

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق
گرایش مخابرات امن و رمزنگاری

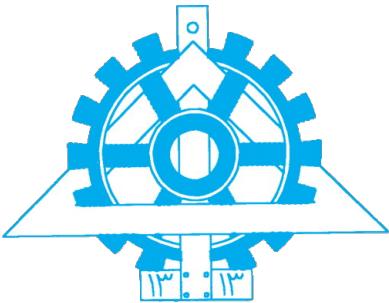
میثم شهبازی دستجرده

استاد راهنما

دکتر محمد علی اخایی

اردیبهشت ۱۴۰۱

سُبْحَانَ رَبِّ الْجَمَلِ



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیووتر



جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره

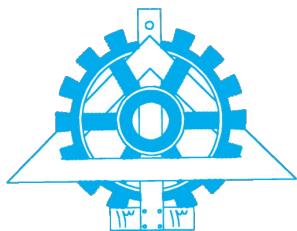
پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق
گرایش مخابرات امن و رمزنگاری

میثم شهبازی دستجرد

استاد راهنما

دکتر محمد علی اخایی

اردیبهشت ۱۴۰۱



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیو تر

گواهی دفاع از پایان نامه کارشناسی ارشد

هیأت داوران پایان نامه کارشناسی ارشد آقای / خانم میثم شهبازی دستجرده به شماره ۱۹۷۲۸۹ در رشته مهندسی برق - گرایش مخابرات امن و رمزنگاری را در تاریخ با عنوان «جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره» به عدد به حروف

با نمره نهايی

و درجه ارزیابی کرد.

امضا	دانشگاه یا مؤسسه	مرتبه دانشگاهی	نام و نام خانوادگی	مشخصات هیأت داوران	ردیف
	دانشگاه تهران	استادیار	دکتر محمد علی اخایی	استاد راهنمای	۱
	دانشگاه تهران	دانشیار	دکتر داور داخلی	استاد داور داخلی	۲
	دانشگاه داور خارجی	دانشیار	دکتر داور خارجی	استاد مدعو	۳
	دانشگاه تهران	دانشیار	دکتر نماینده	نماینده تحصیلات تمکیلی، دانشکده	۴

نام و نام خانوادگی معاون تحصیلات تکمیلی
و پژوهشی دانشکده / گروه:
تحصیلات تکمیلی پردیس دانشکده‌های فنی:

تعهدنامه اصالت اثر

با اسمه تعالیٰ

اینجانب میثم شهبازی دستجرده تأیید میکنم که مطالب مندرج در این پایان نامه حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتری ارائه نشده است.

نام و نام خانوادگی دانشجو: میثم شهبازی دستجرده

تاریخ و امضای دانشجو:

کلیه حقوق مادی و معنوی این اثر
متعلق به دانشگاه تهران است.

این اثر ناچیز تقدیم می‌شود به :

۱۷۶ امید

۹

آرزوی پرپر شده ...

قدردانی

این پایان نامه در زمان همه‌گیری ویروس کرونا، انجام شده است. در زمانی که محدودیت‌های کرونایی موجب غیرحضوری شدن آموزش‌های دانشگاهی شده است. در این شرایط دشوار، حمایت‌های بی‌دریغ جناب آقای دکتر محمدعلی اخایی، پیش از پیش به چشم آمد. بر خود لازم می‌دانم از ایشان به‌دلیل پی‌گیری‌های مرتب جهت پیشبرد پایان‌نامه در این شرایط کرونایی تشکر و قدردانی کنم. همچنین از آقایان رامین طوسی و سید امین حبیبی به‌علت مشاوره و راهنمایی‌های ارزنده تشکر می‌کنم. همچنین از آقای پویا نریمانی به‌علت مساعدت در اتصال از راه دور به رایانه‌های موجود در آزمایشگاه مخابرات امن و رمزنگاری، تشکر می‌کنم

و در پایان، بوسه می‌زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می‌کنم وجود مقدس‌شان را و تشکر می‌کنم از خانواده عزیزم به پاس عاطفه سرشار و گرمای امیدبخش وجودشان، که بهترین پشتیبان من بودند.

میثم شهبازی دستجرده

اردیبهشت ۱۴۰۱

چکیده

یکی از روش‌های احراز هویت خودکار، استفاده از چهره کاربر است. با توجه به پیشرفت‌های چشم‌گیر در حوزه تشخیص چهره، استفاده از چهره محبوبیت خاصی پیدا کرده است. در عین حال، استفاده از چهره برای احراز هویت، روشی به‌طور کامل امن نیست و فرد مهاجم می‌تواند با استفاده از چاپ کردن چهره فرد هدف، یا بازپخش ویدیویی از او، به جای فرد هدف، احراز هویت انجام دهد. از این رو روش‌ها و الگوریتم‌هایی در این حوزه برای بهبود امنیت سیستم‌های احراز هویت با چهره، در تحقیقات دانشگاهی و صنعتی توسعه داده شده است. هدف از این پژوهشها تشخیص و تمیز تصویر چهره واقعی از تصویر چهره تقلیبی ارائه شده توسط فرد مهاجم است. با رشد استفاده از روش‌های یادگیری عمیق در مسائل بینایی ماشین، در این حوزه نیز از الگوریتم‌های یادگیری عمیق برای طبقه‌بندی تصویر واقعی در مقابل تصاویر تقلیبی ارائه شده توسط فرد مهاجم، استفاده شده است. در این پایان‌نامه با ترکیب روش کلاسیک بینایی ماشین و روش‌های یادگیری عمیق، یک عملگر جدید برای جایگزین کردن در یکی از لایه‌های کانولوشن ارائه شده است. همچنین برای افزایش دقت طبقه‌بندی بین دو دسته تصویر واقعی و تقلیبی تابع هزینه‌های برای دسته‌بندی دودویی با حاشیه ارائه شده است که افزودن این حاشیه باعث می‌شود نمونه‌های دو کلاس از یکدیگر فاصله داشته باشند. علاوه بر این برای افزایش قابلیت تعمیم‌پذیری شبکه، تابع هزینه‌ی متریک اختصاصی برای مسئله کشف تقلب در چهره، با کمک گرفتن از شناسه اشخاص پیشنهاد شده است. همچنین نتایج روی برخی از دیتاست‌های معروف در این حوزه، گزارش شده و عملکرد کلی الگوریتم پیشنهادی به همراه سرعت اجرا بحث شده است.

واژگان کلیدی احراز هویت، استفاده از چهره، امینت سیستم‌های احراز هویت، ترکیب روش‌های بینایی ماشین با یادگیری عمیق، تابع هزینه با حاشیه، بایومتریک، تابع هزینه متریک اختصاصی

فهرست مطالب

ت	فهرست تصاویر
ج	فهرست جداول
چ	فهرست الگوریتم‌ها
ح	فهرست برنامه‌ها
۱	فصل ۱: مقدمه
۱	۱.۱ پیشگفتار
۳	۲.۱ اهداف
۴	۳.۱ دستاوردهای پژوهش
۵	۴.۱ ساختار پایان‌نامه
۷	فصل ۲: مروری بر مطالعات انجام شده
۷	۱.۲ مقدمه
۸	۲.۲ روش‌های کلاسیک
۹	۱.۲.۲ تحلیل ریز بافت و عملگر LBP
۱۳	۳.۲ روش‌های مبتنی بر یادگیری عمیق
۱۴	۱.۳.۲ ترکیب روش‌های یادگیری عمیق و ویژگی‌های دستی
۱۶	۲.۳.۲ استفاده از تخمین سیگنال کمکی
۲۲	۳.۳.۲ استفاده از شبکه‌های مولد تخصصی و تابع هزینه‌های مختلف

۴.۲	دیتاست‌های مورد استفاده	۲۸
۱.۴.۲	دیتاست Replay	۲۸
۲.۴.۲	دیتاست CASIA	۲۹
۳.۴.۲	دیتاست MSU	۲۹
۴.۴.۲	دیتاست OULU	۳۰
۵.۴.۲	دیتاست SIW	۳۱
فصل ۳: روش پیشنهادی		
۱.۳	مقدمه	۳۳
۲.۳	مروری بر عملگر کانولوشن	۳۴
۳.۳	عملگر LBP قابل آموزش	۳۵
۴.۳	ساختار شبکه	۳۷
۵.۳	تابع هزینه ARCB	۳۹
۶.۳	تابع هزینه بر اساس شناسه‌ی شخص	۴۲
۷.۳	مقایسه‌ی روش پیشنهادی با پژوهش‌های قبلی	۴۶
فصل ۴: نتایج		
۱.۴	مقدمه	۴۹
۲.۴	ملاحظات پیاده‌سازی	۴۹
۱.۲.۴	پیاده‌سازی LBP قابل آموزش	۵۰
۲.۲.۴	پیاده‌سازی تابع هزینه	۵۰
۳.۲.۴	بارگذاری داده‌ها برای آموزش	۵۰
۳.۴	معیارهای ارزیابی	۵۲
۴.۴	عملکرد مدل در دیتاست‌ها	۵۵
۱.۴.۴	اثر عملگر LBP قابل آموزش در دیتاست Replay	۵۵
۲.۴.۴	اثر تابع هزینه ARCB در دیتاست Replay	۵۷
۳.۴.۴	اثر تابع هزینه بر پایه شناسه‌ی اشخاص در دیتاست Replay	۵۸

۵۹	نتایج روی دیتاست‌های MSU و CASIA ۴.۴.۴
۶۰	دقت در دیتاست SIW ۵.۴.۴
۶۱	دقت در دیتاست OULU ۶.۴.۴
۶۲	نتایج در آزمون بین دیتاست ۷.۴.۴
۶۵	فصل ۵: نتیجه‌گیری و کارهای آینده
۶۵	۱.۵ نتیجه‌گیری
۶۶	۲.۵ پیشنهاد کارهای آینده
۶۷	مراجع

فهرست تصاویر

۱.۱	نمونه‌ای از تصاویر واقعی و تقلبی در حوزه چهره [۱]	۳
۱.۲	ساختار کلی الگوریتم‌های کشف تقلب در چهره	۸
۲.۲	مثالی از محاسبه LBP [۲]	۱۰
۳.۲	روش تصمیم‌گیری بر اساس استفاده از LBP [۲]	۱۲
۴.۲	روش تحلیل ریزبافت در نواحی مختلف تصویر [۳]	۱۳
۵.۲	حالات مختلف ترکیب ویژگی‌های دستی و ویژگی‌های شبکه عمیق [۴]	۱۴
۶.۲	استفاده از شبکه تنظیم دقیق شده و اعمال PCA روی ویژگی‌های عمیق [۵]	۱۵
۷.۲	روش ترکیب LBP و کانولوشن [۶]	۱۵
۸.۲	روش‌های مختلف یادگیری عمیق در حوزه‌ی کشف تقلب چهره [۴]	۱۶
۹.۲	استفاده از عمق برای کشف تقلب در چهره [۷]	۱۷
۱۰.۲	روش استفاده از عمق و تخمین rPPG [۸]	۱۸
۱۱.۲	استفاده از ویژگی‌های عمیق در طول زمان [۹]	۱۸
۱۲.۲	نحوه محاسبه تابع هزینه CDL [۱۰]	۱۹
۱۳.۲	عملگر کانولوشن تغییر یافته [۱۰]	۲۰
۱۴.۲	روش استفاده از فیلتر bilateral در شبکه عمیق [۱۱]	۲۱
۱۵.۲	تابع هزینه BCE روی یک صفحه مسطح به جای یک نورون [۱۲]	۲۲
۱۶.۲	ساختار بر پایه استفاده از شبکه مولد برای تخمین علائم تقلب در سطوح مختلف	۲۳
۱۷.۲	نحوه عملکرد تابع هزینه سه‌گانه روی فاصله بردارهای ویژگی [۱۴]	۲۳

۱۸.۲	نحوه اثر تابع هزینه روی فاصله نمونه‌ها در دیتاست‌های مختلف [۱۵]	۲۴
۱۹.۲	تابع هزینه نامتقارن برای کاهش فاصله نمونه‌های از یک کلاس [۱۶]	۲۵
۲۰.۲	ساختار U-net و تابع هزینه سه‌گانه [۱۷]	۲۵
۲۱.۲	کاهش فاصله نمونه‌های واقعی تا مرکز و افزایش فاصله نمونه‌های تقلبی تا مرکز [۱۸]	۲۶
۲۲.۲	استفاده از LBP در کنار عمق برای یافتن ویژگی‌های خوش ساخت [۱۹]	۲۷
۲۳.۲	نمونه‌هایی از دیتاست Replay [۳]	۲۸
۲۴.۲	نمونه‌هایی از دیتاست CASIA [۲۰]	۲۹
۲۵.۲	نمونه‌هایی از دیتاست MSU [۲۱]	۳۰
۲۶.۲	نمونه‌های واقعی در دیتاست OULU [۱]	۳۰
۲۷.۲	نمونه‌های تقلبی در دیتاست OULU [۱]	۳۱
۲۸.۲	نمونه‌های از دیتاست SIW [۸]	۳۲
۱.۳	شمای کلی روش پیشنهادی	۳۸
۲.۳	مقایسه تابع هزینه BCE کلاسیک با نسخه‌ی حاشیه‌دار ARCB	۴۲
۳.۳	حالتی که دو نمونه متعلق به یک شخص ولی یکی واقعی و دیگری تقلبی است	۴۳
۴.۳	حالتی که دو نمونه متعلق به اشخاص مختلف ولی برچسب یکسان هستند	۴۵
۱.۴	نحوه برش زدن تصادفی چهره با مقداری از پس‌زمینه	۵۲
۲.۴	نمودار میزان خطای برابر	۵۴
۳.۴	نمودار خطای برابر برای شبکه ALEXNET و تابع هزینه BCE	۵۶
۴.۴	نمودار خطای برابر هنگام استفاده از عملگر LBP پیشنهادی	۵۶
۵.۴	نمودار خطای برابر هنگام استفاده از شبکه EfficientNet B0	۵۷
۶.۴	نمودار خطای برابر هنگام استفاده از تابع هزینه ARCB پیشنهادی	۵۸
۷.۴	نمودار خطای برابر با استفاده از تابع هزینه مبتنی بر شناسه اشخاص	۵۹

فهرست جداول

۱.۳	ساختار شبکه پیشنهادی	۳۹
۱.۴	خطای برابر روی دیتاست‌های CASIA و MSU	۵۹
۲.۴	نرخ در پروتکل اول دیتاست SIW	۶۰
۳.۴	نرخ در پروتکل دوم دیتاست SIW	۶۱
۴.۴	دقت در پروتکل اول دیتاست OULU	۶۱
۵.۴	دقت در پروتکل دوم دیتاست OULU	۶۲
۶.۴	نتایج روی آزمون بین دیتاست	۶۳

فهرست الگوریتم‌ها

فهرست برنامه‌ها

فصل ۱

مقدمه

۱.۱ پیشگفتار

یک سیستم احراز هویت به وسیله چهره را در نظر بگیرید که کاربر در مقابل دوربین قرار گرفته و سیستم از طریق تایید مشخصات چهره، به او اجازه دسترسی می‌دهد. حال فرض کنید کاربر غیر مجاز تصویر کاربر قبلًا تایید شده در سیستم را روی کاغذ چاپ کند و کاغذ را در مقابل دوربین سیستم قرار دهد. در این صورت کاربر غیر مجاز می‌تواند خود را به جای کاربر مجاز به سیستم بشناساند و به اطلاعات محرومانه فرد دیگری، به کمک تنها یک تصویر چاپ شده، دسترسی پیدا کند. این یک مثال ساده برای تداعی مشکل امنیتی سیستم‌های احراز اصالت با چهره است.

هر چه محramانگی و اهمیت اطلاعات ذخیره شده درون سیستم بیشتر باشد، مشکل امنیتی ذکر شده توجه بیشتری می‌طلبد. برای مثال فرض کنید سیستم مذبور به اطلاعات حساب بانکی یا اوراق بهادر یا داده‌های محرومانه یک شرکت تجاری مرتبط باشد؛ در این صورت تمامی این اطلاعات حیاتی در معرض خطر آسیب پذیری فرآیند تشخیص و تایید چهره خواهد بود.

این مشکل امنیتی موجب پیدایش زمینه‌ای از تحقیقات در دانشگاه و صنعت شده است که در ادبیات موضوع «جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره^۱» نام دارد. در این عنوان، قسمت احراز هویت مبتنی بر تشخیص چهره در واقع شاخه از بایومتریک^۲ است و قسمت جلوگیری از تقلب، به مسائل امنیتی کار می‌پردازد. هدف از بایومتریک، تشخیص خودکار افراد بر

¹Anti-spoofing for authentication based on face recognition

²Biometric

اساس ویژگی‌های زیست‌شناختی و یا رفتار اشخاص است. برای مثال چهره، عنایت، اثر انگشت، صدا و طرز راه رفتن نمونه از ویژگی‌هایی است که هر فرد را به صورت منحصرًا از فرد دیگر تمیز می‌دهد. تأکید بایومتریک بر «خودکار بودن» فرآیند تشخیص فرد است؛ به همین دلیل لازم است که دخالت انسان در این فرآیند حداقل شود و سیستم به صورت غیر نظارتی^۳ فرد را تشخیص دهد.

در میان شاخصه‌های ذکر شده برای کاربرد بایومتریک، استفاده از چهره اهمیت خاصی دارد. روش‌های بینایی ماشین برای تشخیص چهره سابقه طولانی دارند و به تازگی راه حل‌های استفاده از هوش مصنوعی، تشخیص چهره را دقیق‌تر و متداول‌تر کرده است. از طرفی چهره در مقایسه با اثر انگشت یا صدا و... نمایان‌گر آشناتر برای شناسایی یک فرد است. این ویژگی‌های چهره چه در ابزار شناسایی چه در قرابت استفاده، موجب شده است تشخیص چهره، کاربردهای دیگری نظیر پزشکی قانونی، دوربین‌های مدار بسته، اجازه کنترل و دسترسی به سیستم، و دولت و تجارت الکترونیک داشته باشد.

این کاربرد گسترده و رشد استفاده از چهره در سیستم‌ها، مسائل امنیتی را نیز به همراه دارد. فرد مهاجم به راحتی و با هزینه‌ی کمی می‌تواند تصویر فرد مورد نظر خود را از طریق شبکه‌های اجتماعی یا تصویربرداری از فاصله‌ی دور به دست آورد و اقدامات لازم برای حمله را به عمل آورد.

این نوع حمله با ابزارهای مختلفی می‌تواند صورت بگیرد. برای مثال مهاجم می‌تواند تصویر فرد هدف را روی کاغذ چاپ کند، یا از یک فیلم یا تصویر ذخیره شده در نمایشگر دیجیتال استفاده کند. همچنین با استفاده از گریم یا ماسک می‌تواند چهره خود را شبیه به چهره فرد هدف کند. در میان انواع حمله ذکر شده استفاده از چاپ تصویر و استفاده از نمایشگر دیجیتال متداول‌تر است. استفاده از ماسک به دلیل هزینه بالا و سختی اجرا، چندان متداول نیست.

با توجه به اهمیت موضوع و نگرانی در مورد امنیت سیستم‌های احراز هویت مبتنی بر تشخیص چهره، تحقیقات فراوانی در دانشگاه برای فائق آمدن بر این چالش انجام شده است. که دامنه وسیعی از روش‌های مبتنی بر بینایی ماشین کلاسیک و روش‌های جدیدتر مبتنی بر هوش مصنوعی و یادگیری عمیق را شامل می‌شود.

این چالش امنیتی می‌تواند از دید یک مسئله‌ی بینایی ماشین تعریف شود؛ به گونه‌ای که ورودی مسئله، تصویر از چهره یک فرد است و خروجی سیستم، یک برچسب چهره واقعی یا تقلیبی است. دقت الگوریتم برای اعلام این برچسب‌گذاری، سهم مهمی در امینت کلی سیستم خواهد داشت. در برخی از روش‌ها از اطلاعات بیشتری نظیر سنسور حرارتی و یا مادون قرمز در کنار تصویر استفاده می‌شود اما این امر موجب افزایش هزینه خواهد شد. همچنین الگوریتم‌ها بر اساس استفاده از تنها

³Unsupervised

یک تصویر یا یک دنباله ویدیویی نیز قابل تقسیم هستند.

با وجود تلاش‌های تحقیقاتی در این زمینه که بیش از یک دهه قدمت دارد همچنان مسئله کشف تقلب در تشخیص چهره یک مسئله چالشی می‌باشد. یکی از دلایل چالشی بودن آن، خلاقیت فرد مهاجم برای اعمال حمله جدید است؛ به‌گونه‌ای که این حمله جدید قبلًا در داده‌های مورد استفاده برای توسعه الگوریتم وجود نداشته باشد. یک چالش دیگر تفاوت کیفیت و رزولوشن ابزارهای حمله، نظیر صفحه نمایش و کاغذ چاپ است. این مسئله زمانی بغرنج‌تر می‌شود که حتی برای کاربر انسانی نیز تمیز چهره واقعی و تقلبی دشوار خواهد شد. برای مثال در شکل ۱.۱ یکی از تصاویر تقلبی و دیگری واقعی است. همانطور که مشاهده می‌شود تشخیص چهره واقعی از تقلبی به آسانی میسر نیست.



[۱]: نمونه‌ای از تصاویر واقعی و تقلبی در حوزه چهره [۱]

۲.۱ اهداف

در این پایان‌نامه برای کشف تقلب در تصویر چهره، تمرکز بر روش‌هایی است که تنها از یک تصویر رنگی به جای دنباله ویدیویی یا اطلاعات اضافی نظیر سنسور حرارتی و مادون قرمز، به عنوان ورودی استفاده می‌شود. این رویکرد موجب کاهش هزینه سیستم و قابل استفاده بودن بیشتر خواهد شد. همچنین از انواع حمله‌های مختلف موجود، تنها موارد چاپ روی کاغذ و بازپخش ویدیو بررسی

می‌گردد. با آن که حمله‌های دیگری نظیر استفاده از ماسک سه بعدی نیز وجود دارد اما اعمال چنین حمله‌هایی هزینه‌بر و دشوارتر از نظر اجرا است. بنابرین توجه پایان‌نامه روی حملاتی است که متدالوئر و بیشتر قابل اجرا است.

در این پایان‌نامه با ترکیب روش کلاسیک بینایی ماشین و روش‌های جدید یادگیری عمیق ساختاری برای طبقه‌بندی دقیق‌تر ارائه شده است. این ساختار شامل یک عملگر جدید است که از عملگر LBP کلاسیک الهام گرفته شده است، با این تفاوت که این عملگر همانند عملگر کانولوشن در شبکه‌های عمیق دارای پارامتر برای یادگیری عملگر بهینه با توجه به داده‌های ورودی است. همچنین دوتابع هزینه جدید ارائه شده است. تابع هزینه اول با افزودن یک حاشیه به طبقه‌بندی موجب می‌شود ویژگی‌های دو کلاس با فاصله از یک دیگر قرار بگیرند که موجب افزایش دقت رو داده‌های دیده نشده می‌گردد. تابع هزینه دوم بر اساس شناسه اشخاص مختلف موجود در دیتابست توسعه داده شده است و موجب می‌شود که شبکه عصبی تمرکز بیشتری روی ویژگی‌های تقلب موجود در چهره داشته باشد و به ویژگی‌های ظاهری افراد توجه نکند. که این موجب افزایش قابلیت تعمیم‌پذیری شبکه روی داده‌های آزمون دیده نشده می‌شود.

۳.۱ دستاوردهای پژوهش

در این پایان‌نامه، پس از بیان روش پیشنهادی به صورت ریاضی با آزمایش‌های مختلف روی دیتابست‌های در دسترس و محاسبه نرخ خطای استاندارد در این حوزه، نشان داده می‌شود روش ارائه شده شامل عملگر تحلیل ریزبافت و تابع هزینه جدید موجب افزایش دقت طبقه‌بندی و تعمیم‌پذیری آن می‌شود. قسمت‌های مختلف روش پیشنهادی هر کدام به صورت مجزا، ابتدا روی یک دیتابست کوچک تست شده است و اثر بخشی هر قسمت بررسی شده است. سپس تمام روش پیشنهادی روی دیتابست‌های بزرگ‌تر پیاده شده و معیار خطا با مقادیر به دست آمده در برخی از پژوهش‌های مهم در این حوزه مقایسه شده است. این مقایسه نشان می‌دهد روش پیشنهادی به نتایج رقابتی با نتایج این پژوهش‌ها می‌رسد.

