

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

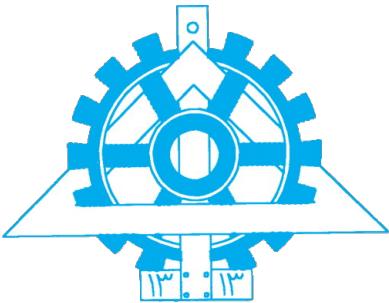
میثم شهبازی دستجرد

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 کامپیوتر



گواهی دفاع از پایان نامه کارشناسی ارشد

هیأت داوران پایان نامه کارشناسی ارشد آقای / خانم میثم شهبازی دستجرده به شماره دانشجویی ۸۱۰۱۹۷۲۸۹ در رشته مهندسی برق - گرایش مخابرات امن و رمزگاری را در تاریخ با عنوان

«جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره»
به حروف

		با نمره نهایی
--	--	---------------

ارزیابی کرد.

--

و درجه

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

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

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

۱۴۰۱ اردیبهشت

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 تهاجمی و تابع هزینه‌های مختلف

۴.۲	دیتاستهای مورد استفاده	۲۷
۱.۴.۲	دیتاست Replay	۲۸
۲.۴.۲	دیتاست CASIA	۲۹
۳.۴.۲	دیتاست MSU	۲۹
۴.۴.۲	دیتاست OULU	۲۹
۵.۴.۲	دیتاست SIW	۳۱
فصل ۳: روش پیشنهادی		
۱.۳	مقدمه	۳۳
۲.۳	مروری بر عملگر کانولوشن	۳۴
۳.۳	عملگر تحلیل ریزبافت قابل آموزش	۳۵
۴.۳	تابع هزینه ARCB	۳۶
۵.۳	تابع هزینه بر اساس شناسهی شخص	۳۹
۶.۳	مقایسهی روش پیشنهادی با پژوهش‌های قبلی	۴۳
فصل ۴: نتایج		
۱.۴	مقدمه	۴۵
۲.۴	ملاحظات پیاده‌سازی	۴۵
۱.۲.۴	پیاده سازی LBP قابل آموزش	۴۶
۲.۲.۴	پیاده‌سازی تابع هزینه	۴۶
۳.۲.۴	بارگذاری داده‌ها برای آموزش	۴۶
۳.۴	معیارهای ارزیابی	۴۸
۴.۴	عملکرد مدل در دیتاستها	۵۰
۱.۴.۴	اثر عملگر LBP قابل آموزش در دیتاست Replay	۵۰
۲.۴.۴	اثر تابع هزینه ARCB در دیتاست Replay	۵۲
۳.۴.۴	اثر تابع هزینه بر پایه شناسهی اشخاص در دیتاست Replay	۵۳
۴.۴.۴	نتایج روی دیتاست‌های CASIA و MSU	۵۳

۵۴	دقت در دیتاست SIW	۵.۴.۴
۵۵	دقت در دیتاست OULU	۶.۴.۴
۵۵	نتایج روی آزمون بین دیتاست	۷.۴.۴
۵۷	فصل ۵: نتیجه‌گیری و کارهای آینده	
۵۷	نتیجه‌گیری	۱.۰
۵۸	پیشنهاد کارهای آینده	۲.۰
۵۹	مراجع	

فهرست تصاویر

۱.۱	نمونه‌ای از تصاویر واقعی و تقلبی در حوزه چهره [۴]	۳
۱.۲	ساختار کلی الگوریتم‌های کشف تقلب در چهره	۸
۲.۱	مثالی از محاسبه LBP [۲۹]	۱۰
۳.۲	روش تصمیم‌گیری بر اساس استفاده از LBP [۲۹]	۱۲
۴.۲	روش تحلیل ریزبافت در نواحی مختلف تصویر [۵]	۱۳
۵.۲	حالات مختلف ترکیب ویژگی‌های دستی و ویژگی‌های شبکه عمیق [۵۳]	۱۴
۶.۲	استفاده از شبکه تنظیم دقیق شده و اعمال PCA روی ویژگی‌های عمیق [۲۵]	۱۵
۷.۲	روش ترکیب LBP و کانولوشن [۳۴]	۱۵
۸.۲	روش‌های مختلف یادگیری عمیق در حوزه‌ی کشف تقلب چهره [۵۳]	۱۶
۹.۲	استفاده از عمق برای کشف تقلب در چهره [۲]	۱۷
۱۰.۲	روش استفاده از عمق و تخمین rPPG [۲۷]	۱۸
۱۱.۲	استفاده از ویژگی‌های عمیق در طول زمان [۴۴]	۱۸
۱۲.۲	نحوه محاسبه تابع هزینه CDL [۵۶]	۱۹
۱۳.۲	عملگر کانولوشن تغییر یافته [۵۶]	۲۰
۱۴.۲	روش استفاده از فیلتر bilateral در شبکه عمیق [۵۲]	۲۱
۱۵.۲	تابع هزینه BCE روی یک صفحه مسطح به جای یک نورون [۱۳]	۲۲
۱۶.۲	ساختار بر پایه استفاده از شبکه مولد برای تخمین علاطم تقلب در سطوح مختلف	۲۳
۱۷.۲	نحوه عملکرد تابع هزینه سه‌گانه روی فاصله بردارهای ویژگی [۳۶]	۲۴

۱۸.۲	نحوه اثر تابع هزینه روی فاصله نمونه‌ها در دیتاست‌های مختلف [۳۸]	۲۴
۱۹.۲	تابع هزینه نامتقارن برای کاهش فاصله نمونه‌های از یک کلاس [۱۸]	۲۵
۲۰.۲	ساختار $\square-\square-\square$ و تابع هزینه سه‌گانه [۹]	۲۵
۲۱.۲	کاهش فاصله نمونه‌های واقعی تا مرکز و افزایش فاصله نمونه‌های تقلبی تا مرکز [۱۴]	۲۶
۲۲.۲	استفاده از LBP در کنار عمق برای یافتن ویژگی‌های خوش ساخت [۵۸]	۲۷
۲۳.۲	نمونه‌هایی از دیتاست Replay [۵]	۲۸
۲۴.۲	نمونه‌هایی از دیتاست CASIA [۵۹]	۲۹
۲۵.۲	نمونه‌هایی از دیتاست MSU [۴۵]	۳۰
۲۶.۲	نمونه‌های واقعی در دیتاست OULU [۴]	۳۰
۲۷.۲	نمونه‌های تقلبی در دیتاست OULU [۴]	۳۱
۲۸.۲	نمونه‌های از دیتاست SIW [۲۷]	۳۲
۱.۳	مقایسه تابع هزینه BCE کلاسیک با نسخه‌ی حاشیه‌دار	۳۹
۲.۳	حالتی که دو نمونه متعلق به یک شخص ولی یکی واقعی و دیگری تقلبی است	۴۱
۳.۳	حالتی که دو نمونه متعلق به اشخاص مختلف ولی برچسب یکسان هستند	۴۲
۱.۴	نحوه برش زدن تصادفی چهره با مقداری از پس‌زمینه	۴۸
۲.۴	نمودار میزان خطای برابر	۴۹
۳.۴	نمودار خطای برابر برای شبکه ALEXNET و تابع هزینه BCE	۵۱
۴.۴	نمودار خطای برابر هنگام استفاده از عملگر LBP پیشنهادی	۵۱
۵.۴	نمودار خطای برابر هنگام استفاده از شبکه efficient net B0	۵۲
۶.۴	نمودار خطای برابر هنگام استفاده از تابع هزینه ARCB پیشنهادی	۵۲
۷.۴	نمودار خطای برابر با استفاده از تابع هزینه مبتنی بر شناسه اشخاص	۵۳

فهرست جداول

۱.۴	خطای برابر روی دیتاستهای CASIA و MSU	۵۴
۲.۴	نرخ در پروتکل اول دیتاست SIW	۵۴
۳.۴	نرخ در پروتکل دوم دیتاست SIW	۵۵
۴.۴	دقت در پروتکلهای اول و دوم دیتاست OULU	۵۶
۵.۴	تایج روی آزمون بین دیتاست	۵۶

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

¹ Anti-spoofing for authentication based on face recognition

² Biometric

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؛ به‌گونه‌ای 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.



شکل ۱.۱: نمونه‌ای از تصاویر واقعی و تقلبی در حوزه چهره [۴]

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 دیتاست 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 دیتاست‌های 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 دیتاست کوچک تست شده است و اثر بخشی هر قسمت بررسی شده است. سپس تمام روش پیشنهادی روی دیتاست‌های 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

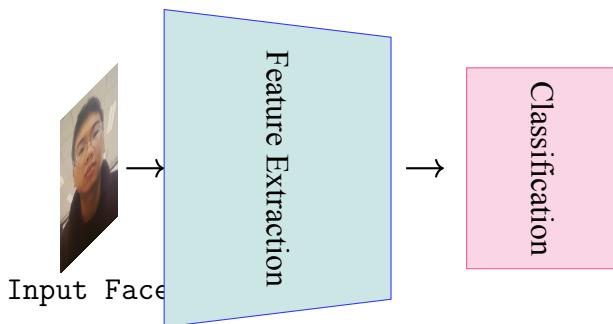
۱.۲ مقدمه

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 به طور کلی بسیاری از مسائل بینایی ماشین ابتدا استخراج ویژگی از تصویر یا ویدیوی ورودی است و سپس طبقه‌بندی ویژگی‌های به دست 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 عمدۀ 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. [۳۳]

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 [۳۳].

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 [۵، ۲۹]. همچنین از روش‌های استخراج ویژگی نظری SIFT [۳۱] و SURF [۳] استفاده شده است. یک متدهای کلاسیک دیگر در دسته روش‌های ایستان استفاده از هیستوگرام گرادیان‌های جهت دار^۳ (HoG) است [۴۹، ۳۷].

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 [۵۱، ۱]. 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 [۴۲، ۷]. 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 [۱۱].

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

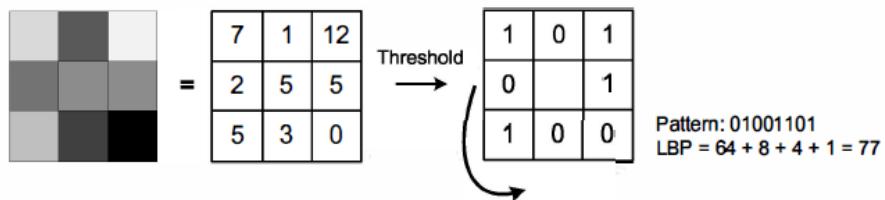
³Histogram of oriented gradients

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.

[۲۹] 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 ۲×۲ 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 [۲۹]

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 خطی⁶، 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.

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

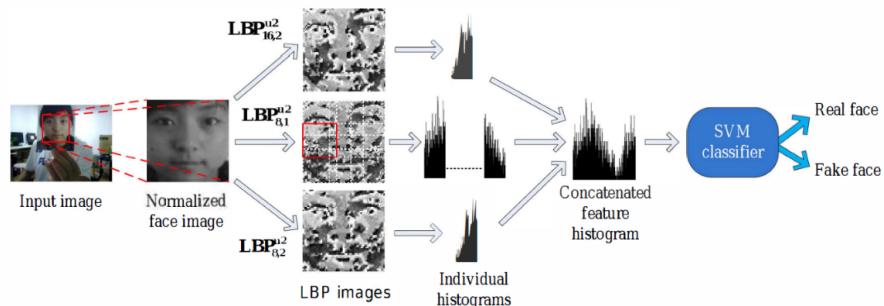
$$I \rightarrow I + \beta \rightarrow s((I_p + \beta) - (I_c + \beta)) = s(I_p - I_c) \quad (3.2)$$

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.

