



ORIGINAL RESEARCH PAPER

Large Scale Crop Type Mapping Using Multi Temporal Sentinel-2 Imagery and a New Attention U-Net-Vit

M. HeidariGholanlo¹, R. Javanmard Alitappeh^{*1}, E. Ghanbari Parmehr²¹ Dept. of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran² Dept. of Geomatics Engineering, Babol Noshirvani University of Technology, Babol, Iran

ABSTRACT

Received: 03 July 2024
Reviewed: 19 September 2024
Revised: 30 September 2024
Accepted: 01 November 2024

KEYWORDS:

Remote Sensing
Deep Learning
Satellite Image Processing
Agricultural Products Detection

* Corresponding author

✉ Rezajavanmard@mazust.ac.ir

☎ (+98 11) 34150000

Background and Objectives: With the advancement of technology and the emergence of multifunctional satellites, a significant amount of real time information is now transmitted from the Earth's surface. These satellites are equipped with sensors that obtain critical information by sending signals at various frequencies to the Earth's surface. The data received from these satellites are utilized in various scientific and military applications, including aviation, geographical studies, meteorology, agriculture, and other research fields. The agricultural sector and crop monitoring have especially benefited from remote sensing methods compared to traditional methods, becoming a primary tool for collecting environmental information for area monitoring applications, attracting researchers' attention. One such application is regional monitoring to examine agricultural products over cultivated areas. The use of remote sensing tools and satellite images is highly efficient due to their wide regional coverage. To automatically analyze these images, classify, and segment cultivated areas, machine learning methods are currently employed. Among these methods, deep learning offers superior performance and higher speed compared to other learning methods, such as manual or semi-automatic methods.

Methods: This paper utilizes deep learning models suitable for segmenting agricultural areas. Generally, these models produce an output of equivalent dimensions for each input. Therefore, for working with satellite images, an improved U-Net model is proposed in this study. The proposed model is developed using Vision Transformers (ViT) in the model's bottleneck for classifying and segmenting four types of agricultural products: rice, wheat, canola, and corn. Compared to convolutional layers, ViT is more efficient in terms of conceptual and algorithmic implementation and requires less computational power. This model addresses the problems and weaknesses of the base U-Net model that arise with complex datasets, diverse in shape, size, and texture, enabling more accurate and reliable segmentation results. Additionally, the proposed improvements enhance the model's robustness and adaptability to various agricultural scenarios.

Findings: In the numerous experiments conducted, the proposed method achieved an accuracy of 83.84 and Precision of 70.69%, providing a better classification of the five target products compared to other methods. The qualitative outputs also indicate better segmentation of the input images when applying the proposed method. Alongside the accuracy metric, other metrics such as focal loss, recall, precision, and MIoU were examined, with the proposed method reaching acceptable levels in most cases. Notably, as the target area was in Iran, data collection and labeling were also carried out in this research, providing a suitable dataset for training other models

Conclusion: This research presents an end-to-end model for learning features related to the segmentation of satellite images. The results indicate that the proposed method can be effectively used for segmenting satellite images received from Sentinel-2 for the target products, such as various crops. Therefore, the results obtained can play a crucial role in water consumption management, planting structure adjustment, loss estimation, and agricultural performance evaluation, providing significant insights for stakeholders. By utilizing these methods, it is possible to achieve improved efficiency and accuracy in agricultural management and optimize resource use in this area, contributing to sustainable agricultural practices and better decision-making processes.



NUMBER OF REFERENCES

52



NUMBER OF FIGURES

8



NUMBER OF TABLES

3

مقاله پژوهشی

نقشه‌برداری نوع محصول در مقیاس بزرگ با استفاده از تصاویر چندزمانی سنتینل-۲ با مدل Attention U-Net-Vit جدید

مجید حیدری قولانلو^۱، رضا جوانمرد علی تپه^{۱*}، عبادت قنبری پرمهر^۲^۱ دانشکده مهندسی برق و کامپیوتر، دانشگاه علم و فناوری مازندران، بهشهر، ایران^۲ دانشکده مهندسی عمران، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

چکیده

پیشینه و اهداف: با پیشرفت تکنولوژی و پیدایش ماهواره‌های چندمنظوره، اطلاعات لحظه‌ای زیادی از سطح زمین مخابره می‌شود. ماهواره‌ها به سنجده‌هایی مجهز هستند که می‌توانند با ارسال سیگنالهایی در فرکانسهای مختلف به سطح زمین به اطلاعات مهمی دست یابند. داده‌های دریافتی از این ماهواره‌ها در کاربردهای مختلف علمی و نظامی از جمله: هوانوردی، مطالعات جغرافیایی، هواشناسی، کشاورزی و دیگر حوزه‌های تحقیقاتی قابل استفاده است. حوزه‌ی کشاورزی و پایش سطوح کشت نیز یکی از حوزه‌هایی است که با توجه به مزیت‌های روش‌های سنجش‌ازدور در مقایسه با روش‌های سنتی، به‌عنوان یکی از ابزارهای اصلی در جمع‌آوری اطلاعات محیطی برای کاربردهای پایش نواحی، مورد توجه پژوهشگران قرار گرفته است. یکی از این موضوعات، پایش منطقه‌ای برای بررسی محصولات کشاورزی در مساحت سطح زیرکشت است که استفاده از ابزارهای سنجش‌ازدور و تصاویر ماهواره‌ای به جهت پوشش منطقه‌ای وسیع بسیار کارا است. جهت بررسی خودکار این تصاویر، طبقه‌بندی و بخش‌بندی نواحی سطح زیرکشت، امروزه از روش‌های یادگیری ماشین استفاده می‌شود. در میان این روش‌ها، یادگیری عمیق در مقایسه با دیگر روش‌های یادگیری مانند روش‌های دستی و یا روش‌های نیمه‌خودکار، عملکرد بهتر و سرعت بالاتری دارد.

روش‌ها: در این مقاله مدل‌های یادگیری عمیق که برای بخش‌بندی نواحی مناسب هستند مورد استفاده قرار گرفته است. عموماً این مدلها بازای هر ورودی، خروجی معادل آن را با همان ابعاد تولید می‌کنند. لذا جهت کار بر روی تصاویر ماهواره‌ای، در این پژوهش مدل U-Net بهبود یافته‌ای ارائه شده است. مدل پیشنهادی با استفاده از ViT در گلوگاه مدل برای طبقه‌بندی و بخش‌بندی چهار نوع محصول کشاورزی شامل برنج، گندم، کلزا و ذرت توسعه داده شده است. استفاده از ViT در مقایسه با لایه‌های کانولوشن از لحاظ ایده و پیاده‌سازی الگوریتمی کارا تر است و حجم محاسباتی کمتری دارد. این مدل مشکلات و نقاط ضعف مدل پایه U-Net را که برای مجموعه داده‌های پیچیده، متنوع در شکل، اندازه و بافت به وجود می‌آید، برطرف می‌نماید.

یافته‌ها: در نتایج آزمایشات انجام شده روش پیشنهادی توانسته است با رسیدن به دقت ۸۳.۸۴ و صحت ۷۰.۶۹، بهتر از دیگر روش‌ها دسته‌بندی درستی از ۵ محصول مورد نظر را ارائه دهد. همچنین خروجی‌های کیفی نیز نشان‌دهنده‌ی بخش‌بندی بهتر تصاویر ورودی با اعمال روش پیشنهادی می‌باشد. در کنار معیار دقت، دیگر معیارها مانند افت کانونی، بازیابی و MIoU نیز مورد بررسی قرار گرفت که در اکثر موارد روش پیشنهادی به مقدار قابل قبولی رسیده است. لازم به ذکر است که با توجه به اینکه منطقه‌ی مورد نظر در ایران در نظر گرفته شد، جمع‌آوری و برچسب‌گذاری داده‌ها نیز در این پژوهش انجام شده است که می‌تواند بعنوان مجموعه داده‌ی مناسبی برای آموزش دیگر مدلها استفاده شود.

نتیجه‌گیری: این تحقیق یک مدل سرتاسری برای یادگیری ویژگی‌های مرتبط با بخش‌بندی تصاویر ماهواره‌ای ارائه داده است. نتایج این تحقیق نشان می‌دهد که روش ارائه شده می‌تواند برای بخش‌بندی تصاویر ماهواره‌ای دریافتی از سنتینل-۲ برای محصولات مورد نظر مورد استفاده قرار گیرد. لذا نتایج حاصل می‌تواند در مدیریت مصرف آب، تنظیم ساختار کاشت، تخمین تلفات و ارزیابی عملکردهای زراعی نقش مهمی را ایفا نماید. با بهره‌گیری از این روش‌ها، می‌توان به بهبود کارایی و دقت در مدیریت کشاورزی دست یافت و از منابع این حوزه بهینه‌تر استفاده کرد.

تاریخ دریافت: ۱۳ تیر ۱۴۰۳
تاریخ داوری: ۲۹ شهریور ۱۴۰۳
تاریخ اصلاح: ۰۹ مهر ۱۴۰۳
تاریخ پذیرش: ۱۱ آبان ۱۴۰۳

واژگان کلیدی:

سنجش از دور
یادگیری عمیق
پردازش تصاویر ماهواره‌ای
تشخیص محصولات کشاورزی

* نویسنده مسئول

Rezajavanmard@mazust.ac.ir

۰۱۱-۳۴۱۵۰۰۰۰

مقدمه

در دنیای امروز حوزه‌ی کشاورزی یکی از بزرگترین مصرف‌کنندگان منابع آبی در سطح جهانی است [۲ و ۱] و بیش از ۷۰ درصد از آب سطحی و زیرزمینی برای مصارف کشاورزی استفاده می‌شود. با توجه به افزایش جمعیت و تغییرات اقلیمی، موضوع کمبود آب به یک معضل جهانی تبدیل شده است. از سوی دیگر رشد بی‌رویه مناطق شهری باعث کاهش زمین‌های زراعی شده است که این موضوع امنیت غذایی دنیا را تهدید می‌کند [۳ و ۴]. در این شرایط، مدیریت کارآمد مصرف آب در بخش کشاورزی به یکی از اولویت‌های اصلی تبدیل شده است. انتخاب روش‌های مناسب آبیاری، استفاده بهینه از منابع آب و مدیریت کاشت و برداشت محصولات، همگی نقش مهمی در کاهش مصرف آب و افزایش بهره‌وری آن دارند.

یکی از ابزارهای پیشرفته‌ای که به مدیریت مصرف آب و بهینه‌سازی فرآیندهای کشاورزی کمک می‌کند، استفاده از تصاویر ماهواره‌ای است [۵ و ۶]. تصاویر ماهواره‌ای امکان رصد و پایش مداوم و دقیق اراضی کشاورزی را فراهم می‌کنند و اطلاعات ارزشمندی درباره وضعیت خاک، رشد محصولات، نیاز آبی گیاهان، و تغییرات اقلیمی ارائه می‌دهند [۷ و ۸]. با تحلیل این تصاویر، کشاورزان می‌توانند تصمیمات بهتری درباره زمان و میزان آبیاری، کوددهی و برداشت محصولات بگیرند. همچنین، این داده‌ها به شناسایی مناطق دارای تنش آبی یا آفت که نیاز به مدیریت ویژه دارند کمک می‌کنند. لذا در این مقاله، تحلیل تصاویر ماهواره‌ای برای تشخیص و دسته‌بندی نوع محصولات کشاورزی مطالعه شده است. نتایج حاصل از روش پیشنهادی در بهبود بهره‌وری منابع آبی، کاهش هزینه‌های تولید، و افزایش بازده محصولات کشاورزی و در نهایت به حفظ منابع طبیعی و پایداری کشاورزی منجر می‌شوند.

