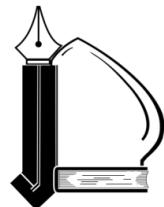


بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ





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

شناخته: ب/ک/3	<p>صورتجلسه دفاعیه پایان نامه دانشجوی دوره کارشناسی ارشد</p> <p>دانشجوی کارشناسی ارشد</p>	<p>دانشگاه فنی شهرضا مدیریت تحصیلات تکمیلی</p>
		جلسه دفاعیه پایان نامه تحصیلی آقای/خانم: سید سجاد اعمی رشته/گرایش: مهندسی کامپیوتر/هوش مصنوعی تحت عنوان: تشخیص بی درنگ چهره در محیط های بدون محدودیت و تعداد واحد: ۶ در تاریخ ۱۳۹۷/۰۷/۳۰ با حضور اعضای هیأت داوران (به شرح ذیل) تشکیل گردید. پس از ارزیابی توسط هیأت داوران، پایان نامه با نمره به عدد <b>و درجه</b> به حروف <b>مورد تصویب قرار گرفت.</b>
	<u>امضاء</u>	<u>نام و نام خانوادگی</u>
	دکتر	دکتر حمید رضا پور رضا
		استاد/ استادان راهنمای:
		استاد/ استادان مشاور:
	دکتر	متخصص و صاحب نظر داخلی:
		متخصص و صاحب نظر خارجی:
		ناینده تحصیلات تکمیلی دانشگاه (ناظر)
		نام و نام خانوادگی:
		امضاء:

## چکیده

در 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

ت	فهرست جداول
ث	فهرست تصاویر
ح	فهرست نمادها
۱	۱ مروری بر کارهای گذشته
۲	۱.۱ مقدمه . . . . .
۲	۲.۱ یافتن چهره . . . . .
۲	۱.۲.۱ رویکردهای مبتنی بر دانش . . . . .
۳	۲.۲.۱ رویکردهای مبتنی بر تطبیق کلیشه یا الگو . . . . .
۳	۳.۲.۱ رویکردهای مبتنی بر ویژگی . . . . .
۶	۴.۲.۱ رویکرد مبتنی بر عامل‌های آماری . . . . .
۱۲	۵.۲.۱ رویکردهای مبتنی بر تصویر . . . . .
۱۳	۳.۱ شناسایی چهره . . . . .
۱۴	۱.۳.۱ رویکردهای سنتی . . . . .
۱۴	۲.۳.۱ رویکرد مبتنی بر فیلتر گابور . . . . .
۱۴	۳.۳.۱ رویکردهای سه بعدی . . . . .
۱۵	۴.۳.۱ رویکردهای تجزیه و تحلیل بافت پوست . . . . .
۱۶	۵.۳.۱ رویکردهای مبتنی بر دوربین حرارتی . . . . .

۱۷	تشخیص چهره مبتنی بر ویدیو	۶.۳.۱
۱۷	رویکرد مبتنی بر چهره ویژه	۷.۳.۱
۱۹	رویکرد مبتنی بر ماشین بردار پشتیبان	۸.۳.۱
۲۱	رویکردهای مبتنی بر شبکه عصبی	۹.۳.۱
۲۵	رویکردهای مبتنی بر نقاط راهنمایی	۱۰.۳.۱
۲۸	جمع‌بندی	۴.۱
۲۹	مروری بر کارهای گذشته در شرایط کنترل نشده	۲
۳۰	مقدمه	۱.۲
۳۰	چالش حالت	۲.۲
۳۹	چالش روش‌نایابی	۳.۲
۴۱	چالش انسداد	۴.۲
۴۳	چالش کمبود تصاویر آموزشی	۵.۲
۴۷	چالش منابع محدود	۶.۲
۵۰	نتیجه‌گیری	۷.۲
۵۳	منابع و مأخذ	

## فهرست جداول

۲۰	..... مقایسه دقیق الگوریتم NDA + SVM با سایر رویکردهای مشابه . . . . .	۱.۱
۲۸	..... مقایسه شبکه عصبی موبایل نت با گوگل نت و VGG . . . . .	۲.۱

## فهرست تصاویر

۱.۱	رویکرد مبتنی بر تطبیق کلیشه [؟].	۳
۲.۱	استفاده از لبه یا ب های معروف برای استخراج ویژگی های مبتنی بر لبه از چهره [؟].	۴
۳.۱	اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می باشند [؟].	۵
۴.۱	نمونه هایی از مستطیل های ویژگی های Haar [؟].	۶
۵.۱	مستطیل های ویژگی های Haar متناسب با بخش های چهره می باشند [؟].	۷
۶.۱	هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [؟].	۷
۷.۱	یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [؟].	۸
۸.۱	نحوه مقدار دهی به پیکسل های تصویر یکپارچه [؟].	۸
۹.۱	بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [؟].	۹
۱۰.۱	بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [؟].	۹
۱۱.۱	سطح روشانایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روش به سمت تاریک برگزیده می شود [؟].	۱۰
۱۲.۱	برای یافتن چهره ها، بخش هایی از تصویر که به الگوی HOG شبیه تر است را پیدا می کنیم [؟].	۱۰
۱۳.۱	نتیجه اجرای روش مبتنی بر NPD [؟].	۱۱
۱۴.۱	نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره [؟].	۱۳
۱۵.۱	یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].	۱۳
۱۶.۱	ویژگی های هندسی (رنگ سفید) مورد استفاده در آزمایش های تشخیص چهره [؟].	۱۴
۱۷.۱	(الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [؟].	۱۵
۱۸.۱	مدل سازی سه بعدی چهره با اشعه فرو سرخ [؟].	۱۵

- ۱۸ ..... ۱۹.۱ نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [؟].
- ۱۸ ..... ۲۰.۱ (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [؟].
- ۲۱ ..... ۲۱.۱ شبکه عصبی عمیق برای شناسایی چهره [؟].
- ۲۲ ..... ۲۲.۱ سه گانه تطبیق - عدم تطبیق [؟].
- ۲۳ ..... ۲۲.۱ (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [؟].
- ۲۴ ..... ۲۴.۱ معماری کلی شبکه GoogLeNet [؟].
- ۲۴ ..... ۲۵.۱ کادر سبز رنگ یکی از بخش های موازی شبکه را نشان می دهد [؟].
- ۲۴ ..... ۲۶.۱ بخش آغازگر شبکه GoogLeNet [؟].
- ۲۵ ..... ۲۷.۱ معماری شبکه VGGFace [؟].
- ۲۵ ..... ۲۸.۱ نتیجه موقعیت ۱۹۴ شاخص روی چهره [؟].
- ۲۶ ..... ۲۹.۱ علامت گذاری خطوط راهنمای بر روی چهره که در هر تکرار با کاهش خطای همراه می باشد [؟].
- ۲۶ ..... ۳۰.۱ (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری TCNN [؟].
- ۲۷ ..... ۳۱.۱ کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن point-wise و depth-wise هست (سمت راست). [؟].
- ۳۲ ..... ۱.۲ رویکردهای مختلف هم ترازی چهره [؟].
- ۳۲ ..... ۲.۲ رویکرد مبتنی بر تطبیق کلیشه [؟].
- ۳۳ ..... ۳.۲ معماری OpenFace [؟].
- ۳۳ ..... ۴.۲ جریان یادگیری در معماری OpenFace [؟].
- ۳۴ ..... ۵.۲ رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [؟].
- ۳۵ ..... ۶.۲ مقدار دهی اولیه و بهینه سازی AAM [؟].
- ۳۵ ..... ۷.۲ نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [؟].
- ۳۶ ..... ۸.۲ معماری شبکه پیشنهادی VS2VI [؟].
- ۳۷ ..... ۹.۲ (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [؟].

- ۱۰.۲ مقایسه روش ارائه شده با سایر روش ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش های دیگر (d) خروجی روش ارائه شده [۴].
- ۱۱.۲ ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۴].
- ۱۲.۲ اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۴].
- ۱۳.۲ تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۴].
- ۱۴.۲ رویکرد کلی الگوریتم GD-HASLR [۴].
- ۱۵.۲ ردیابی، یافتن چهره ها و برچسب زنی [۴].
- ۱۶.۲ تصاویر با حاشیه قرمز رنگ، به اشتباہ برچسب زنی شده اند [۴].
- ۱۷.۲ استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [۴].
- ۱۸.۲ تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۴].
- ۱۹.۲ ساختار شبکه GAN - شبکه AD - شبکه GN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۴].
- ۲۰.۲ معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۴].
- ۲۱.۲ نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۴].
- ۲۲.۲ چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۴].
- ۲۳.۲تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مزدیین دسته ها ایجاد می نماید [۱].

## فهرست نمادها

فصل ۱

## مروری بر کارهای گذشته

## ۱.۱ مقدمه

همانطور که در فصل مقدمه بیان شد، دو موضوع یافتن چهره در تصویر و شناسایی چهره، 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 [۲] 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 موبروی پیشانی ریخته باشد یا در حضور عینک، به درستی عمل نمی‌کند. یا به عنوان مثال:

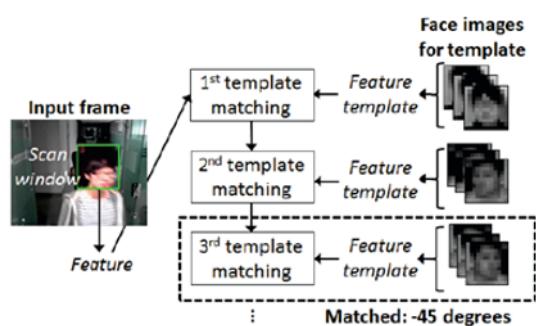
- هر چهره داری دو فرو رفتگی برای چشم‌ها است و چیزی شبیه به ابرو روی این فرو رفتگی‌ها قرار دارد.
- صورت شامل بینی، چشم‌ها و دهان در فاصله‌ها و موقعیت‌های خاصی با یکدیگر می‌باشد.
- چهره مانند یک ناحیه کوچکتر است که بر روی یک ناحیه بزرگتر مانند شانه‌ها قرار گرفته است.

- چهره انسان متقارن 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 تنها ی کافی نیست و موفق به یافتن چهره‌ها در شرایط کنترل نشده با تعداد زیادی چهره نمی‌باشد.

### ۲.۲.۱ رویکردهای مبتنی بر تطبیق کلیشه یا الگو

این رویکرد با استفاده از قالب‌های 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 ۲-۱ 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 ۱.۱: رویکرد مبتنی بر تطبیق کلیشه [۹].

### ۳.۲.۱ رویکردهای مبتنی بر ویژگی

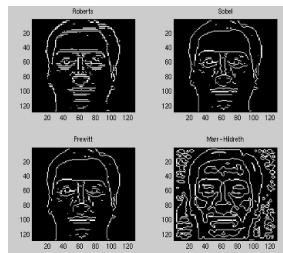
این رویکردها با استخراج ویژگی‌های ساختاری چهره، چهره‌ها را پیدا می‌نمایند. ابتدا به عنوان یک طبقه‌بند، آموزش دیده و سپس برای تمايز میان نواحی شامل چهره و بدون چهره در تصویر استفاده می‌شوند. ایده این رویکرد، غلبه بر محدودیت دانش ما از چهره‌ها می‌باشد. در [۲] تعدادی از این رویکردها مورد بررسی قرار گرفته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

### ۱.۳.۲.۱ رویکرد مبتنی بر حرکت

اگر یک توالی از چند فریم در اختیار باشد، می‌توان به کمک اطلاعات حرکت، 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 می‌شود. بدین ترتیب می‌توان 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 می‌شود.

### ۲.۳.۲.۱ رویکرد مبتنی بر لبه

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



شکل ۲.۱: استفاده از لبه یاب‌های معروف برای استخراج ویژگی‌های مبتنی بر لبه از چهره [?].

### ۳.۳.۲.۱ رویکرد مبتنی بر اطلاعات سطح خاکستری

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



شکل ۳.۱: اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می باشد [۴].

#### ۴.۳.۲.۱ رویکرد مبتنی بر اطلاعات رنگی

رنگ در تصویر اطلاعات با ارزشی به ما می دهد و می توان از رنگ پوست انسان برای یافتن چهره در تصویر استفاده کرد. برای این کار ابتدا رنگ ها هنجار سازی می شوند تا اثر نور پردازی از بین برود. گستردگی ترین فضای رنگ مورد استفاده RGB می باشد.

$$r = \frac{R}{R + G + B} \quad (۱.۱)$$

$$g = \frac{G}{R + G + B} \quad (۲.۱)$$

$$b = \frac{B}{R + G + B} \quad (۳.۱)$$

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

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

$$r + g + b = 1 \quad (۴.۱)$$

پس می توان فقط با داشتن مقدار r و g مقدار b را بدست آورد. با توجه به بافت نگار رنگ سبز و قرمز تصویر، رنگ پوست انسان، بخش کوچکی از بافت نگار را اشغال می کند. بنابراین با بررسی پیکسل های تصویر، می توان با دقت بالایی احتمال حضور چهره در تصویر را تشخیص داد و رنگ پوست انسان را می توان 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 می باشد از ویژگی های قدرتمندتر چهره استفاده نماییم.

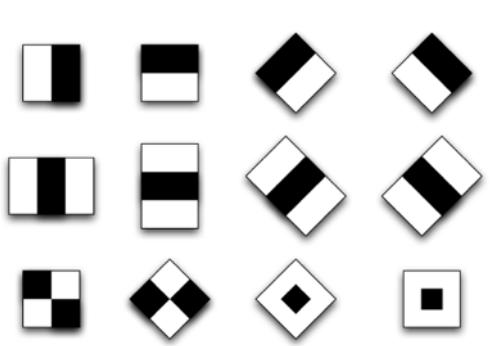
## ٤.٢.١ رویکرد مبتنی بر عامل‌های آماری

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 برای ویژگی‌ها و ارائه یک مدل احتمالی برای چهره، باعث انعطاف پذیری بیشتر سامانه می‌شود. این رویکرد قادر است در مواجه با جا به جایی، چرخش و تغییر مقیاس با دقت بیشتری عمل نماید. در بک رویکرد دقیق‌تر از شبکه‌های بیز 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 و Jones [٤] در [٤] 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، برش، تارشدن و تیزکردن لبه‌های تصویر 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 نشان‌دهنده بخش‌های مختلف صورت می‌باشند.

شکل ٤-٢ نمونه‌هایی از مستطیل‌های ویژگی‌های Haar را نشان می‌دهد.



شکل ٤.١: نمونه‌هایی از مستطیل‌های ویژگی‌های Haar [؟].

