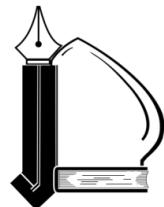


بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



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 پایاننامه دانشجوی دوره کارشناسی ارشد	 دانشگاه فرودی شهروند مدیریت تحصیلات تکمیلی
دانشجوی کارشناسی ارشد	جلسه دفاعیه پایاننامه تحصیلی آقای/خانم: سید سجاد اعمی رشته/گرایش: مهندسی کامپیوتر/هوش مصنوعی تحت عنوان: تشخیص بیدرنگ 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. پس از ارزیابی توسط هیأت داوران، پایاننامه با نمره به عدد و درجه	به حروف
		عنوان مورد تصویب قرار گرفت.
<u>امضاء</u>	<u>نام و نام خانوادگی</u>	<u>عنوان</u>
	دکتر حمید رضا پور رضا	استاد/استادان راهنما:
		استاد/استادان مشاور:
	دکتر	متخصص و صاحبنظر داخلی:
		متخصص و صاحبنظر خارجی:
	ناینده تحصیلات تکمیلی دانشگاه (ناظر)	
	نام و نام خانوادگی:	
	امضاء:	

چکیده

در سال های اخیر، به دلیل استفاده از یادگیری عمیق، تشخیص چهره شاهد پیشرفت های چشمگیری بوده است. با این حال، استراتژی داده محور همچنین یک چالش را به همراه می آورد: تصاویر ارسال شده به سیستم های تشخیص چهره همیشه برای تشخیص مناسب نیستند. به دلیل سناریوهای غیرقانونی و عدم تشخیص چهره، ممکن است چهره هایی با وضوح کم، چهره های تار در حرکت، صورت های مسدود و حتی وصله های پس زمینه وجود داشته باشد. متأسفانه، از آنجا که موتور تشخیص چهره قبل‌اچنین چهره های بی کیفیتی را نمیده است، احتمالاً تصمیمات نامعقولی در مورد این ها می گیرد 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 افزار یک عینک هوشمند برای افراد نایبنا می‌باشد. هنگامی که فرد نایبنا 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 نایبنا 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 پردازشگر تلفن همراه، باید محدودیت‌هایی مانند منابع را در نظر گرفت و الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد. در این پایان نامه محوریت اصلی مطالعات بر روی مرحله شناسایی چهره در شرایط بدون محدودیت است. علاوه بر موارد گفته شده در بالا، ما فرض کردیم که برای آموزش الگوریتم یادگیری عمیق، داده های محدودی از هر دسته در اختیار داریم. برای مقابله با این چالش با استفاده از شبکه های مولد تخصصی تصاویر بیشتری تولید می‌کنیم. برای این منظور ما از ۴ شبکه تولید کننده در کنار هم استفاده کرده ایم. با بررسی نتایج حاصل از این پژوهش بر روی تصاویر مجموعه داده LFW و YouTube Faces، دقت روش پیشنهادی ما به ترتیب برابر با ۹۶٪ و ۹۴٪ شد که دقت بالاتری نسبت به روش‌های مشابه می‌باشد.

فهرست مطالب

ج	فهرست جداول
ج	فهرست تصاویر
ذ	فهرست نمادها
۱	۱ مقدمه
۲	۱.۱ مقدمه
۲	۲.۱ ویژگی‌های بیومتریک
۲	۲.۱.۱ ارزیابی ویژگی‌های بیومتریک انسان
۳	۳.۱ سامانه تشخیص چهره
۴	۴.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره
۵	۵.۱ نمای کلی یک سامانه تشخیص چهره
۵	۵.۱.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر
۶	۶.۱ خوشه‌بندی
۷	۷.۱ طبقه‌بندی
۸	۸.۱ مسئله تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده
۹	۹.۱.۱ چالش‌های سامانه تشخیص چهره
۱۲	۱۲.۱ مروری بر کارهای گذشته
۱۳	۱۳.۱ مقدمه

۱۳	یافتن چهره	۲.۲
۱۳	رویکردهای مبتنی بر دانش	۱.۲.۲
۱۴	رویکردهای مبتنی بر تطبیق کلیشه یا الگو	۲.۲.۲
۱۴	رویکردهای مبتنی بر ویژگی	۳.۲.۲
۱۷	رویکرد مبتنی بر عامل‌های آماری	۴.۲.۲
۲۳	رویکردهای مبتنی بر تصویر	۵.۲.۲
۲۵	شناسایی چهره	۳.۲
۲۵	رویکردهای سنتی	۱.۳.۲
۲۶	رویکرد مبتنی بر فیلتر گابور	۲.۳.۲
۲۷	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
۲۹	تشخیص چهره مبتنی بر ویدیو	۶.۳.۲
۲۹	رویکرد مبتنی بر چهره ویژه	۷.۳.۲
۳۱	رویکرد مبتنی بر ماشین بردار پشتیبان	۸.۳.۲
۳۲	رویکردهای مبتنی بر شبکه عصبی	۹.۳.۲
۳۸	رویکردهای مبتنی بر نقاط راهنمایی	۱۰.۳.۲
۴۱	جمع‌بندی	۴.۲
۴۲	مروی بر کارهای گذشته در شرایط کنترل نشده	۳
۴۳	مقدمه	۱.۳
۴۳	چالش حالت	۲.۳
۵۳	چالش روشنایی	۳.۳
۵۴	چالش انسداد	۴.۳

۵۶	چالش کمبود تصاویر آموزشی	۵.۳
۶۰	چالش منابع محدود	۶.۳
۶۴	نتیجه‌گیری	۷.۳
۶۶	روش پیشنهادی	۴
۶۷	مقدمه	۱.۴
۶۷	پیش‌پردازش	۲.۴
۶۷	همسان‌سازی بافت‌نگار	۱۰.۲.۴
۶۸	یافتن چهره	۲۰.۲.۴
۶۹	تراز کردن تصویر	۳۰.۲.۴
۷۰	دسته‌بندی	۳۰.۴
۷۰	مدل پیشنهادی پایه	۱۰.۳.۴
۷۱	آموزش شبکه جهت استخراج ویژگی	۲۰.۳.۴
۷۲	تابع ضرر	۳۰.۳.۴
۷۵	فناوری‌های استفاده شده	۴۰.۴
۷۶	ارزیابی روش پیشنهادی	۵
۷۷	مقدمه	۱.۵
۷۷	معیار ارزیابی	۲.۵
۷۸	مجموعه داده	۳.۵
۷۸	مجموعه داده‌های آموزش	۱۰.۳.۵
۷۹	مجموعه داده‌های آزمون	۲۰.۳.۵
۸۰	پیکربندی الگوریتم	۴.۵
۸۱	نتایج آزمون	۵.۵
۸۳	نتیجه‌گیری و پیشنهادات	۶

۸۴	مقدمه	۱.۶
۸۴	بحث و نتیجه‌گیری	۲.۶
۸۵	پیشنهادات	۳.۶
۸۶	منابع و مأخذ	

فهرست جداول

۱.۲	مقایسه دقیقت الگوریتم NDA + SVM با سایر رویکردهای مشابه	۳۳
۲.۲	مقایسه شبکه عصبی موبایل نت با گوگل نت و VGG	۴۰
۱.۵	مجموعه‌داده‌های ارزیابی رایج در زمینه بازشناسی چهره	۸۰
۲.۵	دقیقت معماری‌های مختلف بر روی مجموعه‌داده‌های ارزیابی رایج در زمینه بازشناسی چهره	۸۱

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 [？]	۸
۵.۱	مقایسه تفاوت‌های ظاهری ناشی از روشانی 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 [？]	۱۵
۱۲	اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می‌باشند [？]	۱۶
۱۳	نمونه‌هایی از مستطیل‌های ویژگی‌های Haar [？]	۱۸
۱۴	مستطیل‌های ویژگی‌های Haar متناسب با بخش‌های چهره می‌باشند [？]	۱۸
۱۵	هر مستطیل در اندازه‌های مختلف بر روی بخش‌های مختلف تصویر قرار می‌گیرد [？]	۱۹
۱۶	یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [？]	۱۹
۱۷	نحوه مقدار دهی به پیکسل‌های تصویر یکپارچه [？]	۲۰
۱۸	بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [？]	۲۰

- ۲۱ ۱۰.۲ بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [؟].
- ۲۱ ۱۱.۲ سطح روشنایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می شود [؟].
- ۲۲ ۱۲.۲ برای یافتن چهره ها، بخش هایی از تصویر که به الگوی HOG شبیه تر است را پیدا می کنیم [؟].
- ۲۳ ۱۳.۲ نتیجه اجرای روش مبتنی بر NPD [؟].
- ۲۴ ۱۴.۲ نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره [؟].
- ۲۵ ۱۵.۲ یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].
- ۲۶ ۱۶.۲ ویژگی های هندسی (رنگ سفید) مورد استفاده در آزمایش های تشخیص چهره [؟].
- ۲۶ ۱۷.۲ (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [؟].
- ۲۷ ۱۸.۲ مدل سازی سه بعدی چهره با اشعه فروسرخ [؟].
- ۳۰ ۱۹.۲ نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [؟].
- ۳۱ ۲۰.۲ (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [؟].
- ۳۴ ۲۱.۲ شبکه عصبی عمیق برای شناسایی چهره [؟].
- ۳۴ ۲۲.۲ سه گانه تطبیق - عدم تطبیق [؟].
- ۳۶ ۲۳.۲ (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [؟].
- ۳۶ ۲۴.۲ معماری کلی شبکه GoogLeNet [؟].
- ۳۷ ۲۵.۲ کادر سبز رنگ یکی از بخش های موازی شبکه را نشان می دهد [؟].
- ۳۷ ۲۶.۲ بخش آغازگر شبکه GoogLeNet [؟].
- ۳۷ ۲۷.۲ معماری شبکه VGGFace [؟].
- ۳۸ ۲۸.۲ نتیجه موقعیت ۱۹۴ شاخص روی چهره [؟].
- ۳۸ ۲۹.۲ علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطای همراه می باشد [؟].
- ۳۹ ۳۰.۲ (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری TCNN [؟].
- ۴۰ ۳۱.۲ کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن depth-wise و point-wise هست (سمت راست). [؟].

- ۴۵ ۱.۳ رویکردهای مختلف هم ترازی چهره [؟].
- ۴۶ ۲.۳ رویکرد مبتنی بر تطبیق کلیشه [؟].
- ۴۷ ۳.۳ معماری OpenFace [؟].
- ۴۷ ۴.۳ جریان یادگیری در معماری OpenFace [؟].
- ۴۸ ۵.۳ رویکرد کلی الگوریتم مبتنی بر AAM برای رو به روسازی چهره [؟].
- ۴۸ ۶.۳ مقدار دهی اولیه و بهینه سازی AAM [؟].
- ۴۹ ۷.۳ نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [؟].
- ۴۹ ۸.۳ معماری شبکه پیشنهادی VS2VI [؟].
- ۵۰ ۹.۳ (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [؟].
- 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 [؟].
- ۵۲ ۱۱.۳ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [؟].
- ۵۳ ۱۲.۳ اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [؟].
- ۵۴ 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 [؟].
- ۵۶ ۱۴.۳ رویکرد کلی الگوریتم GD-HASLR [؟].
- ۵۷ ۱۵.۳ ردیابی، یافتن چهره ها و برچسب زنی [؟].
- ۵۷ ۱۶.۳ تصاویر با حاشیه قرمز رنگ، به اشتباہ برچسب زنی شده اند [؟].
- ۵۸ ۱۷.۳ استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [؟].
- ۵۹ ۱۸.۳ تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [؟].
- ۶۰ ۱۹.۳ ساختار شبکه AD GAN - شبکه GN برای رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [؟].
- ۶۱ ۲۰.۳ معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [؟].

- ۲۱.۳ نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [؟]. ۶۲
- ۲۲.۳ چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [؟]. ۶۳
- ۲۲.۳ تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مربن دسته ها ایجاد می نماید [۴۹]. ۶۳
- ۱.۴ نمای کلی از روش پیشنهادی [؟]. ۶۷
- ۲.۴ نتیجه اعمال یکسان سازی بافت نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می باشند [۵۰]. ۶۸
- ۳.۴ نمونه از خروجی الگوریتم یافتن چهره retina [۵۱]. ۶۹
- ۴.۴ نمونه از خروجی الگوریتم یافتن چهره retina [۵۱]. ۷۰
- ۵.۴ رویکردهای مختلف هم ترازی چهره [؟]. ۷۱
- ۶.۴ رویکردهای مختلف هم ترازی چهره [؟]. ۷۴
- ۷.۴ رویکردهای مختلف هم ترازی چهره [؟]. ۷۵
- ۱.۵ منحنی FROC مقالات در مقایسه با روش پیشنهادی برای مجموعه داده E-Ophtha-MA [؟]. ۸۲
- ۲.۵ منحنی FROC مقالات در مقایسه با روش پیشنهادی برای مجموعه داده DIARETDB1 [؟]. ۸۲

فهرست نمادها

فصل ١

مقدمة

۱.۱ مقدمه

استفاده روز افزون از سامانه‌های تشخیص هوشمند چهره موجب اهمیت این شاخه از علم هوش مصنوعی شده است. 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 و... 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 مسئله خاص در زمینه تشخیص چهره به صورت بی‌درنگ و در محیط کنترل نشده پرداخته و نیازها و ابزارهای مورد نیاز بررسی شده است.

۲.۱ ویژگی‌های بیومتریک

امروزه در زمینه‌های فراوانی نیاز به سامانه‌ای می‌باشد که هویت اشخاص را بر اساس ویژگی‌های بدن آن‌ها شناسایی کند. این زمینه علمی علاقه مندان فراوانی پیدا کرده و استفاده از ویژگی‌های بیومتریک^۱ در سال‌های اخیر به صورت گسترش موردن 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 بدار ویژگی‌های خاص فیزیولوژیک بازشناسی می‌کند.

۱.۲.۱ ارزیابی ویژگی‌های بیومتریک انسان

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

- عمومیت: هر شخصی باید دارای آن ویژگی بیومتریک باشد.

¹Biometric

- یکتایی: آن ویژگی بیومتریک باید برای هر شخصی منحصر بفرد 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 دقت، 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.

۳.۱ سامانه تشخیص چهره

تشخیص چهره همواره یکی از موضوعات مورد مطالعه در علوم کامپیوتر و هوش مصنوعی بوده است. هدف این است که عکسی به کامپیوتر داده شود و سامانه تشخیص دهد که آیا چهره‌ای در تصویر مشاهده می‌کند یا خیر، و در صورت وجود یک چهره، در صورت امکان آن را شناسایی نماید. 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. 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 تصاویر است که رتبه بندی شده است.

۱.۳.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره

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

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

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

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

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

¹ Access Control

² Tradeoff

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 پیش پردازش قرار می‌گیرند. سپس ارزیابی محتوایی شده و داده‌های نامربوط از

قبيل پس زمينه، 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.

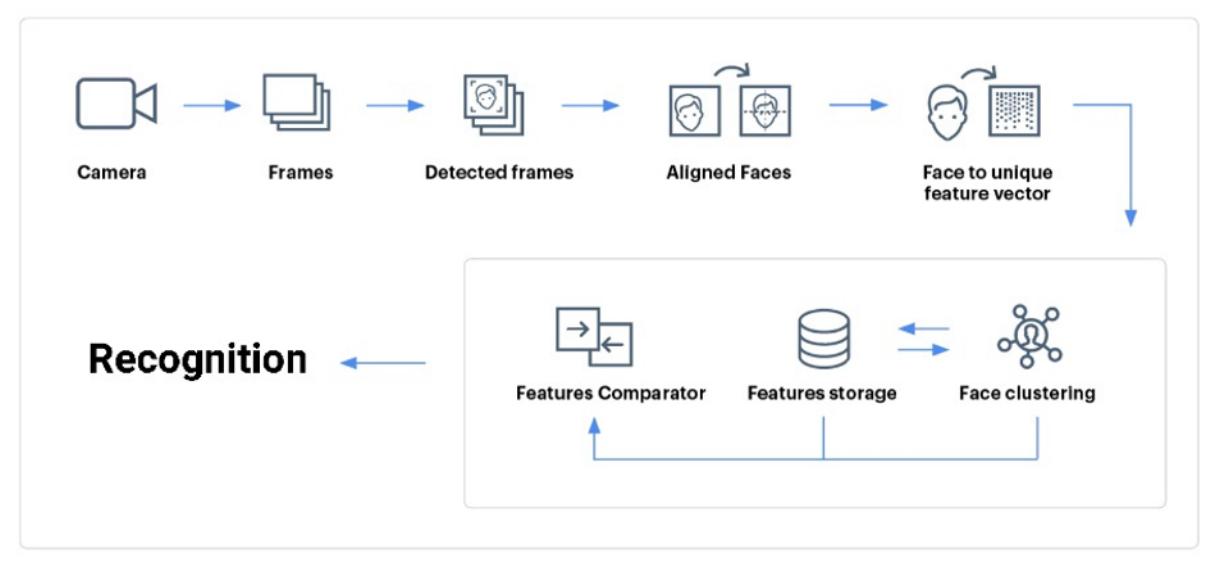
۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر

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

¹Deep Learning



[۱]: نمای کلی یک سامانه تشخیص چهره [۴]

عیق^۱ به همراه اعمال الگوریتم‌های مختلف خوشه‌بندی^۲، وابازش^۳ یا طبقه‌بندی^۴ داده‌ها است که سعی در شناسایی و تشخیص تصاویر مختلف دارند. عنصر اصلی این گونه روش‌ها، داده‌های مورد استفاده در فرایند یادگیری می‌باشد. نوع و فراوانی داده‌ها در افزایش دقت سامانه‌های تشخیص تصاویر مبتنی بر شبکه عصبی بسیار مهم می‌باشد.

۱.۴.۱ خوشه‌بندی

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

خوشه‌بندی سخت^۵: برچسب گذاری به صورت صفر و یک بر روی داده‌ها می‌باشد و مشخص می‌کند داده مورد نظر مربوط به خوشه می‌باشد یا خیر.

خوشه‌بندی نرم^۶: به خوشه‌بندی فازی معروف است و میزان تعلق یک داده به خوشه‌ها را مشخص می‌کند و هر داده می‌تواند با توجه به وزن‌های اختصاص داده شده به آن، به چندین خوشه تعلق داشته باشد.

¹Deep Neural Network

²Clustering

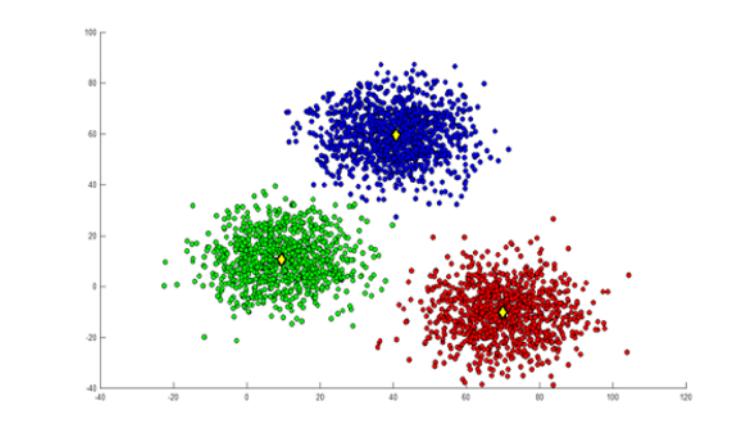
³Regression

⁴Classification

⁵Hard Clustering

⁶Soft Clustering

داده‌های استفاده شده در فرایند خوشبندی، بدون برچسب بوده و یادگیری بدون ناظر^۱ در فرایند خوشبندی داده‌ها استفاده می‌شود. شکل ۱-۲ نمونه‌ای از خوشبندی سخت و یافتن شاخص برای هر خوشبندی را نشان می‌دهد.



شکل ۱-۲: 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^۲ هستند که با توجه به یادگیری داده‌های آموزشی به همراه برچسب آن‌ها، داده‌های آزمایشی را نیز برچسب زنی می‌کنند. در طبقه‌بندی داده‌ها می‌توان از توابع سنجش مختلفی برای سنجش میزان تعلق یک داده به هر دسته استفاده کرد. شکل ۳-۱ یک طبقه‌بند غیر خطی ساده را نشان می‌دهد.

در این نوع از الگوریتم‌ها که از لحاظ تعداد، بار اصلی یادگیری ماشین را بر دوش می‌کشنند، دونوع 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. فرآیند 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.

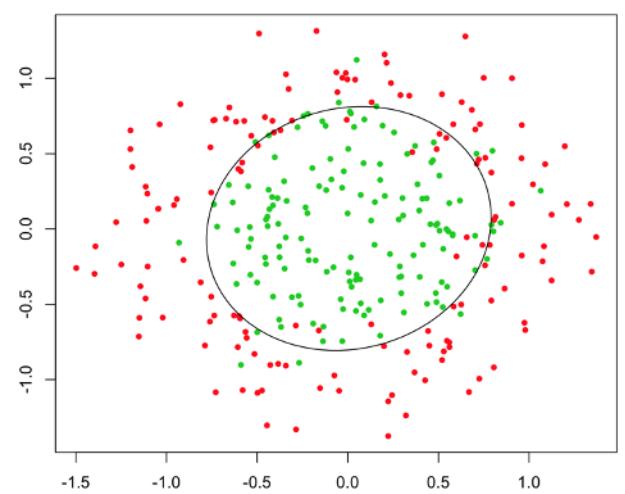
الگوریتم‌های مختلفی برای طبقه‌بندی داده‌ها وجود دارد که در این میان می‌توان به SVM^۳ و KNN^۴ اشاره کرد.

