

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

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

چکیده

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 تخصصی تصاویر بیشتری تولید می‌کنیم. برای این منظور ما از ۴ شبکه تولید کننده در کنار هم استفاده کرده ایم. با بررسی نتایج حاصل از این پژوهش بر روی تصاویر مجموعه داده LFW و YouTube Faces، دقت روش پیشنهادی ما به ترتیب برابر با ۹۶٪ و ۹۴٪ شد. که دقت بالاتری نسبت به روش‌های مشابه می‌باشد.

فهرست مطالب

پ	فهرست جداول
ت	فهرست تصاویر
ث	فهرست نمادها
۱	۱ مروری بر کارهای گذشته در شرایط کنترل نشده
۲	۱.۱ مقدمه
۲	۲.۱ چالش حالت
۱۱	۳.۱ چالش روشنایی
۱۳	۴.۱ چالش انسداد
۱۵	۵.۱ چالش کمبود تصاویر آموزشی
۱۶	۶.۱ چالش منابع محدود
۱۸	۷.۱ نتیجه گیری
۲۳	۸.۱ نتیجه گیری
۲۶	منابع و مأخذ

فهرست جداول

فهرست تصاویر

۱.۱	رویکردهای مختلف هم ترازی چهره [۷].	۴
۲.۱	رویکرد مبتنی بر تطبیق کلیشه [۷].	۴
۳.۱	معماری OpenFace [۷]	۵
۴.۱	جریان یادگیری در معماری OpenFace [۷].	۵
۵.۱	رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۷].	۶
۶.۱	رویکرد مبتنی بر تطبیق کلیشه [۷].	۷
۷.۱	نتیجه آزمایش بر روی مجموعه داده FERET در زاویه های متفاوت [۷].	۷
۸.۱	معماری شبکه پیشنهادی VS2VI [۷]	۸
۹.۱	(a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [۷].	۹
۱۰.۱	مقایسه روش ارائه شده با سایر روش ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش های دیگر (d) خروجی روش ارائه شده [۷].	۱۰
۱۱.۱	۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۷].	۱۰
۱۲.۱	اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۷].	۱۲
۱۳.۱	تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۷].	۱۳
۱۴.۱	رویکرد کلی الگوریتم GD-HASLR [۷]	۱۴
۱۵.۱	ردیابی، یافتن چهره ها و برچسب زنی [۷].	۱۵
۱۶.۱	تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند [۷].	۱۶

۱۷.۱ استفاده از روش خوش بندی گراف و تعیین تصویر شاخص [۷].

۱۸.۱ معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۷].

۱۹.۱ نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۷].

۲۰.۱ چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۷].

۲۱.۱ تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۷].

۲۲.۱ رویکرد مبتنی بر تطبیق کلیشه [۷].

۲۳.۱ حفظ محتوی و انتقال style از تصاویر مجموعه ۱ به تصاویر مجموعه ۲ و برعکس [۷].

۲۴.۱ ساختار شبکه CycleGAN و نحوه محاسبه خطا به صورت انتزاعی [۷].

۲۵.۱ ساختار شبکه AD GAN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۷].

فهرست نمادها

فصل ۱

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 بنده، 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	مشکل ها
حالات چهره	تبديل دو بعدی	استخراج نقاط ویژه باید دقیق تر باشد
حالات چهره	استفاده از شبکه عصبی عمیق	پیچیدگی محاسباتی قابل قبول
حالات چهره	تبديل مدل دو بعدی به سه بعدی	دقت بالا در شرایط کنترل نشده
حالات چهره	تبديل مدل سه بعدی به دو بعدی	پیچیدگی محاسباتی، واستگی به داده های آموزش
روشنایی	همسان سازی بافت-نگار	دقت بالا پیچیدگی محاسباتی
انسداد	استفاده از روش های شناسایی الگو و واپاش	قابل استفاده در تصاویر خاکستری
حدوده دید داده	دقت بالا در انسداد شدید	تھیه مجموعه داده با ردبایی چهره در ویدیو
حدوده دید منابع	استفاده از رایانش ابری	نیاز به یک مرحله طولانی استفاده از تصاویر ویدیو تجهیزات پیشرفته و زمان تاخیر ناهمگن

۲.۱ چالش حالت

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

۱. رویکردهای دو بعدی با پیچیدگی محاسباتی قابل قبول (بیشتر ایده های مبتنی بر نشانه گذاری قدیمی) مانند [۲۳-۲۶].

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

دو بعدی نیاز به توجه بیشتری دارد. پیشرفتهای اخیر مانند [۲۳] است. مزیت: این الگوریتم‌ها از نظر پیچیدگی محاسباتی

قابل قبول هستند و کاملاً برای شرایط این پروژه مناسب می‌باشند.

۲. رویکردهای مبتنی بر شبکه عصبی برای تخمین و اصلاح موقعیت چهره (آموزش و آزمایش با تصاویر دو بعدی) مانند [۲۲، ۲۷-۳۰]. مشکل: این الگوریتم‌ها، به طور متوسط، کندر از دسته پیشین می‌باشند. اما بستگی به این دارد که عمق شبکه

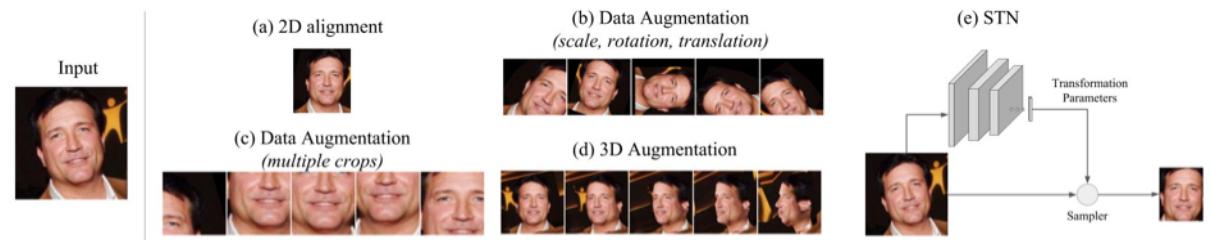
عصبی چه مقدار باشد. وابستگی آن‌ها به داده‌های آموزش می‌باشد و مراحل مجازی رو به رو سازی و هم ترازی چهره ندارند. مزیت: بدون نیاز به تصمیم‌گیری در مورد مجموعه بهینه‌ای از نشانه‌های چهره و دارای دقت بیشتر در شرایط کنترل نشده با انسداد و... . ایده‌هایی مانند [۴۲] برای تخمین موقعیت چهره ممکن است به زمان محاسبات کمک کند.

۳. رویکردهای سبک سه بعدی بدست آمده از تصاویر دو بعدی، مانند [۱۱، ۳۱-۳۴]. مشکل: زمان محاسباتی بالا. یکی از امیدوار کننده‌ترین این الگوریتم‌ها 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 می‌باشند، اما برای محدودیت‌های این پروژه قابل استفاده نمی‌باشند. مزیت: با استفاده از اطلاعات سه بعدی، این روش‌ها به بالاترین دقت تصمیم‌گیری در میان سه نفر رسید.

۴. رویکردهای تبدیل مدل سه بعدی چهره به مدل دو بعدی چهره (روش‌های مبتنی بر پنجره بر اساس چند نمایش دو بعدی مختلف از چهره) مانند [۳۵] مشکل: زمان محاسباتی (نه به اندازه الگوریتم‌های دسته سوم). مزیت: عملکرد بهتر در رو به رو سازی چهره در شرایط کنترل نشده نسبت به الگوریتم‌های دسته اول. ممکن است برای شرایط این پروژه مناسب باشند.

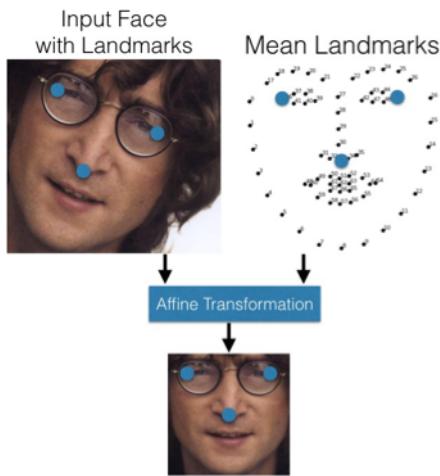
مرجع [۴۳] در سال ۲۰۱۸ خلاصه‌ای از رویکردهای مختلف برای حل مسئله هم ترازی را در شکل ۱-۳ نشان داده است. تصویر سمت چپ، چهره ورودی می‌باشد. (a) هم ترازی با استفاده از تبدیلات دو بعدی ساده می‌باشد. (b) داده افزایی با تنییر مقیاس، تغییر زاویه و جا به جایی می‌باشد. (c) برش‌های چندگانه می‌باشد. (d) داده افزایی مبتنی بر روش‌های سه بعدی می‌باشد. (e) از هیچ ابزاری برای هم ترازی مستقیم استفاده نمی‌نماید. اما یک شبکه را آموزش می‌دهد تا عامل‌های مورد نیاز برای تبدیل هم ترازی را بدست آورد.

