تشخیص جنسیت با توجه به تصویر چهره با استفاده از شبکه FaceNet

خدیجه آقاجانی

استادیار، دانشکده فنی مهندسی دانشگاه مازندران، بابلسر

چکیده

سیستم هایی که به طور خودکار با استفاده از تصاویر چهره جنسیت را مشخص میکنند، به سرعت در حال رشد هستند. از کاربردهای تشخیص جنسیت از چهره میتوان به نظارت تصویری، کنترل دسترسی و تعامل انسان و رایانه اشاره کرد. در سال‌های اخیر، تلاش‌های متعددی برای حل این مساله صورت گرفته است. در این مقاله در ابتدا با استفاده از روش ویلاجونز چهره فرد از تصویر استخراج میشود. سپس از یک شبکه از پیش آموزش داده شده تحت عنوان FaceNet برای استخراج ویژگی از چهره استفاده میشود. در ادامه با استفاده از یک دسته بند از ویژگی استخراج شده برای تشخیص جنسیت فرد استفاده شده است. در این مقاله از دو پایگاه داده UTKFaceو GENDER-FERET جهت ارزیابی مدل پیشنهادی استفاده شده است. مشاهده شده استفاده بردار های استخراج شده و دسته بند SVM در پایگاه های داده UTKFace,GENDER-FERET به ترتیب با دقت ۹۷/۶۴٪ و ۹۸/۳٪ قادر به تشخیص درست جنسیت میباشد.

واژه های کلیدی

تشخیص جنسیت از تصویر، انتقال یادگیری، ماشین بردار پشتیبان.

مقدمه

تشخیص خودکار جنسیت با توجه به تصویر چهره یک کار اساسی در بینایی کامپیوتر است که اخیراً توجه زیادی را به خود جلب کرد و نقش اساسی در طیف گسترده ای از برنامه های کاربردی در دنیای واقعی مانند علم پز شکی قانونی، نظارت بصری، سیستم های تعامل انسان و کامپیوتر و غیره را ایفا میکند[۱]. در حال حاضر تشخیص جنسیت بر اساس چهره، به دلیل تغییرات مختلف در زوایای بصری، حالات چهره، ژست، سن و پس زمینه هنوز یک کار دشواری است.

بسته به نوع ویژگی‌های مورد استفاده در الگوریتم های تشخیص جنسیت، میتوان رویکرد های مختلف برای حل این مساله را به دو دسته تقسیم نمود. رویکرد اول شامل روش های مبتنی بر استخراج ویژگی‌های دست‌ساز است. در این رویکرد نیاز به دانش خاص در زمینه انتخاب ویژگی مناسب از قبیل شدت، بافت، شکل و هندسه است. دسته دوم مبتنی بر استخراج ویژگی های قابل آموزش است، به این صورت که ویژگی های متمایزکننده جنسیت را می توان به طور خودکار از داده های آموز شی آموخت. مزیت استفاده از روش دوم این است که به کمک آنها میتوان جنبه‌هایی از چهره را به تصویر کشید که انسان نمی‌تواند متوجه آن‌ها شود. علاوه بر این، روش استخراج چنین ویژگی‌هایی به دانش خاصی متکی نیست.

