به نام خدا

شناسایی چهره مبتنی بر یادگیری عمیق در محیطIoT-cloud

استاد مربوطه:

خانم دکتر جاسبی

تهیه کننده:

حمید فــرزین پور

علوم تحقیقات بهار1400

چکیده:

در سال‌های اخیر ، فناوری اینترنت اشیا در بسیاری از زمینه های کاربردی مانند مراقبت های بهداشتی، نظارت تصویری، حمل و نقل و غیره مورد استفاده قرار می گیرد. پذیرش و رشد گسترده اینترنت اشیا در این زمینه‌ها حجم عظیمی از داده ها را تولید می کند. به عنوان مثال، دستگاه های اینترنت اشیا مانند دوربین ها هنگام استفاده در سناریوهای نظارت در بیمارستان، حجم عظیمی از تصاویر را تولید می کنند. در اینجا، تشخیص چهره یک عنصر مهم است که می تواند برای تأمین امکانات بیمارستان، تشخیص علائم و تجزیه و تحلیل علائم بیماران، تشخیص تقلب بیمار و تجزیه و تحلیل الگوی ترافیک بیمارستان مورد استفاده قرار گیرد. سیستم های تشخیص چهره به صورت خودکار و هوشمند در محیط کنترل شده از دقت بالایی برخوردار هستند. با این حال، آنها در یک محیط کنترل نشده دقت کمی دارند. همچنین، سیستم ها باید در بسیاری از برنامه های کاربردی مانند مراقبت های بهداشتی هوشمند به صورت بلادرنگ کار کنند. این مقاله یک مدل عمیق مبتنی بر درخت را برای تشخیص خودکار صورت در یک محیط ابر پیشنهاد می کند. مدل عمیق پیشنهادی از نظر محاسباتی بدون به خطر انداختن دقت کم هزینه تر است. در مدل ، یک حجم ورودی به چندین جلد تقسیم می شود ، جایی که برای هر حجم یک درخت ساخته می شود. یک درخت با عامل انشعاب و ارتفاع آن تعریف می شود. هر شاخه با یک تابع باقیمانده نشان داده می شود ، که توسط یک لایه کانولوشن ، یک نرمال سازی دسته ای و یک تابع غیر خطی تشکیل شده است. مدل پیشنهادی در پایگاه های داده مختلف موجود در دسترس عموم ارزیابی می شود. مقایسه عملکرد همچنین با پیشرفته ترین مدل های عمیق برای تشخیص چهره انجام می شود. نتایج آزمایشات نشان می دهد که مدل پیشنهادی به ترتیب، به دقت 98.65٪ ، 99.19٪ ، 95.84٪ در پایگاه داده های FEI، ORL و LFW دست یافته است.

مقدمه

تشخیص چهره یکی از قدیمی ترین موضوعات تحقیقاتی است که در عین حال پویا است. برای امنیت و کاربردهای بیومتریک ضروری است. سیستم های تشخیص چهره اولیه به ویژگی های دستی و طبقه بندی کننده سنتی متکی بودند. برخی از ویژگی های دست ساز شامل الگوی دودویی محلی (LBP) ، توصیف کننده های محلی وبر (WLD) ، تجزیه و تحلیل مولفه اصلی (PCA) و هیستوگرام شیب های گرا (HOG) است. طبقه بندی کننده های سنتی شامل ماشین های بردار پشتیبانی (SVM) ، تجزیه و تحلیل تفکیک خطی (LDA) و برخی از طبقه بندی های حداقل مبتنی بر دور است. این ویژگی ها و طبقه بندی ها در یک محیط کنترل شده ، جایی که چهره ها اغلب از جلو و با بیان خنثی هستند و دارای تنوع کمتری از نور هستند ، به خوبی کار می کنند. با این حال ، در بسیاری از برنامه های کاربردی مانند آنهایی که به نظارت مربوط می شوند ، تصاویر چهره ممکن است پنهان باشند ، از جلو نباشند و دارای وضوح پایین و تنوع زیاد نور باشند. برای این برنامه ها ، سیستم های تشخیص چهره سنتی ممکن است به درستی کار نکنند.

یادگیری عمیق یک تکنیک قدرتمند یادگیری ماشین است که با موفقیت در بسیاری از برنامه های پردازش سیگنال استفاده شده است. این برنامه ها شامل تشخیص گفتار و بلندگو ، تشخیص تصویرو تشخیص فیلم هستند. از زمان معرفی یادگیری عمیق ، بسیاری از معماری ها در حوزه ارائه شده است. این معماری ها از جنبه های مختلف از جمله

تعداد لایه ها ، تعداد فیلترها ، اندازه فیلترها و ترتیب لایه ها با هم متفاوت هستند. اگر تعداد لایه ها زیاد باشد ، می توان معماری را عمیق نامید، در غیر این صورت می توان آن را کم عمق نامید. اگر اندازه فیلتر بزرگ باشد ، معماری گسترده نامیده می شود، در غیر این صورت ، نازک است. از طرف دیگر ، معماری کم عمق و گسترده ای مانند شبکه گسترده باقیمانده (**WRN**) بر اندازه فیلتر بیشتر از عمق شبکه تأکید می کند. هر یک از این معماری ها مزایا و معایب خاص خود را دارند. یک شبکه بسیار عمیق و نازک ممکن است عملکرد خوبی داشته باشد. با این حال ، تعداد پارامترهای بسیار بالایی خواهد داشت. به طور معمول ، یک شبکه یادگیری عمیق با تعداد زیادی پارامتر ممکن است برای یک برنامه در زمان واقعی مناسب نباشد. بنابراین ، نیاز به تعادل بین دقت و پارامترهای مورد نیاز برای برنامه های کاربردی در زمان واقعی مربوط به امنیت وجود دارد .

