



ادغام بدون نظارت تصاویر سنجش از دور با استفاده از شبکه‌ی مخالف مولد چند مقیاس

علی آقاگل زاده*

مهرشاد بریمانی

استادگروه مهندسی برق-مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل
دانشجوی دکترای گروه مهندسی مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل

چکیده

در پهنابندی، یک تصویر چندطیفی با وضوح پایین (LRMS) و یک تصویر تک‌رنگ با وضوح بالا (PAN) برای تولید یک تصویر چندطیفی با وضوح بالا (HRMS) ترکیب می‌شوند. مطالعات اخیر نشان داده‌اند که شبکه‌های عصبی کانولوشن می‌توانند برای پهنابندی تصاویر سنجش از دور با نتایج عالی مورد استفاده قرار گیرند. با این حال، همچنان دو مشکل اساسی وجود دارد که باید برطرف شوند. از یک سو، با وجود نبود تصاویر ایدئال HRMS برای یادگیری، اکثر روش‌های فعلی نیاز به تلاش بیشتری برای تولید داده‌های شبیه‌سازی شده دارند. از سوی دیگر، این روش‌ها معمولاً اطلاعات مکانی غنی موجود در تصاویر تک‌رنگ را نادیده می‌گیرند. برای پرداختن به این مسائل، در این مقاله، با استفاده از شبکه‌های مخالف مولد، یک چهارچوب ادغام بدون نظارت مبتنی بر یک شبکه متراکم چندمقیاسی به نام UMP-GAN برای پهنابندی تصاویر سنجش از دور طراحی شده است که می‌تواند مستقیماً بر روی تصاویر با وضوح کامل بدون نیاز به هیچ گونه پیش‌پردازشی آموزش ببیند. ابتدا، یک شبکه مولد متراکم چندمقیاسی برای استخراج ویژگی‌ها از تصاویر ورودی اصلی برای تولید تصاویر HRMS پیشنهاد می‌شود. در ادامه، دو شبکه تفکیک‌کننده مجزا برای حفاظت از اطلاعات طیفی و جزئیات مکانی موجود در تصاویر ورودی استفاده می‌گردد. بنابراین، روش پیشنهادی امکان آموزش دو شبکه تفکیک‌کننده را فراهم می‌کند، که هر کدام وظیفه متفاوت و مکملی دارند. در نهایت، برای بهبود کیفیت تصویر ادغام شده و اجرای آموزش تحت تنظیمات بدون نظارت، توابع هزینه جدیدی پیشنهاد می‌شود. این روش می‌تواند تبادل اطلاعات گرادینان میان شبکه مولد و شبکه‌های تفکیک‌کننده را عمیق‌تر کند، فرآیند آموزش شبکه را بهینه نماید و هم‌گرایی شبکه را تسریع کند. تصاویر ماهواره‌های WorldView-2 (WV2)، GaoFen-2 (GF2) و QuickBird (QB) برای شبیه‌سازی و تایید روش پیشنهادی مورد استفاده قرار گرفته‌اند. نتایج نشان داده‌اند که روش پیشنهادی در ادغام تصاویر سنجش از دور بهتر از سایر مدل‌های معتبر پیشین عمل می‌کند.

کلمات کلیدی: ادغام تصاویر سنجش از دور، بلوک متراکم چندمقیاسی، پنشارپنینگ، شبکه‌های عصبی کانولوشن، شبکه‌های مخالف مولد، یادگیری بدون نظارت.

۱- مقدمه

در سنجش از دور، محدودیت‌های زیادی در تصویربرداری ماهواره‌ای براساس سطح انرژی دریافتی حسگرها و نسبت سیگنال به نویز وجود دارد (Maneshi et al., 2020). بسیاری از ماهواره‌ها، از جمله WorldView، QuickBird و GeoFen معمولاً تنها می‌توانند دو حالت تصویربرداری را در یک زمان ارائه دهند. این حالت‌ها عبارتند از: تصویر چندطیفی با وضوح پایین^۱ (LRMS)، که پهنای باند باریک‌تر اما طیف گسترده‌ای از طول موج‌ها را پوشش می‌دهد و دارای وضوح مکانی درشت‌تر است و تصویر تک‌رنگ با وضوح بالا^۲ (PAN)، که محدوده طیفی وسیع‌تری از مرئی تا نزدیک مادون قرمز را با وضوح مکانی بالا پوشش می‌دهد اما فاقد جزئیات طیفی است (Zhou et al., 2021). بخش مهمی از کاربردهای سنجش از دور، استفاده از تصاویر چندطیفی با وضوح بالا^۳ (HRMS) است که هر دو ویژگی‌های طیفی و مکانی را حفظ می‌کنند. تصاویر HRMS به طور گسترده‌ای در زمینه‌های تحقیقاتی مانند به‌روزرسانی نقشه‌ها، نظارت بر محیط زیست، استخراج اطلاعات جاده‌ها و طبقه‌بندی پوشش گیاهی مورد استفاده قرار می‌گیرند. وضوح‌های مکانی و طیفی را نمی‌توان به صورت هم‌زمان جمع‌آوری کرد. بنابراین، ادغام تصاویر LRMS و تصویر PAN، که به خوبی به عنوان پنشارپنینگ^۴ شناخته می‌شود، توسط بسیاری از محققان مورد توجه قرار گرفته است.

در طول چند دهه گذشته، طیف وسیعی از روش‌های ادغام تصاویر برای حفظ محتوای طیفی و مکانی تصاویر ادغام شده، طراحی شده‌اند. در مقایسه با روش‌های یادگیری عمیق که به تازگی مطرح شده‌اند، این روش‌ها به عنوان روش‌های پنشارپنینگ سنتی یاد می‌شوند. این روش‌ها به طور کلی می‌توانند به دو دسته عمده طبقه‌بندی شوند: روش‌های مبتنی بر جانشینی مؤلفه‌ها (CS) و روش‌های مبتنی بر تجزیه و تحلیل چند وضوح (MRA). روش CS براساس یک تبدیل برگشت‌پذیر است که به وسیله آن تصاویر چندطیفی به فضای رنگ دیگری تبدیل می‌شوند. روش‌های اشباع-پرده رنگ-شدت (IHS)، تجزیه و تحلیل مؤلفه‌های اصلی (PCA)، گرام اشمیت (GS)، تبدیل برووی و روش جزئیات مکانی وابسته به باند (BDSD) از رایج‌ترین روش‌های مبتنی بر جانشینی مؤلفه هستند (Tu et al., 2004; Laben et al., 2000). رویکرد CS دارای مزایای پیاده‌سازی آسان،

¹ Low-Resolution Multi-Spectral image (LRMS)

² High-Resolution Panchromatic image (PAN)

³ High-Resolution Multi-Spectral image (HRMS)

⁴ Pansharpening

پیچیدگی محاسباتی کم و حفظ اطلاعات مکانی قابل قبول تصاویر PAN در تصاویر ادغام شده است. در روش‌های مبتنی بر MRA، ابتدا تصویر PAN به وسیله‌ی الگوریتم تجزیه چندمقیاسی به کانال‌های فرکانس بالا و فرکانس پایین تجزیه می‌شود و سپس جزئیات فرکانس بالا آن به تصویر LRMS تزریق می‌گردد تا تصویر نهایی ایجاد شود. روش‌های تبدیل هرم لاپلاسی (LPT)، تبدیل موجک (WT) و تبدیل کانتورلت (CT) نمونه‌هایی از روش‌های پناش‌پنینگ مبتنی بر آنالیز چند وضوح هستند (Shah et al., 2008) در فرآیند ادغام تصویر، رویکرد MRA جزئیات طیفی را بهتر از رویکرد CS حفظ می‌کند، اما دارای مشکل اعوجاج مکانی است.

در چند سال اخیر با ظهور روش‌های یادگیری عمیق، شبکه‌های عصبی کانولوشن (CNNs) به طور گسترده‌ای در بینایی کامپیوتری، تشخیص الگو و تحقیقات پردازش تصویر به کار گرفته شده‌اند و ثابت شده است که می‌توانند به نتایج بهتری نسبت به روش‌های سنتی دست یابند (Naji et al., 2020; Naji et al., 2019). این روش‌ها توانایی زیادی در مدل‌سازی نگاشت‌های غیرخطی پیچیده دارند و برای ادغام تصاویر چندطیفی به کار گرفته شده‌اند. در نتیجه منجر به توسعه بسیاری از مدل‌های یادگیری عمیق در حوزه‌ی پناش‌پنینگ شده است. علاوه بر این، پیاده‌سازی مدل‌های تحت نظارت پناش‌پنینگ مبتنی بر فراتفکیک‌پذیری (Ledig et al., 2017) با استفاده از مدل‌های مولدی هم‌چون خودرمنزنگارهای متغیر (VAEs) (Kingma et al., 2014) و شبکه‌های مخالف مولد (GANs) (Goodfellow et al., 2014) انجام می‌شوند. ماسی و همکارانش (Masi et al., 2016) یک شبکه CNN سه لایه به نام PNN را برای اولین بار در حوزه‌ی پناش‌پنینگ پیشنهاد کرده‌اند. معماری این شبکه در اصل بر مبنای مسأله فراتفکیک‌پذیری پیشنهاد شده است. در مسأله پناش‌پنینگ، حفظ اطلاعات هندسی و رنگ تصاویر ورودی در تصویر ادغام شده از اهمیت بالایی برخوردار است. یانگ و همکارانش (Yang et al., 2017) برای دستیابی به این هدف شبکه PanNet را ارائه کرده‌اند. این روش دارای یک معماری شبکه عمیق است که از مفهوم اتصال‌های پرشی ResNet (He et al., 2016) برای ایجاد شبکه‌های عمیق‌تر و آموزش مدل شبکه خود در حوزه‌های فرکانس بالا با هدف حفظ ویژگی‌های هندسی و ساختاری تصویر PAN استفاده می‌کند. وانگ و همکارانش (Wang et al., 2019) مدل

DCCNN را که یک مدل مبتنی بر بلوک‌های CNN متراکم متصل و یادگیری باقی‌مانده است برای پنهان‌پنینگ تصاویر چندطیفی پیشنهاد کرده‌اند. پنگ و همکارانش (Peng et al., 2021) یک روش جدید پنهان‌پنینگ را براساس یک شبکه متراکم چندمقیاسی ارائه کرده‌اند، که از فیلترهای با اندازه‌های مختلف برای استخراج ویژگی‌های تصاویر ورودی استفاده می‌کند. علی‌رغم عمل کرد ادغامی قابل قبول مدل‌های فوق، برخی چالش‌ها همچنان برطرف نشده‌اند. به منظور تنظیم پارامترهای شبکه، این روش‌ها نیاز به یک تصویر مرجع^۱ دارند که در عمل موجود نیست. به بیانی دیگر، تمام روش‌های مبتنی بر یادگیری عمیق مذکور با استفاده از اطلاعات تصاویر مرجع پیاده‌سازی می‌شوند. در واقع این مدل‌ها در مرحله آموزش، بازسازی یک تصویر HRMS از تصاویر PAN و LRMS با وضوح کاهش یافته به عنوان ورودی را با استفاده از پروتکل والد یاد می‌گیرند (Wald et al., 1997). در این پروتکل به منظور بهینه‌سازی شبکه‌ها، این روش‌ها ابتدا تصاویر PAN و LRMS ورودی را کوچک می‌کنند و سپس تصویر LRMS اصلی را به عنوان تصویر مرجع به منظور تشکیل نمونه‌های آموزشی در نظر می‌گیرند. در مرحله آزمایش، ارزیابی این مدل‌های پیشنهادی بر روی تصاویر کوچک شده انجام می‌شود. استفاده از این پروتکل برای تصاویر سنجش از دور ممکن است منجر به شکاف بین تصاویر کوچک‌شده و تصاویر اصلی گردد. برخلاف تصاویر طبیعی، تصاویر سنجش از دور معمولاً در عمق بیت عمیق‌تر گرفته می‌شوند و دارای توزیع پیکسلی متفاوتی هستند. در نتیجه، اگرچه این مدل‌های پنهان‌پنینگ مبتنی بر CNN نظارت شده ممکن است عمل کرد خوبی در هنگام آزمایش بر روی تصاویر کوچک شده داشته باشند، اما آن‌ها به تصاویر با وضوح کامل تعمیم نمی‌یابند. چالش موجود دیگر این است که مدل‌های مذکور عمدتاً اطلاعات مکانی غنی تصاویر PAN را در تولید تصاویر HRMS نادیده می‌گیرند.