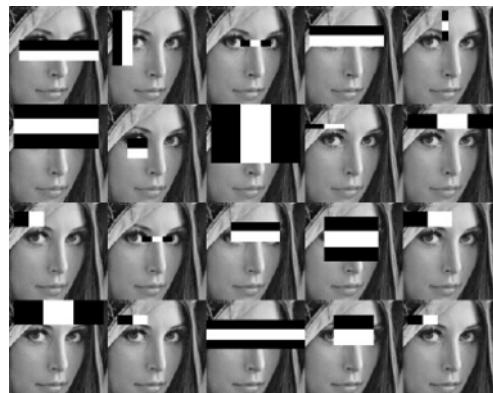
ویژگی‌های Haar برای تشخیص چشم، بینی، دهان و... با کمک تشخیص لبه، تشخیص خط و تشخیص مرکز در تصویر و استخراج ویژگی برای یافتن چهره استفاده می‌شود. مستطیل‌های ویژگی‌های Haar، متناسب با بخش‌های چهره می‌باشند که مثالی از آن در



شکل ۵.۱: مستطیل های ویژگی های Haar متناسب با بخش های چهره می باشند [؟].

شکل ۵-۲ نشان داده شده است.

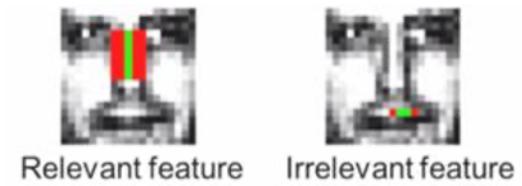
همانطور که در شکل ۶-۲ مشاهده می شود، هر مستطیل در بخش های مختلف چهره در روند تکراری با اندازه های مختلف قرار می گیرد و نتیجه نهایی از کم کردن مجموع سطح روشنایی پیکسل های زیر بخش های زیر بخش های سیاه از مجموع سطح روشنایی پیکسل های زیر بخش های سفید به دست می آید که یک عدد می باشد. از یک پنجره با اندازه  $24 \times 24$  برای قرار دادن مستطیل ها بر روی تصویر استفاده می شود که تعداد زیاد و اندازه های مختلف آن ها باعث می شود برای محاسبه نتیجه نهایی نیاز به انجام بیش از ۱۶۰۰۰۰ محاسبه باشد که زمان زیادی برای هر تصویر خواهد گرفت.



شکل ۱.۶: هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [؟].

Ada Boost: همانطور که در شکل ۷-۲ 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.

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



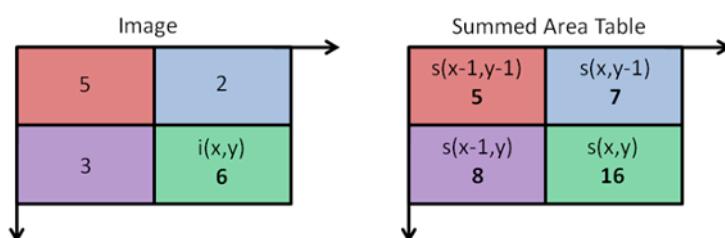
شکل ۷.۱: یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [؟].

عنوان طبقه‌بندهای ضعیف معرفی می‌شوند. Ada Boost در طبقه‌بند قوی، تعداد زیادی طبقه‌بند ضعیف را با هم ترکیب می‌کند. رابطه کلی آن به صورت زیر می‌باشد.

$$F(x) = {}_1f_1(x) + {}_2f_2(x) + \dots \quad (5.1)$$

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

تصویر یکپارچه: تصویر یکپارچه، یا جدول محدوده مجتمع، به منظور ارزیابی سریع تر 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 ۸.۱: نحوه مقدار دهی به پیکسل‌های تصویر یکپارچه [؟].

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

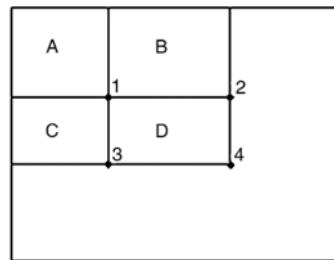
- مقدار پیکسل ۱ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل  $A$ .

- 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  $C$ .

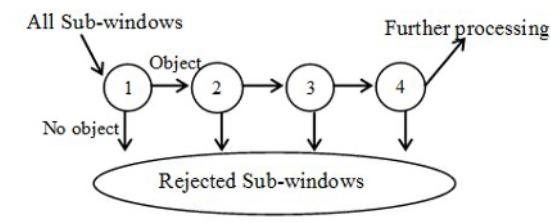
- مقدار پیکسل ۴ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل A و B و C و D.

- مجموع پیکسل‌ها در مستطیل D می‌تواند به صورت  $(3+4) - (1+4)$  محاسبه شود.



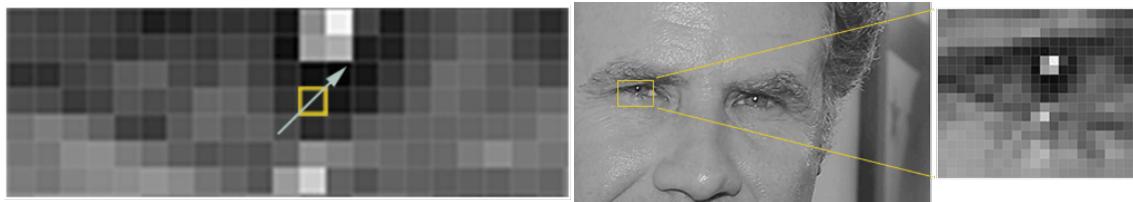
شکل ۹.۱: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [؟].

مراحل آبشاری: اگر تصویر به مربع‌های  $24 \times 24$  پیکسلی تقسیم شده و با پردازش هر بخش که  $2500$  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.  $2500$  ویژگی هر مربع  $24 \times 24$  به دسته‌بندی‌های مختلف تقسیم می‌شود. برای مثال  $10$  ویژگی در دسته اول،  $20$  ویژگی در دسته دوم،  $100$  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 آبشاری را نشان می‌دهد. مجموعه‌ای از طبقه‌بندها به هر زیر‌پنجره 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.

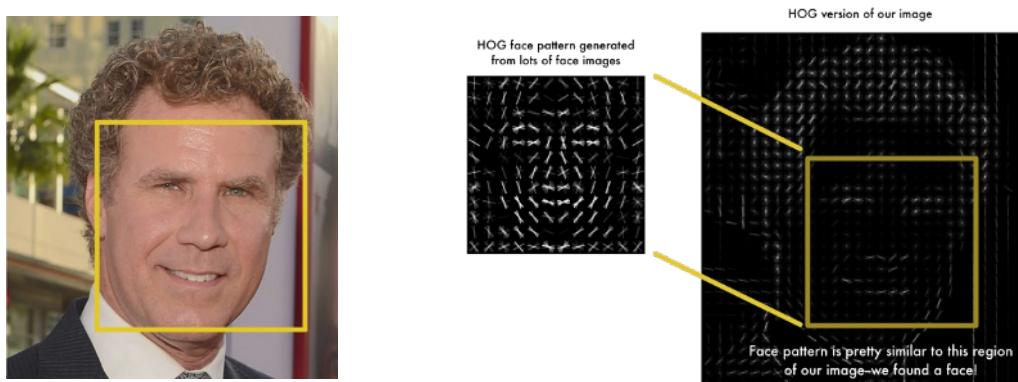


شکل ۱۰.۱: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می‌شود [؟].

در سال ۲۰۰۵ Dalaal و همکاران در [۵] روشی به نام بافت نگار شیب‌های جهت دار 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



شکل ۱۱.۱: سطح روشانایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می شود [۹].



شکل ۱۲.۱: برای یافتن چهره ها، بخش هایی از تصویر که به الگوی HOG شبیه تر است را پیدا می کنیم [۹].

روند برای همه پیکسل های تصویر انجام می شود و به ازای هر پیکسل یک جهت ذخیره خواهد شد که روندی از روشانایی به تاریکی را در تصویر نمایش می دهند.

دلیل استفاده از جهت ها این است که اگر پیکسل ها به طور مستقیم استفاده شوند، تصویر تاریک و تصویر روشن از یک چهره مشخص، دارای سطح روشانایی متفاوتی خواهند بود. اما با در نظر گرفتن جهتی تغییر روشانایی، هم تصویر تیره و هم تصویر روشن، نمایش یکسانی خواهند داشت که حل مسئله را آسان تر می کند. ذخیره جهت ها برای تمام پیکسل ها باعث افزایش جزئیات می شود. لذا روند اصلی روشانایی و تاریکی در سطحی بالاتر در نظر گرفته می شود، به طوری که بتوان الگوی اصلی تصویر را دید. تصویر به بخش های  $16 \times 16$  پیکسل تقسیم می شود و در هر بخش تعداد جهت های به سمت بالا، پایین، چپ و راست شمارش می شود. سپس بخش های درون تصویر با جهت هایی که بزرگتر بودند، جایگزین می شود. نتیجه نهایی، تبدیل تصویر به یک نمایش ساده از ساختار چهره است که در شکل ۱۲-۲ مشاهده می شود. برای یافتن چهره ها در این الگوریتم، بخش هایی از تصویر که به الگوی HOG شبیه تر است، مشخص می شود. با استفاده از این روش، می توان چهره ها را در تصویر به سادگی پیدا کرد.

در سال ۲۰۱۵ Shengcai Liao و همکاران در [۶] یک روش دقیق و سریع برای یافتن چهره ارائه دادند که در آن از اختلاف پیکسل هنجارسازی شده یا NDP برای یافتن چهره استفاده می شود. ارزیابی ویژگی NPD بسیار سریع است و دسترسی به حافظه تنها با استفاده از یک جدول جستجو می باشد. در این روش نشانه گذاری یا خوش بندی 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:

$$f(x, y) = \frac{x - y}{x + y} \quad (6.1)$$

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  $p = s^2$  تعداد پیکسل‌ها باشد، آنگاه تعداد ویژگی‌های استخراج شده برابر  $d = p(p-1)/2$  می‌باشد. سپس علامت 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 تیوه‌تر می‌باشد. همچنین این ویژگی‌ها به خدشه 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 ۱۳.۱: نتیجه اجرای روش مبتنی بر NPD [؟].

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

## ۵.۲.۱ 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 چهره و نمونه‌هایی از غیر چهره به طبقه‌بند داده می‌شود تا از روی این تصویرها عمل یادگیری انجام شود. به طور تجربی دقیق نتایج رویکردهای مبتنی بر تصویر بهتر از سایر رویکردها می‌باشد. این رویکردها به چند دسته تقسیم می‌شوند که در ادامه شرح داده شده است.

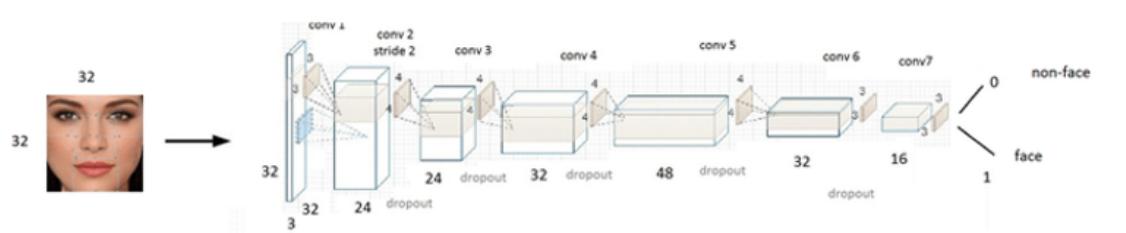
### ۱.۵.۲.۱ رویکرد مبتنی بر ماشین بردار پشتیبان

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

### ۲.۵.۲.۱ رویکرد مبتنی بر شبکه عصبی

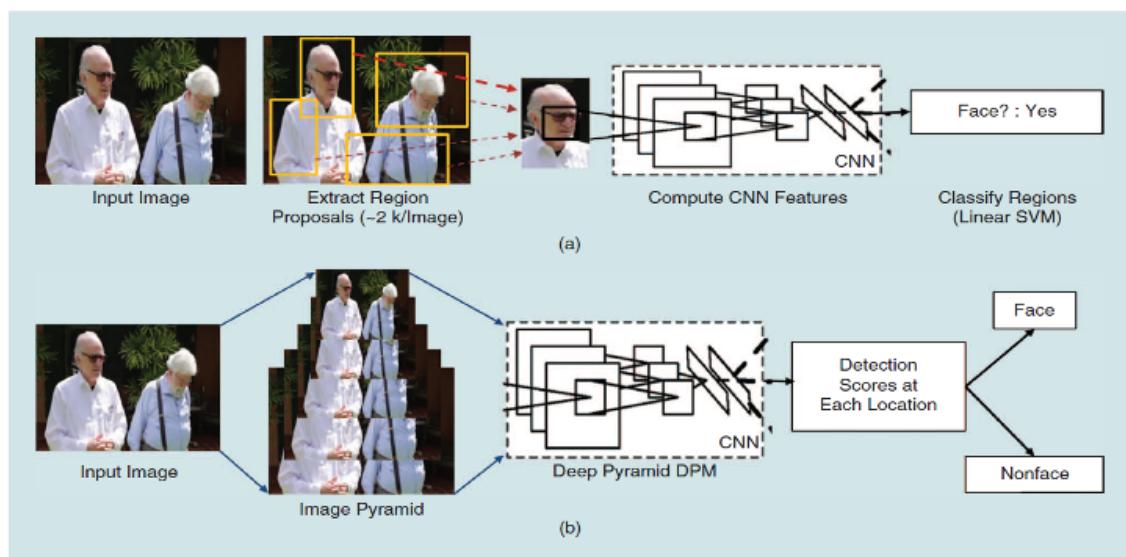
یک راه غیر خطی برای یافتن چهره، استفاده از شبکه‌های عصبی است. اولین رویکردهای عصبی در یافتن چهره بر اساس MLP بود که در مجموعه داده‌های ساده، امیدوار کننده بود. شبکه‌های عصبی با معماری پیمانه ای امروزی بسیار پیچیده‌تر از MLP ساده هستند. نوع خاصی از شبکه‌های عصبی عمیق برای پردازش تصاویر استفاده می‌شوند که شبکه عصبی پیچشی 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 پیچشی برای یافتن چهره را نشان می‌دهد. عبارتند از:

لایه‌های پیچشی برای لغزاندن یک پنجره بر روی ورودی  
لایه‌های تمام متصل برای محاسبه مجموع وزن دار تمام واحد های ورودی  
لایه‌های رای گیری به منظور کاهش حجم داده‌ها با محاسبه مقدار بیشینه، میانگین یا اندازه اقلیدسی هر بخش با توجه به آنچه در [۸] 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



شکل ۱۴.۱: نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره [؟].

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



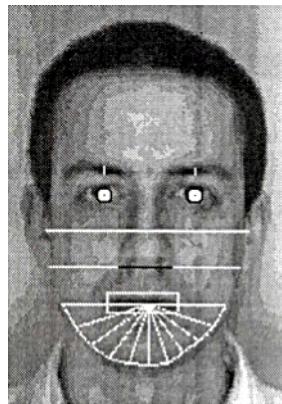
شکل ۱۵.۱: یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].

