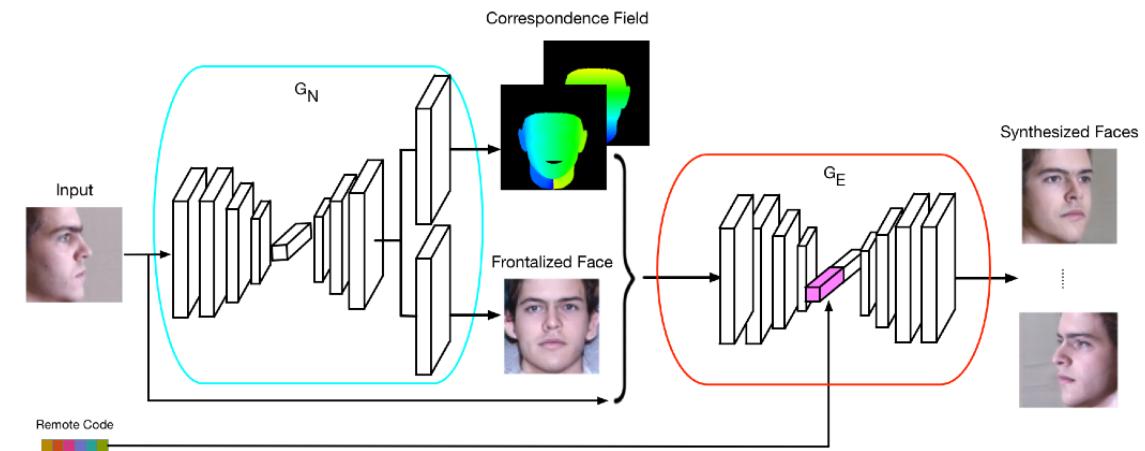
۶.۳ چالش منابع محدود

در سال ۲۰۱۲ Tolga Soyata و همکاران در [۳۱] یک روش تشخیص چهره بی‌درنگ مبتنی بر بینایی ابری^۱ با استفاده از معماری MOCHA ارائه کردند (شکل ۲۰.۳). با فرآگیر 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 بهینه پردازش بین تلفن همراه و فضای ابری با توجه به زمان تأخیر ناهمگن، توانایی محاسبه را به طور قابل توجهی افزایش می‌دهند.

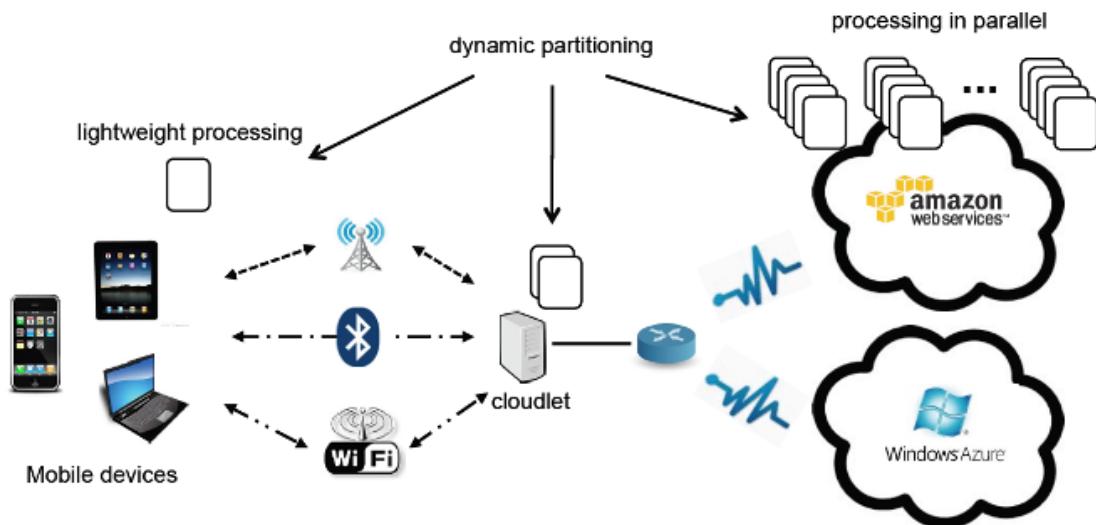
این سامانه از لحاظ ساختار به سه بخش تقسیم می‌شود:

دستگاه همراه: تلفن‌های همراه و iPad‌ها نقش تهیه و ارسال تصاویر را دارند. تصاویر با فرمت RAW فرستاده می‌شوند تا قابلیت پیش پردازش بهتری داشته باشند. اگر سرور ابر به دستگاه همراه نزدیک باشد و ارتباط با سرعت بالا امکان پذیر باشد، تصاویر پیش پردازش به سرور فرستاده می‌شوند. در غیر این صورت مرحله پیش پردازش در دستگاه همراه انجام می‌شود و فقط اطلاعاتی همچون ویژگی‌های Haar و طبقه بندها به سرور فرستاده می‌شوند. پس از اتمام فرایند تشخیص چهره، نتیجه نهایی برای تلفن همراه فرستاده

¹Cloud Vision



شکل ۲۱.۳: ساختار شبکه AD GAN - شبکه GN برای رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۳۰].



شکل ۲۲.۳: معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۳۱].

می شود.

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

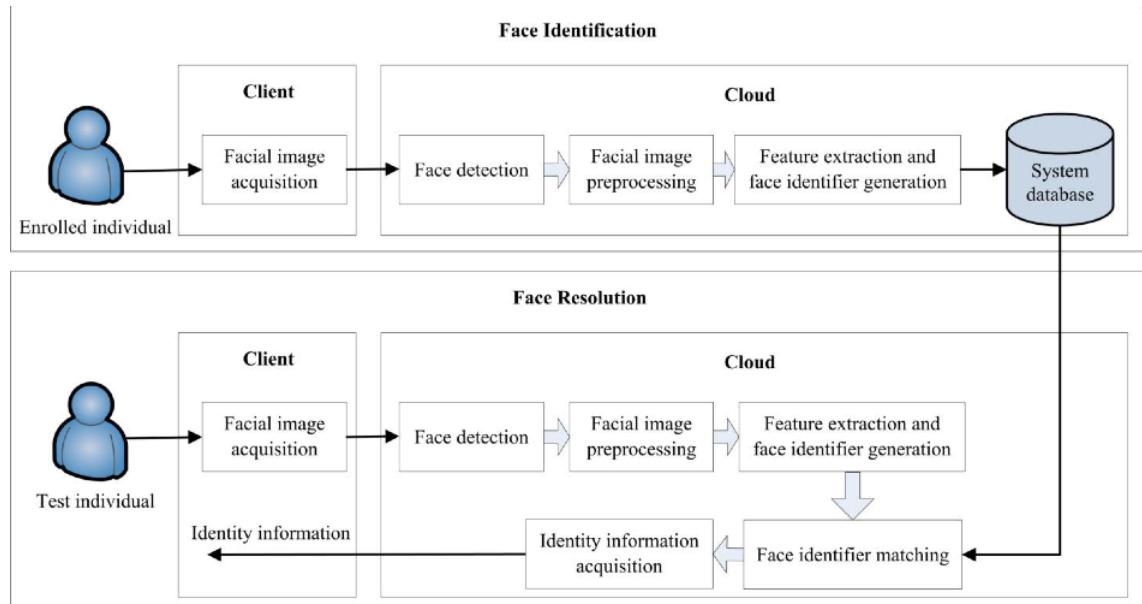
مطلوبی انجام دهند.

