Research Article

Agricultural Engineering, 47(1) (2024) 129-142
DOI: 10.22055/agen.2024.45449.1697

ISSN (E): 2588-526X
ISSN (P): 2588-5944

Weed detection in cereal fields using convolutional neural network based on deep learning

A. Taheri Hajivand[1*], K. Shirini[2] and S. Samadi Gharehveran[3]

1. Department of Biosystems Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
2. Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University Of Tabriz, Tabriz, Iran
3. Department of Electrical Engineering, Faculty of Electrical and Computer Engineering, University Of Tabriz, Tabriz, Iran

Received: 23 November 2023 Accepted: 25 February 2024 *Corresponding Author: a.taheri@tabrizu.ac.ir

Abstract

Introduction: In many countries, on average, more than 50% of people's food comes from grains, and nearly 70% of the cultivated area of one billion hectares of the world is dedicated to grains. A variety of weeds grow along with cereals in the fields, which can reduce crop yield due to competition for light, water, and nutrients. To eliminate weeds accurately and with minimal problems, in time detection with high accuracy and speed is required to be done.

In the agriculture, controlling and eliminating of weeds in grain fields are done. Weeds are one of the most important factors affecting the performance of agricultural products, which are their most important competitors in conventional agriculture. The farmers spray the entire field to eliminate weeds, while weeds appear scattered and patchy in the field; which shows the necessity of using precise operatons to solve this type of heterogeneity. In addition to causing economic damage, the conventional method of fighting can cause pollution of the environment and even the human food chain. Researches show that the losses caused by pests, diseases, and weeds can reach 40% of the global crop every year and it is predicted that this percentage will increase significantly in the coming years. Besides, according to the research of Goktoan et al., the annual cost of weeds for the Australian economy is estimated to be around $4 billion as a loss in agricultural income.

Materials and Methods: Among the new methods for weeds detection is the use of machine vision technology and related methods such as deep learning object detection algorithms and convolutional neural networks (CNN). The steps related to the implementation of the method include preparing data for training and evaluating networks, using new object detection algorithms, using different convolutional neural networks with different characteristics to extract image features in algorithms, and using the feature pyramid network (FPN) method in object detection algorithms. The output of the networks are evaluated in terms of the number of detections, the exact location of detection, and the time of detection in the field. the transform model (ViTs) is based on the transformer architecture that was originally developed for NLP tasks. Transformers use self-awareness mechanisms that allow the model to capture complex relationships between elements in a sequence. In the case of ViTs, sequence elements are image patches. In using the transformer architecture for visual data, it divides the image into small and non-interfering parts. Each patch typically consists of a grid of pixels. These patches are considered the "words" of the image sequence. Spatial embeddings are added to image patches to provide spatial information to the model. Spatial embeddings are necessary because transformers do not have built-in notions of order or spatial relationships. ViTs use multi-series self-awareness mechanisms to capture relationships between different image patches, and the representation of each patch is updated by attention to other patches. Data separation is very important in data watch transformers for two reasons a) the model needs data to learn,

© 2024 The Author(s). Published by Shahid Chamran University of Ahvaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

and b) data needing to measure the model because the model may not be able to extract the information correctly.
Results and Discussion: The best network in terms of positioning accuracy was the ViTs with an average accuracy of 0.95. In addition to this, the network considered in this research managed to recognize 503 of the 535 target weeds, means that our network can recognize 95% of the weeds. The presented method has been able to reach the highest accuracy compared to other existing methods and has been able to detect existing grasses in the much shorter period. Compared to other methods, the reset50 algorithm has been able to detect more than 88%, although its execution time was about 2.5 times that of the proposed method. In comparing the efficiency of algorithms, execution time is as important as accuracy. By making comparisons and considering 70% of the data as training data and 30% as test data, the presented algorithm has been able to detect weeds in the field with an accuracy of over 90% in just 13 seconds.
Conclusion: In this study, recent pre-trained architectures, techniques, and models for weed detection in high-density cereal fields were compared. Different convolutional neural networks were used as model feature extractors; Because deep convolutional network architectures provide better performance and enable faster application development. In addition, a method to solve the problem of weed images with high density has been proposed to increase various parameters of accuracy in detecting and locating weeds. The results of this research show that the best performance in terms of detection time is related to the presented algorithm, which can also be used in mobile phones. The Vision Transfer model has high accuracy in detecting weeds in dense and complex farm environments. The proposed model is interesting and acceptable in terms of performance and reliability. The weed detection rate of the proposed network is more than 95%. In future work, the presented models can be combined with other networks and methods to improve the detection of weed species in field conditions. It is also possible to use other models and networks in the field of deep learning, such as pixel detection and threshold-based intersection.

Keywords: Optimization, Accuracy, Agriculture, Weed, Deep learning.

مهندسی زراعی (مجله علمی کشاورزی) جلد ۴۷ شماره۱، بهار ۱۴۰۳

مقاله پژوهشی

تشخیص علف هرز در مزارع غلات با استفاده شبکه عصبی کانولوشنی مبتنی بر یادگیری عمیق

عادل طاهری حاجی وند [۱*]، کیمیا شیرینی [۲] و سینا صمدی قره ورن [۳]

۱- استادیار گروه بیوسیستم، دانشکده کشاورزی ، دانشگاه تبریز، ایران
۲- دانشجوی دکتری مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، ایران
۳- دانشجوی دکتری مهندسی برق، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، ایران

تاریخچه مقاله

دریافت: ۱۴۰۲/۰۹/۰۲

پذیرش نهایی: ۱۴۰۲/۱۲/۰۶

کلمات کلیدی:

بهینه‌سازی،

دقت،

کشاورزی،

علف هرز،

یادگیری عمیق.

