



Semnan University

# Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>

ISSN: 2783-2538



Type of Article (e.g. Research Article)

## Facial Thermal Image Translation to RGB Visible Light using GAN

Nastaran Malekpoor , Mohammad Javad Fadaeieslam 

Faculty of Electrical and Computer Engineering, Semnan University.

### PAPER INFO

#### Paper history:

Received: 2024-06-21  
Revised: 2022-05-12  
Accepted: 2022-10-19

#### Keywords:

Image-to-image translation;  
Thermal image;  
RGB visible light image;  
Generative Adversarial Network;  
Self attention.

### ABSTRACT

Thermal imaging systems, due to their unique features, such as the ability to record images in different weather conditions, recording images at night, or having anti-counterfeiting properties, have special military, security and judicial applications. However, the images recorded by thermal cameras cannot be recognized by the human eye, and it is very difficult for humans to recognize the faces of thermal images. Converting thermal images to visible light images is in the field of image-to-image content transfer or image to image translation. So far, many deep learning models have been introduced to convert thermal images into visible light. Among these models, adversarial networks have been able to achieve significant progress in this field. In this paper, an attempt was made to improve the ClawGAN network, which is specifically designed to convert thermal images into visible light. Our method is based on the integration of effective techniques such as Unet++, Unet3+, self-attention network in the generator of the base model. In this way, the network will be able to produce higher quality images that can be recognized by the human eye and have minimal distortion, blur and noise when transferring content from the thermal domain to the visible light domain. The obtained results showed that the proposed generator was able to significantly improve the evaluation criteria such as MSE, PSNR, RMSE, UQI and PSNR-B. DOI: <https://doi.org/>

© 2024 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

\* Mohammad Javad Fadaeieslam.  
E-mail address: [fadaei@semnan.ac.ir](mailto:fadaei@semnan.ac.ir)

How to cite this article:

## ترجمه تصویر حرارتی به نور مرئی با استفاده از شبکه‌های مولد تخصصی

نسترن ملکپور<sup>۱</sup>، محمدجواد فدائی اسلام<sup>۲\*</sup>

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۴۰۳/۰۴/۱	سیستم‌های تصویربرداری حرارتی، با توجه به ویژگی‌های منحصر به فرد خود، مانند توانایی ثبت تصاویر در شرایط آب‌وهوای مختلف، ثبت تصاویر در شب و یا دارا بودن خاصیت ضدجعل، کاربردهای نظامی، امنیتی و قضایی ویژه‌ای دارند. با این حال، تصاویر ثبت شده توسط دوربین‌های حرارتی، با استفاده از چشم انسان قابل تشخیص نبوده و شناسایی چهره تصاویر حرارتی، برای انسان بسیار سخت است. تبدیل تصاویر حرارتی به تصاویر نور مرئی، در حوزه انتقال محتوای تصویر یا ترجمه تصویر به تصویر قرار دارد. تاکنون، مدل‌های یادگیری عمیق بسیاری برای تبدیل تصاویر حرارتی به نور مرئی معرفی شده‌اند. از بین این مدل‌ها، شبکه‌های مولد تخصصی توانسته‌اند به پیشرفت قابل توجهی در این زمینه دست پیدا کنند. در این مقاله، سعی شد تا شبکه ClawGAN که به‌طور خاص برای تبدیل تصاویر حرارتی به نور مرئی ارائه شده‌است، بهبود داده‌شود. راهکار ما بر پایه‌ی ادغام تکنیک‌های موثر نظیر Unet++، Unet3+ و شبکه خودتوجه در مولد معماری پایه است. بدین صورت، شبکه قادر خواهد بود تا در زمان انتقال محتوا از دامنه حرارتی به دامنه‌ی نور مرئی، تصاویر با کیفیت بالاتری را تولید کند که قابل تشخیص از طریق چشم انسان بوده و دارای کمترین اعوجاج، تاری و نویز باشند. نتایج بدست آمده نشان داد که مولد پیشنهادی توانست باعث بهبود قابل توجه معیارهای ارزیابی مانند MSE، PSNR، RMSE، UQI و PSNR-B شود.
بازنگری مقاله: ۱۳۹۵/۰۶/۲۰	
پذیرش مقاله: ۱۳۹۵/۰۷/۱۵	
<b>واژگان کلیدی:</b> ترجمه تصویر به تصویر، تصاویر حرارتی، تصاویر نور مرئی، شبکه‌های مولد تخصصی، شبکه خودتوجه	

DOI: <https://doi.org/>

© 2024 Published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

۱- مقدمه<sup>۱</sup>

ثبت شبانه‌روزی تصاویر چهره با وضوح بالا، همواره یکی از چالش‌های سیستم‌های تصویربرداری بوده است. یافتن دوربینی که قادر باشد تصاویر نور مرئی را با کیفیت و وضوح بالا در شرایط مختلف جوی و روشنایی ثبت کند و دارای هزینه‌ی مقرون به‌صرفه‌ای باشد، مشکل است. دوربین‌های حرارتی از ابزارهای تصویربرداری برای ثبت تصاویر چهره هستند که علاوه بر توانایی ثبت تصویر در انواع شرایط جوی و روشنایی، قابلیت ضدجعل نیز دارند. از همین روی در مراکز امنیتی و نظامی که ثبت تصاویر تحت تمام شرایط بسیار مهم است، مورد استفاده قرار می‌گیرند. اما، تصاویر ثبت‌شده توسط این دوربین‌ها برای چشم انسان قابل درک نیست.