در سال ۲۰۱۶ و همکاران در [۲۵] یک روش شناسایی چهره به نام OpenFace Brandon Amos ارائه دادند که ویژگی اصلی آن، آموزش شبکه عصبی عمیق در کمترین زمان و قابلیت اجرا بر روی دستگاه‌های قابل حمل مانند تلفن همراه با در نظر گرفتن منابع محدود می‌باشد. یک تصویر شامل تعدادی چهره به الگوریتم داده می‌شود. پس از یافتن چهره‌ها و مجزا کردن آن



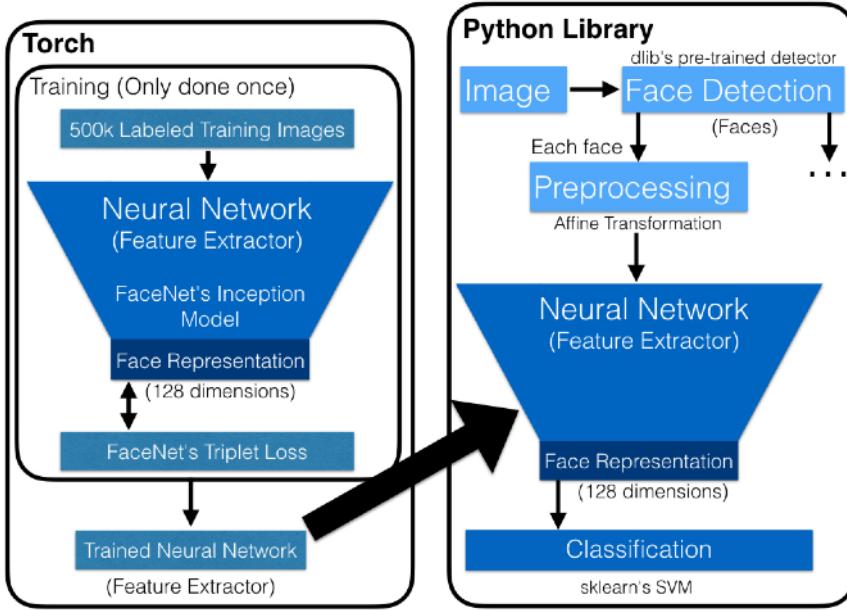
شکل ۱.۱: رویکردهای مختلف هم ترازی چهره [۷].

ها از یکدیگر، هر چهره به طور جداگانه مورد پیش پردازش قرار می گیرد و حجم آن کاهش می یابد. کاهش حجم تصویر برای عملکرد مناسب یک طبقه بندی بهینه بسیار مهم می باشد. تصاویر چهره ها باید هنجارسازی شده و ابعاد آن ها ثابت گردد تا به بخش شناسایی چهره راه یابند. هر تصویر چهره باید مورد تبدیل قرار بگیرد تا چشم ها، بینی و دهان، در مکان مشخصی قرار گیرند. بدین منظور از یک تبدیل هم نسبی دوبعدی ساده استفاده می گردد. ابتدا باید چهره توسط ۶۸ نقطه ویژه، نشانه گذاری شود. سپس نشانه های اطراف چشم ها و بینی (شکل ۲-۳) برای محاسبه عامل های تبدیل هم نسبی استفاده می شوند. پس از انجام تبدیل هم نسبی، تصاویر چهره برش زده شده و اندازه آن ها 96×96 پیکسل می شود. شکل ۲-۳ - تبدیل هم نسبی OpenFace براساس



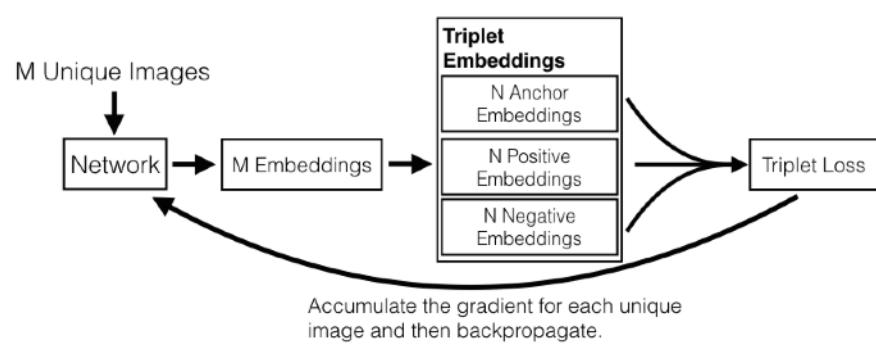
شکل ۲.۱: رویکرد مبتنی بر تطبیق کلیشه [۷].

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 تعلم شبکه از مجموعه داده کوچکی با 500 هزار تصویر چهره استفاده می کند که از ادغام دو مجموعه داده بزرگ برچسب گذاری شده به نام FaceScrub و CASIA-WebFace بدست آمده است. شبکه مورد استفاده در این الگوریتم یک نسخه اصلاح شده از شبکه nn4 الگوریتم FaceNet می باشد. شبکه nn4 مبتنی بر معماری GoogLeNet می باشد. برای تعیین میزان شباهت نتیجه، از فاصله اقلیدسی استفاده شده است. هر تصویر از یک شبکه یکتا به یک سه گانه نگاشت داده می شود.



.[۷] OpenFace معماری

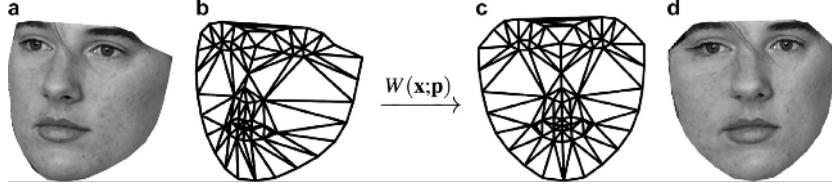
گرادیان خطای سه گانه برای هر تصویر محاسبه شده و به عقب انتشار می‌یابد. در هر دسته کوچک، P تصویر برای هر نفر از Q نفر، در مجموعه داده انتخاب می‌شود. سپس $M \approx PQ$ تصویر به شبکه داده می‌شود تا عملیات forward انجام پذیرد. در این مقاله از $N = Q \binom{P}{2}$ و $Q=15$ استفاده شده است. تمام جفت‌های anchor-positive برای بدست آوردن سه گانه‌های anchor-positive استفاده شده است. تمام جفت‌های anchor-negative برای سنجیدن میزان دقیقیت الگوریتم‌های تشخیص چهره آموزش شبکه را نشان می‌دهد. مجموعه داده LFW یک معیار استاندارد برای سنجیدن دقیقیت الگوریتم‌های تشخیص چهره



.[۷] OpenFace در معماری

می‌باشد که از ۱۳۲۳۳ تصویر چهره از ۵۷۵۰ شخص تشکیل شده است. الگوریتم OpenFace بر روی این مجموعه داده مورد سنجش قرار گرفت که به دقیقیت $0.9292 \pm 0.0134\%$ رسید. در سال ۲۰۱۶ Mohammad Haghhighat و همکاران در [۲۲] یک روش برای هنجارسازی حالت چهره بر اساس تنظیم کردن مدل ظاهری فعال یا AAM یک مدل

پارامتری است که برای ارائه یک شکل مانند چهره انسان استفاده می‌شود. در این الگوریتم ابتدا یک AAM بر روی تصویر چهره قرار گرفته، با روندی تکراری و به صورت بهینه شونده، بر روی چهره تنظیم می‌شود. سپس با استفاده از یک تبدیل هم نسبی، مرحله رو به رو سازی بر روی چهره انجام می‌پذیرد. شکل ۳-۵ رویکرد کلی این الگوریتم را نشان می‌دهد. در این مدل یک تصویر



شکل ۳-۵: رویکرد کلی الگوریتم مبتنی بر AAM برای رو به رو سازی چهره [۷].

چهره با مجموعه‌ای از نقاط ویژه هنجارسازی شده مدل می‌شود که به صورت $\mathbb{A}, \mathbb{B}, \mathbb{C}$ تعریف می‌شود که در آن $\mathbb{A} = 1, 2, \dots$. چهره با مجموعه‌ای از نقاط ویژه هنجارسازی شده مدل می‌شود که به صورت $\mathbb{A}, \mathbb{B}, \mathbb{C}$ تعریف می‌شود که در آن $\mathbb{A} = 1, 2, \dots$. برای انجام این کار یک مرحله یادگیری نیاز است. سپس الگوریتم PCA اعمال می‌شود تا کاهش میزان وابستگی میان نقاط ویژه در هر مجموعه انجام می‌شود و نتیجه یک مدل خطی است که یک مدل نمونه را به صورت زیر نمایش می‌دهد.

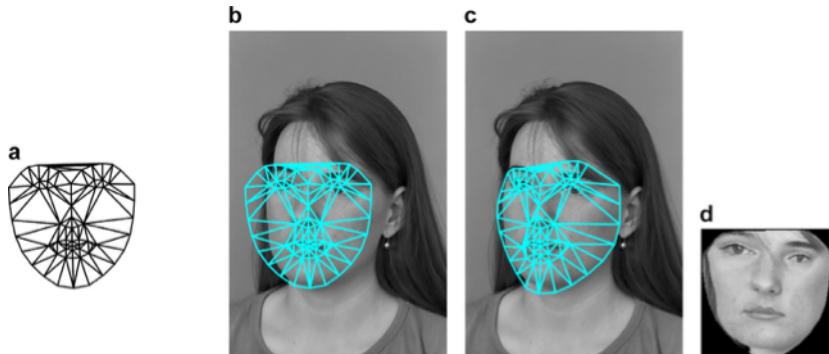
$$S = s_0 + \sum_{i=1}^n p_i s_i \quad (1.1)$$

که در آن \mathbb{A} . شکل پایه، \mathbb{B} نشان دهنده \mathbb{A} امین شکل پایه و $\mathbb{C}, \mathbb{D}, \mathbb{E}$ عامل‌های شکل می‌باشند. ظاهر مدل یک تصویر (x) می‌باشد که در آن x مجموعه پیکسل‌های داخل شکل پایه \mathbb{A} . می‌باشد. مدل ظاهر یک چهره خاص از یک ظاهر پایه \mathbb{A} . و ترکیب خطی از بردارهای ویژه $\mathbb{B}, \mathbb{C}, \mathbb{D}$ تشکیل می‌شود که به صورت زیر تعریف می‌گردد.

$$A(x) = a_0(x) + \sum_{i=1}^m q_i a_i(x) \quad (2.1)$$

که در آن $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}$ عامل‌های ظاهر می‌باشند. عامل‌های شکل و ظاهر برای هر تصویر در فرایند AAM بدست می‌آید.

