

تشخیص و شناسایی چهره در پهپادها با یادگیری عمیق

امیر حمزه فرج‌الهی^{۱*}، محسن رستمی^۲، هاشم پروین^۳، بهزاد نظرپور^۴، مجتبی لک^۵

۱، ۲، ۴ و ۵- استادیار دانشگاه امام علی (ع)، ۳- دانشجوی دکتری دانشگاه اصفهان

(دریافت: ۱۴۰۱/۰۳/۰۹، پذیرش: ۱۴۰۱/۰۸/۰۸)

چکیده

شناسایی چهره انسان با پهپادها، برای کاربردهای مختلف، مانند نظارت، جستجو و امنیت ضروری است. روش‌های قبلی برای تشخیص و شناسایی چهره حساسیت بالایی به محدودیت‌هایی مانند، ارتفاع، زاویه و فاصله از چهره دارند. در این مقاله رویکرد جدیدی برای تشخیص و شناسایی چهره با یادگیری عمیق ارائه می‌شود. روش پیشنهادی در سه مرحله انجام می‌شود. در مرحله اول، با الگوریتم جستجوی انتخابی، ناحیه‌بندی تصاویر انجام می‌شود. در مرحله دوم، یک شبکه عمیق برای عملیات پالایش جعبه‌ها پیشنهاد می‌شود تا جعبه‌های هدف با دقت و سرعت بالایی شناسایی شوند. در واقع، یک مسئله طبقه‌بندی دو کلاسه توسط یادگیری عمیق انجام می‌شود تا چهره‌ها مکان‌یابی شوند. در مرحله سوم، تصاویر مکان‌یابی شده به شبکه عمیق پیشنهادی آموزش داده می‌شوند تا شناسایی چهره‌ها انجام شود. در معماری روش پیشنهادی از خاصیت شبکه‌های عمیق پرکاربرد به صورت ترکیبی استفاده می‌شود و مقایسه کمی روش پیشنهادی با روش‌های جدید از نظر پیچیدگی محاسباتی نشان می‌دهد که آموزش مدل پیشنهادی نسبت به سایر روش‌ها زمان اجرای کمتری لازم دارد. به علاوه، ارزیابی روش پیشنهادی روی مجموعه داده DroneFace نشان می‌دهد که برای فاصله و ارتفاع مختلف نسبت به هدف، روش پیشنهادی میانگین نرخ تشخیص چهره ۷۵/۹ و میانگین نرخ شناسایی چهره ۸۴/۶ را دارد. بنابراین، روش پیشنهادی نسبت به روش‌های جدید در این حوزه دقت و کارایی بالاتری دارد و می‌تواند برای کاربردهای نظارتی و امنیتی استفاده شود.

کلیدواژه‌ها: مکان‌یابی چهره، شناسایی چهره، یادگیری عمیق، پهپادها، پردازش تصویر، ناحیه‌بندی تصویر.

Face Detection and Identification in Drones with Deep Learning

A. H. Farajollahi*, M. Rostami, H. Parvin, B. Nazarpour, M. Lak

University of Imam Ali

(Received: 30/05/2022, Accepted: 30/10/2022)

Abstract

Human face recognition by drones is necessary for various applications, such as surveillance, search and security. The previous methods for face recognition are highly sensitive to limitations such as height, angle and distance from the face. In this article, a new approach for face detection and identification by deep learning is presented. The proposed method is done in three steps. In the first step, images are zoned with the selective search algorithm. In the second step, a deep network is proposed for box refinement operation to identify the target boxes with high accuracy and speed. Actually, a two-class classification problem is performed by deep learning to locate faces. In the third step, the localized images are trained to the proposed deep network to perform face recognition. In the architecture of the proposed method, the properties of widely used deep networks are used in combination, and a quantitative comparison of the proposed method with new methods in terms of computational complexity shows that training the proposed model requires less execution time than other methods. In addition, the evaluation of the proposed method on the DroneFace dataset shows that for different distance and height from the target, the proposed method has an average face recognition rate of 75.9 and an average face recognition rate of 84.6. Therefore, the proposed method has higher accuracy and efficiency than the new methods in this field and can be used for surveillance and security applications.

Keywords: Face Localization, Face Recognition, Deep Learning, Drones, Image Processing, Image Segmentation.

*Corresponding Author E-mail: a.farajollahi@sharif.edu

۱. مقدمه

امروزه هواپیماهای بدون سرنشین (پهپادها)^۱ تأثیر زیادی در افزایش توان امنیتی کشور و کاربردهای نظارتی دارند. پهپادها به دلیل پرواز در ارتفاع و مسافت طولانی، برابردیابی انسان‌ها، سنجش از دور، نظارت و ردیابی استفاده می‌شوند [۱]. تشخیص^۲ و شناسایی^۳ چهره افراد در مکان‌های مختلف، یکی از کاربردهای مهم برای پهپادها محسوب می‌شود که امروزه توسط هوش مصنوعی^۴ انجام می‌شود [۲]. قبل از این مأموران امنیتی مجبور بودند به صورت شخصی فرآیند تشخیص هویت را انجام دهند که این امر زمان زیاد لازم دارد و خطای تشخیص نیز بیشتر است.

لذا، با استفاده از این فناوری در پهپادها، فرآیند شناسایی افراد فراریا افرادی که غیرقانونی در مکانی حضور دارند، روندی مستقل و مشخص دارد و زمان بسیار کوتاهی لازم است تا با درصد خطای بسیار کمی فرآیند شناسایی انجام شود.

امروزه، الگوریتم‌های هوش مصنوعی و پردازش تصویر به پهپادها کمک می‌کنند تا تشخیص و شناسایی افراد به صورت خودکار انجام شود. پهپادها در محیط‌های مختلفی استفاده می‌شوند و این محیط‌ها چالش‌های زیادی برای پهپاد ایجاد می‌کند [۳]. لذا، تشخیص و شناسایی چهره در ارتفاع و فاصله زیاد، بر کیفیت تصاویر چهره تأثیر دارد و دقت تشخیص و شناسایی چهره را کاهش می‌دهد. الگوریتم‌های مختلفی برای تشخیص و شناسایی چهره در دهه‌های گذشته ارائه شده‌اند که در دو دسته تقسیم‌بندی می‌شوند: روش‌های مبتنی بر پردازش تصویر و روش‌های مبتنی بر یادگیری ماشین [۴].

در روش‌های پردازش تصویر، ویژگی‌های دست‌ساز^۵ برای استخراج ویژگی از تصویر استفاده می‌شود. به علاوه، فیلترهای مکانی و فرکانسی مختلفی برای تمایز بین تصاویر چهره و طبقه‌بندی آن‌ها استفاده می‌شود. این روش‌ها، حساسیت بالایی به نویز و کاهش کیفیت تصویر دارند. لذا، تغییرات در تصویر، کاهش نور، نویز و تغییر مقیاس به شدت دقت و کارایی این روش‌ها را کاهش می‌دهد [۶]. در روش‌های مبتنی بر یادگیری ماشین، ویژگی‌های تصویر به صورت خودکار با شبکه عصبی عمیق استخراج می‌شود و هم‌زمان استخراج ویژگی و طبقه‌بندی تصویر انجام می‌شود. در این روش‌ها، با مجموعه داده مشخصی عملیات یادگیری ویژگی انجام شده و بردارهای ویژگی برای آموزش طبقه‌بند استفاده می‌شوند [۷]. روش‌های مبتنی بر یادگیری ماشین و یادگیری عمیق نسبت به روش‌های پردازش تصویر در

مقابل تغییرات مقاوم‌اند و می‌توانند ویژگی‌های مناسب را استخراج کنند [۸]. علاوه، روش‌های یادگیری عمیق نسبت به روش‌های یادگیری ماشین ویژگی‌های بهتری استخراج می‌کنند و عملیات استخراج ویژگی و طبقه‌بندی به صورت هم‌زمان انجام می‌شود [۹].

در این پژوهش، به دلیل کارایی بالای روش‌های یادگیری عمیق در استخراج ویژگی و طبقه‌بندی داده، روشی مبتنی بر شبکه‌های عمیق برای شناسایی چهره روی تصاویر گرفته‌شده با پهپادها ارائه می‌شود. روش پیشنهادی در سه مرحله انجام می‌شود، در مرحله اول، پیش‌پردازش تصاویر ورودی انجام شده و تصاویر چهره با الگوریتم ناحیه‌بندی انتخابی^۶ می‌شوند تا جعبه‌های^۷ هدف و غیر هدف تولید شوند. در مرحله دوم، یک شبکه عمیق برای طبقه‌بندی جعبه‌ها پیشنهاد می‌شود و مکان‌یابی چهره‌ها براساس جعبه‌های هدف انجام می‌شود. در مرحله سوم، تصاویر مکان‌یابی شده به شبکه عمیق پیشنهادی آموزش داده می‌شوند تا شناسایی چهره انجام شود. مهم‌ترین نوآوری‌های این مقاله به شرح زیر است:

- روش پیشنهادی از الگوریتم ناحیه‌بندی انتخابی برای مکان‌یابی چهره استفاده می‌کند؛ بنابراین، روش پیشنهادی در مقابل تغییرات در تصویر، کاهش نور، نویز و تغییر مقیاس مقاوم است. در حالی که، روش‌های قبلی حساسیت بالایی به نویز و تغییرات در تصویر دارند. لذا، روش پیشنهادی کارایی بهتری نسبت به روش‌های قبلی [۳، ۱۰ و ۱۱] دارد.

- در روش پیشنهادی پالایش و طبقه‌بندی نواحی تصویر، مبتنی بر ماژول Inception است که خاصیت شبکه‌های GoogleNet و ResNet را برای آموزش و بهینه‌سازی بهتر استفاده می‌کند. همچنین، در معماری پیشنهادی لایه‌های مختلف کانولوشن، کاهش اندازه و تراکم استفاده می‌شود تا تغییرات داده ورودی به مدل آموزش داده شود. لذا، روش پیشنهادی دقت مکان‌یابی چهره بهتری نسبت به روش‌های قبلی دارد [۵، ۱۰، ۱۲ و ۱۳].

