

Vol. 14/ No. 54/Winter 2025

Research Article

Improving Super-Resolution in Face Images by Modeling Image Degradation Using Pairs of High-Quality and Low-Quality Images

Ahmad Dolatkah, Assistant Professor ^{1*} 

¹Department of information and Communication,
Amin Police University, Tehran, Iran
ict.ahmad93@gmail.com

Correspondence

Ahmad Dolatkah, assistant professor,
department of information and Communication,
Amin Police University, Tehran, Iran,
ict.ahmad93@gmail.com

Received: 17 February 2024

Revised: 14 April 2024

Accepted: 17 April 2024

Abstract

Improving image quality for identification and authentication in security and surveillance systems is of particular importance, and today, using artificial intelligence, the quality of images can be significantly improved. In this regard, the present paper, focusing on the details of face images, has improved the image failure detection model in the adversarial generator network, which led to a suitable performance in the meta-dissolving of face images. Most of the CNN networks that have been presented in recent years require a large set of images with appropriate annotations for proper performance, and they usually perform poorly in the case of degradation that have not been trained, which is addressed in this research to improve this challenge. In this work, pairs of high-quality and low-quality images are used to train the image degradation detection model; This information is then transferred to real images. The naturalness of the output images is one of the most important challenges in this field. The obtained results show that the criterion of perceptual similarity of the obtained image is equal to 38.4%, which is comparable to recent researches. As a result, using the proposed model, more natural images were produced.

Keywords: Quality of face image, Adversarial generative network, Super-resolution, Deep learning.

Highlights

- Improving super-resolution in face images by modeling image degradation using pairs of high-quality and low-quality images
- Improving image quality for identification and authentication in security and surveillance systems.
- Using SynNet and DegNet networks, the image damage detection model was improved and the image details were preserved.

Citation: A. Dolatkah, "Improving Super-Resolution in Face Images by Modeling Image Degradation Using Pairs of High-Quality and Low-Quality Images," *Journal of Southern Communication Engineering*, vol. 14, no. 54, pp. 69–82, 2025, doi:10.30495/jce.2025.1993480.1329, [in Persian].

مقاله پژوهشی

بهبود فراتفکیک پذیری در تصاویر چهره به وسیله مدلسازی خرابی تصویر با استفاده از زوج تصاویر با کیفیت و بی کیفیت

احمد دولت خواه*^۱^۱ گروه فناوری اطلاعات و ارتباطات دانشگاه علوم انتظامیامین، تهران
lct.ahmad93@gmail.com

نویسنده مسئول

^۲ احمد دولت خواه، استادیار گروه فناوری اطلاعات و ارتباطات
دانشگاه علوم انتظامی امین، تهران
lct.ahmad93@gmail.com

چکیده:

بهبود کیفیت تصویر جهت شناسایی و احراز هویت در سیستم های امنیتی و نظارتی دارای اهمیت ویژه بوده و امروزه با استفاده از هوش مصنوعی می توان کیفیت تصاویر را به صورت قابل توجهی بهبود داد. در این راستا مقاله حاضر با تمرکز بر جزئیات تصاویر چهره، مدل تشخیص خرابی تصویر در شبکه مولد تخصصی را بهبود داده است که منجر به عملکرد مناسب در فراتفکیک پذیری تصاویر چهره شد. اکثر شبکه های CNN که در سالهای اخیر ارائه شده است، برای عملکرد مناسب نیاز به مجموعه تصاویر بسیار زیاد با حاشیه نویسی مناسب دارند و معمولاً در مورد خرابی هایی که آموزش ندیده اند عملکرد نامناسبی دارند که در این مقاله به بهبود این چالش پرداخته شده است. در این کار برای آموزش مدل تشخیص خرابی تصویر، از جفت تصویرهای با کیفیت و بی کیفیت استفاده شده است؛ سپس این اطلاعات به تصاویر واقعی انتقال داده می شوند. طبیعی بودن تصاویر خروجی از مهم ترین چالش های موجود در این زمینه است. نتایج به دست آمده نشان می دهد که معیار شباهت ادراکی تصویر به دست آمده برابر با $0.38/4$ بوده که نسبت به پژوهش های اخیر قابل مقایسه می باشد. در نتیجه با استفاده از مدل پیشنهادی، تصاویر طبیعی تری تولید شد.

کلیدواژه ها: افزایش کیفیت تصویر چهره، شبکه مولد تخصصی، فراتفکیک پذیری، یادگیری عمیق

تاریخ دریافت: ۲۸ بهمن ۱۴۰۲

تاریخ بازنگری: ۲۶ فروردین ۱۴۰۳

تاریخ پذیرش: ۲۹ فروردین ۱۴۰۳

<https://doi.org/10.30495/jce.2025.1993480.1329>

۱- مقدمه

در حال حاضر، محدودیت های سخت افزاری و هزینه های زیاد در بهبود سخت افزار، مانع افزایش کیفیت تصاویر در سیستم های تصویربرداری می شود. علاوه بر این افزایش کیفیت همواره نیاز به فضای ذخیره سازی بیشتری دارد که خود می تواند مشکل ساز باشد. تاریخچه پردازش تصویر دیجیتال با توسعه کامپیوترهای دیجیتال همراه بوده است. درحقیقت پردازشگرهای دیجیتال نیازمند حافظه و توان محاسباتی زیادی هستند که پیشرفت در حوزه پردازش تصویر دیجیتال متکی بر توسعه کامپیوترهای دیجیتال و شامل ذخیره سازی داده ها، نمایش و انتقال می باشد. از چند دهه گذشته که رایانه ها امکان پیاده سازی الگوریتم های محاسباتی را فراهم ساخته اند، در راستای شبیه سازی رفتار محاسباتی مغز انسان، کارهای پژوهشی بسیاری از سوی متخصصین علوم رایانه، مهندسی و همچنین ریاضیدان ها شروع شده است، که نتایج کار آن ها، در شاخه های از علم هوش مصنوعی و در زیرشاخه هوش محاسباتی تحت عنوان شبکه های عصبی مصنوعی طبقه بندی شده است.

تصاویر با کیفیت و بی کیفیت می توانند با استفاده از معیارهای اوج نسبت سیگنال به نویز^۱ (PSNR) و شباهت ساختاری^۲ (SSIM) تفکیک شوند. تصاویر با کیفیت عالی باید دارای مقادیر PSNR بالا باشند، به این معنی که نسبت سیگنال به نویز بسیار بالا باشد. این نوع تصاویر باید دارای خطاهای کمتر و اختلافات پیکسلی کمتری با تصویر مرجع (با کیفیت) داشته باشند. تصاویر با کیفیت باید دارای SSIM نزدیک به ۱ باشند. این نشان دهنده تشابه ساختاری بالا با تصاویر با کیفیت مرجع است، به این معنی که تصاویر مقدار زیادی از ویژگی های ساختاری و اطلاعات زیادی را با تصویر با کیفیت مشترک دارند. به طور کلی، این معیارها به عنوان ابزارهای مهم برای ارزیابی کیفیت تصاویر استفاده می شوند و می توانند در فرایند بهبود کیفیت و پیشرفت تصاویر مورد استفاده قرار بگیرند. برای افزایش کیفیت تصاویر بی کیفیت، روش های مختلفی وجود دارد برخی از روش ها با استفاده از الگوریتم ها و فیلترهای پردازش تصویر مانند افزایش کنتراست و اصلاح رنگ، بهبود کیفیت تصویر را ایجاد کند. در برخی از موارد که با نویزهای پیچیده سر و کار نداریم با استفاده از فیلترهای حذف نویز، می توانید سطوح نویز را کاهش داده و کیفیت تصویر را افزایش دهید و در انتها، زمانی که با خرابی های پیچیده در تصویر مواجه می شویم باید از تکنیک های پردازش تصویر مانند فراتفکیک پذیری (super-resolution) استفاده شود که این روش می تواند با استفاده از الگوریتم های پیچیده، اطلاعات جدیدی را به تصویر اضافه کرده و وضوح آن را بهبود بخشند. به همین دلیل برای بهبود تصاویری که دچار خرابی های پیچیده هستند، باید از مفاهیم یادگیری عمیق و شبکه های عصبی استفاده شود.

