بهبود تشخیص صحنه در سنجش از راه‌دور با استفاده از یادگیری عمیق و انتخاب‌گر ویژگی

سید علی جهانی[۱]، کیوان محبی[۲*]، فرساد زمانی بروجنی[۳]

کارشناس‌ارشد رشتهٔ مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد اصفهان (خوراسگان)، اصفهان، ایران[۱]

استادیار گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد اصفهان (خوراسگان)، اصفهان، ایران[۲]

استادیار گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران[۳]

چکیده

تصاویر سنجش از دور به‌عنوان منبع‌دادهٔ با ارزش می‌تواند در اندازه‌گیری و مشاهدهٔ ساختارهای دقیق در سطح زمین کمک کند. هدف این پژوهش ارائهٔ راه‌کار تشخیص صحنه در سنجش از راه‌دور با به‌کارگیری روش‌های یادگیری عمیق است. برای رفع محدودیت‌های روش‌های پیشین، یک رویکرد ترکیبی استخراج ویژگی مطرح شده‌است که در آن دو نوع ویژگی عمیق محلی و سراسری و یک نوع محلی دستی با یکدیگر ترکیب شده‌اند. برای استخراج ویژگی‌ها یک شبکهٔ کانولوشن پیش‌آموزش‌دیده با بیست لایهٔ تمام‌متصل پیشنهاد شده‌است؛ همچنین یک مرحله انتخاب ویژگی با استفاده از دو دسته الگوریتم‌های پالایه و بسته‌بند در این مدل تعبیه شده‌است؛ درنهایت با استفاده از الگوریتم‌های طبقه‌بندی مختلف تشخیص صحنه انجام می‌شود. ارزیابی راه‌کار پیشنهادی برای مجموعه‌داده‌های UCM، AID، RSSCN7 و NWPU-RESISC45 به‌ترتیب دقت ۹۹/۲۷٪ ، ۹۷/۹۱٪ ، ۹۹/۰۹٪ و ۹۳/۰۹٪ را کسب کرده‌است.

واژگان کلیدی: سنجش از راه‌دور، یادگیری عمیق، ویژگی عمیق، یادگیری ترکیبی، از پیش‌آموزش‌دیده.

Improving Scene Recognition in Remote Sensing Using Deep Learning and Feature Selector

Seyyed Ali Jahani[1], Keyvan Mohebbi[*1], Farsad Zamani Boroujeni[2]

Master, Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran[1]

Assistant Professor, Department of Computer Engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran[2]

Assistant Professor, Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran[3]

Abstract

Remote sensing images as a valuable data source in Earth observation can help in measuring and observing detailed structures on the Earth's surface. Scene detection in remote sensing images has many applications in various fields such as urban planning, natural hazard detection, environmental monitoring, vegetation mapping, and geographic object detection. One of the key problems in the interpretation of remote sensing images is the scene classification of remote sensing images. Feature extraction is very important in scene detection and classification. Convolutional neural networks are one of the deep learning methods that have significantly increased the performance of tasks such as object recognition and scene classification, but their performance is highly dependent on the number of labeled images available, which are not available enough especially in the field of remote sensing. Recently, transfer learning, especially for the fine-tuning of pre-trained convolutional neural networks, has attracted more attention from researchers as a practical strategy for scene classification in remote sensing. However, the lack of use of local features and global deep model that is trained on the target data set is one of the limitations of current methods. Also, if these networks are not deep enough and the images do not pass through multiple filters, they cannot extract more semantic information, and the

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۴۰۲/۶/۱۸ • تاریخ پذیرش: ۱۴۰۳/۵/۳۱ • تاریخ انتشار: ۱۴۰۳/۱۰/۲۸ • نوع مطالعه: پژوهشی

[DOI: 10.61186/jsdp.21.3.69]

extracted features do not have high discrimination power, and as a result, scene recognition is not performed well. On the other hand, the features extracted through local features are very large, and not using feature selector methods reduces the accuracy of the model. In this research, to solve the mentioned limitations, a hybrid approach of feature extraction has been proposed in which three types of features including two types of deep local and global features and one type of manual local feature are combined with each other. To extract deep features, pre-trained convolutional networks have been used. The pre-trained networks used are: ResNet, InceptionNet, GoogleNet and EfficientNet_b0. In order to extract as much information as possible from the images, a convolutional network with 20 fully connected layers is proposed. Also, a combined feature selection stage consisting of two categories of filtering and packing algorithms is included in this model. Finally, scene detection is performed using several different classification algorithms. The different structure of pre-trained convolutional networks and their appropriate depth can be effective in improving the extraction of deep features. In addition, the combination of three categories of different features can provide a more comprehensive knowledge of images. The evaluation of the proposed solution on the UCM, AID, RSSCN7 and NWPU-RESISC45 datasets has obtained the accuracy of 99.27%, 97.91%, 99.09% and 93.09% respectively in identifying and classifying images. As a result, this solution has shown a better performance compared to the models that used the manual extraction of features, as well as the methods that use normal convolutional models.

Keywords: Remote Sensing, Deep Learning, Deep Feature, Hybrid Learning, Pre-Trained.

۱-مقدمه

تصاویر سنجش از دور به‌عنوان یک منبع دادهٔ بسیار ارزشمند در رصد زمین نقش بسیار مهمی ایفا می‌کنند؛ این تصاویر، ما را قادر می‌سازند تا به‌دقت ساختارها و الگوهای زمینی را بررسی و اندازه‌گیری کنیم. با توجه به رشد روزافزون فناوری در این زمینه، حجم تصاویر سنجش از دور به‌شدت افزایش یافته‌است[۱،۲]. به‌دلیل این افزایش، نیاز به روش‌ها و رویکردهای نوین برای تفسیر و بهره‌برداری این تصاویر بیش از پیش احساس می‌شود. تفسیر تصاویر سنجش از دور به‌منظور طبقه‌بندی صحنه‌ها به یک حوزهٔ پژوهشی فوق‌العاده مهم و پرچالش تبدیل شده‌است؛ در این حوزه، هدف اصلی آن است که به هر تصویر سنجش از دور، برچسبی مناسب با طبقه‌های معنایی از پیش تعریف‌شده اختصاص دهیم. این کاربردها در زمینه‌های متنوعی از جمله برنامه‌ریزی شهری [۳]، برای اینکه بتوان برنامه‌ریزی کرد که کدام قسمت از شهر را چه‌چیزی تشکیل داده و مدیریت و برنامه‌ریزی انجام داد، تشخیص مخاطرات طبیعی [۴, ۵]، از طریق پایش گسل‌ها، پایش محیط [۶]، برای اینکه بدانیم در هر قسمت چه اتفاقاتی در حال رخ‌دادن است، نقشه‌برداری پوشش گیاهی و تشخیص اشیای جغرافیایی [۷] تأثیر می‌گذارند.

با وجود توسعهٔ روزافزون تصویربرداری سنجش از دور، چالش‌هایی نیز وجود دارد؛ یکی از مهم‌ترین این چالش‌ها، طبقه‌بندی صحنه‌های پیچیده در تصاویر سنجش از دور است. در تصاویر با وضوح پایین و متوسط، ممکن است اشیای مختلف منحنی پاسخ طیفی مشابهی داشته باشند و یا یک شیء منحنی‌های متفاوتی داشته باشد؛ بنابراین، روش‌های طبقه‌بندی در سطح پیکسل اشیا، محدودیت‌های زیادی دارند. درمقابل، استفاده از بلوک تصویر (برای مثال صحنه) به‌عنوان واحد اساسی برای تفسیر تصویر می‌تواند از اطلاعات محتوایی برای رفع ابهام تفسیر استفادهٔ مؤثری کند[۸]. یکی از محدودیت‌هایی که در این بین مطرح‌بوده و در مطالعات پیشین مورد توجه قرارنگرفته، تنوع در ویژگی‌های قابل‌استخراج است[۹، ۱۰]؛ به‌عبارت ساده مطالعات موجود فقط متمرکز بر یک نوع ویژگی مثل ویژگی‌های عمیق و یا ویژگی‌های محلی بوده‌اند و هم‌زمان از مزایای استخراج هر دو بهره نبرده‌اند[۱۱]. در پژوهش حاضر، به‌دنبال یک راه‌کار ترکیبی جهت استخراج ویژگی‌های تمایزدهنده برای تفسیر تصاویر سنجش از دور هستیم؛ در این رویکرد، ما از سه نوع ویژگی، شامل دو نوع ویژگی سراسری، همچنین یک نوع از ویژگی‌های محلی دستی بهره برده و به‌منظور بهبود کارایی تشخیص صحنه‌ها، آن‌ها را با هم ترکیب می‌کنیم.

ادامهٔ این مقاله زیربخش‌بندی شده‌است. بخش دو به تعریف اصول و مفاهیم مورد استفاده در این پژوهش می‌پردازد. بخش سه، تعدادی از کارهای انجام‌شدهٔ پیشین در زمینهٔ سنجش از دور را مرور می‌کند. بخش چهار، راه‌کار پیشنهادی در این زمینه را مطرح کرده و به توضیح مراحل مختلف آن اختصاص دارد. بخش پنج، به ارزیابی راه‌کار ارائه‌شده می‌پردازد؛ همچنین مقایسهٔ این راه‌کار با روش‌های پیشین مرتبط در بخش شش انجام و در نهایت، نتیجه‌گیری و پیشنهاد کارهای آینده در بخش هفت مطرح شده‌است.

۲-ادبیات پژوهش

این بخش شامل تعریف مفاهیم پایهٔ مورد استفاده در پژوهش حاضر است.

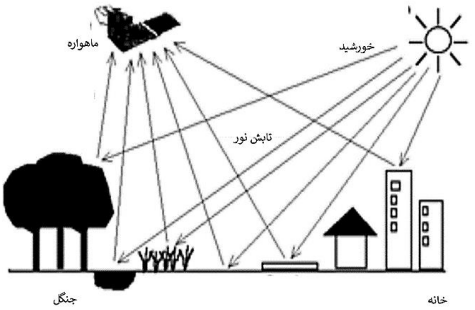
[DOI: 10.61186/jsdp.21.3.69]

۲-۱-سنجش از دور

سنجش از دور (RS) که رصد زمین نیز نامیده می‌شود، به جمع‌آوری اطلاعات در مورد اشیا یا مناطق در سطح زمین بدون تماس مستقیم با جسم یا منطقه اشاره دارد. سنجش از دور روشی برای مشاهدهٔ سطح زمین یا جو از خارج از فضا با استفاده از ماهواره و یا هواپیماهای هوابرد است. سنجش از دور از یک یا چند بخش از طیف الکترومغناطیسی استفاده کرده و انرژی الکترومغناطیسی منعکس‌شده یا ساطع‌شده از سطح زمین را ثبت می‌کند. مقدار تابش از یک شیء، تحت‌تأثیر ویژگی‌ها و تابش به آن شیء است. چشم انسان نور خورشیدی بازتاب‌شده از این اشیا را ثبت و مغز ما رنگ‌ها، تن‌های خاکستری و تغییرات شدت را تفسیر می‌کند. در سنجش از دور انواع مختلفی از ابزارها و وسایل برای ایجاد تابش الکترومغناطیسی در خارج از این محدوده از چهارصد تا هفتصد نانومتر برای چشم انسان، به‌ویژه مادون قرمز نزدیک، مادون قرمز میانی، مادون قرمز حرارتی و مایکروویو استفاده می‌شود.