- برای شناسایی چهره‌های مکان‌یابی شده، در روش پیشنهادی یک شبکه عمیق مجزا طراحی می‌شود که براساس لایه‌های مختلف کانولوشن، کاهش اندازه و نرمال‌سازی، به خوبی طبقه‌بندی و شناسایی چهره‌ها را انجام می‌دهد. لذا، معماری پیشنهادی استخراج ویژگی خودکار مبتنی بر یادگیری ویژگی را استفاده می‌کند و نسبت به روش‌های دست‌ساز^۸ [۲، ۱۴ و ۱۵] ویژگی‌های بهتری از تصویر استخراج می‌کند.

- اغلب روش‌های قبلی نسبت به محدودیت‌های ارتفاع، زاویه و فاصله از چهره بی‌توجه بوده‌اند [۸، ۱۶-۲۰]، اما در روش

^۶ Selective Segmentation

^۷ Boxes

^۸ Hand-craft

^۱ Drones

^۲ Detection

^۳ Recognition

^۴ Artificial Intelligence

^۵ Hand-craft

ماتای و همکارانش [۲۶]، تشخیص چهره را با استفاده از طبقه‌بند آدابوست^۱ و ویژگی‌های هار^۲ انجام دادند که حساسیت بالایی به تغییرات تصویر دارد. خیو و همکارانش [۱]، یک الگوی باینری محلی برای تشخیص چهره روی پهپادهای تجاری برای کاربردهای امنیتی ارائه دادند. این روش هزینه اجرایی پایینی دارد و قابلیت استفاده گسترده دارد. با این حال، این روش میزان اثربخشی و محدودیت‌های فاصله و ارتفاع از هدف را ارزیابی نمی‌کند. گوتو و همکارانش [۲۷]، طراحی و اعتبارسنجی مدل آشکارساز چهره را براساس تبدیل کسینوسی گسسته و تحلیل مؤلفه اصلی ارائه کردند. در این روش، تغییر مقیاس و افزایش نویز دقت و کارایی را کاهش می‌دهد. نویسندگان در [۳] مدلی را ارائه کردند که تشخیص چهره‌ها را براساس استخراج ویژگی‌های قوی تسریع شده، مبتنی بر شناسه چهره در پهپاد انجام می‌دهد. این مدل از روش‌های مختلفی مانند سورف^۳، طبقه‌بند هار و چهره ویژه برای طبقه‌بندی چهره استفاده می‌کند. ارزیابی این مدل در محیط شبیه‌سازی شده انجام می‌شود تا کارایی روش برای شرایط مختلف زاویه، فاصله و ارتفاع بررسی شود. فارلیک و همکارانش [۲۸]، یک روش برای شناسایی چهره براساس ویژگی‌های دست‌ساز روی پهپادها ارائه دادند که پایداری محیطی مناسبی دارد و نسبت به کاهش نویز حساسیت کمتری دارد. در حالی که، این روش نسبت به تغییر مقیاس در تصاویر چهره ناپایدار است. وانگ و همکارانش [۲۹]، ویژگی‌های اضافی را به الگوریتم تشخیصی اضافه کردند و برای نظارت و امنیت استفاده می‌شود. این روش پایداری مطلوبی دارد و تغییرات اندازه و فاصله را به مدل اضافه می‌کند. با این حال، این روش دارای محدودیت‌های خاصی در ارتباط با تغییر رزولوشن^۴ در تصویر دارد. آتاماجا و همکارانش [۲]، روشی مبتنی بر ویژگی‌های هار برای تشخیص چهره روی پهپادها ارائه دادند. در این روش فرآیند شناسایی چهره با سرعت بالایی در فاصله ۵۰ متری از هدف انجام می‌شود، در حالی که نسبت به تغییر زاویه چهره بسیار حساس است. کامپلا و همکارانش [۴]، یک پهپاد جاسوسی با الگوریتم محلی‌سازی چهره براساس اطلاعات ویژه صورت برای تشخیص چهره آموزش دادند. ساختار کلی این روش، شامل استخراج ویژگی، تشخیص چهره و طبقه‌بندی است.

تصاویر گرفته‌شده توسط پهپادها، اغلب چالش‌های بیشتری نسبت به تصاویر معمولی دارند. بنابراین، برخی از عوامل محیطی مانند، نویز، فاصله و ارتفاع باعث کاهش دقت مدل‌های تشخیص چهره در

پیشنهادی با معماری مناسب این چالش‌ها تا حد زیادی برطرف می‌شوند.

• روش پیشنهادی از راهکار بهینه‌سازی مناسب استفاده می‌کند، لذا سرعت همگرایی بالا، پیچیدگی زمانی کمتر و مقیاس‌پذیری بهتری نسبت به سایر روش‌ها دارد. همچنین، پیاده‌سازی روش پیشنهادی آسان است و ساختمان داده خاصی نیاز ندارد. نتایج تجربی روی مجموعه داده مشهور، نشان می‌دهد که روش پیشنهادی، دقت و کارایی بهتری نسبت به روش‌های قبلی دارد.

ساختار این مقاله چنین است. در بخش ۲ مروری بر کارهای پیشین در حوزه مرتبط با این پژوهش ارائه می‌شود. در بخش ۳ روش پیشنهادی همراه با جزئیات کامل ارائه می‌شود. در بخش ۴ برای ارزیابی روش پیشنهادی در مقایسه با سایر روش‌ها، آزمایش‌های تجربی روی مجموعه داده مشخصی انجام شده است. در بخش ۵ نتیجه‌گیری و پیشنهاد برای انجام کارهای آتی آمده است.

۲. پیشینه تحقیق

در این بخش، مروری بر روش‌های قبلی برای تشخیص و شناسایی چهره برای کاربردهای مختلف انجام می‌شود. امروزه یادگیری ماشین و پردازش تصویر کارایی بالایی در مسائل مختلف دسته‌بندی دارند [۲۱، ۲۲]. تشخیص و شناسایی چهره یک فناوری است که برای شناسایی یک شخص از یک منبع ویدیو یا تصویر استفاده می‌شود. ایده تشخیص و شناسایی چهره دارای یک پایه علمی است که به دهه ۱۹۹۰ میلادی باز می‌گردد [۲۳]. پس از آن، این سامانه‌ها مداوم در حال بهینه‌سازی و بهبود بوده و به‌طور گسترده در زندگی روزمره مردم، برای پزشکی قانونی، نظامی و امنیت استفاده می‌شود [۲۴]. مدل‌های تشخیص و شناسایی چهره در دو دسته طبقه‌بندی می‌شوند: روش‌های مبتنی بر پردازش تصویر و روش‌های مبتنی بر یادگیری ماشین. در روش‌های پردازش تصویر، فیلترهای مکانی و فرکانسی مختلفی برای استخراج ویژگی‌های دست‌ساز استفاده می‌شوند. در واقع، این روش‌ها براساس ویژگی‌های ثابتی کلاسه‌بندی تصاویر را انجام می‌دهند. بنابراین، روش‌های مبتنی بر پردازش تصویر، حساسیت زیادی به نویز و کاهش کیفیت تصویر دارند. لذا، تغییراتی مانند، کاهش نور، نویز و تغییر مقیاس به‌شدت کارایی این روش‌ها را کاهش می‌دهد [۱۶].

برای نمونه، گو و همکارانش [۲۵]، یک روش جدید برای افزایش سرعت تشخیص چهره مبتنی بر ویژگی‌های هار ارائه دادند که با ساختار موازی، دقت و عملکرد مطلوبی برای تشخیص چهره دارد.

¹ Adaboost

² Haar

³ SURF

⁴ Resolution

ویژگی‌های مناسب را پیدا نمی‌کند. دیب و همکارانش [۵]، از یادگیری عمیق برای طبقه‌بندی ویژگی‌های چهره برای افزایش کارایی شبکه در تشخیص چهره استفاده می‌کنند. این روش، الگوریتم‌های مختلف یادگیری عمیق را برای تشخیص چهره روی مجموعه داده مشهور ترکیب می‌کند. در این روش تغییر ارتفاع و فاصله چهره از پهپاد، برای طبقه‌بندی استفاده می‌شود. همچنین، با استفاده از شبکه‌های عمیق همبستگی میان برچسب‌ها و تصاویر آموزشی جستجو می‌شود. با این حال، این روش پیچیدگی محاسباتی بالا و مقیاس‌پذیری پایینی دارد. در راستای برطرف نمودن چالش‌های ذکر شده برای مکان‌یابی و شناسایی چهره در ارتفاع و فاصله زیاد، در این مقاله، رویکرد جدیدی برای تشخیص و شناسایی چهره مبتنی بر یادگیری عمیق ارائه می‌شود. در روش پیشنهادی از الگوریتم ناحیه‌بندی انتخابی همراه با دو شبکه عمیق پیشنهادی مکان‌یابی و شناسایی چهره با دقت و سرعت بالا انجام می‌شود.

۳. روش پیشنهادی

در روش پیشنهادی برای رفع چالش‌های موجود در سامانه‌های تشخیص و شناسایی چهره، از راهکار ناحیه‌بندی انتخابی و شبکه‌های عمیق استفاده می‌شود. در واقع، برای تشخیص و شناسایی دقیق‌تر چهره‌ها در فاصله و ارتفاع زیاد، یک روش جدید مبتنی ناحیه‌بندی انتخابی و یادگیری عمیق ارائه می‌شود تا دقت و کارایی سیستم افزایش یابد. روش پیشنهادی در سه مرحله کلی انجام می‌شود. در مرحله اول، براساس الگوریتم ناحیه‌بندی انتخابی، جعبه‌های مختلف پیدا می‌شوند. در مرحله دوم، یک شبکه عصبی عمیق برای پالایش جعبه‌ها طراحی می‌شود و مکان‌یابی چهره‌ها براساس جعبه‌های هدف انجام می‌شود. در مرحله سوم، با استفاده از یک شبکه پیشنهادی، چهره‌های مکان‌یابی شده، شناسایی می‌شوند. با توجه به اینکه پردازش‌های لازم برای تشخیص و شناسایی چهره در بخش هوایی به توان زیادی نیاز دارد. لذا، در روش پیشنهادی شناسایی با استفاده از یادگیری عمیق در بخش زمینی انجام می‌شود که توان کمتری لازم دارد. تصاویر مرتبط با بخش‌های مختلف روش پیشنهادی در شکل (۱) نشان داده شده است. در ادامه مراحل روش پیشنهادی با جزئیات توضیح داده می‌شود.

