

حذف نویز در انتقال محتوای تصویر به تصویر با استفاده از شبکه مولد تخصصی مبتنی بر تبدیل موجک

فرزانه مقصودی قمبوانی^۱، محمدجواد فدائی اسلام^۲ و فرزین یغمایی^۲

چکیده

یکی از چالش برانگیزترین موضوعات در هوش مصنوعی انتقال محتوای تصویر به تصویر است، که اخیراً با استفاده از شبکه‌های مولد تخصصی پیشرفت چشمگیری نموده است. با این حال، روش‌های موجود در انتقال محتوای منبع نویزی به دامنه هدف به صورت سربه‌سر شکست می‌خورند. برای رفع این مشکل، در این مقاله شبکه WTGAN پیشنهاد شده که شامل مولد جدید، جداکننده محلی و سراسری است که شبکه مولد آن بر اساس تبدیل موجک و سازوکار توجه طراحی شده است. از آنجا که تبدیل موجک ابزاری قدرتمند جهت حذف نویز عمومی از تصویر است، در ساختار مولد از آن استفاده شده است. همچنین با استفاده از سازوکار توجه، اتصالات باقیمانده و کنارگذر تناظر بهتر محتوا بین تصویر منبع و هدف به وجود آمده و به بهبود توانایی مولد و عملکرد آن کمک می‌نماید. در این مقاله با استفاده از توابع اتلاف معرفی شده در روش‌های دیگر به قدرتمندی مدل در جهت انتقال محتوای تصویر توأمان با حذف نویز پرداخته شده است. آزمایش‌های انجام شده بر روی مجموعه داده Cityscapes با استفاده از معیارهای ارزیابی PSNR، SSIM و LPIPS نشان دهنده این است که مدل می‌تواند به خوبی اثرات نویز در منبع را کاهش داده، ساختار را به خوبی حفظ نموده و به کیفیت مطلوب دست یابد.

کلیدواژه‌ها

انتقال محتوای تصویر به تصویر، شبکه مولد تخصصی، حذف نویز تصویر، سازوکار توجه، تبدیل موجک.

۱- مقدمه

موضوع مورد توجه بسیاری از محققان قرار دارد [۱]. در سال‌های اخیر روش‌های زیادی در این زمینه ارائه شده است که می‌توان به دو دسته مبتنی بر شبکه‌های مولد تخصصی (GAN) [۱] تا [۶] و غیر GAN [۷]، تقسیم نمود. شبکه‌های مولد تخصصی به خوبی قادر به یادگیری توزیع داده‌ها هستند و به همین سبب در این زمینه نسبت به دیگر روش‌ها برتری دارند. در روش‌های موجود، معمولاً نویز تصاویر ورودی در نظر گرفته نمی‌شود؛ در صورتی که ممکن است، نویز در هر سیستم تصویربرداری به ویژه در شرایط نور کم وجود داشته باشد [۸].

شبکه‌های مولد تخصصی توسط گودفلو و همکارانش [۹] ابداع و در طیف وسیعی از کاربردها از جمله سنتز تصویر، انتقال سبک، ترمیم تصویر و ... استفاده شدند. در GANها دو شبکه به

هدف انتقال محتوای تصویر به تصویر، یادگیری نگاشتی است که تصاویر در دو حوزه مختلف را به یکدیگر تبدیل کند. از آنجایی که طیف گسترده‌ای از کاربردها می‌توانند در این زمینه قرار گیرند، این

این مقاله در مردادماه سال ۱۴۰۱ دریافت شد؛ در آذرماه بازنگری و در دی‌ماه همان سال پذیرفته شد.

^۱ دانشجوی دکتری مهندسی کامپیوتر، گرایش هوش مصنوعی، دانشگاه سمنان، دانشکده مهندسی برق و کامپیوتر

رایانامه: farzane.maghsoudi@semnan.ac.ir

^۲ گروه مهندسی کامپیوتر، دانشگاه سمنان، دانشکده مهندسی برق و کامپیوتر

رایانامه: {fadaei, f_yaghmaee}@semnan.ac.ir

نویسنده مسئول: محمدجواد فدائی اسلام

فیلتر سه بعدی [۱۷]، تجزیه و تحلیل مؤلفه‌های اصلی با گروه‌بندی پیکسل محلی [۱۸]، مبتنی بر یادگیری فرهنگ لغت [۱۹] و ... است؛ آن‌ها بر اساس ویژگی‌های تصاویر و نویز طراحی شده‌اند. در مقابل، روش‌های مبتنی بر یادگیری، مانند شبکه عصبی کانولوشن حذف نویز [۲۰]، از مجموعه داده‌های تصویر مزدوج (به ازای هر تصویر در دامنه منبع یک تصویر متناظر در دامنه هدف وجود دارد) برای نقشه‌برداری از تصاویر نویزی به تصاویر تمیز استفاده می‌کنند که باعث وابستگی عملکرد روش به مجموعه داده‌های آموزشی بزرگ می‌شود. تصاویر نویزی به صورت مصنوعی از تصاویر تمیز با اضافه نمودن نویز شناخته شده (به‌عنوان مثال: گوسی سفید، نمک و فلفل و پواسون) ایجاد می‌شوند. با توجه به موفقیت آن‌ها در بینایی کامپیوتر و برتری نسبت به روش‌های سنتی، سال‌های اخیر پیشرفت سریع شبکه عصبی کانولوشن عمیق در حذف نویز تصویر صورت گرفته است [۲]. در مرجع [۲۱] از شبکه CycleGAN [۱۰] برای کار با تصاویر نویزی واقعی استفاده شد. در این روش تصاویر واضح و مه‌آلود از حوزه‌های مختلف می‌آیند. دو مولد وجود دارد که به‌عنوان مترجم هر دو دامنه عمل می‌کنند. علاوه بر این، آنها از تابع اتلاف سازگاری چرخه جهت دستیابی به نتایج بهتر و حل مشکل فروپاشی حالت^۳ استفاده کردند. در مقاله [۲۲]، با استفاده از توانایی‌های شبکه GAN و تبدیل موجک، شبکه WT-GAN برای بازسازی تصاویر با کیفیت از تصاویر دستگاه‌های توموگرافی کامپیوتری ارائه شد. آن‌ها توانستند با استفاده از تبدیل موجک اطلاعات ساختار و نویز را استخراج کرده و تصاویر با کیفیت بالا را توسط GAN تولید کنند. آن شبکه توانست به خوبی نویز را حذف و جزئیات بافت را حفظ نماید. شبکه‌های GAN در حذف نویزهای مختلف و بازیابی جزئیات بافت تصویر خوب عمل کردند، با این حال به دلیل ناپایداری بلوک جداکننده، جزئیات جعلی و مصنوعات غیرمنتظره در تصویر تولید می‌کنند. برای حل آن مدل HiGAN [۲۳] پیشنهاد شد. در این شبکه از سه مولد استفاده شده است. اولین مولد برای جلوگیری از اتلاف ویژگی‌های فرکانس بالا مانند لبه‌ها و بافت همراه با جداکننده آموزش می‌بیند تا توانایی خود را در حفظ جزئیات ضروری بهبود بخشد. مولد دوم بر حذف اثر ناپایداری‌های ناشی از جداکننده و بازیابی ویژگی‌های فرکانس پایین تمرکز دارد. در نهایت مولد سوم برای کمک به همکاری مؤثر آن دو مولد و افزایش عملکرد بازسازی به کار گرفته شده است. از دیگر روش‌ها، MRDGAN [۲۴] را می‌توان نام برد که برای حذف نویز در فرکانس‌های بالا و افزایش کیفیت تصاویر آژیوگرافی توموگرافی کامپیوتری ارائه شد. این معماری با استفاده از سازوکار توجه^۴، روند استخراج ویژگی در مولد را تسهیل نمود و یک بلوک باقیمانده چند مقیاس را با بلوک کاهش نویز برای حفظ

