CAMBRIDGE
UNIVERSITY PRESS

DATA PAPER

Methodological approaches to identifying and mapping fields of specific crops on a basis of high-resolution satellite images using phenological, geographic, and regional statistical information

Alexey Unagaev[1] [ID], Irina Korotkova[1] and Natalia Efremova[2]

[1]Deep Planet Ltd, London, UK
[2]School of Business and Management, Queen Mary University of London, London, UK
Corresponding author: Natalia Efremova; Email: n.efremova@qmul.ac.uk

Abstract

Currently, methods for mapping agricultural crops have been predominantly developed for a number of the most important and popular crops. These methods are often based on remote sensing data, scarce information about the location and boundaries of fields of a particular crop, and involve analyzing phenological changes throughout the growing season by utilizing vegetation indices, e.g., the normalized difference vegetation index. However, this approach encounters challenges when attempting to distinguish fields with different crops growing in the same area or crops that share similar phenology. This complicates the reliable identification of the target crops based solely on vegetation index patterns. This research paper aims to investigate the potential of advanced techniques for crop mapping using satellite data and qualitative information. These advanced approaches involve interpreting features in satellite images in conjunction with cartographic, statistical, and climate data. The study focuses on data collection and mapping of three specific crops: lavender, almond, and barley, and relies on various sources of information for crop detection, including satellite image characteristics, regional statistical data detailing crop areas, and phenological information, such as flowering dates and the end of the growing season in specific regions. As an example, the study attempts to visually identify lavender fields in Bulgaria and almond orchards in the USA. We test several state-of-the-art methods for semantic segmentation (U-Net, UNet++, ResUnet). The best result was achieved by a ResUnet model (96.4%). Furthermore, the paper explores how vegetation indices can be leveraged to enhance the precision of crop identification, showcasing their advanced capabilities for this task.

Impact Statement

The paper presents advanced methodological approaches and applications of state-of-the-art machine learning algorithms for data collection and mapping of agricultural crops widely used in the food and beauty industries, based on remote sensing data. These approaches use qualitative information about the physical properties of the Earth's surface vegetation, which can appear on satellite and aerial images, and are intended to improve the efficiency of crop mapping, which is usually based on vegetation indices. The proposed methodology of data

collection is based on the combined analysis of remote sensing data with statistical and environmental information and can be useful in particular for mapping crops in regions that differ from each other in geographical and environmental conditions.

1. Introduction

There is an extensive literature on methods for crop mapping and estimating yields of the most common and popular crops using remote sensing data over large territories with scarce ground truth data to calibrate models or validate generated maps (Wardlow et al., 2007; Brinkhoff et al., 2020; Zhong et al., 2015). Currently, methods for identifying agricultural fields often use knowledge about phenological phenomena, which are unique to a particular crop. Phenology[1] is the study of sequential, periodic seasonal events in the life cycle of animal and plant species (such as flowering or leaf fall) and how these events are affected by seasonal and interannual changes in climate and habitat factors. Phenological information is reflected in differences between seasonal variation curves of vegetation indices, such as normalized sifference vegetation index (NDVI) or enhanced vegetation index (EVI), calculated using satellite time series data, and approaches to crop mapping on the basis of this information show quite a high efficiency. Previous studies have shown that such differences are clearly observed for winter and spring crops (Zhong et al., 2015; Masialeti et al., 2010; Belgiu and Csillik, 2018).

There are two features, representing phenological information, that are particularly useful for the accurate identification of certain crops. The first one is the crop calendar, the schedule of the crop growing season from the land preparation to crop establishment, maintenance, and harvest. This tool can be applied for the current season or for several consecutive seasons. The second feature is growing degree days (GDD), an important agroclimatic indicator characterizing the limits of favorable thermal conditions for the development of plants during the season. These two features are widely used to supplement information about the onset of a specific phenological phase, which is useful for more accurate identification of certain crops, especially between winter crops, as well as subsequent yield assessments. The most popular and commonly mapped vegetation cover objects in agricultural research are wheat, maize, sunflower, and rapeseed[2]. For example, Roumenina et al. (2015) used 100 and 300 m Proba-V data to identify 4 crops (wheat, sunflower, rapeseed, and maize) in Bulgaria; Potgieter et al. (2007) and Qiu et al. (2017) used enhanced vegetation indices—EVI and EVI2— calculated from MODIS data, to detect fields of different winter crops in Australia and China respectively; Skakun et al. (2017) used the MODIS NDVI and GDD information time series to obtain apriori knowledge of wheat vegetation in Kansas (USA) and Ukraine, and to estimate yields.

By the end of the second decade of the 21st century, satellite monitoring and agricultural mapping capabilities had improved significantly with the launch of the Sentinel-1 and Sentinel-2 satellites. These satellites provide unique data in the optical and radar ranges, with average revisit rates of 3–6 days and a high resolution of 10 meters per pixel for optical imagery and 5 meters for radar imagery. These satellites were launched in 2015, allowing us to accumulate significant volumes of data. A large volume of data allows us to analyze the condition and assess the yield of the most important crops. Long time series of satellite imagery quantitatively describe seasonal and long-term changes of vegetation cover, providing rich information for comparative analysis of crop conditions across different years. Such a vast amount of imagery can also serve as training data for accurate yield estimates using ML methods.

At the same time, there is a number of difficulties that complicate the process of mapping agricultural crop fields, as well as many other types of vegetation cover, in an arbitrarily chosen territory for both past and current growing seasons. The most common problem is the lack of data or difficulty in collecting field data, such as information about the presence and characteristics of a specific crop at a particular time interval in a certain area. This can be combined with uncertainty related to annual crop rotation or the young age of trees. Annual crop rotation results in a limited time window and therefore a reduced number

---

[1] Encyclopedia Britannica.
[2] Eurostat.

of cloud-free images suitable for mapping annual crops, which causes difficulty in accurately identifying crops for a particular year. New or young crop plantings often cannot be distinguished from other crops due to poorly developed vegetation. Collecting such data remains a laborious and expensive task, particularly in areas with poor transportation and road infrastructure. Therefore, the most important goals of land cover mapping are to improve the efficiency and cost of the field data collection phase and search for alternative sources of relevant information about crop growth in a specific region. Other difficulties are related to remote sensing data and arise from the lack of suitable cloud-free and very high-resolution images and the lack of research and statistical information related to the vegetation cover of the study area.

Despite the increasing volume of high-resolution remote sensing data and sources of detailed statistical information, there is a perceptible lack of highly specialized research on vegetation cover and agricultural crop mapping which takes into account not only vegetation indices but also traditional interpretation signs containing qualitative knowledge about the objects. These signs include features, representing spectral and spatial information: color, tone, shape, texture, and dimensions of objects (Jensen, 2007; Lillesand et al., 2015; Labutina, 2004; Gur'eva et al., 1976). These features allow us to identify crops without additional data collection or to reduce its duration to a minimum.

The contributions of this work are three-fold. First, we explore how the physical characteristics of plants can appear in a satellite image in a natural color or other color band combinations, so that plantings of crops can be reliably identified. Second, we determine the potential range of industrial crops that can be identified on satellite images in different regions in addition to the most popular crops using interpretation signs of satellite images. Finally, we provide an assessment of how accurately crops of interest can be mapped using qualitative information, how it can supplement the methods, based on analysis of vegetation indices, and how efficient are advanced ML algorithms for the accurate mapping of the crops of our interest.

2. Materials and methods

2.1. Subjects of crop detection and mapping

In this work, we investigated several crops, which are widely used in food and beauty industries. We chose crops that belong to different families (Lamiaceae, Rosaceae, and Poaceae) and represent different plant life forms for mapping in order to define the main object interpretation signs and characteristics that contribute to confident identification of the places of cultivation of such crops. We consider lavender (*Lavandula angustifolia*) in Bulgaria and France, almonds (*Prunus dulcis*) in the USA and Australia, and winter barley (*Hordeum vulgare*) in Russia. These crops were selected due to their lower popularity and total area occupied, in comparison to the world's most widespread or similar analogs.

2.2. Materials

In this paper, we used various sources of information facilitating the further process of identifying plantings of specific crops on satellite images. These sources included reports of state ministries of agriculture and commercial analytical organizations; research papers and cartographic materials; web publications about the cultivation of different crops in specific regions and areas occupied by each crop by administrative unit (Skakun et al., 2017); the information about the shape and the average size of fields; the information about approximate annual period of flowering and typical harvest dates (Zipper, 2017).

We also used different sources of information, which helped us narrow down the search area of crop fields and plantings. Some of such sources included websites and addresses of organizations, specializing in the production and marketing of related agricultural products, and travel brochures. Other useful sources, that provided us the information about the shape, size, and colour of crop plantings in a given area and the timing of the beginning of different phenological phases in a specific year, were internet blogs; user photos; and videos (Wu et al., 2021). For the search and validation of farms, fields, and plantings near settlements, we used Google Maps and other mapping services (Esri, Bing, Mapbox, OpenStreetMap) as well as the Google Street View service. For automatic detection of fields across larger areas, we used multitemporal

satellite images of Sentinel-2 obtained from the EO Browser [3], covering Dobrich Province (Bulgaria), Provence (France), California (USA), Victoria (Australia), and Krasnodar Krai (Russian Federation) for some selected years between 2017 and 2022.