الگوریتم‌های POIC و SIC دو الگوریتم شناخته شده برای این منظور می‌باشند. رویکرد SIC نسبت به POIC در شرایطی که تصاویر آزمایشی با تصاویر آموزشی متفاوت باشند، بسیار بهتر عمل می‌کند. اما از طرفی دارای پیچیدگی محاسباتی بیشتری می‌باشد. در این مقاله از یک روش SIC سریع برای حل مسئله بهینه سازی با 100×100 تکرار استفاده شده است. اگر $\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}$ مجموعه عامل‌های بدست آمده باشد، یک تبدیل هم نسبی قطعه‌ای $W(x; p)$ برای رو به رو سازی چهره مورد استفاده قرار می‌گیرد که در آن هریک از مثلث‌های روی توری، به صورت جداگانه به تصویر نتیجه با استفاده از درونیابی نزدیک-ترین همسایه نگاشت پیدا می‌نمایند. برای مقداردهی اولیه از یک مدل پایه \mathbb{A} . استفاده می‌شود که مقدار p در آن صفر می‌باشد (شکل ۳-۶). شکل ۳-۶ - مقدار دهی اولیه و بهینه سازی AAM پس از تنظیم کامل مدل بر روی چهره، یک تبدیل هم نسبی با



شکل ۱.۶: رویکرد مبتنی بر تطبیق کلیشه [۷].

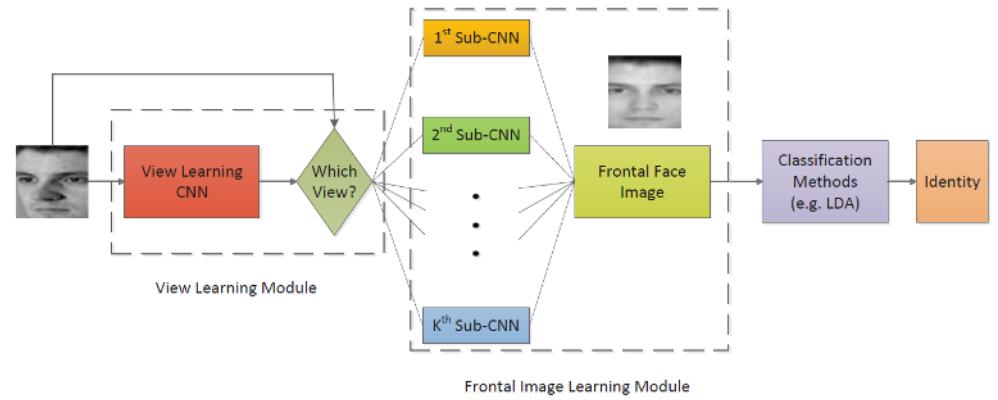
پارامترهای بدست آمده توسط الگوریتم یادگیری یاد شده، می‌تواند حالت چهره را هنجارسازی نماید. در بخش شناسایی چهره، ابتدا بخش چانه از تصویر حذف می‌شود زیرا چانه تقریباً تاثیری در شناسایی یک چهره ندارد. سپس تصویر چهره به اندازه 64×64 پیکسل تبدیل می‌شود و به 64×64 بخش غیر هم پوشان با اندازه 8×8 تقسیم می‌شود. سپس در هر بخش تبدیل DCT انجام می‌شود. ضرایب خروجی تبدیل DCT بر حسب یک پویش زیگزاگی مرتب می‌شوند. اولین ضریب در نظر گرفته نمی‌شود. زیرا نشان دهنده میانگین سطح خاکستری پیکسل های بخش می‌باشد. 10 ضریب بعدی که ضرایب فرکانس پایین می‌باشند، برای ایجاد بردار ویژگی چهره استفاده می‌شوند. برای آموزش و آزمایش از مجموعه داده LFW و FERET استفاده شده است که در آن تصاویر چهره با زوایای چرخش متفاوت وجود دارند. الگوریتم مورد استفاده در این مقاله موفق به دستیابی به شناسایی چهره با دقت $87\% - 93\%$ شده است. در سال ۲۰۱۶، Zhang و همکاران در [۲۹] یک روش رو به رو سازی چهره ارائه دادند که شناسایی چهره را



شکل ۱.۷: نتیجه آزمایش بر روی مجموعه داده FERET در زوایه های متفاوت [۷].

مستقل از نمای چهره انجام می‌دهد. این الگوریتم یادگیری عمیق که VS2VI نامیده می‌شود، از دو بخش اصلی تشکیل شده است. بخش اول یک شبکه عصبی پیچشی برای یادگیری نما و زوایه چهره می‌باشد 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 می‌باشد (شکل ۳-۶). این الگوریتم که می‌تواند با تعداد کمی داده نمونه، به خوبی آموزش بینند، دو بخش تشکیل شده از شبکه عصبی پیچشی

را به هم متصل می نماید تا مشکل نمای چهره در سامانه شناسایی چهره را برطرف نماید. در این معماری برای بازسازی چهره از زاویه رو به رو از لایه های واپیچشی به جای لایه های تمام متصل استفاده شده است. مدل VS2VI از دو بخش اصلی تشکیل شده

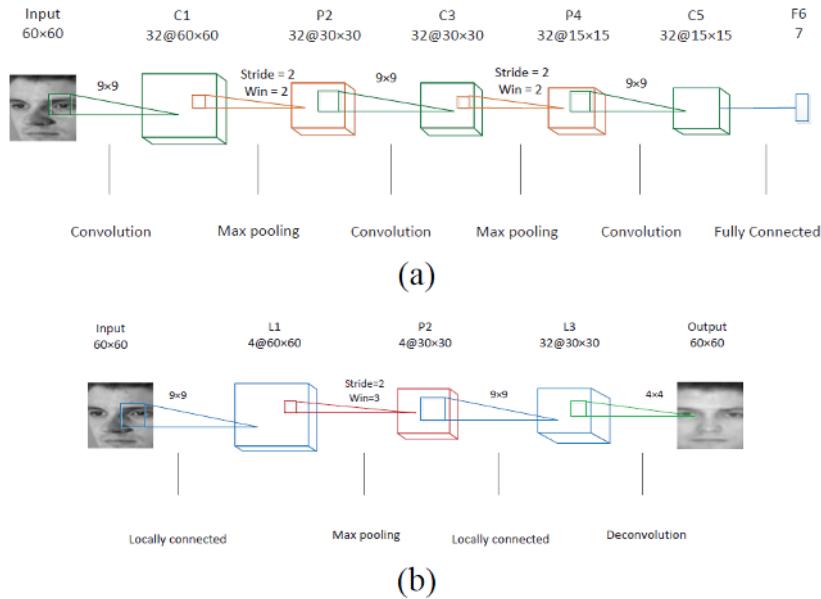


شکل ۸.۱: معماری شبکه پیشنهادی [۷] VS2VI

است. بخش اول به عنوان ورودی یک تصویر خاکستری شامل یک چهره در هر زاویه و نمای دلخواه با ابعاد 60×60 دریافت می کند و آن را با توجه به نمای چهره طبقه بندی می کند. سپس تصویر وارد بخش دوم که از تعدادی شبکه عصبی پیچشی که هر کدام برای یادگیری تناظر بین یک چهره از 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. بخش اول از یک شبکه عصبی پیچشی تشکیل شده است که شامل سه لایه پیچشی، دو لایه رای گیری و یک لایه تمام متصل می باشد. ورودی آن یک تصویر با هر موقعیت و زاویه دلخواه و خروجی آن احتمال قرار داشتن تصویر ورودی در هر دسته از دسته های مربوط به نمای مختلف می باشد. برای لایه های پیچشی از تابع فعالیت ReLU استفاده شده است. و لایه تمام متصل از softmax به عنوان تابع هزینه استفاده کرده است.

$$f(x) = \max(0, x) \quad (3.1)$$

بخش دوم از تعدادی زیر شبکه پیچشی که هر کدام برای یادگیری تناظر بین چهره از رو به رو با یک چهره از یک نمای خاص می باشد، تشکیل شده است. هر یک از این زیر شبکه ها شامل دو لایه با اتصال محلی، یک لایه رای گیری و یک لایه واپیچشی می باشند. سه لایه اول برای استخراج ویژگی ها و لایه آخر برای بازیابی چهره از رو به رو می باشند. ورودی و خروجی این لایه ها تصویر چهره می باشد. لایه آخر به جای لایه تمام متصل از لایه واپیچشی استفاده شده است. زیرا حجم محاسبات را به طور قابل



شکل ۹.۱: (a) معماری مدل یادگیری موقعیت چهره و (b) معماری مدل یادگیری بازسازی چهره از رو به رو [۷].