همچنین برای پیاده‌سازی، برنامه‌نویسی به زبان پایتون انجام شده است و ملاحظات پیاده‌سازی و چالش‌های مربوط به آن، توضیح و تفسیر شده است. علاوه بر این، برای کار کردن با داده‌های ویدیویی و استفاده از آن در شبکه‌هایی که ورودی تصویر دارند، الگوریتمی ارائه شده است که روند آموزش شبکه را تسريع ببخشد. کدهای مرتبط با برنامه در یک مخزن گیت‌هاب^۴ به صورت متن‌باز منتشر

⁴<https://github.com/meysamshahbazi/fas>

شده است. برنامه به گونه‌ای نوشته شده است که نتایج آن قابل بازتولید باشد.

۴.۱ ساختار پایان‌نامه

در فصل دو، ابتدا مروری بر پژوهش‌های انجام شده در حوزه کشف تقلب انجام می‌شود. تحقیقات انجام شده در این حوزه بسیار وسیع است و تنها به مرور روش‌هایی که اهمیت بیشتر در ادبیات موضوع و روش‌هایی که رویکرد مشابهی با این پایان‌نامه داشته‌اند پرداخته می‌شود. در فصل سه، روش پیشنهادی به صورت مبانی نظری گفته می‌شود و در فصل چهار، ابتدا ملاحظات پیاده‌سازی روش ارائه شده بیان می‌گردد و سپس با استفاده از معیارهای ارزیابی متدالول در این حوزه، به بررسی دقیق روش پیشنهادی پرداخته می‌شود. فصل آخر به نتیجه‌گیری و بحث در مورد روش پیشنهادی می‌پردازد.

فصل ۲

مروری بر مطالعات انجام شده

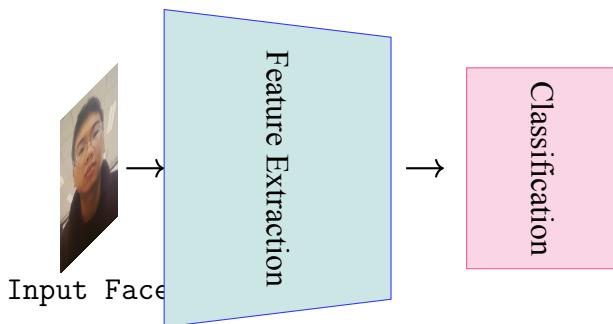
۱.۲ مقدمه

این فصل به مروری بر برخی از مهم‌ترین روش‌های موجود در حوزه کشف تقلب می‌پردازد. در ابتدا دسته‌بندی کلی برای حل مسئله کشف تقلب ارائه می‌شود و سپس دامنه تمرکز روی یک شاخه از این روش‌ها محدود می‌گردد. هرچند که امروزه استفاده از روش‌های یادگیری عمیق گسترش فراوان یافته است و در بسیاری از مسائل بینایی ماشین، روش‌های کلاسیک منسوخ شده‌اند؛ اما این از اهمیت روش‌های کلاسیک نمی‌کاهد. روش‌های کلاسیک بینایی ماشین در مقایسه با روش‌های مبتنی بر یادگیری عمیق، از آنجا که تمرکز بیشتری روی الگوریتم تا تمرکز روی استفاده از داده داشته‌اند، می‌توانند دید میدانی خوبی از نزدیک شدن به مسئله بدهند.

در این پایان‌نامه سعی شده است که از این دید کلاسیک برای حل مسئله با بهره گرفتن از ابزارهای یادگیری عمیق استفاده شود. پس در این فصل ابتدا روش‌های کلاسیک مورد بررسی قرار می‌گیرند و سپس مروری بر روش‌های مبتنی بر یادگیری عمیق انجام می‌گیرد. همانطور که در شکل ۱.۲ مشخص است روش کلی الگوریتم‌های کشف تقلب و به‌طور کلی بسیاری از مسائل بینایی ماشین ابتدا استخراج ویژگی از تصویر یا ویدیوی ورودی است و سپس طبقه‌بندی ویژگی‌های به‌دست آمده است. استخراج ویژگی نقش مهمی در دقت طبقه‌بندی خواهد داشت. یک استخراج ویژگی، یک تابع از تصویر ورودی به یک بردار است و زمانی استخراج ویژگی به درستی انجام گرفته است که بردار خروجی شامل اطلاعات اساسی و مهم برای طبقه‌بندی صحیح باشد.

تفاوت عمدی الگوریتم‌های کلاسیک و یادگیری عمیق در قسمت استخراج ویژگی است. بدین

صورت که در روش‌های کلاسیک، ویژگی‌ها با استفاده از یک روش ایستا انتخاب می‌شوند ولی در روش‌های شبکه عصبی عمیق با استفاده از بهینه‌سازی یک تابع هزینه، روی داده‌های آموزش، استخراج ویژگی‌های مد نظر یاد گرفته می‌شوند.



شکل ۱.۲: ساختار کلی الگوریتم‌های کشف تقلب در چهره

۲.۲ روش‌های کلاسیک

در روش‌های کلاسیک، با استفاده از الگوریتم‌های بینایی ماشین، سعی در یافتن یک مؤلفه‌ی مفید از تصویر است که به آشکار ساختن علائم و استخراج ویژگی‌های مربوط به تقلب در تصویر کمک کند. روش‌های کلاسیک به دو دسته سخت‌افزاری و نرم‌افزاری تقسیم می‌شوند. [۲۲]

در روش‌های سخت‌افزاری یا از یک سخت‌افزار خاص استفاده می‌شود، یا از یک تعامل فیزیکی با کاربر نظیر چشمک زدن و یا پاسخ به یک چالش استفاده می‌گردد. در حالت استفاده از سخت افزار خاص، یک دوربین حرارتی یا چند طیفی به کار برده می‌شود. در این حالت تمایز بین تصویر صورت واقعی و یک کاغذ از طریق بررسی طیف نوری یا حرارت مشخص می‌گردد. در حالاتی دیگر از کاربر خواسته می‌شود یک سری کلمات را ادا کرده یا با دست خود حرکت خاصی را انجام دهد. لازم به ذکر است که در روش‌های سخت‌افزاری، قسمت نرم‌افزار حذف نمی‌شود و پردازش‌ها به صورت خاص متناسب با سخت‌افزار خواهد بود. این بدین معنی است که استفاده از سخت‌افزار، طراحی الگوریتم را حذف نخواهد کرد، بلکه نوع الگوریتم، خاص منظوره بر اساس سخت‌افزار مورد استفاده خواهد شد. مشکل روش‌های سخت‌افزاری این است که هزینه اضافی دارد و تعامل بیشتر کاربر با سیستم را تحمیل می‌کند. تعامل بیشتر، زمان احراز هویت را طولانی‌تر می‌کند که مطلوب نیست. همچنین برای به کار بردن الگوریتم در تلفن همراه، مطلوب این است که الگوریتم‌ها تنها از سنسور دوربین موجود استفاده کنند و نیاز به سخت‌افزار اضافه نباشد [۲۲].

در روش‌های نرم‌افزاری از سخت افزار اضافه‌ای استفاده نمی‌شود؛ و تنها از همان دوربین معمولی، تصویربرداری صورت می‌گیرد؛ اما از یک الگوریتم هوشمند بر پایه‌ی بینایی ماشین استفاده خواهد شد. روش‌های نرم‌افزاری به دو دسته ایستان و پویا تقسیم می‌شود.

در روش‌های ایستان، پردازش تنها روی یک فریم تصویر انجام می‌شود و تقلب را با اطلاعات تک تصویر بررسی می‌کند؛ هر چند که این روش‌ها را در دنباله ویدیویی نیز می‌توان به کار برد و روی هر فریم، این پردازش صورت بگیرد. روش‌های ایستان هزینه محاسباتی کمتری در مقایسه با روش‌های پویا دارند. یکی از معروف ترین روش‌های ایستان استفاده از تحلیل ریز بافت^۱ است که در آن از عملگر الگوهای محلی دودویی^۲ (LBP) استفاده می‌شود. این عملگر می‌تواند از تصویر ویژگی‌های مربوط به بافت تصویر را استخراج کند [۲، ۳]. همچنین از روش‌های استخراج ویژگی نظری SIFT [۲۳] و SURF [۲۴] استفاده شده است. یک متدهای کلاسیک دیگر در دسته روش‌های ایستان استفاده از هیستوگرام گرادیان‌های جهت دار^۳ (HoG) است [۲۵، ۲۶، ۲۷].

در روش‌های پویا از اطلاعات فریم‌های متوالی نیز در کنار هم استفاده می‌شود و برای تحلیل، وابستگی فریم‌های متوالی بررسی می‌شود. روش‌های پویا در مقایسه با روش‌های ایستان زمان پردازش بیشتری دارند اما دقت بهتری را ارائه می‌کنند. در روش پویا از حرکت عضلات صورت به‌وسیله حرکت سر، دهان و چشم بهره برده می‌شود. الگوریتم‌های مورد استفاده در این دسته از روش‌ها در بیشتر موارد بر مبنای الگوریتم جریان نوری^۴ است [۲۸، ۲۹]. علاوه بر استفاده از حرکت چهره، می‌توان با استخراج نقاط کلیدی چهره در فریم‌های متوالی تخمین سه بعدی یا عمق چهره استخراج شود که این عمق تخمین زده شده متفاوت در تصاویر واقعی و تقلبی متفاوت خواهند بود. بدین منظور روش‌های بر پایه تخمین ساختار سه بعدی چهره با استفاده از یک دوربین نیز توسعه داده شده است [۳۰، ۳۱]. تغییرات بافت در بین فریم‌های متوالی نیز می‌تواند یک نشانه مفید برای کشف تقلب باشد که برای این منظور عملگر LBP در سه صفحه عمود بر هم توسعه داده شده است [۳۲].

۱.۲.۲ تحلیل ریز بافت و عملگر LBP

در میان روش‌های نرم‌افزاری ذکر شده، تحلیل ریزبافت در این پایان‌نامه اهمیت بسزایی دارد. یکی از تفاوت‌های بین تصویر واقعی و تقلبی در بررسی بافت اجزای صورت در مقیاس ذره‌بینی^۵

¹Micro texture analysis

²Local binary patterns

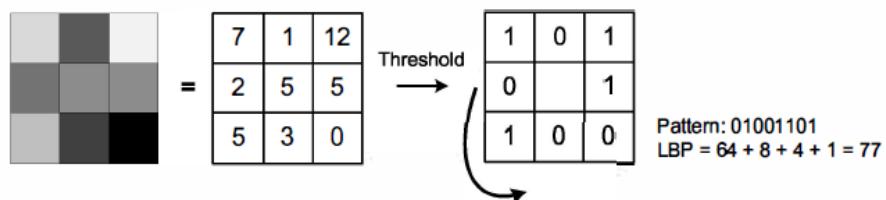
³Histogram of oriented gradients

⁴Optical flow

⁵Microscopy

است. در این مقیاس اثر دانه‌ای چاپ تصویر روی کاغذ منجر به تفاوت با بابت طبیعی چهره انسان می‌شود. همچنین صورت انسان در مقایسه با تصویر نمایش داده شده روی نمایشگر دیجیتال از نظر بابت پیکسلی متفاوت خواهد بود. همچنین صورت واقعی در مقایسه با تصویر چاپ شده یا بازپخش شده روی نمایشگر دیجیتال از نظر انعکاس نور و بازتاب و تشکیل سایه تفاوت دارد. علاوه بر این‌ها تصاویر تقلیلی در مجموع کمی تاری در کیفیت خود دارند. از این رو مسئله کشف تقلب، شباهت‌هایی با مسائل تحلیل کیفیت تصاویر و نهان‌کاوی دارد.

در [۲] برای اولین بار از عملگر الگوهای دودویی محلی یا به اختصار LBP، در حوزه کشف تقلب در چهره استفاده شده است. این عملگر از تعریف بابت از در یک همسایگی در مقیاس محلی الهام گرفته است و یک توصیف‌گر قوی بابت است. بهمنظور آشنایی اولیه، این عملگر ابتدا در یک پنجره سه در سه تعریف می‌شود و سپس رابطه محاسبه آن به صورت کلی تعریف می‌شود. در شکل ۲.۲ مثالی از محاسبه این عملگر در پنجره سه در سه نشان داده شده است.



شکل ۲.۲: مثالی از محاسبه LBP [۲]

ابتدا پیکسل‌های کناری با پیکسل میانی مقایسه می‌شوند، سپس بر مبنای بزرگ‌تر یا کوچک‌تر بودن مقادیر از پیکسل میانی مقدار یک یا صفر به آنها اختصاص داده می‌شود و سپس این دنباله دودویی در یک جهت دایره‌ای خوانده و یک عدد هشت بیتی می‌دهد. در پنجره سه در سه ۸ پیکسل مجاور موجود هست و تعداد حالت‌هایی که خروجی عملگر می‌تواند داشته باشد برابر با $2^8 = 256$ است.

تعریف رسمی این عملگر به صورت کلی برای شعاع R و تعداد نقاط نمونه برداری P در محیط دایره به صورت رابطه ۱.۲ است.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(I_p - I_c)2^p \quad (1.2)$$

که در آن $s(\cdot)$ یک تابع غیر خطی است.

$$s(x) = \begin{cases} 1 & x \geq 0; \\ 0 & \text{otherwise}. \end{cases}$$

این رابطه بیان می‌کند برای محاسبه ریزبافت هر پیکسل در هر نقطه ابتدا یک دایره به شعاع R در نظر گرفته و روی محیط آن P نقطه به فواصل مساوی باید انتخاب شود. در صورتی که برخی نقاط انتخاب شده روی پیکسل خاصی قرار نگیرد باید با استفاده از درون‌یابی دو خطی^۶، مقدار پیکسلی به آن تشخیص داده شود. سپس مقدار این پیکسل‌های روی دایره با پیکسل مرکز دایره مقایسه شده و دنباله دودویی ایجاد می‌گردد. این عمل بدین صورت ادامه می‌یابد که مرکز دایره لغزانده شده و هر بار برای هر پیکسل تصویر ورودی، مقدار LBP محاسبه می‌گردد.

یکی از ویژگی‌های این عملگر، مقاوم بودن در برابر تغییرات یکسان پیکسل‌های تصویر ورودی است. فرض کنید تمامی پیکسل‌ها در یک عدد ثابت ضرب شده یا با یک مقدار ثابت جمع شوند در این صورت به علت اینکه خروجی تابع غیرخطی تغییر نخواهد کرد مقدار نهایی خروجی LBP تغییری نمی‌کند. همچنین این عملگر باز محاسباتی کمی دارد پس سریع است. تفاضل گیری و اعمال تابع غیرخطی $(.)^s$ ساده است و اعمال ضریب 2^p به کمک شیفت، قابل انجام است.

$$I \rightarrow \alpha I \rightarrow s(\alpha I_p - \alpha I_c) = s(I_p - I_c) \quad (2.2)$$

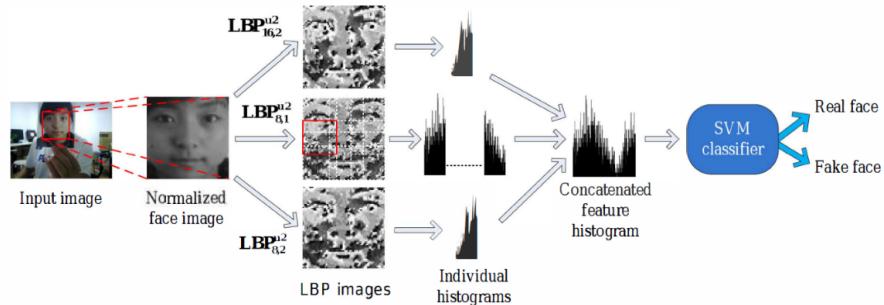
$$I \rightarrow I + \beta \rightarrow s((I_p + \beta) - (I_c + \beta)) = s(I_p - I_c) \quad (3.2)$$

یک نسخه تکامل یافته از LBP، نسخه‌ی یکنواخت این عملگر است که با LBP^{u2} نشان داده می‌شود. این عملگر از این رو معرفی شده است که برخی از الگوهای دودویی بیشتر از سایرین در تصویر متداول‌اند. یک LBP را یکنواخت گویند اگر حداقل دو تغییر از صفر به یک یا بر عکس در نمایش دودویی آن به صورت چرخشی وجود داشته باشد. برای محاسبه برچسب خروجی در حالت یکنواخت، هر الگوی یکنواخت با یک مقدار مجزا نشان داده می‌شود و تمامی حالت‌های غیر یکنواخت به یک مقدار متناظر می‌شوند.

هر خروجی LBP می‌تواند نمایانگر وجود یک نوع الگوی ریزبافت باشد. برای مثال یک LBP با مقدار خاص می‌تواند نشانگر نقطه، گوش، مسطح و... باشد. پس فراوانی این الگوها در تصویر اهمیت دارد. پس از محاسبه LBP به ازای هر پیکسل تصویر، هیستوگرام آن محاسبه می‌شود و از طریق

⁶Bi linear interpolation

توزیع فراوانی الگوهای ریزبافت‌های متفاوت موجود در تصویر، در مورد واقعی یا غیر واقعی بودن آن تصمیم‌گیری می‌شود.

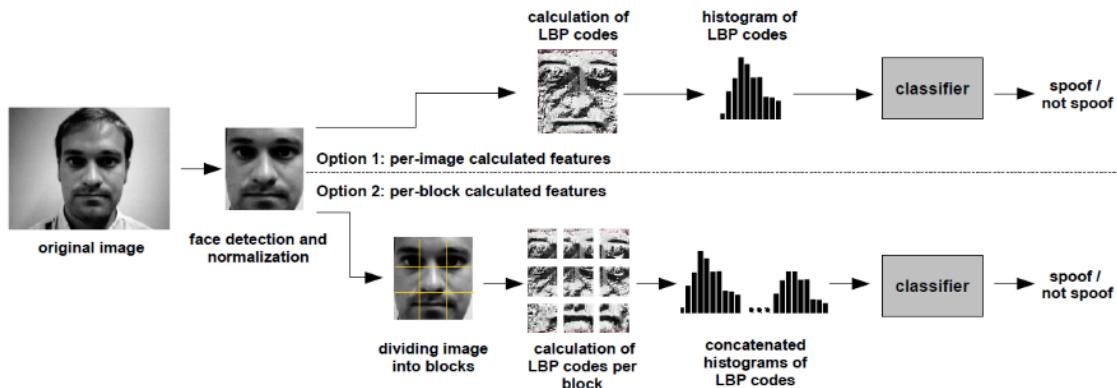


شکل ۳.۲: روش تصمیم‌گیری بر اساس استفاده از LBP [۲]

روش محاسبه و تصمیم‌گیری ارائه شده در [۲] در مورد واقعی یا تقلبی بودن تصویر چهره با استفاده از تحلیل ریزبافت بهصورت شکل ۳.۲ است. ابتدا با استفاده از الگوریتم تشخیص چهره، مختصات صورت انتخاب شده و مقادیر پیکسلی چهره بهصورت نرمالیزه می‌شود. سپس عملگر LBP با شعاع‌های متفاوت اعمال شده و هیستوگرام آنها محاسبه می‌شود، سپس این هیستوگرام‌ها کنار هم گذاشته می‌شود و با الگوریتم SVM طبقه‌بندی صورت می‌گیرد.

در [۳] بر خلاف روش قبلی تنها از عملگر LBP یکنواخت در پنجره سه در سه بهصورت نرمالیزه شده استفاده شده است و از عملگر LBP با شعاع‌های متنوع [۲] استفاده نشده است. همچنین در [۳] به این نکته پرداخته شده است که باید به ریزبافت در نواحی مختلف صورت توجه داشت و توزیع فراوانی ریزبافتها را نباید صرفاً در کل ناحیه صورت بررسی کرد. در این روش در یک حالت هیستوگرام LBP در کل صورت محاسبه می‌شود؛ در حالت دیگر ناحیه صورت به ۹ ناحیه تقسیم شده و در هر کدام بهصورت جداگانه هیستوگرام LBP محاسبه می‌شود و این هیستوگرام‌ها در کنار هم قرار داده می‌شود. سپس هیستوگرام‌ها بهعنوان یک بردار ویژگی به طبقه‌بندی داده می‌شود. در این روش توزیع هر تصویر با توزیع هیستوگرام تصویر چهره واقعی مقایسه می‌شود.

دو روش گفته شده از LBP بهصورت ایستا استفاده کرده‌اند. یعنی ورودی سیستم تنها یک تصویر از چهره فرد است. از آنجا که اطلاعات بین فریم‌ها یعنی تحلیل یک دنباله ویدیویی، می‌تواند به دقت تشخیص کمک کند، پریریا و همکاران عملگر LBP را در فضای سه‌بعدی گسترش داده‌اند تا از اطلاعات بافت در حوزه مکانی تصویر و حوزه زمانی بین فریم‌های متوالی در تصمیم‌گیری استفاده شود [۳۲].



شکل ۴.۲: روش تحلیل ریزبافت در نواحی مختلف تصویر [۳]

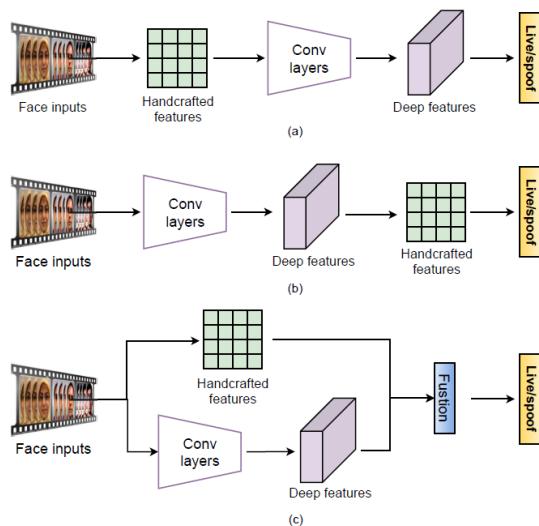
۳.۲ روش‌های مبتنی بر یادگیری عمیق

در عملگر LBP انتخاب ویژگی به صورت دستی انجام می‌گیرد. انگیزه انتخاب ویژگی به صورت هوشمند موجب استفاده از روش‌های یادگیری عمیق برای این کار شده است. ایده استفاده از یادگیری عمیق در حوزه کشف تقلب در تشخیص چهره برای اولین بار توسط پنگ و همکاران مطرح شد [۳۳]. روش ارائه شده در این کار بدین صورت است که ابتدا صورت تشخیص داده می‌شود و پنجره انتخاب شده برای صورت، به‌گونه‌ای در مقیاس‌های مختلف بزرگ می‌شود که شامل پس زمینه صورت نیز باشد. چرا که اطلاعات پس زمینه نیز می‌تواند به کشف تقلب کمک کند. سپس این تصاویر به یک شبکه ALEXNET [۳۴] داده می‌شود و این شبکه کانولوشن ویژگی‌های مدنظر را استخراج می‌کند و در انتهای به‌وسیله SVM طبقه‌بندی صورت می‌گیرد. با اینکه این کار در سال ۲۰۱۴ انجام شده است، اما کاشف به عمل آمده است که استفاده خام از شبکه عصبی عمیق به تنها یک نمی‌تواند به دقت مطلوب برسد. به همین دلیل تاکنون پژوهش‌ها در این حوزه ادامه داشته است و ایده‌های مختلفی برای بهبود عملکرد و افزایش دقت طبقه‌بندی مطرح شده است.

