

الله الرحمن الرحيم



دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

تشخیص چهره دارای ماسک به کمک یادگیری عمیق

نگارنده: سجاد رضوانی

استاد راهنما

دکتر منصور فاتح

استاد مشاور

دکتر حسین خسروی

شهریور ۱۴۰۱

ماحصل آموخته ایم را تقدیم می‌کنم به آنان که مهر آسمانی شان آرام بخش آلام زمینی ام است

به استوارترین تکیه گاهم، دستان پر مهر پدرم

به سبزترین نگاه زندگیم، چشمان سبز مادرم

که هرچه آموختم در مکتب عشق شما آموختم و هرچه بگو شتم قطره ای از دریای بی کران مهربانیان را پاس توانم بگویم .

امروز، مستی ام به امید شماست و فردا کلید باغ بهشت رضای شما

را آوردی کران سنگ تر از این ارزان نداشتم تا به خاک پایتان نثار کنم، باشد که حاصل تلاشم نسیم کوزه غبار خشکیان را

بزداید .

بوسه بر دستان پر مهرتان

اکنون که به یاری پروردگار و یاری و راهبانی اساتید بزرگ موفق به پایان این رساله شده‌ام و خیفه خود داشته‌ام که نهایت

پاسکزاری را از تمامی عزیزانی که در این راه به من کمک کرده‌اند را به‌عل آورم:

در آغاز از استاد بزرگ و دانشمند جناب آقای دکتر فلاح که راهبانی این پایان‌نامه را به عهده داشته‌اند کمال تشکر را دارم.

از جناب آقای خسروی که استاد مشاور این پایان‌نامه بوده‌اند نیز قدردانی می‌نمایم.

از داوران گرامی که زحمت داوری و تصحیح این پایان‌نامه را به عهده داشتند کمال سپاس را دارم.

خالصانه از تمامی اساتید و معلمان و مدرسانی که در مقطع مختلف تحصیلی به من علم آموختند و مرا از سرچشمه دانایی سیراب

کرده‌اند تشکر می‌کنم.

سجاد رضوانی

شهریور ۱۴۰۱

تعمیر نامه

اینجانب سجاد رضوانی دانشجوی دوره کارشناسی ارشد (دکتری) رشته مهندسی کامپیوتر (هوش مصنوعی) دانشکده کامپیوتر دانشگاه صنعتی شاهرود نویسنده پایان نامه تشخیص چهره دارای ماسک به کمک یادگیری عمیق تحت راهنمایی دکتر منصور فاتح متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، 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 ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود . استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

تشخیص چهره یکی از فعال ترین زمینه‌های تحقیقاتی بینایی ماشین و تشخیص الگو است که به طور گسترده در بسیاری از زمینه‌ها مانند شناسایی، کنترل دسترسی و امنیت عمومی کاربرد دارد. توسعه تکنیک‌های یادگیری عمیق، دسترسی به مجموعه داده‌گان چهره در مقیاس بزرگ و تولید سیستم‌های پردازشی با توان بالا، عملکرد سیستم‌های تشخیص چهره به طور قابل توجهی بهبود یافته است. با این حال، سیستم‌های تشخیص چهره در هنگام مواجهه با چالش‌هایی مانند تنوع ژست، روشنایی متفاوت، وضوح کم و انسداد هنوز عملکرد چندان رضایت‌بخشی ندارند. انسداد تصویر یکی از چالش برانگیزترین مشکلات تشخیص چهره است. در این چالش، ظاهر صورت به طور قابل توجهی تغییر می‌کند و ویژگی‌های هویتی چهره از بین می‌رود که تشخیص چهره را دشوار می‌سازد. استفاده از ماسک قسمت بزرگی از صورت از جمله بینی و چانه را مسدود می‌کند. از این رو، سخت‌ترین چالش انسداد صورت در نظر گرفته می‌شود. یکی از روش‌های تشخیص چهره در صورت وجود انسداد، بازسازی و ترمیم چهره است. در سال‌های اخیر، شبکه‌های مبتنی بر GAN عملکرد بسیار خوبی در حوزه ترمیم و بازسازی تصاویر داشته‌اند. در روش پیشنهادی این پایان نامه ابتدا ناحیه ماسک تشخیص داده شده و در ادامه، این بخش بازسازی می‌شود و چهره بدون ماسک ایجاد می‌شود. سپس چهره بازسازی شده به سیستم تشخیص چهره داده می‌شود. معماری شبکه عمیق پیشنهادی در بخش بازسازی چهره، مبتنی بر GAN می‌باشد. خروجی شبکه، علاوه بر تولید تصویر با کیفیت، ویژگی‌های هویتی ناحیه زیر ماسک، یعنی بینی و دهان را حفظ می‌کند. از این رو روش پیشنهادی باعث افزایش دقت حدود ۳۰ درصدی در مقایسه با تصویر با ماسک و حدود ۸ درصدی نسبت به روش‌های مقایسه شده می‌شود. همچنین معیارهای کمی SSIM، PSNR و FID بیانگر عملکرد مناسب روش پیشنهادی در بخش ترمیم ناحیه ماسک است.

کلمات کلیدی: تشخیص چهره، ترمیم تصویر، یادگیری عمیق، شبکه GAN

لیست مقالات مستخرج از پایان نامه

- 1- AttBANet: Attention boundary-aware Network for image segmentation

فهرست مطالب

۵	فهرست جداول
۵	فهرست شکل‌ها
۱	فصل ۱: مقدمه
۲	۱-۱ مقدمه
۳	۱-۲ بیان مسئله و تعاریف
۵	۱-۳ هدف تحقیق
۵	۱-۴ چالش‌ها و راه‌حل‌های آن
۶	۱-۵ نوآوری‌های تحقیق
۷	۱-۶ ساختار پایان‌نامه
۸	فصل ۲: ادبیات موضوع و کارهای پیشین
۹	۲-۱ مقدمه
۹	۲-۲ تاریخچه سیستم‌های تشخیص چهره
۱۱	۲-۳ ساختار سیستم‌های تشخیص چهره
۱۳	۲-۴ روش‌های تشخیص چهره
۱۳	۲-۴-۱ کارهای مرتبط با روشهای کلنگر
۱۷	۲-۴-۲ کارهای مرتبط با روش‌های محلی (هندسی)
۲۰	۲-۴-۳ روش‌های مبتنی بر توصیف‌کننده‌های بافت محلی
۲۴	۲-۴-۴ روش‌های مبتنی بر شبکه‌های عمیق
۲۸	۲-۵ تشخیص چهره تحت انسداد

۲۹	۲-۵-۱ روش‌های تشخیصی چهره تحت انسداد
۴۲	۲-۶ جمع‌بندی
۴۴	فصل ۳: روش پیشنهادی
۴۵	۳-۱ مقدمه
۴۶	۳-۲ بخش بندی ناحیه ماسک
۴۷	۳-۲-۱ دادگان تولیدی
۴۸	۳-۲-۲ شبکه بخش بندی پیشنهادی
۵۱	۳-۲-۳ تابع خطا ترکیبی
۵۳	۳-۳ شبکه بازسازی ناحیه ماسک
۵۴	۳-۳-۱ شبکه مولد
۵۵	۳-۳-۲ شبکه متمایزکننده
۵۶	۳-۳-۳ شبکه بینشی
۵۷	۳-۳-۴ تابع خطای نهایی
۵۷	۳-۳-۵ جمع بندی
۵۹	فصل ۴ ارزیابی و بررسی نتایج
۶۰	۴-۱ مقدمه
۶۰	۴-۲ مشخصات سیستم
۶۱	۴-۳ مجموعه داده‌ها
۶۳	۴-۴ جزئیات پیاده سازی
۶۳	۴-۴-۱ شبکه بخش بندی ناحیه ماسک
۶۴	۴-۴-۲ شبکه ترمیم ناحیه ماسک
۶۵	۴-۵ معیارهای اندازه گیری کمی

- ۶۵..... شبکه بخش‌بندی ناحیه ماسک ۴-۵-۱
- ۶۶..... شبکه ترمیم ناحیه ماسک ۴-۵-۲
- ۶۷..... نتایج ۴-۶
- ۶۷..... شبکه بخش‌بندی ناحیه ماسک ۴-۶-۱
- ۷۰..... شبکه ترمیم ناحیه ماسک ۴-۶-۲
- ۷۲..... بررسی تأثیر بازسازی ناحیه ماسک بر روی دقت شناسایی چهره ۴-۷
- ۷۳..... جمع‌بندی ۴-۸

۷۴ فصل ۵: نتیجه‌گیری و کارهای آینده

- ۷۵..... نتیجه‌گیری ۵-۱
- ۷۶..... کارهای آینده ۵-۲

۷۷ مراجع

۸۵ لغت‌نامه

- ۸۶..... مرتب بر اساس حروف الفبای فارسی
- ۹۰..... مرتب بر اساس حروف الفبای انگلیسی

فهرست جداول

- جدول ۴-۱ مقایسه کمی روش‌های مختلف بر روی مجموعه دادگان MFSD ۶۸
- جدول ۴-۲ مقایسه کمی روش پیشنهادی با روش‌های مختلف. ۷۱
- جدول ۴-۳ مقایسه تأثیر روش‌های مختلف ترمیم ناحیه ماسک، بر روی روش‌های مختلف شناسایی چهره. ۷۳

فهرست شکل‌ها

- شکل ۱-۱ انواع مختلف انسداد چهره. ۳
- شکل ۱-۲ مراحل تشخیص چهره در حالت کلی. ۳
- شکل ۲-۱ روش های تشخیص چهره. ۱۴
- شکل ۲-۲ نمونه ای از استخراج نشانه ها با استفاده از الگوریتم تطبیق گراف دسته ای الاستیک. ۱۸
- شکل ۲-۳ مثالی از محاسبه الگوی باینری محلی صورت. ۲۱
- شکل ۲-۴ مثالی از محاسبه صورت با الگوی باینری محلی یک بعدی. ۲۳
- شکل ۲-۵ انواع مختلف انسداد. ۳۰
- شکل ۲-۶ روش های استخراج ویژگی مقاوم به انسداد. ۳۱
- شکل ۲-۷ روش های بازیابی ناحیه مسدود. ۳۹
- شکل ۲-۸ خروجی تکنیک‌های مبتنی بر نقاشی تصویر. ۴۱
- شکل ۳-۱ بخش بندی ناحیه ماسک. ۴۵
- شکل ۳-۲ ورودی و خروجی شبکه GAN پیشنهادی. ۴۶
- شکل ۳-۳ نمونه تصاویر با ماسک که ماسک به صورت مصنوعی بر روی صورت قرار گرفته است. ۴۶
- شکل ۳-۴ گام های ایجاد مجموعه دادگان MFSD. ۴۷
- شکل ۳-۵ نمونه تصاویر دادگان MSFD. ۴۸
- شکل ۳-۶ معماری شبکه پیشنهادی Att-BANet. ۴۹
- شکل ۳-۷ دو بلوک ResNet. ۵۰
- شکل ۳-۸ ساختار گیت توجه. ۵۱
- شکل ۳-۹ معماری کلی شبکه GAN پیشنهادی. ۵۴
- شکل ۳-۱۰ معماری کلی شبکه متمایزکننده. ۵۶
- شکل ۴-۱ خروجی شبکه در تکرار های مختلف شبکه. ۶۱
- شکل ۴-۲ انواع مختلف ماسک صورت. ۶۲
- شکل ۴-۳ نمونه تصاویر ساخته شده برای آموزش شبکه. ۶۲
- شکل ۴-۴ نمونه تصاویر مجموعه دادگان FERET. ۶۳
- شکل ۴-۵ تصاویر مراحل مختلف شبکه مولد. ۶۵
- شکل ۴-۶ مقایسه روش پیشنهادی و روش های بروز دیگر. ۷۰
- شکل ۴-۷ مقایسه نتیجه روش پیشنهادی با سایر روش های ترمیم ناحیه ماسک. ۷۲

فصل ۱: مقدمه

۱-۱ مقدمه

امروزه تشخیص چهره^۱، یک موضوع مهم و چالش برانگیز در بخش شناسایی الگو و بینایی کامپیوتر است که برای چندین دهه به طور گسترده مورد مطالعه قرار گرفته است [۱،۲]. تصاویر چهره را می توان به سرعت و بدون تماس فیزیکی به دست آورد. از این رو، در مقایسه با سایر بیومتریک‌های محبوب مانند اثر انگشت و عنبیه چشم، تشخیص چهره کاربردی تر است. بنابراین تشخیص چهره در زمینه‌های مختلف مانند کنترل دسترسی، نظارت و سلامتی کاربرد دارد [3].

با گسترش شبکه‌های کانولوشنی^۲، پیشرفت‌های قابل توجهی در تشخیص چهره رخ داده است. بطوری که امروزه دقت سیستم‌های تشخیص چهره تحت شرایط کنترل شده، نزدیک به دقت بازشناسایی انسان است [1]. بر خلاف سیستم عصبی انسان که در شرایط کنترل نشده مانند محیطی با نور کم، تنوع ژست و علی‌الخصوص انسداد^۳، قابلیت بازیابی بسیار خوبی دارد، عملکرد سیستم‌های تشخیص چهره به طور قابل توجهی کاهش می‌یابد. بنابراین امروزه، تمرکز تحقیقات بر روی شرایط کنترل نشده است.

دانش قبلی در مورد قسمت انسدادی وجود ندارد و این انسداد می تواند در هر کجای تصویر و به هر اندازه‌ای باشد. از این رو، انسداد چهره، یکی از سخت‌ترین مشکلات در حوزه تشخیص چهره است. انسداد صورت زمانی اتفاق می‌افتد که روسری، ماسک صورت، عینک، کلاه و غیره استفاده شود. قرار گرفتن اشیاء تصادفی در جلوی صورت نیز موجب انسداد صورت می‌شود. شکل ۱-۱ تصاویر مختلفی از انواع مختلف انسداد را نمایش می‌دهد. همه‌گیری COVID-19 یک بحران غیرمنتظره برای کل دنیا بود که منجر به تعداد زیادی تلفات و مشکلات امنیتی شد. به منظور کاهش شیوع ویروس کرونا، مردم اغلب برای محافظت از خود از ماسک استفاده می‌کنند. این امر سیستم‌های تشخیص چهره را به دلیل وجود حدود ۵۰ درصد انسداد در

¹ Face Recognition

² Convolutional Neural Networks(Cnn)

³ Occlusion

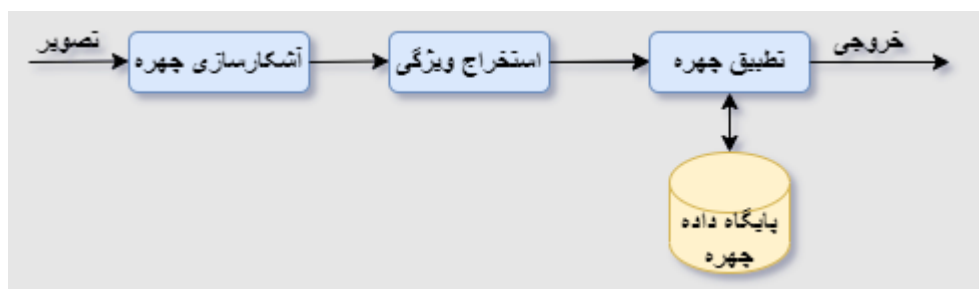
چهره، دچار مشکل می‌کند. بنابراین تمرکز اصلی محققان در طول همه‌گیری ویروس کرونا، ارائه پیشنهادهایی برای رسیدگی به این مشکل از طریق راه‌حل‌های سریع و کارآمد است [۴].



شکل ۱-۱ انواع مختلف انسداد چهره.

۱-۲ بیان مسئله و تعاریف

در حالت کلی، تشخیص چهره از سه قسمت اصلی تشکیل شده است. در شکل (۱-۲) این مراحل نشان داده شده‌اند. این مراحل شامل، آشکارسازی چهره، استخراج ویژگی و تطبیق چهره از چهره آشکارسازی شده و تشخیص چهره هستند که در ادامه هر یک شرح داده شده‌اند [۱۱].



شکل ۱-۲ مراحل تشخیص چهره در حالت کلی.

آشکارسازی چهره^۱ یکی از مهمترین مراحل است که در وهله اول در مسیر تشخیص چهره باید انجام شود. برای اطمینان از عملکرد خوب آشکارسازی چهره، این الگوریتم‌ها باید نسبت به تنوع ژست، تغییرات نور و تفاوت مقیاس کارآمد باشند. در سال‌های اخیر، یادگیری عمیق عملکرد آشکارسازی چهره را نیز

^۱ Face Detection

افزایش داده که در نهایت منجر به استفاده از این روش‌ها در صنعت شده است [۵].

در مرحله استخراج ویژگی^۱، از ویژگی‌هایی با قدرت تمایز بالای چهره استفاده می‌شود. دو روش کلی برای استخراج ویژگی وجود دارد که شامل استخراج ویژگی دستی و استخراج ویژگی به صورت خودکار وجود دارد. در روش اول، ویژگی‌های چهره با پیشنهاد انسان استخراج می‌شود که این ویژگی‌ها، شامل ویژگی‌های هندسی، بافت و شکل هستند. در روش استخراج ویژگی خودکار، ویژگی‌هایی با قدرت تمایز بالا در چهره به کمک شبکه‌های عصبی خودکار به خصوص شبکه‌های عصبی کانولوشنی و مجموعه عظیمی از داده‌های آموزش، استخراج می‌شوند.

در مرحله آخر یعنی تطبیق چهره^۲، از الگوریتم‌های طبقه‌بند برای مقایسه تصویر با پایگاه داده تصاویر کمک گرفته می‌شود و منجر به بدست آوردن هویت فرد می‌شود.

در تشخیص چهره با ماسک علاوه بر مراحل ذکر شده باید از تکنیک‌ها و روش‌های پیشرفته‌تر و به روزتری استفاده شود تا بتوان اثر کاهش ویژگی‌های تصویر را کم کرد و دقت تشخیص بالاتری به دست آورد. این روش‌ها به دو دسته‌ی روش‌های بازنمایی^۳ و بازسازی^۴ تقسیم می‌شوند [۶].

روش‌های بازنمایی سعی در استخراج ویژگی‌هایی دارند که در عین حفظ قابلیت تمایز، کمتر تحت تاثیر نواحی پوشیده‌شده قرار گیرند. این امر، از طریق ایجاد شبکه‌ای مقاوم نسبت به انسداد انجام می‌شود. این روش‌ها بر روی انسدادهایی که نواحی زیادی از تصویر را می‌پوشانند، از دقت تشخیص پایین رنج می‌برند. اما برای انسدادهای جزئی مناسب هستند. برخلاف روش بازنمایی، روش بازسازی سعی در ترمیم تصویر به منظور تولید تصویر بدون ماسک، از تصویر با ماسک دارد. در فصل بعدی به طور مفصل کارهای انجام شده در این حیطة، مورد بررسی قرار می‌گیرد.

¹ Feature Extraction

² Face Matching

³ Representation

⁴ Reconstruction

۱-۳ هدف تحقیق

تکنیک‌های ترمیم تصاویر^۱ وظیفه بازسازی نواحی از دست رفته را دارند، به طوری که تغییرات از نظر بصری واقعی و از نظر معنایی صحیح باشد. برای به دست آوردن دقیق تصویر چهره‌ی کامل از تصویر ماسک‌دار می‌توان از این روش‌ها استفاده کرد. امروزه با گسترش شبکه‌های عصبی عمیق، برای ترمیم تصاویر از این شبکه‌ها استفاده می‌شود. شبکه‌های GAN به دلیل توانایی بالایی که در تولید تصاویر واقعی و استخراج ویژگی‌های سطح بالا دارند، به پیشرفت قابل توجهی در این زمینه دست یافته‌اند. هدف کلی این رساله، ارائه یک روش کاملاً خودکار است که در آن با استفاده از شبکه مبتنی بر GAN کار ترمیم تصویر چهره‌ی ماسک‌دار انجام شود و در نهایت تصویر بدون ماسک به گونه‌ای تولید شود که دقت تشخیص چهره افزایش یابد.

۱-۴ چالش‌ها و راه‌حل‌های آن

استفاده از ماسک صورت باعث مشکلاتی می‌شود که در ادامه به آن‌ها اشاره شده است.

- کلاهبرداران و سارقان با سوء استفاده از نقاب، سرقت و ارتکاب جنایت را بدون شناسایی انجام می‌دهند.
- کنترل دسترسی جامعه و احراز هویت چهره زمانی که بخش بزرگی از صورت توسط ماسک پنهان شده، دشوار است.
- برای نرمال‌سازی صورت و تصحیح ژست از بینی استفاده می‌شود. از این‌رو، در معرض دید قرار دادن ناحیه بینی در تشخیص چهره بسیار مهم است [۷]. استفاده از ماسک، ناحیه بینی را پوشش داده و تشخیص چهره دشوار می‌شود.

¹ Image Inpainting

با توجه به این مشکلات، ماسک صورت به طور قابل توجهی روش‌های تشخیص چهره موجود را به چالش کشیده‌اند. روش ترمیم تصویر با شبکه GAN می‌تواند تا حد زیادی مشکل تشخیص چهره دارای ماسک را حل کند، اما با چالش‌هایی همراه است که در ادامه به آن‌ها اشاره شده است.

- دادگان مناسب، برای بخش‌بندی^۱ تصویر دارای ماسک وجود ندارد.
- در آموزش شبکه‌های مبتنی بر GAN چالش‌هایی نظیر، عدم همگرا شدن شبکه، تولید نمونه‌های محدود توسط مولد^۲، عدم یادگیری توسط مولد به دلیل عملکرد عالی شبکه تمایزگر^۳ و بیش‌برازش^۴ شبکه وجود دارد.
- دستیابی به شبکه GAN در شرایطی که حالت پایدار داشته باشد و نتایج خوبی را به عنوان خروجی ارائه دهد، کمی دشوار است. لذا در این پژوهش سعی شده است با بکارگیری تکنیک‌های موثر زیر، ترمیم تصاویر چهره انجام شود.
- استفاده از لایه‌های کانولوشنی متنوع، مانند کانولوشن گیتی^۵، لایه فراخش^۶ به منظور تولید تصاویر با کیفیت.
- علاوه بر شبکه متمایزکننده، از شبکه بینشی^۷ که در واقع ویژگی‌های استخراج شده از شبکه VGG است نیز استفاده شده تا علاوه بر کمک به بازسازی بهتر، باعث پایداری شبکه شود.
- بکارگیری توابع خطا متنوع مانند خطای L1، MSE و SSIM.

۱-۵ نوآوری‌های تحقیق

به طور کل، نوآوری‌های این پایان‌نامه به شرح ذیل هستند:

¹ Segmentation
² Generator
³ Discriminator
⁴ Overfitting
⁵ Gate Convolution
⁶ Dilation
⁷ Perceptual

۱. مجموعه دادگان تصویر با ماسک که ناحیه ماسک بخش بندی شده باشد تا بحال وجود نداشته، از این رو ابتدا حدود ۱۲۰۰۰ تصاویر با ماسک جمع آوری شد و سپس با صرف زمان زیاد، بر روی ۲۸۵۰ تصویر، به صورت دستی بخش بندی انجام شد و سپس به کمک روش پیشنهادی، بخش بندی با ماسک به صورت خودکار انجام شد.

۲. ارائه روش بخش بندی ناحیه ماسک، به عنوان مرحله پیش پردازش شبکه مبتنی بر GAN.

۳. استفاده از تابع خطای ترکیبی جدید در شبکه پیشنهادی بخش بندی با ماسک.

۴. ارائه شبکه عمیق مبتنی بر GAN به منظور افزایش دقت تشخیص چهره با ماسک.

۱-۶ ساختار پایان نامه

این پایان نامه در پنج فصل تدوین شده است. در فصل دوم، مروری کلی بر سیستم های تشخیص چهره، علی الخصوص تشخیص چهره های دارای انسداد و روش های مختلف ترمیم تصویر انجام شده است. در فصل سوم، روش پیشنهادی برای تشخیص چهره دارای ماسک و در نهایت، مدل جدیدی برای آموزش سیستم تشخیص چهره با ماسک ارائه شده است. در فصل چهارم، نتایج آزمایش های انجام شده در این پایان نامه ارائه شده است. فصل پنجم، به جمع بندی و نتیجه گیری اختصاص داده شده است.

فصل ۲: ادبیات موضوع و کارهای پیشین

۲-۱ مقدمه

تجزیه و تحلیل چهره یک مشکل چالش برانگیز در بینایی کامپیوتر است و بیش از سه دهه است که به طور فعال مورد تحقیق قرار گرفته است. هدف این است که تا حد امکان اطلاعاتی مانند هویت، ژست، جنسیت، سن، احساسات و غیره از یک چهره استخراج شود. انسداد، یک محدودیت کلیدی در روش‌های تشخیص چهره است. به طور کلی، استفاده از کلاه، عینک، ماسک و همچنین هر چیز دیگری که قسمتی از صورت را مسدود کند، باعث انسداد تصویر می‌شود. بنابراین، پوشیدن ماسک سخت‌ترین چالش انسداد صورت در نظر گرفته می‌شود، زیرا بخش بزرگی از صورت از جمله بینی را مسدود می‌کند.

در این فصل، در ابتدا تاریخچه سیستم‌های تشخیص چهره را مرور کرده و سپس به ساختار سیستم‌های تشخیص چهره می‌پردازیم. در قسمت بعد، روش‌های تشخیص چهره بررسی می‌شود. در مرحله بعد، تشخیص چهره تحت انسداد^۱ بررسی و به روش‌های غلبه بر آن به طور مفصل پرداخته می‌شود. در قسمت آخر، به جمع‌بندی خلاصه‌ای از مطالب گفته شده در این فصل خواهیم پرداخت.