2.3. Data processing

All the information that helps to determine the presence of a specific crop in a particular area using satellite images can generally be attributed to two groups—information on the *physical properties* of an object (biometric: height and shape of plants, seasonal changes in their colors; shape and size of plantings) and *statistical information* (the area of the territory occupied by a specific plant species in this region and its dynamics; approximate harvest dates for different crops, etc.). Therefore, the appearance of the plant, the color, shape, and size of the fields are the primary characteristics that are reflected in the satellite image and contribute to the visual identification of agricultural fields and their distinction from natural vegetation, and at the same time, statistical data supplements this information, specifying the growing area and season periods for search fields.

The main visual feature that simplifies the task of identification of fields and plantings occupied by a particular crop on the satellite image is color, which is characterized by the appearance of the plant community, when the dominant plant defines it is color at a certain part of the season. Such visual appearance of the plant community during the flowering period is one of the most effective indicators that could be monitored using satellite images of medium and high resolution (John et al., 2020). One of the efficient ways to accurately identify a crop can be to compare the spectral brightness values in the image and digital photographs in the R, G, and B bands of the visible or infrared ranges, taken in the field on dates close to each other or at identical time periods of different years with the same illumination conditions (the same time of day and weather) and by the same phenological phase of the crop, which together result in similar field colors in the satellite images. For more accurate analysis, it is best to choose images that cover the stages of flowering, wilting, harvesting, etc.

For more effective mapping of fields of a specific crop, it is important to understand the patterns of changing geographical and environmental conditions from area to area, which determine the phenology of each specific crop (Yang et al., 2020). It is also crucial to consider historical, cultural, and economic factors and agricultural traditions that determine the popularity of a particular crop in different regions or countries. These factors can be manifested in individual variations in the appearance and dimensions of plants, different sizes, shapes, structures, and colors of fields, as well as shifts in the timing of harvesting from region to region.

A promising approach to identifying perennial crops and distinguishing similar crops from each other can also be the analysis of differences in the color of plantings at different stages of plant development. On the one hand, satellite images that cover only the period when plants are juvenile, can make it difficult to identify such plantings; but on the other hand, when using time-series images over many years, makes it possible to evaluate how the color of plantings of different crops changes from year to year, which helps to distinguish similar crops from each other. Additionally, the graphs of annual changes and peak values of vegetation indices for several seasons can be analyzed. For example, Chen, Jin, and Brown used Landsat time-series data to determine the planting year and stage of development of almond trees based on the annual maximum values of the NDVI index (Chen et al., 2019a).

2.4. Visual identification of lavender fields in Bulgaria

Using the proposed methodology, we collected data for three main crops: lavender, almonds, and barley. For the lavender identification task, we selected several growing areas in Bulgaria and France, the world's leading countries in the cultivation of lavender for perfumery. The primary source of information about areas of lavender cultivation is a map, demonstrating spatial differences in the patterns of distribution of

---

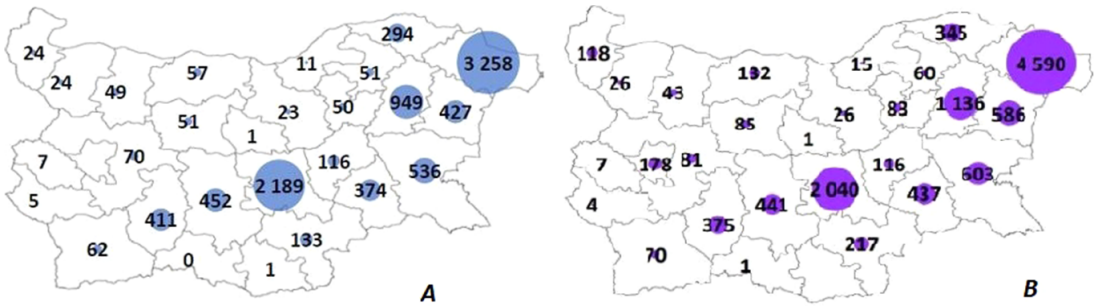[3] Sinergise Ltd., https://apps.sentinel-hub.com/eo-browser

Figure 1. *Areas (hectares) of lavender fields in Bulgaria in 2017 (A) and 2018 (B) according to the State Fund Agriculture and Ministry of Agriculture of Bulgaria (Vylkanov, 2018; InteliAgro, 2019).*

lavender fields. Using the maps of areas, where lavender is cultivated in Bulgaria (Figure 1), it is possible to find out which regions are leaders in lavender production and estimate the share of areas occupied by these fields in relation to the area of all agricultural land. It is also possible to analyze the temporal dynamics of field areas over several years, which will also be reflected in the time series of satellite images through the emergence of new fields or vice versa, rotating lavender with other crops. Therefore, statistical information is an important source of data for tracking changes in areas relative to other crops on spatial and time scales.

Figure 1 shows the leading region in Bulgaria in lavender cultivation in 2018: Dobrich Province, located in the north-east. In this region, the area of lavender growth increased by more than 1300 hectares within a year (about 30%), making up almost half of the total area under this crop in the country. Another large area of lavender production is Stara Zagora Province (central part), with 2040 ha of lavender fields. In this part, however, the area of lavender fields has decreased within a year by more than 100 ha. Knowing the total area occupied by lavender plantings in these areas in a particular year, we can find all fields in the satellite imagery and investigate how effective a particular feature is for their identification. At the same time, to search for new fields of lavender, to which the total area of plantings increases, as mentioned above, it is useful to determine the characteristics of young plants related to their size and shape, and how these features can appear on satellite images. In this case, to reliably identify new plantings, it could be helpful to use very high-resolution satellite imagery for small areas of the region to find some fine details such as row structure, row spacing and other features that are not visible in the lower resolution imagery. Figure 2a–2c shows the visual appearance of lavender fields in Bulgaria in photos and satellite images.

2.5. Visual identification of lavender fields in France

We used maps of tourist routes in lavender-growing areas in Provence to identify lavender fields in France (Figure 3), along with user photographs (Figure 4) and Google Street View, which were used to pinpoint the location of the fields on a high-resolution Google Earth base map.

The colors of the lavender fields both in the photograph and in the satellite images of Bulgaria (Figures 2a–2c) are visually close to each other, which gives reason to interpret these fields with high accuracy as lavender. And although during the entire period of lavender flowering, the color of these fields changes from violet to purple, on some dates it is possible to find many similar areas. The average value of the color intensity according to the RGB color model, calculated in a 3×3 pixel cell, corresponding to a part of a lavender flower taken from one of the rows in the photograph, is 175, 179, and 170, respectively, which produces a magenta/purple color. In the Google image, the average R, G, and B values calculated on a randomly selected lavender field are 95, 73, 75, and in the Sentinel-2 image—100, 65, 65, that is, these values also make up very close shades of magenta.