به عنوان یک راه حل برای چالش‌های موجود، در این مقاله، با الهام از شبکه متراکم چندمقیاسی، یک چهارچوب بدون نظارت برای ادغام تصویر مبتنی بر GAN پیشنهاد می‌گردد که UMP-GAN نامیده می‌شود. در این روش پیشنهادی، متفاوت از مقالات پیشین (Zhou et al., 2021)، از یک شبکه مولد متراکم چندمقیاسی جدید برای استخراج ویژگی‌های تصاویر سنجش از دور استفاده می‌شود. هدف روش پیشنهادی، حفظ اطلاعات مهم تصاویر ورودی با استفاده از یادگیری بدون نظارت و استفاده از یک فرآیند یادگیری رقابتی مداوم بین یک شبکه مولد و دو شبکه تفکیک‌کننده است. به طور خاص، به منظور غلبه بر مشکل اطلاعات مکانی - طیفی از دست رفته، چهارچوب شبکه مولد پیشنهادی شامل دو کدکننده و یک کدگشا است. در این شبکه، ابتدا

¹ Ground Truth

ویژگی‌های مهم از تصاویر ورودی مختلف از طریق کدکننده دوگانه استخراج می‌شوند و سپس از آن‌ها به وسیله‌ی کدگشا، برای بازسازی تصاویر استفاده می‌شود تا یک تصویر HRMS مطلوب تولید گردد. این معماری به طور مؤثر ویژگی‌های عمیق‌تر تصاویر ورودی را با استفاده از لایه‌های کانولوشن و اتصال‌های پرشی استخراج می‌کند. همانگونه که بیان شد این مدل پیشنهادی از دو نوع شبکه تفکیک‌کننده استفاده می‌کند: یک شبکه تفکیک‌کننده طیفی برای متمایز کردن تصویر ادغام شده از تصاویر LRMS ورودی و یک شبکه تفکیک‌کننده مکانی، که تلاش می‌کند تا بین تصویر ادغام شده و تصویر PAN ورودی تمایز ایجاد کند. به منظور استفاده حداکثر از اطلاعات داده‌های مکانی و طیفی ورودی، روش پیشنهادی UMP-GAN بر یادگیری بدون نظارت بدون هیچ مرحله پیش‌پردازشی تمرکز دارد. علاوه بر این، برای بهبود کیفیت تصویر ادغام شده، یک تابع هزینه جدید براساس شاخص ارزیابی غیر مرجع QNR ارائه شده است (Wang et al., 2002). در این مقاله، ما با بهینه‌سازی مدل شبکه خود، برخی از مشکلات مربوط به شبکه‌های مخالف مولد استاندارد، مانند فرآیند آموزش ناپایدار و هم‌چنین ابهام و دشواری در هم‌گرایی را بهبود بخشیده‌ایم. نتایج آزمایش نهایی نیز عمل‌کرد خیلی خوب الگوریتم پیشنهادی را نشان می‌دهد.

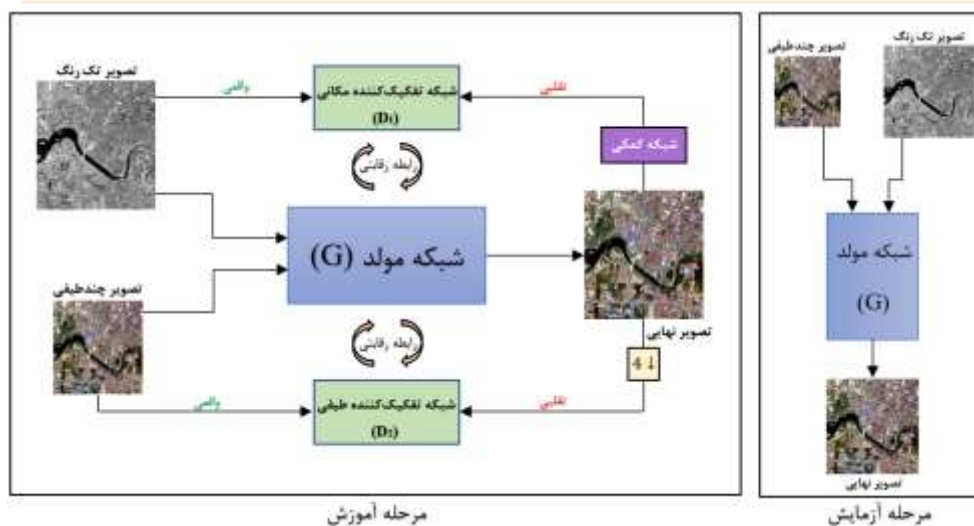
۲- مواد و روش پیشنهادی

۲-۱- مروری بر مدل پیشنهادی UMP-GAN

در حال حاضر، روش پناش‌پینگ مبتنی بر یادگیری عمیق به یک موضوع داغ تحقیقاتی تبدیل شده است. برخلاف روش‌های سنتی ادغام تصاویر سنجش از دور، روش‌های مبتنی بر یادگیری عمیق بر استخراج حداکثر ویژگی‌ها تمرکز دارند، بنابراین می‌توانند تصاویر HRMS بهتری به دست آورند. با این حال، روش‌های پناش‌پینگ مبتنی بر یادگیری عمیق موجود، مشکلاتی دارند که نمی‌توانند اطلاعات را به طور کامل در مقیاس‌های مختلف تصاویر استخراج کنند و برای آموزش به تصاویر مرجع نیاز دارند. به منظور غلبه بر اشکالات روش‌های موجود مبتنی بر CNN، در این مقاله، یک الگوریتم پناش‌پینگ جدید را با استفاده از ایده تزریق جزئیات چندمقیاسی معرفی می‌شود که ویژگی‌های مکانی و طیفی ارائه شده توسط تصاویر ورودی را به دقت در تصویر نهایی حفظ می‌کند. این کار با توسعه یک روش یادگیری بدون نظارت برای تصاویر سنجش از دور انجام می‌شود که در آن تصاویر اصلی برای آموزش شبکه عصبی به کار گرفته می‌شوند. به طور کلی، تصاویر منبع مورد استفاده در این مقاله تصاویر PAN یک بانده و تصاویر LRMS چهار بانده هستند. همانطور که

در شکل ۱ نشان داده شده است، مدل یادگیری UMP-GAN پیشنهادی ما بر حفظ جزئیات مکانی تصویر PAN اصلی و هم‌چنین اطلاعات طیفی تصاویر LRMS اصلی با فرموله کردن این مسأله به دو مسأله‌ی مجزا و استفاده از یک شبکه مخالف مولد برای حل آن تمرکز می‌کند. با استفاده از روش پیشنهادی ما، به جای استفاده از شبکه‌های مخالف مولد استاندارد برای ایجاد یک بازی رقابتی میان یک شبکه مولد و یک تفکیک کننده، ما یک بازی کمینه - بیشینه را بسته به وظایف خاصی که باید حل شوند بین یک شبکه مولد (G) و دو شبکه تفکیک کننده مجزا، یعنی شبکه تفکیک کننده طیفی (D₁) و شبکه تفکیک کننده مکانی (D₂)، می‌سازیم. مدل پیشنهادی از یک شبکه متراکم چندمقیاسی در طراحی شبکه مولد برای استخراج اطلاعات بیشتر استفاده می‌کند.

در طول فرآیند آموزش، G می‌تواند به صورت هم‌زمان با D₁ و D₂ همگرا شود. برای ایجاد یک تصویر HRMS که اطلاعات طیفی آن با تصاویر LRMS اصلی سازگار است، G تلاش می‌کند تا یک بازی رقابتی دو نفره با D₁ ایجاد کند. با استفاده از درون‌یابی دوقطبی، تصاویر HRMS تولید شده با وضوح مشابه تصویر LRMS اصلی درون‌یابی و به D₁ وارد می‌شوند تا توزیع طیفی تصویر HRMS تولید شده با توزیع تصویر LRMS اصلی مطابقت داشته باشد. علاوه بر این، به منظور حفظ بهتر اطلاعات طیفی، یک تطبیق هیستوگرام بین نتایج درون‌یابی شده و تصاویر LRMS انجام می‌گیرد و به‌طور مشخص‌تر، هیستوگرام تصاویر ادغام شده باید مشابه با تصاویر LRMS اصلی باشد. در مرحله بعد، یک بازی دو نفره دیگر بین G و D₂ ایجاد می‌شود. در نتیجه، D₂ تصویر HRMS را مجبور خواهد کرد تا اطلاعات مکانی بیشتری از تصویر PAN اصلی به دست آورد. با الهام از PGMAN (Zhou et al., 2021)، از یک شبکه کمکی تجزیه طیفی برای به دست آوردن یک تصویر تک کاناله از تصویر HRMS تولید شده توسط شبکه مولد استفاده می‌شود. این شبکه متشکل از تنها یک لایه کانولوشن ۳ × ۳ برای برآورد تبدیل تصویر HRMS به تصاویر PAN است. شبکه کمکی به صورت جداگانه با تصاویر LRMS و PAN کوچک شده آموزش داده می‌شود و در طول آموزش مدل UMP-GAN ثابت خواهد ماند. این تصویر تک کاناله و تصویر PAN وارد D₂ می‌شوند، که به ما اجازه می‌دهد تا توزیع مکانی تصویر ادغام شده مشابه تصویر PAN اصلی را حفظ نماییم.



شکل (۱): چهارچوب کلی مدل پیشنهادی UMP-GAN برای پنه‌شارپ‌نینگ.

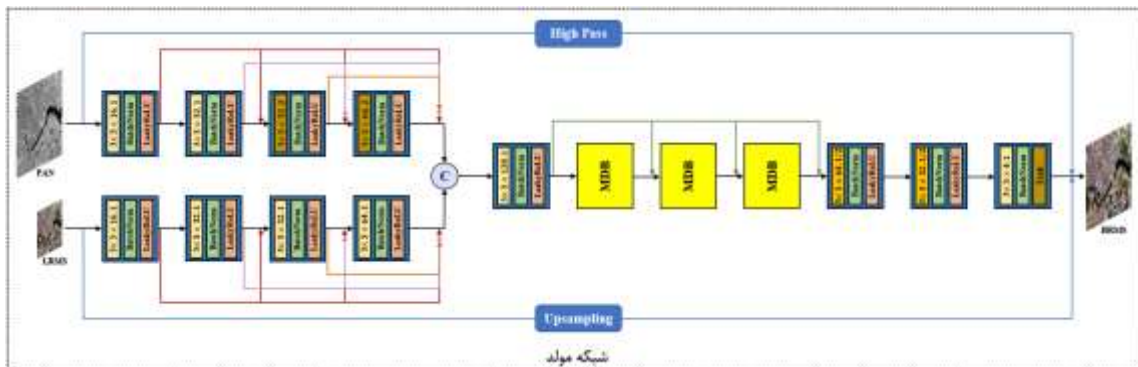
Fig. (1): General framework of the proposed UMP-GAN model for pensharpening

به این ترتیب، G اطلاعات طیفی تصاویر LRMS اصلی را حفظ می‌کند در حالی که هم‌زمان توزیع مکانی تصویر PAN را در تصویر HRMS از طریق یادگیری رقابتی متناسب می‌سازد. اگر D_1 نتواند بین نسخه کوچک شده تصویر تولید شده و تصویر LRMS اصلی تمایز قائل شود و هم‌چنین D_2 نتواند بین تصویر تولید شده و تصویر PAN اصلی تمایز قائل شود، نتیجه گرفته می‌شود که تصویر تولید شده تا حد امکان توانسته است جزئیات مکانی و طیفی تصاویر PAN و LRMS ورودی را در تصویر HRMS تزریق نماید. فرآیند استنتاج تنها به مدل شبکه مولد آموزش دیده نیاز دارد. بنابراین در طول مرحله آزمایش، ویژگی‌های استخراج شده از تصاویر ورودی PAN و LRMS برای تولید تصاویر HRMS وارد G می‌گردند. در ادامه، ابتدا توضیح مفصلی از معماری‌های شبکه مولد چندمقیاسی و شبکه‌های تفکیک‌کننده به صورت جداگانه ارائه می‌شود. سپس توابع هزینه معرفی می‌گردند. درنهایت، برخی جزئیات پیاده‌سازی را در فرآیند آموزش به اشتراک می‌گذاریم.

۲-۲- معماری شبکه مولد پیشنهادی

چندین معماری مبتنی بر CNN وجود دارند که می‌توانند برای پیاده‌سازی شبکه مولد ما انتخاب شوند. به منظور حفظ جزئیات مکانی و طیفی بیشتر در تصویر ادغام شده، در این مقاله از یک معماری اصلاح شده TFNet (Liu et al., 2020) استفاده می‌شود. علاوه بر این، ما معماری PGMAN (Zhou et al., 2021) را با اضافه کردن برخی اتصالات های پرشی الهام گرفته از DenseNet (Huang et al., 2017) به روز کرده‌ایم. این اتصالات می‌توانند اطلاعات بیشتری را به لایه‌های بعدی انتقال دهند تا جزئیات مکانی و طیفی به طور مؤثر به تصاویر HRMS تزریق شوند و فرآیند آموزش را کارآمدتر نمایند. به بیانی دیگر، اتصال Dense به این معنی است که ورودی هر لایه نتیجه آبخاری از تمام خروجی‌های لایه‌های قبلی است. نتایج تجربی نیز اعتبار این اتصالات های پرشی را تایید می‌کنند. همانطور که در شکل ۲ نشان داده شده است، شبکه مولد پیشنهادی از یک ساختار دوگانه کدکننده - تک کدگشا استفاده می‌کند. برای استفاده بهینه از جزئیات مکانی و اطلاعات طیفی و هم‌چنین با توجه به اینکه تصاویر PAN و LRMS اندازه‌های متفاوتی دارند، کدکننده دوگانه شامل دو مسیر مستقل است که برای استخراج ویژگی‌های سلسله مراتبی دو تصویر ورودی، استفاده می‌شود. در مدل پیشنهادی چهار لایه برای استخراج ویژگی در هر مسیر کدکننده در نظر گرفته می‌شوند. در گام بعد خروجی‌های کدکننده دوگانه ترکیب شده و نتیجه به بخش کدگشا وارد می‌شود تا تصویر ادغام شده بازسازی شود. ویژگی‌های استخراج شده در بخش کدکننده‌ها به عنوان ورودی در بخش کدگشا استفاده می‌شوند. بخش کدگشا برای یادگیری ویژگی‌های سلسله مراتبی چندمقیاسی برای ایجاد یک تصویر ادغام شده با وضوح مکانی و طیفی بالا طراحی شده است. با الهام از مقاله (Peng et al., 2021) در این بخش، لایه‌های استخراج ویژگی چندمقیاسی از سه بلوک متراکم چندمقیاسی (MDB) برای استخراج اطلاعات غنی و متراکم از ویژگی‌های سطحی استفاده می‌نماید. معماری شبکه MDB در شکل ۳ نشان داده شده است. هر MDB شامل چهار لایه است. سه لایه اول لایه اتصال متراکم چندمقیاسی هستند و لایه آخر یک لایه ادغام ویژگی محلی است.