تصاویر سنجش از دور کاربردهای زیادی در نقشه‌برداری کاربری و پوشش، کشاورزی، نقشه‌برداری خاک، جنگل‌داری، برنامه‌ریزی شهری، پژوهش‌های باستان‌شناسی، مشاهدات نظامی و بررسی ژئومورفولوژیکی، تغییرات پوشش زمین، جنگل‌زدایی، پویایی پوشش گیاهی، پویایی کیفیت آب و رشد شهری دارد[۱۲].

فرایند سنجش از دور در شکل (۱) نشان داده شده‌است. نور خورشید به سطح زمین می‌تابد و ماهواره انعکاس تابش نور از اشیا را تصویربرداری می‌کند و بدین‌صورت می‌توان با طبقه‌بندی این تصاویر، سنجش‌ازراه‌دور را انجام داد.



(شکل-۱): فرایند سنجش از دور[۵]
(Figure-1): Remote sensing process [5]

۲-۲-شبکه‌های عصبی از پیش آموزش‌دیده

شبکه‌های عمیق از پیش آموزش‌دیده به شبکه‌های عصبی عمیق اطلاق می‌شوند که قبل از استفاده در یک وظیفهٔ خاص، روی یک مجموعه‌دادهٔ بزرگ پیش‌آموزش دیده‌اند؛ این مرحله پیش‌آموزش به شبکه امکان می‌دهد تا ویژگی‌های عمومی را یاد بگیرد که سپس برای انتقال یادگیری به وظایف مشابه دیگر مورد استفاده قرار می‌گیرند[۱۳]. مزایای اصلی استفاده از شبکه‌های عمیق از پیش آموزش‌دیده عبارت‌اند از:

۱. **آموزش سریع‌تر**: از آنجا که لایه‌های شبکه قبلاً ویژگی‌های مفیدی را یاد گرفته‌اند، فقط لایه‌های جدید باید برای وظیفهٔ خاص آموزش داده شوند؛ این منجر به آموزش بسیار سریع‌تر نسبت به آموزش شبکه از ابتدا می‌شود.
۲. **عملکرد بهتر**: ویژگی‌های آموخته‌شده از یک مجموعه‌دادهٔ بزرگ در طول فرایند پیش‌آموزش به‌خوبی به وظایف مشابه دیگر منتقل می‌شوند و درنتیجه بهبود عملکرد بیشتری در مقایسه با آموزش از ابتدا دارند، به‌خصوص زمانی‌که از مجموعه‌داده‌های کوچک‌تر برای تنظیم دقیق استفاده می‌شود.
۳. **نیاز به دادهٔ کمتر**: ویژگی‌های از پیش آموزش‌دیده به شبکه امکان تعمیم بهتر با استفاده از نمونه‌های کمتر در طول تنظیم دقیق برای کار خاص را می‌دهند.

معمول‌ترین شبکه‌های پیش‌آموزش‌دیده شامل AlexNet، VGGNet، ResNet و Inception هستند که بر روی مجموعه‌دادهٔ ImageNet آموزش دیده‌اند. ویژگی‌های آموخته‌شده از تصاویر طبیعی در این شبکه‌های عصبی کانولوشنال را می‌توان برای طبقه‌بندی تصاویر برای وظایف بینایی کامپیوتری دیگر استفاده کرد[۱۴].

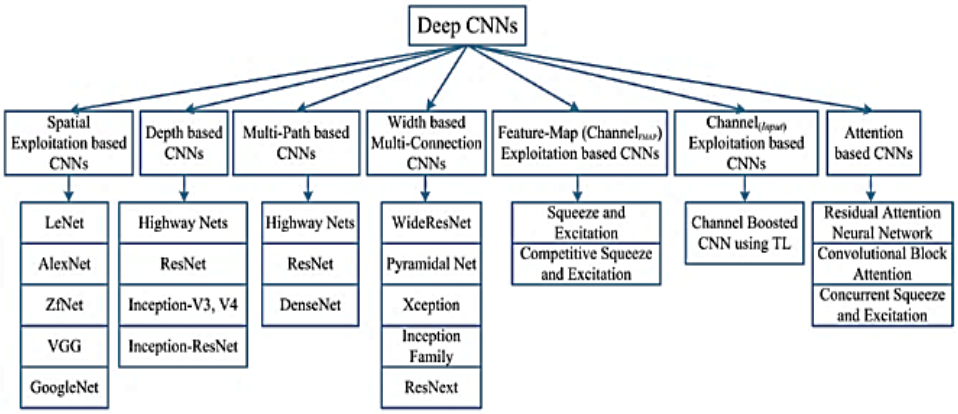
شبکه‌های از پیش آموزش‌دیده، با تثبیت لایه‌های قبلی که ویژگی‌های عمومی را ارائه می‌دهند و فقط تطبیق لایه‌های بعدی برای وظایف جدید مورد استفاده قرارمی‌گیرند. ویژگی‌های لایه‌های قبلی به‌عنوان یک استخراج‌کنندهٔ ویژگی عمومی عمل می‌کنند که دوباره مورد استفاده قرارمی‌گیرد؛ درحالی‌که لایه‌های ویژه به‌مرور در طول تنظیم دقیق تطبیق می‌یابند. این روش نسبت به آموزش شبکه از ابتدا نیاز به داده و زمان آموزش کمتری دارد. تصویری از طبقه‌بندی معماری‌های عمیق کانولوشن در هفت‌دستهٔ مختلف در شکل (۲) آورده شده‌است.

۲-۳-استخراج‌کنندهٔ ویژگی محلی

استخراج‌کننده‌های ویژگی محلی[1] الگوریتم‌هایی هستند که در تصاویر، نقاط مهم و مناطق برجسته را تشخیص‌داده و از آن‌ها ویژگی‌های محلی و توصیف‌گرها را استخراج می‌کنند[۱۵]. این ویژگی‌ها به‌ویژه از تغییرات مقیاسی، چرخشی، تغییرات روشنایی و غیره در تصویر نسبت به تبدیل‌های مختلف مستقل‌اند. استفاده از استخراج‌کننده‌های ویژگی محلی باعث توانایی بالاتر در تشخیص اشیا، تطبیق تصاویر و انجام وظایف طبقه‌بندی می‌شود؛

[1] Local Feature Extractor

[DOI: 10.61186/jsdp.21.3.69]



(شکل-۲): طبقه‌بندی معماری‌های عمیق کانولوشن[۱۴]

(Figure-2): Classification of deep convolution architectures [16]

یکی از معروف‌ترین استخراج‌کننده‌های ویژگی محلی، الگوریتم ORB[1] است، که به‌صورت زیر عمل می‌کند:

تشخیص نقاط کلیدی با استفاده از FAST: این الگوریتم ابتدا با استفاده از الگوریتم FAST نقاط کلیدی پتانسیل را در تصویر تشخیص می‌دهد.

محاسبهٔ جهت‌های نقاط کلیدی: برای هر نقطهٔ کلیدی، جهت‌هایی محاسبه می‌شود تا بتوانیم نقاط کلیدی را به‌صورت چرخشی توصیف کنیم.

محاسبهٔ توصیف‌گرهای BRIEF: در این مرحله با استفاده از الگوریتم BRIEF، توصیف‌گرهای دودویی برای نقاط کلیدی محاسبه می‌شود.

توصیف‌گرهای ORB: توصیف‌گرهای BRIEF به‌صورت چرخشی تصویر برای نقاط کلیدی ایجاد می‌شوند و نقاط کلیدی به‌وسیلهٔ توصیف‌گرهای ORB توصیف می‌شوند؛ بنابراین، استفاده از ORB به تشخیص نقاط کلیدی و توصیف ویژگی‌های محلی در تصاویر کمک می‌کند. این به‌ویژه در وظایفی مانند تطبیق تصاویر و طبقه‌بندی بسیار مفید است. یکی دیگر از استخراج‌کننده‌های محلی معروف، LBP[2] است، که به‌صورت زیر عمل می‌کند:

تولید الگوهای دودویی محلی: برای هر پیکسل در تصویر، پیکسل‌های همسایگی آن با توجه به مقدار پیکسل مرکزی در آستانه‌ای تبدیل به یک عدد دودویی می‌شوند. این الگوهای دودویی نمایانگر الگوهای محلی در تصویرند.

ساخت هیستوگرام الگوهای محلی: در این مرحله، یک هیستوگرام از الگوهای محلی در یک منطقه محاسبه می‌شود. این توصیف‌گرهای LBP می‌توانند به‌صورت محلی برای نقاط کلیدی یا به‌صورت متراکم در تمام تصویر محاسبه شوند.

بنابراین، استخراج‌کننده‌های ویژگی محلی به تشخیص نقاط مهم در تصاویر و توصیف ویژگی‌های آن‌ها پرداخته و در وظایفی مانند تطبیق تصاویر و طبقه‌بندی بسیار مفیدند. SIFT و LBP دو مثال از این استخراج‌کننده‌های ویژگی محلی هستند که به‌ترتیب بر اساس تفاوت‌های گرادیان و الگوهای محلی دودویی عمل می‌کنند[۱۴].

۲-۴-الگوریتم‌های انتخاب‌گر ویژگی

الگوریتم‌های انتخاب‌گر ویژگی[3] یکی از مهم‌ترین مراحل در پردازش داده‌ها و مدل‌سازی ماشینی هستند. این الگوریتم‌ها برای انتخاب ویژگی‌های مهم و معنی‌دار از داده‌ها استفاده می‌کنند تا مدل‌های بهتر و کارآمدتری ایجاد شوند[۱۶]؛ دو الگوریتم معروف در این زمینه، مربع کای[4] و بسته‌بند[5] هستند. الگوریتم مربع کای به‌طور معمول برای مسائل دسته‌بندی و الگوریتم و بسته‌بند برای مسائل پیش‌بینی (رگرسیون) استفاده می‌شوند. در ادامه توضیح مختصری در مورد هر یک ارائه می‌شود:

الگوریتم مربع کای برای انتخاب ویژگی‌هایی استفاده می‌شود که با متغیر هدف (به‌طور معمول یک متغیر دسته‌بندی) رابطهٔ آماری معنی‌دار دارند. مربع کای به‌طور معمول در مسائل دسته‌بندی و تحلیل متغیرها بر اساس فراوانی‌های مشترک در جداول تراکنشی یا جداول ترکیبی استفاده می‌شود؛ در این الگوریتم ابتدا یک جدول تراکنشی برای ویژگی و متغیر هدف ساخته می‌شود، سپس مقدار کای برای هر ویژگی محاسبه می‌شود. ویژگی‌هایی که مربع کای بزرگ‌تری دارند، به‌عنوان ویژگی‌های مهم انتخاب می‌شوند[۱۴].

