پاییز ۱۴۰۰، دوره ۶، شماره ۳، پیاپی ۲۲
DOR :20.1001.1.2538161.1400.6.3.7.9

دریافت مقاله: ۱۳۹۹/۰۷/۰۴
پذیرش مقاله: ۱۳۹۹/۱۲/۲۰

نوع مقاله: پژوهشی

ارائه کاربردی جدید در یافتن اشخاص گم شده در دوربین‌های نظارتی با استفاده از یادگیری عمیق

علی‌محمد لطیف*
دانشیار دانشکده مهندسی کامپیوتر- دانشگاه یزد- یزد- ایران
پست الکترونیکی: alatif@yazd.ac.ir

کامبیز طباطبائی اردکانی
کارشناس ارشد مهندسی فناوری اطلاعات-گرایش سیستم‌های چندرسانه‌ای
پست الکترونیکی: Kambiz.tabatabaei@gmail.com

چکیده

در ایـــن مقاله کاربـــردی جدید جهت یافتـــن چهره در تصاویر دوربین‌های نظارتی با استفاده از شبکه‌های عصبی عمیق ارائه گردیده است. در مکان عمومی در صورتی که یـــک کودک مفقود گردد با توجه بـــه این که در اکثر موارد والدیـــن روی تلفن همراه تصویری از کـــودک خود دارند می‌توانند تصویر کودک خود را به این سامانه تحویل دهند و سامانه با دریافت تصاویر دوربین‌های نظارتی واقع در محل، آخرین مکانی که کودک دیده شده است را مشخص نماید. برای یافتن چهـــره از الگوریتم vggface2 در محیط پایتورچ استفاده شده است. نتایج آزمایش‌ها نشان می‌دهد دقت حاصل شده برای بازیابی تصاویر چهره حدود ۹۹ و برای قاب‌های ویدئو در حدود ۹۶/۶ درصد است و همچنین ایـــن روش در برابر روش‌هایی همچـــون الگوی دودویی محلی که با پردازش تک تک پیکســـل‌ها در ارتباط است از سرعت قابل توجهی برخوردار است.

واژه‌های کلیـــدی: بازیابی تصویر چهره، شـــبکه‌های عصبی عمیق جفتی، دوربین‌های نظارتی

مقدمه

عمل بازیابی تصویـــر چهره با اهداف مختلفی از جمله تاییـــد هویت افراد، شناســـایی افراد، با اســـتفاده از پایگاه داده‌ای از تصاویر چهره اشـــخاص صـــورت می‌گیرد که باعث کاهش هزینه‌های جانبـــی از قبیل عدم صدور کارت پرسنلی، کاهش نیروی انسانی و زمان برای جستجوی فرد مجـــرم، پیدا کردن افراد گم شـــده در اماکن عمومی با یک نمونه تصویر از شخص می‌شود.

یکی از مواردی که در عمل بازیابی و شناســـایی چهره افراد حائز اهمیت اســـت دقت بازیابی است که هرچه تعداد نمونه‌های پایگاه داده بیش‌تر باشـــد دقت بالاتر ولی زمان پردازش هم به موازات آن افزایش می‌یابد.

در تحقیق‌های انجام شـــده در دهه گذشته هنوز مسئله بازیابی تصویر چهره دارای مسائل و پیچیدگی‌های خاصی از قبیـــل نورپـــردازی، حالت چهره و پوشـــیدگی صورت می‌باشـــد. از این رو عمل بازیابی برای تحقق افزایش دقت بازیابی، مجبور به ذخیره تصاویر متعددی از شـــخص در حالت‌های مختلف اســـت که با هزینه فضای اشـــغالی زیاد حافظـــه و افزایش زمـــان پردازش روبرو اســـت. همچنین