¹Unsupervised Learning

²Supervised Learning

³Support Vector Machine

⁴K Nearest Neighbor



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 ۱-۴ برای افراد نابینا می‌باشد. هنگامی که شخص نابینا عینک را بر روی چشم‌اش 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 و زیر بخش‌های آن‌ها در فصل دوم بررسی قرار خواهد گرفت.



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

۱.۵.۱ چالش‌های سامانه تشخیص چهره

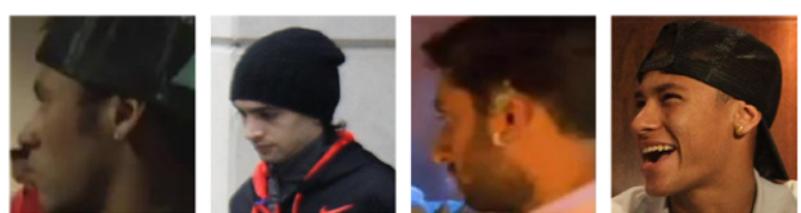
سامانه‌های تشخیص چهره به سطح مشخصی از بلوغ رسیده‌اند، اما توسعه آن‌ها در شرایط کنترل نشده و کاربردهای واقعی هنوز مسیر طولانی در پیش دارد. برای مثال، تشخیص چهره در محیط با تغییرات شدید نورپردازی و حالت چهره، انسداد صورت، وضوح پایین تصویر و...، وردیابی آن در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی و... مشکل می‌باشد. دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چسب نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. در ادامه مهم‌ترین چالش‌ها بر شمرده شده است.

نورپردازی^۱: روشنایی محیط در شب و روز و یا در محیط داخلی و خارجی به شدت تغییر می‌کند. با توجه به ساختار سه بعدی چهره، یک منبع نور مستقیم می‌تواند سایه‌های قوی بر روی چهره ایجاد کند که برخی ویژگی‌های چهره را تغییر می‌دهد. همانطور که در شکل ۱-۵ نشان داده شده است که گاهی تفاوت‌های ظاهری ناشی از روشنایی، بیشتر از تفاوت بین چهره افراد مختلف می‌باشد.



شکل ۱-۵: مقایسه تفاوت‌های ظاهری ناشی از روشنایی و تفاوت بین چهره افراد [؟]

حالت^۲: زاویه چهره نسبت به حسگر دوربین، می‌تواند سامانه را چالش مواجه نماید. چهره با زاویه تند و چهره‌های نیم رخ مانند شکل ۱-۶ 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 ۱-۶: زاویه شدید چهره نسبت به دوربین باعث کاهش دقیقت سامانه می‌شود [؟]

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

¹Lighting

²Pose

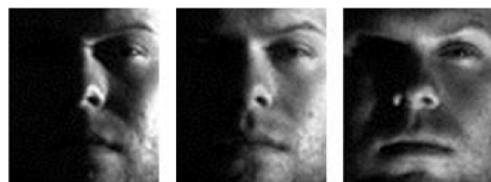
³Out Of Focus Blur

رخ خواهد داد (شکل ۱-۷). 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 دردرس ساز شود.



شکل ۱.۷: تاری خارج از تمرکز به علت عمق کم میدان دوربین [۹]

انسداد^۱: اگر در حال تصویر برداری از یک جمع باشیم، احتمال اینکه چهره فردی توسط فرد دیگری مقداری پوشانده شود، بسیار بالاست. همچنین اگر موهای فرد بر روی صورتش ریخته باشد، از عینک آفتابی استفاده کند، یا نورپردازی غیر یکنواخت باشد (شکل ۱-۸)، انسداد چهره رخ خواهد داد.



شکل ۱.۸: انسداد شدید در اثر نورپردازی [۹]

سن^۲: بیشتر روش‌های مرسوم تشخیص چهره تغییرات سن را نادیده می‌گیرند. بعضی از رویکردها به طور منظم پایگاه داده تصاویر را به روز رسانی کرده و بازآموزی سامانه را انجام می‌دهند. این راه حل فقط برای سامانه‌هایی مناسب است که اغلب وظیفه احراز هویت کارمندان را انجام می‌دهد. در شرایط دیگر سن موضوع را باید جدی گرفت و تلاش کرد تا سامانه نسبت به این نوع تغییرات قوی تر شود.

کمبود داده‌های آموزشی: سامانه‌های تشخیص چهره در کاربردهای واقعی دارای مشکل کمبود داده‌های آموزشی برای آموزش سامانه می‌باشند. تعداد افراد در محیط کنترل نشده زیاد می‌باشد و 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

¹Occlusion

²Aging

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

۲.۱ یافتن چهره

یافتن مکان چهره در تصویر، اولین گام در فرایند تشخیص چهره می‌باشد که نقشی کلیدی در سامانه ایفا می‌نماید. هدف اصلی الگوریتم یافتن چهره این است که تعیین کند آیا چهره‌ای در تصویر وجود دارد یا خیر و در صورت وجود چهره، مکان آن را بیابد. یافتن چهره در تصویر، امری پیچیده است. زیرا چهره انسان همواره دستخوش تغییراتی مانند شرایط روشنایی ووضوح تصویر، حالت چهره، رنگ پوست، حضور عینک یا موی صورت و... می‌شود. در سال ۲۰۰۲ Yang و همکاران در [۱] یک دسته‌بندی برای روش‌های یافتن چهره ارائه کردند که در ادامه شرح داده می‌شود.

۱.۲.۱ رویکردهای مبتنی بر دانش

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 موبایل روی پیشانی ریخته باشد یا در حضور عینک، به درستی عمل نمی‌کند. یا به عنوان مثال:

- هر چهره داری 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.

¹Face Detection

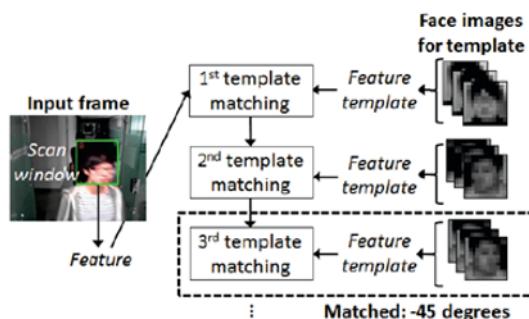
²Face Recognition

- چهره مانند یک ناحیه کوچکتر است که بر روی یک ناحیه بزرگتر مانند شانه‌ها قرار گرفته است.
- چهره انسان متقاض است.

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 تنهایی کافی نیست و موفق به یافتن چهره‌ها در شرایط کنترل نشده با تعداد زیادی چهره نمی‌باشد.

۲.۲.۲ رویکردهای مبتنی بر تطبیق کلیشه یا الگو

این رویکرد با استفاده از قالب^۱ های از پیش تعیین شده برای یافتن چهره‌ها با همبستگی بین الگوها و تصاویر ورودی استفاده می‌نماید. 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 ۱-۲ 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.



شکل ۱.۲: رویکرد مبتنی بر تطبیق کلیشه [؟]

۳.۲.۲ رویکردهای مبتنی بر ویژگی

این رویکردها با استخراج ویژگی^۲ های ساختاری چهره، چهره‌ها را پیدا می‌نمایند. ابتدا به عنوان یک طبقه‌بند، آموزش دیده و سپس برای تمایز میان نواحی شامل چهره و بدون چهره در تصویر استفاده می‌شوند. ایده این رویکرد، غلبه بر محدودیت دانش ما از چهره‌ها

¹Template

²Feature

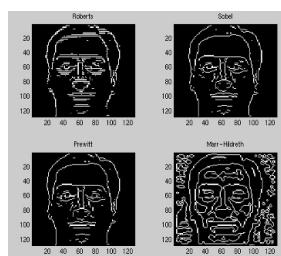
می باشد. در [۲] تعدادی از این رویکردها مورد بررسی قرار گرفته است:

۱.۳.۲.۲ رویکرد مبتنی بر حرکت

اگر یک توالی از چند فریم در اختیار باشد، می توان به کمک اطلاعات حرکت^۱، یافتن چهره را انجام داد. برای این کار به کمک تفاصیل فریمها، قسمت متحرک نسبت به پس زمینه شناسایی شده و بخش بالای آن جدا می شود. بدین ترتیب می توان با احتمال بالا، مکان چهره یک فرد را در یک تصویر پیدا نمود. این رویکرد در مواجهه با اجسام متحرک دیگری مانند اتومبیل دچار اشتباه می شود.

۲.۳.۲.۲ رویکرد مبتنی بر لبه

ابتدا به کمک یک الگوریتم لبه یاب^۲، لبه های تصویر بدست می آید، سپس نازک سازی می شوند و شاخه های اضافه حذف می گردد (شکل ۲-۲). گوشی لبه ها تشخیص داده می شوند و هر مولفه متصل^۳ به شاخه مرکزی آن کاهش می یابد. اجزایی که ویژگی چهره در آن ها نیست حذف می شوند و اجزای نهایی به عنوان سمت چپ چهره، خط مو، یا سمت راست چهره بر چسب گذاری می شوند. در یک آزمایش که ۶۰ تصویر دارای پس زمینه پیچیده حاوی ۹۰ چهره به این سامانه داده شده است، سامانه قادر به یافتن ۷۶٪ چهره ها بوده است.



شکل ۲.۲: استفاده از لبه یاب های معروف برای استخراج ویژگی های مبتنی بر لبه از چهره [۲].

¹Motion

²Edge Detection

³Connected Component

۴.۳.۲.۲ رویکرد مبتنی بر اطلاعات سطح خاکستری

سطوح خاکستری^۱ تصویر چهره شامل اطلاعات مفیدی می‌باشد. برای مثال ابروها، مردمک چشم و لب‌ها معمولاً تاریک‌تر از سایر نواحی صورت هستند. این ویژگی‌ها می‌تواند به یافتن یک چهره در تصویر کمک نماید. در این رویکرد ابتدا بر روی تصویر ورودی، عملیات بسط تباین^۲ و عملیات مورفولوژی^۳ مبتنی بر سطح خاکستری انجام می‌شود تا تصویر بهبود پیدا کند و یافتن نواحی تیره‌تر، راحت‌شود. سپس تصویر به چندین بخش تقسیم می‌شود و سطوح خاکستری بخش‌ها مورد بررسی قرار می‌گیرد. مزیت این رویکرد، کارایی در تصاویر با وضوح پایین می‌باشد (شکل ۳-۲).



شکل ۳.۲: اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می‌باشند [؟].

۴.۳.۲.۳ 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^۴ 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.

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

¹Gray level

²Contrast Stretching

³Morphological Operations

⁴Normalization

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

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

۴.۲.۲ 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 داده شود، بسیاری از این رویکردها شکست می‌خورند. استفاده از عامل‌های آماری^۲ برای ویژگی‌ها و ارائه یک مدل احتمالی برای چهره، باعث انعطاف‌پذیری بیشتر سامانه می‌شود. این رویکرد قادر است در مواجه با 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. تا به امروز صدها رویکرد برای یافتن چهره ارائه شده است تا آن را پیشرفت‌هه تر و دقیق‌تر نماید، اما انقلاب الگوریتم‌های یافتن چهره در سال ۲۰۰۱ Viola Jones در [۳] 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، تارشدن^۶ و تیزکردن لبه‌های تصویر

¹Histogram

²Statistical Parameters

³Bayes Rule

⁴Resizing

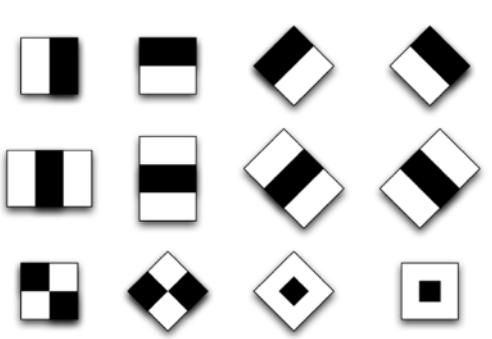
⁵Cropping

⁶Blurring

^۱ 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.

شکل ۴-۲ نمونه‌هایی از مستطیل‌های ویژگی‌های Haar را نشان می‌دهد.



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

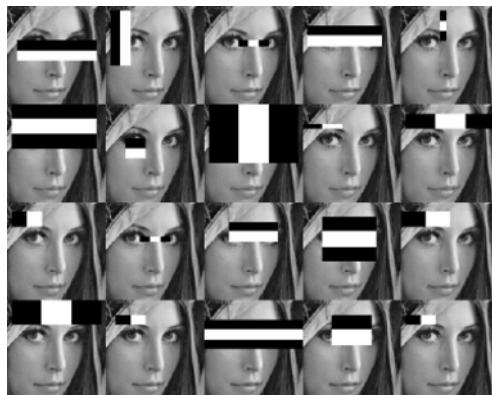
ویژگی‌های Haar برای تشخیص چشم، بینی، دهان و... با کمک تشخیص لبه، تشخیص خط و تشخیص مرکز در تصویر و استخراج ویژگی برای یافتن چهره استفاده می‌شود. مستطیل‌های ویژگی‌های Haar، متناسب با بخش‌های چهره می‌باشند که مثالی از آن در شکل ۵-۲ نشان داده شده است.

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 روشناکی پیکسل‌های زیر بخش‌های سیاه از مجموع سطح روشناکی 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.

^۱Sharpening

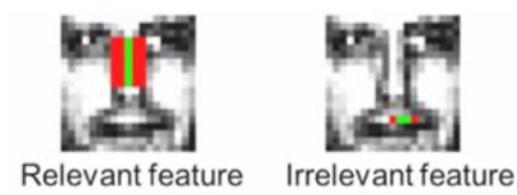


شکل ۵.۲: مستطیل‌های ویژگی‌های Haar متناسب با بخش‌های چهره می‌باشند [?].



شکل ۲: هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [۹].

همانطور که در شکل ۲ مشاهده می شود، تمام ویژگی های Haar برای تصویر ورودی مناسب نخواهد بود. بعضی از این ویژگی ها باید نادیده گرفته شوند و فقط ویژگی های مرتبط انتخاب شوند تا در زمان صرفه جویی شود. این کار به صورت خودکار به کمک عنصر Ada Boost انجام می شود.



شکل ۷.۲: یک ویژگی مرتبط در مقابل یک ویژگی نامرتب [۹].

یک الگوریتم مبتنی بر یادگیری ماشین می باشد که ویژگی های کاربردی را از میان تعداد زیادی ویژگی پیدا می کند. بعد از شناسایی ویژگی های مختلف، مشخص می گردد که هر یک از پنجره ها برای بخشی از چهره مناسب می باشد یا خیر. هر کدام از ضرایب انتخاب شده مثبت در نظر گرفته 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.

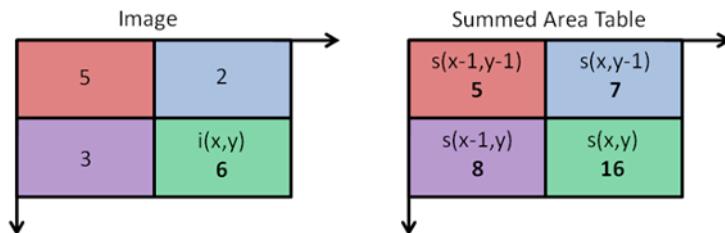
$$F(x) = {}_1f_1(x) + {}_2f_2(x) + \dots \quad (5.2)$$

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. وزن مربوط به طبقه بند می باشد. با استفاده از این الگوریتم و وزن دادن به ویژگی ها، بیش از ۱۶۰۰۰ ویژگی قبلی به کمتر از ۲۵۰ ویژگی کاهش پیدا می کند.

¹Weak Classifier

²Strong Classifier

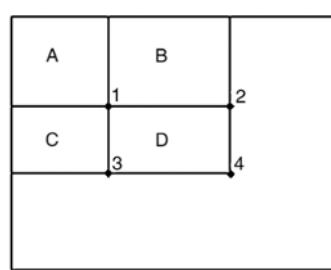
تصویر یکپارچه^۱: تصویر یکپارچه، یا جدول محدوده مجتمع^۲، به منظور ارزیابی سریع‌تر ویژگی‌هایی که در بخش اول معرفی شد، استفاده می‌شود. با توجه به شکل ۸-۲ در یک تصویر یکپارچه مقدار پیکسل در مکان x و y برابر با جمع مقادیر پیکسل‌های بالا و چپ پیکسل x و y می‌باشد.



شکل ۸.۲: نحوه مقدار دهی به پیکسل‌های تصویر یکپارچه [۹].

به عنوان مثال در شکل ۹-۲ مقدار پیکسل‌ها به صورت زیر محاسبه می‌شود:

- مقدار پیکسل ۱ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل A.
- مقدار پیکسل ۲ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل A و B.
- مقدار پیکسل ۳ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل A و C.
- مقدار پیکسل ۴ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل A و B و C و D.
- مجموع پیکسل‌ها در مستطیل D می‌تواند به صورت $(3+2)-(1+4)$ محاسبه شود.



شکل ۹.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [۹].

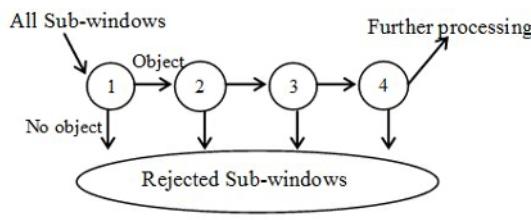
مراحل آبشاری^۳: اگر تصویر به مربع‌های 24×24 پیکسلی تقسیم شده و با پردازش هر بخش که ۲۵۰۰ ویژگی دارد، تشخیص داده شود که چهره‌ای در تصویر وجود دارد یا خیر، حجم محاسبات بسیار زیاد خواهد بود. مراحل آبشاری این فرایند را سریع‌تر انجام می‌دهد.

¹Integral Image

²Summed area table

³Cascading

۲۵۰۰ ویژگی هر مربع 24×24 به دسته‌بندی‌های مختلف تقسیم می‌شود. برای مثال ۱۰۰ ویژگی در دسته دوم، ۱۰۰ ویژگی در دسته سوم و... می‌توان بعد از پردازش هر دسته، در ارتباط با وجود یا عدم وجود چهره در آن دسته تصمیم گرفت تا بخش‌هایی که چهره در آن وجود ندارد زودتر حذف شوند. شکل ۱۰-۲ یک نمای کلی از روند تشخیص آشنا را نشان می‌دهد. مجموعه‌ای از طبقه‌بندها به هر زیر پنجه اعمال می‌شود. طبقه بند 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.

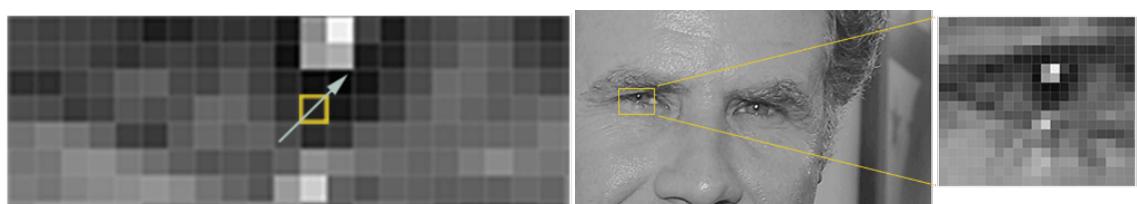


شکل ۱۰.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [۴].

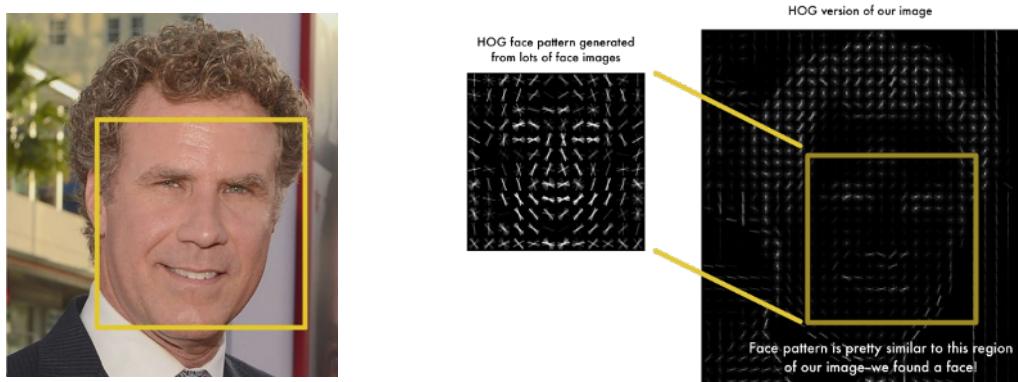
در سال ۲۰۰۵ Dalal و همکاران در [۴] روشی به نام بافت نگار شبیه‌های جهت دار^۱ ارائه کردند که به اختصار HOG نامیده می‌شود. در این روش ابتدا تصویر خاکستری می‌شود، زیرا نیازی به رنگ نیست. پیرامون هر پیکسل بررسی می‌شود تا مشخص شود نسبت به پیکسل‌های پیرامونش چقدر تاریک می‌باشد. مطابق شکل ۱۱-۲ جهتی انتخاب می‌شود 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

¹Histograms Of Oriented Gradients



شکل ۱۱.۲: سطح روشنایی پیکسل‌های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می‌شود [۴].



شکل ۱۲.۲: برای یافتن چهره‌ها، بخش‌هایی از تصویر که به الگوی HOG شبیه‌تر است را پیدا می‌کنیم [۴].