¹ Multi-scale Dense Block (MDB)



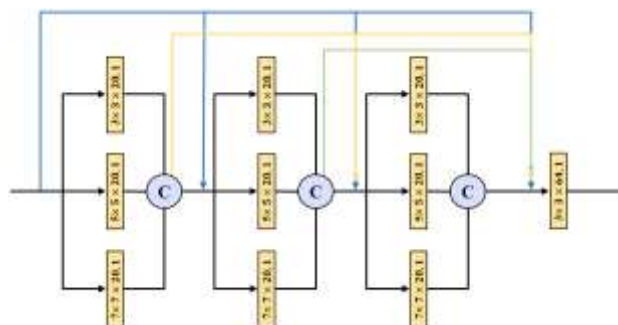
شکل (۲): معماری شبکه مولد پیشنهادی

Fig. (2): Proposed generator network architecture

همانطور که در شکل ۳ نشان داده شده است، در هر لایه اتصال متراکم چندمقیاسی هسته‌های کانولوشن در اندازه‌های 3×3 ، 5×5 و 7×7 تنظیم شده‌اند و ۲۰ نگاشت ویژگی برای هر یک در نظر گرفته شده است. ویژگی‌های به دست آمده در طول بُعد طیفی ترکیب می‌شوند تا خروجی هر لایه را تشکیل دهند. به این ترتیب، ویژگی‌های تصویر در مقیاس‌های مختلف می‌تواند به طور تطبیقی توسط هسته‌های کانولوشن در اندازه‌های مختلف آموخته شوند، در نتیجه غنای ویژگی‌های استخراج شده افزایش می‌یابد. برای استخراج ویژگی‌های کارآمدتر و همچنین استفاده مجدد از ویژگی‌های محلی قبلی و فعلی، اتصال متراکم محلی بین لایه‌های مجاور قرار داده می‌شود که امکان اتصال مستقیم لایه قبلی به تمام لایه‌های بعدی را فراهم می‌کند. سپس، ویژگی‌های سطح پایین و بالا استخراج شده به لایه آخر داده می‌شوند تا تعداد ویژگی‌ها به ۶۴ کاهش یابد. لایه آخر در MDB دارای یک لایه کانولوشن با هسته 3×3 است که می‌تواند به طور تطبیقی ویژگی‌های MDB قبلی و هر لایه MDB فعلی را ترکیب کند. به این ترتیب با کمک بلوک‌های MDB، ویژگی‌های سطح پایین می‌توانند به خوبی با عمیق‌تر شدن شبکه حفظ شوند.

پس از بلوک‌های MDB، دو لایه کانولوشن متوالی که هر دو با هسته 3×3 و گام $\frac{1}{2}$ هستند برای رسیدن به اندازه تصویر HRMS مورد نظر اعمال می‌شوند. در نهایت یک لایه کانولوشن با هسته 3×3 و گام ۱ به بازسازی تصویر ادغام شده کمک می‌نماید. یادگیری باقی‌مانده میان تصاویر ورودی و تصویر نهایی می‌تواند فرآیند آموزش را تثبیت کند و همچنین سرعت هم‌گرایی شبکه و کاهش مصرف حافظه ناشی از ساختارهای عمیق‌تر را بهبود بخشد. در نتیجه به طور خاص برای حفاظت مکانی و طیفی بهتر، تصاویر LRMS بزرگنمایی

شده و اطلاعات فرکانس بالا از تصاویر PAN، به خروجی کدگشا اضافه می‌شود. به منظور سرعت بخشیدن به آموزش و هم‌گرایی مدل پیشنهادی، هر عملیات کانولوشن شامل نرمال‌سازی دسته‌ای^۱ و تابع فعال‌سازی LeakyReLU برای جلوگیری از ناپدید شدن گرادیان و کاهش حساسیت به داده‌های اولیه است. به منظور نگاشت شدت‌های تصویر ادغام شده به بازه $[-1, 1]$ ، تابع فعال‌سازی Tanh در لایه آخر به کار گرفته می‌شود. با استفاده از Tanh در مدل پیشنهادی، می‌توان شبکه‌های مخالف مولد را سریع‌تر و مؤثرتر آموزش داد. به این ترتیب، یک تصویر مطلوب از شبکه مولد پیشنهادی به عنوان خروجی تولید می‌شود.



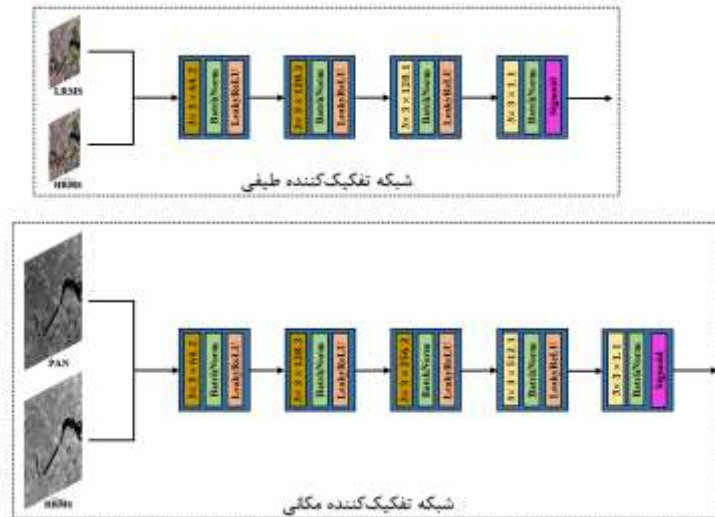
شکل (۳): معماری بلوک‌های MDB

Fig. (3): Architecture of MDB blocks

۳-۲- معماری شبکه‌های تفکیک‌کننده پیشنهادی

برخلاف شبکه مولد، شبکه‌های تفکیک‌کننده تلاش می‌کنند تا تصویر واقعی را از تصویر جعلی متمایز نمایند، در نتیجه ساختار آن‌ها ساده‌تر است. در این مقاله، برای حفظ اطلاعات ورودی و اطمینان از سازگاری در طول فرآیند پناش‌پنینگ، از دو شبکه تفکیک‌کننده در معماری پیشنهادی، یعنی شبکه تفکیک‌کننده طیفی (D1) و شبکه تفکیک‌کننده مکانی (D2)، مطابق با ویژگی‌های خاص ادغام تصاویر سنجش از دور استفاده می‌شود. شکل ۴ معماری هر یک از شبکه‌های تفکیک‌کننده را نشان می‌دهد. شبکه تفکیک‌کننده D1 را در نظر بگیرید که تصویر ایجاد شده و تصویر هدف را با یکدیگر مقایسه می‌کند تا رنگ و وضوح طیفی در تصویر نهایی حفظ گردد.

¹ Batch Normalization



شکل (۴): معماری شبکه‌های تفکیک‌کننده پیشنهادی

Fig. (4): Proposed network architecture for discriminators

اولین گام این است که تصویر HRMS را به همان وضوح مکانی تصویر LRMS اصلی کوچک نماییم تا تصویر HRMS به دست آید و سپس D_1 برای اطمینان از داشتن اطلاعات طیفی مشابه به کار برده شود. D_1 ارتفاع و عرض را کاهش می‌دهد در حالی که عمق را با استفاده از دو عملیات کانولوشن 3×3 با گام دو افزایش می‌دهد. پس از آن، عمق تصویر با عملیات کانولوشن به مقدار ۱ کاهش می‌یابد. در گام بعد، شبکه تفکیک‌کننده D_2 برای مطابقت ساختار مکانی تصویر HRMS تک کاناله (تصویر HRMS) با ساختار مکانی تصویر PAN ورودی به کار می‌رود. در D_2 ، سه لایه کانولوشن با اندازه 3×3 و گام دو، و سپس دو لایه کانولوشن وجود دارد که عمق را به مقدار ۱ کاهش می‌دهد. از آنجا که تصاویر PAN اندازه بزرگتری دارند، ساختار شبکه‌ای D_2 عمیق‌تر از D_1 طراحی شده است. برای سرعت بخشیدن به هم‌گرایی در فاز آموزش، از نرمال‌سازی دسته‌ای در هر لایه کانولوشن استفاده می‌شود و تابع فعال‌سازی LeakyReLU برای جلوگیری از پراکندگی گرادیان است. با اعمال تابع فعال‌سازی سیگموئید در لایه‌ی آخر، فاصله به بازه‌ی [۰، ۱] کاهش می‌یابد. معماری پیشنهادی UMP-GAN دارای شبکه‌های تفکیک‌کننده‌ای مشابه با معماری patchGAN شرطی است که از تکه‌های تصویر به عنوان ورودی استفاده می‌کند. این نوع شبکه‌های تفکیک‌کننده که بر مبنای تکه‌های تصاویر نتیجه‌گیری می‌کنند در مقایسه با شبکه‌های تفکیک‌کننده تصویر کامل، پارامترهای کمتری دارند. از

آنجا که این تفکیک‌کننده‌ها دو ویژگی مکمل تصاویر ورودی را شناسایی می‌نمایند، تصویر تولید شده برای وفاداری و مطابقت مکانی و طیفی براساس توابع هزینه ارائه شده بررسی می‌شود. هیچ تداخل رقابتی میان این دو شبکه تفکیک‌کننده وجود ندارد. هر دو شبکه تفکیک‌کننده خروجی یکسانی دارند، که همان نتایج طبقه‌بندی است، بنابراین شبکه مولد می‌تواند یک تصویر HRMS مطلوب تولید نماید. همان طور که مشاهده می‌شود، زیربنای شبکه مولد و شبکه‌های تفکیک‌کننده پیشنهادی، یک شبکه عصبی کاملاً کانولوشنی است، که باعث می‌شود مدل UMP-GAN پیشنهاد شده، به راحتی آموزش ببیند و به اندازه کافی انعطاف‌پذیر باشد تا تصاویر PAN و LRMS را با اندازه‌های دل‌خواه در مرحله آزمایش بپذیرد.

۲-۴- تابع هزینه مدل پیشنهادی

تابع هزینه:

تابع هزینه مدل UMP-GAN از دو جزء تشکیل شده است: تابع هزینه شبکه مولد (\mathcal{L}_G) و تابع هزینه شبکه‌های تفکیک‌کننده (\mathcal{L}_{D_1} و \mathcal{L}_{D_2}). در ادامه به ترتیب این دو بخش معرفی می‌شوند.

(۱) **تابع هزینه شبکه مولد:** برای یادگیری وزن‌های شبکه G ، مدل پیشنهادی از تابع هزینه زیر استفاده می‌کند:

$$\mathcal{L}_G = \mathcal{L}_{spectral} + \mathcal{L}_{spatial} + \mathcal{L}_{QNR} \quad (1)$$

بخش اول $\mathcal{L}_{spectral}$ در رابطه \mathcal{L}_G نشان دهنده تابع هزینه طیفی بین جزئیات طیفی تصویر ادغام شده و تصویر LRMS ورودی است، که می‌تواند به شکل زیر نوشته شود:

$$\mathcal{L}_{spectral} = \alpha \frac{1}{N} \sum_{n=1}^N (D_1(\hat{P}^{(n)}) - c)^2 + \frac{1}{N} \sum_{n=1}^N \|\hat{P}^{(n)} - X^{(n)}\|_F^2 \quad (2)$$

که در آن X و $\hat{P}^{(n)}$ به ترتیب نسخه تجزیه شده طیفی تصویر تولید شده و تصویر LRMS ورودی با $n \in \mathbb{N}_N$ و \mathbb{N} تعداد نمونه‌ها را نشان می‌دهند، و $D_1(n)$ نتیجه حاصل از شبکه تفکیک‌کننده D_1 است. α نشان دهنده پارامتر تنظیم‌کننده‌ای است که برای متعادل کردن دو تابع هزینه استفاده می‌شود و c ابر پارامتری است که برای نشان دادن مقدار جعلی استفاده می‌شود که G می‌خواهد D_1 داده‌های غلط تولید شده را به عنوان مقادیر واقعی تشخیص دهد، و همچنین $\|\cdot\|_F$ نرم ماتریس فروبنیوس است. در این معادله، چندجمله‌ای اول در سمت راست تابع هزینه رقابتی طیفی و چندجمله‌ای دوم تابع هزینه اطلاعات طیفی است.