نام‌های مولد و جداکننده در یک بازی حداقل-حداکثر با یکدیگر رقابت می‌کنند.

دو معماری CycleGAN [۱۰] و DualGAN [۱۱] از روش‌های پیشگام در انتقال محتوای تصویر به تصویر هستند. آنها به صورت چرخه‌ای از مولد و جداکننده استفاده می‌کنند تا تصاویر مصنوعی را ایجاد نمایند. آنها قادر به تبدیل اطلاعات سطح پایین هستند در حالی که در انتقال معنای سطح بالای تصاویر موفق نیستند. در مقاله EGAN [۲] چالش جدید انتقال محتوای تصویر همراه با نویز مطرح شده است. نویسندگان برای بهبود توانایی بیان مدل ابداعی و یادگیری بهتر ویژگی‌های مهم تصویر، از شبکه U-Net [۱۲] بهره‌جسته‌اند. آنها از تابع اتلاف ادراک^۱ با استفاده از شبکه VGG [۱۳] که حساس به نویز است و توانایی کاهش نویز را دارد، نیز بهره‌برده‌اند. اسکندر و همکاران در [۳] به منظور مطابقت توزیع فرکانس بالای تصاویر واقعی، یک معماری مولد جدید در حوزه موجک به نام USIS ارائه دادند تا بتواند به خوبی به بخش‌بندی معنایی تصاویر پردازد. در حالی که برخی روش‌ها از کاهش کیفیت برای محتوای با فرکانس بالا رنج می‌برند، راهکار SWAGAN [۱۴] مبتنی بر تبدیل موجک برای انتقال سبک ارائه شد تا مولد را در حوزه فرکانس پیاده‌سازی کند. موجک‌ها در سرتاسر معماری‌های مولد و جداکننده ترکیب می‌شوند و یک نمایش پنهان آگاه به فرکانس^۲ را در هر مرحله ارائه می‌دهند. این روش باعث بهبود کیفیت بصری تصاویر تولیدشده و افزایش قابل توجه پیچیدگی محاسباتی می‌شود. به دلیل مقادیر بسیار زیاد پارامترها، شبکه‌های مولد تخصصی پیشرفته معمولاً به حافظه زیاد نیاز دارند و کارایی آنها پایین است. برای مقابله با این چالش، عملکرد GAN‌ها در سطوح مختلف فرکانس در مرجع [۱۵] بررسی شد. نتایج پیاده‌سازی آنها نشان می‌دهد GAN‌های کوچک، قادر به تولید اطلاعات با فرکانس و کیفیت بالا نبوده بنابراین روش تقطیر دانش موجک را پیشنهاد کرده که در این روش ابتدا تصاویر با تبدیل موجک گسسته به باندهای فرکانسی مختلف تجزیه و سپس تنها باندهای فرکانس بالا تقطیر می‌شوند. بنابراین GAN به خوبی می‌تواند باندهای فرکانس بالا را یاد بگیرد. UNIT-DDPM یک مدل احتمالی انتشار حذف نویز بدون نیاز به آموزش تهاجمی است [۱۶]. در این روش توزیع مشترک تصاویر در هر دو حوزه به‌عنوان یک زنجیره مارکوف تعریف می‌شود که با هدف به حداقل رساندن نویز شروع به استنباط در هر دو حوزه می‌کند. این روش منجر به پایداری آموزش و تولید خروجی‌های با کیفیت بالا می‌شود.

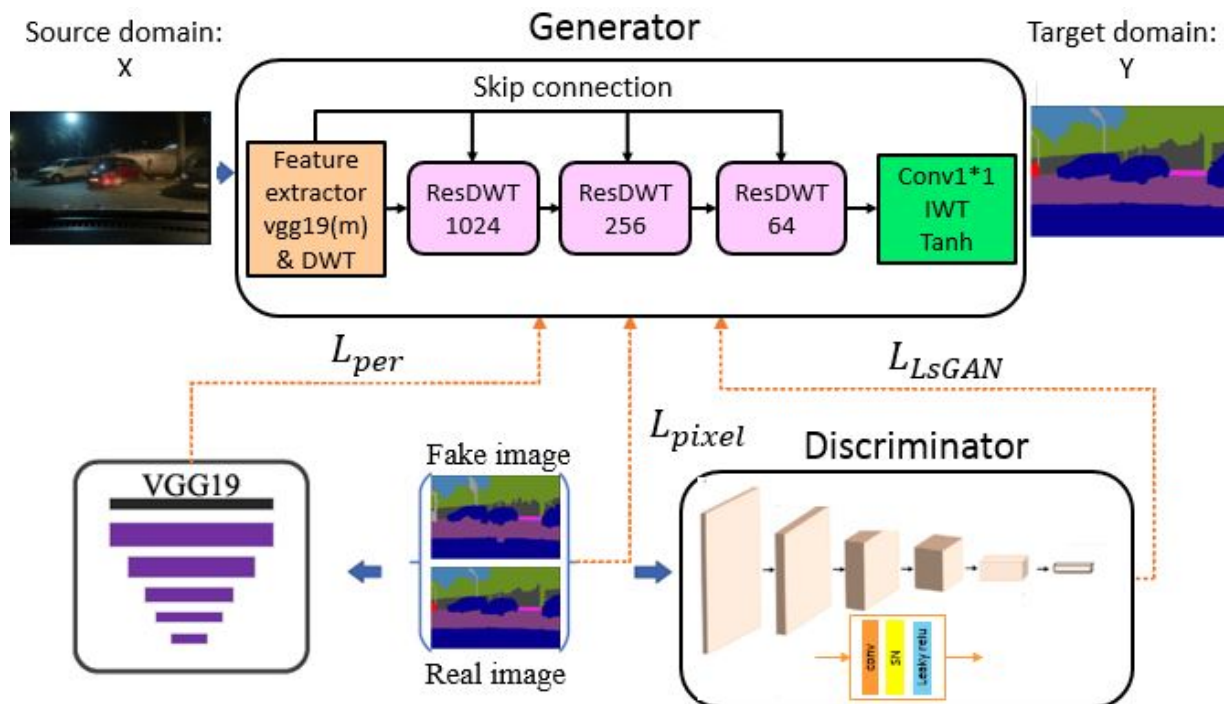
حذف نویز از رایج‌ترین مسائل بینایی کامپیوتر است که به‌عنوان مرحله پیش‌پردازش در بسیاری از کارهای پردازش تصویر استفاده می‌شود. روش‌های سنتی حذف نویز شامل تطبیق بلوک و

³ Mode collapse

⁴ Attention mechanism

¹ Perceptual loss

² frequency-aware latent representation



شکل (۱): مدل پیشنهادی WTGAN. همانطور که دیده می‌شود ورودی ما یک تصویر و خروجی آن بخش بندی تصویر است.