یکسانی خواهد داشت که حل مسئله را آسان‌تر 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 16×16 پیکسل تقسیم می‌شود و در هر بخش تعداد جهت‌های به سمت بالا، پایین، چپ و راست شمارش می‌شود. سپس بخش‌های درون تصویر با جهت‌هایی که بزرگتر بودند، جایگزین می‌شود. 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 [۵] 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 خوشه‌بندی در مرحله آموزش نیز لازم نیست و در برابر تغییرات نور، حالت، انسداد، تصاویر با وضوح پایین و... مقاوم است. NPD بین دو پیکسل در یک تصویر به صورت زیر تعریف شده است:

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

که در آن x و y بزرگتر از صفر هستند و $f(0, 0)$ 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 $p = s \times s$ 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

¹Sign

خدشه^۱ 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.



شکل ۱۳.۲: نتیجه اجرای روش مبتنی بر NPD [۹].

تمام رویکردهای ارائه شده که یافتن چهره را با مدل سازی صريح از ویژگی‌های صورت انجام می‌دهند، در برابر تغییرات غیر قابل پیش‌بینی چهره و شرایط محیطی دچار مشکل می‌شوند. اگرچه بعضی از تلاش‌های اخیر مبتنی بر ویژگی، توانایی مقابله با شرایط کنترل نشده را بهبود 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 روش-هایی هست که بتوانند در شرایط خصم‌مانه تر مانند تشخیص چهره‌های متعدد در زمینه‌های شلوغ به خوبی عمل کنند.

۵.۲.۲ رویکردهای مبتنی بر تصویر

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 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.

¹Noise

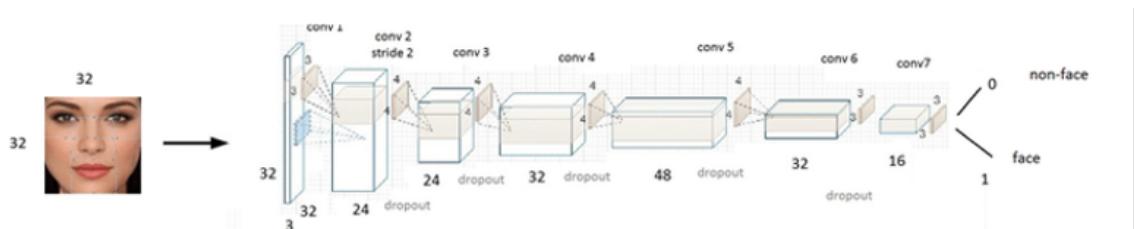
۱.۵.۲.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

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 [۶] از این طبقه‌بند برای یافتن چهره استفاده کردند.

۲.۵.۲.۲ 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 بود که در مجموعه داده‌های ساده، امیدوار کننده بود. شبکه‌های عصبی با معناری پیمانه ای^۲ امروزی بسیار پیچیده‌تر از MLP^۳ 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. شکل ۱۴-۲ نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره را نشان می‌دهد. 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 کاهش حجم داده‌ها با محاسبه مقدار بیشینه، میانگین یا اندازه اقلیدسی

هر بخش



شکل ۱۴.۲: نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره [?].

با توجه به آنچه در [۷] 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

¹Neural Network

²Modular Architecture

³Multi Layer Perceptron

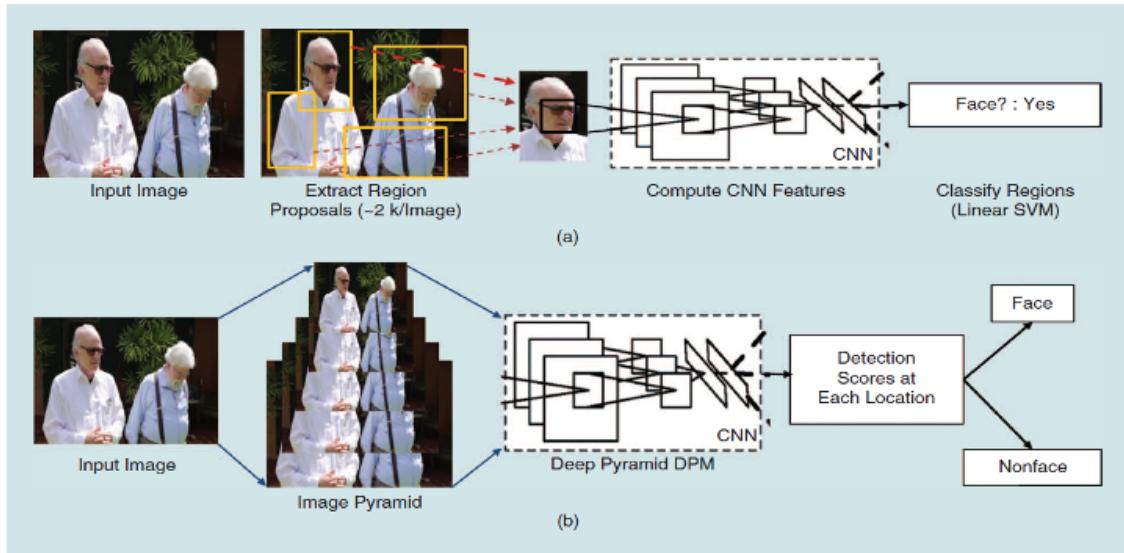
⁴Convolutional Neural Network

⁵Convolution Layer

⁶Fully Connected Layer

⁷Pooling Layer

بیشتر به دلیل ویژگی‌های استفاده شده در طبقه‌بندها است. با پیشرفتهای اخیر در رویکردهای یادگیری عمیق و در دسترس بودن پردازنده‌های گرافیکی، استفاده از شبکه‌های عصبی پیچشی عمیق برای استخراج ویژگی امکان پذیر شده است. ویژگی‌های عمیق به دست آمده به طور گستردۀ ای برای یافتن چهره استفاده می‌شود. با توجه به شکل ۱۵-۲ روش‌های مبتنی بر شبکه عصبی پیچشی عمیق برای یافتن چهره به دو زیر شاخه تقسیم می‌شود: رویکرد مبتنی بر ناحیه و رویکرد پنجره لغزان.



شکل ۱۵.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].

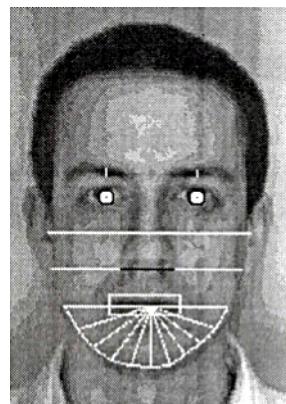
۳.۲ شناسایی چهره

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 الگویی نمایند و الگوهای مقایسه می‌نمایند. 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 [۸] یکی از اولین الگوریتم‌ها در این زمینه را ارائه دادند که رویکرد موفقی مبتنی بر روش‌های تطبیق الگو داشت (شکل ۱۶-۲). این رویکرد در شرایط کنترل شده به دقت ۹۰٪ رسید.



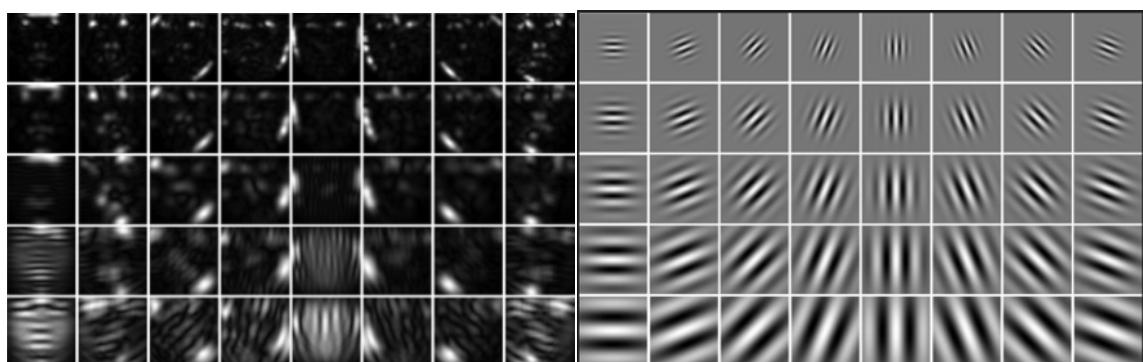
شکل ۱۶.۲: ویژگی‌های هندسی (رنگ سفید) مورد استفاده در آزمایش‌های تشخیص چهره [۹].

۲۰.۳.۲ رویکرد مبتنی بر فیلتر گابور

همانطور که در [۹] آمده است، در این رویکرد ابتدا تصویر را بخش بندی^۱ کرده، سپس بر روی بخش‌های مختلف آن، فیلتر گابور اعمال می‌شود و نتیجه بدست آمده با یک طرح از پیش آمده شده، با یک آستانه گذاری مطابقت داده می‌شود. شکل ۱۷-۲ فیلترهای چندگانه گابور^۲ و تاثیر این فیلترها بر روی تصویر چهره انسان را نشان می‌دهد. دلیل استفاده از فیلتر گابور این است که عملکرد این فیلتر به سامانه بصری انسان بسیار شباهت دارد.

¹Grid

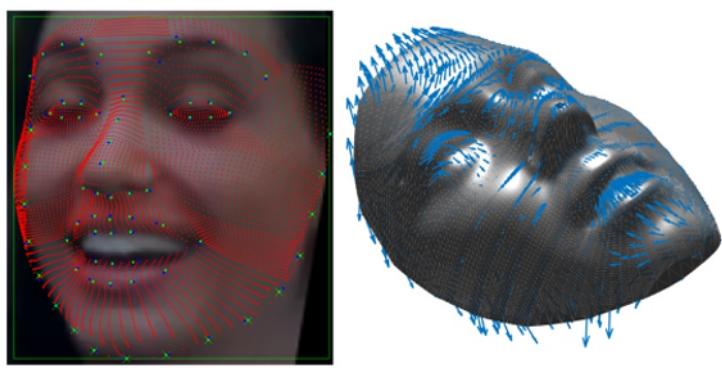
²Gabor



شکل ۱۷.۲: (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۹].

۳.۳.۲ رویکردهای سه بعدی

همانطور که در [۱۰] آمده است، داده‌های سه بعدی دقت تشخیص چهره را به شدت بهبود می‌بخشد، اختلاف داده‌های ورودی با داده‌های ذخیره شده زیادتر است و سامانه با دقت بیشتری عمل می‌کند. روش تشخیص سه بعدی چهره از یک منتشر کننده نور فرو سرخ و یک حسگر به عنوان دریافت کننده استفاده می‌کند. شبکه‌ای از نورهای فرو سرخ که برای انسان قابل رویت نیست، روی چهره تابانده می‌شود. سپس یک حسگر ویژه، پرتوهای بازتاب را دریافت کرده و اطلاعات عمق تصویر پردازش می‌شود. این دسته از الگوریتم‌ها برای شناسایی دقیق اشخاص، برای هر نفر برداری‌های سه بعدی می‌سازند. عیب رویکردهای سه بعدی، نیاز به تجهیزات پیشرفته و غیر قابل استفاده بودن در شرایط کنترل نشده مانند خیابان و معابر پیاده می‌باشد. شکل ۱۸-۲ یک مدل سازی سه بعدی چهره با اشعه فرو سرخ را نشان می‌دهد.



شکل ۱۸.۲: مدل سازی سه بعدی چهره با اشعه فرو سرخ [۹].

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 اندازه گیری 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 [۱۱] 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. برخلاف روش تشخیص دو بعدی، روش سه بعدی و تجزیه و تحلیل بافت پوست به تجهیزات بسیار پیچیده تری نیاز دارد، و با توجه به آنکه تمرکز ما بر روی تشخیص چهره به صورت بی‌درنگ در شرایط کنترل نشده مانند معابر پیاده و خیابان می‌باشد، به توضیح مختصر رویکردهای سه بعدی و تجزیه و تحلیل بافت پوست بسنده می‌کنیم.

۵.۳.۲ رویکردهای مبتنی بر دوربین حرارتی

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

در سال ۲۰۰۳ Diego Socolinsky و همکاران در [۱۲] از شناسایی چهره مبتنی بر دوربین حرارتی در کاربردهای واقعی بهره برداری کردند و یک مجموعه داده جدید از تصاویر حرارتی چهره ایجاد کردند. آن‌ها از حسگرهای الکتریکی فروسرخ با حساسیت کم و با توانایی جذب حرارت طولانی مدت یا LWIR^۱ استفاده کردند. نتایج نشان می‌دهد که تلفیق LWIR و دوربین‌های معمولی، نتایج بهتری در شرایط کنترل نشده دارد. در این مطالعه ۲۴۰ چهره مجزا در طی ۱۰ هفته برای ایجاد پایگاه داده جدید استفاده شده است. داده‌ها در روزهای آفتابی، بارانی و ابری جمع آوری شد. در شرایط کنترل شده دوربین معمولی دقیق ۹۷٪ دارد، در حالی که روش LWIR دارای دقیق ۹۳٪ می‌باشد و ترکیب این دو دارای دقیق ۹۸٪ است. در شرایط کنترل نشده دوربین معمولی دقیق ۶۷٪، دوربین LWIR دقیق ۸۳٪ و ترکیب این دو دارای دقیق ۸۹٪ است.

^۱Longwave Infrared

٦.٣.٢ تشخیص چهره مبتنی بر ویدیو

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

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

چهره و پیگیری آن در طول زمان، موجب انتخاب فریم‌های خوب می‌شود که حاوی چهره‌های رو به رو یا نشانه‌های ارزشمند

است که شرایط نور، انسداد، حالت چهره و... در آن رضایت بخش می‌باشد.

۲. مطالعات روان‌پزشکی 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 ۱۹-۲ نمای کلی این رویکرد را نشان می‌دهد.

٧.٣.٢ رویکرد مبتنی بر چهره ویژه

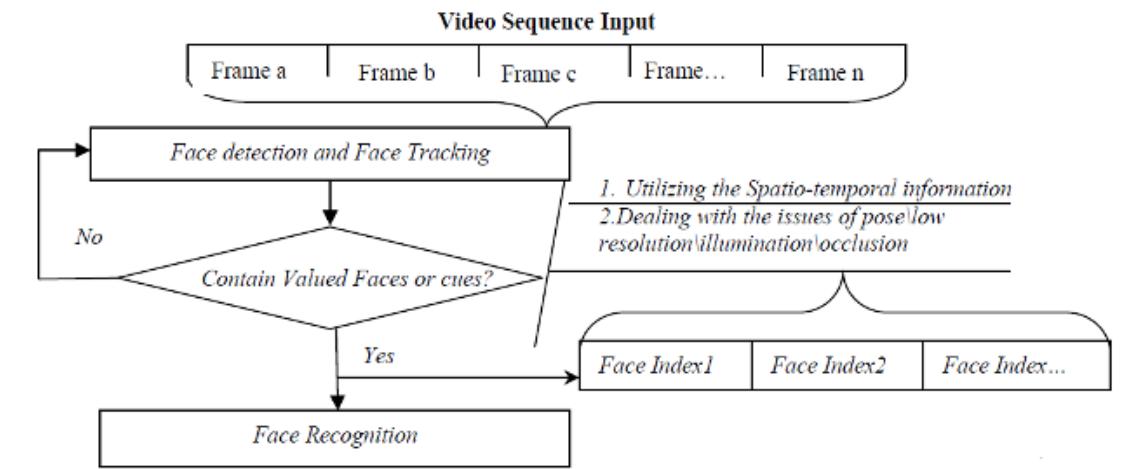
با افزایش حجم داده‌های موجود، نیاز به کاهش ابعاد داده‌ها می‌باشد. تحلیل مؤلفه‌های اساسی یا PCA^۳ یک روش کاهش ابعاد

داده‌ها است که در برخی مسئله‌ها مانند پردازش تصویر به خوبی و با سرعت بالا عمل می‌کند. استفاده از این روش در پردازش سریع

¹Frame

²Detection And Tracking

³Principle Component Analysis



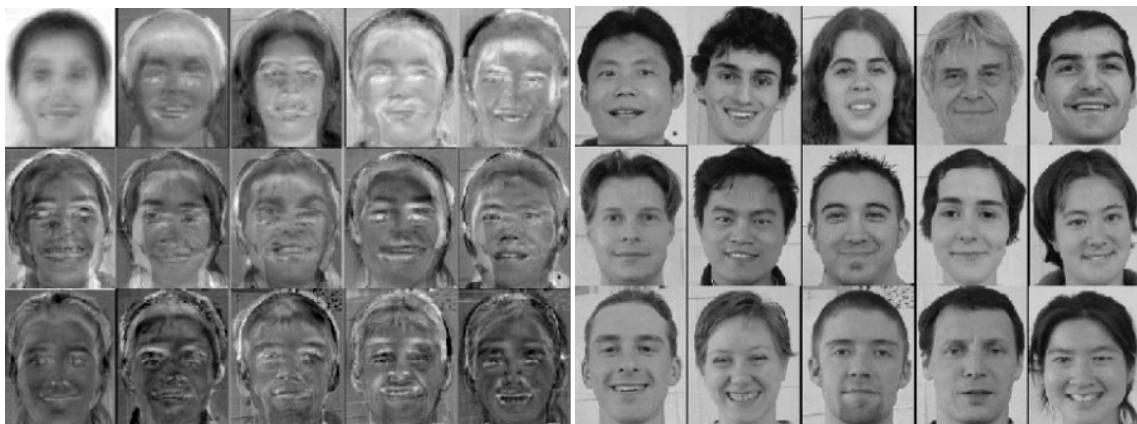
شکل ۱۹.۲: نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [۴].

ترداده‌ها کمک می‌کند و از رخداد مشکل بیش برآزندن^۱ جلوگیری می‌نماید. اگر یک پایگاه داده عظیم از تصاویر چهره اشخاص با وضوح بالا موجود باشد که هر کدام دارای تعدادی زیاد ویژگی هستند، و بخواهیم یک تصویر آزمایش را با این پایگاه داده مقایسه کرده و شخص شبیه به آن را پیدا کنیم، مقایسه تصاویر بسیار زمان بر و در مواردی غیر ممکن خواهد بود. PCA در این مسئله به خوبی عمل می‌کند. با اعمال تکنیک کاهش بعد به تصویر و با به دست آوردن تصویر ویژه چهره‌ها^۲ می‌توان ویژگی‌ها را کاهش داد و نتیجه مطلوب را در زمان بسیار کم گرفت. شکل ۲۰-۲ تعدادی چهره و چهره‌های ویژه متناظر با آن‌ها را نشان می‌دهد.

در سال ۲۰۱۴ Xiao Luan و همکاران در [۱۴] یک روش تشخیص چهره مبتنی بر PCA ارائه دادند که تا حدی در برابر تغییرات نورپردازی و انسداد مقاوم می‌باشد. PCA راستای بیشترین تغییرات را با توجه به تعداد ویژگی‌ها و نوع آن‌ها به ما می‌دهد. به همین دلیل در برخی موارد که تنها محوری که بیشترین تغییرات یا پراکندگی را دارد برای ما مهم است، راه حل مناسبی خواهد بود. در سال ۲۰۱۶ K. R. Sreelakshmi و همکاران در [۱۵] یک روش شناسایی چهره مبتنی بر چهره‌های ویژه ارائه دادند که در آن ابتدا تصویر ورودی با استفاده از ماتریس بردارهای ویژه، به فضای دیگری منتقل می‌شود، سپس در فضای کاهش بعد یافته با داده‌های موجود مقایسه شده و شبیهترین تصویر به آن انتخاب می‌شود. برای مقایسه از معیارهایی مانند معیار اقلیدسی و منتهن 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 نشدن بین تصاویر مختلف یک شخص در پایگاه و نیاز به بروز رسانی تمامی اطلاعات موجود با ورود یک تصویر جدید به پایگاه داده از معایب این روش است. محاسبات ریاضی و مراحل انجام آن‌ها:

¹Overfitting

²Eigenface



شکل ۲۰.۲: (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [؟].

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 رویکرد مبتنی بر ماشین بردار پشتیبان

همان طور که در بخش ۲-۲-۱-۴ گفته شد، ماشین های بردار پشتیبان، طبقه بند های خطی هستند که حاشیه بین ابر صفحه تصمیم و

نمونه های مجموعه آموزش را به حد اکثر می رسانند. در سال ۲۰۱۲ N.M.Khan و همکاران در [۱۶] این طبقه بند را برای تشخیص

چهره مورد استفاده قرار دادند. یک نسخه مبتنی بر هسته برای SVM معرفی شده که K-SVM نام گذاری شده است.

$$a = b \quad (7.2)$$

که در آن w بردار وزن ها، x داده های ورودی و w_0 مقدار پیش قدر^۱ می باشد. سپس یک مدل بی پارامتر از LDA^۲ به نام NDA^۳

¹Bias

²Linear Discriminant Analysis

³Nonparametric Discriminant Analysis

معرفی شده، سپس یک نسخه مبتنی بر هسته^۱ به نام KNDA^۲ معرفی شده است.

$$S_{bk}^{\phi} = \frac{1}{(N_1+N_2)} \sum_{i=1}^{N_1} \omega'_i (\Phi(x_i) - M_2^k(\Phi(x_i))) (\Phi(x_i) - M_2^k(\Phi(x_i)))^T + \frac{1}{(N_1+N_2)} \sum_{i=N_1+1}^{N_1+N_2} \omega'_i (\Phi(x_i) - M_1^k(\Phi(x_i))) (\Phi(x_i) - M_1^k(\Phi(x_i)))^T \quad (8.2)$$

که در آن N1 و N2 تعداد نمونه‌های آموزشی در هر دسته می‌باشد. سپس از ترکیب دو رویکرد فوق، مدلی به نام SVM + NDA معرفی شد که در هر دسته می‌باشد. سپس از ترکیب دو رویکرد فوق، مدلی به نام طراحی نموده است.

$$\begin{aligned} \min_{w \neq 0, w_0, C > 0} \quad & \frac{1}{2} w^T (\lambda S_w^{\phi} + \beta I)^{-1} w + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & t_i \left(\sum_{j=1}^N w_j K(x_i; x_j) + w_0 \right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, N \end{aligned} \quad (9.2)$$

که در آن I ماتریس تنظیم می‌باشد و ضریبی برای تنظیم کنترل میان SVM و NDA می‌باشد. رابطه بالا یک مسئله بهینه‌سازی می‌باشد که به صورت تکراری قابل حل می‌باشد. دقت این روش در مقایسه با سایر روش‌های مشابه در جدول ۱-۲ آورده شده است.