رویکردهای متعلق به دسته اول از انواع مختلفی از ویژگی ها بر اساس اطلاعات رنگ ، بافت و شکل استفاده می کنند. تقریباً همه آنها معماری مشابهی دارند که شامل سه مرحله است: ۱) تشخیص و برش چهره با استفاده از الگوریتم شناخته شده ویلا جونز [۲] ، ۲) پیش پردازش تصویر، به منظور عادی سازی صورت از نظر ابعاد، ژست و روشنایی،۳)استخراج ویژگی های مورد استفاده برای تشخیص جنسیت. در مقاله [۳] ، از ویژگی HOG بعنوان بردار ویژگی و از دسته بند SVM جهت تشخیص استفاده شده است. آقای فکری در [۴] از ویژگی بافت مبتنی بر ویژگی های باینری محلی بهبود یافته (ILBP) برای استخراج ویژگی و از Kullback-Leiblerبعناون دسته بند استفاده نموده است. در مقاله [۵]، پس از استخراج و ترکیب ویژگی های باینری محلی (LBP) و توصیفگر های مرحله محلی (LPQ) ، از ماشین بردار پشتیبان (SVM) به عنوان طبقه‌بندی کننده برای انجام طبقه‌بندی جنسیتی استفاده می‌شود.در [۶]، پس از استخراج ویژگی وفقی از SVM برای دسته بندی استفاده شده است. خانم شبیرینژاد وهمکاران در [۷] ، از یک نمایش ویژگی جدید، یعنی الگوهای مشتق درهم (IDP) که یک تکنیک مبتنی بر مشتق برای استخراج ویژگی‌های چهره است جهت دسته بندی جنسیت ها استفاده کرده اند. لیان و لو [۸] هیستوگرام های الگوی باینری محلی (LBP) را از نواحی مختلف صورت، در یک بردار مشخصه استخراج و به هم الحاق می کنند و یک طبقه بندی SVM را برای تشخیص جنسیت آموزش دادند. منطق این رویکرد این است که یک توصیفگر بافت می‌تواند تفاوت‌های بین پوست صاف‌تر یک زن و پوست زبرتر یک مرد را نشان دهد. در مقاله [۹] از توصیفگرهای SURF استخراج شده از ۵۱ علامت برجسته صورت مربوط به چشم ها، بینی و دهان به عنوان ویژگی های وابسته به دامنه و فیلترهای COSFIRE به عنوان ویژگی های قابل آموزش استفاده است. در ادامه از دسته بند SVM جهت تشخیص کلاس استفاده شده است.

استفاده از یادگیری عمیق در این حوزه نیز در سالهای اخیر بسیار مورد توجه قرار گرفته شده است. در مقالات متعددی از شبکه های مبتنی بر CNN جهت تشخیص جنسیت استفاده شده است[۱۰-

۱۶]. در این مقالات از شبکه هایی متشکل از لایه های کانولوشن و پولینگ یک چارچوب end-to-end ای جهت تشخیص جنسیت ارایه شده است. همینطور استفاده از شبکه های از پیش آموزش داده شده نیز جهت استخراج ویژگی مورد توجه زیادی قرار گرفته شده است. در مقاله [۱۷] از شبکه VGGFace برای استخراج ویژگی استفاده شده است.

در این مقاله از یک شبکه از پیش آموزش داده شده تحت عنوان FaceNet [۱۸] برای استخراج ویژگی چهره استفاده شده است. خروجی این شبکه یک بردار ۱۲۸ تایی به ازا هر تصویر است . در مرحله بعدی کافیست بردار ایجاد شده را به یک دسته بند اعمال نمود . در این پژوهش کارایی KNN و SVM بعنوان دسته بند مقایسه شده ا ست. در این مقاله از دو پایگاه داده ت صاویر چهره جهت ارزیابی ا ستفاده شده ا ست. دقت به د ست آمده در ت شخیص، ن شان دهنده دقت بالای روش پی شنهادی در ت شخیص جن سیت ا ست. ادامه مقاله دارای بخش های زیر ا ست. در بخش دوم روش پی شنهادی ارایه شده است. بخش سوم و چهارم به مشاهدات و نتیجه گیری اختصاص دارد.

۲.روش پیشنهادی

در این بخش به ارایه روش پیشنهادی جهت تشخیص جنسیت افراد از روی تصویر میپردازیم . به دلیل استفاده از روش FaceNet جهت استخراج ویژگی در ادامه ابتدا با این الگوریتم آشنا خواهیم شد.