بخش دوم $\mathcal{L}_{spatial}$ در رابطه \mathcal{L}_G نشان دهنده تابع هزینه مکانی بین جزئیات مکانی تصویر ادغام شده و تصویر PAN ورودی است، که می‌تواند به شکل زیر نوشته شود:

$$\mathcal{L}_{spatial} = \beta \frac{1}{N} \sum_{n=1}^N (D_2(\tilde{P}^{(n)}) - d)^2 + \frac{1}{N} \sum_{n=1}^N \|\nabla \tilde{P}^{(n)} - \nabla Y^{(n)}\|_F^2 \quad (3)$$

که در آن $\tilde{P}^{(n)}$ و Y به ترتیب نسخه تجزیه شده مکانی تصویر تولید شده و تصویر PAN ورودی را نشان می‌دهند، β پارامترهای تنظیم‌کننده مورد استفاده برای تعادل دو تابع هزینه را نشان می‌دهد. $D_2(n)$ نتیجه حاصل از شبکه تفکیک‌کننده D_2 است و ∇ نشان دهنده عملیات فیلترینگ بالاگذر است که برای استخراج اطلاعات مکانی در فرکانس‌های بالا به کار می‌رود، و همچنین از ابر پارامتر d برای نشان دادن مقدار جعلی که G می‌خواهد D_2 داده‌های غلط تولید شده را به عنوان مقادیر واقعی تشخیص دهد، استفاده می‌گردد. در این معادله، چندجمله‌ای اول در سمت راست تابع هزینه رقابتی مکانی و چندجمله‌ای دوم تابع هزینه اطلاعات مکانی است.

برای آموزش یک شبکه، رویکردهای مبتنی بر یادگیری نظارت شده معمولاً تنها از توابع هزینه $L1$ یا $L2$ استفاده می‌کنند. با این حال، در حالت آموزش بدون نظارت، هیچ تصویر ایده‌آلی برای مقایسه در دسترس نیست. در این مقاله، یک راه‌حل جایگزین پیشنهاد میشود تا کیفیت تصاویر ادغام شده با ارجاع به ورودی‌ها به جای تصویر مرجع، سنجیده شود. در نتیجه، L_{QNR} به عنوان تابع هزینه مکمل شبکه G اضافه می‌شود تا کیفیت طیفی و اطلاعات مکانی تصویر تولید شده بهبود یابد. در این روش، فرض می‌شود که پس از ادغام، کیفیت طیفی تصویر بدون تغییر باقی می‌ماند. با استفاده از شاخص QNR (Wang et al., 2002) به عنوان مبنای تابع هزینه به صورت زیر تعریف می‌شود:

$$L_{QNR} = 1 - QNR \quad (4)$$

در مسأله پهنارپنینگ QNR (با مقدار بهینه ۱) به طور گسترده‌ای به کار گرفته می‌شود که شامل هر دو شاخص اعوجاج طیفی (D_λ) و شاخص اعوجاج مکانی (D_s) است (Wang et al., 2002):

$$QNR = (1 - D_\lambda)(1 - D_s) \quad (5)$$

که در آن D_λ به صورت زیر تعریف می‌شود:

$$D_\lambda = \sqrt{\frac{2}{L(L-1)} \sum_{i=1}^L \sum_{\substack{j=1 \\ i \neq j}}^L |Q(P_i, P_j) - Q(X_i, X_j)|} \quad (6)$$

که P_i و X_i به ترتیب نشان دهنده باند نام تصویر تولید شده و تصویر LRMS ورودی و L نشان دهنده تعداد باندها است. Q اشاره به شاخص کیفیت تصویر (QI) دارد که می‌تواند به شکل زیر نوشته شود:

$$Q = \frac{4\sigma_{xy} \cdot \bar{x} \cdot \bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)} \quad (7)$$

که در آن σ_{xy} نشان دهنده کوواریانس میان تصاویر X و Y است، \bar{x} و \bar{y} میانگین هستند، و σ_x^2 و σ_y^2 به ترتیب نشان دهنده واریانس آن‌ها است. شاخص QI (Wang et al., 2002) شباهت‌های آماری بین دو تصویر تک‌رنگ را اندازه‌گیری می‌کند. ایده این است که پس از ادغام، زمانی که جزئیات طیفی از مقیاس‌های درشت به ریز تبدیل می‌شوند، مقادیر QI باید بدون تغییر باقی بمانند. D_s به صورت زیر تعریف می‌شود:

$$D_s = \sqrt{\frac{1}{L} \sum_{i=1}^L |Q(P_i, Y) - Q(X_i, \tilde{Y})|} \quad (8)$$

که Y نشان دهنده تصویر PAN ورودی و \tilde{Y} نسخه با وضوح پایین و تجزیه شده آن است. با استفاده از این تابع هزینه، می‌توان کیفیت تصاویر تولید شده را براساس تصاویر ورودی PAN و LRMS بدون استفاده از تصاویر مرجع HRMS تعیین نمود.

(۲) تابع هزینه شبکه‌های تفکیک‌کننده: در این مقاله از یک چهارچوب شبکه مبتنی بر (LSGAN) Mao (et al., 2017) استفاده می‌شود که از دو شبکه تفکیک‌کننده برای بهبود فرآیند آموزش استفاده شود: شبکه D_1 برای حفظ اطلاعات طیفی و شبکه D_2 برای حفظ اطلاعات مکانی. شبکه G یاد می‌گیرد که جزئیات طیفی بیشتری را برای به اشتباه انداختن شبکه D_1 حفظ کند، در نتیجه می‌تواند بین تصاویر LRMS واقعی و جعلی تمایز قائل شود، و همچنین جزئیات مکانی بیشتری را برای به اشتباه انداختن شبکه D_2 حفظ کند تا بتواند بین تصویر PAN واقعی و تصویر نهایی طیفی تمایز قائل گردد. توابع هزینه‌ی شبکه‌های تفکیک‌کننده به صورت زیر تعریف می‌شوند:

$$\mathcal{L}_{D_1} = \frac{1}{N} \sum_{n=1}^N (D_1(\hat{P}^{(n)}) - b)^2 + \frac{1}{N} \sum_{n=1}^N (D_1(X^{(n)}) - a)^2 \quad (9)$$

$$\mathcal{L}_{D_2} = \frac{1}{N} \sum_{n=1}^N (D_2(\tilde{P}^{(n)}) - b)^2 + \frac{1}{N} \sum_{n=1}^N (D_2(Y^{(n)}) - a)^2 \quad (10)$$

که در آن‌ها \hat{P} ، \tilde{P} ، X و Y به ترتیب نسخه‌های تجزیه شده طیفی و مکانی تصویر تولید شده، تصویر LRMS ورودی و تصویر PAN ورودی هستند. $D(n)$ نتیجه حاصل از شبکه تفکیک‌کننده D است که نتایج طبقه‌بندی تصویر هدف و تصویر ادغام شده را نشان می‌دهد و a و b به ترتیب برچسب‌های آن‌ها هستند. شبکه D_1 برای تمایز بین نسخه تجزیه شده تصویر ادغام شده و تصاویر LRMS ورودی طراحی شده است. علاوه بر این، براساس این فرض که توزیع جزئیات طیفی با مقیاس تغییر نخواهد کرد، جزئیات طیفی تصویر ادغام شده باید یک توزیع قابل مقایسه با تصویر LRMS داشته باشد. همچنین شبکه D_2 برای تمایز بین نسخه تجزیه شده تصویر ادغام شده و تصویر PAN ورودی طراحی شده است.

هدف ما طراحی توابع هزینه‌ای است که به طور عینی و مؤثر فاصله بین داده‌های واقعی و جعلی را کاهش دهد. هنگامی که این فاصله به نقطه‌ای کاهش یابد که شبکه‌های تفکیک‌کننده دیگر نتواند تصویر ادغام شده را از تصاویر ورودی تشخیص دهند، آموزش شبکه به پایان می‌رسد.

۲-۵- جزئیات پیاده‌سازی روش پیشنهادی

در این مقاله، الگوریتم UMP-GAN بر روی Tensorflow نسخه ۱/۵ پیاده‌سازی شد و روی یک GPU NVIDIA GeForce RTX3060 با حافظه ۳۲ گیگابایتی آموزش داده شده است. برخلاف روش‌های آموزشی سنتی، ابتدا شبکه‌های G و D₁ آموزش داده می‌شوند تا فرآیند آموزش سریعتر شود. پس از اینکه شبکه D₁ به هم‌گرایی دست یافت، شبکه D₂ به مدل پیشنهادی اضافه می‌شود و به‌طور هم‌زمان آموزش داده می‌شوند. مدل پیشنهادی UMP-GAN در بیش از ۸۰ دوره^۱ آموزش داده شده و پس از اتمام آموزش، بالاترین وضعیت عمل‌کرد مدل در آخرین دوره‌ها انتخاب شد. اندازه دسته^۲ و نرخ یادگیری^۳ در آزمایش‌های انجام گرفته به ترتیب ۱۶ و ۰/۰۰۰۱ تعیین شده‌اند. بهینه‌ساز adam (Kingma et al., 2017) با ابر پارامترهای ثابت β_1 برابر با ۰/۵ و β_2 برابر با ۰/۹ به منظور بهینه‌سازی وزن شبکه‌های به‌کار گرفته شده‌اند. در مدل پیشنهادی از کاهش وزن^۴ استفاده نشده است، زیرا استفاده از آن موجب کاهش عمل‌کرد ادغام تصویر می‌گردد. با توجه به مقاله (Ma et al., 2020) و مطالعات تجربی، ابر پارامترها در معادلات ۲ و ۳ به صورت $\alpha = 0.002$ و $\beta = 0.001$ تنظیم شده‌اند. برچسب‌های a, b, c و d که به عنوان برچسب‌های نرم نیز شناخته می‌شوند، ثابت نیستند. در معادلات ۹ و ۱۰، برچسب a از تصویر نهایی بین ۰ تا ۰/۳ است، در حالی که برچسب b از تصویر هدف بین ۰/۷ تا ۱/۲ انتخاب می‌شود. برچسب c و d در معادلات ۲ و ۳ مقادیر تصادفی بین ۰/۷ تا ۱/۲ تعیین شده‌اند. در تمام لایه‌ها، از تابع فعال‌سازی LeakyReLU با شیب ۰/۲ استفاده شده است. پس از کوچک‌نمایی، با استفاده از مقیاس بزرگ‌نمایی، تصویر به صورت خودکار تا اندازه ۲۵۶ بزرگ‌نمایی می‌شود. از درون‌یابی دو قطبی برای هر دو عملیات بزرگ‌نمایی و کوچک‌نمایی استفاده شده است.

¹ Epoch

² Checkpoints

³ Batch size

⁴ Learning rate

⁵ Weight decay

۲-۶- داده‌های تحقیق

برای ارزیابی کامل عمل‌کرد روش پیشنهادی، سه مجموعه داده پرکاربرد که از ماهواره‌های WorldView-2 (WV2)، GaoFen-2 (GF2) و QuickBird (QB) به‌دست آمده‌اند، در آزمایش‌های به‌عمل‌آمده به‌کار گرفته شده‌اند. برای مقایسه میان مدل‌های نظارت شده و بدون نظارت، مجموعه داده‌ها تحت هر دو وضوح کاهش یافته (براساس پروتکل والد (Wald et al., 2000)) و تنظیمات وضوح کامل ساخته شده‌اند. تحت پروتکل والد، مدل‌های تحت نظارت را می‌توان بر روی تصاویر با وضوح کاهش یافته با استفاده از تصاویر LRMS اصلی به‌عنوان برچسب آموزش داد. در نتیجه روش ما مشابه روش‌های قبلی است (Ma et al., 2020). مجموعه داده‌های وضوح کامل را با استفاده از یک فیلتر گاوسی محو می‌کند و سپس آن‌ها را به ترتیب با ضریب مقیاس‌گذاری ۴ کوچک می‌نماید. از طرف دیگر تحت تنظیمات وضوح کامل، مدل‌ها به شیوه‌ای بدون نظارت آموزش داده می‌شوند. اگرچه محیط آموزشی محدود به نوع مدل‌ها است، اما محیط آزمایش بدون محدودیت است. این امکان وجود دارد که تمامی مدل‌ها را صرف نظر از نیاز یا عدم نیاز به برچسب‌های نظارتی، روی تصاویر در وضوح کاهش یافته و هم در وضوح کامل آزمایش نمود.

علاوه‌براین، به دلیل اندازه بسیار بزرگ تصاویر سنجش از دور که برای تغذیه به یک شبکه عصبی بسیار بزرگ هستند، این تصاویر به قطعه‌های کوچک برای تشکیل مجموعه‌های آموزشی و آزمایش تقسیم می‌شوند. تصاویر مورد استفاده در مدل UMP-GAN به ترتیب در قطعه‌های کوچک، 64×64 برای تصاویر LRMS و 256×256 برای تصاویر PAN تقسیم شده‌اند. همچنین، مناطق و فصل‌های به‌دست آوردن تصاویر در محدوده گسترده‌ای قرار دارند، که به مدل اجازه می‌دهد تا با داده‌هایی از انواع شرایط روشنایی و شرایط هندسی آموزش ببیند. به منظور حذف اثرات مرزی برای مجموعه آزمایش، از روش استاندارد پیروی شده و تصاویر را در مناطق با همپوشانی کوچک بین تکه‌های مجاور تقسیم می‌شوند. باید تاکید نمود که میان مجموعه‌های آموزش و آزمون هیچ همپوشانی وجود ندارد. اولین مجموعه داده‌های ما شامل ۱۲۰۰۰ جفت قطعه است که از سه صحنه تصویر ماهواره‌ای WV2 تهیه شده توسط شرکت دیجیتال گلوب (Maxar Technologies) استخراج شده‌اند. تصویر PAN در ماهواره WV2 دارای وضوح مکانی $0/5$ متر است و تصاویر LRMS متناظر دارای وضوح $1/8$ متر هستند. برای دستیابی به تصاویر Worldview، سیستم نمایش UTM با مناطق مناسب به‌کار گرفته شده است. مجموعه داده دوم، شامل ۹ صحنه تصویر از ماهواره GF2 است، و تصاویر PAN و LRMS را در ۳۲۰۰۰ جفت قطعه تصویر تقسیم شده‌اند. ماهواره GF2 تصویر PAN را با وضوح مکانی $0/8$ متر فراهم می‌کند و تصاویر LRMS متناظر را با وضوح $3/2$ متر ارائه می‌دهد. در نهایت، مجموعه داده‌های QB شامل شش صحنه تصویر است که در ۱۷۰۰۰ جفت قطعه تصویر تقسیم می‌شوند. وضوح مکانی تصاویر PAN ثبت شده توسط ماهواره QB برابر با $0/6$ متر است، در حالی که تصاویر LRMS متناظر آن‌ها وضوح

مکانی ۲/۴ متر دارند. برای داده‌های آموزش و داده‌های آزمایش، قطعه‌های تصویر به صورت تصادفی به دو دسته ۹۰٪ و ۱۰٪ تقسیم شده‌اند. علاوه‌براین، بدون از دست دادن کلیت در آزمایش‌های انجام گرفته، چهار باند طیفی قرمز، سبز، آبی و نزدیک به مادون قرمز در هر مجموعه داده در نظر گرفته شده‌اند.