استفاده می‌نماید که باعث حذف کارای نویز، بهبود توانایی بیان مدل، حفظ ساختار تصویر و افزایش کیفیت می‌شود ([۲۶] تا [۲۸]).

(۳) با استفاده از بلوک‌های توجه به جزئیات ریز تصویر توجه نموده و به یادگیری بهتر ویژگی‌ها و حذف نویز کمک می‌شود. از طریق آزمایش‌های کمی و کیفی، برتری WTGAN در مقابل روش‌های دیگران نشان داده شد؛ WTGAN تصاویر واقعی‌تر و با کیفیت‌تر تولید نمود و توانست به خوبی نویز را حذف کند.

۲- روش پیشنهادی

شکل (۱) نمای کلی معماری پیشنهادی را نشان می‌دهد که از EGAN الهام گرفته شده است. هدف ما آموزش شبکه مولد و جداکننده جدید برای نگاشت تصاویر از دامنه منبع نویزی (X) به دامنه هدف تمیز (Y) با استفاده از نمونه‌های غیرمزدوج (به ازای هر تصویر در دامنه منبع نیازی به وجود تصویر متناظر در دامنه هدف نیست، این دو دامنه دارای تصاویر مجزا هستند) است.

ساختار GAN پیشنهادی دارای یک مولد (G) و یک جداکننده (D) است. مولد پیشنهادی در طراحی از تبدیل موجک بهره جسته، زیرا عملکرد خوبی در حذف نویزهای عمومی از خود نشان داده است. در ساختار درونی بلوک‌های مولد، سازوکار توجه برای ارتقاء کیفیت تصاویر تولید شده و حذف نویز بهتر تعبیه شده است. بعلاوه اتصالات باقیمانده و کنارگذر، به حفظ بهتر کیفیت تصویر و حذف نویز کمک می‌کنند. روش پیشنهادی با الهام از

اطلاعات رگ‌ها و حذف نویز در تصاویر تولید شده، ترکیب کرد. این روش به یک راه‌حل کم‌هزینه برای به دست آوردن تصاویر آنژیوگرافی توموگرافی کامپیوتری با وضوح بالا رسید.

علی‌رغم ابداع روش‌های پیشرفته در زمینه حذف نویز و انتقال محتوای تصویر به تصویر این موضوع همچنان جای کار بسیار دارد. بنابراین هدف این مقاله، انتقال محتوای یک تصویر نویزی از دامنه منبع به یک تصویر بخش‌بندی شده بدون نویز در دامنه هدف است [۲]. برای چالش مطرح شده، ابتدایی‌ترین ایده این است که ابتدا حذف نویز، سپس انتقال محتوا انجام شود. این روش خطاهای انتقال محتوا را اندازه‌گیری یا به حداقل نمی‌رساند و سبب کاهش کیفیت خواهد شد [۲]. بنابراین با رهیافتی متفاوت به بررسی این چالش پرداخته و یک مدل جدید به نام WTGAN^۱ پیشنهاد داده‌ایم. این مقاله اولین تلاش برای انتقال محتوای تصویر به تصویر با وجود نویز با استفاده از GAN مبتنی بر تبدیل موجک است. پس هدف اصلی روش پیشنهادی، نادیده گرفتن نویز و آموزش ویژگی‌های کلیدی بوده تا بخش بندی معنایی را انجام دهد. به طور خلاصه، جنبه‌های مهم این مقاله در سه بخش زیر است:

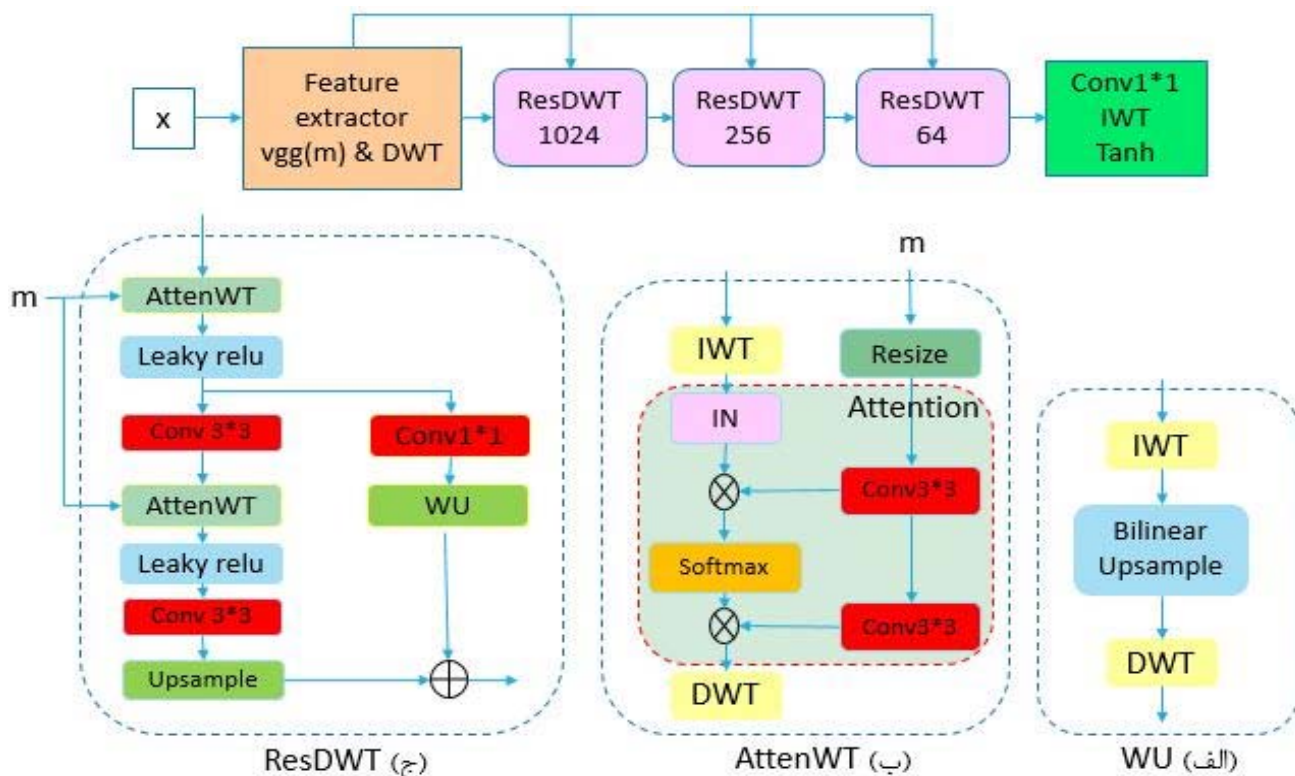
- (۱) یک روش سربه‌سر^۲ مبتنی بر GAN و موجک به نام WTGAN برای انتقال محتوای تصویر به تصویر با نویز ارائه شد.
- (۲) در روش پیشنهادی، مولد و جداکننده با استفاده از شبکه آموزش دیده مشترک به استخراج ویژگی‌های غنی‌تر می‌پردازند [۲۵]. مولد از تبدیل موجک، اتصالات باقیمانده^۳ و کنارگذر^۴