برنامه های کاربردی در یک شهر هوشمند به یک خروجی دقیق در زمان واقعی نیاز دارند. به عنوان مثال ، تراکم ترافیک و مسیرهای جایگزین باید در زمان واقعی تعیین شود ، دسترسی ایمن نیاز به تأیید دقیق دارد و تشخیص پزشکی باید بدون خطا باشد. این شهر هوشمند ممکن است اجزای زیادی مانند خانه های هوشمند ، مراقبت های بهداشتی هوشمند ، مدارس هوشمند ، وسیله نقلیه متصل هوشمند و خریدهای هوشمند داشته باشد. به دلیل افزایش **IoT** ها ، حجم داده ها توسط منیفولدها افزایش یافته است. این حجم ناهمگن و فیلتر نشده در نظر گرفته شده است. این حجم عظیم داده باید با دقت پردازش شود تا خروجی دقیق باشد ، امنیت شکسته نشود و پردازش زمان زیادی را صرف نکند. دستیابی به همه این موارد در یک زمان بسیار دشوار است. با این حال ، باید تعادلی بین آنها برقرار باشد.

صورت یک مولفه مهم بیومتریک در تأیید و شناسایی انسان است. می تواند به عنوان یک فرآیند تأیید ایمن استفاده شود. طی چند دهه گذشته بسیاری از سیستم های شناسایی چهره ارائه شده اند. تحقیقات تشخیص چهره از شناسایی چهره های گرفته شده در یک محیط کنترل شده به عکس گرفتن در یک محیط کنترل نشده ، از تکنیک های استخراج ویژگی های سنتی گرفته تا تکنیک های استخراج ویژگی های عمیق ، تکامل یافته است.

مطالعات مرتبط

در این بخش ، پیشرفت شناخت چهره با استفاده از یادگیری عمیق مورد بحث قرار گرفته است. تقریباً در همه کارهای قبلی از شبکه های عصبی کانولوشن (**CNN**) به عنوان معماری اصلی استفاده شده است. از یک **CNN** سبک با واحدهای نقشه حداکثر ویژگی در برنامه های بیومتریک تشخیص چهره در استفاده شد. یک روش ارزیابی کیفیت بیومتریک در **CNN** جاسازی شده بود. سان و همکاران از ماشین های محدود بولتزمن محدود برای تشکیل یک شبکه کانولوشن عمیق برای تأیید چهره استفاده کرد. این روش با استفاده از چهره های دارای برچسب در پایگاه داده **LFW** به دقت متوسط **93.83**٪ دست یافت شبکه عمیق با استفاده از **VGGNet** با ترکیب ویژگی های تصویر نور مرئی و تصویر نزدیک به مادون قرمز برای تشخیص چهره. یک استراتژی همجوشی برای فیوز نمرات پیشنهاد شد. آنها با استفاده از پایگاه داده **LFW** به دقت **97.35**٪ دست یافتند.

تشخیص چهره مبتنی بر CNN عمیق برای نوزادان در پیشنهاد شد. در پایگاه داده ، 2100 تصویر چهره از 210 نوزاد وجود داشت که در آن هر نوزاد 10 تصویر داشت. نویسندگان دریافتند که افزایش تعداد لایه ها در تشخیص چهره نوزاد عملکرد خوبی ندارد. در آزمایشات ، آنها 91.03 دقت پیدا کردند. گوا و همکاران با ادغام ویژگی های تصویر نور مرئی و تصویر مادون قرمز نزدیک برای شناسایی چهره ، یک شبکه عمیق را با استفاده از VGGNet طراحی کرد. یک استراتژی همجوشی برای فیوز نمرات پیشنهاد شد. آنها با استفاده از پایگاه داده LFW به دقت 97.35٪ دست یافتند.

هو و همکاران [26] با استفاده از دو مدل مختلف CNN عملکرد تشخیص چهره D2 و D3 را مطالعه کرد. آنها دریافتند که مدل عمیق تر CNN عملکرد بهتری نسبت به مدل CNN دیگر دارد. در آزمایشات ، دقت 95٪ با استفاده از پایگاه داده آزمایشگاه تحقیقات Olivetti (ORL) بدست آمد. یک سیستم تشخیص چهره چند حالته با استفاده از CNN های مخصوص حالت (RGB و عمق) در پیشنهاد شد. برخی از ویژگی های دست ساز مانند LBP، HOG و ویژگی های مشابه Haar نیز برای بهبود عملکرد با ویژگی های عمیق آموخته شده ترکیب شده اند.

مدل ResNet همراه با عمیق متشکل از یک شبکه تنه و دو شبکه شاخه برای تشخیص چهره در پیشنهاد شد. از شبکه های شاخه ای برای تبدیل تصاویر با وضوح بالا به تصاویر با وضوح پایین مورد نظر استفاده شد. این مدل در پایگاه داده LFW به دقت 99٪ رسیده است.

برخی از مدل های معروف برای تشخیص چهره شامل مدل های DeepFace، DeepID، DeepID2 و DeepID2 + هستند. تایگمن و همکاران مدل DeepFace را توسعه داده است که دارای هشت لایه است. سه لایه اول دارای لایه های جمع کننده کانولوشن هستند ، سه لایه بعدی لایه های متصل محلی هستند ، در حالی که دو لایه آخر لایه های کاملاً متصل هستند. در پایگاه داده LFW ، مدل بیش از 97٪ دقت داشت. DeepID و انواع آن از مجموعه CNN های کوچک استفاده کرده و آنها را ذوب می کند. هر CNN کوچک دارای چهار لایه کانولوشن ، سه لایه استخر و دو لایه کاملاً متصل بود. این تغییرات در تعداد فیلترهای لایه ها رخ داده است. دقت 99.47٪ با استفاده از مدل DeepID2 + با استفاده از پایگاه داده LFW بدست آمده است.

