نشریه علمی مهندسی فناوری اطلاعات مکانی

سال سیزدهم، شماره اول، بهار ۱۴۰۴

Vol.13, No.1, Spring 2025

۹۵ -۱۲۰

مقاله پژوهشی

شناسایی تغییرات چند کلاسه ساختمانی با استفاده از شبکه‌های یادگیری عمیق مبتنی بر داده‌های سه‌بعدی هوایی و ماهواره‌ای

اکرم افتخاری[1*]، فرهاد صمدزادگان[2]، فرزانه دادرس جوان[3]

۱- دانشجوی دکتری سنجش از دور، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی، دانشگاه تهران

۲- استاد دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی، دانشگاه تهران

۳- استادیار دانشکده علوم اطلاعات جغرافیایی و رصد زمین (*ITC*)، دانشگاه توئنته

تاریخ دریافت مقاله: ۱۴۰۳/۱۰/۱۵ تاریخ پذیرش مقاله: ۱۴۰۴/۰۳/۱۸

چکیده

تشخیص تغییرات سه‌بعدی در ساختمان‌ها نقشی حیاتی در نظارت شهری، توسعه پایدار، و مدیریت بلایا دارد. این پژوهش روشی نوآورانه برای تشخیص تغییرات چند کلاسه ساختمان‌ها ارائه می‌کند. روش پیشنهادی با ترکیب دو معماری پیشرفته شامل شبکه مبدل سیامی و مکانیزم‌های توجه مکانی - کانالی، قابلیت تشخیص دقیق تغییرات ساختمانی را به‌صورت خودکار فراهم می‌سازد. نوآوری اصلی این کار در طراحی یک بلوک کدگذار مبتنی بر توجه دوگانه است که به‌صورت هم‌زمان روابط مکانی محلی و وابستگی‌های کانالی را برای شناسایی تغییرات تحلیل می‌کند . یکی از چالش‌های کلیدی در زمینه شناسایی تغییرات، توزیع نامتعادل کلاس‌ها (مانند ساختمان‌های بدون تغییر، سازه‌های جدید و ساختمان‌های تخریب‌شده) است. برای حل این مشکل، از روش‌هایی نظیر تقویت مؤثرتر داده‌ها و استخراج قطعات همپوشان در مرحله پیش‌پردازش استفاده شده است. روش پیشنهادی بر روی یک مجموعه‌داده استریو از ماهواره *GeoEye-1* با قدرت تفکیک ۰٫۵ متر و یک مجموعه‌داده استریو هوایی با قدرت تفکیک ۰٫۰۸ متر پیاده‌سازی شده است. با اعمال روش پیشنهادی در آزمایش‌ها، ضرایب کاپای ۹۴٪ و ۹۳٪ برای مجموعه‌داده‌های ماهواره‌ای و هوایی به‌دست‌آمده است که در مقایسه با سایر روش‌های پیشرفته مانند *ChangeFormer* با ضریب کاپا ۹۱٪ برای هر دو مجموعه‌داده، بهبود قابل‌ملاحظه‌ای (افزایش ۳ درصدی ضریب کاپا) حاصل شده است. این مدل با بهبود استخراج ویژگی‌ها و عملکرد مناسب روی داده‌های متنوع، به ابزاری قدرتمند برای نظارت بر محیط‌های شهری تبدیل شده و راهکاری مقیاس‌پذیر و قابل‌اعتماد برای برنامه‌ریزی و مدیریت شهری ارائه می‌دهد.

کلیدواژه‌ها : داده‌های سنجش‌ازدور سه‌بعدی، تشخیص تغییر چند کلاسه ساختمانی، شبکه‌های مبدل، بلوک‌های توجه مکانی و کانالی، افزایش داده.

* نویسنده مکاتبه‌کننده: گروه فتوگرامتری و سنجش‌ازدور، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران.

تلفن: ۰۲۱۶۱۱۱۴۲۵۶

Email: a.eftekhari@ut.ac.ir

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

۱- مقدمه

تشخیص تغییرات در ساختمان‌ها یکی از موضوعات اساسی و پراهمیت در تحقیقات سنجش از دور محسوب می‌شود. این فرایند شامل شناسایی و اندازه‌گیری تغییرات مکانی در ساختمان‌ها با استفاده از دو یا چند تصویر ماهواره‌ای یا هوایی هم‌ثبت‌شده[1] است [۱]. شناسایی تغییرات نقش حیاتی در برنامه‌ریزی و توسعه شهری ایفا می‌کند، زیرا اطلاعات ارزشمندی درباره تغییرات و توسعه شهری ارائه می‌دهد [۲]. با گسترش افقی و عمودی شهرها به‌دلیل رشد سریع شهرنشینی و افزایش جمعیت، تقاضا برای اطلاعات دقیق و به‌موقع درباره تغییرات ساختمان‌ها روزبه‌روز افزایش می‌یابد. این داده‌ها از کاربردهای متعددی مانند واکنش به بلایا [۳] ، برنامه‌ریزی کاربری [۴]، و نظارت محیطی [۵] پشتیبانی کرده و به‌عنوان زیربنای مدیریت پایدار شهری شناخته می‌شوند.

داده‌های سنجش‌ازدور سه‌بعدی، شامل تصاویر استریو[2] از سکوهای ماهواره‌ای و هوایی و مدل‌های رقومی سطح (*DSM*)[3]مشتق‌شده از لیدار، نقشی اساسی در ثبت اطلاعات جامع درباره تغییرات ساختمان‌ها ایفا می‌کنند. این داده‌ها تغییرات ارتفاعی و مسطحاتی ساختمان‌ها را با دقت بالا شناسایی و تحلیل می‌کنند [۶]. روش‌های سنتی تشخیص تغییرات سه‌بعدی به طور کلی به دو دسته اصلی تقسیم می‌شوند: مقایسه هندسی و تحلیل طیفی- هندسی[4][۷]. روش‌های مقایسه هندسی، مانند اختلاف ارتفاع و محاسبات فاصله اقلیدسی، حساسیت بالایی به خطاهای ثبت دارند و معمولاً به سناریوهای ۲/۵ بعدی محدود می‌شوند. در مقابل، روش‌های مبتنی بر تفاوت‌های تصویری، اگرچه مقاومت بیشتری در برابر خطاهای تطابق دارند، در مناطق همگن با مشکلاتی مواجه می‌شوند و به‌شدت به اطلاعات دقیق سه‌بعدی وابسته هستند [۸]. از سوی دیگر، روش‌های تحلیل طیفی-هندسی با ادغام داده‌های طیفی و هندسی از طریق پالایش پس‌پردازش، ادغام مستقیم ویژگی‌ها، و تکنیک‌های پس‌طبقه‌بندی، تلاش می‌کنند دقت تشخیص را بهبود بخشند. با وجود پتانسیل بالای این روش‌ها برای افزایش دقت، آن‌ها همچنان به دقت اولیه مقایسه‌های هندسی وابسته‌اند و با چالش‌هایی مانند تنظیم پارامترها و ادغام داده‌ها مواجه هستند که می‌تواند کارایی کلی آن‌ها را محدود کند [۹ و ۱۰]. روش‌های سنتی شناسایی تغییرات سه‌بعدی با چالش‌های عمده‌ای از جمله جابجایی داده‌ها، تأثیرات سایه، انسدادها، و ناهماهنگی‌های طیفی روبه‌رو هستند که این عوامل اغلب منجر به کاهش دقت و نارسایی در تشخیص تغییرات می‌شوند [۱۱].

در سال‌های اخیر، یادگیری عمیق به‌عنوان ابزاری قدرتمند برای تشخیص تغییرات در داده‌های سنجش از دور معرفی شده است. روش‌های تشخیص تغییر مبتنی بر یادگیری عمیق به روش‌های مبتنی بر طبقه‌بندی و مبتنی بر متریک طبقه‌بندی می‌شوند [۱۲]. روش‌های مبتنی بر طبقه‌بندی، تصاویر زمان اول و دوم را در یک بردار مشترک ترکیب می‌کنند و ویژگی‌های عمیق از تصاویر دو زمانی تولید شده استخراج می‌شوند. در چنین روش‌هایی، تصاویر دو زمانی وارد یک شبکه تک‌شاخه می‌شوند و سپس تغییرات بر اساس امتیاز تغییرشان دسته‌بندی می‌شوند [۱۳]. به‌عنوان مثال، شبکه *Unet++* که توسط آقای پنگ و همکارن (۲۰۱۹) معرفی شد، از یک شبکه *Unet* بهبود یافته سرتاسری برای استخراج ویژگی‌های کلی و جزییات و ترکیب آنها برای ایجاد نقشه تغییرات نهایی با دقت بالا توسعه دادند [۱۰]. در روش مبتنی بر متریک، تغییرات با مقایسه فاصله بین ویژگی‌های دو تصویر به دست می آید. در این روش معمولا هر تصویر وارد یک شبکه مجزا می‌شود و شبکه به‌صورت دو شاخه است. فضای

---

[1] *Co-registered*
[2] *Stereo*
[3] *Digital surface model*
[4] *Geometric-spectrum analysis*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

ویژگی طوری آموزش داده می‌شود که بردارهای تعبیه شده مشابه نزدیکتر شوند و بردارهای تعبیه شده ناهمسانی از یکدیگر دور شوند [۱۴]. این روش برای ارزیابی جزئیات جفت تصویر مؤثرتر از روش مبتنی بر طبقه بندی است [۱۵]. فضای ویژگی را می توان با شبکه های سیامی[1]شکل داد که شامل دو شبکه با وزن مشترک یا دو شبکه مجزا با وزن مستقل است. ژانگ و لو، از شبکه یادگیری مشترک طیفی-فضایی (*SSJLN*)[2]و روش‌های شبکه عصبی پیچشی(*CNN*) [3] سیامی برای تشخیص تغییر تصاویر چندطیفی استفاده کردند [۱۶]. آقای ژو و همکاران (۲۰۲۱) یک شبکه تشخیص تغییر ساختمان را پیشنهاد کردند که لایه‌های بهبود چند ویژگی را از اطلاعات شاخه های مرکزی و جانبی استخراج می کرد [۱۵]. در تحقیق دیگر شبکه *SNUNET* معرفی شده که یک شبکه سیامی متصل متراکم است که برای انتقال اطلاعات فشرده بین لایه های شبکه برای کاهش از دست دادن اطلاعات محلی سازی در لایه های عمیق معرفی شد [۱۷].

به طور کلی در این سال‌ها شبکه‌های بسیار زیادی جهت شناسایی تغییرات از تصاویر سنجش از دوری معرفی شده اند که اغلب از نوع روش‌های مبتنی بر متریک مانند *FC-EF* ، *STANet*، *DASNet*، *ChangeFormer* و *IDET* بوده که در تشخیص تغییرات ساختمانی دو یعدی نتایج قابل‌توجهی به دست آورده‌اند [۱، ۱۸، ۱۹، ۲۰ و ۲۱]. با این حال، کاربرد شبکه‌های عمیق در تشخیص تغییرات سه‌بعدی همچنان در مراحل اولیه توسعه قرار دارد. به‌عنوان مثال، محمدی و صمدزادگان (۲۰۲۰) اطلاعات دوبعدی و سه‌بعدی با استفاده از روش‌های شبکه *CNN* و ادغام داده‌های سلسله‌مراتبی برای تجزیه و تحلیل تغییرات چندسطحی و طبقه‌بندی دقیق ترکیب کردند [۲۲]. همچنین، ژلیس و همکاران (۲۰۲۳) از شبکه‌های سیامی و شبکه پیچشی مبتنی بر نقاط کلیدی (*KPConv*[4]) محلی برای تشخیص تغییرات چندگانه در داده‌های سه‌بعدی استفاده کرده‌اند. *KPConv* به دلیل ساختار خاص خود، توانایی استخراج ویژگی‌های مکانی دقیق‌تر از داده‌های ابرنقطه را دارد [۲۳]. در تحقیقی دیگر کو و همکاران (۲۰۲۱)، شبکه *SiamGCN* که یک شبکه سیامی مبتنی بر گراف است را معرفی کردند. این شبکه با استفاده از عملگرهای پیچشی لبه ای، ویژگی‌های تغییر در ابر نقاط سه‌بعدی را استخراج می‌کند. نتایج نشان داد که روش‌های مبتنی بر یادگیری عمیق در ترکیب با تکنیک‌های هندسی مانند استخراج لبه عملکرد بهتری در مواجهه با پیچیدگی داده‌های واقعی دارند [۲۴]. مارسوکی و همکاران (۲۰۲۳) نیز روش یادگیری عمیق مبتنی بر مبدل چندوظیفه‌ای تصاویر دو زمانه (*MTBIT*[5]) را توسعه داده‌اند که بدون وابستگی مستقیم به داده‌های ارتفاع، نقشه‌های تغییرات دوبعدی و سه‌بعدی را از تصاویر اپتیکی دو زمانه تولید می‌کند. مدل پیشنهادی از معماری‌های شبکه‌ای پیشرفته برای استخراج ویژگی‌های مکانی-زمانی و پیش‌بینی تغییرات سه‌بعدی استفاده کرده است. نتایج نشان داده که این روش می‌تواند با دقت بالا تغییرات ارتفاعی و مکانی را در غیاب داده‌های لیدار یا ابرنقطه شناسایی کند [۲۵]. گمرکی و همکاران (۲۰۲۳) به طراحی مدلی خودکار برای تشخیص تغییرات چندگانه سه‌بعدی در ساختمان‌ها با استفاده از معماری شبکه کدگذار-کدگشا[6]پرداختند. رویکرد پیشنهادی آن‌ها، ویژگی‌های مکانی و ارتفاعی داده‌های ورودی را استخراج و با بهره‌گیری از مکانیسم‌های توجه چندگانه و وزن‌دهی تطبیقی کلاس‌ها تغییرات را شناسایی می‌کند. نتایج

[1] *Siamese*
[2] *spectral-spatial joint learning network*
[3] *Convolution neural network*
[4] *Key points convolution*
[5] *MultiTask bitemporal images transformer*
[6] *Encoder-Decoder*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

نشان داد که این مدل می‌تواند تغییرات مختلف مانند تخریب، ساخت‌وساز، یا جابه‌جایی ساختمان‌ها را با دقت بالا و پایداری در مجموعه‌داده‌های نامتوازن شناسایی کند [۲۶].

