

بررسی تأثیر طیف رنگی تصاویر ورودی در فرآیند افزایش کیفیت تصاویر با استفاده از یک روش جدید بر پایه شبکه‌های مولد متخاصم

علیرضا الهی خالدي^۱؛ سعید جمشیدی^{۲*}؛ محسن مسیحی^۳

۱- دانشکده مهندسی شیمی و نفت، دانشگاه صنعتی شریف

۲- دانشیار، دانشکده مهندسی و شیمی و نفت، دانشگاه صنعتی شریف (jamshidi@sharif.edu)

۳- استاد، دانشکده مهندسی و شیمی و نفت، دانشگاه صنعتی شریف

چکیده

بازسازی تصاویر با وضوح بالا از تصاویر بی کیفیت، یکی از چالش‌های اساسی در پردازش تصویر است. در مدل‌سازی محیط متخلخل با استفاده از تصاویر بدست آمده از میکروتوموگرافی کامپیوتری، به دلایل متفاوتی از جمله هزینه‌بر بودن و پیچیدگی محاسباتی استفاده از تصاویر با کیفیت، معمولاً تصاویر با کیفیتی از این محیط‌ها در دسترس نیست. با توجه به پیشرفت توسعه مدل‌هایی از جمله شبکه‌های چندمقیاسی امکان استفاده از جزئیات بیشتر تصاویر در بخش‌هایی از مدل فراهم شده و نیاز به تصاویر با کیفیت بیش از پیش احساس می‌شود. امروزه شبکه‌های مولد متخاصم به عنوان ابزاری کاربردی در افزایش کیفیت تصاویر استفاده می‌شوند. این شبکه‌ها با استفاده از جفت تصاویر با کیفیت و بی کیفیت آموزش دیده و در ادامه با ورودی گرفتن تصاویر بی کیفیت، تصاویر با کیفیت را تولید می‌کنند. تصاویر نهایی مورد استفاده در مدل‌سازی شبکه حفرات تصاویر باینری بوده ولی در صورت استفاده از تصاویر باینری به عنوان ورودی فرآیند افزایش کیفیت تصاویر ممکن است منجر به حذف جزئیات شود. بدین منظور در این پژوهش تأثیر استفاده از تصاویر خاکستری یا باینری به عنوان ورودی در فرآیند آموزش مدل بررسی می‌شود. تصاویر خاکستری یک نمونه ماسه‌سنگ Brea به عنوان ورودی اصلی در نظر گرفته شده و تصاویر باینری با استفاده از تصاویر خاکستری و تحقق تخلخل واقعی نمونه سنگ تولید می‌شوند. در این پژوهش از یک مدل جدید با عنوان RealESRGAN جهت افزایش کیفیت تصاویر استفاده می‌شود.

واژگان کلیدی: شبکه مولد متخاصم، مدل RealESRGAN، تصویر خاکستری، تصویر باینری، مدل‌سازی محیط متخلخل

مقدمه

خواص محیط متخلخل به طور قابل توجهی به ریزساختارها بستگی دارد. با دسترسی به نمایش‌های سه‌بعدی دقیق آن‌ها و مدل‌سازی محیط متخلخل می‌توان پدیده‌های مختلفی را در این محیط شبیه‌سازی کرد. مدل‌سازی مقیاس منافذ به طور گسترده به عنوان بستری برای مطالعه جریان چندفازی در مهندسی نفت، هیدرولوژی و مهندسی محیط زیست کاربرد دارد [۱]. فرآیند مدل‌سازی شبکه حفرات در محیط‌های متخلخل یکی از ابزارهای اساسی در تحلیل و شبیه‌سازی جریان سیالات در مخازن نفت و گاز و دیگر سیستم‌های متخلخل است. این فرآیند با هدف شناسایی ویژگی‌های ساختاری پیچیده این محیط‌ها انجام می‌شود و روش‌های مختلفی برای پیاده‌سازی آن وجود دارد از جمله مدل‌سازی مستقیم با استفاده از تصاویر میکروتوموگرافی کامپیوتری^۱ یا میکروسکوپ الکترونی رویشی^۲، مدل‌سازی با استفاده از شبیه‌سازی

¹ Micro-Computed Tomography (Micro-CT)

² Scanning Electron Microscopy (SEM)

ساختارهای ساده شده و مدل شبکه حفرات^۳، مدل‌های آماری و تحلیلی و روش‌های ترکیبی. در روش میکروتوموگرافی با استفاده از تصاویر با وضوح بالا شبکه‌ای از واکسل‌ها^۴ که نمایانگر حفرات و گلوگاه‌ها هستند، ایجاد می‌شود تا ساختار سه‌بعدی محیط متخلخل بازسازی شود. این روش دقت بالایی دارد اما به دلیل نیاز به تجهیزات گران‌قیمت و پردازش سنگین داده‌ها، هزینه‌بر است [۲].

در روش دوم محیط متخلخل به شبکه‌ای از گره‌ها (نمایانگر حفرات) و یال‌ها (نمایانگر گلوگاه‌ها) ساده می‌شود. این مدل، ویژگی‌هایی نظیر اندازه حفرات، ابعاد گلوگاه‌ها، و توپولوژی شبکه را با دقت معقولی بازسازی می‌کند و در کاربردهایی مانند شبیه‌سازی جریان سیال و تحلیل فشار موئینگی مورد استفاده قرار می‌گیرد. این روش به دلیل سرعت و انعطاف‌پذیری، کاربرد گسترده‌ای در پژوهش‌های مهندسی نفت دارد [۱]. مدل‌های آماری از داده‌های آماری یا معادلات تحلیلی برای تخمین خواص محیط متخلخل استفاده می‌کنند. به عنوان مثال معادله کوزنی-کارمن^۵ ساختار محیط متخلخل را به مجموعه‌ای از لوله‌های موئینه ساده شده تبدیل می‌کند. این مدل‌ها به دلیل سرعت محاسباتی بالا مناسب‌اند اما جزئیات هندسی دقیق تر محیط و ناهماهنگی‌های پیچیده را به خوبی نشان نمی‌دهند [۳].

همانطور که اشاره شد یکی از روش‌های مدل‌سازی محیط متخلخل استفاده از روش‌های ترکیبی می‌باشد که یکی از انواع پرکاربرد این مدل‌ها استخراج ویژگی‌های هندسی اولیه از تصاویر میکروتوموگرافی و استفاده از مدل شبکه حفرات برای شبیه‌سازی جریان می‌باشد [۲]. به‌طور کلی، انتخاب روش مناسب برای مدل‌سازی شبکه حفرات به عواملی نظیر مقیاس محیط، دقت مورد نظر، هزینه محاسباتی و دسترسی به داده‌های ورودی بستگی دارد. پژوهش‌های اخیر نشان داده‌اند که ترکیب روش‌های مختلف می‌تواند به بهبود کیفیت نتایج و کاهش هزینه‌ها کمک کند [۴].

محدودیت کیفیت تصاویر محیط متخلخل ناشی از چالش‌هایی مانند دسترسی محدود به ابزارهای پیشرفته، پیچیدگی نمونه‌ها و محدودیت‌های پردازشی، مانع از نمایش جزئیات دقیق در مدل شبکه حفرات می‌شود. نتایج آزمایشگاهی نشان داده است که مدل‌های شبکه حفرات همیشه با نتایج تجربی مطابقت ندارند، به‌خصوص در مواردی که پیچیدگی محیط زیاد است یا ساختار مدل ساده‌سازی شده است [۵]. وجود این چالش‌ها نیاز به بهبود کیفیت تصاویر و بازیابی جزئیات از دست رفته در تصاویر را نشان می‌دهد. بهبود کیفیت معمولاً شامل کاهش اختلال^۶، افزایش وضوح^۷ و شناسایی ویژگی‌ها^۸ است [۶]. روش‌های سنتی مثل فیلتر کردن و شناسایی لبه‌ها معمولاً در حفظ جزئیات به ویژه در محیط‌های پیچیده ناتوانند [۷].

در سال‌های اخیر، پیشرفت‌های شگرفی در حوزه بهبود کیفیت تصاویر به کمک شبکه‌های عصبی عمیق صورت گرفته است. این تحولات از مدل‌های ساده شبکه‌های عصبی آغاز شد و به توسعه روش‌های پیشرفته‌تر مانند شبکه‌های مولد متخاصم^۹ انجامید که نقش مهمی در تغییر مسیر این حوزه داشتند. شبکه‌های عصبی اولیه با تمرکز بر مدل‌های چندلایه، عملکردی محدود داشتند و تنها قادر به انجام وظایف ابتدایی مانند حذف نویز یا افزایش وضوح بودند [۸]. با معرفی شبکه‌های عصبی کانولوشنی^{۱۰}، این روند تکامل یافت و مدل‌هایی مانند VGG و ResNet توانستند ساختارهای

³ Pore Network Model (PNM)

⁴ Voxel

⁵ Kozeny-Carman

⁶ Noise Reduction

⁷ Resolution Enhancement

⁸ Feature Detection

⁹ Micro-Computed Tomography (Micro-CT)

¹⁰ Convolutional Neural Networks (CNN)

پیچیده‌تری را تحلیل و اطلاعات محلی تصاویر را استخراج کنند، اما این روش‌ها همچنان محدودیت‌هایی در بازسازی جزئیات دقیق تصاویر داشتند [۹].

