

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ





دانشکده مهندسی



دانشگاه فردوسی مشهد  
دانشکده مهندسی - گروه 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

مهرماه ۱۳۹۷



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 برابر هر که بخواهد موهبت می‌فرماید. رحمت و لطف او را بینهایت سپاس می‌گوییم چرا که فهم و درک مطالب این پژوهش را برابر 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. به راستی 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

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

در پیاده سازی این سامانه که باید در مکان‌های عمومی، معابر پیاده و محیط‌های کنترل نشده مورد استفاده قرار بگیرد، چند چالش مهم مانند نورپردازی غیر یکنواخت، انسداد، تاری خارج از تمرکز 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، داده‌های محدودی از هر دسته در اختیار داریم. برای مقابله با این چالش با استفاده از شبکه‌های مولد تخاصمی 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

## فهرست جداول

## فهرست تصاویر

## فهرست نمادها

فصل ۱

مقدمه

## ۱.۱ مقدمه

استفاده روز افزون از سامانه‌های تشخیص هوشمند چهره موجب اهمیت این شاخه از علم هوش مصنوعی شده است. عملکرد این سامانه‌ها در شرایط کنترل شده و آزمایشگاهی به حد مناسبی از بلوغ رسیده است. اما تشخیص چهره در شرایط کنترل نشده، موضوعی چالش برانگیز و در حال پیشرفت می‌باشد. در شرایط مختلفی مانند تابش نور غیر یکنواخت، زاویه نامناسب چهره در مقابل دوربین، وضوح پایین حسگر و... گاهی چهره‌ای یافت نمی‌شود و یا چهره یافت شده، قابل شناسایی نمی‌باشد. این مشکلات در سامانه‌های تشخیص چهره مبتنی بر ویدیو، به دلیل عدم ثبات شرایط محیطی و انسانی، تاثیر بیشتری داشته و در نتیجه، باعث کاهش دقت سامانه در تشخیص افراد می‌شود. در این بخش به منظور آشنایی کلی با موضوع مورد پژوهش، ابتدا توضیح مختصراً در مورد ویژگی‌های بیومتریک انسان و ارزیابی آن‌ها آورده شده است. پس از آن به طور خاص بر روی ویژگی بیومتریک چهره متمرکز شده و در مورد کاربردها، انواع، مراحل، بخش‌ها، مزایا و چالش‌های آن به طور کامل بحث شده است. سپس به تعریف یک مسئله خاص در زمینه تشخیص چهره به صورت بی‌درنگ و در محیط کنترل نشده پرداخته و نیازها و ابزارهای مورد نیاز بررسی شده است.

## ۲.۱ ویژگی‌های بیومتریک

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

### ۱.۲.۱ ارزیابی ویژگی‌های بیومتریک انسان

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

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

- یکتایی: آن ویژگی بیومتریک باید برای هر شخصی منحصر بفرد باشد.
- دوام: معیاری برای سنجش آنکه یک ویژگی بیومتریک چه مدت بدون تغییر باقی می‌ماند.
- ارزیابی: ویژگی بیومتریک مورد نظر باید سادگی کافی را در استفاده برای ارزیابی نمونه‌های متفاوت داشته باشد.
- کارایی: استفاده از ویژگی بیومتریک مورد نظر باید دقت، سرعت و پایداری مطلوب داشته باشد.
- مقبولیت: فناوری استفاده از ویژگی بیومتریک مورد نظر باید در میان جامعه پذیرش شده باشد.
- تصدیق هویت: ویژگی فرد به سامانه ارسال شود و سامانه پاسخی مثبت یا منفی برای تصدیق هویت فرد ارائه نماید.
- تشخیص هویت: ویژگی فرد به سامانه ارسال شود و سامانه با جستجو در پایگاه داده، هویت فرد را استخراج نماید.

### ۳.۱ سامانه تشخیص چهره

تشخیص چهره همواره یکی از موضوعات مورد مطالعه در علوم کامپیوتر و هوش مصنوعی بوده است. هدف این است که عکسی به کامپیوتر داده شود و سامانه تشخیص دهد که آیا چهره‌ای در تصویر مشاهده می‌کند یا خیر، و در صورت وجود یک چهره، در صورت امکان آن را شناسایی نماید. پس از موفقیت سامانه شناسایی از طریق اثر انگشت در چند سال اخیر، فناوری تشخیص چهره یکی از مهم‌ترین فناوری‌های بیومتریک برای شناسایی افراد محسوب می‌شود. این اهمیت و توسعه کاربرد به دو دلیل مهم می‌باشد:

۱. تشخیص چهره برای استفاده در کاربردهای مختلف مانند کاربردهای امنیتی، قابلیت شناسایی خودکار سریع و بدون دخالت

شخص را دارد، سرعت پردازش را بالا برد و خطرا کاهش داده است.

۲. سامانه تشخیص چهره نسبت به سامانه‌های بیومتریک قابل اعتمادی مانند تشخیص اثر انگشت و عنبیه چشم، ارتباط راحت

تری با کاربر ایجاد کرده و بدون تماس عضوی از بدن با سامانه، عملیات تشخیص انجام می‌گیرد. البته توسعه کاربردهای

دوربین‌های دیجیتالی پیشرفته عامل موثری در توسعه و بالا رفتن طرفداران این سامانه بوده است.

سامانه تشخیص چهره بر اساس الگوریتم‌های شناسایی و مقایسه تصاویر کار می‌کند. اساس و پایه این الگوریتم‌ها شناسایی و تجزیه و تحلیل ویژگی‌های مربوط به اندازه، شکل و موقعیت چشم، بینی، گونه‌ها و اعضای چهره می‌باشد. تصاویر رقمی برای سامانه

ارسال می‌شود و سامانه به طور خودکار چهره شخص را در تصویر پیدا می‌نماید و ویژگی‌های آن را استخراج و با نمونه‌های دیگر مقایسه می‌کند. نتیجه این پردازش، لیستی از تصاویر است که رتبه بندی شده است.

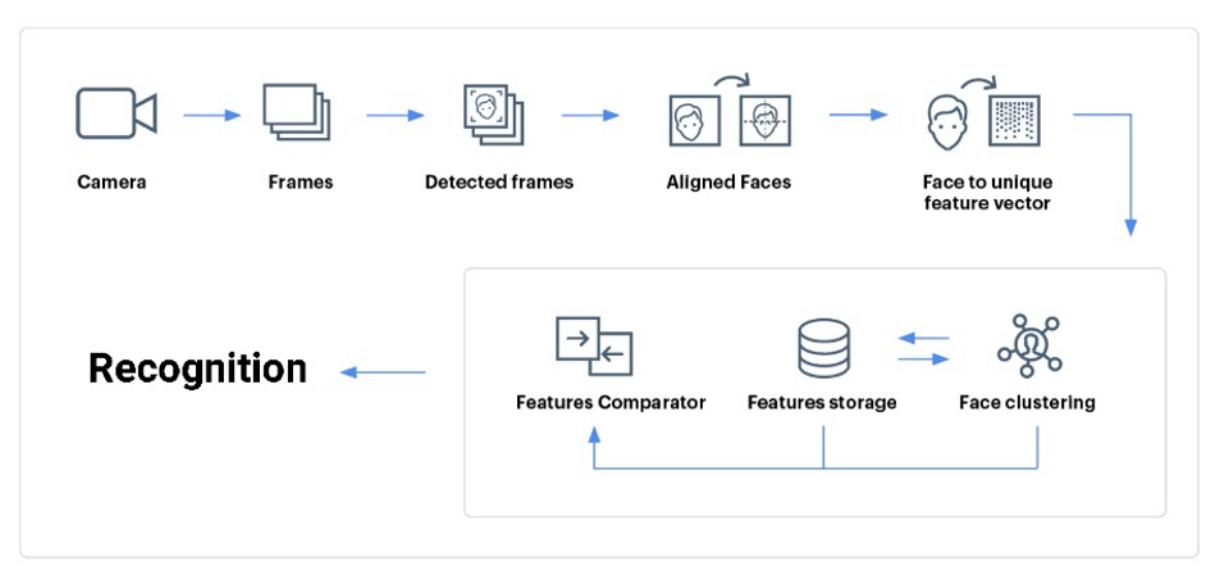
### ۱.۳.۱ کاربردها و ویژگی‌های مهم سامانه تشخیص چهره

فناوری تشخیص چهره که دارای مزایایی چون دقت بالا و سطح پایین دخالت فرد می‌باشد، در مواردی مانند کنترل دسترسی، امنیت اطلاعات، اجرا و نظارت بر قانون، شناسایی مجرمین، کنترل و ثبت تردد در سامانه‌های حضور و غیاب، کنترل نامحسوس و ایجاد امنیت در بانک، فروشگاه، فرودگاه و... مورد استفاده قرار می‌گیرد و در صنعت و علم مورد توجه قرار گرفته است. علاوه بر کاربردهای فوق، شناسایی و پردازش چهره کاربردهای دیگری هم دارند که ارتباطی با تشخیص هویت ندارند. دنبال کردن خط دید چشم و تعیین نژاد، جنس، سن و حالت صورت از جمله این کاربردها هستند که بعضی از آن‌ها در ارتباط بین انسان و کامپیوتر مفید هستند. کاربردهای زیادی برای مباحث شناسایی چهره می‌توان متصور شد که محدوده وسیعی از تصاویر متحرک تا تصاویر ثابت و از کاربردهای امنیتی تا کاربردهای تجاری را شامل می‌شود. این کاربردها را بر اساس نوع تصاویری که استفاده می‌کنند، می‌توان به دو گروه تصاویر ثابت و متحرک تقسیم کرد که در مواردی همچون کیفیت تصویر، زمینه تصویر، در دسترس بودن معیار انطباق و... با یکدیگر تفاوت دارند. دو ویژگی مهم یک سامانه تشخیص چهره عبارتند از: سرعت تشخیص: بدین معنا که یک الگوریتم تشخیص چهره از لحظه قرارگیری فرد در مقابل دوربین، در چه بازه زمانی می‌تواند هویت فرد درون تصویر را تشخیص دهد. دقت تشخیص: بدین معنا که یک الگوریتم تشخیص چهره با چه ضریب اطمینانی می‌تواند هویت یک فرد درون تصویر را تشخیص دهد. هرچه تعداد افراد مختلفی که در پایگاه داده ثبت نام شده اند، بیشتر باشد، احتمال خطا در سامانه بیشتر می‌شود و به یک الگوریتم دقیق تر نیاز داریم. بین سرعت تشخیص و دقت تشخیص، بده بستان وجود دارد. یک الگوریتم کارا باید هر دو ویژگی بالا را در نظر بگیرد.

### ۲.۳.۱ نمای کلی یک سامانه تشخیص چهره

یک سامانه بیومتریک تشخیص چهره شامل بخش‌های مختلفی می‌باشد که در شکل ۱-۱ نشان داده شده است. پنج بخش مهم یک سامانه تشخیص چهره عبارتند از: حسگر دوربین: این بخش وظیفه گرفتن تصویر چهره را بر عهده دارد. دستگاه گیرنده بسته به نیاز و کاربرد می‌تواند یک دوربین سیاه و سفید، رنگی، یک دوربین مخصوص با قابلیت استخراج اطلاعات عمق یا یک دوربین مادون قرمز باشد. یافتن چهره: تصاویر ورودی به این بخش ابتدا مورد پیش پردازش قرار می‌گیرند. سپس ارزیابی محتوایی شده و داده‌های

نامریوط از قبیل پس زمینه، موها و گردن و شانه و... حذف می‌شوند و تنها ناحیه چهره باقی می‌ماند. استخراج اطلاعات: در این بخش ویژگی‌های چهره مورد بررسی قرار گرفته و اطلاعات مورد نیاز از تصویر استخراج می‌گردد تا با تصاویر موجود در پایگاه داده مقایسه گردد. پایگاه داده: این بخش وظیفه ثبت نام، نگهداری و واکنشی ویژگی چهره کاربران را بر عهده دارد. پایگاه داده مجموعه‌ای از تصاویر است که در مرحله طبقه‌بندی مورد استفاده قرار می‌گیرد. در بیشتر روش‌های تشخیص چهره چندین نمای متفاوت از یک شخص در حالت‌های مختلف روحی خنده، اخم، عصبانیت، عادی و یا با عینک از کاربر گرفته می‌شود که موجب بالا رفتن ضربه شناسایی سامانه می‌شود. طبقه‌بندی: در این بخش ویژگی‌های استخراج شده با ویژگی‌های موجود در پایگاه داده تصاویر مقایسه می‌گردد و مشخص می‌شود که آیا چهره مورد نظر در بین چهره‌های موجود می‌باشد یا خیر، و در صورت مثبت بودن جواب، هویت شخص را تایید می‌کند. بر اساس امتیاز بدست آمده از مقایسه که همان درصد تطابق بردار ویژگی گرفته شده با بردارهای ویژگی موجود می‌باشد، چهره مورد نظر مورد تایید قرار گرفته و یا پذیرفته نمی‌شود.



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

## ۴.۱ الگوریتم‌های استخراج ویژگی و برچسب گذاری تصاویر

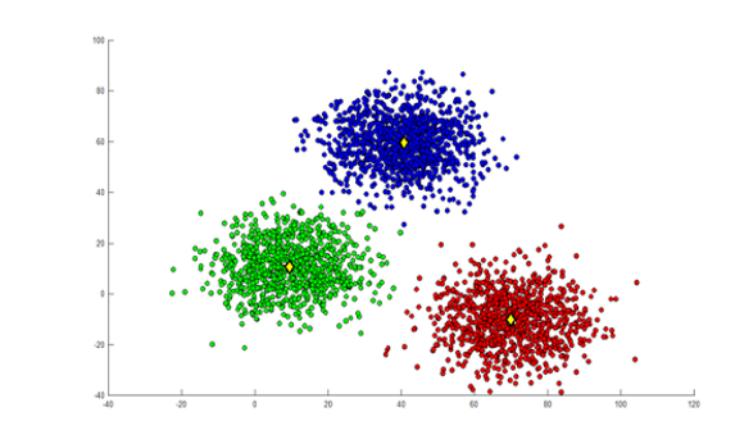
سامانه شناسایی تصویر یکی از مهم‌ترین بخش‌های یک سامانه بینایی ماشین به حساب می‌آید. این سامانه با استفاده از یادگیری عمیق، انواع اطلاعات در تصاویر مانند افراد، متون، اشیا و سایر اطلاعات موجود را شناسایی می‌کند. الگوریتم‌های مورد استفاده در فرایند تشخیص تصاویر بر مبنای استفاده از ویژگی‌های بصری تصویر مانند لبه و رنگ و یا ویژگی‌های استخراج شده از شبکه عصبی

عمیق به همراه اعمال الگوریتم‌های مختلف خوشبندی، واپاش یا طبقه‌بندی داده‌ها است که سعی در شناسایی و تشخیص تصاویر مختلف دارند. عنصر اصلی این گونه روش‌ها، داده‌های مورد استفاده در فرایند یادگیری می‌باشد. نوع و فراوانی داده‌ها در افزایش دقیق سامانه‌های تشخیص تصاویر مبتنی بر شبکه عصبی بسیار مهم می‌باشد.

#### ۱.۴.۱ خوشبندی

روش خوشبندی بر اساس محاسبه میزان و معیار شباخت در داده‌ها، آن‌ها را در خوشه‌های مختلف قرار می‌دهد. به طور کلی دو نوع روشن خوشبندی وجود دارد: خوشبندی سخت: برچسب گذاری به صورت صفر و یک بر روی داده‌ها می‌باشد و مشخص می‌کند داده مورد نظر مربوط به خوشه می‌باشد یا خیر. خوشبندی نرم: به خوشبندی فازی معروف است و میزان تعلق یک داده به خوشه‌ها را مشخص می‌کند و هر داده می‌تواند با توجه به وزن‌های اختصاص داده شده به آن، به چندین خوشه تعلق داشته باشد. داده‌های استفاده شده در فرایند خوشبندی، بدون برچسب بوده و یادگیری بدون ناظر در فرایند خوشبندی داده‌ها استفاده می‌شود.

شکل ۱-۲ نمونه‌ای از خوشبندی سخت و یافتن شاخص برای هر خوشه را نشان می‌دهد.

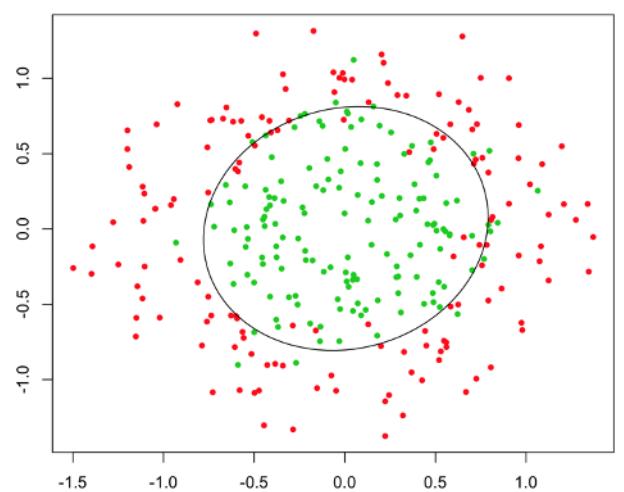


شکل ۱-۲: خوشبندی سخت و یافتن شاخص نمونه با توزیع گوسی بر روی داده‌ها [۹]

#### ۲.۴.۱ طبقه‌بندی

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

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



شکل ۳.۱: یک طبقه‌بند غیر خطی ساده [؟]

#### ۳.۴.۱ مسئله تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده

این پایان نامه با محوریت موضوع تشخیص بی‌درنگ چهره افراد در محیط‌های کنترل نشده با در نظر گرفتن شرایط سخت می‌باشد. هدف ما، طراحی و ساخت یک عینک هوشمند مانند شکل ۴-۱ برای افراد نابینا می‌باشد. هنگامی که شخص نابینا عینک را بر روی چشم‌مانش قرار داده و در محیط‌های عمومی راه می‌رود، دوربینی که بر روی عینک نصب شده است، شروع به بررسی چهره افرادی می‌کند که در زاویه دید آن قرار دارند. در صورت یافتن یک چهره آشنا، فرد مورد نظر شناسایی شده و نام فرد برای شخص نابینا خوانده می‌شود. این مسئله شامل دو بخش اصلی یافتن چهره و شناسایی چهره می‌شود. هریک از این بخش‌ها و زیر‌بخش‌های آن‌ها در فصل دوم مورد بررسی قرار خواهند گرفت.



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

#### ۴.۴.۱ چالش‌های سامانه تشخیص چهره

سامانه‌های تشخیص چهره به سطح مشخصی از بلوغ رسیده اند، اما توسعه آن‌ها در شرایط کنترل نشده و کاربردهای واقعی هنوز مسیر طولانی در پیش دارد. برای مثال، تشخیص چهره در ویدیویی در محیطی با تغییرات شدید نورپردازی و حالت چهره، انسداد صورت، وضوح پایین تصویر و...، و ردیابی آن در فریم‌های ویدیو با در نظر گرفتن تناظر بین فرمی و... مشکل می‌باشد. دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چسب نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. در ادامه مهم ترین چالش‌ها بر شمرده شده است. نورپردازی: روشنایی محیط در شب و روز و یا در محیط داخلی و خارجی به شدت تغییر می‌کند. با توجه به ساختار سه بعدی چهره، یک منبع نور مستقیم می‌تواند سایه‌های قوی بر روی چهره ایجاد کند که برخی ویژگی‌های چهره را تغییر می‌دهد. همانطور که در شکل ۱-۵ نشان داده شده است که گاهی تفاوت‌های ظاهری ناشی از روشنایی، بیشتر از تفاوت بین چهره افراد مختلف می‌باشد.



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

حال: زاویه چهره نسبت به حسگر دوربین، می‌تواند سامانه را با چالش مواجه نماید. چهره با زاویه تندر و چهره‌های نیم رخ مانند شکل ۱-۶ باید برای الگوریتم تشخیص چهره، قابل شناسایی باشند. بنابراین ویژگی‌های استخراج شده توسط الگوریتم باید به گونه‌ای باشند که در هر زاویه‌ای امکان استخراج آن‌ها وجود داشته باشد.



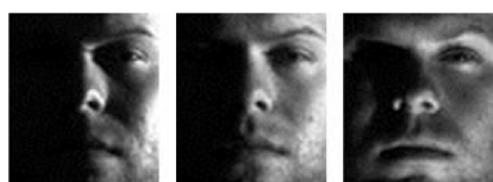
شکل ۱.۶: زاویه شدید چهره نسبت به دوربین باعث کاهش دقت سامانه می‌شود [؟]

تاری خارج از تمرکز: اگر عمق میدان دوربین کم باشد و چهره‌ها با فاصله‌های مختلف از دوربین باشند، مشکل تاری خارج از تمرکز رخ خواهد داد (شکل ۱-۷). اگر دوربین طوری تنظیم شود که چهره نزدیک‌تر، واضح‌تر دیده شود، در مقابل باعث می‌شود که چهره دورتر، مقداری مات شود و برعکس. این مسئله می‌تواند برای سامانه تشخیص چهره دردرس ساز شود.



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

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



شکل ۱.۸: انسداد شدید در اثر نورپردازی [؟]

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

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

منابع محدود: در صورت اجرای پردازش‌های سامانه توسط تلفن همراه، باید محدودیت منابع را در نظر گرفت. تلفن همراه نسبت به رایانه، دارای قدرت پردازنده پایین‌تر و منبع انرژی محدودتر می‌باشد که باعث می‌شود الگوریتم‌هایی با پیچیدگی محاسباتی بالا بر روی این دستگاه‌ها قابل اجرا نباشد. بنابراین الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد. زمان: یکی از چالش‌های موجود، فضایی پر از چهره‌های مختلف در مکان‌های عمومی و در مقابل، نیاز به واکنش سریع توسط سامانه است. مشابه شکل ۹-۱ در فضاهای عمومی و معابر پیاده مردم با سرعت از کنار دوربین عبور می‌کنند و سامانه باید قابلیت تشخیص چهره آن‌ها در چند ثانیه را داشته باشد. اگر کاربر سامانه با افراد جدید دیدار داشته باشد، سامانه باید به سرعت یاد بگیرد که چهره افراد جدید را تشخیص دهد.



شکل ۹.۱: نمونه‌ای از یک سامانه تشخیص چهره بی‌درنگ در محیط کنترل نشده [؟]

فصل ۲

## مروری بر کارهای گذشته

## ۱.۲ 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 بباید. 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 [\[۱\]](#) و همکاران در یک دسته بندی برای روش‌های یافتن چهره ارائه کردند که در ادامه شرح داده می‌شود.

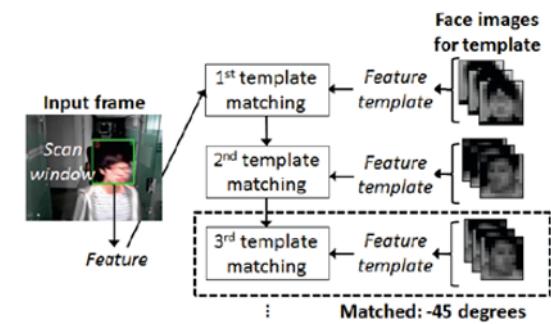
## ۱.۲.۲ رویکردهای مبتنی بر دانش

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 مو بر روی پیشانی ریخته باشد یا در حضور عینک، به درستی عمل نمی‌کند. یا به عنوان مثال:  $\square$  هر چهره داری دو فرو رفتگی برای چشم‌ها است و چیزی شبیه به ابرو روی این فرو رفتگی‌ها قرار دارد.  $\square$  صورت شامل بینی، چشم‌ها و دهان در فاصله‌ها و موقعیت‌های خاصی با یکدیگر می‌باشد.  $\square$  چهره مانند یک ناحیه کوچکتر است که بر روی یک ناحیه بزرگتر مانند شانه‌ها قرار گرفته است.  $\square$  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.

## ۲.۲.۲ 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 انعطاف پذیر برای مقابله با این مشکلات پیشنهاد شده‌اند.



شکل ۲.۱: رویکرد مبتنی بر تطبیق کلیشه [؟].

## ۳.۲.۲ رویکردهای مبتنی بر ویژگی

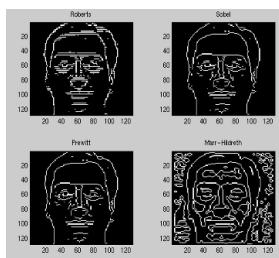
این رویکردها با استخراج ویژگی‌های ساختاری چهره، چهره‌ها را پیدا می‌نمایند. ابتدا به عنوان یک طبقه‌بند، آموزش دیده و سپس برای تمایز میان نواحی شامل چهره و بدون چهره در تصویر استفاده می‌شوند. ایده این رویکرد، غلبه بر محدودیت دانش ما از چهره‌ها می‌باشد. در [۲] تعدادی از این رویکردها مورد بررسی قرار گرفته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

## ۴.۲.۲ 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 مواجه با اجسام متحرک دیگری مانند اتومبیل چار اشتباه می‌شود.

## ۵.۲.۲ 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



شکل ۲-۲: استفاده از لبه یاب‌های معروف برای استخراج ویژگی‌های مبتنی بر لبه از چهره [؟].

## ۶.۲.۲ رویکرد مبتنی بر اطلاعات سطح خاکستری

سطوح خاکستری تصویر چهره شامل اطلاعات مفیدی می‌باشد. برای مثال ابروها، مردمک چشم و لب‌ها معمولاً تاریک‌تر از سایر نواحی صورت هستند. این ویژگی‌ها می‌توانند به یافتن یک چهره در تصویر کمک نماید. در این رویکرد ابتدا بر روی تصویر ورودی، عملیات بسط تباین 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 ۳.۲: اطلاعات سطح خاکستری حتی در وضوح پایین نیز قابل دسترسی می باشند [؟].

#### ۷.۲.۲ 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. گسترده ترین فضای رنگ مورد استفاده ۰۰۰ می باشد.

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

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

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

در روابط بالا  $r$  میزان سطح رنگ قرمز،  $g$  میزان سطح رنگ سبز و  $b$  میزان سطح رنگ آبی در هر پیکسل از تصویر می باشد.