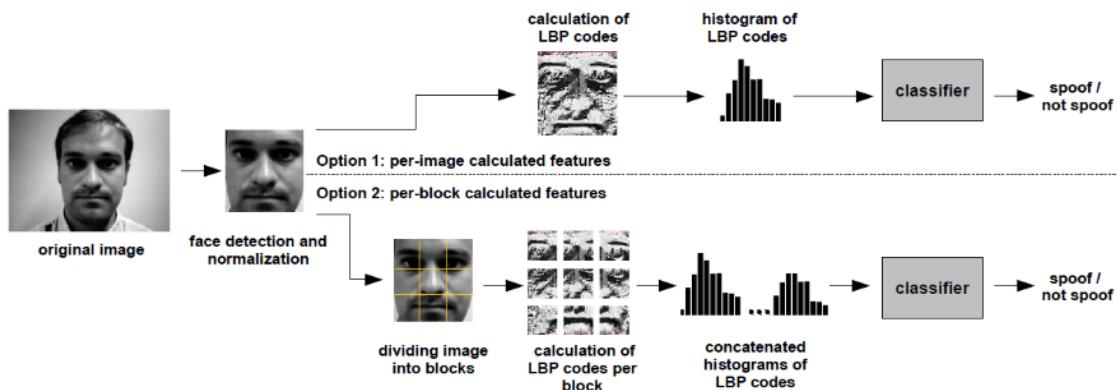


[۲۹]: روش تصمیم‌گیری بر اساس استفاده از LBP

روش محاسبه و تصمیم‌گیری ارائه شده در [۲۹] در مورد واقعی یا تقلیبی بودن تصویر چهره با استفاده از تحلیل ریزبافت به صورت شکل ۳.۲ است. ابتدا با استفاده از الگوریتم تشخیص چهره، مختصات صورت انتخاب شده و مقادیر پیکسلی چهره به صورت 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 [۵] بر خلاف روش قبلی تنها از عملگر LBP یکنواخت در پنجره سه در سه به صورت نرمالیزه شده استفاده شده است و از عملگر LBP با شعاع‌های متنوع [۲۹] استفاده نشده است. همچنین در [۵] 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 به صورت جداگانه هیستوگرام LBP محاسبه می‌شود و این هیستوگرام‌ها در کنار هم قرار داده می‌شود. سپس هیستوگرام‌ها به عنوان 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 به صورت 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 [۱۱].



شکل ۴.۲: روش تحلیل ریزبافت در نواحی مختلف تصویر [۵]

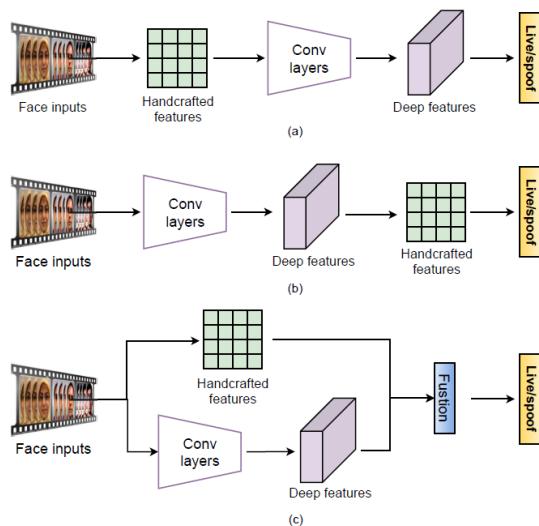
۳.۲ روش‌های مبتنی بر یادگیری عمیق

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 [۴۸]. روش ارائه شده در این کار بدين 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 [۲۲] 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 [۱۶، ۲۳، ۱۲]. شیوه دیگر برای کمک گرفتن اطلاعات فریم‌های متوالی استفاده از ساختار LSTM [۱۶] پس از شبکه کانولوشن است که کارهای [۴۷، ۵۰] 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 [۵۳]. 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.



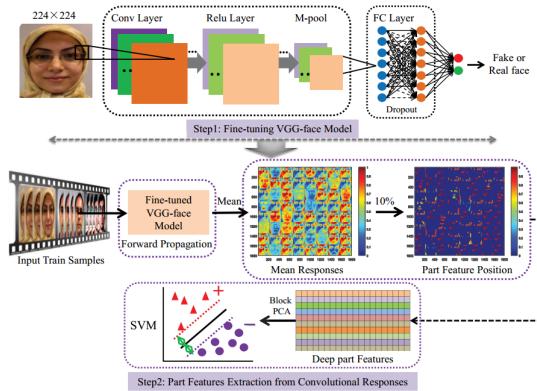
شکل ۵.۲: حالت‌های مختلف ترکیب ویژگی‌های دستی و ویژگی‌های شبکه عمیق [۵۳]

برای مثال فنگ و همکاران [۲۵] 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 [۳۰] 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.

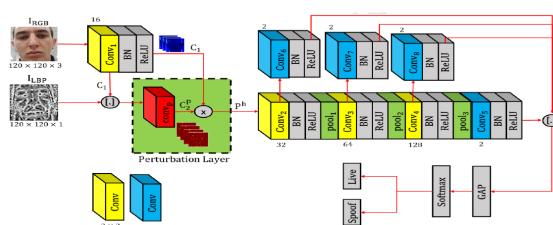


شکل ۶.۲: استفاده از شبکه تنظیم دقیق شده و اعمال PCA روی ویژگی‌های عمیق [۲۵]

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 [۲۴].

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 [۳۴]. این ایده در شکل ۷.۲ نشان داده شده است.

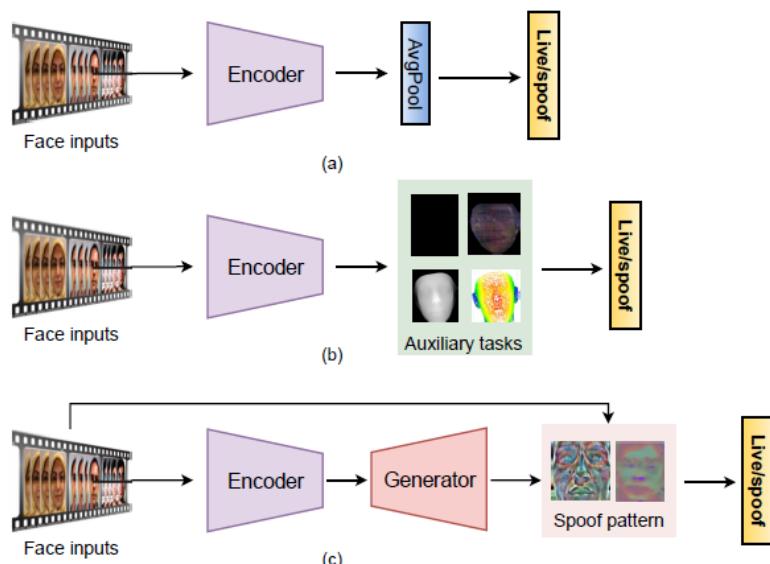
روش‌های ترکیبی بین ویژگی‌های دستی و ویژگی‌های یادگیری عمیق دارای یک قسمت ایستا هستند که حین آموزش شبکه تغییری نخواهند کرد. برای روش‌هایی مبتنی بر شبکه عصبی مطلوب این است که تمامی قسمت‌های شبکه به صورت انتها به انتها یاد گرفته شود.



شکل ۷.۲: روش ترکیب LBP و کانولوشن [۳۴]

۲.۳.۲ استفاده از تخمین سیگنال کمکی

در روش‌های بیان شده روال آموزش شبکه عصبی بهینه کردنتابع هزینه آنتروپی متقاطع 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 ۲: روش‌های مختلف یادگیری عمیق در حوزه‌ی کشف تقلب چهره [۵۳]

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 نظارت کمکی^{۱۲} رسانده است [۲۷]. در روش‌های نظارت کمکی سعی می‌شود از تخمین یک مورد کمکی برای استنتاج تقلیلی یا واقعی بودن چهره استفاده شود. یکی از موارد مهم کمکی در این حوزه تخمین عمق صورت است.

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

⁹Binary cross entropy

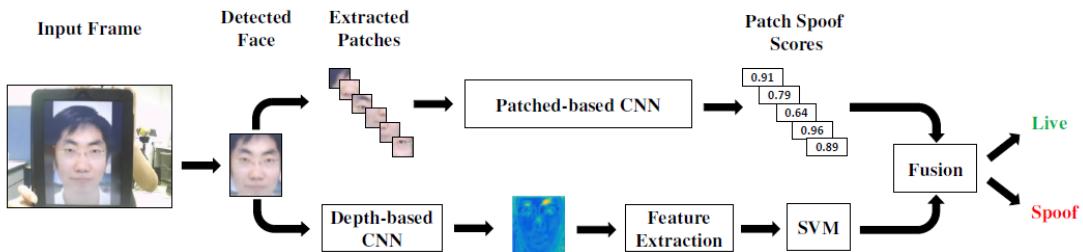
¹⁰Over fitting

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.

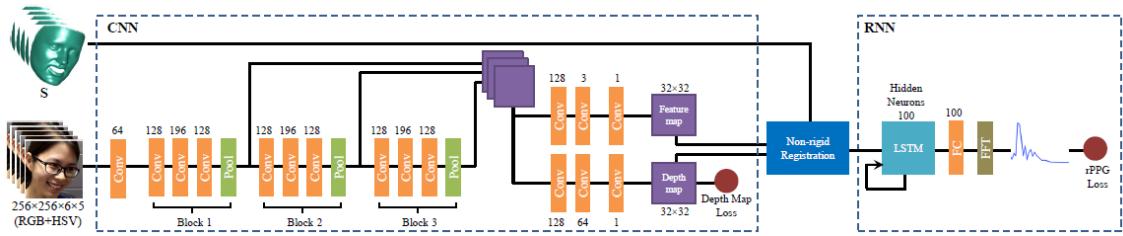
استفاده از عمق از این شهود گرفته شده است که مغز انسان چهره واقعی را دارای عمق می‌بیند، برای مثال بینی نزدیک‌تر از گونه‌ها است، اما چهره تقلبی که روی صفحه نمایش یا کاغذ چاپ شده قرار دارد دارای عمقی مسطح است. در روش‌هایی که از عمق به عنوان یک سیگنال کمکی استفاده کرده‌اند، پیش از آموزش شبکه‌ی کشف تقلب، از یک شبکه تخمین عمق مثل PRNet [۱۰] 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 [۲، ۳۸، ۴۳، ۴۴، ۵۶].



شکل ۹.۲: استفاده از عمق برای کشف تقلب در چهره [۲]

اتوم و همکاران [۲] 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 [۲۷] 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 [۲] ابتدا

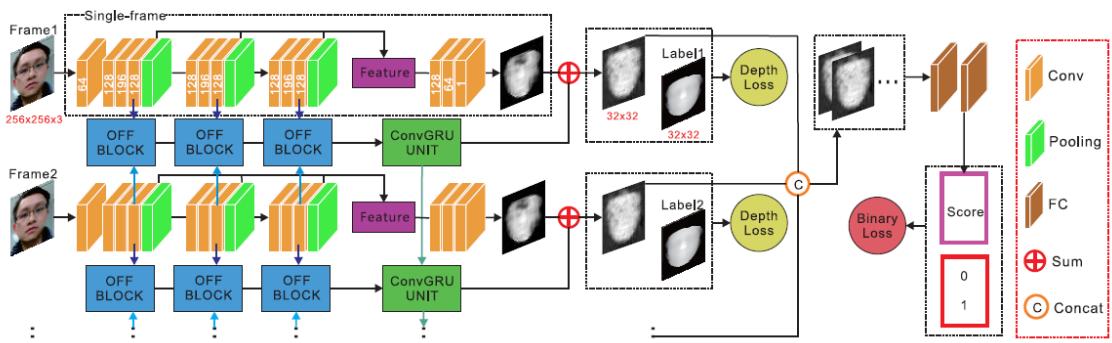


شکل ۱۰.۲: روش استفاده از عمق و تخمین rPPG

برچسب عمق واقعی برای چهره زنده و عمق صفر برای چهره تقلیلی 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 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 [۴۴] ساختاری را به کمک optical flow 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 [۶] 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.



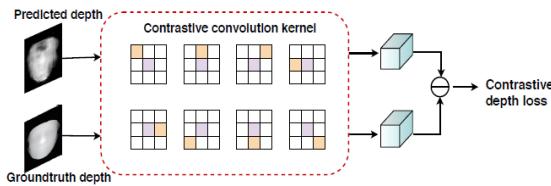
شکل ۱۱.۲: استفاده از ویژگی های عمیق در طول زمان [۴۴]

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

پیوستگی عمق در پیکسل‌های مجاور توسعه داده شده است [۴۳، ۴۴] 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 نشانگر عملگر کانولوشن است. در شکل ۱۲.۲ مربع بنفس متناظر با عدد ۱ و مربع زرد متناظر با عدد ۰ و مربع‌های سفید عدد ۰ را در هسته نشان می‌دهند. یو و همکاران [۵۶] ساختاری تغییر یافته از شبکه‌های کانولوشنی



شکل ۱۲.۲: نحوه محاسبه تابع هزینه CDL [۵۶]

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.