تحلیل تصاویر ماهواره‌ای دقیق برای تشخیص نوع محصول به مجموعه داده‌های جامع و دقیق نیاز دارد. جمع‌آوری تصاویر از مناطق کشت با استفاده از ماهواره سنتینل-۲ امکان‌پذیر است، که تصاویر با وضوح طیفی، مکانی، و زمانی مناسب را برای شناسایی و تحلیل دقیق‌تر محصولات فراهم می‌کند [۹ و ۱۰]. با این حال، نقشه‌برداری نوع محصول با استفاده از تصاویر ماهواره‌ای به دلیل چرخه‌های سریع و متنوع فنولوژی، شیوه‌های مختلف کشاورزی، و تغییرات زمانی و طیفی زیاد با چالش‌هایی مانند عوامل محیطی در مشاهدات ماهواره‌ای، تنوع زمان برداشت محصولات، نیاز به مشاهدات میدانی، داده‌های عظیم ماهواره‌ها همراه است [۱۱].

با توجه به اهمیت موضوع این مقاله روش‌های مختلفی جهت شناسایی نوع محصولات کشاورزی از روی تصاویر ماهواره‌ای ارائه شده است. از جمله این روش‌ها می‌توان به روش‌های سنتی یادگیری ماشین از جمله: ماشین بردار پشتیبان [۱۶]، جنگل تصادفی [۱۴ و ۱۷] یا به روش‌های یادگیری عمیق مانند شبکه‌های حافظه طولانی کوتاه مدت [۱۸] و شبکه‌های عصبی پیچشی اشاره کرد. روش‌های یادگیری

عمیق دقت نگاشت نوع محصول را با استفاده از ویژگی‌های طیفی افزایش می‌دهند [۱۳ و ۱۵ و ۱۹]. با این حال، آموزش مجموعه داده‌هایی با مقیاس بزرگ چالش‌برانگیز است. اکثر مطالعات نقشه‌برداری محصولات کشاورزی بر روی روش‌های رایج یادگیری ماشین سنتی متمرکز شده‌اند. معمولاً روش‌های یادگیری به دلیل عوامل متعددی مانند شرایط آب و هوایی و نوسانات زمان کاشت، دقت بالایی را ارائه نمی‌دهند. برخلاف روش‌های قبلی، بخش‌بندی مبتنی بر پیکسل/بخش روشی است که یک تصویر سنجش‌ازدور را بر اساس بخش‌های تصویر بخش‌بندی می‌کند. بخش‌بندی فرآیند تعریف پیکسل‌های همگن در بخش‌های تصویری مشابه از یک طیف است [۲۰]. در این مقاله برای دسته‌بندی از روش‌های یادگیری عمیق استفاده شده است؛ که بخش‌بندی نقشه‌ها را براساس محصولات کاشت شده نشان می‌دهد. لذا علاوه بر به کار بردن روش‌های معمول که در مدل U-Net برای بخش‌بندی معنایی استفاده شده است؛ با اضافه کردن شبکه ترنسفورمر بینایی [۲۱]، در گلوگاه مدل، و مکانیسم توجه در اتصالات برش باعث بهبود طبقه‌بندی و شناسایی محصولات کشاورزی شده است. مدل پیشنهادی توانسته است چهار محصول اصلی و یا هشت کلاس تعریف شده را دسته‌بندی نموده و نواحی هر دسته را به صورت پیکسلی تفکیک کند.

هدف این پژوهش ارائه یک مدل بهبودیافته از U-Net برای تهیه نقشه دقیق انواع مختلف محصولات کشاورزی از تصاویر در مقیاس بزرگ با استفاده از تصاویر چند زمانی از ماهواره سنتینل-۲ از نه استان کشور است. بهبودها شامل دو جنبه اصلی است. برای جمع‌آوری داده ابتدا تصاویر ماهواره‌ای با مقیاس اصلی از Google Earth Engine (GEE) دریافت شده است. سپس با اعمال پیش پردازش‌هایی مانند حذف ابر و تاثیرمیزان تابش خورشید، از ابزار Q-GIS برای تقسیم بندی تصاویر بزرگ به ابعاد 256×256 استفاده شده است. در ادامه نیز برای حاشیه نویسی (رنگ بندی محصولات و پس زمینه) از ابزار QGIS استفاده شده است. در نهایت نیز تکنیک‌های داده‌افزایی جهت افزایش تعداد نمونه‌ها اعمال شده‌اند. سپس برای بهره‌برداری از اطلاعات طیفی، زمانی و مکانی داده‌های جمع‌آوری شده، مدل بخش‌بندی معنایی جهت طبقه بندی محصولات کشاورزی معرفی می‌شود. که شامل یک ماژول ViT در گلوگاه مدل اصلی Attention U-Net در جهت بهینه سازی فرایند آموزش و کاهش مشکل گرادیان ناپدید شونده می باشد. در نهایت عملکرد مدل با معیارهای مختلف از جمله Dice coefficient که مبتنی بر پیکسل است، و همچنین دقت، صحت و MIOU ارزیابی می‌شود تا دقت نقشه‌برداری محصولات کشاورزی را بهبود بخشد. با استفاده از روش پیشنهادی، بخش‌بندی معنایی تصاویر ماهواره‌ای تسهیل شده است. لازم به ذکر است که برخلاف روش‌های قبلی که برای یک محدوده خاص طراحی شده و سپس به نواحی دیگر تعمیم داده می‌شدند، در این مقاله توزیع داده‌ها در سطح کشور انجام شده است.

کارهای مرتبط

کاربردهای سنجش‌ازدور در ابتدا با مدل‌های آماری ارائه شده‌اند و با پیشرفت روش‌های یادگیری‌ماشینی باعث بهبود طبقه‌بندی و شناسایی در کاربردهای مختلف به کمک تصاویر بدست آمده در تصاویر ماهواره‌ای شده‌اند. طبقه‌بندی با استفاده از تصاویر فراطیفی سنجش‌ازدور با روش‌های یادگیری‌ماشین سنتی مانند تحلیل مؤلفه اصلی و بردار ماشین پشتیبان به مرور جای خود را به روش‌های یادگیری عمیق داده است. اخیراً تکنیک‌های یادگیری عمیق به‌عنوان یک روش مفید برای نقشه‌برداری پوشش زمین از طریق طبقه‌بندی تصاویر سنجش‌ازدور نشان داده شده است. یکی از کاربردهای سنجش‌ازدور شناسایی و طبقه‌بندی در محصولات کشاورزی می‌باشد، که در ادامه در جدول ۱ به جدیدترین روش‌ها اشاره می‌شود خلاصه جدیدترین مدل‌های یادگیری عمیق که به بخش‌بندی و یا طبقه‌بندی محصولات کشاورزی و یا به هر دو پرداخته‌اند و همه آن‌ها از مجموعه دیتاست ماهواره سنتینل-۲ استفاده کرده‌اند؛ آورده شده است.

در میان داده‌های مختلف، روش فضایی، طیفی و زمانی بالاترین نتایج طبقه‌بندی را به دست آورده‌اند؛ و نسبت به سایر روش‌های آزمایش شده مانند مکانی و رادیومتریک بهتر عمل کرده‌اند [۲۲]. بر این اساس مجموعه داده‌هایی که در این پژوهش به کار می‌روند، ویژگی‌های فضایی، طیفی و زمانی را دارند. در روش‌های یادگیری عمیق از معماری‌هایی مانند شبکه عصبی پیچشی و شبکه‌های عصبی بازگشتی تکامل یافته استفاده شده است؛ که عملکرد بالایی را در کاربردهای گوناگون از جمله بخش‌بندی معنایی اشیاء داشته‌اند [۲۴ و ۲۳]. مطالعات نشان می‌دهند که طبقه‌بندی محصولات کشاورزی با استفاده از بخش‌بندی معنایی به کمک روش‌های یادگیری عمیق و داده‌های سنجش از دور، ابزار مهم برای تصمیم‌گیری در شناسایی محصولات کشاورزی در سراسر جهان است [۲۵]. عاصم و همکاران در تحقیقی روشی را برای تشخیص نوع محصولات کشاورزی با استفاده از تصاویر ماهواره‌ای ارائه داده‌اند، که با استفاده از تصاویر چندزمانی به عنوان یک تصویر واحد در نظر گرفته و برای بخش‌بندی محصولات، ترکیبات مناسبی از ۱۰ باند ماهواره سنتینل-۲ را انتخاب کرده‌اند.

معماری [۴۸] TransU-Net بر روی داده‌های شش محصول مختلف ارزیابی شده است و دقت ۸۵٪ را در طبقه‌بندی انواع مختلف محصولات داشته است در این مطالعه ابعاد تصویر ۶۴×۶۴ پیکسل در نظر گرفته شده است [۲۶]. در مطالعه‌ای دیگری که توسط آقای سیدی و همکارانش در چارچوب جدیدی را با ترکیب شبکه عصبی پیچشی و یک مکانیسم توجه دوگانه، برای طبقه‌بندی دقیق محصولات با استفاده از داده‌های سری زمانی سنتینل-۲ پیشنهاد شد [۲۷]. ایده اصلی پشت مکانیسم توجه این است که تمرکز شبکه را بر استخراج ویژگی‌های معنی‌دار به جای ویژگی‌های غیرضروری هدایت می‌کند [۲۸]. در این مطالعه شش نوع مختلف زراعی (یونجه، لوبیا، گندم، جو، کلزا و باغ) و سه کلاس غیر زراعی (توده، بایر و آب) مورد بررسی قرار گرفته است. که بهترین نتیجه مربوط به محصول یونجه با دقت ۹۸.۷٪ بوده است. یکی از دلایل نتایج خوب در این مطالعه داشتن نمونه‌های زیاد و بررسی در مدت زمان طولانی می‌باشد؛ که جمع‌آوری داده برای هر محصول حدوداً از چهار ماه قبل بوده است. محصولات زراعی به دلیل بازه رشد در طول عمرشان ماهیتی پویا دارند؛ بنابراین، استفاده از مجموعه داده‌های سری زمانی یک راه‌حل مؤثر و مناسب برای نقشه‌برداری محصولات است. شبکه ترنسفورمر [۲۹] معماری یادگیری عمیق بر اساس مکانیسم توجه چند سر توسعه یافته است در ابتدا برای پردازش زبان طبیعی مورد استفاده قرار گرفت، سپس ترنسفورمر بینایی (ViT) یک ترنسفورمر است که برای بینایی کامپیوتر طراحی شد و مورد استفاده قرار گرفت. زی انزه و همکاران مدل Segformer [۳۰]، که ترکیبی از معماری ترنسفورمر و شبکه‌های عصبی کانولوشنال ترکیب شده است، ارائه داده‌اند. در مطالعاتی که توسط وی چنگ سونگ و همکاران برای نقشه‌برداری دو هدفه محصولات کشاورزی که هدف ایجاد تعادل بین نقشه‌برداری دقیق انواع محصولات و کاهش خطاهای طبقه‌بندی است، با استفاده از تصاویر سنتینل-۲ بر پایه چندین شبکه یادگیری عمیق نتایج ارزشمندی به دست آورده‌اند [۳۱]، آزمایشات نشان داد که مدل Segformer در مقایسه با سایر روش‌های یادگیری عمیق و یادگیری‌ماشین سنتی رایج، نتایج: نرخ پایین خطا، استخراج ویژگی دقیق در مقیاس کوچک و دقت طبقه‌بندی بهتر را داده است.