چالش‌های کلیدی این روش‌ها شامل موارد زیر است:

۱- مدیریت مجموعه‌داده‌های نامتوازن: بسیاری از رویکردهای فعلی شناسایی تغییرات سه‌بعدی به طور کافی به مسئله مجموعه‌داده‌های به‌شدت نامتوازن نمی‌پردازند که می‌تواند منجر به شناسایی ضعیف کلاس‌های نادر مانند ساختمان‌های جدید یا تخریب‌شده شود. این محدودیت تأثیر آنها را در وظایف نظارت واقعی شهری کاهش می‌دهد، جایی که انواع خاصی از تغییرات ساختمانی کمتر رایج هستند.

۲- محدودیت در قابلیت استخراج ویژگی: اگرچه برخی روش‌ها مکانیزم‌های توجه یا *CNN* ها را برای استخراج ویژگی به کار می‌گیرند، اغلب نمی‌توانند طیف کامل ویژگی‌های مکانی و زمینه‌ای مورد نیاز برای تشخیص دقیق تغییرات سه‌بعدی را ثبت کنند.

۳- ادغام داده‌های چندمنبعی: روش‌هایی که تنها به یک نوع داده (مانند تصاویر اپتیکی یا ابرنقاط) متکی هستند، در تعمیم‌پذیری برای سناریوهای مختلف شهری با مشکل مواجه می‌شوند.

در پاسخ به این چالش‌ها، این مطالعه روشی نوین برای شناسایی تغییرات چند کلاسه ساختمان‌ها به‌صورت سه‌بعدی ارائه می‌دهد که از شبکه دوشاخه سیامی مبتنی بر مبدل استفاده می‌کند و بلوک‌های توجه مکانی و کانالی را در بر دارد. این روش به‌منظور شناسایی دقیق انواع مختلف تغییرات ساختمان‌ها مانند ساختمان‌های جدید و تخریب‌شده با استفاده از داده‌های ماهواره‌ای و هوایی طراحی شده است. ترکیب *CNN*ها با مکانیسم‌های توجه امکان استخراج ویژگی‌های مؤثر را فراهم کرده و دقت و استحکام شناسایی تغییرات ساختمان را در محیط‌های شهری پیچیده بهبود می‌بخشد.

یکی از جنبه‌های این مطالعه، استفاده از یک مجموعه‌داده سه‌بعدی جدید است که از طریق سکو هواپیما جمع‌آوری شده است، از جمله تصاویر و *DSM* این مجموعه‌داده اطلاعات مکانی با وضوح خیلی بالا از مناطق شهری ارائه می‌دهد و تنوع منابع داده‌ای استفاده شده در این تحقیق را افزایش می‌دهد. ادغام این مجموعه‌داده جدید به ارزیابی جامع مطالعه می‌افزاید، زیرا مکمل داده‌های ماهواره‌ای استفاده‌شده است و بدین ترتیب ارزیابی جامع‌تری از تغییرات ساختمان‌ها در مناطق شهری ارائه می‌دهد.

اهداف خاص این تحقیق به شرح زیر است:

۱- توسعه یک شبکه دوشاخه سیامی مبتنی بر مبدل با مکانیزم‌های توجه مکانی و کانالی به‌منظور ارتقای دقت شناسایی تغییرات ساختمان‌ها به‌صورت سه‌بعدی.

۲- پرداختن به مشکل مجموعه‌داده‌های بسیار نامتعادل با استفاده از تکنیک‌های تقویت مؤثرتر داده‌ها و همپوشانی بخش‌ها، به‌گونه‌ای که شناسایی بهتر کلاس‌هایی با نمونه کم مانند ساختمان‌های جدید و تخریب‌شده ممکن گردد.

۳- اعتبارسنجی روش پیشنهادی بر روی دو مجموعه‌داده شهری متمایز ماهواره‌ای و هوایی، از جمله مجموعه‌داده هوایی جدید معرفی‌شده، و ارزیابی عملکرد آن با استفاده از دقت، بازخوانی، امتیاز اف-۱ و ضریب کاپا.

۴- مقایسه روش پیشنهادی با تکنیک‌های پیشرفته موجود، به‌گونه‌ای که برتری و استحکام آن در شناسایی تغییرات چند کلاسه ساختمان‌ها در محیط‌های شهری مختلف به‌وضوح نشان داده شود.

۲- مجموعه‌داده‌ها

در این تحقیق، روش پیشنهادی برای شناسایی تغییرات سه‌بعدی چند کلاسه ساختمان‌ها بر روی دو

[1] *Patch*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

مجموعه‌داده متمایز که از سکوهای ماهواره‌ای و هوایی جمع‌آوری شده‌اند، اعمال می‌شود. در ادامه، ویژگی‌های این مجموعه‌داده‌ها به طور دقیق توصیف می‌شوند تا نقش و اهمیت آن‌ها در برجسته‌سازی این مطالعه مشخص گردد.

۲-۱- مجموعه‌داده تصاویر استریو ماهواره‌ای *GeoEye-1*

اولین مجموعه‌داده شامل تصاویر استریو ماهواره‌ای از شهر تهران است که از وب‌سایت *rslab.ut.ac.ir* به طور رایگان دانلود شده و توسط گمرکی و همکاران (۲۰۲۳) معرفی شده است [۲۶]. این داده‌ها شامل دو تصویر استریو از ماهواره *GeoEye-1* با قدرت تفکیک مکانی ۰/۵ متر مربوط به سال‌های ۲۰۰۹ و ۲۰۱۳ هستند. منطقه مورد مطالعه، منطقه ۲۲ تهران، یک ناحیه شهری تازه‌توسعه‌یافته در شمال غرب شهر است که طی سال‌های اخیر شاهد تغییرات عمده‌ای مانند احداث برج‌های مسکونی، مراکز تجاری، و زیرساخت‌های جدید بوده است. این مجموعه‌داده شامل تصاویر با سه باند طیفی سبز، قرمز و آبی (*RGB*)، *DSM* استخراج‌شده از ابرنقاط، و داده‌های حقیقت‌زمینی است. تغییرات به چهار لایه تقسیم شده‌اند: "غیرطبقه‌بندی‌شده، بدون تغییر، ساختمان‌های جدید و ساختمان‌های تخریب‌شده"، که با استفاده از نرم‌افزار گلوبال‌مپر (*Global Mapper*) ایجاد شده‌اند. شکل (۱) نمای بصری از این مجموعه‌داده را ارائه می‌دهد.

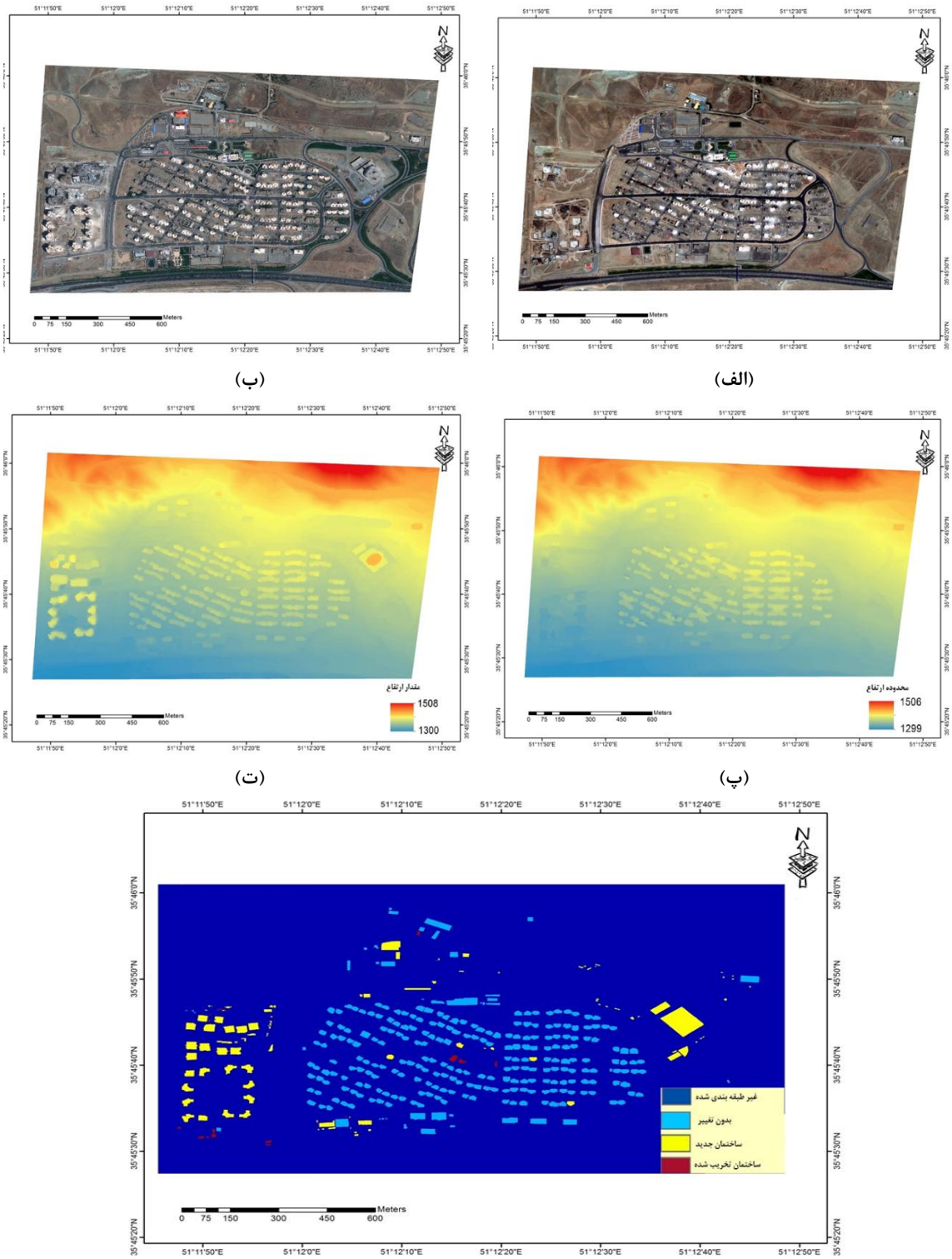
۲-۲- مجموعه داده تصاویر استریو هوایی

مجموعه‌داده دوم شامل تصاویر استریو هوایی است که توسط هواپیما از منطقه ۱۸ تهران در سال‌های ۲۰۰۶ و ۲۰۱۴ ثبت شده‌اند. این منطقه که دارای ساختار شهری متراکم است، در این دوره به دلیل رشد جمعیت تغییرات قابل‌توجهی را در ارتفاع ساختمان‌ها تجربه کرده است. از جمله این تغییرات می‌توان به تبدیل ساختمان‌های یک‌طبقه به چندطبقه اشاره کرد. این مجموعه داده مساحتی برابر با ۱۶/۶ هکتار را پوشش می‌دهد و شامل تصاویر قائم (*Orthophoto*) *RGB* و *DSM* با قدرت تفکیک مکانی بسیار بالای ۰/۰۸ متر است. داده‌های حقیقت زمینی با مقایسه دقیق تصاویر و *DSM*ها استخراج شده و در چهار کلاس متمایز "غیر طبقه‌بندی‌شده، بدون تغییر، ساختمان‌های جدید و ساختمان‌های تخریب‌شده" در محیط نرم‌افزاری *QGIS* دسته‌بندی شده‌اند. شکل (۲) نمایی از تصاویر، *DSM*ها و داده‌های حقیقت زمینی این مجموعه‌داده را نشان می‌دهد. جدول (۱) ویژگی‌های اصلی هر دو مجموعه‌داده را به طور خلاصه بیان می‌کند.

جدول۱: جزئیات جامع مجموعه‌داده‌ها در این مقاله

مجموعه‌داده	زمان	ورودی‌ها به شبکه	قدرت تفکیک مکانی (متر)	مساحت (کیلومترمربع)
تصاویر استریو ماهواره‌ای *GeoEye-1* از تهران	۲۰۰۹ و ۲۰۱۳	تصویر *RGB* و *DSM* در زمان اول و زمان دوم	۰/۵	۱/۵۳
تصاویر استریو هوایی از تهران	۲۰۰۶ و ۲۰۱۴	تصویر *RGB* و *DSM* در زمان اول و زمان دوم	۰/۰۸	۰/۱۶۶

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]



شکل۱: مجموعه‌داده ماهواره‌ای *GeoEye-1* از منطقه ۲۲ شهر تهران، (الف) تصویر *RGB* زمان اول، (ب) تصویر *RGB* زمان دوم، (پ) *DSM* زمان اول، (ت) *DSM* زمان دوم، (ث) داده حقیقت زمینی.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

شکل۲: مجموعه‌داده هوایی از منطقه ۲۲ شهر تهران، (الف) تصویر *RGB* زمان اول، (ب) تصویر *RGB* زمان دوم، (پ)*DSM* زمان اول، (ت) *DSM* زمان دوم، (ث) داده حقیقت زمینی.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

۳- روش پیشنهادی

رویکرد پیشنهادی ما در این تحقیق برای شناسایی تغییرات چند کلاسه ساختمان‌ها با فرایند پیش‌پردازش شروع می‌شود. این مرحله شامل مدیریت داده‌های ورودی، استخراج بخش‌های تصویر و تقویت مؤثرتر داده‌ها به‌منظور مقابله با عدم تعادل کلاس‌ها است. پس از آن، داده‌ها به مجموعه‌های آموزشی، آزمایشی و اعتبارسنجی تقسیم می‌شوند. در مرحله بعد، یک شبکه مبدل مبتنی بر توجه معرفی می‌شود. این شبکه دوشاخه سیامی با بلوک‌های توجه مکانی و کانالی است که به طور مؤثر ویژگی‌ها را برای شناسایی تغییرات مختلف ساختمان استخراج می‌کند. در نهایت، در مرحله پیش‌بینی و ارزیابی مدل، عملکرد شبکه را بر روی دو مجموعه‌داده بررسی شده و اثربخشی این رویکرد را در سناریوهای شهری متنوع نشان داده می‌شود.