* عهده دار مکاتبات:

Email: a.taheri@tabrizu.ac.ir

چکیده

در بسیاری از کشورها به طور میانگین بیش از ۵۰ درصد غذای مردم از غلات تامین شده و غلات نزدیک به ۷۰ درصد سطح زیر کشت جهان را تشکیل می‌دهند. علف‌های هرز مختلفی همراه با غلات در مزارع رشد می کنند که عملکرد محصول را به دلیل رقابت برای نور، آب و مواد مغذی کاهش داده است. برای از بین بردن علف‌های‌هرز به صورت دقیق و با حداقل مشکلات، باید تشخیص به موقع، با دقت و سرعت بالا صورت گیرد. در این مقاله از داده‌هایی که شامل ۲۱۰ تصویر علف هرز در یک قطعه زمین استفاده شد. این تصاویر با استفاده از دوربین Canon EOS R5 که دوربین با کیفیتی است، گرفته شده و برای برچسب‌گذاری تصاویر از ابزار برچسب گذاری مبتنی بر زبان پایتون استفاده شد.از جمله روش‌های نوین در این حوزه، استفاده از فناوری یادگیری عمیق و روش‌های مرتبط با آن نظیر الگوریتم‌های تشخیص اشیاء، یادگیری عمیق و شبکه‌های عصبی کانولوشنی است. مراحل مربوط به انجام پروژه شامل آماده‌سازی داده‌ها جهت آموزش و ارزیابی شبکه‌ها از شبکه‌های عصبی کانولوشنی مختلف با خصوصیات متفاوت برای استخراج ویژگی‌های تصاویر می‌باشد. این مراحل از روش شبکه هرمی ویژگی در الگوریتم‌های تشخیص اشیا بهره می‌گیرند. خروجی شبکه‌ها از نظر تعداد تشخیص، مکان دقیق تشخیص و زمان تشخیص در مزرعه مورد بررسی و ارزیابی قرار گرفتند. بهترین شبکه از نظر دقت مکان‌یابی، مدل ترنسفورم با میانگین دقت ۹۵ /۰ شناخته شد. علاوه بر این نتایج نشان می‌دهد شبکه مدنظر توانست از میان ۵۳۵ علف هرز، ۵۰۳ مورد را به درستی تشخیص دهد.

مقدمه

تحقیقات و پیش‌بینی‌ها نشان می‌دهند که تا سال ۲۰۵۰ میلادی، جمعیت جهان به تعداد حدود ۱۰ میلیارد نفر خواهد رسید و به طبع آن نیاز به مواد غذایی و محصولات کشاورزی نیز افزایش خواهد یافت؛ بنابراین، تولیدات کشاورزی باید حدود ۷۰ درصد افزایش یابند این نیازها تامین شوند. کشاورزی با چالش‌های متعددی مثل تغییرات آب‌وهوا و کمبود منابع آب شیرین مواجه است (۲۶). بنابراین، با بهره‌گیری از فناوری‌های مدرن و روش‌های مدیریت مکانی خاص، کشاورزان می‌توانند با کشاورزی دقیق مشکلات را حل یا کاهش دهند.

علف‌های هرز یکی از مهم‌ترین عوامل مؤثر بر کاهش میزان تولید محصولات کشاورزی هستند. همین امر باعث شده است که در حوزه کشاورزی کنترل و ازبین‌بردن علف‌های هرز در مزارع غلات امری مهم به شمار بیاید. یکی از روش‌های مرسوم مبارزه با علف‌های هرز سم‌پاشی می‌باشد. این روش با اینکه مزیت‌هایی دارد اما علاوه بر واردکردن آسیب‌های اقتصادی می‌تواند باعث آلودگی محیط‌زیست و حتی زنجیره غذایی انسان شود. تحقیقات نشان دهنده این است که تلفات ناشی از آفات، بیماری‌ها و علف‌های هرز می‌تواند به ۴۰ درصد محصول تولیدی جهان در هر سال برسد. پیش‌بینی شده است که این درصد در سال‌های آینده به میزان قابل‌توجهی افزایش یابد (۱۸). کشاورزی دقیق را می‌توان به‌عنوان ابزار راهبردی برای مدیریت بهینه صنعت کشاورزی شامل بذرها و کودها دانست. به دلیل کشت وسیع غلات، عوامل خسارت‌زا همواره محصولات را به طور جدی تهدید و مورد حمله قرار می‌دهند. پرداختن به این امر با توجه به این‌که هزاران سال است غلات در تأمین غذای انسان نقش به سزایی داشته و علاوه بر این حدود ۴۰ درصد کالری مصرفی مردم از غلات به‌خصوص گندم تأمین می‌شود، امری بسیار مهم است (۲). در این راستا، استفاده از زمینه‌های تحقیقاتی جدید نقش مهمی را برای جبران خسارت‌های ناشی از وجود علف‌های هرز، آفات و بیماری‌ها و در نتیجه افزایش تولید ایفا می‌کند. اغلب غلات گیاهانی یک ساله بوده که دانه‌های ریز آنها مصرف خوراکی دارند و دارای دو نوع گرمسیری و سردسیری می‌باشند. از گونه‌های گرمسیری می‌توان به برنج، ذرت، ارزن و از گونه‌های سردسیری به گندم، جو و چاودار اشاره کرد (۱۹).

روش‌های مرسوم برای کنترل و ازبین‌بردن علف‌های هرز به‌صورت مکانیکی و شیمیایی طبقه‌بندی می‌شوند. روش‌های مکانیکی شامل استفاده از ماشین‌های مخصوص کشاورزی از قبیل انواع کولتیواتورها، وجین‌کن‌ها، روش‌های شعله افکنی و روش‌های شیمیایی نیز شامل استفاده از انواع مواد شیمیایی با ماشین‌های پخش‌کن و سم‌پاش‌ها می‌باشند.