در الگوریتم بسته‌بند، ویژگی‌ها به‌صورت گروهی بررسی می‌شوند و هدف انتخاب ویژگی‌هایی است که در

[1] Oriented FAST and Rotated BRIEF
[2] Local Binary Patterns

[3] Feature Selection Algorithms
[4] Chi Squared
[5] Wrapper

[DOI: 10.61186/jsdp.21.3.69]

کنار هم به‌بهترین نتیجه برسند؛ مثال بارزی از این الگوریتم، روش‌های حذف بازگشتی ویژگی[1] هستند؛ این روش به حذف تدریجی ویژگی‌های کم‌اهمیت از مجموعه‌داده و انتخاب ویژگی‌های مهم‌تر برای مدل می‌پردازد و به‌طور معمول در مسائل پیش‌بینی و مدل‌سازی استفاده می‌شود؛ به‌ویژه زمانی‌که تعداد ویژگی‌ها زیاد باشد و نیاز به ساده‌ترکردن مدل و کاهش ابعاد داده وجود داشته باشد. الگوریتم بسته‌بند از یک مدل ماشینی (به‌طور معمول مدل دسته‌بندی یا رگرسیون) شروع می‌کند، سپس ویژگی‌ها را به‌تدریج حذف می‌کند و عملکرد مدل را مورد ارزیابی قرارمی‌دهد. ویژگی‌هایی که حذف‌شده‌اند و مدل هنوز به‌خوبی عمل نکند، به‌تدریج بازگردانده نمی‌شوند.

درنهایت، انتخاب بین این الگوریتم‌ها و یا استفاده از الگوریتم‌های دیگر بستگی به مسئلهٔ مورد نظر و خصوصیات داده دارد؛ همچنین، به‌طور معمول نیاز به تنظیمات و تجربیات مکرر برای انتخاب بهترین ویژگی‌ها وجود دارد.

۲-۵ شبکه‌های عصبی گراف (GCN)

شبکه‌های عصبی گراف[2] (GNN) یا شبکه‌های کانولوشن گراف (GCNs) مدل‌های یادگیری ماشینی هستند که از داده‌هایی با ساختار نموداری مانند شبکه‌های اجتماعی یا مولکول‌ها برای مسائلی که وابستگی‌های میان نقاط داده مهم‌اند، بهره می‌برند؛ این مدل‌ها به گره‌ها و ارتباطات بین آن‌ها در نمودار توجه‌کرده و تلاش می‌کنند نمایش‌های اطلاعاتی از هر گره با درنظرگرفتن اطلاعات از گره‌های مجاورشان را یاد بگیرند. GNN/GCN به تشخیص الگوها و ویژگی‌های مهم در داده‌های گرافیکی کمک می‌کنند و در بسیاری از وظایف مانند طبقه‌بندی گره‌ها، پیش‌بینی ارتباطات و سایر وظایف مشابه مؤثرند. این مدل‌ها با استفاده از مراحل مختلفی مانند نمایش ورودی، پیچیدگی نمودار، تجمع چندلایه و لایهٔ خروجی آموزش می‌بینند.

۳-کارهای انجام‌شده

ژانگ و همکاران[3] یک شبکهٔ کانولوشن کامل ساده و کارآمد مبتنی بر DenseNet برای طبقه‌بندی صحنه‌های سنجش از دور پیشنهاد دادند؛ در این مدل تعداد زیادی اتصالات متراکم وجود دارد که شبکه را عمیق‌تر می‌کند، اما تعداد پارامترها را به میزان قابل توجهی افزایش نمی‌دهد. شبکهٔ طراحی‌شده بیش از صدلایه دارد؛ بااین‌حال، تنها حدود هفت‌میلیون پارامتر دارد که بسیار کمتر از تعداد پارامترهای VGG است. این شبکه می‌تواند تصاویر ورودی را با اندازه‌های مختلف بپذیرد؛ علاوه‌براین، آن‌ها به‌جای لایهٔ به‌طورکامل متصل، لایهٔ کانولوشن را در شبکه درج کردند تا بتوان ویژگی‌های خروجی شبکه را بدون عملیات مسطح طبقه‌بندی کرد، که عملیات طبقه‌بندی را ساده می‌کند. روش آن‌ها عملکرد طبقه‌بندی را در مجموعه‌دادهٔ UCM، مجموعه‌دادهٔ AID، مجموعه‌دادهٔ OPTIMAL-31 و مجموعه‌دادهٔ NWPU-RESISC45 بهبود قابل توجهی می‌دهد و به دقتی بین ۹۲٪ تا ۹۵٪ رسیده‌است[۱۷].

سان و همکاران[4]، یک شبکهٔ دوجهته دروازه‌ای[5] برای ادغام ویژگی سلسله‌مراتبی و حذف اطلاعات تداخلی پیشنهاد دادند؛ در این روش ابتدا عملکرد هر یک از ویژگی‌های کانولوشن تحلیل کمّی و ترکیبی برتر از ویژگی‌های کانولوشن انتخاب می‌شود؛ سپس، یک اتصال دوطرفه برای تجمیع سلسله‌مراتبی ویژگی‌های کانولوشن چندلایه پیشنهاد می‌شود. هم جهت بالابه‌پایین و هم جهت پایین‌به‌بالا برای جمع‌کردن ویژگی‌های کانولوشن چندلایه درنظرگرفته می‌شوند و یک تابع دروازه‌ای برای حذف اطلاعات تداخل در اتصال دو جهته استفاده می‌شود. درنهایت، ویژگی‌های معنایی و ویژگی‌های ظاهری برای طبقه‌بندی ادغام می‌شوند. روش پیشنهادی آن‌ها در چهار مجموعه‌دادهٔ طبقه‌بندی صحنهٔ سنجش‌دور (AID، UC-Merced، WHU-RS19، و OPTIMAL-31) بررسی شده و به‌دقت ۹۵٪ رسیده‌است[۱۸].

ژی و همکاران[6] یک شبکهٔ کانولوشن بدون‌مقیاس (SF-CNN)[7] برای طبقه‌بندی صحنهٔ سنجش از دور معرفی کردند. به‌طورخاص، در این شبکه لایه‌های تمام‌متصل[8] به لایه‌های کانولوشنی تبدیل می‌شوند، که نه‌تنها اجازه می‌دهد تصاویر ورودی دارای اندازه‌های دل‌خواه باشند؛ بلکه توانایی استخراج ویژگی‌های متمایز با استفاده از یک استراتژی سنتی مبتنی بر پنجرهٔ کشویی[9] را نیز حفظ می‌کنند؛ سپس، یک لایهٔ ادغام میانگین سراسری (GAP)[10] بعد از لایهٔ کانولوشن نهایی اضافه می‌شود تا تصاویر ورودی با اندازهٔ دل‌خواه را بتوان برای نمایش ویژگی‌هایی با اندازهٔ یک‌نواخت نگاشت کرد؛ درنهایت، از ویژگی‌های به‌دست‌آمده برای ایجاد

[1] Recursive Feature Elimination
[2] Graph Convolutional Networks
[3] Zhang et al.
[4] Sun et al.
[5] Gated bidirectional network
[6] Xie et al.
[7] Scale-free CNN
[8] Fully Connected
[9] Sliding-window
[10] Global average pooling (GAP)

[DOI: 10.61186/jsdp.21.3.69]

یک لایهٔ تمام‌متصل جدید استفاد که برای طبقه‌بندی نهایی به لایهٔ SoftMax تغذیه می‌شود. نتایج تجربی که با استفاده از چندین مجموعه‌دادهٔ واقعی انجام شده، برتری روش پیشنهادی آن‌ها را نسبت به چندین روش طبقه‌بندی معروف، از جمله روش‌های مبتنی بر CNN از پیش آموزش‌دیده نشان می‌دهد که توانسته در مجموعه‌دادهٔ UC MERCED LAND-USE به دقت ٪۹۹ دست یابد[۱۹].

شی و همکاران[1] یک روش طبقه‌بندی سبک‌وزن کانولوشن مبتنی بر هم‌جوشی شاخه‌ای ویژگی (LCNN-BFF)[2] برای طبقه‌بندی صحنهٔ سنجش از دور پیشنهاد دادند؛ برخلاف ساختار کانولوشن معمولی، مدل پیشنهادی آن‌ها ساختار استخراج ویژگی دوخطی[3] دارد. مدل آن‌ها پیچیدگی را تا حد زیادی کاهش داده‌است. ارزیابی روش روی چهار مجموعه‌دادهٔ NWPU-RESISC45، RSSCN7،UC و AID نشان داد که مدل به دقتی بین ۹۱٪ تا ٪۹۷ رسیده‌است[۲۰].

هوانگ و همکاران[4]، یک معماری کانولوشن جدید با نام RSSCNet [5] با قابلیت تعمیم بالا ارائه دادند؛ در این مدل از سیاست نرخ یادگیری چرخه‌ای دو مرحله‌ای[6] و روش یادگیری انتقالی بدون انجماد[7] برای سرعت‌بخشیدن به آموزش مدل و افزایش دقت استفاده شده‌است؛ همچنین از الگوریتم t-SNE برای کاهش ابعاد و الگوریتم LIME[8] برای بهبود نتایج در مواردی که مدل پیش‌بینی‌های اشتباهی انجام‌داده استفاده شده‌است. مقایسهٔ نتایج در سه مجموعه‌دادهٔ RSSCN7 ، UC Merced و WHU-RS19 نشان می‌دهد که مدل آن‌ها توانسته به دقت ۹۸٪ دست یابد[۲۱].

تانگ و همکاران[9] یک شبکهٔ کانولوشن جدید با نام (ACNet)[10] پیشنهاد دادند که ساختاری مشابه با شبکهٔ سیامی دارد. ساختار شبکه‌های سیامی از دو زیرشبکهٔ کانولوشن تشکیل شده‌است که در طی فرایند یادگیری این دو شبکه وزن‌های خود را به اشتراک می‌گذارند و به‌همین‌دلیل فرایند یادگیری آن‌ها دارای پیچیدگی محاسباتی کمتری است؛ در این مدل از ساختار سیامی جدید برای استخراج ویژگی‌های دقیق از تصاویر سنجش از دور استفاده شده‌است. روش آن‌ها بر روی سه مجموعه‌دادهٔ UCM، AID و NWPU ارزیابی شده و توانسته‌است به دقتی بین ۹۱٪ تا ٪۹۹ دست یابد[۲۲].

انور و همکاران[11] ادغام چندین مدل رنگ عمیق را برای طبقه‌بندی صحنه‌های هوایی بررسی کردند. آن‌ها نشان دادند که ترکیب چندین مدل رنگ عمیق عملکرد تشخیص را در مقایسه با استفاده از شبکهٔ RGB به‌تنهایی بهبود قابل توجهی می‌بخشد. آن‌ها از یک رویکرد فشرده‌سازی نظری اطلاعات[12] برای مقابله با ابعاد بالای داده استفاده کردند، که منجر به مجموعه‌ای از ویژگی‌های رنگ عمیق فشرده بدون افت قابل توجهی در دقت می‌شود. آزمایش‌های جامع بر روی پنج مجموعه‌دادهٔ صحنهٔ سنجش از دور شامل: UC-Merced با ۲۱ طبقه صحنه، WHU-RS19 با نوزده نوع صحنه، RSSCN7 با هفت دسته، AID با سی طبقه صحنهٔ هوایی، و NWPU-RESISC45 با ۴۵ صحنه انجام شده‌است. نتایج نشان می‌دهد که ادغام ویژگی‌های رنگ عمیق همیشه عملکرد طبقه‌بندی کلی را در مقایسه با ویژگی‌های عمیق RGB استاندارد بهبود داده و به‌دقتی بین ۹۰٪ تا ۹۷٪ دست یافته‌است[۲۳].