### ۳.۱ شناسایی چهره

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

### ۱.۳.۱ 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 [۹] 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.



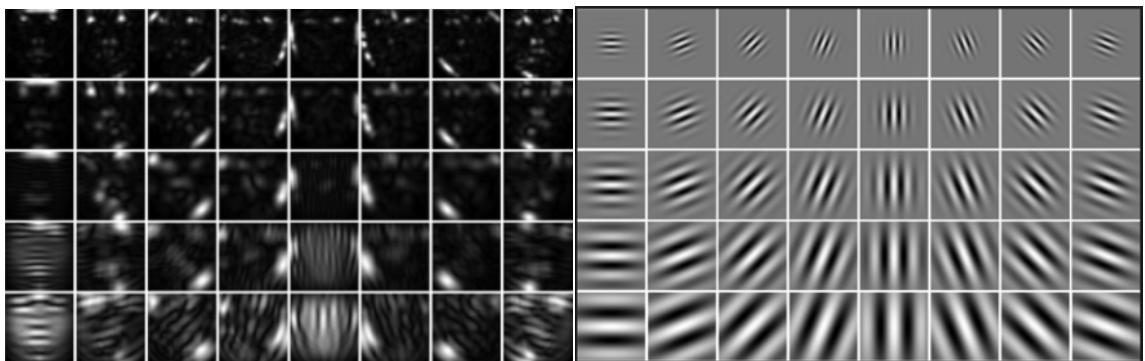
شکل ۱۶.۱: ویژگی‌های هندسی (رنگ سفید) مورد استفاده در آزمایش‌های تشخیص چهره [۹].

### ۲.۳.۱ رویکرد مبتنی بر فیلتر گابور

همانطور که در [۱۰] آمده است، در این رویکرد ابتدا تصویر را بخش بندی کرده، سپس بر روی بخش‌های مختلف آن، فیلتر گابور اعمال می‌شود و نتیجه بدست آمده با یک طرح از پیش آماده شده، با یک آستانه گذاری مطابقت داده می‌شود. شکل ۱۷-۲ فیلترهای چندگانه گابور و تاثیر این فیلترها بر روی تصویر چهره انسان را نشان می‌دهد. دلیل استفاده از فیلتر گابور این است که عملکرد این فیلتر به سامانه بصری انسان بسیار شباهت دارد.

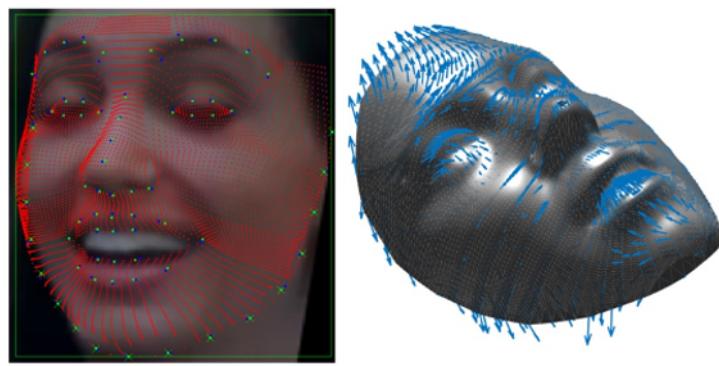
### ۳.۳.۱ رویکردهای سه بعدی

همانطور که در [۱۱] 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 فروسرخ و یک حسگر به عنوان دریافت کننده استفاده می‌کند. شبکه‌ای از نورهای فروسرخ که برای انسان قابل رویت نیست، روی



شکل ۱۷.۱: (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [؟].

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 فروسرخ را نشان می‌دهد.



شکل ۱۸.۱: مدل سازی سه بعدی چهره با اشعه فروسرخ [؟].

#### ۴.۳.۱ 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

نهایی 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 [۱۲] 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 خطرا 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 بیاده و خیابان می‌باشد، به توضیح مختصر رویکردهای سه بعدی و تجزیه و تحلیل بافت پوست بسنده می‌کنیم.

### ۵.۳.۱ 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 [۱۳] 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 الکترونیکی فروسرخ با حساسیت کم و با توانایی جذب حرارت طولانی مدت یا LWIR 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 دقیقاً ۵۰.۹۷٪ دارد، در حالی که روش LWIR دارای دقیقاً ۹۳٪.۹۳٪ می‌باشد و ترکیب این دو دارای دقیقاً ۴۰٪.۹۸٪ است. در شرایط کنترل نشده دوربین معمولی دقیقاً ۰.۶٪، دوربین LWIR دقیقاً ۰.۳٪.۸۳٪ و ترکیب این دو دارای دقیقاً ۰.۲٪.۸۹٪ است.

### ٦.٣.١ تشخیص چهره مبتنی بر ویدیو

در سال ۲۰۰۹ Wang Huafeng و همکاران در [۱۴] 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 افروزنگی موجود در توالی ویدیو برای بهبود عملکرد تشخیص نسبت به تصاویر ثابت وجود دارد. تشخیص

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

۲. مطالعات 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.

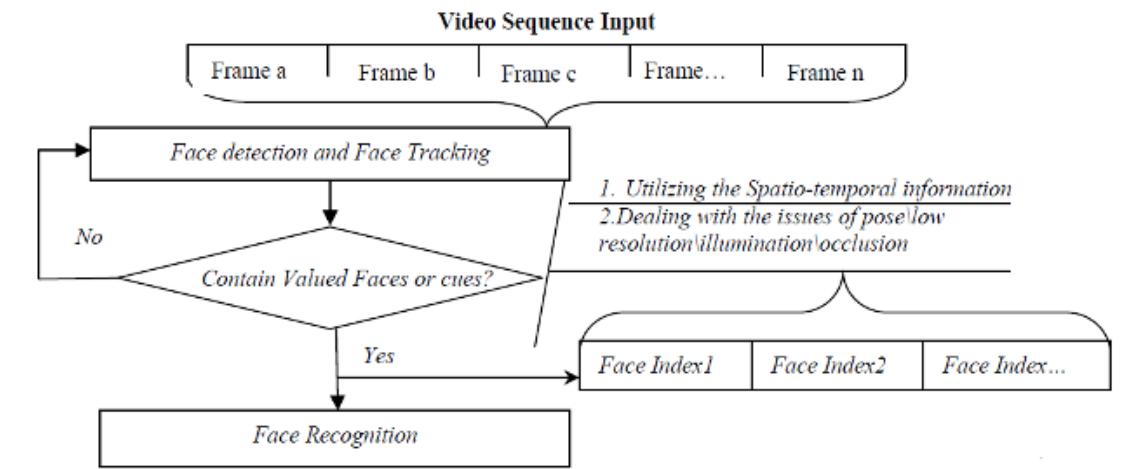
برای تشخیص چهره در تصاویر ویدیویی، دو رویکرد کلی وجود دارد:

مبتنی بر قاب : 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.

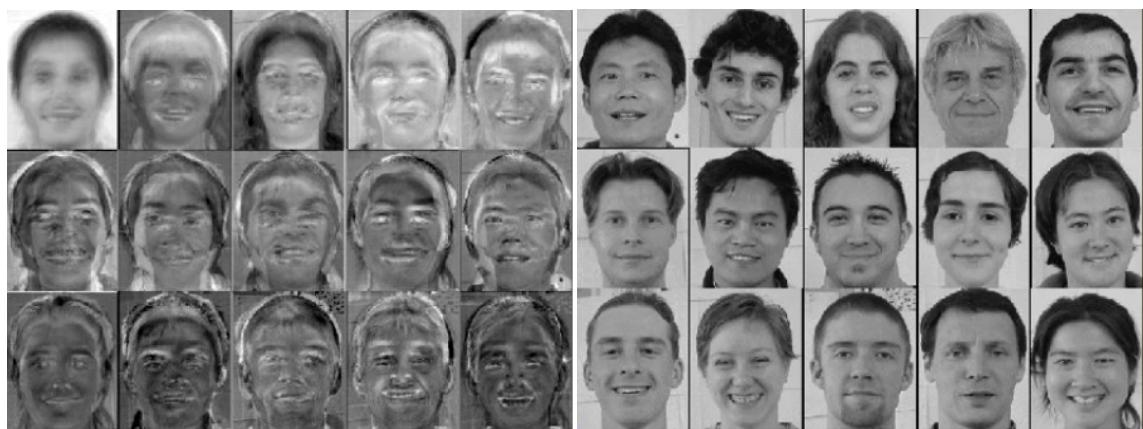
یافتن و ردیابی : یافتن چهره در اولین قاب و سپس ردیابی آن از طریق توالی قاب‌ها. شکل ۱۹-۲ نمای کلی این رویکرد را نشان می‌دهد.

### ٧.٣.١ رویکرد مبتنی بر چهره ویژه

با افزایش حجم داده‌های موجود، نیاز به کاهش ابعاد داده‌ها می‌باشد. تحلیل مؤلفه‌های اساسی یا PCA یک روش کاهش ابعاد داده‌ها است که در برخی مسئله‌ها مانند پردازش تصویر به خوبی و با سرعت بالا عمل می‌کند. استفاده از این روش در پردازش سریع تر داده‌ها کمک می‌کند و از رخ دادن مشکل بیش‌برازاندن 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



شکل ۱۹.۱: نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [۹].



شکل ۲۰.۱: (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۹].

و شخص شبیه به آن را پیدا کنیم، مقایسه تصاویر بسیار زمان برو در مواردی غیر ممکن خواهد بود. PCA در این مسئله به خوبی عمل می‌کند. با اعمال تکنیک کاهش بعد به تصویر و با به دست آوردن تصویر ویژه چهره‌ها 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 [۱۵] 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 براکندگی را دارد برای ما مهم است، راه حل مناسبی خواهد بود.

در سال ۲۰۱۶ K. R. Sreelakshmi و همکاران در [۱۶] 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 منهتن

میتوان استفاده کرد. از مزایای این روش میتوان به سهولت پیادهسازی و استفاده، کاهش حجم دادهها و سرعت بالا اشاره کرد. در نظر نگرفتن پراکندگی درون کلاسی و بین کلاسی دادهها و عدم توجه به برچسب تصویر برای شناسایی و تمایز قابل نشنیدن بین تصاویر مختلف یک شخص در پایگاه و نیاز به بروز رسانی تمامی اطلاعات موجود با ورود یک تصویر جدید به پایگاه داده از معایب این روش است. محاسبات ریاضی و مراحل انجام آن‌ها:

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

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

۳. محاسبه ماتریس کوواریانس بردارها و مقادیر ویژه آن

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

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

#### ۸.۳.۱ رویکرد مبتنی بر ماشین بردار پشتیبان

همان طور که در بخش ۲-۴-۱ گفته شد، ماشین‌های بردار پشتیبان، طبقه‌بندی‌های خطی هستند که حاشیه بین ابرصفحه تصمیم و نمونه‌های مجموعه آموزش را به حداقل می‌رسانند. در سال ۲۰۱۲ N.M.Khan و همکاران در [۱۷] این طبقه‌بند را برای تشخیص چهره مورد استفاده قرار دادند. یک نسخه مبتنی بر هسته برای SVM معرفی شده که K-SVM نام گذاری شده است.

$$a = b \quad (7.1)$$

که در آن  $w$  بردار وزن‌ها،  $x$  داده‌های ورودی و  $w_0$  مقدار پیش قدر می‌باشد. سپس یک مدل بی‌پارامتر از LDA به نام NDA معرفی شده، سپس یک نسخه مبتنی بر هسته به نام K-NDA معرفی شده است.

$$\begin{aligned} S_{bk}^{\phi} &= \frac{1}{(N_1+N_2)} \sum_{i=1}^{N_1} \omega'_i (\Phi(x_i) - M_2^k(\Phi(x_i))) (\Phi(x_i) - M_2^k(\Phi(x_i)))^T \\ &+ \frac{1}{(N_1+N_2)} \sum_{i=N_1+1}^{N_1+N_2} \omega'_i (\Phi(x_i) - M_1^k(\Phi(x_i))) (\Phi(x_i) - M_1^k(\Phi(x_i)))^T \end{aligned} \quad (8.1)$$

جدول ۱.۱: مقایسه دقت الگوریتم NDA + SVM با سایر رویکردهای مشابه

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

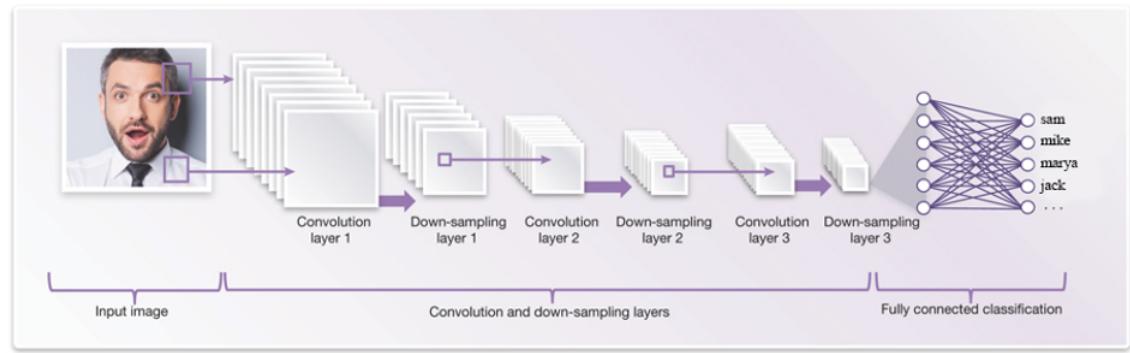
که در آن  $N_1$  و  $N_2$  تعداد نمونه‌های آموزشی در هر دسته می‌باشند. سپس از ترکیب دو رویکرد فوق، مدلی به نام SVM + NDA که در آن ماتریس تنظیم می‌باشد و ضریبی برای تنظیم کنترل میان SVM و NDA می‌باشد. رابطه بالا یک مسئله بهینه‌سازی طراحی نموده است.

$$\begin{aligned}
 & \min_{w \neq 0, w_0, C > 0} \quad \frac{1}{2} w^T (\lambda S_w^\Phi (S_{bk}^\Phi + \beta I)^{-1} + I) w + C \sum_{i=1}^N \xi_i \\
 & \text{s.t.} \quad t_i \left( \sum_{j=1}^N w_j K(x_i; x_j) + w_0 \right) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, \dots, N
 \end{aligned} \tag{۹.۱}$$

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.

### ۹.۳.۱ رویکردهای مبتنی بر شبکه عصبی

یک راه حل غیر خطی برای شناسایی چهره، استفاده از شبکه عصبی پیچشی است که به طور شگفت انگیزی در طبقه-بندی تصاویر چهره خوب کار می کند و ویژگی های ارزشمندی را از تصویر چهره استخراج می کند. بنابراین می توان 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 می دهد.