تشخیص مکان علف‌های هرز در مزارع با سرعت و دقت بالا می‌تواند پیش‌نیاز اساسی برای کنترل آن‌ها به‌صورت شیمیایی باشد که از آن به‌عنوان سم‌پاشی نقطه‌ای یاد می‌شود. چرا که کاهش استفاده سموم با تشخیص دقیق و درست علف هرز رابطه مستقیم دارد. تشخیص مکان دقیق علف‌های هرز نیاز به انواع علف‌کش‌ها را کاهش داده و اثرات مخرب را از بین می‌برد. بدیهی است که در کشاورزی در سطح گسترده در مزارع کنونی امکان پایش و مدیریت فردی هر گیاه از نظر وجود و عدم وجود علف هرز بدون استفاده از فناوری‌های مدرن امکان‌پذیر نخواهد بود (۱۹).

با پیشرفت هوش مصنوعی در حوزه بینایی رایانه به‌ویژه یادگیری عمیق[1]، تشخیص و کنترل انواع آفات در مزارع کشاورزی پیشرفت چشم‌گیری دارد.در این زمینه، تحقیقات بسیاری به خصوص در زمینه پاشش بهینه علف‌کش‌ها و همچنین استفاده از یادگیری عمیق برای

1- Deep learning

تشخیص آفت‌های موجود در مزرعه انجام گرفته است (۷، ۱۱، ۱۴، ۲۱ و ۲۳).

شبکه‌های عصبی قدرت پردازشی بالا و همچنین امکان آموزش مجموعه بزرگی از تصاویر در زمان کمتری دارد و همچنین نتایج به‌دست‌آمده به انتخاب استخراج‌کننده‌های ویژگی وابسته نیست. پژوهش دوس سانتوس و همکاران[1] (۱۰) از مهم‌ترین مطالعاتی است که درباره تشخیص علف‌های هرز با استفاده از یادگیری عمیق و شبکه‌های عصبی کانولوشنی صورت‌گرفته است. هدف از این تحقیقات تشخیص دو نوع مختلف از علف هرز در مزرعه‌های سویا بوده است که با استفاده از علف‌کشی صورت‌گرفته بود. نتایج بادقت ۹۸٪ در طبقه‌بندی رده‌ها به دست آورده شده است (۱۱).

الگوریتم Detect Net از دیگر روش‌های استفاده شده برای تشخیص علف‌های هرز در مزارع غلات است. این الگوریتم دو بخش شبکه کانولوشنی و تابع خوشه‌بندی را شامل می‌شود که طی این الگوریتم با برچسب‌گذاری تصاویر شبکه کانولوشنی آموزش داده شده تا تلفات این آموزش را به کمترین میزان خود و دقیق‌ترین مختصات برسانند. در مطالعات دیرمان و همکاران[2]، یک دقت ۸۷٪ از این الگوریتم حاصل شد (۱۲).

معماری VGG16 و ResNet5 و inceptionv3 از جمله معماری‌های محبوب و پرکاربرد اخیر هستند که در مطالعه احمد و همکاران[3] (۱۳) در کلاس‌بندی علف‌های هرز مزارع ذرت و سویا استفاده نیز شده است. طی این تحقیقات که عملیات پیاده‌سازی آنها در چهارچوب‌های کراس و پایتورچ صورت گرفت و از مدل YOLOV3 برای تشخیص علف‌های هرز موجود در این مزرعه‌ها استفاده شد، بهترین عملکرد مربوط به شبکه VGG16 بادقت ۹۸,۹۰٪ بوده است که این عملکرد بالاترین سرعت آموزش و دقت را در چهارچوب پایتورچ نشان می‌دهد (۱).می‌توان از شبکه تک مرحله‌ای YOLO به عنوان الگوریتم تشخیص گونه‌های مختلف علف‌های هرز در مزرعه‌های گندم استفاده کرد. این روش نسبت به سایر روش‌ها هم از نظر زمان هم از نظر هزینه بسیار قابل اتکا بوده و همچنین امکان کنترل از راه دور از طریق رایانه یا تلفن همراه را نیز دارد.همان‌طور که گفته شد، امروزه روش‌های یادگیری عمیق جایگزین روش‌های یادگیری ماشین و روش‌های سنتی شده است که برای مدیریت علف‌های هرز و سم‌پاشی در مزارع بسیار کاربردی هستند. همچنین مدل‌های یادگیری عمیق مثل شبکه‌های عصبی کانولوشن مبتنی بر ناحیه RCNN و دیگر مدل‌ها بسیار محبوب و بسیار کارآمد هستند. هدف از مقاله حاضر تشخیص علف‌های هرز موجود در مزارع غلات با استفاده از الگوریتم ترنسفورمرهای دیدبان است. بدین منظور از داده‌های جمع‌آوری شده توسط دوربین از مزرعه غلات برای آموزش و ارزیابی استفاده می‌شود.

مواد و روش‌ها

در یادگیری عمیق مهم‌ترین مورد آماده‌سازی مجموعه‌داده‌ها است که در اجرای هر مدل یادگیری عمیق لازم و ضروری است. ایده اصلی این است که تصاویر را به‌عنوان دنباله‌های پچ‌ها[4] موردتوجه قرار دهند و از مکانیزم‌های خودتوجهی[5] برای پردازش آن‌ها در مدل ترانسفورمرهای بینایی[6] (ViTs) استفاده کنند.