VGGFace معرفی شده در سال 2015 از مدل VGGNet-16 استفاده کرده و حدود 99٪ دقت در پایگاه داده LFW بدست آورده است [32]. گوگل FaceNet را معرفی کرد که از معماری GoogleNet-24 استفاده می کرد ، بیش از 99٪ دقت در همان پایگاه داده داشت. سیستم های شناسایی چهره مبتنی بر ResNet مانند از دست دادن AMS و CosFace اخیراً ساخته شده اند. این سیستم ها در چندین پایگاه داده به دقت بالایی دست یافتند. در جدول 1 خلاصه ای از کارهای قبلی آورده شده است.

تقریباً تمام مدل های ذکر شده در بخش به طور عمده بر بهبود دقت تشخیص چهره در محدودیت های مختلف تمرکز داشتند. تعداد کمی از افراد سعی در ایجاد مدل عمیق با پیچیدگی کم برای تشخیص چهره داشتند.

مواد و روش ها

چارچوب شهر هوشمند

چارچوب شهر هوشمند از چندین مولفه شامل خانه های هوشمند ، سیستم های هوشمند ترافیک ، خریدهای هوشمند ، مراقبت های بهداشتی هوشمند ، شبکه های بی سیم پرسرعت و سرورهای ابری تشکیل شده است. سیگنالهای مختلفی که توسط اینترنت اشیا و دستگاههای هوشمند گرفته می شوند برای پردازش و تصمیم گیری به سرور ابری ارسال می شوند. پس از پردازش ، تصمیم برای ذینفعان ارسال می شود. ذینفعان می توانند بر این اساس اقدامات لازم را انجام دهند. یک سناریو را فرض کنید که فردی بخواهد به پارک دسترسی پیدا کند. یک دستگاه هوشمند با استفاده از فناوری بی سیم 5G ، تصویر چهره فرد را گرفته و تصویر را به ابر می فرستد.

در ابر ، چندین ماشین مجازی (VM) وجود دارد که می توانند به طور موازی کار کنند. این ماشین های مجازی به قدرت پردازش بالایی مجهز هستند. تصویر صورت در یکی از این ماشین های مجازی در زمان واقعی پردازش می شود و تأیید انجام می شود. این تصمیم که آیا شخص می تواند به پارک دسترسی داشته باشد یا نه به دروازه پارک ارسال می شود. این ابر دارای حافظه ذخیره سازی بزرگی است که می تواند حجم زیادی از داده های چندرسانه ای را ذخیره کند. ماشین های مجازی می توانند به طور موازی کار کنند تا ابر بتواند همزمان چندین درخواست را کنترل کند

این چارچوب ممکن است از یک امکان محاسبه لبه تشکیل شده باشد. محاسبه لبه در لبه ماشین های محلی و قبل از شبکه های دسترسی رادیویی قرار می گیرد. هدف از محاسبه لبه پردازش و فیلتر سیگنال ها است تا حجم داده هایی که باید به ابر منتقل شوند به حداقل برسد. ماشین های محلی در محاسبه لبه می توانند تلفن های هوشمند یا دستگاه هایی باشند که برای پردازش حداقل سیگنال ها بین خود ارتباط برقرار می کنند. این دستگاه ها قدرت پردازشی کمی دارند و عمر آنها توسط باتری های محدود شده محدود می شود. وظیفه پردازش با استفاده از الگوریتم بهینه سازی در دستگاه ها توزیع می شود. ویژگی هایی که برای بهینه سازی در نظر گرفته می شود عمر باتری ، قدرت پردازش ، بار فعلی و حجم داده است. پس از پایان پردازش ، داده های پردازش شده به ابر ارسال می شوند.

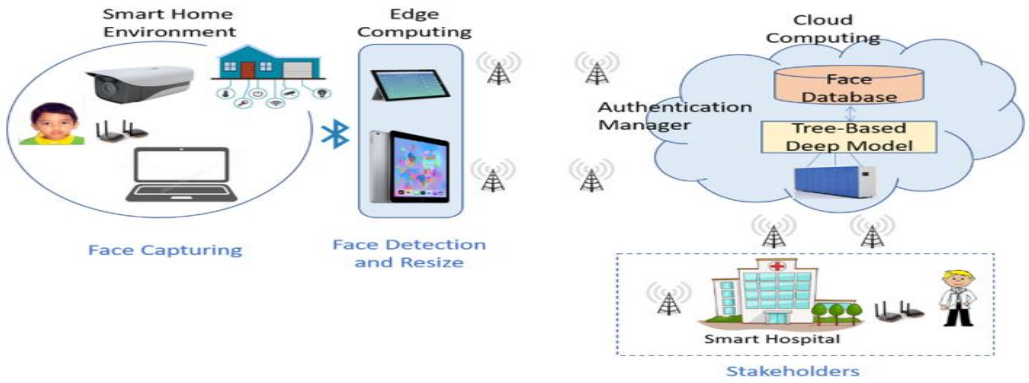

Fig. 1. Structure of a smart city framework for face recognition.