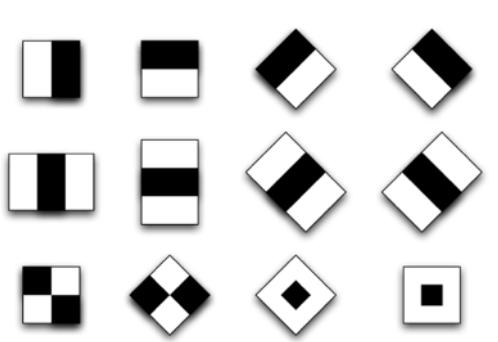
$r$  و  $g$  به ترتیب مقدار هنجار سازی شده برای رنگ قرمز، سبز و آبی می باشد. واضح است که

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

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،  $\text{YUV}$ ،  $\text{L*a*b*}$ ،  $\text{HSL}$ . 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.

## ۸.۲.۲ رویکرد مبتنی بر عامل‌های آماری

رویکردهای بخش‌های قبیل 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 یک 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 پیشرفته تر و دقیق‌تر نماید، اما انقلاب الگوریتم‌های یافتن چهره در سال ۲۰۰۱ بود. زمانی که  $\text{CNN}$  و  $\text{RBM}$  در [۳] یک الگوریتم چهره یاب بی‌درنگ معرفی کردند که قادر به یافتن چهره با دقت بالا بود. در ادامه به شرح این الگوریتم می‌پردازیم. پیش پردازش: ابتدا تصویر از فضای  $\mathbb{R}^3$  به تصویر سطح خاکستری تبدیل می‌شود. زیرا تشخیص چهره‌ها در تصویر خاکستری برای سامانه آسان است. سپس در صورت نیاز، پیش پردازش‌هایی مانند تغییر اندازه، برش، تار شدن و تیزکردن لبه‌های تصویر انجام می‌شود. ویژگی‌های  $\text{CNN}$  تمام چهره‌های انسانی ویژگی‌های مشترکی دارند. برای مثال ناحیه چشم تاریک‌تر از پیکسل‌های همسایه آن است و ناحیه بینی از چشم روشن‌تر است. ویژگی‌های  $\text{RBM}$  مستطیل‌هایی هستند که نشان دهنده بخش‌های مختلف صورت می‌باشند. شکل ۴-۲ نمونه‌هایی از مستطیل‌های ویژگی‌های  $\text{CNN}$  را نشان می‌دهد.



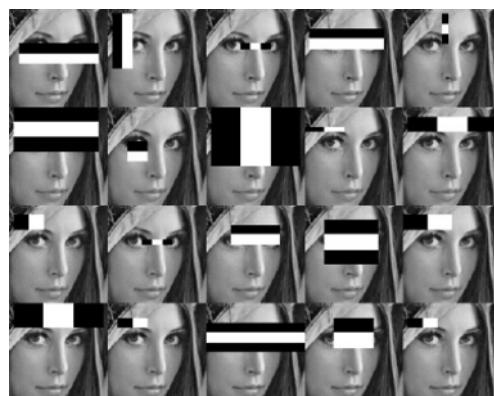
شکل ۴.۲: نمونه‌هایی از مستطیل‌های ویژگی‌های  $\text{CNN}$ . [۹].

ویژگی‌های  $\text{CNN}$  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  $\text{CNN}$  متناسب با بخش‌های چهره می‌باشند که مثالی از آن در شکل ۵-۲ نشان داده شده است.



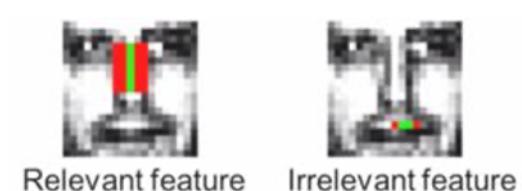
شکل ۵.۲: مستطیل های ویژگی های  $100 \times 100$  مناسب با بخش های چهره می باشند [؟].

همانطور که در شکل ۶-۲ مشاهده می شود، هر مستطیل در بخش های مختلف چهره در روندی تکراری با اندازه های مختلف قرار می گیرد و نتیجه نهایی از کم کردن مجموع سطح روشنایی پیکسل های زیر بخش های سفید به دست می آید که یک عدد می باشد. از یک پنجره با اندازه  $24 \times 24$  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 ۶.۲: هر مستطیل در اندازه های مختلف بر روی بخش های مختلف تصویر قرار می گیرد [؟].

همانطور که در شکل ۷-۲ مشاهده می شود، تمام ویژگی های  $100 \times 100$  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  $\text{ReLU}$  انجام می شود.



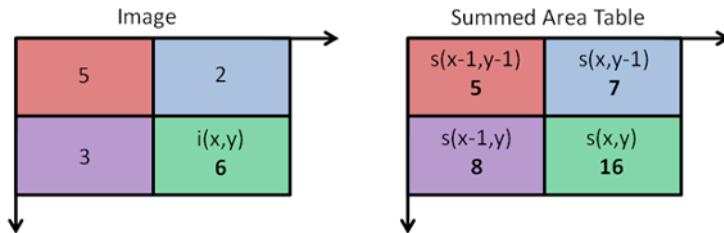
شکل ۷.۲: یک ویژگی مرتبط در مقابل یک ویژگی نامرتبط [؟].

یک الگوریتم مبتنی بر یادگیری ماشین می‌باشد که ویژگی‌های کاربردی را از میان تعداد زیادی ویژگی پیدا می‌کند. بعد از شناسایی ویژگی‌های مختلف، مشخص می‌گردد که هر یک از پنجره‌ها برای بخشی از چهره مناسب می‌باشد یا خیر. هر کدام از ضرایب انتخاب شده مثبت در نظر گرفته می‌شود، در صورتی که حداقل بتواند بیش از نیمی از موارد را تشخیص دهد. این ویژگی‌ها با عنوان طبقه‌بندی‌های ضعیف معرفی می‌شوند. در طبقه‌بند قوی، تعداد زیادی طبقه‌بند ضعیف را با هم ترکیب می‌کند. رابطه کلی آن به صورت زیر می‌باشد.

$$F(x) = {}_1f_1(x) + {}_2f_2(x) + \dots \quad (\text{d.2})$$

که در آن طبقه‌بند قوی می‌باشد که از تعدادی  $\Theta(n)$  که طبقه‌بند ضعیف می‌باشد، تشکیل شده است. هر کدام از طبقه‌بندهای ضعیف، یک خروجی صفر یا یک تولید می‌کنند.  $\Theta(n)$  وزن مربوط به طبقه‌بند می‌باشد. با استفاده از این الگوریتم و وزن دادن به ویژگی‌ها، بیش از  $160000$  ویژگی قبلی به کمتر از  $2500$  ویژگی کاهش پیدا می‌کند.

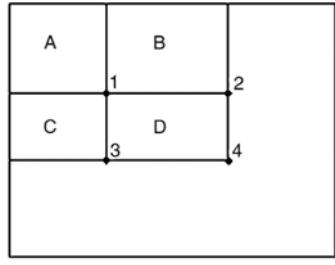
تصویر یکپارچه: تصویر یکپارچه، یا جدول محدوده مجتمع، به منظور ارزیابی سریع‌تر ویژگی‌هایی که در بخش اول معرفی شد، استفاده می‌شود. با توجه به شکل ۲-۸ در یک تصویر یکپارچه مقدار پیکسل در مکان  $\square$  و  $\square$  برابر با جمع مقادیر پیکسل‌های بالا و چپ پیکسل  $\square$  و  $\square$  می‌باشد.



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

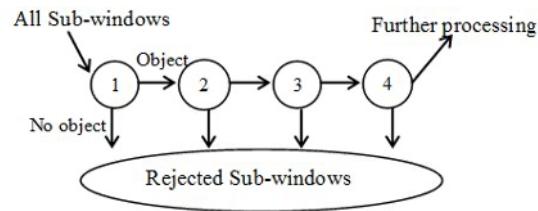
به عنوان مثال در شکل ۹-۲ مقدار پیکسل‌ها به صورت زیر محاسبه می‌شود: مقدار پیکسل ۱ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل ۱. مقدار پیکسل ۲ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل ۱ و ۲. مقدار پیکسل ۳ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل ۱ و ۳. مقدار پیکسل ۴ در تصویر یکپارچه برابر است با مجموع پیکسل‌ها در مستطیل ۱ و ۲ و ۳. مجموع پیکسل‌ها در مستطیل ۱ می‌تواند به صورت  $(3+2) - (1+4)$  محاسبه شود.

**مراحل آبشاری:** اگر تصویر به مربع‌های  $24 \times 24$  پیکسلی تقسیم شده و با پردازش هر بخش که ۲۵۰۰ ویژگی دارد، تشخیص



شکل ۹.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [۴].

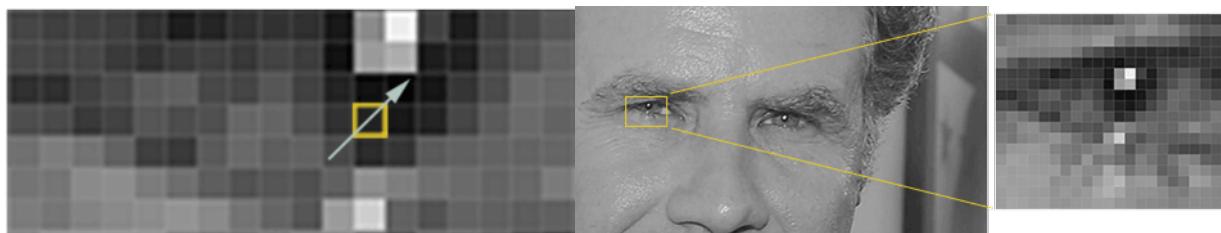
می دهد. ۲۵۰۰ ویژگی هر مربع  $24 \times 24$  به دسته بندی های مختلف تقسیم می شود. برای مثال ۱۰ ویژگی در دسته اول، ۲۰ ویژگی در دسته دوم، ۱۰۰ ویژگی در دسته سوم و ... می توان بعد از پردازش هر دسته، در ارتباط با وجود یا عدم وجود چهره در آن دسته تصمیم گرفت تا بخش هایی که چهره در آن وجود ندارد زودتر حذف شوند. شکل ۱۰-۲ یک نمای کلی از روند تشخیص آیشواری را نشان می دهد. مجموعه ای از طبقه بندها به هر زیر پنجره اعمال می شود. طبقه بند اولیه تعداد زیادی از نمونه های منفی را حذف می کند و پردازش کمی دارد. لایه های بعد، منفی های اضافی را حذف می کنند که نیاز به محاسبات بیشتری دارند.



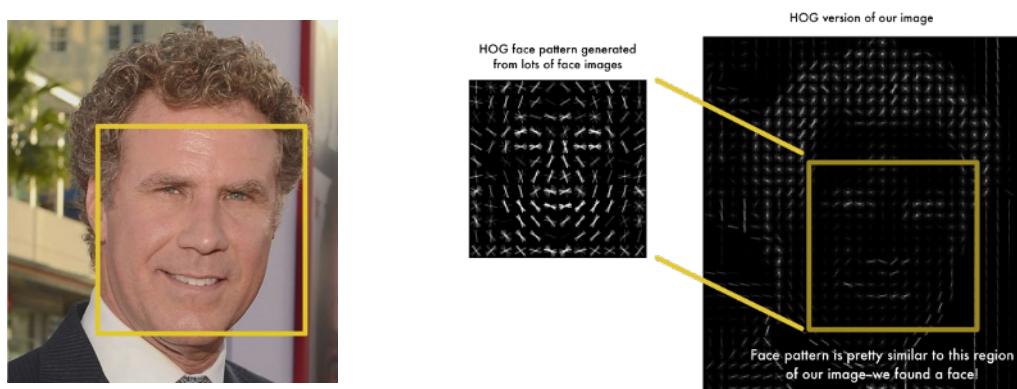
شکل ۱۰.۲: بخشی از یک تصویر که تصویر یکپارچه آن محاسبه می شود [۴].

این الگوریتم عملکرد بسیار خوبی در برنامه های کاربردی بی درنگ و در حضور پس زمینه های شلوغ نشان داده است. اما هنوز در برای چهره هایی که رو به روی دوربین نیستند، تغییرات شدید نور، انسداد و ... دارای چالش می باشد. در سال ۲۰۰۵ میلادی و همکاران در [۴] روشی به نام بافت نگار شبیه های جهت دار ارائه کردند که به اختصار CNN نامیده می شود. در این روش ابتدا تصویر خاکستری می شود، زیرا نیازی به رنگ نیست. پیرامون هر پیکسل بررسی می شود تا مشخص شود نسبت به پیکسل های پیرامونش چقدر تاریک می باشد. مطابق شکل ۱۱-۲ جهتی انتخاب می شود که به سمت پیکسل های تاریکتر باشد. این روند برای همه پیکسل های تصویر انجام می شود و به ازای هر پیکسل یک جهت ذخیره خواهد شد که روندی از روشنایی به تاریکی را در تصویر نمایش می دهدند.

دلیل استفاده از جهت ها این است که اگر پیکسل ها به طور مستقیم استفاده شوند، تصویر تاریک و تصویر روشن از یک چهره مشخص، دارای سطح روشنایی متفاوتی خواهند بود. اما با در نظر گرفتن جهتی تغییر روشنایی، هم تصویر تیره و هم تصویر روشن،



شکل ۱۱.۲: سطح روشنایی پیکسل های اطراف هر پیکسل بررسی شده و راستای روشن به سمت تاریک برگزیده می شود [۹].



شکل ۱۲.۲: برای یافتن چهره ها، بخش هایی از تصویر که به الگوی ۱۶x۱۶ شبیه تر است را پیدا می کنیم [۹].

نمایش یکسانی خواهند داشت که حل مسئله را آسان تر می کند. ذخیره جهت ها برای تمام پیکسل ها باعث افزایش جزئیات می شود.

لذا روند اصلی روشنایی و تاریکی در سطحی بالاتر در نظر گرفته می شود، به طوری که بتوان الگوی اصلی تصویر را دید. تصویر به

بخش های  $16 \times 16$  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 شده از

ساخтар چهره است که در شکل ۱۲-۲ مشاهده می شود. برای یافتن چهره ها در این الگوریتم، بخش هایی از تصویر که به الگوی ۱۶x۱۶

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 [www.cs.toronto.edu/~guerquin/paper/HOGFace.pdf](http://www.cs.toronto.edu/~guerquin/paper/HOGFace.pdf) و همکاران در [۵] یک روش دقیق و سریع برای یافتن چهره ارائه دادند که در آن از

اختلاف پیکسل هنجارسازی شده یا HOG برای یافتن چهره استفاده می شود. ارزیابی ویژگی HOG 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. HOG بین دو پیکسل در یک تصویر به صورت زیر تعریف

شده است:

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

که در آن  $m$  و  $n$  بزرگتر از صفر هستند و  $(m, n) = 1$  باشد، برابر صفر است. این عمل بر روی هر جفت پیکسل از تصویر اجرا می‌شود. اگر تصویر ورودی مربعی با ابعاد  $m \times m$  باشد و  $m \times n$  تعداد پیکسل‌ها باشد، آنگاه تعداد ویژگی‌های استخراج شده برابر  $\frac{m(m-1)}{2}$  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 تیره‌تر 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  $mn^2$  در شکل ۱۳-۲ آمده است.



شکل ۱۳.۲: نتیجه اجرای روش مبتنی بر  $mn^2$  [۹].

تمام رویکردهای ارائه شده که یافتن چهره را با مدل سازی صریح از ویژگی‌های صورت انجام می‌دهند، در برابر تغییرات غیرقابل 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 خصم‌مانه تر مانند تشخیص چهره‌های متعدد در زمینه‌های شلوغ به خوبی عمل کنند.

## ۹.۲.۲ رویکردهای مبتنی بر تصویر

رویکردهای مبتنی بر تصویر نیاز به مجموعه‌ای از تصاویر آموزشی برای پیدا کردن مدل‌های چهره دارند و بر اساس استخراج ویژگی و یادگیری ماشین عمل می‌نمایند. مجموعه‌ای از تصویرهای مختلف چهره طبقه‌بندی می‌شوند و برای تشخیص چهره جدید از این طبقه‌بندی استفاده می‌شود. نمونه‌هایی از غیر چهره به طبقه‌بند داده می‌شود تا از روی این تصویرها عمل یادگیری انجام شود. به طور تجربی دقت نتایج رویکردهای مبتنی بر تصویر بهتر از سایر رویکردها می‌باشد. این رویکردها به چند

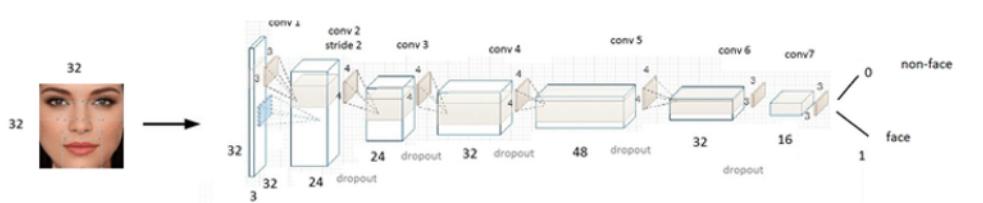
دسته تقسیم می‌شوند که در ادامه شرح داده شده است.

## ۱۰.۲.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

ماشین بردار پشتیبان طبقه‌بندی خطی می‌باشد که حاشیه بین ابرصفحه تصمیم و داده‌های آموزش را به حداقل می‌رساند. برای اولین بار در سال ۱۹۹۷  $\text{OSS}$  و همکاران در [۶] از این طبقه‌بند برای یافتن چهره استفاده کردند.

## ۱۱.۲.۲ رویکرد مبتنی بر شبکه عصبی

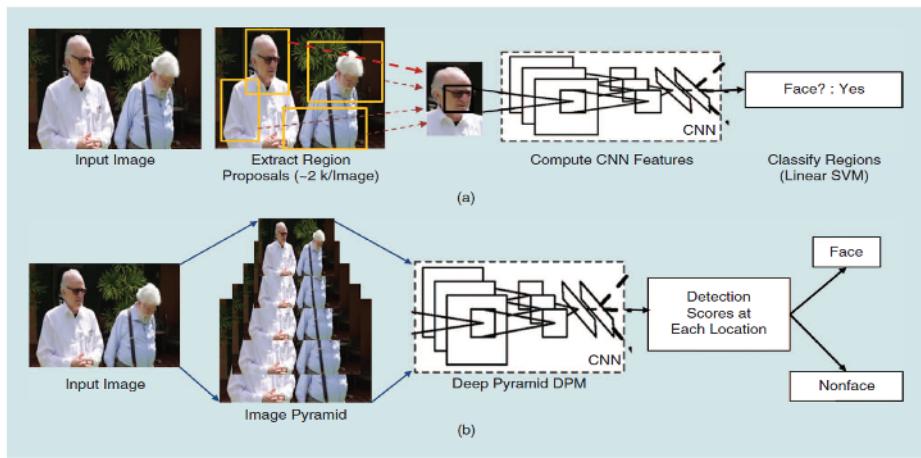
یک راه حل غیر خطی برای یافتن چهره، استفاده از شبکه‌های عصبی است. اولین رویکردهای عصبی در یافتن چهره بر اساس  $\text{OSS}$  بود که در مجموعه داده‌های ساده، امیدوار کننده بود. شبکه‌های عصبی با معماری پیمانه ای امروزی بسیار پیچیده‌تر از  $\text{OSS}$  ساده هستند. نوع خاصی از شبکه‌های عصبی عمیق برای پردازش 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 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



شکل ۱۴.۲: نمای کلی یک شبکه عصبی پیچشی برای یافتن چهره [۹].

با توجه به آنچه در [۷] آمده است، چالش اصلی در یافتن چهره این است که ویژگی‌هایی مانند  $\text{OSS}$  و  $\text{OSS}$  اطلاعات برجسته چهره را در شرایط مختلف نما، نورپردازی، رنگ پوست، انسداد، استفاده از لوازم آرایشی و... استخراج نمی‌کنند. این محدودیت بیشتر به دلیل ویژگی‌های استفاده شده در طبقه‌بندها است. با پیشرفت‌های اخیر در رویکردهای یادگیری عمیق و در دسترس بودن پردازنده‌های گرافیکی، استفاده از شبکه‌های عصبی پیچشی عمیق برای استخراج ویژگی امکان پذیر شده است. ویژگی‌های عمیق به دست آمده به طور گستردگایی برای یافتن چهره استفاده می‌شود. با توجه به شکل ۱۵-۲ روش‌های مبتنی بر

شبکه عصبی پیچشی عمیق برای یافتن چهره به دو زیر شاخه تقسیم می‌شود: رویکرد مبتنی بر ناحیه و رویکرد پنجره لغزان.



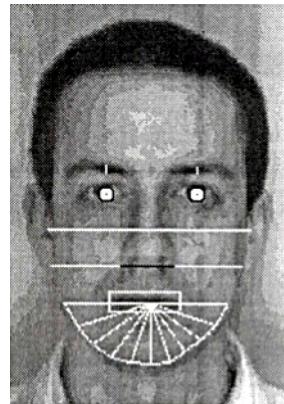
شکل ۱۵.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].

## ۳.۲ شناسایی چهره

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

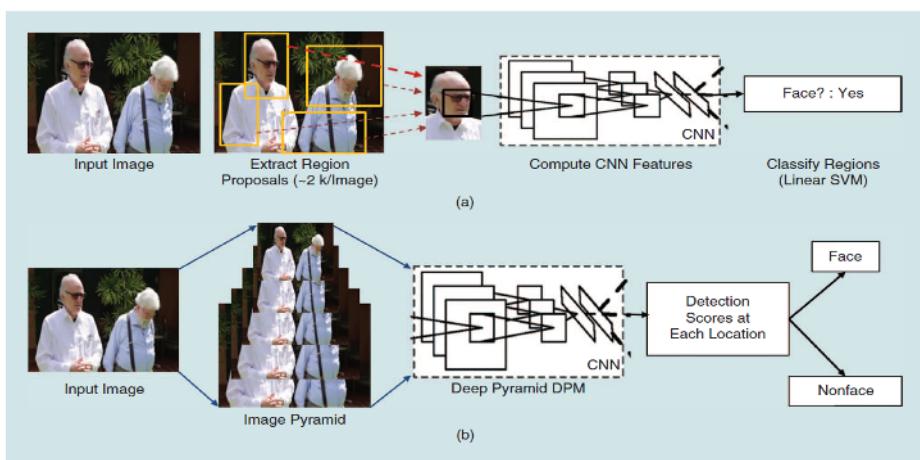
## ۴.۲ رویکردهای سنتی

این رویکردها ویژگی‌های چهره را با علامت‌ها و اندازه‌ها از تصویر استخراج می‌کنند. برای مثال موقعیت نسبی، اندازه و یا شکل چشم‌ها، بینی، گونه‌ها و فک را محاسبه کرده و تجزیه و تحلیل می‌کنند. سپس از این ویژگی‌ها برای جستجوی تصاویر دیگر در پایگاه داده استفاده می‌کنند. در سال ۱۹۹۳ **Levoy et al.** و همکاران در [۸] یکی از اولین الگوریتم‌ها در این زمینه را پایه داده استفاده می‌کنند. این رویکرد در شرایط کنترل شده به دقت ۹۰ رویکرد مبتنی بر فیلتر گابور همانطور که در [۹] 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 آمده شده، با یک آستانه گذاری مطابقت



شکل ۱۶.۲: ویژگی های هندسی (رنگ سفید) مورد استفاده در آزمایش های تشخیص چهره [؟].

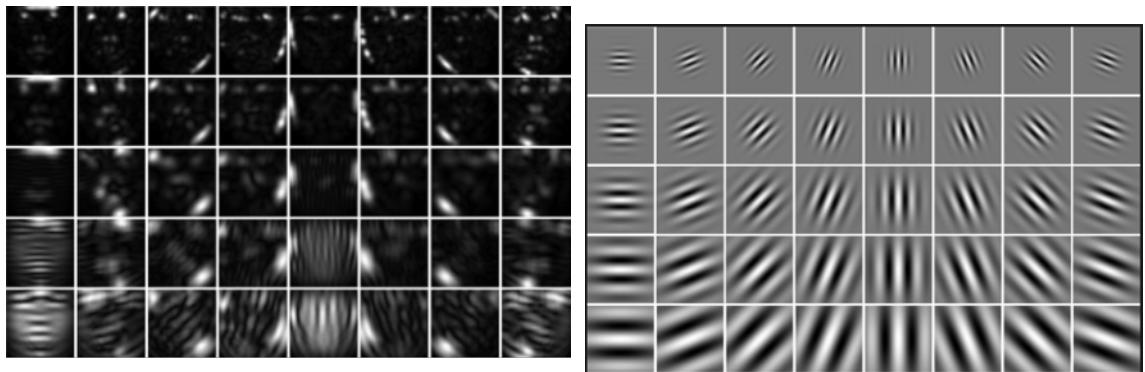
داده می شود. شکل ۱۷-۲ فیلترهای چندگانه گابور و تاثیر این فیلترها بر روی تصویر چهره انسان را نشان می دهد. دلیل استفاده از فیلتر گابور این است که عملکرد این فیلتر به سامانه بصری انسان بسیار شباهت دارد.



شکل ۱۷.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (a) رویکرد مبتنی بر ناحیه و (b) رویکرد پنجره لغزان [؟].