۲-۲ تاریخچه سیستم‌های تشخیص چهره

در این بخش مهم‌ترین مراحل تاریخی را که به پیشرفت فناوری تشخیص چهره کمک کرده است را مرور می‌کنیم:

- در سال ۱۹۶۴، اولین تلاش‌ها برای ساختن برنامه کامپیوتری تشخیص چهره توسط گروهی از محققان آمریکایی زده شد. آن‌ها سعی داشتند با اندازه‌گیری مختصات نواحی مختلف صورت، مانند اندازه چشم، دهان، مرکز مردمک چشم و چندین ویژگی دیگر، برنامه تشخیص چهره ایجاد کنند. تعداد این ویژگی‌ها به ۲۰ ویژگی می‌رسید.
- در سال ۱۹۷۷، محققان با اضافه کردن ۲۱ ویژگی دیگر به سیستم طراحی شده سعی در

^۱ Occluded Face Recognition

پیشرفت آن داشتند.

- در سال ۱۹۸۸، هوش مصنوعی برای توسعه ابزارهای نظری که قبلاً مورد استفاده قرار می‌گرفتند معرفی شد، که ضعف‌های بسیاری را نشان داد. از ریاضیات ("جبر خطی") برای تفسیر متفاوت تصاویر و یافتن راهی برای ساده‌سازی و دستکاری آن‌ها، مستقل از نشانه گذاری انسانی استفاده شد.
- در سال ۱۹۹۱، الکس پنتلند و متیو ترک^۱ از دانشگاه MIT اولین نمونه موفق فناوری تشخیص چهره، Eigenfaces [۹] را ارائه کردند، که از روش تجزیه و تحلیل مولفه‌های اصلی آماری (PCA) استفاده می‌کند.
- در سال ۱۹۹۸، آژانس پروژه‌های تحقیقاتی پیشرفته دفاعی برای تشویق صنعت و دانشگاه‌ها جهت تحقیق و توسعه در حوزه تشخیص چهره، یک پایگاه داده بزرگ و چالش‌برانگیز، متشکل از ۲۴۰۰ تصویر از ۸۵۰ نفر را در اختیار جهان قرار داد.
- در سال ۲۰۰۵، مسابقه بزرگ تشخیص چهره (FRGC) برای تشویق و توسعه فناوری تشخیص چهره برگزار شد، که در نهایت باعث حمایت، از ابتکارات تشخیص چهره شد.
- در سال ۲۰۱۱، همه چیز به دلیل روش یادگیری ماشین مبتنی بر شبکه‌های عصبی مصنوعی، که امروزه از آن به عنوان یادگیری عمیق یاد می‌کنیم، به یکباره شتاب می‌گیرد و محققان به این نتیجه می‌رسند که هر چه سیستم، تصاویر بیش‌تری ببیند، بهتر یاد می‌گیرد.
- در سال ۲۰۱۴، شرکت فیسبوک از الگوریتم Deepface [۱۰]، رونمایی می‌کند. این الگوریتم دقتی نزدیک به دقت بازشناسایی انسان یعنی حدود ۹۷ درصد داشت.

¹ Alex Pentland And Matthew Turk

۲-۳ ساختار سیستم‌های تشخیص چهره

همان طور که در شکل (۲-۱) از فصل قبل مشاهده شد، سیستم‌های تشخیص چهره از سه قسمت اصلی تشکیل شده اند [۱۱]. در مرحله اول، موقعیت صورت در تصویر یا ویدیو مشخص می‌شود، سپس نقاط عطف^۱ چهره، به منظور تراز کردن چهره^۲، استخراج شده و در مرحله بعد، از این تصویر استفاده می‌شود. در این پژوهش روش‌های آشکارسازی تصویر بررسی نمی‌شود.

استخراج ویژگی یک مرحله بسیار مهم در تشخیص چهره است و نرخ تشخیص سیستم به داده‌های معنی‌دار استخراج شده از تصویر چهره بستگی دارد. ویژگی‌های استخراج شده، باید قدرت تمایز بالایی داشته باشند. تکنیک‌ها و الگوریتم‌های زیادی برای استخراج ویژگی در تشخیص چهره استفاده می‌شود و به دو روش تقسیم می‌شوند:

• روش‌های سنتی

ویژگی‌های دستی^۳، از نواحی مشخص و تعریف شده صورت استخراج می‌شوند، که نیازی به بهینه‌سازی یا مرحله یادگیری ندارند. به طور کلی این روش‌ها نیاز به تراز بودن صورت، براساس مختصات چشم‌ها دارند. در ادامه به برخی از این روش‌ها اشاره می‌کنیم.

الگوهای باینری محلی^۴ [۱۲]، برای استخراج ویژگی‌های کارآمد از صورت استفاده می‌شود و به طور گسترده در برنامه‌های کاربردی مختلف استفاده شده است. ایده اصلی این است که تصویر چهره را به چندین منطقه تقسیم کنیم تا توزیع ویژگی‌های LBP به طور مستقل از آن استخراج شود. سپس این توصیف‌گرها به هم متصل می‌شوند تا یک توصیف‌گر کلی^۵ از چهره را تشکیل دهند.

¹ Landmark

² Face Alignment

³ Handcraft Features

⁴ Local Binary Pattern (Lbp)

⁵ Global Descriptor

روش SIFT^۱ [۱۳]، در تشخیص اشیا، محبوب است، همچنین می تواند برای تشخیص چهره نیز اعمال شود. این روش تا حد زیادی نسبت به تغییرات در مقیاس و چرخش ثابت است و همچنین کمتر تحت تأثیر تغییرات روشنایی قرار می گیرد.

• ویژگی های مبتنی بر یادگیری

در مقایسه با ویژگی هایی که از روش های سنتی بدست می آید، این روش ها انعطاف پذیرتر هستند. ویژگی هایی که از داده های آموزشی به دست می آیند می توانند مؤثر باشند و قدرت تمایز بالقوه بالایی برای تشخیص چهره داشته باشند.

رویکردهایی مانند تجزیه و تحلیل مؤلفه اصلی^۲ [۹] و تجزیه و تحلیل تفکیک خطی^۳ [۱۴] دو روش معروف در یادگیری فضای فرعی^۴، برای استخراج ویژگی هستند.

معروف ترین و پر استفاده ترین روش یادگیری، که قدرت تمایز بسیار بالاتری نسبت به بقیه روش ها دارد، شبکه عصبی عمیق است. عملکرد تشخیص چهره به دلیل معماری های پیشرفته شبکه های کانولوشنی در چند سال اخیر افزایش یافته است. برای استخراج ویژگی ها، شبکه های عمیق مختلفی مانند VGGNet، ResNet، GoogleNet و... معرفی شدند. همچنین در برخی مقالات، با ترکیب این شبکه ها نتایج بهتری دست یافتند [۱۲۰].

در مرحله آخر فرآیند تشخیص چهره، تصویر شخصی را داریم و می خواهیم بررسی کنیم که آیا این فرد در پایگاه داده افراد وجود دارد و یا خیر. در مرحله قبل ویژگی افراد در پایگاه داده را به دست آوردیم. یعنی به طور مثال اگر از شبکه عمیق استفاده کردیم، شبکه با پایگاه داده افراد آموزش دیده است.

¹ The Scale Invariant Feature Transform

² Principal Component Analysis (PCA)

³ Linear Discriminant Analysis (LDA)

⁴ Subspace

برخلاف طبقه بندی اشیا، در تشخیص چهره، امکان دارد هویت افراد در هنگام تست از داده‌های آموزشی متفاوت باشند که باعث می‌شود طبقه‌بندی‌کننده آموخته شده برای تشخیص چهره‌ها در زمان تست مورد استفاده قرار نگیرد. بنابراین، الگوریتم‌های تطبیق چهره یک بخش اساسی در FR هستند.

۲-۴ روش‌های تشخیص چهره

یک سیستم تشخیص چهره بر روی تصاویر و یا ویدیوهای بدست آمده از سیستم‌های نظارتی، دوربین مدار بسته و یا سخت افزارهای روزمره عمل می‌کند. در یک سیستم تمام اتوماتیک تشخیص چهره، مراحل به صورت زیر است:

ابتدا صورت باید در تصویر شناسایی شود، سپس تصویر صورت تراز^۱ می‌شود. حال تصویر مناسب برای استخراج ویژگی و طبقه‌بندی آماده شده که در نهایت تشخیص هویت فرد انجام می‌شود. همانطور که در شکل (۲-۱) مشاهده می‌شود، براساس ماهیت روش‌های استخراج و طبقه بندی بکار رفته، روش‌های تشخیص چهره دو بعدی را به چهار زیر کلاس مختلف تقسیم می‌کنیم که عبارتند از: (۱) روش‌های کل نگر^۲، (۲) روش‌های محلی (هندسی)، (۳) روش‌های مبتنی بر توصیف‌کننده‌های بافت محلی^۳ و (۴) روش‌های مبتنی بر یادگیری عمیق.

۲-۴-۱ کارهای مرتبط با روش‌های کل نگر

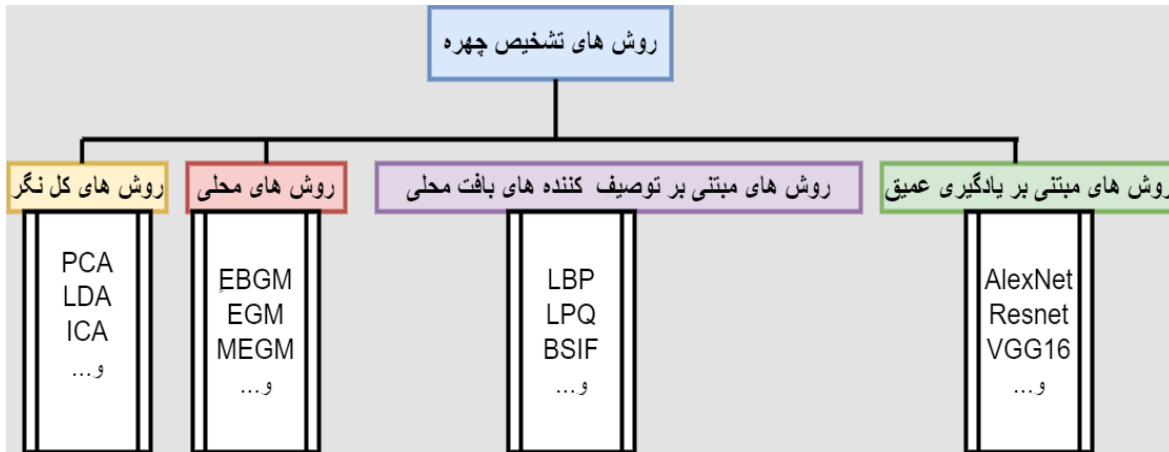
رویکردهای کلی‌نگر برای تشخیص چهره، اطلاعات کلی از مجموعه داده‌گان چهره را برای انجام تشخیص و تأیید چهره در نظر می‌گیرد. اطلاعات جهانی^۴ در درجه اول با تعداد بسیار کمی از ویژگی‌ها نشان داده می‌شود که مستقیماً از پیکسل‌های تصاویر چهره مشتق می‌شوند. این ویژگی‌ها وظیفه شناسایی و نمایش

^۲ Holistic Methods

^۳ Local Texture Descriptors-Based Methods

^۴ Global Information

تفاوت‌های بین تصاویر مختلف چهره را دارند و از این رو به طور منحصربه فرد قابلیت شناسایی سوژه را دارند [۱۵]. در ادامه به برخی از شناخته شده ترین روش‌های این دسته اشاره می‌کنیم.



شکل ۱-۲ روش‌های تشخیص چهره

روش eigenface [۹]، پایه بیش‌تر روش‌هایی است که در این بخش مطرح می‌شوند. دو جز اصلی آن، تجزیه و تحلیل مؤلفه‌های اصلی (PCA) و Eigenvector است که جزء اول، صورت را به بردارهای مشخصه به شکل ماتریس کوواریانس تقسیم می‌کند. سپس از این بردارها برای محاسبه تغییرات بین چند چهره استفاده می‌شود. چهره‌ها با ترکیب خطی بالاترین مقادیر ویژه^۱ مشخص می‌شوند. در مرجع [۱۵]، فضای صورت با استفاده از یک پایگاه داده خصوصی، حاوی ۲۵۰۰ تصویر از ۱۶ نفر، نویسندگان به نرخ طبقه بندی صحیح^۲ ۹۶، ۸۵ و ۶۴ درصد، تحت تغییرات در جهت‌گیری، اندازه و روشنایی رسیدند. در شرایط نامناسب نوری، این روش دقت بسیار پایینی دارد. از این رو نرمال‌سازی روشنایی بسیار ضروری است.

برای غلبه بر مشکل کاهش عملکرد به دلیل تغییرات شرایط نوری، ژائو و یانگ [۱۶]، روشی را برای محاسبه ماتریس کوواریانس با استفاده از سه تصویر، که هر یک در شرایط مختلف نوری گرفته شده‌اند، ارائه کردند. این روش نتوانست به طور کامل مشکل روشنایی را حل کند.

^۱ Eigenvalues

^۲ Correct Classification Rate(Ccr)

در سال ۱۹۹۹ محققانی که روش Eigenface را مطرح کردند، کار اولیه خود را گسترش و روش Eigenfeature [۱۷]، را ابداع کردند. در این روش از اجزا مختلف صورت، مانند بینی، دهان و چشم‌ها استفاده کردند. به این صورت که، یک EigenSpace متشکل از ویژگی‌های صورت، به طور مثال EigenNose، EigenEyes و EigenMouth ساخته، یعنی برای هر کدام، بردار ویژگی منحصر به فرد ایجاد شد. در مقایسه با روش اصلی EigenFace، این روش توسعه یافته حساسیت کمتری به تغییرات ظاهری نشان داد. در مجموعه داده FERET متشکل از ۷۵۶۲ تصویر از حدود ۳۰۰۰ نفر، نویسندگان به نرخ طبقه بندی ۹۵ درصد دست یافتند. EigenFace در مقایسه با EigenFeatures یک روش ساده، سریع و کاربردی بود، اما با این وجود، در برابر تغییرات در شرایط روشنایی و مقیاس، پایداری نداشت.

مرجع [۱۸]، سیستمی را پیشنهاد کرد که نسبت به نورهای مختلف و تغییرات حالت چهره غیر حساس باشد. آن‌ها به جای PCA از روش LDA برای کاهش ابعاد بردار ویژگی‌ها استفاده کردند. در این روش با هدف به حداکثر رساندن قدرت تمایز بین کلاس‌ها در فرآیند آموزش، تصویر را به صورت خطی در یک زیرفضا مستقل بررسی می‌کند. آزمایش‌های مختلف انجام شده بر روی پایگاه‌های اطلاعاتی Harvard و Yale Face نشان داد که FisherFace نسبت به EigenFace نرخ خطای کمتری دارد.

در سال ۲۰۰۲، بارلت و همکارانش [۱۸]، با اشاره به این‌که، روش پایه تجزیه و تحلیل مؤلفه‌های اصلی، تنها به روابط زوجی بین پیکسل‌ها در مجموعه داده تصاویر وابسته هستند و برای استخراج ویژگی‌ها، اطلاعات حیاتی تصویر در بردار ویژگی یافت می‌شود، سعی در استفاده از روش تجزیه و تحلیل مؤلفه‌های مستقل^۱ [۲۰]، در حوزه تشخیص چهره داشتند. این روش تعمیم روش تجزیه و تحلیل مؤلفه‌های اصلی است. آن‌ها دو معماری مختلف با پایگاه داده FERET برای آزمایش عملکرد روش تجزیه و تحلیل مؤلفه‌های مستقل پیاده سازی کردند. اولی تصاویر را به عنوان متغیرهای تصادفی و پیکسل‌ها را به عنوان نتیجه پردازش می‌کند، در حالی که دومی پیکسل‌ها را به عنوان متغیرهای تصادفی و تصاویر را به عنوان نتیجه

^۱ Independent Component Analysis (Ica)

پردازش می‌کند. نتایج هر دو معماری تجزیه و تحلیل مؤلفه‌های مستقل از روش تجزیه و تحلیل مؤلفه‌های اصلی بهتر بود. علاوه بر این، بهترین عملکرد با ادغام هر دو معماری ICA به دست آمد.

فیلتر گابور، یک فیلتر خطی است که برای تحلیل بافت استفاده می‌شود، به این معنی که اساساً تحلیل می‌کند که آیا محتوای فرکانس خاص در تصویر در جهت خاص در یک منطقه محلی در اطراف نقطه یا منطقه تجزیه و تحلیل وجود دارد. بسیاری از دانشمندان دیدگاه معاصر ادعا می‌کنند که فرکانس و جهت گیری نمایش‌های فیلترهای گابور شبیه به سیستم بصری انسان می‌باشند، هرچند هیچ شواهد تجربی و هیچ منطقی عملی برای حمایت از این ایده وجود ندارد. فیلترهای گابور برای نمایش و تمییز بافت مناسب هستند. همچنین در حوزه فضایی، یک فیلتر گابور دوبعدی، یک تابع هسته گاوسی است که توسط یک موج مسطح سینوسی مدولاسیون شده است [۲۱]. در سال ۲۰۱۵، در مرجع [۲۲]، روشی مبتنی بر فیلتر گابور، برای استخراج ویژگی‌ها و افزایش عملکرد سیستم‌های تشخیص چهره پیشنهاد کرد. این فیلترها برای ثبت ویژگی‌های صورت تراز در زوایای خاص استفاده می‌شوند. روش پیشنهادی تحت چندین پایگاه داده مانند ORL و FERET مورد آزمایش قرار گرفت و نتایج خوبی در برابر تغییرات وضعیت نور نشان داد. تبدیل کسینوس گسسته^۱ [۲۳] و تبدیل موجک گسسته^۲ [۲۴، ۲۵]، تکنیک‌های خطی دیگری هستند که برای تجزیه و تحلیل چهره به کار گرفته شده‌اند. هر دو روش عمدتاً برای فشرده سازی تصویر [۲۶] و انتخاب ویژگی استفاده می‌شود. مرجع [۲۷]، یک الگوریتم استخراج ویژگی مبتنی بر دو روش گفته شده در بالا، برای تشخیص چهره پیشنهاد کرد، که تصویر صورت با استفاده از تبدیل موجک گسسته دوبعدی تجزیه می‌شود و سپس تبدیل کسینوس گسسته، برای تقریب تصویر مرحله قبل، به کار برده می‌شود. در نهایت، ضرایب تبدیل کسینوس گسسته برای تطبیق استفاده می‌شود. نتایج تجربی پایگاه داده ORL برتری این الگوریتم را نسبت به روش تجزیه و تحلیل مؤلفه‌های اصلی نشان داد.

به طور خلاصه، قبل از معرفی شبکه‌های عمیق، روش‌های کل نگر در اجرای سیستم‌های تشخیص چهره

¹ Discrete Cosine Transform (Dct)

² Discrete Wavelet Transform (Dwt)

رایج بودند. با این وجود، آنها در برابر تغییرات زمینه و ناهماهنگی‌ها بسیار حساس هستند. به همین دلیل، صورت در اکثر موارد باید به صورت دستی از تصویر برش داده شود.

۲-۴-۲ کارهای مرتبط با روش‌های محلی (هندسی)

توجه و تثبیت نقش مهمی در تشخیص چهره انسان دارد. فرآیندهای توجه معمولاً توسط ویژگی‌های شاخص بومی سازی شده^۱، در نظر گرفته می‌شود. از شاخص‌ها می‌توان در تشخیص چهره استفاده کرد. در تصویر، نواحی مختلف صورت، اطلاعات یکسانی را ارائه نمی‌دهند. به عنوان مثال، پیشانی و گونه‌ها در مقایسه با بینی یا چشم‌ها ساختاری صاف و الگوهای متمایز کمتری دارند. از نقاط شاخص‌ساز برای تشخیص چهره و تشخیص احساس بر اساس توزیع هندسی می‌توان استفاده کرد.

بوکشتاین [۲۸]، نقاط عطف را چنین توصیف کرد: «به شکلی اشاره می‌کند که مشابه‌های بیولوژیکی آن

در مجموعه‌ای از داده‌ها، که از نظر عینی قابل تولید هستند و در تمام اشکال دیگر رخ می‌دهند.»

متداول‌ترین نشانه‌های مورد استفاده در صورت عبارتند از: نوک‌بینی، نوک‌چشم‌ها، نوک‌گوشه‌های دهان، ابروها، وسط‌عنیه، بالای گوش، سوراخ‌های بینی و بینی. لازم به ذکر است که در برخی مواقع به نواحی متمایزکننده صورت، مانند چشم‌ها یا دهان، "ویژگی‌های صورت" نیز گفته می‌شود. گاهی این اصطلاحات منجر به ابهام می‌شود.

در واقع، در تشخیص الگو، اصطلاح ویژگی اغلب برای مشخص کردن یک نمایش خاص استخراج شده از یک الگو به کار می‌رود. به عنوان مثال، بردارهای EigenFace ویژگی نیز نامیده می‌شوند. علاوه بر این، نمایش‌های عددی جمع‌آوری‌شده توسط فیلتر گابور چند کانالی که در یک تصویر سطح خاکستری پیاده‌سازی شده‌اند نیز ویژگی نامیده می‌شوند. برای این منظور، الگوهایی که از موقعیت‌های خاص و متمایز به دست می‌آیند، در این بخش به‌عنوان نقطه‌های شاخص شناخته می‌شوند.

¹ Landmark Characteristics Localized

توزیع نقطه‌های شاخص، در روش‌های مبتنی بر هندسه، برای پیدا کردن فواصل، زوایا و مناطق، استفاده می‌شود [۳۰،۲۹]. برای مثال، در الگوریتم تطبیق گراف دسته‌ای الاستیک^۱ [۳۱]، یک نمودار، موقعیت‌های مربوط به شاخص را مدل‌سازی می‌کند، به این صورت که هر گره یک نقطه از صورت را نشان می‌دهد و یال‌ها مطابق با فواصل، وزن‌دهی می‌شوند (شکل ۲-۲). از یک سری مدل برای تعیین شباهت مشخصه محلی برای هر گره استفاده می‌شود. اگرچه تغییر شکل‌های احتمالی اغلب به نقطه‌های شاخص بستگی دارد. به عنوان مثال، گوشه‌های دهان خیلی بیشتر از نوک بینی تغییر شکل می‌دهند، لذا اطلاعات شاخص‌ها را می‌توان در مدل ساختاری متصل کرد [۳۱].



شکل ۲-۲ نمونه‌ای از استخراج نشانه‌ها با استفاده از الگوریتم تطبیق گراف دسته‌ای الاستیک روش‌های مختلفی برای استخراج نمایش‌های چهره از چندین مؤلفه یا تصویر فرعی صورت پیشنهاد شد. در سال ۱۹۹۴ پنتلند و همکاران [۳۳] روش، تجزیه و تحلیل مؤلفه‌های اصلی را بر اساس اجزایی پیشنهاد کرد که زیرفضای صورت آن شامل برخی از زیرفضاهای ساخته‌شده از تصاویر جزئی از تصاویر اولیه صورت بود. نقاط عطف نیز بین دهان و چشم انتخاب می‌شود.

یکی از اشکال رایج تشخیص چهره بر اساس نشانه‌ها، تطبیق نمودار الاستیک^۲ است [۳۵،۳۴]. این روش یک پیاده‌سازی واقع‌بینانه از ساخت گراف پویا برای شناسایی شی است که با قرار دادن یک نمودار پراکنده، الاستیک و مستطیل شکل بر روی تصویر شی و تعیین موجک گابور به هر گره، نمودار ایجاد می‌شود.

¹ Elastic Bunch Graph Matching (Ebgm)

² Elastic Graph Matching (Egm)

بهینه‌سازی تصادفی^۱ تابع خطا^۲، از روی شباهت بین یال‌ها و تغییر شکل گره‌ها محاسبه می‌شود که از این مرحله به عنوان فرآیند تطبیق گراف^۳ نیز یاد می‌شود. در سال ۱۹۹۳، لادز و همکاران [۳۶]، نتایج هیجان‌انگیزی را بر روی یک مجموعه داده خصوصی از ۸۷ نفر، تحت تغییرات حالت چهره و چرخش تصویر صورت، گزارش کرد.

یکی دیگر از اشکال رایج تشخیص چهره بر اساس نشانه‌ها، تطبیق گراف دسته‌ای الاستیک است [۳۸،۳۷]. این روش توسعه‌ای از روش تطبیق نمودار الاستیک است. در ساختار گراف پشته‌ای^۴ این روش، مجموعه‌ای از یال‌ها برای نمونه‌های مختلفی از یک چهره در هر گره (به عنوان مثال، با دهان و چشم باز یا بسته) محاسبه می‌شود. بنابراین نمایش نمودار پشته می‌تواند چندین تغییر در ظاهر صورت را کنترل کند. در سال ۲۰۲۰، یک سیستم تشخیص چهره پیشنهاد شد که از یک توصیف‌گر^۵ جدید به نام ساختار نمودار محلی متراکم^۶ استفاده می‌کرد. علاوه بر این، توصیف‌گر از درون یابی دوخطی برای بهبود تراکم پیکسلی، هنگام تولید گراف تصویر استفاده می‌کند. این الگوریتم در پایگاه داده LFW و ORL به نتایج خوبی دست یافت [۳۹].