^۱ Wavelet Transform GAN

^۲ End to end

^۳ Residual

^۴ Skip-connection



شکل (۲): مولد پیشنهادی. الف) معماری WU [۳]. ب) معماری AttenWT (IN: نرمال‌سازی نمونه). ج) معماری داخلی ResDWT.

گسسته^۱ (DWT) فرکانس‌های بالای تصویر از بین نمی‌روند، علاوه بر این اجسام کوچک در حوزه مکان دارای ضرایب موجک با مقدار بزرگ هستند که با استفاده از تبدیل موجک می‌توان آن‌ها را مشاهده نمود. تبدیل موجک معکوس^۲ (IWT) نیز به‌عنوان عملیات معکوس تعریف می‌شود که دامنه موجک را به حوزه مکان نگاشت می‌دهد. DWT و IWT به ترتیب به channelwise- نگاهت می‌دهند. همانند USIS تصویر از یک سری فیلترهای پایین‌گذر و بالاگذر عبور می‌کند تا همبستگی‌های موجک‌ها را ایجاد کند. ویژگی‌های استخراج‌شده با استفاده از شبکه پیش‌آموزش دیده با اعمال یک مرحله DWT (m) به بلوک‌های ResDWT منتقل می‌شوند. ResDWT بر اساس روش USIS برای تولید ضرایب موجک به‌جای استفاده مستقیم از پیکسل‌ها طراحی شده است [۳]. در این بلوک دو پیمانانه تعریف شده است: نمونه موجک (WU) و تبدیل موجک مبتنی بر توجه (AttenWT). این معماری تمام ویژگی‌های حوزه موجک را یاد می‌گیرد.

در شکل (۲). الف) معماری WU به تصویر کشیده شده است. با توجه به شکل (۳). ج) در این معماری لایه‌های نمونه‌برداری رو به بالا به هم متصل نیستند، در عوض این لایه‌ها به داخل ResDWT منتقل می‌شوند و از روش‌های مقیاس‌گذاری

روش‌های EGAN [۲]، USIS [۳] و NiceGAN [۲۵] طراحی شده است.

۲-۱- شبکه مولد پیشنهادی

مولد پیشنهادی بر اساس تبدیل موجک و سازوکار توجه طراحی شده است. x نمونه‌ای از دامنه منبع و y نمونه‌ای از دامنه هدف را نشان می‌دهد. مولد شامل یک شبکه استخراج ویژگی و بلوک‌های ResDWT است.

در ابتدا با الهام از شبکه NiceGAN، برای استخراج ویژگی‌های غنی‌تر از تصویر ورودی x ، از شبکه پیش‌آموزش دیده استفاده شد. به‌کارگیری شبکه پیش‌آموزش دیده، علاوه بر کاهش تعداد پارامترها و زمان آموزش، سبب استخراج ویژگی‌های غنی‌تر از تصویر می‌شود. ویژگی‌های به‌دست‌آمده، نه تنها در مولد بلکه در جداکننده هم استفاده شده‌اند که این مسئله مجدداً از تعداد پارامترها کاسته و شبکه را کوچک‌تر می‌کند. از آنجایی که شبکه‌های VGG به نویز حساس هستند، از VGG19 استفاده شد که سبب کاهش اثرات نویز می‌شود.

در مولد، اتصالات باقیمانده و کنارگذر [۲۶] و [۲۷] از ناپدید شدن گرادینان جلوگیری و منجر به حفظ اطلاعات سطوح پایین در سطوح بالاتر می‌گردند. اتصالات کنارگذر از ویژگی‌های استخراج شده از تصویر ورودی نشأت می‌گیرند، بنابراین به حفظ بهتر ساختار و محتوای تصویر ورودی و حذف نویز کمک بسزایی می‌کنند. تبدیل موجک به خوبی می‌تواند به حذف نویز کمک کند، بنابراین در ساختار داخلی مولد از آن استفاده شد. با تبدیل موجک

¹ Discrete Wavelet Transform

² Inverse Wavelet Transform

³ Haar

مختلف در داخل هر شاخه استفاده می‌شود. شاخه هویت با بلوک نمونه‌برداری رو به بالا موجک نمونه‌برداری می‌شود، در حالی که شاخه باقیمانده با درون‌یابی نزدیک‌ترین همسایه نمونه‌برداری می‌شود. این کار سبب می‌شود نگاشت اصلی در اتصال میانبر حفظ شود، زیرا برای هدایت شیب‌های مناسب به لایه‌های اولیه شبکه که ویژگی‌های درشت تولید می‌کنند ضروری است. لایه WU با الهام از SWGAN [۴] از یک IWT و به دنبال آن نمونه‌برداری دوخطی و DWT تشکیل شده است. بنابراین WU با استفاده از نمونه‌برداری نزدیک‌ترین همسایه در شاخه باقیمانده ResDWT وضوح ضرایب موجک را حفظ و به حذف نویز کمک می‌کند. در ادامه از لایه AttenWT استفاده شد. از آنجایی که سازوکار توجه، تأثیر بسیار زیادی بر عملکرد روش دارد و قادر است نقاط با اهمیت را تشخیص داده و به حذف نویز از آن نقاط کمک کند، در معماری پیشنهادی به کار برده شده است. سازوکار توجه در حوزه مکان طراحی شده است از این جهت نمی‌توان در حوزه فرکانس استفاده نمود، در نتیجه از ترکیب DWT + IWT + AttenWT استفاده شد. همان‌طور که در تصویر (۲.ب)) مشاهده می‌کنید ابتدا از IWT و سپس از یک سازوکار توجه معمولی با توجه به ویژگی‌های تصویر ورودی، نتایج مرحله قبل و DWT استفاده شد. این لایه با ترکیب تبدیل موجک و توجه به حفظ کیفیت و ساختار تصویر و حذف نویز بهتر تصویر کمک نمود (هر دو ورودی ResDWT1024 همان m است).