پهپادها می‌شوند. مدل‌های تشخیص چهره قبلی روی پهپادها، با محدودیت‌هایی مانند فاصله، ارتفاع و زاویه چهره‌ها را شناسایی می‌کنند که باعث کاهش دقت و کارایی آن‌ها می‌شود [۳۰]. در این راستا، روش‌های تشخیص چهره با یادگیری ماشین/عمیق توسعه داده می‌شوند. در روش‌های مبتنی بر یادگیری ماشین و یادگیری عمیق، ویژگی‌های تصویر به صورت خودکار با شبکه عمیق استخراج می‌شود. لذا، استخراج ویژگی و طبقه‌بندی تصاویر هم‌زمان انجام می‌شود. در روش‌های یادگیری عمیق با مجموعه داده مناسب، یادگیری ویژگی‌ها انجام می‌شود و بردارهای ویژگی برای آموزش مدل استفاده می‌شوند. روش‌های مبتنی بر یادگیری ماشین/عمیق نسبت به روش‌های پردازش تصویر در مقابل تغییرات مقاوم بوده و ویژگی‌های بهتری را استفاده می‌کنند. برای نمونه، لین و همکارانش [۳۱]، یک تصمیم‌گیری احتمالی برای تشخیص چهره براساس شبکه عمیق ارائه دادند. در این روش، ویژگی‌های خودکار با مدل عمیق احتمالاتی برای تشخیص چهره استخراج می‌شوند. نایر و همکارانش [۳۲]، روشی پایدار و دقیق برای بخش‌بندی و تشخیص چهره‌ها، بر اساس تکه‌های چهره و برآزش اطلاعات ارائه دادند. این روش، الگوریتم توزیع نقطه سه‌بعدی بدون وابستگی به داده‌های بافت و جهت را استفاده می‌کند. یانگ و همکارانش [۷]، روشی مبتنی بر یادگیری عمیق برای تشخیص چهره به صورت باناظر براساس ویژگی‌های چهره ارائه دادند که با یک آشکارساز ویژگی‌های تصاویر چهره به صورت صریح تشخیص و کلاسه‌بندی می‌شوند. فانگ و همکارانش [۳۳]، از الگوریتم مشهور یولو^۱ برای تشخیص چهره استفاده کردند تا ویژگی‌های مناسب از تصویر با لایه‌های کانولوشن و کاهش اندازه استخراج شود و کلاسه‌بندی ویژگی‌ها انجام شود. این روش سرعت تشخیص چهره بالایی دارد، اما دقت شناسایی چهره پایین است.

آلمادی و همکارانش [۳۴]، یک شبکه پیش‌آموز^۲، برای افزایش کارایی تشخیص چهره براساس شبکه عمیق الکسنت^۳ طراحی کردند. این روش، ویژگی‌های عمیق استخراج‌شده را به طبقه‌بند بردار پشتیبان^۴ آموزش می‌دهد و سپس با یادگیری انتقالی مبتنی بر کانولوشن طبقه‌بندی انجام می‌شود. لی و همکارانش [۳۵]، یک مدل استخراج ویژگی مبتنی بر تمرکز^۵ برای استخراج داده‌های معنایی ارائه دادند که محلی‌سازی چهره را به صورت چند مقیاسی انجام می‌دهد. همچنین، یک نمایش مطلوب برای ترکیب ویژگی‌های سطح بالا و سطح پایین ارائه می‌کند. این روش نسبت به تغییر ارتفاع چهره حساسیت بالایی دارد و در این حالت،

¹YOLO

²Pre-train

³AlexNet

⁴SVM

⁵Attention

⁶Representation

$$S_{color}(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k) \quad (1)$$

که در آن، مقدار هیستوگرام برای k امین بین^۴ در توصیف‌گر رنگی است.

شباهت بافتی: ویژگی‌های بافت با استخراج مشتقات گاوسی در ۸ جهت برای هر کانال محاسبه می‌شود. برای هر جهت و هر کانال رنگی، یک هیستوگرام ۱۰ تایی محاسبه می‌شود که منجر به یک توصیف‌گر ویژگی ۱۰×۸×۳ = ۲۴۰ بعدی می‌شود. تشابه بافت دو ناحیه نیز با استفاده از تقاطع‌های هیستوگرام با رابطه (۲) محاسبه می‌شود.

$$S_{texture}(r_i, r_j) = \sum_{k=1}^n \min(t_i^k, t_j^k) \quad (2)$$

که در آن، مقدار هیستوگرام برای k امین بین در توصیف‌گر بافت است.

شباهت اندازه: شباهت اندازه مناطق کوچک‌تر را تشویق می‌کند تا زودتر ادغام شوند، لذا انتخاب ناحیه در تمام مقیاس‌های تصویر انجام می‌شود. اگر این معیار شباهت در نظر گرفته نشود، یک ناحیه واحد تمام مناطق کوچک‌تر را ترکیب می‌کند و انتخاب ناحیه در مقیاس‌های متعدد فقط در این مکان ایجاد می‌شود. شباهت اندازه به‌صورت رابطه (۳) تعریف می‌شود:

$$S_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size(r_{im})} \quad (3)$$

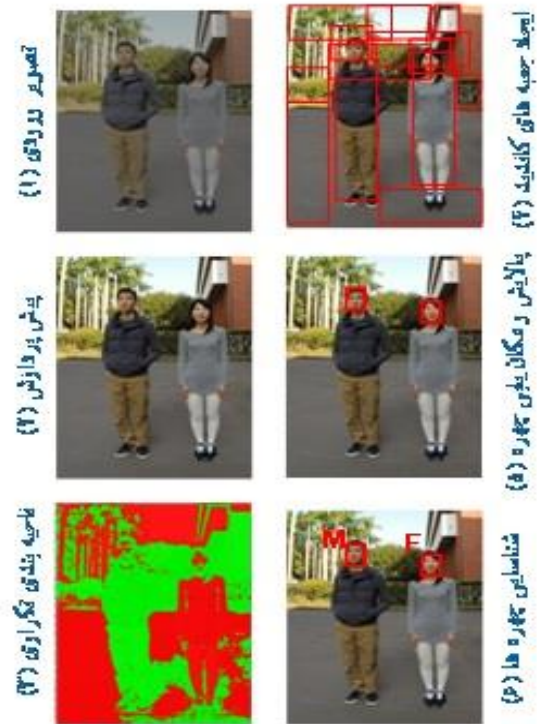
که در آن، اندازه $size(im)$ اندازه تصویر بر حسب پیکسل است.

سازگاری شکل: سازگاری شکل، میزان قرارگیری دو ناحیه r_i و r_j درون هم را اندازه‌گیری می‌کند. اگر r_i در r_j قرار می‌گیرد، باید آن‌ها را ادغام کرد تا شکاف‌ها پر شوند. بعلاوه، اگر یکدیگر را لمس نمی‌کنند، نباید ادغام شوند. سازگاری شکل به‌صورت رابطه (۴) تعریف می‌شود:

$$S_{shape}(r_i, r_j) = 1 - \frac{size(BB_{ij}) - size(r_i) - size(r_j)}{size(r_{im})} \quad (4)$$

که در آن، $size(BB_{ij})$ اندازه یک کادر محدود در اطراف r_i و r_j است.

معیار شباهت کلی: برای اینکه بتوان براساس رنگ، بافت و اندازه ناحیه‌بندی تکراری را برای تشخیص جعبه‌ها انجام داد، یک معیار



شکل ۱. مراحل روش پیشنهادی

۳-۱. ناحیه‌بندی با جستجوی انتخابی

در این بخش، ناحیه‌بندی تصویر با الگوریتم ناحیه‌بندی انتخابی انجام می‌شود. ناحیه‌بندی انتخابی یک الگوریتم پیشنهاد ناحیه است که در تشخیص اشیاء استفاده می‌شود. این الگوریتم براساس محاسبات سلسله مراتبی، نواحی مشابه را با استفاده از رنگ، بافت، اندازه و شکل پیدا می‌کند. هدف این روش تولید جعبه‌ها^۱ برای اشیایی است که شبیه‌سازی هدف هستند که مبتنی بر ناحیه‌بندی چندگانه تصویر عمل می‌کند؛ بنابراین، در این روش ترکیب نواحی کوچک براساس چهار معیار شباهت، رنگ، بافت، اندازه و شکل انجام می‌شود. در واقع، این روش براساس ناحیه‌بندی تصویر جعبه‌های زیادی پیدا می‌کند که با فیلتر کردن این جعبه‌ها نواحی هدف پیدا می‌شوند. رابطه مربوط به این معیارها در ادامه تعریف شده است.

شباهت رنگی: برای ترکیب قطعه‌ها براساس معیار شباهت، یک هیستوگرام^۲ رنگی ۲۵ تایی برای هر کانال تصویر محاسبه می‌شود و هیستوگرام همه کانال‌ها برای پیدا کردن توصیف‌گر^۳ رنگ به یک توصیف‌گر رنگ ۲۵×۳ = ۷۵ متصل می‌شوند. شباهت رنگ دو ناحیه r_i و r_j بر اساس تقاطع هیستوگرام با رابطه (۱) محاسبه می‌شود.

¹ Boxes

² Histogram

³ Descriptor

⁴ Bin

در شبکه پیشنهادی از ماژول Inception براساس شکل (۲) استفاده می‌شود. درواقع، فیلترهای کانولوشن با اندازه های ۳×۳ و ۱×۱ روی لایه قبلی اعمال شده و نتایج به‌صورت موازی با هم ترکیب می‌شوند. این ویژگی از شبکه (خاصیت GoogleNet)، باعث می‌شود تا حجم محاسبات کمتر شود و شبکه سریع‌تر آموزش داده شود. بعلاوه، عمیق‌تر کردن شبکه با اضافه کردن ماژول Residual برای بهینه‌سازی بهتر در این معماری انجام می‌شود تا از پدیده صفر شدن گرادیان‌ها جلوگیری شود. لذا، برای اینکه بهینه‌سازی شبکه سریع‌تر و بهتر انجام شود، از ابتدای بلوک به انتهای بلوک یک اتصال مستقیم انجام می‌شود. خروجی شبکه در لایه آخر یک طبقه‌بندی دو کلاسه را برای تشخیص جعبه هدف و غیر هدف انجام می‌دهد.