به طور خلاصه، عیب عمده همه روش‌های مبتنی بر هندسه این است که باید شامل تصاویر کاملاً هم‌تراز چهره باشند، تا تمام نقاط صورت، مانند بینی، چشم‌ها و دهان، در بردار ویژگی، در مکان مربوطه نمایش داده شوند. لذا، تصاویر صورت اغلب به صورت دستی برای این منظور مرتب می‌شوند. الگوریتم تطبیق گراف الاستیک نسبت به بقیه روش‌های هندسی که در این بخش اشاره شد، برای عملکرد خوب به تراز دقیق نیاز ندارد. هرچند به زمان زیادی برای بررسی و تطبیق نیاز دارد.

¹ Stochastic Optimization

² Loss Function

³ Graph Matching

⁴ Heap Graph

⁵ Descriptor

⁶ Dense Local Graph Structure (D-Lgs)

۲-۴-۳ روش‌های مبتنی بر توصیف‌کننده‌های بافت محلی

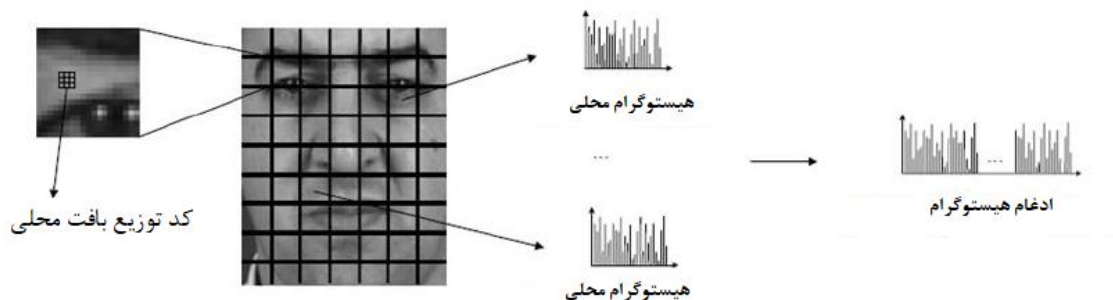
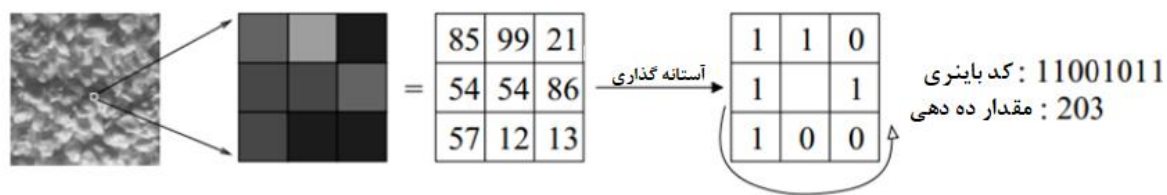
روش‌های استخراج ویژگی متمرکز بر بافت تصاویر، نقش مهمی در تشخیص الگو و بینایی رایانه ایفا می‌کنند. الگوریتم‌های استخراج بافت پیشنهاد شده در ادبیات می‌توانند به روش‌های آماری و ساختاری تقسیم شود [۴۰]. توصیف‌گرهای بافت محلی توجه بیشتری را به خود جلب کردند و برای بسیاری از کاربردها، مانند طبقه‌بندی بافت، تشخیص چهره یا نمایه‌سازی تصویر^۱ معرفی شدند. همچنین این روش‌ها در برابر نور ضعیف، تنوع در روشنایی مقاوم هستند و نیازی به تقسیم‌بندی ندارند [۴۱]. در توصیف‌گر محلی هدف، تبدیل اطلاعات در سطح پیکسل به شکل مناسبی است که در نهایت محتوای غیر حساس به جنبه‌های مختلف را خروجی دهد. لذا بر خلاف توصیف‌گرهای جهانی که ویژگی‌ها را مستقیماً از کل تصویر محاسبه می‌کنند، توصیف‌گر محلی، تصویر را به قسمت‌های کوچک‌تر تقسیم می‌کند. این کار باعث می‌شود که در شرایط نامحدود^۲ کارآمدتر شود [۴۲].

در سال ۲۰۰۶ آهونن و همکاران [۴۳]، یک نمایش جدید و مؤثر از تصویر صورت را براساس توصیف‌گر بافت محلی به نام: الگوی باینری محلی^۳ ارائه کردند. در این روش، تصویر صورت به قسمت‌های مختلف تقسیم می‌شود، سپس برای هر قسمت توزیع بافت محلی حساب شده و با یک هیستوگرام بهبودیافته، ترکیب می‌شود شکل (۲-۳). از این هیستوگرام به عنوان توصیف‌گر چهره استفاده می‌شود.

¹ Image Indexing

² Unconstrained

³ Local Binary Pattern (Lbp)



شکل ۲-۳ مثالی از محاسبه الگوی باینری محلی صورت.

مرجع [۴۴]، یک روش مولد^۱ برای تأیید چهره، براساس روش الگوی باینری محلی پیشنهاد کردند. آن‌ها به کمک هیستوگرام قسمت‌های مختلف تصویر و توزیع احتمال هر یک از بلاک‌ها، به یک مدل کلی از چهره دست‌یافتند. در مرحله بعد، با استفاده از تکنیک تطبیق حداکثری پسینی^۲، مدل عمومی چهره را ایجاد و در نهایت با تصویر ورودی مقایسه کردند. در نهایت آن‌ها متوجه شدند که روش پیشنهادی تحت هر شرایطی نسبت به روش پایه یعنی الگوی باینری محلی، بهتر عمل می‌کند.

هیستوگرام‌های استخراج‌شده از بلاک‌ها، عمدتاً در نمایش اصلی صورت، پراکنده^۳ هستند. علاوه بر این، بلوک‌های بزرگ هیستوگرام‌های متراکمی ارائه می‌کنند که در نمایش تغییرات صورت مؤثر نیستند. در پژوهش [۴۵]، یک کوانتیزاسیون برداری^۴ را روی هر بلوک اعمال کردند تا یک بردار ویژگی مفید به دست آورند. به عنوان مثال، الگوهای هر بلوک در چندین گروه گروه‌بندی می‌شوند، و صورت توسط مجموعه‌ای از برچسب‌های معتبر الگوی باینری محلی مدل می‌شود و سایر برچسب‌های ناکارآمد در نظر گرفته نمی‌شوند.

¹ Generative

² Maximum A Posteriori (Map) Adaptation

³ Sparse

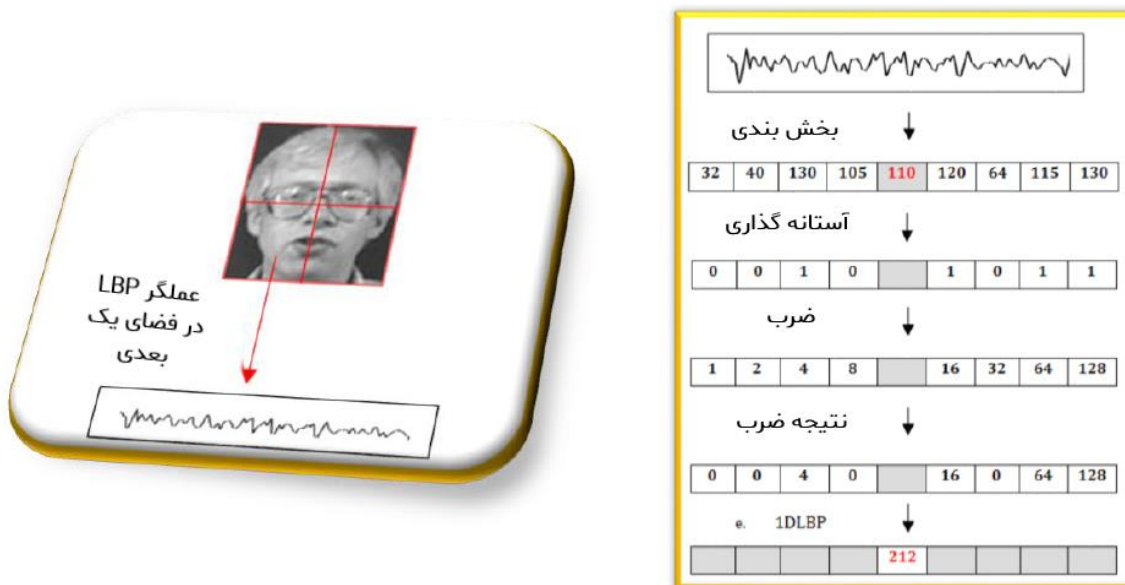
⁴ Vector Quantization (Vq)

در سال ۲۰۱۴-۲۰۱۳، روش الگوی باینری محلی یک بعدی^۱ معرفی شد [۴۶]. این روش یک نمایش جدید از عملگر LBP بود، که الگوها به فضای یک بعدی انتقال^۲ داده می‌شد. این الگوریتم کار خود را در پنج مرحله انجام می‌دهد شکل (۲-۴): ابتدا تصویر ورودی به تعداد بلاک مساوی تقسیم می‌شود. یک انتقال عمودی برای هر بلوک تجزیه شده، در فضای یک بعدی اعمال شده و توصیف‌گر پیشنهادی بر روی هر بلوک پیش‌بینی شده اعمال می‌شود. سپس، بردارهای تولید شده از هر بلوک را به هم متصل کرده تا یک بردار سراسری ایجاد شود. در انتها از روش شناخته شده تجزیه و تحلیل مؤلفه اصلی برای گروه بندی مجدد بردارهای کلی، کاهش ابعاد و حفظ تنها اطلاعات مربوط به هر فرد استفاده شد. برای محاسبه شباهت بین تصاویر صورت از فاصله کای دو^۳ استفاده شد. آن‌ها چندین آزمایش را روی مجموعه داده‌های ORL و AR انجام دادند. آن‌ها به دقت ۹۶.۹، بر روی پایگاه داده AR دست‌یافتند. نویسندگان همچنین کار خود را با اضافه کردن طبقه بندی الگوریتم K-NN و ترکیب بردارهای عمودی و افقی پیش‌بینی شده از هر بلوک گسترش دادند.

¹ One-Dimensional Local Binary Pattern (1dlbp)

² Project

³ Chi-Square



شکل ۴-۲ مثالی از محاسبه صورت با الگوی باینری محلی یک بعدی.

محققان در سال ۲۰۱۲ عملگر کوانتیزه‌سازی فاز محلی^۱ را معرفی کردند [۴۷]. این عملگر برای حل مسئله تشخیص چهره در موقعیت‌های تاریک توسعه داده شد. عملگر LPQ بر اساس کوانتیزاسیون در مناطق محلی فاز تبدیل فوریه است. در روش پیشنهاد آن‌ها، هیستوگرام‌های برچسب LPQ محاسبه شده، و به‌عنوان توصیف‌گر چهره به همان روشی که در روش توصیف چهره LBP استفاده شد. مجموعه دادگان CMU PIE و FRGC، با تارهای گاوسی القایی مصنوعی برای تست این روش استفاده شد. این مرحله حاکی از این بود که این روش تشخیص چهره در محیط‌های تاریک حدود ۵ درصد بهبود داده است.

تشخیص چهره تصاویر با وضوح پایین^۲، یک مشکل چالش‌برانگیز در نظر گرفته می‌شود. لی و همکاران [۴۸]، یک توصیف‌گر بسامد محلی مناسب^۳ برای غلبه بر این مسئله پیشنهاد کرد. مانند توصیف‌گر قبلی، توصیف‌گر پیشنهادی، مبتنی بر اطلاعات فرکانس محلی است که آن را در برابر وضوح پایین و تاریک مقاوم می‌کند. همچنین برخلاف روش قبل، توصیف‌گر، اطلاعات هم‌فاز را به کار می‌گیرد، که این کار باعث

^۱ Local Phase Quantization (Lpq)

^۲ Low-Resolution

^۳ Local Frequency Descriptor (Lfd)

می‌شود اطلاعات بیش‌تری بدست آورد.

Kannala و Rahtu [۴۹]، روشی را برای ساخت توصیفگرهای تصویر محلی پیشنهاد کردند که اطلاعات بافت را به طور کارآمد، رمزگذاری می‌کند، و برای توصیف مناطق تصویر بر اساس هیستوگرام مناسب است. روش پیشنهادی که به نام، ویژگی‌های تصویر آماری باینری شده^۱ شناخته می‌شود، به طور خودکار، با آموزش مجموعه ثابتی از فیلترها از تعداد محدودی از تصاویر اصلی به عنوان جایگزینی برای استفاده از فیلترهای دست‌ساز مانند LBP و LPQ است. برای به دست آوردن شکل آماری معنی‌داری از تصاویر، BSIF از یادگیری به عنوان جایگزینی برای تنظیم دستی استفاده می‌کند و امکان رمزگذاری اطلاعات خاص را با استفاده از کمی‌سازی عنصری فراهم می‌کند. یادگیری، همچنین یک روش راحت و همه‌کاره برای تطبیق اندازه توصیف‌گر و کنترل آن با برنامه‌هایی با ویژگی‌های تصویر غیرعادی ارائه می‌دهد. آن‌ها از FRGC در آزمایشات استفاده کردند. نتایج روش BSIF عملکردی مشابه با روش‌های پیشرفته داشت با این حال، در برخی از تصاویر تراز نشده و تار عملکرد بهتری از خود نشان داد.

به‌طور خلاصه روش‌های مبتنی بر توصیف‌کننده‌های بافت محلی، نسبت به دو روش قبلی، نرخ تشخیص بالاتری دارند و می‌توان به صورت بلادرنگ از این روش‌ها استفاده کرد. علاوه بر این، آن‌ها نسبت به مقیاس^۲ تصویر، و ناهماهنگی پایدارتر هستند. با این حال، این روش‌ها نتوانستند چالش‌های مطرح‌شده مانند تغییر وضعیت^۳، وضوح کم، حالت چهره و شرایط نوری متفاوت، به‌طور قابل قبولی رفع کنند.

۴-۴-۲ روش‌های مبتنی بر شبکه‌های عمیق

شبکه‌های عصبی مصنوعی عمیق، معروف به یادگیری عمیق، در چند سال گذشته برنده رقابت‌های

¹ Binarized Statistical Image Features(Bsif)

² Scale

³ Variations Of Posture

فراوانی در زمینه یادگیری ماشین و تشخیص الگو شده‌اند [۵۰]. یادگیری عمیق که به زیرمجموعه یادگیری ماشین تعلق دارد، از لایه‌های پنهان^۱ پی در پی استفاده می‌کند که به صورت سلسله مراتبی برای نمایش یا طبقه‌بندی الگو و یادگیری ویژگی، به صورت خودکار، به طور گسترده به کار گرفته می‌شوند [۵۱]. سه دلیل اصلی برای برجسته شدن یادگیری عمیق وجود دارد [۵۲]: شروع رشد شدید توانایی‌های پردازش (به عنوان مثال، واحدهای GPU)، و هزینه سخت‌افزارهای محاسباتی به شدت پایین، و در نهایت پیشرفت اخیر در مطالعات یادگیری ماشین باعث شد، بسیاری از محققان نتایج موفقیت آمیزی در یادگیری عمیق و کاربردهای متنوعی از آن بدست آورند.

یادگیری عمیق با توجه به نحوه استفاده از تکنیک و معماری می‌تواند به سه کلاس اصلی طبقه بندی شود:

- یادگیری بدون نظارت یا مولد (رمزگذار خودکار^۲ [۵۳]، ماشین بولتزمن^۳ [۵۴]، شبکه عصبی تکرارشونده^۴ [۵۵])
- یادگیری تحت نظارت یا تمییزدهنده (شبکه عصبی کانولوشنال^۵)
- روش‌های ترکیبی

در فصل بعد بیش‌تر با این شبکه‌ها آشنا می‌شویم و به انواع مختلف شبکه‌های عمیق پرداخته می‌شود. در ادامه، چندین روش تشخیص چهره عمیق مبتنی بر CNN را مورد بحث قرار می‌دهیم.

در سال ۲۰۱۴، شبکه DeepFace [۱۱]، توسط محققین فیسبوک و با آموزش شبکه عصبی ۹ لایه و بر روی ۴ میلیون تصویر توانست، به دقت بسیار بالایی در مجموعه داده LFW دست یابد و برای اولین بار به عملکرد انسان در شرایط بدون محدودیت نزدیک شد. با الهام از این مطالعه، تمرکز تحقیقات تشخیص

¹ Hidden Layer

² Auto Encoder (Ae)

³ Boltzman Machine (Bm)

⁴ Recurrent Neural Network (Rnn)

⁵ Convolutional Neural Network (Cnn)

چهره به رویکردهای مبتنی بر یادگیری عمیق منتقل شده و تنها در سه سال، دقت به طور چشمگیری افزایش یافت.

در مقالات [۵۸-۵۶]، با توسعه روش قبل، توانستند به طور پیوسته عملکرد تشخیص چهره را در LFW افزایش دهند. در [۵۶]، آنها ویژگی‌های^۱ (DeepID) را برای یادگیری ویژگی‌های سطح بالا چهره پیشنهاد کردند. ویژگی‌ها از نواحی مختلف صورت استخراج می‌شوند تا نمایش‌های بیش از حد کامل و مکمل را شکل دهند و عملکرد LFW را با چهره‌هایی که تراز ضعیفی دارند در مقایسه با DeepFace بهبود بخشند. یکی از چالش‌های اصلی تشخیص چهره، ایجاد یک نمایش ویژگی کارآمد، برای کاهش تغییرات درون فردی، در عین افزایش تغییرات بین فردی است، که با روش (DeepID2)^۲ [۵۷] تا حدی حل شد. در این روش، ویژگی‌ها با تغییرات CNNهای عمیق تحت دو سیگنال نظارتی یادگرفته شدند. (۱) سیگنال شناسایی چهره، تغییرات بین فردی را با استفاده از طبقه‌بند softmax افزایش داد. (۲) سیگنال تأیید چهره^۳، تغییرات درون فردی را با کمک تابع خطای متضاد کاهش^۴ می‌دهد.

نویسندگان در [۵۸]، یک شبکه عصبی کانولوشن عمیق با کارایی بالا (DeepID2+) برای تشخیص چهره ارائه کردند که نسبت به DeepID2 [۵۷]، با افزایش ابعاد لایه‌های پنهان و افزودن نظارت بر لایه‌های کانولوشن اولیه، بهبود یافته است.

در [۵۹]، دو معماری شبکه عصبی عمیق‌تر برای تشخیص چهره، به نام DeepID3 پیشنهاد شد. این دو معماری، یعنی GoogLeNet [۶۰] و VGGNet [۶۱]، اجزاء اصلی آن‌ها عبارت‌اند از: لایه‌های کانولوشن پشته‌ای^۵ و لایه inception. همچنین سیگنال‌های نظارتی مشترک شناسایی و تأیید چهره در طول آموزشی، به استخراج ویژگی نهایی و لایه‌های میانی اضافه شدند.

در سال ۲۰۱۵، مقاله [۶۲]، روشی برای جمع‌آوری مجموعه دادگان بزرگ VGGface، با کمک تعداد

¹ Deep Hidden Identity

² Deep Identification Verification Features

³ Verification

⁴ Contrastive Loss

⁵ Stacked Convolution

کمی داده نویزی توسعه دادند. همچنین مجموعه دادگان خود را با معماری‌های مختلف آموزش، و بر روی مجموعه دادگان LFW ارزیابی انجام دادند.

هدف تقویت داده^۱، افزایش مجموعه داده با ایجاد تغییراتی بر روی تصاویر بدون تغییر برچسب‌ها است که معمولاً برای بهبود عملکرد CNN و استفاده می‌شود، و از بیش برداشش جلوگیری می‌کند. در همین راستا مقاله [۶۳]، داده‌های آموزشی خود را با تولید تصاویر چهره جدید و از طریق تغییرات ظاهری خاص، از جمله شکل، ژست، و بیان با استفاده از یک چهره عمومی سه بعدی، افزایش داده اند.

لیو و همکاران [۶۴]، تابع خطای Softmax با حاشیه بزرگ^۲ (L-Softmax) را پیشنهاد کردند، که عموماً پرکاربردترین مؤلفه‌ها را در معماری‌های عمیق CNN ترکیب می‌کند، که عبارتند از: تابع خطای آنتروپی متقاطع^۳، Softmax و لایه آخر شبکه تماماً متصل. این روش با تعریف حاشیه قابل تنظیم باعث، جلوگیری از بیش برداشش شده و عملکرد را در کارهای تأیید و طبقه‌بندی افزایش می‌دهد.

FaceNet مدلی از گوگل است که توسط Schroff و همکارانش در سال ۲۰۱۵ پیشنهاد شد [۶۵]، که از نمایش‌های ۱۲۸ بعدی از شبکه‌های کانولوشن عمیق استفاده می‌کند، و روی ۲۰۰ میلیون تصویر چهره، با استفاده از یک تابع خطا سه‌گانه در لایه نهایی آموزش داده شده است. این تابع خطا شامل دو صورت منطبق^۴ و یک صورت منطبق است، و با محاسبه فاصله بین این سه تصویر محاسبه می‌شود. این تابع بیشتر برای تأیید چهره مناسب است.

ون و همکاران [۶۶]، تابع خطا مرکزی^۵ را، معرفی کردند، که براساس ویژگی‌های چهره، برای هر کلاس یک مرکز می‌آموزد، در حالی که به طور همزمان فاصله بین ویژگی‌ها و مراکز کلاس منطبق را کاهش می‌دهد. بنابراین، قدرت تمایز ویژگی‌های چهره آموخته شده افزایش می‌یابد و تغییرات ویژگی درون کلاسی به حداقل می‌رسد.

¹ Data Augmentation

² Large Margin

³ Cross-Entropy

⁴ Match

⁵ Center Loss Function

وو و همکاران [۶۷] در سال ۲۰۱۸ یک ساختار سبک CNN ارائه کرده اند که روی مجموعه داده های عظیم کار می کند. نویسندگان ابتدا max-feature-map (MFM) را معرفی کردند، که نوعی از فعالساز حداکثری^۱ است، و از یک رابطه رقابتی استفاده می کند. آن ها همچنین سه شبکه را معرفی کرده اند که هزینه های محاسباتی و تعداد پارامترها را کاهش می دهد. آنها کار خود را با ارائه یک شبکه به پایان رسانده اند که پیش بینی می کند کدام شبکه با برچسب های نويز سازگارتر است.

به طور خلاصه، شبکه های عصبی کانولوشنال عمیق با یادگیری ویژگی های متمایزتر در مجموعه داده های گسترده و عملکردهای تشخیص بهتر در مقایسه با رویکردهای کل نگر، هندسی و بافت محلی، تشخیص چهره فوق العاده ای را ارائه کرده اند. آن ها همچنین نسبت به تغییرات در ژست، جهت گیری، انسداد جزئی، ناهماهنگی و بیان استحکام نشان دادند. اگرچه پیشرفت های قابل توجهی با تشخیص چهره مبتنی بر یادگیری عمیق انجام شده است، برخی از چالش ها وجود دارد: آموزش کارآمد CNN به داده های آموزشی در مقیاس بزرگ نیاز دارد، به پیشرفت های سخت افزاری مانند GPU نیاز دارد و به داده های باکیفیت زیادی نیاز دارد.

۲-۵ تشخیص چهره تحت انسداد

انسداد یک مسئله رایج و چالش برانگیز در هنگام تشخیص چهره در برنامه های کاربردی مانند نظارت تصویری است. با وجود پیشرفت های چشم گیری که توسط شبکه های عمیق، در حوزه تشخیص چهره بدون انسداد، به وجود آمده است، با این حال، دقت سیستم های تشخیص چهره در مواجهه با انسداد به شدت پایین می آید. تشخیص چهره تحت انسداد به چندین دلیل می تواند چالش برانگیز باشد:

- انسداد می تواند در هر جایی و با هر اندازه یا شکلی در تصویر صورت باشد.