روش گفته شده روی یک فریم کار می‌کند. برای بهره بردن از اطلاعات بین فریم‌های مختلف استفاده از کانولوشن سه بعدی پیشنهاد شده است [۳۵، ۳۶]. شیوه دیگر برای کمک گرفتن اطلاعات فریم‌های متوالی استفاده از ساختار LSTM [۳۷] پس از شبکه کانولوشن است که کارهای [۳۸، ۳۹] از این ساختار استفاده کرده‌اند.

۱.۳.۲ ترکیب روش‌های یادگیری عمیق و ویژگی‌های دستی

یک ایده برای افزایش دقت شبکه عصبی پیشنهاد ترکیب ویژگی‌های لایه‌های کانولوشن با ویژگی‌های دستی^۷ است. نمای کلی حالت‌های مختلفی که می‌توان برای این کار، ساختار ارائه کرد در شکل ۵.۲ نشان داده شده است [۴]. حالت‌های مختلف این روش بدین صورت است که می‌توان ابتدا ویژگی دستی را استخراج کرد و این ویژگی‌ها را به یک شبکه عمیق داد. یا می‌توان ابتدا از شبکه عمیق برای استخراج ویژگی استفاده کرد و سپس روی ویژگی‌های عمیق به‌دست آمده از روش‌های استخراج ویژگی دستی استفاده کرد یا آن‌که ویژگی‌های عمیق و ویژگی‌های دستی را با هم ادغام کرده و سپس به طبقه‌بند داده شود.



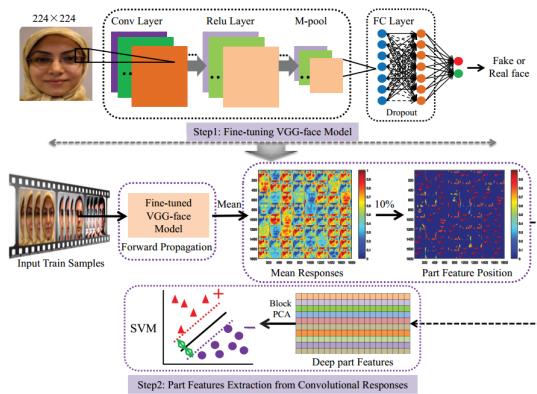
شکل ۵.۲: حالت‌های مختلف ترکیب ویژگی‌های دستی و ویژگی‌های شبکه عمیق [۴]

برای مثال فنگ و همکاران [۵] پیشنهاد داده‌اند که از شبکه‌ی از قبل آموزش داده شده استفاده شود. این ایده که در شکل ۶.۲ نشان داده شده است بدین صورت که از شبکه VGG-face [۴۰] که برای تشخیص چهره، روی حجم زیادی داده آموزش داده شده است، استفاده می‌شود و این شبکه روی داده‌های مربوط به کشف تقلب، تنظیم دقیق^۸ می‌گردد. در مرحله بعد از وزن‌های بهبود یافته استفاده می‌شود و تصاویر نمونه به شبکه داده می‌شود و سپس مقادیر لایه‌های میانی شبکه، به صورت ماتریسی روی هم قرار داده می‌شوند و میانگین گرفته می‌شود سپس مقادیری که مقدار زیادی دارند نگه داشته می‌شوند و بعد آن‌ها با الگوریتم PCA کاهش داده می‌شود. سپس ماتریس کاهش بعد داده

⁷Hand crafted features

⁸Fine tune

شده به یک طبقه بند SVM داده می‌شود و تصمیم‌گیری انجام می‌شود.

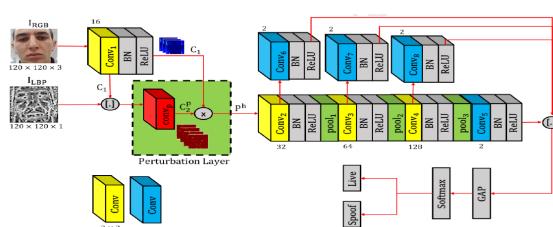


شکل ۶.۲: استفاده از شبکه تنظیم دقیق شده و اعمال PCA روی ویژگی‌های عمیق [۵]

لی و همکاران ابتدا یک شبکه عصبی VGG-face را روی داده‌های مربوط به تشخیص تقلب تنظیم دقیق کرده‌اند و سپس روی کانال‌های مختلف در لایه‌های شبکه، عملگر LBP را اعمال کرده‌اند. با گرفتن هیستوگرام روی آن از SVM برای طبقه بندی استفاده کرده‌اند [۴۱].

رحمان و همکاران روی تصویر ورودی عملگر LBP زده‌اند و با ترکیب ویژگی‌های لایه اول کانولوشن و خروجی LBP را به ادامه شبکه عصبی داده‌اند [۶]. این ایده در شکل ۷.۲ نشان داده شده است.

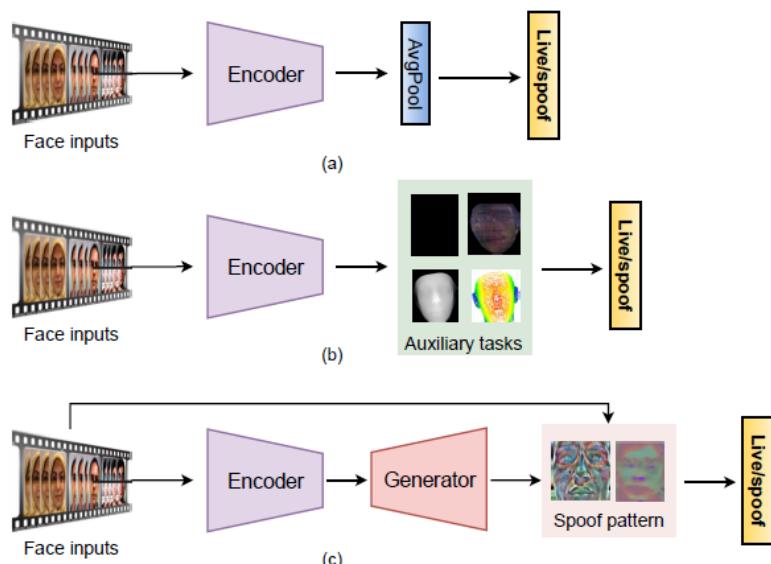
روش‌های ترکیبی بین ویژگی‌های دستی و ویژگی‌های یادگیری عمیق دارای یک قسمت ایستا هستند که حین آموزش شبکه تغییری نخواهند کرد. برای روش‌هایی مبتنی بر شبکه عصبی مطلوب این است که تمامی قسمت‌های شبکه به صورت انتها به انتها یاد گرفته شود.



شکل ۷.۲: روش ترکیب LBP و کانولوشن [۶]

۲.۳.۲ استفاده از تخمين سیگنال کمکی

در روش‌های بیان شده روال آموزش شبکه عصبی بهینه کردنتابع هزینه آنتروپویی متقطع دودویی^۹ است. با این رویکرد که در انتهای شبکه یک نورون برای تصمیم‌گیری وجود دارد و تابع هزینه روی این نورون اعمال می‌شود. مشکل این روش این است که شبکه ممکن است ویژگی‌های غیر مطلوبی را پیدا کند که هر چند در جداسازی داده‌های آموزش مفید است اما ممکن است مشابه این ویژگی‌ها در داده‌های آزمون وجود نداشته باشد. این مشکل با عنوان بیش‌برازش^{۱۰} در علم یادگیری ماشین شناخته می‌شود.



شکل ۸.۲: روش‌های مختلف یادگیری عمیق در حوزه‌ی کشف تقلب چهره [۴]

برای مثال ممکن است شبکه در حین آموزش به قاب صفحه نمایشی که برای حمله استفاده شده است توجه کند، اما در داده‌های آزمون مشابه این قاب وجود نداشته باشد. بدین منظور تلاش محققان برای یافتن ویژگی‌های خوش‌ساخت^{۱۱} به ایده ناظارت کمکی^{۱۲} رسانده است [۸]. در روش‌های ناظارت کمکی سعی می‌شود از تخمين یک مورد کمکی برای استنتاج تقلبی یا واقعی بودن چهره استفاده شود. یکی از موارد مهم کمکی در این حوزه تخمين عمق صورت است.

به طور کلی روش دقیق برای محاسبه عمق، استفاده از دوربین مخصوص است که برای هر پیکسل

⁹Binary cross entropy

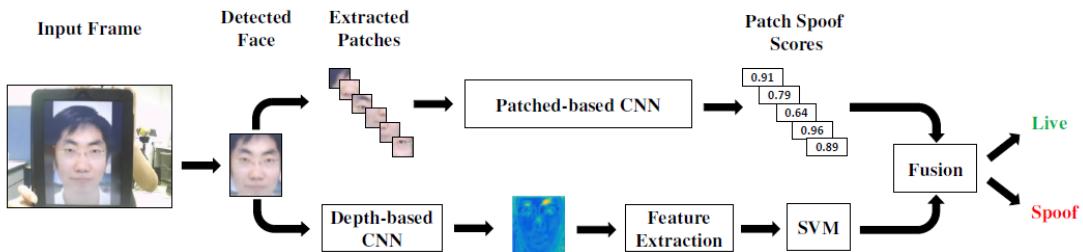
¹⁰Over fitting

¹¹Fine grained features

¹²Auxiliary supervision

مقدار متناظر با عمق آن پیکسل را نیز بدهد. همچنین با استفاده از روش‌های سه‌بعدی‌سازی و استفاده از حداقل دو دوربین، بازسازی مدل سه‌بعدی امکان پذیر است. اما در کشف تقلب در حالت نرم‌افزاری مطلوب این است که این کار به وسیله‌ی تنها یک دوربین ساده انجام شود. لذا در این حالت تنها می‌توان تخمینی از عمق را داشت.

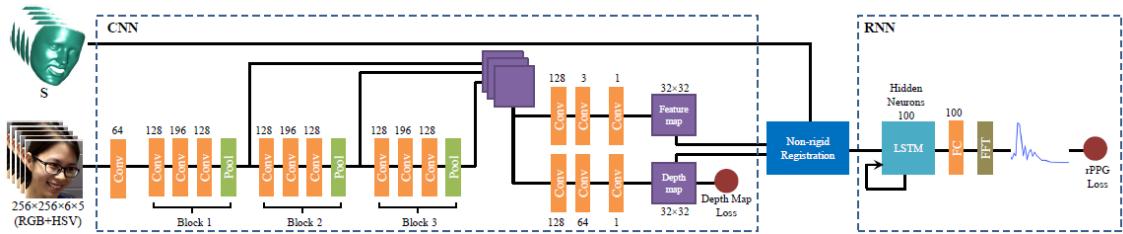
استفاده از عمق از این شهود گرفته شده است که مغز انسان چهره واقعی را دارای عمق می‌بیند، برای مثال بینی نزدیک‌تر از گونه‌ها است، اما چهره تقلبی که روی صفحه نمایش یا کاغذ چاپ شده قرار دارد دارای عمقی مسطح است. در روش‌هایی که از عمق به عنوان یک سیگنال کمکی استفاده کرده‌اند، پیش از آموزش شبکه‌ی کشف تقلب، از یک شبکه تخمین عمق مثل PRNet [۴۲] استفاده می‌شود. و عمق به دست آمده را بین صفر و یک نرمالایز می‌شود. برای تصاویر واقعی این تصویر به عنوان عمق ذخیره شده و برای تصاویر تقلبی، عمق مسطح صفر در نظر گرفته می‌شود. اکنون از این برچسب عمق ایجاد شده برای آموزش ساختار شبکه عصبی توسعه داده شده استفاده می‌شود [۷، ۸، ۹، ۱۵، ۴۳، ۱۰].



شکل ۹.۲: استفاده از عمق برای کشف تقلب در چهره [۷]

اتوم و همکاران [۷] برای اولین بار در این حوزه از عمق به عنوان سیگنال کمکی استفاده کرده‌اند. روش ارائه شده بدین صورت است که ابتدا از تصویر ورودی، صورت تشخیص داده شده و تصویر صورت به دو شبکه داده می‌شود. در مسیر بالایی شکل ۹.۲ قسمت‌های مختلف صورت به صورت تصادفی انتخاب شده و به یک شبکه عصبی کانولوشنی داده می‌شود و در مسیر پایین از طریق یک شبکه عصبی، عمق تصویر تخمین زده می‌شود. سپس اطلاعات دو مسیر با یکدیگر ترکیب شده و در مورد واقعی یا غیرواقعی بودن تصویر تصمیم‌گیری می‌شود.

همچنین لیو و همکاران [۸] علاوه بر استفاده از سیگنال کمکی عمق از تخمین سیگنال rPPG در طول فریم‌های متوالی به عنوان سیگنال حیات چهره بهره برده‌اند. در قسمت عمق مشابه [۷] ابتدا

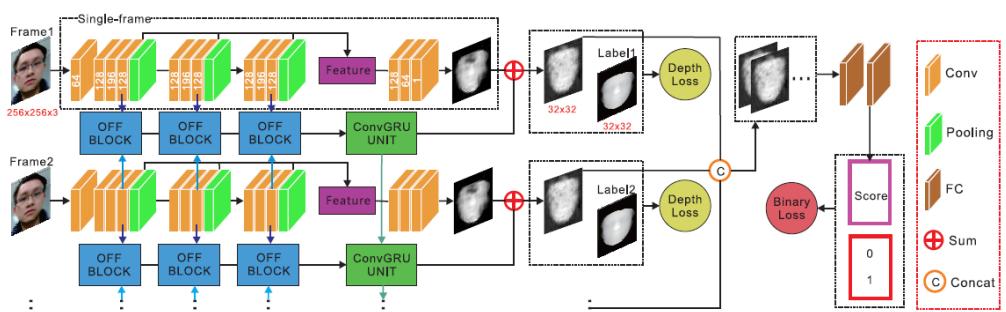


شکل ۱۰.۲: روش استفاده از عمق و تخمین rPPG [۸]

برچسب عمق واقعی برای چهره زنده و عمق صفر برای چهره تقلیبی تخمین زده شده و ازتابع هزینه رابطه ۴.۲ برای بهینه سازی شبکه استفاده می شود. که در آن D_i عمق متناظر با تصویر و Θ مجموعه پارامترهای شبکه است.

$$\Theta_D = \arg \min_{\Theta} \sum_{i=1}^{N_d} \|CNN_D(I_i; \Theta) - D_i\|_1^2 \quad (4.2)$$

همچنین ونگ و همکاران [۹] ساختاری را به کمک الگوریتم جریان نوری ^{۱۳} روی ویژگی های شبکه عصبی برای تخمین عمق توسعه داده اند، به گونه ای که اطلاعات حرکتی بین فریم های متوالی نیز در نظر گرفته می شود. همچنین از ترکیب ساختار GRU [۴۴] با کانولوشن، بلوکی به نام ConvGRU معرفی کرده اند که در آن در رابطه GRU به جای ضرب های ماتریسی از عملگر کانولوشن استفاده شده است و کاربرد آن توجه به ویژگی های بلند مدت در میان فریم های متوالی ورودی است.



شکل ۱۱.۲: استفاده از ویژگی های عمیق در طول زمان [۹]

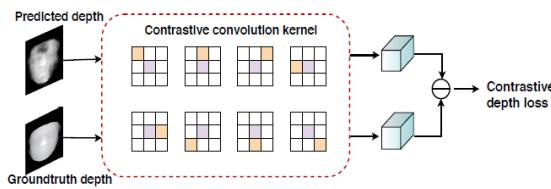
در استفاده از سیگنال کمکی عمق در شبکه نه تنها مقدار عمق می تواند مهم باشد بلکه پیوستگی عمق بین پیکسل های مجاور نیز اهمیت دارد. بدین منظور تابع هزینه CDL برای در نظر گرفتن این

¹³Optical flow

پیوستگی عمق در پیکسل‌های مجاور توسعه داده شده است [۹، ۴۳] در تابع هزینه CDL به جای محاسبه فاصله اقلیدسی عمق تخمینی و برچسب عمق به صورت پیکسل به پیکسل مشابه رابطه ۵.۲ از تفاوت عمق بین پیکسل‌های مجاور نیز استفاده می‌شود.

$$L_{CDL} = \sum_i \|K_i^{CDL} \odot D_P - K_i^{CDL} \odot D_G\| \quad (5.2)$$

که در آن D_P عمق تخمین زده شده توسط شبکه و D_G عمق برچسب واقعی است و K_i^{CDL} هسته‌های کانولوشن دارای ۰ و ۱ هستند که در شکل ۱۲.۲ نشان داده شده است و \odot نشانگر عملگر کانولوشن است. در شکل ۱۲.۲ مربع بنفش متناظر با عدد ۱ و مربع زرد متناظر با عدد ۱ و مربع‌های سفید عدد ۰ را در هسته نشان می‌دهند.



شکل ۱۲.۲: نحوه محاسبه تابع هزینه CDL [۱۰]

یو و همکاران [۱۰] ساختاری تغییر یافته از شبکه‌های کانولوشنی با تأکید بر پیکسل مرکزی پنجره کانولوشن توسعه داده‌اند که در شکل ۱۳.۲ نشان داده شده است. این ساختار با الهام از LBP ایجاد شده است، به گونه‌ای که در هر بار انجام عملگر کانولوشن، پیکسل مرکزی از پیکسل‌های مجاور کم خواهد شد. که رابطه ۶.۲ این عملگر را نشان می‌دهد.

$$y(p_0) = \sum_{p \in R} w(p_n) \cdot (x(p_0 + p_n) - x(p_0)) \quad (6.2)$$

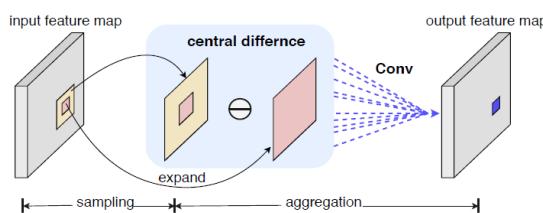
برای آنکه از خاصیت کانولوشن نیز استفاده شود ترکیب خطی رابطه ۶.۲ با رابطه کانولوشن حساب می‌گردد.

$$y(p_0) = \theta \sum_{p \in R} w(p_n) \cdot (x(p_0 + p_n) - x(p_0)) + (1 - \theta) \sum_{p \in R} w(p_n) \cdot x(p_0 + p_n) \quad (7.2)$$

که در آن θ یک هایپر پارامتر است و قسمت اول رابطه ۷.۲ کانولوشن تفاضلی مرکزی و قسمت دوم

کانولوشن کلاسیک است. این رابطه در نهایت به صورت رابطه ۸.۲ ساده می‌گردد.

$$y(p_0) = \sum_{p \in R} w(p_n).x(p_0 + p_n) + \theta(-x(p_0)) \sum_{p \in R} w(p_n) \quad (8.2)$$



شکل ۱۳.۲: عملگر کانولوشن تغییر یافته [۱۰]

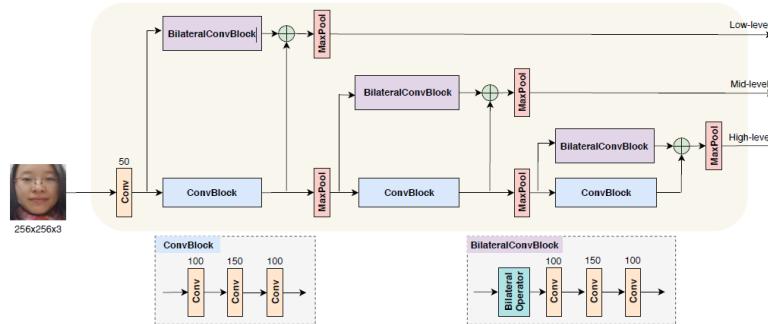
که همانطور که مشاهده می‌شود حاصل نهایی همان کانولوشن کلاسیک خواهد بود که پیکسل مرکزی وزن متفاوتی نسب به کانولوشن کلاسیک خواهد داشت. از این ساختار برای تخمین سیگنال کمکی عمق با نظارت تابع هزینه CDL استفاده می‌شود. همچنین برای یافتن اندازه‌ی شبکه از روش جستجوی معماری شبکه^{۱۴} [۴۵] استفاده شده است.

در جستجوی معماری شبکه برخلاف روش‌های کلاسیک که طراحی معماری شبکه با مهندسی و سعی و خطا انجام می‌شود، تلاش می‌شود معماری بهینه برای کاربرد مورد نظر به صورت خودکار با یادگیری تقویتی و مفاهیم یادگیری ماشین پیدا شود. در حوزه کشف تقلب علاوه بر [۴۵] کارهای [۴۶، ۴۷] متدهایی بر پایه این ابزار برای یافتن شبکه بهینه پیشنهاد داده‌اند.

لی و همکاران به جای تخمین عمق در یک صفحه دو بعدی، از ابر نقاط در فضای سه‌بعدی به عنوان سیگنال کمکی استفاده کرده‌اند و ساختاری به نام 3DPC-NET پیشنهاد کرده‌اند [۴۸].

یو و همکاران [۱۱] مسئله تشخیص تقلب در چهره را یک مسئله تشخیص ماده فرض کرده‌اند. این فرض با توجه به این واقعیت استفاده شده است که جنس پوست صورت با جنس کاغذ چاپ شده و جنس صفحه نمایش متفاوت است. برای تشخیص جنس ماده با الهام از فیلتر bilateral روی ویژگی‌های شبکه عمیق از این فیلتر استفاده کرده‌اند. فیلتر bilateral میانگین وزن دار روی پیکسل‌های مجاور است که با افزایش فاصله تأثیر آن به صورتی تابعی گوسی کاسته می‌شود و روی

¹⁴Network architecture search



شکل ۱۴.۲: روش استفاده از فیلتر bilateral در شبکه عمیق [۱۱]

هر پیکسل به مختصات p و تصویر I به صورت رابطه ۹.۲ تعریف می‌شود.

$$BiBase(I_p) = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(|p - q|) g_{\sigma_r}(|I_p - I_q|) I_q \quad (9.2)$$

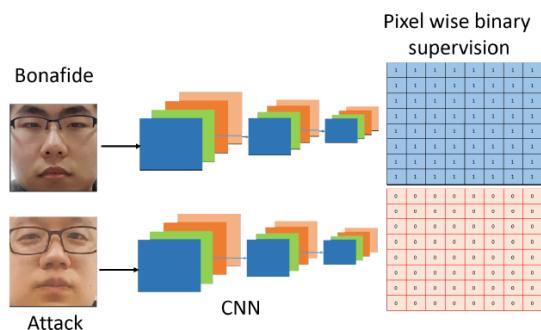
$$\text{with } k = \sum_{q \in I} g_{\sigma_s}(|p - q|) g_{\sigma_r}(|I_p - I_q|)$$

که در آن $g_{\sigma}(x) = \exp(-\frac{x^2}{\sigma^2})$ تابع گوسی است. در این روش ساختار شبکه مشابه [۸] است ولی روی ویژگی‌های کانولوشن این فیلتر اعمال شده است.

با وجود آن‌که سیگنال کمکی عمق در ادبیات موضوع به‌طور گستردگی استفاده شده است اما پر هزینه است و نیاز به پردازش بیشتر برای تخمین عمق دارد. جدای از آن‌که عمق، یک سیگنال کامل برای تشخیص تقلب نیست و فرض مسطح در نظر گرفتن عمق در چهره‌های تقلبی، فرض همیشه برقرار نیست. برای مثال فرض کنید مهاجم ابزار حمله مثل صفحه نمایش یا کاغذ چاپ شده را به صورت مایل قرار دهد در این صورت عمق به صورت یکنواخت در همه‌جا صفر نخواهد بود.

جرج و مارسل روشی را برای پیدا کردن ویژگی‌های خوش‌ساخت بدون استفاده از عمق پیشنهاد کرده‌اند [۱۲]. در این روش از چند لایه اول شبکه DENSNET [۴۹] برای نشان‌کردن ^{۱۵} تصویر ورودی به یک صفحه ۱۴*۱۴ استفاده کرده‌اند. و قرارداد کرده‌اند که برچسب واقعی به جای یک عدد صفر و یک، یک ماتریس دو بعدی به طول کامل صفر یا یک است و تابع هزینه آنتروپی متقاطع دودویی را به جای یک نورون روی یک صفحه دو بعدی در نظر گرفته‌اند. با این روش دیگر نیازی به تخمین عمق نخواهد بود.