در شکل (۲.ج)) معماری داخلی ResDWT ارائه شده است. در ابتدا با استفاده از AttenWT و در پی آن LeakyRelu به بهینه‌سازی ویژگی‌های خروجی پرداخته شد. سپس با الهام از شبکه ResNet [۲۶] با توجه به اینکه در بیشتر موارد تصویر خروجی باید ساختاری را با تصویر ورودی به اشتراک بگذارد با استفاده از اتصالات باقیمانده و لایه‌های کانولوشن 3×3 به عنوان لایه گلوگاه به یادگیری تابع هویت طراحی شده پرداخت. AttenWT به حفظ بهتر کیفیت تصویر و انتقال بهتر اطلاعات از سطوح پایین به بالا و جلوگیری از ناپدید شدن گرادیان پرداخت. علاوه بر این با استفاده از نمونه‌برداری رو به بالا در شاخه هویت و WU در شاخه باقیمانده به حفظ نگاشت تصویر و بهتر شدن نتایج خروجی کمک شد. برای دستیابی به خروجی نهایی پس از بلوک‌های متوالی ResDWT یک لایه کانولوشن 1×1 و IWT و Tanh استفاده شد.

۲-۲- شبکه جداکننده

جداکننده بخش مهمی از چارچوب است، زیرا مسئول تشخیص داده‌های واقعی و جعلی است و از یادگیری نگاشت‌های بی‌اهمیت توسط مولد جلوگیری می‌کند. برای جداکننده از Markov PatchGAN استفاده شد [۱]. این روش استقلال پیکسل‌ها را در

۲-۳- تابع اتلاف

در روش پیشنهادی سه تابع اتلاف استفاده شده است: اتلاف حداقل مربعات (L_{LSGAN}) برای مطابقت توزیع بین تصاویر تولید شده و داده‌های واقعی و اتلاف پیکسل (L_{pixel}) و اتلاف ادراک (L_{per}) برای کاهش اثرات نامطلوب نویز اعمال شد.

L_{LSGAN} [۵]: این اتلاف جریمه بیشتری به خطاهای بزرگ‌تر می‌دهد و منجر به تصحیح بزرگ به جای عدم همگرایی می‌شود. در رابطه (۱) D سعی می‌کند کل رابطه به حداقل برساند در حالی که G با هدف به حداقل رساندن $(1-D(G(x)))^2$ آموزش می‌بیند.

$$L_{LSGAN}(G,D) = E_y [(D(y))^2] + E_x [(1-D(G(x)))^2] \quad (1)$$

L_{pixel} [۲]: روش‌های قبلی نشان داده‌اند که نرم ۱ در مقایسه با نرم ۲ باعث تاری کمتر می‌شود [۱]. این اتلاف برای جلوگیری از تغییرات ناگهانی و مصنوعی تصویر به کار می‌رود (رابطه (۲)). اتلاف سعی در به حداقل رساندن اختلاف توزیع رنگ تصویر دامنه هدف و تصویر تولید شده دارد.

$L_{pixel}(G) = E_{x,y} [|y - G(x)|_1] \quad (2)$
 L_{per} [۲]: اتلاف ادراک (رابطه (۳)) برای تطبیق ویژگی‌ها در شبکه ادراک بصری معرفی شود که بر روی تصویر تولید شده و تصویر دامنه هدف اعمال می‌شود. اتلاف ادراک سبب تأکید مدل بر محتوا و نادیده گرفتن نویز می‌شود. در این اتلاف VGG19 به عنوان شبکه ادراک بصری استفاده و با متغیر Ψ نشان داده می‌شود. فرا پارامتر α_k (K نشان دهنده لایه‌های شبکه VGG19 است) سهم لایه مربوطه را در اتلاف متعادل می‌کند. همانند EGAN، فرا پارامتر α_k برابر یک است.

$L_{per}(G) = E_{x,y} [\sum_k \alpha_k |\Psi_k(y) - \Psi_k(G(x))|_1] \quad (3)$
 در نهایت مولد و جداکننده به‌طور مشترک با استفاده از رابطه (۴) برای بهینه‌سازی هدف نهایی آموزش داده شدند.

$G^* = \arg \min_G \max_D (L_{LSGAN}(G,D) + \lambda_1 L_{pixel}(G) + \lambda_2 L_{per}(G)) \quad (4)$

در رابطه (۴) λ_1 و λ_2 فرا پارامترهای کنترل اهمیت اتلاف پیکسل و ادراک است که در این مقاله $\lambda_1=100$ و $\lambda_2=1$ است (مقادیر حاصل با استفاده از 10-fold به دست آمده است).

¹ Patch

۳- نتایج پیاده‌سازی

در این بخش، آزمایش‌های گسترده‌ای برای ارزیابی عملکرد WTGAN پیشنهادی انجام شد. ابتدا مجموعه داده‌ها و معیارهای ارزیابی معرفی شدند سپس، جزئیات پیاده‌سازی بیان و آزمایش‌های ارزیابی کمی و کیفی ارائه شد.

۳-۱- مجموعه داده‌ها و معیارهای ارزیابی

درک بصری صحنه‌های پیچیده خیابان ویژگی توانمند برای طیف وسیعی از کاربردهاست. بنابراین در این مقاله آزمایش‌هایی بر روی مجموعه داده Cityscapes که مجموعه‌ای بزرگ و کامل از تصاویر شهری در زمینه بخش‌بندی معنایی و انتقال محتوای تصویر به تصویر شامل ۲۹۷۵ نمونه برای آموزش به همراه ۵۰۰ نمونه برای آزمایش است، انجام شد [۲۹]. تمام نمونه‌های تصویر در هر دو حوزه تصویر تمیز و مزدوج هستند. با توجه به اینکه هدف انتقال محتوای تصویر به تصویر با نویز است، باید در آزمایش‌ها به صورت مصنوعی به تصویر ورودی نویز اضافه شود. بنابراین، برای شبیه‌سازی نویز واقعی در تصاویر طبیعی، نویزهای گاوسی مستقل و یکسان با میانگین صفر و انحراف معیار پنج همانند روش EGAN به تصاویر اصلی اضافه شد تا به عنوان تصاویر ورودی نویزی در نظر گرفته شوند [۲].

معیارهای ارزیابی معروف حداکثر نسبت سیگنال به نویز^۱ (PSNR)، شاخص تشابه ساختاری^۲ (SSIM) و شباهت تکه‌ای تصویر ادراکی^۳ (LPIPS) به عنوان شاخص‌های ارزیابی عملکرد انتخاب شدند. هرچه مقادیر PSNR و SSIM بزرگ‌تر باشد، دقت پیکسلی الگوریتم بالاتر خواهد بود.