تیان و همکاران[13] یک‌شبکهٔ متراکم چندمقیاسی با نام SEMSDNet [14]را پیشنهاد دادند. چارچوب شبکهٔ کانولوشن متراکم چندمقیاسی با طبقه‌بندی کننده‌های متعدد و اتصالات متراکم می‌تواند خودکار به یک شبکهٔ کوچک و یک شبکهٔ عمیق باتوجه به پیچیدگی نمونه‌های آزمایشی و محدودیت منابع محاسباتی تبدیل شود؛ علاوه‌براین، به‌منظور استخراج ویژگی‌های مؤثرتر، شبکهٔ فشرده‌سازی و تحریک[15] در چارچوب درج شده‌است تا ویژگی‌های صحنه‌های مختلف را به‌صورت خودسازگار استخراج کند. نتایج تجربی بر روی مجموعه‌داده‌های UC Merced, NWPU-RESISC45, AID نشان می‌دهد که روش SEMSDNet نسبت به روش‌های پیشرفته در عملکرد و کارایی برتری دارد و دقتی در حدود ۹۱٪ تا ٪۹۸ ارائه داده‌است[۲۴].

بی و همکاران[16] یک‌شبکهٔ تجمیع توجه چندمقیاسی (MS2AP)[17] را برای مقابله با محدودیت‌های شبکه‌های کانولوشن ساده پیشنهاد دادند که سه بخش اصلی دارد. نخست، می‌توان آن را به‌راحتی در مدل‌های کانولوشن فعلی جاسازی کرد تا قابلیت نمایش ویژگی برای

[1] Shi et al
[2] Lightweight CNN Classification Method Based on Branch Feature Fusion
[3] Bilinear feature extraction
[4] Hung et al
[5] Remote Sensing Scene Classification Network
[6] Two-stage cyclical learning rate policy
[7] No-freezing transfer learning
[8] Local Interpretable Model, Agnostic Explanation
[9] Tang et al
[10] Attention Consistent Network
[11] Anwer et al
[12] Information theoretic compression
[13] Tian et al
[14] Multiscale Dense Networks with Squeeze and Excitation Attention
[15] Squeeze-and-Excitation
[16] Bi et al
[17] Multi-Scale Staking Attention Pooling

[DOI: 10.61186/jsdp.21.3.69]

صحنه‌های سنجش از راه دور را افزایش دهد. دوم، یک ماژول توجه کانال-فضایی[1] جدید برای استخراج محلی ویژگی‌های کلیدی در بردار ویژگی پیشنهاد دادند که در مقایسه با ماژول‌های فعلی، می‌تواند ویژگی‌های متمایز از بالابه‌پایین و ویژگی‌های پایین‌به‌بالا را هم از حوزهٔ کانال و هم از حوزهٔ فضایی ترکیب کند. سوم، یک اپراتور کانولوشن چندمقیاسی پیشنهاد دادند که می‌تواند نقشه‌های ویژگی چندمقیاسی را استخراج کند و اندازه‌های آن‌ها را ثابت نگه دارد. آزمایش‌های گسترده نشان می‌دهد که مدل آن‌ها در UCM، AID و NWPU به دقتی بین ٪۹۰ تا ٪۹۹ دست می‌یابد[۲۵].

می و همکاران[2] یک چارچوب جدید مبتنی بر نمایش پراکنده برای طبقه‌بندی صحنه در سنجش‌ازراه‌دور با ترکیب ویژگی‌های عمیق پیشنهاد دادند؛ در این مدل، ویژگی‌های چندسطحی ابتدا از لایه‌های مختلف شبکهٔ کانولوشن استخراج می‌شوند تا به‌طورکامل از توانایی یادگیری ویژگی‌های کانولوشن بهره‌برداری کنند. در این مدل از شبکه‌های کانولوشن آموزش‌دیدهٔ موجود برای مثال AlexNet، VGGNet، و ResNet50، برای استخراج ویژگی استفاده شده‌است، که در آن‌ها هیچ نمونهٔ برچسب‌گذاری‌شده‌ای مورد نیاز نیست؛ سپس، طبقه‌بندی مبتنی بر نمایش پراکنده برای ترکیب ویژگی‌های چندسطحی طراحی می‌شود، که به‌ویژه زمانی مؤثر است که فقط تعداد کمی از نمونه‌های آموزشی در دسترس باشد. نتایج تجربی بر روی دو مجموعه‌دادهٔ UC-Merced و WHU-RS19، نشان داد که روش پیشنهادی آن‌ها به دقت ٪۸۷ و ٪۸۸ رسیده‌است[۲۶].

شن و همکاران[3] یک معماری دوگانهٔ جدید با یک استراتژی گروه‌بندی-توجه-تلفیقی[4] برای بهبود عملکرد طبقه‌بندی صحنه ارائه دادند. به‌طورخاص، این مدل از دو شبکهٔ عصبی کانولوشن مختلف برای استخراج ویژگی استفاده می‌کند که در آن استراتژی گروه‌بندی-توجه-هم‌جوشی برای ترکیب ویژگی‌های کانولوشن‌ها به شیوه‌ای خوب و چندمقیاسی استفاده می‌شود؛ به‌این‌ترتیب، نمایش ویژگی‌های حاصل از صحنه افزایش می‌یابد؛ علاوه‌براین، برای پرداختن به مسئلهٔ ظواهر مشابه بین صحنه‌های مختلف، یک تابع هزینه تعریف شده‌است که تنوع‌های کوچک درون طبقاتی و فواصل بین طبقاتی بزرگ را افزایش می‌دهد. آزمایش‌های گسترده روی چهار مجموعه‌دادهٔ UCM، WHU-RS19، AID و OPTIMAL-31 نشان می‌دهد که دقتی در حدود ٪۹۴ تا ٪۹۶ حاصل شده‌است[۲۷].

ژو و همکاران[5] اشبکهٔ عصبی عمیق پیش‌آموزش‌دیدهٔ کانولوشن[6] VggNet16 که بر روی ImageNet [۲۸] آموزش دیده‌است را پیشنهاد دادند. از این مدل از شبکهٔ کانولوشن با عمق کم استفاده کرده‌است که این مورد از جمله محدودیت‌های آن محسوب می‌شود؛ اگر شبکه به اندازهٔ کافی عمیق نباشد، نمی‌تواند اطلاعات معنایی انتزاعی بیشتری استخراج کند؛ برای استخراج ویژگی از تصاویر سنجش از دور استفاده شده‌است. ویژگی‌های تصاویر در لایه‌های کانولوشن سوم لایهٔ سوم، کانولوشن چهارم لایهٔ سوم و کانولوشن پنجم لایهٔ سوم استخراج و این ویژگی‌ها بر روی شبکهٔ عصبی گراف آموزش داده شد؛ سپس خروجی‌های شبکهٔ گراف با خروجی لایهٔ تمام‌متصل در مدل VggNet16 ادغام و طبقه‌بندی خطی انجام شده‌است. این روش بر روی چهار پایگاه دادهٔ NWPU-RESISC45, AID, RSSCN7 و UCM به دقتی بین ٪۸۹ تا ٪۹۴ رسیده‌است[۱۱].

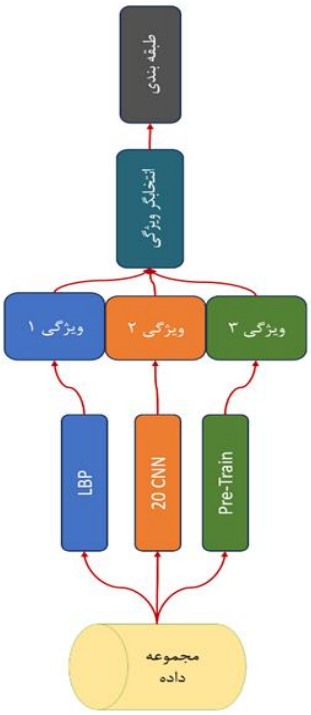
۴-روش پیشنهادی

در این پژوهش، روشی برای سنجش و طبقه‌بندی تصاویر ازراه‌دور با استفاده از یادگیری عمیق ارائه شده‌است. برای رفع مشکلاتی همچون عدم به‌کارگیری انواع مختلف ویژگی‌ها در تشخیص صحنه و نیز رفع محدودیت‌های راه‌کار های پیشین، یک رویکرد ترکیبی برای استخراج ویژگی ارائه شده‌است؛ در این رویکرد، سه نوع ویژگی شامل دو ویژگی عمیق محلی و سراسری و یک ویژگی محلی دستی، با یکدیگر ترکیب می‌شوند تا دقت تشخیص صحنه بهبود یابد. روش پیشنهادی شامل مراحل زیر در شکل (۳) نشان داده شده‌است.

۱. طراحی یک مدل بیست لایهٔ عمیق سراسری و آموزش این مدل با مجموعه‌دادهٔ هدف.
۲. استخراج ویژگی‌های تصاویر مجموعه‌داده با استفاده از مدل آموزش داده شدهٔ بیست لایهٔ عمیق.
۳. استخراج ویژگی‌های تصاویر با استفاده از مدل‌های از پیش آموزش‌دیدهٔ ResNet، InceptionNet، GoogleNet و EfficientNet_b0.
۴. استخراج ویژگی‌های تصاویر با استفاده از الگوریتم‌های استخراج‌کنندهٔ ویژگی‌های محلی.
۵. ادغام ویژگی‌های استخراج‌شده از دو مدل یادگیری عمیق و یک ویژگی‌های محلی.
۶. استفاده از الگوریتم انتخاب‌گر ویژگی برای پیداکردن مهم‌ترین ویژگی‌های هر تصویر.
۷. طبقه‌بندی تصاویر با استفاده از الگوریتم‌های طبقه‌بندی‌کنندهٔ ویژگی.
۸. ارزیابی کارایی مدل.

[1] residual channel-spatial attention module
[2] Mei et al
[3] Shen et al.
[4] grouping-attention-fusion
[5] Xu et al.
[6] Pretrained Convolutional Neural Networks

[DOI: 10.61186/jsdp.21.3.69]



(شکل-۳): روش پیشنهادی

(Figure-3): Proposed approach

در راه‌کار پیشنهادی ابتدا داده‌ها از منبع مربوطه دریافت و مراحل پیش‌پردازش شامل نرمال‌سازی تصاویر و تغییر ابعاد آن‌ها اعمال می‌شود؛ سپس یک مدل کانولوشن بیست‌لایهٔ تمام‌متصل طراحی می‌شود که هدف از آن، استفاده از یک مدل عمیق تمام‌متصل است که بتواند ویژگی‌های مهم را با توجه به عمیق‌بودن از تصاویر استخراج کند و با مجموعه‌دادهٔ هدف مورد آموزش قراربگیرد که این می‌تواند منجر به افزایش کیفیت ویژگی‌های استخراج‌شده شود. شبکه‌های عمیق ازپیش آموزش‌دیده به شبکه‌های عصبی عمیقی اطلاق می‌شوند که پیش از استفاده در یک وظیفهٔ خاص، روی یک مجموعه‌دادهٔ بزرگ آموزش دیده‌اند تا بتوانند به‌طور مؤثری ویژگی‌های مفید برای هر تصویر را استخراج کنند. برای هر پایگاه‌داده، ویژگی‌های تصاویر با استفاده از مدل‌های پیش‌آموزش‌دیده ResNet، InceptionNet، GoogleNet و EfficientNet_b0 استخراج می‌شود؛ سپس ویژگی‌های هر پایگاه داده با مدل بیست لایهٔ تمام‌متصل که ساختار آن در جدول (۱) آمده‌است، استخراج می‌شود.