شبکه‌های مولد متخاصم توسط گودفیلو و همکاران در سال ۲۰۱۴ معرفی شدند [۱۰]. این مدل‌ها از دو بخش مولد^{۱۱} و تفکیک‌کننده^{۱۲} تشکیل شده‌اند که با رقابت با یکدیگر، کیفیت تولید تصاویر را بهبود می‌بخشند. شبکه مولد وظیفه تولید داده‌هایی مشابه داده‌های واقعی را دارد و تلاش می‌کند داده‌های جعلی‌ای تولید کند که از داده‌های واقعی غیرقابل تشخیص باشند. تفکیک‌کننده یک شبکه طبقه‌بند است که تلاش می‌کند داده‌های واقعی را از داده‌های تولیدشده توسط مولد تمایز دهد. این دو شبکه در فرآیندی رقابتی آموزش می‌بینند، به طوری که مولد بهبود می‌یابد تا تفکیک‌کننده را فریب دهد، و تفکیک‌کننده تلاش می‌کند دقیق‌تر عمل کند [۱۰]. از نخستین کاربردهای موفق شبکه مولد متخاصم در بهبود وضوح تصاویر، مدل شبکه مولد متخاصم ابروضوح^{۱۳} (SRGAN) بود که توانست تصاویری با وضوح پایین را به نسخه‌هایی با وضوح بالا و جزئیات دقیق تبدیل کند [۱۱]. مدل SRGAN با استفاده از معماری عمیق و یک تابع زیان پرسپترون^{۱۴}، تغییرات قابل توجهی در بازایی جزئیات تصاویر ایجاد کرد و به نقطه عطفی در این زمینه تبدیل شد.

در ادامه، مدل شبکه مولد متخاصم ابروضوح پیشرفته^{۱۵} (ESRGAN) به عنوان نسخه پیشرفته‌تری از SRGAN معرفی شد. این مدل با بهره‌گیری از بلوک‌های متراکم^{۱۶}، دقت بیشتری در بازسازی جزئیات تصاویر ارائه داد و توانست کیفیت بصری تصاویر را تا حد زیادی افزایش دهد. ESRGAN در مسابقات معتبر، مانند NTIRE 2018، عملکرد برجسته‌ای از خود نشان داد و استاندارد جدیدی در بازسازی تصاویر با وضوح بالا تعیین کرد [۱۲]. مدل شبکه مولد متخاصم با وضوح فوق‌العاده پیشرفته در شرایط واقعی (RealESRGAN) به عنوان گام بعدی در این مسیر، با تمرکز بر بهبود کیفیت تصاویر در شرایط واقعی، مانند تصاویر دارای نویز یا اعوجاج شدید، توسعه یافت. این مدل با استفاده از یک فرآیند تخریب مرتبه بالاتر، تخریب‌های پیچیده دنیای واقعی مانند نویزهای ناشناخته و اعوجاج‌های غیرمعمول را شبیه‌سازی می‌کند [۱۳]. این قابلیت‌ها می‌توانند برای بازسازی و بهبود کیفیت تصاویر میکروسکوپی از ساختارهای متخلخل سنگ‌ها و شبیه‌سازی جریان در مخازن نفتی بسیار کمک کننده باشند.

در این پژوهش از مدل RealESRGAN برای افزایش کیفیت تصاویر استفاده می‌شود. یکی از مسائلی که در پژوهش‌های اخیر به آن پرداخته نشده تاثیر نوع تصاویر ورودی به مدل می‌باشد. مقالات متعددی از روش‌های مختلفی برای افزایش کیفی تصاویر استفاده کرده‌اند و هر یک بر اساس نوع داده‌های در دسترس یکی از انواع تصاویر خاکستری^{۱۷} یا باینری^{۱۸} را به عنوان ورودی مدل خود در نظر گرفته‌اند. شمس و همکاران با استفاده هم‌زمان از شبکه مولد متخاصم و رمزگذار خودکار^{۱۹}، ساختار سه‌بعدی تخلخل بین دانه‌ای و درون دانه‌ای موجود در نمونه تصاویر باینری ماسه‌سنگی و کربناته را بازسازی کردند [۱۴]. رمزگذار خودکار برای اضافه کردن تخلخل درون دانه‌ای به داده‌های خروجی شبکه مولد متخاصم به کار گرفته شد. آن‌ها از تصاویر ماسه‌سنگ Brea برای تایید عملکرد شبکه مولد متخاصم استفاده کردند و نمونه‌های

¹¹ Generator

¹² Discriminator

¹³ Super Resolution Generative Adversarial Network

¹⁴ Perceptron Loss Function

¹⁵ Enhanced Super Resolution Generative Adversarial Network

¹⁶ Residual-in-Residual Dense Block (RRDB)

¹⁷ Binary

¹⁸ Grayscale

¹⁹ Auto Encoder (AE)

ماسه‌سنگی و کربناته از میداین ایران را جهت انجام بازسازی استفاده نمودند.

شمس و همکاران یک روش ترکیبی از رویکرد آماری و شبکه‌های عصبی مولد متخاصم شرطی^{۲۰} (ST-CGAN) برای بازبازی محیط متخلخل سه‌بعدی از تصاویر باینری دوبعدی معرفی کردند [۱۵]. روش آماری داده‌های آماری پیش‌پردازش شده را برای کمک به یادگیری عمیق تولید می‌کند و شبکه مولد متخاصم شرطی تصاویر دو بعدی اولیه را به ساختارهای سه‌بعدی واقعی‌تر تبدیل می‌کند. نمونه ماسه‌سنگی Brea و سنگ کربناته Estailades برای آموزش و ارزیابی مدل استفاده شده‌اند. لیانگ و همکاران از یک مدل سه‌بعدی گام به گام بر پایه شبکه مولد متخاصم استفاده نمودند و تصاویر باینری یک نمونه سنگ شیل با کیفیت پایین را به عنوان ورودی مدل در نظر گرفتند [۱۶]. این مدل توانست شبیه‌سازی‌های سه‌بعدی را بهبود ببخشد و اطلاعات دقیق‌تری برای تحلیل‌های لازم فراهم کند. لی و همکاران از شبکه مولد متخاصم پیش‌رونده آبشاری^{۲۱} (CPGAN) استفاده کردند. آن‌ها در این کار از تصاویر خاکستری با هدف افزایش کیفیت استفاده نمودند [۱۷]. در پژوهش‌های معرفی شده اشاره‌ای به دلیل استفاده از نوع تصویر باینری یا خاکستری نشده است، بنابراین بررسی اثر جنس تصویر ورودی می‌تواند اطلاعات مفیدی در اختیار ما قرار دهد.

روش تحقیق

روش‌های مبتنی بر شبکه‌های مولد متخاصم برای بهبود کیفیت تصاویر و ایجاد تصاویر مصنوعی که از نظر بصری قابل تمایز با تصاویر واقعی نیستند مورد استفاده قرار می‌گیرند. همانطور که گفته شد این شبکه‌ها از دو بخش مولد و تفکیک‌کننده تشکیل شده‌اند. مولد با نمونه‌گیری از یک توزیع تصادفی نرمال، تلاش می‌کند تصاویری مشابه تصاویر واقعی ایجاد کند. مولد، یک نوفه تصادفی^{۲۲} را از یک توزیع احتمالی نرمال نمونه‌برداری کرده و تلاش می‌کند تصاویری مشابه نمونه‌های واقعی تولید کند. این نوفه تصادفی از طریق یک تابع مشتق‌پذیر به فضای داده‌های تصویری نگاشت می‌شود که پارامترهای آن توسط مولد تعیین می‌شوند. هدف مولد، تولید تصاویری است که به‌طور موثری ویژگی‌های تصاویر واقعی را تقلید کنند [۱۳]. مولد تلاش می‌کند تصاویر تولیدشده را به نمونه‌های واقعی نزدیک کند، درحالی‌که تفکیک‌کننده به دنبال بیشینه کردن تابع زیان برای افزایش دقت تمایز میان تصاویر واقعی و جعلی است. این تعامل باعث می‌شود مولد به تدریج توانایی تولید تصاویر واقعی‌تری پیدا کند.

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

در این رابطه، x بیانگر داده‌های واقعی، z نوفه ورودی با توزیع نرمال، $G(z)$ تصویر تولیدشده توسط مولد، $D(x)$ احتمال واقعی بودن تصویر ورودی و \mathbb{E} تابع امید ریاضی است که برای محاسبه میانگین عملکرد تفکیک‌کننده و مولد روی تمام نمونه‌های ممکن داده‌ها (واقعی و جعلی) استفاده می‌شود. بخش اول رابطه، عملکرد تفکیک‌کننده روی داده‌های واقعی را نشان می‌دهد؛ به این صورت که احتمال واقعی بودن هر تصویر واقعی x را محاسبه کرده و امید ریاضی، میانگین این احتمالات را بر روی کل داده‌های واقعی به دست می‌آورد. بخش دوم نیز مربوط به محاسبه عملکرد تفکیک‌کننده روی داده‌های جعلی تولید شده توسط مولد می‌باشد. $G(x)$ معادل مولد است که تصویر جعلی $G(z)$ را با استفاده از ورودی نوفه تصادفی z تولید می‌کند و $D(G(z))$ احتمال واقعی بودن آن است. مولد تلاش می‌کند که