تصاویر از نقطه‌های مشخص در تاریخ مشخص و در برخی نقاط در یک تاریخ دیگر از مزرعه گرفته شده‌اند. برای برچسب گذاری تصاویر از ابزار برچسب گذاری مبتنی بر زبان پایتون استفاده شد. در حوزه تشخیص اشیا از برچسب گذاری تصاویر گرافیکی با عنوان برچسب

1- Santos Ferreira *et al*
2- Dyrmann *et al*
3- Ahmad *et al*
4- Patches
5- Self-attention
6- Vision transformers

گذاری استفاده می‌شود که تصاویر را در کلاسهای مختلف از هم جداسازی می‌کند. در روش برچسب گذاری پس از برچسب گذاری تصویرها فایل‌های مربوطه با قالب xml به وجود می‌آید که برخی از اطلاعات راجع به هر تصویر را داراست. اطلاعات مربوط به فایل‌ها باید شامل قالب، اندازه، نوع تصویر، تعداد و مختصات مستطیل‌های محیطی مربوط به علف‌های هرز باشد. شکل ۱ نمونه ای از دیتای استفاده شده را نشان می‌دهد.

جمع‌آوری داده‌ها

جمع‌آوری داده‌ها شامل ۲۱۰ تصویر علف هرز در یک قطعه زمین می‌باشد که با استفاده از دوربین Canon EOS R5 با رزولوشن تقربی بینابینی ۳۰ مگاپیکسل از ارتفاع ۳۰ سانتی ، انجام شد.

ترنسفورمرهای دیدبان

ترنسفورمرهای دیدبان یک نوع معماری شبکه‌های عصبی عمیق هستند که برای کاربردهای بینایی ماشین و تحلیل تصاویر استفاده می‌شوند. این معماری، مبتنی بر تبدیل است که در ابتدا برای پردازش متون مورداستفاده قرار می‌گرفته و به‌تازگی برای وظایف دیدبانی نیز به کار گرفته شده است.

ViTs[1] بر مبنای معماری ترنسفورمر است که ابتدا برای وظایف NLP [2] توسعه داده شده است. ترنسفورمرها از مکانیزم‌های خود توجهی استفاده می‌کنند

که به مدل امکان می‌دهد روابط پیچیده بین عناصر در یک دنباله را به دست آورد. در مورد ViTs، عناصر دنباله پچ‌های تصویر هستند. در استفاده از معماری ترنسفورمر برای داده‌های بصری تقسیم تصویر به بخش‌های کوچک و بدون تداخل است. از ViTs برای تشخیص علف‌های هرز از گیاهان در این پژوهش استفاده شده است. برای این کار، ابتدا یک مجموعه‌داده متشکل از تصاویر مختلف از گیاهان و علف‌های هرز آماده شده و سپس ViTsرا بر روی این داده‌ها آموزش داده شدند تا بتواند علف‌های هرز را از گیاهان تشخیص دهد. روش کار ViTs برای تحلیل تصاویر به این صورت است که تصاویر را به توکن‌های ورودی تبدیل می‌کند و از این توکن‌ها برای توجه به اطلاعات مختلف در تصویر استفاده می‌کند. با استفاده از بلاک‌های تبدیل[3]، اطلاعات تصویر به صورت مکانی در نظر گرفته می‌شود و سپس برای تصمیم‌گیری درباره دسته‌بندی تصویر، از لایه‌های کاملاً متصل[4] یا لایه‌های خروجی استفاده می‌شود. مزیت اصلی استفاده از ViTs برای تشخیص علف‌های هرز از گیاهان این است که این مدل‌ها قابلیت یادگیری الگوهای پیچیده در تصاویر را دارند و می‌توانند بدون نیاز به ویژگی‌های دستی یا پیش‌پردازش‌های مختلف، ویژگی‌های مفید را از داده‌های ورودی استخراج کنند.

هر پچ به طور معمول شامل شبکه‌ای از پیکسل‌ها است. این پچ‌ها به‌عنوان "کلمات" دنباله تصویر در نظر گرفته می‌شوند. تعبیه‌های مکانی به پچ‌های تصویر اضافه می‌شوند تا اطلاعات مکانی به مدل ارائه شود. تعبیه‌های مکانی ضروری هستند چرا که ترنسفورمرها مفاهیم ترتیب یا روابط مکانی را دارا نیستند. ترنسفورمرهای دیدبان از مکانیزم‌های خودتوجهی چندسری استفاده می‌کنند تا روابط بین پچ‌های مختلف تصویر را ایجاد کنند و هر پچ توسط توجه به پچ‌های دیگر به‌روزرسانی می‌شود. تفکیک داده در ترنسفورمرهای دیدبان به دو دلیل حائز اهمیت است، نیازمندی به داده‌های آموزشی برای یادگیری و نیازمندی به داده‌های آزمایش جهت سنجش مدل.

قابل‌تعمیم بودن یک مدل به این معناست که مجموعه‌داده آموزش‌دیده روی مجموعه دیگری آزمایش شود، یعنی مدل قابلیت تطبیق و واکنش مناسب

1- Vision transformers
2- Natural language processing
3- Transformer blocks
4- Fully connected layers

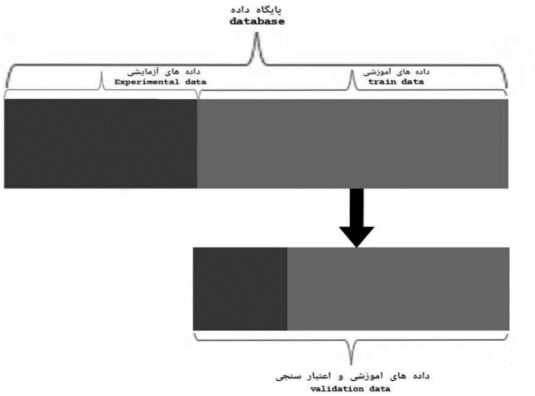
نسبت به داده‌های قبلی و جدید را داشته باشد. روش برگزاری یکی از کارآمدترین روش‌ها برای تفکیک داده‌ها است که زمانی مورداستفاده قرار می‌گیرد که مجموعه‌داده‌ها محدود کم باشند. پایگاه‌داده به دو مجموعه آموزش و آزمایش در روش برگزاری تقسیم می‌شود که نسبت داده‌های آموزش به آزمایش معمولاً به‌صورت ۷۰ درصد به ۳۰ درصد می‌باشد.

