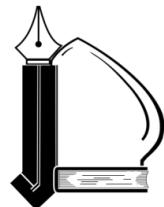


بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشکده مهندسی



دانشگاه فردوسی مشهد
دانشکده مهندسی کروه کامپیوتر
آزمایشگاه بینایی ماشین

دانشگاه فردوسی مشهد

دانشکده مهندسی

گروه مهندسی کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد
هوش مصنوعی

تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت

استاد راهنما: دکتر حمید رضا پور رضا

پژوهش و نگارش: سید سجاد اعمی

۱۳۹۷ مهرماه



Ferdowsi University of Mashhad
m v l a b . u m . a c . i r

آزمایشگاه بینایی ماشین

تعهدنامه

اینجانب سید سجاد اعمی دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر دانشکده مهندسی دانشگاه فردوسی مشهد نویسنده پایان‌نامه تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت تحت راهنمایی دکتر حمید رضا پور رضامتعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود و یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه فردوسی مشهد می‌باشد و مقالات مستخرج با نام "دانشگاه فردوسی مشهد" و یا "Ferdowsi University of Mashhad" به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تاثیرگذار بوده‌اند در مقالات مستخرج از رساله رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است، اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ
امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه فردوسی مشهد می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

تقدیم به

پدر و مادر عزیزم

و همه کسانی که درست اندیشیدن را به من آموختند.

سپاس‌گزاری

سپاس خداوند یکتای عزتمندی که رحمت و دانش او در سراسر گیتی گسترده شده، آسمان‌ها و زمین همه از آن اوست و علم و دانش حقیقی را برو هر که بخواهد موهبت می‌فرماید. رحمت و لطف او را بی‌نهایت سپاس می‌گوییم چرا که فهم و درک مطالب این پژوهش را بر من ارزانی داشت و مرا به این اصل رساند که علم و ایمان دو بال یک پروازند. توفیق تلاش به من داد و هر بار که خطا کردم فرصتی دوباره، تا با امید، تلاشی تازه را آغاز کنم و به خواست او به نتیجه‌ی مطلوب نائل آیم. به‌راستی که همه چیز از آن اوست و همه‌چیز به خواست اوست.

بسمه تعالیٰ

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

چکیده

در سال های اخیر، به دلیل استفاده از یادگیری عمیق، تشخیص چهره شاهد پیشرفت های چشمگیری بوده است. با این حال، استراتژی داده محور همچنین یک چالش را به همراه می آورد: تصاویر ارسال شده به سیستم های تشخیص چهره همیشه برای تشخیص مناسب نیستند. به دلیل سناریوهای غیرقانونی و عدم تشخیص چهره، ممکن است چهره هایی با وضوح کم، چهره های تار در حرکت، صورت های مسدود و حتی وصله های پس زمینه وجود داشته باشد. متأسفانه، از آنجا که موتور تشخیص چهره قبل‌اچنین چهره های بی کیفیتی را نمیده است، احتمالاً تصمیمات نامعقولی در مورد این ها می گیرد این پایان نامه با محوریت موضوع تشخیص بی‌درنگ چهره در محیط‌های کنترل نشده می‌باشد که از دو بخش اصلی یافتن چهره و شناسایی چهره تشکیل شده است. روش پیشنهادی در بخش شناسایی چهره می‌باشد. هدف نهایی افزار یک عینک هوشمند برای افراد نایبنا می‌باشد. هنگامی که فرد نایبنا از عینک استفاده کرده و در محیط‌های عمومی به راه رفتن پردازد، دوربینی که روی عینک نصب شده است، چهره افرادی که در زاویه دید آن قرار دارند را بررسی کرده و در صورت یافتن یک چهره آشنا، فرد مورد نظر شناسایی شده و نامش از طریق صدا برای فرد نایبنا خوانده می‌شود.

در پیاده سازی این سامانه که باید در مکان‌های عمومی، معابر پیاده و محیط‌های کنترل نشده مورد استفاده قرار بگیرد، چند چالش مهم مانند نورپردازی غیریکنواخت، انسداد، تاری خارج از مرکز دوربین و زاویه نامطلوب چهره نسبت به دوربین وجود دارد. از طرفی این سامانه باید به صورت بی‌درنگ رفتار نماید. زیرا فرصت زیادی برای شناسایی فردی که در خیابان در کنار دوربین می‌گذرد، وجود ندارد. از سوی دیگر به دلیل اجرای پردازش‌ها بر روی پردازشگر تلفن همراه، باید محدودیت‌هایی مانند منابع را در نظر گرفت و الگوریتم استفاده شده باید دارای کمترین پیچیدگی زمانی و حافظه باشد. در این پایان نامه محوریت اصلی مطالعات بر روی مرحله شناسایی چهره در شرایط بدون محدودیت است. علاوه بر موارد گفته شده در بالا، ما فرض کردیم که برای آموزش الگوریتم یادگیری عمیق، داده های محدودی از هر دسته در اختیار داریم. برای مقابله با این چالش با استفاده از شبکه های مولد تخصصی تصاویر بیشتری تولید می‌کنیم. برای این منظور ما از ۴ شبکه تولید کننده در کنار هم استفاده کرده ایم. با بررسی نتایج حاصل از این پژوهش بر روی تصاویر مجموعه داده LFW و YouTube Faces، دقت روش پیشنهادی ما به ترتیب برابر با ۹۶٪ و ۹۴٪ شد که دقت بالاتری نسبت به روش‌های مشابه می‌باشد.

فهرست مطالب

ت	فهرست جداول
ث	فهرست تصاویر
ج	فهرست نمادها
۱	۱ روش پیشنهادی
۲	۱.۱ مقدمه
۲	۲.۱ پیش پردازش
۳	۱.۲.۱ همسان سازی بافت نگار
۳	۲.۲.۱ یافتن چهره
۴	۳.۲.۱ تراز کردن تصویر
۵	۳.۱ دسته بندی
۵	۱.۳.۱ مدل پیشنهادی پایه
۱۰	۲.۳.۱ آموزش شبکه جهت استخراج ویژگی
۱۰	۳.۳.۱ تابع ضرر
۱۴	۴.۱ فناوری های استفاده شده
۱۵	۲ ارزیابی روش پیشنهادی
۱۶	۱.۲ مقدمه
۱۶	۲.۲ معیار ارزیابی

۱۷	مجموعه داده	۳.۲
۱۷	مجموعه داده‌های آموزش	۱.۳.۲
۱۸	مجموعه داده‌های آزمون	۲.۳.۲
۱۹	پیکربندی الگوریتم	۴.۲
۲۰	نتایج آزمون	۵.۲
۲۲	نتیجه‌گیری و پیشنهادات	۳
۲۳	مقدمه	۱.۳
۲۳	بحث و نتیجه‌گیری	۲.۳
۲۴	پیشنهادات	۳.۳
۲۵	منابع و مأخذ	

فهرست جداول

- | | | |
|-----|--|----|
| ۱.۱ | مقایسه و ارزیابی برخی از معماری شبکه های رایج در زمینه بینایی ماشین | ۶ |
| ۱.۲ | مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره | ۱۹ |
| ۲.۲ | دقیق معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره | ۲۰ |

فهرست تصاویر

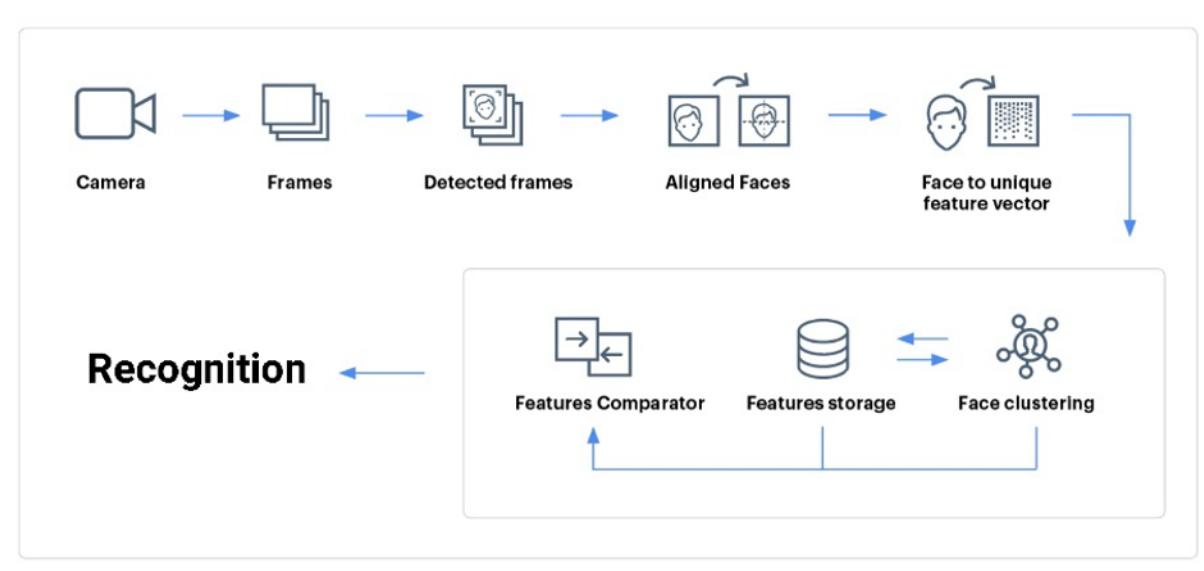
۱.۱	نمای کلی از روش پیشنهادی [۴].	۲
۲.۱	نتیجه اعمال یکسان سازی بافت‌نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می‌باشند [۲].	۳
۳.۱	نمونه از خروجی الگوریتم یافتن چهره [۳] retina	۴
۴.۱	به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد [۴].	۵
۵.۱	مدل پایه مبتنی بر لایه‌های کانولوشن و لایه توجه.	۶
۶.۱	یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن‌های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می‌کند..	۸
۷.۱	معماری ماژول [۴] Attention Shuffle	۸
۸.۱	رویکردهای مختلف هم ترازی چهره [۴].	۱۲
۹.۱	رویکردهای مختلف هم ترازی چهره [۴].	۱۴
۱۰.۲	منحنی FROC مقایسه با روش پیشنهادی برای مجموعه داده E-Ophtha-MA [۴].	۲۱
۱۰.۲	منحنی FROC مقایسه با روش پیشنهادی برای مجموعه داده DIARETDB1 [۴].	۲۱