²⁰ Stacked Conditional Generative Adversarial Network

²¹ Cascaded Progressive Generative Adversarial Network

²² Random Noise

مقدار $\log(1 - D(G(z)))$ را کمینه کند که نشان دهنده نزدیکی تصاویر جعلی به تصاویر واقعی می‌باشد. عبارت سمت چپ تساوی نشان دهنده یک بازی بین مولد و تفکیک‌کننده است. \max_D نشان دهنده تلاش تفکیک‌کننده برای بیشینه کردن تابع هدف $V(G, D)$ بوده و \min_G بیانگر تلاش مولد در کمینه کردن تابع هدف است. شبکه SRGAN یک معماری پیشرفته از GAN است که شامل بلوک‌های باقی‌مانده در مولد و تفکیک‌کننده است. همچنین، مولد در این شبکه از اتصالات میانبر بهره می‌برد تا جزئیات دقیق‌تری در تصاویر تولیدشده استخراج شود. ویژگی منحصر به فرد SRGAN، استفاده از تصاویر با وضوح پایین به جای نوفه تصادفی برای آموزش مولد است. این رویکرد نقطه شروع بهتری برای تولید تصاویر با وضوح بالا فراهم می‌کند و باعث بهبود خروجی نهایی می‌شود. برای تضمین واقعی بودن تصاویر تولید شده و همچنین ارتقای کیفیت بصری آن‌ها، SRGAN از سه نوع تابع زیان استفاده می‌کند: زیان L_1 یا L_2 برای کاهش تفاوت پیکسلی بین تصویر بازسازی شده و تصویر مرجع، زیان VGG یا ادراکی برای بهبود جزئیات و بافت تصاویر، و زیان تقابلی برای واقعی‌تر کردن تصاویر تولیدشده. زیان VGG با بهره‌گیری از ویژگی‌های سطح بالا که توسط معماری پیش‌آموزش‌یافته VGG استخراج می‌شود، به بهبود کیفیت ادراکی تصاویر کمک می‌کند. زیان تقابلی نیز مولد را وادار به تولید تصاویری می‌کند که به‌طور کامل ویژگی‌های تصاویر واقعی را بازتاب دهند. برای تضمین اینکه تصاویر تولیدشده با وضوح بالا هم از نظر بصری واقعی باشند و هم کیفیت ادراکی مطلوبی داشته باشند، شبکه SRGAN از سه نوع تابع زیان در فرآیند آموزش استفاده می‌کند. همان‌طور که در رابطه (۲) آمده است، توابع زیان L_1 (میانگین قدر مطلق خطا) و L_2 (میانگین مربعات خطا) برای کاهش اختلاف پیکسلی بین تصویر تولیدشده و تصویر واقعی با وضوح بالا استفاده می‌شوند. این مرحله تضمین می‌کند که مقادیر شدت روشنایی هر پیکسل در تصویر تولیدشده به‌طور دقیق با مقادیر متناظر در تصویر مرجع تطابق داشته باشد، که این امر برای حفظ شباهت ساختاری میان دو تصویر ضروری است.

$$I_{MSE}^{SR} = \frac{1}{r^2 HW} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{GT} - G(I^{LR}))^2 \quad (2)$$

I^{GT} نمایانگر تصویر مرجع با وضوح بالا^{۲۳} و $G(I^{LR})$ تصویر با وضوح بالای تولید شده توسط مولد از ورودی با وضوح پایین است. پارامترهای W و H نیز به ترتیب ارتفاع و عرض تصویر مرجع، و r میزان بزرگنمایی می‌باشد. این تابع زیان تلاش می‌کند اختلاف مقادیر شدت روشنایی پیکسل‌ها در تصویر تولیدشده و تصویر مرجع را به حداقل برساند، که این امر برای حفظ شباهت ساختاری میان دو تصویر بسیار حیاتی است. علاوه بر دقت پیکسلی، SRGAN بر بهبود کیفیت ادراکی تصاویر تولیدشده تمرکز دارد. برای دستیابی به این هدف، مدل از زیان VGG و زیان تقابلی استفاده می‌کند. زیان VGG از یک معماری از پیش آموزش یافته VGG برای استخراج ویژگی‌های سطح بالا و بافت‌ها بهره می‌گیرد تا نتایج بصری بهتری ارائه دهد [۱۸]. این زیان شباهت بین نقشه‌های ویژگی تصویر تولیدشده و تصویر مرجع را مطابق رابطه زیر اندازه‌گیری می‌کند.

²³ Ground Truth High-Resolution Image

$$l_{VGG}^{SR} = \sum_{k \in layers} \frac{1}{H_k W_k} \sum_{x=1}^{W_k} \sum_{y=1}^{H_k} \left\| \phi_k(I^{HR})_{x,y} - \phi_k(G(I^{LR}))_{x,y} \right\|^2 \quad (3)$$

ϕ_k بیانگر نقشه‌ای ویژگی استخراج شده از لایه k شبکه از پیش آموزش یافته VGG است. H_K و W_K به ترتیب ارتفاع و عرض نقشه‌های ویژگی در لایه k هستند که مقیاس مکانی تصویر را در فضای ویژگی نشان می‌دهند. زبان VGG از مقایسه نقشه‌های ویژگی به جای مقادیر پیکسلی استفاده می‌کند و مولد را به تولید تصاویری با جزئیات دقیق‌تر و بافت‌های طبیعی‌تر هدایت می‌کند. این روش با تمرکز بر ویژگی‌های انتزاعی‌تر در سطوح بالاتر شبکه، کیفیت ادراکی تصاویر را بهبود می‌بخشد و به مولد امکان می‌دهد فراتر از صرفاً شباهت پیکسلی عمل کند.

زبان تقابلی نیز برای وادار کردن مولد به تولید تصاویری که به اندازه تصاویر با وضوح بالا واقعی باشند، استفاده می‌شود. زبان تقابلی با استفاده از یک شبکه تفکیک‌کننده تعریف می‌شود که هدف آن تشخیص تصاویر واقعی از تصاویر جعلی تولید شده توسط مولد است که در رابطه (۱) مشخص شده است. این آموزش تقابلی باعث می‌شود مولد بتواند تصاویر واقعی‌تری تولید کند [۱۹]. در نهایت با وزن‌دهی مشخص به زبان‌های تعریف شده سهم هر زبان در تابع کلی تنظیم می‌شود. در فرآیند آموزش شبکه، گرادینت‌های هر زبان محاسبه می‌شود و این گرادینت‌ها برای به‌روزرسانی پارامترهای مولد و تفکیک‌کننده استفاده می‌شوند. انتخاب وزن‌ها بسیار مهم هستند و مقادیر آن‌ها معمولاً با آزمایش و خطا تعیین می‌شوند تا توازنی بین دقت، کیفیت بصری و واقع‌گرایی برقرار شود.

وانگ و همکاران در سال ۲۰۱۸ مدل جدیدی را با نام ESRGAN توسعه دادند که چند تغییر مهم نسبت به SRGAN داشت: جایگزینی بلوک‌های باقیمانده پایه با بلوک‌های پیشرفته‌تر RRDB، حذف لایه‌های نرمال‌سازی دسته‌ای و بهبود بخش تفکیک‌کننده [۱۲]. تغییر اول با هدف افزایش توانایی یادگیری شبکه و تسهیل فرآیند آموزش انجام شد. بلوک‌های RRDB به دلیل اتصال متراکم خود، بازنمایی بهتری از ویژگی‌ها ارائه داده و فرآیند آموزش را موثرتر کرده‌اند. تغییر دوم برای جلوگیری از ایجاد مصنوعات که گاهی توسط این لایه‌های به وجود می‌آید انجام شده است که به بهبود فرآیند آموزش و کیفیت تصاویر کمک می‌کند. تفکیک‌کننده در ESRGAN، برخلاف مدل‌های قبلی از زبان RaGAN^{۲۴} استفاده می‌کند. این زبان نه تنها واقعی یا جعلی بودن تصاویر را ارزیابی می‌کند، بلکه میزان شباهت تصاویر تولید شده به تصاویر واقعی را نیز بررسی می‌کند. این رویکرد نسبیتی بازخورد دقیق‌تری فراهم کرده و توانایی تفکیک‌کننده را برای هدایت مولد به سمت تولید تصاویر با کیفیت بالاتر تقویت کرده است. تمام این تغییرات در کنار هم موجب بهبود عملکرد ESRGAN شده و مدل RealESRGAN توانست با ارائه تصاویری با واقع‌گرایی بیشتر، وضوح بالاتر و جزئیات دقیق‌تر از SRGAN پیشی بگیرد [۱۳].

ویژگی برجسته RealESRGAN استفاده از فرآیندهای تخریب مرتبه بالاتر است؛ بدین معنا که تصاویر با کیفیت پایین تنها از طریق کوچک‌سازی ساده ایجاد نمی‌شوند، بلکه ترکیبی از نویز، تاری، فشرده‌سازی و اعوجاج‌های متداول نیز بر آن‌ها اعمال می‌شود تا شرایط واقعی‌تری از کاهش کیفیت تصویر شبیه‌سازی گردد. این رویکرد از تخریب‌های تصادفی مرتبه اول برای شبیه‌سازی بهتر تخریب‌های واقعی استفاده می‌کند و مدل را قادر می‌سازد تا در شرایط متنوع‌تری از تخریب تصاویر عملکرد مطلوبی داشته باشد، که این ویژگی آن را کاربردی‌تر کرده است. اگرچه ساختار کلی RealESRGAN به مدل SRGAN شباهت دارد، اما تفکیک‌کننده آن پیشرفت‌های قابل توجهی داشته است. به جای

²⁴ Relativistic Average GAN

استفاده از شبکه‌های کانولوشنی استاندارد برای تخمین احتمال واقعی بودن تصاویر ورودی، این مدل از معماری UNet بهره می‌برد. این معماری با ثبت جزئیات فضایی بیشتر در تصاویر، بازخورد دقیق‌تری از تفکیک‌کننده ارائه می‌دهد، که این بازخورد برای تنظیم بهینه پارامترهای مولد بسیار اهمیت دارد. با این تغییرات، RealESRGAN توانسته است عملکردی بهتر از سایر مدل‌ها ارائه دهد و با هدایت دقیق‌تر مولد، کیفیت بازسازی تصاویر و ابروضوح را به‌طور قابل‌توجهی ارتقا دهد [۱۳].