^۱ Maxout

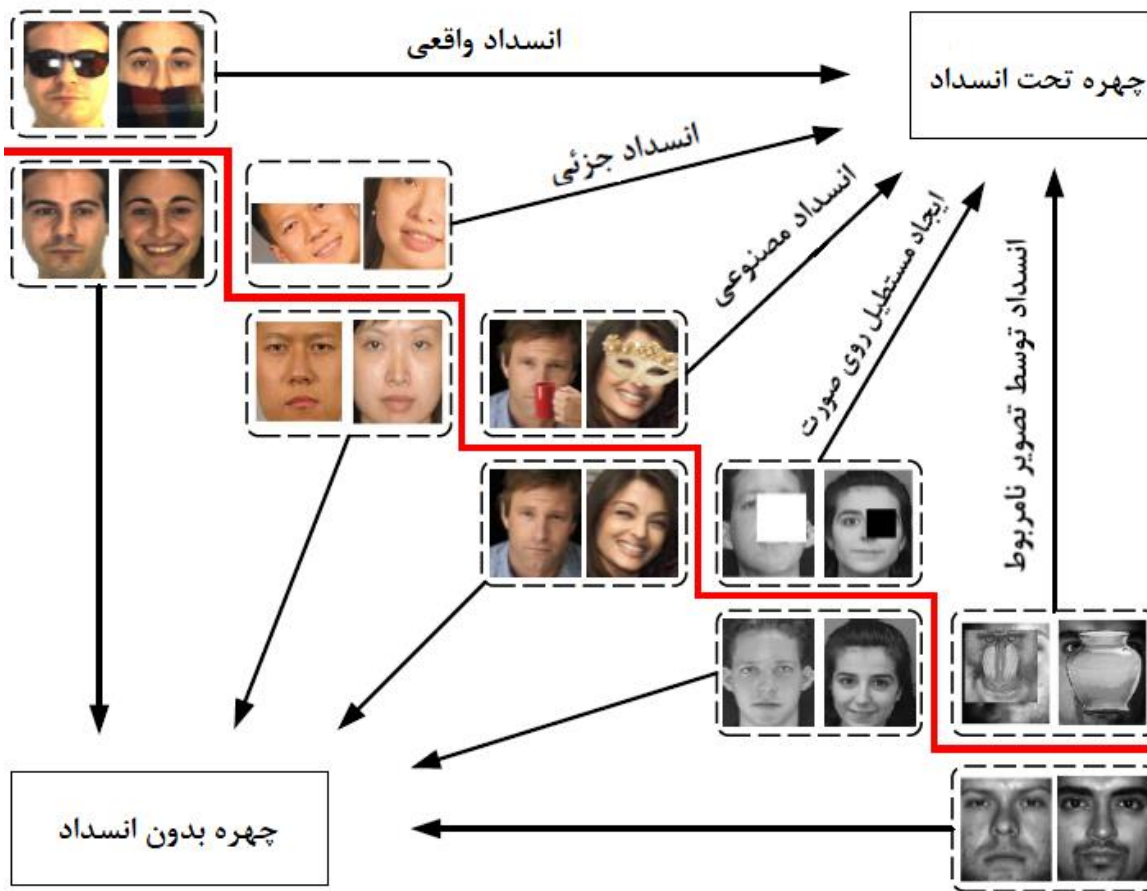
- در انسداد، محتوای بصری ناقص و نشانه‌های هویتی ناکافی می‌باشد.
 - جمع‌آوری مجموعه داده‌های آموزشی در مقیاس بزرگ، با تمام انسدادهای ممکن، در یک سناریوی واقعی، برای استفاده از تکنیک‌های یادگیری عمیق امکان پذیر نیست.
- انسداد چهره، می‌تواند انواع مختلفی داشته باشد، که در شکل (۵-۲) آورده شده است.
- در ادامه به بررسی روش‌های مختلف تشخیص چهره تحت انسداد، و سپس به‌طور خاص روش‌های تشخیص چهره با ماسک، که تمرکز این پژوهش نیز بر روی آن است را بررسی می‌کنیم.

۲-۵-۱ روش‌های تشخیص چهره تحت انسداد

رویکردهای تشخیص چهره تحت انسداد می‌توانند به طور کلی به سه دسته طبقه بندی می‌شوند [۶۸]:

- روش‌های مبتنی بر استخراج ویژگی
- روش‌های تشخیص چهره آگاه از ناحیه انسداد^۱
- روش‌های تشخیص چهره مبتنی بر بازیابی ناحیه انسداد

^۱ Occlusion Aware Face Recognition



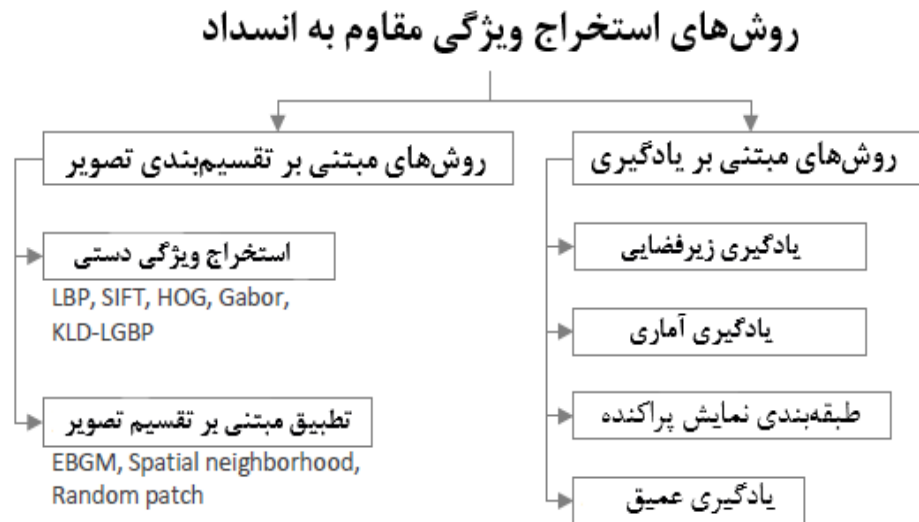
شکل ۵-۲ انواع مختلف انسداد.

۱-۵-۱ روش‌های مبتنی بر استخراج ویژگی

اگر ویژگی‌های استخراج شده به‌طور منطقی در برابر انسداد مقاوم باشند، آنگاه مشکلات در تشخیص چهره تحت انسداد تا حد زیادی حل می‌شود. بنابراین در این روش هدف، استخراج ویژگی‌هایی است که در عین حفظ تمایز، کمتر تحت تأثیر انسداد قرار بگیرند.

ما رویکردها را به ویژگی‌های مهندسی شده و ویژگی‌های مبتنی بر یادگیری تقسیم‌بندی می‌کنیم. اولی به‌طور کلی ویژگی‌های دستی را از نواحی مشخص صورت استخراج می‌کند، که نیازی به بهینه‌سازی یا مرحله یادگیری ندارند. در روش دوم، استخراج ویژگی با استفاده از روش‌های مبتنی بر یادگیری مانند

روش‌های زیرفضای خطی^۱، طبقه‌بندی نمایش پراکنده^۲، یا تکنیک‌های یادگیری عمیق، انجام می‌شود. طبقه‌بندی این روش‌ها در شکل (۲-۶) آمده است.



شکل ۲-۶ روش‌های استخراج ویژگی مقاوم به انسداد.

• روش‌های مبتنی بر تقسیم‌بندی تصویر

توصیف‌گرهای صورت بنا به دلایل زیر، در چندین سال قبل، یعنی قبل از فراگیری روش‌های یادگیری عمیق، بسیار مورد مطالعه و بررسی قرار می‌گرفتند:

- به راحتی از تصاویر خام صورت استخراج می‌شوند.
 - استخراج ویژگی در فضای کم انجام می‌شود و در نتیجه محاسبات کاهش می‌یابد.
 - تا حدی تنوع زیاد در ظاهر صورت را تحمل می‌کنند.
- به طور کلی، ویژگی‌های مهندسی شده (به عنوان مثال، ویژگی‌های صنایع دستی) صورت را به بخش‌های کوچک‌تری تقسیم و بر روی این قسمت‌ها فیلترهای مختلف اعمال می‌کند.

این روش‌ها، نیاز به تراز دقیق نواحی مختلف چهره، مثل مختصات چشم‌ها دارند، که این ویژگی باعث

¹ Linear Subspace Methods

² Sparse Representation Classification

می‌شود از این روش‌ها در کاربردهای واقعی استفاده نشود، زیرا امکان دارد نواحی مختلف چهره مسدود شود و در نتیجه تراز چهره امکان‌پذیر نباشد.

• روش‌های مبتنی بر یادگیری

در مقایسه با ویژگی‌های مهندسی شده، ویژگی‌های آموخته شده زمانی که انواع انسداد در مکان‌های مختلف وجود دارند، انعطاف‌پذیرتر هستند. همچنین ویژگی‌های آموخته شده از داده‌های آموزشی می‌توانند موثر باشند و قدرت تمایز بالقوه بالایی داشته باشند.

• یادگیری زیرفضایی^۱

تصاویر چهره در واقع به یک زیرفضای چهره محدود می‌شوند. بنابراین، این روش‌ها می‌توانند تغییرات اجزای مختلف صورت را که برای تمایز بین افراد ضروری است، حفظ کند.

در بخش‌های قبل با روش‌های PCA، LDA به عنوان شناخته شده ترین روش‌های این دسته آشنا شدیم. مقاله [۶۹]، از PCA برای تشخیص انسداد، و فرموله کردن ضرایب با استفاده از بقیه پیکسل‌ها، استفاده می‌کند. این روش توانست به نتایج طبقه‌بندی خوبی دست یابد و توانایی تشخیص نقاط پرت (انسداد) را داشته باشد.

مقاله [۷۰]، برای مقابله با انسداد جزئی چهره، از تجزیه و تحلیل ویژگی‌های همبستگی در نواحی مختلف صورت، به عنوان مثال، ناحیه چشم، ناحیه بینی و ناحیه دهان استفاده کرد. این روش عملکرد بهتری نسبت به روش‌های LDA و PCA از خود نشان داد.

• یادگیری آماری

در کاربردهای دنیای واقعی، همه انواع انسدادها احتمال وقوع یکسانی ندارند. به عنوان مثال، روسری و عینک آفتابی اغلب احتمال بروز بیشتری دارند. از این رو، می‌توان از یادگیری آماری بر روی تکه‌های

^۱ Subspace Learning

صورت^۱ استفاده شود تا امکان انسداد آنها را در نظر بگیریم. یکی از کارهای اولیه در این راستا، مرجع [۷۱] است، که رویکردی احتمالی دارد، یعنی مخلوطی از گاوسی‌ها برای جبران چهره‌های مسدود. به طور خاص، آن‌ها مناطق محلی تقسیم شده صورت را تجزیه و تحلیل می‌کنند و از رویکرد احتمالی برای پیدا کردن بهترین تطابق استفاده می‌کنند، به طوری که سیستم تشخیص، حساسیت کمتری به انسداد داشته باشد.

در مرجع [۷۲]، زیرفضای چهره با نقشه‌های خودسازماندهی^۲ (SOM) گسترش داده می‌شود. این رابطه شباهت بلوک‌های فرعی را در فضای ورودی در فضای توپولوژیکی SOM نشان می‌دهد. علاوه بر این، آن‌ها از طبقه‌بندی‌کننده نزدیک‌ترین همسایه^۳ استفاده می‌کنند تا رابطه شباهت به دست آمده توسط پیش‌بینی‌های SOM را به طور مؤثر ثبت کنند، که استحکام کل سیستم را بهبود می‌بخشد

مرجع [۷۳]، از هسته گاوسی محلی برای استخراج ویژگی استفاده و سپس با جمع هسته‌های محلی به عنوان ویژگی‌های ترکیبی، طبقه‌بند SVM را آموزش داد.

در مقاله [۷۴]، به کمک یادگیری آماری ویژگی‌های محلی سعی در توسعه یک روش تشخیص چهره داشتند که بر روی انسداد جزئی نیز مقاوم باشد. لذا آنها، احتمال توصیفگرهای ویژگی SIFT در تصاویر آموزشی را بر اساس یک مدل ساده گاوسی تخمین زدند. در مرحله طبقه‌بندی، برای اندازه‌گیری اهمیت هر یک از ویژگی‌های محلی، در تصاویر تست، با محاسبه فاصله وزنی، احتمال تخمین زده شد. بر اساس این کار، آنها یادگیری آماری را بر اساس ویژگی‌های محلی به یک چارچوب کلی گسترش دادند که به کمک، وزن‌های یاد گرفته شده اندازه‌گیری فاصله انجام می‌شود، که وزن‌ها از ویژگی‌های محلی و ویژگی‌های مبتنی بر شباهت بدست می‌آید. استخراج ویژگی از منطقه محلی نمی‌تواند اطلاعات فضایی^۴ چهره را کد کند، که باعث کاهش عملکرد خواهد شد.

¹ Face Patches

² Self Organizing Maps

³ K-Nearest Neighbor

⁴ Spatial Information

• طبقه‌بند نمایش پراکنده^۱

جدا از این روش‌های یادگیری آماری، چندین الگوریتم از طبقه‌بندی نمایش پراکنده (SRC) برای مقابله با انسداد در تشخیص چهره استفاده می‌کنند. روش SRC [۷۵]، از زمان معرفی توجه فرآینده ای را از سوی محققان به خود جلب کرد. این روش قدرت تمایز نمایش پراکنده یک چهره را بررسی می‌کند و از ترکیبی خطی از نمونه‌های آموزشی به‌علاوه خطاهای پراکنده استفاده می‌کند تا بتواند انسداد را شناسایی کند.

یانگ و همکاران [۷۶]، یک طرح بازنمایی و طبقه‌بندی قوی مبتنی بر ویژگی گابور (GRRC) را پیشنهاد کردند که از ویژگی‌های Gabor به جای مقادیر پیکسل برای نمایش تصاویر چهره، استفاده می‌کنند که می‌تواند توانایی تشخیص را افزایش دهد. علاوه بر این، این روش با استفاده فیلتر گابور محاسبات کم هزینه‌تری برای کدگذاری بخش‌های انسداد در مقایسه با SRC اصلی نیاز دارد. برای بررسی روش پیشنهادی، آنها از زیر مجموعه ای از پایگاه داده AR استفاده کردند. این پایگاه داده شامل ۷۹۹ تصویر (حدود هشت نمونه برای هر نفر) از نماهای جلویی غیر مسدود با حالات مختلف چهره برای آموزش است. در مجموعه تست، از ۲۰۰ تصویر همراه با عینک و شال استفاده شد. روش پیشنهادی به دقت تشخیص ۹۳ درصد در عینک آفتابی و ۷۹ درصد دقت در روسری دست یافت. این روش حدود ۱۰ درصد دقت تشخیص بالاتری نسبت به روش SRC داشت.

• یادگیری عمیق

نمایش چهره به دست‌آمده توسط شبکه عمیق عصبی، به دلیل استفاده از مجموعه‌های آموزشی عظیم، بسیار برتر از سایر روش‌های مبتنی بر یادگیری در قدرت تمایز است [۷۷].

^۱ Sparse Representation Classifier (Src)

از نظر عملی، اگر مجموعه داده آموزشی عظیمی ارائه شود که حاوی چهره‌های مسدود شده کافی، برای یک شبکه عمیق باشد، می‌توانیم نمایش چهره مقاوم در برابر انسداد را بدست آوریم [۷۸]. اگرچه جمع‌آوری همچنین پایگاه داده‌ای کار سختی است. به نظر می‌رسد تکنیک افزایش داده‌ها، راه‌حل قابل قبولی برای به‌دست آوردن تعداد زیادی از تصویر صورت دارای انسداد است. مرجع [۷۹]، به‌منظور تقویت تعداد داده‌ها به کمک مدل مو و عینک‌های مختلف، تصاویر چهره مصنوعی ایجاد کرد. هدف از این کار ساخت مدل کانولوشنی بود که بتواند در برابر مدل موهای مختلف و عینک، ویژگی‌های تمییز دهنده استخراج کند. این روش در واقع مشکل کمبود داده را تا حدودی برطرف کرد و منجر به بهبود عملکرد شد. با این حال، استفاده از آن محدود به استفاده از عینک آفتابی و مو در تشخیص است و به طور کلی مشکل انسداد را حل نمی‌کند.

در مرجع [۸۰]، آن‌ها به جای استفاده مستقیم از صورت‌های مسدود شده مصنوعی، ابتدا اهمیت نواحی صورت را در آزمایش حساسیت انسداد تشخیص می‌دهند و سپس یک CNN را با نواحی شناسایی شده صورت مسدود آموزش می‌دهند تا اتکای مدل به این نواحی کاهش یابد. به طور خاص، آن‌ها پیشنهاد می‌کنند که مجموعه آموزشی با تصاویر چهره با انسدادهای واقع در نواحی پر جلوه (قسمت مرکزی صورت) بیشتر از مناطق با اثر کم (قسمت‌های بیرونی صورت) تقویت شود. این کار مدل را مجبور می‌کند تا ویژگی‌های متمایز بیشتری را از قسمت‌های بیرونی صورت بیاموزد و در نتیجه وقتی قسمت مرکزی مسدود می‌شود، مشکل کمتری در تشخیص ایجاد می‌شود. با این حال، وقتی سائز انسدادهای از پیش تعریف شده، با تصاویر چهره مسدود برابر نیست، ممکن است باعث کاهش عملکرد شود.

۲-۵-۱-۲ روش‌های تشخیص چهره آگاه از ناحیه انسداد

اگر فقط از قسمت‌های قابل مشاهده صورت برای تشخیص استفاده می‌شود، پس مشکل انسداد تا

حدودی کاهش می‌یابد. این روش‌ها که به صراحت ناحیه انسداد را حذف می‌کنند روش‌های تشخیص چهره آگاه از انسداد (OAFR) نامیده می‌شوند. این روش‌ها به دو دسته تقسیم می‌شوند. دسته اول، تشخیص چهره مبتنی بر شناسایی انسداد است که ابتدا نواحی مسدود را شناسایی می‌کنند و سپس تنها از قسمت‌های غیرانسدادی را پردازش می‌کنند. دسته دوم فرض می‌کند که یک صورت ناقص^۱ در دسترس است و برخلاف دسته اول، شناسایی انسداد در تشخیص چهره ناقص نادیده گرفته می‌شود.

• روش‌های تشخیص چهره مبتنی بر شناسایی انسداد

یک ایده شهودی برای مقابله با انسداد در تشخیص چهره این است که ابتدا انسداد را شناسایی کنید و سپس چهره را بر اساس قسمت‌های بدون انسداد صورت تشخیص دهید. برخی روش‌ها از انواع انسداد از پیش تعریف شده به عنوان جایگزینی برای انسداد دلخواه در مکان‌های مختلف استفاده می‌کنند تا چالش انسداد را ساده کنند.

برخی از کارها [۸۱]، [۸۲] از یک طبقه‌بندی کننده باینری برای جستجوی ناحیه مسدود شده استفاده می‌کنند و فقط قسمت‌های مسدود نشده را برای مقایسه در نظر می‌گیرند. به طور خاص، آن‌ها ابتدا صورت را به چندین ناحیه غیر همپوشانی تقسیم می‌کنند و سپس یک طبقه‌بندی کننده SVM را آموزش می‌دهند تا تشخیص دهد که آیا هر ناحیه صورت مسدود شده است یا خیر. با حذف مناطق مسدود شده، دقت تشخیص کلی افزایش می‌یابد.

از آنجایی که انسداد می‌تواند ویژگی‌های کل یک تصویر را خراب کند، تکنیک‌های یادگیری عمیق با ارائه یک نمایش بهتر، مشکل را کاهش می‌دهد. در مرجع [۸۳]، یک شبکه عصبی کانولوشن برای شناسایی انسداد در یک محیط چند وظیفه ای طراحی شده است. به طور خاص، چهار ناحیه خاص برای شناسایی انسداد وجود دارد، و هدف هر کدام پیش‌بینی احتمال انسداد یک جزء خاص است. این چهار ناحیه عبارت‌اند از: چشم چپ، چشم راست، بینی و دهان. با این حال، پیش‌بینی فقط انسدادهای

^۱ Partial Face

از پیش تعریف شده انعطاف پذیری را محدود می کند و عدم دقت شناسایی انسداد می تواند در عوض به عملکرد تشخیص آسیب برساند.

برخی از مقالات، قسمت های قابل مشاهده صورت را برای تشخیص انتخاب می کنند و با فرض آگاهی قبلی از انسداد از شناسایی انسداد چشم پوشی می کنند. در مرجع [۸۴]، چندین روش مبتنی بر فضای فرعی از جمله PCA، فاکتورسازی ماتریس غیرمنفی (NMF) [۸۵]، NMF محلی [۸۶]، را مقایسه می کند و از اطلاعات جزئی موجود برای تشخیص چهره استفاده می کند. در هنگام تشخیص چهره، ناحیه چشم زمانی انتخاب می شود که افراد ماسک دارند و زمانی که افراد عینک می زنند از ناحیه پایین استفاده می شود. بدست آوردن نواحی از پیش تعریف شده، به صورت تراز شده، در سناریوی واقعی دشوار است. از این رو در عمل استفاده از این روش در عمل کاربردی نیست. این روش در استفاده از انعطاف پذیری کمبود دارد زیرا به خوبی تراز است بدست آوردن نواحی از پیش تعریف شده در سناریوی واقعی دشوار است.

در مقاله [۸۷]، شبکه پیشنهادی MaskNet به لایه میانی مدل های CNN اضافه می شود، با هدف یادگیری ویژگی های تصویر با کیفیت بالا و نادیده گرفتن مواردی که توسط انسداد تحریف شده اند. MaskNet به عنوان یک شبکه پیچیده کم عمق تعریف می شود که انتظار می رود وزن کمتری را به واحدهای پنهان فعال شده توسط نواحی مسدود صورت اختصاص دهد.

• تشخیص چهره جزئی

چهره های جزئی اغلب در سناریوهای بدون محدودیت ظاهر می شوند، به ویژه تصاویری که توسط دوربین های نظارتی یا دستگاه های دستی (مانند تلفن های همراه) گرفته شده است. از آنجایی که مقایسه ویژگی های بخش های معنایی مختلف صورت بی معنی است، لذا جستجو برای مطابقت معنایی نواحی مشخص صورت و تمامی تصاویر صورت در پایگاه داده ضروری است. مطابقت معنایی را می توان

در مرحله استخراج ویژگی برای استخراج ویژگی‌های چهره مقاوم به انسداد و با قدرت متمایز بالا تکمیل کرد، یا در مرحله مقایسه برای ساخت یک طبقه‌بندی چهره قوی.

شبکه کانولوشنی نظارت دوگانه چند مقیاسی (MDSCNN)^۱ [۸۸]، برای استخراج ویژگی دارای چندین شبکه است که هر کدام بر روی نواحی مختلف صورت در مقیاس‌های مختلف آموزش می‌بینند. تکه‌های چند مقیاسی از کل صورت انتخاب و براساس چشم‌ها تراز می‌شوند. وزن شبکه‌های متعدد برای ایجاد دقت تشخیص نهایی ترکیب می‌شوند. اگرچه این روش می‌تواند نمایش ویژگی‌های خوبی را ارائه دهد، اما آموزش ۵۵ شبکه عصبی کانولوشنال نظارت دوگانه (DSCNN) با توجه به نواحی مقیاس‌بندی شده مشکل‌ساز است و در عمل وقت‌گیر است زیرا برای تولید وصله‌های چند مقیاسی نیاز به پنجره کشویی^۲ است.

مرجع [۸۹]، اولین مطالعه سیستماتیک را برای تشخیص چهره، بدون نیاز اولیه به تراز چهره انجام داد، که به عنوان نقطه عطفی در زمینه تشخیص جزئی چهره در نظر گرفته می‌شود. روش پیشنهادی از توصیفگرهای چند نقطه کلیدی^۳ (MKD) برای نشان دادن یک صورت کلی یا ناقص با طول متغیر استفاده می‌کند. توصیف‌گرها از مجموعه‌ای از تصاویر، دیکشنری می‌سازند، و این امکان را فراهم می‌کنند که به صورت پراکنده توصیفگرهای تصاویر ناقص را نشان دهند و بر این اساس هویت تصویر ناقص را استنتاج کنند.

۲-۵-۱-۳ روش‌های تشخیص چهره مبتنی بر بازیابی ناحیه انسداد

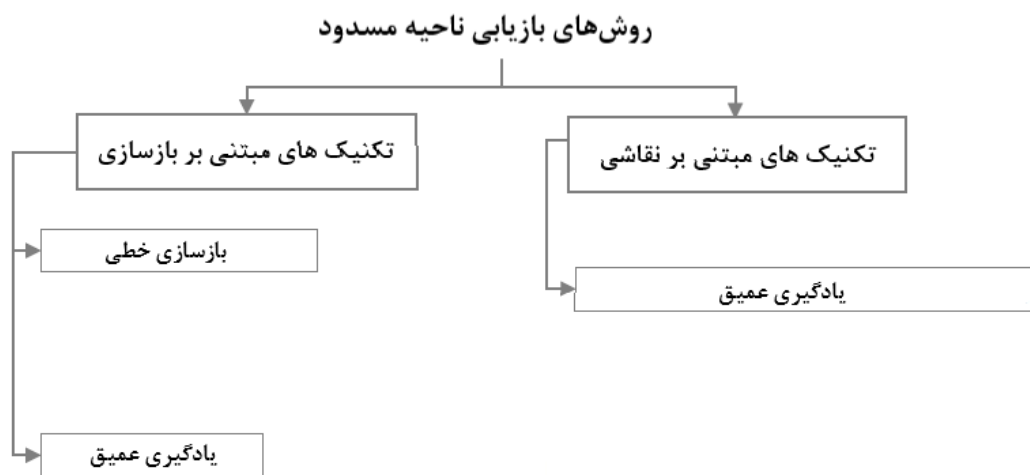
جدا از پرداختن به مشکل انسداد در فضای ویژگی، یک ایده بصری این است که بازیابی صورت مسدود شده را به عنوان جایگزینی برای حل انسداد در فضای پیکسلی تصویر در نظر بگیریم. روش‌های بازیابی

¹ Multiscale Double Supervision Convolutional Neural Network

² Sliding Window

³ Multi-Keypoint Descriptors

انسداد، یک صورت کامل را از صورت مسدود شده بازیابی می‌کند، که امکان استفاده مستقیم از الگوریتم‌های تشخیص چهره معمولی را فراهم می‌کند. به‌طور کلی دو دسته روش برای بازیابی انسداد چهره وجود دارد. روش‌های دسته اول، از تکنیک‌های مبتنی بر بازسازی^۱ برای تشخیص چهره استفاده می‌کنند درحالی که روش‌های دوم، تکنیک‌های نقاشی تصویر^۲ استفاده می‌کنند، که چهره مسدود شده را به عنوان یک مشکل ترمیم تصویر در نظر می‌گیرد. در شکل (۲-۷) طبقه بندی این روش‌ها را مشاهده می‌کنید.



شکل ۲-۷ روش‌های بازیابی ناحیه مسدود.

• تکنیک‌های مبتنی بر بازسازی

تکنیک‌های بازسازی به بررسی رابطه بین تصویر مسدود و تصویر بدون انسداد در فضای پیکسلی تصویر می‌پردازند و سعی دارند نواحی مسدود را به‌گونه‌ای بازسازی کنند که مانند تصویر بدون انسداد شوند. این تکنیک‌ها به دو دسته بازسازی خطی و تکنیک‌های یادگیری عمیق طبقه‌بندی می‌شوند.