The color of lavender fields in satellite images for France differs significantly from the color in photograph (see Figure 4a): for example, in the Sentinel-2 image, it is darker, with a predominance of violet shades, whereas in the Google image, it is lighter than on the Sentinel-2 (Figures 4b and 4c).
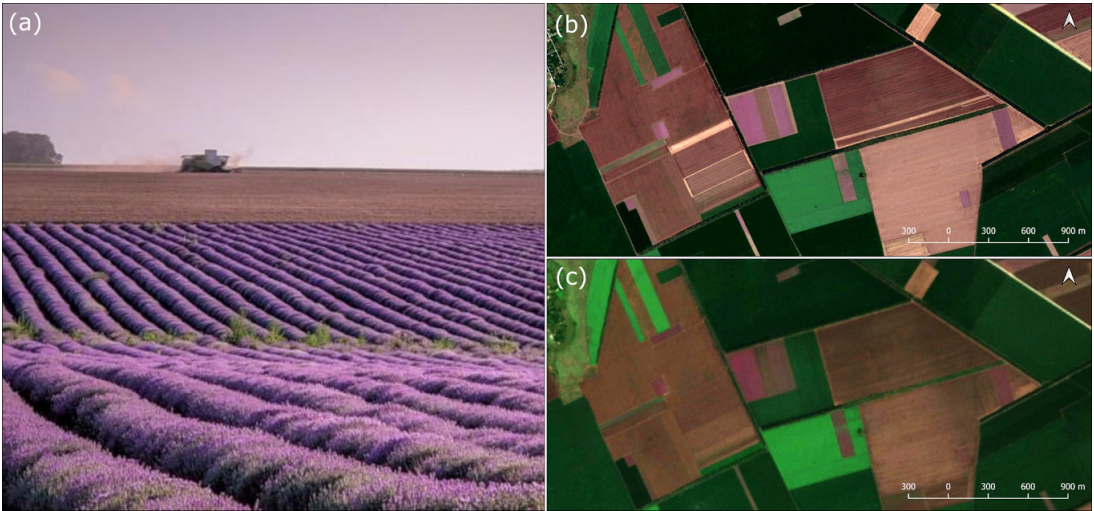
Figure 2. *(a) Lavender field, Dobrich Province, Bulgaria. Photo: (Nedkova, 2021), (b) lavender fields (purple colors), Dobrich Province, Bulgaria. Google Earth. 02.07.2020, (c) lavender fields (purple colors), Dobrich Province, Bulgaria. Sentinel-2 image, true colors. 26.06.2020*
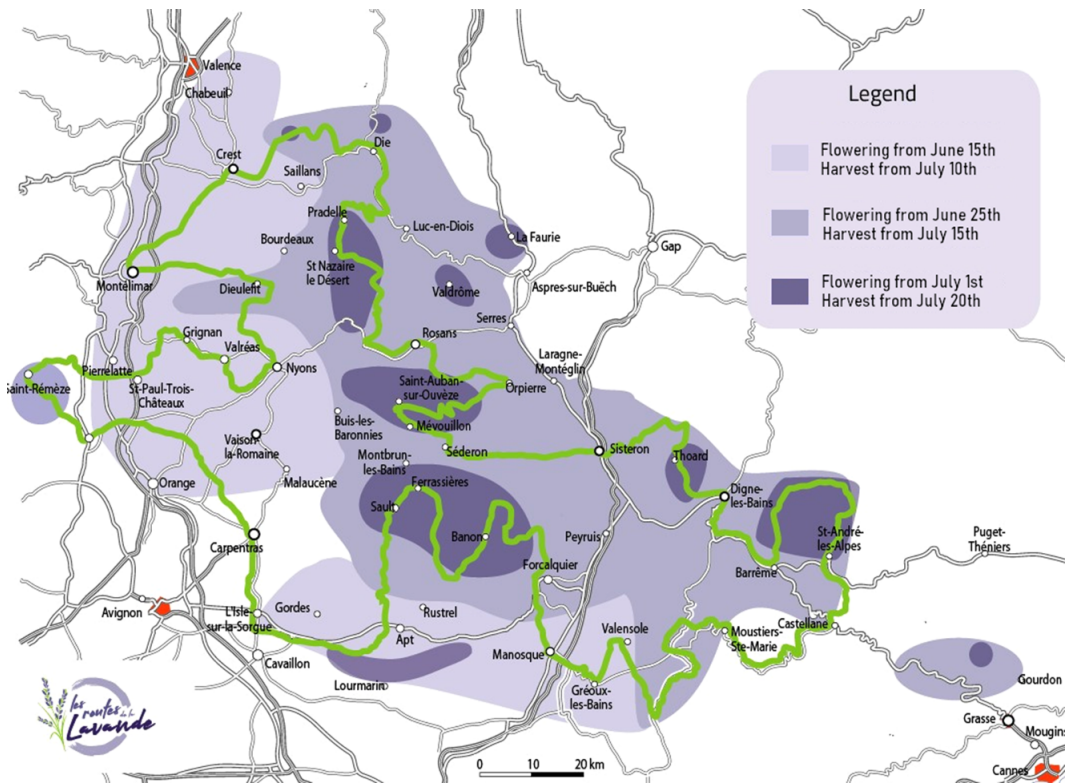

Figure 3. *Map of lavender fields in Provence, France by flowering dates (Bobrowski, n.d.).*
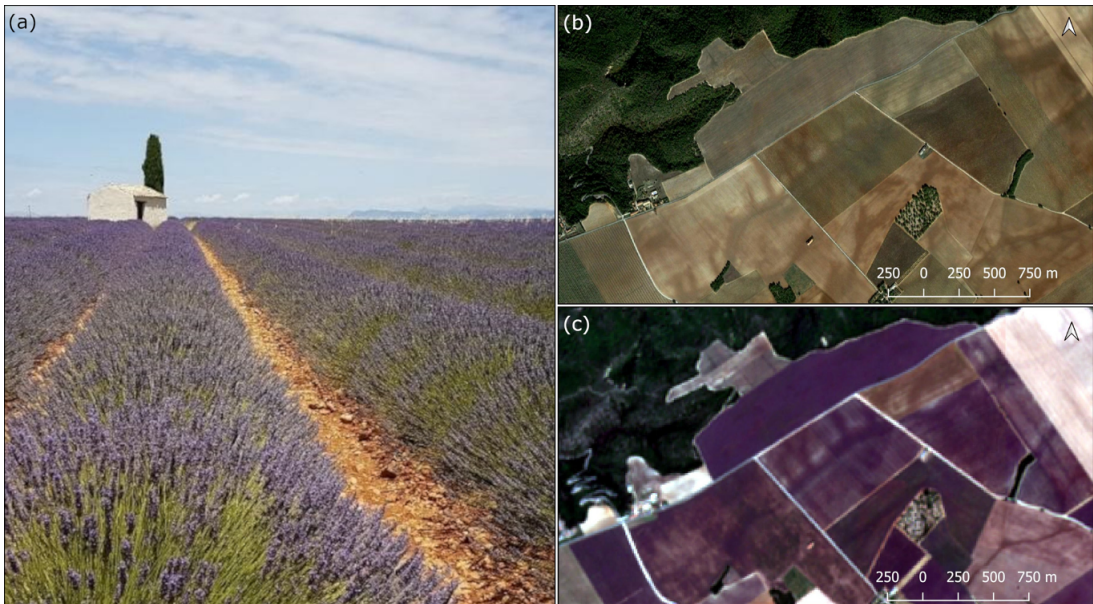
Figure 4. *(a) Lavender field, Provence, France. Photo: Anton Gvozdikov, Google Earth, (b) lavender fields (dark-gray color), Provence, France. Google Earth. 03.05.2016, (c) lavender fields (purple color), Provence, France. Sentinel-2 image, true colors. 03.07.2020.*

The Google image clearly shows that the color of the fields is different and not uniform, with bare soil visible between the individual rows of lavender. In general, the prevalence of the "dirtier" violet colors of lavender fields in the Sentinel-2 image for France compared to the purple colors of the fields in Bulgaria can probably be attributed to the mixing of purple and green colors during the harvest season. This may be due to the local traditions of lavender cultivation, which manifests themselves in the selection of lavender varieties suitable for local climatic and soil conditions, as well as the larger average field size and possibly wider row spacings compared to those in Bulgaria (by simple GIS-based measurements, we found that the average area of an individual lavender field in the Dobrich Province of Bulgaria equals 3.6 ha, whereas in Provence the average area of the field is almost twice higher—6.9 ha).

2.6. *Visual identification of almond orchards in the USA and Australia*

We chose almonds as a second test crop to understand the problems of tree crop planting mapping. The location of the almond orchards was taken from advertisements and catalogs for the sale of the orchards in the United States.

The color of almond orchards during the bloom period is visually similar to the color on the photograph (Figure 5a) and satellite images (Figures 5b and 5c). However, the color of blooming mature almond trees cannot be the only feature that allows the identification of almond orchards of all ages. Since almonds begin to bloom and bear fruit only a few years after they were planted, it is also necessary to search not only for mature trees but also for all new plantings of young trees using statistical information. That means that, by knowing the total area of almond plantings in any territory, we can try to find all plantings of both young and adult trees of all possible varieties. For example, the total area of bearing and non-bearing almonds in California in 2020 amounted to 1.6 million acres, while in the previous season (2019) it was 1.52 million acres (USDA, 2021), that is, it increased by 80,000 acres. At the same time, the area of non-bearing almonds has increased by more than 10,000 acres compared to 2019. The methodology for the identification of young non-bearing almond plantings implies the use of a different method based on the vegetation indices, which will be discussed in a special section of this paper. The approach mentioned
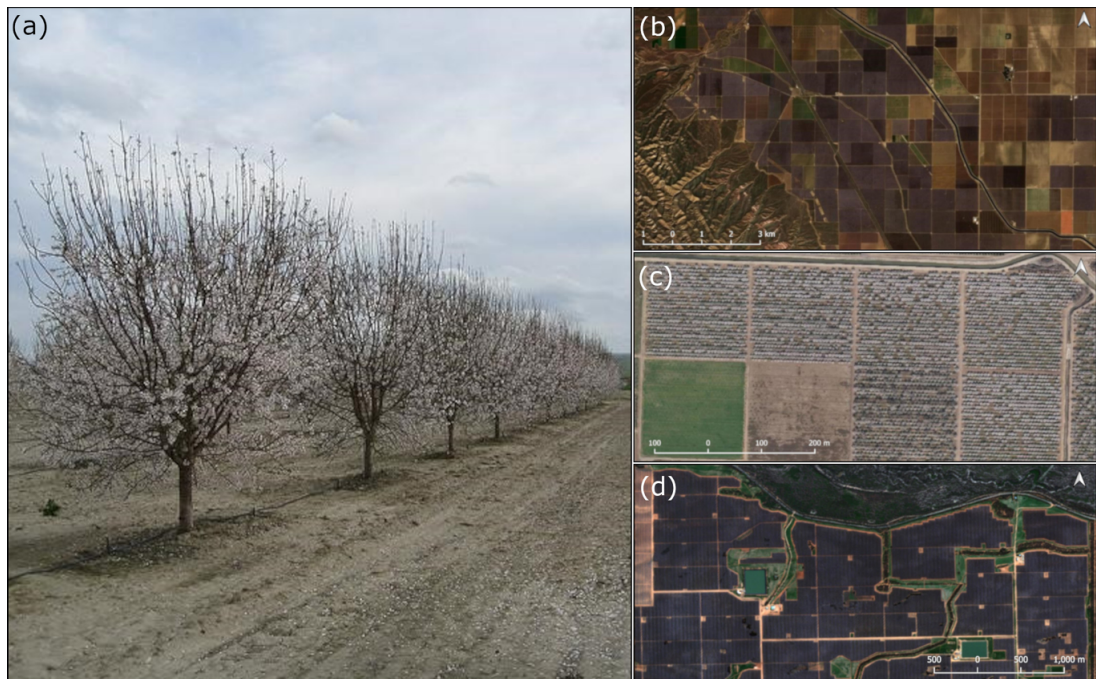
Figure 5. *(a) A blooming almond orchard in California, USA. Photo: (Columbia Climate School, 2018), (b) blooming almond orchards in California (gray-brown color shades), USA. Sentinel-2 image, true colors. 16.02.2018, (c) blooming almond orchards in California, USA. Bing Maps, (d) blooming almond orchards in Victoria, Australia (dark violet and light violet color shades). Sentinel-2 image, true colors. 21.08.2022.*