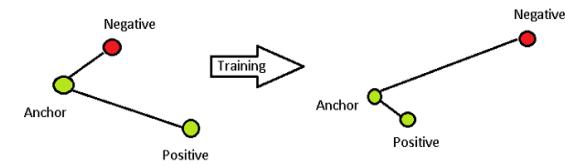
شکل ۲۱.۱: شبکه عصبی عمیق برای شناسایی چهره [؟].

معمولًا به عنوان تابع فعالیت از توابع غیرخطی مانند ReLU استفاده می شود. و عملیات بهینه سازی به روش پس انتشار خطا انجام می گردد. و در لایه خروجی از تابع SoftMax برای طبقه بندی استفاده می شود که خروجی های لایه آخر را هنجار می کند. شبکه های عصبی پیچشی ویژگی های یک چهره را استخراج می کنند که می توان به عنوان یک شناسه برای یک فرد خاص در نظر گرفت. هنگامی که دو تصویر مختلف از چهره یک شخص به عنوان ورودی داده می شود، شبکه باید خروجی های مشابه (ویژگی های نزدیک تر) را برای هر دو تصویر تولید نماید، در حالی که برای چهره دو شخص مختلف، شبکه باید خروجی های بسیار متفاوت برای دو تصویر تولید نماید. شبکه عصبی نیاز به آموزش دارد تا به طور خودکار ویژگی های مختلف چهره ها را شناسایی کند و بر اساس آن محاسبات را انجام دهد. در ادامه چند شبکه عصبی پیچشی معروف مورد بررسی قرار گرفته است.

#### ۱.۹.۳.۱ FaceNet شبکه

در سال ۲۰۱۵ FaceNet و همکاران در [۱۸] یک شبکه عصبی عمیق به نام FaceNet ارائه دادند. FaceNet یک مدل یکپارچه است که می آموزد چگونه تصاویر چهره را به یک فضای اقلیدسی فشرده نگاشت دهد تا فاصله تصاویر به طور مستقیم با میزان شباهت چهره ها مرتبط باشد. هنگامی که این فضا تولید شود، شناسایی چهره، تایید هویت و خوشه بندی می تواند به راحتی با استفاده از روش های استاندارد توسط FaceNet انجام شود. این شبکه برای آموزش از سه گانه تطبیق - عدم تطبیق استفاده می نماید. 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 ۲۲.۱: 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.

$$f(A) - f(P)^2 f(A) - f(N)^2 \quad (10.1)$$

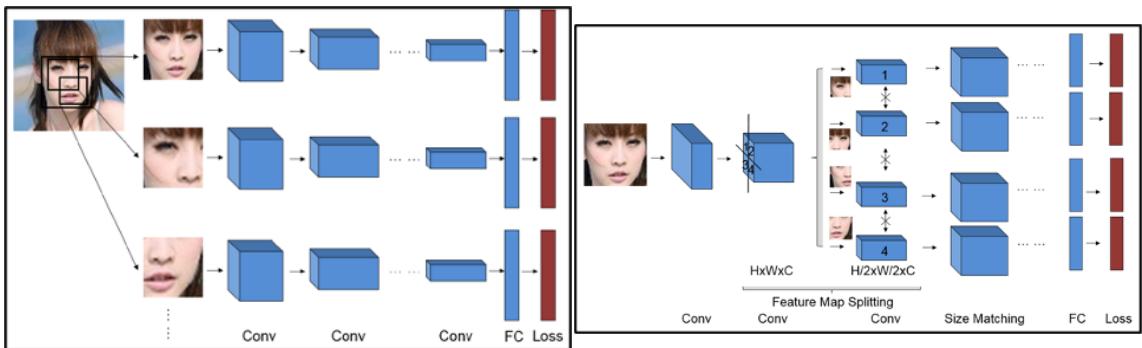
که در آن تابع ویژگی‌های استخراج شده از تصاویر است. شبکه در مرحله آموزش با استفاده از داده‌های برچسب‌گذاری شده، می‌آموزد فاصله بین ویژگی‌های شبیه به هم، کمتر از فاصله بین ویژگی‌های دور باشد و به این ترتیب در مرحله آزمایش می‌تواند داده‌های مشابه و غیر مشابه را به راحتی تفکیک نماید. تابع هزینه در این مدل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

$$L = (i=1)^n f(x_i^A) - f(x_i^P)^2 - f(x_i^A) - f(x_i^N)^2 + \quad (11.1)$$

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 ۲.۹.۳.۱

در سال ۲۰۱۸ و همکاران در [۱۹] 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



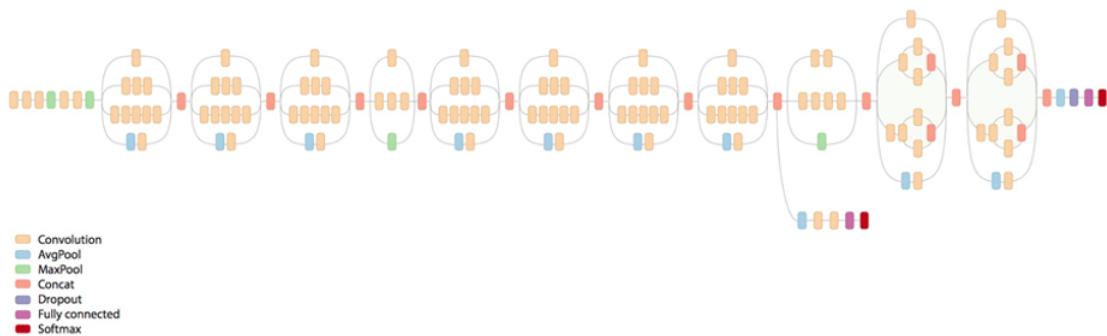
شکل ۲۳.۱: (الف) شبکه عصبی پیچشی SplitNet (ب) شبکه عصبی پیچشی معمولی [؟].

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 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 دقیق تشخیص چهره را با محاسبات کمتر افزایش دهد.

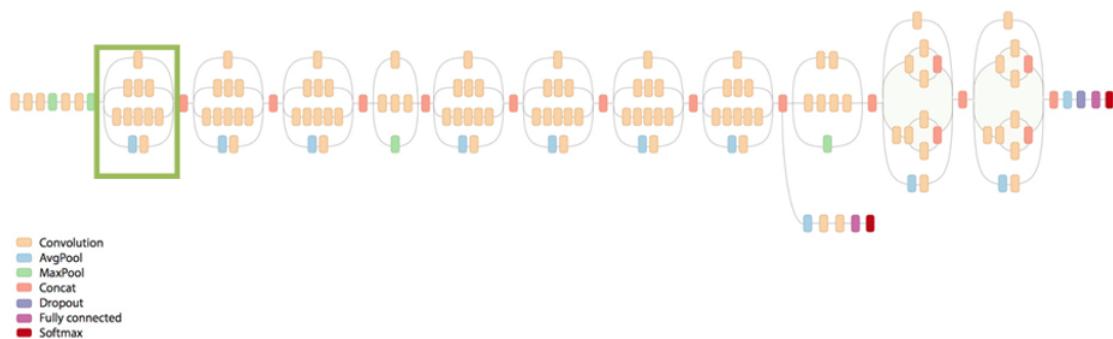
### GoogLeNet شبکه ۳.۹.۳.۱

در سال ۲۰۱۵ Christian Szegedy و همکاران در [۲۰] 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 متوالی 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. در GoogLeNet تمام محاسبات به طور متوالی 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

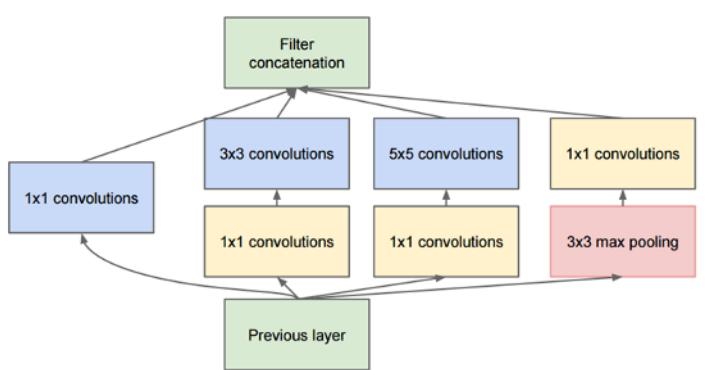


شکل ۲۴.۱: معماری کلی شبکه GoogLeNet.



شکل ۲۵.۱: کادر سبز رنگ یکی از بخش‌های موازی شبکه را نشان می‌دهد [۹].

ایده ساده‌ای بود که نویسنندگان مقاله روی آن تمرکز کردند.



شکل ۲۶.۱: بخش آغازگر شبکه GoogLeNet.

### VGGFace شبکه ۴.۹.۳.۱

در سال ۲۰۱۵ Omkar M. Parkhi و همکاران در [۲۱] شبکه عمیق VGGFace را ارائه کردند که شامل یک توالی طولانی از لایه‌های پیچشی می‌باشد. با توجه به شکل ۲۷-۲ این شبکه که در لایه آخر به عنوان یک طبقه‌بند عمل می‌نماید، هر تصویر آموزشی

چهره را توسط لایه تماماً متصل وتابع ضرر softmax log-loss به یک بردار تبدیل می‌نماید که هر مقدار در این بردار، نشان دهنده احتمال برای یک هویت فردی است. FaceNet مشابه VGGFace از یک تابع ضرر سه گانه در آموزش برای بهبود عملکرد کلی استفاده می‌نماید.

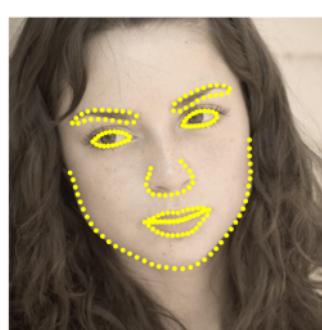


شکل ۲۷.۱: معماری شبکه [۹] VGGFace

### ۱۰.۳.۱ رویکردهای مبتنی بر نقاط راهنمای

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

در سال ۲۰۱۴ وحید کاظمی و جوزفین سالیوان در [۲۲] یک الگوریتم برای یافتن نقاط راهنمای چهره 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

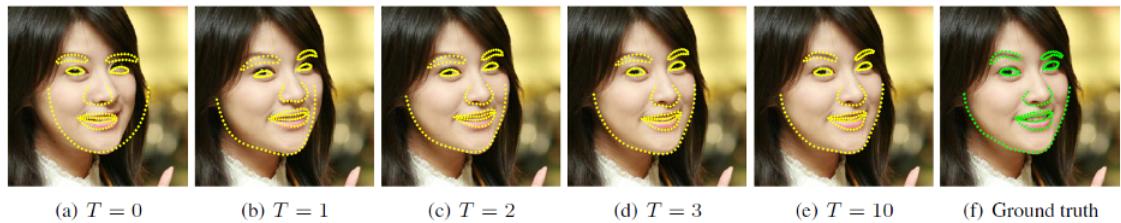


شکل ۲۸.۱: نتیجه موقعیت ۱۹۴ شاخص روی چهره [۹].

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 راهنمایی بر روی چهره را نشان می‌دهد.



شکل ۲۹.۱: علامت گذاری خطوط راهنمایی بر روی چهره که در هر تکرار با کاهش خطای همراه می‌باشد [۴].

در سال ۲۰۱۶ Yue Wu و همکاران در [۲۳] رویکردی برای یافتن نقاط راهنمایی بر روی چهره مبتنی بر شبکه عصبی پیچشی ارائه

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

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

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

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

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

عنوان تابع ضرر معرفی شده است.

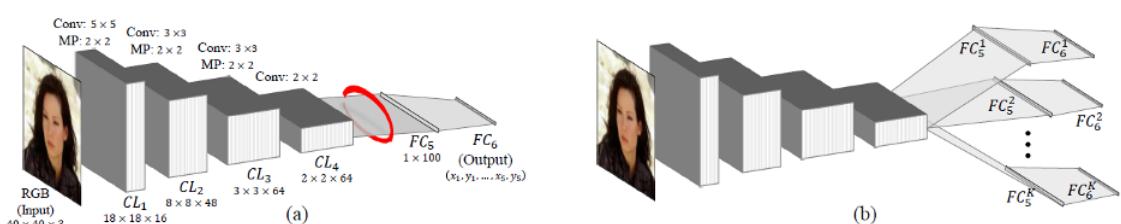
$$L(P_i, \hat{P}_i) = \frac{(P_i - P_{i,2}^2)}{(P_{(i,1)} - P_{(i,2)}^2)} \quad (12.1)$$

که در آن  $P_i$  یک بردار  $2 * m$  برای مختصات پیش‌بینی شده تصویر  $I_i$  و  $P_i$  و مختصات محل دقیق آن نقاط می‌باشد.

مختصات چشم‌ها در تصویر مرجع می‌باشند. در نهایت خروجی لایه تمام متصل توسط الگوریتم GMM به ۶۴ خوشة تقسیم شده

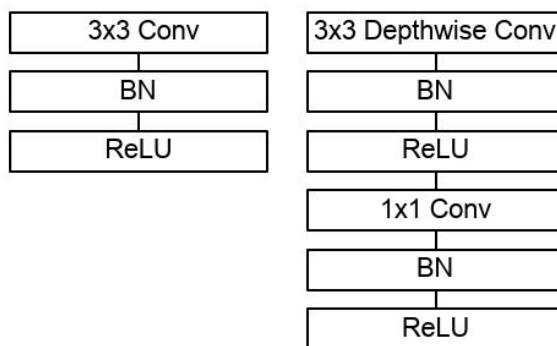
و هریک به صورت جداگانه بررسی شده است. معماری TCNN در شکل ۲۰-۲ قسمت b قابل مشاهده می‌باشد. این شبکه برای

آموزش از مجموعه داده LFW استفاده کرده است.



شکل ۲۰.۱: (a) شبکه عصبی پیچشی معمولی و (b) شبکه عصبی پیچشی با معماری TCNN [۴].

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 [۲۴] 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 یک کرنل  $k$  داریم! حالا با این کرنل، مرحله اول کانولوشن را انجام می‌دهیم. با انجام عمل فیلتر، هر صفحه از کرنل در یک صفحه از فیچرمپ 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 باز هم 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  $1 \times 1$  است. کانولوشن نقطه‌ای همان کانولوشن استاندارد یا رایجی هست که می‌شناسیم و در بسیاری از شبکه‌های کانولوشنی استفاده می‌شود. این کانولوشن  $1 \times 1$  وظیفه مهمی دارد؛ کانولوشن نقطه‌ای  $1 \times 1$  خروجی‌های کانولوشن عمقی (مرحله اول) را بهم ادغام می‌کند. در مرحله قبل بجای تعریف  $M$  کرنل، تنها یک کرنل تعریف کردیم. اما در این مرحله،  $M$  کرنل  $1 \times 1$  تعریف می‌کنیم.



شکل ۳.۱.۱: کانولوشن استاندارد (سمت چپ). کانولوشن dws که شامل دو کانولوشن depth-wise و point-wise هست (سمت راست). [؟].