۹.۳.۲ رویکردهای مبتنی بر شبکه عصبی

یک راه حل غیر خطی برای شناسایی چهره، استفاده از شبکه عصبی پیچشی است که به طور شگفت‌انگیزی در طبقه-بندی 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 ۲۱-۲ ساختار کلی یک شبکه عصبی عمیق برای شناسایی چهره را نشان می‌دهد.

معمولًا به عنوان تابع فعالیت از توابع غیرخطی مانند ReLU^۳ استفاده می‌شود. و عملیات بهینه‌سازی به روش پس انتشار خطای انجام می‌گردد. و در لایه خروجی از تابع SoftMax برای طبقه‌بندی استفاده می‌شود که خروجی‌های لایه آخر را هنجار می‌کند.

شبکه‌های عصبی پیچشی ویژگی‌های یک چهره را استخراج می‌کنند که می‌توان به عنوان یک شناسه برای یک فرد خاص در نظر گرفت. هنگامی که دو تصویر مختلف از چهره یک شخص به عنوان ورودی داده می‌شود، شبکه باید خروجی‌های مشابه (ویژگی‌های

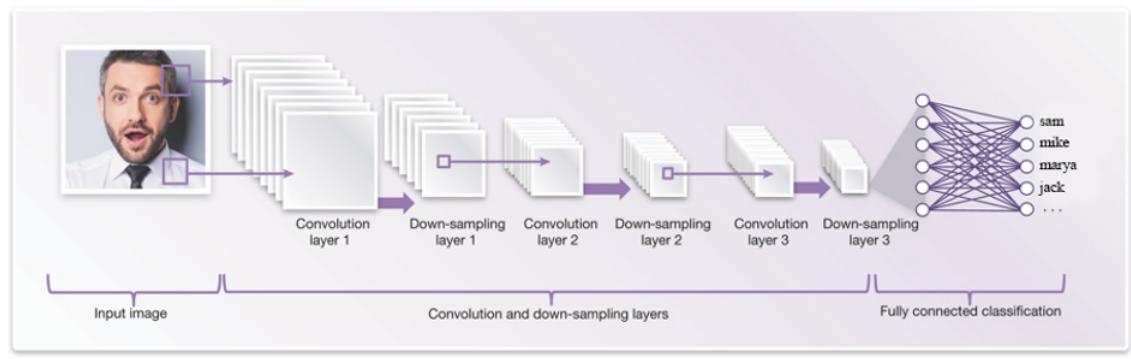
¹Kernel

²Kernel Nonparametric Discriminant Analysis

³Rectified Linear Unit

جدول ۱.۲: مقایسه دقت الگوریتم NDA + SVM با سایر رویکردهای مشابه

Data set	SVM + NDA	LDA	NDA	HLDA	SVM	SVM + LDA
Breast-Cancer	73.8	71.9	72	72.3	72.9	72.9
Flare-Solar	66.9	65.4	65	64.9	66.2	66.5
German	76.7	75.6	74.9	75.9	76.1	75.9
Heart	84.9	83	83.5	83.2	84	83.6
Diabetes	76.7	74.7	74.7	74.2	76.7	76.6
Ringnorm	75.9	75	75.6	75.3	75	74.9
Splice	84.2	82.9	83.3	83.8	83.8	83.2
Thyroid	90.6	87.7	89.4	89.5	88.6	89.2
Twonorm	98	96.9	97	97.3	96.6	96.6
Waveform	86.9	84.6	85.6	86.9	86.7	86.9
Image	85.7	83.6	82.6	82.3	85.2	85.5
Banana	57.3	58	56.7	56.9	55.5	57
Glass	79.4	77.6	79.1	78	78.3	78.7
E-Coli	89.6	88.1	88	90.2	89	89.5
Libras-Movement	93.3	86.6	89.3	88	89.4	90.2
Wine	98	95.2	96.6	96.8	97	97.2
Avg. time	0.86	0.09	0.15	0.22	0.86	0.86
Confidence (%)	-	99.99	100	99.87	99.96	99.95

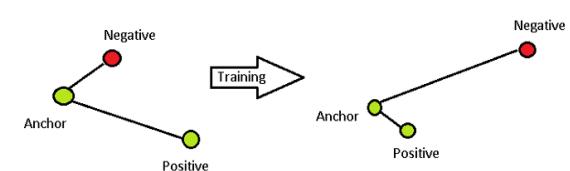


شکل ۲۱.۲: شبکه عصبی عمیق برای شناسایی چهره [؟].

نرده‌یک 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 ۱.۹.۳.۲

در سال ۲۰۱۵ FaceNet یک شبکه عصبی عمیق به نام FaceNet ارائه دادند. Florian Schroff 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 فضا 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 ۲۲-۲، 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 ۲۲.۲: سه گانه تطبیق - عدم تطبیق [؟].

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 زیر برای هر کدام از داده‌های آموزشی برقرار باشد.

$$f(A) - f(P)^2 f(A) - f(N)^2 \quad (10.2)$$

که در آن تابع 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 به صورت زیر تعریف می‌شود:

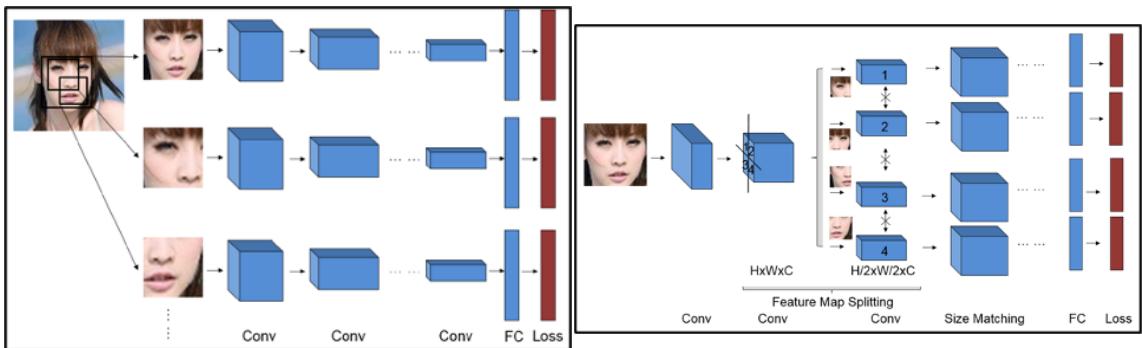
$$L = (i=1)^n f(x_i^A) - f(x_i^P)^2 - f(x_i^A) - f(x_i^N)^2 + \quad (11.2)$$

که در آن x_i^P و x_i^N نمونه‌های مثبت و منفی برای نمونه آموزشی x_i^A می‌باشند و 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.

SplitNet شبکه ۲.۹.۳.۲

در سال ۲۰۱۸ و همکاران در [۱۸] 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 ۲۳-۲ 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.

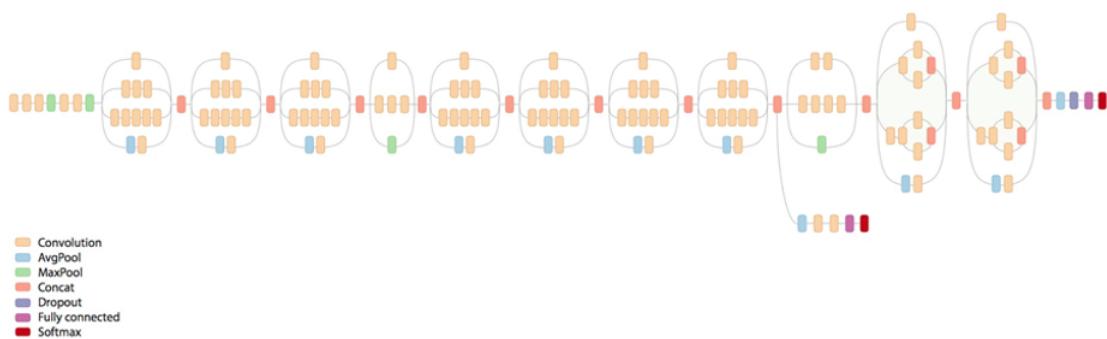
^۱Feed Forward



شکل ۲۳.۲: (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [؟].

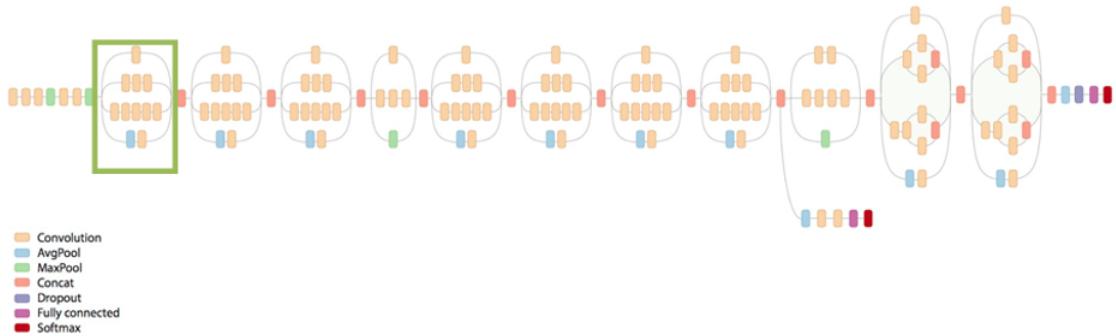
GoogLeNet شبکه ۳.۹.۳.۲

در سال ۲۰۱۵ و همکاران در [۱۹] یک شبکه عصبی عمیق به نام GoogLeNet ارائه دادند. همانطور که در شکل ۲۴-۲ 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.

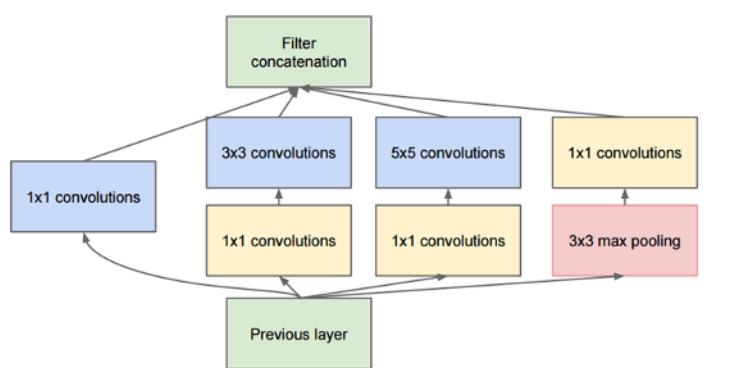


شکل ۲۴.۲: معماری کلی شبکه GoogLeNet.[؟]

در GoogLeNet تمام محاسبات به طور متواالی اتفاق نمی‌افتد، بلکه هر بخش شبکه روندی موازی دارد. کادر سبز رنگ در شکل ۲۵-۲ بخش آغازگر نامیده می‌شود. در ادامه نگاهی دقیق‌تر به این بخش خواهیم داشت. کادر سبز پایین در شکل ۲۶-۲ 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 ۲۵.۲: کادر سبز رنگ یکی از بخش های موازی شبکه را نشان می دهد [۴].



شکل ۲۶.۲: بخش آغازگر شبکه GoogLeNet [۴].

۴.۹.۳.۲ شبکه VGGFace

در سال ۲۰۱۵ Omkar M. Parkhi و همکاران در [۲۰] شبکه عمیق VGGFace را ارائه کردند که شامل یک توالی طولانی از لایه های پیچشی می باشد. با توجه به شکل ۲۷-۲ 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 log-loss softmax 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.

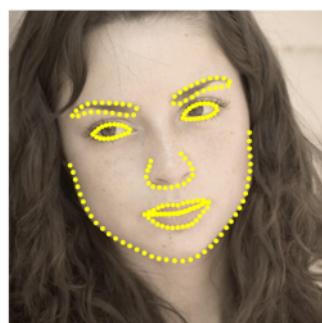


شکل ۲۷.۲: معماری شبکه VGGFace [۴].

^۱Triplet Loss Function

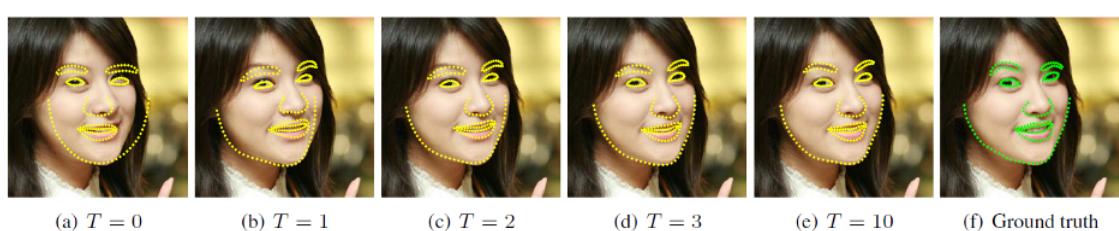
چهره‌هایی که در جهت‌های مختلفی هستند، برای سامانه تشخیص چهره، متفاوت به نظر می‌رسند. برای غلبه بر این چالش در رویکردهای مبتنی بر نقاط راهنمای سعی می‌شود تصویر را چرخانده و جایه جا نمود، بطوریکه چشمها و لب‌ها در یک موقعیت خاص در تصویر قرار بگیرند. بدین ترتیب مقایسه چهره‌ها در مرحله بعد بسیار ساده‌تر خواهد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

در سال ۲۰۱۴ وحید کاظمی و جوزفین سالیوان در [۲۱] یک الگوریتم برای یافتن نقاط راهنمای^۱ بر روی چهره ارائه دادند که از ۱۹۴ نقطه خاص که در هر چهره ای وجود دارد استفاده می‌نماید. شکل ۲۸-۲ 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. پس



شکل ۲۸.۲: نتیجه موقعیت ۱۹۴ شاخص روی چهره [۴].

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 ۲۹-۲ علامت گذاری تکراری خطوط راهنمای بر روی چهره را نشان می‌دهد.



شکل ۲۹.۲: علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطأ همراه می‌باشد [۹].

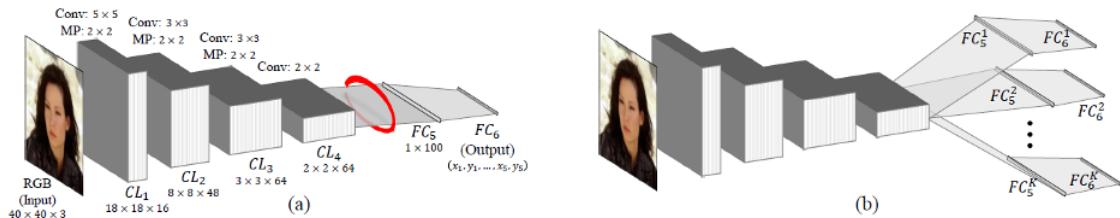
در سال ۲۰۱۶ Yue Wu و همکاران در [۲۲] رویکردی برای یافتن نقاط راهنمای بر روی چهره مبتنی بر شبکه عصبی پیچشی ارائه

^۱Landmark

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 $f(x) = |\tanh(x)|$ در نظر گرفته شده است. و در نهایت تابع زیر به عنوان تابع ضرر معرفی شده است.

$$L(P_i, \hat{P}_i) = \frac{(P_i - P_{i2}^2)}{(P_{(i,1)} - P_{(i,2)}^2)} \quad (12.2)$$

که در آن P_i یک بردار $m * 2$ 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 خوش‌نشانی شده و هریک به صورت جداگانه بررسی شده است. معماری TCNN در شکل ۳۰-۲ قسمت b قابل مشاهده می‌باشد. این شبکه برای آموزش از مجموعه داده LFW استفاده کرده است.



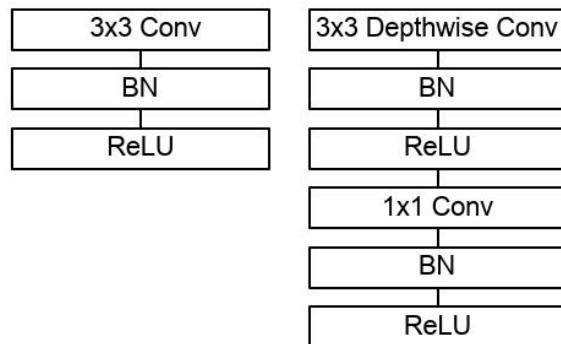
شکل ۳۰.۲: (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری [?] TCNN.

در سال‌های بعد شبکه‌های عمیق‌تر، پهن‌تر و البته پیچیده‌تر مانند ResNet GoogleNet و... 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 کامپیوتوری 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 با هدف طراحی شبکه‌های کارآمد، سبک، سریع و با دقت قابل قبول مطرح شده است. در سال ۲۰۱۹ Howard Andrew و همکاران در [۲۳] معماری MobileNet نسخه ۳ را ارائه دادند. در این مقاله یک نوع کانولوشن جدید به نام convolution separable depth-wise معرفی شد که قلب تپنده شبکه موبایل نت است. در کانولوشن dws ابتدا کانولوشن عمقی اعمال می‌شود و سپس کانولوشن نقطه‌ای که به ترتیب نقش مراحل فیلتر و ادغام در کانولوشن استاندارد را دارند. در کانولوشن استاندارد M کرنل $k \times k$ داشتیم. اما در اینجا تنها

جدول ۲.۲: 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

یک کرنل $k \times k$ داریم! حالا با این کرنل، مرحله اول کانولوشن را انجام می‌دهیم. با انجام عمل فیلتر، هر صفحه از کرنل در یک صفحه از فیچرmap ورودی F کانولوشنی شود. به این مرحله کانولوشن عمقی گفته می‌شود. مرحله دوم، کانولوشن نقطه‌ای (point-wise) هست. این مرحله معادل با مرحله ادغام در کانولوشن استاندارد است. اما بازهم یک تفاوت اساسی بین مرحله ادغام در کانولوشن استاندارد و کانولوشن dws وجود دارد: مرحله ادغام در کانولوشن استاندارد، یک جمع ساده هست، 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؛ کانولوشن نقطه‌ای 1×1 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 ۳.۲: کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن depth-wise و point-wise هست (سمت راست). [?].

شبکه عصبی موبایل نت ۲.۴ میلیون پارامتر دارد. وقتی تعداد پارامترهای این شبکه را با شبکه محبوب ResNet-۱۸ با ۱۱ میلیون پارامتر مقایسه کنیم، متوجه می‌شویم که چقدر میزان پارامترها کمتر است.

۴.۲ جمع‌بندی

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 قبیل نیز بیان شد، برای حل مساله دسته بندی چهره دوروش 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 گرفته اند شامل رویکردهای مبتنی بر شبکه عصبی، بردار پشتیبان و... می‌باشند.

۳ 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
[۲۷.۲۴]	حالت چهره	تبدیل دو بعدی	پیچیدگی محاسباتی قابل قبول	استخراج نقاط ویژه باید دقیق تر باشد
[۳۱.۲۸.۲۲]	حالت چهره	استفاده از شبکه عصبی عمیق	دقت بالا در شرایط کنترل نشده	پیچیدگی محاسباتی، وابستگی به داده های آموزش
[۳۵.۳۲.۱۱]	حالت چهره	تبدیل مدل دو بعدی به سه بعدی	دقت بالا پیچیدگی محاسباتی	دقت بالا در شرایط کنترل نشده
[۲۶]	حالت چهره	تبدیل مدل سه بعدی به دو بعدی	دقت بالا پیچیدگی محاسباتی	قابل استفاده در تصاویر خاکستری
[۲۷.۲۷]	روشنایی	همسان سازی بافت-نگار	پیچیدگی محاسباتی قابل قبول	استخراج نقاط ویژه باید دقیق تر باشد
[۴۹.۴۸]	انسداد	استفاده از روش های شناسایی الگو و ایاش	دقت بالا در انسداد شدید	نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو
[۴۰]	حدودیت داده	تهییه مجموعه داده با رذایی چهره در ویدیو	دقت بالا در انسداد شدید	تجهیزات پیشرفته و زمان تاخیر ناهمگن
[۴۲.۴۱]	حدودیت منابع	استفاده از رایانش ابری	سرعت بالا	

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 اشتباہ 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 [۲۷.۲۴].

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. پیشرفت های اخیر مانند [۲۴] 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، ۳۱-۲۸]

مشکل: این الگوریتم ها، به طور متوسط، کنترل 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 روش زی 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 [۴۳] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند.

۳. رویکردهای سبک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۱۱، ۳۵-۳۲].
مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده ترین این الگوریتم ها در مورد پیچیدگی محاسباتی، [۳۲] است که در شرایط بدون محدودیت آموزش دیده و آزمایش شده است. شامل مراحل مجازی رو به روش زی 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 [۳۶] مشکل: زمان محاسباتی (نه به اندازه الگوریتم های دسته سوم).