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

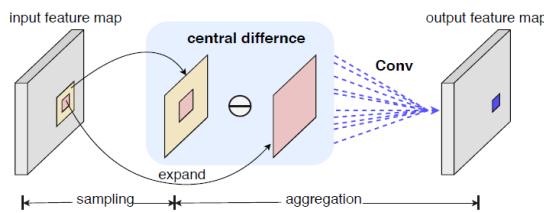
برای آنکه از خاصیت کانولوشن نیز استفاده شود ترکیب خطی رابطه ۶.۲ با رابطه کانولوشن حساب می‌گردد.

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

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.

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



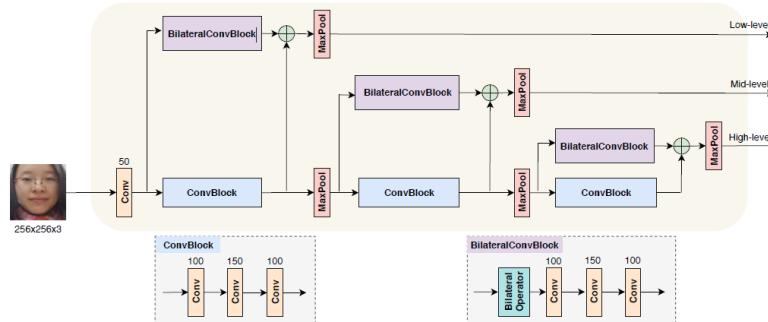
شکل ۱۳.۲: عملگر کانولوشن تغییر یافته [۵۶]

که همانطور که مشاهده می‌شود که 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 کمک استفاده می‌شود. همچنین برای یافتن اندازه‌ی شبکه از روش جستجوی معماری شبکه ۱ [۶۱] 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 [۶۱] کارهای [۵۴، ۵۵] 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 سه بعدی به عنوان سیگنال کمکی استفاده کرده‌اند و ساختاری به نام 3DPC-NET پیشنهاد کرده‌اند [۲۶].

یو و همکاران [۵۲] 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 صفحه‌ی 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



شکل ۱۴.۲: روش استفاده از فیلتر bilateral در شبکه عمیق [۵۲]

هر پیکسل به مختصات p و تصویر I به صورت رابطه ۱۰.۲ تعریف می‌شود.

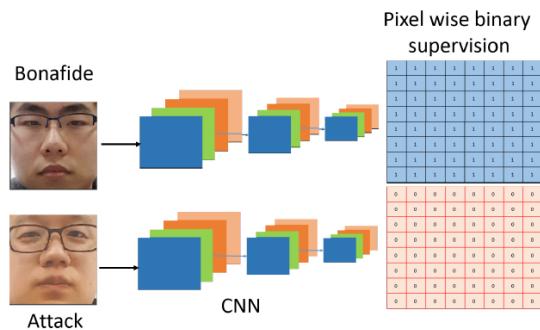
$$BiBase(I_p) = \frac{1}{k} \sum_{q \in I} g_{\sigma_s}(|p - q|) g_{\sigma_r}(|I_p - I_q|) I_q \quad (9.2)$$

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

که در آن $(g_{\sigma}(x) = \exp(\frac{-x^2}{\sigma^2})$ تابع گوسی است. در این روش ساختار شبکه مشابه [۲۷] است ولی روی ویژگی‌های کانولوشن این فیلتر اعمال شده است.

با وجود آن که سیگنال کمکی عمق در ادبیات موضوع به طور گسترده استفاده شده است اما پر هزینه است و نیاز به پردازش بیشتر برای تخمین عمق دارد. جدای از آن که 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 همه جا صفر نخواهد بود.

جرج و مارسل روشی را برای پیدا کردن ویژگی‌های خوش‌ساخت بدون استفاده از عمق پیشنهاد کردند [۱۳]. در این روش از چند لایه اول شبکه DENSNET [۱۷] برای نشان‌کردن تصویر ورودی به یک صفحه $14*14$ استفاده کردند. و قرارداد کردند 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.



شکل ۱۵.۲: تابع هزینه BCE روی یک صفحه مسطح به جای یک نورون [۱۳]

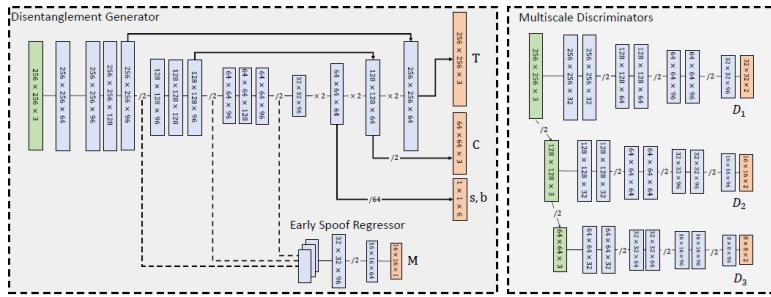
۳.۳.۲ استفاده از شبکه‌های مولد تهاجمی و تابع هزینه‌های مختلف

مسئله کشف تقلب در تشخیص 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 انسان زنده مدل کرد. جورابلو و همکاران [۱۹] برای اولین بار مسئله کشف تقلب در چهره از شبکه‌های مولد تهاجمی (GAN) [۱۵] را مدل کردند و یافتن نویز تصاویر تقلبی استفاده کردند. با تخمین نویز مربوط به کشف تقلب، قدرت استنتاج برای تقلبی بودن تصویر بیشتر خواهد شد.

از آنجا که نویز مربوط به تقلب می‌تواند در سطوح مختلف در تصویر وجود داشته باشد لیو و همکاران [۲۸] ساختاری بر پایه GAN که الگوهای تقلب در ابعاد مختلف تصویر را تخمین بزنند پیشنهاد داده‌اند. در این روش در شبکه مولد dismantlement 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 در GAN در نهایت ویژگی‌های تقلب بهتری تولید خواهد شد.

با وجود اینکه در دو پژوهش اخیر ذکر شده [۲۸، ۱۹] از شبکه مولد تهاجمی برای بهبود دقت در تست درون دیتاست استفاده شده است، توجه پژوهشگران به استفاده از GAN برای تعمیم‌پذیری مدل در دیتاست‌های مختلف جلب شده است [۳۸، ۱۸].

تعمیم‌پذیری مدل در دیتاست‌های مختلف بین معناست که برای مثال از بین چهار دیتاست مختلف، سه دیتاست برای آموزش شبکه استفاده می‌گردد و مدل آموزش داده شده روی دیتاست چهارم آزمایش می‌شود. از آنجا که دیتاست‌های مختلف توزیع‌های متفاوتی دارند، رسیدن به دقت



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

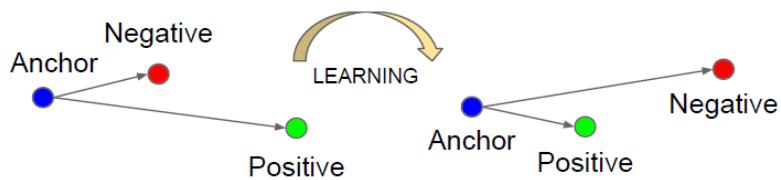
خوب در تست روی دیتاست دیده نشده (که توزیع لزوماً یکسانی با توزیع دیتاستهایی که برای آموزش استفاده شده است ندارد) یک چالش جدی در این حوزه است.

همچنین یک روش برای بهبود قابلیت تعمیم‌پذیری، استفاده از تابع هزینه سه‌گانه^۱ [۳۶] است. در تابع هزینه سه‌گانه هدف این است که استخراج ویژگی به نحوی انجام شود که فاصله ویژگی‌های نمونه‌های مربوط به یک کلاس کوچک و فاصله بین نمونه‌های مربوط به کلاس‌های مختلف زیاد شود. فرض کنید خروجی شبکه استخراج ویژگی بردار باشد. در این صورت برای تشکیل تابع هزینه سه‌گانه لازم است که از خروجی‌های شبکه استخراج ویژگی، یک بردار ویژگی لنگر، یک بردار ویژگی با برچسب یکسان با لنگر و یک بردار ویژگی با برچسب متفاوت با لنگر انتخاب شود. تابع هزینه سه‌گانه به صورت رابطه ۱۰.۲ تعریف می‌شود. که در آن 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 است درون مجموع گیری قرار می‌گیرد.

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

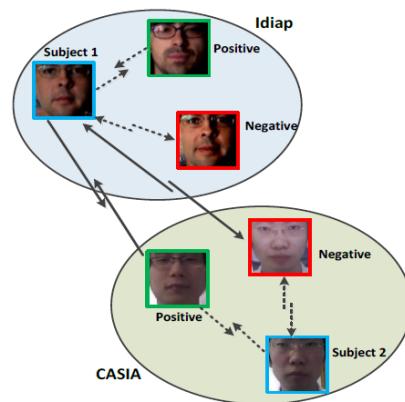
تابع هزینه سه‌گانه به صورت رابطه ۱۱.۲ نیز قابل بیان است. که در آن زمانی که فاصله درون کلاسی کوچکتر از فاصله برون کلاسی به میزان سطح آستانه باشد حاصل \max صفر خواهد بود و در محاسبات تابع هزینه نقش نخواهد داشت.

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



شکل ۱۷.۲: نحوه عملکرد تابع هزینه سه‌گانه روی فاصله بردارهای ویژگی [۳۶]

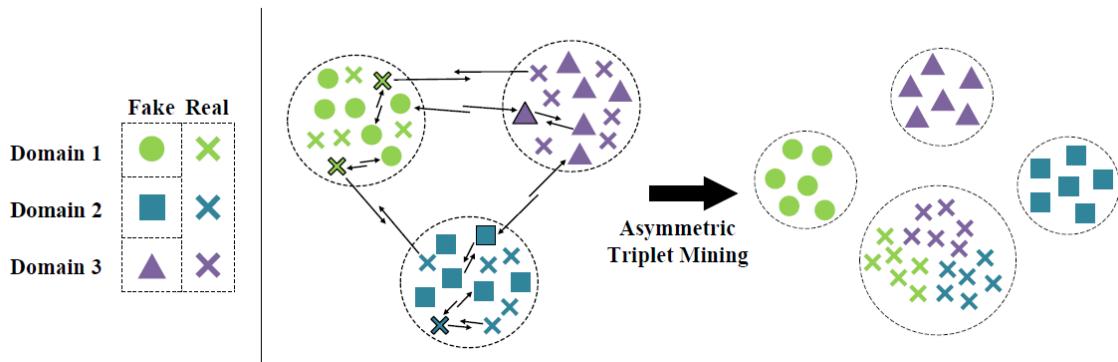
شائو و همکاران [۳۸] از ساختار GAN و ابزار کمکی تخمین عمق و تابع هزینه سه‌گانه برای بهبود تعمیم‌پذیری استفاده کرده‌اند. در این کار یک تابع هزینه بر مبنای تابع هزینه سه‌گانه توسعه داده شده است که فاصله بین نمونه‌ها با برچسب یکسان در دیتاست‌های مختلف را کوچک‌تر کند و فاصله نمونه‌ها با برچسب متفاوت در یک 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.



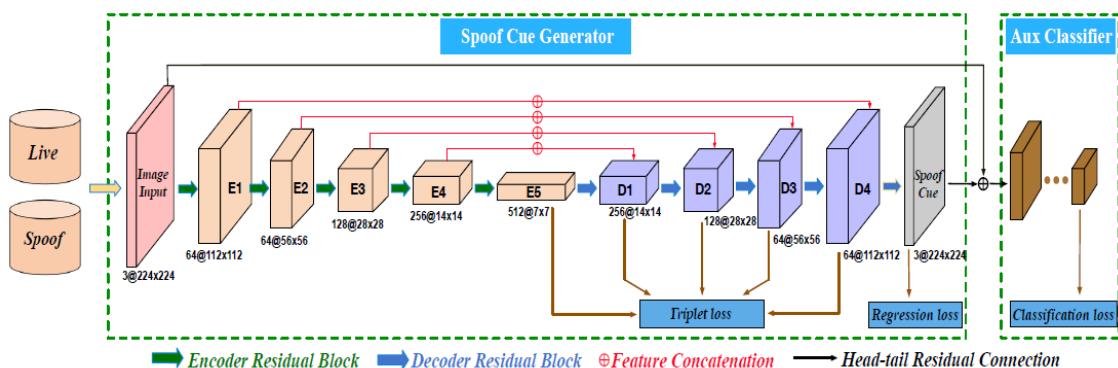
شکل ۱۸.۲: نحوه اثر تابع هزینه روی فاصله نمونه‌ها در دیتاست‌های مختلف [۳۸]

همچنین جیا و همکاران [۱۸] علاوه بر استفاده از GAN صورتی نامتقارنی از تابع هزینه سه‌گانه را پیشنهاد کرده‌اند. به گونه‌ای که نمونه‌های زنده در دیتاست‌های مختلف به یکدیگر نزدیک‌تر شوند و نمونه‌های تقلیبی در دیتاست‌های مختلف از یک دیگر دورتر شده و نمونه‌های واقعی از نمونه‌های تقلیبی با فاصله باشند.