فهرست نمادها

فصل ۱

روش پیشنهادی

همان طور که در فصل گذشته شرح داده شد، شناسایی بی‌درنگ چهره در محیط بدون محدودیت و با دقت بالا با چالش‌های بسیاری همراه است. همچنین دقت بالا و زمان پردازش سریع باهم در تقابل هستند. علاوه بر این‌ها، فرض کمبود داده آموزشی نیز چالش بزرگی محسوب می‌شود. بنابراین در این فصل تلاش می‌کنیم تا روشی برای تشخیص دقیق‌تر و بی‌درنگ چهره توسط شبکه عصبی عمیق در تصاویر بدون محدودیت پیشنهاد دهیم. در ابتدا مرحله پیش‌پردازش شرح داده می‌شود. در قسمت بعد، رویکرد استفاده شده مبتنی بر شبکه‌های پیچشی به منظور استخراج ویژگی از تصاویر چهره شرح داده خواهد شد. نمای کلی روش ارائه شده در شکل ۱.۱ خلاصه شده است که در ادامه هر یک تشریح خواهد شد.



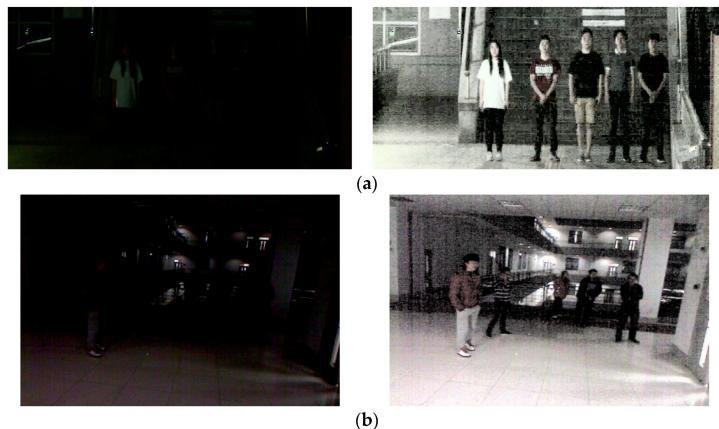
شکل ۱.۱: نمای کلی از روش پیشنهادی [؟].

۲.۱ پیش‌پردازش

بیشتر الگوریتم‌های شناسایی چهره نیاز به اعمال پیش‌پردازش‌هایی بر روی تصویر ورودی دارند. در این روش، پیش‌پردازش شامل همسان‌سازی بافت‌نگار به منظور افزایش تباين، یافتن چهره و تراز کردن تصویر می‌باشد. در ادامه به شرح مراحل پیش‌پردازش می‌پردازیم.

۱.۲.۱ همسان سازی بافت‌نگار

یکی از روش‌های بهبود تصویر، افزایش تباين تصویر است. یکی از روش‌های افزایش تباين تصویر، تکنیک یکنواخت سازی بافت‌نگار است. بطوریکه مقادیر پیکسل‌های تصویر را طوری تغییر می‌دهد تا کل بازه ممکن را تسخیر کند و ایده اساسی آن، نگاشت مقادیر شدت سطوح روشنایی از طریق یکتابع توزیع انتقال است. این عمل باعث افزایش تباين تصویر می‌شود که به معنای بهبود کیفیت تصویر و افزایش دقیق پردازش‌های بعدی است. نمونه ای از این روش را از مقاله [۲] در شکل ۲.۱ مشاهده می‌نمایید.



شکل ۲.۱: نتیجه اعمال یکسان سازی بافت‌نگار بر روی یک تصویر تاریک. تصاویر ورودی در سمت چپ و خروجی در سمت راست می‌باشند [۲].

۲.۲.۱ یافتن چهره

یکی دیگر از مراحل معمول پیش‌پردازش که در طی فرآیند شناسایی چهره انجام می‌شود، مکان یابی و یافتن چهره در تصاویر می‌باشد. برای این منظور از یک روش مبتنی بر شبکه عصبی پیچشی که توسط Deng و همکاران [۳] در سال ۲۰۱۹ معرفی شده است استفاده می‌کنیم. در این روش برای آموزش شبکه پیچشی از یکتابع ضرر مبتنی بر یادگیری چندکاره^۱ استفاده شده است که به صورت زیر می‌باشد.

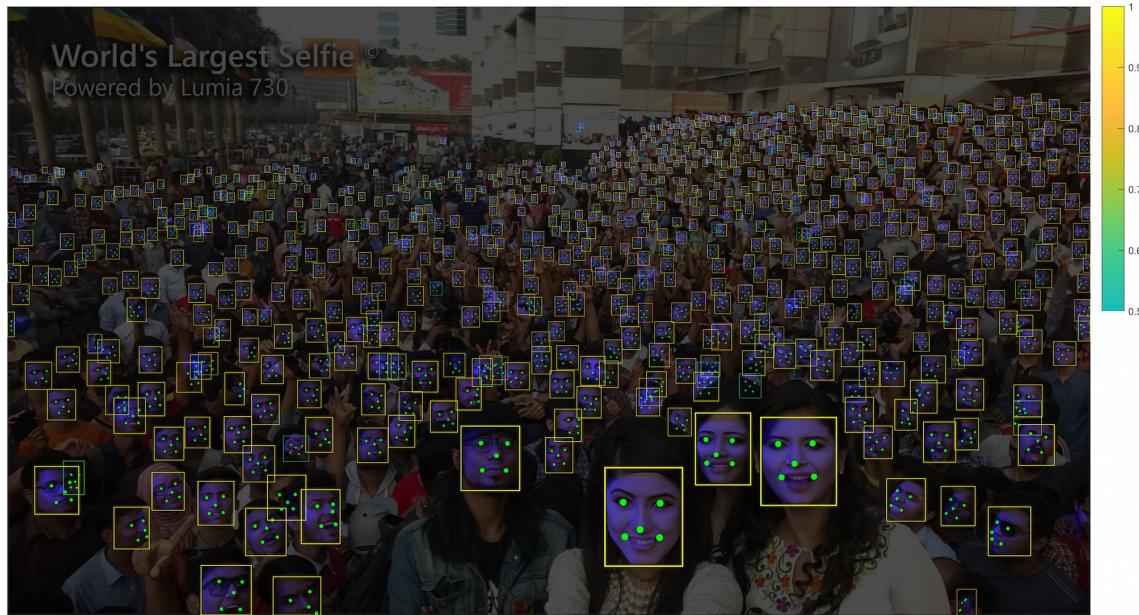
$$L = L_{cls} + L_{box} + L_{pts} + L_{pixel} \quad (1.1)$$

که در آن L_{cls} تابع ضرر مربوط به یافتن یا عدم یافتن چهره می‌باشد. L_{box} تابع ضرر مربوط به مکان چهره می‌باشد. همچنین

تابع ضرر مربوط به یافتن نقاط ویژه روی اجزای چهره می‌باشد، و L_{pixel} تابع ضرر مربوط به یافتن یک مدل سه بعدی مبتنی

¹Multi Task Learning

بر مش از روی چهره می‌باشد. استفاده از تابع ضرر مبتنی بر یادگیری چند کاره، کمک می‌نماید تا فضای مسئله محدود تر شود و الگوریتم بهینه سازی مورد نظر زودتر به سمت نقطه بهینه همگرا شود. ما برای رسیدن به خروجی بی درنگ، این روش را بر روی معماری MobileNet V3 پیاده سازی کرده و آموزش دادیم. نمونه ای از خروجی این روش را در شکل ۳.۱ مشاهده می‌کنید.



شکل ۳.۱: نمونه از خروجی الگوریتم یافتن چهره retina [۳].

۳.۲.۱ تراز کردن تصویر

در ادامه روند پیش‌پردازش، نوبت به تراز کردن تصویر چهره^۱ می‌رسد. پس از یافتن چهره، به تصویر ورودی مناسب شبکه نزدیک تر می‌شویم، اما پس از تراز کردن تصویر جهت آموزش شبکه، بهبود دقت نهایی مشهود است. بدین منظور با استفاده از یک تبدیل غیرخطی، تصویر چهره را به گونه‌ای می‌چرخانیم که چشم، ها در راستای خط افقی قرار بگیرند. روش‌های کلی برای تشخیص چهره از زاویه‌ی روبرو به خوبی عمل می‌کنند اما مقاومت این روش‌ها در مقابل تغییرات زاویه مناسب نیست، به این علت که ویژگی‌های ظاهری با تغییرات زاویه بسیار تغییر پذیر هستند. با تراز کردن تصویر چهره پیش از اعمال طبقه‌بند می‌توان این مشکل را بهبود داد. در طول تراز کردن تصویر، نقاط خاصی از تصویر (مانند نقطه وسط دو چشم و نقاط دو طرف دهان) در نظر گرفته می‌شود و به مختصات مشخصی منتقل می‌شوند. برای این منظور از ۵ نقطه ویژه استخراج شده در مرحله یافتن چهره توسط الگوریتم retina استفاده می‌کنیم. نتیجه اعمال این فرآیند را در شکل ۴.۱ مشاهده می‌کنید. در نهایت تصاویر چهره با اندازه $112 * 112$ * پیکسل ذخیره می‌گردند تا در مرحله

¹Face Alignment

آموزش شبکه، مورد استفاده قرار گیرند.