* نویسندهٔ مسئول

جمع‌آوری داده در این حالت کار دشواری است. از این رو موضوعی که مطرح گردید موضوع روش آموزش شبکه عصبی با استفاده از تعداد محدودی تصویر از هر فرد است که اصطلاحا آن را یادگیری یک‌باره[1] می‌نامند، در این روش محدودیتی برای حالت چهره و نورپردازی در تصاویر آزمون وجود ندارد. با استفاده از این طرح کارآیی روش‌های قبلی کاهش یافت و بسیاری از روش‌های قبلی با شکست مواجه شد. بنابراین این روش بیش‌تر مورد استقبال قرار گرفت. پیش از این برای مرتفع نمودن مسئله بازیابی تصویر چهره با دقت بالا، روش‌های مختلفی از جمله روش تمرکز بر نقاط منحصر به فرد صورت مانند عنبیه چشم و یا نظریه انطباق گراف، الگوی دودویی محلی و انواع آن مورد استفاده قرارگرفت که بایستی پس از تشخیص ناحیه چهره، ناحیه مورد نیاز برای پردازش را تشخیص داده و سپس عمل پردازش انجام شود. در مقابل این روش‌ها که مبتنی بر پردازش تک تک پیکسل‌ها بود و نیاز به حجم عظیمی از داده برای عمل بازیابی داشت، روش‌های مبتنی بر شبکه‌های عصبی عمیقی ارائه شد که با استفاده از تعداد محدودی از تصویر عمل آموزش شبکه انجام میشود و کل تصویر را به‌عنوان ورودی در نظر می‌گیرد، همچنین نیازی به برش ناحیه خاصی از تصویر چهره برای پردازش تک تک پیکسل‌های آن نیست و از دقت نسبتا بالایی برخوردار است.

ساختار این مقاله بدین صورت است که دربخش ۲ به بیان مسئله پرداخته و بخش ۳ پیشینه پژوهش را بررسی نموده است، بخش ٤ روش پیشنهادی را بیان نموده و در پایان در بخش پنجم نتیجه کار را مطرح کرده است.

۲- بیان مسئله

امروزه دوربین‌های نظارتی در اکثر اماکن عمومی و خیابان‌ها وجود دارند و در راستای بازیابی تصویر چهره افراد نقش مهمی را ایفا می‌کنند. پیش از این روند یافتن یک فرد در دوربین‌های نظارتی به این صورت بود که اپراتور با برگرداندن فیلم و تماشای آن بتواند شخص گم شده را ردیابی و پیدا کند. با گذشت زمان الگوریتم‌های هوش مصنوعی در حوزه بازیابی تصویر چهره افراد جایگزین شدند. چالش مهم عمل بازیابی تصویر چهره، بازیابی با دقت بالا است.

در این روش تصویر فرد گم شده توسط مرکز اطلاعات بر روی سیستم شناسایی چهره قرار می‌گیرد و با تعیین یک بازه زمانی خاص، سیستم شروع به جستجوی فرد از قاب‌های ویدئو می‌کند. لازم به ذکر است که امروزه با همه‌گیر شدن گوشی‌های تلفن همراه هوشمند، اغلب تصویری از چهره اشخاص و نزدیکان در آن موجود می‌باشد که باعث تسریع و تغییر در شیوه جستجوی افراد گم شده، گردیده است.

در الگوریتم‌هایی مانند الگوی دودویی محلی، بازیابی چهره با دقت قابل قبولی حاصل شد، اما این الگوریتم‌ها برپایه پردازش تک تک پیکسل‌ها از تصویر چهره بودند. با مطرح شدن شبکه‌های عصبی عمیق و انواع روش‌های یادگیری ماشین، انقلابی در این راستا صورت گرفت، همچنین دقت و سرعت بازیابی را تا حد چشم‌گیری بهبود داد. همچنین سخت‌افزارهای جدید موازی‌سازی (پردازنده کارت گرافیک) روند کار را نسبت به قبل تسریع بخشید.

در این مقاله سعی بر ارائه یک راهکار با دقت بازیابی بالا با استفاده از شبکه‌های عصبی عمیق شده است که با افزایش دقت، زمان پردازش نیز کاهش یابد، تا علاوه بر پردازش‌های غیربرخط، توانایی پردازش و بازیابی تصویر برخط را نیز دارا باشد و همان‌طور که در مقدمه اشاره شد با تعداد محدودی تصویر قابل آموزش باشد. همچنین به‌جای پردازش تک تک پیکسل‌ها کل تصویر چهره را به‌عنوان ورودی در نظر می‌گیرد. از موارد قابل استفاده دیگری که می‌توان برای این روش گفت تطابق چهره فرد حاضر با عکس روی شناسنامه در دروازهٔ ورودی فرودگاه‌ها می‌باشد که با یک نمونه عکس روی شناسنامه توانایی احراز هویت شخص را دارد.