^{۱۵}Map



شکل ۱۵.۲: تابع هزینه BCE روی یک صفحه مسطح به جای یک نورون [۱۲]

۳.۴.۲ استفاده از شبکه‌های مولد تخاصمی وتابع هزینه‌های مختلف

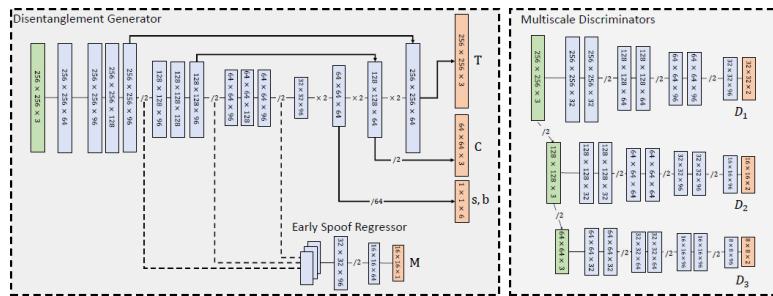
مسئله کشف تقلب در چهره بیشتر شبیه مسئله یافتن یک نویز خاص در تصویر است. ابزارهای حمله نظیر کاغذ چاپ شده و صفحه نمایش گر، بافت و تفکیک پذیری متفاوتی با بافت صورت انسان دارند. که این تفاوت جنس را می‌توان با یک نویز جمع شونده با تصویر چهره زنده مدل کرد. جورابلو و همکاران [۵۰] برای اولین بار در مسئله کشف تقلب چهره از شبکه‌های مولد تخصصی (GAN) [۵۱] برای مدل کردن و یافتن نویز تصاویر تقلیبی استفاده کردند. با تخمین نویز مربوط به کشف تقلب، قدرت استنتاج برای تقلیبی بودن تصویر بیشتر خواهد شد.

از آنجا که نویز مربوط به تقلب می‌تواند در مقیاس‌های مختلف تصویر وجود داشته باشد لیو و همکاران [۱۳] ساختاری بر پایه GAN که الگوهای تقلب در ابعاد مختلف تصویر را تخمین بزند پیشنهاد داده‌اند. این روش که در شکل ۱۶.۲ به تصویر کشیده شده است، در شبکه مولد-disman-element generator ابعاد تصویر در لایه‌های اول کاهش یافته و سپس افزایش می‌یابد و از ویژگی‌های خروجی لایه‌ها با ابعاد مختلف به عنوان ویژگی‌های تقلب تولید شده استفاده می‌شود. در نهایت شبکه multiscale discriminator این ویژگی‌های تقلب در سطوح مختلف را به عنوان ورودی دریافت می‌کند و طی یک بازی رقابتی بین دو شبکه مولد و تفکیک‌دهنده، در نهایت ویژگی‌های تقلب بهتری تولید خواهد شد.

با وجود اینکه در دو پژوهش اخیر ذکر شده [۱۳، ۵۰] از شبکه مولد تخصصی برای بهبود دقیق در تست درون دیتاست استفاده شده است، توجه پژوهشگران به استفاده از GAN برای تعمیم‌پذیری مدل در دیتاست‌های مختلف جلب شده است [۱۵، ۱۶].

تعیین پذیری مدل در دیتاستهای مختلف بدین معناست که برای مثال از بین چهار دیتاست

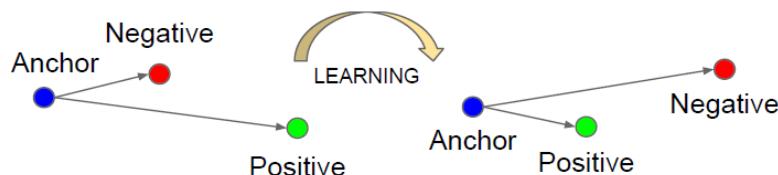
¹⁶Generative adversarial networks



شکل ۱۶.۲: ساختار بر پایه استفاده از شبکه مولد برای تخمین علائم تقلب در سطوح مختلف [۱۳]

مختلف، سه دیتاست برای آموزش شبکه استفاده می‌گردد و مدل آموزش داده شده روی دیتاست چهارم آزمایش می‌شود. از آنجا که دیتاست‌های مختلف توزیع‌های آماری متفاوتی دارند، رسیدن به دقیق خوب در آزمون روی دیتاست دیده نشده (که توزیع لزوماً یکسانی با توزیع دیتاست‌هایی که برای آموزش استفاده شده است ندارد) یک چالش جدی در این حوزه است. از آنجا که دیتاست‌های مربوط به کشف تقلب چهره، با تعداد اشخاص و شرایط تصویربرداری محدود، جمع‌آوری شده‌اند، فراوانی و مشخصه‌های آماری این دیتاست‌ها با نمونه‌های دنیای واقعی یکسان نیست. از این‌رو مدل آموزش دیده شده ممکن است در هنگام استفاده در عمل به دقیق آموزش نرسد.

همچنین یک روش برای بهبود قابلیت تعمیم‌پذیری، استفاده از تابع هزینه سه‌گانه^{۱۷} [۱۴] است. در تابع هزینه سه‌گانه هدف این است که استخراج ویژگی به نحوی انجام شود که فاصله‌ی نمونه‌های مربوط به یک کلاس کوچک و فاصله‌ی بین نمونه‌های مربوط به کلاس‌های مختلف زیاد شود.



شکل ۱۷.۲: نحوه عملکرد تابع هزینه سه‌گانه روی فاصله‌ی بردارهای ویژگی [۱۴]

فرض کنید خروجی شبکه استخراج ویژگی بردار $f(x) \in R^d$ باشد. برای تشکیل تابع هزینه سه‌گانه لازم است که از خروجی‌های شبکه استخراج ویژگی، یک بردار ویژگی لنگر^{۱۸} $f(x^a)$ ، یک بردار ویژگی با برچسب یکسان با لنگر^{۱۹} $f(x^p)$ و یک بردار ویژگی با برچسب متفاوت با لنگر^{۲۰}

¹⁷Triplet loss

¹⁸Anchor

¹⁹Positive

²⁰Negative

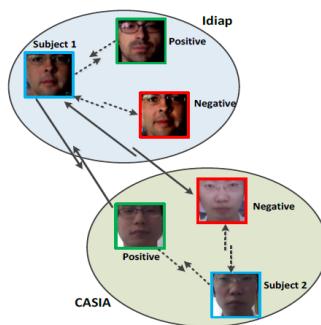
$f(x^n)$ انتخاب شود. در این صورت تابع هزینه سه‌گانه به صورت رابطه ۱۰.۲ تعریف می‌شود. که در آن α یک حاشیه از قبل تعریف شده است. تمام سه‌گانه‌هایی که فاصله درون کلاسی آن‌ها از فاصله برون کلاسی بیشتر از مقدار است درون مجموع گیری قرار می‌گیرد. که در آن یک حاشیه از قبل تعریف شده است. تمام سه‌گانه‌هایی که فاصله درون کلاسی آن‌ها از فاصله برون کلاسی بیشتر از مقدار α است درون مجموع گیری قرار می‌گیرد.

$$L_{trpi} = \sum_i [||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + \alpha]_+ \quad (10.2)$$

تابع هزینه سه‌گانه به صورت رابطه ۱۱.۲ نیز قابل بیان است. که در آن زمانی که فاصله درون کلاسی کوچک‌تر از فاصله برون کلاسی به میزان سطح آستانه باشد حاصل \max صفر خواهد بود و در محاسبات تابع هزینه نقش نخواهد داشت.

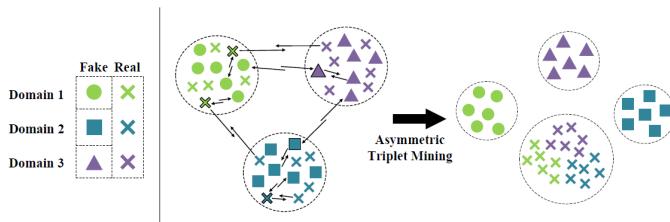
$$L_{trpi} = \sum_i \max(0, ||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + \alpha) \quad (11.2)$$

شائو و همکاران [۱۵] از ساختار GAN و ابزار کمکی تخمین عمق و تابع هزینه سه‌گانه برای بهبود تعمیم‌پذیری استفاده کردند. در این کار یک تابع هزینه بر مبنای تابع هزینه سه‌گانه توسعه داده شده است که فاصله بین نمونه‌ها با برچسب یکسان در دیتاست‌های مختلف را کوچک‌تر کند و فاصله نمونه‌ها با برچسب متفاوت در یک دیتاست را بیشتر کند. با این کار توزیع نمونه‌ها در دیتاست‌های مختلف با یکدیگر مترافق‌تر خواهد شد. در شکل ۱۸.۲ به کارگیری این تابع هزینه را در بین دو دیتاست مختلف نشان می‌دهد.



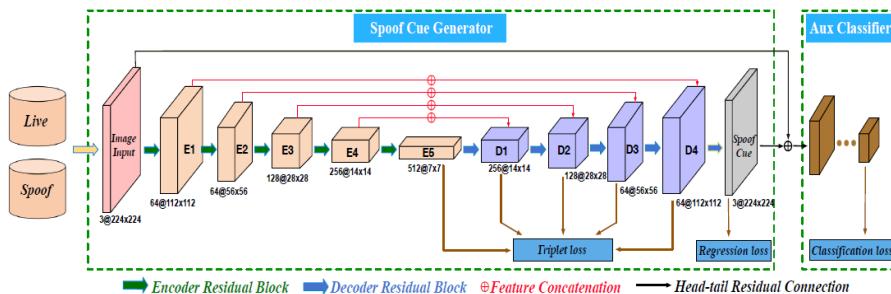
شکل ۱۸.۲: نحوه اثر تابع هزینه روی فاصله نمونه‌ها در دیتاست‌های مختلف [۱۵]

همچنین جیا و همکاران [۱۶] علاوه بر استفاده از GAN صورتی نامتقارن ازتابع هزینه سه‌گانه را پیشنهاد کرده‌اند. به‌گونه‌ای که نمونه‌های زنده در دیتاست‌های مختلف به یکدیگر نزدیک‌تر شوند و نمونه‌های تقلبی در دیتاست‌های مختلف از یکدیگر دورتر شده و نمونه‌های واقعی از نمونه‌های تقلبی با فاصله باشند.



شکل ۱۹.۲: تابع هزینه نامتقارن برای کاهش فاصله نمونه‌های از یک کلاس [۱۶]

فنگ و همکاران [۱۷] یک ساختار U-Net [۵۲] به کار بردند و در میان لایه‌های آخر شبکه مولد الگوهای تقلب از تابع هزینه سه‌گانه استفاده کردند و خروجی این شبکه U-Net را به یک شبکه طبقه‌بند کمکی دادند.



شکل ۲۰.۲: ساختار U-net و تابع هزینه سه‌گانه [۱۷]

پرزکابو و همکاران [۵۳] تابع هزینه سه‌گانه را در فضای نمایی به کار بردند که در رابطه ۱۲.۲ نشان داده شده‌است. که در آن $D_{a,p}$ فاصله درون‌کلاسی و $D_{a,n}$ فاصله برون‌کلاسی است و σ یک هایپر پارامتر است.

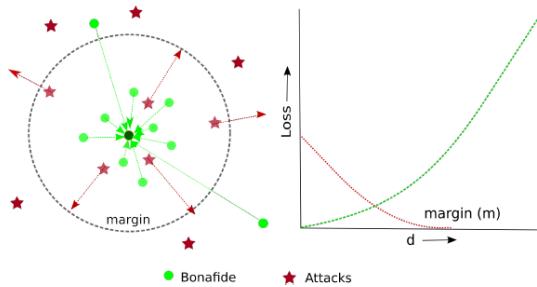
$$L_{tf} = \sum_i \max(0, e^{\frac{D_{a,p}}{\sigma}} - e^{\frac{D_{a,n}}{\sigma}} + \alpha) \quad (12.2)$$

جرج و مارسل [۱۸] تابع هزینه‌ای معرفی کردند که در فضای n بعدی بردارهای ویژگی، نمونه‌های زنده نزدیک به یک مرکز قرار بگیرند و نمونه‌های تقلبی با یک حاشیه از این مرکز فاصله

داشته باشند. مرکز نمونه‌های واقعی در حین آموزش شبکه به روزرسانی می‌شود. فرض کنید مرکز نمونه‌های زنده با C_{BF} نشان داده شود و فاصله بردار ویژگی نمونه i با مرکز با DC_W تعریف شود. در این صورتتابع هزینه تعریف شده به صورت رابطه ۱۳.۲ است.

$$L_{OCCL} = Y \frac{1}{2} DC_W^2 + (1 - Y) \frac{1}{2} \max(0, m - DC_W)^2 \quad (13.2)$$

که در آن Y برچسب واقعی داده است که برابر با یک است اگر نمونه واقعی باشد و صفر است اگر نمونه تقلبی باشد و m یک حاشیه از قبل تعریف شده است.



شکل ۲۱.۲: کاهش فاصله نمونه‌های واقعی تا مرکز و افزایش فاصله نمونه‌های تقلبی تا مرکز [۱۸]

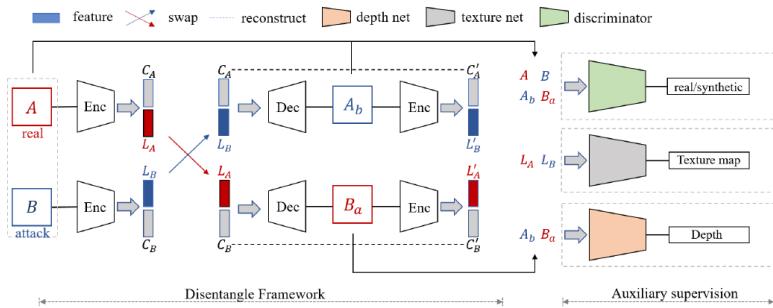
تو و همکاران [۵۴] نیز شبکه VGG-face را به صورت همزمان با دو هدف شناسایی چهره و تشخیص تقلب آموزش داده‌اند و یک تابع هزینه معرفی کرده‌اند که هدف آن منظم‌سازی^{۲۱} و جلوگیری از بیش برآذش شبکه است. در این تابع هزینه فاصله بین هر دو جفت نمونه داده‌ها مستقل از آن که برچسب آن چه باشد کاهش داده می‌شود. تابع هزینه معرفی شده برای این هدف در رابطه ۱۴.۲ بیان شده است. که در آن تابع $(\cdot)\Phi$ نشان دهنده رابطه بین ورودی تصویر و لایه یکی به آخر شبکه است و M تعداد تمام جفت نمونه‌های موجود در دسته آموزش است.

$$L_{tpc} = \sum_{i \neq j}^M \|\Phi(x_i) - \Phi(x_j)\| \quad (14.2)$$

زنگ و همکاران [۱۹] علاوه بر تخمین عمق از تخمین LBP به عنوان سیگنال کمکی استفاده کرده‌اند که در کنار عمق ساختار LBP تصویر ورودی نیز تخمین زده شود. بدین ترتیب که برای تصاویر تقلبی خروجی LBP شبکه باید صفر باشد و برای تصاویر تصاویر واقعی خروجی قسمت

²¹Regularization

باید معادل LBP تصویر ورودی باشد. این شبکه دارای یک شبکه مولد با ساختار U-net و سه شبکه طبقه‌بند برای عمق و LBP و شبکه طبقه‌بندی بر اساس GAN برای تصویر واقعی و ساختگی است.



شکل ۲۲.۲: استفاده از LBP در کنار عمق برای یافتن ویژگی‌های خوش ساخت [۱۹]

ژو و همکاران [۵۵] روی ثبات فضای ویژگی در بین فریم‌های متوالی یک ویدئو تأکید کردند. در این کار به جای استفاده از الگوریتم‌های تشخیص چهره در هر فریم از الگوریتم دنبال‌کننده‌ی چهره استفاده کردند و چهره‌های تخمین زده شده در فریم‌های متوالی را به شبکه تشخیص تقلب دادند. برای این شبکهتابع هزینه‌ای ارائه معرفی کردند که فاصله بین بردارهای ویژگی یک ویدئو در دیتابست را کوچک‌تر کند.

$$L_t = \frac{1}{m} \sum_{i=0}^m \max_{i,j \in v} \|x_i - x_j\|^2 \quad (15.2)$$

که در آن m اندازه دسته آموزش است و x_i, x_j بردارهای فضای ویژگی برای یک ویدئو است. همچنین برای ثبات بردارهای ویژگی در ویدیوهای مختلف، تابع هزینه‌ی دیگری پیشنهاد کردند که فاصله بین بردارهای ویژگی متعلق به یک برچسب واقعی را نیز کوچک‌تر کند.

$$L_t = \frac{1}{m} \sum_{i=0}^m \max_{i,j \in v} y_{ij} \|x_i - x_j\|^2 \quad (16.2)$$

که در آن y_{ij} زمانی که دو بردار ویژگی متعلق به یک کلاس باشند برابر با صفر خواهد بود و در غیر این صورت صفر است.

۴.۲ دیتاست‌های مورد استفاده

مانند بسیاری از مسائل بینایی ماشین، دیتاست نقش حیاتی در توسعه الگوریتم و سنجش میزان دقیقت الگوریتم ایفا می‌کند. از آنجا که تمرکز این پایان‌نامه روی حملات کاغذ چاپ‌شده و بازپخش صفحه نمایش است، به معرفی برخی از دیتاست‌هایی که حاوی این نوع حملات هستند پرداخته می‌شود. حملات نظیر استفاده از ماسک، معمولاً براحتی قابل اجرا نیستند و هزینه‌بر هستند اما دو حمله گفته شده از نظر قابلیت اجرا ساده‌تر، کم هزینه و متداول‌تر هستند.

۱.۴.۲ Replay دیتاست

دیتاست Replay [۳] شامل ویدیوهای از ۵۰ شخص مختلف با نمونه‌های واقعی و تقلبی است. نمونه‌های واقعی در شرایط محیطی نوری کنترل شده با پس‌زمینه یکنواخت و شرایط محیطی با نور کنترل نشده با پس‌زمینه غیر یکنواخت گرفته شده‌اند. برای نمونه‌های تقلبی از صفحه کاغذ چاپ‌شده، استفاده از تلفن همراه برای بازپخش ویدئو و استفاده از تبلت iPad برای پخش ویدئو با کیفیت بالا استفاده شده است. همچنین نمونه‌های تقلبی در دو حالت استفاده از یک پایه ثابت به‌منظور ثابت ماندن ابزار حمله و استفاده از دست که کمی لغزش خواهد داشت گرفته شده‌اند. رزولوشن تمامی نمونه‌ها واقعی و تقلبی با فرمت QVGA یعنی 23.2×32.0 پیکسل است. در ۲۳.۲ نمونه‌هایی از این



شکل ۲۳.۲: نمونه‌هایی از دیتاست Replay [۳]

دیتاست نشان داده شده است. در سطر بالایی نمونه‌ها در محیط کنترل شده از نظر نورپردازی و پس‌زمینه یکنواخت هستند، در حالی که تصاویر سطر پایینی نمونه‌ها دارای نورپردازی غیر کنترل شده

و پس زمینه غیر یکنواخت هستند. تصاویر از سمت چپ به ترتیب تصاویر واقعی، استفاده از کاغذ چاپ شده، استفاده از تلفن همراه برای بازپخش و استفاده از تبلت برای بازپخش ویدئو هستند.

۲.۴.۲ دیتاست CASIA

در دیتاست CASIA [۲۰] نیز از ۵۰ شخص مختلف نمونه های واقعی و تقلبی گرفته شده است. همچنین تصویربرداری با سه نوع دوربین مختلف برای پوشش دادن حالت های مختلف در روزولوشن های مختلف انجام شده است. در این دیتاست حملات نوع کاغذ چاپ شده روی کاغذ گلاسه صورت گرفته است که کیفیت بالاتری نسبت به کاغذ معمولی دارد. همچنین برای حمله بازپخش از تبلت استفاده شده است. در نمونه های واقعی دیتاست از کاربر خواسته شده است که پلک و لب بزند تا ویدیوهای ضبط شده دارای اطلاعات حرکتی صورت باشند. در نمونه حمله های تقلبی قسمت چشم های صورت بریده شده است تا کاربر با پلک زدن بتواند در نمونه های تقلبی اطلاعات حرکت به ویدئو بدهد. همچنین در نمونه هایی که کاغذ بریده نشده است از کاربر خواسته شده که با حرکت دست کاغذ چاپ شده را حرکت بدهد. نمونه هایی از دیتاست CASIA در شکل ۲۴.۲ نشان داده شده است.



[۲۰] نمونه هایی از دیتاست CASIA

۳.۴.۲ دیتاست MSU

در دیتاست MSU [۲۱] از ۵۵ شخص تصویربرداری شده است که ویدیوهای متعلق به ۳۵ فرد در دسترس قرار داده شده است. برای تصویربرداری از دوربین لپ تاپ و دوربین تلفن همراه استفاده شده است که دارای روزولوشن ۷۲۰*۴۸۰ و ۶۴۰*۴۸۰ هستند. استفاده از این نوع دوربین ها به منظور شبیه سازی سناریو احراز هویت از طریق تلفن همراه و لپ تاپ انجام شده است. برای حمله بازپخش از صفحه نمایش تبلت و تلفن همراه استفاده شده است. همچنین برای حمله کاغذ چاپ شده از پرینتر با کیفیت استفاده شده است.



شکل ۲۵.۲: نمونه‌هایی از دیتابست MSU [۲۱]

۴.۴.۲ دیتابست OULU

در دیتابست OULU [۱] از ۵۵ شخص مختلف برای تصویر برداری نمونه‌های واقعی و تقلیبی استفاده شده است. تصویربرداری در سه نشست مختلف، با شش تلفن همراه جدید در زمان جمع‌آوری دیتابست استفاده شده است که باعث تنوع در کیفیت تصویر و محیط پس‌زمینه شده است. برای حمله کاغذ چاپ شده از دو نوع چاپگر با کیفیت و برای حمله بازپخش از یک نمایشگر و صفحه نمایش لپ‌تاپ استفاده است. ویدیوهای ضبط شده با کیفیت Full HD با رزولوشن ۱۹۲۰*۱۰۸۰ گرفته شده‌است.

این دیتابست در مقایسه با دیتابست‌های قبلی تنوع بیشتر و کیفیت بالاتری دارد که باعث چالشی شدن دیتابست شده‌است. نمونه‌های واقعی در دیتابست OULU در شکل ۲۶.۲ نشان داده شده است و نمونه‌های حمله در شکل ۲۷.۲ نشان داده شده است که دو نمونه‌ی سمت چپ دو نوع حمله کاغذ چاپ شده و نمونه‌های سمت راست دو نوع حمله بازپخش را نشان می‌دهد.



شکل ۲۶.۲: نمونه‌های واقعی در دیتابست OULU [۱]



شکل ۲۷.۲: نمونه‌های تقلبی در دیتابست OULU [۱]

همچنین در این دیتابست برای گزارش دقیق پروتکل مختلف به منظور ارزیابی قابلیت تعمیم‌پذیری مدل ارائه شده پیشنهاد شده است. پروتکل اول تنوع نشست را در دقیق مدل بررسی می‌کند، بدین صورت که مدل باید روی داده‌های دو نشست از سه نشست آموزش ببیند و ارزیابی روی نشست سوم خواهد بود. در پروتکل دوم روی یک نوع از حمله کاغذ چاپ شده و یک نوع حمله بازپخش آموزش انجام می‌شود و در ارزیابی از نوع دیگر حمله بازپخش و کاغذ چاپ شده استفاده می‌شود تا تعمیم‌پذیری مدل روی حمله از ابزار دیده نشده بررسی شود. در پروتکل سوم روی ۵ نوع دوربین تلفن همراه از شش نوع آموزش انجام می‌شود و روی نوع ششم ارزیابی صورت می‌گیرد که این حالت برای بررسی قابلیت تعمیم‌پذیری روی نوع سنسور تصویر برداری دیده نشده انجام می‌گردد. در پروتکل چهارم هر سه نوع پروتکل قبلی در هم ادغام می‌شوند تا اثر تنوع نشست، تنوع دوربین و تنوع نوع حمله ملاحظه شود.