استفاده از شبکه های عصبی برای بهبود کیفیت تصاویر دلایل متعددی دارد. شبکه های عصبی، به خوبی قادرند الگوها و ویژگی های پیچیده در تصاویر را تشخیص دهند. این امکان باعث می شود که بتوانند خطاها و بی کیفیتی های موجود در تصاویر را شناسایی کرده و بهبود بخشند. علاوه بر این، با توجه به حجم بالای داده های تصویری موجود، می توان شبکه های عصبی را برای بهبود کیفیت تصاویر با داده های آموزشی غنی و گسترده آموزش داد. این امر می تواند منجر به عملکرد بهتر و کیفیت بالاتر در تصاویر نهایی شود. از طرفی با پیشرفت تکنولوژی، شبکه های عصبی عمیق مانند شبکه های مولد تخصصی^۳ (GAN) امکانات بیشتری برای بهبود کیفیت تصاویر ارائه می دهند. این شبکه ها به طور خاص برای تولید تصاویر با کیفیت بالا و بهبود کیفیت تصاویر واقعی استفاده می شوند.

یادگیری عمیق از جمله مباحثی است که در سال های اخیر بسیار مورد بحث بوده و توجه بسیاری از محققان و پژوهشگران حوزه ی هوش مصنوعی و چه بسا دیگر حوزه ها را به خود معطوف نموده است. در واقع می توان سال ۱۹۹۳ میلادی را سالی دانست که اولین نمونه ی موفق شبکه ی عصبی کانولوشنی^۴ CNN و همچنین شبکه ی LeNet مطرح گردید و در سال ۱۹۹۷ نیز شبکه های عصبی بازگشتی RNN ابداع و مطرح شدند اما در سال های اخیر به دلیل گسترش و پیشرفت در مواردی که ذکر خواهد شد، شاهد رونق بحث یادگیری عمیق بوده ایم [۱].

از جمله موارد مهم در ظهور یادگیری عمیق، بوجود آمدن مجموعه داده های عظیم مانند ImageNet است که عملاً کمبود داده های آموزشی را تا حد قابل قبولی برطرف نمود و مشکلات آموزشی که در زمان عدم جود داده ی زیاد مطرح می شود، مانند بیش برآزش و دقت پایین مدل را ارتقا بخشید. از طرف دیگر پیشرفت های سخت افزاری نظیر مطرح شدن نسل های جدید GPU و پردازشگرها که عملاً فرآیند بسیار طولانی و چه بسا ناممکن محاسبات سنگین در گذشته را جامه عمل پوشاندند، به این امر کمک کردند.

شبکه عصبی کانولوشن یک چارچوب یادگیری عمیق و گسترده است که مورد استفاده قرار می گیرد. این شبکه در ابتدا، بیشتر برای عملیات تشخیص اشیاء استفاده می شد، اما در حال حاضر در زمینه های دیگری مثل ردیابی اشیاء، برآورد موقعیت، کشف و تشخیص متن، کشف برتری بصری، کشف اقدامات، برچسب زدن صحنه، فرا تفکیک پذیری و بسیاری موارد دیگر در حال آزموده شدن می باشد [۲].

ساختار یک شبکه CNN مانند شکل ۱ است که از لایه های مختلف کانولوشن تشکیل شده است و به وسیله عملیات Pooling و توابع ReLue یادگیری اتفاق می افتد. در نهایت خروجی به یک لایه کاملاً متصل وصل می شود. نتیجه ی اعلام شده با نتیجه ی

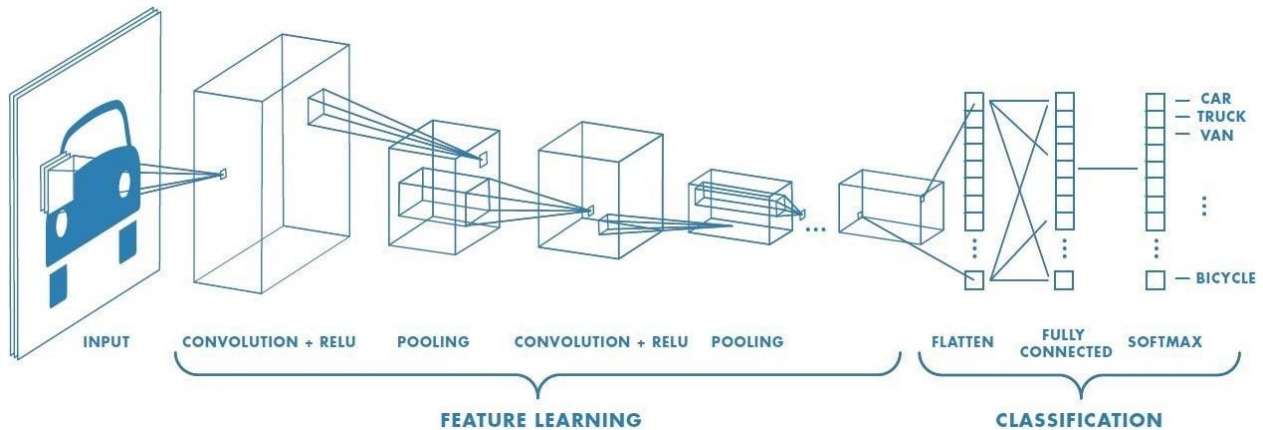
¹ Peak signal-to-noise ratio

² Structural similarity index measure

³ Generative adversarial networks

⁴ Convolutional neural network

مطلوب مقایسه گشته و میزان خطای بوجود آمده توسط تابع هزینه محاسبه می شود سپس توسط الگوریتمی مانند پسا انتشار خطا، روند اصلاح وزن ها انجام می گیرد. به هر بار انجام این مرحله یک تکرار گفته می شود. مراحل اصلاح وزن ها تا زمانی صورت می پذیرد که به بهینه ترین حالت ممکن رسیده شود. بدین منظور از روش های بهینه سازی همچون گرادیان کاهشی Adam، یا انواع دیگر استفاده می شود [۱].