ابر: سرورهای ابر دارای توان پردازشی و پاسخگویی بسیار بالا می باشند که بار 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 در [۳۲] 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 عملا امکان-پذیر است و می تواند سرویس شناسایی موثر چهره را فراهم کند. همانطور که در شکل ۲۳.۳ 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 ۲۳.۳: نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۳۲].

در این سامانه تصویر با فرمت RAW برای ابر 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 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 ۲۴.۳ مشاهده می شود این سامانه ابری از بخش های سرور مدیریت (MS)^۳، سرور اطلاعات (IS)^۴، سرور شناسایی (RS)^۵ و پایگاه داده تشکیل شده است. به علت قدرت بالای پردازش در سرور ابری، امكان پردازش موازی نیز در این سامانه وجود دارد که باعث افزایش سرعت محاسبات و کاهش زمان پاسخ دهی سامانه می گردد. 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

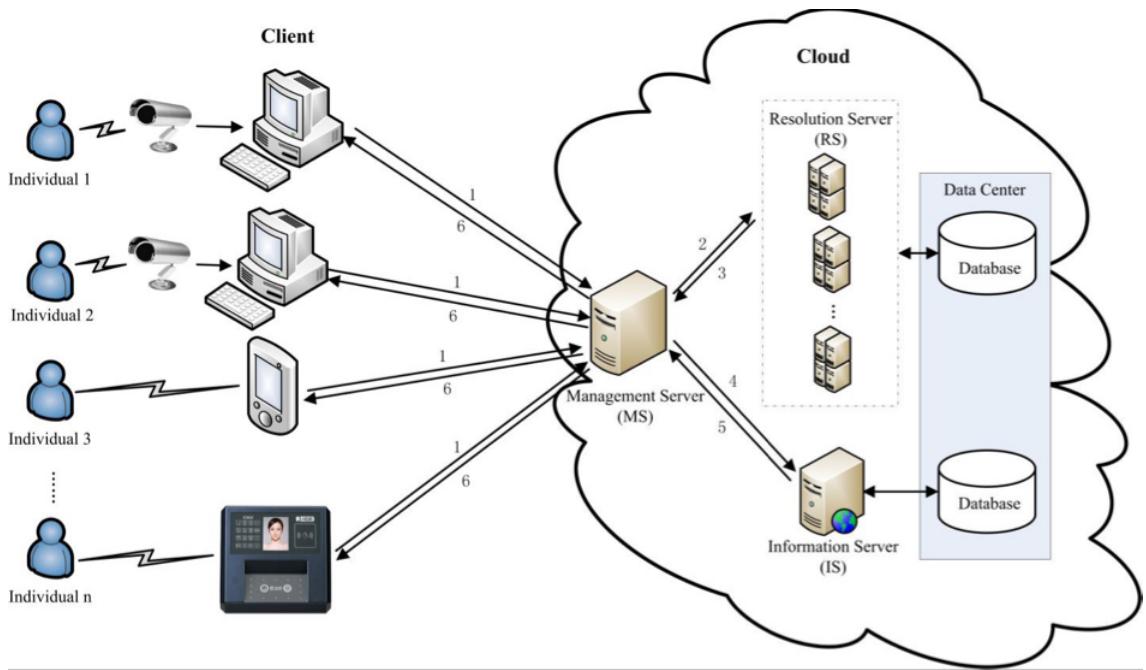
¹Big Data

²Local Binary Patterns

³Management Server

⁴Information Server

⁵Resolution Server



شکل ۲۴.۳: چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۳۲].

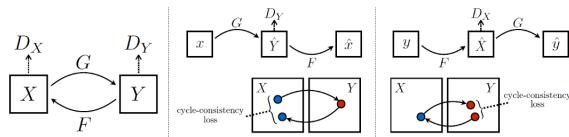
این داشته اند که یک راه حل نسبتاً همه جانبه در مورد مسئله تشخیص چهره و مشکلات آن ارائه دهنند. یکی از این رویکردها، استفاده از تابع ضرر CosFace می‌باشد که در سال ۲۰۱۸ Wang و همکاران در [۳۳] ارائه دادند. این تابع ضرر کسینوسی با حاشیه زیاد^۱ شباهت بسیار زیادی به تابع ضرر softmax دارد. با این تفاوت که به 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 $W_j^T x_i = ||W_j|| ||x_i|| \cos(j)$ تبدیل می‌کند، که 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 ۲۵.۳ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می‌کند اما در مرزهای تصمیم گیری ابهام قابل توجهی به وجود می‌آید، در حالی که تابع ضرر معرفی شده می‌تواند فاصله بیشتری را بین دسته‌های نزدیک اعمال کند.

در سال ۲۰۱۹ Deng و همکاران در [۳۴] یک تابع ضرر پیشنهاد دادند که کمک می‌کند 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 [۳۳] آمده است. سعی در بهبود روش پیشین و افزایش دقت کرده است.

¹Large Margin Cosine Loss



شکل ۲۵.۳: تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مزدین دسته ها ایجاد می نماید [۳۳].

برای آموزش شبکه های عصبی عمیق پیوسته 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 یک تعبیه را یاد می گیرند، مانند triplet loss. بر اساس داده های آموزش در مقیاس بزرگ و معماری DCNN، هر دو روش می توانند عملکرد بسیار خوبی در تشخیص چهره داشته باشند. با این حال، هم رویکرد softmax و هم رویکرد triplet loss اشکالاتی دارد. برای softmax

۱. اندازه ماتریس تبدیل W به طور خطی با افزایش تعداد دسته ها (n) افزایش می یابد.

۲. ویژگیهای 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.

برای triplet loss

۱. برای مجموعه داده های مقیاس بزرگ، رشد شدید در تعداد ترکیب های تعداد تصاویر سه گانه وجود دارد که منجر به افزایش قابل توجه تعداد مراحل تکرار می شود.

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

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 نقطه ای مقادیر موجود در ویژگی های استخراج شده و آخرين لایه کاملاً متصل، برابر با ضرب کسینوسی آنها پس از نرمال سازی می باشد. از تابع مثلثاتی کسینوسی برای محاسبه زاویه بین ویژگی فعلی و وزن هدف استفاده شده است. سپس یک حاشیه زاویه ای به زاویه هدف اضافه شده، در انتهای دوباره مقادیر به فضای خطی برگردانده شده است. مراحل بعدی دقیقاً مانند softmax هستند. مزایای این روش پیشنهادی را می توان به شرح زیر خلاصه کرد:

- در مجموعه داده های تصویر و فیلم در مقیاس بزرگ ، به عملکرد مناسبی دست می یابد.

● فقط به چندین خط کد نیاز دارد و اجرای آن در چارچوب های یادگیری عمیق مبتنی بر Tensorflow و Pytorch آسان است.

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

● هنگام آموزش فقط پیچیدگی محاسباتی ناچیز را اضافه می کند. پردازنه های گرافیکی کنونی می توانند به راحتی از هزاران

دسته مختلف برای آموزش پشتیبانی کنند و مدل به راحتی می تواند هویت های بیشتری را پشتیبانی کند.

