تایید هویت در دستگاه‌های موبایل با استفاده از تایید چهره و تشخیص تصویر ID

Xing Wua,b,, Jianxing Xua, Jianjia Wanga, Yufeng Lia, Weimin Lia, Yike Guoa*

چکیده

احراز هویت در عصر دیجیتال اهمیت زیادی دارد که در آن اطلاعات ID معمولا در زمینه‌های مالی، بیمه، حمل و نقل و سایر زمینه‌ها استفاده می‌شود. چالش‌های احراز هویت در تایید کارت ID ارائه‌شده توسط کاربر و استخراج اطلاعات از کارت ID کاربر قرار دارد. برای مقابله با این چالش، یک چارچوب احراز هویت پیشنهاد شده‌است، که می‌تواند اطلاعات شخصی را از طریق تایید چهره و تشخیص تصویر ID استخراج و تایید کند. احراز هویت توسط مدل تایید چهره پیشنهادی محقق می‌شود که طبقه‌بندی چهره InceptionResNet نامیده می‌شود. رادیو اروپای آزاد از یک ساختار Inception-ResNet برای اطمینان از استخراج ویژگی خوب با هدف تایید دقیق چهره استفاده می‌کند. علاوه بر این، یک روش قوی استخراج کارت ID به نام نگاشت ویژگی تبدیل مورفولوژی (MTFM)برای استخراج اطلاعات ID پیشنهاد شده‌است. نتایج تجربی نشان می‌دهد که Ifree پیشنهادی و MTFM عملکرد بهتری نسبت به روش‌های پیشرفته هم در تایید چهره و هم در استخراج ID دارند.

۱. مقدمه

دستگاه‌های تلفن همراه به طور گسترده برای انجام عملیات از راه دور مورد استفاده قرار می‌گیرند که از قابلیت حمل آن‌ها بهره می‌برند [۱]. کاربردهای تلفن همراه، به ویژه پرداخت‌های تلفن همراه، برای زندگی روزمره ما ضروری هستند. در همین حال، پدیده سرقت هویت (استفاده غیرمجاز از سایر اطلاعات شناسایی افراد یا تقلب تایید شده)به سرعت در بسیاری از کشورها رشد کرده‌است. در بسیاری از سناریوهای تعاملی، به ویژه سناریوهای مالی، اطلاعات هویت کاربر باید در صورت سرقت هویت تایید شود [۲].

سوالی که به طور طبیعی دنبال می‌شود این است که ما می‌توانیم به طور خودکار اطلاعات هویت کاربر را بررسی کنیم تا از سرقت هویت جلوگیری کنیم. برای رسیدن به این هدف، ما باید اطلاعات هویت را از عکس گرفته‌شده توسط کاربر استخراج کنیم، و تایید کنیم که اطلاعات ID متعلق به کاربر است. هر شهروند چینی دارای یک کارت شناسایی است که حاوی اطلاعات هویت دارنده و پرتره چهره او است که در شکل 1a نشان‌داده شده‌است.

شکل ۱. (*a*)نمونه کارت *ID* شهروندان؛(*b*)عکس نمونه‌ای که باید توسط کاربر بر روی دستگاه‌های موبایل برای تایید گرفته شود

با توجه به طراحی کارت‌های شناسایی چینی، یک چارچوب یکپارچه توسعه‌یافته است که می‌تواند احراز هویت شخصی را به طور خودکار با تایید چهره و تشخیص کارت شناسایی انجام دهد. به طور خاص، این چارچوب تنها به یک عکس گرفته‌شده توسط کاربر به عنوان ورودی نیاز دارد و سپس اطلاعات شناخته‌شده و نتیجه تایید چهره را به عنوان خروجی برمی گرداند. انتظار می‌رود که کاربر از یک دستگاه تلفن همراه برای گرفتن عکس با کارت شناسایی دستی خود استفاده کند. عکس نمونه در شکل ۱ (b)نشان‌داده شده‌است. هنگامی که این عکس در سیستم پیشنهادی آپلود می‌شود، نتیجه احراز هویت به صورت خودکار در عرض چند ثانیه به دست خواهد آمد.