1- one-shot-learning

۳- پیشینه پژوهش

داداشی در سال ۱۳۹۰ با استفاده از روش تولید تصاویر مجازی به کمک شبکه‌های عصبی، مسئله بازشناسی چهره با یک تصویر از هر فرد را مورد بررسی قرار داد. بدین صورت که با طراحی ساختارهای مختلفی از شبکه‌های عصبی مصنوعی توانست اطلاعاتی را از تصویر چهره هر فرد و همچنین حالت چهره‌های مجازی ساخته شده از هر فرد را استخراج نماید، به‌طوری که درصد صحت بازیابی در این روش بر روی دادگان آزمون نسبت به مدل مرجع ۱۲/۷۳٪ و نسبت به مدل PCA ۲۶/۳۶٪ دارای بهبود بوده است.

راستگو در سال ۱۳۹۸ با استفاده از تنظیم دقیق مدل از قبل تعمیم داده شده AlexNet و با تبدیل لایه‌های کاملا متصل به لایه‌های هم‌آمیختی[۲] و اعمال فیلترهای مناسب به منظور شناسایی چهره در تصاویر ورودی پرداخت. او با استفاده از برش‌های مختلف عکس ورودی و نیز افزایش تعداد لایه‌های هم‌آمیختی به منظور استخراج ویژگی‌های سطح بالاتر و فیلترهای مناسب در مدل این کار را انجام داد و دقت کار را تا ۹۹/۰۲٪ افزایش داد.

میرزایی در سال ۱۳۹۸ با استفاده از ترکیب انواع مدل‌های چرخشی الگوی دودویی محلی مانند حداقل و حد اکثر پیکسل حاصله از الگوی دودویی محلی با تشخیص لبه تصویر، دقت کار را افزایش داد. همچنین برای مرتفع نمودن مشکل زمان پردازش از پردازشگر گرافیکی مبتنی بر کودا در محیط C++ استفاده نمود. به‌طوری که دقت بازیابی را تا حدود ۹۷/٤٪ افزایش و زمان بازیابی تصویر را تا حدود ۸ ثانیه برای هر چهره رساند.

در سال ۱۳۹۲ قاصری روشی برای بازیابی تصاویر چهره با استفاده از هیستوگرام گرادیان و الگوی دودویی محلی پیشنهاد داد. در این روش ابتدا تصاویر با استفاده از موقعیت مرکز چشم‌ها تنظیم می‌شوند و سپس ناحیهٔ چهره از آن‌ها استخراج می‌گردد. برای استخراج ویژگی در اطراف هر پیکسل سلول‌های کوچکی در نظر گرفته و در هر سلول هیستوگرام گرادیان محاسبه شد و به پیکسل مرکزی سلول اختصاص داده شد. با استفاده از طرح بازخورد ماشین بردار پشتیبان، ویژگی‌های استخراج شده نهایی با هم مقایسه می‌شوند. در نتیجه این کار توانست زمان پردازش را تا ۲۰۰ میلی ثانیه برای هر تصویر برساند.

Koch درسال ۲۰۱۵ توانست با استفاده از ایده یادگیری یک‌باره بر روی شبکه‌های عصبی عمیق سایامیز (در بخش ٤-۱ آمده است) با در دست داشتن تعداد کمی تصویر نمونه بر روی مجموعه داده MNIST که مربوط به نویسه و اعداد است به دقت چشم‌گیری برای بازیابی اعداد دست نوشته برسد.

روش‌هایی که پیش از این مطرح شد، همه از الگوریتم‌هایی استفاده می‌کردند که با پردازش مستقیم تک تک پیکسل‌ها برای استخراج ویژگی از تصویر در ارتباط بوده و لذا از سرعت پایینی برخوردار بودند در مدل پیشنهادی بدون پردازش تک به تک پیکسلی و همچنین با استفاده از تعداد کمی از تصویر جهت آموزش مدل، ویژگی‌های تصویر استخراج شدند که در بخش بعد به آن پرداخته می‌شود.