رابطه ریاضی softmax معروف ترین تابع ضرر طبقه بندی که به طور گسترده استفاده می شود، 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. تعداد ابعاد ویژگی استخراج شده را n در نظر گرفتیم. W_j ستون

زم از وزن W می باشد و b_j بایاس است. مقدار N اندازه دسته و n تعداد دسته ها است. این تابع مستقیماً ویژگی استخراج شده را

برای اعمال شباهت بالاتر برای نمونه های درون کلاس و فاصله بیشتر برای نمونه های بین کلاسی بهینه نمی کند، که منجر به ایجاد

مشکل در عملکرد آن برای تشخیص چهره عمیق تحت تغییرات ظاهری بزرگ درون کلاس می شود (به عنوان مثال تغییرات زاویه چهره

و تغییرات سنی).

ما رابطه فوق را مبنای محاسبات قرار دادیم و تغییرات جزئی به آن اضافه کردیم. برای سادگی مقدار بایاس را صفر در نظر گرفتیم. سپس

حاصل ضرب مقادیر موجود در ویژگی های استخراج شده و آخرین لایه کامل متصل را به صورت (j)

تبديل می کنیم، که j زاویه بین وزن W_j و ویژگی x_i است. نرمال سازی ویژگی ها و وزن ها باعث می شود که خروجی فقط به زاویه بین

ویژگی و وزن بستگی داشته باشد. به کمک نرمال سازی مقادیر وزن $||W_j||$ را برابر ۱ در نظر می گیریم. همچنین ویژگی استخراج

شده $||x_i||$ را نرمال کرده و نام آن را s در نظر می گیریم. بنابرین ویژگی های استخراج شده در یک ابر کره با شعاع s توزیع می شوند.

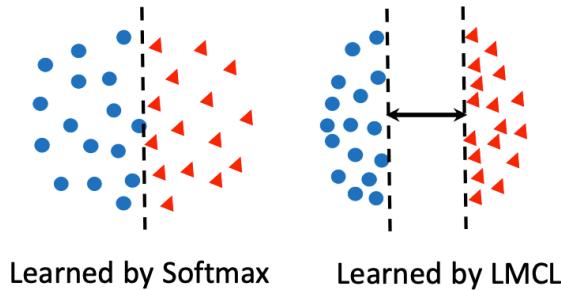
برای افزایش حاشیه بین x_i و W_j یک مقدار lm اضافه می کنیم تا به طور همزمان فشرده سازی درون کلاسی و اختلاف بین کلاسی

را افزایش دهیم.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s(\cos(\theta_j))}} \quad (15.3)$$

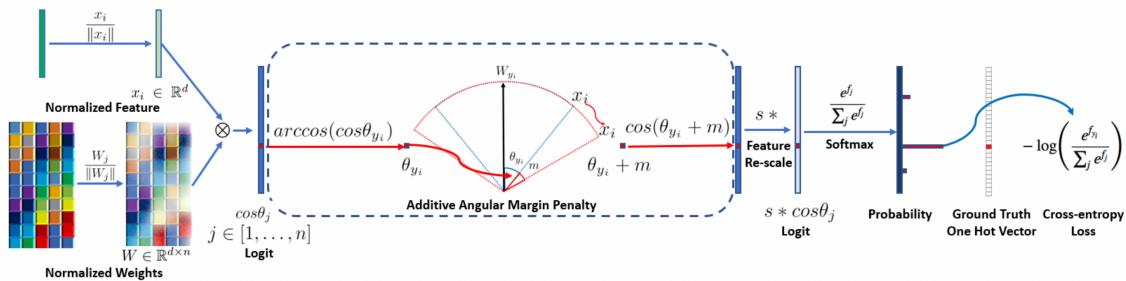
همانطور که در شکل ۲۶.۳ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می کند اما در مرزهای تصمیم

گیری ابهام قابل توجهی به وجود می آید، در حالی که تابع ضرر ما می تواند فاصله بیشتری را بین دسته های نزدیک اعمال کند.



شکل ۲۶.۳: رویکردهای مختلف هم ترازی چهره [۳۴].

این روش دقت ۵۳.۹۹ را بر روی مجموعه داده LFW با معماری ResNet50 بدست آورده است. خلاصه این روش در شکل ۲۷.۳ آمده است.



شکل ۲۷.۳: رویکردهای مختلف هم ترازی چهره [۳۴].

۷.۳ 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 (bluring)

- 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 صلب 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 [۲۱] 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 دقیق تر تشخیص چهره نمی‌نماید.

مقاله [۲۷] روشنی مبتنی بر GAN برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی موقعیت‌های تصویر چهره لطمہ وارد می‌نماید به طوری که شخص مورد نظر قابل شناسایی توسط سامانه یادگیری نمی‌باشد.

مقاله [۲۸] 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 [۳۰] چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره رو به رو بدون عارضه تولید شده است. این چالش به خاطر کمبود تصاویر با عارضه پوستی در مجموعه داده می‌باشد و شبکه در مواجه 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. مایک سوال مطرح می‌کنیم: آیا

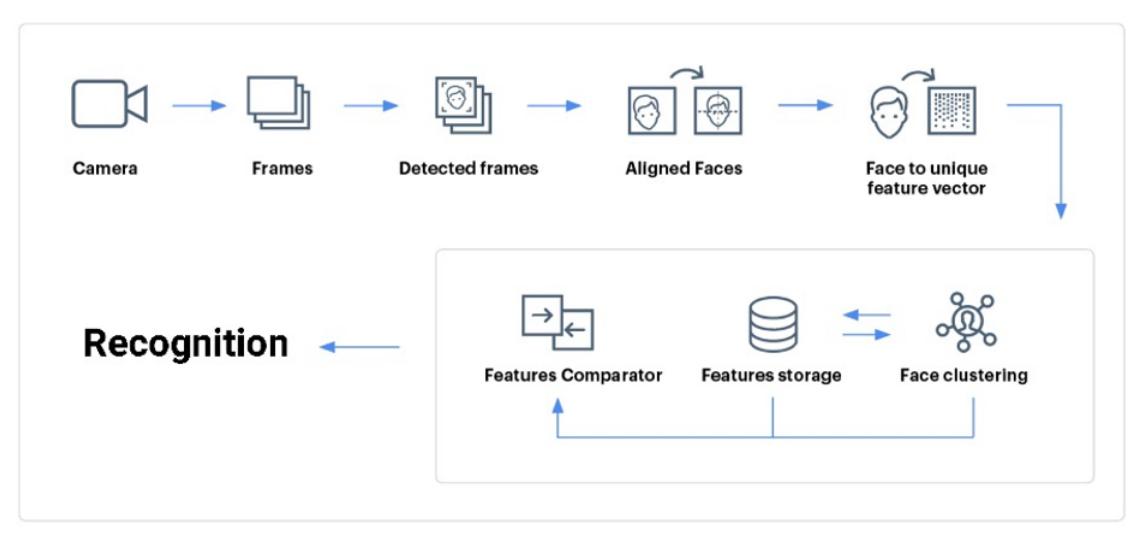
یادگیری عمیق واقعاً مسئله تشخیص چهره را حل می‌کند؟ چالش روش‌های یادگیری عمیق در تشخیص چهره چیست؟ در مقایسه با تشخیص شیء عمومی، تشخیص چهره به دلیل طیف گسترده‌ای از تغییرات در ظاهر چهره‌ها چالش برانگیز است. نورپردازی کنترل نشده، انسداد ناشی از عینک، مو، ریش، کلاه و...، تاری خارج از تمرکز دوربین، کیفیت پایین تصویر، بالا رفتن سن افراد و کمبود داده‌های آموزشی از مواردی می‌باشد که می‌توانند سامانه تشخیص چهره را با مشکل رو به رو 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 کرده اند که لایه‌های