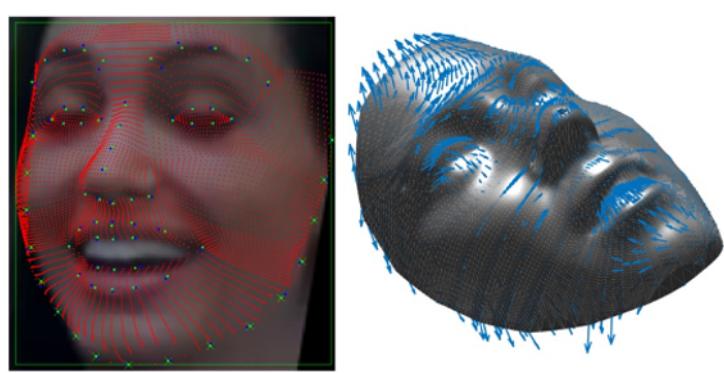
#### ۱.۴.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

رویکردهای سه بعدی همانطور که در [۱۰] آمده است، داده های سه بعدی دقیق تشخیص چهره را به شدت بهبود می بخشد، اختلاف داده های ورودی با داده های ذخیره شده زیادتر است و سامانه با دقیق بیشتری عمل می کند. روش تشخیص سه بعدی چهره از یک منتشر کننده نور فرو سرخ و یک حسگر به عنوان دریافت کننده استفاده می کند. شبکه های از نورهای فرو سرخ که برای انسان قابل رویت نیست، روی چهره تابانده می شود. سپس یک حسگر ویژه، پرتو های بازتاب را دریافت کرده و اطلاعات عمق تصویر پردازش می شود. این دسته از الگوریتم ها برای شناسایی دقیق اشخاص، برای هر نفر برداری های سه بعدی می سازند. عیب رویکردهای سه



شکل ۱۸.۲: (الف) فیلترهای چندگانه گابور (ب) تاثیر این فیلترها بر روی تصویر چهره [۹].

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.



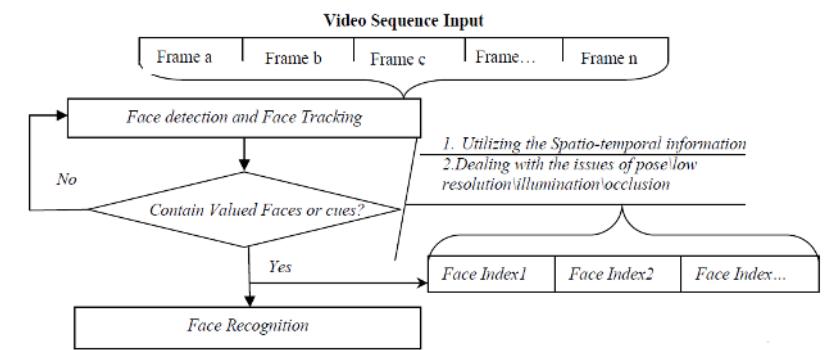
شکل ۱۹.۲: مدل سازی سه بعدی چهره با اشعه فروسرخ [۹].

#### ۲۰.۴.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

رویکردهای تجزیه و تحلیل بافت پوست یکی دیگر از رویکردهای در حال ظهور، استفاده از بافت پوست برای شناسایی چهره می‌باشد که خطوط، الگوها و لکه‌های پوست را به یک فضای ریاضی تبدیل می‌کند. تجزیه و تحلیل بافت بسیار شبیه روش شناسایی چهره است. تصویری از پوست گرفته می‌شود و به بخش‌های کوچکتر تقسیم می‌شود. سپس هر بخش به یک فضای ریاضی قابل اندازه گیری تبدیل می‌شود و خطوط، منافذ و بافت پوست تشخیص داده می‌شود. این رویکرد می‌تواند تفاوت بین دوقلوهای یکسان را شناسایی کند که با استفاده از تشخیص چهره به تنها یک امکان پذیر نیست. آزمایش‌ها نشان دادند که با افزودن تحلیل بافت پوست، عملکرد سامانه تشخیص چهره می‌تواند ۲۵ تا ۲۰۱۷  $\text{m} \times \text{m}$  و همکاران در [۱۱] یک

روش سه بعدی برای توصیف ویژگی‌های چهره ارائه کردند که در آن از مدل سازی سه بعدی چهره به همراه تجزیه و تحلیل بافت پوست استفاده شده است. این سامانه شایستگی استفاده در کاربردهای مختلف امنیتی و نظامی با شناسایی خودکار سریع و بدون دخالت شخص را دارد و سرعت پردازش را بالا و خطأ را کاهش داده است. برتری روش سه بعدی در عدم وابستگی به حرکت و جا به جایی صورت است. انتقال و نصب سامانه تصویر برداری بسیار ساده است. زاویه دید حسگر چندان مهم نیست. همچنین نورپردازی نامناسب تاثیری در این شیوه ندارد و عملیات آن ساده است. بر خلاف روش تشخیص دو بعدی، روش سه بعدی و تجزیه و تحلیل بافت پوست به تجهیزات بسیار پیچیده تری نیاز دارد، و با توجه به آنکه تمرکز ما بر روی تشخیص چهره به صورت بی‌درنگ در شرایط کنترل نشده مانند معابر پیاده و خیابان می‌باشد، به توضیح مختصر رویکردهای سه بعدی و تجزیه و تحلیل بافت پوست بسنده می‌کنیم. 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. بر خلاف 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 [۱۲] از شناسایی چهره مبتنی بر دوربین حرارتی در کاربردهای واقعی بهره برداری کردند و همکاران در [۱۳] از شناسایی چهره ایجاد کردند. آن‌ها از حسگرهای الکتریکی فروسرخ با حساسیت کم و با توانایی جذب حرارت طولانی مدت یا [۰۰۰] استفاده کردند. نتایج نشان می‌دهد که تلفیق [۰۰۰] و دوربین‌های معمولی، نتایج بهتری در شرایط کنترل نشده دارد. در این مطالعه ۲۴۰ چهره مجزا در طی ۱۰ هفته برای ایجاد پایگاه داده جدید استفاده شده است. داده‌ها در روزهای آفتابی، بارانی و ابری جمع آوری شد. در شرایط کنترل شده دوربین معمولی دقیق ۵.۹۷٪ تشخیص چهره مبتنی بر ویدیو در سال ۲۰۰۹ [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: امکان استفاده از افزونگی موجود در توالی ویدیو برای بهبود عملکرد تشخیص نسبت به تصاویر ثابت وجود دارد. تشخیص چهره و پیگیری آن در طول زمان، موجب انتخاب فریم‌های خوب می‌شود که حاوی چهره‌های رو به رو یا نشانه‌های ارزشمند است که شرایط نور، انسداد، حالت چهره و... در آن رضایت بخش می‌باشد. مطالعات روان پزشکی نشان داده است که اطلاعات پویا در فرایند تشخیص فرد بسیار حائز اهمیت است. نمایش‌های موثرتر مانند مدل چهره سه بعدی یا تصاویر [۰۰۰] می‌توانند از اطلاعات فریم‌های ویدیو گرفته شده برای بهبود شناخت استفاده کنند. یادگیری و به روز رسانی مدل در طول زمان امکان پذیر می‌باشد. برای تشخیص

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



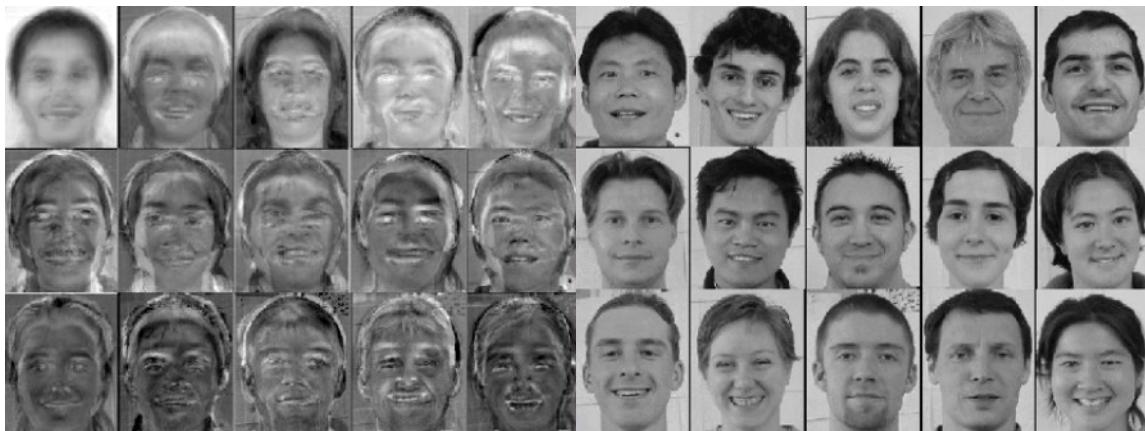
شکل ۲۰.۲: نمای کلی یک سامانه تشخیص چهره مبتنی بر ویدیو [؟].

### ۳.۴.۲ رویکرد مبتنی بر ماشین پشتیبان

روپکرد مبتنی بر چهره ویژه با افزایش حجم داده‌های موجود، نیاز به کاهش ابعاد داده‌ها می‌باشد. تحلیل مؤلفه‌های اساسی یا  $\text{PCA}$  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 سریع‌تر داده‌ها کمک می‌کند و از رخداد مشکل بیش‌برازاندن جلوگیری می‌نماید. اگر یک پایگاه داده عظیم از تصاویر چهره اشخاص باوضوح بالا موجود باشد که هر کدام دارای تعدادی زیاد ویژگی هستند، و بخواهیم یک تصویر آزمایش را با این پایگاه داده مقایسه کرده و شخص شیوه به آن را پیدا کنیم، مقایسه تصاویر بسیار زمان بر و در مواردی غیر ممکن خواهد بود.

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 ۲-۲۰ تعدادی چهره و چهره‌های ویژه متناظر با آن‌ها را نشان می‌دهد. در سال ۲۰۱۴  $\text{PCA}$  و همکاران در [۱۴] یک روش تشخیص چهره مبتنی بر  $\text{PCA}$  ارائه دادند که تا حدی در برابر تغییرات نورپردازی و انسداد مقاوم می‌باشد.  $\text{PCA}$  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 پراکنده‌گی را دارد برای ما مهم است، راه حل مناسبی خواهد بود.

در سال ۲۰۱۶ م. م. همکاران در [۱۵] یک روش شناسایی چهره مبتنی بر چهره‌های ویژه ارائه دادند



شکل ۲۱.۲: (الف) تعدادی چهره و (ب) چهره های ویژه متناظر با آن ها [۴].

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 منهنن میتوان استفاده کرد. از مزایای این روش میتوان به سهولت پیادهسازی و استفاده، کاهش حجم دادهها و سرعت بالا اشاره کرد. در نظر نگرفتن پراکندگی درون کلاسی و بین کلاسی دادهها و عدم توجه به برچسب تصویر برای شناسایی و تمایز قابل نشدن بین تصاویر مختلف یک شخص در پایگاه و نیاز به بروز رسانی تمامی اطلاعات موجود با ورود یک تصویر جدید به پایگاه داده از معایب این روش است. محاسبات ریاضی و مراحل انجام آن‌ها: تبدیل ماتریس تصاویر به بردار و کنار هم قرار دادن آن‌ها برای تشکیل ماتریس دادهها محاسبه میانگین ماتریس بدست آمده و انتقال دادهها به مرکزیت صفر محاسبه ماتریس کوواریانس بردارها و مقادیر ویژه آن انتقال ماتریس دادهها به زیرفضای جدید با استفاده از ماتریس بردارهای ویژه بررسی شباهت بین بردار منتقل شده و بردارهای موجود و انتخاب شبیهترین بردار

#### ۴.۴.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

رویکرد مبتنی بر ماشین بردار پشتیبان همان طور که در بخش ۱-۴-۲-۲ گفته شد، ماشین‌های بردار پشتیبان، طبقه‌بندی‌های خطی هستند که حاشیه بین ابرصفحه تصمیم و نمونه‌های مجموعه آموزش را به حداقل می‌رسانند. در سال ۲۰۱۲ ۰.۰۰۰۰۰ و همکاران در [۱۶] این طبقه‌بند را برای تشخیص چهره مورد استفاده قرار دادند. یک نسخه مبتنی بر هسته برای ۰۰۰ معرفی شده که نام گذاری شده است.

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

که در آن بردار وزن‌ها، داده‌های ورودی و مقدار پیش قدر می‌باشد. سپس یک مدل بی پارامتر از به نام معرفی شده، سپس یک نسخه مبتنی بر هسته به نام معرفی شده است.

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

که در آن ۱۱ و ۱۲ تعداد نمونه‌های آموزشی در هر دسته می‌باشند. سپس از ترکیب دو رویکرد فوق، مدلی به نام + معرفی نموده است.

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

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

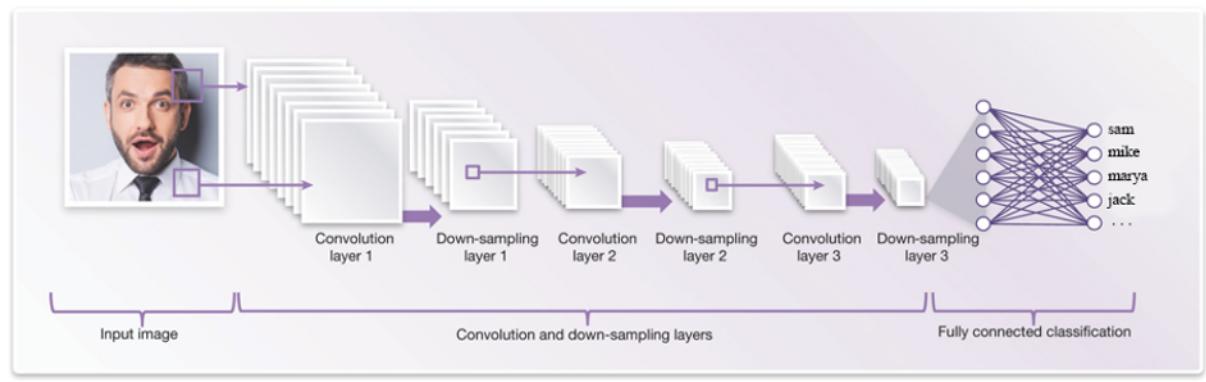
#### ۵.۴.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

رویکردهای مبتنی بر شبکه عصبی یک راه حل غیر خطی برای شناسایی چهره، استفاده از شبکه عصبی پیچشی است که به طور شگفت‌انگیزی در طبقه-بندی تصاویر چهره خوب کار می‌کند و ویژگی‌های ارزشمندی را از تصویر چهره استخراج می‌کند. بنابراین می‌توان از آن در حل مسئله شناسایی چهره و تأیید هویت استفاده کرد. شکل ۲۱-۲ ساختار کلی یک شبکه عصبی عمیق برای شناسایی چهره را نشان می‌دهد.

شکل ۲۱-۲ - شبکه عصبی عمیق برای شناسایی چهره معمولاً به عنوان تابع فعالیت از توابع غیرخطی مانند استفاده می‌شود. و عملیات بهینه‌سازی به روش پس انتشار خطا انجام می‌گردد. و در لایه خروجی از تابع برای طبقه‌بندی استفاده می‌شود که خروجی‌های لایه آخر را هنجار می‌کند. شبکه‌های عصبی پیچشی ویژگی‌های یک چهره را استخراج می‌کنند که

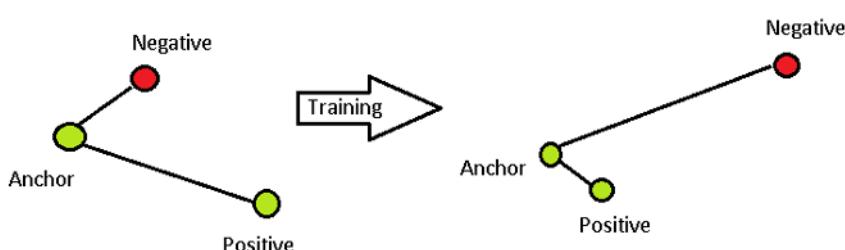
جدول ۱.۲: مقایسه دقت الگوریتم  $\text{NDA} + \text{SVM}$  با سایر رویکرد های مشابه

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>	<b>0.15</b>	0.22	0.86	0.86
Confidence (%)	-	99.99	100	99.87	99.96	99.95



شکل ۲۲.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۱].

می‌توان به عنوان یک شناسه برای یک فرد خاص در نظر گرفت. هنگامی که دو تصویر مختلف از چهره یک شخص به عنوان ورودی داده می‌شود، شبکه باید خروجی‌های مشابه (ویژگی‌های نزدیک تر) را برای هر دو تصویر تولید نماید، در حالی که برای چهره دو شخص مختلف، شبکه باید خروجی‌های بسیار متفاوت برای دو تصویر تولید نماید. شبکه عصبی نیاز به آموزش دارد تا به طور خودکار ویژگی‌های مختلف چهره‌ها را شناسایی کند و بر اساس آن محاسبات را انجام دهد. در ادامه چند شبکه عصبی پیچشی معروف مورد بررسی قرار گرفته است. شبکه  $\text{ResNet}$  در سال ۲۰۱۵ در  $\text{ImageNet}$  و همکاران در [۱۷] یک شبکه عصبی عمیق به نام  $\text{FaceNet}$  ارائه دادند.  $\text{FaceNet}$  یک مدل یکپارچه است که می‌آموزد چگونه تصاویر چهره را به یک فضای اقلیدسی فشرده نگاشت دهد تا فاصله تصاویر به طور مستقیم با میزان شباهت چهره‌ها مرتبط باشد. هنگامی که این فضا تولید شود، شناسایی چهره تایید هویت و خوشبندی می‌تواند به راحتی با استفاده از روش‌های استاندارد توسط  $\text{FaceNet}$  انجام شود. این شبکه برای آموزش از سه گانه تطبیق عدم تطبیق استفاده می‌نماید. با توجه به شکل ۲۲-۲، سه گانه تطبیق عدم تطبیق یک مجموعه از سه تصویر شامل یک تصویر مرجع، یک تصویر منطبق بر تصویر مرجع و یک تصویر غیر منطبق بر تصویر مرجع است که باید فاصله بین تصویر مرجع و تصویر منطبق را به حداقل برساند، زیرا هر دو دارای هویت مشابه هستند و فاصله بین تصویر مرجع و تصویر غیر منطبق را به حداکثر برساند، زیرا این تصاویر دارای هویت متفاوت می‌باشند.



شکل ۲۳.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۱].

شکل ۲۲-۲ - سه گانه تطبیق عدم تطبیق برای هر داده آموزشی (۱) مجموعه‌ای از داده‌های مشابه (۲) و مجموعه‌ای از داده‌های نامرتب (۳) در نظر گرفته می‌شود. سپس داده‌ها با تابع ضرر سه گانه طوری آموزش می‌بینند که رابطه زیر برای هر کدام از داده‌های آموزشی برقرار باشد.

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

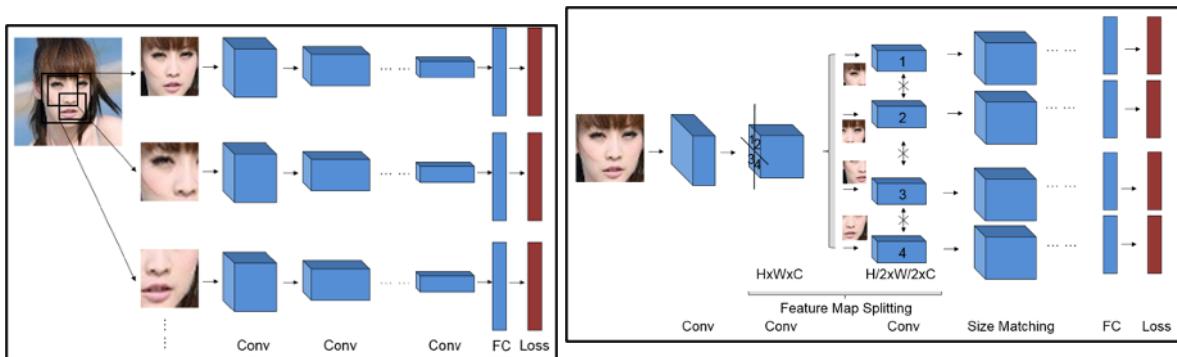
که در آن تابع  $\Box$  ویژگی‌های استخراج شده از تصاویر است. شبکه در مرحله آموزش با استفاده از داده‌های برچسب‌گذاری شده، می‌آموزد فاصله بین ویژگی‌های شبیه به هم، کمتر از فاصله بین ویژگی‌های دور باشد و به این ترتیب در مرحله آزمایش می‌تواند داده‌های مشابه و غیر مشابه را به راحتی تفکیک نماید. تابع هزینه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

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

که در آن  $\Box$  و  $\Box$  نمونه‌های مثبت و منفی برای نمونه آموزشی  $\Box$  می‌باشند و  $\Box$  حاشیه بین داده‌های مثبت و منفی را برای هر داده آموزشی مشخص می‌کند. این شبکه در مجموعه داده برچسب دار  $\Box$  به دقت جدید ۶۳٪.۹۹ رسیده است و در مجموعه داده  $\Box$   $\Box$  دقت آن به ۱۲٪.۹۵ رسیده است.

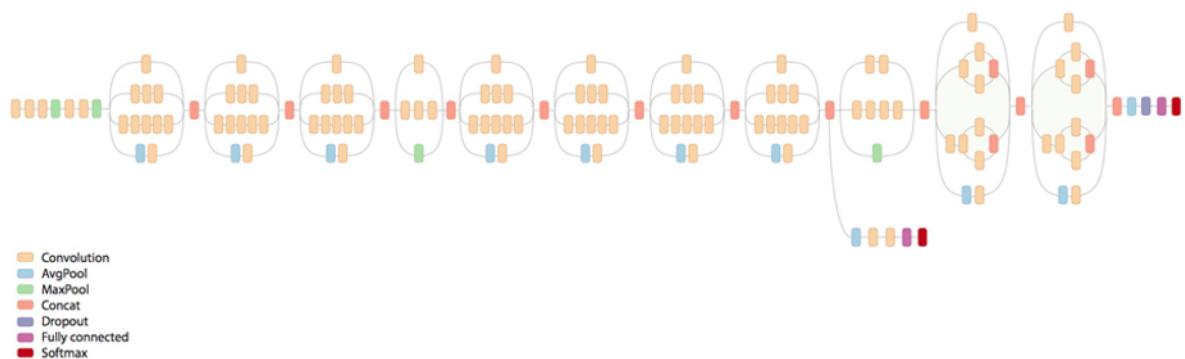
شبکه  $\Box$  در سال ۲۰۱۸  $\Box$  و همکاران در [۱۸] یک شبکه عمیق به نام  $\Box$  برای شناسایی چهره ارائه دادند. با توجه به ساختار معنایی چهره، یک بخش محلی از تصویر چهره همانند تصویر کلی چهره حاوی ویژگی‌ها و اطلاعات مفیدی برای یادگیری عمیق است. به منظور استفاده همزمان از اطلاعات سراسری و محلی، روش‌های یادگیری عمیق موجود برای شناسایی چهره، چندین شبکه  $\Box$  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  $\Box$ - $\Box$  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  $\Box$  را در مقابل شبکه عصبی پیچشی معمولی نشان می‌دهد. نتایج تجربی نشان می‌دهد که این رویکرد می‌تواند به طور موثر دقت تشخیص چهره را با محاسبات کمتر افزایش دهد.

شبکه  $\Box$  در سال ۲۰۱۵  $\Box$  و همکاران در [۱۹] یک شبکه عصبی عمیق به نام  $\Box$  ارائه دادند. همانطور که در شکل ۲۴-۲ مشاهده می‌شود،  $\Box$  یک شبکه عصبی پیچشی با ۲۲ لایه است



شكل ٢٤.٢: (الف) شبکه عصبی پیشی (ب) شبکه عصبی پیشی معمولی [؟].

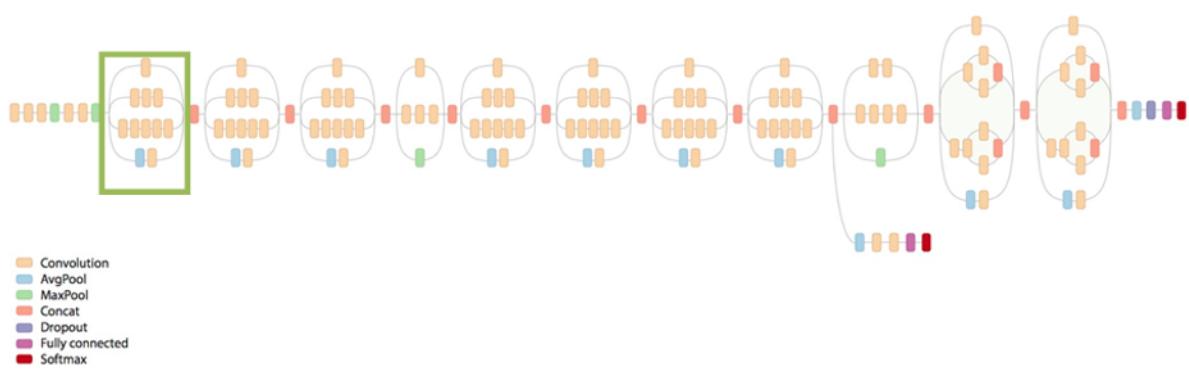
که یکی از اولین معماری های شبکه عصبی پیچشی بود که از رویکرد کلی قرار دادن تعداد زیادی از لایه های پیچشی و رای گیری در کنار هم در یک ساختار متوالی بدست آمد. نویسندها این مقاله همچنین تأکید کردند که این مدل جدید، توجه قابل ملاحظه ای به مصرف حافظه و مصرف انرژی دارد، زیرا کنار هم چیدن تعداد زیادی لایه و فیلتر دارای هزینه محاسباتی و حافظه است که احتمال پیش برازندن را افزایش می دهد.



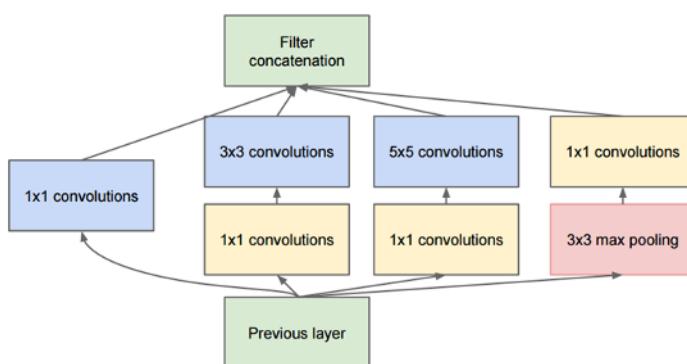
شکل ۲۵.۲: یافتن چهار مبتنی بر شبکه عصبی عمیق (□) رویکرد مبتنی بر ناحیه و (□) رویکرد پنجه لغزان [؟].