جدول ۱: مقایسه جدیدترین روش‌های مبتنی بر تکنیک‌های یادگیری عمیق در کاربردهای بخش‌بندی و طبقه‌بندی محصولات کشاورزی

Table 1. A comparison of the recent deep learning based techniques for crop classification and segmentation

منبع - سال	کاربرد	تعداد محصول	باندهای مورد استفاده	تکنیک	نوع داده‌ها
[۲۲] - ۲۰۲۲	بخش‌بندی	۰	B2, B3, B4, B8	Swin U - Net	فضایی - طیفی - زمانی
[۲۷] - ۲۰۲۲	طبقه‌بندی محصولات	۵	B2, B3, B4, B8	CNN-Attention	فضایی - طیفی - زمانی
[۲۶] - ۲۰۲۲	بخش‌بندی و طبقه‌بندی محصولات	۲	B2, B3, B4, B8, B13	Deep learning	فضایی - طیفی - زمانی
[۴۳] - ۲۰۲۲	طبقه‌بندی محصولات برنج	۱	B1 و B13	CNN-Attention	فضایی - طیفی - زمانی
[۳۷] - ۲۰۲۳	بخش‌بندی محصولات برنج	۱	B2 و B8, B11 و B12	ResU-Net	فضایی - طیفی - زمانی
[۴۷] - ۲۰۲۳	بخش‌بندی و طبقه‌بندی محصولات	۳	B2 و B8, B11 و B12	MSSNET	چند مقیاسی
[۳۱] - ۲۰۲۳	بخش‌بندی برنج	۲	B2, B3, B4, B8	Segformer	فضایی - طیفی - زمانی
۲۰۲۴	بخش‌بندی و طبقه‌بندی محصولات	۴	B2, B3, B4	روش پیشنهادی ما	فضایی - طیفی - زمانی

بخش‌بندی معنایی به نام CropNet [۴۱] توسط پریتپال بوت و همکارانش پیشنهاد شده است که از یک چارچوب رمزگذار-رمزگشا با شبکه ResNet به‌عنوان ستون فقرات رمزگذار و ماژول‌های توجه در رمزگشا برای تمرکز بهبود یافته روی ویژگی‌های حیاتی استفاده می‌کند. مناسب برای مزارع کوچک ارائه شده است. همچنین تکنیک دیگری به کار برده شده است که ویژگی‌های عمیق [۴۲] را از تصاویر سری‌های زمانی SAR استخراج می‌کند. ویژگی‌های عمیق، که اغلب به‌عنوان «ویژگی‌های یادگیری عمیق» شناخته می‌شوند، بازنمایی‌های انتزاعی و پیچیده‌ای هستند که یک مدل یادگیری عمیق، به‌ویژه یک شبکه عصبی، به‌طور خودکار از داده‌های خام در طول فرآیند آموزش استخراج می‌کند. این فرآیند به این صورت است که تصاویر سری‌های زمانی SAR در سطوح چندگانه ترکیب می‌شوند و با استفاده از ماژول مکانیزم توجه و پیش‌گسترش ویژگی‌ها را از مجموعه داده‌ها استخراج می‌کند [۵۰]. روش‌های دیگری هم از تصاویر پهبادها برای تشخیص نواحی یا محصولات کاشت شده استفاده نمودند [۵۱] که بدلیل متفاوت بودن نوع، مرجع و کیفیت داده‌ها مورد بحث این مقاله نبوده‌اند. در این پژوهش روش طبقه‌بندی چند کلاسه بر اساس ViT و مکانیسم توجه به کار برده شده است. اما دلایلی همچون درک فرآیند تصمیم‌گیری مدل و تفسیر نتایج، بخش‌بندی را چالش‌برانگیز می‌کند. دلایل ارائه روش پیشنهادی و نوآوری ما در این پژوهش به شرح زیر است:

- ایجاد مجموعه داده جدید برای محصولاتی که دارای تنوع فصلی و مکانی است.
- یک معماری ترکیبی برای بخش‌بندی و طبقه‌بندی هم‌زمان اراضی و محصولات کشاورزی درون آنها توسعه داده شده است.
- رسیدن به دقت مطلوب و برتری روش پیشنهادی با توجه به مجموعه داده کم نسبت به روش‌های پیشین.
- همچنین مدل پیشنهادی دارای قابلیت تعمیم‌پذیری برای مناطق مختلف است.

مجموعه داده‌ها

داده‌های طبقه‌بندی محصولات معمولاً شامل انواع مختلفی از اطلاعات مانند تصویربرداری ماهواره‌ای، داده‌های اطلاعات میدانی و الگوها و روندهای قبلی کشت می‌باشد. تصاویر ماهواره‌ای با وضوح بالا معمولاً برای طبقه‌بندی محصولات استفاده می‌شود. نوارهای مختلف تصویر می‌توانند اطلاعاتی در مورد سلامت، تراکم و ویژگی‌های پوشش گیاهی ارائه دهند. دیگر معماری مورد استفاده در بخش‌بندی معنایی که بر پایه U-Net مورد توجه قرار گرفته است Attention U-Net [۴۷] می‌باشد. در پژوهشی توسط وانگ زوجی و همکارانش این معماری بر اساس مکانیسم‌های توجه چند مقیاسی انقباضی در U-Net (DMSAU-Net) برای تشخیص آفات حشرات محصول زراعی استفاده شده است. معماری پیشنهادی در این مقاله، دقت تشخیص آفات در محصولات را بهبود

رویکردهای تجزیه و تحلیل تصویر مبتنی بر شی‌روش مؤثری برای پردازش تصاویر چند زمانی سنتینل-۱ و سنتینل-۲ با وضوح فضایی بالا است. تجزیه و تحلیل تصویر مبتنی بر شی شامل پیکسل‌هایی است که اشیا بر اساس شباهت طیفی یا یک متغیر خارجی مانند واحد زمین‌شناسی طبقه‌بندی می‌شوند. نقشه‌برداری دقیق از انواع محصولات در زمین‌های زراعی ناهمگن فضایی برای کاربردهای کشاورزی بسیار اهمیت دارد [۳۳، ۳۴]. با بررسی معماری‌های درخت تصمیم، DeepLabv3+ [۳۸] و Segformer که یک معماری مدل یادگیری عمیق است که مدلی مبتنی بر یک رمزگذار ترنسفورمر و با یک رمزگشای شبکه عصبی پیچشی (CNN) سبک وزن [۴۹] برای انجام طبقه‌بندی پیکسلی ترکیب شده است. CNN سبک وزن به یک مدل شبکه عصبی کانولوشنال اشاره دارد که به گونه‌ای طراحی شده است که از نظر محاسباتی کارآمد، با پارامترهای کمتر و نیازهای محاسباتی کمتر در مقایسه با معماری‌های بزرگتر و پیچیده‌تر CNN باشد. معماری Segformer به مقدار دقت کلی ۹۱.۰۶ درصد دست یافته است [۳۱ و ۳۲]. در پژوهش‌هایی که توسط ای جون و لنگ و همکارانش انجام شد، مجموعه داده‌های که از GEE برای بخش‌بندی شبکه‌های عصبی پیچشی مانند معماری U-Net++ [۳۵] به کار برده شده است. که از طریق بلوک‌های پیچشی Dense به هم متصل شده‌اند و مسیرهای اتصال دوباره طراحی شده است؛ که ارتباط بین نقشه‌های ویژگی رمزگذار و رمزگشا را تقویت می‌کند جریان کاری بدین صورت است که تصاویر کوچک با استفاده از تصاویر نوری چندزمانی، سنتینل-۱ و ۲ با نمونه‌گیری جزئی از کلاس‌ها آموزش داده می‌شوند و قابلیت‌های تعمیم یادگیری عمیق را برای طبقه‌بندی با استفاده از برچسب فراهم می‌سازند. معماری U-Net++ در مقایسه با دیگر معماری‌ها مانند R2U-Net [۳۶]، ResU-Net [۳۷] در طبقه‌بندی شی‌گرا (مبتنی بر چند ضلعی) عملکرد قابل توجه داشته است. این معماری توانسته است نقشه ویژگی با وضوح کامل در چند سطح معنایی ایجاد کند که باعث بالا رفتن دقت مدل شده است. Swin U-Net، اولین معماری U شکل مبتنی بر ترنسفورمر خالص با اتصالات رمزگذار، گلوگاه، رمزگشا و Skip Connection، در بخش بندی تصاویر پزشکی موفقیت آمیز استفاده شده است. برای بهبود بخش‌بندی تصویر و انتخاب طیفی پیشنهاد شده است [۴۰]. جونوان یائو و شوانگن جین در مطالعات خود برای بخش‌بندی چند کلاسی برای تصاویر سنجش‌از‌دور از مجموعه داده‌ای با وضوح متوسط استفاده کرده‌اند. زیرا تصاویر ماهواره‌ای با وضوح متوسط تعداد زیادی از داده‌های سری زمانی را تشکیل می‌دهند. و پوشش کاملی را برای نظارت بر سطح زمین دارند. اجسام مختلف در تصویر ماهواره‌ای ممکن است دارای مقادیر طیفی مشابه و اجسام مشابه مقادیر طیفی متفاوتی داشته باشند، که باعث تاثیرگذاری در دقت طبقه‌بندی شود. برای حل این مشکل مدل پیشنهادی آن‌ها با استفاده از مدل Swin U-Net در یادگیری شبکه‌های عصبی عمیق بهبود داده شده است. یک معماری

مطالعه مورد بررسی قرار گرفته را نشان می‌دهد، که تمرکز آن بر کشاورزی در مقیاس کوچک و عمده محصولات شامل کلزا، گندم، برنج و ذرت است. در جدول ۲ نوع محصولات کشاورزی به همراه تعداد تصاویر به دست آمده در فصل‌های مختلف از مناطق کشت استان‌های مورد مطالعه آمده است. نقشه برداری و بررسی‌های زمینی از مزارع کشت شده بر اساس مختصات GPS به دست آمده انجام شده است. فیلم‌های نقشه برداری شده دارای اندازه‌ها و اشکال مختلفی هستند. مجموعه داده‌ها شامل تصاویر ماهواره‌ای با در نظر گرفتن الگوهای متنوع بارندگی و آب‌وهوا می‌باشد.