۵.۴.۲ دیتابست SIW

در دیتابست SIW [۸] از ۱۶۵ شخص مختلف برای تصویربرداری استفاده شده است. برای تصویربرداری از دو نوع دوربین با کیفیت استفاده شده است. در نمونه‌های واقعی تصویربرداری با فواصل مختلف دوربین تا کاربر انجام شده است تا تنوع فاصله کاربر با دوربین را پوشش دهد. همچنین از کاربر خواسته شده است که حالات مختلف چهره را به خود بگیرد و صورت خود را حرکت بدهد. این موجب تنوع در زاویه چهره نسب به دوربین و تنوع حالات چهره شده است.



شکل ۲۸.۲: نمونه‌های از دیتاست SIW [۸]

همچنین شرایط نورپردازی مختلف در این دیتاست دیده شده است. از دو نوع چاپگر برای حمله کاغذ چاپ شده استفاده شده است و از کاربر خواسته شده است که در دو حالت کاغذ را ثابت نگه دارد و آن را حرکت بدهد. همچنین از تبلت و دو نوع گوشی و صفحه نمایش گر رایانه برای حمله بازپخش استفاده شده است. نمونه‌های این دیتاست در شکل ۲۸.۲ قابل مشاهده است.

این دیتاست دارای سه نوع پروتکل مختلف برای ارزیابی است. در پروتکل اول تنها از ۶۰ فریم اول هر ویدئو برای آموزش استفاده می‌شود و از تمامی فریم‌های ویدیوهای تست برای ارزیابی استفاده می‌شود. از آنجا که در فریم‌های ابتدایی ویدئو کاربر صورت خود را حرکت نمی‌دهد این پروتکل به ارزیابی تغییر حالت چهره می‌پردازد. در پروتکل دوم از سه نوع حمله بازپخش استفاده می‌شود و روی حمله چهارم بازپخش ارزیابی می‌شود تا اثر تنوع ابزار حمله در بازپخش بررسی شود. در پروتکل سوم از یکی از انواع حمله بازپخش یا چاپ برای آموزش استفاده می‌شود و از نوع حمله دیگر برای تست استفاده می‌شود که هدف آن ارزیابی نوع حمله دیده نشده است.

فصل ۳

روش پیشنهادی

۱.۳ مقدمه

در این فصل به توضیح مبانی نظری روش پیشنهادی پرداخته می‌شود. روش پیشنهادی شامل یک عملگر قابل آموزش با فرمول بندی شبیه LBP و قرار دادن این عملگر در لایه اول شبکه کانولوشن کلاسیک است. از آنجا که در مسئله کشف تقلب به جای تمرکز روی ویژگی‌های ظاهری نظیر گوشها، لبه‌ها و... اطلاعات بافت تصاویر اهمیت دارد این لایه مبتنی بر LBP پیشنهاد شده است. ابتدا عملگر LBP قابل آموزش بیان خواهد شد. سپس ساختار شبکه تشریح خواهد شد و در ادامه به توضیح تابع هزینه معرفی شده پرداخته می‌شود.

برای بیان عملگر LBP قابل آموزش ابتدا توضیحی کلی از عملگر کانولوشن و شبکه‌های کانولوشنی همراه با شهود استفاده از این شبکه‌ها در مسائل بینایی ماشین، بیان خواهد شد. سپس رابطه ریاضی عملگر کانولوشن و عملگر LBP ارائه شده و با همانندسازی این دو عملگر، عملگر LBP قابل آموزش به دست خواهد آمد. در ادامه برای بهینه کردن شبکه با هدف بهبود دقیق و افزایش قابلیت تعمیم‌پذیری، دو تابع هزینه معرفی خواهد شد. در تابع هزینه اول هدف تفکیک کردن دو کلاس با حاشیه است و در تابع هزینه دوم هدف مجبور کردن شبکه به توجه به ویژگی‌های تقلب به جای توجه به ویژگی‌های ظاهری افراد است.

۲.۳ مروری بر عملگر کانولوشن

یکی از اجزای اصلی شبکه‌های مبتنی بر یادگیری عمیق عملگر کانولوشن است. این عملگر دارای یک هسته‌ی ضرایب است که به صورت پیچشی در تصویر ورودی ضرب می‌شود و سپس با لغزش بر کل تصویر ورودی، یک تصویر خروجی به دست خواهد آمد.

اعمال عملگر کانولوشن روی سیگنال معادل ضرب تبدیل فوریه عملگر در تبدیل فوریه تصویر ورودی است و با این ضرب می‌توان برخی از فرکانس‌های تصویر ورودی را تقویت یا تضعیف کرد که فیلتر کردن تصویر ورودی خواهد بود. اعمال وزن‌های مختلف به هسته فیلتر می‌تواند خروجی تصویر با مشخصات خاصی را بدهد.

برای مثال با استفاده از وزن‌های خاص در عملگر می‌توان یک فیلتر پایین‌گذر طراحی کرد و با اعمال عملگر کانولوشن این فیلتر پایین‌گذر، یک تصویر که فرکانس‌های بالای آن حذف شده‌اند به دست آورد. طراحی فیلترهای مختلف برای اهداف گوناگونی نظیر یافتن لبه در تصویر یا حذف نویز تصویر می‌تواند کاربرد داشته باشد. اما هر هدف نیازمند یک فیلتر خاص از قبل طراحی شده است.

ایده شبکه‌های عصبی کانولوشنی^۱ (CNN) در این است که وزن‌های فیلتر به صورت پارامتر در نظر گرفته شود و در طی فرآیند بهینه‌سازی تابع هزینه، ضرایب فیلتر به روزرسانی شوند و فیلترهای مد نظر از طریق داده‌های موجود به دست آیند. هر اعمال یک لایه کانولوشن روی تصویر باعث به دست آوردن ویژگی‌ها جدید می‌شود و با اعمال متوالی عملگر پارامتری شده کانولوشن، ویژگی‌های مفهومی‌تر به دست خواهد آمد. این ساختار لایه‌ای در صورتی که با تعداد کافی داده، بهینه شود می‌تواند ویژگی‌های معنایی از تصویر را استخراج کند. که این دریافت معنا از تصویر باعث کاربردهای مختلفی نظیر طبقه‌بندی، تشخیص شیء، شناسایی چهره و... شده است.

در مسئله تشخیص تقلب در تشخیص چهره، بیش از آنکه ویژگی‌های معنایی تصویر مد نظر باشد یافتن ویژگی‌هایی در تصویر که شاخصی برای واقعی یا تقلبی بودن چهره است اهمیت دارد. در واقع هدف این است که شبکه‌ای طراحی شود که نشانه‌های تقلب در تصویر را پیدا کند. یکی از ویژگی‌های نشانه‌های تقلب در تصویر وجود در مقیاس ریز تصویر است به‌گونه‌ای که در نگاه اول تشخیص آن دشوار به نظر می‌آید. یکی دیگر از ویژگی‌های نشانه‌ی تقلب در تصویر وجود آن در بیشتر بخش‌های چهره است. بدین منظور در گام اول عملگری ارائه شده است که هدف آن تحلیل بافت تصویر و کمک گرفتن از ایده‌ی شبکه عصبی بهمنظور یافتن بهترین عملگر با توجه به داده‌های ورودی شبکه است.

¹Convolutional neural network

۳.۳ عملگر LBP قابل آموزش

در این بخش فرمول بندی عملگر LBP قابل آموزش بیان می‌شود. بدین منظور ابتدا رابطه عملگر کانولوشن بیان شده و سپس رابطه ریاضی LBP کلاسیک بیان خواهد شد. سپس با مقایسه رابطه کانولوشن و رابطه LBP کلاسیک، این رابطه به نحوی تغییر داده می‌شود که همانند کانولوشن دارای پارامتر قابل یادگیری باشد اما بر خلاف کانولوشن به جای فیلتر کردن تصویر یک خروجی متناظر با بافت تصویر مطابق فرمول LBP به صورت پارامتری شده بدهد.

رابطه عملگر کانولوشن و تصویر ورودی به صورت رابطه 1.3 است. که در آن I_p مقدار شدت روشنایی تصویر در پیکسل p در یک همسایگی یا پنچره به ابعاد فیلتر است. و W_p مقدار وزن متناظر فیلتر در مختصات p پنجره عملگر است. و تابع $(\cdot)\sigma$ نیز یک تابع غیر خطی است.

$$CNN = \sigma \left(\sum_{p \in N} I_p W_p \right) \quad (1.3)$$

و همچین رابطه عملگر ریزبافت LBP به صورت رابطه 2.3 است. که در آن I_c مقدار روشنایی پیکسل در مرکز پنجره عملگر است. در واقع در این عملگر در یک همسایگی، مقدار هر پیکسل از پیکسل مرکزی کسر می‌گردد و بر اساس خروجی بزرگ‌تر یا کوچک‌تر بودن از صفر، یک وزن 2^p پیدا می‌کند. این وزن به صورت ایستا و طبق تعریف قراردادی عملگر مشخص می‌گردد.

$$LBP = \sum_{p \in N} \sigma(I_p - I_c) 2^p \quad (2.3)$$

$$\sigma(x) = \begin{cases} 1 & x \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

به منظور آن که از ایده یافتن وزن‌های بهینه از طریق داده در این عملگر استفاده شود لازم است که تعریف این عملگر به جای تعریف ایستان به تعریف پارامتری شده برسد. بدین منظور وزن 2^p

به صورت رابطه ۳.۳ تغییر داده می‌شود.

$$2^p = e^{p \ln 2} = e^{w_p} \quad (3.3)$$

که در آن W_p یک پارامتر است که حین بهینه‌سازی تغییر می‌کند تا به بهترین مقدار مناسب برای طبقه‌بندی برسد. با جایگذاری این پارامتر در رابطه LBP کلاسیک، عملگر LBP قابل آموزش به دست خواهد آمد به صورت رابطه ۴.۳ به دست خواهد آمد.

$$LBP_{tr} = \sum_{p \in N} \sigma(I_p - I_c) e^{W_p} \quad (4.3)$$

این عملگر در مقایسه با عملگر کانولوشن نگاه ریزتری به بافت تصویر خواهد داشت. از آن جا که در کانولوشن تمامی پیکسل‌های همسایگی در وزن‌های فیلتر ضرب شده و حاصل جمع آنها در تابع غیر خطی قرار می‌گیرند، در عملگر کانولوشن تمامی پیکسل‌های همسایگی تأثیری به اندازه وزن متناظر خود در خروجی دارند. اما در عملگر LBP از آن جا که تابع غیر خطی بین تفاضل هر پیکسل با پیکسل مرکزی اعمال می‌شود نگاه جزئی‌تری به تصویر خواهد داشت و باعث استخراج ویژگی‌های بافتی تصویر خواهد شد.

تابع σ یک تابع غیر خطی است که نقشی مشابه تابع فعالسازی ^۲ در شبکه‌های عصبی را بازی می‌کند. وظیفه این تابع ایجاد روابط غیرخطی برای عملگر است و تفاوت مهم عملگر LBP قابل آموزش با کانولوشن در اعمال تابع غیرخطی درون عملگر مجموعگیری است در حالی که در کانولوشن‌های شبکه عصبی تابع غیر خطی بیرون عملگر مجموعگیری قرار دارد. هر چند که تعریف کلاسیک برای عملگر LBP استفاده از تابع Heaviside است اما از تابع غیر خطی دیگری نظیر Relu و Sign نیز می‌توان استفاده کرد.

لازم به ذکر است که نحوه بهینه‌سازی شبکه‌های عصبی که دارای لایه‌های متوالی هستند با استفاده از مشتق‌گیری از تابع هزینه که در انتهای شبکه‌ی عصبی قرار دارد با الگوریتم پس انتشار ^۳ است. برای آن که روال بهینه‌سازی به درستی کار کند لازم است که هر کدام از عملگرهای شبکه عصبی مشتق پذیر باشد. برای بررسی مشتق‌پذیری عملگر پیشنهادی، مشتق این عملگر نسبت به یکی از پارامترهای آن یعنی W_p ^{*} طبق رابطه ۵.۳ بررسی می‌گردد.

²Activation function

³Back propagation

$$\begin{aligned}\frac{\partial LBP_{tr}}{\partial W_{p^*}} &= \frac{\partial(\sum_{p \in N} \sigma(I_p - I_c)e^{W_p})}{\partial W_{p^*}} = \sum_{p \in N} \frac{\partial(\sigma(I_p - I_c)e^{W_p})}{\partial W_{p^*}} \\ &= \frac{\partial(\sigma(I_{p^*} - I_c)e^{W_{p^*}})}{\partial W_{p^*}} = \sigma(I_{p^*} - I_c)e^{W_{p^*}}\end{aligned}\quad (5.3)$$

همانطور که مشاهده می‌شود مشتق عملگر پیشنهادی دارای یک مقدار $e^{W_{p^*}}$ درون خود است. برای آنکه فرآیند بهینه‌سازی هموارتر شود می‌توان تعریف عملگر پیشنهادی در رابطه ۴.۳ به جای مقدار مقدار e^{W_p} جایگزین شود در این صورت رابطه عملگر قابل آموزش به صورت رابطه ۶.۳ خواهد شد و مشتق آن نیز به صورت رابطه ۷.۳ خواهد بود.

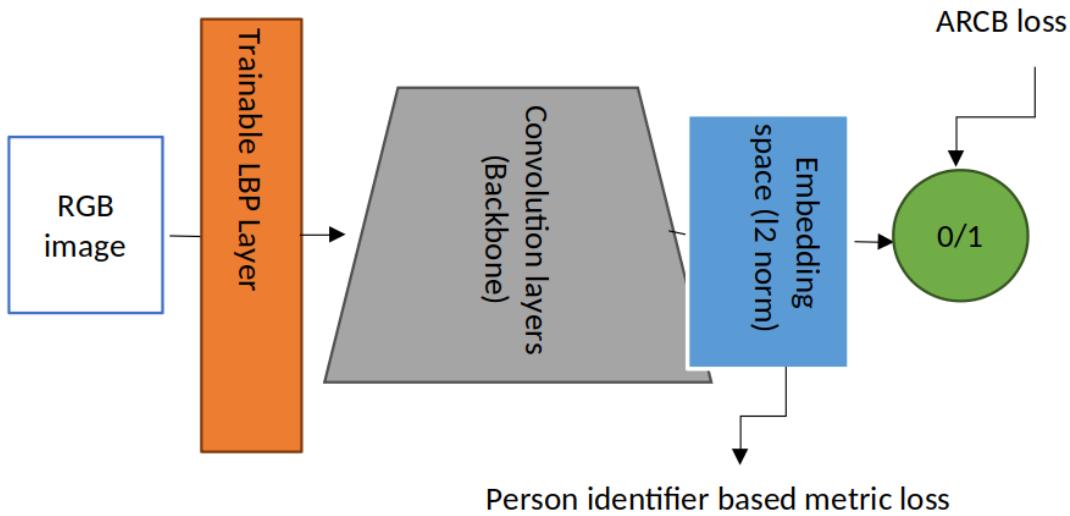
$$LBP_{tr} = \sum_{p \in N} \sigma(I_p - I_c)W_p \quad (6.3)$$

$$\frac{\partial LBP_{tr}}{\partial W_{p^*}} = \sigma(I_{p^*} - I_c) \quad (7.3)$$

۴.۳ ساختار شبکه

از آنجا که عملگر LBP_{tr} یک عملگر تحلیل تصویر در مقیاس ریز است، از این عملگر به عنوان لایه اول شبکه عمیق استفاده می‌شود. پس ساختار شبکه به صورت شکل ۱.۳ خواهد بود. تصویر ورودی به صورت سه کانال رنگی وارد عملگر می‌شود و خروجی آن به یک شبکه متشكل از لایه‌های کانولوشن داده می‌شود که در این پژوهش از شبکه EfficientNet B0 [۵۶] استفاده شده است. شبکه EfficientNet دارای هفت نسخه شبیه به هم اما با ابعاد مختلف است که طی یک الگوریتم جستجوی معماري شبکه به دست آمده است. در این پژوهش از نسخه پایه‌ی آن استفاده می‌شود. نسخه‌های بعدی آن ابعاد بزرگ‌تری دارند که این ابعاد از طریق بهینه‌سازی به دست آمده است و دقیق‌تری نیز ارائه می‌کنند.

جزئیات ساختار ارائه شده در جدول ۱.۳ گزارش شده است. در لایه‌ی اول از عملگر معرفی شده



شکل ۱.۳: شمای کلی روش پیشنهادی

استفاده می‌شود. در لایه‌های بعدی از بلوک‌های MBCConv استفاده شده است که بلوک‌های پایه شبکه Mobilenetv2 [۵۷] است با این تفاوت که squeeze and excitation [۵۸] نیز به آن اضافه شده است.

پس از اعمال لایه‌های کانولوشنی یک بردار مسطح با بعد ۱۲۸۰ خواهد بود که با توجه به تابع هزینه‌های مورد استفاده این خروجی لازم است نرمالایز شود. این خروجی نرمالایز شده با یک لایه خطی دیگر به یک نورون ختم خواهد شد. تک نورون لایه‌ی آخر مقداری بین صفر و یک خواهد داشت که بر حسب مقدار این نورون و انتخاب یک سطح آستانه طبقه‌بندی دو کلاسه صورت خواهد گرفت.

خروجی نرمالایز شده شبکه کانولوشنی در دو تابع هزینه به کار برده می‌شود. تابع هزینه اول که ARCB نام دارد برای طبقه‌بندی با حاشیه در فضای کسینوسی پیشنهاد شده است که با توجه به فرمول‌بندی آن، لازم است که بردار ویژگی نرمالایز باشد. در تابع هزینه دوم که مبتنی بر شناسه اشخاص است از فاصله اقلیدسی بردارهای ویژگی نرمالایز شده استفاده می‌کند.

تابع هزینه متداول در شبکه عصبی برای طبقه‌بندی دو کلاسه تابع آنتروپی متقاطع دودویی^۴ است. اما تحقیقات پیشین در حوزه کشف تقلب نشان داده است که این تابع هزینه به تنها یک مؤثر واقع نخواهد شد. به همین منظور تابع هزینه جدیدی برای طبقه‌بندی معرفی می‌گردد که یک

⁴Binary cross entropy

جدول ۱.۳: ساختار شبکه پیشنهادی

#Channels	Resolution	Operator	Stage
3	224×224	Tainable LBP	0
32	224×224	Conv3x3	1
16	112×112	MBConv1, k3x3	2
24	112×112	MBConv6, k3x3	3
40	56×56	MBConv6, k5x5	4
80	28×28	MBConv6, k3x3	5
112	14×14	MBConv6, k5x5	6
192	14×14	MBConv6, k5x5	7
320	7×7	MBConv6, k3x3	8
1280	7×7	Conv1x1 & Pooling	9
1	1280	Normalization & Linear	10

حاشیه امن برای طبقه بندی ایجاد کند که باعث افزایش قابلیت تعمیم‌پذیری شبکه خواهد شد.

۵.۳ تابع هزینه ARCB

در شبکه‌های عصبی زمانی که خروجی یک طبقه‌بندی چند کلاسه (بیشتر از دو) باشد از تابع فعالسازی سافت‌مکس^۵ در لایه‌ی آخر استفاده می‌شود و در طبقه‌بندی دو کلاسه از تابع فعالسازی سیگموید^۶ استفاده می‌شود. دنگ و همکاران در حوزه تشخیص چهره^۷ که یک طبقه‌بندی چند کلاسه است تابع هزینه آنتروپی متقاطع^۸ (CE) را به فضای کسینوسی برده‌اند و یک حاشیه به تابع هزینه در این فضا اضافه کرده‌اند [۵۹].

با الهام از این کار که ArcFace نام‌گذاری شده است در این پایان‌نامه، تابع هزینه BCE با هدف اعمال حاشیه در فضای کسینوسی بازنویسی می‌شود. فرض کنید خروجی شبکه استخراج ویژگی یک بردار باشد. در تصمیم‌گیری کلاسیک این بردار با بعد d وارد یک لایه شبکه عصبی با ورودی d نورون و خروجی یک نورون خواهد شد. و نهایتاً از تابع سیگموید برای بردن مقدار خروجی به مقدار بین یک و صفر استفاده خواهد شد. در تصمیم‌گیری دو کلاسه رابطه تابع هزینه آنتروپی متقاطع دودویی

⁵SoftMax

⁶Sigmoid

⁷Face recognition

⁸Cross entropy

به صورت رابطه ۸.۳ است.

$$L_{BCE} = -y_i \log P(y_i) - (1 - y_i) \log (1 - P(y_i)) \quad (8.3)$$

که در آن y_i برحسب صحیح متناظر با بردار ویژگی است. و $P(y_i)$ مقدار نورون لایه آخر است، در واقع این مقدار از نوع احتمال است یعنی مقداری بین صفر و یک دارد و هر چه به یک نزدیکتر باشد می‌توان با احتمال بیشتری تصمیم‌گیری کرد که خروجی طبقه‌بندی عدد یک است. رابطه بین نورون خروجی و بردار ویژگی به صورت رابطه ۹.۳ است.

$$P(y_i) = \text{sigmoid}(W^T X_i + b) \quad (9.3)$$

که در آن $W_p \in R^d$ وزن لایه‌ی آخر و b مقدار بایاس است. برای سادگی فرض می‌شود که بایاس صفر است. تابع سیگموید به صورت رابطه $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$ تعریف می‌شود. پیش از لایه آخر شبکه عصبی مقدار وزن‌های W_p نرمالایز کرده بردار ویژگی X_i را نرمالایز کرده و سپس مقیاس s به آن داده می‌شود. این مقیاس‌گذاری برای پایدار کردن فرآیند بهینه سازی صورت گرفته است. با نرمالایز کردن مقدار ضرب داخلی بین وزن و بردار ویژگی معادل کسینوس زاویه بین این دو بردار خواهد شد.

$$W^T X_i = |W^T| |X_i| \cos \theta_i = s \cos \theta_i \quad (10.3)$$

حال با جاگذاری این مقدار در تابع هزینه BCE به صورت رابطه ۱۱.۳ بازنویسی می‌شود.

$$L_{BCE} = -y_i \log \frac{1}{1 + e^{-s \cos \theta_i}} - (1 - y_i) \log (1 - \frac{1}{1 + e^{-s \cos \theta_i}}) \quad (11.3)$$

با توجه به مقداری برحسب واقعی که صفر یا یک است دو حالت رخ می‌دهد.