شکل ۴.۱: به منظور افزایش دقت شبکه، پس از یافتن چهره باید آن را تراز کرد [۲].

۳.۱ دسته بندی

روش پیشنهادی برای تشخیص بی‌درنگ چهره در محیط‌های بدون محدودیت، یک الگوریتم مبتنی بر شبکه‌های پیچشی می‌باشد.

این فرایند را می‌توان به صورت یک مسئله بهینه سازی فرموله کرد که در ادامه به بخش‌های مختلف آن می‌پردازیم.

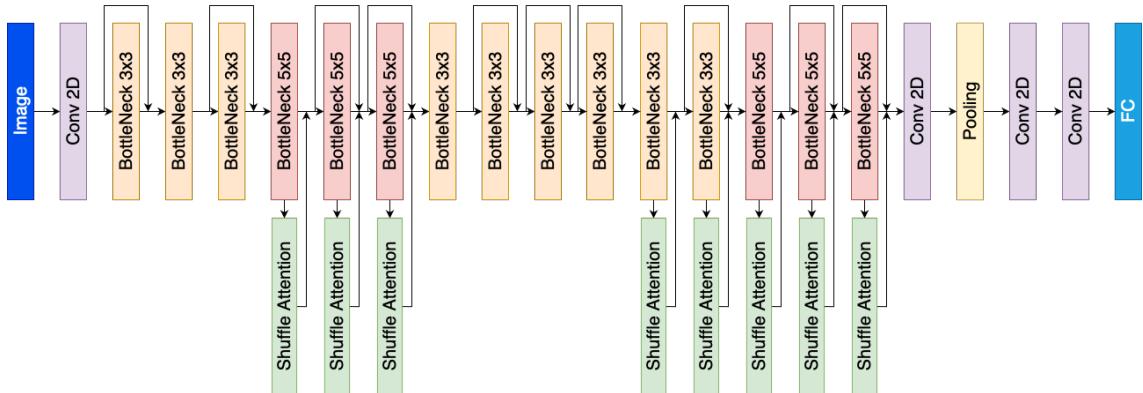
۱.۳.۱ مدل پیشنهادی پایه

در این بخش ابتدا باید معماری مناسب شبکه پایه برای مسئله را به دست آوریم. با بررسی شبکه‌های متقابل و مقایسه دقت و زمان پاسخگویی آن‌ها به کمک یادگیری انتقال، به این نتیجه می‌رسیم که شبکه MobileNetV3 دارای چگالی دقت بالاتری در مقایسه با شبکه‌های دیگر می‌باشد و نسبت دقت دسته بندی به تعداد پارامترهای شبکه در آن بیشتر می‌باشد. بنابرین می‌توان سرعت اجرای مناسب و همچنین دقت مناسب را از این شبکه انتظار داشت. از نتایج در می‌یابیم که بهترین معماری شبکه برای مسئله ما معماری MobileNetV3 است. نتایج آزمایش در جدول ؟؟ آمده است. آزمایش‌هایی بر روی معماری‌های مطرح دیگر نیز انجام شد که به علت ضعیف بودن نتایج یا بالا بودن زمان پاسخ دهی در جدول درج نشده‌اند. تشخیص چهره به دلیل وجود پیکسل‌های مشابه از نظر شدت روشنایی در تصویر بسیار چالش برانگیز بوده است. از آنجا که عملیات کانولوشن به پنجره محلی از شدت روشنایی پیکسل‌های تصویر هدایت می‌شود، بنابراین، این امکان وجود دارد که ویژگی‌های مربوط به پیکسل‌های تصاویر دارای برچسب یکسان، تفاوت‌هایی داشته باشند و یا ویژگی‌های مربوط به پیکسل‌های تصاویر دارای برچسب متفاوت، یکسان باشند. این اختلالات باعث

جدول ۱.۱: مقایسه و ارزیابی برخی از معماری شبکه های رایج در زمینه بینایی ماشین

نام شبکه	تعداد پارامترها	دقت بر روی مجموعه داده LFW
MobileNetV2	۱۳۲۳	۵۷۴۹
MobileNetV3	۱۳۲۳	۵۷۴۹
SqueezeNet	۲۵M.۱	۵۷۴۹
NASNetMobile	۱۳۲۳	۵۷۴۹
EfficientNetB0	۱۳۲۳	۵۷۴۹

کاهش جاپذیری بردارهای ویژگی خروجی می شوند. برای حل این مشکل، از اطلاعات کلی تصاویر به وسیله لایه های توجه استفاده می کنیم. لایه توجه استفاده شده در این پژوهش، شامل لایه توجه وابسته به کanal و لایه توجه وابسته به موقعیت می باشد. معماری شبکه پیشنهادی در شکل ۵.۱ آمده است.



شکل ۵.۱: مدل پایه مبتنی بر لایه های کانولوشن و لایه توجه.

در این معماری مسیر استخراج ویژگی از ۴ لایه کانولوشن، ۱ لایه رای گیری، ۱۵ لایه تنگنا^۱ و ۸ مازول توجه^۲ تشکیل شده است، که در ادامه توضیح داده خواهد شد، تصویر ورودی به بخش استخراج ویژگی داده می شود و مدل در این مسیر به طور خودکار یک سلسله مراتب ویژگی را از تصاویر ورودی آموزش خواهد دید و در نهایت این ویژگی های استخراج شده به عنوان ورودی دسته بند مورد استفاده قرار می گیرد.

¹Bottleneck

²Attention

۱.۱.۳.۱ لایه کانولوشن

۴ لایه کانولوشن دو بعدی همراه با گام یک یا دو استفاده شده است. در اولین لایه کانولوشن اندازه پنجره فیلترها 3×3 و در لایه های کانولوشن بعدی اندازه پنجره فیلترها 1×1 می باشد. دلیل انتخاب سایز کوچک پنجره فیلترها کاهش پیچیدگی محاسباتی و همچنین عملکرد خوب آن ها در استخراج ویژگی می باشد. چگونگی عملکرد یک لایه کانولوشن از رابطه ۲.۱ بدست می آید.

$$x_j^{(l)} = \sum_{i=0}^c w_{c_{ij}}^{(l)} * x_i^{(l-1)} + b_j^{(l)} \quad (2.1)$$

که در آن l نشان دهنده شماره لایه کانولوشن، b و w پارامترهای مدل، x خروجی هر لایه $[1, n]$ \in زیانگر شماره فیلتر موجود در لایه l و همچنین n بیانگر تعداد کل فیلترها در لایه l و نشان دهنده عملکرد کانولوشن می باشد.

۲.۱.۳.۲ لایه تنگنا و واحد باقیمانده

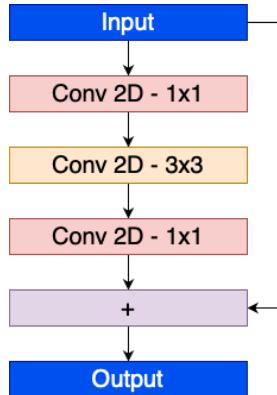
به عنوان مثال به جای پردازش یک نگاشت ویژگی عظیم با ۲۵۶ عمق، ابتدا همه این اطلاعات را در نگاشتهای ویژگی ۶۴ بعدی فشرده می کنیم. زمانی که این فشردگی انجام شد، از کانولوشن 3×3 استفاده می کنیم که وقتی روی ۶۴ نگاشت ویژگی به جای نگاشت اعمال شود، بسیار سریعتر است و چنین پردازشی می تواند همان نتایج یا نتایج بهتری نسبت به پشتلهای معمولی 3×3 ارائه کند. در نهایت با استفاده از 1×1 مجدداً به نگاشت اصلی ۲۵۶ خود بازمی گردیم. از سوی دیگر شبکه های عمیق با واحدهای باقیمانده^۱ بر روی پایگاه داده های مختلف مانند COCO، ImageNet دقت و همگرایی خوبی را از خود نشان داده اند. با استفاده از مسیرهای پرش^۲، واحدهای باقیمانده می توانند به سیگنال ها اجازه دهنده که مستقیماً از یک بلوک به بلوک های دیگر منتقل شوند. به طور کلی، واحدهای باقیمانده را می توان به صورت زیر بیان کرد:

$$x_{l+1} = x_l + R(x_l, w_l) \quad (3.1)$$

در اینجا R نشان دهندهتابع واحد باقیمانده است، x_l ویژگی ورودی به واحد باقیمانده l ام و W_l مجموعه ای از پارامترهای مربوط به واحد باقیمانده l ام می باشد. ایده اصلی شبکه های باقیمانده، عمیق تر کردن یک شبکه به منظور افزایش دقت شبکه مورد نظر می باشد. بنابراین با این عملیات در واقع به شبکه اجازه داده می شود که در صورت نیاز، ویژگی های لایه قبل بدون تغییر و به صورت مستقیم به لایه بعد منتقل شود. در شکل ۱.۶.۱ لایه تنگنا با واحد باقیمانده طراحی شده در این مدل به نمایش گذاشته شده است.