در این مطالعه، داده‌های مربوط به سال ۲۰۱۹ که حدود ۳۰ درصد از کل داده‌ها بودند به عنوان داده آزمایشی و مابقی داده‌ها برای آموزش مورد استفاده قرار گرفتند. از میان همه داده‌های آموزشی ۳۰درصد نیز برای اعتبارسنجی شبکه در نظر گرفته شد.

داده‌افزایی

کم‌بودن تعداد داده‌ها یکی از بزرگ‌ترین چالش‌ها در آموزش مدل‌ها و شبکه‌های تشخیص اشیا است. وقتی یک مدل آموزش داده می‌شود پارامترهای آن را باید به‌گونه‌ای تنظیم کرد که بتواند ورودی خاص را به خروجی تبدیل کند. بهینه‌سازی باهدف تعقیب نقطه مطلوب است که تلفات مدل در آن کم باشد و زمانی این نتیجه حاصل می‌شود که پارامترهای مدل به‌درستی تنظیم شده باشند. شبکه‌های عصبی مدرن معمولاً دارای چندین میلیون پارامتر هستند در نتیجه لازم است تعداد مناسبی از نمونه‌ها به مدل آموزش داده شود تا نتایج مورد قبولی حاصل گردد؛ چراکه تعداد پارامتر موردنیاز باتوجه‌به پیچیدگی مدل باید متناسب باشد. از روش‌های مرسوم افزایش داده می‌توان به تبدیلات هندسی و تبدیلات شدت اشاره نمود.

شکل (۱) نمونه ای از تصاویر موجود در پایگاه داده
Figure (1) An example of the images in the data base



شکل (۲) روش اجرا
Figure (2) The hold out method

فراخوانی مجموعه‌داده

مجموعه‌داده‌ها با اندازه اصلی ۵۱۸۴ × ۳۴۵۶ پیکسیل و الگوریتم‌های آموزشی بر روی پردازنده گرافیکی NVIDIA QUADRO 14000 RTX، با استفاده از پلت فرم برنامه‌نویسی موازی CUDA®، در سیستم عامل لینوکس به کار گرفته شدند.کتابخانه پایتورچ، به عنوان یکی از چهارچوب‌های کاری قدرتمند در حوزه یادگیری عمیق ابزارهای زیادی را برای آسانتر کردن بارگذاری داده‌ها و نحوه خوانش آنها در محیط برنامه نویسی پایتون ارائه کرده‌است. پایگاه داده‌هایی مانند MNIST و CIFAR به کمک این ابزارها فراخوانی می‌شوند. فراخوانی مجموعه داده‌های سفارشی نیازمند پیاده سازی الگوریتمی برای خواندن تصاویر و مشخصات آنها می‌باشد.

ساخت و پیکربندی مدل

در محیط برنامه‌نویسی پایتون و چهارچوب پایتورچ برای ساخت مدل می‌توان با ایجاد کلاس و توابع مختلف در آن لایه‌های شبکه را ایجاد نمود. مدل‌های از پیش آموزش‌دیده شده بر روی مجموعه‌داده‌های معیار بزرگی مانند ImageNet، به‌عنوان منبعی برای جامعه یادگیری عمیق بسیار مفید هستند. این مدل‌ها نقش اصلی را در پیشرفت سریع تحقیقات بینایی رایانه‌ای ایفا می‌کنند. استفاده از یک مدل آموزش‌دیده شده بر روی مجموعه‌داده‌های مختلف باعث می‌شود که شبکه سریع‌تر همگرا شود. ترنسفورمرهای دیدبان نیازمند مجموعه‌داده‌های بزرگی برای پیش آموزش هستند و معمولاً بر روی مجموعه‌های بزرگ تصاویر آموزش می‌بینند تا ویژگی‌های مفیدی را یاد بگیرند. بعد از پیش آموزش، می‌توانند روی مجموعه‌داده‌های کوچک‌تر و وظایف خاص تنظیم شوند.

مجموعه تصاویر برچسب‌دار با اندازه اصلی ۵۱۸۴×۳۴۵۶ پیکسل به‌عنوان ورودی برای آموزش این مدل‌ها استفاده شده‌اند. آموزش تمام مدل‌ها با استفاده از سیستم‌عامل Ubuntu انجام گردیده است. برای آموزش داده‌ها از سخت‌افزار شامل پردازنده مرکزی AMD Ryzen 73800 8-core و پردازنده گرافیکی NVIDIA QUADRO RTX 4000-8GB استفاده شده است. برای انجام آموزش از پلت فرم برنامه‌نویسی موازی CUDA استفاده شد. همه مدل‌ها بر روی چهارچوب کاری پایتورچ پیاده‌سازی شده‌اند.

بهینه‌سازی

بهینه‌سازی شبکه برای یک دوره آموزشی شبکه و تحویل آن به یک تابع ضروری است. برای بهینه‌سازی از یک کلاس پایتونی با توابع مختلف برای محاسبه خطا استفاده شده است و تابع بهینه‌ساز پس از محاسبه خطای شبکه به‌وسیله تابع اتلاف وارد عمل شده و سعی در به‌روزرسانی پارامترهای آموزشی شبکه یعنی وزن و بایاس می‌کند. برای نمونه با کمک مشتق‌گیری می‌توان به نقطه کمینه تابع اتلاف رسید تا از مقدار اتلاف تاحدامکان کاسته و آنرا به صفر نزدیک کرد. تابع بهینه-ساز به پارامترهایی مانند مومنتوم و نرخ یادگیری نیز وابسته است. مومنتوم مسیر را برای حرکت تابع بهینه‌ساز هموارتر می‌کند. نرخ یادگیری سرعت حرکت تابع بهینه‌ساز را مشخص می‌کند و با داده‌های ورودی هر دوره آموزش رابطه مستقیم دارد. شکل ۳ برای نمایش این تابع است.