• بازسازی خطی

¹ Reconstruction

² Image Inpainting

در مرجع [۹۰]، برای حذف انسداد چشم ناشی از عینک از PCA استفاده شده است. آن‌ها با ترکیب مدل تصادفی مارکوف و نمایش پراکنده صورت‌های مسدود شده سعی در بهبود بازسازی نواحی عینک داشتند. انواع زیادی از روش‌ها [۹۱]، [۹۲]، [۹۳] وجود دارد که از PCA برای تشخیص نقاط پرت یا انسداد و سپس بازسازی تصاویر صورت بدون انسداد استفاده می‌کنند.

● یادگیری عمیق

برخی از مقالات از تکنیک‌های یادگیری عمیق برای بازسازی انسداد استفاده می‌کنند. یکی از این مقالات [۹۴]، از رمزگذار پشته‌ای حذف نویز پراکنده^۱ در یک کانال دوگانه برای حذف انسداد صورت استفاده می‌کند. الگوریتم لایه‌ای را برای یادگیری نمایش تصویر بدون انسداد به کار می‌گیرد به این صورت که پارامترهای رمزگذاری یادگرفته شده بتوانند داده‌های نویز را بازسازی کنند. کار دیگر [۹۵] ترکیب معماری LSTM و رمزگذار خودکار^۲ را برای حل مشکل انسداد چهره پیشنهاد می‌کند. LSTM پیشنهادی از دو جزء تشکیل شده است، یک شبکه LSTM، تکه‌های چهره را در مقیاس‌های مختلف و به‌طور متوالی رمزگذاری^۳ می‌کند. شبکه دو کاناله LSTM دیگر از طریق رمزگشایی^۴ سعی در شناسایی انسداد و در نهایت بازسازی ناحیه مسدود دارد.

● تکنیک‌های مبتنی بر نقاشی تصویر

تکنیک‌های نقاشی درون تصویر به طور گسترده‌ای برای به دست آوردن دقیق تصاویر بدون انسداد استفاده می‌شود و محدود به تصاویر چهره نیستند. تکنیک‌های Inpainting بر ترمیم تصاویر مسدود شده تمرکز می‌کنند و تشخیص چهره را دور از توجه قرار می‌دهند.

با توجه به اینکه در روش پیشنهادی از این تکنیک استفاده می‌کند، بنابراین بررسی کامل و دقیقی از روش‌های این دسته خواهیم داشت. در شکل (۸-۲) نتایج این روش‌ها را مشاهده می‌کنید.

¹ Stacked Sparse Denoising Autoencoder

² Autoencoder

³ Encoding

⁴ Decoding



شکل ۸-۲ خروجی تکنیک های مبتنی بر نقاشی تصویر.

مدل های مولد به دلیل توانایی تولید نمونه های جدید از توزیع یکسان مجموعه داده آموزشی، شناخته شده اند. برجسته ترین مدل های مولد، از رمزگذار خودکار متغیر^۱ (VAE) [۹۶] و شبکه مولد تخصصی^۲ (GAN) [۹۷]، نشأت گرفته اند. در VAE، بر خلاف رمزگذار خودکار سنتی^۳، داده های ورودی از توزیع احتمالی پیروی می کنند. این امر موجب می شود این شبکه ها قادر به تولید تصاویر جدید باشند. با این حال، تصاویر تولید شده کیفیت پایینی دارند. در عوض، شبکه GAN می آموزد که تصویر جدید از توزیع داده های آموزش و با استفاده از بازی MINMAX بین شبکه مولد و شبکه تفکیک کننده تولید کند. تمایزکننده می خواهد بین تصاویر واقعی و جعلی تمایز قائل شود، در حالی که تولیدکننده می خواهد با تولید تصاویری واقعی، متمایزکننده را فریب دهد. مرجع [۹۸]، در سال ۲۰۱۵ توانست از معماری های کانولوشنال در GAN استفاده می کند و نمونه های واقع تولید کند. در سال ۲۰۱۷ مرجع [۹۹]، شبکه GAN را برای انجام نقاشی چهره پیشنهاد کرد. این روش توانست با بخش بندی تصویر و در نهایت استفاده

^۱ Variational Autoencoder

^۲ Generative Adversarial Network

^۳ Traditional Autoencoder

از آن در مراحل آموزش دقت تأیید چهره را بهبود می‌بخشد.

ژائو و همکاران [۱۰۰] مدل GAN با عامل دوگانه را برای بهبود واقع‌گرایی تصاویر چهره پیشنهاد کردند. همچنین با معرفی تابع خطا ادراک هویت^۱ اطلاعات هویتی در طول زمان حفظ شده و این شبکه عملکرد جذابی ارائه دهد. مرجع [۱۰۱]، شبکه تنوع هویت^۲ (IDGAN) رویکردی را برای تسهیل ظرفیت، شناسایی چهره‌های مسدود شده ارائه کرد. در این روش، یک شبکه مجزا تحت عنوان شناساگر، به عنوان بازیکنی در نظر گرفته می‌شود که با مولد رقابت می‌کند، که در نهایت منجر به نتایج با کیفیت بالا و حفظ هویت می‌شود. در سال ۲۰۲۲ مرجع [۱۰۲] یک چارچوب جدید با استفاده از تکنیک نقاشی تصویر و برای حذف انسداد ناشی از ماسک صورت ارائه کرد. آن‌ها یک مازول بازسازی چهره سه بعدی ایجاد کردند، که تصویر سه بعدی بدون ماسک را پیشبینی و همراه با نگاشت باینری ناحیه مسدود، خروجی می‌دهد. سپس خروجی این قسمت به یک شبکه کانولوشنی داده می‌شود و در نهایت تصویر بدون ماسک تولید می‌شود.

۲-۶ جمع بندی

در این فصل ابتدا تاریخچه سیستم‌های تشخیص چهره را مشاهده و سپس ساختار کلی تشخیص چهره را بررسی کردیم. در بخش بعد به روش‌های تشخیص چهره پرداختیم و دیدیم که یادگیری عمیق تحول عظیمی در این حوزه به وجود آورد و با توجه به دقت بالاتر و انعطاف‌پذیری بهتری که داشتند در عمل نیز بجای روش‌های سنتی مورد استفاده قرار گرفتند. در بخش آخر نیز تکنیک‌های بکارگرفته شده، در هنگام مواجهه با انسداد را بررسی کردیم. بطور کلی تکنیک‌ها یا در فضای ویژگی و یا در فضای پیکسلی تصویر عمل می‌کردند و با توسعه روش‌های یادگیری عمیق، علی‌الخصوص شبکه‌های GAN، دیدیم که برای تشخیص چهره دارای انسداد می‌توان ابتدا به کمک این شبکه‌ها، عملیات ترمیم تصویر انجام داده و سپس خروجی شبکه، یعنی تصویر بدون انسداد است را به یک شبکه تشخیص چهره داده و بدین‌گونه دقت

¹ Identity Perception

² Identity-Diversity Gan

تشخیص چهره دارای انسداد را افزایش دهیم.

فصل ۳: روش مشاهدی

۳-۱ مقدمه

تشخیص چهره یکی از موضوعات پرطرفدار و قدیمی در بینایی کامپیوتر است. در چند سال اخیر، با توسعه تکنیک‌های یادگیری عمیق مبتنی بر شبکه عصبی کانولوشنی، پیشرفت قابل توجهی داشته است. زدن ماسک صورت، باعث از دست رفتن بیش از نیمی از اطلاعات هویتی و درنهایت عدم تشخیص فرد می‌شود. بسیاری از روش‌ها با استخراج ویژگی‌های بهتر، به‌طور مستقیم سعی در بهبود سیستم تشخیص چهره دارند. در این سیستم‌ها، تصویر با ماسک به عنوان ورودی گرفته می‌شود و سیستم تشخیص چهره هویت شخص را پیش‌بینی می‌کند. برخلاف روش‌های ذکر شده، در این تحقیق، چالش تشخیص چهره با ماسک در فضای پیکسلی بررسی شده است. کار اصلی ما در این پایان‌نامه از دو قسمت بخش‌بندی^۱ ناحیه ماسک (شکل ۳-۱) و استفاده از شبکه GAN به‌منظور ترمیم ناحیه ماسک، تشکیل شده است. در بخش اول شبکه جدید Att-BANet در کنار دادگان تصاویر با ماسک معرفی و پس از اجرای شبکه، از خروجی آن در شبکه دوم استفاده می‌شود. در بخش دوم، خروجی بخش اول و تصویر با ماسک، به‌عنوان ورودی به شبکه GAN پیشنهادی داده می‌شود و خروجی تصویر بدون ماسک تولید می‌شود (شکل ۳-۲).



شکل ۳-۱ بخش بندی ناحیه ماسک.

^۱Segmentation



شکل ۲-۳ ورودی و خروجی شبکه GAN پیشنهادی.
 ورودی ورودی خروجی

۳-۲ بخش بندی ناحیه ماسک

بخش بندی ناحیه ماسک، یک مرحله مقدماتی برای مقابله با مشکل انسداد از طریق بازسازی تصویر است. برچسب گذاری^۱ ناحیه ماسک، یک عملیات زمان بر است. از این رو، دادگان های تصاویر با ماسک، فاقد بخش بندی ناحیه ماسک هستند و فقط مختصات^۲ ماسک وجود دارد. در نتیجه روش های فعلی، تصویر ماسک را بر روی چهره به صورت مصنوعی^۳ قرار می دهند. تصاویر مصنوعی ایجاد شده کمی متفاوت از تصاویر واقعی هستند و با استفاده از این تصاویر در سناریوهای واقعی، دقت تشخیص چهره کاهش می یابد. نمونه ای از این تصاویر در شکل (۳-۳) نشان داده شده اند.



شکل ۳-۳ نمونه تصاویر با ماسک که ماسک به صورت مصنوعی بر روی صورت قرار گرفته است.

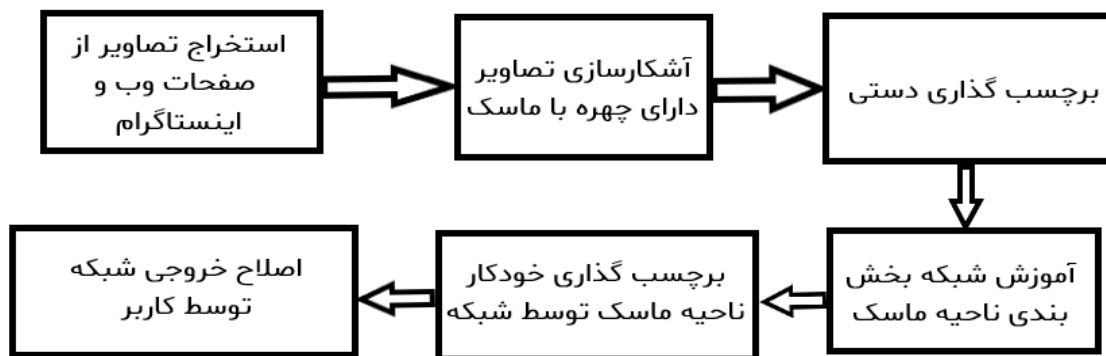
¹ Labeling

² Bounding Box

³ Synthesize

۳-۲-۱ دادگان تولیدی

برای حل مشکل قرار دادن ماسک بر روی صورت، به صورت مصنوعی، که باعث کاهش دقت تشخیص چهره با ماسک می‌شود، دادگان جدیدی تحت عنوان^۱ MFSD معرفی کردیم که شامل حدود ۱۲۰۰۰ تصاویر با ماسک به همراه بخش‌بندی ناحیه ماسک است که در شکل (۳-۵) تعدادی از تصاویر نشان داده شده است. به این منظور، با استفاده از ابزار خزنده^۲ در پایتون، حدود ۲۵۰۰۰ تصویر چهره از صفحات وب و اینستاگرام استخراج شد. سپس به کمک ابزار شناسایی چهره با ماسک AIZOO [۱۰۳]، حدود ۸۵۰۰ تصویر شناسایی شد که در آن حداقل یک نفر یا بیشتر، با ماسک وجود دارند. از این رو، در نهایت، تصویر با ماسک ۱۲۰۰۰ نفر استخراج شد. فرآیند برچسب‌گذاری به صورت نیمه خودکار انجام شد. بدین صورت که ابتدا در حدود ۲۸۵۰ تصویر ناحیه ماسک به صورت دستی توسط کاربر و با ابزار LabelMe مشخص شد. سپس از این تصاویر برای آموزش شبکه بخش‌بندی پیشنهادی، استفاده شد. در نهایت مابقی تصاویر به کمک شبکه آموزش داده شده، بخش‌بندی شد. در ادامه، کاربر تصاویر را بررسی کرد و در صورت مشاهده خطا، آن‌ها را اصلاح نمود. گام‌های ایجاد مجموعه دادگان ساخته شده در شکل (۳-۴) مشاهده می‌شود.



شکل ۳-۴ گام‌های ایجاد مجموعه دادگان MFSD.

¹ Masked Face Segmentation Database

² Crawler

تصویر استخراج شده

بخش بندی ناحیه ماسک



شکل ۵-۳ نمونه تصاویر دادگان MSFD.

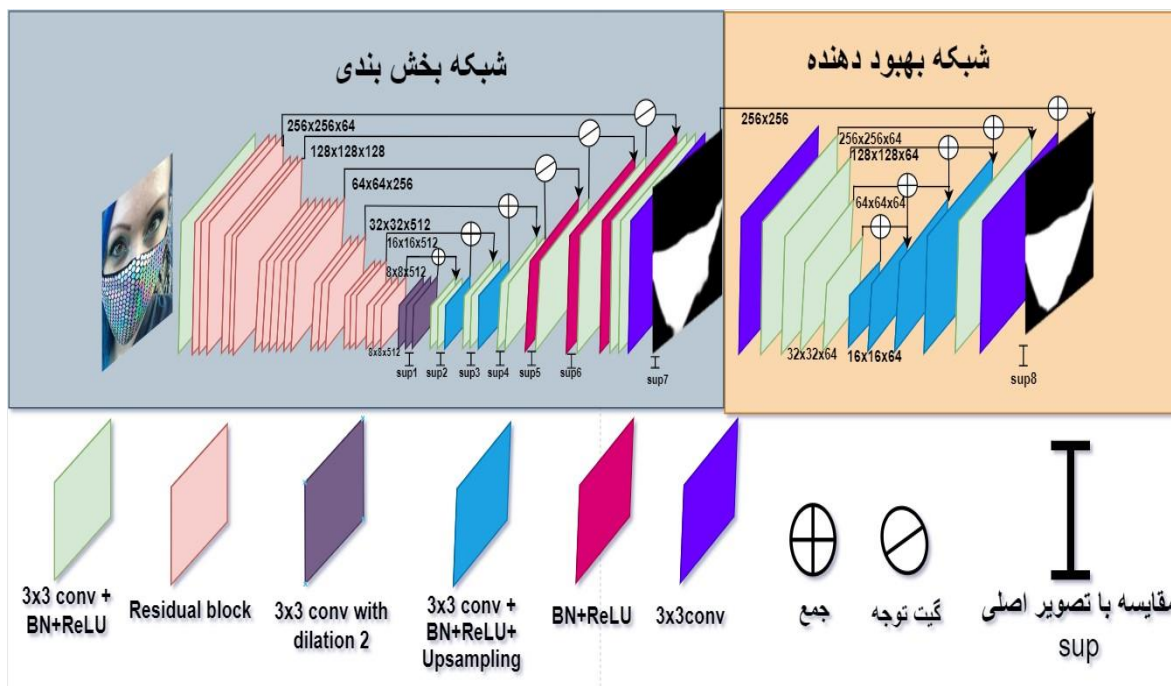
۲-۲-۳ شبکه بخش بندی پیشنهادی

معماری شبکه پیشنهادی جهت بخش بندی و تشخیص ناحیه‌ی ماسک در شکل (۶-۳) نشان داده شده است. مدل پیشنهادی از دو شبکه تشکیل شده است که شامل شبکه بخش بندی و شبکه بهبود دهنده^۱ است. در این مدل، یک تصویر به عنوان ورودی دریافت شده و ناحیه ماسک به عنوان خروجی، پیش بینی می شود. زیر شبکه بخش بندی یک شبکه رمزگذار-رمزگشا^۲ با مکانیزم توجه^۳ برای شناسایی مرز بین مناطق ماسک و غیر ماسک است. در مرحله دوم، مرز ماسک با استفاده از یک شبکه کانولوشنی U-Net بهبود می یابد. در ادامه، هر یک از شبکه های بخش بندی و بهبود دهنده به تفصیل شرح داده شده اند.

¹ Refinement Network

² Encoder-Decoder

³ Attention Mechanism



شکل ۳-۶ معماری شبکه پیشنهادی Att-BANet.

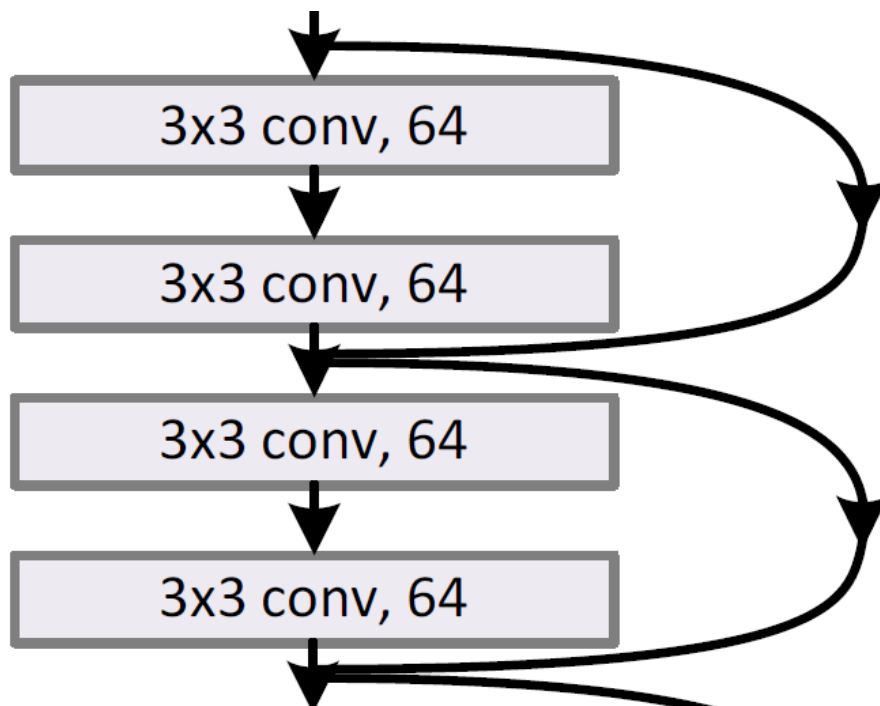
۱-۲-۳ شبکه بخش بندی

شبکه بخش بندی پیشنهادی شامل سه بخش اصلی است که این بخش‌ها شامل یک بخش رمزگذار، یک بخش رمزگشا و یک بخش گیتی با مکانیسم توجه است. قسمت کدگذاری شده از ۷ بلوک^۱ تشکیل شده است. ابتدا تصویر به عنوان ورودی به لایه کانولوشنی 3×3 داده می‌شود. چهار مرحله بعد مشابه بلوک‌های ResNet-34 [۱۰۴] است. همان‌طور که در شکل (۳-۷) مشخص شده است، هر بلوک ResNet-34 از دو لایه کانولوشنی 3×3 تشکیل شده است که ابتدا و انتهای هر لایه با یکدیگر جمع شده و به بلوک بعد انتقال پیدا می‌کند. دو بلوک آخر نیز از بلوک‌های باقی‌مانده^۲ با 512 فیلتر ایجاد شده‌اند. همان‌طور که در شکل (۳-۶) مشخص شده است، بین رمزگذار و رمزگشا پل ساخته شده است، که از سه لایه کانولوشن فراخ^۳ با 512 فیلتر 3×3 تشکیل شده است.

^۱ Block

^۲ Residual Block

^۳ Dilation



شکل ۷-۳ دو بلوک ResNet

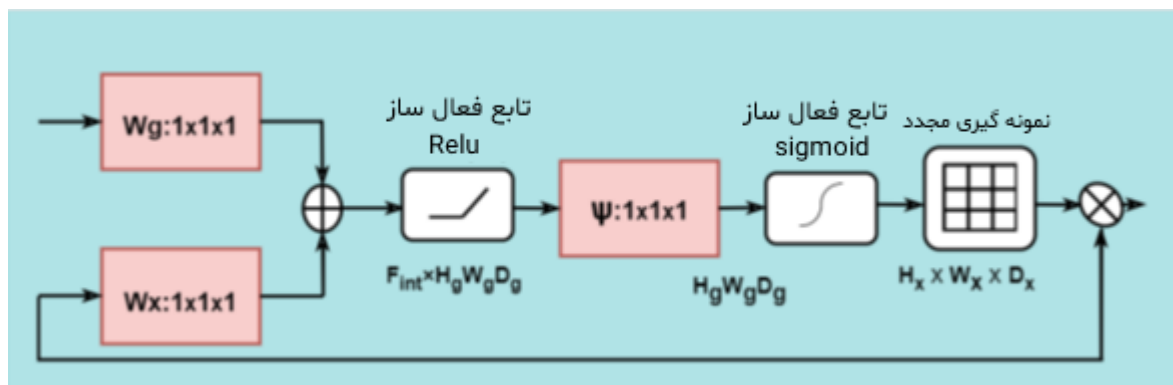
رمزگشا همانند رمزگذار دارای شش مرحله است. دو مرحله اول متشکل از سه لایه کانولوشنی 3×3 هستند که به دنبال آن یک نمونه برداری دوخطی^۱ انجام می شود. در مرحله سوم، از دو لایه کانولوشنی 3×3 استفاده شده است. در مراحل چهارم و پنجم، با هدف جلوگیری از بیش بردازش و همچنین سرعت در فرآیند یادگیری، در ابتدا، از یک لایه نرمال سازی دسته‌ای و تابع فعال ساز Relu به همراه کانولوشنی 3×3 استفاده شده است. مرحله پایانی مشابه مراحل چهارم و پنجم است، با این تفاوت که سه لایه کانولوشنی 3×3 در انتها اضافه شده است. در آخرین لایه کانولوشنی مرحله پایانی، تنها تعداد کانال تصویر به یک کانال تغییر پیدا کرده و تصویر باینری تولید می شود. در سه مرحله ابتدایی قسمت رمزگشا، لایه اول هر مرحله از جمع^۲ لایه متناظر در قسمت رمزگذار و لایه آخر مرحله قبل تشکیل شده است. به منظور بهبود عملکرد شبکه، در ۳ مرحله انتهایی قسمت رمزگشا از گیت توجه^۳ استفاده کردیم. در شکل (۸-۳)، ساختار گیت

^۱ Bilinear Upsampling

^۲ Concatination

^۳ Attention Gate

توجه، نشان داده شده است. گیت توجه، توجه مدل را به مناطق مهم هدایت می‌کند. همچنین به دلیل طراحی سبک وزن، به طور قابل ملاحظه ای قدرت نمایش مدل را بدون افزایش قابل توجهی در هزینه محاسباتی یا تعداد پارامترهای مدل، افزایش می‌دهد.



شکل ۸-۳ ساختار گیت توجه.

۳-۲-۲-۲ شبکه بهبود دهنده

شبکه بازسازی بر اساس معماری رمزگذار-رمزگشای باقیمانده^۱ ساخته شده است شکل (۳-۶). در این معماری، بخش‌های رمزگذار و رمزگشا هر یک، چهار مرحله دارند. هر مرحله دارای یک لایه کانولوشنی با ۶۴ فیلتر ۳×۳ است که به دنبال آن یک نرمال‌سازی دسته ای^۲ و یک تابع فعال‌سازی ReLU وجود دارد. ادغام ویژگی‌های سطح بالا با ویژگی‌های سطح پایین برای بازیابی لبه نواحی ماسک مفید است.

۳-۲-۳ تابع خطا ترکیبی^۳

برای به دست آوردن بخش‌بندی منطقه‌ای با کیفیت بالا و مرزهای دقیق، ما یک تابع خطا ترکیبی را پیشنهاد کرده‌ایم که از سه بخش تشکیل شده است. این سه بخش شامل تابع خطا کانونی^۴، تابع خطا

^۱ Encoder-Decoder Residual

^۲ Batch Normalization

^۳ Hybrid Loss Function

^۴ Focal Loss

¹SSIM و تابع خطا ²IOU هستند.

همان‌طور که در شکل (۳-۶) نشان داده شده است، شبکه در مراحل مختلف خروجی‌های مختلفی نیز دارد که در مجموعه ۸ خروجی دارد. بنابراین تابع خطا نهایی به صورت رابطه‌ی (۳-۱) تعریف می‌شود.

$$Loss = \sum_{k=1}^K Loss^{(k)}$$

$$Loss^{(k)} = Loss_{focal}^{(k)} + Loss_{ssim}^{(k)} + Loss_{iou}^{(k)} \quad (۳-۱)$$

که در آن $Loss^{(k)}$ تابع خطا k امین خروجی است.