ترجمه‌ی محتوای تصاویر حرارتی به تصاویر نور مرئی می‌تواند این محدودیت را برطرف نماید. تاکنون، تعداد زیادی مدل شبکه عصبی برای ترجمه‌ی تصاویر حرارتی به نور مرئی معرفی شده است. وانگ و همکاران [۱] یک شبکه آشکارساز پیشنهاد کردند که می‌تواند ویژگی‌های برجسته چهره را برای تصاویر قابل مشاهده واقعی تر استخراج کند. شبکه مولد تخصصی که برای اولین بار در سال ۲۰۱۴ معرفی شد [۴]، روشی کارآمد برای ترجمه تصویری است؛ آن برای فرار از مسائل مرتبط با آموزش زنجیره مارکوف مونت کارلو، از تکنیک پس‌انتشار خطا<sup>۲</sup> استفاده می‌کند [۲۲ و ۲۳]. این شبکه یاد می‌گیرد که توزیع نهفته ساده را به توزیع داده پیچیده تر نگاهت کند. برای ثبت اطلاعات پیچیده توزیع داده، معماری شبکه مولد تخصصی باید ظرفیت کافی داشته‌باشد. این شبکه، براساس مفهوم یک بازی رقابتی از دو شبکه مولد (G) و یک جداکننده (D) تشکیل شده‌است که آن‌ها با هدف یادگیری با یکدیگر رقابت می‌کنند.

ایمیدستی و همکاران [۲] شبکه Axial-GAN را برای سنتز تصاویر قابل مشاهده با وضوح بالا برای تطبیق بهتر پیشنهاد کردند. لیو و همکاران [۳] ابتدا تصاویر مادون قرمز

حرارتی ثبت‌شده در شب را به تصاویر مرئی در مقیاس خاکستری (GVI) ترجمه کردند و سپس GVI را به تصاویر قابل مشاهده رنگی (CVI) ترجمه کردند. در مقاله [۵] محققان جلوه بصری و عملکرد تشخیص چهره تصاویر چهره را با ترجمه تصویر مادون قرمز نزدیک به نور مرئی بهبود بخشیدند. ژانگ و همکاران Pix2pix [۶] را با اضافه کردن تابع اتلاف سنتز و تابع اتلاف هویت بهبود بخشیدند؛ اما شبکه آن‌ها این مورد را در نظر نمی‌گیرد که داده‌های مزدوج<sup>۳</sup>، دقیقاً از نظر پیکسلی تراز شده نیستند.

شبکه ClawGAN [۷] بر اساس چارچوب CycleGAN [۸] ساخته شد. این شبکه از اتلاف سنتز شده، اتلاف تولیدی-بازسازی، اتلاف خصمانه و اتلاف ثبات چرخه استفاده می‌کند تا یک تابع هدف جدید را تشکیل دهد. همچنین، یک شبکه متصل به claw برای جایگزینی ساختار شبکه U-net مولد به منظور حفظ بیشتر ویژگی‌ها فراخوانی می‌شود. تفاوت اصلی این شبکه با شبکه مولد تخصصی وانیلی و دیگر شبکه‌های موجود، شکل مولد منحصر به فرد آن است. در شکل (۱) معماری مولد این شبکه ترسیم شده‌است. ClawGAN، ویژگی‌های ساختاری تصاویر حرارتی را حفظ می‌کند و نه تنها کیفیت تصاویر را افزایش می‌دهد، بلکه به طور موثر نتایج ترجمه تصویر را در نور روشن و تاریک بهبود می‌بخشد.

شبکه ClawGAN [۷] با وجود پیشرفت‌های قابل توجه در کار ترجمه تصویر حرارتی به تصویر نور مرئی، در برخی شرایط، نمی‌تواند حالات صحیح لباس‌ها و چهره‌ها را زمانی که چهره‌های ناشناخته وارد می‌شوند، به خوبی حفظ کند. بنابراین، تعمیم این شبکه نیاز به بهبود دارد. همچنین، در آزمایش‌ها مشخص شد که زمانی که تصویر هدف چرخش داشته باشد، تصویر تولیدشده توسط این شبکه، تا حدی مخدوش می‌شود.