عمیق CNN ها می توانند اطلاعات انتزاعی بیشتری مانند هویت، ظاهر و ویژگی ها را رمزگذاری کنند؛ با این حال هنوز هنوز کاملاً مطالعه نشده است که لایه ها دقیقاً با ویژگی های محلی برای تشخیص مطابقت دارند. برای شناسایی چهره، عملکرد یادگیری را می توان با یادگیری یک معیار اندازه گیری فاصله متمایز کننده بهبود داد. با این حال، با توجه به محدودیت های حافظه کارت گرافیک ها، نحوه انتخاب جفت ها یا سه گانه های اطلاعاتی و روش های آموزش آنلاین (به عنوان مثال، گرادیان نزولی) در مجموعه داده های بزرگ، هنوز یک مشکل باز است. یکی دیگر از مشکلات چالش برانگیز این است که پردازش ویدیو در شبکه های عمیق را برای استفاده از تجزیه و تحلیل چهره مبتنی بر ویدئو ترکیب کند.

فصل ۴

روش پیشنهادی

همان طور که در فصل گذشته شرح داده شد، 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 ۱.۴ خلاصه شده است که در ادامه هر یک تشریح خواهد شد.



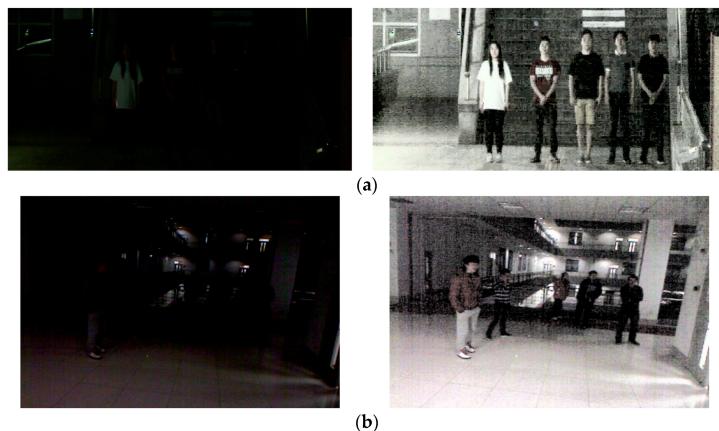
شکل ۱.۴: نمای کلی از روش پیشنهادی [۶].

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. یکی از روش‌های افزایش تباين تصویر، تکنیک یکنواخت سازی بافت‌نگار است. بطوریکه مقادیر پیکسل‌های تصویر را طوری تغییر می‌دهد تا کل بازه ممکن را تسخیر کند و ایده اساسی آن، نگاشت مقادیر شدت سطوح روشنایی از طریق یکتابع توزیع انتقال است. این عمل باعث افزایش تباين تصویر می‌شود که به معنای بهبود کیفیت تصویر و افزایش دقت پردازش‌های بعدی است. نمونه ای از این روش را از مقاله [۳۵] در شکل ۲.۴ مشاهده می‌نمایید.



شکل ۲.۴: نتیجه اعمال یکسان سازی بافت‌نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می‌باشند [۳۵].

۲.۲.۴ یافتن چهره

یکی دیگر از مراحل معمول پیش‌پردازش که در طی فرآیند شناسایی چهره انجام می‌شود، مکان یابی و یافتن چهره در تصاویر می‌باشد. برای این منظور از یک روش مبتنی بر شبکه عصبی پیچشی که توسط Deng و همکاران [۵] در سال ۲۰۱۹ معرفی شده است استفاده می‌کنیم. در این روش برای آموزش شبکه پیچشی از یکتابع ضرر مبتنی بر یادگیری چندکاره^۱ استفاده شده است که به صورت زیر می‌باشد.

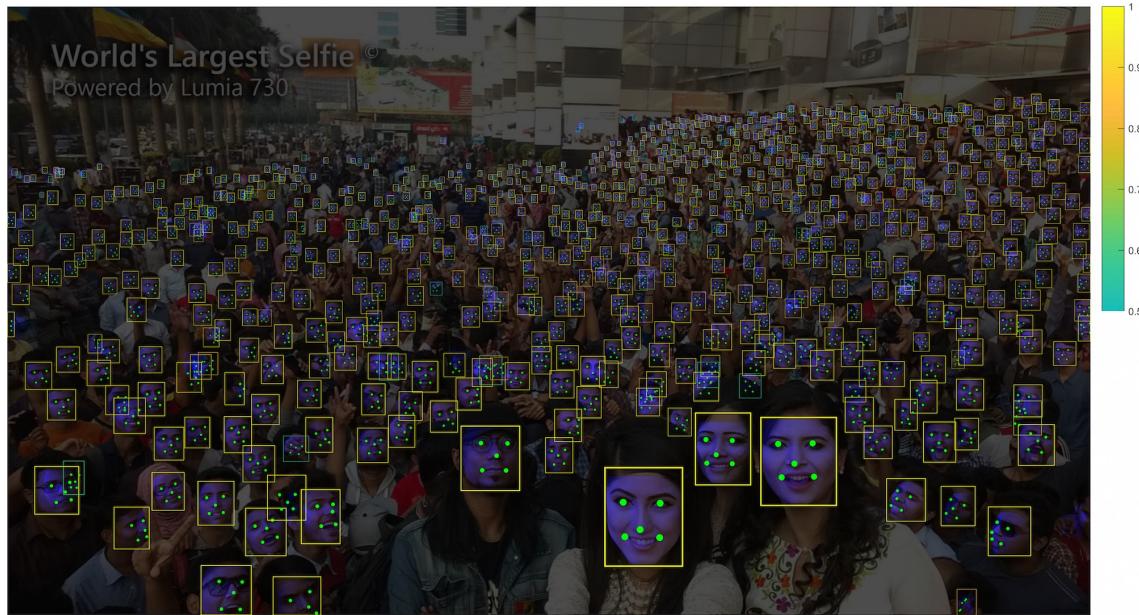
$$L = L_{cls} + L_{box} + L_{pts} + L_{pixel} \quad (1.4)$$

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

تابع ضرر مربوط به یافتن نقاط ویژه روی اجزای چهره می‌باشد، و L_{pixel} تابع ضرر مربوط به یافتن یک مدل سه بعدی مبتنی

¹Multi Task Learning

بر مش از روی چهره می‌باشد. استفاده از تابع ضرر مبتنی بر یادگیری چند کاره، کمک می‌نماید تا فضای مسئله محدود تر شود و الگوریتم بهینه سازی مورد نظر زودتر به سمت نقطه بهینه همگرا شود. ما برای رسیدن به خروجی بی درنگ، این روش را بر روی معماری MobileNet V3 پیاده سازی کرده و آموزش دادیم. نمونه ای از خروجی این روش را در شکل ۳.۴ مشاهده می‌کنید.