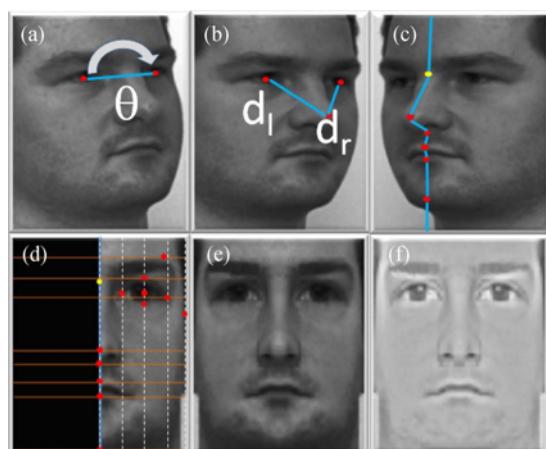
توجهی کاهش می دهد. یک لایه تماماً متصل به 10^3 میلیون پارامتر نیاز دارد، در حالی که لایه واپیچشی به ۴۶۰ هزار پارامتر نیاز دارد. لایه اول که اتصال محلی دارد، از تابع PreLU به عنوان تابع فعالیت استفاده کرده است. لایه واپیچشی برای نمونه افزایی از درون یابی دو خطی استفاده کرده و تابع هزینه آن $loss - \ell_2$ می باشد. برای یادگیری شبکه از الگوریتم پس انتشار خطای Andrey VS2VI به دقت ۹۵٪ در تشخیص چهره با زاویه ۴۵ درجه رسیده است. در سال ۲۰۱۸ استفاده شده است. الگوریتم V.Savchenko و همکاران در [۴۴] یک روش مبتنی بر ML برای شناسایی چهره در محیط های کنترل نشده با تعداد کم نمونه ها بر اساس محاسبه فاصله بین ویژگی های با ابعاد 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 Kullback–Leibler بین ویژگی های غیرمنفی تخمین زده شده است. این رویکرد بر روی مجموعه داده های IJB-A, YTF و LFW مورد آزمایش قرار گرفته است. با توجه به شکل ۹.۳ رویکرد پیشنهادی می تواند با استفاده از فواصل سنتی، افزایش دقت ۳.۰ تا ۵.۵ درصد در مقایسه با روش های شناخته شده داشته باشد، به ویژه اگر تصاویر آموزش و آزمایش تفاوت زیادی داشته باشند. در سال ۲۰۱۳ Marsico و همکاران در [۲۶] یک روش رو به رو سازی چهره ارائه دادند. در ابتدا از الگوریتم STASM برای به دست آوردن ۶۸ نقطه ویژه بر روی چهره استفاده شده است. سپس برای هر تصویر ورودی، شاخص حالت نمونه (SP) محاسبه می شود و در صورتی که مقدار آن کمتر از یک آستانه باشد، تصویر مردود شده و در غیر این صورت به مرحله بعد برای هنجارسازی حالت فرستاده می شود. هرچه مقدار شاخص SP بالاتر باشد، تصویر چهره به



شکل ۱۰.۱: مقایسه روش ارائه شده با سایر روش ها (a) تصویر آزمایشی (b) و (c) خروجی نادرست روش های دیگر (d) خروجی روش ارائه شده [۷].

حالت تمام رخ نزدیکتر است و اصلاح زاویه کمتری نیاز دارد. شکل ۱۱-۳ قسمت a تا c معیارهای مورد نیاز برای محاسبه شاخص

را نشان می دهد. چرخش: چرخش سر در جهت عقربه های ساعت یا عکس آن می باشد. و به صورت زاویه θ تعریف می شود



شکل ۱۱.۱: ۶ مرحله اصلی در فرایند هنجارسازی حالت و روشنایی چهره [۷].

که زاویه بین خط عبوری از مرکز چشم ها و محور افقی θ می باشد.

$$roll = \min\left(\left|\frac{2\theta}{\pi}\right|, 1\right) \quad (4.1)$$

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

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

$$yaw = \frac{\max(d_l, d_r) - \min(d_l, d_r)}{\max(d_l, d_r)} \quad (5.1)$$

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

$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)} \quad (6.1)$$

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

$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch) \quad (7.1)$$

که در آن

$$\alpha + \beta + \gamma = 1 \quad (8.1)$$

که مقادیر این ضرایب از طریق آزمون و خطابه دست می آیند. سپس در مرحله تمام رخ کردن تصویر چهره، بین دو فاصله dr و dl

هر کدام بزرگتر باشند، نشان می دهد آن سمت از چهره بیشتر در دید دوربین است. اگر نیمه سمت راست صورت به طرف دوربین

باشد ($dl \geq dr$)، تصویر بدون تغییر باقی می ماند. در غیر این صورت، تصویر حول محور عمودی بر عکس می شود که باعث می

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

۱۱-۳ قسمت d و e نیمه سمت چپ تصویر حذف شده و از روی تصویر نیمی از چهره، نیمه دیگر نیز ساخته می شود و تصویر تمام

رخ چهره به دست می آید.

۳.۱ چالش روشنایی

متعادل سازی بافت نگار یکی از الگوریتم های مهم در پردازش تصویر است که هدف آن افزایش وضوح تصویر با یکنواخت سازی بافت

نگار تصویر است، به گونه ای که بخش های از تصویر که به علت روشنایی کم یا زیاد، پنهان می باشند، قابل مشاهده شوند. متعادل

سازی بافت نگار قدرتمندترین و رایج ترین روش برای اصلاح روشنایی تصاویر است. اما ضعف این روش، سراسری بودن آن می باشد.

برای رفع این مشکل باید از الگوریتم های محلی استفاده کرد. در سال ۲۰۱۳ Marsico و همکاران در [۲۶] یک روش هنجار سازی

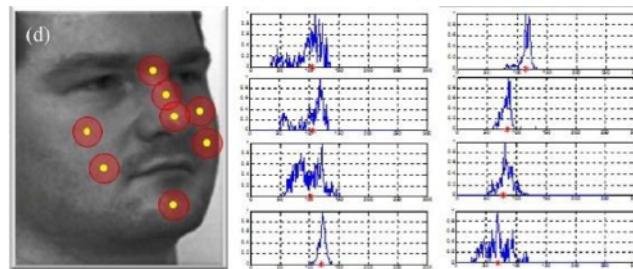
نورپردازی برای تصاویر چهره ارائه دادند و از این روش برای محاسبه شاخص روشنایی نمونه (SI) استفاده کردند. زمانی که تصویر

روشنایی یکنواخت دارد، بیشتر بخش های چهره توزیع یکنواخت سطح خاکستری دارند. اما وقتی روشنایی یکنواخت نباشد، برخی

از نواحی خاص چهره، توزیع یکنواخت سطح خاکستری ندارند. برای مثال جلوی بینی، گونه ها و چانه معمولاً نور را منعکس می

کنند. ۸ ناحیه در شکل ۱۲-۳ با توجه به چنین اصلی انتخاب شده اند. ۸ بافت نگار با رنگ آبی و مرکز آن ها با رنگ قرمز مشاهده

می شود. ۸ بافت نگار فوق به یک توزیع یکنواخت با انحراف معیار کم در همسایگی از مرکز حجم بافت نگار اشاره دارد. بافت نگار



شکل ۱۲.۱: اندازه گیری روشنایی و بافت نگار ۸ نقطه خاص [۷].

هر یک از ناحیه ها بدست آمده و مرکز ثقل آن محاسبه می شود:

$$mc(w) = \frac{\sum_{i=0}^{255} i \times h_w(i)}{\sum_{i=0}^{255} h_w(i)} \quad (9.1)$$

که در آن w نشان دهنده یکی از نواحی ۸ گانه می باشد. ۸ مرکز جرم محاسبه شده، بردار mc را تشکیل می دهند. با توجه به

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

کیفیت روشنایی محاسبه می شود که یک عدد می باشد:

$$SI = 1 - F(std(mc)) \quad (10.1)$$

هرچه مقدار SI بیشتر باشد، یعنی تصویر روشنایی یکنواخت تری دارد. اگر این شاخص به اندازه کافی رضایت بخش نباشد، تصویر رد می شود. در غیر این صورت برای هنجارسازی روشنایی وارد بخش بعدی خواهد شد. در صورت رد شدن تصویر، سیاست های جایگزین برای رسیدگی به این موضوع در دسترس هستند. برای مثال ممکن است یک نمونه جدید درخواست شود که در شرایط برون خط امکان پذیر نیست. یا مداخله انسانی می تواند به صورت دستی نمونه را طبقه-بندی کند. در هر صورت بیشتر بار طبقه بندی بر دوش سامانه خواهد بود. اگر تصویر به مرحله بعد وارد شد، با استفاده از الگوریتم SQI توسط یک ماسک مربعی با اندازه 8×8 مقدار هر پیکسل بر مقدار میانگین همسایگانش تقسیم می شود و نتیجه نهایی حاصل می شود. نتیجه به صورت قسمت f در شکل می باشد. در سال ۲۰۱۵ Jamal Hussain Shah و همکاران در [۳۶] رویکردی برای تشخیص چهره در تغییرات شدید روشنایی پیشنهاد دادند که به سه مرحله تقسیم شده است:

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

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

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

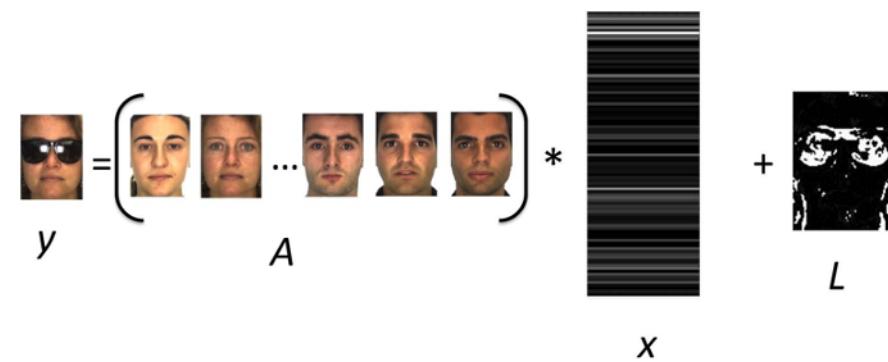
۴.۱ چالش انسداد

در سال ۲۰۱۸ Cho Ying Wu و همکاران در [۳۸] یک رویکرد مبتنی بر واپاژش با جهت گرادیان برای شناسایی چهره های در معرض انسداد ارائه دادند. در کاربردهای واقعی، تعداد داده های آموزش بسیار کم می باشد (شاید یک تصویر به ازای هر شخص). این رویکرد توانایی برخورد با این شرایط را دارد و در مقابل تصاویری که نزدیک به ۸۰ درصد از چهره در شرایط انسداد قرار دارد، به خوبی عمل می کند. نتایج نشان می دهد که با تعداد بسیار کمی از تصاویر آموزشی، مدل پیشنهاد شده GD-HASLR بهترین عملکرد را در مقایسه با سایر روش های پیشفرته، از جمله روش های مبتنی بر شبکه عصبی پیچشی دارد. مجموعه داده آموزشی $A = \mathbb{R}^{d \times n}$ در نظر گرفته شده که در آن n تعداد داده های آموزشی و d حاصل ضرب تعداد پیکسل های طول و عرض تصاویر می باشد. داده های آموزشی چهره های طبیعی و بدون انسداد می باشند. $y = \mathbb{R}^d$ 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 خطای $L = \mathbb{R}^d$ نیز می باشد.