پارامترهای قابل‌تنظیم برای آموزش

مقداردهی پارامترها بر اساس آزمون‌وخطا و با درنظرگرفتن آموزش هرچه بهتر و بهینه‌تر شبکه انتخاب شده‌اند. پارامترها شامل دوره یا تکرار آموزش، داده‌های ورودی هر دوره و تابع بهینه‌ساز هستند. مقادیر پارامترهای قابل‌تنظیم در جدول ۱ ارائه شده‌اند. برای ارزیابی دقت پیش‌بینی و توانایی شبکه‌های عصبی در تشخیص نمونه‌ها، از پارامتر دقت استفاده گردیده است. همچنین باتوجه‌به نتایج، زمان مناسب برای کنترل علف‌های هرز با حمایت شبکه برآورد گردید.

روند کلی تشخیص اشیا به کمک یادگیری عمیق و شبکه‌های عصبی کانولوشنی در شکل ۴ نشان داده شده است.

معیارهای ارزیابی برای تشخیص اشیا

متداول‌ترین معیار ارزیابی شبکه‌های تشخیص اشیا مبتنی بر یادگیری عمیق، دقت Precision و یادآوری Recall و Accuracy است. در این پژوهش جهت بررسی دقت مدل ارائه شده از شاخص دقت استفاده شد. رابطه ۱ نشان‌دهنده نحوه محاسبه دقت است.

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

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (۲)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (۳)$$

آموزش و ارزیابی شبکه

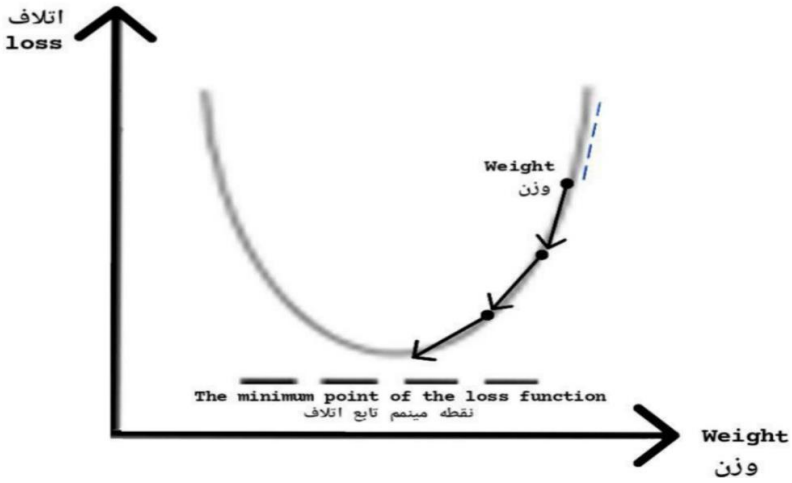
در یادگیری ماشین آخرین مرحله آموزش و ارزیابی شبکه است که مهم‌ترین رکن آن است. در این پژوهش از حلقه پایتونی برای انجام عملیاتی که برای یک دوره آموزش و ارزیابی نیازمند است، استفاده شده است. عمده کار این حلقه دریافت دسته‌ای داده‌های آموزشی و تغذیه آن به شبکه و سپس گذراندن از توابع اتلاف و بهینه‌ساز برای به‌روزرسانی پارامترها است.

همچنین ارزیابی شبکه نیز توسط کتابخانه pycocotools انجام گرفت. برای ارزیابی شبکه از پارامترهای میانگین دقت و میانگین فراخوانی استفاده گردید. میانگین دقت یک معیار عددی است که هم دقت و هم یادآوری را محصور می‌کند و منحنی دقت – فراخوانی را با میانگین دقت در مقادیر فراخوانی از ۰ تا ۱ خلاصه می‌کند. میانگین دقت برابر با مساحت زیر منحنی است.

نتایج عملکرد شبکه ترنسفورمرهای دیدبان

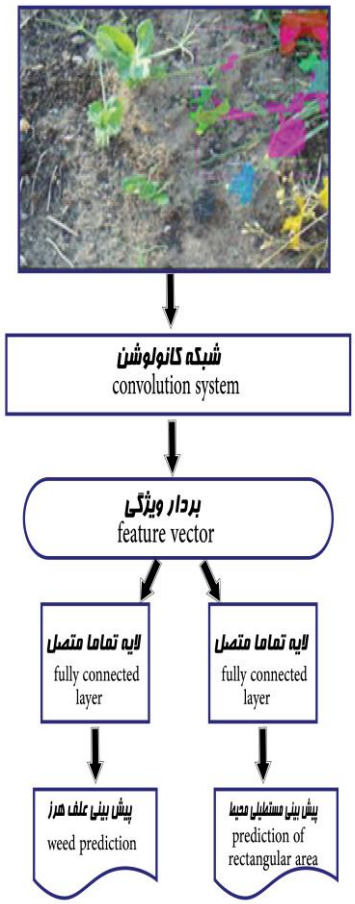
ترنسفورمرهای دیدبان عملکردی رقابتی با شبکه‌های عصبی کانولوشنی در وظایف مختلف بینایی رایانه‌ای نشان داده‌اند. آن‌ها قادر به محدودکردن وابستگی‌های بلندمدت در تصاویر هستند و بنابراین برای وظایفی که نیاز به درک متناسب با کلیات تصویر دارند مناسب هستند.