باتوجه به برخی از مزایای شبکه ClawGAN نسبت به سایر روش‌های ترجمه تصویر حرارتی به نور مرئی، این شبکه به‌عنوان مدل پایه انتخاب شد. در این مقاله،

\* پست الکترونیک نویسنده مسئول: fadaei@semnan.ac.ir

۱. فارغ التحصیل کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

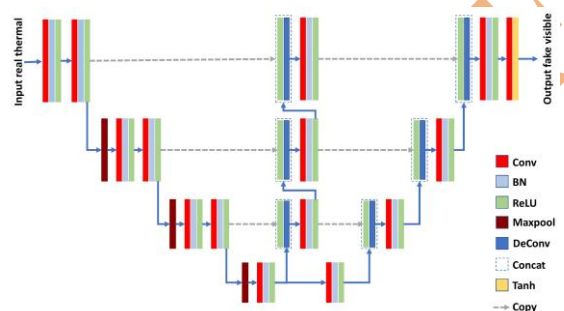
۲. استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

<sup>2</sup> backpropagation<sup>3</sup> paired

استخراج ویژگی‌های کافی از تصاویر حرارتی دارد تا جزئیات چهره را تکمیل کند. اگر ویژگی‌های کم‌عمق مستقیماً با یک شکاف بزرگ به ویژگی‌های عمیق متصل شوند، تفاوت‌های بیش‌از حد اطلاعات معنایی، دشواری یادگیری شبکه را افزایش می‌دهد و ویژگی‌های لایه کم‌عمق نمی‌توانند ویژگی‌های عمیق را تکمیل کنند. شکل (۲) چهارچوب مدل ClawGAN را نشان می‌دهد.

مطابق این شکل، تصاویر حرارتی و مرئی مزدوج که با تطبیق الگو (TM) برش داده می‌شوند، به ترتیب وارد مدل می‌شوند. تصویر حرارتی جعلی  $T_{fake}$  و تصویر مرئی جعلی  $V_{fake}$  به ترتیب توسط مولد  $Generator_{VT}$  و  $Generator_{TV}$  از اتصالات تودرتو تشکیل شده‌اند. به طور مشابه، تصویر حرارتی جعلی  $T_{fake}$  و تصویر مرئی جعلی  $V_{fake}$  به ترتیب به مولدهای  $Generator_{TV}$  و  $Generator_{VT}$  وارد می‌شوند تا تصویر مرئی بازسازی شده چرخه‌ای و تصویر حرارتی بازسازی شده چرخه‌ای را بدست آورند.

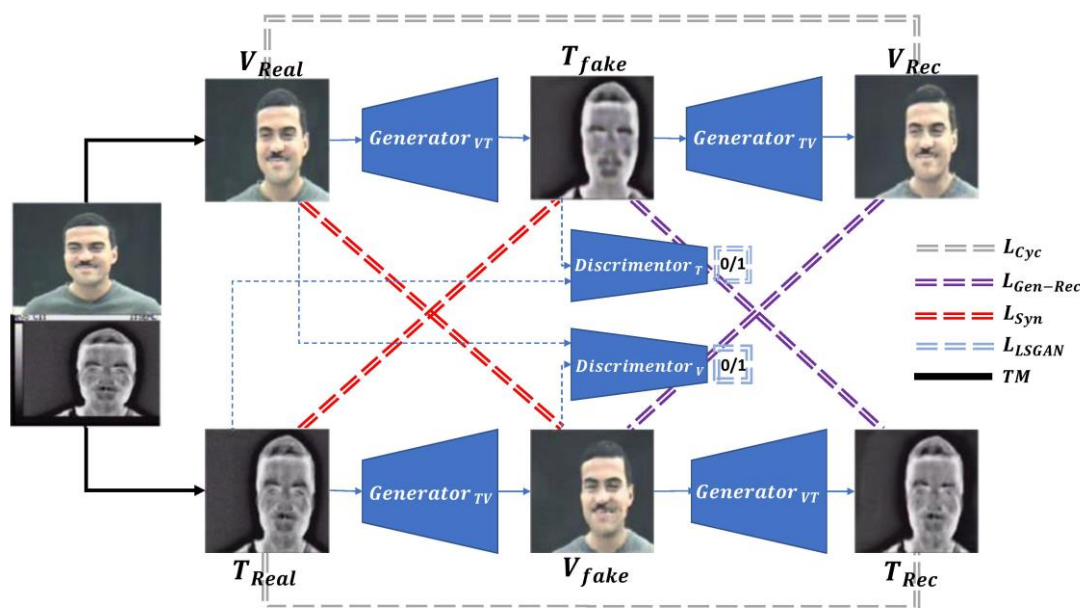
ماژول‌های خودتوجه<sup>۴</sup> جهت ارتقای شبکه ClawGAN به این معماری اضافه شدند. از سوی دیگر در راستای بهبود تصویر ترجمه شده، لایه‌ها و ارتباطات بین آن‌ها با استفاده ساختارهای Unet [۹]، Unet++ [۱۰] و Unet3+ [۱۱] تغییر داده شد که در فصل روش پیشنهادی جزئیات آن آمده است.



شکل (۱) معماری مولد شبکه ClawGAN [۷]

## ۲- روش مقاله پایه

بسیاری از مولدهای GAN از معماری شبکه U-net [۹] استفاده کرده‌اند. در کار ترجمه تصویر صورت، مدل نیاز به



شکل (۲) چهارچوب مدل ClawGAN [۷]

تشخیص تصویر واقعی از تصویر تولیدشده عمل می‌کنند.  $L_{Cyc}$  نشان‌دهنده اتلاف ثبات چرخه است،  $L_{Gen-Rec}$

$Discriminator_V$  و  $Discriminator_T$  جداکننده‌هایی هستند که در حوزه دامنه حرارتی و دامنه مرئی برای

<sup>4</sup> self-attention

ماژول خودتوجه در شبکه مولد، بهترین نتیجه زمانی حاصل شد که بلاک‌های کانولوشنی لایه میانی حذف شد و ماژول‌های خودتوجه جایگزین شدند. همچنین، اتصالات شبکه مولد بازسازی شدند و اتصالات بین بخش نمونه‌برداری پایین و قسمت میانی با الهام از اتصالات Unet3+ ساخته شدند.

