مجله پژوهش های معاصر در علوم و تحقیقات
سال پنجم، شماره ۴۸، تیر ۱۴۰۲
ISSN: ۲۶۷۶-۵۷۶٤
www.jocrisar.ir

مروری بر شبکه های عصبی شناسایی چهره در تصاویر

آناهیتا دهشید [۱] ، فرساد زمانی [۲]

[۱] دانشجوی دکترا، گروه کامپیوتر، واحد اصفهان(خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران (نویسنده مسئول)
[۲] استادیار، گروه کامپیوتر، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران

چکیده

امروزه تصاویر متنوعی در دنیا وجود دارد که شامل اشیا و چهره های مختلفی می باشد. از این تصاویر در زمینه های مختلف مانند شناسایی چهره جهت احراز هویت فرد یا برای شناسایی افرادی که ماسک نزده اند، استفاده می شود. شبکه های عصبی گوناگونی برای این کار به وجود آمده است. در این مقاله یک بررسی سیستماتیک در مورد برخی از این شبکه ها جهت شناسایی چهره در تصاویر و همچنین نحوه ساختار و معماری، مزایا و معایب هرکدام صورت گرفته و از جهت پارامترهای مختلفی چون سرعت، دقت ، زمان اجرا و... مورد بررسی قرار گرفته اند.

واژه‌های کلیدی: شناسایی اشیا، شناسایی چهره، روش یادگیری عمیق، شبکه عصبی

۱. مقدمه

شناسایی چهره از موضوعات مهم در زمینه بینایی کامپیوتر بوده است. امروزه شرکت های مختلف از شناسایی چهره در جهت اطمینان از ورود و خروج کارمندان خود استفاده می کنند (Zhao et al., 2003)، تا از روی تصویر بتوان به تجزیه و تحلیل رفتار چهره پرداخت (Pantic & Rothkrantz, 2000). از این رو برای این هدف، از روش های مختلفی استفاده شده است. در سال های اخیر، روش های یادگیری ماشین و یادگیری عمیق در زمینه های مختلف از جمله شناسایی اشیا (Wu et al., 2020)، شناسایی عابران پیاده (Brunetti et al., 2018)، شناسایی خودرو (Ammour et al., ۲۰۱۷) و غیره استفاده شده است. از این روش ها جهت شناسایی چهره نیز استفاده شده است. مقالات مروری نیز در جهت استفاده از این روش ها برای شناسایی چهره استفاده کرده اند (Zafeiriou et al., 2015).

در این مقاله به بررسی روش های یادگیری عمیق مانند روش های Mask RCNN، روش های مختلف Yolo و دیگر روش ها پرداخته و آن ها را با یکدیگر مقایسه کرده ایم.

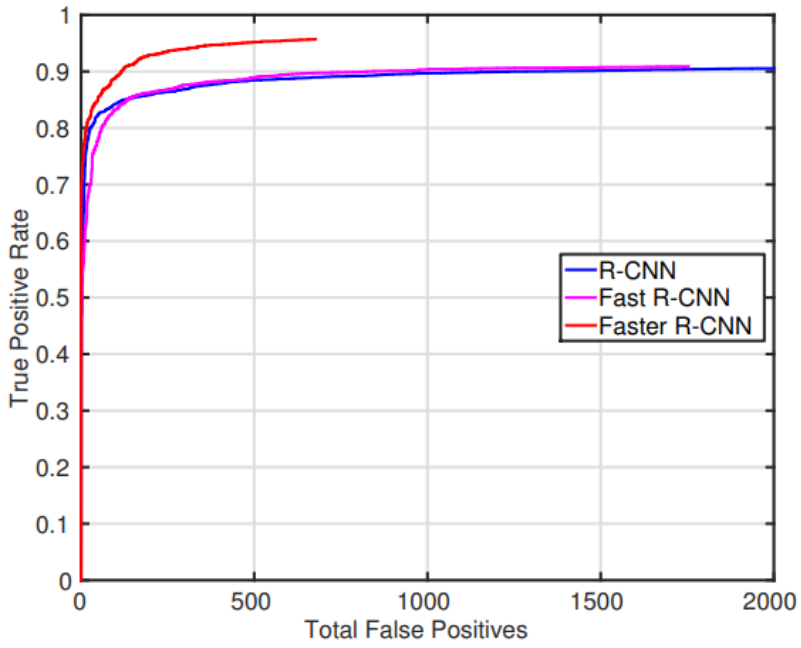
۲. روش تحقیق

در این قسمت به بررسی خلاصه مقالاتی در زمینه شناسایی تصویر[۱]پرداخته و از روش های مختلفی که جهت شناسایی تصویر استفاده کرده اند، می پردازیم.