حالت اول. زمانی که برحسب یک باشد در این صورت تنها عبارت اول در رابطه ۱۱.۳ ظاهر می‌شود. در این حالت مطلوب این است که مقدار داخل لگاریتم بیشینه شود که این معادل این است که زاویه بین بردار ویژگی و وزن لایه آخر به صفر نزدیک شود. برای آنکه بهینه‌سازی با یک حاشیه

انجام شود یک مقدار حاشیه m را به آن افزوده می‌شود.

$$y_i = 1 \rightarrow \theta_i = \theta_i + m \quad (12.3)$$

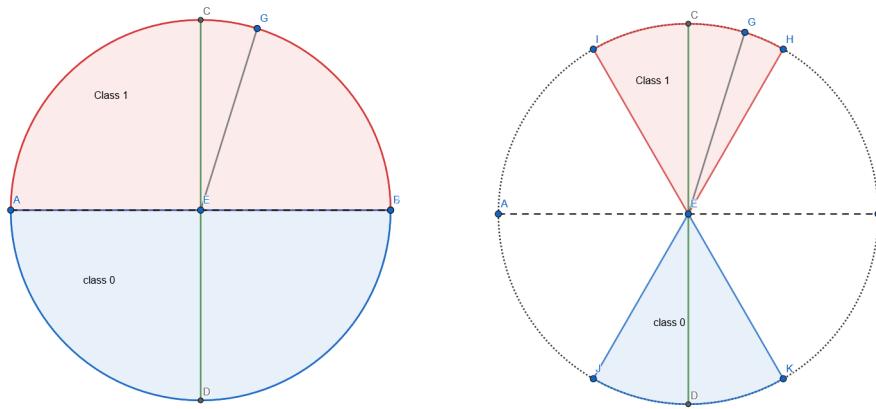
حالت دوم. زمانی که مقدار برچسب واقعی صفر باشد در این صورت عبارت دوم در رابطه ۱۱.۳ ظاهر می‌گردد. بدین منظور لازم است که عبارت داخل لگاریتم کمینه شود که معادل این است که زاویه بین وزن و بردار ویژگی به مقدار π نزدیک شود. برای آنکه بهینه‌سازی با حاشیه انجام شود یک مقدار ثابت حاشیه m از زاویه بین دو بردار کم می‌شود.

$$y_i = 0 \rightarrow \theta_i = \theta_i - m \quad (13.3)$$

با جایگذاری زاویه‌های حاشیه دار شده در روابط ۱۲.۳ و ۱۳.۳ در رابطه تابع هزینه BCE بازنویسی شده در فضای کسینوسی ۱۱.۳ تابع هزینه ARCB به صورت رابطه ۱۴.۳ به دست خواهد آمد.

$$L_{BCE} = -y_i \log \frac{1}{1 + e^{-s \cos(\theta_i + m)}} - (1 - y_i) \log \left(1 - \frac{1}{1 + e^{-s \cos(\theta_i - m)}}\right) \quad (14.3)$$

این تابع هزینه نه تنها ویژگی‌های بین دو کلاس را جدا می‌کند بلکه یک حاشیه به اندازه $2 * m$ نیز به ویژگی‌های دو کلاس مختلف در فضای کسینوسی اضافه می‌کند. این حاشیه باعث می‌شود که در فرآیند بهینه‌سازی، وزن‌های شبکه به گونه‌ای تغییر کنند که قابلیت تعمیم‌پذیری شبکه بیشتر شود. در شکل ۲.۳ تفاوت بین تابع هزینه کلاسیک و تابع هزینه با حاشیه نشان داده شده است. در صورتی که تابع هزینه به درستی بهینه شود باعث می‌شود که در لایه آخر بردارهای ویژگی در فضای کسینوسی به نحوی قرار بگیرند که زاویه بین نمونه‌ی جدید با بردار وزن در حالتی که برچسب یک است به سمت صفر میل کند و در حالتی که برچسب صفر است زاویه به سمت میل کند. و همچنین اثر افزودن حاشیه در تقسیم پذیری بین دو کلاس قابل مشاهده است. در شکل ۲.۳ در سمت چپ نتیجه جداسازی بردارهای ویژگی در حالت استفاده از تابع هزینه BCE را نشان می‌دهد و در سمت راست تابع هزینه ARCB باعث جدا شدن بردارهای ویژگی با یک حاشیه در فضای کسینوسی شده است.



شکل ۲.۳: مقایسه تابع هزینه BCE کلاسیک با نسخه‌ی حاشیه‌دار ARCB

۶.۳ تابع هزینه بر اساس شناسه‌ی شخص

در دیتاست‌های موجود در حوزه کشف تقلب در چهره، برای هر فرد چند نمونه‌ی زنده و چند نمونه تقلیبی وجود دارد. یعنی یک فرد که از چهره او برای جمع‌آوری داده استفاده شده است نمونه فیلم زنده و تقلیبی او ضبط شده است. در ویدیو واقعی و تقلیبی فرد در دیتاست یک ویژگی ظاهری یکسان شامل مشخصه‌های چهره‌ی او وجود دارد که این مشخصه‌ها با فرد دیگر متفاوت است. از طرفی در فرآیند آموزش شبکه مطلوب این است که شبکه به جای تمرکز روی ویژگی‌های ظاهری چهره‌ی افراد روی علائم مربوط به وجود یا عدم وجود تقلب در چهره تأکید داشته باشد.

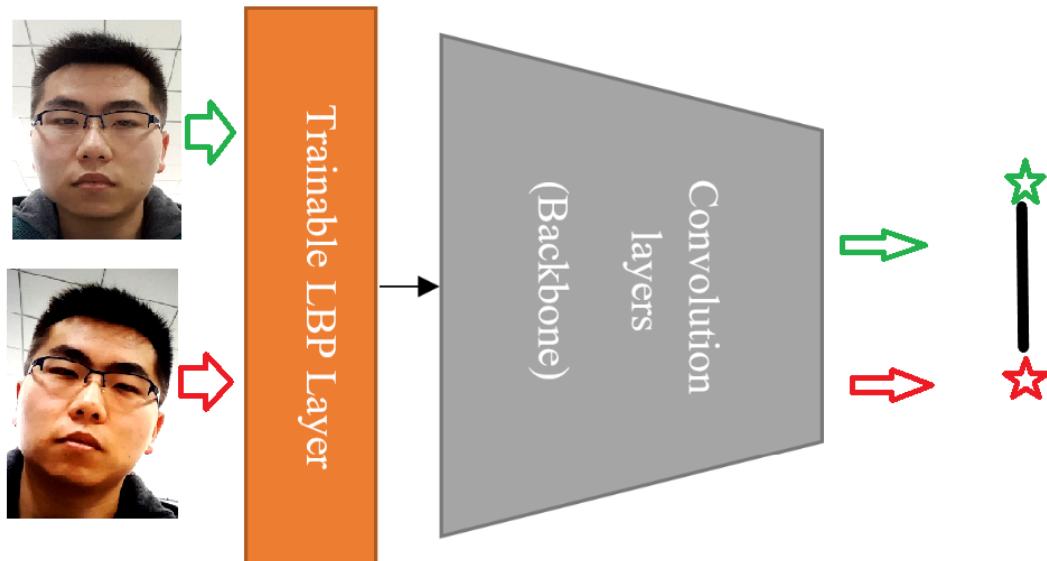
از آنجا که عمدۀ تصویر ورودی به شبکه شامل چهره و مشخصات چهره می‌شود شبکه برای نمونه‌های مختلف از یک فرد دچار چسبندگی به روی ویژگی‌های چهره او خواهد شد که این امر مطلوب نیست. بدین جهت در این بخش یک جرمیمه برای این مورد در تابع هزینه قرار داده می‌شود که هدف شبکه این باشد که به ویژگی‌های ظاهری افراد توجه نکند و توجه آن به ویژگی‌های مربوط به تقلب باشد. فرض کنید بردار ویژگی خروجی قسمت استخراج ویژگی برای فرد k ام با برچسب l به صورت $\{X_k^l \in R^d, l \in \{0, 1\}, K \in \{1, 2, \dots, M\}\}$ باشد. در حین آموزش در هر گام تعداد دسته N فرض می‌شود. در میان این N بردار ویژگی تعداد $\binom{N}{2}$ جفت بردار ویژگی وجود دارد که در میان این تعداد جفت دو حالت مهم است.

حالت اول زمانی که دو بردار ویژگی در جفت، متعلق به یک فرد ولی دارای برچسب مختلف هستند. یعنی $k_1 = k_2, l_1 \neq l_2$. در این حالت با توجه به اینکه مشخصه‌های ظاهری فرد که عمدۀ

⁹Batch size

تصویر ورودی است یکسان است لازم است که فاصله این دو نمونه بیشینه شود. با بیشینه کردن این فاصله شبکه مجبور می‌شود توجه خود را به جای مشخصه‌های ظاهری افراد به سمت ویژگی‌ای که تفاوت این دو نمونه است ببرد و این یعنی تفاوت برچسب این دو بردار ویژگی که یکی واقعی و یکی تقلبی است.

این حالت در شکل ۳.۳ نشان داده شده است. در این شکل تصویر بالایی یک تصویر زنده و تصویر دومی تقلبی است. از آنجا که این دو تصویر شبیه هستند لذا خروجی بردارهای ویژگی آنها ممکن است که نزدیک هم باشند. بردارهای ویژگی به صورت ستاره در فضای d بعدی نشان داده شده‌اند. لازم است که این فاصله بیشینه شود.



شکل ۳.۳: حالتی که دو نمونه متعلق به یک شخص ولی یکی واقعی و دیگری تقلبی است

پس در حالت اول هدف شبکه به صورت رابطه ۱۵.۳ است.

$$\max_{\Theta} d(X_{k_1}^{l_1}, X_{k_2}^{l_2}) = \min_{\Theta} \max(0, M - d(X_{k_1}^{l_1}, X_{k_2}^{l_2})) \quad (15.3)$$

که در آن Θ مجموعه وزن‌های شبکه را نشان می‌دهد و تابع d فاصله اقلیدسی بین دو بردار ویژگی

نرمالایز شده است و به صورت رابطه ۱۶.۳ تعریف می‌شود.

$$d(X_1, X_2) = \left\| \frac{X_1}{\|X_1\|} - \frac{X_2}{\|X_2\|} \right\| \quad (16.3)$$

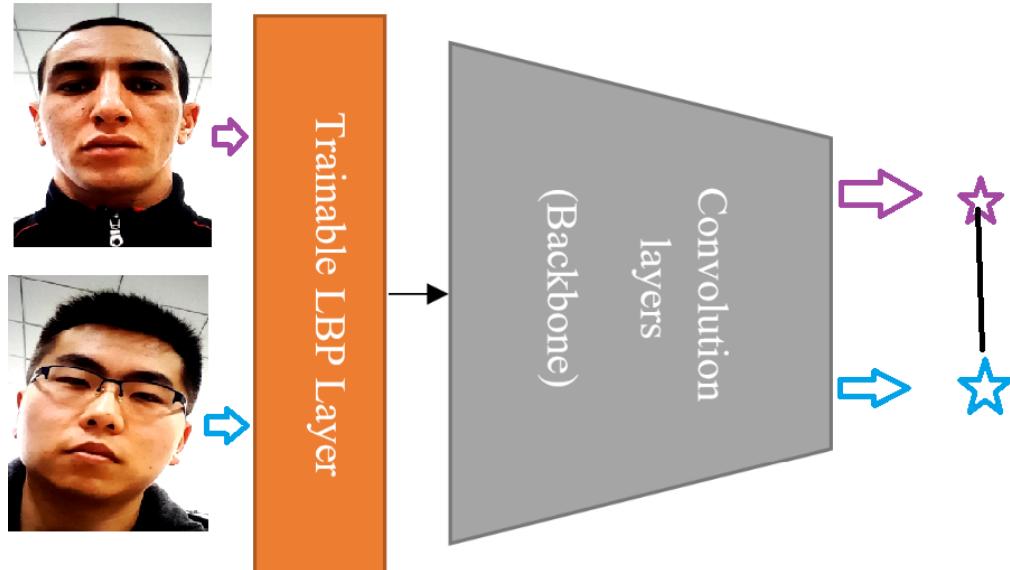
از آنجا که باید در بهینه‌سازی تابع هزینه کمینه شود بیشینه‌سازی فاصله دو بردار ویژگی معادل کمینه سازی مقدار $\max(0, M - d(X_{k_1}^{l_1}, X_{k_2}^{l_2}))$ خواهد بود. در این رابطه M یک هایپر پارامتر است که در صورتی که فاصله دو بردار ویژگی از این مقدار بیشتر باشد مقدار خروجی صفر خواهد بود و در صورتی که کمتر باشد میزان فاصله تا این مقدار M به عنوان مقدار هزینه خواهد بود. با توجه به نامساوی رابطه ۱۷.۳ بیشترین فاصله‌ای که دو بردار ویژگی در فضای نرمالایز شده خواهند داشت عدد ۲ خواهد بود و در پیاده سازی این تابع هزینه مقدار M عدد ۲ در نظر گرفته شده است.

$$\left\| \frac{X_1}{\|X_1\|} - \frac{X_2}{\|X_2\|} \right\| \leq \left\| \frac{X_1}{\|X_1\|} \right\| + \left\| \frac{X_2}{\|X_2\|} \right\| \rightarrow d(X_1, X_2) \leq 2 \quad (17.3)$$

حالت دوم زمانی که دو بردار ویژگی در جفت دارای یک برجسب ولی متعلق به اشخاص مختلفی هستند. به بیان ریاضی یعنی $k_1 \neq k_2, l_1 = l_2$. در این حالت با توجه به تفاوت مشخصه‌های ظاهری اشخاص این دو بردار ویژگی ممکن است فاصله محسوسی در فضای ویژگی داشته باشند. در این حالت مطلوب این است که فاصله این دو بردار ویژگی کم شود. در این صورت شبکه مجبور خواهد شد که به گونه‌ای از تصویر ویژگی انتخاب کند که فاصله این دو بردار ویژگی کم باشد و با رسیدن به این هدف ویژگی‌های استخراج شده بیشتر روی ویژگی‌های کشف تقلب تا ویژگی‌های ظاهری افراد تأکید دارند.

در شکل ۴.۳ این حالت نشان داده شده است. در این مثال دو تصویر ورودی هر دو از نوع تقلبی هستند ولی متعلق به اشخاص مختلفی هستند. در این شکل ستاره نشان‌گر موقعیت بردار ویژگی متناظر با این دو ورودی در فضای d بعدی است. از آنجا که دو فرد ویژگی‌های ظاهری متفاوتی دارند ممکن است فاصله بردارهای ویژگی متناظر با آن‌ها فاصله محسوسی داشته باشد. در این حالت کم کردن این فاصله مدل نظر است. پس به بیان ریاضی در این حالت تابع هزینه به صورت رابطه ۱۸.۳ خواهد بود.

$$\min_{\Theta} d(X_{k_1}^{l_1}, X_{k_2}^{l_2}) \quad (18.3)$$



شکل ۴.۳: حالتی که دو نمونه متعلق به اشخاص مختلف ولی برچسب یکسان هستند

و در نهایتتابع هزینه بر اساس شناسه اشخاص موجود در دیتابست به صورت رابطه ۱۹.۳ خواهد بود. که در آن N_i تعداد جفت نمونه‌ها با ویژگی برچسب یکسان و شخص متفاوت در دسته است و N_j تعداد جفت نمونه با ویژگی برچسب متفاوت و شناسه یکسان است.

$$L_{PiD} = \sum_{l_1 \neq l_2, k_1 \neq k_2} \frac{1}{N_i} d(X_{k_1}^l, X_{k_2}^l) + \frac{1}{N_j} \max(0, M - d(X_k^{l_1}, X_k^{l_2})) \quad (19.3)$$

نحوه تشکیل این تابع هزینه بدین صورت است که در هر گام آموزش از میان N نمونه‌ی موجود در دسته تمامی جفت‌هایی که شرط شناسه متفاوت برچسب یکسان و یا شرط شناسه یکسان و برچسب متفاوت دارند انتخاب شده و فاصله اقلیدسی آن‌ها در رابطه ۱۹.۳ قرار داده می‌شود. این تابع هزینه وقتی کمینه شود شبکه به سمتی حرکت می‌کند که ویژگی‌های مطلوب برای کشف تقلب شناسایی شده و ویژگی‌هایی مرتبط به چهره افراد نادیده گرفته شود.

تابع هزینه معرفی شده در این قسمت بر خلاف تابع هزینه ARCB وظیفه‌ی طبقه‌بندی ندارد و

تنها به عنوان یک نقش کمکی در کنار طبقه‌بندی کمک می‌کند. لذا هرچند که تابع هزینه ARCB به تنهایی قابل استفاده است اما تابع هزینه مبتنی بر شناسه اشخاص به تنهایی قابل استفاده نیست. این تابع یک تابع هزینه متريک است که با استفاده از شناسه اشخاص محدودیت خاصی روی فاصله‌ی جفت نمونه‌های خاص ارائه می‌کند. تابع هزینه متريک به صورت غیر مستقیم باعث افزایش دقت طبقه‌بندی و بهبود قابلیت تعمیم‌پذیری شبکه خواهد شد. در واقع این تابع هزینه باعث می‌شود که شبکه‌ی استخراج ویژگی به نحوی عمل کند که بردارهای ویژگی به نحو مناسبی قرار بگیرند که دقت طبقه‌بندی بیشتر شود و شبکه ویژگی‌های اساسی برای تقلب را فارغ از ویژگی‌های ظاهری افراد استخراج کند. لازم به ذکر است که ایده استفاده از تابع هزینه متريک در کنار طبقه‌بندی روشی کارآمد است که در حوزه کشف تقلب در چهره نیز با فرمول‌بندی‌های مختلف استفاده شده است [۱۵، ۱۶، ۱۷، ۵۳، ۵۴، ۵۵]. اما تابع هزینه پیشنهادی در این پایان‌نامه از شناسه اشخاص برای انتخاب جفت نمونه استفاده می‌کند که از این نظر با روش‌های قبلی متفاوت است.

در نهایت تابع هزینه کلی برای آموزش شبکه شامل ترکیب خطی از دو تابع هزینه معرفی شده و به صورت رابطه $\lambda_3 = \lambda_1 + \lambda_2$ هایپر پارامتر هستند که میزان تأکید بر هر کدام را نشان خواهند داد.

$$L_{overall} = \lambda_1 L_{ArcB} + \lambda_2 L_{PiD} \quad (20.3)$$

۷.۳ مقایسه‌ی روش پیشنهادی با پژوهش‌های قبلی

ایده استفاده از LBP در مقایسه با سایر روش‌های کلاسیک اهمیت ویژه‌ای در حوزه کشف تقلب دارد [۲، ۳، ۳۲، ۶، ۴۱، ۱۰، ۱۹]. روش LBP قابل آموزش در این پژوهش، در مقایسه با روش‌هایی که از ترکیب کانولوشن و عملگر LBP استفاده کرده‌اند [۶، ۴۱] از این نظر متفاوت است که در روش‌های قبلی عملگر LBP به صورت ایستا و بدون پارامتر بوده است اما روش پیشنهادی، یک عملگر قابل آموزش است که دارای پارامترهای یادگیری می‌باشد و در طول آموزش این پارامترها با توجه به داده‌های آموزش بهینه خواهند شد.

در [۶۰] از ایده عملگر LBP قابل آموزش برای کاهش تعداد وزن‌های شبکه استفاده شده است. در واقع روش [۶۰] روی خاصیت تنک بودن ^{۱۰} عملگر LBP تمرکز کرده است و قسمتی از وزن‌های شبکه را به صورت ثابت و با الهام از عملگر LBP در نظر گرفته است و با این روش تعداد وزن‌های

¹⁰Sparsity

قابل آموزش را در شبکه کاهش داده است و نشان داده است که سرعت اجرای شبکه بهبود می‌یابد و دقت شبکه افت کمی خواهد کرد. روش ارائه شده در این پایان‌نامه تعداد وزن‌ها را کم نمی‌کند و دارای فرمول بندی به‌گونه‌ای است که از تفاوت تمامی پیکسل‌های مجاور با پیکسل مرکزی استفاده شود. رابطه ارائه شده در [۶۰] به صورت رابطه ۲۱.۳ است. که در آن b_i^{st} وزن‌های ثابت به صورت تنک هستند و $V_{l,i}^t$ پارامترهای قابل آموزش است.

$$X_{l+1}^t = \sum_{i=1}^m \sigma \left(\sum_s b_i^{st} * X_l^s \right) V_{l,i}^t \quad (21.3)$$

در [۱۰] نیز عملگر کانولوشن با الهام از عملگر LBP تغییر داده شده است به گونه که در رابطه نهایی وزن متفاوتی به پیکسل مرکزی پنجره کانولوشن داده می‌شود و اعمال تابع غیرخطی بیرون مجموعگیری است. که به کلی از نظر فرمول بندی با عملگر ارائه شده در این پایان‌نامه متفاوت است.

$$y(p_0) = \sum_{p \in R} w(p_n) \cdot x(p_0 + p_n) + \theta(-x(p_0)) \sum_{p \in R} w(p_n) \quad (22.3)$$

فصل ۴

نتایج

۱.۴ مقدمه

در این فصل ابتدا ملاحظات پیاده‌سازی روش پیشنهادی بیان می‌شود. سپس معیارهای ارزیابی که در پژوهش‌های مرتبط در این حوزه برای توصیف میزان دقت شبکه وجود دارد تعریف می‌گردد. و در ادامه ابتدا هر قسمت از روش‌های پیشنهادی شامل عملگر LBP قابل آموزش، تابع هزینه ARCB و تابع هزینه‌ی مبتنی بر شناسه‌ی اشخاص روی یک دیتاست کوچک اجرا می‌گردد تا میزان تأثیر هر روش به تنها‌ی مشخص گردد. در انتهای از تمام روش پیشنهادی برای دیتاست‌های بزرگ‌تر استفاده شده و دقت‌های بدست آمده با دقت برخی از معروف‌ترین روش‌های موجود در این حوزه مقایسه شود.

۲.۴ ملاحظات پیاده‌سازی

در این پایان‌نامه از زبان برنامه‌نویسی پایتون و کتابخانه Pytorch^۱ استفاده شده است. این کتابخانه ابزاری قدرتمند برای مدل‌سازی شبکه‌های عمیق است. از آنجا که Pytorch انعطاف‌پذیری بیشتری نسبت به ابزارهای مشابه دارد، پیاده‌سازی توابع جدید و عملگرهای غیر متداول در آن راحت‌تر است. در این پایان‌نامه یک عملگر جدید LBP و تابع هزینه‌ی خاصی معرفی شده است که مشابه آن در ابزارهای یادگیری عمیق به صورت ماثول آمده وجود ندارد؛ اما توسط جریان محاسباتی Pytorch

¹<https://pytorch.org>

قابل پیاده سازی است.

۱.۲.۴ پیاده سازی LBP قابل آموزش

برای پیاده سازی یک عملگر جدید که دارای پارامتر قابل یادگیری باشد لازم است که یک کلاس با ارث بری از `nn.Module` نوشته شود. با این کار این کلاس دارای قابلیت `forward` و `backward` خواهد بود و قابل استفاده در جریان محاسباتی شبکه عمیق خواهد بود. برای آنکه این کلاس دارای پارامترهای یادگیرنده باشد لازم است که متغیر پارامترهای کلاس با استفاده از `nn.Parameter` نوشته شود. با این کار در صورت استفاده از این عملگر به عنوان یک لایه در شبکه، پارامترهای عملگر LBP در میان پارامترهای شبکه قرار می گیرند و بهینه سازی، منجر به بهروزرسانی این پارامترها در کنار سایر پارامترهای شبکه به صورت خودکار خواهد شد.

۲.۲.۴ پیاده سازی تابع هزینه

در هر بار `forward` داده ها به شبکه پس از بلوک استخراج ویژگی یک بردار به دست خواهد آمد که لازم است این بردار در هر مرحله برای استفاده در دو تابع هزینه معرفی شده نرمالایز شوند. در حین تست شبکه از آنجا که تغییری در وزن ها رخ نخواهد داد یک بار نرمالایز کردن کافی خواهد بود. پیاده سازی تابع ARCB با استفاده از توابع Pytorch برای پایدار بودن محاسبات انجام شده است. پارامتر s در تابع هزینه ARCB در رابطه 14.3 برابر با 64 در نظر گرفته شده است. مقادیر λ_1 و λ_2 در رابطه 20.3 هر کدام برابر با 0.5 در نظر گرفته شده اند. به منظور جلوگیری از بیش برآذش داده ها از $[61]$ در لایه آخر پس از نرمالایز کردن بردار ویژگی و پیش از طبقه بند استفاده شده است. $drop\ out$ برای پیاده سازی تابع هزینه مبتنی بر شناسه اشخاص نیاز است به غیر از تصویر ورودی و برچسب تصویر، یک عدد به عنوان شناسه نیز در اختیار باشد. در دیتاست های موجود یافتن عدد شناسه از روی نام فایل ویدئو قابل تشخیص است. برای بهینه سازی شبکه از الگوریتم آدام $[62]$ استفاده شده است.