شکل ۲۴-۲ - معماری کلی شبکه در تمام محاسبات به طور متواالی اتفاق نمی‌افتد، بلکه هر بخش شبکه روندی موازی دارد. کادر سبز رنگ در شکل ۲۵-۲ بخش آغازگر نامیده می‌شود. در ادامه نگاهی دقیق‌تر به این بخش خواهیم داشت.

شکل ۲- کادر سبز رنگ یکی از بخش های موازی شبکه را نشان می دهد کادر سبز پایین در شکل ۲۶-۲ ورودی این بخش و کادر بالایی خروجی می باشد. در هر لایه شبکه های پیچشی معمولی، باید بین یک لایه پیچشی یا رای گیر، یکی را انتخاب نمود. در حالی که اینجا می توان تمام این عملیات را به صورت موازی انجام داد. این همان ایده ساده ای بود که نویسنده گان مقاله روی آن تمرکز کردند.



شکل ۲۶.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۹].



شکل ۲۷.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۹].

شکل ۲۶.۲ - بخش آغازگر شبکه  $\text{ResNet}_{152}$  شبکه  $\text{ImageNet}$  در سال ۲۰۰۹ م. ۱۰۰۰۰۰۰۰۰ و همکاران در

[۲۰] شبکه عمیق  $\text{ResNet}_{152}$  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

این شبکه به یک بردار تبدیل می‌نماید که هر مقدار در این بردار، نشان دهنده احتمال برای یک هویت فردی است.

شبکه  $\text{ResNet}_{152}$  از یک تابع ضرر سه گانه در آموزش برای بهبود عملکرد کلی استفاده می‌نماید.

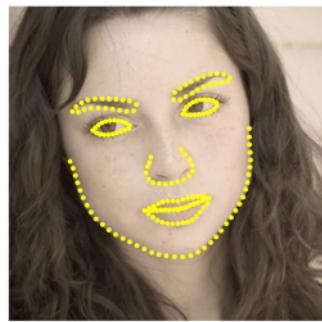


شکل ۲۸.۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۹].

شکل ۲۷.۲ - معماری شبکه  $\text{ImageNet}$

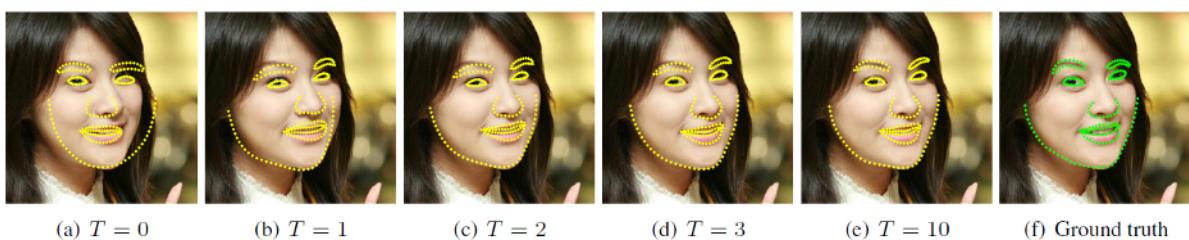
#### ۶.۴.۲ رویکرد مبتنی بر ماشین بردار پشتیبان

رویکردهای مبتنی بر نقاط راهنمایی چهره‌هایی که در جهت‌های مختلفی هستند، برای سامانه تشخیص چهره، متفاوت به نظر می‌رسند. برای غلبه بر این چالش در رویکردهای مبتنی بر نقاط راهنمایی می‌شود تصویر را چرخانده و جایه جا نمود، بطوریکه چشم‌ها و لب‌ها در یک موقعیت خاص در تصویر قرار بگیرند. بدین ترتیب مقایسه چهره‌ها در مرحله بعد بسیار ساده‌تر خواهد شد. در سال ۲۰۱۴ وحید کاظمی و جوزفین سالیوان در [۲۱] یک الگوریتم برای یافتن نقاط راهنمایی بر روی چهره ارائه دادند که از ۱۹۴ نقطه خاص که در هر چهره‌ای وجود دارد استفاده می‌نماید. شکل ۲-۲ مکان این نقاط را بر روی گونه، لبه‌های بیرونی چشم، کناره ابرو و... نشان می‌دهد. سپس این ۱۹۴ نقطه به سامانه آموزش داده می‌شود تا در هر چهره‌ای آن‌ها را تشخیص دهد.



شکل ۲-۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (□) رویکرد مبتنی بر ناحیه و (□) رویکرد پنجره لغزان [؟].

شکل ۲-۲ - نتیجه موقعیت ۱۹۴ شاخص روی چهره پس از این که دانستیم چشم‌ها، دهان و ... کجاست، به راحتی می‌توانیم تناسب تصویر را تغییر داده و آن را چرخانده یا برش بزنیم. به طوری که چشم‌ها و دهان در بهترین حالت ممکن در مرکز قرار گیرد. با استفاده از تغییرات اساسی و اصلی تصویر، مانند تغییر اندازه، چرخش، خطوط موازی را حفظ می‌کنیم که در ریاضی به آن تغییرات نسبت یا افاین می‌گویند. شکل ۲-۲ علامت گذاری تکراری خطوط راهنمایی بر روی چهره را نشان می‌دهد.



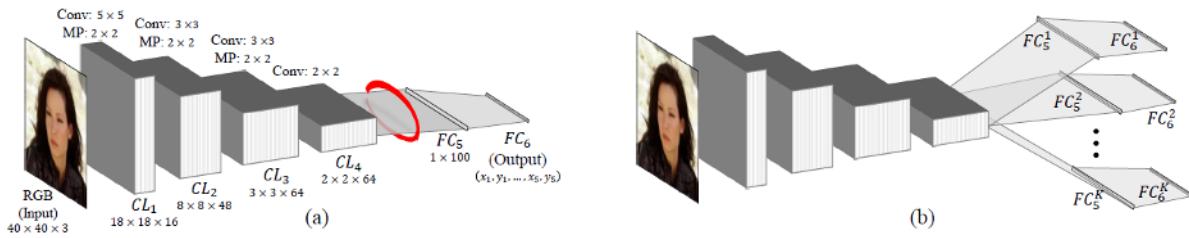
شکل ۲-۳: یافتن چهره مبتنی بر شبکه عصبی عمیق (□) رویکرد مبتنی بر ناحیه و (□) رویکرد پنجره لغزان [؟].

شکل ۲-۲ - علامت گذاری خطوط راهنمایی بر روی چهره که در هر تکرار با کاهش خطای همراه می‌باشد در سال ۲۰۱۸ [۲۲] و همکاران در [۲۲] 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 **FCN** پیشنهاد شده است که به اختصار **FCN** نامیده می‌شود. این شبکه عصبی عمیق از ۴ لایه پیچشی (۱، ۲، ۳، ۴) با لایه‌های رایگیری در میان آن‌ها تشکیل شده است و در انتهای یک لایه تمام متصل هست و پس از آن یک لایه خروجی با اندازه  $2 \times 2$  آمده است که مختصات نقطه ویژه را بر روی چهره مشخص می‌کند. در این مقاله  $\square$  برابر با ۵ در نظر گرفته شده است.تابع فعالیت برای هریک از لایه‌های پیچشی  $\square = \text{ReLU}(\cdot)$  و تابع فعالیت برای لایه تمام متصل  $\square = \text{Sigmoid}(\cdot)$  در نظر گرفته شده است. در نهایت تابع زیر به عنوان تابع ضرر معرفی شده است.

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

که در آن  $\square$  یک بردار  $2 \times 2$  برای مختصات پیش‌بینی شده تصویر  $\square$  و  $\square$  مختصات محل دقیق آن نقاط می‌باشد.  $\square$  و  $\square$  مختصات چشم‌ها در تصویر مرجع می‌باشند. در نهایت خروجی لایه تمام متصل توسط الگوریتم **FCN** به ۶۴ خوش تقسیم شده و هریک به صورت جداگانه بررسی شده است. معماری **FCN** در شکل ۳۰-۲ قسمت  $\square$  قابل مشاهده می‌باشد. این شبکه برای آموزش از مجموعه داده **FCN** استفاده کرده است.



شکل ۳۰-۲: یافتن چهره مبتنی بر شبکه عصبی عمیق (۱) رویکرد مبتنی بر ناحیه و (۲) رویکرد پنجره لغزان [۹].

شکل ۳۰-۲ - (۱) شبکه عصبی پیچشی معمولی و (۲) شبکه عصبی پیچشی با معماری **FCN**

$$M(x) = \begin{cases} 1, & I_{FLAIR}(x) > \theta_{FLAIR} \cap \text{Dilate}_R(WM(x)) > \theta_{WM} \\ 0, & \text{otherwise} \end{cases} \quad (13.2)$$

## ۵.۲ جمع‌بندی

در این فصل مفاهیم پایه در تقسیم‌بندی تصاویر MRI و انواع الگوریتم‌های تقسیم‌بندی و تشخیص ضایعه بیماری MS از روی تصاویر MRI بررسی شد. همانطور که در قبیل نیز بیان شد برای حل مساله تقسیم‌بندی ضایعات MS دو روش کلی، تقطیع و روش‌های مبتنی بر استخراج ویژگی وجود دارد. همچنین روش‌های تقطیع خود دارای رویکردهای مختلفی از جمله روش‌های مبتنی

بر خوشبندی، روش‌های مبتنی بر اطلس و روش‌های مبتنی بر اطلس می‌باشد که در جدول؟؟ مزایا و معایب هر کدام از آن‌ها مورد بحث و بررسی قرار گرفته شده است.

### فصل ۳

## مروری بر کارهای گذشته در شرایط کنترل نشده

در سال های اخیر روش های تشخیص چهره بسیار زیادی به منظور باقتن و شناسایی چهره افراد در تصویر پیشنهاد شده است که توانایی مقاومت در برابر مشکلات و چالش های رایج مانند تغییرات شدید روشنایی، تغییر حالت و زاویه چهره، انسداد، تاری خارج از تمرکز، سالخوردگی و... را ندارند و در کاربردهایی نظیر شرایط کنترل نشده قابل استفاده نیستند. در بخش مقدمه در مورد چالش های موجود در فرایند تشخیص چهره صحبت شد. برای رفع این چالش ها و بهبود طبقه بندی، راه حل هایی پیشنهاد شده است که در این بخش مورد بررسی قرار گرفته اند. جدول ۳- خلاصه ای از روش های مقابله با شرایط کنترل نشده

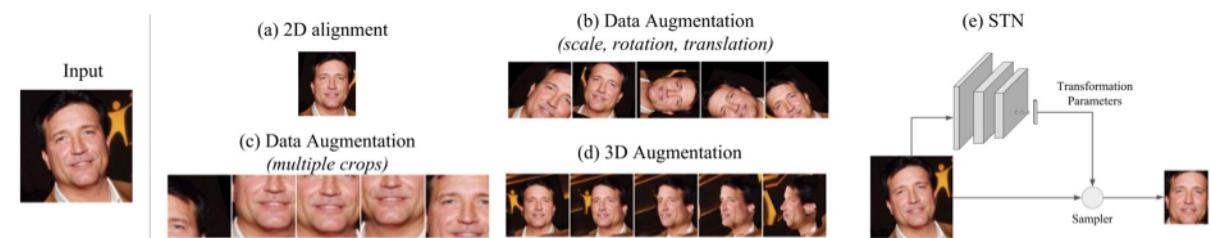
چالش مورد نظر	رویکرد	مزیت ها	مشکل ها
حالت چهره	تبديل دو بعدی	پیچیدگی محاسباتی قابل قبول	استخراج نقاط ویژه باید دقیق تر باشد
حالت چهره	استفاده از شبکه عصبی عمیق	پیچیدگی محاسباتی، وابستگی به داده های آموزش	دقت بالا در شرایط کنترل نشده
حالت چهره	تبديل مدل دو بعدی به سه بعدی	پیچیدگی محاسباتی	دقت بالا
حالت چهره	تبديل مدل سه بعدی به دو بعدی	پیچیدگی محاسباتی	دقت بالا
روشنایی	همسان سازی بافت نگار	پیچیدگی محاسباتی قابل قبول	قابل استفاده در تصاویر خاکستری
انسداد	استفاده از روش های شناسایی الگو و واپاش	دقت بالا در انسداد شدید	استخراج نقاط ویژه
محددیت داده	تهیه مجموعه داده با ردیابی چهره در ویدیو	تهیه مجموعه داده با دقیق تر	نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو
محددیت متابع	استفاده از رایانش ابری	سرعت بالا	تجهیزات پیشرفته و زمان تاخیر ناهمگن

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 چالش، شرایط کنترل نشده را دسته بندی می کنیم.

۱- رویکرد های دو بعدی با پیچیدگی محاسباتی قابل قبول (بیشتر ایده های مبتنی بر نشانه گذاری قدیمی) مانند [۲۶-۲۳]. مشکل: در محیط های بدون محددیت مانند آنچه که در این پروژه داریم، استخراج دقیق مکان نشانه های صورت از تصاویر دو بعدی نیاز به توجه بیشتری دارد. پیشرفته های اخیر مانند [۲۳] است. مزیت: این الگوریتم ها از نظر پیچیدگی محاسباتی قابل قبول هستند و کاملا برای شرایط این پروژه مناسب می باشند.

۲- رویکردهای مبتنی بر شبکه عصبی برای تخمین و اصلاح موقعیت چهره (آموزش و آزمایش با تصاویر دو بعدی) مانند [۲۲، ۲۷-۳۰]. مشکل: این الگوریتم ها به طور متوسط، کندتر از دسته پیشین می باشند. اما بستگی به این دارد که عمق شبکه عصبی چه مقدار باشد. وابستگی آن ها به

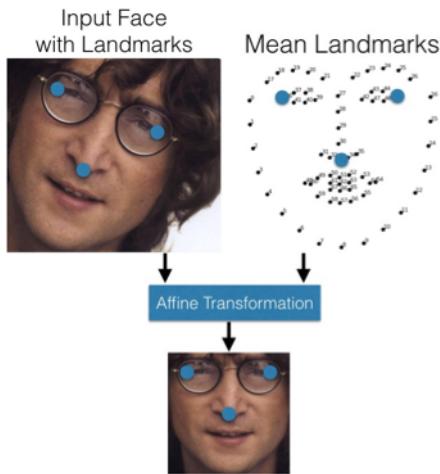
داده های آموزش می باشد و مراحل مجازی رو به رو سازی و هم ترازی چهره ندارند. مزیت: بدون نیاز به تصمیم گیری در مورد مجموعه بهینه ای از نشانه های چهره و دارای دقت بیشتر در شرایط کنترل نشده با انسداد و... . ایده هایی مانند [۴۲] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند. ۳- رویکردهای سبک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۳۱، ۳۴-۳۶]. مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده ترین این الگوریتم ها در مورد پیچیدگی محاسباتی، [۳۱] است که در شرایط بدون محدودیت آموزش دیده و آزمایش شده است. شامل مراحل مجازی رو به رو سازی و تراز بندی چهره می باشد، اما برای محدودیت های این پروژه قابل استفاده نمی باشد. مزیت: با استفاده از اطلاعات سه بعدی، این روش ها به بالاترین دقت تصمیم گیری در میان سه نفر رسید. ۴- رویکرد های تبدیل مدل سه بعدی چهره به مدل دو بعدی چهره (روش های مبتنی بر پنجره بر اساس چند نمایش دو بعدی مختلف از چهره) مانند [۳۵] مشکل: زمان محاسباتی (نه به اندازه الگوریتم های دسته سوم). مزیت: عملکرد بهتر در رو به رو سازی چهره در شرایط کنترل نشده نسبت به الگوریتم های دسته اول. ممکن است برای شرایط این پروژه متناسب باشد. مرجع [۴۳] در سال ۲۰۱۸ خلاصه ای از رویکردهای مختلف برای حل مسئله هم ترازی را در شکل ۱-۳ نشان داده است. تصویر سمت چپ، چهره ورودی می باشد. (۱) هم ترازی با استفاده از تبدیلات دو بعدی ساده می باشد. (۲) داده افزایی با تغییر مقیاس، تغییر زاویه و جا به جایی می باشد. (۳) برش های چندگانه می باشد. (۴) داده افزایی مبتنی بر روش های سه بعدی می باشد. (۵) از هیچ ابزاری برای هم ترازی مستقیم استفاده نمی نماید. اما یک شبکه را آموزش می دهد تا عامل های مورد نیاز برای تبدیل هم ترازی را بدست آورد.



شکل ۱.۳: رویکرد مبتنی بر تطبیق کلیشه [۴].

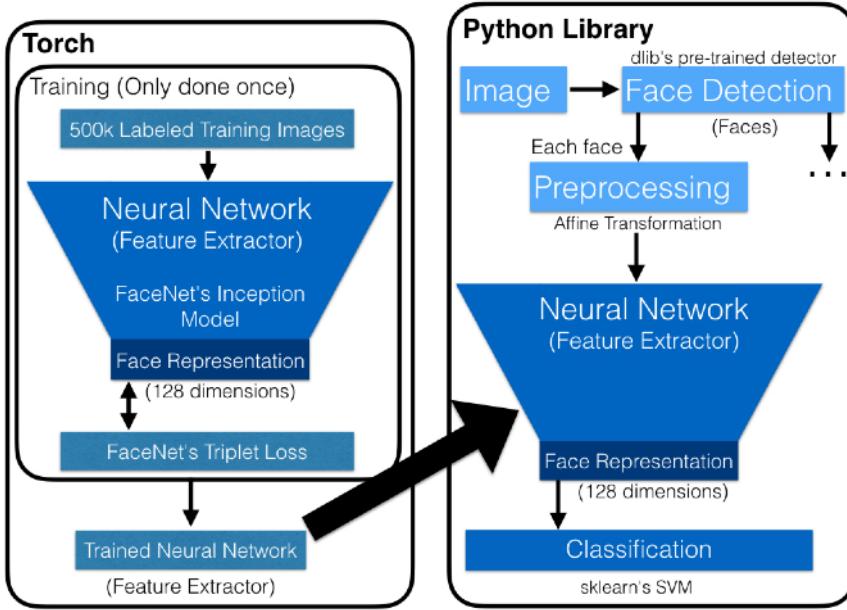
شکل ۳ - رویکردهای مختلف هم ترازی چهره در سال ۲۰۱۶، [www.cs.toronto.edu/~guerquin/](http://www.cs.toronto.edu/~guerquin/) و همکاران در [۲۵] یک روش شناسایی چهره به نام [www.cs.toronto.edu/~guerquin/](http://www.cs.toronto.edu/~guerquin/) ارائه دادند که ویژگی اصلی آن، آموزش شبکه عصبی عمیق در کمترین زمان و قابلیت اجرا بر روی دستگاه های قابل حمل مانند تلفن همراه با در نظر گرفتن منابع محدود می باشد. یک تصویر شامل تعدادی چهره به الگوریتم داده می شود. پس از یافتن چهره ها و مجزا کردن 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 دو بعدی ساده استفاده می گردد. ابتدا باید چهره توسط ۶۸ نقطه ویژه، نشانه گذاری شود. سپس نشانه های اطراف چشم ها و بینی (شکل ۳-۲) برای محاسبه عامل های تبدیل هم نسبی استفاده می شوند. پس از انجام تبدیل هم نسبی، تصاویر چهره برش زده شده و اندازه آن ها  $96 \times 96$  پیکسل می شود. شکل ۳ -

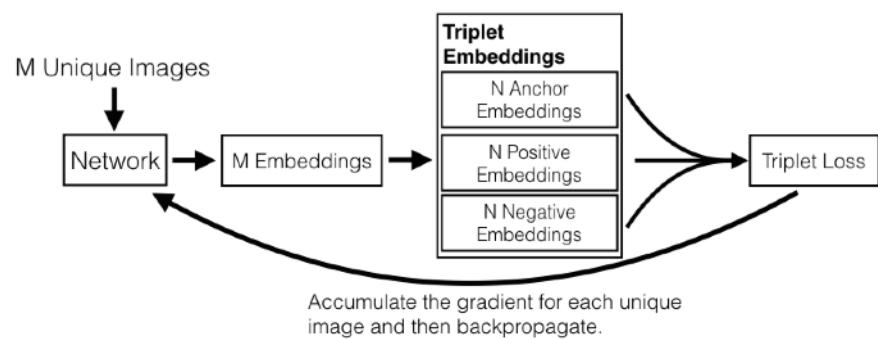


شکل ۳-۲: رویکرد مبتنی بر تطبیق کلیشه [؟].

تبدیل هم نسبی براساس نقاط ویژه آبی پس از پیش پردازش، تصاویر چهره ها به عنوان ورودی به یک شبکه عصبی پیچشی داده می شوند (شکل ۳-۳). این الگوریتم برای تعلیم شبکه از مجموعه داده کوچکی با  $500$  هزار تصویر چهره استفاده می کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری شده به نام  $\text{LFW}$ - $\text{DFW}$  و  $\text{ARFace}$  بدست آمده است. شبکه مورد استفاده در این الگوریتم یک نسخه اصلاح شده از شبکه  $\text{VGG16}$  الگوریتم می باشد. شبکه  $\text{VGG16}$  مبتنی بر معماری  $\text{ReLU}$  می باشد. برای تعیین میزان شباهت نتیجه، از فاصله اقلیدسی استفاده شده است. شکل ۳ - معماری  $\text{ReLU}$  هر تصویر از یک شبکه یکتا به یک سه گانه نگاشت داده می شود. گرادیان خطای سه گانه برای هر تصویر محاسبه شده و به عقب انتشار می یابد. در هر دسته کوچک،  $M$  تصویر برای هر نفر از  $N$  نفر، در مجموعه داده انتخاب می شود. سپس  $M \approx PQ$  تصویر به شبکه داده می شود تا عملیات  $\text{ReLU}$  انجام پذیرد. در این مقاله از  $N=15$  و  $P=20$  استفاده شده است. تمام جفت های  $\text{ReLU}$  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. شکل ۳ - جریان یادگیری در معماری  $\text{ReLU}$  مجموعه داده  $\text{LFW}$  یک معیار استاندارد برای سنجیدن میزان دقیق الگوریتم های تشخیص چهره می باشد که از ۱۳۲۳۳ تصویر چهره از ۵۷۵۰ شخص تشکیل شده است. الگوریتم  $\text{ReLU}$  بر روی این مجموعه داده مورد



شکل ۳.۳: رویکرد مبتنی بر تطبیق کلیشه [۴].



شکل ۴.۳: رویکرد مبتنی بر تطبیق کلیشه [۴].

سنجهش قرار گرفت که به دقت  $0.9292 \pm 0.0134\%$  رسید. در سال ۲۰۱۶ میلادی مذکور شده و همکاران در [۲۳]

یک روش برای هنجارسازی حالت چهره بر اساس تنظیم کردن مدل ظاهری فعال یا **active appearance model** یک مدل پارامتری است

که برای ارائه یک شکل مانند چهره انسان استفاده می‌شود. در این الگوریتم ابتدا یک **template** بر روی تصویر چهره قرار گرفته، با روندی

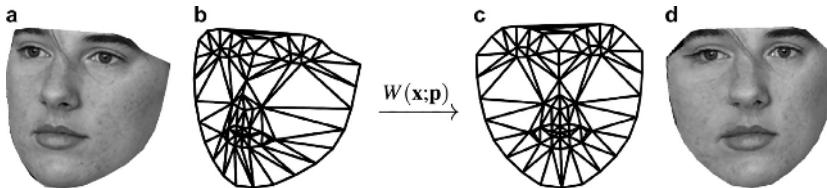
تکراری و به صورت بهینه شونده، بر روی چهره تنظیم می‌شود. سپس با استفاده از یک تبدیل هم نسبی، مرحله رو به رو سازی بر

روی چهره انجام می‌پذیرد. شکل ۵-۳ رویکرد کلی این الگوریتم را نشان می‌دهد. شکل ۳ - ۵ - رویکرد کلی الگوریتم مبتنی بر

برای رو به رو سازی چهره در این مدل یک تصویر چهره با مجموعه‌ای از نقاط ویژه هنجارسازی شده مدل می‌شود که به صورت

۵۰، ۵۰] تعریف می‌شود که در آن  $1 = 0.2$ . برای انجام این کار یک مرحله یادگیری نیاز است. سپس الگوریتم ۵۰۰ اعمال

می‌شود تا کاهش میزان وابستگی میان نقاط ویژه در هر مجموعه انجام می‌شود و نتیجه یک مدل خطی است که یک مدل شکل



شکل ۵.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

نمونه را به صورت زیر نمایش می دهد.

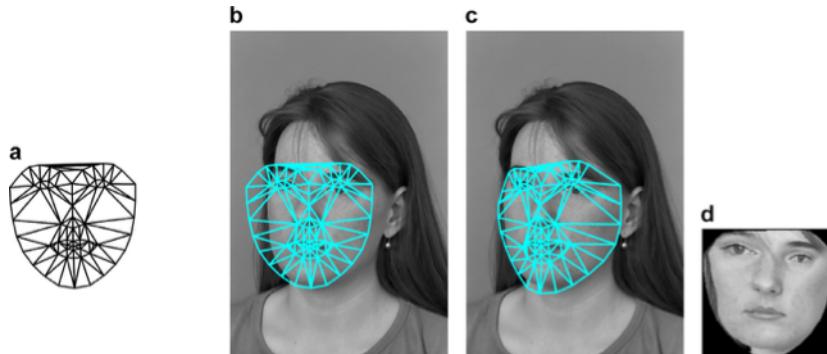
$$S = s_0 + \sum_{i=1}^n p_i s_i \quad (1.3)$$

که در آن  $\mathbf{x}$ . شکل پایه،  $\mathbf{p}$  نشان دهنده  $\mathbf{x}$  امین شکل پایه و  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$  عامل های شکل می باشند. ظاهر مدل یک تصویر ( $\mathbf{I}$ ) می باشد که در آن  $\mathbf{I}$  مجموعه پیکسل های داخل شکل پایه  $\mathbf{x}$ . می باشد. مدل ظاهر یک چهره خاص از یک ظاهر پایه  $\mathbf{x}$ . و ترکیب خطی از بردارهای ویژه  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  تشکیل می شود که به صورت زیر تعریف می گردد.