۳-۱- پیش پردازش داده‌ها

ساخت و تخریب ساختمان‌ها در مناطق شهری عموماً فرایندهای تدریجی هستند و به طور سریع اتفاق نمی‌افتند. در نتیجه، حجم ساختمان‌های بدون تغییر به طور قابل‌توجهی بیشتر از ساختمان‌هایی است که تغییر کرده‌اند. این پدیده در مجموعه‌داده‌های استفاده شده در این مطالعه مشهود بود. به همین دلیل، عدم تعادل در داده‌های حقیقت زمینی، به‌ویژه در کلاس‌های ساختمان‌های تازه ساخته شده و تخریب‌شده، منجر به کاهش دقت مدل شد. بدین ترتیب با تعیین وزن هر کلاس از رابطه (۱) که فرمول پیشنهادی کینگ و ژنگ است، نادر بودن یا نبودن هر کلاس در داده حقیقت زمینی مشخص گردید[۲۳].

$$W_i = \frac{n_{sample}}{n_{class} \times F_i} \qquad \text{رابطه(۱)}$$

در رابطه (۱)، W_i نمایانگر وزن هر کلاس، n_{sample} تعداد کل نمونه ها، n_{class} تعداد کلاس‌ها و F_i فراوانی هر کلاس در مجموعه داده‌ها را نشان می‌دهد. در داده‌های ماهواره‌ای، وزن داده‌های حقیقت‌زمینی برای کلاس‌های غیرطبقه‌بندی‌شده، بدون تغییر ساختمان، ساختمان‌های جدید و ساختمان تخریب‌شده به ترتیب ۰٫۲۸، ۳٫۱۷، ۸٫۰۱ و ۱۷۰٫۶۱ است. بنابراین، کلاس‌های ساختمان‌های جدید و تخریب‌شده نادر هستند. برای داده‌های هوایی، این نسبت‌ها برای چهار کلاس به ترتیب ۰٫۵۳، ۰٫۶۷، ۱٫۸۷ و ۷۲٫۲۶ است. در اینجا، کلاس تخریب‌شده نادر است. مقادیر وزن نشان‌دهنده عدم تعادل قابل توجه در داده‌های آموزشی است. برای حل مشکل عدم تعادل داده‌ها و به‌عنوان یک راه‌حل بهتر، تکنیک تقویت مؤثرتر بر روی بخش‌های داده‌ای که کلاس‌ها در آن‌ها نادر بودند، اعمال شد.

تقویت مؤثرتر که در این مطالعه اعمال شد، بهبود تکنیک *RandAugment* است که توسط کوبک و همکاران (۲۰۲۰) توسعه یافته است [۲۴]. این رویکرد به طور خاص برای ویژگی‌های تصاویر سنجش از دور سفارشی‌سازی شده و منجر به ایجاد ۱۵ روش تقویت مختلف گردیده است. در جدول (۲)، فهرست دقیقی از این روش‌ها، شامل دامنه شدت‌های اعمال شده و توضیحات خاص هر یک شرح داده شده است. در طول آموزش شبکه، تقویت‌های مختلف به جفت‌های مختلف داده‌های دوزمانه در همان دوره اعمال می‌شود، و همان جفت تصاویر ممکن است در دوره‌های مختلف تقویت‌های متفاوتی دریافت کنند تا از تصادفی بودن کافی اطمینان حاصل شود. با این حال، تقویت مشابه باید به طور مداوم به جفت داده‌های دوزمانه داده شده اعمال شود. علاوه بر تقویت مؤثرتر، از همپوشانی بخش‌ها برای افزایش بیشتر اندازه مجموعه‌داده‌ها و بهبود کارایی آموزش استفاده شد. همپوشانی بخش‌ها شامل تقسیم تصاویر ورودی به بخش‌های کوچکتر با همپوشانی است که این امر اطمینان می‌دهد اطلاعات مکانی بیشتری در هر بخش حفظ می‌شود و توانایی یادگیری مدل افزایش می‌یابد. درجه همپوشانی برای مجموعه‌داده‌های مختلف تنظیم شد تا عملکرد بهینه شود و استخراج ویژگی بهتر و دقت بالاتری حاصل گردد. این رویکرد به کاهش اثرات عدم تعادل داده‌ها کمک کرده و امکان استفاده حداکثری از داده‌های موجود در مراحل آموزش، اعتبارسنجی و ارزیابی را فراهم می‌آورد.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

جدول۲: لیست تکنیک‌های تقویت مؤثرتر داده‌ها، درجه و توضیحات آن‌ها

تقویت	درجه	توضیحات
روشنایی	(۰/۰۵، ۰/۹۵)	کنترل روشنایی تصویر
رنگ	(۰/۰۵، ۰/۹۵)	تنظیم تعادل رنگی تصویر
کنتراست	(۰/۰۵، ۰/۹۵)	کنترل کنتراست تصویر
همسان‌سازی	/	همسان‌سازی هیستوگرام تصویر
هویت	/	بازگشت تصویر اصلی
پستریزه	(۴، ۸)	کاهش تعداد بیت‌ها برای هر کانال رنگی
چرخش	(۳۰-، ۳۰)	چرخش تصویر
شارپنس	(۰/۰۵، ۰/۹۵)	تنظیم وضوح تصویر
شیب X	(۰/۳-، ۰/۳)	شیب دادن تصویر در محور افقی
شیب Y	(۰/۳-، ۰/۳)	شیب دادن تصویر در محور عمودی
سولاریزه	(۰، ۲۵۶)	معکوس کردن مقادیر پیکسل‌ها بالای یک آستانه
جابه‌جایی X	(۰/۳-، ۰/۳)	جابه‌جایی تصویر در محور افقی
جابه‌جایی Y	(۰/۳-، ۰/۳)	جابه‌جایی تصویر در محور عمودی
برش	(۰/۲۵، ۰/۳۵)	برش دادن ناحیه‌ای از تصویر
ترکیب برش‌ها	پارامتر کسری: ۰/۳ و ۰/۵	ترکیب بخش‌های تصاویر و برچسب‌های آن‌ها، با ترکیب نواحی تصویر برای ارتقای یادگیری ویژگی‌ها

۳-۲- شبکه مبدل مبتنی بر توجه

شبکه پیشنهادی این مطالعه یک شبکه دو شاخه‌ای سیامی است که تصاویر RGB و DSM ها را از دو نقطه زمانی مختلف به‌عنوان ورودی برای هر شاخه می‌گیرد. شبکه سیامی نوعی معماری شبکه عصبی است که شامل دو یا چند زیرشبکه یکسان با وزن‌های مشترک است. این شبکه‌ها ورودی‌های متفاوتی را می‌گیرند و ویژگی‌های آن‌ها را در فضای تعبیه[1]مقایسه می‌کنند. هدف اصلی این معماری، یادگیری شباهت یا تفاوت بین دو داده ورودی است. لذا شبکه سیامی در موضوع شناسایی تغییرات بسیار پرکاربرد هستند. معماری سیامی به دلیل اشتراک وزن‌ها و قابلیت تمرکز بر شباهت یا تفاوت بین داده‌ها، در شناسایی تغییرات عملکرد بسیار مؤثری دارد. با الهام از معماری‌های مبدل، شبکه ابتدا ویژگی‌ها را با استفاده از یک شبکه عصبی کانولوشنی ۳×۳ استخراج می‌کند. سپس یک بلوک کدگذاری‌کننده مبدل مبتنی بر توجه قرار دارد که شامل دو مکانیسم توجه است: توجه مکانی برای تمرکز بر روابط مکانی محلی و توجه کانالی برای استخراج وابستگی‌ها در میان کانال‌های ویژگی. این ترکیب استخراج ویژگی و پردازش توجه چهار بار تکرار می‌شود.
پس از هر بلوک کدگذار، یک لایه مکس‌پولینگ[2]برای کاهش ابعاد ویژگی‌ها استفاده می‌شود که به شبکه کمک می‌کند تا بر جزئیات مهم مرتبط با شناسایی

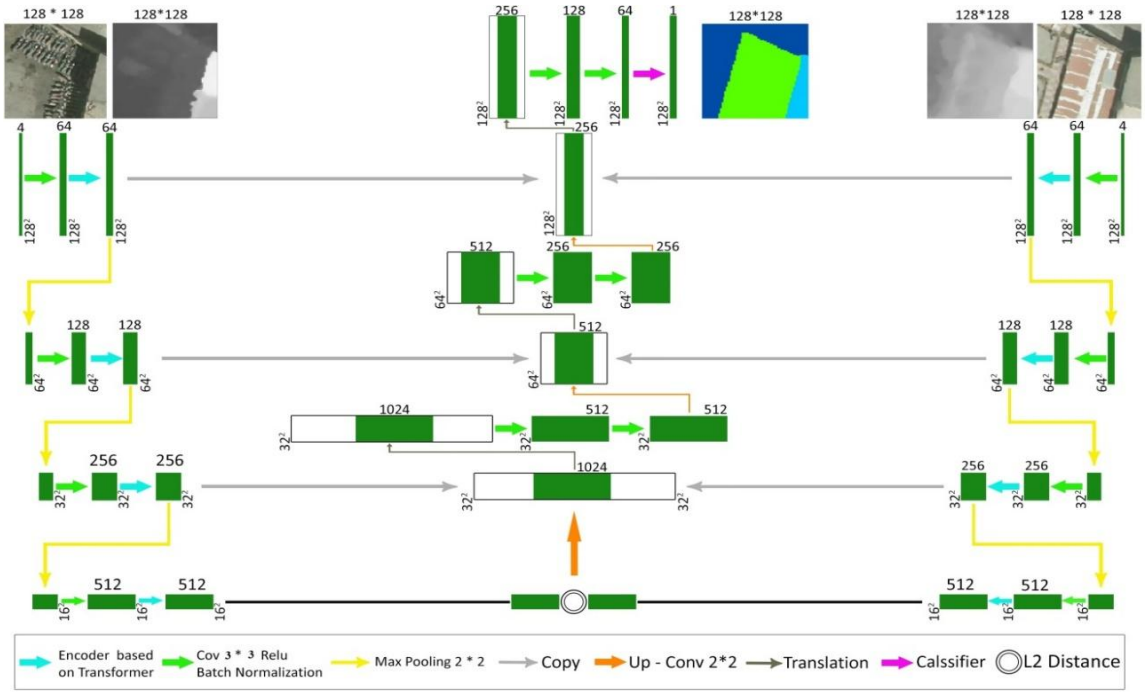
[1] *Embedding space*

[2] *Max Pooling*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

تغییرات متمرکز شود. ویژگی‌ها از هر شاخه از طریق چهار مرحله کدگذاری مقایسه می‌شوند و سپس فاصله اقلیدسی محاسبه می‌شود. فرایند افزایش وضوح (*Upsampling*) در شبکه کدگشا جزئیات سطح پایین را از کدگذار با ویژگی‌های سطح بالا ترکیب کرده و سپس از بازگشت پیچشی (*Deconvolution*) برای اصلاح خروجی استفاده می‌شود. عملیات بازگشت پراکنش (*backpropagation*) امکان انتقال مؤثر سیگنال خطا از تابع زیان را فراهم می‌کند و وزن‌ها را تنظیم کرده و داده‌های چندمنبعی از نقاط زمانی مختلف را به طور هم‌زمان پردازش می‌کند. در لایه آخر نیز از تابع فعال‌سازی سافت‌مکس (*Softmax*) جهت طبقه‌بندی ویژگی‌ها استفاده شده است. شکل(۳) یک نمایش دقیق از معماری شبکه را ارائه می‌دهد. در معماری شبکه، چهار لایه *CNN* و چهار لایه کدگذار مبتنی بر توجه برای استخراج ویژگی‌های عمیق از هر یک از داده‌های زمان اول و دوم استفاده شده است. همچنین، در بخش افزایش وضوح، هشت لایه *CNN* و چهار لایه بازگشت پیچشی استفاده شده است.



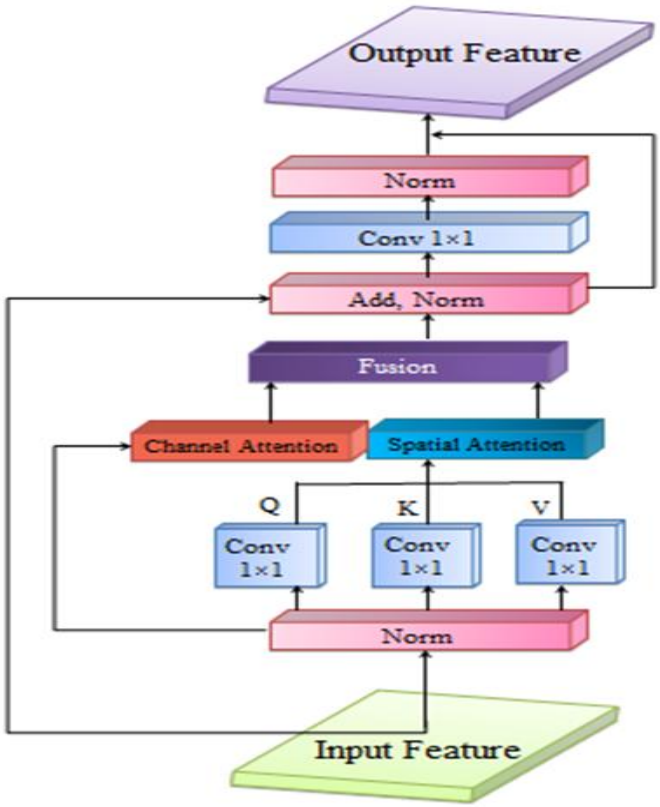
شکل۳: نمای کلی روش پیشنهادی شناسایی تغییرات چند کلاسه ساختمان‌ها مبتنی بر شبکه مبدل دو شاخه‌ای و مکانیزم توجه.