شبکه عمیق مبتنی بر درخت

یک شبکه عصبی عمیق (DNN) مزایای بسیاری را در یادگیری ماشین به دست آورده است. از زمان اختراع آن ، DNN در بسیاری از برنامه ها خصوصاً در پردازش سیگنال تصویر و گفتار مورد استفاده قرار گرفته است. DNN در کاربردهای

مربوط به پردازش تصویر و گفتار به دقت بسیار بالایی دست یافته است. شبکه عمیق برای آموزش به مقدار زیادی داده نیاز دارد.

بسیاری از معماری های DNN وجود دارد. برخی از آنها بر روی عمق تمرکز می کنند و برخی دیگر به اندازه فیلترها تمرکز می کنند. بسته به نوع کاربرد ، ممکن است از یکی از این نوع معماری ها استفاده کنیم. اخیراً شبکه عمیق درختی ارائه شده است ایده مدل عمیق درختی توزیع پردازش در یک ساختار درخت مانند است.

در این مقاله دو مدل عمیق ارائه شده است. این مدل ها یک مدل تک درخت و یک مدل درخت موازی هستند. ابتدا ساختار مدل تک درخت توصیف می شود.

در مدل تک درخت ، ابتدا با افزایش تعداد کانال ها ، حجم ورودی به یک حجم جدید ترسیم می شود. فرض کنید که حجم جدید $H \times W \times C$ باشد ، جایی که $H \times W$ بعد فضایی تصویر است و C شماره کانال است. این حجم به تعدادی از گروه های برابر با عامل انشعاب درخت تقسیم می شود. یک عملکرد باقیمانده برای هر یک از اعضای گروه اعمال می شود. عملکرد باقیمانده ، ساختمان اصلی مدل عمیق درختی است. تابع باقیمانده دارای سه عملیات متوالی است که عبارتند از: کانولوشن ، نرمال سازی دسته ای و یک تابع فعال سازی غیرخطی در قالب یک واحد خطی یکسو کننده (ReLU). این تابع باقیمانده برای هر یک از اعضای گروه اعمال می شود. در هر گره درخت ، الگوریتم تقسیم یکسان برای ساخت یک درخت تا رسیدن به تعداد مشخص شده از ارتفاع درخت (l) اعمال می شود.

در مورد مدل درخت موازی ، هنگامی که حجم ورودی به $H \times W \times C$ نگاشت می شود ، به شاخه های تعداد g تقسیم می شود. حجم هر یک از تقسیم ها $H \times W \times C / g$ است. در حال حاضر تعداد g درختان موازی به همان روشی ساخته شده است که مدل درخت منفرد ساخته شده است. هر یک از درختان موازی یک درخت واحد هستند. حجم $H \times W \times C / g$ از طریق یک تابع باقیمانده و به دنبال آن یک ساختار درخت واحد منتقل می شود. بنابراین ، مدل درخت موازی امتداد مدل تک درخت است. شکل 3 معماری مدل درخت موازی را نشان می دهد.

مدل عمیق درختی برای یک محیط شهر هوشمند مفید است. چندین مزیت نسبت به مدلهای عمیق سنتی دارد. مزایا شامل (i) تعادل بین تعداد پارامترها و دقت ، (ii) زمان محاسبات کمتر در مقایسه با سایر مدل های بسیار عمیق ، (III) محاسبه موازی درختان و (IV) تراکم اطلاعات بالا.

خروجی شاخه های یک درخت برای تولید حجم ورودی اصلی به درخت بهم پیوسته است. سپس تابع باقیمانده دیگری بر روی خروجی متصل شده اعمال می شود. برای درختان موازی ، اگر چهار درخت موازی داشته باشیم ، چهار خروجی بهم پیوسته و به دنبال آن توابع باقیمانده مربوطه داریم. سپس خروجی ها مجدداً به هم متصل می شوند تا حجم نهایی حاصل شود.

شکل 2 (b) ساختار مدلهای عمیق را نشان می دهد. هر دو مدل (تک درخت و درختان موازی) در سه مرحله طراحی شده اند. یک فیلتر کانولوشن به حجم ورودی تصویر تبدیل می شود تا به تصویری از تعداد کانالهای از پیش تعیین شده

تبدیل شود. این تصویر جدید وارد مرحله اول می شود. هر مرحله دارای سه ماژول درختی است که در شکل نشان داده شده است. تعداد فیلترهای یک مرحله با تعداد فیلترهای مرحله قبل دو برابر می شود. مرحله اول دارای 64 فیلتر کانولوشن است.

به دلیل ماهیت مستقل ماژول های درخت ، عملیات جمع شدن در این ماژول ها می تواند به طور موازی اجرا شود تا زمان اجرای کلی کاهش یابد.

مراحل باعث انعطاف پذیری در مدل می شوند به این معنا که یک ساختار مرحله ای به هنگام تغییر فرکانس های دیگر به مراحل دیگر بستگی ندارد. افزایش تعداد کانال ها باعث افزایش تعداد پارامترها می شود که ممکن است باعث نصب بیش از حد شود. بنابراین ، ما باید نمونه برداری کنیم. در بین مراحل ، لایه های کانولوشن وجود دارد که وظیفه آنها کاهش تعداد پارامترها است. این کار با انتخاب یک گام برابر با دو انجام می شود که برابر با یک نمونه پایین با ضریب دو است. این نوع نمونه برداری کمتر از یک لایه جمع شدن است. نقشه های خروجی این سه مرحله به ترتیب 32 ، 16 و 8 است. تعداد کانالهای این مراحل به ترتیب 64 ، 128 و 256 کانال است. ابعاد فضایی بین مراحل به نصف کاهش یافت.