فنگ و همکاران [۹] یک ساختار U-Net [۳۵] به کار برده‌اند و در میان لایه‌های آخر شبکه تولید کننده الگوهای تقلب از تابع هزینه سه‌گانه استفاده کرده‌اند و خروجی این شبکه U-Net را به یک شبکه طبقه بند کمکی داده‌اند.



شکل ۱۹.۲:تابع هزینه نامتقارن برای کاهش فاصله نمونه‌های از یک کلاس [۱۸]



شکل ۲۰.۲: ساختار ۰۰۰-۰-۰ و تابع هزینه سه‌گانه [۹]

پرزکابو و همکاران [۳۲] تابع هزینه سه‌گانه را در فضای نمایی به کار بردند که در رابطه ۱۲.۲ نشان داده شده است.

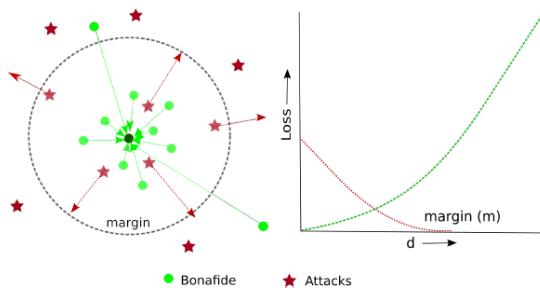
$$L_{tf} = \sum_i \max(0, e^{\frac{D_{a,p}}{\sigma}} - e^{\frac{D_{a,n}}{\sigma}} + \alpha) \quad (12.2)$$

که در آن $D_{a,p}$ فاصله درون‌کلاسی و $D_{a,n}$ فاصله بروん‌کلاسی است و σ یک هایپر پارامتر است. جرج و مارسل [۱۴] 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 نشان داده شود و فاصله بردار ویژگی نمونه \mathbf{x} با مرکز با تعریف شود. در این صورت تابع هزینه

تعریف شده به صورت رابطه ۱۳.۲ است.

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

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.

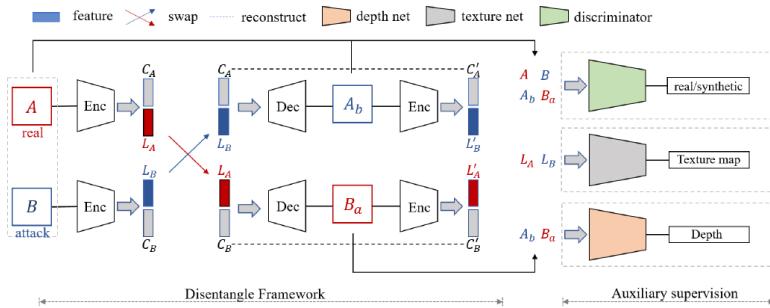


شکل ۲۱.۲: کاهش فاصله نمونه‌های واقعی تا مرکز و افزایش فاصله نمونه‌های تقلبی تا مرکز [۱۴]

تو و همکاران [۴۱] نیز شبکه VGG-face را به صورت همزمان با دو هدف شناسایی چهره و تشخیص تقلب آموزش داده‌اند و یکتابع هزینه معرفی کردند که هدف آن منظم‌سازی ۱ و جلوگیری از بیش برآذش شبکه است. در این تابع فاصله بین هر دو جفت نمونه داده‌ها مستقل از آنکه برچسب آنچه 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 تمام جفت نمونه‌های موجود در دسته آموزش است.

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

ژنگ و همکاران [۵۸] 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 LBP باید معادل 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. ژو و همکاران [۴۶] روی ثبات فضای ویژگی در بین فریم‌های متوالی یک ویدئو تأکید کرده‌اند.



شکل ۲۲.۲: استفاده از LBP در کنار عمق برای یافتن ویژگی‌های خوش ساخت [۵۸]

در این کار به جای استفاده از الگوریتم‌های تشخیص چهره در هر فریم از الگوریتم دنبال‌کننده‌ی چهره استفاده کرده و چهره‌های تخمین زده شده در فریم‌های متوالی را به شبکه تشخیص تقلب داده‌اند. برای این شبکهتابع هزینه‌ای ارائه معرفی کرده‌اند که فاصله بین بردارهای ویژگی یک ویدئو در دیتاست را کوچک‌تر کند.

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

که در آن $\|\cdot\|$ 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 کرده‌اند که فاصله بین بردارهای ویژگی متعلق به یک برچسب واقعی را نیز کوچک‌تر کند.

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

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.

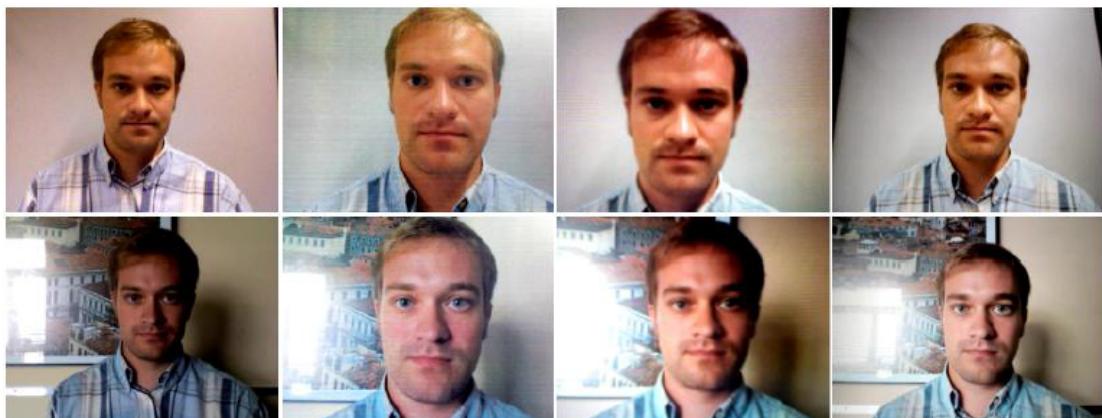
۴.۲ دیتاست‌های مورد استفاده

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 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 دیتاست Replay

دیتاست Replay شامل ویدیوهای از ۵۰ شخص مختلف با نمونه‌های واقعی و تقلبی است [۵] نمونه‌های واقعی در شرایط محیطی نوری کنترل شده با پس زمینه یکنواخت و شرایط محیطی با نور کنترل نشده با پس زمینه غیر یکنواخت گرفته شده‌اند. برای نمونه‌های تقلبی از صفحه کاغذ چاپ شده، 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 به منظور 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 240×320 پیکسل است. در ۲۳.۲ نمونه‌هایی از این



[۵]: نمونه‌هایی از دیتاست Replay

دیتاست نشان داده شده است. در سطر بالایی نمونه‌ها در محیط کنترل شده از نظر نورپردازی و پس زمینه 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 نیز از ۵۰ شخص مختلف نمونه‌های واقعی و تقلبی گرفته شده است [۵۹]. همچنین تصویربرداری با سه نوع دوربین مختلف برای پوشش دادن حالت‌های مختلف در رزوولوشن‌های 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 نمونه‌های واقعی دیتاست از کاربر خواسته شده است که پلک و لب بزند تا ویدیوهای ضبط شده دارای اطلاعات حرکتی صورت باشند. در نمونه حمله‌های تقلبی قسمت چشم‌های صورت بریده شده است تا کاربر با پلک زدن بتواند در نمونه‌های تقلبی اطلاعات حرکت به ویدئو بدهد. همچنین در نمونه‌هایی که کاغذ بریده نشده است از کاربر خواسته شده که با حرکت دست کاغذ چاپ شده را حرکت بدهد. نمونه‌هایی از دیتاست CASIA در شکل ۲۴.۲ نشان داده شده است.



شکل ۲۴.۲: نمونه‌هایی از دیتاست CASIA [۵۹]

۳.۴.۲ دیتاست MSU

در دیتاست MSU [۴۵] از ۵۵ شخص تصویربرداری شده است که ویدیوهای متعلق به ۳۵ فرد در دسترس قرار داده شده است. برای تصویربرداری از دوربین لپ‌تاپ و دوربین تلفن همراه استفاده شده است که دارای رزوولوشن ۶۴۰*۴۸۰ و ۷۲۰*۴۸۰ هستند. استفاده از این نوع دوربین‌ها به منظور شبیه‌سازی سناریو احراز هویت از طریق تلفن همراه و لپ‌تاپ 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.

۴.۴.۲ دیتاست OULU

در دیتاست OULU [۴] از ۵۵ شخص مختلف برای تصویربرداری نمونه‌های واقعی و تقلبی استفاده شده است. تصویربرداری در سه نشست مختلف، با شش تلفن همراه جدید در زمان جمع‌آوری



شکل ۲۵.۲: نمونه‌هایی از دیتاست MSU [۴۵]

دیتاست استفاده شده است که باعث تنوع در کیفیت تصویر و محیط پس زمینه شده است. برای حمله کاغذ چاپ شده از دو نوع چاپگر با کیفیت و برای حمله باز پخش از یک نمایشگر و صفحه نمایش لپ‌تاپ استفاده است. ویدیوهای ضبط شده با کیفیت full HD با رزولوشن 1920×1080 گرفته شده است.

این دیتاست در مقایسه با دیتاست‌های قبلی تنوع بیشتر و کیفیت بالاتری دارد که باعث چالشی شدن دیتاست شده است. نمونه‌های واقعی در دیتاست 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 باز پخش را نشان می‌دهد.



شکل ۲۶.۲: نمونه‌های واقعی در دیتاست OULU [۴]

همچنین در این دیتاست برای گزارش دقیق چهار پروتکل مختلف به منظور ارزیابی قابلیت تعمیم‌پذیری مدل ارائه شده در حالت‌های مختلف 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 بینند و ارزیابی روی نشست سوم خواهد بود. در پروتکل دوم روی یک نوع از حمله کاغذ چاپ شده و یک



شکل ۲۷.۲: نمونه‌های تقلبی در دیتابست OULU [۴]

نوع حمله‌ی بازپخش آموزش انجام می‌شود و در ارزیابی از نوع دیگر حمله باز پخش و کاغذ چاپ شده استفاده می‌شود تا تعیین پذیری مدل روی حمله از ابزار دیده نشده بررسی شود. در پروتوكل 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 تعیین پذیری روی نوع سنسور تصویر برداری دیده نشده انجام می‌گردد. پروتکل چهارم هر سه نوع پرتکل 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 [۲۷] از ۱۶۵ شخص مختلف برای تصویربرداری استفاده شده است. برای تصویربرداری از دو نوع دوربین با کیفیت استفاده شده است. در نمونه‌های واقعی تصویر برداری 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.

همچنین شرایط نورپردازی مختلف در این دیتابست 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 گرایانه برای حمله بازپخش استفاده شده است. نمونه‌های این دیتابست در شکل ۲۸.۲ قابل مشاهده است.



شکل ۲۸.۲: نمونه‌های از دیتابست [۲۷] SIW

این دیتابست 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 نظریه 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 (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 نظری 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 تقلیبی 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.

۳.۳ عملگر تحلیل ریزبافت قابل آموزش

رابطه عملگر کانولوشن و تصویر ورودی به صورت رابطه ۱.۳ 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 ۲.۳ 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.

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

بهمنظور آنکه 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 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.

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

این عملگر در مقایسه با عملگر کانولوشن نگاه ریزتری به بافت تصویر خواهد داشت. از آن‌جا که

در کانولوشن تمامی پیکسل‌های همسایگی در وزن‌های فیلتر ضرب شده و حاصل جمع آنها در 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 آن‌جا 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 $(\cdot)^{\sigma}$ یک تابع غیر خطی است که نقشی مشابه تابع فعالسازی ۱ 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 حاصل جمع است در حالی که در کانولوشن‌های شبکه عصبی تابع غیر خطی بیرون عملگر حاصل جمع قرار دارد. هر چند که تعریف کلاسیک برای عملگر LBP استفاده از تابع Heaviside است اما از توابع غیر خطی دیگری نظیر Relu و Sign نیز می‌توان استفاده کرد.