نمای نزدیک‌تر از بلوک کدگذار مبدل مبتنی بر توجه در شکل(۴) نشان داده شده است. نقش اصلی لایه کدگذار استخراج ویژگی‌ها بر اساس عواملی مانند چرخش، موقعیت و مقیاس اشیا است. این امر با استفاده از یک بلوک توجه مکانی برای دریافت اطلاعات موقعیتی و یک بلوک توجه کانالی برای تحلیل توزیع ویژگی‌ها در میان کانال‌ها حاصل می‌شود. یک شبکه *CNN* پارامترهای پرس‌وجو (*Q*)[1]، کلید (*K*)[2] و مقدار (*V*)[3] را تولید کرده که با خروجی‌های بلوک‌های توجه مکانی و توجه کانالی از طریق یک تابع جمعی ترکیب

[1] *Query*
[2] *Key*
[3] *Value*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

می‌شوند. این یکپارچگی باعث تقویت توانایی شبکه در استخراج ویژگی‌های معنادار و افزایش قدرت آن می‌شود. برای اطمینان از دامنه خروجی ثابت، یک بلوک نرمال‌سازی لایه به کار گرفته می‌شود تا مقادیر به طور معمول در بازه ۰ تا ۱ نگه داشته شوند و خروجی آن با نتایج قبلی ترکیب شود. بخش‌های بعدی، بلوک‌های توجه مکانی و توجه کانالی استفاده شده در این فرایند را به‌تفصیل توضیح خواهند داد.



شکل۴: معماری بلوک کدگذار مبتنی بر توجه

۳-۲-۱- بلوک توجه مکانی

در شبکه‌های عصبی پیچشی، بلوک توجه مکانی (*SAB*)[1] از روابط میان ویژگی‌های درون فضایی برای ایجاد نقشه توجه مکانی استفاده می‌کند [۲۹]. *SAB* به مدل کمک می‌کند تا بر روی نواحی مهم در تصویر تمرکز کرده و پیکسل‌های تغییر یافته و تغییر نیافته را بهتر تمایز دهد.

با بهره‌گیری از زمینه‌های غنی اطراف ویژگی‌های محلی، *SAB* به‌عنوان یک مکانیزم خود - توجه[2] عمل می‌کند و اطلاعات بلندمدت را برای بهبود کیفیت ویژگی‌های استخراج‌شده فراهم می‌کند. خود - توجه که به‌عنوان توجه درون‌سری نیز شناخته می‌شود، بر روی بخش‌های مهم درون یک توالی تمرکز می‌کند تا نمای کلی آن را محاسبه کند. این تکنیک که در استخراج وابستگی‌های بلندمدت موفق عمل می‌کند،

[1] *Spatial attention block*

[2] *Self-Attention*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

سه عنصر کلیدی دارد :پرس‌وجو، کلید، و مقدار. مدل، مجموع وزنی مقادیر را بر اساس میزان تطابق پرس‌وجو با کلیدها محاسبه کرده و به بخش‌های مرتبط از توالی توجه می‌کند. رابطه دقیق میان این عناصر به طور ریاضی در رابطه (۲) تعریف شده است.

$$att(q.s) = \sum_{i=1}^{T} \alpha_i(q.k_i)\, V_i$$ رابطه(۲)

در رابطه (۲) هر عنصر درون توالی منبع که با s نمایش داده می‌شود، یک بردار کلید (k_i) و یک بردار مقدار (V_i) ارائه می‌دهد. مدل امتیاز شباهت میان بردار پرس‌وجو (q) و هر کلید را با استفاده از تابع سافت‌مکس محاسبه می‌کند. این امتیاز اساساً نشان‌دهنده این است که چقدر پرس‌وجو با یک کلید خاص مطابقت دارد. این تنسورها به طور مستقل و مجزا توسط سه لایه پیچشی متفاوت تولید می‌شوند. این فرایند به مدل اجازه می‌دهد که ویژگی‌های مختلف را به شکل انعطاف‌پذیر و کارآمد پردازش کند [۳۰].

همان‌طور که در شکل (۵) نشان داده شده است، مکانیزم توجه مکانی یک نقشه ویژگی (F) با سه بعد: کانال‌ها(C) ، ارتفاع (H) و عرض (W) را به‌عنوان ورودی می‌گیرد. سه لایه پیچشی با ساختار مشابه، F را پردازش می‌کنند تا نقشه‌های ویژگی جدیدی تولید کنند: مقدار، کلید و پرس‌وجو همان نقشه‌های ویژگی جدید هستند که هر سه دارای بعد اصلی ($C\times H\times W$) می‌باشند. برای محاسبه ماتریس شباهت، مدل به تعامل هر عنصر (i) در پرس‌وجو با هر عنصر (j) در کلید می‌پردازد. در این مرحله، نقشه‌های ویژگی K و V به ابعاد جدیدی ($C\times N$) تغییر شکل می‌دهند، که در آن N نمایانگر تعداد کل عناصر مکانی ($H\times W$) است. سپس کلید تغییر شکل یافته ترانهاده شده و با مقدار تغییر شکل یافته (V) ضرب می‌شود. در نهایت، خروجی این ضرب از طریق یک لایه سافت‌مکس (بر اساس رابطه(۳)) پردازش شده و به یک نقشه توجه مکانی (Fs) با ابعاد ($N\times N$) تبدیل می‌شود. این نقشه توجه، نواحی مرتبط‌تر درون نقشه ویژگی را برجسته می‌کند و اطلاعات مکانی مهم را تقویت می‌نماید.

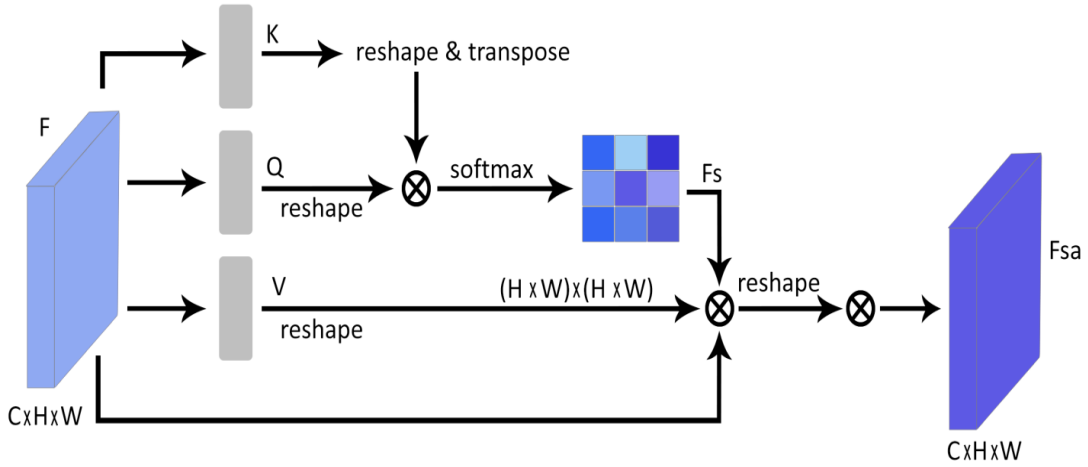
$$Fs_{ji} = softmax\left(\frac{K^T Q}{\sqrt{C}}\right)$$ رابطه(۳)

هر عنصر در نقشه توجه مکانی (Fs_{ji}) اهمیت ویژگی مقدار در موقعیت i را نسبت به ویژگی کلید در موقعیت j نشان می‌دهد. برای جلوگیری از غلبه تعداد کانال‌ها بر امتیازات، یک عامل نرمال‌سازی ($\sqrt{C}$) به‌کار گرفته می‌شود [۱۳]. مقادیر بالاتر در Fs_{ji} نشان‌دهنده ارتباط قوی‌تر بین این ویژگی‌ها است. سپس، بردار Q تغییر شکل داده شده و با Fs ضرب می‌شود، که به طور مؤثر نقشه ویژگی اصلی را بر اساس امتیازهای توجه وزن‌دهی می‌کند. در گام پایانی، یک عامل مقیاس‌گذاری (γ) به نتیجه وزن‌دهی شده اعمال می‌شود. این نتیجه به‌صورت عنصر به عنصر به نقشه ویژگی اصلی اضافه می‌شود تا تأثیر مکانیزم توجه مکانی تکمیل شود، همان‌طور که در رابطه (۴) نشان داده شده است.

$$Fsa_j = \gamma \sum_{i=1}^{N} (Fs_{ji} Q_j) + F_j$$ رابطه(۴)

عامل مقیاس ابتدا برابر صفر قرار می‌گیرد و به‌تدریج در طول فرایند آموزش افزایش می‌یابد. این عامل به کنترل تأثیر مکانیزم توجه بر نقشه ویژگی اصلی کمک می‌کند. طبق رابطه (۴) نشان می‌دهد که خروجی نهایی (Fsa) نسخهای بهبودیافته از نقشه ویژگی اصلی است که در آن هر عنصر، یک جمع وزنی از ویژگی‌ها است و بر اساس نمرات توجه مکانی محاسبه می‌شود. این فرایند به طور مؤثر، زمینه‌ای گسترده‌تر را در نظر گرفته و به‌صورت انتخابی بر نواحی اطلاعاتی که مکانیزم توجه شناسایی کرده، تأکید می‌کند. این کار باعث بهبود توانایی مدل در شبیه‌سازی ویژگی‌های معنایی مشابه می‌شود و منجر به فشردگی و انسجام بیشتر در نمایش کلاس‌ها می‌گردد. در نتیجه، شبکه قادر به تمایز بهتر بین تغییرات واقعی و تغییرات کاذب می‌شود.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]



شکل۵: معماری کلی بلوک توجه مکانی

۳-۲-۲- بلوک توجه کانالی

بلوک توجه کانالی *(CAB)* یکی از مؤلفه‌های کلیدی در معماری پیشنهادی است که مکانیزم توجه کانال را به مدل یادگیری عمیق وارد می‌کند[۲۹]. در شبکه‌های عمیق، کانال‌های مختلف در نقشه ویژگی‌ها اغلب با اشیاء یا مفاهیم مختلف مطابقت دارد. توجه کانال با تغییر اهمیت هر کانال، به‌عنوان ابزاری برای تعیین اولویت تمرکز در شناسایی یک شیء عمل می‌کند. همان‌طور که در شکل (۶) نشان داده شده است، *CAB* برای تولید ویژگی‌های جدید از عملیات پیچش استفاده نمی‌کند. به جای آن، ویژگی ورودی (D) با ابعاد $C\times H\times W$ به شکل $C\times N$ تغییر شکل داده می‌شود، که در آن N حاصل‌ضرب H و W است. سپس D تغییر شکل یافته با ترانهاده خود ضرب می‌شود تا مکانیزم توجه کانالی با اندازه $N\times N$ ایجاد شود. برای اعمال توجه کانالی، از تابع سافتمکس استفاده می‌شود، که در رابطه (۵) تعریف شده است. این فرایند به مدل اجازه می‌دهد تا بر کانال‌های کلیدی تمرکز کرده و دقت تشخیص را بهینه کند.

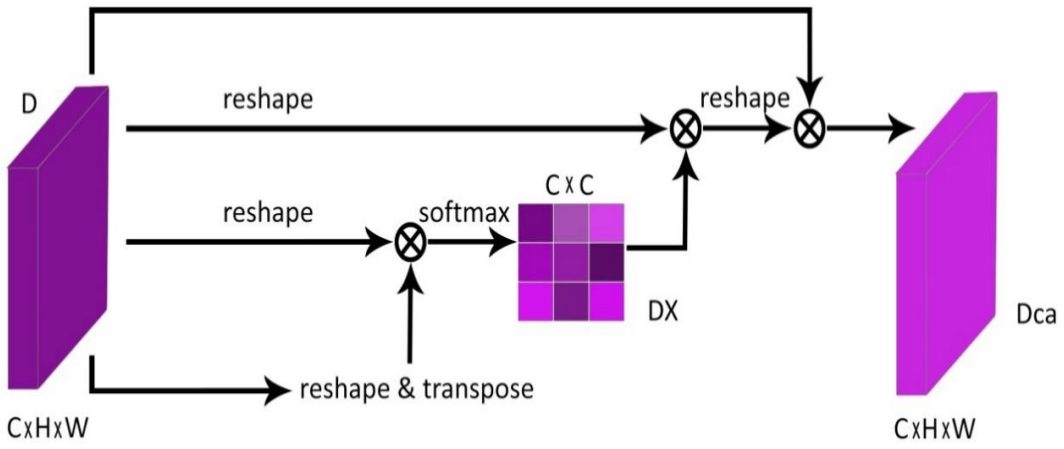
$$Dx_{ji} = softmax(\frac{D_i^T D_j}{\sqrt{C}})$$ رابطه(۵)

در رابطه (۵)، Dx_{ji} تأثیر کانال j ام را بر کانال i ام اندازه‌گیری می‌کند، به طوری که مقادیر بالاتر نشان‌دهنده اتصال قوی‌تر بین آن‌ها هستند. سپس، D تغییر شکل یافته با $C\times N$ ضرب می‌شود با Dx_{ji}، که خروجی نهایی را به‌دست می‌دهد. برای بهینه‌سازی خروجی، یک ضریب مقیاس‌گذاری (δ) به نتیجه اضافه می‌شود. این ضریب، به تنظیم تأثیر مکانیزم توجه کمک می‌کند. همان‌طور که در رابطه (۶) نشان داده شده است، این فرایند به تولید خروجی نهایی کمک می‌کند که اهمیت کانال‌های مرتبط را تقویت می‌کند

$$Dca_j = \delta \sum_{i=1}^{C}(Dc_{ji}Di) + D_j$$ رابطه(۶)