برای کاهش ابعاد فضایی ، پس از مرحله نهایی ، از یک لایه متوسط استخر جهانی (GAP) استفاده می شود. بنابراین ، تولید لایه GAP با در نظر گرفتن میانگین در هر نقشه ، ابعادی 1×1 سانتیگراد دارد. بعد از لایه GAP یک لایه softmax وجود دارد. اندازه فیلتر همه لایه های کانولوشن به غیر از لایه آخر 3×3 است. اندازه فیلتر آخرین لایه کانولوشن 1×1 است. لایه GAP در برنامه های پردازش تصویر موفقیت آمیز به اثبات رسید تعداد پارامترهای مدل ها را کاهش می دهد. به عنوان مثال ، GAP به جای استفاده از فیلتر 3×3 ، از فیلتر 1×1 استفاده می کند ، که برای عملکردهای بعدی تعداد پارامترها را یک دهم کاهش می دهد.

پایگاه داده

در این بخش ، ما به پایگاه داده هایی که در آزمایشات استفاده شده اند می پردازیم. برای آزمایش از سه پایگاه داده در آزمایشات استفاده شده است. بانک های اطلاعاتی ORL، FEI و LFW هستند. پایگاه داده ORL برای هر موضوع 10 تصویر دارد و 40 موضوع وجود دارد. بنابراین ، تعداد تصاویر 400 است. تصاویر صورت از نظر زاویه ای ، نور و حالت های چهره متفاوت هستند. تقریباً همه چهره ها دارای نمای قائم از جلو هستند که گاهی اوقات با اندکی چرخش به چپ یا راست مواجه هستند. این پایگاه داده در دانشگاه کمبریج ایجاد شده است. اندازه تصاویر 112 92 92 است.

پایگاه داده FEI بزرگتر از پایگاه داده ORL است. پایگاه داده FEI شامل تصاویر چهره 100 مرد و 100 زن است. هر موضوع دارای 14 تصویر است. بنابراین ، 2800 تصویر چهره وجود دارد. اندازه تصاویر 440 6 640 است. چهره ها یا از جلو و یا از زاویه هستند و دارای عبارات هستند. تصاویر رنگارنگ هستند و چهره ها در مقابل پس زمینه یکنواخت گرفته شده اند.

Labelled Faces in the Wild (LFW) یک پایگاه داده بزرگ است که برای شناسایی چهره ها در محیط های غیرقانونی طراحی شده است . **13233** تصویر از **5749** نفر وجود دارد که **1680** نفر دارای دو یا چند تصویر هستند. اندازه تصاویر **250 250 250** است. مدل های پیشنهادی عمیق درختی با استفاده از پایگاه داده موسسه تحقیقات پیشرفته کانادا (**CIFAR-10**) آموزش داده شدند. سیاست آموزش از به تصویب رسید. فقط ورق زدن و ترجمه با توجه به انجام شد ، و هیچ افزودنی دیگری اعمال نشد. اندازه تصاویر **32 × 32** است.

نتایج و بحث تجربی

برای ارزیابی مدلها در آزمایش از دقت و چگالی اطلاعات استفاده شد. مدلهای آموزش دیده توسط زیرمجموعه کوچکی از یک پایگاه داده مربوطه تنظیم شده اند. در مورد پایگاه داده **ORL** ، سه تصویر برای هر موضوع برای تنظیم دقیق انتخاب شده است. در مورد پایگاه داده **FEI** ، چهار تصویر برای هر موضوع برای راه اندازی مجدد انتخاب شد. از آنجا که پایگاه داده **LFW** تکرار زیادی از تصویر چهره سوژه ندارد ، فقط یک تصویر برای هر موضوع در تنظیم دقیق انتخاب شد. آزمون توسط تصاویر باقیمانده در هر موضوع انجام شد. بنابراین ، تنظیم دقیق و تصاویر آزمایشی از یکدیگر جدا بودند.

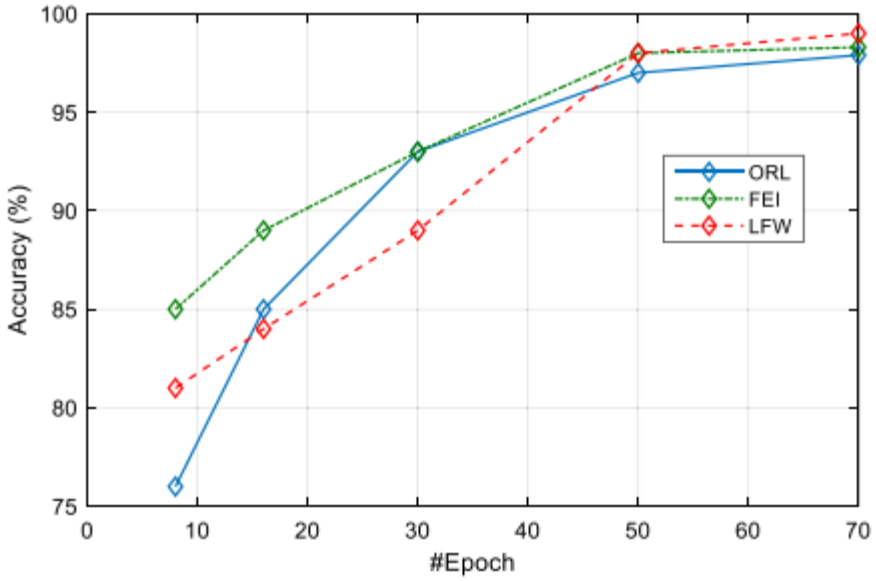

Fig. 4. Accuracy of the single tree model at different epochs using three databases.