در ساختار شبکهٔ بیست لایه، لایه‌های یک تا هجده لایه‌های کانولوشن هستند و هر کدام دارای تعداد مشخصی پالایه است. اندازهٔ پالایه در تمام لایه‌های کانولوشن سه در سه و لایه‌های سه، شش، ده، چهارده و هجده لایه‌های MaxPooling2D هستند. لایهٔ نوزده، لایهٔ Flatten، داده‌ها را به یک بردار تبدیل می‌کند. لایهٔ بیست لایهٔ Dense است که برای طبقه‌بندی استفاده می‌شود.

(جدول-۱): ساختار شبکهٔ بیست لایه

(Table-1): The structure of 20-layer network

لایه	نوع	تعداد پالایه	اندازهٔ پالایه
۱	Conv2D	۶۴	۳*۳
۲	Conv2D	۶۴	۳*۳
۳	MaxPooling2D	-	۲*۲
۴	Conv2D	۱۲۸	۳*۳
۵	Conv2D	۱۲۸	۳*۳
۶	MaxPooling2D	-	۲*۲
۷	Conv2D	۲۵۶	۳*۳
۸	Conv2D	۲۵۶	۳*۳
۹	Conv2D	۲۵۶	۳*۳
۱۰	MaxPooling2D	-	۲*۲
۱۱	Conv2D	۵۱۲	۳*۳
۱۲	Conv2D	۵۱۲	۳*۳
۱۳	Conv2D	۵۱۲	۳*۳
۱۴	MaxPooling2D	-	۲*۲
۱۵	Conv2D	۵۱۲	۳*۳
۱۶	Conv2D	۵۱۲	۳*۳
۱۷	Conv2D	۵۱۲	۳*۳
۱۸	MaxPooling2D	-	۲*۲
۱۹	Flatten	-	-
۲۰	Dense	۷	

در مرحلهٔ بعدی، ویژگی‌های محلی هرتصویر با الگوریتم‌های ORB و LBP استخراج می‌شود؛ ویژگی‌های استخراج‌شده ترکیب و پس از نرمال‌سازی بین صفر و یک، بهترین ویژگی‌ها برای هر تصویر با استفاده از الگوریتم‌های مربع کای و بسته‌بند انتخاب می‌شود. درنهایت با استفاده از چندین الگوریتم طبقه‌بندی مختلف شامل روش‌های انفرادی مثل ماشین بردار پشتیبان[1] و روش‌های گروهی مثل جنگل تصادفی[2] تشخیص صحنه انجام می‌شود.

۵-مجموعه‌داده

مرحلهٔ نخست در راه‌کار پیشنهادی دریافت داده‌ها است. در این پژوهش، از چهار مجموعه‌داده به‌شرح زیر استفاده شده‌است:

- مجموعه‌دادهٔ نخست، پایگاه دادهٔ RSSNC7 است. این مجموعه‌داده را دانشگاه ووهان از Google Earth استخراج کرد و در مجموع ۲۸۰۰ تصویر دارد که به هفت دسته تقسیم می‌شوند و ابعاد هر تصویر ۴۰۰ × ۴۰۰ پیکسل است. هر دسته صحنه در چهار مقیاس مختلف (۱:۷۰۰، ۱:۱۳۰۰، ۱:۲۶۰۰ و ۱:۵۲۰۰) نمونه‌برداری می‌شوند. شکل (۴) برخی از تصاویر نمونه از این مجموعه را ارائه می‌دهد.

[1] Support Vector Machine
[2] Random Forest

[DOI: 10.61186/jsdp.21.3.69]

(شکل-۴): نمونه‌هایی از مجموعه‌دادهٔ RSSNC7
(Figure-4): RSSNC7 Dataset Samples

- مجموعه‌دادهٔ دوم، پایگاه‌دادهٔ NWPU-RESISC45 ، یک مجموعه‌داده در مقیاس بزرگ با تنوع تصویر فراوان است؛ این مجموعه‌داده شامل ۳۱۵۰۰ تصویر رنگی در فضای رنگی RGB است که به ۴۵ دستهٔ معنایی تقسیم شده‌اند. در NWPU-RESISC45، هر کلاس شامل هفتصد تصویر با اندازهٔ ثابت ۲۵۶ × ۲۵۶ پیکسل و وضوح فضایی از حدود ۰/۲ به سی متر متغیر است. شکل (۵) برخی از تصاویر نمونه از این مجموعه‌داده را نشان می‌دهد.
- مجموعه‌دادهٔ سوم، موسوم به پایگاه‌دادهٔ AID را دانشگاه ووهان از Google Earth استخراج کرده و در مجموع شامل ده‌هزار تصویر و هر تصویر بر روی ۶۰۰ × ۶۰۰ پیکسل با وضوح بین ۰.۵ تا هشت متر ثابت شده‌است. مجموعه‌دادهٔ AID شامل سی دستهٔ معنایی و تعداد تصاویر در هر کلاس از ۲۲۰ تا ۴۲۰ متغیر است. در شکل (۶) تعدادی از تصاویر این پایگاه‌داده ارائه شده‌است.

(شکل-۵): نمونه‌هایی از مجموعه‌دادهٔ NWPU-RESISC45
(Figure-5): NWPU-RESISC45 Dataset Samples

(شکل-۶): نمونه‌هایی از مجموعه‌دادهٔ AID
(Figure-6): AID Dataset Samples

- مجموعه‌دادهٔ چهارم؛ یعنی پایگاه‌دادهٔ UCM را آزمایشگاه بینایی کامپیوتری دانشگاه کالیفرنیا از نقشهٔ ملی سازمان زمین‌شناسی ایالات‌متحده (USGS) جمع‌آوری کرده و در مجموع شامل ۲۱۰۰ تصویر است که به ۲۱ دسته تقسیم می‌شوند و هر تصویر دارای ابعاد ۲۵۶ × ۲۵۶ پیکسل با وضوح یک فوت است. شکل (۷) برخی از تصاویر نمونه از این مجموعه‌داده‌های معیار را نشان می‌دهد.

(شکل-۷): نمونه‌هایی از مجموعه‌دادهٔ UCM
(Figure-7): UCM Dataset Samples

۶- طراحی آزمایش‌ها

روش پیشنهادی با استفاده از زبان پایتون[1] و کتابخانهٔ پایتورچ[2] پیاده‌سازی شده‌است؛ همچنین برای استفاده از توصیف‌گرهای محلی و روش‌های انتخاب ویژگی از کتابخانه‌های opencv, scikit learn, numpy, pandas استفاده و تمامی آزمایش‌ها در محیط Google Colab پیاده‌سازی شده‌است.

به‌منظور ارزیابی کارایی روش پیشنهادی از معیار دقت[3] که به‌وسیلهٔ رابطهٔ (۱) محاسبه می‌شود، استفاده شده‌است:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۱)$$

(جدول-۲): پارامترهای مدل بیست لایه برای آموزش
(Table-2): The parameters of 20-layer model for training

پارامترها	مقدار
تکرار[4]	۳۳۰
اندازه دسته[5]	۶۴
تابع فعال ساز[6]	Relu
بهینه‌ساز[7]	Adam
نرخ یادگیری[8]	۰.۰۰۱
درصد داده‌های آموزش[9]	۸۰

پارامترهای مربوط به مدل مورد استفاده در راه‌کار پیشنهادی در جدول (۲) آورده شده‌است.

[1] Python
[2] Pytorch
[3] Accuracy
[4] epochs
[5] Batch size
[6] Activation function
[7] Optimizer function
[8] Learning rate
[9] Training rate

[DOI: 10.61186/jsdp.21.3.69]

برای تکرار، عدد ۳۳۰ انتخاب شد؛ زیرا استفاده از تعداد بیشتر برای تکرار آموزش مدل، موجب خطای بیش‌برازش[1] می‌شود. برای سرعت بالا در آموزش مدل، اندازهٔ دسته ۶۴ انتخاب شد؛ همین‌طور از تابع فعال‌ساز Relu و بهینه‌ساز Adam با نرخ یادگیری ۰/۰۰۱ استفاده شده‌است؛ به این علت که درنرخ ۰/۰۱ مدل نمی‌توانست به‌خوبی آموزش ببیند و همچنین نرخ ۰/۰۰۰۱ منجر به‌کندشدن روند یادگیری شد.

(جدول-۳): پارامترهای مربع کای
(Table-3): Chi-squared parameters

نام مدل	آزمون آماری	تعدادویژگی برتر
SelectKBest	Chi-squared	۵۰

پارامترهای بعدی در جدول (۳) مربوط به انتخاب‌گر ویژگی بر اساس پالایه است که با استفاده از آزمون آماری مربع کای و انتخاب پنجاه‌ویژگی برتر برای هر تصویر آورده شده‌است.

در جدول (۳) تعداد ویژگی برتر به معنای این است که در نهایت پنجاه ویژگی برتر برای هر تصویر از پایگاه‌داده از بین ۱۵۰۰ ویژگی برای هر تصویر انتخاب می‌شود؛ همچنین آزمون آماری مورد استفاده در این مدل، آزمون مربع کای وابستگی بین دو متغیر را اندازه‌گیری کرده و فرضیهٔ صفر مبنی بر مستقل‌بودن دو متغیر را ارزیابی می‌کند.

در زمینهٔ انتخاب ویژگی، آزمون مربع کای برای امتیازدهی به هر ویژگی براساس وابستگی آن به متغیر هدف (یعنی برچسب‌ها) استفاده می‌شود. تابع chi2 در scikit-learn به‌عنوان تابع امتیازدهی برای انتخاب ویژگی با SelectKBest استفاده می‌شود. ورودی این تابع دو آرایهٔ ماتریس ویژگی و متغیر هدف است. ماتریس ویژگی فقط مقادیر غیرمنفی و متغیر هدف حاوی دسته‌بندی است.

پارامترهای مربوط به استفاده از روش بسته‌بند برای انتخاب ویژگی در جدول (۴) آورده شده‌است.