۲-۷- معیارهای ارزیابی طیفی و مکانی تصاویر ادغام شده

برای بررسی برتری مدل پیشنهادی و مقایسه آن با سایر روش‌ها، از معیارهای با مرجع SAM، ERGAS و SSIM برای تصاویر با وضوح کاهش یافته و معیارهای بدون مرجع D_s ، D_λ و QNR برای تصاویر با وضوح کامل استفاده می‌شود. لازم به ذکر است که از کد MATLAB ارائه شده در مخزن سنجش از دور برای محاسبه تمام معیارها استفاده شده است.

۳- پیاده‌سازی، نتایج و بحث

به منظور ارزیابی عمل‌کرد UMP-GAN، از پانزده روش معتبر اخیر برای مقایسه کیفی و کمی، شامل شش روش سنتی، شش روش نظارت شده و سه روش بدون نظارت استفاده شده‌اند. روش‌های سنتی پناش‌پنینگ مانند BDS (Vivone et al., 2019)، GS (Laben et al., 2000)، IHS (Tu et al., 2004)، Brovey (Gillespie et al., 1987)، HPF (Liu et al., 2000) و SFIM (Shao et al., 2020) با مدل پیشنهادی مقایسه شده‌اند. علاوه‌براین، مدل‌های پناش‌پنینگ تحت نظارت شامل PNN (Masi et al., 2016)، DRPNN (Wei et al., 2017)، MSD-CNN (Yuan et al., 2018)، PanNet (Yang et al., 2017)، PSGAN (Liu et al., 2021) و RED-cGAN (Shao et al., 2020) می‌شوند. چهار مدل اول از معماری CNN استفاده می‌نمایند و مدل‌های PSGAN و RED-cGAN مدل مبتنی بر GAN هستند. روش‌های بدون نظارت را مدل‌های Pan-GAN (Ma et al., 2020)، PGMAN (Zhou et al., 2021) و مدل پیشنهادی در این مقاله تشکیل می‌دهند. کدها و پارامترهای تمام این روش‌ها در دسترس عموم هستند و به منظور مقایسه منصفانه، تمام پارامترهای این مدل‌ها براساس مقادیر پیش فرض آن‌ها تنظیم شده‌اند.

۳-۱- ارزیابی کمی

با توجه به معیارهای بدون مرجع برای نتایج وضوح کامل، شاخص D_λ بر روی اعوجاج طیفی و اطلاعات رنگ در تصاویر ادغام شده تمرکز دارد و شاخص D_s بر روی اعوجاج مکانی و جزئیات بافت تمرکز می‌کند، در حالی که شاخص QNR ترکیبی از هر دو معیار است. برای معیارهای با مرجع تحت تنظیمات وضوح کاهش یافته،

شاخص SAM برای مطالعه اعوجاج‌های طیفی در تصاویر ادغام شده به کار می‌رود، اعوجاج‌های مکانی توسط شاخص ERGAS اندازه‌گیری می‌شوند و شاخص SSIM یک معیار جامع است. جدول ۱ آزمایش‌های کمی بر روی تصاویر ادغام شده از داده‌های ماهواره‌ای WorldView-2 را نشان می‌دهد. برای معیارهای بدون مرجع، مدل UMP-GAN پیشنهادی از یادگیری بدون نظارت بهره می‌برد و بهترین مقادیر را در شاخص‌های D_s ، D_λ و QNR به دست می‌آورد که بسیار بهتر از سایر روش‌های مورد مقایسه است. این بدان معنی است که نتایج به دست آمده از نظر توزیع اطلاعات طیفی و مکانی به تصاویر ورودی نزدیک‌تر هستند. سایر روش‌های بدون نظارت برای معیارهای با مرجع عمل کرد نسبتاً ضعیفی دارند. مدل پیشنهادی بهترین عمل کرد را بر روی شاخص‌های SAM و SSIM دارد که نشان می‌دهد این روش به خوبی می‌تواند ویژگی‌های رنگ و ساختارهای محلی را حفظ نماید.

جدول (۱): نتایج کیفی بر روی داده‌های ماهواره‌ای WV2

Table (1): Quantitative results on WV2 satellite data

QNR	معیارهای بدون مرجع			معیارهای با مرجع			روش	نوع
	D_s	D_λ	SSIM	ERGAS	SAM			
۰/۸۸۱۱±۰/۰۴۱۲	۰/۰۵۷۴±۰/۰۰۸۶	۰/۰۶۵۲±۰/۰۱۰۳	۰/۸۸۲۸±۰/۰۰۹۷	۲/۰۹۲۱±۰/۵۱۱۰	۱/۴۳۵۳±۰/۱۲۷۴	Brovey	سنجی	
۰/۹۰۶۹±۰/۰۱۸۹	۰/۰۵۱۰±۰/۰۱۰۳	۰/۰۴۴۳±۰/۰۰۹۷	۰/۸۵۰۶±۰/۰۰۳۷	۳/۳۱۳۹±۰/۵۱۴۳	۱/۴۵۹۱±۰/۱۶۷۳	GS		
۰/۹۴۸۹±۰/۰۰۲۷	۰/۰۴۲۱±۰/۰۰۲۷	۰/۰۰۹۳±۰/۰۰۱۱	۰/۸۱۰۷±۰/۰۱۹۵	۳/۶۳۷۰±۰/۰۲۷۳	۱/۳۳۴۱±۰/۰۹۴۸	SFIM		
۰/۸۵۲۵±۰/۰۲۲۵	۰/۰۶۴۱±۰/۰۰۹۷	۰/۰۸۹۱±۰/۰۱۴۶	۰/۸۱۷۰±۰/۰۲۰۲	۳/۶۰۰۴±۰/۰۹۸۳	۱/۶۸۳۶±۰/۳۸۳۱	IHS		
۰/۹۵۵۱±۰/۰۰۵۰	۰/۰۴۳۹±۰/۰۰۵۰	۰/۰۰۹۹±۰/۰۰۰۴	۰/۸۴۹۳±۰/۰۰۸۹	۳/۳۲۰۱±۰/۴۴۴۱	۱/۴۴۴۹±۰/۱۵۰۶	HPF		
۰/۹۴۹۴±۰/۰۰۱۸	۰/۰۳۰۲±۰/۰۰۰۹	۰/۰۲۱۰±۰/۰۰۳۱	۰/۹۱۹۳±۰/۰۰۸۴	۲/۸۳۰۹±۰/۰۹۱۷	۱/۴۹۹۲±۰/۱۰۳۱	BDSD-PC		
۰/۹۱۱۸±۰/۰۴۵۴	۰/۰۴۹۹±۰/۰۰۳۰	۰/۰۴۰۳±۰/۰۱۷۲	۰/۹۲۷۰±۰/۰۳۳۹	۲/۴۰۱۹±۰/۵۱۲۷	۱/۴۶۰۷±۰/۱۷۰۱	PNN	نظارت شده	
۰/۸۸۸۱±۰/۰۱۵۲	۰/۰۵۴۴±۰/۰۰۷۱	۰/۰۶۰۸±۰/۰۰۹۰	۰/۹۳۸۱±۰/۰۰۹۲	۲/۱۷۳۸±۰/۴۹۶۰	۱/۴۳۰۶±۰/۰۹۵۲	DRPNN		
۰/۹۳۲۸±۰/۰۳۱۰	۰/۰۴۶۷±۰/۰۰۲۵۷	۰/۰۳۱۹±۰/۰۰۶۴	۰/۹۳۳۶±۰/۰۴۷۳	۲/۰۸۹۰±۰/۳۰۰۱	۰/۹۹۹۷±۰/۰۰۴۳	PanNet		
۰/۸۷۷۴±۰/۰۲۹۹	۰/۰۵۹۱±۰/۰۱۱۶	۰/۰۶۷۴±۰/۰۲۰۳	۰/۹۲۹۴±۰/۰۴۲۰	۲/۳۹۰۵±۰/۰۴۲۶	۲/۰۰۰۴±۰/۳۰۶۵	MSD-CNN		
۰/۹۵۰۰±۰/۰۰۵۳	۰/۰۴۱۹±۰/۰۰۴۵	۰/۰۰۸۴±۰/۰۰۰۹	۰/۹۴۴۷±۰/۰۷۰۶	۱/۹۹۰۰±۰/۰۱۰۸	۰/۹۵۴۱±۰/۰۱۰۶	PSGAN		
۰/۹۵۰۱±۰/۰۱۰۱	۰/۰۴۲۲±۰/۰۰۱۳	۰/۰۰۸۰±۰/۰۰۴۲	۰/۹۴۲۸±۰/۰۵۷۱	۲/۰۷۰۳±۰/۰۵۱۰	۰/۹۲۰۴±۰/۱۶۳۵	RED-cGAN		
۰/۹۵۸۷±۰/۰۰۰۷	۰/۰۳۰۹±۰/۰۰۰۵	۰/۰۱۰۷±۰/۰۰۰۲	۰/۹۳۷۲±۰/۰۱۷۴	۱/۷۶۸۲±۰/۰۲۱۳	۱/۰۷۶۴±۰/۰۵۷۰	Pan-GAN	بدون نظارت	
۰/۹۶۲۰±۰/۰۰۴۹	۰/۰۳۰۴±۰/۰۰۱۲	۰/۰۰۷۸±۰/۰۰۳۷	۰/۹۱۵۱±۰/۰۳۰۸	۱/۷۳۰۹±۰/۰۴۴۹	۱/۰۳۵۹±۰/۰۹۴۴۲	PGMAN		
۰/۹۶۵۱±۰/۰۰۰۲	۰/۰۲۸۱±۰/۰۰۰۱	۰/۰۰۶۹±۰/۰۰۱۳	۰/۹۴۸۸±۰/۰۳۳۱	۱/۸۰۰۴±۰/۰۳۰۷	۰/۹۱۹۸±۰/۰۹۵۳	UMP-GAN		
۱	۰	۰	۱	۰	۰	مقدار مطلوب		

توجه: مقادیر پررنگ نشان دهنده بهترین نتایج هستند. ردیف آخر مقدار مطلوب هر معیار را بیان می‌کند.

آزمایش‌های کمی بر روی داده‌های ماهواره‌ای GaoFen-2 در جدول ۲ ارائه شده است. نتایج نشان می‌دهد که UMP-GAN هم‌چنان برتر است و بالاترین مقادیر را در تمام معیارهای بدون مرجع به دست می‌آورد. روش‌های نظارت نشده در مقایسه با روش‌های نظارت شده از نظر معیارهای با مرجع عمل کرد ضعیفی دارند، زیرا الگوریتم‌های نظارت شده مدل را تحت هدایت تصاویر مرجع آموزش می‌دهند. با این حال، مدل پیشنهادی

هنوز هم بهتر از سایر روش‌های بدون نظارت بر معیارهای با مرجع عمل می‌کند. مدل UMP-GAN هم‌چنین نسبت به رویکردهای سنتی در شاخص‌های SSIM و ERGAS عمل کرد بهتری دارد. آزمایش‌های کمی بر روی پایگاه داده QuickBird در جدول ۳ نشان داده شده است. مشابه نتایج مجموعه داده‌های GaoFen-2، در اینجا نیز مدل پیشنهادی در مقایسه با روش‌های سنتی و نظارت شده برتری خود را حفظ می‌کند و بالاترین مقادیر را در تمام معیارهای بدون مرجع به دست می‌آورد. علاوه بر این، مدل پیشنهادی UMP-GAN بر اساس شاخص‌های با مرجع نسبت به رویکردهای سنتی و هم‌چنین مدل‌های بدون نظارت به بالاترین مقادیر دست پیدا می‌کند و بهترین عمل کرد را بر روی شاخص‌های SSIM دارد. اگرچه مقادیر شاخص‌های SAM و ERGAS در روش پیشنهادی بهترین نبودند، اما رقابتی هستند. در نتیجه، مدل UMP-GAN می‌تواند ویژگی‌های مکانی و طیفی را به دلیل ساختار متراکم و استخراج ویژگی چندمقیاسی بهتر حفظ نماید.