ترنسفورمرهای دیدبان بسیار قابل‌تطبیق هستند و می‌توانند به طیف گسترده‌ای از وظایف بینایی رایانه‌ای با تغییرات معمول در معماری به کار روند. آموزش آنها به منابع محاسباتی قابل‌توجهی نیاز دارد؛ زیرا پردازش تعداد زیادی از پچ‌های تصویر است. انتخاب اندازه پچ و هایپرپارامترهای مدل می‌تواند بر عملکرد مدل تأثیر بگذارد. ترنسفورمرهای دیدبان ممکن است در وظایفی که نیاز به جزئیات دقیق دارند، به‌عنوان‌مثال به دلیل تقسیم‌بندی تصاویر به پچ‌ها، با مشکل مواجه شوند و برخی اطلاعات را از دست دهند. آن‌ها جایگزینی برای رویکردهای مبتنی بر CNN هستند و توانایی پردازش تعداد گسترده‌ای از وظایف بینایی را با بهره‌گیری از توانایی معماری ترنسفورمر ارائه می‌دهند. شبکه‌های عصبی پیچشی محدودیت‌هایی مانند هزینه محاسباتی بالا و نیاز به مجموعه‌داده‌های برچسب‌دار بزرگ دارند. مدل‌های خودتوجهی مانند ترنسفورمرهای دیدبان می‌توانند جایگزینی باشند که با این محدودیت‌ها مقابله کنند. مدل‌های ترنسفورمرهای دیدبان تصاویر را به‌عنوان دنباله‌ای از قطعه‌ها پردازش کرده و از کدگذارهای ترانسفورمر و خودتوجهی برای درک روابط بین قطعه‌ها استفاده می‌کنند. این امکان یادگیری با نمونه‌های برچسب‌دار کمتر نسبت به شبکه‌های عصبی پیچشی را فراهم می‌کند.



شکل (۳) بهینه سازی تابع اتلاف

Figure (3) Optimization of the loss function



شکل (٤) روند کلی تشخیص اشیا به کمک شبکه عصبی کانولوشن

Figure (4) The schematic of object recognition using convolutional neural network

شکل (۵) تشخیص شبکه ترنسفورمر دیده‌بان
Figure (5) Detection of the watchdog transformer network

جدول (۱) تنظیم پارامترها
Table (1) Setting of the parameters

پارامتر **Parameter**	**مقدار** **Value**
دوره آموزش Training course	۳۰۰۰ تکرار 3000 iterations
داده‌های ورودی هر دوره (دسته) Input data of each course (Category)	۲ دسته Two category
تابع بهینه‌ساز Optimizer function	نزول گرادیان تصادفی SGD (Stochastic gradient descent)
مونتوم Montum	۰/۸۳
نرخ آموزش Training rate	۰/۰۰۳۲

جدول (۲) ارزیابی عملکرد مدل با سایر الگوریتم‌ها
Table (2) Evaluation of the model performance with other algorithms

زمان تشخیص (ثانیه) Detection time (s)	میانگین دقت Accuracy	صحت Precision	پوشش Recall	مدل Model
13.3	0.95	0.93	0.94	Vision transfer
26	0.87	0.83	0.84	VGG
32	0.84	0.75	0.76	شبکه عصبی کانولوشنی Convolutional neural network
33	0.74	0.71	0.77	Detect net
41	0.77	0.71	0.70	VGG net و google net
28	0.88	0.81	0.82	ResNet50
27	0.78	0.77	0.76	RCNN

نتایج و بحث

مقدار میانگین دقت و زمان اجرا این مدل در جدول ۲ نشان داده شده است. شکل ۳، تشخیص این مدل را بر روی تصویر انتخاب شده برای ارزیابی بصری را نشان میدهد.

همان‌طور که از جدول ۲ مشخص است روش ارائه شده توانسته است به بالاترین دقت در مقایسه با سایر روش‌های موجود برسد و توانسته در مدت زمان بسیار کمتر علفهای موجود را تشخیص دهد. در مقایسه با سایر روش‌ها الگوریتمreset50 نیز قادر به تشخیص بالای ۸۸ درصد بوده است، هرچند زمان اجرای آن حدود ۲,۵ برابر روش ارائه شده است. در مقایسه کارایی الگوریتم‌ها زمان اجرا نیز به‌اندازه دقت حائز اهمیت است. با مقایسه-های انجام شده و درنظرگرفتن ۷۰ درصد از داده‌ها به عنوان داده آموزش و ۳۰ درصد به عنوان داده آزمایش، الگوریتم ارائه شده توانسته است تنها در ۱۷ ثانیه علف‌های هرز موجود در مزرعه را با دقت بالای ۹۰ درصد تشخیص دهد.

نتیجه‌گیری

در این مطالعه معماری‌ها، تکنیک‌ها و مدل‌های اخیر از پیش آموزش‌دیده شده برای شناسایی علف‌های هرز در مزارع غلات با تراکم بالا، بررسی مقایسه شده است. شبکه عصبی کانولوشن مختلف به‌عنوان استخراج‌کننده ویژگی مدل‌ها، مورداستفاده قرار گرفتند؛ چراکه معماری‌های عمیق شبکه کانولوشن عملکرد بهتری را ارائه می‌دهند و توسعه سریع‌تر برنامه را امکان‌پذیر می‌سازند. علاوه بر این روشی برای حل مشکل تصاویر علف‌های هرز با تراکم بالا، به‌منظور افزایش پارامترهای مختلف دقت در تشخیص و مکان‌یابی علف‌های هرز پیشنهاد شده است. برای تشخیص علف‌های هرز در محیط‌های مزرعه پیچیده، مدل پیشنهادی از نظر عملکرد و قابلیت اطمینان بسیار مناسب است. علف‌های هرز در مزارع را می‌توان به‌صورت زمان واقعی و از طریق انواع مختلف مدل‌ها شناسایی کرد و حتی می‌تواند یک تکنیک کلیدی برای تشخیص برخط علف‌های هرز با پهپاد باشد.