این مدل پیشنهادی از ساختار نگاشت ویژگی چهره جدید با نمایش ویژگی قوی برای تایید چهره استفاده می‌کند. روش MTFM پیشنهادی قادر به استخراج اطلاعات ID با دقت بالا می‌باشد. بنابراین، رادیو اروپای آزاد و MTFM برای دستیابی به نتایج امیدوار کننده در تایید هویت با هم ترکیب می‌شوند.

ادامه این مقاله به صورت زیر سازماندهی شده‌است. در بخش ۲، کارهای مرتبط مورد بحث قرار می‌گیرند. در بخش ۳، توصیف دقیقی از چارچوب احراز هویت پیشنهادی ارائه شده‌است. در بخش ۴، جزییات پیاده‌سازی و نتایج تجربی نشان‌داده شده‌است. بخش ۵ نتیجه کار ما است.

۲. کارهای مرتبط

استخراج اطلاعات هویت به خواندن متن در تصاویر کارت شناسایی اشاره دارد. این موضوع به زمینه تحقیقاتی شناسایی و تشخیص متن تعلق دارد که در چند دهه اخیر مورد مطالعه قرار گرفته‌است [۳، ۴، ۵]. تحقیقات اخیر بر روی جهت افقی استاندارد کارت ID تمرکز دارند و مساحت کارت ID کل تصویر را اشغال می‌کند. بنابراین بیشتر روش‌های تشخیص مبتنی بر افکنش افقی و افکنش عمودی هستند [۴، ۱۰]. هیچ تحقیق قبلی برای تشخیص تصویر ID اریب در پس‌زمینه‌های طبیعی وجود ندارد. با توجه به فقدان مجموعه داده‌های کارت شناسایی، یک روش مبتنی بر مورفولوژی تصویر برای تشخیص متون بر روی کارت‌های شناسایی پیشنهاد شده‌است. وظیفه شناخت شخصیت چینی توسط روش یادگیری عمیق مبتنی بر شبکه‌های عصبی کانولوشن (CNNs)به دست می‌آید. برنامه ResNet [۱۱] یک معماری جدید در مدل‌های عمیق CNNاست، که پیشرفت‌های زیادی را در حوزه دید کامپیوتری ایجاد کرده‌است. - ۵۰ResNet در روش تشخیص کاراکتر ادغام می‌شود.

تصدیق چهره یکی از حوزه‌های فرعی تشخیص چهره است. این روش بر بررسی این موضوع تمرکز دارد که آیا دو تصویر چهره متعلق به یک شخص هستند یا نه که به عنوان تایید یک به یک چهره نیز شناخته می‌شود [۱۲]. تشخیص چهره بر انتخاب این که این فرد در مجموعه داده‌های موجود چه کسی است تمرکز دارد. همچنین شناسایی چهره یک به چند