¹Deep Residual Network

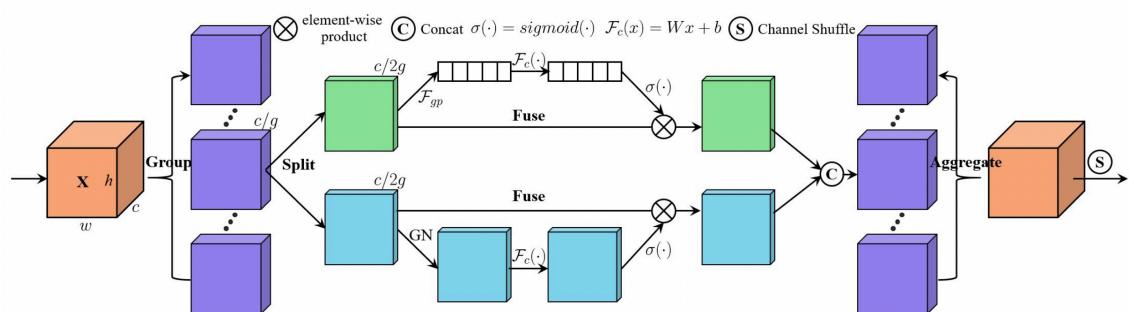
²Shortcut Pathway



شکل ۱: یک لایه تنگنا با واحد باقیمانده که با استفاده از کانولوشن های 1×1 اقدام به کاهش ابعاد نگاشت ویژگی می کند..

۳.۱.۳.۱ واحد توجه

لایه توجه استفاده شده در این پژوهش یک واحد توجه SA^۱ است. این لایه ابتدا ابعاد کanal را به چندین ویژگی فرعی قبل از پردازش موازی آنها تقسیم می کند. سپس، برای هر زیر ویژگی از یک واحد Shuffle برای به تصویر کشیدن وابستگی های ویژگی در هر دو بعد مکانی و کanal استفاده می کند. پس از آن، همه زیر ویژگی ها جمع می شوند و یک عملگر تغییر کanal برای امکان برقراری ارتباط اطلاعاتی بین ویژگی های فرعی مختلف به کار گرفته می شود. معماری این مازول در شکل ۷.۱ آمده است. این لایه شامل دو بخش اصلی وابسته به کanal^۲ و وابسته به موقعیت^۳ می باشد.



شکل ۷.۱: معماری مازول .[۹] Attention Shuffle

برای یک نقشه ویژگی^۴ به نام X با ابعاد $C \times H \times W$ که در آن C و H و W به ترتیب عمق کanal، ارتفاع و عرض هستند، توجه وابسته به کanal هر کanal از ویژگی های استخراج شده سطح بالا می تواند به عنوان یک پاسخ کلاس خاص در نظر گرفته شود و پاسخ های معنایی مختلف با یکدیگر مرتبط هستند. با بهره گیری از وابستگی های متقابل بین کanal ها، می توان بر کanal های ویژگی

¹Shuffle Attention

²Channel Attention Module

³Spacial Attention Module

⁴feature map

وابسته، تأکید کرده و استخراج ویژگی‌های معانی خاص را بهبود بخشد. بنابراین در این پژوهش یک مازول توجه وابسته به کanal ایجاد شده تا به طور واضح و استتگی بین کanal‌ها را مدل کند. ساختار کلی مازول توجه وابسته به کanal در شکل؟؟ نمایش داده شده است، که در آن نقشه توجه وابسته به کanal مستقیماً توسط یک لایه پولینگ^۱ میانگین و دو لایه تماماً متصل بدست خواهد آمد. اگر فرض شود $FM = [fm_1, fm_2, \dots, fm_c]$ نقشه ویژگی‌های ورودی به ابعاد H و W باشد که در آن c نشان دهنده تعداد کanal‌ها است، رابطه ۴.۱ نشان دهنده چگونگی ایجاد بردار نقشه توجه وابسته به کanal می‌باشد.

$$R_k = \frac{1}{H \times W} \sum_i^H \sum_j^W fm_k(i, j) \quad (4.1)$$

با انجام این عملیات، اطلاعات مکانی کلی برای هر کanal به صورت جداگانه محاسبه می‌شود و در بردار R قرار می‌گیرد. سپس این بردار بدست آمده توسط دو لایه تماماً متصل مطابق رابطه زیر ۵.۱ نگاشت می‌شود.

$$\hat{R} = \mathbf{W}_1(\delta(\mathbf{W}_2 R)) \quad (5.1)$$

که در آن $W_1 \in R^{c \times \frac{c}{2}}$ و همچنین $W_2 \in R^{\frac{c}{2} \times c}$ می‌باشد. W_1 و W_2 وزن‌های لایه‌های تماماً متصل و همچنین δ نشان دهنده واحد خطی اصلاح شده می‌باشد. در نهایت نقشه توجه وابسته به کanal بدست آمده در نقشه ویژگی‌های ورودی ضرب شده و خروجی حاصل را ارائه می‌دهد.

توجه وابسته به موقعیت:

وجود ویژگی‌های متمایز برای درک تصویر ورودی ضرری می‌باشد، که این ویژگی‌ها می‌توانند در بازه بزرگی از تنوع قرار بگیرند. با این حال روش‌های زیادی نشان داده اند که ویژگی‌های محلی استخراج شده توسط شبکه‌های تماماً متصل مرسوم، می‌توانند منجر به تقسیم‌بندی نادرست شود [؟]. بنابراین به منظور مدل سازی روابط متنی بر روی ویژگی‌های محلی، مازول توجه وابسته به موقعیت استفاده می‌شود. مطابق شکل؟؟ مازول توجه وابسته به موقعیت، ماتریس ویژگی محلی $A \in \mathbb{R}^{C \times H \times W}$ را به عنوان ورودی دریافت کرده سپس با استفاده از یک لایه کانولوشن دو نقشه ویژگی جدید $C \in \mathbb{R}^{C \times H \times W}$ و $B \in \mathbb{R}^{C \times H \times W}$ را تولید می‌کند. در مرحله بعد دو نقشه ویژگی بدست آمده به سایز $\mathbb{R}^{C \times N}$ تغییر شکل داده می‌شود، که در آن $N = H \times W$ بیانگر تعداد کل پیکسل‌ها می‌باشد. پس از آن یک ضرب ماتریسی بین ترانهاده C و B انجام می‌شود، و یک لایه Softmax برای محاسبه نقشه توجه وابسته به موقعیت اعمال شده و نتیجه نهایی S را تولید می‌کند. در همین حال، با استفاده از یک لایه کانولوشن دیگر و همچین نقشه ویژگی A ، یک نقشه ویژگی $D \in \mathbb{R}^{C \times H \times W}$ ساخته خواهد شد. سپس این نقشه ویژگی D به ابعاد $\mathbb{R}^{C \times N}$ تغییر

¹Pooling

سایز داده خواهد شد. مرحله بعد یک ضرب ماتریسی بین نقشه ویژگی S و ترانهاده نقشه ویژگی D صورت می‌گیرد و نتیجه حاصل به سایز $O \in \mathbb{R}^{C \times H \times W}$ تغییر سایز داده می‌شود. درنهایت نتیجه نهایی ۶.۱ مطابق رابطه محاسبه خواهد شد.

$$O_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j \quad (6.1)$$

از رابطه بالا می‌توان استنباط کرد که نقشه ویژگی E بدست آمده برای هر موقعیت، یک جمع وزنی از ویژگی‌ها، در تمام موقعیت‌ها و ویژگی‌های ورودی A است.

۲.۳.۱ آموزش شبکه جهت استخراج ویژگی

از آن جایی که شبکه MobileNet معماری بسیار سبک تری نسبت به معماری‌های شناخته شده دیگر که در فصل ۲ معرفی شدند، دارد؛ بنابراین استخراج ویژگی‌ها از چهره و دسته بندی تصاویر چهره با دقت بالا برای این شبکه بسیار سخت و دشوار است. همچنین در مواردی شباهت چهره افراد به یکدیگر کار را از آن چه هست سخت‌تر خواهد کرد. بنابراین ما روشی برای استخراج ویژگی از تصاویر چهره پیشنهاد کرده‌ایم که کمک می‌کند ویژگی‌های استخراج شده که متعلق به دو دسته متفاوت هستند، فاصله بیشتری از هم داشته باشند و در مقابل ویژگی‌های استخراج شده برای دو تصویر از چهره یک فرد یکسان، فاصله کمتری از هم داشته باشد؛ تا این طریق بتوان به کاهش مشکلات ذکر شده کمک کرد. این روش شامل بخش‌های تولید چهره، یافتن چهره، آموزش شبکه عصبی پیچشی و استخراج ویژگی می‌باشد.

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

۳.۳.۱ تابع ضرر

یکی از چالش‌های اصلی در یادگیری ویژگی‌ها با استفاده از شبکه‌های عصبی عمیق پیوسته (DCNN) برای شناسایی چهره در مقیاس بزرگ، طراحی تابع ضرر مناسب است که قدرت تفکیک را افزایش می‌دهد. هدف ما کم کردن فاصله بین ویژگی‌های عمیق و مراکز کلاس‌های آن‌ها در فضای اقلیدسی برای دستیابی به فشردگی درون کلاسی بیشتر می‌باشد. ما یک تابع ضرر برای افزایش زاویه ای حاشیه برای دستیابی به ویژگی‌های بسیار متمایز برای تشخیص چهره پیشنهاد می‌کنیم که عملکرد بهتر برخوردار است و می‌توان آن را به راحتی با هزینه‌های محاسباتی ناچیز پیاده سازی کرد. برای آموزش شبکه‌های عصبی عمیق پیوسته برای تشخیص چهره،

دو رویکرد اصلی وجود دارد. روش اول دسته بندی را آموزش می دهند که می تواند هویت های مختلف را در مجموعه آموزش از هم جدا کند، مانند با استفاده از طبقه بندی softmax، و رویکرد دوم که مستقیماً یک تعییه را یاد می گیرند، مانند triplet loss. بر اساس داده های آموزش در مقیاس بزرگ و معماری DCNN، هر دو روش می توانند عملکرد بسیار خوبی در تشخیص چهره داشته باشند. با این حال، هم رویکرد softmax و هم رویکرد triplet loss اشکالاتی دارد. برای :

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

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

:triplet loss برای

۱. برای مجموعه داده های مقیاس بزرگ، رشد شدید در تعداد ترکیب های تعداد تصاویر سه گانه وجود دارد که منجر به افزایش قابل توجه تعداد مراحل تکرار می شود.