پارامترهای مدل های عمیق با استفاده از الگوریتم نزولی شیب مینی دسته ای بهینه سازی شدند. مقدار حرکت به **0.9** تنظیم شد و از دست دادن آنتروپی به عنوان عملکرد از دست دادن استفاده شد. اندازه دسته **128** و میزان یادگیری در ابتدا **0.1** بود و در هر **20** دوره به یک دهم مقدار قبلی کاهش یافت. تعداد عامل شاخه و ارتفاع درخت متنوع بود و به ترتیب به چهار و سه ثابت شد ، زیرا نتایج بهینه را ارائه می دهند.

ابتدا ، ما عملکرد مدلهای عمیق را از نظر دوره ها گزارش می کنیم. به طور معمول ، اگر تعداد دوره ها را افزایش دهیم ، دقت تا تعداد معینی از دوره ها افزایش می یابد. سپس ، دقت یا کاهش می یابد یا همان ثابت می ماند. همه آزمایشات با استفاده از سه پایگاه داده ای که قبلاً ذکر شد انجام شد.

تصویر 4 و 5 به ترتیب دقت مدل عمیق را با استفاده از تک درخت و درختان موازی برای تعداد مختلف دوره نشان می دهد. از هر دو شکل ، تعداد مطلوب دوره 50 می دانیم. در دوره 70 ، اگرچه دقت کمی افزایش یافته است ، اما زمان بیشتری نسبت به دوره 50 طول کشیده است. در مورد برنامه های زمان واقعی ، زمان یک عامل مهم است. در آزمایش های بعدی ، شماره دوره را 50 تعیین کردیم.

هر دو مدل در هر سه پایگاه داده در دوره 50 به دقت مشابهی دست یافتند. این دقت بین 97٪ و 98.5٪ بود. دقت استفاده از پایگاه داده FEI بهتر از استفاده از دو پایگاه داده دیگر بود.

دوم ، ما عملکرد مدل ها را با استفاده از تعداد مختلف فیلترهای اولیه گزارش می دهیم. اگر فیلترهای زیادی داشته باشیم ، تعداد پارامترها افزایش می یابد که پیچیدگی محاسباتی سیستم را افزایش می دهد. انجیر. 6 و 7 به ترتیب دقت مدل ها را با استفاده از درختان منفرد و درختان موازی نشان می دهند. در هر دو مدل ، بالاترین دقت زمانی بدست آمد که تعداد فیلترهای اولیه 64 عدد بود. لطفاً توجه داشته باشید که تعداد فیلترها روی 16 ، 32 ، 64 یا 128 تنظیم شده است. وقتی تعداد فیلترها 128 بود ، دقت ها اندکی کاهش یافت . دقت استفاده از پایگاه داده LFW بهتر از استفاده از دو پایگاه داده دیگر بود.

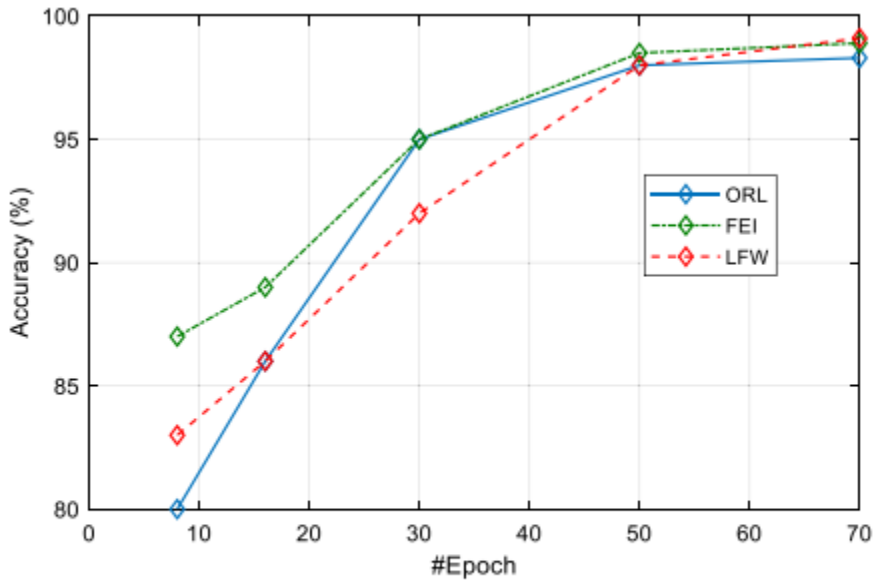

Fig. 5. Accuracy of the parallel tree model at different epochs using three databases.

مدل درخت موازی را می توان با استفاده از تعداد مختلف ماژول های درخت در هر درخت ساخت. اگر تعداد ماژول ها زیاد باشد ، سیستم می تواند در پردازنده های موازی اجرا شود و زمان پردازش کاهش می یابد. ما با استفاده از سه پایگاه داده ، اثر تعداد ماژولها در هر درخت را بر دقت بررسی کردیم. شکل 8 دقت مدل را با تعداد متفاوت ماژول نشان می دهد. تعداد ماژول ها را به 4 ، 6 ، 8 یا 10 تغییر دادیم. از شکل می بینیم که 8 ماژول یا 10 ماژول بهترین دقت را به دست آوردند. به عنوان مثال ، با استفاده از 8 ماژول ، دقت برای پایگاه داده LFW بالاترین بود. با استفاده از 10 ماژول برای این پایگاه داده دقت کمی کاهش یافت. برای دو پایگاه داده دیگر ، دقت در 10 ماژول از 8 ماژول بهتر بود. دقت استفاده از پایگاه داده ORL حداقل بود.