در ماجول Inception پیشنهادی از maxpooling استفاده نمی‌شود، زیرا maxpooling حداکثر مقادیر پیکسل را انتخاب می‌کند و سایر مقادیر نقشه‌های ویژگی را حذف می‌کند. به‌علاوه، اینکار باعث کاهش تعداد پارامترها برای جلوگیری از مشکل بیش‌برازش در طول آموزش شبکه می‌شود و پیچیدگی محاسباتی شبکه نیز کاهش پیدا می‌کند. همچنین، شبکه به‌طور مؤثر ویژگی‌های سطح بالا را استخراج می‌کند و توانایی شبکه برای یادگیری اطلاعات دقیق افزایش می‌یابد.

ابعاد ورودی لایه‌ها نشان می‌دهد که معماری پیشنهادی پارامترهای آموزشی کم و سرعت آموزش بالایی دارد. بنابراین، شبکه عمیق پیشنهادی با استفاده از مجموعه داده چهره آموزش داده می‌شود تا جعبه‌های هدف شناسایی شوند.

در جدول (۱) هر لایه ممکن است دارا یا فاقد پارامتر باشد (به عنوان مثال لایه‌های کانولوشن دارای پارامتر و لایه‌های فعال‌ساز و کاهش اندازه فاقد پارامتر هستند). همچنین، هر لایه ممکن است دارا یا فاقد ابرپارامتر اضافی نیز باشد. شبکه کانولوشن مقادیر پیکسل‌های خام تصویر اصلی را لایه به لایه به امتیاز دسته‌ها در انتهای شبکه تبدیل می‌کند. بنابراین، بعضی از لایه‌ها پارامتر دارند و بعضی دیگر فاقد پارامتر هستند. به‌طور خاص لایه‌های کاهش اندازه و کانولوشن لایه‌هایی هستند که تبدیلاتی را انجام می‌دهند که نه تنها تابعی از فعال‌سازی‌ها در توده ورودی‌اند، بلکه تابع پارامترهایی نظیر وزن و بایاس نورون‌ها نیز می‌باشند. از طرف دیگر لایه‌های فعال‌ساز و کاهش اندازه تنها یک تابع ثابت را پیاده‌سازی می‌کنند. پارامترهای موجود در لایه‌های کانولوشن توسط روش گرادیان نزولی آموزش داده می‌شوند تا امتیازات دسته‌هایی که شبکه عمیق محاسبه می‌کند با برچسب‌های هر تصویر در مجموعه آموزشی سازگار باشد. چون که مقادیر مربوط به این پارامترها به‌صورت خودکار تنظیم می‌شود (عملیات یادگیری).

شباهت کلی با ترکیب وزنی سایر معیارها ایجاد می‌شود؛ بنابراین، شباهت نهایی بین دو ناحیه به‌صورت ترکیب خطی از چهار شباهت فوق به‌صورت رابطه (۵) محاسبه می‌شود.

$$S(r_i, r_j) = z_1 S_{color}(r_i, r_j) + z_2 S_{texture}(r_i, r_j) + z_3 S_{size}(r_i, r_j) + z_4 S_{shape}(r_i, r_j) \quad (5)$$

که در آن، r_i, r_j دو ناحیه یا بخش در تصویر هستند و $z_i \in \{0,1\}$ نشان می‌دهد که آیا معیار تشابه استفاده می‌شود یا خیر. انتخاب ضرایب Z براساس آزمون و خطا (به‌صورت تجربی) تعیین می‌شود و مقدار Z برای هر معیار شباهت ۰ یا ۱ است. اگر معیار شباهت استفاده شود، $Z=1$ و در غیراینصورت $Z=0$

۲-۳. دسته‌بندی باکس‌ها و مکان‌یابی چهره

در این مرحله از روش پیشنهادی، جعبه‌های ایجادشده توسط الگوریتم ناحیه‌بندی انتخابی، پالایش می‌شوند. لذا، در روش پیشنهادی یک شبکه عصبی عمیق مبتنی بر GoogleNet و ResNet برای پالایش جعبه‌ها پیشنهاد می‌شود تا جعبه‌های هدف و غیر هدف به شبکه عصبی آموزش داده شود؛ بنابراین، یک طبقه‌بندی دو کلاسه با یادگیری عمیق با شبکه عمیق پیشنهادی انجام می‌شود تا جعبه‌های هدف شناسایی شوند. شبکه عمیق پیشنهادی در جدول (۱) نشان داده شده است.

جدول ۱. شبکه عمیق پیشنهادی برای دسته‌بندی جعبه‌ها

نوع	گام/اندازه	ابعاد ورودی
کانولوشن	۳×۳/۲	۲۲۹×۲۲۹×۳
کانولوشن	۳×۳/۱	۱۴۹×۱۴۹×۳۲
کانولوشن	۳×۳/۱	۱۴۷×۱۴۷×۳۲
کاهش اندازه	۳×۳/۲	۱۴۷×۱۴۷×۶۴
کانولوشن	۳×۳/۱	۷۳×۷۳×۶۴
کانولوشن	۳×۳/۲	۷۱×۷۱×۸۰
کانولوشن	۳×۳/۱	۳۵×۳۵×۱۹۲
Inception×۳	طبق شکل ۲	۳۵×۳۵×۲۸۸
Inception×۵	طبق شکل ۲	۱۷×۱۷×۷۶۸
Inception×۲	طبق شکل ۲	۸×۸×۱۲۸۰
کاهش اندازه	۸×۸	۸×۸×۱۲۸۰
لایه خطی	logits	۱×۱×۱۲۸۰
لایه سافت‌مکس	classifier	۱×۱×۲

در معماری پیشنهادی، لایه خطی^۱، کانولوشن^۲ و کاهش اندازه^۳ با اندازه^۴ و طول گام^۵ مشخص استفاده می‌شوند. همچنین،

^۴ Size

^۵ Stride

^۱ Linear

^۲ Convolution

^۳ Pooling

استخراج می‌شود. پس از آن، با لایه نرمال‌سازی، خروجی‌های لایه کانولوشن، نرمال می‌شوند که باعث کاهش بیش‌برازش^۱ شبکه و کاهش گرادینان می‌شود. در واقع، لایه نرمال‌سازی شبکه باعث جلوگیری از تغییر توزیع داده در هر لایه می‌شود [۳۷]. بعلاوه، در هر مرحله، بعد از اعمال تابع فعال‌ساز، عملیات کاهش اندازه تصویر نرمال‌شده انجام می‌شود تا شبکه محاسبات کمتری انجام داده و نسبت به تغییر مقیاس مقاوم باشد. در نهایت، پس از چندین لایه کانولوشن، نرمال‌سازی و کاهش اندازه، ویژگی‌های مناسب‌استخراج شده و توسط لایه تمام‌متصل^۲ و تابع softmax به بردار ۱۱ تایی تبدیل می‌شود.

برای افزایش انطباق‌پذیری روش پیشنهادی، راهکار پس‌انتشارخطا^۳ برای تنظیم پارامترها به صورت معکوس استفاده می‌شود. لذا، پارامترهای شبکه با استفاده از تابع اتلاف آنتروپی متقابل^۴ و بهینه‌ساز آدام^۵ به‌روزرسانی می‌شود [۹]؛ بنابراین، در روش پیشنهادی برای مجموعه داده با اندازه T ، به صورت $e = [(I(1), q(1)), \dots, (I(T), q(T))]$ تابع اتلاف به صورت رابطه (۷) محاسبه می‌شود.

$$\Gamma(\theta) = -\frac{1}{T} \sum_{i=1}^T [q^i \log(O^i) + (1-q^i) \log(1-O^i)] \quad (7)$$

که در آن، O^i خروجی واقعی مربوط به تصویر $I(i)$ است و $q(i)$ نشان‌دهنده برچسب تصویر i است، گرادینان پس انتشار پارامترهای کانولوشنی b و w با رابطه‌های (۸) و (۹) محاسبه می‌شوند:

$$\frac{\partial}{\partial w^i} \Gamma(\theta) = \frac{1}{T} \sum_{i=1}^T [x^i (O^i - q^i)] \quad (8)$$

$$\frac{\partial}{\partial b^i} \Gamma(\theta) = \frac{1}{T} \sum_{i=1}^T [x^i (O^i - q^i)] \quad (9)$$

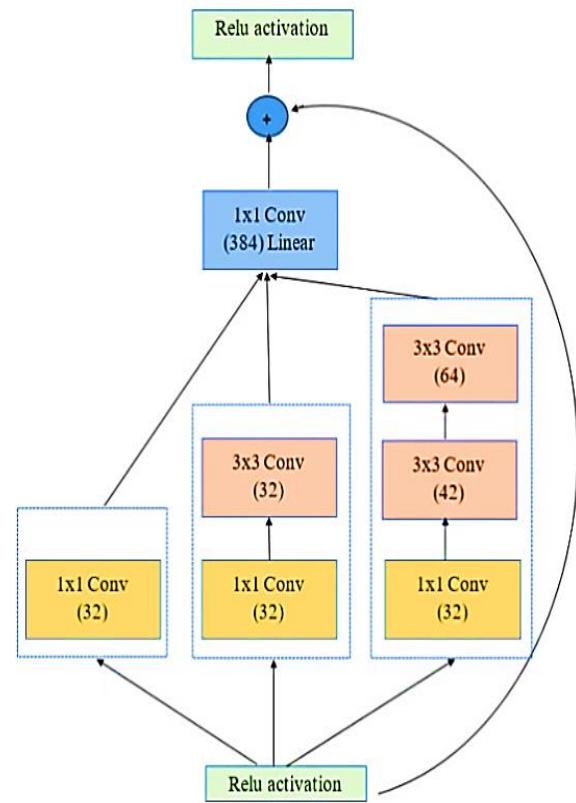
معادلات پارامترهای قابل یادگیری w^i و b^i لایه l به صورت زیر به‌روزرسانی می‌شوند:

$$w_i^l = w_i^l - \eta \frac{\partial E}{\partial w_i^l} \quad (10)$$

$$b_i^l = b_i^l - \eta \frac{\partial E}{\partial b_i^l} \quad (11)$$

که در آن، η نشان‌دهنده نرخ یادگیری^۶ است و E نشان‌دهنده خطای داده‌های آموزشی برای نمونه‌های فعلی است.