شکل ۳.۴: نمونه از خروجی الگوریتم یافتن چهره retina [۵].

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 نزدیک تر 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 رویه‌رو 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 ۴.۴ مشاهده می‌کنید. در نهایت تصاویر چهره با اندازه $112 * 112$ * پیکسل ذخیره می‌گردند تا در مرحله

¹Face Alignment

آموزش شبکه، مورد استفاده قرار گیرند.



شکل ۴.۴: به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد [۶].

۳.۴ دسته بندی

روش پیشنهادی برای تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت، یک الگوریتم مبتنی بر شبکه‌های پیچشی می‌باشد.

این فرایند را می‌توان به صورت یک مسئله بهینه سازی فرموله کرد که در ادامه به بخش‌های مختلف آن می‌پردازیم.

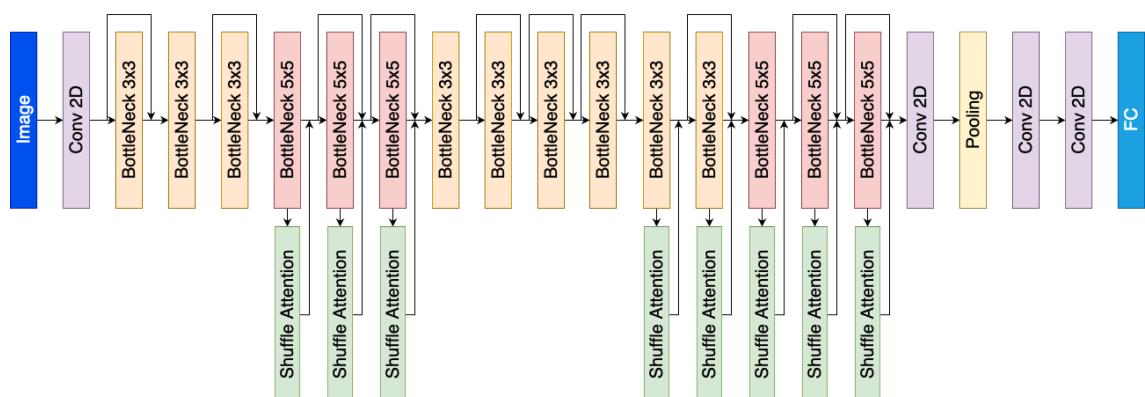
۱.۳.۴ مدل پیشنهادی پایه

در این بخش ابتدا باید معماری مناسب شبکه پایه برای مسئله را به دست آوریم. با بررسی شبکه‌های متداول و مقایسه دقت و زمان پاسخگویی آن‌ها به کمک یادگیری انتقال، به این نتیجه می‌رسیم که شبکه MobileNetV3 دارای چگالی دقت بالاتری در مقایسه با شبکه‌های دیگر می‌باشد و نسبت دقت دسته بندی به تعداد پارامترهای شبکه در آن بیشتر می‌باشد. بنابرین می‌توان سرعت اجرای مناسب و همچنین دقت 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 ؟؟ آمده است. آزمایش‌هایی بر روی معماری‌های مطرح دیگر نیز انجام شد که به علت ضعیف بودن نتایج یا بالا بودن زمان پاسخ دهی در جدول درج نشده‌اند. تشخیص چهره به دلیل وجود پیکسل‌های مشابه از نظر شدت روشناهی در تصویر سیار چالش برانگیز بوده است. از آنجا که عملیات کانولوشن به پنجره محلی از شدت روشناهی 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

جدول ۱.۴: مقایسه و ارزیابی برخی از معماری شبکه های رایج در زمینه بینایی ماشین

نام شبکه	تعداد پارامترها	دقت بر روی مجموعه داده LFW
MobileNetV2	۵۳M.۳	۵۰.۹۳
MobileNetV3	۵M.۲	۸.۹۵
SqueezeNet	۲۵M.۱	۲.۸۹
NASNetMobile	۳۲M.۵	۶۰.۹۰
EfficientNetB0	۳M.۵	۵۰.۸۵

می کنیم. لایه توجه استفاده شده در این پژوهش، شامل لایه توجه وابسته به کanal و لایه توجه وابسته به موقعیت می باشد. معماری شبکه پیشنهادی در شکل ۵.۴ آمده است.



شکل ۵.۴: مدل پایه مبتنی بر لایه های کانولوشن و لایه توجه.

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 به بخش استخراج ویژگی داده می شود و مدل در این مسیر به طور خودکار یک سلسله مراتب ویژگی را از تصاویر ورودی آموزش خواهد دید و در نهایت این ویژگی های استخراج شده به عنوان ورودی دسته بند مورد استفاده قرار می گیرد.

¹Bottleneck

²Attention

۱.۱.۳.۴ لایه کانولوشن

۴ لایه کانولوشن دو بعدی همراه با گام یک یا دو استفاده شده است. در اولین لایه کانولوشن اندازه پنجره فیلترها 3×3 و در لایه های کانولوشن بعدی اندازه پنجره فیلترها 1×1 می باشد. دلیل انتخاب سایز کوچک پنجره فیلترها کاهش پیچیدگی محاسباتی و همچنین عملکرد خوب آن ها در استخراج ویژگی می باشد. چگونگی عملکرد یک لایه کانولوشن از رابطه 2.4 بدست می آید.

$$x_j^{(l)} = \sum_{i=0}^c w_{c_{ij}}^{(l)} * x_i^{(l-1)} + b_j^{(l)} \quad (2.4)$$

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

۲.۱.۳.۴ لایه تنگنا و واحد باقیمانده

به عنوان مثال به جای پردازش یک نگاشت ویژگی عظیم با 256×256 عمق، ابتدا همه این اطلاعات را در نگاشتهای ویژگی 64×64 بعدی فشرده می کنیم. زمانی که این فشردگی انجام شد، از کانولوشن 3×3 استفاده می کنیم که وقتی روی 64×64 نگاشت ویژگی به جای نگاشت اعمال شود، بسیار سریعتر است و چنین پردازشی می تواند همان نتایج یا نتایج بهتری نسبت به پشتلهای معمولی 3×3 ارائه کند. در نهایت با استفاده از 1×1 مجدداً به نگاشت اصلی 256×256 خود بازمی گردیم.