نامیده می‌شود [۱۳]. در چارچوب احراز هویت پیشنهادی، تایید چهره ضروری است. به منظور اطمینان از این که اطلاعات هویت ادعا شده به کاربر تعلق دارد، لازم است که چهره‌ها در عکس ID ضبط‌شده بررسی شوند تا از حمله کنندگان جعلی جلوگیری شود.
روش‌های مدرن تشخیص چهره [۱۴، ۱۵] اغلب CNNها را به عنوان استخراج‌کننده ویژگی قوی در نظر می‌گیرند. یک CNNبا یک مجموعه داده تصویر چهره بزرگ در [۱۶] آموزش داده شده‌است و به عنوان یک استخراج‌کننده ویژگی برای تایید چهره کار می‌کند. FaceNet یک نگاشت یکپارچه برای تشخیص چهره و خوشه‌بندی است [۱۲]، و به طور مستقیم یک تصویر چهره را به یک فضای اقلیدسی فشرده نگاشت می‌کند که در آن فواصل مستقیما با یک معیار شباهت چهره متناظر هستند. این ایده توسط latum پیشنهادی به امانت گرفته شده‌است که در آن استخراج کننده ویژگی با جایگزین کردن GoogLeNet از FaceNet با Inception-ResNet [۱۷] به روز رسانی می‌شود و آن را برای استخراج ویژگی کارآمدتر می‌سازد. یک چارچوب از سبک CNN در [۱۸] ارائه شده‌است تا یک دسته‌بندی فشرده بر روی داده‌های چهره در مقیاس بزرگ با برچسب‌های بزرگ پر سر و صدا را یاد بگیرد. یک چارچوب چند وظیفه‌ای آبشاری عمیق، MTCNN، در [۱۹] برای بهره‌برداری از همبستگی ذاتی در کارهای رو در رو پیشنهاد شده‌است. مزایا و معایب تحقیقات قبلی وجود دارد، نقاط قوت آن‌ها اتخاذ شده و نقاط ضعف آن‌ها برای توسعه چارچوب تایید هویت یکپارچه نادیده گرفته شده‌است.

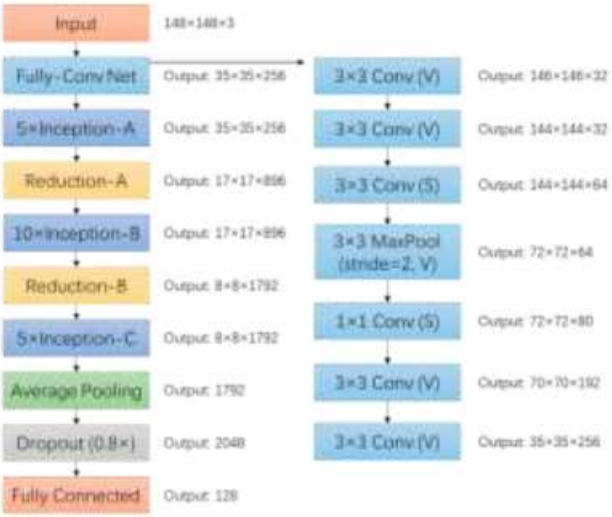
۳. چارچوب احراز هویت

۳.۱ مرور چارچوب

به منظور تایید اطلاعات هویت کاربر قبل از سرقت هویت، وظیفه کلیدی اطمینان از این است که اطلاعات هویت ارائه‌شده توسط کاربر به او تعلق دارد. رادیو اروپای آزاد برای تحقق تاییدیه و تضمین امنیت اجرا می‌شود. بخش تایید چهره با قضاوت شباهت بین چهره کاربر و چهره بر روی کارت ID او انجام می‌شود، که در شکل. پس از آن متون روی کارت شناسایی استخراج می‌شوند. برای مقابله با کارت شناسایی چوله شده در عکس، یک روش تصحیح چوله براساس MTCNN پیشنهاد شده‌است. علاوه بر این، یک روش تشخیص متن براساس مورفولوژی تصویر برای تشخیص متون بر روی تصاویر ID اجرا می‌شود. سپس از الگوریتم تصویر برای بخش‌بندی هر کاراکتر از کارت ID استفاده می‌کنیم. در نهایت، یک مدل عمیق CNNبرای شناسایی بیش از ۳۵۰۰ شخصیت چینی ساخته شده‌است که اغلب شخصیت‌های چینی مورد استفاده را پوشش می‌دهند. با پیشنهاد رادیو اروپای آزاد و ام تی اف ام، هویت شخصی تایید می‌شود تا از سرقت هویت جلوگیری شود.