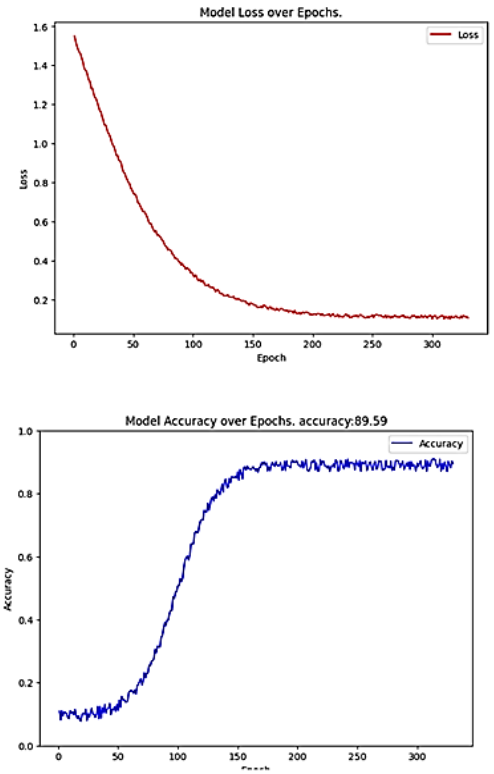
(جدول-۴): پارامترهای الگوریتم بسته‌بند
(Table-4): Wrapper algorithm parameters

انتخاب‌گر	طبقه‌بندی	گام	اعتبارسنجی متقابل
RFECV	LogisticRegression	۳۳	۱۰۰

در این جدول، پارامتر طبقه‌بندی، رگرسیون لجستیک است که دسته‌بندی ویژگی‌ها در انتخاب‌گر را انجام می‌دهد. پارامتر گام[2] درنمونهٔ RFECV مشخص می‌کند که در هر تکرار چند ویژگی حذف شود تا زمانی‌که به تعداد بهینهٔ ویژگی‌ها برسد. پارامتر اعتبارسنجی متقابل، تعداد فولدهای اعتبارسنجی متقاطع را برای استفاده در فرایند انتخاب ویژگی مشخص می‌کند. اعتبارسنجی متقابل، روشی برای ارزیابی عملکرد یک مدل یادگیری ماشینی با تقسیم داده‌ها به چندین زیرمجموعه و استفاده از هر زیرمجموعه برای آموزش و آزمایش است.

۶-۱-آزمایش نخست: بررسی اثر استفاده از مدل بیست لایهٔ تمام‌متصل بر روی کارایی تشخیص صحنه ازراه‌دور

در مرحلهٔ نخست این آزمایش، مجموعه‌دادهٔ RSSNC7 به تعداد ۳۳۰ دور روی مدل آموزش داده می‌شود. همان‌طورکه در شکل (۸) نمایش داده شده، ابتدای آموزش دقت مدل از ۱۲٪ شروع و پس از حدود صد دور آموزش به بالای ۸۰ درصد و تابع ضرر نیز پس از صد دور آموزش کمتر شده و پس از ۳۳۰ دور، دقت به ۸۹/۵۰٪ رسیده‌است.

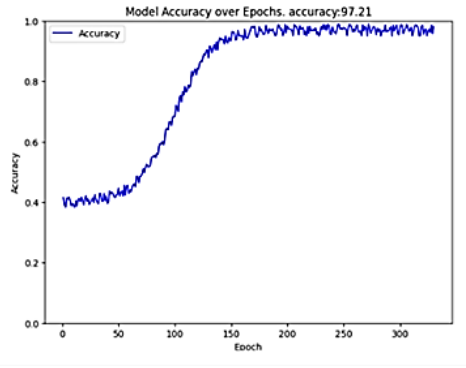


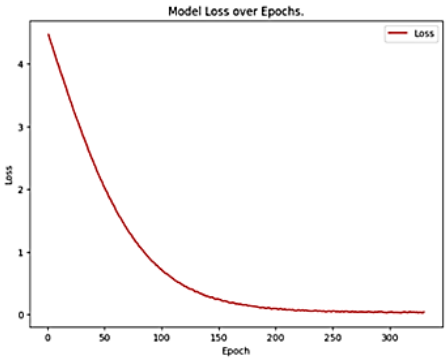
(شکل-۸): مقادیر دقت و ضرر برای مجموعه‌دادهٔRSSNC7
(Figure-8): Accuracy and Loss values for RSSNC7 dataset

در مرحلهٔ بعد، مدل با مجموعه‌دادهٔ UCM به تعداد ۳۳۰ دور آموزش داده می‌شود. همان‌طورکه در شکل (۹) نمایش داده‌شده، ابتدای آموزش دقت مدل از ۴۱٪ شروع و پس از حدود صد دور آموزش به بالای ۹۰٪ و تابع ضرر پس از صد دور آموزش کمتر شده و پس از ۳۳۰ دور دقت به ۹۷/۲۱٪ رسیده‌است.

[1] overfitting
[2] Step

[DOI: 10.61186/jsdp.21.3.69]

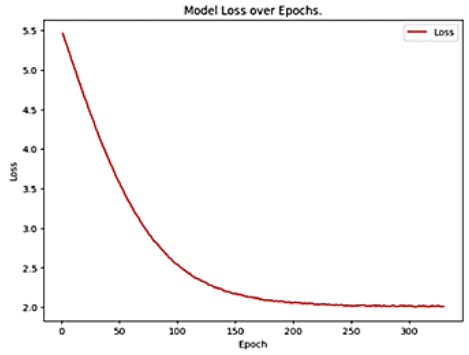


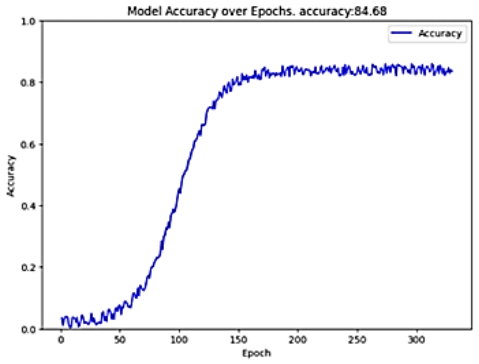


(شکل-۹): مقادیر دقت و تابع ضرر برای مجموعه‌دادهٔUCM

Figure-9): Accuracy and Loss values for UCM dataset

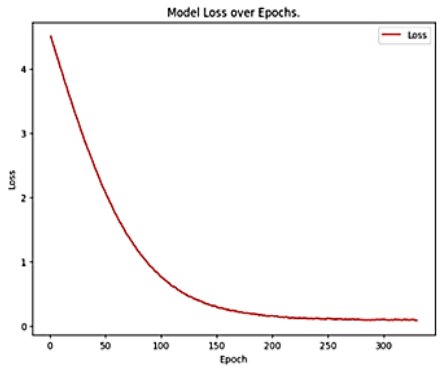
سپس مدل با مجموعه‌دادهٔ NWPU45 به تعداد ۳۳۰ دور آموزش داده می‌شود. همان‌طورکه در شکل (۱۰) نمایش داده‌شده، ابتدای آموزش دقت مدل از ۷٪ شروع و پس از حدود صد دور آموزش به بالای ۷۵٪ و تابع ضرر پس از صد دور آموزش کمتر شده و پس از ۳۳۰ دور، دقت به ۸۴/۶۸٪ رسیده‌است.

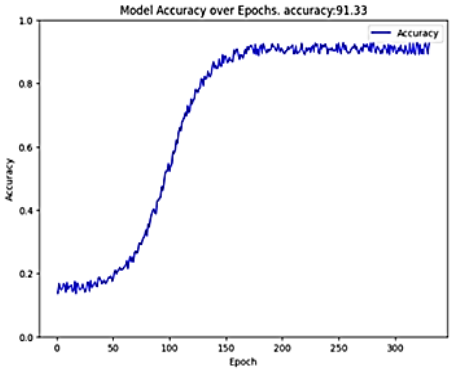




(شکل-۱۰): مقادیر دقت و ضرر برای مجموعه‌دادهٔ NWPU45

(Figure-10): Accuracy and Loss values for NWPU45 dataset





(شکل-۱۱): مقادیر دقت‌وضرر برای مجموعه‌دادهٔ AID

(Figure-11): Accuracy and Loss values for AID dataset

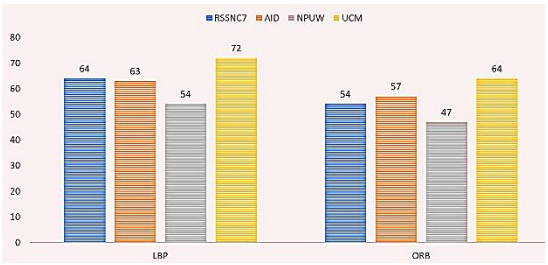
در مرحلهٔ آخر، مدل با مجموعه‌دادهٔ AID به تعداد ۳۳۰ دور آموزش داده می‌شود. همان‌طورکه در شکل (۱۱) نمایش داده شده، ابتدای آموزش دقت مدل از ۱۷٪ شروع و پس از حدود صد دور آموزش به بالای ۸۵ درصد رسید. تابع ضرر پس از صد دور آموزش کمتر شده و پس از ۳۳۰ دور، دقت به ۹۱/۳۳٪ رسیده‌است.

۶-۲-آزمایش دوم: بررسی مدل‌های از پیش آموزش‌داده‌شده بر روی کارایی تشخیص صحنه از راه‌دور

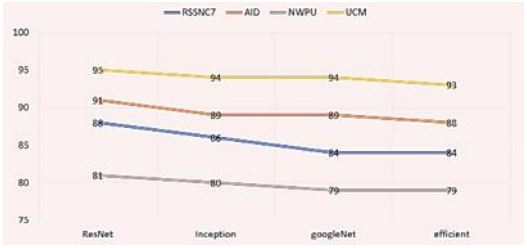
در این آزمایش به بررسی مدل‌های ازپیش آموزش‌دیده بر روی کارایی تشخیص صحنه می‌پردازیم. در این مرحله با استفاده از چهار مدل ResNet, GoogleNet, EfficientNet, InceptionNet_b0 استخراج ویژگی‌های تصاویر انجام می‌شود؛ ابتدا تمام تصاویر مجموعه‌داده‌ها به مدل‌های ازپیش آموزش‌داده‌شده وارد و برای هر تصویر هزار ویژگی استخراج می‌شود.

دقت ویژگی‌های استخراج‌شده برای مجموعه‌داده‌های مختلف با هر مدل در شکل (۱۲) نشان داده شده‌است. با توجه به نتایج به‌دست‌آمده مدل ResNet توانست ویژگی‌های دقیق‌تری نسبت به سایر مدل‌ها استخراج کند و پس از آن InceptionNet_b0 و سپس GoogleNet و در آخر EfficeintNet قراردارد.

[DOI: 10.61186/jsdp.21.3.69]



(شکل-۱۲): دقت ویژگی‌های مدل‌های ازپیش آموزش‌دیده
(Figure-12): Accuracy of features of pre-trained models



(شکل-۱۳): دقت ویژگی‌های توصیف‌گرهای محلی
(Figure-13): Accuracy of features of local descriptors

۶-۳-آزمایش سوم: بررسی اثر استفاده از توصیف‌گرهای محلی بر روی کارایی تشخیص صحنه از راه‌دور

در این آزمایش به بررسی اثر استفاده از توصیف‌گرهای محلی بر روی کارایی تشخیص صحنه ازراه‌دور می‌پردازیم؛ در این مرحله از دو الگوریتم ORB و LBP استفاده می‌کنیم. هر کدام از این دو الگوریتم، تصاویر مجموعه‌داده‌ها را دریافت و الگوریتم LBP برای هر تصویر ۴۲۰ و ORB تعداد شش‌هزار ویژگی استخراج می‌کند؛ پس از آن دقت ویژگی‌های استخراج‌شده با الگوریتم ماشین بردار پشتیبان به‌دست‌آمده که در شکل (۱۳) نمایش داده شده‌است. همان‌طورکه مشاهده می‌شود ویژگی‌های استخراج‌شده به‌وسیلهٔ الگوریتم LBP دقیق‌ترند.