شکل ۱: شماتیک کلی شبکه CNN با هدف دسته بندی تصاویر

Figure 1. General schematic of CNN network with the purpose of image classification

عیب اصلی CNN این است که گاهی اوقات آموزش آن زمان زیادی صرف می کند. با این حال، پس از آموزش، کارایی زمان تست آن ها بسیار سریع تر از سایر روش ها مانند SVM^۱ یا KNN^۲ است. یکی دیگر از عیب های مهم، نیاز به مجموعه داده های بزرگ (یعنی صدها یا هزاران عکس) و حاشیه نویسی مناسب آن هاست [۳].

شبکه های عصبی پیچشی (CNN در مدیریت داده هایی که آموزش دیده اند به خوبی عمل می کنند؛ اما در مورد آن ها که از مجموعه های آموزشی منحرف می شوند عملکرد ضعیف تری دارند [۴]. این ویژگی باعث می شود که در مواجهه با تصاویر بی کیفیت دنیای واقعی که معمولاً با تخریب پیچیده روبرو شده اند، کارآمد نباشد. با این حال، ساختن این جفت مجموعه داده های با کیفیت و بی کیفیت (LQ و HQ) در دنیای واقعی نه امکان پذیر است و نه عملی، زیرا انواع تخریب در دنیای واقعی بسیار متنوع هستند و برخی از آن ها توسط سیستم تصویربرداری اعمال نشده اند. برای کاهش مشکلات در بازیابی تصاویر LQ در دنیای واقعی، برخی از پژوهش ها تلاش می کنند تا پارامترهای تخریب را پیش بینی کنند [۹-۵]. با این حال، تخریب واقعی معمولاً با انواع مختلفی ترکیب می شود که هر یک ویژگی های ذاتی خود را از دست داده اند.

همان طور که گفته شد، CNN پرکاربردترین شبکه عصبی در حوزه پردازش تصویر می باشد. در این مقاله قصد داریم از CNN در شبکه مولد تخصصی (GAN) استفاده کنیم. همان طور که میدانیم، شبکه GAN از نوع شبکه های مولد بوده و امروزه به عنوان یکی از محبوب ترین موضوعات در حوزه یادگیری ماشین محسوب می شود. شبکه GAN از دو بخش اصلی شبکه مولد یا ژنراتور و شبکه تفکیک کننده تشکیل شده است. هدف مولد، فریب دادن تفکیک کننده است و هدف تفکیک کننده فهمیدن این است که نمونه ی تولیدی مصنوعی است یا از توزیع داده های واقعی نمونه برداری شده است. به عبارت دیگر خروجی تفکیک کننده احتمال واقعی بودن نمونه را نشان می دهد. به این صورت که احتمال بیشتر معادل این است که نمونه ی تولیدی مشابه نمونه واقعی و احتمال نزدیک به صفر به معنی نزدیک بودن به نمونه تقلبی می باشد. زمانی که احتمال خروجی تفکیک کننده به ۰.۵ می رسد، راه حل بهینه به دست آمده است. به این معنی که تفکیک کننده قادر به تمیز دادن نمونه های اصلی از تقلبی نمی باشد. هر دو شبکه ی مولد و تفکیک کننده، به صورت ضمنی توابعی هستند که توسط شبکه های عصبی عمیق پیاده سازی شده اند.

¹ Support vector machine

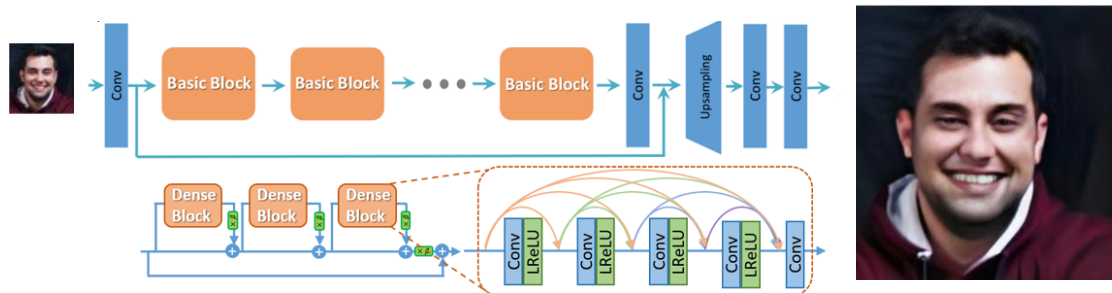
² k-nearest neighbors

یکی از مشکلات شبکه GAN موضوع بزرگنمایی و مقیاس‌پذیری تصاویر بود که با شبکه ESRGAN^۱ این موضوع حل شد. در ESRGAN با استفاده از شبکه VGG^۲ لایه‌های کانولوشن و پولینگ فراوانی تشکیل شد تا بتواند عمل سوپررزولوشن را به خوبی انجام دهد [۱۰].

در این مقاله هدف اصلی، گسترش شبکه ESRGAN به منظور بازسازی تصاویر واقعی با وضوح پایین از طریق ترکیب جفت‌های آموزشی با یک فرآیند خرابی عملی‌تر است. خرابی‌های پیچیده واقعی معمولاً ناشی از ترکیب‌های پیچیده‌ای از فرآیندهای مختلف خرابی هستند، مانند سیستم تصویربرداری دوربین‌ها، ویرایش تصویر و انتقال اینترنتی. به عبارت دیگر، خرابی‌ها با چندین فرآیند تکرار شونده مدل‌سازی می‌شوند، هر فرآیند به عنوان مدل خرابی کلاسیک است. به صورت تجربی، از یک فرآیند مدل‌سازی خرابی درجه دوم برای توازن مناسب بین سادگی و کارایی استفاده شده است.

مدل خرابی کلاسیک فقط شامل یک فرآیند ثابت است در حالی که ممکن است تصویر شامل خرابی‌های پیچیده شود. چنین فرآیند خرابی پیچیده‌ای نمی‌تواند با مدل کلاسیک اولیه مدل‌سازی شود. بنابراین، در این پژوهش یک مدل خرابی بهینه ارائه شده است. یک مدل مشتمل بر n فرآیند تکراری خرابی است، که هر فرآیند خرابی از مدل خرابی کلاسیک با همان روش اما با پارامترهای مختلف استفاده می‌کند. اخیراً، روش‌های مبتنی بر داده برای طراحی یک مدل تخریب عملی با ساخت ترکیبات پیچیده، نمونه‌برداری، نویز و فشرده‌سازی پیشنهاد شده‌اند. اگرچه این روش‌ها انواع تخریب متنوع‌تری دارند و در بیشتر موارد، تعمیم‌زبانی در مدیریت تصاویر LQ در دنیای واقعی نشان می‌دهند، اما هنوز نمی‌توانند برخی از تخریب‌های واقعی پیچیده را پوشش دهند [۱۸-۱۱].

ژنراتور ESRGAN استفاده شده در این طرح یک شبکه عمیق با چندین بلوک RRDB است که در شکل ۲ شماتیک این شبکه دیده می‌شود. همچنین معماری اصلی ESRGAN به منظور اجرای فرآیند قرا تفکیک‌پذیری با ضریب اندازه یک و دو برابر، گسترش یافته است. از آنجا که ESRGAN یک شبکه سنگین است، ابتدا از pixel-unshuffle برای کاهش اندازه فضایی و افزایش اندازه کانال قبل از تغذیه ورودی‌ها به معماری اصلی ESRGAN استفاده شده است. بنابراین، بیشتر محاسبات در یک فضای کوچک‌تر انجام می‌شود که می‌تواند مصرف حافظه GPU و منابع محاسباتی را کاهش دهد.