معیار LPIPS برای قضاوت در مورد شباهت ادراکی بین دو تصویر استفاده شده است. LPIPS با استفاده از نرم ۱ بین ویژگی‌های استخراج شده از AlexNet پیش آموزش دیده بین تصویر ورودی ($P_{Alex}(x)$) و تصویر تولید شده ($P_{Alex}(y)$) محاسبه می‌شود. نمره LPIPS می‌تواند برای اندازه‌گیری تنوع روش‌های چندوجهی استفاده شود. علاوه بر این LPIPS معیاری از تشابه بین دو تصویر است و به خوبی با ادراک انسان مطابقت دارد. LPIPS کمتر به معنای کیفیت ادراکی بهتر و شباهت بیشتر تکه‌های تصویر است [۳۰].

۳-۲- جزئیات پیاده‌سازی

همه مدل‌ها با استفاده از Adam با مقادیر $\beta_1=0.5$ و $\beta_2=0.999$ آموزش دیده‌اند. تصاویر با احتمال ۰,۵ به صورت افقی چرخانده، اندازه آن‌ها به 286×286 تغییر داده شده و به طور تصادفی به اندازه

256×256 برش زده شدند. اندازه دسته برای همه آزمایش‌ها بر روی یک تنظیم شد. روش با بسته Tensorflow پیاده‌سازی و بر روی کارت گرافیک Google Colab آموزش داده شد.

۳-۳- ارزیابی کمی و کیفی

روش پیشنهادی با روش‌های جدید EGAN، USIS و Pix2Pix مقایسه شد. برای مقایسه از پیاده‌سازی‌های رسمی آن‌ها که در دسترس عموم است استفاده شد. در ادامه، مدل با استفاده از معیارهای ارزیابی کمی ارزیابی شد.

نتایج ارزیابی کمی در جدول (۱) نشان داده شده است. نتایج نشان دهنده برتری مدل پیشنهادی در مقایسه با روش‌های پایه است. هر چه LPIPS کوچک‌تر باشد، نتایج به تصاویر طبیعی نزدیک‌تر است.

جدول (۱) مقایسه روش پیشنهادی با دیگر روش‌ها

معیارها	[۱] Pix2pix	[۲] EGAN	[۳] USIS	WTGAN پیشنهادی
PSNR	۲۷,۲۴	۲۷,۹۶	۲۸,۶۴	۲۹,۱۳
SSIM	۰,۸۲۱۴	۰,۸۳۰۹	۰,۸۲۹۴	۰,۸۶۰۲
LPIPS	۰,۱۴۳۲	۰,۱۳۱۹	۰,۱۲۹۶	۰,۱۲۸۴

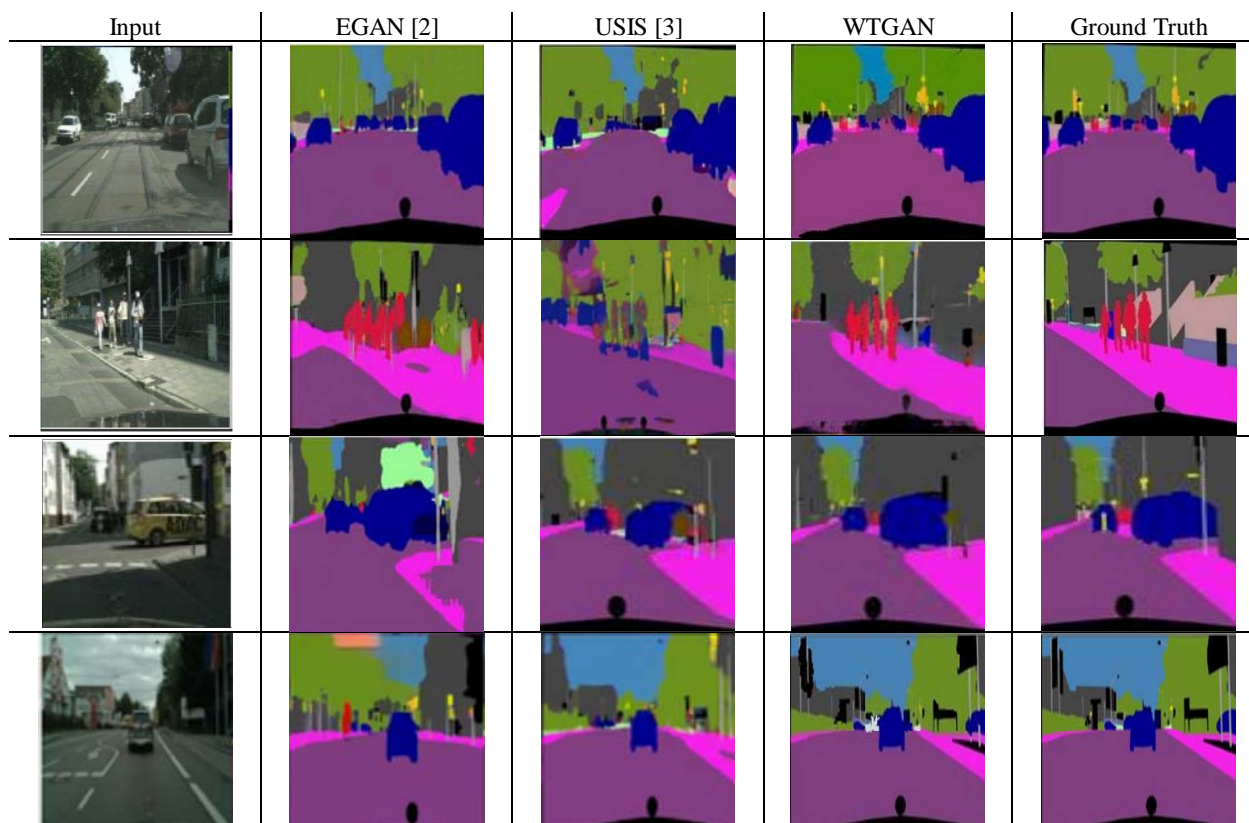
روش USIS به طور کلی عملکرد بهتری از EGAN نشان می‌دهد و در مدیریت انتقال محتوای تصویر به تصویر با وظیفه حذف نویز در شرایط پیچیده نور و در سبک‌های غنی‌تر تصویری بهتر عمل می‌کند. با مقایسه روش پیشنهادی با دیگر روش‌ها به این نتیجه رسیدیم که توانسته بهتر از آن‌ها عمل کند. WTGAN به خوبی می‌تواند نویز تصویر ورودی را حذف و محتوا را به خوبی به دامنه هدف انتقال دهد.