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

که در آن  $a_0(\mathbf{x}), a_1(\mathbf{x}), \dots, a_m(\mathbf{x})$  عامل های ظاهر می باشند. عامل های شکل و ظاهر برای هر تصویر در فرایند  $\mathbf{I}$  بدست می آید. الگوریتم های  $\mathbf{PCA}$  و  $\mathbf{LDA}$  دو الگوریتم شناخته شده برای این منظور می باشند. رویکرد  $\mathbf{PCA}$  نسبت به  $\mathbf{LDA}$  در شرایطی که تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می باشد. در این مقاله از یک روش  $\mathbf{PCA}$  سریع برای حل مسئله بهینه سازی با  $100$  تکرار استفاده شده است. اگر  $\mathbf{I} = \mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n$  مجموعه عامل های بدست آمده باشد، یک تبدیل هم نسبی قطعه ای،  $\mathbf{H}(\mathbf{I})$  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  $\mathbf{x}$ . استفاده می شود که مقدار  $\mathbf{x}$  در آن صفر می باشد (شکل ۳-۶ قسمت).

شکل ۳ - مقدار دهی اولیه و بهینه سازی  $\mathbf{PCA}$  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  $64 \times 64$  بخش غیر هم پوشان با اندازه  $8 \times 8$  تقسیم می شود. سپس در هر بخش تبدیل  $\mathbf{PCA}$  اجرا می شود. ضرایب خروجی تبدیل  $\mathbf{PCA}$  بر حسب یک پویش زیگزاگی مرتب می شوند. اولین ضریب در نظر گرفته نمی شود. زیرا نشان دهنده میانگین سطح



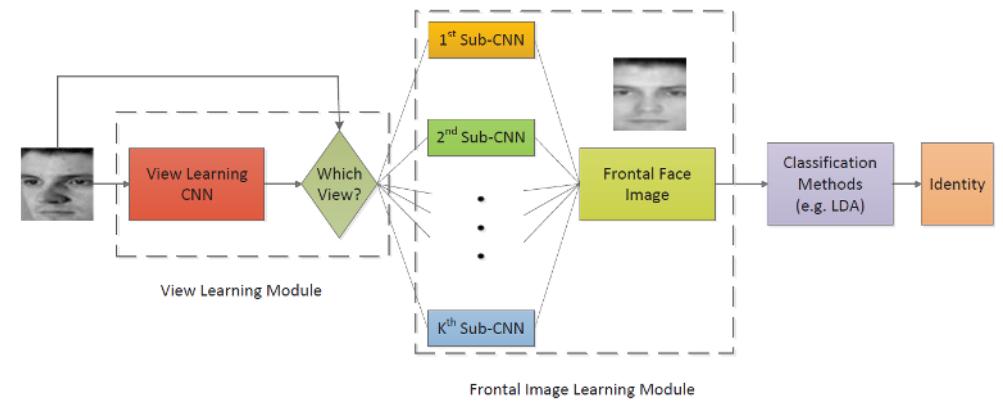
شکل ۶.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

خاکستری پیکسل های بخش می باشد. ۱۰ ضریب بعدی که ضرایب فرکانس پایین می باشند، برای ایجاد بردار ویژگی چهره استفاده می شوند. برای آموزش و آزمایش از مجموعه داده ۲۰۰۰۰ و ۲۰۰ استفاده شده است که در آن تصاویر چهره با زوایای چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این مقاله موفق به دستیابی به شناسایی چهره با دقت ۸۷.۳٪ شده است. شکل ۳ -



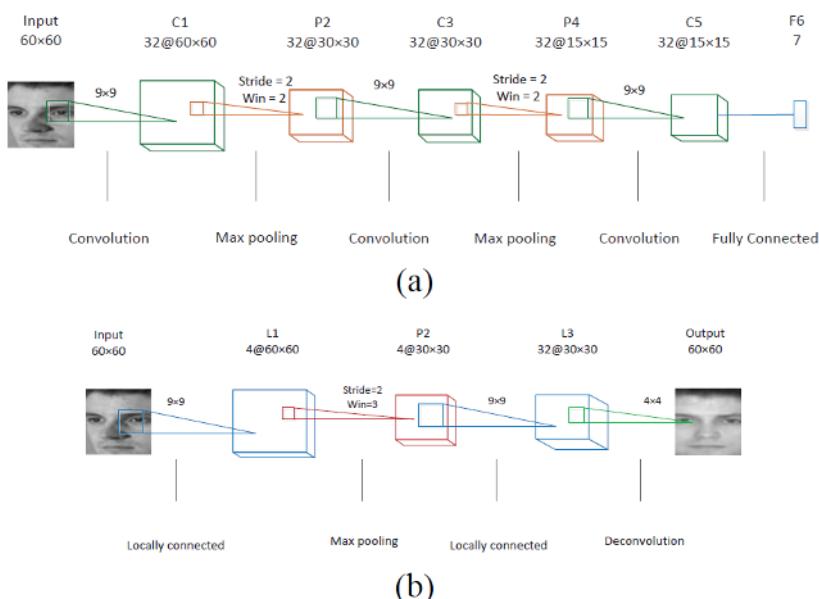
شکل ۷.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

نتیجه آزمایش بر روی مجموعه داده ۲۰۰۰۰ زاویه های متفاوت در سال ۲۰۱۶، ۲۰۰۰۰ و همکاران در [۲۹] یک روش رو به رو سازی چهره ارائه دادند که شناسایی چهره را مستقل از نمای چهره انجام می دهد. این الگوریتم یادگیری عمیق که ۲۰۰ 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 ۶-۳). این الگوریتم که می تواند با تعداد کمی داده نمونه، به خوبی آموزش ببیند، دو بخش تشکیل شده از شبکه عصبی پیچشی را به هم متصل می نماید تا مشکل نمای چهره در سامانه شناسایی چهره را برطرف نماید. در این معماری برای بازسازی چهره از زاویه رو به رو از لایه های واپیچشی به جای لایه های تمام متصل استفاده شده است. شکل ۳ - معماری شبکه پیشنهادی ۲۰۰۰۰ مدل ۲۰۰۰۰ از دو بخش اصلی تشکیل شده است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با بعد ۶۰ دریافت می کند و آن را با توجه به نمای چهره طبقه بندی می



شکل ۸.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

کند. سپس تصویر وارد بخش دوم که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از رو به رو با یک چهره از یک زاویه و نمای خاص می 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. شکل ۹.۳



شکل ۹.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

(a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو 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 های مربوط به نماهای مختلف

می باشد. برای لایه های پیچشی از تابع فعالیت  $\text{F}(x)$  استفاده شده است. و لایه تمام متصل از  $\text{F}(x)$  به عنوان تابع هزینه استفاده کرده است. (۳-۳)  $= \text{F}(x) = \text{loss}$  بخش دوم از تعدادی زیر شبکه پیچشی که هر کدام برای یادگیری تنباط بین چهره از رو به رو با یک چهره از یک نمای خاص می باشد، تشکیل شده است. هر یک از این زیر شبکه ها شامل دو لایه با اتصال محلی ، یک لایه رای گیری و یک لایه واپیچشی می باشند. سه لایه اول برای استخراج ویژگی ها و لایه آخر برای بازیابی چهره از رو به رو می باشند. ورودی و خروجی این لایه ها تصویر چهره می باشد. لایه آخر به جای لایه تمام متصل از لایه واپیچشی استفاده شده است.

زیرا حجم محاسبات را به طور قابل توجهی کاهش می دهد. یک لایه تماماً متصل به  $10^3$  میلیون پارامتر نیاز دارد، در حالی که لایه واپیچشی به  $460$  هزار پارامتر نیاز دارد. لایه اول که اتصال محلی دارد، از تابع  $\text{F}(x)$  به عنوان تابع فعالیت استفاده کرده است. لایه واپیچشی برای نمونه افرایی از درون یابی دو خطی استفاده کرده و تابع هزینه آن  $\text{loss} = \ell_2$  می باشد. برای یادگیری شبکه از الگوریتم پس انتشار خطأ استفاده شده است. الگوریتم  $\text{BP}$  به دقت  $64\% - 95\%$  در تشخیص چهره با زاویه  $45$  درجه رسیده است.

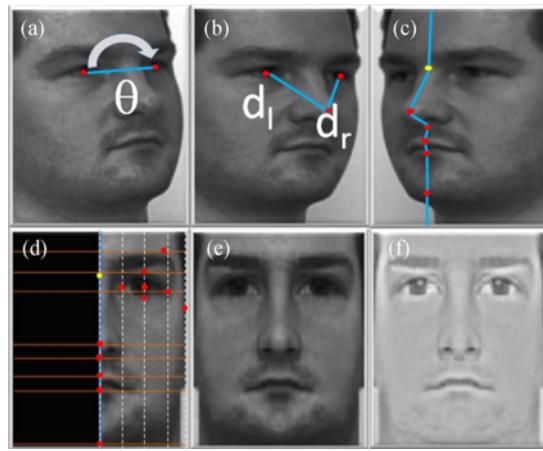
در سال  $2018$   $\text{BP}$  و همکاران در [۴۴] یک روش مبتنی بر  $\text{BP}$  برای شناسایی چهره در محیط های کنترل نشده با تعداد کم نمونه ها بر اساس محاسبه فاصله بین ویژگی های با بعد بالا که توسط شبکه عصبی پیچشی عمیق استخراج شده است ارائه دادند. این روش جدید شناسایی آماری، احتمال فاصله ها را نسبت به تمام تصاویر مجموعه داده ها با استفاده از قانون بیز به حداقل می رساند. این احتمال با تخمین توزیع هنجار طبیعی  $\text{N}(x|\mu, \sigma^2)$  بین ویژگی های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده های  $\text{MNIST}$  و  $\text{FCRN}$ - $\text{Gated}$  مورد آزمایش قرار گرفته است. با توجه به شکل  $14-3$  رویکرد پیشنهادی می تواند با استفاده از فواصل سنتی، افزایش دقت  $3.0$  تا  $5.5$  درصد در مقایسه با روش های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش، تفاوت زیادی داشته باشند. شکل  $3$  - مقایسه روش ارائه شده با سایر روش ها ( $\text{BP}$ )



شكل ١٠.٣: رویکرد مبتنی بر تطبیق کلیشه [؟].

[۲۶] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم  $\text{PCA}$  برای به دست آوردن  $68$  نقطه ویژه بر روی چهره تصویر آزمایشی (۱) و (۲) خروجی نادرست روش های دیگر (۳) خروجی روش ارائه شده در سال ۲۰۱۳  $\text{PCA}$  و همکاران در

استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه ( $\Theta$ ) محاسبه می شود و در صورتی که مقدار آن کمتر از یک آستانه باشد، تصویر مردود شده و در غیر این صورت به مرحله بعد برای هنجارسازی حالت فرستاده می شود. هرچه مقدار شاخص  $\Theta$  بالاتر باشد، تصویر چهره به حالت تمام رخ نزدیکتر است و اصلاح زاویه کمتری نیاز دارد. شکل ۱۱-۳ قسمت ۱۱ تا ۱۰ معیارهای مورد نیاز برای محاسبه شاخص  $\Theta$  را نشان می دهد. شکل ۱۱-۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره



شکل ۱۱.۳: رویکرد مبتنی بر تطبیق کلیشه [؟].

چرخش: چرخش سر در جهت عقربه های ساعت یا عکس آن می باشد. و به صورت زاویه  $\Theta$  تعریف می شود که زاویه بین خط عبوری از مرکز چشم ها و محور افقی  $\Theta$  می باشد.

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

انحراف: چرخش در راستای محور افقی است و  $\Theta$  و  $\Phi$  فاصله مرکز چشم چپ و راست از نوک بینی می باشد. اندازه گیری این فاصله ها در صورت برابر بودن، برای تشخیص تمام رخ بودن تصویر چهره مورد استفاده قرار می گیرد.

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

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

$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)} \quad (A.3)$$

با محاسبه ۳ شاخص فوق، شاخص  $\text{MA}$  محاسبه می شود:

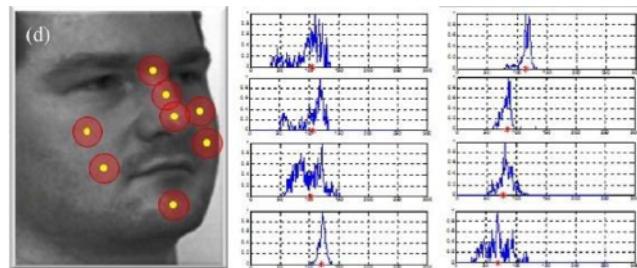
$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch) \quad (6.3)$$

که در آن

$$\alpha + \beta + \gamma = 1 \quad (V.3)$$

که مقادیر این ضرایب از طریق آزمون و خطا به دست می‌آیند. سپس در مرحله تمام رخ کردن تصویر چهره، بین دو فاصله ۰۰ و ۰۰ هر کدام بزرگتر باشند، نشان می‌دهد آن سمت از چهره بیشتر در دید دوربین است. اگر نیمه سمت راست صورت به طرف دوربین باشد (۰۰)، تصویر بدون تغییر باقی می‌ماند. در غیر این صورت، تصویر حول محور عمودی برعکس می‌شود که باعث می‌شود همیشه نیمه سمت راست تصویر پردازش شود. سپس برای ثابت کردن طول سطرها، سطرها بسط داده می‌شوند. مطابق شکل ۱۱-۳ قسمت ۰ و ۰ 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. 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 [۲۶] و همکاران در ۰۰۰۰۰۰۰۰ یک روش هنجار سازی نورپردازی برای تصاویر چهره ارائه دادند و از این روش برای محاسبه شاخص روشنایی نمونه (۰۰) 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، گونه ها و چانه معمولاً نور را منعکس می کنند. ۸ ناحیه در شکل ۱۲-۳ با توجه به چنین اصلی انتخاب شده اند. ۸ بافت نگار با رنگ آبی و مرکز آن ها با رنگ قرمز مشاهده می شود. شکل ۱۲-۳ - اندازه گیری روشنایی و بافت



شکل ۱۲-۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

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 ثقل آن محاسبه می شود:

$$mc(w) = \frac{\sum_{i=0}^{255} i \times h_w(i)}{\sum_{i=0}^{255} h_w(i)} \quad (8.3)$$

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

$$SI = 1 - F(std(mc)) \quad (9.3)$$

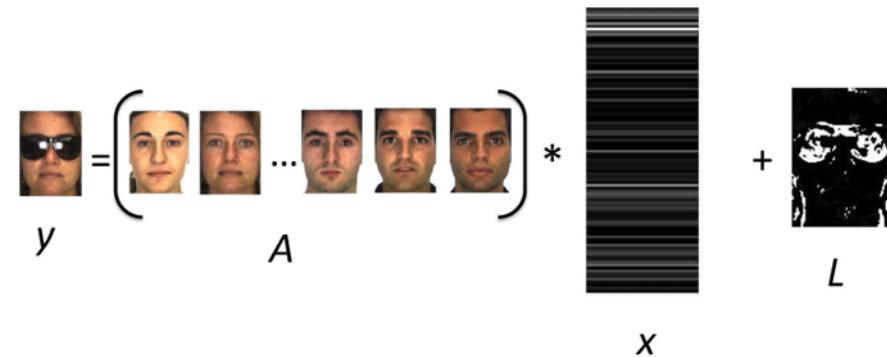
هرچه مقدار  $w$  بیشتر باشد، یعنی تصویر روشنایی یکنواخت تری دارد. اگر این شاخص به اندازه کافی رضایت بخش نباشد، تصویر رد می شود. در غیر این صورت برای هنجارسازی روشنایی وارد بخش بعدی خواهد شد. در صورت رد شدن تصویر، سیاست های جایگزین برای رسیدگی به این موضوع در دسترس هستند. برای مثال ممکن است یک نمونه جدید 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  $\mathbf{H}$  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

در شکل می باشد. در سال ۲۰۱۵ میلادی این تحقیق و همکاران در [۳۶] رویکردی برای تشخیص چهره در تغییرات شدید روشنایی پیشنهاد دادند که به سه مرحله تقسیم شده است: برای اصلاح روشنایی غیر یکنواخت، همسان سازی بافت نگار براساس بر اساس ناحیه استفاده می شود. ویژگی های مبتنی بر  $\mathbb{R}^d$  از تصویر چهره استخراج می شود. فرایند طبقه بندی بر اساس مدل  $\mathbb{R}^{d \times d}$  انجام می شود. چالش انسداد در سال ۲۰۱۸ میلادی این تحقیق و همکاران در [۳۸] 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 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  $\mathbb{R}^d$  تعداد داده های آموزشی و  $\mathbb{R}^n$  حاصل ضرب تعداد پیکسل های طول و عرض تصاویر می باشد. داده های آموزشی چهره های طبیعی و بدون انسداد می باشند.  $y = \mathbb{R}^d$  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

$$\text{خطای } L = \mathbb{R}^d \text{ نیز می باشد. (شکل ۱۶-۳)}$$

$$y = Ax + L \quad (10.3)$$

که در آن  $\mathbb{R}^d$  بردار ضرایب با  $\mathbb{R}^n$  بعد می باشد. شکل ۱۳-۳ - تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده



شکل ۱۳.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

و یک تصویر  $\mathbb{R}^d$  که نشان دهنده انسداد است، تشکیل شده است برای آنکه شرط تنک بودن به رابطه بالا اضافه شود، مسئله به صورت زیر نوشته می شود:

$$\operatorname{argmin}_{x \in \mathbb{R}^d} \|s.t.y - Ax\| \quad (11.3)$$

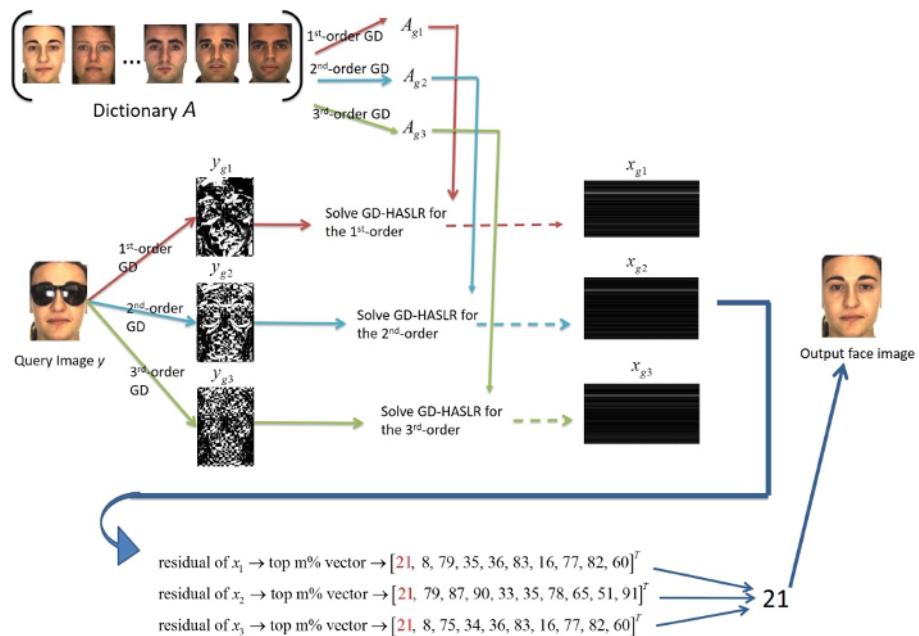
که در آن  $\epsilon$  یک آستانه خطای می باشد. برای تصویر و ورودی و تصاویر مجموعه داده آموزش، گرادیان مرتبه اول، دوم و سوم محاسبه شده و به عنوان ویژگی هر تصویر در نظر گرفته می شود. در ادامه شرط کم رتبه بودن ماتریس ویژگی ها نیز به این رابطه اضافه می شود. با استفاده از روش ضرایب لاغرانژ، رابطه بالا را می توان به صورت یک مسئله بهینه سازی بدون محدودیت نوشت و حل نمود.

$$L(x, L, z) = \alpha LM + xi + zT(y - Ax - L) + 2y - Ax - L22 \quad (12.3)$$

که در آن  $\alpha$  ضریب لاغرانژ و  $\beta$  عامل مجازات می باشد. پس از بدست آوردن بردار تنک  $\beta$  می توان باقیمانده دسته  $\alpha$  را به صورت زیر محاسبه نمود:

$$r_i = y - Ai(x)2 \quad (13.3)$$

که در آن  $(x)\beta$  نشان دهنده  $\alpha$  امین انتخاب کننده دسته می باشد که فقط ورودی های مربوط به دسته  $\alpha$  را حفظ می کند و در سایر قسمت ها برابر با صفر می باشد. در نهایت دسته ای که کمترین باقیمانده را داشته باشد، انتخاب می شود. رویکرد کلی الگوریتم در شکل ۱۴-۳-رویکرد کلی الگوریتم  $\alpha-\beta-\gamma$  در سال ۲۰۱۴.  $\alpha$  و همکاران در [۳۷] یک روش تشخیص چهره پوشیده شده در پس زمینه پیچیده ارائه کردند. 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 استفاده شده است. از آنجا که سرعت  $\alpha$  نسبتاً پایین است، از  $\beta$  به همراه آن نیز استفاده کرده اند. در مرحله دوم از طبقه بند  $\gamma$  برای طبقه بندی چهره های پوشیده شده استفاده شده است که برای انواع پوشیدگی آموزش داده شده است. چالش کمبود تصاویر آموزشی در سال ۲۰۱۷  $\alpha$  و همکاران در [۳۹] 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



شکل ۱۴.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

است، نوآوری این رویکرد، ساختن یک مجموعه داده استاندارد برای شبکه عصبی از روی دوربین های ناظارتی 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  $\text{MTCNN}$  مجموعه تصاویری از آن به عنوان یک دسته طبقه بندی می شود. شکل ۱۵.۳ - ردیابی، یافتن چهره ها و برچسب زنی برخی تصاویر در هر دسته به اشتباه در مرحله اول به عنوان تصویر



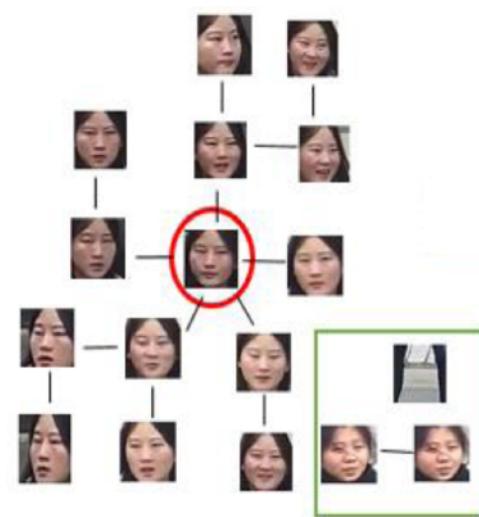
شکل ۱۵.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

یک فرد در نظر گرفته شده اند (شکل ۱۶-۳). شکل ۱۶-۳ - تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند با استفاده از روش خوشه بندی گراف روی ویژگی های استخراج شده از شبکه،  $\text{MTCNN}$ - $\text{ResNet}$  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



شکل ۱۶.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

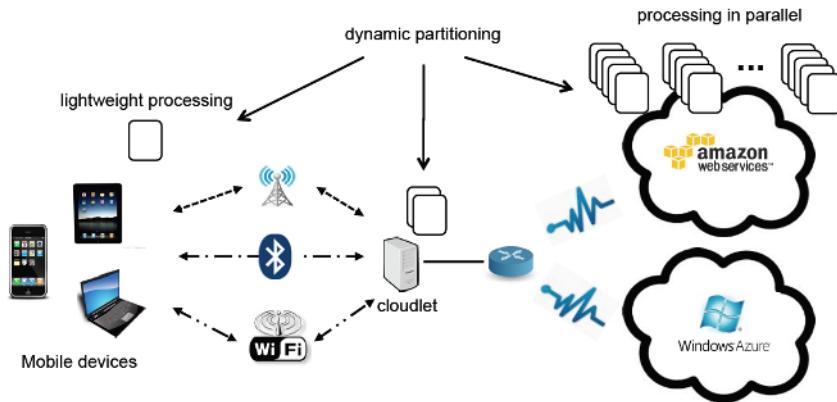
باشد، این تصاویر متعلق به یک فرد هستند. با توجه به شکل ۱۷-۳ تصویری که بیشترین شباهت را به تصاویر دیگر دارد، به عنوان شاخص برای آن شخص انتخاب می‌شود. شکل ۱۷-۳ - استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص با استفاده از



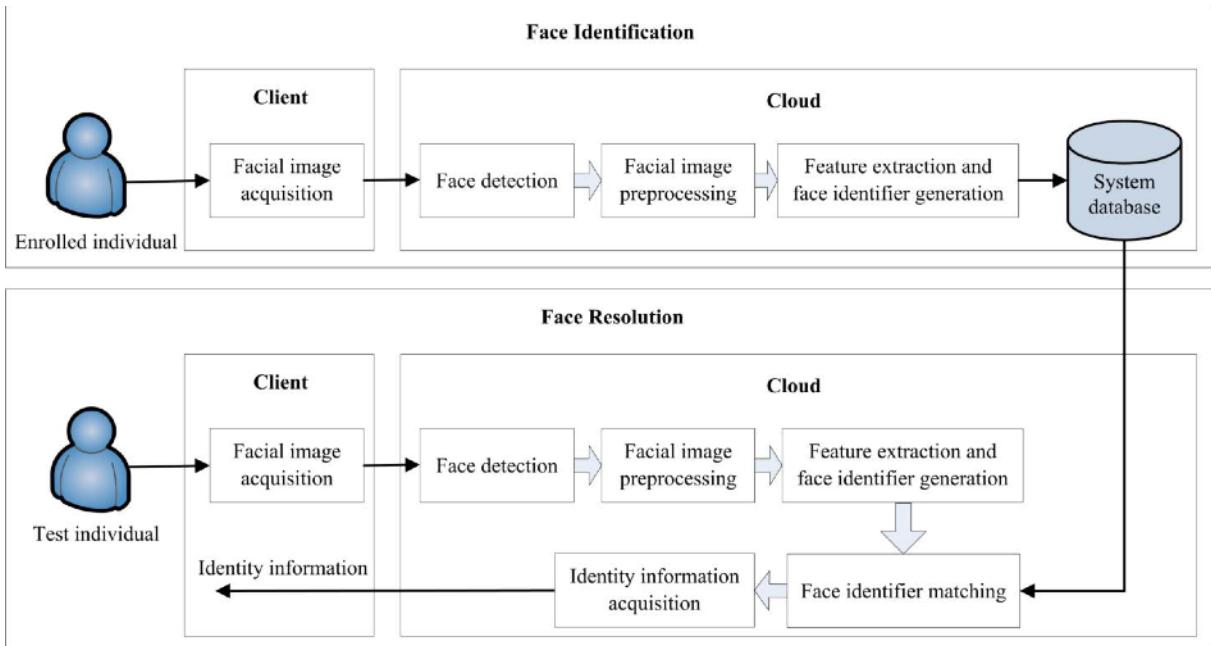
شکل ۱۷.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