ضریب δ در ابتدا مقدار صفر دارد و به‌تدریج در طول فرایند آموزش تعیین می‌شود. این تنظیم تدریجی به مدل اجازه می‌دهد تا تأثیر مکانیزم توجه را به‌مرور یاد بگیرد و اعمال کند. ویژگی نهایی هر کانال، ترکیبی وزنی از تمامی کانال‌ها و ویژگی اولیه است. این ترکیب به طور مؤثر اطلاعات مربوط به کانال‌های کلیدی را تقویت می‌کند و همان‌طور که در معادله پیشین توضیح داده شد، ساختاری منسجم و دقیق برای نمایش ویژگی‌ها ارائه می‌دهد.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]



شکل۶: ساختار کلی بلوک توجه کانال

۳-۳- مقایسه با روش‌های پیشرفته موجود

این مقاله رویکردی نوین برای شناسایی تغییرات ارائه می‌دهد که از ساختار شبکه دوشاخه‌ای ترکیب‌شده با مبدل‌ها بهره می‌برد. این روش با استفاده از بلوک‌های توجه کانالی و توجه مکانی، به طور مؤثر ویژگی‌های تمایزدهنده را از تصاویر استخراج می‌کند و دقت شناسایی تغییرات را بهبود می‌بخشد. برای ارزیابی اثربخشی این رویکرد، عملکرد شبکه پیشنهادی با سایر روش‌هایی که مبتنی بر مکانیزم‌های توجه یا معماری مبدل هستند، مقایسه شده است. نتایج این مقایسه، توانایی برتر مدل را در شناسایی تغییرات نشان می‌دهد. به طور خاص، عملکرد شبکه پیشنهادی با روش‌هایی که از مکانیزم‌های توجه استفاده می‌کنند مقایسه شده است. یکی از این روش‌ها *DDCNN* است که توسط پنگ و همکاران (۲۰۲۰) معرفی شده و از مکانیزم توجه متراکم در مراحل کدگذاری و کدگشایی برای پردازش تصاویر دوگانه‌زمانی بهره می‌برد. همچنین، این روش یک واحد تفاضلی بین لایه‌های ورودی را برای بهبود کارایی شبکه در نظر می‌گیرد [۱]. همچنین، عملکرد شبکه پیشنهادی با روش *AGCDetNet* مقایسه شده است. این روش از توجه مکانی به همراه ماژولی که شامل فیلترگذاری مداخله هدایت‌شده باتوجه به کانال است، بهره می‌برد. این رویکرد ویژگی‌ها را در سطوح و زمینه‌های مختلف و در مقیاس‌های متفاوت بهبود می‌بخشد و در نتیجه توانایی‌های شناسایی تغییرات هدایت‌شده با توجه را ارتقا می‌دهد [۳۱].

در روش‌های مبتنی بر مبدل‌ها، باندارا و پاتل (۲۰۲۱) شبکه *ChangeFormer* را معرفی کردند که از یک معماری شبکه سیامی مبتنی بر مبدل استفاده می‌کند که ترکیبی از کدگذار مبدل و کدگشا *MLP* را به کار می‌برد [۲۱]. هم چنین شبکه *TinyViT* چالش‌های مرتبط با مبدل‌های حجیم بینایی[1] *(ViTs)* را با معرفی مجموعه‌ای از مدل‌های فشرده و کارآمد حل می‌کند. این مدل‌های کوچک از دانش مدل‌های بزرگ پیش‌آموزش‌دیده مانند مدل‌های آموزش‌دیده بر روی مجموعه داده *ImageNet* از طریق فرایند فشرده‌سازی مدل بهره می‌برند و به آن‌ها این امکان را می‌دهد که عملکرد خوبی داشته باشند در حالی که اندازه‌ای مناسب برای کاربردهای محدود منابع دارند [۳۲]. چن و همکاران (۲۰۲۲) از شبکه *BIT* برای مدل‌سازی مؤثر زمینه‌های مکانی-زمانی استفاده کردند و تصاویر دوگانه‌زمانی را به‌عنوان توکن‌های معنایی نشان می‌دهد، به طوری که کدگذار مبدل زمینه را در فضای

[1] *Vision Transformers*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

زمان-توکن فشرده شده ضبط می‌کند [۳۳]. در روش دیگر، محمدی و قادری (۲۰۲۳) شبکه *SiamixFormer* را معرفی کردند که یک مدل سیامی نوآورانه برای شناسایی تغییرات است که از یک تنظیم‌سازی دوکدگذاری‌کننده در یک ساختار مبدل سلسله‌مراتبی در مرحله کدگذاری استفاده می‌کند.گنجاندن مبدل‌های زمانی در هم‌جواری ویژگی‌ها به حفظ میدان‌های دید وسیع ایجادشده توسط کدگذارهای مبدل کمک می‌کند. خروجی مبدل زمانی سپس از طریق یک کدگشا با روش *MLP* ساده در هر مرحله عبور می‌کند [۳۴].

۴- آزمایش‌ها و نتایج

این بخش نتایج روش پیشنهادی را ارائه می‌دهد و آن را با روش‌های پیشرفته موجود مقایسه می‌کند. در این بخش، جزئیات پیاده‌سازی فنی، ارزیابی دقت و نتایج آزمایش‌ها به طور کامل بررسی شده است.

۴-۱- تنظیمات اجرایی

بسته برنامه‌نویسی *PyTorch* بر روی یک *GPU Nvidia RTX 3070* با ۸ گیگابایت حافظه در تمام آزمایش‌ها برای اجرای استراتژی پیشنهادی استفاده شد. در طول آموزش، از بهینه‌ساز *Adam* با نرخ یادگیری که به طور تدریجی از ۰/۱ تا ۰/۰۰۰۱ کاهش می‌یابد، استفاده کردیم. اندازه دسته ورودی (بچ)[1]را برای واردکردن داده‌های آموزشی به شبکه، برابر با ۸ تنظیم کردیم. در این مطالعه، به دلیل عدم تعادل بین داده‌های تغییر کرده و تغییر نکرده در مجموعه‌داده شناسایی تغییرات ساختمان، از تابع زیان آنتروپی باینری متقابل وزن‌دار *(WBCE)*[2] استفاده شد. در جدول (۳) به طور خلاصه هایپرپارامترهایی که پس از بهینه‌سازی به دست آمده-اند، آورده شده است.

در دو مجموعه داده موجود، ۷۲٪ آن برای آموزش در نظر گرفته شده است، ۱۸٪ برای تست و ۱۰٪ باقی‌مانده برای ارزیابی مدل. باتوجه به محدودیت‌های کارت گرافیک و بر اساس مدل نمایش داده شده در شکل (۳)، اندازه ورودی را ۱۲۸ × ۱۲۸ پیکسل انتخاب کردیم. همان‌طور که قبلاً ذکر شد، به دلیل کمبود داده‌های آموزشی در برخی کلاس‌ها، از تقویت موثرتر داده‌ها استفاده شده است که در جدول (۲) آورده شده است. علاوه بر این، با افزایش تعداد بخش‌های داده با هم‌پوشانی، اندازه مجموعه‌داده را افزایش دادیم: ۶۴ پیکسل برای مجموعه‌داده هوایی و ۳۲ پیکسل برای مجموعه‌داده ماهواره‌ای. این تقویت مؤثرتر داده منجر به تقسیم‌بندی‌های زیر برای آموزش، تست و ارزیابی شد که به‌صورت تصادفی از بین کل داده‌ها انتخاب شده‌اند.

- مجموعه‌داده ماهواره‌ای: ۵۱۷۶ نمونه برای آموزش، ۱۲۹۶ نمونه برای تست و ۷۲۰ نمونه برای ارزیابی
- مجموعه‌داده هوایی: ۴۴۱۹ نمونه برای آموزش، ۱۱۰۵ نمونه برای تست و ۶۱۴ نمونه ارزیابی

[1] *Batch*

[2] *weighted binary cross entropy*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

جدول۳: مقادیر بهینه هایپرپارامترهای مدل پیشنهادی

دسته‌بندی	هایپرپارامتر	مقدار	توضیحات
پارامترهای آموزش	نرخ یادگیری اولیه	۰/۱	با استفاده از کاهش پلکانی تا ۰,۰۰۰۱ در طول آموزش تنظیم شد
	اندازه بچ	۸	بهینه‌شده باتوجه‌به محدودیت‌های حافظه *GPU (NVIDIA RTX 3070)*
	تعداد اپک	۱۰۰	باتوجه‌به نمودار همگرایی انتخاب شد
	بهینه‌ساز	*Adam*	پارامترهای پیش‌فرض(β_1=0.9, β_2=0.999)
معماری شبکه	اندازه ورودی	۱۲۸ × ۱۲۸ پیکسل	متناسب با محدودیت‌های محاسباتی انتخاب شد
	تعداد لایه‌های کدگذار	۴	هر لایه شامل *CNN* + بلوک توجه مکانی - کانالی
	تعداد فیلترهای اولیه	۶۴	در هر لایه کدگذار دوبرابر می‌شود
	اندازه کرنل‌های *CNN*	۳×۳	برای تمام لایه‌های کانولوشنی
	ضریب مقیاس γ و δ	۰ تا ۱	به‌صورت تدریجی در طول آموزش افزایش می‌یابد
پیش‌پردازش	اندازه همپوشانی پچ‌ها	۶۴ پیکسل (هوایی) ۳۲ پیکسل (ماهواره‌ای)	برای افزایش حجم داده‌های آموزشی
	روش‌های تقویت داده	۱۵ روش	مطابق جدول(۲)
تابع زیان	نوع تابع	*WBCE*	وزن‌دهی خودکار بر اساس فراوانی کلاس‌ها

۴-۲- معیارهای ارزیابی

جنبه‌های مختلف اثربخشی روش پیشنهادی با استفاده از چهار معیار ارزیابی زیر مورد ارزیابی قرار می‌گیرد:

معیار دقت، به نسبت نمونه‌های مثبت به کل پیش‌بینی‌های مثبت اشاره دارد (رابطه (۷)). به‌عبارت‌دیگر، دقت نشان می‌دهد که چقدر پیش‌بینی‌های مدل برای کلاس مثبت درست بوده است.

$$Precision = \frac{TP}{TP+FP}$$ رابطه(۷)

معیار بازخوانی[1] که به آن حساسیت نیز گفته می‌شود، نشان‌دهنده توانایی مدل در شناسایی تمامی نمونه‌های مثبت موجود در داده‌ها است. به‌عبارت‌دیگر، بازخوانی محاسبه می‌کند که مدل از کل نمونه‌های مثبت واقعی چه درصدی را به‌درستی شناسایی کرده است. مقادیر زیاد این معیار نشان‌دهنده نرخ منفی کاذب پایین است (رابطه (۸)).

$$Recall = \frac{TP}{TP+FN}$$ رابطه(۸)

[1] *Recall*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

معیار امتیاز اف-۱[1] نشان‌دهنده میانگین متعادل بین دقت و معیار بازخوانی است که به‌عنوان میانگین هارمونیک محاسبه می‌شود (رابطه (۹)).

$$F1-score=\frac{2\times Precision\times Recall}{Precision+Recall}$$ رابطه(۹)

ضریب کاپا[2] نشان‌دهنده میزان توافق واقعی مدل با برچسب‌های واقعی است و به‌ویژه در شرایطی که تعداد کلاس‌ها نامتوازن است، بسیار مفید است. این معیار نشان می‌دهد که چقدر توافق واقعی بیشتر از توافق تصادفی است (رابطه (۱۰)).

رابطه(۱۰)

$$Kappa\ \text{Coefficient (KC)}=\frac{2\times(TP\times TN-FN\times FP)}{(TP+FP)\times(FP+TN)+(TP+FN)\times(FN+TN)}$$

در روابط (۷) تا (۱۰)، مقدار مثبت واقعی *(TP)* به پیکسل‌های تغییر یافته‌ای اشاره دارد که به‌درستی شناسایی شده‌اند. مثبت کاذب *(FP)* به پیکسل‌های تغییر نیافته‌ای اشاره دارد که به‌اشتباه به‌عنوان پیکسل‌های تغییریافته شناسایی شده‌اند. منفی واقعی *(TN)*به مواردی اشاره دارد که مدل به‌درستی پیکسل‌های تغییر نیافته را شناسایی کرده است. منفی کاذب *(FN)* به پیکسل‌های تغییر یافته‌ای اشاره دارد که به‌اشتباه به‌عنوان پیکسل‌های تغییر نیافته طبقه‌بندی شده‌اند. در زمینه شناسایی تغییرات، کاهش تعداد *FP* باعث افزایش دقت می‌شود، زیرا به معنی شناسایی کمتر پیکسل‌های تغییر نیافته به‌عنوان تغییر یافته است. همچنین، کاهش تعداد *FN* موجب افزایش بازیابی می‌شود، زیرا پیکسل‌های تغییر یافته به‌درستی شناسایی می‌شوند. معیار اف-۱ به‌عنوان یک معیار جامع برای ارزیابی عملکرد مدل عمل می‌کند که میانگین هارمونیک دقت و بازیابی است و مقادیر بالاتر آن نشان‌دهنده نتایج بهتری است.

۴-۳- نتایج آزمایش‌ها

چندین آزمایش در این مقاله طراحی و اجرا شده تا تأثیر اجزای مختلف معماری پیشنهادی را در عملکرد مدل ارزیابی شود. در معماری دوشاخه‌ای ارائه‌شده در شکل (۳)، ابتدا یک مدل پایه بر اساس شبکه *CNN* با هسته‌های ۳×۳ پیاده‌سازی شد. سپس، تأثیر استفادهٔ مستقل از بلوک‌های توجه مکانی و توجه کانالی در درون کدگذار مورد بررسی قرار گرفت. در نهایت، روش پیشنهادی با استفادهٔ هم‌زمان از هر دو بلوک *CAB* و *SAB* در هر کدگذار اجرا شد تا توانایی این ترکیب در استخراج ویژگی‌های متمایز و مؤثر ارزیابی گردد.