above was also applied to the visual identification of almond orchards in Australia—the leading country in commercial almond production in the Southern Hemisphere (Austrade, 2022). This location was selected arbitrarily as an area for which the information about the exact locations of the several largest almond farms in the area, as well as photos of the orchards and dates of the approximate bloom period, was publicly available (Almond Board, 2023). Finding out the approximate timing of almond bloom in Australia (from late July to early September) made it possible to determine the area occupied by its orchards near the known places of its cultivation. The "striped" texture of the objects, which were identified in the image as crop plantings (Figure 5d), also helps to conclude that these objects are orchards. In general, the texture of the orchards allows easy manual and automated (with Machine Learning algorithms) identification on satellite imagery.

2.7. Visual identification of winter barley fields

Our choice of winter crop type was winter barley: we analyzed the dynamics of harvested areas of winter barley during the 2018–2019 season in Krasnodar Krai, Russia. According to the dates of publication of articles in local online newspapers, the stages of the beginning, middle, and end of the harvest, as well as the area of the harvested areas on the corresponding dates, were quite clearly traced. For example, on 10.06.2019, the "VK Press news" agency announced the upcoming beginning of the barley harvest (Bezruk, 2019), and on 28.06.2019, an article was published in "Yug Times", which announced the completion of the barley harvest at 100% of the sown area (Yug Times, 2019), that is, 136,900 ha (see Table 1).

Figure 6a shows the beginning and Figure 6b shows the end of barley harvesting in the Krasnodar Krai in the second and third decades of June.

Table 1. *Dates of publications on the progress of the harvest in the Krasnodar Krai and the percentage of harvested fields of winter barley and wheat for a given date (% of the total area of crop fields and hectares)*

Date	Winter barley	Winter wheat
10.06.2019	0%	–
13.06.2019	5% (7,300 ha)[a]	–
17.06.2019	44% (57,000 ha)[b]	0%
20.06.2019	81% (113,000 ha)[c]	10% (140,000 ha)
28.06.2019	100% (136,900 ha)[d]	29% (426,000 ha)

[a]https://agro-bursa.ru/gazeta/obzor-rynka-zerna/2019/06/17/obzor-rynka-zerna-pshenicy-muki.html.
[b]https://kubnews.ru/obshchestvo/2019/06/17/ozimyy-yachmen-ubirayut-vo-vsekh-munitsipalitetakh-krasnodarskogo-kraya/.
[c]https://sovecon.ru/blog/2019/06/20/v_krasnodarskom_krae_pshenitsu_uberut_s_ploschadi_poryadka_1_5_mln_gektarov_ministerstvo_selskogo_
hozyajstva__i_pererabatyvajuschej_promyshlennosti_krasnodarskogo_kraya/.
[d]https://yugtimes.com/news/49278/.

To improve the efficiency of the considered approach, more frequent high-resolution time-series images are required, for example, on a daily basis. To increase the reliability of field detection, it is also necessary to have additional data on all varieties of barley grown in a given territory and the differences in their vegetation (growing degree days, precipitation requirements, etc.).

2.8. *Additional steps. Auxiliary calculated vegetation indices*

Often, plantings of the crop of interest need to be distinguished from the more important and popular crops. If it is a perennial crop, in order to distinguish it from annual crops it is necessary to determine the age of planting. In this case, there are often not enough single images, and it becomes necessary to detect fields using additional methods based on long time series of satellite imagery. Vegetation index analysis is one of the most commonly used approaches to solve the problem of distinguishing between crops.

When comparing the curves of vegetation indices, it is also necessary to take into account the ecological growing conditions of each crop and its individual varieties in a given territory, which is manifested in the difference in the dates of the onset of each subsequent phase of plant development, causing differences in the time of the beginning and end of the harvest. This data must be compared with statistical information, as well as with the percentage of the area occupied by the crop of interest versus the entire area of the region and "secondary" crop in the area (e.g., winter barley vs winter wheat), as well as the dynamics of the area of fields in different years. As an example, the average per field NDVI values, derived from Sentinel-2 images, were calculated for several almond orchards–for bearing (Figure 7a) and non-bearing orchards (Figure 7b).

The graph in Figure 7a has a unique shape for bearing almond orchards, and the segments of the graph, in which the index has maximum and minimum absolute values, represent the dates of the onset of a particular annual phenological phase (for example, dormancy from November to February or blooming from February to mid-March).

If we look at the graph in Figure 7b, we can see that the maximum annual absolute values of NDVI on this orchard tend to increase each year from the smallest values (0.2–0.3) to about 0.7, which means that a tree perennial crop is growing in this field, the total biomass of trees tends to increase every year, which would be significantly different from the NDVI curves for fields of annual crops, on which there is a constant crop rotation, with an annual change in the shape of the curves. In general, to solve the problem of detecting fields of perennial crops, an important step is to search for areas where such changes have occurred, based on a clear change in the color of the field in the spring-summer period, which can occur due to the regular change of annual crops in the process of crop rotation or complete plowing of the field and planting of perennial crops. To distinguish the field of perennial crops from fields of similar crops, it is

Figure 6. *(a) Sentinel-2 satellite image, true colors. 15.06.2019. The harvesting of winter barley in the region started on 10.06-11.06.2019, the barley fields are in light and orange-brown color shades. Ripening wheat fields are shown in green-brown shades. Krasnodar Krai, Russia; (b) Sentinel-2 image, true colours. 28.06.2019. The harvesting of winter barley in the region is officially over and the fields are being prepared for the next season (left part of the image). Wheat harvesting continues, its fields have gray-brown shades. Krasnodar Krai, Russia.*

useful to use satellite images with a large number of spectral channels. For the detection of a specific annual crop growth stage or a phenological phase, more specialized indices are often used that can characterize the phenology of different crops. For example, Chen, Jin, and Brown used the enhanced bloom index to study the phenology of almonds in California (Chen et al., 2019b).

3. Description of the dataset

The results of the proposed approach can be used on a large scale (e.g., an administrative unit or a country) as a set of training samples for one of the algorithms that use ML methods for automatic identification of the fields. Neural networks and, specifically, semantic segmentation methods are widely used for the task of automatic classification of satellite imagery (Mou et al., 2017; Hong et al., 2022; He et al., 2022; Vaze

Table 2. *Semantic segmentation results for lavender fields*

Model	Depth	Filters	Learning rate	Dropout rate	Test DC
Random forest	—	—	—	—	0.9271
U-Net	5	32	0.0005	0	0.9522
UNet++	4	32	0.0005	0.3	0.8283
ModSegNet	4	64	0.001	0.3	0.8476
ResUnet	4	16	0.001	0.3	**0.9604**

Table 3. *Semantic segmentation results for almond fields*

Model	Depth	Filters	Learning rate	Dropout rate	Test DC
Random forest	—	—	—	—	0.9114
U-Net	5	32	0.001	0	0.9487
UNet++	4	32	0.0005	0.3	0.8306
ModSegNet	4	64	0.001	0.3	0.8308
ResUnet	4	16	0.0005	0	**0.9549**

et al., 2020; Cai et al., 2018). However, as we mentioned earlier, gathering a sufficient amount of data for image classification with neural networks is a time-consuming and laborious process. The method, proposed in this paper, allows fast and easy data collection for such tasks. To demonstrate the high performance of the proposed approaches we attempted to identify and map two industrial crops—lavender and almond in different geographic regions.

First, we visually identified all lavender fields in Bulgaria's leading lavender cultivation region, Dobrich Province, in 2019. We mapped 7,734 hectares of lavender fields using cloud-free Sentinel-2 satellite images on two summer dates (26.06.2019 and 02.07.2019) on the basis of a "true color" band combination. We also checked the validity of the identified fields using Google, Esri, and Bing base maps. According to the publications, it was officially declared that the area occupied by lavender in Dobrich Province in 2019 was 7,223 ha (Agri.bg, 2019).