از آنجا که عملگر 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 (۱.۱.۳.۳) خواهد بود. تصویر ورودی به صورت سه کanal رنگی وارد عملگر می‌شود و خروجی آن به یک شبکه متشکل از لایه‌های کانولوشن داده می‌شود که در این پژوهش از شبکه EfficientNet B0 [۴۰] استفاده شده است. و خروجی آن یک بردار مسطح 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 برحسب مقدار این نورون و انتخاب یک سطح آستانه طبقه‌بندی دو کلاسه صورت خواهد گرفت. تابع هزینه متدائل در شبکه عصبی برای طبقه‌بندی دو کلاسه تابع آنتروپی متقاطع دودویی (BCE) است. اما تحقیقات پیشین در حوزه کشف تقلب نشان داده است که این تابع هزینه به تنها یک مؤثر واقع خواهد شد. به همین منظور تابع هزینه جدیدی برای طبقه‌بندی معروفی می‌گردد که یک حاشیه امن برای طبقه‌بندی ایجاد کند که باعث افزایش قابلیت تعمیم‌پذیری شبکه خواهد شد.

۴.۳ تابع هزینه ARCB

در شبکه‌های عصبی زمانی که خروجی یک طبقه‌بندی چند کلاسه (بیشتر از دو) باشد از تابع فعالسازی سافت‌مکس ۱ در لایه‌ی آخر استفاده می‌شود و در طبقه‌بندی دو کلاسه از تابع فعالسازی سیگموید ۲ استفاده می‌شود. دنگ و همکاران در حوزه تشخیص چهره^۳ که یک طبقه‌بندی چند

کلاسه است تابع هزینه آنتروپی متقاطع ۴ (CE) را به فضای کسینوسی برده‌اند و یک حاشیه به تابع هزینه در این فضا اضافه کرده‌اند [۸].

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 فضای بین یک و صفر استفاده خواهد شد. در تصمیم‌گیری دو کلاسه رابطه تابع هزینه آنتروپی متقاطع دودویی به صورت رابطه ۵.۳ است.

$$L_{BCE} = -y_i \log P(y_i) - (1 - y_i) \log (1 - P(y_i)) \quad (5.3)$$

که در آن y_i برچسب صحیح متناظر با بردار ویژگی است. و 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 ۷.۳ است.

$$P(y_i) = \text{sigmoid}(W^T X_i + b) \quad (6.3)$$

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.

$$W^T X_i = |W^T| |X_i| \cos \theta_i = s \cos \theta_i \quad (7.3)$$

حال با جاگذاری این مقدار در تابع هزینه BCE به صورت رابطه ۸.۲ بازنویسی می‌شود.

$$L_{BCE} = -y_i \log \frac{1}{1 + e^{-s \cos \theta_i}} - (1 - y_i) \log \left(1 - \frac{1}{1 + e^{-s \cos \theta_i}}\right) \quad (8.3)$$

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

حالت اول. زمانی که برچسب یک باشد در این صورت تنها عبارت اول در رابطه ۸.۳ 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.

$$y_i = 1 \rightarrow \theta_i = \theta_i + m \quad (9.3)$$

حالت دوم. زمانی که مقدار برچسب واقعی صفر باشد در این صورت عبارت دوم در رابطه ۸.۳ ظاهر می‌گردد. بدین منظور لازم است که عبارت داخل لگاریتم بیشینه شود که معادل این است که زاویه بین وزن و بردار ویژگی به مقدار نزدیک شود. برای آنکه بهینه‌سازی با حاشیه انجام شود یک مقدار ثابت حاشیه m از زاویه بین دو بردار کم می‌شود.

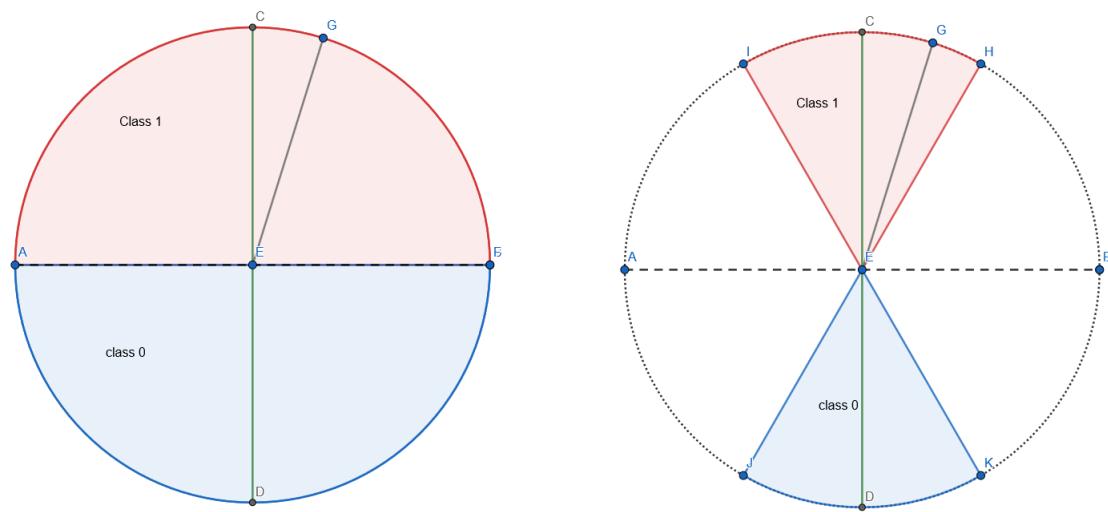
$$y_i = 0 \rightarrow \theta_i = \theta_i - m \quad (10.3)$$

با جایگذاری زاویه‌های حاشیه دار شده در روابط ۹.۳ و ۱۰.۳ در رابطه تابع هزینه BCE بازنویسی شده در فضای کسینوسی ۸.۳ تابع هزینه ARCB به صورت رابطه ۱۱.۳ به دست خواهد آمد.

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

این تابع هزینه نه تنها ویژگی‌های بین دو کلاس را جدا می‌کند بلکه یک حاشیه به اندازه 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 تعمیم، پذیری شبکه بیشتر شود. در شکل ۱.۳ 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 ۱.۳ در سمت چپ بهینه نتیجه جداسازی بردارهای ویژگی در حالت استفاده ازتابع 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 ۱.۳: مقایسه تابع هزینه BCE کلاسیک با نسخه‌ی حاشیه‌دار

۵.۳ تابع هزینه بر اساس شناسه‌ی شخص

در دیتاستهای موجود در حوزه کشف تقلب در تشخیص چهره، برای هر فرد چند نمونه 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 ویدئو 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 چهره 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 جریمه 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 $X_k^l \in R^d$, $I \in \{0, 1\}$, $K \in \{1, 2, \dots, M\}$ باشد. در حین آموزش در هر گام تعداد دسته ۱ 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 $k_1 \neq k_2$, $l_1 \neq l_2$. در این حالت با توجه به اینکه مشخصه‌های ظاهری فرد که عده 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 تقلیبی است.

این حالت در شکل ۲.۳ نشان داده شده است. در این شکل تصویر بالایی یک تصویر زنده و تصویر دومی تقلیبی 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.

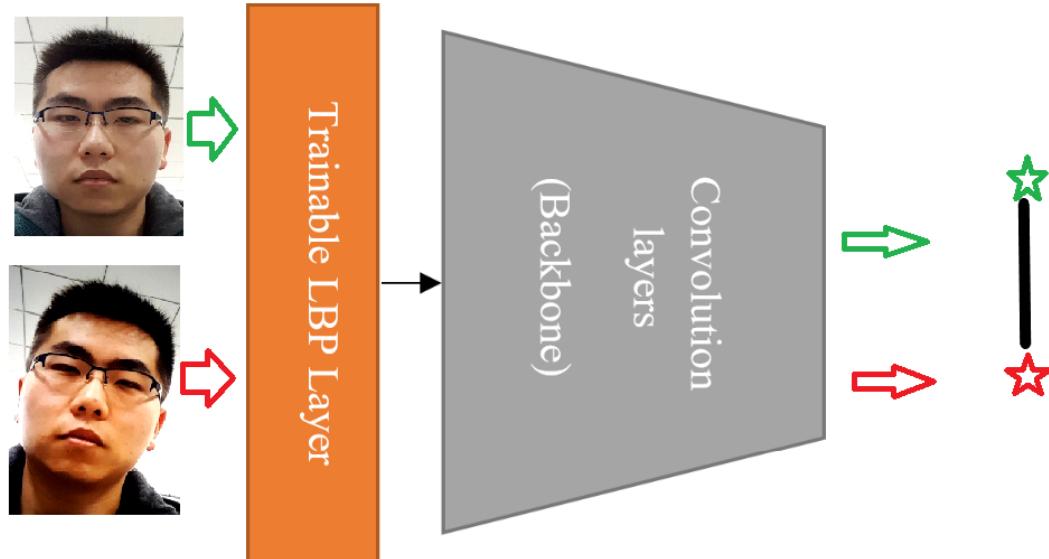
پس در حالت اول هدف شبکه به صورت رابطه ۱۲.۳ است.

$$\max_{\Theta} d(X_{k_1}^{l_1}, X_{k_2}^{l_2}) = \min_{\Theta} \max(0, M - d(X_{k_1}^{l_1}, X_{k_2}^{l_2})) \quad (12.3)$$

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 ۱۳.۳ تعریف می‌شود.

$$d(X_1, X_2) = \left\| \frac{X_1}{\|X_1\|} - \frac{X_2}{\|X_2\|} \right\| \quad (13.3)$$

از آنجا که باید در بهینه سازی 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



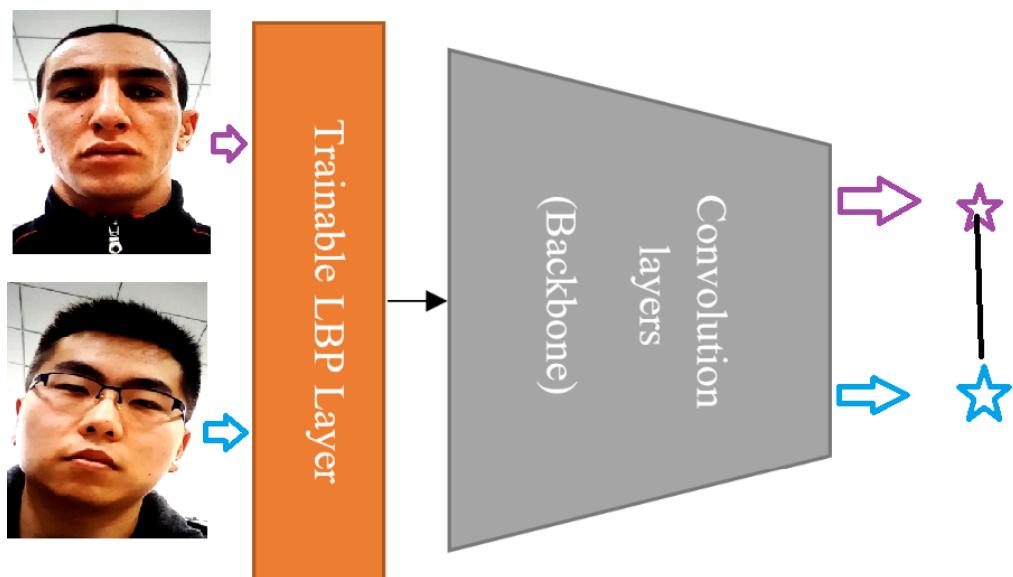
شکل ۲.۳: حالتی که دو نمونه متعلق به یک شخص ولی یکی واقعی و دیگری تقلبی است

و در صورتی که کمتر باشد میزان فاصله تا این مقدار M به عنوان مقدار هزینه خواهد بود. با توجه به نامساوی رابطه ۱۴.۳ 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.

$$\left\| \frac{X_1}{\|X_1\|} - \frac{X_2}{\|X_2\|} \right\| \leq \left\| \frac{X_1}{\|X_1\|} \right\| + \left\| \frac{X_2}{\|X_2\|} \right\| \rightarrow d(X_1, X_2) \leq 2 \quad (14.3)$$