بعد از آموزش مدل و بدست آمدن نمونه باکیفیت تصاویر نیاز به یک معیار ارزیابی برای بررسی تصاویر تولیدی می‌باشد. بدین منظور از معیارهای مختلفی استفاده می‌شود. این معیارها معمولاً به دو دسته کلاسیک و پیشرفته (مبتنی بر یادگیری عمیق) تقسیم می‌شوند. معیارهای کلاسیک شامل میانگین مربعات خطا (MSE^{25})، ریشه میانگین مربعات خطا ($RMSE^{26}$)، نسبت سیگنال به نویز پیک ($PSNR^{27}$)، شاخص شباهت ساختاری ($SSIM^{28}$) هستند. MSE میانگین مربع تفاوت پیکسل‌های تصویر بازسازی‌شده و تصویر اصلی را محاسبه می‌کند. این معیار ساده است اما معمولاً همبستگی ضعیفی با ادراک انسانی دارد. $RMSE$ که نسخه مقیاس‌شده از MSE است، خطا را به شکلی قابل‌درک‌تر اندازه‌گیری می‌کند. $PSNR$ نسبت سیگنال به نویز را در تصویر بازسازی‌شده اندازه‌گیری می‌کند، گرچه این معیارها هم‌راستایی کمی با ارزیابی بصری انسانی دارند. در مقابل، $SSIM$ شباهت ساختاری را براساس روشنایی، کنتراست و ساختار ارزیابی می‌کند، که این ویژگی آن را برای مقایسه‌های ساختاری پیچیده مناسب‌تر می‌سازد [۲۰]. معیارهای پیشرفته از یادگیری عمیق برای ارزیابی بهتر شباهت ادراکی استفاده می‌کنند. به‌عنوان مثال، معیار $LPIPS^{29}$ با استفاده از استخراج ویژگی‌های عمیق، شباهت را به‌طور مؤثرتری با ادراک انسانی هماهنگ می‌کند، اما نیاز به شبکه‌های از پیش آموزش‌شده دارد. معیار FID^{30} توزیع ویژگی‌ها را مقایسه می‌کند تا شباهت بین تصاویر تولیدی و واقعی را ارزیابی کرده و تمرکز آن بر کیفیت و ظاهر واقع‌گرایانه تصاویر است. معیار IS^{31} نیز برای سنجش کیفیت و تنوع تصاویر تولیدشده است. هرچه تصاویر واضح‌تر باشند و مجموعه تصاویر تنوع بیشتری داشته باشد، مقدار IS بالاتر خواهد بود. با این حال، از آنجا که این معیار بر پایه شبکه‌های آموزش‌دیده روی تصاویر عمومی تعریف شده، برای داده‌های خاکستری و باینری محیط متخلخل چندان مناسب نیست. هر یک از این معیارها مزایا و معایب خاص خود را دارند و انتخاب بهترین معیار بستگی به اهداف و نیازهای خاص کاربرد دارد [۲۰].

با توجه به اینکه در فرآیند آموزش مدل از معیارهای کلاسیک MSE و $RMSE$ استفاده می‌شود و تصاویر به‌دست‌آمده تا حد خوبی از نظر این معیارها به تصاویر واقعی نزدیک هستند، نیاز به استفاده از معیارهای تکمیلی برای بررسی کیفیت تصاویر وجود دارد. در این پژوهش علاوه بر معیارهای مذکور، از دو معیار دیگر نیز استفاده شد. معیار $PSNR$ به‌عنوان یکی از معیارهای کلاسیک پرکاربرد در حوزه پردازش تصویر، نسبت بیشینه توان سیگنال به توان نویز را می‌سنجد. این معیار بر پایه MSE تعریف شده و از رابطه زیر محاسبه می‌شود.

²⁵ Mean Squared Error

²⁶ Root Mean Squared Error

²⁷ Peak Signal-to-Noise Ratio

²⁸ Structural Similarity Index

²⁹ Learned Perceptual Image Patch Similarity

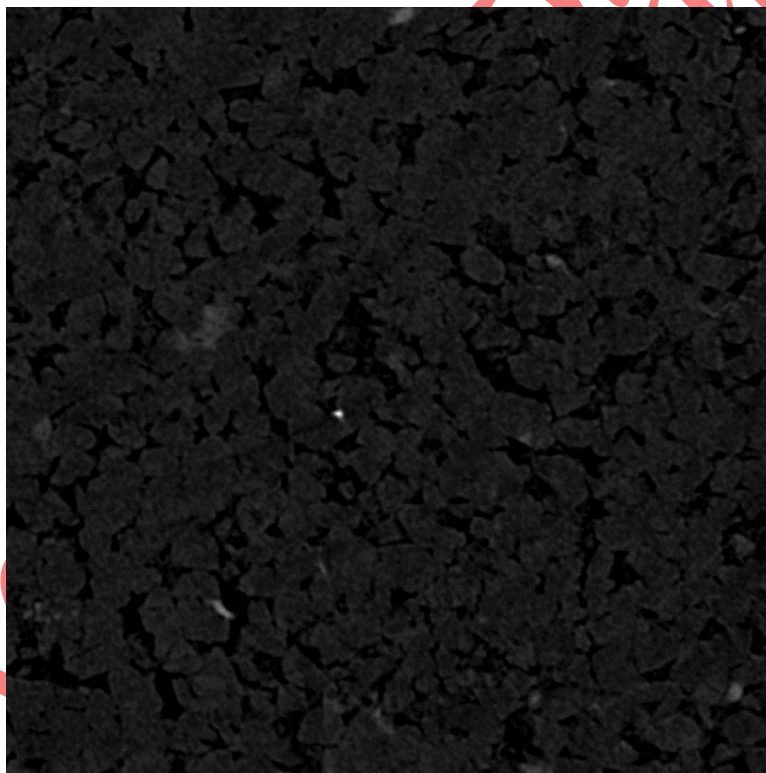
³⁰ Fréchet Inception Distance

³¹ Inception Score

$$PSNR = 20\log_{10}(MAX_I) - 10\log_{10}(MSE) \quad (4)$$

MAX_I بیشینه مقدار ممکن پیکسل در تصویر (برای تصاویر این پژوهش برابر ۲۵۵) و MSE همان میانگین مربعات خطا بین تصویر بازسازی شده و تصویر اصلی است. مقدار بالاتر این معیار نشان‌دهنده شباهت بیشتر تصویر بازسازی شده به تصویر اصلی است.

همچنین معیار LPIPS به‌عنوان یک معیار پیشرفته ادراکی مطرح است. برخلاف معیار FID که بر اساس توزیع‌های آماری تعریف می‌شود و فرض می‌کند ویژگی‌ها از توزیع گاوسی پیروی می‌کنند، LPIPS بر پایه تحلیل شباهت‌های ادراکی در فضای ویژگی‌های شبکه‌های از پیش‌آموزش‌دیده عمل می‌کند. این ویژگی باعث می‌شود LPIPS تطابق بیشتری با سیستم بینایی انسان داشته باشد و نتایج دقیق‌تری در ارزیابی کیفیت ادراکی تصویر ارائه دهد [۲۱]. برای آموزش شبکه و محاسبه این معیار از کتابخانه‌ی متن‌باز lpips در پایتون استفاده شد. این معیار فاصله بین دو تصویر را بیان می‌کند و مقدار کمتر آن نشان‌دهنده شباهت ادراکی بیشتر بین تصویر بازسازی شده و تصویر اصلی است.