شبکه عصبی موبایل نت ۲۰۴ میلیون پارامتر دارد. وقتی تعداد پارامترهای این شبکه را با شبکه محبوب ResNet-۱۸ با ۱۱ میلیون

جدول ۲.۱: 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

پارامتر مقایسه کنیم، متوجه می شوید که چقدر میزان پارامترها کمتر است.

## ۴.۱ 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 قبل نیز بیان شد، برای حل مساله دسته بندی چهره دو روش کلی، مبتنی بر تصویر و روش‌های مبتنی بر استخراج ویژگی

وجود دارد. همچنین روش‌های مبتنی بر تصویر خود دارای رویکردهای مختلفی از جمله روش‌های مبتنی بر رنگ بندی، روش‌های

مبتنی بر شکل و روش‌های مبتنی بر گرادیان می‌باشد: همچنین روش‌های مبتنی بر استخراج ویژگی که در سال‌های اخیر بسیار

مورد توجه قرار گرفته اند شامل رویکردهای مبتنی بر شبکه عصبی، بردار پشتیبان و... می‌باشند.

## ۲ 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 تمکر، 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 ۱.۲ خلاصه ای از روش های مقابله با شرایط کنترل نشده

مشکل ها	مزیت ها	رویکرد	چالش مورد نظر	مقاله ها	
استخراج نقاط ویژه باید دقیق تر باشد	پیچیدگی محاسباتی قابل قبول	تبديل دو بعدی	حالات چهره	[۲۸-۲۵]	
پیچیدگی محاسباتی، واستگی به داده های آموزش	دقت بالا در شرایط کنترل نشده	استفاده از شبکه عصبی عمیق	حالات چهره	[۲۲-۲۹، ۲۳]	
		دقت بالا پیچیدگی محاسباتی	تبديل مدل دو بعدی به سه بعدی	حالات چهره	[۲۶-۳۳، ۱۲]
		دقت بالا پیچیدگی محاسباتی	تبديل مدل سه بعدی به دو بعدی	حالات چهره	[۷۷]
قابل استفاده در تصاویر خاکستری	همسان سازی بافت-نگار	پیچیدگی محاسباتی قابل قبول	روشنایی	[۳۸، ۲۸]	
		استفاده از روش های شناسایی الگو و ایجاد	دقت بالا در انسداد شدید	انسداد	[۴۰، ۳۹]
نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو	تهییه مجموعه داده با ردبایی چهره در ویدیو	حدودیت داده	تهییه مجموعه داده با دقت بالا	[۴۱]	
تجهیزات پیشرفته و زمان تاخیر ناهمگن	سرعت بالا	استفاده از رایانش ابری	حدودیت منابع	[۴۳، ۴۲]	

## ۲.۱ چالش حالت

چالش حالت زمانی پیش می آید که چهره فرد کاملا رو به روی دوربین قرار نگیرد و دارای زاویه زیادی باشد. در این شرایط با توجه به ساختار سه بعدی چهره، ممکن است سامانه نتواند ویژگی های درستی از چهره استخراج نماید و در تشخیص هویت دچار اشتباه شود. گرچه شبکه عصبی پیچشی توانایی مقابله با این چالش را از طریق استفاده از مجموعه داده های بزرگ و آموزش تصاویر مختلف از حالات چهره دارد، اما این کار باعث بزرگ شدن پایگاه داده و کند شدن سامانه می شود. استفاده از یک پی برنده 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 و هم ترازی 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. رویکردهای دو بعدی با پیچیدگی محاسباتی قابل قبول (بیشتر ایده های مبتنی بر نشانه گذاری قدیمی) مانند [۲۸-۲۵].

مشکل: در محیط های بدون محدودیت مانند آنچه که در این پروژه داریم، استخراج دقیق مکان نشانه های صورت از تصاویر

دو بعدی نیاز به توجه بیشتری دارد. پیشرفتهای اخیر مانند [۲۵] است.

مزیت: این الگوریتم‌ها از نظر پیچیدگی محاسباتی قابل قبول هستند و کاملاً برای شرایط این پروژه مناسب می‌باشند.

۲. رویکردهای مبتنی بر شبکه عصبی برای تخمین و اصلاح موقعیت چهره (آموزش و آزمایش با تصاویر دو بعدی) مانند [۲۳،

[۳۲-۲۹].

مشکل: این الگوریتم‌ها، به طور متوسط، کندر از دسته پیشین می‌باشند. اما بستگی به این دارد که عمق شبکه عصبی چه

مقدار باشد. وابستگی آن‌ها به داده‌های آموزش می‌باشد و مراحل مجازی رو به روش‌سازی و هم ترازی چهره ندارند.

مزیت: بدون نیاز به تصمیم‌گیری در مورد مجموعه بهینه‌ای از نشانه‌های چهره و دارای دقت بیشتر در شرایط کنترل نشده

با انسداد و... ایده‌هایی مانند [۴۴] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند.

۳. رویکردهای سبک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۱۲، ۳۳-۳۶].

مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده ترین این الگوریتم‌ها در مورد پیچیدگی محاسباتی، [۳۳] است که در

شرایط بدون محدودیت آموزش دیده و آزمایش شده است. شامل مراحل مجازی رو به روش‌سازی و تراز بندی چهره می‌باشند،

اما برای محدودیت‌های این پروژه قابل استفاده نمی‌باشند.

مزیت: با استفاده از اطلاعات سه بعدی، این روش‌ها به بالاترین دقت تصمیم‌گیری در میان سه نفر رسید.

۴. رویکردهای تبدیل مدل سه بعدی چهره به مدل دو بعدی چهره (روش‌های مبتنی بر پنجره بر اساس چند نمایش دو بعدی

مختلف از چهره) مانند [۳۷] مشکل: زمان محاسباتی (نه به اندازه الگوریتم‌های دسته سوم).

مزیت: عملکرد بهتر در رو به روش‌سازی چهره در شرایط کنترل نشده نسبت به الگوریتم‌های دسته اول. ممکن است برای شرایط

این پروژه مناسب باشند.

مرجع [۴۳] در سال ۲۰۱۸ خلاصه‌ای از رویکردهای مختلف برای حل مسئله هم ترازی را در شکل ۱-۳ نشان داده است.

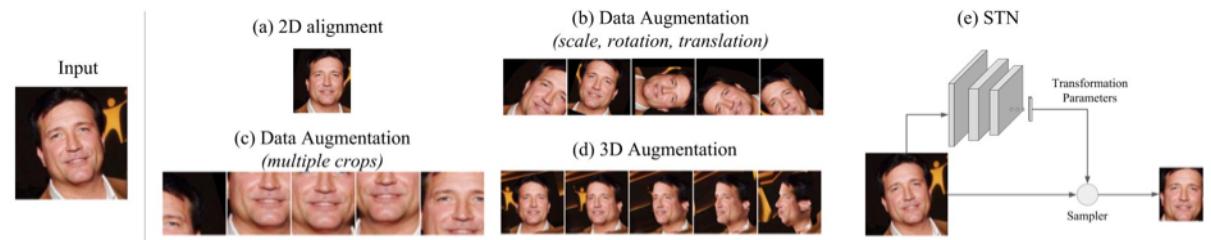
تصویر سمت چپ، چهره ورودی می‌باشد. (a) هم ترازی با استفاده از تبدیلات دو بعدی ساده می‌باشد. (b) داده افزایی با تغییر

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

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

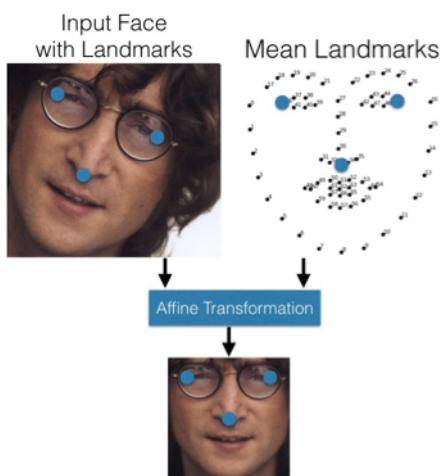
تبدیل هم ترازی را بدست آورد.

در سال ۲۰۱۶ و همکاران در [۲۷] یک روش شناسایی چهره به نام OpenFace Brandon Amos ارائه دادند که ویژگی اصلی



شکل ۱.۲: رویکردهای مختلف هم ترازی چهره [۹].

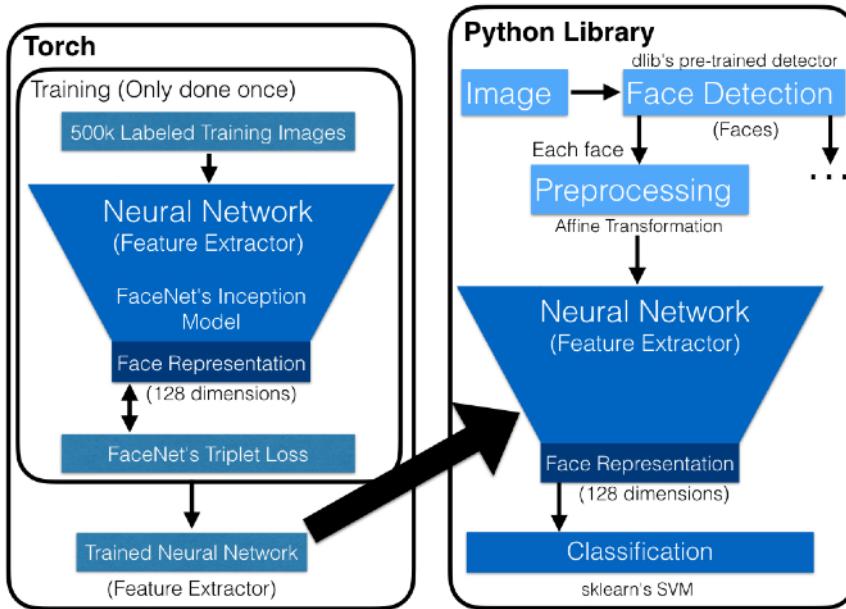
آن، آموزش شبکه عصبی عمیق در کمترین زمان و قابلیت اجرا بر روی دستگاه های قابل حمل مانند تلفن همراه با در نظر گرفتن منابع محدود می باشد. یک تصویر شامل تعدادی چهره به الگوریتم داده می شود. پس از یافتن چهره ها و مجزا کردن آن ها از یکدیگر، هر چهره به طور جداگانه مورد پیش پردازش قرار می گیرد و حجم آن کاهش 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 دوبعدی 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 شده و اندازه آن ها  $96 \times 96$  پیکسل می شود. شکل ۲-۳ - تبدیل هم نسبی OpenFace براساس نقاط ویژه آبی پس از پیش



شکل ۲.۲: رویکرد مبتنی بر تطبیق کلیشه [۹].

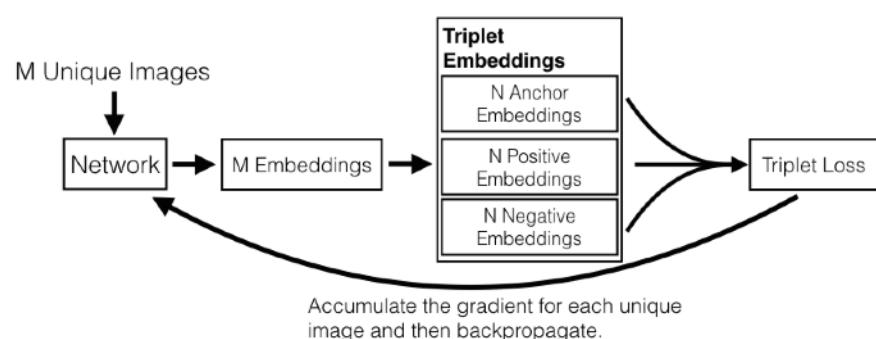
پردازش، تصاویر چهره ها به عنوان ورودی به یک شبکه عصبی پیچشی داده می شوند (شکل ۳-۳). این الگوریتم برای تعلیم شبکه از مجموعه داده کوچکی با  $500$  هزار تصویر چهره استفاده می کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری شده به نام FaceScrub و CASIA-WebFace بدست آمده است. شبکه مورد استفاده در این الگوریتم یک نسخه اصلاح شده از شبکه

الگوریتم 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 گانه



.[۴] OpenFace معماری

برای هر تصویر محاسبه شده و به عقب انتشار می یابد. در هر دسته کوچک،  $P$  تصویر برای هر نفر از  $Q$  نفر، در مجموعه داده انتخاب  $Q=15$  و  $P=20$  انجام پذیرد. در این مقاله از  $M \approx PQ$  تصویر به شبکه داده می شود تا عملیات forward استفاده شده است. تمام جفت های anchor-positive برای بدست آوردن سه گانه های  $N = Q \binom{P}{2}$  مورد استفاده قرار می گیرند. خطای سه گانه محاسبه شده و مشتق آن برای پس انتشار خطا استفاده می شود. شکل ۳-۴ چگونگی آموزش شبکه را نشان می دهد. مجموعه داده LFW یک معیار استاندارد برای سنجیدن میزان دقیقیت الگوریتم های تشخیص چهره می باشد که از ۱۳۲۳۳



.[۴] OpenFace یادگیری در معماری

تصویر چهره از ۵۷۵۰ شخص تشکیل شده است. الگوریتم OpenFace بر روی این مجموعه داده مورد سنجش قرار گرفت که به دقیقت

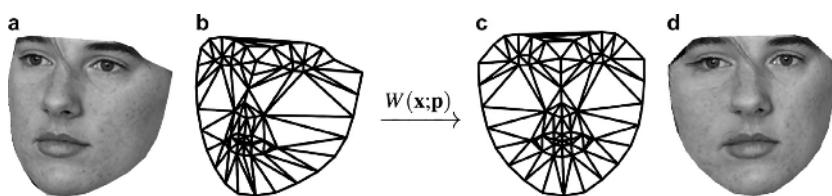
$0.9292 \pm 0.0134\%$  رسید. در سال ۲۰۱۶ Mohammad Haghigat و همکاران در [۲۵] یک روش برای هنجارسازی

حالت چهره بر اساس تنظیم کردن مدل ظاهری فعال 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 می‌

پذیرد. شکل ۵-۳ رویکرد کلی این الگوریتم را نشان می‌دهد. در این مدل یک تصویر چهره با مجموعه ای از نقاط ویژه هنجارسازی



شکل ۵.۲: رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۲].

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.

$$S = s_0 + \sum_{i=1}^n p_i s_i \quad (1.2)$$

که در آن  $s$ . شکل پایه،  $s$ ; نشان دهنده  $\alpha$  امین شکل پایه و  $p_1, p_2, \dots, p_n$  عامل های شکل می‌باشند. ظاهر مدل AAM

یک تصویر  $(x)$  می‌باشد که در آن  $x$  مجموعه پیکسل های داخل شکل پایه  $s$ . می‌باشد. مدل ظاهر یک چهره خاص از یک ظاهر

پایه  $a$ . و ترکیب خطی از بردارهای ویژه  $a_1, a_2, \dots, a_m$  تشکیل می‌شود که به صورت زیر تعریف می‌گردد.

$$A(x) = a_0(x) + \sum_{i=1}^m q_i a_i(x) \quad (2.2)$$

که در آن  $q_1, q_2, \dots, q_m$  عامل های ظاهر می‌باشند. عامل های شکل و ظاهر برای هر تصویر در فرایند AAM بدست می‌آید.