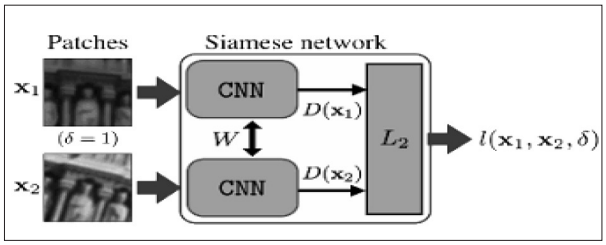
۴- روش پیشنهادی

٤-۱- شبکه عصبی عمیق سایامیز[۳]

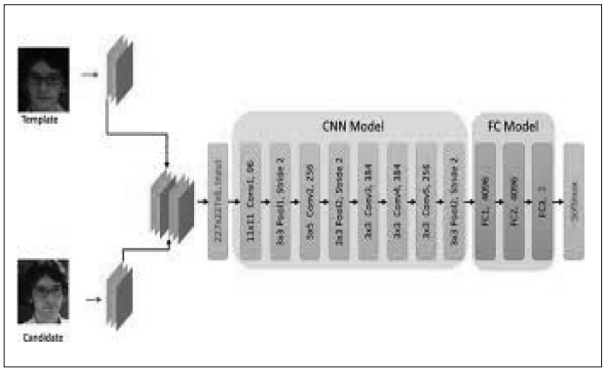
شبکه عصبی عمیق سایامیز شبکه عصبی جفتی نیز نامیده می‌شود که در عملیات مقایسه بین دو بردار ویژگی کاربرد دارد. روش کار این شبکه‌ها به این صورت است که وزن‌های بهینه برای یک بردار مرجع محاسبه شده و به‌عنوان خط مبنا در نظر گرفته می‌شود. سپس برای تک‌تک بردارهای ویژگی وزن‌ها محاسبه و با وزن‌های بردار مرجع مقایسه می‌شوند. این نوع شبکه در پروژه‌های تشخیص هویت از جمله اثرانگشت، تشخیص چهره و تشخیص دست‌خط و امضاء جعلی کاربرد زیادی دارد

2- convolutional

3- siamese



شکل ۱: نمای شبکه سایامیز
www.semanticscholar.org



شکل۲: لایه‌های هم‌آمیختی شبکه سایامیز
www.intechopen.com

(شکل۱). از آنجایی که عمل مقایسه در این شبکه‌ها دوبه‌دو صورت می‌گیرد، لذا از دقت بالائی برخوردار است.

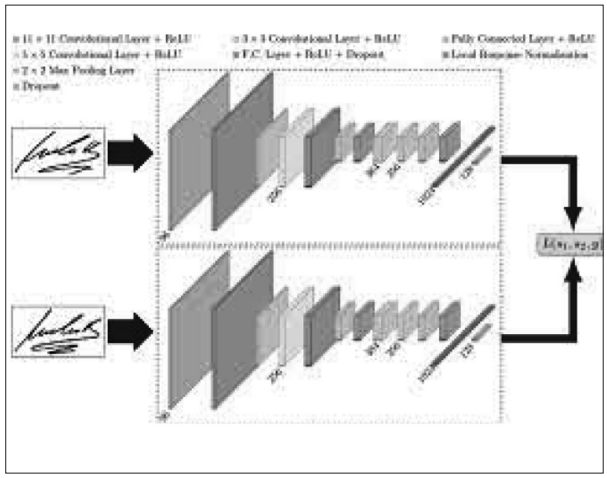
به این شبکه‌ها یادگیری یک‌باره هم گفته می‌شود؛ زیرا با دیدن یک تصویر نمونه توانایی یادگیری را دارند.

٤-۲- مدل VGGFACE2

در این مقاله، شـــبکه بر روی مجموعه داده vggface2 شامل سه میلیون تصویر از ۹۱۳۱ شخص که از هرشخص به‌طور متوسط ۳٦۲ عدد تصویر وجود دارد، آموزش دیده شـــده است. این تصاویر با اســـتفاده جستجوی تصویری گوگل قابل دســـترس هستند که شـــامل محدوده وسیعی از حالت‌های مختلف از جمله روشـــنایی، سن، حالت چهره را دارا می‌باشـــدکه به دو گروه ۸٦۳۱ تایی برای داده‌های آموزش و گروه ۵۰۰ تایی برای داده‌های آزمون تفکیک و تصاویر به اندازه ۲۲٤ × ۲۲٤ از وســـط برش داده شدند. ساختار مدل vggface2 در شکل ٤ آورده شده است.