Second, we checked our approach for the identification of almond plantings in the United States. We visually identified and mapped 70,439 ha (173,943 acres) of almond orchards in the leading region by almond acreage in the USA—Kern County, CA—as of the 2018 year using cloud-free Sentinel-2 images on two dates (21.02.2018 and 08.03.2018). We checked the validity of the identified fields using Google, Esri, and Bing base maps. We compared the results of identification with the officially reported estimate, which was 65,269 ha (USDA, 2019) and with the USDA Cropland Data Layer for 2018. While the Cropland Data Layer is claimed to have very high user accuracy of identification of almond fields for California—97.2% (USDA, 2018), the proposed approach may provide slightly less accuracy, but it is based entirely on data that is publicly available.

The accuracy of the visual identification can be significantly increased if we use a set of very high-resolution images (e.g., WorldView-3) on a small patch of land with relatively uniform geographic conditions.

The collected dataset consists of two parts. The first part includes the vector GIS files, containing all lavender fields in Dobrich Province (Bulgaria) and in Valensole Commune (Provence, France) as well as all almond fields in Kern County (USA) and the fields on several large farms in Victoria (Australia) respectively. The second part consists of the fragments of Sentinel-2 images and binary masks of lavender and almond fields in the regions mentioned above.

After collecting the data, we conducted experiments with the automatic identification of lavender and almonds for other geographic regions: France and Australia respectively. For this task, we developed several ML models using the data obtained as the result of visual identification of fields (see Section 4.1).
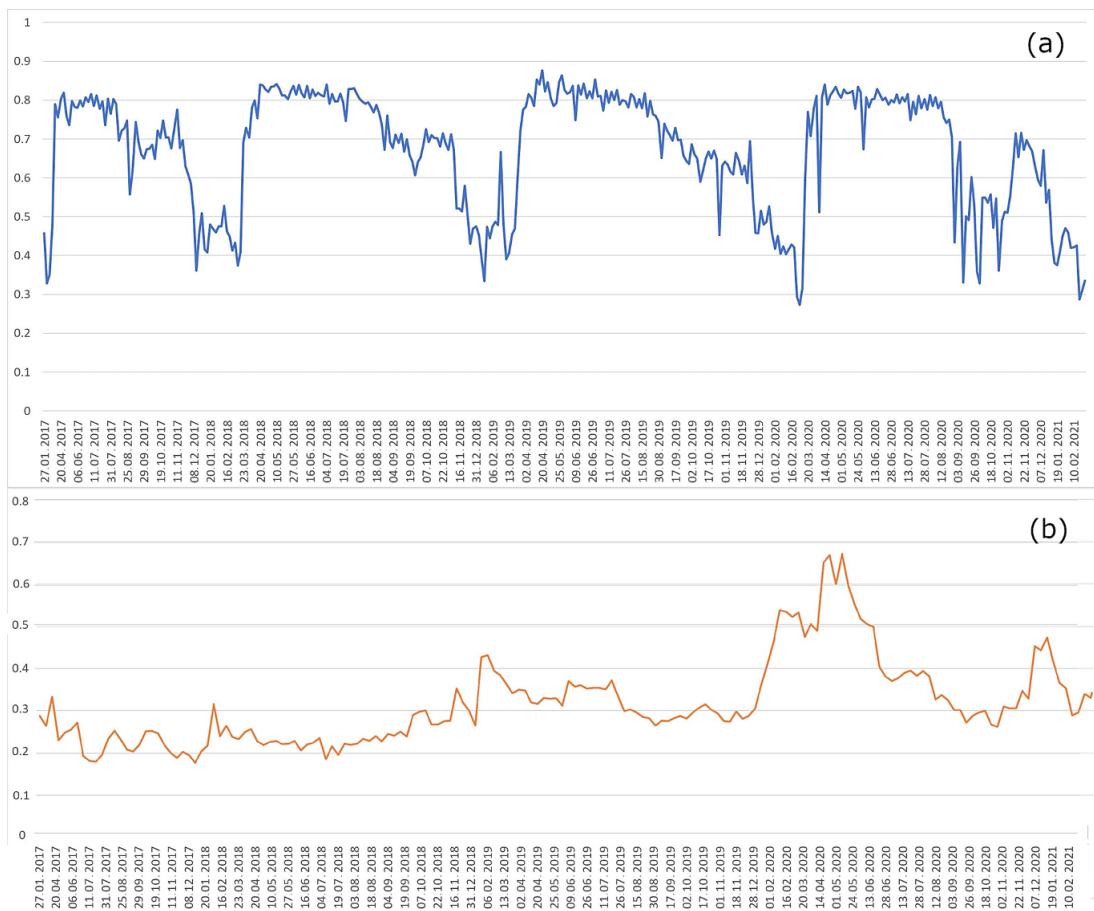
Figure 7. *(a) NDVI for bearing almonds' orchard for the 2017–2020 period in California (source: EO Browser Sentinel-Hub), (b) NDVI for non-bearing almonds' orchard in California for the 2017–2020 period (source: EO Browser Sentinel-Hub). The NDVI values for each date were calculated as the average of the index values for the entire field.*

4. Application and results

4.1. Semantic segmentation of lavender fields

This section briefly discusses the methods we used for automatic field segmentation and provides a comparison of these frameworks for the lavender and almond segmentation tasks.

We used the approach, proposed in Section 2.5 for ground-truth data collection of lavender field data. We downloaded the Sentinel-2 satellite images for the flowering period (June-July) over the entire province as described above. This imagery was used to create masks for the training and testing datasets (Figure 8A). We explored a variety of state-of-the-art models to perform a semantic segmentation task on the resulting dataset.

As the inputs of the U-Net model, we used 12 bands of Sentinel images, complemented with manually calculated vegetation indexes (in this case, we used the NDVI index, added as an additional layer to the input data). As a baseline method, we used Random Forest (Breiman, 2001). We explored the majority of the state-of-the-art semantic segmentation methods: U-Net (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018), ResUnet (Zhang and Liu, 2017), and ModSegNet (Barros et al., 2022) models. We explored the range of parameters, including learning rates, number of filters, and dropout rates. Our experiments
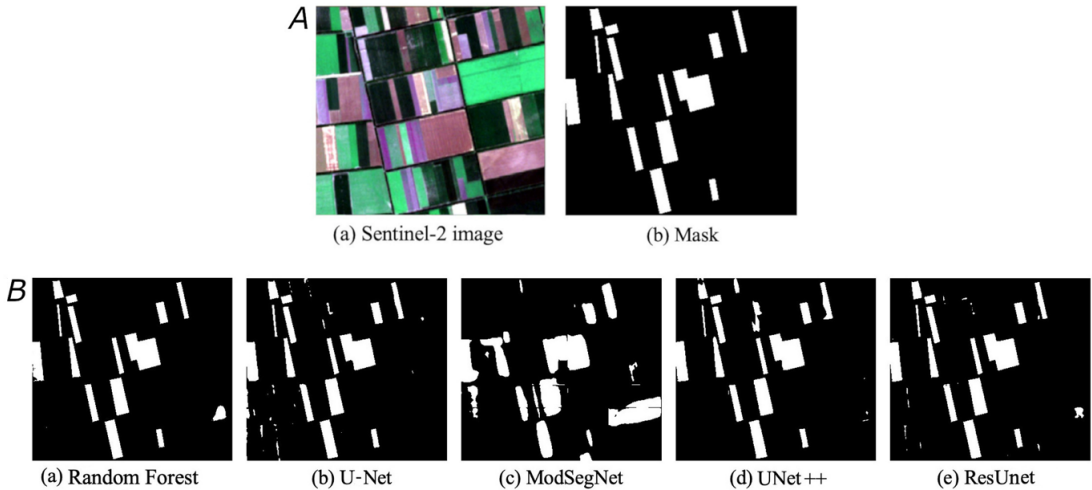
Figure 8. A - Sample data, prepared for the task of semantic segmentation of lavender fields in Bulgaria. Source: Sentinel-2. (a) original Sentinel-2 12-band stack image (true colors), (b) binary segmentation mask. B - Results of lavender segmentation in Bulgaria. (a) Random Forest, (b) U-Net, (c) ModSegNet, (d) UNet++, (e) ResUnet.

included various U-Net model hyperparameters: 16 or 32 filters, and various dropout rates. ResUnet uses key features of both U-Net and ResNet models, where deep U-shape architecture combines with residual connections, preventing the vanishing and exploding of the gradient. We tested several models with 16 or 32 filter numbers, 0.0005–0.001 learning rates, and a range of dropouts. While training U-Net models, we tested different numbers of filters (16/32), and dropout rates (0/0.3). Larger dropout or larger number of filters did not always provided higher dice scores. In ResUnet architectures, we also did not find the correlation between dice and the number of filters (16/32), or dropout (0/0.3) (Table 2). The results demonstrate that the differences between the applied models' performance are relatively small. Visual assessment of models' performance showed that many of the tested models can be used for land cover segmentation. The best-performing model for lavender segmentation with Dice coefficient 0.9604 was ResUnet on Sentinel-2A product.