لذا، شبکه عصبی گام‌به‌گام با یادگیری این پارامترها قادر به انجام وظیفه محول شده است.



شکل ۲. ماژول Inception

۳-۳. شناسایی چهره

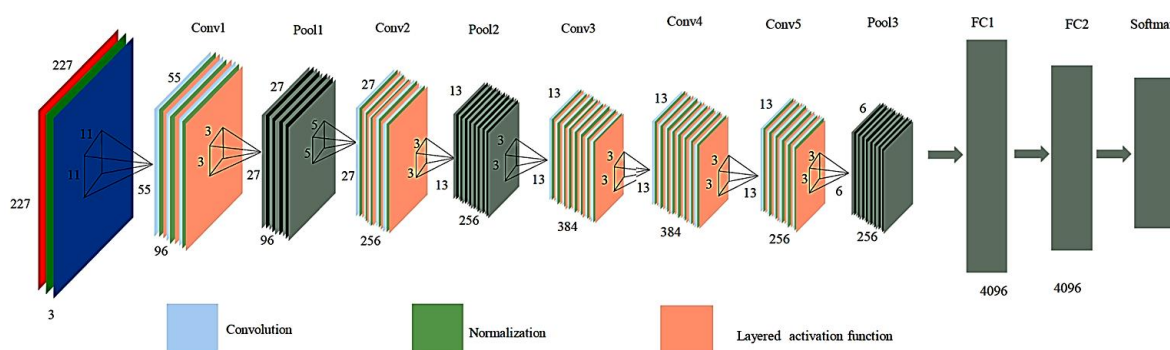
در این بخش، پس از مکان‌یابی چهره‌ها، شناسایی چهره با شبکه عمیق پیشنهادی انجام می‌شود. چارچوب شبکه پیشنهادی برای شناسایی چهره در شکل (۳) نشان داده شده است. در شبکه پیشنهادی، ابتدا استخراج ویژگی با استفاده از مجموعه داده‌های بزرگ انجام می‌شود و شبکه پیش‌آموز بعد از حذف لایه softmax به‌عنوان یک استخراج‌کننده ویژگی برای استخراج ویژگی‌های چهره استفاده می‌شود. برای پیش‌آموزش شبکه پیشنهادی با تصاویر بزرگ، از مجموعه داده‌ای [۳۶] استفاده می‌شود. خروجی ۴۰۹۶ بعدی آخرین لایه تمام متصل به‌عنوان بردار ویژگی برای تصویر مکان‌یابی شده، استفاده می‌شود و با رابطه (۶) تعیین می‌شود:

$$O^{l+1} = \varphi(W^{l+1} O^l + b^{l+1}) \quad (6)$$

که در آن، O^l, O^{l+1} به ترتیب بردار خروجی هر دو لایه l و $(l+1)$ است w وزنی از ضرایب خطی، b بردار بایاس و φ تابع غیرخطی (*) است.

تصاویر مکان‌یابی شده ابتدا به اندازه ۲۲۴×۲۲۴ تغییر می‌کنند و با استفاده از فیلترهای کانولوشن، ویژگی‌های تصویر

¹ Over-fitting
² Fully Connected
³ Back propagation
⁴ Cross entropy loss
⁵ ADAM
⁶ Learning Rate



شکل ۳. شبکه شناسایی چهره

دنیای واقعی بهتر نشان می‌دهد. همچنین این مجموعه داده‌ای حاوی تصاویر گرفته‌شده توسط پهپاد و برچسب آن‌ها است. مشخصات آماری این مجموعه داده در جدول (۳) نشان داده شده است. ابتدا، مجموعه داده DroneFace برای آموزش و ارزیابی مدل تقسیم‌بندی می‌شود. در آزمایش‌ها ۲۵٪ از داده‌ها برای اعتبارسنجی و ۷۵٪ برای آموزش مدل استفاده می‌شود.

جدول ۲. شبه کد روش پیشنهادی

Algorithm 1. Pseudo code of our method

Inputs: I : Images, B = boxes L : labels

Outputs: F = faces, R = recognized faces

Begin

۱: Initialize parameters

Face Localization (I, B):

۲: For each image in i

۳: For each bounding box in b

۴: $P = \text{Forward}(i, b)$ o Localization Network

۵: Compute Loss (P, GT) using Eq. (۷)

۶: Loss. Backward()

۷: Optimizer. Step()

۸: Return F

Face Recognition (F, L):

۹: For each face in f

۱۰: For each label in l

۱۱: $P = \text{Forward}(f, l)$ to Recognition Network

۱۲: Compute Loss (P, GT) using Eq. (۷)

۱۳: Loss. Backward()

۱۴: Optimizer. Step()

۱۵: Return R

End

جدول ۳. مشخصات مجموعه داده استفاده شده

ویژگی‌ها	مقادیر
نوع افراد	۴ زن و ۷ مرد
تعداد تصاویر	۱۳۶۴
ابعاد تصاویر (پیکسل)	۲۷۶۰×۲۶۸۰
ابعاد تصاویر چهره (پیکسل)	۳۸۴×۲۳ و ۳۸۴×۳۱
ارتفاع تصاویر (متر)	۱.۵، ۳، ۴، ۵
فاصله دوربین تا چهره (متر)	۲ تا ۱۷ متر

۳-۴. طراحی الگوریتم

در این بخش، الگوریتم کلی روش پیشنهادی معرفی می‌شود و در جدول (۲) شبه‌کد روش پیشنهادی نشان داده شده است. ورودی‌های الگوریتم شامل تصاویر ورودی I ، باکس‌های چهره B مجموعه برچسب‌ها L است. در خط ۱ مقادیر اولیه پارامترها به صورت تجربی تعیین می‌شوند. پس از آن، تابع مکان‌یابی چهره اجرا می‌شود و بدنه تابع در خطوط ۲-۸ فرایند آموزش مدل را با الگوریتم پیشنهادی انجام می‌دهد و وزن‌های شبکه با مقادیر جدید در هر تکرار براساس خطا و بهینه‌ساز به‌روز می‌شوند. این مراحل ادامه می‌یابد تا تعداد تکرارها تمام شود یا اینکه خطای مطلق از یک مقدار آستانه کمتر شود. پس از پایان الگوریتم، تصاویر چهره مکان‌یابی می‌شوند و به‌عنوان خروجی برگشت داده می‌شوند. پس از استخراج چهره‌ها، تابع شناسایی چهره تعریف می‌شود و بدنه تابع در خطوط ۹-۱۵ فرایند آموزش مدل را بر اساس الگوریتم پیشنهادی انجام می‌دهد و وزن‌های شبکه با مقادیر جدید در هر تکرار براساس خطا و بهینه‌ساز به‌روز می‌شوند. پس از پایان الگوریتم، چهره‌های شناسایی‌شده به‌عنوان خروجی برگشت داده می‌شوند.

۴. آزمایش‌های تجربی

در این بخش، روش پیشنهادی بر اساس معیارهای ارزیابی مختلف با سایر روش‌های جدید در این حوزه مقایسه می‌شود. ابتدا مجموعه داده همراه مشخصات آماری آن معرفی می‌شود، سپس معیارهای ارزیابی پرکاربرد در پیش‌بینی و شناسایی چهره ارائه می‌شوند. نتایج عملی روش پیشنهادی در مقایسه با سایر روش‌ها، مقایسه و بررسی می‌شوند.

۴-۱. مجموعه‌های داده‌ای

در این مقاله آزمایش‌های تجربی روی مجموعه داده مشهور DroneFace انجام شده است [۱۲]. این مجموعه داده به دلیل ویژگی‌های مناسب، در اغلب کارهای تحقیقاتی استفاده می‌شود. لذا، ارزیابی روش‌ها روی این داده عملکرد روش را در کاربردهای

۲-۴. معیارهای ارزیابی

برای ارزیابی سامانه‌های تشخیص و شناسایی معیارهای مختلفی وجود دارد. در این پژوهش، دو معیار شناخته‌شده برای ارزیابی دقت پیش‌بینی استفاده می‌شود که عبارت‌اند از نرخ تشخیص چهره^۱ و نرخ شناسایی چهره^۲ که به‌صورت زیر محاسبه می‌شوند.

$$DR = RR = \frac{TP}{TP + FN} \quad (۱۲)$$

که در آن، DR نرخ تشخیص چهره‌هاست و با تقسیم تعداد تصاویر چهره درست مکان‌یابی شده بر تعداد کل تصاویر چهره محاسبه می‌شود. همچنین، RR نرخ شناسایی چهره‌هاست که تعداد تصاویر درست شناسایی شده تقسیم بر تعداد کل تصاویر چهره است. TP نشان‌دهنده چهره‌هایی است که به‌درستی مکان‌یابی یا شناسایی شده‌اند و FN چهره‌هایی هستند که به غلط مکان‌یابی یا شناسایی شده‌اند. برای ارزیابی دقیق و مقایسه روش پیشنهادی با سایر روش‌ها از کتابخانه قدرتمند پایتورچ^۳ استفاده می‌شود. برای انجام آزمایش‌ها از سرور گوگل‌کولب^۴ با حافظه ۱۶ گیگابایتی استفاده شده است.