نتایج نشان می‌دهند که بهترین عملکرد از نظر زمان تشخیص مربوط به الگوریتم ارائه شده است که قابلیت استفاده در تلفن‌های همراه را نیز دارد. مدل Vision transfe، دقت بالایی در تشخیص علف‌های هرز در محیط‌های مزرعه پرتراکم و پیچیده دارد. مدل پیشنهادی از نظر عملکرد و قابلیت اطمینان جالب و قابل‌قبول است. میزان شناسایی علف‌های هرز شبکه پیشنهادی با اطمینان بالای ۹۵ درصد است. به عنوان کارهای آتی می‌توان مدل‌های ارائه شده را با سایر شبکه‌ها و روش‌ها برای بهبود تشخیص گونه‌های علف هرز در شرایط مزرعه ترکیب کرد. همچنین می‌توان از دیگر مدل‌ها و شبکه‌ها در حوزه یادگیری عمیق، از نوع تشخیص پیکسل‌ها و تقطیع مبتنی بر آستانه استفاده کرد.

References

1. Ahmad, A., Saraswat, D., Aggarwal, V., Etienne, A., and Hancock, B. 2021. Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. Computers and Electronics in Agriculture, 184, 106081.

2. Ahmadi, M., Hooshmand, A. R., BoroomandNasab, S., and Sharifi, M. A. 2019. Calibration and Validation of WOFOST Model for Wheat in Qazvin Plain. Iranian Journal of Soil and Water Research, 50(2) 329-338. (In Persian with English Abstract.)

3. Alom, M. Z. Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin M. S., and Asari, V. K. 2019. A state- of-the-art survey on deep learning theory and architectures. Electronics, 8(3): 292 .

4. Alzubaidi, L., Zhang, J., Humaidi. A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O.,... and Farhan, L. 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8(1): 1-74.

5. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

6. Asad, M. H., and Bais, A. 2020. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. Information Processing in Agriculture, 7(4): 535-545.

7. Ashok Kumar, D., and Prema, P. 2013. A review on crop and weed segmentation based on digital images. Multimedia processing, communication and computing applications, 213: 279-291 .

8. Barker J, Sarathy S and Tao A .2016. "DetectNet Deep Neural Network for Object Detection in DIGITS". Nvidia, (retrieved: 2016-11-30) .

9. dos Santos Ferreira, A., Freitas, D. M., da Silva, G. G., Pistori, H., and Folhes, M. T. 2017. Weed detection in soybean crops using Cony Nets. Computers and Electronics in Agriculture, 143: 314-324 .

10. Dyrmann, M., Jørgensen, R. N., and Midtiby, II. S. 2017. RoboWeedSupport-Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. Advances in Animal Biosciences, 8(2): 842-847.

11. Eyre, M. D., Critchley, C. N. R., Leifert, C., and Wilcockson, S. J. 2011. Crop sequence, crop protection and fertility management effects on weed cover in an organic،conventional farm management trial. European Journal of Agronomy, (3)153-162.

12. Ferentinos, K. P. 2018. Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture. 145: 311-318 .

13. Gao, J., French, A. P., Pound, M. P., He, Y., Pridmore, T. P., and Pieters, J. G. 2020. networks for image-based Convolvulus sepium detection in sugar beet fields. Plant Methods, 16(1): 1-12 .

14. Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., and De, D. 2020. Fundamental concepts of convolutional neural network. In Recent trends and advances in artificial intelligence and Internet of Things, 172: 519-567.

15. Gianessi, L. P., and Reigner, N. P. 2007. The value of herbicides in US crop production. Weed Technology, 21(2): 559-566 .

16. Gupta, G. R., Oomman, N., Grown, C., Conn, K., Hawkes, S., Shawar, Y. R., ... and Darmstadt, G. L. 2019. Gender equality and gender norms: framing the opportunities for health. The Lancet, 393(10190) : 2550-2562.

17. Hamuda, E., Glavin, M., and Jones, E. 2016. A survey of image processing techniques for plant extraction and segmentation in the field. Computers and electronics in agriculture, 125: 184-199 .

18. He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

19. Hernández-Hernández, J. L., Garcia-Mateos, G., González-Esquiva, J. M., Escarabajal-Henarejos, D., Ruiz-Canales. A., and Molina-Martinez, J. M. 2016. Optimal color space selection method for plant،soil segmentation in agriculture. Computers and Electronics in Agriculture, 122: 124-132 .

20. Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In international conference on machine learning,37: 448-456.

21. Islam, N., Rashid, M. M., Wibowo, S., Xu, C. Y., Morshed, A., Wasimi, S. A.,... and Rahman, S. M. 2021. Early weed detection using image processing and machine learning techniques in an Australian chilli farm. Agriculture, 11(5): 387.

22. Jabir, B., and Falih, N. 2022. Deep learning-based decision support system for weeds detection in wheat fields. International Journal of Electrical and Computer Engineering, 12(1): 816 .

23. Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

24. Le. V. N. T. Truong, G., and Alameh, K. 2021. Detecting weeds from crops under complex field environments based on Faster RCNN. Eighth International Conference on Communications and Electronics, 21: 350-355.

25. Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., and Moscholios, I. 2020. A compilation of UAV applications for precision agriculture. Computer Networks, 172, 107148.

26. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. 2016. Rethinking the inception architecture for computer vision. IEEE conference on computer vision and pattern recognition, 2818-2826.