الگوریتم FaceNet:

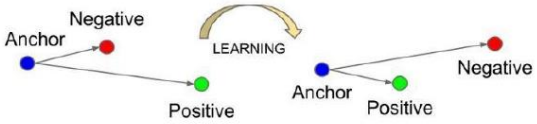
الگوریتم FaceNet یک الگوریتم تشخیص چهره است که هدف از ارایه آن استخراج بردار ویژگی است که بتوان به کمک آن عملیات تشخیص را با دقت بالایی انجام داد. به کمک این الگوریتم هر چهره به یک فضای اقلید سی ۱۲۸ بعدی تحت عنوان embedding نگا شت میشود. بعد از استخراج این بردار، عملیاتی نظیر تشخیص چهره، صحت سنجی و دسته بندی قابل انجام است. در این روش تابع هزینه تابعی تحت عنوان تابع هزینه سه تایی است. در این تابع داده ها به مجموعه ای از تصاویر سه تایی تبدیل میشود. در هر سه گانه یک تصویر تحت عنوان anchor از یک کلاس، تصویری تحت عنوان تصویر مثبت از همان کلاس و تصویر منفی از یک کلاس دیگر وجود دارد. هدف تابع هزینه در اینجا این ا ست که اختصاص بردار ویژگی به تصاویر به گونه ای انجام شود که فاصله بردارهای ویژگی استخراج شده از ت صاویر متعلق به یک کلاس، کمینه شود در حالی که فا صله بردار های ویژگی ت صاویر متعلق به کلاس های متفاوت بی شینه شود. بطور خلاصه تابع هزینه سه گانه را میتوان به صورت زیر نوشت :

$$triplet = \sum_{i=1}^{N}[\left\|f(x_i^a) - f(x_i^p)\right\|_2^2 - \left\|f(x_i^a) - f(x_i^n)\right\|_2^2 + \alpha] \quad (۱)$$

در این رابطه x_i^a ، x_i^p و x_i^p به ترتیب ت صاویر anchor، مثبت و منفی است. $f(x_i)$ نشان دهنده embedding استخراج شده برای ت صویر x_i ا ست. α نیز مقدار حا شیه را ن شان میدهد به این معنی که اگر برای مثال مقدار آن برابر ۰/۵ باشد به دنبال این هستیم که اختلاف فاصله بین anchor و ت صویر مثبت و فاصله بین anchor و ت صویر منفی حداقل نیم با شد. در شکل (۱) نمونه ای از یک سه گانه نمایش داده شده ا ست. در هر سه گانه برای انتخاب ت صویر مثبت از تصاویر با فاصله بالا در کلاسی به anchor به آن متعلق ا ست و برای ت صاویر منفی از ت صاویر خارج از کلاس با فا صله کم ا ستفاده می شود. در شکل (۲) نیز نحوه عملکرد تابع هزینه سه تایی نمایش داده شده است .

ج ب الف

شکل ۱: نمونه ای از یک سه گانه در الگوریتم FaceNet . الف) تصویر anchor. ب) تصویر منفی و ج) تصویر مثبت



شکل ۲: نمایشی از عملکرد تابع هدف سه کانه.همانطور که در این شکل مشاهده میشود هدف از استفاده از این تابع این است که بردار ویژگی دو تصویر متعلق به یک کلاس به هم نزدیک و بردار ویژگی تصاویر متعلق به کلاس های مختلف از هم دور تر شود.

الگوریتم پیشنهادی :

دیاگرام الگوریتم پیشنهادی جهت تشخیص جنسیت در شکل (۳) نمایش داده شده ا ست. در روش پی شنهادی ابتدا در هر ت صویر چهره اشخاص از داخل تصویر استخراج میشود. برای استفاده ازمدل از پیش آموزش دیده شده FaceNet نیاز ا ست که ابعاد ت صویر ورودی برای ا ستفاده از این مدل ۱۶۰*۱۶۰ با شد. بنابراین هر چهره به یک تصویر ۱۶۰*۱۶۰ تغییر سایز داده میشود و قبل از ارسال آن به شبکه FaceNet تصویر نرمالیزه میشود. بردار ویژگی به ازا هر چهره با گذشتن از شبکه مربوط به FaceNet استخراج میشود.