در جداول (۱)، (۲) و (۳) اتصالات و پارامترهای لایه‌های کدگذار، کدگشا و لایه میانی نشان داده شده‌است.

جدول (۱) ترتیب لایه‌های کدگذار

نام لایه	ورودی	پارامترها	شکل خروجی
Input	None	0	256×256×3
Conv2d_10	Input	896	256×256×32
LN	Conv2d_10	64	256×256×32
Conv2d_11	LN	9248	256×256×32
LN_1	Conv2d_11	64	256×256×32
Maxpool2d	LN_1	0	128×128×32
Conv2d_12	Maxpool2d	18496	128×128×64
LN_2	Conv2d_12	128	128×128×64
Conv2d_13	LN_2	36928	128×128×64
LN_3	Conv2d_13	128	128×128×64
Maxpool2d_1	LN_3	0	64×64×64
Conv2d_14	Maxpool2d_1	73856	64×64×128
LN_4	Conv2d_14	256	64×64×128
Conv2d_15	LN_4	147584	64×64×128
LN_5	Conv2d_15	256	64×64×128
Maxpool2d_2	LN_5	0	32×32×128
Conv2d_16	Maxpool2d_2	295168	32×32×256
LN_6	Conv2d_16	512	32×32×256

در مدل پایه، از ReLU [۱۳] به عنوان تابع فعالسازی استفاده شده‌است و نرمالسازی دسته‌ای [۱۴] در همه لایه‌ها به جز آخرین لایه کانولوشنی انجام شده‌است. در این مقاله، برای اطمینان از ویژگی‌های کانولوشنی محلی خوب و همگرایی سراسری گرادین کاهشی، ترکیب فعالساز و نرمالساز تغییر داده شد. برای عملکرد بهتر شبکه از نرمالسازی لایه‌ای [۱۵] به جای نرمالسازی دسته‌ای و از فعالساز GeLU [۱۶] به جای ReLU استفاده شد.

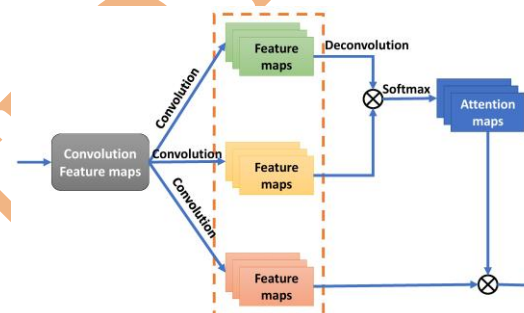
#### ۴- نتایج پیاده‌سازی

آزمایش‌های مختلفی برای بررسی عملکرد شبکه‌ی پیشنهادی انجام شد؛ که نتایج آن در این بخش آمده‌است. در ابتدا، مجموعه داده‌های بکاررفته در پیاده‌سازی توضیح داده شده و سپس، جزئیات پیاده‌سازی و نتایج ارزیابی‌های کمی و کیفی ارائه شده‌است.

نشان‌دهنده اتلاف تولیدی-بازسازی،  $L_{syn}$  نشان‌دهنده اتلاف سنتز شده، و  $L_{LSGAN}$  نشان‌دهنده اتلاف خصمانه است. طبق این شکل، عملکرد مدل در کار ترجمه تصویر صورت عمدتاً به عملکرد مولد برای تبدیل تصاویر حرارتی/مرئی بستگی دارد. شبکه جداکننده، تصویری به ابعاد  $۲۵۶ \times ۲۵۶ \times ۳$  را می‌گیرد و امتیازی از این تصویر جعلی یا واقعی ایجاد می‌کند (۰ برای جعلی و ۱ برای واقعی). از چهار تابع اتلاف مورد استفاده، سه مورد برای برزرسانی پارامترهای مولد و یک تابع اتلاف برای برزرسانی پارامترهای جداکننده استفاده شده‌اند.