(شکل ۱۶-۳)

$$y = Ax + L \quad (11.1)$$

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



شکل ۱۳.۱: تصویر انسداد از ترکیب خطی تمام چهره های آموزشی در مجموعه داده و یک تصویر L که نشان دهنده انسداد است، تشکیل شده است [۷].

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \|s.t. y - Ax\| \quad (12.1)$$

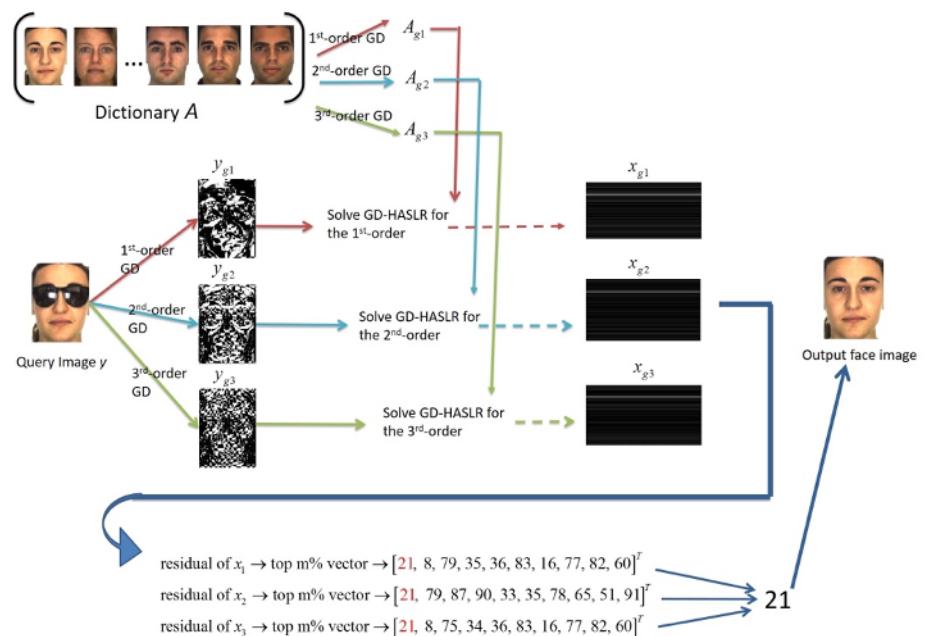
که در آن ϵ یک آستانه خطای خطا می باشد. برای تصویر و ورودی و تصاویر مجموعه داده آموزش، گرادیان مرتبه اول، دوم و سوم محاسبه شده و به عنوان ویژگی هر تصویر در نظر گرفته می شود. در ادامه شرط کم رتبه بودن ماتریس ویژگی ها نیز به این رابطه اضافه می شود. با استفاده از روش ضرایب لاغرانژ، رابطه بالا را می توان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

$$L(x, L, z) = \alpha LM + xi + zT(y - Ax - L) + 2y - Ax - L22 \quad (13.1)$$

که در آن α ضریب لاغرانژ و β عامل مجازات می باشد. پس از بدست آوردن بردار تک x می توان باقیمانده دسته Ω را به صورت زیر محاسبه نمود:

$$r_i = y - Ai(x)2 \quad (14.1)$$

که در آن (x) نشان دهنده Ω امین انتخاب کننده دسته می باشد که فقط ورودی های مربوط به دسته Ω را حفظ می کند و در سایر قسمت ها برابر با صفر می باشد. در نهایت دسته ای که کمترین باقیمانده را داشته باشد، انتخاب می شود. رویکرد کلی الگوریتم در شکل ۱۴-۳ آمده است. در سال ۲۰۱۴ J. Li و همکاران در [۳۷] یک روش تشخیص چهره پوشیده شده در پس زمینه پیچیده



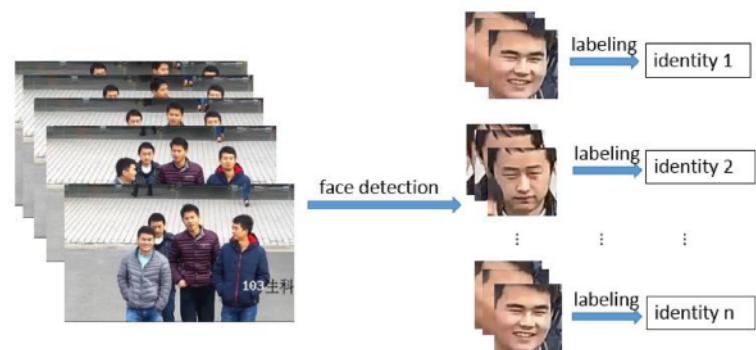
شکل ۱۴.۱: رویکرد کلی الگوریتم [۷] GD-HASLR

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

۵.۱ چالش کمبود تصاویر آموزشی

در سال ۲۰۱۷ Ya Wang 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 است، نوآوری این رویکرد، ساختن یک مجموعه داده استاندارد برای شبکه عصبی از روی دوربین های نظارتی در محیط است که در چهار مرحله به صورت زیر ساخته می شوند. با توجه به اینکه تصاویر مورد نظر برای هر فرد در مجموعه فریم های پشت سر هم از یک دوربین موجود است، می توان مجموعه تصاویر یک فرد را بوسیله ترکیب الگوریتم تشخیص چهره و ردیابی چهره جمع آوری کرد. پس از شناسایی یک چهره، با ردیابی آن به وسیله الگوریتم KCF، مجموعه تصاویری از آن به عنوان یک دسته طبقه بندی می شود. برخی تصاویر در هر دسته



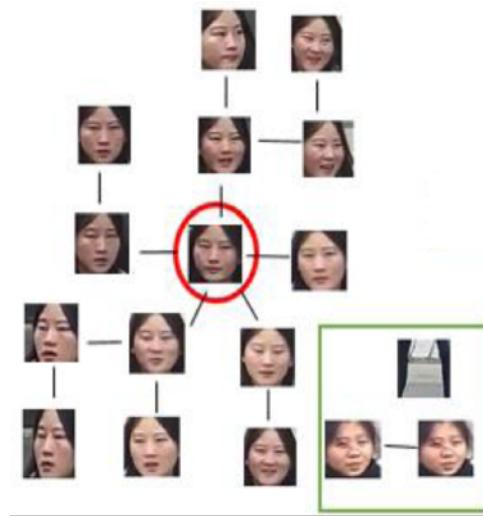
شکل ۱۵.۱: ردیابی، یافتن چهره ها و برچسب زنی [۷].

به اشتباه در مرحله اول به عنوان تصویر یک فرد در نظر گرفته شده اند (شکل ۱۶-۳). با استفاده از روش خوش بندی گراف روی ویژگی های استخراج شده از شبکه VGG-Face، تشخیص و پاک سازی تصاویر اشتباه انجام می شود. فاصله کسینوسی بین ویژگی های تصاویر چهره محاسبه می شود و اگر این فاصله برای هر دو تصویر کمتر از یک مقدار آستانه باشد، این تصاویر متعلق به یک فرد هستند. با توجه به شکل ۱۷-۳ تصویری که بیشترین شباهت را به تصاویر دیگر دارد، به عنوان شاخص برای آن شخص



شکل ۱۶.۱: تصاویر با حاشیه قرمز رنگ، به اشتباه برچسب زنی شده اند [۷].

انتخاب می شود. با استفاده از محاسبه فاصله بین هر داده با داده مرکزی و در نظر گرفتن یک آستانه، داده های تکراری در هر دسته



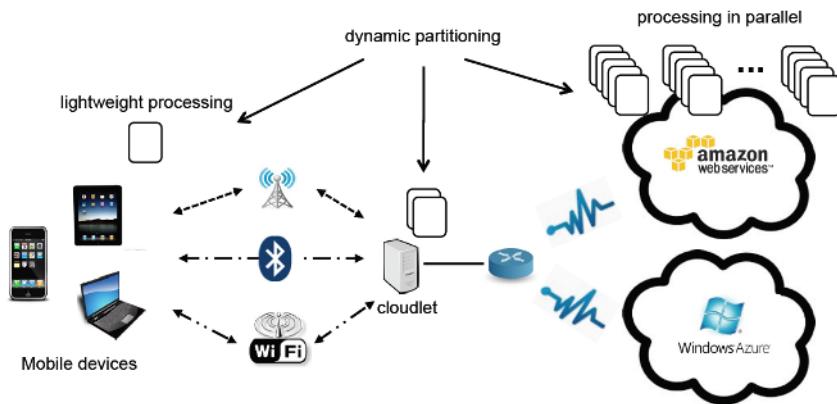
شکل ۱۷.۱: استفاده از روش خوشه بندی گراف و تعیین تصویر شاخص [۷].

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 VGG-Face همراه با Fine-tuning برای طبقه بندی تصاویر آزمایشی استفاده شده است که به دقت ۹۲٪ رسیده است.