شکل ۲: شماتیک ژنراتور در شبکه ESRGAN [۱۰]

Figure 2. Generator schematic in ESRGAN network [10]

در این مقاله از تفکیک‌کننده U-Net با نرمال‌سازی طیفی (SN) استفاده شده است. U-Net ارزش‌های واقعی برای هر پیکسل را تشخیص می‌دهد و می‌تواند بازخورد جزئی به تولیدکننده ارائه دهد. هم‌زمان، ساختار U-Net و خرابی‌های پیچیده نیز ناپایداری آموزش را افزایش می‌دهد. در این طرح از نرمال‌سازی طیفی برای استقرار پویایی‌های آموزش استفاده شده است. تخریب واقعی تصویر با مدل پیشنهادی کشف شده است، که شامل الف- یادگیری تخریب واقعی از جفت تصاویر چهره واقعی LQ و HQ با DegNet و ب- انتقال آن به HQ است.

ساختار مقاله بدین صورت است که ابتدا به بررسی پژوهش‌های گذشته پرداخته می‌شود که در این بخش نتایج پژوهش‌های اخیر و مزایا و معایب آن‌ها بررسی می‌شود. سپس روش پیشنهادی شرح داده می‌شود. برای بیان روش پیشنهادی، ابتدا

¹ Enhanced Super-Resolution

² Visual Geometry Group

مجموعه داده‌ها و جزئیات پیاده‌سازی و سپس مدل پیشنهادی ارائه می‌شود. سپس نتایج شبیه‌سازی و مقایسه با سایر روش‌ها ارائه شده و در انتها ارزیابی و نتیجه‌گیری بیان خواهد شد.

۲- پژوهش‌های گذشته

تعدادی از تحقیقات انجام شده در زمینه افزایش کیفیت تصویر در جدول ۱ دسته‌بندی شده‌اند. همان‌طور که مشاهده می‌شود، دقت پایین در جزئیات تصویر در اکثر پژوهش‌ها یک نقطه ضعف به حساب می‌آید. مهم‌ترین چالشی که در پژوهش‌های گذشته دیده می‌شود، عملکرد نامناسب الگوریتم‌های ارائه شده برای خرابی‌های واقعی می‌باشد. این امر باعث شده است که الگوریتم‌های ارائه شده دارای معایبی از جمله محو شدگی تصویر، جابه‌جایی پیکسل‌ها، وجود نویز و مشخص نبودن مرز سوژه‌ها باشد. دلیل اصلی وجود این معایب، پروسه یادگیری خرابی تصاویر در این الگوریتم‌ها می‌باشد. در واقع تمرکز اصلی این پژوهش‌ها بر بهبود کیفیت تصاویر با خرابی‌های از پیش مشخص شده می‌باشد لذا این الگوریتم‌ها معمولاً در مواجهه با خرابی‌های واقعی دچار مشکل می‌شوند.

جدول ۱: پژوهش‌های اخیر در زمینه افزایش کیفیت تصویر

Table 1. Recent researches in the field of increasing image quality

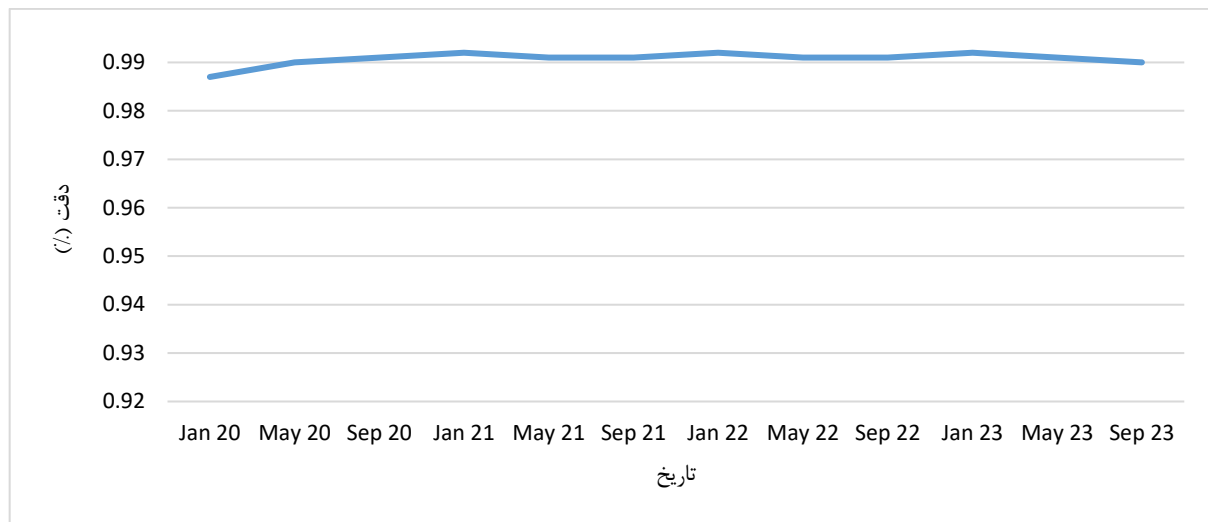
ردیف	مدل استفاده شده	SSIM	PSNR	سوژه	ویژگی تصویر خروجی	مرجع
۱	PGAN	۰/۷۲	۲۴/۹۰	چهره انسان	به هم ریختگی در پیکسل‌های اجزای تشکیل دهنده صورت	۱۹
۲	FSRGAN	۰/۴۴	۱۶/۸۰	چهره انسان	محوشدگی کل تصویر	۲۰
۳	ESRGAN	۰/۶۲	۲۲/۸۰	چهره انسان	مناسب برای زوم بر جزئیات صورت	۱۰
۴	Wang	۰/۷۷	۲۷/۵۷	چهره انسان	یکپارچه نبودن مرز بین پیکسل‌ها (روی تصاویر (سیاه‌وسفید تست شده است	۲۱
۵	Staged-CNN	۰/۶۴	۲۳	چهره انسان	زمان‌بر بودن زمان بازسازی تصویر	۲۲
۶	FSRNet	۰/۶۷	۲۳/۰۷	چهره انسان	از بین رفتن حالت چهره	۲۳
۷	SRFlow	۰/۷۱	۲۵/۲۰	چهره انسان	محوشدگی جزئی	۲۴
۸	Att-FH	۰/۷۱	۲۵/۱۰	چهره انسان	جابه‌جایی پیکسل‌ها در جزئیات تصویر	۲۵
۹	Panini-Net	-	۱۸	چهره انسان	بازسازی کاملاً بی‌دقت تصاویر	۲۶
۱۰	HiFaceGAN	۰/۷۹	۲۲/۴۰	چهره انسان	وجود نویز	۲۷
۱۱	pSp	۰/۶۸	۱۸/۹۰	چهره انسان	از بین رفتن جزئیاتی مثل عینک فرد	۲۸
۱۲	SSResNet	۰/۷۹	۲۵/۳۰	چهره انسان	محوشدگی کل تصویر	۲۹
۱۳	Cycle GAN	-	۱۶/۱۰	چهره انسان	به‌هم‌ریختگی شدید پیکسل‌ها	۳۰
۱۴	HLLH	-	۲۳/۲۲	چهره انسان	از بین رفتن پیکسل‌ها در جزئیات تصویر	۳۰
۱۵	ANR	-	۲۷/۷۰	چهره انسان	جابه‌جایی پیکسل‌ها (فقط برای تصاویر (سیاه‌وسفید تست شده است	۳۱
۱۶	RAISR	۰/۸۶	۲۴/۲۰	چهره انسان	تصویر محو می‌شود اما جزئیات حفظ می‌شود	۳۲