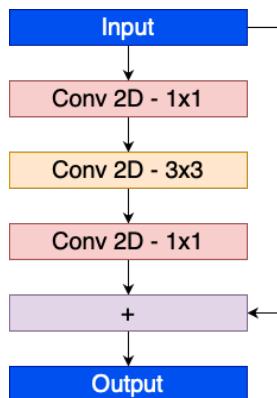
از سوی دیگر شبکه های عمیق با واحدهای باقیمانده^۱ بر روی پایگاه داده های مختلف مانند COCO، ImageNet دقت و همگرایی خوبی را از خود نشان داده اند. با استفاده از مسیرهای پرش^۲، واحدهای باقیمانده می توانند به سیگنال ها اجازه دهند که مستقیماً از یک بلوک به بلوک های دیگر منتقل شوند. به طور کلی، واحدهای باقیمانده را می توان به صورت رابطه 3.4 بیان کرد.

$$x_{l+1} = x_l + R(x_l, w_l) \quad (3.4)$$

در اینجا R نشان دهندهتابع واحد باقیمانده است، x_l ویژگی ورودی به واحد باقیمانده l ام و W_l مجموعه ای از پارامترهای مربوط به واحد باقیمانده l ام می باشد. ایده اصلی شبکه های باقیمانده، عمیق تر کردن یک شبکه به منظور افزایش دقت شبکه مورد نظر می باشد. بنابراین با این عملیات در واقع به شبکه اجازه داده می شود که در صورت نیاز، ویژگی های لایه قبل بدون تغییر و به صورت مستقیم به لایه بعد منتقل شود. در شکل ۶.۴ لایه تنگنا با واحد باقیمانده طراحی شده در این مدل به نمایش گذاشته شده است.

¹Deep Residual Network

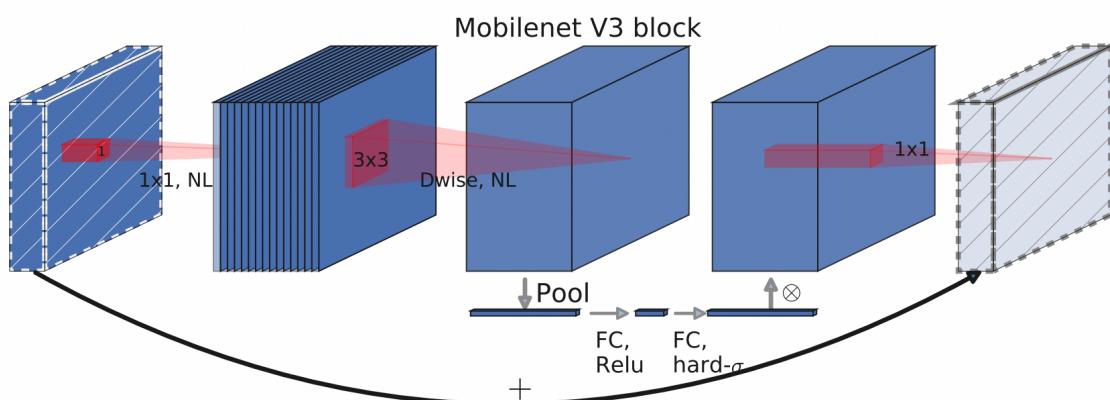
²Shortcut Pathway



شکل ۶.۴: یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می کند..

۳.۱.۳.۴ واحد توجه

لایه توجه استفاده شده در این پژوهش یک واحد توجه SA^۱ است که در مقاله [۱۷] معرفی شده است. این لایه ابتدا ابعاد کanal را به چندین ویژگی فرعی قبل از پردازش موازی آنها تقسیم می کند. سپس، برای هر زیر ویژگی از یک واحد Shuffle برای به تصویر کشیدن وابستگی های ویژگی در هر دو بعد مکانی و کanal استفاده می کند. پس از آن، همه زیر ویژگی ها جمع می شوند و یک عملگر تغییر کanal برای امکان برقراری ارتباط اطلاعاتی بین ویژگی های فرعی مختلف به کار گرفته می شود. معماری این مازول در شکل ۷.۴ آمده است. این لایه شامل دو بخش اصلی وابسته به کanal^۲ و وابسته به موقعیت^۳ می باشد.



. [۱۷] معماری مازول

یک نقشه ویژگی^۴ 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

¹Shuffle Attention

²Channel Attention Module

³Spacial Attention Module

⁴feature map

ورودی واحد توجه در نظر گرفته می‌شود. ابتدا X به G گروه در طول کانال تقسیم می‌شود. رابطه ۴.۴ این موضوع را نشان می‌دهد. سپس هر X_k در طول کانال به دو نیم تقسیم شده و X_{k1} و X_{k2} را تشکیل می‌دهند. یکی از این دو زیر نقشه ویژگی^۱ برای توجه وابسته به کانال و دیگری برای توجه وابسته به موقعیت مورد استفاده قرار می‌گیرد. این عملیات توجه به مفهوم چه چیزی؟ و کجا؟ معنا می‌بخشد.

$$X = [X_1, \dots, X_G], X_k \in \mathbb{R}^{C/G \times H \times W} \quad (4.4)$$

برای اتصال صحیح مازول واحد توجه به معماری MobileNetV3 مقدار G را برابر ۴ در نظر گرفتیم. توجه وابسته به کانال با بهره‌گیری از وابستگی‌های بین کانال‌ها، می‌توان بر ویژگی‌های وابسته، تأکید کرده و استخراج ویژگی‌های معانی خاص را بهبود بخشید. بنابراین در این پژوهش یک مازول توجه وابسته به کانال ایجاد شده تا به طور واضح وابستگی بین کانال‌ها را مدل کند. ابتدا نقشه ویژگی X_{k1} که از مرحله قبل بدست آمده است، توسط یک لایه پولینگ میانگین گیر سراسری^۲ به یک نقشه ویژگی s با ابعاد $1 \times 1 \times C/2G$ تبدیل می‌شود. رابطه ۵.۵ نشان دهنده چگونگی ایجاد بردار نقشه توجه وابسته به کانال می‌باشد.

$$s = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j) \quad (5.4)$$

با انجام این عملیات، اطلاعات مکانی کلی برای هر کانال به صورت جداگانه محاسبه می‌شود و در s قرار می‌گیرد. سپس خروجی لایه توجه وابسته به کانال مطابق رابطه ۶.۴ بدست می‌آید.

$$X'_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \quad (6.4)$$

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 کانال 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 ضرری می‌باشد، که این ویژگی‌ها می‌توانند در بازه بزرگی از تنوع قرار بگیرند. به منظور مدل سازی روابط مبتنی بر روی ویژگی‌های محلی، مازول توجه وابسته به موقعیت استفاده می‌شود.

¹sub feature map

²Global Averaging Pooling

ابتدا ورودی X_{k2} توسط یک عملگر GN^۱ نرمال سازی می‌شود و خروجی این مرحله طبق رابطه ۷.۴ بدست می‌آید.

$$X'_{k2} = \sigma(W_2GN(X_{k2}) + b_2).X_{k2} \quad (7.4)$$

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'_{k1} و X'_{k2}

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 معماری

های مدرن را آسان می‌کند.^[۱۷].

۲.۳.۴ تابع ضرر

یکی از چالش‌های اصلی در یادگیری ویژگی‌ها با استفاده از شبکه‌های عصبی عمیق پیوسته (DCNN) 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^[۳۴] 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

کمک می‌نماید. رابطه ریاضی softmax معروف ترین تابع ضرر طبقه بندی که به طور گسترده استفاده می‌شود، به صورت رابطه ۸.۴

است.