شکل ۳: نحوه آماده سازی داده جهت ورودی به شبکه در روش پیشنهادی . الف) نمونه تصویر ورودی. ب) استخراج چهره. ج) نرمالیزه کردن تصویر. د) تغییر سایز تصویر به اندازه مطلوب شبکه.FaceNet

گام بعدی اســـتفاده از یک مدل یادگیری ماشـــین برای آموزش دسته بند جهت تشخیص مرد یا زن بودن است. در این پژوهش از دو دسته بند SVM و K-NN استفاده شده است.

دسته بندی به روش K-NN

دســته بندی به روش K-NNیک رویکرد بانظارتی اســت که در آن داده ها دارای برچســب هســتند. داده ها (ورودی های مدل برای مثال بردار ویژگی منتسب به تصاویر ورودی) در یک فضای چند بعدی در نظر گرفته میشود. هر بعد این بردار ویژگی نشان دهنده اطلاعات خاصی از داده ورودی است. در اینجا با توجه به برچســب های زن ومرد فضــای مورد نظر به دو بخش تقســیم میشود. در هنگام تست با ورود داده جدید، برچسب آن بر اساس نتیجه رای گیری بین برچسب های k همسایه آن داده در فضای در نظر گرفته شده، تعیین میشود.برای به دست آوردن همسایه های هر داده از فاصــله اقلیدســی بین بردار ویژگی داده مذکور و سایر داده ها استفاده میشود.

دسته بندی به روش SVM :

ماشین بردار پشتیبان یکی از روش های ساده ای است که در دسته بندی و یا رگرسیون قابل استفاده است. در این روش با استفاده از یک ابرصفحه با رابطه زیر مرزی بین بردارهای منتسب به داده ها ترسیم میشود.

$$y = b + W^T X \qquad (۲)$$

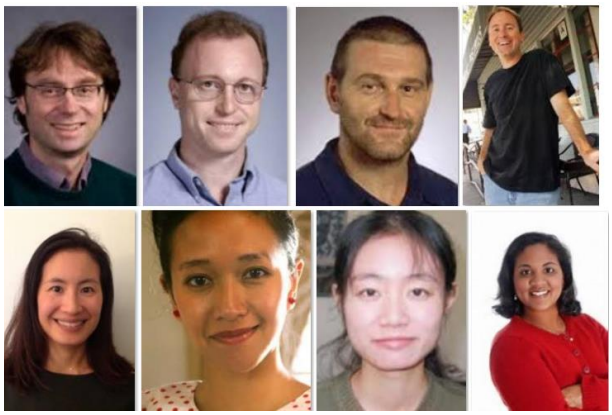
در این رابطه $W = [w_0, w_1, ..., w_{127}]$ و b پارامتر های مربوط به ابرصــفحه و $X = [x_0, x_1, ..., x_{127}]$ بردار ویژگی منتســب به هر تصویر است. در فرایند آموزش، پارامترهای مربوط به ابر صفحه در جهت تمایز بهتر بین داده ها بر اســاس برچســب هایشــان تخمین زده میشود. برای به دست آوردن مرزهای دقیق تر بین داده ها گاها از SVM دارای کرنل اســتفاده میشــود که نتیجه اســتفاده از آن ایجاد مرز غیر خطی بین داده ها میگردد. به این منظور به نظر میرســد که در ابتدا داده ها به یک فضــای با ابعاد بالاتر نگاشت میشود و در فضای جدید به کمک یک ابر صفحه دســته بندی انجام میشــود. کرنل در اینجا یک معیار شــباهت بین نقاط داده در فضای اولیه است. از توابع کرنل رایج میتوان به RBF اشاره نمود. در این تابع، شباهت بین دونقطه در فضای تبدیل تابع نمایی میرایی از فاصله بین بردار ویژگی ها ست. تابع RBF به ازا دو بردار ویژگی x,x' به صورت رابطه زیرست.