۳.۲.۴ بارگذاری داده ها برای آموزش

برای بارگذاری و آماده سازی داده ها، توابع و کلاس های آماده در کتابخانه `Pythoch` وجود دارد که به صورت خودکار تصاویر موجود در یک پوشه استفاده خواهد کرد اما به دلیل ماهیت ویدیویی داده ها و همچنین تابع هزینه خاص معرفی شده نمی توان از توابع آماده استفاده کرد. در برخی دیتاست ها

فایلی برای مختصات چهره وجود دارد که می‌توان در هر فریم ویدیو، قسمت مربوط به چهره را برش زد و به جای استفاده از کل فریم تنها قسمت چهره به همراه کمی از قسمت پس‌زمینه تصویر به عنوان ورودی به شبکه داده شود. در دیتاست‌هایی که این فایل مختصات وجود ندارد با استفاده از روش MTCCN [۶۳] چهره فریم‌ها پیدا شده و در یک فایل متنی ذخیره شده است.

دیتاست‌های معرفی شده همگی به صورت ویدیو هستند. از آنجا که روش ارائه شده روی تک تصویر کار می‌کند یکی از نکات عملی در خصوص آموزش روی داده‌های ویدیویی، نحوه آماده‌سازی داده‌ها برای آموزش است. یک روش تبدیل ویدیو به تصویر و ذخیره آن روی دیسک است. اما این کار موجب مصرف شده حجم زیادی از دیسک خواهد شد و از آنجا که در حین آموزش لازم است که تصاویر مجدداً از دیسک به حافظه RAM بارگذاری شوند روال آموزش کند خواهد شد. از طرفی از آنجا که نمونه‌های موجود در دو کلاس با یک دیگر برابر نیستند به منظور پایدار شدن تابع هزینه ARCB لازم است که در هر دسته به تعداد نزدیک هم تصویر از هر کلاس وجود داشته باشد. از طرفی برای آنکه تابع هزینه مبتنی بر شناسه اشخاص به درستی عمل کند لازم است که پراکندگی تصاویر در هر دسته به اندازه کافی باشد تا حالت‌های مختلف از اشخاص با شناسه‌های متفاوت و برچسب متفاوت در دسته وجود داشته باشد. همچنین لازم است که ترتیب داده‌ها تا حد ممکن تصادفی باشند تا غیر یقینی بیشتری در حین آموزش، برای شبکه وجود داشته باشد.

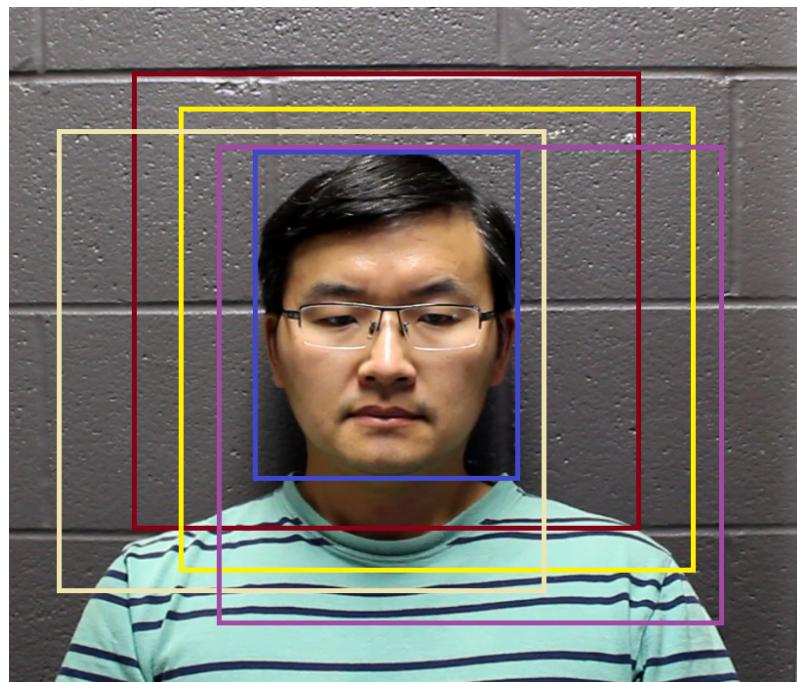
در پایه‌سازی روش این پایان‌نامه ابتدا به تعداد دسته، ویدیو در حافظه RAM بارگذاری خواهد شد و در هر مرحله یک فریم به صورت تصادفی از هر ویدیو انتخاب می‌شود که در نهایت به تعداد دسته، فریم برای آموزش وجود خواهد داشت. با این کار این فریم‌ها هر کدام از ویدیوهای متفاوتی هستند که موجب می‌شود تصاویر موجود در دسته پراکندگی لازم را داشته باشند. در مراحل بعدی از همین ویدیوها که در حافظه RAM بارگذاری شده‌اند استفاده خواهد شد و این روال تا زمانی که فریم در ویدیوها وجود داشته باشد ادامه خواهد داشت. سپس دسته ویدئو دیگری انتخاب خواهد شد و آموزش روی همه ویدیوها ادامه دارد.

از آنجا که پس از انتخاب تعدادی ویدیو، به تعداد فریم‌های آن و به صورت متوالی مرحله‌های آموزش تکرار می‌شود و فریم‌های متوالی یک ویدئو از نظر ظاهری نزدیک به هم هستند لازم است که غیر یقینی داده‌ها بیشتر شود بدین منظور از روش‌های افزایش داده^۲ به صورت تصادفی استفاده می‌شود. بدین منظور از تبدیلاتی که هر تصویر ورودی را به صورت تصادفی چرخش می‌دهند استفاده می‌شود.

به منظور جلوگیری از بیش برآش از روش پاک کردن تصادفی قسمتی از تصویر ورودی استفاده

²Data augmentation

شده است [۶۴]. همچنین هنگامی که قرار است قسمت چهره به همراه پس زمینه برش زده شود این کار به صورت یک پنجره تصادفی انجام می‌شود؛ بدین ترتیب در هر بار بارگذاری داده‌ها موقعیت چهره در تصویر برش زده تصادفی خواهد بود و لزوماً همیشه در مرکز تصویر نخواهد بود. در شکل ۱.۴ چهره در تصویر برش زده تصادفی چهره به همراه پس زمینه نشان داده شده است. در این تصویر مستطیل آبی چهره فرد را نشان می‌دهد و مستطیل‌هایی به صورت تصادفی برای هر بار انتخاب چهره انتخاب می‌شوند. و در انتهای تصویر انتخابی به 224×224 تغییر اندازه می‌شود.



شکل ۱.۴: نحوه برش زدن تصادفی چهره با مقداری از پس‌زمینه

برای پیاده سازی کلاس بارگذاری داده یک Data loader سفارشی نوشته شده است و همچنین برای آنکه استراتژی ترتیب تصادفی انتخاب ویدیو و استفاده مجدد از فریم‌های ویدیو متوالی پیاده شود یک تابع Batch sampler سفارشی نوشته شده است. در پیاده‌سازی این تابع از مفهوم Iteration در زبان برنامه‌نویسی پایتون استفاده شده است.

۳.۴ معیارهای ارزیابی

مسئله کشف تقلب یک مسئله طبقه‌بندی دو کلاسه است که در هنگام آزمون، معمولاً تعداد نمونه‌های واقعی و تقلبی یکسان نیستند. به همین دلیل معیار دقت شبکه یعنی تعداد نمونه‌های

درست پیش‌بینی شده تقسیم بر تعداد کل نمونه‌ها ملاک خوبی برای قضاوت در مورد عملکرد شبکه نیست.

بدین منظور از معیاری به نام نرخ خطای برابر^۳ (*EER*) و ترسیم آن به ازای آستانه‌های مختلف، در قالب نمودار نرخ خطای برابر استفاده می‌شود. دو حالت برای تشکیل این نمودار مهم است. نرخ خطای قبول کردن^۴ (*FAR*) نمونه، که به معنی این است که برچسب واقعی چهره زنده بوده است اما به عنوان چهره تقلیبی پیش‌بینی شده است. و نرخ خطای رد کردن^۵ (*FRR*) که به معنی این است که نمونه برچسب تقلیبی دارد ولی به عنوان چهره زنده پیش‌بینی شده است.

$$FAR = \frac{\text{number of false accepted samples}}{\text{total number of fake samples}} \quad (1.4)$$

$$FRR = \frac{\text{number of false rejected samples}}{\text{total number of real samples}} \quad (2.4)$$

معمولًاً این مقدار بر اساس یک آستانه که یک پارامتر است محاسبه می‌گردد. برای مثال در شبکه‌ی عصبی مقدار تک نورون لایه آخر باتابع فعالسازی سیگموید، مقداری بین صفر و یک خواهد داشت. و با انتخاب یک سطح آستانه و مقایسه مقدار نورون لایه‌ی آخر با این سطح آستانه تصمیم‌گیری در مورد پیش‌بینی برچسب نمونه انجام می‌شود. نرخ خطای برابر، برابر با مقداری است که *FAR* با *FRR* برابر شود.

$$\tau_{EER} = \arg \min_{\tau} |FAR(\tau) - FRR(\tau)| \quad (3.4)$$

$$EER = FAR(\tau_{EER}) = FRR(\tau_{EER}) \quad (4.4)$$

در شکل ۲.۴ این معیار را در قالب نمودار به ازای سطوح مختلف آستانه نشان می‌دهد. در دیتاست‌هایی که داده دارای سه قسمت آموزش، توسعه و آزمون است، معمولًاً روی داده‌های آموزش

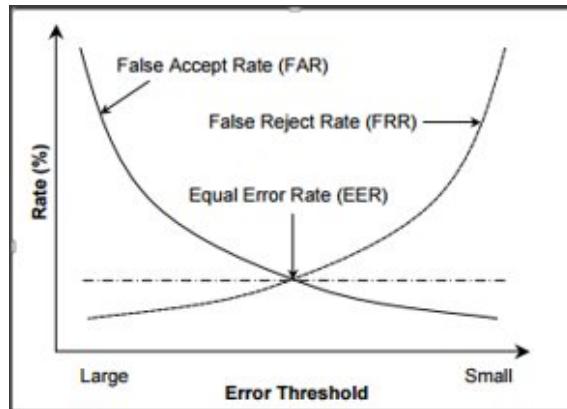
³Equal Error Rate

⁴False acceptance rate

⁵False rejection rate

وزن‌های شبکه به دست می‌آید و روی قسمت توسعه، پارامتر τ_{EER} به دست خواهد آمد. و روی قسمت آزمون معیار نصف کل نرخ خطای^۶ به صورت رابطه ۵.۴ تعریف می‌شود.

$$HTER = \frac{FAR(\tau_{EER}) + FRR(\tau_{EER})}{2} \quad (5.4)$$



شکل ۲.۴: نمودار میزان خطای برابر

با تحلیل نمودار نرخ خطای برابر، می‌توان در مورد میزان عملکرد شبکه بحث کرد. هر چه که مقدار تقاطع منحنی FAR و FRR

پایین‌تر باشد، شبکه دقیق‌تر باشد. همچنین مقدار FAR و FRR در نزدیکی‌های محل تقاطع نشان می‌دهد که شبکه چه میزان دو کلاس را از هم جدا کرده است. این یعنی نه تنها مقدار نرخ خطای برابر اهمیت دارد بلکه مطلوب این است که با تغییرات جزئی میزان سطح آستانه، نرخ خطای در اطراف آن نیز کم باشد. هر چه مقدار خطای برابر در اطراف محل تقاطع دو منحنی کوچک‌تر باشد شبکه قابلیت تعمیم‌پذیری بیشتری روی داده‌های دیده نشده خواهد داشت.

یک معیار دیگر برای ارزیابی استفاده از استاندارد ISO/IEC 30107-3 است که در آن از نرخ خطای طبقه‌بندی ارائه حمله^۷ ($APCER$) و نرخ خطای طبقه‌بندی ارائه خوب^۸ ($BPCER$)^۹ تعریف می‌شود که در آن $BPCER$ معادل $APCE$ است ولی $APCE$ معادل بیشترین FAR به ازای ابزارهای حمله مختلف است. منظور از ابزار حمله، حمله کاغذ چاپ شده یا حمله بازپخش است.

⁶Half total error rates

⁷Attack Presentation Classification Error Rate

⁸Bona Fide Presentation Classification Error Rate

رابطه ۶.۴ نحوه محاسبه $APCER$ را نشان می‌دهد که در آن PAI معادل ابزار حمله ارائه^۹ است. در دیتاست‌های مورد استفاده در این پایان‌نامه PAI پارامتر دو مقدار خواهد داشت که یکی برای حمله کاغذ چاپ شده و دیگری برای حمله بازپخش است. همچنین متوسط نرخ خطای طبقه‌بندی $APCER$ ^{۱۰} به صورت میانگین $BPCER$ و $APCER$ تعریف می‌شود.

$$APCER = \max_{PAI=1,\dots,C} FAR_{PAI} \quad (6.4)$$

$$ACER = \frac{APCER + BPCER}{2} \quad (7.4)$$

۴.۴ عملکرد مدل در دیتاست‌ها

این بخش به بررسی دقیق روش پیشنهادی روی دیتاست‌های مختلف می‌پردازد. در ابتدا برای بررسی اثر بخشی روش پیشنهادی روی دیتاست Replay که دیتاست نسبتاً کوچکی است، روش پیشنهادی بررسی می‌شود. این کار با هدف اثبات مفهوم^{۱۱} انجام می‌شود. و سپس روی دیتاست‌های دیگر دقیق گزارش می‌شود.

۱.۴.۴ اثر عملگر LBP قابل آموزش در دیتاست Replay

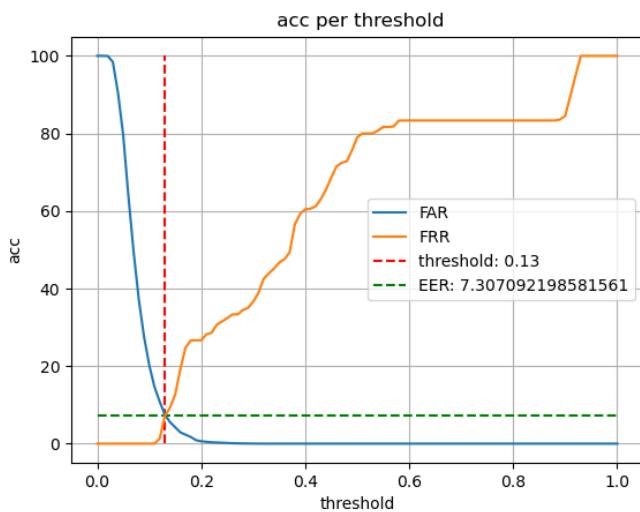
به منظور مقایسه روش‌های پیشنهادی و تأثیر آنها در بهبود دقیق ابتدا یک شبکه ALEXNET بدون عملگر LBP با تابع هزینه BCE به کار برد و نتایج نرخ خطای برابر برای این مورد به صورت شکل ۳.۴ است.

همانطور که مشاهده می‌شود با سطح آستانه متناظر با محل تقاطع دو نمودار برابر $\tau_{EER} = 0.13$ برای نورون آخر است که در این سطح آستانه نرخ خطای برابر $EER = 7.3\%$ روی داده دیده نشده به دست می‌آید. اما لازم است توجه شود تنها مقدار خطای مهم نیست و عملکرد نمودار در سایر مقادیر سطح آستانه نیز مهم است و در سطح آستانه $\tau = 0.6$ مقدار خطای $FRR = 80\%$ است که بسیار زیاد است. هر چند که در این ناحیه خطای $FAR = 0$ است. این تفاوت بین دو مقدار خطای نشان می‌دهد

⁹Presentation Attack Instrument

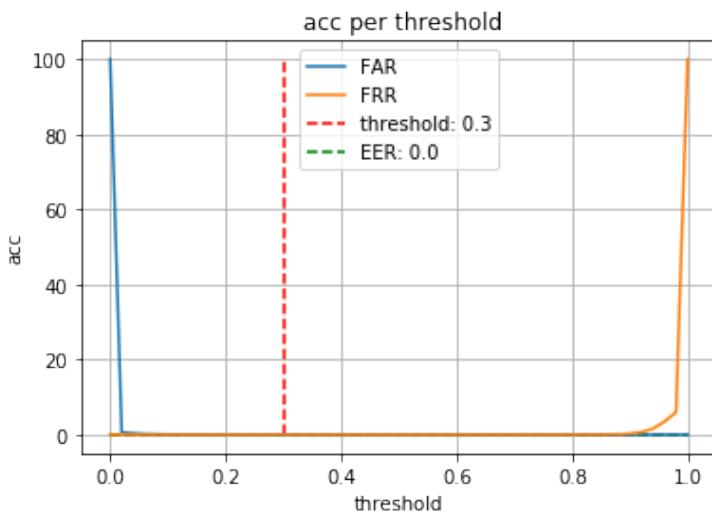
¹⁰Average Classification Error Rate

¹¹Proof of concept



شکل ۳.۴: نمودار خطای برابر برای شبکه ALEXNET و تابع هزینه BCE

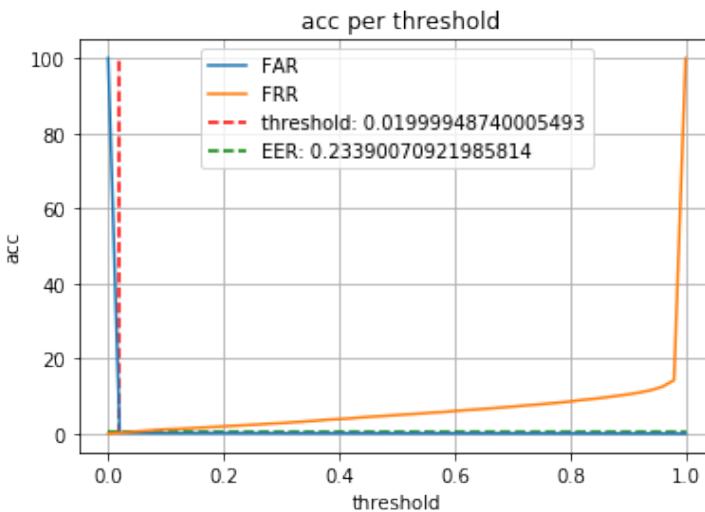
که دو کلاس از یک دیگر جدا نشده اند و استخراج ویژگی به درستی صورت نگرفته است. همچنین در اطراف سطح آستانه $\tau_{EER} = 0.13$ با کمی تغییر در سطح آستانه مقدار خطا بزرگ می‌شود. با استفاده از عملگر LBP قابل آموزش پیش از ALEXNET و تابع هزینه نیز کماکان BCE نمودار شکل ۴.۴ به دست می‌آید. همانطور که مشاهده می‌شود استفاده از تنها یک لایه LBP پیش



شکل ۴.۴: نمودار خطای برابر هنگام استفاده از عملگر LBP پیشنهادی

از ALEXNET مقدار خطا را به صفر درصد رسانده است. همچنین وضعیت خطا در اطراف آستانه نیز بهبود یافته است. از آنجا که افزودن یک لایه عملگر LBP قابل آموزش، بار محاسباتی به شبکه

اضافه می کند برای مقایسه دیگر نمودار آموزش شبکه با تابع هزینه BCE و شبکه EfficientNet B0 با شروع از وزن های تصادفی، به صورت شکل ۵.۴ است.



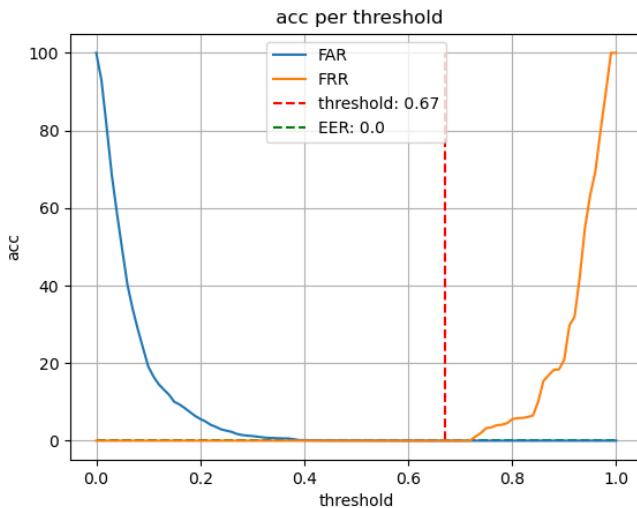
شکل ۵.۴: نمودار خطای برابر هنگام استفاده از شبکه EfficientNet B0

این نمودار نشان می دهد لزوماً استفاده از شبکه پیچیده نمی تواند به نتیجه مطلوب برساند. لازم است توجه شود این نمودار بدین معنی نیست که لایه LBP به همراه ALEXNET قدرت بیشتری نسبت به شبکه Efficient net دارد. بلکه در این کاربرد خاص و دیتاست Replay که حجم داده کمی دارد استفاده از شبکه ساده تر اما هوشمندانه با توجه به مسئله، دقت بهتری را ایجاد می کند.

۲.۴.۴ اثر تابع هزینه ARCB در دیتاست Replay

اکنون تنها از شبکه ALEXNET بدون عملگر LBP استفاده می شود ولی تابع هزینه ARCB معرفی شده به جای تابع BCE استفاده می شود. نمودار شکل ۶.۴ نشان می دهد تغییر تابع هزینه بدون تغییری در ساختار می تواند تاثیرگذار باشد. نمودار در مقایسه با نمودارهای قبلی متقارن تر شده است. در این شکل میزان خطای در اطراف سطح آستانه صفر است ولی با دور شدن از سطح آستانه و نزدیک شدن به مقدار τ و $1 - \tau$ خطای بیشتر می شود. این تأثیر حاشیه در تابع هزینه ARCB است که موجب شده است دو کلاس با یک حاشیه از یک دیگر جدا شوند. در این حالت اگر مقدار آستانه در اطراف $\tau_{EER} = 0.67$ و در بازه $0.39 \leq \tau \leq 0.75$ قرار داشته باشد مقدار خطای FAR و FRR هر دو صفر خواهد بود. این باند اطمینان حاصل افزودن m در رابطه تابع هزینه ARCB است که موجب جداسازی دو کلاس شده است. همچنین این تابع هزینه در مقایسه با تابع BCE میزان خطای

برابر را نیز کاهش داده است که بدین معناست که تابع هزینه، شبکه استخراج ویژگی را مجبور کرده است که به دنبال ویژگی‌های اساسی برای جداسازی دو کلاس با حاشیه باشد.



شکل ۴.۶: نمودار خطای برابر هنگام استفاده از تابع هزینه ARCB پیشنهادی

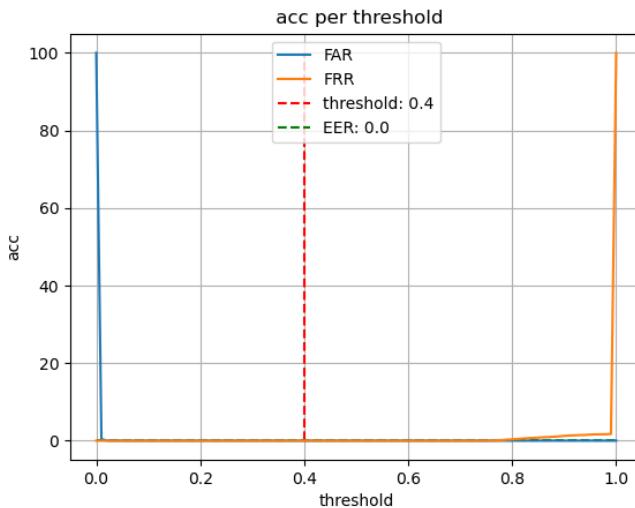
۳.۴.۴ اثر تابع هزینه بر پایه شناسه‌ی اشخاص در دیتاست Replay

اکنون از ساختار ساده ALEXNET استفاده می‌شود و تابع هزینه برای طبقه‌بند تابع BCE است ولی تابع هزینه مبتنی بر شناسه اشخاص نیز به آن افزوده شده است. نمودار این حالت به صورت شکل ۷.۴ است. همانطور که مشاهده می‌شود خطا در آستانه‌های $0.8 \leq \tau \leq 0.01$ به صورت مطلق صفر است. این خطای صفر در این باند آستانه، نشان دهنده تاثیر تابع هزینه مبتنی بر شناسه اشخاص است. با به کار بردن این تابع هزینه و بدون هیچ تغییری در ساختار شبکه، وزن‌های شبکه به گونه‌ای تغییر یافته‌اند که ویژگی‌های اساسی مربوط به تقلب را پیدا کنند و به ویژگی‌های ظاهری چهره افراد توجه نکنند.

$$L_{PiD} = \sum_{l_1 \neq l_2, k_1 \neq k_2} \frac{1}{N_i} d(X_{k_1}^l, X_{k_2}^l) + \frac{1}{N_j} \max(0, M - d(X_k^{l_1}, X_k^{l_2})) \quad (8.4)$$

چنان که در رابطه ۸.۴ مشاهده می‌شود عبارت اول این تابع هزینه باعث می‌شود جفت نمونه‌های دارای یک برچسب و شناسه شخص متفاوت به یکدیگر نزدیک‌تر شوند و در عبارت دوم جفت نمونه‌های

یک شخص و متعلق به کلاس‌های متفاوت از یکدیگر به مقدار M در فضای نرمالایز شده دور شوند.