٤-۳- شرح کار انجام شده

در مـــدل vggface2 کـــه در اینجا مورد اســـتفاده قرار گرفته، تصویری به‌عنوان تصویر مرجع به ورودی شبکه



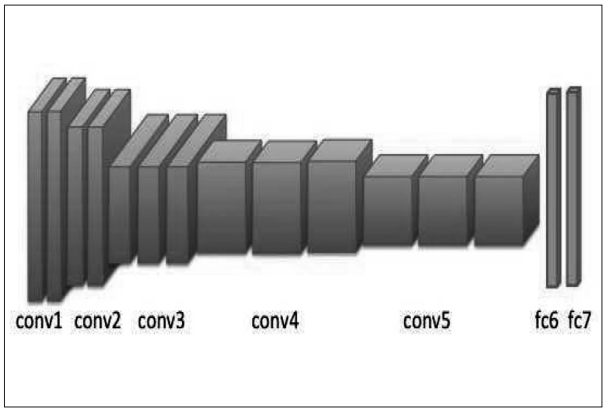
شکل۳: نمونه‌ای از تایید امضا با استفاده از شبکه سایامیز
www.semanticscholar.org

داده می‌شود و پارامترهای آن تصویر به‌عنوان خط مبنا در نظر گرفته می‌شوند. سپس برای کلیه تصاویر موجود این عمل انجام شـــده و با هم مقایسه می‌شوند. از آن جائی که در این مدل کلیه تصاویر دوبه‌دو با هم مقایسه می‌گردند، از دقت حائز اهمیتی برخوردار است. اگر عدد حاصل شده از فاصله تشـــابه‌یابی صفر بود، به معنای تشـــابه وگرنه به‌معنای عدم تشابه است.

نکته حائز اهمیت افزایش دقت و زمان پردازش اســـت. برای مرتفع نمودن این مشـــکل، ورودی‌ها و مدل شبکه به حافظه پردازنده گرافیکی، جهت پردازش موازی داده شـــد که زمان پردازش را به‌طور قابل توجهی کاهش داد. از این رو این روش برای پردازش‌های برخط کاربرد زیادی دارد.

در ایـــن مقاله، آزمون بـــر روی دو نوع عکس و ویدئو انجام گرفت، بدین صورت کـــه ناحیه تصویر چهره را بر روی عکس و یـــا قاب ویدئو تشـــخیص و به‌عنوان بردار ویژگی مرجع در نظر گرفته شد و عمل مقایسه بردار مرجع با تک‌تک بردارهای ویژگی تصاویر موجود در پوشه انجام شد، همچنین عمل تشابه‌یابی بین تک‌تک تصاویر صورت گرفت.

نکته‌ای که در این پژوهش برای پردازش ویدئوی زنده از اهمیت خاصی برخوردار اســـت، تنظیم نمودن زمان پر شـــدن میانگیر و یا همان پوشـــه‌ای که قرار است قاب‌های ویدئوی زنده در آن قرار گیرد، می‌باشـــد، به‌طوری‌که نرخ