تصاویر تولید شده توسط روش پیشنهادی و دیگر روش‌ها در شکل (۳) نمایش داده شده است. تصاویر ورودی با اضافه نمودن نویز گاوسی با میانگین صفر و انحراف معیار ۵ بدست آمده‌اند. همان‌طور که مشاهده می‌شود الگوریتم پیشنهادی نسبت به دیگران به خوبی توانسته بخش‌بندی معنایی انجام دهد. علاوه بر این توانسته به خوبی نویز را نادیده گرفته و سبک و محتوا را حفظ نماید. با توجه به تصویر ورودی نویزی نتیجه حاصل از روش EGAN دارای اختلاف بیشتری نسبت به دیگران با تصویر بخش‌بندی هدف است. با این حال، USIS توانست به نتایج بهتری دست یابد. روش پیشنهادی نسبت به دو روش دیگر به نتیجه نزدیک‌تر به تصویر بخش‌بندی هدف دست یافت. این‌ها نشان دهنده برتری روش ما نسبت به سایر روش‌هاست و می‌تواند اثبات کند روش پیشنهادی قادر است اثرات نویز را نادیده گرفته و به بخش‌بندی معنایی مطلوب دست یابد.

¹ Peak signal to noise ratio

² Structural similarity index

³ Learned perceptual image patch similarity



شکل (۳): تصاویر تولیدشده توسط روش پیشنهادی و دیگر روش‌ها (تصاویر ورودی دارای سطح نویز $\sigma = 5$ و میانگین $\sigma = 0$ هستند).

۴- نتیجه‌گیری

در این مقاله روش WTGAN برای انتقال محتوای تصویر به تصویر در حضور نویز پیشنهاد شده است. این روش برای یادگیری از فرآیند آموزش غیر مزدوج بهره می‌گیرد. همچنین روش ارائه شده از معماری متخاصم طراحی شده با بلوک‌های ResDWT برای حذف نویز و بهبود توانایی سازماندهی محتوای مولد استفاده می‌کند. به کارگیری شبکه از پیش آموزش دیده VGG19 به کاهش پارامتر، زمان آموزش و استخراج ویژگی‌های غنی‌تر از تصویر انجامید. در ادامه، استفاده از سازوکار توجه، اتصالات باقیمانده و کنارگذر به بهبود کیفیت تصویر کمک بسزایی نمود و در حذف نویز از تصویر تأثیر بسیاری داشت. WTGAN با استفاده از آزمایش‌های کمی و کیفی ارزیابی شد. نتایج نشان داد روش پیشنهادی نسبت به مقاله پایه بهتر عمل می‌کند. اگر چه در این مسیر به نتایج خوبی دست یافتیم اما مسیر طولانی تا به بلوغ رسیدن روش در این زمینه و ارتقاء کیفیت وجود دارد.

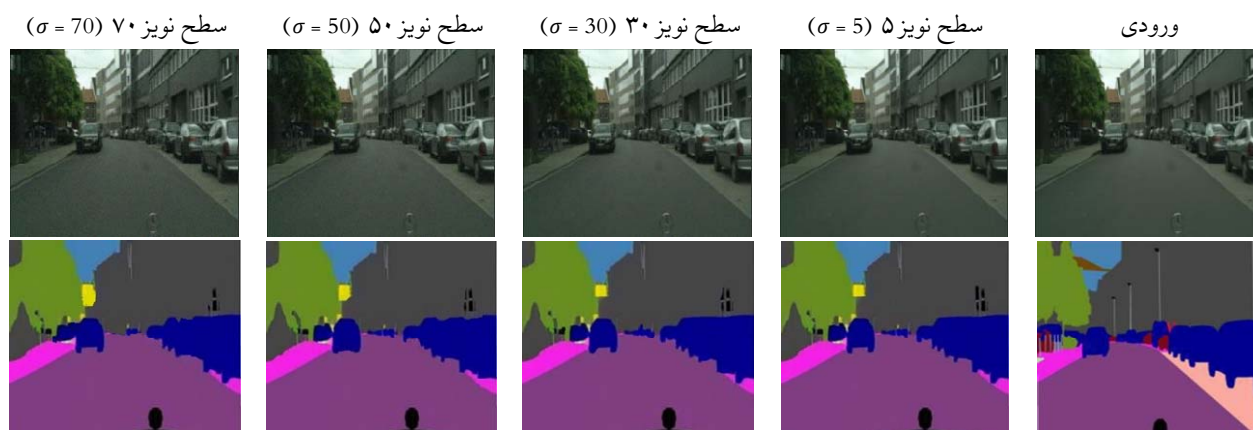
مراجع

- [1] Isola, Phillip, J. Zhu, T. Zhou, A. Efros, "Image-to-image translation with conditional adversarial networks.", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134, 2017.

در جدول (۲) نتایج حاصل از مدل پیشنهادی و دیگر روش‌ها با سطوح نویز متفاوت ارائه شده است. علاوه بر این به صورت بصری اثربخشی روش پیشنهادی در سطوح مختلف نویز مورد بررسی قرار گرفت. شکل (۴) نتایج بخش‌بندی معنایی در سطوح نویز مختلف را نشان می‌دهد. زمانی که سطح نویز بسیار بالا است تأثیر زیادی بر روی بخش‌بندی معنایی ندارد. با جمع‌بندی نتایج جدول (۲) و شکل (۴) در می‌یابیم مدل ما توانست بخش‌بندی معنایی را در تصویر ورودی نویزی اعمال کند و به نتایج بهتری دست یابد. علاوه بر این با بالا رفتن سطح نویز همچنان روش پیشنهادی عملکرد بهتری از خود نشان می‌دهد. این بدین معناست که روش ما در صورت گسترش نویز نیز مؤثر است.

جدول (۲) میانگین نمره PSNR و SSIM در روش‌های مختلف با واریانس‌های مختلف.

$\sigma = 70$	$\sigma = 50$	$\sigma = 30$	$\sigma = 5$	سطح نویز	
۲۱,۳۶	۲۳,۸۲	۲۵,۵۶	۲۷,۹۶	PSNR	[۲] EGAN
۰,۵۰۴۵	۰,۶۰۳۴	۰,۷۵۰۶	۰,۸۳۰۹	SSIM	
۲۲,۱۴	۲۳,۹۱	۲۵,۷۸	۲۸,۶۴	PSNR	[۳] USIS
۰,۵۳۴۱	۰,۶۵۲۴	۰,۷۷۴۷	۰,۸۲۹۴	SSIM	
۲۲,۳۸	۲۴,۹۲	۲۵,۹۲	۲۹,۱۳	PSNR	WTGAN پیشنهادی
۰,۵۵۷۶	۰,۶۸۲۳	۰,۷۹۱۲	۰,۸۶۰۲	SSIM	



شکل (۴): تصاویر تولیدشده توسط روش پیشنهادی با سطوح نویز گوسی مختلف (در تمامی سطوح میانگین برابر ۰ است). تصویر بخش بندی شده معنایی مرجع پایین تصویر ورودی نشان دهنده است.