جیانگ و همکاران[۲] در سال ۲۰۱۷ جهت کاهش بار محاسباتی برای شناسایی تصویر از روش Faster RCNN استفاده کرده‌اند. در این روش از دو ماژول RPN و CNN جهت تولید شیء، از SGD برای هر دو شاخه طبقه بندی و رگرسیون و همچنین از مجموعه داده چهره WIDER که دارای ۱۲۸۸۰ تصویر و ۱۵۹۴۲۴ چهره می باشد، استفاده شده است. RPN و Fast R-CNNبصورت سرتا سر آموزش داده می شوند به گونه ای که ابتدا Fast RCNN و سپس RPN استفاده می شود و ورودی Fast R-CNN در واقع به خروجی RPN وابسته است. آن ها از آموزش تشخیص چهره براساس دو مدل از پیش آموزش دیده VGG و ImageNet بهره برده و برای RPN نیز فقط از ۳۰۰ پیشنهاد چهره برتر برای متعادل کردن کارایی و دقت استفاده شده است. نتایج این مقاله نشان دهنده در دیتاست WIDER حاشیه نویسی صورت به طور محکم در اطراف ناحیه صورت مشخص شده اند در حالی که حاشیه نویسی در IJB-A شامل مناطق بزرگتر (مانند مو) می شود، به دلیل اشتراک لایه های کانولوشن بین RPN و آشکارسازFast R-CNN ، زمان فرآیند کل سیستم کمتر است، RPNبه سایر روش‌های پیشنهاد شی، مانندEdgeBox ، متکی نیست و همچنین نتایج خود را با روش های ، Fast RCNN و Faster RCNN در دیتاست FDDB مقایسه کرده و نتایج را در شکل ۱ به صورت نمودار نشان داده اند. در شکل ۲ نیز نتایج تشخیص نمونه در مجموعه داده IJB-A ، که در آن جعبه‌های مرزی سبز حاشیه نویسی ground-truth و جعبه‌های مرز قرمز نتایج Faster R-CNN هستند را نشان داده اند (Jiang & Learned-Miller, 2017).

[۱] Face Detection

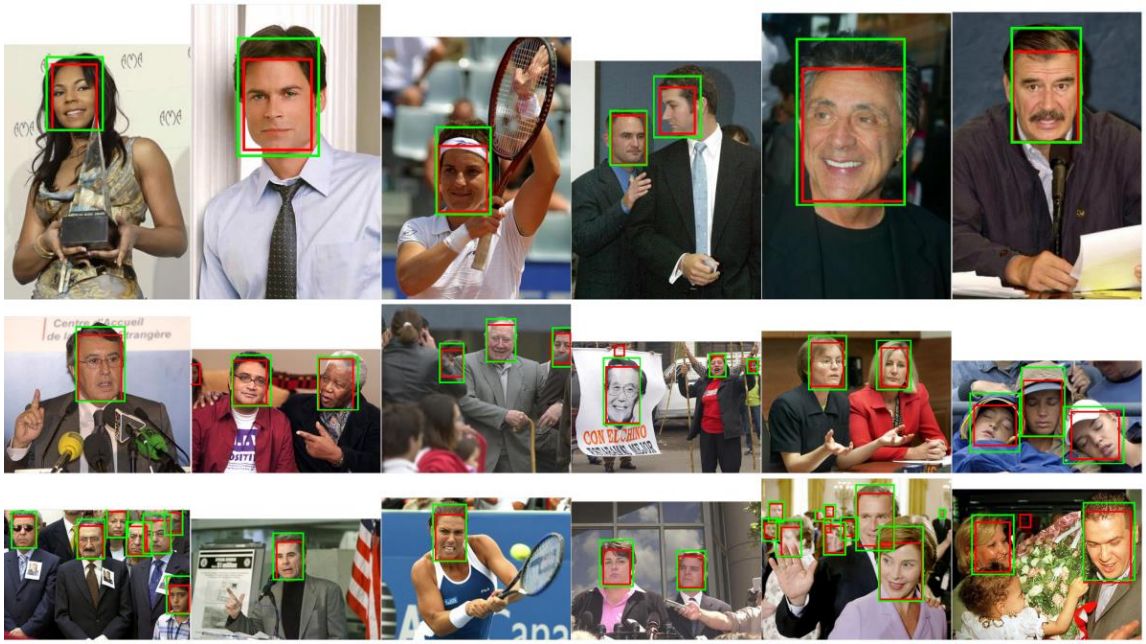
[۲] Jiang et al.