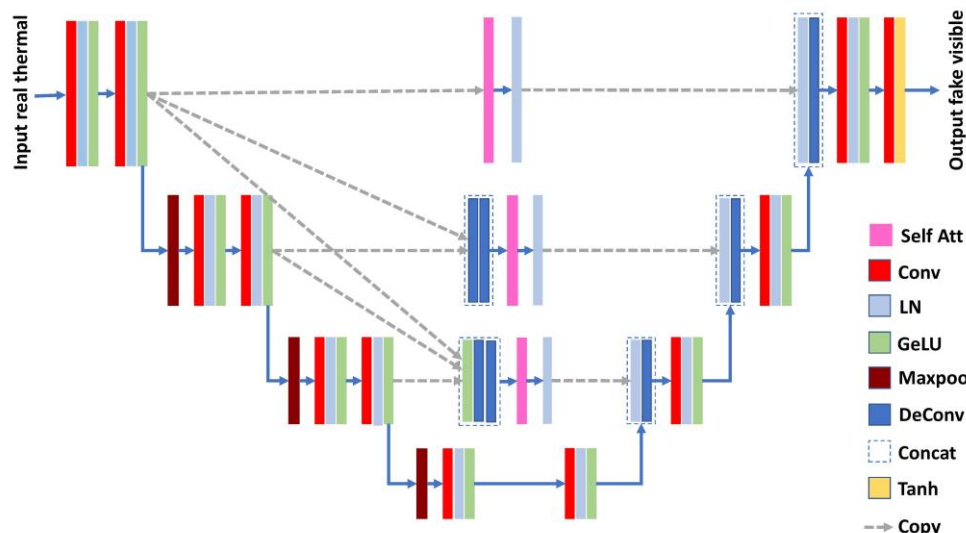
#### ۳- روش پیشنهادی

در این مقاله، از معماری‌های شبکه خودتوجه [۱۲] و ساختار شبکه اتصال نودرتو برای ارتقای معماری ClawGAN [۷]، استفاده شد. شکل (۳) معماری شبکه خودتوجه به کار رفته در این پژوهش را نشان می‌دهد.



شکل (۳) معماری شبکه خودتوجه.  $\otimes$  نشان‌دهنده ضرب ماتریس است و عملیات softmax در هر ردیف انجام می‌شود. این مکانیسم را می‌توان به سادگی به‌عنوان یک نگاهت ویژگی در نظر گرفت که در deconvolution آن ضرب می‌شود و اجازه می‌دهد پیکسل‌های هر دو مکان، مستقیماً با همدیگر در ارتباط باشند [۱۲]

شکل (۴) معماری شبکه مولد پیشنهادی را نشان می‌دهد. برای تمرکز شبکه مولد به تصاویر با جزئیات بیشتر و همچنین انتقال ویژگی‌های مهم بین لایه‌های نمونه‌برداری پایین و نمونه‌برداری بالا، لایه‌های شبکه خودتوجه بین این دو قسمت قرار گرفتند. به عبارتی دیگر، برای پیاده‌سازی ماژول خودتوجه، از بخش میانی شبکه مولد که ویژگی‌ها را از لایه‌های پایین به لایه‌های بالاتر منتقل می‌کند، استفاده شد. در نهایت، پس از آزمایش مدل‌های متفاوت قرارگیری



شکل (۴) مولد مدل پیشنهادی

که ماهیت آماری دارند، وفاداری اطلاعات بصری (VIF) [۲۱] و شاخص کیفیت عمومی (UQI) [۲۲] که براساس نظریه اطلاعات هستند، تشابه ساختاری (SSIM) [۲۳] و نسبت سیگنال به نویز شامل جلوه‌های مسدودکننده (PSNR-B) [۲۴] که براساس اطلاعات ساختاری هستند؛ و همچنین FID [۲۵] که براساس استخراج ویژگی‌های شبکه است، استفاده شد. FID بردارهای ۲۰۴۸ بعدی را بین لایه‌های کاملاً متصل به‌عنوان ویژگی‌های تصویر توسط شبکه Inception V3 از پیش آموزش‌دیده استخراج می‌کند و توزیع ویژگی‌ها با توزیع گاوسی چندمتغیره را برای محاسبه فاصله FID بین توزیع نرمال تصویر واقعی و توزیع نرمال تصویر تولیدشده متناسب می‌کند. FID می‌تواند یک معیار ارزیابی عالی برای شبکه‌های GAN باشد.

#### ۲-۴- جزئیات پیاده‌سازی

تصاویر ورودی، همانند شبکه ClawGAN، با اندازه ۲۵۶×۲۵۶ به مدل وارد می‌شوند. به نوبت، تثبیت وزن‌های یک طرف مولد و جداکننده در آموزش، برورسانی طرف دیگر و بهینه‌سازی شبکه در دستور کار قرار دارد. مولد و جداکننده هر دو از Adam به‌عنوان بهینه‌ساز استفاده می‌کنند. اندازه دسته روی ۱ تنظیم شده‌است؛ زیرا در این کار، برای انتخاب اندازه دسته بزرگتر، به منابع بیشتری نیاز بود و محیط کولب قادر به برآورده کردن این منابع نیست. در این کار همانند ClawGAN، نرخ یادگیری را روی

#### ۴-۱- مجموعه داده‌ها و معیارهای ارزیابی

در این پژوهش، برای آزمایش‌ها از پایگاه داده صورت حرارتی امرئی (IEEE OTCBVS WS Series Bench, 2000 [۱۹]) استفاده شد که در آن، تصاویر حرارتی توسط دوربین Raytheon Palm-IR-Pro و تصاویر قابل مشاهده توسط دوربین Panasonic WV-CP234 گرفته شده‌است. این مجموعه داده، یک مجموعه داده معمولی از تصاویر چهره مرئی حرارتی و RGB است و شامل ۴۲۲۸ تصویر مزدوج (جمعاً ۸۴۵۶) حرارتی/مرئی ۳۲۰×۲۴۰ متشکل از ۲۹ فرد است. همه‌ی افراد دارای حالت‌ها، زوایا، سیما و نور مختلف هستند. از آنجایی که در این مجموعه داده، تصاویر با ژست‌ها، روشنایی و سوژه‌های تکراری زیادی وجود دارد، در این کار مانند کارهای مشابه، موارد تکراری حذف شدند و تصاویر غیرتکراری و با وضوح بالاتر انتخاب شده‌اند. مشابه [۱۹]، برای ارزیابی میزان دقیق بهبود مدل مورد بررسی، تصاویر مزدوج مربوط به ۲۹ فرد استخراج شد و تعداد ۶۲۲ تصویر مزدوج به دست آمد. مجموعه آموزش و آزمایش با نسبت ۸۰ و ۲۰ از داده اصلی استخراج شد.