محاسبه فاصله بین هر داده با داده مركزی و در نظر گرفتن یک آستانه، داده های تکراری در هر دسته مشخص شده و حذف می شوند. با توجه به مقدار داده های درون هر دسته، تصفیه بین دسته ای انجام می شود. اگر مجموعه داده های درون هر دسته کمتر از ۱۰۰ تصویر باشد، آن دسته از مجموعه داده حذف می شود. دقت خوشه بندی و جمع آوری مجموعه داده ۲٪.۹۹ شده است. در نهایت از یک مدل پیش آموزش دیده شده شبکه همراه با ۰۰۰۰۰۰۰۰-۰۰۰۰۰۰۰۰ همراه با طبقه بندی تصاویر آزمایشی استفاده شده است که به دقت ۹۲٪ رسیده است. چالش منابع محدود در سال ۲۰۱۲ ۰۰۰۰۰۰۰۰ و همکاران در [۴۰] یک روش تشخیص چهره بی درنگ مبتنی بر بینایی ابری با استفاده از معماری ۰۰۰۰۰۰۰۰ ارائه کردند (شکل ۱۸-۳). با فراغت 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

قابل توجهی افزایش می دهند. شکل ۳ - معماری : دستگاه های تلفن همراه از طریق اتصال چندگانه با کامپیوتر

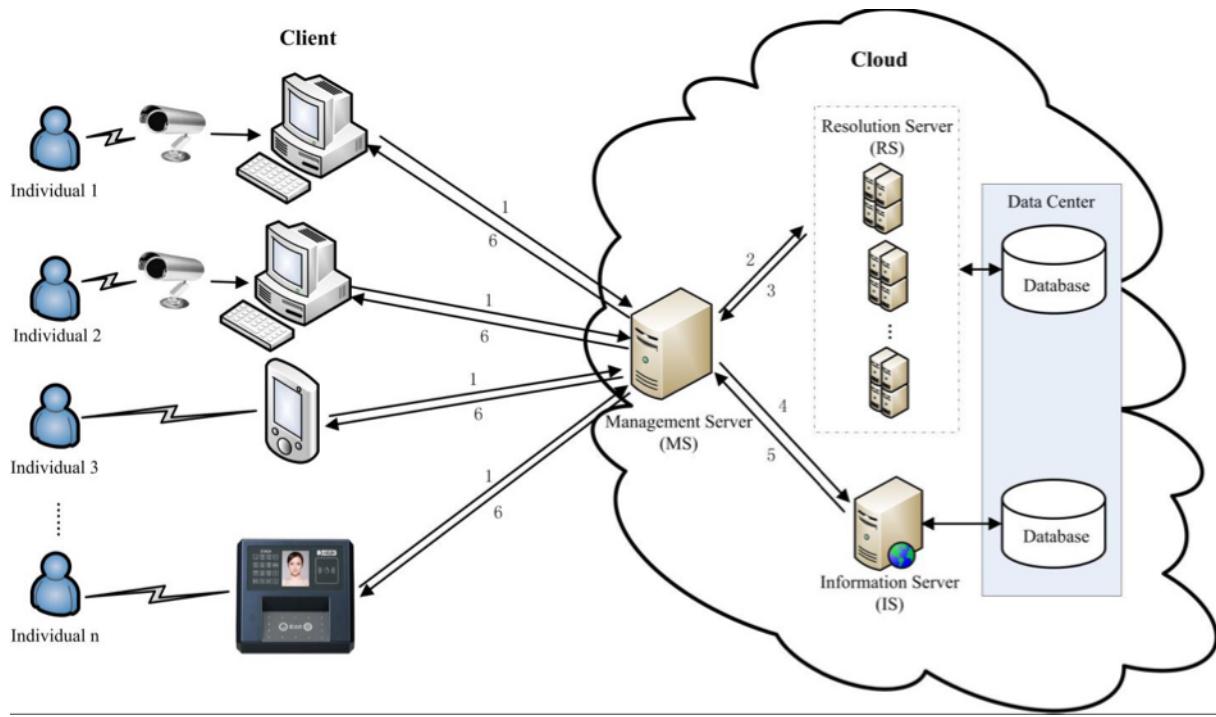


شکل ۱۸.۳: رویکرد مبتنی بر تطبیق کلیشه [؟].



شکل ۱۹.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری در این سامانه تصویر با فرمت  $\text{JPEG}$  برای ابر ارسال شده و برای یافتن چهره از ویژگی های  $\text{RGB}$  استفاده شده است. سپس عملیات همسان سازی بافت نگار بر روی تصویر چهره اعمال می شود تا بهبود جزئی حاصل شود. سپس از الگوریتم  $\text{PCA}$  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 ۳-۲۰ مشاهده می شود این سامانه ابری از بخش های سرور مدیریت ( $\text{DB}$ )، سرور اطلاعات ( $\text{DB}$ )، سرور شناسایی ( $\text{DB}$ ) و پایگاه داده تشکیل شده است. به علت قدرت بالای پردازش در سرور ابری، امکان پردازش موازی نیز در این سامانه وجود دارد که باعث افزایش سرعت محاسبات و کاهش زمان پاسخ دهی سامانه می گردد. شکل ۳-۲۰ چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری نتیجه گیری به تازگی یادگیری عمیق در تشخیص چهره و سیاری از زمینه های هوش مصنوعی به راه حل غالب تبدیل شده است. ما یک سوال مطرح می کنیم: آیا یادگیری عمیق واقعاً مسئله تشخیص چهره به دلیل طیف گسترده ای از تغییرات در ظاهر چهره ها تشخیص چهره چیست؟ در مقایسه با تشخیص شیء عمومی، تشخیص چهره به دلیل طیف گسترده ای از تغییرات در ظاهر چهره ها 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 داده های آموزشی از مواردی می باشد که می توانند سامانه تشخیص چهره را با مشکل رو به رو نمایند. از طرفی اکثر مجموعه داده ها تنها شامل چند هزار عکس می باشد. یک مجموعه داده حاوی اطلاعات بدون محدودیت و مقیاس بزرگ، سامانه چارچوب چهره را به چالش های همچون گراییش های شدید، نور کم و تصاویر کوچک و تاریک چهره تبدیل



شکل ۳: رویکرد مبتنی بر تطبیق کلیشه [؟].

می کند. محققان فرض کرده اند که لایه های عمیق  $\text{CNN}$  ها می توانند اطلاعات انتزاعی بیشتری مانند هویت، ظاهر و ویژگی ها را

رمزگذاری کنند؛ با این حال هنوز هنوز کاملا مطالعه نشده است که لایه ها دقیقا با ویژگی های محلی برای تشخیص مطابقت دارند.

برای شناسایی چهره، عملکرد یادگیری را می توان با یادگیری یک معیار اندازه گیری فاصله متمایز کننده بهبود داد. با این حال، با

توجه به محدودیت های حافظه کارت گرافیک ها، نحوه انتخاب جفت ها یا سه گانه های اطلاعاتی و روش های آموزش آنلاین (به

عنوان مثال، گرادیان نزولی) در مجموعه داده های بزرگ، هنوز یک مشکل باز است. یکی دیگر از مشکلات چالش برانگیز این است

که پردازش ویدیو در شبکه های عمیق را برای استفاده از تجزیه و تحلیل چهره مبتنی بر ویدئو ترکیب کند.

دلیل اصلی به وجود آمدن چالش ها 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 سیاری 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

کمیعد 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.

مقاله [۳] از شبکه‌های مولد تخصصی برای تولید داده‌ها استفاده کرده است که به اختصار GAN نامیده می‌شوند. GAN از دو شبکه مستقل تولید کننده و تمیز دهنده استفاده تشکیل شده است. شبکه تولید کننده از روی بردار  $\theta$  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  $\theta$  معرفی می‌شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می‌شود. به عنوان مثال بردار  $\theta$  برای تصویر خانمی که عینک آفتابی نزدی است از بردار  $\theta$  برای تصویر خانمی که عینک آفتابی زده است، کم می‌شود و حاصل آن، بردار مربوط به یک عینک آفتابی می‌باشد. سپس این بردار با بردار تصویر آقایی که عینک نزدی است جمع می‌شود. نتیجه نهایی تصویر همان آقا با عینک آفتابی می‌باشد. به عنوان مثالی دیگر، با میانگین گیری بردارهای مربوط به دو تصویر از چهره شخصی که به سمت راست و چپ متمایل است، توانسته چهره رو به روی شخص را بازسازی نماید. اما کیفیت کار هنوز تا حالت مطلوب فاصله دارد.



شکل ۳: رویکرد مبتنی بر تطبیق کلیشه [۴].

شکل ۲ GAN تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۳] مقاله [۴] یک روش مبتنی بر GAN 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 و ResNet استخراج شده است ارائه دادند. این روش جدید

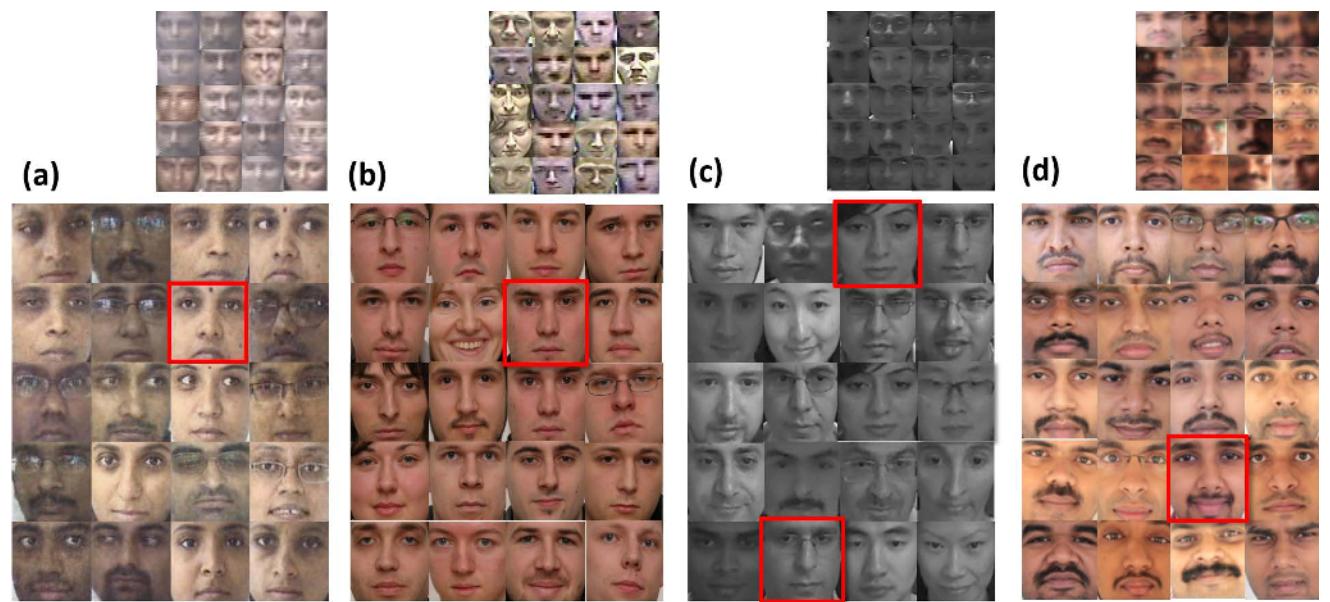
شناسایی آماری، احتمال فاصله‌ها را نسبت به تمام تصاویر مجموعه داده‌ها با استفاده از قانون بیز به حداکثر می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی  $\pi_{\text{non}}(\theta)$  بین ویژگی‌های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده‌های  $\mathcal{D}_1$  و  $\mathcal{D}_2$  مورد آزمایش قرار گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقت ۳۰ تا ۵۵ درصد در مقایسه با روش‌های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند.

مقاله [۶] یک شبکه عمیق مبتنی بر  $\mathcal{D}_1$  را با نام  $\pi_{\text{non}}(\theta)$  پیشنهاد می‌دهد، که تصاویر واقع گرایانه باوضوح بالا از روی تصاویر باوضوح پایین بازسازی می‌کند. این تصاویر چهره غیر واقعی اما واقع گرایانه و با کیفیت، باعث عملکرد بهتر سامانه شناسایی چهره برای مقایسه تصاویر می‌شود. رویکرد اصلی مقاله در روش آموزش تخصصی  $\pi_{\text{non}}(\theta)$  بهینه سازی تابع ضرر بازسازی چند مقیاسی است، بر اساس شاخص‌های مانند: شاخص شباهت ساختاری چند مقیاسی،  $(\mathcal{D}_1)$  میانگین مربعات خطای هر قسمت  $(\mathcal{D}_1)$  و اگرایی جنسن شانون اصلاح شده  $(\mathcal{D}_1)$  و تنوع متقابل در اطلاعات.  $(\mathcal{D}_1)$  شبکه تمیز دهنده در  $\mathcal{D}_1$  بر اساس اطلاعات طبقه بندی که به طور ضمنی در طول آموزش آموخته می‌شود، هویت هر شخص را حفظ می‌کند. این رویکرد سریعتر از شبکه‌های مبتنی بر  $\mathcal{D}_1$  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 باوضوح بالا از روی تصاویر باوضوح پایین در ۴ مجموعه داده مختلف. موارد با حاشیه قرمز خروجی اشتباه هستند - [۶] شبکه‌های  $\mathcal{D}_1$  یاد می‌گیرند تصاویر جدیدی تولید کنند که شبیه به تصاویر واقعی باشند. اما این شبکه‌ها معمولاً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند. مقاله [۷] یک شبکه  $\mathcal{D}_2$  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 غالباً 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 ردیف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند - [۷] مقاله [۸] یک شبکه  $\mathcal{D}_2$  جدید پیشنهاد می‌دهد که ورودی آن یک تصویر و خروجی آن نیز یک تصویر می‌باشد. این روش پیشنهادی که به روش  $\mathcal{D}_1$

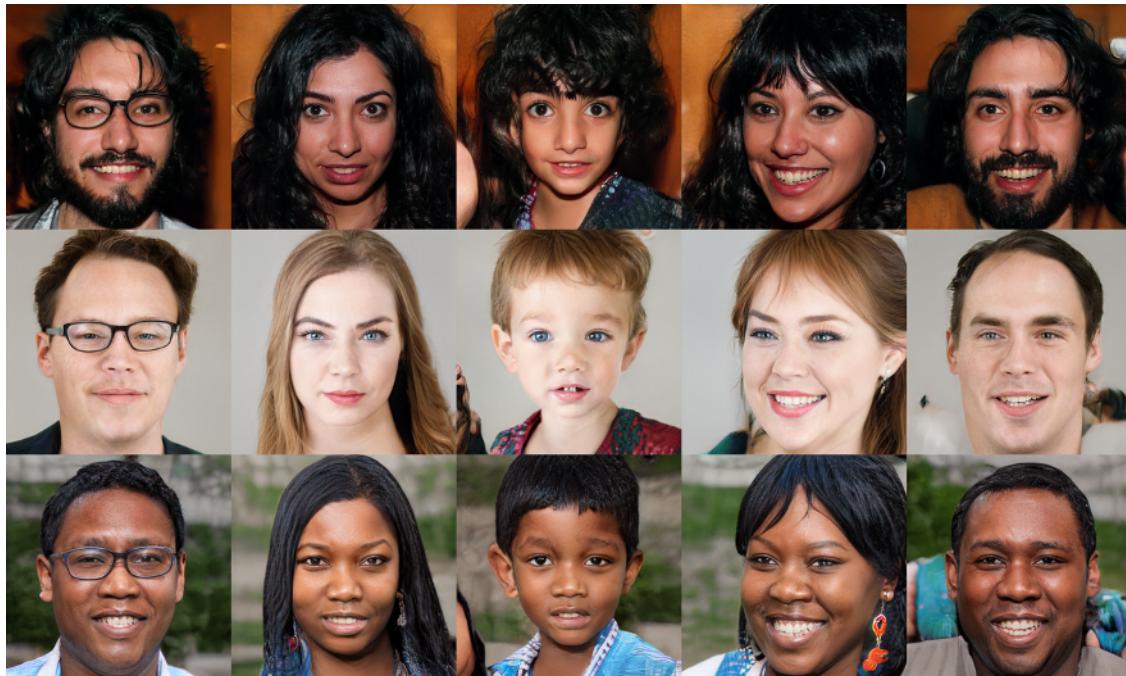
آن نیز یک تصویر می‌باشد که تغییرات مورد نظر بر روی آن اعمال شده باشد. این رویکرد از دو شبکه تولید کننده و تمیز دهنده تشکیل شده است. مجموعه داده‌های آموزشی آن به گونه‌ای است که به ازای هر تصویر، یک تصویر برچسب نیز موجود می‌باشد. هدف از آموزش شبکه تولید کننده آن است که برای هر تصویر ورودی، تصویر خروجی را سیار نزدیک به برچسب همان تصویر تولید نماید.



شکل ۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

شکل ۳ به ترتیب هر چپ به راست: تصویر ورودی برچسب تصویر ورودی تصویر خروجی - [۸] شبکه تولید کننده از معماری استفاده می‌نماید که ورودی آن تصویر و خروجی آن نیز تصویر می‌باشد. معماری شبکه تمیز دهنده به گونه‌ای می‌باشد که تصویر تولید شده و برچسب اصلی را به عنوان ورودی گرفته و با یکدیگر مقایسه می‌نماید. مقاله [۹] یک شبکه جدید پیشنهاد می‌دهد که ورودی آن یک تصویر و خروجی آن نیز یک تصویر می‌باشد. این روش پیشنهادی که به روش [۸] یا [۹] شناخته می‌شود، مشابه مقاله [۸] به گونه‌ای طراحی شده که به عنوان ورودی یک تصویر را دریافت می‌نماید و خروجی آن نیز یک تصویر می‌باشد که تغییرات مورد نظر بر روی آن اعمال شده باشد. مجموعه داده‌های آموزشی آن شامل دو مجموعه مجزا می‌باشد. به گونه‌ای که مجموعه اول شامل تصاویری می‌باشد که میخواهیم تغییر کنند و شبیه تصاویر مجموعه دوم شوند.

شکل ۳ حفظ محتوى و انتقال از تصاویر مجموعه ۱ به تصاویر مجموعه ۲ و برعکس - مقاله [۹] شبکه تولید

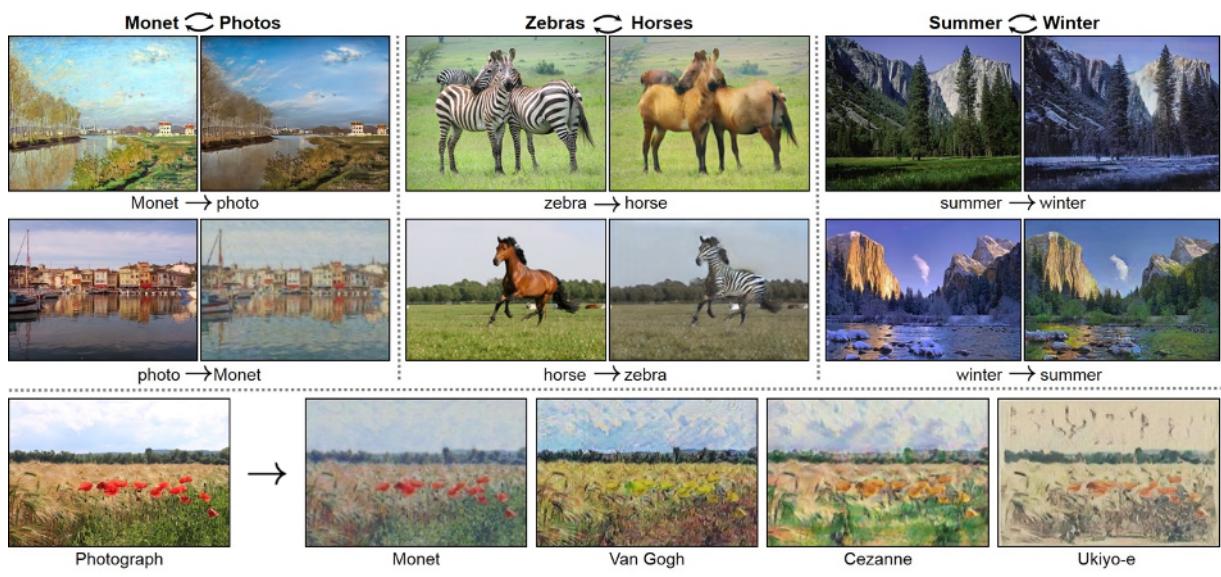


شکل ۳: ۲۳.۳: رویکرد مبتنی بر تطبیق کلیشه [۹].

کننده از دو شبکه  $\Phi_{\text{FaceNet}}$  تشکیل شده است که به صورت موازی با یکدیگر آموزش داده خواهد شد که هدف آن دستیابی به وزن‌هایی می‌باشد که بتوان به کمک آن تصاویر مجموعه اول را به تصاویری شبیه به تصاویر مجموعه دوم تبدیل نمود. یک تصویر از مجموعه اول وارد شبکه اول شده و تبدیل به یک تصویر خروجی می‌شود. تصویر خروجی وارد شبکه دوم شده و دوباره به تصویر مشابه تصویر اولیه تبدیل می‌شود. شبکه تمیز دهنده باید این دو تصویر را با یکدیگر مقایسه کرده و خطای مناسب را برای به روز رسانی وزن‌های شبکه برگرداند. همین اتفاق برای تصاویر مجموعه دوم تکرار می‌شود.

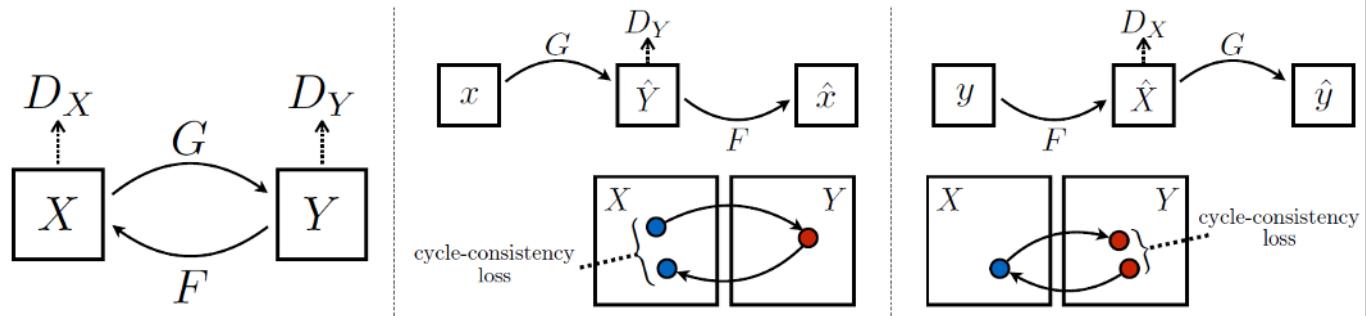
شکل ۹: ساختار شبکه  $\Phi_{\text{FaceNet}}$  و نحوه محاسبه خطا به صورت انتزاعی - [۹]

در مقاله [۱۰] به موضوع تولید چهره در زوایای دلخواه پرداخته شده است. در این مقاله از دو شبکه  $\Phi_{\text{FaceNet}}$  استفاده شده است که در شبکه اول از روی چهره زاویه‌دار، چهره روبرو تولید شده است. سپس با استفاده از شبکه  $\Phi_{\text{FaceNet}}$  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



شكل ٢٤.٣: رویکرد مبتنی بر تطبیق کلیشه [؟].

می کند.



شکل ۲۵.۳: رویکرد مبتنی بر تطبیق کلیشه [؟].

شکل ۱۰ ساختار شبکه برای رو به رو سازی چهره و شبکه برای تولید چهره از زوایای مختلف

[۱۰]- چالش‌های موجود بیشتر سامانه‌های تشخیص چهره عملکردهای قابل قبولی را در محیط‌های کنترل شده ارائه می‌دهند.

اما در محیط‌های بدون محدودیت و در مععرض تحریب شدید تصاویر چهاره، عملکرد خوبی، ندارند و در کاربردهای واقعی، هنوز مسیری

طولانی، پرای بجهود در پیش دارند. از جمله جالترهای مهم، اساسی، و عمومی، در سامانه‌های تشخیص چهاره‌گی، تواند به موارد زیر

اشری نمود:

تغییر زاویه و حالت جهت نسبت به دو ریزین (۱۰۰۰)

انسداد صورت توسط اشیاء مانند عنک آفتاب و شاگرد (شکر)

تغییرات اساسی در چهره با گذر زمان، مانند رشد موها و ریش‌ها و یا بالا رفتن سن مانند سفید شدن موها (Figure 1)

تاری خارج از تمرکز دوربین (Out-of-focus)

وضوح پایین تصویر (Blur)

ردیابی چهره در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی (Frame-to-frame)

دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چشم 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 سیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویربرداری شده باشد.

مقاله [۵] روشنی برای رو به رو سازی تصویر چهره پیشنهاد کرده بود که در برخی موارد، چهره را به خوبی می‌چرخاند، اما در نیمی

از موقع نیز نتیجه خروجی الگوریتم، تصویر چهره را دچار اعوجاج‌هایی می‌نماید که روند تشخیص چهره را با مشکل بیشتری مواجه

می‌سازد. از این رو فرایند رو به رو سازی به طور میانگین کمک شایانی به بالا رفتن دقت تشخیص چهره نمی‌نماید. مقاله [۳] روشنی

مبتنی بر [۱] برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی موقع به تصویر چهره لطمہ وارد می‌نماید

به طوری که شخص مورد نظر قابل شناسایی توسط سامانه یادگیری نمی‌باشد. مقاله [۶] 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. مقاله [۸] نیازمند مجموعه داده‌های آموزشی است که به ازای هر تصویر، یک تصویر برچسب نیز در آن موجود باشد.