شکل ۱- مقایسه روش مقاله با روش های دیگر (Jiang & Learned-Miller, 2017)

شکل ۲- نتایج مقاله با ground truth در مجموعه داده IJB-A (Jiang & Learned-Miller, 2017)

همانطور که در شکل ۲ نشان داده شده است، روش Faster RCNN توانسته نتایج بهتری را در مجموعه داده IJB-A نشان دهد. شکل ۳ نیز نتایج تشخیص نمونه در مجموعه داده FDDB، که در آن جعبه‌های مرزی سبز حاشیه نویسی ground-truth و جعبه‌های مرز قرمز نتایج Faster R-CNN هستند را نشان می دهد.



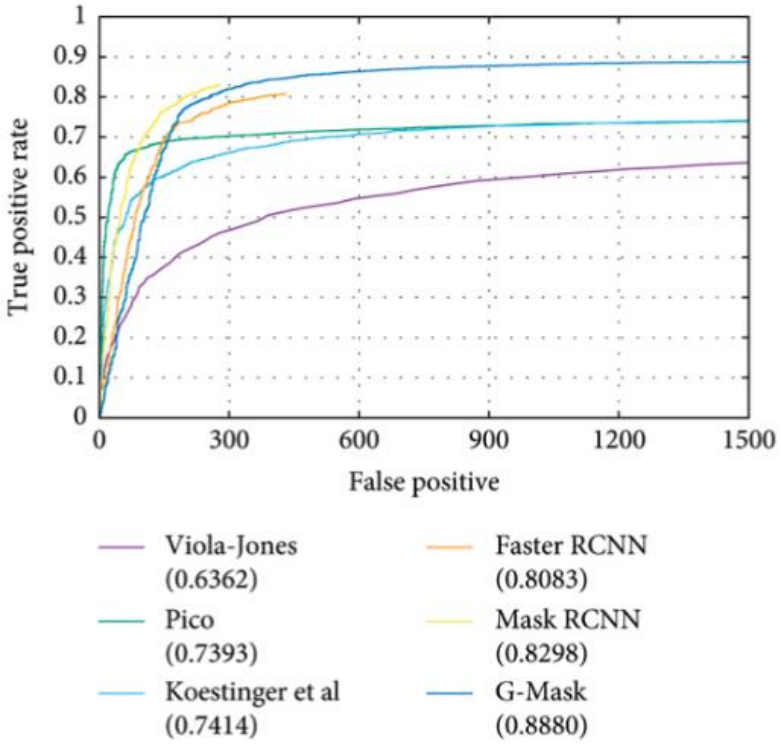
شکل ۳- نتایج مقایسه مقاله با روش ground truth در مجموعه داده FDDB (Jiang & Learned-Miller, 2017)

همانگونه که در شکل ۳ نیز نشان داده شده است، روش Faster RCNN فقط نواحی صورت در تصویر را تشخیص داده است در صورتی که روش دیگر، تواحی بیشتری مانند مو را هم در کادر قرار داده است که نشان دهنده عملکرد بهتر مدل Faster RCNN بوده است (Jiang & Learned-Miller, 2017).