تابع خطا کانونی به‌عنوان تابع خطای سطح پیکسلی استفاده شده است. این تابع، مشابه تابع خطای BCE است، با این تفاوت که پارامترهای α و β برای کنترل مشکل عدم تعادل کلاس^۳ استفاده می‌شود. در بخش‌بندی ناحیه ماسک، با توجه به این که بیش‌تر نواحی تصویر، مقدار سیاه دارد. از این‌رو امکان ایجاد عدم تعادل در دو کلاس سیاه و سفید وجود دارد بنابراین، از این تابع خطا استفاده شد.

$$L_{focal} = \begin{cases} -\alpha(1-y')^{\gamma} \log(y'), & y = 1 \\ -(1-\alpha)(y')^{\gamma} \log(1-y'), & y = 0 \end{cases} \quad (۳-۲)$$

در رابطه‌ی (۳-۲)، y تصویر حقیقی^۴ و y' تصویر پیش‌بینی شده توسط شبکه است. همچنین α برابر

۰.۳ و β برابر ۳ قرار داده ایم. مقادیر α و β به صورت تجربی به دست آمده‌اند.

از تابع خطا SSIM برای گرفتن اطلاعات ساختاری و پیش‌بینی بهتر نواحی لبه، در تابع خطای نهایی

استفاده کردیم، رابطه (۳-۳).

¹ Structural Similarity Index Measure (ssim)

² Intersection Over Union

³ Class Imbalance

⁴ Ground Truth

$$L_{SSIM} = 1 - \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3-3)$$

در رابطه (3-3) μ_x ، μ_y و σ_x ، σ_y به ترتیب میانگین و انحراف معیار x و y ، σ_{xy} کواریانس است. همچنین، $C_1 = 0.012$ و $C_2 = 0.032$ است که به منظور جلوگیری از خطای تقسیم بر صفر اضافه شده است.

از تابع خطای IOU نیز به عنوان تابع خطای سطح نگاشت¹ به صورت استفاده کردیم.

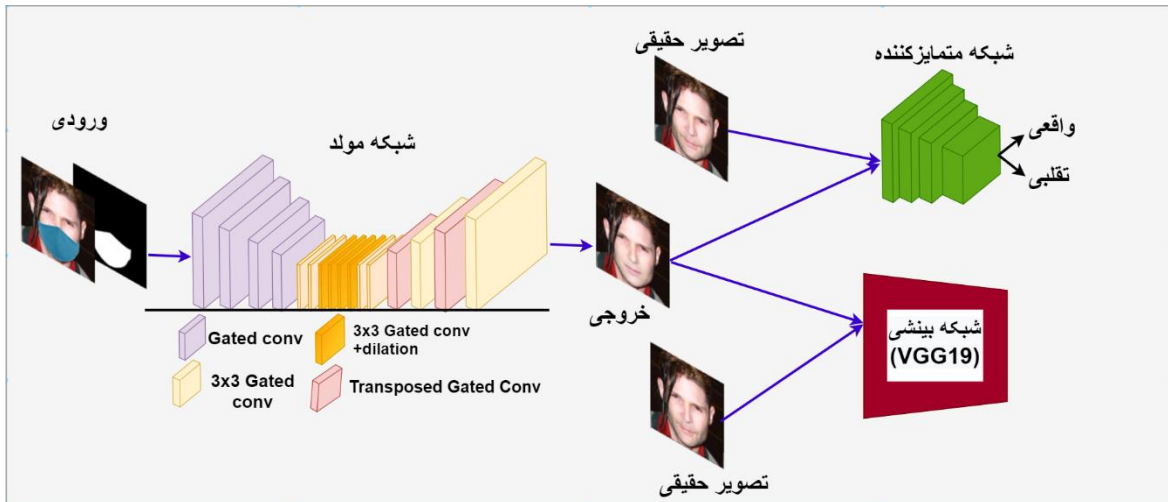
$$L_{IOU} = 1 - \frac{\sum_{r=1}^H \sum_{c=1}^W S(r,c)G(r,c)}{\sum_{r=1}^H \sum_{c=1}^W [S(r,c) + G(r,c) - S(r,c)G(r,c)]} \quad (4-3)$$

در رابطه بالا، $G(r,c)$ تصاویر حقیقی در پیکسل (r,c) است و $S(r,c)$ احتمال پیش‌بینی شده خروجی شبکه است.

3-3 شبکه بازسازی ناحیه ماسک

هدف این بخش، حذف ناحیه ماسک و بازسازی ناحیه پشت آن است. این بازسازی باید به گونه‌ای باشد که ویژگی‌های ظاهری و ساختاری ناحیه بازسازی با تصویر اصلی مطابقت داشته باشد. بدین منظور تصویر اصلی با ماسک و ناحیه بخش‌بندی دودویی حاصل از بخش قبل به عنوان ورودی به شبکه GAN پیشنهادی داده می‌شود و در خروجی تصویر بدون ماسک تولید می‌شود. شبکه پیشنهادی از سه قسمت مولد، متمایز کننده و شبکه بینشی تشکیل شده است که در شکل (3-9) نشان داده شده است.

¹ Map-Level



شکل ۹-۳ معماری کلی شبکه GAN پیشنهادی

۳-۳-۱ شبکه مولد

ساختار شبکه مولد به صورت U شکل است که در قسمت رمزگذار از چهار بلوک با لایه‌های کانولوشنی گیتی^۱ [۱۰۷] استفاده شده است. در بلوک اول، از فیلتر کانولوشنی 7×7 با گام^۲ ۱ استفاده شده و تعداد کانال خروجی روی ۶۴ تنظیم شده است. در بلوک آخر، از فیلتر کانولوشنی 3×3 با گام ۲ استفاده شده و تعداد کانال خروجی روی ۲۵۶ تنظیم شده است.

ارتباط بین رمزگذار و رمزگشا از طریق ۸ بلوک با لایه‌های کانولوشنی گیتی است که در تمامی بلوک‌ها از فیلتر کانولوشنی 3×3 با گام ۱ استفاده شده است. همچنین تعداد ورودی و خروجی فیلترها برابر ۵۱۲ است و به منظور دریافت اطلاعات بیشتر، در ۴ لایه میانی این قسمت، به ترتیب از فراخس^۳ ۲، ۴، ۸ و ۱۶ استفاده شده است.

کدگشا از ۴ بلوک تشکیل شده است. در بلوک اول و سوم از لایه‌های کانولوشنی ترانهاد^۴ جهت انجام عملیات Up-sampling استفاده شده است. در لایه دوم، از فیلتر کانولوشنی 3×3 با گام ۱ استفاده شده

¹ Gated Convolution
² Stride
³ Dilation
⁴ Transposed Convolution

است. در بلوک آخر، تعداد خروجی برابر تعداد کانال تصویر نهایی و تابع فعال‌ساز \tanh به‌منظور قرار دادن خروجی شبکه در بازه $[-1, 1]$ تنظیم می‌شود.

بنابراین مدل مولد، تصویر حقیقی با ماسک I_{input} و تصویر باینری ناحیه ماسک (خروجی شبکه بخش بندی)، I_{mask} را به‌عنوان ورودی دریافت می‌کند و تصویر بازسازی‌شده که ناحیه ماسک در آن حذف شده است را تولید می‌کند $I_{g_{out}}$ (۵-۳).

$$I_{g_{out}} = G_{edit}(I_{input}, I_{mask}) \quad (۵-۳)$$

تابع خطا، در روند آموزش نقش کلیدی دارد. از این رو، این بخش، بخش مهم شبکه‌مولد است. از این رو، برای تولید تصاویر واقع‌بینانه^۱ که بتوان هویت افراد را تمیز داد، از ترکیب توابع $ssim$ و $L1$ به‌عنوان تابع خطای بازسازی استفاده کرده‌ایم. ترکیب دو تابع خطای ذکر شده، باعث می‌شود، هم شباهت ساختاری و هم شباهت پیکسلی تصویر خروجی حفظ شود.

$$\tau_{rc} = \tau_{l_1} + \tau_{ssim} \quad (۶-۳)$$

خطای $L1$ اختلاف پیکسل بین تصویر حقیقی I_{gt} و تصویر تولید شده I_{edit} را حساب می‌کند.

$$\tau_{l_1} = |I_{g_{out}} - I_{gt}| \quad (۷-۳)$$

خطای $ssim$ شباهت ساختاری بین تصویر حقیقی I_{gt} و تصویر تولید شده I_{edit} را محاسبه می‌کند.

$$\tau_{ssim} = SSIM(I_{gt}, I_{g_{out}}) \quad (۸-۳)$$

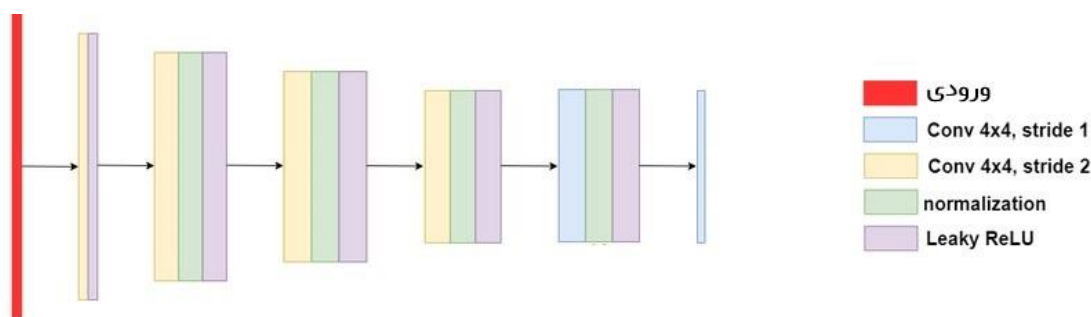
۲-۳-۳ شبکه متمایزکننده

ساختار شبکه متمایزکننده مانند ساختار استفاده شده در شبکه $pix2pix$ [۱۰۸] است. شبکه $pix2pix$ از انواع شبکه‌های GAN شرطی^۲ است که با هدف ترجمه تصویر در تصویر استفاده می‌شود. در این فرآیند،

^۱ Realistic

^۲ Conditional Gan

مولد، مجموعه دادگان دوتایی A و B را دریافت می کند و در نهایت یاد می گیرد چگونه تصاویر به شکل A را به B تبدیل کند. وظیفه شبکه متمایزکننده، مجبور کردن مولد به تولید تصاویر نزدیک به تصویر حقیقی است. شبکه متمایزکننده از پنج لایه کانولوشنی با تابع فعال ساز Leaky Relu تشکیل شده است. اندازه هسته، کانولوشن 4×4 است و در نهایت خروجی یک کاناله و به اندازه ابعاد تصویر حقیقی است. از تابع خطای mse در شبکه متمایزکننده استفاده شده است. در شکل (۱۰-۳) می توان ساختار شبکه متمایزکننده را مشاهده کرد.



شکل ۱۰-۳ معماری کلی شبکه متمایزکننده

۳-۳-۳ شبکه بینشی

در فرآیند آموزش شبکه GAN پیشنهادی، به منظور دستیابی به نمایش ویژگی^۱ نزدیک به تصویر حقیقی، از لایه های میانی شبکه از پیش آموزش دیده^۲ vgg19 [۱۰۹] استفاده کرده ایم. شبکه vgg در عین استخراج ویژگی های مناسب، تعداد پارامتر کمتری نسبت به دیگر شبکه های به روز دارد و زمان آموزش شبکه تا حدی کاهش می یابد. تابع خطای این مرحله، از اندازه گیری فاصله ویژگی بین ۴ لایه میانی شبکه vgg در تصویر حقیقی و تصویر تولید شده، بدست می آید (۳-۹). ϕ_i نداشت ویژگی در لایه i ام است.

¹ Feature Representation

² Pre-Trained

$$\tau_{prec} = \sum_i \|\phi_i(I_{gout}) - \phi_i(I_{gt})\| \quad (9-3)$$

۳-۳-۴ تابع خطای نهایی

در نهایت، توابع خطای عنوان شده در مراحل قبل را باهم جمع می‌کنیم و به هر کدام ضریب (λ) داده می‌شود. که این ضرایب نیز به طور تجربی به دست می‌آید.

$$Loss_{total} = \lambda_{rc1} \tau_{rc1} + \lambda_{rc2} \tau_{rc2} + \lambda_{prec} \tau_{prec} + \lambda_G \tau_G \quad (10-3)$$

در رابطه (۱۰-۳)، τ_G خطای mse تابع مولد است. همچنین مقادیر ضرایب $\lambda_G, \lambda_{rc1}, \lambda_{rc2}, \lambda_{prec}$ به ترتیب برابر ۱، ۱۰، ۱۰۰ و ۱۰۰۰ است.

۳-۳-۵ جمع بندی

در این فصل، به توضیح دو روش پیشنهادی بخش‌بندی ناحیه ماسک و ترمیم ناحیه ماسک پرداخته شد. همچنین مجموعه دادگان MFSD معرفی شد. شبکه بخش‌بندی ناحیه ماسک، از دو قسمت اصلی تشکیل شده است. قسمت اول از دو شبکه تشکیل شده است که شامل شبکه بخش‌بندی و شبکه بهبود دهنده است. در این مدل، یک تصویر به عنوان ورودی دریافت شده و ناحیه ماسک به عنوان خروجی، پیش‌بینی می‌شود. شبکه ترمیم ناحیه ماسک، با هدف حذف ناحیه ماسک و بازسازی ناحیه پشت آن، از ساختار شبکه GAN استفاده شده است. شبکه پیشنهادی ترمیم ناحیه ماسک از سه قسمت مولد، متمایز کننده و شبکه بینشی تشکیل شده است. همچنین در این فصل برای حل مشکل قرار دادن ماسک بر روی صورت، به صورت مصنوعی، که باعث کاهش دقت تشخیص چهره با ماسک می‌شود، دادگان جدیدی تحت عنوان ^۱ MFSD معرفی کردیم که شامل حدود ۱۲۰۰۰ تصاویر با ماسک به همراه بخش‌بندی ناحیه ماسک است.

^۱ Masked Face Segmentation Database

فصل ۴ ارزیابی و بررسی نتایج

۴-۱ مقدمه

در فصل قبل روش پیشنهادی به طور کامل شرح داده شد. در این فصل، در ابتدا نگاهی به مشخصات سیستم می‌پردازیم. سپس دادگان مورد استفاده و پیش‌پردازش‌های مورد نیاز مربوط به آن بررسی می‌شود. سپس به جزئیات مربوط به دو روش پیشنهادی بخش‌بندی ناحیه ماسک و ترمیم ناحیه ماسک خواهیم پرداخت. به منظور ارزیابی روش پیشنهادی، ارزیابی‌های مختلفی از لحاظ کمی و کیفی انجام شده است و روش پیشنهادی با جدیدترین روش‌ها مقایسه شده است. در پایان، میزان بهبود دقت شناسایی چهره، بعد و قبل از بازسازی ناحیه ماسک با استفاده از روش پیشنهادی در مقایسه با جدیدترین روش‌های موجود، بررسی شده است.

۴-۲ مشخصات سیستم

مدل پیشنهادی در این پایان‌نامه به زبان برنامه‌نویسی پایتون و با استفاده از pytorch نسخه ۱.۱۰، پیاده‌سازی شده است. برای آموزش مدل از کارت گرافیک Geforce GTX 3070 استفاده شده است. به دلیل محدودیت سخت‌افزاری، اندازه دسته ۱ انتخاب شد. دوره آموزش ۲۰۰ تنظیم شده است. نرخ یادگیری در ابتدای فرآیند یادگیری برابر 10^{-3} بوده و بعد از تکرار^۱های ۱۰۰۰۰۰، ۷۵۰۰۰، ۵۰۰۰۰ در ۰.۱ ضرب می‌شود. در شکل (۴-۱) خروجی شبکه، در تکرارهای مختلف نشان داده شده است.

^۱ Iteration



شکل ۴-۱ خروجی شبکه در تکرارهای مختلف شبکه.

۴-۳ مجموعه داده‌ها

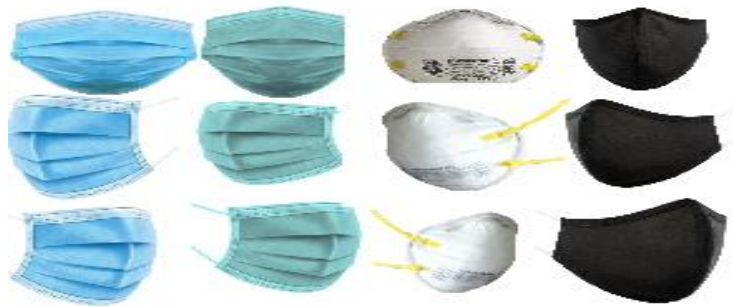
باتوجه به این که شبکه پیشنهادی ترمیم ناحیه ماسک به تصاویر با ماسک و بدون ماسک افراد نیاز دارد و تا کنون هیچ‌یک از مجموعه داده‌ها همچین ویژگی را ندارند، ما مجموعه داده مصنوعی ایجاد کردیم که در ادامه روش تولید آن را بررسی می‌کنیم.

در فرآیند تولید مجموعه داده، از مجموعه دادگان CelebA [۱۱۰] استفاده کردیم. CelebA یک مجموعه داده ویژگی‌های چهره در مقیاس بزرگ با بیش از ۲۰۰ هزار عکس از افراد مشهور است. در این مجموعه دادگان از هر شخص تنها یک عکس وجود دارد.

ابتدا به صورت تصادفی تعداد ۱۵ هزار تصویر چهره افراد مختلف را از مجموعه داده CelebA انتخاب کردیم. تصاویر مجموعه داده CelebA ابعاد ۲۱۸×۱۷۸ دارد که ابعاد پایینی به منظور ورود به شبکه پیشنهادی است. لذا ابتدا به کمک قطعه کدی [۱۱۲]، کیفیت تصاویر بهتر شد و سپس به ابعاد ۵۱۲×۵۱۲ تغییر پیدا کرد. سپس چندین نمونه ماسک، که از نظر شکل، رنگ و سایز باهم تفاوت دارند، را انتخاب (شکل ۴-۲) و با بکارگیری روش [۱۱۱]، ماسک را بر روی تصاویر انتخاب شده از مجموعه دادگان CelebA قرار می‌دهیم (شکل ۴-۳).

در بخش تشخیص چهره، از مجموعه داده FERET استفاده شده است. این مجموعه داده شامل حدود ۱۴۰۰۰ تصویر از حدود ۱۲۰۰ نفر است که برای کاربردهای مختلف، زیرمجموعه‌های مختلفی دارد. به منظور ارزیابی و مقایسه دقت تشخیص چهره، در حالت‌های با ماسک، بدون ماسک و بازسازی شده، از

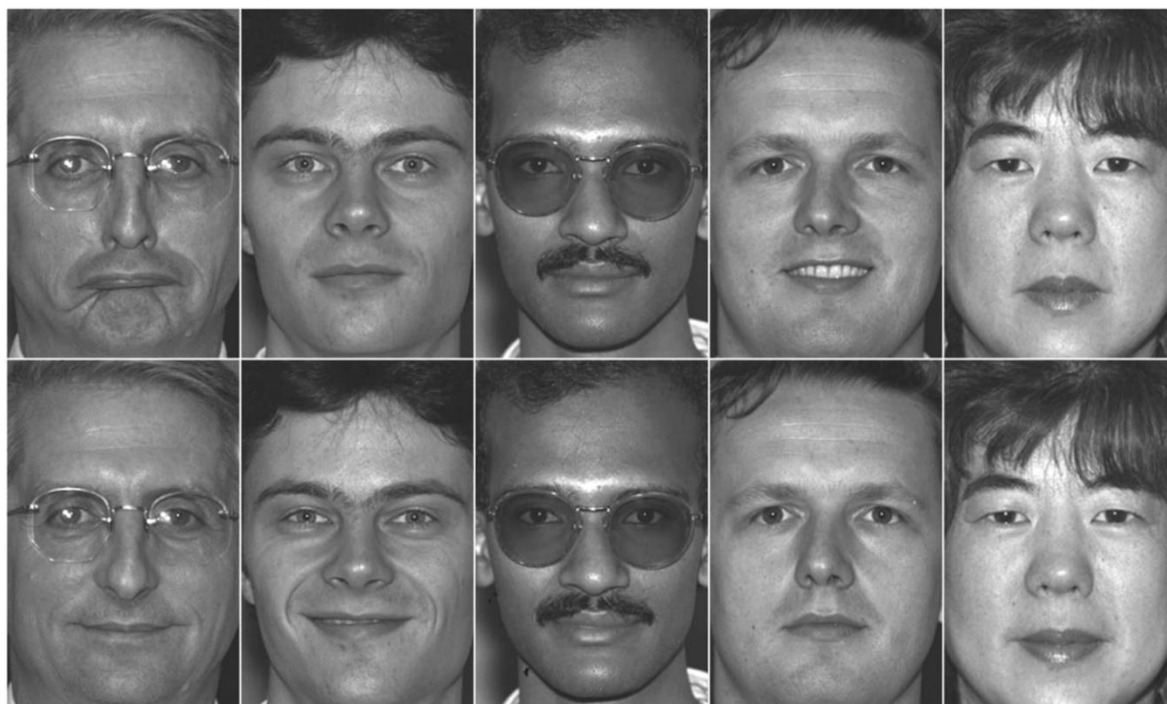
زیرمجموعه‌های FA , FB آن استفاده کردیم. تصاویر دو زیرمجموعه مربوط به افراد یکسان است، با این تفاوت که از نظر روشنایی و حالات چهره، تفاوت دارند شکل (۴-۴).



شکل ۲-۴ انواع مختلف ماسک صورت



شکل ۳-۴ نمونه تصاویر ساخته شده برای آموزش شبکه.



شکل ۴-۴ نمونه تصاویر مجموعه دادگان FERET .

۴-۴ جزئیات پیاده سازی

۴-۴-۱ شبکه بخش بندی ناحیه ماسک

در فرآیند معیارسنجی، ما ۲۸۵۰ تصویر برچسب گذاری شده دستی را به سه گروه تقسیم کردیم که شامل مجموعه آموزشی (۲۰۰۰ تصویر)، مجموعه اعتبارسنجی (۵۵۰ تصویر)، و مجموعه آزمایشی (۳۰۰ تصویر) هستند. همه مدل ها با رمزگذارهای از پیش آموزش دیده ImageNet [۱۰۵] آموزش داده شده اند. همچنین برای مقایسه ی صحیح بین مدل ها، دوره آموزش^۱ ۴۰، اندازه دسته^۲ ۸، بهینه ساز Adam [۱۰۶] برای بهینه سازی پارامترها و نرخ یادگیری 10^{-4} لحاظ شده اند. به منظور افزایش دقت ۹ مدل مقایسه شده، در قسمت کدگذار آن ها از ساختار^۳ های vgg19، resnet50، efficientNet-B7 استفاده کردیم.

¹ Epoch

² Batch Size

³ Backbone

۲-۴-۴ شبکه ترمیم ناحیه ماسک

در در فرآیند آموزش شبکه پیشنهادی، ابتدا مولد، تصویر چهره با ماسک (I_{m_gt}) و تصویر باینری ناحیه ماسک (I_{mask}) را به عنوان ورودی دریافت می‌کند و سپس تصویر بازسازی شده (I_{g_out1})، را در خروجی تولید می‌کند. به منظور قرار دادن ناحیه ماسک بازسازی شده، بر روی تصویر اصلی، عملیات (۴-۱) انجام می‌شود. این عملیات از دو قسمت تشکیل شده است. در قسمت اول، پیکسل‌های تصویر باینری ناحیه ماسک (I_{mask}) را از یک کم کرده تا ناحیه ماسک، سیاه شود. و با ضرب نظیر به نظیر مقادیر پیکسل‌های آن با تصویر چهره با ماسک (I_{m_gt})، ناحیه ماسک در تصویر (I_{m_gt}) سیاه می‌شود. در قسمت دوم، با ضرب نظیر به نظیر مقادیر پیکسل‌های تصویر باینری ناحیه ماسک (I_{mask}) و تصویر بازسازی شده (I_{g_out1})، ناحیه ماسک در تصویر باینری، مشابه خروجی شبکه می‌شود. در نهایت با جمع دو قسمت گفته شده، ناحیه ماسک بازسازی شده، بر روی تصویر اصلی قرار می‌گردد.

$$I_{edit1} = I_{m_gt} * (1 - I_{mask}) + I_{g_out1} * I_{mask} \quad (۱-۴)$$

حال، مجدد تصویر باینری ناحیه ماسک I_{mask} و تصویر I_{edit1} را به شبکه مولد می‌دهیم و خروجی I_{g_out2} تولید می‌شود. این کار باعث بهبود در عملکرد بازسازی ناحیه ماسک می‌شود و تصویر نهایی شبکه مولد تولید می‌شود. در شکل (۴-۵) خروجی شبکه نشان داده شده است. عملیات (۴-۱) را بر روی خروجی نهایی شبکه I_{g_out2} ، انجام می‌دهیم و I_{edit2} ایجاد می‌شود. شبکه مولد، سعی در بازسازی ناحیه ماسک دارد و فیلترهای مراحل مختلف شبکه به گونه‌ای مقدار دهی می‌شود که تنها بر روی ناحیه ماسک اعمال شود. از این رو، در تصویر خروجی شبکه تنها ناحیه ماسک مشخص شده است و بقیه نواحی نامشخص است.



تصویر نهایی (I_{edit2}) خروجی نهایی مولد ($I_{g,out2}$) خروجی اول مولد ($I_{g,out1}$) تصویر با ماسک (I_{m_gt}) تصویر حقیقی (I_{gt})

شکل ۴-۵ تصاویر مراحل مختلف شبکه مولد.