حالت دوم زمانی که دو بردار ویژگی در جفت دارای یک برچسب 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 ۳.۳ این حالت نشان داده شده است. در این مثال دو تصویر ورودی هر دو از نوع تقلیب هستند ولی متعلق به اشخاص مختلفی هستند. در این شکل ستاره نشان گر موقعیت بردار ویژگی متناظر با این دو ورودی در فضای \mathbb{H} بعدی است. از آنجا که دو فرد ویژگی های ظاهری متفاوتی دارند ممکن است فاصله بردارهای ویژگی متناظر با آنها فاصله محسوسی داشته باشد. در این حالت کم کردن این فاصله مد نظر است. پس به بیان ریاضی در این حالت تابع هزینه به صورت رابطه ۱۵.۳



شکل ۳.۳: حالتی که دو نمونه متعلق به اشخاص مختلف ولی برچسب یکسان هستند

خواهد بود.

$$\min_{\Theta} d(X_{k_1}^{l_1}, X_{k_2}^{l_2}) \quad (15.3)$$

و در نهایت تابع هزینه بر اساس شناسه اشخاص موجود در دیتابست به صورت رابطه ۱۶.۳ 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 سمپل با ویژگی برچسب متفاوت و شناسه یکسان است.

$$L_{PiD} = \sum_{l_1 \neq l_2, k_1 \neq k_2} \frac{1}{N_i} d(X_{k_1}^l, X_{k_2}^l) + \frac{1}{N_j} \max(0, M - d(X_k^{l_1}, X_k^{l_2})) \quad (16.3)$$

نحوه تشکیل اینتابع 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-برچسب متفاوت دارند انتخاب شده و فاصله اقلیدسی آن‌ها در رابطه ۱۶.۳ 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. در نهایت تابع هزینه کلی برای آموزش شبکه به صورت رابطه ۱۷.۳ خواهد بود. که در آن λ_1 و λ_2 هایپر پارامتر هستند که میزان تأکید بر هر کدام را نشان خواهند داد.

$$L_{overall} = \lambda_1 L_{ArcB} + \lambda_2 L_{PiD} \quad (17.3)$$

۶.۳ مقایسه‌ی روش پیشنهادی با پژوهش‌های قبلی

روش LBP قابل آموزش در این پژوهش، در مقایسه با روش‌هایی که از ترکیب کانولوشن و عملگر LBP استفاده کرده‌اند [۱۰، ۳۱] از این نظر متفاوت است که در متدهای 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 خواهد شد. در [۷] از ایده عملگر LBP قابل آموزش برای کاهش تعداد وزن‌های شبکه استفاده شده است. در واقع متدهای [۷] روی خاصیت تنک بودن ۱ 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 قابل آموزش را در شبکه کاهش داده است و نشان داده است که سرعت اجرای شبکه بهبود می‌یابد و دقیق شدن افکار کمی خواهد کرد. متدهای 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 [۷] به صورت رابطه (۲۰.۳) است.

در آن وزن‌های ثابت به صورت تنک هستند و پارامترهای قابل آموزش است. در [۸] نیز عملگر کانولوشن با الهام از عملگر LBP تغییر داده شده است به گونه که در رابطه نهایی وزن متفاوتی به

فصل ۳: روش پیشنهادی

جلوگیری از تقلب برای احراز هویت مبتنی بر تشخیص چهره

پیکسل مرکزی پنجره کانولوشن داده می‌شود و اعمال تابع غیرخطی بیرون مجموع گیری است. که به کلی از نظر فرمول بندی با عملگر ارائه شده در این پایان‌نامه متفاوت است.

٤ فصل

نتایج

۱.۴ مقدمه

در این فصل ابتدا ملاحظات پیاده‌سازی روش پیشنهادی بیان می‌شود. سپس معیارهای ارزیابی که در پژوهش‌ها برای توصیف میزان دقت شبکه وجود دارد تعریف می‌گردد. و در ادامه ابتدا هر قسمت از روش‌های پیشنهادی 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 بهدست آمده با دقت روش‌های موجود در این حوزه مقایسه شود.

۲.۴ ملاحظات پیاده‌سازی

در این پایان‌نامه از زبان برنامه‌نویسی پایتون و کتابخانه Pytorch 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 آماده وجود ندارد؛ اما توسط جریان محاسباتی Pytorch قابل پیاده‌سازی 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 قبل 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 به روز رسانی این پارامترها خواهد شد.

۲.۲.۴ پیاده سازیتابع هزینه

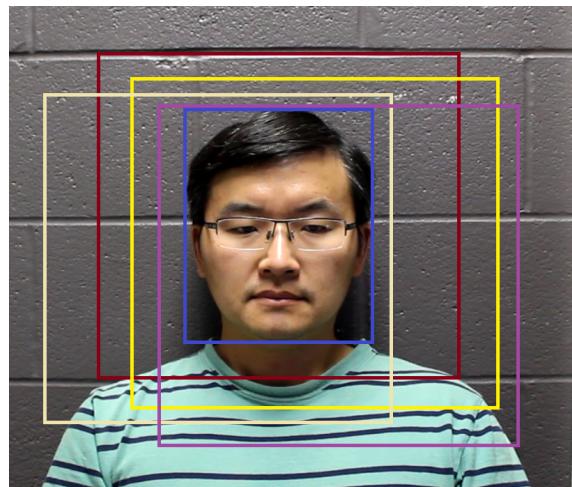
در هر بار forward داده ها به شبکه پس از بلوک استخراج ویژگی یک بردار به دست خواهد آمد که لازم است این بردار در هر مرحله برای استفاده در دوتابع هزینه معرفی شده نرمالایز شوند. در حین تست شبکه از آنجا که تغییری در وزن ها رخ نخواهد داد یک بار نرمال سازی کافی خواهد بود. پیاده سازی تابع ARCB با استفاده از توابع Pytorch برای پایدار بودن محاسبات انجام شده است. به منظور جلوگیری از بیش برازش داده ها از drop out [۳۹] در لایه آخر پس از نرمالایز کردن بردار ویژگی و پیش از طبقه بند استفاده شده است. 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 یافت ن عدد شناسه از روی نام فایل ویدئو قابل تشخیص است. برای بهینه سازی شبکه از الگوریتم آدام [۲۰] 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 خود کار 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 می توان در هر فریم ویدئو، قسمت مربوط به چهره را برش زد و به جای استفاده از کل فریم تنها قسمت چهره به همراه کمی از قسمت پس زمینه تصویر به عنوان ورودی به شبکه داده شود. در دیتاست هایی که این فایل مختصات وجود ندارد با استفاده از روش MTCCN [۵۷] چهره فریم ها پیدا شده و در یک فایل متنی ذخیره شده است. دیتاست های معرفی شده همگی به صورت ویدئو هستند. از آنجا که روش ارائه شده روی تک تصویر کار می کند یکی از

نکات عملی در خصوص آموزش روی داده‌های ویدیویی، نحوه آماده سازی داده‌ها برای آموزش است. یک روش تبدیل ویدئو 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 ویدئو از هر کلاس وجود داشته باشد. از طرفی برای آنکه تابع هزینه مبتنی بر شناسه اشخاص به درستی عمل کند لازم است که پراکنده‌گی ویدیوها 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 پیاده‌سازی روش این پایان‌نامه ابتدا به تعداد دسته، ویدئو در حافظه RAM بارگذاری خواهد شد و در هر مرحله یک فریم به صورت رندوم از هر ویدئو انتخاب داده می‌شود که در نهایت به تعداد دسته، فریم برای آموزش وجود خواهد داشت. 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 خواهد داشت. از آنجا که پس از انتخاب تعدادی ویدئو، 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 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 برازش از روش پاک کردن تصادفی قسمتی از تصویر ورودی استفاده شده است [۶۰]. 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 نحوه برش زدن تصادفی چهره به همراه پس زمینه نشان داده شده است. در این تصویر مستطیل آبی چهره فرد را نشان می‌دهد و مستطیل‌های رنگی به صورت تصادفی برای هر بار انتخاب چهره انتخاب می‌شوند.

برای پیاده‌سازی کلاس بارگذاری داده یک data loader سفارشی نوشته شده است و همچنین برای آنکه استراتژی ترتیب تصادفی انتخاب ویدئو و استفاده مجدد از فریم‌های ویدئو متوالی پیاده شود یک تابع `iteration` سفارشی نوشته شده است. در پیاده‌سازی این تابع از مفهوم `iteration` در زبان برنامه نویسی پایتون استفاده شده است.



شکل ۱.۴: نحوه برش زدن تصادفی چهره با مقداری از پس زمینه

۳.۴ معیارهای ارزیابی

مسئله کشف تقلب یک مسئله طبقه‌بندی دو کلاسه است که در هنگام آزمون، معمولاً تعداد نمونه‌های واقعی و تقلیلی 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 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 ۱ نمونه، که به معنی این است که برچسب واقعی چهره زنده بوده است اما به عنوان چهره تقلیلی پیش‌بینی شده است. و نرخ خطای رد کردن ۲ که به معنی این است که نمونه برچسب تقلیلی دارد ولی به عنوان چهره زنده پیش‌بینی شده است.

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

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

ممولاً این مقدار بر اساس یک آستانه که یکی از یک پارامتر است محاسبه می‌گردد. برای مثال در شبکه‌ی عصبی مقدار تک نورون لایه آخر با تابع فعالسازی سیگموید، مقداری بین صفر و یک

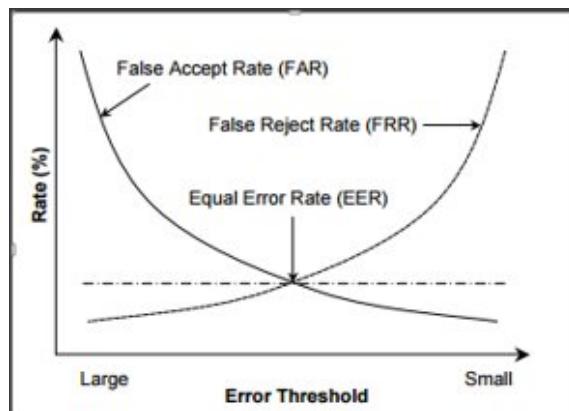
خواهد داشت. و با انتخاب یک سطح آستانه مقایسه مقدار نورون لایه‌ی آخر با این سطح آستانه تصمیم‌گیری در مورد پیش‌بینی برچسب نمونه انجام می‌شود. نرخ خطای برابر با مقداری است که FRR با FAR برابر شود.

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

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

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، معمولاً روی داده‌های آموزش وزن‌های شبکه به دست می‌آید و روی قسمت توسعه، پارامتر به دست خواهد آمد. و روی قسمت آزمون معیار نصف کل نرخ خطای 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 FRR و FAR پایین‌تر باشد، شبکه دقیق‌تر بود. همچنین مقدار FRR و

در نزدیکی‌های محل تقاطع نشان می‌دهد که شبکه چه میزان دو کلاس را از هم جدا کرده است.

یک معیار دیگر برای ارزیابی استفاده از استاندارد ISO/IEC 30107-3 است که در آن از نرخ خطای طبقه‌بندی ارائه حمله ۱ (APCER) و نرخ خطای طبقه‌بندی ارائه خوب ۲ (BPCER) تعریف می‌شود که در آن FRR معادل APCER است ولی APCER معادل بیشترین FAR به ازای ابزارهای حمله مختلف است. منظور از ابزار حمله، حمله کاغذ چاپ شده یا حمله بازپخش است. همچنین متوسط نرخ خطای طبقه‌بندی به صورت میانگین APCER و BPCER تعریف می‌شود.

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

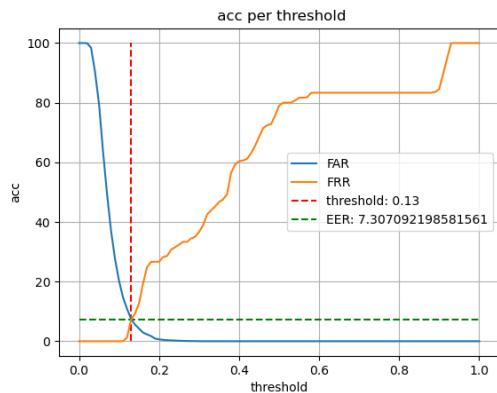
$$ACER = \frac{APCER + BPCER}{2} \quad (7.4)$$

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 انجام می‌شود. و سپس روی دیتاست‌های دیگر دقیق‌گزارش می‌شود.