۶.۱ چالش منابع محدود

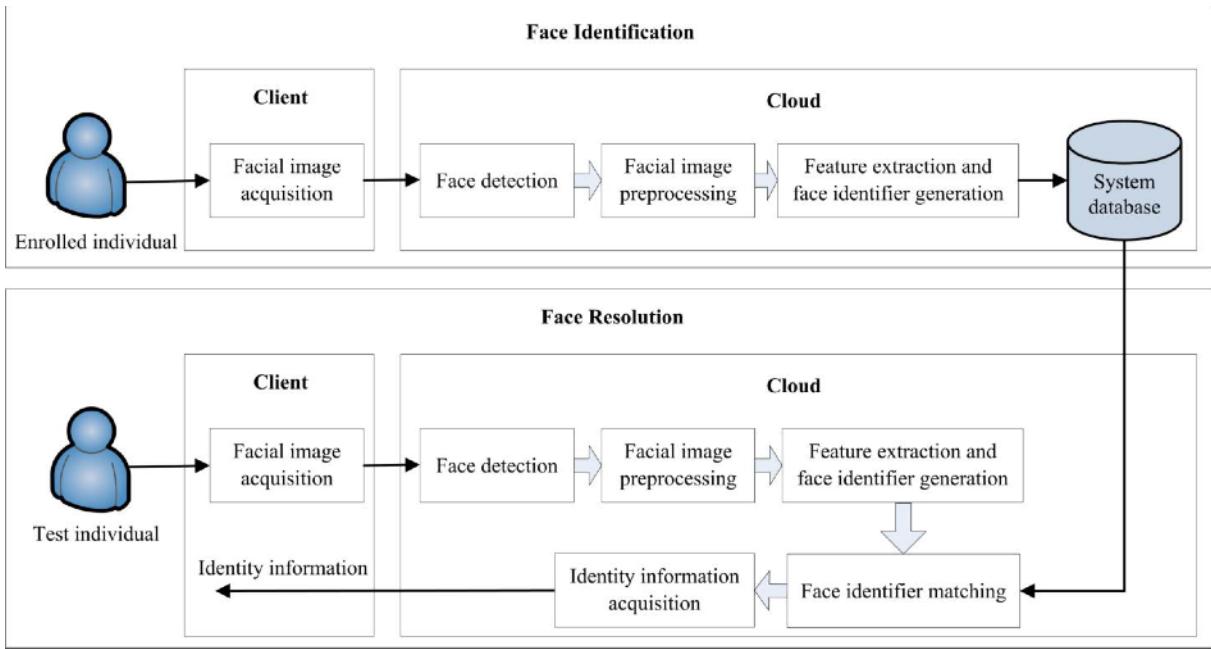
در سال ۲۰۱۲ Tolga Soyata و همکاران در [۴۰] یک روش تشخیص چهره بی درنگ مبتنی بر بینایی ابری با استفاده از معماری MOCHA ارائه کردند (شکل ۱۸-۳). با فرآگیر شدن تلفن همراه هوشمند در میان شهروندان، سامانه تشخیص چهره می تواند از همکاری مشترک محاسبات تلفن همراه و رایانش ابری استفاده کند. چالش این سامانه، چگونگی 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 تأثیر ارتباطات مختلف و قدرت

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



شکل ۱۸.۱: معماری MOCHA: دستگاه های تلفن همراه از طریق اتصال چندگانه با cloudlet و ابر ارتباط برقرار می کنند [۷].

شود: دستگاه همراه: تلفن های همراه و iPad ها نقش تهیه و ارسال تصاویر را دارند. تصاویر با فرمت RAW فرستاده می شوند تا قابلیت پیش پردازش بهتری داشته باشند. اگر سرور ابر به دستگاه همراه نزدیک باشد و ارتباط با سرعت بالا امکان پذیر باشد، تصاویر پیش پردازش به سرور فرستاده می شوند. در غیر این صورت مرحله پیش پردازش در دستگاه همراه انجام می شود و فقط اطلاعاتی همچون ویژگی های Haar و طبقه بندها به سرور فرستاده می شوند. پس از اتمام فرایند تشخیص چهره، نتیجه نهایی برای تلفن همراه فرستاده می شود. ابر کوچک : سرورها و رایانه هایی که توانایی پردازشی متوسطی دارند، ابر کوچک یا cloudlet برای تلفن همراه فرستاده می شوند. این دستگاه ها که به عنوان میان دستگاه های همراه و سرورهای ابری اصلی قرار دارند، مجهز به GPU می باشند نامیده می شوند. این دستگاه ها که به دوش می کشند و تصمیم گیری نهایی بر روی آن انجام می پذیرد. در سال ۲۰۱۸ Pengfei Hu و همکاران در [۴۱] یک رویکرد تشخیص چهره مبتنی بر رایانش ابری ارائه کردند. افزایش برنامه های کاربردی در زمینه کلان داده ها باعث افزایش تقاضای سامانه های شناسایی چهره برای محاسبات قدرتمند و ظرفیت ذخیره سازی بالا می شود. این سامانه به طور کامل از مزایای محاسبات ابری بهره می برد تا به طور موثر توانایی محاسبات و ظرفیت ذخیره سازی را بهبود بخشد. نتایج تجربی نشان می دهد که طرح پیشنهادی عملکردی ارائه شده است و می تواند سرویس شناسایی موثر چهره را فراهم کند. همانطور که در شکل ۱۹-۳ مشاهده می شود، تنها تهیه تصویر بر عهده دستگاه سرویس گیرنده می باشد و تمام محاسبات باقتن و شناسایی چهره بر روی ابر انجام می شود. در این سامانه تصویر با فرمت RGB برای ارسال شده و برای یافتن چهره از ویژگی های

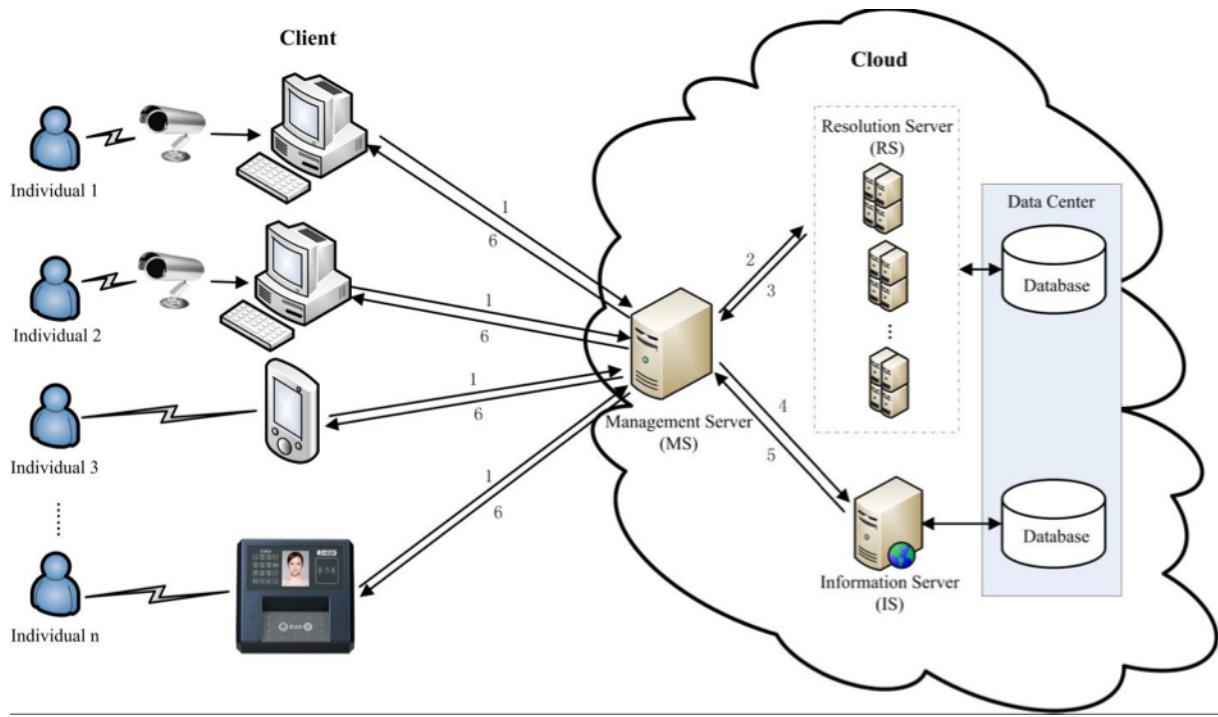


شکل ۱۹.۱: نمای کلی سامانه تشخیص چهره مبتنی بر رایانش ابری [۷].

استفاده شده است. سپس عملیات همسان سازی بافت نگار بر روی تصویر چهره اعمال می شود تا بهبود جزئی حاصل شود. سپس از الگوریتم HOG برای استخراج ویژگی های چهره استفاده شده، برای هر تصویر یک شناسه تولید می گردد و در نهایت با استفاده از فاصله اقلیدسی با شناسه تصاویر موجود در پایگاه داده مطابقت داده می شود. همانطور که در شکل ۲۰-۳ مشاهده می شود این سامانه ابری از بخش های سرور مدیریت (DB)، سرور اطلاعات (DB)، سرور شناسایی (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.

۷.۱ نتیجه گیری

به تازگی یادگیری عمیق در تشخیص چهره و بسیاری 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 را حل می کند؟ چالش روش های یادگیری عمیق در 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 مرکز دوربین، کیفیت پایین تصویر، بالا رفتن سن افراد و کمبود داده های آموزشی از مواردی می باشند که می توانند سامانه تشخیص چهره را با مشکل رو به رو نمایند.



شکل ۲۰.۱: چارچوب سامانه شناسایی چهره مبتنی بر رایانش ابری [۷].