در مرحله بعد، I_{edit2} را به شبکه متمایزکننده می‌دهیم و سپس خطای خروجی آن، L_{fake} را محاسبه می‌کنیم. همچنین برای محاسبه خطای نهایی شبکه متمایزکننده، تصویر بدون ماسک را هم به شبکه متمایزکننده داده و خطای آن، L_{real} را نیز محاسبه می‌کنیم. سپس میانگین این دو خطا را گرفته و به عنوان خطای نهایی شبکه متمایزکننده در نظر می‌گیریم و به منظور یادگیری شبکه، گرادیان حساب می‌شود. در نهایت تابع خطا محاسبه و گرادیان آن جهت بروزرسانی شبکه، اعمال می‌شود.

۴-۵ معیارهای اندازه گیری کمی

۴-۵-۱ شبکه بخش بندی ناحیه ماسک

برای ارزیابی عملکرد شبکه، از پنج معیار استاندارد تقسیم‌بندی معنایی استفاده می‌شود. IoU معیار ارزیابی اولیه است که برای ارزیابی این کار استفاده می‌شود. علاوه بر این، سایر معیارهای استاندارد مانند

دقت^۱ (Pr)، یادآوری^۲ (Re)، امتیاز F1^۳ و دقت پیکسل^۴ (Pa) را محاسبه می‌کنیم. در رابطه (۲-۴) این معیارها به صورت ریاضی فرموله شده‌اند:

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 F1 - score &= \frac{2 * TP}{2 * TP + FP + FN} \\
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 IoU &= \frac{TP}{TP + FP + FN}
 \end{aligned} \tag{۲-۴}$$

۲-۵-۴ شبکه ترمیم ناحیه ماسک

برای اندازه‌گیری کمی عملکرد بازسازی ناحیه ماسک، از سه معیار رایج یعنی PSNR، SSIM و FID [۱۱۳] استفاده شده است.

(۱) PSNR: این معیار برحسب دسی‌بل است و نسبت سیگنال به نویز را محاسبه می‌کند و طبق رابطه زیر محاسبه می‌شود (۳-۴).

$$PSNR_{(x,y)} = \frac{10 \log_{10}[\max(\max(x), \max(y))]^2}{MSE} \tag{۳-۴}$$

^۱ Precision

^۲ Recall

^۳ F1-Score

^۴ Pixel Accuracy

به طوری که MSE مجموع مربعات خطا می باشد.

(۲) SSIM: شباهت ساختاری بین تصویر حقیقی و تصویر تولید شده را محاسبه می کند. در فصل سوم و در رابطه (۳-۳) آورده شده است.

(۳) Frechet Inception (FID): امتیاز فاصله اولیه Frechet یا به اختصار FID معیاری است که فاصله بین بردارهای ویژگی محاسبه شده از مدل Inception v3، برای تصاویر واقعی و تولید شده را محاسبه می کند. نمرات پایین تر نشان می دهد که دو گروه از تصاویر شبیه تر هستند یا آمار مشابه بیشتری دارند. همچنین نمره کامل ۰ نشان دهنده یکسان بودن دو گروه از تصاویر است. فرمول FID در رابطه (۴-۴) آورده شده است. در این فرمول، μ_1 و μ_2 میانگین بردار ویژگی، تصویر حقیقی و خروجی شبکه است. همچنین C_1 و C_2 ماتریس کوواریانس برای بردارهای ویژگی تصاویر واقعی و تولید شده است.

$$FID = \|\mu_1 - \mu_2\|^2 + Tr(C_1 + C_2 - 2 * \text{sqrt}(C_1 * C_2)) \quad (4-4)$$

۴-۶ نتایج

۴-۶-۱ شبکه بخش بندی ناحیه ماسک

در جدول (۱-۴) نتایج نهایی مدل های مختلف و روش پیشنهادی مشاهده می شود. همان طور که در جدول (۱-۴) نشان داده شده است، روش پیشنهادی با ۱۰ مدل به روز، Unet [۱۲۲]، Unet++ [۱۲۳]، MANet [۱۲۴]، Linknet [۱۲۵]، FPN [۱۲۶]، PSPnet [۱۲۷]، PAN [۱۲۸]، Deep LabV3 [۱۲۹]، Deep LavV3++ [۱۳۰] و BASNet [۱۳۱] مقایسه شده است. همچنین روش پیشنهادی نتایج بهتری

نسبت به روش‌های دیگر در معیارهای IOU، F1-score، Precision و Recall دارد. شکل (۴-۶) نتایج

مقایسه بخش‌بندی ناحیه ماسک در روش پیشنهادی و روش‌های پیشرفته دیگر را نشان می‌دهد.

جدول ۴-۱ مقایسه کمی روش‌های مختلف بر روی مجموعه دادگان MFSD.

IOU	F1	Precision	Recall	Acc	ساختار کدگذار	شبکه
۸۹.۵۱۶	۹۴.۳۲۱	۹۱.۵۰۷	۹۷.۶۳	۹۶.۱۱۵	efficientB7	[۱۲۲] Unet
۸۹.۱۸۸	۹۴.۱۲۰	۹۱.۴۶۳	۹۷.۳۱۸	۹۶.۰۳۲	ResNet50	
۸۹.۲۰۸	۹۴.۱۷۵	۹۱.۲۸	۹۷.۵۳۳	۹۶.۰۱۲	Vgg19	
۸۹.۵۲۴	۹۴.۳۳۶	۹۱.۵۶۵	۹۷.۵۶۷	۹۶.۱۲۹	efficientB7	[۱۲۳] Unet++
۸۸.۲۵۶	۹۳.۷۵	۹۰.۹۵۲	۹۷.۱۰۶	۹۶.۶۷۸	ResNet50	
۸۹.۲۴۲	۹۴.۱۹۵	۹۱.۶۰۲	۹۷.۲۲۵	۹۶.۹۹۸	Vgg19	
۸۹.۵۸۷	۹۴.۳۷۳	۹۱.۶۶۷	۹۷.۵۲۹	۹۷.۱۵۹	efficientB7	[۱۲۴] MAnet
۸۹.۶۱	۹۴.۰۷۸	۹۱.۲۱۸	۹۷.۴۰۵	۹۵.۹۳۴	ResNet50	
۸۹.۲۴۲	۹۴.۲۱۲	۹۱.۴۵	۹۷.۴۱۸	۹۵.۹۹۹	Vgg19	
۸۹.۴۸۴	۹۴.۳۲	۹۱.۶۲۹	۹۷.۴۸۲	۹۶.۱۱۳	efficientB7	[۱۲۵] LinkNet
۸۹.۰۱۱	۹۴.۰۴	۹۱.۴۲۷	۹۷.۱۵۲	۹۵.۹۰۴	ResNet50	
۸۹.۰۶۴	۹۴.۰۸۲	۹۱.۶۵۲	۹۶.۹۷۸	۹۵.۹۰۵	Vgg19	
۸۹.۱۸۷	۹۴.۱۹۸	۹۱.۱۴۸	۹۷.۶۴۵	۹۵.۹۹۶	ResNet50	[۱۲۶] FPN
۸۹.۳۳۵	۹۴.۲۵	۹۱.۴۹۳	۹۷.۴۹	۹۶.۰۴۳	Vgg19	

۸۹.۰۸۵	۹۴.۰۲۹	۹۱.۳۹۹	۹۷.۱۹۶	۹۵.۹۴۹	efficientB7	[۱۲۷] PSPNet
۸۸.۸۳۴	۹۳.۹۳۴	۹۱.۰۹	۹۷.۳۵۲	۹۵.۸۷۵	Vgg19	
۸۹.۳۱۲	۹۴.۲۴۱	۹۴.۵۲۲	۹۷.۴۳۵	۹۶.۰۳۳	ResNet50	[۱۲۸] PAN
۸۸.۴۶۶	۹۴.۲۹۸	۹۱.۵۲	۹۷.۵۹	۹۶.۱۰۷	Vgg19	
۸۸.۹۱۲	۹۳.۹۶	۹۱.۱۹	۹۷.۲۳۴	۹۵.۸۸۹	ResNet50	[۱۲۹] DeepLabV3
۸۹.۲۴۲	۹۴.۲۱۲	۹۱.۴۵	۹۷.۴۴۸	۹۵.۹۹۹	efficientB7	[۱۳۰] DeepLabV3++
۸۸.۴۶۳	۹۴.۳۲۰	۹۱.۴۶	۹۷.۶۶۵	۹۶.۰۸۳	ResNet50	
۹۱.۶۸	۹۵.۲۰۲	۹۴.۷۵۴	۹۶.۲۳۵	۹۷.۸۷۸	–	[۱۳۱] BASNet
۹۳.۸۱۴	۹۶.۶۰۶	۹۷.۱۶۴	۹۷.۷۷۴	۹۷.۷۲۵		روش پیشنهادی



شکل ۴-۶ مقایسه روش پیشنهادی و روش‌های بروز دیگر

۲-۶-۴ شبکه ترمیم ناحیه ماسک

ما ۱۰۰۰ تصویر ماسک‌دار از دادگان CelebA را به کمک روش مطرح شده در فصل قبل، تولید کرده‌ایم. سپس از روش‌های [۱۱۴]، [۱۱۵] و [۱۱۶]، برای بازسازی ناحیه ماسک و در نهایت مقایسه با روش پیشنهادی، استفاده کرده‌ایم. مقاله [۱۱۴] با معرفی شبکه متخاصم مولد خود توجه^۱، سعی در بهبود

^۱ Self-Attention Generative Adversarial Network

بازسازی تصاویر دارد. مقاله [۱۱۵] از طریق پیش‌بینی نقاط شاخص چهره و شبکه GAN، عملیات بازسازی تصویر را انجام می‌دهد. مقاله [۱۱۶] از نظر ساختار شبکه GAN، از دو متمایزکننده استفاده کرده است که یکی از متمایزکننده‌ها کل ناحیه صورت را بررسی می‌کند و دیگری مختص ناحیه ماسک است. همان‌طور که در جدول (۲-۴) مشاهده می‌شود، روش پیشنهادی عملکرد بهتری در تمامی معیارهای ارزیابی، بجز معیار FID، نسبت به بقیه روش‌های مقایسه شده داشته است. منبع کد روش [۱۱۶] در دسترس نبود. از این‌رو روش آزمایش بر اساس این پژوهش انجام شد. در شکل (۷-۴) خروجی روش‌های مقایسه شده و روش پیشنهادی قابل مشاهده است. همان‌طور که در شکل (۷-۴) مشخص است، در تصویر خروجی شبکه پیشنهادی، ناحیه ماسک به خوبی بازسازی شده است، یعنی چهره طبیعی‌تر به نظر می‌رسد و ساختار چهره به خوبی حفظ شده است. از این‌رو، خروجی رویکرد ما به طور قابل توجهی بهتر عمل می‌کند.

جدول ۲-۴ مقایسه کمی روش پیشنهادی با روش‌های مختلف.

روش	FID	SSIM	PSNR
[114]	۱۲/۸۰۸	۰/۸۹۳	۲۴/۷۰۹
[115]	۱۲/۴۶۵	۰/۸۹۷	۲۵/۵۶۹
[116]	۳/۵۴۸	۰/۸۶۴	۲۶/۱۹
روش پیشنهادی	۷/۵۴۸	۰/۹۱۸	۲۷/۲۴۸



شکل ۴-۷ مقایسه نتیجه روش پیشنهادی با سایر روش های ترمیم ناحیه ماسک.

۴-۷ بررسی تأثیر بازسازی ناحیه ماسک بر روی دقت شناسایی چهره

هدف اصلی بازسازی ناحیه ماسک، بهبود دقت شناسایی چهره است، لذا در این قسمت قصد داریم تأثیر ماسک زدن را بر چندین روش شناسایی چهره بررسی کنیم. سپس به بررسی بهبود دقت پس از اعمال برخی از روش‌های بازسازی ناحیه ماسک، از جمله روش پیشنهادی می‌پردازیم. آزمایشات بر روی روش‌های شناسایی چهره، Face-Net [۱۱۷] و Arc-Face [۱۱۸]، انجام شده است. به منظور شناسایی چهره، تصویر ۱۶۰۰ نفر از مجموعه دادگان FERET را انتخاب کرده و دو روش شناسایی چهره بر روی آن آزمایش شده است. در مرحله بعد ماسک بر روی تصاویر قرار گرفته و دوباره عملیات شناسایی انجام شده است. همانطور که در جدول ۳-۴ مشاهده می‌شود، دقت تشخیص کاهش یافته است. در نهایت عملیات بازسازی بر روی تصاویر تست انجام شده و مجدداً دقت شناسایی چهره محاسبه شد.

جدول ۳-۴ مقایسه تأثیر روشهای مختلف ترمیم ناحیه ماسک، بر روی روشهای مختلف شناسایی چهره.

شناسایی چهره		تصویر	تصویر	[۱۱۴]	[۱۱۵]	روش
		بدون ماسک	با ماسک			پیشنهادی
	Face-Net	۹۸/۶۵	۶۷/۴۸	۸۵/۵۵	۸۸/۶۸	۹۵/۲۷
	ArcFace	۹۸/۸	۷۸/۱۱	۸۰/۲۷	۸۶/۴	۹۲/۹

۴-۸ جمع بندی

در این فصل ابتدا مشخصات سیستمی که شبکه اجرا شده است را بررسی کردیم. سپس جزئیات مربوط به پیاده سازی، انجام شده و معیارهای ارزیابی استفاده شده در هر دو شبکه پیشنهادی بررسی شدند. با انجام پیاده سازیها و آزمایشات مشاهده شد که استفاده از شبکه GAN پیشنهادی، به واسطه استفاده از توابع خطای متنوع و ساختارهای استفاده شده، منجر به بهبود کیفیت ترمیم ناحیه ماسک شده است. از طرفی هدف نهایی انجام عملیات ترمیم ناحیه ماسک، بهبود دقت شناسایی چهره است. از این رو در بخش هفتم به بررسی اثر ترمیم تصویر در تشخیص چهره پرداخته شد. همچنین روش پیشنهادی با روشهای به روز دیگر مقایسه شد.

فصل ۵: نتیجه‌گیری و کارهای آینده

۵-۱ نتیجه‌گیری

یکی از مشکلاتی که سیستم‌های تشخیص چهره با آن مواجه هستند، تشخیص چهره با ماسک است. در این پایان‌نامه، برخلاف روش‌های مرسوم که سعی در بهبود سیستم تشخیص چهره دارند، برای حل مشکل تشخیص چهره با ماسک، سعی در حذف ماسک و بازسازی ناحیه آن، به‌گونه‌ای که علاوه بر تولید تصویر باکیفیت، ویژگی‌های هویتی ناحیه زیر ماسک، یعنی بینی و دهان حفظ شود. به منظور ناحیه ماسک، در این مقاله، دو شبکه و یک مجموعه دادگان پیشنهاد شد. در شبکه اول، ناحیه ماسک بخش‌بندی می‌شود و تصویری باینری، خروجی می‌دهد. در این قسمت، شبکه Att-BANet معرفی شد. این شبکه از دو قسمت تشکیل شده است که در قسمت اول با به کارگیری گیت توجه، ویژگی‌های بهتری استخراج شد. قسمت دوم در بهبود خروجی قسمت اول، یعنی تفکیک بهتر نواحی مرزی ناحیه ماسک نقش موثری دارد. تابع خطای پیشنهادی در این بخش از توابع خطای SSIM، Focal loss و IOU تشکیل شده است. نتایج تجربی نشان می‌دهد که شبکه پیشنهادی دقت بالاتری نسبت به روش‌های بروز دیگر دارد.

در شبکه دوم با هدف ترمیم ناحیه ماسک، از شبکه GAN استفاده شده است. معماری پیشنهادی از سه قسمت مولد، متمایزکننده و شبکه بینشی تشکیل شده است. شبکه مولد با تولید تصاویر بدون ماسک جدید، سعی در گول زدن شبکه متمایزکننده دارد. از طرف دیگر شبکه متمایزکننده در هر مرحله تصاویر واقعی و جعلی را از هم متمایز می‌کند. با تکرار این مراحل، در نهایت شبکه یاد می‌گیرد، تصاویر با ماسک را به تصاویر بدون ماسک تبدیل کند. شبکه بینشی نیز در این مراحل کمک می‌کند تصاویر واقعی‌تری تولید شود. در فصل ۴، عملکرد روش پیشنهادی از جنبه‌های مختلف کمی، کیفی و میزان تأثیر بر روی دقت شناسایی چهره با کارهای مشابه مقایسه شده است. مقایسه کیفی و هم کمی نشان می‌دهد که در مقایسه با روش‌های بروز دیگر، مدل ما قادر به تولید نتایج با کیفیت ادراکی بالا بوده و عملکرد بهتری نسبت به روش‌های مقایسه شده نیز دارد. همچنین دقت شناسایی چهره با تصاویر تولید شده توسط روش

پیشنهادی در چندین روش شناسایی چهره با سایر روش‌های ترمیم چهره مقایسه شده است که روش پیشنهادی دقت شناسایی چهره بالاتری را ارائه کرده است. همچنین به منظور توانمندسازی شبکه نسبت به انواع مختلف ماسک، مجموعه دادگان MFSD معرفی شد. این مجموعه دادگان شامل، حدود ۱۲۰۰۰ تصویر با ماسک به همراه بخش‌بندی ناحیه ماسک متناظر با تصاویر افراد است.

۵-۲ کارهای آینده

با توجه به این‌که مجموعه دادگانی وجود ندارد که تصویر با ماسک و بدون ماسک افراد جمع‌آوری شده باشد، از این‌رو، تمامی روش‌های بازسازی صورت، برای آموزش شبکه GAN از تصاویر مصنوعی استفاده می‌کنند. این امر باعث می‌شود در برخی از مواقع عملیات بازسازی تصویر به خوبی انجام نشود. لذا جمع‌آوری مجموعه دادگانی که این نیاز را برآورده کند، می‌تواند باعث افزایش کیفیت تصویر خروجی و در نهایت بالا رفتن دقت سیستم‌های تشخیص چهره شود. علاوه بر دقت سیستم‌های تشخیص چهره، سرعت پردازش آن نیز مهم است. شبکه‌های GAN و در کل شبکه‌هایی که تعداد پارامتر زیادی دارند، باعث می‌شود سرعت پردازش به‌طور چشمگیری پایین بیاید. لذا انتظار می‌رود طراحی و پیاده‌سازی شبکه‌ای که علاوه بر افزایش دقت، سرعت پردازش تصویر را افزایش دهد، نیز در آینده پیشنهاد شود.



- [1] M. Wang and W. Deng. Deep face recognition: A survey. arXiv preprint arXiv:1804.06655, 2018
- [2] [2]Adjabi I., Ouahabi A., Benzaoui A., Taleb-Ahmed A. (2020). Past, Present, and Future of Face Recognition: A Review. *Electronics*, 9(8), 1188. <https://doi.org/10.3390/electronics9081188>
- [3] Zeng D, Veldhuis R, Spreeuwens L (2020) A survey of facerecognition techniques under occlusion. arXiv
- [4] Walid Hariri. Efficient masked face recognition method during the covid-19 pandemic. arXiv:2105.03026, 2021
- [5] Li Y, Sun B, Wu T, Wang Y (2016) Face detection with end-to-end integration of a convnet and a 3d model. In: European conference on computer vision. Springer, Cham, pp 420–436
- [6] Ge, S., et al.: Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Trans. Circ. Syst. Video Technol.* 30(10), 3387–3397 (2020)
- [7] Walid Hariri. Efficient masked face recognition method during the covid-19 pandemic. arXiv:2105.03026, 2021
- [8] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” in Proc. ICML, 2017
- [9] Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* 1991, 3, 71–86. [CrossRef]
- [10] Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA
- [11] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J. C. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, “Deep learning for understanding faces: Machines may be just as good, or better, than humans,” *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 66–83, 2018
- [12] Timo Ahonen, Abdenour Hadid, and Matti Pietik`ainen. Face recognition with local binary patterns. In European conference on computer vision, pages 469–481. Springer, 2004.
- [13] David G Lowe. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision, volume 2, pages 1150–1157. IEEE, 1999.
- [14] Peter N Belhumeur, Jo~ao P Hespanha, and David J Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Technical report, Yale University New Haven United States, 1997.
- [15] Zafaruddin GM, Fadewar HS. Face recognition: a holistic approach review. Paper presented at: 2014 International Conference on Contemporary Computing and Informatics (IC3I); 2014; Mysore, India:175-178. doi:101109/IC3I.2014.7019610.

- [16] Zhao, L.; Yang, Y.H. Theoretical analysis of illumination in PCA-based vision systems. *Pattern Recognit.* 1999
- [17] Pentland, A.; Moghaddam, B.; Starner, T. View-Based and modular eigenspaces for face recognition. In *Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 21–23 June 1994; pp. 84–91.
- [18] Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* 1997, 19, 711–720. [CrossRef]
- [19] Bartlett, M.; Movellan, J.; Sejnowski, T. Face Recognition by Independent Component Analysis. *IEEE Trans. Neural Netw.* 2002, 13, 1450–1464. [CrossRef]
- [20] Stone, J.V. Independent component analysis: An introduction. *Trends Cogn. Sci.* 2002, 6, 59–64. [CrossRef]
- [21] <https://behsanandish.com/learning/generalipmv/%D9%81%DB%8C%D9%84%D8%AA%D8%B1-%DA%AF%D8%A7%D8%A8%D9%88%D8%B1-gabor-filter-%DA%86%DB%8C%D8%B3%D8%AA%D8%9F/>
- [22] Abhishree, T.M.; Latha, J.; Manikantan, K.; Ramachandran, S. Face recognition using Gabor Filter based feature extraction with anisotropic diffusion as a preprocessing technique. *Procedia Comput. Sci.* 2015, 45, 312–321. [CrossRef]
- [23] Zehani, S.; Ouahabi, A.; Oussalah, M.; Mimi, M.; Taleb-Ahmed, A. Trabecular bone microarchitecture characterization based on fractal model in spatial frequency domain imaging. *Int. J. Imaging Syst. Technol.* accepted.
- [24] Ouahabi, A. *Signal and Image Multiresolution Analysis*, 1st ed.; ISTE-Wiley: London, UK, 2012
- [25] Guetbi, C.; Kouame, D.; Ouahabi, A.; Chemla, J.P. Methods based on wavelets for time delay estimation of ultrasound signals. In *Proceedings of the 1998 IEEE International Conference on Electronics, Circuits and Systems*, Lisbon, Portugal, 7–10 September 1998; pp. 113–116.
- [26] Ferroukhi, M.; Ouahabi, A.; Attari, M.; Habchi, Y.; Taleb-Ahmed, A. Medical video coding based on 2nd-generation wavelets: Performance evaluation. *Electronics* 2019, 8, 88. [CrossRef]
- [27] Wang, M.; Jiang, H.; Li, Y. Face recognition based on DWT/DCT and SVM. In *Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM)*, Taiyuan, China, 22–24 October 2010; pp. 507–510.
- [28] Bookstein, F.L. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* 1989, 11, 567–585. [CrossRef]
- [29] Shih, F.Y.; Chuang, C. Automatic extraction of head and face boundaries and facial features. *Inf. Sci.* 2004, 158, 117–130. [CrossRef]
- [30] Zobel, M.; Gebhard, A.; Paulus, D.; Denzler, J.; Niemann, H. Robust facial feature localization by coupled features. In *Proceedings of the 2000 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, Grenoble, France, 26–30 March 2000; pp. 2–7.

- [31] Wiskott, L.; Fellous, J.M.; Malsburg, C.V.D. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* 1997, 19, 775–779. [CrossRef]
- [32] Xue, Z.; Li, S.Z.; Teoh, E.K. Bayesian shape model for facial feature extraction and recognition. *Pattern Recognit.* 2003, 36, 2819–2833. [CrossRef]
- [33] Pentland, A.; Moghaddam, B.; Starner, T. View-Based and modular eigenspaces for face recognition. In *Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 21–23 June 1994; pp. 84–91.
- [34] Lades, M.; Vorbuggen, J.C.; Buhmann, J.; Lange, J.; Malsburg, C.V.D.; Wurtz, R.P.; Konen, W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 1993, 42, 300–311. [CrossRef]
- [35] Wiskott, L. Phantom faces for face analysis. *Pattern Recognit.* 1997, 30, 837–846. [CrossRef]
- [36] Lades, M.; Vorbuggen, J.C.; Buhmann, J.; Lange, J.; Malsburg, C.V.D.; Wurtz, R.P.; Konen, W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 1993, 42, 300–311. [CrossRef]
- [37] Duc, B.; Fischer, S.; Bigun, J. Face authentication with Gabor information on deformable graphs. *IEEE Trans. Image Process.* 1999, 8, 504–516. [CrossRef] [PubMed]
- [38] Kotropoulos, C.; Tefas, A.; Pitas, I. Frontal face authentication using morphological elastic graph matching. *IEEE Trans. Image Process.* 2000, 9, 555–560. [CrossRef] [PubMed]
- [39] Kumar, D.; Garaina, J.; Kisku, D.R.; Sing, J.K.; Gupta, P. Unconstrained and Constrained Face Recognition Using Dense Local Descriptor with Ensemble Framework. *Neurocomputing* 2020.
- [40] Ait Aouit, D.; Ouahabi, A. Nonlinear Fracture Signal Analysis Using Multifractal Approach Combined with Wavelet. *Fractals Complex Geom. Patterns Scaling Nat. Soc.* 2011, 19, 175–183.
- [41] Davies, E.R. Introduction to texture analysis. In *Handbook of Texture Analysis*; Mirmehdi, M., Xie, X., Suri, J., Eds.; Imperial College Press: London, UK, 2008; pp. 1–31.
- [42] Benzaoui, A.; Hadid, A.; Boukrouche, A. Ear biometric recognition using local texture descriptors. *J. Electron. Imaging* 2014, 23, 053008.
- [43] Ahonen, T.; Hadid, A.; Pietikäinen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006, 28, 2037–2041.
- [44] Rodriguez, Y.; Marcel, S. Face authentication using adapted local binary pattern histograms. In *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, Graz, Austria, 7–13 May 2006; pp. 321–332.
- [45] Boutella, E.; Harizi, F.; Bengherabi, M.; Ait-Aoudia, S.; Hadid, A. Face verification using local binary patterns and generic model adaptation. *Int. J. Biomed.* 2015, 7, 31–44.