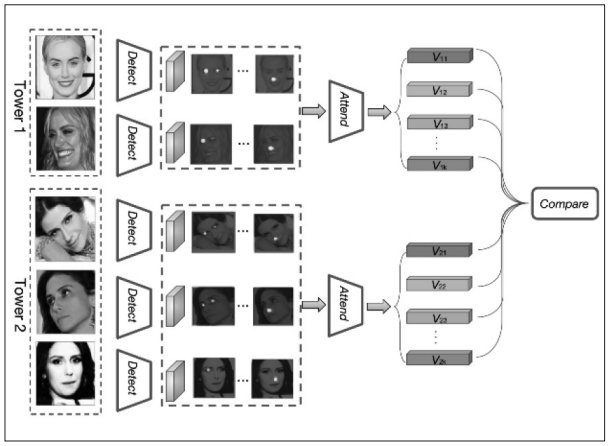


شکل۴: معماری شبکه VGGFACE2
www.researchgate.net

خواندن قاب از میانگیر و یا همان پوشـــه بایستی با اندکی تاخیر باشـــد تا سرعت خواندن قاب از سرعت ذخیره قاب بالاتر نرود، در صورت وقوع این مشکل سیستم با کمبود قاب و تداخل مواجه خواهد شـــد، بدین صورت با انتخاب ضریب خواندن قاب ٤٨ قاب در هر دفعه این مشکل مرتفع گردیـــد. (بدین صورت کـــه در دفعه اول قاب شـــماره ٤٨ خوانده و سپس قاب شماره ٩٦ و الی آخر) به‌طوری‌که هم پردازش با سرعت قابل قبولی به‌صورت برخط انجام شد و هم با کمبود قاب مواجه نشد. همچنین زمان پردازش هر چهره در حدود ۳ ثانیه طول کشید.

۵- نتیجه‌گیری و کارهای آتی

نتایجـــی که بر طبق تحقیقات این پژوهش حاصل شـــد در جـــدول ۱ برای عکس، و قاب‌هـــای ویدئو به‌طور مجزا بیان شـــده اســـت. لذا از کارهایی که در آینده می‌توان در راســـتای این پژوهش انجام داد، عمل قطعه‌بندی تصویر با استفاده از یادگیری عمیق، برای یافتن ناحیه چهره بر روی تصویر است که در ابتدا باید تصویر و ماسک را به شبکه آموزش داد تا شـــبکه خود قادر به یافتن تصویر چهره از عکس و یا قاب ویدئو باشد تا در مقایسه با روش‌های قبلی همچون Haarclassification که در کتابخانه opencv قرار داد مجبـــور به پیمایش تک تک پیکســـل‌های تصویر برای یافتن ناحیه نباشد.



شکل۵: نمای کلی شبکه مقایسه‌گر
www.robots.ox.ac.uk

جدول ۱: نتیجه درصد دقت بازیابی برای ویدئو و عکس

نوع تصویر	تعداد کل چهره	TN	FP	FN	TP	دقت به درصد
تصویر	۱۰۰	۲۱	۲	۰	۷۸	۹۹
قاب ویدئو	۵۹	۰	۰	۲	۵۷	۹۶/۶

منابع

۱- م. قاصـــری و ح. ابراهیم‌نـــژاد، «بازیابی تصاویر چهره با اســـتفاده از ترکیب هیســـتوگرام گرادیان و الگوی باینری محلی»، ماشین بینایی و پردازش تصویر، جلد ۱، شماره ۱، pp. ۵۸-۶۸، ۱۳۹۲.

۲- ن. داداشـــی , ف. عبدالعلی و س. ع. ســـیدصالحی، «بهبود بازشناسی چهره با یک تصویر از هر فرد به روش تولید تصاویر مجازی توسط شبکه‌های عصبی»، پردازش علائم و داده‌ها، جلد ۱، شماره ۱۵، pp. ۳۳-۴۴، ۱۳۹۰.

۳- ر. راســـتگو و ک. کیانی، «شناسایی چهره با استفاده از تنظیم دقیق شبکه‌های کانولوشنی عمیق و رویکرد یادگیری انتقالی»، مجله مدل‌سازی در مهندسی، جلد ۵۸، pp. ۱۰۳-۱۱۱، ۱۳۹۸.

۴- ک. میرزایی و ک. طباطبائی اردکانی، «شناســـایی چهره با اســـتفاده از الگوی دودویی محلی ترکیبی برپایه پردازنده گرافیکی جهت تسریع امر شناسایی افراد در پایگاه‌های نظامـــی»، فصلنامه علمی ترویجی علوم و فنـــاوری دریا، جلد ۹۱، شماره ۲۳، pp. ۱۷-۲۳، ۱۳۹۸.

5. R. Shiv, K. Satish and K. Rajat, “Local SVD based NIR face retrieval,” journal of visual communication and image representation, pp. 141-152, 2017.

6. C. Bor-Chun, C. Yan-Ying and K. Yin-HSI, “Scalable Face Image Retrieval using Attribute-Enhanced Sparse Codewords,” IEEE, pp. 1163-1173, 2013.

7. G. Koch, R. Zemel and R. Salakhutdinov, “Siamese Neural Networks for One-shot Image Recognition,” in International Conference on Machine Learning, Lille,France, 2015.