۳.۲ فاز تایید صورت

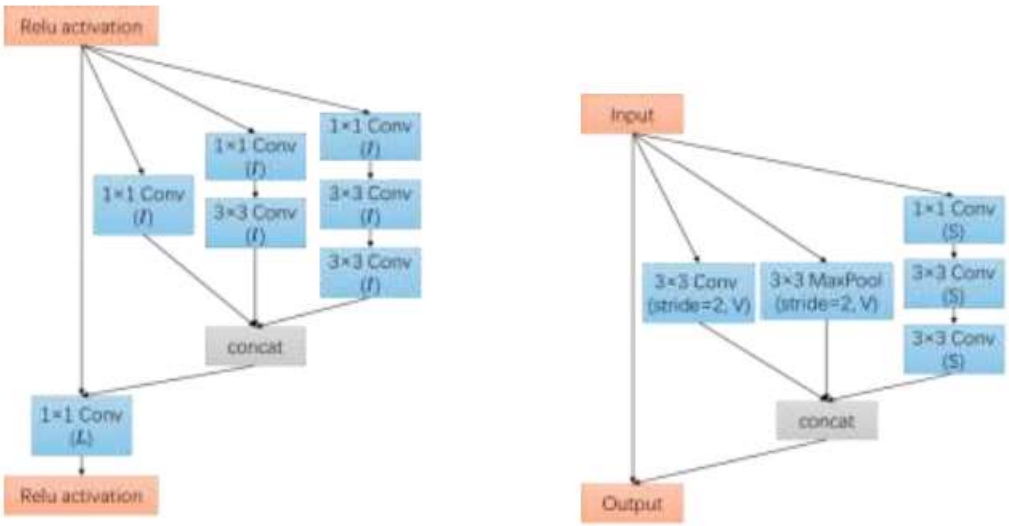
مرحله تایید چهره ابتدا در چارچوب پیشنهادی برای تضمین قابلیت اطمینان اطلاعات هویت انجام می‌شود. تشخیص چهره پیش‌نیاز تایید چهره است، چون عکس گرفته‌شده شامل دو چهره است که در شکل ۱ b نشان‌داده شده‌است. هنگامی که دو چهره شناسایی شوند، با یکدیگر مقایسه خواهند شد.



شکل ۲. ساختار کلی مدل جایگزاری چهره

تصدیق چهره می‌تواند یک مشکل طبقه‌بندی در قضاوت شباهت بین دو چهره باشد. بنابراین معماری شبکه عصبی کانولوشن عمیق تشریح شده در FaceNet را بهبود می‌دهیم تا به بیان ویژگی چهره دست یابیم. ساختار ادراک GoogLeNet با Inception-ResNet برای ساخت مدل Ifree و ایجاد ویژگی‌های چهره بهتر (برای مثال تعبیه صورت)جایگزین شده‌است. هنگامی که دسته‌بندی چهره به دست می‌آید، سوال به اندازه‌گیری فاصله بین دو دسته‌بندی چهره تبدیل می‌شود. نتیجه تایید با آستانه گذاری فاصله بدست می‌آید.

جزییات شبکه کامل - Conv در شکل ۲ نشان‌داده شده‌است. اندازه تصویر ورودی ۱۴۸ * ۱۴۸ با ۳ کانال است. ساختار شروع و کاهش در شکل ۳ (a)و شکل ۳ (b)نشان‌داده شده‌است. " 5Inception - A " به این معنی است که ۵ ساختار اولیه یکسان در واحد وجود دارد . V به معنای لایه‌گذاری "Valid" و لایه‌گذاری "مشابه" S است. شکل خروجی هر لایه در سمت راست هر لایه خلاصه شده‌است.