۲. استخراج مجموعه تصاویر سه گانه یک مسئله دشوار برای آموزش موثر می باشد.

ما برای افزایش بیشتر قدرت تمایز مدل تشخیص چهره و ایجاد ثبات در روند آموزش، تابع ضرر مبتنی بر توابع مثلثاتی را پیشنهاد می کنیم. همانطور که در شکل ۲ نشان داده شده است، حاصل ضرب نقطه ای مقادیر موجود در ویژگی های استخراج شده و آخرين لایه کاملاً متصل، برابر با ضرب کسینوسی آن ها پس از نرمال سازی می باشد، ما از تابع مثلثاتی کسینوسی برای محاسبه زاویه بین ویژگی فعلی و وزن هدف استفاده می کنیم. سپس یک حاشیه زاویه ای به زاویه هدف اضافه می کنیم، در انتها با استفاده از تابع کسینوس دوباره مقادیر را به فضای خطی برمی گردانیم. مراحل بعدی دقیقاً مانند softmax هستند. مزایای این روش پیشنهادی را می توان به شرح زیر خلاصه کرد:

- در مجموعه داده های تصویر و فیلم در مقیاس بزرگ ، به عملکرد مناسبی دست می یابد.
- فقط به چندین خط کد نیاز دارد و اجرای آن در چارچوب های یادگیری عمیق مبتنی بر Pytorch و Tensorflow آسان است. برای داشتن عملکرد پایدار نیازی به ترکیب با سایر توابع ضرر ندارد و به راحتی همگرا می شود.
- هنگام آموزش فقط پیچیدگی محاسباتی ناچیز را اضافه می کند. پردازنده های گرافیکی کنونی می توانند به راحتی از هزاران دسته مختلف برای آموزش پشتیبانی کنند و مدل به راحتی می تواند هویت های بیشتری را پشتیبانی کند.

رابطه ریاضی softmax معروف ترینتابع ضرر طبقه بندی که به طور گستردۀ استفاده می شود، به شرح زیر است:

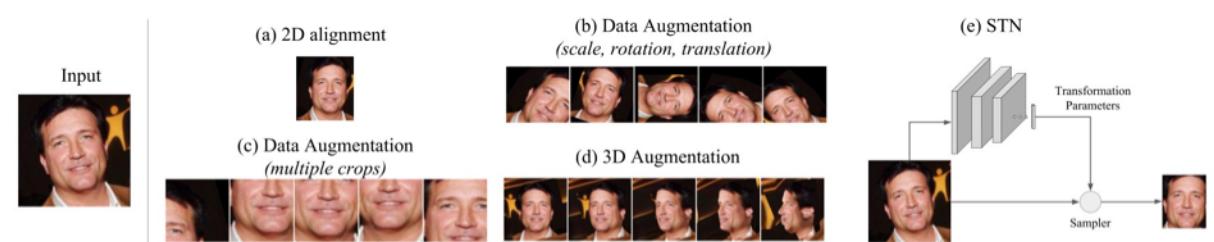
$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (7.1)$$

که در آن x_i نشان دهنده ویژگی عمیق نمونه i از دسته y است. تعداد ابعاد ویژگی استخراج شده را W_j ستون j ام از وزن W می باشد و b_j بایاس است. مقدار N اندازه دسته و n تعداد دسته ها است. این تابع مستقیماً ویژگی استخراج شده را برای اعمال شباهت بالاتر برای نمونه های درون کلاس و فاصله بیشتر برای نمونه های بین کلاسی بینه نمی کند، که منجر به ایجاد مشکل در عملکرد آن برای تشخیص چهره عمیق تحت تغییرات ظاهری بزرگ درون کلاس می شود (به عنوان مثال تغییرات زاویه چهره و تغییرات سنی).

ما رابطه فوق را مبنای محاسبات قرار دادیم و تغییرات جزیی به آن اضافه کردیم. برای سادگی مقدار بایاس را صفر در نظر گرفتیم. سپس حاصل ضرب مقادیر موجود در ویژگی های استخراج شده و آخرین لایه کامل متصل را به صورت (j) $W_j^T x_i = ||W_j|| |x_i| \cos(\theta_j)$ تبدیل می کنیم، که θ_j زاویه بین وزن W_j و ویژگی x_i است. نرمال سازی ویژگی ها و وزن ها باعث می شود که خروجی فقط به زاویه بین ویژگی و وزن بستگی داشته باشد. به کمک نرمال سازی مقادیر وزن W_j را برابر ۱ در نظر می گیریم. همچنین ویژگی استخراج شده $|x_i|$ را نرمال کرده و نام آن را s در نظر می گیریم. بنابرین ویژگی های استخراج شده در یک ابر کره با شعاع s توزیع می شوند. برای افزایش حاشیه بین x_i و W_j یک مقدار l اضافه می کنیم تا به طور همزمان فشرده سازی درون کلاسی و اختلاف بین کلاسی را افزایش دهیم.

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1}^n e^{s(\cos(\theta_j))}} \quad (8.1)$$

همانطور که در شکل ۳ نشان داده شده است، softmax ویژگی های تقریباً قابل تفکیکی ایجاد می کند اما در مرازهای تصمیم گیری ابهام قابل توجهی به وجود می آید، در حالی که تابع ضرر ما می تواند فاصله بیشتری را بین دسته های نزدیک اعمال کند.



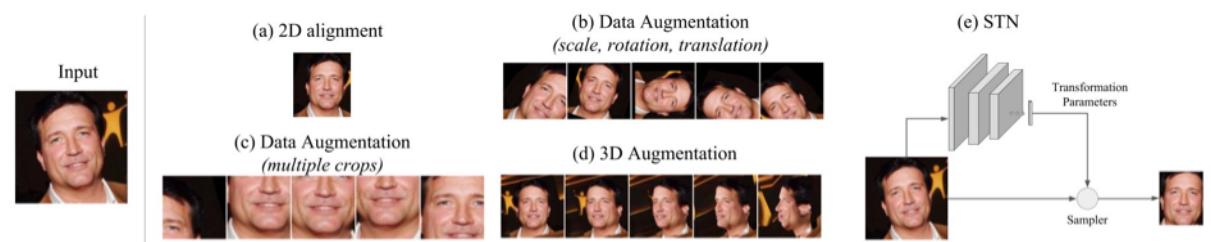
شکل ۸.۱: رویکردهای مختلف هم ترازی چهره [?].

۱.۳.۳.۱ آموزش

در ابتدای روند آموزش، لازم است پارامترهای مدل مقدار دهی اولیه شوند و انتخاب پارامترهای اولیه میتواند تأثیر زیادی در مدل آموزش یافته داشته باشد. در این پژوهش به منظور مقدار دهی اولیه پارامترها از تابع توزیع یکنواخت استفاده شده است. از دیگر چالشهای اساسی برای روشهای بهینه سازی مبتنی بر گرادیان، انتخاب میزان نرخ یادگیری مناسب است. روشهای کلاسیک گرادیان تصادفی از نرخ یادگیری ثابت یا کاهشی استفاده میکنند، که برای همه پارامترهای مدل یکسان است. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف میتوانند از نظر مقدار متفاوت باشند، که میتواند به نرخ یادگیری مختلفی نیاز داشته باشد. با این حال، مشتقات جزئی پارامترهای لایه‌های مختلف میتوانند از نظر مقدار تفاوت قابل توجهی داشته باشند، که میتواند به نرخ یادگیری مختلفی نیاز داشته باشد. در سالهای اخیر، تمایل به توسعه روشهایی برای انتخاب خودکار نرخ یادگیری مستقل افزایش یافته است. اکثر روشهای عنوان مثل، AdaGrad، AdaDelta، RMSprop و [۴۷] [۴۸] [۴۹] [۵۰] آمارهای مختلف مشتقات جزئی را در چندین تکرار جمع آوری میکنند و از این اطلاعات برای تعیین میزان یادگیری سازگار برای هر پارامتر استفاده میکنند. این امر به ویژه برای آموزش شبکه‌های عمیق بسیار مهم است، جایی که نرخ یادگیری مطلوب اغلب برای هر لایه بسیار متفاوت است. در این پژوهش در آزمایشات انجام شده از همه روشهای نام برده استفاده شد ولی روش Adam عملکرد بهتری ارائه داده است.

۲.۳.۳.۱ دسته‌بندی

در مرحله آزمون به منظور تشخیص هویت یک تصویر چهره، پس از پیش‌پردازش تصویر را مطابق با ورودی شبکه تغییر اندازه میدهیم و جهت استخراج ویژگی به آن شبکه می‌دهیم. پس از استخراج ویژگی‌ها توسط شبکه، بردار ۵۱۲ تایی بدست آمده را با بردار‌های مربوط به چهره‌های بانک اطلاعاتی مقایسه کرده و با محاسبه فاصله اقلیدسی بردار‌ها، نزدیک‌ترین شخص مورد نظر انتخاب شده و در صورتی که فاصله میان بردار ویژگی آن‌ها از حد آستانه کمتر باشد. عمل دسته‌بندی انجام شده و هویت چهره مورد نظر تعیین می‌شود. در غیر این صورت اعلام میداریم که شخص مورد نظر قابل شناسایی نمی‌باشد.



شکل ۹.۱: رویکردهای مختلف هم ترازی چهره [۴].

۴.۱ فناوری های استفاده شده

پیاده سازی این الگوریتم به کمک زبان برنامه نویسی پایتون و کتابخانه OpenCV و PyTorch انجام شده است. از کتابخانه های مهم مورد استفاده دیگر در این کار می توان به NumPy برای انجام محاسبات ماتریسی و SciPy و Scikit Learn اشاره کرد. [۳۵] برای آموزش شبکه عصبی مربوط به دسته بندی ۱۲ گیگابایت حافظه اصلی و ۱۲ گیگابایت حافظه گرافیکی در اختیار گرفتیم.