رابطه بین تصویر حرارتی و تصویر مرئی، با نگاشت پیکسل به پیکسل مطابقت دارد. بنابراین، برای بررسی میزان بهبود مدل پیشنهادی با مدل پایه، از معیارهای رایج ارزیابی کیفیت تصویر مانند خطای میانگین مربع خطاها (RMSE) و نسبت سیگنال به نویز پیک (PSNR) [۲۰]

معیارهای ارزیابی، عملکرد بهتری نسبت به مدل ClawGAN دارد. ارزیابی بصری تصویر واقعی و خروجی مدل‌های ClawGAN و مدل پیشنهادی در شکل (۵) نشان داده شده‌است. در جدول (۶)، پارامترهای مدل ClawGAN و مدل پیشنهادی مشاهده می‌شود. همان‌طور که در جدول (۵) مشاهده می‌شود، با اینکه تعداد پارامترهای قابل آموزش افزایش یافته، اما تعداد پارامترهای غیرقابل آموزش در مدل پیشنهادی به صفر رسیده‌است.

با بررسی نتایج، می‌توان اینطور نتیجه گرفت که مدل پیشنهادی از بسیاری جهات نسبت به مدل ClawGAN برتری دارد و می‌تواند به‌عنوان یک مدل مترجم، نتایج بهتری را ارائه دهد.

۰.۰۰۰۲ تنظیم شد. فرآیند آموزش مدل پیشنهادی مشابه مقاله پایه است که در آن ساز و کاری برای کاهش عدم تطابق نیز در نظر گرفته شده‌است.

#### ۳-۴- ارزیابی کمی و کیفی

در این پژوهش، روش پیشنهادی با روش ClawGAN مقایسه شد. به دلیل محدودیت‌های محیط کولب و همچنین سنگین بودن شبکه مولد تخصصی، مدل ClawGAN و مدل پیشنهادی تا ۲۰ اپیک آموزش داده شد و سپس ارزیابی‌های لازم صورت گرفت. در جدول (۴) نتایج معیارها آورده شده‌است. قسمت‌های پررنگ، برتری نسبی را نشان می‌دهند. همان‌طور که در جدول (۴) مشاهده می‌شود، مدل پیشنهادی برطبق بسیاری از

جدول (۲) ترتیب لایه‌های میانی همراه با ماژول خودتوجه

نام لایه	ورودی	تعداد پارامترها	شکل خروجی
DeConv2d_1	LN_3	16448	256×256×64
DeConv2d	LN_5	262272	256×256×128
Concat	LN_1, DeConv2d_1, DeConv2d	0	256×256×224
Conv2d_17	Concat	50400	256×256×224
Conv2d_18	Concat	50400	256×256×224
Conv2d_19	Concat	50400	256×256×224
DeConv2d_18	Conv2d_18	0	256×224×256
Matmul	Self1, DeConv2d_18, Conv2d_17	0	256×224×224
Softmax	Matmul	0	256×224×224
Matmul_1	Conv2d_19, Softmax	0	256×256×224
Conv2d_20	Matmul_1	200928	256×256×224
ADD	Concatenate, $\gamma \times$ Conv2d_20	0	256×256×224
LN_7	ADD	448	256×256×224
Concat_1	LN_1, DeConv2d_1	0	256×256×96
Conv2d_21	Concat_1	9312	256×256×96
Conv2d_22	Concat_1	9312	256×256×96
Conv2d_23	Concat_1	9312	256×256×96
DeConv2d_22	Conv2d_22	0	256×96×256
Matmul_2	Self2, DeConv2d_22, Conv2d_21	0	256×96×96
Softmax_1	Matmul_2	0	256×96×96
matmul_3	Conv2d_23, Softmax_1	0	256×256×96
Conv2d_24	matmul_3	36960	256×256×96
Add_1	Concat_1, $\gamma \times$ Conv2d_24	0	256×32×32
LN_8	Add_1	192	256×256×96
Conv2d_25	LN_1	1056	256×256×32
Conv2d_26	LN_1	1056	256×256×32
Conv2d_27	LN_1	1096	256×256×32
DeConv2d_26	Conv2d_26	0	256×32×256
Matmul_4	Self3, Conv2d_26, Conv2d_25	0	256×32×32
Softmax_2	Matmul_4	0	256×32×32
Matmul_5	Conv2d_27, Softmax_2	0	256×256×32
Conv2d_28	Matmul_5	4128	256×256×32
Add_2	LN_1, $\gamma \times$ Conv2d_28	0	256×256×32
LN_9	Add_2	64	256×256×32