۳-۴. تنظیم پارامترها

روش پیشنهادی دارای چندین پارامتر است که برای رسیدن به نتیجه مطلوب باید به درستی تنظیم شوند. پارامتر Z در رابطه (۵) یک پارامتر وزنی برای تعیین میزان شباهت هر معیار است و مقدار Z برای هر معیار شباهت ۰ یا ۱ است. پارامتر η نرخ یادگیری یا طول گام‌های مدل برای بهینه‌سازی تابع خطا است که ۰/۰۵ تنظیم شد. همچنین، حداکثر تعداد تکرارهای برای همگرایی مدل ۱۰۰ تکرار تنظیم شد.

۴-۴. نتایج عملی برای مکان‌یابی چهره

در این بخش، نتایج مکان‌یابی چهره بر اساس معیار DR ، با روش‌های $Rekognition$ [۱۲]، $Daryahavard$ [۱۳]، $Face++$ [۱۲]، $AdamDeep$ [۵] و $Liwang$ [۱۰] مقایسه می‌شود. اعداد ذکر در جدول‌های (۴) تا (۹) دقت روش‌ها را نشان می‌دهند که برحسب درصد گزارش شده‌اند. نتایج هر کدام از روش‌ها بر اساس میزان فاصله و ارتفاع از چهره‌ها برحسب متر نشان داده شده است. همان‌طور که از نتایج مشخص است، روش پیشنهادی در اغلب موارد عملکرد بهتری نسبت به سایر روش‌ها برای ارتفاع و فاصله از چهره دارد. درواقع، چون روش پیشنهادی از ناحیه‌بندی تکراری همراه ماژول Inception برای

مکان‌یابی چهره استفاده می‌کند، لذا نسبت به تغییرات در تصویر، نویز و تغییر اندازه مقاوم است؛ بنابراین، روش پیشنهادی عملکرد بهتری نسبت به روش‌های دیگر در این حوزه دارد و معیار DR بهتری را تولید می‌کند.

جدول ۴. نتایج مکان‌یابی روش Daryanavard

۱۶	۵/۱۴	۱۳	۱۲	۱۰	۵/۸	۷	۵/۵	۴	۵/۲	
۰	۰	۱۵	۳۱	۲۵	۶۶	۷۷	۹۶	۷۴	۸۰	۵
۰	۰	۶۱	۵۶	۹۱	۸۲	۹۳	۹۰	۸۸	۸۶	۴
۰	۵	۴۸	۷۱	۷۵	۸۱	۸۷	۹۰	۷۵	۹۳	۳
۰	۱۵	۳۵	۳۴	۶۰	۶۶	۷۹	۸۰	۸۸	۹۸	۱/۵

جدول ۵. نتایج مکان‌یابی روش AdamDeep

۱۶	۵/۱۴	۱۳	۱۲	۱۰	۵/۸	۷	۵/۵	۴	۵/۲	
۰	۷	۲۴	۳۸	۳۴	۹۷	۸۳	۹۹	۸۴	۴۵	۵
۰	۱۴	۵۰	۶۷	۱۰۰	۸۹	۱۰۰	۹۷	۹۲	۷۲	۴
۰	۱۶	۷۰	۸۷	۸۶	۱۰۰	۱۰۰	۹۴	۸۲	۹۲	۳
۵	۵۷	۸۲	۷۷	۸۰	۹۷	۱۰۰	۱۰۰	۱۰۰	۹۷	۵/۱

جدول ۶. نتایج مکان‌یابی روش Face++

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۰	۳	۱۲	۴۸	۷۰	۹۴	۸۵	۸۵	۷۰	۲۴	۵
۲	۹	۳۹	۸۲	۱۰۰	۱۰۰	۱۰۰	۹۷	۹۱	۶۱	۴
۰	۶	۴۵	۹۱	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۹۷	۳
۰	۱۵	۶۷	۹۴	۹۷	۹۷	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱/۵

جدول ۷. نتایج مکان‌یابی روش Rekognition

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۲	۱۵	۳۳	۶۴	۷۶	۹۷	۱۰۰	۱۰۰	۹۱	۳۰	۵
۵	۲۱	۵۵	۸۸	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۸۵	۴
۲	۲۴	۷۰	۹۱	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۳
۱۱	۵۵	۸۲	۹۴	۹۷	۹۷	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱/۵

جدول ۸. نتایج مکان‌یابی روش Liwang

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۰	۲	۱۱	۲۷	۳۲	۵۷	۸۱	۶۳	۸۴	۷۷	۵
۰	۰	۵۲	۴۹	۷۳	۸۰	۹۱	۸۹	۷۴	۸۵	۴
۲	۸	۳۲	۵۸	۵۶	۶۳	۷۵	۸۶	۶۲	۷۵	۳
۱۰	۱۵	۲۰	۳۵	۵۵	۷۷	۶۳	۷۹	۸۹	۸۸	۱/۵

جدول ۹. نتایج مکان‌یابی روش پیشنهادی

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۸	۲۰	۴۲	۵۶	۸۰	۹۳	۹۵	۹۵	۷۹	۶۴	۵
۲۰	۲۸	۶۱	۸۳	۹۵	۹۵	۹۸	۹۶	۹۷	۹۵	۴
۸	۳۵	۸۰	۹۴	۹۷	۹۹	۹۸	۹۵	۹۷	۹۹	۳
۱۵	۶۱	۸۴	۹۲	۱۰۰	۹۵	۹۳	۹۷	۱۰۰	۱۰۰	۱/۵

¹ Detection rate

² Recognition rate

³ Pytorch

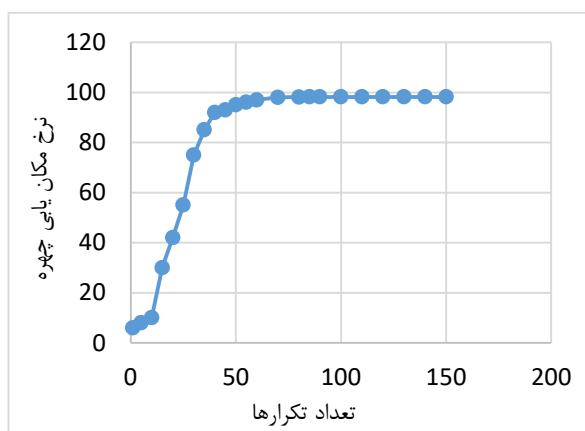
⁴ Google colab

جدول ۱۵. نتایج شناسایی چهره روش پیشنهادی

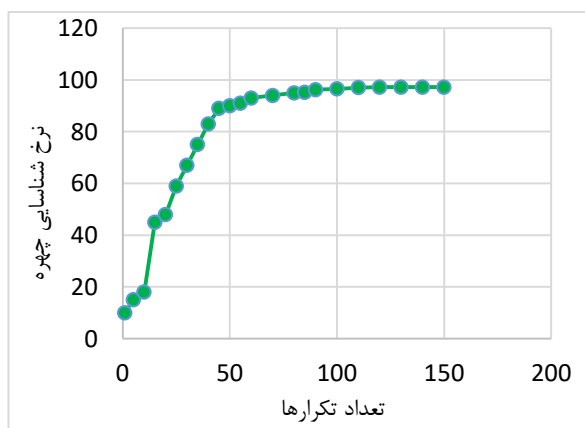
۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۵۵	۵۳	۶۱	۷۷	۸۵	۹۷	۹۶	۹۳	۸۲	۷۹	۵
۵۲	۵۷	۷۲	۸۵	۹۲	۹۶	۱۰۰	۹۶	۱۰۰	۹۳	۴
۵۹	۵۹	۷۸	۸۹	۹۵	۸۹	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۳
۵۷	۷۲	۸۰	۹۱	۱۰۰	۹۷	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱/۵

۴-۶. تحلیل همگرایی

در این بخش، همگرایی روش پیشنهادی بررسی می‌شود. نتایج حاصل در شکل (۴) و (۵) نشان داده شده است. در روش پیشنهادی استفاده از یادگیری عمیق با تابع بهینه‌ساز آدام، باعث همگرایی سریع می‌شود، در واقع، روش پیشنهادی ویژگی‌های شبکه‌های GoogleNet و ResNet را برای دسته‌بندی و بهینه‌سازی استفاده می‌کند. لذا روش پیشنهادی در حدود ۱۰۰ تکرار برای شبکه مکان‌یابی چهره و شبکه شناسایی چهره به جواب مطلوب همگرا می‌شود، لذا همگرایی الگوریتم تأثیر زیادی در سرعت یادگیری مدل پیشنهادی دارد؛ بنابراین، همگرایی مطلوب به خصوص روی داده‌های بزرگ، باعث افزایش مقیاس‌پذیری مدل می‌شود.



شکل ۴. همگرایی شبکه مکان‌یابی چهره



شکل ۵. همگرایی شبکه شناسایی چهره

۴-۵. نتایج عملی برای شناسایی چهره

در این بخش، روش پیشنهادی بر اساس معیار RR ، روی مجموعه داده DroneFace برای شناسایی چهره ارزیابی و مقایسه می‌شود. نتایج حاصل در جدول‌های (۱۰) تا (۱۵) نشان داده شده است. نتایج هر کدام از روش‌ها بر اساس میزان فاصله و ارتفاع از چهره‌ها برحسب متر نشان داده شده است. نتایج نشان می‌دهد که روش پیشنهادی در اغلب موارد عملکرد بهتری نسبت به سایر روش‌ها برای ارتفاع و فاصله از چهره دارد. در واقع، چون روش پیشنهادی از یک شبکه عمیق با ساختار مطلوب (کانولوشن - کاهش اندازه - نرمال ساز) برای شناسایی چهره استفاده می‌کند، لذا نسبت به تغییرات در تصویر، نویز و تغییر اندازه مقاوم است؛ براین، روش پیشنهادی عملکرد بهتری نسبت به روش‌های دیگر در این حوزه دارد و معیار RR بهتری را تولید می‌کند.