۲ فصل

ارزیابی روش پیشنهادی

۱.۲ مقدمه

در این کار سعی بر این داشته‌ایم تا به کمک روش‌های یادگیری ژرف در راستای شناسایی چهره در ویدیوهای بدون محدودیت قدمی برداریم و در این حوزه عملکرد هوش مصنوعی و یادگیری ژرف را بهبود دهیم. بنابراین با تحقیق و آزمایش روشی برای دسته بندی دقیق‌تر چهره افراد در تصاویر ویدیویی پیشنهاد داده‌ایم که شرح آن در فصل قبل انجام شد و نوبت آن است که الگوریتم پیشنهادی را ارزیابی کرده و با بیان نتایج به مقایسه با کارهای دیگر می‌پردازیم.

۲.۲ معیار ارزیابی

یکی از معیارهایی که بسیار در زمینه‌های تحقیقاتی و مسائل دسته بندی حائز اهمیت است، معیار دقت می‌باشد. در این کار ما با تصاویر چهره دارای هویت‌های مختلف سرو کار داریم. بنابراین معیار دقت در این کار به معنی درصد نمونه‌هایی است که هویت آن‌ها به درستی تشخیص داده شده است. فرمول این معیار در رابطه ۱.۲ بیان شده است.

$$Accuracy = \frac{TP}{N} \quad (1.2)$$

که در آن مثبت‌های صحیح (TP) تعداد نمونه‌هایی که به درستی تشخیص داده شده‌اند. و N تعداد کل نمونه‌ها را نشان می‌دهد. همانطور که می‌دانیم، به طور معمول معماری‌های شبکه عصبی عمیق‌تر، دارای دقت بالاتر و همچنین زمان پردازش بیشتری می‌باشند. از طرفی هرچه معماری شبکه عصبی، سبک تر و تعداد پارامترهای آن کمتر باشد، می‌توان انتظار داشت که سرعت بالاتری در زمان اجرا داشته باشد، اما دقت آن با کاهش همراه است. بر همین اساس در مقابل معیار دقت، معیار دیگری به نام تراکم دقت یا چگالی دقت وجود دارد که از تقسیم مقدار دقت بر تعداد پارامترهای معماری شبکه بدست می‌آید و می‌تواند معیار خوبی برای ارزیابی دقت معماری‌های مختلف نسبت به سرعت اجرای آن‌ها باشد.

$$AccuracyDensity = \frac{Accuracy}{Mparams} \quad (2.2)$$

که در آن Mparams تعداد پارامترهای شبکه به میلیون می‌باشد.

۳.۲ مجموعه داده

در زمینه تشخیص چهره، مجموعه داده های بسیار زیادی وجود دارند. تعداد از این مجموعه داده ها، حاوی تصاویر چهره در محیط های آزمایشگاهی و کنترل شده می باشند که در بحث ما گنجانده نمی شوند. با رشد الگوریتم های یادگیری عمیق و بدست آمدن نتایج مناسب در زمینه تشخیص چهره، مجموعه داده های جدیدی منتشر شدند که دارای تصاویر چهره در محیط های بدون محدودیت هستند. برخی از این مجموعه داده ها برای آموزش و برخی را برای آزمون استفاده کرده ایم که به شرح آن ها می پردازیم.

۱.۳.۲ مجموعه داده های آموزش

۱.۱.۳.۲ مجموعه داده CASIA Web-Face

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

۲.۱.۳.۲ مجموعه داده MS-Celeb-1M

مجموعه داده MS-Celeb-1M^۱ که به مراتب مجموعه داده بزرگتری محسوب می شود، شامل بیش از ۸ میلیون تصویر چهره متفاوت از ۱۰۰ هزار فرد است. برخلاف مجموعه داده CASIA Web-Face، این مجموعه داده با نویز و برچسب اشتباه همراه است. این مجموعه داده را شرکت مایکروسافت ایجاد کرده است. [۵۳]

۳.۱.۳.۲ مجموعه داده VGGFace2

محققان دانشگاه آکسفورد نسخه ۲ مجموعه داده VGGFace را با ۳۱۰.۳ میلیون تصویر از ۹۱۳۱ فرد مختلف ارائه کردند. این تصاویر با کمک جستجوی تصویر گوگل جمع آوری شده و شامل تغییرات مختلف برای هر فرد نظیر سن، جهت، نور و ... هستند. این مجموعه داده شامل افراد مختلفی نظیر سیاستمداران، وزشکاران، بازیگران و ... است و به طور تقریبی از هر فرد ۳۶۲ تصویر مختلف موجود است. [۵۴]

¹Microsoft celebrities

۲.۳.۲ مجموعه داده‌های آزمون

۱.۲.۳.۲ LFW مجموعه داده

مجموعه داده LFW^۱ که شامل ۱۳۲۳۳ تصویر چهره متفاوت از ۵۷۴۹ فرد مختلف است، از اولین مجموعه داده‌های منتشر شده برای مسائل تشخیص چهره بدون محدودیت می‌باشد. [۵۵]

۲.۲.۳.۲ PubFig مجموعه داده

مجموعه داده PubFig^۲ که شامل ۵۸۷۹۷ تصویر چهره متفاوت از ۲۰۰ فرد است، از اینترنت جمع آوری شده است. این تصاویر شامل تغییرات مختلف مانند جهت، نور، انسداد و ... هستند. [۵۶]

۳.۲.۳.۲ YouTube Faces مجموعه داده

این مجموعه داده برای استفاده در کارهای تشخیص چهره در تصاویر ویدیویی ایجاد شده است. این مجموعه داده شامل ۳۴۲۵ ویدیو از ۱۵۹۵ فرد مختلف است و تمام ویدیوها از سایت یوتیوب دانلود شده اند. به طور میانگین ۱۵.۲ فیلم برای هر شخص در دسترس است. کوتاه ترین مدت ویدیو ۴۸ فریم، طولانی ترین ویدیو ۶۰۷۰ فریم و متوسط طول یک ویدیو ۳۰۱۸۱ فریم است. [۵۴]

۴.۲.۳.۲ CFP مجموعه داده

مجموعه داده CFP^۳ که شامل ۷۰۰۰ تصویر چهره متفاوت از ۵۰۰ فرد است، تصاویر افراد مشهور را در حالت‌های تمام رخ و نیم رخ جمع آوری کرده است. این مجموعه داده می‌تواند ابزار ارزیابی بسیار خوبی برای چالش زاویه چهره باشد. [۵۵]

¹Labeled Faces in the Wild

²Public Figures

³Celebrities in Frontal-Profile in the Wild

٥.٢.٣.٢ مجموعه داده CACD

مجموعه داده CACD^۱ شامل ۱۶۳۴۴۶ تصویر از ۲۰۰۰ فرد مشهور است که از اینترنت جمع آوری شده است. تصاویر از موتورهای جستجو با استفاده از نام افراد مشهور و سال (۲۰۰۴-۲۰۱۳) به عنوان کلمات کلیدی جمع آوری شده است. بنابراین، می‌توان با تفربیق سال تولد افراد از سال عکس گرفته شده، به سادگی سن افراد در تصاویر را تخمین زد. این مجموعه داده می‌تواند ابزار ارزیابی بسیار خوبی برای چالش تغییرات سن باشد. [۵۷]

٦.٢.٣.٢ مجموعه داده MegaFace

محققان دانشگاه واشنگتن مجموعه داده MegaFace^۲ عظیم را ارائه کرده اند که هم در مسائل تایید چهره و هم در مسائل تشخیص چهره کاربرد دارد. این مجموعه داده شامل یک زیر مجموعه آزمایش است که خود از دو مجموعه FaceScrub شامل تصاویر افراد مشهور و FGNet شامل تصاویر مخصوص چالش سن، تشکیل شده است. اگر الگوریتم مورد نظر بر روی مجموعه دقت بالا بدست آورد، نشان دهنده قوت الگوریتم در تصاویر با تفاوت سن های بالا می باشد [۵۸]. اطلاعات تکمیلی درباره مجموعه داده های آزمایش را در جدول ۱۰ مشاهده می نمایید.

جدول ۱۰: مجموعه داده های ارزیابی رایج در زمینه بازنگشتن چهره

نام	تعداد تصاویر چهره	تعداد افراد	نوع کاربرد
LFW	۱۳۲۳۳	۵۷۴۹	تشخیص چهره
PubFig	۵۸۷۹۷	۲۰۰	تشخیص چهره
YouTube Faces	۳۴۲۵	۱۵۹۵	تشخیص چهره
CFP	۷۰۰۰	۵۰۰	تشخیص چهره
CACD	۱۶۳۴۴۶	۲۰۰۰	تشخیص چهره
MegaFace	۹۷۵ + ۱۴۱۰۰۰ + ۱۰۰۰۰۰	۸۲ + ۶۹۵ + ۶۹۰۰۰	تشخیص چهره و تایید چهره

٤.٢ پیکربندی الگوریتم

به منظور آموزش شبکه، در هر دوره ۲۰٪ تعداد داده های آموزشی را به عنوان داده های ارزیابی در نظر می گیریم تا روند آموزش شبکه را بر اساس عامل های دیگر مورد بررسی قرار دهیم. همچنین در هر تکرار، تعداد دسته هایی که به شبکه برای آموزش داده می شود را

^۱Cross-Age Celebrity Dataset

^۲Million-Scale Face Recognition Dataset

برابر ۱۶ قرار دادیم. توصیه می شود این مقدار توانی از ۲ باشد که مقدار ۱۶ با تجربه و توجه به ظرفیت حافظه پردازنده گرافیکی بدست آمده است. از بهینهساز Adam جهت آموزش استفاده کرده ایم. این بهینهساز نیز نرخ آموزش را بر اساس خطای صورت تطبیقی کم می کند. همچنین میانگین کاهش گرادیان های تکرارهای قبل را نگهداری می کند تا بر اساس آنها جهت گرادیان تکرار جدید را محاسبه کند. برای تابع ضرر از معادله ۸.۱ استفاده کرده ایم. این تابع میزان ضرر را به صورتی که در فصل قبل صحبت شد، محاسبه می کند.