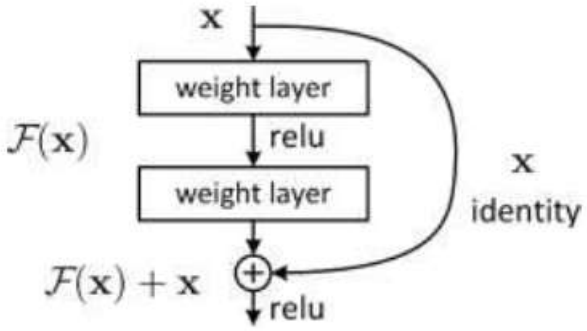


شکل ۳. (a)ساختار دریافتی در مدل جایگزاری چهره؛ (b)ساختار کاهشی در مدل جایگزاری چهره

در شکل ۳ (a)، "I" نشان‌دهنده بخش ادراک مختلف و "L" نشان‌دهنده طول خروجی است. I = 32، L = 256 در هنگام آغاز - A، I = 128، L = 896 هنگام آغاز - B و I = 192 L = 1792 هنگام آغاز - C.
شکل ۳ (b)ساختار کاهشی را نشان می‌دهد که در آن "V" نشان‌دهنده پوشش "Valid" و "S" نشان‌دهنده پوشش "مشابه" است. این یک واحد نمونه‌برداری در مدل است. این واحدها به منظور ساخت مدل جایگزاری چهره متصل شده‌اند.
پس از ساخت مدل عمیق، سپس بر روی مجموعه داده چهره آموزش داده می‌شود تا ویژگی‌های چهره استخراج شود. ۱ بین لایه‌های کانولوشن استاندارد اضافه می‌شوند. ابعاد نگاشت خروجی بر روی ۱ - ۱۲۸ تنظیم شده‌است. خروجی یک نمایش ویژگی از داده‌های ورودی است. با بهره‌گیری از مدل عمیق، تصاویر چهره به طور متراکم در فضای ۱۲۸ بعدی قرار می‌گیرند که منجر به نمایش ویژگی کارآمد می‌شود.

۳. ۳ تشخیص تصویر ID در MTFM

پس از تایید چهره کاربر، اطلاعات کارت شناسایی او باید برای احراز هویت استخراج شود. در اغلب موارد، عکس‌های گرفته‌شده توسط کاربران افقی نیستند، که باعث نرخ تشخیص متن پایین می‌شوند. این ما را به سمت اصلاح افقی کارت شناسایی پیش از تشخیص هدایت می‌کند. روش اصلاح چوله تصویر پیشنهادی براساس MTCNN یک روش بسیار موثر برای اصلاح افقی تصاویر کارت شناسایی است. صورت هم تراز شده برای محاسبه جهت کارت و تعیین مکان محل کارت مورد استفاده قرار می‌گیرد، و تصویر به صورت افقی تصحیح شده را می توان با چرخش زاویه متناظر با تصویر اصلی به دست آورد.
عملیات مورفولوژی تصویر به طور گسترده‌ای در پردازش تصویر مورد استفاده قرار می‌گیرد، که به استخراج مولفه‌های تصویری که برای بیان شکل ناحیه موردنظر در یک تصویر معنادار هستند، کمک می‌کند. با توجه به عملیات مورفولوژی تصویر، تشخیص تصویر بعدی می‌تواند ویژگی‌های شکل بسیار متمایز شی هدف را به دست آورد.
با تشخیص چهره کارت ID، یک تصویر ID کامل بدون پس‌زمینه را می توان با توجه به پیکربندی استاندارد کارت ID چینی به دست آورد. سپس عملگر کلاه بالایی (مقدار اختلاف بین تصویر اصلی و تصویر عملیات نزدیک)بر روی تصویر اعمال می‌شود تا ناحیه پیش‌زمینه سیاه به دست آید. سپس عملیات نزدیک برای کاهش فضای بین کلمات استفاده می‌شود. سپس روش Otsu به آستانه تصویر اعمال می‌شود. سپس با اعمال عملیات بسته و فرسایش، فیلدهای متنی کارت شناسایی جدا می‌شوند. در نتیجه خطوط متنی شناسایی می‌شوند. پس از تشخیص خطوط متن، به منظور به دست آوردن یک کاراکتر واحد در هر خط متن، از روش افکنش تصویر برای بخش‌بندی کاراکترها استفاده می‌شود؛ در نتیجه یک تصویر تک کاراکتری پس از بخش‌بندی متن به دست می‌آید.
در مرحله شناخت کاراکتر، به دلیل تعداد زیاد شخصیت‌های چینی، برخی از مدل‌های شبکه عصبی کانولوشن برای دستیابی به بهترین وظیفه شناسایی بیش از ۳۵۰۰ کاراکتر مورد آزمایش قرار می‌گیرند. ما مدل ResNet - ۵۰ را برای مقابله با این مشکل پیاده‌سازی می‌کنیم و آن به بهترین نرخ تشخیص دست می‌یابد. ResNet با ساختار سنتی CNNبا توجه به ماژول اتصال میان‌بر که در شکل ۴ نشان‌داده شده‌است، متفاوت است.