$$K(x, x') = \exp(-\gamma \|x - x'\|) \qquad (۳)$$

۳.پیاده سازی و مشاهدات :

جهت ارزیابی مدل پیشــنهادی از دو پایگاه داده GENDER-FERET و پایگاه UTKFace اســتفاده شــده اســت. پایگاه داده GENDER-FERET دارای تعداد بالانسی از برچسب های خانم و آقاست. این پایگاه شامل تصاویری از روبرو، دارای احساسات مختلف ، شــدت روشــنایی های متفاوت و پس زمینه های متنوع اســت. در این پایگاه به ازا هر فرد یک تصــویر وجود دارد لذا امکان اینکه از یک فرد هم در تست و هم در آموزش تصویری موجود باشد امکان پذیر نیست. چند نمونه از تصاویر این پایگاه داده در شکل (۴) نمایش داده شده است.

شکل ۴: نمونه های تصادفی از تصاویر پایگاه GENDER-FERET.



شکل ۵: نمونه های تصادفی از تصاویر پایگاه UTKFace.

پایگاه داده دوم پایگاه داده UTKface اســـت . این پایگاه داده دارای تصاویری از رنج سنی گسترده (از ۰ سال تا ۱۱۶ سال) ا ست. در تصاویر تغییرات شدیدی در پس زمینه، شدت روشنایی، رزلوشن و انسداد مشاهده میشود. چند نمونه ار تصاویر متعلق به این پایگاه داده در شـکل (۵) نمایش داده شـده اسـت. در اینجا ،برای اسـتفاده از این پایگاه داده جهت تشخیص جنسیت از تصاویر متعلق به رنج سنی ۲۰ تا ۶۰ سال استفاده شده است.

همانطور که در بخش قبل اشـاره شـد در این پژوهش از الگوریتم FaceNet جهت ا ستخراج ویژگی ا ستفاده شده ا ست. همین طور از الگوریتم های K-NN و SVM جهت دسـته بندی اسـتفاده شـده است. جهت ارزیابی روش پیشنهادی از معیار دقت استفاده شده است. در اینجا در هر آزمایش ۱۵ در صد داده ها را به مجموعه تست و ۸۵٪ داده ها را به مجموعه آموزش اخت صاص داده ایم . با توجه به ت صادفی بودن انتخاب مجموعه تسـت و آموزش در اینجا برای گزارش مقدار دقت ، 5 بار آزمایش انجام و نهایتا میانگین و انحراف از معیار مقدار accuracy محا سبه شده ا ست. در جدول (۱) مقدار دقت به ازا هر مجموعه داده گزارش داده شـده اسـت. همینطور در جداول (۲و۳) ماتریس های در هم ریختگی به ازا هر مجموعه داده نمایش داده شده اسـت . لازم به ذکرسـت که اعداد داخل ماتریس ها در این دو جدول نیز میانگین اعداد به دست آمده در ۵ تکرار است.