شکل- ۱. محدوده‌ی جغرافیایی مورد مطالعه
Fig. 1: Geographical region of interest

جدول ۲. تعداد و موقعیت جغرافیایی محصولات کشاورزی جمع آوری شده

Table 1. Location and number of collected agricultural products data				
مناطق مورد مطالعه	برنج	گندم	کلزا	ذرت
خراسان شمالی		✓		
البرز			✓	
فارس			✓	
اردبیل			✓	
تهران			✓	
کرمانشاه			✓	
خوزستان	✓	✓	✓	
اصفهان	✓	✓	✓	
مازندران	✓	✓	✓	
تعداد تصاویر	۵۵	۵۴	۶۰	۵۵
تعداد مناطق مورد مطالعه	۱۲۵	۲۹۰	۴۱۰	۳۲۵
بازه زمانی	از اردیبهشت تا شهریور	از فروردین تا تیر	از فروردین تا تیر	از مهر تا آذر
تعداد برچسب	۱۰۴۰	۱۱۶۰	۱۶۴۰	۱۳۰۰

در آزمایش‌های انجام شده، داده‌های ماهواره‌ای با داده‌های برچسب‌داری که از داده‌های برداری در نرم‌افزارهای سامانه اطلاعات جغرافیایی (GIS) ایجاد شده‌اند، ادغام شده‌اند. پس از حل چالش‌های جمع‌آوری داده‌ها،

با مداخلات به موقع برای محافظت از عملکرد محصول کمک می‌کند [۴۸]. وضوح فضایی تصاویر ماهواره‌ای برای شناسایی دقیق مزارع و انواع مختلف محصولات بسیار مهم است. تصاویر با وضوح بالاتر می‌توانند اطلاعات دقیق‌تری را ارائه دهند، در حالی که تصاویر با وضوح پایین برای طبقه‌بندی کلی مناسب‌تر می‌باشند. ماهواره‌های سنتینل برای مأموریت‌های رصد زمین هستند که توسط آژانس فضایی اروپا (ESA) به عنوان بخشی از برنامه کوپرنیک توسعه یافته است. این مأموریت‌ها برای ارائه داده‌های با کیفیت بالا برای برنامه‌های مختلف محیطی و امنیتی طراحی شده‌اند. وظیفه ماهواره سنتینل-۲ تصویربرداری نوری برای نظارت بر زمین است که شامل نقشه‌برداری کشاورزی و پوشش زمین می‌شود. ماهواره سنتینل-۲ دارای ۱۳ باند طیفی است. در این مطالعه از باندهای اصلی (باندهای RGB) استفاده شده است که شامل B2 (رنگ آبی)، B3 (رنگ سبز) و B4 (رنگ قرمز) است و دارای وضوح ۱۰ متر می‌باشند.

چالش‌های جمع‌آوری مجموعه داده‌ها

- جمع‌آوری داده‌ها از ماهواره‌هایی مانند سنتینل-۲ می‌تواند چندین چالش را به همراه داشته باشد، از جمله موارد زیر:
 - وضوح فضایی سنتینل-۲ بین ۱۰ تا ۶۰ متر است که ممکن است برای جمع‌آوری اطلاعات دقیق در زمین‌های کشاورزی کوچک یا گیاهان خاص کافی نباشد.
 - دوره تکرار زمانی سنتینل-۲ پنج روز است که ممکن است با مراحل یا فنولوژی کشاورزی همخوانی نداشته باشد.
 - پوشش ابری می‌تواند تصاویر ماهواره‌ای را مبهم کند و بر کیفیت داده‌های جمع‌آوری شده برای نظارت کشاورزی تأثیر منفی بگذارد.
 - پیش‌پردازش تصاویر ماهواره‌ای برای تشخیص محصولات کشاورزی، که شامل تصحیح اثرات جوی، پوشش ابر و سایر عوامل است، برای اطمینان از صحت و قابلیت اطمینان داده‌ها ضروری است اما می‌تواند هزینه‌بر و زمان‌بر باشد.
 - تفسیر تصاویر ماهواره‌ای برای برچسب‌گذاری کشاورزی نیاز به تخصص در زمینه سنجش از دور، فنولوژی محصول و زراعت دارد تا اطلاعات مرتبط مانند سلامت محصول، تخمین عملکرد و تشخیص آفات و بیماری‌ها استخراج شود.
 - یکی از چالش‌های اصلی، مرحله برچسب‌گذاری است. برای انتخاب هر محصول، باید نوع محصولات با شناسایی فنولوژی و استفاده از مختصات به دست آمده و حضور میدانی، به عنوان نمونه‌ای از داده‌ها برچسب‌گذاری شوند.

مناطق جغرافیایی مجموعه داده‌ها

تصاویر ماهواره‌ای مورد مطالعه در این تحقیق از ۹ استان مختلف ایران شامل خراسان شمالی، البرز، فارس، اردبیل، کرمانشاه، تهران، خوزستان، اصفهان و مازندران جمع‌آوری شده است. شکل ۱ محدوده‌ای که در این

- یکپارچه‌سازی یا ادغام دو یا چند قسمت تصویر استفاده می‌شود. یکپارچه‌سازی شامل حالت‌های زیر است:
- تصاویر چندزمانی: از سنجنده‌های یکسان یا چندگانه برای تشخیص تغییرات استفاده می‌کند.
 - تصاویر چندفضایی: از سنجنده‌های یکسان یا چندگانه برای یک محدوده خاص استفاده می‌شود.
 - تصاویر سنجش از دور: هنگامی که با داده‌های جانبی ترکیب شوند، می‌توانند درک جامع‌تری از سطح زمین ارائه دهند.

شکل ۲ نوعی از تصاویر سنجش از دور چندزمانی را نشان می‌دهد که از منطقه‌ای در بهشهر، استان مازندران، در فاصله زمانی ۱۸ خرداد تا ۲۶ مرداد ۱۴۰۲ گرفته شده است.



شکل ۲: نمونه‌های از تصاویر چندزمانی منطقه‌ای از بهشهر استان مازندران توسط سنتینل ۲-
Fig. 2: A sample of multi temporal Sentinel-2 images from Behshahr, Mazandaran province

نقاط مختلف جمع‌آوری شده‌اند و هر چند ضلعی در تصویر ممکن است نمایانگر یک محصول متفاوت باشد و اطلاعات کافی در مورد آن وجود نداشته باشد. این مسأله برچسب‌گذاری را مشکل کرده و باعث ایجاد چالش‌هایی در آموزش مدل‌های یادگیری می‌شود.

پیش‌پردازش داده‌ها

داده‌های سنتینل-۲ باید قبل از پردازش، برای از بین بردن خطاهای جوی و رادیومتری تصحیح شوند. تصحیح اتمسفری یکی از مهم‌ترین مراحل پیش‌پردازش است؛ زیرا می‌تواند بر نتیجه نهایی تأثیر قابل توجهی بگذارد. هدف اصلی تصحیح جوی، اصلاح اثرات جوی بر تصاویر ماهواره‌ای با تعیین ویژگی‌های نوری است. با این کار، تغییرات در زاویه خورشید، اثرات مه و ابرها به طور مؤثری از بین می‌روند. بنابراین، هنگام پردازش داده‌های سنتینل-۲، توصیه می‌شود از تصاویر بدون ابر استفاده شود. پس از این مرحله، داده‌ها برای استفاده در مدل‌های یادگیری آماده می‌شوند. در تصاویر ماهواره‌ای، پوشش ابری می‌تواند کیفیت تصاویر را به میزان قابل توجهی کاهش دهد و بر توانایی تشخیص ویژگی‌ها و جزئیات سطح تأثیر بگذارد. شکل ۳ نشان می‌دهد که پوشش ابری می‌تواند به‌ویژه برای کاربردهایی مانند بخش‌بندی پوشش زمین چالش‌برانگیز باشد.

نیاز به استفاده از تکه‌های همپوشان کاشی‌های ۲۵۶×۲۵۶ برای تجزیه و تحلیل تصاویر ایجاد شد. تقسیم بندی تصاویر اولیه بزرگ در محدوده جغرافیایی مورد نظر با استفاده از ابزار Q-GIS به به ابعاد ۲۵۶×۲۵۶ تبدیل شده است. تصاویر سری زمانی داده‌های بدون ابر، تقریباً در فواصل پانزده روزه از ماهواره سنتینل-۲ برای هر فصل زراعی در بازه زمانی یک اردیبهشت تا ۳۰ آذر ساخته شده‌اند.

داده‌های چند زمانی

داده‌های چندزمانی به مجموعه داده‌هایی اشاره دارد که تغییرات یک منطقه خاص را در طول زمان ثبت می‌کنند. در نقشه‌برداری، در صورتی که سطح پیکسل‌ها دارای اشکال مختلف و متمایز نباشند، از

آماده‌سازی داده‌ها

یکی از مراحل پیش‌پردازش داده در تصاویر ماهواره‌ای، آماده‌سازی داده‌های ماهواره‌ای با تغییر ماهیت داده‌ها است. داده‌های ماهواره سنتینل-۲ به صورت فایل‌های فشرده با فرمت SAFE ارائه می‌شوند. باندهای طیفی در سنتینل-۲ به صورت فایل‌های JPG در فرمت SAFE ذخیره می‌شوند و دارای سه وضوح هندسی مختلف (۱۰ متر، ۲۰ متر و ۶۰ متر) هستند. وضوح ۱۰ متری این فایل‌های JPG را می‌توان در یک فایل Geotiff با اندازه پیکسل یکنواخت ۱۰ متر به صورت ساختار داده‌ی رستری ایجاد کرد. با استفاده از این ساختار می‌توان عملیات همپوشانی را به آسانی و به شکل مؤثر انجام داد. برای استفاده بهینه از داده‌ها، پیش‌پردازش‌هایی انجام می‌شود تا حجم داده به حداقل برسد و زمان محاسبات کاهش یابد. پس از انجام مراحل توضیح داده شده در بخش‌های قبلی این مقاله، نقشه مورد نظر از روی تصاویر ماهواره‌ای سنتینل-۲ ایجاد می‌شود.

در این مطالعه از ابزار Q-GIS برای حاشیه نویسی (رنگ بندی هر محصول و پس زمینه) استفاده شده است. مجموعه داده‌های استفاده شده در این مقاله، در شرایط جوی و مکانی مختلف و همچنین در فصل‌های مختلف سال جمع‌آوری شده‌اند که این امر باعث پیچیدگی و زمان‌بر شدن کار می‌شود. یکی دیگر از مشکلات این است که داده‌ها از

شبکه‌های عصبی پیچشی آبشاری چندمرحله‌ای متکی است. این چارچوب‌های آبشاری ویژگی‌های سطح بالا را استخراج کرده و پیش‌بینی‌های متراکمی را در هر مرحله انجام می‌دهند. پیش‌بینی متراکم به معنای پیش‌بینی یک برچسب یا مقدار رگرسیون برای هر پیکسل در یک تصویر ورودی است؛ به جای ارائه یک برچسب واحد برای کل تصویر، به هر پیکسل یک برچسب یا مقدار اختصاص داده می‌شود. این رویکرد می‌تواند منجر به استفاده بیش از حد از منابع محاسباتی شود، زیرا به طور مکرر ویژگی‌های سطح پایین مانند بافت تصویر را استخراج می‌کند.