- [46] Benzaoui, A.; Boukrouche, A. Face Analysis, Description, and Recognition using Improved Local Binary Patterns in One Dimensional Space. *J. Control Eng. Appl. Inform. (CEAI)* 2014, 16, 52–60.
- [47] Ahonen, T.; Rathu, E.; Ojansivu, V.; Heikkilä, J. Recognition of Blurred Faces Using Local Phase Quantization. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR)*, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
- [48] Lei, Z.; Ahonen, T.; Pietikainen, M.; Li, S.Z. Local Frequency Descriptor for Low-Resolution Face Recognition. In *Proceedings of the 9th Conference on Automatic Face and Gesture Recognition (FG)*, Santa Barbara, CA, USA, 21–25 March 2011; pp. 161–166.
- [49] Kannala, J.; Rahtu, E. BSIF: Binarized statistical image features. In *Proceedings of the 21th International Conference on Pattern Recognition (ICPR)*, Tsukuba, Japan, 11–15 November 2012; pp. 1363–1366.
- [50] Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* 2015, 61, 85–117.
- [51] Deng, L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans. Signal Inf. Process.* 2014, 3, 1–29.
- [52] Deng, L.; Yu, D. Deep Learning: Methods and Applications. *Found. Trends Signal Process.* 2014, 7, 197–387.
- [53] Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* 2010, 11, 3371–3408.
- [54] Salakhutdinov, R.; Hinton, G. Deep Boltzmann machines. In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, Clearwater, FL, USA, 16–19 April 2009; pp. 448–455.
- [55] Sutskever, I.; Martens, J.; Hinton, G. Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, Bellevue, WA, USA, 28 June–2 July 2011; pp. 1017–1024.
- [56] Sun, Y.; Wang, X.; Tang, X. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition* Columbus, OH, USA, 23–28 June 2014; pp. 1891–1898.
- [57] Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 8–13 December 2014; pp. 1988–1996.
- [58] Sun, Y.; Wang, X.; Tang, X. Deeply learned face representations are sparse, selective, and robust. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015; pp. 2892–2900.
- [59] Sun, Y.; Liang, D.; Wang, X.; Tang, X. DeepID3: Face Recognition with Very Deep Neural Networks. *arXiv2015*, arXiv:1502.00873v1.

- [60] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 2nd International Conference on Learning Representations (ICLR), Banff, AB, Canada, 14–16 April 2014.
- [61] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
- [62] Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the 2015 British Machine Vision Conference, Swansea, UK, 7–10 September 2015; pp. 41.1–41.12.
- [63] Masi, I.; Tran, A.T.; Hassner, T.; Leksut, J.T.; Medioni, G. Do we really need to collect millions of faces for effective face recognition? In Proceedings of the 2016 European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 579–596.
- [64] Liu, W.; Wen, Y.; Yu, Z.; Yang, M. Large-margin softmax loss for convolutional neural networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 507–516.
- [65] Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
- [66] Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 499–515.
- [67] Wu, X.; He, R.; Sun, Z.; Tan, T. A Light CNN for Deep Face Representation with Noisy Labels. *IEEE Trans. Inf. Forensics Secur.* 2018, 13, 2884–2896.
- [68] D. Zeng, R. N. J. Veldhuis, L. Spreeuwers, A survey of face recognition techniques under occlusion, arXiv preprint arXiv:2006.11366
- [69] Aleš Leonardis and Horst Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000
- [70] Marios Savvides, Ramzi Abiantun, Jingu Heo, S Park, Chunyan Xie, and BVK Vijayakumar. Partial & holistic face recognition on frgc-ii data using support vector machine. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 48–48. IEEE, 2006.
- [71] Aleix M Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):748–763, 2002.
- [72] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. *IEEE Transactions on Neural Networks*, 16(4):875–886, 2005.

- [73] Kazuhiro Hotta. Robust face recognition under partial occlusion based on support vector machine with local gaussian summation kernel. *Image and Vision Computing*, 26(11):1490–1498, 2008.
- [74] Jeongin Seo and Hyeyoung Park. A robust face recognition through statistical learning of local features. In *International Conference on Neural Information Processing*, pages 335–341. Springer, 2011.
- [75] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009.
- [76] Meng Yang, Lei Zhang, Simon CK Shiu, and David Zhang. Gabor feature based robust representation and classification for face recognition with gabor occlusion dictionary. *Pattern Recognition*, 46(7):1865–1878, 2013.
- [77] acopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan. Deep face recognition: A survey. In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pages 471–478. IEEE, 2018.
- [78] Erjin Zhou, Zhimin Cao, and Qi Yin. Naive-deep face recognition: Touching the limit of lfw benchmark or not? *arXiv preprint arXiv:1501.04690*, 2015.
- [79] Jiang-Jing Lv, Xiao-Hu Shao, Jia-Shui Huang, Xiang-Dong Zhou, and Xi Zhou. Data augmentation for face recognition. *Neurocomputing*, 230:184–196, 2017.
- [80] Daniel S´aez Trigueros, Li Meng, and Margaret Hartnett. Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. *Image and Vision Computing*, 79:99–108, 2018.
- [81] Zhaohua Chen, Tingrong Xu, and Zhiyuan Han. Occluded face recognition based on the improved svm and block weighted lbp. In *Image Analysis and Signal Processing (IASP), 2011 International Conference on*, pages 118–122. IEEE, 2011.
- [82] Rui Min, Abdenour Hadid, and Jean-Luc Dugelay. Improving the recognition of faces occluded by facial accessories. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 442–447. IEEE, 2011.
- [83] Yizhang Xia, Bailing Zhang, and Frans Coenen. Face occlusion detection based on multi-task convolution neural network. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*, pages 375–379. IEEE, 2015.
- [84] Visualization (CGIV), 2010 Seventh International Conference on, pages 142–146. IEEE, 2010.
- [85] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.
- [86] Stan Z Li, Xin Wen Hou, Hong Jiang Zhang, and Qian Sheng Cheng. Learning spatially localized, parts-based representation. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

- [87] Weitao Wan and Jiansheng Chen. Occlusion robust face recognition based on mask learning. In 2017 IEEE International Conference on Image Processing (ICIP), pages 3795–3799. IEEE, 2017.
- [88] Lingxiao He, Haiqing Li, Qi Zhang, Zhenan Sun, and Zhaofeng He. Multiscale representation for partial face recognition under near infrared illumination. In Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, pages 1–7. IEEE, 2016.
- [89] Shengcai Liao, Anil K Jain, and Stan Z Li. Partial face recognition: Alignment-free approach. *IEEE Transactions on pattern analysis and machine intelligence*, 35(5):1193–1205, 2013.
- [90] Jeong-Seon Park, You Hwa Oh, Sang Chul Ahn, and Seong-Whan Lee. Glasses removal from facial image using recursive error compensation. *IEEE transactions on pattern analysis and machine intelligence*, 27(5):805–811, 2005.
- [91] Fernando De La Torre and Michael J Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117–142, 2003.
- [92] Sanja Fidler, Danijel Skocaj, and Ales Leonardis. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):337–350, 2006.
- [93] Ale’s Leonardis and Horst Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1):99–118, 2000.
- [94] Lele Cheng, Jinjun Wang, Yihong Gong, and Qiqi Hou. Robust deep auto-encoder for occluded face recognition. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1099–1102. ACM, 2015.
- [95] Fang Zhao, Jiashi Feng, Jian Zhao, Wenhan Yang, and Shuicheng Yan. Robust lstm-autoencoders for face de-occlusion in the wild. *IEEE Transactions on Image Processing*, 27(2):778–790, 2018.
- [96] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [97] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.
- [98] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [99] Y. Li, S. Liu, J. Yang, and M.-H. Yang, “Generative face completion,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3911–3919
- [100] J. Zhao, L. Xiong, J. Li, and et al., “3d-aided dual-agent gans for unconstrained face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*
- [101] S. Ge, C. Li, S. Zhao, and D. Zeng, “Occluded face recognition in wild by identity-diversity inpainting,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2020

- [102] Non-Deterministic Face Mask Removal Based On 3D Priors arXiv preprint arXiv:2202.09856, 2022
- [103] D.Chiang: Detecting Faces and Determine Whether People are Wearing Mask (2020). : <https://github.com/AIZOOTech/FaceMaskDetection>
- [104] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.770{778 (2016).
- [105] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision 115(3), 211{252 (2015)
- [106] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- [107] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In Proceedings of the IEEE International Conference on Computer Vision, pages 4471–4480, 2019.
- [108] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5967-5976.
- [109] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. ICLR, 2015, pp. 1-14.
- [110] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 3730-3738.
- [111] aqeelanwar.: MaskTheFace . <https://github.com/aqeelanwar/MaskTheFace>
- [112] LynnHo.: HD CelebA Cropper. <https://github.com/LynnHo/HD-CelebA-Cropper>
- [113] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 6626-6637.
- [114] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, “Self-attention generative adversarial networks,” in International conference on machine learning. PMLR, 2019, pp. 7354–7363.
- [115] Yang Yang, Xiaojie Guo, Jiayi Ma, Lin Ma, and Haibin Ling, “Lafin: Generative landmark guided face inpainting,”
- [116] arXiv preprint arXiv:1911.11394, 2019. Nizam Ud Din, Kamran Javed, Seho Bae, and Juneho Yi, “A novel gan-based network for unmasking of masked face,” IEEE Access, vol. 8, pp. 44276–44287, 2020.
- [117] Jose, E., Greeshma, M., Haridas, M.T., and Supriya, M.: ‘Face recognition based surveillance system using facenet and MTCNN on jetson tx2’, in Editor (Ed.)^(Eds.): ‘Book Face recognition based surveillance system using facenet and mtcnn on jetson tx2’ (IEEE, 2019, edn.), pp. 608-613

- [118] Deng, J., Guo, J., Xue, N., and Zafeiriou, S.: ‘Arcface: Additive angular margin loss for deep face recognition’, in Editor (Ed.)^(Eds.): ‘Book Arcface: Additive angular margin loss for deep face recognition’ (2019, edn.), pp. 4690-4699
- [119] M. Wang and W. Deng, “Deep face recognition: A survey,” 2018, arXiv:1804.06655. [Online]. Available: <https://arxiv.org/abs/1804.06655>.
- [120] Xu, S. (2021). Survey of occluded and unoccluded face recognition. In Advances in Computer, Communication and Computational Sciences, pages 1001–1014. Springer.
- [121] Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. Electronics 2020,9, 1188.
- [122] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234{241 (2015). Springer.
- [123] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: U-net++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3{11. Springer.
- [124] Fan, T., Wang, G., Li, Y., Wang, H.: Ma-net A multi-scale attention network for liver and tumor segmentation. IEEE Access 8, 179656{179665} (2020).
- [125] Chaurasia, A., Culurciello, E.: Linknet: Exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP), pp. 1{4 (2017)
- [126] Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117{2125 (2017) IEEE
- [127] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881{2890 (2017).
- [128] Li, H., Xiong, P., An, J., Wang, L.: Pyramid attention network for semantic segmentation. arXiv preprint arXiv:1805.10180 (2018).
- [129] Chen, L.C., Papandreou, G., Schro, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017).
- [130] Chen, L.-C., Zhu, Y., Papandreou, G., Schro, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801{818 (2018)
- [131] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7479{7489 (2019)

مرتب بر اساس حروف الفبای فارسی

عنوان فارسی	عنوان انگلیسی
آشکارسازی چهره	Face Detection
آنتروپی متقاطع	Cross-Entropy
ادراک هویت	Identity Perception
استخراج ویژگی	Feature Extraction
اطلاعات فضایی	Spatial Information
اطلاعات جهانی	Global Information
امتیاز F1	F1-Score
الگوی باینری محلی	Local Binary Pattern (LBP)
الگوی باینری محلی یک بعدی	One-Dimensional Local Binary Pattern (LDLBP)
انتقال	Project
اندازه دسته	Batch Size
انسداد	Occlusion
بازسازی	Reconstruction
بازنمایی	Representation
بخش بندی	Segmentation
برچسب گذاری	Labeling
بلوک	Block
بلوک های باقی مانده	Residual Block
بهینه سازی تصادفی	Stochastic Optimization
بیش برزش	Overfitting
بینشی	Perceptual
پراکنده	Sparse

Sliding Window	پنجره کشویی
Pre-Training	پیش‌آموزش دیده
Loss Function	تابع خطا
Hybrid Loss Function	تابع خطا ترکیبی
Verification	تأیید چهره
Discrete Cosine Transform (DCT)	تبدیل کسینوس گسسته
Discrete Wavelet Transform (DWT)	تبدیل موجک گسسته
Principal Component Analysis (PCA)	تجزیه و تحلیل تفکیک خطی
Linear Discriminant Analysis (LDA)	تجزیه و تحلیل مؤلفه اصلی
Independent Component Analysis (ICA)	تحلیل مؤلفه های مستقل
Face Alignment	تراز کردن چهره
Image Inpainting	ترمیم تصاویر
Face Recognition	تشخیص چهره
Occlusion Aware Face Recognition	تشخیص چهره آگاه از ناحیه انسداد
Occluded Face Recognition	تشخیص چهره تحت انسداد
Ground Truth	تصویر حقیقی
Face Matching	تطبیق چهره
Maximum A Posteriori (Map) Adaptation	تطبیق حداکثری پسینی
Graph Matching	تطبیق گراف
Elastic Bunch Graph Matching (EBGM)	تطبیق گراف دسته‌ای الاستیک
Elastic Graph Matching (EGM)	تطبیق نمودار الاستیک
Variations Of Posture	تغییر وضعیت
Data Augmentation	تقویت داده
Iteration	تکرار
Face Patches	تکه‌های صورت
Discriminator	تمایزگر
Descriptor	توصیف‌گر
Local Frequency Descriptor (LFD)	توصیف‌گر بسامد محلی مناسب
Global Descriptor	توصیف‌گر کلی
Multi-Keypoint Descriptors (MKD)	توصیف‌گرهای چند نقطه کلیدی
Concatination	جمع

Large Margin	حاشیه بزرگ
Maxout	حداکثری
Crawler	خزنده
Center Loss Function	تابع خطا مرکزی
Precision	دقت
Pixel Accuracy	دقت پیکسل
Epoch	دوره آموزش
Stacked Sparse Denoising Autoencoder	رمزگذار پشته‌ای حذف نویز پراکنده
Auto Encoder (AE)	رمزگذار خودکار
Traditional Autoencoder	رمزگذار خودکار سنتی
Variational Autoencoder	رمزگذار خودکار متغیر
Encoder-Decoder Residual	رمزگذار - رمزگشای باقیمانده
Encoding	رمزگذاری
Decoding	رمزگشایی
Holistic Methods	روش های کل نگر
Local Texture Descriptors-Based Methods	روش های مبتنی بر توصیف کننده های بافت - محلی
Backbone	ساختار
Dense Local Graph Structure (D-LGS)	ساختار نمودار محلی مترکم
Map-Level	سطح نگاشت
Landmark Characteristics Localized	شاخص بومی سازی شده
Refinement Network	شبکه بهبود دهنده
(IDGAN) GAN Identity-Diversity	تنوع هویت
Recurrent Neural Network (RNN)	شبکه عصبی تکرار شونده
Convolutional Neural Network (CNN)	شبکه عصبی کانولوشنال
Multiscale Double Supervision Convolutional Neural Network (MDSCNN)	شبکه کانولوشنی نظارت دو گانه چند مقیاسی
Generative Adversarial Network	شبکه مولد تخصصی
Convolutional Neural Networks (CNN)	شبکه های کانولوشنی
Aligned Face	صورت تراز
Partial Face	صورت ناقص
Sparse Representation Classifier (SRC)	طبقه بند نمایش پراکنده

Dilation	فراخش
Subspace	فضای فرعی
Stacked Convolution	کانولوشن پشت‌های
Transposed Convolution	کانولوشنی ترانهاده
Gated Convolution	کانولوشنی گیتی
Chi-Square	کای دو
Vector Quantization (VQ)	کوانتیزاسیون برداری
Local Phase Quantization (LPQ)	کوانتیزه‌سازی فاز محلی
Stride	گام
Heap Graph	گراف پشت‌های
Attention Gate	گیت توجه
Hidden Layer	لایه‌های پنهان
Boltzman Machine (BM)	ماشین بولتزمن
Function Contrastive Loss	تابع خطای متضاد کاهش
Bounding Box	مختصات
Synthesize	مصنوعی
Benchmarking	معیار سنجی
Eigenvalues	مقادیر ویژه
Scale	مقیاس
Attention Mechanism	مکانیزم توجه
Match	منطبق
Generator	مولد
Unconstrained	نامحدود
Correct Classification Rate(CCR)	نرخ طبقه‌بندی صحیح
Batch Normalization	نرمال‌سازی دسته‌ای
K-Nearest Neighbor	نزدیک‌ترین همسایه
Landmark	نقاط عطف
Self Organizing Maps	نقشه‌های خودسازماندهی
Feature Representation	نمایش ویژگی
Image Indexing	نمایه‌سازی تصویر
Upsampling Bilinear	نمونه برداری دوخطی

Realistic	واقع بینانه
Low-Resolution	وضوح پایین
Binarized Statistical Image Features (BSIF)	ویژگی‌های تصویر آماری باینری شده
Handcraft Features	ویژگی‌های دستی
Recall	یادآوری
Subspace Learning	یادگیری زیرفضایی

مرتب بر اساس حروف الفبای انگلیسی

عنوان فارسی	عنوان انگلیسی
صورت تراز	Aligned Face
گیت توجه	Attention Gate
مکانیزم توجه	Attention Mechanism
رمزگذار خودکار	Auto Encoder (AE)
ساختار	Backbone
نرمال‌سازی دسته ای	Batch Normalization
اندازه دسته	Batch Size
معیار سنجی	Benchmarking
نمونه برداری دوخطی	Upsampling Bilinear
ویژگی‌های تصویر آماری باینری شده	Binarized Statistical Image Features (BSIF)
بلوک	Block
ماشین بولتزمن	Boltzman Machine (BM)
مختصات	Bounding Box
تابع خطا مرکزی	Center Loss Function
کای دو	Chi-Square
جمع	Concatination
تابع خطای متضاد کاهش	Function Contrastive Loss
شبکه‌های کانولوشنی	Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN)	شبکه عصبی کانولوشنال
Correct Classification Rate(CCR)	نرخ طبقه بندی صحیح
Crawler	خزنده
Cross-Entropy	آنترپی متقاطع
Data Augmentation	تقویت داده
Decoding	رمزگشایی
Dense Local Graph Structure (D-LGS)	ساختار نمودار محلی متراکم
Descriptor	توصیف‌گر
Dilation	فراخشی
Discrete Cosine Transform (DCT)	تبدیل کسینوس گسسته
Discrete Wavelet Transform (DWT)	تبدیل موجک گسسته
Discriminator	تمایزگر
Eigenvalues	مقادیر ویژه
Elastic Bunch Graph Matching (EBGM)	تطبیق گراف دسته‌ای الاستیک
Elastic Graph Matching (EGM)	تطبیق نمودار الاستیک
Encoder-Decoder Residual	رمزگذار- رمزگشای باقیمانده
Encoding	رمزگذاری
Epoch	دوره آموزش
F1-Score	امتیاز F1
Face Alignment	تراز کردن چهره
Face Detection	آشکارسازی چهره
Face Matching	تطبیق چهره
Face Patches	تکه‌های صورت
Face Recognition	تشخیص چهره
Feature Extraction	استخراج ویژگی
Feature Representation	نمایش ویژگی
Gated Convolution	کانولوشنی گیتی
Generative Adversarial Network	شبکه مولد تخصصی
Generator	مولد
Global Descriptor	توصیف‌گر کلی
Global Information	اطلاعات جهانی

Graph Matching	تطبیق گراف
Ground Truth	تصویر حقیقی
Handcraft Features	ویژگی‌های دستی
Heap Graph	گراف پشته‌ای
Hidden Layer	لایه‌های پنهان
Holistic Methods	روش‌های کل‌نگر
Hybrid Loss Function	تابع خطا ترکیبی
Identity Perception	ادراک هویت
(IDGAN) GAN Identity-Diversity	تنوع هویت
Image Indexing	نمایه‌سازی تصویر
Image Inpainting	ترمیم تصاویر
Independent Component Analysis (ICA)	تحلیل مؤلفه‌های مستقل
Iteration	تکرار
K-Nearest Neighbor	نزدیک‌ترین همسایه
Labeling	برچسب‌گذاری
Landmark	شاخص بومی سازی شده
Landmark Characteristics Localized	شاخص بومی سازی شده
Large Margin	حاشیه بزرگ
Linear Discriminant Analysis (LDA)	تجزیه و تحلیل مؤلفه اصلی
Local Binary Pattern (LBP)	الگوی باینری محلی
Local Frequency Descriptor (LFD)	توصیف‌گر بسامد محلی مناسب
Local Phase Quantization (LPQ)	کوانتیزه‌سازی فاز محلی
Local Texture Descriptors-Based Methods	روش‌های مبتنی بر توصیف‌کننده‌های بافت - محلی
Loss Function	تابع خطا
Low-Resolution	وضوح پایین
Map-Level	سطح نگاشت
Match	منطبق
Maximum A Posteriori (Map) Adaptation	تطبیق حداکثری پسینی
Maxout	حداکثری
Multi-Keypoint Descriptors (MKD)	توصیفگرهای چند نقطه کلیدی

Multiscale Double Supervision Convolutional Neural Network (MDSCNN)	شبکه کانولوشنی نظارت دوگانه چند مقیاسی
Occluded Face Recognition	تشخیص چهره تحت انسداد
Occlusion	انسداد
Occlusion Aware Face Recognition	تشخیص چهره آگاه از ناحیه انسداد
One-Dimensional Local Binary Pattern (LDLBP)	الگوی باینری محلی یک بعدی
Overfitting	بیش‌برازش
Partial Face	صورت ناقص
Perceptual	بینشی
Precision	دقت
Pixel Accuracy	دقت پیکسل
Pre-Training	پیش‌آموزش دیده
Principal Component Analysis (PCA)	تجزیه و تحلیل تفکیک خطی
Project	انتقال
Realistic	واقع‌بینانه
Recall	یادآوری
Reconstruction	بازسازی
Recurrent Neural Network (RNN)	شبکه عصبی تکرارشونده
Refinement Network	شبکه بهبود دهنده
Representation	بازنمایی
Residual Block	بلوک‌های باقی‌مانده
Scale	مقیاس
Segmentation	بخش‌بندی
Self Organizing Maps	نقشه‌های خودسازماندهی
Sliding Window	پنجره کشویی
Sparse	پراکنده
Sparse Representation Classifier (SRC)	طبقه‌بند نمایش پراکنده
Spatial Information	اطلاعات فضایی
Stacked Convolution	کانولوشن پشت‌پشته‌ای
Stacked Sparse Denoising Autoencoder	رمزگذار پشت‌پشته‌ای حذف نویز پراکنده
Stochastic Optimization	بهینه‌سازی تصادفی

Stride	گام
Subspace	فضای فرعی
Subspace Learning	یادگیری زیرفضایی
Synthesize	مصنوعی
Traditional Autoencoder	رمزگذار خودکار سنتی
Transposed Convolution	کانولوشنی ترانهاده
Unconstrained	نامحدود
Variational Autoencoder	رمزگذار خودکار متغیر
Variations Of Posture	تغییر وضعیت
Vector Quantization (VQ)	کوانتیزاسیون برداری
Verification	تأیید چهره

Abstract

Face recognition is one of the most active research fields of machine vision and pattern recognition, which is widely used in many fields such as identification, access control, and public security. The development of deep learning techniques, access to large-scale face datasets and production of high-power processing systems, the performance of face recognition systems has improved significantly. However, face recognition systems still do not perform very satisfactorily when facing challenges such as pose variety, different illumination, low resolution, and occlusion. Image occlusion is one of the most challenging face recognition problems. In this challenge, the appearance of the face changes significantly and the identity features of the face are lost, which makes it difficult to recognize the face. Using a mask blocks a large part of the face, including the nose and chin. Hence, it is considered the most difficult challenge of facial occlusion. One of the methods of face detection in case of obstruction is face reconstruction and restoration. In recent years, GAN-based networks have performed very well in the field of image restoration and reconstruction. In the proposed method of this thesis, first the mask area is detected and then this part is reconstructed and the face without mask is created. Then the reconstructed face is given to the face recognition system. The proposed deep network architecture in face reconstruction is based on GAN. The output of the network, in addition to producing a high-quality image, preserves the identity features of the area under the mask, that is, the nose and mouth. Therefore, the proposed method increases the accuracy by about 30% compared to the masked image and by about 8% compared to the compared methods. Also, the quantitative criteria of SSIM, PSNR and FID indicate the proper performance of the proposed method in the mask area restoration.

Keywords: Face recognition, image inpainting, Deep Learning, GAN network



Shahrood University of
Technology

Faculty of Computer Engineering
M.Sc. Thesis in Artificial Intelligence Engineering

Masked Face Recognition Using Deep Learning

By: Sadjad Rezvani

Supervisors:
Dr. Mansoor Fateh

Advisors:
Dr. Hossein Khosravi

August 2022