محدودیت این معماری‌ها، نیاز به تصاویر جفت برای آموزش می‌باشد. به گونه‌ای که به ازای هر تصویر رو به رو، باید یک تصویر

زاویه دار با همان شرایط از همان شخص نیز داشته باشیم. با توجه به شرایط این مسئله، این موضوع عملی نمی‌باشد. در مقاله [۱۰]

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 مواجهه با این مسئله راهکاری ارائه

نمی‌دهد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند.

فصل ۴

## روش پیشنهادی

همان طور که در فصل اول شرح داده شد، شناسایی میکروآنوریسم به عنوان اولین علامت رتینوپاتی دیابتی بسیار مهم و حیاتی است و تشخیص زودهنگام این عارضه می‌تواند به تسريع درمان و جلوگیری از تداوم بیماری و نابینایی فرد مبتلا کمک کند. بنابراین در این فصل تلاش می‌کنیم تا روشی برای تشخیص بهتر و دقیق‌تر میکروآنوریسم در تصاویر فوندوس رنگی پیشنهاد دهیم. از آن جایی که میکروآنوریسم‌ها بسیار کوچک هستند شناسایی آن‌ها در تصاویر فوندوس بسیار سخت و دشوار است. همچنین در مواردی شباهت این عارضه به عارضه‌های دیگری نظیر هموریج و گاها محو شدگی عارضه در پس‌زمینه تصاویر فوندوس کار را از آن‌چه هست سخت‌تر خواهد کرد. بنابراین ما روشی برای تشخیص میکروآنوریسم در تصاویر فوندوس پیشنهاد کردایم که در ابتدا نواحی کاندید را استخراج کرده و سپس از بین نواحی کاندید، بنا را بر توجه به نواحی گذاشته و نواحی اصلی میکروآنوریسم را مشخص می‌کند تا از این طریق بتوان به کاهش مشکلات ذکر شده کمک کرد. این روش به ترتیب شامل گام پیش‌پردازش، آموزش شبکه عصبی پیچشی اول مبتنی برای استخراج نواحی کاندید، آموزش شبکه عصبی پیچشی دوم برای شناسایی میکروآنوریسم‌ها از نواحی کاندید و تکمیل بردار ویژگی شبکه عصبی پیچشی دوم با ویژگی‌های دست‌ساز است. روش پیشنهادی همان‌طور که شرح داده شد روش پیشنهاد شده در این کار روپرکردی مبتنی بر توجه به نواحی کاندید است تا در دو مرحله بتوانیم نواحی میکروآنوریسم را شناسایی کنیم. به این ترتیب که در ابتدا نواحی مشکوک و کاندید را شناسایی کرده و سپس از بین این نواحی به شناسایی میکروآنوریسم‌های اصلی و حذف مثبت‌های کاذب می‌پردازیم. در شکل ۳ تصویر کلی از روش پیشنهادی برای حل مسئله تشخیص میکروآنوریسم در تصاویر فوندوس رنگی را مشاهده می‌کنید.

### شکل ۳ - نمای کلی از روش پیشنهادی برای تشخیص میکروآنوریسم

پیش‌پردازش شامل تصحیح گاما و اعمال فیلتر دوطرفه به منظور حذف نویز پس‌زمینه است. در ادامه به شرح مراحل پیش‌پردازش می‌پردازیم. در راستای یکسان‌سازی رنگ پس‌زمینه تصاویر فوندوس در مجموعه داده و همچنین حل مشکل نورپردازی غیریکنواخت در تصاویر، سعی در تخمین پس‌زمینه داریم. این کار از طریق اعمال یک فیلتر میانه با هسته ۳۳ در ۳۳ پیکسل بر روی تصویر اصلی انجام می‌شود. پس از تخمین، پس‌زمینه را از تصویر اصلی کم می‌کنیم تا مشکل را رفع کرده باشیم.<sup>[۳۱]</sup> در مرحله بعد نواحی کاملاً تاریک در اطراف دایره تصویر فوندوس به کمک نقاب دایروی حذف شده و تصویر را به شکل مربع در می‌آوریم. در شکل ۳ نتیجه این مراحل پیش‌پردازش نشان داده شده است.

### شکل ۳ - حذف پس‌زمینه و نواحی تاریک در اطراف تصاویر فوندوس

در ادامه روند پیش‌پردازش، نوبت به تصحیح گاما می‌رسد. با این هدف که با استفاده از یک تبدیل غیرخطی روشنایی تصویر بهبود یابد. پس از پیش‌پردازش بالا، تصویر نسبت به قبل بهبود پیدا می‌کند اما پس از تکه کردن تصویر جهت آموزش شکه، تاریک بودن تصاویر مشهود است. برای حل این مشکل از گاما ۸.۰ استفاده می‌کنیم تا تصویر کمی روشن تر شود. همچنین با تکه کردن تصویر، نویزهایی در پس‌زمینه مشاهده می‌شود که ممکن است در فرآیند آموزش شبکه را دچار خطا کند. برای حذف این نویزها و در عین حال حفظ لبه‌ها در تصویر از فیلتر دوطرفه استفاده می‌کنیم. نتیجه اعمال این دو فرآیند را در تصویر زیر مشاهده می‌کنید. این نکته حائز اهمیت است که نتیجه حذف نویز در تکه‌های ایجاد شده از هر تصویر قابل مشاهده است.

شکل ۳-۳- قبل و بعد از اعمال تصحیح گاما و فیلتر دوطرفه

شکل ۳-۴- قبل و بعد از اعمال تصحیح گاما و فیلتر دوطرفه در یک تکه بیمار مبتتنی بر ناحیه استخراج تکه برای آموزش شبکه نیاز داریم تا تصاویر فوندوس پیش‌پردازش شده را تکه تکه کنیم. زیرا میکروآنوریسم بسیار ریز بوده و تشخیص آن در یک تصویر کلی برای شبکه کاری بسیار دشوار است. از آنجایی که در تصویر برچسب‌زده شده، نقاط میکروآنوریسم دقیقاً مشخص شده‌اند می‌توانیم با تکه کردن تصویر و مشخص کردن مناطق میکروآنوریسم در هر تکه، آموزش را برای شبکه ساده‌تر کنیم. زیرا به جای دادن کل تصویر به شبکه، تکه‌های ایجاد شده را جهت آموزش به شبکه می‌دهیم. برچسب‌زدن به تکه‌ها در بسیاری از مقالات در این زمینه و زمینه‌های مشابه برای برچسب زدن به این روش عمل کردند که اگر پیکسل مرکز تکه، یک پیکسل از ناحیه میکروآنوریسم باشد، به کل تکه برچسب بیمار یا یک تخصیص می‌یابد و اگر پیکسل مرکز ناحیه‌ای از میکروآنوریسم نباشد برچسب سالم یا صفر برای تکه در نظر گرفته می‌شود. این روش برچسب زدن یک مشکل عمده دارد. اگر شبکه با این روش برچسب‌زدن آموزش باید، شبکه یاد می‌گیرید که با توجه به نواحی مشاهده شده، برای پیکسل مرکز تصمیم بگیرد که سالم است یا حاوی میکروآنوریسم. بنابراین برای آزمون یک تصویر فوندوس کلی باید به ازای تمام پیکسل‌های یک تصویر و به مرکز آن‌ها تکه انتخاب کرد و به شبکه جهت ارزیابی داد. با این وجود حتی اگر دقت شبکه خیلی بالا هم باشد، به علت زیاد بودن تعداد تکه‌های انتخاب شده، مثبت‌های کاذب و منفی‌های کاذب بسیار زیادی خواهیم داشت. بنابراین در این روش ما برچسب‌زدن را به دو روش انجام خواهیم داد. در ابتدا نواحی کاندید در تصویر فوندوس را با استفاده از آموزش به طریق توجه به وجود میکروآنوریسم در تکه پیدا خواهیم کرد و پس از آن در نواحی کاندید به روش توجه به پیکسل مرکز به نواحی اصلی و دقیق میکروآنوریسم دست پیدا خواهیم کرد. برچسب زدن با توجه به وجود میکروآنوریسم در تکه (مبتتنی بر ناحیه) همان‌طور که گفته شد این روش برچسب زدن به منظور انتخاب کردن نواحی کاندید انجام شده است. به این صورت که در ابتدا سعی می‌کنیم نواحی دارای میکروآنوریسم را مشخص کنیم و پس از آن در نواحی کاندید جستجو کنیم و پیکسل‌های دقیق را پیدا کنیم. برای پیدا کردن نواحی کاندید باید روش برچسب زدن را با توجه به

میکروآنوریسم در تکه انجام دهیم. به این صورت که اگر میکروآنوریسم در داخل تکه قرار داشت برچسب بیمار و اگر وجود نداشت برچسب سالم را به آن تکه می‌دهیم. برخی از پژوهش‌ها مانند آنچه توسط چادزیک و همکاران [۳۴] بیان شده است از این روش استفاده کردند. اما یک مشکل در این روش وجود دارد. اگر میکروآنوریسم فقط در یک یا چند پیسکل حاشیه تکه وجود داشته باشد، تشخیص آن نه تنها برای شبکه بلکه برای انسان نیز سخت خواهد بود. برای رفع این مشکل تدبیری در این کار در نظر گرفته شده است. در این روش به تکه برچسب بیمار را می‌دهیم اگر و فقط اگر عارضه میکروآنوریسم به طور کامل درون پچ قرار گرفته باشد. همچنین ایجاد یک فاصله چند پیکسلی از اطراف تکه نیز می‌تواند به تشخیص آسان‌تر میکروآنوریسم کمک کند. به این صورت که اگر به طور مثال اندازه تکه ۱۰۱ پیکسل باشد، تکه‌ای برچسب بیمار را دریافت می‌کند که میکروآنوریسم به طور کامل در داخل تکه ۸۰ در ۸۰ پیکسلی داخل تکه اصلی قرار گرفته باشد. برچسب زدن با توجه به پیکسل مرکز (مبتنی بر پیکسل مرکز) پس از استخراج نواحی کاندید، بر آنیم که از بین نواحی محتمل، نواحی واقعی میکروآنوریسم را به کمک روش پیکسلی مشخص کنیم. بنابراین نیازمند روش دیگری برای برچسب زدن تکه‌ها هستم. برای این منظور به این صورت عمل می‌کنیم که اگر پیکسل مرکز تکه یک پیکسل میکروآنوریسم بود به کل تکه برچسب بیمار و در غیر این صورت برچسب سالم را می‌دهیم. در مقاله [۳۱] افتخاری و همکاران از همین روش برچسب زدن استفاده شده است. به صورتی که در گام اول به علت زیاد بودن تکه‌های سالم، این تکه‌ها به صورت کاملاً تصادفی انتخاب شده‌اند. اما در این کار ما روشی دیگر را برای انتخاب تکه در نظر گرفته‌ایم. برای انتخاب تکه‌های سالم دو مرحله در نظر گرفته‌ایم. در ابتدا به صورت تصادفی تکه‌های کاملاً سالم را انتخاب می‌کنیم. یعنی تکه‌هایی که اصلاً اثری از میکروآنوریسم در آن‌ها وجود ندارد. این کار باعث می‌شود که یادگیری شبکه برای تشخیص تکه سالم و بیمار ساده‌تر شود. حال در مرحله دوم تکه‌های نیمه سالم را انتخاب می‌کنیم. این تکه‌ها، تکه‌هایی هستند که میکروآنوریسم در آن‌ها وجود دارد اما پیکسل مرکز را شامل نمی‌شود. دلیل نام‌گذاری نیز همین موضوع است که تکه نیمه سالم است، زیرا طبق روش برچسبزنی پیکسلی، مرکز شامل میکروآنوریسم نمی‌باشد و از طرفی اثرهایی از میکروآنوریسم در تکه دیده می‌شود. توجه به این نکته در انتخاب تکه الزاماً است که در انتخاب تکه نیمه سالم باید دقیق کرد که میکروآنوریسم زیاد به پیکسل مرکز نزدیک نشود تا کار یادگیری برای شبکه سخت نشود. انتخاب یک حاشیه امن برای پیکسل مرکز به اندازه بین ۱۰ تا ۱۵ پیکسل مناسب است که میکروآنوریسم از این فاصله به پیکسل مرکز نزدیک‌تر نشود. دلیل انتخاب تکه‌های نیمه سالم این است که شبکه فقط به وجود یا عدم وجود میکروآنوریسم توجه نکند. بلکه وجود میکروآنوریسم در مرکز تکه را در دستور کار قرار دهد. بنابراین وجود این تکه‌ها در داده‌های آموزشی باعث می‌شود که شبکه مراز را بهتر یاد بگیرد. در شکل ۳ نمونه‌ای از تکه بیمار، نیمه سالم و سالم مشاهده می‌کنید.

شکل ۳-۵- انواع تکه‌ها در روش برچسب زدن با توجه به پیکسل مرکز انتخاب نوع شبکه و اندازه مناسب تکه حال نوبت



سالم بیشتر از نواحی نزدیک به تکه‌های بیمار انتخاب شوند. در این صورت به علت شباهت نسبی پس‌زمینه بین تکه سالم و تکه بیمار مجاور، تمرکز شبکه به یادگیری میکروآنوریسم بیشتر جلب خواهد شد. در انتهای، مجموعه داده ما شامل حدود ۴۶۰۰۰ تکه سالم و بیمار آماده برای آموزش شبکه ۵۰-۵۰۰۰۰۰ می‌شود. در آموزش شبکه لایه‌های استخراج بردار ویژگی و دسته‌بندی نهایی به صورت سفارشی مناسب با مسئله خودمان طراحی کرده‌ایم و برای آموزش لایه‌های ابتدایی که تقریباً برای تمام مسائل یکسان هستند را ثابت در نظر می‌گیریم. زیرا این لایه‌ها کار تشخیص خطوط، زوایا و سایر ویژگی‌های یکسان در تمام مجموعه داده‌ها را انجام می‌دهند و نیاز به آموزش مجدد نیست و در انتهای لایه‌های سفارشی خود را اضافه می‌کنیم و آموزش می‌دهیم که به این کار یادگیری انتقال می‌گویند. شکل ۳ نمایی از شبکه عصبی مورد استفاده و لایه‌های سفارشی در انتهای ساختار را نشان می‌دهد.

شکل ۳-۶- یادگیری انتقال در آموزش شبکه عصبی پیچشی مبتنی بر ناحیه

پس از آموزش شبکه در ۳۰ دوره بر روی داده‌های آموزشی، می‌توان از این شبکه برای استخراج پیکسل‌های کاندید استفاده کرد. روند کار برای استخراج پیکسل‌های کاندید به این صورت است که ابتدا بر روی تصویر اصلی با یک تعداد گام کوتاه ۵ پیکسل شروع به حرکت می‌کنیم و تکه‌های ۳۳ در ۳۳ پیکسل استخراج می‌کنیم. تکه‌های استخراج شده را جهت ارزیابی به شبکه می‌دهیم و احتمال بیمار بودن را به تک تک پیکسل‌های داخل آن تکه تخصیص می‌دهیم. با این روش به جز پیکسل‌های مرزی، سایر پیکسل‌ها در چندین تکه حضور دارند. بنابراین دو ماتریس به اندازه طول و عرض تصویر فوندووس تشکیل می‌دهیم. یکی از این ماتریس‌ها مشخص می‌کند که هر پیکسل در چند تکه حضور داشته است. (图 1-3) و دیگری مجموع احتمالات هر پیکسل که در تکه‌های ارزیابی حضور داشته است را در خود نگهداری می‌کند. (图 1-4) در نتیجه میانگین احتمال بیمار بودن هر پیکسل از طریق رابطه (1-۳) به دست می‌آید.

به تک تک پیکسل های آن تکه نسبت می دهنند. ما در این کار به روشی دیگر عمل کردیم. باید این نکته را مدنظر قرار داد که هر چه به پیکسل های حاشیه تکه نزدیک می شویم، تاثیر آن پیکسل در نتیجه احتمال به دست آمده کمتر می شود. به صورتی که می توان گفت پیکسل های مرزی احتمال صفر دارند. بنابراین احتمال به دست آمده را به طور یکسان به تمام پیکسل ها نسبت نمی دهیم. بلکه قبل از این کار از توزیع نرمال استفاده می کنیم. این توزیع نرمال ماتریسی دو بعدی هم اندازه با تکه یعنی  $33 \times 33$  پیکسل است و با نام  $\Sigma_{\text{مشخص}}^{\text{مشخص}}$  مشخص شده است که در شکل ۷۳ مشاهده می کنید. پس برای هر پیکسل، وزن معادل آن مشخص شده است. احتمال به دست آمده از شبکه را در وزن مربوطه هر پیکسل ضرب می کنیم تا احتمال بیمار بودن آن پیکسل در تکه به دست آید. احتمال حاصل را به درایه جدول  $\Sigma_{\text{مشخص}}^{\text{مشخص}}$  متناظر با آن پیکسل اضافه می کنیم و به این صورت این جدول تکمیل خواهد

شد.

### شکل ۳-۷- ماتریس وزن‌ها معادل توزیع نرمال مربوط به تکه‌های ۳۳ در ۳۳ پیکسل

روش معمول مقالات که به صورت ضمنی برای پر کردن جدول استفاده می‌کنند به این صورت است که پس از بررسی هر تکه، برای هر پیکسلی که در تکه حضور دارد، یک واحد به درایه متناظر آن پیکسل در جدول اضافه می‌کنند. اما در این کار به جای یک واحد، وزن آن پیکسل در تکه مورد بررسی را به درایه متناظر در جدول اضافه می‌کنیم. به عنوان مثال اگر در یک تکه پیکسلی در سطر ۱۶ و ستون ۱۵ باشد به درایه آن پیکسل ۹۳۷۵.۰ واحد اضافه می‌کنیم. به طور خلاصه برای پیکسلی که در سطر  $i$  و ستون  $j$  از تصویر اصلی قرار دارد و در داخل تکه ۳۳ در ۳۳ در سطر  $i$  و ستون  $j$  قرار دارد مجموع احتمالات و تعداد (تعداد وزن‌دار) از رابطه (۲-۳) و (۳-۳) به دست می‌آید.

$$= \frac{1}{\text{تکه}} \times \frac{\text{تکه}}{\text{تکه}} + \frac{\text{تکه}}{\text{تکه}} = \frac{\text{تکه}}{\text{تکه}} \quad (2-3)$$
$$\text{تکه} + \text{تکه}$$

پس از استخراج تمام تکه‌ها و دادن آن‌ها به شبکه اول و محاسبه احتمال میانگین برای هر پیکسل، نوبت به مشخص کردن پیکسل‌های کاندید می‌رسد. پیکسل‌هایی که دارای احتمال بالای ۷۰ هستند را به عنوان کاندید در نظر می‌گیریم. در نگاه اول شاید عدد ۷۰ زیاد به نظر آید. زیرا بهتر است که شبکه اول مثبت‌های کاذب بیشتر داشته باشد تا شبکه دوم میکروآنوریسم‌ها را از مثبت‌های کاذب تشخیص دهد. اما با انجام آزمایش‌های مختلف این عدد انتخاب شده است و مثبت‌های کاذب به اندازه کافی تولید می‌کند. چون شبکه به صورتی آموزش دیده است که برای میکروآنوریسم‌های تشخیصی خود احتمال بالای ۹۰ در نظر می‌گیرد. پس عدد ۷۰ شاهد مثبت‌های کاذب زیادی خواهد بود. شناسایی پیکسل‌های میکروآنوریسم از میان پیکسل‌های کاندید حال نوبت آن است که از میان پیکسل‌های کاندید انتخاب شده از مرحله قبل، پیکسل‌های میکروآنوریسم را تشخیص دهیم. برای این منظور به مرکز پیکسل‌های کاندید تکه‌های ۱۰۱ در ۱۰۱ پیکسل تشکیل می‌دهیم و این تکه را به شبکه عصبی آموزش دیده مبتنی بر پیکسل مرکز می‌دهیم تا احتمال بیمار بودن پیکسل مرکز را برای ما مشخص کند. آموزش شبکه جهت استخراج ویژگی مطابق آنچه که پیش‌تر بیان شد به آموزش این شبکه می‌پردازیم. مشابه مرحله قبل از شبکه ۵۰-۵۰ برای آموزش استفاده می‌کنیم و ساختار آن مشابه آن چیزی است که در شکل ۳-۶ مشاهده می‌کنید. مجموعه داده خود را پس از پیش‌پردازش و افزایش داده‌ها، آماده می‌کنیم. تعداد کل داده‌های آموزش برای این مرحله را به حدود ۳۲۰۰۰۰ رساندیم و آموزش را در ۳۰ دوره انجام داده‌ایم. افزودن ویژگی‌های دست‌ساز به بردار ویژگی شبکه پس از آموزش شبکه، می‌توان از این شبکه برای دسته‌بندی استفاده کرد. اما به منظور بهبود ویژگی‌های استخراج شده از این شبکه، تعدادی ویژگی دست‌ساز مهمی را که در [۲۸] ذکر شده است به بردار ویژگی شبکه

اضافه می‌کنیم. زیرا کیفیت ویژگی‌های استخراج شده به طور مستقیم بر روی دقت تشخیص میکروآنوریسم اثر می‌گذارد. ویژگی‌ها را از کاتال سیز و آبی به طور مساوی استخراج می‌کنیم تا مکمل یکدیگر باشند. این ویژگی‌ها که در مقاله مذکور بیان شده‌اند به شرح زیر هستند: ویژگی‌های مبتنی بر نمایه برش تقاطعی نمایه مجموعه‌ای از مقادیر پیکسل‌ها است که از نقاط با فاصله منظم در راستای یک بخش به دست می‌آید و برش تقاطعی مقادیر پیکسل‌ها در امتداد یک بخش، نمایه برش تقاطعی نامیده می‌شود. همان‌طور که در شکل ۳ مشاهده می‌شود ععمولاً نمایه برش تقاطعی میکروآنوریسم‌ها در هر جهت توزیع گاوی مشخص و مشابهی را نشان می‌دهد و در مقابل توزیع شدت نمایه‌های مختلف کاندیدهای اشتباه مثل عروق خونی در جهت‌های مختلف متفاوت هستند. با توجه به این مشخصه مجموعه‌ای از ویژگی‌های آماری از نمایه برش تقاطعی در جهت‌های مختلف قابل استخراج است که برای تشخیص میکروآنوریسم بسیار مفید هستند. زاویه بین خطوط برش تقاطعی، شش درجه در نظر گرفته شده است که بیانگر ۳۰ خط برش تقاطعی است و طول خطوط ۳۱ در نظر گرفته شده است. هر نمایه برش تقاطعی به صورت یک بردار با نام ۱۰ در نظر گرفته شده است و میانگین نمایه برش تقاطعی با یک بردار با نام ۱۰ مشخص می‌شود که هر درایه آن میانگین شدت نمایه مربوطه را نشان می‌دهد. علاوه بر این، یک منحنی گاوی بر روی ۱۰ برازش می‌کنیم که با ۱۰ نمایش می‌دهیم. حال ویژگی‌های زیر را استخراج می‌کنیم:

(الف) اختلاف بین میانگین و کمینه بردار میانگین نمایه برش تقاطعی: (۱۰) ب) میانگین ضرایب همبستگی بین تمام نمایه‌های برش تقاطعی ۱۰ و میانگین نمایه برش تقاطعی. (۱۰) ج) انحراف معیار ضرایب همبستگی بین تمام نمایه‌های برش تقاطعی ۱۰ و میانگین نمایه برش تقاطعی. (۱۰) د) میانگین اختلافهای عادی شده بین تمام نمایه‌های برش تقاطعی ۱۰ و میانگین نمایه برش تقاطعی. (۱۰) ه) اختلاف عادی شده به صورت رابطه (۵-۳) محاسبه می‌شود:

که ۱۰ طول هر نمایه را نشان می‌دهد. ه) انحراف معیار از تمام اختلافهای عادی شده

شکل ۳-۸- نمونه‌ای از خطوط برش تقاطعی و نمایه‌های برش تقاطعی از کاندیدهای میکروآنوریسم بازچاپ از [۵۵]. ۱۰ میکروآنوریسم صحیح و ۱۰ و ۱۰ کاندیدهای اشتباه هستند.

ویژگی‌های مبتنی بر تبدیل برش تقاطعی در مقاله [۲۸] برای تقویت ویژگی‌های کاندیدهای میکروآنوریسم، یک تصویر محلی تبدیل برش تقاطعی محلی (۱۰) تولید شده است و ما از این ویژگی‌ها نیز برای تقویت بردار ویژگی خود استفاده می‌کنیم. برای هر پیکسل کاندید یک تصویر محلی ۳۱ در ۳۱ پیکسلی به مرکز آن پیکسل استخراج می‌کنیم و ۱۰ نمایه برش تقاطعی استخراج می‌کنیم و این بردارها را در یک ماتریس قرار می‌دهیم که هر ستون این ماتریس بیانگر یک نمایه برش تقاطعی است. به این ترتیب یک تصویر ۳۱ در ۳۰ پیکسلی تبدیل برش تقاطعی تولید شده است. شکل ۳ ۹ یک نمونه ۱۰ از میکروآنوریسم واقعی

و دو کاندید اشتباه را نشان می‌دهد. همان‌طور که در شکل می‌بینید، در تصویر اصلی پیش‌پردازش شده تفاوت بین میکروآنوریسم واقعی و کاندیدهای اشتباه از ویژگی‌های بافت‌نگار گرادیان‌های جهت‌دار (( $\perp$ ) حاصل نمی‌شود. در حالی که تصویر  $\perp$  به طور قابل توجهی تفاوت ویژگی‌های  $\perp$  را بین میکروآنوریسم‌های واقعی و اشتباه را افزایش می‌دهد. زیرا نسبت میکروآنوریسم در تصویر  $\perp$  حدود چهار برابر میزان میکروآنوریسم در تصویر اصلی آن است و در این تصویر در میان میکروآنوریسم‌های واقعی گرادیان افقی به صورت قابل ملاحظه‌ای افزایش یافته است و تغییرات گرادیان عمودی حساس‌تر شده است.