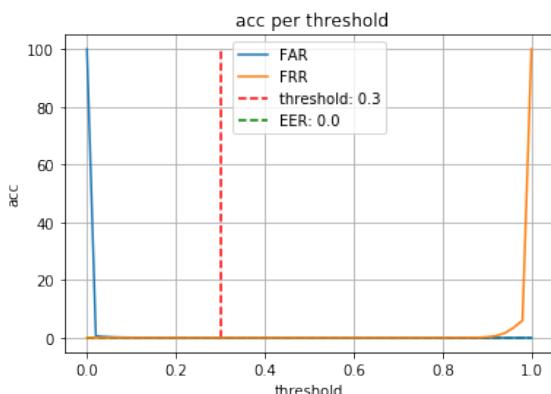
۱۰.۴.۴ اثر عملگر LBP قابل آموزش در دیتاست Replay

بهمنظور مقایسه‌ی 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 بهصورت شکل ۳.۴ است. همانطور که مشاهده می‌شود با در نظر گرفتن سطح آستانه ۱۳.۰ برای نورون آخر به خطای ۳.۷ درصد روی داده دیده نشده می‌رسیم. اما لازم است توجه شود تنها مقدار خطای نیست و عملکرد نمودار در سایر نقاط سطح آستانه نیز مهم است و در سطح آستانه ۶.۰ مقدار خطای FRR حدود ۸۰ درصد است که بسیار زیاد است. همچنین در اطراف سطح آستانه ۱۳.۰ با کمی تغییر در سطح آستانه مقدار خطای بزرگ می‌شود.



شکل ۴.۳: نمودار خطای برابر برای شبکه ALEXNET و تابع هزینه BCE

با استفاده از عملگر LBP قابل آموزش پیش از ALEXNET و تابع هزینه نیز کماکان BCE باشد نمودار شکل ۴.۴ به دست می‌آید. همانطور که مشاهده می‌شود استفاده از تنها یک لایه LBP پیش



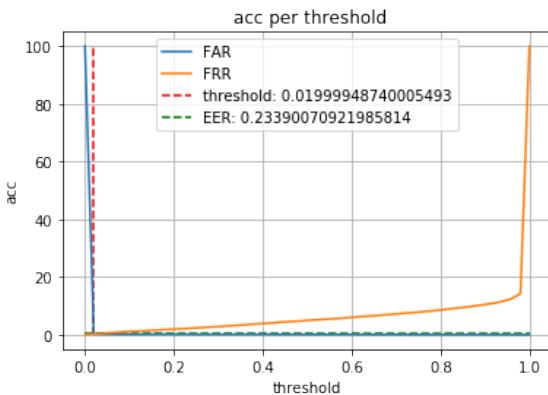
شکل ۴.۴: نمودار خطای برابر هنگام استفاده از عملگر LBP پیشنهادی

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 کمی محاسبات به شبکه اضافه می‌کند برای مقایسه دیگر نمودار آموزش شبکه با تابع هزینه BCE و شبکه efficient net B0

۵.۴ به صورت شکل

است.

این نمودار نشان می‌دهد لزوماً استفاده از شبکه پیچیده نمی‌تواند 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 Replay دارد. بلکه در این کاربرد خاص و دیتابست Replay که حجم داده کمی

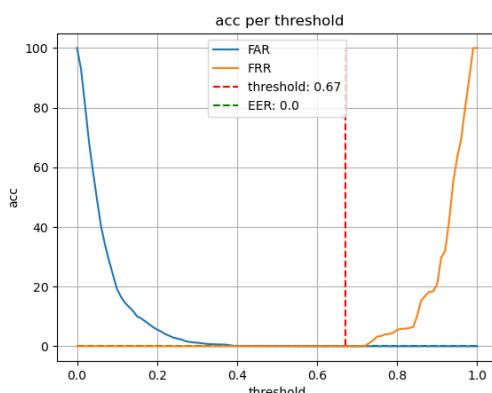


شکل ۴.۵: نمودار خطای برابر هنگام استفاده از شبکه efficient net B0

دان‌ها استفاده از شبکه ساده‌تر اما هوشمندانه با توجه به مسئله، دقت بهتری را ایجاد می‌کند.

۲.۴.۴ اثر تابع هزینه ARCB در دیتاست Replay

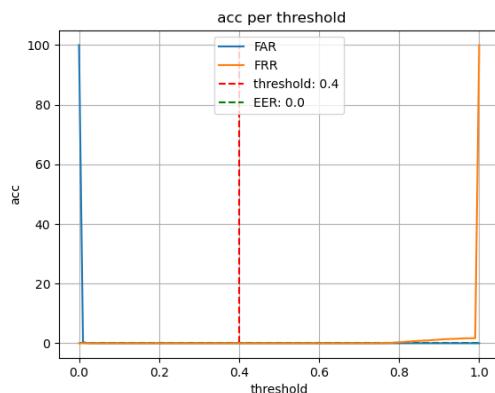
اکنون تنها از شبکه ALEXNET بدون عملگر LBP استفاده می‌شود ولی تابع هزینه ARCB معرفی شده به جای تابع BCE استفاده می‌شود. نمودار شکل ۶.۴ نشان می‌دهد تغییر تابع هزینه بدون تغییری در ساختار می‌تواند تاثیرگذار باشد. نمودار در مقایسه با نمودارهای قبلی متقارن‌تر شده است. در این شکل میزان خطا در اطراف سطح آستانه صفر است ولی با دور شدن از سطح آستانه و نزدیک شدن به مقدار ۰ و ۱ خطا بیشتر می‌شود. این تأثیر حاشیه در تابع هزینه ARCB است که موجب شده است دو کلاس با یک حاشیه از یک دیگر جدا شوند.



شکل ۶.۴: نمودار خطای برابر هنگام استفاده از تابع هزینه ARCB پیشنهادی

۳.۴.۴ اثر تابع هزینه بر پایه شناسه اشخاص در دیتاست Replay

اکنون از ساختار ساده ALEXNET استفاده می‌شود و تابع هزینه برای طبقه بند 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 خطای در آستانه‌های ۰ تا ۰.۸ به صورت مطلق صفر است. که نشان می‌دهد دو کلاس با حاشیه مناسبی از هم جدا شده‌اند. تا این قسمت اثر هر کدام از روش‌های پیشنهادی به تنها‌ی بررسی شده‌اند. برای ادامه فصل تمامی روش‌ها در کنار یکدیگر استفاده می‌شود. و شبکه استخراج ویژگی Efficient net است. همچنین به منظور تسريع در همگرا شدن شبکه، قسمت استخراج ویژگی از وزن‌های آموزش دیده روی دیتاست image-ne استفاده می‌شود ولی این وزن‌ها حین آموزش تغییر می‌کند.



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

۴.۴.۴ نتایج روی دیتاست‌های MSU و CASIA

دیتاست‌های MSU و CASIA نسبت به دیتاست Replay دارای رزولوشن تصویر بیشتری هستند. این دیتاست‌ها بر خلاف دیتاست 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 دیتاست گزارش شده است.

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 دقیق دیتاست‌های SIW و OULU است. این دو دیتاست نسبت به دیتاست‌های قبلی جدیدتر و دارای حجم بیشتری هستند. به همین دلیل

جدول ۱.۴: خطای برابر روی دیتاست‌های CASIA و MSU

EER (%)	Dataset
0.54	CASIA
0.0	MSU

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 تعیین‌پذیری 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 ویدئو استفاده می‌شود ولی برای تست از تمامی فریم‌های ویدیوهای تست استفاده می‌شود. از آنجا که در فریم‌های ابتدایی هر ویدئو، 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 تعیین‌پذیری 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

نمایشگر	۰	۱	۲
[۲۰] ۰۰۰۰۰۰۰۰۰۰	۵۸.۳	۵۸.۳	۵۸.۳
[۱۴] ۰۰۰۰	۰	۵۰.۰	۲۵.۰
[۲۵] ۰۰۰۰۰	-	-	۱
[۸] ۰۰۰۰	۰۷.۰	۱۷.۰	۱۲.۰
[۲۴] ۰۰۰۰	۶۴.۰	۱۷.۰	۴.۰
[۲۸] ۰۰۰-۰۰۰۳	۶۹.۰	۱۷.۰	۴.۰
۰۰۰۰+۰۰۰	۱۴.۰	۱۲.۰	۱۳.۰

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 عمکرد 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

	۱	۲	۳	۴
[۲۰] ۱	۶۹.۰ ۵۷.۰	۶۹.۰ ۵۷.۰	۶۹.۰ ۵۷.۰	
[۱۴] ۲	۰.۰	۰.۰	۰.۰	
[۲۵] ۳	-	-	۰۵.۰ ۲۸.۰	
[۸] ۴	۰.۰	۰.۹.۰ ۰	۰.۵.۰ ۰۴.۰	
[۲۴] ۵	۰.۰ ۰.۰	۰.۸.۰ ۰۴.۰	۰.۴.۰ ۰۲.۰	
[۲۸] ۶	۲۸.۰ ۴۶.۰	۰۶.۰ ۰۴۳.۰	۱۴.۰ ۰۴۵.۰	
۷	۰۱۲۹.۰ ۰۰۷۵.۰	۰۱۷۳.۰ ۰۱.۰	۰۱۵۱.۰ ۰۰۸۷.۰	

۶.۴.۴ دقต 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 تست انجام می‌گیرد. این پروتکل با این هدف ارائه شده است که قابلیت روش پیشنهادی با تغییر مکان تصویربرداری ارزیابی شود. 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 جدول ۴.۴ در پروتکل اول و دوم گزارش شده است.

۷.۴.۴ نتایج روی آزمون بین دیتاست

هماویدیویی 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

جدول ۴.۴: دقต در پروتکل‌های اول و دوم دیتاست OULU

	۲ ۰۰۰۰۰۰۰۰۰			۱ ۰۰۰۰۰۰۰۰۰		
دیتاست	۰۰۰۰۰	۰۰۰۰۰	۰۰۰۰۰	۰۰۰۰۰	۰۰۰۰۰	۰۰۰۰۰
[۱۵] ۰۰۰	۵.۲	۳.۱	۹.۱	۵.۲	۹.۸	۷.۵
[۲۰] ۰۰۰۰۰۰۰۰۰	۷.۲	۷.۲	۷.۲	۶.۱	۶.۱	۶.۱
[۱۶] ۰۰۰۰۰۰	۲.۴	۴.۴	۳.۴	۲.۱	۷.۱	۵.۱
[۱۴] ۰۰۰۰۰	۸.۰	۶.۰	۷.۰	۸.۰	۰	۴.۰
[۲۵] ۰۰۰۰۰	۲.۴	۳.۰	۲.۲	۲.۱	۵.۲	۹.۱
[۸] ۰۰۰۰۰	۸.۱	۸.۰	۳.۱	۴.۰	۰	۲.۰
[۲۴] ۰۰۰۰۰	۵.۲	۳.۱	۹.۱	۰.۲	۰.۰	۰.۱
[۲۶] ۰۰۰۰۰۰۰۰۰۰۰	۴.۱۱	۶.۰	۰.۶	۸۳.۰	۰	۴۲.۰
[۱۷] ۰۰۰۰۰	۳.۲	۶.۱	۹.۱	۸.۰	۳.۱	۱.۱
[۲۸] ۰۰۰-۰۰۰۲	۱.۳	۸.۲	۰.۳	۳.۲	۰	۲.۱
۰۰۰۰۰+۰۰۰۰	۹۷.۰	۹۷.۰	۹۷.۰	۵۸.۲	۲	۲۹.۲

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 تست بین دو دیتاست مختلف است. بدین منظور مدل روی یک دیتاست آموزش داده می‌شود و روی دیتاست دیگر تست می‌شود. برای بررسی دقت مدل در تست بین دیتاست، شبکه روی دیتاست CASIA آموزش داده شده است و روی دیتاست Replay تست شده است. نتایج این حالت در جدول [۴]

به همراه دقت پژوهش‌های دیگر گزارش شده است. با مقایسه نتایج دقت در آزمون بین دیتاست

جدول ۵.۴: تایج روی آزمون بین دیتاست

دیتاست	۰۰۰۰۰
[۲۵] ۰۰۰۰۰	31.5
[۲۴] ۰۰۰۰۰	۱۷
[۲۰] ۰۰۰۰۰۰۰۰۰	۶.۲۷
[۱۶] ۰۰۰۰۰۰	۵.۲۸
[۱۵] ۰۰۰	۴.۲۱
[۱۴] ۰۰۰۰۰	۴.۲۷

و درون دیتاست تفاوت قابل ملاحظه خطأ، دیده می‌شود.