The Dice coefficient is defined as

$$DC = \frac{2|A \cap B|}{|A| + |B|}, \tag{1}$$

where A is a set representing the ground truth and B represents the computed segmentation (Crum et al., 2006). Both images (sets) are binary with values '0' or '1' at each of their pixels. In segmentation tasks, Dice Coeff (Dice loss = $1 -$ Dice Coeff) is used as a Loss function.

The segmentation results are displayed in Figure 8B.

4.2. Semantic segmentation of almond orchards

We used the approach, proposed in Section 2.6 for ground-truth data collection of almond orchard data. We downloaded Sentinel-2 satellite imagery for the bloom period (February–March) over the entire county. These images were used to create masks for the training and testing datasets (Figure 9A). For the almond orchard detection task, we used the models from the previous task. The best-performing models were U-Net and ResUnet. Similarly to the lavender segmentation task, we explored a range of parameters, including learning rates, dropout rates, the number of downsampling–upsampling stages, and various depths of U-Net models. The best-performing model for almond segmentation with Dice coefficient of 0.9549 was ResUnet and U-Net with Dice coefficient of 0.9487 (Figure 9B).
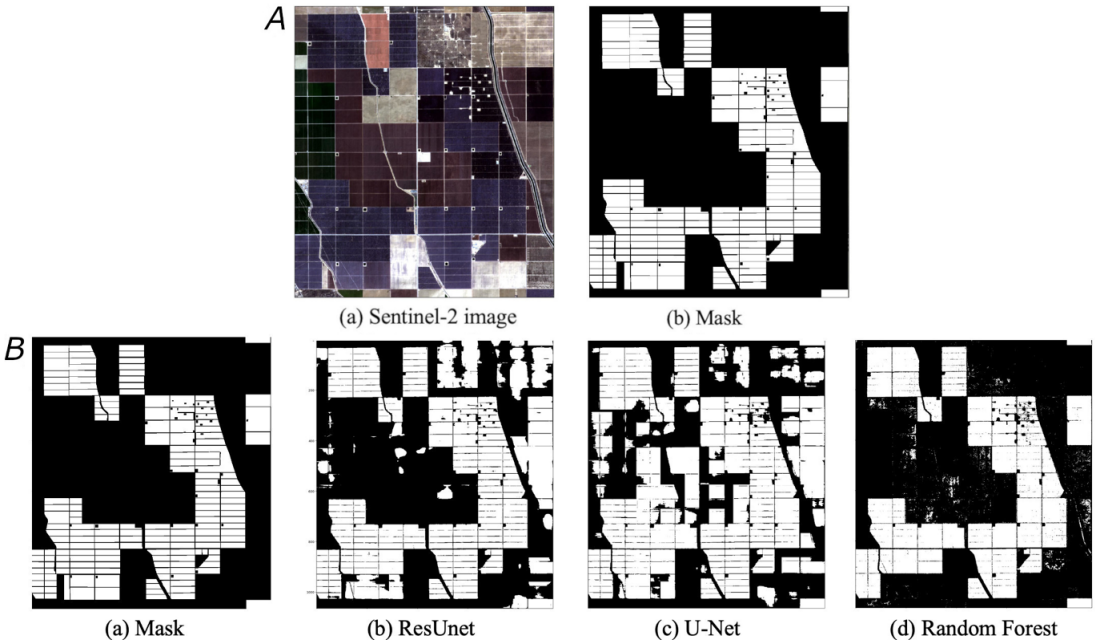
Figure 9. A—Sample data, prepared for the task of semantic segmentation of almond orchards in the United States. Source: Sentinel-2. (a) Original Sentinel-2 12-band stack image (true colors). (b) Binary segmentation mask. B—Almond fields segmentation task. We use random forest model (92% accuracy) as a baseline model. We further compare semantic segmentation methods with the baseline model: ResUnet, U-Net.

4.3. Transfer learning

We further evaluated the performance of pre-trained models on our dataset. We compared the performance of the random weight initialized model, pre-trained ResUnet model, and our adversarially trained ResUnet model on our dataset. Each model was trained for 30 epochs. The learning rate was set to 0.00001 for transfer learning experiments. We also compared the results to the predictions of the pre-trained models without fine-tuning. We provide the results for the best-performing models for almond fields in Table 4.

We conducted a few experiments with pre-trained models. In the first experiment, we initialized the model with the ResUnet model with ResNet-50 backbone, pre-trained on the ImageNet data (Deng et al., 2009). In the second experiment, we used the same model, pre-trained on the SeCo dataset (Manas et al., 2021). Finally, we pre-trained the same model on a single crop on an Australian dataset. This model was trained on a large number of images (over 60000 patches) of another crop in South-East Australia.

Our transfer learning experiments show there is a limited gain from using transferred models. The best performance increase after fine-tuning resulted in a gain of 1% accuracy (from 0.9549 to 0.9684), with ResUnet model, pre-trained on the Australian dataset. One explanation is that the pre-trained model must "unlearn" the priors that were only applicable to the Australian agricultural landscape and a different type of crop. The visual inspection of the results shows better detection of the field boundaries themselves with

Table 4. Transfer learning experiments

Pre-training	Backbone	Crop	Test DC
ImageNet	ResNet–50	Almonds	0.8544
SeCo	ResNet–50	Almonds	0.8939
Australian	ResNet–50	Almonds	**0.9684**

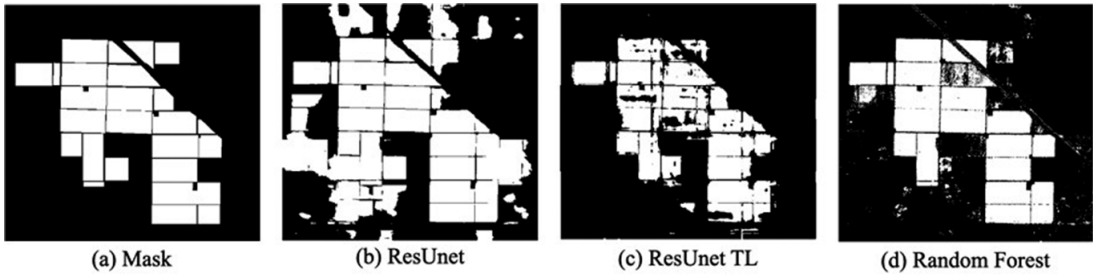(a) Mask   (b) ResUnet   (c) ResUnet TL   (d) Random Forest

Figure 10. *Almond fields segmentation task with transfer learning. The best-performing model is pre-trained on binary crop segmentation tasks in another region.*

more false positives (Figure 10). Regardless, in general, the models, pre-trained on the Sentinel-2 imagery, outperform the model trained from random weight initialization, indicating that some degree of transfer learning was successful.

4.4. Additional considerations

The approach presented above is crop-specific and territory-specific. Multiple factors contribute to this restriction, including the size of the fields, availability of very high-resolution imagery, cloud cover, and difficulties in determining the flowering window and capturing the dates of sowing and harvest. It is also useful to perform primary crop categorization according to the difficulty of their detection: crops that are simple to detect (popular crops, large area and amount of data, and frequent field surveys); crops with medium detection difficulty (large set of statistical data and a small choice of available satellite images for key dates and vice versa); and crops that are difficult to detect (small size of fields, annual crops, lack of field and statistical data, and a small set of available images for the region of interest). In this regard, a very important stage is the collection of high-quality, significant, and reliable statistical information on the location and area of fields, including online information and data in different languages, as well as the organization of a convenient format for the exchange of such information between local residents and those conducting the research. It is also necessary to perform a primary analysis of the possibility of mapping crop fields based on very high-resolution images which are used by the most popular web mapping platforms.
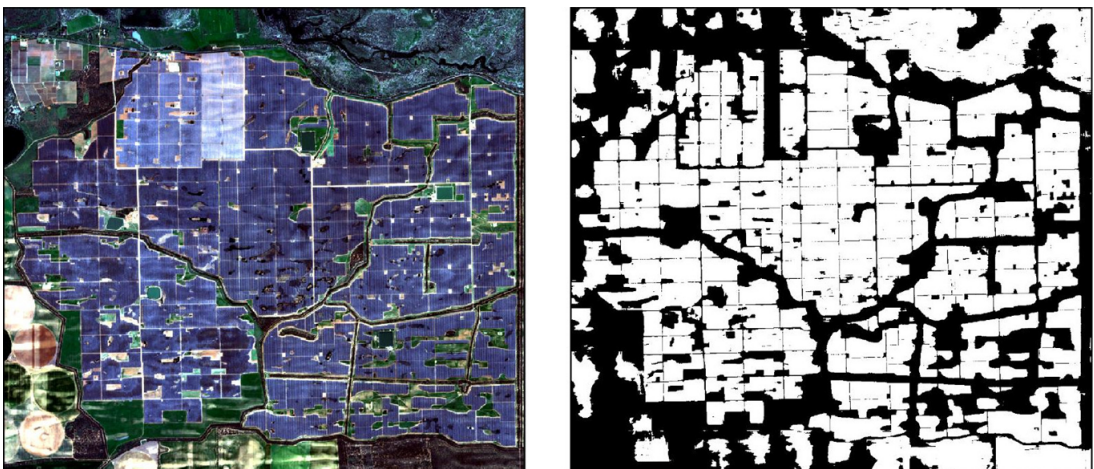

Figure 11. *Results of the pre-trained model on a new region: ResUnet model results in Australia. Left: original Sentinel-2 12-band stack image (true colors); right: the results of semantic segmentation.*

For both lavender and almond orchards, we have performed an experiment, where we used a pre-trained best-performing model developed using lavender and almond plantings for Bulgaria and the USA, on the other region (France for lavender and Australia for almonds, respectively). From the literature, we know that this approach is not readily applicable in the majority of cases: pre-trained models are typically fine-tuned on the local labels to achieve the best performance in the new region (Kerner et al., 2020; Wang et al., 2022). Similarly to the research conducted in the literature, we did not find promising results while running the lavender model on the France region, potentially due to the differences in atmospheric properties that change the color of lavender fields in satellite images. However, the almond classification performed reasonably well in Australia (potentially due to the similar climatic conditions between the two regions of choice: Figure 11). This result can be improved by fine-tuning the model on the small amount of the data from the new region (Wang et al., 2022).