مدل استاندارد U-Net با هدف بهبود حساسیت مدل به پیکسل‌های پیش‌زمینه، به روشی جدید به نام Attention U-Net ارتقا یافته است. این روش به شبکه اجازه می‌دهد تا ساختارهای بصری عمومی‌تر را بیاموزد و در برابر ورودی‌های نویزدار مقاوم باشد. همان‌طور که در شکل ۵ الف نشان داده شده است، معماری Attention U-Net از عملگر الحاق به جای عملگر جمع بین مسیر رمزگذاری (downsampling) و مسیر رمزگشا (upsampling) استفاده می‌کند. این اتصالات پرش به شبکه امکان می‌دهند تا اطلاعات سراسری را در حین نمونه‌برداری به اطلاعات محلی منتقل کنند. در مسیر نمونه‌برداری، به دلیل تقارن شبکه، تعداد زیادی نقشه ویژگی وجود دارد که امکان انتقال اطلاعات را فراهم می‌کنند.

معماری Attention U-Net دارای دروازه‌های توجه است که به شبکه اجازه می‌دهند بر روی مناطق مهم تمرکز کند و اطلاعات نامربوط را فیلتر کند. به عبارت دیگر، این دروازه‌ها اطلاعات کم‌اهمیت‌تر را سرکوب می‌کنند [۴۵]. مکانیسم توجه با کاهش تفاوت معنایی بین مسیرهای رمزگذاری و رمزگشایی و حفظ جزئیات لبه‌ها، به بهبود دقت بخش‌بندی کمک می‌کند. علاوه بر این، استفاده از دروازه‌های توجه، شبکه را قادر می‌سازد تا به‌طور مؤثر مشکل تداخل نویز را برطرف کرده و منجر به نتایج بخش‌بندی قوی‌تری شود. با گنجاندن مکانیسم‌های توجه در معماری U-Net، مدل می‌تواند به‌صورت پویا اهمیت مکان‌های مختلف فضایی در نقشه‌های ویژگی را در طول مراحل رمزگذاری و رمزگشایی تنظیم کند.

در پردازش تصویر، اطلاعات زمینه به رابطه بین اشیاء، اشکال و بافت‌های درون یک تصویر اشاره دارد. شناخت این روابط به درک صحنه کلی کمک می‌کند. در بحث تشخیص اشیاء، اشیاء اطراف می‌توانند به بهبود دقت شناسایی و طبقه‌بندی اشیاء در یک تصویر کمک کنند.

در مدل پیشنهادی برای بهبود Attention U-Net، از یک ترنسفورمر بینایی پیکسلی (ViT) در گلوگاه مدل استفاده شده است. ادغام ViT معماری می‌تواند به مدل کمک کند تا وابستگی‌های فضایی دوربرد را کنترل کند [۴۶]. وابستگی‌های فضایی دوربرد به روابط یا همبستگی بین پیکسل‌ها یا مناطق دوردست در یک تصویر یا یک ساختار داده مکانی اشاره دارد. شکل ۵ الف) معماری مدل پیشنهادی را نشان می‌دهد.



شکل ۳: میزان تاثیر پوشش ابری در بخش‌بندی (مستطیل راست بالا نشان‌دهنده ی ابر است)

Fig. 3: The impact of cloud cover on segmentation (top right rectangle indicates the cloud).

روش پیشنهادی

چارچوب کلی روش پیشنهادی در شکل ۴ نشان داده شده است که شامل دو بخش آموزش و تست می‌باشد و در پنج مرحله انجام شده است. در مرحله اول، مجموعه داده‌ها جمع‌آوری شدند. در مراحل دوم و سوم، تصاویر ماهواره‌ای بخش‌بندی شده و برچسب‌گذاری محصولات انجام می‌شود. آموزش مدل در مرحله چهارم با استفاده از داده‌های جمع‌آوری شده بر روی معماری بهبودیافته U-NET انجام گردید. در مرحله پنجم، با استفاده از مدل آموزش‌دیده، تصاویر جدید بخش‌بندی و کلاس‌بندی شده‌اند.

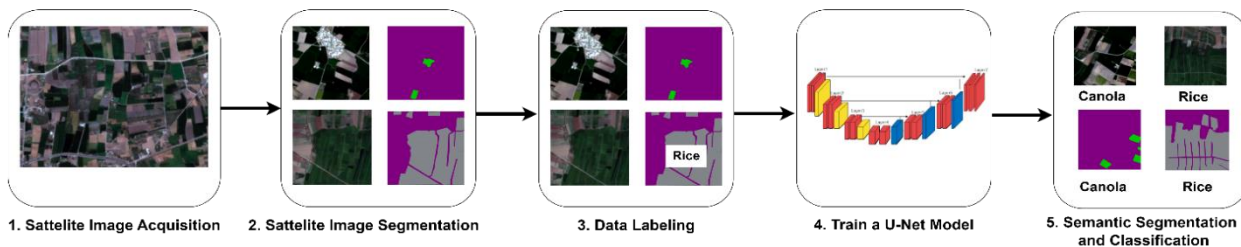
در این تحقیق، تصاویر جمع‌آوری شده یک ماه قبل از برداشت محصول موردنظر تهیه شده‌اند تا وضعیت فیزیولوژیکی پوشش گیاهی هر محصول را به دقت بازتاب دهند. نکته قابل توجه دیگر این است که با تغییرات فیزیولوژیکی رشد محصول و تغییر رنگ‌دانه‌ها، رنگ‌های متفاوتی ایجاد می‌شود. برای حل این موضوع، محدوده‌ی طیف رنگی برای هر محصول شامل دو رنگ در نظر گرفته شده است.

بخش‌بندی معنایی

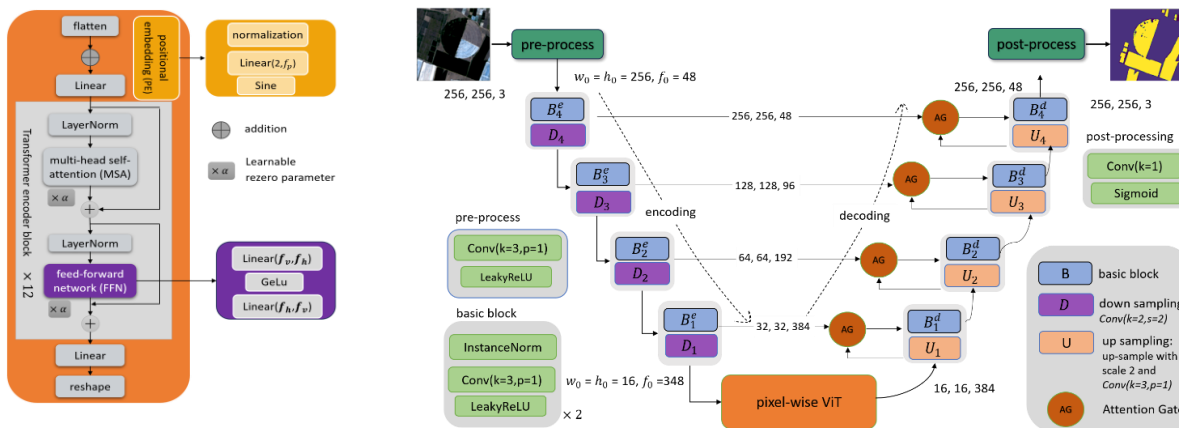
بخش‌بندی معنایی یک روش برای بخش‌بندی پیکسلی تصویر بر اساس ویژگی‌های مشترک پیکسل‌ها، مانند رنگ، است. به عبارت دیگر، هدف تولید یک نقشه بخش‌بندی پیکسلی متراکم از یک تصویر است که در آن هر پیکسل به یک کلاس یا شیء خاص اختصاص داده می‌شود. مدل‌های مختلف U-Net به دلیل عملکرد بالا و استفاده کارآمد از حافظه GPU معمولاً برای کارهای بخش‌بندی تصویر استفاده می‌شوند. هدف این مدل‌ها دستیابی به دقت بالا با استفاده از تعداد کم نمونه‌های آموزشی است.

معماری Attention U-Net-ViT

مدل U-Net یکی از محبوب‌ترین مدل‌ها برای بخش‌بندی تصاویر است. این معماری به طور کلی از دو قسمت رمزگذار و رمزگشا تشکیل شده است. دیتای ورودی ابتدا وارد رمزگذار می‌شود و پس از اعمال عملگرهای مختلف لایه‌های کانولوشن، به مسیر دوم یعنی رمزگشا منتقل می‌شود. مدل U-Net در لایه‌های رمزگذار و رمزگشا به



شکل ۴: مراحل اجرای فرآیند روش پیشنهادی در این تحقیق
 Fig. 4: The steps of implementing the proposed method in this paper



شکل ۵: الف) معماری روش پیشنهادی بر روی مدل U-Net
 a) The proposed architecture based on U_Net

ب) بلوک رمزگذار ترنسفورمر
 Fig. 5: b) Decoder block transformer

معیارهای ارزیابی

استفاده از معیارهای ارزیابی در یادگیری ماشین برای ارزیابی عملکرد یک مدل یادگیری ماشینی بسیار مهم است. در ادامه به بررسی معیارهای ارزیابی استفاده شده از جمله: FL، MIOU، دقت، فراخوانی، صحت و Dice خواهیم پرداخت.

معیار ارزیابی افت کانونی

افت کانونی (FL) یک نسخه بهبود یافته از اتلاف متقاطع آنتروپی (CE) است که با اختصاص وزن‌های بیشتر به نمونه‌هایی که به اشتباه طبقه‌بندی شده‌اند (مانند پس‌زمینه با بافت پرنویز یا اشیاء جزئی یا موضوعات مورد علاقه ما) و وزن کمتر به نمونه‌های درست طبقه‌بندی شده، عملکرد مدل را بهبود می‌بخشد. فرمول محاسبه افت کانونی به صورت زیر است:

$$\text{Focal Loss} = - \sum_{i=1}^n (i - p_i)^{\gamma} \log_b p_i \quad (1)$$

معیار ارزیابی صحت

صحت به این معناست که مدل تا چه اندازه خروجی را درست پیش‌بینی می‌کند. برای اندازه‌گیری این معیار از فرمول ۲ بدست می‌آید.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{total Examples}} \quad (2)$$

در شکل ۵ ب) بلوک دیاگرام یک ترنسفورمر بینایی نمایش داده شده است. این مدل بینایی یک نوع معماری شبکه عصبی است که تصاویر را به تکه‌هایی با اندازه ثابت تقسیم می‌کند. برای حفظ موقعیت تکه‌ها از موقعیت‌یابی توکار (position embedding) استفاده شده است. سپس هر تکه به صورت خطی در موقعیت اصلی خود قرار گرفته و نرمال می‌شود و نهایتاً تکه‌های تصاویر وارد مکانیسم چندسره توجه-خود (multi-head self-attention) می‌شوند. منظور از توجه-خود این است که هر تصویر ورودی به ویژگی‌های خود توجه می‌کند و با تصاویر خروجی یا تصاویر دیگر توجهی ندارد. در فاز خروجی نهایی معماری ViT، پیش‌بینی کلاس با عبور خروجی آخرین بلوک ترنسفورمر از یک سر طبقه‌بندی که معمولاً از یک لایه کاملاً متصل تشکیل شده، به دست می‌آید.

عملکرد یک مدل ترنسفورمر بینایی به عواملی مانند بهینه‌ساز، عمق شبکه و فرآیندهای خاص مجموعه داده بستگی دارد. ViT مبتنی بر پیکسل عمدتاً از یک پشته بلوک‌های رمزگذار ترنسفورمر تشکیل شده است و اندازه پشته ثابت است. برای ایجاد یک ورودی به پشته ViT، ابتدا رمزگذاری در امتداد ابعاد فضایی صاف (flat) می‌شود تا دنباله‌ای از تکه‌های ترنسفورمر تشکیل شود.