جدول(۴) نتایج حاصل از این آزمایش‌ها را بر روی دو مجموعه‌داده مستقل (ماهواره‌ای و هوایی) نمایش می‌دهد. همان‌گونه که مشاهده می‌شود، استفاده از بلوک *SAB* منجر به افزایش بازخوانی از ۹۴٫۵۶٪ (مدل پایه) به ۹۵٫۸۳٪ گردید. این بهبود ۱٫۲۷٪نشان‌دهنده‌ی نقش مؤثر *SAB* در تقویت توانایی مدل برای شناسایی نواحی دارای تغییر است، چرا که این بلوک با مدل‌سازی روابط مکانی، کمک می‌کند تا تغییرات محلی، جزئی یا غیرخطی بهتر شناسایی شوند. از سوی دیگر، استفاده از بلوک *CAB* باعث افزایش امتیاز اف-۱ از ۹۵٫۲۸٪به ۹۶٫۷۶٪شد (بهبود ۱٫۴۸٪)، که این امر گویای اثربخشی آن در کاهش خطاهای مثبت و منفی کاذب است؛ به‌ویژه از طریق تمرکز بر ویژگی‌های مهم در سطح کانال.

ترکیب دو بلوک *SAB* و *CAB* در مدل نهایی پیشنهادی، بیشترین بهبود را به همراه داشت: دقت کلی به ۹۸٫۶۳٪، بازخوانی به ۹۸٫۵۰٪ و امتیاز اف-۱ به ۹۸٫۵۷٪ افزایش یافت. این پیشرفت به‌ترتیب برابر با ۲٫۷٪، ۴٫۲٪ و ۳٫۵٪ نسبت به مدل پایه است که از نظر آماری و کاربردی، قابل‌توجه محسوب می‌شوند. به‌ویژه افزایش هم‌زمان بازخوانی و اف-۱، نشان‌دهنده‌ی توانمندی شبکه در تشخیص تغییرات با حساسیت بالا و تعادل مناسب بین دقت و پوشش است. همچنین، افزایش ۵٫۶٪ در ضریب کاپا بیانگر هم‌خوانی بیشتر

[1] *F1-Score*
[2] *Kappa Coefficient*

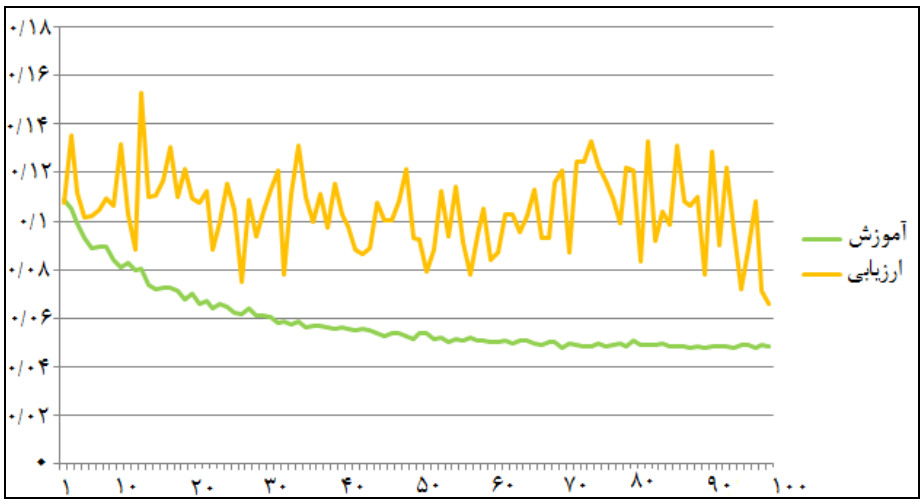
[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

پیش‌بینی‌ها با واقعیت زمینی و کاهش تأثیر شانس در نتایج مدل است.

در مجموع، این تحلیل‌ها مؤید آن است که طراحی پیشنهادی، با بهره‌گیری از دو مکانیزم توجه مکمل، موجب بهبود معنادار در استخراج ویژگی‌های مکانی و کانالی شده و عملکرد مدل را در شناسایی دقیق و چندکلاسه‌ی تغییرات به‌شکل مؤثری ارتقا داده شده است. در نهایت، با ادغام بلوک‌های توجه مکانی و توجه کانالی در معماری شبکه، تعداد کل پارامترهای مدل به ۱۸،۵۷۰،۱۶۸ (۵۷/۱۸ میلیون) پارامتر رسید. باتوجه‌به پیکربندی سخت‌افزاری شامل کارت گرافیک *NVIDIA RTX 3070 GDDR6* با ظرفیت ۸ گیگابایت و اندازه بچ ۸، زمان متوسط آموزش برای هر اپک ۱۰/۱ دقیقه محاسبه شد. همان‌گونه که در شکل(۷) دیده می‌شود، نمودار همگرایی آموزش شبکه برای مجموعه‌داده هوایی بیانگر آن است که استفاده از روش پیشنهادی موجب تسریع در فرآیند آموزش و دستیابی سریع‌تر به همگرایی شده است.

جدول۴: بررسی نتایج روش پیشنهادی بر روی دو مجموعه‌داده مختلف

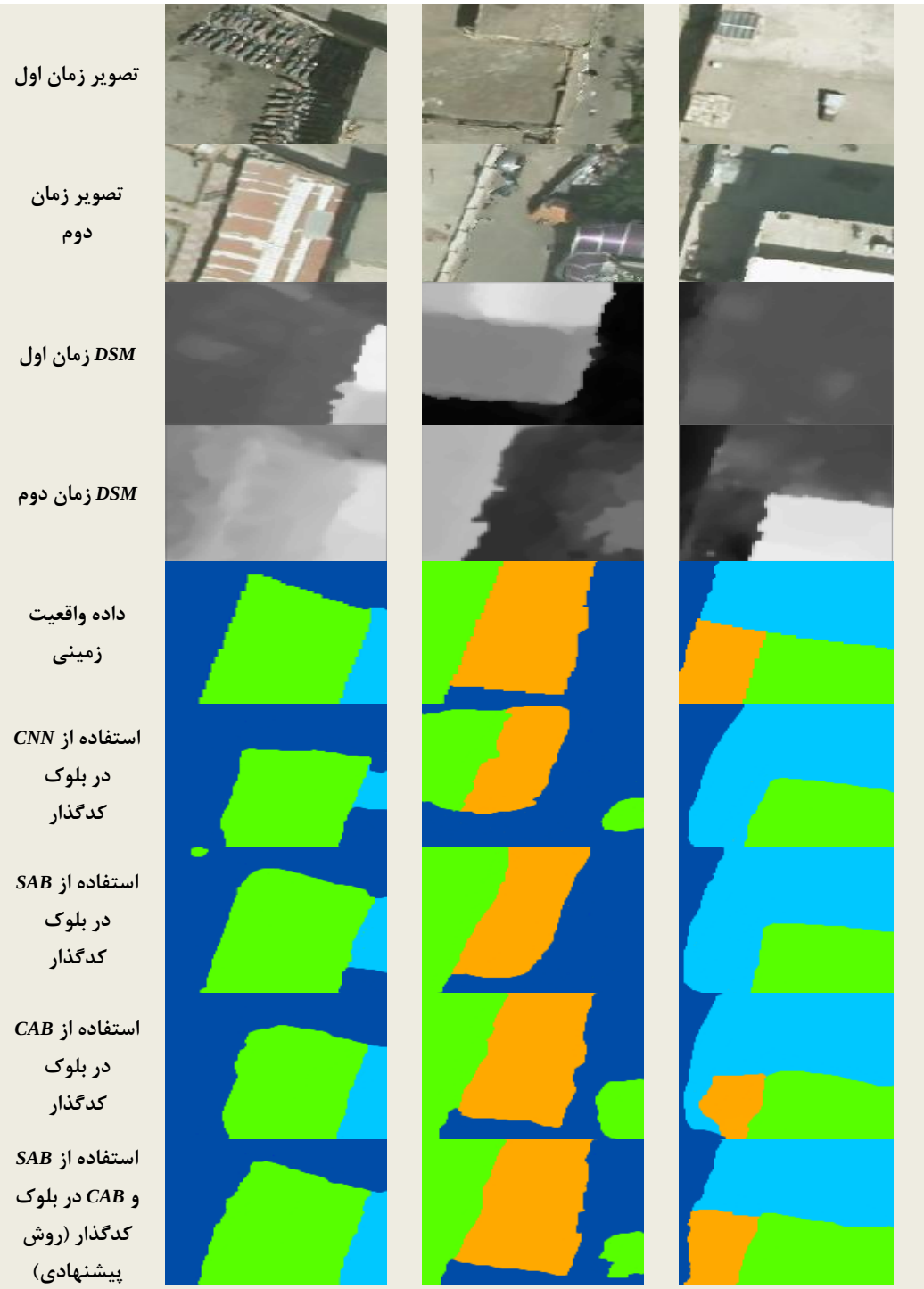
روش	دقت (درصد)	بازخوانی (درصد)	اف-۱ (درصد)	ضریب کاپا
استفاده از *CNN* ۳×۳ به‌جای بلوک کدگذار	۹۶/۰۲	۹۴/۵۶	۹۵/۲۸	۰/۸۸
SAB در هر بلوک کدگذار	۹۷/۲۱	۹۵/۸۳	۹۶/۵۲	۰/۹۰
CAB در هر بلوک کدگذار	۹۷/۵۲	۹۶/۰۲	۹۶/۷۶	۰/۹۱
CAB + SAB در هر بلوک کدگذار (روش پیشنهادی) - داده هوایی	۹۸/۶۳	۹۸/۵۰	۹۸/۵۷	۰/۹۳
CAB + SAB در هر بلوک کدگذار (روش پیشنهادی) –داده ماهواره‌ای	۹۸/۱۲	۹۸/۲۶	۹۸/۱۹	۰/۹۴



شکل۷: نمودار تغییرات تابع زیان آموزش و ارزیابی در حین آموزش شبکه برای داده هوایی

شکل(۸) نتایج بصری حاصل از اعمال تدریجی مدل پیشنهادی بر روی مجموعه‌داده هوایی را نشان می‌دهد. در خطوط ۶ تا ۹ شکل(۸)، نتایج ناشی از رویکرد اجرایی پیوسته، به‌ویژه در مناطقی که تغییرات با اندازه‌های مختلف مشاهده می‌شود، ارائه شده است. این نتایج محدودیت‌های استفاده از یک شبکه دوشاخه‌ای مدل‌سازی‌شده با *CNN* را نشان می‌دهند، زیرا این مدل قادر به شبیه‌سازی جامع تغییرات نیست.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]



شکل۸: نتایج بصری حاصل از اعمال تدریجی روش پیشنهادی بر روی داده هوایی . رنگ آبی نشان‌دهنده نواحی غیرطبقه‌بندی‌شده است، رنگ فیروزه‌ای نشان‌دهنده عدم تغییر ساختمان، رنگ سبز نشان‌دهنده ساختمان‌های جدید و رنگ نارنجی نشان‌دهنده ساختمان‌های تخریب‌شده است.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

با این حال، استفاده از *SAB*، جزئیات مکانی، به‌ویژه در لبه‌ها را بهبود می‌بخشد و شناسایی تغییرات دوبعدی را ارتقا می‌دهد. از طرف دیگر، استفاده از *CAB* باعث بهینه‌سازی استخراج جزئیات خاص باند می‌شود. باتوجه به اینکه *DSM* یکی از باندهای ورودی است، این امر به استخراج دقیق‌تر جزئیات ارتفاع و تغییرات سه‌بعدی کمک می‌کند. در نتیجه، ترکیب *SAB* و*CAB* باعث بهبود نتایج در شناسایی تغییرات دوبعدی و سه-بعدی می‌شود. در ستون دوم شکل(۸)، نمونه‌ای از یک پیچیدگی مشابه به ساختمان در تصویر دوم، همراه با تغییرات مثبت ارتفاع در *DSM*، ارائه شده است. این تغییرات به دلیل حساسیت مدل به تغییرات ارتفاع، به طور دقیق توسط مدل پیش‌بینی شده‌اند.

۴-۴- نتایج مقایسه با روش‌های پیشرفته

جدول (۵) مقایسه‌ای جامع از عملکرد روش پیشنهادی در مقایسه با روش‌های مختلف در دو مجموعه‌داده ماهواره‌ای و هوایی ارائه می‌دهد. نتایج مقایسه پیاده‌سازی مدل پیشنهادی با روش‌های پیشرفته دیگر بر روی دو مجموعه‌داده نشان‌دهنده عملکرد برتر این مدل است. روش پیشنهادی در این مجموعه داده‌ها، از نظر امتیاز اف-۱ عملکرد بهتری را نسبت به سایر روش‌ها ارائه می‌دهد و ضریب کاپا نیز در مقایسه با دیگر تکنیک‌ها برابر یا بالاتر است.

در مقایسه با سایر روش‌ها، *ChangeFormer* به‌عنوان نزدیک‌ترین رقیب، در هر دو مجموعه داده نتایج قابل‌توجهی ارائه می‌دهد، اما در مقایسه با روش پیشنهادی، امتیاز اف-۱ و ضریب کاپا کمی پایین‌تر هستند. *DDCNN* و *AGCDetNet* عملکرد خوبی دارند، اما از نظر هر دو معیار نسبت به روش پیشنهادی پایین‌تر هستند. *Tiny ViT* و *BIT* در مقایسه با روش پیشنهادی، امتیازات کمتری در هر دو مجموعه‌داده به دست آورده‌اند. مدل *SiamixFormer* نیز در این جدول، امتیازات کمتری نسبت به روش پیشنهادی نشان می‌دهد. در داده هوایی، روش پیشنهادی با امتیاز اف-۱ تقریباً ۱٪ بالاتر از روش*ChangeFormer*، نتایجی رقابتی به دست آورده است. در دیتای ماهواره‌ای، افزایش ۰٫۳ در ضریب کاپا برتری این روش را نشان می‌دهد. روش پیشنهادی از ترکیب *SAB* و *CAB*در هر لایه کدگذار استفاده می‌کند که این ترکیب، ویژگی‌های متمایز و جامعی را از تصاویر ورودی و *DSM* ها در هر دو سمت شبکه استخراج می‌کند. امتیاز بالای اف-۱، بیانگر عملکرد متعادل مدل است که به طور هم‌زمان دقت و بازخوانی بالایی را به دست می‌آورد. این بهبود به دلیل استفاده استراتژیک از مکانیزم‌های توجه مکانی و کانالی است که استخراج ویژگی‌های برد بلند را تقویت می‌کند و در نتیجه، دقت کلی مدل افزایش می‌یابد. علاوه بر این، ضریب کاپای بالا نشان‌دهنده توانایی مدل در شناسایی دقیق تغییرات بین کلاس‌ها است و اثربخشی مدل پیشنهادی را به‌خوبی نشان می‌دهد.