الگوریتم های POIC و SIC دو الگوریتم شناخته شده برای این منظور می‌باشند. رویکرد SIC نسبت به POIC در شرایطی که

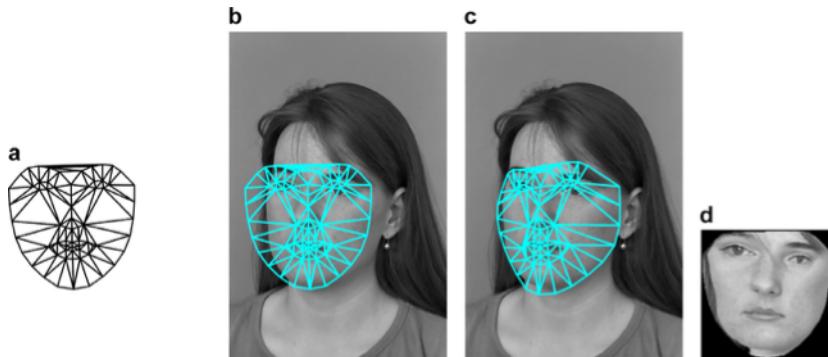
تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می‌کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می-

باشد. در این مقاله از یک روش SIC سریع برای حل مسئله بهینه سازی با  $100 \times 100$  تکرار استفاده شده است. اگر  $p_1 = p_2 = \dots, p_m$

مجموعه عامل های بدست آمده باشد، یک تبدیل هم نسبی قطعه ای  $W(x; p)$  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 نگاشت

پیدا می نمایند. برای مقداردهی اولیه از یک مدل پایه  $s$ . استفاده می شود که مقدار  $p$  در آن صفر می باشد (شکل ۶-۳ قسمت a).



شکل ۶.۲: مقدار دهی اولیه و بهینه سازی [۹] AAM.

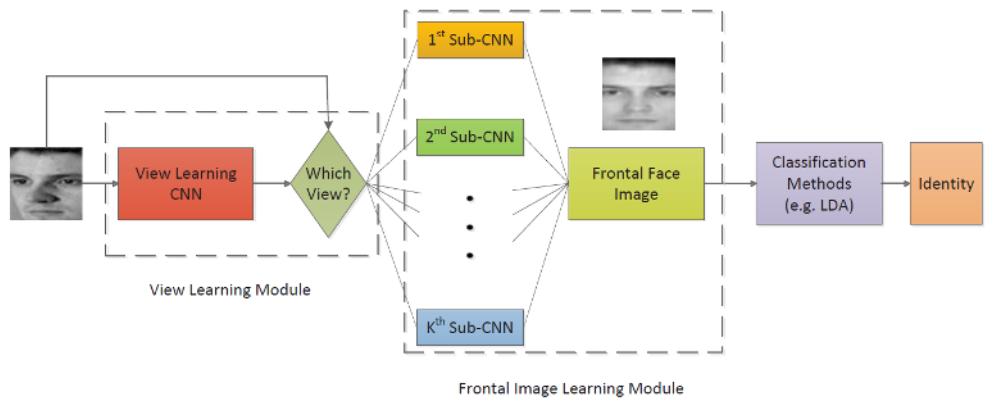
پس از تنظیم کامل مدل بر روی چهره، یک تبدیل هم نسبی با پارامترهای بدست آمده توسط الگوریتم یادگیری یاد شده، می تواند حالت چهره را هنجارسازی نماید. در بخش شناسایی چهره، ابتدا بخش چانه از تصویر حذف می شود زیرا چانه تقریباً تاثیری در شناسایی یک چهره ندارد. سپس تصویر چهره به اندازه  $64 \times 64$  پیکسل تبدیل می شود و به  $64 \times 8$  بخش غیر هم پوشان با اندازه تقسیم می شود. سپس در هر بخش تبدیل DCT 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.  $10$  ضریب بعدی که ضرایب فرکانس پایین می باشند، برای ایجاد بردار ویژگی چهره استفاده می شوند. برای آموزش و آزمایش از مجموعه داده LFW و FERET استفاده شده است که در آن تصاویر چهره با زوایای چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این مقاله موفق به دستیابی به شناسایی چهره با دقیق  $87.3\%$  شده است. در سال  $2016$  Zhang و همکاران در [۳۱] یک روش رو به رو



شکل ۷.۲: نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۹].

سازی چهره ارائه دادند که شناسایی چهره را مستقل از نمای چهره 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 بین یک چهره از رو به رو با یک چهره از یک

زاویه و نمای خاص می باشد (شکل ۳-۶). 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 به جای لایه های تمام متصل استفاده شده است. مدل VS2VI از دو بخش

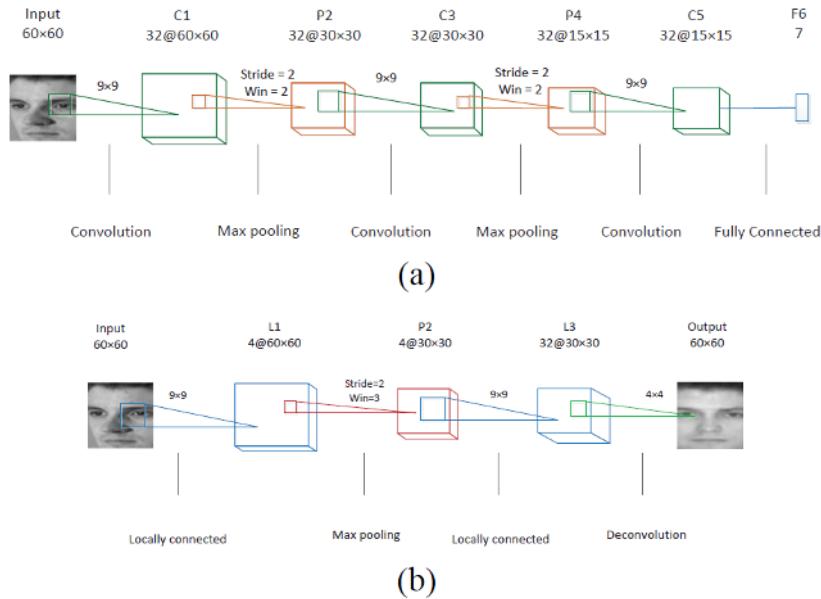


شکل ۸.۲: معماری شبکه پیشنهادی VS2VI.

اصلی تشکیل شده است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با ابعاد  $60 \times 60$  دریافت می کند و آن را با توجه به نمای چهره طبقه بندی می کند. سپس تصویر وارد بخش دوم می شود که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از رو به رو با یک چهره از زاویه و نمای خاص می باشد، تشکیل شده است. در این بخش چهره با نمای رو به رو بدست می آید و را مورد شناسایی قرار می دهیم تا هویت فرد مشخص شود. برای این منظور نیز از الگوریتم LDA 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 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 نمای های 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.

$$f(x) = \max(0, x) \quad (3.2)$$

بخش دوم از تعدادی زیر شبکه پیچشی که هر کدام برای یادگیری تناظر بین چهره از رو به رو با یک چهره از زاویه 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.



شکل ۹.۲: (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [؟].

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  $460$  هزار پارامتر نیاز دارد. لایه اول که اتصال محلی دارد، از تابع PreLU به عنوان تابع فعالیت استفاده کرده است. لایه واپیچشی برای نمونه افزایی از درون یابی دو خطی استفاده کرده و تابع هزینه آن  $loss - \ell_2$  می باشد. برای یادگیری شبکه از الگوریتم پس انتشار خطا استفاده شده است. الگوریتم VS2VI دقیق  $66\%$ . $95$  در تشخیص چهره با زاویه  $45$  درجه رسیده است. در سال  $2018$  Andrey V.Savchenko و همکاران در [۴۵] یک روش مبتنی بر ML 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 حد اکثر می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی Kullback–Leibler بین ویژگی‌های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده‌های IJB-A، LFW و YTF مورد آزمایش قرار گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقیق  $30.5$  تا  $5.5$  درصد در مقایسه با روش‌های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند.

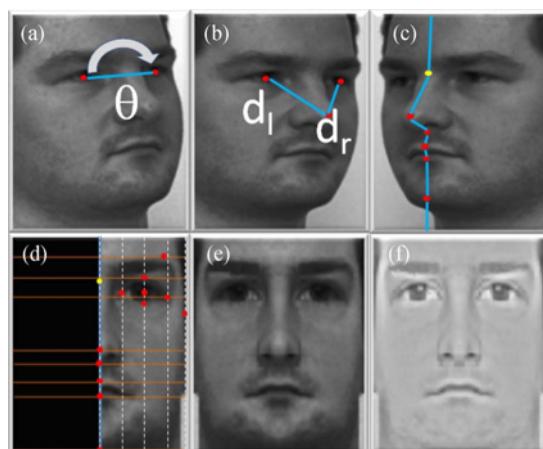
در سال  $2013$  Marsico و همکاران در [۲۸] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم STASM برای به دست آوردن  $68$  نقطه ویژه بر روی چهره استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه (SP) محاسبه می



شکل ۱۰.۲: مقایسه روش ارائه شده با سایر روش‌ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش‌های دیگر (d) خروجی روش ارائه شده [؟].

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 می‌شود. هرچه مقدار شاخص SP بالاتر باشد، تصویر چهره به حالت تمام رخ نزدیکتر است و اصلاح زاویه کمتری نیاز دارد.

شکل ۱۱-۳ قسمت a تا c معیارهای مورد نیاز برای محاسبه شاخص SP را نشان می‌دهد. چرخش: چرخش سر در جهت عقربه‌های



شکل ۱۱.۲: ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [؟].

ساعت یا عکس آن می‌باشد. و به صورت زاویه تعريف می‌شود که زاویه بین خط عبوری از مرکز چشم‌ها و محور افقی x می‌باشد.

$$roll = \min\left(\left|\frac{2\theta}{\pi}\right|, 1\right) \quad (4.2)$$

انحراف: چرخش در راستای محور افقی است و r و ۱ فاصله مرکز چشم چپ و راست از نوک بینی می‌باشد. اندازه گیری این فاصله ها در صورت برابر بودن، برای تشخیص تمام رخ بودن تصویر چهره مورد استفاده قرار می‌گیرد.

$$yaw = \frac{\max(d_l, d_r) - \min(d_l, d_r)}{\max(d_l, d_r)} \quad (5.2)$$

شیب: چرخش سر در راستای محور عمودی را اندازه‌گیری می‌کند.

$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)} \quad (6.2)$$

با محاسبه ۳ شاخص فوق، شاخص SP محاسبه می‌شود:

$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch) \quad (7.2)$$

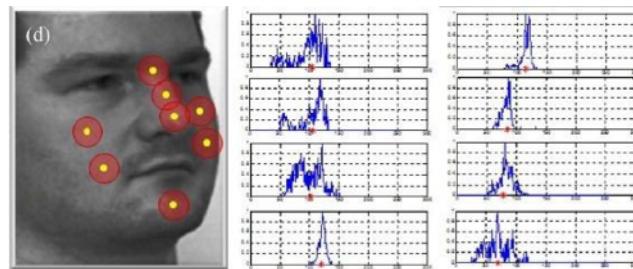
که در آن

$$\alpha + \beta + \gamma = 1 \quad (8.2)$$

که مقادیر این ضرایب از طریق آزمون و خطابه دست می‌آیند. سپس در مرحله تمام رخ کردن تصویر چهره، بین دو فاصله  $dr$  و  $dl$  هر کدام بزرگتر باشند، نشان می‌دهد آن سمت از چهره بیشتر در دید دوربین است. اگر نیمه سمت راست صورت به طرف دوربین باشد ( $dl \geq dr$ )، تصویر بدون تغییر باقی می‌ماند. در غیر این صورت، تصویر حول محور عمودی برعکس می‌شود که باعث می‌شود همیشه نیمه سمت راست تصویر پردازش شود. سپس برای ثابت کردن طول سطراها، سطراها بسط داده می‌شوند. 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 می‌شود و تصویر تمام رخ چهره به دست می‌آید.

## ۳.۲ چالش روشنایی

متعادل سازی بافت نگار یکی از الگوریتم‌های مهم در پردازش تصویر است که هدف آن افزایش وضوح تصویر با یکنواخت سازی بافت نگار تصویر است، به گونه‌ای که بخش‌های از تصویر که به علت روشنایی کم یا زیاد، پنهان می‌باشند، قابل مشاهده شوند. متعادل سازی بافت نگار قدرتمندترین و رایج‌ترین روش برای اصلاح روشنایی تصاویر است. اما ضعف این روش، سراسری بودن آن می‌باشد. برای رفع این مشکل باید از الگوریتم‌های محلی استفاده کرد. در سال ۲۰۱۳ Marsico و همکاران در [۲۸] یک روش هنجار‌سازی نورپردازی برای تصاویر چهره ارائه دادند و از این روش برای محاسبه شاخص روشنایی نمونه (SI) استفاده کردند. زمانی که تصویر روشنایی یکنواخت دارد، بیشتر بخش‌های 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، گونه‌ها و چانه معمولاً نور را منعکس می‌کنند. ناحیه در شکل ۱۲-۳ با توجه به چنین اصلی انتخاب شده‌اند. ۸ بافت نگار بارنگ آبی و مرکز آن‌ها با رنگ قرمز مشاهده می‌شود. ۸



شکل ۱۲.۲: اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [؟].

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

$$mc(w) = \frac{\sum_{i=0}^{255} i \times h_w(i)}{\sum_{i=0}^{255} h_w(i)} \quad (9.2)$$

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  $mc$  توسطتابع سیگموید  $F$  در بازه  $[1, 0]$  هنجارسازی می شوند و شاخص کیفیت روشنایی محاسبه می شود که یک عدد می باشد:

$$SI = 1 - F(std(mc)) \quad (10.2)$$

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  $8 \times 8$  مقدار هر پیکسل بر مقدار میانگین همسایگانش تقسیم می شود و نتیجه نهایی حاصل می شود. نتیجه به صورت قسمت  $\mathbb{F}$  در شکل می باشد. در سال ۲۰۱۵ Jamal Hussain Shah و همکاران در [۳۸] 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 ناحیه استفاده می شود.

۲. ویژگی های مبتنی بر LDA از تصویر چهره استخراج می شود.