جدول (۲): نتایج کیفی بر روی داده‌های ماهواره‌ای GF2

Table (2): Quantitative results on GF2 satellite data

QNR	معیارهای بدون مرجع		معیارهای با مرجع			روش	نوع
	D_s	D_λ	SSIM	ERGAS	SAM		
0.9141 ± 0.0436	0.0614 ± 0.0283	0.0261 ± 0.0170	0.8393 ± 0.1855	$3/585 \pm 0.3152$	$1/697 \pm 0.2038$	Brovey	سنتی
0.9034 ± 0.0575	0.0598 ± 0.0315	0.0391 ± 0.0292	0.8495 ± 0.196	$3/806 \pm 0.3712$	$2/0337 \pm 0.2036$	GS	
0.9476 ± 0.0213	0.0334 ± 0.0075	0.0196 ± 0.0144	0.8206 ± 0.197	$4/0967 \pm 0.5006$	$2/025 \pm 0.1996$	SFIM	
0.9096 ± 0.0501	0.0599 ± 0.0298	0.0324 ± 0.0226	0.8431 ± 0.200	$3/8162 \pm 0.4666$	$2/2007 \pm 0.2129$	IHS	
0.9447 ± 0.0473	0.0320 ± 0.0073	0.0240 ± 0.0161	0.8331 ± 0.165	$3/9606 \pm 0.5140$	$2/2504 \pm 0.2473$	HPF	
0.9341 ± 0.0078	0.0392 ± 0.0054	0.0277 ± 0.0105	0.6219 ± 0.108	$4/9670 \pm 0.2983$	$1/9084 \pm 0.1891$	BDSD-PC	
0.9355 ± 0.0219	0.0498 ± 0.0099	0.0154 ± 0.0127	0.9605 ± 0.0073	$1/5135 \pm 0.3866$	$1/3028 \pm 0.2027$	PNN	نظارت شده
0.9118 ± 0.0393	0.0713 ± 0.0266	0.0181 ± 0.0142	0.9666 ± 0.0065	$1/3519 \pm 0.3498$	$1/2357 \pm 0.1650$	DRPNN	
0.8388 ± 0.0283	0.1524 ± 0.0255	0.0103 ± 0.0035	0.9685 ± 0.0064	$1/3480 \pm 0.3511$	$1/1002 \pm 0.1007$	PanNet	
0.9281 ± 0.0149	0.0602 ± 0.0098	0.0124 ± 0.0055	0.9617 ± 0.0072	$1/5151 \pm 0.3860$	$1/5057 \pm 0.2048$	MSD-CNN	
0.8647 ± 0.0249	0.1127 ± 0.0228	0.0094 ± 0.0027	0.9700 ± 0.0071	$1/3559 \pm 0.3530$	$1/0889 \pm 0.0928$	PSGAN	
0.9534 ± 0.0117	0.0341 ± 0.0055	0.0129 ± 0.0102	0.9691 ± 0.0078	$1/3492 \pm 0.3207$	$1/3871 \pm 0.1076$	RED-cGAN	
0.9131 ± 0.0248	0.0468 ± 0.0081	0.0420 ± 0.0179	0.7424 ± 0.0226	$4/6713 \pm 0.4526$	$2/4895 \pm 0.2394$	Pan-GAN	بدون نظارت
0.9688 ± 0.0094	0.0219 ± 0.0107	0.0095 ± 0.0051	0.9482 ± 0.0152	$3/2379 \pm 0.4501$	$2/6481 \pm 0.1991$	PGMAN	
0.9768 ± 0.0031	0.0154 ± 0.0016	0.0079 ± 0.0013	0.9598 ± 0.0080	$1/9984 \pm 0.2585$	$1/8405 \pm 0.2390$	UMP-GAN	
۱	۰	۰	۱	۰	۰	مقدار مطلوب	

توجه: مقادیر پررنگ نشان دهنده‌ی بهترین نتایج هستند. ردیف آخر مقدار مطلوب هر معیار را بیان می‌کند.

۳-۲- ارزیابی بصری

شکل‌های ۴، ۵ و ۶ نتایج پنشارپینگ بر روی تصاویر واقعی سنجش از دور را نشان می‌دهند که به ترتیب از مجموعه داده‌های ماهواره‌ای WorldView-2، GaoFen-2 و QuickBird گرفته شده‌اند. این مجموعه از داده آزمایشی از تصاویر ورودی PAN و LRMS اصلی حاصل شده‌اند. مجموعه داده آزمایشی حاوی ساختمان‌ها و پوشش گیاهی با رنگ‌های متنوع و اطلاعات طیفی و همچنین اطلاعات بافتی غنی هستند که فرصتی عالی برای مقایسه اثربخشی روش‌های مختلف در حفظ اطلاعات طیفی و مکانی را فراهم می‌کنند. به منظور ارزیابی کیفیت طیفی، تصاویر LRMS بزرگنمایی شده به عنوان مرجع بصری به کار گرفته می‌شوند. در همین حال، تصاویر PAN در وضوح کامل به عنوان مرجع مکانی نشان داده شده‌اند. همانطور که در مقاله (He et al., 2016) اشاره شده است، حتی اگر برخی از روش‌ها مقادیر کمی بالاتری تولید نمایند لزوماً نتایج بصری بهتری را نمایش نخواهند داد.

جدول (۳): نتایج کمی بر روی داده‌های ماهواره‌ای QB

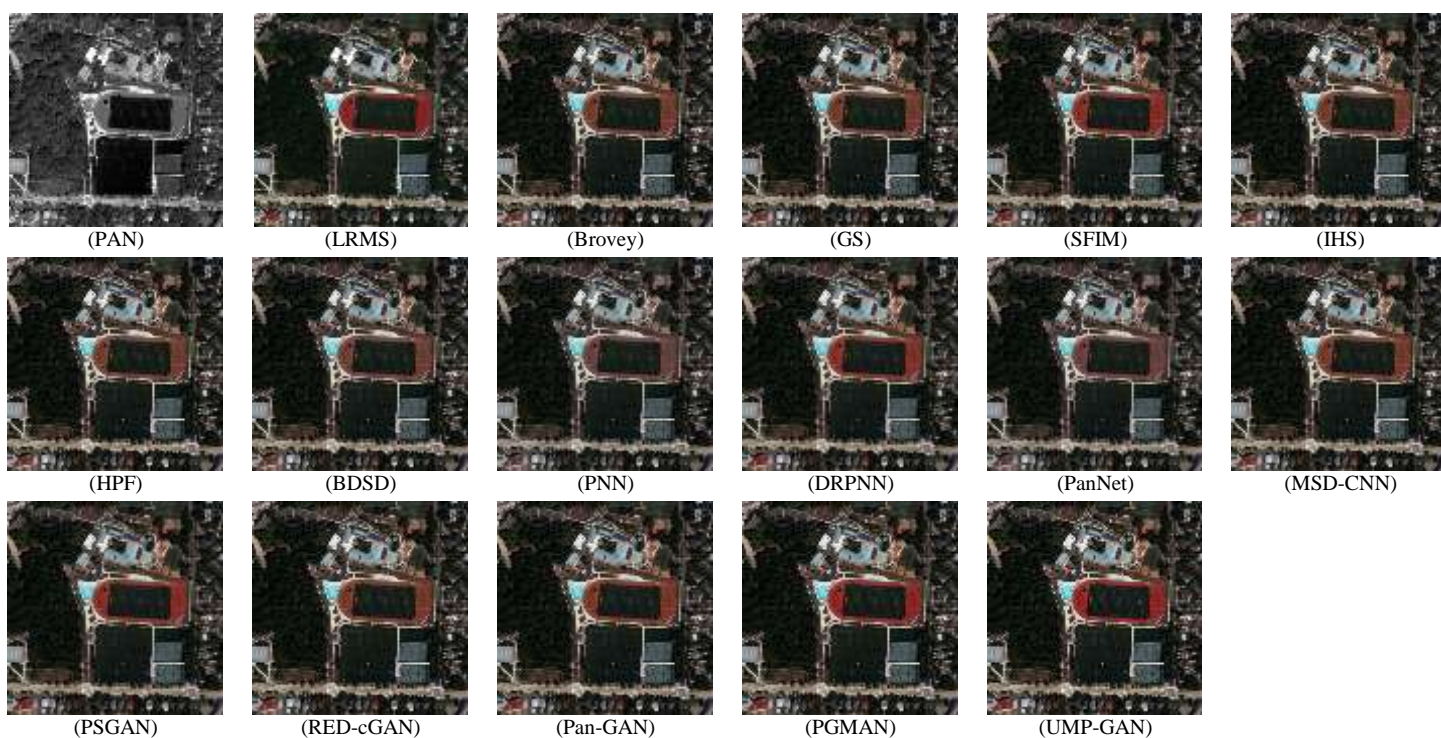
Table (3): Quantitative results on QB satellite data

QNR	معیارهای بدون مرجع		SSIM	معیارهای با مرجع			روش	نوع
	D _s	D _λ		ERGAS	SAM			
۰/۹۳۰۰±۰/۰۰۰۷	۰/۰۴۵۳±۰/۰۰۰۴	۰/۰۲۵۸±۰/۰۰۰۳	۰/۸۷۲۹±۰/۰۰۲۰	۲/۳۰۵۲±۰/۳۰۵۰	۱/۶۸۳۳±۰/۱۰۴۲	Brovey	سنجی	
۰/۹۳۰۹±۰/۰۰۰۵	۰/۰۴۵۱±۰/۰۰۰۳	۰/۰۲۵۱±۰/۰۰۰۲	۰/۸۷۴۹±۰/۰۰۴۶	۲/۳۹۱۹±۰/۴۰۰۲	۱/۶۰۳۸±۰/۰۹۵۷	GS		
۰/۹۴۴۴±۰/۰۰۱۳	۰/۰۴۶۳±۰/۰۰۰۵	۰/۰۰۹۷±۰/۰۰۰۸	۰/۸۵۷۵±۰/۰۰۱۸	۲/۷۷۶۸±۰/۴۶۱۱	۱/۶۹۲۶±۰/۱۰۴۱	SFIM		
۰/۹۲۷۶±۰/۰۰۱۰	۰/۰۴۸۰±۰/۰۰۰۷	۰/۰۲۵۶±۰/۰۰۰۴	۰/۸۷۲۵±۰/۰۰۳۰	۲/۴۸۰۱±۰/۴۰۸۵	۱/۸۶۰۹±۰/۱۲۵۸	IHS		
۰/۹۴۶۲±۰/۰۰۰۴	۰/۰۴۴۲±۰/۰۰۰۳	۰/۰۱۰۰±۰/۰۰۰۱	۰/۸۵۵۲±۰/۰۰۲۹	۲/۷۱۶۴±۰/۴۲۰۵	۱/۷۳۴۱±۰/۰۸۴۶	HPF		
۰/۹۶۹۷±۰/۰۰۰۷	۰/۰۱۹۸±۰/۰۰۰۵	۰/۰۱۰۷±۰/۰۰۰۸	۰/۸۳۱۰±۰/۰۰۲۷	۳/۰۹۴۴±۰/۴۱۳۷	۲/۱۰۷۵±۰/۱۱۰۱	BDS-PC		
۰/۸۳۴۵±۰/۰۰۰۳	۰/۰۲۲۲±۰/۰۰۰۲	۰/۰۷۰۵±۰/۰۰۰۶	۰/۹۱۳۳±۰/۰۰۰۹	۲/۰۶۲۲±۰/۳۹۶۵	۲/۵۲۷۶±۰/۰۷۴۰	PNN	نظارت شده	
۰/۹۲۵۴±۰/۰۰۰۶	۰/۰۵۵۶±۰/۰۰۰۵	۰/۰۲۰۱±۰/۰۰۰۲	۰/۹۲۹۵±۰/۰۰۰۲	۲/۴۱۸۹±۰/۲۲۴۲	۲/۱۰۹۶±۰/۱۰۸۴	DRPNN		
۰/۹۴۴۴±۰/۰۰۰۴	۰/۰۴۹۴±۰/۰۰۰۳	۰/۰۰۶۵±۰/۰۰۰۱	۰/۹۲۹۹±۰/۰۰۰۳	۲/۲۱۶۳±۰/۲۰۷۲	۱/۵۲۳۷±۰/۰۶۱۹	PanNet		
۰/۸۴۹۴±۰/۰۰۱۳	۰/۰۱۰۴±۰/۰۰۰۸	۰/۰۵۲۰±۰/۰۰۰۱	۰/۸۹۵۶±۰/۰۰۱۴	۲/۸۶۶۲±۰/۴۰۹۵	۳/۳۹۷۴±۰/۱۹۸۶	MSD-CNN		
۰/۸۹۷۵±۰/۰۰۰۳	۰/۰۷۱۱±۰/۰۰۰۱	۰/۰۳۳۸±۰/۰۰۰۲	۰/۹۲۶۵±۰/۰۰۰۷	۱/۴۸۵۲±۰/۰۸۴۵	۲/۳۹۷۱±۰/۱۰۳۵	PSGAN		
۰/۸۸۵۸±۰/۰۰۱۲	۰/۰۷۹۷±۰/۰۰۰۳	۰/۰۳۷۴±۰/۰۰۰۹	۰/۹۲۳۹±۰/۰۰۱۱	۱/۹۹۵۰±۰/۱۰۶۳	۲/۴۴۵۰±۰/۰۶۶۳	RED-cGAN		
۰/۹۵۷۹±۰/۰۰۱۹	۰/۰۳۲۷±۰/۰۰۱۹	۰/۰۰۹۹±۰/۰۰۰۳	۰/۹۱۷۸±۰/۰۰۱۹۷	۲/۴۴۹۸±۰/۲۴۲۴	۲/۵۹۸۶±۰/۰۹۸۰	Pan-GAN	بدون نظارت	
۰/۹۶۴۰±۰/۰۰۰۱	۰/۰۲۷۵±۰/۰۰۰۸	۰/۰۰۸۷±۰/۰۰۰۶	۰/۹۰۰۳±۰/۰۰۴۱	۲/۴۸۸۰±۰/۱۶۳۷	۱/۷۴۰۶±۰/۰۹۰۶	PGMAN		
۰/۹۷۵۱±۰/۰۰۰۵	۰/۰۱۹۱±۰/۰۰۰۶	۰/۰۰۵۹±۰/۰۰۰۴	۰/۹۳۰۷±۰/۰۰۱۱۴	۱/۹۷۴۲±۰/۱۰۰۶	۱/۶۹۹۳±۰/۰۵۰۱	UMP-GAN		
۱	۰	۰	۱	۰	۰	مقدار مطلوب		

توجه: مقادیر پررنگ نشان دهنده بهترین نتایج هستند. ردیف آخر مقدار مطلوب هر معیار را بیان می‌کند.