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

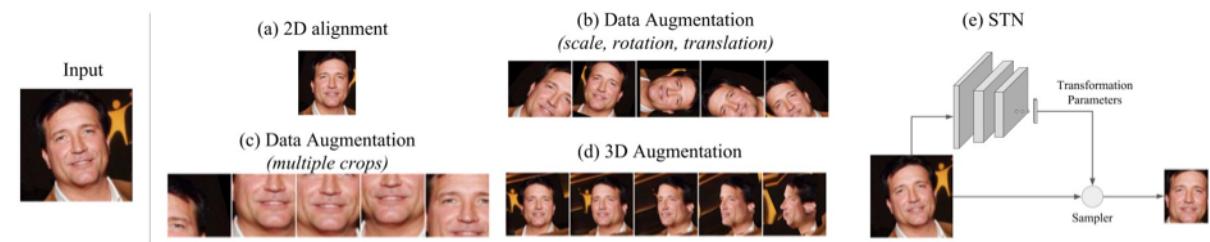
مرجع [۴۳] در سال ۲۰۱۸ خلاصه ای از رویکردهای مختلف برای حل مسئله هم ترازی را در شکل ۱-۳ نشان داده است.
تصویر سمت چپ، چهره ورودی می باشد. (a) هم ترازی با استفاده از تبدیلات دو بعدی ساده می باشد. (b) داده افزایی^۳ با تغییر

¹Computational Complexity

²Patch-Based

³Data Augmentation

مقیاس، تغییر زاویه و جا به جایی می باشد. (c) برش های چندگانه می باشد. (d) داده افزایی مبتنی بر روش های سه بعدی می باشد. (e) از هیچ ابزاری برای هم ترازی مستقیم استفاده نمی نماید. اما یک شبکه را آموزش می دهد تا عامل های مورد نیاز برای تبدیل هم ترازی را بدست آورد.



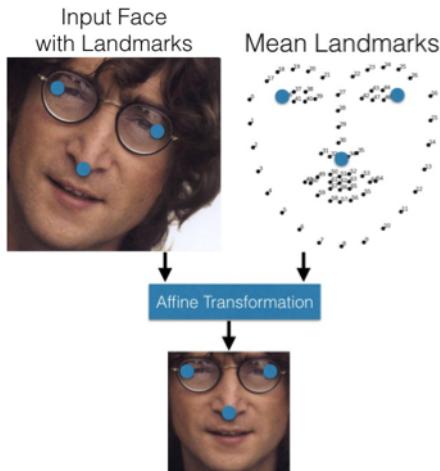
شکل ۱.۳: رویکردهای مختلف هم ترازی چهره [۴].

در سال ۲۰۱۶ Brandon Amos و همکاران در [۲۶] یک روش شناسایی چهره به نام OpenFace ارائه دادند که ویژگی اصلی آن، آموزش شبکه عصبی عمیق در کمترین زمان و قابلیت اجرا بر روی دستگاه های قابل حمل مانند تلفن همراه با در نظر گرفتن منابع محدود می باشد. یک تصویر شامل تعدادی چهره به الگوریتم داده می شود. پس از یافتن چهره ها و مجزا کردن^۱ آن ها از یکدیگر، هر چهره به طور جداگانه مورد پیش پردازش^۲ قرار می گیرد و حجم آن کاهش می یابد. کاهش حجم تصویر برای عملکرد مناسب یک طبقه بندی بهینه بسیار مهم می باشد. تصاویر چهره ها باید هنجارسازی شده و ابعاد آن ها ثابت گردد تا به بخش شناسایی چهره راه یابند. هر تصویر چهره باید مورد تبدیل قرار بگیرد تا چشم ها، بینی و دهان، در مکان مشخصی قرار گیرند. بدین منظور از یک تبدیل هم نسبی^۳ دو بعدی ساده استفاده می گردد. ابتدا باید چهره توسط ۶۸ نقطه ویژه، نشانه گذاری شود. سپس نشانه های اطراف چشم ها و بینی (شکل ۲-۳) برای محاسبه عامل های تبدیل هم نسبی استفاده می شوند. پس از انجام تبدیل هم نسبی، تصاویر چهره برش زده و اندازه آن ها 96×96 پیکسل می شود. شکل ۲-۳ - تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی پس از پیش پردازش، تصاویر چهره ها به عنوان ورودی به یک شبکه عصبی پیچشی داده می شوند (شکل ۳-۳). این الگوریتم برای تعلیم شبکه از مجموعه داده کوچکی با 500 هزار تصویر چهره استفاده می کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری شده به نام FaceScrub و CASIA-WebFace بدست آمده است. شبکه مورد استفاده در این الگوریتم یک نسخه اصلاح شده از شبکه nn4 FaceNet می باشد. شبکه nn4 مبتنی بر معماری GoogLeNet می باشد. برای تعیین میزان شباهت نتیجه، از فاصله اقلیدسی استفاده شده است. هر تصویر از یک شبکه یکتا به یک سه گانه نگاشت داده می شود. گرادیان خطای سه

¹Isolate

²Preprocessing

³Affine Transformation



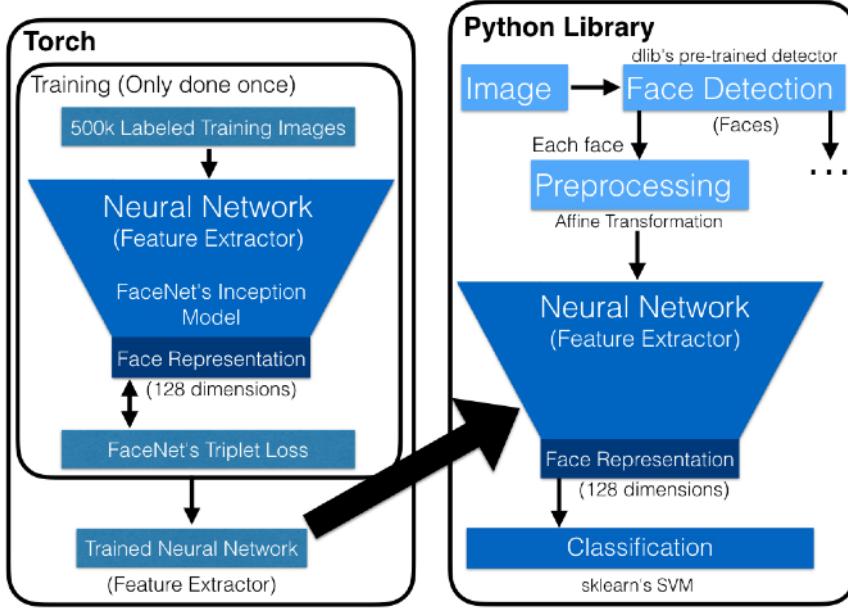
شکل ۲.۳: رویکرد مبتنی بر تطبیق کلیشه [؟].

گانه برای هر تصویر محاسبه شده و به عقب انتشار می‌یابد. در هر دسته کوچک^۱، P تصویر برای هر نفر از Q نفر، در مجموعه داده انتخاب می‌شود. سپس $M \approx PQ$ تصویر به شبکه داده می‌شود تا عملیات forward انجام پذیرد. در این مقاله از $P=20$ و $Q=15$ استفاده شده است. تمام جفت‌های anchor-positive برای بدست آوردن سه گانه‌های $N = Q \binom{P}{2}$ مورد استفاده قرار می‌گیرند. خطای سه گانه محاسبه شده و مشتق آن برای پس انتشار خطا استفاده می‌شود. شکل ۴-۳ چگونگی آموزش شبکه را نشان می‌دهد. مجموعه داده LFW یک معیار استاندارد برای سنجیدن میزان دقیقت الگوریتم‌های تشخیص چهره می‌باشد. الگوریتم OpenFace بر روی این مجموعه داده مورد سنجش قرار گرفت که به دقیقت $0.9292 \pm 0.0134\%$ رسید. در سال ۲۰۱۶ Mohammad Haghhighat و همکاران در [۲۴] 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 می‌شود. در این الگوریتم ابتدا یک AAM بر روی تصویر چهره قرار گرفته، با روندی تکراری و به صورت بهینه شونده، بر روی چهره تنظیم می‌شود. سپس با استفاده از یک تبدیل هم نسبی، مرحله رو به رو سازی بر روی چهره انجام می‌پذیرد. شکل ۵-۳ رویکرد کلی این الگوریتم را نشان می‌دهد. در این مدل یک تصویر چهره با مجموعه‌ای از نقاط ویژه هنجارسازی شده مدل می‌شود که به صورت $[x_i, y_i]$ تعریف می‌شود که در آن $i=1, 2, \dots, n$. برای انجام این کار یک مرحله یادگیری نیاز است. سپس الگوریتم PCA اعمال می‌شود تا کاهش میزان وابستگی میان نقاط ویژه در هر مجموعه انجام می‌شود و نتیجه یک مدل خطی است که یک مدل شکل نمونه را به صورت زیر نمایش می‌دهد.

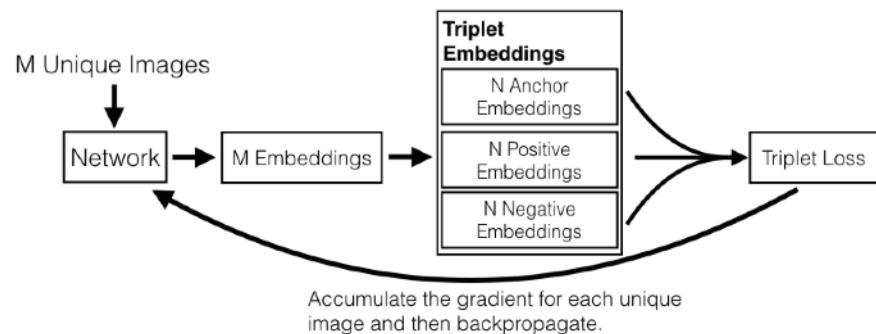
$$S = s_0 + \sum_{i=1}^n p_i s_i \quad (1.3)$$

¹Mini Batch

²Active Appearance Model



شکل ۳.۳: معماری [۹]OpenFace.



شکل ۴.۳: جریان یادگیری در معماری [۹]OpenFace.

که در آن s_0 شکل پایه، s_i نشان دهنده i امین شکل پایه و $[p_n, p_2, p_1]$ عامل های شکل می-باشند. ظاهر^۱ مدل

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 a_0 و ترکیب خطی از بردارهای ویژه a_i ، $i=1, 2, \dots, m$ تشکیل می شود که به صورت زیر تعریف می گردد.

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

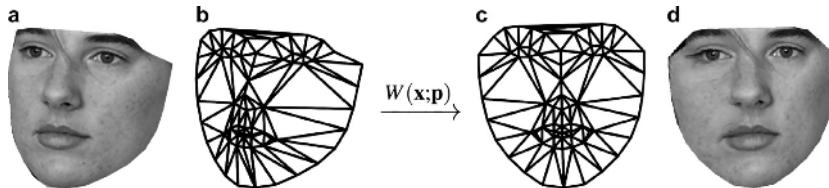
که در آن $[q_m, q_2, q_1]$ عامل های ظاهر می باشند. عامل های شکل و ظاهر برای هر تصویر در فرایند AAM بدست می آید.

الگوریتم های POIC^۲ و SIC^۳ دو الگوریتم شناخته شده برای این منظور می باشند. رویکرد SIC نسبت به POIC در شرایطی که

¹Appearance

²Project-Out Inverse Compositional

³Simultaneous Inverse Compositional



شکل ۵.۳: رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۹].

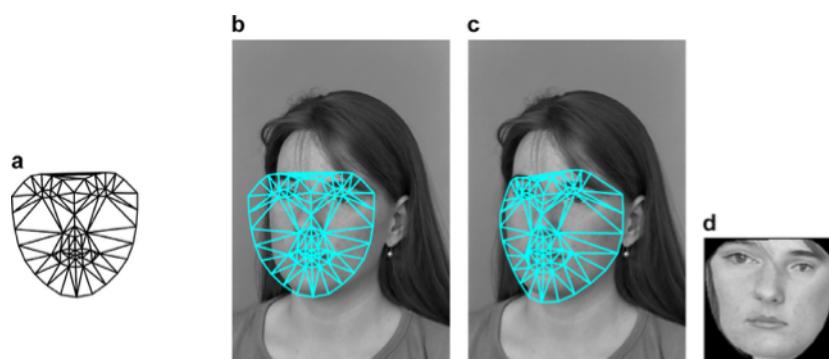
تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می باشد.

در این مقاله از یک روش SIC سریع برای حل مسئله بهینه سازی با 100×100 تکرار استفاده شده است. اگر $p = [p_1, p_2, \dots, p_n]^T$ مجموعه

عامل های بدست آمده باشد، یک تبدیل هم نسبی قطعه ای ^۱ $W(x; p)$ برای رو به رو سازی چهره مورد استفاده قرار می گیرد که

در آن هریک از مثلث های روی توری، به صورت جداگانه به تصویر نتیجه با استفاده از درونیابی نزدیک ترین همسایه ^۲ 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). پس



شکل ۶.۳: مقدار دهی اولیه و بهینه سازی AAM [۹].

از تنظیم کامل مدل بر روی چهره، یک تبدیل هم نسبی با پارامترهای بدست آمده توسط الگوریتم یادگیری یاد شده، می تواند حالت

چهره را هنجارسازی نماید. در بخش شناسایی چهره، ابتدا بخش چانه از تصویر حذف می شود زیرا چانه تقریباً تاثیری در شناسایی

یک چهره ندارد. سپس تصویر چهره به اندازه 64×64 پیکسل تبدیل می شود و به 64×64 بخش غیر هم پوشان با اندازه 8×8 تقسیم می

شود. سپس در هر بخش تبدیل DCT ^۳ انجام می شود. ضرایب خروجی تبدیل DCT بر حسب یک پویش زیگزاگی مرتب می

شوند. اولین ضریب در نظر گرفته نمی شود. زیرا نشان دهنده میانگین سطح خاکستری پیکسل های بخش می باشد. 10×10 ضریب

بعدی که ضرایب فرکانس پایین می باشند، برای ایجاد بردار ویژگی چهره استفاده می شوند. برای آموزش و آزمایش از مجموعه داده

FERET و LFW استفاده شده است که در آن تصاویر چهره با زوایای چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این

¹Piecewise Affine Transformation

²Nearest Neighbor Interpolation

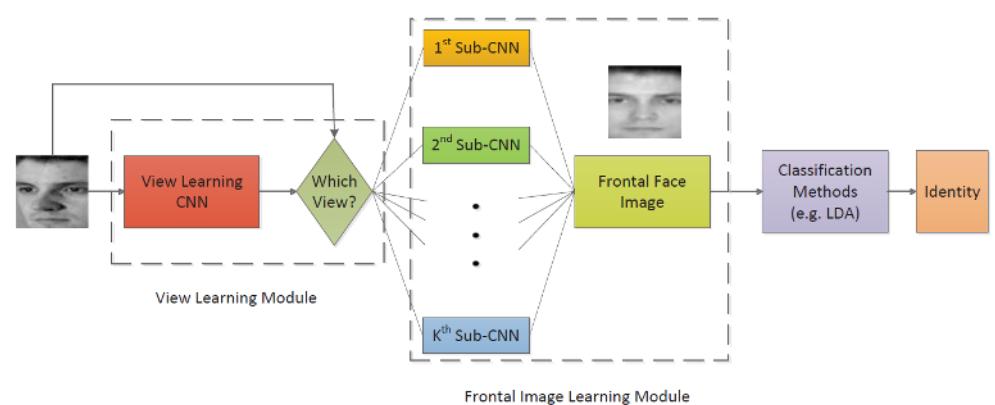
³Discrete Cosine Transform

مقاله موفق به دستیابی به شناسایی چهره با دقت ۳۰.۸۷٪ شده است. در سال ۲۰۱۶، Zhang و همکاران در [۳۰] یک روش رو به



شکل ۷.۳: نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۹].

رو سازی چهره ارائه دادند که شناسایی چهره را مستقل از نمای چهره^۱ انجام می دهد. این الگوریتم یادگیری عمیق که VS2VI نامیده می شود، از دو بخش اصلی تشکیل شده است. بخش اول یک شبکه عصبی پیچشی برای یادگیری نما و زاویه چهره می باشد و بخش دوم از تعدادی شبکه عصبی پیچشی تشکیل شده است که هر کدام برای یادگیری تناظر^۲ بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص می باشد (شکل ۶-۳). این الگوریتم که می تواند با تعداد کمی داده نمونه، به خوبی آموزش ببیند، دو بخش تشکیل شده از شبکه عصبی پیچشی را به هم متصل می نماید تا مشکل نمای چهره در سامانه شناسایی چهره را برطرف نماید. در این معماری برای بازسازی چهره از زاویه رو به رو از لایه های واپیچشی^۳ به جای لایه های تمام متصل استفاده شده است. مدل



شکل ۸.۳: معماری شبکه پیشنهادی [۹] VS2VI.

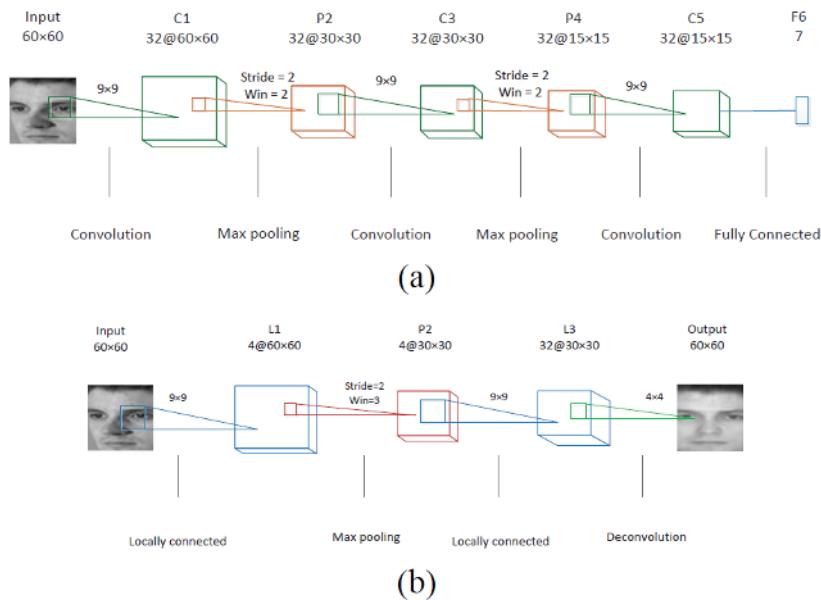
VS2VI از دو بخش اصلی تشکیل شده است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با ابعاد ۶۰×۶۰ دریافت می کند و آن را با توجه به نمای چهره طبقه بندی می کند. سپس تصویر وارد بخش دوم می شود که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص می

¹Facial View

²correspondence

³deconvolutional

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. بخش اول از یک شبکه عصبی پیچشی تشکیل شده



شکل ۹.۳: (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [۴].

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

$$f(x) = \max(0, x) \quad (3.3)$$

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

¹linear discriminant analysis

هزینه آن ℓ_2 -loss می باشد. برای یادگیری شبکه از الگوریتم پس انتشار خطا^۱ استفاده شده است. الگوریتم VS2VI به دقت ۶۵٪ در تشخیص چهره با زاویه ۴۵ درجه رسیده است. در سال ۲۰۱۸ Andrey V.Savchenko و همکاران در [۴۴] یک روش مبتنی بر ML^۲ برای شناسایی چهره در محیط‌های بدون محدودیت با تعداد کم نمونه‌ها بر اساس محاسبه فاصله بین ویژگی‌های با ابعاد بالا که توسط شبکه عصبی پیچشی عمیق مانند VGG، ResNet و SENet استخراج شده است ارائه دادند. این روش جدید شناسایی آماری، احتمال فاصله‌ها را نسبت به تمام تصاویر مجموعه داده‌ها با استفاده از قانون بیز به حداقل می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی Kullback–Leibler^۳ بین ویژگی‌های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده‌های IJB-A و YTF و LFW مورد آزمایش قرار گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقت ۳۰٪ تا ۵۰٪ درصد در مقایسه با روش‌های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند.



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 [؟].

در سال ۲۰۱۳ Marsico و همکاران در [۲۷] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم STASM^۴ برای به دست آوردن ۶۸ نقطه ویژه بر روی چهره استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه (SP)^۴ محاسبه می شود 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 می شود. 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 می دهد. چرخش: چرخش سر در جهت عقربه‌های ساعت یا عکس آن می باشد. و به صورت زاویه تعریف می شود که زاویه بین خط عبوری از مرکز چشم‌ها و محور افقی x می باشد.

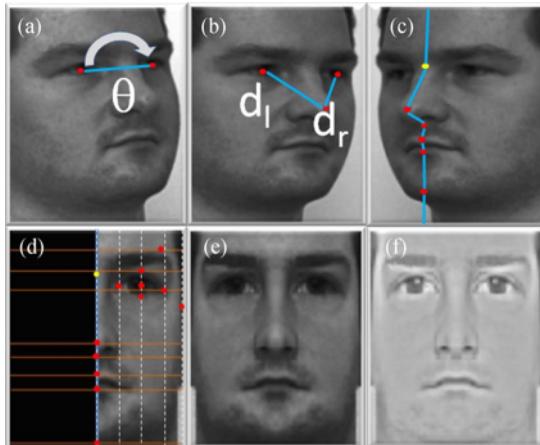
$$roll = \min\left(\left|\frac{2\theta}{\pi}\right|, 1\right) \quad (4.3)$$

¹Backpropagation

²Maximum Likelihood

³Extended Active Shape Model

⁴Simple Pose



شکل ۱۱.۳: ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۹].