¹Group Norm

²channel shuffle

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (8.4)$$

که در آن \tilde{x} نشان دهنده ویژگی عمیق نمونه i از دسته y است. تعداد ابعاد ویژگی استخراج شده را n در نظر گرفتیم. W_j ستون j از وزن W می‌باشد و b_j بایاس است. مقدار N اندازه دسته و n تعداد دسته‌ها است. از آنجا کهتابع $softmax$ ذات زاویه‌ای و قطبی دارد، بنابرین می‌توان آن را به صورت رابطه ۹.۴ بازنویسی کرد:

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\|W_{y_i}\| \|x_i\| \cos(\theta_{y_i})}}{\sum_{j=1}^n e^{\|W_j\| \|x_i\| \cos(\theta_j)}} \quad (9.4)$$

در این تابع ابتدا بردار ویژگی x که خروجی آخرین لایه شبکه می‌باشد، نرمال می‌شود. همچنین وزن‌های W مربوط به آخرین لایه شبکه نیز نرمال‌سازی می‌شوند. بنابراین می‌توان برای سادگی اندازه $\|x_i\|$ را برابر مقدار ثابت s و اندازه $\|W_j\|$ را برابر صفر در نظر گرفت. بنابرین ویژگی‌های استخراج شده در یک ابر کره با شعاع s توزیع می‌شوند. همچنین مقدار b_j را برابر صفر در نظر می‌گیریم. تابع نهایی به صورت رابطه ۱۰.۴ نوشته می‌شود.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i}) + m)}}{\sum_{j=1}^n e^{s(\cos(\theta_j) + m)}} \quad (10.4)$$

حاصل ضرب وزن‌ها در ویژگی‌های استخراج شده محاسبه می‌گردد، که برابر $\cos(\theta_j) arccos$ می‌شود. سپس آن محاسبه شده که مقدار b_j را به ما می‌دهد. سپس برای افزایش حاشیه بین \tilde{x} و W_j یک مقدار حاشیه زاویه‌ای m به زاویه هدف اضافه می‌کنیم تا به طور همزمان فشرده سازی درون کلاسی و اختلاف بین کلاسی را افزایش دهیم، در انتها دوباره $\cos(\theta_j)$ محاسبه شده و در ثابت s ضرب می‌شود. مراحل بعدی دقیقاً مانند $softmax$ هستند. در این روش بنا به پیشنهاد مقاله [۳۴] مقدار $m = 0.5$ و مقدار $s = 64$ در نظر گرفته شده است. مزایای این روش پیشنهادی را می‌توان به شرح زیر خلاصه کرد:

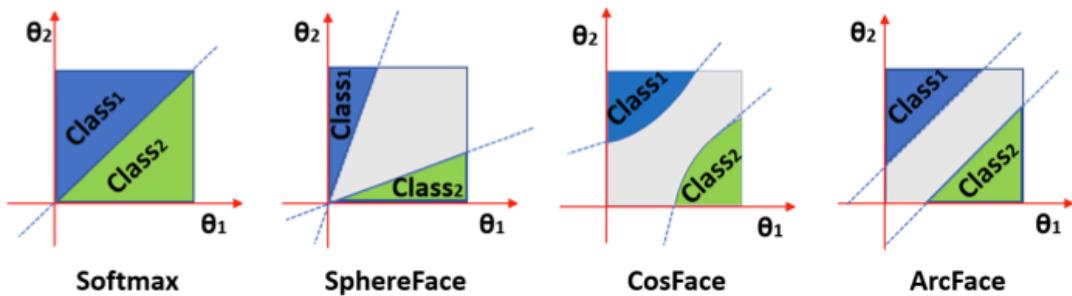
- در مجموعه داده‌های تصویر و فیلم در مقیاس بزرگ، به عملکرد مناسبی دست می‌یابد.

- فقط به چندین خط کد نیاز دارد و اجرای آن در چارچوب‌های یادگیری عمیق مبتنی بر Tensorflow و Pytorch 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 می کند. پردازنده های گرافیکی کنونی می توانند به راحتی از هزاران

دسته مختلف برای آموزش پشتیبانی کنند و مدل به راحتی می تواند هویت های بیشتری را پشتیبانی کند.

همانطور که در شکل ۸.۴ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می کند اما در مراتب های تصمیم گیری ابهام قابل توجهی به وجود می آید، در حالی که تابع ضرر ArcFace می تواند فاصله بیشتری را بین دسته های نزدیک اعمال کند.



شکل ۸.۴: تابع ضرر ArcFace در مقایسه با توابع ضرر مشهور دیگر [۳۴].

۳.۳.۴ آموزش مدل و استخراج ویژگی

استفاده از این روش کمک می کند تا تعداد دسته ها پس از آموزش قابل تغییر باشد و همچنین برای learning shot one بسیار مناسب است. همانطور که پیشتر بیان شد به آموزش این شبکه می پردازیم. مجموعه داده خود را پس از پیش پردازش و افزایش داده ها، آماده می کنیم. داده های آموزش متفاوت را بارگزاری کردیم و آموزش را در ۳۰ دوره انجام داده ایم. در ابتدای روند آموزش، لازم است پارامترهای مدل مقدار دهی اولیه شوند و انتخاب پارامترهای اولیه مبتنی بر تأثیر زیادی در مدل آموزش یافته داشته باشد. در این پژوهش به منظور مقدار دهی اولیه پارامترها از تابع توزیع یکنواخت استفاده شده است.

از دیگر چالش های اساسی برای روش های بهینه سازی مبتنی بر گرادیان، انتخاب میزان نرخ یادگیری مناسب است. روش های کلاسیک گرادیان تصادفی از نرخ یادگیری ثابت یا کاهشی استفاده می کنند، که برای همه پارامترهای مدل یکسان است. با این حال، مشتقات جزئی پارامترهای لایه های مختلف می توانند از نظر مقدار متفاوت باشند، که می تواند به نرخ یادگیری مختلفی نیاز داشته باشد. با این حال، مشتقات جزئی پارامترهای لایه های مختلف می توانند از نظر مقدار تفاوت قابل توجهی داشته باشند، که می تواند به نرخ یادگیری مختلفی نیاز داشته باشد. در سال های اخیر، تمایل به توسعه روش هایی برای انتخاب خودکار نرخ یادگیری مستقل

افزایش یافته است. اکثر روش‌ها به عنوان مثال، AdaGrad، 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 ۵۱۲ تایی بدست آمده را با بردارهای مربوط به چهره‌های بانک اطلاعاتی مقایسه کرده و با محاسبه فاصله اقلیدسی بردارها، نزدیک ترین شخص مورد نظر انتخاب شده و در صورتی که فاصله میان بردار ویژگی آن‌ها از حد آستانه کمتر باشد، عمل دسته‌بندی 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.^[۳۵] برای آموزش شبکه عصبی مربوط به دسته‌بندی 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 سرو کار 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.

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

که در آن مثبت‌های صحیح (TP) تعداد نمونه‌هایی که به درستی تشخیص داده شده‌اند. و N تعداد کل نمونه‌ها 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، سبک تر 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 به میلیون می‌باشد.

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 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 و بدون نویز و اشتباه است. این مجموعه داده در مسائل تایید چهره و تشخیص چهره کاربرد دارد. [۵۶]

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. [۵۷]

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. [۵۸]

¹Microsoft 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. [۵۹]

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. [۶۰]

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 شده اند. به طور میانگین ۱۵.۲ فیلم برای هر شخص در دسترس است. کوتاه ترین مدت ویدیو ۴۸ فریم، طولانی ترین ویدیو ۶۰۰ فریم و متوسط طول یک ویدیو ۱۸۱ فریم است. [۵۸]

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. [۵۹]

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. [۶۱]

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 سن های بالا می باشد [۶۲]. اطلاعات تکمیلی درباره مجموعه داده های آزمایش را در جدول ۹ مشاهده می نمایید.