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

۵.۲ نتایج آزمون

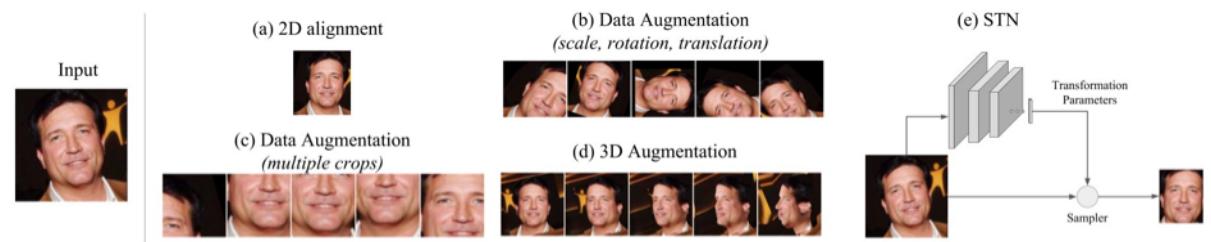
با توجه به پیکربندی بیان شده و همچنین عامل هایی که در فصل سوم توضیح داده شد، مدل را آموزش داده ایم و به سراغ مجموعه داده های آزمون می رویم. مقدار دقت را طبق ۱.۲ برای معماری های مرتبط و مشهور و همچنین روش پیشنهادی محاسبه کردیم. در جدول ۲ نتایج حاصل دقت را در معماری های مختلف بر روی مجموعه داده داده های مشاهده می کنید.

جدول ۲: دقت معماری های مختلف بر روی مجموعه داده های ارزیابی رایج در زمینه بازشناسی چهره

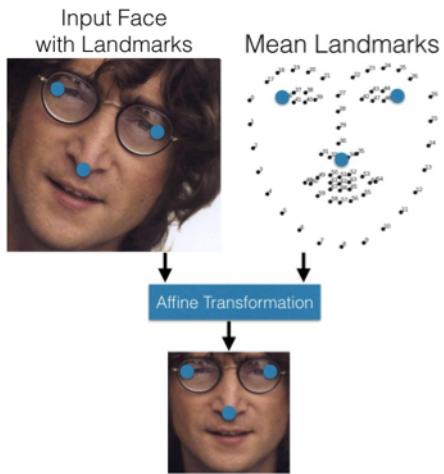
LFW	PubFig	YouTube Faces	CFP	CACD	MegaFace	معماری مدل	مجموعه داده آموزش
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	Web-Face CASIA
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	
۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	۵۵.۹۹	NBM	MS-Celeb-1M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

نمودار FROC که مقدار دقت و میزان خطای صورت تطبیقی را نشان می دهد، برای مجموعه داده های Web-CASIA و MS-Celeb-1M Face در شکل ۱.۲ و شکل ۲.۲ مشاهده می کنید.

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



. [?] E-Ophtha-MA مقالات در مقایسه با روش پیشنهادی برای مجموعه داده FROC منحنی



. [?] DIARETDB1 مقالات در مقایسه با روش پیشنهادی برای مجموعه داده FROC منحنی

این نشان می‌دهد که ترکیب ویژگی‌های ژرف (استخراج شده توسط انسان) و ویژگی‌های معنادار (استخراج شده توسط انسان) باعث می‌شود که سامانه مزیت هر دو دسته ویژگی را با هم داشته باشد و در به دست آوردن نتایج بهتر کمک کننده باشد.

٣ فصل

نتیجه‌گیری و پیشنهادات

در این بخش جمع‌بندی و نتیجه‌گیری این پژوهش بیان می‌شود. هدف کار ارائه شده در این پژوهش ارائه و بهبود روش‌های خودکار مبتنی بر هوش مصنوعی و یادگیری ژرف به منظور تشخیص چهره به صورت بی‌درنگ در شرایط بدون محدودیت در تصاویر ویدیویی است.

۲.۳ بحث و نتیجه‌گیری

همانطور که در فصل‌های قبل نیز بیان شد، روش‌های مختلفی برای یافتن چهره و شناسایی چهره ارائه شده‌اند. برای حل مساله دسته‌بندی چهره دو روش کلی، مبتنی بر تصویر و روش‌های مبتنی بر استخراج ویژگی وجود دارد. روش‌های مبتنی بر تصویر خود دارای رویکردهای مختلفی از جمله روش‌های مبتنی بر رنگ بندی، روش‌های مبتنی بر شکل و روش‌های مبتنی بر گرادیان می‌باشد. دسته‌ای از این روش‌ها سعی کرده‌اند تا به کمک روش‌های غیر ژرف و استخراج ویژگی‌های معنادار از پیکسل‌های چهره به این هدف دست یابند و در نهایت یک دسته‌بند را با ویژگی‌های ارائه شده آموزش دهند و از این دسته‌بند برای آزمون استفاده کنند.

همچنانی روش‌های مبتنی بر استخراج ویژگی که در سال‌های اخیر بسیار مورد توجه قرار گرفته اند شامل رویکردهای مبتنی بر شبکه عصبی، بردار پشتیبان و... می‌باشند. این روش‌ها با تکیه بر ویژگی‌های ژرف استخراج شده از شبکه‌های عصبی پیچشی سعی دارند چهره‌ها را در تصاویر و ویدیو شناسایی کنند. به این صورت که تصاویر آموزشی را به شبکه می‌دهند و شبکه در مرحله آموزش سعی می‌کند ویژگی‌های لازم و تفکیک‌پذیر بین چهره افراد مختلف را شناسایی کند. واضح است که ویژگی‌های استخراج شده توسط شبکه با ویژگی‌های معنادار استخراج شده توسط انسان همیشه برای نیستند و حتی ممکن است ویژگی‌های شبکه اصلاً قابل تفسیر نباشند. در سال‌های گذشته، انواع شبکه‌های عمیق در زمینه دسته‌بندی تصاویر چهره به شدت مورد استفاده قرار گرفته است. مدل‌های مختلف ارائه شده با گرفتن تصاویر مختلف چهره به عنوان ورودی، هویت شخص را به عنوان خروجی تولید می‌کنند. اما روش‌های ارائه شده، دارای محدودیت‌هایی نظیر پایین بودن دقت، تعداد زیاد پارامترهای شبکه‌های ارائه شده، سرعت کم پردازش و همچنانی تعداد زیاد تصاویر ورودی برای آموزش می‌باشند. ما با توجه به شرایط بی‌درنگ و محدودیت‌های دیگر مسئله، در میان روش‌های استخراج ویژگی توسط یادگیری ژرف به دنبال معماری سبک‌تر و سریع‌تر بودیم که موفق شدیم با تحلیل و آزمایش به معماری مورد نظر دست پیدا کنیم که علاوه بر پردازش‌های سبک و سرعت بالا، دارای دقت لازم و کافی در دسته‌بندی نیز باشد.

به نظر می‌رسد افزودن لایه‌های خاص منظوره به معماری شبکه‌های عصبی معروف، باعث افزایش دقت الگوریتم در کاربردهای خاص می‌شود. به منظور بهبود در نتایج باید در تغییر لایه‌های معماری شبکه دقت بیشتری داشت. بنابراین برای بهبود ویژگی‌های ژرف، استفاده از ویژگی‌های چند مدل شبکه در کنار هم می‌تواند کمک کننده باشد. به این صورت که از هر مدل ، لایه‌های سودمندتر را انتخاب کرده و از کنار هم قرار دادن آن‌ها، بردار ویژگی ما غنی‌تر خواهد شد.

همچنین می‌توان تابع ضرر را بروز رسانی و بهینه تر کرد و به این ترتیب بردار ویژگی‌هایی به دست خواهد آمد که از فاصله درون دسته‌ای کمتر و فاصله برون دسته‌ای بیشتری برخوردار باشند. از طرفی در این میان ممکن است ویژگی‌هایی یکسان از شبکه‌ها استخراج شود، و یا این که ویژگی‌هایی وجود داشته باشند که در تشخیص هویت چهره مورد نظر زیاد نقش موثری نداشته باشند. به این ترتیب استفاده از روش‌های کاهش بردار ویژگی برای استخراج ویژگی‌های اصلی برای ایجاد مرز بهتر بین دسته‌ها می‌توان مفید باشد. در مقابل با تحقیق می‌توان ویژگی‌های متمایز کننده معنادار دیگری برای چهره به دست آورد تا به بردار ویژگی اضافه شوند. روش‌های مختلفی برای پیش‌پردازش تصاویر چهره وجود دارند. استفاده از روش‌های جدید پیش‌پردازش به منظور کاهش بیشتر نویز، یکسان سازی روش‌نایی تصاویر بافت‌نگار و ... می‌تواند به استخراج ویژگی‌های مفیدتر کمک کند تا بتوان نتایج بهتر برای این منظور به دست آورد.

در بحث استخراج ویژگی می‌توان با استفاده از لایه‌های بیشتر و همچنین استفاده از مدل‌های دیگری از لایه توجه، ویژگی‌های بهتری استخراج کرد و در نهایت ویژگی‌های استخراج شده منجر به دسته بندی بهتر تصاویر خواهد شد. همچنین به دلیل وجود دسته های زیاد در پایگاه داده، می‌توان از تابع ضری که مبتنی بر تعداد زیاد دسته‌ها باشد استفاده کرد که منجر به آموزش بهتر مدل و همچنین کم شدن اشتباہات دسته بندی می‌شود.