فرآیند آموزش در مدل‌های یادگیری ماشینی برای تشخیص الگوها و پیش‌بینی یا تصمیم‌گیری براساس داده‌ها صورت می‌گیرد. یکی از چالش‌های اصلی در معماری‌های ترنسفورمر بینایی، کمبود داده برای

معیار ارزیابی دقت

زمانی که ارزش مثبت های کاذب بالا باشد، معیار دقت معیار مناسبی خواهد بود. نتیجه این امر این است که این مدل، نقاط درست بسیاری را به اشتباه تشخیص می دهد.

(۳)

$$\text{Precision} = \frac{\text{True Positiv}}{\text{True Positive} + \text{False Positive}}$$

معیار ارزیابی فراخوانی

در نقطه مقابل صحت این معیار، در مواقعی که دقت تشخیص کلاس منفی حائز اهمیت باشد معیار فراخوانی مناسب می باشد.

(۴)

$$\text{Recall} = \frac{\text{True Positiv}}{\text{True Positive} + \text{False Negative}}$$

معیار ارزیابی اشتراک بر اجتماع

اشتراک بر اجتماع (IoU) دو نقطه برابر است با نسبت مساحت هم پوشانی به مساحت کل. مقدار این نسبت بین ۰ تا ۱ تغییر می کند، به طوری که اگر مختصات نقطه پیش بینی شده توسط مدل و نقطه حقیقی دقیقاً یکسان باشند، مقدار IoU برابر با ۱ خواهد بود. این معیار را می توان برای هر کلاس به صورت جداگانه محاسبه کرد یا با گرفتن میانگین IoU در تمام کلاس ها، یک IoU متوسط به دست آورد که به آن mIoU گفته می شود.

$$\text{IoU} = \frac{\text{Area Of Overlap}}{\text{Area Of Union}} \quad (۵)$$

معیار ارزیابی ضریب تشابه تاس

Dice یکی از معیار های اصلی برای ارزیابی نتایج بخش بندی تصاویر است. مقدار Dice نشان می دهد که نتایج مدل بخش بندی، چقدر به ماسک حقیقی بافت، شباهت دارند. برای به دست آوردن مقدار متوسط Dice از میانگین گرفته می شود.

$$\text{Dice} = \frac{2 \text{ True Positivite}}{2 \text{ True Positivite} + \text{False Positivite} + \text{False Negative}} \quad (۶)$$

نتایج شبیه سازی

جدول ۳: مقایسه معماری ها و نتایج بدست آمده

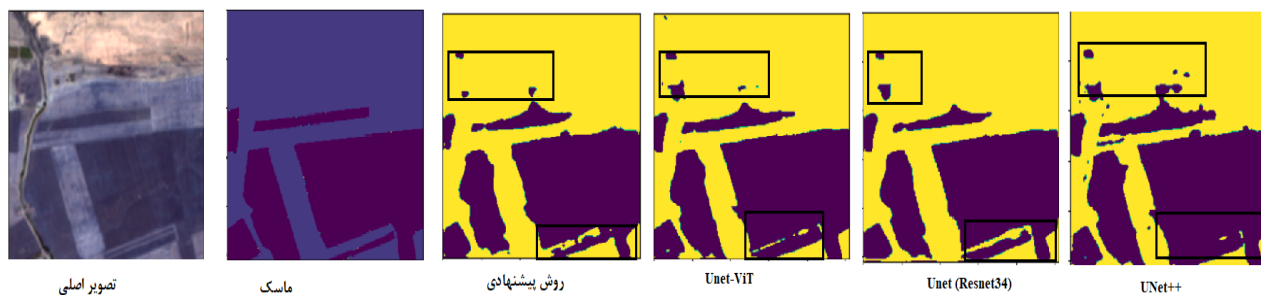
Table 3: Comparison of different architectures' result

مدل	تعداد پارامتر	صحت	دقت	فراخوانی	m(Iou)	DiceyH
UNet (Resnet34)	۳۶۹۴۴۷۷۶	۶۶.۴۰	۸۰.۷۴	۵۶.۹۲	۴۰٪	۰.۴۵۵
UNet	۳۱۸۶۳۴۹۶	۶۸.۳۵	۸۲.۰۰	۵۶.۸۴	۳۹.۹٪	۰.۴۳۱
UNet++	۳۱۸۶۳۴۹۶	۶۷.۳۵	۸۲.۵۶	۵۹.۷۷	۴۳.۵۰	۰.۴۱۴
Attention UNet	۳۴۸۷۹۴۸۳	۶۹.۵۵	۸۲.۸۳	۵۷.۷۳	۳۴.۶٪	۰.۴۵۶
Swin UNet	۱۷۹۱۷۷۴۰	۵۲.۵۰	۳۲.۵۷	۳۰.۳۱	۱۹.۶٪	۰.۱۶۶
UNet (backbone ViT)	۱۰۱۷۵۶۸۷۶	۵۶.۴۶	۴۳.۴۲	۴۸.۴۲	۲۵.۳٪	۰.۲۵۶
UNet-ViT	۸۸۴۰۵۵۷۹	۶۹.۹۰	۸۳.۳۴	۵۱.۶۳	۴۳.۹٪	۰.۴۵۸
روش	۸۸۷۵۶۸۷۶	۷۰.۶۹	۸۳.۸۴	۵۹.۳۸	۴۶.۸٪	۰.۴۶۶

پیشنهادی

مجموعه داده آموزشی شامل ۲۴۶ تصویر همراه با برچسب با اندازه ۲۵۶×۲۵۶ از تصاویر سنتینل-۲ استفاده شده است. برای پیاده سازی طبقه بندی کننده یادگیری عمیق، باید مدل را برای دسته های مختلف کلاس ها آموزش داد. جدول ۳ نتایج طبقه بندی را با استفاده از سه باند سنتینل-۲ نشان می دهد مدل U-Net با ستون فقرات Resnet34، که از بیش برآزش جلوگیری کرده و نتایج بهتری نسبت به U-Net ساده به دست آورده است. همچنین، معماری پیشنهادی توانسته است پارامترهای زیاد خود را به خوبی تنظیم کرده و نتایج قابل توجهی به دست آورد، در حالی که مدل های دیگر به دلیل فقدان وزن های اولیه نتایج خوبی نداشته اند. مدل پیشنهادی، با وجود داشتن پارامترهای بیشتر نسبت به معماری Swin-UNet، به نتایج بهتری دست یافته است. در مقالات مشابه، Swin-UNet نتایج خوبی داشته است، اما به دلیل تطبیق بیش از حد داده به نسبت تعداد و توزیع کلاس ها در تصاویر ماهواره ای، دچار عدم یادگیری شده و بیشتر به پس زمینه توجه کرده است [۴۴]. در جدول ۳ مشاهده می شود که معماری Swin-UNet، که در مقالات بخش ۲ به عنوان یکی از بهترین مدل ها معرفی شده بود، به دلیل تطبیق بیش از حد با داده ها و کم بودن تعداد نمونه نسبت به توزیع کلاس ها در تصاویر ماهواره ای، دچار عدم یادگیری شده و بیشتر به پس زمینه توجه کرده است، بنابراین نتوانسته به خوبی آموزش ببیند. مدل پیشنهادی در معیارهایی مانند صحت و Dice بهترین نتایج را به دست آورده است و همچنین از نظر دقت و mIoU نزدیک به معماری U-Net با ستون فقرات Resnet34 که از پیش آموزش دیده است، عمل کرده است. این امر نشان دهنده کارایی بالای معماری پیشنهادی است. نتایج هر مدل از محاسبه ی میانگین ۴ اجرای مختلف آن بدست آمده است. که همواره عدد نهایی میانگین با انحراف معیار ± 0.5 نشان دهنده ی ثبات نتیجه ی تست مدل در این دیتاست بوده است. در شکل ۶ مشاهده می شود که روش پیشنهادی در مقایسه با سایر معماری ها شباهت بیشتری به ماسک دارد و همچنین نسبت مساحت هم پوشانی به مساحت کل نتایج بهتری ارائه داده است. با این حال، معماری U-Net با ستون فقرات Resnet34 در مقایسه با دیگر معماری ها نقاط ضعفی نیز دارد.

شکل ۷ نشان می دهد که عدم تعادل افت کانونی در مدل های Attention UNet و SwinUNet در طول آموزش و اعتبارسنجی، باعث بیش برآزش می شود. U-Net با استفاده از ستون فقرات Resnet34، در تعداد دورهای کمتر به تعادل رسیده است، در حالی که SwinUNet نتوانسته است به تعادل برسد. به همین دلیل، در آموزش و اعتبارسنجی، مدل پیشنهادی نتایج قابل توجهی از خود نشان می دهد. شکل ۸ نتایج کیفی طبقه بندی تصاویر ماهواره سنتینل-۲ را با برچسب و به همراه خروجی های بدست آمده بر اساس معماری های بررسی شده نشان می دهد. ویژگی های طیفی تولید شده توسط پوشش گیاهی، زاویه تابش و موقعیت ابرها بر دقت طبقه بندی تأثیر گذار است.



شکل ۶: مقایسه نتایج کیفی مدل‌ها
Fig. 6: Qualitative model comparison

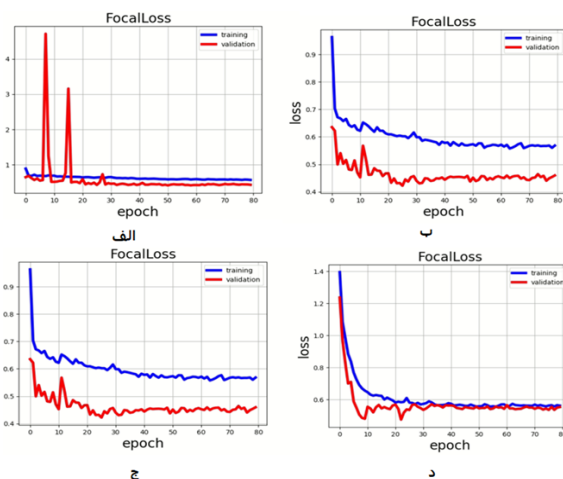
علاوه بر این، به دلیل نداشتن اطلاعات کافی در برچسب‌گذاری، طبقه‌بندی‌های نادرست در میان زمین‌های بدون کشت و مناطق شهری وجود دارد. همچنین این شکل نشان می‌دهد که تصاویر با وضوح بالا قابلیت استخراج قوی‌تری در جزئیات فضایی مانند: تشخیص مزرعه در قطعات کوچک در بخش‌بندی مزارع را دارند. طبقه‌بندی نادرست در زمین‌هایی که کمبود اطلاعات وجود دارد، باعث کاهش دقت و صحت عملکرد مدل‌ها شده است، این امر در مدل SwinUnet به دلیل کم بودن تعداد داده باعث شده است که بیشتر تصاویر به نادرستی شناسایی شوند.