شکل ۵ نتایج ادغام در تصاویر وضوح کامل WorldView-2 را نشان می‌دهد. نتایج به دست آمده از روش‌های GS، Brovey، IHS و BDS-PC اطلاعات فرکانس بالای تمیز تقریباً مشابه با تصاویر PAN ورودی ارائه می‌دهند و به طور مؤثر اطلاعات مکانی را تزریق می‌نمایند. با این حال، این روش‌ها از اعوجاج طیفی رنج

می‌برند، به خصوص روش‌های GS، Brovey و IHS که رنگ آن‌ها تیره‌تر از تصاویر LRMS ورودی است. برای مثال، در اطلاعات طیفی حاصل از روش BSDS نواحی ساختمانی تحریف شده‌اند. در تصویر حاصل شده از روش SFIM اعوجاج واضح و تاری مکانی وجود دارد.



شکل (۵): نتایج بصری روش‌های مختلف ادغام تصاویر در مجموعه داده‌های با وضوح کامل متعلق به سنجنده WV2

Fig. (5): Qualitative comparison of different methods for pansharpening under the full-resolution test on the data from WV2

از آنجاکه روش‌های پنشارپنینگ تحت نظارت بر روی مجموعه داده‌های وضوح کاهش یافته آموزش داده می‌شوند، تعمیم مدل آموزش دیده به تصاویر واقعی با وضوح کامل منجر به نتایج ضعیف ادغام می‌گردد. به عنوان مثال، نتیجه حاصل از روش PNN اعوجاج طیفی قابل توجه و همچنین اثرات تاری جزئی را نشان می‌دهد. این اعوجاج طیفی ممکن است نتیجه معماری ساده PNN نیز باشد، که ثبت روابط طیفی میان

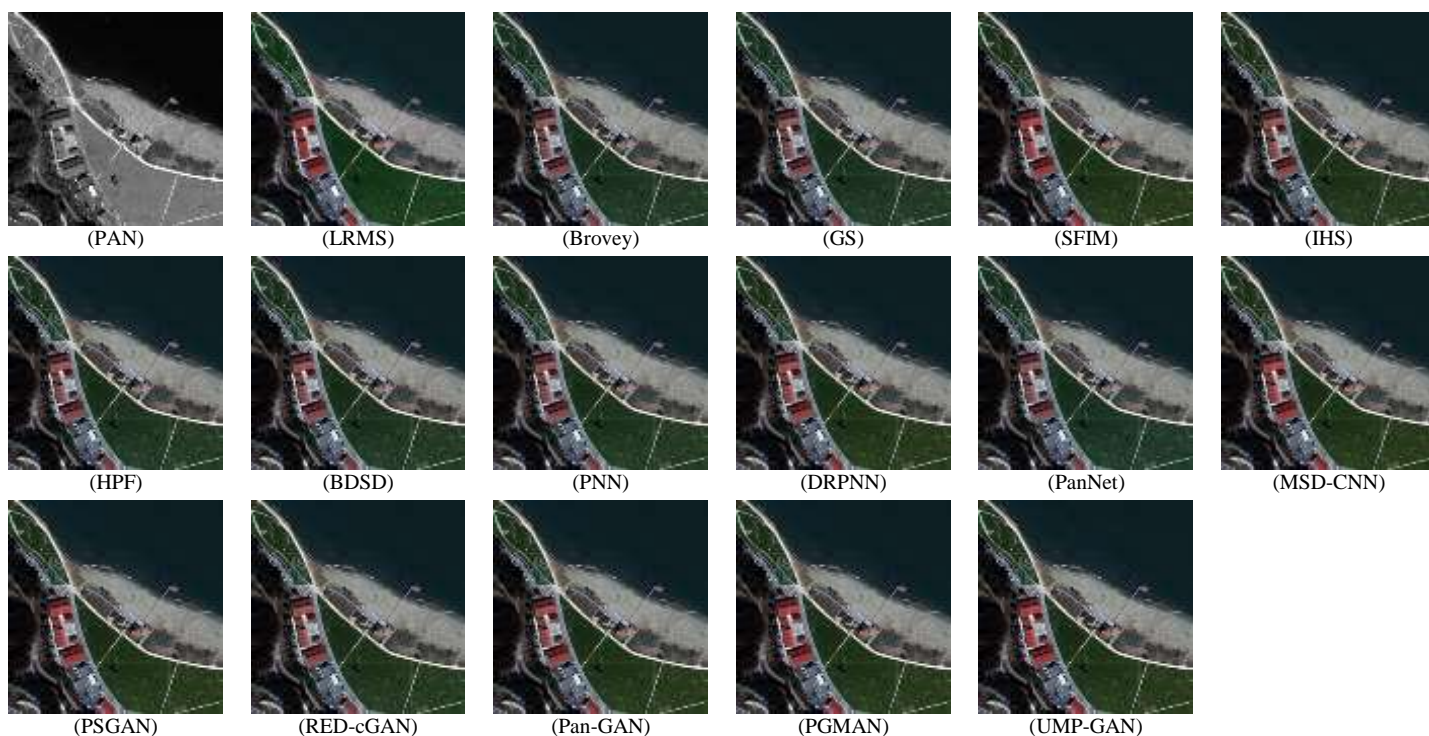
E-mail: mehrshad_brm@nit.ac.ir; aghagol@nit.ac.ir

باند‌های تصاویر MS را دشوار می‌کند. همانگونه که در شکل نشان داده شده است در مقایسه با روش‌های ادغام تحت نظارت، روش پیشنهادی که مبتنی بر یادگیری عمیق بدون نظارت است بهتر می‌تواند اطلاعات طیفی و مکانی را در وضوح کامل حفظ نماید. تصویر حاصل شده از روش PanNet دارای اعوجاج‌های مکانی و طیفی با اثرات مات‌کننده را نشان می‌دهد، که در نتیجه باقی‌مانده‌های آموخته شده توسط ResNet ایجاد می‌شوند. واضح است که نتایج روش‌های DRPNN و PSGAN تاری جزئیات و اعوجاج طیفی را نشان می‌دهند. نتایج کمی و بصری مزایای استفاده از چند شبکه تفکیک‌کننده در مقایسه با تنها یک شبکه تفکیک‌کننده را نشان می‌دهند. Pan-GAN در حفظ جزئیات طیفی بهتر عمل می‌کند، اگرچه از نظر حفظ اطلاعات مکانی ضعیف است و تصاویر نسبتاً تاری تولید می‌نماید.

در میان نتایج، تصویر نهایی به دست آمده با استفاده از روش PGMAN شبیه به تصویر به دست آمده از روش پیشنهادی است؛ با این حال، برخی جزئیات بافت از دست رفته واضح است. روش پیشنهادی خطوط واضح‌تری از اشیا زمینی نسبت به سایر روش‌های بدون نظارت ارائه می‌کند، که به این معنی است که این روش می‌تواند اطلاعات ساختار مکانی را بهتر از آن‌ها حفظ نماید. این را می‌توان با این واقعیت توضیح داد که UMP-GAN می‌تواند اطلاعات دقیق در تصاویر ورودی را با استفاده از بلوک‌های MDB پیشنهادی به طور کامل استخراج کند. بنابراین، نتیجه حاصل از ادغام مدل پیشنهادی به تصاویر ورودی نزدیک‌تر بوده و در حفظ جزئیات مکانی و ویژگی‌های طیفی مربوطه عمل کرد بهتری دارد.

شکل ۶ نتایج ادغام از مجموعه داده‌های وضوح کامل GaoFen-2 را نشان می‌دهد. نتیجه حاصل از روش Brovey نمی‌تواند جزئیات طیفی تصویر LRMS ورودی را به درستی حفظ نماید و منجر به اعوجاج‌های رنگی قابل توجهی می‌شود. هم‌چنین واضح است که روش‌های تحت نظارت PNN، DRPNN، PanNet، MSD-CNN و PSGAN جزئیات مکانی را با کیفیت پایین نشان می‌دهند. به نظر می‌رسد که تعمیم مؤثر اطلاعات آموخته شده از تصاویر با وضوح پایین به کاربردهای عملی دشوار است. به این دلیل که تصاویر سنجش از دور معمولاً عمق بیت بیشتر و توزیع پیکسل منحصر به فردی نسبت به تصاویر طبیعی دارند. هنگامی که تصاویر PAN و LRMS ورودی به صورت تحت نظارت کوچک‌نمایی می‌شوند، تغییری در توزیع داده‌ها رخ خواهد داد و جزئیات اصلی از دست خواهند رفت. از طرف دیگر، مدل پیشنهادی UMP-GAN زمانی که بر روی تصاویر با وضوح کامل آموزش داده می‌شود، به طور قابل توجهی مؤثرتر است. هم‌چنین با توجه به نتایج تجربی، روش پیشنهادی تاری و اعوجاج طیفی کمتری نسبت به روش‌های مقایسه شده ایجاد می‌کند و می‌تواند

تعادل خوبی میان حفظ رنگ و بافت به دست آورد. بنابراین تمرکز ما بر توسعه یک مدل بدون نظارت پنشارپنینگ است زیرا در این حوزه ادغام معمولاً بر روی تصاویر وضوح کامل در کاربردهای واقعی انجام می‌شود.



شکل (۶): نتایج بصری روش‌های مختلف ادغام تصاویر در مجموعه داده‌های با وضوح کامل متعلق به سنجنده GF2

Fig. (6): Qualitative comparison of different methods for pansharpening under the full-resolution test on the data from GF2

شکل ۷ نتایج ادغام از مجموعه داده‌های وضوح کامل QuickBird را نشان می‌دهد. همانطور که در رنگ بام ساختمان‌ها مشهود است، روش‌های Brovey، GS، IHS و HPF جزئیات طیفی را از تصویر LRMS تحریف می‌کنند. علاوه‌براین، تصاویر ادغام شده حاصل از روش BDSB دارای مقداری اعوجاج رنگی هستند. نتایج نامطلوب با برخی اعوجاج‌ها نیز توسط روش نظارت شده MSD-CNN مشاهده می‌شوند. روش PNN نیز

انحراف طیفی قابل توجهی را نشان می دهد. به عنوان مثال، رنگ بامها به طور قابل توجهی با اطلاعات طیفی تصویر LRMS متفاوت است و رنگ زمین تحریف شده است.



شکل (۷): نتایج بصری روش های مختلف ادغام تصاویر در مجموعه داده های با وضوح کامل متعلق به سنجنده QB
Fig. (7): Qualitative comparison of different methods for pansharpening under the full-resolution test on the data from QB

می توان مشاهده نمود که نتیجه حاصل از روش RED-cGAN کیفیت مکانی خوب اما کیفیت طیفی پایینی را به دلیل تغییرات جزئی در رنگ های ساختمان نشان می دهد، که با نتایج سایر مجموعه داده ها مطابقت دارد. البته باید توجه داشت که این روش های نظارت شده اطلاعات مکانی را بهتر از سایر روش های سنتی حفظ می کنند. روش های سنتی اعوجاج ها و مصنوعات قابل توجهی ایجاد می نمایند، در حالی که روش های دیگر نتایج نسبتاً قابل قبولی را ارائه می دهند. اگرچه نتیجه حاصل از روش Pan-GAN جزئیات طیفی را به خوبی حفظ می کند، اما منجر به اعوجاج مکانی جدی نیز می شود. نتیجه به دست آمده توسط روش PGMAN رقابتی است. در مقابل، مدل پیشنهادی UMP-GAN تصویری را با کیفیت مکانی و طیفی مطلوب تولید می نماید.

بر اساس این سه مجموعه داده آزمایش، نتایج بصری برتری مدل پیشنهادی UMP-GAN را نشان می‌دهند و مزایای آموزش روش‌های بدون نظارت بر توزیع داده‌های اصلی را برجسته می‌نمایند. نتایج ادغام به دست آمده از مدل UMP-GAN از نظر توزیع مکانی و طیفی به تصاویر ورودی نزدیک‌تر هستند. این به این دلیل است که مدل پیشنهادی ویژگی‌های ادغام شده را مستقیماً از تصویر ورودی بدون هیچ پیش‌پردازشی در طول فرآیند آموزش به دست می‌آورد. علاوه بر این، مدل ارائه شده به طور کامل از اطلاعات چندمقیاسی از طریق بلوک‌های MDB استفاده می‌کند تا ویژگی‌های عمیق‌تر تصاویر ورودی را به طور کامل حفظ نماید.