یکی دیگر از مشکلات موجود در حوزه تشخیص چهره، نبود داده کافی به منظور آموزش مدل می‌باشد. به منظور حل این مشکل می‌توان مدلی به منظور تولید تصاویر مصنوعی ارائه داد تا با مشکل نبود داده کافی مقابله کرد و در نهایت مدلی کارآمدتر و بهتر برای تشخیص ارائه داد. و همچنان راه حل‌ها و ایده‌های دیگری می‌توانند در جهت افزایش دقت این سامانه کمک کننده باشند

منابع و مأخذ

- [1] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “Cosface: Large margin cosine loss for deep face recognition,” 2018.
- [2] S. W. Cho, N. R. Baek, M. C. Kim, J. H. Koo, J. H. Kim, and K. R. Park, “Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network,” *Sensors*, vol.18, no.9, 2018.
- [3] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” 2019.
- [4] Ming-Hsuan Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: a survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.1, pp.34–58, 2002.
- [5] E. Hjelmås and B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding*, vol.83, no.3, pp.236–274, 2001.
- [6] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol.1, pp.I–I, 2001.
- [7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol.1, pp.886–893 vol. 1, 2005.
- [8] S. Liao, A. K. Jain, and S. Z. Li, “A fast and accurate unconstrained face detector,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.211–223, 2016.
- [9] E. Osuna, R. Freund, and F. Girosit, “Training support vector machines: an application to face detection,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.130–136, 1997.
- [10] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, “Deep learning for understanding faces: Machines may be just as good, or better, than humans,” *IEEE Signal Processing Magazine*, vol.35, no.1, pp.66–83, 2018.
- [11] R. Brunelli and T. Poggio, “Face recognition: features versus templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.15, no.10, pp.1042–1052, 1993.

- [12] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, “2d and 3d face recognition: A survey,” *Pattern Recognition Letters*, vol.28, no.14, pp.1885–1906, 2007. Image: Information and Control.
- [13] JafriRabia and A. R., “A survey of face recognition techniques,” *Journal of Information Processing Systems*, vol.5, no.2, pp.41–68, 2009.
- [14] G. Hu, F. Yan, J. Kittler, W. Christmas, C. H. Chan, Z. Feng, and P. Huber, “Efficient 3d morphable face model fitting,” *Pattern Recognition*, vol.67, pp.366–379, 2017.
- [15] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, “Face recognition with visible and thermal infrared imagery,” *Computer Vision and Image Understanding*, vol.91, no.1, pp.72–114, 2003. Special Issue on Face Recognition.
- [16] H. Wang, Y. Wang, and Y. Cao, “Video-based face recognition: A survey,” *World Academy of Science, Engineering and Technology*, vol.60, pp.293–302, 2009.
- [17] X. Luan, B. Fang, L. Liu, W. Yang, and J. Qian, “Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion,” *Pattern Recognition*, vol.47, no.2, pp.495–508, 2014.
- [18] K. R. Sreelakshmi, R. Anitha, and K. R. Rebitha, “Multiple media based face recognition in unconstrained environments using eigenfaces,” in *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, pp.1–6, 2016.
- [19] N. Khan, R. Ksantini, I. Ahmad, and B. Boufama, “A novel svm+nda model for classification with an application to face recognition,” *Pattern Recognition*, vol.45, no.1, pp.66–79, 2012.
- [20] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.815–823, 2015.
- [21] G. Wen, Y. Mao, D. Cai, and X. He, “Split-net: Improving face recognition in one forwarding operation,” *Neurocomputing*, vol.314, pp.94–100, 2018.
- [22] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1–9, 2015.
- [23] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [24] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1867–1874, 2014.
- [25] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, “Facial landmark detection with tweaked convolutional neural networks,” 2016.
- [26] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for mobilenetv3,” 2019.
- [27] M. Haghhighat, M. Abdel-Mottaleb, and W. Alhalabi, “Fully automatic face normalization and single sample face recognition in unconstrained environments,” *Expert Systems with Applications*, vol.47, pp.23–34, 2016.

- [28] J.-J. Lv, C. Cheng, G.-D. Tian, X.-D. Zhou, and X. Zhou, “Landmark perturbation-based data augmentation for unconstrained face recognition,” *Signal Processing: Image Communication*, vol.47, pp.465–475, 2016.
- [29] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, “Openface: A general-purpose face recognition library with mobile applications,” tech. rep., CMU-CS-16-118, CMU School of Computer Science, 2016.
- [30] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, “Robust face recognition for uncontrolled pose and illumination changes,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.43, no.1, pp.149–163, 2013.
- [31] W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, J. Choi, J. Lekust, J. Kim, P. Natarajan, R. Nevatia, and G. Medioni, “Face recognition using deep multi-pose representations,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp.1–9, 2016.
- [32] I. Masi, S. Rawls, G. Medioni, and P. Natarajan, “Pose-aware face recognition in the wild,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.4838–4846, 2016.
- [33] T. Zhang, Q. Dong, and Z. Hu, “Pursuing face identity from view-specific representation to view-invariant representation,” in *2016 IEEE International Conference on Image Processing (ICIP)*, pp.3244–3248, 2016.
- [34] Junho Yim, Heechul Jung, ByungIn Yoo, Changkyu Choi, Dusik Park, and Junmo Kim, “Rotating your face using multi-task deep neural network,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.676–684, 2015.
- [35] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” 2014.
- [36] Xiangyu Zhu, Z. Lei, Junjie Yan, D. Yi, and S. Z. Li, “High-fidelity pose and expression normalization for face recognition in the wild,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.787–796, 2015.
- [37] C. Ding, C. Xu, and D. Tao, “Multi-task pose-invariant face recognition,” *IEEE Transactions on Image Processing*, vol.24, no.3, pp.980–993, 2015.
- [38] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, “Unconstrained face recognition: Identifying a person of interest from a media collection,” *IEEE Transactions on Information Forensics and Security*, vol.9, no.12, pp.2144–2157, 2014.
- [39] C. Ding and D. Tao, “Pose-invariant face recognition with homography-based normalization,” *Pattern Recognition*, vol.66, pp.144–152, 2017.
- [40] J. Hussain Shah, M. Sharif, M. Raza, M. Murtaza, and Saeed-Ur-Rehman, “Robust face recognition technique under varying illumination,” *Journal of Applied Research and Technology*, vol.13, no.1, pp.97–105, 2015.
- [41] J. Li, B. Li, Y. Xu, K. Lu, K. Yan, and L. Fei, “Disguised face detection and recognition under the complex background,” in *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pp.87–93, 2014.

- [42] C. Y. Wu and J. J. Ding, “Occluded face recognition using low-rank regression with generalized gradient direction,” *Pattern Recognition*, vol.80, pp.256–268, 2018.
- [43] Ya Wang, Tianlong Bao, Chunhui Ding, and Ming Zhu, “Face recognition in real-world surveillance videos with deep learning method,” in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pp.239–243, 2017.
- [44] T. Soyata, R. Muraleedharan, C. Funai, M. Kwon, and W. Heinzelman, “Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture,” in *2012 IEEE Symposium on Computers and Communications (ISCC)*, pp.000059–000066, 2012.
- [45] P. Hu, H. Ning, T. Qiu, Y. Xu, X. Luo, and A. K. Sangaiah, “A unified face identification and resolution scheme using cloud computing in internet of things,” *Future Generation Computer Systems*, vol.81, pp.582–592, 2018.
- [46] L. Di Martino, J. Preciozzi, F. Lecumberry, and A. Fernández, “Face matching with an a contrario false detection control,” *Neurocomputing*, vol.173, 08 2015.
- [47] A. V. Savchenko and N. S. Belova, “Unconstrained face identification using maximum likelihood of distances between deep off-the-shelf features,” *Expert Systems with Applications*, vol.108, pp.170–182, 2018.
- [48] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2016.
- [49] S. Banerjee and S. Das, “Lr-gan for degraded face recognition,” *Pattern Recognition Letters*, vol.116, pp.246–253, 2018.
- [50] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” 2019.
- [51] J. Cao, Y. Hu, B. Yu, R. He, and Z. Sun, “3d aided duet gans for multi-view face image synthesis,” *IEEE Transactions on Information Forensics and Security*, vol.14, no.8, pp.2028–2042, 2019.
- [52] “Casia web-face dataset,” <https://paperswithcode.com/dataset/casia-webface>. Accessed: 2021-04-24.
- [53] “Ms-celeb-1m,” <https://www.microsoft.com/en-us/research/project/ms-celeb-1m-challenge-recognizing-one-million-celebrities-real-world>. Accessed: 2021-04-24.
- [54] “Vggface2 dataset,” https://www.robots.ox.ac.uk/~vgg/data/vgg_face/. Accessed: 2021-04-24.
- [55] “Labeled faces in the wild dataset,” <http://vis-www.cs.umass.edu/lfw/>. Accessed: 2021-04-24.
- [56] “Public figures face database,” <https://www.cs.columbia.edu/CAVE/databases/pubfig/>. Accessed: 2021-04-24.
- [57] “Cross-age celebrity dataset,” <https://paperswithcode.com/dataset/cacd>. Accessed: 2021-04-24.
- [58] “Megaface: Million-scale face recognition dataset,” <http://megaface.cs.washington.edu>. Accessed: 2021-04-24.

Abstract

Ferdowsi University Mashhad (FUM)
Department of Computer

Thesis submitted
for the degree of M.Sc.

Title:

Realtime Face Recognition in Unconstraint Environments

Supervisor: DR. Hamid Reza Pour Reza

By: Sajjad Aemmi

June 2021