جدول ۱:دقت روش پیشنهادی در داده های استفاده شده

	faceNet+K-NN	faceNet+SVM
Gender-Feret	٪۹۶/۸۲	٪۹۸/۳
UTKface	٪۹۷/۴۱	٪۹۷/۶۴

جدول ۲ : ماتریس های در هم ریختگی در پایگاه داده UTK

		پیش بینی	دسته بند SVM
		مرد	زن
واقعی	مرد	٪۹۷/۹۸	٪۲/۰۲
	زن	٪۲/۸	٪۹۷/۲

		پیش بینی	دسته بند Knn
		مرد	زن
واقعی	مرد	٪۹۷/۲۵	٪۲/۷۵
	زن	٪۲/۳۹	٪۹۷/۶۱

جدول ۲ : ماتریس های در هم ریختگی در پایگاه داده Gender-FERET

		پیش بینی	دسته بند SVM
		مرد	زن
واقعی	مرد	98/73%	1/27%
	زن	2/12%	97/88%

		پیش بینی	دسته بند Knn
		مرد	زن
واقعی	مرد	95/76%	4/24%
	زن	2/12%	97/88%

۴.نتیجه گیری

در این مقاله روشـی جهت تشـخیص جنسـیت از روی چهره با ا ستفاده از شبکه FaceNet ارایه شده ا ست. در روش ارایه شده در ابتدا ، چهره فرد با کمک الگوریتم ویلاجونز اسـتخراج میشـود. سـپس پس از نرمالیزه کردن و تغییر سایز به سایز مطلوب شبکه FaceNet بردار ویژگی از آن ا ستخراج می شود. با ورود بردار های به د ست آمده به دسـته بند های KNN و SVM عملیات تشـخیص در انتها انجام شده است. مشاهده شده استفاده بردار های استخراج شده و دسته بند SVM در پایگاه های داده GENDER-FERET, UTK به ترتیب با دقت ٪۹۷/۶۴ و ٪۹۸/۳ قادر به تشخیص درست جنسیت میباشد.

منابع :

[1] Khan, K., Attique, M., Syed, I. and Gul, A., 2019. Automatic gender classification through face segmentation. Symmetry, 11(6), p.770.

[2] P. Viola, P. and Jones, M.J., 2004. Robust real-time face detection. International journal of computer vision, 57(2), pp.137-154.

[3] Carcagnì, P., Cazzato, D., Coco, M.D., Leo, M., Pioggia, G. and Distante, C., 2014, October. Real-time gender based behavior system for human-robot interaction. In International Conference on Social Robotics (pp. 74-83). Springer, Cham.

[4] Fekri-Ershad, S., 2019. Gender classification in human face images for smart phone applications based on local texture information and evaluated Kullback-Leibler divergence. Traitement du Signal, 36(6), pp.507-514.

[5] Zhang, C., Ding, H., Shang, Y., Shao, Z. and Fu, X., 2018. Gender classification based on multiscale facial fusion feature. Mathematical Problems in Engineering, 2018.

[6] Shmaglit, L. and Khryashchev, V., 2013. Gender classification of human face images based on adaptive features and support vector machines. Optical Memory and Neural Networks, 22(4), pp.228-235.

[7] Shobeirinejad, A. and Gao, Y., 2010, August. Gender classification using interlaced derivative patterns. In 2010 20th International Conference on Pattern Recognition (pp. 1509-1512). IEEE.

[8] Lian, H.C. and Lu, B.L., 2006, May. Multi-view gender classification using local binary patterns and support vector machines. In International Symposium on Neural Networks (pp. 202-209). Springer, Berlin, Heidelberg.

[9] Azzopardi, G., Greco, A., Saggese, A. and Vento, M., 2018. Fusion of domain-specific and trainable features for gender recognition from face images. IEEE access, 6, pp.24171-24183.

[10] Levi, G. and Hassner, T., 2015. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-42).

[11] Jia, S. and Cristianini, N., 2015. Learning to classify gender from four million images. Pattern recognition letters, 58, pp.35-41.

[12] van de Wolfshaar, J., Karaaba, M.F. and Wiering, M.A., 2015, December. Deep convolutional neural networks and support vector machines for gender recognition. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 188-195). IEEE.

[13] Ranjan, R., Patel, V.M. and Chellappa, R., 2017. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE transactions on pattern analysis and machine intelligence, 41(1), pp.121-135.

[14] Levi, G. and Hassner, T., 2015. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-42).

[15] Srinivas, N., Atwal, H., Rose, D.C., Mahalingam, G., Ricanek, K. and Bolme, D.S., 2017, May. Age, gender, and fine-grained ethnicity prediction using convolutional neural networks for the East Asian face dataset. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 953-960). IEEE.

[16] Kumar, S., Singh, S. and Kumar, J., 2019, January. Gender classification using machine learning with multi-feature method. In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0648-0653). IEEE.

[17] Smith, P. and Chen, C., 2018, December. Transfer learning with deep CNNs for gender recognition and age estimation. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 2564-2571). IEEE.

[18] Schroff, F., Kalenichenko, D. and Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).