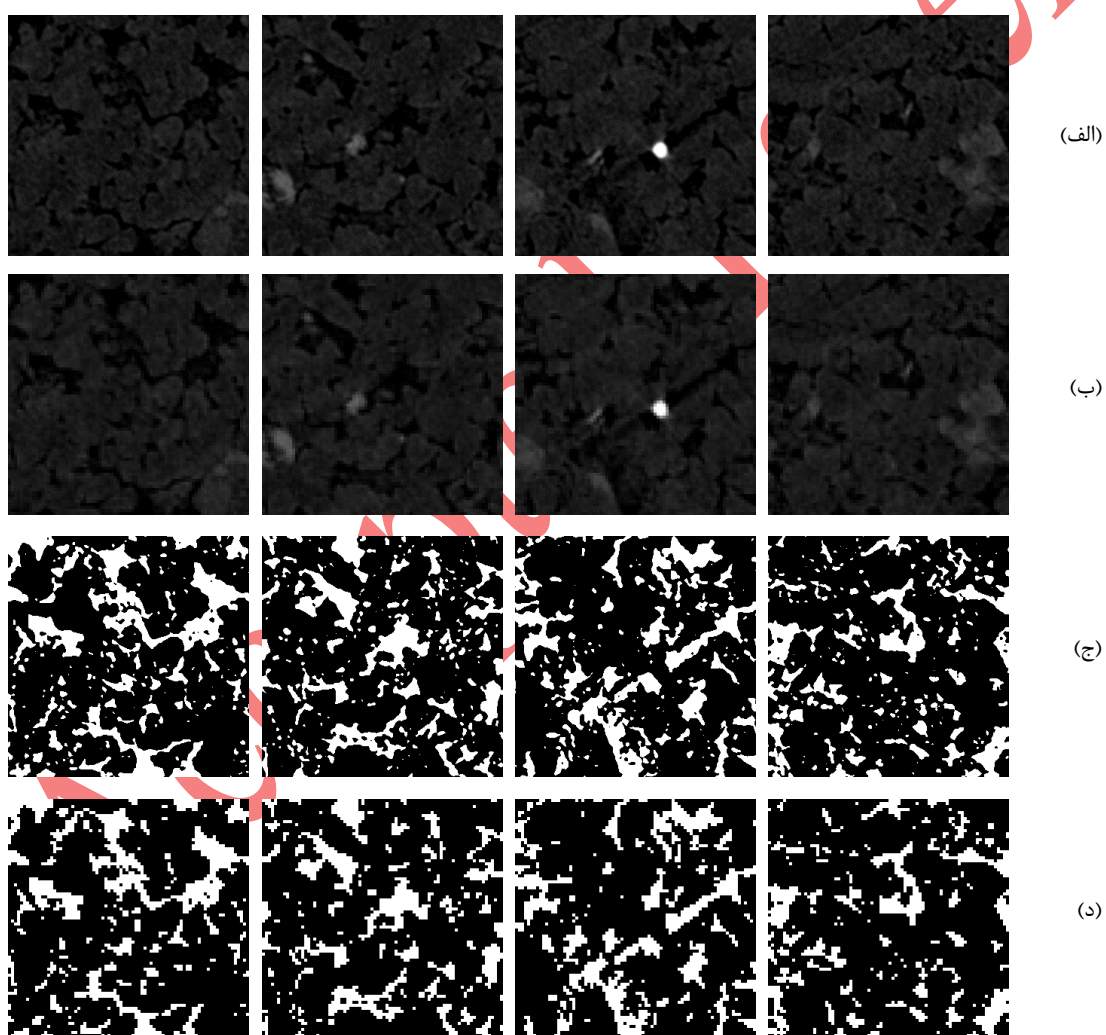


شکل ۱. نمونه تصویر ماسه‌سنگ Breccia

هدف از این پژوهش بررسی اثر استفاده از نوع تصاویر ورودی به مدل‌های یادگیری عمیقی است که جهت افزایش کیفیت تصاویر ورودی استفاده می‌شوند. برای بررسی این اثر نیاز به یک مجموعه داده مناسب می‌باشد. بدین منظور از داده‌های ماسه‌سنگ Breccia استفاده می‌شود که تخلخلی در حدود ۱۸.۹۶ دارد [۲۲]. شکل ۱ یک نمونه تصویر از این سنگ را نمایش می‌دهد. کیفیت این تصویر ۲.۲۵ میکرون می‌باشد. ۱۰۰۰ تصویر از این محیط با ابعاد ۱۰۰۰ در ۱۰۰۰ پیکسل

موجود می‌باشد. نقاط سیاه نشان دهنده فضای خالی حفرات و با کاهش میزان رنگ هر پیکسل به سمت سفید سنگ مشخص می‌شود.

با دقت در شکل ۱ مشاهده می‌شود که حتی پیکسل‌هایی که نشان دهنده سنگ هستند نیز از طیف رنگی تیره‌ای برخوردار هستند. بنابراین تعیین یک حد مناسب برای باینری کردن تصاویر محیط متخلخل نیاز می‌باشد. به طور کلی هر پیکسل در یک تصویر خاکستری مقداری بین ۰ تا ۲۵۵ دارد. بهترین راه برای تعیین حد باینری تصاویر استفاده از نمودار توزیع مقادیر پیکسل‌ها در تمامی تصاویر محیط متخلخل می‌باشد. انتخاب حد باید به گونه‌ای انجام شود که مقدار تخلخل محاسبه شده با استفاده از تصاویر باینری با تخلخل بدست آمده از آزمایشگاه مطابقت داشته باشد. در بخش نتایج در مورد حد مناسب بدست آمده توضیح داده می‌شود.



شکل ۲. نمونه تصاویر (الف) خاکستری با کیفیت، (ب) خاکستری با ۴ مرتبه کیفیت کاهش یافته، (ج) باینری با کیفیت، (د) باینری با ۴ مرتبه کیفیت کاهش یافته، از ماسه سنگ Brea

همانطور که در بخش مقدمه اشاره شد، ورودی این پژوهش دو مجموعه داده خاکستری و باینری می‌باشد. بعد از تعیین

حد مناسب، باینری کردن تصاویر انجام می‌شود. این مورد هم برای تصاویر با کیفیت بالا و هم تصاویر با کیفیت پایین انجام می‌شود. مدل RealESRGAN تا ۲ و ۴ برابر قابلیت افزایش کیفیت تصاویر را دارد، برای ارزیابی بهتر هدف مورد نظر از نسبت ۴ برابری جهت افزایش کیفیت تصاویر استفاده می‌شود. بنابراین نیاز به تصاویری با کیفیت ۴ برابر کمتر نسبت به تصاویر واقعی داریم. برای تولید تصاویر با کیفیت پایین، از یک روش ساده مبتنی بر میانگین‌گیری استفاده شد. به این صورت که هر بلوک 4×4 پیکسل به یک پیکسل جدید تبدیل شد. در تصاویر خاکستری، مقدار پیکسل جدید برابر با میانگین مقادیر آن ۱۶ پیکسل قرار داده شد. در تصاویر باینری نیز ابتدا میانگین این ۱۶ پیکسل محاسبه گردید و سپس با اعمال یک آستانه ۰.۵، مقدار جدید تعیین شد؛ به گونه‌ای که اگر میانگین کمتر از ۰.۵ بود مقدار صفر و در غیر این صورت مقدار یک به پیکسل اختصاص یافت. این فرآیند موجب کاهش وضوح تصویر به یک‌چهارم مقدار اصلی شد و داده‌های با کیفیت پایین لازم برای آموزش مدل را فراهم نمود.

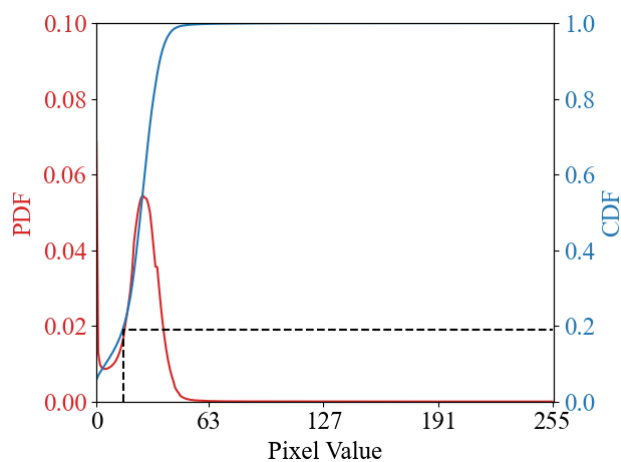
از آنجایی که مدل‌های هوش مصنوعی به حجم داده زیادی نیاز دارند و با توجه به اهمیت جزئیات موجود در تصاویر بهتر است که تصاویر اولیه به تصاویر کوچک‌تری برای ورود به مدل تبدیل شوند. بنابراین هر یک از تصاویر اولیه به سائز ۹۶۰ در ۹۶۰ تبدیل می‌شوند که قابلیت تبدیل به سائزهای کوچک‌تر را داشته باشد. هر یک از ابعاد مدل سه‌بعدی اولیه به ۳ بخش تقسیم می‌شود. این کار باعث می‌شود که ۲۷ مدل کوچک‌تر سه‌بعدی در اختیار داشته باشیم. از این مکعب‌ها مکعب وسط به عنوان داده مورد نظر برای اعتبارسنجی نهایی انتخاب می‌شود. در بین ۲۶ مکعب دیگر ۷۰ درصد تصاویر به عنوان داده آموزشی و باقی تصاویر به عنوان داده تست استفاده می‌شوند. در شکل ۲ نمونه‌ای از این تصاویر آورده شده است (برای نشان دادن تنوع تصاویر در هر مورد چهار تصویر آورده شده است که تصاویر هر ستون معادل یکدیگر می‌باشند). با توجه به حد باینری تعیین شده به ازای تمامی تصاویر باکیفیت و بی‌کیفیت موجود نمونه باینری آن‌ها نیز تهیه می‌شود. جدول ۱ تعداد تصاویر استفاده شده در این پژوهش را نشان می‌دهد.

جدول ۱. تعداد تصاویر مورد استفاده در فرآیند آموزش و تست مدل

مجموعه داده	تعداد کل تصاویر	تعداد تصاویر آموزشی	تعداد تصاویر تست	تعداد تصاویر اعتبارسنجی
تصاویر خاکستری	۸۶۴۰	۵۸۲۴	۲۴۹۶	۳۲۰
تصاویر باینری	۸۶۴۰	۵۸۲۴	۲۴۹۶	۳۲۰

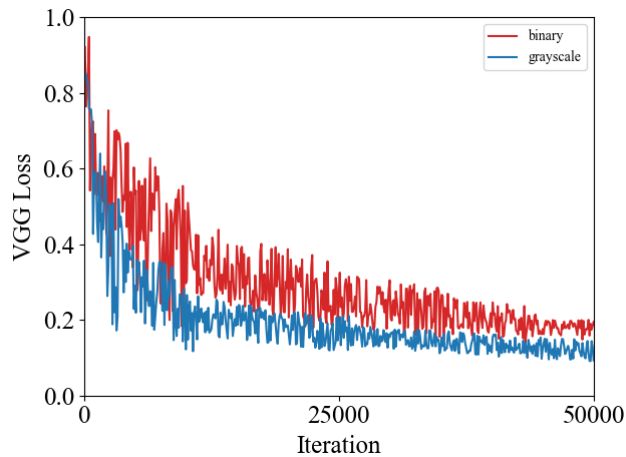
بحث و نتایج

همانطور که در بخش قبل اشاره شد یکی از مراحل آماده‌سازی داده‌های ورودی، باینری کردن تصاویر ورودی می‌باشد. تعیین حد مناسب برای باینری کردن تصاویر یکی از مراحل حائز اهمیت می‌باشد. برای این هدف از نمودار توزیع فراوانی مقادیر پیکسل تصاویر خاکستری استفاده می‌شود که در شکل ۳ نشان داده شده است. همانطور که در این شکل مشاهده می‌شود توزیع مقادیر پیکسل دارای دو مد می‌باشد. از آنجایی که در تصاویر ماسه‌سنگ Brea رنگ سیاه نشان دهنده سنگ و رنگ سفید نشان دهنده فضای خالی است، در این نمودار مد اول مربوط به حفرات و مد دوم مربوط به سنگ می‌باشد. بنابراین مقداری حدود ۰.۱۸۹۶ که برابر تخلخل سنگ است برای تعیین حد باینری در نمودار فراوانی تجمعی در نظر گرفته شده و مقدار معادل آن در محور x به مقدار ۱۵ می‌رسد. بنابراین مقدار ۱۵ به عنوان حد باینری در نظر گرفته می‌شود. تصاویر باینری موجود در شکل ۲ نیز با همین حد باینری به دست آمده‌اند.