بحث و نتیجه‌گیری

در این تحقیق، یک چارچوب مدل‌سازی برای بخش‌بندی معنایی تصاویر ماهواره‌ای با وضوح بالا، به منظور نقشه‌برداری کاربری زمین و مدیریت کشاورزی معرفی شده است. این مطالعه به منظور ارزیابی عملکرد مدل و بهبود دقت طبقه‌بندی، از شاخص‌های ویژگی تولید شده توسط سنجش‌ازدور استفاده می‌کند. عملکرد مدل با استفاده از معیارهای ضرر مشترک شامل تابع افت کانونی، بازیابی، Dice، صحت و MIOU ارزیابی شده است. معماری پیشنهادی نسبت به مدل‌های دیگر نشان دهنده بهبود عملکرد در طبقه‌بندی است. همچنین، مدل پیشنهادی به جز در معماری UNet++ نیز عملکرد بهتری ارائه داده است.

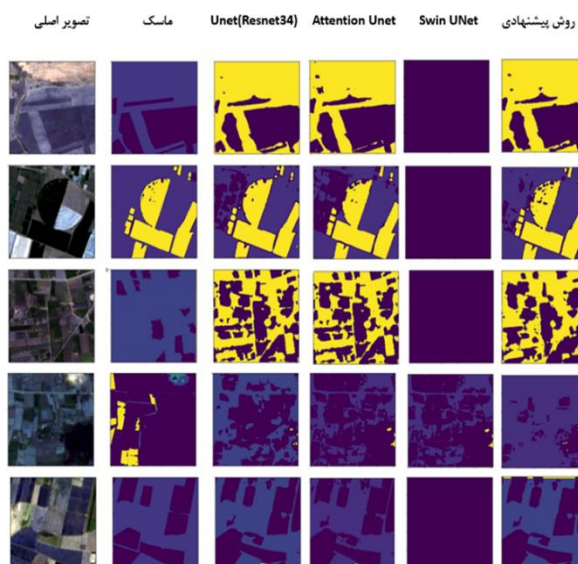
این تحقیق یک مدل سرتاسری برای یادگیری ویژگی‌های مرتبط با بخش بندی تصاویر ماهواره‌ای ارائه داده است. نتایج این تحقیق نشان می‌دهد که روش ارائه شده می‌تواند برای بخش بندی تصاویر ماهواره‌ای دریافتی از سنسینل-۲ برای محصولات مورد نظر مورد استفاده قرار گیرد. لذا نتایج حاصل می‌تواند در مدیریت مصرف آب، تنظیم ساختار کاشت، تخمین تلفات و ارزیابی عملکردهای زراعی نقش مهمی را ایفا نماید. با بهره‌گیری از این روش‌ها، می‌توان به بهبود کارایی و دقت در مدیریت کشاورزی دست یافت و از منابع بهینه‌تر استفاده کرد.

در مقایسه با مدل‌های مرسوم فعلی، رویکرد توصیف شده امکان یادگیری خودکار و کارآمد ویژگی‌های طبیعی، شاخص‌های پوشش گیاهی و ویژگی‌های بافتی را فراهم می‌کند. این مدل نیز برای طبقه‌بندی محصول و کاربری زمین در فصول مختلف سال مناسب است و قابلیت تعمیم فضایی زیادی را برای تولید نقشه‌های پوشش محصولات



شکل ۷: مقایسه FL برای هر یک از مدل‌های الف) U-Net (backbone ResNet34)، ب) Attention U-Net، ج) Swin U-Net و د) مدل پیشنهادی

Fig. 7: Comparison of FL for each of the models left-up) U-Net(ResNet34 backbone) right-up) Attention U-Net down-left)Swin U-Net down-right) Proposed model



شکل ۸: مقایسه نتایج کیفی مدل پیشنهادی با مدل‌های یادگیری عمیق مختلف در طبقه‌بندی تصاویر ماهواره‌ی سنسینل-۲

Fig. 8: Comparison of qualitative result comparison between proposed and other deep models in Sentinel-2 satellite image

[2] D. Wisser, S. Frolking, E.M. Douglas, B. M. Fekete, C. J. Vörösmarty, and A. H. Schumann, "Global irrigation water demand: Variability and uncertainties arising from agricultural and climate data sets," *Geophysical Research Letters*, vol. 35, no. 24, 2008. <https://doi.org/10.1029/2008GL035296>

[3] M. Flörke, C. Schneider, and R. I. McDonald, "Water competition between cities and agriculture driven by climate change and urban growth," *Nature Sustainability*, vol. 1, no. 1, pp. 51-58, 2018. <https://doi.org/10.1038/s41893-017-0006-8>

[4] F. Alimohammadi et al., "Hyperspectral imaging coupled with multivariate analysis and artificial intelligence to the classification of maize kernels," *International Agrophysics*, vol. 36, no. 2, 2022. <https://doi.org/10.31545/intagr/147227>

[5] H. R. Khan et al., "Early identification of crop type for smallholder farming systems using deep learning on time-series sentinel-2 imagery," *Sensors*, vol. 23, no. 4, p. 1779, 2023.

[6] B. Usowicz and J. Lipiec, "Spatial variability of thermal properties in relation to the application of selected soil-improving cropping systems (SICS) on sandy soil," *Int. Agrophys*, vol. 36, pp. 26.2022,9-284. doi: 10.31545/intagr/152122

[7] I. Becker-Reshef et al., "Crop type maps for operational global agricultural monitoring," *Scientific Data*, vol. 10, no. 1, p. 172, 2023. <https://doi.org/10.1038/s41597-023-02047-9>

[8] K. Luo, L. Lu, Y. Xie, F. Chen, F. Yin, and Q. Li, "Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning," *Computers and Electronics in Agriculture*, vol. 205, p. 107577, 2023. <https://doi.org/10.1016/j.compag.2022.107577>

[9] J. Bramich, C. J. Bolch, and A. Fischer, "Improved red-edge chlorophyll-a detection for Sentinel 2," *Ecological Indicators*, vol. 120, p. 106876, 2021.

[10] L. Feng, Y. Wang, Z. Zhang, and Q. Du, "Geographically and temporally weighted neural network for winter wheat yield prediction," *Remote Sensing of Environment*, vol. 262, p. 112514, 2021. <https://doi.org/10.1016/j.rse.2020.106876>

[11] L. Blickensdörfer, M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmí, and P. Hostert, "Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany," *Remote sensing of environment*, vol. 269, p. 112831, 2022. <https://doi.org/10.1016/j.rse.2021.112831>

[12] Feng Gao, Xiaoyang Zhang. Mapping Crop Phenology in Near Real-Time Using Satellite Remote Sensing: Challenges and Opportunities. *J Remote Sens*. 2021. <https://spj.science.org/doi/full/10.34133/2021/8379391>

[13] K. Zhang, D. Yuan, H. Yang, J. Zhao, and N. Li, "Synergy of Sentinel-1 and Sentinel-2 imagery for crop classification based on DC-CNN," *Remote Sensing*, vol. 15, no. 11, p. 2727, 2023. <https://doi.org/10.3390/rs15112727>

[14] A. Orynbaikyzy, U. Gessner, and C. Conrad, "Spatial transferability of random forest models for crop type classification using Sentinel-1 and Sentinel-2," *Remote Sensing*, vol. 14, no. 6, p. 1493, 2022. <https://doi.org/10.3390/rs14061493>

[15] M. Huang et al., "Remote sensing image fusion algorithm based on two-stream fusion network and residual channel attention mechanism," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1-14, 2022. <https://doi.org/10.1155/2022/8476000>

[16] F. Ahmed, H. A. Al-Mamun, A. H. Bari, E. Hossain, and P. Kwan, "Classification of crops and weeds from digital images: A support vector machine approach," *Crop Protection*, vol. 40, pp. 98-104, 2012. <https://doi.org/10.1016/j.cropro.2012.04.024>

در مناطق کشاورزی بدون برچسب مرجع زمینی قابل اعتماد در زمان واقعی دارد.

با توجه به دیتاست کوچک استفاده شده در این مقاله، در نگاه اول UNet مستعد overfitting می باشد، اما در ادامه به برخی از راهکارهای جلوگیری از آن اشاره خواهیم نمود:

الف) U-Net از یک مسیر انقباضی برای گرفتن زمینه و یک مسیر متقارن در حال گسترش استفاده می کند که محلی سازی دقیق را امکان پذیر می سازد. این معماری به شبکه اجازه می دهد تا با جمع آوری اطلاعات زمینه در مقیاس های چندگانه، از داده های محدود به طور مؤثرتری یاد بگیرد.

ب) U-Net شامل اتصالات پرش است که نقشه های ویژگی را از مسیر انقباض مستقیماً به مسیر در حال گسترش مربوطه منتقل می کند. این اتصالات جریان اطلاعات گرادیان را در طول آموزش تسهیل می کند و به مدل کمک می کند حتی با داده های محدود به طور مؤثر یاد بگیرد. و با اضافه کردن Attention در معماری U-Net برای تقویت مناطق اطلاعاتی در نقشه های ویژگی و سرکوب اطلاعات نامربوط استفاده می شود.

ج) راهکار دیگر ما جهت جلوگیری از overfitting در مدل U-Net استفاده از تکنیک های داده افزایی، مانند چرخش، برعکس کردن (Flip) و مقیاس بندی در نمونه های آموزشی اولیه بوده است. لذا مدل نهایی می تواند ویژگی های قوی تری را بیاموزد و بهتر تعمیم دهد. U-Net می تواند از تکنیک های منظم سازی مانند حذف تصادفی یا عادی سازی دسته های بهره مند شود، که برای جلوگیری از تطبیق تعمیم بیش از حد و بهبود تعمیم، به ویژه هنگام کار با داده های محدود کمک کند.

در ادامه ی این تحقیق نویسندگان علاقه مند به افزایش نمونه ها و انواع محصولات کشاورزی، افزایش دقت و همچنین یافتن یک محصول خاص در دوره های زمانی مختلف مثل کاشت دوم برنج، می باشند.

مشارکت نویسندگان

میزان مشارکت نویسندگان در نگارش مقاله برابر می باشد.

تشکر و قدردانی

نویسندگان از تمام افرادی که در جمع آوری داده ها ما را در این تحقیق یاری کردند تشکر و تقدیر می نماید.

تعارض منافع

«هیچگونه تعارض منافع توسط نویسندگان بیان نشده است.»