شکل ۴. بلوک اتصال کوتاه *ResNet*

مدل‌های عمیق CNNپیچیده هستند زیرا آن‌ها اغلب میلیون‌ها پارامتر دارند. پیچیدگی همچنین تناسب مدل‌های عمیق CNNبا داده‌های آموزشی را دشوار می‌سازد. با افزایش عمق مدل، پارامترها به سرعت افزایش می‌یابند. بلوک اتصال کوتاه برای کاهش پارامترها و در عین حال افزایش عمق شبکه طراحی شده‌است. همچنین، مدل به خوبی طراحی‌شده قابل آموزش است. نتایج نشان می‌دهد که این ایده مختصر و موثر است.

۴. اجرا و آزمایش‌ها چارچوب پیشنهادی

۴.۱. مجموعه داده‌های چهره و شخصیت

حالات برچسب دار در بیابان (LFW)یک مجموعه آزمایشی عملی برای تایید چهره است [۲۰]. این روش عمدتا برای مطالعه تشخیص چهره در موقعیت‌های نامحدود استفاده می‌شود. پایگاه‌داده LFW عمدتا تصاویر را از اینترنت جمع‌آوری می‌کند، نه از آزمایشگاه‌ها. این تصویر حاوی بیش از ۱۳،۰۰۰ تصویر چهره است. هر تصویر با نام فرد مربوطه شناسایی می‌شود. در میان آن‌ها، ۱،۶۸۰ نفر با بیش از یک تصویر مطابقت دارند، یعنی حدود ۱،۶۸۰ نفر بیش از دو چهره دارند. VGGFace2 [۲۱] مجموعه داده شناسایی چهره در مقیاس بزرگ دیگری است که شامل بیش از ۹۰۰۰ هویت و ۳.۳ میلیون چهره است که طیف گسترده‌ای از قومیت‌ها، لهجه‌ها، حرفه‌ها و سنین مختلف را پوشش می‌دهد. کاسا - وب فیس [۲۲] نیز یک مجموعه داده بزرگ عمومی با حدود ۱۰۰۰۰ نفر و بیش از ۵۰۰۰۰۰ تصویر است. مدل جایگزاری چهره ما بر روی این بزرگ‌ترین مجموعه داده آموزش داده شده‌است.

به منظور شناخت خوب کاراکترها، یک مجموعه داده از کاراکترها مورد نیاز است. با این حال، به دست آوردن مجموعه داده کارت ID دشوار است، زیرا کارت‌های ID دارایی شخصی هستند. برای رسیدن به این هدف، ما از روش تولید تصویر متنی برای تولید مجموعه داده متنی استفاده‌شده در آزمایش استفاده می‌کنیم. مجموعه داده تولید شده شامل ۳۵۶۲ دسته کاراکتر، شامل ۳۵۰۰ کاراکتر چینی، ۵۲ کاراکتر انگلیسی بزرگ و کوچک، و ۱۰ رقم عربی است. مجموعه داده شامل ۳۶۳،۲۱۰ تصویر متنی از داده‌های تک کاناله است که هر کدام ۳۲ پیکسل اندازه دارند. در نسبت ۵: ۱، مجموعه داده کل به طور تصادفی به یک مجموعه آموزش و یک مجموعه آزمون تقسیم می‌شود. در عین حال، ما از عملیات بهبود داده‌ها در هنگام ساخت داده‌های آموزشی نیز استفاده می‌کنیم. چرخش تصادفی تصویر، برش تصادفی، تبدیل پرسپکتیو، تغییر شکل و دیگر موارد به هر تصویر اعمال می‌شود که تا حد زیادی داده‌های آموزشی را افزایش می‌دهد. بهبود داده‌ها همچنین باعث می‌شود که مجموعه آموزشی شامل ویژگی‌های تصویر بیشتری باشد، که مدل آموزش‌دیده را قوی‌تر می‌کند.