جدول (۳) ترتیب لایه‌های کدگشا

نام لایه	ورودی	تعداد پارامترها	شکل خروجی
Conv2d_29	LN_6	590080	32×32×256
LN_10	Conv2d_29	512	32×32×256
De_Conv2d	LN_10	524416	128×128×128
De_Conv2d_3	De_Conv2d_2	65664	256×256×128
Concat_2	De_Conv2d_3, LN_7	0	256×256×352
Conv2d_30	Concat_2	405632	256×256×128
LN_11	Conv2d_30	256	256×256×128
Maxpool2d_3	LN_11	0	128×128×128
De_Conv2d_4	Maxpool2d_3	32832	256×256×64
Concat_3	De_Conv2d_4, LN_8	0	256×256×160
Conv2d_31	Concat_3	92224	256×256×64
LN_12	Conv2d_31	128	256×256×64
Maxpool2d_4	LN_12	0	128×128×64
De_Conv2d_5	Maxpool2d_4	8224	256×256×32
Concat_4	De_Conv2d_5, LN_9	0	256×256×64
Conv2d_32	Concat_4	18464	256×256×32
LN_13	Conv2d_32	64	256×256×32
Conv2d_33	LN_13	1539	256×256×3

جدول (۴) نتایج ارزیابی کمی مدل‌های پایه و مدل پیشنهادی

FID↓	VIF↑	PSNR-B↑	PSNR↑	SSIM↑	UQI↑	RMSE↓	MSE↓	مدل
186	3.58	-36.1	8.17	-0.016	0.001	87.1	8798	ClawGAN
<b>173</b>	<b>5.47</b>	-36.62	<b>8.93</b>	-0.029	<b>0.007</b>	<b>79.09</b>	<b>7205</b>	Our Method



شکل (۵) ارزیابی بصری مدل پایه و مدل پیشنهادی

جدول (۵) پارامترهای مدل ClawGAN و مدل پیشنهادی

Non-trainable params	Trainable params	Total params	روش‌ها
2,816	2,295,715	2,298,531	ClawGAN
0	3,027,363	3,027,363	روش پیشنهادی

شد. از آنجایی که شبکه‌های مولد تخصصی، پیچیدگی‌های زیادی دارند و عامل بهبود در یک شبکه مولد تخصصی، لزوماً باعث بهبود در شبکه‌های مولد تخصصی دیگر نمی‌شود، برای انجام این تحقیق، مدل‌های مختلف

### ۵- نتیجه‌گیری

انگیزه اصلی در این مقاله، بررسی تأثیر روش‌های بهبود رایج در شبکه‌های مولد تخصصی بود و برای انجام این بررسی‌ها، شبکه ClawGAN به‌عنوان مدل پایه انتخاب



نسبت به شبکه ClawGAN بدست آمد. نتایج بدست آمده نشان داد که استفاده از ماژول خودتوجه و تغییر اتصالات در مولد، می‌تواند باعث بهبود قابل توجه معیارهای ارزیابی مانند MSE، PSNR، RMSE، UQI و PSNR-B شود. با این حال، شبکه بازسازی شده نسبت به مدل پایه دارای پیچیدگی‌ها و پارامترهای بیشتری است و همین عامل باعث بالا رفتن زمان اجرای مدل در محیط کولب می‌شود. به همین دلیل، در این کار از تعداد تصاویر کمتری نسبت به مقاله [۷] استفاده شد. بررسی‌های بعدی می‌تواند شامل بررسی روش‌های جدید برای کمتر کردن تعداد پارامترهای شبکه با حفظ ویژگی‌های مهم آن و در نتیجه بیشتر کردن تصاویر مجموعه داده آموزشی باشد.

پایاده‌سازی ماژول‌های خودتوجه و تغییر اتصالات مدل مولد با الهام از شبکه unet3+ بررسی شدند. در بررسی‌های اولیه، مشخص شد که استفاده از نرمالسازی لایه‌ای به جای نرمالسازی دسته‌ای و استفاده از فعالساز GeLU به جای ReLU باعث بهبود جزئی شبکه می‌شود. به همین دلیل، در ادامه پژوهش، این دو تغییر در تمام بررسی‌ها لحاظ شدند. برای پایاده‌سازی ماژول خودتوجه، از بخش میانی شبکه مولد که ویژگی‌ها را از لایه‌های پایین به لایه‌های بالاتر منتقل می‌کند، استفاده شد. در نهایت، بهترین نتیجه در زمان پایاده‌سازی ماژول خودتوجه، زمانی حاصل شد که بلاک‌های کانولوشنی لایه میانی حذف شده و ماژول‌های خودتوجه جایگزین شدند. در ادامه، با پایاده‌سازی خاصی از ماژول خودتوجه و تغییر برخی از اتصالات با الهام از معماری Unet3+، بهبود نسبی نتایج