شکل‌های(۹) و (۱۰) نتایج بصری برای هر دو مجموعه‌داده را با روش‌های پیشرفته مقایسه کرده و به طور مؤثری عملکرد روش پیشنهادی را نشان می‌دهند. شکل (۹) نمونه‌ای از مجموعه‌داده‌های هوایی را نمایش می‌دهد. می‌توان مشاهده کرد که برخی مدل‌ها مانند *SiamixFormer* و *Tiny ViT*در شناسایی کلاس "تخریب‌شده" با مشکل مواجه هستند. علاوه بر این، مدل‌هایی مانند *AGCDetNet* و *BIT* در تشخیص صحیح محدوده تغییرات مثبت ساخت‌و ساز و عدم تغییر خطاهایی دارند. در این نمونه، تغییرات مثبت ساخت و سازی در *DSM* مربوط به زمان دوم به‌وضوح مشخص هستند. بنابراین، مدل باید بتواند ویژگی‌های مناسب را از این تصاویر استخراج کند. روش پیشنهادی ما نتایجی مشابه با مدل‌های *DDCNN* و *ChangeFormer* به دست آورده است. با این حال، مدل *DDCNN* منطقه‌ای که با دایره قرمز مشخص شده را به اشتباه به‌عنوان "غیرطبقه‌بندی‌شده" طبقه‌بندی کرده و فقط به تصاویر متکی بوده است و *DSM* مربوط به زمان دوم را نادیده گرفته است. این خطا احتمالاً به دلیل وجود سایه ساختمان در سمت

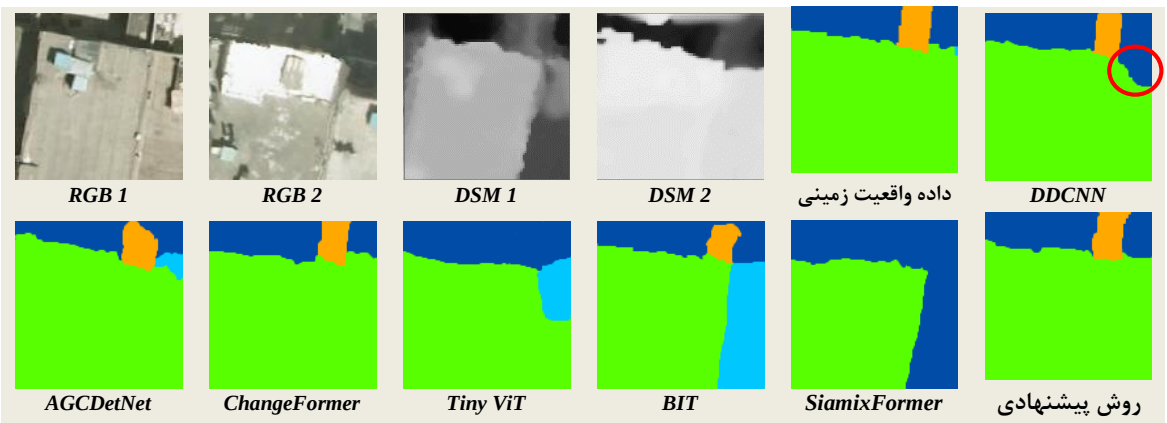
[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

راست تصویر زمان دوم ایجاد شده است که منطقه‌ای تاریک به وجود آورده و مدل را دچار اشتباه کرده است. بنابراین، مدل‌هایی که ویژگی‌های مرتبط را به طور مؤثر از هر دو تصویر و *DSM* استخراج می‌کنند، عملکرد موفق‌تری خواهند داشت.

شکل(۱۰) نمونه‌ای از مجموعه داده‌های ماهواره‌ای را نمایش می‌دهد. در اینجا، بخشی از ساختمان‌ها در تصویر دوم تخریب شده‌اند، درحالی‌که سایر ساختمان‌ها بدون تغییر باقی مانده‌اند. برخی روش‌ها مانند *TinyViT* و *BIT* حساسیت کمی نسبت به تغییرات ساختاری منفی (تخریب‌ها) نشان داده و در شناسایی این تغییرات، به‌ویژه برای ساختمان‌هایی با رنگ‌های مشابه با زمین، دقت پایینی دارند. مدل‌های *DDCNN*، *AGCDetNet* و *SiamixFormer* حساسیت کمتری نسبت به روش‌هایی که از داده‌های *DSM* استفاده می‌کنند دارند و بیشتر به استخراج ویژگی‌ها از تصاویر *RGB* متکی هستند که منجر به تشخیص ناقص تغییرات می‌شود. روش پیشنهادی ما در مقایسه با *ChangeFormer* عملکردی نزدیک به آن دارد و هر دو نتایجی بسیار مشابه با واقعیت زمینی به دست می‌آورند. با این حال، روش پیشنهادی در استخراج لبه‌ها و جزئیات با دقت بیشتری عمل می‌کند

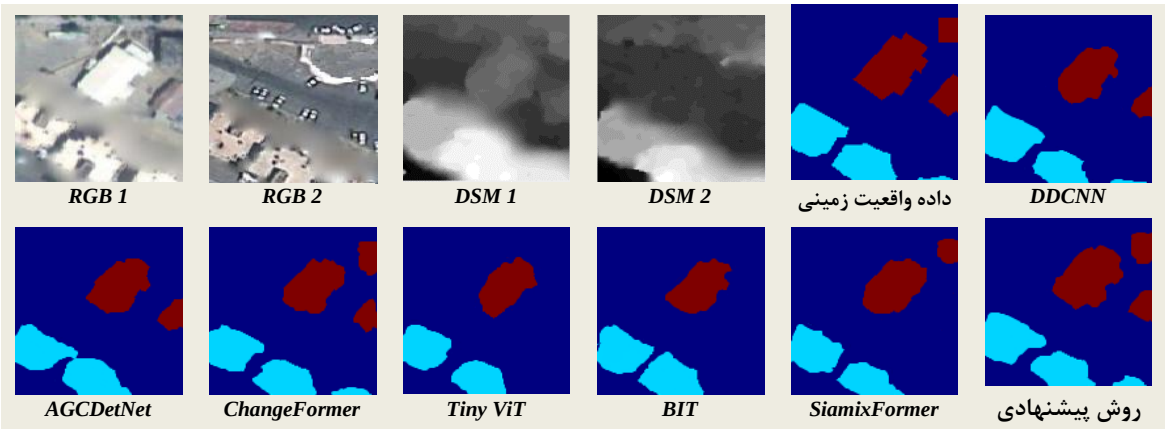
جدول ۵: ارزیابی تطبیقی روش پیشنهادی برای تشخیص تغییرات چندگانه در ساختمان‌ها با استفاده از داده‌های هوایی و ماهواره‌ای

روش‌ها	هوایی		ماهواره‌ای	
	کاپا	اف-۱ (درصد)	کاپا	اف-۱ (درصد)
DDCNN	۰٫۹۰	۹۷٫۴۴	۰٫۸۹	۹۵٫۹۱
AGCDetNet	۰٫۸۹	۹۶٫۰۷	۰٫۹۰	۹۶٫۹۷
ChangeFormer	۰٫۹۱	۹۷٫۷۲	۰٫۹۱	۹۷٫۸۹
Tiny ViT	۰٫۸۶	۹۳٫۸۱	۰٫۸۷	۹۴٫۳۵
BIT	۰٫۸۹	۹۶٫۹۳	۰٫۸۸	۹۵٫۶۶
SiamixFormer	۰٫۸۷	۹۵٫۸۷	۰٫۸۸	۹۵٫۱۶
روش پیشنهادی	۰٫۹۳	۹۸٫۵۷	۰٫۹۴	۹۸٫۱۹



شکل۹: ارزیابی بصری مقایسه روش‌های پیشرفته با روش پیشنهادی در مجموعه‌داده‌های هوایی. رنگ آبی نشان‌دهنده نواحی غیرطبقه‌بندی‌شده است، رنگ فیروزه‌ای نشان‌دهنده عدم تغییر ساختمان، رنگ سبز نشان‌دهنده ساختمان‌های جدید و رنگ نارنجی نشان‌دهنده ساختمان‌های تخریب‌شده است.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]



شکل۱۰: ارزیابی بصری مقایسه روش‌های پیشرفته با روش پیشنهادی در مجموعه‌داده‌های ماهواره‌ای. رنگ آبی نشان‌دهنده مناطق " غیرطبقه‌بندی‌شده "، رنگ فیروزه‌ای نشان‌دهنده "عدم تغییر ساختمان"، رنگ قرمز نشان‌دهنده "ساختمان‌های تخریب‌شده" و رنگ زرد نشان‌دهنده "ساختمان‌های جدید" است.

۵- بحث و بررسی نتایج

یکی از نقاط قوت برجسته روش پیشنهادی این تحقیق، استفاده از روش‌های تقویت مؤثرتر داده‌ها و بخش‌های همپوشان است که برای مقابله با چالش کمبود داده‌های آموزشی در برخی کلاس‌ها به کار گرفته شده است. این استراتژی، همان‌طور که در بخش جزئیات پیاده‌سازی نیز توضیح داده شد، با هدف بهبود توزیع داده‌ها و تقویت مدل در شناسایی دقیق کلاس‌هایی با داده‌های آموزشی محدود طراحی شده است. جدول (۶) تأثیر این روش را بر روی مجموعه‌داده ماهواره‌ای نشان می‌دهد و بهبود معیارهای ارزیابی مدل را برجسته می‌کند.

در مقایسه با حالتی که تقویت داده و بخش‌های غیر همپوشان استفاده نشده است، استفاده از تقویت داده‌های مؤثرتر و بخش‌های همپوشان باعث افزایش چشمگیر تمامی معیارها شده است. در حالت بدون تقویت داده، ضریب کاپا برابر ۰٫۸۱است، درحالی‌که با استفاده از تقویت داده‌های مؤثرتر و پچ‌های همپوشان، این مقدار به ۰٫۹۴ افزایش یافته است. به طور مشابه، امتیاز اف-۱ از ۹۵٫۰۶ ٪ به ۹۸٫۱۹ ٪، دقت بازخوانی از ۹۵٫۸۹ ٪ به ۹۸٫۲۶ ٪ و دقت از ۹۴٫۲۵ ٪ به ۹۸٫۱۲ ٪ بهبود یافته است. این نتایج نشان‌دهنده تأثیر مثبت و قابل‌توجه این تکنیک‌ها بر عملکرد مدل است.

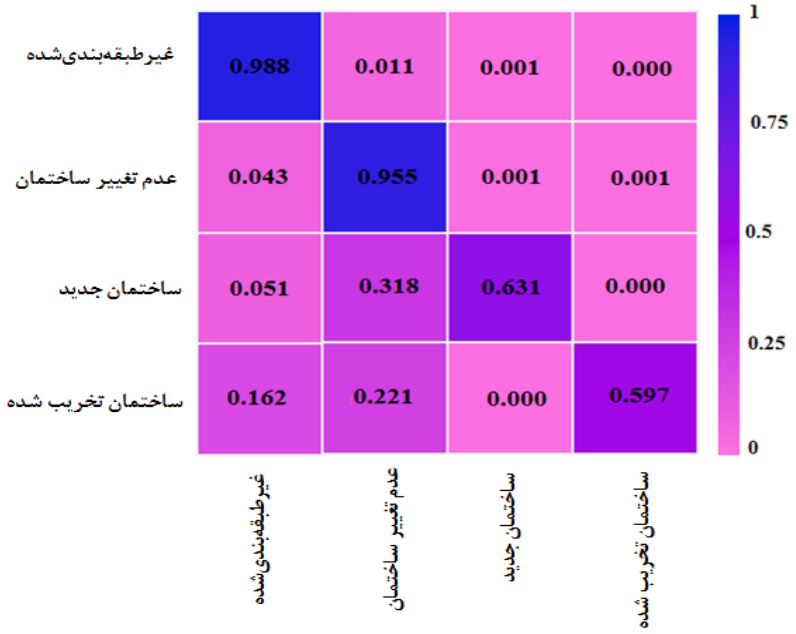
جدول ۶: تأثیر تکنیک تقویت مؤثرتر داده‌ها و بخش‌های همپوشان بر عملکرد مدل در مجموعه داده ماهواره‌ای

روش‌ها	دقت (درصد)	بازیابی (درصد)	اف-۱ (درصد)	کاپا
بدون تقویت داده و بخش‌های غیر همپوشان	۹۴٫۲۵	۹۵٫۸۹	۹۵٫۰۶	۰٫۸۱
با تقویت مؤثرتر داده‌ها و بخش‌های همپوشان	۹۸٫۱۲	۹۸٫۲۶	۹۸٫۱۹	۰٫۹۴