4.5. Directions for further research and open questions

In this paper, we discussed approaches providing essential information for accurately identifying various crop fields. However, these methods fall short in detecting specific crops of interest. Additional strategies include capturing images more frequently (e.g., daily) to determine phenological phases' timing and selecting cloud-free images (Pan et al., 2015). Simultaneously using satellite images with different spectral ranges enhances identification accuracy (Haack et al., 2000; Liu et al., 2014).

Accurate data on seasonal plant development, considering environmental conditions, aids in distinguishing field, and planting characteristics (Wardlow et al., 2007; Qiu et al., 2017; Yang et al., 2020) are some of the possible directions for future work. Searching for photos and videos aids in identifying specific crops in an area, especially for perennial crops (Wu et al., 2021). Collaborating with organizations, exchanging information with local stakeholders, and utilizing statistical data contribute to accurate crop mapping (Laso Bayas et al., 2017; Skakun et al., 2017; Zipper, 2017).

To help identify fields and plantings in areas with similar conditions, one might consider applying similar approaches to comparable regions and creating libraries of spectral and phenological curves, using data alongside hyperspectral imagery (Qiu et al., 2017; Skakun et al., 2017; Nidamanuri et al., 2007). A field verification stage assesses overall detection and identification reliability.

Specific questions regarding improving crop mapping accuracy on cloudy days, the efficiency of radar data versus optical data (Haack et al., 2000; Skakun et al., 2016), and the real-time identification and yield estimation of less popular annual crops (Gao and Zhang, 2021; Vicente-Guijalba et al., 2015) are raised. Additionally, addressing the absence of scientific and statistical information on a territory and the lack of up-to-date high-resolution images requires further exploration (Wang et al., 2019; Lin et al., 2022).

The rapid development of foundation models for remote sensing (Jakubik et al., 2023; Tuia et al., 2023) brings a new lens into the research in crop mapping. Although these models are typically able to classify only major crops, such as soybean, corn, maize, etc., they can be fine-tuned to achieve better results on regional data and rare crops. Foundation models are an important emerging class of ML models, which use very large ML models, are trained with extremely large and broad data sets, and require considerable computing resources during training. For example, in Jakubik et al. (2023), the authors built a large-scale foundation model using satellite and remote sensing data, including the Harmonized Landsat and Sentinel-2 (HLS) data. This model was fine-tuned on a range of Earth observation tasks involving multi-temporal cloud gap imputation, flood mapping, wildfire scar segmentation, and multi-temporal crop segmentation. Such fine-tuning can be performed for various downstream tasks with small labeled datasets. The methodology, presented in this paper, can be used to gather the data for fine-tuning foundation models in the future.

5. Conclusion

The paper describes several methodological approaches to collecting the data and mapping the fields of crops from satellite images using phenological, geographic, ecological, and statistical information in

addition to the traditional approaches based on vegetation indices. In the example of three crops (lavender, almond, and winter barley), it was shown that the color in the satellite image is an important feature that allows the identification of fields of the specific crop among other types of land cover, in particular in areas where similar and more popular crops are cultivated. Such data as statistical information (i.e., the area occupied by the crop fields or sown area in different years), the growing season, and the start and end dates of the harvest are of great importance for crop mapping. When used together with the color of the fields, they make it possible to visually distinguish the fields of the mapped crop from other vegetation cover types and fields of similar crops, and, in addition, they help to validate the accuracy of the mapping. To support our methodological considerations, we identified and mapped all lavender fields using satellite imagery in Bulgaria's leading lavender cultivation region—Dobrich Province, and all almond orchards in one of the leading almond cultivation regions of the USA—Kern County, California.

Based on this data, we developed several advanced ML segmentation models and applied them for mapping lavender and almond crops in other regions, France and Australia. The obtained data about the color changes of the mapped crop fields during the growing season provides valuable information for searching the crop of interest in areas with similar geographic or environmental conditions, but different in terms the availability of geographic data related to the mapping of the interested crop. We discovered that information about field color, statistical, geographic and environmental conditions, and long time-series data, such as curves of vegetation indices (e.g., seasonal and/or long-term changes in NDVI values), when used together, can significantly improve crop mapping accuracy; in addition, we identified a variety of other useful directions, approaches, and questions that can further improve the process of mapping the fields of the crop of interest. Conducted experiments demonstrate the principle feasibility of identifying and detecting represented crops visually and using ML algorithms for different regions of the world.

The methodology, presented in this paper, can be used for a variety of tasks, such as collecting fine-grained data for training small and medium-scale ML models, including all types of semantic segmentation tasks, and for fine-tuning large geospatial foundation models.

References

Agri.bg (2019) Lavandula 2019: Poveche ploshti i po-visoki dobivi [lavender 2019: more areas and higher yields]. Available at: https://agri.bg/novini/lavandula-2019-poveche-ploshti-i-po-visoki-dobivi; last accessed 21 November 2023.

Almond Board (2023) Almond insights 2022/2023. Available at: https://viewer.joomag.com/aba-almond-insights-22-23/0921304001697669494?short&; last accessed 21 November 2023.

Austrade (2022) Almonds: The little nut that cracked the big export game. Available at: https://www.austrade.gov.au/en/news-and-analysis/news/almonds-the-little-nut-that-cracked-the-big-export-game#:~:text=There%20are%20over%2017%20million,a%20hit%20around%20the%20world.; last accessed 21 November 2023.

Barros T, **Conde P**, **Gonçalves G**, **Premebida C**, **Monteiro M**, **Ferreira CSS and Nunes UJ** (2022) Multispectral vineyard segmentation: a deep learning comparison study. *Computers and Electronics in Agriculture*, *195*:106782.

Belgiu M and Csillik O (2018) Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing Environment 204*:509–523.

Bezruk F (2019) Urozhay-2019: Kuban' nachinayet uborku zernovykh kolosovykh kul'tur [harvest-2019: Kuban begins harvesting grain crops]. Available at: https://web.archive.org/web/20210621195620/https://www.vkpress.ru/life/urozhay-2019-kuban-nachinaet-uborku-zernovykh-kolosovykh-kultur/?id=125655; last accessed 21 November 2023.

Bobrowski T (n.d.) When lavender bloom: Provence. Available at: https://frenchriviera.travel/when-lavender-bloom-provence/; last accessed 21 November 2023.

Breiman L (2001) Random forests. *Machine Learning*, 5–32.

Brinkhoff J, **Vardanega J and Robson A J** (2020) Land cover classification of nine perennial crops using sentinel-1 and -2 data. *Remote Sensing 12*(1).

Cai Y, **Guan K**, **Peng J**, **Wang S**, **Seifert C**, **Wardlow B and Li Z** (2018) A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach. *Remote Sensing of Environment 210*: 35–47.

Chen B, **Jin Y and Brown P** (2019a) Automatic mapping of planting year for tree crops with landsat satellite time series stacks. *ISPRS Journal of Photogrammetry and Remote Sensing 151*:176–188.

Chen B, **Jin Y and Brown P** (2019b) An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS Journal of Photogrammetry and Remote Sensing 156*:108–120.

Columbia Climate School (2018) Climate, earth, and society. Available at: https://news.climate.columbia.edu/2018/09/05/almonds-california-healthy-hazardous/; last accessed 21 November 2023.

Crum W, **Camara O and Hill D** (2006) Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging 25*:1451–1461.

Deng J, **Dong W**, **Socher R**, **Li L-J**, **Li K and Fei-Fei L** (2009) Imagenet: a large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. pp. 248–255.

Gao F and Zhang X (2021) Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *Journal of Remote Sensing 2021*:1–14.

Gur'eva ZI, **Petrov KM and Sharkov VV** (1976) *Aerofotometody geologo-geomorfologicheskogo issledovaniya vnutrennego shel'fa i beregov morey. Atlas annotirovannykh aerofotosnimkov [Geological and geomorphological photointerpretation of inner shelf and sea coasts. Collection of annotated aerial photographs]*. Nedra, Leningrad.