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 efficient net 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 توسعه تابع 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 تست شبکه تغییری نمی‌کند

۲.۵ پیشنهاد کارهای آینده

در این پژوهش از efficient net 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 برای افزایش دقت با یک ساختار جدید می‌تواند به افزایش دقت کمک کند. بهمنظور آنالیز بهتر بافت در تصویر، عملگر LBP می‌تواند توسعه بیشتری داده شود به‌گونه‌ای که در تمامی لایه‌های شبکه بهجای کانولوشن قرار بگیرد. همچنین تابع هزینه ARCB می‌تواند مشابه روش [۱۳] 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 باز نویسی شود. همچنین استفاده از عمق در کنار روش پیشنهادی ممکن است دقت بهتری به‌دست آورد.

در این پایان‌نامه تمرکز 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 آدام و شبکه با وزن‌های آموزش دیده شده استفاده شده است. پژوهش بعدی می‌تواند شامل استفاده از بهینه سازی SGD و شروع با وزن‌های تصادفی و آموزش روی تعداد ایپاک زیاد باشد که ممکن است نقطه بهینه بهتری را پیدا کند.

مراجع

- [1] Anjos, André, Chakka, Murali Mohan, and Marcel, Sébastien. Motion-based countermeasures to photo attacks in face recognition. *IET biometrics*, 3(3):147–158, 2014.
- [2] Atoum, Yousef, Liu, Yaojie, Jourabloo, Amin, and Liu, Xiaoming. Face anti-spoofing using patch and depth-based cnns. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 319–328. IEEE, 2017.
- [3] Boulkenafet, Zinelabidine, Komulainen, Jukka, and Hadid, Abdenour. Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24(2):141–145, 2017.
- [4] Boulkenafet, Zinelabinde, Komulainen, Jukka, Li, Lei, Feng, Xiaoyi, and Hadid, Abdenour. 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)*, pages 612–618. IEEE, 2017.
- [5] Chingovska, Ivana, Anjos, André, and Marcel, Sébastien. 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)*, pages 1–7. IEEE, 2012.
- [6] Cho, Kyunghyun, Van Merriënboer, Bart, Gulcehre, Caglar, Bahdanau, Dzmitry, Bougares, Fethi, Schwenk, Holger, and Bengio, Yoshua. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [7] De Marsico, Maria, Nappi, Michele, Riccio, Daniel, and Dugelay, Jean-Luc. Moving face spoofing detection via 3d projective invariants. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 73–78. IEEE, 2012.
- [8] Deng, Jiankang, Guo, Jia, Xue, Niannan, and Zafeiriou, Stefanos. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

- [9] Feng, Haocheng, Hong, Zhibin, Yue, Haixiao, Chen, Yang, Wang, Keyao, Han, Junyu, Liu, Jingtuo, and Ding, Errui. Learning generalized spoof cues for face anti-spoofing. *arXiv preprint arXiv:2005.03922*, 2020.
- [10] Feng, Yao, Wu, Fan, Shao, Xiaohu, Wang, Yanfeng, and Zhou, Xi. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 534–551, 2018.
- [11] Freitas Pereira, Tiago de, Anjos, André, Martino, José Mario De, and Marcel, Sébastien. Lbp-based countermeasure against face spoofing attacks. In *Asian Conference on Computer Vision*, pages 121–132. Springer, 2012.
- [12] Gan, Junying, Li, Shanlu, Zhai, Yikui, and Liu, Chengyun. 3d convolutional neural network based on face anti-spoofing. In *2017 2nd international conference on multimedia and image processing (ICMIP)*, pages 1–5. IEEE, 2017.
- [13] George, Anjith and Marcel, Sébastien. Deep pixel-wise binary supervision for face presentation attack detection. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.
- [14] George, Anjith and Marcel, Sébastien. Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks. *IEEE Transactions on Information Forensics and Security*, 16:361–375, 2020.
- [15] Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, and Bengio, Yoshua. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [16] Hochreiter, Sepp and Schmidhuber, Jürgen. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [17] Huang, Gao, Liu, Zhuang, Van Der Maaten, Laurens, and Weinberger, Kilian Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [18] Jia, Yunpei, Zhang, Jie, Shan, Shiguang, and Chen, Xilin. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020.
- [19] Jourabloo, Amin, Liu, Yaojie, and Liu, Xiaoming. Face de-spoofing: Anti-spoofing via noise modeling. In *Proceedings of the European conference on computer vision (ECCV)*, pages 290–306, 2018.

- [20] Kingma, Diederik P and Ba, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [21] Komulainen, Jukka, Hadid, Abdenour, and Pietikäinen, Matti. Context based face anti-spoofing. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2013.
- [22] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [23] Li, Haoliang, He, Peisong, Wang, Shiqi, Rocha, Anderson, Jiang, Xinghao, and Kot, Alex C. Learning generalized deep feature representation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(10):2639–2652, 2018.
- [24] Li, Lei and Feng, Xiaoyi. Face anti-spoofing via deep local binary pattern. In *Deep Learning in Object Detection and Recognition*, pages 91–111. Springer, 2019.
- [25] Li, Lei, Feng, Xiaoyi, Boulkenafet, Zinelabidine, Xia, Zhaoqiang, Li, Mingming, and Hadid, Abdenour. An original face anti-spoofing approach using partial convolutional neural network. In *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2016.
- [26] Li, Xuan, Wan, Jun, Jin, Yi, Liu, Ajian, Guo, Guodong, and Li, Stan Z. 3dpc-net: 3d point cloud network for face anti-spoofing. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2020.
- [27] Liu, Yaojie, Jourabloo, Amin, and Liu, Xiaoming. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 389–398, 2018.
- [28] Liu, Yaojie, Stehouwer, Joel, and Liu, Xiaoming. On disentangling spoof trace for generic face anti-spoofing. In *European Conference on Computer Vision*, pages 406–422. Springer, 2020.
- [29] Määttä, Jukka, Hadid, Abdenour, and Pietikäinen, Matti. Face spoofing detection from single images using micro-texture analysis. In *2011 international joint conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011.
- [30] Parkhi, Omkar M, Vedaldi, Andrea, and Zisserman, Andrew. Deep face recognition. 2015.

- [31] Patel, Keyurkumar, Han, Hu, and Jain, Anil K. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10):2268–2283, 2016.
- [32] Pérez-Cabo, Daniel, Jiménez-Cabello, David, Costa-Pazo, Artur, and López-Sastre, Roberto J. Deep anomaly detection for generalized face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [33] Ramachandra, Raghavendra and Busch, Christoph. Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 50(1):1–37, 2017.
- [34] Rehman, Yasar Abbas Ur, Po, Lai-Man, and Komulainen, Jukka. Enhancing deep discriminative feature maps via perturbation for face presentation attack detection. *Image and Vision Computing*, 94:103858, 2020.
- [35] Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [36] Schroff, Florian, Kalenichenko, Dmitry, and Philbin, James. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
- [37] Schwartz, William Robson, Rocha, Anderson, and Pedrini, Helio. Face spoofing detection through partial least squares and low-level descriptors. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2011.
- [38] Shao, Rui, Lan, Xiangyuan, Li, Jiawei, and Yuen, Pong C. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10031, 2019.
- [39] Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [40] Tan, Mingxing and Le, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

- [41] Tu, Xiaoguang, Ma, Zheng, Zhao, Jian, Du, Guodong, Xie, Mei, and Feng, Jiashi. Learning generalizable and identity-discriminative representations for face anti-spoofing. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–19, 2020.
- [42] Wang, Tao, Yang, Jianwei, Lei, Zhen, Liao, Shengcai, and Li, Stan Z. Face liveness detection using 3d structure recovered from a single camera. In *2013 international conference on biometrics (ICB)*, pages 1–6. IEEE, 2013.
- [43] Wang, Zezheng, Yu, Zitong, Zhao, Chenxu, Zhu, Xiangyu, Qin, Yunxiao, Zhou, Qiusheng, Zhou, Feng, and Lei, Zhen. Deep spatial gradient and temporal depth learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5042–5051, 2020.
- [44] Wang, Zezheng, Zhao, Chenxu, Qin, Yunxiao, Zhou, Qiusheng, Qi, Guojun, Wan, Jun, and Lei, Zhen. Exploiting temporal and depth information for multi-frame face anti-spoofing. *arXiv preprint arXiv:1811.05118*, 2018.
- [45] Wen, Di, Han, Hu, and Jain, Anil K. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.
- [46] Xu, Xiang, Xiong, Yuanjun, and Xia, Wei. On improving temporal consistency for online face liveness detection system. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 824–833, 2021.
- [47] Xu, Zhenqi, Li, Shan, and Deng, Weihong. Learning temporal features using lstm-cnn architecture for face anti-spoofing. In *2015 3rd IAPR asian conference on pattern recognition (ACPR)*, pages 141–145. IEEE, 2015.
- [48] Yang, Jianwei, Lei, Zhen, and Li, Stan Z. Learn convolutional neural network for face anti-spoofing. *arXiv preprint arXiv:1408.5601*, 2014.
- [49] Yang, Jianwei, Lei, Zhen, Liao, Shengcai, and Li, Stan Z. Face liveness detection with component dependent descriptor. In *2013 International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2013.
- [50] Yang, Xiao, Luo, Wenhan, Bao, Linchao, Gao, Yuan, Gong, Dihong, Zheng, Shibao, Li, Zhifeng, and Liu, Wei. Face anti-spoofing: Model matters, so does data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3507–3516, 2019.

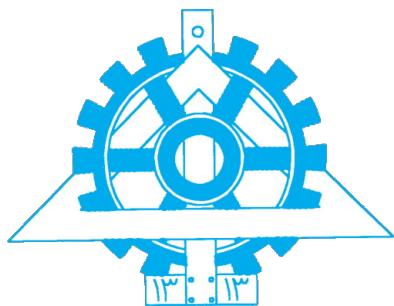
- [51] Yin, Wenze, Ming, Yue, and Tian, Lei. A face anti-spoofing method based on optical flow field. In *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pages 1333–1337. IEEE, 2016.
- [52] Yu, Zitong, Li, Xiaobai, Niu, Xuesong, Shi, Jingang, and Zhao, Guoying. Face anti-spoofing with human material perception. In *European Conference on Computer Vision*, pages 557–575. Springer, 2020.
- [53] Yu, Zitong, Qin, Yunxiao, Li, Xiaobai, Zhao, Chenxu, Lei, Zhen, and Zhao, Guoying. Deep learning for face anti-spoofing: A survey. *arXiv preprint arXiv:2106.14948*, 2021.
- [54] Yu, Zitong, Qin, Yunxiao, Xu, Xiaqing, Zhao, Chenxu, Wang, Zezheng, Lei, Zhen, and Zhao, Guoying. Auto-fas: Searching lightweight networks for face anti-spoofing. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 996–1000. IEEE, 2020.
- [55] Yu, Zitong, Wan, Jun, Qin, Yunxiao, Li, Xiaobai, Li, Stan Z, and Zhao, Guoying. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3005–3023, 2020.
- [56] Yu, Zitong, Zhao, Chenxu, Wang, Zezheng, Qin, Yunxiao, Su, Zhuo, Li, Xiaobai, Zhou, Feng, and Zhao, Guoying. Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5295–5305, 2020.
- [57] Zhang, Kaipeng, Zhang, Zhanpeng, Li, Zhifeng, and Qiao, Yu. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.
- [58] Zhang, Ke-Yue, Yao, Taiping, Zhang, Jian, Tai, Ying, Ding, Shouhong, Li, Jilin, Huang, Feiyue, Song, Haichuan, and Ma, Lizhuang. Face anti-spoofing via disentangled representation learning. In *European Conference on Computer Vision*, pages 641–657. Springer, 2020.
- [59] Zhang, Zhiwei, Yan, Junjie, Liu, Sifei, Lei, Zhen, Yi, Dong, and Li, Stan Z. A face antispoofing database with diverse attacks. In *2012 5th IAPR international conference on Biometrics (ICB)*, pages 26–31. IEEE, 2012.
- [60] Zhong, Zhun, Zheng, Liang, Kang, Guoliang, Li, Shaozi, and Yang, Yi. Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13001–13008, 2020.

- [61] Zoph, Barret and Le, Quoc V. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016.

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