همان طور که در جدول ۱ دیده می‌شود، پژوهش‌های انجام شده در زمینه افزایش کیفیت تصاویر با استفاده از شبکه GAN از سال ۲۰۱۸ مورد بررسی قرار گرفته‌اند. شبکه‌های HLLH, Cycle GAN و SSResNet^۲ از جمله الگوریتم‌های قدیمی‌تر هستند که در آن‌ها خرابی‌های پیچیده لحاظ نشده‌اند و الگوریتم فقط تحت شرایط خاص، خروجی مطلوب ارائه می‌دهد [۲۹-۳۰]. همچنین برخی از این شبکه‌ها مانند ANR^۳ فقط برای تصاویر سیاه‌وسفید آموزش داده شده است [۳۱]. در نتیجه، خروجی این شبکه‌ها معمولاً با جابه‌جایی پیکسل همراه بوده و به هیچ وجه دقت مناسبی در جزئیات ندارند. در سال‌های اخیر، ESRGAN با PSNR=22/8 یکی از بهترین مدل‌های ارائه شده برای فراتفکیک پذیری تصاویر چهره (با حفظ جزئیات) می‌باشد [۱۰]. اما همانطور که بیان شد، این مدل همچنان در تشخیص خرابی‌های پیچیده دچار مشکل می‌شود. لذا در این مقاله مدلی برای تشخیص خرابی‌های پیچیده تصویر بر مبنای شبکه ESRGAN ارائه خواهد شد.

۳- روش پیشنهادی

۳-۱- داده‌ها و جزئیات پیاده‌سازی

تصاویر چهره با کیفیت پایین با استفاده از دیتاست FFHQ استخراج شده است و سپس از شبکه GPEN برای به دست آوردن نمونه‌های مصنوعی با کیفیت بالا مرتبط با آن‌ها استفاده شد [۳۳]. در شکل ۳ دقت این دیتاست دیده می‌شود.



شکل ۳: دقت دیتاست FFHQ در طول زمان

Figure 3: Accuracy of the FFHQ dataset over time

تصاویر استفاده شده برای آموزش مدل تشخیص خرابی تصویر شامل انواع مختلف از افت کیفیت هستند؛ از ناچیز تا شدید و از قدیمی تا حال. از این تصاویر، ۱۰۰۰۰ عکس برای آموزش، ۱۰۰۰ عکس برای اعتبارسنجی و ۵۰۰۰ عکس برای تست استفاده می‌شود. بهینه‌ساز Adam با پارامترهای $\beta_1 = 0/5$ و $\beta_2 = 0/999$ برای الگوریتم پیشنهادی به کار گرفته می‌شود. نرخ یادگیری اولیه به مقدار $0/00002$ تنظیم شده و هنگامی که تابع خطای MSE بر روی مجموعه اعتبارسنجی به ثباتی نزدیک می‌شود، این نرخ به $0/50$ کاهش می‌یابد.

۳-۲- مدل پیشنهادی

در این طرح، تلاش اولیه‌ای برای ایجاد خرابی در تصاویر انجام شده است که شامل یادگیری خرابی واقعی از جفت‌های تصاویر باکیفیت و بی‌کیفیت با استفاده از DegNet، و انتقال این یادگیری به تصاویر HQ جهت تولید نمونه‌های واقعی LQ آن‌ها با SynNet است. در مورد اول، به جای اینکه از یک تصویر LQ تنها برای پیش‌بینی پارامترهای خرابی آن استفاده شود، از تصاویر

¹ High-Low-Low-High

² Scale Selective Residual Network

³ Adaptive Non-local Residual Network

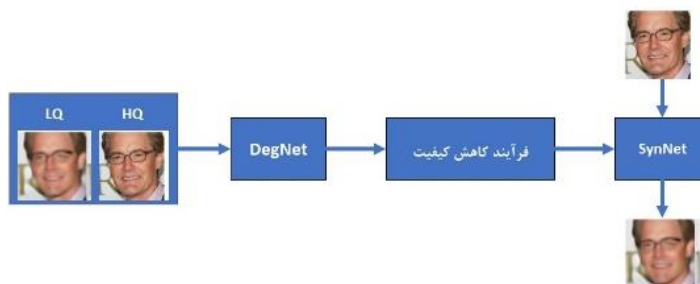
واقعی LQ و HQ مصنوعی آن به عنوان ورودی استفاده شده است تا خرابی را ایجاد کند. برای جداسازی محتوای تصویر و نوع خرابی، یک چارچوب که خرابی را از طریق چند لایه متصل به صورت دقیق پیش‌بینی می‌کند تا وزن‌های کانولوشن را تولید کند، طراحی شده است. سپس این اطلاعات به تصاویر طبیعی انتقال داده می‌شود. بنابراین، این طرح دو زیرشبکه اصلی دارد؛ یعنی DegNet برای یادگیری نمایش افت و SynNet برای سنتز تصاویر با کیفیت پایین با ورودی با کیفیت بالا و پارامتر Ω داده شده از مرحله قبل. پس از دریافت Ω ، با الهام از Style-GANs که سبک تصویر تولیدی را با یک بردار در فضای W کنترل می‌کنند، در این کار ساختار مشابهی برای نگاشت نمایش افت Ω به فضای W از طریق چندین لایه کاملاً متصل انتخاب شده است. سپس، به جای ورود نویز گسترده در StyleGAN، محتوای تصویر SynNet توسط ویژگی‌های تصاویر HQ ورودی فراهم می‌شود. در نهایت، با استفاده از تبدیل D و محتوای تصویر C ، تصویر تضعیف شده با عملیات پیچش مدوله شده (MC Block) بازسازی می‌شود که در آن استایل‌های افت به عنوان وزن‌های پیچشی برای کنترل فرآیند افت تصویر ورودی داده شده عمل می‌کنند. شکل ۵ نمای کلی این مدل را نمایش می‌دهند.