شکل ۳. نمودار توزیع مقادیر پیکسل تصاویر ورودی در ماسه‌سنگ Brea

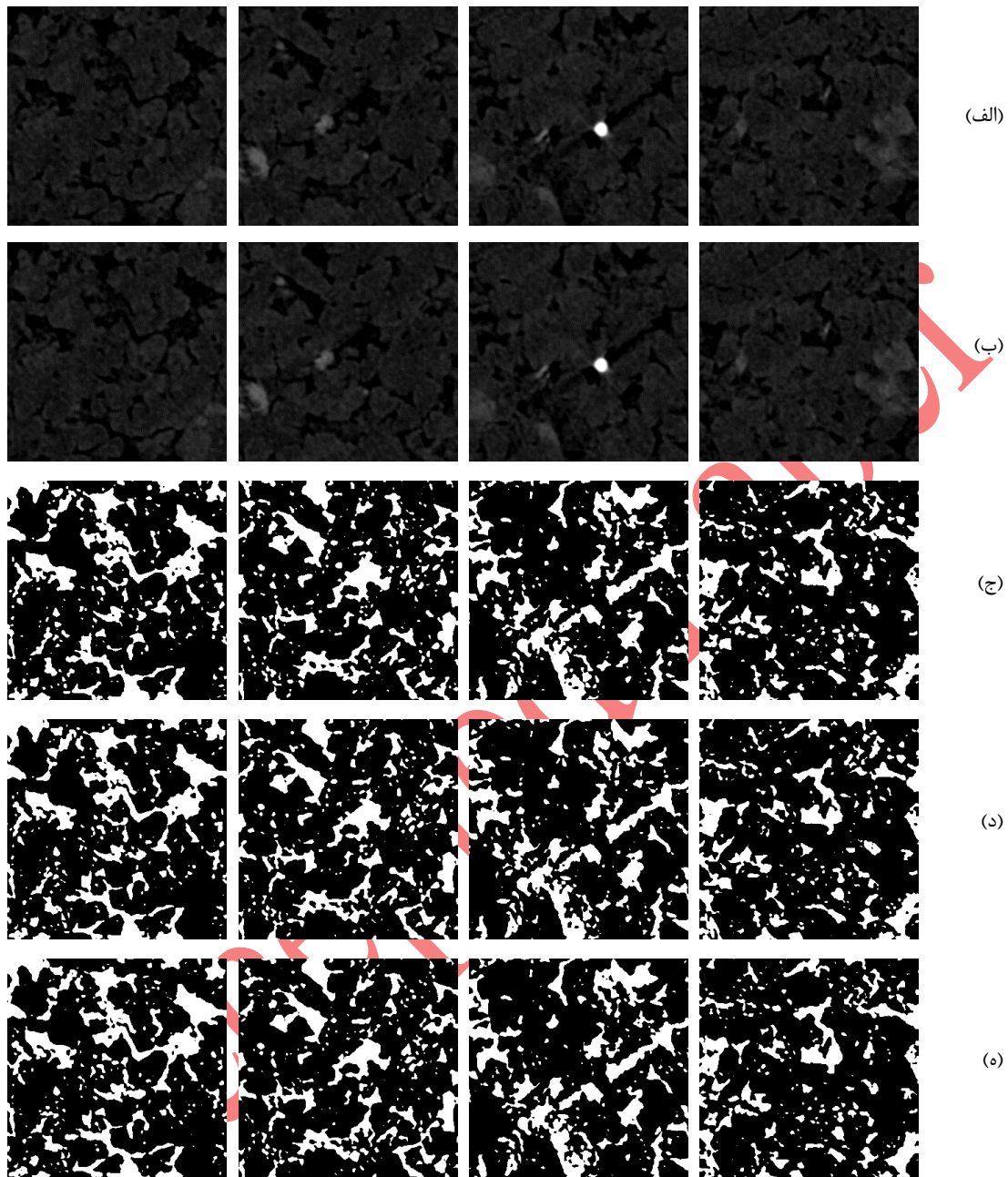
بعد از آماده‌سازی تصاویر فرآیند آموزش مدل انجام می‌شود. برای آموزش مدل از بهینه‌ساز Adam استفاده شد. این روش نوعی الگوریتم بهینه‌سازی مبتنی بر گرادیان است که با ترکیب دو مفهوم مومنتوم و تخمین تطبیقی نرخ یادگیری عمل می‌کند. در این الگوریتم، میانگین نمایی گرادیان‌ها (مرتب اول) و میانگین نمایی مربع گرادیان‌ها (مرتب دوم) به صورت هم‌زمان محاسبه و به‌روزرسانی می‌شوند. این موضوع باعث می‌شود الگوریتم در مقایسه با روش‌های ساده‌تری مانند گرادیان نزولی تصادفی، سرعت همگرایی بالاتر و پایداری بیشتری داشته باشد. در این پژوهش، مقادیر اولیه نرخ یادگیری برابر با 0.0001 ، پارامتر مومنتوم اول (β_1) و برابر 0.9 و مومنتوم دوم (β_2) برابر 0.999 در نظر گرفته شد. اجرای برنامه با استفاده از گرافیک Nvidia GeForce RTX 3090 GPU انجام شده است که نسبت به سیستم‌های معمولی توانایی بالاتری دارد و اندازه دسته روی ۸ تنظیم شده است به طوری که همزمان ۸ تصویر در فرآیند آموزش به طور موازی بررسی می‌شوند که سرعت آموزش را بالا می‌برد (این امکان در گرافیک‌های پایین‌تر وجود ندارد). با توجه به محدودیت‌های موجود و زمان زیاد مورد نیاز برای آموزش مدل‌ها و با توجه به اینکه دو بار آموزش (یک بار با تصاویر خاکستری و یک بار با تصاویر باینری) نیاز به آموزش مدل بود، تصمیم گرفته شد که آموزش مدل با 5000 بار حدس و خطا انجام شود. البته اجرای بیشتر این حدس و خطا نیز امکان‌پذیر است ولی نیازمند صرف زمان بیشتری خواهد بود ولی در این پژوهش هدف از آموزش و ساخت مدل مقایسه عملکرد مدل با دو نوع متفاوت از تصاویر ورودی است. فرآیند آموزش بعد از تنظیم پارامترهای مدل در حدود ۸.۵ ساعت برای تصویرهای خاکستری به طول انجامید و برای تصاویر باینری این عدد به حدود ۱.۵ ساعت کاهش یافت که میزان کاهش قابل ملاحظه‌ای داشت. شکل ۴ تابع زیان محاسبه شده در فرآیند آموزش دو مدل با ورودی‌های باینری و خاکستری را نمایش می‌دهد.



شکل ۴. تابع زیان VGG در فرایند آموزش دو مدل باینری و خاکستری

با دقت در شکل ۴ مشخص می‌شود که استفاده از تصاویر باینری باعث می‌شود در ابتدا مدل خطای بیشتری داشته باشد و عدم قطعیت مدل نیز افزایش می‌یابد به گونه‌ای که در مقادیر تلاش ابتدایی تغییرات زیادی در تابع زیان اتفاق افتاده و نوسان زیادی دارد. البته یکی از مزیت‌های استفاده از تصاویر باینری این است نوسان‌های انتهایی کمتری دارند، هر چند برای رسیدن به این حد تعداد حدس و خطای بیشتری لازم دارد. در تصاویر خاکستری رسیدن به حد خطای پایین تا حدود ۱۲۰۰۰ تلاش اتفاق می‌افتد که از نظر زمانی در حدود ۲ ساعت زمان می‌برد، در حالی که با استفاده از تصاویر باینری این زمان به حدود ۱.۵ ساعت کاهش می‌یابد و همان دقت را به ما می‌دهد. البته دقت نهایی تصاویر خاکستری بهتر از تصاویر باینری می‌باشد که نیازمند صرف زمان بیشتر و سخت‌افزار قوی‌تری برای آموزش مدل می‌باشد. شکل ۵ نمونه‌ای از تصاویر پیش‌بینی شده با استفاده از دو مدل مختلف نشان می‌دهد (تصاویر معادل تصاویر موجود در شکل ۲ می‌باشند). با توجه به اینکه نتایج بدست آمده از تصاویر تست نتایج بهتر و دقیق‌تری هستند و تعداد بالای این تصاویر باعث افزایش حجم نمودارها می‌شود، صرفاً از این به بعد بر روی نتایج بدست آمده از مجموعه اعتبار سنجی استفاده می‌شود. بخاطر اینکه داده‌های تست به طور تصادفی از بین تصاویر دیگر استخراج شده‌اند و تصاویر کنارشان در فرآیند آموزش دیده شده‌اند، این مورد باعث می‌شود که نتایج بدست آمده در این مجموعه نتایج بهتری باشد. همانطور که در بخش قبل اشاره شد مجموعه اعتبار سنجی به عنوان یک بلوک جدا از مجموعه ۲۷ بلوک مجزای موجود در داده‌های ورودی انتخاب شده است و عدم تطابق دقیق الگوهای این بلوک با سایر بلوک‌ها باعث می‌شود پدیده بیش‌برازش^{۳۲} در این داده‌ها دیده نشود. با دقت در شکل ۶ مشخص می‌شود که به تصاویر باینری تولیدشده دو نوع هستند. دسته‌ای که مستقیماً از طریق مدل آموزش دیده با تصاویر باینری تولید می‌شوند و دسته دیگر از تصاویر خاکستری بدست آمده توسط مدل آموزش دیده با تصاویر خاکستری تولید شده‌اند. با دقت بصری در تصاویر تولید شده نمی‌توان به وضوح کیفیت تصاویر بدست آمده را مشخص نمود، بنابراین نیاز به استفاده از معیاری جهت ارزیابی کیفیت تصاویر می‌باشد.