منابع و مآخذ

[1] D. Torbunov et al., "Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 702-712. <https://doi.org/10.1109/wacv56688.2023.00077>

- [31] W. Song et al., "Bi-Objective Crop Mapping from Sentinel-2 Images Based on Multiple Deep Learning Networks," *Remote Sensing*, vol. 15, no. 13, p. 3417, 2023. <https://doi.org/10.3390/rs15133417>
- [32] L. B. da Costa, O. L. F. de Carvalho, A. O. de Albuquerque, R. A. T. Gomes, R. F. Guimarães, and O. A. de Carvalho Júnior, "Deep semantic segmentation for detecting eucalyptus planted forests in the Brazilian territory using sentinel-2 imagery," *Geocarto International*, vol. 37, no. 22, pp. 6538-6550, 2022. <https://doi.org/10.1080/10106049.2021.1943009>
- [33] L. Wang, G. Jin, X. Xiong, H. Zhang, and K. Wu, "Object-based automatic mapping of winter wheat based on temporal phenology patterns derived from multitemporal Sentinel-1 and Sentinel-2 imagery," *ISPRS International Journal of Geo-Information*, vol. 11, no. 8, p. 424, 2022. <https://doi.org/10.3390/ijgi11080424>
- [34] W. Chen and G. Liu, "A novel method for identifying crops in parcels constrained by environmental factors through the integration of a Gaofen-2 high-resolution remote sensing image and Sentinel-2 time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023. <https://doi.org/10.1109/JSTARS.2023.3329987>
- [35] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMI 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, 2018: Springer*, pp. 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [36] W. Song et al., "Hierarchical extraction of cropland boundaries using Sentinel-2 time-series data in fragmented agricultural landscapes," *Computers and Electronics in Agriculture*, vol. 212, p. 108097, 2023. <https://doi.org/10.1016/j.compag.2023.108097>
- [37] A. O. Onojeghuo, Y. Miao, and G. A. Blackburn, "Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery," *Remote Sensing*, vol. 15, no. 6, p. 1517, 2023. <https://doi.org/10.3390/rs15061517>
- [38] Oktay, Ozan, et al. "Attention u-net: Learning where to look for the pancreas." *arXiv preprint arXiv:1804.03999*(2018). <https://doi.org/10.7717/peerj-cs.1767/fig-2>
- [39] Wang, Xuqi, Shanwen Zhang, and Ting Zhang. "Crop insect pest detection based on dilated multi-scale attention U-Net." *Plant Methods* 20.1 2024. <https://doi.org/10.1186/s13007-024-01163-w>
- [40] Cao, Hu, et al. "Swin-unet: Unet-like pure transformer for medical image segmentation." *European conference on computer vision*. Cham: Springer Nature Switzerland, 2022. https://doi.org/10.1007/978-3-031-25066-8_9
- [41] P. K. Buttar and M. K. Sachan, "Semantic segmentation of satellite images for crop type identification in smallholder farms," *The Journal of Supercomputing*, vol. 80, no. 2, pp. 1367-1395, 2024. <https://doi.org/10.1007/s11227-023-05533-4>
- [42] He, Sheng, et al. "Segmentation ability map: Interpret deep features for medical image segmentation." *Medical image analysis* 84 (2023): 102726. <https://doi.org/10.1016/j.media.2022.102726>
- [43] M. Saadat, S. T. Seydi, M. Hasanlou, and S. Homayouni, "A Convolutional Neural Network Method for Rice Mapping Using Time-Series of Sentinel-1 and Sentinel-2 Imagery," *Agriculture*, vol. 12, no. 12, p. 2083, 2022. <https://doi.org/10.3390/agriculture12122083>
- [44] A. M. Obeso, J. Benois-Pineau, A. Á. R. Acosta, and M. S. G. Vázquez, "Architectural style classification of Mexican historical buildings using
- [17] R. Saini and S. K. Ghosh, "Crop classification on single date sentinel-2 imagery using random forest and support vector machine," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 683-688, 2018. <https://doi.org/10.5194/isprs-archives-XLII-5-683-2018, 2018>
- [18] G.-H. Kwak et al., "Combining 2D CNN and bidirectional LSTM to consider spatio-temporal features in crop classification," *Korean Journal of Remote Sensing*, vol. 35, no. 5_1, pp. 681-692, 2019. <https://doi.org/10.7780/kjrs.2019.35.5.1.5>
- [19] Y. Wang, L. Feng, Z. Zhang, and F. Tian, "An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 199, pp. 102-117, 2023. <https://doi.org/10.1016/j.isprsjprs.2023.04.002>
- [20] J. Lee and T. Warner, "Segment based image classification," *International Journal of Remote Sensing*, vol. 27, no. 16, pp. 3403-3412, 2006. <https://doi.org/10.1080/01431160600606866>
- [21] M. I. H. Shihab, N. Tasnim, H. Zunair, L. K. Rupty, and N. Mohammed, "VISTA: Vision transformer enhanced by U-Net and image colorfulness frame filtration for automatic retail checkout," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 3183-3191. <https://doi.org/10.1109/cvprw56347.2022.00359>
- [22] J. Yao and S. Jin, "Multi-category segmentation of Sentinel-2 images based on the Swin UNet method," *Remote Sensing*, vol. 14, no. 14, p. 3382, 2022. <https://doi.org/10.3390/rs14143382>
- [23] B. Niu, Q. Feng, B. Chen, C. Ou, Y. Liu, and J. Yang, "HSI-TransUNet: A transformer based semantic segmentation model for crop mapping from UAV hyperspectral imagery," *Computers and Electronics in Agriculture*, vol. 201, p. 107297, 2022. <https://doi.org/10.1016/j.compag.2022.107297>
- [24] M. Gao, T. Lu, and L. Wang, "Crop Mapping Based on Sentinel-2 Images Using Semantic Segmentation Model of Attention Mechanism," *Sensors*, vol. 23, no. 15, p. 7008, 2023. <https://doi.org/10.3390/s23157008>
- [25] A. Joshi, B. Pradhan, S. Gite, and S. Chakraborty, "Remote-sensing data and deep-learning techniques in crop mapping and yield prediction: A systematic review," *Remote Sensing*, vol. 15, no. 8, p. 2014, 2023. <https://doi.org/10.3390/rs15082014>
- [26] A. H. Khan, Z. Zafar, M. Shahzad, K. Berns, and M. M. Fraz, "Crop Type Classification using Multi-temporal Sentinel-2 Satellite Imagery: A Deep Semantic Segmentation Approach," in *2023 International Conference on Robotics and Automation in Industry (ICRAI)*, 2023: IEEE, pp. 1-6. <https://doi.org/10.1109/ICRAI57502.2023.10089586>
- [27] S. T. Seydi, M. Amani, and A. Ghorbanian, "A dual attention convolutional neural network for crop classification using time-series Sentinel-2 imagery," *Remote Sensing*, vol. 14, no. 3, p. 498, 2022. <https://doi.org/10.3390/rs14030498>
- [28] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48-62, 2021. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [29] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30, (2017). <https://doi.org/10.7717/peerjcs.1946/fig-10>
- [30] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12077-12090, 2021. <https://doi.org/10.1109/cvpr52688.2022.00134>

Heidari Gholanlo M at Department of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran

✉ majid.heidari.gholanlo@gmail.com



رضا جوانمرد علی تپه (دکتری ۲۰۱۶)

استادیار دانشگاه علم و فناوری مازندران ایران است. او از سال ۲۰۱۹ به عنوان محقق همکار با دانشگاه تکنیکو لیسبون در زمینه بیورباتیک همکاری داشته است. در سال ۲۰۱۹ به عنوان پژوهشگر در

آزمایشگاه هوش مصنوعی تصویری دانشگاه آکسفورد بروکس (انگلستان) در زمینه خودروهای خودران فعالیت داشت. او دکترای خود را در زمینه هوش مصنوعی و رباتیک از دانشگاه فدرال میناس گرای (UFMG)، برزیل به اتمام رساند. علاقه‌مندی‌های پژوهشی او شامل رباتیک، یادگیری ماشین و بینایی ماشین می‌باشند.

Javanmard Alitappeh, R. Assistant Professor at Department of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran

✉ Rezavanmard@mazust.ac.ir



عبادت قنبری پرمهر دارای مدرک دکتری

تخصصی مهندسی نقشه برداری (گرایش فتوگرامتری) از دانشگاه ملبورن استرالیا می‌باشد. پس از اخذ دکتری، ایشان به مدت دو سال به عنوان پژوهشگر پسادکتر در دانشگاه RMIT استرالیا و از سال ۱۳۹۵ تا کنون به

عنوان استادیار در گروه مهندسی نقشه برداری دانشگاه صنعتی نوشیروانی بابل مشغول به فعالیت بودند. ایشان تا کنون موفق به چاپ بیش از ۳۰ مقاله در مجلات و کنفرانس‌های معتبر بین‌المللی شده‌اند. زمینه‌های تخصصی ایشان عبارتند از: فتوگرامتری، سنجش از دور، پردازش تصاویر، کاربرد یادگیری ماشین و هوش مصنوعی در کشاورزی و سنجش از دور.

Parmehr, E. G. Assistant Professor at Department of Geomatics Engineering, Babol Noshirvani University of Technology, Babol, Iran

✉ parmehr@nit.ac.ir

deep convolutional neural networks and sparse features," Journal of Electronic Imaging, vol. 26, no. 1, pp. 011016-011016, 2017. <https://doi.org/10.1117/1.jei.26.1.011016>

[45] Mountrakis, Giorgos, Jungho Im, and Caesar Ogole. "Support vector machines in remote sensing: A review." ISPRS journal of photogrammetry and remote sensing 66.3 (2011): 247-259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>

[46] Torbunov, Dmitrii, et al. "Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023. <https://doi.org/10.1109/wacv56688.2023.00077>

[47] T. Lu, M. Gao, and L. Wang, "Crop classification in high-resolution remote sensing images based on multi-scale feature fusion semantic segmentation model," Frontiers in Plant Science, vol. 14, p. 1196634, 2023. <https://doi.org/10.3389/fpls.2023.1196634>

[48] Chen, Jieneng, et al. "Transunet: Transformers make strong encoders for medical image segmentation." arXiv preprint arXiv:2102.04306 (2021). <https://doi.org/10.20944/preprints202411.2377.v1>

[49] Nikouei, Seyed Yahya, et al. "Real-time human detection as an edge service enabled by a lightweight cnn." 2018 IEEE International Conference on Edge Computing (EDGE). IEEE, 2018. <https://doi.org/10.1109/edge.2018.00025>

[50] Zhang, Anjun, et al. "Region level SAR image classification using deep features and spatial constraints." ISPRS journal of photogrammetry and remote sensing 163 (2020): 36-48. <https://doi.org/10.1016/j.isprsjprs.2020.03.001>

[51] K. Borooshan, S. Behzadi, "Detection of Rice Seedlings in UAV Image Using DenseNet Network", Journal of Remote Sensing and Geoinformation Research (JRSGR), 2023. <https://doi.org/10.3390/rs15010196>

[52] E. Fernandes, P. Wildemberg, J. dos Santos, "Water Tanks and Swimming Pools Detection in Satellite Images: Exploiting Shallow and Deep-Based Strategies", Anais do XVI Workshop de Visão ComputacionalDO, Brasil, 2020. <https://doi.org/10.5753/wvc.2020.13491>

معرفی نویسندگان

AUTHOR(S) BIOSKETCHES



مجید حیدری قولانلو در سال ۱۳۹۶ مدرک

کارشناسی مهندسی برق گرایش کنترل را از دانشگاه پیام نور مشهد اخذ نموده و پس از آن در سال ۱۴۰۳ کارشناسی ارشد هوش مصنوعی را از دانشگاه علم و فناوری مازندران دریافت کرده است. زمینه تخصصی ایشان بینایی ماشین، تصاویر ماهواره‌ای و پردازش تصویر می‌باشد.

Citation (Vancouver): HeidariGholanlo M, Javanmard Alitappeh R, Ghanbari Parmehr E. [Large Scale Crop Type Mapping Using Multi Temporal Sentinel-2 Imagery and a New Attention U-Net-Vit]. *J. RS. GEOINF. RES.* 2024; 2(2): 233-246

doi <https://doi.org/10.22061/jrsgr.2024.11047.1071>



COPYRIGHTS

© 2024 The Author(s). This is an open-access article distributed under the terms and conditions of the Creative Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>)