پروسه یادگیری DegNet و SynNet می‌تواند به صورت روابط ۲ و ۱ بیان شوند (که به ترتیب با F_{Deg} و F_{Syn} نمایش داده شده‌اند):

$$\Omega_f^R = F_{Deg}(I_f^R, I_f^P; \theta_{Deg}) \quad (1)$$

$$I_f^L = F_{Syn}(I_f^P, \Omega_f^R; \theta_{Syn}) \quad (2)$$

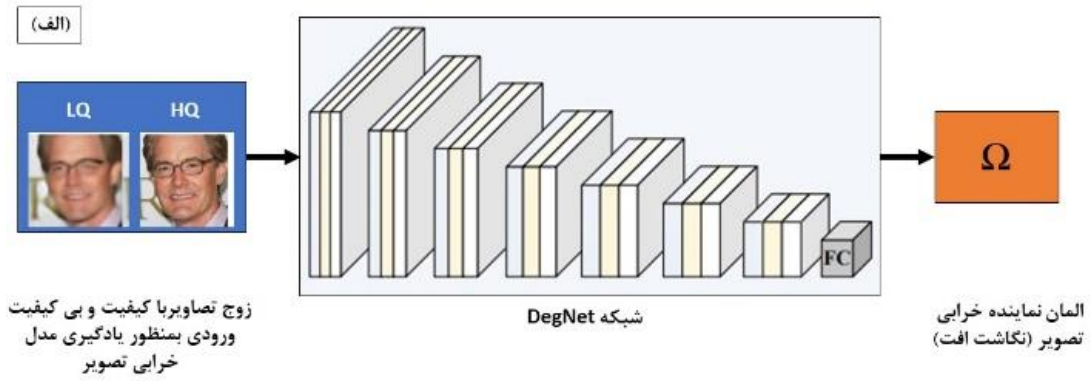
که در این رابطه θ_{Deg} و θ_{Syn} پارامترهای قابل آموزش شبکه‌های DegNet و SynNet هستند، Ω_f^R پارامتر خرابی تصویر بوده و I_f^R ، I_f^P و I_f^L به ترتیب تصاویر بی کیفیت ساخته شده، تصاویر با کیفیت ساخته شده و تصاویر بی کیفیت واقعی هستند. در شکل ۴ پروسه تولید تصویر بی کیفیت واقعی دیده می‌شود.



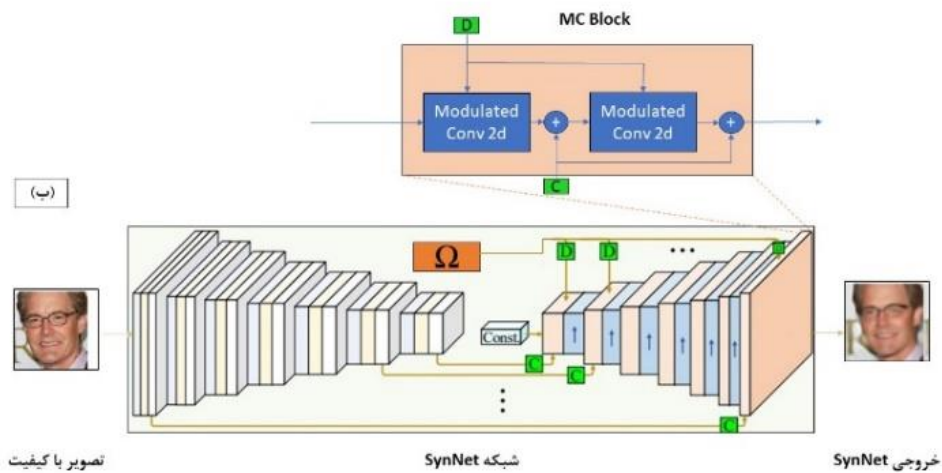
شکل ۴: پروسه تولید تصویر بی کیفیت واقعی

Figure 4. Process of real LQ image production

با تشریح مدل پیشنهادی می‌توان به نوآوری در این مقاله پی برد. امروزه، یکی از چالش‌ها در پروژه‌های افزایش کیفیت تصاویر، مدل‌سازی خرابی تصویر است. ایده اصلی این مقاله استفاده از جفت تصاویر با کیفیت و بی کیفیت است تا خرابی تصویر مدل شود. شبیه‌سازی‌ها نشان می‌دهند که این الگوریتم می‌تواند فرآیند واقعی خرابی تصویر را یاد بگیرد. در این طرح، ابتدا سعی می‌شود خرابی در تصاویر ایجاد شود. در این کار برای یادگیری خرابی از جفت تصاویر با کیفیت پایین و کیفیت بالا استفاده می‌شود. در واقع، به جای اینکه از یک تصویر LQ تنها برای پیش‌بینی پارامترهای خرابی آن استفاده شود از تصاویر واقعی LQ و HQ مصنوعی آن به عنوان ورودی استفاده می‌شود تا خرابی را ایجاد کند. برای جداسازی محتوای تصویر و نوع خرابی، یک شبکه عصبی که خرابی را از طریق چند لایه متصل به صورت دقیق پیش‌بینی می‌کند تا وزن‌های کانولوشن را تولید کند، طراحی شده است. توجه داشته شود که روش پیشنهادی ممکن است در سناریوهای بدون چهره کارایی محدودی داشته باشد چرا که مدل‌سازی برای تصاویر چهره انجام شده است.



(الف)



(ب)

شکل ۵: نمای کلی مدل پیشنهادی (الف) DegNet و (ب) SynNet
Figure 5. Overview of the proposed model (a) DegNet and (b) SynNet

۴- نتایج شبیه سازی

در شکل ۶ خروجی مدل BSRGAN، ESRGAN و مدل پیشنهادی دیده می شود. معیارهای مقایسه مورد بررسی در این مقاله عبارتند از اوج نسبت سیگنال به نویز (PSNR) که بر حسب دسی بل بیان می شود، شباهت ساختاری (SSIM)، شباهت ادراکی (LPIPS) و دقت تایید چهره در دیتاست که بر حسب درصد هستند. نتایج شبیه سازی برای مدل پیشنهادی در جدول ۲ دیده می شود.

جدول ۲: نتایج شبیه سازی مدل پیشنهادی

Table 2. Simulation results of the proposed model

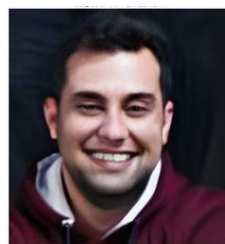
معیار مقایسه	PSNR (dB)	SSIM (%)	LPIPS (%)	دقت تایید چهره در دیتاست (%)
مدل پیشنهادی	۲۵/۴۲	۷۱/۵۰	۳۸/۴۰	۹۸/۸۰

بهبود فرانتفیک پذیري در تصاویر چهره به وسیله .../احمد دولت خواه

همانطور که در تصویر ۷ دیده می‌شود، این مدل در فرانتفیک پذیري و نمایش جزئیات تصویر چهره بسیار کارآمد می‌باشد. با توجه شکل ۸ و ۹ و توجه به مقدار LPIPS دیده می‌شود که تصاویر تولید شده با مدل پیشنهادی بسیار طبیعی‌تر از مدل ESRGAN است.