## مراجع

- [1] Wang, Chaoyue, Chang Xu, Chaohui Wang, and Dacheng Tao. 2018. "Perceptual Adversarial Networks for Image-to-Image Transformation." *IEEE Transactions on Image Processing* 27 (8): 4066–79. <https://doi.org/10.1109/TIP.2018.2836316>.
- [2] Immidisetti, Rakhil, Shuowen Hu, and Vishal M. Patel. 2021. "Simultaneous Face Hallucination and Translation for Thermal to Visible Face Verification Using Axial-GAN." *2021 IEEE International Joint Conference on Biometrics, IJCB 2021*. <https://doi.org/10.1109/IJCB52358.2021.9484353>.
- [3] Liu, Shuo, Mingliang Gao, Vijay John, Zheng Liu, and Erik Blasch. 2021. "Deep Learning Thermal Image Translation for Night Vision Perception." *ACM Transactions on Intelligent Systems and Technology* 12 (1): 1–18. <https://doi.org/10.1145/3426239>.
- [4] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. "Generative Adversarial Networks." *Communications of the ACM* 63 (11). <https://doi.org/10.1145/3422622>.
- [5] Wu, Fangyu, Weihang You, Jeremy S. Smith, Wenjin Lu, and Bailing Zhang. 2019. "Image-Image Translation to Enhance Near Infrared Face Recognition." *Proceedings - International Conference on Image Processing, ICIP 2019-Sept*: 3442–46. <https://doi.org/10.1109/ICIP.2019.8804414>.
- [6] Isola, Phillip, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. "Image-to-Image Translation with Conditional Adversarial Networks." *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*: 5967–76. <https://doi.org/10.1109/CVPR.2017.632>.
- [7] Luo, Yi, Dechang Pi, Yue Pan, Lingqiang Xie, Wen Yu, and Yufei Liu. 2022. "ClawGAN: Claw Connection-Based Generative Adversarial Networks for Facial Image Translation in Thermal to RGB Visible Light." *Expert Systems with Applications* 191 (April 2021): 116269. <https://doi.org/10.1016/j.eswa.2021.116269>.
- [8] Dou, Hao, Chen Chen, Xiyuan Hu, and Silong Peng. 2019. "ASYMMETRIC CYCLEGAN FOR UNPAIRED NIR-TO-RGB FACE IMAGE TRANSLATION Institute of Automation, Chinese Academy of Sciences, Beijing, China University of Chinese Academy of Sciences, Beijing, China Beijing Visystem Co. Ltd, Beijing, China," 1757–61.

- [9] Weng, Weihao, and Xin Zhu. 2021. "UNet: Convolutional Networks for Biomedical Image Segmentation." *IEEE Access* 9: 16591–603. <https://doi.org/10.1109/ACCESS.2021.3053408>.
- [10] Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. "Unet++: A Nested u-Net Architecture for Medical Image Segmentation." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11045 LNCS: 3–11. [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1).
- [11] Huang, Huimin, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen Wei Chen, and Jian Wu. 2020. "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation." *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 2020-May (ii)*: 1055–59. <https://doi.org/10.1109/ICASSP40776.2020.9053405>.
- [12] Zhang, Han, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. 2019. "Self-Attention Generative Adversarial Networks." *36th International Conference on Machine Learning, ICML 2019 2019-June*: 12744–53.
- [13] Zou, Difan, Yuan Cao, Dongruo Zhou, and Quanquan Gu. 2020. "Gradient Descent Optimizes Over-Parameterized Deep ReLU Networks." *Machine Learning* 109 (3): 467–92. <https://doi.org/10.1007/s10994-019-05839-6>.
- [14] Ioffe, Sergey, and Christian Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." *32nd International Conference on Machine Learning, ICML 2015 1*: 448–56.
- [15] Ba, Jimmy Lei, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. "Layer Normalization." <http://arxiv.org/abs/1607.06450>.
- [16] Hendrycks, Dan, and Kevin Gimpel. 2016. "Gaussian Error Linear Units (GELUs)," 1–10. <http://arxiv.org/abs/1606.08415>.
- [17] Zhu, Jun Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." *Proceedings of the IEEE International Conference on Computer Vision 2017-October*: 2242–51. <https://doi.org/10.1109/ICCV.2017.244>.
- [18] Mao, Xudong, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, and Stephen Paul Smolley. 2017. "Least Squares Generative Adversarial Networks." *Proceedings of the IEEE International Conference on Computer Vision 2017-October*: 2813–21. <https://doi.org/10.1109/ICCV.2017.304>.
- [19] "Otcbvs 2004." n.d. <http://vcip1-okstate.org/pbvs/bench/Data/03/download.html>.
- [20] Willmott, Cort J., and Kenji Matsuura. 2005. "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30 (1): 79–82. <https://doi.org/10.3354/cr030079>.
- [21] Sheikh, Hamid Rahim, and Alan C. Bovik. 2006. "Image Information and Visual Quality." *IEEE Transactions on Image Processing* 15 (2): 430–44. <https://doi.org/10.1109/TIP.2005.859378>.

[۲۲] شیلاندری، آرش؛ حسین مروی؛ حسین خسروی، "افزایش داده و انتخاب مؤثر ویژگی در شبکه های مولد متخاصمی جهت تشخیص احساس از گفتار،" مدل سازی در مهندسی، ۲۱، ۷۲، (۱۴۰۲): ۱-۱۷.

[۲۳] جباری، حامد؛ نوشین بیگدلی، "یک شبکه مولد متخاصم کپسولی جدید برای طبقه بندی نامتوازن تصاویر اسپریم انسان"، ۲۱، ۷۳، (۱۴۰۲): ۲۷۹-۲۹۴.