روش های تشخیص چهره موجود، فقط با استفاده از یک کادر محدود، هر چهره را بومی‌سازی می‌کنند، که نمی‌تواند هر چهره را به طور همزمان از تصویر پس زمینه جدا کند. برای حل این مشکل، لین و همکاران در سال ۲۰۲۰ یک روش تشخیص چهره و بخش‌بندی مبتنی بر Mask R-CNN بهبود یافته به نام G-Mask ارائه کرده‌اند. این روش تشخیص چهره و تقسیم‌بندی را در یک چارچوب با هدف به دست آوردن اطلاعات دقیق‌تر از چهره ترکیب می‌کند. در این روش،ResNet-۱۰۱ برای استخراج ویژگی‌ها استفاده می‌شود،RPN برای تولید RoI استفاده می‌شود و RoIAlign به طور صادقانه مکان‌های مکانی دقیق را برای تولید ماسک باینری از طریق شبکه کاملاً پیچشی (FCN) حفظ می‌کند. برای بهبود دقت تشخیص نیز از GIoU[۲] استفاده می شود. در این مقاله به منظور افزایش قابلیت اطمینان نمونه ها، ۵۱۱۵ نمونه را از مجموعه داده های FDDB و ChokePoint انتخاب کرده و آنها را با برچسب های ماسک حاشیه نویسی کرده است و پس از کار حاشیه نویسی، مدل G-Mask روی این مجموعه داده آموزش داده شده است. از شرایط مورد استفاده در این مقاله نیز به این صورت است که پیشنهاد منطقه تنها در صورتی چهره محسوب می شود که امتیاز اطمینان بیشتر از ۰,۷ باشد. در مقایسه با Faster R-CNN،Mask R-CNN وMultitask Cascade CNN ، روش G-Mask به نتایج امیدوارکننده‌ای در معیارهای FDDB ، AFW، و WIDER FACE دست یافته است.

چارچوب پیشنهادی قادر است چهره ها را به درستی تشخیص دهد و در عین حال هر چهره را در یک تصویر به طور دقیق تقسیم کند. مدل G-Mask نه تنها به محلی‌سازی جعبه مرزی هدف چهره پی برد، بلکه اطلاعات چهره را با ماسک باینری از تصویر پس زمینه جدا کرد، به طوری که اطلاعات چهره با جزئیات بیشتری از طریق فرآیند فوق به دست آمد. نتایج مقایسه این روش با روش های دیگر نیز در شکل ۴ آمده است. در شکل ۵ نیز نتایج مقایسه این روش با روش پایه Mask RCNN آمده است (Lin et al., 2020).

[۲] Generalized Intersection over Union



شکل ۴- نمودار مقایسه روش G-Mask RCNN با روش های دیگر (Lin et al., 2020)

شکل ۵- مقایسه نتایج Mask RCNN و G-Mask RCNN (Lin et al., 2020)

در شکل ۵ ابتدا نتایج استخراج شده از مدل G-Mask را نشان داده و سپس نتایج مدل Mask. همانگونه که در شکل نیز نشان داده شده است، روش G-Mask توانسته همه چهره های موجود در تصاویر را تشخیص دهد در صورتی که روش Mask برخی چهره ها را نتوانسته تشخیص دهد.

مانند مسائل در زمینه های دیگر، در روش های تشخیص چهره در تصاویر نیز با مشکل دقت کم تشخیص چهره در شرایط انسداد پیچیده روبرو هستیم که جهت حل این مشکل لی و همکاران[۴] در سال ۲۰۲۲ به ارائه روش CNN دو کاناله با انسداد پرسپترون پرداخته اند. این روش به این صورت عمل می کند که ابتدا واحد ناحیه occlusion judgment مشخص شده و سپس شبکه VGG16 یکپارچه شده است تا یک شبکه عصبی پرسپترون انسداد را تشکیل دهد سپس ویژگی‌های مناطق مسدود نشده و مناطق کمتر مسدود شده در تصاویر چهره توسط شبکه عصبی ادراکی استخراج شده و دو مجموعه داده قابل دسترسی AR و MAFA آزمایش شده است. همچنین از الگوریتم یادگیری انتقال برای پیش‌آموزش پارامترهای لایه کانولوشن برای کاهش مشکل اضافه برازش ناشی از نمونه‌های داده آموزشی ناکافی استفاده شده و استخراج ویژگی‌های چهره از کل چهره با بهینه سازی شبکه و سپس ویژگی‌های چهره شبکه عصبی پرسپترون مسدود کننده و شبکه باقی مانده وزن گذاری و ترکیب شده اند. از معایب این روش می توان به زمان زیاد جهت آموزش الگوریتم در این مقاله اشاره کرد (Li, ۲۰۲۲).