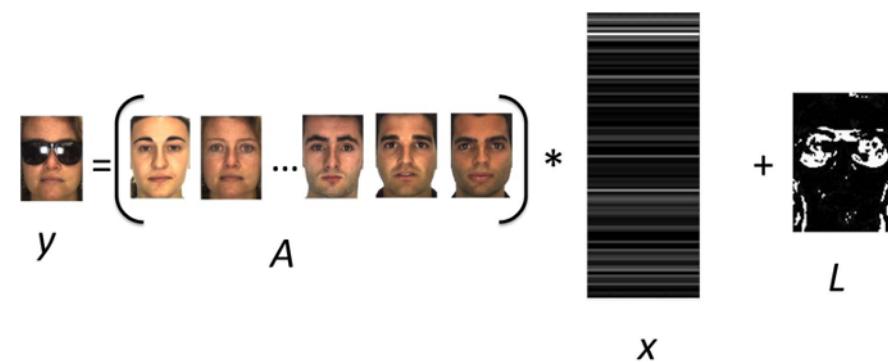
۳. فرایند طبقه بندی بر اساس مدل OPPM انجام می شود.

## ۴.۲ چالش انسداد

در سال ۲۰۱۸ Cho Ying Wu و همکاران در [۴۰] یک رویکرد مبتنی بر واپاکش 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 خطی  $L = \mathbb{R}^d$  نیز می باشد. (شکل ۱۶-۳)

$$y = Ax + L \quad (11.2)$$

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

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|s.t. y - Ax\| \quad (12.2)$$

که در آن  $\epsilon$  یک آستانه خطای باشد. برای تصویر و 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.

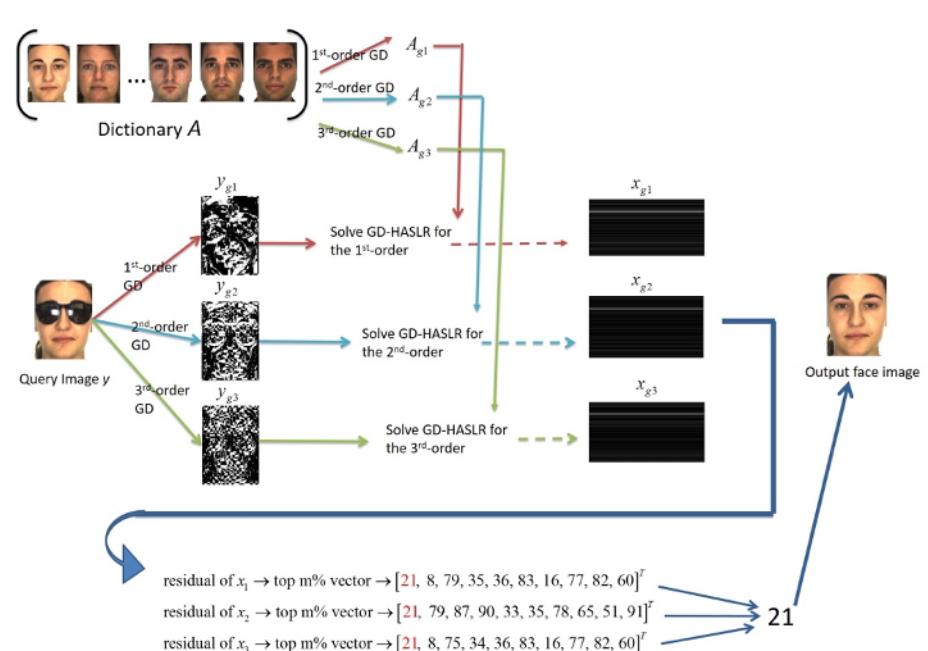
$$L(x, L, z) = \alpha LM + xi + zT(y - Ax - L) + 2y - Ax - L22 \quad (13.2)$$

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

محاسبه نمود:

$$r_i = y - Ai(x)2 \quad (14.2)$$

که در آن  $(x)_\delta$  نشان دهنده  $\delta$  امین انتخاب کننده دسته می باشد که فقط ورودی های مربوط به دسته  $\delta$  ام را حفظ می کند و در سایر قسمت ها برابر با صفر می باشد. در نهایت دسته ای که کمترین باقیمانده را داشته باشد، انتخاب می شود. رویکرد کلی الگوریتم در شکل ۱۴-۳ آمده است. در سال ۲۰۱۴ J. Li و همکاران در [۳۹] یک روش تشخیص چهره پوشیده شده در پس زمینه پیچیده ارائه



شکل ۱۴.۲: رویکرد کلی الگوریتم [۴].

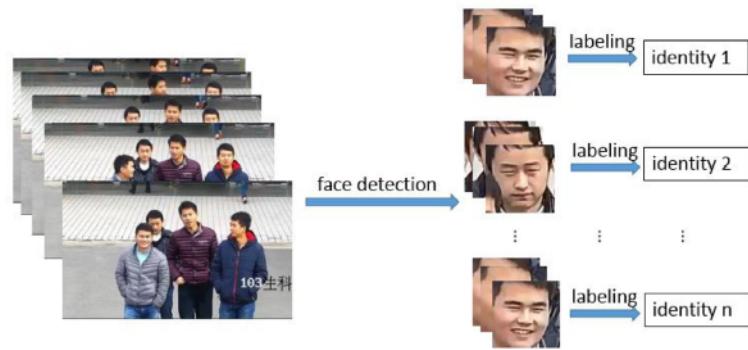
کردن. این الگوریتم از دو مرحله تشکیل شده است. در مرحله اول تعیین می کنند که آیا شی یک شخص می باشد یا خیر و در مرحله دوم بررسی می شود که آیا چهره پوشیده شده می باشد یا خیر و در صورت پوشش چهره، نوع پوشش و اینکه پوشیدگی با ماسک، کلاه،

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 بایین است، از LBP به همراه آن نیز استفاده کرده اند. در مرحله دوم از طبقه بند Adaboost چهره های پوشیده شده استفاده شده است که برای انواع پوشیدگی آموزش داده شده است.

## ۵.۲ 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 که در شرایط نورپردازی، زاویه و حالت های مختلفی تصویربرداری شده باشد. در مقابل فرض 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 پردازش ها بر روی تلفن همراه وجود دارد و الگوریتم ارائه شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد.

در سال ۲۰۱۷ Ya Wang [۴۱] روشی برای تشخیص چهره در دوربین های نظارتی در محیط بدون محدودیت به وسیله شبکه عصبی پیچشی عمیق ارائه دادند. از آنجایی که داده های آموزشی ورودی به مدل از اهمیت بالایی برای تشخیص برخوردار هستند و همچنین به تعداد زیادی از داده های هر دسته برای بهبود عملکرد سامانه نیاز است، نوآوری این رویکرد، ساختن یک مجموعه داده استاندارد برای شبکه عصبی از روی دوربین های نظارتی در محیط است که در چهار مرحله به صورت زیر ساخته می شوند. با توجه به اینکه تصاویر مورد نظر برای هر فرد در مجموعه فریم های پشت سر هم از یک دوربین موجود است، می توان مجموعه تصاویر یک فرد را بوسیله ترکیب الگوریتم تشخیص چهره و ردیابی چهره جمع آوری کرد. پس از شناسایی یک چهره، با ردیابی آن به وسیله الگوریتم KCF، مجموعه تصاویری از آن به عنوان یک دسته طبقه بندی می شود. برخی تصاویر در هر دسته به اشتباه در مرحله اول به عنوان تصویر یک فرد در نظر گرفته شده اند (شکل ۱۶-۳). با استفاده از روش خوش بندی گراف روی ویژگی های استخراج شده از شبکه VGG-Face، تشخیص و پاک سازی تصاویر اشتباه انجام می شود. فاصله کسینوسی بین ویژگی های تصاویر چهره محاسبه می شود و اگر این فاصله برای هر دو تصویر کمتر از یک مقدار آستانه باشد، این تصاویر متعلق به یک فرد هستند. با توجه به شکل ۱۷-۳

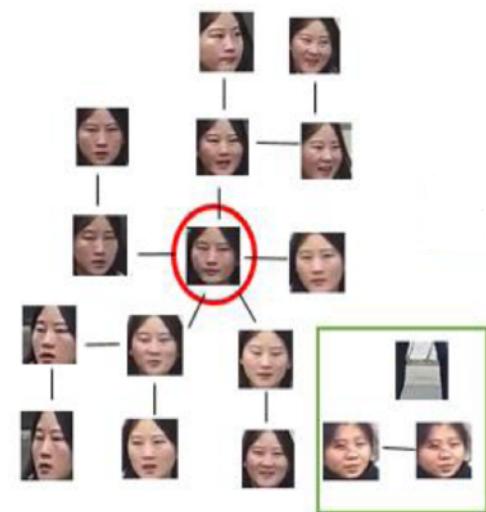


شکل ۱۵.۲: ردیابی، یافتن چهره ها و برچسب زنی [؟].



شکل ۱۶.۲: تصاویر با حاشیه قرمز، به اشتباه برچسب زنی شده اند [؟].

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



شکل ۱۷.۲: استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [؟].

بین هر داده با داده مرکزی و در نظر گرفتن یک آستانه، داده های تکراری در هر دسته مشخص شده و حذف می شوند. با توجه به

مقدار داده های درون هر دسته، تصفیه بین دسته ای انجام می شود. اگر مجموعه داده های درون هر دسته کمتر از ۱۰۰ تصویر باشد،

آن دسته از مجموعه داده حذف می شود. دقیت خوشه بندی و جمع آوری مجموعه داده ۲٪.۹۹ دشده است. در نهایت از یک مدل

پیش آموزش دیده شده شبکه VGG-Face برای طبقه بندی تصاویر آزمایشی استفاده شده است که به

دقت ۹۲٪ رسیده است.

مقاله [۴۶] از شبکه‌های مولد تخصصی برای تولید داده‌ها استفاده کرده است که به اختصار GAN نامیده می‌شوند. شبکه مستقل تولید کننده و تمیز دهنده استفاده شده است. شبکه تولید کننده از روی بردار  $Z$  که می‌تواند یک نویز تصادفی باشد، یک تصویر تولید می‌کند و شبکه تمیز دهنده وظیفه دارد تصاویر واقعی را از تصاویر تولید شده توسط شبکه تولید کننده تشخیص دهد. بنابراین هر تصویر با یک بردار  $Z$  معرفی می‌شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می‌شود. به عنوان مثال بردار  $Z$  برای تصویر خانمی که عینک آفتابی نزدیک است از بردار  $Z$  برای تصویر خانمی که عینک آفتابی نزدیک است، کم می‌شود و حاصل آن، بردار مربوط به یک عینک آفتابی می‌باشد. سپس این بردار با بردار تصویر آقایی که عینک نزدیک است جمع می‌شود. نتیجه نهایی تصویر همان آقا با عینک آفتابی می‌باشد. به عنوان مثالی دیگر، با میانگین گیری بردارهای مربوط به دو تصویر از چهره شخصی که به سمت راست و چپ متمایل است، توانسته چهره رو به روی شخص را بازسازی نماید. اما کیفیت کار هنوز تا حالت مطلوب فاصله دارد.



شکل ۱۸.۲: تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۴۷].

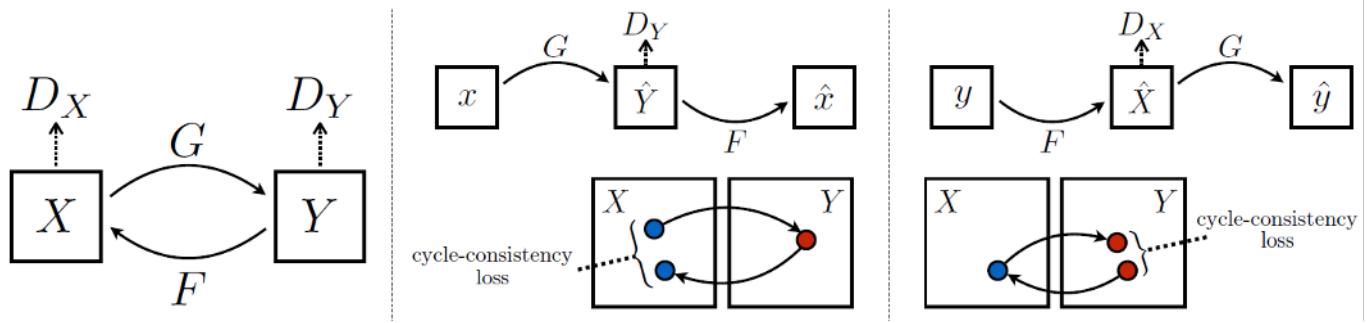
مقاله [۴۷] یک شبکه عمیق مبتنی بر GAN را با نام LR-GAN پیشنهاد می‌دهد، که تصاویر واقع گرایانه با وضوح بالا از روی تصاویر با وضوح پایین بازسازی می‌کند. این تصاویر چهره غیر واقعی اما واقع گرایانه و با کیفیت، باعث عملکرد بهتر سامانه شناسایی

چهره برای مقایسه تصاویر می‌شود. رویکرد اصلی مقاله در روش آموزش تخاصمی LR-GAN بهینه سازی تابع ضرر بازسازی چند مقیاسی است، بر اساس شاخص‌های مانند: شاخص شباهت ساختاری چند مقیاسی (SSIM)، میانگین مربعات خطای هر قسمت (PMSE)، واگرایی جنسن شانون اصلاح شده (JSD) و تنوع متقابل در اطلاعات (MVI). شبکه تمیز دهنده در LR-GAN، بر اساس اطلاعات طبقه‌بندی که به طور ضمنی در طول آموزش آموخته می‌شود، هویت هر شخص را حفظ می‌کند. این رویکرد سریعتر از شبکه‌های مبتنی بر GAN اخیر به یک همگرایی می‌رسد. این مدل که به دقت بالای ۹۰٪ رسیده است، رتبه اول را در ۴ مجموعه داده شرایط بدون محدودیت کسب کرده است.

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 عموماً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند. مقاله [۴۸] یک شبکه GAN جدید پیشنهاد می‌دهد که بخش تولید کننده آن به طور خودکار یاد می‌گیرد بدون هیچ ناظر انسانی ویژگی‌های بصری متفاوت تصاویر را از یکدیگر جدا نماید. پس از اتمام مرحله یادگیری، ما می‌توانیم این ویژگی‌های بصری را به دلخواه خود ترکیب نماییم. برای مثال ویژگی‌های اساسی مانند جنسیت، سن، طول مو، وجود عینک و زاویه چهره را از تصویر ۱ با ویژگی‌های دیگری از تصویر ۲ ترکیب کرد و یک چهره جدید تولید نمود. نگاه این شبکه تولید کننده به هر تصویر، مجموعه ای از ویژگی‌های بصری می‌باشد. هر ویژگی بصری با اندازه مشخص، جلوه‌های تصویر را کنترل می‌کند. ویژگی‌های بصری غالب مانند زاویه چهره، مو، شکل صورت؛ ویژگی‌های بصری میانی مانند فرم لب و چشم‌ها و ویژگی‌های سبک تر مانند رنگ. ما می‌توانیم این ویژگی‌های بصری را با ضرایب دلخواه خود ترکیب نماییم.