جدول ۱۰. نتایج شناسایی چهره روش Daryanavard

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۲۳	۴۲	۵۵	۵۹	۶۲	۷۹	۷۸	۸۰	۶۲	۳۱	۵
۳۴	۵۵	۶۱	۶۳	۶۲	۷۶	۷۲	۶۵	۷۷	۸۸	۴
۲۰	۴۸	۵۴	۵۵	۶۰	۶۳	۶۵	۷۲	۷۵	۸۴	۳
۴۰	۳۱	۴۶	۶۵	۷۴	۸۲	۷۵	۸۵	۹۰	۸۸	۱/۵

جدول ۱۱. نتایج شناسایی چهره روش AdamDeep

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۳۱	۵۰	۵۶	۶۳	۷۶	۸۰	۸۹	۹۲	۷۴	۳۴	۵
۳۴	۵۷	۷۰	۷۵	۸۴	۸۴	۸۵	۸۹	۹۵	۹۵	۴
۴۸	۵۵	۶۸	۸۴	۹۲	۵۶	۹۶	۹۹	۹۱	۹۱	۳
۴۴	۶۲	۵۶	۸۵	۹۳	۹۳	۹۹	۱۰۰	۹۴	۹۲	۱/۵

جدول ۱۲. نتایج شناسایی چهره روش Face++

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۵۰	۵۰	۴۹	۵۵	۶۵	۷۶	۷۴	۷۶	۶۷	۵۰	۵
۵۰	۴۹	۵۰	۷۳	۹۵	۹۶	۱۰۰	۹۵	۸۶	۶۳	۴
۵۰	۵۰	۵۵	۸۳	۸۵	۹۹	۱۰۰	۱۰۰	۱۰۰	۹۴	۳
۵۰	۵۱	۶۳	۸۳	۸۴	۹۴	۹۸	۱۰۰	۱۰۰	۱۰۰	۱/۵

جدول ۱۳. نتایج شناسایی چهره روش Rekognition

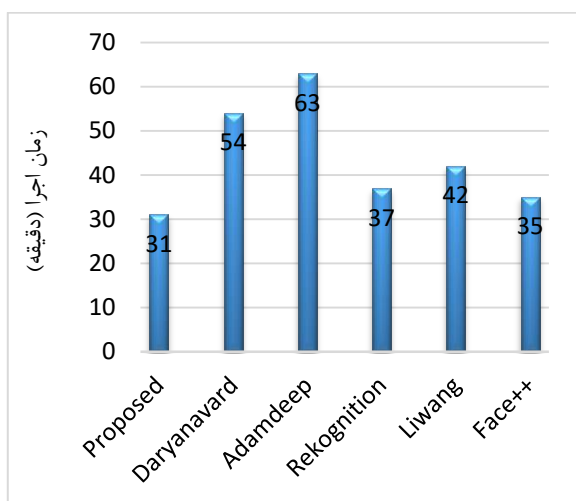
۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۵۰	۵۱	۵۵	۶۹	۷۷	۹۷	۹۹	۱۰۰	۸۵	۵۴	۵
۵۰	۵۲	۶۴	۸۹	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۸۵	۴
۵۰	۵۳	۷۰	۹۱	۹۸	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۳
۵۱	۶۰	۷۹	۹۲	۹۷	۹۷	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱/۵

جدول ۱۴. نتایج شناسایی چهره روش Liwang

۱۶	۱۴/۵	۱۳	۱۲	۱۰	۸/۵	۷	۵/۵	۴	۲/۵	
۴۸	۴۹	۵۹	۶۵	۶۸	۷۲	۷۶	۷۰	۷۶	۵۰	۵
۴۵	۵۰	۵۵	۵۷	۵۹	۶۵	۶۹	۷۹	۸۵	۹۰	۴
۵۴	۵۴	۵۹	۶۲	۷۰	۸۰	۷۷	۹۰	۸۶	۸۵	۳
۵۳	۵۵	۵۲	۷۴	۸۰	۸۳	۸۱	۸۹	۸۸	۹۵	۱/۵

۴-۷. مقیاس‌پذیری

بازگشتی و عدم استفاده از لایه پولینگ در Inception باعث کاهش بار محاسباتی مدل پیشنهادی می‌شود.



شکل ۷. پیچیدگی محاسباتی روش‌های مختلف در فاز آموزش مدل.

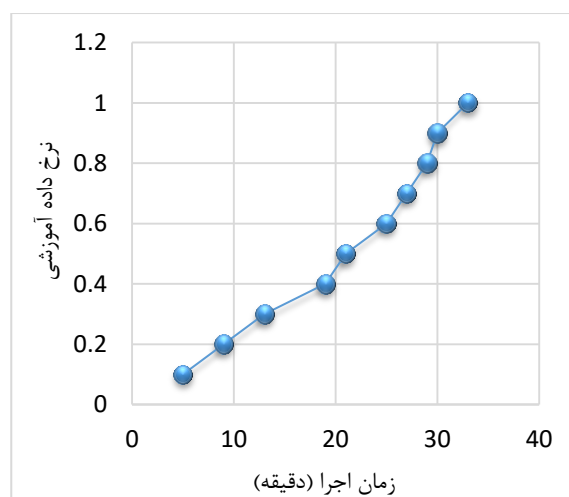
۴-۹. آزمون آماری

در این بخش، برای ارزیابی دقیق‌تر، روش پیشنهادی نسبت به سایر روش‌ها از آزمون آماری تی‌تست^{۱۳} براساس معیار MAE استفاده می‌شود. آزمون آماری مشخص می‌کند، آیا از لحاظ آماری دقت روش پیشنهادی نسبت به سایر روش‌ها در سطح معنی‌داری ۰/۰۵ تفاوت معنی‌داری دارد یا خیر؟ نتایج حاصل از آزمون آماری در جدول (۱۶) نشان داده شده است. نتایج نشان می‌دهد در اکثر موارد، مقدار p-value روش‌های دیگر در مقایسه با روش پیشنهادی کمتر از سطح معنی‌داری است و این نشان می‌دهد که فرض صفر رد می‌شود و فرض یک قابل قبول است. بنابراین، خطای روش پیشنهادی نسبت به سایر روش‌ها به صورت معناداری کمتر است. همچنین، روش پیشنهادی در اکثر موارد واریانس کمتری نسبت به بقیه روش‌ها دارد و ثبات نتایج آن برای مجموعه داده استفاده شده بیشتر است.

جدول ۱۶. مقایسه روش‌ها با آزمون آماری

مقادیر	ویژگی‌ها	روش‌ها
۰/۰۰۰۳	واریانس	Proposed
-	مقدار p	
۰/۰۰۰۴۱	واریانس	Daryanavard
۰/۰۰۰۰۶۴	مقدار p	
۰/۰۰۰۰۷۲	واریانس	Adamdeep
۰/۰۰۰۰۳۰۵	مقدار p	
۰/۰۰۰۰۳۴	واریانس	Rekognition
۰/۰۰۰۰۳۰۲	مقدار p	
۰/۰۰۰۰۱۱	واریانس	Liwang
۰/۰۰۰۰۲۱۹	مقدار p	
۰/۰۰۰۰۳۲	واریانس	Face++
۰/۰۰۰۰۴۲۱	مقدار p	

یکی از ویژگی مهم برای یک الگوریتم یادگیری ماشین کارایی و مقیاس‌پذیری آن در مقابله با رشد داده است. در واقع باید مدل با افزایش تصاویر و برچسب‌ها در زمان مناسبی پیش‌بینی را انجام دهد و مقیاس‌پذیری مناسبی را داشته باشد. همان‌طور که در شکل (۶) نشان داده شده است، روش پیشنهادی با درصدهای مختلفی در بازه ۰/۱ تا ۱ با طول گام ۰/۱ آموزش داده شده است و زمان اجرای روش پیشنهادی برای هر حالت مشخص شده است. نتایج نشان می‌دهد که روش پیشنهادی برای مجموعه داده DroneFace با افزایش اندازه داده زمان اجرای آن به صورت خطی رشد می‌کند، لذا مدل پیشنهادی دارای مقیاس‌پذیری مطلوب است. این مقیاس‌پذیری ناشی از ساختار مناسب شبکه و بهینه‌سازی مناسب وزن‌های شبکه است، لذا روش پیشنهادی به دلیل مقیاس‌پذیری بودن می‌تواند برای کاربردهای دنیای واقعی روی داده‌های بزرگ استفاده شود.



شکل ۶. مقیاس‌پذیری مدل

۴-۸. ارزیابی پیچیدگی محاسباتی

در این بخش، پیچیدگی محاسباتی روش پیشنهادی در مقایسه با روش‌های مختلف ارزیابی می‌شود. زمان اجرای روش پیشنهادی در مقابل سایر روش‌ها روی مجموعه داده DroneFace در شکل (۷) نشان داده شده است. نتایج نشان می‌دهد که در مرحله آموزش مدل، روش پیشنهادی در اغلب موارد زمان اجرای کمتری برای رسیدن به نقطه بهینه لازم دارد. پیچیدگی محاسباتی مطلوب در روش پیشنهادی به دلیل استفاده از ساختار شبکه عمیق مناسب و استفاده از الگوریتم بهینه‌سازی مناسب است. در واقع، روش پیشنهادی در طول فرایند آموزش مدل به دلیل اینکه از ساختمان داده یا آرایه کمکی استفاده نمی‌کند، زمان اجرای کمتری دارد. به علاوه، در ساختار شبکه عمیق پیشنهادی وجود لایه‌های