Haack B, **Herold N and Bechdol M** (2000) Radar and optical data integration for land-use/land-cover mapping. *Photogrammetric Engineering and Remote Sensing*, 66.

He X, **Zhou Y**, **Zhao J**, **Zhang D**, **Yao R and Xue Y** (2022) Swin transformer embedding unet for remote sensing image semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15.

Hong D, **Han Z**, **Yao J**, **Gao L**, **Zhang B**, **Plaza A and Chanussot J** (2022) Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing 60*:1–15.

InteliAgro (2019) Proizvodstvo na eterichno-masleni kulturi i masla [production of essential oil crops and oils]. Available at: http://inteliagro.bg/sites/default/files/free_files/Essential_Oils_Market_2018_BG_revised_0.pdf; last accessed 21 November 2023.

Jakubik J, **Roy S**, **Phillips C**, **Fraccaro P**, **Godwin D**, **Zadrozny B**, **Szwarcman D**, **Gomes C**, **Nyirjesy G**, **Edwards B** et al (2023) Foundation models for generalist geospatial artificial intelligence. *arXiv preprint arXiv:2310.18660*.

Jensen JR (2007) *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River NJ: Prentice-Hall.

John A, **Ong J**, **Theobald EJ**, **Olden JD**, **Tan A and HilleRisLambers J** (2020) Detecting montane flowering phenology with cubesat imagery. *Remote Sensing 12*(18).

Kerner H, **Tseng G**, **Becker-Reshef I**, **Nakalembe C**, **Barker B**, **Munshell B**, **Paliyam M and Hosseini M** (2020) Rapid response crop maps in data sparse regions. *KDD '20 Humanitarian Mapping Workshop, August 24, 2020, San Diego, CA*.

Labutina IA (2004) *Deshifrirovaniye aerokosmicheskikh snimkov [Interpretation of Satellite Images]*. Aspekt Press, Moscow.

Laso Bayas JC, **See L**, **Perger C**, **Justice C**, **Nakalembe C**, **Dempewolf J and Fritz S** (2017) Validation of automatically generated global and regional cropland data sets: the case of tanzania. *Remote Sensing 9*(8):1–23.

Lillesand TM, **Kiefer RW and Chipman JW** (2015) *Remote Sensing and Image Interpretation*. New York: Wiley.

Lin C, **Zhong L**, **Song X-P**, **Dong J**, **Lobell DB and Jin Z** (2022) Early- and in-season crop type mapping without current-year ground truth: generating labels from historical information via a topology-based approach. *Remote Sensing of Environment 274*: 112994.

Liu M W, **Ozdogan M and Zhu X** (2014) Crop type classification by simultaneous use of satellite images of different resolutions. *IEEE Transactions on Geoscience and Remote Sensing 52*(6):3637–3649.

Manas O, **Lacoste A**, **Giro-i Nieto X**, **Vazquez D and Rodriguez P** (2021) Seasonal contrast: unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9414–9423.

Masialeti I, **Egbert S and Wardlow B** (2010) A comparative analysis of phenological curves for major crops in Kansas. *GISci Remote and Sensing 47*:241–259.

Mou L, **Ghamisi P and Zhu X X** (2017) Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing 55*:3639–3655.

Nedkova G (2021) Shte otrovi li aromatnata lavandula dobrich? [will the fragrant lavender poison dobrich?]. Available at: https://agri.bg/novini/shche-otrovi-li-aromatnata-lavandula-dobrich; last accessed 21 November 2023.

Nidamanuri, **R Garg P and Ghosh S** (2007) Development of an agricultural crops spectral library and classification of crops at cultivar level using hyperspectral data. *Precision Agriculture 8*:173–185.

Pan Z, **Huang J**, **Zhou Q**, **Wang L**, **Cheng Y**, **Zhang H**, **Blackburn G A**, **Yan J and Liu J** (2015) Mapping crop phenology using ndvi time-series derived from hj-1 a/b data. *International Journal of Applied Earth Observation and Geoinformation 34*: 188–197.

Potgieter A, **Apan A**, **Dunn P and Hammer G** (2007) Estimating crop area using seasonal time series of enhanced vegetation index from modis satellite imagery. *Australian Journal of Agricultural Research 58*:316–325.

Qiu B, **Luo Y**, **Tang Z**, **Chen C**, **Lu D**, **Huang H**, **Chen Y**, **Chen N and Xu W** (2017) Winter wheat mapping combining variations before and after estimated heading dates. *ISPRS Journal of Photogrammetry and Remote Sensing 123*:35–46.

Ronneberger O, **Fischer P and Brox T** (2015) U-net: convolutional networks for biomedical image segmentation. In Navab N, Hornegger J, Wells WM, and Frangi AF, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham: Springer International Publishing, pp. 234–241.

Roumenina E, **Atzberger C**, **Vassilev V**, **Dimitrov P**, **Kamenova I**, **Banov M**, **Filchev L and Jelev G** (2015) Single- and multi-date crop identification using proba-v 100 and 300 m s1 products on zlatia test site, bulgaria. *Remote Sensing 7*:13843–13862.

Skakun S, **Franch B**, **Vermote E**, **Roger J-C**, **Becker-Reshef I**, **Justice C and Kussul N** (2017) Early season large-area winter crop mapping using modis ndvi data, growing degree days information and a gaussian mixture model. *Remote Sensing of Environment 195*:244–258.

Skakun S, **Kussul N**, **Shelestov A Y**, **Lavreniuk M and Kussul O** (2016) Efficiency assessment of multitemporal c-band radarsat-2 intensity and landsat-8 surface reflectance satellite imagery for crop classification in ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9*:3712–3719.

Tuia D, **Schindler K**, **Demir B**, **Camps-Valls G**, **Zhu X X**, **Kochupillai M**, **Džeroski S**, **van Rijn J N**, **Hoos H H**, **Del Frate F**, et al (2023) Artificial intelligence to advance earth observation: a perspective. *arXiv preprint arXiv:2305.08413*.

USDA, N. A. S. S. (2018) 2018 California cropland data layer — nass/usda. Available at: https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/metadata_ca18.htm; last accessed 23 January 2024.

USDA, N. A. S. S. (2019) 2018 California almond acreage report. Available at: https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Almond/Acreage/201904almac.pdf; last accessed 21 November 2023.

USDA, N. A. S. S. (2021) 2020 California almond acreage report. Available at: https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Almond/Acreage/202104_almac.pdf; last accessed 21 November 2023.

Vaze S, **Foley CJ**, **Seddiq MEA**, **Unagaev A and Efremova N** (2020) Optimal use of multi-spectral satellite data with convolutional neural networks. *AI for Social Good Harvard CRCS Workshop.*

Vicente-Guijalba, F Martinez-Marin T and Lopez-Sanchez JM (2015) Dynamical approach for real-time monitoring of agricultural crops. *IEEE Transactions on Geoscience and Remote Sensing 53*:3278–3293.

Vylkanov N (2018) Eterichno-masleni kulturi –razmisli vurkhu opita na bulgariya [essential oil crops - reflect on the experience of Bulgaria]. Available at: http://inteliagro.bg/sites/default/files/free_files/Essential_oils_market_Dobrich_presentation.pdf; last accessed 21 November 2023.

Wang S, **Azzari G and Lobell D** (2019) Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment 222*:303–317.

Wang S, **Waldner F and Lobell DB** (2022) Unlocking large-scale crop field delineation in smallholder farming systems with transfer learning and weak supervision. *Remote Sensing 14*:5738.

Wardlow B, **Egbert S, and Kastens J** (2007) Analysis of time-series modis 250 m vegetation index data for crop classification in the U.S. central great plains. *Remote Sensing of Environment 108*:290–310.

Wu F, **Wu B**, **Zhang M**, **Zeng H and Tian F** (2021) Identification of crop type in crowdsourced road view photos with deep convolutional neural network. *Sensors 21*:1165.

Yang Y, **Ren W**, **Tao B**, **Ji L**, **Liang L**, **Ruane A C**, **Fisher J B**, **Liu J**, **Sama M**, **Li Z and Tian Q** (2020) Characterizing spatiotemporal patterns of crop phenology across north america during 2000–2016 using satellite imagery and agricultural survey data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *170*:156–173.

Yug Times (2019) Na kubani urozhay ozimogo yachmenya vyros pochti na 15 proc [in kuban the winter barley harvest increased by almost 15 perc.]. Available at: https://yugtimescom/news/49278/; last accessed 21 November 2023.

Zhang Z and Liu Q (2017) Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*. vol. *15*, no. 5, pp. 749–753, May 2018, doi: 10.1109/LGRS.2018.2802944.

Zhong C, **Wang C and Wu C** (2015) Modis-based fractional crop mapping in the u.s. midwest with spatially constrained phenological mixture analysis. *Remote Sensing 7*:512–529.

Zhou Z, **Rahman Siddiquee M M**, **Tajbakhsh N and Liang J** (2018) 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, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, Springer, pp. 3–11.

Zipper S (2017) Agricultural research using social media data. *Agronomy Journal 110*:349–358.