در سال های اخیر نیز استفاده از نسخه های مختلف روش YOLO جهت تشخیص چهره در تصاویر استفاده شده است برای مثال خلیلی و همکاران[۵] در سال ۲۰۲۲ از روش YOLO برای شناسایی چهره استفاده کرده است. این روش از معیار WIDER FACEبرای آموزش YOLOv1 تا v4 در چارچوب Darknet و ترکیب نتایج آنها با دو روش، یعنی همجوشی جعبه‌های وزنی[۶] و غیر حداکثر وزنی[۷] استفاده کرده است. تجزیه و تحلیل تجربی نشان داد که mAP در مجموعه WBF مدل‌ها برای همه تصاویر آسان، متوسط و سخت در مجموعه داده‌ها به ترتیب ۷,۸۱، ۲۲,۹۱ و ۱۲,۹۶ درصد افزایش می‌یابد در حالیکه این اعداد برای گروه NMW برابر ۶,۲۵ ٪، ۲۰,۸۳ ٪ و ۱۱,۱۱٪ است (Khalili & Shakiba, 2022).

Gao و همکاران[۸] در سال ۲۰۲۲ از روش TinyYOLOv3 بهبود یافته و مکانیزم توجه جهت تشخیص چهره در تصاویر استفاده کرده است. این روش با توجه به نرخ تشخیص کم روش‌های تشخیص چهره سنتی در پس زمینه پیچیده و زمان تشخیص طولانی روش‌های تشخیص چهره موجود مبتنی بر یادگیری عمیق به وجود آمده است. استفاده از شبکه اصلی TinyYOLOv3و طراحی مجدد آن برای استخراج اطلاعات معنایی فراوان تر، که باعث افزایش تعداد لایه های شبکه می شود. استفاده از تقاطع کامل روی اتحاد برای بهبود این از دست دادن پیش‌بینی اولیه مختصات جعبه مرزی. افزایش سرعت همگرایی شبکه با بهینه سازی هایپرپارامتر و روش خوشه بندی جعبه پیشینی استفاده شده. آزمایش‌های تشخیص چهره بر روی مجموعه داده‌های Wider Face انجام شده و نتایج تجربی نشان دهنده دقت تشخیص بالای الگوریتم پیشنهادی در صحنه های پیچیده می باشد همچنین الگوریتم پیشنهادی از نظر سرعت تشخیص و اندازه مدل بهتر از سایرین است (Gao & Yang, 2022).

وو[۹] و همکاران در سال ۲۰۲۲ از روش FMD-Yolo جهت پیشگیری و کنترل موثر بیماری کرونا استفاده کرده اند. این روش جهت نظارت بر اینکه آیا افراد در مکان‌های عمومی ماسک می‌پوشند یا خیر استفاده شده که راهی موثر برای جلوگیری از انتقال ویروس است. در این روش ابتدا از استخراج کننده ویژگی از Im-Res2Net-101 استفاده می کند که ماژول Res2Net و شبکه باقیمانده عمیق را ترکیب می کند سپس یک شبکه تجمیع مسیر پیشرفته En-PAN برای

---

[۴] Li et al.

[۵] Khalili et al.

[۶] WBF

[۷] NMW

[۸] Gao et al.

[۹] Wu et al.

ترکیب ویژگی‌ها اعمال می‌شود همچنین از روش ماتریس NMS در مرحله استنتاج برای بهبود کارایی و دقت تشخیص استفاده می شود. ارزیابی روش بر روی دو پایگاه داده عمومی با نتایج در مقایسه با هشت الگوریتم تشخیص پیشرفته دیگر انجام می شود. نتایج نشان می دهد که در سطح IoU=0.5 ، FMD-Yolo پیشنهادی به بهترین دقت ٪۹۲,۰ و ٪۸۸,۴ در دو مجموعه داده دست یافته است و در IoU=0.75 به ترتیب ٪۵,۵ و ٪۳,۹ در مقایسه با دومی بهبود یافته است که نشان دهنده برتری FMD-Yolo در تشخیص ماسک می باشد (Wu et al., 2022).