۵. نتیجه‌گیری

در این مقاله، یک روش تشخیص و شناسایی چهره برای فاصله و ارتفاع زیاد از چهره روی تصاویر گرفته‌شده توسط پهپادها ارائه شد. هر چند روش‌های مختلفی برای تشخیص و شناسایی چهره ارائه شده است، اما اغلب این روش‌ها حساسیت بالایی به فاصله و ارتفاع از چهره دارند. بنابراین، برای مقابله با این چالش، در روش پیشنهادی از ناحیه‌بندی تکراری و شبکه عصبی عمیق برای مکان‌یابی چهره استفاده شد. علاوه، یک شبکه عصبی عمیق با ساختار مناسب برای شناسایی چهره ارائه شد که عملیات کانولوشن، کاهش اندازه و نرمال‌سازی را برای شناسایی درست چهره‌ها استفاده می‌کند. مدل پیشنهادی در مقابل داده‌های بزرگ به‌خوبی عمل می‌کند و همگرایی مطلوبی دارد. نتایج آزمایش‌های تجربی روی مجموعه داده DroneFace نشان می‌دهد که روش پیشنهادی در اغلب مواقع عملکرد بهتری نسبت به سایر روش‌ها دارد. در معماری روش پیشنهادی از خاصیت شبکه‌های عمیق پرکاربرد به‌صورت ترکیبی استفاده می‌شود و مقایسه کمی روش پیشنهادی با روش‌های جدید از نظر پیچیدگی محاسباتی نشان می‌دهد که آموزش مدل پیشنهادی نسبت به سایر روش‌ها زمان اجرای کمتری لازم دارد. با اینکه روش پیشنهادی برای فاصله و ارتفاع زیاد از چهره‌ها با دقت بالایی تشخیص و شناسایی چهره را انجام می‌دهد. اما روش پیشنهادی و سایر روش‌های قبلی، نسبت به تغییر زاویه چهره کارایی مطلوبی ندارند. برای کارهای آتی، باید روشی ارائه شود که نسبت به تغییر زاویه چهره حساسیت کمتری داشته باشد و تشخیص و شناسایی چهره با دقت مطلوبی انجام شود. در این راستا، با طراحی یک شبکه عصبی عمیق می‌توان تغییرات زاویه چهره را براساس ویژگی‌های هندسی چهره به مدل آموزش داد که تحقیقات آتی ما در راستای برطرف نمودن این چالش است

۶. مراجع

- [5] Deeb, A.; Roy, K.; Edoh, K. D. "Drone-Based Face Recognition Using Deep Learning"; Int. Conf. Advanced Machine Learning Technologies and Applications. 2020, 197-206.
- [6] Yang, M. H.; Kriegman, D. J.; Ahuja, N. "Detecting Faces in Images: A Survey"; IEEE Trans. Pattern Anal. Mach. Intell. 2002, 24, 1, 34-58.
- [7] Yang, S.; Luo, P.; Loy, C. C.; Tang, X. "Faceness-Net: Face Detection through Deep Facial Part Responses"; IEEE Trans. Pattern Anal. Mach. Intell. 2017, 40, 8, 1845-1859.
- [8] Zhang, L.; Sun, L.; Yu, L.; Dong, X.; Chen, J.; Cai, W.; Ning, X. "Arface: Attention-Aware and Regularization for Face Recognition with Reinforcement Learning"; IEEE Trans. Biom. Behavior. Iden. 2021.
- [9] Zhu, Y.; Jiang, Y. "Optimization of Face Recognition Algorithm Based on Deep Learning Multi Feature Fusion Driven by Big Data"; Image Vision Comput. 2020, 104, 104023.
- [10] Wang, L.; Siddique, A. A. "Facial Recognition System Using LBPH Face Recognizer for Anti-Theft and Surveillance Application Based on Drone Technology"; Meas. Control. 2020, 53, 1070-1077.
- [11] Cheng, E. J.; Chou, K. P.; Rajora, S.; Jin, B. H.; Tanveer, M.; Lin, C. T.; Prasad, M. "Deep Sparse Representation Classifier for Facial Recognition and Detection System"; Pattern Recogn. Lett. 2019, 125, 71-77.
- [12] Hsu, H. J.; Chen, K. T. "Face Recognition on Drones: Issues and Limitations"; Int. J. Science Humanities Management and Technology. 2018, 4, 39 - 47.
- [13] Xun, Z.; Wang, L.; Liu, Y. "Improved Face Detection Algorithm Based on Multitask Convolutional Neural Network for Unmanned Aerial Vehicles View"; J. Electron Imaging. 2022, 31, 061804.
- [14] He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H. J. "Face Recognition Using Laplacian Faces"; IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 328-340.
- [15] Suri, S.; Sankaran, A.; Vatsa, M.; Singh, R. "Improving Face Recognition Performance Using TECS2 Dictionary"; Pattern Recogn. Lett. 2021, 145, 88-95.
- [16] Cai, D.; He, X.; Han, J.; Zhang, H. J. "Orthogonal Laplacianfaces for Face Recognition" IEEE Trans. Image Process. 2006, 15, 3608-3614.
- [17] Naseem, I.; Togneri, R.; Bennamoun, M. "Linear Regression for Face Recognition"; IEEE Trans. Pattern Anal. Mach. Intell. 2010, 32, 11, 2106-2112.
- [18] Qiu, H.; Gong, D.; Li, Z.; Liu, W.; Tao, D. "End2End Occluded Face Recognition by Masking Corrupted Features"; IEEE Trans. Pattern Anal. Mach. Intell. 2021, 0162-8828.
- [19] Fu, C.; Wu, X.; Hu, Y.; Huang, H.; He, R. "DVG-Face: Dual Variational Generation for Heterogeneous Face Recognition"; IEEE Trans. Pattern Anal. Mach. Intell. 2021, 0162-8828.
- [20] Tripathi, R. K.; Jalal, A. S. "Novel Local Feature Extraction for Age Invariant Face Recognition"; Expert Syst. Appl. 2021, 175, 114786.
- [21] Teimouri, M.; Rezaei, M. "Blind Classification of Space-Time Codes Using Machine Learning"; Adv. Defence Sci. & Technol. 2019, 10, 1-10 (In Persian).
- [1] Xu, Y.; Yu, G.; Wu, X.; Wang, Y.; Ma, Y. "An Enhanced Viola-Jones Vehicle Detection Method from Unmanned Aerial Vehicles Imagery"; IEEE Trans. Intell. Transp. Syst. 2017, pp. 1-5.
- [2] Atmaja, A. P.; Setyawan, S. B.; Setia, L. D.; Yulianto, S. V.; Winarno, B.; Lestariningsih, T. "Face Recognition System Using Micro Unmanned Aerial Vehicle"; J. Phys. Conf. Ser. 2021, 012043.
- [3] Bold, S.; Sosorbaram, B.; Lee, S. R. "Implementation of Autonomous Unmanned Aerial Vehicle With Moving-Object Detection and Face Recognition"; Info. Sci. Appl. 2016, 361-370.
- [4] Kompella, A.; Kulkarni, R. V. "A Semi-Supervised Recurrent Neural Network for Video Salient Object Detection"; Neural Comput. Appl. 2021, 33, 2065-2083.

- [32] Nair, P.; Cavallaro, A. "3-D Face Detection, Landmark Localization, and Registration Using a Point Distribution Model"; IEEE Trans. Multimedia. 2009, 11, 611-623.
- [33] Fang, W.; Wang, L.; Ren, P. "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments"; IEEE Access. 2019, 8, 1935-1944.
- [34] Almabdy, S.; Elrefaei, L. "Deep Convolutional Neural Network-Based Approaches for Face Recognition"; Appl. Sci. 2019, 9, 20, 4397.
- [35] Li, Z.; Tang, X.; Wu, X.; Liu, J.; He, R. "Progressively Refined Face Detection Through Semantics-Enriched Representation Learning"; IEEE Trans. Inf. Forensics Security 2019, 15, 1394-1406.
- [36] Cao, J.; Li, Y.; Zhang, Z. "Celeb-500k: A Large Training Dataset for Face Recognition"; IEEE Image Proc. 2018, 2406-2410.
- [37] Mishra, N. K.; Dutta, M.; Singh, S. K. "Multiscale Parallel Deep CNN (MpdCNN) Architecture for the Real Low-Resolution Face Recognition for Surveillance"; Image Vision Comput. 2021, 104290.
- [38] Ning, X.; Shaohui, X.; Fangzhe, N.; Qingliang, Z.; Chen, W.; Weiwei, C.; Weijun, L.; Yizhang, J. "Face Editing Based on Facial Recognition Features"; IEEE Trans. Cogn. Develop. 2022, 2379-8920.
- [39] Li, Y.; Lao, L.; Cui, Z.; Shan, S.; Yang, J. "Graph Jigsaw Learning for Cartoon Face Recognition"; IEEE Trans. Image Process. 2022, 1057-7149.
- [40] Li, P.; Tu, S.; Xu, L. "Deep Rival Penalized Competitive Learning for Low-Resolution Face Recognition"; Neural Networks 2022, 0893-6080.
- [22] Sabeteghlidi, A.; Latif, A.; Esmailizaini, A. "Presenting a New Photographic CAPTCHA Using Morphology"; Adv. Defence Sci. & Technol. 2017, 8, 235-241 (In Persian).
- [23] Daugman, J. "Face and Gesture Recognition: Overview"; IEEE Trans. Pattern Anal. Mach. Intell. 1997, 19, 675-676.
- [24] Hjelmas, E.; Low, B. K. "Face Detection: A survey"; Comput. Vis. Image Und. 2001, 83, 236-274.
- [25] Gao, C.; Lu, S. L. "Novel FPGA Based Haar Classifier Face Detection Algorithm Acceleration"; Int. Conf. Field Programmable Logic and Applications. 2008, 373-378.
- [26] Matai, J.; Irturk, A.; Kastner, R. "Design and Implementation of an Fpga-Based Real-Time Face Recognition System"; IEEE Int. Conf. Symposium. Field-Programmable Custom Computing Machines 2011, 97-100.
- [27] Gottumukkal, R.; Asari, V. K. "An Improved Face Recognition Technique Based on Modular PCA Approach"; Pattern Recogn. Lett. 2004, 25, 429-436.
- [28] Farlik, J.; Kratky, M.; Casar, J.; Stary, V. "Multispectral Detection of Commercial Unmanned Aerial Vehicles"; Sensors 2019, 19, 1517.
- [29] Wang, L.; Siddique, A. A. "Facial Recognition System Using LBPH Face Recognizer for Anti-Theft and Surveillance Application Based on Drone Technology"; Meas. Control. 2020, 53, 1070-1077.
- [30] Pu, Y. H.; Chiu, P. S.; Tsai, Y. S.; Liu, M. T.; Hsieh, Y. Z.; Lin, S. S. "Aerial Face Recognition and Absolute Distance Estimation Using Drone and Deep Learning"; J. Supercomput. 2021, 1-21.
- [31] Lin, S. H.; Kung, S. Y.; Lin, L. J. "Face Recognition/Detection by Probabilistic Decision-Based Neural Network"; IEEE Trans. Neural Networks 1997, 8, 114-132.