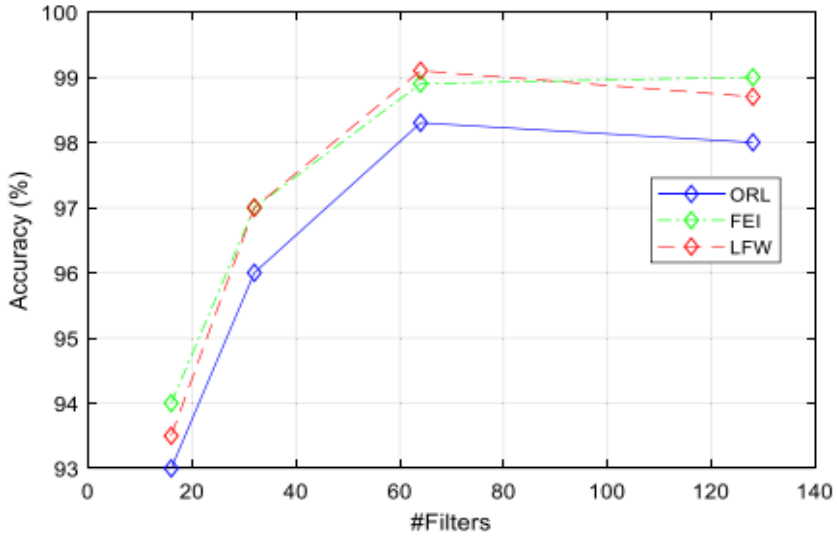

Fig. 7. Accuracy of the parallel tree model for different numbers of initial filters using three databases.

یکی از اهداف اصلی مدل های پیشنهادی عمیق مبتنی بر درخت ، دستیابی به دقت بالا در زمان کمتری است. هنگامی که تعداد پارامترهای سیستم کمتر باشد ، یک سیستم پردازش را در زمان کمتری انجام می دهد. برنامه های شهر هوشمند به یک سیستم برای کار در زمان واقعی و با دقت بالا نیاز دارند. ما مدل های پیشنهادی را از نظر تعداد پارامترها و دقت بررسی کردیم. برای ارزیابی مدلها از معیاری به نام تراکم اطلاعات به عنوان شاخص پایداری بین دقت و تعداد پارامترها استفاده شد. تراکم اطلاعات با دقت بیش از تعداد پارامترها در میلیون تعریف می شود. اگر دقت بالا و تعداد پارامترها کم باشد ، تراکم اطلاعات زیاد است که برای یک مدل مناسب است. ما از نظر تراکم اطلاعات مدلهای مختلفی را مقایسه کردیم. شکل 9 عملکرد مدلهای مختلف را نشان می دهد. چهار مدل معروف عمیق برای مقایسه مقایسه در نظر گرفته شد. برای این آزمایش خاص ، تنظیمات را به گونه ای تغییر دادیم که هم آموزش و هم آزمایش در پایگاه داده CIFAR-10 انجام شود. بنابراین ، حساسیت ها برای تشخیص چهره نیست بلکه برای تشخیص تصویر است. با این حال ، بر معیار چگالی اطلاعات مدل ها تأثیر نمی گذارد. مدل های مقایسه شده ResNext [44] ، شبکه باقیمانده گسترده

[9] ، VGG 16 [6] و MobileNet [45] هستند. مدل VGG بسیار عمیق و دارای دقت بالایی است. با این حال ، پارامترهای زیادی دارد. MobileNet عمیق نیست و از دقت قابل مقایسه ای برخوردار است. ResNext و شبکه گسترده باقیمانده از نظر عمق و دقت بین مدل VGG و MobileNet قرار دارد. با مقایسه این چهار مدل ، شبکه VGG و مدل MobileNet از تراکم اطلاعات بسیار بالاتری برخوردار هستند ، در حالی که دو مدل دیگر دارای تراکم اطلاعات کمتری هستند. مدل های پیشنهادی عمیق مبتنی بر درخت دارای تراکم اطلاعات بالاتری نسبت به چهار مدل قبلی هستند. مدل عمیق مبتنی بر یک درخت از سایر مدل ها بهتر عمل کرد. به عنوان مثال ، مدل پیشنهادی مبتنی بر درخت منفرد دارای تراکم اطلاعاتی 0/58 است ، در حالیکه مدل پیشنهادی متکی بر درخت دارای تراکم اطلاعاتی 0/28 است. تراکم اطلاعاتی VGG Net و MobileNet به ترتیب 0.26 و 0.25 است.

شکل 10 دقت تشخیص چهره با پایگاه داده LFW را با استفاده از مدلهای مختلف شبکه عصبی عمیق نشان می دهد. در اینجا ، ما پنج مدل موجود را که به دقت خوبی رسیده اند و دو مدل پیشنهادی را در نظر گرفتیم. از شکل ، در می یابیم که مدل عمیق مبتنی بر درخت موازی پیشنهادی ، دقت قابل مقایسه ای را با Face Net بدست آورده است [33]. Face Net دارای دقت 99.6٪ بود ، در حالی که مدل مبتنی بر درخت موازی پیشنهادی دارای دقت 99.4٪ بود.