³² Overfitting

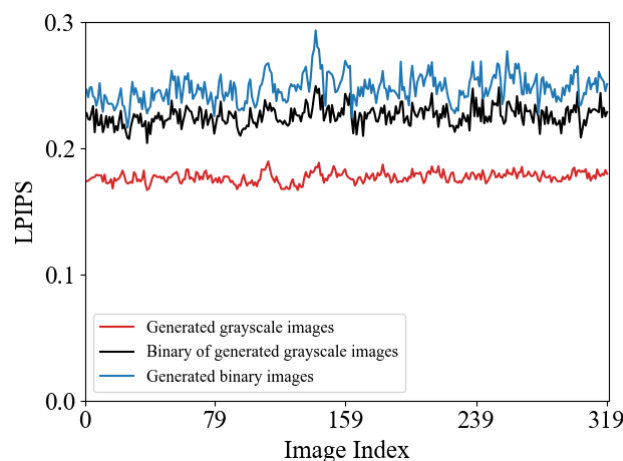


شکل ۵. نمونه تصاویر الف) خاکستری باکیفیت، ب) خاکستری تولید شده توسط مدل، ج) باینری باکیفیت، د) باینری تولید شده توسط مدل، ه) باینری حاصل از تصاویر خاکستری تولید شده توسط مدل

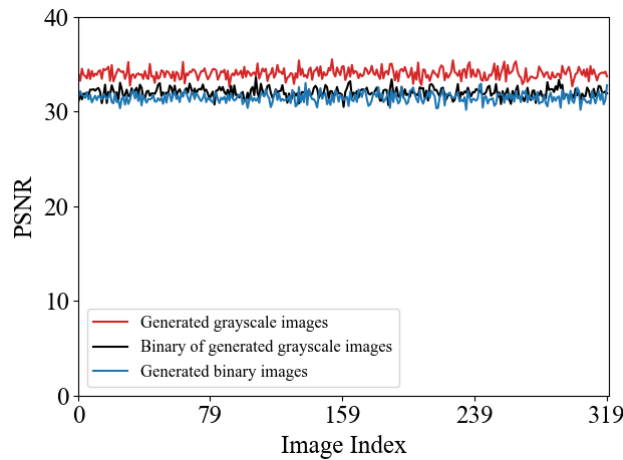
همانطور که در بخش قبل اشاره شد، برای ارزیابی تصاویر نهایی می‌توان از معیارهای ارزیابی نظیر LPIPS و PSNR نیز استفاده نمود. شکل ۶ مقایسه بین ۳۲۰ تصویر موجود در مجموعه تست را با استفاده از معیار LPIPS نشان می‌دهد. این تصاویر توسط دو مدل آموزش دیده با ورودی‌های خاکستری و باینری بدست آمده‌اند. همانطور که قبلاً ذکر شد، هر چه میزان این معیار بیشتر باشد نشان دهنده تفاوت بیشتر تصاویر می‌باشد. با توجه به شکل ۶ مشخص می‌شود که تصاویر خاکستری شباهت بیشتری به یکدیگر دارند که نشان دهنده عملکرد بهتر مدل می‌باشد. همچنین میزان تغییر و نوسان

این معیار در مقایسه با تصاویر باینری کمتر است که نشان می‌دهد به طور کلی استفاده از تصاویر خاکستری نتایج بهتری دارد. اما نکته‌ای که در اینجا اهمیت دارد این است که تصاویر خاکستری به طور مستقیم در فرآیند مدل‌سازی شبکه حفرات کاربردی ندارند. در واقع تصاویر خاکستری به تصاویر باینری تبدیل می‌شوند و پس از آن از این تصاویر جهت استخراج اطلاعات شبکه حفرات استفاده می‌شود. بنابراین بهترین راه، مقایسه نتیجه تصاویر باینری بدست آمده از دو مدل می‌باشد. یکی از مجموعه‌ها تصاویر باینری بدست آمده از مدلی است که توسط تصاویر باینری آموزش دیده و مجموعه دیگر تصاویر باینری است که از تصاویر خاکستری تولیدشده توسط مدل آموزش‌دیده با استفاده از تصاویر خاکستری بدست آمده‌اند. هر یک از این مجموعه‌ها با تصاویر باینری بدست آمده از تصاویر خاکستری واقعی مقایسه می‌شوند. شکل ۶ نشان می‌دهد که به طور کلی استفاده از تصاویر خاکستری نتایج بهتری دارد ولی به نسبت زمان ۶ برابری که در فرآیند آموزش دارد تفاوت بسیار زیادی در نتایج ایجاد نمی‌کند. بنابراین به نظر می‌رسد که در حوزه مدل‌سازی شبکه حفرات به دلیل ماهیت مسئله و استفاده از تصاویر باینری با توجه به زمان بیشتر آموزش مدل استفاده از تصاویر باینری منطقی‌تر به نظر می‌رسد.

در ادامه، شکل ۷ نتایج مقایسه همین تصاویر را با استفاده از معیار PSNR نشان می‌دهد. همانطور که مشاهده می‌شود، مقادیر PSNR برای تمامی مجموعه‌ها بالاتر از ۳۰ دسی‌بل بوده و به یکدیگر نزدیک هستند. این موضوع نشان می‌دهد که PSNR به‌عنوان یک معیار کلاسیک مبتنی بر اختلاف پیکسل به پیکسل نمی‌تواند تفاوت‌های کیفی میان تصاویر را به خوبی آشکار سازد و بیشتر جنبه‌ی کلی شباهت را می‌سنجد. مقایسه شکل‌های ۶ و ۷ بیانگر این است که هرچند PSNR تصویری کلی از نزدیکی بازسازی به مرجع ارائه می‌دهد، اما LPIPS به دلیل ماهیت ادراکی خود توانایی بیشتری در نمایش تفاوت‌های کیفی میان تصاویر بازسازی‌شده دارد.



شکل ۶. معیار LPIPS محاسبه شده بین تصاویر خاکستری تولید شده و واقعی، باینری بدست آمده از تصاویر خاکستری تولید شده و واقعی، و باینری تولید شده و باینری بدست آمده از تصاویر واقعی



شکل ۷. معیار PSNR محاسبه شده بین تصاویر خاکستری تولید شده و واقعی، باینری بدست آمده از تصاویر خاکستری تولید شده و واقعی، و باینری تولید شده و باینری بدست آمده از تصاویر واقعی

همان طور که در جدول ۲ مشاهده می شود، مقادیر PSNR در هر سه حالت بالاتر از ۳۰ دسی بل بوده و اختلاف اندکی با یکدیگر دارند، که نشان می دهد این معیار کلاسیک توانایی محدودی در نمایش تفاوت های کیفی میان تصاویر دارد. در مقابل، معیار LPIPS تفاوت های بیشتری میان حالت های مختلف نشان داده و بیانگر آن است که تصاویر خاکستری شباهت ادراکی بیشتری به تصاویر مرجع دارند. با این حال، اختلاف کلی میان حالت ها چندان زیاد نیست و این موضوع تأیید می کند که استفاده از تصاویر باینری نیز می تواند نتایج قابل قبولی ارائه دهد، به ویژه با در نظر گرفتن زمان محاسباتی کمتری که دارند.

جدول ۲. مقادیر میانگین و انحراف معیار برای معیارهای PSNR و LPIPS برای مجموعه تست در سه حالت مختلف

معیار	حالت خاکستری	باینری از خاکستری	خاکستری
LPIPS	0.0041 ± 0.1770	0.0109 ± 0.2475	0.0076 ± 0.2257
PSNR	0.5373 ± 34.0395	0.5130 ± 32.0202	0.5210 ± 31.04787

جدول ۳ مقادیر تخریب به دست آمده از سناریوهای مختلف را نشان می دهد. همان طور که مشاهده می شود، استفاده از تصاویر باینری حتی پس از کاهش کیفیت و بازسازی، تنها خطایی در حدود ۳ درصد نسبت به مقدار مرجع ایجاد کرده است که می توان آن را خطای قابل قبول دانست. در مقابل، اگرچه استفاده از تصاویر خاکستری و سپس تبدیل آن ها به باینری می تواند خطا را اندکی کاهش دهد، اما این مسیر نیازمند بار محاسباتی سنگین تری است. بنابراین، این نتایج می تواند تاییدی بر استفاده مستقیم از تصاویر باینری در فرآیند ارتقای کیفیت باشد که با وجود سادگی و سرعت بیشتر، رویکردی منطقی و قابل اتکا برای مدل سازی شبکه حفرات محسوب می شود.