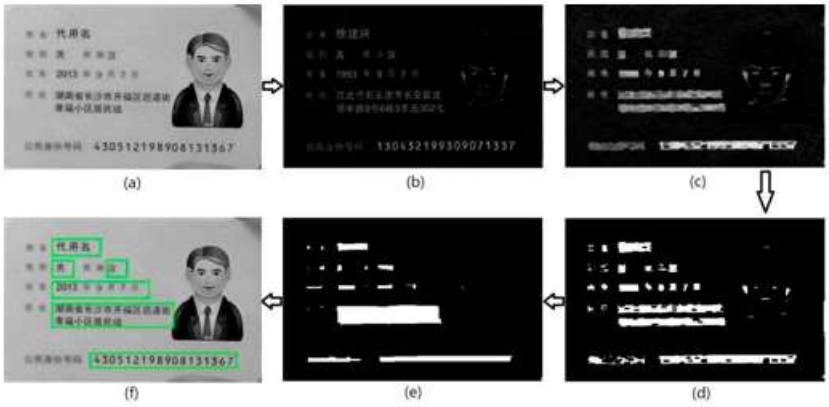
۴.۲ تایید صورت

به منظور قابل آموزش ساختن مدل جایگزاری چهره، پس از اتصال کامل لایه، یک لایه SFTmax دنبال می‌شود. تابع اتلاف به صورت اتلاف آنتروپی متقابل تعریف می‌شود، و سپس از پس انتشار برای به روز رسانی پارامترهای مدل استفاده می‌شود. پس از آموزش بر روی مجموعه داده‌های CASIA - WebFace و VGFace2، می‌توانیم یک مدل به خوبی آموزش‌دیده به دست آوریم.

جدول ۱: نتایج دقت تست در مجموعه داده‌های عمومی

نرخ اعتبارسنجی در LFW	مجموعه داده های آموزشی	شبکه
89.4%	CASIA-WebFace	FaceNet
96.1%	CASIA-WebFace	IRFE
90.3%	VGGFace2	FaceNet
97.5%	VGGFace2	IRFE

برای آزمون‌های قابل‌مقایسه‌ای، نتایج تایید چهره بر روی مجموعه داده مقیاس بزرگ LFW آزمایش می‌شود. فیس نت اصلی اجرا شده‌است. عملکرد صحت سنجی در جدول ۱ نشان می‌دهد که روش پیشنهادی نمایش ویژگی چهره را افزایش داده و از روش‌های تایید سطح پیشرفته با نرخ اعتبار ۹۷.۵ ٪ بهتر عمل می‌کند.



شکل ۵. پردازش محلی‌سازی متن

۴.۳ تشخیص متن

برای تصاویری که در ناحیه ID قرار گرفته‌اند، یک سری تحولات مورفولوژیکی برای تشخیص نواحی متنی در کارت ID اجرا می‌شود. اول، عملیات کلاه سیاه مورفولوژیکی را بر روی تصویر اعمال می‌کنیم تا نواحی تاریک را در یک پس‌زمینه روشن پیدا کنیم (شکل ۵ (b)). دوم، هسته مستطیلی برای اعمال عملیات نزدیک به منظور از بین بردن شکاف بین حروف استفاده می‌شود (شکل ۵ (c)). سوم، آستانه Otsu برای به دست آوردن تصویر دودویی اعمال می‌شود (شکل ۵ (d)). در نهایت عملیات خاتمه و فرسایش دیگری برای به دست آوردن ناحیه متنی نهایی اعمال می‌شود.