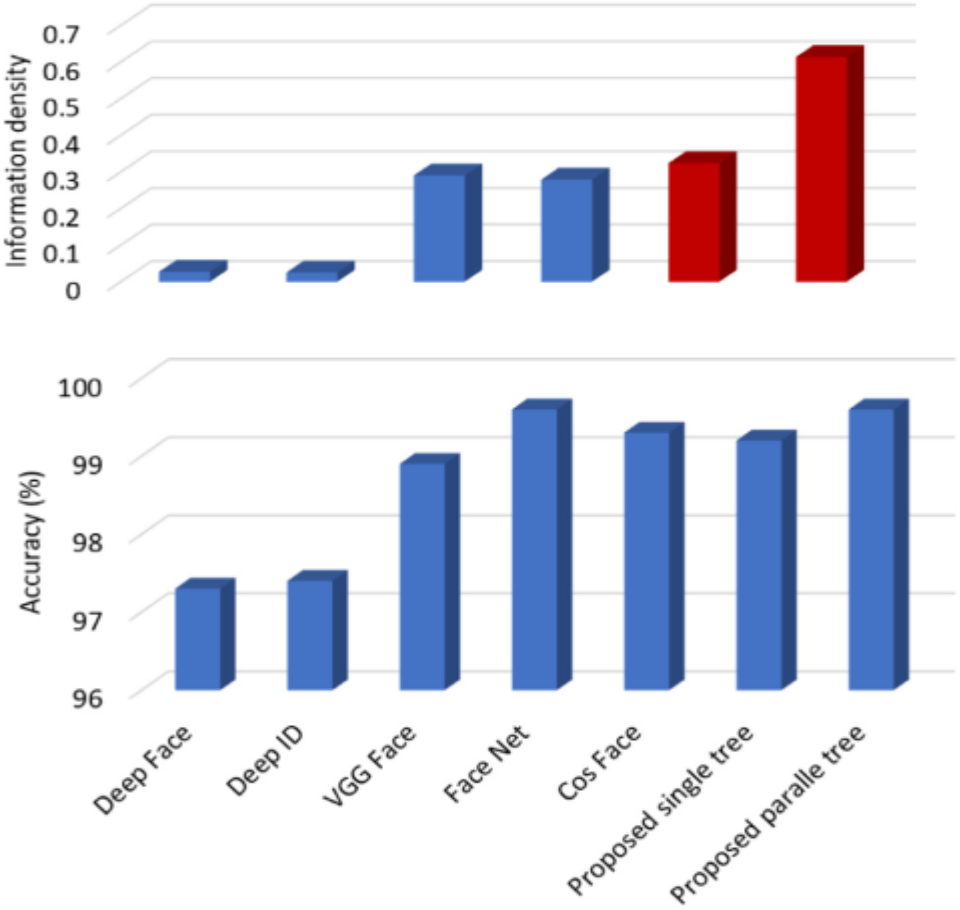

Fig. 10. Accuracy comparison between the models.

Face Net شبکه ای بسیار عمیق است که دارای لایه های مختلفی از تجمع است. بنابراین ، زمان اجرای آن بسیار بیشتر از مدل پیشنهادی است. شبکه **Cos Face** همچنین پارامترهای زیادی دارد. با در نظر گرفتن تمام این جنبه ها می توان گفت که مدل عمیق مبتنی بر درخت برای اهداف تشخیص چهره بسیار کارآمد است.

مدلهای پیشنهادی دارای شاخه هستند ، بنابراین محاسبه را می توان به شاخه ها توزیع کرد تا به صورت موازی اجرا شوند. عملیات در هر شاخه به عملیات شاخه های دیگر وابسته نیست. بنابراین ، زمان محاسبه در مدل پیشنهادی کمتر است. تعداد کل پارامترها در مدل عمیق مبتنی بر یک درخت و در مدل عمیق درخت پایه موازی به شرح زیر است.

$$O\left(n.c^2.\left(\frac{1}{b^L}+m^2.\left(\frac{1-b^L}{1-b}+\frac{1}{b^L}\right)\right)\right)$$

برای مدل مبتنی بر تک درخت

$$O\left(n.c^2.\left(b^L+m^2.\left(b^{L+1}-2-b\right)\right)\right)$$

برای مدل درختی موازی

که **n** تعداد فیلترهایی است که اندازه آنها **m × m**، **b** و **L** به ترتیب عامل انشعاب و ارتفاع درخت است و **c** کانالهای تصویر است.

نتیجه گیری

مدل های عمیق مبتنی بر درخت برای تشخیص چهره در این مقاله ارائه شده است. دو مدل از مدل پیشنهادی وجود داشت: تک درخت و درختان موازی. در مدل درختان موازی ، چندین درخت منفرد به طور موازی مرتب شده اند. از تابع باقیمانده به عنوان سنگ بنای کل معماری استفاده شد. آزمایشات با استفاده از سه پایگاه داده چهره در دسترس عموم انجام شد. مدل های پیشنهادی با استفاده از این پایگاه ها به دقت **99**٪ دست یافتند. تراکم اطلاعات مدل پیشنهادی مبتنی بر تک درخت نزدیک به **0.6** بود که برای یک مدل عمیق عالی در نظر گرفته می شود. در مقایسه با مدلهای عمیق موجود ، مدلهای پیشنهادی دارای دقت قابل مقایسه با تعداد کمتری پارامترها بودند. این یافته ها ثابت می کند که مدل های پیشنهادی می توانند به طور موثر به عنوان یک سیستم تشخیص چهره در زمان واقعی برای اهداف امنیتی استفاده شوند. مسیرهای آینده کار پیشنهادی می تواند به شرح زیر باشد. اول ، مدلهای پیشنهادی را می توان به جای ترکیبهای استفاده شده در مدلها ، شامل اجزای بیشتری مانند توصیفگرهای جمع شده محلی [**46**] کرد. دوم ، پایگاه داده های بیشتری می تواند در آزمایشات گنجانده شود. سوم ، مدل های پیشنهادی می توانند در برنامه های دیگری مانند تشخیص جنسیت و تشخیص احساسات صورت نیز مورد استفاده قرار گیرند.