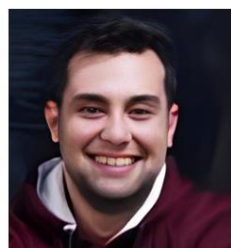
الف - ورودی



ب - ESRGAN



ج - BSRGAN



د - مدل پیشنهادی



الف - ورودی



ب - ESRGAN



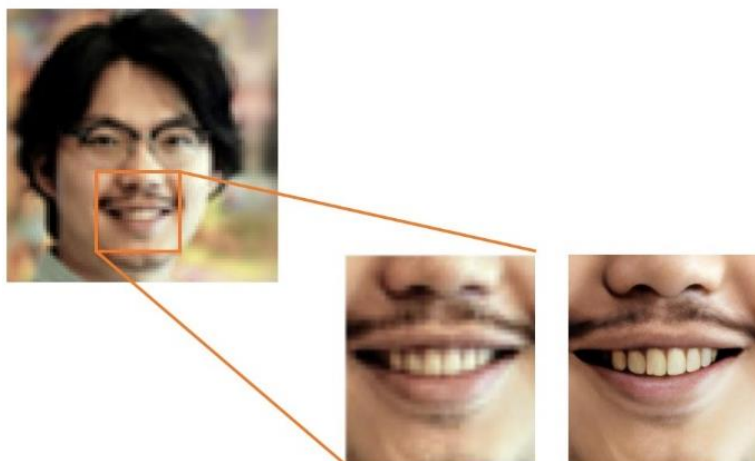
ج - BSRGAN



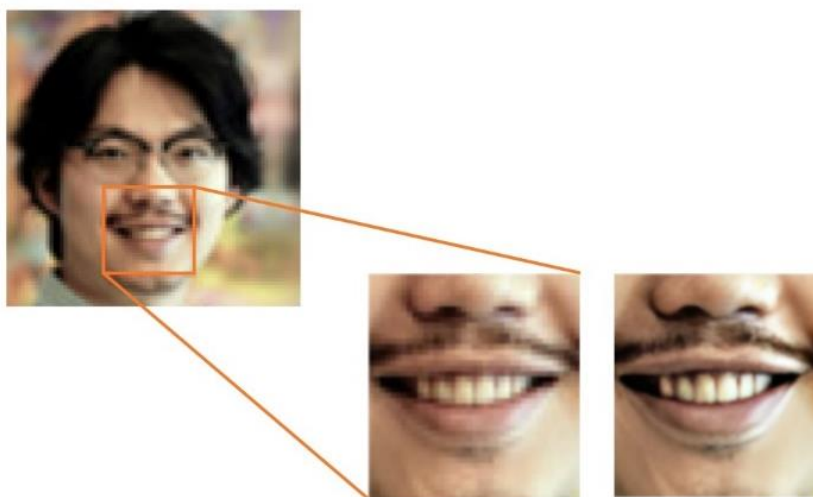
د - مدل پیشنهادی

شکل ۶: تصویر (الف) ورودی و مقایسه خروجی مدل‌های (ب) ESRGAN [۱۰]، (ج) BSRGAN [۱۵] و (د) مدل پیشنهادی

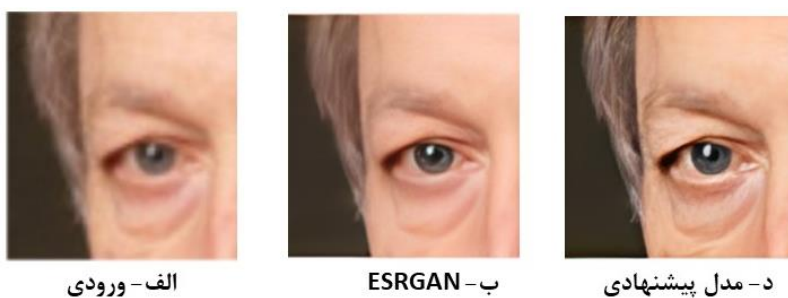
Figure 6. (a) input and output comparison of (b) ESRGAN [10], (c) BSRGAN [15] and (d) proposed model



شکل: فراتفکیک پذیری در جزئیات تصویر صورت با مدل پیشنهادی [۱۰]
 Figure 7. Super-resolution in the details of the face image with the proposed model [10]



شکل ۸: فراتفکیک پذیری در جزئیات تصویر صورت با مدل ESRGAN [۱۰]
 Figure 8. Super-resolution in face image details with ESRGAN model [10]



شکل ۹: مقایسه جزئیات تصویر در مدل ESRGAN [۱۰] و مدل پیشنهادی
 Figure 9. Comparison of image details in the ESRGAN model [10] and the proposed model

مقایسه کمی این نتایج در جدول ۳ ارائه شده است. همان طور که دیده می شود مدل پیشنهادی با حفظ SSIM مقداری PSNR بیشتری نسبت به پژوهش های اخیر داشته است. مهم تر این که مقدار معیار LPIPS کاهش یافته است که بیانگر این است که مدل پیشنهادی شباهت ادراکی بیشتری ارائه می دهد.

جدول ۳: مقایسه کمی مدل پیشنهادی با پژوهش های اخیر

Table 3. Quantitative comparison of the proposed model with recent research

مدل	PSNR (dB)	SSIM (%)	LPIPS (%)	دقت تائید چهره در دیتاست (%)	مرجع
BSRGAN	۲۴/۴۸	۷۱	۳۹/۱۰	۹۸/۷۰	[۱۵]
ESRGAN	۲۴/۵۰	۷۱/۵۰	۳۸/۷۰	۹۸/۸۰	[۱۰]
RGI	۱۹/۸۵	۸۱	۳۸/۶۰	۹۸/۸۰	[۳۲]
MI-GAN	۲۴/۶۵	۷۲/۵۰	۳۹	۹۸/۷۰	[۳۳]
مدل پیشنهادی	۲۵/۴۲	۷۱/۵۰	۳۸/۴۰	۹۸/۸۰	مقاله حاضر

نکته قابل توجه در این پژوهش، کاهش مقدار LPIPS است که باعث تولید تصاویر واقعی تر چهره و حفظ جزئیات چهره شده است. نتایج این کار می تواند به پژوهش های حوزه احراز هویت بیومتریک، کمک شایانی کند.

۵- نتیجه گیری

در این کار خرابی های پیچیده تصویر چهره مدل سازی شده است و زوج تصاویر با کیفیت و بی کیفیت بعنوان ورودی مدل تشخیص خرابی تصویر اعمال شد. با این کار و با استفاده از شبکه های SynNet و DegNet، مدل تشخیصی خرابی تصویر در شبکه مولد تخصصی بهبود یافت و جزئیات تصویر حفظ شد. در نتیجه عملکرد مناسبی در فرانتفیک پذیرگی شاهد بودیم. نتایج نشان می دهد که با ارائه این مدل، پارامترهای تصویر به PSNR=25/42 dB ، SSIM=71/5% و LPIPS=38/4% بهبود یافته که نسبت به پژوهش های اخیر پیشرفت قابل توجهی را شاهد هستیم.