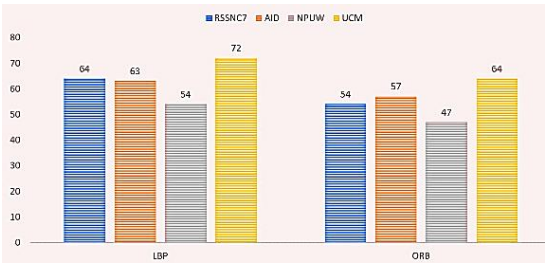
۶-۴-آزمایش چهارم: بررسی اثر استفاده از بسته‌بند و مربع کای به‌عنوان انتخاب‌گر ویژگی بر روی کارایی تشخیص صحنه ازراه‌دور

برای ادغام ویژگی‌های استخراج‌شده، ابتدا تمام ویژگی‌ها برای هر تصویر به‌صورت یک سطر درآمده‌است و سپس آرایه‌ها به هم متصل می‌شوند. در نهایت برای هر تصویر تمامی ویژگی ها در یک سطر قرار می‌گیرند. قبل از استفاده از الگوریتم‌های انتخاب‌گر ویژگی، ویژگی‌های به‌دست‌آمده از مدل بیست لایه و ویژگی‌های مدل‌های از پیش آموزش‌دیده و ویژگی‌های توصیف‌گر محلی با هم ترکیب و برای هر تصویر در یک سطر ذخیره شده‌اند.

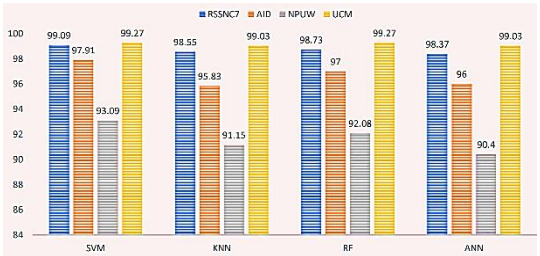
در این آزمایش به بررسی اثر استفاده از الگوریتم‌های بسته‌بند و مربع کای به‌عنوان انتخاب‌گر ویژگی برای ویژگی‌های ترکیب‌شده بر روی کارایی تشخیص صحنه ازراه‌دور می‌پردازیم. دقت و نتایج به‌دست‌آمده برای هر الگوریتم در شکل (۱۴) مشاهده می‌شود. هر دو الگوریتم، پنجاه ویژگی برتر را برای هر تصویر از ۱۵۰۰ ویژگی کلی انتخاب می‌کنند؛ دلیل انتخاب پنجاه ویژگی آن است که این تعداد، بیشترین ویژگی مؤثر را برای هر تصویر دربردارد و با تعداد کمتر یا بیشتر از این مقدار، دقت ویژگی‌های انتخاب‌شده کاهش پیدا می‌کند. برای طبقه‌بندی ویژگی‌ها نیز از روش ماشین بردار پشتیبان استفاده شده‌است. با توجه به نتایج به‌دست‌آمده می‌توان گفت که الگوریتم بسته‌بند، روش بهتری برای انتخاب ویژگی‌های تصویر برای تشخیص صحنه از راه‌دور است.

۶-۵-آزمایش پنجم: بررسی اثر استفاده از الگوریتم‌های طبقه‌بندی بر کارایی تشخیص صحنه از راه‌دور

الگوریتم‌های طبقه‌بندی، ویژگی‌های استخراج‌شده به‌وسیلهٔ انتخاب‌گر ویژگی را دسته‌بندی می‌کنند. پس از استخراج ویژگی‌ها از دو مدل و توصیف‌گر ویژگی محلی و ترکیب آن‌ها، سپس استفاده از انتخاب‌گر ویژگی و طبقه‌بندی تصاویر، خروجی نهایی بر روی چهار پایگاه‌داده به‌دست آمده‌است.



(شکل-۱۴) دقت الگوریتم‌های انتخاب‌گر ویژگی
(Figure-14) Accuracy of feature selection algorithms



(شکل-۱۵): دقت الگوریتم‌های طبقه‌بندی
(Figure-15): Accuracy of classification algorithms

در این آزمایش به بررسی دقت مدل با استفاده از چهار الگوریتم طبقه‌بندی ماشین بردار پشتیبان[1]، جنگل

[1] Support Vector Machine

[DOI: 10.61186/jsdp.21.3.69]

تصادفی[1]، شبکهٔ عصبی مصنوعی آبشاری[2] و k نزدیک‌ترین همسایه[3] می‌پردازیم. نتایج ارزیابی در شکل (۱۵) نمایش داده شده‌است.

به‌منظور نمایش خروجی روش پیشنهادی، برای هر پایگاه‌داده، ترکیبی که بیشترین دقت را دارد آورده شده‌است. طبق نتایج، الگوریتم ماشین بردار پشتیبان، بالاترین دقت را در بین طبقه‌بندها دارد.

۷-ارزیابی روش پیشنهادی

به‌منظور ارزیابی راه‌کار پیشنهادی، عملکرد آن با تعدادی پژوهش مشابه پیشین مقایسه شد. در بیشتر روش‌هایی که تاکنون ارائه شده‌اند، تنها از ویژگی‌های سراسری عمیق استخراج‌شده به‌وسیلهٔ شبکه‌های کانولوشن استفاده شده‌است؛ در واقع بیشتر پژوهش‌ها مبتنی بر یک نوع خاص ویژگی مثل سراسری و یا محلی بوده‌اند و برای رسیدن به دانش کافی از تصاویر، بیشتر از شبکه‌های کانولوشن پیش‌آموزش‌دیده استفاده کرده‌اند. در این حالت اگر عمق شبکه به اندازهٔ کافی مناسب نباشد ویژگی‌های استخراج‌شده قدرت افتراق بالایی ندارند و درنتیجه تشخیص صحنه به‌خوبی انجام نمی‌شود؛ ازاین‌رو، در این پژوهش از ترکیب ویژگی‌های عمیق سراسری که از شبکه‌های کانولوشن استخراج می‌شود در کنار ویژگی‌های محلی دستی و عمیق که از توصیف‌گرهای محلی به‌دست می‌آید، استفاده شده‌است؛ همچنین در این پژوهش، یک مرحله انتخاب ویژگی ترکیبی مبتنی بر روش‌های مربع کای و بسته‌بند نیز برای رسیدن به ویژگی‌های مناسب در مدل وجود دارد. مطالعات نشان می‌دهد تاکنون از مدل ترکیبی بالا برای تشخیص صحنه در سنجش از دور استفاده نشده‌است.

(جدول-۵): درصد دقت برای سنجش تصاویر ازراه‌دور برای پایگاه دادهٔ RSSNC7 (نسبت تمرین ۵۰٪)

(Table-5): Percentage of accuracy for remote sensing images for RSSNC7 database (training ratio 50%)

روش	دقت
BoVW(LBP) [28]	۸۱.۶۹
BoVW(SIFT) [28]	۸۱.۳۴
VGG19+WSPM-CRC [29]	۹۱.۳۱
Resnet+Hybrid-KCRC [30]	۹۲.۶۲
Two-stage-deep-feature [31]	۹۲.۳۷
DFAGCN [14]	۹۴.۱۴
Proposed Approach	۹۹.۰۹

نتایج ارزیابی‌ها روی چهار مجموعه‌دادهٔ RSSNC7، UCM، NWPU و AID به‌ترتیب در جدول‌های (۵) تا (۸) آمده‌است. با توجه به نتایج نشان داده‌شده، راه‌کار پیشنهادی برای تمام مجموعه‌داده‌ها دقت بالاتری نسبت به روش‌های مورد مقایسه دارد.

[1] Random Forest
[2] Cascade ANN
[3] KNN

(جدول-۶): درصد دقت برای سنجش تصاویر ازراه‌دور برای پایگاه‌دادهٔ UCM (نسبت تمرین ۸۰٪)

(Table-6): Percentage of accuracy for remote sensing images for UCM database (training ratio 80%)

روش	دقت
BoVW(LBP) [28]	۷۷.۱۲
BoVW(SIFT) [28]	۷۴.۱۲
VGG19+WSPM-CRC [29]	۹۶.۱۴
Resnet+Hybrid-KCRC [30]	۹۷.۲۹
GoogleNet [32]	۹۴.۳۱
VGG-VD-16 [29]	۹۵.۲۱
VGG-M+IFK [28]	.۹۶.۹
MLDS [26]	۹۷.۷۸
DFAGCN [14]	۹۸.۴۸
Proposed Approach	۹۹.۲۷

(جدول-۷): درصد دقت برای سنجش تصاویر ازراه‌دور برای پایگاه‌دادهٔ NWPU (نسبت تمرین ۲۰٪)

(Table-7): Percentage of accuracy for remote sensing images for NWPU database (training ratio 20%)

روش	دقت
BoVW(LBP) [28]	۴۴.۹۷
BoVW(SIFT) [28]	۳۲.۹۶
AlexNet+SAFF [33]	۸۴.۰۰
VGG-VD16+SAFF [33]	۸۷.۸۶
AlexNet [12]	۷۹.۸۵
VGGNET [12]	۷۹.۷۹
GoogleNET [32]	۷۸.۴۸
MIDV-Net_CS [33]	۸۷.۹۹
RADC-Net [34]	۸۷.۶۳
DFAGCN [14]	۸۹.۲۹
Proposed Approach	۹۳۰۹

(جدول-۸): درصد دقت برای سنجش تصاویر ازراه‌دور برای پایگاه‌دادهٔ AID (نسبت تمرین ۵۰٪)

(Table-8): Percentage of accuracy for remote sensing images for AID database (training ratio 50%)

روش	دقت
BoVW(LBP) [28]	۶۴.۲۵
BoVW(SIFT) [28]	۶۷.۶۵
VGG19+WSPM-CRC [29]	۹۲.۵۷
Resnet+Hybrid-KCRC [30]	۹۲.۸۷
CaffeNet [31]	۸۹.۵۲
VGG-VD-16 [31]	۸۹.۶۴
GoogleNET [31]	۸۷.۳۹
VGG-VD16+SAFF [33]	۹۳.۸۳
MF2NET [35]	۹۴.۸۴
VGG-16-CapsNet [36]	۹۴.۷۱
DFAGCN [14]	۹۴.۸۸
Proposed Approach	۹۷.۹۱

[DOI: 10.61186/jsdp.21.3.69]

در نهایت، مقایسه‌ای بین استخراج‌کنندهٔ ویژگی محلی LBP، مدل از پیش تمرین‌داده‌شده Resnet، شبکهٔ بیست لایهٔ تمام‌متصل با راه‌کار پیشنهادی، شامل ترکیب این سه مدل، انجام شده‌است. نتیجه در جدول (۹) آمده و برای هر چهار مجموعه‌داده دقت راه‌کار پیشنهادی مقدار بهتری را نشان می‌دهد. دلیل این برتری، استفاده از ترکیب ویژگی‌های محلی و سراسری استخراج‌شده و معماری مدل بیست لایهٔ ازپیش آموزش‌دیده است؛ همچنین برای استخراج ویژگی‌ها از مدل‌های ResNet101,CNN20layer,LBP و برای انتخاب ویژگی‌های ترکیب‌شده، از الگوریتم بسته‌بند استفاده شده و طبقه‌بندی برعهدهٔ ماشین بردار پشتیبان بوده‌است.