جدول ۱.۵: مجموعه داده های ارزیابی رایج در زمینه بازنگشتن چهره

نام	تعداد تصاویر چهره	تعداد افراد	نوع کاربرد
LFW	۱۳۲۳۳	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

۴.۵ پیکربندی الگوریتم

به منظور آموزش شبکه، در هر دوره ۲۰٪ تعداد داده های آموزشی را به عنوان داده های ارزیابی در نظر می گیریم تا روند آموزش شبکه را بر اساس عامل های دیگر مورد بررسی قرار دهیم. همچنین در هر تکرار، اندازه دسته هایی که به شبکه برای آموزش داده می شود را

¹Cross-Age Celebrity Dataset

²Million-Scale Face Recognition Dataset

برابر ۶۴ قرار دادیم. توصیه می‌شود این مقدار توانی از ۲ باشد که مقدار ۶۴ 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 100.4 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 پنج مدل شبکه برای آموزش خواهیم داشت و بهترین نتیجه را برای آزمون بر روی مجموعه داده‌های آزمون انتخاب می‌کنیم.

۵.۵ نتایج آزمون

با توجه به پیکربندی بیان شده و همچنین عامل‌هایی که در فصل سوم توضیح داده شد، مدل را آموزش داده‌ایم و به سراغ مجموعه داده‌های آزمون می‌رویم. مقدار دقت را طبق رابطه 1.5 برای معماری‌های مرتبط و مشهور و همچنین روش پیشنهادی محاسبه کردیم. در جدول ۲.۵ نتایج حاصل از محاسبه دقت را در معماری‌های مختلف بر روی مجموعه داده داده‌های مشاهده می‌کنید.

جدول ۲.۵: دقت معماری‌های مختلف بر روی مجموعه داده‌های ارزیابی رایج در زمینه بازشناسی چهره

ImageNet	Cifar100	Mnist	MegaFace	AgeDB-30	LFW	Custom Dataset	روش
۲.۷۵	۰۱.۸۰	۷.۹۹	۵۹.۹۰	۰۵.۹۳	۱۰.۹۸	۵۰.۹۲	ArcFace
۸.۷۹	۴۷.۸۲	۹.۹۹	۵۰.۹۳	۰۲.۹۶	۶۵.۹۹	۸۰.۹۹	MobileNetV۳
							SA-MobileNetV۳

در جدول ۳.۵ نتایج حاصل از محاسبه سرعت را در معماری‌های مختلف بر روی مجموعه داده داده مشاهده می‌کنید.

جدول ۳.۵: سرعت معماری‌های مختلف بر روی مجموعه داده‌های ارزیابی رایج در زمینه بازشناسی چهره

Madds	Accuracy density	Number of Parameters	Accuracy (imageNet)	روش
M ۸۲۲۰	۳۲.۳	M ۲۵	۰.۸۳	ResNet۵۰
M ۶۹.۶۲۷	۷۳.۲۰	M ۵.۳	۵۶.۷۲	MobileNetV۲
M ۶۹.۴۴۸	۹۲.۱۳	M ۴.۵	۲.۷۵	MobileNetV۳
M ۶۸.۴۴۵	۰۶.۲۱	M ۸.۳	۸.۷۹	SA-MobileNetV۳

با توجه به نتایجی که در جدول‌ها و شکل‌ها مشاهده می‌کنید متوجه می‌شویم که هر چند حساسیت روش پیشنهادی ما در تعدادی از

مقادیر مثبت کاذب نسبت به موارد مشابه در دیگر مقاله‌ها کمتر شده است اما در کل معیار Fscore آن نسبت به بقیه بهتر شده است و این نشان می‌دهد که ترکیب ویژگی‌های ژرف (استخراج شده توسط انسان) و ویژگی‌های معنادار (استخراج شده توسط انسان) باعث می‌شود که سامانه مزیت هر دو دسته ویژگی را با هم داشته باشد و در به دست آوردن نتایج بهتر کمک کننده 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 شبکه‌هایی ارائه شده، سرعت کم پردازش و همچنانی تعداد زیاد تصاویر ورودی برای آموزش می‌باشند. ما با توجه به شرایط بی‌درنگ و محدودیت‌هایی 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 بهبود در نتایج باید در تغییر لایه‌های معماری شبکه دقت بیشتری داشت. بنابراین برای بهبود 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 تر کرد و به این ترتیب بردار ویژگی‌هایی به دست خواهد آمد که از فاصله درون دسته‌ای کمتر و فاصله برون دسته‌ای بیشتری برخوردار باشند. 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 می‌توان 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 روش‌نایی 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 دسته های زیاد در پایگاه داده، می‌توان از تابع ضری که مبتنی بر تعداد زیاد دسته‌ها باشد استفاده کرد که منجر به آموزش بهتر مدل و همچنین کم شدن اشتباہات 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

منابع و مأخذ

- [1] E. Hjelmås and B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding*, vol.83, no.3, pp.236–274, 2001.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” 2019.
- [6] 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.
- [7] 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*.
- [8] JafriRabia and A. R., “A survey of face recognition techniques,” *Journal of Information Processing Systems*, vol.5, no.2, pp.41–68, 2009.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.

- [13] 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.
- [14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [15] 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.
- [16] 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.
- [17] Q.-L. Z. Y.-B. Yang, “Sa-net: Shuffle attention for deep convolutional neural networks,” 2021.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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 (ICIWC)*, pp.239–243, 2017.
- [27] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2016.

- [28] S. Banerjee and S. Das, “Lr-gan for degraded face recognition,” *Pattern Recognition Letters*, vol.116, pp.246–253, 2018.
- [29] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” 2019.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” 2019.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] E. Osuna, R. Freund, and F. Girosit, “Training support vector machines: an application to face detection,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.130–136, 1997.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, “Facial landmark detection with tweaked convolutional neural networks,” 2016.

- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” 2014.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” *Pattern Recognition*, vol.66, pp.144–152, 2017.
- [53] 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.
- [54] 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.
- [55] L. Di Martino, J. Preciozzi, F. Lecumberry, and A. Fernández, “Face matching with an a contrario false detection control,” *Neurocomputing*, vol.173, 08 2015.
- [56] “Casia web-face dataset,” <https://paperswithcode.com/dataset/casia-webface>. Accessed: 2021-04-24.
- [57] “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.

- [58] “Vggface2 dataset,” https://www.robots.ox.ac.uk/~vgg/data/vgg_face/. Accessed: 2021-04-24.
- [59] “Labeled faces in the wild dataset,” <http://vis-www.cs.umass.edu/lfw/>. Accessed: 2021-04-24.
- [60] “Public figures face database,” <https://www.cs.columbia.edu/CAVE/databases/pubfig/>. Accessed: 2021-04-24.
- [61] “Cross-age celebrity dataset,” <https://paperswithcode.com/dataset/cacd>. Accessed: 2021-04-24.
- [62] “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