۳-۳- بررسی‌های فراتر

همانگونه که در بخش‌های پیشین نیز بیان شد، مدل پیشنهادی از بلوک‌های MDB برای استخراج ویژگی‌های سلسله‌مراتبی متراکم از تصاویر LRMS و PAN ورودی استفاده می‌کند. لایه‌های کانولوشن چندمقیاسی و اتصال‌های متراکم محلی دو بخش مهم تشکیل دهنده این بلوک‌های متراکم هستند. از طرفی، ویژگی‌های غنی‌تر در مقیاس‌های مختلف را می‌توان با لایه‌های کانولوشن چندمقیاسی به دست آورد. از طرف دیگر، ساختار اتصال متراکم محلی امکان استفاده مجدد از ویژگی‌ها را برای کاهش اتلاف اطلاعات مکانی و طیفی تصاویر ورودی فراهم می‌نماید. این کار منجر به بهبود عمل‌کرد ادغام تصاویر سنجش از دور می‌گردد. آزمایشی برای تحلیل تأثیر تعداد بلوک‌های MDB بر روی مدل پیشنهادی UMP-GAN انجام شده است. سایر بخش‌های مدل بدون تغییر باقی می‌مانند، در حالی که تعداد بلوک‌های MDB از ۱ تا ۴ تغییر می‌کنند. نتایج تجربی کمی بر روی مجموعه داده‌های ماهواره‌ای WorldView-2 و GaoFen-2 در جدول ۴ برای تحلیل اثر تعداد بلوک‌های MDB در کارایی مدل پیشنهادی ارائه شده است. عمل‌کرد هم‌گرایی UMP-GAN نیز در تعداد مختلف بلوک‌های MDB در شکل ۸ نمایش داده شده است.

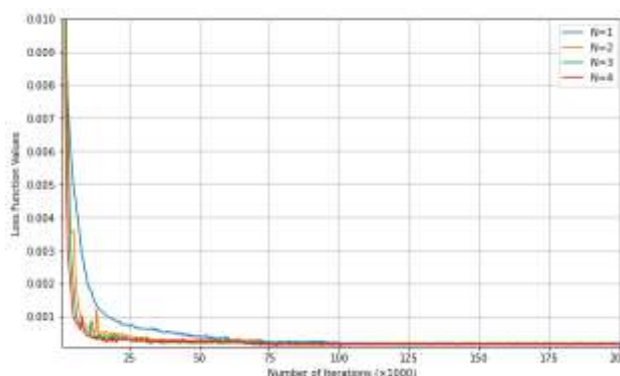
جدول (۴): نتایج آزمایش مدل UMP-GAN در تعداد مختلف MDBها

Table (4): Experimental results of the UMP-GAN model in different numbers of MDBs

تعداد MDBها	WorldView-2			GaoFen-2		
	D_λ	D_s	QNR	D_λ	D_s	QNR
$N=1$	0.0113±0.0020	0.0397±0.0012	0.9494±0.0014	0.0155±0.0023	0.0283±0.0028	0.9566±0.0037
$N=2$	0.0088±0.0012	0.0340±0.0009	0.9574±0.0009	0.0106±0.0017	0.0204±0.0015	0.9692±0.0030
$N=3$	0.0069±0.0013	0.0281±0.0001	0.9651±0.0002	0.0079±0.0013	0.0154±0.0016	0.9768±0.0031
$N=4$	0.0052±0.0010	0.0283±0.0002	0.9666±0.0004	0.0068±0.0015	0.0143±0.0013	0.9789±0.0027

توجه: مقادیر پررنگ نشان دهنده‌ی بهترین نتایج هستند.

می‌توان در جدول ۴ مشاهده کرد که وقتی تعداد بلوک‌های MDB از ۱ به ۳ تغییر می‌کند، تمام معیارهای بدون مرجع به طور قابل توجهی بهبود می‌یابند. مقدار QNR در $N=4$ کمی بالاتر از $N=3$ است. با در نظر گرفتن بار محاسباتی، مدل پیشنهادی UMP-GAN با سه بلوک MDB بیش‌ترین اطلاعات مکانی و طیفی را از تصویر LRMS و PAN ورودی می‌تواند استخراج نماید و به نتایج رضایت بخشی در آزمایش‌ها دست یابد. عمل کرد هم‌گرایی، همانطور که در شکل ۷ نشان داده شده است، نتیجه مشابهی را نشان می‌دهد. به طور خلاصه، تعداد بیشتر از بلوک‌های MDB منجر به پیچیده‌تر شدن شبکه و فرآیند آموزش می‌گردد. بنابراین، در مدل پیشنهادی از سه بلوک MDB استفاده خواهد شد.



شکل (۸): عمل کرد هم‌گرایی مدل UMP-GAN در تعداد مختلف MDBها

Fig. (8): Convergence performance of the UMP-GAN model in different numbers of MDBs

۴- نتیجه‌گیری

در این مقاله، یک مدل متراکم چندمقیاسی بدون نظارت مبتنی بر شبکه مخالف مولد به نام UMP-GAN برای مسئله پنشارپینگ تصاویر سنجش از دور پیشنهاد شده است. روش پیشنهادی پنشارپینگ را به عنوان یک مسئله چند وظیفه‌ای، یعنی حفظ هم‌زمان اطلاعات طیفی از یک تصویر چندطیفی (LRMS) و جزئیات مکانی از یک تصویر تک‌رنگ (PAN) فرموله می‌کند. این سیستم شامل یک شبکه مولد و دو شبکه تفکیک‌کننده برای کاهش اعوجاج‌های طیفی و مکانی در تصویر ادغام شده است. به منظور دستیابی به آموزش بدون نظارت و هم‌چنین بهبود هر چه بیشتر عمل‌کرد مدل پیشنهادی، شاخص بدون مرجع QNR به تابع هزینه شبکه مولد اضافه می‌شود. به عنوان یکی از مزایای روش‌های بدون نظارت، UMP-GAN می‌تواند بر روی تصاویر با وضوح کاهش یافته و یا تصاویر با وضوح کامل بدون نیاز به تصویر مرجع آموزش داده شود.

هم‌چنین، مدل ارائه شده بر روی تصاویر اصلی PAN و LRMS بدون نیاز به هیچ مرحله پیش‌پردازش برای حفظ سازگاری با محیط کاربرد عملی توسعه داده شده است. آزمایش‌های گسترده بر روی مجموعه داده‌های Worldview-2 (WV2)، GaoFen-2 (GF2) و QuickBird (QB) نشان می‌دهند که الگوریتم پیشنهادی معیارهای ارزیابی بهتری را به دست می‌آورد و توزیع اطلاعات طیفی و جزئیات مکانی پایه را به خوبی با از دست دادن اندک اطلاعات تحت شرایط بدون نظارت در تصویر ادغام شده، حفظ می‌نماید. علاوه‌براین، یک بررسی فراتر نیز برای نشان دادن اعتبار روش پیشنهادی انجام شده است. عمل‌کرد مناسب در آزمایش بر روی تصاویر با وضوح کامل و نتایج رضایت‌بخش در تصاویر با مقیاس کاهش یافته، توانایی قدرتمند مدل UMP-GAN را نشان می‌دهد.

۵- منابع

- M. Maneshi, H. Ghassemian and M. Imani. (2020). Sparse Representation of Injected Details for MRA-Based Pansharpening. IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Ahmedabad, India, 86–89.
- H. Zhou, Q. Liu and Y. Wang. (2021). PGMAN: An Unsupervised Generative Multiadversarial Network for Pansharpening. in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 6316–6327.
- Te-Ming Tu, P. S. Huang, Chung-Ling Hung and Chien-Ping Chang. (2004). A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. in IEEE Geoscience and Remote Sensing Letters, 1(4), 309–312.
- C. A. Laben and B. V. Brower. (2000). Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. U.S. Patent 6011875.
- V. P. Shah, N. H. Younan and R. L. King. (2008). An Efficient Pan-Sharpener Method via a Combined Adaptive PCA Approach and Contourlets. in IEEE Transactions on Geoscience and Remote Sensing, 46(5), 1323–1335.
- M. Amin-Naji, A. Aghagolzadeh and M. Ezoji, M. (2020). CNNs hard voting for multi-focus image fusion. Journal of Ambient Intelligence and Humanized Computing, 11, 1749–1769.
- Johnson, A. Alahi, and L. Fei-Fei. (2016). Perceptual losses for real-time style transfer and super-resolution. in Proc. Eur. Conf. Comput. Vis., 694–711.
- M. Amin-Naji, A. Aghagolzadeh and M. Ezoji, M. (2019). Ensemble of CNN for multi-focus image fusion. Information Fusion, 51, 201–214.
- C. Ledig et al. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 105–114.
- D. P. Kingma and M. Welling. (2014). Auto-Encoding Variational Bayes,” in Int. Conf. Learn. Representations, 1–14.
- I. Goodfellow et al. (2014). Generative adversarial nets,” in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q.

Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2672–2680. [Online]. Available: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>

G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa. (2016). Pansharpening by convolutional neural networks. *Remote Sens.*, 8(7), 594–615.

J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley. (2017). PanNet: A deep network architecture for pan-sharpening. in *Proc. IEEE Int. Conf. Comput. Vis.*, 5449–5457.

K. He, X. Zhang, S. Ren, and J. Sun. (2016). Deep residual learning for image recognition. in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 770–778.

Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang. (2018). A multiscale and multidepth convolutional neural network for remote sensing imagery pansharpening. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 11(3), 978–989.

D. Wang, Y. Li, L. Ma, Z. W. Bai, and J. C. W. Chan. (2019). Going deeper with densely connected convolutional neural networks for multispectral pansharpening. *Remote Sens.*, 11(22), 19.

J. Peng, L. Liu, J. Wang, E. Zhang, X. Zhu, et al. (2021). PSMD-Net: A Novel Pan-Sharpener Method Based on a Multiscale Dense Network. in *IEEE Transactions on Geoscience and Remote Sensing*, 59(6), 4957–4971.

X. Liu, Q. Liu, and Y. Wang. (2020). Remote sensing image fusion based on two-stream fusion network. *Inf. Fusion*, 55, 1–15.

L. Wald, T. Ranchin, and M. Mangolini. (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.*, 63(6), 691–699.

Q. Liu, H. Zhou, Q. Xu, X. Liu and Y. Wang. (2021). PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpener. in *IEEE Transactions on Geoscience and Remote Sensing*, 59(12), 10227–10242.

X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley. (2017). Least squares generative adversarial networks. in *Proc. IEEE Int. Conf. Comput. Vis.*, 2794–2802.

E-mail: mehrshad_brm@nit.ac.ir; aghagol@nit.ac.ir

- J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang. (2020). Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion*, 62, 110–120.
- G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. (2017). Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2261–2269.
- Zhou Wang and A. C. Bovik. (2002). A universal image quality index,” in *IEEE Signal Processing Letters*, 9(3), 81–84.
- D. P. Kingma and J. Ba. (2017). Adam: A method for stochastic optimization. arXiv:1412.6980.
- L. Wald. (2000). Quality of high resolution synthesised images: Is there a simple criterion?. in *Proc. 3rd Conf. “Fusion Earth Data: Merging Point Meas., Raster Maps Remotely Sensed Images”*, 99–103.
- G. Vivone. (2019). Robust band-dependent spatial-detail approaches for panchromatic sharpening. *IEEE Trans. Geosci. Remote Sens.*, 57(9), 6421–6433.
- A. R. Gillespie, A. B. Kahle, and R. E. Walker. (1987). Color enhancement of highly correlated images. II. Channel ratio and ‘chromaticity’ transformation techniques. *Remote Sensing of Environment*, 22(3), 343–365.
- J. Liu. (2000). Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18), 3461–3472.
- Z. Shao, Z. Lu, M. Ran, L. Fang, J. Zhou and Y. Zhang. (2020). Residual Encoder–Decoder Conditional Generative Adversarial Network for Pansharpening. in *IEEE Geoscience and Remote Sensing Letters*, 17(9), 1573–1577.



Unsupervised fusion of remote sensing images using a multiscale generative adversarial network

Abstract

In pansharpening, a low-resolution multispectral image (LRMS) and a high-resolution panchromatic image (PAN) are fused to produce a high-resolution multispectral image (HRMS). Recent studies have shown that convolutional neural networks can be used for sharpening remote sensing images with excellent results. However, there are still two major problems that need to be resolved. Despite the lack of ideal HRMS images for learning, most current methods require more effort to generate simulated data. On the other hand, these methods usually ignore the rich spatial information in panchromatic images. To address these issues, we have proposed an unsupervised fusion framework based on a multiscale dense network called UMP-GAN for pansharpening. This framework, which employs generative adversarial networks, can be trained directly on full-resolution images without requiring any preprocessing step. First, a multiscale dense generator network is proposed to extract features from original input images to generate HRMS images. In the following, two separate discriminator networks are used to protect the spectral information and spatial details of the input images in the final image. Therefore, the proposed method provides the possibility of training two discriminating networks, each of which has a different and complementary task. Finally, new loss functions are proposed to improve the quality of the fused image and perform training under unsupervised settings. This method can deepen the exchange of gradient information between the generator network and the discriminator networks, optimize the network training process, and accelerate network convergence. WorldView-2 (WV2), GaoFen-2 (GF2), and QuickBird (QB) satellite images have been used to simulate and validate the proposed UMP-GAN method. The results show that the proposed method performs better than other state-of-the-art models in the fusion of remote sensing images.

Keywords: Remote Sensing Image Fusion, Multiscale Dense Block, Pansharpening, Convolutional Neural Networks, Generative Adversarial Networks, Unsupervised Learning.