(جدول-۹): مقایسهٔ بین مدل‌های مختلف
(Table-9): Comparison between different models

پایگاه‌داده	روش	دقت
RSSNC7	LBP	۶۴.۳۳
	Pre-Train	۸۸.۶۸
	20CNN	۸۹.۵۰
	Proposed Approach	۹۹.۰۹
NWPU	LBP	۵۴.۳۵
	Pre-Train	۸۱.۹۳
	20CNN	۸۴.۶۸
	Proposed Approach	۹۳.۰۹
UCM	LBP	۷۲.۲۵
	Pre-Train	۹۵.۶۳
	20CNN	۹۷.۲۱
	Proposed Approach	۹۹.۲۷
AID	LBP	۶۳.۳۸
	Pre-Train	۹۱.۱۲
	20CNN	۹۱.۳۳
	Proposed Approach	۹۴.۳۱

۸-نتیجه‌گیری و پیشنهادهای آینده

در حال حاضر، روش‌هایی برای سنجش تصاویر ازراه‌دور ارائه شده که مبتنی بر استفاده از روش‌های یادگیری عمیق و شبکه‌های ازپیش آموزش‌دیده‌اند. در روش پیشنهادی که بر مبنای ترکیب دو نوع ویژگی عمیق و یک نوع ویژگی محلی است، با استخراج و ترکیب این ویژگی‌ها و سپس ادغام آن‌ها مهم‌ترین ویژگی‌های هر تصویر با استفاده از الگوریتم‌های انتخاب‌گر ویژگی انتخاب می‌شود. ارزیابی این راه‌کار نشان داد که معماری پیشنهادی به‌همراه استفاده از مدلی که با مجموعه‌دادهٔ هدف آموزش دیده باشد باعث افزایش عملکرد در تشخیص صحنه در سنجش‌ازراه‌دور می‌شود.

به‌منظور توسعهٔ پژوهش جاری می‌توان ترکیب ویژگی‌های استخراج‌شده از مدل دستی کانولوشنی بیست لایه و پیش‌آموزش این مدل با تصاویر مجموعه‌دادهٔ هدف را که باعث استخراج ویژگی‌های بهتر و دقیق‌تری از تصاویر می‌شود، درنظرگرفت.

۹-مراجع 9-Reference

[1] Q. Hu et al., "Exploring the use of Google Earth imagery and object-based methods in land use/cover mapping," Remote Sensing, vol. 5, no. 11, pp. 6026-6042, 2013.

[2] L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," Proceedings of the IEEE, vol. 103, no. 9, pp. 1560-1584, 2015.

[3] A. Tayyebi, B. C. Pijanowski, and A. H. Tayyebi, "An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran," Landscape and Urban Planning, vol. 100, no. 1-2, pp. 35-44, 2011.

[4] Z. Y. Lv, W. Shi, X. Zhang, and J. A. Benediktsson, "Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation," IEEE journal of selected topics in applied earth observations and remote sensing, vol. 11, no. 5, pp. 1520-1532, 2018.

[5] T. R. Martha, N. Kerle, C. J. Van Westen, V. Jetten, and K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," IEEE transactions on geoscience and remote sensing, vol. 49, no. 12, pp. 4928-4943, 2011.

[6] F. Ghazouani, I. R. Farah, and B. Solaiman, "A multi-level semantic scene interpretation strategy for change interpretation in remote sensing imagery," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 11, pp. 8775-8795, 2019.

[7] Y. Li, Y. Zhang, X. Huang, and A. L. Yuille, "Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images," ISPRS journal of photogrammetry and remote sensing, vol. 146, pp. 182-196, 2018.

[8] Y. Gu, Y. Wang, and Y. Li, "A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection," Applied Sciences, vol. 9, no. 10, p. 2110, 2019.

[9] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 3735-3756, 2020.

[DOI: 10.61186/jsdp.21.3.69]

[10] W. Wang, Y. Chen, and P. Ghamisi, "Transferring CNN With Adaptive Learning for Remote Sensing Scene Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-18, 2022.
[11] K. Xu, H. Huang, P. Deng, and Y. Li, "Deep feature aggregation framework driven by graph convolutional network for scene classification in remote sensing," IEEE Transactions on Neural Networks and Learning Systems, 2021.
[12] S. Aggarwal, "Principles of remote sensing," Satellite remote sensing and GIS applications in agricultural meteorology, vol. 23, no. 2, pp. 23-28, 2004.
[13] M. a. Ahmadi and R. Dianat, "Introducing a method for extracting features from facial images based on applying transformations to features obtained from convolutional neural networks," (in eng), Signal and Data Processing, Research vol. 17, no. 3, pp. 141-156, 2020.
[14] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artificial intelligence review, vol. 53, no. 8, pp. 5455-5516, 2020.
[15] M. A. Zare Chahooki and z. khalifeh zadeh, "A General Investigation on the Combination of Local and Global Feature Selection Methods for Request Identification on Telegram," (in eng), Signal and Data Processing, Applicable vol. 19, no. 2, pp. 175-196, 2022.
[16] M. Imani and H. Ghassemian, "Supervised Feature Extraction of Face Images for Improvement of Recognition Accuracy," (in eng), Signal and Data Processing, Research vol. 16, no. 1, pp. 158-172, 2019.
[17] J. Zhang, C. Lu, X. Li, H.-J. Kim, and J. Wang, "A full convolutional network based on DenseNet for remote sensing scene classification," Mathematical Biosciences and Engineering, vol. 16, no. 5, pp. 3345-3367, 2019.
[18] H. Sun, S. Li, X. Zheng, and X. Lu, "Remote sensing scene classification by gated bidirectional network," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 1, pp. 82-96, 2019.
[19] J. Xie, N. He, L. Fang, and A. Plaza, "Scale-free convolutional neural network for remote sensing scene classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6916-6928, 2019.
[20] C. Shi, T. Wang, and L. Wang, "Branch feature fusion convolution network for remote sensing scene classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp 5194-5210, 2020.
[21] S.-C. Hung, H.-C. Wu, and M.-H. Tseng, "Remote sensing scene classification and explanation using RSSCNet and LIME," Applied Sciences, vol. 10, no. 18, p. 6151, 2020.
[22] X. Tang, Q. Ma, X. Zhang, F. Liu, J. Ma, and L. Jiao, "Attention consistent network for remote sensing scene classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2030-2045, 2021.
[23] R. M. Anwer, F. S. Khan, and J. Laaksonen, "Compact deep color features for remote sensing scene classification," Neural Processing Letters, vol. 53, no. 2, pp. 1523-1544, 2021.
[24] T. Tian, L. Li, W. Chen, and H. Zhou, "SEMSDNet: A multiscale dense network with attention for remote sensing scene classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5501-5514, 2021.
[25] Q. Bi, H. Zhang, and K. Qin, "Multi-scale stacking attention pooling for remote sensing scene classification," Neurocomputing, vol. 436, pp. 147-161. 2021.
[26] S. Mei, K. Yan, M. Ma, X. Chen, S. Zhang, and Q. Du, "Remote sensing scene classification using sparse representation-based framework with deep feature fusion," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5867-5878, 2021.
[27] J. Shen, T. Zhang, Y. Wang, R. Wang, Q. Wang, and M. Qi, "A Dual-Model Architecture with Grouping-Attention-Fusion for Remote Sensing Scene Classification," Remote Sensing, vol. 13, no. 3, p. 433, 2021.
[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255: Ieee.
[29] B.-D. Liu, J. Meng, W.-Y. Xie, S. Shao, Y. Li, and Y. Wang, "Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification," Remote Sensing, vol. 11, no. 5, p. 518, 2019.
[30] B.-D. Liu, W.-Y. Xie, J. Meng, Y. Li, and Y. Wang, "Hybrid collaborative representation for remote-sensing image scene classification," Remote Sensing, vol. 10, no. 12, p. 1934, 2018.
[31] Y. Liu, Y. Liu, and L. Ding, "Scene classification based on two-stage deep feature fusion," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 2, pp. 183-186, 2017.
[32] T. Gong, X. Zheng, and X. Lu, "Remote Sensing Scene Classification with Multi-task Learning," in Proceedings of the 7th China High Resolution Earth Observation Conference (CHREOC 2020), 2022, pp. 403-418: Springer.
[33] R. Cao, L. Fang, T. Lu, and N. He, "Self-attention-based deep feature fusion for remote sensing scene classification," IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 1, pp. 43-47, 2020.

[DOI: 10.61186/jsdp.21.3.69]

[34] Q. Bi, K. Qin, H. Zhang, Z. Li, and K. Xu, "RADC-Net: A residual attention based convolution network for aerial scene classification," Neurocomputing, vol. 377, pp. 345-359, 2020.
[35] K. Xu, H. Huang, Y. Li, and G. Shi, "Multilayer feature fusion network for scene classification in remote sensing," IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 11, pp. 1894-1898, 2020.
[36] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-CapsNet," Remote Sensing, vol. 11, no. 5, p. 494, 2019.

سید علی جهانی دارای مدرک کارشناسی‌ارشد رشتهٔ مهندسی کامپیوتر گرایش هوش مصنوعی است. زمینهٔ اصلی فعالیت و موضوعات پژوهشی مورد علاقهٔ ایشان، هوش مصنوعی، پردازش تصویر و یادگیری عمیق است.
نشانی رایانامهٔ ایشان عبارت‌است از:
Alijahani1919@gmail.com

کیوان محبی استادیار گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان) و دارای مقالات متعددی در کنفرانس‌های ملی و بین‌المللی و نشریات معتبر علمی است و عضویت در کمیتهٔ علمی و داوری کنفرانس‌ها و نشریات علمی را در کارنامهٔ خود دارد. موضوعات پژوهشی مورد علاقهٔ ایشان عبارت‌اند از: مهندسی نرم‌افزار، معماری نرم‌افزار، آزمون نرم‌افزار، وب معنایی و شبکه‌های اجتماعی.
نشانی رایانامهٔ ایشان عبارت‌است از:
K.mohebbi@khuisf.ac.ir

فرساد زمانی بروجنی در سال ۲۰۱۳ مدرک دکترای خود را در رشتهٔ کامپیوتر از دانشگاه پوترا مالزی دریافت کرد. ایشان هم‌اکنون استادیار گروه مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران است. زمینه‌های پژوهشی مورد علاقهٔ ایشان داده‌کاوی، یادگیری ماشین و تجزیه و تحلیل داده‌های حجیم است.
نشانی رایانامهٔ ایشان عبارت‌است از:
Farsad.zamani@srbiau.ac.ir

[DOI: 10.61186/jsdp.21.3.69]