جدول ۳. مقادیر تخلخل شبکه حفرات حاصل از تصاویر مختلف

خطا (%)	تخلخل	نوع داده
مرجع	۱۸.۹۶	تصاویر خاکستری
۰.۹۵	۱۸.۷۸	تصاویر باینری حاصل از تصاویر خاکستری با کیفیت (باینری اصلی)
۵.۴۹	۱۷.۹۲	تصاویر باینری حاصل از تصاویر خاکستری بی کیفیت
۲.۲۷	۱۸.۵۳	تصاویر باینری حاصل از تصاویر خاکستری با کیفیت افزایش یافته
۷.۱۲	۱۷.۶۱	تصاویر باینری بی کیفیت حاصل از تصاویر باینری اصلی
۲.۹۵	۱۸.۴۰	تصاویر باینری با کیفیت افزایش یافته

نتیجه گیری

در این پژوهش مدل RealESRGAN به عنوان یک مدل جدید و کاربردی جهت افزایش کیفیت تصاویر استفاده شد. برای ارزیابی عملکرد مدل علاوه بر معیار LPIPS به عنوان یکی از معیارهای پیشرفته ارزیابی و مقایسه تصاویر، از معیار PSNR نیز به عنوان یک معیار کلاسیک استفاده گردید. نتایج نشان داد که در حالی که PSNR اختلاف چندانی میان مجموعه تصاویر مختلف نشان نمی دهد و مقادیر آن برای همه تصاویر بالاتر از ۳۰ دسی بل و نزدیک به یکدیگر است، معیار LPIPS توانسته تفاوت های ادراکی میان تصاویر بازسازی شده را با دقت بیشتری آشکار سازد. همچنین نتایج محاسبه تخلخل بر روی تصاویر باینری بیانگر آن است که خطای بازسازی در مقایسه با مقدار واقعی تنها حدود ۳ درصد بوده است که می توان آن را خطای قابل قبول دانست.

به عنوان یکی از چالش های اساسی در فرآیند آموزش مدل، نتیجه استفاده از تصاویر باینری یا خاکستری در تصاویر نهایی مورد بررسی قرار گرفت. برای این هدف از یک نمونه ماسه سنگ Brea استفاده شد. با توجه به محدودیت تصاویر و با هدف افزایش سرعت آموزش، تصاویر اولیه به بلوک های کوچک تر تقسیم شدند و یک مجموعه داده مصنوعی اولیه شامل ۲۷ بلوک تولید شد که یکی از آن ها به عنوان داده اعتبارسنجی برای بررسی نتایج نهایی مورد استفاده قرار گرفت. نتایج به دست آمده تأیید می کند که به طور کلی استفاده از تصاویر خاکستری می تواند دقت بیشتری به همراه داشته باشد، اما با توجه به ماهیت و نوع کاربرد این تصاویر در فرآیند مدل سازی شبکه حفرات که تنها به صورت باینری استفاده می شوند، و همچنین با در نظر گرفتن زمان آموزش ۶ برابری برای تصاویر خاکستری، ارزش افزوده چندانی در نهایت ایجاد نمی کند. بنابراین با توجه به طولانی بودن روند آموزش مدل و خطای محدود به دست آمده برای تصاویر باینری، توصیه می شود که در اینگونه مسائل از تصاویر باینری به عنوان ورودی استفاده شود.

البته در مسائل دیگر که تصاویر خاکستری به عنوان ورودی استفاده می شوند نوع نتیجه گیری می تواند متفاوت باشد. به عنوان پیشنهادی برای پژوهش های آینده این بررسی می تواند بر روی نمونه سنگ های دیگری نیز انجام شود. علاوه بر این، مطالعه نمونه های مختلف این امکان را می دهد که اثر طول همبستگی بین داده ها در نتایج نهایی بررسی گردد. باید توجه داشت که این مطالعه صرفاً بر یک نمونه ماسه سنگ و یک مدل خاص در مقیاس چهار برابر متمرکز بوده است و معیارهای ارزیابی نیز محدود به LPIPS و PSNR بوده اند. بنابراین تعمیم نتایج به سایر انواع سنگ ها، مقیاس های متفاوت یا مدل های افزایش کیفیت دیگر نیازمند پژوهش های تکمیلی است که می تواند در آینده دنبال شود.

- [1] Dong, H. and Blunt, M.J., 2009. Pore-network extraction from micro-computerized-tomography images. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 80(3), p.036307. DOI: 10.1103/PhysRevE.80.036307.
- [2] Blunt, M.J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., Paluszny, A. and Pentland, C., 2013. Pore-scale imaging and modelling. *Advances in Water resources*, 51, pp.197-216. DOI: 10.1016/j.advwatres.2012.03.003.
- [3] Carman, P.C., 1939. Permeability of saturated sands, soils and clays. *The Journal of Agricultural Science*, 29(2), pp.262-273. DOI: 10.1017/S0021859600051789.
- [4] Wu, Y., Tahmasebi, P., Lin, C., Munawar, M.J. and Cnudde, V., 2019. Effects of micropores on geometric, topological and transport properties of pore systems for low-permeability porous media. *Journal of Hydrology*, 575, pp.327-342. DOI: 10.1016/j.jhydrol.2019.05.014.
- [5] Wu, X., Wang, F., Xiao, Z., Zhang, Y., Zhao, J., Fang, C. and Wei, B., 2024. Multiscale pore network modeling and flow property analysis for tight sandstone: a case study. *Journal of Geophysics and Engineering*, 21(1), pp.47-59. DOI: 10.1093/jge/gxad092.
- [6] Liu, Y., Guo, C., Cao, J., Cheng, Z., Ding, X., Lv, L., Li, F. and Gong, M., 2020. A new resolution enhancement method for sandstone thin-section images using perceptual GAN. *Journal of Petroleum Science and Engineering*, 195, p.107921. DOI: 10.1016/j.petrol.2020.107921.
- [7] Singh, K. and Kapoor, R., 2014. Image enhancement using exposure based sub image histogram equalization. *Pattern Recognition Letters*, 36, pp.10-14. DOI: 10.1016/j.patrec.2013.08.024.
- [8] Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F. and Zheng, Y., 2019. Recent progress on generative adversarial networks (GANs): A survey. *IEEE access*, 7, pp.36322-36333. DOI: 10.1109/ACCESS.2019.2905015.
- [9] Krichen, M., 2023, July. Generative adversarial networks. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE. DOI: 10.1109/ICCCNT56998.2023.10306417.
- [10] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27. DOI: 10.1145/3422622.
- [11] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690). DOI: 10.48550/arXiv.1609.04802.
- [12] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C., 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0). DOI: 10.48550/arXiv.1809.00219.
- [13] Wang, X., Xie, L., Dong, C. and Shan, Y., 2022. Realesrgan: Training real-world blind super-resolution with pure synthetic data supplementary material. *Computer Vision Foundation open access*, 1(2), p.2. DOI: 10.1109/ICCVW54120.2021.00217.
- [14] Shams, R., Masihi, M., Boozarjomehry, R.B. and Blunt, M.J., 2020. Coupled generative adversarial and auto-encoder neural networks to reconstruct three-dimensional multi-scale porous media. *Journal of Petroleum Science and Engineering*, 186, p.106794. DOI: 10.1016/j.petrol.2019.106794.
- [15] Shams, R., Masihi, M., Boozarjomehry, R.B. and Blunt, M.J., 2021. A hybrid of statistical and conditional generative adversarial neural network approaches for reconstruction of 3D porous media (ST-CGAN). *Advances in Water Resources*, 158, p.104064. DOI: j.advwatres.2021.104064.

- [16] Liang, Y., Wang, S., Feng, Q., Zhang, M., Cao, X. and Wang, X., 2024. Ultrahigh-resolution reconstruction of shale digital rocks from FIB-SEM images using deep learning. *SPE J.*, 29(3), pp.1434-1450. DOI: 10.2118/218397-PA.
- [17] Li, Y., Jian, P. and Han, G., 2022. Cascaded progressive generative adversarial networks for reconstructing three-dimensional grayscale core images from a single two-dimensional image. *Frontiers in Physics*, 10, p.716708.
- [18] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. DOI: 10.48550/arXiv.1409.1556.
- [19] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690). DOI: 10.48550/arXiv.1609.04802.
- [20] Arabboev, M., Begmatov, S., Rikhsivoev, M., Nosirov, K. and Saydiakbarov, S., 2024. A comprehensive review of image super-resolution metrics: classical and AI-based approaches. *Acta IMEKO*, 13(1), pp.1-8. DOI: 10.21014/actaimeko.v13i1.1679.
- [21] Valdebenito Maturana, C.N., Sandoval Orozco, A.L. and García Villalba, L.J., 2023. Exploration of Metrics and Datasets to Assess the Fidelity of Images Generated by Generative Adversarial Networks. *Applied Sciences*, 13(19), p.10637. DOI: <https://doi.org/10.3390/app131910637>.
- [22] Neumann, R.F., Barsi-Andreeta, M., Lucas-Oliveira, E., Barbalho, H., Trevizan, W.A., Bonagamba, T.J. and Steiner, M.B., 2021. High accuracy capillary network representation in digital rock reveals permeability scaling functions. *Scientific reports*, 11(1), p.11370. DOI: 10.1038/s41598-021-90090-0.

Investigating the effect of the color spectrum of input images in the process of increasing image quality using a new method based on adversarial generative networks

challenges in image processing. In modeling porous media using images obtained from computed microtomography (Micro-CT), high-quality images of these media are usually unavailable for different reasons, including the cost and computational complexity of high-quality imaging. With the advancement in the development of multiscale networks, it is possible to use more image details in these networks, and there is a further need for high-quality images. Today, adversarial generative networks are used as a practical tool in increasing image quality. These networks are trained using pairs of high-quality and low-quality images and then, they produce high-quality images by taking low-quality images as input. The final images used in modeling the porous media are binary images, but using binary images as input in the image quality enhancement process may result in detail loss. For this purpose, this study investigates the effect of using grayscale or binary images as input in the model training process. Grayscale images of a Brea sandstone sample are taken as the main input and binary images are generated using grayscale images and the realization of the real porosity of the rock sample. In this research, a new model called RealESRGAN is used to enhance the quality of the images

Keywords: Generative Adversarial Network (GAN), Real ESRGAN Model, Grayscale Image, Binary Image, Pore Network Modeling