انحراف: چرخش در راستای محور افقی است و d_l و d_r فاصله مرکز چشم چپ و راست از نوک بینی می باشد. اندازه گیری این فاصله

ها در صورت برابر بودن، برای تشخیص تمام رخ بودن تصویر چهره مورد استفاده قرار می گیرد.

$$yaw = \frac{\max(d_l, d_r) - \min(d_l, d_r)}{\max(d_l, d_r)} \quad (5.3)$$

شیب: چرخش سر در راستای محور عمودی را اندازه گیری می کند.

$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)} \quad (6.3)$$

با محاسبه ۳ شاخص فوق، شاخص SP محاسبه می شود:

$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch) \quad (7.3)$$

که در آن

$$\alpha + \beta + \gamma = 1 \quad (8.3)$$

که مقادیر این ضرایب از طریق آزمون و خطابه 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

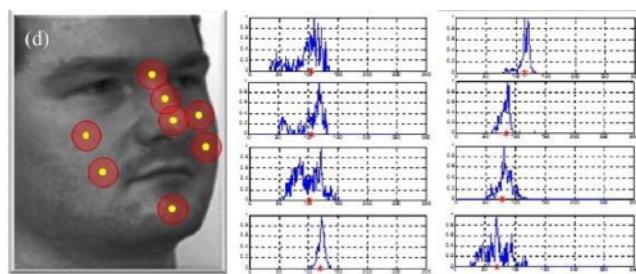
همیشه نیمه سمت راست تصویر پردازش شود. سپس برای ثابت کردن طول سطراها، سطراها بسط داده می شوند. مطابق شکل ۱۱-۳

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 [۲۷] یک روش هنجار سازی نورپردازی برای تصاویر چهره ارائه دادند و از این روش برای محاسبه شاخص روشنایی نمونه (SI)^۱ استفاده کردند. زمانی که تصویر روشنایی یکنواخت دارد، بیشتر بخش های چهره توزیع یکنواخت سطح خاکستری دارند. اما وقتی روشنایی یکنواخت نباشد، برخی از نواحی خاص چهره، توزیع یکنواخت سطح خاکستری ندارند. برای مثال جلوی بینی، گونه ها و چانه معمولاً نور را منعکس می کنند. ۸ ناحیه در شکل ۱۲-۳ با توجه به چنین اصلی انتخاب شده اند. ۸ بافت نگار بارنگ آبی و مرکز آن ها بارنگ قرمز مشاهده می شود.



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

بافت نگار فوق به یک توزیع یکنواخت با انحراف معیار کم در مرکز حجم بافت نگار اشاره دارد. بافت-نگار هر یک از ناحیه ها پدست آمده و مرکز ثقل آن محاسبه می شود:

$$mc(w) = \frac{\sum_{i=0}^{255} i \times h_w(i)}{\sum_{i=0}^{255} h_w(i)} \quad (9.3)$$

که در آن w نشان دهنده یکی از نواحی ۸ گانه می باشد. ۸ مرکز جرم محاسبه شده، بردار mc را تشکیل می دهند. با توجه به فرض تشابه ذکر شده در میان نواحی صورت در نظر گرفته شده، انتظار می رود هیچ تنوع قابل توجهی در میان عناصر بردار وجود نداشته باشد و توزیع های یکسانی از سطوح خاکستری را نمایش دهند. برای دستیابی به این منظور پراکندگی مرکز حجم ها از ۸ نمودار بافت نگار محاسبه شده است. سپس عناصر بردار mc توسط تابع سیگموید F در بازه $[1, 0]$ هنجار سازی می شوند و شاخص کیفیت

¹Sample Illumination

روشنایی محاسبه می شود که یک عدد می باشد:

$$SI = 1 - F(std(mc)) \quad (10.3)$$

هرچه مقدار SI بیشتر باشد، یعنی تصویر روشنایی یکنواخت تری دارد. اگر این شاخص به اندازه کافی رضایت بخش نباشد، تصویر رد می شود. در غیر این صورت برای هنجارسازی روشنایی وارد بخش بعدی خواهد شد. در صورت رد شدن تصویر، سیاست های 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 می تواند 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 می شود و نتیجه نهایی حاصل می شود. نتیجه به صورت قسمت f در شکل می باشد. در سال ۲۰۱۵ Jamal Hussain Shah و همکاران در [۳۷] رویکردی برای تشخیص چهره در تغییرات شدید روشنایی پیشنهاد دادند که به سه مرحله تقسیم شده است:

۱. برای اصلاح روشنایی غیر یکنواخت، همسان سازی بافت نگار براساس بر اساس ناحیه استفاده می شود.
۲. ویژگی های مبتنی بر LDA از تصویر چهره استخراج می شود.
۳. فرایند طبقه بندی بر اساس مدل OPPM انجام می شود.

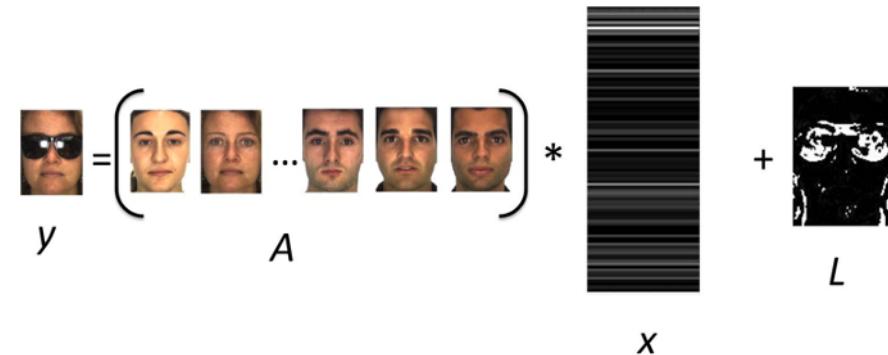
۴.۳ چالش انسداد

در سال ۲۰۱۸ Cho Ying Wu و همکاران در [۳۹] یک رویکرد مبتنی بر واپاش با جهت گرادیان برای شناسایی چهره های در معرض انسداد ارائه دادند. در کاربردهای واقعی، تعداد داده های آموزش بسیار کم می باشد (شاید یک تصویر به ازای هر شخص). این رویکرد توانایی برخورد با این شرایط را دارد و در مقابل تصاویری که نزدیک به ۸۰ درصد از چهره در شرایط انسداد قرار دارد، به خوبی عمل می کند. نتایج نشان می دهد که با تعداد بسیار کمی از تصاویر آموزشی، مدل پیشنهاد شده GD-HASLR بهترین عملکرد را در مقایسه با سایر روش های پیشرفتی، از جمله روش های مبتنی بر شبکه عصبی پیچشی دارد. مجموعه داده آموزشی $A = \mathbb{R}^{d \times n}$ در نظر گرفته شده که در آن n تعداد داده های آموزشی و d حاصل ضرب تعداد پیکسل های طول و عرض تصاویر می باشد. داده های آموزشی چهره های طبیعی و بدون انسداد می باشند. $\mathbb{R}^d =$ یک داده آزمایشی می باشد. می توان از یک ترکیب خطی داده های

آموزش برای تخمین زدن داده آزمایش استفاده کرد که شامل یک عبارت خطای $L = \mathbb{R}^d$ نیز می باشد. (شکل ۱۶-۳)

$$y = Ax + L \quad (11.3)$$

که در آن x بردار ضرایب با n بعد می باشد. برای آنکه شرط تنک بودن به رابطه بالا اضافه شود، مسئله به صورت زیر نوشته می شود:



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 [?].

$$\operatorname{argmin}_{x \in \mathbb{R}^n} s.t. y - Ax \quad (12.3)$$

که در آن ϵ یک آستانه خطای می 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.

$$L(x, L, z) = \alpha LM + xi + zT(y - Ax - L) + 2y - Ax - L \quad (13.3)$$

که در آن Z ضریب لاغرانژ و عامل مجازات می باشد. پس از بدست آوردن بردار تنک x می توان باقیمانده دسته $\hat{\alpha}$ را به صورت زیر

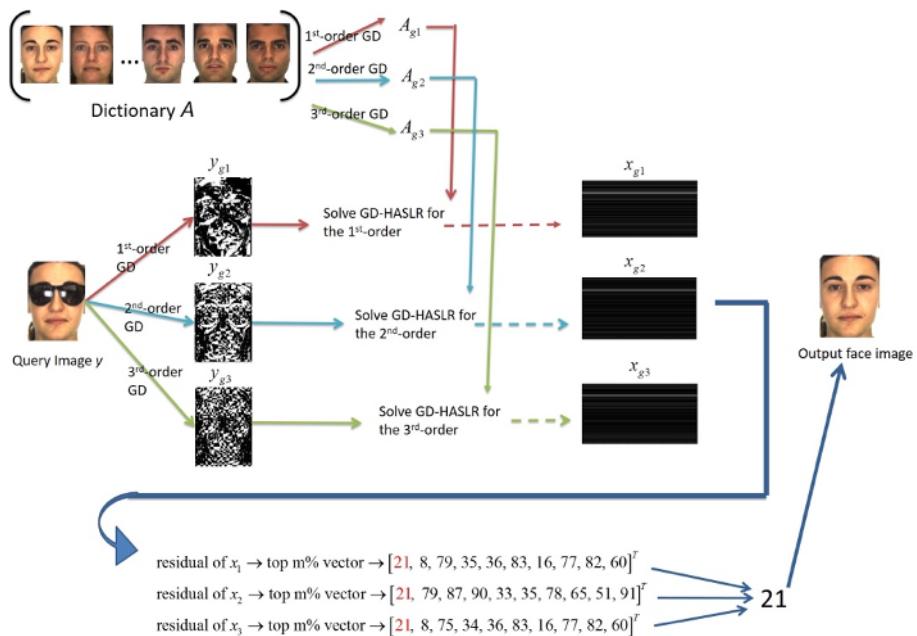
محاسبه نمود:

$$r_i = y - Ai(x) \quad (14.3)$$

که در آن $(x, \hat{\alpha})$ نشان دهنده $\hat{\alpha}$ امین انتخاب کننده دسته می باشد که فقط ورودی های مربوط به دسته $\hat{\alpha}$ 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. در سال ۲۰۱۴ J. Li و همکاران در [۳۸] یک روش تشخیص چهره پوشیده شده در پس زمینه پیچیده ارائه



شکل ۱۴.۳: رویکرد کلی الگوریتم [۹] GD-HASLR

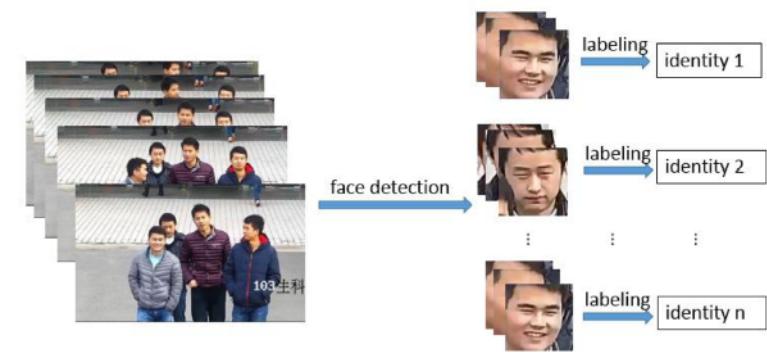
کردن. این الگوریتم از دو مرحله تشکیل شده است. در مرحله اول تعیین می کنند که آیا شی یک شخص می باشد یا خیر و در مرحله دوم بررسی می شود که آیا چهره پوشیده شده می باشد یا خیر و در صورت پوشش چهره، نوع پوشش و اینکه پوشیدگی با ماسک، کلاه، عینک یا ... است را مشخص می کند. در مرحله اول یک رویکرد تشخیص شی در پیش 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 استفاده 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 نمی باشد و ساختار سه بعدی و پیچیده ای دارد و ممکن است تصویر از هر زاویه ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده های آموزشی بسیاری می باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز بر این است که داده های آموزش بسیار کم هستند. از این رو مسئله تشخیص چهره باید در شرایطی حل شود که داده های آموزشی کافی در اختیار نمی باشد. بنابراین به الگوریتمی نیاز داریم که به ما کمک کند با تولید داده های غیر واقعی، مشکل کمبود داده های

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 Wang Ya و همکاران در [۴۰] 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 بهبود 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 طبقه‌بندی می‌شود. برخی تصاویر در هر دسته به اشتباه در مرحله اول به عنوان



شکل ۱۵.۳: ردیابی، یافتن چهره‌ها و برچسب زنی [۴].

تصویر یک فرد در نظر گرفته شده اند (شکل ۱۶-۳). با استفاده از روش خوش‌بندی گراف^۱ روی ویژگی‌های استخراج شده از شبکه

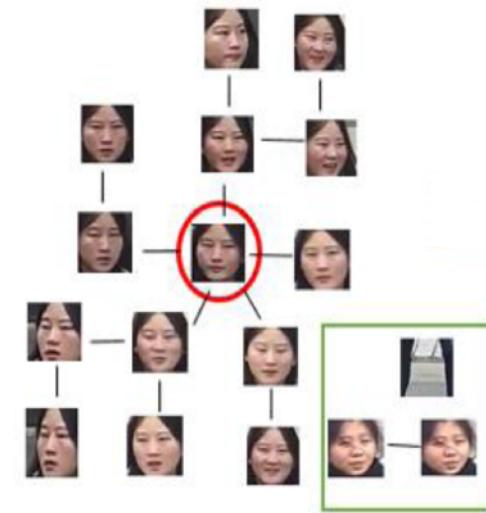


شکل ۱۶.۳: تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند [۴].

VGG-Face، تشخیص و پاک سازی تصاویر اشتباه انجام می‌شود. فاصله کسینوسی بین ویژگی‌های تصاویر چهره محاسبه می‌شود 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

^۱Graph Clustering

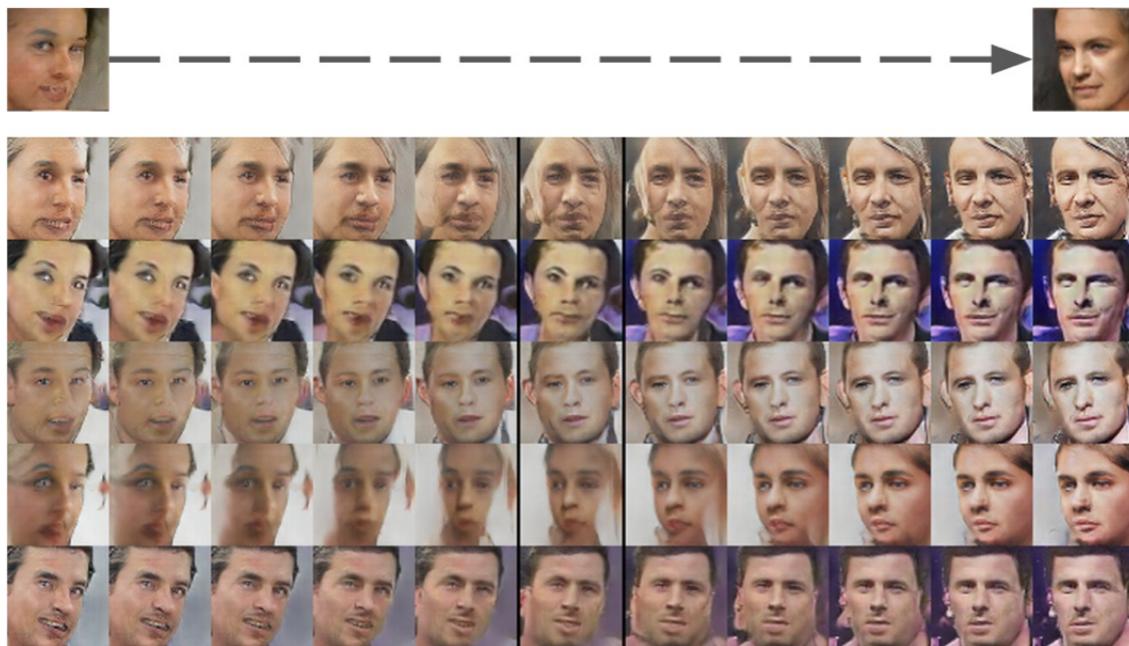
تصویری که بیشترین شباهت را به تصاویر دیگر دارد، به عنوان شاخص برای آن شخص انتخاب می‌شود. با استفاده از محاسبه فاصله



شکل ۱۷.۳: استفاده از روش خوشه‌بندی گراف و تعیین تصویر شاخص [؟].

بین هر داده با داده مرکزی و در نظر گرفتن یک آستانه، داده‌های تکراری در هر دسته مشخص شده و حذف می‌شوند. با توجه به مقدار داده‌های درون هر دسته، تصفیه بین دسته ای انجام می‌شود. اگر مجموعه داده‌های درون هر دسته کمتر از ۱۰۰ تصویر باشد، آن دسته از مجموعه داده حذف می‌شود. دقیق خوشه‌بندی و جمع آوری مجموعه داده ۲٪.۹۹ شده است. در نهایت از یک مدل پیش آموزش دیده شده شبکه VGG-Face برای طبقه‌بندی تصاویر آزمایشی استفاده شده است که به دقیق ۹۶٪ رسیده است.

مقاله [۴۵] از شبکه‌های مولد تخصصی برای تولید داده‌ها استفاده کرده است که به اختصار GAN نامیده می‌شوند. GAN از دو شبکه مستقل تولید کننده و تمیز دهنده استفاده تشكیل شده است. شبکه تولید کننده از روی بردار Z که می‌تواند یک نویز تصادفی باشد، یک تصویر تولید می‌کند و شبکه تمیز دهنده تصاویر واقعی را از تصاویر تولید شده توسط شبکه تولید کننده تشخیص دهد. بنابراین هر تصویر با یک بردار Z معرفی می‌شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می‌شود. به عنوان مثال بردار Z برای تصویر خانمی که عینک آفتابی نزدی است از بردار Z برای تصویر خانمی که عینک آفتابی زده است، کم می‌شود و حاصل آن، بردار مربوط به یک عینک آفتابی می‌باشد. سپس این بردار با بردار تصویر آقایی که عینک نزدی 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.



شکل ۱۸.۳: تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۴].

مقاله [۴۶] 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 شانون اصلاح شده (JSD) و تنوع متقابل در اطلاعات (MVI). شبکه تمیز دهنده در LR-GAN، بر اساس اطلاعات طبقه بندی که به طور ضمنی در طول آموزش آموخته می‌شود، هویت هر شخص را حفظ می‌کند. این رویکرد سریعتر از شبکه‌های مبتنی بر GAN اخیر به یک همگرایی می‌رسد. این مدل که به دقت بالای ۹۰٪ رسیده است، رتبه اول را در ۴ مجموعه داده شرایط بدون محدودیت کسب کرده است.

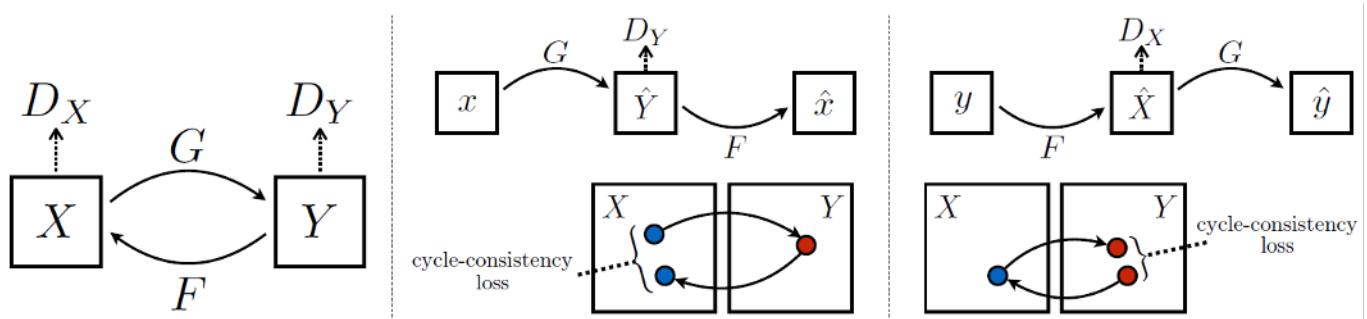
شکل ۳ تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین در ۴ مجموعه داده مختلف. موارد با حاشیه قرمز خروجی اشتباه هستند - [۶] شبکه‌های GAN یاد می‌گیرند تصاویر جدیدی تولید کنند که شبیه به تصاویر واقعی باشند. اما این شبکه‌ها معمولاً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند.

مقاله [۴۷] 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، مامی‌توانیم 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.

شکل ۳ - تصاویر هر دوستون دارای برخی ویژگی‌های دیداری مشابه هستند - [۷]

در مقاله [۴۸] 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 روبرو تولید شده است. سپس با استفاده از شبکه GAN دوم از روی تصویر چهره روبرو، تصویر با زاویه دلخواه با استفاده از یک پارامتر کنترلی تولید می‌شود. چالشی که در این مقاله به آن اشاره شده است، مسئله عدم توان اداده‌ها 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 روبرو 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 مواجه 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.



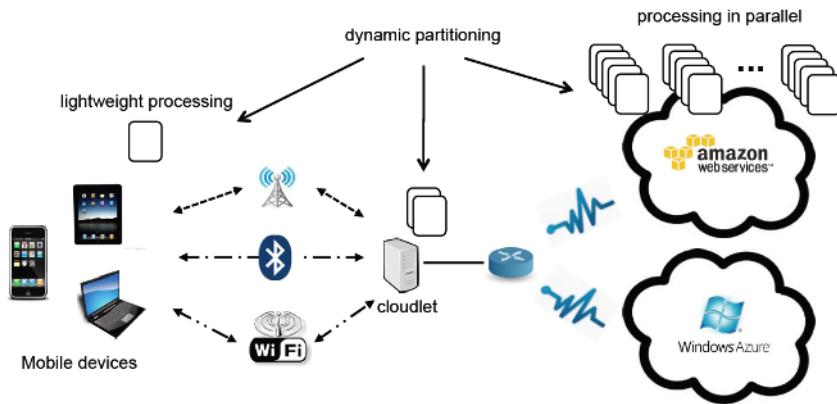
شکل ۱۹.۳: ساختار شبکه AD-GAN - شبکه GE برای تولید چهره از زوایای مختلف [۷].