شکل ۳-۹- تجسم ویژگی  $\perp$  از تصویر پیش‌پردازش شده در مقابل تصویر تبدیل برش تقاطعی بازچاپ از [۵۵].

میکروآنوریسم صحیح و  $\perp$  و  $\square$  کاندیدهای اشتباه هستند.

با در نظر گرفتن کل تصویر محلی به عنوان سلول ویژگی‌های زیر را استخراج می‌کنیم. الف):  $\square$  استخراج ویژگی‌های  $\perp$  از تصویر پیش‌پردازش شده محلی. ابتدا اندازه و جهت گرادیان هر پیکسل را محاسبه می‌کنیم. سپس آمار بافت‌نگار  $\perp$  تصویر را در ۹ جهت محاسبه می‌کنیم. بنابراین ویژگی‌های  $\perp$  را در یک بردار نه بعدی استخراج کردہ‌ایم. ب):  $\square$  استخراج همان ویژگی‌های  $\perp$  بر روی تصویر  $\perp$  آموزش  $\perp$  برای دسته‌بندی کاندیدها تا اینجا توانسته‌ایم سه نوع ویژگی را استخراج کنیم. نوع اول، ویژگی‌هایی است که از طریق آموزش شبکه عصبی پیچشی به دست می‌آید. نوع دوم ویژگی‌های مبتنی بر نمایه  $\perp$  برش تقاطعی و نوع سوم ویژگی‌های مبتنی تبدیل برش تقاطعی است. حال از تمام ویژگی‌های منتخب مطابق شکل ۱۰ ۳  $\perp$  برای آموزش یک ماشین بردار پشتیبان (SVM) استفاده می‌کنیم و در مرحله آزمون، ویژگی‌های عنوان شده را استخراج کرده و به کمک یک ماشین بردار پشتیبان سالم و یا بیمار (دارای میکروآنوریسم) بودن آن تکه را مشخص می‌کنیم.

شکل ۳-۱۰- دسته‌بندی ویژگی‌های استخراج شده از تکه به کمک ماشین بردار پشتیبان

دسته‌بندی در مرحله آزمون به منظور بررسی یک تصویر فوندوس، پس از پیش‌پردازش تصویر را مطابق با ورودی شبکه اول تکه می‌کنیم و جهت استخراج میکروآنوریسم‌های کاندید به آن شبکه می‌دهیم. پس از استخراج پیکسل‌های کاندید، به مرکز این پیکسل‌ها تکه‌ای به اندازه ورودی شبکه دوم می‌زنیم تا این شبکه ویژگی‌های لازم را استخراج کند. همچنین ویژگی‌های دست‌ساز را هم از این کاندیدها استخراج کرده و جهت دسته‌بندی به ماشین بردار پشتیبان می‌دهیم. در شکل ۱۱ ۳ یک تصویر فوندوس را مشاهده می‌کنید که در تصویر سمت چپ مکان اصلی میکروآنوریسم‌ها را نشان می‌دهد و در سمت راست نتیجه الگوریتم به منظور شناسایی میکروآنوریسم‌ها را می‌بینید.

شکل ۳-۱۱- مکان میکروآنوریسم‌های اصلی و میکروآنوریسم‌های شناسایی شده توسط روش پیشنهادی در یک تصویر

فوندوس

فصل ۵

## ارزیابی روش پیشنهادی

مقدمه در این کار سعی بر این داشته‌ایم تا به کمک روش‌های یادگیری ژرف در راستای شناسایی میکروآنوریسم‌ها در تصاویر فوندوس قدم برداریم و در حوزه پزشکی عملکرد هوش مصنوعی و یادگیری ژرف را بهبود دهیم. بنابراین با تحقیق و آزمایش روشی برای شناسایی میکروآنوریسم در تصاویر فوندوس پیشنهاد داده‌ایم که شرح آن در فصل سوم انجام شد و نوبت آن است که الگوریتم پیشنهادی را با نمونه‌های مشابه مقایسه کنیم. پیاده‌سازی این الگوریتم به کمک زبان برنامه نویسی پایتون و کتابخانه `OpenCV` انجام شده است.<sup>[۳۵]</sup>. محیط برنامه نویسی این کار بر روی سرویس ابری `Amazon AWS` می‌باشد که حدود ۲۵ گیگابایت حافظه و حدود ۷۰ گیگابایت فضای دیسک در اختیار گرفته‌ایم. از کتابخانه‌های مهم مورد استفاده در این کار می‌توان به کتابخانه `OpenCV` و `TensorFlow`<sup>[۳۶]</sup> با هدف آموزش و ارزیابی مدل‌های یادگیری ژرف، `TensorBoard`<sup>[۳۷]</sup> به منظور پردازش تصاویر، `Keras`<sup>[۳۸]</sup> برای انجام محاسبات ریاضی و ماتریسی و کار با آرایه‌ها اشاره کرد. پس از بیان معیارهای ارزیابی، به بیان نتایج و مقایسه با کارهای دیگر می‌پردازیم. معیارهای ارزیابی مفاهیم اولیه قبل از پرداختن به معیارهای ارزیابی نمادها و مفاهیم اولیه را بیان می‌کنیم: مثبت‌های صحیح: (+) نمونه‌های مثبت (بیمار) که به درستی بیمار تشخیص داده شده‌اند. مثبت‌های کاذب: (-) نمونه‌های منفی (سالم) که به اشتباه بیمار تشخیص داده شده‌اند. منفی‌های صحیح: (-) نمونه‌های منفی (سالم) که به درستی سالم تشخیص داده شده‌اند. منفی‌های کاذب: (-) نمونه‌های مثبت (بیمار) که به اشتباه سالم تشخیص داده شده‌اند. دقت یکی از معیارهایی که بسیار در پزشکی و سایر زمینه‌های تحقیقاتی حائز اهمیت است، معیار دقت می‌باشد. در این کار ما با تصاویر دارای میکروآنوریسم و تصاویر سالم سر و کل داریم. بنابراین معیار دقت در این کار به معنی درصد نمونه‌هایی است که بیمار و یا سالم بودن آن‌ها به درستی تشخیص داده شده است. فرمول این معیار در رابطه (۱-۴) بیان شده است.

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

که  $N$  تعداد کل نمونه‌ها را نشان می‌دهد. حساسیت در حوزه پزشکی و تشخیص ضایعه‌ها، عارضه‌ها و بیماری‌ها، معیار حساسیت بسیار اهمیت ویژه و بالاتری نسبت به دقت دارد. به این معنی که هر چقدر درصد شناسایی عارضه یا بیماری از بین نمونه‌های آزمون بیمار بیشتر باشد الگوریتم و روش ما ارزشمندتر خواهد بود. حساسیت نیز بیانگر همین موضوع است و درصد تشخیص صحیح نمونه‌های بیمار را نشان می‌دهد. دلیل ارزشمند بودن این معیار این است که نمونه‌های بیمار باید حتماً به درستی تشخیص داده شوند و اگر نمونه بیماری به اشتباه سالم تشخیص داده شود می‌تواند تبعات جیران ناپذیری برای بیمار به همراه داشته باشد. رابطه (۲-۴) رابطه حساسیت را نشان می‌دهد.

ویژگی در مقابل معیار حساسیت، معیار تحت عنوان ویژگی وجود دارد. این ویژگی درصد نمونه‌های سالمی که به درستی سالم تشخیص داده شده‌اند را نشان می‌دهد. فرمول این معیار را در رابطه (۴-۳) عنوان کرده‌ایم.

با توجه به سه معیاری که ذکر شد مشخص می‌شود که معیار حساسیت از اهمیت بیشتری نسبت به دو معیار دیگر برخوردار است. اما اگر سیستمی را در نظر بگیریم که تمام نمونه‌ها را بیمار تشخیص دهد، این سیستم حساسیت ۱۰۰٪ به همین منظور به دنبال معیاری خواهیم بود که معیارهای حساسیت و ویژگی را به هم در کنار یکدیگر داشته باشد. برای این کار مقدار یکی از عامل‌ها، نظیر آستانه تشخیص بیماری را آن قدر عوض می‌کنیم تا میانگین تعداد مثبت‌های کاذب در هر تصویر آزمون به مقدار یک، دو، چهار و هشت برسد. حال در این مقدار مشخص میزان حساسیت به دست آمده را محاسبه می‌کنیم. سپس به کمک تخمین، میزان حساسیت را به ازای مقدار انتزاعی یک دوم، یک چهارم و یک هشتم محاسبه می‌کنیم و بین حساسیت‌های به دست آمده میانگین می‌گیریم. مقدار به دست آمده معیار (۴-۳) را مشخص می‌کند که میانگین حساسیت را به ازای مقدار مشخص مثبت‌های کاذب در هر تصویر نشان می‌دهد. این معیار مثبت‌های کاذب را در خود جای داده است و حساسیت را در شرایط خاص یکسان بررسی می‌کند. فرمول محاسبه این معیار را به صورت رابطه (۴-۴) ذکر کرده‌ایم.

که در این رابطه بیان کننده حساسیت در حالتی است که تعداد مثبت‌های کاذب برابر با مقدار ۰ باشد. ۰ تعداد کل آستانه‌هایی است که به ازای آن مقدار حساسیت را محاسبه کرده‌ایم. مجموعه داده در این کار از دو مجموعه داده برای آموزش و آزمون استفاده شده است که به شرح آن‌ها می‌پردازیم. مجموعه داده (۴-۳) این بانک شامل دو مجموعه داده زیر مجموعه با نام‌های (۴-۳)-۱-۱ برای میکروآنوریسم‌ها و همچنین مجموعه داده (۴-۳)-۱-۲ برای اگزودا است. مجموعه داده میکروآنوریسم شامل ۱۴۸ تصویر دارای میکروآنوریسم و هموریج‌های ریز و ۲۳۳ تصویر بدون عارضه است. [۳۹]. مجموعه داده (۴-۳)-۱ این مجموعه داده شامل ۸۹ تصویر است که به صورت دستی در پنج دسته قرار گرفته‌اند که این پنج دسته معرف حالتها و درجه‌های مختلف رتینوپاتی دیابتی است. ۲۷ تصویر سالم، ۷ تصویر با رتینوپاتی دیابتی خفیف، ۲۸ تصویر با رتینوپاتی دیابتی متوسط و شدید بدون نورگزایی و در نهایت ۲۷ تصویر رتینوپاتی دیابتی شدید به همراه نورگزایی است. در کل، مجموعه داده به دو دسته تصاویر آموزش شامل ۲۸ تصویر و تصاویر آزمون شامل ۶۱ تصویر تقسیم شده‌اند. در برچسب زدن به هر تصویر چهار خبره نقش داشته‌اند. [۴۰]. در این کار ما مناطقی که دارای بیش از ۷۵ درصد نظر موافق برای میکروآنوریسم وجود دارد را به عنوان مناطق اصلی میکروآنوریسم در نظر می‌گیریم.

پیکربندی الگوریتم به منظور آموزش شبکه‌ها، در هر دوره ۲۰ تابع ضرر از تابع آنتروپویی تقاطعی دسته‌ای استفاده کرده‌ایم. این تابع میزان ضرر را به صورت رابطه (۴-۵) محاسبه می‌کند.

که در این تابع  $\Delta$  اندازه خروجی شبکه (در کار ما  $\Delta$  برابر با دو است) را نشان می‌دهد. بنابراین مقدار از برچسب اصلی امین  $\Delta$  مقدار از بردار خروجی پیش‌بینی شده توسط شبکه را نشان می‌دهد. همچنین برای به دست آوردن بهترین نتیجه از آموزش، از روش ارزیابی تقاطعی استفاده کردند. به این صورت که در هر مرتبه آموزش، تعداد داده‌های آموزش را به پنج دسته تقسیم می‌کنند. چهار قسمت را برای آموزش و یک قسمت را برای آزمون در نظر می‌گیرند. به این ترتیب پنج مدل شبکه برای آموزش خواهیم داشت و بهترین نتیجه را برای آزمون بر روی داده‌های آزمون انتخاب می‌کنند. نتایج آزمون با توجه به پیکربندی بیان شده و همچنین عامل‌هایی که در فصل سوم توضیح داده شد، مدل را آموزش دادند و به سراغ آزمون داده‌های آزمون می‌روند. آستانه شبکه عصبی پیچشی دوم که ویژگی‌های مبتنی بر پیکسل را برای دسته‌بندی نهایی استخراج می‌کند آن قدر تغییر دادند. آستانه شبکه عصبی مثبت‌های کاذب در هر تصویر آزمون برابر با مقادیر یک، دو، چهار و هشت شود. حال در هر یک از چهار آستانه به دست آمده، به طور مجزا میانگین حساسیت بر روی تصاویر آزمون به دست آورده‌اند و سپس با تخمین، میانگین حساسیت را در شرایطی که تعداد مثبت‌های کاذب برای هشت، چهار و یک دوم باشد حساب کردند. در جدول ۱ و جدول ۲ نتایج حاصل از محاسبه حساسیت را در مقادیر مختلف مثبت‌های کاذب در دو مجموعه داده  $\text{I}_{\text{train}}-\text{I}_{\text{test}}$  و  $\text{II}_{\text{train}}-\text{II}_{\text{test}}$  مشاهده می‌کنند.

نمودار ۴-۱ که مقدار حساسیت را به ازای مقادیر مختلف مثبت‌های کاذب نشان می‌دهد، برای مجموعه داده‌های  $\text{I}_{\text{train}}$  در شکل ۴ و شکل ۱ در شکل ۴ و شکل ۲ مشاهده می‌کنید.

شکل ۴-۱- منحنی  $\text{I}_{\text{train}}$  مقالات در مقایسه با روش پیشنهادی برای مجموعه داده  $\text{I}_{\text{train}}-\text{I}_{\text{test}}$

شکل ۴-۲- منحنی  $\text{I}_{\text{train}}$  مقالات در مقایسه با روش پیشنهادی برای مجموعه داده  $\text{I}_{\text{train}}-\text{I}_{\text{test}}$

با توجه به نتایجی که در جدول‌ها و شکل‌ها مشاهده می‌کنید متوجه می‌شویم که هر چند حساسیت روش پیشنهادی ما در تعدادی از مقادیر مثبت کاذب نسبت به موارد مشابه در دیگر مقاله‌ها کمتر شده است اما در کل معیار  $\text{I}_{\text{train}}-\text{I}_{\text{test}}$  آن نسبت به بقیه بهتر شده است و این نشان می‌دهد که ترکیب ویژگی‌های ژرف (استخراج شده توسط انسان) و ویژگی‌های معنادار (استخراج شده توسط انسان) باعث می‌شود که سامانه مزیت هر دو دسته ویژگی را با هم داشته باشد و در به دست آوردن نتایج بهتر کمک کننده باشد.

## ٦ فصل

### نتیجه‌گیری و پیشنهادات

۱-۵- مقدمه هدف کار ارائه شده در این پژوهش ارائه و بهبود روش‌های خودکار مبتنی بر هوش مصنوعی و یادگیری ژرف به منظور

کمک به جامعه پزشکی در تشخیص زودهنگام رتینوپاتی دیابتی است و همچنان راه حل‌ها و ایده‌های دیگری می‌تواند در جهت

پیشرفت تشخیص بیماری از جمله رتینوپاتی دیابتی کمک کننده باشند. ۲-۵- بحث و نتیجه‌گیری روش‌های مختلفی برای شناسایی

زودهنگام رتینوپاتی دیابتی و شناسایی میکروآنوریسم ارائه شده‌اند. دسته‌ای از این روش‌ها سعی کردۀ‌اند تا به کمک روش‌های غیر

ژرف و استخراج ویژگی‌های معنادار بین پیکسل‌های میکروآنوریسم و غیرمیکروآنوریسم تفاوت قائل شوند و در نهایت یک دسته‌بند

را با ویژگی‌های ارائه شده آموزش دهند و از این دسته‌بند برای آزمون استفاده کنند. در مقابل دسته‌ای دیگر از روش‌ها وجود

دارند که با تکیه بر ویژگی‌های ژرف استخراج شده از شبکه‌های عصبی پیچشی سعی دارند پیکسل‌های میکروآنوریسم را شناسایی

کنند. به این صورت که تصاویر آموزشی را به شبکه می‌دهند و شبکه در مرحله آموزش سعی می‌کند ویژگی‌های لازم و تفکیک‌پذیر

بین میکروآنوریسم‌ها و غیرمیکروآنوریسم‌ها را شناسایی کند. واضح است که ویژگی‌های استخراج شده توسط شبکه با ویژگی‌های

معنادار استخراج شده توسط انسان همیشه برابر نیستند و حتی ممکن است ویژگی‌های شبکه اصلاً قابل تفسیر نباشند. پس در این

جا با دو دسته از ویژگی‌ها رو به رو هستیم. در این پژوهش سعی شده است که از هر دو دسته ویژگی استفاده شود. بنابراین ما

ویژگی‌های معنادار مطلوب را در کنار ویژگی‌های استخراج شده توسط شبکه قرار داده‌ایم تا میزان دقت و حساسیت شبکه را بالا

بریم. ۳-۵- پیشنهادات به نظر می‌رسد استفاده از ویژگی‌های معنادار در کنار ویژگی‌های ژرف در تشخیص میکروآنوریسم بسیار بهتر

عمل می‌کند. به منظور بهبود در نتایج باید در انتخاب ویژگی‌ها از هر دو دسته دقت بیشتری داشت. بنابراین برای بهبود ویژگی‌های

ژرف، استفاده از ویژگی‌های چند مدل شبکه در کنار هم می‌تواند کمک کننده باشد. به این صورت که هر مدل ویژگی‌های مختلف را

استخراج می‌کند و از کنار هم قرار دادن چند دسته ویژگی ژرف، بردار ویژگی ما غنی‌تر خواهد شد. همچنین می‌توان بهترین اندازه

تکه را برای هر یک از مدل‌ها به دست آورد و به این ترتیب یک شبکه ترکیبی چند مقیاسی به دست خواهد آمد که از چند مدل

قدرتمند و با اندازه تکه‌های مختلف برای تشخیص میکروآنوریسم استفاده می‌کند. از طرفی در این میان ممکن است ویژگی‌هایی

یکسان از شبکه‌ها استخراج شود. و یا این که ویژگی‌هایی وجود داشته باشند که در تشخیص میکروآنوریسم زیاد نقش موثری نداشته

باشند. به این ترتیب استفاده از روش‌های کاهش بردار ویژگی برای استخراج ویژگی‌های اصلی برای ایجاد مرز بین پیکسل‌های

میکروآنوریسم و غیر میکروآنوریسم می‌توان مفید باشد. در مقابل با تحقیق می‌توان ویژگی‌های متمایز کننده معنادار دیگری برای

پیکسل‌های میکروآنوریسم به دست آورد تا به بردار ویژگی اضافه شوند. روش‌های مختلفی برای پیش‌پردازش تصاویر فوندوس وجود

دارند. استفاده از روش‌های جدید پیش‌پردازش به منظور کاهش بیشتر نویز، یکسان سازی روشنایی تصاویر بافت‌نگار و ... می‌تواند به

استخراج ویژگی‌های مفیدتر کمک کند تا بتوان نتایج بهتر برای تشخیص زودهنگام رتینوپاتی دیابتی و میکروآنوریسم به دست آورد.

## منابع و مأخذ

- [1] A. Danelakis, T. Theoharis, and D. A. Verganelakis, “Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging,” *Computerized Medical Imaging and Graphics*, vol.70, pp.83–100, 2018.
- [٢] حمید ایران پور، ”زیست.“ ۲۰۱۵ <http://zistshenasi94.blogfa.com/post/4>
- [3] R. Milo and E. Kahana, “Multiple sclerosis: geoepidemiology, genetics and the environment,” *Autoimmunity reviews*, vol.9, no.5, pp.A387–A394, 2010.
- [4] R. D. Fields, “White matter matters,” *Scientific American*, vol.298, no.3, pp.54–61, 2008.
- [5] X. Lladó, A. Oliver, M. Cabezas, J. Freixenet, J. C. Vilanova, A. Quiles, L. Valls, L. Ramió-Torrentà, and À. Rovira, “Segmentation of multiple sclerosis lesions in brain mri: a review of automated approaches,” *Information Sciences*, vol.186, no.1, pp.164–185, 2012.
- [6] M. Cabezas Grebol *et al.*, “Atlas-based segmentation of multiple sclerosis lesions in magnetic resonance imaging,” 2013.
- [7] I. Despotović, B. Goossens, and W. Philips, “Mri segmentation of the human brain: challenges, methods, and applications,” *Computational and mathematical methods in medicine*, vol.2015, 2015.
- [8] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, “A nonparametric method for automatic correction of intensity nonuniformity in mri data,” *IEEE transactions on medical imaging*, vol.17, no.1, pp.87–97, 1998.
- [9] M. A. Balafar, “Review of intensity inhomogeneity correction methods for brain mri images,” *Issue*, vol.13, pp.60–66, 2012.
- [10] A. Birenbaum and H. Greenspan, “Multi-view longitudinal cnn for multiple sclerosis lesion segmentation,” *Engineering Applications of Artificial Intelligence*, vol.65, pp.111–118, 2017.
- [11] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, “Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation,” *IEEE transactions on medical imaging*, vol.35, no.5, pp.1229–1239, 2016.

- [12] G. Kang, B. Hou, Y. Ma, F. Labeau, Z. Su, *et al.*, “Acu-net: A 3d attention context u-net for multiple sclerosis lesion segmentation,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.1384–1388, IEEE, 2020.
- [13] S. Jain, D. M. Sima, A. Ribbens, M. Cambron, A. Maertens, W. Van Hecke, J. De Mey, F. Barkhof, M. D. Steenwijk, M. Daams, *et al.*, “Automatic segmentation and volumetry of multiple sclerosis brain lesions from mr images,” *NeuroImage: Clinical*, vol.8, pp.367–375, 2015.
- [14] S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T.-H. Kim, “An effective method for computerized prediction and segmentation of multiple sclerosis lesions in brain mri,” *Computer methods and programs in biomedicine*, vol.140, pp.307–320, 2017.
- [15] A. C. da Silva Senra Filho, “A hybrid approach based on logistic classification and iterative contrast enhancement algorithm for hyperintense multiple sclerosis lesion segmentation,” *Medical & biological engineering & computing*, vol.56, no.6, pp.1063–1076, 2018.
- [16] O. Ghribi, L. Sellami, M. B. Slima, C. Mhiri, M. Dammak, and A. B. Hamida, “Multiple sclerosis exploration based on automatic mri modalities segmentation approach with advanced volumetric evaluations for essential feature extraction,” *Biomedical Signal Processing and Control*, vol.40, pp.473–487, 2018.
- [17] X. Gui, C. Chuansheng, L. Zhong-Lin, and D. Qi, “Brain imaging techniques and their applications in decision-making research,” *Xin li xue bao. Acta psychologica Sinica*, vol.42, no.1, p.120, 2010.
- [18] A. Ceccarelli, R. Bakshi, and M. Neema, “Mri in multiple sclerosis: a review of the current literature,” *Current opinion in neurology*, vol.25, no.4, pp.402–409, 2012.
- [19] I. Dekker and M. P. Wattjes, “Brain and spinal cord mr imaging features in multiple sclerosis and variants,” *Neuroimaging Clinics*, vol.27, no.2, pp.205–227, 2017.
- [20] U. Vovk, F. Pernus, and B. Likar, “A review of methods for correction of intensity inhomogeneity in mri,” *IEEE transactions on medical imaging*, vol.26, no.3, pp.405–421, 2007.
- [21] Z. Hou, “A review on mr image intensity inhomogeneity correction,” *International journal of biomedical imaging*, vol.2006, 2006.
- [22] B. M. Dawant, A. P. Zijdenbos, and R. A. Margolin, “Correction of intensity variations in mr images for computer-aided tissue classification,” *IEEE transactions on medical imaging*, vol.12, no.4, pp.770–781, 1993.
- [23] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging*, vol.29, no.6, pp.1310–1320, 2010.
- [24] C. Li, J. C. Gore, and C. Davatzikos, “Multiplicative intrinsic component optimization (mico) for mri bias field estimation and tissue segmentation,” *Magnetic resonance imaging*, vol.32, no.7, pp.913–923, 2014.

- [25] D. García-Lorenzo, S. Francis, S. Narayanan, D. L. Arnold, and D. L. Collins, “Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging,” *Medical image analysis*, vol.17, no.1, pp.1–18, 2013.
- [26] S. M. Smith, “Fast robust automated brain extraction,” *Human brain mapping*, vol.17, no.3, pp.143–155, 2002.
- [27] N. Abraham and N. M. Khan, “A novel focal tversky loss function with improved attention u-net for lesion segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp.683–687, IEEE, 2019.
- [28] S. H. Gheshlaghi, A. Madani, A. Suratgar, and F. Faraji, “Multiple sclerosis diagnosis with fuzzy cmeans,” *Computer Science & Information Technology*, p.97.
- [29] Y. Zhao, S. Guo, M. Luo, Y. Liu, M. Bilello, and C. Li, “An energy minimization method for ms lesion segmentation from t1-w and flair images,” *Magnetic resonance imaging*, vol.39, pp.1–6, 2017.
- [30] Y. Zhao, S. Guo, M. Luo, X. Shi, M. Bilello, S. Zhang, and C. Li, “A level set method for multiple sclerosis lesion segmentation,” *Magnetic resonance imaging*, vol.49, pp.94–100, 2018.
- [31] M. H. Alshayeji, M. A. Al-Rousan, H. Ellethy, and S. Abed, “An efficient multiple sclerosis segmentation and detection system using neural networks,” *Computers & Electrical Engineering*, vol.71, pp.191–205, 2018.

## **Abstract**



**Ferdowsi University Mashhad (FUM)**  
**Department of Computer**

Thesis submitted  
for the degree of M.Sc.

**Title:**  
**MRI Segmentation**

**Supervisor: DR. Hamid Reza Pour Reza**

**By: Mehdi Sadeghi Bakhi**

January 2018