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 هایی همچون گرایش های شدید، نور کم و تصاویر کوچک و تاریک چهره تبدیل می

کند. محققان فرض کرده اند که لایه های عمیق 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 چسب نمی باشد و ساختار سه بعدی و پیچیده ای دارد و ممکن است تصویر از هر

زاویه ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده های

آموزشی بسیاری می باشد که در شرایط نورپردازی، زاویه و حالت های مختلفی تصویربرداری شده باشد. در مقابل فرض بر این است

که داده های آموزش بسیار کم هستند. از این رو مسئله تشخیص چهره باید در شرایطی حل شود که داده های آموزشی کافی در

اختیار نمی باشد. بنابراین به الگوریتمی نیاز داریم که به ما کمک کند با تولید داده های غیر واقعی، مشکل کمبود داده های آموزشی

را حل نماییم. از سویی دیگر محدودیت منابع برای اجرای پردازش‌ها 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 \mathbb{Z} که می‌تواند یک نویز تصادفی باشد، یک تصویر تولید می‌کند و شبکه تمیز دهنده وظیفه دارد تصاویر واقعی را از تصاویر تولید شده توسط شبکه تولید کننده تشخیص دهد. بنابراین هر تصویر با یک بردار Z معرفی می‌شود. در این مقاله محاسبات در فضای برداری انجام شده و بردار حاصل، تبدیل به تصویر خروجی می‌شود. به عنوان مثال بردار Z برای تصویر خانمی که عینک آفتابی نزدیک است از بردار Z برای تصویر خانمی که عینک آفتابی زده است، کم می‌شود و حاصل آن، بردار مربوط به یک عینک آفتابی می‌باشد. سپس این بردار با بردار تصویر آقایی که عینک نزدیک است جمع می‌شود. نتیجه نهایی تصویر همان آقا با عینک آفتابی می‌باشد. به عنوان مثالی دیگر، با میانگین گیری بردارهای مربوط به دو تصویر از چهره شخصی که به سمت راست و چپ متمایل است، توانسته چهره را به روی شخص را بازسازی نماید. اما کیفیت کار هنوز تا حالت مطلوب فاصله دارد. مقاله [۴] یک روش مبتنی



شکل ۲۱.۱: تولید تصاویر چهره از زوایای مختلف با استفاده از درونیابی بردارهای تصاویر چپ و راست [۷].

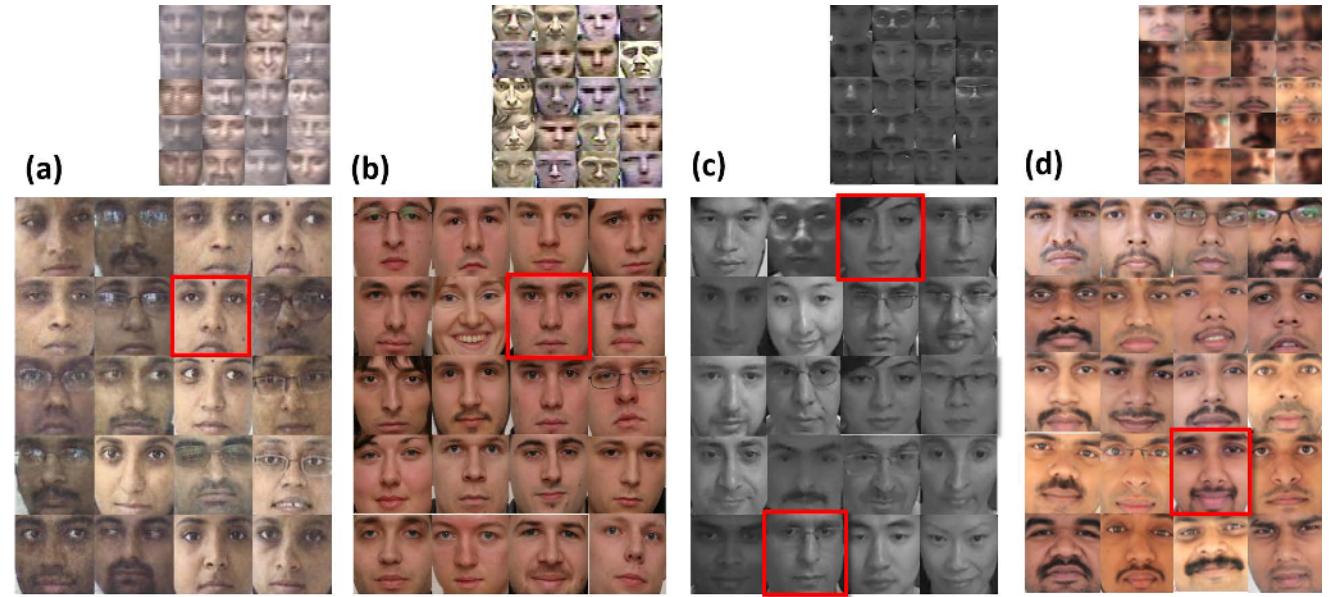
برای شناسایی چهره در محیط‌های بدون محدودیت با تعداد کم نمونه‌ها بر اساس محاسبه فاصله بین ویژگی‌های با بعد بالا که توسط شبکه عصبی پیچشی عمیق مانند SENet، ResNet و VGG استخراج شده است ارائه دادند. این روش جدید

شناسایی آماری، احتمال فاصله‌ها را نسبت به تمام تصاویر مجموعه داده‌ها با استفاده از قانون بیز به حداقل می‌رساند. این احتمال با تخمین توزیع هنجار طبیعی Kullback–Leibler می‌باشد که بروی مجموعه داده‌های IJB-A و YTF-LFW مورد آزمایش قرار گرفته است. رویکرد پیشنهادی می‌تواند با استفاده از فواصل سنتی، افزایش دقیق 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 GAN را با نام ۲۰۱۷-۲۰۱۸ پیشنهاد می‌دهد، که تصاویر واقع گرایانه با وضوح بالا را از روی تصاویر با وضوح پایین بازسازی می‌کند. این تصاویر چهره غیر واقعی اما واقع گرایانه و با کیفیت، باعث عملکرد بهتر سامانه شناسایی چهره برای مقایسه تصاویر می‌شود. رویکرد اصلی مقاله در روش آموزش تخصصی LR-GAN بهینه سازی تابع ضرر بازسازی چند مقیاسی است، بر اساس شاخص‌های مانند: شاخص شباهت ساختاری چند مقیاسی (SSIM)، میانگین مربعات خطای برای هر قسمت (PMSE)، و اگرایی جنسن شanon اصلاح شده (JSD) و تنوع متقابل در اطلاعات (MVI). شبکه تمیز دهنده در LR-GAN، بر اساس اطلاعات طبقه بنده که به طور ضمنی در طول آموزش آموخته می‌شود، هویت هر شخص را حفظ می‌کند. این رویکرد سریعتر از شبکه‌های مبتنی بر GAN اخیر به یک همگرایی می‌رسد. این مدل که به دقیق ۹۰٪ رسیده است، رتبه اول را در ۴ مجموعه داده شرایط بدون محدودیت کسب کرده است.

شکل ۳ تولید تصویر با وضوح بالا از روی تصاویر با وضوح پایین در ۴ مجموعه داده مختلف. موارد با حاشیه قرمز خروجی اشتباه هستند - [۶] شبکه‌های GAN یاد می‌گیرند تصاویر جدیدی تولید کنند که شبیه به تصاویر واقعی باشند. اما این شبکه‌ها معمولاً کنترل کمی روی ویژگی‌های بصری تصاویر خروجی دارند. مقاله [۷] یک شبکه GAN جدید پیشنهاد می‌دهد که بخش تولید کننده آن به طور خودکار یاد می‌گیرد بدون هیچ ناظر انسانی ویژگی‌های بصری متفاوت تصاویر را از یکدیگر جدا نماید. پس از اتمام مرحله یادگیری، ما می‌توانیم این ویژگی‌های بصری را به دلخواه خود ترکیب نماییم. برای مثال ویژگی‌های اساسی مانند جنسیت، سن، طول مو، وجود عینک و زاویه چهره را از تصویر ۱ با ویژگی‌های دیگری از تصویر ۲ ترکیب کرد و یک چهره جدید تولید نمود. نگاه این شبکه تولید کننده به هر تصویر، مجموعه ای از ویژگی‌های بصری می‌باشد. هر ویژگی بصری با اندازه مشخص، جلوه‌های تصویر را کنترل می‌کند. ویژگی‌های بصری غالب مانند زاویه چهره، مو، شکل صورت؛ ویژگی‌های بصری میانی مانند فرم لب و چشم‌ها و ویژگی‌های سبک تر مانند رنگ. ما می‌توانیم این ویژگی‌های بصری را با ضرایب دلخواه خود ترکیب نماییم.

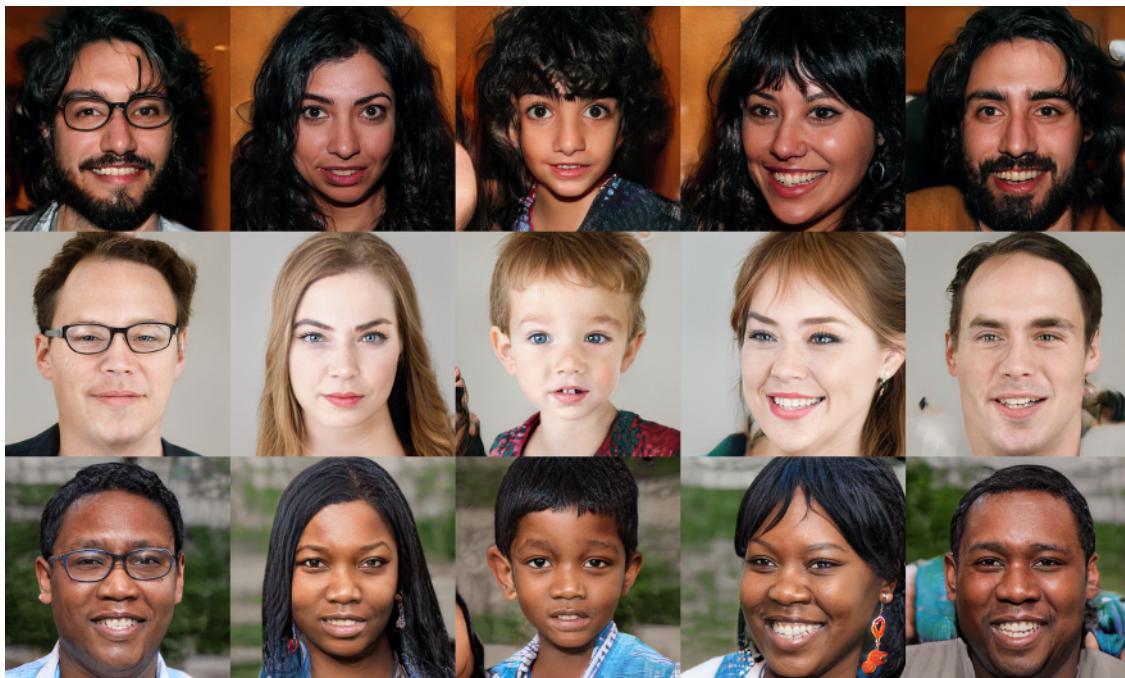
شکل ۳ ۲۰ تصاویر هر ردیف و هر ستون دارای برخی ویژگی‌های دیداری مشابه هستند - [۷] مقاله [۸] یک شبکه GAN جدید پیشنهاد می‌دهد که ورودی آن یک تصویر و خروجی آن نیز یک تصویر می‌باشد. این روش پیشنهادی که به روش ۲۰۱۷-۲۰۱۸ یا ۲۰۱۸-۲۰۱۹ شناخته می‌شود، به گونه‌ای طراحی شده که به عنوان ورودی یک تصویر را دریافت می‌نماید و خروجی

آن نیز یک تصویر می‌باشد که تغییرات مورد نظر بر روی آن اعمال شده باشد. 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 بسیار نزدیک به برچسب همان تصویر تولید نماید.