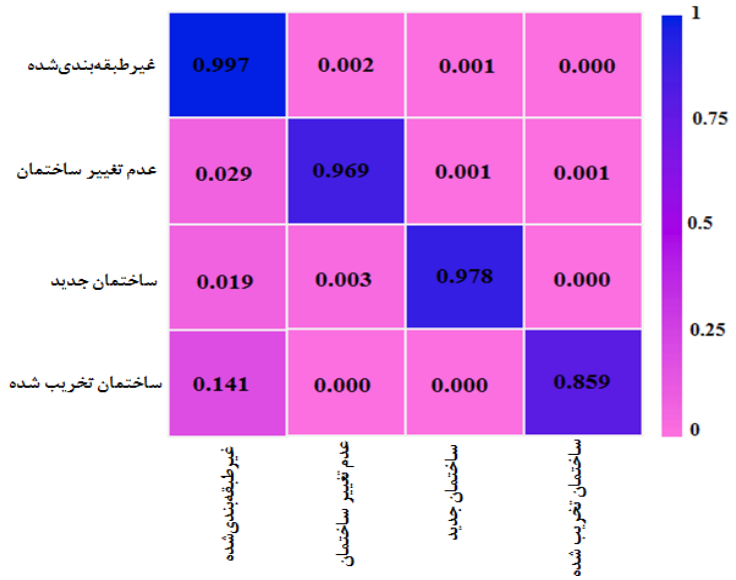
علاوه بر این، شکل (۱۱) تأثیر این تقویت داده‌ها را بر روی ماتریس ابهام نمایش می‌دهد. این شکل نشان می‌دهد که روش پیشنهادی توانسته عملکرد مدل را به‌ویژه در کلاس‌هایی که داده‌های آموزشی کمی دارند، به طور قابل توجهی بهبود دهد. به‌عنوان مثال، کلاس‌هایی مانند ساختمان‌های جدید و تخریب‌شده که

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

داده‌های آموزشی محدودی دارند، پس از اعمال این تقویت داده‌ها دقت بسیار بالاتری در شناسایی تجربه کرده‌اند. این نتایج نشان می‌دهند که استفاده از تقویت مؤثرتر داده‌ها و بخش‌های همپوشان نه‌تنها توزیع متوازن‌تری از داده‌ها را فراهم می‌کند، بلکه به طور مستقیم باعث افزایش دقت مدل و بهبود تعمیم‌پذیری آن در شناسایی تغییرات میان کلاس‌ها می‌شود. این روش می‌تواند به‌عنوان یک استراتژی مؤثر برای بهبود عملکرد مدل در کاربردهایی با داده‌های آموزشی محدود در نظر گرفته شود.



(الف)



(ب)

شکل۱۱: ماتریس ابهام مجموعه‌داده ماهواره‌ای، (الف) بدون هیچ گونه تقویت و بخش‌های غیر همپوشان، (ب) با تقویت مؤثرتر و بخش‌های همپوشان

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]
[DOI: 10.61882/jgit.13.1.95]

۶- نتیجه‌گیری

در این مطالعه، رویکرد نوآورانه‌ای برای تشخیص تغییرات چند کلاسه ساختمان‌ها به‌صورت سه‌بعدی ارائه شده که از شبکه سیامی دوشاخه‌ای مبتنی بر مبدل و مکانیزم‌های توجه مکانی و کانالی بهره می‌گیرد. این روش به چالش‌های اساسی پایش شهری، مانند عدم تعادل داده‌ها و یکپارچه‌سازی داده‌های متنوع سنجش‌ازدور سه‌بعدی از سکوهای ماهواره‌ای و هوایی پرداخته است. معماری دوشاخه‌ای پیشنهادی با ترکیب مؤثر *CNN* ها و کدگذارهای مبتنی بر مکانیزم توجه، امکان استخراج ویژگی‌های غنی و تفکیک دقیق کلاس‌های مختلف تغییرات ساختمانی، از جمله ساختمان‌های جدید و تخریب‌شده را فراهم کرده است. نتایج تجربی بر روی دو مجموعه‌داده، شامل تصاویر استریو ماهواره‌ای و *DSM* و تصاویر استریو هوایی و *DSM*، نشان داد که روش پیشنهادی با دستیابی به ضریب کاپا ۹۳٪ و ۹۴٪ و ضریب اف-۱ ۹۸٫۵۷٪ و ۹۸٫۱۹٪ در مجموعه داده‌های هوایی و ماهواره‌ای، عملکردی قابل‌قبول در مقایسه با سایر روش‌های پیشرفته مانند *ChangeFormer* (با ضریب کاپا ۹۱٪ و ضریب اف-۱ ۹۷٫۷۲ % و ۹۷٫۸۹ % برای هر دو مجموعه داده) داشته است. اگرچه اختلاف دقت بین روش پیشنهادی و سایر روش‌ها در برخی موارد کم است (به‌عنوان‌مثال، اختلاف ضریب کاپا ۰٫۰۲)، این بهبودها نشان‌دهنده پتانسیل روش پیشنهادی در ارائه نتایج دقیق‌تر و پایدارتر در سناریوهای مختلف شهری است.

یکی از عوامل کلیدی موفقیت، استفاده از استراتژی‌های تقویت مؤثرتر داده‌ها و بخش‌های همپوشان در مرحله پیش‌پردازش بود که تأثیر مشکل رایج عدم تعادل کلاس‌ها را کاهش داده و دقت مدل را بهبود بخشیده است. همچنین، بهره‌گیری از مکانیزم‌های توجه مکانی و کانالی در شبکه مبدل، امکان استخراج جزئیات پیچیده از صحنه‌های شهری و ارائه عملکرد دقیق در تشخیص تغییرات را فراهم کرده است.

یافته‌های این پژوهش، پتانسیل شبکه‌های مبتنی بر مبدل و توجه را در پایش شهری و برنامه‌ریزی استراتژیک نشان می‌دهند و ابزاری مقیاس‌پذیر و قابل انطباق برای مدیریت پایدار شهرها ارائه می‌دهند. این تحقیق نه تنها پیشرفتی در حوزه تشخیص تغییرات سه‌بعدی ساختمان‌ها محسوب می‌شود، بلکه زمینه‌ای برای تحقیقات آینده در یکپارچه‌سازی تکنیک‌های پیشرفته یادگیری عمیق با داده‌های سنجش از دور سه‌بعدی فراهم کرده است. کارهای آینده می‌توانند بر گسترش قابلیت‌های مدل برای مدیریت سناریوهای شهری متنوع‌تر، از جمله یکپارچه‌سازی منابع داده جدید نظیر *SAR* و لیدار و بهبود مکانیزم‌های توجه برای استخراج بهتر ویژگی‌ها از محیط‌های پیچیده شهری متمرکز شوند. علاوه بر این، بررسی تکنیک‌های یادگیری انتقالی می‌تواند انطباق مدل را با شهرها و داده‌های جدید افزایش داده و استفاده از آن را به‌عنوان ابزاری چندمنظوره برای پایش جهانی شهری تسهیل کند.

مراجع

[1] X. Peng, L. Zhang, H. Sun, and L. Ma, "Optical remote sensing image change detection based on attention mechanism and image difference", IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 9, 2020.

[2] L. Li, Y. Bian, H. Zhu, and Z. Shao, "Urban building change detection in SAR images using combined differential image and residual U-Net network", Remote Sensing, vol. 11, no. 9, pp. 1091–1091, 2019.

[3] Z. Zheng, J. Wu, A. Zhen, et al., "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters", Remote Sensing of Environment, vol. 265, 2021.

[4] A. H. Chughtai, H. Abbasi, and I. R. Karas, "A review on change detection method and accuracy assessment for land use land

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

cover", Remote Sensing Applications: Society and Environment, vol. 22, 2021.

[5] J. C. Padró, A. Muñoz, A. Pons, et al., "Comparison of four UAV georeferencing methods for environmental monitoring purposes focusing on the combined use with airborne and satellite remote sensing platforms", International Journal of Applied Earth Observation and Geoinformation, vol. 75, 2019.

[6] S. Pang, X. Yang, L. Zhang, and H. Lin, "Object-based analysis of airborne LiDAR data for building change detection", Remote Sensing, vol. 6, no. 11, pp. 10733–10749, 2014.

[7] R. Qin, J. Tian, and P. Reinartz, "3D change detection – Approaches and applications", 2016.

[8] A. Sasagawa, R. Shibasaki, M. Nakagawa, and A. Ichikawa, "Investigation on automatic change detection using pixel-changes and DSM-changes with ALOS-PRISM triplet images", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 40, no. 7/W2, pp. 213–217, 2013.

[9] J. Xu, X. Zhang, Y. Zhao, et al., "Remote sensing change detection based on multidirectional adaptive feature fusion and perceptual similarity", Remote Sensing, vol. 13, no. 15, pp. 3053–3053, 2021.

[10] D. Peng, Y. Zhang, and H. Guan, "End-to-end change detection for high resolution satellite images using improved UNet++", Remote Sensing, vol. 11, no. 11, 2019.

[11] A. Shafique, I. Shafique, and T. M. Khan, "Deep learning-based change detection in remote sensing images: A review", 2022.

[12] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis", IEEE Access, 2020.

[13] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection", Remote Sensing, 2020.

[14] Y. Zhan, L. Wang, Z. Liu, et al., "Change detection based on deep Siamese convolutional network for optical aerial images", IEEE Geoscience and Remote Sensing Letters, 2017.

[15] J. Xue, X. Liu, C. Ma, et al., "Multi-feature enhanced building change detection based on semantic information guidance", Remote Sensing, vol. 13, no. 20, 2021.

[16] W. Zhang and X. Lu, "The spectral-spatial joint learning for change detection in multispectral imagery", Remote Sensing, 2019.

[17] S. Fang, Z. Li, Y. Wang, and J. Wang, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images", IEEE Geoscience and Remote Sensing Letters, vol. 19, 2021.

[18] Q. Guo, Z. Lin, C. Yang, and Z. Wang, "IDET: Iterative difference-enhanced transformers for high-quality change detection", arXiv preprint, arXiv:2207.09240, 2022.

[19] J. Chen, Y. Chen, X. Liu, and M. Du, "DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, 2021.

[20] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers", IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–14, 2021.

[21] W. G. C. Bandara and V. M. Patel, "A transformer-based Siamese network for change detection", IEEE, 2022.

[22] H. Mohammadi and F. Samadzadegan, "An object-based framework for building change analysis using 2D and 3D information of high-resolution satellite images", Advances in Space Research, vol. 66, no. 6, pp. 1386–1404, 2020.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

[23] I. de Gélis, S. Lefèvre, and T. Corpetti, "Siamese KPConv: 3D multiple change detection from raw point clouds using deep learning", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 197, pp. 274–291, 2023.

[24] T. Ku, M. L. Nguyen, P. Nguyen, et al., "SHREC 2021: 3D point cloud change detection for street scenes", Computer Graphics, vol. 99(C), pp. 192–200, 2021.

[25] V. Marsocci, A. Corsini, and L. Barbato, "Inferring 3D change detection from bitemporal optical images", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 196, 2023.

[26] M. Gomroki, M. Hasanlou, and J. Chanussot, "Automatic 3D multiple building change detection model based on encoder-decoder network using highly unbalanced remote sensing datasets", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023.

[27] G. King and L. Zeng, "Logistic regression in rare events data", Political Analysis, vol. 9, no. 2, pp. 137–163, 2001.

[28] E. D. Cubuk, B. Zoph, D. Mane, et al., "Randaugment: Practical automated data augmentation with a reduced search space", in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops, 2020.

[29] M. H. Guo, J. Lu, C. Dong, et al., "Attention mechanisms in computer vision: A survey", 2022.

[30] P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations", arXiv preprint, arXiv:1803.02155, 2018.

[31] K. Song and J. Jiang, "AGCDetNet: An attention-guided network for building change detection in high-resolution remote sensing images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, 2021.

[32] K. Wu, Z. Xu, Z. Wang, et al., "TinyViT: Fast pretraining distillation for small vision transformers", Springer, 2022.

[33] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers", IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022.

[34] A. Mohammadian and F. Ghaderi, "SiamixFormer: A fully-transformer Siamese network with temporal fusion for accurate building detection and change detection in bi-temporal remote sensing images", International Journal of Remote Sensing, vol. 44, no. 12, 2023.

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.13, No.1, Spring 2025

Research Paper

Multiclass Building Change Detection Using Deep Learning Networks Based on 3D Aerial and Satellite Datasets

Akram Eftekhari [1]*, Farhad Samadzadegan [2], Farzaneh Dadrass Javan [3]

1- Ph.D. Student in Remote Sensing, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran
2- Professor, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran
3- Assistant Professor, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente

Abstract

Detecting three-dimensional changes in buildings plays a vital role in urban monitoring, sustainable development, and disaster management. This research presents an innovative method for detecting multi-class building changes. The proposed approach combines two advanced architectures—a Siamese transformer network and spatial-channel attention mechanisms—to enable the automated, precise detection of the structural changes. The primary innovation lies in the design of a dual-attention encoder block that simultaneously analyzes the local spatial relationships and channel dependencies to identify the changes.

A key challenge in change detection is the class imbalance (e.g., the unchanged buildings, the new constructions, and the demolished structures). To address this problem, methods such as the enhanced data augmentation and the overlapping patch extraction during preprocessing phase been employed. The proposed method was implemented on a stereo dataset from the GeoEye-1 satellite (0.5-meter resolution) and an aerial stereo dataset (0.08-meter resolution).

In the experiments, the proposed method achieved Kappa coefficients of 94% and 93% for the satellite and aerial datasets, respectively. This marks a significant improvement (3% increase in Kappa coefficient) compared to the state-of-the-art methods like ChangeFormer, which achieved 91% for both datasets. By enhancing feature extraction and performing robustly on diverse data, the model emerges as a powerful tool for urban environment monitoring, offering a scalable and reliable solution for urban planning and management.

Key words: *3D Remote Sensing Data, Multiclass Building Change Detection, Transformer Networks, Spatial and Channel Attention Blocks, Data Augmentation.*

Correspondence Address*: School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*
Tel*: +98 21 61114256*
Email*: a.eftekhari@ut.ac.ir*

[Downloaded from jgit.kntu.ac.ir on 2026-01-29]

[DOI: 10.61882/jgit.13.1.95]