CNN ها هنوز به طور کامل از اطلاعات ساختار چهره بهره برداری نمی‌کنند که باعث غیرقابل تشخیص شدن چهره هایی با زوایای چرخش شدیدتر می شود از این رو، ژو و همکاران[۱] در سال ۲۰۲۲ از روش MTCNet برای تشخیص چهره هایی که چرخش دارند استفاده کرده است. در این روش برای تشخیص چهره چرخشی غیر متغیر راه حلی ارائه شده که به طور کامل از نشانه های چهره برای بهبود عملکرد تشخیص با استفاده از همکاری بین تشخیص چهره و تراز چهره استفاده می کند. این روش از یک معماری آبشاری با سه مرحله برای پیش‌بینی چهره‌ها با دامنه‌های چرخش در صفحه در حال کاهش تدریجی در یک فرآیند درشت به ریز استفاده می‌کند همچنین از زاویه چرخش برای بهبود عملکرد تراز چهره و یکپارچه سازی زمینه استفاده می کنیم. همچنین اطلاعات زمینه‌ای را بررسی می‌کند تا چهره های چالش برانگیز را از سناریوهای نامحدود تشخیص دهد. این روش بر روی هر دو مجموعه داده جهت گیری و چرخش چندگانه انجام شده است. همچنین آزمایش‌های گسترده‌ای را بر روی چندین مجموعه داده معیار محبوب از جمله مجموعه دادهWIDER FACE ، FDDB و AFLW انجام شده است. نتایج نشان می دهد این روش نتایجی رقابتی با پیشرفته‌ترین آشکارسازهای چهره دست می‌یابد در حالی که از نظر زمان کارآمد است. از مزایای این روش نیز می توان این روش قادر به پیش‌بینی زاویه RIP و محلی‌سازی نقطه عطف با هزینه زمانی کمتر می باشد و ابتدا بین چهره‌ها و غیر چهره‌ها تمایز قائل می‌شود، سپس پیش‌بینی زاویه و مکان‌یابی نقطه عطف چهره را انجام می‌دهد (Zhou et al., 2022).

۳. نتیجه‌گیری

در این مقاله به بررسی و مقایسه شبکه های عصبی شناسایی چهره پرداخته شد که از جهات مختلفی از جمله سرعت و دقت، ساختار و معماری و زمان اجرای هر کدام از شبکه ها مورد مقایسه و ارزیابی قرار گرفتند. تحقیقات انجام گرفته نشان میدهد شبکه عصبی Faster RCNN سریعتر از سایر شبکه ها و همچنین با توجه به اهمیت زیاد پارامتر دقت یا صحت شبکه عصبی YOLO بیشترین دقت و شبکه G-Mask RCNN بیشترین زمان اجرا را نسبت به سایر شبکه ها دارند. با توجه به ویژگی و مزیت منحصر به فرد هرکدام از این شبکه ها، به نظر می رسد به عنوان کارهای آتی ترکیب و استفاده مشترک از این شبکه ها می تواند عملکرد بهتری را از جهت سرعت و دقت جهت شناسایی چهره به همراه داشته باشد.

منابع

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. Remote Sensing, 9(4), 312 .

Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, 300, 17-3۳ .

Gao, J., & Yang, T. (2022). Face detection algorithm based on improved TinyYOLOv3 and attention mechanism. Computer Communications, 181, 329-33۷ .

Jiang, H., & Learned-Miller, E. (2017). Face detection with the faster R-CNN. 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) ,

[۱] Zhou et al.

Khalili, S., & Shakiba, A. (2022). A face detection method via ensemble of four versions of YOLOs. ۲۰۲۲ International Conference on Machine Vision and Image Processing (MVIP) ,

Li, Y. (2022). Face Detection Algorithm Based on Double-Channel CNN with Occlusion Perceptron. Computational Intelligence and Neuroscience, 2022 .

Lin, K., Zhao, H ,.Lv, J., Li, C., Liu, X., Chen, R., & Zhao, R. (2020). Face detection and segmentation based on improved mask R-CNN. Discrete dynamics in nature and society, 2020 .

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. IEEE Transactions on pattern analysis and machine intelligence, 22(12), 1424-1445 .

Wu, P., Li, H., Zeng, N., & Li, F. (2022). FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. Image and vision computing, 117, 104341 .

Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. Neurocomputing, 396, 39-64 .

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding, 138, 1-24 .

Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), 399-458 .

Zhou, L., Zhao, H., & Leng, J. (2022). MTCNet: Multi-task collaboration network for rotation-invariance face detection. Pattern Recognition, 124, 108425 .