شکل ۲۲.۱: رویکرد مبتنی بر تطبیق کلیشه [۷].

شبکه تولید کننده از معماری U-Net 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 GAN جدید پیشنهاد می‌دهد که ورودی آن یک تصویر و خروجی آن نیز یک تصویر می‌باشد. این روش پیشنهادی که به روش Unpaired Image-to-Image یا 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. شبکه تولید کننده از دو شبکه autoencoder تشکیل شده است که به صورت موازی با یکدیگر آموزش داده خواهند 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 تصویری مشابه تصویر اولیه تبدیل می‌شود. شبکه تمیز دهنده باید این دو

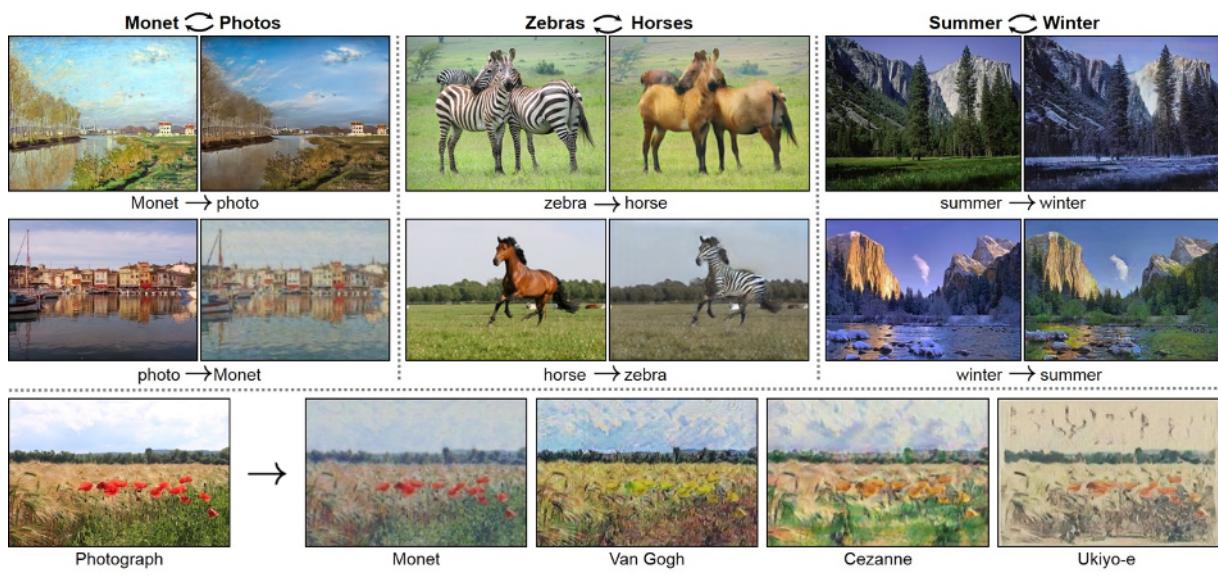


شکل ۱: حفظ محتوى و انتقال style از تصاویر مجموعه ۱ به تصاویر مجموعه ۲ و برعکس [۷].

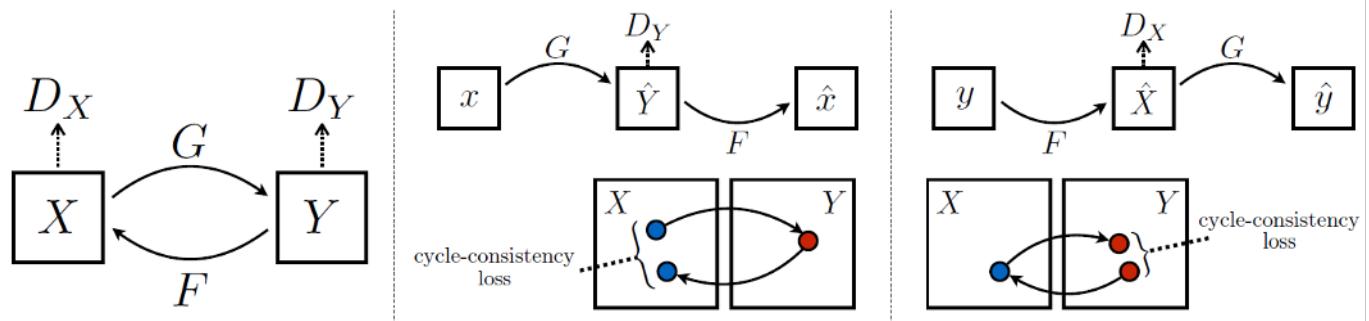
تصویر را با یکدیگر مقایسه کرده و خطای مناسب را برای به روز رسانی وزن‌های شبکه برگرداند. همین اتفاق برای تصاویر مجموعه دوم تکرار می‌شود. در مقاله [۱۰] به موضوع تولید چهره در زوایای دلخواه پرداخته شده است. در این مقاله از دو شبکه GAN استفاده شده است که در شبکه اول از روی چهره زاویه‌دار، چهره روبرو تولید شده است. سپس با استفاده از شبکه GAN دوم از روی تصویر چهره روبرو، تصویر با زاویه دلخواه با استفاده از یک پارامتر کنترلی تولید می‌شود. چالشی که در این مقاله به آن اشاره شده است، مسئله عدم توازن داده‌ها در وجود برخی ویژگی‌ها در تصاویر می‌باشد. این مقاله در برخی تصاویر چهره زاویه‌دار به مشکل برخورد می‌کرد. به عنوان مثال چهره‌هایی که دارای عارضه‌های پوستی می‌باشند توسط شبکه‌ها نادیده گرفته شده و تصویر چهره روبرو بدون عارضه تولید شده است. این چالش از جایی نشات می‌گیرد که تصاویر با عارضه پوستی در مجموعه داده بسیار کم می‌باشند و شبکه در مواجه با این مسئله ایده‌ای برای آن ندارد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند.

۸.۱ نتیجه گیری

بیشتر سامانه‌های تشخیص چهره عملکردهای قابل قبولی را در محیط‌های کنترل شده ارائه می‌دهند، اما در محیط‌های بدون محدودیت و در معرض تخریب شدید تصاویر چهره، عملکرد خوبی ندارند و در کاربردهای واقعی هنوز مسیری طولانی برای بهبود در



شکل ۲۴.۱: ساختار شبکه CycleGAN و نحوه محاسبه خطا به صورت انتزاعی [۷].



شکل ۲۵.۱: ساختار شبکه AD-GAN - شبکه GN برای رو به رو سازی چهره و شبکه GE برای تولید چهره از زوایای مختلف [۷].

پیش دارند. از جمله چالش‌های مهم، اساسی و عمومی در سامانه‌های تشخیص چهره می‌توان به موارد زیر اشاره نمود:

- تشخیص چهره در محیطی با تغییرات شدید نورپردازی مانند روز و شب (晴朗和夜間)
- تغییر زاویه و حالت چهره نسبت به دوربین (面部姿勢)
- انسداد صورت توسط اشیایی مانند عینک آفتابی و شال گردن (面部遮擋)
- تغییرات اساسی در چهره با گذر زمان، مانند رشد موها و ریش‌ها و یا بالا رفتن سن مانند سفید شدن موها (面部老齡化)
- تاری خارج از تمرکز دوربین (面部離焦)
- وضوح پایین تصویر (面部模糊)

● ردیابی چهره در فریم‌های ویدیو با در نظر گرفتن تناظر بین فریمی (FFM) (FFM)

دلیل اصلی به وجود آمدن چالش‌ها این است که چهره انسان یک شی چیز نمی‌باشد و ساختار سه بعدی و پیچیده‌ای دارد و ممکن است تصویر از هر زاویه‌ای گرفته شده باشد. بنابراین برای آموزش یک الگوریتم یادگیری که بتواند چهره افراد را از یکدیگر تمیز دهد، نیاز به داده‌های آموزشی بسیاری می‌باشد که در شرایط نورپردازی، زاویه و حالت‌های مختلفی تصویربرداری شده باشد. مقاله [۵] روشنی برای رو به رو سازی تصویر چهره پیشنهاد کرد که در برخی موارد، چهره را به خوبی می‌چرخاند، اما در نیمه 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 GAN برای تغییر زاویه چهره پیشنهاد داده بود که این الگوریتم نیز در برخی مواقع 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 مواجه با این مسئله راهکاری ارائه نمی‌دهد و فقط جهت چهره را تغییر می‌دهد و بافت غالب صورت را بر روی صورت خروجی اعمال می‌کند.

۲ فصل

نتیجه‌گیری و پیشنهادات

۱.۲ 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.

۲.۲ بحث و نتیجه‌گیری

روش‌های مختلفی برای شناسایی زودهنگام رتینوپاتی دیابتی و شناسایی میکروآنوریسم ارائه شده‌اند. دسته‌ای از این روش‌ها سعی کرده‌اند 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 اینجا 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 ببریم.

۳.۲ 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 متماز 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