مراجع

- [1] P. Kaur and H. S. Pannu, "Comparative analysis of continuous and discrete orthogonal moments for face recognition," *Proc. Int. Conf. Electron. Commun. Aerosp. Technol. ICECA*, 2017, pp. 449–453, 2017, doi: 10.1109/ICECA.2017.8203724.
- [2] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," *International Conference on Communication and Signal Processing (ICCSPP)*, Chennai, India, 2017, pp. 0588–0592, doi: 10.1109/ICCSPP.2017.8286426.
- [3] A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.
- [4] J. Kim, J. K. Lee and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.
- [5] S. Guo, Z. Yan, K. Zhang, W. Zuo and L. Zhang, "Toward convolutional blind denoising of real photographs," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722, doi: 10.1109/CVPR.2019.00181.
- [6] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li and F. Huang, "Real-world super-resolution via kernel estimation and noise injection," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, 2020, pp. 1914–1923, doi: 10.1109/CVPRW50498.2020.00241.
- [7] Z. Luo, Y. Huang, S. Li, L. Wang and T. Tan, "Unfolding the alternating optimization for blind super

- resolution,” *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 5632-5643, 2020.
- [8] J. Gu, H. Lu, W. Zuo, and C. Dong, “Blind super-resolution with iterative kernel correction,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1604–1613, doi: 10.1109/CVPR.2019.00170.
- [9] X. Wang, L. Xie, C. Dong and Y. Shan, “Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data,” *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 1905–1914, doi: 10.1109/ICCVW54120.2021.00217.
- [10] K. Zhang, J. Liang, L. Van Gool and R. Timofte, “Designing a Practical Degradation Model for Deep Blind Image Super-Resolution,” *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 4771–4780, doi: 10.1109/ICCV48922.2021.00475.
- [11] L. Liu and S. Liu, “Remote detection of human vital sign with stepped-frequency continuous wave radar,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 3, pp. 775–782, 2014, doi: 10.1109/JSTARS.2014.2306995.
- [12] M. Elad and A. Feuer, “Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, 1997, doi: 10.1109/83.650118.
- [13] C. Liu and D. Sun, "On Bayesian Adaptive Video Super Resolution," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 346-360, Feb. 2014, doi: 10.1109/TPAMI.2013.127.
- [14] K. Zhang, W. Zuo and L. Zhang, "Deep Plug-And-Play Super-Resolution for Arbitrary Blur Kernels," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 1671-1681, doi: 10.1109/CVPR.2019.00177.
- [15] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo and L. Zhang, “Blind Face Restoration via Deep Multi-scale Component Dictionaries,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2020, pp. 399–415, doi: 10.1007/978-3-030-58545-7_23.
- [16] Z. Wei, Y. Huang, Y. Chen, C. Zheng and J. Gao, “A-ESRGAN: Training Real-World Blind Super-Resolution with Attention U-Net Discriminators,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2023, pp. 16–27, doi: 10.1007/978-981-99-7025-4_2.
- [17] M. Zhang and Q. Ling, “Supervised Pixel-Wise GAN for Face Super-Resolution,” *IEEE Trans. Multimed.*, vol. 23, pp. 1938–1950, 2021, doi: 10.1109/TMM.2020.3006414.
- [18] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet and M. Norouzi, "Image Super-Resolution via Iterative Refinement," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713-4726, 1 April 2023, doi: 10.1109/TPAMI.2022.3204461.
- [19] J. Jiang, Y. Yu, S. Tang, J. Ma, A. Aizawa and K. Aizawa, “Context-Patch Face Hallucination Based on Thresholding Locality-Constrained Representation and Reproducing Learning,” *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 324–337, 2020, doi: 10.1109/TCYB.2018.2868891.
- [20] Y. Yin, J. P. Robinson, Y. Zhang and Y. Fu, “Joint super-resolution and alignment of tiny faces,” *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 12693–12700, 2020, doi: 10.1609/aaai.v34i07.6962.
- [21] A. Lugmayr, M. Danelljan, L. Van Gool and R. Timofte, “SRFlow: Learning the Super-Resolution Space with Normalizing Flow,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2020, pp. 715–732, doi: 10.1007/978-3-030-58558-7_42.
- [22] C. Chen, D. Gong, H. Wang, Z. Li, and K. Y. K. Wong, “Learning Spatial Attention for Face Super-Resolution,” *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021, doi: 10.1109/TIP.2020.3043093.
- [23] Y. Wang, Y. Hu and J. Zhang, “Panini-Net: GAN Prior Based Degradation-Aware Feature Interpolation for Face Restoration,” *Proc. AAAI Conf. Artif. Intell. AAAI 2022*, vol. 36, pp. 2576–2584, 2022, doi: 10.1609/aaai.v36i3.20159.
- [24] T. Wang *et al.*, “A Survey of Deep Face Restoration: Denoise, Super-Resolution, Deblur, Artifact Removal,” 2022, [Online]. Available: <http://arxiv.org/abs/2211.02831>, doi: 10.48550/arXiv.2211.02831.
- [25] Y. Wang, Y. Hu, J. Yu and J. Zhang, “GAN Prior Based Null-Space Learning for Consistent Super-

- resolution,” *Proc. AAAI Conf. Artif. Intell. AAAI 2023*, vol. 37, pp. 2724–2732, 2023, doi: 10.1609/aaai.v37i3.25372.
- [26] Y. Liu, Z. Dong, K. Pang Lim and N. Ling, "A Densely Connected Face Super-Resolution Network Based on Attention Mechanism," in *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Kristiansand, Norway, 2020, pp. 148-152, doi: 10.1109/ICIEA48937.2020.9248111.
- [27] R. Ghorbandoost and F. Razzazi, "High Fidelity Reversible information Steganography in Images using Difference Expansion and Appropriate Region Selection," *Journal of Information and Communication Technology in Policing*, vol. 3, no. 9, pp. 1-16, 2022, doi: 10.22034/pitc.2022.1265729.1101[in persian].
- [28] I. Hadinejad, M. A. Amiri and M. H. Fahimifar, "An Optimum Method for Noise Reduction and Quality Improvement of the Passive Millimeter Wave Images Based on Nonsubsampled Shearlet Transform and Improved Adaptive Median Filter," *Journal of Information and Communication Technology in Policing*, vol. 3, no. 12, pp. 30-43, 2022, doi: 10.22034/pitc.2023.1271283.1179 [in Persian].
- [29] M. S. Kalami Yazdi, M. Nezhadshahbodaghi and M. R. Mosavi, "INS/Image Integrated Navigation System based on Deep Learning in order to Monitor the Places Traveled by Drivers," *Journal of Information and Communication Technology in Policing*, vol. 3, no. 11, pp. 35-46, 2022, doi: 10.22034/pitc.2022.1270483.1174 [in persian].
- [30] P. Samadinia, , K. Rahbar and A. Broumandnia, "Efficient Multi-Focus Image Fusion via Depthmap," *Journal of Information and Communication Technology in Policing*, vol. 3, no. 10, pp. 59-70, 2022, doi: 10.22034/pitc.2022.1269046.1148 [in persian].
- [31] M. Shokoohi, "Introducing an intelligent system for detecting traffic signs with deep learning to reduce road accidents," *Journal of Information and Communication Technology in Policing*, vol. 3, no. 10, pp. 47-58, 2022, doi: 10.22034/pitc.2022.1268864.1145 [in persian].
- [32] Z. Liu *et al.*, "Fine-Grained Face Swapping Via Regional GAN Inversion," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 8578-8587, doi: 10.1109/CVPR52729.2023.00829.
- [33] A. Sargsyan, S. Navasardyan, X. Xu and H. Shi, "MI-GAN: A Simple Baseline for Image Inpainting on Mobile Devices," *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 7301–7311, doi: 10.1109/ICCV51070.2023.00674.

COPYRIGHTS

©2025 by the authors. Published by the Islamic Azad University Bushehr Branch. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0>