شکل ۷.۴: نمودار خطای برابر با استفاده ازتابع هزینه مبتنی بر شناسه اشخاص

تا این قسمت اثر هر کدام از روش‌های پیشنهادی به تنها یی بررسی شده‌اند. برای ادامه فصل تمامی روش‌ها در کنار یکدیگر استفاده می‌شود. و شبکه استخراج ویژگی EfficientNet B0 است. همچنین بهمنظور تسريع در همگرا شدن شبکه، قسمت استخراج ویژگی از وزن‌های آموزش دیده روی دیتابست ImageNet [۶۵] استفاده می‌شود ولی این وزن‌ها حین آموزش تغيير می‌کند.

۴.۴.۴ نتایج روی دیتابست‌های CASIA و MSU

دیتابست‌های MSU و CASIA نسبت به دیتابست Replay دارای رزولوشن تصویر بیشتری هستند. این دیتابست‌ها بر خلاف دیتابست replay که دارای سه قسمت آموزش، توسعه و آزمون است تنها دارای دو قسمت آموزش و آزمون می‌باشد. در جدول ۱.۴ مقدار نرخ خطای برابر در قسمت آزمون دیتابست گزارش شده است.

جدول ۱.۴: خطای برابر روی دیتابست‌های CASIA و MSU

EER (%)	Dataset
0.54	CASIA
0.0	MSU

از آنجا که این دو دیتاست کمی قدیمی هستند رسیدن به نرخ خطای صفر چندان دشوار نیست. در پژوهش‌های اخیر در این حوزه، عمدۀ گزارش‌های دقت روی دیتاست‌های SIW و OULU است. این دو دیتاست نسبت به دیتاست‌های قبلی جدیدتر و دارای حجم بیشتری هستند. به همین دلیل در پژوهش‌های اخیر بیشتر از این دو دیتاست استفاده شده است. هر کدام از این دو دیتاست دارای پروتکل‌های مختلفی هستند که حالت‌های مختلف برای بررسی تعمیم‌پذیری مدل را نشان می‌دهد.

۵.۴.۴ دقت در دیتاست SIW

در پروتکل اول دیتاست SIW به بررسی تغییر حالت چهره می‌پردازد. بدین منظور برای آموزش از ۶۰ فریم اول هر ویدیو استفاده می‌شود ولی برای آزمون از تمامی فریم‌های ویدیوهای تست استفاده می‌شود. از آنجا که در فریم‌های ابتدایی هر ویدیو، کاربر صورت خود را تکان نمی‌دهد پس داده‌های آموزش تنها شامل تصاویر صورت با موقعیت ثابت در مقابل دوربین است. ولی داده‌های تست شامل همه حالت‌های حرکت چهره در ویدئو است. این پروتکل قابلیت تعمیم‌پذیری مدل ارائه شده را در حالت‌های مختلف چهره نشان می‌دهد. نتایج این حالت در جدول ۲.۴ همراه با مقایسه با برخی روش‌های معروف ذکر شده است.

جدول ۲.۴: نرخ در پروتکل اول دیتاست SIW

Method	APCER	BPCER	ACER
Auxiliary [۸]	3.58	3.58	3.58
LGSC [۱۷]	0	0.50	0.25
STASN [۳۹]	-	-	1
CDCN [۱۰]	0.07	0.17	0.12
SGTD [۴۳]	0.64	0.17	0.4
3DPC-NET [۴۸]	0.69	0.17	0.4
ARCB+PID	0.14	0.12	0.13

در پروتکل دوم از چهار نوع حمله‌ی بازپخش، هر بار یک حمله برای تست کنار گذاشته می‌شود و آموزش شبکه روی سه حمله‌ی بازپخش دیگر انجام می‌شود. پس برای این پروتکل چهار حالت مختلف وجود دارد که میانگین و واریانس دقت روی چهار حالت گزارش می‌شود. این پروتکل با هدف بررسی عدمکرد روش پیشنهادی روی نوع حمله بازپخش دیده نشده طراحی شده است. نتایج در جدول ۲.۴ گزارش شده است.

جدول ۳.۴: نرخ در پروتکل دوم دیتاست SIW

Method	APCER	BPCER	ACER
Auxiliary [۸]	0.57 ± 0.69	0.57 ± 0.69	0.57 ± 0.69
LGSC [۱۷]	0 ± 0	0 ± 0	0 ± 0
STASN [۳۹]	-	-	0.28 ± 0.05
CDCN [۱۰]	0 ± 0	0 ± 0.09	0.04 ± 0.5
SGTD [۴۳]	0.0 ± 0.0	0.04 ± 0.08	0.02 ± 0.04
3DPC-NET [۴۸]	0.46 ± 0.28	0.43 ± 0.06	0.45 ± 0.14
ARCB+PID	0.0075 ± 0.0129	0.01 ± 0.0173	0.0087 ± 0.0151

۶.۴.۴ دقت در دیتاست OULU

دیتاست OULU نیز دارای چهار پروتکل مختلف است که در این پایاننامه دقت روی پروتکل اول و دوم گزارش شده است. دیتاست OULU در سه مکان مختلف تصویر برداری شده است. در پروتکل اول روی ویدیوهای مربوط به مکان اول و دوم آموزش صورت می‌گیرد و در ویدیوهای مکان سوم آزمون انجام می‌گیرد. این پروتکل با این هدف ارائه شده است که قابلیت روش پیشنهادی با تغییر مکان تصویربرداری ارزیابی شود. نتایج به دست آمده به همراه مقایسه با روش‌هایی که روی این دیتاست روش خود را ارزیابی کرده‌اند در جدول ۴.۴ گزارش شده است.

جدول ۴.۴: دقت در پروتکل اول دیتاست OULU

Method	APCER	BPCER	ACER
GFA[۵۴]	2.5	8.9	5.7
Auxiliary [۸]	1.6	1.6	1.6
FaceDs [۵۰]	1.2	1.7	1.5
LGSC [۱۷]	0.8	0	0.4
STASN [۳۹]	1.2	2.5	1.9
CDCN [۱۰]	0.4	0	0.2
SGTD [۴۳]	2.0	0.0	1.0
DeepPixBis [۱۲]	0.83	0	0.42
STDN[۱۳]	0.8	1.3	1.1
3DPC-NET [۴۸]	2.3	0	1.2
ARCB+PID	2.58	2	2.29

در پروتکل دوم از دو حمله کاغذ چاپ شده و دو حمله بازپخش موجود در دیتاست یک حمله چاپ و یک حمله بازپخش برای آموزش و حمله چاپ و بازپخش دیگر برای تست استفاده می‌شود. هدف این پروتکل ارزیابی ابزار حمله دیده نشده در آموزش است. نتایج مربوط به دقت مدل ارائه شده در این پروتکل در جدول ۵.۴ آورده شده است.

جدول ۵.۴: دقت در پروتکل دوم دیتاست OULU

Method	APCER	BPCER	ACER
GFA[۵۴]	2.5	1.3	1.9
Auxiliary [۸]	2.7	2.7	2.7
FaceDs [۵۰]	4.2	4.4	4.3
LGSC [۱۷]	0.8	0.6	0.7
STASN [۳۹]	4.2	0.3	2.2
CDCN [۱۰]	1.8	0.8	1.3
SGTD [۴۳]	2.5	1.3	1.9
DeepPixBis [۱۲]	11.4	0.6	6.0
STDN[۱۳]	2.3	1.6	1.9
3DPC-NET [۴۸]	3.1	2.8	3.0
ARCB+PID	0.97	0.97	0.97

۷.۴.۴ نتایج در آزمون بین دیتاست

همانطور که در قسمت‌های قبلی مشاهده شده است با روش‌های جدید یادگیری عمیق، رسیدن به نرخ خطای نزدیک صفر، دور از انتظار نیست. اما نحوه عملکرد مدل ارائه شده روی داده‌های دیده نشده با توزیع متفاوت همچنان موضوع چالشی و مهم در تحقیقات دانشگاهی است. یک مدل ممکن است روی یک دیتاست با توزیع خاص به دقت بسیار بالایی برسد ولی هنگام استفاده از این مدل در دنیای واقعی، ضعیف عمل کند.

نتایج ارائه شده تا اینجا دقت مدل درون دیتاست بوده است. یکی دیگر از مسائل مهم در حوزه کشف تقلب، بررسی دقت در آزمون بین دو دیتاست مختلف است. بدین منظور مدل روی یک دیتاست آموزش داده می‌شود و روی دیتاست دیگر ارزیابی می‌شود. برای بررسی دقت مدل در تست بین دیتاست، شبکه روی دیتاست CASIA آموزش داده شده است و روی دیتاست Replay ارزیابی شده است. نتایج این حالت در جدول ۶.۴ به همراه دقت پژوهش‌های دیگر گزارش شده است.

جدول ۶.۴: نتایج روی آزمون بین دیتاست

Method	HTER %
STASN [۳۹]	31.5
SGTD [۴۳]	17
Auxiliary [۸]	27.6
FaceDs [۵۰]	28.5
GFA[۵۴]	21.4
LGSC [۱۷]	27.4
3DPC-NET [۴۸]	23.4
ARCB+PID	21.25

با مقایسه نتایج دقت در آزمون بین دیتاست و درون دیتاست تفاوت قابل ملاحظه خطأ، دیده می شود.

فصل ۵

نتیجه‌گیری و کارهای آینده

۱.۵ نتیجه‌گیری

در این پایان‌نامه به بررسی روش‌های موجود در حوزه امنیت سیستم‌های احراز هویت با استفاده از چهره پرداخته شد. روش‌های موجود به صورت عمده از سیگنال‌های کمکی نظیر عمق استفاده کرده‌اند. همچنین در بسیاری از روش‌ها از فریم‌های متوالی ویدئو برای استنتاج در مورد زنده یا تقلیبی بودن چهره استفاده شده است. در این پایان‌نامه روشی مبتنی بر استفاده از تنها یک فریم توسعه داده شده است. همچنین روش پیشنهادی نیازی به عمق به عنوان سیگنال کمکی ندارد. با این وجود روش پیشنهادی در پروتکل‌های اول و دوم در دو دیتاست بزرگ و جدید در این حوزه به دقت‌های رقابتی با روش‌های دیگر رسیده است.

از آنجا که قسمت اصلی پردازش در روش پیشنهادی بر پایه شبکه EfficientNet B0 است حجم محاسباتی روش پیشنهادی بهینه است. از نظر زمان پاسخ، به دلیل استفاده از یک فریم، سریع است. در این پایان‌نامه عملگری جدید بر پایه LBP پیشنهاد شده است که خاصیت آموزش پذیری شبکه‌های CNN را دارد. همچنین به علت توسعهتابع هزینه با حاشیه، قابلیت تفکیک پذیری شبکه بیشتر شده است و استفاده از تابع هزینه مبتنی بر شناسه اشخاص موجب افزایش تعمیم‌پذیری شبکه شده است. مزیت استفاده از تابع هزینه در این است که افزایش دقت بدون افزودن بار محاسباتی به شبکه حاصل می‌شود. لذا در روش پیشنهادی با وجود آنکه زمان آموزش بیشتری نیاز دارد اما زمان ارزیابی و استفاده از شبکه تغییری نمی‌کند.

۲.۵ پیشنهاد کارهای آینده

در این پژوهش از EfficientNet B0 استفاده شده است. پژوهش‌های بعدی می‌تواند شامل استفاده از ساختار از ابتدا طراحی شده باشد. همچنین بهمنظور افزایش دقت استفاده از ساختار توجه^۱ در شبکه می‌تواند مفید باشد. استفاده از دنباله ویدیویی بهجای یک فریم با یک ساختار جدید می‌تواند به افزایش دقت کمک کند. بهمنظور آنالیز بهتر بافت در تصویر، عملگر LBP می‌تواند توسعه بیشتری داده شود به گونه‌ای که در تمامی لایه‌های شبکه بهجای کانولوشن قرار بگیرد. همچنین تابع هزینه ARCB می‌تواند مشابه روش [۱۲] روی یک صفحه مسطح بهجای یک نورون نوشته شود. تابع هزینه مبتنی بر شناسه اشخاص می‌تواند بهجای استفاده از شناسه اشخاص روی ویژگی‌های دیگر نظری ابزار حمله بازنویسی شود. همچنین استفاده از عمق در کنار روش پیشنهادی ممکن است دقت بهتری به دست آورد.

در این پایان‌نامه مرکز روی حملات چاپ و بازپخش بوده است. در این حوزه دیتاست‌هایی وجود دارند که شامل حملات استفاده از ماسک هستند. استفاده از روشی مشابه روش پیشنهادی روی دیتاست‌هایی که دارای تصاویر RGB و IR هستند نیز می‌تواند پژوهش بعدی باشد. علاوه بر این، در این پایان‌نامه بهمنظور افزایش سرعت همگرایی، از روش بهینه‌سازی آدام و شبکه با وزن‌های آموزش دیده شده استفاده شده است. پژوهش بعدی می‌تواند شامل استفاده از بهینه‌سازی SGD و شروع با وزن‌های تصادفی و آموزش روی تعداد ایپاک زیاد باشد که ممکن است نقطه بهینه بهتری را پیدا کند.

¹Attention

مراجع

- [1] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, “Oulu-npu: A mobile face presentation attack database with real-world variations,” in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*. IEEE, 2017, pp. 612–618.
- [2] J. Määttä, A. Hadid, and M. Pietikäinen, “Face spoofing detection from single images using micro-texture analysis,” in *2011 international joint conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–7.
- [3] I. Chingovska, A. Anjos, and S. Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*. IEEE, 2012, pp. 1–7.
- [4] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, “Deep learning for face anti-spoofing: A survey,” *arXiv preprint arXiv:2106.14948*, 2021.
- [5] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, “An original face anti-spoofing approach using partial convolutional neural network,” in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2016, pp. 1–6.
- [6] Y. A. U. Rehman, L.-M. Po, and J. Komulainen, “Enhancing deep discriminative feature maps via perturbation for face presentation attack detection,” *Image and Vision Computing*, vol. 94, p. 103858, 2020.
- [7] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, “Face anti-spoofing using patch and depth-based cnns,” in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 319–328.
- [8] Y. Liu, A. Jourabloo, and X. Liu, “Learning deep models for face anti-spoofing: Binary or auxiliary supervision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 389–398.

- [9] Z. Wang, C. Zhao, Y. Qin, Q. Zhou, G. Qi, J. Wan, and Z. Lei, “Exploiting temporal and depth information for multi-frame face anti-spoofing,” *arXiv preprint arXiv:1811.05118*, 2018.
- [10] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, “Searching central difference convolutional networks for face anti-spoofing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5295–5305.
- [11] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, “Face anti-spoofing with human material perception,” in *European Conference on Computer Vision*. Springer, 2020, pp. 557–575.
- [12] A. George and S. Marcel, “Deep pixel-wise binary supervision for face presentation attack detection,” in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8.
- [13] Y. Liu, J. Stehouwer, and X. Liu, “On disentangling spoof trace for generic face anti-spoofing,” in *European Conference on Computer Vision*. Springer, 2020, pp. 406–422.
- [14] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.
- [15] R. Shao, X. Lan, J. Li, and P. C. Yuen, “Multi-adversarial discriminative deep domain generalization for face presentation attack detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 023–10 031.
- [16] Y. Jia, J. Zhang, S. Shan, and X. Chen, “Single-side domain generalization for face anti-spoofing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8484–8493.
- [17] H. Feng, Z. Hong, H. Yue, Y. Chen, K. Wang, J. Han, J. Liu, and E. Ding, “Learning generalized spoof cues for face anti-spoofing,” *arXiv preprint arXiv:2005.03922*, 2020.
- [18] A. George and S. Marcel, “Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 361–375, 2020.
- [19] K.-Y. Zhang, T. Yao, J. Zhang, Y. Tai, S. Ding, J. Li, F. Huang, H. Song, and L. Ma, “Face anti-spoofing via disentangled representation learning,” in *European Conference on Computer Vision*. Springer, 2020, pp. 641–657.
- [20] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” in *2012 5th IAPR international conference on Biometrics (ICB)*. IEEE, 2012, pp. 26–31.

- [21] D. Wen, H. Han, and A. K. Jain, “Face spoof detection with image distortion analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.
- [22] R. Ramachandra and C. Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey,” *ACM Computing Surveys (CSUR)*, vol. 50, no. 1, pp. 1–37, 2017.
- [23] K. Patel, H. Han, and A. K. Jain, “Secure face unlock: Spoof detection on smartphones,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2268–2283, 2016.
- [24] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face antispoofing using speeded-up robust features and fisher vector encoding,” *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 141–145, 2017.
- [25] W. R. Schwartz, A. Rocha, and H. Pedrini, “Face spoofing detection through partial least squares and low-level descriptors,” in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–8.
- [26] J. Komulainen, A. Hadid, and M. Pietikäinen, “Context based face anti-spoofing,” in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2013, pp. 1–8.
- [27] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor,” in *2013 International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–6.
- [28] W. Yin, Y. Ming, and L. Tian, “A face anti-spoofing method based on optical flow field,” in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*. IEEE, 2016, pp. 1333–1337.
- [29] A. Anjos, M. M. Chakka, and S. Marcel, “Motion-based counter-measures to photo attacks in face recognition,” *IET biometrics*, vol. 3, no. 3, pp. 147–158, 2014.
- [30] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection using 3d structure recovered from a single camera,” in *2013 international conference on biometrics (ICB)*. IEEE, 2013, pp. 1–6.
- [31] M. De Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, “Moving face spoofing detection via 3d projective invariants,” in *2012 5th IAPR International Conference on Biometrics (ICB)*. IEEE, 2012, pp. 73–78.

- [32] T. d. Freitas Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, “Lbp- top based counter-measure against face spoofing attacks,” in *Asian Conference on Computer Vision*. Springer, 2012, pp. 121–132.
- [33] J. Yang, Z. Lei, and S. Z. Li, “Learn convolutional neural network for face anti-spoofing,” *arXiv preprint arXiv:1408.5601*, 2014.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [35] J. Gan, S. Li, Y. Zhai, and C. Liu, “3d convolutional neural network based on face anti-spoofing,” in *2017 2nd international conference on multimedia and image processing (ICMIP)*. IEEE, 2017, pp. 1–5.
- [36] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, “Learning generalized deep feature representation for face anti-spoofing,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2639–2652, 2018.
- [37] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [38] Z. Xu, S. Li, and W. Deng, “Learning temporal features using lstm-cnn architecture for face anti-spoofing,” in *2015 3rd IAPR asian conference on pattern recognition (ACPR)*. IEEE, 2015, pp. 141–145.
- [39] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, and W. Liu, “Face anti-spoofing: Model matters, so does data,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3507–3516.
- [40] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” 2015.
- [41] L. Li and X. Feng, “Face anti-spoofing via deep local binary pattern,” in *Deep Learning in Object Detection and Recognition*. Springer, 2019, pp. 91–111.
- [42] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, “Joint 3d face reconstruction and dense alignment with position map regression network,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 534–551.
- [43] Z. Wang, Z. Yu, C. Zhao, X. Zhu, Y. Qin, Q. Zhou, F. Zhou, and Z. Lei, “Deep spatial gradient and temporal depth learning for face anti-spoofing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5042–5051.

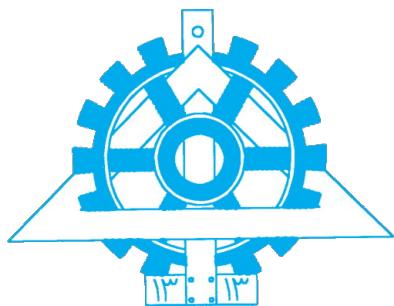
- [44] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.
- [45] B. Zoph and Q. V. Le, “Neural architecture search with reinforcement learning,” *arXiv preprint arXiv:1611.01578*, 2016.
- [46] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, “Nas-fas: Static-dynamic central difference network search for face anti-spoofing,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 9, pp. 3005–3023, 2020.
- [47] Z. Yu, Y. Qin, X. Xu, C. Zhao, Z. Wang, Z. Lei, and G. Zhao, “Auto-fas: Searching lightweight networks for face anti-spoofing,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 996–1000.
- [48] X. Li, J. Wan, Y. Jin, A. Liu, G. Guo, and S. Z. Li, “3dpc-net: 3d point cloud network for face anti-spoofing,” in *2020 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2020, pp. 1–8.
- [49] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [50] A. Jourabloo, Y. Liu, and X. Liu, “Face de-spoofing: Anti-spoofing via noise modeling,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 290–306.
- [51] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [52] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [53] D. Pérez-Cabo, D. Jiménez-Cabello, A. Costa-Pazo, and R. J. López-Sastre, “Deep anomaly detection for generalized face anti-spoofing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [54] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, and J. Feng, “Learning generalizable and identity-discriminative representations for face anti-spoofing,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–19, 2020.

- [55] X. Xu, Y. Xiong, and W. Xia, “On improving temporal consistency for online face liveness detection system,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 824–833.
- [56] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [57] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [58] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [59] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “Arcface: Additive angular margin loss for deep face recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.
- [60] F. Juefei-Xu, V. Naresh Boddeti, and M. Savvides, “Local binary convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 19–28.
- [61] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [62] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [63] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [64] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.
- [65] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

Abstract

An automated authentication method that makes use of the user's face is one option. Because of substantial advancements in face recognition technology, facial recognition has become increasingly common. Face authentication is not totally safe, however, and an attacker can authenticate by printing the target person's face or replaying a video of him / her instead of the target person, which is a known vulnerability. Academic and industrial research have therefore developed methods and algorithms in this field to increase the security of face authentication systems, which have been tested and proven to work. The goal of this investigation is to determine the difference between the real face image and the phony face image supplied by the attacker. Deep learning algorithms have been used to classify the real image against the fake images provided by the attacker as a result of the increased use of deep learning methods in machine vision problems. Deep learning algorithms have been used to classify the real image against the fake images provided by the attacker. In this dissertation, a novel operator is presented to replace one of the convolution layers in a machine vision system by integrating the classical way of machine vision with deep learning methods. Additionally, in order to improve the classification accuracy between the two categories of real and counterfeit images, a cost function for binary classification with a margin has been proposed, which adds a margin to the samples of the two classes in order to space the samples of the two classes apart. In addition, in order to improve the network's scalability, a specific metric cost function for the problem of face fraud detection has been presented, which makes use of the identities of persons to do this. Furthermore, on certain well-known datasets in this sector, the results are presented, and the overall performance of the suggested approach is reviewed, as well as the execution speed of the algorithm under consideration.

Keywords Authentication, face use, security of authentication systems, combination of machine vision methods with deep learning, marginal cost function, biometric, proprietary metric cost function



University of Tehran
College of Engineering

Faculty of Electrical and
Computer Engineering
Faculty of Electrical and
Computer Engineering



Anti-spoofing for authentication based on face recognition

A Thesis submitted to the Graduate Studies Office
In partial fulfillment of the requirements for
The degree of Master of Science
in Electrical Engineering - Cryptography and Secure Communication

By:

مهدیه احمدی

Supervisor:

Dr Mohammad Ali Akhaee

May 2022