دو شبکه کانولوشن اجرا شده و نتایج تجربی با روش‌های قبلی مقایسه شده‌اند. همانطور که در جدول ۲ نشان‌داده شده‌است، مقایسه دقت تشخیص تصویر ID لیست شده‌است. عملکرد تشخیص در مجموعه داده جمع‌آوری‌شده ما که شامل ۴۷ تصویر کارت شناسایی است، مورد آزمایش قرار گرفته‌است.

LeNet - 5 [۲۳] با اعمال تابع فعالسازی ReLU به جای تابع سیگموئید اجرا می‌شود. سایر جزئیات پیاده‌سازی LeNet - 5 همان چیزی است که در مقاله اصلی توضیح داده شده‌است.

جدول ۲: مقایسه نتایج تشخیص

نام	روش	دقت
Fang et al. [9]	SVM	71.1%
LeNet-5	CNN	73.9%
Cheng et al. [15]	Local Similarity Voting	92.0%
MTFM	Deep CNN	**96.83%**

کارهای قبلی مانند [۱۵] تنها می‌توانند با شناسایی PIN سر و کار داشته باشند زیرا این مشکل بسیار ساده‌تر است: اعداد شناسایی شخصی تنها ده دسته از کاراکترها را دارند. هنگام مواجهه با شناخت شخصیت چینی، که بیش از ۳۵۰۰ مقوله شخصیت دارد، این ماموریت دشواری است. با بهره‌گیری از توسعه یادگیری عمیق، به راحتی می توان آن را با MTFM به خوبی طراحی‌شده حل کرد. شبکه ResNet یک مدل عمیق در دید کامپیوتری است و به صورت انتها به انتها قابل آموزش است و بسیاری از مشکلات بینایی را بهبود می‌بخشد. برنامه - ۵۰ResNet به طرح ما اعمال می‌شود و به بهترین نتیجه تشخیص در میان کارهای مشابه دست می‌یابد. با توجه به جدول ۲، ما می‌توانیم با اطمینان نتیجه‌گیری کنیم که MTFM پیشنهادی بهتر از روش‌های تشخیص کاراکتر سطح بالا با بالاترین دقت ۹۶.۸۳ ٪ عمل می‌کند.

۵. بحث و نتیجه‌گیری

برای جلوگیری از سرقت هویت، طرح احراز هویت سنتی دارای یک فرآیند بسیار خسته‌کننده است. برای مقابله با این چالش، یک چارچوب احراز هویت شخصی خودکار پیشنهاد شده‌است، که می‌تواند هویت را از طریق یک عکس استخراج و تایید کند. معمولا کاربر باید ابتدا یک تصویر ID بگیرد تا هویت خود را ارائه دهد، و سپس تشخیص پویایی با چشمک زدن یا سر تکان دادن تایید می‌شود. در نهایت، تصویر چهره گرفته‌شده توسط کاربر و تصویر ID مقایسه می‌شوند تا اطمینان حاصل شود که هویت درست است. تأیید هویت توسط IRFE انجام می شود که در آن جاسازی چهره با استفاده از Inception-ResNet افزایش می یابد.روش شناسایی ID پیشنهادی MTFM یک نتیجه استخراج اطلاعات دقیق را برای بهبود دقت تأیید هویت ارائه می دهد.. روش تشخیص ID پیشنهادی MTFM یک نتیجه استخراج اطلاعات دقیق را برای بهبود دقت تایید هویت فراهم می‌کند. این روش می‌تواند با شرایط پیچیده‌ای مانند تصویر ID مورب، روشنایی و غیره سر و کار داشته باشد. نتایج تجربی نشان می‌دهند که چارچوب پیشنهادی برای تضمین قابلیت اطمینان تایید هویت، بهتر از روش‌های سطح بالا با دقت بالا عمل می‌کند.