۶.۳ چالش منابع محدود

در سال ۲۰۱۲ Tolga Soyata و همکاران در [۴۱] یک روش تشخیص چهره بی‌درنگ مبتنی بر بینایی ابری^۱ با استفاده از معماری MOCHA ارائه کردند (شکل ۱۸-۳). 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

¹Cloud Vision

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 می باشد. نتایج نشان می دهد که الگوریتم های بخش بندی بهینه پردازش بین تلفن همراه و فضای ابری با توجه به زمان تأخیر ناهمگن، توانایی محاسبه را به طور قابل توجهی افزایش می دهند. این سامانه از لحاظ ساختار به سه بخش تقسیم می

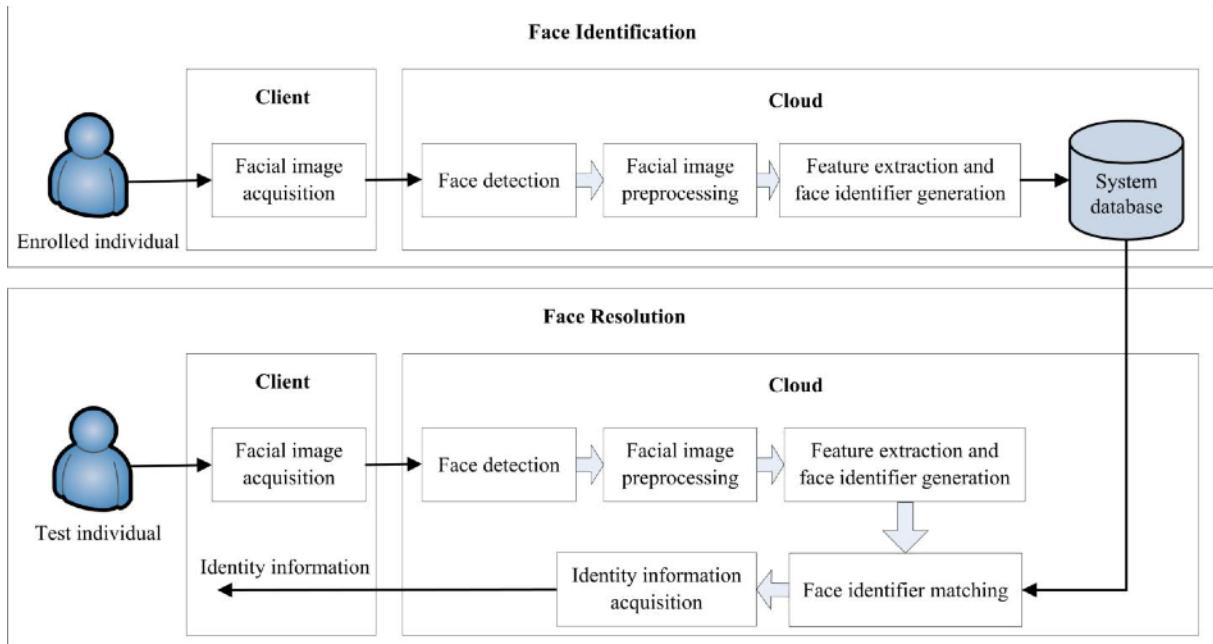


شکل ۲۰.۳: معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [؟].

شود: دستگاه همراه: تلفن های همراه و iPad ها نقش تهیه و ارسال تصاویر را دارند. تصاویر با فرمت RAW فرستاده می شوند 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 می شوند. در غیر این صورت مرحله پیش پردازش در دستگاه همراه انجام می شود و فقط اطلاعاتی همچون ویژگی های Haar و طبقه بندها به سرور فرستاده می شوند. پس از اتمام فرایند تشخیص چهره، نتیجه نهایی برای تلفن همراه فرستاده 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 باز محاسبات سنتگین سامانه را به دوش می کشند و تصمیم گیری نهایی بر روی آن انجام می پذیرد. در سال ۲۰۱۸ Pengfei Hu [۴۲] یک رویکرد تشخیص چهره مبتنی بر رایانش ابری ارائه کردند. افزایش برنامه های کاربردی در زمینه کلان داده ها 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 عمل امکان-پذیر است و می تواند سرویس شناسایی موثر چهره را فراهم کند. همانطور که در شکل ۱۹-۳ مشاهده می شود، تنها تهیه تصویر بر عهده دستگاه سرویس گیرنده می باشد و تمام محاسبات یافتن و شناسایی چهره بر روی

¹Big Data

ابر انجام می شود.



شکل ۲۱.۳: نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۴].

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^۱ 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 ۲۰-۳ مشاهده می شود این سامانه ابری از بخش های سور

مدیریت (MS)^۲، سورور اطلاعات (IS)^۳، سورور شناسایی (RS)^۴ و پایگاه داده تشکیل شده است. به علت قدرت بالای پردازش در سور

ابری، امکان 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 این داشته اند که یک راه حل نسبتاً همه جانبه در مورد مسئله تشخیص چهره و مشکلات آن ارائه دهند. یکی از این رویکردها،

استفاده از تابع ضرر CosFace^۵ می باشد که در سال ۲۰۱۸ Wang و همکاران در [۴۹] ارائه دادند. این تابع ضرر کسینوسی با حاشیه

زیاد^۶ شباهت بسیار زیادی به تابع ضرر softmax دارد. با این تفاوت که به جای ضرب ماتریس ضرب های W در بردار ویژگی x،

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

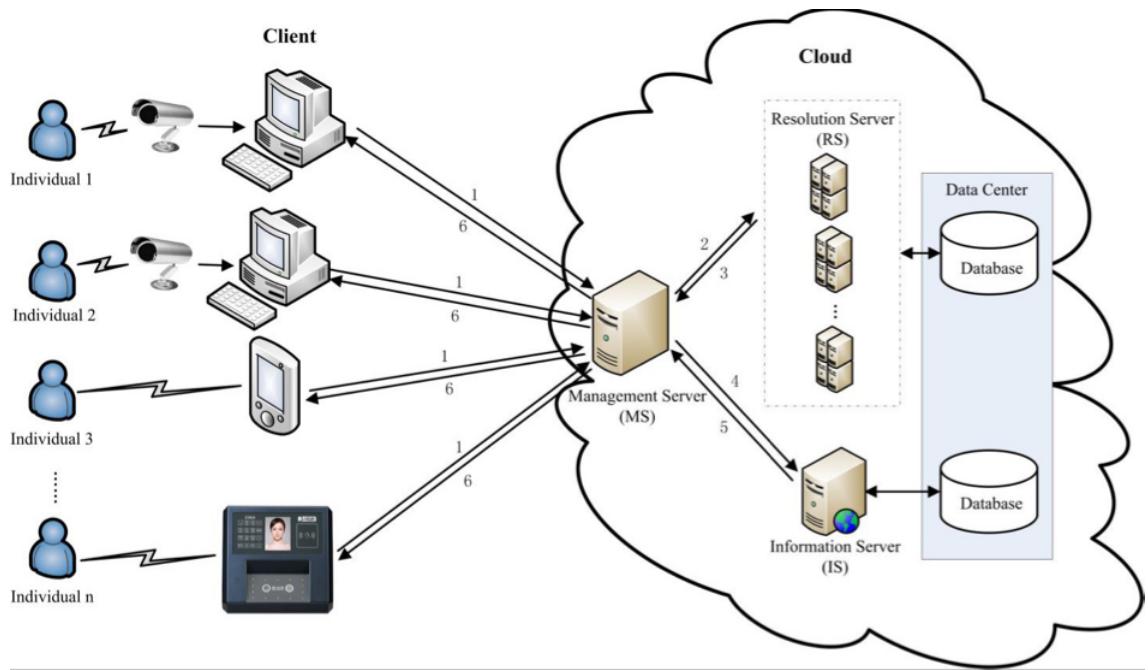
¹Local Binary Patterns

²Management Server

³Information Server

⁴Resolution Server

⁵Large Margin Cosine Loss

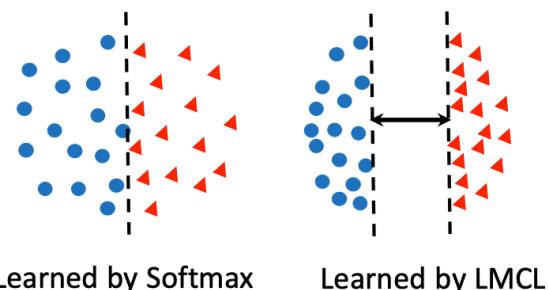


شکل ۲۲.۳: چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۹].

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

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

همانطور که در شکل ۲۳.۳ 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 معرفی شده می تواند فاصله بیشتری را بین دسته های نزدیک اعمال کند.



شکل ۲۳.۳: تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مرز بین دسته ها ایجاد می نماید [۴۹].

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 (blurring)
- وضوح پایین تصویر (low resolution)
- ردیابی چهره در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی (face tracking)

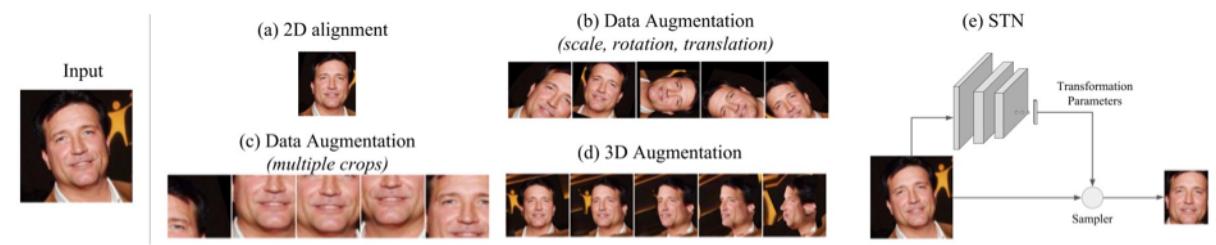
دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چشم نیمی باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده‌های آموزشی بسیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویربرداری شده باشد. مقاله [۲۴] روشنی برای رو به رو سازی تصویر چهره پیشنهاد کرده بود که در برخی موارد، چهره را به خوبی می‌چرخاند، اما در نیمی از موقع نیز نتیجه خروجی الگوریتم، تصویر چهره را دچار اعوجاج‌هایی می‌نماید که روند تشخیص چهره را با مشکل بیشتری مواجه می‌سازد. از این رو فرایند رو به رو سازی به طور میانگین کمک شایانی به بالا رفتن دقت تشخیص چهره نمی‌نماید. مقاله [۴۵] روشنی مبتنی بر GAN برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی مواقع به تصویر چهره لطمه وارد می‌نماید به طوری که شخص مورد نظر قابل شناسایی توسط سامانه یادگیری نمی‌باشد. مقاله [۴۶] 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 [۴۸] چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره روبرو بدون عارضه تولید شده است. این چالش به خاطر کمود تصاویر با عارضه پوستی در مجموعه داده می‌باشد و شبکه در مواجهه با این مسئله راه کاری ارائه نمی‌دهد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند. به تازگی یادگیری عمیق در تشخیص چهره و بسیاری از زمینه‌های هوش مصنوعی به راه حل غالب تبدیل شده است. ما یک سوال مطرح می‌کنیم: آیا یادگیری عمیق واقعاً مسئله تشخیص چهره را حل می‌کند؟ چالش روش‌های یادگیری عمیق در تشخیص چهره چیست؟ در مقایسه با تشخیص شیء عمومی، تشخیص چهره به دلیل طیف گسترده‌ای از تغییرات در ظاهر چهره‌ها چالش برانگیز است. نورپردازی کنترل نشده، انسداد ناشی از عینک، مو، ریش، کلاه و...، تاری خارج از تمرکز دوربین، کیفیت پایین تصویر، بالا رفتن سن افراد و کمبود داده‌های آموزشی از مواردی می‌باشد که می‌توانند سامانه تشخیص چهره را مشکل روبرو نمایند. از طرفی اکثر مجموعه داده‌ها تنها شامل چند هزار عکس می‌باشد. یک مجموعه داده حاوی اطلاعات بدون محدودیت و مقیاس بزرگ، سامانه چارچوب چهره را به چالش‌هایی همچون گرایش‌های شدید، نور کم و تصاویر کوچک و تاریک چهره تبدیل می‌کند. محققان فرض کرده‌اند که لایه‌های عمیق CNN‌ها می‌توانند اطلاعات انتزاعی بیشتری مانند هویت، ظاهر و ویژگی‌ها را رمزگذاری کنند؛ با این حال هنوز کاملاً مطالعه نشده است که لایه‌ها دقیقاً با ویژگی‌های محلی برای تشخیص مطابقت دارند. برای شناسایی چهره، عملکرد یادگیری را می‌توان با یادگیری یک معیار اندازه‌گیری فاصله متمایز کننده بهبود داد. با این حال، با توجه به محدودیت‌های حافظه کارت گرافیک ها، نحوه انتخاب جفت‌های این سه گانه‌های اطلاعاتی و روش‌های آموزش آنلاین (به عنوان مثال، گرادیان نزولی) در مجموعه داده‌های بزرگ، هنوز یک مشکل باز است. یکی دیگر از مشکلات چالش برانگیز این است که پردازش ویدیو در شبکه‌های عمیق را برای استفاده از تجزیه و تحلیل چهره مبتنی بر ویدئو ترکیب کند.

فصل ۴

روش پیشنهادی

همان طور که در فصل گذشته شرح داده شد، شناسایی چهره در محیط بدون محدودیت به صورت بی‌درنگ و با دقت بالا با چالش‌های بسیاری همراه است. همچنین دقت بالا و زمان پردازش پایین 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 رنگی 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.



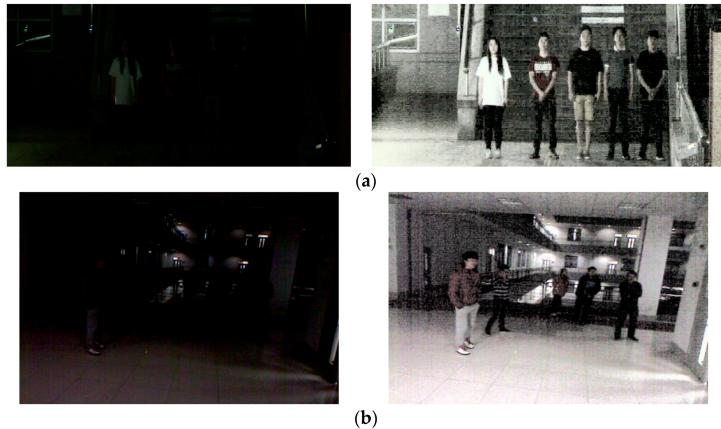
شکل ۱.۴: نمای کلی از روش پیشنهادی [۹].

۲.۴ پیش‌پردازش

بیشتر الگوریتم‌های شناسایی چهره نیاز به اعمال پیش‌پردازش‌هایی بر روی تصویر ورودی دارند. 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 همسان‌سازی بافت‌نگار

یکی از روش‌های بھبود تصویر، بھبود تباين تصویر است. یکی از روش‌های افزایش تباين تصویر، تکنیک یکتواخت 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 بھبود کیفیت تصویر و افزایش دقت پردازش‌های بعدی است. نمونه‌ای از این روش را از مقاله [۵۰] در شکل ۲.۴ مشاهده می‌نمایید.



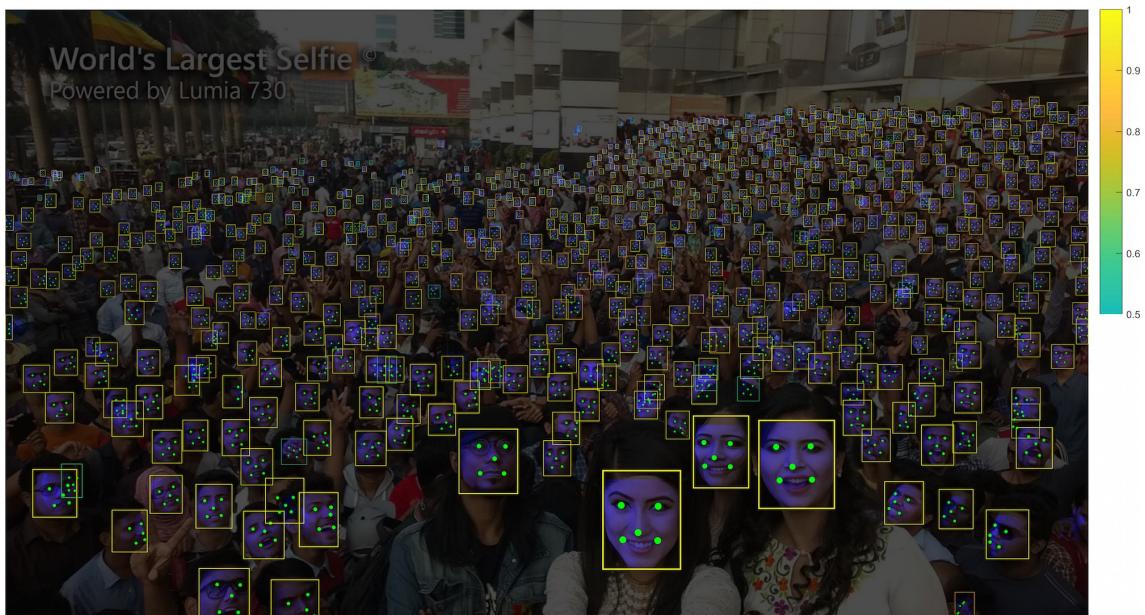
شکل ۲.۴: نتیجه اعمال یکسان سازی بافت‌نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می‌باشند [۵۰].

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 [۵۱] در سال ۲۰۱۹ معرفی شده است استفاده می‌کنیم. در این روش برای آموزش شبکه پیچشی از یکتابع ضرر مبتنی بر یادگیری چندکاره استفاده شده است که به صورت زیر می‌باشد.

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

L_{cls} که در آن تابع ضرر مربوط به یافتن یا عدم یافتن چهره می‌باشد. L_{box} تابع ضرر مربوط به مکان چهره می‌باشد. همچنین L_{pts} تابع ضرر مربوط به یافتن نقاط ویژه روی اجزای چهره می‌باشد، و L_{pixel} تابع ضرر مربوط به یافتن یک مدل سه بعدی مبتنی بر مشاهده سازی کرده و آموزش دادیم. نمونه‌ای از خروجی این روش را در شکل ۳.۴ مشاهده می‌کنید.

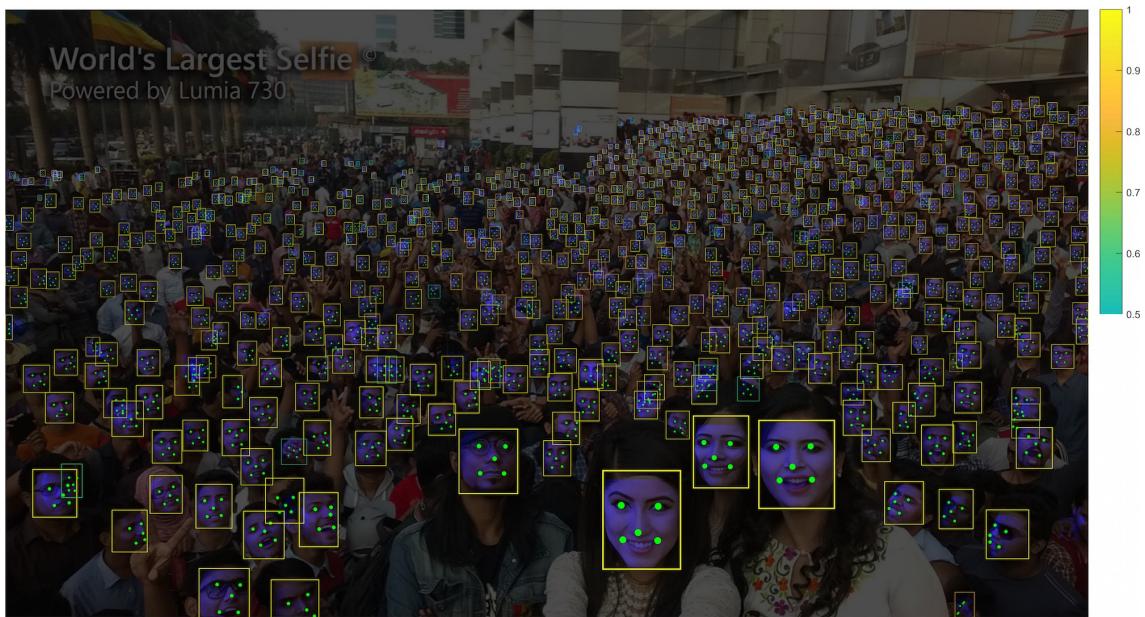


شکل ۳.۴: نمونه از خروجی الگوریتم یافتن چهره retina [۵۱]

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 دفت نهایی مشهود است. بدین منظور با استفاده از یک تبدیل غیرخطی، تصویر چهره را به گونه‌ای بپردازیم 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 هم‌تراز کردن تصاویر چهره با یک تصویر مرجع، پیش از اعمال طبقه‌بند می‌توان تا حدی این مشکل را بهبود داد. در طول هم‌تراز کردن تصویر، نقاط خاصی از تصویر (مانند نقطه‌ی 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.