شکل ۳ - تصاویر هر دویف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند - [۷]

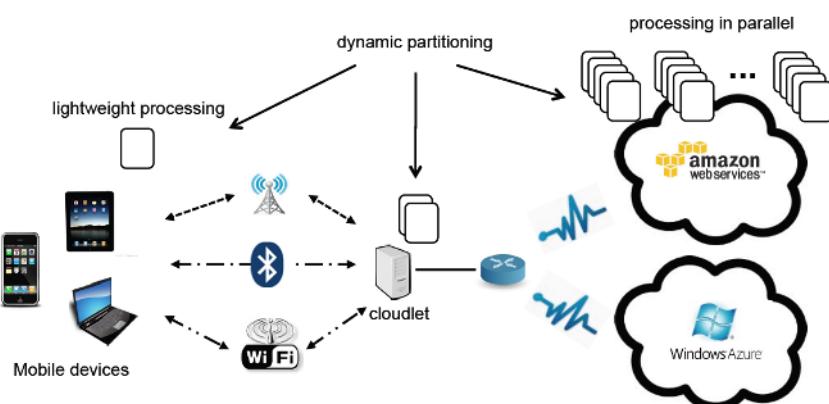
در مقاله [۴۹] به موضوع تولید چهره در زوایای دلخواه پرداخته شده است. در این مقاله از دو شبکه GAN استفاده شده است که در شبکه اول از روی چهره زاویه‌دار، چهره روبرو تولید شده است. سپس با استفاده از شبکه GAN دوم از روی تصویر چهره روبرو، تصویر با زاویه دلخواه با استفاده از یک پارامتر کنترلی تولید می‌شود. چالشی که در این مقاله به آن اشاره شده است، مسئله عدم توازن داده‌ها در وجود برخی ویژگی‌ها در تصاویر می‌باشد. این مقاله در برخی تصاویر چهره زاویه‌دار به مشکل برخورد می‌کرد. به عنوان مثال چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره 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 ۱۹.۲: ساختار شبکه AD GAN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۴].

## ۶.۲ چالش منابع محدود

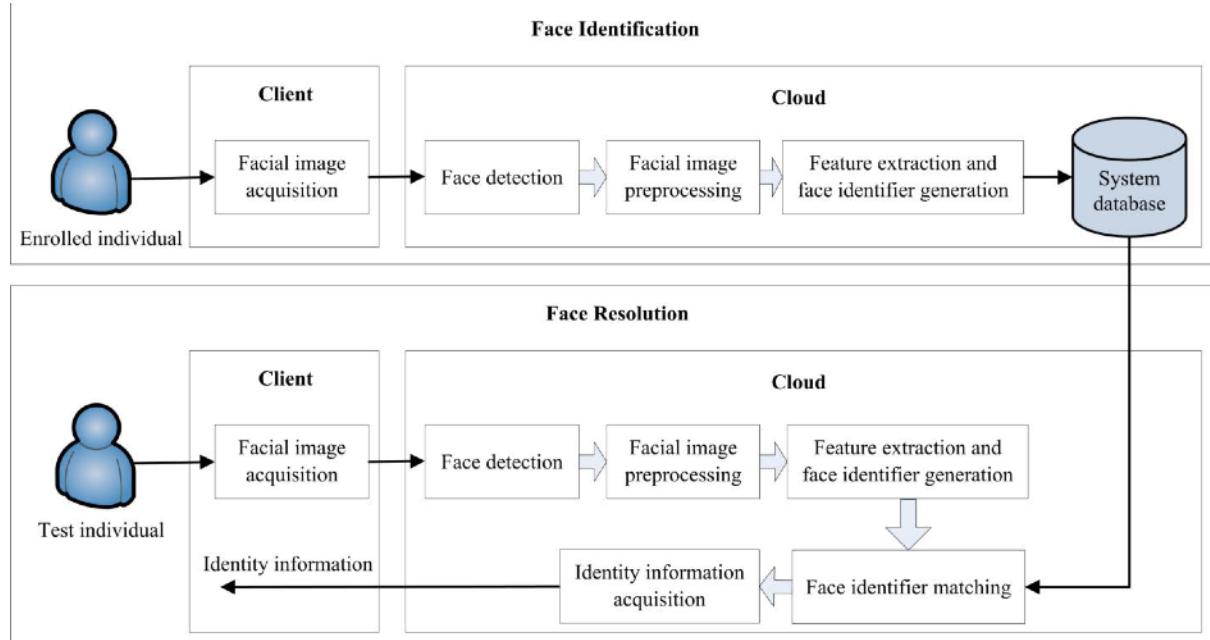
در سال ۲۰۱۲ Tolga Soyata و همکاران در [۴۲] یک روش تشخیص چهره بی درنگ مبتنی بر بینایی ابری با استفاده از معماری MOCHA ارائه کردند (شکل ۱۸-۳). با فراگیر شدن تلفن همراه هوشمند در میان شهروندان، سامانه تشخیص چهره می تواند از همکاری مشترک محاسبات تلفن همراه و رایانش ابری استفاده کند. چالش این سامانه، چگونگی تجزیه انجام وظیفه بین تلفن همراه و فضای ابری، توزیع بار محاسبه در میان سرورهای ابر برای به حداقل رساندن زمان پاسخ با توجه به تأخیر ارتباطات مختلف و قدرت محاسبه سرور 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



شکل ۲۰.۲: معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۴].

تقسیم می 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

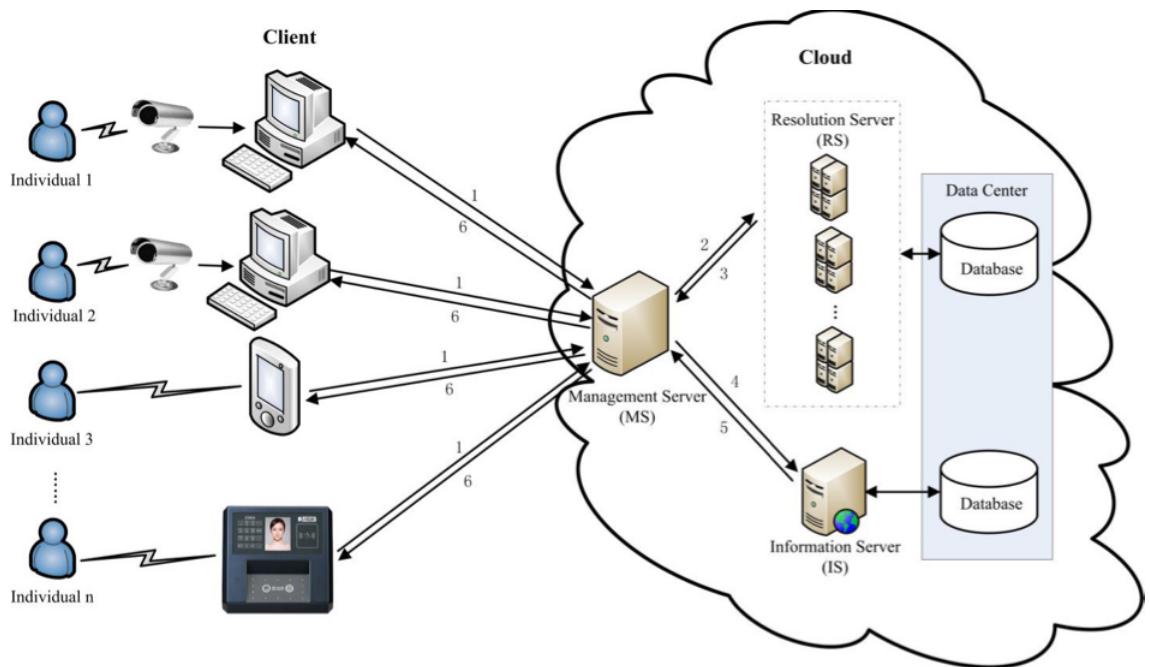
فقط اطلاعاتی همچون ویژگی های Haar و طبقه بندها به سرور فرستاده می شوند. پس از اتمام فرایند تشخیص چهره، نتیجه نهایی برای تلفن همراه فرستاده می شود. ابر کوچک : سرورها و رایانه هایی که توانایی پردازشی متوسطی دارند، ابر کوچک یا cloudlet نامیده می شوند. این دستگاه ها که به عنوان میان دستگاه های همراه و سرورهای 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 با محاسبات سنگین سامانه را به دوش می کشند و تصمیم گیری نهایی بر روی آن انجام می پذیرد. در سال ۲۰۱۸ Pengfei Hu و همکاران در [۴۳] 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 بپسند 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 ابر انجام می شود.



شکل ۲۱.۲: نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۹].

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 بپسند جزئی حاصل شود. سپس از الگوریتم LBP 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 سرور مدیریت، سرور اطلاعات، (IS) سرور شناسایی (RS) و پایگاه داده تشکیل شده است. به علت قدرت بالای پردازش در سرور ابری، امکان پردازش موازی نیز در این سامانه وجود دارد که باعث افزایش سرعت محاسبات و کاهش زمان پاسخ دهی سامانه می گردد. علاوه بر



شکل ۲۰.۲: چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۹].

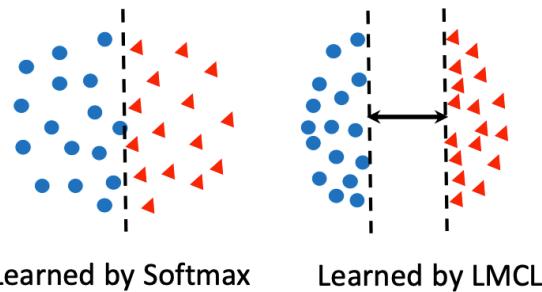
رویکردهای بالا که هر یک بر روی حل یک مسئله خاص تمرکز کرده بودند، برخی روش‌هایی که اخیراً معرفی شده اند، سعی بر این داشته اند که یک راه حل نسبتاً همه جانبه در مورد مسئله تشخیص چهره و مشکلات آن ارائه دهند. یکی از این رویکردها، استفاده از تابع ضرر CosFace می‌باشد که در سال ۲۰۱۸ Wang و همکاران در [۱] ارائه دادند. این تابع ضرر کسینوسی با حاشیه زیاد<sup>۱</sup> شباهت بسیار زیادی به تابع ضرر softmax دارد. با این تفاوت که به جای ضرب ماتریس ضرب های  $W$  در بردار ویژگی  $x$ ، حاصل ضرب مقادیر موجود در ویژگی های استخراج شده و آخرین لایه کامل متصل را به صورت  $W_j^T x_i = ||W_j|| ||x_i|| \cos(j)$  تبدیل می‌کند، که  $\theta_j$  زاویه بین وزن  $W_j$  و ویژگی  $x_i$  است.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} - m))}}{e^{s(\cos(\theta_{y_i} - m))} + \sum_{j=1}^n e^{s(\cos(\theta_{j,i}))}} \quad (15.2)$$

همانطور که در شکل ۲۳.۲ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می‌کند اما در مرزهای تصمیم

<sup>1</sup>Large Margin Cosine Loss

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



شکل ۲۳.۲: تابع ضرر CosFace حاشیه بیشتری نسبت به SoftMax در مز بین دسته‌ها ایجاد می‌نماید [۱].

## ۷.۲ نتیجه گیری

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

دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چلب نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و

ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده‌های آموزشی بسیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویرپردازی شده باشد.

مقاله [۲۵] روشی 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 چهار اعوجاج‌هایی می‌نماید که روند تشخیص چهره را 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 [۴۶] روشی مبتنی بر GAN برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی موقعیت‌ها به تصویر چهره لطمه وارد می‌نماید به طوری که شخص مورد نظر قابل شناسایی توسط سامانه یادگیری نمی‌باشد. مقاله [۴۷] 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 [۴۹] چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره روبرو بدون عارضه تولید شده است. این چالش به خاطر کمبود تصاویر با عارضه پوستی در مجموعه داده می‌باشد و شبکه در مواجه با این مسئله راه کاری ارائه نمی‌دهد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند. به تازگی یادگیری عمیق در تشخیص چهره و بسیاری از زمینه‌های هوش مصنوعی به راه حل غالب تبدیل شده است. مایک سوال مطرح می‌کنیم: آیا یادگیری عمیق واقعاً مسئله تشخیص چهره را حل می‌کند؟ چالش روش‌های یادگیری عمیق در تشخیص چهره چیست؟ در مقایسه با تشخیص شیء عمومی، تشخیص چهره به دلیل طیف گسترده‌ای از تغییرات در ظاهر چهره‌ها چالش برانگیز است. نورپردازی کنترل نشده، انسداد ناشی از عینک، مو، ریش، کلاه و...، تاری خارج از تمرکز دوربین، کیفیت پایین تصویر، بالا رفتن سن افراد و کمبود داده‌های آموزشی از مواردی می‌باشد که می‌توانند سامانه تشخیص چهره را با مشکل 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 کرده اند که لایه‌های عمیق CNN‌ها می‌توانند اطلاعات انتزاعی بیشتری مانند هویت، ظاهر و ویژگی‌ها را رمزگذاری کنند؛ با این حال هنوز هنوز کاملاً

مطالعه نشده است که لایه‌ها دقیقاً با ویژگی‌های محلی برای تشخیص مطابقت دارند. برای شناسایی چهره، عملکرد یادگیری را می‌توان با یادگیری یک معیار اندازه‌گیری فاصله متمایز کننده بهبود داد. با این حال، با توجه به محدودیت‌های حافظه کارت گرافیک ها، نحوه انتخاب جفت‌های 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.

## منابع و مأخذ

- [1] 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.
- [2] 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.
- [3] E. Hjelmås and B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding*, vol.83, no.3, pp.236–274, 2001.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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*.
- [11] JafriRabia and A. R., “A survey of face recognition techniques,” *Journal of Information Processing Systems*, vol.5, no.2, pp.41–68, 2009.
- [12] 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.

- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [22] 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.
- [23] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, “Facial landmark detection with tweaked convolutional neural networks,” 2016.
- [24] 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.
- [25] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, “Fully automatic face normalization and single sample face recognition in unconstrained environments,” *Expert Systems with Applications*, vol.47, pp.23–34, 2016.
- [26] 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.
- [27] 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.

- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” 2014.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” *Pattern Recognition*, vol.66, pp.144–152, 2017.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.

- [42] 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.
- [43] 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.
- [44] L. Di Martino, J. Preciozzi, F. Lecumberry, and A. Fernández, “Face matching with an a contrario false detection control,” *Neurocomputing*, vol.173, 08 2015.
- [45] 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.
- [46] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2016.
- [47] S. Banerjee and S. Das, “Lr-gan for degraded face recognition,” *Pattern Recognition Letters*, vol.116, pp.246–253, 2018.
- [48] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” 2019.
- [49] 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.

## **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