- [11] Yi, Zili, H. Zhang, P. Tan, M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation.", In Proceedings of the IEEE international conference on computer vision, pp. 2849-2857, 2017.
- [12] Ronneberger, Olaf, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation.", In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241, 2015.
- [13] Simonyan, Karen, A. Zisserman, "Very deep convolutional networks for large-scale image recognition.", arXiv preprint arXiv:1409.1556, 2014.
- [14] Gal, Rinon, D. C. Hochberg, A. Bermano, D. Cohen-Or, "SWAGAN: A style-based wavelet-driven generative model.", ACM Transactions on Graphics (TOG) 40, vo. 4, pp. 1-11, 2021.
- [15] Zhang, Linfeng, X. Chen, X. Tu, P. Wan, N. Xu, K. Ma, "Wavelet knowledge distillation: Towards efficient image-to-image translation.", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12464-12474, 2022.
- [16] Sasaki, Hiroshi, C. G. Willcocks, T. P. Breckon, "Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models.", arXiv preprint arXiv:2104.05358, 2021.
- [17] Aharon, Michal, M. Elad, A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.", IEEE Transactions on signal processing 54, vo. 11, pp. 4311-4322, 2006.
- [18] Zhang, Lei, W. Dong, D. Zhang, G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping.", Pattern recognition 43, vo. 4, pp. 1531-1549, 2010.
- [19] Elad, Michael, M Aharon, "Image denoising via sparse and redundant representations over learned dictionaries.", IEEE Transactions on Image processing 15, pp. 3736-3745, 2006.
- [20] Zhang, Kai, W. Zuo, Y. Chen, D. Meng, L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn
- [2] Yan, Lan, W. Zheng, F. Wang, C. Gou, "Joint image-to-image translation with denoising using enhanced generative adversarial networks.", Signal Processing: Image Communication, vo. 91, pp. 116072, 2021.
- [3] Eskandar, George, M. Abdelsamad, K. Armanious, B. Yang, "USIS: Unsupervised Semantic Image Synthesis.", arXiv preprint arXiv:2109.14715, 2021.
- [4] Hong, Lin, X. Wang, Z. Xiao, G. Zhang, J. Liu, "WSUIE: Weakly supervised underwater image enhancement for improved visual perception.", IEEE Robotics and Automation Letters 6, vo. 4, pp. 8237-8244, 2021.
- [5] Mao, Xudong, Q. Li, H. Xie, R. YK Lau, Z. Wang, S. P. Smolley, "Least squares generative adversarial networks.", In Proceedings of the IEEE international conference on computer vision, pp. 2794-2802, 2017.
- [6] Tang, Hao, D. Xu, N. Sebe, Y. Yan. "Attention-guided generative adversarial networks for unsupervised image-to-image translation.", In 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2019.
- [7] Gokay, Dilara, E. Simsar, E. Atici, A. Ahmetoglu, A. E. Yuksel, P. Yanardag, "Graph2Pix: A Graph-Based Image to Image Translation Framework.", In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2001-2010, 2021.
- [8] Chen, Chen, Q. Chen, J. Xu, V. Koltun, "Learning to see in the dark.", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3291-3300, 2018.
- [9] Goodfellow, Ian, J. Pouget-Abadie, M. i Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial networks.", Communications of the ACM 63, vo. 11, pp. 139-144, 2020.
- [10] Zhu, Jun-Yan, T. Park, P. Isola, A. A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks.", In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232, 2017.



فرزانه مقصودی قمبوانی کارشناسی ارشد خود را در گرایش هوش مصنوعی و با پژوهش در زمینه فراکتال‌ها و آشوب تحقیقاتی در دانشگاه سمنان گذراند. در حال حاضر ایشان دانشجوی مقطع دکترای دانشگاه سمنان در زمینه شبکه‌های مولد تخصصی و انتقال محتوای تصویر به تصویر است.



محمدجواد فدائی اسلام استادیار گروه مهندسی کامپیوتر دانشگاه سمنان می‌باشد. ایشان دوره کارشناسی مهندسی کامپیوتر-سخت‌افزار، کارشناسی ارشد و دکترای خود را در دانشگاه علم و صنعت ایران گذرانده است. زمینه‌های مورد علاقه ایشان، یادگیری عمیق در حوزه پردازش متن و تصویر است.



فرزین یغمایی دوره کارشناسی کامپیوتر را در دانشگاه صنعتی امیرکبیر، کارشناسی ارشد و دکترای خود را در رشته مهندسی کامپیوتر-هوش مصنوعی در دانشگاه صنعتی شریف به پایان رسانده است و در حال حاضر دانشیار گروه مهندسی کامپیوتر دانشگاه سمنان می‌باشد. زمینه‌های مورد علاقه ایشان پردازش تصویر و ویدئو، بازشناسی الگو و متن‌کاوی است.

- for image denoising.", IEEE transactions on image processing 26, vo. 7, pp. 3142-3155, 2017.
- [21] Shi, Changyou, J. Lu, Q. Sun, S. Cheng, X. Feng, Wei Huang, "Image Dehazing Via Cycle Generative Adversarial Network.", In 2021 3rd International Conference on Advanced Information Science and System (AISS), pp. 1-7, 2021.
- [22] Su, Wanqi, Y. Qu, C. Deng, Y. Wang, F. Zheng, Z. Chen, "Enhance generative adversarial networks by wavelet transform to denoise low-dose CT images.", In 2020 IEEE International Conference on Image Processing (ICIP), pp. 350-354, 2020.
- [23] Vo, Duc My, D. M. Nguyen, T. P. Le, S. Lee, "HI-GAN: A hierarchical generative adversarial network for blind denoising of real photographs.", Information Sciences, vo. 570, pp. 225-240, 2021.
- [24] Li, Pengcheng, Z. Li, X. Pang, H. Wang, W. Lin, W. Wu, "Multi-scale residual denoising GAN model for producing super-resolution CTA images.", Journal of Ambient Intelligence and Humanized Computing 13, vo. 3, pp. 1515-1524, 2022.
- [25] Chen, Runfa, W. Huang, B. Huang, F. Sun, B. Fang, "Reusing discriminators for encoding: Towards unsupervised image-to-image translation.", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8168-8177, 2020.
- [26] He, Kaiming, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition.", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
- [27] Amirkolaei, H. Amini, D. O. Bokov, H. Sharma, "Development of a GAN architecture based on integrating global and local information for paired and unpaired medical image translation.", Expert Systems with Applications 203, pp. 117421, 2022.
- [28] Abu-Srhan, Alaa, I. Almallahi, M. AM. Abushariah, W. Mahafza, O. S. Al-Kadi, "Paired-unpaired Unsupervised Attention Guided GAN with transfer learning for bidirectional brain MR-CT synthesis.", Computers in Biology and Medicine 136, pp. 104763, 2021.
- [29] Cordts, Marius, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, "The cityscapes dataset for semantic urban scene understanding.", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213-3223, 2016.
- [30] Zhang, Richard, P. Isola, A. A. Efros, E. Shechtman, O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric.", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586-595, 2018.