^۱Face Alignment



شکل ۴.۴: نمونه از خروجی الگوریتم یافتن چهره retina [۵۱]

۳.۴ دسته بندی

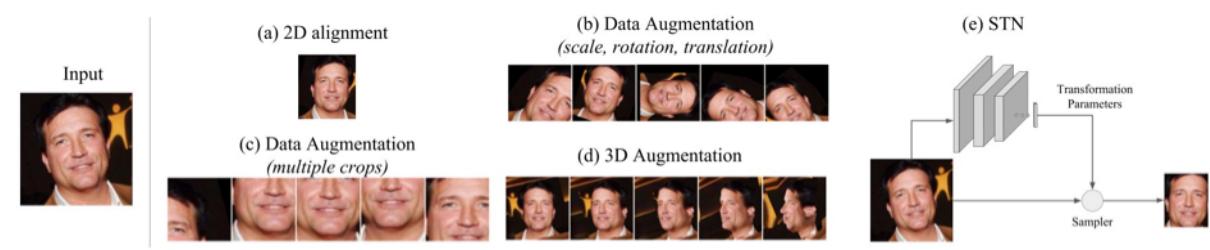
روش پیشنهادی برای تشخیص چهره به صورت بی درنگ در محیط های بدون محدودیت، یک الگوریتم مبتنی بر شبکه های پیچشی می باشد. این فرایند را می توان به صورت یک مسئله بهینه سازی فرموله کرد که در ادامه به بخش های مختلف آن می پردازیم.

۱.۳.۴ مدل پیشنهادی پایه

ابتدا نوبت آن است که معماری مناسب شبکه برای مسئله را به دست آوریم. با بررسی شبکه های به روز از جمله MobileNetV2، EfficientNetB0، ResNet-50، VGG19، SqueezeNet، NASNetMobile، MobileNetV3 و زمان پاسخگویی آن ها به کمک یادگیری انتقال، به این نتیجه می رسیم که شبکه های MobileNetV3 و SqueezeNet دارای چگالی دقت بالاتری می باشند و نسبت دقت دسته بندی به تعداد پارامترهای شبکه در آن ها بیشتر می باشد. بنابرین می تواند سرعت اجرای مناسب و همچنین دقت مناسب را از آن ها انتظار داشت. نتایج بررسی را در شکل زیر آمده اند.

آزمایش هایی بر روی معماری های مطرح دیگر نیز انجام شد که به علت ضعیف بودن نتایج یا بالا بودن زمان پاسخ دهی در شکل درج نشده اند. از نتایج در می یابیم که بهترین معماری شبکه برای مسئله ما معماری MobileNetV3 است. در این معماری مسیر استخراج ویژگی از پنج لایه کانولوشن، چهار لایه خطی و نرمال ساز چهار لایه دیکانولوشن^۳ و شش مازول تشکیل شده است، که

در ادامه توضیح داده خواهد شد، تصویر ورودی I 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 از لایههای مختلف با یکدیگر ترکیب شده و به عنوان ورودی تقسیم‌بند مورد استفاده قرار میگیرد [۱۴]. لایه کانولوشن سه لایه کانولوشن همراه با گام دو، اندازه حجم ورودی را با ضریبی از دو کاهش میدهد. سایز پنجره فیلترها 3×3 میباشد. دلیل انتخاب سایز کوچک پنجره فیلترها کاهش بیچیدگی محاسباتی و همچنین عملکرد خوب آنها در استخراج ویژگی میباشد. چگونگی عملکرد یک لایه کانولوشن از رابطه $3 \cdot 2$ بدست میآید.



شکل ۴: رویکردهای مختلف هم ترازی چهره [۴].

۲.۳.۴ آموزش شبکه جهت استخراج ویژگی

از آن جایی که شبکه MobileNet معماری بسیار سبک تری نسبت به معماری های شناخته شده دیگر که در فصل ۲ معرفی شدند، دارد؛ بنابراین استخراج ویژگی ها از چهره و دسته بندی تصاویر چهره با دقت بالا برای این شبکه بسیار سخت و دشوار است. همچنین در مواردی شباهت چهره افراد به یکدیگر کار را از آن چه هست سخت‌تر خواهد کرد. بنابراین ما روشی برای استخراج ویژگی از تصاویر چهره پیشنهاد کرده‌ایم که 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؛ تا این طریق بتوان به کاهش مشکلات ذکر شده کمک کرد. این روش شامل بخش‌های تولید چهره، یافتن چهره، آموزش شبکه عصبی پیچشی و استخراج ویژگی می‌باشد.

همانطور که پیش‌تر بیان شد به آموزش این شبکه می‌پردازیم. مجموعه داده خود را پس از پیش‌پردازش و افزایش داده‌ها، آماده می‌کنیم. تعداد کل داده‌های آموزش برای این مرحله را به حدود ۳۲۰۰۰۰ رساندیم و آموزش را در ۳۰ دوره انجام داده‌ایم.

یکی از چالش های اصلی در یادگیری ویژگی ها با استفاده از شبکه های عصبی عمیق پیوسته (DCNN) برای شناسایی چهره در مقیاس بزرگ، طراحی تابع ضرر مناسب است که قدرت تفکیک را افزایش می دهد. هدف ما کم کردن فاصله بین ویژگی های عمیق و مرکز کلاس های آنها در فضای اقلیدسی برای دستیابی به فشردگی درون کلاسی بیشتر می باشد. ما یک تابع ضرر برای افزایش زاویه ای حاشیه برای دستیابی به ویژگی های بسیار متمایز برای تشخیص چهره پیشنهاد می کنیم که عملکرد بهتر برخوردار است و می توان آن را به راحتی با هزینه های محاسباتی ناچیز پیاده سازی کرد. برای آموزش شبکه های عصبی عمیق پیوسته برای تشخیص چهره، دو رویکرد اصلی وجود دارد. روش اول دسته بندی را آموزش می دهنده که می تواند هویت های مختلف را در مجموعه آموزش از هم جدا کند، مانند با استفاده از طبقه بندی softmax، و رویکرد دوم که مستقیماً یک تعییه 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 :

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

۲. ویژگی های آموخته شده برای مسئله های طبقه بندی با مجموعه بسته قابل تفکیک هستند اما به اندازه کافی برای مسئله تشخیص چهره که یک مسئله باز می باشد، مناسب نیستند.

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.

ما برای افزایش بیشتر قدرت تمایز مدل تشخیص چهره و ایجاد ثبات در روند آموزش، تابع ضرر مبتتنی بر توابع مثلثاتی را پیشنهاد می کنیم. همانطور که در شکل ۲ نشان داده 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، ما از تابع مثلثاتی کسینوسی برای محاسبه زاویه بین ویژگی فعلی و وزن هدف استفاده می کنیم. سپس یک حاشیه زاویه ای به زاویه هدف اضافه می کنیم، در انتها با استفاده از تابع کسینوس دوباره مقادیر را به فضای خطی برمی گردانیم. مراحل بعدی دقیقاً مانند softmax هستند. 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 می‌کند. پردازنهای گرافیکی کنونی می‌توانند به راحتی از هزاران دسته مختلف برای آموزش پشتیبانی کنند و مدل به راحتی می‌تواند هویت‌های بیشتری را پشتیبانی کند.

رابطه ریاضی softmax معروف ترینتابع ضرر طبقه‌بندی که به طور گسترده استفاده می‌شود، به شرح زیر است:

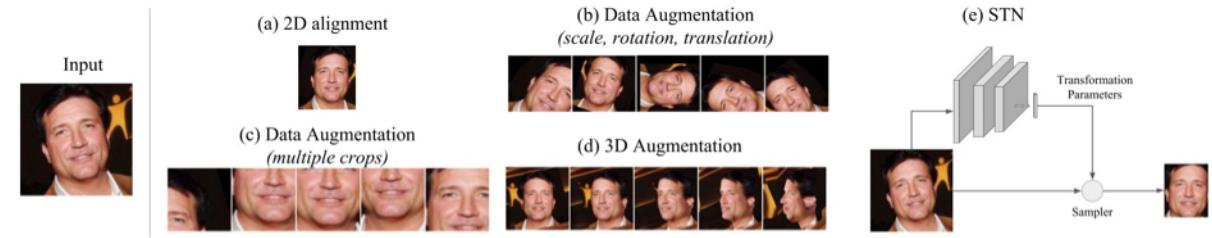
$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_j^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (2.4)$$

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 j از وزن W می‌باشد و b_j بایاس است. مقدار N اندازه دسته و n تعداد دسته‌ها است. این تابع مستقیماً ویژگی استخراج شده را برای اعمال شباهت بالاتر برای نمونه‌های درون کلاس و فاصله بیشتر برای نمونه‌های بین کلاسی بهینه نمی‌کند، که منجر به ایجاد مشکل در عملکرد آن برای تشخیص چهره عمیق تحت تغییرات ظاهری بزرگ درون کلاس می‌شود (به عنوان مثال تغییرات زاویه چهره و تغییرات سنی).

ما رابطه فوق را مبنای محاسبات قراردادیم و تغییرات جزئی به آن اضافه کردیم. برای سادگی مقدار بایاس را صفر در نظر گرفتیم. سپس حاصل ضرب مقادیر موجود در ویژگی‌های استخراج شده و آخرین لایه کامل متصل را به صورت (j) تبدیل می‌کنیم، که j زاویه بین وزن W_j و ویژگی x_i است. نرمال سازی ویژگی‌ها و وزن‌ها باعث می‌شود 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 می‌گیریم. همچنین ویژگی استخراج شده $|x_i|$ را نرمال کرده و نام آن را s در نظر می‌گیریم. بنابرین ویژگی‌های استخراج شده در یک ابر کره با شعاع s توزیع می‌شوند. برای افزایش حاشیه بین x_i و W_j یک مقدار m اضافه می‌کنیم تا به طور همزمان فشرده سازی درون کلاسی و اختلاف بین کلاسی را افزایش دهیم.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1}^n e^{s(\cos(\theta_j))}} \quad (3.4)$$

همانطور که در شکل ۳ نشان داده شده است، softmax ویژگی‌های تقریباً قابل تفکیکی ایجاد می‌کند اما در مرزهای تصمیم گیری ابهام قابل توجهی به وجود می‌آید، در حالی کهتابع ضرر ما می‌تواند فاصله بیشتری را بین دسته‌های نزدیک اعمال کند.

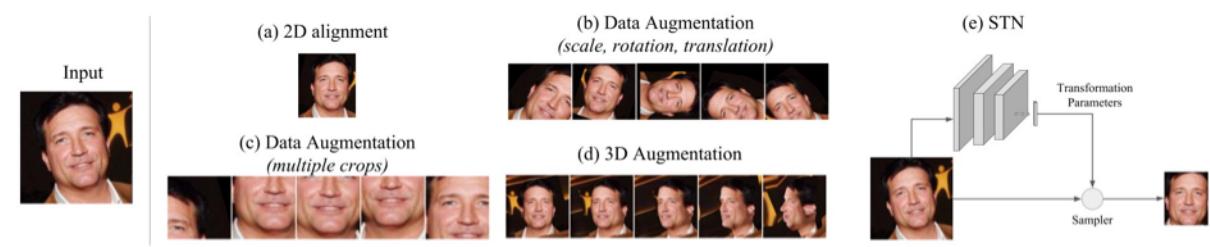


شکل ۶.۴: رویکردهای مختلف هم ترازی چهره [۹].

۱.۳.۳.۴ آموزش

در ابتدای روند آموزش، لازم است پارامترهای مدل مقدار دهی اولیه شوند و انتخاب پارامترهای اولیه میتواند تأثیر زیادی در مدل آموزش یافته داشته باشد. در این پژوهش به منظور مقدار دهی اولیه پارامترها ازتابع توزیع یکنواخت استفاده شده است. از دیگر چالشهای اساسی برای روشهای بهینه سازی مبتنی بر گرادیان، انتخاب میزان نرخ یادگیری مناسب است. روشهای کلاسیک گرادیان تصادفی از نرخ یادگیری ثابت یا کاهشی استفاده میکنند، که برای همه پارامترهای مدل یکسان است. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف میتوانند از نظر مقدار متفاوت باشند، که میتواند به نرخ یادگیری مختلفی نیاز داشته باشد. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف میتوانند از نظر مقدار تفاوت قابل توجهی داشته باشند، که میتواند به نرخ یادگیری مختلفی نیاز داشته باشد. در سالهای اخیر، تمايل به توسعه روشهایی برای انتخاب خودکار نرخ یادگیری مستقل افزایش یافته است. اکثر روشهای آمارهای مختلف مشتقات جزئی را در چندین تکرار عنوان مثال، AdaGrad [۴۷] و [۴۸] AdaDelta، RMSprop، Adam [۵۰] و [۴۹] آمده اند. 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 روشن Adam عملکرد بهتری ارائه داده است.

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 ۵۱۲ تایی بدست آمده را با بردارهای مربوط به چهره‌های بانک اطلاعاتی مقایسه کرده و با محاسبه فاصله اقلیدسی بردارهای 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.



شکل ۷.۴: رویکردهای مختلف هم ترازی چهره [۹].

۴.۴ فناوری‌های استفاده شده

پیاده‌سازی این الگوریتم به کمک زبان برنامه نویسی پایتون و کتابخانه OpenCV و PyTorch انجام شده است. از کتابخانه‌های NumPy برای انجام محاسبات ماتریسی و Scikit Learn و SciPy اشاره کرد. [۳۵] مهم مورد استفاده دیگر در این کار می‌توان به ۱۲ گیگابایت حافظه اصلی و ۱۲ گیگابایت حافظه گرافیکی در اختیار گرفتیم. 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 را برای آموزش و برخی را برای آزمون استفاده کرده ایم که به شرح آن ها می پردازیم.

۱.۳.۵ مجموعه داده های آموزش

۱.۱.۳.۵ CASIA Web-Face مجموعه داده

مجموعه داده Web-Face CASIA که شامل ۴۹۴۴۱۴ تصویر چهره متفاوت از ۱۰۵۷۵ فرد است، پاک و بدون نویز و اشتباه است. این مجموعه داده در مسائل تایید چهره و تشخیص چهره کاربرد دارد. [۵۲]

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. [۵۳]

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. [۵۴]

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. [۵۵]

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. [۵۶]

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 ۶۰۷۰ فریم و متوسط طول یک ویدیو ۳۰۱۸۱ فریم است. [۵۴]

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. [۵۵]

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. [۵۷]

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 [۵۸]. 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، تعداد دسته هایی که به شبکه برای آموزش داده می شود را

¹Cross-Age Celebrity Dataset

²Million-Scale Face Recognition Dataset

برابر ۱۶ قرار دادیم. توصیه می شود این مقدار توانی از ۲ باشد که مقدار ۱۶ با تجربه و توجه به ظرفیت حافظه پردازنده گرافیکی بدست آمده است. از بهینهساز Adam جهت آموزش استفاده کرده ایم. این بهینهساز نیز نرخ آموزش را بر اساس خطای صورت تطبیقی کم می کند. همچنین میانگین کاهش گرادیان های تکرارهای قبل را نگهداری می کند تا بر اساس آنها جهت گرادیان تکرار جدید را محاسبه کند. برای تابع ضرر از معادله ۳.۴ استفاده کرده ایم. این تابع میزان ضرر را به صورتی که در فصل قبل صحبت شد، محاسبه می کند.

همچنین برای به دست آوردن بهترین نتیجه از آموزش، از روش ارزیابی تقاطعی استفاده کرده ایم. به این صورت که در هر مرتبه آموزش، تعداد داده های آموزش را به پنج دسته تقسیم می کنیم. چهار قسمت را برای آموزش و یک قسمت را برای ارزیابی در نظر می گیریم. به این ترتیب پنج مدل شبکه برای آموزش خواهیم داشت و بهترین نتیجه را برای آزمون بر روی مجموعه داده های آزمون انتخاب می کنیم.

۵.۵ نتایج آزمون

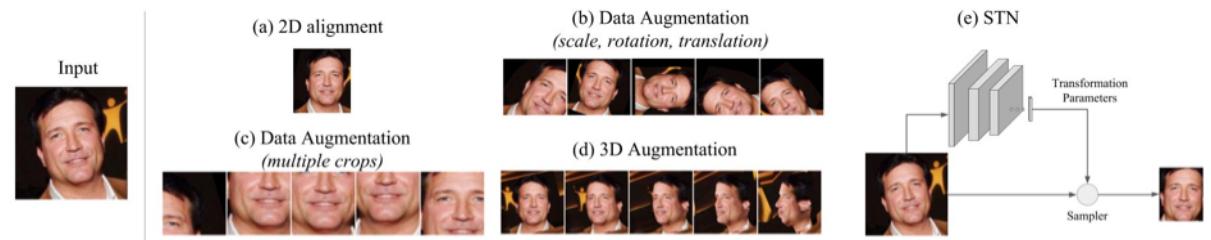
با توجه به پیکربندی بیان شده و همچنین عامل هایی که در فصل سوم توضیح داده شد، مدل را آموزش داده ایم و به سراغ مجموعه داده های آزمون می رویم. مقدار دقت را طبق رابطه ۱.۵ برای معماری های مرتبط و مشهور و همچنین روش پیشنهادی محاسبه کردیم. در جدول ۲.۵ نتایج حاصل از محاسبه دقت را در معماری های مختلف بر روی مجموعه داده داده های مشاهده می کنید.

جدول ۲.۵: دقت معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره

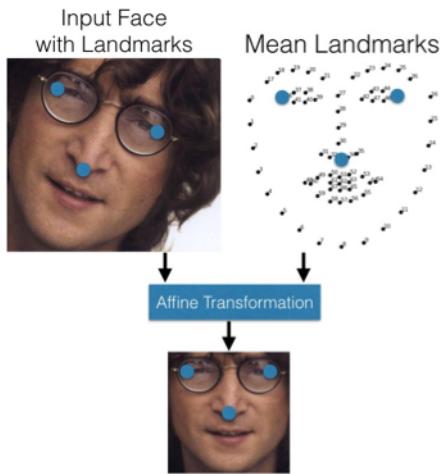
LFW	PubFig	YouTube Faces	CFP	CACD	MegaFace	معماری مدل	مجموعه داده آموزش
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	Web-Face CASIA
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	MS-Celeb-1M
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	

نمودار FROC که مقدار دقت و میزان خطای صورت فرایند آموزش شبکه نشان می دهد، برای مجموعه داده های Web-CASIA در شکل ۱.۵ و شکل ۲.۵ مشاهده می کنید.

با توجه به نتایجی که در جدول ها و شکل ها مشاهده می کنید متوجه می شویم که هر چند حساسیت روش پیشنهادی ما در تعدادی از مقادیر مثبت کاذب نسبت به موارد مشابه در دیگر مقاله ها کمتر شده است اما در کل معیار Fscore آن نسبت به بقیه بهتر شده است و



شکل ۱.۵: منحنی FROC مقایسه در مقالات برای مجموعه داده E-Ophtha-MA.



شکل ۲.۵: منحنی FROC مقایسه در مقالات برای مجموعه داده DIARETDB1.

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

نتیجہ گیری 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 ویژگی‌ما غنی‌تر خواهد شد.

همچنین می‌توان تابع ضرر را بروز رسانی و بهینه‌تر 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] 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.
- [2] E. Hjelmås and B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding*, vol.83, no.3, pp.236–274, 2001.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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*.
- [10] JafriRabia and A. R., “A survey of face recognition techniques,” *Journal of Information Processing Systems*, vol.5, no.2, pp.41–68, 2009.
- [11] 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.

- [12] 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.
- [13] H. Wang, Y. Wang, and Y. Cao, “Video-based face recognition: A survey,” *World Academy of Science, Engineering and Technology*, vol.60, pp.293–302, 2009.
- [14] X. Luan, B. Fang, L. Liu, W. Yang, and J. Qian, “Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion,” *Pattern Recognition*, vol.47, no.2, pp.495–508, 2014.
- [15] 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.
- [16] N. Khan, R. Ksantini, I. Ahmad, and B. Boufama, “A novel svm+nda model for classification with an application to face recognition,” *Pattern Recognition*, vol.45, no.1, pp.66–79, 2012.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [21] 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.
- [22] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, “Facial landmark detection with tweaked convolutional neural networks,” 2016.
- [23] 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.
- [24] M. Haghhighat, 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.
- [25] 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.
- [26] 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.

- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” 2014.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” *Pattern Recognition*, vol.66, pp.144–152, 2017.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.

- [41] 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.
- [42] 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.
- [43] L. Di Martino, J. Preciozzi, F. Lecumberry, and A. Fernández, “Face matching with an a contrario false detection control,” *Neurocomputing*, vol.173, 08 2015.
- [44] 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.
- [45] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2016.
- [46] S. Banerjee and S. Das, “Lr-gan for degraded face recognition,” *Pattern Recognition Letters*, vol.116, pp.246–253, 2018.
- [47] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” 2019.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” 2019.
- [52] “Casia web-face dataset,” <https://paperswithcode.com/dataset/casia-webface>. Accessed: 2021-04-24.
- [53] “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.
- [54] “Vggface2 dataset,” https://www.robots.ox.ac.uk/~vgg/data/vgg_face/. Accessed: 2021-04-24.
- [55] “Labeled faces in the wild dataset,” <http://vis-www.cs.umass.edu/lfw/>. Accessed: 2021-04-24.
- [56] “Public figures face database,” <https://www.cs.columbia.edu/CAVE/databases/pubfig/>. Accessed: 2021-04-24.

- [57] “Cross-age celebrity dataset,” <https://paperswithcode.com/dataset/cacd>. Accessed: 2021-04-24.
- [58] “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